DeGroot's ENDOCRINOLOGY

Basic Science and Clinical Practice

EIGHTH EDITION

Editor-in-Chief

R. Paul Robertson, MD
Emeritus Professor
Pennock Chair for Diabetes Research
Section Head, Endocrinology and Metabolism
Medicine and Cell Biology
University of Minnesota
Minneapolis, Minnesota;
Division Head, Clinical Pharmacology
Medicine and Pharmacology
University of Washington
Seattle, Washington

Associate Editors

Linda C. Giudice, MD, PhD
Distinguished Professor and The Robert B Jaffe MD Endowed Professor in the Reproductive Sciences
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco
San Francisco, California

Ashley B. Grossman, BA, BSc, MD, PhD, FRCP, FMedSci
Professor of Endocrinology
Green Templeton College
University of Oxford
Oxfordshire, United Kingdom;
Professor Neuroendocrinology
Barts and the London School of Medicine
University of London
London, United Kingdom

Gary D. Hammer, MD, PhD
Millie Schembechler Professor of Adrenal Cancer
Director - Endocrine Oncology Program
Professor of Internal Medicine – Metabolism, Endocrinology and Diabetes
Professor of Molecular and Integrative Physiology
Professor of Cell and Developmental Biology
University of Michigan
Ann Arbor, Michigan

Michael D. Jensen, MD
Distinguished Mayo Clinic Investigator
Tomas J. Watson, Jr. Professor in Honor of Dr. Robert L. Frye
Division of Endocrinology, Diabetes, Metabolism, and Nutrition
Mayo Clinic
Rochester, Minnesota

George J. Kahaly, MD, PhD
Professor of Medicine and Endocrinology/Metabolism
Department of Medicine I
Johannes Gutenberg University Medical Center
Mainz, Germany

Ronald S. Swerdloff, MD, MACP
Senior Investigator
The Lundquist Institute at Harbor-UCLA;
Distinguished Professor of Medicine
David Geffen School of Medicine at UCLA;
Chief, Division of Endocrinology
Harbor-UCLA Medical Center
Torrance, California

Rajesh V. Thakker, MD, ScD, FRCP, FMedSci, FRS
May Professor of Medicine
Academic Endocrine Unit
Radcliffe Department of Medicine
University of Oxford
Oxfordshire, United Kingdom

ELSEVIER

Elsevier
1600 John F. Kennedy Blvd.
Ste 1800
Philadelphia, PA 19103-2899

DEGROOT'S ENDOCRINOLOGY: BASIC SCIENCE
AND CLINICAL PRACTICE, EIGHTH EDITION

ISBN: 978-0-323-69412-4
Volume 1: 9780443107801
Volume 2: 9780443107818

Copyright © 2023 by Elsevier Inc. All rights reserved.

Chapter 84 by Nicole Vietor and Henry Burch, Chapter 99 by Karel Pacak, and Chapter 120 by Francesco DeMayo are in the public domain.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notice

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Previous editions copyrighted 2016, 2010, 2006, 2001, 1995, 1989, 1979.

Content Strategist: Nancy Duffy/Humayra R. Khan
Content Development Specialist: Grace Onderlinde
Publishing Services Manager: Shereen Jameel
Project Manager: Beula Christopher
Design Direction: Brian Salisbury

Printed in India

Last digit is the print number: 9 8 7 6 5 4 3 2 1

ELSEVIER
Book Aid International
Working together
to grow libraries in
developing countries
www.elsevier.com • www.bookaid.org

Table of Contents

Front Matter
Copyright
Contributors
Preface
1. Principles of Endocrinology
2. Principles of Endocrine Measurements
3. Endocrine Rhythms, the Sleep-Wake Cycle, and Biological Clocks
4. Prolactinomas and Disorders of Prolactin Secretion
5. Hypopituitarism Including Growth Hormone Deficiency
6. Acromegaly
7. Cushing Disease
Hormone–Producing Pituitary Tumors
10. Pituitary Surgery
11. Craniopharyngiomas and Parasellar Masses
12. Disorders of Sodium, Diabetes Insipidus and Hyponatremia
13. Anatomy and Physiology of the Hypothalamus and Pituitary
14. Genetic Disorders of the Pituitary
15. Aggressive Pituitary Tumors
16. Radiotherapy in Pituitary Tumors
Other Growth Factors
18. Growth Hormone Deficiency in Children
19. Basic and Clinical Appetite Regulation and Energy Expenditure
20. Obesity : isk, Risk Factors, and Medical Management
21. Genetic Syndromes Associated With Obesity
22. Regulation of Intermediary Metabolism During Fasting, Feeding, and Exercise
23. Adipose Tissue

25. Lipoprotein Metabolism and the Treatment of Lipid Disorders
26. Metabolic Syndrome
27. Bariatric Procedures and Operations
28. Development of the Endocrine Pancreas
29. Regulation of Islet Hormone Synthesis and Secretion
30. Mechanisms of Insulin Action
31. Classification and Diagnosis of Diabetes Mellitus
32. Autoimmune (Type 1) Diabetes Mellitus : Etiology,
33. Type 2 Diabetes Mellitus : Etiology, Pathogenesis, and Natural History
34. Diabetes Mellitus and Pregnancy
35. Hyperglycemia Secondary to Non-Traditional Diabetic Conditions
36. β Cell Glucose Toxicity and Oxidative Stress
37. Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State
38. Hypoglycemia and Hypoglycemic Syndromes
39. Treatment of Type 1
of Type 2 Diabetes Mellitus
41. Pancreas and Islet Transplantation
42. Diabetes and Long-Term Complications
43. Diabetic Eye Disease
44. Diabetic Neuropathy
45. Diabetic Kidney Disease
46. Diabetic Foot and Vascular Complications
47. Parathyroid Hormone and the Parathyroid Hormone Receptor Type 1 in the Regulation of Calcium and Phosphate Homeostasis and Bone Metabolism

49. Calcitonin
50. Vitamin D : From Photosynthesis, Metabolism, and Action to Clinical Applications
51. Bone Development and Remodeling
52. Regulation of Calcium Homeostasis and Genetic Disorders That Affect Calcium Metabolism
53. Genetic Disorders of Phosphate Homeostasis
54. Primary Hyperparathyroidism
55. Malignancy-Associated Hypercalcemia
56. Surgical Management of Hyperparathyroidism
Progressive Osseous Heteroplasia : Disorders Caused by GNAS Mutations that Reduce Gsα Activity
58. Genetic Defects in Vitamin D Metabolism and Action
59. Hereditary Disorders of the Skeleton
60. Osteoporosis
61. CKD-MBD Syndrome : Chronic Kidney Disease Produced Disorders in Skeletal, Vascular, and Cardiac Tissues and Mineral Metabolism
62. Disorders of
of the Thyroid
65. Thyroid-Stimulating Hormone and TSH Receptor
66. Thyroid Hormone Transporters and Metabolism
67. Thyroid Hormone Action
68. Thyroid Function Testing (Thyrotropin, Triiodothyronine, and Thyroxine)
69. Thyroid Imaging
70. Pathology of the Thyroid Gland
71. Graves Disease
72. Graves' Orbitopathy
73. Thyroiditis
74. Hypothyroidism

the Thyroid

77. Euthyroid and Hyperthyroid Nodules and Goiter

78. Differentiated Thyroid Cancer – Streamlining Diagnosis and Optimizing Management

79. Medullary Thyroid Cancer

80. Anaplastic (Undifferentiated) Thyroid Carcinoma

81. Resistance to Thyroid Hormone and Genetic Defects of the Thyroid

82. Surgical Management of Thyroid Disease

83. Nonthyroidal Illness Syndrome

and Homeostasis

86. Genetic Disorders of the Adrenal Cortex

87. Enzymes and Pathways of Human Steroidogenesis

88. Mineralocorticoids : Physiology, Metabolism, Receptors, and Resistance

89. Glucocorticoid Receptorse : Mechanisms of Action in Health and Disease

90. Adrenal Androgens, Adrenarche, and Adrenopause

91. Adrenal Pathology

92. Adrenal Gland Imaging

93. Adrenal

95. Adrenal Genomics II : Familial and Sporadic Neoplasia

96. Primary Aldosteronism

97. Adrenocorticotropic Hormone–Independent Cushing Syndrome

98. Adrenocortical Carcinoma

99. Pheochromocytoma

100. Adrenal Surgery

101. Differences of Sex Development

102. Endocrinology of Sexual Maturation and Puberty

103. Endocrinology of Pubertal Disorders

104. Transgender Healthcare

105. Gonadotropin Regulation

Diagnosis, and Treatment
107. Testosterone Deficiency: Aging and Chronic Disease
108. Gynecomastia
109. Erectile Dysfunction
110. Misuse and Abuse of Anabolic Hormones
111. Regulation of Spermatogenesis
112. Prevalence and Causes of Male Infertility
113. Clinical Management of Male Infertility
114. Male Contraception
115. Testicular Dysgenesis Syndrome and Testicular Tumors
117. The Menstrual Cycle and Disorders of Ovulation
118. Folliculogenesis, Ovulation, and Luteogenesis
119. Ovarian Hormone Synthesis
120. Estrogen and Progesterone Action
121. Contraception
122. Menopause and Perimenopause
123. Female Infertility : Evaluation and Management
124. Androgen Excess Disorders in Women
125. Endometriosis
126. Uterine Fibroids and Adenomyosis
127. Hormonal
129. Hormonal Control of Human Pregnancy and Parturition
130. Fetal and Neonatal Endocrinology
131. Hormonal Changes and Pregnancy Testing
132. Autoimmune Polyglandular Syndromes
133. Multiple Endocrine Neoplasia Type 1
134. Multiple Endocrine Neoplasia Types 2 and 3, and Medullary Thyroid Carcinoma
135. Neuroendocrine Tumor Syndromes
136. Ectopic Hormone Syndromes

1

Principles of Endocrinology

J. Larry Jameson

OUTLINE

Definition and Scope of Endocrinology, 2
Historical Perspectives, 3
Principles of Hormone Action, 6
  Hormone Biosynthesis and Secretion, 6
  Feedback Regulation, 6
  Paracrine and Autocrine Regulation, 7
  Hormonal Rhythms and Pulsatility, 7
  Hormone Transport and Degradation, 8
Hormone Action Through Receptors, 8
  Membrane Receptors, 9
  Nuclear Receptors, 10
Role of the Clinical Endocrinologist, 10
Major Unsolved Problems, 12

DEFINITION AND SCOPE OF ENDOCRINOLOGY

The term *endocrine* was coined by Starling to contrast the actions of hormones secreted internally (endocrine) with those secreted externally (exocrine) or into a lumen, such as the gastrointestinal tract.[1] This terminology continues today but makes the specialty somewhat opaque to the general public, who are more familiar with the term *hormone* and with particular disorders of the endocrine system. The term *hormone* is derived from the Greek verb *hormao,* which means "to set in motion." This phrase captures the dynamic properties of hormones and their ability to elicit a cascade of physiologic responses by acting on specific target tissues. Reminiscent of Newton's third law of motion, which states that "for every action, there is an equal and opposite reaction," hormone action is typically counteracted by physiologic responses that restore the system to equilibrium.

The major physiologic processes controlled by hormones include (1) growth and maturation, (2) intermediary metabolism, and (3) reproduction. However, the clinical specialty of endocrinology is most clearly delineated by diseases that afflict the classic glands, that is, hypothalamus, pituitary, thyroid, parathyroid, pancreatic islets, adrenal gland, testis, and ovary. In various parts of the world, additional clinical disorders, such as hypertension, nutrition, obesity, osteoporosis, and hyperlipidemia, also fall within the scope of endocrinology.

The basic science of endocrinology has evolved from studies of hormone identification and structure to a greater focus on hormone action. Concepts of receptors and intracellular signaling, as well as many aspects of transcriptional regulation, remain an essential component of the field. Endocrinology is ultimately the study of intercellular communication. In some cases, communication occurs within the same tissue, as exemplified by autocrine and paracrine actions of insulin-like growth factor-1 (IGF-1). More classically, hormones mediate communication between organs, as exemplified by the actions of parathyroid hormone (PTH) on bone or kidney. In this era of genomics, proteomics, and metabolomics, the traditional lines that separate endocrinology from other physiologic disciplines are becoming blurred. Erythropoietin is a classic hormone. Because it is produced by the kidney and regulates erythrocyte production, erythropoietin's clinical role is relevant primarily to the fields of nephrology and hematology. Similarly, blood cell–stimulating factors such as granulocyte colony-stimulating factor (G-CSF) are studied and used by hematologists and oncologists. The receptors for colony-stimulating growth factors such as G-CSF and granulocyte-macrophage colony-stimulating factor are, however, members of a superfamily that includes the growth hormone (GH) and prolactin (PRL) receptors. These receptors share similar intracellular signaling systems, including the Janus kinase (JAK)–signal transduction and activators of transcription (STAT) pathway. Growth factors with more pleomorphic functions, such as cytokines, are being investigated and used in almost every specialty.

Principles of endocrinology are readily transferable to other clinical disciplines. For example, hormones play a crucial role in blood pressure maintenance, intravascular volume regulation, and peripheral vascular resistance tone in the cardiovascular system. Angiotensin II, catecholamines, endothelins, and other vasoactive substances act via specific receptors to mediate dynamic changes in vascular tone. The heart produces hormones, such as atrial natriuretic peptide, in response to volume overload, resulting in compensatory natriuresis. The gastrointestinal tract is a remarkably rich source of peptide hormones, such as ghrelin, gastrin, glucagon-like peptide, cholecystokinin, secretin, and vasoactive intestinal peptide, among many others. Some of these factors, such as ghrelin and cholecystokinin, modulate appetite and perform local actions in the gastrointestinal tract; others, such as gastrin and secretin, act mainly in the gastrointestinal tract to induce physiologic responses to meals.

With the discovery of new hormones (e.g., PTH-related peptide, leptin, ghrelin, activin, atrial/brain natriuretic peptide, fibroblast growth factor 21, fibroblast growth factor 23), the scope of investigative and clinical endocrinology continues to expand. In addition, many areas of traditional endocrinology have been "spun off" and transformed into other disciplines. For example, although hypothalamic regulation of the pituitary gland remains a core element of endocrinology, neuroendocrinology is rapidly becoming a distinct discipline. Similarly, calcium regulation is inextricably linked to bone metabolism. Some bone disorders, such as osteoporosis or rickets, are treated mainly by endocrinologists, whereas others, such as renal osteodystrophy or phosphate wasting disorders, are often managed by nephrologists. Reproductive endocrinology has become a subspecialty of gynecology and urology, primarily because of the skills needed to perform procedures related to the evaluation and treatment of infertility. Ovulation induction protocols and various forms of assisted reproductive technology are increasingly used to manage infertility, which affects 10% to 15% of reproductive-age couples. Intracytoplasmic sperm injection has revolutionized the approach to male infertility. A new discipline of "oncofertility" addresses the need for fertility preservation in women, as well as men, with cancers that occur during the reproductive time frame. Common endocrine diseases, such as autoimmune thyroid disease and type 1 diabetes mellitus, are caused by abnormal regulation of immune surveillance and tolerance. Less common diseases, such as polyglandular failure, Addison disease, and lymphocytic hypophysitis, also have an immunologic basis. Although immunology is an independent discipline, the interface with endocrinology is important for understanding the pathogenesis of these disorders. Cytokines and interleukins have profound effects on the functions of the pituitary, adrenal, thyroid, and gonads. Thus, the boundaries of endocrinology change constantly, spawning new disciplines and expanding into new scientific realms.

KEY POINTS

- Although individual hormones regulate specific physiologic processes, such as growth, metabolism, and reproduction, they typically act in a concerted or integrated manner with other hormones.

HISTORICAL PERSPECTIVES

Although concepts of fertility and reproduction can be traced to ancient times, most of our current understanding of endocrinology has evolved during the past 150 years.[2] The structures of the major glands and ducts were initially captured in drawings by Renaissance anatomists and artists. The publication of *De Humani Corporis Fabrica* in 1543 by Vesalius provided a turning point in studies of human anatomy. Fallopio, also of the Padova School, published *Observationes Anatomicae* in 1561, which included a detailed description of the "slender and narrow seminal passage that arises from the horn of the uterus."

A timeline for selected advances in endocrinology is depicted in Figure 1.1. Berthold recorded the physiologic consequences of castration in 1849. He demonstrated that castration of a cock caused regression of secondary sex characteristics and mating behavior. Transplantation of the testes into the abdominal cavity restored these features, proving a role for the gonads in sexual differentiation and illustrating basic principles of hormone withdrawal and replacement. In 1855, Claude Bernard noted that the liver produced two secretions, an external secretion (bile) and an internal secretion (glucose), which passed directly into the circulation. This concept was later extended by Bayliss and Starling, who discovered that secretin, a substance extracted from duodenal mucosa, induced pancreatic exocrine secretion after intravenous injection. This observation distinguished the properties of circulating hormones from physiologic reflexes mediated by the nervous system.

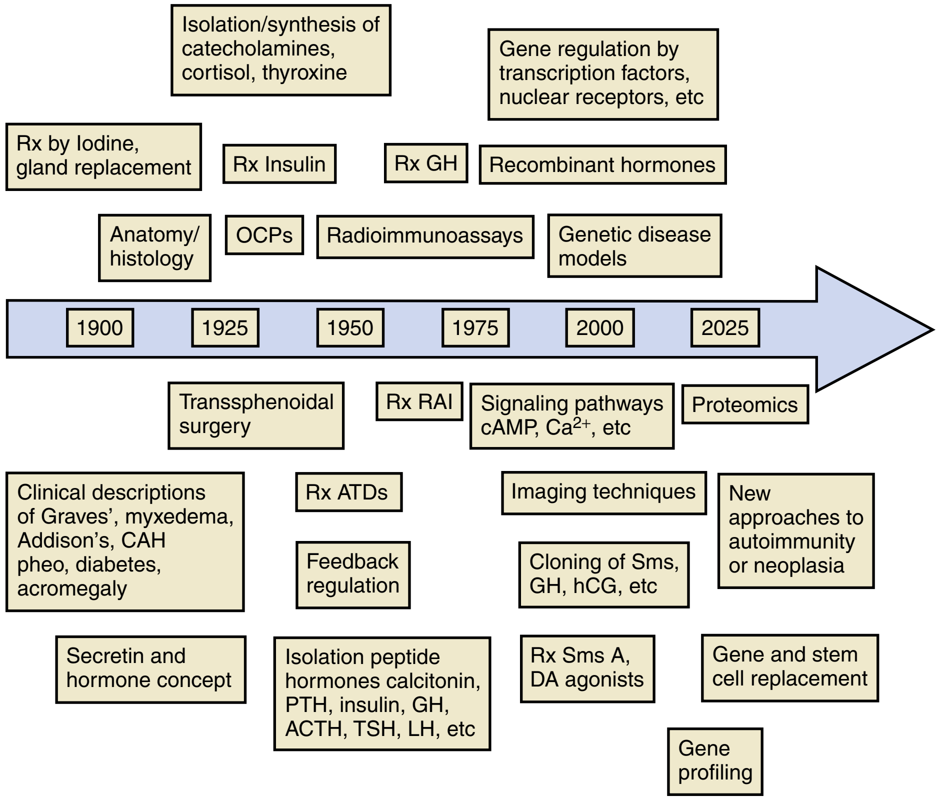
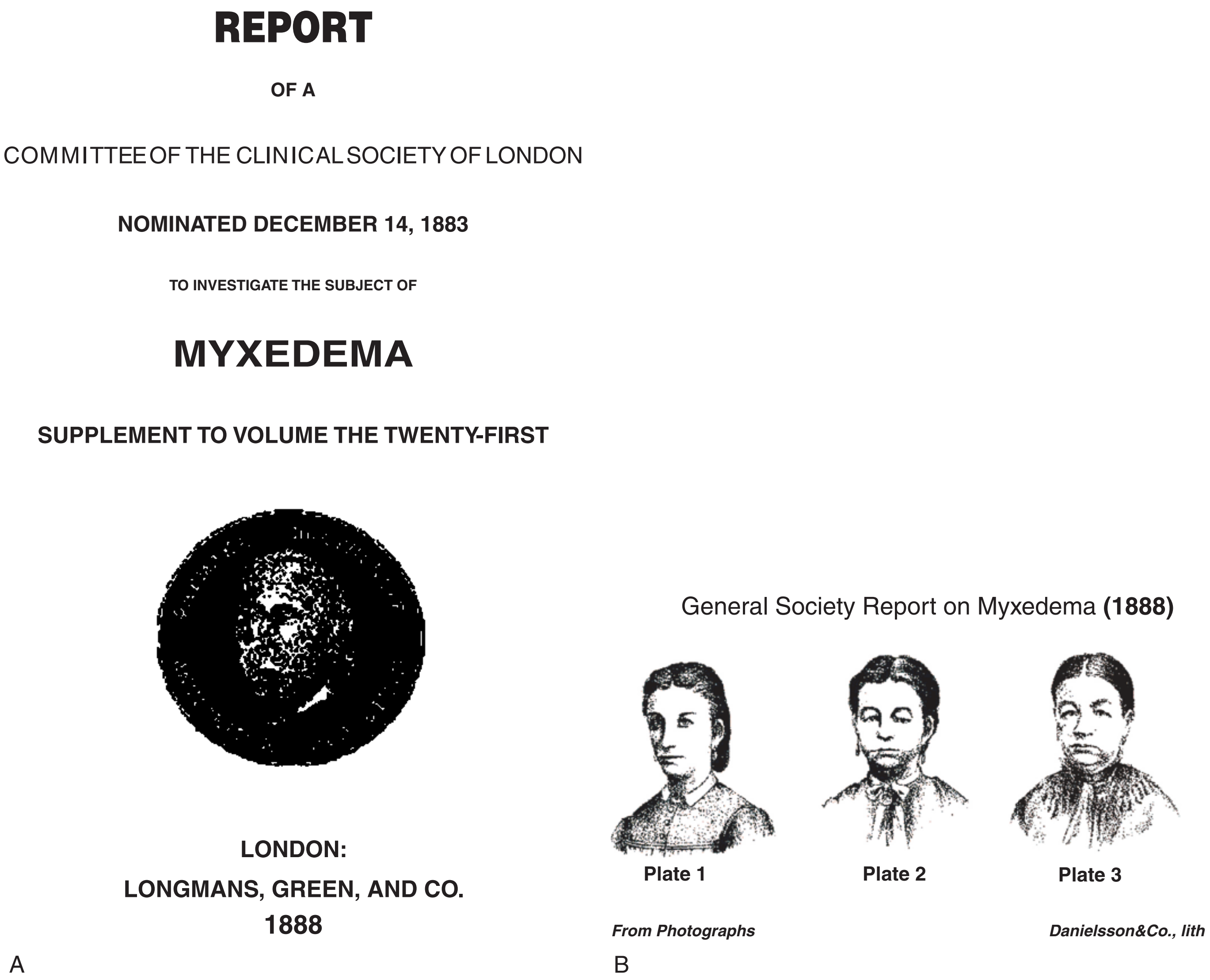

Isolation/synthesis of catecholamines, cortisol, thyroxine
Gene regulation by transcription factors, nuclear receptors, etc
Rx by Iodine, gland replacement
Rx Insulin
Rx GH
Recombinant hormones
Anatomy/ histology
OCPs
Radioimmunoassays
Genetic disease models
1900
1925
1950
1975
2000
2025
Transsphenoidal surgery
Rx RAI
Signaling pathways cAMP, Ca2+, etc
Proteomics
Clinical descriptions of Graves', myxedema, Addison's, CAH pheo, diabetes, acromegaly
Rx ATDs
Imaging techniques
New approaches to autoimmunity or neoplasia
Feedback regulation
Cloning of Sms, GH, hCG, etc
Secretin and hormone concept
Isolation peptide hormones calcitonin, PTH, insulin, GH, ACTH, TSH, LH, etc
Rx Sms A, DA agonists
Gene and stem cell replacement
Gene profiling

Fig. 1.1 Timeline of selected advances in endocrinology. *ACTH,* Adrenocorticotropic hormone; *ATDs,* antithyroid drugs; *Ca^{2+}*, calcium; *CAH*, congenital adrenal hyperplasia; *DA,* dopamine; *GH*, growth hormone; *hCG*, human chorionic gonadotropin; *LH*, luteinizing hormone; *OCPs,* oral contraceptive pills; *pheo,* pheochromocytoma; *PTH*, parathyroid hormone; *RAI,* radioactive iodine; *Rx,* treatment; *Sms A,* somatostatin analogs, *TSH,* thyroid-stimulating hormone.

In the late 1800s, the clinical manifestations of many endocrine disorders were described. The Report on Myxedema by the Clinical Society of London (1888) is a remarkable example of the power of astute clinical observation. In addition to recognition that the adult disorder of myxedema shared certain clinical features of cretinism,[3] a tenuous connection to thyroid gland dysfunction was proposed. The plates shown in Figure 1.2 illustrate some of the clinical manifestations of hypothyroidism as described by William Ord, who coined the term *myxedema* (mucinous edema).[4] Several years later, George Murray tested the role of the thyroid gland in myxedema by demonstrating that repeated subcutaneous injections of sheep thyroid extract corrected the disorder.[5] This was probably the first example of successful hormone replacement and spawned parallel efforts for other glandular diseases. By the turn of the century, the clinical manifestations of Graves disease, acromegaly, Addison disease, diabetes mellitus, and pheochromocytoma were well established. Hormone isolation and replacement strategies became a major research effort, culminating in the characterization of corticosteroids, thyroid hormones, and sex steroids. The history of endocrinology is replete with colorful renditions of hormone isolation and discovery. A recurring theme is teamwork and parallel observations by different teams working on the same problem—a testimony to the impact of scientific communication and the need for technology to drive advances.

The discovery of insulin in 1921 has been chronicled extensively and is a true inflection point in endocrinology.[6] The pancreatic islets are clusters of endocrine cells that are embedded within the exocrine pancreas. Early experiments in dogs by Minkowski[7] showed that pancreatectomy caused diabetes, demonstrating the pancreas as the organ responsible for regulating glucose. Banting and Best set out to isolate insulin, a process that was greatly aided by the expertise of Collip, a



REPORT

OF A

COMMITTEE OF THE CLINICAL SOCIETY OF LONDON

NOMINATED DECEMBER 14, 1883

TO INVESTIGATE THE SUBJECT OF

MYXEDEMA

SUPPLEMENT TO VOLUME THE TWENTY-FIRST

LONDON:
LONGMANS, GREEN, AND CO.
1888

A
General Society Report on Myxedema (1888)
Plate 1
Plate 2
Plate 3
From Photographs
Danielsson&Co., lith
B

Fig. 1.2 A, Cover page from the 1888 Clinical Society of London Report on Myxoedema. B, Clinical manifestations of myxedema. Plates taken from serial photographs of a woman with untreated hypothyroidism. Plate 1: At 21 years of age, before onset of myxedema. Plate 2: At 28 years of age, showing early features of myxedema. Plate 3: At 32 years of age, illustrating overt features of myxedema. (From *Clinical Society of London report on myxedema.* Boston: Francis A. Countway Library of Medicine; 1888. Photographs originally published in Ord WM. On myxoedema, a term proposed to be applied to an essential condition in the "cretinoid" affection occasionally observed in middle-aged women. *Medico-Chirurgical Trans.* 1978;61:57–78.)

protein chemist who isolated several other peptide hormones, including PTH.[8,9] Despite erratic initial results in diabetic dogs, Banting and Best soon achieved unequivocal success using partially purified insulin. At the time of insulin isolation, children with type 1 diabetes had no treatment options aside from starvation therapy, which could not prevent their ultimate demise from hyperglycemia and ketoacidosis. The initial insulin treatment results were stunningly successful, providing immediate clinical benefits soon followed by the ability to achieve long-term management with repeated use of insulin injections (Fig. 1.3). This dramatic treatment strategy was stymied initially by the limited supply of purified insulin, a problem that ultimately was solved by the development of recombinant human insulin. In this current era, pancreas and islet transplantation represent alternative treatment approaches. However, limited human donor tissue and ongoing challenges with immunosuppression have restricted the use of transplantation to patients with severe type 1 diabetes. There is hope, however, that stem cell biology or the ability to regenerate pancreatic islet β cells might ultimately overcome these limitations.

Recognition that the hypothalamus produces a variety of pituitary regulatory factors was another major advance. In addition to establishing a link between the brain and the "master gland," the hypothalamic-pituitary system underscored the critical importance of anatomic proximity and vascular delivery for the regulation of hormone action. It is now appreciated that discrete pulses of hypothalamic gonadotropin-releasing hormone (GnRH), GH-releasing hormone, thyrotropin-releasing hormone (TRH), and corticotropin-releasing hormone (CRH) act locally on the pituitary gland and exert little, if any, physiologic effect at more distal sites in the body.

Following the isolation of many steroid and peptide hormones during the first half of the twentieth century, a conceptual framework was outlined for mechanisms of hormone action. For peptide hormones, Sutherland established the idea of a second messenger system in which a hormone binds to a membrane receptor, thereby activating intracellular second messenger pathways such as cyclic adenosine monophosphate (cAMP).[10] For steroid and thyroid hormones, Tata established the concept of hormone action at the nuclear level, acting via intracellular receptors that altered gene expression, which in turn caused changes in protein levels.[11] The development of the radioimmunoassay (RIA) by Berson and Yalow revolutionized endocrine physiology and diagnosis by allowing accurate measurement of minute amounts of circulating hormones.[12] The impact of RIA on physiology, endocrinology, and clinical medicine cannot be overemphasized. RIA and related assays are now used routinely for almost all hormone measurements and have replaced many less-sensitive chemical methods and bioassays. RIA was once the province of specialty endocrine laboratories but has gradually become automated and integrated into clinical pathology laboratories. Mass spectroscopy methods are being used increasingly as a means to measure steroids and peptides. These methods not only are sensitive and highly quantitative, but also do not necessarily depend upon antibodies to detect specific molecules.

Important advances in therapeutic modalities have accompanied our improved understanding of endocrine diseases. Hormone replacement strategies have been refined, along with advances in surgical approaches for endocrine tumors. Many hormone excess syndromes are primarily managed surgically, including transsphenoidal surgery for pituitary tumors or excision of parathyroid, adrenal, and pancreatic tumors. Many glandular surgeries are now performed via minimally invasive techniques, such as laparoscopy or video-assisted resection through very small incisions. In addition to hormonal replacements, important medical therapies that have been developed include the use of radioactive iodine[13] and antithyroid drugs[14] for hyperthyroidism, dopamine agonists for prolactinomas,[15] somatostatin analogues for acromegaly, oral hypoglycemics for diabetes,[16] gonadal steroids as

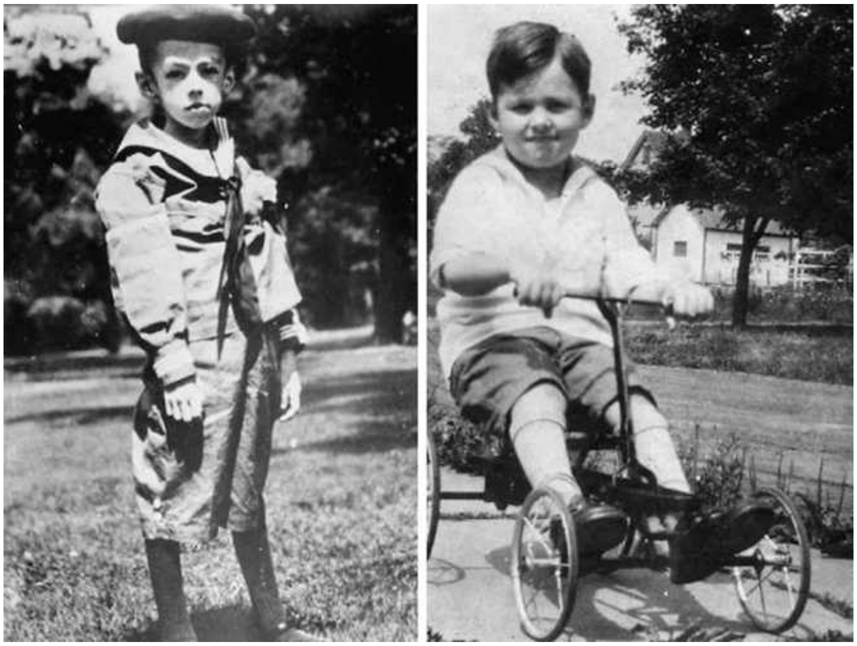


Fig. 1.3 Treatment of type 1 diabetes mellitus with insulin. Teddy Ryder was one of the first patients treated by Dr. Banting. After undergoing "starvation treatment" *(left panel),* which was the only therapy available at the time, he began insulin treatment at 5 years of age (July 10, 1922) *(right panel).* One year later (July 10, 1923), he is seen "cured." Teddy Ryder lived to 76 years of age. (Modified with permission from the University of Toronto Libraries Discovery and Early Development of Insulin online collection, http://digital.library.utoronto.ca/insulin/.)

contraceptives,[17] and somatostatin analogues for tumors of the gastrointestinal tract.[18]

In recent years, the tools of molecular genetics have dramatically accelerated our understanding of endocrinology. DNA sequences encoding hormones such as somatostatin,[19] GH,[20] insulin,[21] and chorionic gonadotropin[22] were among the first human complementary DNAs (cDNAs) cloned. Recombinant DNA techniques are now used routinely to identify new hormones and receptors and to elucidate hormone function.

Hormone genes have provided important models for understanding mechanisms of transcriptional regulation. Hormones typically are expressed in a cell-specific manner (e.g., GH, thyroglobulin), providing prototypes for identifying transcription factors (e.g., Pit-1, TTF-1) that restrict expression to particular cells or tissues. Hormone-regulated pathways have provided experimental variables that can be switched on or off, thereby revealing highly regulated target genes that can be used as experimental models. Thus, studies of the cAMP signaling system have unraveled the protein kinase A cascade and transcription factor targets, such as cAMP response element binding protein. Nuclear receptor pathways have been particularly illuminating. In addition to identifying target genes regulated by hormones such as estrogen, glucocorticoid, or thyroid hormone, detailed analyses of these pathways have helped to define how DNA binding specificity is encoded in promoters and how transcription factors suppress or enhance gene expression by recruiting corepressor or coactivator complexes. Transcription by nuclear receptors is arguably the best understood paradigm for how transcription factors initiate transcription, assemble a transcription complex, and renew the process to ensure multiple rounds of RNA synthesis. The genetic basis for several hundred endocrine disorders has been determined through the use of molecular biological approaches, and these tests are being used increasingly in clinical practice.

In addition to technical advances such as RIAs and recombinant DNA technology, endocrinology has contributed disproportionately to pivotal conceptual advances in science and medicine. Almost every aspect of physiology is tied to rhythms. The endocrine system has provided models for rapid rhythms such as luteinizing hormone (LH) or GH pulsatility, circadian rhythms such as cortisol or vasopressin production, and longer rhythms such as the menstrual cycle or bone remodeling. Concepts of hormone–receptor interaction and second messengers established signal transduction paradigms that proliferated into innumerable signaling networks. Polypeptide precursors, such as pro-opiomelanocortin (POMC), preproglucagon, preproparathyroid hormone, and others established pathways for protein processing, transport, and secretion. Studies of growth factors helped to refine concepts of autocrine and paracrine action, which can be viewed as an extension of classic endocrine action. Hormone replacement formed the foundation for the use of biologic agents such as factor VIII, G-CSF, and erythropoietin. The genetic basis for cancer has been elucidated by studies of the multiple endocrine neoplasia syndromes, types I and II.

KEY POINTS

- The field of endocrinology has pioneered many technological advances, including protein sequencing, RIAs, ligand–receptor interactions, and recombinant DNA technology, including the large-scale synthesis of hormones such as insulin or GH.

PRINCIPLES OF HORMONE ACTION

The principles of hormone action include fundamental concepts such as hormone biosynthesis and secretion, feedback regulation, hormone–receptor binding, and initiation of intracellular signaling. These principles are broadly applicable to other subspecialties in addition to endocrinology.

Hormone Biosynthesis and Secretion

Hormones can be divided into five major classes: (1) amino acid derivatives such as dopamine, catecholamines, and thyroid hormone; (2) small neuropeptides such as GnRH, TRH, somatostatin, and vasopressin; (3) large proteins such as insulin, LH, and PTH that are produced by classic endocrine glands; (4) steroid hormones such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) vitamin derivatives such as retinoids (vitamin A) and vitamin D. As a rule, amino acid derivatives and peptide hormones interact with cell-surface membrane receptors. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and interact with intracellular nuclear receptors.

Many peptide hormones are produced from precursor polypeptides. Characteristic signal or leader sequences target these peptides for extracellular transport via secretory granules. Some precursors, such as POMC or preproglucagon, encode multiple biologically active peptides that are generated by specific processing enzymes; other precursors, such as preproinsulin and vasopressin, encode single hormones that are excised from larger proteins. The secretion of peptide hormones is tightly controlled by intracellular signals that regulate vesicle transport and fusion with the plasma membrane, resulting in hormone release into the extracellular milieu. Steroid hormones such as progesterone, cortisol, and testosterone are synthesized from cholesterol derivatives through a series of enzymatic steps. These enzymes are expressed specifically in steroidogenic tissues such as the adrenal gland and gonads. Their enzymatic activities are regulated in response to trophic hormones such as adrenocorticotropic hormone (ACTH), LH, or follicle-stimulating hormone (FSH). Thyroid hormone is produced by modifications (iodination) of tyrosines in thyroglobulin. Vitamin D and retinoic acid are derived in part from dietary sources but can also be generated and activated by endogenous synthetic pathways.

KEY POINTS

- Hormones can be divided into five major classes: (1) amino acid derivatives such as dopamine, catecholamines, and thyroid hormone; (2) small neuropeptides such as GnRH, TRH, somatostatin, and vasopressin; (3) large proteins such as insulin, LH, and PTH that are produced by classic endocrine glands; (4) steroid hormones such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) vitamin derivatives such as retinoids (vitamin A) and vitamin D.

Feedback Regulation

The elucidation of negative feedback has had a profound impact on endocrinology. This principle holds that hormones have a particular set point that is controlled by downregulating stimulatory pathways when the set point is exceeded and upregulating stimulatory pathways when hormone levels fall below the set point. Probably every hormone is regulated in this manner, although the regulatory pathways might not be immediately evident for new hormones. These regulatory loops are well illustrated by the major hypothalamic-pituitary-hormone axes and include both stimulatory (e.g., TRH stimulates thyroid-stimulating hormone [TSH]; TSH stimulates T4/T3 production) and inhibitory components (e.g., T4/T3 suppress TRH and TSH) (Fig. 1.4). Feedback regulation also occurs for endocrine systems that do not involve the pituitary gland. For example, calcium feeds back to inhibit PTH, glucose inhibits insulin secretion, and leptin acts on hypothalamic

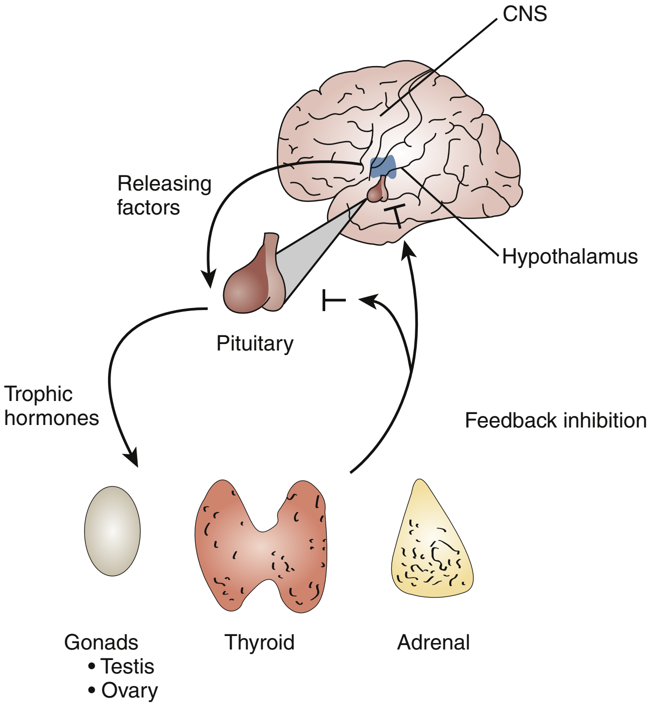


Fig. 1.4 Feedback regulation of the hypothalamic-pituitary axis. *CNS*, Central nervous system.

pathways to suppress appetite. Although the descriptions of these feedback mechanisms oversimplify the complex physiologic pathways that regulate hormone levels, they provide useful insight into endocrine testing paradigms. For example, hypothyroidism is characterized by elevated TSH, an appropriate physiologic response to deficient thyroid hormone levels. Dexamethasone suppression of the CRH/ACTH axis is used to diagnose Cushing disease, which is characterized by impaired negative feedback regulation. A deficient adrenal response to exogenous ACTH is used to document primary adrenal insufficiency.

KEY POINTS

- Feedback regulation is a powerful means to achieve homeostasis. It is also the foundation for important endocrine tests, such as the ACTH stimulation test or the dexamethasone suppression test.

Paracrine and Autocrine Regulation

Whereas feedback mechanisms control many classic endocrine pathways, local regulatory systems, often involving growth factors, play critical roles in all tissues (Fig. 1.5). Paracrine regulation refers to factors released by one cell that act on an adjacent cell in the same tissue. For example, somatostatin secretion by pancreatic islet delta cells inhibits insulin secretion from nearby β cells. The oocyte produces growth and differentiation factor–9, which acts on adjacent granulosa cells to stimulate the transition of primary follicles to secondary follicles. The anatomic relationships of cells have an important influence on paracrine regulation. Seminiferous tubules are exposed to a very high testosterone concentration from the interstitial Leydig cell compartment. On the other hand, the Sertoli cell product, androgen-binding protein, helps to retain high local testosterone concentrations. Activin exerts paracrine effects in the pituitary gland, where it stimulates FSH production. However, activin also exerts biologic activity in many other tissues, perhaps explaining why it is regulated locally and is neutralized by binding proteins such as follistatin. Autocrine regulation describes the action of a factor on the same cell from which it is produced. IGF-1 acts on many cells that produce it, including chondrocytes, breast epithelium, and gonadal cells. Intracrine regulation refers to effects within a cell. The term is not commonly used but captures the important concept that many signaling and enzymatic pathways are influenced by other pathways or by substrate or product concentrations. For example, 3-hydroxy-3-methylglutaryl coenzyme A reductase, the rate-limiting enzyme in cholesterol biosynthesis, is inhibited by the end product, cholesterol.

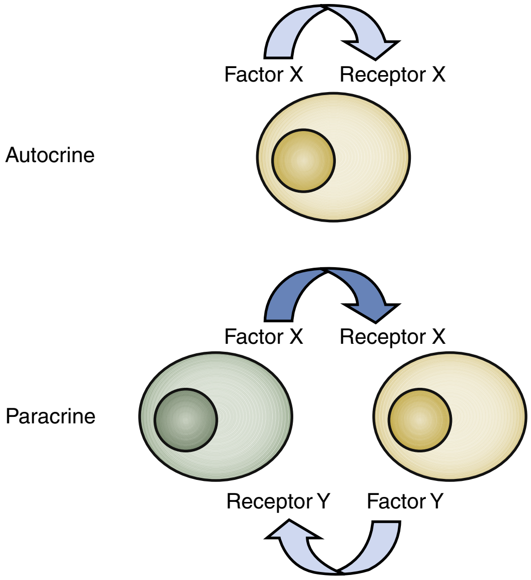


Fig. 1.5 Autocrine and paracrine regulation. Many growth factors act locally to regulate cell growth, differentiation, and function. Autocrine regulation describes the action of a factor that acts on the same cell, whereas paracrine regulation describes a circumstance in which the product of one cell acts on a different cell type.

KEY POINTS

- Paracrine and autocrine regulation allow growth factors and cytokines to act locally, minimizing systemic biological effects.

Hormonal Rhythms and Pulsatility

Hormonal rhythms are used to adapt to environmental changes, such as seasons of the year, the daily light-dark cycle, sleep, meals, and stress. In many species, reproduction is seasonal, presumably a mechanism to ensure survival of the offspring. In the extreme northern and southern hemispheres, calcium absorption and bone remodeling decline during winter, when vitamin D production is reduced. The human menstrual cycle is repeated on average every 28 days, reflecting the time required for follicular maturation and ovulation. In some species, estrus cycles are intimately linked to mating behavior induced by behavioral cues and the production of pheromones. Essentially all pituitary hormone rhythms are entrained to sleep and the circadian cycle, which in turn is dictated by sunlight exposure. The hypothalamic-pituitary-adrenal (HPA) axis, for example, exhibits characteristic peaks of ACTH and cortisol production before dawn and a nadir between late afternoon and midnight. Recognition of these rhythms is important for endocrine testing and treatment. Patients with Cushing syndrome exhibit inappropriately increased midnight cortisol levels. The HPA axis is more susceptible to suppression by glucocorticoids administered at night because they blunt the early morning rise of ACTH. Understanding this diurnal rhythm provides the basis for increased physiologic hormone replacement through the use of larger glucocorticoid doses in the morning than in the afternoon.

Although circadian rhythms were identified initially in the context of sleep cycles and hormonal rhythms, there is mounting evidence for circadian periodicity at the level of organs and individual cells.[23] For example, a number of genes expressed by hepatocytes, such as *REV-ERB,* exhibit rhythmic expression, thereby exerting epigenetic regulatory effects on metabolic pathways such as gluconeogenesis and lipid biosynthesis.

Many peptide hormones are secreted in discrete pulses, often reflecting regulation by the nervous system. For example, hypothalamic GnRH induces LH pulses once every 1 to 2 hours. Intermittent hypothalamic GnRH pulses are required to maintain pituitary gonadotrope sensitivity, whereas continuous GnRH exposure causes desensitization. This feature of gonadotropin regulation serves as the basis for using long-acting GnRH agonists to treat central precocious puberty or to decrease testosterone levels in the management of prostate cancer.

The pulsatile nature of hormone secretion and the rhythmic pattern of hormone production have important implications for the measurement of circulating hormone levels, as levels can change dramatically over several hours. For some hormones, integrated markers have been developed to circumvent hormonal fluctuations. For example, a 24-hour collection of urinary free cortisol integrates cortisol production throughout a diurnal cycle. IGF-1 provides a relatively stable biologic marker of GH action. Glycosylated hemoglobin is used as an index of long-term (weeks to months) circulating blood glucose, which is linked covalently to hemoglobin in a concentration-dependent manner.

KEY POINTS

- Virtually every hormonal system is linked directly or indirectly to circadian rhythms of sleep, light exposure, and feeding. These circadian rhythms not only influence physiologic responses but have practical implications for interpreting the measurements of many hormones.

Hormone Transport and Degradation

The level of a hormone is determined by its rate of secretion and its circulating half-life. After protein biosynthesis and precursor processing, peptide hormones are stored in secretory granules. These granules undergo progressive maturation and sequential translocation before arriving at the plasma membrane for imminent release into the circulation. The stimulus for hormone secretion is typically a releasing factor or neural signal that induces rapid changes in intracellular calcium concentration, which leads secretory granules to fuse with the plasma membrane and releases their contents into the extracellular environment and bloodstream. In contrast, steroid hormones usually diffuse into the circulation as they are synthesized. Thus, their secretion closely mirrors rates of synthesis. For example, ACTH and LH induce steroidogenesis by stimulating the activity of steroidogenic acute regulatory protein, which transports cholesterol into the mitochondrion. In parallel, ACTH and LH stimulate the production of other rate-limiting enzymes in the steroidogenic pathway, such as the cholesterol side-chain cleavage enzyme.

Hormone-binding proteins can affect volume of distribution, level of unbound or “free” hormone, and rates of hormone clearance. Most steroid hormones and many peptide hormones circulate in association with binding proteins. T4 and T3 bind to thyroxine-binding globulin (TBG), albumin, and thyroxine-binding prealbumin. Similarly, cortisol binds to cortisol-binding globulin, and androgens and estrogens bind to sex hormone–binding globulin (SHBG). IGF-1 and IGF-2 bind to multiple IGF-binding proteins. GH interacts with GH-binding protein, a circulating fragment of the GH receptor extracellular domain. Abnormal binding proteins can significantly alter total hormone concentrations but usually have little clinical consequence, as the regulatory feedback systems respond to unbound or “free” hormone levels. For example, TBG deficiency greatly reduces total thyroid hormone levels, but the free concentrations of T4 and T3 remain normal. Liver disease and medications can also influence binding protein levels (e.g., estrogen increases TBG). Nonetheless, these abnormalities can create diagnostic confusion, and some alterations (e.g., increased SHBG) may shift ratios of hormones (e.g., testosterone, estradiol) that bind with different affinities.

Knowledge of hormone half-life is important for achieving physiologic hormone replacement, as the frequency of dosing and the time required to reach steady state are determined by rates of hormone decay. T4, for example, has a half-life of about 7 days. Consequently, more than 1 month is required to reach a new steady state, and single daily doses are sufficient to achieve constant hormone levels. T3, in contrast, has a half-life of about 1 day. Its administration is associated with more dynamic serum levels, and it must be administered two to three times per day to generate more constant blood levels. Synthetic glucocorticoids vary widely in their half-lives. Analogues with a longer half-life (e.g., dexamethasone) are associated with greater suppression of the HPA axis. Most protein hormones (e.g., ACTH, GH, PRL, PTH, LH) have relatively short half-lives (<20 minutes), leading to sharp peaks of secretion and decay. These dynamic hormone fluctuations must be considered in analysis of hormone levels, which may vary widely over short time intervals and between patient visits. Although frequent hormone sampling can trace these pulses, this is not practical outside of a research setting. In some instances, clinicians elect to pool several samples to obtain a more representative hormone level. More often, clinicians recognize that the normal ranges are relatively wide, in part because of hormonal pulses. Rapid hormone decay is useful in certain clinical settings. Because PTH has a short half-life, intraoperative PTH can be used to confirm successful removal of a parathyroid adenoma. This is particularly valuable diagnostically when there is a possibility of multicentric disease or parathyroid hyperplasia, as occurs with multiple endocrine neoplasia or renal insufficiency.

KEY POINTS

- Precursor hormones provide another level of regulation of hormone bioavailability and action. For many hormones, such as POMC or preproglucagon, these precursors can generate multiple hormone products.

HORMONE ACTION THROUGH RECEPTORS

Hormone receptors can be divided broadly into membrane receptors and nuclear receptors. Membrane receptors primarily bind peptide hormones and small molecules that cannot traverse the plasma membrane (e.g., catecholamines, dopamine). Nuclear receptors bind small, lipid-soluble molecules that diffuse or are transported across the cell membrane (e.g., thyroid hormone, steroids, vitamin D). Hormones bind to both classes of receptors with specificity and high affinity. These characteristics are often described by Scatchard plots, which allow estimation of equilibrium dissociation constants and maximum binding (Fig. 1.6).

Binding affinity generally coincides with the concentration of circulating hormones and is typically in the subnanomolar range. Receptor occupancy at any given moment is a function of hormone

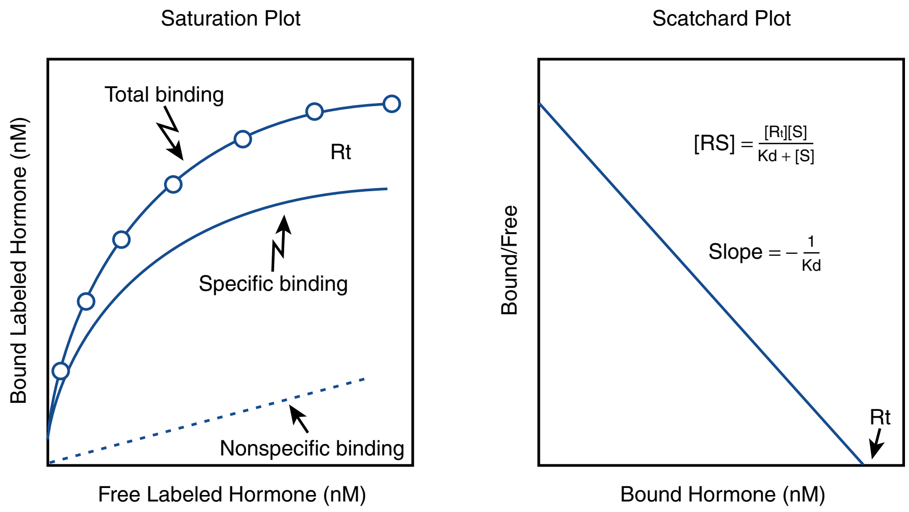


Fig. 1.6 **Plots of ligand–receptor interactions.** *Left panel,* Theoretical equilibrium saturation plot. As increasing amounts of labeled hormone are added, the amount of receptor-bound hormone increases until the binding sites are saturated (total binding). The addition of a large amount of unlabeled hormone allows determination of nonspecific or unsaturable binding. The hormone concentration at which half-maximal binding occurs provides a measure of binding affinity. *Right panel,* Theoretical Scatchard plot. The *x*-axis represents specific binding, and the *y*-axis denotes specific binding divided by free radioligand concentration. Maximal binding is estimated by the *x*-intercept, and the dissociation constant (K_d) is estimated as the negative reciprocal of the slope. *Rt,* receptor, total; *S,* substrate.

concentration and the receptor's affinity for the hormone. Receptor numbers vary greatly in different target tissues, providing one of the major determinants of specific cellular responses to circulating hormones. For example, ACTH receptors are located almost exclusively in the adrenal cortex, and FSH receptors are found predominantly in the gonads. In contrast, insulin and thyroid hormone receptors are widely distributed, reflecting the need for metabolic responses in all tissues. Tissue-specific knockouts of insulin receptors have revealed distinct metabolic actions of insulin in various tissues.[24] For example, deletion in muscle causes increased free fatty acids and adipose tissue, whereas insulin receptor deletion in adipose tissue reduces fat mass and enhances insulin sensitivity. Deletion of the receptor in pancreatic islet β cells impairs insulin secretion, apparently because of reduced islet cell mass.

KEY POINTS

- In general, peptide hormones interact with cell-surface membrane receptors, which in turn activate second messenger signaling pathways. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and interact with intracellular nuclear receptors, which act predominantly by modulating gene transcription.

Membrane Receptors

Membrane receptors can be divided into several major groups (Fig. 1.7): (1) seven transmembrane domain G protein–coupled receptors (GPCRs), (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) the transforming growth factor–β (TGF-β) family of serine kinase receptors. There are several hundred GPCRs.[25] They bind a broad array of hormones, including large proteins (e.g., TSH, PTH), small peptides (e.g., TRH, somatostatin), catecholamines (epinephrine, dopamine), and even minerals (e.g., calcium). These receptors possess seven transmembrane-spanning regions composed of hydrophobic α-helical domains that are connected by extracellular and intracellular loops. After the receptor binds a hormone, these transmembrane domains undergo conformational changes that alter interactions with intracellular G proteins. The G proteins provide a link to intracellular signaling pathways such as adenyl cyclase, phospholipase C, mitogen-activated protein kinases (MAPK), and others. G proteins form a heterotrimeric complex that is composed of various Gα and Gβ-γ subunits. The α subunit contains the guanine nucleotide–binding site and hydrolyzes guanosine triphosphate (GTP) to guanosine diphosphate (GDP). Gα

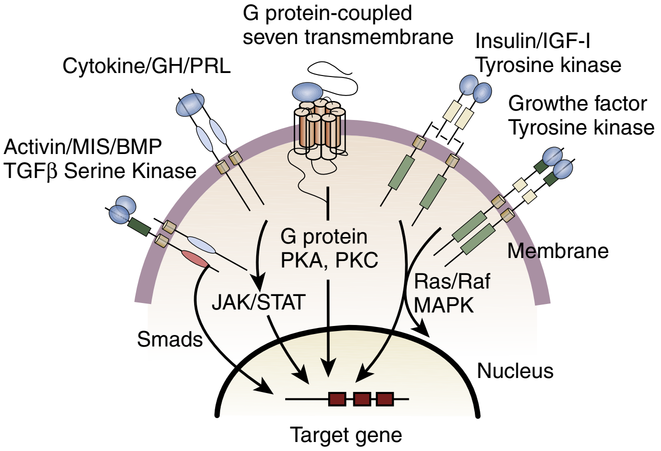


Fig. 1.7 **Membrane receptors.** Classes of membrane receptors can be defined on the basis of structural similarities and signaling pathways. The figure depicts major classes of membrane receptors, although these categories are somewhat arbitrary. Each of these receptor classes is described in a separate chapter. *BMP,* Bone morphogenetic protein; *GH*, growth hormone; *IGF*, insulin-like growth factor; *JAK,* Janus kinase; *MAPK,* mitogen-activated protein kinase; *MIS,* Mullerian inhibiting substance; *PKA,* protein kinase A; *PKC,* protein kinase C; *PRL,* prolactin; *STAT,* signal transduction and activators of transcription; *TGF,* transforming growth factor. (Modified with permission from *Harrison's principles of internal medicine.* 15th ed. New York: McGraw-Hill; 2001.)

is active when GTP is bound and inactive after hydrolysis to GDP. The Gβ-γ subunits modulate activity of the α subunit and mediate their own effector-signaling pathways. A variety of endocrinopathies result from G protein mutations or from mutations in receptors that modify their interactions with G proteins. For example, McCune–Albright syndrome is caused by somatic mutations in Gα that prevent GTP hydrolysis, thereby causing constitutive activation of the Gα signaling pathway. Selected mutations in the transmembrane domains of GPCRs can mimic hormone-induced conformational changes, leading to activation of G proteins independent of hormone binding. This type of mutation in the TSH receptor accounts for a significant fraction of solitary autonomously functioning thyroid nodules. Activating mutations in the LH receptor cause LH-independent precocious puberty in boys.

Tyrosine kinase receptors transmit signals for insulin and a variety of growth factors, such as IGF-1, epidermal growth factor, platelet-derived growth factor, and fibroblast growth factors. Ligand binding induces autophosphorylation, leading to interactions with intracellular adaptor proteins such as Shc and insulin receptor substrates. Depending on the receptor and adaptor complexes, one or more kinase pathways, including the Raf-Ras-MAPK and the Akt/protein kinase B pathways, are activated. The tyrosine kinase receptors play a prominent role in cell growth and differentiation, as well as in intermediary metabolism.

The GH and PRL receptors belong to the cytokine receptor family. Ligand binding induces receptor interactions with intracellular kinases such as the JAKs, which phosphorylate members of the STAT family, and other signaling pathways (Ras, PI3-K, MAPK). The activated STAT proteins translocate to the nucleus and stimulate expression of target genes.

The TGF-β receptor family mediates the actions of TGF-βs, activins, Müllerian inhibiting substance (MIS; also known as anti-Müllerian hormone), and bone morphogenic proteins. This receptor family consists of type I and II subunits, which undergo autophosphorylation after ligand binding. The phosphorylated receptors bind intracellular Smads (named for a fusion of terms for *Caenorhabditis elegans* sma + mammalian mad). Like the STAT proteins, the Smads serve a dual role of transducing the receptor signal and acting as transcription factors.

KEY POINTS

- Membrane receptors for hormones fall into four major groups that differ in their structures and signaling pathways: (1) seven transmembrane domain GPCRs, (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) TGF-β family serine kinase receptors.

Nuclear Receptors

The nuclear receptor superfamily can be divided into receptors with known ligands (type I and II) and so-called orphan nuclear receptors, for which ligands have not been identified or might not exist.[26] Type I nuclear receptors include classic steroid receptors (e.g., glucocorticoid, estrogen, progesterone, androgen, mineralocorticoid) that bind to DNA as homodimers. Type II nuclear receptors bind to DNA as heterodimers, typically using the retinoid X receptor as a partner. Type II receptors bind a variety of molecules, including thyroid hormones, vitamin D, retinoic acid, and bile acids. Many orphan receptors bind DNA as monomers. Nuclear receptors have a highly conserved central zinc-finger DNA binding domain (Fig. 1.8). The carboxyterminal domain is more variable and includes a ligand-binding pocket, dimerization surfaces, and transcription-activating motifs. The aminoterminal domain varies greatly in length and in some receptors contains transcription-activating domains.

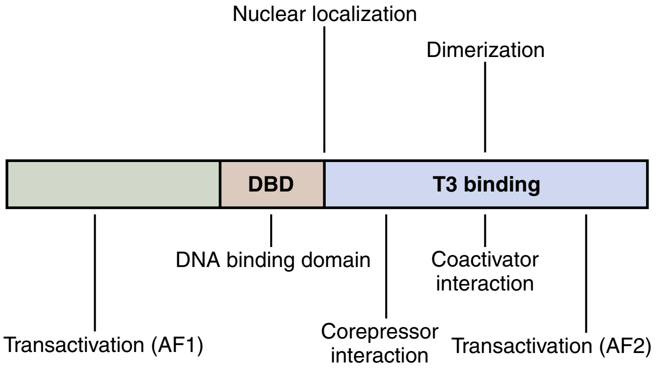


Fig. 1.8 Nuclear receptor structure. The central zinc-finger DNA binding domain is highly conserved among nuclear receptors. The carboxyterminus includes regions involved in ligand binding, transcriptional repression and activation, and dimerization. The aminoterminus is highly variable in length and, for some receptors, contains transcriptional regulatory sequences.

Nuclear receptors act primarily by altering rates of gene transcription, although there is growing evidence for interactions with other cellular signaling pathways, such as the MAPK pathway. Hormone binding induces conformational changes that induce receptor interactions with transcriptional cofactors. For type I receptors, ligand binding creates a bend in the extreme carboxyterminus (AF-2 domain) such that a hydrophobic cleft is created, providing a docking site for coactivators (e.g., steroid receptor coactivators). For type II receptors, hormone binding induces similar conformational changes. However, in the absence of hormone, corepressors are bound and silence gene transcription. Hormone binding induces the dissociation corepressors and the recruitment of coactivators (Fig. 1.9). Consequently, gene expression can shift from being silenced in the absence of hormone to being stimulated in the presence of hormone.

KEY POINTS

- Nuclear receptors mediate gene expression by recruiting an array of transcriptional corepressors or coactivators to alter rates of gene transcription or silencing.

ROLE OF THE CLINICAL ENDOCRINOLOGIST

Clinicians are often attracted to the field of endocrinology because it integrates physiology, biochemistry, and cell signaling with patient care. The fact that many endocrine disorders are amenable to cure or effective treatment also makes the practice of endocrinology especially satisfying.

As medicine becomes more specialized, physician roles are changing. Many general medicine inpatient services are now managed by hospitalists, who work in conjunction with primary care physicians and consultants. Inpatient length of stay continues to decrease, and this setting is now rarely used for extensive endocrine evaluation and testing. In fact, many endocrine pathways are perturbed by the stress associated with acute illness or hospitalization (e.g., sick euthyroid; false-positive dexamethasone suppression tests). These changes have resulted in a shift of an endocrinologist's activity to the outpatient setting.

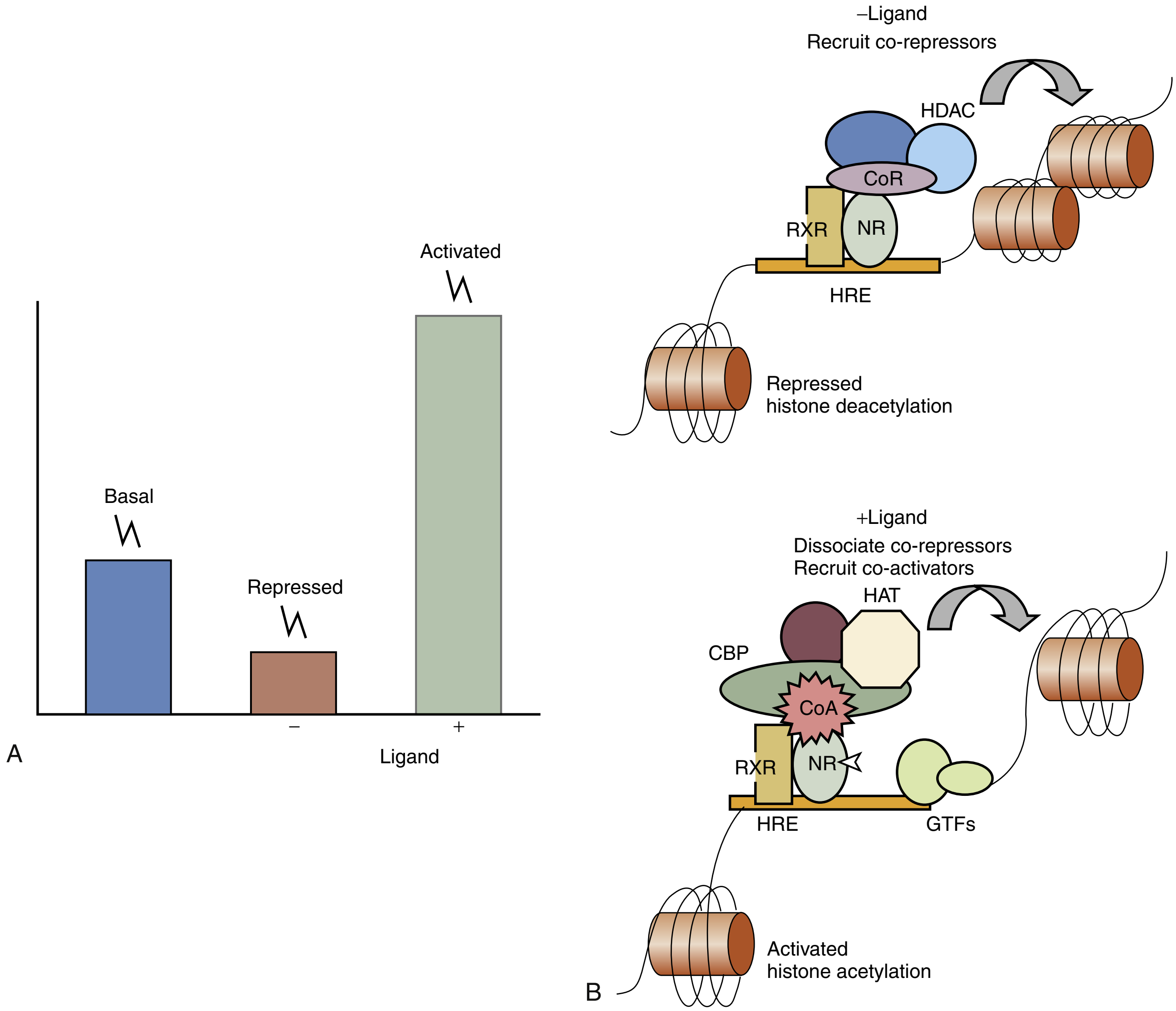


Fig. 1.9 Pathways of nuclear receptor transcriptional repression and activation. A, Nuclear receptors act primarily by altering rates of gene transcription. Many nuclear receptors, particularly type II receptors, cause transcriptional repression or silencing in the absence of ligand. After ligand binding, repression is relieved, and transcription is stimulated above the basal state. **B,** Although the exact mechanism of transcriptional repression or activation is variable, the model shown here illustrates ligand-induced relief of repression by corepressors (CoRs), followed by the recruitment of coactivators (CoAs). Many CoR complexes possess histone deacetylase activity, which is thought to induce gene silencing. CoA complexes possess histone acetylase activity, which induces changes in chromatin structure that facilitate the recruitment of additional transcriptional activators, which ultimately stimulate RNA polymerase. *CBP,* CREB-binding protein; *GTFs,* general transcription factors; *HAT,* histone acetyl transferase; *HDAC,* histone deacetylase; *HRE,* hormone response element; *NR,* nuclear receptor; *RXR,* retinoid X receptor.

Clinical endocrinologists serve three main roles: (1) to serve as consultants for patients who present with clinical endocrine conundrums that stretch the knowledge base and expertise of general internists or physicians in nonmedical specialties; (2) to provide specific short-term services for the treatment of endocrine disorders such as Graves disease, management of a thyroid nodule, or evaluation and treatment of hyperparathyroidism; and (3) to provide long-term management of challenging disorders such as brittle diabetes mellitus, congenital adrenal hyperplasia, or hypocalcemia. For many years, endocrinologists developed and performed specialized tests, including RIAs. In addition to performing appropriate stimulation or suppression protocols, special expertise was required to run and interpret these hormone assays. However, sensitive immunoradiometric assays, such as those for TSH, PTH, and most other hormones, are now commercially available. In some countries, these changes have influenced reimbursement, in that hormone testing generated revenue and compensated for other services, such as diabetes education. In the United States, endocrinologists are currently reimbursed primarily on the basis of evaluation and management coding, which takes into account the diagnosis and complexity of the evaluation. Endocrine procedures include thyroid aspiration biopsies, bone density measurements, and radioiodine treatment. However, wide institutional variation is seen in terms of who performs these procedures.

In addition to these classical roles, clinicians often make important contributions by recognizing new diseases or variants of old themes. For example, thyroid hormone resistance, lack of the GH receptor, and leptin deficiency states were identified in clinical practice. Often this role is played by the physician who sees patients and is involved in research: the clinical investigator.

Although practices vary considerably, the most common disorders seen by most endocrinologists are diabetes mellitus, thyroid disorders, metabolic bone disease, pituitary disorders, and reproductive abnormalities and infertility.[27] Pediatric endocrinologists also see a large number of patients with growth deficiency, pubertal delay, and a variety of inherited endocrine diseases such as congenital adrenal hyperplasia and Turner syndrome. Many academic centers have begun to structure multidisciplinary clinics for the management of diabetes mellitus, pituitary tumors, disorders of sex determination, and thyroid nodules.[28]

Increasingly, the clinical challenge is to identify endocrine disorders at their earliest stages before clinical manifestations are obvious. Terms such as *subclinical hypothyroidism, impaired glucose tolerance,* and *incidental adrenal* or *pituitary adenoma* have crept into our vocabulary and have changed our approach to patients. Similarly, endocrine disorders such as osteoporosis, hyperparathyroidism, hypertension, or hyperlipidemia rarely present with specific symptoms or signs. Because the clinical features of subclinical endocrine disorders are subtle or absent, laboratory testing takes on added importance as we attempt to diagnose more subtle forms of disease.

Available treatment options strongly influence the practice of endocrinology. Many new drugs have been discovered and approved during the past decade. Type 2 diabetes now can be treated by using drugs that influence insulin sensitivity (thiazolidinediones, metformin) and by administering agents that enhance insulin release (sulfonylureas, exenatide, repaglinide). Insulin derivatives with altered pharmacokinetics (lispro, aspart, glulisine, insulin glargine, detemir) facilitate intensive blood glucose control with a reduced risk for hypoglycemia. Bisphosphonates provide a novel mechanism for inhibiting osteoclast function and are widely used to treat osteoporosis and hypercalcemia. PTH analogs are being used to stimulate osteoblast function to enhance bone mass. Medical therapy for prolactinomas with dopamine agonists established a new paradigm for nonsurgical management of pituitary tumors. More recently, somatostatin analogues were developed, including depot formulations, for adjunctive or primary treatment of acromegaly. Pegvisomant is a novel GH analogue that blocks the GH receptor and is highly efficacious for reducing IGF-1 levels in patients with acromegaly. A variety of new formulations are available for delivering sex steroids. These include estrogen and testosterone patches. Testosterone gels and gum patches are also available. Building on the observation that tamoxifen exhibits mixed agonist/antagonist activity in various tissues, an active search is under way for additional selective estrogen receptor modulators (e.g., raloxifene) that exhibit unique profiles of estrogen action. It is expected that similar drugs might be identified for androgen, glucocorticoid, thyroid, and other nuclear receptors. Inhibitors of enzymes such as aromatase (anastrozole, letrozole) and 5α-reductase (finasteride) allow selective reduction of steroid levels. The potential for producing recombinant hormones has been fully realized, providing unlimited supplies of uncontaminated hormones or their derivatives. Examples of recombinant hormones in general use include insulin, GH, TSH, LH, and FSH. There is now great interest in the activities and clinical use of a variety of others, including leptin, GLP 1-37, YY 3-36, ghrelin, and others.

KEY POINTS

- Like most medical specialties, endocrinology is complex and requires highly specialized knowledge and experience. The effective consultant should gather information independently and decide whether to assume primary management of a patient if the clinical situation demands it, or to provide advice to the primary physician if the management is more straightforward.

MAJOR UNSOLVED PROBLEMS

Despite many impressive advances in endocrinology, a surprising number of fundamental problems remain incompletely solved. Although it is not possible to review all of these, it is provocative to identify a few, if only as a matter of perspective.

What is the basis of variability in normal ranges of hormones? The normal range for most hormones is remarkably broad. Testosterone, for example, varies between 300 and 1000 ng/dL, and total T4 ranges between 4 and 12 ng/dL. However, for any given individual, hormone levels are relatively constant, suggesting a defined set point. Some of this variability results from pulsatile or rhythmic hormone secretion. In some cases, differences in plasma-binding proteins account for variability. It seems likely that tissue responses, mediated by receptors and signaling pathways, also may define set points. From a practical perspective, the wide normal ranges create challenges for determining when subtle changes in hormone levels occur for a particular individual.

How should we optimally deliver hormone replacement? We currently have reasonable hormone replacements for most glandular deficiency syndromes. However, hormone replacement does not perfectly recapitulate normal physiology in any case. In type 1 diabetes, ideal insulin replacement would always maintain euglycemia without hypoglycemia. This challenge is underscored by the sometimes dramatic improvement in metabolic control after pancreatic islet transplantation. Thyroid hormone replacement, because of its relatively long half-life, approximates physiologic replacement, at least as assessed by TSH levels. Nonetheless, controversy about the need to replace T3 and T4 is ongoing. Glucocorticoid replacement is notoriously challenging, and patients often exhibit mild cushingoid features to avoid adrenal insufficiency. Ideally, we would use a physiologic marker, analogous to TSH, to assess adequate adrenal replacement. Similar issues exist for sex steroid replacement. Testosterone and estrogen replacement has been improved with the advent of patches and gels, but these delivery systems remain nonphysiologic. GH replacement in GH-deficient children rarely completely corrects height or even growth velocity. This might reflect relatively late diagnosis, or we might not have optimized hormone administration such that it mimics physiologic GH secretion or integrated actions with other hormones such as sex steroids. Many possible solutions to these challenges, including modified hormones and new formulations to alter absorption or pharmacokinetics, have been identified. In addition, careful clinical studies will be needed to correlate the physiologic effects of various hormone regimens.

How can osteoporosis be prevented? Currently, osteoporosis is identified after a fracture or on the basis of markedly reduced bone density. Treatment strategies can prevent further bone loss or can induce modest increases in bone mass. Clearly, it would be preferable to generate greater maximum bone mass early in life and to identify those who are at risk for accelerated bone loss.

What causes common autoimmune endocrinopathies such as type 1 diabetes mellitus, Hashimoto thyroiditis, Graves disease, and

Addison disease? Treatments are available for these disorders, but they are not directed at the underlying autoimmune process. If effective treatments could prevent, interrupt, or suppress the autoimmune process without causing significant side effects, the lifelong consequences of these disorders might be avoided. For autoimmune diseases associated with significant morbidity, such as type 1 diabetes, strategies directed toward immune tolerance should be considered, even if they are associated with moderate risk for complications or side effects.

What causes nodularity in endocrine glands? One or more nodules develop in the thyroid, pituitary, and adrenal glands in about 25% of patients when assessed at autopsy or with the use of sensitive techniques such as ultrasound or magnetic resonance imaging. The detection of "incidentalomas" is becoming commonplace as imaging methods are used more widely. Although these nodules are rarely malignant, they may produce excess hormones. Rarely, somatic mutations are found in nodules, and these may cause clonal expansion. More commonly, the nodules are polyclonal, and the causes are unknown. Understanding this process is likely to provide fundamental insight into neoplasia and create novel treatment approaches.

What regulates appetite and energy expenditure set points? The current obesity epidemic underscores the importance of understanding these metabolic processes to develop risk profiles and new treatments. The hypothalamic control of appetite is gradually being dissected, largely on the basis of mutations associated with severe, early-onset obesity (e.g., MC4R, leptin receptor)[29] and the identification of new appetite-regulatory hormones such as ghrelin. These pathways are potential targets for drugs that suppress appetite and/or increase energy expenditure.

As we look at the last 150 years of endocrinology and consider likely advances over the next decade, we have every reason to expect the pace of discovery to accelerate. Many advances have been driven by new technologies such as RIAs, recombinant DNA technology, and imaging. Current technology focuses on genomics, proteomics, metabolomics, nanotechnology, epigenetics, bioinformatics, and the use of computerized information systems in clinical practice. As scientists and clinicians, we will harness these and other technologies to tackle the many unanswered questions that remain in endocrinology.

KEY POINTS

- The field of endocrinology is dynamic and changes continuously, emphasizing the importance of lifelong learning, rigorous literature review, and participation in seminars and conferences to keep up with the latest advances.

REFERENCES

1. Bayliss WM, Starling EH. The mechanism of pancreatic secretion. *J Physiol*. 1902;28:325–353.
2. Medvei VC. *A history of endocrinology*. MTP Press: Hingham, MA. 1982.
3. Clinical Society of London. *Report on Myxoedema*. Francis A. Countway Library of Medicine: Boston. 1888.
4. Ord WM. On myxoedema, a term proposed to be applied to an essential condition in the "cretinoid" affection occasionally observed in middle-aged women. *Medico-Chirurgical Trans.* 1978;61:57–78.
5. Murray G. Note on the treatment of myxoedema by hypodermic injections of an extract of the thyroid gland of a sheep. *Br Med J*. 1891;2:796–797.
6. Bliss M. *The Discovery of Insulin*. The University of Chicago Press: Chicago. 1982.
7. Von Mering J, Minkowski O. Diabetes mellitus nach pankreasextirpation. *Arch Exp Path Pharmakol*. 1890;26:371–387.
8. Collip JB. The original method as used for the isolation of insulin in semi pure form for the treatment of the first clinical cases. *J Biol Chem*. 1923;55:50–51.
9. Collip JB. The extraction of a parathyroid hormone that will prevent or control parathyroid tetany and which regulates the level of blood calcium. *J Biol Chem*. 1925;63:395–438.
10. Sutherland EW, Oye I, Butcher RW. The action of epinephrine and the role of the adenyl cyclase system in hormone action. *Recent Prog Horm Res*. 1965;21:623–646.
11. Tata JR, Widnell CC. Ribonucleic acid synthesis during the early action of thyroid hormone. *Biochem J*. 1966;98:604–620.
12. Berson SA, Yalow RS. Radioimmunoassays of peptide hormones in plasma. *N Engl J Med*. 1967;277:640–647.
13. Means JH. Historical background of the use of radioactive iodine in medicine. *N Engl J Med*. 1955;252:936–940.
14. Astwood EB. Treatment of hyperthyroidism with thiourea. *J Am Med Assoc*. 1943;122:78–81.
15. Thorner MO, McNeilly AS, Hagan C, et al. Long-term treatment of galactorrhoea and hypogonadism with bromocriptine. *Br Med J*. 1974;2:419–422.
16. Loubatieres A. *These doctorat no. 86. Sci naturelles*. Montpellier: Montpellier, France. 1969.
17. Pincus G. Control of conception by hormonal steroids. *Science*. 1966;153:493–500.
18. Lamberts SW, Uitterlinden P, Verschoor L, et al. Long-term treatment of acromegaly with the somatostatin analogue SMS 201–995. *N Engl J Med*. 1985;313:1576–1580.
19. Itakura K, Hirose T, Crea R, et al. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*. 1977;198:1056–1063.
20. Seeburg PH, Shine J, Martial JA, et al. Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. *Nature*. 1977;270:486–494.
21. Ullrich A, Shine J, Chirgwin J, et al. Rat insulin genes: construction of plasmids containing the coding sequences. *Science*. 1977;196:1313–1319.
22. Fiddes JC, Goodman HM. Isolation, cloning and sequence analysis of the cDNA for the alpha-subunit of human chorionic gonadotropin. *Nature*. 1979;281:351–356.
23. Feng D, Lazar MA. Clocks, metabolism, and the epigenome. *Mol Cell*. 2012;47:158–167.
24. Minokoshi Y, Kahn CR, Kahn BB. Tissue-specific ablation of the GLUT4 glucose transporter or the insulin receptor challenges assumptions about insulin action and glucose homeostasis. *J Biol Chem*. 2003;278:33609–33612.
25. Katrich V, Cherezov V, Stevens RC. Structure-function of the G protein-coupled receptor superfamily. *Ann Rev Pharmacol Toxicol*. 2013;53:531–556.
26. Kelly DP. Introduction to the nuclear receptor review series. *Circ Res*. 2010;106:1557–1558.
27. Brennan MD, Miner KM, Rizza RA. The mayo clinic. *J Clin Endocrinol Metab*. 1988;83:3427–3434.
28. Biller BMK, Swearingen B, Zervas NT, et al. A decade of the Massachusetts General Hospital neuroendocrine clinical center. *J Clin Endocrinol Metab*. 1997;82:1668–1674.
29. List JF, Habener JF. Defective melanocortin 4 receptors in hyperphagia and morbid obesity. *N Engl J Med*. 2003;348:1160–1163.

2

Principles of Endocrine Measurements

Hershel Raff, Joely Straseski, and Brian Keevil

OUTLINE

Introduction and Brief History of Endocrine Assays, 14
  Radioimmunoassay, 14
  Nonisotopic Immunoassays, 16
  Monoclonal Antibodies, 16
  Structural Assays, 16
  Preanalytic Methods and Harmonization of Assays, 16
Overview of General Concepts, 16
  Preanalytic Variables, 16
  Analytical Parameters, 17
  Clinical Parameters, 17
  Reference Intervals, 17
  General Interferences, 19
  Summary, 19
Competitive Binding Assays, 19
  Principle, 19
  Modern Example Analytes, 21
  Advantages, 21
  Disadvantages, 21
Noncompetitive Immunoassays, 21
  Principle, 22
  Advantages, 22
  Disadvantages, 23
  Special Circumstances, 25
Free Serum/Plasma Hormone Assays, 25
  Methods for Measurement of Free Hormones, 26
  Additional Techniques, 26
  Special Circumstances, 27
Plasma Renin Assays, 27
  Plasma Renin Activity Assay and its Modifications, 27
  Advantages, 28
  Disadvantages, 28
  Direct Renin Concentration, 28
Structural Assays—Mass Spectrometry, 28
  Gas Chromatography—Mass Spectrometry, 28
  Liquid Chromatography—Tandem Mass Spectrometry, 29
  Liquid Chromatography—Tandem Mass Spectrometry Applications, 30
  Multiplexed Liquid Chromatography—Tandem Mass Spectrometry, 30
  Special Circumstances, 31
Summary and Future Directions, 31

INTRODUCTION AND BRIEF HISTORY OF ENDOCRINE ASSAYS

KEY POINTS

- The invention of the radioimmunoassay was a seminal event in the evolution of modern clinical and experimental endocrinology.
- Nonisotopic immunoassays have better safety and throughput than radioactive assays.
- Structural assays (e.g., chromatography and mass spectrometry) represent a major diagnostic advance, particularly in the measurement of small molecules such as steroids.

The accurate measurement of hormones in body fluids is critical to the practice of clinical endocrinology and is necessary for proper diagnosis and monitoring of treatment. Clinicians depend on the reliability, sensitivity, specificity, and reproducibility of the laboratory results they receive. This requires a symbiotic partnership between the clinical endocrinologist and the clinical laboratory.

This chapter will provide practical explanations for clinical endocrinologists to understand the tests that they order and will avoid greatly detailed explorations of method techniques. For more detailed technical descriptions of individual endocrine assays, we recommend the excellent chapter by Sluss and Hayes.[1]

Radioimmunoassay

The seminal event in the genesis of the modern hormone assay was the groundbreaking work of Yalow and Berson in the development of the radioimmunoassay (RIA), for which Dr. Yalow received the Nobel Prize.[2] (Dr. Berson was deceased when this Nobel Prize was awarded in 1977.) Before the development of the RIA, hormone measurements were primarily performed using bioassays and chemical-based assays.[3] These assays were very labor-intensive and slow, required significant technical expertise, and were notoriously unreliable. Examples of early cellular bioassays include the use of ovarian ascorbic acid depletion for the measurement of serum luteinizing hormone (LH)[4] and adrenal ascorbic acid depletion for plasma adrenocorticotropic hormone (ACTH).[5,6] Another example is the plasma vasopressin bioassay, which measured antidiuresis in the water-loaded rat.[7,8] There were no reliable bioassays for steroid and thyroid hormones, so chemical methods such as the fluorometric assay were developed.[9]

The development of the RIA for peptides in the late 1950s liberated us from the need for bioassays and chemical assays; this method was novel in its reliance on the use of antibodies for the specific detection of an analyte. The history of the development of the plasma insulin RIA has been extensively documented and provides a good example of one type of RIA design.[10,11] The basic concept was the incubation of fixed concentrations of radiolabeled human insulin and a polyclonal antibody specific for insulin with increasing concentrations of unlabeled insulin. The unlabeled insulin and radiolabeled insulin compete for the same binding site on the antibody, giving a classic inverse hyperbolic curve after separation of the bound and free radiolabeled insulin and the measurement of the bound to free fraction ratio (Fig. 2.1). Thus, this is called a competitive binding assay, where the final analyte concentration is inversely proportional to the signal produced by the assay. An RIA's sensitivity and specificity depended on the quality of the antibody used and the availability of a pure hormone standard. Also, the serum, plasma, or urine sample often required a tedious and time-consuming extraction step to avoid matrix effects, improve sensitivity, and eliminate as many interfering substances as possible.

Another example of the revolutionary nature of this method was the development of the RIA for measurement of plasma ACTH[12,13] (Fig. 2.1). Notice that the "sweet spot" of the assay ranged from approximately 5 to 80 pg/mL and encompassed the results from a "normal subject" when accounting for the 5-fold dilution of plasma. Also note the ability to differentiate the diurnal nadir and peak in a normal subject and the identification of the very high ACTH result in a patient with primary adrenal insufficiency (Addison disease). The plasma ACTH RIA was the state of the art until it was replaced by the noncompetitive two-site "sandwich" immunometric assay.[14]

RIA methods are reliant upon the production of specific antibodies, taking advantage of amino acid sequence and structural differences in peptides from different species to allow the production of specific antibodies by immunizing animals. However, adapting RIA methods to the measurement of human steroids was challenging, as steroid hormones are

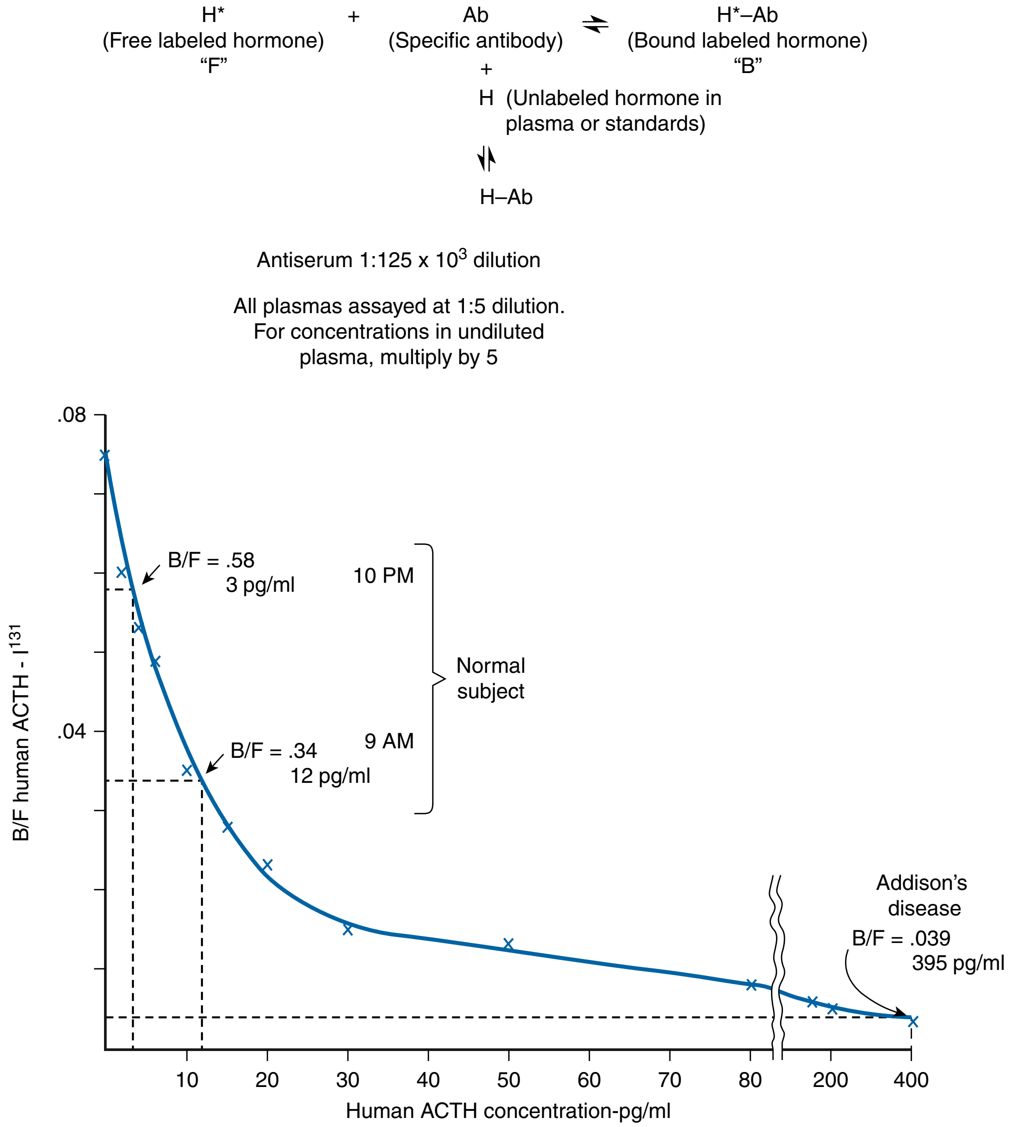


Fig. 2.1 Principle of the competitive binding radioimmunoassay. An unlabeled hormone competes with labeled hormone for binding to a specific antibody. The more unlabeled hormone in the sample, the less labeled hormone is bound to the antibody. Separation and measurement of the bound (B) and free (F) fractions generates a standard curve from which the concentration of the hormone can be interpolated. *ACTH*, Adrenocorticotropic hormone. (Modified from Berson SA, Yalow RS. General principles of radioimmunoassay. *Clin Chim Acta*. 1968;22:51–69.)

not immunogenic by themselves. To develop useful antibodies for steroid hormone assays, the steroid had to be conjugated to immunogenic substances like bovine serum albumin.[15,16] Whereas noncompetitive immunoassays have supplanted competitive binding assays for larger peptides, competitive binding assays based on the original design of the RIA are still used for smaller substances such as steroids and thyroid hormones.

Nonisotopic Immunoassays

There were several reasons for the desire to use methods other than radioisotopes to detect hormone binding. First, there are obvious regulatory and safety issues. In addition, the gamma-emitting isotopes of iodine used for many RIAs (e.g., ^{125}I) have a relatively short half-life, thus limiting their shelf life. The development of alternated methods to label antigen, and subsequently antibody, was obviously needed. The development of the enzyme immunoassay (EIA) and enzyme-linked immunosorbent assay (ELISA) was a major advance in this regard.[16,17] Immobilization of antibody or antigen in a well or tube format also facilitated simple separation of bound from free components, if necessary. The basic method of generating a measurable signal in these assays is via the conjugation of antigen or antibody to an enzyme rather than a radioisotope. Subsequent addition of the appropriate chemicals produces a detectible color change that is proportional (either directly or indirectly, based on the assay design) to the amount of binding. This color change is easily measured with a relatively inexpensive spectrophotometer (i.e., “plate reader”) available in virtually any laboratory. Since then, other methods such as chemiluminescent and fluorometric labels have been developed that improve automation, sensitivity, throughput, and costs, but the basic concepts of the methods are the same.

Monoclonal Antibodies

The quality of an immunoassay is dependent upon the generation of sufficient antibody with high sensitivity and specificity for the antigen (hormone) in question. Polyclonal animal-generated antibodies were the state of the art until the development of monoclonal antibodies from mouse hybridoma cells, for which Millstein and Kohler received the Nobel Prize in 1984.[18] This method allows the generation of large quantities of highly specific antibodies and is currently used for many clinical assays.

Structural Assays

Almost simultaneous with the development of immunoassays was the development of structural methods such as ultraviolet detection and mass spectrometry that measures the ratio of mass to charge of ions.[19-21] The ion trap method of mass spectrometry was developed by Dehmelt and Paul, who received the Nobel Prize in Physics in 1989. Additional advances included the electrospray ionization method developed by Fenn and the soft laser desorption method developed by Tanaka, who shared the Nobel Prize in Chemistry in 2002. Since then, the development of tandem mass spectrometry (MS/MS) has revolutionized the use of this method in clinical laboratories.[22,23] Of course, none of these methods would be useful clinically without the development of high-performance liquid chromatography (HPLC, or LC) in the 1940s through the 1970s.[24,25] The LC-MS/MS method has replaced the slower and more challenging gas chromatography–MS in most clinical applications.[26]

Preanalytic Methods and Harmonization of Assays

Two additional matters merit a brief discussion. The first is the concept of *preanalytic error*.[27] The results generated in the clinical laboratory are only as good as the samples that are provided for analysis. Improper sample tube type (e.g., serum vs. ethylenediaminetetraacetic acid (EDTA) plasma), careless sample handling and processing, and lack of attention to patient preparation (e.g., discontinuation of interfering drugs such as biotin, fasting state, and assuring the correct sampling time) can negate a result generated by even the best analytic method. Any root cause analysis of a problematic result requires assessment of potential preanalytic errors in addition to analytic imprecision.

Secondly, it is important to realize that very few hormone assays have undergone harmonization or standardization, or require accuracy-based quality assessment schemes.[28] This is a costly and time-consuming process and requires cooperation of governmental accrediting bodies and reagent manufacturers.[29,30] Many clinicians are familiar with the variability of results they receive for the same patient from different laboratories using different methods.[31-33] Importantly, this fact reinforces the need to use laboratory-specific reference intervals. When confronted with an unexpected or unexplainable laboratory result, one can confirm that the sample was drawn and processed properly, and, if necessary, repeat the test or refer the sample to an alternate laboratory.

OVERVIEW OF GENERAL CONCEPTS

KEY POINTS

- Preanalytic variables such as patient preparation, sample collection, and sample processing are critical to generate reliable hormone measurements.
- Understanding that reference intervals are specific to the method and laboratory is essential for the proper interpretation of test results.
- Most assays are susceptible to interferences that must be considered when interpreting laboratory results.

Whereas laboratory testing represents a relatively small portion of overall healthcare expenditure (~5%), laboratory test results contribute to an estimated 70% of overall critical decision-making, including diagnosis, medication management, and admitting/discharging decisions.[34] The “brain-to-brain loop,” first published by Lundberg in 1981,[35] identified the numerous steps that occur between clinicians ordering a test and their interpretation and eventual action on a result: collection, identification, transportation, preparation, analysis, and reporting. Appreciation of issues surrounding the integrity of the sample itself, the analytical and clinical performance of a laboratory measurement, comparison of patient results to reference values, and known assay interferences are all critical to interpreting the final result.

Preanalytic Variables

Patient Preparation: Any test result is only as good as the sample used for the measurement. This simple, yet often overlooked, fact cannot be overstated. A patient’s posture, diet, medications and supplements, fasting status, stress, race, gender, activity level, and hydration status, as well as the sampling time and collection method, can all have physiologic effects on certain analytes or may interfere with their analytic measurement.

Stress is known to increase ACTH, cortisol, and catecholamine concentrations.[36-38] Exercise before sample collection can increase the concentrations of hormones such as prolactin, growth hormone, cortisol, glucagon, and catecholamines.[37,39] Moving from recumbent to upright positioning decreases overall plasma volume and concentrates protein as fluid moves toward the lower extremities.[40] This hemoconcentration affects protein-bound analytes such as calcium and sex steroids.[40] Food intake or fasting status naturally affects blood glucose and lipid concentrations, but also gastrin, insulin, renin, aldosterone, and cortisol concentrations.[41] These situations should be limited or noted during the sample collection process to aid eventual interpretation of results.

Timing of sample collection is particularly critical for endocrine analytes. A common example is measuring analytes associated with female reproduction and the menstrual cycle. Many hormones also have a daily rhythm. It is critical that clinicians be very specific about timing when placing a laboratory test order and educating the patient about the importance of sampling at the correct time. A rule of thumb is that the sample should be obtained at the circadian peak if a hormone deficiency is suspected and at the circadian nadir if hormone excess is suspected. For example, evaluation of male hypogonadism should measure testosterone in a morning blood sample (circadian peak), whereas salivary cortisol for the diagnosis of Cushing syndrome must be performed on samples collected at bedtime (circadian nadir).[42-44] Although the time of collection should be reported along with each laboratory result, making this problem obvious in retrospect, it saves a lot of time and confusion if the sample is collected at the correct time in the first place.

Avoiding errors in patient identification is central to the patient safety mission of most institutions, because of the potential negative outcomes. Updated sample collection protocols and electronic solutions are examples of methods used to limit these errors.

Sample Type and Collection: The sample types that can be used for laboratory measurements are almost limitless. Each sample type has limitations, challenges in collection and processing, and certain clinical benefits.

Serum is the aqueous portion of whole blood remaining after clot formation, whereas plasma is the aqueous portion remaining when an anticoagulated blood sample is separated into aqueous and cellular portions. Whole-blood samples are exposed to anticoagulant but not separated. Plasma and serum differ in their composition; for example, serum does not contain fibrinogen (consumed by clot formation) and has slightly higher potassium concentrations (released by platelets during clot formation) than plasma.

Urine samples may be collected at a single time (random "spot" urine) or over a full 24-hour period to normalize for variation in secretion and clearance over time.[45] For 24-hour collections, it is best to measure urinary creatinine simultaneously to verify an adequate collection and control for variations in urine flow rates.[46,47]

Saliva or oral fluid is considered an ultrafiltrate of the plasma and represents the "free" or unbound portion of some analytes. The most common use of saliva samples is for the measurement of cortisol in the diagnosis of Cushing syndrome (late-night salivary cortisol) and during cosyntropin stimulation testing in patients with poor venous access.[38,48] The advantages are particularly notable for cortisol measurements; the noninvasive sampling of saliva and the ability to perform sampling at home in a low-stress environment reduce the chance of artificially elevated cortisol values because of unfamiliar surroundings and overall stress.[49] In fact, a new, noninvasive cosyntropin stimulation test has been developed with nasal ACTH administration and salivary cortisol sampling.[50] The two main potential problems with home saliva sampling are the inability to verify that the sample was obtained at the correct time and assurance that the sample has not been contaminated (e.g., with topical steroids, blood).[51]

For blood samples, different tubes and collection devices are available, and their use depends on factors such as the analyte(s) and sample type(s) to be measured, how quickly a sample must be tested, and the laboratory methods that will be used. Most frequently encountered are blood collection tubes that contain additives to either accelerate (serum) or prevent (plasma) clot formation; however, special tubes may also be required for uses such as preserving specimens for cell counts, trace metals or toxicological assessments, or blood cultures. Additives and preservatives used in some collection tubes may interfere with laboratory measurements (e.g., lithium heparin additive will interfere with lithium measurements, and potassium EDTA chelates calcium and magnesium). Therefore, tubes are collected in a specific order to avoid contamination of additives and preservatives from one tube to the next.

Analytical Parameters

Reliable test results start with reliable methods. The clinician's dependence on accurate test results dictates the need for stringent evaluation of a method before its use. The eventual clinical application of a test result, along with the robustness of the analytical method, help drive the evaluation and acceptance criteria used. Most frequently encountered laboratory tests are either commercially developed or developed by an individual laboratory (often referred to as a Laboratory-Developed Test). Each requires review of different criteria to determine acceptable performance. Metrics associated with both analytical and clinical performance are reviewed and compared against the amount of error that would be considered clinically acceptable.

Analytical performance metrics include imprecision, accuracy, sensitivity, and specificity. Imprecision refers to the reproducibility of an individual result. It estimates the random error, or dispersion around the mean, associated with the analytic error of the method. In other words, if you were to test the same sample multiple times, what amount of variation would you expect? The standard deviation (SD) and coefficient of variation associated with those measurements provide a quantitative measure of imprecision. Accuracy describes the difference between the "true" value and that obtained by your method; that difference describes the bias of the testing method. Any existing bias between your method and whatever is established as "truth" (e.g., a gold standard or reference method) can be determined by comparison, recovery, or interference studies. Figure 2.2 illustrates imprecision and accuracy in terms of a bullseye, with the center of the bullseye representing "truth." Data points that are close to each other are considered precise, and those that are close to the center of the bullseye are considered accurate. The best clinical laboratory tests are both precise and accurate.

Analytical sensitivity studies determine the lower limit of detection for a particular assay, whereas analytical specificity evaluates whether similar compounds may cross-react with your method. These measures should not be confused with clinical sensitivity and specificity, described later.

Clinical Parameters

The clinical diagnostic accuracy of a laboratory test requires comparison of it to an index, or reference, method of diagnosis; this may be another well-defined laboratory test or a more direct measure of disease, such as biopsy. Comparisons between the index test and the test in question reveal "true" (index and testing methods match) and "false" (index and testing methods do not agree) positive and negative results. Mathematical calculation of those individual values provides the test's clinical sensitivity (ability to accurately provide a positive result in the presence of disease) and clinical specificity (ability to accurately provide a negative result in the absence of disease). When the prevalence of a disease is factored in, positive and negative predictive values can also be calculated, which describe the overall reliability of a positive or negative value.

Reference Intervals

Reference intervals are defined as the range of values observed in a reference population for the analyte of interest. Most laboratory results are accompanied by analyte- and laboratory- or method-specific reference intervals which allow for comparison of the patient's result to the expected results obtained from a similar population. Intervals are typically generated from a population of "healthy" individuals whose age, sex, race, and/or ethnicity represent the population of interest.

Identifying an appropriate reference population is complex and central to defining an appropriate reference interval. Defining "healthy" in respect to the analyte you are measuring and obtaining samples from "healthy" patients can be particularly challenging. Strictly defined inclusion and exclusion criteria are needed for

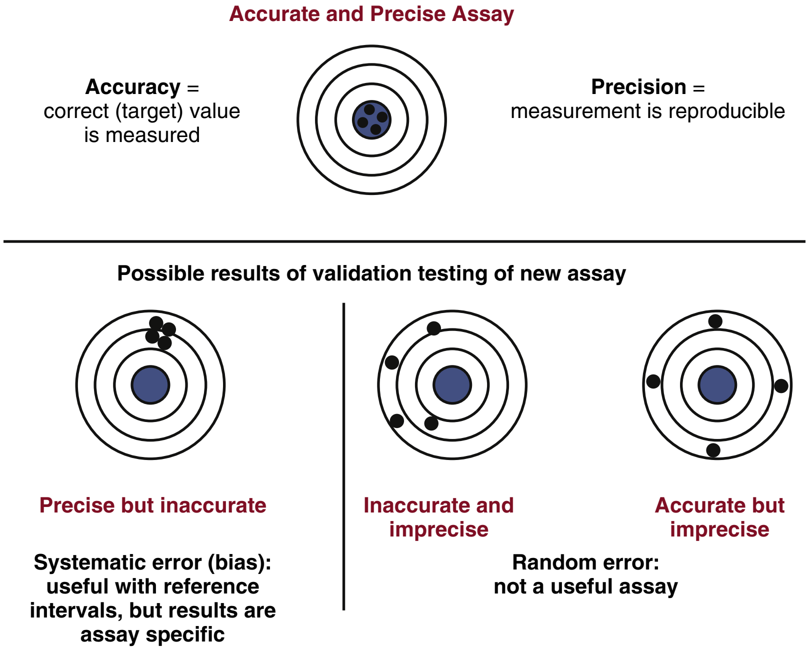

Accurate and Precise Assay
Accuracy = correct (target) value is measured
Precision = measurement is reproducible
Possible results of validation testing of new assay
Precise but inaccurate
Inaccurate and imprecise
Accurate but imprecise
Systematic error (bias): useful with reference intervals, but results are assay specific
Random error: not a useful assay

Fig. 2.2 Assay accuracy and precision are related parameters that are used to assess the quality and reproducibility of an assay. (From Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, Goldfine, RJ, Koenig, AB, et al., eds. *Williams Textbook Of Endocrinology.* 14th ed. Philadelphia, Elsevier; 2020:62–90.e66.)

the reference population to ensure the resulting values provide an accurate picture of the analyte of interest. An example of the need for a well-defined reference population is the reference interval for thyroid-stimulating hormone (TSH). If not screened carefully, the reference population may contain individuals with subclinical thyroid disease.[52-54] Established intervals from those individuals may thus be skewed, resulting in possible misdiagnosis of clinical patients once implemented.

For certain subpopulations, identifying a true reference population is challenging at best and impossible at worst. Examples include scenarios where samples from healthy individuals are extremely limited in availability or volume, such as pediatric populations. Samples are rarely taken from healthy children; if samples are obtained, it is usually as part of a clinical investigation, or the child is ill. That makes it very difficult to obtain samples to determine pediatric reference intervals. National and international efforts have attempted to address this gap by establishing well-characterized repositories of samples from healthy children and using these to develop subsets of pediatric reference intervals.[55-58] Another extremely challenging example relevant to endocrinology is reference intervals for dynamic (suppression or stimulation) testing. These investigations are performed in individuals suspected of having disease. To replicate that scenario for the purposes of a reference population, suppressing or stimulating substances would need to be administered to healthy individuals which would not routinely be approved by modern institutional research review boards. Another example is reference intervals for analytes measured in body fluids or extracted tissue. These sample types are associated with pathology and would not be obtained from healthy individuals to measure reference intervals.

Reference intervals are often partitioned into groups that have statistically distinct values. Examples include estradiol and testosterone in adult men versus women or bone markers and creatinine in children versus adults. In endocrinology, many analytes are partitioned by sex, age, menstrual cycle phase, trimester of pregnancy, Tanner stage in children, etc. There are differences between races for measures such as hemoglobin A1c, with Black Americans having significantly higher values than non–Black Americans.[59] Establishing reference intervals *de novo* requires many healthy individuals, so the need for multiple partitions rapidly increases the number of volunteers or samples needed.

Samples are often obtained from laboratory personnel for reference interval determinations. This approach does have limitations. One example is determining reference intervals for analytes that are dependent upon ambulatory status or stress such as plasma catecholamines. Slight increases in plasma catecholamine concentrations are a common finding in patients without pheochromocytoma.[60] This is likely resulting from the stress of a having a blood sample collected from a patient compared to a healthy volunteer used to having routine blood draws performed.

There are a few analytes (e.g., hemoglobin A1c, cholesterol, 25-hydroxyvitamin D) for which the laboratory provides a "desired" or "optimal" level rather than a reference interval. These desired levels are typically opinions of experts, based on clinical and actuarial data, and can lead to controversy.[61,62]

A common mistake is confusing the terms "reference interval" and "normal range." There really is no such thing as a normal range; reference intervals should be viewed as a guidance, because they rarely distinguish between health and disease on their own. Consider a patient with Cushing disease in whom a plasma ACTH concentration within the reference interval is clearly not "normal." The preferred term in this case is "inappropriately not suppressed" for the degree of hypercortisolism. Another example is a serum TSH concentration within the reference interval in a patient with hypothyroidism, that is, "inappropriately not increased" for the degree of hypothyroidism that can lead to the diagnosis of secondary hypothyroidism (i.e., hypopituitarism). These concepts are challenging to explain to nonexperts. Unfortunately, the archaic term “normal range” is ingrained in clinicians.

Statistics applied to the data generated from the reference populations define the lower and upper limits of these intervals. Parametric statistics are used when the data follow a Gaussian (normally distributed or bell-shaped) curve, whereas nonparametric statistics are not limited by the data’s distribution and rank the values from lowest to highest. Because many biological analytes are skewed (non-Gaussian) in their distribution, guidelines for determining reference intervals recommend the use of nonparametric statistics.[63] Establishing a reference interval *de novo* also requires a considerable number of individuals. Statistical significance requires a minimum of 120 individuals per partition to establish a new reference interval; a smaller number is required if a laboratory is verifying an existing interval.[63] Reference intervals are frequently based on the central 95% of measured values (mean ±2 SD for normally distributed analytes, 2.5th to 97.5th percentiles for nonnormally distributed analytes), which naturally indicates that 5% of values obtained from a healthy population would not be represented in a typical reference interval. All laboratory methods have imprecision associated with their measurements; thus, clinical results that are very close to the reported reference limits should be interpreted within the context of the clinical situation, as these reference intervals are not absolute cutoffs.

General Interferences

Each assay method is susceptible to different types of interference, which will be discussed for individual methods later in this chapter. However, there are some general concepts that are important.

Patients may be taking a medication or supplement that interferes with a particular laboratory test. This could be an analog of a hormone, which would falsely elevate the measurement of that hormone. A classic example is prednisone (prednisolone) cross-reactivity in serum total cortisol immunoassays.[64] This is typically not a problem in current structural assays for cortisol (i.e., LC-MS/MS) because of the assay design. There are also drugs that can falsely decrease a hormone result. A relatively common example is related to the ACTH stimulation test used to evaluate the cortisol response in a patient suspected of having adrenal insufficiency. Baseline assessment of the patient’s plasma ACTH concentration is needed to determine their ability to produce ACTH.[38] However, incorrectly collecting the sample after injecting the standard 250-μg cosyntropin/synacthen ($ACTH_{1-24}$) dose occurs fairly frequently, particularly in hospitalized patients. Because plasma $ACTH_{1-24}$ concentrations are enormously high after injection, collecting the sample after cosyntropin administration paradoxically leads to an abnormally low plasma $ACTH_{1-39}$ result. This is because of the design of the noncompetitive two-antibody sandwich assay used to measure ACTH concentrations (discussed later in this chapter).[65] This could result in the erroneous diagnosis of hypopituitarism (suppressed plasma ACTH) rather than primary adrenal insufficiency (increased plasma ACTH). A third, now well-known, interfering substance biotin, which is used in mega-doses for cosmetic reasons or as a treatment for multiple sclerosis[66,67] or other disorders.[68,69] Biotin can either falsely decrease or increase the hormone result in methods using biotin-streptavidin conjugation; the effect is dependent upon the assay design.[70] These interferences will be discussed where appropriate as this chapter proceeds.

Another common problem occurs in the measurement of serum total thyroid and steroid hormones, which normally circulate in a large part bound to plasma proteins. Obviously, an increase in binding proteins (e.g., increase in corticosteroid-binding globulin [CBG] and thyroxine-binding globulin [TBG] by estrogen therapy or contraceptives) or decrease in binding proteins (e.g., during critical illness, nephrotic syndrome) can confuse the total hormone measurement.[36,71,72] This will be discussed in the section on free hormones later.

Summary

A laboratory test is only as reliable as the sample on which it was performed. Similarly, interpreting a result requires a well-characterized and reliable reference interval. The possibility of interferences with laboratory measurements should be considered when results are incongruous with the clinical scenario. Practical approaches to these situations include repeating the sample collection with close attention to preanalytic variables or, depending on the analyte, requesting measurement by a different method, if possible. These clinical decisions should be made in consultation with a clinical laboratory director familiar with the testing that was performed.

COMPETITIVE BINDING ASSAYS

KEY POINTS

- Competitive binding assays typically use an antibody to bind the unknown hormone of interest in competition with a labeled form of the hormone.
- Simplicity and high-throughput automation are the main advantages of the competitive binding assay.
- Disadvantages include a limited analytic range and cross-reactivity to and interference by other factors, homologous hormones, and metabolites.

The competitive binding assay (typically performed by immunoassay) revolutionized clinical endocrinology and was briefly described in Figure 2.1. The key to the performance of this assay is the use of a moiety that has specific binding sites for the analyte to be measured. Typically, this is a specific polyclonal or, subsequently, monoclonal antibody (hence the term “*immuno*assay”). However, there are other less common approaches such as the use of corticosteroid-binding globulin (CBG; transcortin) as the *in vitro* binding moiety.[73] The use of a single antibody dictates the need for a small target; epitope size, overall confirmation, and folding of the target must allow for binding by a single antibody.

Principle

Figure 2.3 depicts the general design of the competitive binding assay. The (unlabeled) patient sample is mixed with a constant and limited amount of labeled analyte (antigen); analyte-specific antibody is added, and the unlabeled and labeled analytes compete for binding to the same (limited) antibody binding sites. The amount of labeled analyte bound to the antibody is inversely proportional to the amount of unlabeled analyte present in the patient sample. Hormones are the most frequent analytes or antigens tested in endocrinology. However, other peptides and molecules may be measured by the same methods.

The requirements for a successful competitive binding assay are as follows:

1. *Antibody with excellent sensitivity and specificity for the analyte of interest.* Antibodies specific to the analyte of interest are traditionally raised in animal hosts. The sensitivity of the antibody must be sufficient to measure the target hormone at physiological or even subphysiological concentrations. Plasma or serum from the animal is typically used at a very large dilution (e.g., 1:250,000), so it may be used for many assays and many years. Monoclonal antibodies produced in large amounts are now commonly used for this purpose.

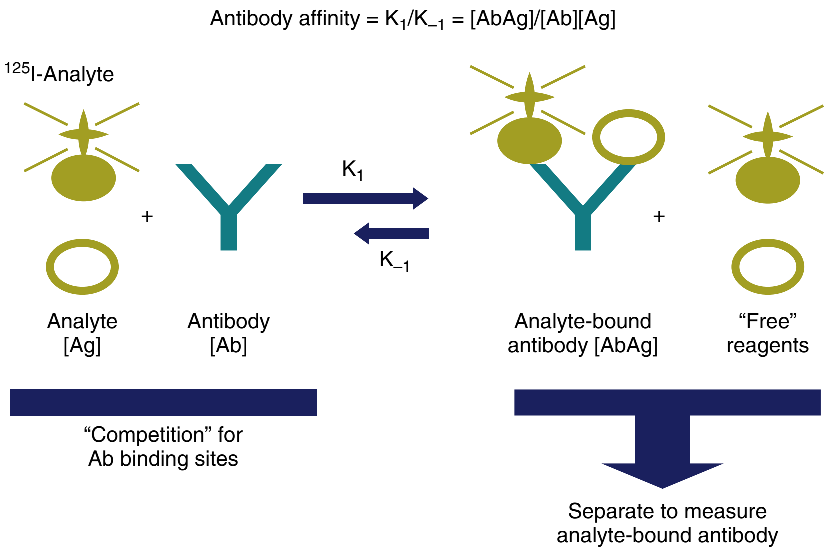

Antibody affinity = K_1/K_{-1} = [AbAg]/[Ab][Ag]
^{125}I-Analyte
+
K_1
K_{-1}
+
Analyte [Ag]
Antibody [Ab]
Analyte-bound antibody [AbAg]
"Free" reagents
"Competition" for Ab binding sites
Separate to measure analyte-bound antibody

Fig. 2.3 Components of a radioimmunoassay as an example of a competitive binding assay. *Ag,* Antigen; *Ab,* antibody. (From Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, Goldfine, RJ, Koenig, AB, et al., eds. *Williams Textbook of Endocrinology.* 14th ed. Philadelphia, Elsevier; 2020:62–90.e66.)

2. *Pure hormone to use for the standard curve* (i.e., "unlabeled hormone" in Fig. 2.1; see also Fig. 2.4). In the early days of these assays, hormone for this purpose was typically purified directly from tissue (e.g., pituitary gland). With few exceptions, highly purified hormone is now produced via chemical synthesis. The hormone is diluted to concentrations encompassing the subphysiological to pathophysiological range to generate a standard (calibration) curve against which the unknown (clinical sample) is compared, and the result derived by interpolation. A "zero" concentration tube or well is typically included in the assay to assess maximum binding of the labeled hormone to the antibody in the absence of unlabeled ("competing") hormone (see left corner of the standard curve in Fig. 2.1).
3. *Labeled hormone.* The pure hormone described in #2 must be labeled in some way so it can be subsequently detected. The original RIA used radioactive isotopes to label the hormone. Although radioactive labels are still used in specific cases, many modern assays employ nonradioactive labels such as enzymes or fluorescent or luminescent compounds. Biotin-streptavidin binding is used by some methods, primarily to increase assay specificity and sensitivity.[74] One conjugated streptavidin molecule can tightly bind four biotin "signal" molecules to separate bound from free hormone, thus amplifying the final signal.
4. *Patient sample.* In early assays, the biological materials being tested had to undergo a chemical extraction and occasionally a chromatographic separation before measurement. This was particularly true if the antibody had significant cross-reactivity with similar substances, or if the matrix, or sample itself, contained compounds that could affect the kinetics of the assay. One example of such a matrix effect is plasma proteins. Many modern assays have overcome these matrix effects and now allow measurement directly in the biological material from the patient. When assaying for "total hormone" in the sample (total hormone refers to the combination of hormone bound to plasma proteins (e.g., steroids and thyroid hormone) and unbound hormone), the hormone in the biological sample must be "released" from the binding protein before testing, which can be done in several ways (e.g., addition of chemicals or heat).
5. *Incubation.* The antibody and labeled hormone are added in the same concentration and volume to each assay tube or well. Standards of known concentrations or patient samples are added in the same volume to each tube or well. This combination is mixed well and incubated at the optimal temperature and time to promote antibody–antigen binding; these parameters are determined during initial assay development.
6. *Separation.* Labeled hormone bound to the detection antibody is separated from any remaining unbound (i.e., free). There are a number of ways to do this, including adsorption to activated charcoal, use of a secondary antibody raised against (specific to) the first antibody (typically anti-IgG of the species in which the antibody was raised), or immobilization of the antibody on a solid support, such as the surface of a test tube, well, or bead.
7. *Measurement and calculation.* Once separation is complete, the amount of antibody-bound label present in the tube or well is measured. In some cases, both bound and free labeled hormone are measured from the same tube, and their ratio is used as the signal value (see y-axis of Fig. 2.1). The final step is to compare the bound labeled hormone in the standards to the amount present in the patient sample. The simplest way to conceptualize this is by looking at the dotted lines in Figure 2.1. A bound-to-free labeled hormone ratio of 0.34 interpolated to a plasma concentration of about 12 pg/mL in this case. The final plasma concentration of ACTH in this example would be 60 pg/mL, after accounting for the 1:5 plasma dilution before testing.

There are many variations on this theme, including the way the hormone is labeled and detected, how the antibody is generated, and the approach to separating labeled hormone bound to the antibody from free. However, the inverse, hyperbolic standard curve is ubiquitous. Mathematical algorithms can now fit the standard curve to an equation such that the result is automatically calculated. There is also typically a correction for nonspecific binding (tubes/wells incubated without antibody).

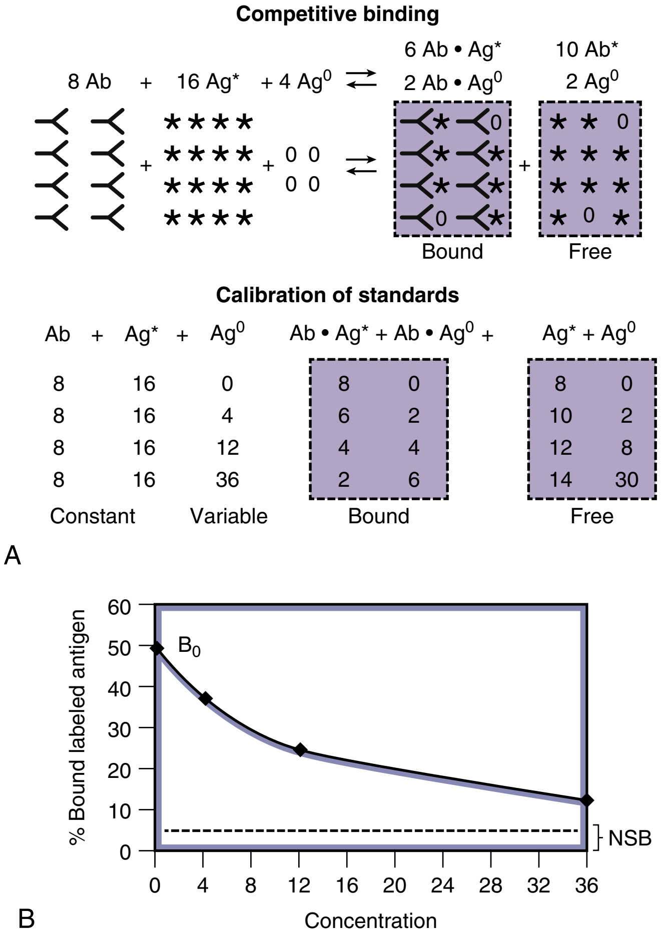


Fig. 2.4 Details about competitive binding assays. **A,** Competition of the labeled antigen (Ag*, in endocrinology, typically a hormone) and unlabeled antigen (Ag^{o}, either a calibrator or unknown sample) for a specific antibody (*Ab*) to the antigen (*Ag*) generates a dose (calibrator) response curve. **B,** B_0 represents maximum binding of Ag* in the absence of Ag^{o}. Nonspecific binding (NSB) is binding to anything except Ag. (From Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, Goldfine, RJ, Koenig, AB, et al., eds. *Williams Textbook of Endocrinology.* 14th ed. Philadelphia, Elsevier; 2020:62–90.e66.)

Modern Example Analytes

Many clinical assays no longer use a single antibody competitive binding approach, particularly in the case of peptides and proteins. If the hormone is large enough to have at least two epitopes against which two different antibodies can be generated, the noncompetitive immunoassay is preferable. Smaller peptide hormones, steroids, and thyroid hormones are still measured using competitive binding assays. This difference is critically important because the interpretation of the final concentrations and the interferences to which these assays are susceptible are quite different from the noncompetitive assays that will be described in the next section.

Advantages

The main advantage of the competitive binding assay is its simplicity. The basic assay design can be employed by RIA, chemiluminescent, enzyme immunoassay, and other methods. Simple, inexpensive colorimetric plate readers make the use of EIA/ELISA accessible to small laboratories. Automation on large instrument platforms provides fast and reliable results. The availability of monoclonal assays has made the development of novel competitive binding assays simple and relatively inexpensive.

Disadvantages

1. *Standard curve shape and analytic range.* Looking again at Figures 2.1 and 2.4, one can see that the dynamic range of the assay is limiting. At low unlabeled hormone concentrations, the binding of the labeled hormone can be very close to the zero calibrator. Many remember when early generations of the TSH assay could not differentiate a low-normal result and a truly low result (as in Graves disease).[75] At high concentrations, the curve becomes very flat, such that a small error in binding translates to a large error in determining the final hormone concentration. The range of 20% to 80% binding is the "sweet spot" of the curve. Low hormone concentrations may require extracting a larger volume of sample to increase assay sensitivity, whereas high concentrations may require error-prone dilutions of the sample.
2. *Cross-reactivity and interferences.* Some competitive binding assays employ the relationship between biotin and streptavidin. Biotin supplementation at supraphysiological concentrations may interfere with binding and final measurement of the analyte of interest.[76-79] Both competitive binding and noncompetitive immunoassays that use biotin-streptavidin interactions can be affected by exogenous biotin; however, they differ in the observed effects on analyte concentrations. Streptavidin is often used to bind complexes to a solid support or microparticle (Fig. 2.5). Excess biotin in the patient sample may bind the streptavidin, thereby preventing biotin-labeled antibodies from binding and creating further complexes with the analyte of interest. The detectable signal will subsequently be low. Competitive binding assays exhibit an inverse relationship between signal and analyte concentration; thus, the analyte concentration will be erroneously high. Biotin supplementation should be ruled out if laboratory results are not congruent with the clinical scenario. Note that not all assays use biotin-streptavidin interactions for analyte detection, so communication with the testing laboratory is necessary when investigating biotin interference.

The presence of cross-reacting antibodies may interfere with these types of assays. Clinical scenarios such as autoimmune disease,[80,81] alterations in protein binding concentrations,[82] and antireagent antibodies[83-86] may all contribute to assay interference. Heterophilic antibodies may also interfere with competitive immunoassays but affect noncompetitive immunoassays more frequently and are discussed later.

NONCOMPETITIVE IMMUNOASSAYS

KEY POINTS

- Two antibodies specific to different, noncompeting epitopes on the antigen (hormone) are used to label, capture, and specifically measure the analyte of interest.
- The main advantages of this approach compared to competitive immunoassays are improved specificity, sensitivity, efficiency, and availability of monoclonal antibodies.
- Its main disadvantages are interference by other factors (e.g., endogenous antibodies, biotin) and similar analytes, and a "high-hook" effect in which very high concentrations of the analyte can falsely lower the result.

Noncompetitive immunoassays, also known as immunometric assays, are primarily used to measure larger polypeptides.[3,18] The most common type of noncompetitive immunoassay is the two-site, or "sandwich," assay. These assays were originally developed as

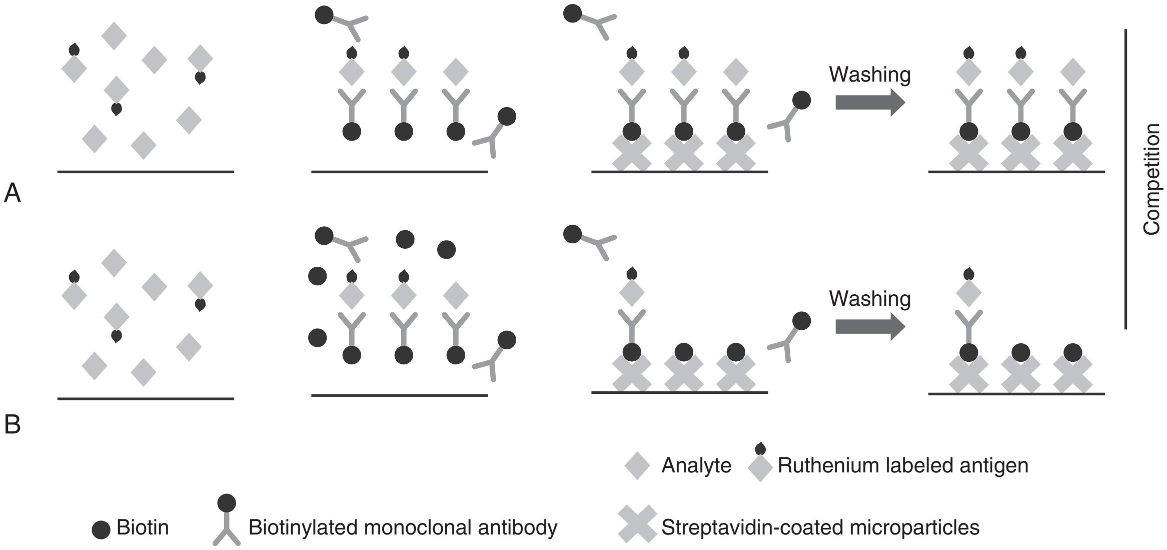

A
Washing
B
Washing
Competition
Analyte
Ruthenium labeled antigen
Biotin
Biotinylated monoclonal antibody
Streptavidin-coated microparticles

Fig. 2.5 Mechanism of biotin interference with a competitive immunoassay. **A,** Sample without significant biotin contamination; signal is inversely proportional to analyte concentration. **B,** Excess biotin in the sample takes up streptavidin binding sites, thereby blocking further complex formation. Decreased signal is generated, leading to falsely increased analyte concentrations. Note that not all competitive immunoassays use biotin for signal generation. (Modified from Li J, Wagar EA, Meng QH. Comprehensive assessment of biotin interference in immunoassays. *Clin Chim Acta.* 2018;487:293–298.)

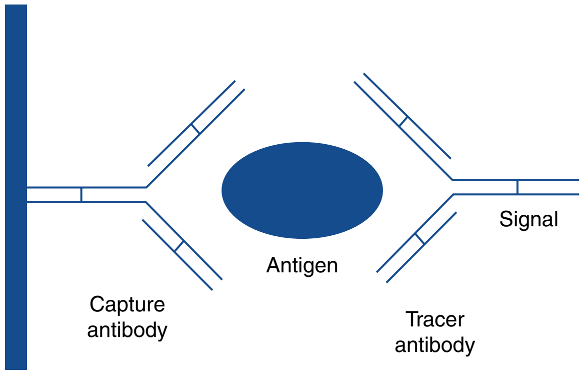


Fig. 2.6 Noncompetitive immunoassay (informally known as sandwich assays). Noncompetitive immunoassays use a capture antibody fixed to a solid phase and a tracer antibody, which has a measurable signal fixed to it, in a liquid phase. Ideally, each of these antibodies will bind a different epitope on the antigen (i.e., hormone of interest). (From Salah AK, Gharad SM, Bodiwala K, et al. You can assay that again! *Am J Med.* 2007;120:671–672.)

immunoradiometric assays, although few laboratories continue to use radioactivity as the tracer signal.[14,87]

Principle

The concept is relatively straightforward (Fig. 2.6). The analyte of interest—here called the "antigen"—is typically a polypeptide hormone or protein and must be of sufficient size (e.g., number of amino acids) to contain at least two epitopes against which two different antibodies can be generated. Antibodies may be raised against the antigen (e.g., hormone) of interest by either polyclonal/purification techniques, or, more commonly, monoclonal antibody technology (see discussion earlier in this chapter). One of the antibodies (specific for one epitope) is immobilized and is called the "capture" antibody. Mechanisms of immobilization include adsorption onto larger particles, wells, or tubes, or precipitation out of solution. The second antibody (specific for a different epitope) is in the liquid phase and is labeled so it can be detected. It is therefore called the "tracer" antibody. Frequently used labels include fluorescence, chemiluminescence, or coupled enzyme reactions. The capture and tracer antibodies are present in excess.

A standard (calibration) curve for a noncompetitive immunoassay is composed of a series of increasing concentrations of the antigen (e.g., hormone or other analyte) just as it is for competitive binding assays (Figs. 2.1 and 2.4). After incubation (for as little as 10–20 minutes!), the unbound tracer antibody is washed away, and the captured antigen is quantified via the bound tracer. The relationship between antigen and labeled antibody is directly proportional; the more hormone/analyte in the reaction, the greater the signal (Fig. 2.7). This is because the amount of "sandwich" (capture antibody–antigen–tracer antibody) formed increases in direct proportion to amount of hormone/analyte (the "meat" inside the antibody "sandwich"). Note that this is the opposite of the relationship described for competitive binding assays. Examples of endocrine analytes that may be detected by noncompetitive immunoassays are prolactin, TSH, and ACTH.

Advantages

1. *Improved specificity.* Noncompetitive immunoassays are typically more specific than competitive binding assays. This is because of the use of two antibodies in this assay design: only the full-length antigen of interest binds to both the capture and tracer antibodies. A similar antigen or antigen fragment may be bound by one (capture or tracer), but not both, and therefore would not be detected. This is particularly relevant when measuring hormones (see ACTH discussion later).
2. *Improved sensitivity.* Noncompetitive immunoassays allow for almost complete capture of antigen present, which results in superb sensitivity and analytic measurement range. Use of monoclonal antibodies in assay design also contributes to the enhanced specificity, and this combination provides better sensitivity overall.
3. *Linear standard curve.* The standard (calibration) curve (Fig. 2.7) is more linear than those of competitive binding assays (see

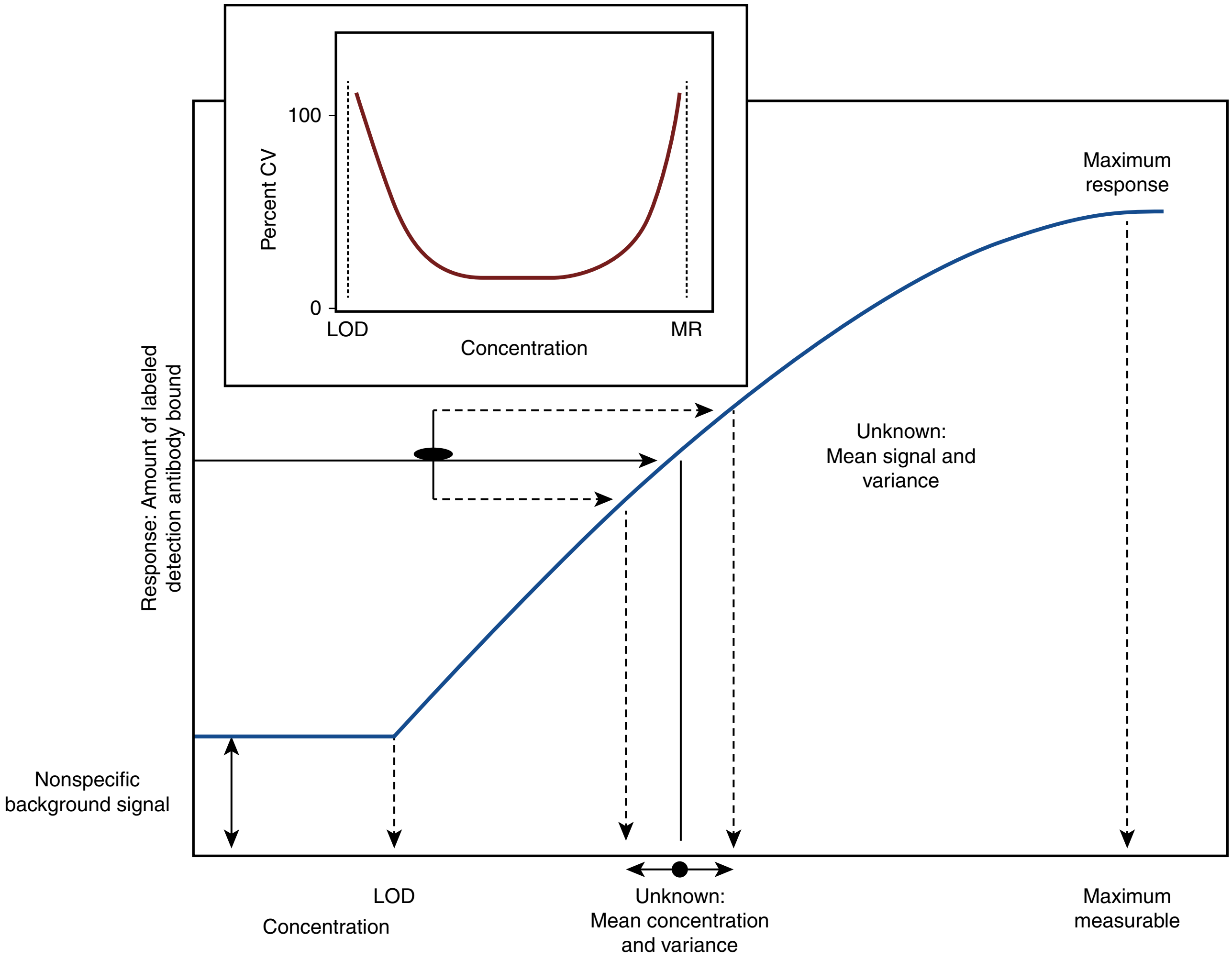


Fig. 2.7 Characteristics of the signal generated by a noncompetitive immunoassay. Signal generated by the amount of detection antibody bound to the capture–analyte complex is directly proportional to the concentration of analyte in a noncompetitive immunoassay. The concentration can be extrapolated from response (signal measured) by measuring known concentrations of "calibrators." As shown in the inset, the variance associated with measurement is heteroscedastic and increases significantly as the upper or lower limits of the assay are approached. *LOD*, Limit of detection; *CV*, coefficient of variation; *MR*, maximum response. (From Sluss PM. Methodologies for measurement of cardiac markers. *Clin Lab Med*. 2014;34:167–185, viii.)

Figs. 2.1 and 2.4), so the data reduction scheme is simpler and more reliable and has less error at the extremes of hormone concentrations. This translates into better sensitivity and analytic measurement range.

4. *Improved efficiency*. Noncompetitive immunoassays are typically fast and easily automated. These benefits can affect turnaround times and overall patient care, particularly when used for applications such as intraoperative PTH measurements during minimally invasive parathyroid adenomectomies.[88]
5. *Monoclonal antibodies are plentiful.*

Disadvantages

1. *Cross-reactivity with similar compounds*. Although noncompetitive immunoassays have improved specificity compared to competitive binding assays, immunoassays are only as specific as the antibodies they use. Cross-reactivity with similar compounds (e.g., similar hormone structure, fragment of an intact polypeptide, metabolite) can result in either positive (falsely elevated concentration) or negative (falsely low concentration) interference (Fig. 2.8). Some well-known examples include human chorionic gonadotropin assay detection of intact and/or subunits, insulin assay detection of endogenous hormone compared to exogenous analogs, and PTH

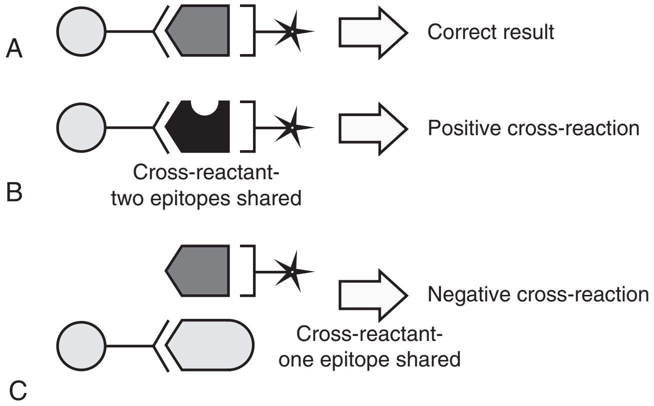


Fig. 2.8 Interfering cross-reactions in noncompetitive, two-site (sandwich) immunoassays. **A,** Correct result. **B,** Cross-reactant binds to both antibodies (shares both epitopes with antigen). **C,** Cross-reactant only binds one of the two antibodies (shares only one epitope with the antigen). (From Sturgeon CM, Viljoen A. Analytical error and interference in immunoassay: minimizing risk. *Ann Clin Biochem*. 2011;48:418–432 and From Seth J, Sturgeon CM. Pitfalls in immunoassays in endocrinology. Endocrinology and Metabolism In-Service Training and Continuing Education. 1993;11(4):89–99. with permission from the American Association for Clinical Chemistry, Inc.)

assay detection of biointact molecule or fragments. Another good example of negative assay interference by fragments is the measurement of plasma ACTH after injection of ACTH$_{1-24}$ (cosyntropin/synacthen) for ACTH stimulation testing. The high concentration of ACTH$_{1-24}$ can bind one of the assay antibodies, essentially removing it from the final "sandwich" and causing falsely low results.[65]

2. *Antibody interferences.* Patients may produce antibodies against components that are used in the immunoassays.[83-86] This interference is particularly noted for the noncompetitive, sandwich immunoassay design, but interference with competitive assays has also been documented.[89] Human antimouse antibodies in patient samples are the most frequently encountered type of human anti-animal antibody (HAAA). HAAAs may result from monoclonal antibody therapies and imaging modalities or from workplace or hobby exposure (e.g., to mice). Heterophilic antibodies, in contrast, are directed against poorly defined antigens and have weaker avidity for their target than HAAAs (Fig. 2.9).[90] Interference by either antibody type may cause *either* false positive or false negative results, depending on the specificity of the antibody. Their binding may allow for bridging across the capture and tracer antibodies (falsely elevated concentration) or may block further binding or sandwich formation (falsely low concentration). Many assays include nonimmune animal serum to "soak up" any HAAA or heterophilic antibodies that may be present in a patient sample and avoid further assay interference. However, antibody interference can still occur if they are present in large enough quantities.
3. *Biotin interference.* The ability of supraphysiological concentrations of biotin to interfere with assays that depend on biotin-streptavidin interactions for analyte detection was introduced in the competitive binding assay section. Excess biotin in a patient sample may bind to streptavidin reagent found on microparticles or solid supports (Fig. 2.10). This blocks binding of biotin-labeled antibody and further complex formation. In other words, when considering noncompetitive immunoassays, the "sandwich" cannot be

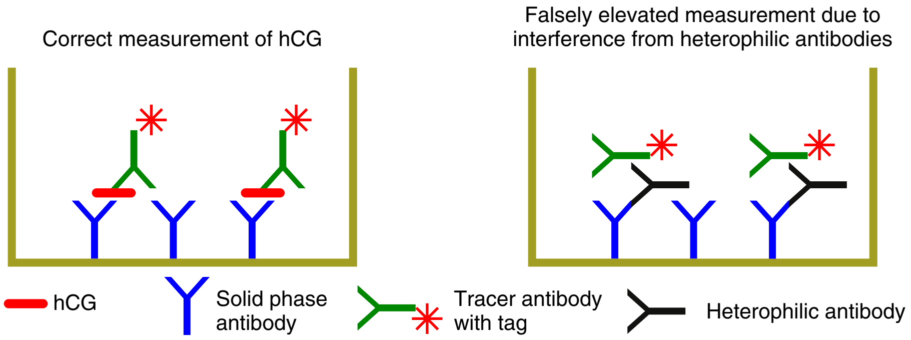

Correct measurement of hCG
Falsely elevated measurement due to interference from heterophilic antibodies
hCG
Solid phase antibody
Tracer antibody with tag
Heterophilic antibody

Fig. 2.9 Heterophilic antibody interference in noncompetitive immunoassays. *Left*: A schematic illustration of a noncompetitive immunoassay for human chorionic gonadotropin (*hCG*). *Right*: Falsely elevated result caused by interference from heterophilic antibodies. (From Bolstad N, Warren DJ, Nustad K. Heterophilic antibody interference in immunometric assays. *Best Pract Res Clin Endocrinol Metab*. 2013;27:647–661.)

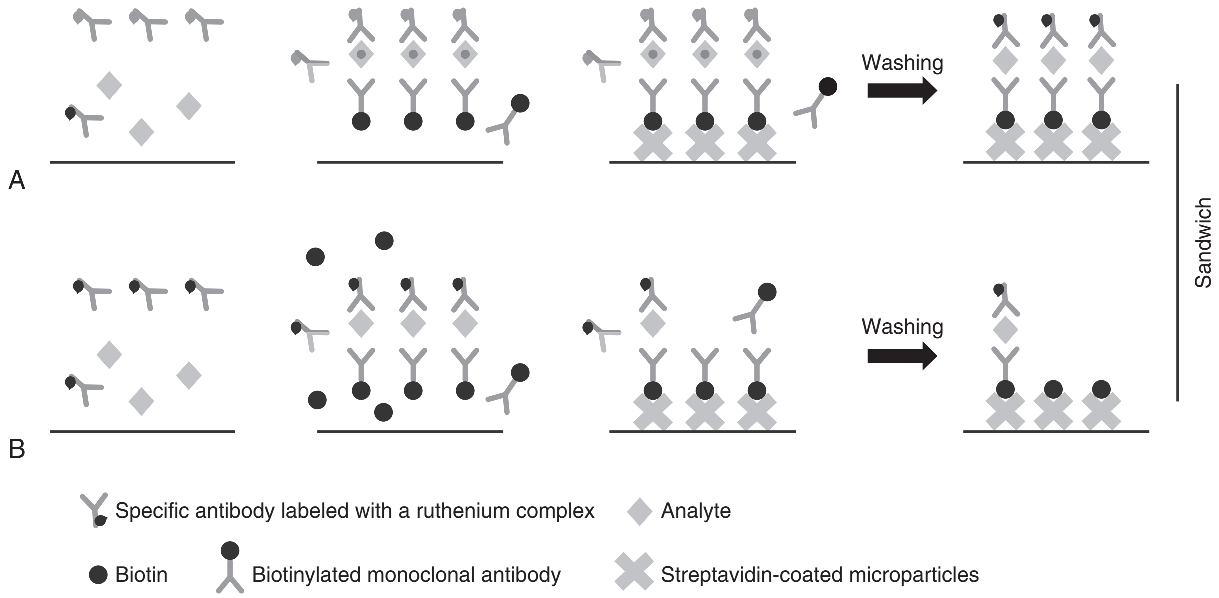

A
B
Washing
Washing
Sandwich
Specific antibody labeled with a ruthenium complex
Analyte
Biotin
Biotinylated monoclonal antibody
Streptavidin-coated microparticles

Fig. 2.10 Mechanisms of biotin interference with a noncompetitive (two-antibody sandwich) immunoassay. **A,** Sample without significant biotin contamination; signal is directly proportional to analyte concentration. **B,** Excess biotin in the sample takes up streptavidin binding sites, thereby blocking further complex formation. Decreased signal is generated, leading to falsely decreased analyte concentrations. Note: not all noncompetitive immunoassays utilize biotin for signal generation. (Modified from Li J, Wagar EA, Meng QH. Comprehensive assessment of biotin interference in immunoassays. *Clin Chim Acta*. 2018;487:293–298.)

formed, and the resulting signal will be low. Because of the direct relationship between signal detection and analyte concentration for noncompetitive immunoassays, the final analyte concentration will be falsely low. As with competitive binding assays, biotin supplementation should be ruled out if low laboratory results are not congruent with the clinical scenario. Note that not all assays use biotin-streptavidin interactions for analyte detection, so communication with the testing laboratory is necessary when investigating biotin interference.

4. *"High-dose hook" effect.* The assumption of the design of the noncompetitive immunoassay is that both the capture and tracer antibodies are in excess, so the more analyte in the reaction, the more "sandwich" is formed and detected. However, if the analyte concentration is dramatically increased (typically >10-fold higher than the highest standard (Fig. 2.11)), the amount of antibody can become limiting, and the assay essentially converts to a competitive binding assay. Paradoxically, this results in a (falsely) lowered result, even though very high amounts of the analyte are present in the patient specimen. This can be resolved by serially diluting the sample to lower the analyte concentration to the point where the two antibodies are now in excess (~1000 miU/mL in Fig. 2.11). Overall, this is a rare occurrence because commercial immunoassay manufacturers test for assay susceptibility to hook effect and alter their assay formulations accordingly. Although it is theoretically possible that the hook effect could lower the result into the reference interval, we have never observed this in the many assays we have performed.

Special Circumstances

Noncompetitive immunoassays are not currently amenable to the measurement of smaller molecules like steroids and thyroid hormones, because they are too small allow for generation of two antibodies to distinct parts of the molecule. As a result, immunoassays for steroids, thyroid hormones, and small peptides (e.g., oxytocin) are still measured by competitive binding assay.

The noncompetitive immunoassay design permits rapid turnaround times. The rapid PTH assay is used to intraoperatively verify parathyroid adenoma removal during minimally invasive parathyroidectomies.[88] A similar modification has been made to the noncompetitive ACTH immunoassays, allowing rapid verification of intraoperative corticotropin-secreting tumor removal.[91,92]

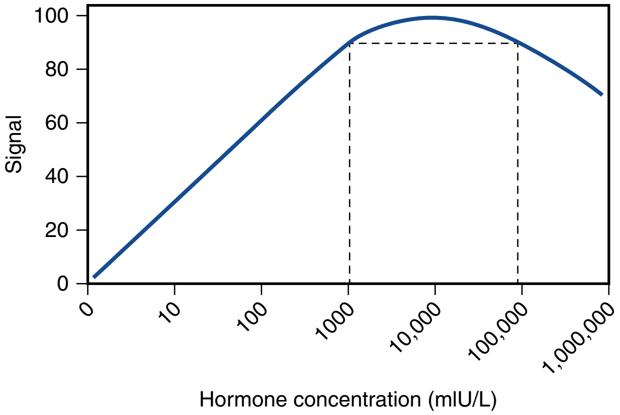


Fig. 2.11 High-dose hook effect. The response signal reaches a maximum and then decreases when the antigen concentration exceeds the limit of the noncompetitive immunoassay. (From Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, Goldfine, RJ, Koenig, AB, et al., eds. *Williams Textbook of Endocrinology.* 14th ed. Philadelphia, Elsevier; 2020:62–90.e66.)

Not all noncompetitive immunoassays are created equally. Each vendor may use a different set of proprietary antibodies, which can lead to discordant results between methods. For example, one noncompetitive immunoassay platform has been shown to frequently give falsely increased plasma ACTH levels, leading to unnecessary testing and even unnecessary surgery.[31-33]

A very interesting application of noncompetitive immunoassays is the use of combinations of antibodies to different epitopes on larger peptides and proteins to evaluate the presence of precursor peptides. An example is the evaluation of patients with the potential for high circulating levels of proopiomelanocortin and its posttranslational products (i.e., ACTH precursors). These typically arise from neuroendocrine tumors (e.g., small cell lung cancer) that can cause ectopic ACTH syndrome.[93] A research group has developed antibody combinations that can quantify these ACTH precursors,[87] thus enabling early detection and treatment. Unfortunately, this assay is not commercially available at this time.

FREE SERUM/PLASMA HORMONE ASSAYS

KEY POINTS

- Steroid and thyroid hormones circulate in the blood mainly bound to plasma proteins, with a small fraction in the free (biologically active) form.
- Measurement of free serum thyroid hormones and testosterone are particularly useful for clinical care.
- Methods that physically separate the free hormone from the protein bound are the gold standard.
- Calculation of the free hormone from the serum concentration of total hormone and binding protein concentration may be useful but should be used with caution.

Catecholamines and most peptides and protein hormones circulate in the plasma compartment in the dissolved (free) state. Direct measurement in serum/plasma therefore accurately reflects their biological activity. Conversely, hydrophobic thyroid and steroid hormones circulate mostly bound to proteins that have varied specificity, capacity, and affinity for the hormone (Table 2.1). The unbound (free) form binds its receptor on target cells to exert its biological effects and comprises a smaller fraction of the overall concentration. The extreme example

TABLE 2.1 Specific Plasma Binding Proteins for Steroid and Thyroid Hormones

Major Specific Binding Protein	Hormone
Corticosteroid-binding globulin (CBG)[a]	Cortisol and corticosterone
CBG and sex hormone–binding globulin (SHBG) [a,b]	Progesterone
CBG [a]	Aldosterone
SHBG[b]	Testosterone and estrogen
Vitamin D–binding protein	25OHD and 1,25(OH)$_2$D
Thyroid-binding globulin (TBG) and transthyretin	Thyroxine
TBG	Triiodothyronine

Note: All of the hormones listed bind albumin to some extent (low-affinity/high-capacity binding protein).

[a] CBG binds cortisol with higher affinity than progesterone and aldosterone.

[b] SHBG binds testosterone with higher affinity than progesterone and estrogen.

is thyroxine, which is more than 99.5% bound to plasma proteins in the circulation.[94,95] Overall, up to 90% to 95% of any given steroid hormone may actually be protein-bound. Binding proteins may be specific, high-affinity carriers (Table 2.1) or low-affinity proteins with large binding capacity (e.g., albumin, prealbumin).

Most standard assays for the measurement of steroid and thyroid hormones determine their total concentration in serum/plasma; these "total" measurements assume that the concentrations and kinetics of binding proteins are normal. There are several circumstances in which that assumption is not valid, including obesity, throughout the aging process, and during critical illness. Some additional scenarios include estrogen therapy, birth control pills containing estrogen, and pregnancy, all of which increase corticosteroid-binding globulin (CBG) concentrations and thyroxine-binding globulin (TBG) concentrations. Critical illness may decrease serum binding. Hyperthyroidism induces hepatic production of sex hormone–binding globulin (SHBG), which binds both testosterone and estrogen. These scenarios make total hormone measurements suspect and may affect reference intervals and dynamic testing cutoffs.[48,71] For example, the cutoff for a normal baseline or $ACTH_{1\text{-}24}$–stimulated serum/plasma cortisol can be significantly increased in the context of estrogen-induced increases in CBG.[96]

Serum/plasma–free hormone assays have been developed because of the issues described and because measurement of "free" hormones, at least in theory, provides the best evidence for changes in the bioactivity of the endogenous serum/plasma hormones. Some have also advocated for using salivary steroid measurements as a valid surrogate for measurement of serum free hormone concentrations.[48] Free hormone assays can be technically challenging and therefore may have limitations that must be appreciated when interpreting results in the context of the clinical scenario. Their challenging nature also dictates that free hormone tests are frequently performed in reference or specialty laboratories.

Methods for Measurement of Free Hormones

The most frequently used methods for estimating free hormone concentrations are physical separation of free from protein-bound hormone and calculations based on binding constants. Additional or historical techniques include selective precipitation, calculations including ratios, and direct antibody-based binding (immuno) assays targeted toward free hormone. Overall, these latter techniques are not recommended, and concerns regarding their use will be described later.

Free hormone is typically present in very small concentrations; thus, sensitive measurement techniques are essential. An appreciation for the complexity of measuring free steroid hormones is also critical for accurate result interpretation. Protein binding kinetics, competitive binding, and equilibrium with other binding proteins all make these measurements particularly challenging.

Physical Separation - Equilibrium Dialysis and Ultrafiltration: Physical separation of free from protein-bound fractions allows for direct measurement of the biologically active hormone. Physical separation methods such as equilibrium dialysis (ED) and ultrafiltration (UF) have long been considered the gold standard for these measurements. The analytic principle of ED is based on two solution-filled chambers separated by a semipermeable membrane; the patient specimen is located in one chamber, and an appropriate dialysis buffer is located in the other (Fig. 2.12). The membrane pores are of a sufficient diameter to allow the unbound hormone to pass through, while retaining serum-binding proteins and bound hormone. After a period of time, the free hormone from the patient sample will come into equilibrium between the two chambers. Measuring the amount of hormone that has passed from the patient sample into the dialysis buffer chamber provides the free hormone concentration. UF similarly separates free from bound hormone via a semipermeable membrane but uses centrifugation to force the molecules through. Both techniques are routinely paired with high-sensitivity MS-based methods for the final measurement of free hormone. Temperature, pH, and protein adsorption to or leakage through the membrane can all have a negative impact on the results from the ED and UF methods. Importantly, the time allowed to reach equilibrium (often 12–24 hours!) and any dilution of the sample may have a significant influence on ED measurements.

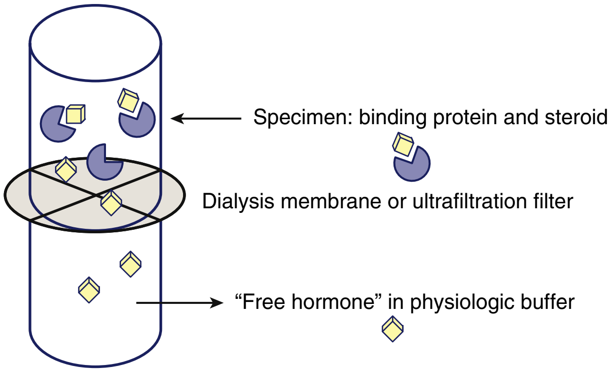

Specimen: binding protein and steroid
Dialysis membrane or ultrafiltration filter
"Free hormone" in physiologic buffer

Fig. 2.12 Free hormone assay design using a dialysis membrane (often referred to as equilibrium dialysis). This allows separation of free hormone from hormone bound to plasma/serum protein. (From Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, Goldfine, RJ, Koenig, AB, et al., eds. *Williams Textbook of Endocrinology.* 14th ed. Philadelphia, Elsevier; 2020:62–90.e66.)

Calculation of Free Hormone Fraction: Testosterone provides the best example of estimated free (and/or weakly bound to albumin, or "bioavailable") fractions by use of calculations. Formulas for the determination of free estradiol, cortisol, and others also exist[97,98] but are not in routine use. The most frequently used formulas use measured values of (1) the hormone of interest, (2) its main binding protein (e.g., SHBG, CBG), and (3) albumin, along with their respective binding constants, to determine unbound versus bound concentrations. The accuracy of the calculated value is dependent on the accuracy of the individual measurements and the binding constants used. Automated methods exist for most of the individual components that would be included in these calculations; therefore, this approach is often more accessible for the typical hospital laboratory. Several free testosterone formulas have been published, and many studies report good correlation with ED methods.[99,100] However, differences in measurement, formulas, and constants used can lead to significant disparities among methods, and calculated values might best be considered estimates of true free hormone concentrations.[101,102] Direct (noncalculated) measurements are recommended in cases with known or suspected alternation in binding proteins.[30]

Additional Techniques

Selective precipitation of SHBG from a patient sample using ammonium sulfate has been described for the determination of free testosterone concentrations.[103] This relies on the addition of radiolabeled testosterone to the patient sample. After ammonium sulfate is added and the sample is centrifuged, the supernatant contains all non–SHBG-bound fractions (free and bioavailable). The nonbound radioactive testosterone is measured, and that percentage of the total testosterone concentration approximates the amount of bioavailable testosterone present in the sample. This method is not in routine use, largely because of the availability of ED and UF methods.

Direct (analog) immunoassay methods have been used for the measurement of free testosterone; however, these assays are unacceptable because of problems with bias, reproducibility, and hormone recovery.[29,30,99,100]

The free androgen index (FAI) is the ratio of testosterone to SHBG multiplied by 100 to obtain a percentage. Although it is a simple calculation, caution should be exercised when interpreting the FAI. Like calculations for free testosterone, accuracy of the individually measured components is central to the overall utility of the calculated value. Although there is good correlation between the FAI and physical separation techniques in females, this relationship is not ideal and not useful in males. Yet, the performance of FAI in females has been used to identify and manage patients with polycystic ovary disease.[104]

Special Circumstances

Comment on Free Vitamin D Metabolites: Like other steroid hormones described earlier, 25 hydroxyvitamin D (25OHD) circulates bound to plasma proteins. Vitamin D–binding protein (DBP) carries approximately 90% of the total plasma vitamin D.[105] Albumin carries approximately 10%, but with much lower affinity than DBP.[105] The loose binding of 25OHD to albumin renders that portion "bioavailable," which is similar to the fraction of testosterone that binds to albumin.[30] Less than 0.05% of plasma 25OHD circulates in the free form.

One may then wonder why good, direct free 25OHD or DBP assays have not been developed to allow determination of free and bioavailable serum 25OHD in a manner similar to calculated free and bioavailable testosterone.[30] This is an area of significant controversy, particularly because changes in DBP were purported to account for differences in total 25OHD observed between Black Americans and White Americans.[106] Subsequent studies questioned the validity of the DBP measurements used in this prior study[107] and, in fact, questioned the need for measuring or calculating free 25OHD in any circumstance.[108]

Comment on Salivary Hormone Measurements as a Surrogate for Free Plasma/Serum Hormones: Salivary cortisol is useful as a surrogate for free plasma/serum cortisol for the diagnosis of disorders of the hypothalamic-pituitary-adrenal axis, particularly Cushing syndrome and adrenal insufficiency.[43,109,110] Salivary cortisol may be measured by immunoassay or LC-MS/MS. For Cushing syndrome screening, immunoassay may have a higher sensitivity because of the cross-reactivity of most antibodies with salivary cortisol metabolites.[111] Immunoassay does have poorer specificity compared to LC-MS/MS, particularly in patients with mild cortisol excess owing to adrenal (ACTH-independent) Cushing syndrome.[111] The measurement of salivary cortisone by LC-MS/MS is useful for identifying contamination of the salivary sample with hydrocortisone[112] and for monitoring the adequacy of cortisol replacement therapy in patients with adrenal insufficiency.[113] Other salivary hormone analyses have been suggested, including testosterone and estrogen in the evaluation of gonadal function and 17-alpha-hydroxyprogesterone in the early evaluation of congenital adrenal hyperplasia because of 21-hydroxylase deficiency.[114-117] Furthermore, a new, completely noninvasive nasal ACTH stimulation test using salivary cortisol as an outcome has been proposed.[50] This could make the evaluation of adrenal insufficiency simpler and available to primary care physicians in any clinical setting.

PLASMA RENIN ASSAYS

KEY POINTS

- Plasma renin can be measured as plasma renin activity (PRA; measurement of angiotensin I generation *in vitro*) or direct renin concentration (DRC) by noncompetitive immunoassay.
- The main advantage of PRA is that it assesses the actual endogenous biological activity of renin.
- The main advantages of DRC are that it is simple and high-throughput, and is not influenced by changes in plasma angiotensinogen concentrations.

Renin is not a hormone in the classic sense: it is an enzyme that catalyzes the cleavage of angiotensin I (a small peptide) from angiotensinogen ("renin substrate"). There is considerable confusion about how plasma renin is measured.

Endocrinologists should take the lead in teaching the medical community the proper pronunciation of the word "*renin*" (i.e., REE-nin; as in "REE-nal" artery). The mispronunciation of the word renin as "WREN-in" is incorrect and imprecise, and it is up to us to correct this common (annoying) error.

Plasma Renin Activity Assay and its Modifications

Plasma renin *activity* (PRA) is measured by a hybrid assay that is part bioassay (generation of angiotensin I *in vitro*) and typically part immunoassay (the measurement of angiotensin I generated *in vitro*) (Fig. 2.13).[118] The first step is the *in vitro* 37° C generation of angiotensin I from the plasma renin and angiotensinogen contained in the patient's sample. Reagents are added that prevent the breakdown of angiotensin

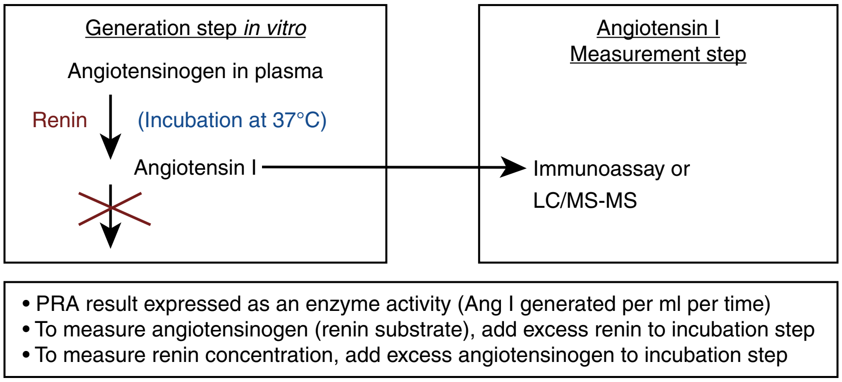

Generation step in vitro
Angiotensinogen in plasma
Renin
(Incubation at 37°C)
Angiotensin I
Angiotensin I Measurement step
Immunoassay or LC/MS-MS
• PRA result expressed as an enzyme activity (Ang I generated per ml per time)
• To measure angiotensinogen (renin substrate), add excess renin to incubation step
• To measure renin concentration, add excess angiotensinogen to incubation step

Fig. 2.13 Simplified version of the plasma renin activity assay. The first step generates angiotensin I from endogenous (patient) angiotensinogen and renin. The X represents inhibitors that are added to prevent degradation of angiotensin I during the generation step. The second step is the measurement of angiotensin I concentration. *LC-MS/MS*, liquid chromatography tandem—mass spectroscopy; *Ang I*, angiotensin 1; *PRA*, plasma renin activity.

I *in vitro*. A parallel sample incubation on ice (4° C) is commonly done to correct for the endogenous angiotensin I in the plasma sample. Once the angiotensin I generation step is complete and the reaction is stopped, the next step involves angiotensin I concentration measurement by immunoassay or LC-MS/MS.[119,120] The result is expressed as a biological activity (e.g., pmol of angiotensin I generated per minutes of incubation per liter of plasma). Because angiotensinogen concentration is in excess in the typical patient sample, the angiotensin I generated (renin activity) is dependent only on the concentration of renin in the sample. The PRA assay can be modified to measure plasma angiotensinogen "concentration" (Fig. 2.13). This is accomplished by adding renin in excess in the incubation step.[119] Plasma renin concentration can be measured by adding excess angiotensinogen in the incubation step.[119] In both cases, the measurement of angiotensin I concentration generated is the final step.

Advantages

The main advantages of the PRA assay are (1) the decades of experience with the interpretation of the results and (2) assessment of the actual bioactivity of renin within the sample.

Disadvantages

The PRA assay is tedious and time-consuming. If there is an alteration in angiotensinogen concentration, it can become rate-limiting in the generation step. This can also be considered an advantage because if angiotensinogen is rate-limiting *in vitro*, it is likely to be rate limiting *in vivo*.

Direct Renin Concentration

With the development of the two-site immunometric (noncompetitive) assay, it is now possible to directly measure plasma renin concentration.[121] This method is fast and inexpensive and correlates well with PRA.[122-124] Many reference laboratories offer both PRA and direct renin concentration because many clinicians are used to PRA in the assessment of primary aldosteronism using the plasma aldosterone to renin ratio.[124,125] The main disadvantage of the method for measuring direct renin concentration is that is it not a bioassay and is unaffected by potentially rate-limiting low concentrations of angiotensinogen in the patient's circulation.

STRUCTURAL ASSAYS—MASS SPECTROMETRY

KEY POINTS

- After an initial physical separation of the compound of interest (typically by liquid chromatography), mass spectrometry is used for detection.
- Mass spectrometry uses transformation of the compound into charged molecular ions, and then separation, fragmentation, and measurement of the ions according to their mass-to-charge ratio.
- Advantages include the ability to measure only the compound of interest (high specificity) and the ability to detect low concentrations (sensitivity).
- Disadvantages include instrument and maintenance costs and the technical expertise required.

Mass spectrometry instruments used in clinical laboratories include two main types: gas chromatography-mass spectrometry (GC-MS) and LC-MS/MS. Both techniques rely on the physical separation of compounds on chromatography columns, the transformation of these compounds into charged molecular ions, and then separation, fragmentation, and measurement of these in the mass spectrometer.[19-23] Ions are separated according to their mass-to-charge ratio (m/z), which is a dimensionless quantity calculated by dividing the mass number of an ion by its charge state.

Gas Chromatography—Mass Spectrometry

GC-MS has been widely used in steroid chemistry and is still the main technique used for the identification and characterization of inborn errors of steroid metabolism.[126] Steroid profiling developed over 40 years ago, opening up the field of urinary steroid profiling, which is now sometimes referred to as urinary steroid metabolomics.[127] GC-MS owes its success to the efficient separating capability of very long fused silica capillary columns coupled with the specificity of a mass detector. With this combination, the identification of structurally-related steroids can be achieved. Chromatographic separation of steroids is crucial, because there are many steroids with similar m/z values that will appear identical to the mass spectrometer. Good examples of this include 11-deoxycortisol, corticosterone, and 21-deoxycortisol, which share a parent ion m/z of 347 and also fragment to give the same daughter ions.[128]

The main problem with GC-MS is the need to chemically derivatize steroids before analysis, because only highly volatile molecular ions will travel through the instrument. To achieve this, derivatized steroids are vaporized and are carried by inert gas through a heated silica column and interact with the chemically bonded stationary phase attached to the walls of the column. Gaseous to stationary phase interactions based on the differences in hydrophobicity of the individual steroids cause differences in retention of steroids on the column, so they arrive at the detector at different times. Derivatization also improves steroid stability and allows optimal sensitivity and chromatographic separation.[129] GC-MS has an advantage over other techniques in that it is able to measure a wide variety of steroids and their metabolites simultaneously within the same analytical run. Run times are typically 30 minutes or longer, and the sample preparation steps are long and tedious, but the data that are eventually generated are unmatched by any other technique, because novel and unknown compounds can be identified.[126] It is the ability to search for previously unidentified steroids and generate metabolic fingerprints from the interrelationship of steroids that make this technique so powerful for the investigation of steroid disturbances.

The main weakness of GC-MS is the inherent complexity of sample preparation, which makes it a difficult technique to use. The introduction of electrospray ionization and improvements in liquid chromatography have propelled LC-MS/MS to the forefront of separation techniques. Indeed, it has largely superseded GC-MS in both routine and research laboratories. Differences between the two techniques are marked, and each have specific advantages. GC-MS confers the ability to screen for all compounds in a sample, whereas LC-MS/MS relies on targeted analysis of known compounds, so in the case of LC-MS/MS, the analyst will only detect what they expect. This is an important consideration because although LC-MS/MS is unmatched for quantitation of targeted steroids, so far it cannot replace GC-MS for discovery research.[26] This is a situation that may change with improvements in LC-MS/MS technology, including supercritical fluid chromatography[26] and the use of accurate mass instruments.[130] However, LC-MS/MS generally does not need derivatization, so sample preparation is simpler, and the run times are much shorter, allowing for faster throughput.

Some sample preparation is still required, because neat serum, urine, or saliva samples cannot be injected directly into the LC-MS/MS system. The main ways of preparing samples to remove proteins and other interfering substances include simple protein precipitation and liquid or solid phase extraction.[131] The latter two techniques are more suitable for high-sensitivity work, where significant sample clean-up, matrix removal, and possibly analyte concentration are needed.[128] The workflow is therefore much simpler than the GC-MS workflow, and this makes LC-MS/MS much more amenable for use in a routine clinical laboratory. However, the high cost of instrumentation and greater technical complexity, requiring highly skilled technical staff, has impeded the wider adoption of LC-MS/MS. As a consequence, some laboratories tend not to use LC-MS/MS and still depend on manual or fully automated immunoassays for many analytes.

Liquid Chromatography—Tandem Mass Spectrometry

LC-MS/MS uses a liquid mobile phase to achieve chromatographic separation on a short packed column. The mobile phase may contain salts or weak acids to enhance ionization in the source, and the column packing material is chemically bonded with a variety of ligands to make it hydrophobic in nature. Separation occurs by hydrophobic interactions between the analyte, the column packing material, and the mobile phase. Samples are injected as a liquid onto the column, and different analytes are separated according to their relative retention on the column and consequently reach the detector at different times. Chromatographic separation is still important to discriminate between compounds with identical mass.[131] Electrospray ionization was a breakthrough, because it converts the column effluent from a liquid containing molecules into gaseous ions, thus enabling their passage through the mass spectrometer. The free ions generated are drawn into the mass spectrometer and separated according to their m/z using quadrupoles, fragmented in a collision cell, and then detected.[132]

Analytical specificity in GC-MS is determined by chromatographic separation on a very long column and measurement of molecular ions, whereas for LC-MS/MS specificity is achieved in a three-stage process. First, some crucial separation is achieved on the much shorter analytical column sufficient to separate compounds with the same m/z. Second, specificity is augmented by selecting the molecular ion formed in the ionization source from other potentially interfering ions using a quadrupole mass analyzer. Third, the molecular ion is selectively fragmented in the collision cell, after which molecular ion fragments are selected using the second quadrupole mass analyzer and then measured in the detector. The capacity to select a parent and a daughter ion after a chromatographic separation makes LC-MS/MS both highly sensitive and specific.

The molecular ion is often capable of forming several abundant fragment ions, and additional quality checks are now commonly incorporated by monitoring a second qualifier fragment ion to assess interference in the method. It is the fragment ions that are measured in the detector, having been produced by selectively fragmenting a molecular ion that has undergone prior separation on a chromatography column. (Fig. 2.14). This may seem complicated, but the instrument is under full software control, and each step is fully automated and highly reproducible. Once set up, the instrument can be left unattended for

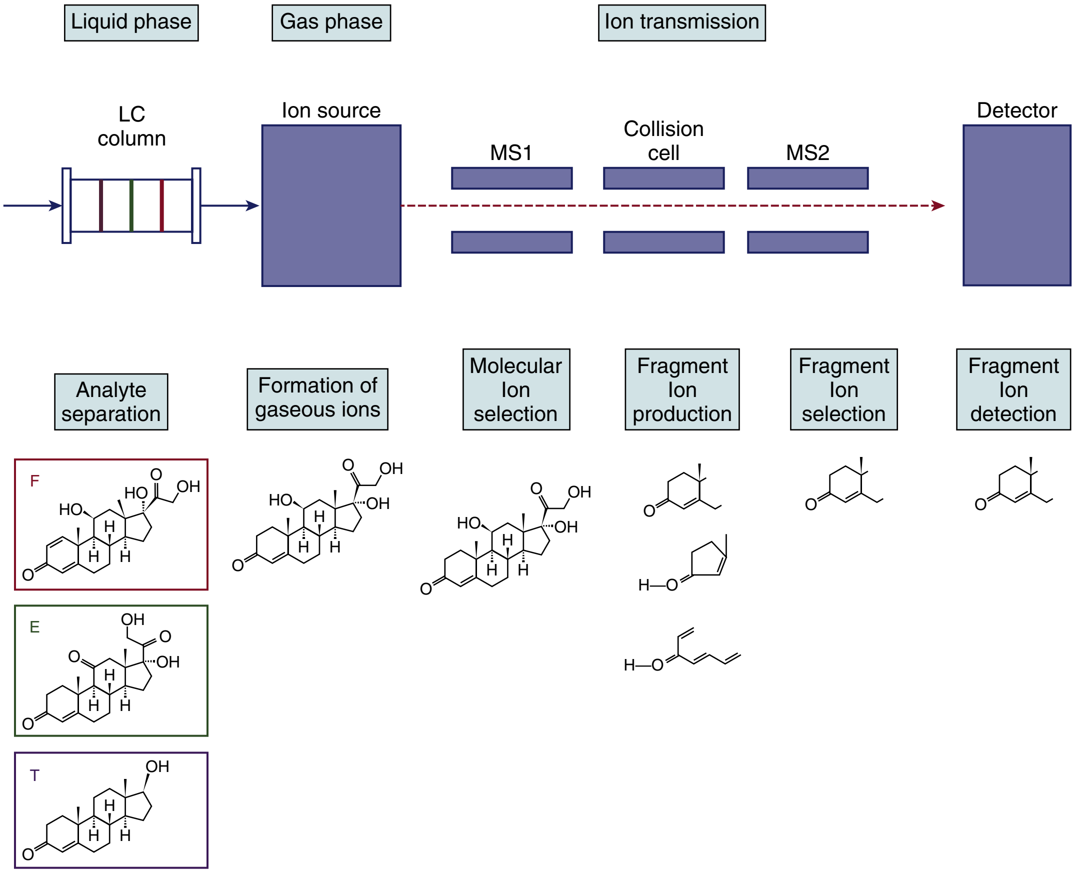


Fig. 2.14 Diagram showing the measurement of cortisol using liquid chromatography tandem-mass spectroscopy. Cortisol (F), estradiol (E), and testosterone (T) are first separated using liquid chromatography. Cortisol elutes first from the column, it is ionized in the mass spectrometer source, and the cortisol ions are selectively transferred through the first quadrupole to a collision cell. Cortisol is fragmented in the collision cell, and an abundant fragment ion is selectively transferred through the second quadrupole for measurement in the detector. *LC*, liquid chromatographer; *MS*, mass spectrometer.

long periods of time, including overnight and weekends, and the results downloaded to the laboratory information system for reporting. Both MS techniques are superior to immunoassay methods in terms of specificity and also because of the sample clean-up steps.

Between-assay variability: LC-MS/MS is well placed for the development of niche assays that are performed less frequently, are unavailable, or perform poorly on immunoassay platforms. Some of these niche assays include the less commonly measured steroids such as 11-deoxycorticosterone (11-DOC), aldosterone, 17-hydroxyprogesterone (17-OHP), dehydroepiandrosterone, and dihydrotestosterone. There is the possibility of an immunoassay being withdrawn abruptly from the market because of quality or regulatory issues. Another important consideration is the reproducibility of immunoassays over time because of calibration or reagent changes, which can cause problems with diagnostic cutoffs in routine laboratory procedures and in longitudinal research studies.[133] However, just as there are poorly performing immunoassays, there are also poorly performing LC-MS/MS methods, and this has been shown in interlaboratory surveys for a number of steroids.[134] The main problems are related to the choices of sample preparation technique, chromatography columns, and internal standards, all of which can hinder minimizing sample matrix effects. Interlaboratory variability can often be improved by standardizing sample clean-up and chromatography conditions, because it is typically not the MS measurement technique that is at fault. Improvements in interlaboratory performance have also been achieved by the availability of commercially available calibration material.[135] The steroids that are infrequently measured, such as 17-OHP, tend to have the worst comparative performance between laboratories.[136] Improvements in interlaboratory variability have been achieved with accuracy-based initiatives such as the Centers for Disease Control and Prevention (CDC) hormone standardization program and vitamin D standardization certification program.[137,138] Central to this strategy is the use of LC-MS/MS to assign target values to single-donor patient serum or plasma samples that can then be distributed to laboratories for analysis. Having a true assigned target concentration allows assessment of bias from the true result in a clinical sample. LC-MS/MS assigned target values have since been adopted by many national external quality assessment schemes, but, as in the CDC program, only for a limited number of steroids.

Liquid Chromatography—Tandem Mass Spectrometry Applications

Cortisol: In most laboratories, cortisol is measured by immunoassay, but LC-MS/MS is a useful alternative technology, as it is not affected by cross-reactivity of synthetic exogenous glucocorticoids such as prednisolone/prednisone, methylprednisolone, and cortisol precursors and metabolites such as 11-DOC. Some cortisol immunoassays are also less effective at freeing cortisol from CBG, as demonstrated by recovery in women taking the oral contraceptive pill.[139] Variability between assays will affect how the cortisol result is interpreted against existing reference intervals, so the clinician should be aware of the assay characteristics of their local laboratory. From a practical standpoint, rapid LC-MS/MS assays for the routine measurement of serum cortisol are now available, with the added benefit of identifying samples that show evidence of exogenous steroid use.[140] LC-MS/MS can also provide the added value of simultaneously measuring dexamethasone with cortisol to determine if a patient has absorbed the drug or metabolized it quickly, when interpreting dexamethasone suppression tests.[141]

Estradiol: Many estradiol (E2) immunoassays perform poorly for the measurement of samples from men, children, and postmenopausal women,[142] because of a lack of sensitivity and the cross-reactivity with some drugs, particularly the estrogen receptor antagonist fulvestrant.[143] Achieving the necessary sensitivity to measure E2 using LC-MS/MS has been challenging, but improvements in instrumentation have enabled the development of assays capable of measuring concentrations as low as 0.6 pmol/L,[144] which meets the required sensitivity of 1 pg/mL (3.7 pmol/L) advocated by the Endocrine Society.[97] It is also important to measure E2 in postmenopausal women taking aromatase inhibitors, because treatment can reduce serum E2 by 85%; such depletion may be an efficacy or safety biomarker of adjuvant aromatase inhibitor treatment.[145]

Vitamin D: Automated immunoassays have proved unreliable in a number of patient groups, including pregnant women, intensive care patients, patients with liver failure, hemodialysis patients, and osteoporotic patients.[146] LC-MS/MS is the recommended method in these patients because of its superior specificity. In addition, immunoassays do not distinguish between 25(OH)D_2 and 25(OH)D_3 and suffer from cross-reactivity with other vitamin D metabolites recognized by the antibody, such as 24,25$(OH)_2$D.[146] Many of the earlier LC-MS/MS methods for vitamin D were developed for speed and sample throughput, and also suffered from some interference. Separation of 25(OH)D_3 from the C3 epimer of 25(OH)D_3 could not be achieved. This may be a problem in neonates, where concentrations can be high but typically return to adult levels of less than 10% of total vitamin D by 3 months of age.[147]

Peptides: Smaller peptides such as angiotensin 1 (Ang 1) are easily measured using LC-MS/MS and can be used in a routine hospital laboratory setting. Larger peptides such as insulin-like growth factor 1 and thyroglobulin offer a greater analytical challenge, because they invariably need to be digested with proteolytic enzymes to produce manageable fragments for the mass spectrometer to measure.[148] Thyroglobulin measurement by LC-MS/MS after protein digestion is the method of choice for many large contract laboratories because it is free from thyroid antibody interference.[149]

As mentioned previously, PRA is typically measured by immunoassay of Ang I generation but may also be measured, following a suitable incubation step, by LC-MS/MS.[120] LC-MS/MS methods provide sensitive interference-free measurement of Ang I and offer a good alternative to immunoassay-based methods.[150]

Multiplexed Liquid Chromatography—Tandem Mass Spectrometry

Another advantage of LC-MS/MS is the capability to simultaneously measure several different analytes to produce multiplexed test panels. Performing multiplexed analysis is often a minimal extra expense, because the sample preparation and LC-MS/MS conditions are the same. Selective testing panels have been developed to investigate different clinical conditions, such as second-tier testing for congenital adrenal hyperplasia (CAH)[151] to supplement newborn screening programs and for the investigation of adrenal insufficiency.[152] Commercially available kits have been developed for the measurement of steroid panels (www.Chromsystems.com), and the introduction of easy-to-use kits has already simplified the introduction and service delivery of LC-MS/MS in routine laboratory procedures.

Multiplexed LC-MS/MS panels have recently allowed a more thorough investigation of conditions such as polycystic ovary syndrome (PCOS)[153] and CAH.[154] There is increasing interest in the measurement of not only testosterone, but other androgens such as androstenedione and the 11-oxygenated steroids such as 11 ketotestosterone (11KT).[155] The evidence for the use of other androgens in screening for CAH is also compelling, and 11KT may prove to be an effective androgenic marker in this condition.[156]

Some have also advocated the measurement of 21-deoxycortisol as a discriminator between 21-hydroxylase deficiency and PCOS in women with hyperandrogenism.[157] Diagnostic uncertainty often arises in these cases, because the clinical presentations of nonclassical CAH and PCOS are similar, and 17-OHP (the traditional marker for CAH) may also be elevated in PCOS.

Multiplexing tests in a single assay may confer benefits, but it can also generate large amounts of complicated data that need to be carefully interpreted. The development of urine steroid metabolomics for discriminating benign from malignant adrenal tumors has shown that this problem can be solved with the use of artificial intelligence.[127,158] Urine steroid profiling, first developed using GC-MS, has been shown to predict adrenocortical carcinoma with 95% sensitivity and specificity and can achieve an early diagnosis. The analysis has since been transferred from GC-MS to LC-MS/MS using fewer steroids in the panel with an increase in sample throughput but, crucially, with no detriment to the test performance. Other diagnostic areas likely to benefit from this approach include the differentiation of adrenal from pituitary Cushing syndrome, and the diagnosis of CAH.[130] There is also some promising work on the use of multiplexed LC-MS/MS assays coupled with machine learning in the investigation of Cushing syndrome[159] and pheochromcytoma.[160] Machine learning represents a great advance that reduces the need for expert highly subjective interpretative skills.

Special Circumstances

Urine used to be the sample of choice for measuring steroids and biogenic amines, because the larger concentrations found in urine were necessary to overcome sensitivity problems in older assays. However, 24-hour urine samples can be challenging for the patient to collect correctly and for the laboratory to process, so some urine tests are gradually being replaced by serum tests (e.g., for metanephrines[161] and 5-hydroxyindoleacetic acid[162]). Indeed, plasma metanephrines are now the method of choice for the investigation of paraganglioma,[163] because metanephrines are continuously secreted by the tumor and are readily measured using LC-MS/MS.

Saliva offers benefits because it is easy to collect, it can be collected by the patient at any time of day and can be collected frequently. However, saliva collection is not without contamination problems, arising mainly from topical steroid use.[112] Metabolism of steroids occurs in the salivary glands, allowing the use of salivary cortisone as a surrogate marker for serum free cortisol.[113] Salivary cortisone is technically easier to measure than cortisol by LC-MS/MS, because it is present in greater than 4-fold concentrations, closely reflects free serum cortisol after adrenal stimulation, and is unaffected by CBG changes. LC-MS/MS analysis of late-night salivary cortisol or cortisone does not improve the already excellent sensitivity of late-night cortisol by immunoassay, but does improve specificity of the test.[111] Post dexamethasone salivary cortisone may be a better discriminatory marker between healthy subjects and patients with Cushing syndrome or with autonomous cortisol secretion than salivary cortisol.[141] In addition, salivary cortisone is an excellent marker for contamination of the saliva sample by topical hydrocortisone (cortisol).[112] Although technically inferior to LC-MS/MS, cross reactivity with other corticosteroids may actually improve the diagnostic performance of immunoassays.[164] Metabolomics programs are beginning to generate new candidate diagnostic markers based on multiplexed LC-MS/MS methods.[165]

Fully automated clinical mass spectrometry analyzers are finally beginning to appear on the market, but currently have a limited test repertoire and are relatively expensive, a position analogous to the introduction of laboratory automation for immunoassay 30 years ago. These instruments are less technically demanding, will bring LC-MS/MS to a wider range of laboratories, and should eventually improve the availability of high-quality tests to the clinician.

SUMMARY AND FUTURE DIRECTIONS

Clinicians depend on accurate and reproducible hormone measurements for the diagnosis, treatment, and management of patients with endocrine diseases. Since the development of the RIA, many improvements have been made that increase throughput and reduce errors. However, it is incumbent on clinicians to be aware of the potential for preanalytic and analytic errors in all measurements obtained. Clinicians should be attuned to the reference intervals for the result obtained, as they can vary significantly between laboratories and methods. When a result is obtained that does not fit with the clinical scenario, clinicians should repeat the measurement and, if needed, work with their clinical laboratory partners to resolve the issue. When possible, referring a sample to another laboratory for analysis by an alternate method may be useful. All assays have strengths and weaknesses, including susceptibility to interferences such as similar hormones and metabolites and other factors such as biotin and heterophile antibodies.

LC-MS/MS is emerging as the method of choice for small molecules—particularly steroids—although this approach is very labor-intensive, requires sophisticated instrumentation and technical expertise, and also has analytic strengths and weaknesses. The development of methods to measure multiple steroid metabolites in a single sample by multiplexed LC-MS/MS has promise to evaluate diseases that are difficult to diagnose and manage, such as adrenocortical cancer and PCOS. The development of accuracy-based quality control schemes for more analytes will clearly improve agreement between laboratories and between different methods. Finally, education of clinicians as to the strengths and weaknesses of the methods for measuring hormones is a mission that will continue to improve clinical care.

REFERENCES

1. Sluss PM, Hayes FJ. Laboratory techniques for recognition of endocrine disorders. In: Melmed SA, RJ, Goldfine AB, et al., eds. *Williams Textbook Of Endocrinology*. 14th ed. Philadelphia, PA: Elsevier; 2020:62–90.e66.
2. de Herder WW. Heroes in endocrinology: Nobel Prizes. *Endocr Connect*. 2014;3:R94–R104.
3. Wheeler MJ. A short history of hormone measurement. *Methods Mol Biol*. 2013;1065:1–6.
4. Mills JM, Schwartz NB. Ovarian ascorbic acid as an endogenous and exogenous assay for cyclic proestrous LH release. *Endocrinology*. 1961;69:844–850.
5. Sayers G, Swallow RL, Giordano ND. An improved technique for the preparation of isolated rat adrenal cells: a sensitive, accurate and specific method for the assay of ACTH. *Endocrinology*. 1971;88:1063–1068.
6. Hodges JR. A modified adrenal ascorbic acid depletion technique for the bioassay of ACTH. *J Endocrinol*. 1955;12:152–158.
7. Kappagoda CT, Linden RJ, Pashley M. Increased sensitivity of ADH bioassay in rats by change in diet. *J Physiol*. 1980;299:425–435.
8. Forsling ML, Jones JJ, Lee J. Factors influencing the sensitivity of the rat to vasopressin. *J Physiol*. 1968;196:495–505.
9. Mattingly D. A simple fluorimetric method for the estimation of free 11-hydroxycorticoids in human plasma. *J Clin Pathol*. 1962;15:374–379.
10. Yalow RS. Remembrance project: origins of RIA. *Endocrinology*. 1991;129:1694–1695.
11. Yalow RS. Radioimmunoassay: past and potential role in clinical medicine. *Proc Rudolf Virchow Med Soc City N Y*. 1979;33:16–21.
12. Berson SA, Yalow RS. Radioimmunoassay of ACTH in plasma. *J Clin Invest*. 1968;47:2725–2751.
13. Berson SA, Yalow RS. General principles of radioimmunoassay. *Clin Chim Acta*. 1968;22:51–69.

14. Raff H, Findling JW. A new immunoradiometric assay for corticotropin evaluated in normal subjects and patients with Cushing's syndrome. *Clin Chem*. 1989;35:596–600.
15. Bartke A. The early days of steroid radioimmunoassays at the Worcester Foundation. *Endocrinology*. 1991;129:2277–2278.
16. Lieberman S. How steroid-specific antibodies came about: a personal history. *Steroids*. 1994;59:512–513.
17. Lequin RM. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clin Chem*. 2005;51:2415–2418.
18. Wu AH. A selected history and future of immunoassay development and applications in clinical chemistry. *Clin Chim Acta*. 2006;369:119–124.
19. Mann M. The ever expanding scope of electrospray mass spectrometry—a 30 year journey. *Nat Commun*. 2019;10:3744.
20. McLafferty FW. A century of progress in molecular mass spectrometry. *Annu Rev Anal Chem (Palo Alto Calif)*. 2011;4:1–22.
21. Zhou M, Veenstra T. Mass spectrometry: m/z 1983-2008. *Biotechniques*. 2008;44:667–668. 670.
22. Shackleton C. Clinical steroid mass spectrometry: a 45-year history culminating in HPLC-MS/MS becoming an essential tool for patient diagnosis. *J Steroid Biochem Mol Biol*. 2010;121:481–490.
23. Wudy SA, Choi MH. Steroid LC-MS has come of age. *J Steroid Biochem Mol Biol*. 2016;162:1–3.
24. Snyder LR. HPLC: past and present. *Anal Chem*. 2000;72:412A–420A.
25. Engelhardt H. One century of liquid chromatography. From Tswett's columns to modern high speed and high performance separations. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2004;800:3–6.
26. Shackleton C, Pozo OJ, Marcos J. GC/MS in recent years has defined the normal and clinically disordered steroidome: will it soon be surpassed by LC/tandem MS in this role? *J Endocr Soc*. 2018;2:974–996.
27. Raff H, Sluss PM. Pre-analytical issues for testosterone and estradiol assays. *Steroids*. 2008;73:1297–1304.
28. Sturgeon CM, Viljoen A. Analytical error and interference in immunoassay: minimizing risk. *Ann Clin Biochem*. 2011;48:418–432.
29. Rosner W, Vesper H, Endocrine Society, et al. Toward excellence in testosterone testing: a consensus statement. *J Clin Endocrinol Metab*. 2010;95:4542–4548.
30. Rosner W, Auchus RJ, Azziz R, et al. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2007;92:405–413.
31. Donegan DM, Algeciras-Schimnich A, Hamidi O, et al. Corticotropin hormone assay interference: A case series. *Clin Biochem*. 2019;63:143–147.
32. Greene LW, Geer EB, Page-Wilson G, et al. Assay-specific spurious ACTH results lead to misdiagnosis, unnecessary testing, and surgical misadventure—a case series. *JES*. 2019;3:763–772.
33. Shi J, Dhaliwal P, Zi Zheng Y, et al. An intact ACTH LC-MS/MS assay as an arbiter of clinically discordant immunoassay results. *Clin Chem*. 2019;65:1397–1404.
34. Forsman RW. Why is the laboratory an afterthought for managed care organizations? *Clin Chem*. 1996;42:813–816.
35. Lundberg GD. Acting on significant laboratory results. *JAMA*. 1981;245:1762–1763.
36. Ramamoorthy S, Cidlowski JA. Corticosteroids: mechanisms of action in health and disease. *Rheum Dis Clin North Am*. 2016;42:15–31. vii.
37. Kvetnansky R, Sabban EL, Palkovits M. Catecholaminergic systems in stress: structural and molecular genetic approaches. *Physiol Rev*. 2009;89:535–606.
38. Raff H, Sharma ST, Nieman LK. Physiological basis for the etiology, diagnosis, and treatment of adrenal disorders: Cushing's syndrome, adrenal insufficiency, and congenital adrenal hyperplasia. *Compr Physiol*. 2014;4:739–769.
39. Moghetti P, Bacchi E, Brangani C, et al. Metabolic effects of exercise. *Front Horm Res*. 2016;47:44–57.
40. Felding P, Tryding N, Hyltoft Petersen P, et al. Effects of posture on concentrations of blood constituents in healthy adults: practical application of blood specimen collection procedures recommended by the Scandinavian Committee on Reference Values. *Scand J Clin Lab Invest*. 1980;40:615–621.
41. Simundic AM, Cornes M, Grankvist K, et al. Standardization of collection requirements for fasting samples: for the Working Group on Preanalytical Phase (WG-PA) of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). *Clin Chim Acta*. 2014;432:33–37.
42. Bhasin S, Brito JP, Cunningham GR, et al. Testosterone therapy in men with hypogonadism: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2018;103:1715–1744.
43. Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2008;93:1526–1540.
44. Raff H, Carroll T. Cushing's syndrome: from physiological principles to diagnosis and clinical care. *J Physiol*. 2015;593:493–506.
45. Waikar SS, Sabbisetti VS, Bonventre JV. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney Int*. 2010;78:486–494.
46. Newman DJ, Pugia MJ, Lott JA, et al. Urinary protein and albumin excretion corrected by creatinine and specific gravity. *Clin Chim Acta*. 2000;294:139–155.
47. Nieman LK. Diagnostic tests for Cushing's syndrome. *Ann N Y Acad Sci*. 2002;970:112–118.
48. Raff H, Brock S, Findling JW. Cosyntropin-stimulated salivary cortisol in hospitalized patients with hypoproteinemia. *Endocrine*. 2008;34:68–74.
49. Raff H, Raff JL, Findling JW. Late-night salivary cortisol as a screening test for Cushing's syndrome. *J Clin Endocrinol Metab*. 1998;83:2681–2686.
50. Elder CJ, Vilela R, Johnson TN, et al. Pharmacodynamic studies of nasal tetracosactide with salivary glucocorticoids for a noninvasive Short Synacthen Test. *J Clin Endocrinol Metab*. 2020;105.
51. Carroll T, Raff H, Findling JW. Late-night salivary cortisol measurement in the diagnosis of Cushing's syndrome. *Nat Clin Pract Endocrinol Metab*. 2008;4:344–350.
52. Biondi B. The normal TSH reference range: what has changed in the last decade? *J Clin Endocrinol Metab*. 2013;98:3584–3587.
53. Spencer CA, Hollowell JG, Kazarosyan M, et al. National Health and Nutrition Examination Survey III thyroid-stimulating hormone (TSH)-thyroperoxidase antibody relationships demonstrate that TSH upper reference limits may be skewed by occult thyroid dysfunction. *J Clin Endocrinol Metab*. 2007;92:4236–4240.
54. Kratzsch J, Fiedler GM, Leichtle A, et al. New reference intervals for thyrotropin and thyroid hormones based on National Academy of Clinical Biochemistry criteria and regular ultrasonography of the thyroid. *Clin Chem*. 2005;51:1480–1486.
55. Colantonio DA, Kyriakopoulou L, Chan MK, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. *Clin Chem*. 2012;58:854–868.
56. Kohse KP, Thamm M. KiGGS-the German survey on children's health as data base for reference intervals. *Clin Biochem*. 2011;44:479.
57. Ridefelt P. Population-based pediatric reference intervals in general clinical chemistry: a Swedish survey. *J Med Biochem*. 2015;34:64–65.
58. Kant AK, Graubard BI. Race-ethnic, family income, and education differentials in nutritional and lipid biomarkers in US children and adolescents: NHANES 2003-2006. *Am J Clin Nutr*. 2012;96:601–612.
59. Kirk JK, D'Agostino Jr RB, Bell RA, et al. Disparities in HbA1c levels between African-American and non-Hispanic white adults with diabetes: a meta-analysis. *Diabetes Care*. 2006;29:2130–2136.
60. Eisenhofer G, Goldstein DS, Walther MM, et al. Biochemical diagnosis of pheochromocytoma: how to distinguish true- from false-positive test results. *J Clin Endocrinol Metab*. 2003;88:2656–2666.
61. Heaney RP, Holick MF. Why the IOM recommendations for vitamin D are deficient. *J Bone Miner Res*. 2011;26:455–457.
62. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2011;96:1911–1930.
63. Defining CLSI. Establishing, and verifying reference intervals in the clinical laboratory: approved guideline. In: *CLSI Document EP28-A3c*. 3rd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

64. Lin CL, Wu TJ, Machacek DA, et al. Urinary free cortisol and cortisone determined by high performance liquid chromatography in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 1997;82:151–155.
65. Raff H, Findling JW, Wong J. Short loop adrenocorticotropin (ACTH) feedback after ACTH-(1-24) injection in man is an artifact of the immunoradiometric assay. *J Clin Endocrinol Metab*. 1989;69:678–680.
66. Sedel F, Papeix C, Bellanger A, et al. High doses of biotin in chronic progressive multiple sclerosis: a pilot study. *Mult Scler Relat Disord*. 2015;4:159–169.
67. Tourbah A, Lebrun-Frenay C, Edan G, et al. MD1003 (high-dose biotin) for the treatment of progressive multiple sclerosis: a randomised, double-blind, placebo-controlled study. *Mult Scler*. 2016;22:1719–1731.
68. Koutsikos D, Agroyannis B, Tzanatos-Exarchou H. Biotin for diabetic peripheral neuropathy. *Biomed Pharmacother*. 1990;44:511–514.
69. Fernandez-Mejia C. Pharmacological effects of biotin. *J Nutr Biochem*. 2005;16:424–427.
70. Piketty ML, Polak M, Flechtner I, et al. False biochemical diagnosis of hyperthyroidism in streptavidin-biotin-based immunoassays: the problem of biotin intake and related interferences. *Clin Chem Lab Med*. 2017;55:780–788.
71. Hamrahian AH, Oseni TS, Arafah BM. Measurements of serum free cortisol in critically ill patients. *N Engl J Med*. 2004;350:1629–1638.
72. Klose M, Lange M, Rasmussen AK, et al. Factors influencing the adrenocorticotropin test: role of contemporary cortisol assays, body composition, and oral contraceptive agents. *J Clin Endocrinol Metab*. 2007;92:1326–1333.
73. Murphy BE. Some studies of the protein-binding of steroids and their application to the routine micro and ultramicro measurement of various steroids in body fluids by competitive protein-binding radioassay. *J Clin Endocrinol Metab*. 1967;27:973–990.
74. Diamandis EP, Christopoulos TK. The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin Chem*. 1991;37:625–636.
75. Nicoloff JT, Spencer CA. Clinical review 12: The use and misuse of the sensitive thyrotropin assays. *J Clin Endocrinol Metab*. 1990;71:553–558.
76. Wijeratne NG, Doery JC, Lu ZX. Positive and negative interference in immunoassays following biotin ingestion: a pharmacokinetic study. *Pathology*. 2012;44:674–675.
77. Elston MS, Sehgal S, Du Toit S, et al. Factitious Graves' disease due to biotin immunoassay interference-a case and review of the literature. *J Clin Endocrinol Metab*. 2016;101:3251–3255.
78. Siddiqui U, Egnor E, Sloane JA. Biotin supplementation in MS clinically valuable but can alter multiple blood test results. *Mult Scler*. 2017;23:619–620.
79. Li J, Wagar EA, Meng QH. Comprehensive assessment of biotin interference in immunoassays. *Clin Chim Acta*. 2018;487:293–298.
80. Srichomkwun P, Scherberg NH, Jaksic J, et al. Diagnostic dilemma in discordant thyroid function tests due to thyroid hormone autoantibodies. *AACE Clin Case Rep*. 2017;3:e22–e25.
81. Zouwail SA, O'Toole AM, Clark PM, et al. Influence of thyroid hormone autoantibodies on 7 thyroid hormone assays. *Clin Chem*. 2008;54:927–928.
82. Cartwright D, O'Shea P, Rajanayagam O, et al. Familial dysalbuminemic hyperthyroxinemia: a persistent diagnostic challenge. *Clin Chem*. 2009;55:1044–1046.
83. Maharjan AS, Wyness SP, Ray JA, et al. Detection and characterization of estradiol (E2) and unconjugated estriol (uE3) immunoassay interference due to anti-bovine alkaline phosphatase (ALP) antibodies. *Pract Lab Med*. 2019;17:e00131.
84. Rulander NJ, Cardamone D, Senior M, et al. Interference from anti-streptavidin antibody. *Arch Pathol Lab Med*. 2013;137:1141–1146.
85. Lam L, Bagg W, Smith G, et al. Apparent hyperthyroidism caused by biotin-like interference from IgM anti-streptavidin antibodies. *Thyroid*. 2018;28:1063–1067.
86. Lim YY, Ong L, Loh TP, et al. A diagnostic curiosity of isolated androstenedione elevation due to autoantibodies against horseradish peroxidase label of the immunoassay. *Clin Chim Acta*. 2018;476:103–106.
87. Crosby SR, Stewart MF, Ratcliffe JG, et al. Direct measurement of the precursors of adrenocorticotropin in human plasma by two-site immunoradiometric assay. *J Clin Endocrinol Metab*. 1988;67:1272–1277.
88. Lang BH, Fung MMH. Intraoperative parathyroid hormone (IOPTH) assay might be better than the second-generation assay in parathyroidectomy for primary hyperparathyroidism. *Surgery*. 2021;169:109–113.
89. Kricka LJ. Human anti-animal antibody interferences in immunological assays. *Clin Chem*. 1999;45:942–956.
90. Kaplan IV, Levinson SS. When is a heterophile antibody not a heterophile antibody? When it is an antibody against a specific immunogen. *Clin Chem*. 1999;45:616–618.
91. Graham KE, Samuels MH, Raff H, et al. Intraoperative adrenocorticotropin levels during transsphenoidal surgery for Cushing's disease do not predict cure. *J Clin Endocrinol Metab*. 1997;82:1776–1779.
92. Raff H, Shaker JL, Seifert PE, et al. Intraoperative measurement of adrenocorticotropin (ACTH) during removal of ACTH-secreting bronchial carcinoid tumors. *J Clin Endocrinol Metab*. 1995;80:1036–1039.
93. Stovold R, Blackhall F, Meredith S, et al. Biomarkers for small cell lung cancer: neuroendocrine, epithelial and circulating tumour cells. *Lung Cancer*. 2012;76:263–268.
94. Refetoff S, Fang VS, Marshall JS. Studies on human thyroxine-binding globulin (TBG). IX. Some physical, chemical, and biological properties of radioiodinated TBG and partially desialylated TBG. *J Clin Invest*. 1975;56:177–187.
95. Refetoff S, Hagen SR, Selenkow HA. Estimation of the T 4 binding capacity of serum TBG and TBPA by a single T 4 load ion exchange resin method. *J Nucl Med*. 1972;13:2–12.
96. Dichtel LE, Schorr M, Loures de Assis C, et al. Plasma free cortisol in states of normal and altered binding globulins: implications for adrenal insufficiency diagnosis. *J Clin Endocrinol Metab*. 2019;104:4827–4836.
97. Rosner W, Hankinson SE, Sluss PM, et al. Challenges to the measurement of estradiol: an Endocrine Society Position Statement. *J Clin Endocrinol Metab*. 2013;98:1376–1387.
98. Mazer NA. A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. *Steroids*. 2009;74:512–519.
99. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*. 1999;84:3666–3672.
100. Miller KK, Rosner W, Lee H, et al. Measurement of free testosterone in normal women and women with androgen deficiency: comparison of methods. *J Clin Endocrinol Metab*. 2004;89:525–533.
101. Ly LP, Sartorius G, Hull L, et al. Accuracy of calculated free testosterone formulae in men. *Clin Endocrinol (Oxf)*. 2010;73:382–388.
102. Van Uytfanghe K, Stockl D, Kaufman JM, et al. Validation of 5 routine assays for serum free testosterone with a candidate reference measurement procedure based on ultrafiltration and isotope dilution-gas chromatography-mass spectrometry. *Clin Biochem*. 2005;38:253–261.
103. Cumming DC, Wall SR. Non-sex hormone-binding globulin-bound testosterone as a marker for hyperandrogenism. *J Clin Endocrinol Metab*. 1985;61:873–876.
104. Barth JH, Field HP, Yasmin E, et al. Defining hyperandrogenism in polycystic ovary syndrome: measurement of testosterone and androstenedione by liquid chromatography-tandem mass spectrometry and analysis by receiver operator characteristic plots. *Eur J Endocrinol*. 2010;162:611–615.
105. Heureux N. Vitamin D testing-where are we and what is on the horizon? *Adv Clin Chem*. 2017;78:59–101.
106. Powe CE, Evans MK, Wenger J, et al. Vitamin D-binding protein and vitamin D status of black Americans and white Americans. *N Engl J Med*. 2013;369:1991–2000.
107. Boers J, Oldenburg-Ligtenberg PC, Stades AM, et al. Possible pitfalls in the workup of ectopic ACTH secretion illustrated by four rare cases. *BMJ Case Rep*. 2019;12.
108. Bouillon R. Free or total 25OHD as marker for vitamin D status? *J Bone Miner Res*. 2016;31:1124–1127.
109. Laudat MH, Cerdas S, Fournier C, et al. Salivary cortisol measurement: a practical approach to assess pituitary-adrenal function. *J Clin Endocrinol Metab*. 1988;66:343–348.

110. Raff H. Utility of salivary cortisol measurements in Cushing's syndrome and adrenal insufficiency. *J Clin Endocrinol Metab*. 2009;94:3647–3655.
111. Kannankeril J, Carroll T, Findling JW, et al. Prospective evaluation of late-night salivary cortisol and cortisone by EIA and LC-MS/MS in suspected Cushing syndrome. *J Endocr Soc*. 2020;4:bvaa107.
112. Raff H, Singh RJ. Measurement of late-night salivary cortisol and cortisone by LC-MS/MS to assess preanalytical sample contamination with topical hydrocortisone. *Clin Chem*. 2012;58:947–948.
113. Perogamvros I, Keevil BG, Ray DW, Trainer PJ. Salivary cortisone is a potential biomarker for serum free cortisol. *J Clin Endocrinol Metab*. 2010;95:4951–4958.
114. Otten BJ, Wellen JJ, Rijken JC, et al. Salivary and plasma androstenedione and 17-hydroxyprogesterone levels in congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 1983;57:1150–1154.
115. Walker RF, Hughes IA, Riad-Fahmy D. Salivary 17 alpha-hydroxyprogesterone in congenital adrenal hyperplasia. *Clin Endocrinol (Oxf)*. 1979;11:631–637.
116. Wood P. Salivary steroids. *Ann Clin Biochem*. 2009;46:427.
117. Fiers T, Kaufman JM. Management of hypogonadism: is there a role for salivary testosterone. *Endocrine*. 2015;50:1–3.
118. Cartledge S, Lawson N. Aldosterone and renin measurements. *Ann Clin Biochem*. 2000;37(Pt 3):262–278.
119. Sealey JE, Gerten-Banes J, Laragh JH. The renin system: variations in man measured by radioimmunoassay or bioassay. *Kidney Int*. 1972;1:240–253.
120. Carter S, Owen LJ, Kerstens MN, et al. A liquid chromatography tandem mass spectrometry assay for plasma renin activity using online solid-phase extraction. *Ann Clin Biochem*. 2012;49:570–579.
121. Menard J, Guyenne TT, Corvol P, et al. Direct immunometric assay of active renin in human plasma. *J Hypertens Suppl*. 1985;3:S275–278.
122. Findling JW, Raff H, Hansson JH, et al. Liddle's syndrome: prospective genetic screening and suppressed aldosterone secretion in an extended kindred. *J Clin Endocrinol Metab*. 1997;82:1071–1074.
123. Hartman D, Sagnella GA, Chesters CA, et al. Direct renin assay and plasma renin activity assay compared. *Clin Chem*. 2004;50:2159–2161.
124. Rossi GP, Barisa M, Belfiore A, et al. The aldosterone-renin ratio based on the plasma renin activity and the direct renin assay for diagnosing aldosterone-producing adenoma. *J Hypertens*. 2010;28:1892–1899.
125. Magill SB. Pathophysiology, diagnosis, and treatment of mineralocorticoid disorders. *Compr Physiol*. 2014;4:1083–1119.
126. Shackleton CH, Snodgrass GH. Steroid excretion by an infant with an unusual salt-losing syndrome: a gas chromatographic-mass spectrometric study. *Ann Clin Biochem*. 1974;11:91–99.
127. Arlt W, Biehl M, Taylor AE, et al. Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*. 2011;96:3775–3784.
128. Keevil BG. LC-MS/MS analysis of steroids in the clinical laboratory. *Clin Biochem*. 2016;49:989–997.
129. Halket JM, Waterman D, Przyborowska AM, et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot*. 2005;56:219–243.
130. Hines JM, Bancos I, Bancos C, et al. High-resolution, accurate-mass (HRAM) mass spectrometry urine steroid profiling in the diagnosis of adrenal disorders. *Clin Chem*. 2017;63:1824–1835.
131. French D. Advances in bioanalytical techniques to measure steroid hormones in serum. *Bioanalysis*. 2016;8:1203–1219.
132. Grebe SK, Singh RJ. LC-MS/MS in the clinical laboratory - where to from here? *Clin Biochem Rev*. 2011;32:5–31.
133. Binkley N, Dawson-Hughes B, Durazo-Arvizu R, et al. Vitamin D measurement standardization: the way out of the chaos. *J Steroid Biochem Mol Biol*. 2017;173:117–121.
134. Buttler RM, Martens F, Ackermans MT, et al. Comparison of eight routine unpublished LC-MS/MS methods for the simultaneous measurement of testosterone and androstenedione in serum. *Clin Chim Acta*. 2016;454:112–118.
135. Carter GD, Jones JC. Use of a common standard improves the performance of liquid chromatography-tandem mass spectrometry methods for serum 25-hydroxyvitamin-D. *Ann Clin Biochem*. 2009;46:79–81.
136. Greaves RF, Ho CS, Loh TP, et al. Working Group 3 "Harmonisation of Laboratory Assessment" European Cooperation in S, Technology Action BMD. Current state and recommendations for harmonization of serum/plasma 17-hydroxyprogesterone mass spectrometry methods. *Clin Chem Lab Med*. 2018;56:1685–1697.
137. Cao ZT, Botelho JC, Rej R, et al. Impact of testosterone assay standardization efforts assessed via accuracy-based proficiency testing. *Clin Biochem*. 2019;68:37–43.
138. Erdman P, Palmer-Toy DE, Horowitz G, et al. Accuracy-based vitamin D survey: six years of quality improvement guided by proficiency testing. *Arch Pathol Lab Med*. 2019;143:1531–1538.
139. Hawley JM, Owen LJ, Lockhart SJ, et al. Serum cortisol: an up-to-date assessment of routine assay performance. *Clin Chem*. 2016;62:1220–1229.
140. Owen LJ, Adaway JE, Davies S, et al. Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine service laboratory. *Ann Clin Biochem*. 2013;50:345–352.
141. Ueland GA, Methlie P, Kellmann R, et al. Simultaneous assay of cortisol and dexamethasone improved diagnostic accuracy of the dexamethasone suppression test. *Eur J Endocrinol*. 2017;176:705–713.
142. Middle JG, Kane JW. Oestradiol assays: fitness for purpose? *Ann Clin Biochem*. 2009;46:441–456.
143. Owen LJ, Monaghan PJ, Armstrong A, et al. Oestradiol measurement during fulvestrant treatment for breast cancer. *Br J Cancer*. 2019;120:404–406.
144. Bertelsen BE, Kellmann R, Viste K, et al. An ultrasensitive routine LC-MS/MS method for estradiol and estrone in the clinically relevant sub-picomolar range. *J Endocr Soc*. 2020;4:bvaa047.
145. Handelsman DJ, Gibson E, Davis S, et al. Ultrasensitive serum estradiol measurement by liquid chromatography-mass spectrometry in post-menopausal women and mice. *J Endocr Soc*. 2020;4.
146. Dirks NF, Ackermans MT, Lips P, et al. The when, what & how of measuring vitamin D metabolism in clinical medicine. *Nutrients*. 2018;10:482.
147. Ooms N, van Daal H, Beijers AM, et al. Time-course analysis of 3-epi-25-hydroxyvitamin D3 shows markedly elevated levels in early life, particularly from vitamin D supplementation in preterm infants. *Pediatr Res*. 2016;79:647–653.
148. Rauh M. LC-MS/MS for protein and peptide quantification in clinical chemistry. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2012;883–884:59–67.
149. Kushnir MM, Rockwood AL, Roberts WL, et al. Measurement of thyroglobulin by liquid chromatography-tandem mass spectrometry in serum and plasma in the presence of antithyroglobulin autoantibodies. *Clin Chem*. 2013;59:982–990.
150. Van Der Gugten JG, Holmes DT. Quantitation of plasma renin activity in plasma using liquid chromatography-tandem mass spectrometry (LC-MS/MS). *Methods Mol Biol*. 2016;1378:243–253.
151. Schwarz E, Liu A, Randall H, et al. Use of steroid profiling by UPLC-MS/MS as a second tier test in newborn screening for congenital adrenal hyperplasia: the Utah experience. *Pediatr Res*. 2009;66:230–235.
152. Peitzsch M, Dekkers T, Haase M, et al. An LC-MS/MS method for steroid profiling during adrenal venous sampling for investigation of primary aldosteronism. *J Steroid Biochem Mol Biol*. 2015;145:75–84.
153. O'Reilly MW, Kempegowda P, Jenkinson C, et al. 11-oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2017;102:840–848.
154. Fiet J, Le Bouc Y, Guechot J, et al. A liquid chromatography/tandem mass spectometry profile of 16 serum steroids, including 21-deoxycortisol and 21-deoxycorticosterone, for management of congenital adrenal hyperplasia. *J Endocr Soc*. 2017;1:186–201.
155. Keevil B. Steroid mass spectrometry for the diagnosis of PCOS. *Med Sci (Basel)*. 2019;7:78.
156. Turcu AF, El-Maouche D, Zhao L, et al. Androgen excess and diagnostic steroid biomarkers for nonclassic 21-hydroxylase deficiency without cosyntropin stimulation. *Eur J Endocrinol*. 2020;183:63–71.
157. Oriolo C, Fanelli F, Castelli S, et al. Steroid biomarkers for identifying non-classic adrenal hyperplasia due to 21-hydroxylase deficiency in a population of PCOS with suspicious levels of 17OH-progesterone. *J Endocrinol Invest*. 2020;43:1499–1509.

158. Wilkes EH, Rumsby G, Woodward GM. Using machine learning to aid the interpretation of urine steroid profiles. *Clin Chem*. 2018;64: 1586–1595.
159. Masjkur J, Gruber M, Peitzsch M, et al. Plasma steroid profiles in subclinical compared with overt adrenal Cushing syndrome. *J Clin Endocrinol Metab*. 2019;104:4331–4340.
160. Erlic Z, Kurlbaum M, Deutschbein T, et al. Metabolic impact of pheochromocytoma/paraganglioma: targeted metabolomics in patients before and after tumor removal. *Eur J Endocrinol*. 2019;181:647–657.
161. Eisenhofer G, Lenders JW, Timmers H, et al. Measurements of plasma methoxytyramine, normetanephrine, and metanephrine as discriminators of different hereditary forms of pheochromocytoma. *Clin Chem*. 2011;57:411–420.
162. Miller AG, Brown H, Degg T, et al. Measurement of plasma 5-hydroxyindole acetic acid by liquid chromatography tandem mass spectrometry—comparison with HPLC methodology. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2010;878:695–699.
163. Darr R, Kuhn M, Bode C, et al. Accuracy of recommended sampling and assay methods for the determination of plasma-free and urinary fractionated metanephrines in the diagnosis of pheochromocytoma and paraganglioma: a systematic review. *Endocrine*. 2017;56:495–503.
164. Casals G, Hanzu FA. Cortisol measurements in Cushing's syndrome: immunoassay or mass spectrometry? *Ann Lab Med*. 2020;40:285–296.
165. Bancos I, Taylor AE, Chortis V, et al. Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. *Lancet Diabetes Endocrinol*. 2020;8:773–781.

3

Endocrine Rhythms, the Sleep-Wake Cycle, and Biological Clocks

Stafford L. Lightman and Thomas J. Upton

OUTLINE

Introduction, 36
  What are Biological Rhythms?, 36
  The Sleep-Wake Cycle as a Rhythm, 36
  Clocks in the Brain and the Periphery, 36
Endocrine Rhythms, 37
  Examples of Endocrine Rhythms in Humans, 37
  The Hypothalamo-Pituitary-Adrenal Axis as a Model Endocrine Rhythm, 37
  Glucocorticoid Pulsatility, 39
  Rhythms in Endocrine Disease, 39
  Limitations of Single Timepoint Assessments, 39
Cell Clocks, 39
The Sleep-Wake Cycle, 40
  Sleep Regulation, 40
  Sleep Stages, 40
Melatonin as the Biological Maker of the Night, 40
Consequences of Rhythm Disturbance on Metabolic Health, 40
  Sleep Mistiming and Human Health, 40
  Jet Lag and Social Jet Lag, 41
Novel technology for the Study of Biological Rhythms, 41
Summary, 41
Conclusions, 41

INTRODUCTION

In 2017, the Nobel Prize in Physiology or Medicine was jointly awarded for the discovery of the molecular mechanisms that control the circadian clock. In announcing the prize the Nobel Committee described how the work of Jeffrey C. Hall, Michael Rosbash, and Michael W. Young explains "how plants, animals and humans adapt their biological rhythm so that it is synchronized with the Earth's revolutions."[1] Adaptation of life to the length of each day had proven to be literally "in our genes."

What Are Biological Rhythms?

Biological rhythms are natural oscillations that occur in physiology and behavior that respond to periodic changes in the environment. They can be described by their length and by the circumstances in which they occur. Humans and many other animals have a *diurnal* behavior rhythm reflecting more activity during daylight hours, and rest or sleep at night. By contrast, *nocturnal* species such as rodents are active at night and rest during the day. Rhythms with longer periods such as annual migration or the human menstrual cycle are *infradian,* whereas rhythms that occur within a period of minutes to hours are *ultradian.*

Rhythms that are endogenous, self-sustained, and approximate a 24-hour day, are termed *circadian*. In mammals this includes secretion of many hormones, most notably cortisol and melatonin. Circadian rhythms are entrained or synchronized to external cues (*Zeitgebers*, from the German "time giver"), and in humans and other mammals the most potent of these are light and timing of food.[2] The most important point is that circadian rhythms are *intrinsic*—they continue to oscillate even in the absence of a *zeitgeber* or synchronizing input.

KEY POINTS

- Measurement of melatonin in dim light conditions is considered the gold standard measure of circadian phase. Melatonin secretion is extremely sensitive to light, which may alter the phase of the rhythm or temporarily suppress secretion.

The Sleep-Wake Cycle as a Rhythm

The sleep-wake cycle is perhaps the most obvious and visible behavioral manifestation of a circadian rhythm in humans, having evolved as a consequence of the earth's 24-hour rotation. Characterized by a temporary suspension of consciousness, sleep is nevertheless associated with a multitude important and dynamic processes.[3] Bordered at the onset by the rise of the "sleep hormone" melatonin and at the offset by an anticipatory rise in the hormone cortisol (Fig. 3.1), sleep occurs as a series of *ultradian* stages, with each cycle lasting approximately 90 to 120 minutes.[3] Accompanying these cycles is a complex array of physiological events, including changes in sympathetic output, fluctuations in body temperature and muscle tone, and the secretion of hormones like growth hormone and aldosterone.[4,5] Disturbance of the sleep rhythm, either in time or structure, results in adverse health outcomes for the individual.[6]

Clocks in the Brain and the Periphery

We have already noted that circadian rhythms are *intrinsic,* and therefore their period persists in the absence of external cues. The mechanisms by which internal time is maintained depend on a network of intracellular clocks, coordinated by a central pacemaker in the suprachiasmatic nucleus of the hypothalamus (SCN). The SCN consists of tiny bilateral nuclei of a few thousand cells each—the cells are tightly

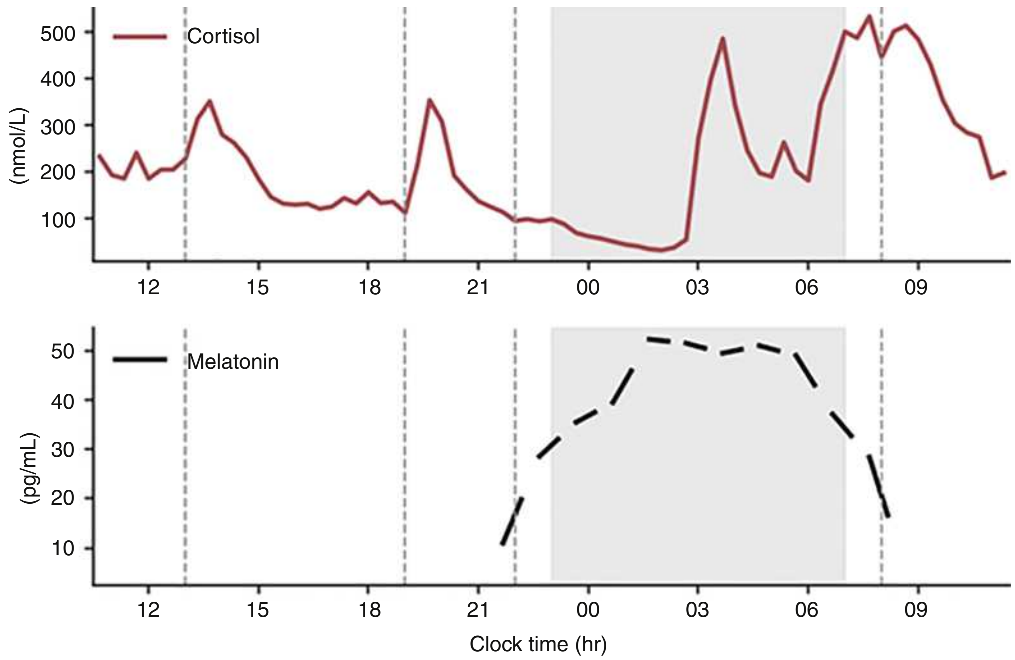


Fig. 3.1 Illustrative example of the daily rhythms of two circadian hormones—cortisol and melatonin. In this experiment, a healthy male volunteer was admitted to the research facility, and plasma samples were taken every 20 minutes during a 25-hour routine in which ambient light, meals *(vertical dashed lines)*, and sleep times *(shaded areas)* were controlled. Both the diurnal and ultradian rhythms of cortisol *(top panel, red)* can be observed. Following the nadir, pulsatile secretion begins around 3:00 AM, with peak levels just after waking. In dim light conditions (<10 lux ambient light) melatonin *(bottom panel, dashed lines)* begins to rise in the evening, peaks during sleep, and falls to undetectable levels soon after waking. (Source: Author's own data [unpublished].)

coupled, and output oscillates with a free-running period of close to 24 hours. Light information conveying time of day reaches the SCN via the direct synaptic connections from projections of specialized photosensitive retinal ganglion cells along the retinohypothalamic tract. The traditional view of the circadian timing system in mammals was to assume that the SCN was the only tissue capable of generating self-sustaining oscillations. In this model, the vast majority of peripheral tissues acted as subservient "slaves" to the master clock.[7] It has only been in the very recent past that scientists have discovered that, in fact, molecular clocks are present in almost all cells and tissues of the body, and that the SCN functions to coordinate and synchronize these rhythms.[8]

ENDOCRINE RHYTHMS

An endocrinologist's job is to concern themselves with the diagnosis and management of conditions affecting endocrine glands and hormones. Even though almost all hormones show rhythmic time of day variation, even in disease states, clinicians are typically required to interpret and extrapolate information about disorders based on single or limited time series samples. These timepoints for sampling almost always converge in the morning, which is generally the most convenient for clinicians and laboratory staff, but do not necessarily correspond with the best time for diagnostic value—the peaks and troughs in the hormone or hormones being considered. Interindividual variation in circadian and behavioral timing results in very broad "normal" ranges that can be very difficult to interpret.

Examples of Endocrine Rhythms in Humans

The examples of human endocrine rhythms are numerous. In this review we will focus on the oscillations of glucocorticoids that have both *circadian* and *ultradian* rhythmicity, but it is worth mentioning a few others by way of example. Melatonin is a *circadian* hormone, secreted by the pineal gland and in humans heralds the onset of sleep. Secretion begins in the evening, peaks during sleep, and falls to low or undetectable levels during daylight waking hours. Measurement of melatonin, when the onset is measured under dim light conditions,[9] is considered the benchmark method of circadian phase determination in humans. Melatonin has effects that go substantially beyond conveying timing information to body tissues. There is an increasing evidence to suggest that melatonin plays an important role in the modulation of key metabolic processes, including glucose regulation and energy balance.[10,11] Glucose and insulin themselves have *diurnal* and *ultradian* periods, and evidence from studies in humans and rodents suggests that "eating against the clock"—that is, at metabolically unfavorable clock times—contributes to the onset of metabolic dysfunction and diabetes.[12]

Ultradian rhythms are very frequently superimposed upon rhythms of longer periods. In endocrinology, pulsatile ultradian secretion is a characteristic feature of most hormones, including growth hormone,[13] insulin,[14] and the reproductive axis,[15,16] as well as the pulsatile secretion of glucocorticoids.[17–20]

The Hypothalamo-Pituitary-Adrenal Axis as a Model Endocrine Rhythm

The synthesis and secretion of glucocorticoid hormones occurs under the control of positive and negative feedback loops that exist between the hypothalamus, pituitary, and adrenal glands—the so-called hypothalamo-pituitary-adrenal (HPA) axis. Stimulated by the release of adrenocorticotropic hormone (ACTH), glucocorticoids exert multiple effects, including regulation of growth, metabolism, and the

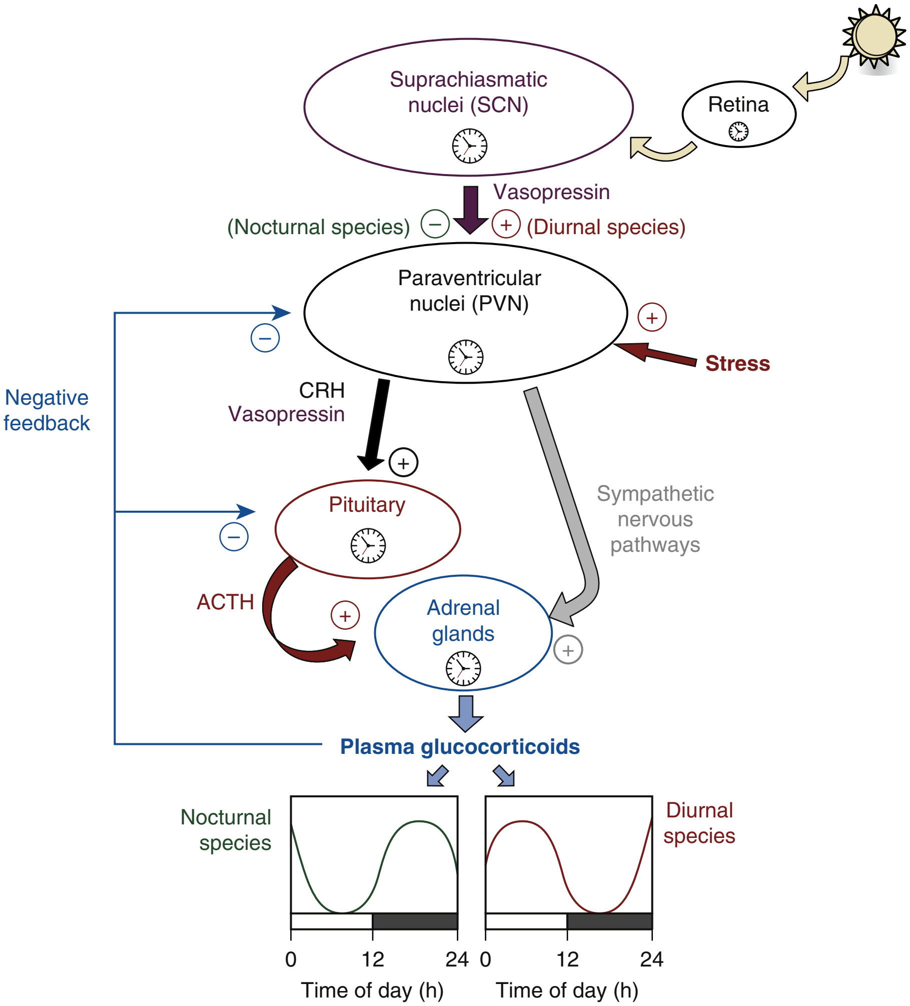


Fig. 3.2 Representation of the control of circadian glucocorticoid release in mammals. The circadian secretion of glucocorticoids is dependent on the rhythmic release of adrenocorticotropic hormone (*ACTH*) and a gating process by the adrenal clock. Clock symbols represent self-sustained oscillators. Rhythmic activity of the hypothalamo-pituitary-adrenal axis regulated by the central clock in the suprachiasmatic nuclei (*SCN*), which is entrained by ambient light that is received via the retina. *PVN,* Paraventricular nuclei; *CRH,* corticotrophin-releasing hormone. (Modified from Oster H, Challet E, Ott V, et al. The functional and clinical significance of the 24-hour rhythm of circulating glucocorticoids. *Endocr Rev*. 2016;38:3–45.)

stress response—their action is mediated through the near-ubiquitous presence of glucocorticoid[21] and less widely distributed mineralocorticoid[22] receptors in central and peripheral tissues. In humans, the circadian rhythm of cortisol peaks just before waking, anticipating the start of the new day.[23]

KEY POINTS

- The synthesis and secretion of glucocorticoid hormones is under the control of the hypothalamo-pituitary-adrenal axis, which exhibits both a diurnal circadian and an ultradian rhythm. The expression of genes and gene transcripts is influenced by the pattern of glucocorticoid exposure, and several human diseases are associated with abnormal glucocorticoid pulsatility.
- Sleep is characterized by cycles of rapid eye movement and nonrapid eye movement stages, whose gold standard measurement is polysomnography. Sleep is regulated both by a homeostatic process, driven by sleep debt, and a circadian process, regulated by the central pacemaker in the suprachiasmatic nuclei.

The ACTH secretory rhythm is coordinated by inputs from the SCN, which project to the paraventricular nucleus of the hypothalamus (PVN). The SCN is thus able to regulate corticotrophin-releasing hormone (CRH) from neurons in the PVN, and therefore release of ACTH from the anterior pituitary.[24] Further, in the presence of metyrapone, which temporarily suppresses endogenous cortisol production, and under conditions of constant CRH infusion, there is still a circadian rhythm of ACTH secretion, suggesting a role for other hypothalamic mediators such as vasopressin.[25]

The responsiveness of the adrenal gland to ACTH stimulation depends on the phase of the circadian cycle,[26] and, in addition, peripheral "clocks" located in the adrenal cortex can be entrained by light, likely via neuronal connections of the autonomic nervous system.[26,27] Thus, the circadian rhythm of glucocorticoids, entrained by the SCN, is modulated at the level of the adrenal both through inputs from the sympathetic nervous system, and by intrinsic adrenal mechanisms that "gate" sensitivity to ACTH stimulus depending on the time of day (Fig. 3.2).

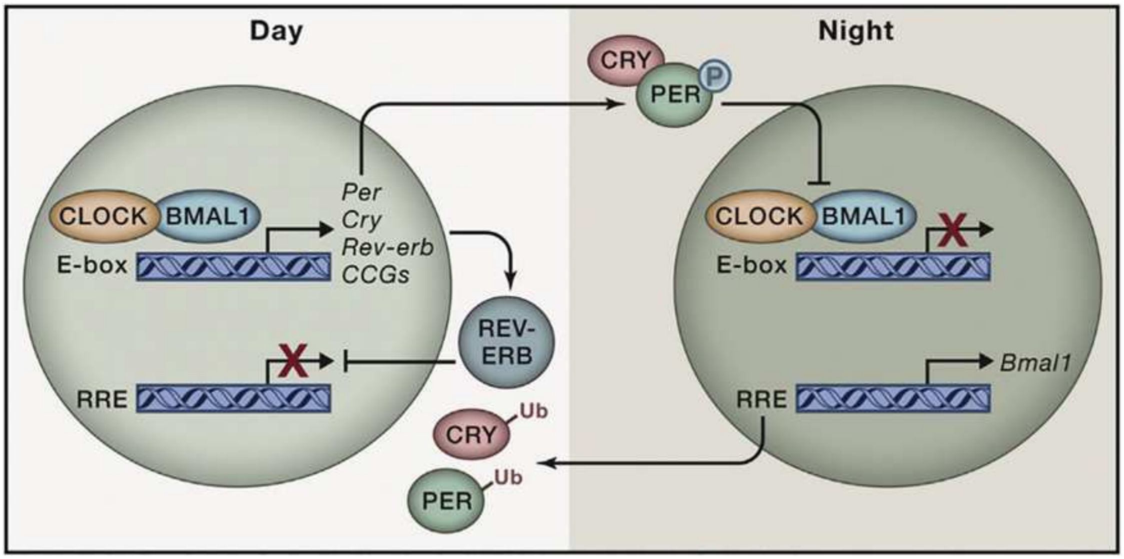


Fig. 3.3 The machinery of the intracellular circadian clock, present in almost all mammalian cells. *CLOCK* and *BMAL1* are transcriptional activators that dimerize to activate core clock genes (*CCGs*) including the expression of *Per, Cry* and *Rev-erb*. Levels of PER and CRY proteins increase to maximal levels during the night, where they themselves form dimer complexes, translocate to the cell nucleus, and suppress *CLOCK:BMAL1* transcription. Post-translational modification of *PER* and *CRY* proteins facilitate their degradation, a necessary step to start a new circadian cycle. In a second loop, high concentrations of *REV-ERB* proteins during the day inhibit *Bmal1* transcription, while at night, low levels of *REV-ERB* allow transcription to take place. (From Asher G, Sassone-Corsi P. Time for Food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell.* 2015;161:84–92.)

In contrast to the daily circadian rhythm, the *ultradian* rhythm of glucocorticoids is characterized by the pulsatile secretion of hormone every 60 to 90 minutes, as determined in studies of both rodent animals[18] and humans.[20,28] There is no evidence of any rhythms of this frequency in the hypothalamus, and hypotheses from mathematical modeling predict that ultradian oscillations in ACTH and cortisol secretion occur at a subhypothalamic level as a consequence of dynamic interactions between positive feedforward from pituitary ACTH on the adrenal and negative feedback generated by adrenal cortical glucocorticoids[29] on pituitary corticotrophs. These predictions have been validated in rodent models where ultradian oscillations of ACTH and corticosterone (the dominant glucocorticoid in rodents) persist even under conditions of constant CRH infusion,[30] thus demonstrating that pulsatile CRH secretion is not required for the ultradian rhythm of glucocorticoids.

Glucocorticoid Pulsatility

The importance of ultradian glucocorticoid secretion has for the most part been investigated in animal experiments. These have shown that expression of genes and gene transcripts in the rodent liver[31,32] and brain[33,34] are differentially influenced by the altering patterns of exposure to the principal glucocorticoid in rodents, corticosterone. In contrast to rats exposed to constant levels of glucocorticoid, ultradian pulses induce cyclical glucocorticoid receptor–mediated gene transcription, or "gene pulsing."[31] In humans, evidence for the biological consequence of patterns of glucocorticoid exposure is hinted at by differences in HPA axis pulsatility in adults with depression[35] or sleep apnea[36] and following major cardiac surgery.[37,38] Specific evidence of the potential importance of glucocorticoid pulsatility on normal cognitive function has emerged from recent experiments in healthy adults. In a protocol in which endogenous cortisol production was suppressed, only participants who received pulsatile replacement were able to perform well in cognitively demanding tasks, compared with those participants who received standard oral or continuous infusion replacement regimens.[39]

Rhythms in Endocrine Disease

Pathological rhythmicity driven by the secretion of hormones from tumors is the hallmark of many endocrine diseases. For example, high-frequency blood sampling in patients with pituitary adenomas reveals pulsatility of growth hormone, prolactin and cortisol secretion rather than simply tonic hypersecretion.[40] Disorganized and irregular but pulsatile secretion of glucocorticoid has been identified in patients with both Cushing disease[41] and adrenal Cushing syndrome.[42] A diurnal rhythm with superimposed ultradian pulsatility also characterizes the rhythm of aldosterone secretion in health and in primary aldosteronism, differentiated by higher basal release and secretion irregularity in the disease condition.[43]

Limitations of Single Timepoint Assessments

Given that hormonal processes are dynamic, occasional assessment of single hormones at restricted timepoints provides incomplete, and potentially misleading, information. Single-timepoint measurements also fail to capture the significant and important metabolic and hormonal changes that occur during sleep. The major fluxes of both melatonin and cortisol occur during this time, and the sleep period is typically the only part of the day in which the fasting state can be observed without interference with or from any daily activities.

CELL CLOCKS

The intrinsic cell clock machinery present both in neurons of the SCN and in peripheral tissues consists of transcriptional-translational feedback loops that oscillate with a period of approximately 24 hours (Fig 3.3). The principal loop is made up of two transcription activators: brain and muscle ARNTL-like protein 1 (BMAL1) and circadian locomotor output cycles kaput (CLOCK). The BMAL1/CLOCK dimer complex activates numerous so-called "clock-controlled genes," as well as inducing transcription of *period* (*PER*) and *cryptochrome* (*CRY*). Increasing levels of PER and CRY proteins act via negative feedback to inhibit CLOCK and BMAL transcription, PER and CRY are degraded, and the cycle begins once more.[44] In addition to these core components, feedback is also regulated by a number of posttranslational mechanisms, including phosphorylation and glycosylation of clock proteins.[45] Studies in rodents have demonstrated that malfunction of clock genes leads to alterations of circadian period,[46] while deactivating mutations of *Per2/Cry1* result in the complete dysregulation of the circadian rhythms of ACTH, corticosterone, and clock gene expression in mice.[26] In humans, a hereditary delayed sleep phase disorder has been associated with gain-of-function mutation in the *CRY1* gene.[47]

KEY POINTS

- Cells in the brain and in peripheral tissues contain clock machinery consisting of transcriptional-translational feedback loops that oscillate with a period of approximately 24 hours.

THE SLEEP-WAKE CYCLE

Sleep can be defined as a reversible, temporary state of altered consciousness, characterized by reduced responsiveness, altered muscle activity, and changes in metabolism.[48] Sleep is a highly complex process involving multiple interacting systems ranging from gene expression to whole-body physiological regulation, and, despite its regular occurrence and the considerable time that it occupies in our lifetime, there remains much that is not understood (for reviews see Dijk & Landholt[3] and Carskadon & Dement[49]).

Sleep Regulation

The regulation of sleep in humans has been described by a two-part model consisting of a *homeostatic* process, represented by sleep debt, and a *circadian* process that is controlled by the SCN circadian pacemaker.[50] In the model, homeostatic sleep pressure increases with increasing time awake and abates again during sleep. The sleep period oscillates with a circadian period that is normally entrained to the light cycle of day and night.[50] In animal experiments, lesions to the SCN do not disrupt sleep homeostasis,[51] indicating that the two processes are distinct. Uncoupling the sleep-wake cycle from the circadian rhythm in humans can be achieved using *forced desynchrony* protocols in which day length is gradually extended to 28 hours, a period beyond which the circadian pacemaker can entrain. This allows researchers to separate the individual influences of the circadian and homeostatic processes on sleep,[52] as well as the ability to assess the "free-running" periods of circadian rhythms independent of the sleep-wake cycle.[53,54]

Sleep Stages

Laboratory studies of human sleep typically measure multiple physiological variables, which are standardized into a protocol known as *polysomnography*.[55] Measurement of the electroencephalogram (EEG) simultaneously with eye movement and muscle tone represents the core features by which stages of sleep can be defined. Broadly, sleep can be divided into cyclical periods of rapid eye movement (REM) sleep that occur every 90 to 120 minutes on the background of periods of non-REM (NREM) sleep, each defined by distinct patterns of EEG and muscle activity.[56] The transition between wakefulness and the onset of sleep is classified as stage N1 sleep. A sleep period that contains large amounts of N1 indicates frequent arousals including, for example, those caused by sleep apnea.[56] Stage N2 constitutes the majority of NREM sleep, whereas deep sleep (N3) is more predominant in the earlier part of the night. REM sleep accounts for approximately 25% of sleep time,[56] becomes dominant in last third of the night,[49] and is closely correlated with the circadian rhythm of core body temperature.[57]

MELATONIN AS THE BIOLOGICAL MAKER OF THE NIGHT

In controlled, dimly lit conditions, melatonin begins to rise in the evening, before habitual bed time, and remains elevated during the night, before falling to very low levels during the day.[58] Profiles of melatonin vary between individuals and across gender, age, and chronotype[9,59-62] but are generally stable within an individual. This means that measurement of melatonin is considered a robust and reliable marker of circadian phase.[9] Light is the most potent synchronizer of the circadian clock, and light exposure has a profound effect on the secretion of melatonin. Bright light exposure in the morning leads to a phase advance (earlier melatonin onset), whereas light exposure in the evening delays or suppresses melatonin secretion.[63-66]

Measurement of circulating plasma melatonin remains the gold standard for assessment of the endogenous circadian phase in humans.[9] Full profiling of melatonin typically requires admission to a specialist research unit and drawing of multiple blood tests. The onset of melatonin in ambulatory settings is usually estimated by obtaining a partial profile in saliva during the hours leading up to habitual bed time,[9] which is less invasive than blood sampling, but requires active participation from the participant (physical collection of samples, altering light and body position if collecting in dim light conditions).[67]

Although there is unequivocal evidence that exposure to artificial light influences melatonin rhythms,[68-71] there is as yet no strong mechanistic link between this disruption and any specific pathology. In a case-control study of nurses, those with the lowest excretion of urinary melatonin metabolites were significantly more likely to develop type 2 diabetes than those with the highest levels.[72] In a study of 21 healthy female volunteers, 5 mg of exogenous melatonin resulted in impaired glucose tolerance at both morning and evening timepoints, when glucose was measured acutely after dose administration.[73] Further, melatonin receptor mutations have been linked with an increased risk of impaired glucose tolerance and type 2 diabetes in gene-wide association and prospective studies,[74-76] but in an analysis of data from more than 189,000 participants in the UK Biobank, the presence of the mutation was not associated with increased risk of diabetes in shift workers.[77] Untangling the complex relationships between evening bright light–induced melatonin suppression with associated behaviors such as late meals, natural interindividual variation in chronotype and light sensitivity, and any effect on sleep duration or quality remains a significant challenge.

CONSEQUENCES OF RHYTHM DISTURBANCE ON METABOLIC HEALTH

Robust evidence of the importance of coordinated rhythms in the maintenance of health comes from the study of clock-absent animal models—disabling mutations in core clock genes results in loss of not only behavioral rhythmicity but also the development of abnormal metabolic phenotypes that include obesity, impaired glucose metabolism, lipid disorders, and reduced lifespan.[78,79] In the wild-type condition, factors that normally function to synchronize and entrain daily rhythms, such as hormones, light, and food timing, can become instead "desynchronizers"—that is, exert harmful phase shifting effects—when presented at the wrong time of day, or without coordination.

KEY POINTS

- Factors that act to desynchronize rhythms include mistiming of sleep and meals, exposure to bright light during normal periods of sleep and rest, and the pathological changes in rhythmicity that occur in endocrine diseases.

Sleep Mistiming and Human Health

There is increasing evidence that misalignment of sleep is associated with metabolic dysfunction. Epidemiological studies have linked late chronotype (evening preference) with increased risk of multiple comorbidities and all-cause mortality.[80] Short sleep duration has been associated with an increased risk of diabetes and obesity,[80,81] and sleep fragmentation with impaired glucose tolerance.[82] Following 7 days of partial sleep deprivation, transcriptome analysis of plasma obtained from 26 adult volunteers found that over 700 genes were either up- or downregulated, including a significant reduction in the number of genes with circadian expression. These included *PER*, *CRY*, and *CLOCK*, as well as upregulation of genes associated with immune and stress responses.[83]

Shift workers are a particularly vulnerable group. Successive night shifts are associated with an accumulation of sleep debt,[87] and during recovery sleep quality may still be affected for many days after completing night shift work.[87]

A metaanalysis of 26 studies involving over 300,000 participants found that shift work was associated with a significant increased risk of overweight and obesity,[84] and a review of 38 metaanalyses and 24 systematic reviews additionally found links with shift work and increased risk of type 2 diabetes, coronary artery disease, stroke, and cancer.[85] A systematic review of 33 observational studies found that shift workers overall consumed a less healthy diet, were less regular with meal timing, and were more likely to consume food at night.[86]

Jet Lag and Social Jet Lag

Rapid shifts in time zone associated with air travel frequently results in a collection of symptoms commonly known as “jet lag,” a syndrome that has been formally acknowledged as a sleep disorder.[88] Jet lag occurs as a direct consequence of acute misalignment of internal timing rhythms with day/night cycles. Symptoms include difficulty initiating and/or maintaining sleep, excessive sleepiness, altered alertness and performance, and somatic symptoms of altered gastrointestinal function.[88]

Direction of travel influences the effect on the circadian system, with eastward flight, which requires a phase advance of the circadian timing system, being generally more difficult to adapt to than westward travel, which requires a phase delay.[89] Phase adjustment may be hastened by optimizing the timing of bright daylight exposure in the new time zone,[90] and a systematic review of randomized trials concluded that there is good evidence to recommend the use of melatonin supplementation to reduce or prevent the symptoms of jet lag, where travel occurs over at least five time zones, especially in an eastward direction.[89] Where the duration of a trip is short (<3 days) it may be best to maintain a “home-based” sleep pattern rather than attempt to adapt to the new time zone.[91]

“Social jet lag” describes the discrepancy in sleep time that may occur between work and free days.[92] Societal pressures around work and education schedules may conflict with the naturally selected sleep patterns of many individuals, leading to an accumulation of sleep debt. The amount of social jet lag can be quantified by calculating the difference between the midpoint of sleep on work days compared with nonwork days.[92] Thus, people with an evening chronotype (who prefer later bed and wake up times) may be disproportionately affected by the requirement to adhere to a social timetable that requires rising early in the morning for work or school. Large-scale surveys in European populations suggest that two-thirds of working-age people suffer from at least 1 hour of social jet lag, and one-third from at least 2 hours.[93] There is an increasing body of evidence that, like shift work, social jet lag is associated with potentially unhealthy lifestyle choices[94] and other health problems, including smoking,[92] obesity,[95] and metabolic disturbance.[96,97]

NOVEL TECHNOLOGY FOR THE STUDY OF BIOLOGICAL RHYTHMS

Accurate and confident diagnosis of endocrine disease is important and remains a challenge, especially for diseases such as Cushing syndrome and primary aldosteronism. To make progress in understanding the significance of rhythmic changes in human endocrinology and metabolism, more data need to be collected at high resolution, including during undisturbed sleep. In-laboratory protocols involving the sampling of blood are the gold standard, but are typically laborious, invasive, and expensive. Further, the circumstances in which data are captured in laboratory settings are unlikely to reflect the life conditions that patients inhabit when they present for assessment at the clinic.

Novel technology in the form of continuous ambulatory microdialysis presents an opportunity to study the daily rhythms of hormones and metabolites at high frequency and within a framework of normal life. Clinical microdialysis involves the placement of a semipermeable membrane (probe) into the tissue of interest, and then perfusion of the probe at a slow flow rate, producing a microdialysate containing dialyzed substances of interest. Coupling microdialysis with a portable collection device developed by the Lightman group,[98] frequent (20-minute) automated sampling in subcutaneous tissue has become possible over a full 24-hour period, during which the participant is able to continue normal activities.[99,100] Work using the prototype device has so far demonstrated that circadian profiles of free cortisol are not only detectable in subcutaneous fluid,[98] but remain robust over time and correlate with total plasma concentrations.[99] With the emergence of mass spectrometry as an analytical method,[101] simultaneous measurement of multiple analytes has now become achievable. For example, using this approach it has been possible for the first time to demonstrate the presence and rhythmicity of tissue free melatonin, measured simultaneously with cortisol and its metabolite cortisone (Fig. 3.4). A major multicenter clinical trial of the technology to measure daily rhythms of multiple tissue adrenal steroids in health and in endocrine conditions (ULTRADIAN, https://clinicaltrials.gov/ct2/show/NCT02934399), with the aim of providing clinical benefit through the provision of informative dynamic data without the need for blood sampling or hospital admission, has recently completed recruitment (Fig. 3.5).

KEY POINTS

- Capturing the dynamics of daily rhythms may provide an opportunity to improve the understanding of disease processes, as well as facilitating faster and more accurate diagnosis of endocrine conditions.

SUMMARY

As humans have evolved to live in a constantly changing environment, rhythmicity has become an intrinsic feature of our physiology. Rhythms of different periods and amplitude coordinate and synchronize with one another to create systems of dynamic equilibrium that can adapt to the unique circumstances of the individual. Key rhythms in humans include the diurnal circadian rhythm of glucocorticoids, which also exhibit ultradian pulsatility, and the circadian rhythm of melatonin, a marker of the biological night.

Aside from these two hormones, almost all hormone and metabolic processes are rhythmic, coordinated by cellular clocks and feedback/feedforward loops that optimize physiological functions for the appropriate time of day. Endocrine and metabolic disease occurs when the timing or oscillation of these rhythms becomes pathological or desynchronized.

CONCLUSIONS

Although the importance of rhythms is become increasingly recognized, there remains a major knowledge gap around the mechanisms

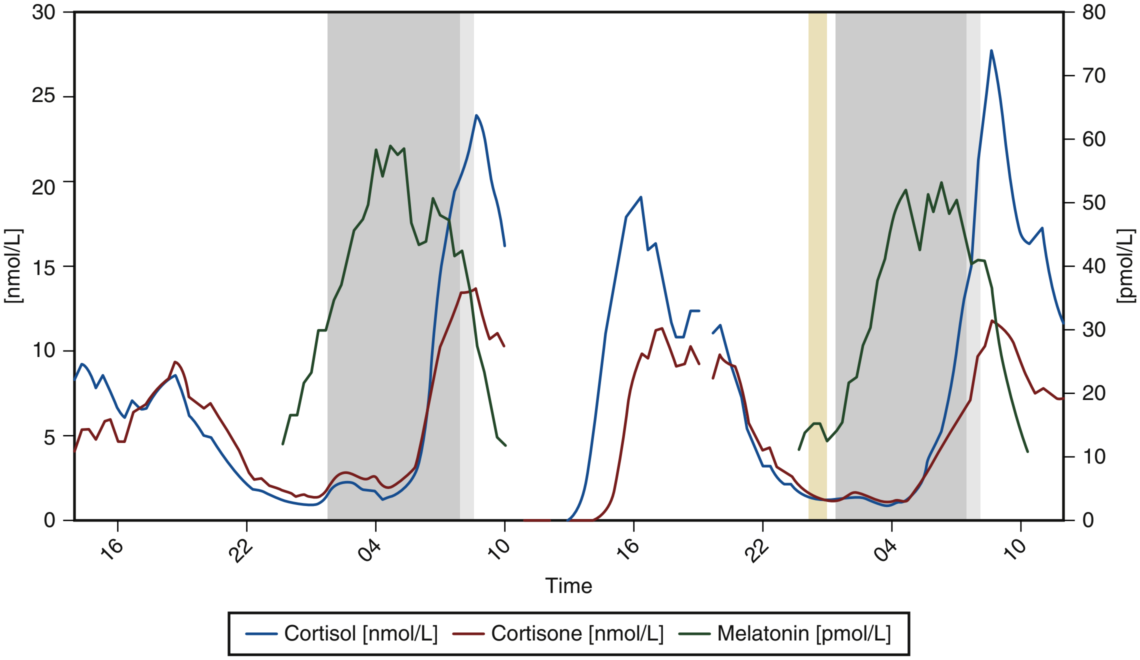


Fig. 3.4 Free tissue cortisol, cortisone, and melatonin profiles collected over 2 consecutive days using ambulatory microdialysis technology. In this illustrative example from a healthy male volunteer, samples were collected during normal living conditions in which the participant was free to choose ambient light conditions and sleep timing. As in the plasma example (Fig. 3.1), diurnal rhythms in cortisol *(blue)* and cortisone *(red)* can clearly be observed, with nadir levels during sleep *(dark grey-shaded area)* and peaks anticipating habitual rising times *(light grey)*. Deliberate exposure to bright blue–spectrum light on the second night of sampling *(yellow shaded area)* is here associated with temporary suppression of melatonin and a delay reaching peak night-time levels. (Source: Author's own unpublished data.)

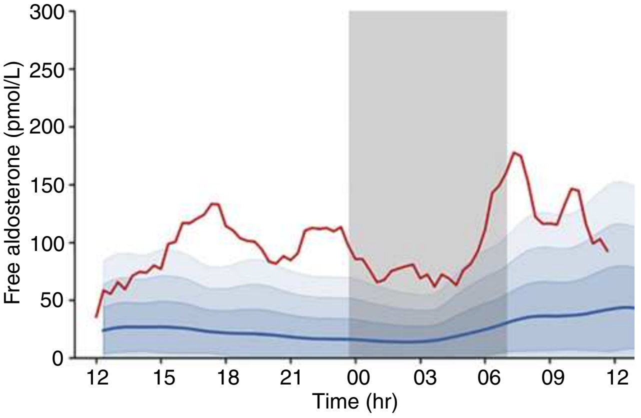


Fig. 3.5 **Example profile of the ultradian secretion of tissue free aldosterone in primary aldosteronism.** Free aldosterone *(red)* was measured over 24 hours in abdominal subcutaneous tissue using ambulatory microdialysis technology in a young adult undergoing investigation for hypertension. Mean aldosterone concentrations in 268 healthy adults *(blue, standard deviation in blue-shaded areas)* and the participant's sleep period *(grey)* are shown for reference. Adrenal vein sampling later confirmed unilateral overproduction of aldosterone. (Source: Unpublished data from the ULTRADIAN study [https://clinicaltrials.gov/ct2/show/NCT02934399), kindly provided by Dr. Eder Zavala, University of Birmingham.)

by which rhythm disturbance results in disease. Further studies in clinical populations are needed, and data from ambulatory high-frequency tissue sampling may provide new insights into the understanding of rhythm dynamics, as well as the mechanisms by which disruption results in endocrine and metabolic disease.

REFERENCES

1. Prize Announcement. *Nobel Media AB 2020*; 2017.
2. Asher G, Sassone-Corsi P. Time for food: The intimate interplay between nutrition, metabolism, and the circadian clock. *Cell*. 2015;161:84–92.
3. Dijk DJ, Landolt HP. Sleep physiology, circadian rhythms, waking performance and the development of sleep-wake therapeutics. *Handb Exp Pharmacol*. 2019;253:441–481.
4. Van Cauter E, Leproult R, Plat L. Age-related changes in slow wave sleep and REM sleep and relationship with growth hormone and cortisol levels in healthy men. *JAMA*. 2000;284:861–868.
5. Charloux A, Gronfier C, Lonsdorfer-Wolf E, et al. Aldosterone release during the sleep-wake cycle in humans. *Am J Physiol Endocrinol Metab*. 1999;276:43–49.
6. Medic G, Wille M, Hemels MEH. Short- and long-term health consequences of sleep disruption. *Nat. Sci. Sleep*. 2017;9:151–161.
7. Oster H, Challet E, Ott V, et al. The functional and clinical significance of the 24-h rhythm of circulating glucocorticoids. *Endocr Rev*. 2016;38:er20151080.
8. Reinke H, Asher G. Crosstalk between metabolism and circadian clocks. *Nat Rev Mol Cell Biol*. 2019;20:227–241.
9. Benloucif S, Burgess HJ, Klerman EB, et al. Measuring melatonin in humans. *J Clin Sleep Med*. 2008l;4:66–69.
10. Karamitri A, Jockers R. Melatonin in type 2 diabetes mellitus and obesity. *Nat Rev Endocrinol*. 2019;15:105–125.
11. Olcese JM, Skwarlo-Sonta K, Tosini G, et al. Melatonin signaling a key regulator of glucose homeostasis and energy metabolism. *Front Endocrinol*. 2019;1:488.
12. Qian J, Scheer FAJL. Circadian system and glucose metabolism: implications for physiology and disease. *Trends Endocrinol Metab*. 2016;27:282–293.
13. Tannenbaum GS, Martin JB. Evidence for an endogenous ultradian rhythm governing growth hormone secretion in the rat. *Endocrinology*. 1976;98:562–570.

14. Simon C, Brandenberger G. Ultradian oscillations of insulin secretion in humans. *Diabetes*. 2002;51:3–6.
15. Campos P, Herbison AE. Optogenetic activation of GnRH neurons reveals minimal requirements for pulsatile luteinizing hormone secretion. *Proc Natl Acad Sci*. 2014;111:18387–18392.
16. Choe HK, Kim H-D, Park SH, et al. Synchronous activation of gonadotropin-releasing hormone gene transcription and secretion by pulsatile kisspeptin stimulation. *Proc Natl Acad Sci U S A*. 2013;110:5677–5682.
17. Veldhuis JD, Iranmanesh A, Lizarralde G, et al. Amplitude modulation of a burstlike mode of cortisol secretion subserves the circadian glucocorticoid rhythm. *Am J Physiol*. 1989;257:E6–E14.
18. Jasper MS, Engeland WC. Synchronous ultradian rhythms in adrenocortical secretion detected by microdialysis in awake rats. *Am J Physiol Regul Integr Comp Physiol*. 1991;261:R1257–R1268.
19. Windle RJ, Wood SA, Shanks N, et al. Ultradian rhythm of basal corticosterone release in the female rat: Dynamic interaction with the response to acute stress. *Endocrinology*. 1998;139:443–450.
20. Henley DE, Leendertz J, Russell GM, et al. Development of an automated blood sampling system for use in humans. *J Med Eng Technol*. 2009;33:199–208.
21. Caratti G, Matthews L, Poolman T, et al. Glucocorticoid receptor function in health and disease. *Clin Endocrinol*. 2015;83:441–448.
22. Lightman SL, Wiles CC, Atkinson HC, et al. The significance of glucocorticoid pulsatility. *Eur J Pharmacol*. 2008;583:255–262.
23. Lightman SL, Conway-Campbell BL. The crucial role of pulsatile activity of the HPA axis for continuous dynamic equilibration. *Nat Rev Neurosci*. 2010;11:710–718.
24. Spiga F, Walker JJ, Terry JR, et al. HPA axis rhythms. *Compr Physiol*. 2014;4:1273–1298.
25. Ur E, Capstick C, McLoughlin L, et al. Continuous administration of human corticotropin-releasing hormone in the absence of glucocorticoid feedback in man. *Neuroendocrinology*. 1995;61:191–197.
26. Oster H, Damerow S, Kiessling S, et al. The circadian rhythm of glucocorticoids is regulated by a gating mechanism residing in the adrenal cortical clock. *Cell Metab*. 2006;4:163–173.
27. Jasper MS, Engeland WC. Splanchnic neural activity modulates ultradian and circadian rhythms in adrenocortical secretion in awake rats. *Neuroendocrinology*. 1994;59:97–109.
28. Weitzman ED, Fukushima D, Nogeire C. Twenty-four hour pattern of the episodic seretion of cortisol in normal subjects. *J Clin Endocrinol Metab*. 1971;33:14–22.
29. Walker JJ, Terry JR, Lightman SL. Origin of ultradian pulsatility in the hypothalamic-pituitary-adrenal axis. *Proc Biol Sci*. 2010;277:1627–1633.
30. Walker JJ, Spiga F, Waite E, et al. The origin of glucocorticoid hormone oscillations. *PLoS Biol*. 2012;10:e1001341.
31. Stavreva D, Wiench M, John S, et al. Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat Cell Biol*. 2009;11:1093–1102.
32. Flynn BP, Conway-Campbell BL, Lightman SL. The emerging importance of ultradian glucocorticoid rhythms within metabolic pathology. *Ann Endocrinol (Paris)*. 2018;79:112–114.
33. George CL, Birnie MT, Flynn BP, et al. Ultradian glucocorticoid exposure directs gene-dependent and tissue-specific mRNA expression patterns in vivo. *Mol Cell Endocrinol*. 2017;439:46–53.
34. Conway-Campbell BL, et al. Glucocorticoid ultradian rhythmicity directs cyclical gene pulsing of the clock gene period 1 in rat hippocampus. *J Neuroendocrinol*. 2010;22:1093–1100.
35. Young EA, Ribeiro SC, Ye W. Sex differences in ACTH pulsatility following metyrapone blockade in patients with major depression. *Psychoneuroendocrinology*. 2007;32:503–507.
36. Henley DE, Russell GM, Douthwaite JA, et al. Hypothalamic-pituitary-adrenal axis activation in obstructive sleep apnea: the effect of continuous positive airway pressure therapy. *J Clin Endocrinol Metab*. 2009;94:4234–4242.
37. Gibbison B, Spiga F, Walker JJ, et al. Dynamic pituitary-adrenal interactions in response to cardiac surgery. *Crit Care*. 2014;43:791–800.
38. Gibbison B, Keenan DM, Roelfsema F, et al. Dynamic pituitary-adrenal interactions in the critically ill after cardiac surgery. *J Clin Endocrinol Metab*. 2020;105:1–16.
39. Kalafatakis K, Russell GM, Harmer CJ, et al. Ultradian rhythmicity of plasma cortisol is necessary for normal emotional and cognitive responses in man. *Proc Natl Acad Sci U. S. A.* 2018:201714239.
40. Roelfsema F, Pereira AM, Biermasz NR, et al. Hormone secretion by pituitary adenomas is characterized by increased disorderliness and spikiness but more regular pulsing. *J Clin Endocrinol Metab*. 2014;99:3836–3844.
41. Roelfsema F. Patients with Cushing's disease secrete adrenocorticotropin and cortisol jointly more asynchronously than healthy subjects. *J Clin Endocrinol Metab*. 1998;83:688–692.
42. van Aken MO, Pereira AM, van Thiel SW, et al. Irregular and frequent cortisol secretory episodes with preserved diurnal rhythmicity in primary adrenal Cushing's syndrome. *J Clin Endocrinol Metab*. 2005;90:1570–1577.
43. Siragy HM, Vieweg WV, Pincus S, et al. Increased disorderliness and amplified basal and pulsatile aldosterone secretion in patients with primary aldosteronism. *J Clin Endocrinol Metab*. 1995;80:28–33.
44. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet*. 2017;18:164–179.
45. Crane BR, Young MW. Interactive features of proteins composing eukaryotic circadian clocks. *Annu Rev Biochem*. 2014;83:191–219.
46. Hastings M, O'Neill JS, Maywood ES. Circadian clocks: regulators of endocrine and metabolic rhythms. *J Endocrinol*. 2007;195:187–198.
47. Patke A, Murphy PJ, Onat OE, et al. Mutation of the human circadian clock gene CRY1 in familial delayed sleep phase disorder. *Cell*. 2017;169:203–215.e13.
48. Siegel JM. Sleep viewed as a state of adaptive inactivity. *Nat Rev Neurosci*. 2009;10:747–753.
49. Carskadon MA, Dement WC. Normal human sleep: An overview. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. Philadelphia, PA: Elsevier; 2011:16–26.
50. Borbély AA, Daan S, Wirz-Justice A, et al. The two-process model of sleep regulation: A reappraisal. *J Sleep Res*. 2016;25:131–143.
51. Trachsel L, Edgar DM, Seidel WF, et al. Sleep homeostasis in suprachiasmatic nuclei-lesioned rats: effects of sleep deprivation and triazolam administration. *Brain Res*. 1992;589:253–261.
52. Dijk DJ, Czeisler CA. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci*. 1995;15:3526–3538.
53. Wyatt JK, Ritz-De Cecco A, Czeisler CA, et al. Circadian temperature and melatonin rhythms, sleep, and neurobehavioral function in humans living on a 20-h day. *Am J Physiol*. 1999;277:R1152–R1163.
54. Czeisler CA, Duffy JF, Shanahan TL, et al. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*. 1999;284:2177–2181.
55. Vaughn BV, Giallanza P. Technical review of polysomnography. *Chest*. 2008;134:1310–1319.
56. Shrivastava D, Jung S, Saadat M, et al. How to interpret the results of a sleep study. *J Community Hosp Intern Med Perspect*. 2014;4:24983.
57. Czeisler CA, Weitzman ED, Moore-Ede MC, et al. Human sleep: Its duration and organization depend on its circadian phase. *Science*. 1980;210:1264–1267.
58. Lewy AJ, Cutler NL, Sack RL. The endogenous melatonin profile as a marker for circadian phase position. *J Biol Rhythms*. 1999;14:227–236.
59. Klerman H, St Hilaire MA, Kronauer RE, et al. Analysis method and experimental conditions affect computed circadian phase from melatonin data. *PLoS One*. 2012;7:e33836.
60. Gunn PJ, Middleton B, Davies SK, et al. Sex differences in the circadian profiles of melatonin and cortisol in plasma and urine matrices under constant routine conditions. *Chronobiol Int*. 2016;33:39–50.
61. Burgess HJ, Wyatt JK, Park M, et al. Home circadian phase assessments with measures of compliance yield accurate dim light melatonin onsets. *Sleep*. 2015;38:889–897.

62. Cain SW, Dennison CF, Zeitzer JM, et al. Sex differences in phase angle of entrainment and melatonin amplitude in humans. *J Biol Rhythms*. 2010;25:288–296.
63. Zeitzer JM, Duffy JF, Lockley SW, et al. Plasma melatonin rhythms in young and older humans during sleep, sleep deprivation, and wake. *Sleep*. 2007;30:1437–1443.
64. Scheuermaier K, Laffan AM, Duffy JF. Light exposure patterns in healthy older and young adults. *J Biol Rhythms*. 2010;25:113–122.
65. Foret J, Aguirre A, Touitou Y, et al. Effect of morning bright light on body temperature, plasma cortisol and wrist motility measured during 24 hour of constant conditions. *Neurosci Lett*. 1993;155:155–158.
66. Park SJ, Tokura H. Bright light exposure during the daytime affects circadian rhythms of urinary melatonin and salivary immunoglobulin A. *Chronobiol Int*. 1999;16:359–371.
67. Facer-Childs ER, Middleton B, Bagshaw AP, et al. Human circadian phenotyping and diurnal performance testing in the real world. *J Vis Exp*. 2020:(158);e60448. https://doi.org/10.3791/60448.
68. Phillips AJK, Vidafar P, Burns AC, et al. High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proc Natl Acad Sci U S A*. 2019;116:12019–12024.
69. Stone JE, Mcglashan EM, Quin N, et al. The role of light sensitivity and intrinsic circadian period in predicting individual circadian timing. *J Biol Rhythms*. 2020;20:1–13.
70. Prayag AS, Najjar RP, Gronfier C. Melatonin suppression is exquisitely sensitive to light and primarily driven by melanopsin in humans. *J Pineal Res*. 2019;66:1–8.
71. Gooley JJ, Chamberlain K, Smith KA, et al. Exposure to room light before bedtime suppresses melatonin onset and shortens melatonin duration in humans. *J. Clin. Endocrinol. Metab.* 2011;96:463–472.
72. McMullan CJ, Schernhammer ES, Rimm EB, et al. Melatonin secretion and the incidence of type 2 diabetes. *JAMA*. 2013;309:1388–1396.
73. Rubio-Sastre P, Scheer FAJL, Gómez-Abellán P, et al. Acute melatonin administration in humans impairs glucose tolerance in both the morning and evening. *Sleep*. 2014;37:1715–1719.
74. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proença C, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet*. 2009;41:89–94.
75. Lyssenko V, Nagorny CLF, Erdos MR, et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet*. 2009;41:82–88.
76. Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet*. 2009;41:77–81.
77. Dashti HS, Vetter C, Lane JM, et al. Assessment of MTNR1B Type 2 diabetes genetic risk modification by shift work and morningness-eveningness preference in the UK biobank. *Diabetes*. 2020;69:259–266.
78. Turek FW. Obesity and metabolic syndrome in circadian clock mutant mice. *Science*. 2005;308:1043–1045.
79. Eckel-Mahan K, Sassone-Corsi P. Metabolism and the circadian clock converge. *Am Physiol Soc*. 2013;93:107–135.
80. Knutson KL, von Schantz M. Associations between chronotype, morbidity and mortality in the UK Biobank cohort. *Chronobiol Int*. 2018;00:1–9.
81. Knutson KL, Van Cauter E. Associations between sleep loss and increased risk of obesity and diabetes. *Ann New York Acad Sci*. 2008;1129:287–304.
82. Stamatakis KA, Punjabi NM. Effects of sleep fragmentation on glucose metabolism in normal subjects. *Chest*. 2010;137:95–101.
83. Moller-Levet CS, Archer SN, Bucca G, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci*. 2013;110:E1132–E1141.
84. Liu Q, Shi J, Duan P, et al. Is shift work associated with a higher risk of overweight or obesity? A systematic review of observational studies with meta-analysis. *Int J Epidemiol*. 2018;47:1956–1971.
85. Kecklund G, Axelsson J. Health consequences of shift work and insufficient sleep. *BMJ*. 2016;355:1–13.
86. Souza RV, Sarmento RA, de Almeida JC, et al. The effect of shift work on eating habits: A systematic review. *Scand. J Work Environ Heal.* 2019;45:7–21.
87. Garde AH, Nabe-Nielsen K, Jensen MA, et al. The effects of the number of consecutive night shifts on sleep duration and quality. *Scand J Work Environ Heal*. 2020;46:446–453.
88. Thorpy M. International classification of sleep disorders. *Sleep Disord Med*. 2017;35:475–484.
89. Herxheimer A, Petrie KJ. Melatonin for the prevention and treatment of jet lag. *Cochrane Database Syst Rev*. 2002;10:74–78.
90. Herxheimer A. Jet lag. *BMJ Clin. Evid.* 2014;2014:1–16.
91. Lowden A, Åkerstedt T. Retaining home-base sleep hours to prevent jet lag in connection with a westward flight across nine time zones. *Chronobiol Int*. 1998;15:365–376.
92. Wittmann M, Dinich J, Merrow M, et al. Social jetlag: misalignment of biological and social time. *Chronobiol Int*. 2006;23:497–509.
93. Roenneberg T, Kantermann T, Juda M, et al. Light and the human circadian clock. In: Kramer A, Merrow M, eds. *Circadian Clocks*. Heidelberg: Springer Berlin; 2013:311–331.
94. Mota MC, Silva CM, Balieiro LCT, et al. Association between social jetlag food consumption and meal times in patients with obesity-related chronic diseases. *PLoS One*. 2019;14:1–14.
95. Roenneberg T, Allebrandt KV, Merrow M, et al. Social jetlag and obesity. *Curr Biol*. 2012;22:939–943.
96. Koopman ADM, Rauh SP, Van T Riet E, et al. The association between social jetlag, the metabolic syndrome, and type 2 diabetes mellitus in the general population: the New Hoorn Study. *J Biol Rhythms*. 2017;32:359–368.
97. Martínez-Lozano N, Barraco GM, Rios R, et al. Evening types have social jet lag and metabolic alterations in school-age children. *Sci Rep*. 2020;10:1–11.
98. Bhake RC, Leendertz JA, Linthorst ACEE, et al. Automated 24-hours sampling of subcutaneous tissue free cortisol in humans. *J Med Eng Technol*. 2013;37:180–184.
99. Bhake RC, Kluckner V, Stassen H, et al. Continuous free cortisol profiles – circadian rhythms in healthy men. *J Clin Endocrinol Metab*. 2019;104:5935–5947.
100. Bhake R, Russell GM, Kershaw Y, et al. Continuous free cortisol profiles in healthy men. *J Clin Endocrinol Metab*. 2020;105:1749–1761.
101. Wudy SA, Schuler G, Sánchez-Guijo A, et al. The art of measuring steroids: Principles and practice of current hormonal steroid analysis. *J. Steroid Biochem. Mol Biol.* 2018;179:88–103.

4

Prolactinomas and Disorders of Prolactin Secretion

Marcello D. Bronstein and Andrea Glezer

OUTLINE

History, 46
Epidemiology, 46
Pathogenesis, 47
Pathology, 47
  Nontumoral Lesions Associated with Hyperprolactinemia, 49
  Pituitary Prolactin-Secreting Carcinomas, 49
Differential Diagnosis of Hyperprolactinemia, 49
  Physiologic Hyperprolactinemia, 49
  Pharmacologic Hyperprolactinemia, 49
  Pathologic Hyperprolactinemia, 50
Diagnosis of Prolactinomas, 51
  Clinical Features, 51
  Laboratory Evaluation, 52
  Imaging, 53
Treatment, 53
  Medical Therapy, 53
  Surgical Therapy, 58
  Radiotherapy, 59
  Therapeutic Advances and Perspectives, 59
  Prolactinomas and Pregnancy, 59
  Treatment Planning and Follow-Up, 61

HISTORY

Nonphysiological hyperprolactinemia is the most prevalent hypothalamopituitary dysfunction and its main cause is prolactinomas. These are the most common pituitary tumors with a prevalence of 500 cases per million inhabitants.[1] These tumors are classified as microadenomas (diameter <10 mm) or macroadenomas (≥10 mm) and can be enclosed, expansive or invasive.[2] Prolactinomas are more common in women, especially microprolactinomas. Prolactin-secreting pituitary carcinomas are exceedingly rare.[3] As hyperprolactinemia usually is associated with menstrual disturbances, anovulation, and sexual impairment and prolactinomas commonly affect people under their 40s, these tumors are an important cause of infertility. Other causes of hyperprolactinemia, as physiologic and drug-induced, must be ruled out to diagnose prolactinoma properly. Also, awareness of laboratory and imaging pitfalls can avoid incorrect diagnosis and treatment.[4] New therapeutic approaches such as temozolomide for aggressive/resistant prolactinomas have been used.[5]

"If a woman is neither pregnant nor has given birth, and produces milk, her menstruation has stopped." This sentence, attributed to Hippocrates (*Aphorisms,* Section 5, #39), shows that the association between menstrual disturbances and inappropriate milk secretion has been known since ancient times. It was only in the twentieth century, however, that such disturbances were associated with hypersecretion of a pituitary hormone. In 1928, Striker and Grueter[6] identified a pituitary factor that was able to induce milk secretion in rabbits. Early in the eighteenth century, Hunter discovered that "pigeon's milk," a substance secreted by male and female pigeons to feed the young pigeon, is secreted by the crop, and according to Hunter, "the crop behaves like the udder of mammalian females regarding uterine gestation." In 1933, Riddle and colleagues[7] identified the stimulatory effect of a pituitary hormone on pigeons' crop growth and differentiation, which also controlled milk secretion in mammals, and called it *prolactin* (PRL). The pigeon's crop model later was used for the PRL bioassay. Coincidentally, the association of amenorrhea, infertility, and galactorrhea were described better around the 1930s.

The existence of a human PRL, distinct from growth hormone (GH), remained controversial, however, until the development of a specific radioimmunoassay for PRL in the early 1970s,[8] when it was shown that all the above-mentioned syndromes were linked to elevated serum PRL levels. The development of pituitary microsurgery and, later on, high-resolution imaging techniques showed that most of the cases with a normal sella on plain x-rays, described as nontumoral, were small pituitary tumors (microadenomas). There are many other causes of hyperprolactinemia that must be differentiated from prolactinomas so that a correct therapeutic approach can be instituted.

Cloning the genes encoding PRL and its receptor brought a better understanding of hyperprolactinemia physiology and pathophysiology.[9,10] More recently, kisspeptin has been shown to be involved in hypogonadism as a result of hyperprolactinemia,[11] pointing to future approaches for hyperprolactinemia-induced hypogonadism.[12]

KEY POINTS

- Prolactinomas usually are sporadic and benign pituitary tumors. Microprolactinomas frequently occur among women between 30 and 40 years old and mainly impact gonadal status and fertility, whereas macroprolactinomas are more common in men, and tumor mass effects such as headache and visual impairment are important issues.

EPIDEMIOLOGY

Hyperprolactinemia is the most prevalent hypothalamopituitary dysfunction, with prolactinomas being the main cause. Although prolactinomas represent nearly 50% of adenohypophyseal tumors in autopsy series, they correspond roughly to 25% of surgically-removed pituitary adenomas,[13] which is probably best explained by the excellent results of medical treatment of prolactinomas with dopaminergic agonists. Their prevalence is estimated at 500 cases per million individuals in the United States, with an incidence of 27 cases/million/year.[13] Nevertheless, prolactinoma standardized incidence rates could be higher, up 5.41 cases/100,000/year.[1] Vroonen et al. reviewed prolactinoma frequency in seven recent epidemiologic studies, confirming that prolactinoma is the most common pituitary adenoma, accounting for between 40% to 66% of tumors (Table 4.1).[1]

Prolactinomas present a sex bias, occurring ten times more frequently in females compared with males until the age of 50 years, when the proportion is 1:1.[14] Microprolactinomas (diameter <10 mm) represent about 60% of PRL-secreting adenomas and are more common in women than in men (20:1), whereas macroadenomas have roughly the same prevalence in both genders.[14] There are no sex-related differences in autopsy series. Prolactinomas occur at all ages, with the diagnosis made predominantly in the 20s and 30s in both sexes.[13] They are the most frequent pituitary adenomas encountered in childhood and adolescence.[13] Prolactinomas represent the minority of pituitary tumors diagnosed after age 65 years[13]; however, their diagnosis may be underestimated, and macroadenomas were much more frequent than microadenomas when prolactinomas were diagnosed after the menopause.[15] Prolactinomas are the most prevalent pituitary adenoma in the multiple endocrine neoplasia type 1 (MEN1) syndrome.[13] Isolated familial prolactinomas have been described[16] (see also Chapter 14). Also, sporadic prolactinomas in young patients and familial pituitary tumors are related to germline mutations in the *aryl hydrocarbon receptor interacting protein* (AIP).[17] Prolactinomas can also be associated with pheochromocytomas/paragangliomas in patients harboring germline mutations in the succinate dehydrogenase complex subunits or mutations/intragenic deletions in *MAX*.[18] Pituitary carcinoma, defined by the presence of metastasis, is a vey rare entity. Recently, 72 cases were reviewed in the literature, with those with PRL secretion being the second most frequent type (23%).[3] (See also Chapter 15.)

KEY POINTS

- Hyperprolactinemia is the most common neuroendocrinology disturbance, occurring in 70% of women with irregular menses and galactorrhea.

TABLE 4.1 Epidemiological Studies of the Prevalence of Pituitary Adenomas and Prolactinomas (%)

	Country	Pituitary Adenoma Prevalence	Prolactinomas %
Daly et al., 2006	Belgium	1/1064	66.2
Fontana et al., 2009	Switzerland	1/1241	56.0
Fernandez et al., 2010	United Kingdom	1/1289	57.0
Raappana et al., 2010	Finland	1/1471	51.0
Gruppetta et al., 2013	Malta	1/1321	47.0
Agustsson et al., 2015	Iceland	1/865	40.0
Fainstein Day et al., 2016	Argentina	1/1030	57.5

Reproduced with the permission of S. Karger AG, Basel. From Vroonen L, Daly A, F, Beckers A: Epidemiology and Management Challenges in Prolactinomas. *Neuroendocrinology* 2019;109:20-27. doi: 10.1159/000497746.

PATHOGENESIS

Similar to other pituitary adenomas, prolactinomas are probably monoclonal in origin. To date, the exact mechanisms leading to the development of lactotroph adenomas are not well established. Drug treatment is effective in 80% to 90% of cases; therefore, the minority of patients undergo surgical treatment, which means that the tumor samples available for research studies are not representative of prolactinomas in general. Among the genes that contain candidate pathological mutations are the pituitary tumor transforming gene (*PTTG*) and the heparin-binding secretory transforming gene (*HST*), both of which induce angiogenesis through fibroblast growth factors (FGF-2 and FGF-4). Estrogens seem to play a pivotal role in these mechanisms, because estrogen-modulating drugs (e.g., tamoxifen and raloxifene) and an antiestrogen drug inhibit PTTG expression in prolactinomas *in vitro* and prolactinoma growth *in vivo*. Reduction in the expression of the leukemia inhibitory factor and nerve growth factor and overexpression of bone morphogenetic protein 4 and the high-mobility group A2 (*HMGA2*) gene are among other events that may potentially be involved in prolactinoma tumorigenesis.[19]

A decrease in dopaminergic inhibition seems to play a role, at least a permissive one, in prolactinoma development. Several studies have shown the development of prolactinomas in mice with D2 receptor knockout, with this phenotype being more severe and developing more quickly in female animals and in animals treated with estrogens.[20] To date, no "natural" mutations have been found in the gene encoding the D2 receptor.[21] $PRLr^{-/-}$ mice exhibited more intense hyperprolactinemia and larger tumors than did age-matched $Drd2^{-/-}$ mice, and there were cumulative effects in compound homozygous mutant male mice. This fact suggests that PRL inhibits lactotrophs not only by the activation of hypothalamic dopamine neurons, but also directly within the pituitary in a dopamine-independent fashion.

Polymorphisms in tumor suppressor genes (genes that regulate cell cycle and receptor growth factors) have already been described, as those coding for TGF-beta type II receptor[22] and DR2.[23]

Prolactinomas associated with MEN1 present inactivating mutations characterized by loss of heterozygosity at locus 11q13 and mutations in the gene encoding menin. It is generally claimed that these prolactinomas are larger and more aggressive than their sporadic counterparts.[13] The recently described inactivating mutations in the gene encoding AIP, located on chromosome 11q13.3, are frequently found in isolated familial pituitary adenomas (mainly somatotropinomas, but also prolactinomas),[17] but rarely in sporadic pituitary tumors.[24]

PATHOLOGY

The terms *microadenoma* and *macroadenoma,* coined by Hardy,[25] denote pituitary adenomas measuring less or more than 1 cm, respectively. These adenomas are generally enclosed within a pseudocapsule, but also can be invasive. Macroprolactinomas also can be enclosed but usually expand to the optic chiasmal region or invade local structures such as the cavernous sinus or the sellar floor and sphenoid sinus. Prolactinomas can be classified according to pathology subtypes as sparsely-granulated lactotroph adenomas, densely-granulated lactotroph adenomas, and acidophilic stem cell adenomas, with the last two subtypes being more aggressive.[26] The sparsely-granulated lactotroph is the most common type, composed of medium-sized cells with chromophobic to slightly eosinophilic cytoplasm and a central, oval nucleus. Microcalcifications are commonly seen in lactotroph adenomas, and amyloid-like substance, forming small hyaline bodies, can also be found.[27]

Acidophilic stem cell and sparsely-granulated lactotroph adenomas are well characterized by immunohistochemistry[27] (Fig. 4.1).

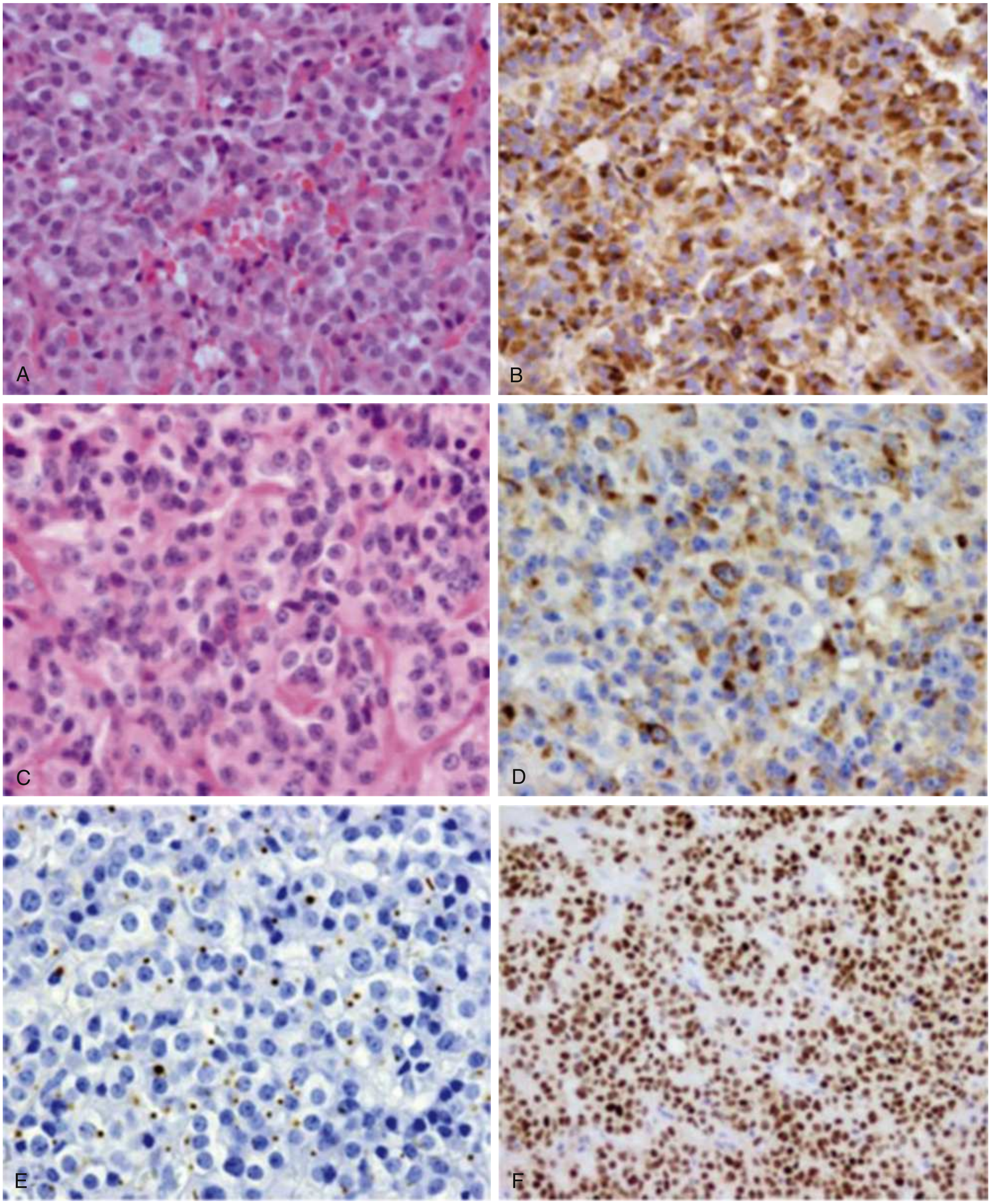


Fig. 4.1 **A,** Sparsely granulated lactotroph adenomas are composed of slightly acidophilic or chromophobic cells containing a central nucleus with well-dispersed chromatin (hematoxylin and eosin stain (H&E). **B,** Immunohistochemistry (IHC) for prolactin (PRL) in sparsely granulated lactotroph adenomas is characteristic, with paranuclear location in a dot-like pattern (Golgi-like pattern) (PRL-IHC). **C,** Acidophilic stem cell adenomas are characterized by cells with acidophilic cytoplasm with focal oncocytic changes (H&E). **D,** Immunoexpression of PRL is variable in the tumor and in the cells in acidophilic stem cells adenomas (PRL-IHC). **E,** Cytokeratin immunostaining highlights small fibrous bodies that are not present in either sparsely or densely granulated lactotroph adenomas (cytokeratin-IHC). **F,** All pathological variants of prolactinomas express the acidophilic lineage transcription factor Pit-1, here exemplified in an acidophilic stem cell adenoma (Pit-1–IHC). (From Lopes MBS. Pathology of prolactinomas: any predictive value? *Pituitary*. 2020;23:3–8.)

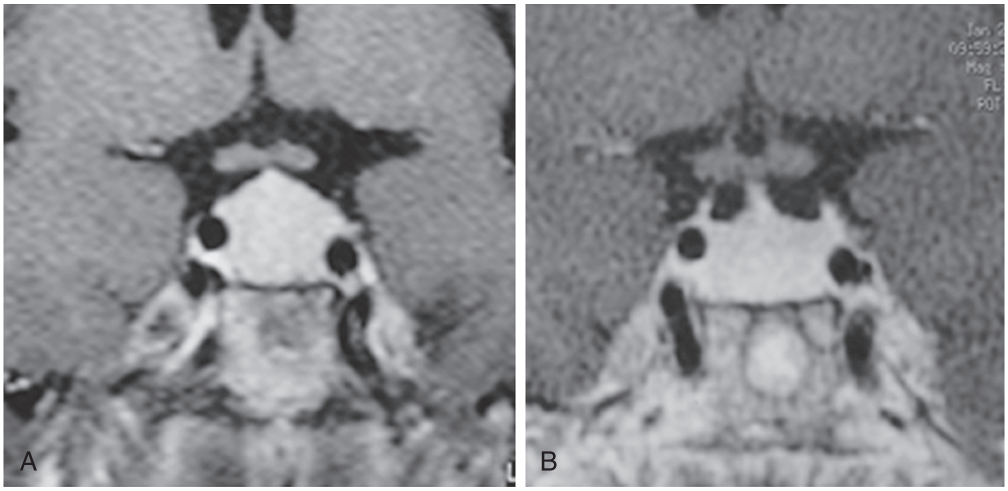


Fig. 4.2 Magnetic resonance imaging (gadolinium-enhanced Tl-weighted coronal views) of a 41-year-old woman with primary hypothyroidism **A**, before and **B**, after 1 month on levothyroxine replacement therapy. (From Bronstein MD. Problems in the differential diagnosis of the hyperprolactinemic patient. Clinical Endocrinology Update 2003. Miami: Endocrine Society; 2003:241–247.)

Nontumoral Lesions Associated With Hyperprolactinemia

Many nontumoral conditions can be mistaken for prolactinomas. Thyrotroph hyperplasia is associated with primary hypothyroidism as a result of loss of feedback (Fig. 4.2).[28] The correct diagnosis is important, because this condition regresses with thyroid hormone replacement.[28] Idiopathic lactotroph hyperplasia is a rare cause of hyperprolactinemia that can be mistaken for an expanding macroprolactinoma. The pituitary mass shape can help in the differential diagnosis. Inflammatory lesions such as lymphocytic hypophysitis (occurring mainly during pregnancy and the puerperium) and sarcoidosis can be misdiagnosed as a lactotroph adenoma.[29]

Pituitary Prolactin-Secreting Carcinomas

Pituitary PRL-secreting carcinomas are exceedingly rare tumors that exhibit morphologic features indistinguishable from those of PRL-secreting adenomas (see also Chapter 15). In addition to local invasiveness, distant intracranial and extracranial metastases in the nervous system and visceral metastases are necessary for the diagnosis of malignancy. Regarding intracranial lesions, sometimes it is difficult to distinguish between contiguity of an invasive prolactinoma and true metastasis from a PRL-secreting carcinoma.[3]

DIFFERENTIAL DIAGNOSIS OF HYPERPROLACTINEMIA

The main causes of hyperprolactinemia are listed in Table 4.2.

Physiologic Hyperprolactinemia

Throughout pregnancy, owing to estrogen-induced hypertrophy and hyperplasia of lactotrophs, the size of a normal pituitary increases up to 136%, and serum PRL levels progressively increase. A recent study including 378 healthy pregnant women showed the following median PRL serum levels in each trimester: 79.0, 144.6, and 306.6 ng/mL in the first, second, and third trimester, respectively.[30] Serum PRL levels decline quickly after delivery but remain slightly increased in nursing women for several months, especially with breastfeeding.

Because exercise and nonspecific stress are physiologic causes of hyperprolactinemia, there is a concern that the stress-induced PRL increase could lead to hormone elevation during venipuncture, and a period of rest before drawing blood is recommended if those conditions are suspected.[31]

TABLE 4.2 Causes of Hyperprolactinemia

Physiologic
- Pregnancy and puerperium
- Neonatal period
- Physical activity
- Stress

Drug-Induced
- Dopamine receptor blockers
  - Sulpiride
  - Chlorpromazine
  - Haloperidol
  - Risperidone
  - Metoclopramide
  - Domperidone
- Serotonin reuptake inhibitors
- Cimetidine
- Tricyclic antidepressants
- Verapamil
- Methyldopa
- Protease inhibitors

Pathologic
- Pituitary disease
  - Prolactinomas
  - Acromegaly
  - Cushing disease
  - Nelson syndrome
  - Lymphocytic hypophysitis
  - "Empty sella" syndrome
- Hypothalamic disease and hypothalamus/pituitary disconnection
  - Nonfunctioning pituitary adenomas
  - Meningiomas
  - Dysgerminomas
  - Craniopharyngiomas
  - Inflammatory/granulomatous
  - Sarcoidosis
  - Histiocytosis
  - Stalk section
  - Vascular
  - Actinic
- Neurogenic
  - Chest wall lesions
  - Spinal cord lesions
- Miscellaneous
  - Primary hypothyroidism
  - Adrenal insufficiency
  - Uremia
  - Cirrhosis
  - Paraneoplastic
  - Idiopathic

KEY POINTS
- Pregnancy, drugs, hypothyroidism, and uremia should be ruled out before diagnosing prolactinomas.

Pharmacologic Hyperprolactinemia

Among medications that increase serum PRL, dopamine receptor blockers are the most potent. Neuroleptics (e.g., sulpiride,

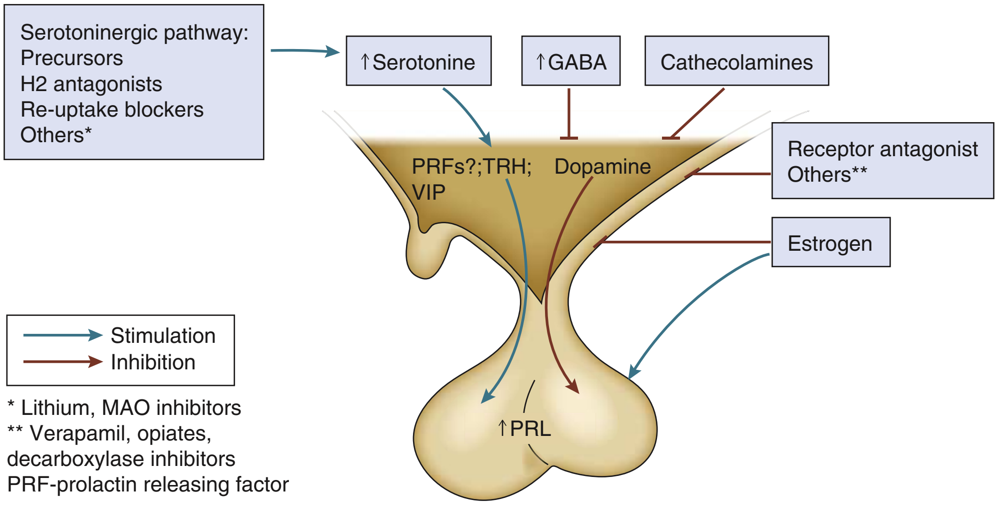


Fig. 4.3 Mechanisms related to drug-induced hyperprolactinemia. *GABA,* Gamma aminobutyric acid; *MAO,* monoamine oxidase; *PRF,* prolactin-releasing factor; *PRL, prolactin; TRH,* thyrotropin-releasing hormone; *VIP,* vasoactive intestinal polypeptide.

haloperidol, chlorpromazine, risperidone) and antiemetic drugs (e.g., metoclopramide, domperidone) can elevate serum PRL to levels that usually are detected with prolactinomas. Serotoninergic and antihistaminergic drugs are less potent than antidopaminergic medications. The calcium channel blocker verapamil elevates serum PRL levels, probably by decreasing central dopamine generation, and possibly through N-type calcium channels. It was shown that protease inhibitors used for the treatment of acquired immunodeficiency syndrome can cause hyperprolactinemia, but the mechanism is unknown. A detailed inquiry about drug use is mandatory for all hyperprolactinemic patients.[32]

Hyperprolactinemia may occur in patients on opioids, although its prevalence and the impact of dose, route, and type of drug have not been clearly established. Fig. 4.3 summarizes mechanisms related to drug-induced hyperprolactinemia.[33]

The Endocrine Society guidelines recommend treating hypogonadism induced by hyperprolactinemia related to antipsychotic drugs with hormonal replacement, if there is no contraindication.[31] Nevertheless, the association of psychiatric drugs inducing hyperprolactinemia and the use of dopamine agonists (DAs) has been reported. Such combined treatment must only be used with the agreement of the psychiatrist and with close follow-up.[34]

Pathologic Hyperprolactinemia

Hyperprolactinemia is present in about 40% of acromegalic patients as a result of GH/PRL cosecretion by the same or by different tumor cells, or secondary to hypothalamus-pituitary disconnection. Thus, the Endocrine Society recommends screening all prolactinoma patients for GH hypersecretion at the time of diagnosis.[31] As some studies have pointed to the development of acromegaly in about 4% of prolactinoma patients during long-term follow-up under DA therapy, regular monitoring of insulin-like growth factor–1 (IGF-1) levels is suggested by some authors.[35]

Hyperprolactinemia secondary to impaired hypothalamiotuberoinfundibular dopamine secretion or to stalk or even intrapituitary disconnection can be caused by tumors, inflammatory diseases, or trauma. In these cases, PRL is produced by the normal lactotrophs and rarely exceeds 150 μg/L. The differential diagnosis of macroprolactinomas is mainly clinical and is seen in nonfunctioning pituitary adenomas ("pseudoprolactinomas") and, to a lesser extent, with craniopharyngiomas.[36] Other tumoral lesions, such as meningiomas and chordomas, and nontumoral conditions, such as "empty sella" syndrome and even intrasellar aneurysms, can be associated with hyperprolactinemia, however.[37] The differential diagnosis between macroprolactinomas and pseudoprolactinomas is crucial regarding primary treatment, which is medical for macroprolactinomas and surgical for pseudoprolactinomas. Patients with nonfunctioning tumors treated with DAs were previously considered to have resistant macroprolactinomas, owing to the absence of tumor shrinkage even with the (obvious) PRL decrease to very low or undetectable levels (Fig. 4.4).[38] A clue to differentiate pseudoprolactinomas from true prolactinomas is the dramatic early PRL decrease with a low dose of DA.

Primary hypothyroidism can be associated with hyperprolactinemia, presumably as a result of high thyrotropin-releasing hormone levels that stimulate PRL release and presumably reduce PRL metabolic clearance. Thyrotroph hyperplasia may occur, leading to pituitary enlargement mimicking a pituitary adenoma (see Fig. 4.2). Cushing disease and adrenal insufficiency can be associated with hyperprolactinemia, which also can be present in Nelson syndrome.

Uremia can be associated with hyperprolactinemia, mainly in patients with end-stage renal disease. The mechanism probably is related to reduced PRL clearance and to a presumably reduced dopaminergic tonus. Hyperprolactinemia is present in up to 20% of patients with liver cirrhosis, probably because of an unbalanced estrogen-to-androgen ratio and to an altered dopaminergic tone. Nipple manipulation and chest-wall lesions, such as herpes zoster and surgical scars, may increase serum PRL via stimulation of neuron pathways going through the spinal cord.[29] Ectopic production of PRL has rarely been described.[39] Finally, when all the abovementioned causes are ruled out, hyperprolactinemia is considered idiopathic or functional.[40] It is likely, however, that most patients with this condition may harbor tiny microprolactinomas undetected by imaging. De Bellis et al.[41] pointed out the presence of antipituitary antibodies in patients with idiopathic hyperprolactinemia, suggesting an autoimmune etiology for a subset of cases. Also, Newey et al. reported three sisters with hyperprolactinemia attributed to a heterozygous loss-of-function mutation in the PRL receptor gene, resulting in prolactin insensitivity.[42]

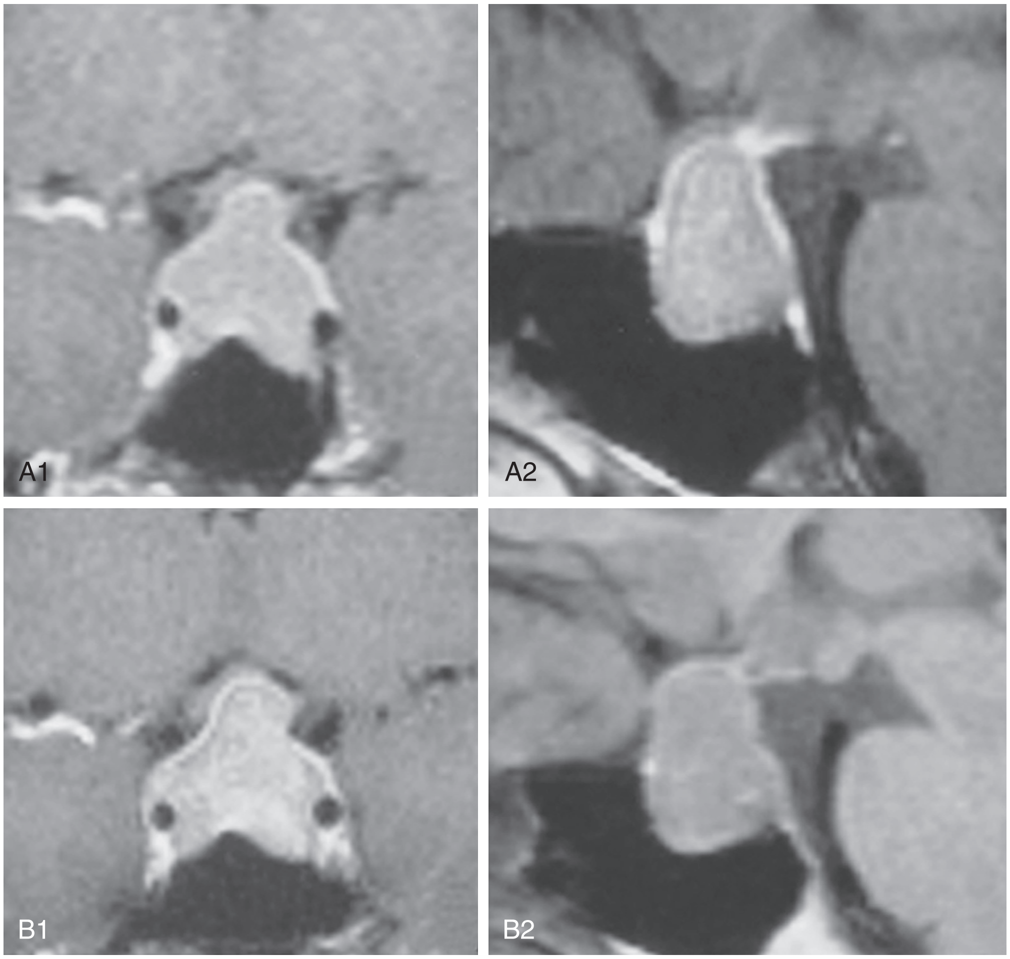


Fig. 4.4 Magnetic resonance imaging of a 41-year-old woman with a nonfunctioning pituitary adenoma and serum prolactin level of 51 ng/mL before (**A1,** coronal view; **A2,** sagittal view) and during the twelfth month of cabergoline treatment (**B1,** coronal view; **B2,** sagittal view). No tumor shrinkage was observed, despite serum prolactin decrease to 1.2 ng/mL. (From Bronstein MD. Problems in the differential diagnosis of the hyperprolactinemic patient. Clinical Endocrinology Update 2003. Miami: Endocrine Society; 2003:241–247.)

KEY POINTS

- Hyperprolactinemia causes hypogonadotrophic hypogonadism, especially because of impairment in gonadotropin-releasing hormone secretion and pulsatility, via kisspeptine. The gold standard treatment for prolactinoma is dopamine agonists, because of their efficacy in normalizing serum prolactin levels, restoring the gonadal axis, and reducing tumor dimensions in 80% to 90% of cases. Neurosurgery, radiotherapy, and chemotherapy can be indicated in partially/completely resistant tumors.

DIAGNOSIS OF PROLACTINOMAS

Clinical Features

Hyperprolactinemia and its impact on the gonadotropic axis are the hallmark of microprolactinomas and macroprolactinomas and their clinical manifestations. Macroprolactinomas also may cause neurologic and visual disturbances and impairment of other pituitary functions owing to the tumor mass effect (Table 4.3).[43] Loss of visual fields and impairment of visual acuity are the main ophthalmologic manifestations. Headache is the most common neurologic presentation; this is typically attributed to dural stretch or cavernous sinus invasion, and less frequently to the clinical symptoms of pituitary apoplexy. Other rare presentations of invasive macroprolactinomas

TABLE 4.3 Clinical Manifestations of Prolactinomas

Related to Hyperprolactinemia

- Gonadal impairment
  - Menstrual disturbances
  - Infertility
  - Sexual dysfunction
  - Osteoporosis
- Galactorrhea

Related to Tumor Mass Effect

- Visual disturbances
  - Visual field defects
  - Reduced visual acuity
  - Disturbances of ocular motility
- Headache
- Rare manifestations
  - Trigeminal neuralgia
  - Hydrocephalus
  - Otoneurologic
  - Neuropsychiatric
- Impairment of other pituitary functions

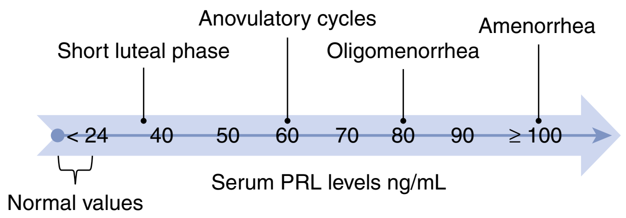


Fig. 4.5 Serum prolactin elevation and gonadal impairment. *PRL,* Prolactin.

are hydrocephalus, neuropsychiatric manifestations, and otoneurologic manifestations.[44]

Women generally present with a correlation between the degree of serum PRL elevation and gonadal impairment, ranging from a short luteal phase to amenorrhea (Fig. 4.5).[45] The main mechanism involved in hypogonadism related to hyperprolactinemia is the suppression of the pulsatile secretion of gonadotropin-releasing hormone, via kisspeptin.[11] Its administration reverses gonadal status and restores fertility in animals and in hyperprolactinemic women.[12]

In addition to menstrual disturbances, hyperprolactinemic women complain of loss of libido, vaginal dryness, dyspareunia, and psychological distress.[46] Similar to other disorders associated with hypogonadism, osteoporosis is a frequent finding in women with hyperprolactinemia/prolactinomas, mainly in prolactinomas with longstanding hypogonadism. Although a direct effect of PRL on bone has been considered, there is evidence that hyperprolactinemia is not a risk factor in itself for the development of osteoporosis, with the associated hypoestrogenism being the major determinant.[47]

Galactorrhea is a frequent finding in women with hyperprolactinemia, but it can be absent, mainly in patients with macroprolactinomas associated with severe hypogonadism. Breast examination, if not adequate, also may fail to disclose mild galactorrhea. These facts may explain the discrepancy among series, which show a prevalence of galactorrhea ranging from 30% to 84%.[48] Galactorrhea also may occur in normoprolactinemic women. A study including 235 patients with galactorrhea associated with diverse conditions showed that serum PRL was normal in 86% of women with idiopathic galactorrhea without amenorrhea.[49] The mechanism is unknown, and is attributed to different causes such as mammary hypersensitivity to PRL or "occult" transient hyperprolactinemia.

Hyperprolactinemia in men is predominantly associated with macroprolactinomas. The main clinical manifestations are hypogonadism, which is usually severe, along with loss of libido and reduced body hair growth, visual impairment, and headache.[50] Testosterone replacement without serum PRL normalization seldom restores libido, an observation that points to a direct effect of PRL on sexual behavior, as previously suggested by animal models.[51] Galactorrhea is far less frequent in men than in women with prolactinomas. However, when present in men harboring pituitary tumors, it strongly suggests the presence of a prolactinoma.[52] Osteoporosis is also present in hyperprolactinemic men.[47]

Although the prevalence of prolactinomas in both genders is higher in the second and third decades of life, prolactinomas can occur in the elderly and in younger individuals.[53,54] The predominance of larger tumors in children and adolescents points to molecular mechanisms influencing proliferation rather than the time course of the disease influencing the progression of prolactinoma size and invasiveness.

Laboratory Evaluation

Except in cases where venipuncture induced significant stress, a single determination of high serum PRL level confirms hyperprolactinemia.[31] Ruling out other causes of hyperprolactinemia, the diagnosis of prolactinoma can be established. Microprolactinomas usually present PRL levels of 100 to 200 ng/mL, although values that are only two times the upper limit of the normal range can also be found (normal range <24–30 ng/mL).[55] Macroprolactinomas are usually associated with PRL values above 250 ng/mL, and giant prolactinomas with levels higher than 1000 ng/mL.[31,56] Mainly in macroprolactinomas, the assessment of other pituitary functions is mandatory. IGF-1 levels must be monitored either to verify GH deficiency or to identify the rare cases of prolactinomas that progress with PRL/GH cosecretion.[57] As restoration of pituitary function may occur after adenoma removal by surgery or tumor shrinkage attained by medical therapy,[58] these assessments have to be repeated after such procedures to avoid unnecessary hormone replacement.

Concerning other causes of hyperprolactinemia, serum PRL levels rarely exceed 150 ng/mL in hypothalamo-pituitary disconnection, as lesions that mimic macroprolactinomas, especially nonfunctioning pituitary adenomas, generally exhibit serum PRL levels below this value (pseudoprolactinomas).[36] Nevertheless, a laboratory artefact may lead to an erroneous differential diagnosis between macroprolactinomas and pseudoprolactinomas: when serum PRL is evaluated by two-site immunometric assays, large amounts of antigen—PRL in this case—saturate both capture and signal antibodies, impairing their binding and causing serum PRL to be underestimated (the so-called high-dose "hook effect"). Patients bearing macroprolactinomas with extremely high serum PRL levels (generally >10,000 ng/mL, depending on the assay measuring range) may present with falsely lower levels, within the 30 to 150 ng/mL range, causing the patient to be misdiagnosed as harboring a nonfunctioning pituitary adenoma. To avoid unnecessarily performing surgery (which is the treatment of choice for nonfunctioning tumors), PRL assays with serum dilution or using two-step incubation are recommended in patients with macroadenomas who may harbor a prolactinoma.[59]

If such assays are not readily available, clinical clues pointing to prolactinomas are patient age younger than 50 years, presence of galactorrhea in male patients, and tumor shrinkage in response to treatment with DA drugs, as shown by rapid visual improvement in cases with chiasmal compression or rapid tumor reduction evidenced by magnetic resonance imaging (MRI).[52]

Another laboratory pitfall concerns the presence of high serum PRL levels in subjects with few or no symptoms related to PRL excess. Human PRL in the circulation manifests a marked size heterogeneity, with three forms (23, 50, and 150–170 kD) that are indistinguishable by routine assays. The 23-kD form (little PRL) is the most common form, but serum PRL can be elevated secondary to the presence of 150- to 170-kD aggregates with low biological activity (big-big PRL), leading to *macroprolactinemia.*[60] The presence of molecular aggregates with low biological activity, such as big-big PRL, should be suspected when high serum PRL levels are detected in patients with no or scarce signs and symptoms related to hyperprolactinemia.[4,61] As the gold standard assay, gel filtration chromatography, is costly and time-consuming, precipitation with polyethylene glycol is usually employed to rule out macroprolactinemia.[31] Macroprolactinemia is a common finding, occurring in 8% to 42% of all cases of hyperprolactinemia.[4]

Despite controversies concerning the biological activity of PRL aggregates,[62] most patients with macroprolactinemia and normal monomeric PRL levels[63] do not manifest clinical features related to hyperprolactinemia and do not need any treatment. To avoid unnecessary medical or surgical procedures, macroprolactin screening is mandatory when clinical features and serum PRL assay results are conflicting.

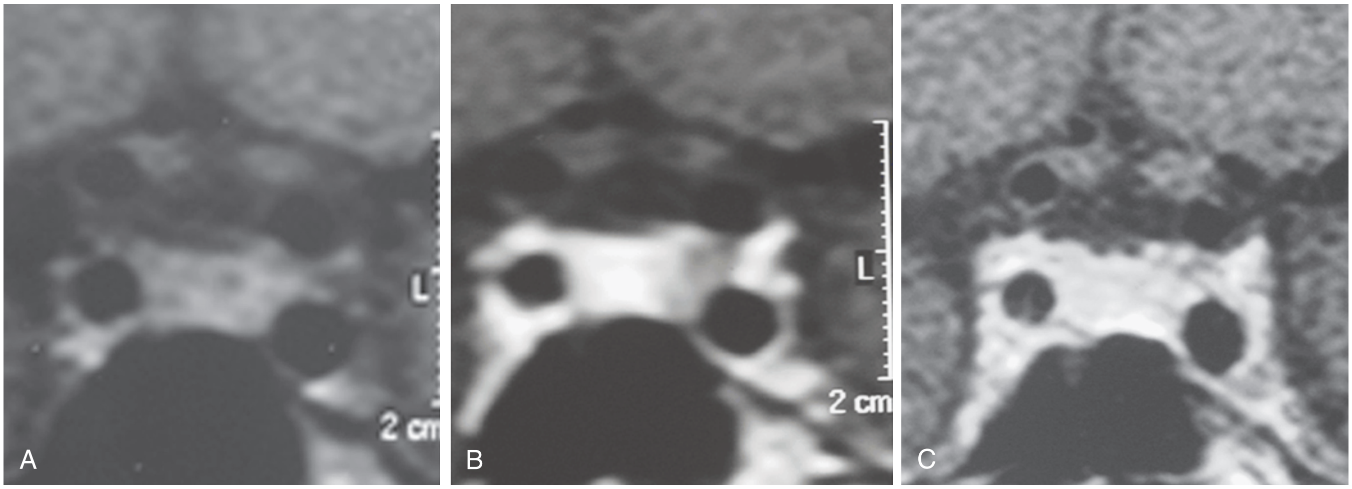


Fig. 4.6 Magnetic resonance imaging (coronal views) of a microprolactinoma disclosed only during dynamic gadolinium-enhanced sequence. **A,** Before contrast injection. **B,** Lack of contrast enhancement on the left side of the pituitary gland, suggestive of microadenoma. **C,** Whole gland contrasted.

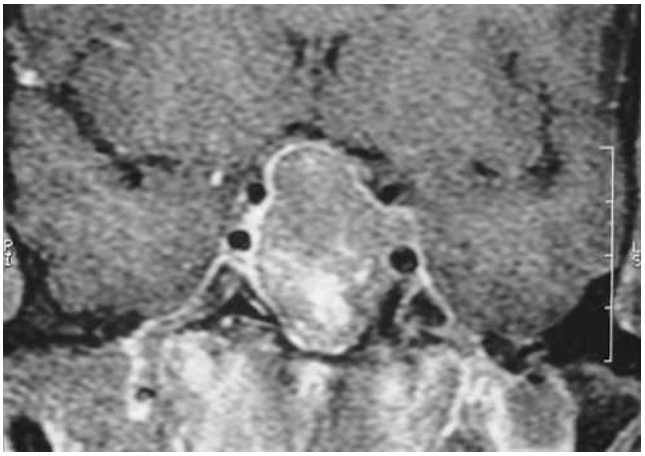


Fig. 4.7 Magnetic resonance imaging (gadolinium-enhanced coronal view) showing optic chiasm compression and left cavernous sinus invasion by a macroadenoma.

KEY POINTS

- Macroprolactinemia is a frequent laboratory finding and, when associated with normal monomeric values, is a benign condition, not needing specific treatment. Hook effect, which underestimates serum PRL levels, can be avoided with serum dilution and is important to consider in the differential diagnosis of prolactinomas.

Imaging

MRI is currently the gold-standard imaging method for diagnosis and treatment follow-up of micro- and macroprolactinomas. MRI is superior to computed tomography (CT) for tumor boundary delineation, especially regarding its associations with the optic chiasm and cavernous sinus (Figs. 4.6 and 4.7). Additionally, tumor consistency, presence of hemorrhage, and presence of cystic lesions are better shown by MRI. CT still has its place, as it can be used to assess bone erosion and calcification and may help surgical planning.[64] Although MRI is currently the most sensitive imaging method for microprolactinomas, tumors measuring less than 2 mm may remain undetected by this technique, supporting the concept that the so-called idiopathic hyperprolactinemia actually is related to tiny microprolactinomas.[40]

It should be stressed that pitfalls such as macroprolactinemia and pituitary incidentalomas may lead to an erroneous prolactinoma diagnosis, leading to unnecessary medical, or even surgical, procedures.[4] Fig. 4.8 shows a flow diagram for the diagnostic evaluation of hyperprolactinemia.

TREATMENT

Therapy for hyperprolactinemia/prolactinomas aims to reverse the symptoms caused by hormone hypersecretion and the neurologic and visual manifestations attributed to tumor mass effect. The ideal treatment must spare or even improve other pituitary dysfunctions if present.

Treatment of secondary hyperprolactinemia is intended to treat or remove the cause of the disorder, such as primary hypothyroidism, surgical removal of a nonfunctioning pituitary adenoma, and withdrawal of drugs such as risperidone, when possible, to bring serum PRL down to normal levels.[32,36]

Regarding prolactinoma treatment two powerful therapeutic advances came about almost simultaneously in the 1970s: the improvement of pituitary microsurgery[2] and the development of ergot derivatives with potent dopaminergic agonistic activity.[65] Additionally, radiation therapy, mainly carried out by stereotactic techniques, has a place, albeit restricted (Table 4.4). The therapeutic strategy must consider several aspects, such as the patient's clinical presentation, the differences in natural history between microadenomas and macroadenomas, the desire for pregnancy, and the patient's treatment preference, if applicable.

Medical Therapy

Bromocriptine. Bromocriptine (BRC), an ergot derivative, was the first DA used in clinical practice. BRC binds and stimulates the seven–membrane-spanning domain dopamine D2 receptors in normal and in adenomatous lactotrope cells, inducing activation of the G_i receptor (negatively coupled to adenylate cyclase), which ultimately inhibits PRL synthesis and secretion.

Pharmacokinetic studies with BRC show that, after a single oral dose of 2.5 mg, serum levels peak, and maximal suppressive action

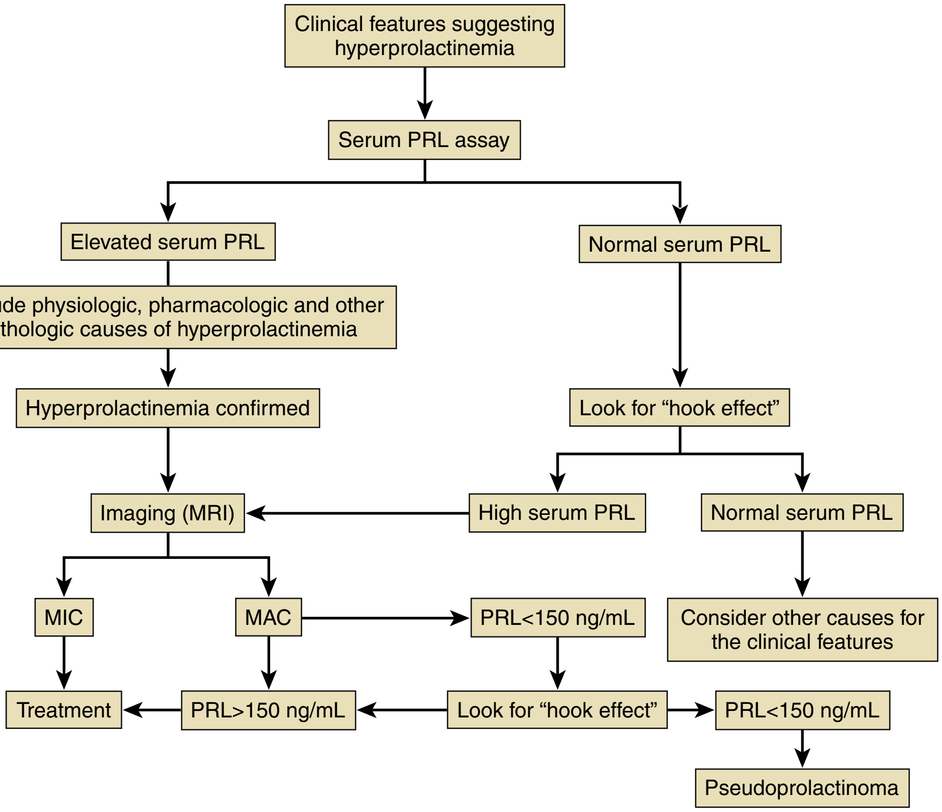

Clinical features suggesting hyperprolactinemia
Serum PRL assay
Elevated serum PRL
Normal serum PRL
Exclude physiologic, pharmacologic and other pathologic causes of hyperprolactinemia
Hyperprolactinemia confirmed
Look for "hook effect"
Imaging (MRI)
High serum PRL
Normal serum PRL
MIC
MAC
PRL<150 ng/mL
Consider other causes for the clinical features
Treatment
PRL>150 ng/mL
Look for "hook effect"
PRL<150 ng/mL
Pseudoprolactinoma

Fig. 4.8 Flow diagram for the diagnostic evaluation of hyperprolactinemia. *MAC,* Macroadenoma; *MIC,* microadenoma; *MRI,* magnetic resonance imaging; *PRL,* prolactin.

TABLE 4.4 Treatment of Prolactinomas

Medical
Dopamine agonist drugs
Sexual steroid replacement (for selective microprolactinomas)

Surgical
Transsphenoidal approach
  Sublabial
  Endonasal (with or without endoscope)

Radiotherapy
Conventional
Stereotactic
  Cobalt-60 or linear acceleration
  Single shot ("radiosurgery") or multiple shots
Proton beam

occurs between 1 and 3 hours. The drug has a relatively short mean elimination half-life (~6.2 hours) and generally is administered twice a day, although there are data indicating the efficacy of once-daily administration.[65]

BRC treatment results in serum PRL normalization and clinical improvement in most patients with hyperprolactinemia/prolactinomas. A compilation of 13 series from the literature, encompassing 286 women treated with BRC, showed PRL normalization and return of menses in 80% and 73 to 100% of cases, respectively.[66] Although less frequent, microprolactinoma in men also can be treated successfully with DAs. We have treated 11 men with microprolactinoma with DAs, and successfully achieved serum PRL and serum testosterone normalization in 73% and 86%, respectively.[67]

Because of the mass-reducing effect of BRC, this DA has also become a primary treatment for macroprolactinomas, as it can help avoid surgery, which is often noncurative.[68] Patients with giant prolactinomas (diameter >4 cm) also achieved disease control with primary BRC therapy.[69] In our series, serum PRL was normalized in 60% of women with macroprolactinomas (mean BRC dose 7 mg/day); 72% of these women had recovery of their menses. Menses also were restored in 44% of patients who remained hyperprolactinemic.[70] Regarding macroprolactinoma in men, serum PRL and testosterone levels normalized in 67% and 85%, respectively, of 66 patients bearing macroprolactinomas, showing that, in men also, the gonadotropic axis may normalize even when PRL is not fully normalized. We obtained 80% tumor reduction in primary-treated macroprolactinoma patients.[67] The time course of macroprolactinoma shrinkage is highly variable. Some patients experience a dramatic improvement of visual acuity 12 hours after BRC has been introduced, with tumor shrinkage documented within 1 week. Therefore, visual improvement usually occurs early after drug introduction and should suggest continuing medical treatment instead of performing surgery (Fig 4.9). In others, significant tumor reduction is observed only 1 year after medical therapy is started.[68]

Macroprolactinoma reexpansion during successful medical therapy, albeit uncommon, may occur, and can lead to visual impairment or headache recurrence.[71] Another rare situation is visual deterioration despite initial improvement, provoked by herniation of the optic chiasm secondary to tumor shrinkage. These two situations generally require surgical intervention, although drug reduction or discontinuation may repair the chiasmal herniation.[72] In addition, pituitary surgery is often needed for a third rare situation, cerebrospinal fluid (CSF) leakage, occurring in macroprolactinomas invading the sphenoid sinus in cases treated with DAs (Fig. 4.10). In such cases, the tumor shrinkage behaves as the cork removed from a wine bottle, opening a pathway for CSF flow. CSF leakage, which carries a risk of meningitis, often occurs shortly after the beginning of medical therapy but may manifest several months later.[73,74]

BRC-induced shrinkage of macroprolactinomas can restore other impaired anterior pituitary functions as a result of the restoration of hypothalamopituitary connection.[75] To avoid the maintenance of unnecessary hormonal replacement, pituitary function must be reassessed in patients with macroprolactinoma and previous hypopituitarism successfully treated with DAs.[75]

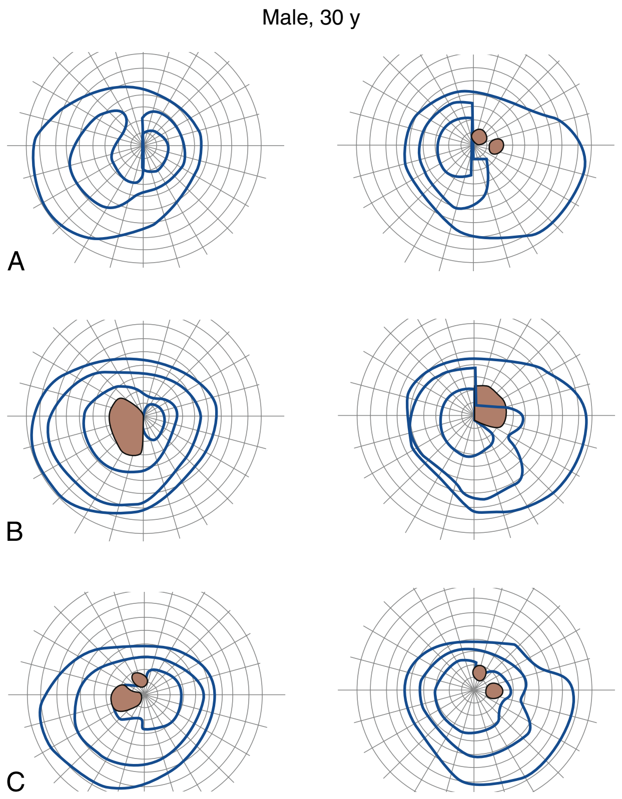


Fig. 4.9 Visual fields (Goldman perimetry) of a 30-year-old man harboring a macroprolactinoma. Impaired visual fields **A**, before treatment, **B**, showed amelioration 3 days after the beginning of bromocriptine treatment and **C**, were almost normalized after 28 days of drug therapy.

One of the drawbacks of medical therapy for prolactinomas is the need for long-term treatment. Nevertheless, a retrospective study from our group showed that 25.8% of 62 patients with microprolactinomas and 15.9% of 69 patients with macroprolactinomas treated with BRC for a median of 47 months continued to be normoprolactinemic for a median of 44 months after drug withdrawal.[76] There were no statistically significant differences regarding age, gender, BRC initial dose and duration of use, tumor size, pregnancy during treatment, or previous pituitary surgery or radiotherapy among patients who continued to be normoprolactinemic and patients who did not. Other reports from the literature point to a percentage of patients who remained normoprolactinemic after BRC cessation ranging from 6.6% to 50%.[76,77] The difference in long-term findings versus short-term findings may be explained by microscopic alterations of the lactotrope during BRC administration, suggesting a cytostatic effect related to short-term therapy and a cytocidal effect related to long-term treatment, which could explain the maintenance of normoprolactinemia after drug withdrawal.[78]

Another factor that may influence remission of prolactinomas is their natural history. A study including 25 women with untreated hyperprolactinemia (18 microadenomas, 7 macroadenomas) for a mean period of 11.3 years (mean initial PRL levels 225 ng/mL) showed that 7 of 22 patients with amenorrhea resumed menses spontaneously. Galactorrhea resolved completely in eight of 19 patients. Only one patient had a slight progression of sellar abnormality. At the reevaluation, mean PRL levels had decreased to 155 ng/mL.[79] Finally, there is evidence that women with hyperprolactinemia who experience the menopause have a significant chance of normalizing their PRL levels, pointing to the influence of estrogen.[80]

Among the problems associated with medical treatment of prolactinomas, drug side effects are the most prevalent. BRC is generally well tolerated at doses between 2.5 and 20 mg/day. The most frequent side effects are nausea, vomiting, orthostatic hypotension, and, to a lesser extent, nasal congestion, headache, constipation, and psychotic events, such as auditory hallucinations, delusional ideas, and mood changes. Care should be taken when prescribing the drug to psychiatric patients.[34] There have been isolated reports of leukopenia, hepatitis, headaches, and cardiac arrhythmias. Higher doses (20–140 mg daily), used for patients with Parkinson disease, may lead to pleural effusions, thickening, and parenchymal lung changes, as well as retroperitoneal fibrosis.[65]

The side effects of BRC can be minimized by starting therapy at 1.25 mg at bedtime after food intake and progressively increasing it

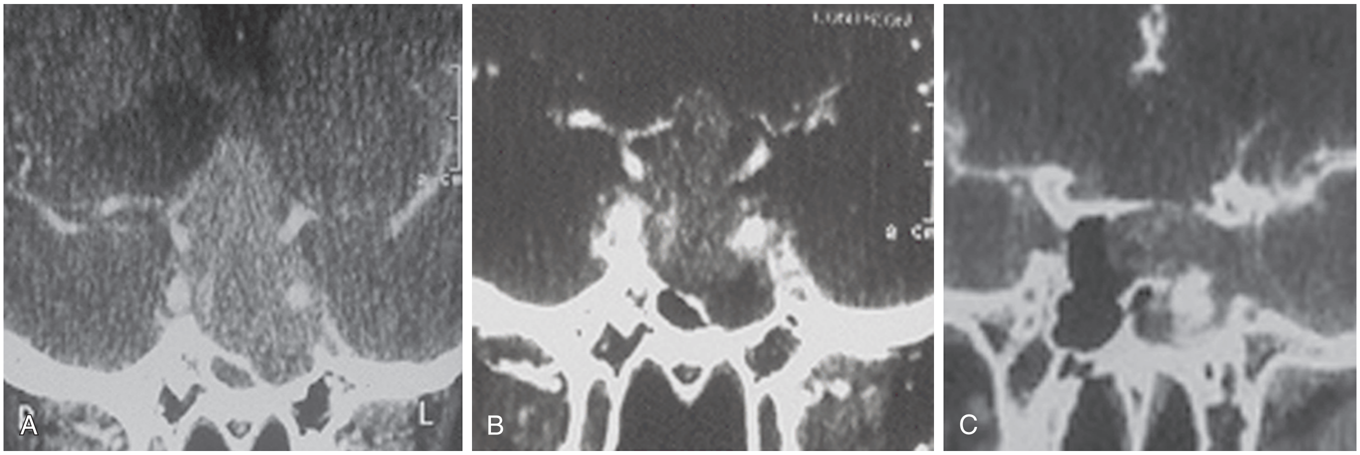


Fig. 4.10 Computed tomography (coronal views) of a man harboring a macroprolactinoma invading the sphenoid sinus **A**, before **B**, and 6 days after bromocriptine treatment, showing shrinkage mainly of the sphenoid sinus mass, **C**, with cerebrospinal fluid leakage confirmed by metrizamide injection.

to include 1.25 mg a day after breakfast and at bedtime, according to individual tolerance, until the therapeutic dose is achieved. When normoprolactinemia is obtained, the dose can be reduced in many cases.[77] A subset of patients remains intolerant, however, even when all such recommendations are followed.

An even more vexing problem in the medical treatment of prolactinomas is partial or complete lack of responsiveness to BRC. The definition of BRC resistance is arbitrary, with the failure of the drug in normalizing serum PRL levels or in reducing the tumor size using a dose equal to or greater than 15 mg/day for at least 3 months being the most currently accepted. According to this principle, about 10% of prolactinomas are resistant to BRC.[31] This criterion has limitations in clinical practice, however. Patients who are intolerant to BRC do not attain the dose established as the limit, and, although they are different conceptually, BRC intolerance and resistance overlap. Patients harboring expanding tumors with visual impairment should not be made to wait for 3 months without any improvement to confirm resistance. As previously mentioned, clinical amelioration and tumor reduction may be achieved even if serum PRL is still above the normal range.[70] Pragmatically, *BRC resistance* can be defined as the failure to obtain adequate clinical results in patients using the highest tolerable drug dose.

The mechanisms of prolactinoma BRC resistance have not yet been fully elucidated. *In vivo* and *in vitro* studies using radiolabeled dopamine antagonists as markers[81,82] or assessing D2 receptor expression by RT-PCR[83] show that dopamine D2 receptor density is reduced in resistant prolactinomas compared with tumors that are responsive to BRC. When the D2 receptor density in lactotrope cell membranes is extremely low, the DA drug may paradoxically lead to tumor growth.[82] The paucity of dopamine D2 receptors seems to be the main mechanism of BRC resistance in prolactinomas. Additionally, other events have been described such as a decreased ratio between the short and long isoforms of the dopamine D2 receptor, which are derived from alternative splicing,[84] and low filamin-A (cytoskeleton protein) expression.[85]

It is of note that some prolactinomas that are initially responsive to BRC become resistant during therapy. In such situations, especially in aggressive tumors, the diagnosis of the rare PRL-secreting carcinoma must be considered.[86]

Several other DA drugs have been employed for the treatment of hyperprolactinemia/prolactinomas. Many shows therapeutic efficacy, but two of them in particular merit detailed description, aiming mainly at overcoming intolerance and resistance to BRC.

Quinagolide. Quinagolide is a nonergot DA drug that has specific affinity for the dopamine D2 receptor. Its efficacy in normalizing serum PRL levels is similar to that of BRC, ranging from 45% to 100%, with most studies of patients with macroprolactinomas showing PRL normalization in between 58% and 75% of cases, paralleled by clinical improvement.[65] Quinagolide is administered once a day at doses ranging from 0.075 to 0.6 mg. Some studies point to better tolerance compared with BRC. Additionally, there is evidence that around 50% of patients with BRC-resistant prolactinomas attained normoprolactinemia when switched to quinagolide.[87] Such tumors were partially responsive to BRC, however, with the further decrease attributed to better patient tolerance and a higher affinity to the already reduced but still present dopamine D2 receptors. Tumor size evaluation in 105 patients showed more than 50% shrinkage and 25% to 50% reduction in tumor size in 48.1% and 20.2% of cases, respectively.[87] Quinagolide is commercially available only in Europe.

Cabergoline. Cabergoline (CAB) is a synthetic ergoline that shows high specificity and affinity for the dopamine D2 receptor. It is a potent and long-acting inhibitor of PRL secretion, with an elimination half-life ranging between 63 and 109 hours.[88] PRL-lowering effects occur rapidly within 3 hours and were evident after a single-dose administration at the end of follow up (21 days) in puerperal women and at 14 days in patients with hyperprolactinemia. This pharmacologic profile allows CAB to be administered once or twice a week in most patients, usually at a weekly dose of 0.5 to 2 mg. In our experience, normoprolactinemia was maintained in some patients taking one tablet (0.5 mg) of CAB every 10 days and even every 2 weeks. Many studies have shown the efficacy of CAB in normalizing serum PRL levels and reducing tumor size, with consequent improvement of clinical and visual manifestations (Fig. 4.11).

Many studies pointed to the better efficacy and tolerance of CAB compared with BRC.[89,90] In a multicenter Brazilian study, data collected from 388 patients (320 women and 68 men) bearing prolactinomas (220 microadenomas and 68 macroadenomas) showed that CAB was significantly more effective than BCR in inducing normoprolactinemia (81.9% vs. 67.1%), without any difference according to gender.[55] In a large, multicenter European comparative study including 459 women with hyperprolactinemic amenorrhea, 0.5 to 1 mg of CAB twice weekly was more effective than 2.5 to 5 mg of BRC twice daily in the treatment of hyperprolactinemic amenorrhea, restoring ovulatory cycles in 72% of women and normalizing plasma PRL levels in 83%, compared with 52% and 59% for BRC. Adverse effects were recorded in 68% of women taking CAB and 78% of women taking BRC, but only 3% discontinued CAB, whereas 12% stopped taking BRC because of drug intolerance.[89] Regarding CAB's effect on macroprolactinoma size, in a compilation of 12 series including 320 patients, 28.4% had a greater than 50% tumor shrinkage, 28.4% had a reduction between 25% and 50%, and 43% had less than 25% or no reduction at all.[87] Because many of these patients had been previously intolerant or resistant to other DAs, the poorer CAB results for macroprolactinoma shrinkage compared with other drugs may have been biased. This bias can be illustrated by a study showing that the rate of macroprolactinoma shrinkage greater than 80% after CAB treatment was higher in naive patients (92.3%) than in patients who were previously intolerant (42.1%), resistant (30.3%), or responsive (38.4%) to BRC or quinagolide.[91] As far as DA resistance in prolactinomas is concerned, some studies indicate that CAB normalized serum PRL levels in patients who were resistant to BRC and even to quinagolide. In two series dealing with BRC-treated prolactinomas without full PRL normalization, normoprolactinemia was achieved in 70% and 85% of patients when they were switched to CAB.[92,93] As previously pointed out (see section on quinagolide), such tumors did respond partially to BRC, however, and the additional decrease induced by CAB is probably related to its better tolerability and higher affinity to the reduced but still present dopamine D2 receptors. Consequently, severely resistant prolactinomas would not be expected to respond to any DA drug, and should be treated by different means.

Some studies emphasized the maintenance of normoprolactinemia in 23% of patients after CAB withdrawal for 12 months, on average. A long-term prospective study including 200 hyperprolactinemic patients showed that, 2 to 5 years after CAB cessation, normoprolactinemia still persisted in 76% of patients with "nontumoral" hyperprolactinemia, 70% of patients with microprolactinoma, and 65% of patients with macroprolactinoma.[94] These results suggest that CAB also compares favorably with BRC in terms of maintenance of normal PRL levels after drug withdrawal.[76]

In last decade, three metaanalyses evaluated normoprolactinemia after DA withdrawal (Table 4.5) and reported rates of 21% in 743 patients (BRC or CAB) in 2010,[95] 35% in 637 patients (only CAB) in 2015,[96] and 37% in more than 1100 patients in 2018 (BRC or CAB).[97]

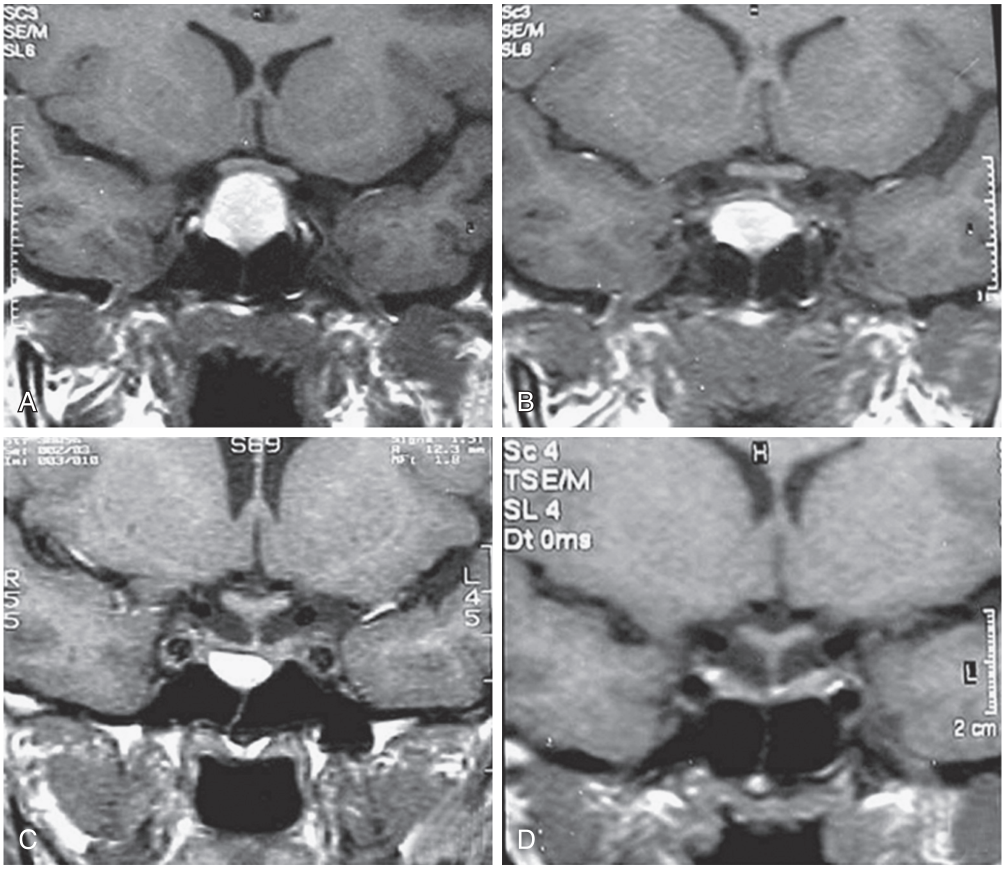


Fig. 4.11 Magnetic resonance imaging (T1-weighted with gadolinium enhancement) of a man bearing a large macroprolactinoma. **A,** Coronal view and **B,** sagittal view before treatment. **C,** Coronal view and **D,** sagittal view after 1 month of cabergoline treatment, showing significant tumor shrinkage that was paralleled by visual improvement.

TABLE 4.5 Summarized Data of Prolactinoma Remission From Metaanalyses

Author	Drug	Studies included	*n*	Remission %	Prognostic factors
Dekkers et al., 2010	BRC/CAB	19	743 353 MIC 159 MAC	21	Length of treatment, normoprolactinemia and tumor reduction
Hu et al., 2015	CAB	11	637 492 MIC 123 MAC	35	Tumoral reduction and low dose of CAB at withdrawal
Xia et al., 2019	BRC/CAB	24	1106 727 MIC 306 MAC	36.6	CAB, length of treatment, normoprolactinemia, tumor reduction and low dose of CAB at withdrawal

BRC, Bromocriptine; *CAB,* cabergoline; *MAC,* macroadenoma; *MIC,* microadenoma.

These increasing remission rates could be associated with greater experience with CAB, a longer duration of treatment, and the implementation of better criteria for DA cessation.[98] In these different reviews, prognostic factors associated with remission were DA treatment at least for 2 years, normoprolactinemia, significant tumor volume reduction,[95,97] and a low maintenance dose of CAB.[96,97] Recurrence occurred especially in the first year after treatment cessation,[31] and a second attempt to withdraw CAB after 2 additional years of treatment is associated with lower remission rates.[99,100]

One concern regarding DA treatment is related to a potential risk of cardiac valvopathy, raised by two studies pointing to a high prevalence of cardiac valve dysfunction, compared with the general population, in patients treated for Parkinson disease with high doses of CAB (up to 3 mg/day), but not other DA drugs.[101] The supposed mechanism is related to an agonist effect on the $5HT_{2B}$ receptor, leading to inappropriate mitogenic stimulation on valve and chordae tendineae. Although this dose is much higher than that used for prolactinomas, the cumulative dose derived from the potential long-term use of CAB by patients harboring prolactinomas has raised concerns about the potentially harmful effects of this drug on the cardiac valves in this disease context.

Recently, a metaanalysis was performed that included 13 case-control studies with 836 hyperprolactinemic patients who received CAB for at least 6 months and excluded single case reports and previous metaanalyses. The authors found that more tricuspid regurgitation was observed in the CAB-treated patients compared with controls (although this was not associated with symptoms), but that there was no significant increase in any other valvulopathy.[102] The risk of clinically significant valvular disease (moderate tricuspid insufficiency) was observed only in one study, with a high mean CAB cumulative dose.[103] The Endocrine Society guideline on hyperprolactinemia treatment recommends transthoracic echocardiogram surveillance only for patients on CAB doses higher than 2 mg per week.[31] The British Committee recommends performance of a transthoracic echocardiogram before initiating DA treatment, and repeating annually if the CAB dose is higher than 2 mg weekly or at 5 years if a lower dose is used.[104] Our policy is to perform a transthoracic echocardiogram before and yearly during DA treatment, irrespective of the dose used.

A variety of psychiatric side effects as psychosis, mania, anxiety, depression, and, very importantly, impulse control disorders (ICDs) have been reported.[34]

ICD is defined as a failure to resist an impulse or the temptation to perform an act that is harmful to oneself or others, and can be manifested as pathological gambling, hypersexuality, binge eating, compulsive shopping, compulsive medication use, and punding (repetitive acts). This side effect is probably related to dopamine D3 receptor drug affinity, as observed with pramipexole and ropinirole in Parkinson disease.[105]

Five studies evaluated case series of patients with ICD harboring prolactinomas, pointing to hypersexuality in individuals taking DA.[34] Current smoking, alcohol use, and gambling history were associated with ICD.[106] These factors could be apparent from the medical history, and clinicians need to be aware of these potentially devastating side effects.

Despite such potential side effects, medical therapy seems to be the first option for prolactinoma therapy, being more effective than surgery, especially for macroprolactinomas,[107] including giant prolactinomas.[56] To date, CAB seems to be the first-choice drug for prolactinoma treatment in view of its remarkable tolerance, as well as its capacity to normalize serum PRL levels, reduce tumor size, and induce high rates of persisting normoprolactinemia after drug withdrawal.

Surgical Therapy

Pituitary surgery by the transsphenoidal approach was used at the beginning of the twentieth century and was reintroduced in the early 1960s, and greatly improved when the surgical microscope was introduced.[2,108] More recently, endonasal endoscopic surgery has become available. Compared to the classic transsphenoidal approach, endoscopic pituitary surgery seems to reduce the time of hospitalization, but a clear improvement in surgical results remains to be demonstrated.[109] Besides serum PRL normalization, this surgical modality aims at reducing or eliminating the mass effect of expanding macroadenomas, often leading to the resolution of neurologic and visual manifestations. The transcranial surgical approach is reserved solely for tumors with a predominant extrasellar location expanding out of the midline, when extended endoscopic transsphenoidal surgery is not indicated or available.[110]

Surgical success in normalizing serum PRL levels depends on the experience and ability of the surgeon and on the tumor size and invasiveness. Preoperative serum PRL levels, usually associated with tumor dimension and location, were found to be paramount in predicting surgical remission, being the only predictive factor in multivariate analysis in some reports.[107] Consequently, the best results were achieved in microprolactinomas with a preoperative serum PRL of 100 ng/mL or less.

A compilation of 50 published series showed that postsurgical serum PRL normalization was achieved in 73.7% of 2137 patients with microprolactinomas and 33.9% of 2226 patients with macroprolactinomas.[107] Comparison among the series is difficult, however, as many authors do not mention preoperative serum PRL levels, the tumor size, or the degree of invasiveness. The grade of clinical improvement in cured patients is high, mainly in smaller tumors. Menses and fertility restoration rates are high, and pregnancy rates range from 75% to 90% in different series. Some patients can achieve menstrual regulation and even become pregnant without full PRL normalization.[70,111] Sexual improvement in men is less likely to occur, probably owing to irreversible gonadotropic damage caused by the higher frequency of extrasellar and invasive prolactinomas compared with women. Nonetheless, pituitary function generally is preserved in intrasellar adenomas and is restored by the removal of suprasellar adenomas causing hypothalamopituitary disconnection.[111] In addition, neurologic and visual amelioration is frequently achieved after pituitary surgery.

Patient preference for surgical therapy should be considered in well-circumscribed micro and intrasellar macroprolactinomas, in a experienced and well-resourced pituitary center.[112] Also, in young women harboring microprolactinomas, surgery in experienced hands can be cost effective in tumor remission compared to long-term DA treatment.[113]

On the other hand, surgical debulking has been shown to improve further medical therapy in patients with prolactinomas that are partially resistant to DA.[114,115]

An important caveat of surgical treatment is prolactinoma recurrence. A literature compilation reported recurrence rates of 18.2% of 809 microprolactinomas and 22.8% of 465 macroprolactinomas.[107] In our series, the recurrence of hyperprolactinemia in surgically-cured patients was 27% for microprolactinomas and 17% for macroprolactinomas.[70] Median time to recurrence varied among the different surgical series, from 1 to 7 years.

Although the occurrence of relapses contributes to the decrease of long-term surgical cure rate of prolactinomas, many patients with recurrent hyperprolactinemia remain asymptomatic and have no evidence of tumor regrowth.[116] Relapse can be transient. From a cohort of 44 patients with surgically cured microprolactinomas, eight (18.2%) experienced recurrence and were followed up. Only two of the eight patients experienced permanent relapse.[117] Additional therapy is reserved only for patients with symptomatic recurrence.

Surgical treatment for prolactinomas is indicated for cases with persistent intolerance or hormonal or tumor resistance to DAs, pituitary apoplexy,[118] tumor growth during medical therapy,[71] CSF leakage attributed to DA-induced tumor shrinkage of invasive macroprolactinomas,[73,74] and (rarely) visual loss on medical therapy, secondary to optic chiasm herniation resulting from tumor retraction.[72] Additionally, surgery is an excellent alternative for patients harboring microprolactinomas, especially patients with serum PRL of 100 ng/mLor less, with poor compliance with medical therapy.[119] Table 4.6 summarizes the indications for in prolactinomas.

TABLE 4.6 Surgical Treatment Indications in Prolactinomas

Intolerance/resistance to dopamine agonists
Apoplexy
Cerebrospinal fluid leakage owing to dopamine agonist-induced tumor shrinkage
Visual loss on medical therapy, secondary to optic chiasm herniation resulting from tumor retraction
Microprolactinoma and patient preference

Radiotherapy

The efficacy of conventional radiation therapy for prolactinomas is lower than that of medical or surgical treatment. Our results for 19 patients who underwent radiotherapy after medical/surgical failure showed that only three (16%) achieved serum PRL normalization 5, 6, and 15 years after the procedure.[70] A compilation from the literature shows serum PRL normalization in less than one third of patients 5 to 15 years after conventional radiotherapy, with hypopituitarism occurring in 5.5% to 93.3% of cases.[107] The prevalence is probably greater in reality, because many series underestimate GH deficiency in adults. Other complications associated with conventional radiotherapy include optic nerve lesions, neurologic dysfunctions, cerebrovascular disease, and the development of other intracranial tumors.

As conventional external beam radiotherapy (EBRT) can result in these significant complications, it has largely been replaced by stereotactic radiosurgery (SRS), mainly single-shot gamma knife "radiosurgery" (GKRS). According to the Clinical Practice Guidelines published by the European Society of Endocrinology,[5] GKRS is indicated for tumors with a maximal diameter less than 3 cm and a distance from the optic chiasm of at least 3 to 5 mm. Otherwise, fractionated EBRT is the better option. In addition, EBRT is preferred for tumors with irregular anatomy, including suprasellar or brainstem extension and diffuse local infiltration.[120]

In a compilation of 17 studies including 416 patients, the success rate of GKRS was reported to be 4.5% to 83%.[121] Different biochemical control criteria and different follow-up intervals probably explain this wide variation in the literature regarding remission and complication rates.

A recent multicenter study evaluated the role of SRS in prolactinomas, including 289 patients treated between 1997 and 2016 at 10 institutions: endocrine remission (normal PRL without DA) occurred in 28%, 41%, and 54% of patients at 3, 5, and 8 years, respectively, whereas endocrine control (PRL level ≤30 ng/mL on DA) was the more common endpoint, obtained in 63% of cases. Regarding complications, 25% presented a new hormone deficiency, 3% a new visual complication, and 5% tumor progression. A pretreatment PRL level of 270 ng/mL or less was a predictor of endocrine remission.[122] It is important to note that tumor growth after radiotherapy is a rare event,[123] and its occurrence in this study could be the result of selection bias because only patients with medically refractory or residual/recurrent tumors that were not amenable to resection were included.

Whether or not to maintain DA during radiotherapy is a controversial point. In three studies, the remission rate was low in patients on DA, whereas in another three studies, medical treatment did not influence the results.[121]

More data and increased follow-up time are required to assess the superiority of GKRS to conventional radiotherapy. In the authors' opinion, at present radiation therapy for prolactinomas should be reserved for cases with medical and surgical failures, especially for invasive tumors.

Therapeutic Advances and Perspectives

As a result of the efficacy of medical and surgical therapies, prolactinomas can be adequately controlled in most patients. A subset of patients who are resistant to DAs, especially patients harboring invasive macroadenomas who are not expected to be surgically cured, pose an important therapeutic problem. Radiotherapy use is generally limited, is efficacious only in the long term, and usually leads to hypopituitarism.

Many studies did not find tumor enlargement in patients with "idiopathic" hyperprolactinemia or microprolactinomas, before or after the menopause.[124,125] Sexual steroid replacement can be considered for patients for whom pregnancy is not a concern, and galactorrhea is not disturbing; these patients must then be followed carefully. Additionally, testosterone may be given to men harboring resistant microprolactinomas and even partially resistant macroprolactinomas, with PRL level monitoring and imaging. The introduction of an aromatase-inhibitor drug may be considered if there is evidence of tumor growth or a serum PRL increase.[126]

The use of clomiphene citrate as an alternative to testosterone therapy for men with hypogonadism has been described for patients both with and without normal PRL levels. The main advantage is fertility maintenance or restoration.[127]

Prolactinomas usually do not respond to somatostatin.[128] It has been shown, however, that a selective analog of the somatostatin receptor subtype 5 (SSTR5) inhibits PRL secretion in a human prolactinoma cell culture.[129] This *in vitro* finding was confirmed *in vivo*: two cases of prolactinoma resistant to CAB treated with Pasireotide LAR, a multiligand somatostatin analog with high affinity for SSTR5, exhibited control of hormonal secretion,[130,131] and tumor reduction was also achieved in one of these cases.[131]

The development of PRL receptor antagonists could also represent an option for medical treatment of resistant macroprolactinomas without mass effects, because they should be able to block the actions of PRL, if not its production.[132] One therapeutic approach would be the association of DAs and PRL receptor-antagonist drugs in cases of macroprolactinomas exhibiting tumor shrinkage but not PRL normalization when on DA monotherapy.

The alkylating agent temozolomide (TMZ) has been experimentally used in many tumors (see also Chapter 15). The drug effect can be reversed by the enzyme methylguanine-DNA methyltransferase (MGMT). TMZ can be given orally and is currently approved for refractory anaplastic astrocytoma and glioblastoma multiforme. Since 2006, it has been used in aggressive/resistant pituitary tumors, mostly in a treatment schedule involving TMZ 150 to 200 mg/m^2/day for 5 days every 28 days. Monitoring of hematological parameters, liver function tests, and careful clinical observation for potential adverse effects during treatment are necessary during chemotherapy. The 2016 European Survey included 25 aggressive prolactinomas and 15 PRL-secreting pituitary carcinomas: nine showed progression, 10 stability, and 19 complete/partial response with TMZ.[133] The European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumors and carcinomas recommend TMZ monotherapy as first-line chemotherapy following documented tumor growth, or in association with radiotherapy (Stupp protocol). Although low MGMT immunohistochemistry expression is associated with a positive response to TMZ, a drug trial may be considered even in cases with high MGMT.[5]

Tyrosine kinase and mTOR inhibitors are potential drugs for resistant/aggressive prolactinomas. Lapatinib, a dual tyrosine kinase inhibitor of both EGF receptor type 1 (EGFR/ErbB1) and 2 (HER2) was able to suppress serum PRL levels and to cause tumor shrinkage in rats with prolactinomas and in two patients harboring aggressive and resistant prolactinomas.[134]

In vitro studies pointed to the effectiveness of antiestrogen treatment in inhibiting PRL secretion and cellular proliferation.[135]

Fig. 4.12 depicts a proposed flowchart for the management of resistant prolactinomas.

Prolactinomas and Pregnancy

The development of efficacious medical and surgical therapies for prolactinomas has made pregnancy possible for women bearing such tumors. Gestational risks owing to the possibility of tumor growth during pregnancy, mainly in women with macroadenomas, raised a

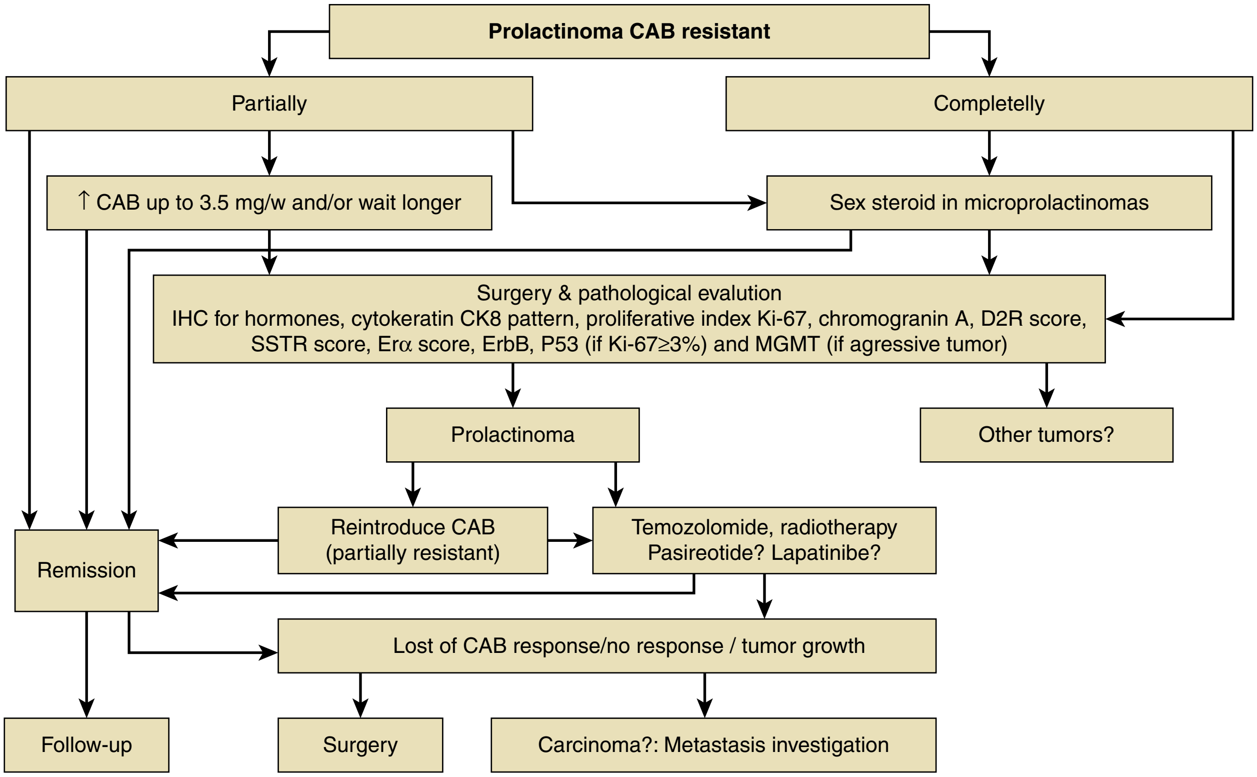

Prolactinoma CAB resistant
Partially
Completelly
↑ CAB up to 3.5 mg/w and/or wait longer
Sex steroid in microprolactinomas
Surgery & pathological evalution
IHC for hormones, cytokeratin CK8 pattern, proliferative index Ki-67, chromogranin A, D2R score, SSTR score, Erα score, ErbB, P53 (if Ki-67≥3%) and MGMT (if agressive tumor)
Prolactinoma
Other tumors?
Reintroduce CAB (partially resistant)
Temozolomide, radiotherapy Pasireotide? Lapatinibe?
Remission
Lost of CAB response/no response / tumor growth
Follow-up
Surgery
Carcinoma?: Metastasis investigation

Fig. 4.12 Flowchart depicts prolactinoma-resistant management. *CAB,* Cabergoline; *D2R,* dopamine receptor type 2; *Erα*, estrogen receptor α; *ErbB*, ErbB receptor family, also known as the EGF receptor family or type I receptor family, includes the epidermal growth factor (EGF) receptor (EGFR) or ErbB1/Her1, ErbB2/Her2, ErbB3/Her3, and ErbB4/Her4; *IHC,* immunohistochemistry; *MGMT,* O(6)-Methylguanine-DNA methyltransferase; *SSTR5*, somatostatine receptor type 5.

concern, however. The management of pregnancy in patients with prolactinomas who have undergone different therapies is brought into focus here.

To date, the number of pregnancies in women treated with BRC is greater those in women treated with CAB, and therefore BRC is still the only ”authorized” DA for this purpose.[31] In a compilation study of pregnancies in 246 women with microprolactinomas treated with BRC alone, tumor growth symptoms were reported in 1.6% of patients, although an asymptomatic increase in tumor size was shown in 4.5% of the cases. None of the patients needed surgical intervention during pregnancy.[136] We have followed 71 term pregnancies, and the results were similar.[137]

The risk of complications during pregnancy is much greater with macroprolactinomas. A review from the literature pointed to symptoms related to tumor growth in 15.5% of 45 patients with macroprolactinomas treated with BRC alone. The rate of complications was only 4.3% in the 46 patients who underwent surgery or radiotherapy before pregnancy. Additionally, asymptomatic tumor growth was seen in 8.9% of patients without prior surgery or radiotherapy.[136] We followed 51 term pregnancies in patients with macroprolactinomas.[137,138] Of these patients, 21 had had previous surgery, and none presented with symptoms or signs of tumor growth. Of the 30 patients treated with pregestational BRC only, 11 (37%) manifested complaints related to tumor growth: all of them presented with headaches, and seven had visual alterations (Fig. 4.13).

The safety of BRC reintroduction or even maintenance during pregnancy is supported by substantial experience with this DA, as reported in the literature. A large review[139] included 2587 pregnancies and did not show an increase in maternal or fetal morbidity or mortality. Since then, some authors have favored the maintenance of BRC in pregnancy to prevent complications in patients with macroprolactinomas without previous surgery or radiotherapy.[140] The number of pregnancies following treatment with CAB, which has become the DA of choice for prolactinoma treatment, has substantially increased. A recent review did not show any differences in pregnancy outcomes such as abortion, preterm delivery, or fetal malformation in women taking BRC ($n = 6272$) or CAB ($n = 1061$), as compared to the normal population.[141] Recently, we reported that the maternal-fetal outcomes of 233 pregnancies in women being treated with CAB were similar to previous data reported in the literature. Nevertheless, we found a higher miscarriage rate in the subgroup of patients with maintenance of CAB after pregnancy confirmation.[142] The causality of this finding is not established, and more data are needed. The use of CAB for permitting pregnancy is still not approved by the US Food and Drug Administration or and the European Medicines Agency; however, it can be an alternative for BRC-intolerant/resistant cases,[143,144,145] and it should be withdrawn in cases with tumor within sellar boundaries. Of 52 pregnancies in which CAB was maintained, there were three miscarriages, three abortions (one with neural tube malformation), five cases of prematurity, one case of femoral hip dysplasia, and one fetal death because of preeclampsia.[142,146,147,148,149]

In data from 16 series, symptomatic prolactinoma enlargement during pregnancy occurred in 2.5% of cases of microadenomas, 18.1% of cases of macroadenomas without previous surgery or radiotherapy, and 4.7% of cases of macroadenomas with these previous treatments.[141] These data show the higher risk of tumor growth in macroprolactinomas during pregnancy, necessitating stricter follow-up (Fig. 4.14). The first recommendation should be the use of a nonhormonal contraceptive along with a DA until tumor shrinkage to within sellar boundaries has been shown. The duration of previous treatment with DA might be

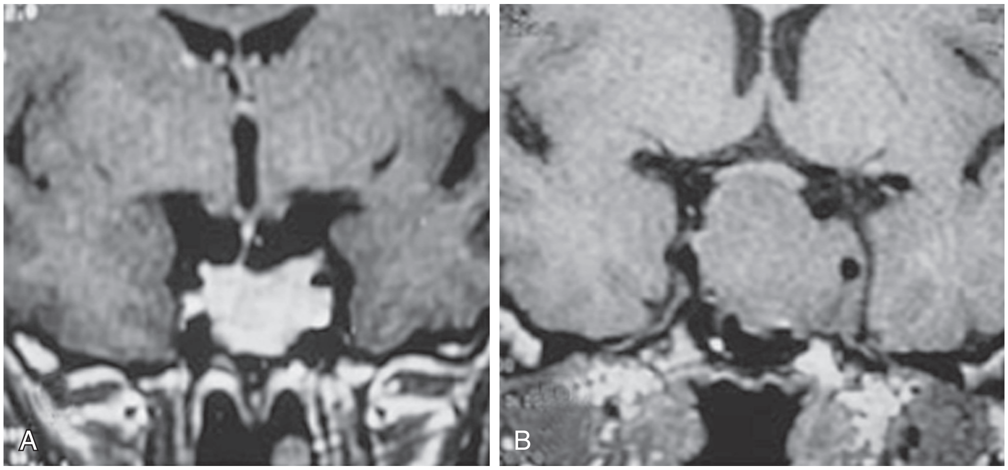


Fig. 4.13 Magnetic resonance imaging (coronal views) of a patient with macroprolactinoma. **A,** Tumor is limited to sellar boundaries during bromocriptine treatment before pregnancy. **B,** Tumor growth during the fourth month of pregnancy without bromocriptine use. (From Musolino NRC, Bronstein MD. Prolactinomas and pregnancy. In Bronstein MD (ed.). Pituitary tumors in pregnancy. Boston: Kluwer Academic Publishers; 2001:91–108.)

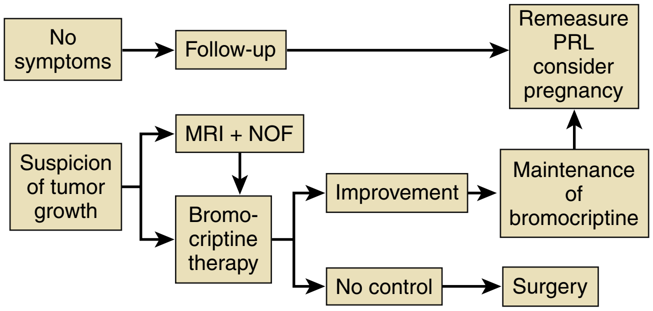

No symptoms
Follow-up
Remeasure PRL consider pregnancy
Suspicion of tumor growth
MRI + NOF
Bromo-criptine therapy
Improvement
Maintenance of bromocriptine
No control
Surgery

Fig. 4.14 Suggested algorithm for the follow-up of patients with macroprolactinoma during pregnancy. *MRI,* Magnetic resonance imaging; *NOF,* neuroophthalmologic examination; *PRL,* prolactin. (From Musolino NRC, Bronstein MD. Prolactinomas and pregnancy. In Bronstein MD (ed.). Pituitary tumors in pregnancy. Boston: Kluwer Academic Publishers; 2001:91–108.)

important.[142,150] After pregnancy has been confirmed, the DA can be withdrawn, and the patient must be closely monitored for symptoms related to tumor growth. If there is suspected tumor expansion, confirmation can be made through noncontrast MRI after the fourth month of gestation, and by visual field testing (see Fig. 4.14). Monitoring serum PRL levels during pregnancy does not seem to be useful, because they are not always related to tumor behavior during gestation. The reintroduction of DA in such cases often leads to clinical amelioration and tumor reduction. In nine of 11 patients who exhibited complications during pregnancy, BRC reintroduction led to complete resolution of the symptoms related to tumor growth.[137] Surgery also can be employed as treatment for symptomatic tumor growth in pregnancy. Several authors have reported good results, although the increased risk of spontaneous abortion in patients who undergo surgery is well known.[137] Regarding DA maintenance throughout gestation, it is our policy to use such an approach only when the patient becomes pregnant after a short treatment period, mainly without confirmation of tumor shrinkage or when the tumor is outside sellar boundaries.

Follow-Up After Delivery. Breastfeeding does not increase the risk of tumor growth in patients who progressed well during pregnancy.[137,144] However, it is contraindicated in patients maintained on DA after delivery, owing to tumor-growth signs.

There have been several reports in the literature regarding reduction or normalization of serum PRL levels after delivery.[137] Remission is reported in 10% to 68% of cases.[137,144,150,151,152] In our multicenter study, normal levels of PRL without DA occurred in 10% of cases. When all patients were considered, 42% presented reduced serum PRL levels, and 46% presented tumor reduction, compared to before pregnancy.[142] The explanation for this "curative" effect of pregnancy remains to be clarified. It may be related partly to modifications in the vasculature of the adenoma attributed to the estrogen stimulation, resulting in necrosis or microinfarctions of the adenomatous tissue. Hemorrhagic zones in prolactinoma patients receiving estrogen therapy have already been described.[153]

Another issue of concern is the outcomes of children whose mothers took DA drugs during pregnancy. One study reported 4-month to 9-year follow-up of 546 children exposed to BRC *in utero*.[139] The authors did not find any developmental impairment in the children. We followed 70 children born to mothers who conceived while taking BRC.[137] At a mean follow-up time of 67 months (range: 12–240 months), only two children presented with disorders of neuropsychomotor development, including one case of idiopathic hydrocephaly and one case of tuberous sclerosis. We did not find any similar reports in the literature. Fifteen of these children already had started puberty, one of them precociously. Regarding the risk of neurodevelopmental impairment attributed to fetal exposure to CAB, no abnormalities were found in two series of five[137] and 99[145] children, respectively. In the other studies, a slight retardation in verbal fluency in two boys and difficulty achieving complete continence in one child were detected among 88 children,[151] seizures in two children and pervasive developmental disorder in two additional children were detected amongst 61 children,[154] and four cases of epilepsy and four cases of delayed neurocognitive development were detected among 145 children.[142]

Treatment Planning and Follow-Up

Based on the previously described data, the current gold-standard therapy for either microprolactinomas or macroprolactinomas is medical treatment with DA drugs. Physicians must motivate patients to embark on such long-term treatment based on the overall better results compared with surgery and, mainly based on more recent data,[76,94] with maintenance of normoprolactinemia after drug withdrawal in a substantial number of patients. Surgical treatment of prolactinomas is indicated for patients with persistent intolerance or hormonal or tumor resistance to more than one DA—particularly if pregnancy is desired—or if the tumor has grown during medical treatment. In cases of resistance, rare conditions such as acidophil stem cell adenoma or PRL-secreting carcinoma should be considered. Malignant prolactinomas respond only temporarily, if they respond at all, when patients are switched to another DA drug or undergo surgery. Surgical therapy is also indicated in selected cases of pituitary apoplexy, in the event that DA–induced CSF leakage occurs in invasive macroprolactinomas, and in the exceedingly rare occurrence of visual loss secondary to optic chiasm herniation during medical therapy. Additionally, surgery in skilled hands may be considered for patients who are not willing to undergo long-term medical therapy, especially patients harboring microprolactinomas and serum PRL levels less than 100 ng/mL.

Radiotherapy, either conventional or stereotactic, is reserved for prolactinomas that not responsive to medical or surgical treatment, particularly invasive ones.

There is evidence that, in general, estrogen replacement is not harmful for women harboring microprolactinomas. This approach may be used when fertility is not an issue for patients with intolerance or resistance to medical therapy and who are not cured by or not willing to undergo pituitary surgery. In addition, women who are successfully medically treated may use hormonal contraceptives if they have not adapted to barrier methods. Menopausal women bearing microprolactinomas can interrupt DA drug use and are allowed to start hormone replacement therapy if is indicated.[125] There is evidence from animal models for a role for PRL in carcinogenesis, mainly carcinoma of the mammary gland, although human data are still highly controversial.[155] It is now apparent that human mammary epithelial cells can synthesize PRL endogenously, permitting autocrine/paracrine actions within the mammary gland that are independent of pituitary PRL and are probably not responsive to DA drugs. At present, the maintenance of high serum PRL levels in premenopausal and postmenopausal women is not a concern with regard to carcinogenesis. Finally, for macroprolactinomas that are not adequately controlled, use of estrogens is generally discouraged, given their potential for growth induction.

There are many therapeutic perspectives for resistant/aggressive prolactinomas: specific somatostatin or somatostatin/dopamine analogs, selective antiestrogen drugs, cytotoxic drugs such as TMZ, and novel PRL-receptor antagonists. Nevertheless, more studies are needed to assess their places in the treatment algorithm.

The follow-up planning for patients with prolactinomas depends on the tumor size and clinical, laboratory, and imaging response to therapy. Many patients are clinically controlled even if serum PRL levels are still above the normal range, and do not need further drug dose increases. For microprolactinomas, and especially for macroprolactinomas, significant tumor shrinkage or "disappearance" on MRI is a good prognostic marker, and in such cases, imaging reassessment can be performed sporadically. Medically or surgically controlled patients must be reassessed periodically, both clinically and hormonally, to identify patients who may discontinue the DA drug with maintenance of normoprolactinemia or patients with recurrence of hyperprolactinemia.

REFERENCES

1. Vroonen L, Daly AF, Beckers A. Epidemiology and management challenges in prolactinomas. *Neuroendocrinology*. 2019;109:20–27.
2. Hardy J. Transsphenoidal microsurgery of the normal and pathological pituitary. *Clin Neurosurg*. 1969:185–217.
3. Yoo F, Kuan EC, Heaney AP, et al. Corticotrophic pituitary carcinoma with cervical metastases: case series and literature review. *Pituitary*. 2018;21:290–301.
4. Glezer A, D'Alva CB, Salgado LR, et al. Pitfalls in pituitary diagnosis: peculiarities of three cases. *Clin Endocrinol (Oxf)*. 2002;57:135–139.
5. Raverot G, Burman P, McCormack A, et al. European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumours and carcinomas. *Eur J Endocrinol*. 2018;178:G1–G24.
6. Striker P, Grueter F. Action du lobe antérieur de l'hypophyse sur la montée laiteuse. *C R Soc Biol Paris*. 1929;99:1978–1980.
7. Riddle O, Bartes R, Dykshorn D. The preparation, identification and assay of prolactin—a hormone of the anterior pituitary. *Am J Physiol*. 1933;105:191–216.
8. Hwang P, Guyda H, Friesen H. A radioimmunoassay for human prolactin. *PNAS*. 1971;68:1902–1906.
9. Owerbach D, Rutter WJ, Cooke NE, et al. The prolactin gene is located on chromosome 6 in humans. *Science*. 1981;212:815–816.
10. Boutin JM, Jolicoeur C, Okamura H, et al. Cloning and expression of the rat prolactin receptor, a member of the growth hormone/prolactin receptor gene family. *Cell*. 1988;53:69–77.
11. Sonigo C, Bouilly J, Carré N, et al. Hyperprolactinemia-induced ovarian acyclicity is reversed by kisspeptin administration. *J Clin Invest*. 2012;122:3791–3795.
12. Millar RP, et al. Hypothalamic-pituitary-ovarian axis reactivation by kisspeptin-10 in hyperprolactinemic women with chronic amenorrhea. *J Endocr Soc*. 2017;1:1362–1371.
13. Ciccarelli A, Daly AF, Beckers A. The epidemiology of prolactinomas. *Pituitary*. 2005;8:3–6.
14. Colao A, et al. Gender differences in the prevalence, clinical features and response to cabergoline in hyperprolactinemia. *Eur J Endocrinol*. 2003;148:325–331.
15. Shimon I, Bronstein MD, Shapiro J, et al. Women with prolactinomas presented at the postmenopausal period. *Endocrine*. 2014;47:889–894.
16. Berezin M, Karasik A. Familial prolactinoma. *Clin Endocrinol (Oxf)*. 1995;42:483–486.
17. Daly AF, Beckers A. Familial isolated pituitary adenomas (FIPA) and mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene. *Endocrinol Metab Clin North Am*. 2015;44:19–25.
18. Daly AF, Castermans E, Oudijk L, et al. Pheochromocytomas and pituitary adenomas in three patients with. *Endocr Relat Cancer*. 2018;25:L37–L42.
19. Bronstein MD, Melmed S. [Pituitary tumorigenesis]. *Arq Bras Endocrinol Metabol*. 2005;49:615–625.
20. Cruz-Soto ME, Scheiber MD, Gregerson KA, et al. Pituitary tumorigenesis in prolactin gene-disrupted mice. *Endocrinology*. 2002;143:4429–4436.
21. Friedman E, Adams EF, Höög A, et al. Normal structural dopamine type 2 receptor gene in prolactin-secreting and other pituitary tumors. *J Clin Endocrinol Metab*. 1994;78:568–574.
22. Shah SS, Aghi MK. The role of single-nucleotide polymorphisms in pituitary adenomas tumorigenesis. *Cancers (Basel)*. 2019;11:1977.
23. Bueno C, et al. Cabergoline and prolactinomas: lack of association between DRD2 polymorphisms and response to treatment. *Pituitary*. 2017;20:295–300.
24. Yu R, Bonert V, Saporta I, et al. Aryl hydrocarbon receptor interacting protein variants in sporadic pituitary adenomas. *J Clin Endocrinol Metab*. 2006;91:5126–5129.
25. Hardy J. Transsphenoidal surgery of hypersecreting pituitary tumors. In: Kolhler GRG, ed. *Diagnosis and Treatment of Pituitary Tumors*. New York: Elsevier; 1973:179–194.
26. Lloyd RV, Osamura RY, Klöppel G, et al., eds. *WHO classification of tumours of endocrine organs*. 4th edition. Lyon (France): IARC Press; 2017.
27. Lopes MBS. Pathology of prolactinomas: any predictive value? *Pituitary*. 2020;23:3–8.
28. Ozbey N, Sariyildiz E, Yilmaz L, et al. Primary hypothyroidism with hyperprolactinaemia and pituitary enlargement mimicking a pituitary macroadenoma. *Int J Clin Pract*. 1997;51:409–411.
29. Vilar L, Fleseriu M, Bronstein MD. Challenges and pitfalls in the diagnosis of hyperprolactinemia. *Arq Bras Endocrinol Metabol*. 2014;58:9–22.
30. Hu Y, et al. Serum prolactin levels across pregnancy and the establishment of reference intervals. *Clin Chem Lab Med*. 2018;56:803–807.
31. Melmed S, et al. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:273–288.
32. Molitch ME. Drugs and prolactin. *Pituitary*. 2008;11:209–218.

33. Madhusoodanan S, Parida S, Jimenez C. Hyperprolactinemia associated with psychotropics—a review. *Hum Psychopharmacol*. 2010;25:281–297.
34. Ioachimescu AG, Fleseriu M, Hoffman AR, et al. Psychological effects of dopamine agonist treatment in patients with hyperprolactinemia and prolactin-secreting adenomas. *Eur J Endocrinol*. 2019;180:31–40.
35. Manuylova E, Calvi LM, Hastings C, et al. Late presentation of acromegaly in medically controlled prolactinoma patients. *Endocrinol Diabetes Metab Case Rep*. 2016. [Epub ahead of print].
36. Karavitaki N, Thanabalasingham G, Shore HC, et al. Do the limits of serum prolactin in disconnection hyperprolactinaemia need re-definition? A study of 226 patients with histologically verified non-functioning pituitary macroadenoma. *Clin Endocrinol (Oxf)*. 2006;65:524–529.
37. Glezer A, Paraiba DB, Bronstein MD. Rare sellar lesions. *Endocrinol Metab Clin North Am*. 2008;37:195–211.
38. Greenman Y. Dopaminergic treatment of nonfunctioning pituitary adenomas. *Nat Clin Pract Endocrinol Metab*. 2007;3:554–555.
39. Shuman W, Loewenstern J, Pai A, et al. Variability in clinical presentation and pathologic implications of ectopic pituitary tumors: critical review of literature. *World Neurosurg*. 2019;122:397–403.
40. Glezer A, Bronstein MD. Approach to the patient with persistent hyperprolactinemia and negative sellar imaging. *J Clin Endocrinol Metab*. 2012;97:2211–2216.
41. De Bellis A, Colao A, Pivonello R, et al. Antipituitary antibodies in idiopathic hyperprolactinemic patients. *Ann N Y Acad Sci*. 2007;1107:129–135.
42. Newey PJ, Gorvin CM, Cleland SJ, et al. Mutant prolactin receptor and familial hyperprolactinemia. *N Engl J Med*. 2013;369:2012–2020.
43. Mah PM, Webster J. Hyperprolactinemia: etiology, diagnosis, and management. *Semin Reprod Med*. 2002;20:365–374.
44. Bronstein M, Musolino N, Cardim CEA. Treatment of macroprolactinomas with a long-acting, parenteral and repeatable new form of bromocriptine. In: Landolt AM, PUH ZJ, et al., eds. *Advances in Pituitary Adenoma Research*. Oxford: Pergamon Press; 1988.
45. Serri O, Chik CL, Ur E, et al. Diagnosis and management of hyperprolactinemia. *CMAJ*. 2003;169:575–581.
46. Kars M, van der Klaauw AA, Onstein CS, et al. Quality of life is decreased in female patients treated for microprolactinoma. *Eur J Endocrinol*. 2007;157:133–139.
47. Di Filippo L, Doga M, Resmini E, et al. Hyperprolactinemia and bone. *Pituitary*. 2020;23:314–321.
48. Bronstein M, Marino RJ, Pereira D. Therapeutic alternatives for hyperprolactinemia: the role of bromo-ergocriptine. *Rev Bras Ginecol Obstet*. 1983;5:193.
49. Kleinberg DL, Noel GL, Frantz AG. Galactorrhea: a study of 235 cases, including 48 with pituitary tumors. *N Engl J Med*. 1977;296:589–600.
50. Duskin-Bitan H, Shimon I. Prolactinomas in males: any differences? *Pituitary*. 2020;23:52–57.
51. Drago F, Scapagnini U. Side effects of drugs stimulating prolactin secretion on the behavior of male rats. *Arch Int Pharmacodyn Ther*. 1985;276:271–278.
52. Molitch ME, Thorner MO, Wilson C. Management of prolactinomas. *J Clin Endocrinol Metab*. 1997;82:996–1000.
53. Cannavò S, Venturino M, Curtò L, et al. Clinical presentation and outcome of pituitary adenomas in teenagers. *Clin Endocrinol (Oxf)*. 2003;58:519–527.
54. Jallad R, Goic M, Musolino N, et al. Prolactinomas in children and adolescents: retrospective study on 47 patients. *Arq Bras Endocrinol Metabol*. 2000;44:300.
55. Vilar L, Freitas MC, Naves LA, et al. Diagnosis and management of hyperprolactinemia: results of a Brazilian multicenter study with 1234 patients. *J Endocrinol Invest*. 2008;31:436–444.
56. Shimon I, Sosa E, Mendoza V, et al. Giant prolactinomas larger than 60 mm in size: a cohort of massive and aggressive prolactin-secreting pituitary adenomas. *Pituitary*. 2016;19:429–436.
57. Andersen M, Hagen C, Frystyk J, et al. Development of acromegaly in patients with prolactinomas. *Eur J Endocrinol*. 2003;149:17–22.
58. Karavitaki N, Dobrescu R, Byrne JV, et al. Does hypopituitarism recover when macroprolactinomas are treated with cabergoline? *Clin Endocrinol (Oxf)*. 2013;79:217–223.
59. Frieze TW, Mong DP, Koops MK. "Hook effect" in prolactinomas: case report and review of literature. *Endocr Pract*. 2002;8:296–303.
60. Sinha YN. Structural variants of prolactin: occurrence and physiological significance. *Endocr Rev*. 1995;16:354–369.
61. Vallette-Kasic S, Morange-Ramos I, Selim A, et al. Macroprolactinemia revisited: a study on 106 patients. *J Clin Endocrinol Metab*. 2002;87:581–588.
62. Glezer A, Soares CR, Vieira JG, et al. Human macroprolactin displays low biological activity via its homologous receptor in a new sensitive bioassay. *J Clin Endocrinol Metab*. 2006;91:1048–1055.
63. Bronstein MD. Editorial: Is macroprolactinemia just a diagnostic pitfall? *Endocrine*. 2012;41:169–170.
64. Rennert J, Doerfler A. Imaging of sellar and parasellar lesions. *Clin Neurol Neurosurg*. 2007;109:111–124.
65. Auriemma RS, Pirchio R, De Alcubierre D, et al. Dopamine agonists: from the 1970s to today. *Neuroendocrinology*. 2019;109:34–41.
66. Vance ML, Evans WS, Thorner MO. Drugs five years later. *Bromocriptine. Ann Intern Med*. 1984;100:78–91.
67. Bronstein M. Prolactinoma in men. *Arq Bras Endocrinol Metab*. 1999;43:338.
68. Molitch ME, et al. Bromocriptine as primary therapy for prolactin-secreting macroadenomas: results of a prospective multicenter study. *J Clin Endocrinol Metab*. 1985;60:698–705.
69. Wu ZB, Su ZP, Zhuge QC, et al. Bromocriptine treatment of invasive giant prolactinomas involving the cavernous sinus: results of a long-term follow up. *J Neurosurg*. 2006;104:54–61.
70. Bronstein M, Musolino N, Cunha-Neto MEA, et al. Hyperprolactinemia therapy: lessons learned from long-term follow-up. *Eur J Endocrinol*. 1994;130:116.
71. Adler I, Barsi P, Czirják S, et al. Rapid re-enlargement of a macroprolactinoma after initial shrinkage in a young woman treated with bromocriptine. *Gynecol Endocrinol*. 2005;20:317–321.
72. Alvarez Berastegui GR, Raza SM, Anand VK, et al. Endonasal endoscopic transsphenoidal chiasmapexy using a clival cranial base cranioplasty for visual loss from massive empty sella following macroprolactinoma treatment with bromocriptine: case report. *J Neurosurg*. 2016;124:1025–1031.
73. Bronstein MD, Musolino NR, Benabou S, et al. Cerebrospinal fluid rhinorrhea occurring in long-term bromocriptine treatment for macroprolactinomas. *Surg Neurol*. 1989;32:346–349.
74. Lam G, Mehta V, Zada G. Spontaneous and medically induced cerebrospinal fluid leakage in the setting of pituitary adenomas: review of the literature. *Neurosurg Focus*. 2012;32:E2.
75. Sibal L, Ugwu P, Kendall-Taylor P, et al. Medical therapy of macroprolactinomas in males: I. Prevalence of hypopituitarism at diagnosis. II. Proportion of cases exhibiting recovery of pituitary function. *Pituitary*. 2002;5:243–246.
76. Passos VQ, Souza JJ, Musolino NR, et al. Long-term follow-up of prolactinomas: normoprolactinemia after bromocriptine withdrawal. *J Clin Endocrinol Metab*. 2002;87:3578–3582.
77. Biswas M, Smith J, Jadon D, et al. Long-term remission following withdrawal of dopamine agonist therapy in subjects with microprolactinomas. *Clin Endocrinol (Oxf)*. 2005;63:26–31.
78. Gen M, Uozumi T, Ohta M, et al. Necrotic changes in prolactinomas after long term administration of bromocriptine. *J Clin Endocrinol Metab*. 1984;59:463–470.
79. Koppelman MC, Jaffe MJ, Rieth KG, et al. Hyperprolactinemia, amenorrhea, and galactorrhea. A retrospective assessment of twenty-five cases. *Ann Intern Med*. 1984;100:115–121.
80. Greenman Y. Prolactinomas and menopause: any changes in management? *Pituitary*. 2020;23:58–64.
81. Scillitani A, Dicembrino F, Di Fazio P, et al. In vivo visualization of pituitary dopaminergic receptors by iodine-123 methoxybenzamide (IBZM) correlates with sensitivity to dopamine agonists in two patients with macroprolactinomas. *J Clin Endocrinol Metab*. 1995;80:2523–2525.
82. Pellegrini I, Rasolonjanahary R, Gunz G, et al. Resistance to bromocriptine in prolactinomas. *J Clin Endocrinol Metab*. 1989;69:500–509.
83. Passos VQ, et al. Genes differentially expressed in prolactinomas responsive and resistant to dopamine agonists. *Neuroendocrinology*. 2009;89:163–170.
84. Caccavelli L, Feron F, Morange I, et al. Decreased expression of the two D2 dopamine receptor isoforms in bromocriptine-resistant prolactinomas. *Neuroendocrinology*. 1994;60:314–322.

85. Peverelli E, Mantovani G, Vitali E, et al. Filamin-A is essential for dopamine d2 receptor expression and signaling in tumorous lactotrophs. *J Clin Endocrinol Metab*. 2012;97:967–977.
86. Kars M, Roelfsema F, Romijn JA, et al. Malignant prolactinoma: case report and review of the literature. *Eur J Endocrinol*. 2006;155:523–534.
87. Molitch ME. Medical management of prolactin-secreting pituitary adenomas. *Pituitary*. 2002;5:55–65.
88. Del Dotto P, Bonuccelli U. Clinical pharmacokinetics of cabergoline. *Clin Pharmacokinet*. 2003;42:633–645.
89. Webster J, et al. A comparison of cabergoline and bromocriptine in the treatment of hyperprolactinemic amenorrhea. Cabergoline Comparative Study Group. *N Engl J Med*. 1994;331:904–909.
90. Pascal-Vigneron V, Weryha G, Bosc M, et al. [Hyperprolactinemic amenorrhea:treatment with cabergoline versus bromocriptine. Results of a national multicenter randomized double-blind study]. *Presse Med*. 1995;24:753–757.
91. Colao A, Di Sarno A, Landi ML, et al. Macroprolactinoma shrinkage during cabergoline treatment is greater in naive patients than in patients pretreated with other dopamine agonists: a prospective study in 110 patients. *J Clin Endocrinol Metab*. 2000;85:2247–2252.
92. Verhelst J, Abs R, Maiter D, et al. Cabergoline in the treatment of hyperprolactinemia: a study in 455 patients. *J Clin Endocrinol Metab*. 1999;84:2518–2522.
93. Colao A, Di Sarno A, Sarnacchiaro F, et al. Prolactinomas resistant to standard dopamine agonists respond to chronic cabergoline treatment. *J Clin Endocrinol Metab*. 1997;82:876–883.
94. Colao A, Di Sarno A, Cappabianca P, et al. Withdrawal of long-term cabergoline therapy for tumoral and nontumoral hyperprolactinemia. *N Engl J Med*. 2003;349:2023–2033.
95. Dekkers OM, Lagro J, Burman P, et al. Recurrence of hyperprolactinemia after withdrawal of dopamine agonists: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2010;95:43–51.
96. Hu J, Zheng X, Zhang W, et al. Current drug withdrawal strategy in prolactinoma patients treated with cabergoline: a systematic review and meta-analysis. *Pituitary*. 2015;18:745–751.
97. Xia MY, Lou XH, Lin SJ, et al. Optimal timing of dopamine agonist withdrawal in patients with hyperprolactinemia: a systematic review and meta-analysis. *Endocrine*. 2018;59:50–61.
98. Souteiro P, Belo S, Carvalho D. Dopamine agonists in prolactinomas: when to withdraw? *Pituitary*. 2020;23:38–44.
99. Kwancharoen R, Auriemma RS, Yenokyan G, et al. Second attempt to withdraw cabergoline in prolactinomas: a pilot study. *Pituitary*. 2014;17:451–456.
100. Vilar L, Albuquerque JL, Gadelha PS, et al. Second attempt of cabergoline withdrawal in patients with prolactinomas after a failed first attempt: is it worthwhile? *Front Endocrinol (Lausanne)*. 2015;6:11.
101. Schade R, Andersohn F, Suissa S, et al. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med*. 2007;356:29–38.
102. Stiles CE, Tetteh-Wayoe ET, Bestwick J, et al. A meta-analysis of the prevalence of cardiac valvulopathy in hyperprolactinemic patients treated with Cabergoline. *J Clin Endocrinol Metab*. 2019;104:523–538.
103. Colao A, Galderisi M, Di Sarno A, et al. Increased prevalence of tricuspid regurgitation in patients with prolactinomas chronically treated with cabergoline. *J Clin Endocrinol Metab*. 2008;93:3777–3784.
104. Steeds R, Stiles C, Sharma V, et al. Echocardiography and monitoring patients receiving dopamine agonist therapy for hyperprolactinaemia: a joint position statement of the British Society of Echocardiography, the British Heart Valve Society and the Society for Endocrinology. *Clin Endocrinol (Oxf)*. 2019;90:662–669.
105. Moore TJ, Glenmullen J, Mattison DR. Reports of pathological gambling, hypersexuality, and compulsive shopping associated with dopamine receptor agonist drugs. *JAMA Intern Med*. 2014;174:1930–1933.
106. Dogansen SC, Cikrikcili U, Oruk G, et al. Dopamine agonist-induced impulse control disorders in patients with prolactinoma: a cross-sectional multicenter study. *J Clin Endocrinol Metab*. 2019;104:2527–2534.
107. Gillam MP, Molitch ME, Lombardi G, et al. Advances in the treatment of prolactinomas. *Endocr Rev*. 2006;27:485–534.
108. Grosvenor AE, Laws ER. The evolution of extracranial approaches to the pituitary and anterior skull base. *Pituitary*. 2008;11:337–345.
109. Prevedello DM, Doglietto F, Jane JA, et al. History of endoscopic skull base surgery: its evolution and current reality. *J Neurosurg*. 2007;107:206–213.
110. Dumont AS, Kanter AS, Jane JAJ, et al. Extended transsphenoidal approach. *Front Horm Res*. 2006;34:29–45.
111. Losa M, Mortini P, Barzaghi R, et al. Surgical treatment of prolactin-secreting pituitary adenomas: early results and long-term outcome. *J Clin Endocrinol Metab*. 2002;87:3180–3186.
112. Casanueva FF, Barkan AL, Buchfelder M, et al. Criteria for the definition of Pituitary Tumor Centers of Excellence (PTCOE): A Pituitary Society Statement. *Pituitary*. 2017;20:489–498.
113. Jethwa PR, Patel TD, Hajart AF, et al. Cost-effectiveness analysis of microscopic and endoscopic transsphenoidal surgery versus medical therapy in the management of microprolactinoma in the United States. *World Neurosurg*. 2016;87:65–76.
114. Primeau V, Raftopoulos C, Maiter D. Outcomes of transsphenoidal surgery in prolactinomas: improvement of hormonal control in dopamine agonist-resistant patients. *Eur J Endocrinol*. 2012;166:779–786.
115. Vroonen L, Jaffrain-Rea ML, Petrossians P, et al. Prolactinomas resistant to standard doses of cabergoline: a multicenter study of 92 patients. *Eur J Endocrinol*. 2012;167:651–662.
116. Tyrrell JB, Lamborn KR, Hannegan LT, et al. Transsphenoidal microsurgical therapy of prolactinomas: initial outcomes and long-term results. *Neurosurgery*. 1999;44:254–261; discussion 261–263.
117. Thomson JA, Gray CE, Teasdale GM. Relapse of hyperprolactinemia after transsphenoidal surgery for microprolactinoma: lessons from long-term follow-up. *Neurosurgery*. 2002;50:36–39; discussion 39–40.
118. Nawar RN, AbdelMannan D, Selman WR, et al. Pituitary tumor apoplexy: a review. *J Intensive Care Med*. 2008;23:75–90.
119. Kreutzer J, Buslei R, Wallaschofski H, et al. Operative treatment of prolactinomas: indications and results in a current consecutive series of 212 patients. *Eur J Endocrinol*. 2008;158:11–18.
120. Dai C, Liu X, Ma W, et al. The treatment of refractory pituitary adenomas. *Front Endocrinol (Lausanne)*. 2019;10:334.
121. Ježková J, Hána V, Kosák M, et al. Role of gamma knife radiosurgery in the treatment of prolactinomas. *Pituitary*. 2019;22:411–421.
122. Hung YC, Lee CC, Yang HC, et al. The benefit and risk of stereotactic radiosurgery for prolactinomas: an international multicenter cohort study. *J Neurosurg*. 2019:1–10.
123. Phillips J, East HE, French SE, et al. What causes a prolactinoma to be aggressive or to become a pituitary carcinoma? *Hormones (Athens)*. 2012;11:477–478.
124. Karunakaran S, Page RC, Wass JA. The effect of the menopause on prolactin levels in patients with hyperprolactinaemia. *Clin Endocrinol (Oxf)*. 2001;54:295–300.
125. Testa G, egetti W, Motta T, et al. Two-year treatment with oral contraceptives in hyperprolactinemic patients. *Contraception*. 1998;58:69–73.
126. Gillam MP, et al. The novel use of very high doses of cabergoline and a combination of testosterone and an aromatase inhibitor in the treatment of a giant prolactinoma. *J Clin Endocrinol Metab*. 2002;87:4447–4451.
127. Ribeiro RS, Abucham J. Recovery of persistent hypogonadism by clomiphene in males with prolactinomas under dopamine agonist treatment. *Eur J Endocrinol*. 2009;161:163–169.
128. Bronstein MD, Knoepfelmacher M, Liberman B, et al. Absence of suppressive effect of somatostatin on prolactin levels in patients with hyperprolactinemia. *Horm Metab Res*. 1987;19:271–274.
129. Hofland LJ, van der Hoek J, van Koetsveld PM, et al. The novel somatostatin analog SOM230 is a potent inhibitor of hormone release by growth hormone- and prolactin-secreting pituitary adenomas in vitro. *J Clin Endocrinol Metab*. 2004;89:1577–1585.
130. Lasolle H, et al. Pasireotide: A potential therapeutic alternative for resistant prolactinoma. *Ann Endocrinol (Paris)*. 2019;80:84–88.
131. Coopmans EC, van Meyel SWF, Pieterman KJ, et al. Excellent response to pasireotide therapy in an aggressive and dopamine-resistant prolactinoma. *Eur J Endocrinol*. 2019;181:K21–K27.
132. Ortiz LD, Syro LV, Scheithauer BW, et al. Temozolomide in aggressive pituitary adenomas and carcinomas. *Clinics (Sao Paulo)*. 2012;67:119–123.
133. McCormack A, Dekkers OM, Petersenn S, et al. Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*. 2018;178:265–276.

134. Cooper O, Mamelak A, Bannykh S, et al. Prolactinoma ErbB receptor expression and targeted therapy for aggressive tumors. *Endocrine*. 2014;46:318–327.
135. Li C, Sun Z, Gui S, et al. Effects of fulvestrant, an estrogen receptor antagonist, on MMQ cells and its mechanism. *Neuro Endocrinol Lett*. 2009;30:268–274.
136. Molitch ME. Pregnancy and the hyperprolactinemic woman. *N Engl J Med*. 1985;312:1364–1370.
137. Bronstein MD. Prolactinomas and pregnancy. *Pituitary*. 2005;8:31–38.
138. Musolino NR, Bronstein M. Prolactinomas and pregnancy. In: Bronstein M, ed. *Pituitary Tumors in Pregnancy*. Boston: Kluwer Academic Publishers; 2001:91–108.
139. Krupp P, Monka C. *Bromocriptine in pregnancy: safety aspects*. Klin Wochenschr; 1987;65:823–827.
140. Konopka P, et al. Continuous administration of bromocriptine in the prevention of neurological complications in pregnant women with prolactinomas. *Am J Obstet Gynecol*. 1983;146:935–938.
141. Huang W, Molitch ME. Pituitary tumors in pregnancy. *Endocrinol Metab Clin North Am*. 2019;48:569–581.
142. Sant' Anna BG, Musolino NRC, Gadelha MR, et al. A Brazilian multicentre study evaluating pregnancies induced by cabergoline in patients harboring prolactinomas. *Pituitary*. 2019;23:120–128.
143. Bronstein MD, Paraiba DB, Jallad RS. Management of pituitary tumors in pregnancy. *Nat Rev Endocrinol*. 2011;7:301–310.
144. Auriemma RS, et al. Results of a single-center observational 10-year survey study on recurrence of hyperprolactinemia after pregnancy and lactation. *J Clin Endocrinol Metab*. 2013;98:372–379.
145. Ono M, Miki N, Amano K, et al. Individualized high-dose cabergoline therapy for hyperprolactinemic infertility in women with micro- and macroprolactinomas. *J Clin Endocrinol Metab*. 2010;95:2672–2679.
146. Glezer A, Bronstein MD. Prolactinomas, cabergoline, and pregnancy. *Endocrine*. 2014;47:64–69.
147. Rastogi A, Bhadada SK, Bhansali A. Pregnancy and tumor outcomes in infertile women with macroprolactinoma on cabergoline therapy. *Gynecol Endocrinol*. 2017;33:270–273.
148. Karaca Z, et al. How does pregnancy affect the patients with pituitary adenomas: a study on 113 pregnancies from Turkey. *J Endocrinol Invest*. 2018;41:129–141.
149. Lambert K, et al. Macroprolactinomas and nonfunctioning pituitary adenomas and pregnancy outcomes. *Obstet Gynecol*. 2017;129:185–194.
150. Holmgren U, Bergstrand G, Hagenfeldt K, et al. Women with prolactinoma--effect of pregnancy and lactation on serum prolactin and on tumour growth. *Acta Endocrinol*. 1986;111:452–459.
151. Lebbe M, et al. Outcome of 100 pregnancies initiated under treatment with cabergoline in hyperprolactinaemic women. *Clin Endocrinol (Oxf)*. 2010;73:236–242.
152. Domingue ME, et al. Outcome of prolactinoma after pregnancy and lactation: a study on 73 patients. *Clin Endocrinol (Oxf)*. 2014;80:642–648.
153. Peillon F, Racadot J, Moussy DEA. Prolactin-secreting adenomas: a correlative study of morphological and clinical data. In: Fahlbuch R, KV W, eds. *Treatment of Pituitary Adenomas*. Stuttgart: Thieme; 1978:114.
154. Stalldecker G, Mallea-Gil MS, Guitelman M, et al. Effects of cabergoline on pregnancy and embryo-fetal development: retrospective study on 103 pregnancies and a review of the literature. *Pituitary*. 2010;13:345–350.
155. Goffin V. Prolactin receptor targeting in breast and prostate cancers: New insights into an old challenge. *Pharmacol Ther*. 2017;179:111–126.

5

Hypopituitarism Including Growth Hormone Deficiency

Gudmundur Johannsson and Oskar Ragnarsson

OUTLINE

History, 66
Etiology of Hypopituitarism, 66
  Hypophysitis, 67
Morbidity and Mortality, 67
Clinical Features, 69
Diagnosis and Endocrine Evaluation, 69
  Adrenocorticotropic Hormone Deficiency: Secondary Adrenal Insufficiency, 69
  Thyroid-Stimulating Hormone Deficiency: Central Hypothyroidism, 70
  Luteinizing Hormone/Follicle-Stimulating Hormone Deficiency: Hypogonadotropic Hypogonadism, 71
  Growth Hormone Deficiency, 71
  Hypopituitarism Following Pituitary Surgery, 72
Hormone replacement Therapy, 72
  Adrenocorticotropic Hormone Deficiency: Secondary Adrenal Insufficiency, 72
  Thyroid-Stimulating Hormone Deficiency Central Hypothyroidism, 74
  Luteinizing Hormone/Follicle-Stimulating Hormone Deficiency: Hypogonadotropic Hypogonadism, 75
  Growth Hormone Deficiency, 76
  Childhood-Onset Growth Hormone Deficiency in Adults, 77
  Hormone–Hormone Interactions, 77
Pregnancy, 77
Patient Care–Multidisciplinary Approach and Patient Education, 78

HISTORY

Pituitary insufficiency was first convincingly described in the beginning of the nineteenth century by Alfred Frölich, who presented a man with a pituitary tumor, abnormal fat accumulation, and underdeveloped genitalia, a state he called "adiposogenital dystrophy."[1] At about the same time, Joseph Babinski also described a patient with a pituitary tumor and similar clinical features.[2] Undoubtedly, both of these patients suffered from hypopituitarism, at that time called the Babinski–Frölich syndrome.

Since these first descriptions of hypopituitarism, a growing body of scientific discoveries has led us to today's knowledge of pituitary physiology and pathophysiology. In fact, several of the major discoveries that have been awarded the Nobel prize in physiology or medicine are related to the hypothalamopituitary medical field: Houssay in 1947 for discovering the influence of anterior pituitary hormones on glucose metabolism; Kendal, Reichstein, and Hench in 1950 for the discovery of cortisol; Guillemin and Schally in 1977 for the detection and isolation of the hypothalamic hormones thyrotropin-releasing hormone and gonadotropin-releasing hormone (GnRH); Yalow, also in 1977, for the development of radioimmunoassays for peptide hormones; and Lauterbur and Mansfield in 2003 for their pioneering work concerning magnetic resonance imaging.

ETIOLOGY OF HYPOPITUITARISM

A large number of medical conditions can cause hypopituitarism (Table 5.1). The most common cause is a pituitary adenoma, mainly nonfunctioning pituitary adenomas, but also hormone-producing adenomas. In Sweden, the annual incidence of pituitary adenomas between 2001 and 2011 was 3.9 cases per 100,000 inhabitants, and the lifetime risk for developing an adenoma was 0.3%.[3] Approximately 50% of all pituitary adenomas are nonfunctioning, and one third are prolactinomas. Between 40% and 85% of patients with nonfunctioning pituitary adenoma have some degree of pituitary insufficiency at diagnosis, varying from deficient production of one pituitary hormone to panhypopituitarism.[4,5,6] Although pituitary function in some patients can improve following pituitary surgery, development of new hormone deficits is also common.[4,6]

Hypopituitarism is less common in patients with hormone-producing adenomas because the symptoms and signs associated with the hypersecretion draws the attention toward the pituitary at an earlier stage when the adenoma has not reached a sufficient size to affect the function of the pituitary gland. Of note is that pituitary adenomas, both functioning and nonfunctioning, almost never present with diabetes insipidus.

Neoplasms other than pituitary adenomas in the sella region are less common. Many of these, however, like craniopharyngiomas and pituitary metastases, frequently present with multiple pituitary hormone deficiencies, including diabetes insipidus.[7,8] Similarly, when the pituitary gland is affected by a systemic disease such as neurosarcoidosis, Langerhans cell histiocytosis, granulomatosis with polyangitis, and hemochromatosis, panhypopituitarism and diabetes insipidus are also common. Other symptoms related to the underlying disease itself may, however, complicate the clinical picture and delay the detection of hypopituitarism (see also Chapters 6, 7 and 9).

Pituitary cysts are common, especially Rathke's cysts, which are seen in more than 10% of the population.[9] Pituitary cysts are usually small and infrequently causes hypopituitarism. However, large pituitary cysts do occur that may cause pituitary insufficiency. In these cases, the greatest clinical challenge is to differentiate the lesion from a craniopharyngioma or a pituitary adenoma with a large cystic component.

TABLE 5.1 Etiology of Acquired Hypopituitarism in Adults

Pituitary adenomas
Nonfunctioning
Prolactinoma
Acromegaly
Cushing disease
TSHoma

Other neoplasms in the sellar/suprasellar region
Craniopharyngioma
Meningioma
Metastasis
Lymphoma
Germinoma
Teratoma
Pituicytoma
Glioma
Chordoma
Chondrosarcoma
Ependymoma

Pituitary cysts
Rathke's cleft cyst
Epidermoid cyst
Subarachnoid cyst

Hypophysitis
Lymphocytic hypophysitis (autoimmune hypophysitis)
IgG4-related
Checkpoint inhibitor–induced

Pituitary infarction
Sheehan syndrome
Apoplexy

Infiltrative diseases
Sarcoidosis
Langerhans cell histiocytosis
Hemochromatosis
Granulomatosis with polyangitis (Wegener granulomatosis)

Infectious diseases
Abscess
Tuberculoma

Miscellaneous
Pituitary radiotherapy
Pituitary surgery
Radiotherapy for intracerebral tumors and nasopharyngeal cancer
Empty sella
Traumatic brain injury
Opioid-induced

Depending on etiology, patients with hypopituitarism may present with various symptoms other than symptoms related to the hormone deficiencies themselves.

Pituitary insufficiency develops insidiously in most cases. However, there are some important exceptions when this happens acutely. One of these is Sheehan syndrome, an acute infarction of the pituitary gland in women following childbirth, caused by blood loss and hypovolaemia.[10] Another important cause of acute pituitary insufficiency is pituitary apoplexy,(i.e., bleeding into a pituitary adenoma).[11] In both these scenarios, symptoms and signs because of acute secondary adrenal insufficiency dominate the clinical picture. Patients with pituitary apoplexy may also present with local symptoms because of the sudden expansion of the tumour, that is, optic nerve and/or chiasmal compression, ocular pares owing to pressure on the cranial nerves in the cavernous sinus, as well as severe headache, fever, and impaired consciousness resulting from meningeal irritation.

Hypophysitis

Hypophysitis has until recently been considered a rare cause of hypopituitarism. Several different forms of hypophysitis exist; lymphocytic hypophysitis (also called autoimmune hypophysitis) and IgG4-related hypophysitis are the most common.[12] Although hypophysitis can occur at any age in both men and women, lymphocytic hypophysitis typically affects women during late pregnancy or during the first 3 months after delivery. The most common clinical presentation of hypophysitis is headache and hypopituitarism, including ineffective or absent lactation. Symptoms because of cranial nerve compression occurs less often.

With the arrival of new biological drugs, the incidence of hypophysitis has increased substantially.[13] Immune checkpoint molecules can either have an inhibiting or stimulating effect on T-cell function in solid tumors. Both cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) play a pivotal role in the maintenance of immunological tolerance, thereby preventing development of autoimmune disorders. Therefore, when immune checkpoint inhibitors are used for the treatment of cancer, the development of autoimmune disorder is a common side effect. Indeed, autoimmune endocrine disorders are among the most common adverse events related to use of these drugs, including hypothyroidism, hyperthyroidism, hypophysitis, primary adrenal insufficiency, and insulin-dependent diabetes mellitus. Among the commonly used immune checkpoint inhibitors is ipilimumab, a CTLA-4 inhibitor used for treatment of advanced malignant melanoma that causes hypophysitis in up to 10% of patients.[14] Although immune checkpoint inhibitors acting through PD-1 are associated with a much lower incidence of hypophysitis, the combination of CTLA-4 and PD-1 inhibitors seem to be associated with greatest risk of endocrine-related adverse events, including hypophysitis.[15] Of interest is that isolated adrenocorticotropic hormone (ACTH) deficiency is common in patients with immune checkpoint inhibitor–induced hypophysitis, in contrast to patients with pituitary adenoma, who almost always have at least one other pituitary hormone deficiency before they develop secondary adrenal insufficiency.

MORBIDITY AND MORTALITY

Patients with hypopituitarism have increased mortality,[16] as first demonstrated in two seminal studies from Sweden and the United Kingdom.[17,18] Both studies showed a markedly increased mortality (Fig. 5.1), suggesting that management of the underlying disorder, as well as endocrine replacement therapy, could improve. In a metaanalysis including data from six observational studies with over 19,000 hypopituitary patients, the standardized mortality ratio (SMR) was doubled in patients with pituitary insufficiency compared with a reference group, most often the background population (Fig. 5.2).[19] Furthermore, women with hypopituitarism have a higher SMR than men, and young age is associated with higher SMR.[19,20] Other factors that have been found to be associated with increased mortality are inadequate hormone replacement therapy and pituitary radiotherapy, as well as the underlying etiology, such as a craniopharyngioma.[18,21,22] A limitation of most epidemiological studies on mortality (and morbidity) is the inclusion of patients with different causes of pituitary insufficiency. Nevertheless, in a study including only patients with nonfunctioning pituitary adenoma, patients with hypopituitarism still

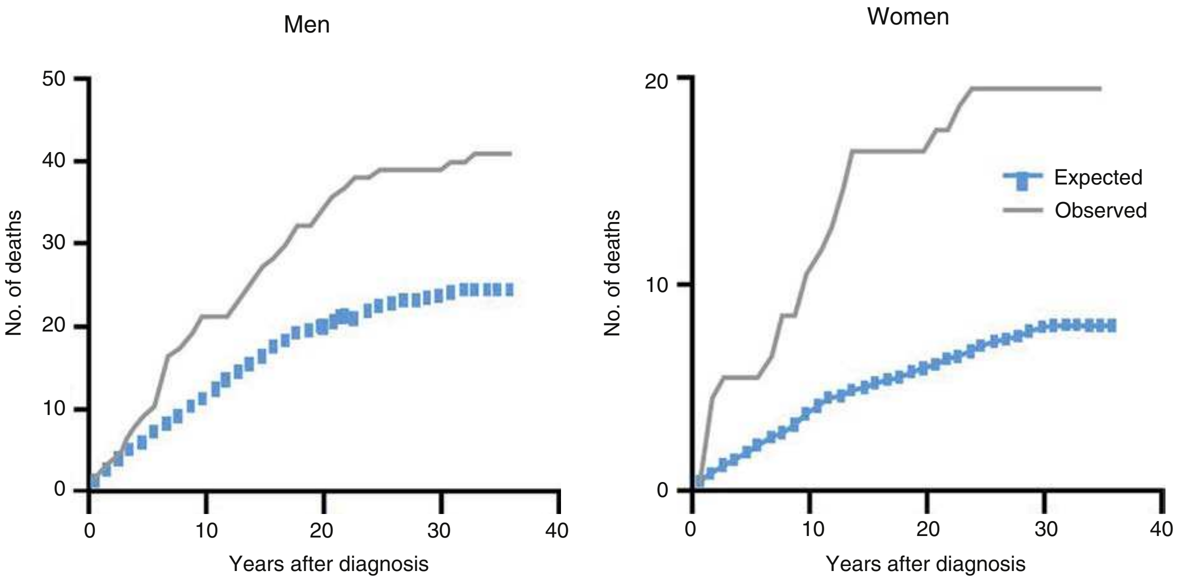


Fig. 5.1 **Observed Number of Deaths Among 333 Patients With Hypopituitarism Compared With the Expected Number in Age- And Gender-Matched Controls.** (From Rosen T, Bengtsson BA. Premature mortality due to cardiovascular disease in hypopituitarism. *Lancet*. 1990;336:285–288. Reprinted with permission from Elsevier.)

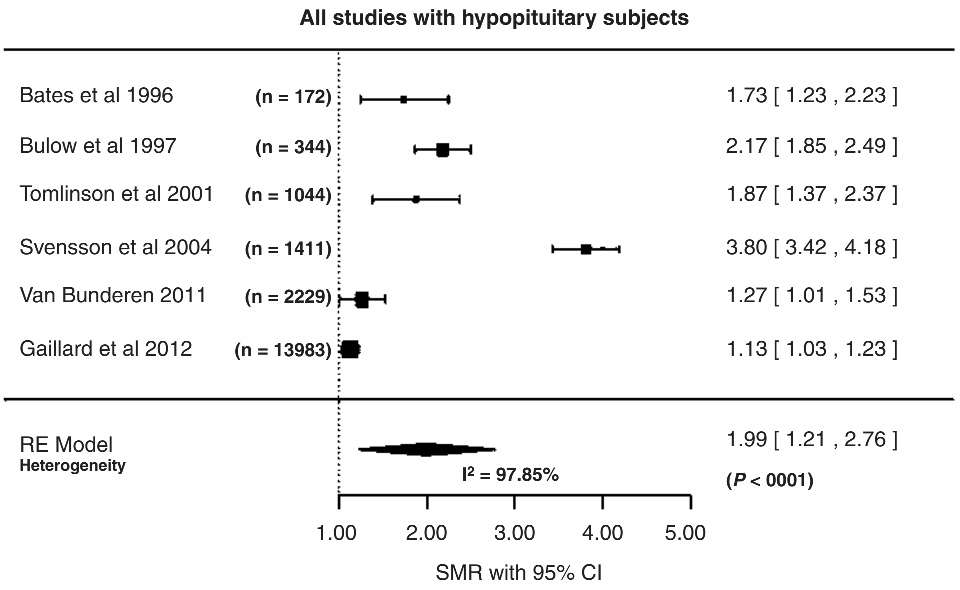


Fig. 5.2 **Forest Plot Showing the Pooled Standardized Mortality Ratio (SMR) in Patients with Hypopituitarism From a Metaanalysis of Six Observational Studies Including 19,153 Patients.** *CI*, Confidence interval; *RE*, random effects. (From Pappachan JM, Raskauskiene D, Kutty VR, et al. Excess mortality associated with hypopituitarism in adults: a meta-analysis of observational studies. *J Clin Endocrinol Metab*. 2015;100:1405–1411. Reprinted with permission from Oxford University Press.)

had an increased mortality rate compared with patients with unaffected pituitary function.[20]

Inadequate hormone replacement such as untreated hypogonadotropic hypogonadism, untreated growth hormone (GH) deficiency, as well as nonphysiologic glucocorticoid replacement, may contribute to the increased mortality in patients with hypopituitarism. In fact, patients with untreated hypogonadotropic hypogonadism have a doubled mortality rate compared with their counterparts.[18] It is therefore unfortunate that one-fourth of women of reproductive age with hypogonadotropic hypogonadism are still not receiving estrogen replacement therapy.[23] Similarly, in observational studies, patients with untreated GH deficiency have more than doubled mortality rate compared with patients receiving GH replacement who have an SMR comparable to the background population.[19] Furthermore, patients with ACTH deficiency, especially patients receiving relatively high glucocorticoid doses, have an adverse cardiometabolic profile,[24] as well as increased mortality.[21,22,25,26] Although the excess mortality has mainly been related to the increased risk of cardiovascular and cerebrovascular diseases, deaths from infections are also of importance.[27-29] In fact, patients with secondary adrenal insufficiency have more than 4-fold

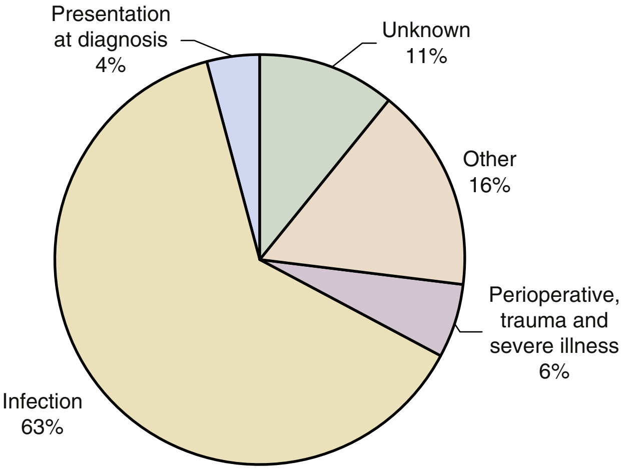


Fig. 5.3 Precipitating Factors of Adrenal Crisis in Patients with Secondary Adrenal Insufficiency. (Modified from Smans LC, Van der Valk ES, Hermus AR, et al. Incidence of adrenal crisis in patients with adrenal insufficiency. *Clin Endocrinol (Oxf)*. 2016;84:17–22.)

increased inpatient admissions because of infections compared with controls,[27] four of 100 patients develop adrenal crisis per year, where infections are the most common cause (Fig. 5.3),[28] and the mortality rate because of infections is six times higher than in the general population.[29]

CLINICAL FEATURES

Clinical features related to hypopituitarism occur insidiously because of gradual loss of anterior pituitary function.

KEY POINTS: Clinical Features in Patients with Hypopituitarism:

Adrenocorticotropic hormone deficiency	**Chronic:** Fatigue, lethargy, weight loss, nausea, and anorexia **Acute:** Nausea, vomiting, hypotension, decreased growth of pubic and axillary hair in women, and hyponatraemia
Thyroid-stimulating hormone deficiency	Fatigue, lethargy, cold intolerance, constipation, dry skin, hair loss, reduced energy, and concentration difficulties
Luteinizing hormone/ follicle-stimulating hormone deficiency	**Women:** Oligomenorrhoea or amenorrhoea **Men:** Erectile dysfunction, decreased muscle mass, decreased hair growth, and anemia **Both:** Sexual dysfunction, infertility, osteoporosis, poor concentration and energy
Growth hormone deficiency	Abdominal obesity, reduced muscle mass, lack of energy and initiative, and impaired quality of life
Prolactin deficiency	Ineffective or absent lactation

However, the clinical presentation can also be dominated by the underlying disorder, such as a pituitary tumor that causes pressure on nearby structures such as the optic nerves, chiasma, and/or the hypothalamus. As symptoms and signs related to a gradual loss of pituitary function are insidious, and often not recognized, the first symptom leading to the diagnosis is therefore frequently loss of visual acuity and/or visual field defects.

When hypopituitarism is suspected, the medical history is of paramount importance, and should include assessment of fatigue and general well-being, metabolic risk factors, changes in body weight and body fat distribution, muscle function, sexual function and fertility, and menstrual history, as well as symptoms related to adrenal insufficiency, in particular the response to previous infections and the recovery time. On physical examination, body composition should be noted, including evaluation of fat distribution and muscle mass, as well as color of the skin, body hair, and testicular volume.

The most common cause of hypopituitarism is a pituitary adenoma. These are mostly slowly growing, and, through pressure on the pituitary, loss of function occurs gradually. Interestingly, this usually occurs in a certain order, which is the same or similar order as in hypopituitarism caused by radiotherapy,[30] where GH and luteinizing hormone (LH)/follicle-stimulating hormone (FSH) deficiencies develop first, followed by thyroid-stimulating hormone (TSH) and ACTH deficiencies, and lastly prolactin deficiency. The reason for this is not clear, but likely reflects the different localization of the hormone-secreting cells in the pituitary and their resistance to pressure and irradiation. Symptoms that occur during the early stage of hypopituitarism are therefore related to GH deficiency and hypogonadotropic hypogonadism.

Isolated GH deficiency is likely a common feature of hypopituitarism during the early stage of its development. GH deficiency in adults is mainly associated with fatigue and abdominal obesity, symptoms that are common in various other and more common conditions, as well as in the general population. Isolated GH deficiency is therefore unlikely to be detected unless other coexisting symptoms and signs related to the underlying disease are also present. Oligo- or amenorrhea caused by hypogonadotropic hypogonadism is also an early and common feature of hypopituitarism in women of fertile age. Again, because other causes than hypopituitarism are much more common, the correct diagnosis is often delayed. Thus, symptoms because of GH deficiency and hypogonadotropic hypogonadism are often nonspecific and less distinct than the clinical features caused by secondary adrenal insufficiency. Diagnostic delay is therefore common.

Although the clinical presentation is most often insidious, more acute presentations may occur. These can be either related to the underlying cause where the order of loss of pituitary function is different from what is seen in a slowly growing pituitary adenoma, or when the cause of hypopituitarism develops suddenly. In both these cases the rapid development of adrenal insufficiency usually makes the clinical course acute. Sheehan syndrome and pituitary apoplexy are examples of disorders that often present with acute panhypopituitarism. Hypophysitis, either autoimmune or because of immune checkpoint inhibitors, and pituitary metastasis, often present with isolated ACTH deficiency. In both of these examples the presentation is more or less acute and may be life-threatening because of an adrenal crisis.

DIAGNOSIS AND ENDOCRINE EVALUATION

Adrenocorticotropic Hormone Deficiency: Secondary Adrenal Insufficiency

Secondary adrenal insufficiency is a potentially fatal condition and should therefore be ruled out in all patients with suspected hypopituitarism and/or a lesion in the hypothalamopituitary region.[31]

KEY POINTS: Biochemical Evaluation of Patients with Suspected Hypopituitarism

Adrenocorticotropic hormone (ACTH) deficiency	Stable patients: • Measure serum cortisol at 8:00 AM. • Serum cortisol >400–450 nmol/L rules out secondary adrenal insufficiency • Serum cortisol <50–100 nmol/L confirms secondary adrenal insufficiency after exogenous glucocorticoid use has been ruled out in patients with symptoms compatible with hypocortisolism • Perform ACTH stimulation test in patients with inconclusive serum cortisol Instable patients: • Collect blood samples for analysis of cortisol and ACTH in acutely ill and/or hemodynamically instable patients, and start empirical treatment with parenteral glucocorticoids without a delay Acute secondary adrenal insufficiency: • ACTH stimulation cannot be used when acute secondary adrenal insufficiency is suspected • Evaluation of suspected acute secondary adrenal insufficiency should be based on clinical features and nonstimulated serum cortisol concentrations
Thyroid-stimulating hormone (TSH) deficiency	• Measure free T4 and TSH • Low or inappropriately normal TSH, together with low free T4, is compatible with central hypothyroidism
Luteinizing hormone (LH)/follicle-stimulating hormone (FSH) deficiency	• Measure LH, FSH and estrogen (women) or testosterone (men) • Low or inappropriately normal LH and FSH, in combination with low concentrations of estrogen in women and low testosterone in men, is compatible with hypogonadotropic hypogonadism • Measure serum prolactin to rule out hyperprolactinemia as a cause for hypogonadotropic hypogonadism
Growth hormone (GH) deficiency	GH stimulation tests (GH peak is test-specific): • Insulin tolerance test • Arginine–GH–releasing hormone test • Glucagon stimulation test • MacimorelinIn patients with more than three anterior pituitary hormone deficiencies and low serum insulin-like growth factor 1, a GH stimulation test may not be needed

Also, secondary adrenal insufficiency has a pronounced negative impact on general well-being and quality of life that improves substantially with glucocorticoid replacement therapy.

In hemodynamically stable patients, secondary adrenal insufficiency can be ruled out if morning serum cortisol is greater than 400 to 450 nmol/L.[32,33] On the contrary, in patients with serum cortisol less than 50 to 100 nmol/L, secondary adrenal insufficiency can be considered to be confirmed in patients with symptoms compatible with adrenal insufficiency, and after use of exogenous glucocorticoids has been ruled out. In patients with a serum cortisol between 100 and 400 nmol/L, ACTH stimulation (also called the *synacthen* or *cosyntropin* test) is needed to rule out or confirm secondary adrenal insufficiency. During the test, serum cortisol is measured before, as well as 30 and/or 60 minutes after, a synthetic ACTH is injected intravenously or intramuscularly. Serum cortisol that exceeds 400 to 450 nmol/L can be considered an adequate response in most assays. Patients without an increase in serum cortisol, on the other hand, have adrenal insufficiency. Patients in whom serum cortisol increases following ACTH stimulation compared to the pretest value, but does not exceed 350 to 450 nmol/L, may be considered to have partial adrenal insufficiency. How and whether this should be replaced with daily doses of glucocorticoids, or only during intercurrent illness, should be based on the overall clinical picture of the individual patient. In the clinical setting where it is difficult to differentiate between primary and secondary adrenal insufficiency, plasma ACTH should also be measured. ACTH in patients with primary adrenal insufficiency is usually greatly elevated, whereas in patients with secondary adrenal insufficiency it is either low or inappropriately normal.

Traditionally, 250 µg of ACTH have been used for the stimulation test. However, because 250 µg is a supraphysiological dose, it has been suggested that a lower dose with 1 µg may be more adequate for the evaluation of secondary adrenal insufficiency. However, studies that have compared the two doses have not been able to support the superiority of either dose.[34] Another test used to diagnose secondary adrenal insufficiency is the insulin tolerance test, often considered to be the gold standard for evaluation of the hypothalamus-pituitary-adrenal (HPA) axis. During the test, insulin is administrated intravenously to provoke hypoglycemia, which has a powerful stimulating effect on CRH, arginine vasopressin, ACTH, and, subsequently, cortisol production. The main advantage with the test is that the whole HPA is tested, and that GH secretion, also stimulated by hypoglycemia, can be tested simultaneously. However, the test is cumbersome, is not without risks, and is contraindicated in patients with cardiovascular disease, patients with epilepsy, and in the elderly. Thus, the insulin tolerance test should only be done in units with sufficient experience in performing the test.

In the acutely ill and/or hemodynamically unstable patient, the greatest challenge is to think of secondary adrenal insufficiency as a differential diagnosis. If secondary adrenal insufficiency is suspected, blood samples should be collected immediately for analysis of plasma ACTH and serum cortisol. Empirical treatment with parenteral glucocorticoids should be initiated thereafter without delay and before the results from the biochemical analyses have arrived. No clear cutoff cortisol concentrations are available to confirm or rule out secondary adrenal insufficiency in a critically ill patient. Serum cortisol concentrations increase during stress, and the cutoff used to rule out adrenal insufficiency in otherwise stable patients (400–450 nmol/L) is, therefore, probably too low in the acute setting.

Secondary adrenal insufficiency usually develops gradually. However, there are some exceptions when secondary adrenal insufficiency develops acutely, (e.g., following pituitary surgery and in association with pituitary apoplexy or Sheehan syndrome). ACTH stimulation cannot be used under these circumstances. During chronic secondary adrenal insufficiency, a gradual downregulation of ACTH receptors occurs in the adrenal cortex. Thus, when acute secondary adrenal insufficiency develops, this downregulation has not occurred yet, and serum cortisol increases to normal concentrations following administration of exogenous ACTH. Evaluation of suspected acute secondary adrenal insufficiency should therefore be based on the clinical features and nonstimulated cortisol concentrations in serum.

ACTH also stimulates the production of androgens from the adrenal cortex. Thus, women with ACTH deficiency develop not only secondary adrenal insufficiency, but also androgen deficiency. Low concentrations of dehydroepiandrosterone (DHEA) sulfate in serum support the diagnosis.

Thyroid-Stimulating Hormone Deficiency: Central Hypothyroidism

Central hypothyroidism is, in the vast majority of cases, part of a more extensive hypopituitarism. Isolated central hypothyroidism may, however, occur in relation to specific underlying etiologies, such as hypophysitis induced by checkpoint inhibitors, as well as congenital causes that are most often diagnosed during neonatal screening or in early childhood. Genetic testing should be done in familial cases of central hypothyroidism and in cases where the etiology remains unexplained after initial evaluation, including imaging studies of the pituitary gland.[35] Symptoms and signs that are specifically related to central hypothyroidism may not be as striking as in patients with overt primary hypothyroidism, as these are often masked by symptoms and signs related to other pituitary hormone deficiencies.

Central hypothyroidism is diagnosed biochemically by measuring free T4 and TSH.[35] Subnormal concentrations of free T4, together with low or inappropriately normal serum TSH, should raise the suspicion of central hypothyroidism. In some cases, mainly where the cause is because of hypothalamic dysfunction, TSH concentrations may even be slightly elevated. In these cases, the circulating TSH has normal immunoreactivity but low bioactivity.

In patients with a well-defined hypothalamopituitary disorder who have low free T4 and inappropriately low TSH concentrations, the diagnosis of central hypothyroidism is obvious. In less distinct cases, the free T4 and TSH measurements should be repeated, and other causes such as nonthyroidal illness and drugs that reduce TSH secretion, such as glucocorticoids, dopamine, antiepileptics, antipsychotics, and metformin, should be ruled out. In milder forms of central hypothyroidism, free T4 concentration can be in the low-normal reference range, and a trial of L-thyroxine replacement may be of value. Also, patients with unreplaced GH deficiency may have central hypothyroidism despite a free T4 concentration in the low-normal range, because of reduced conversion of T4 to T3.[36]

Luteinizing Hormone/Follicle-Stimulating Hormone Deficiency: Hypogonadotropic Hypogonadism

The clinical presentation of hypogonadotropic hypogonadism depends on age and gender. In women of premenopausal age, menstrual disorder with either oligo- or amenorrhea and postmenopausal symptoms are typical. Common complaints in men are reduced muscle strength, increased fat accumulation, loss of masculine bodily hair, reduced frequency of shaving, fatigue, and impotence. Reduced libido or infertility can also be dominating symptoms in both genders.

Hypogonadotropic hypogonadism is characterized biochemically by low or inappropriately normal LH and FSH, in combination with low concentrations of estrogen in women and low testosterone in men.[37] Because hyperprolactinaemia, even of modest grade, is a common cause of hypogonadotropic hypogonadism, serum prolactin should always be measured. Biochemical confirmation of hypogonadotropic hypogonadism is essential in men and women of premenopausal age before replacement therapy is started. However, the clinical value of diagnosing hypogonadotropic hypogonadism in postmenopausal women, characterized by the same biochemical picture as in women of reproductive age, is limited.

Growth Hormone Deficiency

In patients with an expanding structural lesion within the hypothalamopituitary region, and in patients who have received pituitary radiation, GH is the first pituitary hormone to suffer. Therefore, GH deficiency is present in most patients with hypopituitarism and should be suspected in all patients with a hypothalamopituitary disorder. The diagnosis of GH deficiency is based on biochemical testing because most symptoms and signs related to GH deficiency, such as fatigue, concentration difficulties, lack of initiative, abdominal fat accumulation, and dyslipidemia, are nonspecific. Because endogenous GH secretion is highly pulsatile, with most of the GH secretion occurring during sleep and low GH concentrations during daytime, a GH stimulation test is needed in most cases to confirm the diagnosis.

GH is a powerful metabolic hormone that, through its action on the GH receptor and generation of endocrine and paracrine insulin-like growth factor 1 (IGF-1), induces effects in most organs in the human body.[38] Clinically, GH treatment is best known for its effects on the growth plate. In children with short stature because of GH deficiency, GH replacement increases growth velocity, and normal final height is frequently obtained (see also Chapters 17 and 18). The biological effects of GH are, however, much more diverse. GH and its interplay with insulin in fasting and fed states regulates metabolism and energy storage.[39] GH secretion is suppressed after ingestion of food and increases in the fasting state. GH increases mobilization of energy by inducing lipolysis and insulin resistance during fasting, and simultaneously promotes protein synthesis. GH also affects lipoprotein and lipid metabolism by upregulation of the low density lipoprotein receptor in the liver and increasing the hepatic output of very low density lipoproteins.[40] GH also increases bone resorption and bone formation, and thereby bone remodeling, through its effects on both osteoclasts and osteoblasts.[41] Furthermore, GH and IGF-1 have multiple effects on the cardiovascular system, including an inotropic effect on the heart muscle, and reducing peripheral vascular resistance, mediated by increased endothelial production of nitric oxide.[42] GH and IGF-1 are also found in the cerebrospinal fluid,[43] and GH and IGF-1 receptors are expressed in cells in the central nervous system, illustrating the importance of GH and IGF-1 for normal brain function.

Symptoms and Signs. Based on the physiological effects of GH in various clinical and experimental systems, some symptoms and signs related to adult GH deficiency may be expected. The best-described consequences of GH deficiency in adults are related to changes in body composition.[44] GH-deficient patients have increased body fat mass with abdominal preponderance, reduced muscle mass and strength, and reduced extracellular water. Reduced bone mineral density has also been associated in adults with GH deficiency, although data related to whether fracture risk is increased are less clear, although this has been suggested in some studies.[41] Dyslipidemia is also common, where triglycerides, total- and low density lipoprotein–cholesterol are increased, and high density lipoprotein–cholesterol is reduced, a lipid pattern associated with increased risk of cardiovascular disease.[44] Reduced cardiac output, endothelial dysfunction, and increased sympathetic nerve activity have also been shown in adults with GH deficiency.[42,45] Furthermore, quality of life and general well-being are frequently compromised.[46] Most of the features mentioned above improve or even normalize in response to GH replacement. It should, however, be recognized that most studies on clinical features in adults with GH deficiency have included subjects with severe hypopituitarism. Therefore, some of the clinical features considered to be associated with adult GH deficiency my actually be caused by inadequate replacement with other pituitary hormones, or effects of the underlying disease itself.[47]

Diagnosis. Diagnosis of GH deficiency in adults should be performed in the context of a well-defined hypothalamopituitary disorder (Table 5.1) where the probability of GH deficiency is high.[48-50] The main reason is that the specificity of the tests in an unselected population is often low, partially explained by a strong negative impact of obesity, in particular abdominal obesity, on GH secretion, and overdiagnosis of GH deficiency can therefore be avoided. Diagnosis of GH deficiency should also only be performed in patients where other pituitary function has been thoroughly evaluated and hormone deficiencies adequately replaced.

Various stimulation tests have been used and validated for the diagnosis of GH deficiency in adult, including the insulin tolerance test, arginine (Arg)-GH releasing hormone (GHRH), glucagon, Arg alone, GH-releasing peptide (GHRP) 2 and 6, as well as GHRP in combination with GHRH. The insulin tolerance test, previously used primarily to diagnose adrenal insufficiency, was evaluated early for the diagnosis of GH deficiency in adults.[51] In patients with well-established hypopituitarism, the insulin tolerance test convincingly identifies patients with GH deficiency by using a cutoff of 3 μg/L, and 5 μg/L in younger subjects. Also, results from the test are not markedly affected by obesity. However, it is labor-intensive, the hypoglycemia that is induced is inconvenient for the patient, and, because of safety, should not be used in the elderly and in patients with ischemic heart disease or epilepsy. In experienced centers its use is safe and the repeatability is high.

The combined Arg-GHRH test is also well-validated for the diagnosis of GH deficiency in adults. The GH peak obtained with the test has a high correlation with the results from the insulin tolerance test, and the repeatability is high. The cutoff used to diagnose GH deficiency with Arg-GHRH is body mass index (BMI)-dependent, with lower peaks in obese than in lean subjects (BMI <25, 11.5 μg/L; 25< BMI ≤30, 8 μg/L; BMI ≥30, 4.2 μg/L).[52] The test is well-tolerated by patients, but is limited by frequent false-negative results in patients who have received radiation therapy toward the hypothalamopituitary area, for up to 5 years after treatment, because GHRH directly stimulates the pituitary gland.[53] Other commonly used agents are glucagon and Arg alone. Both are well tolerated but limited by weight-dependent peak GH response. Therefore, by using the traditional cutoff of 3 μg/L, overdiagnosis of GH deficiency in obese subjects is common, and a more stringent criterion of 1 μg/L has been suggested, at least for the glucagon stimulation test.[48]

The GH secretagogue macimorelin was recently developed for the diagnosis of GH deficiency in adults.[54] The test allows oral administration of macimorelin after 8 hours of fasting, with collection of blood samples for GH in serum before administration and at 30, 45, 60, and 90 minutes after. The most frequent side effect is mild and transient dysgeusia. The correlation between the peak GH response in response to insulin tolerance test and macimorelin is very high, as is the repeatability. The suggested cutoff for the diagnosis of GH deficiency is 5.1 μg/L. The main limitations of macimorelin is that it is contraindicated in patients taking drugs that prolong the QT interval on electrocardiogram and cannot be used together with drugs that induce CYP3A4 (such as carbamazepine, phenytoin, rifampin, St. John's wort), because increased CYP3A4 activity may reduce the serum concentration of macimorelin and, therefore, give false-positive results.

Hypopituitarism Following Pituitary Surgery

Although pituitary function can improve following pituitary surgery in some patients, the development of new hormone deficiencies is more common.[55] Most important is to rule out secondary adrenal deficiency by measuring serum cortisol after stopping all glucocorticoid treatment for 24 hours. A very low cortisol (<100 nmol/L) is compatible with adrenal insufficiency, while values above 400 to 450 nmol/L rule it out. The remaining patients, with values between 100 and 400 nmol/L, should be followed closely until the adrenocortical function can be tested adequately with a synacthen test or insulin tolerance test approximately 6 weeks postoperatively. With this approach, new postoperative adrenal insufficiency, as well as restoration of normal cortisol production in some cases, can be detected.

Thyroid function should also be tested during the early postoperative period. However, because thyroxine has a relatively long half-life, a significant decrease in free T4 concentrations is typically first seen 1 week after surgery in patients who have developed central hypothyroidism. Because hypopituitarism following pituitary surgery develops acutely, dose titration of L-thyroxine with a low starting dose is not necessary.

Evaluation of gonadal function and GH secretion should not be done in the early postoperative period, because the anesthesia and the operation itself may have a significant effect on their secretions. Instead, evaluation of the gonadal axis should be reevaluated approximately 3 months postoperatively, and GH production 6 to 12 months following surgery, when other hormone deficiencies have been adequately evaluated and treated.[55]

HORMONE REPLACEMENT THERAPY

Adrenocorticotropic Hormone Deficiency: Secondary Adrenal Insufficiency

Glucocorticoid replacement therapy is essential for all patients with secondary adrenal insufficiency.[32] Hydrocortisone is the most commonly used glucocorticoid. Hydrocortisone has the same chemical structure as endogenously produced cortisol and has a relatively short half-life (1.5 hours), which makes it suitable for replacement where the dose and timing of intake can be used to mimic the physiological diurnal variation of cortisol production (Fig. 5.4). In some countries, cortisone

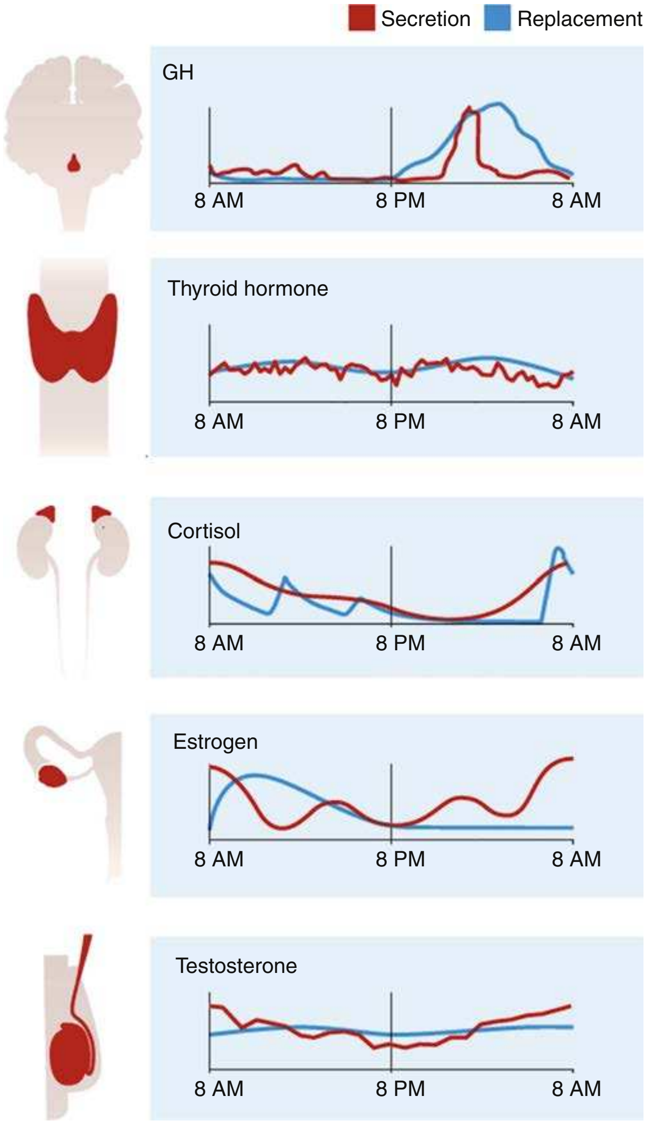


Fig. 5.4 Schematic Illustration of the Different Hormone Profiles in Serum During Normal Physiological Secretion *(Red Lines)* and Hormone Replacement *(Blue Lines)*. *GH,* Growth hormone.

acetate, also with a relatively short half-life, is the preferred glucocorticoid for patients with adrenal insufficiency. Two observational studies have demonstrated a better cardiometabolic profile in patients receiving cortisone acetate compared with hydrocortisone.[24,56] This may be explained by the fact that the glucocorticoid potency of hydrocortisone in relation to cortisone acetate is probably underestimated, that is, the assumption that 20 mg hydrocortisone equals 25 mg cortisone acetate is not correct. Replacement therapy with glucocorticoids such as prednisolone and dexamethasone, with a considerably longer half-life and higher potency, is rarely advocated.[33]

Hydrocortisone is usually administrated in two or three divided doses per day.

KEY POINTS: Treatment of Hypopituitarism

Adrenocorticotropic hormone deficiency	• Hydrocortisone (or cortisone acetate) in two or three divided doses, 15–20 mg per day, with the highest dose taken upon awakening • Administer increased glucocorticoid dose during stress (see following Key Points box) • Consider androgen replacement in women with secondary adrenal insufficiency
Thyroid-stimulating hormone (TSH) deficiency	• L-thyroxine is the preferred treatment • Low starting dose with dose titration aiming at free T4 concentration in the upper-normal reference range • TSH concentrations cannot be used to guide dose titration
Luteinizing hormone/follicle-stimulating hormone deficiency	• Women of premenopausal age: estrogen, together with progesterone if the uterus is intact • Women of postmenopausal age: no treatment needed • Men: testosterone
Growth hormone (GH) deficiency	• The initial daily GH dose should be no higher than 0.1–0.3 mg to avoid side effects • Higher starting dose can be used in the young, in particular in young women receiving oral estrogen replacement • The dose titration should be done cautiously to avoid side effects • Serum insulin-like growth factor 1 concentration should be normalized and preferably targeted at the upper-normal reference range

The first dose is taken upon waking, and the subsequent dose 4 to 6 hours later. A higher dose is administered in the morning than in the noon/afternoon to mimic the circadian cortisol production and exposure. Based on studies on cortisol production in healthy adults, 10 to 15 mg of hydrocortisone and 15 to 20 mg of cortisone acetate can be considered to be near-physiologic doses.[57,58] Thus, a typical dose regime for patients taking 15 mg hydrocortisone per day would be 10 mg upon awakening and 5 mg 6 hours later. In patients who experience symptoms because of hypocortisolemia before the second dose, a three-dose regime may be more appropriate, (e.g., 10+5+2.5 mg with ~4 hours in between). Although 10 to 20 mg of hydrocortisone per day is sufficient for most patients, some will need higher doses to achieve a satisfactory effect on quality of life. Doses of 30 mg or higher should, however, be avoided. In fact, high glucocorticoid replacement doses in patients with secondary adrenal insufficiency are associated with a higher BMI and low density lipoprotein, greater waist circumference, worse quality of life, and increased mortality (Fig. 5.5).[21,22,24,25,59] Recently, a new dual-release hydrocortisone formulation has been approved in Europe.[60] The tablet, administered once daily, consists of two components: a coating that is released immediately after ingestion, and a modified-release core. In comparison to conventional hydrocortisone replacement, this dual-release formulation provides a more physiological glucocorticoid exposure (Fig. 5.6) and seems to have beneficial effects on quality of life as well as cardiometabolic outcome.[61]

Patients with partial secondary adrenal insufficiency, (i.e., where serum cortisol increases following ACTH stimulation but does not exceed 350 to 450 nmol/L, depending on the assay), should receive individualized glucocorticoid replacement. Some of these patients need the same regime as patients with complete secondary adrenal insufficiency, others need chronic treatment but with lower doses (5–10 mg per day), and some may only need glucocorticoids during episodes of intercurrent illness.

Women with ACTH deficiency not only lack adequate cortisol production, but also have inadequate androgen production. Androgen replacement in women with secondary adrenal insufficiency may have beneficial effects on mood, sexual health, and general well-being.[62,63] For this purpose, orally administered DHEA (12.5–25 mg per day) is often used. However, the treatment response varies greatly. Some patients experience a clear beneficial effect, while others do not. Also, side effects such as acne, oily hair, and greasy skin are common, even with low doses. Current clinical guidelines recommend against routine replacement with androgens to women with adrenal insufficiency.[64] Thus, until more data on safety and effectiveness are available, decisions about androgen replacement should be made on an individual level.

The renin-angiotensin-aldosterone system in patients with secondary adrenal insufficiency is intact. Thus, unlike patients with primary adrenal insufficiency, patients with secondary adrenal insufficiency do not need aldosterone replacement.

Patients with secondary adrenal insufficiency, just like patients with Addison disease, need higher doses of glucocorticoids during times of increased physiological and psychological stress (Fig. 5.7).

KEY POINTS: Glucocorticoid Treatment During Stress

Fever	Double or triple the glucocorticoid dose until recovery
Gastroenteritis	50–100 mg hydrocortisone intramuscularly (IM) or intravenously (IV), followed by 100–200 mg per 24 h as continuous infusion or boluses of 50 mg every 6 h, together with liberal amounts of isotonic fluids
Minor procedures (e.g., tooth extraction)	Extra glucocorticoid dose before the procedure, e.g., 20 mg hydrocortisone
Invasive procedures	50–100 mg hydrocortisone IM or IV before start of the procedure
Major surgery	50–100 mg hydrocortisone IV before anesthesia, followed by 100–200 mg per 24 hours, together with IV fluids
Acute adrenal crisis	100 mg hydrocortisone IV, followed by 100–200 mg per 24 hours as continuous infusion or boluses of 50 mg every 6 hours, together with liberal amounts of isotonic fluids

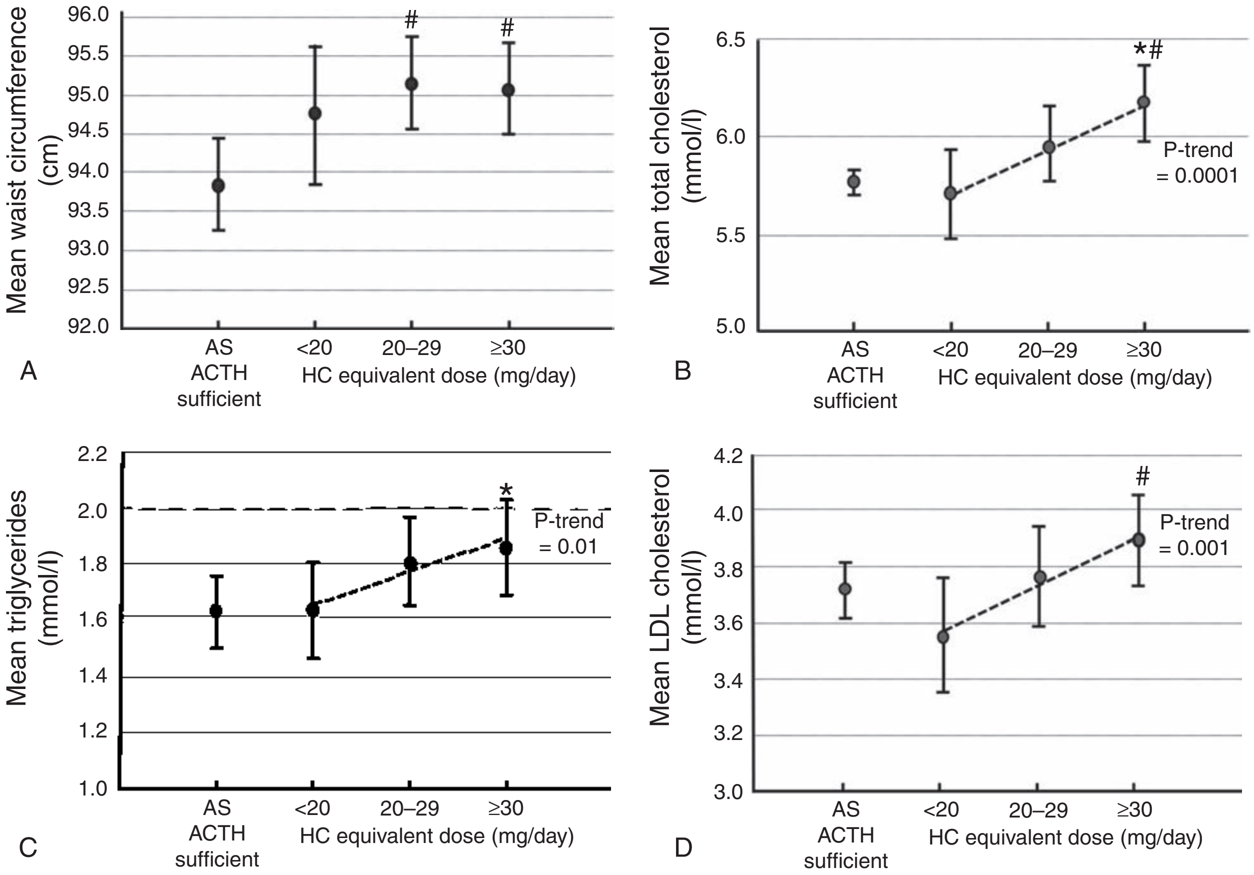


Fig. 5.5 The mean waist circumference (**A**), total cholesterol (**B**), triglycerides (**C**), and LDL-cholesterol (**D**) in adrenocorticotropic hormone (*ACTH*)-sufficient (*AS*) patients, compared with patients with secondary adrenal insufficiency receiving <20 mg, 20–29 mg, or ≥30 mg hydrocortisone per day. *HC,* hydrocortisone. (From Filipsson H, Monson JP, Koltowska-Haggstrom M, et al. The impact of glucocorticoid replacement regimens on metabolic outcome and comorbidity in hypopituitary patients. *J Clin Endocrinol Metab.* 2006;91:3954–3961. Reprinted with permission from Oxford University Press.)

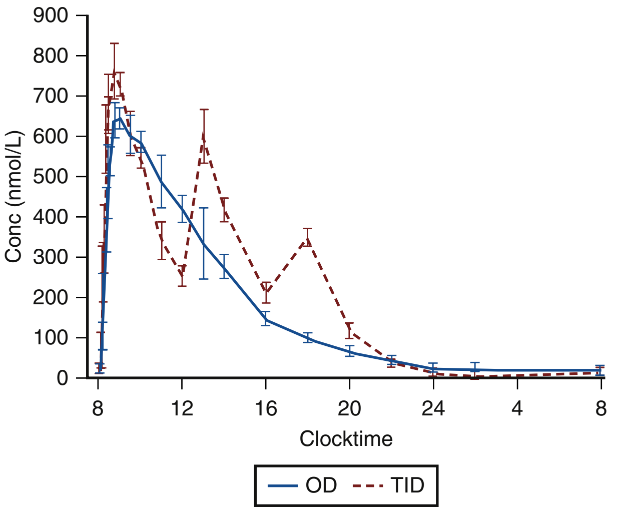
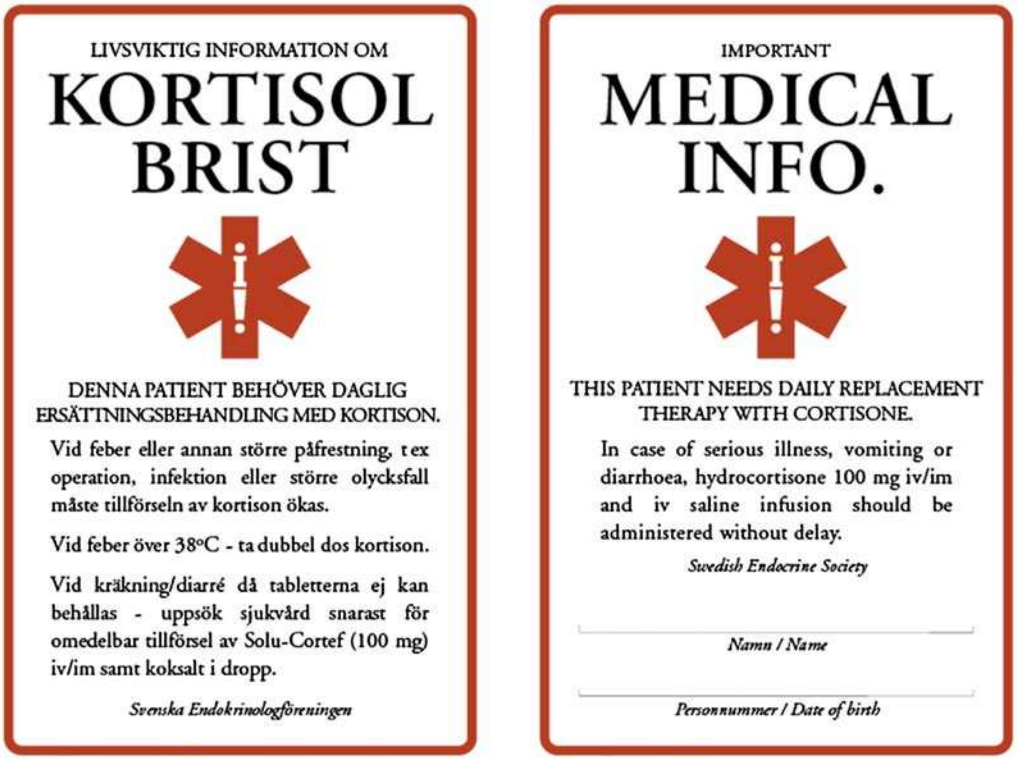


Fig. 5.6 **Mean Serum Cortisol Concentration and the 95Th Confidence Interval After Single Dose of Dual-Release Hydrocortisone *(Blue Line)* and Three Times Dosing of Conventional Hydrocortisone in 64 Patients with Primary Adrenal Insufficiency.** *OD,* Once daily; *TID,* three times daily. (From Johannsson G, Nilsson AG, Bergthorsdottir R, et al. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab.* 2012;97:473–481. Reprinted with permission from Oxford University Press.)

This is of paramount importance, because death from adrenal crisis is an important, and potentially avoidable, contributor to the increased mortality observed in patients with pituitary insufficiency.[29] A cornerstone in the management of patients with adrenal insufficiency during intercurrent illness is to prevent acute adrenal crisis by increasing the regular replacement dose in a timely manner. Although different regimes exist, doubling the glucocorticoid dose is usually sufficient, (i.e., patients receiving 10 mg hydrocortisone in the morning and 5 mg in the afternoon should take 20+10 mg instead). In patients with recurrent vomiting and/or diarrhea, where the gastrointestinal uptake of hydrocortisone is uncertain, glucocorticoids have to be administrated parenterally. Patients who present with acute adrenal crisis also need parenteral treatment with glucocorticoids. In addition, because most patents with adrenal crisis are dehydrated, abundant amounts of intravenous fluids are needed.

Thyroid-Stimulating Hormone Deficiency Central Hypothyroidism

L-thyroxine is the standard replacement therapy for patients with central hypothyroidism. Because of its long half-life (7 days), once-daily administration is sufficient, where a peak serum concentration of free T4 is reached 2 to 3 hours after ingestion.[35] L-thyroxine replacement does not mimic the normal small diurnal variations in free T4



LIVSVIKTIG INFORMATION OM

KORTISOL BRIST

DENNA PATIENT BEHÖVER DAGLIG ERSÄTTNINGSBEHANDLING MED KORTISON.

Vid feber eller annan större påfrestning, t ex operation, infektion eller större olycksfall måste tillförseln av kortison ökas.

Vid feber över 38°C - ta dubbel dos kortison.

Vid kräkning/diarré då tabletterna ej kan behållas - uppsök sjukvård snarast för omedelbar tillförsel av Solu-Cortef (100 mg) iv/im samt koksalt i dropp.

Svenska Endokrinologföreningen

IMPORTANT

MEDICAL INFO.

THIS PATIENT NEEDS DAILY REPLACEMENT THERAPY WITH CORTISONE.

In case of serious illness, vomiting or diarrhoea, hydrocortisone 100 mg iv/im and iv saline infusion should be administered without delay.

Swedish Endocrine Society

Namn / Name

Personnummer / Date of birth

Fig. 5.7 The Swedish Medical Emergency Card for Patients with Adrenal Insufficiency. Cards with the Same Design are Available in More than 30 Different Languages.

concentrations entirely, with peaks between 8:00 AM and 12:00 AM and troughs between 11:00 PM and 3:00 AM. Whether this difference is of clinical significance is not known.

Initiating L-thyroxine replacement should not be done unless adrenal insufficiency has first been ruled out. The reason is that L-thyroxine increases the metabolic clearance rate of cortisol, and thereby increases the risk of adrenal crisis in subjects with untreated adrenal insufficiency. Careful evaluation of HPA functions should therefore be done in all patients with central hypothyroidism, because the probability of coexisting adrenal insufficiency is high.

A common starting dose of L-thyroxine is 25 µg per day. However, the initial dose for the elderly and patients with cardiovascular disease should be lower. The dose of L-thyroxine should then be titrated, aiming at a serum free T4 concentration in the mid or upper part of the normal reference range. A common strategy is to increase the dose by 25 µg every 4 to 6 weeks until the goal has been reached. Again, dose titration in the elderly and subjects with cardiovascular disease should be done at a slower pace and in smaller dose steps. In prospective dose titration studies, an optimal L-thyroxine dose in subjects younger than 60 years was 1.4 to 1.7 µg/kg per day, and 1.1 to 1.4 µg/kg per day in patients older than 60 years.[65] A careful dose titration, where free T4 reaches the mid- or upper-normal reference range, is supported by studies showing beneficial effects on cardiovascular risk factors (Fig. 5.8).[66,67] Thus, to obtain optimal L-thyroxine dose, free T4 should be measured before the morning dose is taken, aiming at a concentration in the mid or upper part of the normal range, and symptoms and signs related to over- or under-treatment should be avoided. TSH, used to adjust L-thyroxine dose in patients with primary hypothyroidism, cannot, for obvious reasons, be used for patients with central hypothyroidism.

Luteinizing Hormone/Follicle-Stimulating Hormone Deficiency: Hypogonadotropic Hypogonadism

Women of reproductive age with hypogonadotropic hypogonadism should receive treatment with estrogen, and progesterone if the uterus is intact.[32,37] In many cases, a relatively high doses of estrogen are needed to relieve the postmenopausal symptoms, (e.g., 2 mg orally or 75 to 100 µg with transdermal patches). The goals with treatment are to improve quality of life, preserve bone mineral density, and prevent premature death. Women of postmenopausal age with hypogonadotropic hypogonadism do usually not require any specific replacement therapy.

Men with hypogonadotropic hypogonadism should receive replacement therapy with testosterone, either administrated intramuscularly at regular intervals, or with a transdermal gel formulation once daily, or an oral formulation that was recently approved by the US Food and Drug Administration.[68] The goals with treatment are to improve quality of life and sexual function, increase bone mineral density, muscle mass, and muscle strength, and regain normal fat distribution. Before the treatment is initiated, undiagnosed prostate cancer should be ruled out by clinical examination of the prostate gland and by measuring prostate-specific antigen. During testosterone replacement therapy, testosterone in serum should be measured at regular intervals to ensure adequate concentrations. In addition, because secondary polycythemia is the most common side effect of testosterone treatment, hematocrit should also be measured at least annually. In patients who develop polycythemia (hematocrit >48%–50%), the testosterone dose should be reduced, and other reasons such as obstructive sleep apnea should be considered.

Premenopausal women with hypogonadotropic hypogonadism who wish to become pregnant should be offered reproductive treatment,

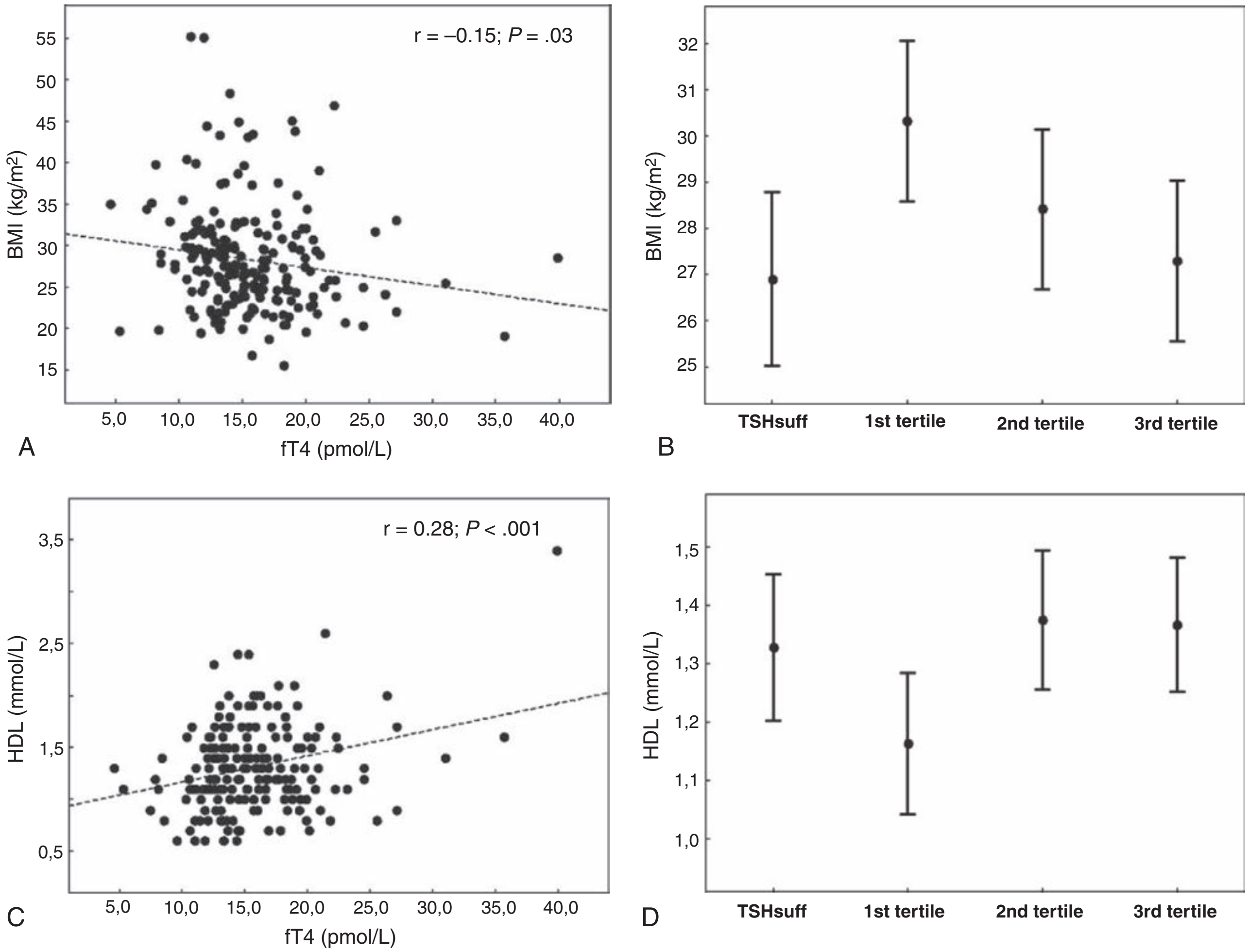


Fig. 5.8 The association between free T4 (fT4) concentrations and body mass index (*BMI*) (**A**) and high density lipoprotein (*HDL*)-cholesterol (**C**), as well as the relationship between L-thyroxine treatment and BMI (**B**) and HDL-cholesterol (**D**) in 208 patients with hypopituitarism. *TSHsuff,* Thyroid-stimulating hormone–sufficient. (From Klose M, Marina D, Hartoft-Nielsen ML, et al. Central hypothyroidism and its replacement have a significant influence on cardiovascular risk factors in adult hypopituitary patients. *J Clin Endocrinol Metab.* 2013;98:3802–3810. Reprinted with permission from Oxford University Press.)

with gonadotropins administered subcutaneously or with GnRH analogs administered via pump. Although fertility is greatly reduced in women with hypopituitarism, ovulation can be induced in up 75%.[69] Similar treatment should be offered to men with hypogonadotropic hypogonadism, where sufficient spermatogenesis can be induced in the majority of patients with hypopituitarism acquired during adulthood and successful pregnancy in approximately one-third.[70]

Growth Hormone Deficiency

Although GH has been approved for adults with GH deficiency for more than 20 years, and international guidelines recommend treatment to eligible patients, far from all patients with hypopituitarism and GH deficiency receive GH replacement therapy.[48-50] In many countries, the main reason is economical restrictions and lack of reimbursement. The reason why GH is not prescribed to GH-deficient patients in many developed countries is less clear. Nevertheless, this means that there is a selection of adult patients who receive GH replacement, which should be considered when data on efficacy and outcome are interpreted.

Most studies that have investigated the effects of GH replacement on body composition, lipid metabolism, quality of life, and general well-being indicate that patients with the most severe phenotype improve the most. There are no strong data supporting the idea that the beneficial effects of GH replacement are affected by age, gender, or underlying disease. Thus, clinically applicable scoring systems, or cutoffs, for selection of patients who theoretically benefit most from GH replacement have not yet been developed. Therefore, current guidelines state that all eligible patients with a well-defined hypothalamopituitary disease and high likelihood of GH deficiency should be tested and offered treatment if they fail the test.[48-50]

Before initiating GH treatment, the underlying condition, other pituitary hormone deficiencies, and their replacement, the risk of diabetes mellitus, and cardiometabolic status, should be carefully evaluated. GH replacement should be initiated with a low daily dose, and thereafter dose-titrated according to age-adjusted serum IGF-1 concentrations and the clinical response. A higher starting dose and a faster dose titration can be used in the young. The reason is that young patients, particularly young women using oral estrogen, are less sensitive to GH and, therefore, less prone to develop dose-related side effects such as arthralgia, muscle stiffness, and peripheral edema caused by the salt- and water-retaining effects of GH and IGF-1. Typically, the dose titration takes 3 to 12 months, depending on the patient's sensitivity to GH. Because many adults with GH deficiency have a baseline serum IGF-1 that is already in the low-normal range, the target is usually a serum IGF-1 concentration in the high-normal range. Once a stable dose has been reached, IGF-1 should be measured once or twice yearly, at the same time as the clinical response, side effects, and safety are evaluated, including the effect on glucose metabolism.

Both randomized controlled trials and open longitudinal studies have demonstrated various beneficial and sustained effects of GH replacement in adults.[71,72] The most robust effects are on body composition, with reduced fat mass, in particular reduced abdominal fat, increased lean body mass, and increased extracellular water, (i.e., changes toward a normal body composition). GH has also positive effects on bone mineral density. After a temporary reduction during the first year after treatment is started, bone mineral density gradually increases.[73] Other beneficial effects are increased muscle mass and muscle strength,[74] as well as increased cardiac output and exercise capacity,[42,75] (i.e., effects that probably explain the self-reported increase in energy level, vitality, and well-being). In a prospective randomized crossover trial, quality of life improved during GH replacement, as reported by both the patients and their partners.[76] In another placebo-controlled trial where adults receiving chronic GH replacement for more than 10 years were randomized in a blinded fashion to either receive placebo or continued GH replacement, quality of life deteriorated during placebo treatment.[77]

The safety with GH replacement has been extensively studied, showing that the risk for *de novo* cancer or recurrence of an underlying pituitary tumor and cardiovascular risk is not increased.[78] Although an increased risk for developing diabetes mellitus during the initial phase of GH replacement was found in one study, other studies have not shown an increased risk. Diabetes is, therefore, not considered to be a contraindication for GH treatment. Similarly, having a residual pituitary tumor after surgery is not a contraindication, and tumor surveillance should not be influenced by treatment with GH.

One of the obstacles with GH is the daily subcutaneous injections that some patients consider unappealing and prevent them from initiating treatment. Although treatment adherence is high during the first year of treatment, it decreases markedly thereafter, with negative effect on outcome. Long-acting GH products allowing once-weekly injections have therefore been developed, mainly to improve convenience for the patients, increase the number of patients willing to start GH replacement, and improve adherence during long-term replacement.[79] Currently available data indicate that the efficacy and safety profile of long-acting GH is similar to daily GH injections.

Childhood-Onset Growth Hormone Deficiency in Adults

Approximately 10% to 20% of adults with GH deficiency are diagnosed during childhood. The main aim of GH replacement during childhood is to obtain normal growth rate and normal final height. In children with coexisting deficiencies of other pituitary hormones, careful glucocorticoid and thyroid hormone replacement is mandatory, as well as careful puberty induction, to obtain normal final height. When final height has been reached, endogenous GH production should be reevaluated in all patients who have received GH replacement during childhood, as only a minority remain GH-deficient in adulthood[80] (see also Chapter 18).

How patients with childhood-onset GH deficiency should be retested depends on the underlying etiology and how high the probability of continuing chronic GH deficiency is. Patients who likely remain GH-deficient during adulthood are those who are diagnosed with severe GH deficiency during childhood, patients with two or more additional pituitary hormone deficits, patients with structural hypothalamopituitary abnormalities, patients with central nervous system tumors, or patients who have received high-dose cranial irradiation, as well as patients with well-defined genetic causes. Reevaluation in patients with high probability of chronic GH deficiency can be performed by discontinuing GH replacement for a month and thereafter measuring serum IGF-1. If the IGF-1 standard deviation score falls below –2, GH deficiency is confirmed. In patients with a low probability of permanent GH deficiency, serum IGF-1 should be measured, and a GH stimulation test should be performed. The choice of test should depend on the experience of the center and if there is a need to test the HPA axis at the same time. Subjects with continuing GH deficiency should be offered restart of GH replacement.[48-50,80]

Hormone–Hormone Interactions

Most patients with hypopituitarism require replacement therapy with two or more hormones. Some of these interact with each other, (i.e., administration of one hormone may influence the concentration of another hormones).[47,81] In many cases this can be of clinical significance, and adequate dose adjustments are required.

KEY POINTS: Hormone–Hormone Interactions

Thyroxine–Cortisol	• Treatment with thyroxine increases the metabolism of cortisol • Initiation of thyroxine replacement in patients with untreated adrenal insufficiency can therefore induce adrenal crisis
Growth hormone (GH)–Cortisol	• GH inhibits 11ß–hydroxysteroid dehydrogenase type 1 activity • Initiation of GH replacement may induce overt adrenal insufficiency in untreated patients • Glucocorticoid dose may need to be increased in patients receiving a relatively low replacement dose when GH replacement is initiated
GH–Thyroxine	• GH reduces free T4 concentrations • Presence of central hypothyroidism may first become evident after GH replacement is initiated
Estrogen–Cortisol	• Estrogen increases production of cortisone-binding globulin, and subsequently concentrations of total cortisol in serum • Basal and stimulated cortisol in serum may be falsely high
Estrogen–GH/ Insulin-like growth factor 1 (IGF-1)	• Oral, but not transdermal, estrogen replacement attenuates the IGF-1 response of GH replacement • Higher GH doses are need in women on oral estrogen treatment

PREGNANCY

Extensive structural changes occur in the pituitary gland during pregnancy, mainly because of the dramatically increased prolactin production. Consequently, a hyperplasia of the lactotropic cells occurs, and the size of the pituitary gland increases by at least 2-fold. The biochemical alterations are also substantial and must be considered when pregnant patients with hypopituitarism are being monitored.[32]

KEY POINTS: Management of Women with Hypopituitarism During Pregnancy

Adrenocorticotropic hormone deficiency	• Exclude secondary adrenal insufficiency in all women with hypopituitarism before pregnancy • Optimize glucocorticoid doses before conception • Avoid long-acting synthetic steroids like dexamethasone that are not inactivated in the placenta • Increase the glucocorticoid replacement dose in women who experience symptomatic hypocortisolemia during pregnancy • Administer stress-doses of glucocorticoids during labor and delivery
Thyroid-stimulating hormone (TSH) deficiency	• Optimize the thyroxine dose before pregnancy by clinical evaluation and by measuring free T4, which should be in the mid-upper range of normal • TSH concentrations cannot be used to guide adjustments of the thyroxine dose in women with central hypothyroidism • Consider 25% increment of the thyroxine dose when pregnancy has been confirmed • Measure free T4 concentrations every 4–6 weeks and adjust the thyroxine dose when needed according to clinical findings and pregnancy-specific reference intervals
Growth hormone (GH) deficiency	• The safety of GH replacement during pregnancy has not been studied sufficiently • Available observations have, however, not demonstrated any difference concerning complications, or number of live births, in women who continue with GH during the first and second trimester, or throughout pregnancy, compared with those who discontinue • Based on the physiologic GH placental production, continued GH replacement should be considered during the first trimester, gradual tapering of the dose during the second trimester, and discontinuation thereafter

Serum concentrations of (total) cortisol increase during pregnancy, mainly because of the estrogen-driven increased production of corticosteroid-binding globulin.[82] Measurement of cortisol in serum has therefore a limited value in the diagnostic work-up of pregnant women suspected to have adrenal insufficiency. Concentrations of urinary free cortisol increase as well, indicating a slightly increase endogenous production during pregnancy. Nevertheless, most women with secondary adrenal insufficiency do not have to increase their regular glucocorticoid dose during pregnancy. However, some women may experience symptomatic hypocortisolemia, most commonly during the third trimester, where the dose should be increased, in particular in women on low daily doses. Long-acting synthetic steroids that are not inactivated by placental 11β–hydroxisteroid dehydrogenase type 2, like dexamethasone, should be avoided during pregnancy. Instead, short-acting glucocorticoids such as hydrocortisone are preferable. During labor and delivery, all women should receive stress-doses of glucocorticoids.

Human chorionic gonadotropin (hCG) increases during pregnancy. The chemical structure of hCG is similar to that of TSH, which allows hCG to stimulate the TSH receptor. Therefore, during the first trimester of pregnancy, the thyroid gland is mainly stimulated by hCG, and TSH concentrations are frequently low.[82] Similarly, free T4 concentrations are often high in early pregnancy and gradually decrease thereafter. In fact, free T4 concentrations during the third trimester of pregnancy are significantly lower than in nonpregnant women. TSH concentrations cannot be used to guide adjustments of the thyroxine dose in women with central hypothyroidism. Instead, the dose should be optimized before pregnancy by clinical evaluation and by measuring free T4, which should be in the mid-upper range of normal. Because the thyroid gland in patients with central hypothyroidism is intact, and hCG is the main stimulus of thyroxine production, it is uncertain whether the dose should be increased directly when pregnancy is confirmed in the same way as in patients with primary hypothyroidism. Nevertheless, a 25% increment of the thyroxine dose in early pregnancy is a common, and most probably safe, practice.[35] Thereafter, free T4 concentrations should be measured every 4 to 6 weeks and further dose adjustments performed when needed according to clinical findings and pregnancy-specific reference intervals.

Gradually increasing amounts of placental GH are produced in the placenta throughout pregnancy.[82] Increased IGF-1 concentrations are, therefore, commonly observed in the healthy pregnant woman. GH replacement seems to be safe during pregnancy. In a prospective multicenter study on women with GH deficiency, including 173 who became pregnant, no difference concerning complications or number of live births were observed between women who continued with GH throughout pregnancy compared with those who discontinued.[69] However, because data on safety are still limited, current clinical guidelines recommend that GH replacement should be discontinued during pregnancy in women with GH deficiency. Nevertheless, at some centers, based on physiologic GH placental production, GH replacement is continued during the first trimester, gradually tapered during the second trimester, and thereafter discontinued.

PATIENT CARE–MULTIDISCIPLINARY APPROACH AND PATIENT EDUCATION

To achieve the best possible treatment outcome for the individual patient with hypopituitarism, a multidisciplinary approach is necessary, including a team of experts experienced in diagnosing and treating the underlying disorder, as well as experience in managing hormone deficiencies. The majority of patients with hypopituitarism have multiple pituitary hormone deficiencies. Although some replacement therapies can be guided by serum markers such as free T4, testosterone, and serum IGF-1, the exact target for each individual patient can be difficult to determine. In addition, for other replacement therapies, such as glucocorticoid replacement, a useful biomarker for guidance of optimal dosing is lacking. Thus, adjustment of the replacement doses should be done in close collaboration between the patient and the treating physician, and symptoms of over- or underreplacement should be evaluated regularly. Indeed, patient education and increased knowledge of hypopituitarism and cardiometabolic risk factors among patients and healthcare providers, as well as the introduction of adult GH replacement therapy, are probably some of the most important explanations for the seemingly improved outcome recently reported in patients with hypopituitarism.[37,78]

REFERENCES

1. Fröhlich A. Ein fall von tumor der hypophysis cerebri ohne akromegalie. *Wien Klin Rundschau*. 1901;15:883–886.
2. Babinski J. Tumeur du corps pituitaire sans acromégalie et avec arrêt de développement des organes génitaux. *Rev Neurol*. 1900;8:531–533.
3. Tjornstrand A, Gunnarsson K, Evert M, et al. The incidence rate of pituitary adenomas in western Sweden for the period 2001-2011. *Eur J Endocrinol*. 2014;171:519–526.

4. Drange MR, Fram NR, Herman-Bonert V, et al. Pituitary tumor registry: a novel clinical resource. *J Clin Endocrinol Metab*. 2000;85:168–174.
5. Nomikos P, Ladar C, Fahlbusch R, et al. Impact of primary surgery on pituitary function in patients with non-functioning pituitary adenomas -- a study on 721 patients. *Acta Neurochir (Wien)*. 2004;146:27–35.
6. Webb SM, Rigla M, Wagner A, et al. Recovery of hypopituitarism after neurosurgical treatment of pituitary adenomas. *J Clin Endocrinol Metab*. 1999;84:3696–3700.
7. Muller HL. The diagnosis and treatment of craniopharyngioma. *Neuroendocrinology*. 2020;110:753–766.
8. Schill F, Nilsson M, Olsson DS, et al. Pituitary metastases: a nationwide study on current characteristics with special reference to breast cancer. *J Clin Endocrinol Metab*. 2019;104:3379–3388.
9. Trifanescu R, Ansorge O, Wass JA, et al. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*. 2012;76:151–160.
10. Diri H, Tanriverdi F, Karaca Z, et al. Extensive investigation of 114 patients with Sheehan's syndrome: a continuing disorder. *Eur J Endocrinol*. 2014;171:311–318.
11. Briet C, Salenave S, Bonneville JF, et al. Pituitary apoplexy. *Endocr Rev*. 2015;36:622–645.
12. Joshi MN, Whitelaw BC, Carroll PV. Mechanisms in Endocrinology: Hypophysitis: diagnosis and treatment. *Eur J Endocrinol*. 2018;179: R151–R163.
13. Chiloiro S, Capoluongo ED, Tartaglione T, et al. The changing clinical spectrum of hypophysitis. *Trends Endocrinol Metab*. 2019;30:590–602.
14. de Filette J, Andreescu CE, Cools F, et al. A systematic review and meta-analysis of endocrine-related adverse events associated with immune checkpoint inhibitors. *Horm Metab Res*. 2019;51:145–156.
15. Barroso-Sousa R, Barry WT, Garrido-Castro AC, et al. Incidence of endocrine dysfunction following the use of different immune checkpoint inhibitor regimens: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4:173–182.
16. Sherlock M, Ayuk J, Tomlinson JW, et al. Mortality in patients with pituitary disease. *Endocr Rev*. 2010;31:301–342.
17. Rosen T, Bengtsson BA. Premature mortality due to cardiovascular disease in hypopituitarism. *Lancet*. 1990;336:285–288.
18. Tomlinson JW, Holden N, Hills RK, et al. Association between premature mortality and hypopituitarism. West Midlands Prospective Hypopituitary Study Group. *Lancet*. 2001;357:425–431.
19. Pappachan JM, Raskauskiene D, Kutty VR, et al. Excess mortality associated with hypopituitarism in adults: a meta-analysis of observational studies. *J Clin Endocrinol Metab*. 2015;100:1405–1411.
20. Olsson DS, Nilsson AG, Bryngelsson IL, et al. Excess mortality in women and young adults with nonfunctioning pituitary adenoma: a Swedish nationwide study. *J Clin Endocrinol Metab*. 2015;100:2651–2658.
21. Sherlock M, Reulen RC, Alonso AA, et al. ACTH deficiency, higher doses of hydrocortisone replacement, and radiotherapy are independent predictors of mortality in patients with acromegaly. *J Clin Endocrinol Metab*. 2009;94:4216–4223.
22. Zueger T, Kirchner P, Herren C, et al. Glucocorticoid replacement and mortality in patients with nonfunctioning pituitary adenoma. *J Clin Endocrinol Metab*. 2012;97:E1938–E1942.
23. Olivius C, Landin-Wilhelmsen K, Olsson DS, et al. Prevalence and treatment of central hypogonadism and hypoandrogenism in women with hypopituitarism. *Pituitary*. 2018;21:445–453.
24. Filipsson H, Monson JP, Koltowska-Haggstrom M, et al. The impact of glucocorticoid replacement regimens on metabolic outcome and comorbidity in hypopituitary patients. *J Clin Endocrinol Metab*. 2006;91:3954–3961.
25. O'Reilly MW, Reulen RC, Gupta S, et al. ACTH and gonadotropin deficiencies predict mortality in patients treated for nonfunctioning pituitary adenoma: long-term follow-up of 519 patients in two large European centres. *Clin Endocrinol (Oxf)*. 2016;85:748–756.
26. Hammarstrand C, Ragnarsson O, Hallen T, et al. Higher glucocorticoid replacement doses are associated with increased mortality in patients with pituitary adenoma. *Eur J Endocrinol*. 2017;177:251–256.
27. Stewart PM, Biller BM, Marelli C, et al. Exploring inpatient hospitalizations and morbidity in patients with adrenal insufficiency. *J Clin Endocrinol Metab*. 2016;101:4843–4850.
28. Smans LC, Van der Valk ES, Hermus AR, et al. Incidence of adrenal crisis in patients with adrenal insufficiency. *Clin Endocrinol (Oxf)*. 2016;84:17–22.
29. Burman P, Mattsson AF, Johannsson G, et al. Deaths among adult patients with hypopituitarism: hypocortisolism during acute stress, and de novo malignant brain tumors contribute to an increased mortality. *J Clin Endocrinol Metab*. 2013;98:1466–1475.
30. Clayton PE, Shalet SM. Dose dependency of time of onset of radiation-induced growth hormone deficiency. *J Pediatr*. 1991;118:226–228.
31. Bancos I, Hahner S, Tomlinson J, et al. Diagnosis and management of adrenal insufficiency. *Lancet Diabetes Endocrinol*. 2015;3:216–226.
32. Fleseriu M, Hashim IA, Karavitaki N, et al. Hormonal replacement in hypopituitarism in adults: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:3888–3921.
33. Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:364–389.
34. Ospina NS, Al Nofal A, Bancos I, et al. ACTH stimulation tests for the diagnosis of adrenal insufficiency: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2016;101:427–434.
35. Persani L, Brabant G, Dattani M, et al. 2018 European Thyroid Association (ETA) guidelines on the diagnosis and management of central hypothyroidism. *Eur Thyroid J*. 2018;7:225–237.
36. Agha A, Walker D, Perry L, et al. Unmasking of central hypothyroidism following growth hormone replacement in adult hypopituitary patients. *Clin Endocrinol (Oxf)*. 2007;66:72–77.
37. Higham CE, Johannsson G, Shalet SM. Hypopituitarism. *Lancet*. 2016;388:2403–2415.
38. Moller N, Jorgensen JO. Effects of growth hormone on glucose, lipid, and protein metabolism in human subjects. *Endocr Rev*. 2009;30:152–177.
39. Rabinowitz D, Zierler KL. A metabolic regulating device based on the actions of human growth hormone and of insulin, singly and together, on the human forearm. *Nature*. 1963;199:913–915.
40. Angelin B, Rudling M. Growth hormone and hepatic lipoprotein metabolism. *Curr Opin Lipidol*. 1994;5:160–165.
41. Ohlsson C, Bengtsson BA, Isaksson OG, et al. Growth hormone and bone. *Endocr Rev*. 1998;19:55–79.
42. Boger RH, Skamira C, Bode-Boger SM, et al. Nitric oxide may mediate the hemodynamic effects of recombinant growth hormone in patients with acquired growth hormone deficiency. A double-blind, placebo-controlled study. *J Clin Invest*. 1996;98:2706–2713.
43. Johansson JO, Larson G, Andersson M, et al. Treatment of growth hormone-deficient adults with recombinant human growth hormone increases the concentration of growth hormone in the cerebrospinal fluid and affects neurotransmitters. *Neuroendocrinology*. 1995;61:57–66.
44. Johannsson G. Management of adult growth hormone deficiency. *Endocrinol Metab Clin North Am*. 2007;36:203–220.
45. Sverrisdottir YB, Elam M, Herlitz H, et al. Intense sympathetic nerve activity in adults with hypopituitarism and untreated growth hormone deficiency. *J Clin Endocrinol Metab*. 1998;83:1881–1885.
46. Koltowska-Haggstrom M, Mattsson AF, Monson JP, et al. Does long-term GH replacement therapy in hypopituitary adults with GH deficiency normalise quality of life? *Eur J Endocrinol*. 2006;155:109–119.
47. Filipsson H, Johannsson G. GH replacement in adults: interactions with other pituitary hormone deficiencies and replacement therapies. *Eur J Endocrinol*. 2009;161:S85–S95.
48. Yuen KCJ, Biller BMK, Radovick S, et al. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for management of growth hormone deficiency in adults and patients transitioning from pediatric to adult care. *Endocr Pract*. 2019;25:1191–1232.
49. Molitch ME, Clemmons DR, Malozowski S, et al. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1587–1609.
50. Ho KK. 2007 GH Deficiency Consensus Workshop Participants. Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, Japan Endocrine Society, and Endocrine Society of Australia. *Eur J Endocrinol*. 2007;157:695–700.

51. Hoffman DM, O'Sullivan AJ, Baxter RC, et al. Diagnosis of growth-hormone deficiency in adults. *Lancet*. 1994;343:1064–1068.
52. Corneli G, Di Somma C, Baldelli R, et al. The cut-off limits of the GH response to GH-releasing hormone-arginine test related to body mass index. *Eur J Endocrinol*. 2005;153:257–264.
53. Darzy KH, Aimaretti G, Wieringa G, et al. The usefulness of the combined growth hormone (GH)-releasing hormone and arginine stimulation test in the diagnosis of radiation-induced GH deficiency is dependent on the post-irradiation time interval. *J Clin Endocrinol Metab*. 2003;88:95–102.
54. Garcia JM, Biller BMK, Korbonits M, et al. Macimorelin as a diagnostic test for adult GH deficiency. *J Clin Endocrinol Metab*. 2018;103:3083–3093.
55. Esposito D, Olsson DS, Ragnarsson O, et al. Non-functioning pituitary adenomas: indications for pituitary surgery and post-surgical management. *Pituitary*. 2019;22:422–434.
56. Ekstrand E, Esposito D, Ragnarsson O, et al. Metabolic effects of cortisone acetate vs hydrocortisone in patients with secondary adrenal insufficiency. *J Endocr Soc*. 2020;4:bvaa160.
57. Esteban NV, Loughlin T, Yergey AL, et al. Daily cortisol production rate in man determined by stable isotope dilution/mass spectrometry. *J Clin Endocrinol Metab*. 1991;72:39–45.
58. Kerrigan JR, Veldhuis JD, Leyo SA, et al. Estimation of daily cortisol production and clearance rates in normal pubertal males by deconvolution analysis. *J Clin Endocrinol Metab*. 1993;76:1505–1010.
59. Ragnarsson O, Mattsson AF, Monson JP, et al. The relationship between glucocorticoid replacement and quality of life in 2737 hypopituitary patients. *Eur J Endocrinol*. 2014;171:571–579.
60. Johannsson G, Nilsson AG, Bergthorsdottir R, et al. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab*. 2012;97:473–481.
61. Isidori AM, Venneri MA, Graziadio C, et al. Effect of once-daily, modified-release hydrocortisone versus standard glucocorticoid therapy on metabolism and innate immunity in patients with adrenal insufficiency (DREAM): a single-blind, randomised controlled trial. *Lancet Diabetes Endocrinol*. 2018;6:173–185.
62. Arlt W, Callies F, van Vlijmen JC, et al. Dehydroepiandrosterone replacement in women with adrenal insufficiency. *N Engl J Med*. 1999;341: 1013–1020.
63. Johannsson G, Burman P, Wiren L, et al. Low dose dehydroepiandrosterone affects behavior in hypopituitary androgen-deficient women: a placebo-controlled trial. *J Clin Endocrinol Metab*. 2002;87:2046–2052.
64. Wierman ME, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:3489–3510.
65. Ferretti E, Persani L, Jaffrain-Rea ML, et al. Evaluation of the adequacy of levothyroxine replacement therapy in patients with central hypothyroidism. *J Clin Endocrinol Metab*. 1999;84:924–929.
66. Filipsson Nystrom H, Feldt-Rasmussen U, Kourides I, et al. The metabolic consequences of thyroxine replacement in adult hypopituitary patients. *Pituitary*. 2012;15:495–504.
67. Klose M, Marina D, Hartoft-Nielsen ML, et al. Central hypothyroidism and its replacement have a significant influence on cardiovascular risk factors in adult hypopituitary patients. *J Clin Endocrinol Metab*. 2013;98:3802–3810.
68. Swerdloff RS, Wang C, White WB, et al. A new oral testosterone undecanoate formulation restores testosterone to normal concentrations in hypogonadal men. *J Clin Endocrinol Metab*. 2020;105:2515–2531.
69. Vila G, Fleseriu M. Fertility and pregnancy in women with hypopituitarism: a systematic literature review. *J Clin Endocrinol Metab*. 2020;105:dgz112.
70. Liu PY, Baker HW, Jayadev V, et al. Induction of spermatogenesis and fertility during gonadotropin treatment of gonadotropin-deficient infertile men: predictors of fertility outcome. *J Clin Endocrinol Metab*. 2009;94:801–808.
71. Maison P, Griffin S, Nicoue-Beglah M, et al. Impact of growth hormone (GH) treatment on cardiovascular risk factors in GH-deficient adults: a metaanalysis of blinded, randomized, placebo-controlled trials. *J Clin Endocrinol Metab*. 2004;89:2192–2199.
72. Gotherstrom G, Bengtsson BA, Bosaeus I, et al. A 10-year, prospective study of the metabolic effects of growth hormone replacement in adults. *J Clin Endocrinol Metab*. 2007;92:1442–1445.
73. Johannsson G, Rosen T, Bosaeus I, et al. Two years of growth hormone (GH) treatment increases bone mineral content and density in hypopituitary patients with adult-onset GH deficiency. *J Clin Endocrinol Metab*. 1996;81:2865–2873.
74. Widdowson WM, Gibney J. The effect of growth hormone (GH) replacement on muscle strength in patients with GH-deficiency: a meta-analysis. *Clin Endocrinol (Oxf)*. 2010;72:787–792.
75. Widdowson WM, Gibney J. The effect of growth hormone replacement on exercise capacity in patients with GH deficiency: a metaanalysis. *J Clin Endocrinol Metab*. 2008;93:4413–4417.
76. Burman P, Broman JE, Hetta J, et al. Quality of life in adults with growth hormone (GH) deficiency: response to treatment with recombinant human GH in a placebo-controlled 21-month trial. *J Clin Endocrinol Metab*. 1995;80:3585–3590.
77. Filipsson Nystrom H, Barbosa EJ, Nilsson AG, et al. Discontinuing long-term GH replacement therapy—a randomized, placebo-controlled crossover trial in adult GH deficiency. *J Clin Endocrinol Metab*. 2012;97:3185–3195.
78. Stochholm K, Johannsson G. Reviewing the safety of GH replacement therapy in adults. *Growth Horm IGF Res*. 2015;25:149–157.
79. Allen DB, Backeljauw P, Bidlingmaier M, et al. GH safety workshop position paper: a critical appraisal of recombinant human GH therapy in children and adults. *Eur J Endocrinol*. 2016;174:P1–P9.
80. Clayton PE, Cuneo RC, Juul A, et al. Consensus statement on the management of the GH-treated adolescent in the transition to adult care. *Eur J Endocrinol*. 2005;152:165–170.
81. Giavoli C, Libe R, Corbetta S, et al. Effect of recombinant human growth hormone (GH) replacement on the hypothalamic-pituitary-adrenal axis in adult GH-deficient patients. *J Clin Endocrinol Metab*. 2004;89:5397–5401.
82. Karaca Z, Tanriverdi F, Unluhizarci K, et al. Pregnancy and pituitary disorders. *Eur J Endocrinol*. 2010;162:453–475.

6

Acromegaly

Shlomo Melmed

OUTLINE

Pathogenesis, 81
  Excess Growth Hormone Secretion, 81
  Excess Growth Hormone–Releasing Hormone Secretion, 83
  Role of the Hypothalamus in the Etiology of Acromegaly, 83
  Intrinsic Pituitary Lesions, 83
  Candidate Genes in the Etiology of Acromegaly, 84
  Epidemiology, 85
Diagnosis, 85
  Documenting Growth Hormone Hypersecretion, 85
  Localizing the Source of Excess Growth Hormone, 85
  Clinical Manifestations, 86
  Laboratory Findings, 88
  Mortality, 88
Management, 88
  Treatment Goals, 88
  Surgery, 89
  Pituitary Irradiation, 90
  Pharmacologic Management (Fig. 6.5), 90
Integrated Treatment Approach to the Management of Acromegaly, 94
  Patients With Likelihood of Good Surgical Outcome, 94
  Patients With a Poor Likelihood of Successful Surgical Outcome, 94
  Follow-Up, 94
  Additional Management Considerations, 95

Acromegaly is a disease of spectacular growth and metabolic disorders that has fascinated physicians for centuries. The natural history of the disorder, if left untreated, results in gross acral and facial disfigurement, musculoskeletal disability, cardiac failure, respiratory dysfunction, diabetes, and accelerated mortality.[1,2] If the disease occurs before epiphyseal closure, gigantism results.[1] After the first modern description of the disease in 1886 by Marie, it was subsequently recognized that the disorder is associated with a growth hormone (GH)-secreting adenohypophyseal adenoma, resulting in both a central mass lesion and the protean peripheral effects of sustained tissue exposure to high GH levels.[1,3]

KEY POINTS

- Acromegaly is usually caused by a GH-secreting pituiatry adenoma.

PATHOGENESIS

The pathogenetic events that underlie the etiology of pituitary acromegaly include excessive pituitary somatotroph cell proliferation and unrestrained GH hypersecretion (Fig. 6.1).[4] GH is secreted by somatotroph cells, the largest differentiated compartment of the anterior pituitary. GH secretion is under dual hypothalamic inhibitory control: somatotropin release–inhibiting factor (SRIF) inhibits secretion, and GH–releasing hormone (GHRH) stimulates both GH synthesis and secretion. Ghrelin, a gut-derived peptide, binds to the GH secretagogue receptor and acts primarily at the hypothalamus to induce GH release. Insulin-like growth factor 1 (IGF-1), the peripheral target molecule for GH action, participates in negative GH feedback inhibition by acting both at the hypothalamus to induce SRIF and directly at the pituitary to inhibit *GH* gene transcription.[5] Peripheral sex and adrenal steroids also regulate GH secretion.[6] GH binds to peripheral GH receptors that elicit signaling by JAK/STAT (Janus kinase/signal transducer and activator of transcription) intracellular phosphorylation cascades.[7] GH acts directly to attenuate insulin action and induce lipolysis. Most growth-promoting actions of GH are indirectly mediated by IGF-1, which is synthesized in the liver, kidney, pituitary, gastrointestinal tract, muscle, and cartilage.[7] GH actions mediated by IGF-1 include protein synthesis; amino-acid transportation; muscle, cartilage, and bone growth; DNA and RNA synthesis; and cell proliferation.[8] Local production of IGF-1 may be under autocrine and paracrine regulation, acting in concert with circulating IGF-1 and GH to elicit a final tissue impact.

Excess Growth Hormone Secretion

Tumors may arise from clonal expansion of one or more of the anterior pituitary differentiated cell types[9] and thereby result in specific hormone hypersecretory syndromes.[10] The most common cause of acromegaly is a somatotroph (GH-secreting) adenoma of the anterior pituitary, which account for 30% of all hormone-secreting pituitary adenomas (Fig. 6.2). GH-secreting adenomas arise from differentiated cells secreting *GH* gene products[1,9] (Table 6.1). These cells include somatotrophs, mixed mammosomatotrophs (secreting both GH and prolactin [PRL]), or more primitive acidophilic stem cells. Regardless of their cellular origin, transformation and subsequent replication of these cells result in adenoma formation as well as unrestrained GH secretion.[1] Most patients harbor densely granulated GH cell adenomas, which are commonly encountered in older patients with indolent disease progression.[9] Sparsely granulated GH cell adenomas occur in younger patients with more aggressive disease onset and higher GH levels.[1,2,11] Mammosomatotroph cell tumors, or discrete mammotroph and somatotroph tumors, reflect the common stem cell origin of the somatotroph cell lineage. Although acidophil

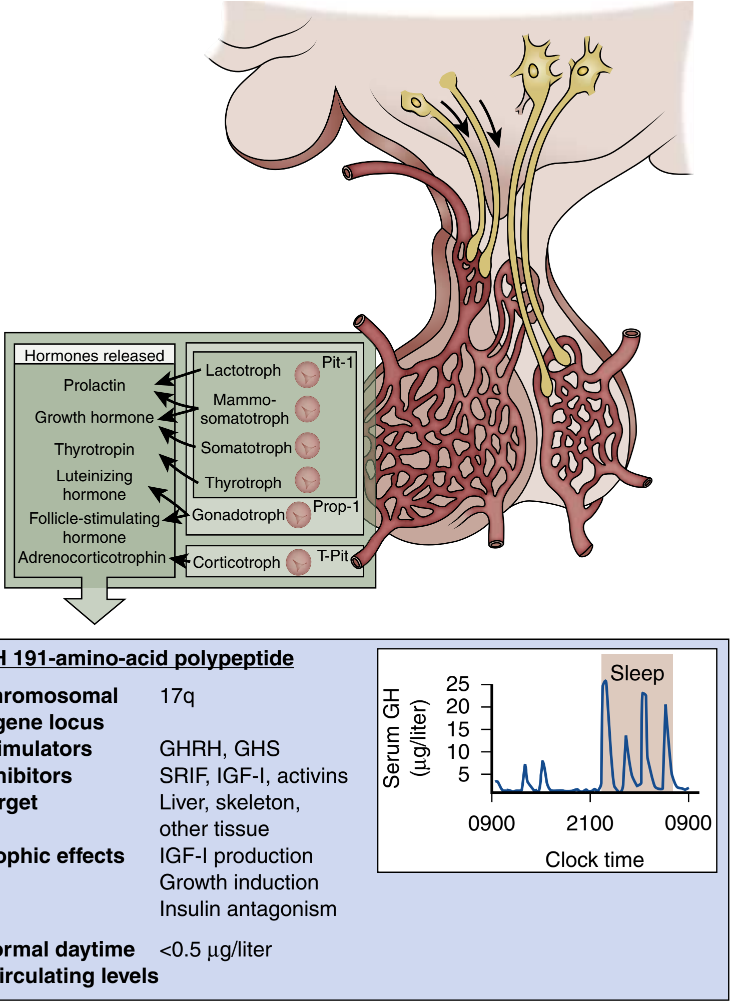

Hormones released
Prolactin
Growth hormone
Thyrotropin
Luteinizing hormone
Follicle-stimulating hormone
Adrenocorticotrophin
Lactotroph
Mammo-somatotroph
Somatotroph
Thyrotroph
Pit-1
Gonadotroph
Prop-1
Corticotroph
T-Pit
GH 191-amino-acid polypeptide
Chromosomal gene locus 17q
Stimulators GHRH, GHS
Inhibitors SRIF, IGF-I, activins
Target Liver, skeleton, other tissue
Trophic effects IGF-I production
Growth induction
Insulin antagonism
Normal daytime circulating levels <0.5 µg/liter
Sleep
Serum GH (µg/liter)
25
20
15
10
5
0900
2100
0900
Clock time

Fig. 6.1 Synthesis and secretion of growth hormone by the anterior pituitary gland. *GHS,* Growth hormone secretagogue; *GHRH*, growth hormone–releasing hormone; *IGF-1*, insulin-like growth factor 1; *SRIF*, somatotropin release–inhibiting factor. (Modified from Melmed S. Acromegaly. *N Engl J Med.* 2006; 355:2558-2573.)

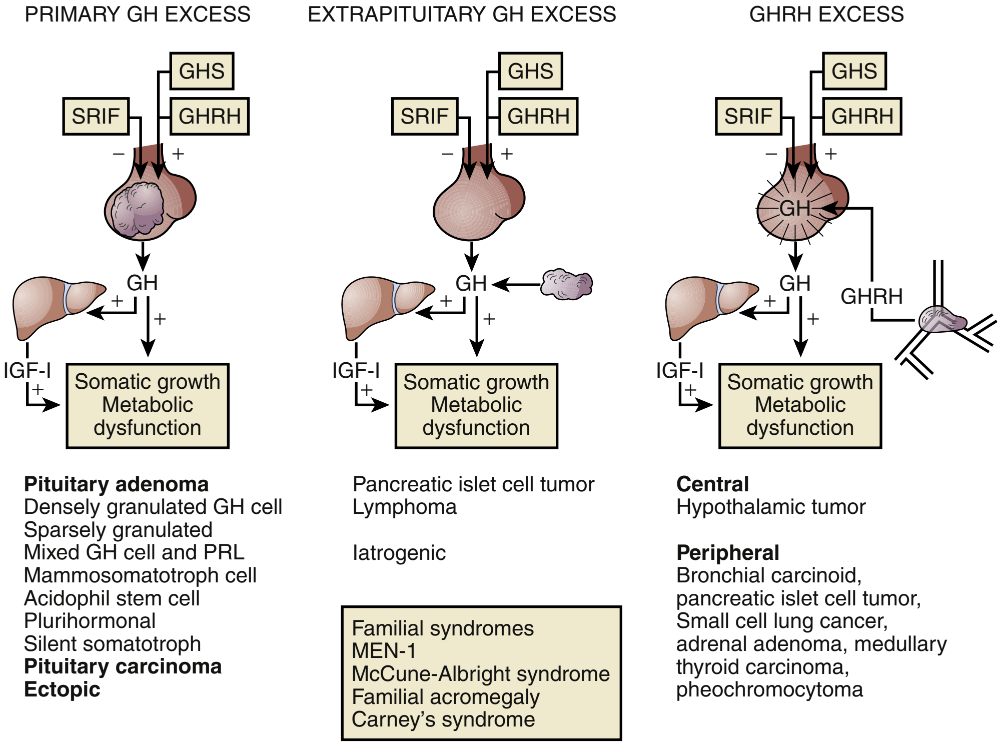

PRIMARY GH EXCESS
EXTRAPITUITARY GH EXCESS
GHRH EXCESS
GHS
SRIF
GHRH
−
+
GH
IGF-I
Somatic growth
Metabolic
dysfunction
GHRH
Pituitary adenoma
Densely granulated GH cell
Sparsely granulated
Mixed GH cell and PRL
Mammosomatotroph cell
Acidophil stem cell
Plurihormonal
Silent somatotroph
Pituitary carcinoma
Ectopic
Pancreatic islet cell tumor
Lymphoma
Iatrogenic
Familial syndromes
MEN-1
McCune-Albright syndrome
Familial acromegaly
Carney's syndrome
Central
Hypothalamic tumor
Peripheral
Bronchial carcinoid,
pancreatic islet cell tumor,
Small cell lung cancer,
adrenal adenoma, medullary
thyroid carcinoma,
pheochromocytoma

Fig. 6.2 Pathogenesis of acromegaly. *GHRH*, Growth hormone–releasing hormone; *GHS*, growth hormone secretagogue, including ghrelin; *IGF-1*, insulin-like growth factor 1; *MEN1,* multiple endocrine neoplasia type 1; *SRIF*, somatotropin release–inhibiting factor. (Modified from Melmed S. Acromegaly. *N Engl J Med.* 2006; 355:2558-2573.)

TABLE 6.1 **Clinical and Pathologic Characteristics of Growth Hormone–Secreting Pituitary Tumors**

Cell Type	Hormonal Products	Clinical Features	Histologic Features
Densely granulated somatotroph	GH	Slow growing	Numerous somatotrophs with large secretory granules
Sparsely granulated somatotroph	GH	Rapidly growing Often invasive	Cellular pleomorphism
Mixed cell (somatotroph/lactotrope)	GH and PRL	Variable	Densely and sparsely granulated somatotrophs and lactotrophs
Mammosomatotroph	GH and PRL	Common in children Gigantism Mild hyperprolactinemia	Both GH and PRL in same cell, often same secretory granule
Acidophil stem cell	PRL and GH	Rapidly growing/invasive Hyperprolactinemia dominant	Distinctive ultrastructure Giant mitochondria
Plurihormonal cell	GH (PRL) with α-GSU, FSH/LH, TSH, or ACTH	Secondary hormonal products are often clinically silent Rarely hyperthyroidism or Cushing disease	Variable: either monomorphous or plurimorphous
Somatotroph carcinoma	GH	Aggressive and invasive	Rigorously documented extracranial metastasis

ACTH, Adrenocorticotropic hormone; *FSH*, follicle-stimulating hormone; *GH*, growth hormone; *α-GSU*, α-glycoprotein subunit; *LH*, luteinizing hormone; *PRL*, prolactin; *TSH*, thyroid-stimulating hormone. (Modified from Melmed S. Pathogenesis of pituitary tumors. *Endocrinol Metab Clin North Am* 1992; 21:553-574.)

stem cell adenomas secrete GH, their predominant product is PRL, thus accounting for the high incidence of hyperprolactinemic symptoms (galactorrhea, amenorrhea, infertility) initially seen in these patients. Patients with McCune–Albright syndrome also may have acromegaly, although the presence of a discrete GH-cell adenoma has been inconsistently reported in these cases.[12] Rarely, acromegaly may occur in patients with a partially empty sella. The rim of pituitary tissue surrounding the empty sella may harbor a small endocrinologically active GH-secreting adenoma not visible on magnetic resonance imaging (MRI) (i.e., <2 mm in diameter). Very rarely, ectopic GH production by pancreatic, lung, ovarian, or lymphocytic neoplasms may result in acromegaly.[13]

Excess Growth Hormone–Releasing Hormone Secretion

Excessive circulating levels of GHRH may overstimulate the pituitary and cause somatotroph hyperplasia, GH hypersecretion, and acromegaly.[14] Central overproduction of GHRH may occur in patients harboring hypothalamic hamartomas or gangliocytomas. These rare tumors are usually diagnosed by pathologic examination of a surgically resected sellar mass causing GH hypersecretion and acromegaly. Ectopic GHRH production by neuroendocrine tumors, although rare, accounts for most cases of ectopic acromegaly.

Pituitary somatotroph hyperplasia with acromegaly associated with ectopic GHRH production has been reported in more than 100 patients, and the original isolation of GHRH was from a pancreatic neuroendocrine tumor.[15] Because the peripheral features of hypersomatotropism are quite similar in all forms of pituitary and nonpituitary acromegaly, diagnosis of the etiology of the disease may be clinically challenging.

Role of the Hypothalamus in the Etiology of Acromegaly

Hypothalamic GHRH and SRIF selectively regulate *GH* gene expression and GH protein secretion. Mice bearing an overexpressing *GHRH* transgene are subject to somatotroph hyperplasia, and ultimately to pituitary adenomas.[16] In patients with neuroendocrine tumors and ectopic GHRH production, somatotroph hyperplasia and occasionally adenomas also may develop, which suggests that disordered endocrine or paracrine GHRH or SRIF action may be permissive for pituitary tumor growth.[10] GHRH signaling defects also have been identified in acromegaly. Constitutive activation of the GHRH receptor G-protein signaling unit facilitates ligand-independent induction of *GH* gene expression. This *gsp* mutation results in guanosine triphosphatase (GTPase) inactivation, with subsequent elevated cyclic adenosine monophosphate (cAMP) levels and GH hypersecretion. Excessive cAMP response element binding (CREB) protein serine phosphorylation also may account for activation of the CREB–pituitary specific transcription factor 1 signaling unit in a subset of GH-cell adenomas.[17] cAMP activation mediates both DNA damage and GH hypersecretion in these adenomas.[18]

Intrinsic Pituitary Lesions

Virtually all GH-cell adenomas arise as discrete clonal expansions of a transformed cell[19] (Table 6.2). This monoclonal origin implies that intrinsic genetic alterations account for tumorigenic initiating events and supports observations that resection of small, well-circumscribed adenomas usually results in surgical cure of GH-secreting adenomas.

Because adenohypophyseal tissue surrounding the pituitary adenoma is histologically normal, it is unlikely that multiple independent cellular growth events (e.g., generalized hyperplasia) precede adenoma formation. Increasing evidence points to complex molecular cascades accounting for the cellular progression, resulting in pituitary-cell transformation and, ultimately, tumor formation. Multistep development of pituitary acromegaly involves a spectrum of genetic alterations associated with dysregulation of cell proliferation, differentiation, and GH production[10] (see Table 6.2).

Candidate Genes in the Etiology of Acromegaly

Generally, somatic oncogene mutations have not been found in GH-secreting adenomas. Several transgenic animal models have shown that disruption of tumor-suppressor genes (including retinoblastoma [*RB*] and *p27*) results in a high incidence of pituitary tumor formation in affected mice (Table 6.3). Loss of heterozygosity (LOH) involving chromosomes 11q13, 13, and 9 occurs in up to 20% of sporadic[20] pituitary tumors. Chromosome 13q LOH occurs in proximity to the *RB* locus and was found in 13 aggressive pituitary tumors, whereas small, well-circumscribed tumors exhibit intact *RB* alleles.[21] Chromosomal instability in GH-secreting tumors is associated with chromosomal copy number variations, and these aberrations were linked to the cAMP pathway inducing GH production.[18] Germline aryl hydrocarbon receptor interacting protein (*AIP*) gene mutations occur in a subset of young patients with familial acromegaly[22,23] (see Chapter 14). Moreover, transgenic mice heterozygous for AIP inactivation develop GH-expressing pituitary adenomas with complete penetrance.[24]

TABLE 6.2 Evidence for an Intrinsic Pituitary Defect in the Pathogenesis of Acromegaly

Growth hormone (GH)-secreting adenomas are monoclonal in origin
Somatotroph hyperplasia is rarely seen in normal pituitary tissue surrounding pituitary adenomas
Successful surgical cure of well-circumscribed GH cell adenomas is achieved in >75% of patients
Generalized somatotroph hyperplasia is rare
Unrestrained GH hypersecretion occurs independent of physiologic hypothalamic feedback control
Normalization of GH pulsatility may occur after complete adenoma resection

Modified from Melmed S. Pituitary-tumor endocrinopathies. *N Engl J Med* 2020; 382:937-950.

Activating Mutations. GTPase acts to inactivate stimulatory G (Gs) proteins that induce adenyl cyclase and intracellular cAMP accumulation.[25] Missense mutations replacing residue 201 (Arg → Cys or His) or 227 (Gln → Arg or Leu), termed *gsp* mutations, result in persistently elevated ligand-independent Gs activity and constitutively elevated cAMP and GH hypersecretion. *Gsp* mutations occur in a subset of GH adenomas, with a prevalence ranging from 30% to 40% in White patients to 10% in Japanese patients with acromegaly. X-linked acrogigantism has been reported with trisomy X, associated with overexpressed somatotroph GPR101 leading to GH overproduction mediated by protein kinase C activation.[26,27]

Development of true GH-cell carcinoma with documented extracranial metastases, however, is exceedingly rare,[28,29] and classic oncogene mutations have not been observed in large series analyzed.[18] A pituitary tumor transforming gene (*PTTG*) isolated from rat GH-secreting pituitary tumor cells[30] is homologous to yeast securin, which regulates sister chromatid separation during mitosis.[31] *PTTG* overexpression results in experimental pituitary tumor formation *in vivo*. PTTG mRNA is abundant in GH-producing tumors, with more than 10-fold increases evident in larger tumors.[32] The transforming potential of PTTG indicates a role in early induction of GH cell transformation by inappropriate regulation of the pituitary cell cycle.[33] Consistent with these observations, DNA damage with chromosomal copy number variations reported in these tumors has been linked to cAMP-mediated GH overproduction.[18] More than 70% of GH-secreting adenomas exhibit features of DNA damage and premature proliferative arrest, including increased p21 and beta-galactosidase expression, all features of cellular senescence, which may underlie the invariably benign nature of these tumors.[34]

Genetic Syndromes (see also Chapter 14). Acromegaly may occur as a component of multiple endocrine neoplasia (MEN) syndromes, including the Carney complex or MEN1. The Carney complex consists of myxomas, spotty skin pigmentation, and testicular, adrenal,

TABLE 6.3 Selected Genes Associated With Molecular Pathogenesis of Pituitary Adenomas

Gene	Function	Mode of Activation or Inactivation	Clinical Context
GNAS	Oncogene	Activating, imprinting	Nonfamilial, syndromic or sporadic
CREB	Transcription factor	Constitutive phosphorylation	Sporadic
AIP	Tumor suppressor	Inactivating	Familial, syndromic
MEN1	Tumor suppressor	Inactivating	Familial, syndromic
PRKAR1A	Tumor suppressor	Inactivating	Familial, syndromic
H-RAS	Oncogene	Activating	Invasive or malignant
CCNB2	Cyclin	Induced by *HMGA*	Sporadic
CCND1	Oncogene	Overexpression	Sporadic
CDKN1B	Cyclin-dependent kinase inhibitor	Inactivating	Sporadic, syndromic
HMGA2	Oncogene	Overexpression	Sporadic
FGFR4	Oncogene	Alternative transcription	Sporadic
PTTG	Securin	Overexpression	Sporadic
RB	Tumor suppressor	Epigenetic silencing	Sporadic
CDKN2A	Cyclin-dependent kinase inhibitor	Epigenetic silencing	Sporadic
BRG1	Tumor suppressor	Glucocorticoid receptor function	Sporadic
GADD45G	Proliferation inhibitor	Epigenetic silencing	Sporadic
MEG3	Proliferation inhibitor	Epigenetic silencing	Sporadic
USP8	Pro-proliferative (corticotrophs)	Ubiquitination	Sporadic
GPR101	Hypersecretion (GH)	Duplication	Sporadic
STAT3	Hypersecretion (GH)	Transcriptional	Sporadic

Modified from Melmed S. Pathogenesis of pituitary tumors. *Nat Rev Endocrinol.* 2011; 7:257-266.
GH, Growth hormone.

and pituitary tumors.[35] About 20% of patients with this autosomal dominant syndrome associated with chromosome 2p16 harbor GH-secreting pituitary tumors.

MEN1, an autosomal dominant syndrome, consists of hyperplastic or adenomatous parathyroid glands, endocrine pancreas, and anterior pituitary.[36] Pituitary adenomas develop in almost half these patients, with GH cell adenomas reported in about 10% of afflicted subjects. Inactivation of the *MEN1* tumor suppressor gene likely accounts for the syndrome, in accordance with Knudson's "two hit" theory, whereby both inherited allelic germline mutations and a somatic deletion are required for inactivation of both specific alleles and subsequent tumor formation.

Low-prevalence germline mutations of the *AIP* gene (located on 11q 13.3) were reported to be predisposing in a subset of patients with familial acromegaly and gigantism[22]; 15% of families with isolated familial acromegaly exhibit *AIP* mutations, and tumors are encountered with gigantism at a younger age in subjects harboring a mutation.[37,38] *AIP* has been proposed as a tumor suppressor gene for pituitary adenomas[24]; however, the clinical significance of these findings to the broad adult population harboring sporadic GH-secreting tumors remains to be elucidated, as somatic *AIP* mutations have not been reported in sporadic tumors.

Epidemiology

Acromegaly is a rare disease, and accurate assessment of its prevalence in the community has been difficult to ascertain. Data from the United Kingdom and Belgium point to an approximately 10-fold higher incidence than previously considered.[39] In the United States, an estimated 3000 new cases are diagnosed each year.[40] The mean age at diagnosis is 40 to 45 years, and its insidious onset may cause the disease to not be diagnosed until 10 to 12 years after symptom onset.[41,42] This long delay in diagnosis is often attributed to the subtle and slow onset of common symptoms, including headache, joint pains, jaw malocclusion, or mild type 2 diabetes.[43] Furthermore, this relatively long-term delay allows prolonged exposure of peripheral tissues to unacceptably elevated GH and IGF-1 levels.

DIAGNOSIS

Persistent GH hypersecretion is the hallmark of acromegaly. Excess GH stimulates hepatic production of IGF-1, which is responsible for most of the clinical manifestations of acromegaly.[44] The diagnosis is often delayed because of slow clinical progression over many years. Heightened clinical awareness may result in earlier diagnosis. Although serum GH and IGF-1 concentrations are both increased in virtually all patients with acromegaly, serum IGF-1 levels may be discordant with GH increases.[45] When acromegaly is suspected, biochemical testing is required to confirm the clinical diagnosis, and imaging techniques are used to localize the cause of excess GH secretion (Table 6.4).

Documenting Growth Hormone Hypersecretion

The diagnosis of acromegaly is confirmed by measurement of serum GH after a glucose load and by assessing levels of GH-dependent IGF-1.[3,46,47] IGF-1 levels reflect the integrated bioeffects of GH hypersecretion, and age-matched elevated IGF-1 levels are pathognomonic of acromegaly.[48]

Measurement of the serum IGF-1 concentration is the most precise screening test for acromegaly. Unlike those of GH, serum IGF-1 concentrations do not fluctuate hourly according to food intake, exercise, or sleep, but rather reflect integrated GH secretion during the preceding day or longer. Plasma IGF-I levels correlate with mean 24-hour basal GH concentrations measured every 10 to 20 minutes.[49] Serum IGF-1 concentrations are elevated in virtually all patients, thus providing excellent discrimination from subjects without acromegaly. In normal subjects, serum IGF-1 concentrations are highest during puberty and decline gradually thereafter; values are significantly lower in adults older than 60 years than in younger subjects. Females have higher levels than males, and pregnancy may also be associated with elevated IGF-1 levels.[8] Thus, an inappropriately controlled "normal" IGF-1 value in an elderly male patient may in fact be truly elevated and indicative of acromegaly.

TABLE 6.4 Diagnosis of Acromegaly

Biochemical Testing
GH nadir >1 ng/mL during oral glucose load
Elevated age-matched IGF-1 level
MRI
Visualization of pituitary adenoma

GH, Growth hormone; *IGF-1*, insulin-like growth factor 1; *MRI*, magnetic resonance imaging.

Although all patients have increased GH secretion, it may be difficult to distinguish elevated random GH levels from normal. As GH levels fluctuate widely throughout the day and night, measuring random GH levels rarely provides useful information for diagnosis of the disorder. Short-term fasting, exercise, stress, and sleep are associated with elevated GH, and the availability of ultrasensitive GH assays has indicated that this pulsatile GH rhythm may occur at levels below the detectable sensitivity of previously available assays. Serum GH concentrations fluctuate widely, from less than 0.5 ng/mL (with ultrasensitive assays) during most of the day to as high as 20 or 30 ng/mL at night or after vigorous exercise. Random serum GH concentrations may be elevated in patients with uncontrolled diabetes mellitus, liver disease, and malnutrition, so dynamic tests have been proposed to confirm pituitary GH hypersecretion.[46]

The diagnostic hallmark of excess GH hypersecretion is failure to suppress GH levels to 0.4 μg/L or less (using a chemiluminescent immunoradiometric assay) during a 2-hour period after a 75-g oral glucose load.[46] Several factors determine measured serum GH values, including age, sex, body mass index (BMI), and the type of assay employed. Spontaneous GH secretion is attenuated by 50% to 70% in subjects aged 65 years or more, and IGF-1 levels also decline progressively with age. GH levels correlate inversely with BMI, and lean subjects exhibit higher GH values, as do female subjects. Accordingly, criteria for the diagnosis of acromegaly requires demonstrating a mean 24-hour GH level of greater than 2.5 μg/L, a nadir GH of greater than 1 μg/L after a glucose load, and/or an elevated age- and sex-matched serum IGF-1 level.

When nadir GH was measured during a 75-g oral glucose tolerance test (OGTT) in 46 acromegaly patients using three different commercial assays, GH concentrations were shown to be dependent on the specific assay employed, sex, age, and BMI.[50] Values obtained with one commercial assay were approximately 2.3-fold higher than those determined using a second assay. In one assay, postglucose values of less than 1 μg/L were associated with disease control, whereas with the other assay, the proposed cutoff value was 0.5 μg/L. Thus, interpretation of biochemical control should also be determined by knowledge of the specific GH assay employed, as well as the appropriateness of assay controls. About 10% of patients or fewer may have apparently "normal" GH or IGF-1 levels or both at the time of diagnosis. Repeating the assays in a reputable laboratory may often resolve an apparent clinical/biochemical discordance.

Localizing the Source of Excess Growth Hormone

Once a biochemical diagnosis of GH hypersecretion is confirmed, MRI of the pituitary to localize the source of hormone excess is indicated. MRI effectively delineates soft-tissue pituitary masses, and gadolinium-enhanced MRI may detect adenomas 2 mm in diameter. In about 75% of patients, the tumor is a macroadenoma (tumor diameter of >10 mm) and may extend to parasellar or suprasellar regions or invade the cavernous sinus. More than 90% of patients exhibit a discrete pituitary adenoma on MRI, whereas about 10% of patients may harbor a partial or even apparently total empty sella. Functional GH-secreting adenomas may arise in the remnant rim of pituitary tissue surrounding the empty sella and may not be visible on MRI. Rarely, other nonpituitary causes of acromegaly (see earlier) will require abdominal or chest imaging to localize the source of ectopic GHRH or, more rarely, GH production.

Nonpituitary Acromegaly. Rare nonpituitary causes of acromegaly include a hypothalamic tumor secreting GHRH,[14] a nonendocrine tumor secreting GHRH,[15] or ectopic GH secretion by a nonendocrine tumor.[13,51] If pituitary MRI findings are normal, abdominal and chest imaging should be performed, followed by catheterization studies in an attempt to demonstrate an arteriovenous GHRH gradient over the suspected tumor bed. In patients with ectopic GHRH secretion, serum GHRH and GH concentrations are both elevated, and pituitary MRI reveals a normal-sized or enlarged hyperplastic gland.[52]

An algorithm for the diagnostic evaluation of patients suspected of having acromegaly is shown in Fig. 6.3. A normal age-controlled serum IGF-1 concentration is strong evidence excluding the diagnosis. If the serum IGF-1 concentration is high (or equivocal), serum GH should be measured over 2 hours after oral glucose administration. If pituitary MRI fails to reveal the presence of a discrete adenoma in the presence of clear-cut biochemical evidence of hypersomatotropism, studies to identify the rarely encountered GHRH- or GH-secreting tumors should be undertaken.

Clinical Manifestations

The somatotroph adenoma itself, especially if a macroadenoma, may cause local symptoms such as headache, visual-field defects (classically bitemporal hemianopia), and cranial-nerve palsies. These compressive features are not unique to acromegaly and may occur with any enlarging sellar mass. Nevertheless, the headache associated with acromegaly is uniquely debilitating and may not be exclusively caused by pressure effects.

The systemic clinical features of acromegaly occur as a consequence of the deleterious impact of elevated serum concentrations of both GH and IGF-1 on peripheral tissues (Table 6.5). The somatic impact of elevated GH includes growth stimulation of a variety of tissues, such as skin, connective tissue, cartilage, bone, and many epithelial tissues, including mucosal surfaces.[3,4,53] The metabolic effects of excess GH include nitrogen retention, insulin antagonism, and enhanced lipolysis.[54]

KEY POINTS

- High levels of GH and IGF-1 lead to acral and metabolic dysfunction.

The onset of acromegaly is insidious, and disease progression is usually slow. At diagnosis, about 75% of patients are shown to harbor macroadenomas (tumor diameter of >10 mm), and some tumors extend to the parasellar or suprasellar regions.[55] Headaches are the initial symptom in approximately 60% of patients, and 10% have visual symptoms. Disease persistence is adversely determined by sparse GH tumor granularity, younger age of onset, larger and more extrasellar tumor mass extension, and higher initial GH and IGF-I levels.[11]

Acral Overgrowth. Acral and soft-tissue overgrowth is invariably a feature of acromegaly. Characteristic findings include an enlarged, protruding jaw (macrognathia) with associated mandibular overbite and enlarged, swollen hands and feet, resulting in increasing shoe and glove size and the need to enlarge rings. Facial features are coarse, with enlargement of the nose and frontal bones, as well as the jaw; the upper incisors are consequently spread apart. Despite the prominence of these findings, the rate of change is so slow that patients may not seek care because their appearance has changed. Automated facial classification photographic software has also been employed to detect the disorder.[54]

Rheumatologic Features. Musculoskeletal symptoms are leading causes of morbidity and serious functional disability in patients with acromegaly.[56] At least half of all patients exhibit minor arthralgias, and severe, debilitating arthritic features may ultimately develop. The pathogenesis of joint disease generally begins with a noninflammatory osteoarthritic disorder and culminates in severe secondary joint and cartilage degeneration.[57-59] Excess GH and IGF-1 exposure leads to uneven cartilage proliferation that results in a mechanically unstable joint surface. Subchondral cysts and osteophytes then develop in an irreversible self-perpetuating process. Severe physical deformity and functional disability result from these inexorable pathologic and mechanical stresses. The incidence of vertebral fractures is increased, approximately 40% in active acromegaly, especially in the face of lowered bone mineral density.[58,60] Hypogonadism and decreased femoral neck bone mineral density are significant risk factors. Although symptomatic and functional relief of arthritic disorders is observed in most patients after reducing GH levels, structural changes are not reversible.

Skin and Soft Tissues. The skin thickens, and multiple recurrent skin tags may appear. Hyperhidrosis at rest is common (present in 50% of patients)[2,61] and often malodorous. Hair growth increases, and some women have hirsutism. Other manifestations of soft-tissue overgrowth include macroglossia, deepening of the voice, and paresthesias of the hands (carpal tunnel syndrome) from nerve entrapment.[62,63] A symmetric sensorimotor peripheral (rarely hypertrophic) neuropathy may also occur unrelated to entrapment. Skeletal muscle mass was unchanged from controls in 27 acromegaly patients undergoing dual-photon x-ray absorptiometry.[64]

Thyroid. Thyroid enlargement may be diffuse or multinodular. Thyroid function is, however, usually normal.

Cardiovascular. Impaired cardiovascular function is an important determinant of morbidity and mortality,[3,65] with exacerbated cardiovascular risk factors.[66] The deleterious direct impact of excess GH and IGF-1 on the heart, as well as the effect of hypertension, which is present in 30% of patients, contribute to the disorder.[62,67] The severity of the cardiomyopathy correlates with the duration of exposure to hypersomatotropism. Mean left ventricular mass may be significantly increased to more than 200 g, as opposed to a normal mean weight of 140 g, and end-systolic and diastolic volumes are attenuated. Concentric ventricular hypertrophy is associated with interstitial fibrosis and lymphocytic infiltration. Resting diastolic blood pressure and left and right ventricular peak filling rates are elevated. Postexercise systolic and diastolic blood pressure may also be elevated, and the left ventricular ejection fraction is attenuated. GH induces distal nephron epithelial sodium channel–mediated sodium transport, likely accounting for the enhanced sodium retention characteristic of acromegaly.[68]

Sleep Apnea. Peripheral airway obstruction caused by macroglossia, mandible deformation, mucosal hypertrophy, and inspiratory laryngeal collapse has long been recognized as causing airway obstruction snoring, and sleep apnea.[67,69,70] Sleep apnea afflicts most patients with acromegaly and has been documented in up to 80% of cases.[71] Macroglossia and enlargement of the soft tissues of the pharynx and larynx lead to obstructive sleep apnea in about 50% of patients; others have central sleep apnea, possibly resulting from altered central respiratory control.[72] Sleep apnea may be an important cause of mortality in these patients. A central form of sleep apnea was recognized in

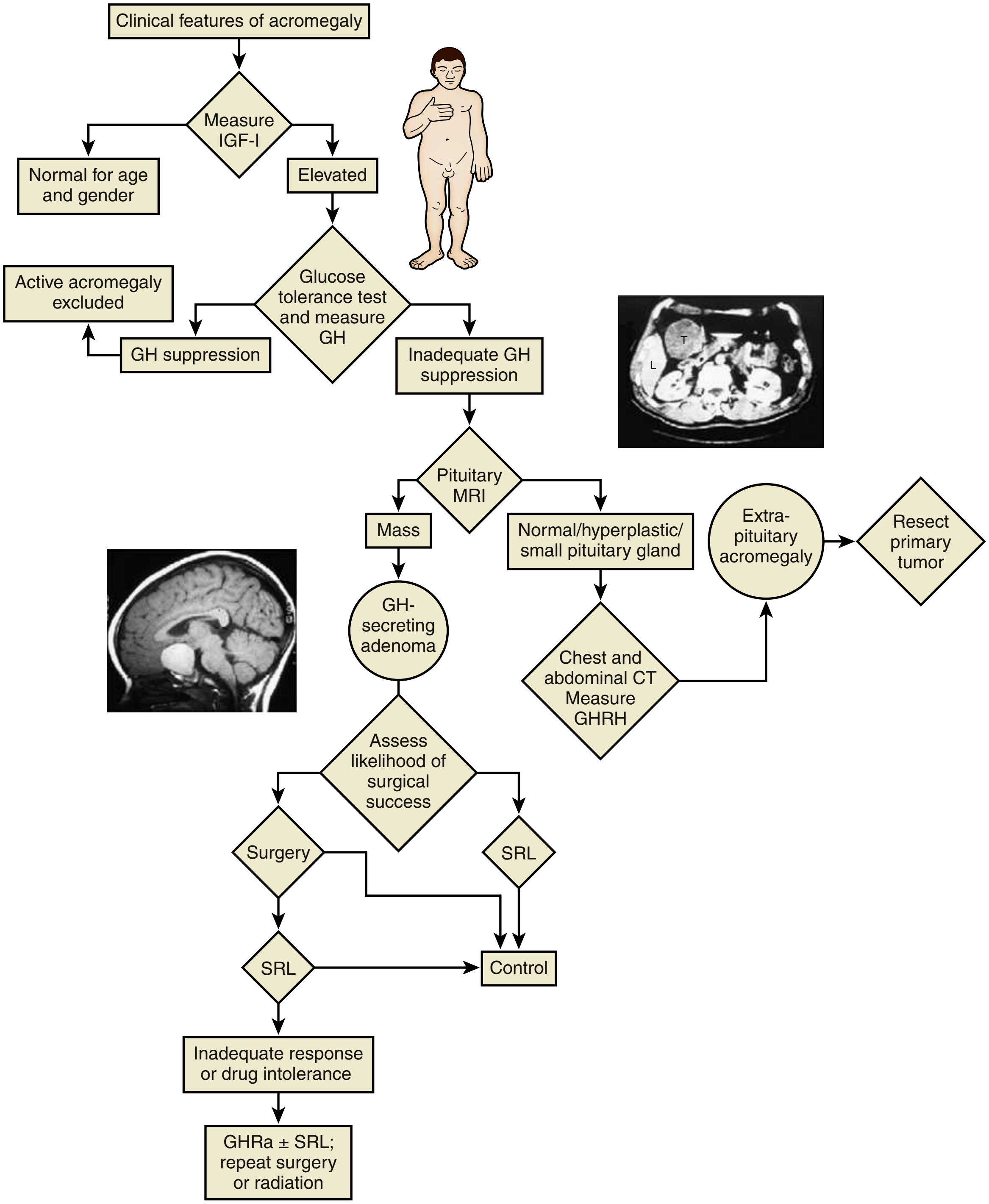

Clinical features of acromegaly
Measure IGF-I
Normal for age and gender
Elevated
Glucose tolerance test and measure GH
Active acromegaly excluded
GH suppression
Inadequate GH suppression
T
L
Pituitary MRI
Mass
Normal/hyperplastic/ small pituitary gland
Extra-pituitary acromegaly
Resect primary tumor
GH-secreting adenoma
Chest and abdominal CT Measure GHRH
Assess likelihood of surgical success
Surgery
SRL
SRL
Control
Inadequate response or drug intolerance
GHRa ± SRL; repeat surgery or radiation

Fig. 6.3 Diagnosis and management of acromegaly. *CT*, Computed tomography; *GH*, growth hormone; *GHRH*, growth hormone–releasing hormone; *IGF-1*, insulin-like growth factor 1; *MRI*, magnetic resonance imaging; *OGTT*, oral glucose tolerance test; *GHRA*, growth hormone receptor antagonist; *SRL*, somatostatin receptor ligand. (Modified from Melmed S. Acromegaly. *N Engl J Med*. 2006; 355:2558-2573.)

acromegaly[72] that appears to correlate more closely with elevated GH and IGF-1 levels and may reflect central respiratory suppression caused by the dysregulated hypothalamic–GH axis. Disease control improves or abrogates sleep apnea, with improved indices for apnea, hypopnea, and oximetry after treatment.[69,73] After 6 months of treatment of 14 apneic acromegalic patients with octreotide, a 40% decrease in the number of apneic events per hour was seen, as well as a decrease in total apneic time from 28% to 15%. Maximum oxygen saturation rose from 76% to 84%, accompanied by a decline in daytime sleepiness, as well as improvement in central and obstructive apneic parameters.[72] Nevertheless, sleep apnea may persist despite acromegaly control, and posttreatment evaluation is still required.[71]

Diabetes. GH is a potent antagonist of insulin action. Approximately 25% of patients have frank diabetes at diagnosis, and glucose values greater than 200 mg/dL (11 mmol/L) on OGTT in those without a

TABLE 6.5 Risk of Long-Term Exposure to Elevated Growth Hormone Levels

Arthropathy
Unrelated to age at onset or to growth hormone (GH) levels
Usually occurs with acromegaly of long duration
Reversibility
  Rapid symptomatic improvement with treatment
  Irreversible bone and cartilage lesions
Bone Disorders
Vertebral fractures
Neuropathy
Peripheral nerves affected
  Intermittent anesthesias, paresthesias
  Sensorimotor polyneuropathy
  Impaired sensation
Reversibility
  Onion bulbs (whorls) do not regress with lowered GH levels
Cardiovascular Disease
Cardiomyopathy
  Left ventricular diastolic function decreased
  Left ventricular mass increased; arrhythmias
  Fibrous hyperplasia of connective tissue
Hypertension
  Exacerbates cardiomyopathic changes
Reversibility
  May progress even with normalized GH levels
Respiratory Disease
Sleep apnea
  Caused by soft-tissue overgrowth and decreased pharyngeal muscle tone
Reversibility
  Improved with reduction of GH levels
Malignancy
Increased risk of soft tissue tumors and perhaps malignancy
Increased colon polyps
Reversibility
  Effect of therapy on risk of malignancy unclear
Carbohydrate Intolerance
Occurs in one fourth of acromegalics, more often with family history of diabetes mellitus
Reversibility
  Improves with reduced GH levels

Modified from Melmed S, Dowling RH, Frohman L, et al. Acromegaly: consensus for cure. *Am J Med.* 1994; 97:468.

diabetes diagnosis are not uncommon.[74] Diabetes is an important systemic complication of hypersomatotropism and a major determinant of mortality, with only 30% of patients with diabetes at the time acromegaly is diagnosed surviving 20 years.[75]

Gonadal Function. Women may have amenorrhea, with or without galactorrhea, and some have hot flashes and vaginal atrophy. Men may have erectile failure, loss of libido, decreased facial hair growth, and testicular atrophy. Hypogonadism is caused either by hyperprolactinemia (present in about 30% of patients) or by impaired gonadotropin secretion as the expanding pituitary tumor compresses normal pituitary gonadotroph cells. Asymptomatic, reversible prostatic enlargement also is common, even in men with hypogonadism.

Neoplasms. Acromegaly is associated with an enhanced risk for development of colonic polyps, as evidenced by retrospective studies,[76] whereas prospective studies have reported premalignant adenomatous colonic polyps in up to 30% of patients. Colon length is increased, and apoptotic activity is decreased. Multiple adenomatous polyps are more likely, as well as polyps proximal to the splenic flexure, underscoring the need for full-length colonoscopy.[77] Although a metaanalysis suggests a doubling of colon cancer risk in acromegaly,[78] the relationship between excess cancer-related mortality and serum GH levels is less clear.[79,80] Screening for colorectal and thyroid cancer is recommended as for the general public.[70]

Laboratory Findings

Patients exhibit increased serum GH and IGF-1 concentrations and may have hyperglycemia, with frank diabetes occurring in 25% of patients. Some patients have hypertriglyceridemia. Hypercalciuria and hyperphosphatemia (>5.5 mg/dL) occur in approximately 70% of patients as a result of direct stimulation of renal tubular phosphate reabsorption by IGF-1. Hyperprolactinemia occurs in about 30% of patients and may be caused by cosecretion of PRL and GH by the tumor or may be due to stalk interference with hypothalamic-pituitary portal delivery of dopamine. Secretion of other pituitary hormones, especially gonadotropins, also may be decreased because of mass compression.

KEY POINTS

- If uncontrolled, the disorder is associated with increased morbidity and mortality, mainly from cardiovascular, respiratory, and metabolic disorders.

Mortality

The overall mortality rate in acromegaly is about two to four times that of the general population.[81,82] In a metaanalysis of 16 studies, the overall weighted mean standardized mortality ratio (SMR) was 1.72.[83] Even after tumor surgical resection, mortality remained high (SMR 1.32).[83] Survival is reduced an average of 10 years in comparison to age-matched controls. When 18 studies of mortality outcomes in acromegaly were analyzed,[81] there appeared to be a 1.9-fold increase over expected mortality rates (range 1.16 to 3.3) for patients having undergone heterogeneous treatments. Although long-term standardized mortality outcomes have not been reported for untreated acromegaly, several significant mortality determinants are evident. These include GH greater than 2.5 μg/L, or elevated IGF-1 level, the prior use of radiotherapy, time delay from diagnosis, duration of the disease, and the presence of hypertension.[84] Most significant predictors of survival include a postglucose GH level of less than 1 μg/L, a random serum GH of less than 2 to 2.5 μg/L, or a normal serum IGF-1 level (Fig. 6.4). Clearly, attaining tight biochemical control should therefore be a goal of therapy. Aggressive management of comorbidities including hypertension, heart failure, sleep apnea, and diabetes likely also contributes to improved mortality rates.[82,85] Retrospective studies indicate that survival may be normalized to a control age-matched rate by controlling GH levels; in particular, serum IGF-1 standard deviation score life-table analysis showed that GH levels less than 2.5 μg/L were associated with survival rates equal to those of the general population.[82] A metaanalysis of 18 studies showed that controlling random GH levels to less than 2.5 μg/L and normalizing IGF-I levels restored adverse SMR to 1.1 (95% confidence interval [CI]:0.9–1.4).[86] Tight control of GH through aggressive multimodal therapy appears to reduce mortality risk to that expected for control subjects.[46]

MANAGEMENT

Treatment Goals

Treatment goals embody principles that apply to treating hormonal hypersecretory tumors (Table 6.6). Treatment should be both safe and efficacious. GH and IGF-1 levels should be normalized, especially

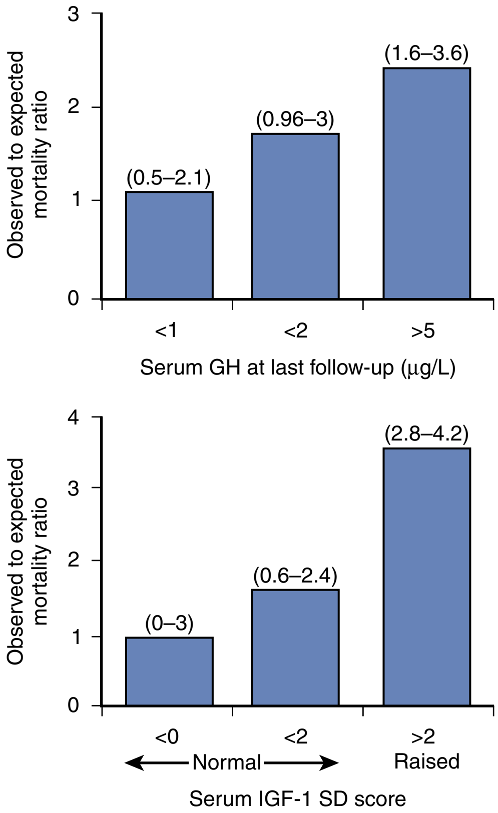


Fig. 6.4 Growth hormone and insulin-like growth factor 1 levels as mortality determinants. *GH,* Growth hormone; *IGF-1*, insulin-like growth factor 1; *SD*, standard deviation. (Data from Holdaway IM, Rajasoorja RC, Gamble GD. Factors influencing mortality in acromegaly. *J Clin Endocrinol Metab*. 2004; 89:667-674.)

because elevated GH levels have been associated with mortality in these patients. Thus, tight biochemical control is an important therapeutic endpoint. Tumor mass effects, especially central compression of visual tracts, should be alleviated. Importantly, the integrity of pituitary function should be preserved, and if hypopituitarism develops, patients require lifelong pituitary replacement. Clinical features of the disease that lead to the characteristic morbidity, and ultimately to mortality, should be ameliorated. Several treatment options are available. Transsphenoidal surgical resection of the adenoma, pharmacologic therapy with somatostatin receptor ligands (SRLs), dopamine agonists, and GH receptor antagonist (GHRA), and radiation therapy may all be used in the treatment of acromegaly. The challenge of tight GH control can now be met with greater stringency by using single or multimodal forms of therapy. Effective disease control should thus include sustained hormone suppression, a contained adenoma mass, improved systemic morbidity, and ultimately, normalized mortality (see Table 6.6).

KEY POINTS

- Treatment options include surgical excision, medical therapies, and radiotherapy.

Surgery

The most important determinant of a successful surgical outcome is the experience of the surgeon. Endoscopic approaches to pituitary adenoma resection are efficacious and safe.[87-90] Surgery rapidly alleviates acromegaly symptoms, removes the tumor mass, relieves optic-tract pressure effects, and relieves headache. Tumor debulking, even if partial, also may be helpful in enhancing the effectiveness of subsequent therapy. Surgical control also is inversely correlated with initial tumor invasiveness and GH levels.[91,92] Not all patients are appropriate surgical candidates, usually because of coexisting cardiovascular and pulmonary disease, which may be a contraindication to anesthesia. If the tumor is sufficiently large or invasive and portends intraoperative

TABLE 6.6 Treatment Goals for Acromegaly

	ACROMEGALY TREATMENT				
	Surgery	**Somatostatin Receptor Ligands**	**GH Receptor Antagonist**	**Dopamine Agonist**	**Radiotherapy**
Biochemical Control					
GH <1.0 μg/L	Macros <50% Micros >80%	~65%	0	<15%	60%, 10 yr
IGF-1 normal	Macros <50% Micros >80%	~65%	>90%	<15%	60%, 10 yr
Onset	Rapid	Rapid	Rapid	Slow	Slow
Compliance	One-time consent	Sustained	Sustained	Good	Good
Tumor mass	Debulked or resected	No growth or shrinkage	Unknown	Unchanged	Ablated
Disadvantages					
Cost	One time	Ongoing	Ongoing	Ongoing	One time
Hypopituitarism	10%	None	Low IGF-1	None	>50%
Other	Tumor persistence or recurrence Diabetes insipidus Local complications	Gallstones Nausea; diarrhea Hyperglycemia (pasireotide) Injection site symptoms (injectables)	Elevated liver enzymes	Nausea Sinusitis High dose required	Local nerve damage Visual and CNS disorders Cerebrovascular risk

CNS, Central nervous system; *GH*, growth hormone; *IGF-1*, insulin-like growth factor 1. (Data from Melmed S. Acromegaly pathogenesis and treatment. *J Clin Invest*. 2009; 119:3189-3202.)

damage to vital structures, the benefits of the procedure should be weighed against surgical risks.

Remission (defined as normal IGF-1 and either GH <0.4 ng/mL during an OGTT, or random GH <1 ng/mL) was achieved in all 14 microadenoma patients and 28 of 46 macroadenoma patients in one series.[89] Using remission criteria of normalized IGF-1 levels or GH suppression to less than 2 μg/L after glucose loading, more than 90% of patients with microadenomas successfully achieve control. Unfortunately, most tumors encountered at diagnosis are large macroadenomas, and fewer than 50% of patients with macroadenomas are biochemically controlled. For patients with these invasive tumors, surgical resection is invariably followed by persistent GH and/or IGF-1 hypersecretion. In patients with persistent disease, even after adjuvant irradiation or medical therapy, mortality remains significantly increased (almost 2-fold). Thus, the level of GH attained postoperatively is the most important determinant of mortality outcome. GH levels obtained 24 hours after surgery showed the highest predictive power for long-term remission, which was sustained in 125 of 127 patients with early GH levels less than 1 μg/L.[93] Regardless of the treatment mode, normalization of GH restores mortality risk to that of age-matched population controls, whereas postoperative disease persistence is associated with a 3.5-fold relative mortality risk.[82] Surgical debulking may improve subsequent disease control by SRLs.[94] Surgical resection of at least 75% of the tumor mass appeared to enhance responsiveness of residual adenoma tissue to postoperative SRL treatment.[95]

Surgical success is based largely on the skill and experience of the neurosurgeon and on tumor size or invasiveness. Surgery is useful for prompt reduction of GH levels, and tumor debulking may enhance the effectiveness of medical therapy. After apparent successful resection, however, up to 8% of tumors recur within 10 years. GH levels should be measured in the immediate postoperative period; evidence of GH hypersecretion at this time portends either disease persistence or long-term recurrence. Overall, for all tumor types, and using a highly sensitive immunoradiometric GH assay, postglucose GH values less than 1.0 μg/L are found in 50% of patients after surgery, whereas in 39% of patients with normalized IGF-1 levels, GH levels still fail to achieve suppression.[96]

Side Effects. The most important adverse surgical event is failure to totally resect invasive tumor, with consequent persistent hormonal hypersecretion. Postoperative complications occur in approximately 10% of patients, and their incidence is largely dependent on the experience of the operating surgeon. Complications include permanent diabetes insipidus, cerebrospinal fluid leaks requiring repair, meningitis, severe sinusitis, and hypopituitarism.[89] Perioperative morbidity and residual pituitary failure remain of concern in patients with invasive tumors, especially when operated on by less-experienced surgeons.

Pituitary Irradiation

Techniques for pituitary radiotherapy include external radiation with either a cyclotron or a cobalt-60 source. Maximal tumor irradiation with minimal damage to nontumorous surrounding tissue has been achieved by advances in stereotactic MRI-directed tumor localization, focused-beam direction, field-size simulation, head immobilization, and isocentral rotational techniques.

Tumor growth is invariably arrested after fractionated radiotherapy, but GH decline is slow, decreasing by approximately 20% per year. Within 18 years, 90% of patients have random serum GH concentrations lower than 5 μg/L. The degree and rapidity of GH attenuation are highly dependent on pretreatment GH levels.

Stereotactic ablation of GH-secreting adenomas by Gamma Knife radiosurgery has demonstrated improved efficacy. Over a 10-year follow-up period, approximately 50% achieve remission.[97,98] Proton-beam therapy also decreases GH secretion but is not widely available.

Side Effects (see also Chapter 16). Pituitary failure develops in approximately 50% of patients undergoing radiation therapy by 10 years and requires thyroid, gonadal, or adrenal steroid replacement or a combination of these.[99] Rarely, optic-tract damage results in visual deficits. Ten years after conventional radiotherapy, patients have a small risk of a secondary brain malignancy. Radiation may rarely induce brain parenchymal changes manifested as depression, decreased memory, loss of vision, and cranial-nerve palsies. Gamma Knife radiosurgery is associated with fewer local side effects. Metaanalyses show visual deficits in 3% of patients and hypopituitarism in 27%.[97] The risk for stroke and vascular injury is enhanced (SMR 4.42) in patients undergoing pituitary radiotherapy, with dosage being an adverse determinant.[100]

Thus, radiotherapy is effective in acromegaly, although its benefits are dose- and time-dependent, and GH reduction is delayed by 10 to 15 years. Radiation therapy may be useful for patients with growing pituitary tumors whose condition is not controlled by surgery or who are resistant to medical therapy.

Pharmacologic Management (Fig. 6.5)

Somatostatin Receptor Ligands. Octreotide is a synthetic octapeptide analogue of native, naturally-occurring somatostatin. This 8-amino-acid analog (molecular weight 1019 g/mol) binds selectively to somatostatin receptor 2 (SST2).[101] After subcutaneous injection, octreotide is rapidly absorbed, and peak drug concentrations (5.5 ng/mL) are achieved within 24 minutes of a 100-μg injection. The plasma distribution is about 12 minutes, and the elimination half-life is 1.5 hours, as compared with 2 minutes for natural SRIF.[101] The drug inhibits pituitary GH secretion, and also directly suppresses hepatic IGF-1 production,[102] controls tumor growth, and relieves soft-tissue symptoms. Pasireotide binds predominantly to somatostatin receptor 5 (SST5) and less so to SST2.[103] Hypopituitarism does not develop during SRL treatment, because SRIF analogs bind selectively to the somatostatin receptor subtype that regulates GH secretion.[104]

Responses to Long-Term Depot Somatostatin Receptor Ligand Preparations. Long-term depot SRL preparations include octreotide long-acting release (LAR) and lanreotide, which bind predominantly to SST2 receptors, and pasireotide LAR, which binds predominantly to SST5 receptors.[105-107] Octreotide LAR and pasireotide are delivered intramuscularly, whereas lanreotide is delivered by deep subcutaneous injection. Oral octreotide capsules (OOCs) contain unmodified octreotide suspended within a lipophilic medium in an enteric-coated gelatin capsule to enable octreotide absorption via transient openings in tight junctions between intestinal epithelial cells.[108] Administration of capsules timed to meals ensures effective octreotide bioavailability.[109,110]

Fig. 6.6 depicts the pharmacokinetic response to octreotide LAR in acromegaly. Drug levels peak at about 28 days after a single injection of octreotide LAR and decrease slowly thereafter. GH levels decline after the injection and by day 14 are suppressed to less than 2 μg/L. GH suppression (as determined by measurement of integrated secretion over a 4-hour period) is sustained through day 49 and starts increasing thereafter. From the pharmacokinetic curve, it is apparent that a single injection administered every 30 days will allow GH levels to be persistently suppressed throughout the month. GH suppression appears to be sustained as long as patients receive monthly SRL injections. A metaanalysis showed that about 55% of patients overall normalize both GH and IGF-1 levels while taking SRLs[111] (Fig. 6.7). Testosterone levels and eugonadism are also restored in most hypogonadal men. Several published studies include subjects preselected for prior GH responsivity to SRLs, and a more rigorous assessment of the literature suggests an overall response rate of 40% to 50% in unselected patients.[112] SRL therapy controls tumor growth in the majority of treated patients, and continued tumor expansion while receiving these drugs has rarely been reported. A metaanalysis of 41 eligible studies showed that 53% (95%

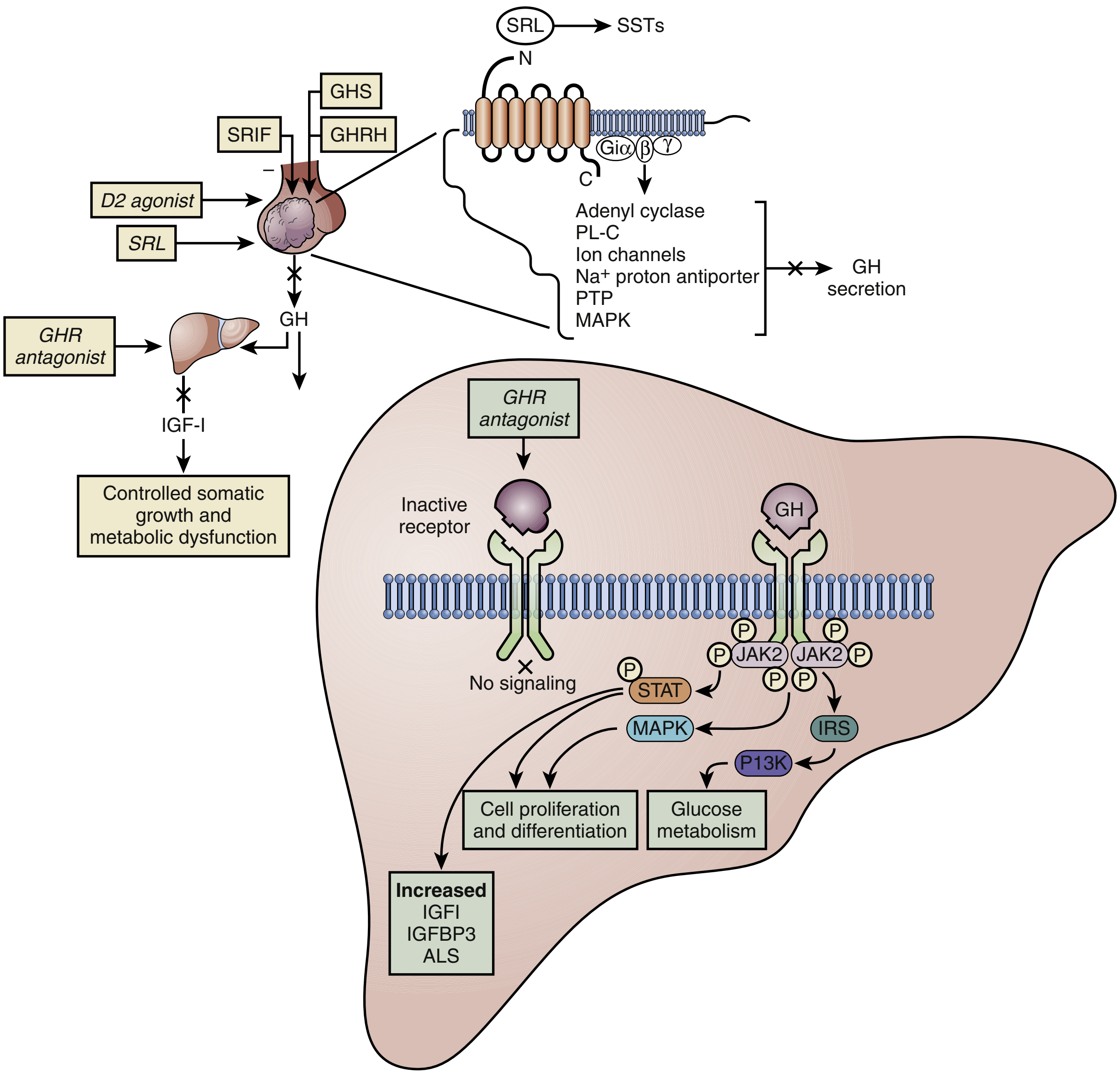

SRL
SSTs
N
GHS
SRIF
GHRH
−
Giα
β
γ
C
D2 agonist
SRL
Adenyl cyclase
PL-C
Ion channels
Na+ proton antiporter
PTP
MAPK
GH
secretion
GH
GHR
antagonist
IGF-I
Controlled somatic
growth and
metabolic dysfunction
GHR
antagonist
Inactive
receptor
GH
No signaling
P
JAK2
JAK2
STAT
MAPK
IRS
P13K
Cell proliferation
and differentiation
Glucose
metabolism
Increased
IGFI
IGFBP3
ALS

Fig. 6.5 Acromegaly pharmacotherapy. Somatostatin receptor ligands (*SRLs*) signal via G-protein coupled receptors to suppress growth hormone (*GH*) secretion and action and control tumor growth. Growth hormone receptor (*GHR*) antagonist blocks peripheral GH signaling to suppress insulin-like growth factor 1. *GHRH*, Growth hormone–releasing hormone; *GHS*, growth hormone secretagogue; *SRIF*, somatotropin release–inhibiting factor; *SRL*, somatostatin receptor ligands; *SST*, somatostatin receptor. (Modified from Melmed S. Acromegaly. *N Engl J Med*. 2006; 355:2558-2573.)

CI:45%–61%) of patients receiving SRL therapy experience a significant reduction in pituitary tumor mass.[113] Treatment with octreotide LAR resulted in 66% (95% CI:51%–74%) of patients experiencing tumor mass shrinkage. Overall, the weighted mean percentage reduction in tumor size for all SRLs was 37.4% (95% CI:22.4%–52.4%), increasing to 50.6% (95% CI:42.7%–58.4%) for octreotide LAR. Several unresolved issues remain to be determined, including the effectiveness and safety of SRL therapy in patients exhibiting optic chiasm compression, the effects of SRL withdrawal on subsequent tumor mass reexpansion, and the effect of medically-induced tumor shrinkage on subsequent surgical outcomes.[114,115] In a randomized study of 62 patients, 6 months of presurgical treatment with octreotide resulted in enhanced surgical cure rates for patients with macroadenomas. Overall, 50% of medically pretreated versus 16% of untreated patients were cured by surgery.[115] However, in another study, long-term surgical cure rates were not altered.[114]

Several biomarkers determine patient responsiveness to SRLs and have enabled disease classification for personalized therapeutic approaches.[11] Adverse determinants include higher baseline GH levels, sparse tumor GH granularity, T2-weighted tumor hyper-intensity on MRI, and low levels of tumor SST2 expression[11,116-118] (Table 6.7). Loss of specific tumor signaling molecules, including raf kinase inhibitory protein and[116] e-cadherin,[119] and germline *AIP* mutation[120] are associated with decreased octreotide responsiveness.

Extended dosing intervals (every 6 weeks) may be effective in patients well-controlled with long-acting SRL, whereas those with persistent IGF-1 elevations may benefit from higher doses and/or more frequent dosing.[46,47,121] These patients may also benefit from a switch

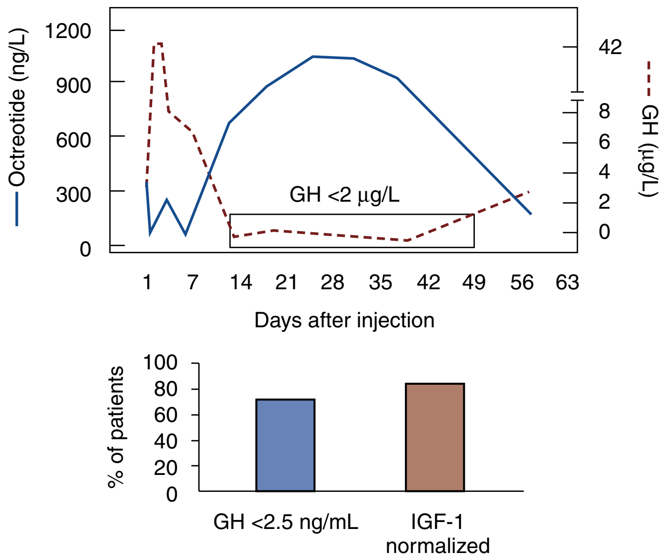


Fig. 6.6 Pharmacokinetics of octreotide long-acting release. *GH*, Growth hormone; *IGF-1*, insulin-like growth factor 1. (From Lancranjan I, Bruns C, Grass P, et al. *Metabolism*. 1996;45:67-71.)

to pasireotide,[122] but the high rate of drug-induced hyperglycemia requires careful monitoring (see later).

OOCs may be another option for patients controlled on octreotide LAR or lanreotide.[109,110] A phase 3 randomized, placebo-controlled, double-blinded study showed that 58% of patients maintained normalized IGF-1 levels after switching from injectable to oral SRL, and IGF-1 levels returned to baseline after a single SRL injection in those who required reversion to injectable SRL because of treatment failure or adverse events.[110]

OOC is initiated at a dose of 40 mg/day, but a starting dose of 60 mg/day may be an optimal starting dose for most patients.[47] Treatment is started at the time of the previously scheduled SRL injection, and can be uptitrated more frequently (every 2–4 weeks) than injectable SRLs.[47]

Side Effects. Although SRLs are relatively safe in the long term, several important adverse events have been reported. Side effect profiles are similar for all SRL formulations, except for increased hyperglycemia with pasireotide.

Asymptomatic echogenic gallbladder lesions develop in about 25% of patients.[99,101] These lesions, including sludge and gallstones, are usually diagnosed within the first 2 years of treatment, with few if any new echogenic events encountered thereafter. The prevalence of

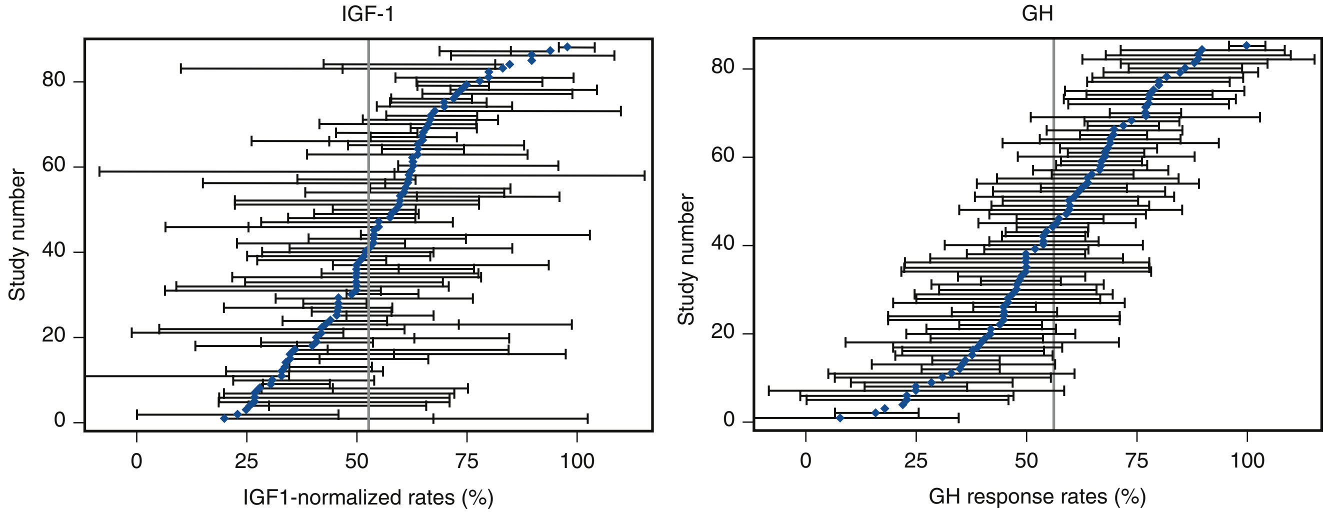


Fig. 6.7 Metaanalysis of somatostatin receptor ligand responsiveness in acromegaly. *GH*, Growth hormone; *IGF-1*, insulin-like growth factor 1. (Data from Carmichael JD, Bonert VS, Nuno M, et al. Acromegaly clinical trial methodology impact on reported biochemical efficacy rates of somatostatin receptor ligand treatments: a meta-analysis. *J Clin Endocrinol Metab*. 2014; 99:1825-1833.)

TABLE 6.7 Classification of Acromegaly

		Type 1	Type 2	Type 3
Frequency		50%	19%	31%
Tumor	*Size*	Micro	Macro	Macro
	Invasiveness	No	Intermediate	Yes
	GH granulation	Dense	Both	Sparse
Immunoreactivity	*Ki67 >3%*	90%	33%	42%
	SST2	58%	30%	22%
Biochemistry	*IGF-1*	Lower	Intermediate	Higher
Outcomes	*No. of surgeries*	1	1 or 2	≥2
	Control	Frequent	Intermediate	Rare

n = 338

GH, Growth hormone; *IGF-1*, insulin-like growth factor 1; *SSTR2,* somatostatin receptor 2.

Data from Cuevas-Ramos D, Carmichael JD, Cooper O, et al. A structural and functional acromegaly classification. *J Clin Endocrinol Metab*. 2015; 100:122-131.

gallstones during octreotide therapy appears to vary geographically. Dietary and/or other environmental factors play a role in their pathogenesis. Transient gastrointestinal symptoms, including anorexia, nausea, vomiting, flatulence, and loose stools, may occur, especially during the first 2 weeks of therapy; these symptoms may be ameliorated by injecting the medication between meals or at night. Rarely, fat malabsorption and bradycardia have been reported. The side effect profile of oral octreotide is not different from that of injectable octreotide.[110]

Unlike octreotide LAR and lanreotide, pasireotide is associated with high rates of drug-induced hyperglycemia, with up 40% of treated patients showing new-onset hyperglycemia and 25% developing diabetes mellitus.[122] Poor glycemic control at baseline, particularly elevated fasting blood glucose, is likely a predictive factor.[123] Treatment with antidiabetic medications ameliorates the effect in most patients, but careful patient selection and monitoring of glycemic effects are critical.[47,70]

Summary. Injectable and oral SRLs are effective and safe in managing GH hypersecretion, especially in patients in whom surgical resection has failed to achieve a stringent biochemical remission. Indications for SRL therapy include: postoperative persistence of GH secretion from residual tumor tissue; during the latency period of GH nonresponsiveness after radiation; and primary therapy of patients who decline surgery or are medically unfit for surgery or who exhibit a low probability of surgical cure.[124,125] Most patients with macroadenomas have persistent elevated postsurgical GH. Surgical debulking may enable long-term efficacy of SRLs in attaining biochemical control.[95] Oral octreotide may be particularly effective in preventing breakthrough symptoms that occur toward the end of injection cycles as injectable drug levels wane.[126,127]

Further indications that have not been verified by prospective controlled studies include the use of preoperative SRL therapy to enhance subsequent surgical results. Whether or not sustained long-term SRL therapy with or without combination with a GHRA to tightly control GH and IGF-1 levels improves ultimate mortality outcomes is not yet proven.

Dopamine Agonists. High doses of cabergoline have been used in management, but long-term results are not compelling.[128] After 18 weeks of dose titration, cabergoline did not significantly reduce mean IGF-1 levels in 24 patients and only normalized IGF-1 levels in two patients.[129] Because high doses are required to achieve even moderate efficacy, the incidence of adverse events is far higher than that usually seen when treating patients with prolactinomas. Patients receiving high doses of dopamine agonists complain of gastrointestinal symptoms, including nausea, vomiting, and abdominal cramps. Rarely, arrhythmias have been reported. Nasal stuffiness and sleep disturbances are common complaints.

Growth Hormone Receptor Antagonist. Pegvisomant, a GHRA, directly inhibits GH action in the periphery.[130] Unlike SRLs and dopamine agonists that act centrally to inhibit GH secretion through somatotroph cell somatostatin and dopamine receptors, pegvisomant interferes with GH receptor signaling, thereby suppressing peripheral IGF-1 generation in almost all patients with GH-secreting pituitary tumors treated for up to 36 months. Pegvisomant binds one GH receptor unit at site 1 but cannot bind the mutated site 2.[130] The site 2 mutation in pegvisomant blocks initiation of subsequent GH signal transduction.

In an initial study, daily doses of pegvisomant (10, 15, or 20 mg) given for 12 weeks normalized IGF-1 levels in 38%, 75%, and 82% of patients with acromegaly, respectively[131] (Fig. 6.8). Subsequent longer-term follow-up in ACROSTUDY patients monitored for 5 years reported IGF-1 normalization in 63% of patients with a mean dose of 18 mg/day.[132] Concomitant dose-dependent reduction in serum IGF-1 levels is accompanied by dose-dependent regression of soft-tissue swelling, excessive perspiration, and fatigue, with no significant improvement in arthralgias or headache. Pegvisomant also improves insulin sensitivity and glucose tolerance, reducing fasting serum insulin and glucose levels and decreasing glycated hemoglobin levels.

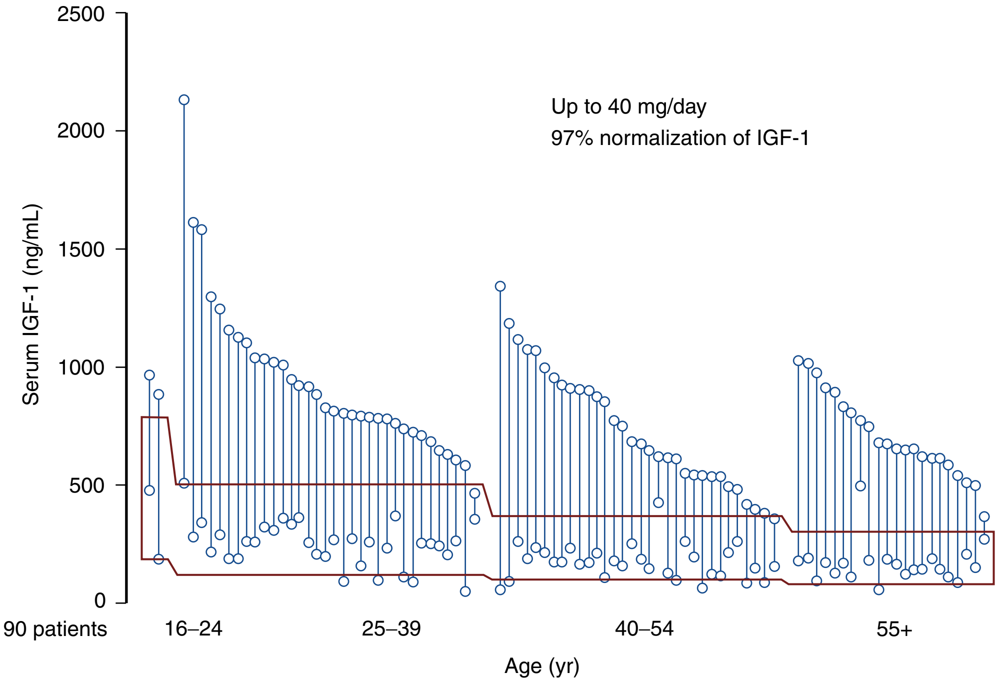


Fig. 6.8 Insulin-like growth factor 1 (IGF-1) levels at baseline and after 12 months of pegvisomant therapy. (Data from van der Lely AJ, Hutson RK, Trainer PJ, et al. Long-term treatment of acromegaly with pegvisomant, a growth hormone-receptor antagonist. *Lancet*. 2001; 358:1754-1759.)

Short-term (<3 months of treatment) side effects encountered in pegvisomant-treated patients included reversible injection-site reactions (11%) and diarrhea (14%) in an 18-month study.[133] GH levels are reversibly increased by approximately 2-fold, mirroring the IGF-1 decrease. Despite detection of anti-GH antibodies in 17% of patients, no evidence of tachyphylaxis was found. Pegvisomant also enhances insulin sensitivity in patients exhibiting glucose intolerance. In a follow-up of 1288 subjects treated for a mean of 3.7 years, serious adverse events were reported in 12% of subjects. Persistent pituitary tumor growth is observed infrequently (~3% of patients), and transiently elevated (3 times the upper limit of normal) liver transaminase levels are observed in up to 2.5% of patients, with rarely reported drug-induced hepatitis[132]. Other side effects include lipohypertrophy[134] and features of hypopituitarism if IGF-1 levels are attenuated to below normal ranges.

GH levels are elevated in patients receiving pegvisomant therapy, so IGF-1 levels are measured to monitor efficacy, emphasizing the need to standardize commercial IGF-1 assays and develop age-specific reference ranges. IGF-1 levels should be titrated to midnormal ranges. Possible overtreatment with pegvisomant may create clinical features of GH deficiency, as pegvisomant may suppress IGF-1 levels below the lower limit of normal.

Pegvisomant alone or in combination with SRL can normalize IGF-1 levels in SRL-resistant patients[46] with weekly, twice-weekly, or daily pegvisomant injections and SRL cotreatment,[135,136] and less frequent dosing may be more cost effective[137] (Fig. 6.9). Combination with cabergoline may also offer additive efficacy.[129]

INTEGRATED TREATMENT APPROACH TO THE MANAGEMENT OF ACROMEGALY

Patients With Likelihood of Good Surgical Outcome

After diagnosis, the likelihood of surgical cure is assessed. For small, well-circumscribed tumors, surgical excision by an experienced pituitary surgeon is the treatment of choice.[1] Surgical cure rates are maximal for noninvasive, well-encapsulated smaller tumors.[88,89] After surgery, patients are monitored to ensure that GH responses to a glucose load are less than 1 μg/L and that IGF-1 levels are normalized. If, however, after surgery, hormone levels are not controlled, which is indicative of disease persistence or recurrence, SRLs are indicated.[46] Pegvisomant as monotherapy or combined with an SRL is indicated after failure of other therapy. Patients should be followed by measuring age-matched IGF-1 levels, aiming for the midnormal range levels. Liver function tests should be evaluated monthly during the first 6 months of pegvisomant therapy. If the patient is still not biochemically controlled, a dopamine agonist is added, or reoperation or irradiation is considered.

Patients With a Poor Likelihood of Successful Surgical Outcome

Patients with large adenomas are likely to have a poor surgical outcome, and fewer than half will be biochemically controlled. Most newly diagnosed patients have large macroadenomas, which portend a poor surgical outcome with persistent postoperative GH hypersecretion.[92] They can be offered primary medical treatment, as would patients in whom surgery is contraindicated or those who decline surgery.[138] Use of preoperative octreotide to enhance subsequent surgical outcomes has been advocated,[115] as the drug reduces tumor bulk. If biochemical control is not achieved, drug efficacy may be improved by dose increases, use of monotherapy or combination GHRA therapy, or addition of a dopamine agonist.[46] If patients are resistant to medical treatment, second surgery or irradiation is indicated, depending on the size and location of the tumor remnant and the skill of the neurosurgeon.

KEY POINTS

- Long-term surveillance and management are required to achieve optimal control of comorbidities and mortality.

Follow-Up

Laboratory follow-up of patients includes attaining stringent responses of GH less than 1 μg/L, accompanied by a normalized IGF-1 level. If these goals are not achieved, medical treatment should be initiated. If already being administered, efficacy may be improved by increasing medication dose frequency, combining a GHRA with an SRL, using GHRA as monotherapy, or adding a dopamine agonist. Reoperation and radiation therapy are further adjuvant options for resistant patients. In patients with tumors that have been effectively excised,

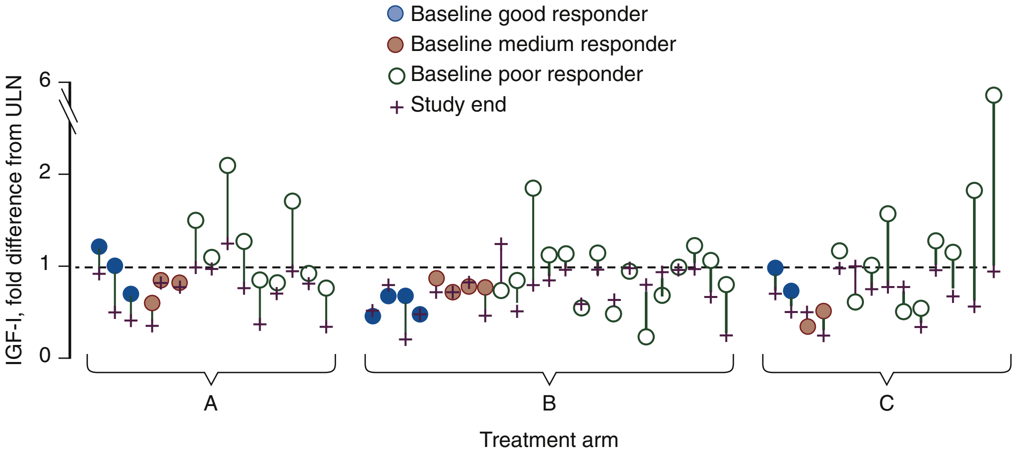


Fig. 6.9 Fold difference from upper limit of normal in IGF-1 at baseline and study end in 52 participants who completed at least 24 weeks combination SRL plus pegvisomant therapy stratified as good, medium, or poor responders to prior SRL and grouped by treatment arm. Arm A, high-dose SRL plus weekly pegvisomant; arm B, low-dose SRL plus weekly pegvisomant; arm C, low-dose SRL plus daily pegvisomant. *IGF-1*, Insulin-like growth factor 1; *SRL*, somatostatin receptor ligand; *ULN*, upper limit of normal. (Reprinted with permission from Bonert V, Mirocha J, Carmichael J, et al. Cost-effectiveness and efficacy of a novel combination regimen in acromegaly: a prospective, randomized trial. *J Clin Endocrinol Metab*. 2020; 105:e3236-e3245.)

serial MRI is warranted only once every 2 years after surgery. Invasive residual tumor tissue mass or recurrent tumors require more frequent MRI evaluation. Although tumor mass may not invariably shrink with SRL therapy, further progressive tumor growth rarely occurs while patients are taking the medication.

Additional Management Considerations

The clinical, economic, and quality of life burden for these patients is significant, and optimizing disease control to improve adverse outcomes is imperative.[126,127,139] Patients with acromegaly require management of multiple associated comorbidities. Colonoscopy should be performed at diagnosis. Because cardiovascular morbidity is so high, aggressive management of hypertension, left ventricular hypertrophy, cardiac failure, and arrhythmias should be pursued. Pulmonary function and sleep evaluation should be undertaken in all patients early in the course of the disorder, and debilitating arthritis requires aggressive rheumatologic management. Screening and therapy for insulin resistance and diabetes are important, and SRLs usually improve diabetes control dramatically, except for pasireotide, which usually exacerbates hyperglycemia. Insulin requirements may immediately decrease to 90% of pretreatment needs as GH is effectively suppressed. Headache is an extremely common symptom and usually improves with SRLs; if not, potent analgesics may be indicated. Maxillofacial disorders may require dental, maxillary, and facial cosmetic surgery. Fertility is commonly of concern to patients, and reports of successful pregnancies in women treated with octreotide provide optimistic guidelines for pregnancy management.[140] As patients with acromegaly may be depressed and have low self-esteem and other psychosocial sensitivity,[133] individual or group counseling may be indicated. Finally, all available treatment modes for acromegaly are associated with therapy-specific complications, which, if they arise, should be promptly addressed.

Long-acting depot and oral preparations of SRLs as well as GHRA have changed the approach to management, inasmuch as efficacy, patient compliance and medication safety are attained.

REFERENCES

1. Melmed S. Medical progress: acromegaly. *N Engl J Med*. 2006;355:2558–2573.
2. Ribeiro-Oliveira Jr A, Barkan A. The changing face of acromegaly--advances in diagnosis and treatment. *Nat Rev Endocrinol*. 2012;8:605–611.
3. Colao A, Grasso LFS, Giustina A, et al. Acromegaly. *Nat Rev Dis Primers*. 2019;5:20.
4. Melmed S. Pituitary-tumor endocrinopathies. *N Engl J Med*. 2020;382:937–950.
5. Yamashita S, Melmed S. Insulinlike growth factor I regulation of growth hormone gene transcription in primary rat pituitary cells. *J Clin Invest*. 1987;79:449–452.
6. Melmed S. Pathogenesis and diagnosis of growth hormone deficiency in adults. *N Engl J Med*. 2019;380:2551–2562.
7. Rotwein P. Mapping the growth hormone--Stat5b--IGF-I transcriptional circuit. *Trends Endocrinol Metab*. 2012;23:186–193.
8. Clemmons DR. Metabolic actions of insulin-like growth factor-I in normal physiology and diabetes. *Endocrinol Metab Clin North Am*. 2012;41:425–443 (vii-viii).
9. Melmed S, Braunstein GD, Horvath E, et al. Pathophysiology of acromegaly. *Endocr Rev*. 1983;4:271–290.
10. Melmed S. Pathogenesis of pituitary tumors. *Nat Rev Endocrinol*. 2011;7:257–266.
11. Cuevas-Ramos D, Carmichael JD, Cooper O, et al. A structural and functional acromegaly classification. *J Clin Endocrinol Metab*. 2015;100:122–131.
12. Weinstein LS, Shenker A, Gejman PV, et al. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*. 1991;325:1688–1695.
13. Melmed S, Ezrin C, Kovacs K, et al. Acromegaly due to secretion of growth hormone by an ectopic pancreatic islet-cell tumor. *N Engl J Med*. 1985;312:9–17.
14. Sano T, Asa SL, Kovacs K. Growth hormone-releasing hormone-producing tumors: clinical, biochemical, and morphological manifestations. *Endocr Rev*. 1988;9:357–373.
15. Thorner MO, Perryman RL, Cronin MJ, et al. Somatotroph hyperplasia. Successful treatment of acromegaly by removal of a pancreatic islet tumor secreting a growth hormone-releasing factor. *J Clin Invest*. 1982;70:965–977.
16. Mayo KE, Hammer RE, Swanson LW, et al. Dramatic pituitary hyperplasia in transgenic mice expressing a human growth hormone-releasing factor gene. *Mol Endocrinol*. 1988;2:606–612.
17. Bertherat J, Chanson P, Montminy M. The cyclic adenosine 3',5'-monophosphate-responsive factor CREB is constitutively activated in human somatotroph adenomas. *Mol Endocrinol*. 1995;9:777–783.
18. Ben-Shlomo A, Deng N, Ding E, et al. DNA damage and growth hormone hypersecretion in pituitary somatotroph adenomas. *J Clin Invest*. 2020;130:5738–5755.
19. Herman V, Fagin J, Gonsky R, et al. Clonal origin of pituitary adenomas. *J Clin Endocrinol Metab*. 1990;71:1427–1433.
20. Thakker RV, Pook MA, Wooding C, et al. Association of somatotrophinomas with loss of alleles on chromosome 11 and with gsp mutations. *J Clin Invest*. 1993;91:2815–2821.
21. Pei L, Melmed S, Scheithauer B, et al. Frequent loss of heterozygosity at the retinoblastoma susceptibility gene (RB) locus in aggressive pituitary tumors: evidence for a chromosome 13 tumor suppressor gene other than RB. *Cancer Res*. 1995;55:1613–1616.
22. Vierimaa O, Georgitsi M, Lehtonen R, et al. Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science*. 2006;312:1228–1230.
23. Beckers A, Lodish MB, Trivellin G, et al. X-linked acrogigantism syndrome: clinical profile and therapeutic responses. *Endocr Relat Cancer*. 2015;22:353–367.
24. Raitila A, Lehtonen HJ, Arola J, et al. Mice with inactivation of aryl hydrocarbon receptor-interacting protein (Aip) display complete penetrance of pituitary adenomas with aberrant ARNT expression. *Am J Pathol*. 2010;177:1969–1976.
25. Spada A, Lania A, Ballare E. G protein abnormalities in pituitary adenomas. *Mol Cell Endocrinol*. 1998;142:1–14.
26. Abboud D, Daly AF, Dupuis N, et al. GPR101 drives growth hormone hypersecretion and gigantism in mice via constitutive activation of Gs and Gq/11. *Nat Commun*. 2020;11:4752.
27. Trivellin G, Daly AF, Faucz FR, et al. Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. *N Engl J Med*. 2014;371:2363–2374.
28. Karga HJ, Alexander JM, Hedley-Whyte ET, et al. Ras mutations in human pituitary tumors. *J Clin Endocrinol Metab*. 1992;74:914–919.
29. Pei L, Melmed S, Scheithauer B, et al. H-ras mutations in human pituitary carcinoma metastases. *J Clin Endocrinol Metab*. 1994;78:842–846.
30. Pei L, Melmed S. Isolation and characterization of a pituitary tumor-transforming gene (PTTG). *Mol Endocrinol*. 1997;11:433–441.
31. Melmed S. Mechanisms for pituitary tumorigenesis: the plastic pituitary. *J Clin Invest*. 2003;112:1603–1618.
32. Zhang X, Horwitz GA, Heaney AP, et al. Pituitary tumor transforming gene (PTTG) expression in pituitary adenomas. *J Clin Endocrinol Metab*. 1999;84:761–767.
33. Quereda V, Malumbres M. Cell cycle control of pituitary development and disease. *J Mol Endocrinol*. 2009;42:75–86.
34. Chesnokova V, Zonis S, Kovacs K, et al. p21(Cip1) restrains pituitary tumor growth. *Proc Natl Acad Sci U S A*. 2008;105:17498–17503.
35. Stratakis CA. Carney complex: a familial lentiginosis predisposing to a variety of tumors. *Rev Endocr Metab Disord*. 2016;17:367–371.
36. Teh BT, Kytola S, Farnebo F, et al. Mutation analysis of the MEN1 gene in multiple endocrine neoplasia type 1, familial acromegaly and familial isolated hyperparathyroidism. *J Clin Endocrinol Metab*. 1998;83:2621–2626.
37. Daly AF, Vanbellinghen JF, Khoo SK, et al. Aryl hydrocarbon receptor-interacting protein gene mutations in familial isolated pituitary adenomas: analysis in 73 families. *J Clin Endocrinol Metab*. 2007;92:1891–1896.

38. Georgitsi M, Raitila A, Karhu A, et al. Molecular diagnosis of pituitary adenoma predisposition caused by aryl hydrocarbon receptor-interacting protein gene mutations. *Proc Natl Acad Sci U S A*. 2007;104:4101–4105.
39. Fernandez A, Karavitaki N, Wass JA. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin Endocrinol*. 2010;72:377–382.
40. Broder MS, Chang E, Cherepanov D, et al. Incidence and prevalence of acromegaly in the United States: a claims-based analysis. *Endocr Pract*. 2016;22:1327–1335.
41. Reid TJ, Post KD, Bruce JN, et al. Features at diagnosis of 324 patients with acromegaly did not change from 1981 to 2006: acromegaly remains under-recognized and under-diagnosed. *Clin Endocrinol*. 2010;72:203–208.
42. Nachtigall L, Delgado A, Swearingen B, et al. Changing patterns in diagnosis and therapy of acromegaly over two decades. *J Clin Endocrinol Metab*. 2008;93:2035–2041.
43. Lavrentaki A, Paluzzi A, Wass JA, et al. Epidemiology of acromegaly: review of population studies. *Pituitary*. 2017;20:4–9.
44. Clemmons DR, Van Wyk JJ, Ridgway EC, et al. Evaluation of acromegaly by radioimmunoassay of somatomedin-C. *N Engl J Med*. 1979;301:1138–1142.
45. Carmichael JD, Bonert VS, Mirocha JM, et al. The utility of oral glucose tolerance testing for diagnosis and assessment of treatment outcomes in 166 patients with acromegaly. *J Clin Endocrinol Metab*. 2009;94:523–527.
46. Melmed S, Bronstein MD, Chanson P, et al. A Consensus Statement on acromegaly therapeutic outcomes. *Nat Rev Endocrinol*. 2018;14:552–561.
47. Fleseriu M, Biller BMK, Freda PU, et al. A Pituitary Society update to acromegaly management guidelines. *Pituitary*. 2021;24:1–13.
48. Barkan AL, Beitins IZ, Kelch RP. Plasma insulin-like growth factor-I/somatomedin-C in acromegaly: correlation with the degree of growth hormone hypersecretion. *J Clin Endocrinol Metab*. 1988;67:69–73.
49. Faje AT, Barkan AL. Basal, but not pulsatile, growth hormone secretion determines the ambient circulating levels of insulin-like growth factor-I. *J Clin Endocrinol Metab*. 2010;95:2486–2491.
50. Arafat AM, Mohlig M, Weickert MO, et al. Growth hormone response during oral glucose tolerance test: the impact of assay method on the estimation of reference values in patients with acromegaly and in healthy controls, and the role of gender, age, and body mass index. *J Clin Endocrinol Metab*. 2008;93:1254–1262.
51. Beuschlein F, Strasburger CJ, Siegerstetter V, et al. Acromegaly caused by secretion of growth hormone by a non-Hodgkin's lymphoma. *N Engl J Med*. 2000;342:1871–1876.
52. Melmed S, Ziel FH, Braunstein GD, et al. Medical management of acromegaly due to ectopic production of growth hormone-releasing hormone by a carcinoid tumor. *J Clin Endocrinol Metab*. 1988;67:395–399.
53. Molitch ME. Diagnosis and treatment of pituitary adenomas: a review. *J Am Med Assoc*. 2017;317:516–524.
54. Schneider HJ, Kosilek RP, Gunther M, et al. A novel approach to the detection of acromegaly: accuracy of diagnosis by automatic face classification. *J Clin Endocrinol Metab*. 2011;96:2074–2080.
55. Syro LV, Rotondo F, Ramirez A, et al. Progress in the diagnosis and classification of pituitary adenomas. *Front Endocrinol*. 2015;6:97.
56. Mazziotti G, Frara S, Giustina A. Pituitary diseases and bone. *Endocr Rev*. 2018;39:440–488.
57. Mazziotti G, Lania AGA, Canalis E. Management of endocrine disease: bone disorders associated with acromegaly: mechanisms and treatment. *Eur J Endocrinol*. 2019;181:R45-R56.
58. Gadelha MR, Kasuki L, Lim DST, et al. Systemic complications of acromegaly and the impact of the current treatment landscape: an update. *Endocr Rev*. 2019;40:268–332.
59. Constantin T, Tangpricha V, Shah R, et al. Calcium and bone turnover markers in acromegaly: a prospective, controlled study. *J Clin Endocrinol Metab*. 2017;102:2416–2424.
60. Mazziotti G, Bianchi A, Porcelli T, et al. Vertebral fractures in patients with acromegaly: a 3-year prospective study. *J Clin Endocrinol Metab*. 2013;98:3402–3410.
61. Ben-Shlomo A, Melmed S. Skin manifestations in acromegaly. *Clin Dermatol*. 2006;24:256–259.
62. Melmed S. Acromegaly pathogenesis and treatment. *J Clin Invest*. 2009;119:3189–3202.
63. Colao A, Ferone D, Marzullo P, et al. Systemic complications of acromegaly: epidemiology, pathogenesis, and management. *Endocr Rev*. 2004;25:102–152.
64. Freda PU, Shen W, Reyes-Vidal CM, et al. Skeletal muscle mass in acromegaly assessed by magnetic resonance imaging and dual-photon x-ray absorptiometry. *J Clin Endocrinol Metab*. 2009;94:2880–2886.
65. Pereira AM, van Thiel SW, Lindner JR, et al. Increased prevalence of regurgitant valvular heart disease in acromegaly. *J Clin Endocrinol Metab*. 2004;89:71–75.
66. Berg C, Petersenn S, Lahner H, et al. Cardiovascular risk factors in patients with uncontrolled and long-term acromegaly: comparison with matched data from the general population and the effect of disease control. *J Clin Endocrinol Metab*. 2010;95:3648–3656.
67. Pivonello R, Auriemma RS, Grasso LF, et al. Complications of acromegaly: cardiovascular, respiratory and metabolic comorbidities. *Pituitary*. 2017;20:46–62.
68. Kamenicky P, Viengchareun S, Blanchard A, et al. Epithelial sodium channel is a key mediator of growth hormone-induced sodium retention in acromegaly. *Endocrinology*. 2008;149:3294–3305.
69. Parolin M, Dassie F, Alessio L, et al. Obstructive sleep apnea in acromegaly and the effect of treatment: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2020;105.
70. Giustina A, Barkan A, Beckers A, et al. A consensus on the diagnosis and treatment of acromegaly comorbidities: an update. *J Clin Endocrinol Metab*. 2020;105:e937–e946.
71. Attal P, Chanson P. Endocrine aspects of obstructive sleep apnea. *J Clin Endocrinol Metab*. 2010;95:483–495.
72. Grunstein RR, Ho KK, Sullivan CE. Effect of octreotide, a somatostatin analog, on sleep apnea in patients with acromegaly. *Ann Intern Med*. 1994;121:478–483.
73. Wolters TLC, Roerink S, Drenthen LCA, et al. The course of obstructive sleep apnea syndrome in patients with acromegaly during treatment. *J Clin Endocrinol Metab*. 2020;105:290–304.
74. Petrossians P, Daly AF, Natchev E, et al. Acromegaly at diagnosis in 3173 patients from the Liege Acromegaly Survey (LAS) database. *Endocr Relat Cancer*. 2017;24:505–518.
75. Hannon AM, Thompson CJ, Sherlock M. Diabetes in patients with acromegaly. *Curr Diab Rep*. 2017;17:8.
76. Terzolo M, Reimondo G, Gasperi M, et al. Colonoscopic screening and follow-up in patients with acromegaly: a multicenter study in Italy. *J Clin Endocrinol Metab*. 2005;90:84–90.
77. Resmini E, Tagliafico A, Bacigalupo L, et al. Computed tomography colonography in acromegaly. *J Clin Endocrinol Metab*. 2009;94:218–222.
78. Rokkas T, Pistiolas D, Sechopoulos P, et al. Risk of colorectal neoplasm in patients with acromegaly: a meta-analysis. *World J Gastroenterol*. 2008;14:3484–3489.
79. Mercado M, Gonzalez B, Vargas G, et al. Successful mortality reduction and control of comorbidities in patients with acromegaly followed at a highly specialized multidisciplinary clinic. *J Clin Endocrinol Metab*. 2014;99:4438–4446.
80. Ritvonen E, Loyttyniemi E, Jaatinen P, et al. Mortality in acromegaly: a 20-year follow-up study. *Endocr Relat Cancer*. 2016;23:469–480.
81. Biermasz NR, Dekker FW, Pereira AM, et al. Determinants of survival in treated acromegaly in a single center: predictive value of serial insulin-like growth factor I measurements. *J Clin Endocrinol Metab*. 2004;89:2789–2796.
82. Ayuk J, Clayton RN, Holder G, et al. Growth hormone and pituitary radiotherapy, but not serum insulin-like growth factor-I concentrations, predict excess mortality in patients with acromegaly. *J Clin Endocrinol Metab*. 2004;89:1613–1617.
83. Dekkers OM, Biermasz NR, Pereira AM, et al. Mortality in acromegaly: a metaanalysis. *J Clin Endocrinol Metab*. 2008;93:61–67.
84. Sherlock M, Reulen RC, Alonso AA, et al. ACTH deficiency, higher doses of hydrocortisone replacement, and radiotherapy are independent predictors of mortality in patients with acromegaly. *J Clin Endocrinol Metab*. 2009;94:4216–4223.

85. Kauppinen-Makelin R, Sane T, Reunanen A, et al. A nationwide survey of mortality in acromegaly. *J Clin Endocrinol Metab*. 2005;90:4081–4086.
86. Holdaway IM, Bolland MJ, Gamble GD. A meta-analysis of the effect of lowering serum levels of GH and IGF-I on mortality in acromegaly. *Eur J Endocrinol*. 2008;159:89–95.
87. Mamelak AN, Carmichael J, Bonert VH, et al. Single-surgeon fully endoscopic endonasal transsphenoidal surgery: outcomes in three-hundred consecutive cases. *Pituitary*. 2013;16:393–401.
88. Starke RM, Raper DM, Payne SC, et al. Endoscopic vs microsurgical transsphenoidal surgery for acromegaly: outcomes in a concurrent series of patients using modern criteria for remission. *J Clin Endocrinol Metab*. 2013;98:3190–3198.
89. Jane Jr JA, Starke RM, Elzoghby MA, et al. Endoscopic transsphenoidal surgery for acromegaly: remission using modern criteria, complications, and predictors of outcome. *J Clin Endocrinol Metab*. 2011;96:2732–2740.
90. Chen CJ, Ironside N, Pomeraniec IJ, et al. Microsurgical versus endoscopic transsphenoidal resection for acromegaly: a systematic review of outcomes and complications. *Acta Neurochir*. 2017;159:2193–2207.
91. Agrawal N, Ioachimescu AG. Prognostic factors of biochemical remission after transsphenoidal surgery for acromegaly: a structured review. *Pituitary*. 2020;23:582–594.
92. Mercado M, Abreu C, Vergara-Lopez A, et al. Surgical and pharmacological outcomes in acromegaly: real-life data from the Mexican Acromegaly Registry. *J Clin Endocrinol Metab*. 2020;105.
93. Kim EH, Oh MC, Lee EJ, et al. Predicting long-term remission by measuring immediate postoperative growth hormone levels and oral glucose tolerance test in acromegaly. *Neurosurgery*. 2012;70:1106–1113; discussion 1113.
94. Colao A, Attanasio R, Pivonello R, et al. Partial surgical removal of growth hormone-secreting pituitary tumors enhances the response to somatostatin analogs in acromegaly. *J Clin Endocrinol Metab*. 2006;91:85–92.
95. Jallad RS, Musolino NR, Kodaira S, et al. Does partial surgical tumour removal influence the response to octreotide-LAR in acromegalic patients previously resistant to the somatostatin analogue? *Clin Endocrinol*. 2007;67:310–315.
96. Freda PU, Post KD, Powell JS, et al. Evaluation of disease status with sensitive measures of growth hormone secretion in 60 postoperative patients with acromegaly. *J Clin Endocrinol Metab*. 1998;83:3808–3816.
97. Singh R, Didwania P, Lehrer EJ, et al. Stereotactic radiosurgery for acromegaly: an international systematic review and meta-analysis of clinical outcomes. *J Neuro Oncol*. 2020;148:401–418.
98. Knappe UJ, Petroff D, Quinkler M, et al. Fractionated radiotherapy and radiosurgery in acromegaly: analysis of 352 patients from the German Acromegaly Registry. *Eur J Endocrinol*. 2020;182:275–284.
99. Giustina A, Chanson P, Bronstein MD, et al. A consensus on criteria for cure of acromegaly. *J Clin Endocrinol Metab*. 2010;95:3141–3148.
100. Ayuk J. Does pituitary radiotherapy increase the risk of stroke and, if so, what preventative actions should be taken? *Clin Endocrinol*. 2012;76:328–331.
101. Lamberts SW, van der Lely AJ, de Herder WW, et al. Octreotide. *N Engl J Med*. 1996;334:246–254.
102. Murray RD, Kim K, Ren SG, et al. Central and peripheral actions of somatostatin on the growth hormone-IGF-I axis. *J Clin Invest*. 2004;114:349–356.
103. Gadelha MR, Wildemberg LE, Bronstein MD, et al. Somatostatin receptor ligands in the treatment of acromegaly. *Pituitary*. 2017;20:100–108.
104. Shimon I, Yan X, Taylor JE, et al. Somatostatin receptor (SSTR) subtype-selective analogues differentially suppress in vitro growth hormone and prolactin in human pituitary adenomas. Novel potential therapy for functional pituitary tumors. *J Clin Invest*. 1997;100:2386–2392.
105. Colao A, Ferone D, Marzullo P, et al. Long-term effects of depot long-acting somatostatin analog octreotide on hormone levels and tumor mass in acromegaly. *J Clin Endocrinol Metab*. 2001;86:2779–2786.
106. Attanasio R, Baldelli R, Pivonello R, et al. Lanreotide 60 mg, a new long-acting formulation: effectiveness in the chronic treatment of acromegaly. *J Clin Endocrinol Metab*. 2003;88:5258–5265.
107. Melmed S, Cook D, Schopohl J, et al. Rapid and sustained reduction of serum growth hormone and insulin-like growth factor-1 in patients with acromegaly receiving lanreotide Autogel therapy: a randomized, placebo-controlled, multicenter study with a 52 week open extension. *Pituitary*. 2010;13:18–28.
108. Tuvia S, Atsmon J, Teichman SL, et al. Oral octreotide absorption in human subjects: comparable pharmacokinetics to parenteral octreotide and effective growth hormone suppression. *J Clin Endocrinol Metab*. 2012;97:2362–2369.
109. Melmed S, Popovic V, Bidlingmaier M, et al. Safety and efficacy of oral octreotide in acromegaly: results of a multicenter phase III trial. *J Clin Endocrinol Metab*. 2015;100:1699–1708.
110. Samson SL, Nachtigall LB, Fleseriu M, et al. Maintenance of acromegaly control in patients switching from injectable somatostatin receptor ligands to oral octreotide. *J Clin Endocrinol Metab*. 2020;105:e3785–3797.
111. Carmichael JD, Bonert VS, Nuno M, et al. Acromegaly clinical trial methodology impact on reported biochemical efficacy rates of somatostatin receptor ligand treatments: a meta-analysis. *J Clin Endocrinol Metab*. 2014;99:1825–1833.
112. Mercado M, Borges F, Bouterfa H, et al. A prospective, multicentre study to investigate the efficacy, safety and tolerability of octreotide LAR (long-acting repeatable octreotide) in the primary therapy of patients with acromegaly. *Clin Endocrinol*. 2007;66:859–868.
113. Giustina A, Mazziotti G, Torri V, et al. Meta-analysis on the effects of octreotide on tumor mass in acromegaly. *PloS One*. 2012;7:e36411.
114. Shen M, Shou X, Wang Y, et al. Effect of presurgical long-acting octreotide treatment in acromegaly patients with invasive pituitary macroadenomas: a prospective randomized study. *Endocr J*. 2010;57:1035–1044.
115. Carlsen SM, Lund-Johansen M, Schreiner T, et al. Preoperative octreotide treatment in newly diagnosed acromegalic patients with macroadenomas increases cure short-term postoperative rates: a prospective, randomized trial. *J Clin Endocrinol Metab*. 2008;93:2984–2990.
116. Fougner SL, Borota OC, Berg JP, et al. The clinical response to somatostatin analogues in acromegaly correlates to the somatostatin receptor subtype 2a protein expression of the adenoma. *Clin Endocrinol*. 2008;68:458–465.
117. Plockinger U, Albrecht S, Mawrin C, et al. Selective loss of somatostatin receptor 2 in octreotide-resistant growth hormone-secreting adenomas. *J Clin Endocrinol Metab*. 2008;93:1203–1210.
118. Potorac I, Petrossians P, Daly AF, et al. T2-weighted MRI signal predicts hormone and tumor responses to somatostatin analogs in acromegaly. *Endocr Relat Cancer*. 2016;23:871–881.
119. Fougner SL, Lekva T, Borota OC, et al. The expression of E-cadherin in somatotroph pituitary adenomas is related to tumor size, invasiveness, and somatostatin analog response. *J Clin Endocrinol Metab*. 2010;95:2334–2342.
120. Kasuki L, Vieira Neto L, Wildemberg LE, et al. AIP expression in sporadic somatotropinomas is a predictor of the response to octreotide LAR therapy independent of SSTR2 expression. *Endocr Relat Cancer*. 2012;19:L25–L29.
121. Giustina A, Mazziotti G, Cannavo S, et al. High-dose and high-frequency lanreotide autogel in acromegaly: a randomized, multicenter study. *J Clin Endocrinol Metab*. 2017;102:2454–2464.
122. Colao A, Bronstein MD, Brue T, et al. Pasireotide for acromegaly: long-term outcomes from an extension to the Phase III PAOLA study. *Eur J Endocrinol*. 2020;182:583.
123. Schmid HA, Brue T, Colao A, et al. Effect of pasireotide on glucose- and growth hormone-related biomarkers in patients with inadequately controlled acromegaly. *Endocrine*. 2016;53:210–219.
124. Fleseriu M, Hoffman AR, Katznelson L, et al. American Association of Clinical Endocrinologists and American College of Endocrinology disease state clinical review: management of acromegaly patients: what is the role of pre-operative medical therapy? *Endocr Pract*. 2015;21:668–673.
125. Losa M, Bollerslev J. Pros and cons in endocrine practice: pre-surgical treatment with somatostatin analogues in acromegaly. *Endocrine*. 2016;52:451–457.
126. Strasburger CJ, Karavitaki N, Stormann S, et al. Patient-reported outcomes of parenteral somatostatin analogue injections in 195 patients with acromegaly. *Eur J Endocrinol*. 2016;174:355–362.

127. Geer EB, Sisco J, Adelman DT, et al. Patient reported outcome data from acromegaly patients treated with injectable somatostatin receptor ligands (SRLs) in routine clinical practice. *BMC Endocr Disord*. 2020;20:117.
128. Kuhn E, Chanson P. Cabergoline in acromegaly. *Pituitary*. 2017;20:121–128.
129. Higham CE, Atkinson AB, Aylwin S, et al. Effective combination treatment with cabergoline and low-dose pegvisomant in active acromegaly: a prospective clinical trial. *J Clin Endocrinol Metab*. 2012;97:1187–1193.
130. Kopchick JJ, Parkinson C, Stevens EC, et al. Growth hormone receptor antagonists: discovery, development, and use in patients with acromegaly. *Endocr Rev*. 2002;23:623–646.
131. Trainer PJ, Drake WM, Katznelson L, et al. Treatment of acromegaly with the growth hormone-receptor antagonist pegvisomant. *N Engl J Med*. 2000;342:1171–1177.
132. van der Lely AJ, Biller BM, Brue T, et al. Long-term safety of pegvisomant in patients with acromegaly: comprehensive review of 1288 subjects in ACROSTUDY. *J Clin Endocrinol Metab*. 2012;97:1589–1597.
133. van der Lely AJ, Hutson RK, Trainer PJ, et al. Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. *Lancet*. 2001;358:1754–1759.
134. Bonert VS, Kennedy L, Petersenn S, et al. Lipodystrophy in patients with acromegaly receiving pegvisomant. *J Clin Endocrinol Metab*. 2008;93:3515–3518.
135. Neggers SJ, Franck SE, de Rooij FW, et al. Long-term efficacy and safety of pegvisomant in combination with long-acting somatostatin analogs in acromegaly. *J Clin Endocrinol Metab*. 2014;99:3644–3652.
136. Strasburger CJ, Mattsson A, Wilton P, et al. Increasing frequency of combination medical therapy in the treatment of acromegaly with the GH receptor antagonist pegvisomant. *Eur J Endocrinol*. 2018;178:321–329.
137. Bonert V, Mirocha J, Carmichael J, et al. Cost-effectiveness and efficacy of a novel combination regimen in acromegaly: a prospective, randomized trial. *J Clin Endocrinol Metab*. 2020;105:e3236–e3245.
138. Cozzi R, Montini M, Attanasio R, et al. Primary treatment of acromegaly with octreotide LAR: a long-term (up to nine years) prospective study of its efficacy in the control of disease activity and tumor shrinkage. *J Clin Endocrinol Metab*. 2006;91:1397–1403.
139. Ben-Shlomo A, Sheppard MC, Stephens JM, et al. Clinical, quality of life, and economic value of acromegaly disease control. *Pituitary*. 2011;14:284–294.
140. Chanson P, Vialon M, Caron P. An update on clinical care for pregnant women with acromegaly. *Expert Rev Endocrinol Metab*. 2019;14:85–96.

7

Cushing Disease

Antoine Tabarin and Jean-Benoît Corcuff

OUTLINE

Epidemiology, 99
Etiology of Cushing Disease, 99
Pathophysiology, 102
Clinical Features, 102
  Clinical Aspects, 102
  Nonspecific Biochemical and Radiological Investigations, 105
  Clinical Spectrum of Cushing Disease. Who Should Be Tested for Cushing Syndrome?, 105
Biochemical Diagnosis of Cushing Syndrome, 106
  General Considerations, 106
  First-Line Biochemical Investigations, 107
  Second-Line Biochemical Investigations, 108
  Other Biochemical Investigations, 109
Establishing Adrenocorticotropic Hormone–Dependency of Cushing Syndrome, 109
Differential Diagnosis of Adrenocorticotropic Hormone–Dependent Cushing Syndrome, 110
  Hypercortisolemic States Without Cushing Syndrome, 110
  Differentiating Cushing disease From Ectopic Adrenocorticotropic Hormone Syndrome and Ectopic Corticotropin-Releasing Hormone Secretion, 111
Treatment of Cushing Disease, 116
  General Considerations, 116
  Treatment Options After Postoperative Recurrence or Persistence, 119
  Somatostatin Analogs, 122
  Dopamine Agonists, 122
  Glucocorticoid Receptor–Directed Drugs, 124
  Combination of Medical Treatments, 124
  Future Medical Treatments, 124
Prognosis For Cushing Disease: Persistent Morbidity and Mortality, 124
  Metabolic Syndrome and Cardiovascular Risk, 124
  Muscle, 124
  Bone Disease, 124
  Mood and Cognitive Function, 124
  Autoimmunity, 125
  Quality of Life, 125
  Mortality in Cushing Disease, 125
  Implications for Treatment and Long-Term Care, 125
Special Conditions, 125
  Cushing Disease as an Emergency Condition, 125
  Cyclic Cushing Syndrome, 126
  Aggressive Pituitary Corticotroph Adenomas and Corticotroph Carcinomas, 126
  Pediatric Cushing Disease, 126
  Cushing Disease and Pregnancy, 127
  Cushing Disease and Chronic Renal Failure, 127
Future Directions, 128

Cushing syndrome consists of a group of symptoms caused by chronic and excessive tissue exposure to glucocorticoids. Given the paramount importance of glucocorticoids in the regulation of numerous physiological functions, Cushing syndrome results in the disruption of multiple homeostatic processes and is responsible for a significant clinical burden in the majority of patients, including increased mortality and impaired health-related quality of life. Despite significant advances in our understanding of the disease and improvements in sophisticated diagnostic and therapeutic tools, Cushing disease still poses many challenges, ranging from its early diagnosis to the difficulty of curative surgery and limiting the long-term sequelae, decreased quality of life, and decreased life expectancy.

EPIDEMIOLOGY

Cushing syndrome is a rare condition. The most common cause remains iatrogenic, resulting from prolonged corticosteroid therapy. From a pathophysiological standpoint, endogenous Cushing syndrome can be divided into adrenocorticotropic hormone (ACTH)-dependent and ACTH-independent causes. ACTH dependent-causes represent 70% to 82% of cases of Cushing syndrome in adults, and among these Cushing disease accounts for 86% to 93% of cases (Fig. 7.1).[1-3] Precise determination of its epidemiology is limited by the small size of studies, but the incidence of Cushing disease is estimated to be between 1.8 and 2.6/million population/year.[3,4] The incidence may have increased in the last few decades owing to increased medical awareness and refinement of biochemical diagnostic methods.[4] The prevalence of declared cases of Cushing disease in Biscaye, Spain in 1994 was 39.1 per million inhabitants.[5] The prevalence can also be estimated from community-based cross-sectional studies, regional databases and sellar mass registries, and national collections of clinically-relevant pituitary adenomas. A metaanalysis of studies conducted in geographically diverse populations[6] revealed a mean prevalence of pituitary adenomas of 89.1 per 100,000 (range: 76–116 per 100,000). Corticotroph adenomas responsible for Cushing disease accounted for 4% of these. The epidemiology of Cushing disease in children and pregnant women differs significantly from that in nonpregnant adults (see Special Conditions).

ETIOLOGY OF CUSHING DISEASE

Cushing disease is caused by tumoral proliferation of autonomously secreting pituitary corticotrophs. This proliferation depends on a

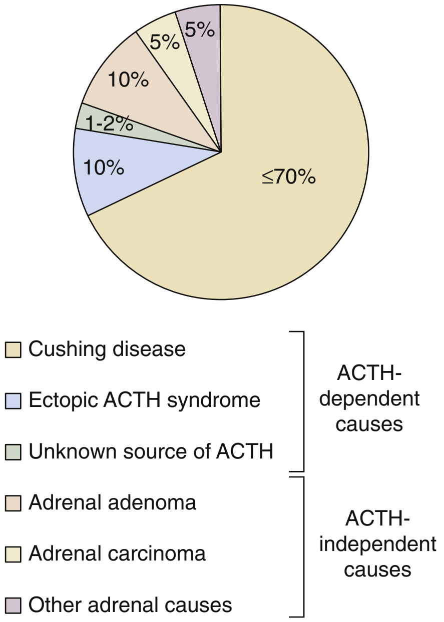


Fig. 7.1 Causes of endogenous Cushing syndrome. Approximate prevalence of the different etiologies of Cushing syndrome according to the literature. *ACTH,* Adrenocorticotropic hormone.

KEY POINTS

- Since 2014, thanks to the availability of modern exome sequencing, somatic mutations have been identified in more than half of the cases of corticotropic adenomas responsible for CD. Identifying these molecular alterations could pave the way for future medical treatments.

monoclonal expansion of cells. Although corticotroph hyperplasia has been reported in the periphery of adenomas, hyperplasia does not seem to be able, *per se,* to elicit a Cushing syndrome (apart from in the very rare context of ectopic corticotropin-releasing hormone [CRH] secretion).[1]

The fourth edition of the World Health Organization (WHO) classification of pituitary tumors takes into account cell lineage and changes in histopathology grading.[7] Applied to corticotropinomas, the classification identifies densely-granulated, sparsely-granulated, or Crooke cell adenomas.[7] Corticotropinomas are from a T-box pituitary restricted transcription factor (T-Pit) lineage, consequently showing specific T-Pit immunoreactivity. Although most corticotropinomas are small microadenomas, some are macroadenomas. Despite their rarity, corticotropinomas are the second most frequent origin of pituitary carcinomas (less than 1% of all pituitary neoplasms).[8] Two specific corticotropinomas are of particular concern for their frequent aggressive behavior: silent corticotroph adenomas and Crook cell adenomas.[7,9-11] Subsequent to the 2017 WHO classification, it has been proposed to change the nomenclature of pituitary adenomas to pituitary neuroendocrine tumors (PitNETs), to provide more uniformity in the classification of all neuroendocrine tumors despite organ-specific differences.[7] For example, investigating aggressive PitNETs using the same approach as has been used in neuroendocrine tumors (NETs), it was found that loss-of-function of the *ATRX* gene, which has been shown to predict malignant tumor phenotype in NETs, is highly prevalent in aggressive corticotroph tumors.[12]

Tumoral corticotrophs synthesize pro-opiomelanocortin (POMC) and secrete POMC-derived peptides, including ACTH, with no apparent qualitatively abnormal processing in most tumors responsible for Cushing disease. ACTH secretion in adenomas exhibits some resistance to glucocorticoid-induced negative feedback. However, no frequent gross abnormality in the glucocorticoid receptor has been demonstrated, and no clear reason for this resistance has yet been uncovered. A subset of corticotroph adenomas are characterized by the absence of Cushing syndrome and normal cortisol dynamics or elevated ACTH associated with mildly abnormal cortisol dynamics. These so-called silent corticotroph adenomas usually present as macroadenomas associated with mass-related symptoms. Various mechanisms may explain why an adenoma that synthesizes ACTH does not cause the Cushing syndrome.[13] These include polymorphisms in the *POMC* gene, predominant secretion of high–molecular weight ACTH, increased intracellular degradation of ACTH, and failure of exocytosis of ACTH from the cell membrane[13,14]: alterations in the processing of the prohormone POMC in corticotrophs is the most likely. Studies have demonstrated a huge decrease in the expression of prohormone convertase 1/3, an enzyme involved in the posttranslational processing of POMC into mature and biologically active ACTH, in silent corticotroph adenomas as compared to corticotroph adenomas associated with hypercortisolism. This abnormal processing, which has also been described in aggressive and corticotroph-secreting adenomas, may result in high circulating levels of POMC.[15] Interestingly, a subset of silent corticotroph adenomas can evolve into Cushing disease after years of inactivity, and this transformation is associated with the resumption of a robust expression of prohormone convertase 1/3.[13]

Until 2014, studies of the pathogenesis of corticotroph adenomas with hypothesis-driven molecular explorations were unsuccessful. Targeted approaches looking for gene mutations in known endocrine familial genetic disorders did not fare any better: only a minor fraction of cases of Cushing disease show alterations associated with germline mutations linked to pituitary adenomas (multiple endocrine neoplasia 1 (*MEN1*), cyclin-dependent kinase inhibitor 1B and 2C (*CDKN1B* and *CDKN2C*, associated with MEN4), and aryl-hydrocarbon receptor-interacting protein (*AIP*)).[16,17] Other potently oncogenic germline mutations of *GNAS, PRKAR1A*, succinate dehydrogenase subunits (*SDHs*), and *TSC1/2* have been occasionally described.[16,17] Disrupted microRNA production in DICER1 syndrome and germline mutations in genes associated with the mismatch repair pathway (*MLH1, PMS2, MSH2, MSH6*) leading to an autosomal dominant inherited cancer syndrome have also been rarely identified in patients with aggressively ACTH-secreting tumors.[16,17]

With the availability of high-throughput exome sequencing, seminal whole-exome studies of corticotroph adenomas identified somatic mutations in the deubiquitinase gene *USP8* in four out of 10[18] and eight out of 12 adenomas.[19] *USP8* gain-of-function mutations were confirmed by subsequent independent studies.[17] A meta-analysis of 677 patients showed that about 32% of female and 14% of male patients with corticotropinomas have a somatic *USP8* mutation.[20] Interestingly, no such mutations in *USP8* were found in other pituitary tumor subtypes. Somatic mutations in *USP8* are also common in Nelson tumors.[21] Lastly, a single case of a 16-year-old patient with a *de novo* germline *USP8* mutation with recurrent Cushing disease and multiple unrelated comorbidities was recently reported.[22]

From a pathophysiological point of view, USP8 is a deubiquitinase, that is, an enzyme that removes the ubiquitin tag that labels proteins for proteasomal or lysosomal degradation (Fig. 7.2). Phosphorylated USP8 binds to the 14-3-3 protein, in hibiting its activity. USP8 mutants fail to bind 14-3-3 protein and thus remain active. Active USP8 deubiquitinates numerous target proteins, thus protecting them from degradation. Epidermal growth factor receptor (EGFR) is an important such target in corticotrophs, stimulating POMC transcription via several pathways. Activating *USP8* mutations rescue the EGF receptor from lysosomal degradation, thus increasing both receptor expression and intracellular signaling. The demonstrated role of the EGFR pathway in *USP8*-mutated tumors presents opportunities for the use of EGFR inhibitors such as gefitinib.

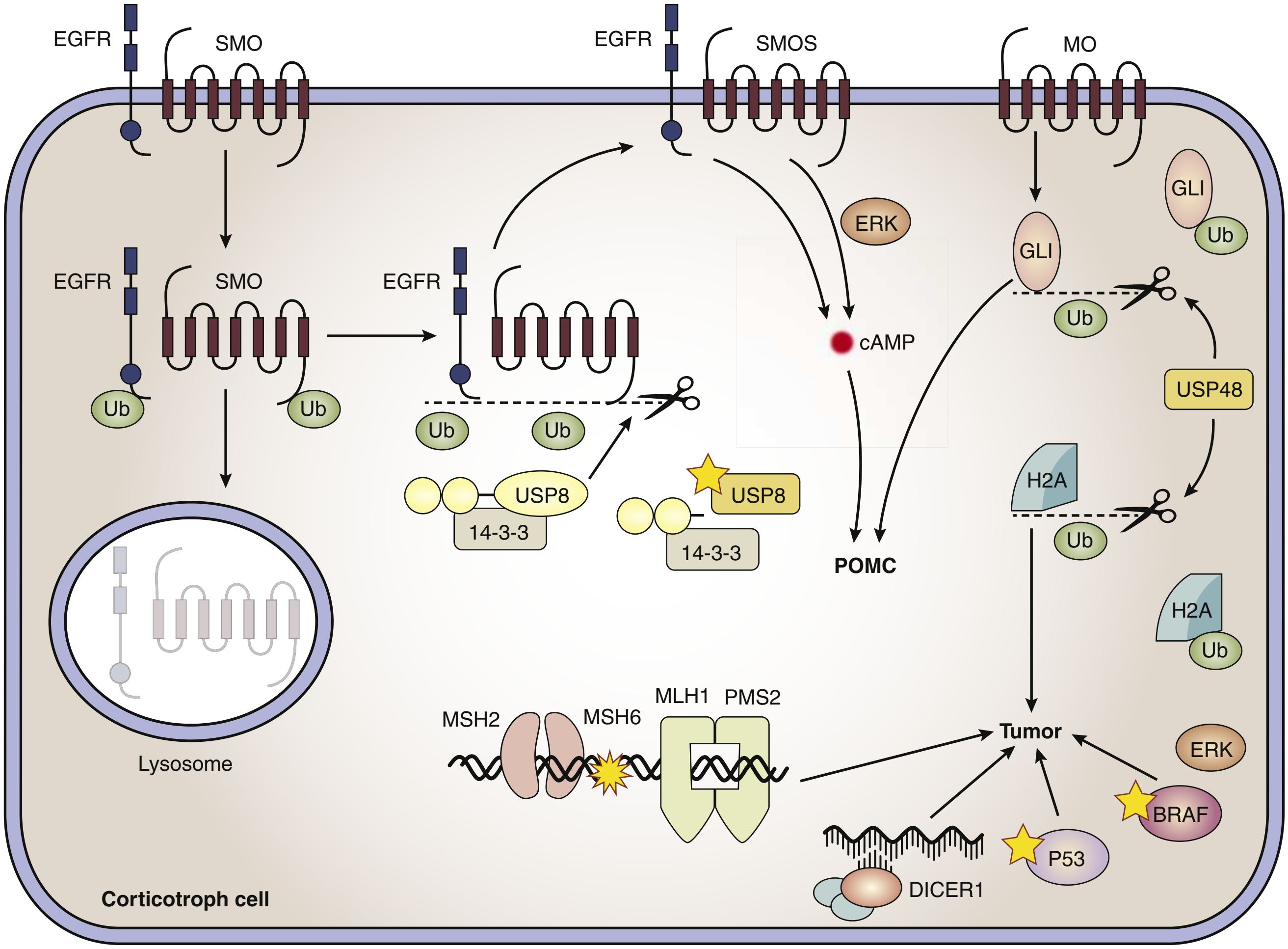


Fig. 7.2 Recurrent gene mutations found in corticotropinomas and their consequences. The deubiquitinase USP8 removes ubiquitin (*Ub*) tags from its targets, such as epidermal growth factor receptor (*EGFR*) and smoothened (*SMO*), preventing these receptors from being degraded in the proteasome and allowing them to be recycled back to the cell surface. Mutated USP8 displays increased enzymatic activity and thus leads to higher expression of the receptors on the cell membrane. Increased EGFR and SMO activity leads to increased cyclic adenosine monophosphate (*cAMP*) signaling and pro-opiomelanocortin (*POMC*) levels. Similarly, *GLI1* and *histone 2a (H2A)* may be targeted by USP48, leading to increased activity with *USP48* mutations. Loss-of-function mutations of *DICER1, TP53, MLH1,* and *MSH2* and gain-of-function mutations of *BRAF* have also been suggested to be sporadically associated with corticotroph tumorigenesis. (Reproduced with permission from Nadhamuni VS and Korbonits M. Novel insights into pituitary tumorigenesis: genetic and epigenetic mechanisms. *Endocrine Reviews.* 2020;41:1-26.)

Metaanalysis of the currently available literature concerning the genotype–phenotype correlations of *USP8* mutants reveals a number of discrepancies between series regarding the age of patients, size of adenomas, prevalence of mutations in microadenomas versus macroadenomas, intensity of ACTH and cortisol secretion, aggressiveness of adenomas, and probability of surgical failure.[20] The lack of a systematic protocol evaluating these patients and selection bias may account for this heterogeneity, preventing any definitive conclusion. Recent whole-genome studies shed new light on the characteristics of *USP8*-mutated adenomas. In a recent pangenomic analysis, *USP8*-wild-type corticotroph PitNETs appeared to be more aggressive than *USP8*-mutated PitNETs, with more frequent sphenoid invasion; this finding was supported by transcriptome analysis, which identified an increased epithelial-mesenchymal transition signature.[10] Similarly, whole-genome sequencing and genomic instability assessment of a series enriched in aggressive corticotroph tumors found that *USP8*-mutated tumors were less aggressive and less prevalent than *USP8*-wild-type tumors among silent corticotroph or Crooke's cell tumors. In accordance with this, most *USP8*-mutated tumors lack somatic inactivating *TP53* driver mutations and exhibit a nondisrupted genome.[11] Interestingly, increased somatostatin receptor 5 (*SSTR5*) expression was found in *USP8*-mutated corticotropinomas compared to their *USP8*-wild-type counterparts.[10] Whether *USP8*-mutated corticotropinomas respond better to the SSTR5 agonist pasireotide remains to be demonstrated.

Further exome-sequencing studies looking for driver mutations have revealed additional infrequent genetic alterations in corticotroph adenomas, including loss-of-function mutations of the glucocorticoid receptor gene in about 5% of *USP8*-mutated corticotropinomas.[17] Another deubiquitinase gene, *USP48*, was found to be mutated in about one-tenth to one-quarter of corticotropinomas.[17] *USP48* mutants enhance transcription of the *POMC* gene through *USP8*-independent pathways.[17] Mutations of *USP8* and *USP48* do not seem to occur simultaneously. The well-known protooncogene *BRAF* may also be mutated in approximately 12% of corticotropinomas.[17] *BRAF* activates a signaling pathway that increases POMC synthesis, and, interestingly, *in vitro* studies show that clinical BRAF inhibitors decrease ACTH production.

Loss-of-function mutations of the *CABLES1* gene were found in a small subset of patients and were preferentially associated with corticotroph macroadenomas.[17]

Mitotic count and the Ki-67 labeling index, along with radiological invasiveness, are the most commonly used markers of pathological aggressiveness in clinical practice.[9] Unfortunately, these criteria do not accurately predict pituitary tumor behavior such as rapid growth, recurrence, and malignant transformation.[7,9,17] Multiomics analysis of corticotroph PitNETs (including analysis of somatic mutations, chromosomal alterations, miRNome, methylome, and transcriptome studies) represent a hope for a better and clinical relevant classification of corticotroph tumors. Subgroups of corticotropinomas based on their transcriptome, miRNA clusters, and changes in histone acetylation that are associated with ACTH secretion, glucocorticoid resistance, and aggressiveness have already been recently identified.[10,11,17]

PATHOPHYSIOLOGY

Excessive ACTH secretion stimulates steroidogenesis in the zona fasciculata and zona reticularis of the adrenal cortex, inducing an immediate release of cortisol, aldosterone, and dehydroepiandrosterone (DHEA), and Δ-4-androstenedione via cyclic adenosine monophosphate (cAMP)-dependent pathways, as well as the expression of key steroidogenesis enzymes, Gs protein, and the ACTH receptor MC2R. ACTH also activates an autocrine network, including insulin-like growth factor 1 (IGF-1), IGF-2, basic fibroblast growth factor, and transforming growth factor β, that is responsible for bilateral adrenocortical hyperplasia and rendering the hyperplastic adrenocortical cells more sensitive to ACTH, thus further increasing the cellular response.

Although aldosterone is secreted in response to ACTH, negative feedback suppresses aldosterone synthesis. A multiplex analysis of 15 steroids showed that ACTH-driven Cushing syndrome exhibits lower aldosterone levels than are seen in cortisol-secreting adrenal adenomas.[23] However, synthesis of mineralocorticoids such as 11-deoxycorticosterone (DOC) and, in many cases, 18-OH-DOC, is increased and can contribute, along with cortisol, to an increase in blood pressure.

CLINICAL FEATURES

Clinical Aspects

Excessive cortisol production has numerous systemic effects, given the widespread expression of its receptors (Table 7.1). Although the full-blown Cushingoïd phenotype is unmistakable, clinical diagnosis may be equivocal in patients with few of the typical characteristics. These cases often correspond to patients with mild hypercortisolism or given the common slow progression of the disease, patients in whom the onset of the disease is recent (Fig. 7.3). Another difficulty is that chronic hypercortisolism is responsible for nonspecific features such as obesity, hypertension, and menstrual irregularity, which are common in the general population. This situation generates two contradictory caveats: performance of unwarranted and costly screening tests for patients unlikely to be affected (see Who Should Be Tested) and, conversely, a mean latency to diagnosis of 38 months.[24]

One useful strategy when considering the diagnosis of Cushing syndrome is to look for evidence of progressive physical changes by examining serial individual photographs taken in the past (Fig. 7.3). Another approach relies on identification of the more specific signs and symptoms of the disorder. Ecchymoses (Fig. 7.4), plethora, myopathy/proximal muscle weakness, and osteopenia/osteoporosis, as well as wide purple striae (Fig. 7.5), which reflect the catabolic and antianabolic effects of cortisol excess on protein metabolism, are associated with the highest positive likelihood ratio (Fig. 7.6).[25,26]

TABLE 7.1 **Prevalence of the Main Symptoms of Cushing Disease at Presentation**

Symptoms	Prevalence (%)
Obesity/weight gain	70–95
Rounded face (moon face)	81–90
Supraclavicular/dorsocervical fat pads	50
Skin	
Hirsutism/alopecia	56-75
Facial plethora	70–90
Violaceous striae	44–50
Easy bruising	35–65
Acne	20–35
Gonads	
Menstrual irregularity	43–80
Male hygonadism	50–75
Neuropsychological	70–85
Emotional lability/depression	54–81
Anxiety disorder/emotional lability	36–66
Sleeping disorders	57–69
Cognitive dysfunction	50–90
Musculoskeletal	
Muscle weakness/atrophy	42–83
Osteopenia or fractures	21–70
Decreased linear growth in children	70–80
Metabolic	
Hypertension	58–85
Glucose intolerance/diabetes mellitus	30–70
Dyslipidemia	38–71

Increased deposition of fat, one of the earliest signs, occurs in almost all patients and is reported as weight gain or difficulty in maintaining weight.[27] Fat distribution is altered with increased amounts in the visceral compartments and subcutaneous sites on the face and neck, resulting in the typical truncal obesity. Increased fat in the face ("moon facies"), as well as in the supraclavicular and the dorsocervical area ("buffalo hump"), is uncommon in healthy people (Fig 7.7). Although these features are common in massive obesity, this redistribution of fat is very evocative when associated with minimal weight gain and when it contrasts with thin arms and legs owing to myopathy (Fig. 7.8). Loss of subcutaneous tissue and thinning of the skin results in skin abnormalities that are rare in the general and middle-aged population. Ecchymoses (Fig. 7.4), often after minimal trauma, and cutaneous atrophy, seen as a fine "cigarette paper" wrinkling over the dorsum of the hand, are also typical (Fig. 7.4). Cutaneous atrophy is physiologically influenced by gender and age, with men and younger individuals having greater skin thickness. Therefore, it is useful to compare the patient's skin with that of a near age- and gender-matched healthy person.

Spontaneous facial plethora (e.g., in the absence of topical drying agents and glucocorticoids), especially over the cheeks, is a characteristic feature that usually decreases rapidly in intensity after cessation of hypercortisolism (Fig 7.7). Facial erythema is also related partly to loss of subcutaneous tissue and is more obvious in pale White individuals. A demarcation line, representing collar,

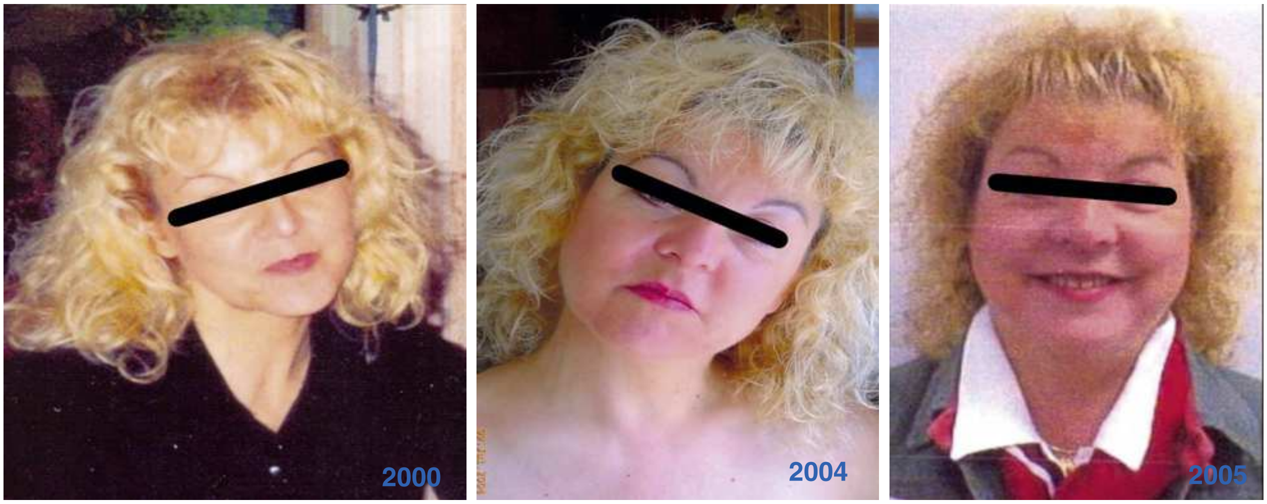


Fig. 7.3 Progression of Cushingoïd features. Note the typical slow progression of facial rounding and plethora over 5 years.

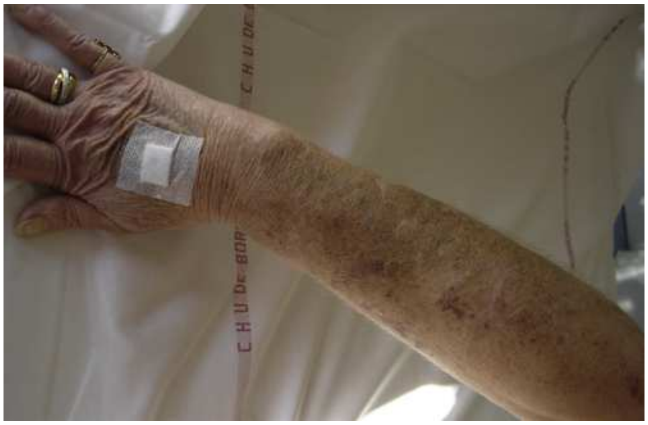


Fig. 7.4 Multiple spontaneous ecchymosis and thinning of the skin in a 63-year-old woman with Cushing disease.

sleeve, or shoulder straps, may differentiate hypercortisolism-related erythema from exogenous causes such as ultraviolet radiation from ultraviolet lamps or sunlight.

Purple striae greater than 1 cm in width are virtually pathognomonic for Cushing syndrome (Fig. 7.5) and appear very different from the silvery, healed striae that are typical postpartum. Although most common over the abdomen, striae also occur in unusual parts of the body, such as over the hips, buttocks, thighs, breasts, and upper arms.

Proximal muscle weakness is a hallmark of Cushing syndrome[28] (Fig. 7.8) and has been reported in 70% of subjects in large cohorts.[2] Histologically, this is reflected in profound atrophy of fibers, or fatty infiltration,[28,29] and body composition analysis has revealed reduction of muscle mass in the arms and legs.[27,29] Weakness can be assessed by enquiring if patients experience difficulty or weakness in climbing stairs, or in getting up from a chair or bed without using their hands. Formal muscle strength testing is useful. Assessment of the strength of the hip flexors is done by asking the patient to get out of a chair or get up after crouching without using his/her arms. Maximal duration of leg extension while seated and handgrip tests are also quantifiable tests of proximal muscle strength.

Osteopenia and osteoporosis are common, owing to the direct effects of cortisol excess on the replication and differentiation of osteoblasts resulting in uncoupling of bone turnover, as well as indirect effects on intestinal and renal calcium homoeostasis, impairment of pituitary hormone secretion, and catabolic effects on muscles.[29,30] A history of fractures, typically of the feet, ribs, or vertebrae after minimal or no trauma, is frequent, especially in men, suggesting that cortisol-induced hypogonadism plays an associated role in the pathogenesis of bone fragility. A number of studies have shown decreased bone mineral density (BMD) in Cushing disease.[29,30] However, BMD underestimates the impact of chronic hypercortisolism that also alters the bone microarchitecture.[31] Consequently, fractures contrasting with normal or minimally altered BMD are common in glucocorticoid-induced osteoporosis.[30,31] Avascular necrosis of bone, a rare complication of endogenous hypercortisolism, is more common in iatrogenic hypercortisolism.

Hirsutism, as well as pustular acne, reflecting increased secretion of androgens, or papular acne, reflecting pure glucocorticoid excess, are common in both women and men.

Most patients experience emotional and cognitive changes including increased fatigue, irritability, depressed mood, decreased libido, insomnia, anxiety, impaired memory and concentration, and changes in appetite and sleep.[32,33] Psychiatric symptoms may be exacerbated in vulnerable individuals.[34] Irritability and sleep disturbances with early falling asleep because of fatigue followed by persistent and early night awakening are often the earliest symptoms. According to current classifications of mental health disorders, between 52% to more than 80% of patients meet the criteria for a major affective disorder.[33] The depressed mood may induce suicide.[34] The time course of psychiatric symptoms is characteristically intermittent, with alternation of depressed phases with periods of excessive arousal, in contrast to the constant dysphoria reported by depressed patients without Cushing syndrome. A minority of patients are manic.[34] Psychiatric symptoms are often refractory to common medical therapies, including antidepressants, but rapidly improve following cessation of hypercortisolism. During intense hypercortisolism, some patients may develop acute steroid psychosis that associates agitation and paranoid symptoms and can prevent adherence to oral therapy and render parenteral administration of steroidogenic inhibitors necessary.[35] Cognitive function is also frequently negatively affected, with impaired memory and impaired concentration being the most frequent symptoms.[33,36] Imaging studies have revealed morphological and functional changes in the whole brain and in cerebral areas that are rich in glucocorticoid receptors, particularly the hippocampus, amygdala, and prefrontal cortex, which are fundamental regions associated with emotional and cognitive functions such as memory.[37]

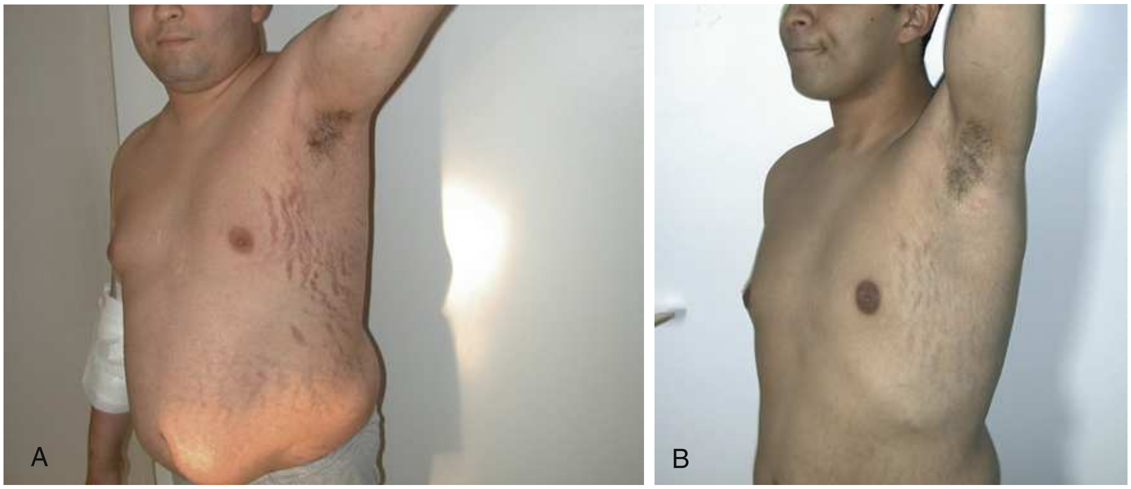


Fig. 7.5 Purple striae before and 3 months after successful pituitary surgery in an 18-year-old man. **A,** Before surgery: typical large violaceous striae. **B,** After surgical remission. Disappearance of the purple appearance of the stretch marks. Also note the weight loss and thinning of the face compared to the preoperative habitus.

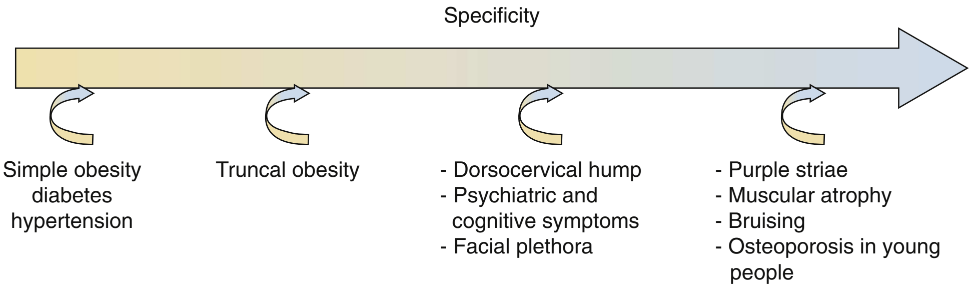

Specificity
Simple obesity diabetes hypertension
Truncal obesity
- Dorsocervical hump
- Psychiatric and cognitive symptoms
- Facial plethora
- Purple striae
- Muscular atrophy
- Bruising
- Osteoporosis in young people

Fig. 7.6 Classification of symptoms of Cushing syndrome in increasing order of specificity.

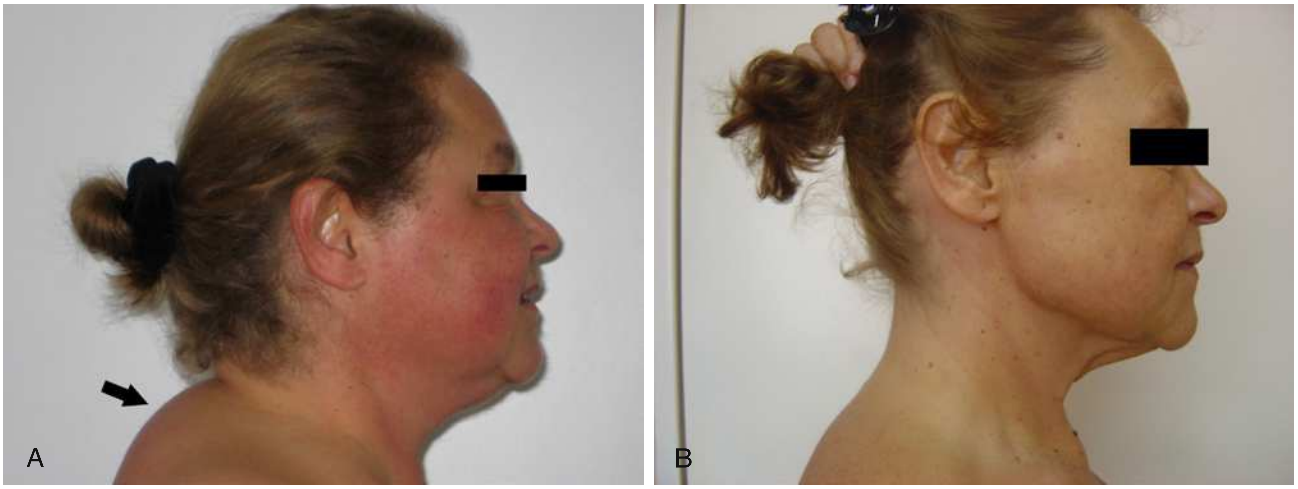


Fig. 7.7 Favorable evolution of buffalo hump, facial rounding, and plethora following successful pituitary surgery in a 60-year-old woman. **A,** Before surgery. **B,** One year after surgical remission. Note the disappearance of the buffalo hump (*arrow*) and the facial plethora.

Hypertension is frequent (Table 7.1),[29,33] and although hypertension is common in the general population, its presence in patients younger than 40 years of age, especially if it is difficult to control, may alert physicians to the syndrome.[2,38] Fungal infections of the skin, such as mucocutaneous candidiasis and pityriasis versicolor, staphylococcus infection of the face after shaving with a razor in men, and increased latency of healing of small everyday skin wounds, are common features. Poor wound healing and wound dehiscence occur less often but are an important consideration in patients with severe Cushing syndrome who are treated surgically.

All of these clinical features are exacerbated in patients with intense hypercortisolism (see Cushing Disease as an Emergency Condition).

Decreased libido associated with erectile dysfunction is common in men. Women usually experience menstrual irregularity, oligomenorrhea, and infertility, reflecting polycystic ovary syndrome when the intensity of hypercortisolism is moderate, or secondary amenorrhea resulting from hypogonadotroph hypogonadism in severe Cushing syndrome.[29]

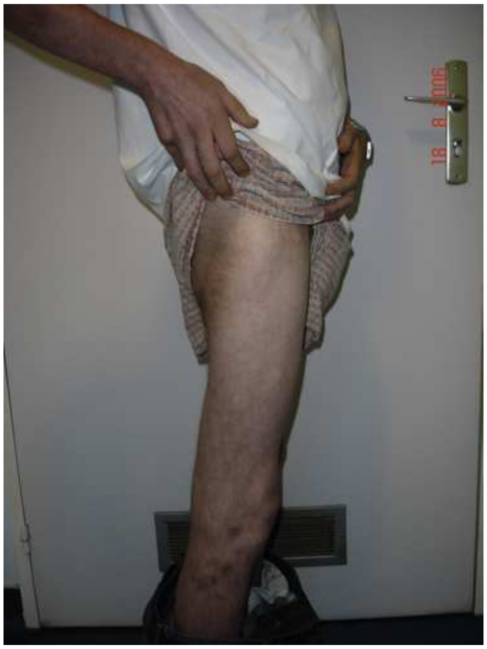


Fig. 7.8 Profound proximal amyotrophia of the thighs of a 23-year-old man with Cushing disease. Note the ecchymosis on the lateral face of the knee and leg.

Nonspecific Biochemical and Radiological Investigations

Hypokalemic metabolic alkalosis is usually observed when hypercortisolism is intense and is therefore more frequent in ectopic ACTH syndrome than in Cushing disease. It is mainly secondary to activation of the mineralocorticoid receptor by cortisol excess at the renal tubule, because of saturation of the enzyme 11β-hydroxysteroid dehydrogenase type 2, which inactivates cortisol to cortisone.[35] Serum albumin inversely correlates with cortisol levels, but this is clinically significant only at very high cortisol levels.[39] Drastic reductions in serum albumin should alert the physician to the possibility of concomitant pathology, such as infection.

Chronic hypercortisolism increases clotting factors, including factor VIII, fibrinogen, and von Willebrand factor, and reduces fibrinolytic activity, resulting in a 4-fold to 10-fold increased risk of thrombotic events.[40-42] Deep vein thrombosis and pulmonary thromboembolism may be more frequent after pituitary surgery for Cushing disease,[41] and retrospective studies suggest a beneficial effect of thromboprophylaxis.[41,42]

Cortisol regulates both lipolytic and adipogenic processes in a complex fashion.[27] Lipid abnormalities include increases in very low-density lipoproteins, low-density lipoproteins, and triglycerides, and usually a decrease in high-density lipoproteins.[33]

Cushing syndrome is characterized by insulin resistance and hyperinsulinemia, with glucose intolerance and diabetes mellitus occurring in 30% to 70% of patients.[27,29,33] In patients with Cushing syndrome (as with other etiologies of secondary diabetes), diabetes mellitus is more accurately diagnosed using an oral glucose tolerance test than fasting blood glucose. The potentiation of multiple cardiovascular risk factors in Cushing syndrome, such as hypertension, insulin resistance and glucose intolerance/diabetes mellitus, dyslipidemia, and hypercoagulability, is responsible for accelerated vascular disease, including cerebrovascular accidents and myocardial infarction.[29,33,43-45] A direct pathogenic effect of cortisol on blood vessels and the myocardium is also suspected.[46] Biomarkers for cardiovascular disorders include increased carotid artery intima-media thickness and atherosclerotic plaques seen on Doppler ultrasonography[47] or computed tomography (CT) coronary angiography.[48] On echocardiography, Cushing syndrome is characterized by left ventricular hypertrophy and diastolic and systolic dysfunction that may be partially explained by increased myocardial fibrosis.[29] Magnetic resonance imaging (MRI) studies have found left ventricle hypertrophy associated with decreased systolic strain and impaired diastolic filling caused by an abnormal relaxation pattern in patients with Cushing syndrome.[49]

Hypercortisolism suppresses the thyroidal, gonadal, and growth hormone (GH) axes. Low triiodothyronine levels attributed to the inhibition of thyroxine deiodination is common. Thyroid-stimulating hormone (TSH) release is also decreased with a loss of its nocturnal surge, resulting in reduced thyroxine and triiodothyronine levels, with reduced TSH levels being correlated with the intensity of hypercortisolism.[50] In both men and women, low levels of luteinizing hormone, follicle-stimulating hormone, and gonadal steroids, consistent with hypogonadotropic hypogonadism, are common in intense hypercortisolism, whereas polycystic ovary syndrome is common in patients with Cushing disease of moderate intensity.[29,33] Hypercortisolemia reduces GH secretion during sleep and blunts the GH response to stimulation tests.[29,51]

Clinical Spectrum of Cushing Disease. Who Should Be Tested for Cushing Syndrome?

KEY POINTS

- Generalized biochemical screening for CD in obese, diabetic and hypertensive patients is not recommended. Clinical suspicion leading to biochemical screening is based on clinical examination for "specific" symptoms of Cushing's syndrome, occurrence of unexplained changes in cognition and mood, and features unusual for the condition of the patient.

The typical patient with Cushing disease is a middle-aged woman who complains of the gradual development of symptoms over several years. However, the clinical presentation of patients with Cushing disease encompasses a wide spectrum and varies with regard to the type, number, and severity of symptoms, which are in turn to some degree related to the intensity and duration of hypercortisolism, patient sex and age (with more severe forms in older patients and male patients),[52] individual sensitivity to glucocorticoids,[53,54] and preexisting sporadic or inherited vulnerabilities such as overweight, hypertension, and diabetes.[55] The clinical presentation of pituitary corticotroph macroadenomas, apart from visual field changes caused by suprasellar expansion, is not unique. In contrast, rare and extremely aggressive corticotroph macroadenomas present at a slightly younger age, and cavernous sinus and dural involvement may result in cranial neuropathies and facial neuralgia. Only a few case reports attest to overt malignancy, with cerebrospinal or extracranial metastasis of ACTH-producing pituitary tumors, but the relative prevalence of corticotropinomas is significant among aggressive and malignant pituitary tumors.[7-9]

As early as 1964, several attempts have been made to develop a diagnostic score that can detect an increased probability of a patient having Cushing syndrome before performing biochemical assays.[25,26] Although none of these attempts achieved any great accuracy, these studies highlight that "catabolic" symptoms such as osteoporosis, muscle atrophy, purple striae, and easy bruising are among the more specific symptoms and are associated with the highest likelihood ratio of Cushing syndrome (Fig 7.6). However, these clinical symptoms are absent in a number of patients with Cushing syndrome. In addition, diagnosing Cushing disease at an early stage is an important challenge, given the deleterious impact of prolonged exposure to excess cortisol. This raises the issue of screening for the disease in poorly symptomatic patients and/or patients lacking specific symptoms, such as diabetic and obese patients. This issue is important given the increased prevalence of these diseases and the fact that Cushing syndrome is a treatable cause of secondary obesity and diabetes. Systematic screening studies using biochemical investigations in unselected patients with type 2 diabetes mellitus referred to specialized centers found a prevalence of Cushing syndrome ranging from 0% to 3.3%.[56] In the largest multicenter study performed to date, involving more than 800 outpatients with diabetes in Italy, the prevalence was 0.6%.[57] Contrary to patients with a clinical suspicion of Cushing syndrome, adrenal adenomas with autonomous cortisol secretion responsible for mild cortisol excess, including normal urinary free cortisol (UFC), were much more prevalent than were cases of Cushing disease.[56] Furthermore, the impact of treatment of mild Cushing syndrome on glycemic control was not mentioned in the majority of cases. The prevalence of Cushing syndrome in obesity has been assessed in a metaanalysis involving 5819 patients.[58] Among obese patients with nonspecific symptoms, including type 2 diabetes, the prevalence of Cushing syndrome ranged between 0% and 1.1%.[58] Studies conducted in patients with hypertension suffer from major methodological drawbacks that prevent accurate evaluation. Rare screening studies in osteoporotic patients have found a prevalence ranging from 1.3% to 3.8% in unselected patients, corresponding in almost all cases to adrenal adenomas associated with mild Cushing syndrome.[56]

A screening strategy is justified is it is supported by epidemiological concerns and sufficient evidence that its efficacy benefit will outweigh its drawbacks.[56] The biological spectrum of “occult” Cushing syndrome and its relatively low prevalence may negatively affect the performance of biochemical diagnostic tests, favoring the occurrence of false positives.[56] Furthermore, whether curing minimal hypercortisolism favorably affects clinical outcomes remains unknown. Last, the feasibility of large-scale screening, in terms of workload and costs, is highly questionable, given the high prevalence of patients with conditions to screen. As recommended by the Endocrine Society, widespread screening appears not to be indicated, and, before considering biochemical tests, it is mandatory to have clinical suspicion based on factors that increase the pretest probability of Cushing syndrome.[59] This includes careful clinical examination to look for specific symptoms of Cushing syndrome, to enquire if symptoms have accumulated and worsened over time, to ascertain the occurrence of unexplained changes in cognition, short-term memory, insomnia and mood, and to look for features that are unusual for the patient’s condition, such as osteoporosis and hypertension in a young person, thin skin in men, and so on.[59]

BIOCHEMICAL DIAGNOSIS OF CUSHING SYNDROME

KEY POINTS

- The choice between first-line biochemical investigations for the diagnosis of Cushing’s syndrome should be based primarily on the patient’s conditions, e.g. individual characteristics that can lead to false positive or negative results of tests.

General Considerations

First-line biochemical tests used for the diagnosis of Cushing syndrome are aimed at establishing the cardinal biochemical features of Cushing syndrome, these being excessive production of cortisol, loss of the normal circadian rhythm of cortisol secretion, and impaired feedback of the hypothalamic-pituitary-adrenal (HPA) axis. As mentioned previously, it is important to test patients with a significant pretest probability of Cushing syndrome (see Clinical Aspects) to ensure the best performance of biochemical tools[56,59,60] and to use a first-line test with high sensitivity to avoid missing milder cases.[59] All tests have limitations related to the individual patient situation (Table 7.2) that should guide the choice of the test. All tests have imperfect sensitivity and specificity.[60,61] Imperfect sensitivity implies the need for repeated evaluation if the patient presents with convincing clinical features. Imperfect specificity implies that any abnormal test should be confirmed by a second first-line test to confirm the diagnosis (Fig. 7.9).[59] This strategy increases confidence in the test results when consistently normal or abnormal results are obtained. Second-line tests should be performed in expert centers in ambiguous cases.

As with all hormone assays, physicians must be aware that results for a single sample measured in different assays may differ.[62] Caution should be exerted when considering results near the cutoff value on a single measurement, as subnormal or slightly increased results can often be explained by assay variability. It also important to consider that no “universal” threshold for any biochemical measurement exists, and that normal ranges vary substantially depending on the method used, implying the need to interpret test results in the context of an appropriate and accurately defined normal range. Additionally, the performance of tests in real-life practice is often worse than that quoted in the literature.[63]. This may be because of the greater expertise of laboratories in tertiary expert centers but also attributed to the retrospective design of studies and the use of healthy subjects in the control group, which results in decreases in both the specificity and the positive likelihood ratio.[60,61] Lastly, it is important to keep in mind that any intercurrent illness will appropriately stimulate HPA axis function and result in increased cortisol secretion, emphasizing the importance of testing patients in stable clinical conditions.

TABLE 7.2 **Main Causes of False Results in Biochemical Tests Used for the Diagnosis of Cushing Syndrome**

Screening test	False positive	False negative
Urine free cortisol	Polyuria Medications Overcollection Crossreactivity/immunoassays Nonneoplastic hypercortisolism/stress	Undercollection Moderate to severe renal insufficiency
Low-dose dexamethasone suppression test	Increased CBG (pregnancy, chronic active hepatitis, oral estrogens) Multiple medications Rapid metabolism of dexamethasone Impaired absorption of dexamethasone Nonneoplastic hypercortisolism/stress	Decreased CBG (liver cirrhosis, nephrotic syndrome) Slow metabolism of dexamethasone
Late-night salivary cortisol	Age, hypertension, diabetes/dependent on the assay Shift workers Tooth bleeding Nonneoplastic hypercortisolism/stress	Improper collection

CBG, Corticosteroid-binding globulin.
(Data from references 3,29,33,34,235.)

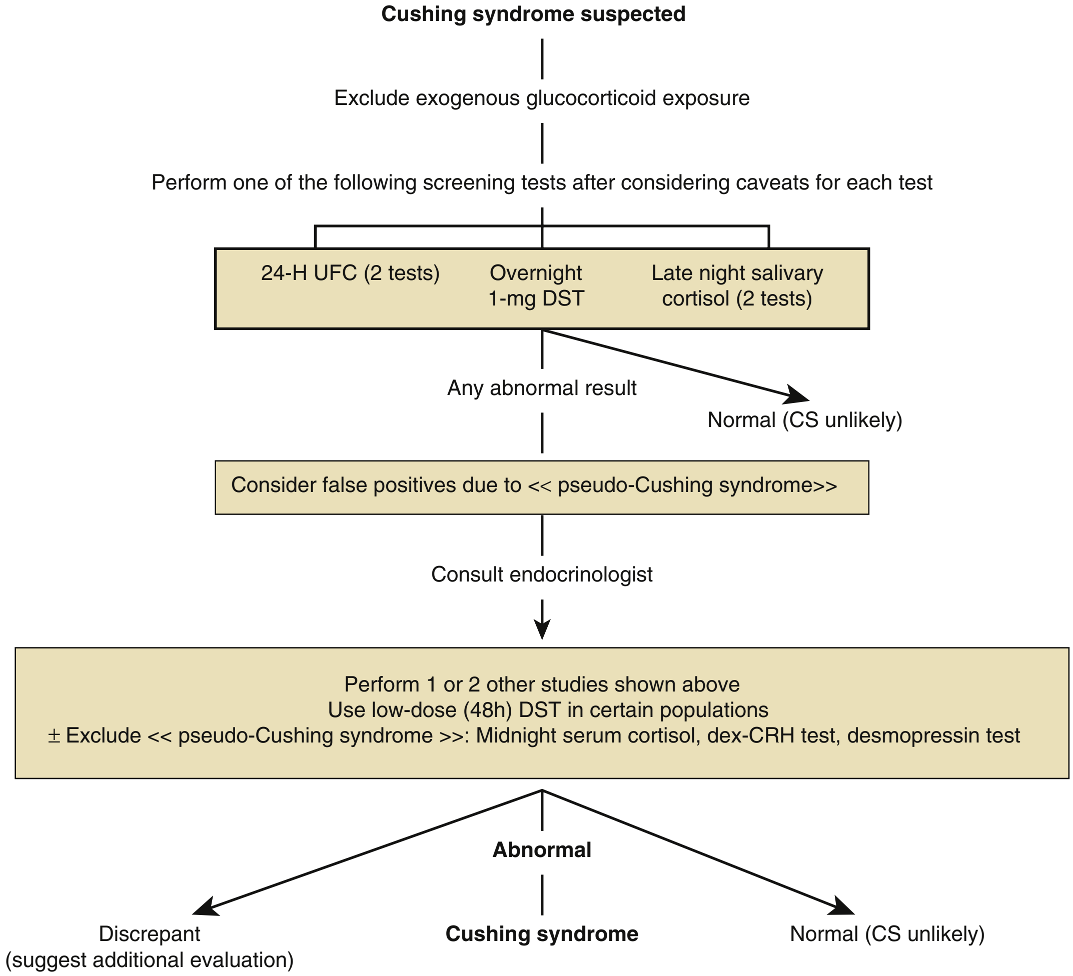

Cushing syndrome suspected
Exclude exogenous glucocorticoid exposure
Perform one of the following screening tests after considering caveats for each test
24-H UFC (2 tests)
Overnight 1-mg DST
Late night salivary cortisol (2 tests)
Any abnormal result
Normal (CS unlikely)
Consider false positives due to << pseudo-Cushing syndrome>>
Consult endocrinologist
Perform 1 or 2 other studies shown above
Use low-dose (48h) DST in certain populations
± Exclude << pseudo-Cushing syndrome >>: Midnight serum cortisol, dex-CRH test, desmopressin test
Abnormal
Discrepant (suggest additional evaluation)
Cushing syndrome
Normal (CS unlikely)

Fig. 7.9 Algorithm for testing patients suspected of having Cushing syndrome, modified from Endocrine Society recommendations. *CRH,* Corticotropin-releasing hormone; *CS*, Cushing syndrome, *DST*, dexamethasone suppression test; *UFC,* urinary free cortisol. (Reproduced with permission from Nieman LK, Biller BMK, Findling JW, et al. The diagnosis of Cushing syndrome: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2008;93:1526-1540.)

First-Line Biochemical Investigations

Urinary Free Cortisol. Under normal conditions, 5% to 10% of plasma cortisol is free or unbound and physiologically active. The kidneys filter unbound cortisol, with most being reabsorbed, while the remainder is excreted unmodified. A 24-hour UFC collection produces an integrated measure of circulating serum cortisol during the day. The major drawback of the test is the potential for overcollection or undercollection of the 24-hour specimen. Consequently, written instructions must be given to the patient, and creatinine excretion in the collected urine should be measured to assess completeness (~1 g/24 hours in a 70-kg patient with normal kidney function). Creatinine measurement cannot be used to correct for incomplete collection, as rates of cortisol and creatinine excretion are not parallel over the 24-hour period. Various groups have attempted to overcome collection issues by proposing shorter collection periods focusing on periods when the loss of circadian rhythm differs most from normal controls, such as at night (from 8:00 PM to 8:00 AM), but this approach has not been widely accepted.[64,65] High-performance liquid chromatography and tandem mass spectrometry (LC-MS/MS) is now commonly used to measure UFC, which overcomes the problem of crossreactivity of other endogenous or exogenous glucocorticoids that occurs in some, but not all, immunoassays. Consequently, UFC values are invariably lower when measured by LC-MS/MS as compared with immunoassays because of the lack of crossreactivity with interfering glucocorticoid metabolites.[66] However, aside from the rare situation of spurious results owing to exogenous steroid intake, direct head-to-head comparison studies show that the diagnostic performance of mass spectrometry and selected immunoassays is comparable,[66] assuming that correctly established assay-specific normal ranges are used.

A number of conditions influence urinary cortisol excretion (Table 7.2). Diuresis stimulates cortisol excretion, and mildly increased UFC is observed in patients with polydipsia.[67] Of more clinical relevance, UFC is significantly decreased with moderate and severe renal insufficiency and so should not be used in this situation.[68] Although values greater than 4-fold of normal are pathognomonic of Cushing syndrome, a recent metaanalysis and several studies show that the sensitivity of UFC for the diagnosis of Cushing syndrome is 92% to 96% and is lower than that of late-night salivary cortisol (LNSC) or the 1-mg dexamethasone suppression test (DST).[61] False negatives are more likely to occur in patients with Cushing syndrome of mild intensity.[69-72] Furthermore, it should be stressed that UFC is highly variable from one day to another and, in a series of 162 patients with overt Cushing disease, 5% had at least one out of four UFC collections within the normal range, emphasizing the need for multiple collections.[73] Consequently, it may be more pertinent to use UFC as a confirmatory test rather than a screening test.[61]

TABLE 7.3 Examples of Drugs That May Interfere With Dexamethasone Suppression Tests

Drugs that accelerate dexamethasone metabolism by induction of CYP3A4	Drugs that impair dexamethasone metabolism by inhibition of CYP3A4	Drugs that increase CBG and may falsely elevate total serum cortisol results
• Carbamazepine	• Aprepitant, fosaprepitant	• Estrogens
• Ethosuximide	• Cimetidine	• Mitotane
• Phenobarbital	• Diltiazem, verapamil, amiodarone	
• Phenytoin	• Fluoxetine	
• Pioglitazone	• Ketoconazole, Itraconazole, fluconazole, etc.	
• Pyrimidone	• Macrolides (erythromycin, clarithromycin, etc.)	
• Rifampin	• Ritonavir	
• Rifapentine		
False positive possible	False negative possible	False positive possible

CBG, Corticosteroid-binding globulin; *CYP*, cytochrome P450.

Late-Night Salivary Cortisol. The rationale for the measurement of LNSC is 2-fold. Firstly, salivary cortisol is a reflection of plasma free cortisol concentration, almost circumventing the changes in total cortisol that are caused by corticosteroid-binding globulin (CBG) alterations, and more specifically the increase in cortisol in women taking estrogens.[74] Secondly, saliva can be sampled at the critical period during which plasma cortisol reaches its nadir.[72] From a practical perspective, LNSC offers the advantage of nontraumatic sampling that can be performed conveniently in both in- and outpatient settings.[75] LNSC is also convenient in children and obese patients, where venipuncture may be difficult. Salivary cortisol is stable for several days at room temperature, for weeks at 4° C, and can be sent to the testing laboratory by mail. Self-sampling can be easily repeated when cyclical Cushing syndrome is suspected.[65] Studies suggest that there is no difference in the diagnostic performance if sampling is performed earlier than 11:00 PM, whereas another study suggested that patients could sample saliva at their normal bedtime.[74,76] The main drawbacks are the possible contamination of samples by blood, meaning teeth should not be brushed just before sampling, night-shift work, which can induce different patterns of the circadian rhythm of cortisol in a subset of individuals (termed “shifters”),[77] and patients having recently changed their “time-zone.” LNSC measurement is performed using immunoassay, providing wide variations in cutoffs depending on the assay used, or by LC-MS/MS. The latter provides the possibility of measuring both cortisol and cortisone. Saliva contains higher levels of cortisone than cortisol owing to conversion by the 11β-hydroxysteroid dehydrogenase (11β-HSD) type 2 enzyme in the salivary glands, which “amplifies” the cortisol signal. The relative performance of immunoassays versus LC-MS/MS and whether specific measurement of salivary cortisone is an important additional diagnostic parameter have been a matter of debate.[63,74] Normal values increase with age and the presence of diabetes and hypertension,[74] therefore requiring validation of normal ranges based on the population being studied to ensure the best performance, especially in the context of mild hypercortisolism.

Two saliva samples are recommended to ensure the best diagnostic performance,[59,63] with sensitivity and specificity ranging from 94% to 97% and 91% to 95%, respectively, in a recent metaanalysis.[61] Several publications suggest that LNSC is more suitable for diagnosing mild forms of Cushing disease with a normal UFC.[69-72,78]

Overnight Dexamethasone Test. The rationale of the test is that, in normal individuals, administration of the synthetic glucocorticoid dexamethasone results in suppression of the HPA axis, whereas patients with Cushing syndrome are resistant, at least partially, to this negative feedback. The overnight DST correspond to 9:00 AM serum cortisol measurement after a single dose of 1 mg of dexamethasone administered at 11:00 PM. Its low cost and simplicity, as well as its performance, explain its widespread use.[59,61] The dose of dexamethasone used and that performance of the test are similar in obese and nonobese subjects.[79] The measurement of salivary rather than serum cortisol has been proposed to circumvent the issue of variations in CBG and increased total cortisol in women taking estrogens.[59,74,80] In the context of screening, a relatively “low” threshold of 50 nmol/L has been adopted to ensure the best sensitivity,[81,82] which is as high as 95%.[59-61] The major drawback of the test is a roughly 10% rate of false positives,[57,82] resulting in a specificity of around 90%.[61] The interpretation of these values, in practical terms, is that impaired suppression does not necessarily mean that a patient has Cushing syndrome. Conversely, adequate suppression is very strong evidence against the diagnosis. Factors such as variable absorption and increased or decreased dexamethasone metabolism by CYP3A4 owing to other compounds[59] (Table 7.2 and 7.3) may be responsible for a number of false results. Therefore, a history of symptoms of malabsorption or bariatric bypass surgery should be noted, and a careful drug history (see http://medicine.iupui.edu/flockhart/table.htm) should be taken before the test is used in a patient.[83] Measurement of plasma dexamethasone has been proposed in patients with conditions that may induce or inhibit CYP3A4 activity. However, this useful adjunct is not widely used because of the added cost and some uncertainty of the threshold to employ.[80] One solution to overcome malabsorption is to use one of the published intravenous DSTs, recognizing that criteria for response in this case have not been standardized.[59] Pregnancy and other causes of increased or decreased CBG (such as exogenous estrogens, chronic hepatitis, and nephrotic syndrome) should also be excluded, because these are likely to result in false-positive and false-negative tests.[62,84] Combined oral contraceptive pill or hormone replacement therapy containing estradiol should be stopped for 4 to 6 weeks before performing tests.[84] If stopping exogenous estrogens is not possible, or if the patient is pregnant, UFC or salivary cortisol should be used, as these tests reflect the free cortisol level.

Second-Line Biochemical Investigations

Low-Dose Dexamethasone Suppression Test. Dexamethasone is given in doses of 0.5 mg for 48 hours, at 6-hourly intervals, beginning at 9:00 AM on day 1, and serum cortisol is obtained at 9:00 AM on day 3. In a recent metaanalysis, both the 2-day and the 1-mg overnight DST analyzed with a postdexamethasone serum cortisol cutoff of 1.8 μg/dL (50 nmol/L) appeared to have approximately comparable performance.[61] No head–to-head comparison of the two tests has been conducted, but a few studies suggest that the low-dose DST (LDDST) is less susceptible to interference owing to dexamethasone malabsorption and stress activation of the HPA axis[57,82,85,86] than the overnight

1-mg test. Given its cumbersome protocol, the LDDST is usually considered as a second- or third-line test for patients with equivocal results. Measurement of salivary cortisol can be a substitute for serum cortisol in women taking the combined oral contraceptive pill.

Midnight Serum Cortisol. Similarly to LNSC, the purpose of midnight serum cortisol (MSC) measurement is to demonstrate elevated concentrations in the period when physiological concentrations reach their nadir. MSC concentration has a roughly similar diagnostic performance to LNSC.[61,71] However, there has been some controversy with regard to the appropriate cortisol threshold to use. A concentration of 50 nmol/L, which was determined in unstressed healthy medical students, effectively excludes Cushing syndrome but results in a false-positive rate of approximately 25%.[59] Therefore, the suggested cutoff criteria are higher, at 7.5 to 8.3 μg/dL (207–229 nmol/L).[82,87,88] Of note, MSC increases in women taking estrogens. The main caveat of this approach is the need to sample blood at midnight, which cannot usually be done in an outpatient setting.

Other Biochemical Investigations

The desmopressin test and dexamethasone-CRH test may be used to differentiate Cushing disease from functional activation of the HPA axis ("pseudo-Cushing") in patients with mild hypercortisolemia (see Differential Diagnosis of Adrenocorticotropic Hormone–Dependent Cushing Syndrome).

Cortisol is incorporated into the hair cells by diffusion from blood capillaries, and as hair has a growth rate of approximately 1 cm/month, it is a suitable matrix for the estimation of mid- to long-term tissular exposure to circulating cortisol. Similar to saliva, 11β-HSD type 2 enzyme in sweat glands of the skin generates cortisone from cortisol in hair. Several studies have reported a similar diagnostic accuracy of cortisol and cortisol plus cortisone measurement in a single hair sample to that of UFC in patients with overt Cushing syndrome.[89,90] Owing to its ability to reflect long-term glucocorticoid exposure, hair cortisol measurement may be particularly useful for diagnosing intermittent/cyclical Cushing syndrome.[89,90] Although hair sampling has a number of practical advantages (noninvasive procedure not requiring specific timing, ability to store hair samples at room temperature and send them by mail to the testing laboratory), it also suffers from several limitations, such as the cumbersome and time-consuming preparation and extraction of hair samples. It may therefore be best reserved for difficult cases.

ESTABLISHING ADRENOCORTICOTROPIC HORMONE–DEPENDENCY OF CUSHING SYNDROME

The strategy for the differential diagnosis of Cushing syndrome begins with measurement of plasma ACTH to distinguish between ACTH-dependent and ACTH-independent causes (Fig. 7.10). In Cushing disease, tumoral corticotrophs show reduced sensitivity to the negative feedback normally exerted by cortisol. Hence, ACTH concentrations are inappropriate, for example, neither low nor undetectable as they are in Cushing syndrome owing to a secreting adrenal tumor. Preanalytical steps (sampling into ethylenediaminetetraacetic acid tubes kept in an ice water bath, rapid centrifugation, aliquoting, and, if necessary, freezing within 1 hour) are critical to avoid ACTH degradation by peptidases and falsely decreased concentrations. Using sensitive two-site immunoradiometric assays, ACTH concentrations less than 10 ng/L in a context of hypercortisolism confirm the ACTH independence of the syndrome, whereas ACTH concentrations greater than 20 ng/L diagnose ACTH dependence. Repeated measurements of cortisol and ACTH in blood samples drawn simultaneously are necessary, because 3% to 4% of patients with Cushing disease have been shown to occasionally have suppressed ACTH levels.[91] When the basal ACTH level is indeterminate (10–20 ng/L [2–4 pmol/L]), a situation that was described in 9% of patients with Cushing disease,[91] a CRH test should be performed. Patients with primary adrenal disease rarely show maximal ACTH values greater than 20 to 30 ng/L (4–6 pmol/L), although patients with Cushing disease usually exceed this value.

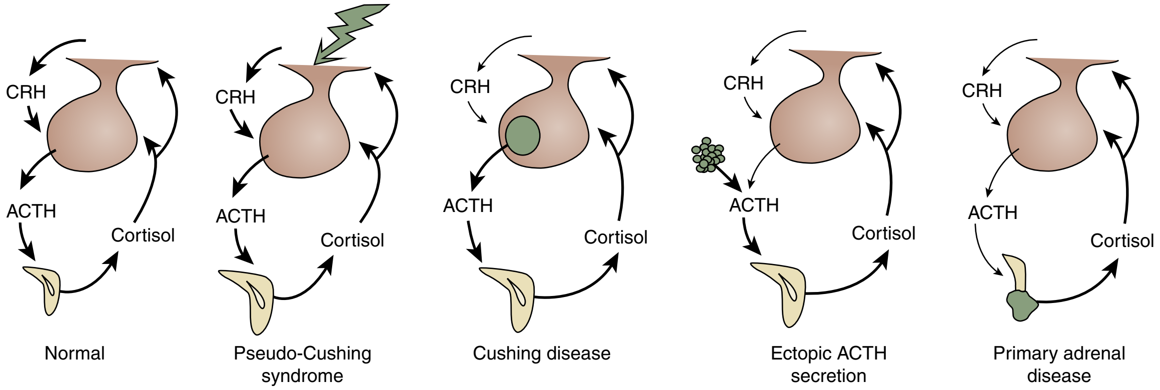


Fig. 7.10 Pathophysiology of the hypothalamic-pituitary-adrenal axis in hypercortisolemic states. Corticotropin-releasing hormone (CRH) secretion from the hypothalamus normally stimulates adrenocorticotropic hormone (ACTH) secretion from the pituitary gland. This in turn results in increased cortisol production from the adrenal glands. The system is modulated by negative feedback inhibition of both CRH and ACTH secretion by cortisol. In pseudo-Cushing syndrome, the CRH neuron is activated by central input (*large shaded arrow*), resulting in increased CRH output that results in hypercortisolism. Increased cortisol production restrains corticotroph activation but does not completely reverse the activation of the CRH neuron, so that mild to moderate hypercortisolism may persist. In Cushing disease, a corticotroph adenoma secretes ACTH in excess and is inhibited only partially by rising cortisol levels. In this setting and that of ectopic ACTH secretion and primary adrenal disease, the CRH neuron is suppressed by hypercortisolism. In the case of ectopic ACTH secretion, excessive secretion of ACTH from a nonpituitary tumor is typically not inhibited by glucocorticoid feedback. In this setting and that of autonomous production of cortisol by the adrenal gland, ACTH secretion by normal corticotrophs is suppressed by hypercortisolism.

TABLE 7.4 Conditions Associated With Chronic Nonneoplastic Functional Hypercortisolism

Conditions associated with chronic nonneoplastic functional hypercortisolism
Intense chronic physical exercise
Anorexia nervosa
Pregnancy
Sleep apnea
Anorexia nervosa
Glucocorticoid resistance syndrome
Uncontrolled diabetes
Alcoholism, especially withdrawal
Psychiatric disorders/severe depression

(Modified from and reproduced with permission from Nieman LK, Biller BMK, Findling JW, et al. The diagnosis of Cushing syndrome: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2008;93:1526-1540.)

In all cases, caution is important in interpreting plasma ACTH results because of marked variability in the results between assays and possible suboptimal assay performance in routine conditions.[92,93] Indeed, a significant number of patients with ACTH-independent Cushing syndrome may exhibit nonsuppressed plasma ACTH concentrations.[93] Use of the CRH test and measurement of plasma ACTH with an additional assay may be important in discordant cases.[91,92] In the same vein, caution is also important in the analysis of abdominal CT scans if these are performed in ambiguous cases, given the 20% prevalence of adrenal macronodules, which may be unilateral and mimic adrenal disease in longstanding Cushing disease.[94]

DIFFERENTIAL DIAGNOSIS OF ADRENOCORTICOTROPIC HORMONE–DEPENDENT CUSHING SYNDROME

The differential diagnosis of ACTH-dependent Cushing syndrome involves two stages: in the presence of moderate hypercortisolism with no specific Cushingoïd symptoms, to rule out causes of functional hypercortisolism; in the presence of overt hypercortisolism with florid Cushing syndrome, to rule out an ectopic ACTH syndrome

KEY POINTS

- Once the diagnosis has been made, the first step in the etiological workup to identify the cause of Cushing's syndrome is the concomitant measurement of serum cortisol and plasma ACTH.
- Functional (e.g. non-neoplastic) causes of Cushing's syndrome such as "pseudo-Cushing's syndrome" and, more rarely, inactivating mutations in the NR3C gene encoding the glucocorticoid receptor, should be excluded in the presence of mild ACTH-dependent hypercortisolism with few or no specific Cushingoïd symptoms.
- CD accounts for approximately 90% of ACTH-dependent cases. This likelihood is increased in an individual with progression of symptoms over a few months, hypokalemia, and large increases in urinary free cortisol, serum cortisol, and plasma ACTH.
- Pituitary MRI is a cornerstone of investigations in ACTH-dependent Cushing's syndrome given the frequency of CD. However, it may be negative in at least a third of patients with CD.
- Although it is not free from false positives and false negatives, BIPSS, performed by an experienced team, is the reference tool for the diagnosis of CD when pituitary imaging is negative. The timing of the BIPSS during the etiological assessment is variable between expert teams.

Hypercortisolemic States Without Cushing Syndrome

Nonneoplastic Hypercortisolism ("Pseudo-Cushing Syndrome"). Activation of the HPA axis is a common adaptive response to situations where homeostasis is challenged. A number of clinical situations and pathologies may cause chronic stimulation of the HPA axis through neural pathways and result in a state of (usually mild) ACTH-dependent hypercortisolism that disappears after resolution of the underlying condition (Fig. 7.10 and Table 7.4). The case of "nonneoplastic hypercortisolism" is not a diagnostic issue when it occurs during situations that are clinically very distinct from Cushing syndrome, such as undernutrition (including anorexia nervosa) or intense athletic activity. Conversely, it can represent a true diagnostic challenge, also termed "pseudo-Cushing syndrome," where patients with nonneoplastic hypercortisolism display nonspecific clinical features of Cushing syndrome, such as visceral obesity, hypertension, hyperglycemia, menstrual disorders, and osteoporosis. This may be the case in patients with chronic alcoholism, psychiatric conditions, polycystic ovary syndrome, and diabetes mellitus with poor glycemic control.[95,96]

As for the etiological diagnosis of Cushing disease, differential diagnosis should be conducted using a progressive probabilistic approach. This should start with a thorough history and physical examination. Patients with depressive illness responsible for pseudo-Cushing syndrome usually have major forms of permanent depression, especially psychotic depression,[95] that do not resemble the most common pleiomorphic and fluctuating psychiatric symptoms seen in Cushing syndrome.[34] Collaboration with expert psychiatrists may be useful for defining the psychiatric picture. Patients with pseudo-Cushing syndrome rarely have specific clinical manifestations. Although their presence strongly argues for true Cushing syndrome, their absence does not exclude mild forms of Cushing disease.[85] In addition, facial fullness with plethora, purple striae, and proximal myopathy may be present in alcoholic patients. The intensity of biochemical abnormalities should also be considered. The 24-hour UFC value is often moderately increased, and elevations of UFC three to four times the upper limit of normal are highly suggestive of Cushing syndrome.[95,96] Chronic stressors are often responsible for false positives of the overnight 1-mg DST associated with the 50-nmol/L cortisol threshold. Several studies suggest that the LDDST may be more specific in identifying pseudo-Cushing syndrome among those not suppressed according to the 1-mg test.[82,85,86] LNSC and MSC determinations are probably the best discriminating biochemical tools because pseudo-Cushing syndrome is characterized by preservation of the circadian rhythm of cortisol secretion, albeit at higher absolute levels.[85,87,95] Increased LNSC has been found in some patients with poorly-controlled diabetes mellitus and hypertension, requiring revised thresholds.[74] It should also be remembered that acute alcohol withdrawal occurring during hospitalization is a major stressor in patients suffering from chronic alcoholism. Second-line investigations in ambiguous cases include the desmopressin test and the dexamethasone-CRH test. The desmopressin test takes advantage of the overexpression of the aberrant expression of its cognate receptor, the vasopressin V2 renal receptor, in approximately 85% of corticotroph adenomas contrary to noncancerous corticotrophs.[97,98] An increase in plasma ACTH of less than 6 pmol/L (>27 ng/L) during the test has been associated with 75% and 87% sensitivities and 90% specificity.[95,96] There is therefore some overlap between the two groups, but the higher the ACTH response the more likely it is to be Cushing disease. The rationale of the dexamethasone-CRH test relies on the different degree of sensitivity to glucocorticoid feedback between patients with pseudo-Cushing syndrome and patients with Cushing disease. In the latter, low doses of dexamethasone are insufficient to inhibit the cortisol increase in response to CRH, whereas patients with pseudo-Cushing syndrome are expected to be unresponsive to CRH while being suppressed by dexamethasone.[95,96] Typically, 2 mg of dexamethasone is given orally as eight doses over 2 days before performing the CRH test. Various thresholds for cortisol responses have been proposed, but the test is far from being completely specific.[95,96] It does not perform better than the dexamethasone part of the protocol, according to some studies,[85,86] and, in comparative studies, has lower accuracy than both the desmopressin test[82] and MSC.[85] Of note, psychiatrists have used the LDDST and the dexamethasone-CRH test to characterize and classify neuropsychiatric disorders,[95] casting doubt on the diagnostic value of the tests in depression-induced activation of the HPA axis.

In ambiguous cases, biochemical reevaluation is needed to make a diagnosis, especially after treatment of the putative causal condition (glycemic control in diabetic patients, treatment of depression, recovery from alcoholism). Although an obvious pituitary adenoma on MRI may be diagnostic, we caution against the use of medical imaging because of the high prevalence of small artefactual pituitary images.[99] Similarly, bilateral inferior petrosal sinus sampling (BIPSS) is discouraged because it cannot distinguish neoplastic from nonneoplastic hypercortisolism, both conditions being pituitary ACTH–driven and associated with high central-to-peripheral ACTH gradients.[95]

NR3C1 Mutations and General Resistance to Glucocorticoids.

Glucocorticoid resistance should be suspected in patients with persistent mild hypercortisolism contrasting with an absence of clinical or metabolic symptoms. Abnormalities in the *NR3C1* gene that encodes the glucocorticoid receptor are associated with generalized glucocorticoid resistance. This rare condition is characterized by decreased tissue sensitivity to glucocorticoids, which, at the pituitary and hypothalamic levels, induce compensatory ACTH secretion and, consequently, biochemical ACTH-dependent hypercortisolism discordant with the absence of Cushing syndrome. The chronic increase in ACTH secretion is responsible for the development of adrenal nodules and hypersecretion of mineralocorticoid precursors and DHEA, which may induce hypertension, hypokalemia, and hirsutism. The spectrum of the disease is, however, very wide, and patients may be minimally symptomatic. Heterozygous *NR3C1* mutations have been identified in more than 5% of patients with incidentally-discovered moderate-sized adrenal nodules and an impaired 1-mg DST.[100] Genetic analysis for *NR3C1* mutations should therefore be considered in patients with mild ACTH-dependent hypercortisolism without specific clinical features of Cushing syndrome, especially in the presence of a moderate decrease in serum potassium associated with low plasma aldosterone and renin concentrations.

Differentiating Cushing disease From Ectopic Adrenocorticotropic Hormone Syndrome and Ectopic Corticotropin-Releasing Hormone Secretion

Epidemiology and Etiologies of Ectopic Adrenocorticotropic Hormone Syndrome.

The prevalence of ectopic adrenocorticotropic hormone syndrome (EAS) established from tertiary referral centers is 9% to 18% of cases of ACTH-dependent Cushing syndrome.[35,101] The vast majority of EAS cases occurs in adulthood, with pediatric and adolescent cases being the exception.[102] EAS results from unregulated ACTH secretion by tumors of various locations, and various degrees of histological differentiation and aggressiveness.[35] Some ectopic ACTH-secreting tumors, such as well-differentiated bronchial neuroendocrine tumors (previously called "carcinoids"), carry an excellent tumor prognosis.[103] These tumors usually exhibit indolent growth and may even be radiologically undetectable for months.[35,101,103] At the opposite end of the spectrum are aggressive neuroendocrine carcinomas associated with locoregional or distant spread.[35] Obviously, these tumors represent different diagnostic challenges. Their respective frequencies vary among series, possibly because of recruitment bias (e.g., cancer vs. endocrine departments). However, the proportion of differentiated neuroendocrine tumors recruited by endocrinologists has increased dramatically over the years.[35,101] Ectopic CRH-secreting tumors are very rare, corresponding to the same etiologies as ectopic ACTH syndrome, and most cases involve concomitant secretion of CRH and ACTH.[104] Rarer, challenging cases of exclusive CRH secretion have also been described.[105]

Clinical Presentation and Baseline Biochemical Investigations.

In patients with evident neoplastic disease treated in various hospital departments, there is no differential diagnosis issue with Cushing disease, and the main diagnostic problem is to not overlook EAS because classical features of Cushing syndrome (such as weight gain, buffalo hump, etc.) may be missing owing to the rapid development of intense hypercortisolism. In an endocrinological clinical setting, and although there is significant overlap in the clinical and biochemical presentation between patients with Cushing disease and EAS,[103] several characteristics can direct the diagnosis towards an EAS. Patients with EAS are significantly older and more likely to be male than patients with Cushing disease. The progression of symptoms is usually accelerated, and hypercortisolism is intense in EAS. Consequently, within months patients develop overt catabolic symptoms (profound muscle weakness, osteoporotic fractures), mild or absent weight gain,[35,101] hypertension, and profound hypokalemia owing to the saturation of the renal 11β-HSD type 2 by excess substrate. Large increases in UFC (>5-fold the normal range), serum cortisol (>1000 nmol/L), and ACTH (>150 pg/mL) concentrations are common.[35,101] Indeed, the pretest probability of Cushing disease in a middle-aged woman with ACTH-dependent Cushing syndrome of moderate biochemical intensity, without hypokalemia or major elevation of plasma ACTH, and having settled gradually over more than a year, is generally more than 90%.[106] This pretest probability may exceed the sensitivity and specificity of some dynamic tests.

Biochemical Investigations

General considerations. The rationale of dynamic tests to differentiate Cushing disease from EAS is to indirectly assess, *in vivo,* the differences in molecular equipment between adenomatous corticotrophs (which typically exhibit the "corticotroph phenotype") and ectopic ACTH-secreting tumor cells. The corticotroph phenotype is characterized by the presence of glucocorticoid receptors that retain the ability to inhibit ACTH secretion in the presence of high doses of dexamethasone; overexpression of the V3R receptor for vasopressin and aberrant expression of the V2R renal receptor for vasopressin and its pharmacological analog desmopressin; and overexpression of the CRH receptor.[97,98] Consequently, a clear suppression of cortisol production following a high-dose DST (HDDST) and an increase in plasma ACTH and cortisol concentrations following intravenous injection of CRH or desmopressin are expected in Cushing disease. Typically, ectopic ACTH-secreting tumors lack this molecular equipment and do not respond to the tests. However, well-differentiated neuroendocrine tumors may express the glucorticoid, V2R/V3R, and CRH receptors and, consequently, respond to tests in a similar way to corticotroph adenomas. Therefore, no single test is free of false positives and false negatives, and any test should be interpreted from a probabilistic perspective, (e.g., the stronger the response, the more likely the diagnosis of Cushing disease). One important caveat for any test is the cyclical Cushing syndrome that may occur both in Cushing disease and EAS, in which the response to tests during periods of tumor inactivity may reflect normal pituitary corticotroph function. Among other considerations, no large prospective study conducted in difficult cases has assessed the usefulness of tests in a "real-life" setting where available information (age, sex, intensity of hypercortisolism, etc.) affects the pretest probability, as mentioned previously. Lastly, time-consuming tests should not be performed in urgent cases where patients have intense hypercortisolism.

Tumoral Markers. An increased plasma concentration of POMC suggests the presence of an aggressive ACTH-secreting tumor such as a pituitary corticotroph macroadenoma or an ectopic ACTH-secreting tumor.[15] However, this assay is not widely available, and increased POMC concentrations are usually associated with obvious pituitary and ectopic ACTH-secreting tumors on imaging.[15]

Measurements of biochemical markers indicating the ectopic source of ACTH are not routinely performed but may be part of the workup for patients where there is a strong suspicion of EAS. These include calcitonin, plasma or urinary metanephrines, and urinary 5-hydroxyindoleacetic acid for ACTH-secreting medullary thyroid carcinoma, pheochromocytoma, and gut neuroendocrine tumors, respectively.[35,101] The measurement of plasma chromogranin, nonspecific enolase, and other nonspecific markers has been abandoned, given their poor diagnostic utility.

High-Dose Dexamethasone Suppression Test. The rationale for the HDDST is that the negative feedback control of ACTH is reset to a higher level than normal in Cushing disease. Consequently, cortisol levels may decrease after high doses of dexamethasone, contrary to what occurs in EAS. In the modern era, the time-consuming and cumbersome standard 2-day test combined with urine collection has been replaced by the overnight HDDST, which has been found to have equivalent diagnostic performance in a direct comparison of the two approaches.[107] Eight milligrams of dexamethasone are given orally as a single dose at midnight, and the 8:00 AM serum cortisol value the following morning is compared with that of the previous day. Although the test is usually well-tolerated, administering a large dose of dexamethasone in patients with intense hypercortisolism may worsen comorbidities, such as hyperglycemia, or precipitate steroid psychosis. In both older and more recent studies, a 68% to 71% suppression of plasma cortisol has been associated with 64% to 71% sensitivity and 99% to 100% specificity for the diagnosis of Cushing disease.[107,108] Consequently, the performance of the HDDST has been greatly questioned, and the test has been abandoned by several expert teams based on several arguments.[35,109,110] These include a complete overlap between the responses in Cushing disease and EAS, with a range of suppression from 0% to 99% in each clinical situation.[106] In the same vein, an Italian study evaluating the performance of the test by receiver operating characteristic analysis, showed that the results obtained using the percentage decrease in cortisol were not different from those

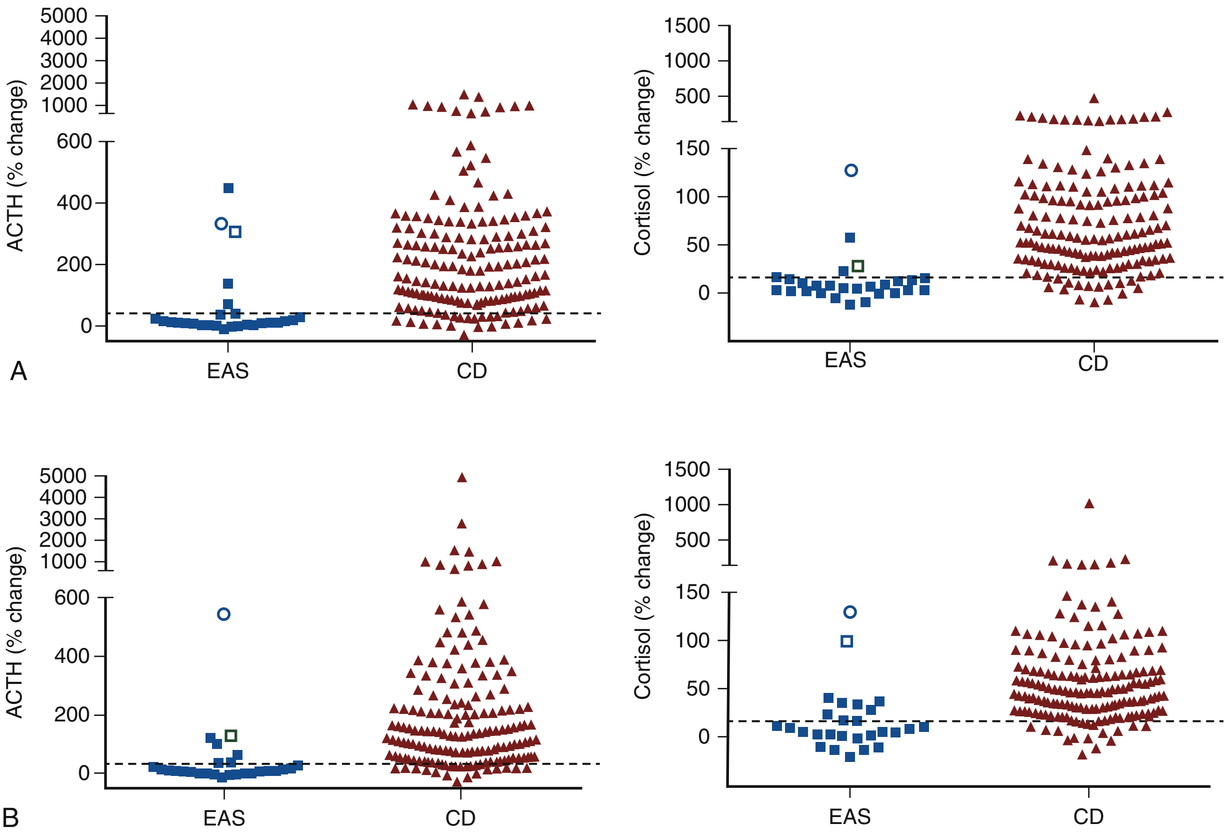


Fig. 7.11 Individual values of relative increases (%) in plasma cortisol and adrenocorticotropic hormone (*ACTH*) **A,** during the corticotropin-releasing hormone (*CRH*) test and **B,** desmopressin test in 167 consecutive patients with Cushing disease (CD) and 27 patients with ectopic ACTH syndrome (*EAS*). The open squares and open circles represent plasma cortisol and ACTH concentrations of two patients with EAS who responded to both the CRH and desmopressin tests. Dashed lines represent thresholds for the positive response to the tests. Note that no test allows complete differentiation between the two groups of patients. (Reproduced with permission from Frete C, Corcuff JB, Kuhn E, et al. Non-invasive diagnostic strategy in ACTH-dependent Cushing syndrome. *JCEM.* 2020;105:1-12.)

obtained by chance.[111] A large study also reported that the diagnostic performance obtained with a 30% plasma cortisol suppression following the LDDST was approximately the same as that of the HDDST.[112]

Corticotropin-Releasing Hormone Stimulation Test. The CRH test is probably the most widely used and has the best-documented performance for the diagnosis of Cushing disease. Synthetic ovine CRH or synthetic human CRH is injected intravenously at a dose of 100 µg or 1 µg/kg body weight. Response is assessed by measuring the percentage increase in circulating ACTH and cortisol concentrations at 15, 30, 45, and 60 minutes after administration of CRH as compared with baseline preinjection values. As the response occurs rapidly after CRH injection, several authors have proposed limiting blood sampling to the first 30 minutes.[108,113,114] The test can be performed irrespective of the time of the day and is well-tolerated, with very few and mild side effects (facial flushing, headaches). A positive response is highly suggestive of Cushing disease, although there is no universal consensus regarding the criteria for interpreting the test. In most large studies, the ACTH response was found to perform better than the cortisol response, and, using variable thresholds of a 35% to 105% increase in plasma ACTH, the test had 83% to 93% sensitivity and 90% to 100% specificity.[108,111,113,114] Other studies have stated that the cortisol response performs better than the ACTH response when using a threshold of a 14% increase.[113] In clinical practice, we recommend considering both the cortisol and ACTH responses. In the largest published study to date, which involved 167 patients with Cushing disease and 27 challenging cases of ectopic ACTH secretion, a combined response of 17% and 37% increase in cortisol and ACTH, respectively, was associated with a positive predictive value for Cushing disease of 97%[110] (Fig. 7.11). The stimulatory action of CRH can be strengthened by the synergistic effect of desmopressin, but the combined desmopressin-CRH test is rarely used.[115]

The Desmopressin Test. Desmopressin is a pharmacological analog of vasopressin that binds to V2R, and the rationale of the test is to take advantage of the aberrant expression of the V2R in adenomatous corticotrophs.[97,98] Desmopressin is injected intravenously at a dose of 10 µg, and response is assessed by measuring the percentage increase in circulating ACTH and cortisol concentrations 15, 30, 45, 60, 90, and 120 minutes after the injection as compared with baseline preinjection values. The test is well-tolerated with mild side effects (flushing, tachycardia, mild decrease in blood pressure, headache), but restriction of fluid intake for 12 hours is recommended to avoid water overload. In a metaanalysis, a positive response based mainly on predetermined arbitrary thresholds for ACTH and cortisol responses was observed in 84% of patients with Cushing disease, but also in 31% of the limited number of included patients with EAS, suggesting that the test had suboptimal specificity.[98] This may be because of the expression of vasopressin receptors in a number of well-differentiated ectopic ACTH-secreting tumors bias in published studies.[98,115] In the largest published study to date, involving 167 patients with Cushing disease and 27 challenging cases of ectopic ACTH secretion, a combination of cortisol response of 18% or more and ACTH increase greater than 33% showed approximately the same performance as the CRH test, with a 97% positive predictive value and only a slightly lower specificity (81% vs. 85%) (Fig. 7.11).[110]

Combination of Tests. In view of the limitations of single tests, the performance of at least two dynamic tests is recommended.[1,101] Concordant positive results reduce the risk of false positives. The combination of HDDST and the CRH test[1,101,108,116] has been the most accepted in this perspective. In a large study, the combination of the LDDST and CRH tests had similar diagnostic accuracy (sensitivity 94%, specificity 97%) to the combination of the HDDST and CRH tests.[112] Some groups have suggested performing the HDDST, CRH, and desmopressin tests as a trio[116] when pituitary imaging is not conclusive, whereas other groups that have abandoned the HDDST have proposed the combination of the CRH and desmopressin tests.[110] However, it should be kept in mind that, although positive responses to several tests clearly shift the probability in favor of Cushing disease, this does not formally exclude EAS.[110,111] In addition, the results of tests have been reported to be discordant in at least one in four patients.[35,101,110]

Pituitary Imaging. Pituitary MRI is the most useful imaging technique for the visualization of corticotroph adenomas. Because more than 80% of corticotroph adenomas are microadenomas, with a median size of 4 to 5 mm,[110,117,118] the technical aspects of pituitary MRI play an important role.[119] Despite optimal technique, it is classically accepted that standard pituitary MRI with spin-echo sequences and gadolinium enhancement identifies approximately 50% of corticotroph adenomas.[119] However, some studies have shown an improvement over time in the rate of detection to 71% to 81%, with refinements in MRI techniques and the growing expertise of dedicated neuroradiologists.[110,117,120] Alternative MRI protocols such as dynamic MRI, combining rapid and early image acquisition following a bolus injection of contrast medium, or fast spoiled gradient recalled echo, which permits imaging with 1-mm slice intervals, offer greater sensitivity for localizing corticotroph microadenomas. However, these techniques also yield more false-positive scans.[119] False positives, which are also observed with spin-echo sequences, include erroneous localization of pituitary adenomas in patients with Cushing disease or the suggestion of a microadenoma in a patient with ectopic ACTH syndrome.[103,120] The frequency of pituitary lures increases inversely with the size of the lesion and has been found in at least 10% of patients without pituitary disease.[99] Consequently, some algorithms recommend consideration of MRI as positive only when pituitary images suggesting an adenoma are greater than 6 mm.[1,101,121] This 6-mm threshold, which originates from a 1994 study,[99] has been debated[110] because it does not take into account more recent technological improvements in MRI.[117] Although lesions 2 mm or smaller should not be considered as positive, given the poor reliability of pituitary MRI at this size, analysis of qualitative characteristics of pituitary lesions by experienced radiologists/neurosurgeons is also a major contributor to the performance of pituitary MRI.[110,122] Increasing the strength of the main magnetic field to 3 Tesla (T) allows greater spatial resolution. However, definitive superiority of 3 T MRI over 1.5 T MRI has not yet been clearly demonstrated.[119]

Promising results have been obtained with positron emission tomography (PET)/CT using ^{11}C-methionine, an amino-acid tracer that accumulates in pituitary adenomas with a low grain uptake. The coupling of PET/CT with MRI, which is currently under evaluation, allowed localization of corticotroph microadenomas, invisible on MRI, with acceptable anatomical precision.[123] PET, using the more specific tracer CRH, has recently been studied using molecular imaging in a small number of patients and may pave the way for future imaging strategies.[124]

Bilateral Inferior Petrosal Sinus Sampling. ACTH is secreted into the two cavernous sinuses that drain into the inferior petrosal sinuses. Assuming that the secretion of normal corticotrophs is suppressed by chronic hypercortisolism, direct measurement of ACTH concentration in the petrosal sinuses after catheterization via the femoral and comparison of its concentration with that in a peripheral vein (central-to-peripheral plasma ACTH gradient) is therefore a direct way to demonstrate that the inappropriate ACTH secretion originates from

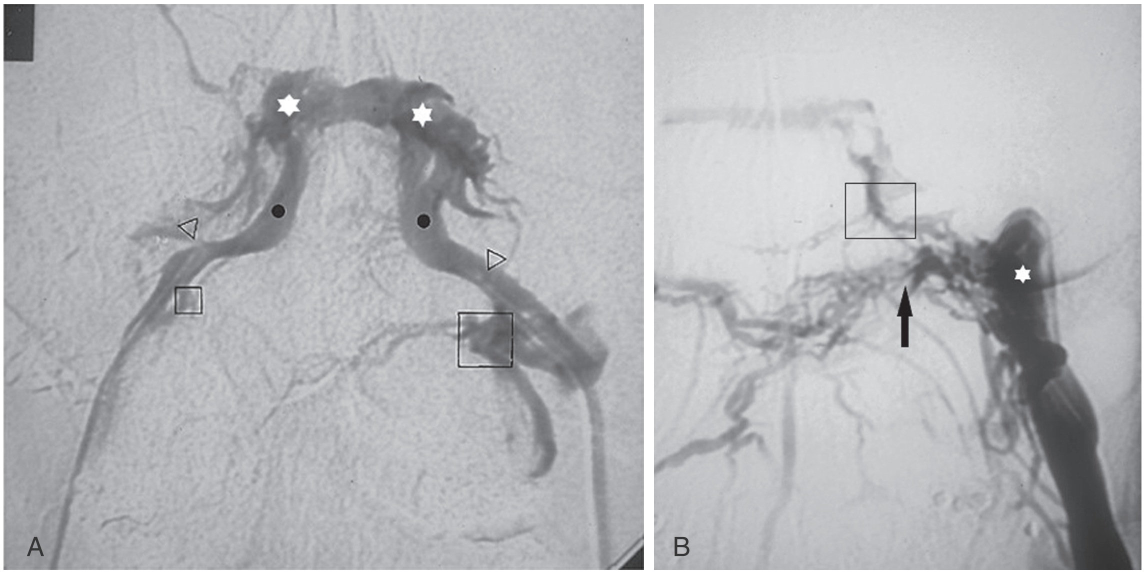


Fig. 7.12 Angiograms performed during bilateral inferior petrosal sinus (IPS) sampling. **A,** Adequate localization of the catheter tips (*arrow head*) in the "horizontal" part of the IPS. Note that the two IPSs have approximately comparable diameters. Dots: vertical part of the IPS situated just below the cavernous sinus (*star*). Square: Basilar venous plexus of the clivus that drains nonpituitary blood into the jugular vein. The correct localization of catheter tips prevents the dilution of samples by nonpituitary blood. **B,** Anatomical variant with atrophic left IPS (*square*) that cannot be catheterized. The catheter tip is located in the jugular bulb (*arrow*), allowing dilution of samples with nonpituitary blood, which may result in false negatives bilateral inferior petrosal sinus sampling results.

the pituitary. To overcome the pulsatility of ACTH secretion, which may result in the absence of a central-to-peripheral ACTH gradient,[125] additional blood samples are drawn at the three sites, usually, 2, 5, and 10 to 15 minutes following peripheral intravenous injection of 100 µg of CRH. Given that some pituitary corticotroph adenomas do not respond to CRH, or CRH may not be available, desmopressin may be used in place of CRH with approximately similar sensitivity and specificity.[125] Other studies have proposed simultaneous administration of CRH and desmopressin, with slightly different thresholds for analysis of the results.[126]

Adequate catheterization of inferior petrosal sinuses (in the "horizontal" segment; see Fig. 7.12) is of major importance to avoid dilution of ACTH in the petrosal sinuses by nonpituitary blood. Cavernous-inferior petrosal sinus venography must be performed during and at the end of the procedure to assess venous anatomy and to ensure correct localization of the catheter tips. In expert hands and following considerable experience, optimal catheterization of inferior petrosal sinuses is obtained in more than 90% of patients.[125] Transient otalgia during the ascent of catheters and groin swelling or hematoma are common. The incidence of serious adverse events, such as a cerebrovascular accident or transient cranial nerve palsy with experienced teams, is less than 1%.[127] To prevent deep venous thrombosis and pulmonary embolism, a rare event in experienced hands and with recent developments in endovascular materials, some groups, although not all, administer heparin for anticoagulation before the procedure.[125]

The universal criterion for the diagnosis of Cushing disease is a central-to-peripheral plasma ACTH gradient of 2.0 or more before CRH administration or 3.0 or more after CRH injection.[125,128] To date, BIPSS has the best performance for the diagnosis of Cushing disease, with sensitivities ranging from 85% to 100% and specificities from 82% to 100%. In 11 published monocentric studies enrolling at least 50 patients, the overall sensitivity based on 1309 patients with Cushing disease was 92%, and the specificity based on 97 patients with EAS was 96%.[125] The baseline sensitivity improved by 2% to 21% following CRH injection.[125,128] False negatives are more common and usually result from incorrect catheter placement or anatomical variations in venous drainage (Fig. 7.12) allowing dilution of pituitary blood with venous flow coming from other areas of the brain stem.[125,129] Expert analysis of angiograms should therefore be performed in conjunction with analysis of ACTH concentrations. Based on analysis of 501 procedures, a study identified that low ACTH concentrations in the petrosal sinuses (400 ng/L) associated with low central-to-peripheral ACTH gradients should raise suspicion of a false negative.[120] Prolactin measurement in inferior petrosal sinus blood samples has been proposed as a means of recognizing whether pituitary venous blood has been accurately sampled, with recalculation of central-to-peripheral ACTH ratios after normalization using central-to-peripheral prolactin ratios.[125] However, the criteria for interpretation vary between centers, and a degree of uncertainty remains.[130] Bilateral internal jugular venous sampling has the advantage of simplicity and has been proposed as an alternative to BIPSS.[131] However, the procedure has not gained wide acceptance, as its sensitivity is clearly inferior to that of petrosal sinus sampling. False positives occur rarely in patients with ectopic ACTH syndrome and are not always fully explained.[125,129] Aside from the rare situation of an ectopic CRH-secreting tumor that is stimulating pituitary corticotrophs,[105] an important prerequisite is the suppression of normal anterior corticotroph function by chronic hypercortisolism. Incomplete suppression of physiological corticotroph function in patients with EAS and intermittent/cyclical cortisol secretion, or following a previous treatment with anticortisolic drugs, may therefore result in false positives. It is therefore mandatory to perform BIPSS in a state of chronic hypercortisolism and to verify this with the help of routine biochemical tools.

The use of BIPSS to localize corticotroph microadenomas within the pituitary has been proposed based on the assumption that each pituitary half drains into the homolateral cavernous sinus. Consequently, an ACTH gradient of 1.4 or more between the right and left petrosal sinus would predict lateralization of the tumor and would favor a hemihypophysectomy in the case of negative surgical exploration.[132] However, venous drainage in the petrosal sinus is asymmetric in a number of patients,[133,134] and most subsequent studies have reported low accuracy that was sometimes no different from chance.[133,134] Using higher thresholds for the intersinus ACTH gradient[120] or restricting the analysis of this gradient to patients with symmetrical petrosal venous drainage resulted in a marginal improvement in performance.[133,134] The accuracy of cavernous sinus sampling for predicting the lateralization of corticotroph adenomas has been explored, but was not better than that of BIPSS.[133,134] In one series, hemihypophysectomy based on lateralization data was curative in only about 50% of patients.[134]

Overall Strategy for Differential Diagnosis. Several algorithms for the differential diagnosis of ACTH-dependent Cushing syndrome have been proposed (Fig. 7.13). However, a number of considerations need

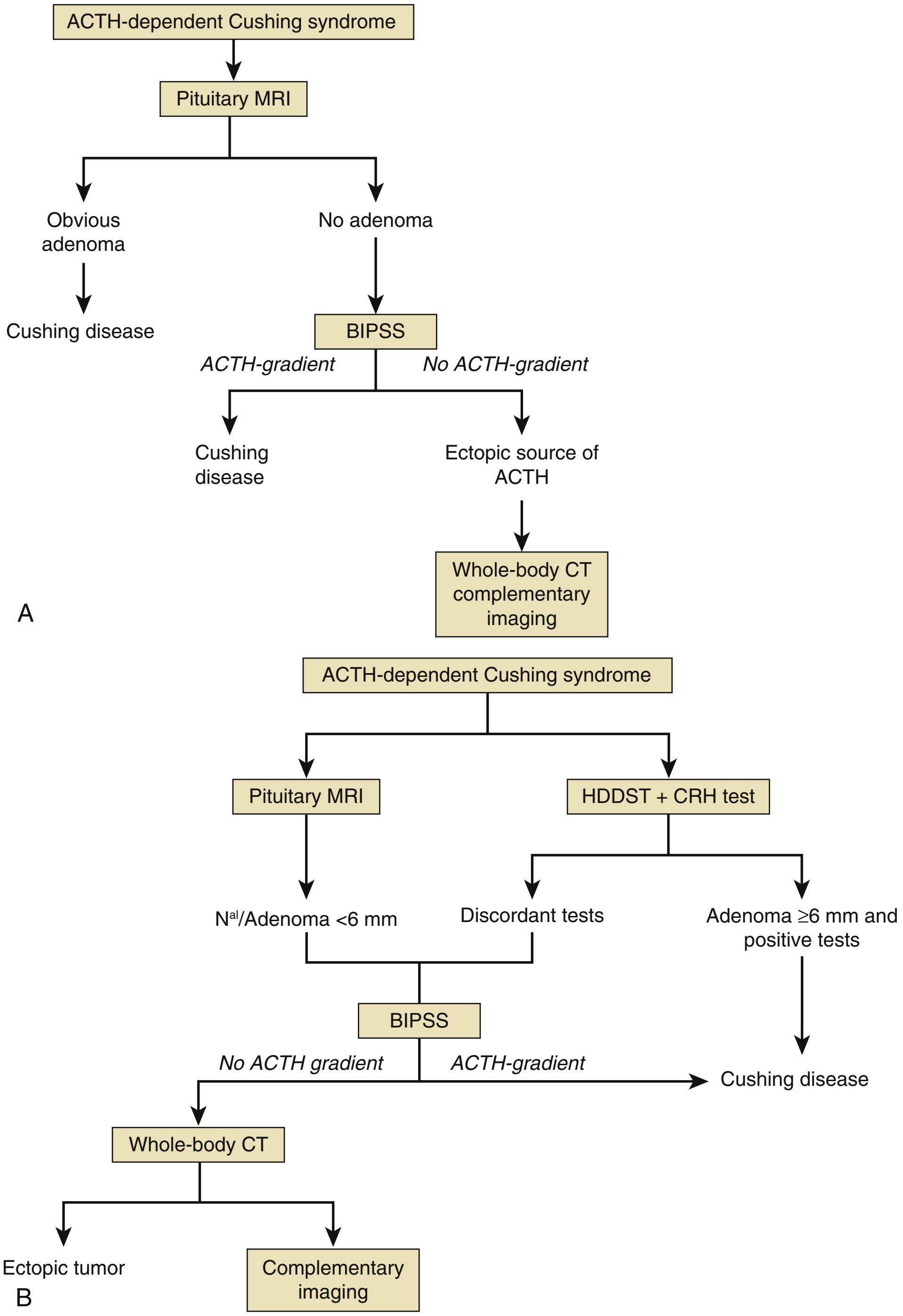

ACTH-dependent Cushing syndrome
Pituitary MRI
Obvious adenoma
No adenoma
Cushing disease
BIPSS
ACTH-gradient
No ACTH-gradient
Cushing disease
Ectopic source of ACTH
Whole-body CT complementary imaging
A
ACTH-dependent Cushing syndrome
Pituitary MRI
HDDST + CRH test
Nal/Adenoma <6 mm
Discordant tests
Adenoma ≥6 mm and positive tests
BIPSS
No ACTH gradient
ACTH-gradient
Cushing disease
Whole-body CT
Ectopic tumor
Complementary imaging
B

Fig. 7.13 Three currently used algorithms for the differential diagnosis of adrenocorticotropic hormone (*ACTH*)-dependent Cushing syndrome in patients with a strong pretest probability of having Cushing disease. Owing to the absence of formal prospective studies comparing diagnostic algorithms with pertinent endpoints, various diagnostic strategies may be used to differentiate Cushing disease from ectopic ACTH syndrome. **A,** Because bilateral inferior petrosal sinus sampling (*BIPSS*) has the best overall diagnostic accuracy, this algorithm does not involve dynamic tests when pituitary magnetic resonance imaging is normal. **B,** Most frequently cited algorithm.

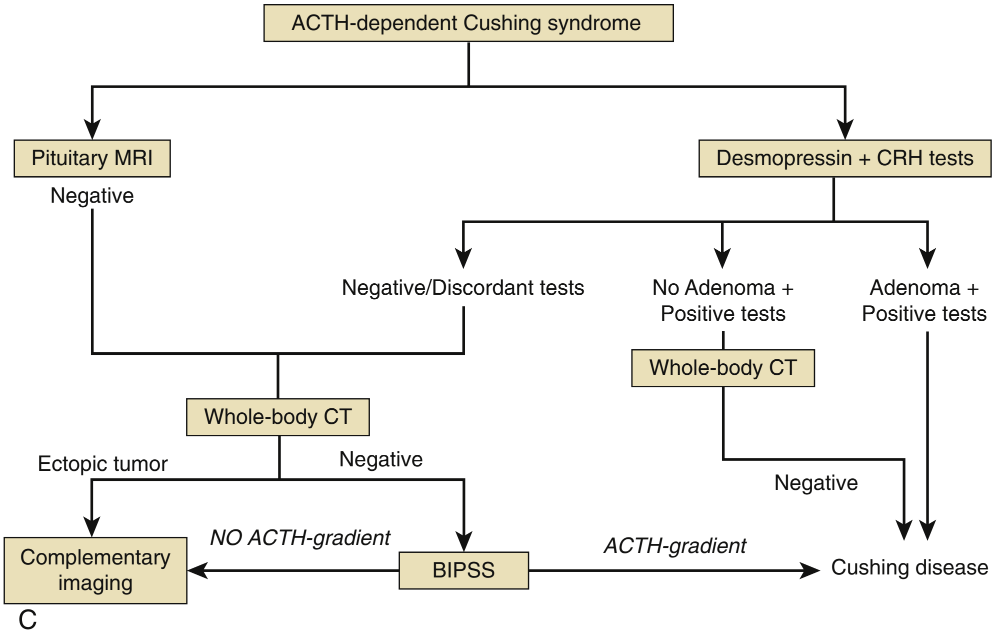


Fig. 7.13, Cont'd C, Alternative algorithm. The debated high-dose dexamethasone suppression test is replaced by the desmopressin test. A neck-to-pelvis thin-slice computed tomography scan is always performed before BIPSS to reduce the need for BIPSS in selected patients. *CRH*, corticotropin-releasing hormone. (**B** reproduced with permission from Lacroix A, Feelders RA, Stratakis CA, et al. Cushing syndrome. *Lancet.* 2015;386:913-27. **C** reproduced with permission from Frete C, Corcuff JB, Kuhn E, et al. Non-invasive diagnostic strategy in ACTH-dependent Cushing Syndrome. *JCEM.* 2020;105:1-12.)

to be taken into account. Firstly, there is no formal prospective study comparing diagnostic algorithms using relevant endpoints such as histological confirmation of the disease or surgical remission. Therefore, no algorithm can be recommended in an evidence-based manner. Secondly, no study has precisely evaluated the benefit and cost effectiveness of different strategies according to the global context, (e.g., the pretest probability of Cushing disease). For example, in a male patient with recent onset of intense hypercortisolism and with greatly increased plasma ACTH concentrations, the probability of EAS may be superior to that of Cushing disease, and performing thin-slice neck-to-pelvis CT first may be more relevant than pituitary MRI or a time-consuming battery of biological tests. Thirdly, the best diagnostic performance is obtained in expert centers with multidisciplinary expert teams.[122] This reinforces the importance of referring patients to expert centers. Fourthly, no algorithm can ensure 100% sensitivity and specificity, and thus all testing procedures should be considered to be probabilistic.[101]

Given the performance of BIPSS, some investigators recommend that all patients with ACTH-dependent Cushing syndrome and negative pituitary MRI undergo BIPSS. Although BIPSS has the best overall diagnostic accuracy, it is an invasive procedure requiring an expert center and is less readily available than noninvasive tests. In the same vein of reducing investigations, but at the opposite extreme in terms of invasiveness, some endocrinologists suggest performing pituitary surgery solely on the basis of an obvious adenoma (>6 mm) seen on pituitary MRI.[109]

The most recommended algorithm starts with a noninvasive approach and includes HDDST and CRH tests combined with pituitary MRI. BIPSS is scheduled when tests are negative/discordant, or if there is no obvious pituitary image on MRI (≥6 mm).[1,101,121] As discussed previously, the 6-mm criterion for positive MRI is debated.[110] Negative results of BIPSS should also be interpreted in relation to biochemical tests. Patients with frank positive responses to both CRH and HDDST have a high probability of having Cushing disease, and a nonsignificant central-to-peripheral ACTH gradient may reflect a BIPSS false negative. Transsphenoidal surgical exploration has been proposed for such patients[135] after negative results from extensive body imaging for an EAS.[35,101,136] Other groups have suggested performing three dynamic tests (HDDST, CRH, and desmopressin) and proceeding with pituitary surgery in patients with negative pituitary MRI when all tests are positive, without performing extensive body imaging, given the high likelihood of Cushing disease.[116,137]

Other publications have proposed modification of this "classical" algorithm to reduce the need for BIPSS while ensuring a 100% positive predictive value for Cushing disease.[110,122] This consists of pituitary MRI, desmopressin, and CRH tests in conjunction with expert radiological interpretation. If MRI shows a convincing pituitary image, and both tests indicate Cushing disease, the patient can proceed to surgery. Otherwise these tests were followed by modern thin-slice neck-to-pelvis CT that allowed rapid identification of an ectopic tumor. In patients with positive tests and negative CT, pituitary surgery was performed despite normal MRI. In the remaining patients with discordant or negative tests, BIPSS was performed if no lesion was seen at CT. The combination of tests, MRI, and possibly CT gave a correct diagnosis in 47% of a cohort including 167 patients with Cushing disease and 27 with challenging cases of EAS, halving the need for BIPSS in those patients where it is currently recommended.

TREATMENT OF CUSHING DISEASE

General Considerations

The duration of exposure to cortisol excess may play a major role in the persistence of comorbidities and reduced life expectancy after remission (see Prognosis for Cushing Disease: Persistent Morbidity and Mortality), suggesting that, once diagnosed, the rapidity with which

KEY POINTS

- Pituitary surgery performed by an expert surgeon with a high number of procedures is the only treatment that can offer a permanent cure with a low rate of adverse effects, and is the first-line treatment for CD.
- Given the frequency of recurrences, life –long follow-up is mandatory in patients in with post-surgical remission of CD
- The choice of second-line treatment when pituitary surgery has failed or is impossible is complex, must be individualized, shared by patients and discussed in an expert multidisciplinary team.
- Comorbidities may persist despite CD remission and explain at least in part the persistent deterioration in quality of life compared to other pituitary pathologies and long-term excess mortality. Treatment of the long-term sequelae of CD is therefore an important part of the treatment of the disease.

intervention controls hypercortisolism is of paramount importance. In addition, careful follow-up and treatment of persisting comorbidities after cure is an important component of therapy.[138] Resection of the "primary" tumor in the first instance, regardless of its etiology, is the rule in the treatment of Cushing syndrome.[138] Various secondary options are available when surgery fails, and significant advances in the development of pharmacological treatments have been made over the last decades. However, the treatment of Cushing disease remains complex, and, despite multiple treatments, biochemical control may not be achieved in up to 30% of patients, as has been reported in large multicenter cohorts.[139] This emphasizes that treatment of Cushing disease requires a dedicated expert team headed by a dedicated endocrinologist.

Pituitary Surgery. Surgical removal of the pituitary adenoma performed by an expert surgeon is the only therapeutic option that can offer definitive cure with a low rate of adverse effects, and is the first-line treatment for Cushing disease.[138]

Results and Safety. Some centers use, either routinely or in patients with more severe clinical features, preoperative medical adrenal blockade to attain a period of eucortisolemia for several weeks to several months before surgery.[140] The aim here is to allow reversal of some of the comorbidities of hypercortisolemia that may inhibit wound repair and cause other complications in the perioperative period. To date, this rationale is mainly empirical and mostly used in patients with severe Cushing disease. No difference in mortality and morbidity has been found between pretreated patients or patients operated on immediately.[140] The disadvantage of this strategy is that the normal corticotroph cells may be disinhibited by the time of surgery, so expected hypocortisolism after successful tumor resection does not occur, and this indicator of remission is thus unreliable.[140] Based on retrospective studies suggesting that preoperative anticoagulation reduces the risk of thromboembolic events, some groups recommend treating all patients.[40,41]

The goal of surgery is a selective and slightly enlarged adenomectomy with the preservation of as much normal pituitary tissue as possible. The procedure should therefore be performed by a neurosurgeon who is experienced in transsphenoidal surgery because the rate of surgical complications is less frequent than when performed by surgeons who seldom carry out this operation.[141,142]

Most pituitary corticotroph adenomas are microadenomas or small macroadenomas,[110,117] therefore, the transnasal transsphenoidal route is the rule in the majority of patients. A metaanalysis including more than 6000 operated patients concluded that endoscopic surgery achieves comparable results for microadenomas, and probably better results for macroadenomas, than does microscopic surgery.[143] If a tumor is identified during surgery, it is removed selectively, ideally including some tissue surrounding the adenoma (the "pseudocapsule"). If no tumor is visible, the pituitary gland is systematically explored after cross-incision. In patients in whom no adenoma is identified, several options are possible, among which partial/subtotal hypophysectomy, (e.g., removal of two-thirds of the lateral and posterior pituitary tissue), appears to be the best compromise between a high success rate and low rate of surgery-related hypopituitarism.[142] Some surgeons resect one half of the pituitary gland on the side of lateralization on BIPSS, but the accuracy of BIPSS for this purpose may not be better than chance, and, in one series, hemihypophysectomy based on lateralization data was curative in only about 50% of patients.[134]

Mortality rates associated with transsphenoidal surgery are approximately 1% to 2%.[138,142] Transient diabetes insipidus is probably the most common complication, being reported in more than 10% of patients. Other perioperative complications, including cerebrospinal fluid leak, meningitis, and profuse bleeding, occur in less than 10% of patients operated on by an experienced neurosurgeon. Permanent complications, such as persistent diabetes insipidus and injury to the optic nerve or nerves of the cavernous sinus (causing ptosis or diplopia), occur much less frequently. However, hypopituitarism, particularly GH deficiency, is common (53%–59% of patients), with other anterior hormone deficits occurring in approximately 35% to 49% of patients.[138,142] Such complications are more common after resection of larger tumors, after repeat surgery, and for small tumors, depending on the aggressiveness and extent of surgery. In a metaanalysis, the mean frequency of hypopituitarism was 7%, 20%, and 80% in patients who had undergone adenomectomy, hemihypophysectomy, and subtotal hypophysectomy, respectively.[142]

Variable criteria of various biochemical tests have been used to characterize postoperative remission.[141,142] This may explain the wide variations in reported remission rates, which ranged between 25% and 97% (mean 78%) in metaanalyses.[141,142] Various favorable prognostic factors have been established for postoperative remission.[118,144] These include microadenoma (vs. macroadenoma), lack of invasion of the dura or cavernous sinus, visualization of the adenoma during surgery, histological confirmation of an ACTH-secreting tumor, and the experience of the neurosurgeon. In rare instances, failure of pituitary surgery may be related to adenomas arising in ectopic sites such as the cavernous sinus or along the pituitary stalk. There is a continuing debate about whether visualization of pituitary microadenomas on MRI is a favorable prognostic factor for remission and whether to perform surgery in patients in whom no pituitary adenoma is seen on pituitary MRI.[118,142] We, and others, favor surgery once the diagnosis of Cushing disease is definitively established, based on the observation that the surgeon's eye is more accurate than MRI, and that the success of surgery does not differ based on the MRI results.[117,145] Based on series that showed a reduced rate of cure when MRI was negative,[118] "wait and see" alternative approaches, including pharmacological control of hypercortisolism and regular follow-up with pituitary MRI, have been suggested.[142,146,147] In our opinion, even though the rate of surgical success is slightly reduced in patients with a negative MRI, surgery should still be the first choice because it offers a definitive success rate of 70% to 85%,[117,118,142,145] and, importantly, all other therapeutic options remain possible in patients for whom surgery fails. However, the probability of identifying a pituitary microadenoma may be lower in patients with negative pituitary MRI, leading to more extensive pituitary resection and possibly to an increase in postoperative partial hypopituitarism.[145]

Despite initial remission, Cushing disease recurs in a significant number of patients. Although the criteria for remission vary among studies, two recent metanalyses of 17 to 68 studies with lengthy follow-up identified reported recurrence rates ranging from 5% to 48%, with a mean of 15% to 19%.[118,141] Recurrence occurred within 3 to 345 months (mean 47–51 months) following surgery. Most recurrences occur during the 15 years following surgery, but some recurrences were observed even 20 years after surgery, explaining an actuarial recurrence rate that may be as high as 25%.[118,141,142] Prolonged, if not lifelong, follow-up is therefore recommended.[138] The highest recurrence rates appeared to occur in studies where 10 patients or fewer were operated on per year, while the lowest rates were seen when 31 to 40 patients were operated on per year.[141] These data reinforce the recommendation that Cushing disease patients must be operated on by an expert surgeon.[138]

Postoperative Evaluation and Management. Various protocols are used at different expert centers. Some give higher than normal glucocorticoid replacement intraoperatively for 24 hours and then taper the dose within 1 to 3 days postoperatively. Where close perioperative supervision is possible, such glucocorticoid "cover" may be omitted, permitting a very early postoperative assessment of cure.[148] Morning (8:00 AM–9:00 AM) serum cortisol measurements 20 to 30 hours after the last glucocorticoid administration are the most commonly used parameter for early postoperative evaluation.[138] Other authors have proposed the use of salivary cortisol as a substitute for serum cortisol to characterize the postoperative status.[149]

What defines remission after transsphenoidal surgery is still debated, as are the prognostic factors for sustained remission and cure. As normal corticotroph cells are suppressed by the chronic hypercortisolism, it is expected that cortisol concentrations will be suppressed following complete resection of the corticotroph adenoma. Consequently, patients with hypocortisolemia less than 5 μg/dL (140 nmol/L) are classified as being in remission.[138] Complementary investigations (midnight serum or salivary cortisol, 24-hour UFC and 1-mg DST) in patients with nonsuppressed morning plasma cortisol will identify those with persistent hypercortisolism (surgical failure) and patients with eucortisolism. Patients with surgical failure and no tumor on pathology examination require reevaluation of the diagnosis of Cushing disease, including BIPSS if not performed before pituitary surgery. Of note, approximately 5% may experience delayed remission 2 to 3 months following surgery, possibly as the result of progressive necrosis of remaining tumor cells.[150]

Patients with documented eucortisolism after complementary investigations can be safely monitored without treatment and reevaluated initially at 3-monthly intervals. Although most patients will have recurrence, eucortisolism is compatible with sustained remission in some individuals.[138,151] Hypocortisolemic patients should be started on glucocorticoid replacement, with hydrocortisone at 12 to 15 mg/m^2 given in two to three divided doses, and the last dose should be taken no later than 5:00 PM to avoid long-term suppression of the HPA axis. Patients receiving hydrocortisone supplementation should be instructed that they are "dependent" on taking glucocorticoids and given appropriate counseling and education for adrenal insufficiency. Adrenal crises tend to occur early after remission of Cushing syndrome, which is likely the period of maximal HPA axis suppression.[152]

Patients should also be told to expect symptoms of malaise (asthenia, dizziness, joint and muscle aches, difficulty in concentrating) during the postoperative months, with these signs being interpreted as a withdrawal syndrome. These symptoms may persist up to 1 year, and no specific medication, including increasing the dose of hydrocortisone, has proven to be effective. The most important intervention is often continuous reassurance by the medical team.

Postoperatively, assessment for deficiencies in other pituitary hormones should be performed and the appropriate replacement regimen initiated as necessary. However, the results of evaluation of pituitary secretion in the early postoperative period may be difficult to interpret, as the suppressive action of chronic hypercortisolism on gonadotrophins, TSH, and GH secretion may take several weeks or months to resolve after remission.[51] Importantly, although GH supplementation in adult patients with persistent profound GH deficiency results in a significant improvement of quality of life and metabolic parameters,[153] other studies have warned of a possible increase in the incidence of metabolic syndrome.[154] There is no consensus on the timing of pituitary secretion testing, but is suggested to perform this evaluation when the patient recovers from corticotroph insufficiency and focus on the comparison of free T4 concentrations obtained within 1 month after surgery with preoperative T4 levels to identify hypothyroidism.[138]

Plasma sodium should thus be assessed several times during the first 2 weeks to diagnose hyponatremia, usually owing to transient inappropriate antidiuretic hormone secretion, which may occur in as many as 10% of patients.[138,142] Some glucocorticoid-induced comorbidities, including hypertension and glucose intolerance, may normalize rapidly during the postoperative period; therefore, treatments for these need to be reassessed, possibly after stopping medication. However, it is mandatory to accurately control cardiovascular risk factors during the postoperative period, given the persisting increase in frequency of acute events during the first postoperative year.[44] Similarly, deep vein thrombosis and pulmonary thromboembolism may be more frequent after pituitary surgery for Cushing disease.[40,42] There is no consensus regarding the indications for and duration of preventive anticoagulation treatment, which that may last for at least 2 months, according to some groups.[40,42]

Prediction of Postoperative Recurrence. Identification of early postoperative predictive factors of recurrence that can define high- and low-risk groups to adapt the duration and frequency of follow-up examinations has been an area of intensive research.

The lack of *USP8* mutations in adenomatous corticotrophs has recently been associated with long-term remission in a small cohort of patients, but requires additional studies.[20] Early postoperative corticotroph insufficiency is a well-established protective factor for recurrence but not an absolute criterion for definitive cure.[138,155-157] The duration of corticotroph insufficiency has also been identified as a significant predictor of recurrence.[155,157] In one study, recovery within 6 months, 1 year, and 2 years had a decreasing positive predictive value for recurrence of 64%, 61%, and 59%, respectively,[155] and no patient with corticotroph insufficiency of more than 3 years in duration experienced a recurrence.

The CRH test, performed in the early postoperative period, added no benefit for prediction of recurrence over early basal cortisol testing.[157] The desmopressin test has also been studied, based on the rationale that ACTH and cortisol responses are secondary to the stimulation of aberrantly expressed V2R, and that these receptors are expressed only in adenomatous corticotroph cells.[98] Although postoperative morning cortisol in patients with corticotroph insufficiency reflects the suppressed function of normal pituitary cells, a persistent "paradoxical" response to desmopressin would indicate the persistence of adenomatous corticotrophs that may induce late recurrence.[98] Two recent large studies have shown that the disappearance of the response to the test within 3 months following successful surgery is associated with a lower rate of recurrence in the long term.[98,156,158] The predictive value of a negative test was independent of other prognostic factors and showed a higher performance than did early postoperative morning cortisol.[156,158] Limitations of this test are that it is informative only in patients with a positive response before surgery, and that variable critical thresholds for the test are reported in different studies.

Postoperative Follow-Up. Various strategies have been proposed for long-term surveillance, but the lack of large-scale prospective comparative studies hampers attempts to define an optimal management strategy.[157] According to the Endocrine Society,[138] morning plasma cortisol concentrations should be measured before the morning intake of hydrocortisone every 3 months in patients with postoperative hypocortisolism, followed by a short Synacthen (Tetracosactrin, Cortrosyn) ACTH stimulation test when cortisol concentrations are greater than or equal to 200 nmol/L. Various strategies for tapering and discontinuing hydrocortisone are suggested when baseline or stimulated concentrations of cortisol are 450 nmol/L or higher. At this stage of recovery, we recommend precise evaluation of HPA function to ensure that the cortisol elevation is not associated with disease recurrence. LNSC or a 1-mg DST can be used to assess the precise status for this purpose and then serve as a new baseline for future assessments. Follow-up should rely on both clinical and biochemical investigations. Although appearance or progression of clinical symptoms, such as weight gain, worsening of hypertension, and hyperglycemia may indicate recurrence, monitoring commonly identifies the recurrence of biochemical abnormalities before that of clinical symptoms.[157] Biochemical evaluation is also warranted when patients are personally convinced that their disease has recurred, based on psychological perception of their state during the hypercortisolism phase.

Several studies suggest that an increase in 24-hour UFC is a late marker of Cushing disease recurrence, and that LNSC is more sensitive for diagnosing recurrence at an early stage. However, early-stage recurrence is a difficult diagnostic situation characterized by mild and widely fluctuating cortisol excess,[157,159,160] and multiple saliva samples may be necessary to ensure optimal diagnostic performance.[160] The accuracy of the 1-mg DST for this purpose has been less well documented,[70] but in our opinion is a valuable tool.[75]

Stimulation testing has also been proposed for diagnosis of recurrence at an early stage. A positive response to the desmopressin test has also been reported as occurring before the increase in UFC and sometimes before increases in LNSC.[159,161] A combined dexamethasone-desmopressin test (1 mg of dexamethasone, administered at midnight before desmopressin testing at 8:00 AM)[162] has also been proposed. The limited number of patients studied and the relatively short follow-up make it difficult to draw definitive conclusions on the usefulness of these investigations. Given its relatively poor sensitivity, pituitary MRI cannot be recommended as part of the regular follow-up in patients with postoperative remission.[138]

In summary, in the absence of consensus on the manner and duration of patient follow-up, the Endocrine Society recommends lifelong evaluation for all patients.[138] This should be specifically mentioned to patients. Individualized follow-up is still difficult to define, owing to problems in accurately stratifying patients at a low versus a high risk of recurrence. In our opinion, surgery performed in an expert center, histological confirmation of a corticotroph microadenoma, followed by a very low early postoperative cortisol level, prolonged corticotroph insufficiency, and lack of persistent postoperative response to the desmopressin test would favor a less rigorous follow-up schedule, with yearly examinations and a warning to patients that consultation is mandatory if clinical symptoms reappear after this period. More stringent follow-up would be preferable for other patients. Using a vigilant strategy to diagnose recurrence at an early stage is supported by small studies that demonstrate a clinical benefit of treating patients with early-stage recurrence despite normal UFC.[163]

Treatment Options After Postoperative Recurrence or Persistence

Second-line treatments include repeat pituitary surgery, pituitary radiotherapy, medical treatments, and bilateral adrenalectomy (BLA). The choice of second-line treatment varies based on several parameters, including the severity of hypercortisolism, pituitary oncological status, previous medical history, age and sex, concomitant medications and potent drug–drug interactions, patient factors such as a wish for fertility, efficacy profile, availability, and cost of treatments.[138] The choice between the different options is complex, as each modality has pros and cons, and there is no “one size fits all” treatment. Consequently, the therapeutic strategy should be discussed by an expert multidisciplinary team.[138] The complexity of second-line strategies is illustrated by a multicenter retrospective study that concluded that, at the last follow-up, biochemical control was not achieved, despite multiple treatments, in 30% of patients.[139] Apart from expert team discussions, shared decision-making with the patient is also highly recommended.[138]

Repeat Transsphenoidal Surgery. The indications for repeat pituitary surgery are slightly different in persistent hypercortisolism versus recurrent hypercortisolism. Persistence of Cushing disease corresponds to three main scenarios: removal of nonadenomatous pituitary tissue (with negative histology), incomplete removal of a microadenoma within the pituitary (with positive histology), or incomplete removal of an adenoma with extrasellar invasion. The expertise of the neurosurgeon performing the first surgery, the type of resection, and postoperative pituitary imaging need to be considered in this situation. Early repeat surgery is advisable in the two first scenarios; that is, if subsequent pituitary MRI is positive, or in those patients previously operated by a nonexpert neurosurgeon. Repeat surgery is not indicated in patients with cavernous sinus or massive dural invasion identified at the initial procedure.[121]

Recurrence of Cushing disease usually corresponds to regrowth of the adenoma from residual remnant cells at the site of the initial resection and the adjacent cavernous sinus dura. Repeat surgery is therefore an option because the adenoma can be found at, or immediately contiguous to, the site at which it was found originally.[121] Some teams, but not all, reoperate only if MRI if positive, as this is supposed to be a positive prognostic factor for the success of a second surgery. However, dural invasion may explain the lower remission rates and second recurrence following repeat surgery as compared to initial surgery.[118,164] More extensive resection is usual in repeat surgery, such as subtotal hypophysectomy, and morbidity, including hypopituitarism, is increased as compared to initial surgery.[118,142,164]

Pituitary Radiotherapy. Radiotherapy can be performed as a first-line treatment when surgery is not feasible, although it is mainly used as second-line treatment after surgical failure for its antisecretory effect or for oncogenic purposes in case of growing macroadenomas and large postsurgical remnants. Regardless of the technique used, the antisecretory effect is delayed, and it is necessary to first ensure that a medical treatment is effective and well-tolerated before planning radiotherapy.

Two types of radiotherapy may be performed: conventional or fractionated radiotherapy and radiosurgery. Both are performed with stereotactic planning to optimize the tumor dose and to minimize radiation to other areas. Conventional radiotherapy uses ionizing x-rays at a dose of 4500 to 5000 cGy, fractionated in 25 to 30 sessions over approximately 5 weeks. Radiosurgery applies a very high energy to the target in only one session. Radiosurgery uses various sources of radiation such as gamma rays from a Cobalt source (Gamma Knife) or from linear accelerators (LINAC, CyberKnife), or radiation with heavy particles such as protons. Proton beam therapy is associated with a sparing of the deeper tissues from unwanted radiation, a condition that might be preferable in children. One limitation of radiosurgery is the risk of inadvertent radiation-induced injury of cranial nerves (II, III, IV) or of the optic chiasm, in case of close proximity to the adenoma, which has been reported in 1.5% (range: 0%–6%) of cases. This does not occur with conventional radiotherapy, because of the fractionation of the dose delivered.[165] Because of its rapidity and convenience, radiosurgery has become a modality of choice for irradiation of small tumor remnants located in the sellar and parasellar regions. Two recent studies suggest that the efficiency of whole-sellar radiosurgery in patients with normal MRI is similar to that of patients with an identifiable target.[166,167] Therefore, conventional radiotherapy is currently used mostly in patients who are not candidates for radiosurgery owing to large tumor size or proximity to the optic structures (distance to the chiasm <5 mm).

A thorough metaanalysis published in 2018 reported biochemical remission rates of 68% (range: 46%–89%), with a mean time to remission of 15 to 24 months, using conventional radiotherapy as adjuvant therapy. Remission rates and time to remission for radiosurgery were roughly similar, at 66% (27%–100%) after 7 to 40 months.[165] However, similar to surgery, the efficiency of pituitary radiotherapy and comparison of the performance between conventional radiotherapy and radiosurgery is difficult because of highly variable criteria for remission of hypercortisolism and variable duration of follow-up across studies.[165] Although recurrence is considered rare following conventional radiotherapy, the largest series published, involving 278 patients treated by Gamma Knife, reported an 18% recurrence rate after apparent control of hypercortisolism, suggesting that care must be taken before drawing any conclusion regarding the long-term effectiveness.[166] Whether the rate of control of hypercortisolism is lower in radiosurgery but occurs earlier than with conventional radiotherapy remains debatable.

There is no consensus on prognostic factors for the success of radiotherapy, but, similar to pituitary surgery, radiotherapist expertise and treatment center patient volume positively influence remission rate.[165] A potential negative effect of ketoconazole treatment through an action at the pituitary level when given during radiosurgery has also been suggested.[168] Similar concerns have been reported for cabergoline during conventional radiotherapy.[169] Repeat radiosurgery in patients with Cushing disease that is refractory to both prior surgery and radiosurgery has been studied in a small series, and biochemical remission was achieved in approximately half of the patients.[170] However, 10% morbidity in terms of visual loss and diplopia was observed.

Tumor control, defined as a decrease in or stability of tumor volume on MRI, is similar between conventional radiotherapy and radiosurgery, and is achieved in approximately 90% of patients.[165]

The development of pituitary insufficiency is a frequent side effect of radiotherapy. The occurrence of new pituitary deficits is reported in up to 40% of patients after conventional radiotherapy[165] but may involve more than 70% of patients over the long term.[142] Furthermore, information on how pituitary function was evaluated, especially GH and sex hormones, is frequently lacking, making it possible that hypopituitarism was underdiagnosed in many studies. In a large multicenter Gamma Knife study, 25% of patients developed new hormone deficiencies during a mean follow-up of 5.6 years,[166] suggesting that the tolerance of radiosurgery is better than that of conventional radiotherapy. These results may be biased by a shorter follow-up than that cited for conventional radiotherapy.[165] The relative risk of developing a second intracranial tumor after pituitary radiotherapy is increased 3- to 4-fold, especially in patients irradiated at a young age, but with a low absolute risk (~1%) that is similar after conventional radiotherapy and radiosurgery.[171] An increased rate of ischemic stroke following radiotherapy has not been described in Cushing disease.[165] The possibility of a mild impact of radiotherapy on cognitive function is still a matter of debate.[168]

Therefore, continuous endocrine follow-up after pituitary radiotherapy remains critical (1) to assess the efficacy of radiotherapy after tapering of pharmacological treatments, assuming that after 2 to 3 years the probability of control of hypercortisolism with radiotherapy is low; (2) to detect and treat hypopituitarism; and (3) to ensure the absence of recurrence in cured patients, especially after radiosurgery.

Bilateral Adrenalectomy. BLA is currently performed using a laparoscopic approach. When performed in expert centers, the mortality rate is zero, and the morbidity is low.[172,173] The major advantage of BLA is its ability to provide immediate resolution of hypercortisolism and major improvement in clinical symptoms and quality of life.[173] Furthermore, its efficacy is permanent in almost all patients, and persistence of mild cortisol secretion owing to adrenal remnants rarely leads to recurrence of hypercortisolism.[172] No accurate study comparing clinical improvement following BLA versus radiotherapy and/or pharmacological strategies is available. The major drawback of BLA is the need for lifelong steroid replacement, with a risk of adrenal crisis that is equivalent to that observed in adrenal insufficiency resulting from other causes.[173] Patient education is therefore of primary importance after BLA.

The choice of BLA versus other second-line therapies must be debated in each individual case, but it is the treatment of choice in young women who are not cured following pituitary surgery and who wish to become pregnant, given the teratogenicity of available drugs. Outside of this situation, BLA is commonly proposed as the last option after numerous therapeutic attempts and many years after the diagnosis of Cushing disease.[173] This common strategy may favor persistent comorbidities because of the duration of exposure to cortisol excess.

The risk of typical Nelson syndrome, which includes the development of a pituitary macroadenoma, has been dramatically reduced with the availability of pituitary MRI.[174] The term "corticotroph tumor progression" is actually more appropriate, implying radiographic evidence of adenoma growth or identification of a new adenoma when previous pituitary imaging was negative. Using this definition, corticotroph tumor progression occurs in approximately 10% to 40% of patients[172,174] after a mean/median period of 2 to 3 years following BLA.[174] In one large study, corticotroph tumor progression occurred as early as 1 year after BLA, whereas its prevalence reached a plateau after 7 years.[175] Corticotroph tumor progression is not linked to *USP8* mutations[21] and has also been observed in patients treated with steroidogenesis inhibitors or glucocorticoid receptor antagonists.[146,147,176] Transformation into metastatic pituitary carcinoma is described rarely.[177]

All of the predictive factors identified for the development of corticotroph tumor progression, such as young age, presence of pituitary adenoma, and increased Ki-67 index, have given rise to contradictory publications.[174] The most consistent predictive factor is a marked increase in ACTH concentrations in the first year after BLA, although various thresholds have been proposed.[174,175] However, less than 50% of patients with these criteria will show tumor growth on MRI during a 10-year follow-up period.[177] Because BLA may be indicated in young women who are pregnant, the finding that pregnancy does not accelerate corticotroph tumor progression is important.[174,175] The protective effect of adjuvant radiotherapy for corticotroph tumor progression has been a matter of intense debate, but has not been confirmed.[174,175] In our experience, it should be considered before BLA in those patients with large and aggressive tumors.

There is no consensus on the monitoring of patients following BLA, which may include plasma ACTH measurement and, most importantly, pituitary MRI 6 to 12 months after BLA and then yearly. After

TABLE 7.5 Main Characteristics of Current Pharmacological Treatments for Cushing Disease

Target	Drugs	Usual dose range	Effectiveness	Time of action	Main adverse effects	Evidence/Comments
Pituitary-directed drugs	Pasireotide LAR	10–30 mg IM/4 weeks	26%–40%	Within 1 month	Frequent hyperglycemia, diarrhea, gallstones	Prospective randomized studies Works better in mild/moderate hypercortisolism Possible adenoma shrinkage
	Cabergoline	2.5–5 mg/week	28%–35%	Poorly known ≥1 month?	Asthenia, nausea, hypotension, psychiatric symptoms	Retrospective studies and limited cohorts Low-cost treatment Well tolerated
Adrenal-directed drugs	Ketoconazole	200–600 mg PO bid-tid	≥50%	2–3 days	Cytolytic hepatitis (severe <5%), gastrointestinal disorders, rash, male hypogonadism and gynecomastia	Retrospective studies Numerous drug interactions Need gastric acidity for absorption Reduced number of pills compared to metyrapone Preferred in woman over metyrapone
	Metyrapone	250–1000 mg PO qid	≥50%	Few hours	Nausea, dyspepsia, hirsutism/acne, hypokalemia, hypertension and edema at high dosage	Retrospective studies Very rapid effect Multiple pills in 3–4 daily intakes Issue in cortisol measurements for monitoring drug dosage Preferred in men over ketoconazole
	Mitotane	0.5–3.0 g PO tid	≥75%	≥3 months	Gastrointestinal disorders, asthenia/dizziness, neurological/cognitive disorders, hypercholesterolemia, biological cholestasis, hypogonadism, central hypothyroidism	Retrospective studies with few patients Frequent intolerance Almost constant adrenal insufficiency Long-term (years) avoidance of pregnancy after withdrawal
	Osilodrosat	2–30 mg PO bid	66%–81%	Days	Nausea, hypokalemia, hypertension, and edema at high dosage, hirsutism? Adrenal insufficiency	Prospective randomized studies/placebo control Issues in cortisol measurement for monitoring Reduced number of pills compared with metyrapone Frequent adrenal insufficiency during the titration period owing to the potency of the drug
	Etomidate	Various protocols ≤0.3 mg/kg/h	≤100%	Hours	Sedation, nephrotoxicity, myoclonus	Short-term treatment emergency treatment/oral therapy impossible Use in intensive care unit preferable
Glucocorticoid receptor–directed drug	Mifepristone	300–1200 mg PO bid	38%–60%	Days	Nausea, asthenia, metrorragia, endometrial hyperplasia, dizziness, hypokalemia	Open prospective study Efficiency mainly based on improved glycemia/hypertension Biochemical monitoring with cortisol impossible/Challenge of measuring disease activity Expensive (United States)

LAR, Long-acting release; *PO*, per os (by mouth).

5 to 7 years, the imaging intervals can be extended.[175] As the adenoma growth rate is usually slow, it is important to compare MRI scans with the first baseline examination.

Management of corticotroph tumor progression is variable and includes imaging surveillance for small tumors without mass effects. Surgery and radiotherapy alone and in combination are the most effective treatments, allowing tumor control rates of approximately 80%.[177] Medical treatments targeting the pituitary adenoma, such as dopamine and somatostatin agonists, have occasionally given good results, but only in isolated case reports.[174] Temozolomide treatment may be advisable in very aggressive tumors.[138,174]

Pharmacological Treatments

General Considerations. Pharmacological treatment is indicated mainly as first-line treatment in inoperable patients or as salvage treatment in rare patients with life-threatening intense hypercortisolism and as adjuvant treatment in patients with surgical failure or recurrence. In this last situation, medical treatment is usually a “bridging” therapy used to control hypercortisolism until a therapeutic option yields definitive cessation of hypercortisolism (repeat surgery, pituitary radiotherapy, BLA). Currently available drugs act through three modes of action: inhibition of cortisol synthesis, modulation of ACTH release by the corticotroph adenoma, and antagonism of cortisol at the level of its receptor.

Several considerations need to be taken into account regarding medical treatment. No drug has 100% potency, but all may induce adrenal insufficiency, and all have various side-effects (Table 7.5). Careful monitoring, including measurement of 8:00 AM plasma cortisol, is therefore necessary to anticipate and prevent acute adrenal failure. Spontaneous day-to-day fluctuations in cortisol production are also frequent in Cushing disease. Consequently, consistent large consistent changes in UFC are required to show the effectiveness of cortisol-lowering drugs and, conversely, to diagnose a true escape phenomenon.[73] All medications, with the exception of cabergoline, are contraindicated or not recommended in pregnant women.[138] Two main treatment strategies are currently employed, with no comparative data indicating the superiority of one strategy over the other. With the exception of glucocorticoid receptor antagonists, the "titration strategy" is the most commonly used and adapts the drug dosage to normalize biochemical markers such as UFC or serial daily measurements of plasma cortisol. The "block and replace" strategy is less frequently used and combines medical treatments at a dosage that induces the lowest cortisol concentrations achievable in association with glucocorticoid replacement therapy.[178] This strategy, which reduces the need for biochemical monitoring once the "blockade" is obtained, has been recommended as first-line therapy in situations of hospital overload, such as during active phases of the COVID-19 pandemic.[178] It should also be kept in mind that normalization of UFC does not imply restoration of the circadian rhythm of cortisol, as demonstrated by persistently elevated late-night salivary cortisol.[179,180] Recently published studies suggest that symptoms of hypercortisolism, such as hypertension, are better controlled in patients with both normalized UFC and midnight cortisol than in those patients with normalized UFC only.[180] Lastly, growth of pituitary adenoma remnants can be observed during all medical treatments and especially with those that do not target the pituitary adenoma.[142,146,147,176,181] Follow-up should therefore include pituitary MRI, bearing in mind that growth of remnants is usually a very slow process. Some authors take advantage of this situation and delay repeat pituitary surgery until a convincing MRI image is found in patients treated with steroidogenesis inhibitors following unsuccessful first surgery.[147]

Pituitary-Directed Drugs. Although inhibition of adenomatous ACTH secretion is ideal from a pathophysiological standpoint, its effectiveness remains only moderate because it depends on the presence and functionality of target receptors in the adenoma. To date, there are no predictive factors of their effectiveness, so a "trial-and-error" approach is required. Pituitary-directed drugs can induce tumor shrinkage and are an option when there are concerns about the oncogenic characteristics of the corticotroph adenoma.

Somatostatin Analogs

Pasireotide is a multireceptor-targeted somatostatin analog with a high binding affinity for the SSTR5 subtype that is expressed in approximately 30% of corticotroph adenomas.[142] Administered subcutaneously, twice-daily, in a large double-blind randomized trial, pasireotide was able to normalize UFC after 3 months in 26% of patients and improve both their clinical symptoms and quality of life.[182] Normalization of UFC was achieved approximately half of the patients with mild hypercortisolism (UFC <2 times the upper limit of normal [ULN]) in a second large trial using a long-acting version of pasireotide administered monthly via intramuscular injection.[183] A decrease in adenoma volume (≥20%) was observed in roughly half of the patients with evaluable tumors, whereas it increased in 10% of patients during the 12-month study.[183]

The main adverse effect associated with pasireotide is the occurrence of hyperglycemia in about 75% of patients, which appears within 2 weeks of starting treatment.[182,183] Its intensity is highest in older and already glucose-intolerant patients and diabetic patients. This side effect is independent of the control of hypercortisolism and is secondary to the inhibition of secretion of insulin and incretins.[142] Anticipation of this side effect with an adapted diet and careful monitoring of hyperglycemia are therefore mandatory. First-line treatments for pasireotide-induced hyperglycemia include metformin, glucagon like peptide 1 agonists, or dipeptidyl peptidase 4 antagonists.[142]

Dopamine Agonists

Cabergoline is a dopamine agonist that targets the type 2 dopamine receptor expressed in approximately 60% of corticotroph adenomas.[142] In the largest multicenter retrospective survey performed to date, normalization of UFC was observed in 40% of 53 patients treated with cabergoline as monotherapy in the short term and in 28% of patients in the long term.[184] The relatively moderate dosage median dosage needed to control UFC (2.5 mg/week), together with reports of corticotroph insufficiency during cabergoline treatment, suggests that some corticotroph adenomas may be exquisitely sensitive to the drug. However, up to one-third of patients who are initially responders can experience definitive escape from the treatment.[142] The delay in the action of cabergoline is not well known, but results from a 6-week prospective study suggest that its cortisol-lowering effect may be a slow process.[185] In some cases of corticotroph macroadenomas an antitumor effect has also been noted.[186] Adverse events may occur, including nausea, vomiting, and orthostatic dizziness. Rare impulsivity disorders have been recognized in patients treated for hyperprolactinemia and should be sought during questioning. Lastly, most studies of cabergoline therapy for hyperprolactinemia have provided reassurance with regard to the occurrence of carcinoid valvulopathy.[186] Overall, there appear to be several benefits of cabergoline, including its low cost, oral route of administration, and good tolerance.

Adrenal-Directed Drugs

Ketoconazole. Ketoconazole inhibits the enzymes CYP17A1, CYP11A1, and CYP11B1.[187] It is used orally at daily doses ranging from 400 to 1200 mg taken in two divided doses.[187] For optimal absorption, the presence of gastric acidity is necessary, and thus concomitant use of proton pump inhibitors may reduce its effectiveness. Small studies have suggested that the nadir of cortisol is obtained after 2 to 3 days of treatment.[188] In the largest retrospective study of ketoconazole in the treatment of Cushing disease,[189] normalization of UFC associated with clinical improvement was obtained in 49% of 200 patients, at an average dose of 700 mg/day of ketoconazole. Baseline UFC was not a predictor of efficacy, but this low dosage suggests frequent undertitration and that the efficacy of ketoconazole may have been underestimated. Indeed, a higher rate of control of hypercortisolism of various causes was observed with adequate and careful titration in single-center series.[142] Escape from treatment occurs in less than 15% of cases.[142,189] A transient and mild increase in liver enzymes is seen in approximately 15% of patients and usually occurs during the first 2 months of treatment or after increasing the drug dosage.[189,190] Major cytolytic hepatitis occurs in only 2.5% to 5.5% of patients. This idiosyncratic event is unrelated to the drug dosage and is reversible with withdrawal of the drug.[189,190] Frequent measurement of hepatic enzymes during the first weeks of treatment is mandatory. Other common adverse effects include gastrointestinal intolerance, male hypogonadism, and gynecomastia, secondary to the inhibition of testicular steroidogenesis that occurs at lower doses than those required to inhibit cortisol production.[187] Although the exact prevalence of ketoconazole-induced male hypogonadism is unknown, this adverse effect suggests that the drug should be preferentially used in women. Importantly, ketoconazole is a potent inhibitor of CYP3A4 and thus may modify the bioavailability of numerous drugs that are metabolized by this enzyme.[187] A detailed analysis of the patient drug profile is thus mandatory before commencing treatment with ketoconazole, as well as close monitoring and dose adjustments of the associated drugs.

Metyrapone. Metyrapone, which mainly inhibits the CYP11B1 and CYP11B2 enzymes,[187] is notable for its rapid onset of action, with a nadir of cortisol levels obtained approximately 2 hours following a single dose.[191] Because of the short half-life of the drug, it should be administered in three to four daily divided doses.[192] A large retrospective UK survey reported control of hypercortisolism in approximately 50% of patients.[192] The average dose of metyrapone was only 1350 mg/day, indicating probable undertitration and underestimation of treatment efficacy. Nausea and abdominal discomfort are the most frequent side effects of the drug. Metyrapone causes the accumulation of steroids upstream from the 11β-hydroxylase blockade, including mineralocorticoids that can be responsible for edema, hypokalemia, and hypertension at high dosages of the drug, and of adrenal androgens, which are responsible for hirsutism and acne in women.[142] The dramatic increase in the immediate cortisol precursor 11-deoxycortisol, which is structurally similar to cortisol, interferes with many commercial cortisol immunoassays, leading to an overestimation of cortisol levels.[187,191] It is therefore essential to monitor the cortisol response using more specific laboratory methods, such as LC-MS/MS.

Mitotane. Mitotane is an adrenolytic drug that inhibits multiple enzymatic steps of cortisol synthesis, including 11β-hydroxylase and P450 side-chain cleavage. In addition, mitotane at relatively high dosage induces cellular necrosis, leading to “chemical adrenalectomy.” A high rate of control of hypercortisolism and improvement in clinical symptoms is therefore obtained.[142] However, the anticortisolic effect is delayed for a median of 6 months, meaning that mitotane used in monotherapy is unsuited to situations requiring rapid control of hypercortisolism.[146] Mitotane has a long half-life owing to its accumulation in adipose tissue, but adrenal insufficiency may be reversible several years after cessation of treatment.[146] Because mitotane is a potent enzyme inducer (CYP3A4), the required doses of hydrocortisone supplementation may need to be increased by approximately one-third in some patients.[142]

A lower dosage of mitotane is required for treatment of Cushing disease than for treatment of adrenocortical carcinoma, and measurement of circulating concentrations of mitotane may help to target the titration of the drug.[146,187] In the largest single-center series, control of hypercortisolism was achieved in 72% of patients with a mean dose of 2.6 g/day.[146]

Mitotane leads to an increase in levels of CBG and therefore an increase in serum cortisol concentrations. Cortisol monitoring should thus be performed by measurement of UFC or LNSC. The main limitation of mitotane is the frequency of adverse effects, including gastrointestinal disorders, asthenia, neurological and cognitive disturbances, increased liver enzymes, hypercholesterolemia, thyrotroph insufficiency, male hypogonadism, and ovarian cysts.[142,146,187,193] Up to 30% of patients with Cushing disease may discontinue treatment because of drug intolerance.[142,146] Lastly, mitotane is a potent teratogen, and pregnancy should be avoided before clearance of circulating drug.

Osilodrostat. Osilodrostat is a more potent inhibitor of CYP11B2 and CYP11B1 than metyrapone, with a prolonged half-life compared to metyrapone allowing two doses per day. In a 22-week, open-label, prospective study (LINC-1), osilodrostat normalized UFC in 79% of a small cohort of patients with Cushing disease.[179] Interestingly, the dosage of the drug could be decreased in some individuals in the long term, during the extension phase of the study, while maintaining adequate control of hypercortisolism. In a recently published large-scale phase 3 study (LINC-3), at least 66% of patients with Cushing disease normalized UFC for at least 6 months of treatment, and approximately 10% more patients had a decrease in UFC greater than 50%.[181] Clinical symptoms and quality of life improved in parallel with the reductions in UFC. The most common adverse events were generally mild and included nausea, abdominal discomfort, and headaches. Side effects related to the production of steroids upstream of the pharmacological blockade, such as hypokalemia, hypertension, and hirsutism in women, were noted in 10% of patients. Importantly, because of the potency of the drug, hypocortisolism occurred at variable intensity in half of patients in the short and mid-term.[181] Adrenal insufficiency should therefore be anticipated during osilodrostat treatment by regular 8:00 AM plasma cortisol measurements, appropriate patient education, and prescription of hydrocortisone supplementation to be taken if needed.

Levoketoconazole. Although ketoconazole comprises a racemic mix of two enantiomers (an L-form and a D-form), levoketoconazole is a pure L-enantiomer that may possess higher potency as a steroidogenesis inhibitor than ketoconazole, allowing a lower drug dosage to be used. In addition, levoketoconazole has reduced liver exposure and interferes less with bile-acid synthesis than ketoconazole, hypothetically resulting in a minimal rate of hepatic dysfunction.

In a recently published multicenter phase III study in patients mostly suffering from Cushing disease,[194] 31% of the 94 patients had normalized UFC at the end of a 6-month maintenance phase, and the response rate was 49% in the 39 patients who had an initial UFC greater than two times and less than five times the ULN. Interestingly, testosterone concentrations were not reduced in men. Transaminases reversibly increased by more than three times the ULN in 11% of patients and more than five times the ULN in 3% of patients. These results are quite similar to those obtained with racemic ketoconazole, and in the absence of a head-to-head trial it is difficult to know whether levoketoconazole is superior in terms of efficacy and safety. Future studies may provide more data on the benefits of this new drug for Cushing disease.

Etomidate. Etomidate is an anesthetic drug administered intravenously in a polyethylene glycol solution that strongly inhibits 11β-hydroxylase, leading to a rapid decrease in cortisol production. Etomidate is usually used for several days while waiting for another intervention in patients with life-threatening hypercortisolism who are unable or unwilling to tolerate oral therapy, such as patients with acute steroid psychosis. Several protocols have been published using etomidate at variable dosages and using a titration procedure (implying four daily serum cortisol measurement) or a “block and replace” regimen.[195,196] Relatively high doses (0.3 mg/kg/h) usually decrease cortisol levels within hours but induce some degree of sedation and are preferably performed in an intensive care unit (ICU) setting. An alternative and convenient protocol, usable outside the ICU and involving reduced dosage of etomidate at 0.020 to 0.030 mg/kg/hour with daily increments of 0.010 mg/kg/hour until a morning cortisol value of 500 nmol/L is achieved, was recently published, and was able to control hypercortisolism within 3 days.[197] Similar to the case of metyrapone and osilodrostat, which inhibit 11-β-hydroxylase, cortisol immunoassays that have some degree of crossreactivity with 11-deoxycortisol may overestimate true cortisol concentrations. Other possible adverse effects include nausea, vomiting, kidney dysfunction, and myoclonus.

Glucocorticoid Receptor–Directed Drugs

Mifepristone is a competitive glucocorticoid receptor antagonist that inhibits cortisol action by binding to its cognate receptor. Mifepristone is not a selective antagonist and also binds to the progesterone receptor. The SEISMIC study evaluated the effect of mifepristone monotherapy in 50 patients treated for at least 30 days and reported an improvement in glucose tolerance and hypertension in 60% and 38% of patients, respectively.[176] In a long-term safety extension study, a beneficial effect on weight loss and other clinical features of Cushing syndrome persisted in patients who remained on therapy for 2 years.[198]

Importantly, by blocking the glucocorticoid receptor, mifepristone increases the secretion of ACTH and consequently increases cortisol concentrations, despite reduced access of cortisol to target tissues. Consequently, evaluation of the efficacy and titration of the drug is complex and is exclusively based on clinical symptoms and intermediate biochemical markers such as blood glucose, whereas serum and urinary measurements of cortisol are useless. This exclusively clinical approach to titrating mifepristone may lead to persistence of some degree of hypercortisolism and, conversely, difficulty in diagnosing hypoadrenalism because cortisol concentrations are not low in these patients. When hypoadrenalism is suspected, mifepristone treatment should be temporarily interrupted, and dexamethasone should be given in higher than normal replacement doses. Adverse effects of mifepristone include metrorragia and endometrial thickening in 20% of patients, which are secondary to antagonism of the progesterone receptor.[176] The increase in cortisol concentration may also exceed the capacity of the renal 11-βHSD enzyme to convert cortisol to metabolically inactive cortisone, and cortisol excess can stimulate the mineralocorticoid receptor, leading to hypokalemia, edema, and hypertension.

Combination of Medical Treatments

Medical treatments can be used in combination to increase their potency and/or reduce their dosage, as well as to limit the adverse effects of a given drug. Various combinations have been described, such as ketoconazole and metyrapone, that act on different target enzymes of adrenal steroidogenesis,[142,187,199,200] or steroidogenesis inhibitors combined with centrally-acting drugs such as pasireotide and cabergoline.[142,184]

Future Medical Treatments

Medical therapy for Cushing disease is an active area of research, and many novel and promising therapeutic targets for the treatment of patients with Cushing disease are currently being investigated, such as drugs targeting the tumoral corticotrophs, the MC2R ACTH- and glucocorticoid-receptor.[201] This research will be enriched by ongoing molecular studies of corticotroph tumors using modern pangenomic approaches (see Etiology).

PROGNOSIS FOR CUSHING DISEASE: PERSISTENT MORBIDITY AND MORTALITY

Remission or cure of Cushing disease is followed by clinical improvement. However, comorbidities may not be completely reversible, despite prolonged biochemical remission. Evaluation and treatment of the long-term negative effects of chronic hypercortisolism are therefore an important part of treatment to reduce morbidity, improve patient quality of life, and, hypothetically, reduce the long-term excess mortality associated with Cushing disease.

Metabolic Syndrome and Cardiovascular Risk

A decrease in body weight and fat mass is usually seen 6 to 12 months after successful treatment of Cushing disease. However, central obesity and increased total and visceral fat remain increased several years after remission,[47,54,202] and a number of patients maintain an unfavorable metabolic profile with an increased prevalence of diabetes and dyslipidemia.[29,33] The persistence of altered body fat distribution plays a key role in the maintenance of metabolic syndrome and is associated with a persistent state of low-grade inflammation.[202] GH supplementation may increase the incidence of diabetes and cardiovascular disease and should be carefully monitored.[154]

An improvement in blood pressure occurs in a majority of patients, but hypertension may persist in at least 30% of patients.[38] Persistence of hypertension has been attributed, at least in part, to microvascular remodeling and persistent abdominal obesity with insulin resistance. The prevalence of atherosclerotic plaques in the coronary and carotid arteries is increased even several years after cure as compared with control populations.[29,47,48]

Several studies suggests that the duration of hypercortisolism is a predictor of the persistence of cardiovascular risk factors and hypertension.[47,203] Importantly, the increased morbidity owing to thromboembolism, stroke, and myocardial infarction was found to be sustained, especially during the first year after cure.[43,44] Similarly, the procoagulative phenotype was not fully recovered 1 year postcure.[42]

Muscle

The natural history of muscle function following cure of Cushing disease is unclear. On MRI, patients may still show reduced limb skeletal muscle mass in the midterm and altered muscle function and muscle quality resulting in decreased muscle strength in the long term.[28,142,204]

Bone Disease

Most studies suggest a relief from glucocorticoid-induced suppression of bone formation following resolution of hypercortisolism with an early increase in plasma osteocalcin, rapid improvement of bone microarchitecture, and major improvements of bone mineral density, both in adults and children/adolescents.[29-31] However, 4 to 5 years may be needed to achieve full recovery, and a subset of patients in long-term remission show persistently lower bone mineral density than controls.[30] This may more specifically concern patients who experienced longstanding hypercortisolism, postmenopausal women, patients with associated hypopituitarism, or those who were treated with an excessive dose of hydrocortisone replacement.[54,202] Acquired spinal damage, including vertebral collapse and geometric abnormalities, persists after remission. A significant decrease in the fracture rate following remission of endogenous hypercortisolism has been described, but the data are scarce.[30] Of note, patients with previous fractures are predisposed to develop new fractures, despite remission, confirming that a history of fractures is a hallmark of skeletal fragility. Overall, considering the uncertainties of the natural history of bone disease following remission, an individualized approach is recommended. This includes detailed fracture assessment with lateral imaging of the spine, evaluation of bone mineral density and trabecular bone score, adequate calcium and vitamin D intake, avoidance of postsurgical excessive glucocorticoid supplementation, and personalized follow-up based on the results of the bone evaluation.

Mood and Cognitive Function

Significant improvement of psychopathology occurs rapidly following successful surgical treatment of Cushing disease or after reduction of hypercortisolism by pharmacological means.[34] However, patients with a previous history of Cushing disease show an increased prevalence in psychopathology symptoms, maladaptive personality, and mental exhaustion associated with a prolonged recovery time after mentally strenuous tasks. Long-term residual cognitive impairment, independent of coexisting chronic fatigue, affective disorders, or associated pituitary deficits, has been reported in several studies[36,205] and is more pronounced than in patients treated for nonfunctioning pituitary adenomas. Although several brain imaging studies have shown improvement in atrophy of key brain areas following remission, persistent structural and functional abnormalities have been described, such as decreased hippocampal volume at MRI and reduced functional responses in the prefrontal cortex in patients with severe verbal and visual memory impairment.[37] Overall, chronic hypercortisolism may have irreversible effects on the central nervous system. An increased rate of suicide has been reported in patients cured of Cushing disease.[206]

Autoimmunity

Cortisol excess inhibits the immune system. Following remission, a rebound in autoimmunity may be observed. Preexisting autoimmune diseases may become active again, and new onset of autoimmune diseases may also occur, with autoimmune thyroiditis being the most frequent endocrine disorder observed.[142]

Quality of Life

Health-related quality of life (QoL) analyzed with nonspecific and specific questionnaires[207] improves after remission. However, large cross-sectional studies report long-term residual impairment in physical and social functioning, as well as limitations in the ability to perform usual activities of daily life, owing to physical and emotional problems, pain, fatigue, sleep problems, and reduced vitality.[36,208] Postoperative hypopituitarism may also be detrimental to the QoL. Together with the residual psychological and cognitive impairments, decreased perceived health-related QoL has significant social and economic consequences, such as an inability to return to work.[208] Interestingly, patients with a history of Cushing disease report more negative illness perceptions compared with patients with other chronic diseases.[209]

Mortality in Cushing Disease

In the original publication by Harvey Cushing in 1932, the median survival in untreated patients with Cushing disease was less than 5 years, with the main causes of mortality being cardiovascular complications, including stroke and myocardial infarction. Multiple retrospective, nationwide, single and multicenter cohort studies have characterized the standardized mortality ratios (SMRs) in Cushing disease compared with both the general population and to patients with other types of pituitary adenomas or causes of Cushing syndrome. Persistent or recurrent Cushing disease after pituitary surgery confers a high risk of mortality that is mainly attributed to cardiovascular disease and infection, as compared to patients in remission. In a metaanalysis of eight studies,[45] the pooled SMR for patients with persistent or recurrent disease was 4.6 (95% confidence interval [Cl]: 2.9–7.3), compared with an SMR of 2.5 (95% Cl: 1.4–4.2) for patients in remission. Although remission is associated with a reduction in SMR, several studies report that it remains increased.[45,206,210,211] Furthermore, patients treated for Cushing disease have increased mortality compared with patients treated for other types of pituitary adenomas or patients with Cushing syndrome owing to adrenal causes.[211] In a multicenter study[210] analyzing long-term survival of patients in remission for at least 10 years, those who achieved remission after pituitary surgery alone had long-term survival comparable to that of the general population (0.9; 95% Cl: 0.6–1.5), whereas patients needing multiple treatments, probably reflecting disease that was more difficult to cure and prolonged exposure to cortisol excess, had a persistently increased SMR (2.6; 95% Cl: 1.9–3.6).[210] In support of this, mortality data from 346 patients over 31 years of follow-up identified that the length of exposure to hypercortisolism was associated with increased risk of death.[203] Other factors associated with increased mortality include older age at diagnosis, male gender, depression, and persisting cardiovascular risk factors such as diabetes and hypertension, and are independent predictors of mortality in most studies.[45,203,210]

Implications for Treatment and Long-Term Care

Taken together, these data highlight the importance of limiting the time of exposure to cortisol excess, emphasizing the importance of early diagnosis, choosing aggressive intervention, and attempting to perform, whenever possible, pituitary surgery as a first-line treatment. In patients with persistent disease or recurrence, hypercortisolism should be controlled as effectively and as soon as possible. Aside from rapid and effective treatment of hypercortisolism, the treatment of persistent comorbidities after cure represents an important therapeutic aspect, and aggressive treatment of hypertension, dyslipidemia, and other cardiovascular risk factors is warranted.[138] In patients with residual adrenal insufficiency, adequate adjustment of the hydrocortisone dose is important to avoid prolonging hypercortisolism, reduce the risk of infections, and allow healing of comorbidities. In addition, treatment of physical comorbidities may contribute to improvement in perceived quality of life. However, there is no effective therapeutic strategy for cognitive deficits except for psychological support and interactions with patients through membership in dedicated associations.

SPECIAL CONDITIONS

Cushing Disease as an Emergency Condition

Undiagnosed longstanding Cushing syndrome or intense hypercortisolism, a condition that is mostly seen in patients with EAS but also in a subset of patients with Cushing disease, is a life-threatening condition in the short term. In such instances the catabolic and metabolic symptoms are exacerbated, including bedridden status owing to profound myopathy and osteoporotic vertebral fractures, severe hypertension and acute heart failure, profound hypokalemia and major hyperglycemia, severe infections caused by bacterial, fungal, and opportunistic pathogens, pulmonary embolism, peritonitis as a result of gut perforation, and untreatable psychosis.[35]

This emergency scenario requires several tasks to be performed simultaneously in the setting of intensive care in collaboration with experienced endocrinologists. Firstly, rapid and systematic evaluation of comorbidities must be performed, and preventive and curative treatments of these comorbidities must be started simultaneously.[35] Management should combine specific treatments of morbid conditions, anticoagulation, and prophylaxis of *Pneumocystis jirovecii* pneumonia with sulfamethoxazole-trimethoprim.[138] At the same time, the diagnosis of Cushing disease must be conducted, but the time devoted to biochemical etiological investigations should be dramatically shortened so as not to delay initiation of specific treatments for hypercortisolism. Biochemical investigations can even be reduced to serial blood samples only, taken over several hours, for cortisol and ACTH measurements to confirm ACTH-dependent Cushing syndrome. Performing time-consuming dynamic tests and BIPSS may be incompatible with both the urgency to treat and the general condition of the patient. Because EAS is the most prevalent etiology in this scenario, we suggest rapidly performing a neck-to-pelvis thin-slice CT, the main priority being to ascertain the absence of an obvious and aggressive ectopic tumor that may require urgent surgical excision and/or antineoplastic medical treatment.[35]

BLA can be performed as a life-saving therapy in critically ill patients. Contemporary studies have reported no deaths, but the rate of perioperative complications may be increased in this situation.[212] Medical treatment using fast-acting steroidogenesis inhibitors, such as ketoconazole and metyrapone, is a valuable alternative, and, given in association at high dosage, ketoconazole and metyrapone have been shown to rapidly and dramatically reduce extremely intense hypercortisolism, giving normalization of UFC within 1 week in 50% of patients.[199] Alternatively, etomidate may be used to rapidly control hypercortisolism in the short term while waiting to perform pituitary surgery or implement other therapeutic alternatives.

Cyclic Cushing Syndrome

A small subset of patients with Cushing disease show alternating periods of normal and increased cortisol secretion. Several terms have been used for this condition, such as “periodic,” “intermittent,” and “cyclical” Cushing syndrome. There is no universal definition for this condition, but demonstration of three peaks and two troughs of cortisol production is usually accepted, with the term “cyclical” being more specifically used when alternating of episodes occurs at a steady rhythm.[213,214] The duration of phases is extremely variable and ranges from days to months. The pathophysiology of this condition, which may occur in all causes of Cushing syndrome, remains unknown.

From a clinical perspective, the clinical course of patients may be unchanged, usually with mild Cushingoïd symptoms, or it may parallel the cortisol production, with episodes of exacerbation and spontaneous improvement or resolution of clinical features such as hypertension, edema, hyperglycemia, and psychiatric symptoms.

The main drawback of this condition is that it may be responsible for aberrant results on standard diagnostic tests, such as normal UFC and normal response to dexamethasone suppression, and thus for an incorrect diagnosis of an absence of Cushing syndrome.[213,214] Biochemical investigations used for the biochemical etiology, including BIPSS, can also lead to erroneous results and should be performed only during a sustained period of hypercortisolism.

Intermittent Cushing disease should be suspected if patients report spontaneous and huge variations in the intensity of clinical symptoms or when clinical symptoms contrast with normal biochemical evaluations. In the latter instance, the possibility of factitious Cushing syndrome because of surreptitious intake of exogenous glucocorticoids should also be ruled out. There is no ideal investigation to diagnose intermittent Cushing disease aside from regular clinical follow-up of patients and repetition of biochemical investigations to identify phases of cortisol hypersecretion.[65] Because of the convenience of sampling, measurement of salivary cortisol is an accepted tool for this purpose.[65] More recently, the measurement of cortisol in hair samples may be useful in the retrospective identification of intermittent oversecretion.[90]

Aggressive Pituitary Corticotroph Adenomas and Corticotroph Carcinomas

Aggressive pituitary tumors correspond to radiologically-invasive tumor with unusually rapid tumor growth rate, or clinically relevant tumor growth despite optimal standard therapies (surgery, radiotherapy, and conventional medical treatments).[9] Aside from local aggressiveness, some tumors metastasize distantly, either hematogenously or through the spinal fluid. Predictors of aggressiveness include the combination of radiological invasion and the presence of proliferative markers (Ki-67 index ≥3% and mitotic count >2) and immunostaining for p53 at histology, plus young age, as well a tumor subtypes such as silent corticotroph adenomas, Crooke cell adenomas, and Nelson syndrome tumors.[7,9-11] Germline genetic testing should be performed in young patients. Treatment of aggressive corticotroph tumors should be discussed by a multidisciplinary expert team and include repeat surgery, pituitary radiotherapy, and medical treatments. Among these, pasireotide and cabergoline may occasionally show transient efficacy.[9,142] The only chemotherapy with established activity is the alkylating agent temozolomide given at a dosage of 150 to 200 mg/m^2 for 5 consecutive days every 28 days and associated with careful monitoring of hematological parameters and liver function tests. Objective radiological responses should be assessed after three cycles and are observed in no more than 50% of patients. High methylguanine-DNA-methyltransferase expression in tumor tissues may predict lack of response, but there are exceptions. In patients with rapid tumor progression on temozolomide treatment, systemic cytotoxic therapy should be administered, although no specific drug can be recommended at the time of writing[9] (see also Chapter 16).

Pediatric Cushing Disease

Cushing disease rarely occurs in children, with approximately 10% of new cases each year occurring in children. Cushing disease is the most common cause of Cushing syndrome and accounts for between 75% and 80% of cases in children and adolescents. However, before the age of 6 years, adrenal diseases are the most common causes. Cushing disease does not show a female predominance in prepubertal children.[215] Pituitary microadenomas are the most common cause of Cushing disease in children. The vast majority of patients with pediatric Cushing disease do not have causative germline genetic defects. Although somatic mutations in *USP8* occur in pediatric corticotroph adenomas,[216] germline mutations of the *AIP, MEN1, CDKN1B, CABLES1,* and *DICER1* genes, responsible for isolated or syndromic Cushing disease, only account for a small percent of cases.[16,17] However, physical characteristics (including birthmarks) and family history should be closely examined to consider whether genetic testing is appropriate.

Disease onset is often insidious but also has some specific clinical differences from adult disease. The most common presentation is a decreased growth rate leading to growth retardation contrasting with weight gain.[215,217] This auxological contrast distinguishes Cushing disease from subjects with simple obesity because most simply obese children are tall. Importantly, and in contrast with the decreased growth velocity, children may have advanced skeletal maturation owing to the secretion of adrenal androgens.[218] Excessive exposure to adrenal androgens combined with gonadotrophin deficiency are also responsible for abnormal virilization, with advanced pubic hair and genital growth in boys associated with prepubertal testicular volumes, or with pubic hair growth in girls showing prepubertal breast development. Easy bruising and muscle weakness may be less common in pediatric patients.[215,217] Psychiatric and cognitive changes can also affect school performance in children.[215,217,219]

The diagnostic algorithm and use of biochemical tools to diagnose Cushing disease is similar to those for adults, but with an emphasis in children on measuring LNSC, taking advantage of its convenience.[215] In children, UFC should be corrected for body surface area (× 1.72 m^2), and in children weighing less than 40 kg the dexamethasone dose is adjusted to 30 μg/kg/day for the 2-day LDDST.[217] Pituitary MRI is as effective in children as it is in adults and, regardless of the methodology used, is positive in only 50% to 60% of cases.[215,220] Given the extreme rarity of EAS in children,[102] the need for BIPSS in ACTH-dependent Cushing syndrome is questionable, especially in children with a clear-cut response to the CRH test.

Transsphenoidal surgery remains the treatment of choice, with roughly similar rates of remission as adults in expert hands.[215]

Recurrences following remission may occur, as in adults, despite postoperative corticotroph insufficiency.[215,221] Conventional radiotherapy used in the setting of failure of pituitary surgery may perform faster than in adults, with reported remission rates as high as 70% within 18 months.[215,219,222] Large pediatric studies looking at the most recent radiosurgery are not currently available; however, proton beam therapy may be considered owing to its ability to limit irradiation of surrounding brain areas. Similarly, until recently there were no medications approved by the US Food and Drug Administration for use in pediatric Cushing disease, leading to off-label prescription of commonly used drugs such as steroidogenesis inhibitors or cabergoline.[142,184] Adrenalectomy is also an option in refractory Cushing disease[219] and, in the opinion of these authors, may ensure better control of hypercortisolism than medical therapy for improving linear growth velocity. Following pituitary surgery, with or without radiotherapy, the incidence of GH deficiency, at least in the short term, is high and prompt evaluation and treatment with human GH is strongly encouraged to ensure growth acceleration and to catch up growth.[215,219] However, despite this, most children do not reach their target height. Pubertal development may also be disturbed owing to the suppressive effect of hypercortisolemia on gonadotropins. Sex steroid replacement should be performed with caution in prepubertal children in order not to compromise final height because of advanced skeletal maturation.[218] Abnormal body composition and fat mass,[223] hypertension,[224] and cardiovascular remodeling may also persist after cure, and management of these is mandatory. Significant psychopathology may persist after remission, and some children may experience a decline in cognitive and school performance 1 year after surgical cure, without associated psychopathology. Residual impairment in quality of life with persistence of forgetfulness, unclear thinking, decreased attention span, and suicide ideation have also been described.[225-227]

Cushing Disease and Pregnancy

Pregnancy rarely occurs in women with Cushing disease owing to dysovulation secondary to cortisol excess. As of 2020, fewer than 300 cases had been published. Causes of Cushing syndrome in pregnancy contrast with those in nonpregnant women. In a review of 122 pregnant women, adrenal causes (mainly adenomas but also carcinomas) represented 62% of cases, and only 35% of women had Cushing disease.[228]

Cushing syndrome during pregnancy is responsible for both maternal and fetal morbidity. The most commonly reported maternal morbidities are hypertension (40%–68%), diabetes or glucose intolerance (25%–37%), and preeclampsia (14%–27%). Maternal death occurred in 2% of cases. The most frequent fetal morbidities were prematurity (43%–66%), intrauterine growth retardation (15%–21%), and spontaneous abortion, intrauterine death, and stillbirth (11%–30%).[228,229]

The diagnosis of Cushing disease during pregnancy is challenging. From a clinical perspective, faciotruncal weight gain, purple striae, diabetes, and hypertension are frequently attributed to the context of pregnancy. The more specific signs of hypercatabolism (bruising, amyotrophia) are rarer. Preeclampsia or intrauterine growth retardation may reveal the disease. From an endocrine perspective, pregnancy is a physiological state of hypercortisolism because of an activation of the HPA axis that is secondary to a progressive increase of CRH and corticotropin secretion by the placenta.[229] The result is a rise in serum, salivary, and urinary cortisol levels that becomes more significant as pregnancy advances. Accordingly, mean serum cortisol levels and 24-hour UFC (measured by LC-MS/MS) in the first, second, and third trimesters of pregnancy were multiplied by 1.7, 2.4, and 3.1, respectively.[230] Nonetheless, a UFC greater than three to four times the ULN should be considered abnormal. In addition, high levels of estrogens increase the synthesis of CBG and cause an overestimation of serum cortisol. The suppression of cortisol following the 1-mg DST is attenuated compared with nonpregnant women, but few data are available concerning the LDDST.[228,229] Serum free cortisol also increases because of pregnancy-induced HPA activation. Cortisol secretion, however, maintains a circadian rhythm despite higher serum levels; measurement of LNSC can therefore be recommended for diagnosis using appropriate thresholds according to the stage of pregnancy.[230,231]

Among other biochemical changes observed during pregnancy, ACTH levels may be less discriminant and may not be suppressed in women with cortisol-secreting adrenal adenomas, possibly owing to pituitary ACTH stimulation by placental CRH or by placental ACTH itself.[229] Consequently, when plasma ACTH concentration is ambiguous with regard to the ACTH-dependency of the disease, it is safe to first perform adrenal imaging by ultrasound to rule out the presence of an adenoma or carcinoma. In ACTH-dependent cases, the performance of dynamic tests (such as the HDDST, CRH, and desmopressin tests) during pregnancy is not precisely known, but ectopic ACTH syndrome is extremely rare (<5% of cases).[228,229] Pituitary MRI without gadolinium injection should be performed, especially if surgery is planned before birth. Bilateral inferior petrosal sinus sampling has been performed in only a few pregnant women with suspected Cushing disease.[229]

Cases of mild and moderate Cushing disease discovered at the end of pregnancy can be managed conservatively by medical treatment of comorbidities. In other cases, pituitary surgery remains the first-line treatment option and should ideally be performed during weeks 12 to 30 of gestation to ensure a minimal rate of maternal and fetal morbidity.[228,229] Radiotherapy is not indicated during pregnancy due both to the delayed results and the potential teratogenicity. Medical therapy, usually initiated during the second or third trimester, is generally a second-line treatment option after surgery. Mitotane is contraindicated during gestation. During pregnancy, metyrapone has been the mostly frequently used medical option.[229] Caution should be exercised with regard to the consequences of the increase in the cortisol precursor 11-deoxycorticosterone, which may worsen hypertension. Rare cases treated successfully with cabergoline at doses of 2.0 to 3.5 mg/week have also been reported, and ketoconazole has been occasionally used with good results, despite its antiandrogenic effect and potential teratogenicity.[229]

Cushing Disease and Chronic Renal Failure

Cushing syndrome has only rarely been described in the setting of chronic renal failure. However, biochemical diagnosis of Cushing syndrome may be difficult for several reasons. Modifications of the serum matrix may be responsible for spuriously increased serum cortisol levels, making it compulsory to use immunoassays using an organic extraction procedure. Decreased glomerular filtration rates result in decreased cortisol excretion, and UFC may be normal despite excessive cortisol production.[68] Chronic renal failure may also activate the HPA axis because of the high inflammatory state and stress of dialysis.[232,233] Consequently, ACTH levels are increased, and, despite the persistence of a normal circadian rhythm of cortisol secretion, LNSC may be increased above normal levels.[232,234] Patients also have abnormal suppression of cortisol after the 1-mg DST.[233,234] This may result from poor dexamethasone absorption, some degree of glucocorticoid resistance, and the chronic inflammatory state. The optimal approach in patients with clinical suspicion includes LDDST and measurement of LNSC because adequate suppression and normal LNSC levels exclude the diagnosis. However, elevated LNSC and abnormal dexamethasone suppression must be viewed with caution. Secondary diagnostic tests for Cushing disease, such as the desmopressin stimulation test, have not been extensively studied in patients with chronic renal failure.

FUTURE DIRECTIONS

Research aims and future expected improvements include:

- Decreasing the latency to diagnosis to reduce the long-term burden of disease.[24]
- Defining efficient and cost-effective algorithms for differential diagnosis, taking into account the pretest probability of Cushing disease and using modern diagnostic tools.[122]
- Improving imaging techniques for the visualization of corticotroph adenomas.[119,123,124]
- Evaluating whether presurgical medical treatment improves surgical outcomes and how long it should be prescribed for.[140]
- Evaluating the utility of venous thromboembolism prophylaxis before and after remission.[42]
- Identifying the best follow-up strategy to diagnose early recurrence at the individual level following successful pituitary surgery.
- Identifying quantitative biological markers of tissular exposure to cortisol to accurately monitor medical therapy doses.
- Evaluating new medical therapies in development and comparing the long-term impact of medical therapies versus bilateral adrenalectomy in patients with surgical failure.
- Defining efficient therapeutic approaches to improve cognitive defects and decreased quality of life in patients with a history of Cushing disease.
- Improving our understanding of the pathophysiology and the classification of corticotroph tumors with multiomics analysis to better anticipate the behavior of individual tumors and identify targeted medical treatments when surgery fails.

REFERENCES

1. Lacroix A, Feelders RA, Stratakis CA, et al. Cushing's syndrome. *Lancet*. 2015;386:913–927.
2. Valassi E, Santos A, Yaneva M, et al. The European Registry on Cushing's syndrome: 2-year experience. Baseline demographic and clinical characteristics. *Eur J Endocrinol*. 2011;165:383–392.
3. Sharma ST, Nieman LK, Feelders RA. Cushing's syndrome: epidemiology and developments in disease management. *Clinical Epidemiol*. 2015;7:281–293.
4. Ragnarsson O, Olsson DS, Chantzichristos D, et al. The incidence of Cushing's disease: a nationwide Swedish study. *Pituitary*. 2019;22: 179–186.
5. Etxabe J, Vazquez JA. Morbidity and mortality in Cushing's disease: an epidemiological approach. *Clin Endocrinol*. 1994;40:479–484.
6. Daly AF, Beckers A. The epidemiology of pituitary adenomas. *Endocrinol Metab Clin North Am*. 2020;49:347–355.
7. Mete O, Lopes MB. Overview of the 2017 WHO classification of pituitary tumors. *Endocr Pathol*. 2017;28:228–243.
8. Yoo F, Kuan EC, Heaney AP, et al. Corticotrophic pituitary carcinoma with cervical metastases: case series and literature review. *Pituitary*. 2018;21:290–301.
9. Raverot G, Burman P, McCormack A, et al. European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumors and carcinomas. *Eur J Endocrinol*. 2018;178:G1–G24.
10. Neou M, Villa C, Armignacco R, et al. Pangenomic classification of pituitary neuroendocrine tumors. *Canc Cell*. 2020;37:123–134 e5.
11. Uzilov AV, Taik P, Cheesman KC, et al. USP8 and TP53 drivers are associated with CNV in a corticotroph adenoma cohort enriched for aggressive tumors. *J Clin Endocrinol Metab*. 2021;106:826–842.
12. Casar-Borota O, Boldt HB, Engstrom BE, et al. Corticotroph aggressive pituitary tumors and carcinomas frequently harbour ATRX mutations. *J Clin Endocrinol Metab*. 2021;106:1183–1194.
13. Drummond J, Roncaroli F, Grossman AB, et al. Clinical and pathological aspects of silent pituitary adenomas. *J Clin Endocrinol Metab*. 2019;104:2473–2489.
14. Garcia-Martinez A, Cano DA, Flores-Martinez A, et al. Why don't corticotroph tumors always produce Cushing's disease? *Eur J Endocrinol*. 2019;181:351–361.
15. Raffin-Sanson ML, Massias JF, Dumont C, et al. High plasma proopiomelanocortin in aggressive adrenocorticotropin- secreting tumors. *J Clin Endocrinol Metab*. 1996;81:4272–4277.
16. Tatsi C, Flippo C, Stratakis CA. Cushing syndrome: old and new genes. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101418.
17. Srirangam Nadhamuni V, Korbonits M. Novel insights into pituitary tumorigenesis: genetic and epigenetic mechanisms. *Endocr Rev*. 2020;41:821–846.
18. Reincke M, Sbiera S, Hayakawa A, et al. Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet*. 2015;47:31–38.
19. Ma ZY, Song ZJ, Chen JH, et al. Recurrent gain-of-function USP8 mutations in Cushing's disease. *Cell Res*. 2015;25:306–317.
20. Wanichi IQ, de Paula Mariani BM, Frassetto FP, et al. Cushing's disease due to somatic USP8 mutations: a systematic review and meta-analysis. *Pituitary*. 2019;22:435–442.
21. Perez-Rivas LG, Theodoropoulou M, Puar TH, et al. Somatic USP8 mutations are frequent events in corticotroph tumor progression causing Nelson's tumor. *Eur J Endocrinol*. 2018;178:57–63.
22. Cohen M, Persky R, Stegemann R, et al. Germline USP8 mutation associated with pediatric cushing disease and other clinical features: a new syndrome. *J Clin Endocrinol Metab*. 2019;104:4676–4682.
23. Eisenhofer G, Masjkur J, Peitzsch M, et al. Plasma steroid metabolome profiling for diagnosis and subtyping patients with cushing syndrome. *Clin Chem*. 2018;64:586–596.
24. Rubinstein G, Osswald A, Hoster E, et al. Time to diagnosis in Cushing's syndrome: a meta-analysis based on 5367 patients. *J Clin Endocrinol Metab*. 2020;105:e12–e23.
25. Leon-Justel A, Madrazo-Atutxa A, Alvarez-Rios AI, et al. A probabilistic model for Cushing's syndrome screening in at-risk populations: a prospective multicenter study. *J Clin Endocrinol Metab*. 2016;101:3747–3754.
26. Nugent CA, Warner HR, Dunn JT, et al. Probability theory in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 1964;24:621–627.
27. Ferrau F, Korbonits M. Metabolic comorbidities in Cushing's syndrome. *Eur J Endocrinol*. 2015;173:M133–M157.
28. Drey M, Berr CM, Reincke M, et al. Cushing's syndrome: a model for sarcopenic obesity. *Endocrine*. 2017;57:481–485.
29. Pivonello R, Isidori AM, De Martino MC, et al. Complications of Cushing's syndrome: state of the art. *Lancet Diabet Endocrinol*. 2016;4:611–629.
30. Mazziotti G, Frara S, Giustina A. Pituitary diseases and bone. *Endocr Rev*. 2018;39:440–488.
31. Vinolas H, Grouthier V, Mehsen-Cetre N, et al. Assessment of vertebral microarchitecture in overt and mild Cushing's syndrome using trabecular bone score. *Clin Endocrinol*. 2018;89:148–154.
32. Starkman MN, Schteingart DE. Neuropsychiatric manifestations of patients with Cushing's syndrome. Relationship to cortisol and adrenocorticotropic hormone levels. *Arch Intern Med*. 1981;141:215–219.
33. Feelders RA, Pulgar SJ, Kempel A, et al. The burden of Cushing's disease: clinical and health-related quality of life aspects. *Eur J Endocrinol*. 2012;167:311–326.
34. Starkman MN. Neuropsychiatric findings in Cushing syndrome and exogenous glucocorticoid administration. *Endocrinol Metab Clin North Am*. 2013;42:477–488.
35. Young J, Haissaguerre M, Viera-Pinto O, et al. Management of endocrine disease: cushing's syndrome due to ectopic ACTH secretion: an expert operational opinion. *Eur J Endocrinol*. 2020;182:R29–R58.
36. Broersen LHA, Andela CD, Dekkers OM, et al. Improvement but no normalization of quality of life and cognitive functioning after treatment of Cushing syndrome. *J Clin Endocrinol Metab*. 2019;104:5325–5337.
37. Andela CD, van Haalen FM, Ragnarsson O, et al. Mechanisms in endocrinology: cushing's syndrome causes irreversible effects on the human brain: a systematic review of structural and functional magnetic resonance imaging studies. *Eur J Endocrinol*. 2015;173:R1–R14.

38. Isidori AM, Graziadio C, Paragliola RM, et al. The hypertension of Cushing's syndrome: controversies in the pathophysiology and focus on cardiovascular complications. *J Hypertens*. 2015;33:44–60.
39. Putignano P, Kaltsas GA, Korbonits M, et al. Alterations in serum protein levels in patients with Cushing's syndrome before and after successful treatment. *J Clin Endocrinol Metab*. 2000;85:3309–3312.
40. Stuijver DJ, van Zaane B, Feelders RA, et al. Incidence of venous thromboembolism in patients with Cushing's syndrome: a multicenter cohort study. *J Clin Endocrinol Metab*. 2011;96:3525–3532.
41. Boscaro M, Sonino N, Scarda A, et al. Anticoagulant prophylaxis markedly reduces thromboembolic complications in Cushing's syndrome. *J Clin Endocrinol Metab*. 2002;87:3662–3666.
42. Wagner J, Langlois F, Lim DST, et al. Hypercoagulability and risk of venous thromboembolic events in endogenous Cushing's syndrome: a systematic meta-analysis. *Front Endocrinol*. 2018;9:805.
43. Dekkers OM, Horvath-Puho E, Jorgensen JO, et al. Multisystem morbidity and mortality in Cushing's syndrome: a cohort study. *J Clin Endocrinol Metab*. 2013;98:2277–2284.
44. Papakokkinou E, Olsson DS, Chantzichristos D, et al. Excess morbidity persists in patients with Cushing's disease during long-term remission: a Swedish nationwide study. *J Clin Endocrinol Metab*. 2020;105:2616–2624.
45. van Haalen FM, Broersen LH, Jorgensen JO, et al. Management of endocrine disease: mortality remains increased in Cushing's disease despite biochemical remission: a systematic review and meta-analysis. *Eur J Endocrinol*. 2014;172:R143–R149.
46. Di Dalmazi G, Vicennati V, Garelli S, et al. Cardiovascular events and mortality in patients with adrenal incidentalomas that are either non-secreting or associated with intermediate phenotype or subclinical Cushing's syndrome: a 15-year retrospective study. *Lancet Diabet Endocrinol*. 2014;2:396–405.
47. Colao A, Pivonello R, Spiezia S, et al. Persistence of increased cardiovascular risk in patients with Cushing's disease after five years of successful cure. *J Clin Endocrinol Metab*. 1999;84:2664–2672.
48. Barahona MJ, Resmini E, Vilades D, et al. Coronary artery disease detected by multislice computed tomography in patients after long-term cure of Cushing's syndrome. *J Clin Endocrinol Metab*. 2013;98:1093–1099.
49. Kamenicky P, Redheuil A, Roux C, et al. Cardiac structure and function in Cushing's syndrome: a cardiac magnetic resonance imaging study. *J Clin Endocrinol Metab*. 2014;99:E2144–E2153.
50. Bartalena L, Martino E, Petrini L, et al. The nocturnal serum thyrotropin surge is abolished in patients with adrenocorticotropin (ACTH)-dependent or ACTH-independent Cushing's syndrome. *J Clin Endocrinol Metab*. 1991;72:1195–1199.
51. Pecori Giraldi F, Andrioli M, De Marinis L, et al. Significant GH deficiency after long-term cure by surgery in adult patients with Cushing's disease. *Eur J Endocrinol*. 2007;156:233–239.
52. Pecori Giraldi F, Moro M, Cavagnini F. Gender-related differences in the presentation and course of Cushing's disease. *J Clin Endocrinol Metab*. 2003;88:1554–1558.
53. Muller LM, Kienitz T, Deutschbein T, et al. Glucocorticoid receptor polymorphisms influence muscle strength in Cushing's syndrome. *J Clin Endocrinol Metab*. 2020;105:305–313.
54. Ragnarsson O, Glad CA, Bergthorsdottir R, et al. Body composition and bone mineral density in women with Cushing's syndrome in remission and the association with common genetic variants influencing glucocorticoid sensitivity. *Eur J Endocrinol*. 2015;172:1–10.
55. Bertagna X. Management of endocrine disease: can we cure Cushing's disease? A personal view. *Eur J Endocrinol*. 2018;178:R183–R200.
56. Tabarin A, Perez P. Pros and cons of screening for occult Cushing syndrome. *Nature Rev*. 2011;7:445–455.
57. Terzolo M, Reimondo G, Chiodini I, et al. Screening of Cushing's syndrome in outpatients with type 2 diabetes: results of a prospective multicentric study in Italy. *J Clin Endocrinol Metab*. 2012;97:3467–3475.
58. van Hulsteijn LT, Pasquali R, Casanueva F, et al. Prevalence of endocrine disorders in obese patients: systematic review and meta-analysis. *Eur J Endocrinol*. 2020;182:11–21.
59. Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing's syndrome: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1526–1540.
60. Elamin MB, Murad MH, Mullan R, et al. Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. *J Clin Endocrinol Metab*. 2008;93:1553–1562.
61. Galm BP, Qiao N, Klibanski A, et al. Accuracy of laboratory tests for the diagnosis of Cushing syndrome. *J Clin Endocrinol Metab*. 2020;105:2081–2094.
62. Klose M, Lange M, Rasmussen AK, et al. Factors influencing the adrenocorticotropin test: role of contemporary cortisol assays, body composition, and oral contraceptive agents. *J Clin Endocrinol Metab*. 2007;92:1326–1333.
63. Kannankeril J, Carroll T, Findling JW, et al. Prospective evaluation of late-night salivary cortisol and cortisone by EIA and LC-MS/MS in suspected Cushing syndrome. *J Endocrine Soc*. 2020;4:bvaa107.
64. Corcuff JB, Tabarin A, Rashedi M, et al. Overnight urinary free cortisol determination: a screening test for the diagnosis of Cushing's syndrome. *Clin Endocrinol*. 1998;48:503–508.
65. Graham UM, Hunter SJ, McDonnell M, et al. A comparison of the use of urinary cortisol to creatinine ratios and nocturnal salivary cortisol in the evaluation of cyclicity in patients with Cushing's syndrome. *J Clin Endocrinol Metab*. 2013;98:E72–E76.
66. Osswald A, Wang R, Beuschlein F, et al. Performance of LC-MS/MS and immunoassay based 24-h urine free cortisol in the diagnosis of Cushing's syndrome. *J Steroid Biochem Mol Biol*. 2019;190:193–197.
67. Mericq MV, Cutler Jr GB. High fluid intake increases urine free cortisol excretion in normal subjects [see comments]. *J Clin Endocrinol Metab*. 1998;83:682–684.
68. Chan KC, Lit LC, Law EL, et al. Diminished urinary free cortisol excretion in patients with moderate and severe renal impairment. *Clin Chem*. 2004;50:757–759.
69. Elias PC, Martinez EZ, Barone BF, et al. Late-night salivary cortisol has a better performance than urinary free cortisol in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 2014;99:2045–2051.
70. Kidambi S, Raff H, Findling JW. Limitations of nocturnal salivary cortisol and urine free cortisol in the diagnosis of mild Cushing's syndrome. *Eur J Endocrinol*. 2007;157:725–731.
71. Putignano P, Toja P, Dubini A, et al. Midnight salivary cortisol versus urinary free and midnight serum cortisol as screening tests for Cushing's syndrome. *J Clin Endocrinol Metab*. 2003;88:4153–4157.
72. Raff H, Raff JL, Findling JW. Late-night salivary cortisol as a screening test for Cushing's syndrome. *J Clin Endocrinol Metab*. 1998;83:2681–2686.
73. Petersenn S, Newell-Price J, Findling JW, et al. High variability in baseline urinary free cortisol values in patients with Cushing's disease. *Clin Endocrinol*. 2014;80:261–269.
74. Backlund N, Brattsand G, Israelsson M, et al. Reference intervals of salivary cortisol and cortisone and their diagnostic accuracy in Cushing's syndrome. *Eur J Endocrinol*. 2020;182:569–582.
75. Nunes ML, Vattaut S, Corcuff JB, et al. Late-night salivary cortisol for diagnosis of overt and subclinical Cushing's syndrome in hospitalized and ambulatory patients. *J Clin Endocrinol Metab*. 2009;94:456–462.
76. Raff H, Phillips JM. Bedtime salivary cortisol and cortisone by LC-MS/MS in healthy adult subjects: evaluation of sampling time. *J Endocrine Soc*. 2019;3:1631–1640.
77. Weibel L, Brandenberger G. The start of the quiescent period of cortisol remains phase locked to the melatonin onset despite circadian phase alterations in humans working the night schedule. *Neurosci Lett*. 2002;318:89–92.
78. Sandouk Z, Johnston P, Bunch D, et al. Variability of late-night salivary cortisol in Cushing disease: a prospective study. *J Clin Endocrinol Metab*. 2018;103:983–990.
79. Lammert A, Nittka S, Otto M, et al. Performance of the 1 mg dexamethasone suppression test in patients with severe obesity. *Obesity*. 2016;24:850–855.
80. Ueland GA, Methlie P, Kellmann R, et al. Simultaneous assay of cortisol and dexamethasone improved diagnostic accuracy of the dexamethasone suppression test. *Eur J Endocrinol*. 2017;176:705–713.
81. Findling JW, Raff H, Aron DC. The low-dose dexamethasone suppression test: a reevaluation in patients with Cushing's syndrome. *J Clin Endocrinol Metab*. 2004;89:1222–1226.

82. Pecori Giraldi F, Pivonello R, Ambrogio AG, et al. The dexamethasone-suppressed corticotropin-releasing hormone stimulation test and the desmopressin test to distinguish Cushing's syndrome from pseudo-Cushing's states. *Clin Endocrinol*. 2007;66:251–257.
83. Valassi E, Swearingen B, Lee H, et al. Concomitant medication use can confound interpretation of the combined dexamethasone-corticotropin releasing hormone test in Cushing's syndrome. *J Clin Endocrinol Metab*. 2009;94:4851–4859.
84. Vastbinder M, Kuindersma M, Mulder AH, et al. The influence of oral contraceptives on overnight 1 mg dexamethasone suppression test. *Netherlands J Med*. 2016;74:158–161.
85. Gatta B, Chabre O, Cortet C, et al. Reevaluation of the combined dexamethasone suppression-corticotropin-releasing hormone test for differentiation of mild Cushing's disease from pseudo-Cushing's syndrome. *J Clin Endocrinol Metab*. 2007;92:4290–4293.
86. Martin NM, Dhillo WS, Banerjee A, et al. Comparison of the dexamethasone-suppressed corticotropin-releasing hormone test and low-dose dexamethasone suppression test in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 2006;91:2582–2586.
87. Papanicolaou DA, Yanovski JA, Cutler Jr GB, et al. A single midnight serum cortisol measurement distinguishes Cushing's syndrome from pseudo-Cushing states. *J Clin Endocrinol Metab*. 1998;83:1163–1167.
88. Reimondo G, Allasino B, Bovio S, et al. Evaluation of the effectiveness of midnight serum cortisol in the diagnostic procedures for Cushing's syndrome. *Eur J Endocrinol*. 2005;153:803–809.
89. Brossaud J, Charret L, De Angeli D, et al. Hair cortisol and cortisone measurements for the diagnosis of overt and mild Cushing's syndrome. *Eur J Endocrinol*. 2021;184:445–454.
90. Wester VL, van Rossum EF. Clinical applications of cortisol measurements in hair. *Eur J Endocrinol*. 2015;173:M1–M10.
91. Ceccato F, Tizianel I, Vedolin CK, et al. Human corticotropin-releasing hormone tests: 10 years of real-life experience in pituitary and adrenal disease. *J Clin Endocrinol Metab*. 2020;105:e3938–e3949.
92. Greene LW, Geer EB, Page-Wilson G, et al. Assay-specific spurious ACTH results lead to misdiagnosis, unnecessary testing, and surgical misadventure-a case series. *J Endocrine Soc*. 2019;3:763–772.
93. Pecori Giraldi F, Saccani A, Cavagnini F. Assessment of ACTH assay variability: a multicenter study. *Eur J Endocrinol*. 2011;164:505–512.
94. Sohaib SA, Hanson JA, Newell-Price JD, et al. CT appearance of the adrenal glands in adrenocorticotrophic hormone-dependent Cushing's syndrome. *AJR Am J Roentgenol*. 1999;172:997–1002.
95. Findling JW, Raff H. Diagnosis of endocrine disease: differentiation of pathologic/neoplastic hypercortisolism (Cushing's syndrome) from physiologic/non-neoplastic hypercortisolism (formerly known as pseudo-Cushing's syndrome). *Eur J Endocrinol*. 2017;176:R205–R216.
96. Scaroni C, Albiger NM, Palmieri S, et al. Approach to patients with pseudo-Cushing's states. *Endocrine Connect*. 2020;9:R1–R13.
97. Dahia PL, Grossman AB. The molecular pathogenesis of corticotroph tumors. *Endocr Rev*. 1999;20:136–155.
98. Vassiliadi DA, Tsagarakis S. Diagnosis of endocrine disease: the role of the desmopressin test in the diagnosis and follow-up of Cushing's syndrome. *Eur J Endocrinol*. 2018;178:R201–R214.
99. Hall WA, Luciano MG, Doppman JL, et al. Pituitary magnetic resonance imaging in normal human volunteers: occult adenomas in the general population. *Ann Intern Med*. 1994;120:817–820.
100. Vitellius G, Trabado S, Hoeffel C, et al. Significant prevalence of NR3C1 mutations in incidentally discovered bilateral adrenal hyperplasia: results of the French MUTA-GR Study. *Eur J Endocrinol*. 2018;178:411–423.
101. Hayes AR, Grossman AB. The ectopic adrenocorticotropic hormone syndrome: rarely easy, always challenging. *Endocrinol Metab Clin North Am*. 2018;47:409–425.
102. More J, Young J, Reznik Y, et al. Ectopic ACTH syndrome in children and adolescents. *J Clin Endocrinol Metab*. 2011;96:1213–1222.
103. Ilias I, Torpy DJ, Pacak K, et al. Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institutes of Health. *J Clin Endocrinol Metab*. 2005;90:4955–4962.
104. Nakhjavani M, Amirbaigloo A, Rabizadeh S, et al. Ectopic Cushing's syndrome due to corticotropin releasing hormone. *Pituitary*. 2019;22:561–568.
105. Young J, Deneux C, Grino M, et al. Pitfall of petrosal sinus sampling in a Cushing's syndrome secondary to ectopic adrenocorticotropin-corticotropin releasing hormone (ACTH-CRH) secretion. *J Clin Endocrinol Metab*. 1998;83:305–308.
106. Aron DC, Raff H, Findling JW. Effectiveness versus efficacy: the limited value in clinical practice of high dose dexamethasone suppression testing in the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 1997;82:1780–1785.
107. Dichek HL, Nieman LK, Oldfield EH, et al. A comparison of the standard high dose dexamethasone suppression test and the overnight 8-mg dexamethasone suppression test for the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 1994;78:418–422.
108. Ritzel K, Beuschlein F, Berr C, et al. ACTH after 15 min distinguishes between Cushing's disease and ectopic Cushing's syndrome: a proposal for a short and simple CRH test. *Eur J Endocrinol*. 2015;173:197–204.
109. Findling JW, Raff H. Cushing's syndrome: important issues in diagnosis and management. *J Clin Endocrinol Metab*. 2006;91:3746–3753.
110. Frete C, Corcuff JB, Kuhn E, et al. Non-invasive diagnostic strategy in ACTH-dependent Cushing's Syndrome. *J Clin Endocrinol Metab*. 2020;105:3273–3284.
111. Reimondo G, Paccotti P, Minetto M, et al. The corticotrophin-releasing hormone test is the most reliable noninvasive method to differentiate pituitary from ectopic ACTH secretion in Cushing's syndrome. *Clin Endocrinol*. 2003;58:718–724.
112. Isidori AM, Kaltsas GA, Mohammed S, et al. Discriminatory value of the low-dose dexamethasone suppression test in establishing the diagnosis and differential diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 2003;88:5299–5306.
113. Newell-Price J, Morris DG, Drake WM, et al. Optimal response criteria for the human CRH test in the differential diagnosis of ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 2002;87:1640–1645.
114. Nieman LK, Oldfield EH, Wesley R, et al. A simplified morning ovine corticotropin-releasing hormone stimulation test for the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 1993;77:1308–1312.
115. Tsagarakis S, Tsigos C, Vasiliou V, et al. The desmopressin and combined CRH-desmopressin tests in the differential diagnosis of ACTH-dependent Cushing's syndrome: constraints imposed by the expression of V2 vasopressin receptors in tumors with ectopic ACTH secretion. *J Clin Endocrinol Metab*. 2002;87:1646–1653.
116. Barbot M, Trementino L, Zilio M, et al. Second-line tests in the differential diagnosis of ACTH-dependent Cushing's syndrome. *Pituitary*. 2016;19:488–495.
117. Cristante J, Lefournier V, Sturm N, et al. Why we should still treat by neurosurgery patients with Cushing's disease and a normal or inconclusive pituitary MRI. *J Clin Endocrinol Metab*. 2019;104:4101–4113.
118. Stroud A, Dhaliwal P, Alvarado R, et al. Outcomes of pituitary surgery for Cushing's disease: a systematic review and meta-analysis. *Pituitary*. 2020;23:595–609.
119. Vitale G, Tortora F, Baldelli R, et al. Pituitary magnetic resonance imaging in Cushing's disease. *Endocrine*. 2017;55:691–696.
120. Wind JJ, Lonser RR, Nieman LK, et al. The lateralization accuracy of inferior petrosal sinus sampling in 501 patients with Cushing's disease. *J Clin Endocrinol Metab*. 2013;98:2285–2293.
121. Lonser RR, Nieman L, Oldfield EH. Cushing's disease: pathobiology, diagnosis, and management. *J Neurosurg*. 2017;126:404–417.
122. Nieman LK. Is it time for a new approach to the differential diagnosis of ACTH-dependent Cushing syndrome? *J Clin Endocrinol Metab*. 2020;105:e4964–e4966.
123. Koulouri O, Steuwe A, Gillett D, et al. A role for 11C-methionine PET imaging in ACTH-dependent Cushing's syndrome. *Eur J Endocrinol*. 2015;173:M107–M120.
124. Walia R, Gupta R, Bhansali A, et al. Molecular imaging targeting corticotropin releasing hormone receptor for corticotropinoma: a changing paradigm. *J Clin Endocrinol Metab*. 2020;106:e1816–e1826.
125. Pecori Giraldi F, Cavallo LM, Tortora F, et al. The role of inferior petrosal sinus sampling in ACTH-dependent Cushing's syndrome: review and joint opinion statement by members of the Italian Society for

Endocrinology, Italian Society for Neurosurgery, and Italian Society for Neuroradiology. *Neurosurg Focus*. 2015;38:E5.
126. Tsagarakis S, Vassiliadi D, Kaskarelis IS, et al. The application of the combined corticotropin-releasing hormone plus desmopressin stimulation during petrosal sinus sampling is both sensitive and specific in differentiating patients with Cushing's disease from patients with the occult ectopic adrenocorticotropin syndrome. *J Clin Endocrinol Metab*. 2007;92:2080–2086.
127. Lefournier V, Gatta B, Martinie M, et al. One transient neurological complication (sixth nerve palsy) in 166 consecutive bilateral inferior petrosal sinus sampling for the etiological diagnosis of ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 1999;84:3401–3406.
128. Oldfield EH, Doppman JL, Nieman LK, et al. Petrosal sinus sampling with and without corticotropin-releasing hormone for the differential diagnosis of Cushing's syndrome [published erratum appears in N Engl J Med 1992 Apr 23;326(17):1172] [see comments]. *N Engl J Med*. 1991;325:897–905.
129. Swearingen B, Katznelson L, Miller K, et al. Diagnostic errors after inferior petrosal sinus sampling. *J Clin Endocrinol Metab*. 2004;89:3752–3763.
130. De Sousa SMC, McCormack AI, McGrath S, et al. Prolactin correction for adequacy of petrosal sinus cannulation may diminish diagnostic accuracy in Cushing's disease. *Clin Endocrinol*. 2017;87:515–522.
131. Ilias I, Chang R, Pacak K, et al. Jugular venous sampling: an alternative to petrosal sinus sampling for the diagnostic evaluation of adrenocorticotropic hormone-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 2004;89:3795–3800.
132. Oldfield EH, Chrousos GP, Schulte HM, et al. Preoperative lateralization of ACTH-secreting pituitary microadenomas by bilateral and simultaneous inferior petrosal venous sinus sampling. *N Engl J Med*. 1985;312:100–103.
133. Lefournier V, Martinie M, Vasdev A, et al. Accuracy of bilateral inferior petrosal or cavernous sinuses sampling in predicting the lateralization of Cushing's disease pituitary microadenoma: influence of catheter position and anatomy of venous drainage. *J Clin Endocrinol Metab*. 2003;88:196–203.
134. Liu C, Lo JC, Dowd CF, et al. Cavernous and inferior petrosal sinus sampling in the evaluation of ACTH-dependent Cushing's syndrome. *Clin Endocrinol*. 2004;61:478–486.
135. Sheth SA, Mian MK, Neal J, et al. Transsphenoidal surgery for Cushing disease after nondiagnostic inferior petrosal sinus sampling. *Neurosurgery*. 2012;71:14–22.
136. Isidori AM, Sbardella E, Zatelli MC, et al. Conventional and nuclear medicine imaging in ectopic Cushing's syndrome: a systematic review. *J Clin Endocrinol Metab*. 2015;100:3231–3244.
137. Testa RM, Albiger N, Occhi G, et al. The usefulness of combined biochemical tests in the diagnosis of Cushing's disease with negative pituitary magnetic resonance imaging. *Eur J Endocrinol*. 2007;156:241–248.
138. Nieman LK, Biller BM, Findling JW, et al. Treatment of Cushing's syndrome: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2015;100:2807–2831.
139. Geer EB, Shafiq I, Gordon MB, et al. Biochemical control during long-term follow-up of 230 adult patients with Cushing disease: a multicenter retrospective study. *Endocr Pract*. 2017;23:962–970.
140. Valassi E, Franz H, Brue T, et al. Preoperative medical treatment in Cushing's syndrome: frequency of use and its impact on postoperative assessment: data from ERCUSYN. *Eur J Endocrinol*. 2018;178:399–409.
141. Petersenn S, Beckers A, Ferone D, et al. Therapy of endocrine disease: outcomes in patients with Cushing's disease undergoing transsphenoidal surgery: systematic review assessing criteria used to define remission and recurrence. *Eur J Endocrinol*. 2015;172:R227–R239.
142. Pivonello R, De Leo M, Cozzolino A, et al. The treatment of Cushing's disease. *Endocr Rev*. 2015;36:385–486.
143. Broersen LHA, Biermasz NR, van Furth WR, et al. Endoscopic vs. microscopic transsphenoidal surgery for Cushing's disease: a systematic review and meta-analysis. *Pituitary*. 2018;21:524–534.
144. Braun LT, Rubinstein G, Zopp S, et al. Recurrence after pituitary surgery in adult Cushing's disease: a systematic review on diagnosis and treatment. *Endocrine*. 2020;70:218–231.
145. Salenave S, Gatta B, Pecheur S, et al. Pituitary magnetic resonance imaging findings do not influence surgical outcome in adrenocorticotropin-secreting microadenomas. *J Clin Endocrinol Metab*. 2004;89:3371–3376.
146. Baudry C, Coste J, Bou Khalil R, et al. Efficiency and tolerance of mitotane in Cushing's disease in 76 patients from a single center. *Eur J Endocrinol*. 2012;167:473–481.
147. Castinetti F, Morange I, Jaquet P, et al. Ketoconazole revisited: a preoperative or postoperative treatment in Cushing's disease. *Eur J Endocrinol*. 2008;158:91–99.
148. Ironside N, Chatain G, Asuzu D, et al. Earlier post-operative hypocortisolemia may predict durable remission from Cushing's disease. *Eur J Endocrinol*. 2018;178:255–263.
149. Amlashi FG, Swearingen B, Faje AT, et al. Accuracy of late-night salivary cortisol in evaluating postoperative remission and recurrence in Cushing's disease. *J Clin Endocrinol Metab*. 2015;100:3770–3777.
150. Valassi E, Biller BM, Swearingen B, et al. Delayed remission after transsphenoidal surgery in patients with Cushing's disease. *J Clin Endocrinol Metab*. 2010;95:601–610.
151. Bochicchio D, Losa M, Buchfelder M. Factors influencing the immediate and late outcome of Cushing's disease treated by transsphenoidal surgery: a retrospective study by the European Cushing's Disease Survey Group. *J Clin Endocrinol Metab*. 1995;80:3114–3120.
152. Broersen LHA, van Haalen FM, Kienitz T, et al. The incidence of adrenal crisis in the postoperative period of HPA axis insufficiency after surgical treatment for Cushing's syndrome. *Eur J Endocrinol*. 2019;181:201–210.
153. Hoybye C, Ragnarsson O, Jonsson PJ, et al. Clinical features of GH deficiency and effects of 3 years of GH replacement in adults with controlled Cushing's disease. *Eur J Endocrinol*. 2010;162:677–684.
154. Webb SM, Mo D, Lamberts SW, et al. Metabolic, cardiovascular, and cerebrovascular outcomes in growth hormone-deficient subjects with previous Cushing's disease or non-functioning pituitary adenoma. *J Clin Endocrinol Metab*. 2010;95:630–638.
155. Alexandraki KI, Kaltsas GA, Isidori AM, et al. Long-term remission and recurrence rates in Cushing's disease: predictive factors in a single-centre study. *Eur J Endocrinol*. 2013;168:639–648.
156. Vassiliadi DA, Balomenaki M, Asimakopoulou A, et al. The desmopressin test predicts better than basal cortisol the long-term surgical outcome of Cushing's disease. *J Clin Endocrinol Metab*. 2016;101:4878–4885.
157. Hinojosa-Amaya JM, Varlamov EV, McCartney S, et al. Hypercortisolemia Recurrence in Cushing's disease; a diagnostic challenge. *Front Endocrinol*. 2019;10:740.
158. Cambos S, Mohammedi K, Castinetti F, et al. Persistent cortisol response to desmopressin predicts recurrence of Cushing's disease in patients with post-operative corticotropic insufficiency. *Eur J Endocrinol*. 2020;182:489–498.
159. Bou Khalil R, Baudry C, Guignat L, et al. Sequential hormonal changes in 21 patients with recurrent Cushing's disease after successful pituitary surgery. *Eur J Endocrinol*. 2011;165:729–737.
160. Danet-Lamasou M, Asselineau J, Perez P, et al. Accuracy of repeated measurements of late-night salivary cortisol to screen for early-stage recurrence of Cushing's disease following pituitary surgery. *Clin Endocrinol*. 2014;82:260–266.
161. Ambrogio AG, Andrioli M, De Martin M, et al. Usefulness of desmopressin testing to predict relapse during long-term follow-up in patients in remission from Cushing's disease. *Endocr Connect*. 2017;6:791–799.
162. Le Marc'hadour P, Muller M, Albarel F, et al. Postoperative follow-up of Cushing's disease using cortisol, desmopressin and coupled dexamethasone-desmopressin tests: a head-to-head comparison. *Clin Endocrinol*. 2015;83:216–222.
163. Carroll TB, Javorsky BR, Findling JW. Postsurgical recurrent Cushing disease: clinical benefit of early intervention in patients with normal urinary free cortisol. *Endocr Pract*. 2016;22:1216–1223.
164. Rubinstein G, Osswald A, Zopp S, et al. Therapeutic options after surgical failure in Cushing's disease: a critical review. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101270.
165. Ironside N, Chen CJ, Lee CC, et al. Outcomes of pituitary radiation for Cushing's disease. *Endocrinol Metab Clin North Am*. 2018;47:349–365.
166. Mehta GU, Ding D, Patibandla MR, et al. Stereotactic radiosurgery for Cushing disease: results of an international, multicenter study. *J Clin Endocrinol Metab*. 2017;102:4284–4291.

167. Shepard MJ, Mehta GU, Xu Z, et al. Technique of whole-sellar stereotactic radiosurgery for Cushing disease: results from a multicenter, international cohort study. *World Neurosurg*. 2018;116:e670–e679.
168. Castinetti F, Brue T, Ragnarsson O. Radiotherapy as a tool for the treatment of Cushing's disease. *Eur J Endocrinol*. 2019;180:D9–D18.
169. Thakkar K, Lila A, Sarathi V, et al. Cabergoline may act as a radioprotective agent in Cushing's disease. *Clin Endocrinol*. 2020;92:55–62.
170. Mehta GU, Ding D, Gupta A, et al. Repeat stereotactic radiosurgery for Cushing's disease: outcomes of an international, multicenter study. *J Neuro Oncol*. 2018;138:519–525.
171. Burman P, van Beek AP, Biller BM, et al. Radiotherapy, especially at young age, increases the risk for de novo brain tumors in patients treated for pituitary/sellar lesions. *J Clin Endocrinol Metab*. 2017;102:1051–1058.
172. Guerin C, Taieb D, Treglia G, et al. Bilateral adrenalectomy in the 21st century: when to use it for hypercortisolism? *Endocr Relat Cancer*. 2016;23:R131–R142.
173. Reincke M, Ritzel K, Osswald A, et al. A critical reappraisal of bilateral adrenalectomy for ACTH-dependent Cushing's syndrome. *Eur J Endocrinol*. 2015;173:M23–M32.
174. Fountas A, Karavitaki N. Nelson's syndrome: an update. *Endocrinol Metab Clin North Am*. 2020;49:413–432.
175. Assie G, Bahurel H, Coste J, et al. Corticotroph tumor progression after adrenalectomy in Cushing's Disease: a reappraisal of Nelson's Syndrome. *J Clin Endocrinol Metab*. 2007;92:172–179.
176. Fleseriu M, Biller BM, Findling JW, et al. Mifepristone, a glucocorticoid receptor antagonist, produces clinical and metabolic benefits in patients with Cushing's syndrome. *J Clin Endocrinol Metab*. 2012;97:2039–2049.
177. Fountas A, Lim ES, Drake WM, et al. Outcomes of patients with Nelson's syndrome after primary treatment: a multicenter study from 13 UK pituitary centers. *J Clin Endocrinol Metab*. 2020;105:1527–1537.
178. Newell-Price J, Nieman LK, Reincke M, et al. Endocrinology in the time of Covid-19: management of Cushing's syndrome. *Eur J Endocrinol*. 2020;183:G1–G7.
179. Fleseriu M, Pivonello R, Young J, et al. Osilodrostat, a potent oral 11beta-hydroxylase inhibitor: 22-week, prospective, Phase II study in Cushing's disease. *Pituitary*. 2016;19:138–148.
180. Newell-Price J, Pivonello R, Tabarin A, et al. Use of late-night salivary cortisol to monitor response to medical treatment in Cushing's disease. *Eur J Endocrinol*. 2020;182:207–217.
181. Pivonello R, Fleseriu M, Newell-Price J, et al. Efficacy and safety of osilodrostat in patients with Cushing's disease (LINC 3): a multicentre phase III study with a double-blind, randomised withdrawal phase. *Lancet Diabetes Endocrinol*. 2020;8:748–761.
182. Colao A, Petersenn S, Newell-Price J, et al. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med*. 2012;366:914–924.
183. Lacroix A, Gu F, Gallardo W, et al. Efficacy and safety of once-monthly pasireotide in Cushing's disease: a 12 month clinical trial. *Lancet Diabetes Endocrinol*. 2018;6:17–26.
184. Ferriere A, Cortet C, Chanson P, et al. Cabergoline for Cushing's disease: a large retrospective multicenter study. *Eur J Endocrinol*. 2017;176:305–314.
185. Burman P, Engstrom BE, Ekman B, et al. Limited value of cabergoline in Cushing's disease: a prospective study of a 6 weeks treatment in 20 patients. *J Clin Endocrinol Metabol*. 2016;174:17–24.
186. Giraldi EA, Ioachimescu AG. The role of dopamine agonists in pituitary adenomas. *Endocrinol Metab Clin North Am*. 2020;49:453–474.
187. Daniel E, Newell-Price JD. Therapy of endocrine disease: steroidogenesis enzyme inhibitors in Cushing's syndrome. *Eur J Endocrinol*. 2015;172:R263–R280.
188. McCance DR, Hadden DR, Kennedy L, et al. Clinical experience with ketoconazole as a therapy for patients with Cushing's syndrome. *Clin Endocrinol*. 1987;27:593–599.
189. Castinetti F, Guignat L, Giraud P, et al. Ketoconazole in Cushing's disease: is it worth a try? *J Clin Endocrinol Metab*. 2014;99:1623–1630.
190. Young J, Bertherat J, Vantyghem MC, et al. Hepatic safety of ketoconazole in Cushing's syndrome: results of a Compassionate Use Programme in France. *Eur J Endocrinol*. 2018;178:447–458.
191. Monaghan PJ, Owen LJ, Trainer PJ, et al. Comparison of serum cortisol measurement by immunoassay and liquid chromatography-tandem mass spectrometry in patients receiving the 11beta-hydroxylase inhibitor metyrapone. *Ann Clin Biochem*. 2011;48:441–446.
192. Daniel E, Aylwin S, Mustafa O, et al. Effectiveness of metyrapone in treating Cushing's syndrome: a retrospective multicenter study in 195 patients. *J Clin Endocrinol Metab*. 2015;100:4146–4154.
193. Salenave S, Bernard V, Do Cao C, et al. Ovarian macrocysts and gonadotrope-ovarian axis disruption in premenopausal women receiving mitotane for adrenocortical carcinoma or Cushing's disease. *Eur J Endocrinol*. 2015;172:141–149.
194. Fleseriu M, Pivonello R, Elenkova A, et al. Efficacy and safety of levoketoconazole in the treatment of endogenous Cushing's syndrome (SONICS): a phase 3, multicentre, open-label, single-arm trial. *Lancet Diabet Endocrinol*. 2019;7:855–865.
195. Carroll TB, Peppard WJ, Herrmann DJ, et al. Continuous etomidate infusion for the management of severe Cushing syndrome: validation of a standard protocol. *J Endocr Soc*. 2019;3:1–12.
196. Preda VA, Sen J, Karavitaki N, Grossman AB. Etomidate in the management of hypercortisolaemia in Cushing's syndrome: a review. *Eur J Endocrinol*. 2012;167:137–143.
197. Constantinescu SM, Driessens N, Lefebvre A, et al. Etomidate infusion at low doses is an effective and safe treatment for severe Cushing's syndrome outside intensive care. *Eur J Endocrinol*. 2020;183:161–167.
198. Fein HG, Vaughan 3rd TB, Kushner H, et al. Sustained weight loss in patients treated with mifepristone for Cushing's syndrome: a follow-up analysis of the SEISMIC study and long-term extension. *BMC Endocr Disord*. 2015;15:63.
199. Corcuff JB, Young J, Masquefa-Giraud P, et al. Rapid control of severe neoplastic hypercortisolism with metyrapone and ketoconazole. *Eur J Endocrinol*. 2015;172:473–481.
200. Kamenicky P, Droumaguet C, Salenave S, et al. Mitotane, metyrapone, and ketoconazole combination therapy as an alternative to rescue adrenalectomy for severe ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 2011;96:2796–2804.
201. Feelders RA, Newell-Price J, Pivonello R, et al. Advances in the medical treatment of Cushing's syndrome. *Lancet Diabet Endocrinol*. 2019;7:300–312.
202. Barahona MJ, Sucunza N, Resmini E, et al. Persistent body fat mass and inflammatory marker increases after long-term cure of Cushing's syndrome. *J Clin Endocrinol Metab*. 2009;94:3365–3371.
203. Lambert JK, Goldberg L, Fayngold S, et al. Predictors of mortality and long-term outcomes in treated Cushing's disease: a study of 346 patients. *J Clin Endocrinol Metab*. 2013;98:1022–1030.
204. Berr CM, Stieg MR, Deutschbein T, et al. Persistence of myopathy in Cushing's syndrome: evaluation of the German Cushing's Registry. *Eur J Endocrinol*. 2017;176:737–746.
205. Broersen LHA, Jha M, Biermasz NR, et al. Effectiveness of medical treatment for Cushing's syndrome: a systematic review and meta-analysis. *Pituitary*. 2018;21:631–641.
206. Ragnarsson O, Olsson DS, Papakokkinou E, et al. Overall and disease-specific mortality in patients with Cushing Disease: a Swedish nationwide study. *J Clin Endocrinol Metab*. 2019;104:2375–2384.
207. Webb SM, Badia X, Barahona MJ, et al. Evaluation of health-related quality of life in patients with Cushing's syndrome with a new questionnaire. *Eur J Endocrinol*. 2008;158:623–630.
208. Santos A, Resmini E, Martinez Momblan MA, et al. Quality of life in patients with Cushing's sisease. *Front Endocrinol*. 2019;10:862.
209. Tiemensma J, Kaptein AA, Pereira AM, et al. Negative illness perceptions are associated with impaired quality of life in patients after long-term remission of Cushing's syndrome. *Eur J Endocrinol*. 2011;165:527–535.
210. Clayton RN, Jones PW, Reulen RC, et al. Mortality in patients with Cushing's disease more than 10 years after remission: a multicentre, multinational, retrospective cohort study. *Lancet Diabet Endocrinol*. 2016;4:569–576.
211. Ntali G, Asimakopoulou A, Siamatras T, et al. Mortality in Cushing's syndrome: systematic analysis of a large series with prolonged follow-up. *Eur J Endocrinol*. 2013;169:715–723.

212. Ritzel K, Beuschlein F, Mickisch A, et al. Clinical review: outcome of bilateral adrenalectomy in Cushing's syndrome: a systematic review. *J Clin Endocrinol Metab*. 2013;98:3939–3948.
213. Alexandraki KI, Kaltsas GA, Isidori AM, et al. The prevalence and characteristic features of cyclicity and variability in Cushing's disease. *Eur J Endocrinol*. 2009;160:1011–1018.
214. Meinardi JR, Wolffenbuttel BH, Dullaart RP. Cyclic Cushing's syndrome: a clinical challenge. *Eur J Endocrinol*. 2007;157:245–254.
215. Storr HL, Savage MO. Management of endocrine disease: Paediatric Cushing's disease. *Eur J Endocrinol*. 2015;173:R35–R45.
216. Faucz FR, Tirosh A, Tatsi C, et al. Somatic USP8 gene mutations are a common cause of pediatric Cushing disease. *J Clin Endocrinol Metab*. 2017;102:2836–2843.
217. Magiakou MA, Mastorakos G, Oldfield EH, et al. Cushing's syndrome in children and adolescents. Presentation, diagnosis, and therapy. *N Engl J Med*. 1994;331:629–636.
218. Lodish MB, Gourgari E, Sinaii N, et al. Skeletal maturation in children with Cushing syndrome is not consistently delayed: the role of corticotropin, obesity, and steroid hormones, and the effect of surgical cure. *J Pediatr*. 2014;164:801–806.
219. Yordanova G, Martin L, Afshar F, et al. Long-term outcomes of children treated for Cushing's disease: a single center experience. *Pituitary*. 2016;19:612–624.
220. Batista DL, Riar J, Keil M, et al. Diagnostic tests for children who are referred for the investigation of Cushing syndrome. *Pediatrics*. 2007;120:e575–e586.
221. Devoe DJ, Miller WL, Conte FA, et al. Long-term outcome in children and adolescents after transsphenoidal surgery for Cushing's disease. *J Clin Endocrinol Metab*. 1997;82:3196–3202.
222. Storr HL, Plowman PN, Carroll PV, et al. Clinical and endocrine responses to pituitary radiotherapy in pediatric Cushing's disease: an effective second-line treatment. *J Clin Endocrinol Metab*. 2003;88:34–37.
223. Valdes N, Tirosh A, Keil M, et al. Pediatric Cushing's syndrome: greater risk of being overweight or obese after long-term remission and its predictive factors. *Eur J Endocrinol*. 2021;184:179–187.
224. Lodish MB, Sinaii N, Patronas N, et al. Blood pressure in pediatric patients with Cushing syndrome. *J Clin Endocrinol Metab*. 2009;94:2002–2008.
225. Keil MF, Merke DP, Gandhi R, et al. Quality of life in children and adolescents 1-year after cure of Cushing syndrome: a prospective study. *Clin Endocrinol*. 2009;71:326–333.
226. Merke DP, Giedd JN, Keil MF, et al. Children experience cognitive decline despite reversal of brain atrophy one year after resolution of cushing syndrome. *J Clin Endocrinol Metab*. 2005;90:2531–2536.
227. Keil MF, Zametkin A, Ryder C, et al. Cases of psychiatric morbidity in pediatric patients after remission of Cushing syndrome. *Pediatrics*. 2016;137:e1–e6.
228. Caimari F, Valassi E, Garbayo P, et al. Cushing's syndrome and pregnancy outcomes: a systematic review of published cases. *Endocrine*. 2017;55:555–563.
229. Brue T, Amodru V, Castinetti F. Management of endocrine disease: management of Cushing's syndrome during pregnancy: solved and unsolved questions. *Eur J Endocrinol*. 2018;178:R259–R266.
230. Jung C, Ho JT, Torpy DJ, et al. A longitudinal study of plasma and urinary cortisol in pregnancy and postpartum. *J Clin Endocrinol Metab*. 2011;96:1533–1540.
231. Lopes LM, Francisco RP, Galletta MA, et al. Determination of nighttime salivary cortisol during pregnancy: comparison with values in non-pregnancy and Cushing's disease. *Pituitary*. 2016;19:30–38.
232. Raff H, Trivedi H. Circadian rhythm of salivary cortisol, plasma cortisol, and plasma ACTH in end-stage renal disease. *Endocrine Connect*. 2013;2:23–31.
233. Raff H, Cohen EP, Findling JW. A commentary on diagnosing Cushing's disease in the context of renal failure. *Eur J Endocrinol*. 2019;181:C9–C11.
234. Cardoso EM, Arregger AL, Budd D, et al. Dynamics of salivary cortisol in chronic kidney disease patients at stages 1 through 4. *Clin Endocrinol*. 2016;85:313–319.
235. Valassi E, Crespo I, Santos A, et al. Clinical consequences of Cushing's syndrome. *Pituitary*. 2012;15:319–329.

8

Clinically Nonfunctioning Sellar Masses

Peter J. Snyder

OUTLINE

Types of Clinically Nonfunctioning Sellar Masses, 134
  Pituitary Adenomas, 134
  Etiology of Pituitary Adenomas, 134
  Other Benign Tumors, 135
  Malignant Tumors, 136
  Pituitary Hyperplasia, 136
  Hypophysitis, 136
  Other Sellar Masses, 136
Clinical Presentations, 137
  Neurologic Presentations, 137
  Incidental Finding, 137
  Endocrinologic Presentations, 137
  Pituitary Apoplexy, 139
Diagnosis, 139
  Utility of Diagnosing a Clinically Nonfunctioning Sellar Mass, 139
  Imaging of the Sellar Region, 139
  Endocrinologic Tests, 139
  Pathologic Evaluation, 140
Treatment, 140
  Surgery, 140
  Radiation, 141
  Pharmacologic Treatment, 142
  Observation, 142

TYPES OF CLINICALLY NONFUNCTIONING SELLAR MASSES

Many types of lesions can present as masses in the region of the sella turcica (Table 8.1). The most common are pituitary adenomas. Other lesions include nonpituitary benign tumors, malignant tumors, hyperplasia of any anterior pituitary cell type, different kinds of hypophysitis, cysts, and infiltrative diseases.

Pituitary Adenomas

Nomenclature—Tumors that arise from cells of the anterior pituitary gland that do not metastasize have been called adenomas. The large size and aggressive behavior of some of these tumors has led to the suggestion that the term adenoma be replaced by the term pituitary neuroendocrine tumor.[1]

Classification of Pituitary Adenomas—Adenomas can arise from any normal anterior pituitary cell type: lactotroph, somatotroph, corticotroph, gonadotroph, and thyrotroph. Any pituitary adenoma type may first be recognized as a sellar mass that is apparently nonfunctioning clinically, even those whose hormonal secretion usually causes a recognizable clinical syndrome, such as corticotroph or somatotroph adenomas. In some situations the clinical syndrome (e.g., Cushing syndrome or acromegaly) is present but not recognized; in other situations, the clinical syndrome is relatively subtle or nonexistent. The latter have been referred to as “silent” adenomas.[2] It is probably more accurate, however, to regard the clinical presentations of these adenomas as a spectrum from clinically obvious (e.g., frank acromegaly), to clinically subtle (e.g., slight increase in ring size but still overall normal appearance), to clinically silent (e.g., no clinical manifestations but high levels of insulin-like growth factor 1 (IGF-1)),[3] to totally silent (e.g., no clinical manifestations and a normal IGF-1 level, but immunostaining of excised tissue for growth hormone) (Table 8.2). Corticotroph adenomas exhibit a similar spectrum.[4]

KEY POINTS

- Gonadotroph adenomas are the most common clinically nonfunctioning pituitary adenoma, but corticotroph and somatotroph adenomas may also be clinically nonfunctioning.

The most common pituitary adenoma that presents as a sellar mass is a gonadotroph adenoma. They comprise 40% to 50% of all macroadenomas and approximately 80% of clinically nonfunctioning adenomas. They are often not recognized as being of gonadotroph cell origin, as they secrete inefficiently, and what they do secrete—intact gonadotropins and their subunits—often does not produce a recognizable clinical syndrome. Consequently, these adenomas are often not recognized until they become so large that they cause neurologic symptoms, or they are recognized incidentally when imaging is performed for an unrelated symptom. Occasionally, however, they do cause recognizable clinical syndromes, as described later under “Clinical Syndromes,” and more often they are recognized by finding abnormal serum concentrations of gonadotropins and their subunits.

Etiology of Pituitary Adenomas

All pituitary adenomas appear to be true neoplasms, arising from a somatic mutation of a single progenitor cell that divides repetitively. The evidence comes from studies that show that virtually all pituitary adenomas are monoclonal, that is, arise from a somatic mutation of a single cell.[5,6]

Mutations of specific genes have been identified in association with hereditary pituitary adenomas that are part of clinical syndromes. In multiple endocrine neoplasia type I (MEN1), a mutation of the *MEN1*

TABLE 8.1 Types of Sellar Masses

- Pituitary Adenomas
  - Gonadotroph
  - Corticotroph
  - Somatotroph
  - Lactotroph
  - Thyrotroph
- Other Benign Tumors
  - Craniopharyngioma
  - Meningioma
  - Pituicytoma
- Malignant Tumors
  - Germ Cell Tumor (Ectopic Pinealoma)
  - Chordoma
  - Lymphoma
  - Metastatic Disease
  - Pituitary Carcinoma
- Pituitary Hyperplasia
  - Lactotroph Hyperplasia of Pregnancy
  - Thyrotroph Hyperplasia owing to Primary Hypothyroidism
  - Gonadotroph Hyperplasia owing to Primary Hypogonadism
- Hypophysitis
  - Lymphocytic
  - Anticytotoxic T Lymphocyte–Associated Antigen Immunotherapy
  - IgG4 (Plasma Cell)
  - Granulomatous
- Other
  - Cysts (Rathke's Cleft, Arachnoid, Dermoid, Epidermoid)
  - Infiltrative and Granulomatous Diseases (Sarcoidosis, Langerhans Cell Histiocytosis, Tuberculoma)
  - Abscess
  - Vascular Abnormalities (Aneurysm, Arteriovenous Fistula)

gene[7] results in decreased expression of the tumor suppressor protein menin and development of adenomas of the pituitary, parathyroids, and pancreas. All pituitary adenoma types can occur in MEN1, most commonly lactotroph and somatotroph adenomas, but gonadotroph adenomas[8,9] and those identified only as clinically nonfunctioning[10,11] have also been reported. Another clinical syndrome, the Carney complex,[12] which often results from germline-inactivating mutations in the gene encoding the regulatory subunit type I of the cyclic adenosine monophosphate (cAMP)-dependent protein kinase A, is associated with somatotroph adenomas but not gonadotroph or other clinically nonfunctioning adenomas.

Pituitary adenomas that are familial but not associated with syndromes are referred to as familial isolated pituitary adenomas (FIPAs).[13] These comprise only about 2% of all patients who have pituitary adenomas. Within some families, the adenomas are of the same type in all family members, and within other families, different family members have different adenomas. Somatotroph and lactotroph adenomas are the most common, but clinically nonfunctioning and gonadotroph adenomas may occur. Most are macroadenomas.

Of FIPAs, 10% to 20% are associated with germline mutations of the aryl hydrocarbon receptor-interacting protein (*AIP*) gene.[13,14] Clinical characteristics more likely to predict an *AIP* mutation include age younger than 18 years and growth hormone hypersecretion leading to gigantism.[14] Screening for the *AIP* mutation in family members of patients with known *AIP* mutations allows detection of previously unrecognized adenomas when they are smaller and therefore less likely to have extended outside of the sella[15] (see also Chapter 14).

TABLE 8.2 Spectrum of Functionality of Pituitary Adenomas

Term	Description
Clinically obvious	Typical physical features of excessive hormonal hypersecretion (e.g., classical acromegaly or Cushing syndrome)
Clinically subtle	Subtle physical features of excessive hormonal hypersecretion (e.g., subtle acromegaly or Cushing syndrome)
Clinically silent	Elevated serum concentration of pituitary hormone (follicle-stimulating hormone, insulin-like growth factor 1, adrenocorticotropic hormone) but not even subtle clinical manifestations
Totally silent	Type of adenoma identifiable only by immunostaining of excised adenoma; normal serum concentration of hormone typically secreted by that cell type

Genetic mutations associated with the development of most isolated pituitary adenomas are not known, but approximately 40% of somatotroph adenomas are associated with mutations of the gene encoding the alpha-subunit of the G stimulatory protein ($G_{s\alpha}$) and, as a consequence, constitutively activate adenylyl cyclase and increase cAMP, which is mitogenic to somatotroph cells, thereby resulting in somatotroph adenomas.[16]

Other Benign Tumors

Craniopharyngiomas. Craniopharyngiomas are solid or mixed solid-cystic epithelial tumors that arise from remnants of Rathke's pouch, either within or above the sella.[17] Two distinct types can be identified: adamantinomatous, which occur either in childhood or adulthood; and papillary, which occur only in adulthood. Adamantinomatous craniopharyngiomas have activating mutations of *CTNNB1*, the gene that encodes β-catenin, and an activated Wnt signalling pathway.[18] Papillary craniopharyngiomas have mutations in the *BRAF* oncogene.[18]

The major presenting symptoms are growth retardation in children and abnormal vision in adults. Hormonal abnormalities are also common, including hyperprolactinemia, deficiencies of anterior pituitary hormones, and diabetes insipidus. Magnetic resonance imaging (MRI) often reveals a heterogeneous signal, and computed tomography (CT) scan often shows calcifications.

Demonstration of the molecular abnormality associated with each type of craniopharyngioma offers the potential of targeted medical treatment.[19] Administration of BRAF inhibitors has already been reported to achieve clinical responses in cases of papillary tumors (see also Chapter 11, where these tumors and related parasellar masses are discussed in more detail).

Meningiomas. These are usually benign and arise from the meninges anywhere within the head. About 20% arise near the sella, causing visual impairment and hormonal deficiencies. By MRI, meningiomas typically emit a low signal on T1-weighted images and a high signal on T2-weighted images and exhibit intense enhancement after gadolinium and a "dural tail," a thickening of the dura adjacent to the meningioma that tapers.

Pituicytomas. Pituicytomas are rare, benign tumors that arise from pituicytes,[26] which are glial cells of the posterior pituitary. They have no hormonal secretory function and can be diagnosed only by the characteristic histologic pattern of elongated cells in bundles and immunostaining for thyroid transcription factor 1.[20]

Malignant Tumors

Some malignant tumors arise within or near the sella, and others metastasize there. Malignancies that arise in the parasellar region include germ cell tumors, sarcomas, chordomas, lymphomas, and pituitary carcinomas.

Germ Cell Tumors (Ectopic Pinealomas). Suprasellar germ cell tumors usually occur through the third decade of life and are histologically and biologically similar to other germ cell tumors, such as germinomas, teratomas, embryonal carcinomas, and choriocarcinomas. They may present with headaches, nausea, vomiting, lethargy (from increased intracranial pressure with pineal lesions), diplopia, paralysis of upward conjugate gaze, hypopituitarism, and diabetes insipidus. If the tumor is in the pineal, imaging will show a mass in the third ventricle; if the tumor is in the infundibulum, it will be thickened. Serum concentrations of human chorionic gonadotropin beta and/or alpha fetoprotein may be increased. Although these lesions are highly malignant and metastasize readily, they are also highly radiosensitive.

Chordomas. Chordomas are slowly growing malignancies that arise from notochord remnants. Chordomas that arise in the clivus, the bone that is the base of the sella turcica, and may present with headaches, visual impairment, and anterior pituitary hormonal deficiencies. MRI often shows a heterogeneous sellar mass associated with osteolytic bony erosion and calcification that may be distinct from the normal pituitary. Histologically, they exhibit markers for epithelial cells, including cytokeratin and vimentin. After surgical excision, local invasion and recurrence commonly occur; mean patient survival is about 5 years. Rarely, they become sarcomatous. A less aggressive subtype is the chondrosarcoma.

Primary Lymphoma. Primary central nervous system lymphoma sometimes involves the pituitary and hypothalamus. A review of 13 patients with primary pituitary lymphoma noted neurologic symptoms, including headaches and visual and oculomotor impairment and/or deficiencies of anterior pituitary hormones and vasopressin.[21] MRI showed a sellar mass with variable extrasellar extension.

Metastatic Disease. Metastases to the hypothalamus and pituitary gland occur most commonly with breast cancer in women and lung cancer in men but are also encountered with other cancers.[22] Symptoms include diabetes insipidus, anterior pituitary dysfunction, visual field defects, retroorbital pain, and ophthalmoplegia. The metastases may be the presenting sign of previously undiscovered malignancy.

Pituitary Carcinoma. Carcinoma arising from anterior pituitary cells is quite rare. When it does occur, the malignancy can arise from any anterior pituitary cell type; lactotroph, somatotroph, corticotroph, thyrotroph, and gonadotroph carcinomas have been reported.[23] Diagnosis is made by finding a distant metastasis (see also Chapter 15).

Pituitary Hyperplasia

Hyperplasia of any anterior pituitary cell type can occur and results in generalized enlargement of the pituitary. The causes are both physiologic and pathologic. During normal pregnancy, increased estradiol secretion causes lactotroph hyperplasia, causing an average increase of about 2.5 mm in each dimension.[24] When a target gland fails, the resulting lack of feedback inhibition on the corresponding pituitary trophic hormone-secreting cells causes hyperplasia. For example, longstanding, untreated primary hypothyroidism causes thyrotroph hyperplasia and symmetrical pituitary enlargement, and treatment with thyroxine reverses the enlargement.[25] Similarly, longstanding, untreated primary hypogonadism causes gonadotroph hyperplasia. A neuroendocrine tumor that secretes growth hormone–releasing hormone causes somatotroph hyperplasia.[26]

Hypophysitis

Several types of hypophysitis have been recognized.[27] Most types are characterized by diffuse enlargement of the pituitary, and often the stalk, that regresses within months of the diagnosis, as well as hypopituitarism that is likely to be permanent. In some series, hypoadrenalism is the most common hormonal deficiency, but in other series other hormonal deficiencies are more common.[28]

One type of hypophysitis is lymphocytic, which sometimes occurs in late pregnancy or postpartum but can also be seen in women who are not pregnant, as well as infrequently in men.[29] It is characterized by headaches whose intensity is out of proportion to the size of the lesion. Another type is caused by immune checkpoint inhibitors, especially ipilimumab, used to treat malignant melanoma and other malignancies.[30] Hypophysitis also occurs because of infiltration of plasma cells, which produce immunoglobulin G4 (IgG4). This type may occur in association with IgG4 inflammation of other organs, most notably the pancreas, causing pancreatitis. Although this condition is not common, it was detected by screening in seven of 170 consecutive patients with hypopituitarism.[31] Granulomatous hypophysitis has been reported in association with systemic granulomatous conditions, such as Wegener granulomatosis (also known as "granulomatosis with polyangiitis") and tuberculosis.

Other Sellar Masses

Rathke's Cleft and Other Cysts. Remnants of the embryonic Rathke's pouch that persist may enlarge during adulthood and present as a sellar mass. These may present with neurologic findings, such as headaches and visual disturbances, or with endocrinologic abnormalities, such as hyperprolactinemia, hypopituitarism, and diabetes insipidus. The majority of Rathke's cleft cysts, however, are small and do not increase in size during observation.[32,33] Other sellar cysts include arachnoid, epidermoid, and dermoid cysts.

Infiltrative and Granulomatous Diseases. Infiltrative diseases, such as sarcoidosis[34] and Langerhans cell histiocytosis,[35] may cause a sellar or hypothalamic mass and/or widening of the pituitary stalk. Tuberculomas of the sellar region may occur as part of systemic infection or may be isolated. Because these lesions primarily affect the hypothalamus and infundibulum, diabetes insipidus is common, whereas anterior pituitary hormone deficiencies are less frequent. Infiltrative sarcoidosis of this area occurs in most patients with central nervous system sarcoid involvement.[38] Typically these patients present with varying degrees of anterior pituitary failure, with or without diabetes insipidus.

Abscess. Pituitary abscesses, which are rare, can occur in a normal or diseased pituitary gland. In immunocompromised subjects, they may be caused by fungi (*Aspergillus, Nocardia,* or *Candida albicans*) or *Pneumocystis carinii.* In a series of 24 patients, 16 presented with symptoms and physical findings consistent with a pituitary mass, whereas only eight had features suggestive of infection, such as fever, leukocytosis, or meningismus.[36] MRI is usually unable to distinguish between pituitary abscess and pituitary adenoma, so most patients are diagnosed at the time of surgical exploration.

Vascular Abnormalities. Aneurysms and arteriovenous fistulas can present as sellar masses. Because they arise from the cavernous sinus, they may cause cranial nerve palsies, eye pain, and headaches. They can usually be recognized by MRI, but a highly vascular meningioma may be confused with an aneurysm. CT or magnetic resonance angiography is essential when these are suspected to avoid inappropriate—and dangerous—surgery.

CLINICAL PRESENTATIONS

Clinically nonfunctioning sellar masses, by definition, do not cause florid syndromes of hormonal excess, and as a result often grow unrecognized until they become so large as to cause neurologic symptoms or symptoms of hormonal deficiencies. An equally common presentation is as an incidental finding when imaging is performed for an unrelated reason, such as head trauma. Least common is hormonal hypersecretion by a pituitary adenoma, although not associated with an obvious clinical syndrome. These may be called clinically subtle if there are mild clinical manifestations, or clinically silent if there are no clinical manifestations at all but the adenoma can be recognized biochemically (Table 8.2). Hormonal hyposecretion is common, but usually not the presenting manifestation. Pituitary apoplexy is a rare presentation.

KEY POINTS

- Clinically nonfunctioning pituitary adenomas are as likely to come to medical attention as an incidental finding as a manifestation of the adenoma.

Neurologic Presentations

Visual Abnormalities. A mass within or near the sella that stretches the optic chiasm or nerves sufficiently may result in visual impairment. Masses that arise below the optic chiasm, such as pituitary adenomas, may extend superiorly and elevate and then stretch the chiasm sufficiently to cause visual field abnormalities, initially of the upper, outer quadrants (bilateral superior quadrantinopia) and then of the outer halves (bilateral hemianopia), before much later affecting central vision and visual acuity. Masses that arise above the chiasm may first cause asymmetric impairment of inferior visual fields. They may be reversible once the masses have been partially or completely excised.

Oculomotor Abnormalities. Masses that arise within the sella and extend into a cavernous sinus may affect one or more of the oculomotor nerves on that side, resulting in diplopia, impaired extraoculomotor movements, and/or ptosis.

Other. Other neurologic manifestations of sellar masses are headaches, cerebrospinal fluid leakage and meningitis, and hydrocephalus. Any kind of sellar mass can cause a headache; those that grow more rapidly are more likely to do so. Cerebrospinal fluid (CSF) rhinorrhea is uncommon. When it does occur, it is the result of an aggressive pituitary adenoma that extends inferiorly and erodes the cribriform plate, allowing leakage of CSF, which in turn predisposes to retrograde infection and meningitis. Hydrocephalus is also uncommon. When it occurs it is often the result of a suprasellar lesion that obstructs the fourth ventricle.

Incidental Finding

As a result of the ready availability of imaging techniques in developed countries, sellar masses are often discovered when imaging is performed for an unrelated reason. Although many lesions discovered in this manner are small, others are larger than 1 cm in diameter, and some are even associated with hormonal hyper- and hyposecretion and with neurologic abnormalities that had not been previously recognized. In a prospective study of 269 patients referred for evaluation of a pituitary adenoma, the adenoma was detected incidentally in 131 (48.7%), of which 111 were macroadenomas (Table 8.3).[37] The reasons for the imaging were legion; the most common were head trauma and neurologic and ear symptoms.

TABLE 8.3 Presentations of 269 Clinically Nonfunctioning Pituitary Adenomas

- Incidental Finding ($n = 131$)
  - Head and neck symptoms
  - Eye symptoms
  - Neurologic symptoms
  - Other
- Symptoms Resulting from the Adenoma ($n = 138$)
  - Mass effect
    - Visual abnormality
    - Headache
  - Pituitary hormonal dysfunction
    - Panhypopituitarism
    - Hypoadrenalism
    - Hypothyroidism
    - Hypogonadism

Modified from Freda PU, Bruce JN, Khandji AG, et al. Presenting features in 269 patients with clinically nonfunctioning pituitary adenomas enrolled in a prospective study. *J Endocr Soc*. 2020;4:bvaa021.

TABLE 8.4 Hormonal Criteria for the Diagnosis of Gonadotroph Adenomas

Men	Women
Supranormal Basal Serum Concentrations of	
FSH	FSH but not LH
α LHβ, or FSHβ subunits	Any subunit relative to intact FSH and intact LH
LH and testosterone	FSH and estradiol (in premenopausal women)
Supranormal Response to TRH of	
FSH	FSH
LH	LH
LHβ (most common)	LHβ (most common)

FSH, Follicle-stimulating hormone; *LH*, luteinizing hormone; *TRH*, thyrotropin-releasing hormone.

Endocrinologic Presentations

Hormonal Excess: Gonadotroph Adenomas

Serum Concentrations of Gonadotropins and Their Subunits. Although gonadotroph adenomas are typically nonfunctioning clinically because they secrete inefficiently and because their secretory products—intact gonadotropins and their subunits—usually do not cause a clinical syndrome, they often can be identified by their basal and stimulated secretory products (Table 8.4). Gonadotroph adenomas often secrete sufficient intact follicle-stimulating hormone (FSH) to result in a supranormal serum FSH concentration from minimal to ten times the upper limit of normal.[38] In contrast, gonadotroph adenomas rarely produce supranormal serum concentrations of intact luteinizing hormone (LH), but when they do, the serum testosterone concentration is elevated.[39] A small percentage of patients who have gonadotroph adenomas have supranormal basal serum concentrations of the common α-subunit.[38] Administration of synthetic thyrotropin-releasing hormone (TRH) to patients who have gonadotroph adenomas often produces an increase in the serum concentrations of intact gonadotropins and their subunits, especially of the LHβ subunit.[38]

KEY POINTS

- A gonadotroph adenoma can often be recognized *in vivo* by a supranormal concentration of follicle-stimulating hormone.

Clinical Syndromes. Although gonadotroph adenomas usually do not result in recognizable clinical syndromes, uncommonly they may cause menstrual irregularities and ovarian hyperstimulation in premenopausal women, testicular enlargement in men, and precocious puberty in children. They are also subject to apoplexy when synthetic TRH is administered for diagnosis or synthetic gonadotropin-releasing hormone (GnRH) or a GnRH analog is administered for treatment[40] (Table 8.5).

Menstrual irregularities and ovarian hyperstimulation result from continuous secretion of intact FSH in a premenopausal woman.[40,41] Continuous secretion of FSH, in contrast to cyclical secretion by normal gonadotroph cells, results in continuous stimulation of the ovaries, with the result that the ovaries become very large because of multiple, large cysts. Continuous stimulation of the ovaries also results in continuous secretion of estradiol, so ovulation does not occur. The resulting clinical presentation is usually oligomenorrhea and infertility. Pelvic ultrasound shows multiple large cysts and a widened endometrial "stripe" (Fig. 8.1). These patients typically have elevated serum FSH, suppressed LH, and elevated α-subunit and estradiol.[40]

TABLE 8.5 Clinical Syndromes Associated With Gonadotroph Adenomas

Ovarian hyperstimulation
Pituitary apoplexy following administration of thyrotropin-releasing hormone, gonadotropin-releasing hormone, or a gonadotropin-releasing hormone analog
Large testicular size
Premature puberty in a prepubertal boy

The estradiol concentration is often higher than 500 pg/mL and sometimes as high as 2000 pg/mL. The clinical picture of oligomenorrhea and multiple ovarian cysts can be mistaken for polycystic ovarian syndrome, but administration of a superactive GnRH analog to a patient with a gonadotroph adenoma results in increased, rather than decreased, FSH secretion and ovarian size and function.[42] Excision of the gonadotroph adenoma can lead to restoration of normal gonadotropin secretion, ovarian function, and pregnancy.[42]

Other clinical presentations of gonadotroph adenomas are even less common (Table 8.5). In men, testicular size may be large.[39,40,43] In boys and girls, isosexual precocity may occur.[40] Administration of TRH diagnostically or a GnRH analog therapeutically, such as to treat prostate cancer, in a patient with a previously unrecognized gonadotroph adenoma can cause pituitary apoplexy.[44]

Hormonal Excess: Other Pituitary Adenomas. Although somatotroph and corticotroph adenomas often secrete efficiently and therefore often result in the classic clinical syndromes of acromegaly and Cushing syndrome, some somatotroph and corticotroph adenomas secrete growth hormone and adrenocorticotropic hormone (ACTH) excessively but do not cause the corresponding clinical syndrome classically or even subtly. These can be called silent adenomas (Table 8.2). Somatotroph adenomas that cause an elevated serum concentration of IGF-1 but not even subtle features of acromegaly can be termed clinically silent, and those that do not elevate IGF-1 can be called totally silent.[3] Similarly, corticotroph adenomas that cause an elevated plasma ACTH concentration and/or elevated urine cortisol excretion can be termed clinically silent, and those that cause neither hormonal abnormality can be called totally silent.[4] In a retrospective study of 100 consecutive surgically-excised pituitary macroadenomas, 24 were somatotroph adenomas, and, of these, eight were clinically silent.[3] In a prospective study of 124 patients with pituitary macroadenomas, 20 were corticotroph adenomas.[4] Of these, eight were clinically silent, and five were totally silent.

Hormonal Deficiencies. The large size of many sellar masses commonly causes hormonal hyposecretion because of compression of the pituitary or stalk, but these deficiencies often do not prompt the patient to seek medical attention. In the prospective study of clinically nonfunctioning pituitary adenomas described above, of the 131 patients

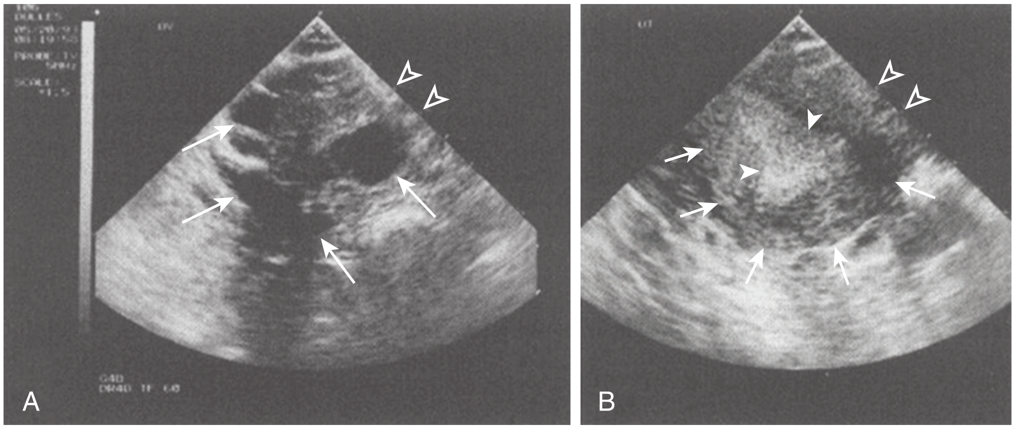


Fig. 8.1 Ultrasound of **A**, ovaries and **B**, uterus in a 39-year-old woman who had a gonadotroph adenoma secreting follicle-stimulating hormone and causing ovarian hyperstimulation. In **A**, the *closed arrowheads* indicate large ovarian cysts. In **B**, the *arrows* indicate the uterus, and the *closed arrowheads* indicate the thickened endometrium. The distance between the *open arrowheads* is 1 cm. (From Djerassi A, Coutifaris C, West VA, et al. Gonadotroph adenoma in a premenopausal woman secreting follicle-stimulating hormone and causing ovarian hyperstimulation. *J Clin Endocrinol Metab*. 1995;80:591-594.)

whose adenomas were detected incidentally, 30 were hypogonadal, 13 hypothyroid, and nine hypoadrenal.[37]

Pituitary Apoplexy

Sudden bleeding into and infarction of the pituitary, called pituitary apoplexy, is a dramatic presentation of an adenoma.[45] It is an uncommon event; its incidence in one study was 0.17 episodes/100,000 persons per year.[46] It may occur in patients known to have an adenoma, but it may be the event that brings the adenoma to medical attention. Precipitating factors include initiation of anticoagulant treatment and administration of TRH diagnostically and a GnRH analog therapeutically.[45] The most common symptom is the sudden onset of severe headache. Visual symptoms include oculomotor palsies because of sudden expansion of the pituitary into a cavernous sinus and visual field deficits because of elevation of the optic chiasm. Hypopituitarism is common, including the sudden onset of cortisol deficiency, which may be life-threatening if not recognized and treated immediately.

DIAGNOSIS

Establishing the diagnosis of a clinically silent sellar mass may proceed from recognizing that a patient's visual abnormality or other neurologic symptom could be caused by a sellar lesion, to confirming the presence of the lesion by an imaging procedure, to attempting to characterize the lesion by its hormonal features. If the sellar mass was found incidentally when imaging was performed because of symptoms unrelated to the mass, the next step is to attempt to characterize the lesion by its hormonal features.

Utility of Diagnosing a Clinically Nonfunctioning Sellar Mass

Identifying the specific type of sellar mass is desirable to distinguish a pituitary adenoma from other lesions and, if a pituitary adenoma, to provide a marker by which to monitor the response to treatment. Identifying the lesion as being of pituitary rather than nonpituitary origin is of value because it may influence treatment. If surgery is indicated, for example, the surgical approach to a pituitary adenoma may be different from that for another lesion. If the lesion is a somatotroph adenoma, identified by an elevated serum concentration of IGF-1, medical treatment may be an option. Finding a tumor marker such as elevated IGF-1, characteristic of a somatotroph adenoma, or elevated serum FSH, characteristic of a gonadotroph adenoma, not only identifies the lesion as being of somatotroph or gonadotroph origin but also provides a means by which to follow the response to treatment. For example, when the serum FSH concentration is elevated before surgery, the decrease after surgery correlates with reduction in adenoma mass seen by imaging.[47]

Imaging of the Sellar Region

MRI is the best imaging technique for the sellar region. Pituitary adenomas tend to be relatively symmetrical, although there are notable exceptions. They enhance on T1-weighted images following the administration of gadolinium, although to a lesser degree than normal pituitary tissue (Fig. 8.2). Several MRI features suggest a greater likelihood of a pituitary adenoma or of another type of lesion, but no feature is pathognomonic. A lesion that appears to be entirely cystic, for example, suggests a lesion such as a Rathke's cleft cyst, but occasionally a pituitary adenoma appears entirely cystic by MRI. A sellar mass that is irregular in shape and emits a heterogeneous signal is more likely to be a craniopharyngioma, but a pituitary adenoma can present similarly. A converse example is an invasive lesion of the clivus, which is typical of a chordoma, but some pituitary adenomas extend primarily into the clivus and therefore can be mistaken for a chordoma. Because of the limitations of imaging in distinguishing types of sellar lesions, endocrinologic testing is invariably necessary.

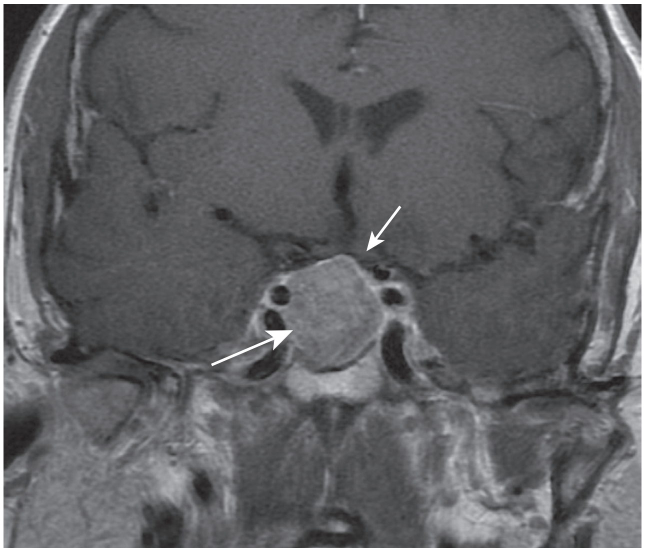


Fig. 8.2 Magnetic resonance image of a sellar mass (*larger arrow*) showing a brighter signal than the surrounding central nervous system tissue after administration of gadolinium. The mass slightly elevates the optic chiasm (*smaller arrow*). The mass was identified as a gonadotroph adenoma when the excised tissue showed strong immunostaining for the transcription factor steroidogenic factor 1 (Fig. 8.3).

Endocrinologic Tests

Sellar mass lesions should be evaluated by measurement of the serum concentrations of pituitary hormones and related target gland hormones to determine if any abnormality suggests that the lesion is a specific type of pituitary adenoma. A prolactin concentration above 100 ng/mL, and especially above 200 ng/mL, suggests a lactotroph adenoma. An elevated IGF-1 concentration suggests a somatotroph adenoma even if the patient does not exhibit clinical features of acromegaly.[3] An elevated plasma ACTH and/or 24-hour urine cortisol suggests a corticotroph adenoma, even if the patient does not exhibit features of hypercortisolemia.[4] An elevated serum T4 associated with a thyroid-stimulating hormone (TSH) concentration that is not suppressed suggests a thyrotroph adenoma (see Chapter 9).

Suspicion that the lesion is a gonadotroph adenoma depends on specific combinations of intact gonadotropins and their subunits (Table 8.4). The combinations differ somewhat in men and women. In a man who has a pituitary macroadenoma, elevated basal serum concentrations of intact gonadotropins and/or their subunits alone or in combination with responses of any of these to TRH is strong evidence that the adenoma is of gonadotroph origin.[38] An elevated basal FSH concentration is common.[38,43] In a woman of postmenopausal age, elevated basal serum concentrations of intact FSH or gonadotropin subunits are usually of little diagnostic value, because either the adenoma or the nonadenomatous postmenopausal gonadotroph cells could be the source. However, a gonadotroph adenoma is likely if intact FSH is markedly elevated but LH is not at all elevated, or if one of the gonadotropin subunits is distinctly elevated but intact FSH and LH are not elevated.[48] In a woman of premenopausal age, ovarian hyperstimulation, including elevated serum estradiol concentration, as discussed earlier, elevated FSH out of proportion to LH levels, or an elevated basal α-subunit concentration all point to the gonadotroph nature of the sellar mass.[40]

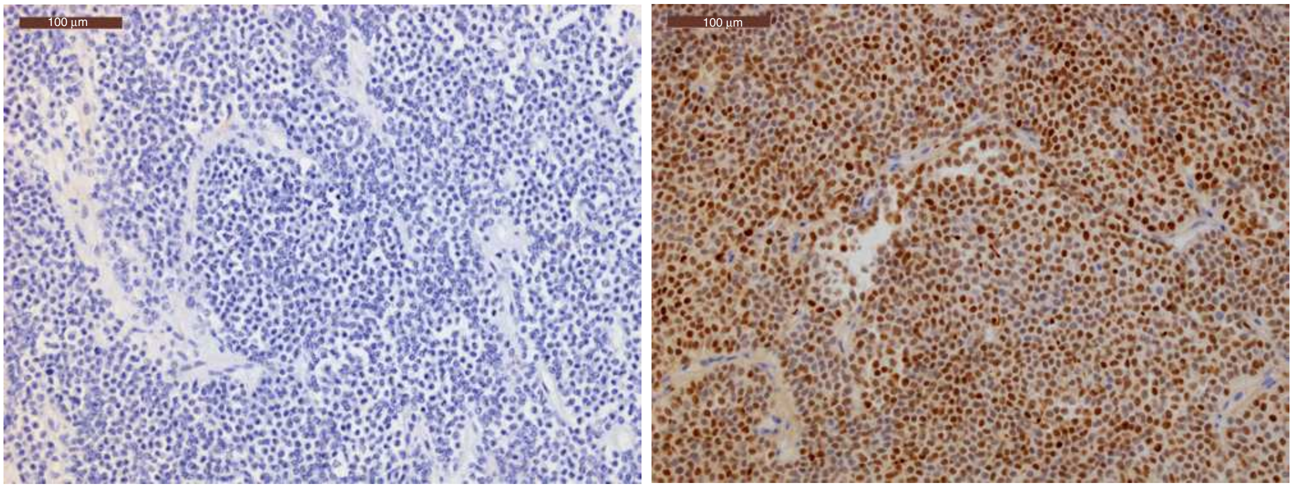


Fig. 8.3 Immunostaining of the excised pituitary adenoma shown by magnetic resonance imaging in Fig. 8.2. Immunostaining for follicle-stimulating hormone β (*left panel*) is negative, but immunostaining for the transcription factor steroidogenic factor 1 (*right panel*) identifies it as a gonadotroph adenoma. Note that the adenoma cells stain uniformly and strongly, but the endothelial cells along the blood vessels do not stain. (Courtesy of MacLean Nasrallah, MD, PhD.)

Gonadotroph adenomas can usually be readily distinguished from pituitary enlargement because of gonadotroph hyperplasia that results from longstanding primary hypogonadism. In primary hypogonadism, the pituitary enlargement is not as prominent, and both gonadotropins are elevated.

Pathologic Evaluation

Even when the identity of a sellar mass cannot be determined *in vivo*, pathologic examination of excised tissue can usually make this determination. By hematoxylin-eosin staining, pituitary adenoma cells do not exhibit the normal pituitary glandular pattern but are arranged in cords or sheets, sometimes interspersed with varying amounts of fibrous tissue. In an adenoma, cells are usually homogeneously similar in size, but cell size varies considerably among adenomas. Immunostaining for pituitary hormones (growth hormone, ACTH, prolactin, FSHβ, LHβ and TSHβ, and α-subunit) often identifies the adenoma type. Most adenomas that do not stain for any pituitary hormones and were previously called "null cell" adenomas can be identified by staining for the transcription factors associated with the differentiation of specific pituitary cell types. Most notably, adenomas previously classified as "null cell" by staining for FSHβ and LHβ can now be identified as gonadotroph adenomas by staining for the transcription factor steroidogenic factor 1 (SF-1) (Fig. 8.3). In a study of 516 consecutive adenomas, 119 could not be identified by hormone immunostaining, but 79 of these could be identified as gonadotroph adenomas by staining for SF-1, 32 as corticotroph adenomas by staining for T-box pituitary restricted transcription factor (T-Pit), and two as being of somatomammotroph lineage by staining for Pit1 (Fig. 8.4).[49]

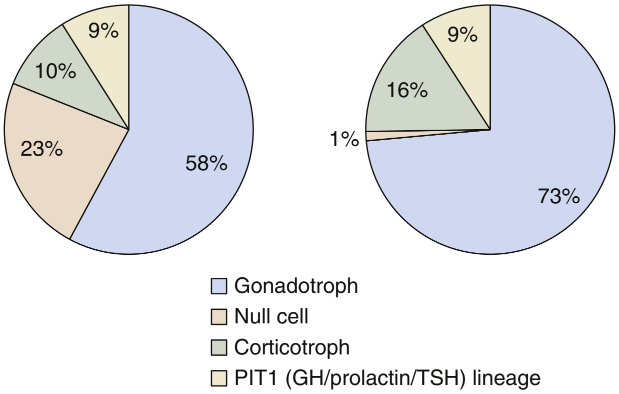


Fig. 8.4 Improved identification of pituitary adenomas by immunostaining for transcription factors compared to staining for pituitary hormones. By immunostaining for pituitary hormones (left), the cell type of 23% of 516 excised pituitary adenomas could not be identified; they were therefore classified as null cell. By immunostaining for pituitary transcription factors (right), the cell type of only 1% could not be identified. *GH,* Growth hormone; *TSH*, thyroid-stimulating hormone. (Figure from Drummond J, Roncaroli F, Grossman AB, et al. Clinical and pathological aspects of silent pituitary adenomas. *J Clin Endocrinol Metab.* 2019;104:2473-89; data from Nishioka H, Inoshita N, Mete O, et al. The complementary role of transcription factors in the accurate diagnosis of clinically nonfunctioning pituitary adenomas. *Endocr Pathol.* 2015;26:349-55.)

TREATMENT

When a clinically nonfunctioning sellar mass is not detected until it becomes so large that it causes significant visual impairment, treatment should reduce the size of the mass as soon as possible. Transsphenoidal surgery is the only treatment that meets this criterion. If substantial residual adenoma tissue remains after surgery, radiation is usually effective in preventing regrowth. Several pharmacologic treatments have been tried, but none reduces the size of clinically nonfunctioning adenomas reliably. Observation alone is a satisfactory approach in the absence of neurologic symptoms.

Surgery

Surgical Approaches. Transsphenoidal surgery using an operating microscope was introduced in the 1970s. Since the early 2000s, transsphenoidal surgery using an endoscope has been used increasingly because it affords a wider, more panoramic field of vision, better illumination, and the ability to see around corners. These advantages are particularly helpful in surgery of clinically nonfunctioning pituitary adenomas, which tend to be large and extend beyond the sella.[50] The endoscopic approach also facilitates surgery on large adenomas, because it allows removal of a piece of the clivus for an "extended" approach beyond the sella. Although few studies have compared the two techniques directly, retrospective series and metaanalyses show that the endoscopic technique results in achieving gross total removal more often and in leaving less residual when the adenoma extends beyond the sella.[50]

KEY POINTS

- Transsphenoidal surgery by a very experienced pituitary surgeon has a high likelihood of removing a large amount of a clinically nonfunctioning pituitary adenoma and a low likelihood of serious adverse events.

Efficacy of Surgery. Efficacy of surgery for a clinically nonfunctioning pituitary adenoma can be judged by the amount of the adenoma tissue excised and the amount remaining, the reduction in serum concentrations of secretory products that were elevated before surgery, improvement in vision, and restoration of pituitary function that was subnormal before surgery. Complete resection of adenoma tissue is generally lower the larger the lesion and the degree of extension outside of the sella, especially into the cavernous sinuses. In a prospective, multicenter study of 260 patients with clinically nonfunctioning adenomas with an average diameter of 2.5 cm, gross total resection was achieved in approximately 80% by both the microscopic and endoscopic techniques.[51] Gross total resection was somewhat less, approximately 70%, when the average diameter was greater than 3.0 cm and approximately 40% (by the microscopic technique) or 60% (by the endoscopic technique) when the adenoma extended into a cavernous sinus, as judged by Knosp grades 3 and 4.[51] Vision that was impaired before surgery improves to some extent in the majority of patients following surgery. Visual fields are more likely to improve the younger the patient and the shorter the duration of visual symptoms, as well as in the preoperative absence of central field involvement or abnormal visual acuity.[52-54] Hormonal function that was abnormal before surgery may recover afterwards in a substantial minority of patients.[55-58] The chance of recovery by axis varies among studies. Recovery has been reported as late as a year or more after surgery.[57,58]

Complications of Surgery. Serious complications of transsphenoidal surgery, whether endoscopic or microscopic, include carotid artery injury, central nervous system injury, loss of vision, ophthalmoplegia, CSF leak and meningitis, and death. Other complications include diabetes insipidus, which can be transient or permanent, syndrome of inappropriate antidiuretic hormone secretion (SIADH), and new anterior pituitary hormone deficiencies. Serious side effects are uncommon when the procedures are performed by surgeons who have great experience. In the study comparing the results of endoscopic and microscopic surgery, in which all participating neurosurgeons had great experience in pituitary surgery, each complication occurred in less than 1% of patients in both groups.[51] The incidence of serious complications is greater, however, when surgery is performed by neurosurgeons with less experience. In a survey taken before the endoscopic technique was widely used, neurosurgeons were asked to report their own experience with microscopic surgery.[59] Neurosurgeons who reported performing fewer than 200 total procedures reported approximately five to six times as many serious complications as neurosurgeons who reported performing more than 500 procedures.

Evaluation of the Results of Surgery. The effects of surgery on vasopressin secretion, anterior pituitary function, residual adenoma volume, and visual function should be evaluated. If diabetes insipidus occurs, it usually will be within 24 to 48 hours and may be transient or permanent. Whether or not diabetes insipidus occurs, SIADH may occur approximately 1 week after surgery. It is possible, although not common, for diabetes insipidus shortly after surgery to be followed by SIADH 1 week afterwards, and for that to be followed by permanent diabetes insipidus (see also Chapter 12).

Any hormone that was hypersecreted before surgery, such as FSH, ACTH, or IGF-1, can be used as a marker to judge the extent of removal of the adenoma. The possibility of anterior pituitary hormone deficiencies should be evaluated because deficiencies before surgery can improve afterwards, but surgery can also cause new deficiencies. This testing should be performed first about a month after surgery, but deficiencies before surgery have been reported to improve up to a year or more afterwards.

The anatomic effects of surgery should be evaluated by MRI at least 1 month afterwards. If visual fields were abnormal before surgery, they should be assessed afterwards.

Radiation

Radiation therapy is used primarily to prevent growth of residual adenoma tissue following surgery, especially when a considerable amount of tissue remains and/or the risk of growth appears to be high. In a retrospective study of 132 patients with clinically nonfunctioning pituitary adenomas treated by transsphenoidal surgery and then observed for a mean of 7.6 years afterwards, recurrence occurred in 10.7% of those who had no residual adenoma visible, 38.8% of those who had intrasellar residual adenoma, and 66.7% of those with residual extrasellar extension.[60] Radiation is not commonly used as primary therapy because a decrease in size may take years.

Types of Radiation. Radiation therapy has been used to treat pituitary adenomas for decades. Current techniques employ radiation from one of three sources: high-energy x-rays (photons) from a linear accelerator, gamma radiation from a ^{60}Co source, and protons from a cyclotron. Radiation from all three sources can be administered as a single high dose (18–20 Gy) or in multiple small doses over 5 to 6 weeks for a total dose of 45 to 50 Gy. Single-dose techniques are often referred to as "stereotactic radiosurgery," which is a misnomer as there is no surgery. All of the techniques use computer-generated models based on MRI so that the radiation conforms to boundaries of the lesion.

The most important factors in choosing between a single, high dose and multiple, fractionated low doses are the distance of the lesion from the optic pathways and the size of the adenoma tissue to be radiated. Adenoma tissue that is not at least 3 to 5 mm from the optic pathway should be treated by fractionated radiation to decrease the risk of blindness. Adenomas that are very large, (e.g., >3 cm in diameter), should also be treated by fractionated radiation to reduce the total dose and decrease the risk of blindness. The choice of the radiation source is based on availability.

Efficacy of Radiation (See also Chapter 16). All modalities of radiation therapy currently employed have been shown to be effective in preventing further growth of clinically nonfunctioning pituitary adenomas. The relative efficacy of the different modalities cannot be compared directly, however, because published studies report on a single modality. In series of patients treated with fractionated radiotherapy using a linear accelerator[61,62] or proton beam[63] and observed for a median of 40 months or more, adenoma size was reduced or stable in 90% to 100%. In a review of 25 studies involving 1621 patients, of whom 452 had clinically nonfunctioning pituitary adenomas and were treated with single-dose radiation from a linear accelerator, gamma source, or proton beam, adenoma size was controlled in about 90%

during follow-up of more than 40 months.[64] In a report from a single center at which 100 clinically nonfunctioning adenomas were treated with single-dose radiation from a gamma source and followed for a median of 45 months, adenoma volume decreased or remained stable in 92% of patients.[65]

Complications of Radiation. Although current radiotherapy techniques offer the theoretical advantage of targeting the tumor by stereotactic means, new deficiencies of anterior pituitary hormones and neurologic complications occur following these treatments. In the series of 100 patients with clinically nonfunctioning adenomas treated with single-dose gamma radiation and followed for a median of 45 months, 20% developed new hypopituitarism.[65] In the review of 35 studies involving 1621 patients, optic neuropathy occurred in about 1%, other cranial neuropathies in 1.3%, and parenchymal brain damage in about 0.8%.[64]

Management of Patients After Radiation. Hormonal evaluation, both for excessive secretion of the hormones that were secreted excessively by the adenoma before treatment and for anterior pituitary hormone deficiencies, should be performed 6 and 12 months after radiation and once a year thereafter. Evaluation of adenoma size by MRI should be performed 1 year after radiation; and if the mass is smaller, less frequently thereafter. Neuroophthalmologic evaluation should be repeated after radiation if findings were abnormal before.

Pharmacologic Treatment

Clinically nonfunctioning pituitary adenomas that are associated with elevated serum concentrations of prolactin or growth hormone can be treated with the pharmacologic treatments typically used for those adenomas.

Many clinically nonfunctioning pituitary adenomas, largely gonadotroph adenomas, have been shown to express dopamine receptors,[66] and some studies suggest that dopamine agonist treatment may be useful to treat residual adenoma tissue following removal by surgery. One study showed that cabergoline treatment reduced residual adenoma size in five of nine patients 1 year following surgery.[66] Another study compared regrowth of clinically nonfunctioning adenomas at one hospital that administered dopamine agonist treatment and a nearby hospital that did not.[67] In 55 patients treated with a dopamine agonist beginning soon after surgery, residual adenoma size decreased in a mean follow-up of 8.8 years in 38% of patients but in none of the 60 patients who were not treated. During the same follow-up period, adenoma size increased in 13% of the treated patients and in 47% of the untreated patients.

Gonadotroph adenomas also express somatostatin receptors, perhaps the subtype 3 most abundantly,[68] but treatment of patients who have residual gonadotroph adenoma tissue after surgery does not have clear effects on adenoma size.[69,70]

Agonist analogs of GnRH produce either an agonist effect or no effect on secretion and have no effect on adenoma size.[71,72] Treatment with the GnRH antagonist Nal-Glu GnRH for 6 months suppressed FSH but did not reduce adenoma size.[73]

Observation

Observation alone is a reasonable course for a patient who has a clinically nonfunctioning sellar mass, even one greater than 1 cm, as long as the mass is not associated with neurologic symptoms. If observation is chosen, however, the patient should be followed carefully for possible enlargement of the mass and development of visual impairment. In several retrospective studies of the natural history of clinically nonfunctioning sellar masses greater than 1 cm, increases in size occurred in 20% to 50% of patients in an average of 4 to 5 years.[60,74-76] Patients should also be followed for a possible change in hormonal function, as an increase in size could lead to decreased hormonal secretion, and an adenoma that did not cause detectable hormonal hypersecretion on initial evaluation could do so subsequently.

REFERENCES

1. Asa SL, Casar-Borota O, Chanson P, et al. From pituitary adenoma to pituitary neuroendocrine tumor (PitNET): an International Pituitary Pathology Club proposal. *Endocr Relat Cancer*. 2017;24:C5–C8.
2. Horvath E, Kovacs K, Killinger DW, et al. Silent corticotropic adenomas of the human pituitary gland: a histologic, immunocytologic, and ultrastructural study. *Am J Pathol*. 1980;98:617–638.
3. Wade AN, Baccon J, Grady MS, et al. Clinically silent somatotroph adenomas are common. *Eur J Endocrinol*. 2011;165:39–44.
4. Guttenberg KB, Mayson SE, Sawan C, et al. Prevalence of clinically silent corticotroph macroadenomas. *Clin Endocrinol (Oxf)*. 2016;85:874–880.
5. Alexander JM, Biller BM, Bikkal H, et al. Clinically nonfunctioning pituitary tumors are monoclonal in origin. *J Clin Invest*. 1990;86:336–340.
6. Herman V, Fagin J, Gonsky R, et al. Clonal origin of pituitary adenomas. *J Clin Endocrinol Metab*. 1990;71:1427–1433.
7. Chandrasekharappa SC, Guru SC, Manickam P, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science*. 1997;276:404–407.
8. Benito M, Asa SL, Livolsi VA, et al. Gonadotroph tumor associated with multiple endocrine neoplasia type 1. *J Clin Endocrinol Metab*. 2005;90:570–574.
9. Livshits A, Kravarusic J, Chuang E, et al. Pituitary tumors in MEN1: do not be misled by borderline elevated prolactin levels. *Pituitary*. 2016;19:601–604.
10. Cohen-Cohen S, Brown DA, Himes BT, et al. Pituitary adenomas in the setting of multiple endocrine neoplasia type 1: a single-institution experience. *J Neurosurg*. 2020:1–7.
11. Verges B, Boureille F, Goudet P, et al. Pituitary disease in MEN type 1 (MEN1): data from the France-Belgium MEN1 multicenter study. *J Clin Endocrinol Metab*. 2002;87:457–465.
12. Correa R, Salpea P, Stratakis CA. Carney complex: an update. *Eur J Endocrinol*. 2015;173:M85–M97.
13. Beckers A, Aaltonen LA, Daly AF, et al. Familial isolated pituitary adenomas (FIPA) and the pituitary adenoma predisposition due to mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene. *Endocr Rev*. 2013;34:239–277.
14. Caimari F, Hernandez-Ramirez LC, Dang MN, et al. Risk category system to identify pituitary adenoma patients with AIP mutations. *J Med Genet*. 2018;55:254–260.
15. Marques P, Caimari F, Hernandez-Ramirez LC, et al. Significant benefits of AIP testing and clinical screening in familial isolated and young-onset pituitary tumors. *J Clin Endocrinol Metab*. 2020;105.
16. Spada A, Arosio M, Bochicchio D, et al. Clinical, biochemical, and morphological correlates in patients bearing growth hormone-secreting pituitary tumors with or without constitutively active adenylyl cyclase. *J Clin Endocrinol Metab*. 1990;71:1421–1426.
17. Muller HL, Merchant TE, Warmuth-Metz M, et al. Craniopharyngioma. *Nat Rev Dis Primers*. 2019;5:75.
18. Brastianos PK, Taylor-Weiner A, Manley PE, et al. Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nat Genet*. 2014;46:161–165.
19. Alexandraki KI, Kaltsas GA, Karavitaki N, et al. The medical therapy of craniopharyngiomas: the way ahead. *J Clin Endocrinol Metab*. 2019;104:5751–5764.
20. Viaene AN, Lee EB, Rosenbaum JN, et al. Histologic, immunohistochemical, and molecular features of pituicytomas and atypical pituicytomas. *Acta Neuropathol Commun*. 2019;7:69.
21. Giustina A, Gola M, Doga M, et al. Clinical review 136: primary lymphoma of the pituitary: an emerging clinical entity. *J Clin Endocrinol Metab*. 2001;86:4567–4575.

22. Castle-Kirszbaum M, Goldschlager T, Ho B, et al. Twelve cases of pituitary metastasis: a case series and review of the literature. *Pituitary*. 2018;21:463–473.
23. Heaney AP. Clinical review: pituitary carcinoma: difficult diagnosis and treatment. *J Clin Endocrinol Metab*. 2011;96:3649–3660.
24. Gonzalez JG, Elizondo G, Saldivar D, et al. Pituitary gland growth during normal pregnancy: an in vivo study using magnetic resonance imaging. *Am J Med*. 1988;85:217–220.
25. Khawaja NM, Taher BM, Barham ME, et al. Pituitary enlargement in patients with primary hypothyroidism. *Endocr Pract*. 2006;12:29–34.
26. Thorner MO, Perryman RL, Cronin MJ, et al. Somatotroph hyperplasia. Successful treatment of acromegaly by removal of a pancreatic islet tumor secreting a growth hormone-releasing factor. *J Clin Invest*. 1982;70:965–977.
27. Joshi MN, Whitelaw BC, Carroll PV. Mechanisms in endocrinology: hypophysitis: diagnosis and treatment. *Eur J Endocrinol*. 2018;179: R151–R163.
28. Honegger J, Schlaffer S, Menzel C, et al. Diagnosis of primary hypophysitis in Germany. *J Clin Endocrinol Metab*. 2015;100:3841–3849.
29. Thodou E, Asa SL, Kontogeorgos G, et al. Clinical case seminar: lymphocytic hypophysitis: clinicopathological findings. *J Clin Endocrinol Metab*. 1995;80:2302–2311.
30. Faje AT, Sullivan R, Lawrence D, et al. Ipilimumab-induced hypophysitis: a detailed longitudinal analysis in a large cohort of patients with metastatic melanoma. *J Clin Endocrinol Metab*. 2014;99:4078–4085.
31. Bando H, Iguchi G, Fukuoka H, et al. The prevalence of IgG4-related hypophysitis in 170 consecutive patients with hypopituitarism and/or central diabetes insipidus and review of the literature. *Eur J Endocrinol*. 2014;170:161–172.
32. Trifanescu R, Ansorge O, Wass JA, et al. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*. 2012;76:151–160.
33. Culver SA, Grober Y, Ornan DA, et al. A case for conservative management: characterizing the natural history of radiographically diagnosed Rathke cleft cysts. *J Clin Endocrinol Metab*. 2015;100:3943–3948.
34. Langrand C, Bihan H, Raverot G, et al. Hypothalamo-pituitary sarcoidosis: a multicenter study of 24 patients. *QJM*. 2012;105:981–995.
35. Kaltsas GA, Powles TB, Evanson J, et al. Hypothalamo-pituitary abnormalities in adult patients with langerhans cell histiocytosis: clinical, endocrinological, and radiological features and response to treatment. *J Clin Endocrinol Metab*. 2000;85:1370–1376.
36. Vates GE, Berger MS, Wilson CB. Diagnosis and management of pituitary abscess: a review of twenty-four cases. *J Neurosurg*. 2001;95:233–241.
37. Freda PU, Bruce JN, Khandji AG, et al. Presenting features in 269 patients with clinically nonfunctioning pituitary adenomas enrolled in a prospective study. *J Endocr Soc*. 2020;4: bvaa021.
38. Daneshdoost L, Gennarelli TA, Bashey HM, et al. Identification of gonadotroph adenomas in men with clinically nonfunctioning adenomas by the luteinizing hormone beta subunit response to thyrotropin-releasing hormone. *J Clin Endocrinol Metab*. 1993;77:1352–1355.
39. Snyder PJ, Sterling FH. Hypersecretion of LH and FSH by a pituitary adenoma. *J Clin Endocrinol Metab*. 1976;42:544–550.
40. Ntali G, Capatina C, Grossman A, et al. Clinical review: functioning gonadotroph adenomas. *J Clin Endocrinol Metab*. 2014;99:4423–4433.
41. Djerassi A, Coutifaris C, West VA, et al. Gonadotroph adenoma in a premenopausal woman secreting follicle-stimulating hormone and causing ovarian hyperstimulation. *J Clin Endocrinol Metab*. 1995;80:591–594.
42. Castelbaum AJ, Bigdeli H, Post KD, et al. Exacerbation of ovarian hyperstimulation by leuprolide reveals a gonadotroph adenoma. *Fertil Steril*. 2002;78:1311–1313.
43. Snyder PJ. Gonadotroph cell adenomas of the pituitary. *Endocr Rev*. 1985;6:552–563.
44. Chanson P, Schaison G. Pituitary apoplexy caused by GnRH-agonist treatment revealing gonadotroph adenoma. *J Clin Endocrinol Metab*. 1995;80:2267–2268.
45. Briet C, Salenave S, Bonneville JF, et al. Pituitary apoplexy. *Endocr Rev*. 2015;36:622–645.
46. Raappana A, Koivukangas J, Ebeling T, Pirila T. Incidence of pituitary adenomas in Northern Finland in 1992-2007. *J Clin Endocrinol Metab*. 2010;95:4268–4275.
47. Harris RI, Schatz NJ, Gennarelli T, et al. Follicle-stimulating hormone-secreting pituitary adenomas: correlation of reduction of adenoma size with reduction of hormonal hypersecretion after transsphenoidal surgery. *J Clin Endocrinol Metab*. 1983;56:1288–1293.
48. Daneshdoost L, Gennarelli TA, Bashey HM, et al. Recognition of gonadotroph adenomas in women. *N Engl J Med*. 1991;324:589–594.
49. Nishioka H, Inoshita N, Mete O, et al. The complementary role of transcription factors in the accurate diagnosis of clinically nonfunctioning pituitary adenomas. *Endocr Pathol*. 2015;26:349–355.
50. Nishioka H. Recent evolution of endoscopic endonasal surgery for treatment of pituitary adenomas. *Neurol Med Chir (Tokyo)*. 2017;57:151–158.
51. Little ASKD, White WL, et al. Results of a prospective multicenter controlled study comparing surgical outcomes of microscopic versus fully endoscopic surgery for nonfunctioning pituitary adenomas: the Transsphenoidal Extent of Resection (TRANSSPHER) Study. *J Neurosurg*. 2020;132:1043–1053.
52. Lee DK, Sung MS, Park SW. Factors influencing visual field recovery after transsphenoidal resection of a pituitary adenoma. *Korean J Ophthalmol*. 2018;32:488–496.
53. Lee S, Kim SJ, Yu YS, et al. Prognostic factors for visual recovery after transsphenoidal pituitary adenectomy. *Br J Neurosurg*. 2013;27:425–429.
54. Luomaranta T, Raappana A, Saarela V, et al. Factors affecting the visual outcome of pituitary adenoma patients treated with endoscopic transsphenoidal surgery. *World Neurosurg*. 2017;105:422–431.
55. Jahangiri A, Wagner JR, Han SW, et al. Improved versus worsened endocrine function after transsphenoidal surgery for nonfunctional pituitary adenomas: rate, time course, and radiological analysis. *J Neurosurg*. 2016;124:589–595.
56. Little AS, Gardner PA, Fernandez-Miranda JC, et al. Pituitary gland recovery following fully endoscopic transsphenoidal surgery for nonfunctioning pituitary adenoma: results of a prospective multicenter study. *J Neurosurg*. 2019:1–7.
57. Pofi R, Gunatilake S, Macgregor V, et al. Recovery of the hypothalamo-pituitary-adrenal axis after transsphenoidal adenomectomy for non-ACTH-secreting macroadenomas. *J Clin Endocrinol Metab*. 2019;104:5316–5324.
58. Harary MDA, DiRisio AC, Dawood HY, et al. Endocrine function and gland volume after endoscopic transsphenoidal surgery for nonfunctional pituitary macroadenomas. *J Neurosurg*. 2019;131:1142–1151.
59. Ciric I, Ragin A, Baumgartner C, et al. Complications of transsphenoidal surgery: results of a national survey, review of the literature, and personal experience. *Neurosurgery*. 1997;40:225–236; discussion 36-37.
60. Levy MJ, Robertson IJ, Khalk N, et al. Long-term follow-up of a large prospective cohort of patients with nonfunctioning pituitary adenomas: the outcome of a conservative management policy. *Clin Endocrinol (Oxf)*. 2018;89:354–359.
61. Colin P, Jovenin N, Delemer B, et al. Treatment of pituitary adenomas by fractionated stereotactic radiotherapy: a prospective study of 110 patients. *Int J Radiat Oncol Biol Phys*. 2005;62:333–341.
62. Mackley HB, Reddy CA, Lee SY, et al. Intensity-modulated radiotherapy for pituitary adenomas: the preliminary report of the Cleveland Clinic experience. *Int J Radiat Oncol Biol Phys*. 2007;67:232–239.
63. Ronson BB, Schulte RW, Han KP, et al. Fractionated proton beam irradiation of pituitary adenomas. *Int J Radiat Oncol Biol Phys*. 2006;64:425–434.
64. Sheehan JP, Niranjan A, Sheehan JM, et al. Stereotactic radiosurgery for pituitary adenomas: an intermediate review of its safety, efficacy, and role in the neurosurgical treatment armamentarium. *J Neurosurg*. 2005;102:678–691.
65. Mingione V, Yen CP, Vance ML, et al. Gamma surgery in the treatment of nonsecretory pituitary macroadenoma. *J Neurosurg*. 2006;104:876–883.
66. Pivonello R, Matrone C, Filippella M, et al. Dopamine receptor expression and function in clinically nonfunctioning pituitary tumors: comparison with the effectiveness of cabergoline treatment. *J Clin Endocrinol Metab*. 2004;89:1674–1683.
67. Greenman Y, Cooper O, Yaish I, et al. Treatment of clinically nonfunctioning pituitary adenomas with dopamine agonists. *Eur J Endocrinol*. 2016;175:63–72.

68. Lee M, Lupp A, Mendoza N, et al. SSTR3 is a putative target for the medical treatment of gonadotroph adenomas of the pituitary. *Endocr Relat Cancer*. 2015;22:111–119.
69. Colao A, Di Somma C, Pivonello R, et al. Medical therapy for clinically non-functioning pituitary adenomas. *Endocr Relat Cancer*. 2008;15:905–915.
70. Fusco A, Giampietro A, Bianchi A, et al. Treatment with octreotide LAR in clinically non-functioning pituitary adenoma: results from a case-control study. *Pituitary*. 2012;15:571–578.
71. Klibanski A, Jameson JL, Biller BM, et al. Gonadotropin and alpha-subunit responses to chronic gonadotropin-releasing hormone analog administration in patients with glycoprotein hormone-secreting pituitary tumors. *J Clin Endocrinol Metab*. 1989;68:81–86.
72. Roman SH, Goldstein M, Kourides IA, et al. The luteinizing hormone-releasing hormone (LHRH) agonist [D-Trp6-Pro9-NEt]LHRH increased rather than lowered LH and alpha-subunit levels in a patient with an LH-secreting pituitary tumor. *J Clin Endocrinol Metab*. 1984;58:313–319.
73. McGrath GA, Goncalves RJ, Udupa JK, et al. New technique for quantitation of pituitary adenoma size: use in evaluating treatment of gonadotroph adenomas with a gonadotropin-releasing hormone antagonist. *J Clin Endocrinol Metab*. 1993;76:1363–1368.
74. Hwang K, Kwon T, Park J, et al. Growth pattern and prognostic factors of untreated nonfunctioning pituitary adenomas. *J Korean Neurosurg Soc*. 2019;62:256–262.
75. Karavitaki N, Collison K, Halliday J, et al. What is the natural history of nonoperated nonfunctioning pituitary adenomas? *Clin Endocrinol (Oxf)*. 2007;67:938–943.
76. Tresoldi AS, Carosi G, Betella N, et al. Clinically nonfunctioning pituitary incidentalomas: characteristics and natural history. *Neuroendocrinology*. 2020;110:595–603.
77. Drummond J, Roncaroli F, Grossman AB, et al. Clinical and pathological aspects of silent pituitary adenomas. *J Clin Endocrinol Metab*. 2019;104:2473–2489.

9

Thyroid-Stimulating Hormone–Producing Pituitary Tumors

Mark Gurnell, Waiel A. Bashari, Russell Senanayake, James MacFarlane, and Olympia Koulouri

OUTLINE

Introduction, 145
Epidemiology, 145
Etiopathogenesis and Pathology, 146
  Etiopathogenesis, 146
  Pathology, 146
Clinical Presentation, 148
Differential Diagnosis, 148
Investigation, 148
  Step 1 – Clinical Assessment, 148
  Step 2 – Exclusion of Laboratory Assay Interference, 148
  Step 3 – Distinguishing a Thyroid-Stimulating Hormone–Secreting Pituitary Adenoma from TR**β** Resistance to Thyroid Hormone, 149
  Step 4 – Pituitary Imaging, 151
Management, 151
Conclusions, 153

INTRODUCTION

In the majority of patients with hyperthyroidism, pituitary thyroid stimulating hormone (TSH, thyrotropin) secretion is suppressed as a consequence of autonomous thyroid hormone (TH, thyroxine [T4], triiodothyronine [T3]) production by the thyroid gland, and most commonly arises in the context of Graves disease, toxic multinodular goiter, or a solitary toxic adenoma. However, in a small but important subgroup of patients, TSH production is not suppressed, but is in fact the driver for hyperthyroidism.[1,2] TSH-producing pituitary adenomas (PAs), commonly referred to as thyrotropinomas or TSHomas, are an important cause of such central hyperthyroidism (i.e., neoplastic inappropriate secretion of TSH) and must be differentiated from nonneoplastic inappropriate TSH secretion, as is seen in the syndrome of resistance to thyroid hormone (RTHβ) caused by loss-of-function mutations in human *THRB* (encoding the β isoform of the thyroid hormone receptor [TRβ]).[3,4]

In recent years, increased clinician awareness, combined with more robust laboratory assays and improved cross-sectional imaging, has led to the recognition of previously undisclosed cases, especially microthyrotropinomas that are associated with only mild perturbations of thyroid function. However, because thyrotropinomas remain a rare cause of hyperthyroidism, many clinicians (even in tertiary centers) have only limited experience of their investigation and management, which can be challenging, especially in cases with atypical clinical features and equivocal or negative radiological findings.[2,5] Indeed, confirming the diagnosis, with exclusion of other conditions that can mimic a TSH-secreting PA, is often the point of greatest uncertainty in the management pathway.[6]

In this chapter, we consider current perspectives on the epidemiology and pathophysiology of TSH-producing PAs, review their varied modes of presentation, and provide guidance on the key steps in the diagnostic pathway and therapeutic options for this rare, but important, subgroup of PAs.

EPIDEMIOLOGY

Thyrotropinomas account for approximately 1% of all pituitary tumors in historical surgical and pathological series.[1] However, just as the overall prevalence of clinically relevant PAs is now recognized to be higher than previously suspected (~1:1200 of the general population), findings from a national registry in Sweden covering the interval 1990 to 2010 have shown (1) an age standardized national incidence of 0.15 per one million inhabitants per year, with an increasing incidence over time (0.05 per million per year in 1990–1994, rising to 0.26 per million per year in 2005–2009), and (2) a prevalence rate in 2010 of 2.8 TSH-secreting PAs per million inhabitants.[7] Several factors are likely to have contributed to recent higher detection rates, including: (1) availability in routine clinical laboratories of ultrasensitive and highly specific TSH assays; (2) greater use of combined measurement of free T4 and TSH to screen for suspected thyroid dysfunction; (3) increased availability of high quality pituitary magnetic resonance (MR) imaging; and (4) increased awareness of the condition. At the same time, the proportion of microadenomas has increased. Early case series reported macroadenomas in greater than 80% of patients with TSH-secreting tumours, with up two thirds manifesting extrasellar extension. However, recent studies indicate that micro-TSHomas are being increasingly diagnosed,[7–9] and it seems likely that the balance will be even further redressed with the advent of alternative (e.g., positron emission tomography [PET]) techniques for detecting small functioning tumours.[10] The majority of patients are diagnosed in middle age, with both sexes affected (median age of diagnosis 56 years in the Swedish cohort, with 60.7% women).[7] However, both paediatric and elderly (>80 years) subjects have been described.[1,5,11]

KEY POINTS

- Although thyrotropinomas are less common than other anterior pituitary adenomas, recent studies point to a higher prevalence than previously suspected (2.8 per million). Most arise sporadically, with a peak age at diagnosis between 40 and 60 years, with both sexes affected.

ETIOPATHOGENESIS AND PATHOLOGY

Etiopathogenesis

Under normal physiological conditions, the hypothalamo-pituitary-thyroid axis is tightly regulated to maintain T4 and T3 levels within a narrow range.[12] Thyrotropin-releasing hormone (TRH, a tripeptide: pyroGlu-His-Pro-NH_2) is secreted from parvocellular neurons in the paraventricular nucleus of the hypothalamus in response to low circulating levels of thyroid hormone (TH) or other triggers (e.g., environmental cold exposure, leptin, neuropeptide Y). Binding to the TRH receptor (TRHR, a calcium-mobilizing, rhodopsin family G protein–coupled receptor [GPCR]) on thyrotroph cells in the anteromedial portion of the adenohypophysis results in the synthesis and secretion of TSH through recruitment of Gq/11, triggering the phospholipase C–inositol 1,4,5-trisphosphate/1,2-diacylglycerol pathway, with resultant mobilization of intracellular calcium and activation of protein kinase C.[13,14]

TSH (a glycoprotein hormone heterodimer comprising alpha- and beta-subunits) binds to its cognate receptor (TSHR, also a GPCR) on the basolateral membrane of thyroid follicular cells to stimulate T4 and T3 synthesis and release. TSH exhibits both circadian and pulsatile patterns of secretion, peaking during the evening and early hours of the morning. Luteinizing hormone (LH) and human chorionic gonadotropin (hCG), which share a common alpha-subunit with TSH, can bind and activate the TSH receptor in certain clinical settings, for example, in early pregnancy when hCG levels are very high, with resultant elevation in T4 and T3 and consequent suppression of pituitary TSH release (see later). Thyrotroph cells in the adenohypophysis are also subject to negative regulation by somatostatin (predominantly acting through somatostatin receptor [SSTR] subtypes 1, 2, and 5) and dopamine (via the dopamine subtype 2 receptor). In addition, exposure to supraphysiological glucocorticoids, treatment with opioids, anorexia nervosa, and nonthyroidal illness have all been linked to reduced basal and TRH-stimulated TSH secretion.

T4 is produced in greater abundance (14:1) compared with T3, but the latter is the biologically active form. Conversion of T4 to T3 in peripheral tissues (in particular the liver) is facilitated by the type 1 deiodinase (DIO1), a member of a family of deiodinases (DIO1–3), which also metabolize T4 and T3 to inactive forms. DIO2 expressed in the hypothalamus and pituitary regulates local T3 production to facilitate inhibition of TRH and TSH production as part of a classical negative feedback loop. TH action in target tissues is mediated by two major receptor subtypes—alpha (TRα) and beta (TRβ) and their splice variants—which are members of the steroid nuclear receptor superfamily.[3,4]

Normal thyrotrophs account for only 5% of all anterior pituitary cells,[15] which may, at least in part, explain why thyrotropinomas are rare compared with other anterior pituitary tumor subtypes. The etiopathogenesis of TSH-secreting PAs remains poorly understood. Although the majority arise within the sella, a small number of TSH-secreting adenomas are located within the sphenoid sinus or nasopharynx and considered ectopic.[16,17] Thyrotropinomas have been observed in the context of multiple endocrine neoplasia type 1,[18] McCune–Albright syndrome (because of a somatic activating mutation in the α-subunit of Gs (encoded by *GNAS*),[19] and familial isolated PA (owing to mutation in the aryl hydrocarbon receptor-interacting protein [AIP]),[20] but such cases are rare, and the majority appear to arise sporadically.

Data from analyses of all pituitary tumor subtypes (including a small number of TSH-secreting adenomas) point to clonal expansion from a single transformed cell.[21,22] Evidence of aberrant transcriptional regulation of thyrotrophs (with resultant TSH hypersecretion and unrestrained growth) has been sought but, so far, most studies have proved largely uninformative, including those examining protooncogenes (e.g., *GSP*, *RAS*) and tumor suppressor genes (e.g., *RB*, *P53*).[23,24] The pituitary-specific transcription factor Pit-1 is overexpressed in some TSH-secreting PAs, and might play a role in cell proliferation as well as promoting hormone cosecretion in this subgroup of adenomas.[1,2]

The hallmark of TSH-secreting PAs is constitutive and autonomous secretion of TSH with failure of negative regulation by T3. Several possible molecular mechanisms for impaired negative feedback have been proposed. For example, a somatic loss-of-function mutation in *THRB* (encoding a dominant negative TRβ with impaired T3-mediated regulation of TSH transcription) was identified in one of five thyrotropinomas[25]; aberrant expression of a TRβ4 isoform (capable of functioning as an endogenous antagonist of normal TRβ signaling) was reported in three thyrotropinomas[26]; and disordered deiodinase function (resulting in reduced local pituitary T3 generation) was most pronounced in thyrotropinomas when compared with other pituitary tumor subtypes.[27] However, perhaps counter to this hypothesis, it is noteworthy that the occurrence of thyrotropinoma in subjects with *THRB* RTH (RTHβ) is extremely rare despite life-long aberrant TRβ signaling in these subjects.[28]

Although TRHR is expressed on the majority of thyrotroph tumor cells, the attenuated/absent response of TSH and α-subunit to exogenous TRH administration in many patients with TSH-secreting PAs suggests the TRHR signaling pathway is impaired/nonfunctional in these tumors.

Most thyrotropinomas express SSTR subtypes 1, 2, 3, and 5, with SSTR2 the dominant isoform. However, it seems unlikely that the primary molecular defect resides within this pathway, as depot somatostatin receptor ligand (SRL) therapy rapidly normalizes thyroid function in the majority of patients and may result in tumor shrinkage.[29]

In a recent study, using whole-exome sequencing in a small number of thyrotropinomas, no common driver mutations (either in genes known to predispose to other PA subtypes or amongst novel candidate genes identified in the tumors) were confirmed, suggesting that discrete primary events may underlie tumorigenesis in different patients.[30] Similarly, studies of copy number variations were largely inconclusive.[30]

Pathology

TSH-secreting PAs are classically chromophobic under light microscopy and demonstrate positive immunohistochemical staining for the unique TSH beta-subunit, the common alpha-subunit (ASU), and the transcription factor Pit-1. TSH-secreting macroadenomas are often accompanied by unbalanced hypersecretion of the common ASU, and double-immunostaining studies have suggested the existence of two different types of cells: one secreting ASU alone and another cosecreting the ASU and the TSH beta-subunit.[31]

Normal thyrotrophs share a common lineage with somatotrophs and lactotrophs, with Pit-1 being a key regulator of differentiation and proliferation for each of these three anterior pituitary cell types. Consistent with this, cosecretion of growth hormone (GH) and/or prolactin (PRL) is observed in 25% to 30% of thyrotropinomas (Fig. 9.1) and may be the dominant clinical manifestation, with hyperthyroxinemia and inappropriate TSH secretion only identified during endocrine evaluation. Intriguingly however, not all patients with biochemical evidence of abnormal GH secretion manifest overt acromegaly (Bashari, MacFarlane, Senanayake, Koulouri, Gurnell, unpublished observation). In addition, a subgroup of tumors demonstrate positive immunohistochemical staining for GH and/or PRL, but do not exhibit clinical or biochemical

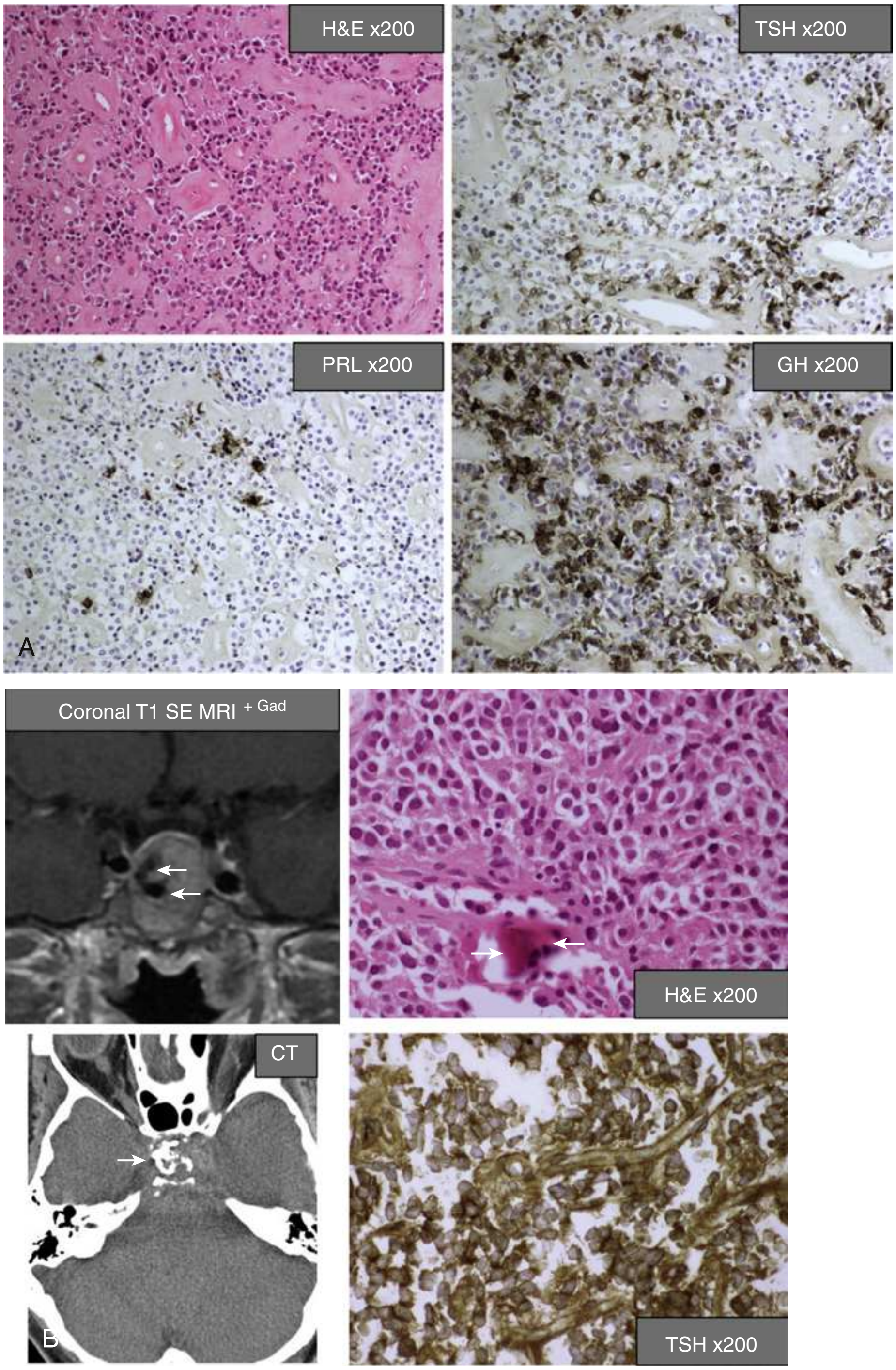


Fig. 9.1 Histological appearances in thyrotropinomas. **A**, Pituitary adenoma demonstrating positive immunohistochemical staining for thyroid-stimulating hormone (*TSH*), prolactin (PRL), and growth hormone (*GH*). **B**, Thyrotropinoma with prominent intratumoral calcification—on magnetic resonance imaging (MRI) a semicircular hypodense area is seen within the right side of the sella *(arrows)*, which could be mistaken for a vascular structure, but which is confirmed as focal calcification on computed tomography (*CT*) *(arrow)*. At surgery, macroscopic calcium deposits were resected, and histology confirmed additional microscopic foci of calcification *(arrows)* with strong immunohistochemical staining for TSH. *H&E*, Hematoxylin and eosin; *SE*, spin echo. (Pathology images provided courtesy of Dr. Kieren Allinson, Department of Neuropathology, Addenbrooke's Hospital, Cambridge.)

acromegaly or hyperprolactinaemia. Similarly, so-called silent thyrotropinomas (clinically nonfunctioning PAs, but with abundant positive TSH beta-subunit staining on immunohistochemistry) are well described,[32] and multihormone-staining (clinically silent) tumors have also been reported.

TSH-secreting PAs are almost always benign, with only a handful of malignant tumors (metastasizing to brain/spine, bone, or lung) reported to date.[33-35] A subgroup of thyrotropinomas are more fibrotic than other pituitary tumors, which has been linked to local secretion of basic fibroblast growth factor by the adenoma cells.[36] In this context, both microscopic and macroscopic calcification ("pituitary stone") may be seen, and can confound the interpretation of MR imaging and make surgical resection more challenging (Fig. 9.1).

Finally, not all TSH-secreting PAs are associated with overt elevation of serum TSH levels. Indeed, it has been estimated that between 30% and 40% of cases have a TSH level that repeatedly falls within the reference range but is inappropriately normal for the ambient raised thyroid hormone levels. This may be explained, at least in part, by increased bioactivity of neoplasm-derived TSH.[1]

KEY POINTS

- The majority of thyrotropinomas occur sporadically, with only a small number of cases arising in the context of genetic disorders known to predispose to pituitary adenomas. Although several molecular mechanisms have been postulated, the etiology of most thyroid-stimulating hormone (TSH)-secreting adenomas remains unknown. Approximately 25% of tumors cosecrete growth hormone and/or prolactin, but not all are associated with clinical sequelae. Ectopic TSH-secreting pituitary adenomas and TSH-secreting pituitary carcinomas are both described, but are extremely rare.

CLINICAL PRESENTATION

Clinical features of hyperthyroidism secondary to a TSH-secreting PA overlap those seen in primary hyperthyroidism (and include palpitations/arrhythmias, tremor, heat intolerance, weight loss, sleep disturbance, etc.), but without disease-specific (e.g., Graves ophthalmopathy or dermopathy) manifestations.[1,2,5] In some patients, symptoms and signs remain relatively unchanged over long intervals (≥10 years), whereas in others they are progressive. Clinical features are often surprisingly modest, reflecting milder perturbations in T4 and T3 levels than are typically seen in primary hyperthyroidism owing to Graves disease, toxic nodular disease, or thyroiditis. Indeed, some individuals appear to be asymptomatic even after direct questioning. Importantly, however, following successful treatment, a number of these patients report an improvement in their well-being, coinciding with resolution of previously unrecognized clinical symptoms (Bashari, MacFarlane, Senanayake, Koulouri, Gurnell, unpublished observation). The estimated mean latency between onset of hyperthyroidism and establishment of the diagnosis is typically around 4 to 5 years (but may exceed a decade). In mixed secreting tumors, clinical features of growth hormone excess (acromegaly) and hyperprolactinemia (hypogonadism, galactorrhea) may be seen or even dominate the presentation.

Goiter and/or thyroid nodules are found in many patients and likely reflect sustained TSH stimulation of thyroid follicular cells over many years.[37,38] The cooccurrence of differentiated thyroid carcinoma has been described in a small number of cases.[39,40] Autoimmune thyroid dysfunction may also coexist and give rise to diagnostic uncertainty.[41,42]

In early case series, compressive symptoms (such as visual field defects [80%] and headaches [20%]) prevailed over those owing to hyperthyroidism. However, this is no longer the case, as many patients present at an earlier stage, and the proportion of microadenomas is increasing.[7-9] Unilateral exophthalmos because of orbital invasion by the pituitary tumor is very rare, but must be distinguished from asymmetric exophthalmos of Graves disease. In macroadenomas, concomitant hypopituitarism (e.g., hypogonadism, hypoadrenalism) may be present. TSH-secreting PAs have been reported in pregnant women, in patients with multiple endocrine neoplasia type 1, and in atypical McCune–Albright syndrome.[18,19,43]

KEY POINTS

- As with other functioning pituitary adenomas, thyroid-stimulating hormone–secreting tumors may manifest with evidence of hormone excess (central hyperthyroidism), local compressive features (e.g., visual loss), and/or hypopituitarism. In contrast to primary hyperthyroidism, symptoms and signs can be surprisingly mild (even absent), and features of hormone cosecretion (e.g., acromegaly, hyperprolactinemia) may predominate. Coexistent thyroid nodules/cancer and autoimmune thyroid dysfunction may complicate the clinical presentation.

DIFFERENTIAL DIAGNOSIS

In theory, establishing the diagnosis of a TSH-secreting PA should be relatively straightforward, given the distinctive laboratory finding of inappropriately nonsuppressed TSH in the face of overt hyperthyroidism (with increased free T4 and free T3). However, this biochemical pattern is not restricted to thyrotropinomas, and it is often necessary to consider and actively exclude other possible causes before proceeding with targeted intervention (Table 9.1).

KEY POINTS

- Thyroid-stimulating hormone (TSH)-secreting pituitary adenomas have a distinctive laboratory fingerprint: raised thyroid hormones with inappropriate, nonsuppressed TSH. However, there are a multitude of other causes of hyperthyroxinemia with detectable TSH, and these must be carefully considered/excluded before concluding that the patient has a thyrotropinoma.

INVESTIGATION

Step 1 – Clinical Assessment

Careful clinical evaluation is an essential first step in the investigation of any patient with a suspected TSH-secreting PA. Specifically, it is important to: (1) identify potential confounding factors (including intercurrent illness, medications/supplements [e.g., amiodarone, heparin, biotin]); (2) make a judgment as to the clinical thyroid status of the patient (i.e., hypothyroid, euthyroid, or hyperthyroid); and (3) assess for other features that might point to a TSH-secreting adenoma (e.g., coexistent acromegaly or hyperprolactinemia, local suprasellar/parasellar mass effect) or, alternatively, suggest the possibility of laboratory assay artifact (e.g., falsely detectable serum TSH in a patient with dysthyroid eye signs and a thyroid bruit owing to Graves disease).

Step 2 – Exclusion of Laboratory Assay Interference

Thyroid function tests are amongst the most commonly requested biochemistry investigations. Both THs and TSH are typically measured using immunoassays on high-throughput automated analyzers. Such platforms are susceptible to multiple different forms of assay interference (Table 9.1), and close liaison with the clinical chemistry laboratory is therefore critically important to avoid inappropriate further investigation and treatment. Confirmation that free thyroid hormone levels (free T4 and T3) are genuinely raised may require referral to a reference laboratory with the ability to perform a two-step (back titration) assay, equilibrium dialysis, or mass spectrometry, which allow the detection of heterophilic or antiiodothyronine antibodies or variant binding proteins (e.g., albumin in familial dysalbuminemic hyperthyroxinemia; transthyretin in dystransthyretinemic hyperthyroxinemia). TSH dilution studies, immunosubtraction (using

TABLE 9.1 **Causes of Elevated Serum Thyroid Hormones With Nonsuppressed Thyroid-Stimulating Hormone**

Mechanism	Laboratory findings	Clinical context/example(s)
Increased thyroid hormone binding capacity	• raised total T4 & T3 • normal free T4 & T3 • nonsuppressed/normal TSH	• pregnancy • estrogen-containing oral contraceptive/hormone replacement therapy
Laboratory assay interference—falsely elevated free TH in a euthyroid individual	• raised free T4 ± T3—but with potential variability across assay platforms • normal TSH	• heterophile antibodies • antiiodothyronine antibodies • familial dysalbuminemic hyperthyroxinemia (FDH)[a] • familial dystransthyretinemic hyperthyroxinemia (DTTRH)[a]
Laboratory assay interference—falsely nonsuppressed TSH in a patient with genuine hyperthyroxinemia	• raised total and free T4 & T3 • detectable or raised TSH – but with potential variability across assay platforms	• heterophile antibodies • human antianimal antibodies • macro-TSH
Requirement for mildly elevated serum T4 to maintain normal serum T3 and TSH while taking exogenous thyroxine therapy	• raised free T4 • normal free T3 • normal TSH	• thyroxine replacement therapy in primary hypothyroidism
Inconsistent TH administration with dosing immediately before laboratory testing	• raised free T4 ± T3—but may vary depending on timing of blood sampling • raised TSH	• intermittent/erratic thyroxine administration
Inhibition of deiodinase function	• raised free T4 • normal free T3 • normal TSH	• amiodarone therapy
Displacement of TH from binding proteins[b]	• Raised free T4 • Normal/raised free T3 • normal TSH	• heparin therapy
Altered hypothalamic-pituitary-thyroid axis physiology; altered binding protein status	• raised free T4 ± T3—but with potential variability across assay platforms and phase of illness • nonsuppressed TSH	• acute psychiatric disorder • neonatal period
Neoplastic inappropriate TSH secretion	• raised total and free T4 & T3 • normal/raised TSH	• TSH-secreting pituitary adenoma
Impaired TH action: loss-of-function mutations in *THRB*	• raised free T4 & T3 • normal TSH	• TRβ resistance to thyroid hormone (*THRB* RTH)
Impaired TH action: loss-of-function mutations in *THRA*	• normal/low free T4 • raised free T3 • normal TSH	• TRα resistance to thyroid hormone (*THRA* RTH)
Impaired TH transport: loss-of-function mutations in *SLC16A2* (*MCT8*)	• normal/low free T4 • raised free T3 • normal TSH	• Allan–Herndon–Dudley syndrome (MCT8 deficiency)
Impaired TH metabolism: loss-of-function mutations in *SECISBP2* (*SBP2*)	• raised free T4 • normal/low free T3 • normal TSH	• SECISBP2 deficiency syndrome (typically associated with widespread selenoprotein deficiencies)

[a] FDH and DTTRH are also both associated with an increase in total TH levels because of the increased binding capacity of mutant albumin and transthyretin, respectively; at the same time, artifactual elevation in free T4 (and in some instances free T3) is observed on some laboratory platforms.

[b] Both fractionated and unfractionated heparin can cause artifactual elevation in measured free T4 (± free T3) owing to displacement from albumin binding sites by free fatty acids (generated in response to heparin-mediated activation of endothelial lipoprotein lipase).

MCT8, Monocarboxylate transporter 8; *SECISBP2*, selenocysteine insertion sequence binding protein 2; *SLC16A2*, solute carrier family 16 member 2; *T3*, triiodothyronine; *T4*, thyroxine; *TH*, thyroid hormone; *TSH*, thyroid-stimulating hormone.

polyethylene glycol or protein G/A), or gel filtration chromatography can be used to exclude interference in the TSH assay.[44]

Step 3 – Distinguishing a Thyroid-Stimulating Hormone–Secreting Pituitary Adenoma From TRβ Resistance to Thyroid Hormone

Once genuine hyperthyroxinemia with nonsuppressed TSH has been confirmed, the main task is to distinguish a TSH-secreting PA from rare genetic disorders of thyroid hormone transport, metabolism or action, in particular RTHβ). The degree of clinical thyrotoxicosis and the magnitude of elevation of thyroid hormone and TSH levels show considerable overlap when comparing cohorts of patients with thyrotropinomas and RTHβ. Accordingly, other investigations are often employed to distinguish between these two entities (Table 9.2).

Family History/Genetic Screening. The majority of TSH-secreting PAs are sporadic, with only occasional cases reported in the context of familial PA syndromes. In contrast, RTHβ is an autosomal dominant disorder, and affected family members can be readily identified through thyroid function testing, which then guides subsequent genetic screening. However, between 10% and 15% of patients with

TABLE 9.2 Distinguishing Between Thyroid-Stimulating Hormone–Secreting Pituitary Adenoma and TRβ Resistance to Thyroid Hormone (TRβ RTH)

	TSH-Secreting Pituitary Adenoma	TRβ RTH
Family history	typically no relevant family history	often positive (either known or identified during investigation)
Genetic analysis (*THRB* gene screen)	negative[a]	positive in 85%–90% of cases
Serum T4 and T3	elevated[b]	elevated[c]
Serum TSH	normal or elevated	normal or mildly elevated
Circadian TSH secretion	absent	preserved
SHBG	elevated (but may be normal in tumors with GH cosecretion and in some microthyrotropinomas)	normal (but may be elevated by other factors, e.g., combined oral contraceptive use)
ASU	normal or elevated	normal
ASU:TSH molar ratio	normal or elevated	normal
Response to exogenous TRH stimulation	absent/attenuated	preserved/exaggerated
Response to exogenous L-T3 suppression	absent/attenuated	retained (although may be incomplete)
Response to trial of depot SRL therapy	rapid and sustained suppression of TSH in majority of cases (with restoration of euthyroidism or induction of hypothyroidism)	no effect[d]
Features of acromegaly and/or hyperprolactinemia	present in 25%–30% of cases	absent
MR imaging	adenoma (although some microadenomas may not be visualized on standard clinical MR imaging)	normal (although incidentalomas may be identified)

[a] Two cases of apparent coexistent RTHβ and a TSH-secreting pituitary adenoma have so far been reported.
[b] Occasionally, serum free T4 and/or free T3 may intermittently fall within the upper part of the laboratory reference range in a patient with a thyrotropinoma.
[c] In a small subgroup of patients with RTHβ, serum free T3 is not overtly elevated outside the reference range.
[d] In RTHβ there may be a transient minor reduction in thyroid hormone and TSH levels immediately after commencing depot SRL therapy, but this is not sustained.
ASU, Alpha-subunit; *GH*, growth hormone; *L-T3*, lio-thyronine; *MR*, magnetic resonance; *SHBG*, sex hormone–binding globulin; *SRL*, somatostatin receptor ligand; *T3*, triiodothyronine; *T4*, thyroxine; *TRH*, thyrotropin-releasing hormone; *TSH*, thyroid-stimulating hormone.

RTHβ do not harbor a demonstrable mutation in *THRB* (see Chapter 81: Resistance to Thyroid Hormone).

Alpha Subunit and Alpha Subunit to Thyroid-Stimulating Hormone Molar Ratio. In traditional case series, a raised serum ASU level and increased ASU:TSH molar ratio were considered important diagnostic markers for thyrotropinoma. The ASU:TSH molar ratio is recommended to mitigate the potential confounding effects of raised gonadotrophin levels (e.g., in postmenopausal women), reflecting the shared common alpha-subunit with LH and FSH. However, in this context, the population reference range for ASU:TSH molar ratio is so wide[45] as to render it of limited value for the detection of all but the most obvious macroadenomas. As a general rule, a raised ASU (and/or ASU:TSH molar ratio) may help to support suspicions of a thyrotropinoma, but a normal result does not exclude the possibility.

Sex Hormone-Binding Globulin. Hepatic production of sex hormone-binding globulin (SHBG) is under direct thyroid hormone (positive) regulation via the TRβ1 isoform of the receptor. In RTHβ the liver is therefore resistant, and SHBG levels typically fall within the age- and gender-related reference range. In contrast, in the presence of a TSH-secreting tumor, elevated thyroid hormone levels cause hepatic thyrotoxicosis with raised SHBG. Unfortunately, however, normal SHBG levels may sometimes be seen, especially in microthyrotropinomas or when there is cosecretion of growth hormone. Similarly, other factors (e.g., oral estrogen therapy, pregnancy, liver dysfunction, anorexia) may cause an elevated SHBG in RTHβ.

Dynamic Testing. The use of dynamic testing can be particularly helpful in discriminating between RTHβ and a TSH-secreting PA.[1,2,44]

1. TRH test (200 mcg intravenously) (not available in all countries):
   - Thyrotropinoma: absent or attenuated (<4-fold) rise in TSH
   - RTHβ preserved or exaggerated TSH response
2. Depot SRL (e.g., lanreotide autogel 90–120 mg or sandostatin LAR 20–30 mg every 4 weeks for two to three injections):
   - Thyrotropinoma: rapid and sustained normalization of free T3, free T4, and TSH levels in 80% to 90% of patients; some patients may even develop central hypothyroidism requiring dose titration and/or levothyroxine supplementation.[29,46]
   - RTHβ no significant change in free TH or TSH levels (although a transient nonsustained diminution may be observed if levels are checked shortly after initiating treatment[46])

   Interestingly, the short-acting octreotide suppression test (single dose of 100 mcg subcutaneously) is less discriminatory, with TSH levels partially suppressing (~50%) in some patients with RTHβ, as well as in the majority of TSH-secreting PAs.[46]
3. Liothyronine (L-T3) suppression test (various protocols exist, including progressive dose escalation or fixed dosing [e.g., 100 mcg/day] over a 10-day period):
   - Thyrotropinoma: absent or only slight reduction of TSH
   - RTHβ: TSH suppression is marked (although not always complete)

In practice, the L-T3 suppression test is often reserved for those cases in which results remain equivocal despite extensive investigation, or for detecting low level residual or recurrent disease following pituitary surgery.[1,6] The test is not suitable for those with known or suspected cardiac disease (arrhythmias, ischemia, heart failure) or a history of major psychiatric disease.

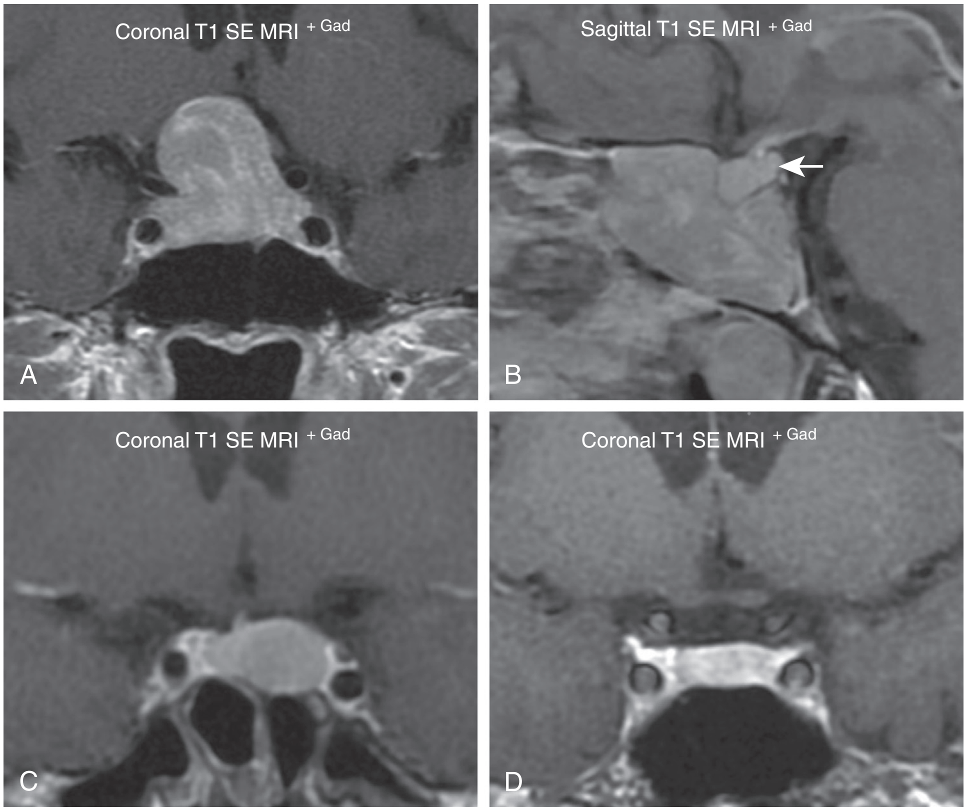


Fig. 9.2 Spectrum of thyroid-stimulating hormone–secreting pituitary adenomas. **A**, Macroadenoma with predominant suprasellar extension compressing the optic chiasm. **B**, Macroadenoma filling the sphenoid sinus—note how the adenoma appears discrete from the normal pituitary gland, which is situated superiorly in the sella *(arrow)*, raising the possibility of an extrasellar ectopic origin. **C**, Left-sided predominantly intrasellar macroadenoma. **D**, No visualized adenoma. *Gad*, Gadolinium; *SE*, spin echo; *T1*, T1-weighted; *MRI*, magnetic resonance imaging.

Table 9.2 summarizes the typical findings in TSH-secreting PA and RTHβ across the various tests.

Step 4 – Pituitary Imaging

Once the diagnosis of a TSH-secreting PA has been confirmed, attention turns to tumor localization. The finding of a pituitary microadenoma in a patient with inappropriate secretion of TSH, although strongly suggestive, is not diagnostic of a TSH-secreting pituitary tumor, as pituitary incidentalomas are found on MR imaging in up to 10% of the general population. In contrast, incidental macroadenomas are much less commonly encountered. Indeed, some clinicians reason that, if a pituitary macroadenoma is found in a patient with raised free thyroid hormone levels and nonsuppressed TSH, then it is reasonable to proceed directly to surgery without additional confirmatory testing. However, in our experience this will occasionally result in unnecessary surgery, especially if due diligence in excluding laboratory assay interference is not undertaken. We therefore recommend that imaging is performed only when there is clear clinical and laboratory evidence of the presence of neoplastic inappropriate TSH secretion (± local mass effect).

In patients with TSH-secreting PAs there is no correlation between serum TSH and tumor size. In older case series, patients commonly presented with macroadenomas with suprasellar and sphenoidal extension, whereas in more recent series microadenomas feature much more frequently.[7-9] Indeed, it is now clear that thyrotropinomas exhibit a spectrum ranging from large invasive (even giant) tumors through to microadenomas that cannot be readily visualized using conventional MR sequences (analogous to some Cushing microadenomas) (Fig. 9.2). For the latter, various approaches have been employed to aid localization of the tumor (including inferior petrosal sinus sampling with TRH stimulation[9]) and octreotide scintigraphy.[9] However, the most promising approach appears to be functional/molecular imaging with ^{11}C-methionine PET–computed tomography (CT) coregistered with volumetric MRI, which may permit precise localization of the tumor (Fig. 9.3), especially when performed pre-and postsomatostatin analog therapy.[10]

KEY POINTS

- Diagnosis of a thyroid-stimulating hormone–secreting pituitary adenoma requires a multifaceted approach that combines clinical, laboratory, genetic, and radiological investigations in four sequential steps: (1) clinical assessment to determine thyroid status, exclude confounding intercurrent illness or medications, and identify associated clinical features (e.g., growth hormone or prolactin excess); (2) exclusion of laboratory assay interference; (3) distinction from TRβ resistance to thyroid hormone; (iv) cross-sectional (magnetic resonance imaging, computed tomography) with or without molecular imaging of the sella and parasellar regions to localize the tumor.

MANAGEMENT

In most patients with TSH-secreting pituitary tumors, the goals of therapy are similar to those for other functioning PAs, namely: (1)

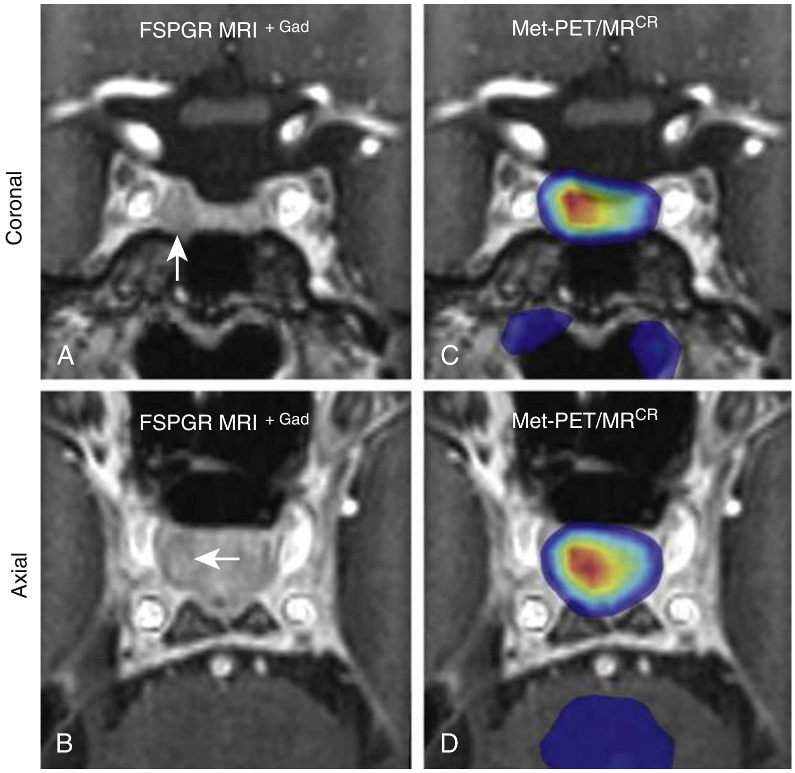


Fig. 9.3 Confirmation of the site of a microadenoma using volumetric magnetic resonance imaging (*MRI*) and molecular (functional) imaging with ^{11}C-methionine positron emission tomography (*PET*) in a patient in whom standard clinical magnetic *SE*, spin echo; (*MR*) sequences pre- and postcontrast had failed to identify a thyrotropinoma. **A** and **B**, Volumetric (*FSPGR*) MR sequences (1-mm slices) identify a possible right-sided microadenoma adjacent to the cavernous sinus *(arrows)*. **C** and **D** ,^{11}C-methionine PET/computed tomography (CT) coregistered with FSPGR MRI demonstrates avid tracer uptake at the site of the suspected lesion. Transsphenoidal surgery and histological analysis confirmed a right-sided thyroid-stimulating hormone–secreting microadenoma. *Gad*, Gadolinium; *FSPGR*, fast spoiled gradient recalled acquisition; *Met-PET/MR*^{CR}, ^{11}C-methionine PET/CT coregistered with volumetric MRI.

restoration of euthyroidism; (2) amelioration/reversal of local compressive symptoms/signs; and (3) correction of hypopituitarism. Early diagnosis helps prevent systemic complications (e.g., cardiac arrhythmias, osteoporosis) and may be associated with improved surgical cure rates.

Transsphenoidal surgery remains the treatment of choice. Historically, pituitary surgery has been reported to achieve normalization of thyroid hormone secretion and resolution of the pituitary mass in approximately 50% of patients, achieve normalization of thyroid parameters despite incomplete tumor removal in approximately 25% of patients, and be unsuccessful (with persistent hyperthyroidism and residual adenoma) in nearly 30% of cases. However, improved surgical remission rates have been reported in more recent series, likely reflecting improved surgical techniques and earlier diagnosis.[47] In addition, surgical cure rates for microadenomas are considerably higher and approach 90% in experienced centers.[48]

The success of treatment is also dependent on the criteria used. It has been proposed that an early indicator of "cure" may be an undetectable TSH concentration 7 days after surgery (when the normal thyrotrophs remain suppressed as a consequence of prolonged exposure to raised circulating thyroid hormone levels).[47] However, many patients are now treated with somatostatin analogs before surgery, often with rapid normalization of hyperthyroidism, which may therefore allow recovery of normal thyrotroph function and, at the same time, confound interpretation of early postoperative testing. Accordingly, normalization of dynamic tests (e.g., TRH test, T3 suppression test) may represent a better (later) test of remission. In any case, long-term follow-up is necessary to detect relapse and recurrence.

Primary medical therapy is increasingly recognized as a viable alternative to pituitary surgery in some patients with TSH-secreting pituitary tumors. Dopamine agonists are an effective treatment in true mixed TSH/PRL-secreting PAs, whereas success is limited when bromocriptine is used in patients with pure TSH-secreting PAs.[49] In contrast, depot SRLs (e.g., octreotide LAR or lanreotide ATG) suppress TSH secretion in approximately 90% of TSH-secreting PAs and normalize thyroid hormone levels in approximately 75% to 90% of patients. Moreover, it has been estimated that approximately 15% of patients will develop central hypothyroidism requiring dose titration and/or levothyroxine supplementation.[29] Interestingly, significant tumor shrinkage (≥20%) is observed in only approximately 50% of cases, and total resolution of a macroadenoma with maintenance of long-term remission in response to somatostatin analog therapy has been reported only rarely.[50] As a general rule, the effects of SRLs are reversible, and treatment must therefore be continued long-term. The development of tachyphylaxis, requiring dose escalation to maintain biochemical control, is seen in up to 10% of cases. True SRL resistance may be observed in a small number of patients.[51] The second-generation SRL pasireotide may be considered for treatment resistant cases.[52] Long-term SRL therapy is expensive and may be associated with side effects such as cholelithiasis and carbohydrate intolerance. Accordingly, although some patients and their clinicians may elect to pursue long-term primary SRL therapy, for many patients it remains a bridge to surgery (allowing anesthesia to proceed safely), and as adjunctive therapy when surgery is unable to achieve full remission, or while awaiting the beneficial effects of radiotherapy.

It is appropriate to briefly mention conventional antithyroid drug therapy, which may be used for a short duration in preparation for surgery. However, concerns remain that longer-term use may predispose to tumor expansion (Nelson-like phenomenon),[1] and for that reason permanent thyroid ablation is also generally best avoided, although may occasionally be indicated in those with recalcitrant arrhythmias or cardiac failure.

Conventional radiotherapy and stereotactic radiosurgery have also been employed in the management of TSH-secreting pituitary tumors, especially when surgery and/or medical therapy has proved unsuccessful.[53] However, the number of reported cases, especially using modern radiotherapy techniques, remains relatively small. As with other pituitary tumor subtypes, allowing for the time lapse required for the full effect to be realized, and the potential for late side effects (e.g., hypopituitarism, second tumor, possible increased cerebrovascular disease), radiotherapy is generally reserved for the small number of cases in which combination surgery and medical therapy are unable to achieve satisfactory biochemical and/or tumor control.

KEY POINTS

- For thyroid-stimulating hormone–secreting pituitary microadenomas, transsphenoidal surgery by an experienced pituitary surgeon is generally preferred and achieves complete/sustained remission in a significant proportion of cases. For macroadenomas, surgical success is influenced by several factors, including surgical expertise, tumor size, and extent of cavernous sinus invasion. Somatostatin receptor ligand therapy may be used before surgery to permit safe anesthesia, or when there is persistent disease—either as long-term treatment or while awaiting the effects of radiotherapy. Antithyroid drugs can help achieve control of hyperthyroidism pending other interventions, but are typically restricted to short-term use to minimize the risk of tumor expansion.

CONCLUSIONS

Although thyrotropinomas remain one of the rarest anterior PA subtypes, recent studies suggest they are more common than previously reported, with a greater proportion of microadenomas now diagnosed. However, understanding of the etiopathogenesis of these tumors remains limited, and further collaborative studies across major centers are likely to be required to improve understanding of the molecular basis for the development of tumoral inappropriate TSH secretion. Combined clinical, laboratory, genetic, and radiological investigation allows the diagnosis of a TSH-secreting PA to be made with confidence, and in the knowledge that other causes of hyperthyroxinemia with nonsuppressed TSH have been robustly excluded. Molecular imaging shows promise for the localization of small tumors that are beyond the resolution of existing MR imaging. Transsphenoidal surgery remains the treatment of choice, although depot somatostatin analog therapy is now established as a key component in the management pathway: either pre- or postsurgery, or increasingly as long-term primary medical therapy.

REFERENCES

1. Beck-Peccoz P, Persani L, Mannavola D, et al. Pituitary tumours: TSH-secreting adenomas. *Best Prac Res Clin Endocrinol Metab*. 2009;23:597–606.
2. Amlashi FG, Tritos NA. Thyrotropin-secreting pituitary adenomas: epidemiology, diagnosis, and management. *Endocrine*. 2016;52:427–440.
3. Moran C, Gurnell M, Chatterjee VK. Syndromes of resistance to thyroid hormone. In: *Oxford Textbook of Endocrinology*. 3rd ed. Oxford: Oxford University Press; 2021.
4. Deleted in review.
5. Tjörnstrand A, Nyström HF. Diagnosis of endocrine disease: diagnostic approach to TSH-producing pituitary adenoma. *Euro J Endocrinol*. 2017;177:R183–R197.
6. Beck-Peccoz P, Lania A, Beckers A, et al. 2013 European thyroid association guidelines for the diagnosis and treatment of thyrotropin-secreting pituitary tumors. *Euro Thyroid J*. 2013;2:76–82.
7. Önnestam L, Berinder K, Burman P, et al. National incidence and prevalence of TSH-secreting pituitary adenomas in Sweden. *J Clin Endocrinol Metab*. 2013;98:626–635.
8. Macchia E, Gasperi M, Lombardi M, et al. Clinical aspects and therapeutic outcome in thyrotropin-secreting pituitary adenomas: a single center experience. *J Endocrinol Invest*. 2009;32:773–779.
9. Socin HV, Chanson P, Delemer B, et al. The changing spectrum of TSH-secreting pituitary adenomas: diagnosis and management in 43 patients. *Euro J Endocrinol*. 2003;148:433–442.
10. Koulouri O, Hoole AC, English P, et al. Localisation of an occult thyrotropinoma with 11C-methionine PET-CT before and after somatostatin analogue therapy. *Lancet Diabetes Endocrinol*. 2016;4:1050.
11. Nakayama Y, Jinguji S, Kumakura S, et al. Thyroid-stimulating hormone (thyrotropin)-secretion pituitary adenoma in an 8-year-old boy: case report. *Pituitary*. 2012;15:110–115.
12. Andersen S, Pedersen KM, Bruun NH, et al. Narrow individual variations in serum T4 and T3 in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab*. 2002;87:1068–1072.
13. Hinkle PM, Gehret AU, Jones BW. Desensitization, trafficking, and resensitization of the pituitary thyrotropin-releasing hormone receptor. *Front Neurosci*. 2012;6:180.
14. Bonomi M, Cangiano B, Persani L. *Encyclopedia of Endocrine Diseases*. Vol 4. New York: Academic Press; 2018.
15. Kovacs K. Tumours of the pituitary gland. In: *Atlas of Tumour Pathology*. Armed Forces Institute of Pathology; 1986. Fascicle 21, series 2.
16. Jin Z, Zhang Z, Li F, et al. Hyperthyroidism caused by an ectopic thyrotropin-secreting tumor of the nasopharynx: a case report and review of the literature. *Thyroid*. 2013;23:1172–1177.
17. Ortiz E, Peldoza M, Monnier E, et al. Ectopic pituitary adenoma of the TSH-secreting sphenoidal sinus with excellent response to somatostatin analogs. Theory of the embryogenesis and literature review from a clinical case. *Steroids*. 2020;154:108535.
18. Burgess J, Shepherd J, Greenaway T. Thyrotropinomas in multiple endocrine neoplasia type 1 (MEN-1). *ANZ J Med*. 1994;24:740–741.
19. Gessl A, Freissmuth M, Czech T, et al. Growth hormone-prolactin-thyrotropin-secreting pituitary adenoma in atypical McCune-Albright syndrome with functionally normal Gs alpha protein. *J Clin Endocrinol Metab*. 1994;79:1128–1134.
20. Hernández-Ramírez LC, Gabrovska P, Dénes J, et al. Landscape of familial isolated and young-onset pituitary adenomas: prospective diagnosis in AIP mutation carriers. *J Clin Endocrinol Metab*. 2015;100:E1242–E1254.
21. Ma W, Ikeda H, Watabe N, et al. A plurihormonal TSH-producing pituitary tumor of monoclonal origin in a patient with hypothyroidism. *Hormone Res*. 2003;59:257–261.
22. Herman V, Fagin J, Gonsky R, et al. Clonal origin of pituitary adenomas. *J Clin Endocrinol Metab*. 1990;71:1427–1433.
23. Sumi T, Stefaneanu L, Kovacs K, et al. Immunohistochemical study of p53 protein in human and animal pituitary tumors. *Endocrine Pathol*. 1993;4:95.
24. Dong Q, Brucker-Davis F, Weintraub BD, et al. Screening of candidate oncogenes in human thyrotroph tumors: absence of activating mutations of the G alpha q, G alpha 11, G alpha s, or thyrotropin-releasing hormone receptor genes. *J Clin Endocrinol Metab*. 1996;81:1134–1140.
25. Ando S, Sarlis NJ, Oldfield EH, et al. Somatic mutation of TRβ can cause a defect in negative regulation of TSH in a TSH-secreting pituitary tumor. *J Clin Endocrinol Metab*. 2001;86:5572–5576.
26. Tagami T, Usui T, Shimatsu A, et al. Aberrant expression of thyroid hormone receptor β isoform may cause inappropriate secretion of TSH in a

TSH-secreting pituitary adenoma. *J Clin Endocrinol Metab*. 2011;96:E948–E952.

27. Tannahill L, Visser T, McCabe C, et al. Dysregulation of iodothyronine deiodinase enzyme expression and function in human pituitary tumours. *Clin Endocrinol*. 2002;56:735–743.
28. Teng X, Jin T, Brent GA, et al. A patient with a thyrotropin-secreting microadenoma and resistance to thyroid hormone (P453T). *J Clin Endocrinol Metab*. 2015;100:2511–2514.
29. Illouz F, Chanson P, Sonnet E, et al. Somatostatin receptor ligands induce TSH deficiency in thyrotropin-secreting pituitary adenoma. *Euro J Endocrinol*. 2020;184:1–8.
30. Sapkota S, Horiguchi K, Tosaka M, et al. Whole-exome sequencing study of thyrotropin-secreting pituitary adenomas. *J Clin Endocrinol Metab*. 2017;102:566–575.
31. Terzolo M, Orlandi F, Bassetti M, et al. Hyperthyroidism due to a pituitary adenoma composed of two different cell types, one secreting α-subunit alone and another cosecreting α-subunit and thyrotropin. *J Clin Endocrinol Metab*. 1991;72:415–421.
32. Kirkman MA, Jaunmuktane Z, Brandner S, et al. Active and silent thyroid-stimulating hormone-expressing pituitary adenomas: presenting symptoms, treatment, outcomes, and recurrence. *World Neurosurg*. 2014;82:1224–1231.
33. Mixson AJ, Friedman TC, Katz DA, et al. Thyrotropin-secreting pituitary carcinoma. *J Clin Endocrinol Metab*. 1993;76:529–533.
34. Brown RL, Muzzafar T, Wollman R, et al. A pituitary carcinoma secreting TSH and prolactin: a non-secreting adenoma gone awry. *Euro J Endocrinol*. 2006;154:639–643.
35. Lee W, Cheung AS, Freilich R. TSH–secreting pituitary carcinoma with intrathecal drop metastases. *Clin Endocrinol*. 2012;76:604–606.
36. Smyth HS, Asa SL, Ezzat S, et al. Basic fibroblast growth factor expression by two prolactin and thyrotropin-producing pituitary adenomas. *Endo Pathol*. 2007;6:125–134.
37. Gesundheit N, Petrick PA, Nissim M, et al. Thyrotropin-secreting pituitary adenomas: clinical and biochemical heterogeneity: case reports and follow-up of nine patients. *Ann Int Med*. 1989;111:827–835.
38. Abs R, Stevenaert A, Beckers A. Autonomously functioning thyroid nodules in a patient with a thyrotropin-secreting pituitary adenoma: possible cause–effect relationship. *Euro J Endocrinol*. 1994;131:355–358.
39. Ünlütürk U, Sriphrapradang C, Erdoğan MF, et al. Management of differentiated thyroid cancer in the presence of resistance to thyroid hormone and TSH-secreting adenomas: a report of four cases and review of the literature. *J Clin Endocrinol Metab*. 2013;98:2210–2217.
40. Perticone F, Pigliaru F, Mariotti S, et al. Is the incidence of differentiated thyroid cancer increased in patients with thyrotropin-secreting adenomas? Report of three cases from a large consecutive series. *Thyroid*. 2015;25:417–424.
41. Arai N, Inaba M, Ichijyo T, et al. Thyrotropin-producing pituitary adenoma simultaneously existing with Graves' disease: a case report. *J Med Case Rep*. 2017;11.
42. Losa M, Mortini P, Minelli R, et al. Coexistence of TSH-secreting pituitary adenoma and autoimmune hypothyroidism. *J Endocrinol Invest*. 2006;29:555–559.
43. Caron P, Gerbeau C, Pradayrol L, et al. Successful pregnancy in an infertile woman with a thyrotropin-secreting macroadenoma treated with somatostatin analog (octreotide). *J Clin Endocrinol Metab*. 1996;81:1164–1168.
44. Koulouri O, Moran C, Halsall D, et al. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Prac Res Clin Endocrinol Metab*. 2013;27:745–762.
45. Beck-Peccoz P, Persani L, Faglia G. Glycoprotein hormone α-subunit in pituitary adenomas. *Trend Endocrinol Metab*. 1992;3:41–45.
46. Mannavola D, Persani L, Vannucchi G, et al. Different response to chronic somatostatin analogues in patients with central hyperthyroidism. *Clin Endocrinol*. 2005;62:176–181.
47. Yamada S, Yamaguchi-Okada M, Takeuchi Y, et al. Clinicopathological characteristics and therapeutic outcomes in thyrotropin-secreting pituitary adenomas: a single-center study of 90 cases. *J Neurosurg*. 2014;121:1462–1473.
48. Cossu G, Daniel RT, Pierzchala K, et al. Thyrotropin-secreting pituitary adenomas: a systematic review and meta-analysis of postoperative outcomes and management. *Pituitary*. 2019;22:79–88.
49. Teramoto A, Sanno N, Tahara S, et al. Pathological study of thyrotropin-secreting pituitary adenoma: plurihormonality and medical treatment. *Acta Neuropathol*. 2004;108:147–153.
50. Fliers E, van Furth WR, Bisschop PH. Cure of a thyrotrophin (TSH)-secreting pituitary adenoma by medical therapy. *Clin Endocrinol*. 2012;77:788–790.
51. Beck-Peccoz P, Brucker-Davis F, Persani L, et al. Thyrotropin-secreting pituitary tumors. *Endo Rev*. 1996;17:610–638.
52. van Eersel MEA, Meeuwisse–Pasterkamp SH, Muller Kobold AC, et al. Treatment of a thyrotropin-secreting pituitary adenoma (TSH-oma) with pasireotide LAR. *Clin Endocrinol*. 2017;87:877–879.
53. Malchiodi E, Profka E, Ferrante E, et al. Thyrotropin-secreting pituitary adenomas: outcome of pituitary surgery and irradiation. *J Clin Endocrinol Metabol*. 2014;99:2069–2076.

Pituitary Surgery

Daniel F. Kelly, Michael B. Avery, and Garni Barkhoudarian

OUTLINE

History and Recent Developments of Transsphenoidal Surgery, 155
Surgical Indications and Goals of Surgery, 155
- Pituitary Adenoma, 155
- Rathke's Cleft Cyst, 156
- Craniopharyngioma, 157
- Other Lesions, 158

Preoperative Management, 159
- Endocrine Status, 159
- Imaging Evaluation, 160

Surgical Technique, 160
- Endoscopic Endonasal Approach for Pituitary Adenoma, Rathke's Cleft Cyst, and Craniopharyngioma, 161

Acute Postoperative Management and Early Observation Postdischarge, 163
- Diabetes Insipidus, 163
- Syndrome of Inappropriate Antidiuretic Hormone, 163
- Early Remission, 163

Outcomes After Transsphenoidal Pituitary Adenoma Surgery, 163
Major Complications, 165
- Permanent Hypopituitarism, 165
- Cerebrospinal Fluid Leak and Meningitis, 165
- Cranial Nerve Injury, 165
- Postoperative Intrasellar Hematoma, 165
- Internal Carotid Artery Injury, 165

Resection and Remission Rates, 166
- Pituitary Adenoma, 166
- Rathke's Cleft Cyst, 166
- Craniopharyngioma, 166

Pituitary Centers of Excellence, 166
Conclusion, 167

HISTORY AND RECENT DEVELOPMENTS OF TRANSSPHENOIDAL SURGERY

The advent of transsphenoidal pituitary adenoma surgery dates back over a century. Since that time, the safety and efficacy have improved tremendously. The first major improvement came in the 1960s with the introduction of the operating microscope and fluoroscopy. These two advances significantly improved operative field illumination and magnification and provided a rudimentary degree of neuronavigation. Selective adenomectomy and pituitary gland preservation became increasingly possible, resulting in a marked improvement in patient outcomes. The era of endoscopic transsphenoidal surgery began in the late 1990s, bringing with it further improvements in illumination, image quality, viewing angle, and dexterity (Fig. 10.1).[1-5] Since then, many other important technical strides have been made, including utilization of the Doppler probe to localize the parasellar internal carotid arteries (ICAs) and avoid ICA injury; pseudocapsular dissection technique for maximizing the probability of remission/resection; greater appreciation of cavernous sinus (CS) anatomy and adenoma invasion patterns; intraoperative neurophysiological monitoring for cranial nerves; and enhanced skull base closure techniques to avoid cerebrospinal fluid (CSF) leaks and meningitis. Since 2013, national practice trends confirm that a majority of pituitary adenoma operations are now performed fully endoscopically.[6] From an organizational standpoint, the concept of multidisciplinary teams including neurosurgery, endocrinology, otolaryngology, neuroophthalmology, radiation oncology, neuroradiology, and neuropathology has been promoted to optimize management for this complex patient population.[7]

SURGICAL INDICATIONS AND GOALS OF SURGERY

The most common pituitary pathologies are pituitary adenomas, Rathke's cleft cysts (RCCs), and craniopharyngiomas. Surgical indications include: (1) pituitary or sellar mass lesions causing mass effects on surrounding structures resulting in symptoms such as hypopituitarism, visual loss, diplopia and/or headache, (2) hypersecreting adenomas leading to Cushing disease, acromegaly/gigantism, secondary hyperthyroidism, or hyperprolactinemia (refractory or intolerant of medical therapy), or (3) pituitary apoplexy, which may occur in endocrine-active and -inactive adenomas, as well as occasionally in RCCs (Table 10.1).

KEY POINTS

- The endoscopic endonasal approach is the predominant technique used for all sellar and parasellar tumors.

Pituitary Adenoma

Pituitary adenomas are broadly categorized based on size, invasiveness, and hormonal production. Microadenomas are less than 1 cm in maximal diameter, macroadenomas are 1 cm or greater, and giant adenomas are 4 cm or greater (Fig. 10.2). Degree of invasiveness into the CS is most commonly categorized with the Knosp grading system, and is based on coronal magnetic resonance imaging (MRI) sellar images. This grading system is increasingly used as an important parameter of surgical resectability, with grade 0 to 2 adenomas having no or minimal CS invasion and grade 3 to 4 adenomas having

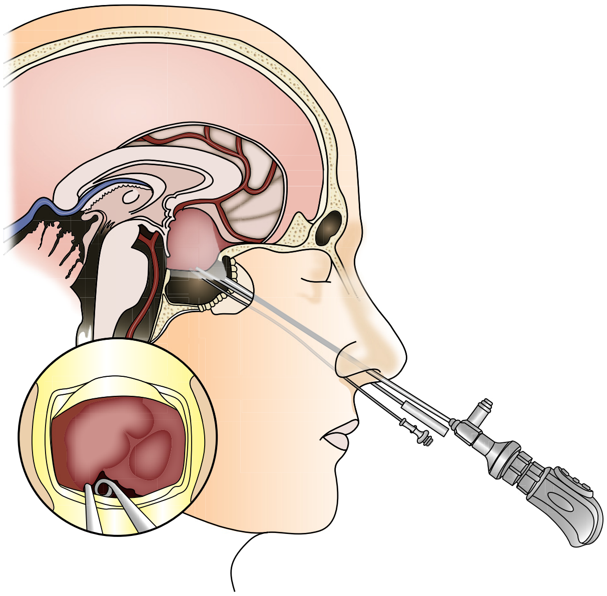


Fig. 10.1 Given the superior panoramic, high-definition view, the endoscopic endonasal approach with a rigid 4-mm endoscope is the preferred technique for removal of pituitary adenomas, Rathke's cleft cysts, and most craniopharyngiomas.

TABLE 10.1 Indications for Pituitary Surgery

1. Relief of mass effect
    - Headache
    - Hypopituitarism
    - Vision loss
    - Other cranial neuropathies
    - Headache
    - Hypothalamic dysfunction
2. Hypersecreting endocrinopathy
    - Acromegaly/gigantism (GH)
    - Cushing disease (ACTH)
    - TSH-secreting adenoma
    - Prolactinoma (PRL; refractory to medical management)
3. Pituitary apoplexy

GH, Growth hormone; *ACTH,* adrenocorticotropic hormone; *PRL*, prolactin; *TSH*, thyroid-stimulating hormone.

progressively greater CS invasion.[8] From an endocrine perspective, endocrine-inactive adenomas are without hormonal hypersecretion, and endocrine-active adenomas include patients with hypersecretion symptoms of Cushing disease with elevated/autonomous adrenocorticotropic hormone (ACTH) and cortisol, acromegaly/gigantism with elevated growth hormone (GH) and insulin like growth factor 1 (IGF-1), prolactinomas with elevated prolactin, and thyrotropinomas with elevated thyroid-stimulating hormone (TSH) and T4/T3. Importantly, as adenoma size increases, so generally does the invasiveness. However, even small endocrine-active adenomas such as those causing Cushing disease or prolactinomas can have grade 3 or 4 CS invasion.

Surgery is considered first-line therapy for all symptomatic endocrine-inactive and endocrine-active pituitary adenomas except prolactinomas, which are generally treated initially with dopamine agonists, either cabergoline or bromocriptine. As described in Chapter 4, medical therapy with dopamine agonists is effective in normalizing serum prolactin levels in approximately 95% and 80% of microprolactinomas and macroprolactinomas, respectively, with resultant tumor shrinkage in most patients.[9] In patients with prolactinomas that are refractory to medical therapy or in whom significant side-effects preclude consistent dosing, endonasal surgery should be considered. Additionally, in patients with prolactinomas who present with apoplexy, particularly those with visual or extraocular movement deficits, urgent endonasal surgery is generally recommended.

Endocrine-inactive microadenomas are often asymptomatic, while macroadenomas typically present with symptoms related to mass effects on adjacent neurological structures. Symptoms may include headache, vision loss (classically bitemporal hemianopia), hypopituitarism, cranial nerve palsy, pituitary apoplexy, and rarely hydrocephalus. Serum prolactin levels may be elevated because of interruption of the inhibitory tuberoinfundibular dopamine pathway, known as the stalk effect. In our recent series on patients aged 65 years or older with pituitary adenomas, approximately 25% did not require surgery and were successfully managed conservatively.[10]

In all operative cases, surgical goals include: (1) maximal safe adenoma resection with biochemical remission and/or relief of mass effect on surrounding structures, (2) preservation and/or restoration of normal pituitary gland function, (3) effective reconstruction of the sellar floor/skull base, and (4) avoidance of complications (Table 10.2, Fig. 10.3).

KEY POINTS

- A multidisciplinary approach to patient selection is paramount, including assessments from neurosurgery, otolaryngology, endocrinology, radiology, neuroophthalmology, and anesthesiology.

Rathke's Cleft Cyst

The majority of patients harboring RCCs are asymptomatic; however, larger cysts can lead to headache, hormonal disturbances (including hypopituitarism or hyperprolactinemia) or visual loss (Fig. 10.4). When

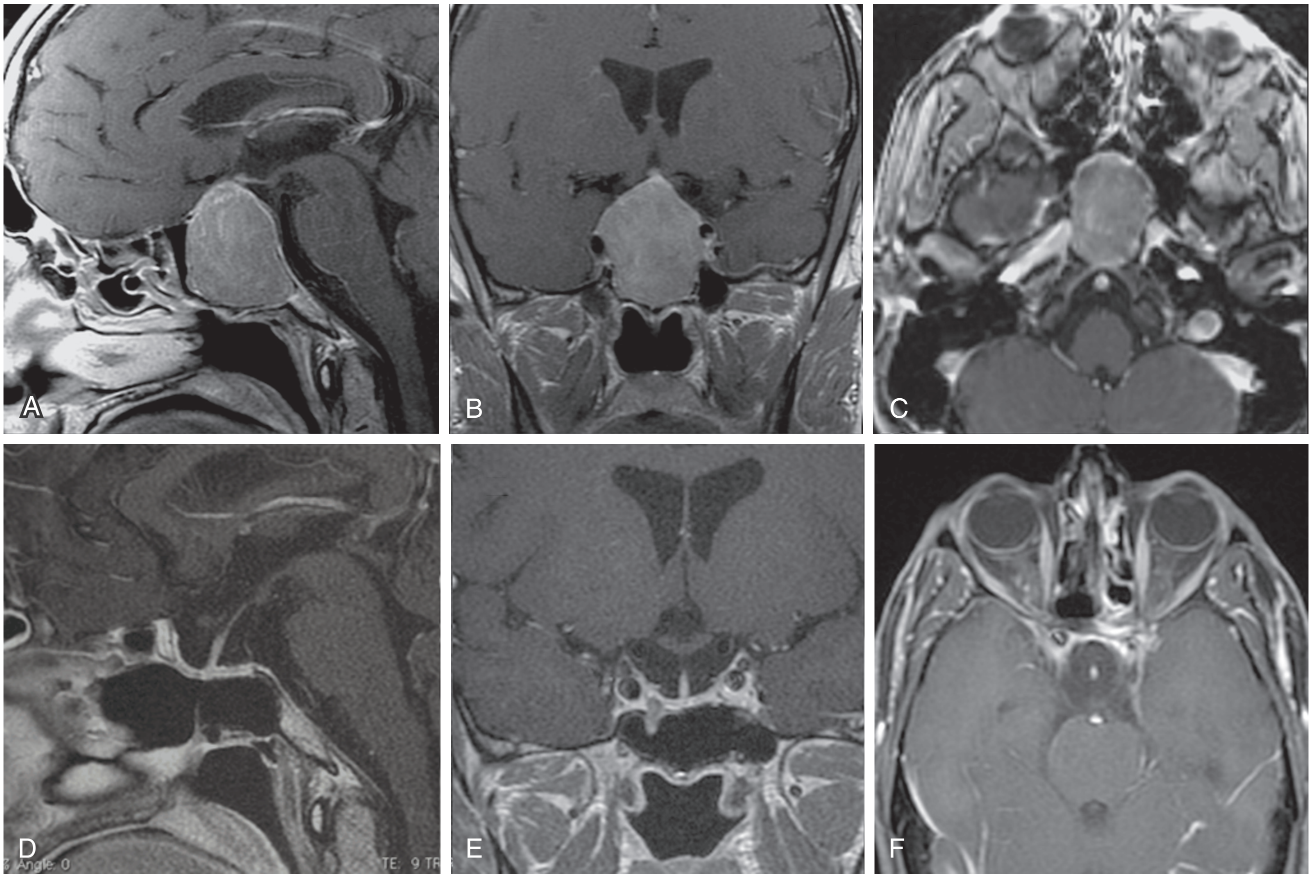


Fig. 10.2 A 51-year old man with an endocrine-inactive pituitary macroadenoma. Preoperative (**A–C**) magnetic resonance imaging (MRI) shows severe gland elevation. His postoperative MRI (**D–F**) 4 years after endoscopic resection shows gross total removal without recurrence and gland reexpansion. His preoperative hypopituitarism resolved after surgery.

TABLE 10.2 Surgical Goals for Pituitary Surgery

- Biochemical remission for hypersecreting endocrinopathies
  - Acromegaly/gigantism (GH)
  - Cushing disease (ACTH)
  - TSH-secreting adenoma
  - Prolactinoma (PRL; refractory to medical management)
- Relief of mass effect
  - Hypopituitarism
  - Vision loss
  - Other cranial neuropathies
  - Headache
  - Hypothalamic dysfunction
  - Pituitary apoplexy
- Preservation of normal pituitary gland function
- Effective closure of sella and/or skull base defect
- Obtain tissue diagnosis
- Avoidance of complications

GH, Growth hormone; *ACTH*, adrenocorticotropic hormone; *PRL*, prolactin; *TSH*, thyroid-stimulating hormone.

symptomatic, RCCs should be considered for surgery, with the goals being to fenestrate and remove the cyst contents, relieve the mass effect, preserve or restore normal pituitary gland function, and avoid surgical complications (Table 10.2). In our experience, only approximately 40% of patients presenting with a RCC warrant surgery, and the remainder can be observed, given the small size of the cysts and lack of symptoms or signs.[11] In our recent study, cyst progression was noted on serial imaging in only 3% of patients who underwent conservative management.

Craniopharyngioma (see also Chapter 11)

Common presenting symptoms include vision loss, hypopituitarism (including diabetes insipidus [DI]), headaches, and cognitive impairment from hydrocephalus, and occasional hypothalamic dysfunction (including hyperphagia and obesity).[12,13] Surgery for craniopharyngiomas is often challenging because of their adherence to vital parasellar structures, including the optic apparatus, vessels of the circle of Willis, infundibulum, and hypothalamus (Fig. 10.5). As these tumors are generally radiosensitive, it is reasonable to attempt maximal safe removal, aim to preserve gland function (if intact preoperatively), and be willing to leave small tumor remnants for relatively neurologically intact patients, followed by adjuvant stereotactic radiotherapy to preserve endocrinological function (Table 10.2). More aggressive attempts at gross total resection with infundibulum removal may be considered for patients with multiple hormonal deficiencies, including DI. In some cases, an Ommaya reservoir may be inserted into a cystic component of the tumor, enabling access for serial cyst aspiration and, in rare cases, delivery of intratumoral chemotherapeutic agents (e.g., interferon-gamma or radioisotopes).[14] As most craniopharyngiomas are located in the retrochiasmal space, an endoscopic endonasal transtubercular approach is ideal for the majority of patients. Otherwise, a transcranial route may need to be considered.

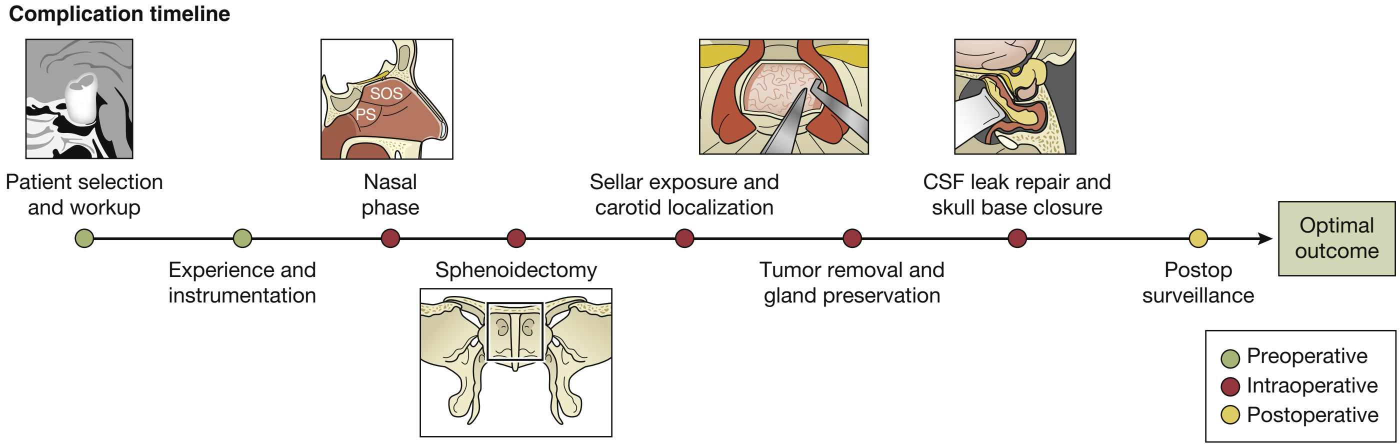

Complication timeline
Patient selection and workup
SOS
PS
Nasal phase
Sellar exposure and carotid localization
CSF leak repair and skull base closure
Optimal outcome
Experience and instrumentation
Sphenoidectomy
Tumor removal and gland preservation
Postop surveillance
Preoperative
Intraoperative
Postoperative

Fig. 10.3 Major steps of transsphenoidal surgery from patient selection and workup to achieving an optimal outcome, with key moments for complication avoidance highlighted. Green circles: preoperative; red circles: intraoperative; yellow circle: postoperative. *CSF*, cerebrospinal fluid.

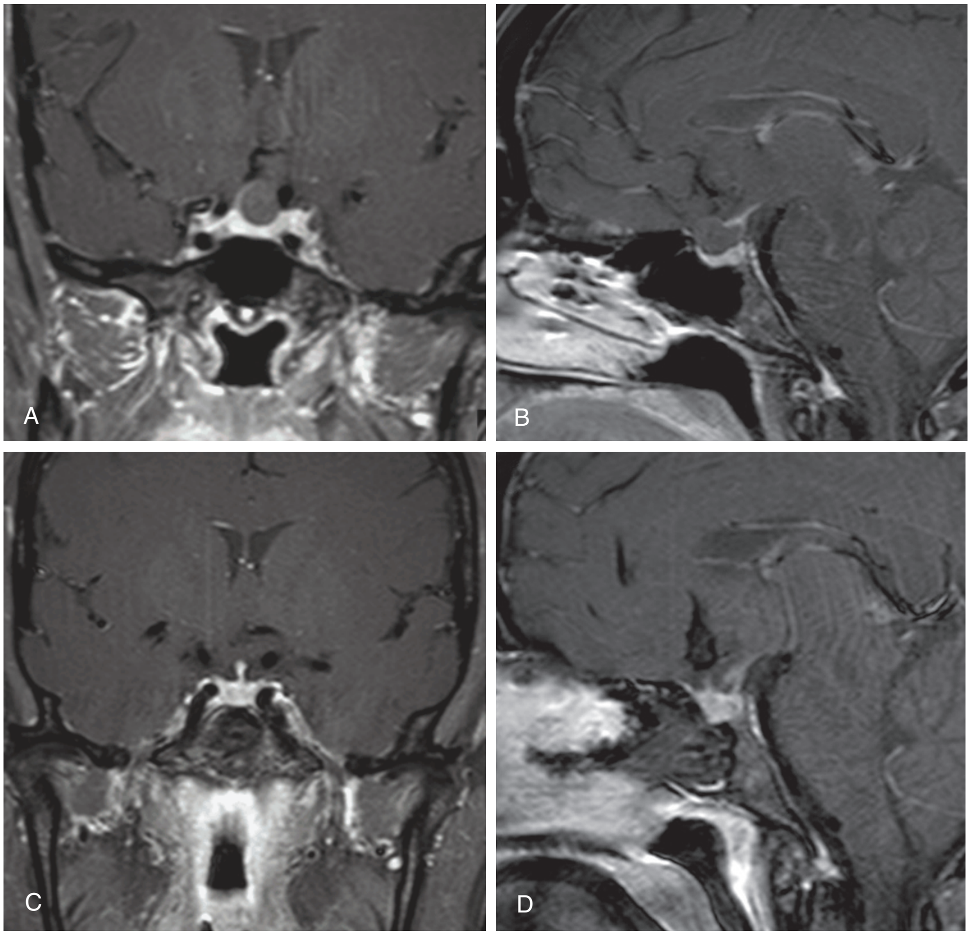


Fig. 10.4 Coronal (**A**) and sagittal (**B**) T1 with gadolinium magnetic resonance imaging (MRI) of an 83-year-old female presenting with intermittent hyponatremia, adrenal insufficiency, low insulin-like growth factor 1, and low gonadotropin levels. A large sellar and suprasellar nonenhancing lesion is seen between the adenohypophysis and neurohypophysis, consistent with a Rathke's cleft cyst. An endoscopic endonasal cyst fenestration was performed, with the postoperative day 1 MRI (**C, D**) showing gland reexpansion. She was placed on adrenal and thyroid replacement, with no further episodes of hyponatremia 2 years postoperatively.

Other Lesions

Although pituitary adenomas, RCCs, and craniopharyngiomas represent the majority of sellar pathologies, other pathologies are important to identify and treat appropriately. Several pathologies may mimic pituitary or sellar tumors, including carotid artery aneurysms, lymphocytic hypophysitis, hypothalamic hamartomas, and pituitary hyperplasia (e.g., primary hypothyroidism) (Fig. 10.6). It is critical to obtain all relevant imaging and consult a neuroradiologist in the context of the clinical presentation to avoid unnecessary surgery.

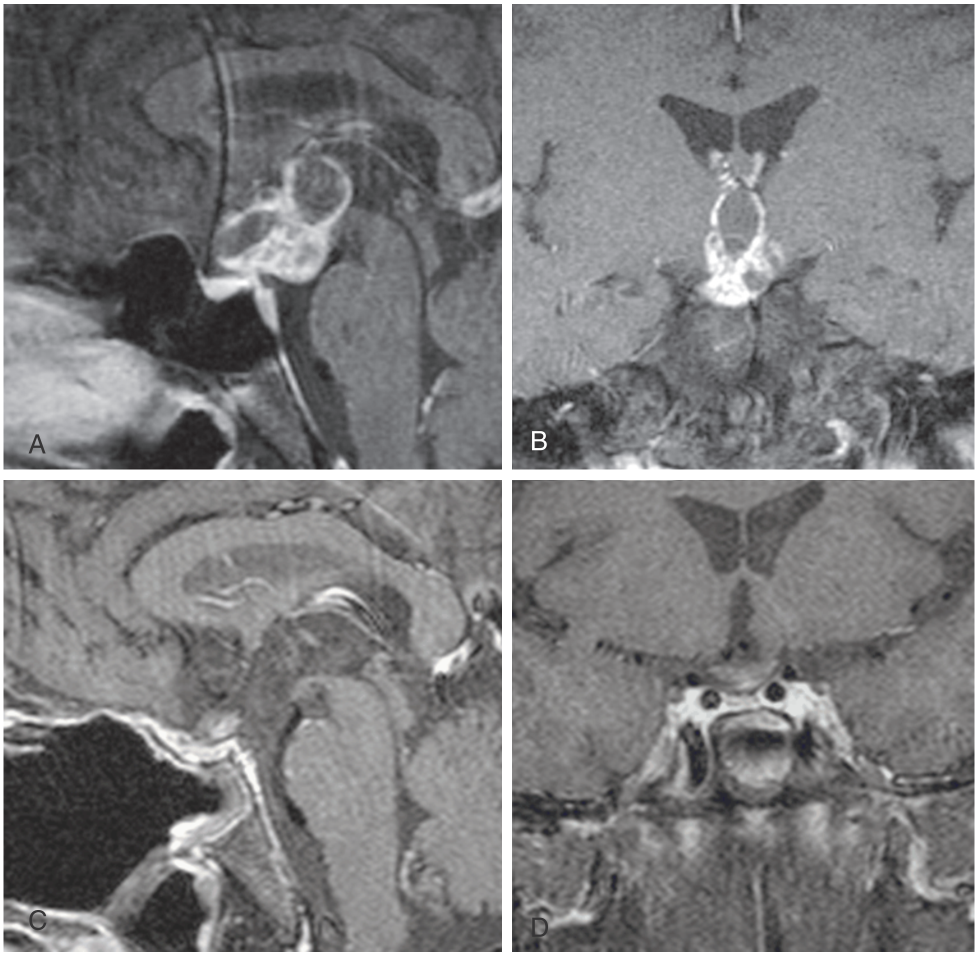


Fig. 10.5 Coronal (**A**) and sagittal (**B**) T1 with gadolinium magnetic resonance imaging (MRI) of a 52-year-old male presenting with persistent cold intolerance, weight gain, recent vision loss, and complete anterior and posterior pituitary failure. A large solid and cystic enhancing mass is seen in the suprasellar and retrochiasmal regions, consistent with craniopharyngioma. An endonasal endoscopic procedure was performed with gross total resection. Postoperative MRI 1 year later (**C, D**) demonstrates no residual or recurrent tumor, with intact nasoseptal flap used for the skull base reconstruction. His vision improved, and he remains on full pituitary hormone physiologic replacement.

PREOPERATIVE MANAGEMENT

Appropriate patient selection contributes directly to surgical outcomes. This includes not only medical fitness for surgery, but also a careful evaluation of any previous interventions, endocrinological status, individual anatomical characteristics, and the surgical goals. A multidisciplinary approach to patient selection is paramount, including assessments from neurosurgery, otolaryngology, endocrinology, radiology, neuroophthalmology, and anesthesiology.

KEY POINTS

- Indications for endoscopic endonasal pituitary surgery include relief of mass effect, hypersecreting endocrinopathy and pituitary apoplexy.

As such, patients diagnosed with a symptomatic pituitary adenoma, RCC, or craniopharyngioma should undergo (1) thorough clinical history and physical examination, (2) comprehensive pituitary hormonal assessment, (3) pituitary MRI with gadolinium (or computed tomography [CT] sellar imaging if MRI is not available or the patient has an MRI-incompatible implant such as a pacemaker), and (4) medical assessment for anesthesia safety, routine preoperative testing, and bloodwork. In patients with chiasmal or optic nerve compression (without or with visual complaints), an ophthalmology evaluation, preferably with a neuroophthalmologist, with visual field testing and ideally optical coherence tomography is obtained. Cardiac or pulmonary function testing may also be indicated depending upon a given patient's comorbidities. In particular, patients with advanced Cushing disease and acromegaly may have significant diabetes mellitus, hypertension, and cardiac dysfunction, including cardiomyopathy, arrhythmias, or valvular disease, warranting additional evaluation.[15] Preoperative panhypopituitarism, including DI, also warrants careful preoperative evaluation. Patients with acromegaly may also pose significant intraoperative challenges associated with difficult intubation from macroglossia, prognathism, and hypertrophy of laryngeal tissue.[16] An individualized approach is necessary to minimize perioperative complications.

Endocrine Status

Regardless of the symptomatology or pathology, all sellar/parasellar lesions that present with a consideration for surgery undergo a comprehensive evaluation of the hypothalamic-pituitary-adrenal axis. This includes assessment of the adenohypophysis (fasting morning cortisol and ACTH, GH, IGF-1, prolactin [PRL], TSH, total and/or free T4, luteinizing hormone, follicle-stimulating hormone, and testosterone/estrogen based on gender) and neurohypophysis (serum sodium, urine

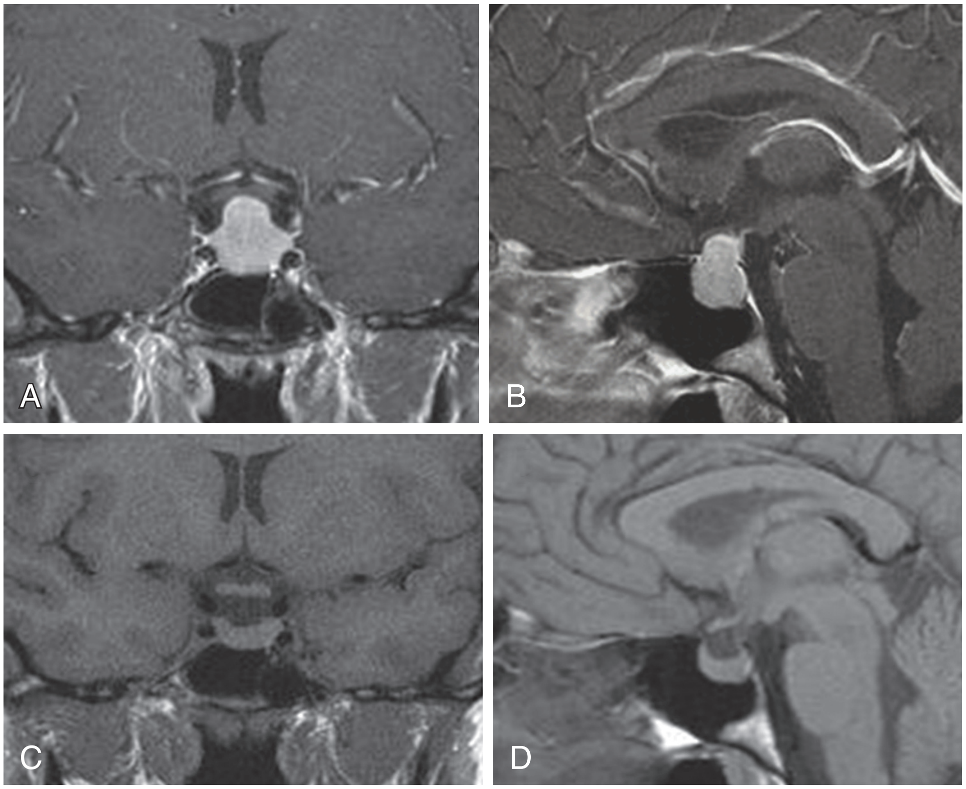


Fig. 10.6 Coronal (**A**) and sagittal (**B**) T1 with gadolinium-enhanced magnetic resonance imaging sequences of a 22-year-old female who presented with fatigue, headache, and galactorrhea and was found to have a serum prolactin level of 139 ng/mL. Preliminary diagnosis was of a pituitary adenoma, but during the workup she was found to have primary hypothyroidism with an elevated thyroid-stimulating hormone level of 1576 mIU/L and free T4 below 0.4 ng/dL. She was started on levothyroxine with normalization of serum free T4 level and resolution of pituitary hyperplasia (**C, D**).

osmolality/specific gravity, urine output). This provides a baseline endocrine status for subsequent postoperative assessments. There are specific endocrinological testing protocols to complete when evaluating patients suspected of having an endocrine-active adenoma, including Cushing disease, acromegaly, prolactinomas, and thyrotropinomas, which are described in other chapters in this text. However, of particular importance for surgery is to evaluate for hypocortisolemia, hypothyroidism, and DI, each of which, if not addressed preoperatively, can have a serious impact on the patient. Perioperative stress dosing with glucocorticoids is mandatory for patients with proven adrenal insufficiency. However, patients with either a normal adrenal axis or Cushing disease generally do not need preoperative glucocorticoids and instead can have cortisol levels monitored in the early postoperative period. Preoperative glucocorticoid administration interferes with the assessment of early biochemical remission in patients with Cushing disease.[17,18] If hypothyroidism is present, supplementation should not occur before corticosteroid supplementation so as to avoid an unsupported increase in metabolic rate leading to Addisonian shock.[19] DI may be treated with desmopressin (DDAVP) to correct hypernatremia, as described in Chapter 12. Chronic sodium aberrations from DI should be corrected no faster than 0.5 mEq/hour to avoid complications of cerebral edema.

Imaging Evaluation

All patients undergoing pituitary surgery should have a pituitary MRI with gadolinium and, in some instances, a thin-slice CT of the paranasal sinuses. The MRI is used primarily to identify the likely pathology, in correlation with the clinical and endocrinological findings, but also to determine the most appropriate surgical route and potential hazards or pitfalls. These include identification of the normal pituitary gland and infundibulum, relationship of critical nearby structures to the pathology, and extent of invasion into the CSs. An optimal MRI protocol includes T1, T1 with gadolinium, and T2 sequences with coronal and sagittal reformats. Fat saturation techniques are important to reduce the air artefact of the adjacent sphenoid sinus. A dynamic pituitary MRI study performs repeated T1 sequences after the administration of gadolinium and is helpful in identifying microadenomas because they tend to absorb the contrast later than normal pituitary tissue.[20] This is particularly helpful in patients with suspected Cushing disease (see also Chapter 7). A sequence with thin-slice thickness and spacing, such as magnetization prepared-rapid gradient echo, is helpful for registering an intraoperative neuronavigation systems. A dedicated thin-cut contrast-enhanced CT of the sella with coronal and sagittal reconstructions should be used in patients with contraindications to MRI.

The paranasal sinuses may be appreciated on a CT scan and are studied for nasal septum deviations, air cell anatomical variations, areas of bony dehiscence, the degree of sphenoid sinus pneumatization, and evidence of sellar expansion or bony changes indicating pathology diagnosis and chronicity. Together, MRI and CT allow for the planning of skull base reconstruction to minimize the risk of a postoperative CSF leak.

SURGICAL TECHNIQUE

Several surgical approaches to the pituitary gland and sellar region have been described. These are broadly divided into endonasal approaches and transcranial approaches. Once a patient has been deemed appropriate for surgery, careful consideration is given toward achieving the specific surgical goals (Table 10.2) and clearly understanding patient-specific anatomical factors. Today, the endoscopic endonasal approach is the predominant technique used for all sellar and parasellar tumors, and skull base neurosurgeons have largely transitioned from microscopic approaches.[6]

KEY POINTS

- Goals for endoscopic endonasal pituitary surgery include biochemical remission for hypersecreting endocrinopathies, relief of mass effect, preservation of normal pituitary gland function, effective closure of sella and/or skull base, obtaining a tissue diagnosis, and avoidance of complications.

In fact, graduates from many surgical training programs today are only trained in endoscopic techniques. Transcranial procedures are rarely indicated for pituitary adenomas and RCCs, but may be preferable in a significant minority of craniopharyngiomas based on anatomical factors, including location relative to the optic chiasm.

Endoscopic Endonasal Approach for Pituitary Adenoma, Rathke's Cleft Cyst, and Craniopharyngioma

The endonasal transsphenoidal approach provides a binostril direct trajectory to the sella with minimal tissue to transgress *en route*, passing through the nasal cavities and sphenoid sinus to reach the sella (see Fig. 10.1). There are many subtle variations to the procedure, but it is increasingly performed with both a neurosurgeon and otolaryngologist working together (Fig. 10.7). The major phases of the operation

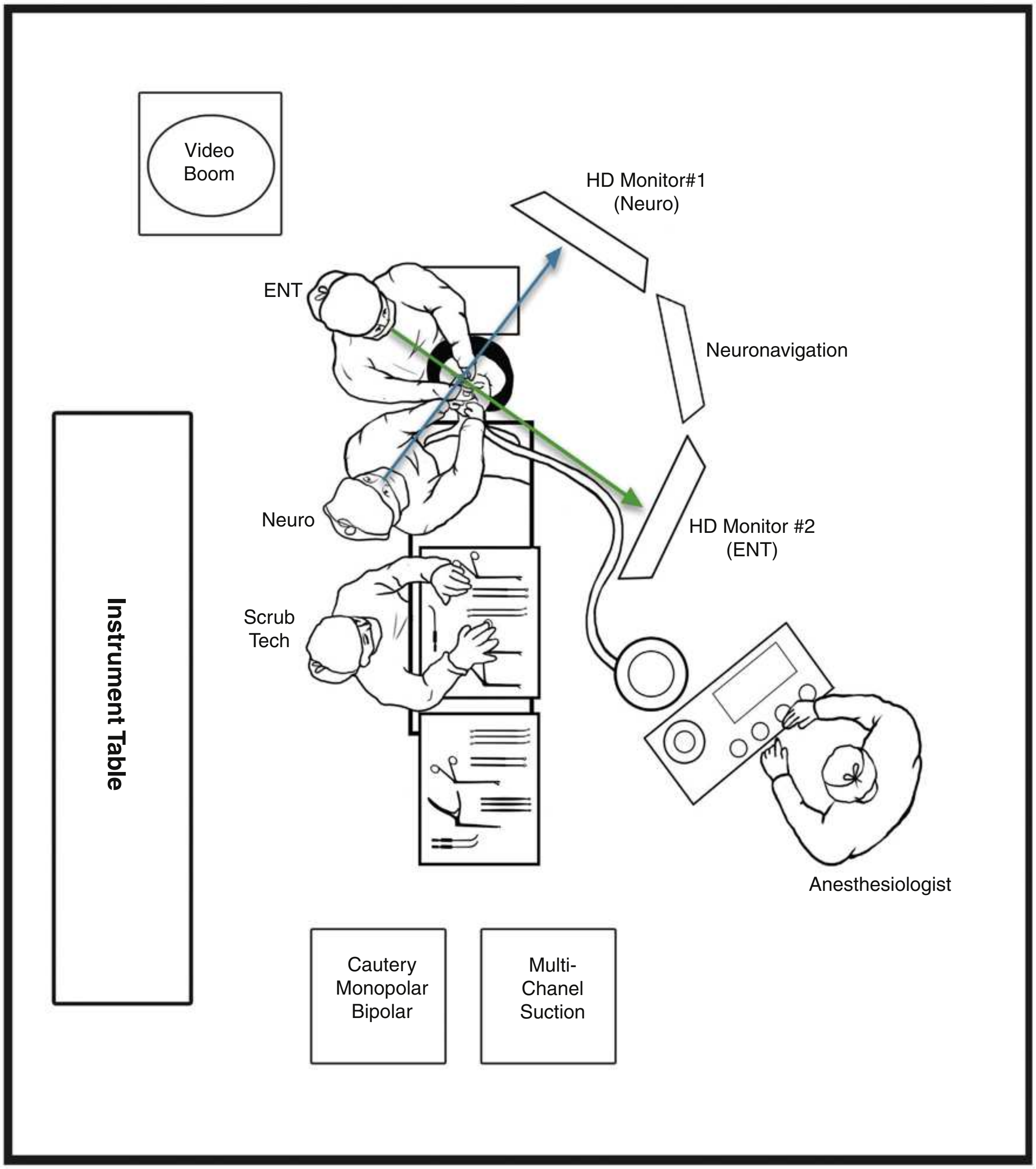


Fig. 10.7 Diagram of the typical operating room setup for endoscopic endonasal surgery. The procedure involves an otolaryngologist and neurosurgeon, both positioned on the patient's right side. Ergonomics are optimized by having an endoscope monitor positioned directly in front of each surgeon, as well as an arm rest for the otolaryngologist's left arm, holding the endoscope. *ENT,* Ear, nose and throat doctor; *HD,* high-definition.

include: (1) the sinonasal approach through the nasal cavities and sphenoid sinus, (2) the sellar exposure, (3) tumor or cyst removal, and (4) skull base reconstruction.

For all sellar and parasellar lesions, the patient is positioned supine on the operating room table with the head being slightly extended and laterally flexed for optimal surgical ergonomics. Most of the operation is performed with a 0-degree rigid 4-mm high-definition endoscope.

For the sinonasal approach, the nasal cavities are gently expanded through each nostril by out-fracturing, but not removing, the middle turbinates, then elevating bilateral nasoseptal mucosal rescue flaps. These strategically placed mucosal incisions allow exposure of the sphenoid keel bone while preserving bilateral posterior nasoseptal arteries and septal olfactory strips, thereby dramatically reducing the risk of early postoperative epistaxis and anosmia (Fig. 10.8).[21] The posterior aspect of the nasal septal bone is removed, and a wide sphenoidotomy is performed, connecting the two nasal corridors, allowing a binostril approach.

Next, removal of the sellar bone is performed with a combination of high-speed drills and rongeurs. This opening is tailored to the specific pathology based on preoperative imaging. A critical phase in this sellar exposure is localization of the parasellar ICAs using a combination of intraoperative neuronavigation and the Doppler probe (Fig. 10.9).

The sellar dura is then opened sharply, and the adenoma is localized and removed. In many cases, an attempt is made to use the adenoma pseudocapsule, which helps facilitate complete, *en bloc* removal. This technique, as well as minimizing traction on the gland, and making selective incisions in attenuated gland, can optimize adenoma access while minimizing risk of pituitary gland injury and postoperative hypopituitarism. For adenomas with likely CS invasion, or high suprasellar extension, visualization and accessing such tumor is performed with 30-degree and 45-degree endoscopes.

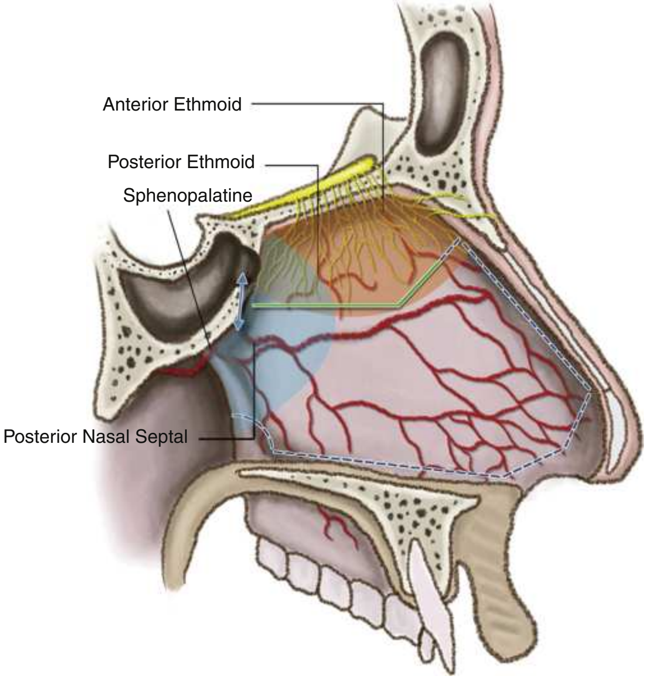


Fig. 10.8 Lateral view of the nasal septum depicting the mucosal cut made for elevating a nasoseptal rescue flap *(solid line)* during a standard endoscopic transsphenoidal surgery, preserving olfaction by sparing the septal olfactory strip, above which contains the olfactory nerve fibers. If needed, a nasoseptal flap may be harvested *(dashed line)*, supplied by the posterior septal artery of the sphenopalatine artery.

Finally, skull base reconstruction is performed to prevent postoperative CSF leak and meningitis. The specific repair protocol is dependent upon the grade of the CSF leak, with larger diaphragmatic defects warranting a more extensive repair.[22] An abdominal fat graft is used frequently during reconstruction to obliterate any sellar dead space, especially when an intraoperative CSF leak is identified. In our practice, nasoseptal flaps are used in only 2% to 3% of pituitary adenoma and RCC cases, but in the majority of craniopharyngioma operations.

Rathke's Cleft Cyst. The sellar approach to RCCs is generally identical to that of pituitary adenomas. Once the sellar dura is opened, a low vertical incision in the adenohypophysis is made to reach the cyst, given that most RCCs are located posterior to the adenohypophysis. The pituitary gland is resilient to such incisions, with generally no resultant increase in permanent hormonal dysfunction, as we recently demonstrated.[11,23] Cyst contents are then evacuated, and angulated endoscopes are used to gently explore the cyst lining to ensure complete removal. No attempt is made to strip the cyst lining, given that this maneuver can lead to permanent hypopituitarism. Skull base reconstruction then commences as described earlier.

Craniopharyngioma. For craniopharyngiomas, the extent of growth into the suprasellar cisterns dictates the extent of bony removal to access the lesion. In many cases, an extended transsellar, transtubercular approach is needed. The transsellar portion is the same as described previously, while the transtubercular component requires additional bone removal of the tuberculum sellae and a small portion of the planum sphenoidale. The dura and arachnoid are opened to expose the pituitary gland, infundibulum, and suprasellar cistern. In most instances of a retrochiasmal craniopharyngioma, tumor removal requires working through a narrow intracranial corridor above the pituitary gland and under the optic chiasm. This approach allows excellent visualization of the infundibulum and superior hypophyseal arteries, which require preservation if the gland and optic apparatus

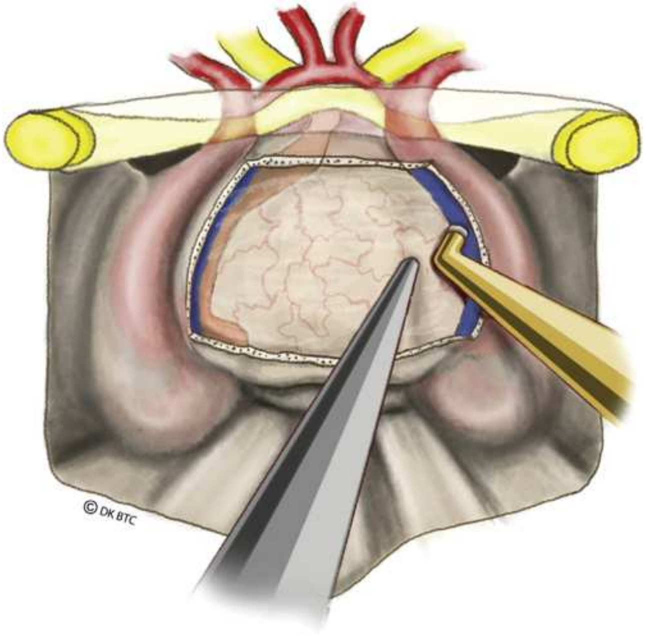


Fig. 10.9 To reduce the risk of vascular injury, the Doppler probe is used frequently during endoscopic endonasal surgery to identify the cavernous carotid arteries and other major vessels in real time during the approach to the sella and parasellar space. In this drawing, the left internal carotid artery is identified here during a transsphenoidal approach.

are to remain functional. After maximal safe tumor removal, skull base reconstruction proceeds, typically using an abdominal fat graft, collagen matrix, nasal septal bone graft, and a vascularized nasoseptal flap. This more extensive skull base repair is needed given the large diaphragmatic grade 3 CSF leak produced by the tumor removal itself.[22] Nasal sponges are generally placed for 5 to 6 days to ensure that the nasoseptal flap adheres well to the skull base. In our experience, lumbar drainage/CSF diversion is not necessary for such extended transsphenoidal approaches, although these are often used at other institutions.

ACUTE POSTOPERATIVE MANAGEMENT AND EARLY OBSERVATION POSTDISCHARGE

At our institution, most postoperative patients undergo a head CT in the recovery room before going to a step-down unit bed, to assess the skull base repair integrity. Patients with pituitary adenomas or RCCs generally go to a step-down unit and, length of stay is typically 1 to 2 nights. Patients with a craniopharyngioma or a giant adenoma with large supradiaphragmatic extension will typically spend one night in the intensive care unit, then another night or two in the step-down unit. A sellar MRI is performed on postoperative day 1. Perioperative antibiotics and nasal care are administered, with varying protocols employed based on whether nasal sponges were inserted during surgery.

After discharge from hospital, patients will have an outpatient assessment of their serum sodium level on postoperative day 5 to 7 to rule out delayed hyponatremia, as well as serum cortisol, ACTH, and any other relevant hormones (e.g., serum PRL for prolactinomas). Follow-up appointments will be arranged with otolaryngology on postoperative day 5 or 10 (depending on the presence or absence of nasal sponges), neurosurgery on postoperative day 10, and endocrinology, neuroophthalmology, neurooncology, and others as needed. A pituitary MRI is generally performed approximately 3 months after surgery, with subsequent imaging surveillance based on pathology and residual disease.

Endocrinological monitoring in the acute postoperative phase is focused on three potential events: (1) DI and sodium fluctuations (syndrome of inappropriate antidiuretic hormone [SIADH]), (2) acute adrenal insufficiency, and (3) early remission in endocrine-active adenomas.

KEY POINTS

- Endocrinological monitoring in the acute postoperative phase is focused on three potential events: (1) diabetes insipidus and sodium fluctuations (syndrome of inappropriate antidiuretic hormone), (2) acute adrenal insufficiency, and (3) early remission in endocrine-active adenomas.

Diabetes Insipidus

All patients are closely monitored for postoperative DI with hourly recordings of fluid intake and output (if a Foley catheter is in place) or with each void. An output of more than 400 mL in a single hour or greater than 200 mL/hour when averaged out between voids prompts an assessment of serum sodium and urine specific gravity. Patients are encouraged to drink to thirst. DDAVP is administered intravenously if the specific gravity is below 1.005 and the patient is hypernatremic or unable to maintain oral fluid intake. DI may be transient or permanent, which may ensue from a triphasic response. The incidence of transient DI after transsphenoidal surgery is highly variable in the literature (between 1% and 60%, and typically around 20%) and often occurs within the first 5 days after surgery, resolving by postoperative day 11 in most cases.[24,25] Persistent DI (greater than 3 months) has been reported in up to 10% (ranging from 0% to 3.8% in our series), with risk factors including increased tumor size and surgery for RCCs or craniopharyngiomas.[23,24,26,27]

Syndrome of Inappropriate Antidiuretic Hormone

SIADH is a relatively common occurrence after pituitary surgery, with a reported incidence of between 2% and 35%.[28] The peak time of onset is on postoperative days 7 and 8, but can occur anytime between days 4 and 12.[29] While the pathogenesis remains uncertain, a recent study by Burke et al. compared a prophylactic fluid restriction of 1 L/day for 1 week after discharge from hospital to a retrospective control cohort without fluid restriction. They found that none of the 203 patients in the fluid restriction cohort required readmission for hyponatremia, compared with 3.41% of the 585 in the control cohort.[30] We have implemented this into our practice with success. Postoperatively, we obtain a baseline serum sodium level, followed by a repeat level on postoperative day 5, at around the time of peak incidence. We ensure that we educate the patient of the symptoms of hyponatremia, including headache, confusion, nausea, vomiting, fatigue, and muscle cramps. Should hyponatremia occur, an evaluation for adrenal insufficiency and hypothyroidism may be necessary, based on the clinical scenario. No consistent risk factors have been identified for the onset of delayed hyponatremia, but some studies have suggested increased tumor size, female gender, and advanced age to be possible risk factors.[31-39]

Early Remission

Patients with acromegaly, Cushing disease, prolactinomas, and thyrotropinomas are assessed for early biochemical remission while in hospital.[18,27,40] For acromegaly, postoperative day (POD) 1 and 2 morning GH levels are drawn, and early remission is a POD 1 or 2 GH level of 1 ng/mL or less and a subsequent normal sex- and age-adjusted IGF-1 at least 3 months after surgery. Patients with Cushing disease have serum cortisol and ACTH levels monitored at 8-hour intervals. Early remission is defined by a POD 1 or 2 morning cortisol level less than 5 g/dL and needing cortisol replacement for at least 3 months postsurgery. Stress-dose replacement hydrocortisone is administered in hospital once this cortisol drop occurs, followed by a taper to physiological replacement, typically within 3 to 4 days of surgery which is then maintained and slowly tapered off over 6 to 12 months postsurgery. For prolactinomas, early remission is defined as a POD 1 or 2 morning PRL level less than 10 ng/dL and normal serum PRL level (<20 ng/mL for men and <25 ng/mL for women) at least 3 months postsurgery without dopamine-agonist therapy. For thyrotropinomas, remission is defined as a normal TSH and free T4 at least 3 months postsurgery.

OUTCOMES AFTER TRANSSPHENOIDAL PITUITARY ADENOMA SURGERY

At high-volume pituitary centers with experienced surgeons, pituitary adenoma surgery generally has a high success rate in terms of endocrine remission rates, preservation of hormonal function, visual recovery, and headache resolution, with low complication rates.

KEY POINTS

- At high-volume pituitary centers with experienced surgeons, pituitary adenoma surgery generally has a high success rate in terms of endocrine remission rates, preservation of hormonal function, visual recovery, and headache resolution, with low complication rates.

Improvement in visual acuity and visual fields is typically seen in 75% to 90% of patients presenting with bitemporal or monocular vision loss associated with an adenoma. Similarly, headache resolution occurs in 70% to 90% of patients who present with preoperative headache.[27,41,42] As shown in Table 10.3, complications rates in recent large retrospective series of pituitary adenoma patients undergoing

TABLE 10.3 **Major outcomes and complication rates after endoscopic pituitary adenoma resection in recent large series.**

Study	Year	Study type	N	EOR (%)	LOS (days)	% of Redo Surgery in cohort	Intrasellar Hematoma (%)	Vascular injury (%)	Stroke (%)	Vision Decline (%)	New CND (%)	CSF Leak (%)	Meningitis (%)	New Endocrine Dysfunction (%)	Epistaxis (%)	Death (%)
Paluzzi[43]	2014	R	555	GTR 65.3	NR	17.5	1.1	0.3	0.2	2.5	0.6	5	0.9	A 3.1; P 2.5	1	0.2
Magro[44]	2016	R	300	GTR 59	NR	14	2	0.3		2.43	2.8	2.7	3.3	A 13.7; P 6.2	2.3	0.7
Eseonu[45]	2017	R	275	Mean EOR 85.1	2.4	16.1	0	0.4	0.73	5.49	0.36	3.6	0.7	Total 3.3; P 0.7	0.36	0
Kim[46]	2018	R	331	GTR 74.9		29.3	2.4	0	0	2.7	NR	2.4	5.4	A NR; P 3	NR	0.3
Nayak[47]	2018	R	271	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	A NR; P 4	NR	NR
Little[48]	2019	Pr	169	GTR 79.3	NR	8.3	NR	NR	NR	NR	NR	NR	NR	Total 9.7; P 2.4	NR	NR
Younus[49]	2019	R	583	GTR 72.9	4	12.8	1.9	NR	NR	3.1	NR	NR	NR	NR	NR	NR
Gerges[50]	2020	R	190	GTR 66.8	NR	10	1			2.6	NR	3.2	NR	A 18.9; P 8.4	NR	NR
Thakur[51]	2021	R	514	GTR 69 EndoR 71	2 (median)	19	0.2	0.2	0.2	0	0.2	1.7	0.8	A 3.5;P 2.3	0.6	0.2

N – number of patients; *EOR* – extent of resection; *LOS* – length of stay; *CSF* – cerebrospinal fluid; *R* – retrospective; *Pr* – prospective; *NR* – not reported; *GTR* – gross total resection; *NTR* – near total resection; *A* – anterior; *P* – posterior; *EndoR* – endocrine remission *CND* - cranial nerve deficit endonasal endoscopic surgery are quite low, ranging from 0% to 2% for carotid artery injury, visual worsening, meningitis, and sellar hematoma; the rate of new hypopituitarism typically is under 8% to 10%.[43-50]

MAJOR COMPLICATIONS

Permanent Hypopituitarism

New permanent anterior and posterior hypopituitarism after pituitary adenoma removal is generally low in experienced hands at high-volume centers, ranging on average from 2% to 7%, and, as we have demonstrated, is most dependent upon tumor size.[27,48,52-59] For adenomas under 2 cm, the risk of new hypopituitarism is typically 1% to 5%, but with adenomas over 3 cm (including giant adenomas), the rate may range from 10% to 15%.[60] Similarly, the rate for new hypopituitarism for RCC is also quite low, ranging from 0% to 4%, and in our recent series of 36 patients, none had new hypopituitarism.[11] The rate of new permanent hypopituitarism is significantly higher for craniopharyngiomas and giant pituitary adenomas, with reported incidences of up to 58%.[26,61,62] We routinely obtain postoperative morning serum cortisol and ACTH levels to monitor for hypocortisolemia, while other hormone levels are checked at a later date as an outpatient as needed. Important surgical concepts that may reduce the risk of anterior hypopituitarism include preoperative identification of the normal gland on MRI, respecting the tumor pseudocapsule (when present), avoiding undue traction to the gland with gentle manipulation, and using selective gland incisions to reduce gland traction and manipulation.

Cerebrospinal Fluid Leak and Meningitis

Avoiding postoperative CSF leaks remains a key goal in transsphenoidal adenoma surgery, and the incidence in most large recent series is under 5%. In our recent large series of endoscopic endonasal skull base repair including 362 pituitary cases, the postoperative CSF leak rate was 1.4%, and the meningitis rate was 1.1%.[22] Our experience, and that of others, suggests that a graded repair protocol based on the size of the skull base defect is highly effective for most patients. However, patients with a higher body mass index (generally >30) are at increased risk of postoperative CSF leak, and we have a low threshold to place nasal packing as a temporary buttress in such patients. Additionally, we place all patients with an intraoperative leak (grade 1, 2, or 3) on acetazolamide 250 mg by mouth every 8 hours for 2 days. Should a postoperative leak be discovered while the patient is in hospital or after discharge, in most instances reoperating to revise the repair is indicated, although in some patients temporary CSF diversion with a lumbar drain is reasonable.[22] This is assessed via a "tip test" or "tilt test", where the patient leans forward to observe for clear nasal rhinorrhea. The fluid is assessed for βeta-2 transferrin, which is very specific for CSF. The separation of blood from the draining fluid on gauze or paper, known as the halo sign, has been shown to be neither sensitive nor specific for CSF.[63] Confirmatory tests may be necessary in some occult cases of persistent postoperative CSF rhinorrhea, including CT myelogram or nuclear scintigraphy myelogram with nasal pledget radionucleotide detection.

Should patients develop signs and symptoms of meningitis postoperatively, such as nuchal rigidity, headache, decreased level of consciousness, photophobia, and fever, a timely workup should commence. This includes serum white blood cell count, erythrocyte sedimentation rate, C-reactive protein, electrolytes, and blood culture, as well as a lumbar puncture to evaluate for cell counts, protein, glucose, gram stain, and culture. MRI with gadolinium of the brain should be obtained to assess for potential infection location, as well as hydrocephalus, which may need to be treated with either an external ventricular drain or a lumbar drain. Ideally, broad-spectrum antibiotics should be administered once serum and CSF samples have been collected for culture, but this should not be delayed. Current guidelines suggest the use of vancomycin and a third- or fourth-generation cephalosporin for adequate CSF penetration.[64] Once an organism is isolated, the antibiotics may be tailored to the susceptibility profile of the pathogen. Duration of antibiotic use is usually 14 days, or 21 days for gram-negative bacilli. In some cases, cultures may return negative, suggesting possible aseptic meningitis. Clinical acumen must be relied upon to decide whether or not to continue antibiotics, which takes into account the clinical scenario, CSF analysis, and imaging findings.

Cranial Nerve Injury

Although an uncommon complication, the optic nerves and chiasm, abducens, oculomotor, and trochlear cranial nerves may be at risk during surgery, especially for tumors such as invasive pituitary adenomas and craniopharyngiomas that may invade the CSs or extend superiorly beyond the sellar diaphragm toward the optic apparatus.[65,66] Every artery of the circle of Willis supplies small branches to the optic apparatus. Most relevant to endoscopic endonasal pituitary surgery are the superior hypophyseal and anterior cerebral arteries, which must be meticulously dissected to prevent visual pathway impairments. Craniopharyngiomas frequently extend through the suprasellar cisterns and may be adherent to these arteries. These challenging tumors are associated with a 2% to 8% risk of new or worsening vision deficit.[61,67-70] For routine pituitary adenomas and RCCs, new or worsened diplopia from a cranial neuropathy occurs in less than 1% of patients.[11] The risk of cranial neuropathy, especially abducens, oculomotor, and trochlear nerves, increases as the pathology and surgical corridor extend further laterally into the CSs or inferiorly along the brainstem, requiring a transclival approach.[71] For such cases, intraoperative cranial nerve monitoring should be used.

Postoperative Intrasellar Hematoma

As with any surgical procedure, meticulous hemostasis is required to minimize the risk of hematoma formation. In pituitary surgery, the risk is greatly increased when residual adenoma remains. Hematoma formation may result in postoperative headache, cranial neuropathy, vision loss, or hydrocephalus. Patients with a coagulopathy are especially prone to hematoma formation and should be optimized medically before surgery. The presence of a hematoma in some cases warrants a return to the operating room for evacuation, particularly if large or symptomatic.

Internal Carotid Artery Injury

Injury to the ICA is a serious, but fortunately rare, complication, and in most large series occurs in less than 1% of patients.[10,43-46] The segment within the CS is the most common site of injury, with a reported injury rate of 1.1%.[72,73] Catastrophic bleeding may occur, risking stroke or death, necessitating immediate and definitive action to mitigate. Once the bleeding has been stabilized, the procedure is aborted, and the patient is brought to the angiography suite for evaluation of arterial injury and/or pseudoaneurysm formation. In many cases either a covered stent or flow-diverting stent is used, with an approximately 90% success rate.[74] At our institution, we have implemented a "carotid injury timeout" in conjunction with a standard operative timeout for high-risk procedures. These include patients with previous surgery, previous radiation, or invasive parasellar tumors that incorporate a segment of the ICA. Additional equipment is made available in the room to deal with a carotid artery injury, including essential instruments, backup equipment, medications, and crossmatched blood. The neurointerventional team is notified before the operation. With the use of an intraoperative navigation system and a Doppler probe, the incidence of ICA injury is dramatically reduced.[75]

TABLE 10.4 **Remission Rates for Pituitary Adenomas After Endoscopic Endonasal Resection**

	SURGICAL REMISSION RATE (%)	
Tumor Type	**Microadenoma**	**Macroadenoma**
Acromegaly	70–90	50–75
Cushing Disease	65–95	55–95
Prolactinoma	80–95	40–90
Thyrotropinoma (Thyroid-Stimulating Hormone-Secreting)	60–90	
Endocrine-Inactive	55–90	

RESECTION AND REMISSION RATES

Pituitary Adenoma

Overall rates of gross total resection for endocrine-inactive adenomas range from approximately 55% to 90% (Table 10.4).[76-79] Risk factors for near-total or subtotal removal include CS invasion, supradiaphragmatic extension, giant adenoma (≥4 cm in maximal diameter), and prior surgery. In these cases, the surgical goal is still maximal safe removal with pituitary gland and optic apparatus decompression. Serial imaging must be performed as growth of the residual tumor has been observed in as many as half of patients after 3 years[80] (see also Chapter 8).

Similar to endocrine-inactive adenomas, for endocrine-active adenomas, remission rates are significantly higher for microadenomas associated with acromegaly, Cushing disease, prolactinomas, and TSH-secreting adenomas than for macroadenomas. CS invasion is the most important factor precluding endocrine remission, even with the use of endoscopy. As described in other chapters, for large macroadenomas or invasive adenomas that cannot be completely removed, adjuvant therapies including medical therapies and/or radiosurgery or stereotactic radiotherapy may be needed to achieve remission and to halt tumor progression.

Rathke's Cleft Cyst

Recent surgical series of RCCs report high rates of improvement of headaches (71%–100%), visual disturbances (64%–100%), and hyperprolactinemia (67%–100%) but less successful rates of recovery of pituitary gland function, both anterior (18%–70%) and posterior (0%–50%).[11,81-83] A recent metaanalysis found an overall recurrence rate for these lesions of 12.5%, with recent series reporting recurrence rates from 0% to 21%.[84-86] Squamous metaplasia found on histopathological examination has been shown to increase the rate of recurrence.[34,82,86] Gross total removal of the cyst wall does not improve recurrence rates compared with cyst drainage with fenestration, but does appear to increase the rates of new endocrinopathy. The rate of new hypopituitarism has been reported at 0% to 67% in various series, but is typically around 5% to 15% when utilizing the cyst fenestration approach.[84-86] We do not remove the cyst wall, and have had no patients develop new permanent DI or hypopituitarism, and have noted a symptomatic recurrence rate of only 8%.[11] We have also noted 100% of patients with visual loss recovered, and over 80% of patients had an improvement in headache and endocrinopathy.

Craniopharyngioma (see also Chapter 11)

Given their invasiveness to surrounding structures, subtotal removal of craniopharyngiomas is common, and most will require radiosurgery or stereotactic radiotherapy for treatment of residual or recurrent tumor. Recent endoscopic series report complete removal rates ranging from 37% to 75%.[61,67,68] Improvement of vision is high, ranging from 75% to 86% of patients. Given the intimate association of the tumor with the infundibulum, the rate of new or worsened hypopituitarism of the anterior gland is high, ranging from 38% to 47%, and the rate of new DI ranges from 42% to 48%. Other complications of craniopharyngioma surgery include worsened vision (2%–2.5%), CSF leak (3.8%–23.4%), and hyperphagia related to hypothalamic injury, which can occur in a significant minority of adult patients.[61,67,68,87]

An important and somewhat recent understanding in managing craniopharyngiomas is that of the two histological subtypes, adamantinomatous and papillary; while the former is much more common, the latter is nearly always associated with a BRAF V600E missense mutation. Several clinical trials are currently underway examining the efficacy of BRAF inhibitors and mitogen-activated protein kinase inhibitors in recurrent papillary craniopharyngiomas. Currently, there are only case reports and small case series published, but randomized control trials are underway.[88] We have incorporated BRAF therapy at our institution for select papillary craniopharyngioma patients, with good preliminary outcomes.

PITUITARY CENTERS OF EXCELLENCE

Pituitary and related sellar tumors present significant diagnostic and therapeutic challenges to endocrinologists, neurosurgeons, and other involved subspecialties, such as radiation oncologists and neuropathologists. Given their complexity, it is increasingly accepted that, for most patients with a pituitary tumor, optimal treatment is achieved with a multidisciplinary approach from an integrated group of highly experienced specialists.

KEY POINTS

- Pituitary and related sellar tumors present significant diagnostic and therapeutic challenges, and optimal treatment is achieved with a multidisciplinary approach from an integrated group of highly experienced specialists in a Pituitary Center of Excellence.

This concept of *pituitary centers of excellence* (COE) has been promoted for over two decades, including in an article published in 2012 in *Neurosurgery*.[7] In that article we stressed the complexity of managing pituitary tumor patients in terms of both achieving the correct diagnosis and delivering optimal management, and the need for a team approach (Fig. 10.10). We also showed data on the relationship between surgical experience to clinical outcomes and drew parallels from other surgical subspecialties on the importance of experience. In particular, we proposed that a pituitary COE "should fulfill 3 key missions: (1) provide comprehensive care and support to patients with pituitary disorders; (2) provide residency training, fellowship training, and/or continuing medical education in the management of pituitary and neuroendocrine disease; and (3) contribute to research in pituitary disorders."[7] Since 2012, multiple other publications have supported this concept of pituitary COEs, including a 2017 *Pituitary Society* Statement paper outlining their proposed criteria to define a pituitary COE.[89] These criteria outline the ideal training and experience required by neurosurgeons and endocrinologists and emphasized the establishment of a group of expert supporting units in neuroradiology, neuropathology, radiation oncology, neurooncology, neuroophthalmology, and specialized clinical and research nursing. Similar to our original paper in 2012, the *Pituitary Society* also proposed that pituitary COEs should contribute scientific research for the advancement of pituitary

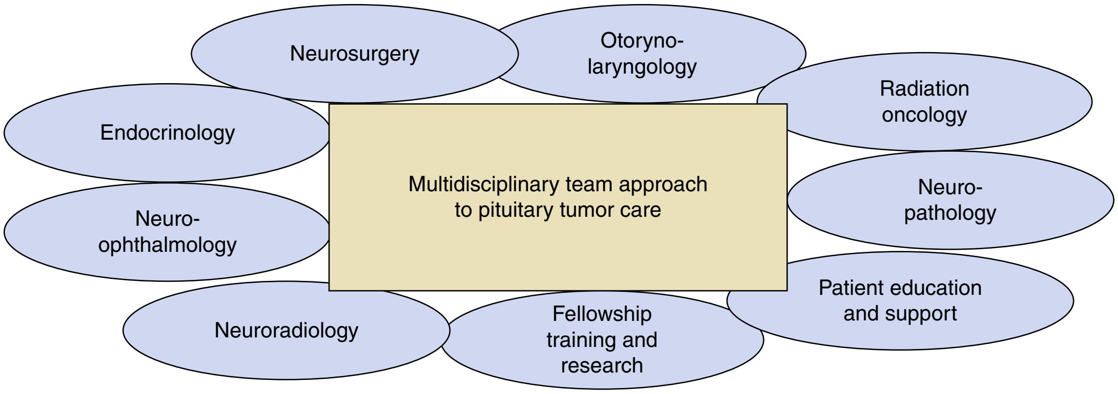

Neurosurgery
Otoryno-laryngology
Radiation oncology
Endocrinology
Multidisciplinary team approach to pituitary tumor care
Neuro-pathology
Neuro-ophthalmology
Patient education and support
Neuroradiology
Fellowship training and research

Fig. 10.10 The concept of pituitary centers of excellence emphasizes the importance of a multidisciplinary approach to pituitary tumor treatment, optimizing patient outcomes. (From McLaughlin N, Laws ER, Oyesiku NM, et al. Pituitary centers of excellence. *Neurosurgery.* 2012;71:916–924; discussion 924–926.)

medicine through clinical trials, peer-reviewed publications, residency and fellowship training, and pituitary-related conferences.

CONCLUSION

Pituitary surgery has evolved substantially since its inception over a century ago. Many advances since then have helped transform the care and outcomes of patients diagnosed with pituitary adenomas, RCCs, craniopharyngiomas, and other pathologies. The endoscope is one such advancement that has become the predominant visualization technique for endonasal pituitary and skull base surgery and has largely supplanted the operating microscope. With an increasing emphasis on a multidisciplinary approach and careful perioperative neuroendocrine monitoring, modern pituitary surgery has well-defined indications and goals that, over recent decades, have resulted in incrementally improved resection and remission rates, visual recovery, and gland preservation rates, as well as reduced incidence of complications. Collectively, these advances, along with close collaboration between neurosurgeons, endocrinologists, and other specialists, have dramatically advanced the field of pituitary medicine.

REFERENCES

1. Elhadi AM, Hardesty DA, Zaidi HA, et al. Evaluation of surgical freedom for microscopic and endoscopic transsphenoidal approaches to the sella. *Neurosurgery*. 2015;11:69–78; discussion 78–79.
2. Mamelak AN. Pro: endoscopic endonasal transsphenoidal pituitary surgery is superior to microscope-based transsphenoidal surgery. *Endocrine*. 2014;47:409–414.
3. Lobo B, Heng A, Barkhoudarian G, et al. The expanding role of the endonasal endoscopic approach in pituitary and skull base surgery: a 2014 perspective. *Surg Neurol Int*. 2015;6:82.
4. McLaughlin N, Eisenberg AA, Cohan P, et al. Value of endoscopy for maximizing tumor removal in endonasal transsphenoidal pituitary adenoma surgery. *J Neurosurg*. 2013;118:613–620.
5. Louis RG, Eisenberg A, Barkhoudarian G, et al. Evolution of minimally invasive approaches to the sella and parasellar region. *Int Arch Otorhinolaryngol*. 2014;18:S136–S148.
6. Asemota AO, Ishii M, Brem H, et al. Comparison of complications, trends, and costs in endoscopic vs microscopic pituitary surgery: analysis from a US health claims database. *Neurosurgery*. 2017;8:458–472.
7. McLaughlin N, Laws ER, Oyesiku NM, et al. Pituitary centers of excellence. *Neurosurgery*. 2012;71:916–924; discussion 924–926.
8. Micko AS, Wohrer A, Wolfsberger S, et al. Invasion of the cavernous sinus space in pituitary adenomas: endoscopic verification and its correlation with an MRI-based classification. *J Neurosurg*. 2015;122:803–811.
9. Colao A. Pituitary tumours: the prolactinoma. *Best Pract Res Clin Endocrinol Metab*. 2009;23:575–596.
10. Thakur JD, Corlin A, Mallari RJ, et al. Pituitary adenomas in older adults (≥65 years): 90-day outcomes and readmissions: a 10-year endoscopic endonasal surgical experience. *Pituitary*. 2020;24:14–26.
11. Barkhoudarian G, Palejwala SK, Ansari S, et al. Rathke's cleft cysts: a 6-year experience of surgery vs. observation with comparative volumetric analysis. *Pituitary*. 2019;22:362–371.
12. Muller HL. Craniopharyngioma and hypothalamic injury: latest insights into consequent eating disorders and obesity. *Curr Opin Endocrinol Diabetes Obes*. 2016;23:81–89.
13. Muller HL, Merchant TE, Warmuth-Metz M, et al. Craniopharyngioma. *Nat Rev Dis Primers*. 2019;5:75.
14. Mrowczynski OD, Langan ST, Rizk EB. Craniopharyngiomas: a systematic review and evaluation of the current intratumoral treatment landscape. *Clin Neurol Neurosurg*. 2018;166:124–130.
15. Mizera L, Elbaum M, Daroszewski J, et al. Cardiovascular complications of acromegaly. *Acta Endocrinol (Buchar)*. 2018;14:365–374.
16. Friedel ME, Johnston DR, Singhal S, et al. Airway management and perioperative concerns in acromegaly patients undergoing endoscopic transsphenoidal surgery for pituitary tumors. *Otolaryngol Head Neck Surg*. 2013;149:840–844.
17. McLaughlin N, Cohan P, Barnett P, et al. Early morning cortisol levels as predictors of short-term and long-term adrenal function after endonasal transsphenoidal surgery for pituitary adenomas and Rathke's cleft cysts. *World Neurosurg*. 2013;80:569–575.
18. Esposito F, Dusick JR, Cohan P, et al. Clinical review: early morning cortisol levels as a predictor of remission after transsphenoidal surgery for Cushing's disease. *J Clin Endocrinol Metab*. 2006;91:7–13.
19. Murray JS, Jayarajasingh R, Perros P. Lesson of the week: deterioration of symptoms after start of thyroid hormone replacement. *BMJ*. 2001;323:332–333.
20. Tabarin A, Laurent F, Catargi B, et al. Comparative evaluation of conventional and dynamic magnetic resonance imaging of the pituitary gland for the diagnosis of Cushing's disease. *Clin Endocrinol (Oxf)*. 1998;49:293–300.
21. Griffiths CF, Cutler AR, Duong HT, et al. Avoidance of postoperative epistaxis and anosmia in endonasal endoscopic skull base surgery: a technical note. *Acta Neurochir (Wien)*. 2014;156:1393–1401.
22. Conger A, Zhao F, Wang X, et al. Evolution of the graded repair of CSF leaks and skull base defects in endonasal endoscopic tumor surgery: trends in repair failure and meningitis rates in 509 patients. *J Neurosurg*. 2018;130:861–875.
23. Barkhoudarian G, Cutler AR, Yost S, et al. Impact of selective pituitary gland incision or resection on hormonal function after adenoma or cyst resection. *Pituitary*. 2015;18:868–875.
24. Lobatto DJ, de Vries F, Zamanipoor Najafabadi AH, et al. Preoperative risk factors for postoperative complications in endoscopic pituitary surgery: a systematic review. *Pituitary*. 2018;21:84–97.

25. Sorba EL, Staartjes VE, Voglis S, et al. Diabetes insipidus and syndrome of inappropriate antidiuresis (SIADH) after pituitary surgery: incidence and risk factors. *Neurosurg Rev*. 2020;44:1503–1511.
26. Zacharia BE, Amine M, Anand V, et al. Endoscopic endonasal management of craniopharyngioma. *Otolaryngol Clin North Am*. 2016;49:201–212.
27. Fatemi N, Dusick JR, Mattozo C, et al. Pituitary hormonal loss and recovery after transsphenoidal adenoma removal. *Neurosurgery*. 2008;63:709–718; discussion 718–719.
28. Yuen KCJ, Ajmal A, Correa R, et al. Sodium perturbations after pituitary surgery. *Neurosurg Clin N Am*. 2019;30:515–524.
29. Bohl MA, Ahmad S, Jahnke H, et al. Delayed hyponatremia is the most common cause of 30-day unplanned readmission after transsphenoidal surgery for pituitary tumors. *Neurosurgery*. 2016;78:84–90.
30. Burke WT, Cote DJ, Iuliano SI, et al. A practical method for prevention of readmission for symptomatic hyponatremia following transsphenoidal surgery. *Pituitary*. 2018;21:25–31.
31. Staiger RD, Sarnthein J, Wiesli P, et al. Prognostic factors for impaired plasma sodium homeostasis after transsphenoidal surgery. *Br J Neurosurg*. 2013;27:63–68.
32. Patel KS, Shu Chen J, Yuan F, et al. Prediction of post-operative delayed hyponatremia after endoscopic transsphenoidal surgery. *Clin Neurol Neurosurg*. 2019;182:87–91.
33. Kelly DF, Laws Jr ER, Fossett D. Delayed hyponatremia after transsphenoidal surgery for pituitary adenoma. Report of nine cases. *J Neurosurg*. 1995;83:363–367.
34. Kinoshita Y, Tominaga A, Arita K, et al. Post-operative hyponatremia in patients with pituitary adenoma: post-operative management with a uniform treatment protocol. *Endocr J*. 2011;58:373–379.
35. Lee JI, Cho WH, Choi BK, et al. Delayed hyponatremia following transsphenoidal surgery for pituitary adenoma. *Neurol Med Chir (Tokyo)*. 2008;48:489–492; discussion 492–494.
36. Sane T, Rantakari K, Poranen A, et al. Hyponatremia after transsphenoidal surgery for pituitary tumors. *J Clin Endocrinol Metab*. 1994;79:1395–1398.
37. Tomita Y, Kurozumi K, Inagaki K, et al. Delayed postoperative hyponatremia after endoscopic transsphenoidal surgery for pituitary adenoma. *Acta Neurochir (Wien)*. 2019;161:707–715.
38. Yoon HK, Lee HC, Kim YH, et al. Predictive Factors for delayed hyponatremia after endoscopic transsphenoidal surgery in patients with nonfunctioning pituitary tumors: a retrospective observational study. *World Neurosurg*. 2019;122:e1457–e1464.
39. Zada G, Liu CY, Fishback D, et al. Recognition and management of delayed hyponatremia following transsphenoidal pituitary surgery. *J Neurosurg*. 2007;106:66–71.
40. Giustina A, Barkhoudarian G, Beckers A, et al. Multidisciplinary management of acromegaly: a consensus. *Rev Endocr Metab Disord*. 2020;21:667–678.
41. Dallapiazza RF, Jane Jr JA. Outcomes of endoscopic transsphenoidal pituitary surgery. *Endocrinol Metab Clin North Am*. 2015;44:105–115.
42. Fatemi N, Dusick JR, de Paiva Neto MA, et al. The endonasal microscopic approach for pituitary adenomas and other parasellar tumors: a 10-year experience. *Neurosurgery*. 2008;63:244–256; discussion 256.
43. Paluzzi A, Fernandez-Miranda JC, Tonya Stefko S, et al. Endoscopic endonasal approach for pituitary adenomas: a series of 555 patients. *Pituitary*. 2014;17:307–319.
44. Magro E, Graillon T, Lassave J, et al. Complications related to the endoscopic endonasal transsphenoidal approach for nonfunctioning pituitary macroadenomas in 300 consecutive patients. *World Neurosurg*. 2016;89:442–453.
45. Eseonu CI, ReFaey K, Rincon-Torroella J, et al. Endoscopic versus microscopic transsphenoidal approach for pituitary adenomas: comparison of outcomes during the transition of methods of a single surgeon. *World Neurosurg*. 2017;97:317–325.
46. Kim JH, Lee JH, Lee JH, et al. Endoscopic transsphenoidal surgery outcomes in 331 nonfunctioning pituitary adenoma cases after a single surgeon learning curve. *World Neurosurg*. 2018;109:e409–e416.
47. Nayak P, Montaser AS, Hu J, et al. Predictors of postoperative diabetes insipidus following endoscopic resection of pituitary adenomas. *J Endocr Soc*. 2018;2:1010–1019.
48. Little AS, Gardner PA, Fernandez-Miranda JC, et al. Pituitary gland recovery following fully endoscopic transsphenoidal surgery for nonfunctioning pituitary adenoma: results of a prospective multicenter study. *J Neurosurg*. 2019:1–7.
49. Younus I, Gerges MM, Dobri GA, et al. Readmission after endoscopic transsphenoidal pituitary surgery: analysis of 584 consecutive cases. *J Neurosurg*. 2019:1–6.
50. Gerges MM, Rumalla K, Godil SS, et al. Long-term outcomes after endoscopic endonasal surgery for nonfunctioning pituitary macroadenomas. *J Neurosurg*. 2020:1–12.
51. Thakur JD, Corlin A, Mallari RJ, et al. Complication avoidance protocols in endoscopic pituitary adenoma surgery: a retrospective cohort study in 514 patients. *Pituitary*. 2021;24:930–942.
52. Berker M, Isikay I, Berker D, et al. Early promising results for the endoscopic surgical treatment of Cushing's disease. *Neurosurg Rev*. 2013.
53. Starke RM, Reames DL, Chen CJ, et al. Endoscopic transsphenoidal surgery for Cushing disease: techniques, outcomes, and predictors of remission. *Neurosurgery*. 2013;72:240–247; discussion 247.
54. Wagenmakers MA, Boogaarts HD, Roerink SH, et al. Endoscopic transsphenoidal pituitary surgery: a good and safe primary treatment option for Cushing's disease, even in case of macroadenomas or invasive adenomas. *Eur J Endocrinol*. 2013;169:329–337.
55. Hazer DB, Isik S, Berker D, et al. Treatment of acromegaly by endoscopic transsphenoidal surgery: surgical experience in 214 cases and cure rates according to current consensus criteria. *J Neurosurg*. 2013;119:1467–1477.
56. Shin SS, Tormenti MJ, Paluzzi A, et al. Endoscopic endonasal approach for growth hormone secreting pituitary adenomas: outcomes in 53 patients using 2010 consensus criteria for remission. *Pituitary*. 2013;16:435–444.
57. Jane Jr JA, Starke RM, Elzoghby MA, et al. Endoscopic transsphenoidal surgery for acromegaly: remission using modern criteria, complications, and predictors of outcome. *J Clin Endocrinol Metab*. 2011;96:2732–2740.
58. Cappabianca P, Cavallo LM, Colao A, et al. Surgical complications associated with the endoscopic endonasal transsphenoidal approach for pituitary adenomas. *J Neurosurg*. 2002;97:293–298.
59. Gondim JA, Almeida JP, de Albuquerque LA, et al. Pure endoscopic transsphenoidal surgery for treatment of acromegaly: results of 67 cases treated in a pituitary center. *Neurosurg Focus*. 2010;29:E7.
60. de Paiva Neto MA, Vandergrift A, Fatemi N, et al. Endonasal transsphenoidal surgery and multimodality treatment for giant pituitary adenomas. *Clin Endocrinol (Oxf)*. 2010;72:512–519.
61. Koutourousiou M, Gardner PA, Fernandez-Miranda JC, et al. Endoscopic endonasal surgery for craniopharyngiomas: surgical outcome in 64 patients. *J Neurosurg*. 2013;119:1194–1207.
62. Koutourousiou M, Gardner PA, Fernandez-Miranda JC, et al. Endoscopic endonasal surgery for giant pituitary adenomas: advantages and limitations. *J Neurosurg*. 2013;11:621–631.
63. Dula DJ, Fales W. The 'ring sign': is it a reliable indicator for cerebral spinal fluid? *Ann Emerg Med*. 1993;22:718–720.
64. Tunkel AR, Hasbun R, Bhimraj A, et al. 2017 Infectious Diseases Society of America's clinical practice guidelines for healthcare-associated ventriculitis and meningitis. *Clin Infect Dis*. 2017;64:e34–e65.
65. Barkhoudarian G, Kelly D. Complications with transsphenoidal surgery: a review. In: Laws Jr ER, Cohen-Gadol AA, Schwartz TH, Sheehan JP, eds. *Transsphenoidal Surgery*. Cham: Springer; 2017.
66. Berker M, Hazer DB, Yucel T, et al. Complications of endoscopic surgery of the pituitary adenomas: analysis of 570 patients and review of the literature. *Pituitary*. 2012;15:288–300.
67. Leng LZ, Greenfield JP, Souweidane MM, et al. Endoscopic, endonasal resection of craniopharyngiomas: analysis of outcome including extent of resection, cerebrospinal fluid leak, return to preoperative productivity, and body mass index. *Neurosurgery*. 2012;70:110–123; discussion 123–124.
68. Cavallo LM, Frank G, Cappabianca P, et al. The endoscopic endonasal approach for the management of craniopharyngiomas: a series of 103 patients. *J Neurosurg*. 2014;121:100–113.
69. Yamada S, Fukuhara N, Yamaguchi-Okada M, et al. Therapeutic outcomes of transsphenoidal surgery in pediatric patients with craniopharyngiomas: a single-center study. *J Neurosurg Pediatr*. 2018;21:549–562.

70. Madsen PJ, Buch VP, Douglas JE, et al. Endoscopic endonasal resection versus open surgery for pediatric craniopharyngioma: comparison of outcomes and complications. *J Neurosurg Pediatr*. 2019;24:1–10.
71. Koutourousiou M, Fernandez-Miranda JC, Wang EW, et al. The limits of transsellar/transtuberculum surgery for craniopharyngioma. *J Neurosurg Sci*. 2018;62:301–309.
72. Padhye V, Valentine R, Wormald PJ. Management of carotid artery injury in endonasal surgery. *Int Arch Otorhinolaryngol*. 2014;18:S173–S178.
73. Chin OY, Ghosh R, Fang CH, et al. Internal carotid artery injury in endoscopic endonasal surgery: a systematic review. *Laryngoscope*. 2016;126:582–590.
74. Ghorbani M, Griessenauer CJ, Shojaei H, et al. Endovascular reconstruction of iatrogenic internal carotid artery injury following endonasal surgery: a systematic review. *Neurosurg Rev*. 2020;44:1797–1804.
75. Dusick JR, Esposito F, Malkasian D, et al. Avoidance of carotid artery injuries in transsphenoidal surgery with the Doppler probe and micro-hook blades. *Neurosurgery*. 2007;60:322–328; discussion 328–329.
76. Dehdashti AR, Ganna A, Karabatsou K, et al. Pure endoscopic endonasal approach for pituitary adenomas: early surgical results in 200 patients and comparison with previous microsurgical series. *Neurosurgery*. 2008;62:1006–1015; discussion 1015–1017.
77. Frank G, Pasquini E, Farneti G, et al. The endoscopic versus the traditional approach in pituitary surgery. *Neuroendocrinology*. 2006;83:240–248.
78. Yano S, Kawano T, Kudo M, et al. Endoscopic endonasal transsphenoidal approach through the bilateral nostrils for pituitary adenomas. *Neurol Med Chir (Tokyo)*. 2009;49:1–7.
79. Gondim JA, Almeida JP, Albuquerque LA, et al. Endoscopic endonasal approach for pituitary adenoma: surgical complications in 301 patients. *Pituitary*. 2011;14:174–183.
80. Dallapiazza RF, Grober Y, Starke RM, et al. Long-term results of endonasal endoscopic transsphenoidal resection of nonfunctioning pituitary macroadenomas. *Neurosurgery*. 2015;76:42–52; discussion 52–53.
81. Frank G, Sciarretta V, Mazzatenta D, et al. Transsphenoidal endoscopic approach in the treatment of Rathke's cleft cyst. *Neurosurgery*. 2005;56:124–128; discussion 129.
82. Kim JE, Kim JH, Kim OL, et al. Surgical treatment of symptomatic Rathke cleft cysts: clinical features and results with special attention to recurrence. *J Neurosurg*. 2004;100:33–40.
83. Han SJ, Rolston JD, Jahangiri A, et al. Rathke's cleft cysts: review of natural history and surgical outcomes. *J Neurooncol*. 2014;117:197–203.
84. Mendelson ZS, Husain Q, Elmoursi S, et al. Rathke's cleft cyst recurrence after transsphenoidal surgery: a meta-analysis of 1151 cases. *J Clin Neurosci*. 2014;21:378–385.
85. Benveniste RJ, King WA, Walsh J, et al. Surgery for Rathke cleft cysts: technical considerations and outcomes. *J Neurosurg*. 2004;101:577–584.
86. Aho CJ, Liu C, Zelman V, et al. Surgical outcomes in 118 patients with Rathke cleft cysts. *J Neurosurg*. 2005;102:189–193.
87. Van Gompel JJ, Nippoldt TB, Higgins DM, et al. Magnetic resonance imaging-graded hypothalamic compression in surgically treated adult craniopharyngiomas determining postoperative obesity. *Neurosurg Focus*. 2010;28:E3.
88. Alexandraki KI, Kaltsas GA, Karavitaki N, et al. The medical therapy of craniopharyngiomas: the way ahead. *J Clin Endocrinol Metab*. 2019;104:5751–5764.
89. Casanueva FF, Barkan AL, Buchfelder M, et al. Criteria for the definition of Pituitary Tumor Centers of Excellence (PTCOE): a Pituitary Society Statement. *Pituitary*. 2017;20(5):489–498.

11

Craniopharyngiomas and Parasellar Masses

Niki Karavitaki

OUTLINE

Types of Parasellar Masses, 170
Craniopharyngiomas, 170
  Epidemiology, 170
  Pathology, 170
  Pathogenesis, 170
  Imaging Features, 171
Clinical Manifestations and Pituitary Function At Presentation, 172
  Management, 172
Long-Term Morbidity and Mortality, 173
Rathke's Cleft Cysts, 174
  Epidemiology, 174
  Pathology, 174
  Pathogenesis, 174
  Imaging Features, 174
  Clinical Manifestations and Pituitary Function at Presentation, 174
  Management, 175
Meningiomas, 176
Chordomas, 177
Gliomas, 177
Germ Cell Tumors, 177
Langerhans Cell Histiocytosis, 178
Hamartomas, 178
Pituitary Metastasis, 178

TYPES OF PARASELLAR MASSES

Various types of nonadenomatous lesions can be detected in the parasellar area (Table 11.1). The most common are craniopharyngiomas, Rathke's cleft cysts, and meningiomas. They generally present with mass effects to surrounding structures, or they may be incidentally found. Their management depends on their etiology, location, presenting manifestations, and potential for aggressive/malignant behavior. Other neoplastic lesions are much more rare; inflammatory lesions will not be discussed further in this chapter.

CRANIOPHARYNGIOMAS

Epidemiology

Craniopharyngiomas account for 2% to 5% of all primary intracranial neoplasms (up to 15% of intracranial tumors diagnosed in childhood).[1] Their annual incidence is around 0.18 cases per 100,000 population[2] and can be detected at any age, including the prenatal and neonatal periods.[3] A bimodal age distribution with a peak incidence rates at ages 5 to 14 years and 50 to 74 years has been described,[2,4] the pathophysiological significance of which is uncertain. No sex differences have been confirmed.[4,5]

Pathology

Craniopharyngiomas are World Health Organization (WHO) grade I epithelial tumors arising along the path of the craniopharyngeal duct (the canal connecting the stomodeal ectoderm with the evaginated Rathke's pouch). Malignant transformation has been rarely reported and is possibly triggered by previous irradiation.[1] Histologically, there are two subtypes, the adamantinomatous and the papillary.[1,6]

The adamantinomatous is the most common subtype and is diagnosed at any age. Macroscopically, it consists of cystic and/or solid components, necrotic debris, fibrous tissue, and calcification. The cystic fluid ranges from "machinery oil" to shimmering cholesterol-laden fluid consisting of desquamated squamous epithelial cells, rich in membrane lipids and cytoskeleton keratin. Their margins can be sharp and irregular, often merging into a peripheral zone of dense reactive gliosis in the surrounding brain tissue. The epithelium is typically composed of three layers of cells: a distinct palisaded basal layer; an intermediate layer of variable thickness composed of loose aggregates of stellate cells (termed "stellate reticulum"); and a top layer facing into the cyst lumen with abruptly enlarged, flattened, and keratinized to flat plate-like squamous cells. The flat squames are desquamated and form nodules of "wet" keratin which often show heavy calcification. The presence of the typical adamantinomatous epithelium or of the "wet" keratin nodules alone are diagnostic for this variant.[1]

The papillary subtype has been almost exclusively described in adults. It consists of mature squamous epithelium forming pseudopapillae and of an anastomosing fibrovascular stroma with membranous beta-catenin immunoreactivity only (in contrast to adamantinomatous, which can also demonstrate nuclear immunopositivity for beta-catenin). The differential diagnosis between a papillary craniopharyngioma and a Rathke's cleft cyst may be difficult; in such cases, absence of a solid component and the presence of extensive ciliation and/or mucin production are suggestive of Rathke's.[1,6]

Pathogenesis

Beta-catenin mutations in exon 3 have been identified in the adamantinomatous subtype. This exon encodes the degradation targeting box of beta-catenin, and the mutant form is resistant to degradation leading to accumulation of nuclear beta-catenin protein (a transcriptional activator of the Wnt signaling pathway). Strong beta-catenin expression has been shown in the adamantinomatous subtype, indicating reactivation of the Wnt signaling pathway and subsequent deregulation of several downstream pathways.[7-9]

TABLE 11.1 Types of Nonadenomatous Parasellar Masses

Tumoral Lesions	Craniopharyngioma Meningioma Chordoma Glioma Langerhans cell histiocytosis Schwannoma Neurofibroma Lymphoma Germ cell tumor Metastasis Chondrosarcoma Paraganglioma Hemangiopericytoma Hamartoma Lipoma Tumors of posterior pituitary
Cystic lesions	Rathke's cleft cyst Epidermoid cyst Dermoid cyst Arachnoid cyst Sphenoid sinus mucocele
Vascular	Aneurysms Cavernous hemangioma Cavernous sinus thrombosis Carotid cavernous fistula
Inflammatory/infectious	Sarcoidosis Wegener's granulomatosis Tolosa–Hunt syndrome Tuberculosis Abscess

In the papillary subtype, activating mutations in *BRAF* (V600E) have been detected (in 81%–100% of cases, depending on the sequencing method used), and expression of the mutant protein has been confirmed by immunohistochemistry implicating activation of the RAS/RAF/MEK/ERK pathway.[8]

Gains or losses of large chromosomal regions are rare in craniopharyngiomas, and the two pathological subtypes have different epigenomic and transcriptomic signatures.[8]

KEY POINTS

- Craniopharyngiomas are very rare tumors (annual incidence around 0.18 cases per 100,000 population) and comprise two distinct pathological subtypes: adamantinomatous (often carrying *beta-catenin* mutations) and papillary [characterized by *BRAF* (V600E) mutations]

Imaging Features

Most craniopharyngiomas are located in the sellar/parasellar region and the majority of them (94%–95%) have a suprasellar component at the time of detection. In terms of consistency, they can be purely or predominantly cystic, purely or predominantly solid, or mixed. Rare locations of these tumors that have been reported are the nasopharynx, the paranasal area, the sphenoid bone, the ethmoid sinus, the intrachiasmatic area, the temporal lobe, the pineal gland, the posterior cranial fossa, the cerebellopontine angle, the midportion of the midbrain, or within the third ventricle.[1]

On computed tomography (CT), they are usually of mixed attenuation; the cyst fluid has low density, and the contrast medium enhances any solid portion, including the cyst capsule. This imaging modality is helpful in identifying calcification.[1] On magnetic resonance imaging (MRI), the appearances depend on the proportion of the solid and cystic components, the content of the cyst(s) (cholesterol, keratin, hemorrhage), and the degree of tumor calcification (Fig. 11.1). On precontrast T1-weighted images, a solid lesion appears as

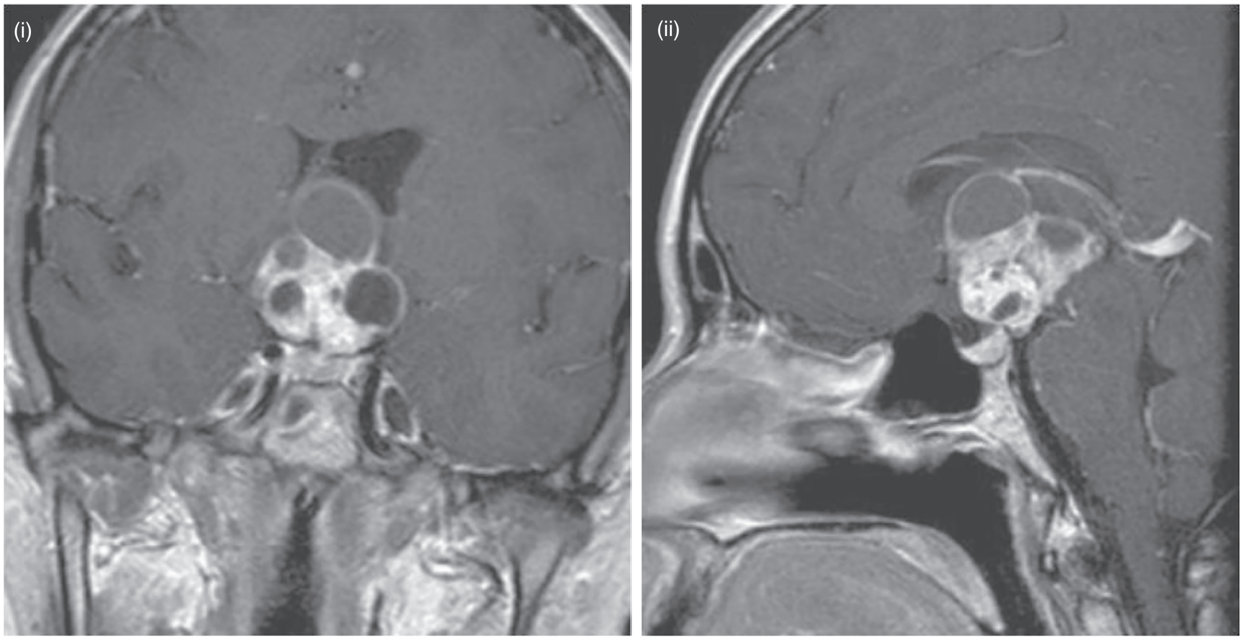


Fig. 11.1 Magnetic resonance imaging of craniopharyngioma. Coronal (i) and sagittal (ii) T1-weighted postgadolinium images with lobulated shape and mixed solid and cystic characteristics. Peripheral contrast-enhancing rim of the cystic components are evident. (From Trifanescu R, Ansorge O, Wass JA, et al. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*. 2012;76:151–160.)

TABLE 11.2 Presenting Clinical Features in Children and Adults With Craniopharyngioma

	Children	Adults	Total
Headaches	78%	56%	64%
Menstrual disorders		57%	
Visual field defects	46%	60%	55%
Decreased visual acuity	39%	40%	39%
Nausea/vomiting	54%	26%	35%
Growth failure	32%		
Poor energy	22%	32%	29%
Impaired sexual function		28%	
Impaired secondary sexual characteristics (patients aged ≥13 years)			24%
Lethargy	17%	26%	23%
Other cranial nerves palsies	27%	9%	15%
Polyuria/polydipsia	15%	15%	15%
Papilledema	29%	6%	14%
Cognitive impairment (memory, concentration)	10%	17%	14%
Anorexia/weight loss	20%	8%	12%
Optic atrophy	5%	14%	10%
Hyperphagia/excessive weight gain	5%	13%	10%
Psychiatric symptoms/change in behavior	10%	8%	8%
Somnolence	5%	10%	8%
Galactorrhea		8%	
Decreased consciousness/coma	10%	4%	6%
Cold intolerance	0%	8%	5%
Unsteadiness/ataxia	7%	3%	4%
Hemiparesis	7%	1%	3%
Blindness	3%	3%	3%
Meningitis	0%	3%	2%

(From Karavitaki N, Brufani C, Warner JT, et al. Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endocrinol (Oxf)*. 2005;62:397–409.)

iso- or hypointense relative to the brain and enhances after gadolinium administration; on T2-weighted images, craniopharyngiomas are usually of mixed hypo- or hyperintensity. Cystic parts are usually hypointense on T1- and hyperintense on T2-weighted sequences, and a thin peripheral contrast-enhancing rim of the cyst can be shown on T1-weighted images. Protein, cholesterol, and methemoglobin relate with high signal on T1-weighted sections, while very concentrated protein and various blood products can be associated with low T2-weighted signal.[1]

The differential diagnosis on imaging includes a number of sellar or parasellar lesions, including Rathke's cleft cyst, dermoid cyst, epidermoid cyst, pituitary adenoma, germinoma, hamartoma, suprasellar aneurysm, arachnoid cyst, suprasellar abscess, glioma, meningioma, sarcoidosis, tuberculosis, and Langerhans cell histiocytosis (LCH). Differentiation from a Rathke's cleft cyst (typically small, round, purely cystic lesions lacking calcification) can occasionally be challenging.[1,10]

CLINICAL MANIFESTATIONS AND PITUITARY FUNCTION AT PRESENTATION

The presenting manifestations are attributed to pressure effects on surrounding structures including the visual pathways, brain parenchyma, ventricular system, major blood vessels, and the hypothalamopituitary system.[10-12] Their severity depends on the location, size, and growth potential of the tumor, and the duration of the symptoms until diagnosis ranges from 1 week to 372 months.[1] The main presenting clinical features (neurological, visual, hypothalamopituitary) are shown in Table 11.2. Headaches, nausea/vomiting, visual disturbances, growth failure (in children), and hypogonadism (in adults) are the most frequently reported. Particularly in childhood-onset craniopharyngioma, a reduced growth rate is already evident in early infancy, whereas weight gain tends to present later and predicts obesity.[13] Various rates of hypothalamopituitary dysfunction have been reported, with growth hormone (GH) deficiency diagnosed in 35% to 100% of the evaluated patients, follicle-stimulating hormone (FSH)/luteinizing hormone (LH) deficiency in 38% to 91%, adrenocorticotropic hormone (ACTH) deficiency in 21% to 68%, thyroid-stimulating hormone (TSH) deficiency in 20% to 42%, and diabetes insipidus in 6% to 38%.[1]

Management

Surgery Combined With or Without Adjuvant Radiotherapy.

Surgery followed or not by radiotherapy is the most widely used first therapeutic approach for craniopharyngiomas. These tumors pose a surgical challenge because of their irregular borders and to their tendency to adhere to vital neurovascular structures, rendering safe complete removal difficult. In cases of large cystic components, fluid aspiration can lead to relief of the obstructive manifestations, but subsequent removal of the solid portion should not be delayed because of the significant risk of cyst refilling.[10,14] The extent of tumor excision has been a subject of debate and is influenced by many factors, including tumor location and presence of brain invasion.[1,15,16] In recent years, partial or limited resection combined with radiotherapy has been adopted in many centers rather than complete tumor removal, aiming to reduce hypothalamic damage and subsequent morbidity.[17] The rate of postoperative mortality in published series ranges between 0% and 5.4%.[18] A metaanalysis of 2955 patients showed an early mortality rate of 2.6% after transsphenoidal and 3.1% after transcranial surgery.[19] Various radiotherapy techniques have been applied for the irradiation of craniopharyngiomas, and doses between 50 and 60 Gy have been used for fractionated irradiation.[20,21] Radiotherapy of cystic craniopharyngiomas carries the risk of enlargement, and this needs to be taken into account during the period of irradiation.[10,20,21]

Recurrent tumors are detected at mean intervals ranging between 1 and 4.3 years following various primary treatment modalities.[1] Series with imaging confirmation of the radicality of resection report 10-year recurrence rates after gross total removal ranging between 0% and 62%, and after partial or subtotal resection between 25% and 100%. In cases of partial removal, adjuvant radiotherapy (conventional fractionated external beam) improves the local control rates (recurrence rates 10%–63% at 10 years).[1] In a systematic review, subtotal or partial resection with adjuvant radiotherapy, and gross total removal, showed similar rates of long-term tumor control.[22] Tumor control rates with newer higher precision techniques, such as fractionated stereotactic conformal radiotherapy, have remained optimal, with 5-year progression-free survival exceeding 90%.[23] Series using proton beam therapy have also provided optimal results,[21,24] but studies with long-term follow-up are necessary.

The growth rate of craniopharyngiomas varies considerably, and reliable clinical, radiological, and pathological criteria predicting their behavior are not available. Attempts to identify prognostic factors for recurrence, other than the type of treatment like age group at diagnosis, sex, imaging features, and pathological subtypes, have not provided consistent results.[1] Surgery for the management of recurrent craniopharyngiomas remains technically challenging, and the beneficial effect of radiotherapy (proceeded or not by surgery) has been clearly shown.[10,25] Recurrent lesions with significant cystic component not amenable to total extirpation may be treated by repetitive aspirations through an indwelling Ommaya reservoir apparatus.

Intracystic Irradiation. This involves stereotactically guided instillation of beta-emitting isotopes (mainly ^{32}phosphate, ^{90}yttrium, ^{186} rhenium, ^{198}gold) into cystic craniopharyngiomas. In studies with mean or median follow-up between 3.1 and 11.9 years after administration of ^{90}yttrium or ^{32}phosphorus at doses of 200 to 270 Gy, complete or partial cyst resolution was reported in 71% to 88%, stable appearances in 3% to 19%, and increase in size in 5% to 10% of cases.[26-31] New cyst formation or increase in the solid component of the tumor was found in between 6.5% and and 20%. Deterioration of vision has been reported in 10% to 58% of cases (attributed to failure of cyst collapse, formation of new cysts, increase in the solid tumor, or possibly radiation damage).[26-31]

Intracystic Administration of Sclerosing Substances. Intracystic installation of the antineoplasmatic agent bleomycin through an Ommaya reservoir connected to a catheter has been used in cystic craniopharyngiomas, with tumor control rates ranging between 0% and 100%. Direct leakage of bleomycin to surrounding tissues during the installation procedure, diffusion of the drug through the cyst wall, or high drug dose have been associated with various toxic (hypothalamic damage, blindness, hearing loss, ischemic attacks, peritumoral oedema), or even fatal, effects.[32,33] The value of this approach in the management algorithm of craniopharyngiomas, as well as optimal protocols for the administration of bleomycin, remain to be established.

Intracystic interferon-alpha has also been applied in cystic lesions, and reports suggest that it can delay disease progression. This effect, combined with the fact that it only has a few significant side effects, makes interferon-alpha a potential option for delaying definitive treatment with surgery or radiotherapy in children with cystic craniopharyngioma.[34]

Stereotactic Radiosurgery. This technique delivers a single fraction of high-dose irradiation to precisely mapped targets. Tumor volume and close attachment to the optic apparatus are limiting factors for its application. Reported 5-year progression-free survival rates range between 61% and 68% for small-volume lesions.[35-37] Stereotactic radiosurgery can be considered for well-defined residual tumors or for the control of small, solid recurrent tumors, particularly after failure of conventional radiotherapy.

Systemic Chemotherapy/Interferon-Alpha. The potential benefit of systemic chemotherapy in craniopharyngiomas has been investigated in a very limited number of patients, with rather inconclusive results.[38,39] Pegylated interferon-a-2b has also been tried,[36] but further studies are needed to clarify the impact of this approach.[40,41]

Targeted Therapy. The identification of the *BRAF* (V600E) mutation in papillary craniopharyngiomas has opened perspectives for the use of agents targeting and inhibiting mutant BRAF in cases resistant to other treatments. Case reports using BRAF and MAPK inhibitors have shown reduction in tumor volume in some patients,[8,42] and results of large trials assessing the efficacy of this treatment modality, particularly for aggressive tumors, are required.

KEY POINTS

- Subtotal or partial resection with adjuvant radiotherapy, and gross total removal, are reported to achieve similar rates of long-term tumor control, but the first option has been adopted by many centers aiming to reduce hypothalamic damage and subsequent morbidity.
- Other management modalities have been applied with variable success rates and adverse effects including intracystic administration of sclerosing substances, stereotactic radiosurgery, and systemic chemotherapy/interferon-alpha. The value of agents targeting and inhibiting mutant *BRAF* in papillary craniopharyngiomas is currently under investigation.

LONG-TERM MORBIDITY AND MORTALITY

Patients with craniopharyngiomas may suffer from significant long-term morbidities (mainly endocrine, visual, hypothalamic, neurobehavioral), attributed to damage of critical structures by the primary or recurrent tumor and/or to the adverse effects of the treatments offered.

At last assessment, pituitary hormone deficits range between 88% and 100% for GH, 80% and 95% for FSH/LH, 55% and 88% for ACTH, 39% and 95% for TSH, and 25% and 86% for vasopressin/antidiuretic hormone.[1] GH replacement in adult patients with craniopharyngioma improves free-fat mass, total and low-density lipoprotein, and quality of life, but it does not significantly decrease total body fat.[43] Based on observational studies, GH replacement is not associated with an increased risk of tumor recurrence.[44,45] A rare, challenging complication is adipsic diabetes insipidus, which confers significant risk of serious electrolyte imbalance and is one of the most difficult morbidities to manage; careful fluid balance with close monitoring of intake/output and regular checking of the weight are vital to avoid recurrent episodes of hyper- or hyponatremia that could further contribute to morbidity and mortality.[1]

Compromised vision has been reported in up to 62.5% of patients treated by surgery combined or not with radiotherapy over observation periods of 10 years.[1] Visual outcome is adversely affected by presence of visual symptoms at presentation and by daily irradiation doses greater than 2 Gy.[1]

Hypothalamic damage can manifest as hyperphagia and uncontrollable obesity, disorders of thirst and water/electrolyte balance, behavioral and cognitive impairment, loss of temperature control, and disordered sleep patterns. Obesity is the most frequent hypothalamic morbidity (reported in 26%–61% of the patients treated by surgery combined or not with radiotherapy) and results from alterations in the mechanisms involved in satiety, hunger, and energy balance.[1,10,46-48] It can be exaggerated by neurological defects, visual failure, somnolence, sleep disturbance, hypopituitarism, and psychosocial disorders.[51] The orexigenic gastric hormone ghrelin has not been found to be elevated.[52] Factors proposed to be associated with significant hypothalamic morbidity are hypothalamic disturbance at diagnosis, hypothalamic invasion, attempts to remove adherent tumor from the region of hypothalamus, and multiple operations for recurrence.[1,53,54] The management of hypothalamic obesity is very challenging and includes provision of dietary and behavioral modifications, encouragement of regular physical activity, psychological counseling, antiobesity drugs, and bariatric surgery. A systematic review and metaanalysis including 21 cases of bariatric surgery for hypothalamic obesity in patients with craniopharyngioma (six with adjustable gastric banding, eight with sleeve gastrectomy, six with Roux-en-Y gastric bypass, and one with biliopancreatic diversion) suggested that the gastric bypass group had the maximal mean weight loss achieved 12 months postsurgery.[55] Medical therapies including dextroamphetamine, the combination of diazoxide and metformin (aiming to reduce the hyperinsulinemia), octreotide (aiming to reduce hyperinsulinemia and simultaneously enhance insulin's action), glucagon-like peptide-1 analogs, and methionine aminopeptidase 2 inhibitors have been also tried, with variable outcomes.[49]

Survival of patients with craniopharyngioma can be compromised.[1,48] and the risk of cardio-/cerebrovascular and respiratory mortality is increased.[1,56,57] Hypothalamic obesity and related metabolic complications[58] may play a role in the cardiovascular morbidity and, indeed, hypothalamic involvement has been shown to have a negative impact on mortality.[59]

RATHKE'S CLEFT CYSTS

Epidemiology

Rathke's cleft cysts (RCCs) are benign, sellar, and/or suprasellar lesions originating from the remnants of Rathke's pouch.[60] They are the most common incidentally-discovered sellar lesions after pituitary adenomas. In routine autopsies, they are found in 13% to 33% of normal pituitary glands.[61,62] They have been described in all age groups, with a peak frequency between the fourth and sixth decades. Symptomatic cases of RCC are rare and account for 5% to 15% of all surgically resected sellar lesions.[63]

Pathology

RCCs are benign cystic lesions ranging from a few mm to 5 cm with usually a thick, mucoid material consisting of cholesterol and protein.[1,63,64] The cyst wall is lined by simple cuboidal or columnar epithelium with or without cilia and occasionally goblet cells secreting mucous. Pseudostratified columnar ciliated respiratory-type epithelium, squamous metaplasia (making them occasionally indistinguishable from craniopharyngiomas), and nodules of mucinous material may also be seen.[65-67] An inflammatory reaction penetrating the cyst epithelium or subjacent stroma has been described in up to 50% of symptomatic RCCs and can result in destruction of the adjacent pituitary gland and the development of hypopituitarism, and thus pituitary hormone deficits.[63-68] Coexistence of RCCs with pituitary adenomas has been reported, but this is probably coincidental.[69]

Pathogenesis

RCCs originate from embryonic remnants of Rathke's pouch, which appears on the twenty-fourth day of embryonic life as a dorsal diverticulum from the ectodermal stomodeum (primitive oral cavity). Rathke's pouch extends cranially to form the craniopharyngeal duct, and at the same time the infundibulum develops from the diencephalon as a downgrowth of the neuroepithelium. Between the third and fifth months, the anterior wall of Rathke's pouch forms the pars anterior of the pituitary, and the posterior wall forms the pars intermedia. Meanwhile, the median eminence, the pituitary stalk, and the posterior pituitary lobe develop from the infundibulum. Obliteration of the craniopharyngeal duct is associated with involution of the pouch, and, if this does not happen, a cyst develops between the pars distalis and the pars nervosa.[63,64]

Imaging Features

RCCs usually appear as well-circumscribed, spherical, ovoid, or dumbbell sellar lesions in the region of the pars intermedia.

On CT, they are usually hypo- or isodense homogeneous lesions or slightly hyperdense relative to the brain parenchyma. The enhancement is ring-like or capsular.[63]

On MRI, they are usually well-demarcated cystic lesions with homogeneous intensity signal and may show thin cyst wall enhancement. The signal intensity varies widely (hypo-, iso-, or hyperintense) and depends on the cyst content[66,70,71] (Fig. 11.2) Cysts demonstrating hypointensity on T1-weighted images and hyperintensity on T2-weighted ones have cerebrospinal fluid (CSF)-like fluid; those with hyperintensity on T1-weighted images and isointensity on T2-weighted ones have mucoid material, whereas hyperintensity on both T1- and T2-weighted images suggests the presence of blood. Cysts with a high protein concentration show high T1 signal intensity. Although hyperintensity on T2 images has been described, homogeneous T2 hypointense signal intensity is highly suggestive of RCC, facilitating the differential diagnosis from chordomas, chondroid tumors, abscesses, mucoceles, and lipomas, all showing a hyperintense T2 signal.[70] Small intracystic nodules attributed to proteinaceous concentrations demonstrate low signal intensity on T2-weighted and high signal intensity on T1-weighted images compared with that of the surrounding cyst fluid; these do not enhance, and are pathognomonic for RCCs.[70,72]

Other cystic lesions included in the differential diagnosis apart from craniopharyngiomas and cystic adenomas are dermoid cysts, epidermoid cysts, arachnoid cysts, and abscesses.

Clinical Manifestations and Pituitary Function at Presentation

Small cysts are asymptomatic, and their detection is usually incidental. When RCCs are symptomatic, the manifestations mostly relate to pressure effects to adjacent structures. Headache is the most prevalent symptom and has been associated with either hyper- or isointense content on T1-weighted MRI, mucous cyst content, or intense chronic cyst wall inflammation. Major visual field defects have been demonstrated in up to 75% of cases and correlate with cyst size, whereas cranial nerve palsies have been rarely described.[63,5-67,73]

Anterior pituitary hormone deficits at presentation vary widely, between 19% and 81%, with hypogonadism and hyperprolactinemia as the most frequent hormonal manifestations. Diabetes insipidus can be present in 0% to 19% of the cases, and inappropriate antidiuretic hormone secretion with hyponatremia has also been rarely reported.[63,65-67,73]

Apoplexy on the background of a RCC is an uncommon form of presentation.[63] Aseptic meningitis has been rarely reported and is

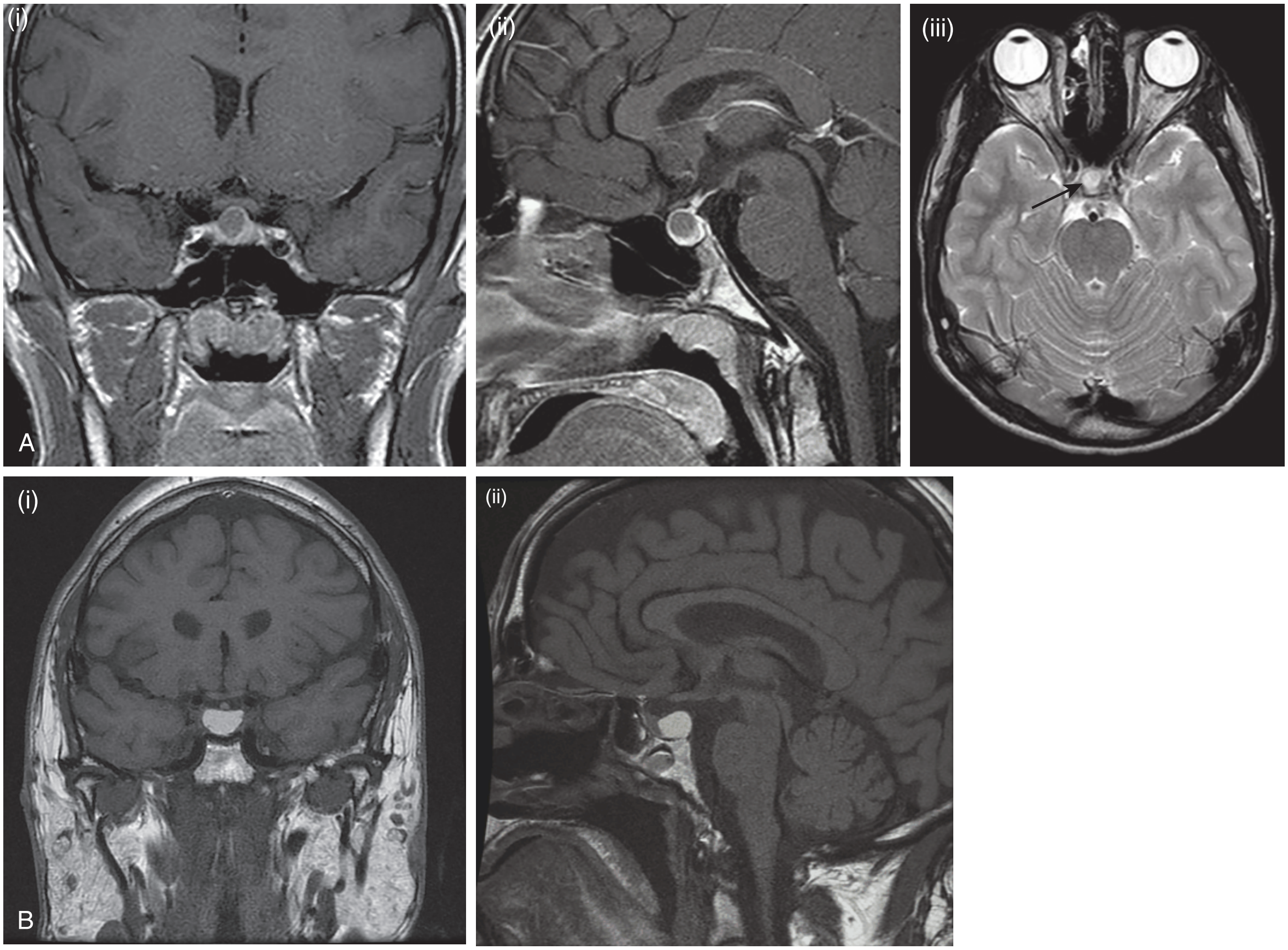


Fig. 11.2 Magnetic resonance imaging of Rathke's cleft cysts. **A**, coronal (i) and sagittal (ii) T1-weighted postgadolinium images demonstrating low signal intensity and rim enhancement. On T2-weighted images this cyst had hyperintense signal *(arrow)* (iii). **B**, Coronal (i) and sagittal (ii) T1-weighted images showing hyperintense signal. (From Trifanescu R, Ansorge O, Wass JA, et al. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*. 2012;76:151–160.)

attributed to rupture of the cystic content into the subarachnoid space or to extension of the inflammation into surrounding structures.[63]

KEY POINTS

- Rathke's cleft cysts are benign lesions often discovered incidentally. Their imaging features depend on the cyst content and when symptomatic, they can present with headaches, visual deterioration and pituitary dysfunction.

Management

Small, asymptomatic RCCs do not require surgical intervention. In symptomatic ones, surgery is indicated, aiming to drain the cyst content and safely remove as much of the capsule as possible. Cyst fenestration and drainage via an endonasal/sublabial transsphenoidal approach are most commonly adopted.[63,74] Radiotherapy is not routinely required but may be considered for recurring lesions.[75]

After surgery, headaches and major visual field defects improve or resolve in a significant number of patients. Partial hypopituitarism may recover postsurgery, whereas panhypopituitarism seldom resolves.[63] Postoperative new diabetes insipidus is reported in between 25% and 67% of the cases.[63] It has been proposed that aggressive surgery with resection of the cyst wall is more commonly associated with new endocrine morbidity.[74]

Data on the endocrine and visual outcome after long-term follow-up are rather limited; in a series of 33 cases of RCCs with mean follow-up of 48 months, 50% of the patients had gonadotrophin deficiency, 42% ACTH deficiency, and 47% TSH deficiency, 39% required desmopressin treatment, and 19% had at least quadrantanopia at last review.[73]

Literature on the natural history of nonoperated presumed RCCs suggests that 26% to 94% of them remained stable (based on imaging studies or on the clinical picture) during observation periods of up to 9 years.[64,67,74,76] In a series of 29 patients with symptomatic cystic sellar lesions and imaging features consistent with RCC, spontaneous involution was reported in 31% of cases.[77] Possible mechanisms include rate of absorption of cystic fluid greater than rate of secretion or repeated cyst ruptures and leakage.[70] Enlargement of the cysts has been reported in 5.3% of incidentally found presumed RCCs during a mean follow-up period of 39 months.[76] The change in the cyst size is probably attributed to imbalance between secretion and absorption of the cyst content, intracystic hemorrhage, or infection.

Reported RCC recurrence rates after surgery range between 0% and 33% during variable follow-up periods and show wide variation

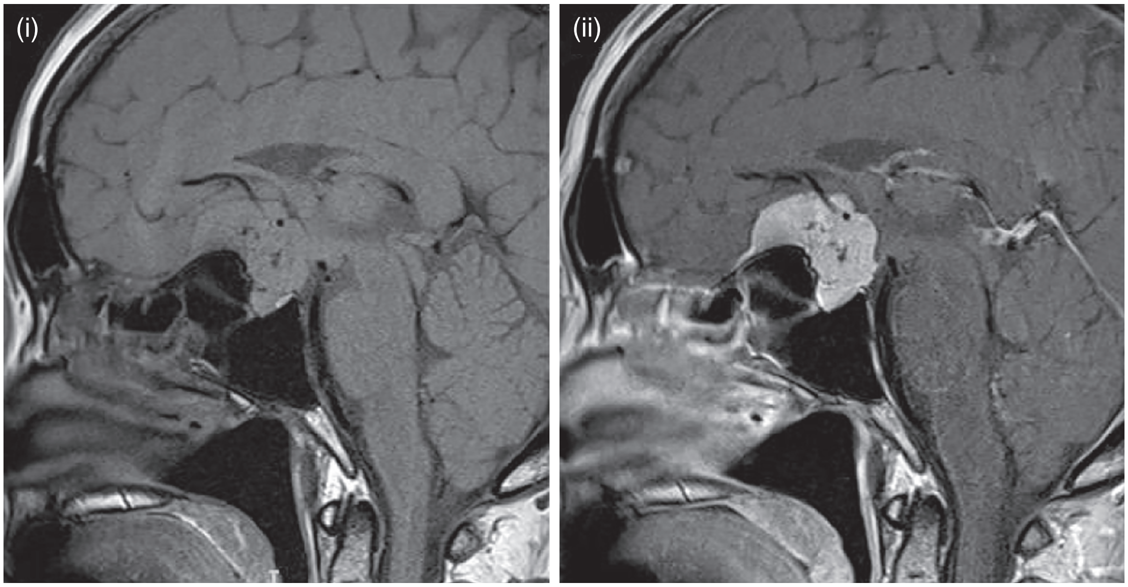


Fig. 11.3 Magnetic resonance imaging (sagittal, T1-weighted) (i) pre- and (ii) postgadolinium demonstrating a meningioma arising from the sella and extending to the suprasellar cistern. There is enhancement after gadolinium, and the characteristic dura tail is also present anteriorly.

(between 0% and 33%).[63,78] This variation may to some extent be explained by the different definitions of relapse (asymptomatic or symptomatic, requiring surgical intervention or not). In a series with mean follow-up of 48 months, the 4-year probability of recurrence was 52% at 4 years.[73] Although rare, multiple relapses have also been described.[63] The presence of squamous metaplasia in the cyst wall has been associated with increased risk of recurrence.[64,66,74] Data on extent of resection/aggressiveness of surgery and risk of relapse are not consistent.[63] The optimal management of recurrent RCCs is unclear. In asymptomatic recurrences, monitoring could be adopted, whereas in symptomatic ones, aspiration and, where possible, removal of the cyst wall have been used.[73,75] Radiotherapy after surgical intervention for a recurrent mass has also been described.[75]

KEY POINTS

- Asymptomatic Rathke's cleft cysts require no active treatment and they usually remain stable or show spontaneous involution. In symptomatic cases, surgery aiming to drain the cyst content and safely remove as much of the capsule is indicated. Recurrence after surgery is not uncommon and long-term endocrine morbidity can be significant, necessitating long-term follow-up.

MENINGIOMAS

Parasellar meningiomas account for 15% of all meningiomas and arise from arachnoidal cells of the anterior clinoid, tuberculum sellae, diaphragma sellae, or inner wing of the sphenoid.[79] Most are WHO grade I with a meningoepithelial subtype, whereas atypical meningiomas are rare.[80,81] They are more common in females. *TRAF7*, *AKT1*, and *KLF4* mutations have been reported to be common in meningiomas of the parasellar region.[82]

On CT, they appear as homogeneous masses, hyper- or isodense compared with gray matter, often with calcification present. On T1-weighted MRI, they are isointense or occasionally hypointense, with homogenous enhancement after gadolinium (Fig. 11.3). On T2-weighted MRI they can be isointense or hyperintense. Adjacent dural contrast enhancement, the dural "tail" is frequently observed and is a pathognomonic feature for meningiomas.[83] Differential diagnosis includes schwannomas, chondrosarcomas, chordomas, metastases, myelomas, and lymphomas, and a biopsy may be needed to confirm the diagnosis.

Depending on their location, they can present with visual deterioration related to optic nerve compression, oculomotor nerve palsy, trigeminal neuralgia, hypoesthesia or dysesthesia, facial palsy, cochlear and vestibular dysfunctions, epilepsy, neurocognitive dysfunction, aphasia, hydrocephalus, or symptoms of brain stem compression.[80] Parasellar meningiomas may cause hyperprolactinemia, but pituitary dysfunction is generally uncommon.

Management includes observation (in cases of nongrowing, asymptomatic lesions), surgery, surgery combined or not with radiotherapy (in cases of residual WHO grade II or WHO grade I meningiomas with aggressive features), or radiotherapy alone (for asymptomatic or mildly symptomatic parasellar meningiomas).[84] Tumor control rates are reported as more than 90% at 5 years after postoperative irradiation (fractionated stereotactic radiotherapy or gamma knife radiosurgery or cyber knife radiosurgery or proton beam radiotherapy).[84] Medical therapies are rarely needed for parasellar meningiomas; they could be considered for aggressive tumors and include the combination of everolimus and octreotide or targeted therapies.[80] Peptide receptor radionuclide therapy could be an alternative to conventional radiotherapy in extended skull base meningiomas with a low growth rate; however, this approach requires validation.

Long-term prognosis is usually favorable, but long-term follow-up is required because of the risk of recurrence. Assessment of the pituitary function is needed postoperatively in cases of surgical removal, whereas long-term endocrine follow-up is recommended when radiotherapy has been offered. Rare aggressive cases with multiple recurrences have been described.

KEY POINTS

- Parasellar meningiomas account for 15% of all meningiomas. Management is individualized, and these tumors overall have favorable long-term prognosis.

CHORDOMAS

Chordomas are low-grade, slow-growing, but locally invasive and locally aggressive tumors arising from remnants of the notochord. Skull base chordomas account for around 35% of the chordomas,[85] and based on the WHO classification they are divided in three groups: conventional, chondroid, and dedifferentiated. Most parasellar chordomas arise from the clivus. They express the transcription factor Brachyury (T), which can be used as a marker to discriminate between these tumors and other skull base lesions.[82] *PI3K* and *LYST* mutations have also been found in a small number of cases, representing plausible driver mutations for the tumor.[83,86]

Presenting manifestations of skull base and clival chordomas include headaches and/or cranial nerve dysfunctions (most frequently cranial nerve VI). Rarely, CSF leak may be detected.[87]

On CT, they are hypodense and may show calcification. On MRI, chordomas are less intense than the clivus in T1 sequences, enhance heterogeneously after gadolinium administration, and are hyperintense on T2-weighted images.[82]

First-line treatment is radical resection, particularly given that the extent of removal is a major prognostic factor.[85] Adjuvant radiotherapy is offered in cases of residual tumor, and this has to be at relatively high doses (up to 70–75 Gy), because they are relatively radioresistant.[84,88] Local control rates at 5-year follow-up have been reported at 65% with intensity-modulated radiation therapy (5-year survival 86%)[89] and 50% with fractionated stereotactic radiation therapy (5-year survival 82%).[90] Stereotactic radiosurgery may be offered for small residual or recurrent tumors, with local control rates between 21% and 72% at 5 years.[84] Currently, surgery with adjuvant proton beam radiotherapy is becoming the treatment of choice for these tumors.[85] Long-term follow-up is needed for these tumors, and in case of local or metastatic recurrence, surgical excision and radiotherapy can be considered. Chordomas are resistant to conventional chemotherapeutic agents.

KEY POINTS

- Chordomas are low-grade, slow-growing but locally invasive and locally aggressive tumors arising from remnants of the notochord. Skull base chordomas account for around 35% of the chordomas. For their management, radical resection is recommended and, when radiotherapy is offered, this requires to be in high doses due to their radioresistance. Currently, surgery with adjuvant proton beam radiotherapy is becoming the treatment of choice.

GLIOMAS

Gliomas in the parasellar area arise in the hypothalamus or optic pathway. Pathological subtypes include pilocytic astrocytoma (the most common, characterized as WHO grade I tumors), or diffuse astrocytomas (WHO grade II tumors, which have a tendency to malignant progression to anaplastic astrocytoma or glioblastoma).[91] When they involve the optic pathways, they may be linked to neurofibromatosis type 1 and can be multiple. Mutation of genes encoding proteins involved the MAPK pathway have been described in pilocytic astrocytomas.[92]

They can present with visual deterioration or retroorbital pain, proptosis, headaches, or hypothalamopituitary dysfunction (including obesity and diabetes insipidus).

On imaging, they appear as mixed tumors with cystic component and heterogeneous enhancement. The solid components are iso- or hypointense on T1-weighted images and hyper- or isointense on T2-weighted images.

The prognosis of pilocytic astrocytomas is generally optimal, with greater than 95% 10-year survival rates following surgery. Patient age and extent of tumor resection are prognostic factors. Progression to malignancy is rare in pilocytic astrocytomas. Chemotherapy does not improve progression-free survival, whereas topoisomerase inhibitors in combination with chemotherapy and new molecular agents targeting the RAS/ERK and related signaling pathways are promising novel therapeutic approaches.[91]

KEY POINTS

- Gliomas in the parasellar area arise in the hypothalamus or optic pathway. The most common pathological subtype is pilocytic astrocytoma, which has optimal prognosis after surgical resection.

GERM CELL TUMORS

Primary intracranial germ cell tumors are mostly diagnosed during the first two decades of life, and the majority arise along the midline axis, from the pineal gland to the neurohypophyseal/infundibular stalk. The most common are germinomas developing in the suprasellar area. Nongerminomatous germ cell tumors include teratomas, yolk sac tumors, embryonal carcinoma, and choriocarcinoma.[91] The main molecular alterations in germinomas include changes in the MAPK and/or phosphatidylinositol 3-kinase pathway, global low DNA methylation and chromosomal instability.[93]

They manifest as large lesions or thickening of the pituitary stalk and cause visual deterioration, diabetes insipidus, or anterior pituitary dysfunction. Precocious puberty may also be a presenting manifestation because of the secretion of human chorionic gonadotropin (hCG) by neoplastic syncytiotrophoblasts.

On CT, they usually appear as hyperdense lesions with inhomogenous enhancement and calcification. On MRI, they show variable signal intensity on T1-weighted images, whereas on T2-weighted images they are slightly isointense or hyperintense. Cystic changes may be also seen.[93]

The diagnosis is suggested by characteristic imaging features and elevated beta-hCG and/or alpha-fetoprotein in the blood or CSF. In cases with normal beta-hCG and/or alpha-fetoprotein levels, surgical biopsy for establishing the diagnosis needs to be considered.[94]

Germinomas are very radiosensitive, and craniospinal irradiation is associated with high long-term survival rates. Neoadjuvant chemotherapy may offer comparable control at lower radiation doses. In cases of relapse or metastasis, two-course chemotherapy followed by radiotherapy to the tumor region and craniospinal irradiation have shown a 98% 5-year progression-free survival rate. Nongerminomatous germ cell tumors have a more aggressive behavior, and 5-year survival rates range between 60% to 80%, after a combination of chemotherapy and irradiation. Cyclin-dependent kinase inhibitors and molecular targeted agents targeting KIT and the AKT/mTOR pathway are management options under investigation.[94]

KEY POINTS

- Primary intracranial germ cell tumors (germinomas or nongerminomatous germ cell tumors) mostly arise along the midline axis, from the pineal gland to the neurohypophyseal/infundibular stalk. They are usually diagnosed in the first two decades of life. Germinomas are very radiosensitive, whereas nonerminomatous germ cell tumors have a more aggressive behavior.

LANGERHANS CELL HISTIOCYTOSIS

LCH is a rare condition characterized by abnormal proliferation of Langerhans cells (dendritic antigen-presenting cells). It can involve any organ, including the hypothalamopituitary system. It can present at any age, but it is more prevalent in children aged below 15 years.

LCH in the parasellar area presents with diabetes insipidus and anterior pituitary dysfunction. There are no specific appearances on imaging (loss of bright spot of posterior pituitary, infundibular enlargement, hypothalamic lesion).

Diagnosis is established by biopsy of the affected area (usually the skin). Once diagnosis is confirmed, evaluation for systematic involvement is necessary.

Various chemotherapy protocols and radiotherapy are management options, with multisystem disease having worse prognosis.[95,96]

KEY POINTS

- Langerhans'cell histiocytosis is more prevalent in children and when affects the hypothalamo-pituitary system, it presents with pituitary dysfunction. Evaluation of systematic involvement is necessary and multisystem disease has worse prognosis. Chemotherapy and radiotherapy are management options.

HAMARTOMAS

Hamartomas are congenital lesions mainly detected in the hypothalamus and are mostly found in young children. Histopathologically, they consist of small neurons intermixed with glia and large neurons.

Their presenting manifestations include central precocious puberty attributed to ectopic release of gonadotropin-releasing hormone and gelastic seizures that may also be related to neurobehavioral symptoms (including cognitive deficits, attention deficit hyperactivity disorder, mood disorder).[97]

Radiologically, they appear as small nodular lesions that are iso- or slightly hypodense on CT and isointense or slightly hypointense on T1- and T2-weighted images.[91]

Observation is recommended when minor clinical manifestations are present (e.g., brief and infrequent episodes of gelastic seizures). Antiepileptic medications are not effective in controlling gelastic seizures. Surgery provides the most optimal results in seizure control and halts neurobehavioral manifestations. In cases where the tumor is difficult to resect without causing neuroendocrine damage, gamma knife radiosurgery is an alternative.[98-99] Precocious puberty is treated successfully with gonadotropin hormone–releasing agonists.

KEY POINTS

- Hamartomas are congenital lesions mainly detected in the hypothalamus. They can present with precocious puberty and gelastic seizures. Main management options include observation, surgery or radiotherapy. Precocious puberty is treated successfully with gonadotropin hormone releasing agonists.

PITUITARY METASTASIS

Pituitary metastasis is a rare manifestation of malignancy, with reported prevalences of 0.4% in radiological, 1% in surgical, and between 0.14% and 28.1% in autopsy series. Pathophysiological mechanisms are direct hematogenous spread, spread via the hypothalamic-portal vessels, meningeal spread via the suprasellar cistern, or extension from juxtasellar or skull base metastases.

Clinical manifestations include visual field impairment or optic neuropathy, cranial nerve palsies, anterior pituitary dysfunction or diabetes insipidus. Diabetes insipidus points toward the diagnosis of pituitary metastasis in a patient with known malignancy; nonetheless, contemporary data suggest that its frequency as a presenting manifestation is reducing, possibly because of earlier detection of pituitary metastatic infiltration, increased awareness of manifestations of hypopituitarism, and better availability and wider use of relevant biochemical testing. Pituitary metastasis is most frequently detected in patients with a known history of cancer (breast and lung are the most common primary sites). It may also be the first manifestation of malignancy or can be detected incidentally.

On imaging, they present as sellar or suprasellar masses or stalk thickening. Their radiological features can be nonspecific and are often not helpful in discriminating from pituitary adenoma or other sellar lesion. On MRI, they can be hypointense on T1-weighted images and hypo- or hyperintense on T2-weighted images, and they show postcontrast enhancement. Rapid growth and invasion of parasellar structures should raise suspicion of pituitary metastatic infiltration. Increased uptake on (^{18}F)-fluorodeoxyglucose positron emission tomography/CT has been detected; this is a nonspecific finding, as it can also be seen in other pathologies or even in normal pituitary tissue.

The management of pituitary metastasis aims to provide symptomatic relief (mainly of visual disturbances), with a favorable impact on quality of life, and to prevent further enlargement of the lesion. It includes surgery and radiotherapy (alone or in combination), as well as (depending on the primary malignancy) chemotherapy, immunotherapy, or hormonal therapy. Survival after the diagnosis of pituitary metastasis depends on the type and stage of primary cancer, and appears to be improving over time, possibly because of earlier diagnosis of pituitary metastatic infiltration and to improved life expectancy of cancer patients.[100]

KEY POINTS

- Pituitary metastasis is a rare manifestation of malignancy and can present as visual field impairment or optic neuropathy, cranial nerve palsies, anterior pituitary dysfunction or diabetes insipidus. Radiological features can be non-specific but rapid growth of a lesion in a patient with known malignancy should raise suspicion of pituitary metastatic infiltration.
- The management of pituitary metastasis aims to provide symptomatic relief and to prevent further enlargement of the lesion. It includes surgery, radiotherapy and depending on the primary malignancy, chemotherapy, immunotherapy or hormonal therapy.

REFERENCES

1. Karavitaki N, Cudlip S, Adams CB, et al. Craniopharyngiomas. *Endocr Rev*. 2006;27:371–397.
2. Nielsen EH, Feldt-Rasmussen U, Poulsgaard L, et al. Incidence of craniopharyngioma in Denmark (n=189) and estimated world incidence of craniopharyngioma in children and adults. *J Neuro Oncol*. 2011;104:755–763.
3. Scagliotti V, Avagliano L, Gualtieri A, et al. Histopathology and molecular characterisation of intrauterine-diagnosed congenital craniopharyngioma. *Pituitary*. 2016;19:50–56.
4. Bunin GR, Surawicz TS, Witman PA, et al. The descriptive epidemiology of craniopharyngioma. *J Neurosurg*. 1998;89:547–551.
5. Sorva R, Heiskanen O. Craniopharyngioma in Finland. A study of 123 cases. *Acta Neurochir (Wien)*. 1986;81:85–89.

6. Crotty TB, Scheithauer BW, Young Jr WF, et al. Papillary craniopharyngioma: a clinicopathological study of 48 cases. *J Neurosurg*. 1995;83:206–214.
7. Buslei R, Nolde M, Hofmann B, et al. Common mutations of beta-catenin in adamantinomatous craniopharyngiomas but not in other tumors originating from the sellar region. *Acta Neuropathol*. 2005;109:589–597.
8. Larkin S, Karavitaki N. Recent advances in molecular pathology of craniopharyngioma. *F1000Res*. 2017;6:1202.
9. Muller HL, Merchant TE, Puget S, et al. New outlook on the diagnosis, treatment and follow-up of childhood-onset craniopharyngioma. *Nat Rev Endocrinol*. 2017;13:299–312.
10. Karavitaki N, Brufani C, Warner JT, et al. Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endocrinol (Oxf)*. 2005;62:397–409.
11. Gautier A, Godbout A, Grosheny C, et al. Markers of recurrence and long-term morbidity in craniopharyngioma: a systematic analysis of 171 patients. *J Clin Endocrinol Metab*. 2012;97:1258–1267.
12. Nielsen EH, Jorgensen JO, Bjerre P, et al. Acute presentation of craniopharyngioma in children and adults in a Danish national cohort. *Pituitary*. 2013;16:528–535.
13. Muller HL, Emser A, Faldum A, et al. Longitudinal study on growth and body mass index before and after diagnosis of childhood craniopharyngioma. *J Clin Endocrinol Metab*. 2004;89:3298–3305.
14. Fahlbusch R, Honegger J, Paulus W, et al. Surgical treatment of craniopharyngiomas: experience with 168 patients. *J Neurosurg*. 1999;90:237–250.
15. De Vile CJ, Grant DB, Kendall BE, et al. Management of childhood craniopharyngioma: can the morbidity of radical surgery be predicted? *J Neurosurg*. 1996;85:73–81.
16. Van Effenterre R, Boch AL. Craniopharyngioma in adults and children: a study of 122 surgical cases. *J Neurosurg*. 2002;97:3–11.
17. Tan TSE, Patel L, Gopal-Kothandapani JS, et al. The neuroendocrine sequelae of paediatric craniopharyngioma: a 40-year meta-data analysis of 185 cases from three UK centres. *Eur J Endocrinol*. 2017;176:359–369.
18. Buchfelder M, Schlaffer SM, Lin F, et al. Surgery for craniopharyngioma. *Pituitary*. 2013;16:18–25.
19. Elliott RE, Jane Jr JA, Wisoff JH. Surgical management of craniopharyngiomas in children: meta-analysis and comparison of transcranial and transsphenoidal approaches. *Neurosurgery*. 2011;69:630–643;discussion 643.
20. Albano L, Losa M, Flickinger J, et al. Radiotherapy of parasellar tumors. *Neuroendocrinology*. 2020;110:848–858.
21. Rutenberg MS, Rotondo RL, Rao D, et al. Clinical outcomes following proton therapy for adult craniopharyngioma: a single-institution cohort study. *J Neuro Oncol*. 2020;147:387–395.
22. Yang I, Sughrue ME, Rutkowski MJ, et al. Craniopharyngioma: a comparison of tumor control with various treatment strategies. *Neurosurg Focus*. 2010;28:E5.
23. Minniti G, Saran F, Traish D, et al. Fractionated stereotactic conformal radiotherapy following conservative surgery in the control of craniopharyngiomas. *Radiother Oncol*. 2007;82:90–95.
24. Luu QT, Loredo LN, Archambeau JO, et al. Fractionated proton radiation treatment for pediatric craniopharyngioma: preliminary report. *Cancer J*. 2006;12:155–159.
25. Jose CC, Rajan B, Ashley S, et al. Radiotherapy for the treatment of recurrent craniopharyngioma. *Clin Oncol (R Coll Radiol)*. 1992;4:287–289.
26. Pollock BE, Lunsford LD, Kondziolka D, et al. Phosphorus-32 intracavitary irradiation of cystic craniopharyngiomas: current technique and long-term results. *Int J Radiat Oncol Biol Phys*. 1995;33:437–446.
27. Voges J, Sturm V, Lehrke R, et al. Cystic craniopharyngioma: long-term results after intracavitary irradiation with stereotactically applied colloidal beta-emitting radioactive sources. *Neurosurgery*. 1997;40:263–269; discussion 269–270.
28. Van den Berge JH, Blaauw G, Breeman WA, et al. Intracavitary brachytherapy of cystic craniopharyngiomas. *J Neurosurg*. 1992;77:545–550.
29. Hasegawa T, Kondziolka D, Hadjipanayis CG, et al. Management of cystic craniopharyngiomas with phosphorus-32 intracavitary irradiation. *Neurosurgery*. 2004;54:813–820;discussion 820–812.
30. Julow JV. Intracystic irradiation for craniopharyngiomas. *Pituitary*. 2013;16:34–45.
31. Kickingereder P, Maarouf M, El Majdoub F, et al. Intracavitary brachytherapy using stereotactically applied phosphorus-32 colloid for treatment of cystic craniopharyngiomas in 53 patients. *J Neuro Oncol*. 2012;109:365–374.
32. Takahashi H, Yamaguchi F, Teramoto A. Long-term outcome and reconsideration of intracystic chemotherapy with bleomycin for craniopharyngioma in children. *Childs Nerv Syst*. 2005;21:701–704.
33. Hukin J, Steinbok P, Lafay-Cousin L, et al. Intracystic bleomycin therapy for craniopharyngioma in children: the Canadian experience. *Cancer*. 2007;109:2124–2131.
34. Kilday JP, Caldarelli M, Massimi L, et al. Intracystic interferon-alpha in pediatric craniopharyngioma patients: an international multicenter assessment on behalf of SIOPE and ISPN. *Neuro Oncol*. 2017;19:1398–1407.
35. Xu Z, Yen CP, Schlesinger D, et al. Outcomes of gamma knife surgery for craniopharyngiomas. *J Neuro Oncol*. 2011;104:305–313.
36. Niranjan A, Kano H, Mathieu D, et al. Radiosurgery for craniopharyngioma. *Int J Radiat Oncol Biol Phys*. 2010;78:64–71.
37. Kobayashi T, Kida Y, Mori Y, et al. Long-term results of gamma knife surgery for the treatment of craniopharyngioma in 98 consecutive cases. *J Neurosurg*. 2005;103:482–488.
38. Bremer AM, Nguyen TQ, Balsys R. Therapeutic benefits of combination chemotherapy with vincristine, BCNU, and procarbazine on recurrent cystic craniopharyngioma. A case report. *J Neuro Oncol*. 1984;2:47–51.
39. Lippens RJ, Rotteveel JJ, Otten BJ, et al. Chemotherapy with Adriamycin (doxorubicin) and CCNU (lomustine) in four children with recurrent craniopharyngioma. *Eur J Paediatr Neurol*. 1998;2:263–268.
40. Yeung JT, Pollack IF, Panigrahy A, et al. Pegylated interferon-alpha-2b for children with recurrent craniopharyngioma. *J Neurosurg Pediatr*. 2012;10:498–503.
41. Goldman S, Pollack IF, Jakacki RI, et al. Phase II study of peginterferon alpha-2b for patients with unresectable or recurrent craniopharyngiomas: a Pediatric Brain Tumor Consortium report. *Neuro Oncol*. 2020;22:1696–1704.
42. Alexandraki KI, Kaltsas GA, Karavitaki N, et al. The medical therapy of craniopharyngiomas: the way ahead. *J Clin Endocrinol Metab*. 2019;104:5751–5764.
43. Verhelst J, Kendall-Taylor P, Erfurth EM, et al. Baseline characteristics and response to 2 years of growth hormone (GH) replacement of hypopituitary patients with GH deficiency due to adult-onset craniopharyngioma in comparison with patients with nonfunctioning pituitary adenoma: data from KIMS (Pfizer International Metabolic Database). *J Clin Endocrinol Metab*. 2005;90:4636–4643.
44. Karavitaki N. Hypopituitarism oddities: craniopharyngiomas. *Horm Res*. 2007;68:151–153.
45. Karavitaki N, Warner JT, Marland A, et al. GH replacement does not increase the risk of recurrence in patients with craniopharyngioma. *Clin Endocrinol (Oxf)*. 2006;64:556–560.
46. Daousi C, Dunn AJ, Foy PM, et al. Endocrine and neuroanatomic features associated with weight gain and obesity in adult patients with hypothalamic damage. *Am J Med*. 2005;118:45–50.
47. Stripp DC, Maity A, Janss AJ, et al. Surgery with or without radiation therapy in the management of craniopharyngiomas in children and young adults. *Int J Radiat Oncol Biol Phys*. 2004;58:714–720.
48. Pereira AM, Schmid EM, Schutte PJ, et al. High prevalence of long-term cardiovascular, neurological and psychosocial morbidity after treatment for craniopharyngioma. *Clin Endocrinol (Oxf)*. 2005;62:197–204.
49. van Iersel L, van Santen HM, Brokke KE, et al. Pathophysiology and individualized treatment of hypothalamic obesity following craniopharyngioma and other suprasellar rumors: a systematic review. *Endocr Rev*. 2018;40:193–235.
50. de Vile CJ, Grant DB, Hayward RD, et al. Obesity in childhood craniopharyngioma: relation to post-operative hypothalamic damage shown by magnetic resonance imaging. *J Clin Endocrinol Metab*. 1996;81:2734–2737.
51. Holmer H, Pozarek G, Erfurth E-M, et al. Reduced energy expenditure and impaired feeding-related signals but not high energy intake reinforces hypothalamic obesity in adults with childhood onset craniopharyngioma. *J Clin Endocrinol Metab*. 2010;95:5395–5402.

52. Goldstone AP, Patterson M, Kalingag N, et al. Fasting and postprandial hyperghrelinemia in Prader-Willi syndrome is partially explained by hypoinsulinemia, and is not due to peptide YY3-36 deficiency or seen in hypothalamic obesity due to craniopharyngioma. *J Clin Endocrinol Metab*. 2005;90:2681–2690.
53. Elowe-Gruau E, Beltrand J, Brauner R, et al. Childhood craniopharyngioma: hypothalamus-sparing surgery decreases the risk of obesity. *J Clin Endocrinol Metab*. 2013;98:2376–2382.
54. Mortini P, Gagliardi F, Bailo M, et al. Magnetic resonance imaging as predictor of functional outcome in craniopharyngiomas. *Endocrine*. 2016;51:148–162.
55. Bretault M, Boillot A, Muzard L, et al. Clinical review: bariatric surgery following treatment for craniopharyngioma: a systematic review and individual-level data meta-analysis. *J Clin Endocrinol Metab*. 2013;98:2239–2246.
56. Erfurth EM, Holmer H, Fjalldal SB. Mortality and morbidity in adult craniopharyngioma. *Pituitary*. 2013;16:46–55.
57. Wijnen M, Olsson DS, van den Heuvel-Eibrink MM, et al. Excess morbidity and mortality in patients with craniopharyngioma: a hospital-based retrospective cohort study. *Eur J Endocrinol*. 2018;178:93–102.
58. Wijnen M, Olsson DS, van den Heuvel-Eibrink MM, et al. The metabolic syndrome and its components in 178 patients treated for craniopharyngioma after 16 years of follow-up. *Eur J Endocrinol*. 2018;178:11–22.
59. Daubenbuchel AM, Hoffmann A, Gebhardt U, et al. Hydrocephalus and hypothalamic involvement in pediatric patients with craniopharyngioma or cysts of Rathke's pouch: impact on long-term prognosis. *Eur J Endocrinol*. 2015;172:561–569.
60. Karavitaki N, Wass JA. Non-adenomatous pituitary tumors. *Best Pract Res Clin Endocrinol Metab*. 2009;23:651–665.
61. Osborn AG, Preece MT. Intracranial cysts: radiologic-pathologic correlation and imaging approach. *Radiology*. 2006;239:650–664.
62. Teramoto A, Hirakawa K, Sanno N, et al. Incidental pituitary lesions in 1,000 unselected autopsy specimens. *Radiology*. 1994;193:161–164.
63. Trifanescu R, Ansorge O, Wass JA, et al. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*. 2012;76:151–160.
64. Zada G, Lin N, Ojerholm E, et al. Craniopharyngioma and other cystic epithelial lesions of the sellar region: a review of clinical, imaging, and histopathological relationships. *Neurosurg Focus*. 2010;28:E4.
65. Billeci D, Marton E, Tripodi M, et al. Symptomatic Rathke's cleft cysts: a radiological, surgical and pathological review. *Pituitary*. 2004;7:131–137.
66. Kim JE, Kim JH, Kim OL, et al. Surgical treatment of symptomatic Rathke cleft cysts: clinical features and results with special attention to recurrence. *J Neurosurg*. 2004;100:33–40.
67. Nishioka H, Haraoka J, Izawa H, et al. Magnetic resonance imaging, clinical manifestations, and management of Rathke's cleft cyst. *Clin Endocrinol (Oxf)*. 2006;64:184–188.
68. Hama S, Arita K, Nishisaka T, et al. Changes in the epithelium of Rathke cleft cyst associated with inflammation. *J Neurosurg*. 2002;96:209–216.
69. Karavitaki N, Scheithauer BW, Watt J, et al. Collision lesions of the sella: co-existence of craniopharyngioma with gonadotroph adenoma and of Rathke's cleft cyst with corticotroph adenoma. *Pituitary*. 2008;11:317–323.
70. Bonneville F, Cattin F, Marsot-Dupuch K, et al. T1 signal hyperintensity in the sellar region: spectrum of findings. *Radiographics*. 2006;26:93–113.
71. Hayashi Y, Tachibana O, Muramatsu N, et al. Rathke cleft cyst: MR and biomedical analysis of cyst content. *J Comput Assist Tomogr*. 1999;23:34–38.
72. Binning MJ, Gottfried ON, Osborn AG, et al. Rathke cleft cyst intracystic nodule: a characteristic magnetic resonance imaging finding. *J Neurosurg*. 2005;103:837–840.
73. Trifanescu R, Stavrinides V, Plaha P, et al. Outcome in surgically treated Rathke's cleft cysts: long-term monitoring needed. *Eur J Endocrinol*. 2011;165:33–37.
74. Aho CJ, Liu C, Zelman V, et al. Surgical outcomes in 118 patients with Rathke cleft cysts. *J Neurosurg*. 2005;102:189–193.
75. Mukherjee JJ, Islam N, Kaltsas G, et al. Clinical, radiological and pathological features of patients with Rathke's cleft cysts: tumors that may recur. *J Clin Endocrinol Metab*. 1997;82:2357–2362.
76. Sanno N, Oyama K, Tahara S, et al. A survey of pituitary incidentaloma in Japan. *Eur J Endocrinol*. 2003;149:123–127.
77. Amhaz HH, Chamoun RB, Waguespack SG, et al. Spontaneous involution of Rathke cleft cysts: is it rare or just underreported? *J Neurosurg*. 2010;112:1327–1332.
78. Lin M, Wedemeyer MA, Bradley D, et al. Long-term surgical outcomes following transsphenoidal surgery in patients with Rathke's cleft cysts. *J Neurosurg*. 2018;130:831–837.
79. FitzPatrick M, Tartaglino LM, Hollander MD, et al. Imaging of sellar and parasellar pathology. *Radiol Clin North Am*. 1999;37:101–121.
80. Graillon T, Regis J, Barlier A, et al. Parasellar meningiomas. *Neuroendocrinology*. 2020;110:780–796.
81. Maiuri F, Mariniello G, Guadagno E, et al. WHO grade, proliferation index, and progesterone receptor expression are different according to the location of meningioma. *Acta Neurochir (Wien)*. 2019;161:2553–2561.
82. Gatto F, Perez-Rivas LG, Olarescu NC, et al. Diagnosis and treatment of parasellar lesions. *Neuroendocrinology*. 2020;110:728–739.
83. Huang RY, Bi WL, Griffith B, et al. Imaging and diagnostic advances for intracranial meningiomas. *Neuro Oncol*. 2019;21:i44–i61.
84. Albano L, Losa M, Flickinger J, et al. Radiotherapy of parasellar tumors. *Neuroendocrinology*. 2020;110:848–858.
85. George B, Bresson D, Herman P, et al. Chordomas: a review. *Neurosurg Clin N Am*. 2015;26:437–452.
86. Tarpey PS, Behjati S, Young MD, et al. The driver landscape of sporadic chordoma. *Nat Commun*. 2017;8:890.
87. Tenny S, Varacallo M. Chordoma. In: *StatPearls [Internet]*. Treasure Island: StatPearls Publishing; 2020.
88. Stacchiotti S, Sommer J, Chordoma Global Consensus Group. Building a global consensus approach to chordoma: a position paper from the medical and patient community. *Lancet Oncol*. 2015;16:e71–e83.
89. Sahgal A, Chan MW, Atenafu EG, et al. Image-guided, intensity-modulated radiation therapy (IG-IMRT) for skull base chordoma and chondrosarcoma: preliminary outcomes. *Neuro Oncol*. 2015;17:889–894.
90. Debus J, Hug EB, Liebsch NJ, et al. Brainstem tolerance to conformal radiotherapy of skull base tumors. *Int J Radiat Oncol Biol Phys*. 1997;39:967–975.
91. Kaltsas GA, Kolomodi D, Randeva H, et al. Nonneuroendocrine neoplasms of the pituitary region. *J Clin Endocrinol Metab*. 2019;104:3108–3123.
92. Jones DT, Hutter B, Jäger N, et al. International cancer genome consortium PedBrain tumor project. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat Genet*. 2013;45:927–932.
93. Fukushima S, Yamashita S, Kobayashi H, et al. Intracranial Germ Cell Tumor Genome Analysis Consortium (The iGCTConsortium). Genome-wide methylation profiles in primary intracranial germ cell tumors indicate a primordial germ cell origin for germinomas. *Acta Neuropathol*. 2017;133:445–462.
94. Murray MJ, Bartels U, Nishikawa R, et al. Consensus on the management of intracranial germ-cell tumors. *Lancet Oncol*. 2015;16:e470–e477.
95. Kaltsas GA, Evanson J, Chrisoulidou A, et al. The diagnosis and management of parasellar tumors of the pituitary. *Endocr Relat Cancer*. 2008;15:885–903.
96. Yeh EA, Greenberg J, Abla O, et al. Evaluation and treatment of Langerhans cell histiocytosis patients with central nervous system abnormalities: current views and new vistas. *Pediatr Blood Cancer*. 2018;65.
97. Mittal S, Mittal M, Montes JL, et al. Hypothalamic hamartomas. Part 1. Clinical, neuroimaging, and neurophysiological characteristics. *Neurosurg Focus*. 2013;34:E6.
98. Mittal S, Mittal M, Montes JL, et al. Hypothalamic hamartomas. Part 2. Surgical considerations and outcome. *Neurosurg Focus*. 2013;34:E7.
99. Régis J, Lagmari M, Carron R, et al. Safety and efficacy of Gamma Knife radiosurgery in hypothalamic hamartomas with severe epilepsies: a prospective trial in 48 patients and review of the literature. *Epilepsia*. 2017;58:60–71.
100. Lithgow K, Siqueira I, Senthil L, et al. Pituitary metastases: presentation and outcomes from a pituitary center over the last decade. *Pituitary*. 2020;23:258–265.

12

Disorders of Sodium, Diabetes Insipidus and Hyponatremia

Christopher J. Thompson and Aoife Garrah

OUTLINE

Introduction, 181
Physiology of Sodium and Water Homeostasis, 181
  Synthesis and Secretion of Vasopressin, 181
  Cellular Action of Vasopressin, 182
  Osmoregulation of Vasopressin Release, 182
  Baroregulation of Vasopressin Release, 184
  Other Mechanisms Regulating Vasopressin Release, 185
  Secretion of Copeptin, 185
Diabetes Insipidus, 185
  Hypothalamic Diabetes Insipidus, 185
  Nephrogenic Diabetes Insipidus, 187
  Primary Polydipsia, 187
  Investigation of Diabetes Insipidus, 187
  Adipsic Diabetes Insipidus, 189
  Treatment of Diabetes Insipidus, 189
Hyponatremia, 190
  Morbidity and Mortality, 190
  Differential Diagnosis of Hyponatremia, 190
The Syndrome of Inappropriate Antidiuresis, 191
  Introduction, 191
  Causes of Syndrome of Inappropriate Antidiuresis, 191
  Pathophysiology of Syndrome of Inappropriate Antidiuresis, 192
  Diagnosis of Syndrome of Inappropriate Antidiuresis, 193
  Differential Diagnosis of Syndrome of Inappropriate Antidiuresis, 193
Treatment of Hyponatremia, 193
  Factors Affecting Therapy, 193
  Treatment of Chronic Hyponatremia Owing to Syndrome of Inappropriate Antidiuresis, 194
  Treatment of Acute Hyponatremia, 195
Conclusions, 196

INTRODUCTION

Normal human physiology depends on the preservation of the tonicity of the extracellular fluid (ECF). A sophisticated process of osmoregulation, which integrates the secretion and action of arginine vasopressin (AVP) and the regulation of thirst and fluid intake, maintains plasma osmolality within tightly defined limits. In healthy individuals, plasma osmolality rarely deviates by more than 2% from baseline, so abnormal plasma sodium concentrations reflect significant disturbances of normal physiology. Abnormal osmoregulation can occur as a result of AVP deficiency, which manifests as diabetes insipidus (DI), inappropriate secretion of AVP, which is the commonest cause of hyponatremia, and a number of other, less common, syndromes.

In this chapter, we will discuss the physiological control of secretion of AVP, the renal antidiuretic actions of AVP, and the interaction with the regulation of thirst. We will then apply these physiological principles to the pathophysiological mechanisms that are the basis for DI and hyponatremia.

PHYSIOLOGY OF SODIUM AND WATER HOMEOSTASIS

Synthesis and Secretion of Vasopressin

AVP is a nonapeptide synthesized in the hypothalamus that has important antidiuretic and vasopressor actions. There is considerable structural homology between AVP and oxytocin, which is also secreted from the posterior pituitary (Fig. 12.1).[1] Vasopressin is synthesized predominantly in the cell bodies of magnocellular neurones of the paraventricular nuclei (PVN) and supraoptic (SON) nuclei of the anterior hypothalamus (Fig. 12.2). The hormone is carried within neurosecretory granules down the predominantly unmyelinated axons of the magnocellular neurons, in the supraoptic-hypophyseal tract, at a speed of 2 mm per hour. The precursor undergoes successive cleavage by basic endopeptidases.[2] These axons terminate in the posterior pituitary, which is rich in fenestrated capillaries.

The gene encoding the AVP precursor is expressed mainly in the hypothalamus, and to a lesser extent in other tissues including the adrenal glands, the gonads, the cerebellum, and pituicytes in the posterior pituitary. The AVP and oxytocin genes are closely linked on chromosome 20. Posttranslational processing forms prepro-AVP-NPII, which is the precursor from which AVP is derived. The precursor consists of a signal peptide, AVP, neurophysin II (NPII, 95 amino acids), and a glycopeptide, copeptin (39 amino acids) (Fig. 12.3).

NPII is cleaved from prepro-AVP-NPII and stored in a molecule complex with AVP, prior to cleavage into the component parts and subsequent secretion into the bloodstream by exocytosis. Vasopressinergic neurons open voltage-gated calcium channels in nerve-terminal membranes. This transient calcium ion influx results in fusion of neurosecretory granules with the nerve terminal membrane and the release of AVP and NPII in equimolar amounts into the circulation.[3] Copeptin, which has unknown biological actions, is cosecreted with AVP into the circulation, and has been shown to be a stable indirect marker of AVP release.

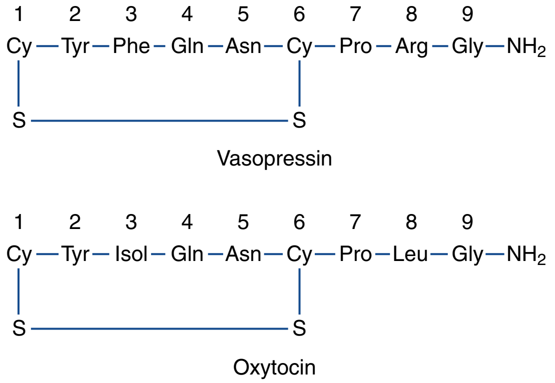

1 2 3 4 5 6 7 8 9
Cy—Tyr—Phe—Gln—Asn—Cy—Pro—Arg—Gly—NH2
S
S
Vasopressin
1 2 3 4 5 6 7 8 9
Cy—Tyr—Isol—Gln—Asn—Cy—Pro—Leu—Gly—NH2
S
S
Oxytocin

Fig. 12.1 Amino Acid Structure of Vasopressin and Oxytocin. Note that only positions 3 and 8 have different amino acids.

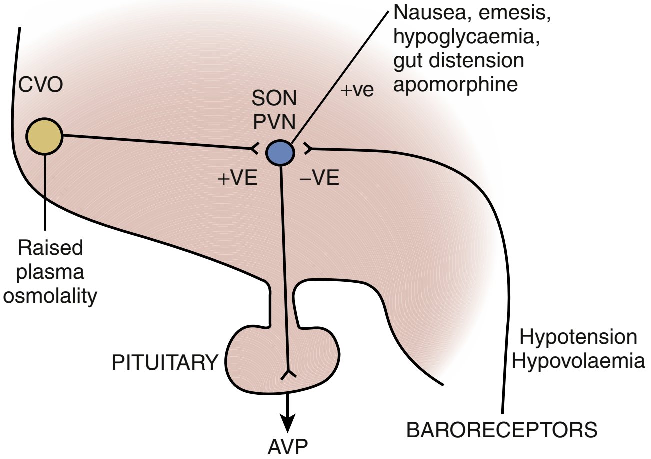


Fig. 12.2 Factors Governing Arginine Vasopressin (*AVP*) Secretion. *CVO,* Circumventricular organ (site of osmoreceptors); *PVN,* paraventricular nuclei; *SON,* supraoptic nuclei.

In addition to magnocellular neurons, which project from the SON and PVN to the posterior pituitary, there are also parvocellular vasopressinergic neurons in the PVN and the suprachiasmatic nucleus. These smaller neurons project via the median eminence to the portal vascular system of the anterior pituitary and secrete AVP and corticotrophin-releasing hormone (CRH), which in turn stimulates secretion of adrenocorticotropic hormone (ACTH) from the anterior pituitary gland. Some parvocellular vasopressinergic neurons also project from the PVN to the forebrain, brain stem, and spinal cord. AVP is also released into the central nervous system, where it acts as a neurotransmitter and/or neuromodulator.

Cellular Action of Vasopressin

AVP binds to seven-transmembrane domain G protein-coupled receptors on the plasma membrane of target cells. Three subtypes of AVP receptor exist (V1–V3), with distinct distributions, actions, and signal transduction.

The V1 Receptor. The V1 receptor is found on the cell surface of cells in vascular smooth muscle, liver, myometrium, and platelets, as well as multiple sites in the central nervous system. AVP binding to the receptor causes activation of Gq/11-mediated phospholipase C, resulting in an increase in intracellular calcium.

Binding of AVP to V1 receptors at physiological plasma concentrations has been shown to exert a weak pressor effect.[4] AVP-mediated vasoconstriction is most noticeable in the splanchnic, renal, and hepatic artery beds. The powerful vasopressor effects of high plasma concentrations of AVP is, however, clinically significant. The pressor effect of AVP is utilized in the management of bleeding esophageal varices to promote vasoconstriction and support arterial blood pressure. Rebound hypotension on withdrawal of treatment, coupled with concerns regarding reduced cardiac output, pulmonary hypertension, hyponatremia, and bowel ischemia, have limited the use of AVP in the setting of septic shock.

The V2 Receptor. The V2 receptor is predominantly expressed on the basolateral membrane of cells in the distal convoluted tubule and collecting ducts of the kidney. Binding of AVP to renal V2 receptors stimulates the recruitment of selective water channels (aquaporins), which allow reabsorption of renal tubular water and concentration of urine. Extrarenal effects of V2 activation include factor VII release from hepatocytes and release of von Willebrand factor from vascular endothelium. AVP also causes activation of liver glycogen phosphorylase, with release of glucose into the circulation.

When AVP binds to the V2 receptors on the basolateral membrane of the cells of the collecting ducts, it triggers an increase in cytoplasmic cyclic adenosine monophosphate, which activates protein kinase (Fig. 12.4). Subsequently, there is movement of stored aquaporin-2 to the luminal membrane of the cell. Aquaporin-2 forms tetramers in the cell membrane, with each monomer acting as a water channel[5] that allows passage of tubular water into the cell. Free water reabsorption from the distal nephron, and thus urine concentration, are directly influenced by serum AVP concentrations. This relationship was established prior to the discovery of aquaporins. The AVP–V2 receptor interaction also simultaneously causes an increase in the expression of the aquaporin-2 gene, with increased generation of aquaporin-2 mRNA.

The V3 Receptor. The V3 receptor, also known as V1b, is found on corticotroph cells in the anterior pituitary. In isolation, AVP is a weak ACTH secretagogue, but when it acts synergistically with CRH it causes significant secretion of ACTH, which is physiologically important. Parvocellular neurons that coexpress AVP and CRH project via the median eminence and terminate in the hypophyseal portal bed, which provides circulation to the anterior pituitary. AVP and CRH expression by these parvocellular neurons is subject to negative feedback control by glucocorticoids. AVP has no secretagogue properties with respect to the other anterior pituitary hormones.

KEY POINTS

- Arginine vasopressin (AVP) is a peptide hormone that is synthesized in the hypothalamus and has important antidiuretic and vasopressor actions. Three subtypes of vasopressin receptor exist (V1–V3), with distinct distribution, actions, and signal transduction. The binding of AVP to renal V2 receptors stimulates the recruitment of selective water channels (aquaporins), which allow reabsorption of renal tubular water and concentration of urine.

Osmoregulation of Vasopressin Release

Water balance in healthy individuals is so carefully regulated that plasma osmolality varies by only 1% to 2% in physiological conditions, as long as there is free access to water.[6] Changes in plasma osmolality are detected by specialized osmoreceptor cells in the anterior hypothalamus (Fig. 12.2). Fenestrations in the blood-brain barrier covering the anterior hypothalamus allow the magnocellular cells of the circumventricular organs—the subfornical organ and the organum vasculosum lamina terminalis—to be bathed in plasma. Changes in plasma

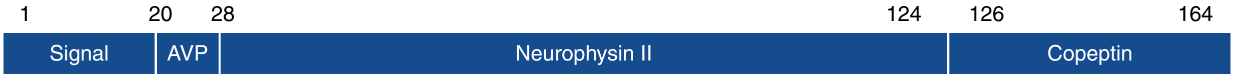


Fig. 12.3 Structure of the arginine vasopressin (*AVP*) preprohormone.

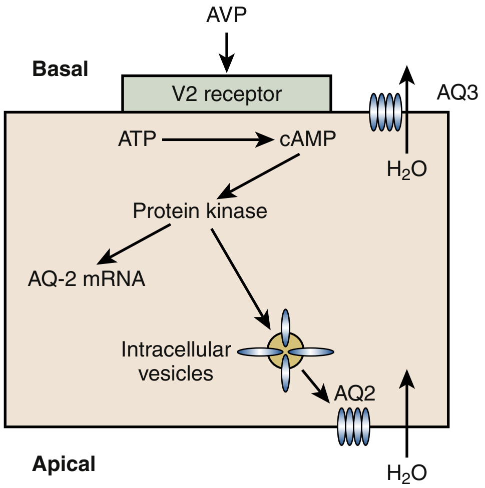


Fig. 12.4 Arginine vasopressin (*AVP*) binds to the V2 receptor and stimulates movement of aquaporin-2 (*AQ2*) to the apical membrane, where it forms a water channel for reabsorption of water (H_2O). Water moves through the apical membrane to the blood via aquaporin-3 (*AQ3*). *ATP,* Adenosine triphosphate; *cAMP,* cyclic adenosine monophosphate.

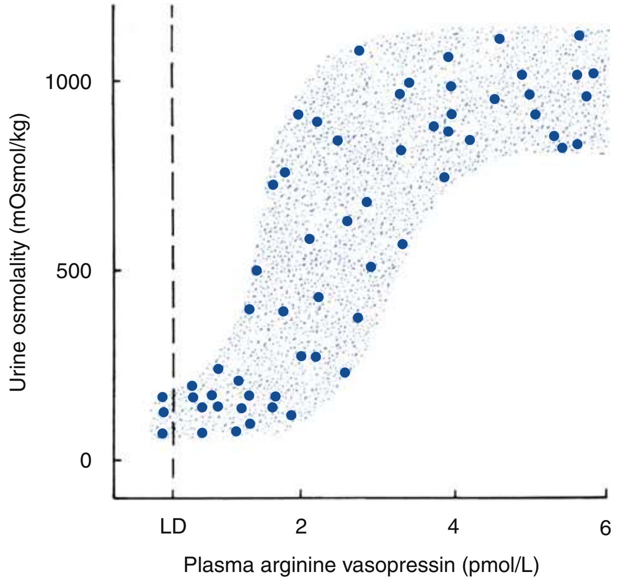


Fig. 12.5 Relationship Between Plasma Arginine Vasopressin and Urine Osmolality. Data Were Obtained During Water Loading and Fluid Restriction in A Group of Healthy Adults. Maximal Urine Concentration Is Achieved by Plasma Arginine Vasopressin Values of 3 To 4 Pmol/L. *Ld,* Limit of Detection of the Arginine Vasopressin Assay, 0.3 Pmol/L.

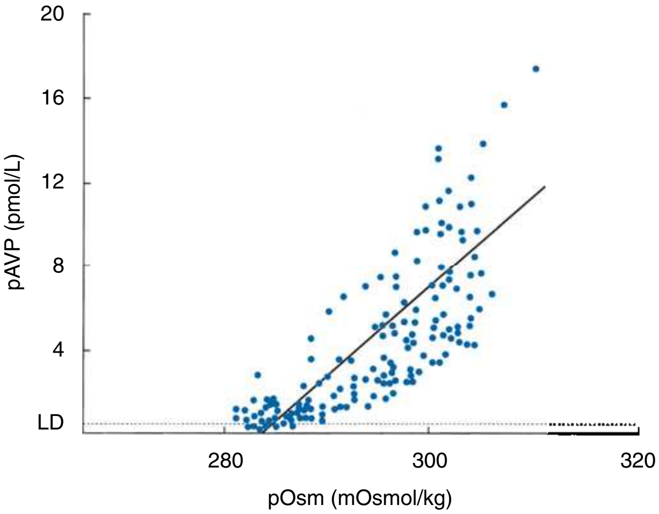


Fig. 12.6 Relationship Between Plasma Osmolality (*pOsm*) and Plasma Arginine Vasopressin (*pAVP*). Increases in pAVP in response to hypertonicity induced by infusion of 855 mmol/L saline in a group of healthy adults. The mean regression line *(dashed line)* is defined by the following equation: pAVP = 0.43 (pOsm 284); *r* = 0.96; *P* < .001. *LD,* Limit of detection of the arginine vasopressin assay, 0.3 pmol/L.

osmolality cause these hypothalamic nuclei to depolarize, sending neural signals via the nucleus medianus to the SON and PVN. An elevation in plasma osmolality causes depolarization of these nuclei, leading to increased synthesis of AVP, and secretion of AVP and copeptin from the posterior pituitary into the circulation. The osmoreceptor cells are solute-specific, in that they respond vigorously to changes in plasma sodium concentration, but less so to variations in blood urea.[7] They respond to osmotic changes induced by the intravenous infusion of mannitol, but are completely insensitive to changes in blood glucose concentrations in experimental situations.[7,8]

There is a linear relationship between plasma osmolality and plasma AVP concentrations throughout the physiological range of plasma tonicity.[9-11] If plasma concentrations are lowered by excessive ingestion of hypotonic fluid to below 280 to 285 mOsm/kg, secretion of AVP is suppressed, and plasma concentrations of the hormone are undetectable on radioimmunoassay measurement. In the absence of AVP-mediated water reabsorption from the kidney, hypotonic polyuria develops.[11] The increase in free water clearance allows plasma osmolality to rise into the normal range. If plasma osmolality rises, owing to dehydration, there is an increase in plasma concentrations of AVP, in proportion to the rise in plasma osmolality. The action of AVP, to concentrate the urine and allow reabsorption of water, restores plasma osmolality to normal. This homeostatic process occurs continuously to maintain plasma osmolality within a narrow reference range. The relationship between plasma AVP concentration and urine osmolality is shown in Fig. 12.5.

The development of sensitive radioimmunoassays has allowed the nature of this relationship to be studied experimentally. Using intravenous infusion of hypertonic sodium chloride solution to increase plasma osmolality in healthy volunteers, plasma AVP can be shown to increase in a linear fashion. Application of linear regression analysis to the data shows that a direct correlation defines the relationship between plasma osmolality and plasma AVP concentration (Fig. 12.6). The qualitative relationship is expressed by the equation pAVP = 0.43 (pOsm - 284), where pAVP indicates plasma AVP concentration and pOsm indicates plasma osmolality.[12]

This mathematical formula describes two important physiologic characteristics of the relationship between plasma osmolality and plasma AVP that are clinically relevant. The first of these is the osmotic threshold for AVP release, which is defined by the abscissal intercept of the regression line. At a mean plasma osmolality of 284 mOsm/kg, plasma AVP is secreted, with increments in secretion as plasma osmolality rises. This represents the “set point,” or the osmotic threshold for AVP release, which is identical in humans to the osmotic threshold for the onset of thirst.[12] This is an important physiological concept, as hypotonic polyuria develops when plasma osmolality is suppressed by water ingestion to concentrations below this set point. AVP detectable in the plasma by radioimmunoassay at plasma osmolalities below this is therefore, by definition, “inappropriately” elevated. Measurements of plasma AVP with ultrasensitive cytochemical methods at plasma osmolalities below the physiological threshold have led to suggestions that secretion cannot be completely suppressed by hypotonicity.[13] However, the very low concentrations demonstrated experimentally seem not to exert a significant biological effect, as they are unable to prevent hypotonic polyuria, and they could simply represent incomplete clearance of presecreted AVP.

The slope of the regression line, which represents the change in plasma AVP concentration per unit change in plasma osmolality, is the second important characteristic of the formula. This defines the sensitivity of the osmoreceptor-AVP releasing unit. Steeper slopes indicate a larger rise in plasma AVP concentrations after osmotic stimulation. Shallow slopes indicate lower rates of AVP secretion in response to hyperosmolality; a pathological example of this can be seen in partial hypothalamic DI.[14]

The osmotic threshold and sensitivity of AVP release vary significantly between individuals, but repeat testing has shown that there is good reproducibility of these parameters within individuals.[15]

Although physiological control of AVP secretion and thirst is almost entirely osmotic, the switch off of both is nonosmotic and is triggered by the act of drinking. In studies in healthy individuals who have been rendered hyperosmolar by hypertonic saline infusion,[16] drinking is associated with an immediate fall in plasma AVP and thirst, before any changes in plasma osmolality can be measured. The fall in plasma AVP is so rapid that it mimics the half-life of AVP,[16] suggesting that a neuroendocrine reflex stimulated by oropharyngeal distension switches off AVP secretion. This reflex prevents overhydration during periods of drinking in response to osmotic stimuli.

The conventional relationship between plasma osmolality and plasma AVP concentration is altered in a number of physiological and pathophysiological situations. Pregnancy causes a lowering of the threshold for AVP secretion, which is responsible for the fall in basal plasma sodium concentration during pregnancy. The sensitivity of the osmoregulatory line remains unchanged. The mechanism is unclear but may be related to increased plasma concentrations of human chorionic gonadotropin (hCG).

Normal physiological aging is associated with complex changes in osmoregulation. Basal circulating AVP concentrations increase with age, and the AVP response to osmotic stimulation has been shown to be enhanced in comparison to younger subjects. Thirst, as measured by visual analog scales, is attenuated, however, and measured water intake after osmotic stimulation intake is reduced.[17] This makes elderly humans more vulnerable to hypernatremia, and many elderly patients in long-term care institutions exhibit permanent hypernatremia, though cognitive impairment may contribute to reduced thirst in these individuals.[18]

Osmoregulation in diabetes mellitus is normal,[8] even in the presence of hyperglycemia,[19] with preserved solute specificity of both AVP release and thirst; hyperglycemia stimulates neither thirst nor AVP secretion. However, chronic hyperglycemia renders the renal tubules resistant to the antidiuretic effects of AVP, due to failure of recruitment of aquaporin-2.[20] Improved glycemic control reverses the state of partial nephrogenic DI.[20] Survivors of hyperglycemic hyperosmolar state show typical osmoregulatory changes of aging, however, with exaggerated AVP secretion and attenuated thirst in response to dehydration.[21]

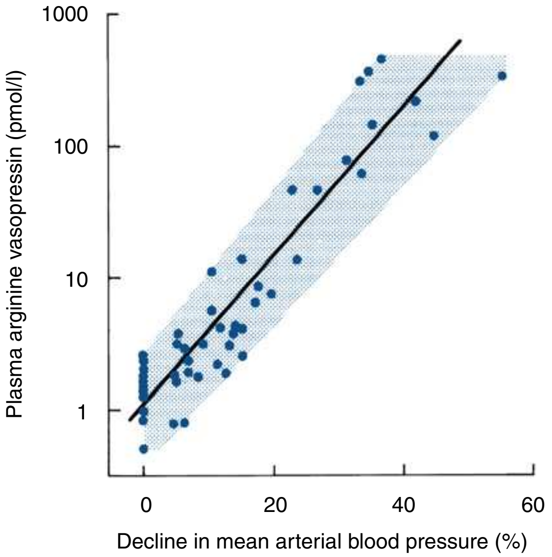


Fig. 12.7 Relationship Between Plasma Arginine Vasopressin (pAVP) and Percentage Decline in Mean Arterial Blood Pressure (MABP). Arterial Blood Pressure Was Reduced by Infusing Increasing Doses of Trimetaphan in Healthy Men. the Regression Line is Defined by the Following Equation: Log (pAVP) = 0.06 (Mabp 0.67); $r = 0.98$; $P < .001$; $n = 48$.

Baroregulation of Vasopressin Release

Physiological fluctuations in blood pressure have almost no effect on plasma AVP secretion. However, pathophysiological falls in both blood volume and pressure can stimulate very marked AVP secretion. Although minor drops in arterial blood pressure (on the order of 10%) slightly increase plasma AVP concentrations (Fig. 12.7), larger falls of 20% to 30% are highly potent stimuli to AVP secretion, causing marked elevations in plasma AVP concentrations. This makes teleological sense, as minor falls in blood pressure are common, and if AVP were to be stimulated by these small falls, humans would be in a permanent state of antidiuresis. The relationship between blood pressure and plasma AVP is shown in Fig. 12.7.

The baroreceptors are situated in the cardiac atria, the carotid sinus, and the aortic arch. They exert a tonic inhibitory effect on AVP secretion in normal circumstances, but when they are unloaded during hypotension or hypovolemia they can stimulate very high plasma concentrations of AVP via pathways quite separate from the osmoreceptors. The relationship between blood pressure and AVP is exponential rather than linear, and can be modified by neurohumoral influences such as atrial natriuretic peptides, which inhibit AVP responses, and norepinephrine (noradrenaline), which augments the baroregulatory AVP response. Baroregulatory inputs also modify osmotic AVP secretion, in that hypovolemia augments the AVP response to hypernatremia.

KEY POINTS

- Elevation in plasma osmolality is detected by osmoreceptor cells in the hypothalamus, leading to increased synthesis of arginine vasopressin (AVP) and secretion of AVP and copeptin from the posterior pituitary into the circulation. This homeostatic process occurs continuously to maintain plasma osmolality within a narrow reference range.

Other Mechanisms Regulating Vasopressin Release

In addition to osmotic and baroregulatory stimuli, AVP secretion occurs in response to nausea, manipulation of abdominal contents at surgery, and hypoglycemia. More recently, arginine infusion has been shown to stimulate the secretion of copeptin and has been suggested to be a useful test in the differential diagnosis of DI.[22]

Secretion of Copeptin

The prepro-AVP precursor is composed of AVP, NPII, and copeptin (Fig. 12.3). During axonal transport from the SON and PVN the posterior pituitary, the precursor hormone is enzymatically cleaved into the three constituent components. Copeptin, which is a larger peptide than AVP, comprising 39 amino acids,[23] is stored in secretory granules in the posterior pituitary and is secreted with AVP in response to osmotic and nonosmotic stimuli.[24] Copeptin has no known biological effects, but as it is secreted in equimolar proportions with AVP, it is a biological marker for AVP release. The half-life for copeptin is much longer than that for AVP, so it is not cleared as quickly from plasma during water loading.

KEY POINTS

- Copeptin has no known biological effects, but, as it is secreted in equimolar proportions with arginine vasopressin (AVP), it is a biological marker for AVP release.

DIABETES INSIPIDUS

Diabetes insipidus is a clinical syndrome characterized by hypotonic polyuria, due to failure to concentrate urine. There are three major causes of hypotonic polyuria, and differentiation between the three disease processes is clinically important, as the diagnosis dictates the treatment recommended to the patient.[25]

1. Central or hypothalamic DI (HDI). Polyuria is caused by insufficient AVP secretion, so there is no effective antidiuretic hormone. There needs to be a loss of at least 80% of the magnocellular neurons, which synthesize and secrete AVP, before polyuria develops.
2. Nephrogenic DI (NDI), in which AVP secretion is normal, but there is renal insensitivity to the antidiuretic effects of the hormone.
3. Primary polydipsia, where excessive fluid intake reflects abnormal perception of thirst.

A full list of the causes of DI is shown in Table 12.1.

Hypothalamic Diabetes Insipidus

HDI (also known as neurogenic, central, or cranial DI) occurs where AVP secretion is impaired such that the patient presents with a hypotonic diuresis and symptoms of polyuria and nocturia. The causative lesion for HDI spares the thirst mechanism in the vast majority of cases[26]; because thirst appreciation is intact, therefore, patients respond to the polyuria with appropriate drinking. The resultant increase in fluid intake nearly always maintains normonatremia; plasma sodium is so rarely elevated at presentation of HDI that the presence of hypernatremia, in circumstances where fluid is freely available, should always raise the suspicion of associated hypodipsia. AVP deficiency may be complete, with no detectable plasma AVP, or partial, where AVP is detectable in the plasma at low concentrations inappropriate to the ambient plasma osmolality.

HDI is rare in the general population, with an estimated prevalence of 1:25,000, but is common in the aftermath of pituitary surgery. It may occur transiently or permanently following other forms of neurosurgical intervention.

TABLE 12.1 **Causes of Diabetes Insipidus**

Hypothalamic Diabetes Insipidus	
Congenital	Hereditary (X-linked or autosomal dominant)
	DIDMOAD
Acquired	Pituitary surgery
	Tumors (craniopharyngioma, germinoma, pinealoma metastases)
	Traumatic brain injury
	Granuloma (tuberculosis, sarcoid, histiocytosis X)
	Infections (encephalitis, meningitis)
	Vascular disorders (Sheehan syndrome, aneurysms, subarachnoid hemorrhage, gastrointestinal bleed)
	Hypophysitis (A/immune, lymphocytic, immune checkpoint inhibitors)
	Idiopathic
	Pregnancy
Nephrogenic DI	
Congenital	Hereditary (X-linked recessive or autosomal dominant)
Acquired	Chronic renal disease (polycystic kidneys, obstructive uropathy)
	Metabolic disease (hypercalcemia, hypokalemia)
	Drugs (lithium, demeclocycline)
	Osmotic diuresis (glucose, mannitol)
	Amyloidosis
	Myelomatosis
Primary Polydipsia	
Compulsive water drinking	
Affective disorders, e.g., schizophrenia	
Craniopharyngiona/Rathke's cleft cyst	
Sarcoidosis	

The majority of cases of HDI are acquired during adulthood, and the commonest cause of DI in most endocrine practices is surgery for pituitary tumors. Pituitary adenomas rarely cause DI without surgical intervention, and where a pituitary mass is associated with polyuria and polydipsia, craniopharyngioma, granuloma, or hypophysitis are far more likely diagnoses. Table 12.2 shows the distribution of causes of permanent HDI in our own large series of patients. The predominance of neurosurgical causes reflects our work in conjunction with the National Neurosurgery Centre.

DI occurs following pituitary surgery in 18% to 30 % of cases,[27-30] usually within 2 days of surgery. The most common natural history is spontaneous resolution over the next 2 to 5 days. The pathophysiology of acute transient DI is thought to be surgical contusion injury to the magnocellular neurons projecting from the SON and PVN to the posterior pituitary. In some patients, however, DI persists, and AVP deficiency becomes permanent. Most neurosurgical series report a low incidence of 1% to 8% for permanent DI following transsphenoidal surgery.[27,29-31] Many published studies are retrospective, however, and have used polyuria alone as the criteria for diagnosis of DI. We found that formal screening for DI using water deprivation testing or hypertonic saline infusion identifies much higher rates of permanent DI, occurring, for instance, in 30% of patients who had transcranial surgery for pituitary adenoma and much higher figures of over 90% following craniopharyngioma surgery.[28] Many of these patients had partial DI with minor symptoms, but nearly all patients benefitted from therapeutic intervention.

TABLE 12.2 Causes of Permanent Hypothalamic Diabetes Insipidus in 231 Cases on the Beaumont Hospital Pituitary Register

Cause of Diabetes Insipidus		*n*	%
Pituitary tumors		167	73
	Surgery of pituitary adenoma	108	47
	Craniopharyngioma	52	23
	Pituitary apoplexy	7	3
Other tumors		13	6
	Germinoma	8	4
	Pinealoma	2	1
	Pituitary/stalk metastases	2	1
	Pituitary stalk meningioma	1	0.5
Idiopathic/autoimmune		19	9
Traumatic brain injury		7	3
Subarachnoid hemorrhage		4	2
Lymphocytic hypophysitis		10	5
Langerhans giant cell histiocytosis		2	1
Sarcoidosis		4	2
Postencephalitis		1	0.5
Postmeningitis		1	0.5
Tuberculoma of pituitary gland		1	0.5
Inherited		2	1

(From Behan LA, Sherlock M, Moyles P, et al. Abnormal plasma sodium concentrations in patients treated with desmopressin for cranial diabetes insipidus: results of a long-term retrospective study. *Eur J Endocrinol.* 2015;172:243–250.)

In a small minority of patients, posthypophysectomy DI follows the triple-phase response. Initially, classical postoperative DI presents in the first 2 days following surgery. However, after 4 to 8 days of hypotonic polyuria, urine output drops, and plasma sodium begins to fall, even after discontinuation of desmopressin therapy. This second, antidiuretic phase is characterized by a rise in plasma AVP concentrations, presumably from uncontrolled release of prestored peptide from damaged magnocellular neurons. If hypotonic intravenous fluids continue to be administered, the fall in plasma sodium concentration can be abrupt. The antidiuretic phase of the triple-phase response usually lasts 5 to 7 days, but there is considerable variation, from 2 to 14 days, before progression to the final phase, the development of permanent DI. It is thought permanent DI occurs when gliosis of damaged neurons prevents further AVP release.

The major factor influencing the risk of developing DI following pituitary surgery is tumor type. Craniopharyngioma is frequently associated with preoperative DI, but in addition, surgical intervention is far more commonly associated with the development of DI than surgery for any other intracranial tumor.[28,32] Surgery for ACTH-secreting pituitary adenomas has been associated with a higher risk of developing DI in a number of series.[27,29,31] The data from the literature are conflicting on the influence of tumor size on the risk of postoperative DI. One review found that a postoperative plasma sodium concentration greater than 145 mmol/L had 98% specificity in predicting the presence of DI.[32] Metastatic deposits in the hypothalamus or pituitary stalk usually have no endocrine effects, but some, particularly from carcinoma of the breast or bronchus, have been reported to produce HDI. Others tumors, such as germinoma, pinealoma, and parasellar meningiomas, can also cause HDI. Interestingly, although radiotherapy for nonpituitary cranial tumors causes anterior pituitary dysfunction commonly, it does not seem to cause DI.[33]

Lymphocytic infiltration of the neurohypophysis (infundibuloneurohypophysitis), recognized by a thickened pituitary stalk and inflammatory infiltrates of T lymphocytes and plasma cells with eosinophils, is a well-described cause of HDI, occurring in 30% of adult cases.[34]

Idiopathic DI is diagnosed less often than previously; better imaging techniques have revealed structural lesions more frequently, and there is greater awareness of autoimmune disease as a cause of DI. It has been estimated that one-third of patients previously diagnosed to have idiopathic HDI may an autoimmune etiology. The characteristics of autoimmune DI include the presence of circulating antibodies against AVP-secreting cells, young age of onset, and thickened pituitary stalk on T1-weighted magnetic resonance imaging (MRI).[35] Our own data show that patients with autoimmune DI have evidence of other organ-specific autoimmune endocrine disease, particularly thyroid disease, in one third of cases.[36]

Traumatic brain injury causes acute DI in 20% of cases,[37] some of which develop a triple-phase response similar to that seen after pituitary surgery. DI following brain injury is nearly always transient, however, with the vast majority recovering on retesting at 6 months.[38] Persistent DI in the acute phase after brain injury is a powerful predictor of acute mortality.[39] Only 7% of long-term survivors of head injury have permanent DI when formally tested with water deprivation.[40]

Sheehan syndrome is now rare in developed countries as a result of improvements in obstetric care. It is not usually associated with HDI, even when there is widespread anterior hypopituitarism, but maximal urine-concentrating ability is impaired in some patients.[41] HDI can occasionally occur in normal pregnancy, termed gestational DI, owing to increased activity of circulating vasopressinase, the placental aminopeptidase.[42] Symptoms resolve postpartum. The actions of vasopressinase can unmask partial HDI in patients with pituitary disease or can worsen symptoms in undiagnosed partial HDI. As a result, patients with pituitary disease need to be monitored during pregnancy for change in requirements for desmopressin. HDI in pregnancy must be differentiated from the transient NDI that is seen very occasionally.

Familial HDI is extremely rare. Autosomal dominant HDI is associated with loss-of-function mutations in the *AVP* gene, resulting in the production of a folding-incompetent peptide precursor that accumulates in the secretory pathway apparatus of magnocellular neurons.[43,44] This mutant precursor is responsible for an autophagocytic process, causing progressive damage and loss of vasopressinergic neurons. A single mutant allele is sufficient for this process to occur. Familial HDI almost always presents in childhood.[45]

Wolfram, or DIDMOAD, syndrome is a rare autosomal recessive, progressive neurodegenerative disorder characterized by the association of HDI with diabetes mellitus, optic atrophy, and bilateral sensorineural deafness, though other manifestations such as gonadal failure, renal outflow tract dilatation (secondary to reduced nerve fibers in the bladder wall), and progressive ataxia with brain stem atrophy may also occur.[46] The syndrome is associated with premature death, usually owing to ascending renal tract infection secondary to atonic bladder and hydronephrosis. Although DI is often reported to occur in only one third of cases, in one series where careful testing with AVP responses to hypertonic saline was performed, subnormal responses were seen in all patients.[47] Wolfram syndrome is associated with loss-of-function mutations in the *WFSI* gene on chromosome 4p16.1 that encodes an 890–amino acid glycoprotein (wolframin) found in the endoplasmic reticulum. An additional locus has been identified recently at chromosome 4q22-24, suggesting that the syndrome may be genetically heterogeneous.[47]

Nephrogenic Diabetes Insipidus

NDI occurs when renal resistance to the antidiuretic effects of AVP allows hypotonic diuresis to develop. An outline of potential causes is given in Table 12.1. Lithium therapy is the commonest cause of NDI in clinical practice: approximately 30% of patients treated with long-term lithium develop NDI. Although polyuria usually disappears with cessation of treatment, a minority of patients develop interstitial nephritis, and permanent NDI occurs. For reasons that remain unexplained, patients with lithium-induced NDI seem more vulnerable to the development of hypernatremia than patients with untreated HDI.[48] This suggests that lithium-induced NDI has coexistent impairment of thirst, with inadequate fluid intake. The mechanism for this has not been adequately defined experimentally.

Other metabolic causes of NDI include hypokalemia and hypercalcemia; NDI is usually reversible with correction of the electrolyte disturbance. Poorly-controlled diabetes mellitus causes renal resistance to AVP, which seems to be owing to failure to generate aquaporin-2 production in response to receptor binding.[8,19,20] The reversible nature of the metabolic causes of NDI illustrates how AVP-stimulated generation of aquaporins is vulnerable to metabolic and pharmacological insults.

Familial NDI is rare. X-linked-recessive familial NDI is caused by inherited loss-of-function mutations in the gene encoding the AVP V2 receptor. More than 70 different mutations have been described, affecting all aspects of receptor physiology, including expression, ligand binding, and G protein coupling.[49] The vast majority cause complete loss of function and present in infancy, though some mutations are associated with partial loss of function.[50] A minority of FNDI families have autosomal recessive NDI, with loss-of-function mutations of the gene encoding AQP2. The majority of mutations occur in the region of the gene encoding the transmembrane domain of the water-channel protein. Other kindreds have been described with an autosomal dominant NDI owing to loss-of-function mutations in AQP2 affecting the C-terminal intracellular domain of the protein. NDI is expressed in these kindreds, because the protein product of the mutant allele forms mixed oligomers with the product of the wild-type allele, resulting in sequestration within the Golgi or mistargeting to the basolateral rather than the apical membrane in a dominant negative manner.[51,52]

Primary Polydipsia

Primary polydipsia occurs when excessive fluid intake lowers the plasma osmolality to levels below the osmotic threshold for AVP secretion.[53] In the absence of circulating AVP, hypotonic polyuria develops. Many patients seem to have no underlying illness, but careful osmoregulatory studies have identified a number of abnormalities of thirst regulation. Patients typically have a low osmotic threshold for thirst that compels them to drink, even when plasma osmolality is lowered below the usual physiological thirst threshold. In addition, they have an exaggerated thirst response to osmotic challenge and an inability to suppress thirst in response to drinking, so the volume of fluid drank is in excess of that required for normal hydration.[54] Some forms of primary polydipsia are associated with psychiatric disorders, and up to 20% of patients with chronic schizophrenia have polydipsia. If these patients are simultaneously prescribed drugs that can cause syndrome of inappropriate antidiuresis (SIAD), hyponatremia is inevitable and occasionally severe. Rarely, the problem may be associated with irritative structural abnormalities in the hypothalamus or posterior pituitary, such as hypothalamic sarcoidosis or craniopharyngioma,[55] but usually brain imaging is normal.

KEY POINTS

- Polyuria can result from abnormalities in any of three fundamental components of the physiological control of water balance:
  - Subnormal arginine vasopressin (AVP) secretion, or central or hypothalamic diabetes insipidus (DI), which reflects hypothalamic or posterior pituitary damage.
  - Impaired renal concentrating ability in response to AVP (nephrogenic DI).
  - Excessive thirst appreciation (primary polydipsia).

The differentiation between these causes of polydipsia is clinically important, as the diagnosis dictates the treatment recommended to the patient.

Investigation of Diabetes Insipidus

Urine flow rates in excess of 40 mL/kg per 24 hours in adults or more than 100 mL/kg per 24 hours in infants are suggestive of DI. A careful clinical history is rarely diagnostic, but often gives valuable information in the investigation of a polyuric patient. Daytime polyuria without nocturia is a suggestive of primary polydipsia, as patients with either form of DI are unable to concentrate urine overnight, with consequent nocturia. Frequent passage of small volumes of urine may suggest prostatism, urinary tract infection, or bladder instability, rather than osmoregulatory dysfunction. A history of lithium or diuretic therapy is relevant. The presence of a mass in the pituitary fossa raises a strong diagnostic likelihood of a craniopharyngioma or germinoma; equally, the presence of DI in a patient with a sellar mass indicates that the mass is not a pituitary "adenoma." A strong family history of autoimmune thyroid disease suggests an autoimmune etiology for the DI.

Initial investigations should aim to confirm that the patient is polyuric; 15% of patients who are referred to our service for investigation of polyuria have normal urine volumes, with frequency of micturition due to urinary tract disease. A 24-hour urine volume of less than 3 L in an adult indicates that osmotic studies are unnecessary. Simple ambulatory investigations include electrolytes, calcium, midstream specimen of urine, and urinalysis; if the patient does not have glycosuria, polyuria is not due to diabetes mellitus. Clinic plasma and urine osmolality are rarely diagnostic unless the urine osmolality exceeds 700 mOsm/kg, which implies normal AVP secretion and action and therefore excludes DI. In this circumstance, the diagnosis is nearly always polydipsia, and formal osmoregulatory investigations are unnecessary. Urine osmolality is characteristically low, irrespective of the cause of polyuria. However, an elevated plasma osmolality is more likely with HDI or NDI, whereas a plasma osmolality at the low end of the reference range is suggestive of primary polydipsia.

In most patients it is necessary to proceed to a water deprivation test.[25] This is a two-step test of osmoregulatory function. The first step is an 8-hour period of dehydration. The normal physiological response is to respond to the rise in plasma osmolality with secretion of AVP, and concentration of urine to over 700 mOsm/kg. Theoretically, patients with primary polydipsia have intact osmoregulation and should therefore respond to dehydration with appropriate urine concentration. In the simplest use of this test therefore, the key measurement is the final urine osmolality, and the result answers the key question "does the patient have true DI, or not?". In patients undergoing water deprivation following hypophysectomy, this is the only step needed to make or refute the diagnosis of HDI. In clinical scenarios with a wider differential diagnosis, the second step of the test is needed.

The second step documents the renal response to desmopressin. Urine osmolality is measured at intervals after intramuscular desmopressin; in patients with HDI, there should be a normal concentrating response, with a rise in urine osmolality to greater than 700 mOsm/

kg. In contrast, patients with NDI will fail to respond to desmopressin, and urine will remain dilute. The rationale for step 2 is therefore to distinguish between HDI and NDI.

In practice, the water deprivation test does not always produce straightforward results. Surreptitious water intake can give spurious results, so the test should be undertaken with strict supervision; the most accurate results are derived from units with experience in performing the test. Partial forms of HDI and NDI are difficult sometimes to differentiate from primary polydipsia.[56] Furthermore, prolonged polyuria from any cause leads to partial resistance to the antidiuretic action of AVP.[57] This can mean that patients with severe polyuria due to primary polydipsia may appear to have partial NDI. The traditional explanation for this is that prolonged polyuria dilutes the high solute in the renal medullary interstitium, upon which water reabsorption depends. A far more likely explanation is that chronic AVP deficiency, from either HDI or long-term suppression from overdrinking, leads to failure to generate mRNA for luminal cell aquaporin-2, so that binding of desmopressin to the V2 receptor does not immediately lead to optimum water reabsorption from the urine.

The measurement of AVP, and more recently copeptin, during hypertonic (3%–5%) saline infusion has been shown to discriminate between partial cranial DI and primary polydipsia with greater diagnostic accuracy than the water deprivation test.[58] Hypertonic saline infusion leads to maximal stimulation of AVP secretion. Fig. 12.8 indicates the relationship of plasma osmolality to plasma AVP during hypertonic saline infusion in patients with varying causes of polyuria. When plasma AVP concentrations are measured with sensitive radioimmunoassays, they are a very useful diagnostic tool in this context. However, commercially available radioimmunoassays are almost uniformly unfit for purpose. Many have limits of detection that are higher than the physiological reference range for plasma AVP concentration. Access to highly sensitive and specific radioimmunoassays is limited to a few specialized centers. In addition, measurement of AVP in plasma needs an extraction step and a prolonged, two-step, double-antibody technique.[59] The assay is therefore difficult to access and difficult to perform and does not give a quick turnover of results.

The measurement of plasma copeptin offers an accessible, practical surrogate for the measurement of plasma AVP. Published data show that there is a close correlation between the two measurements.[60] Copeptin only requires small plasma volumes and uncomplicated sample handling and is easy to measure in the laboratory. It is likely to replace the measurement of plasma AVP in the approach to differential diagnosis of polyuric states, and indeed has done so in many laboratories. The plasma half-life of copeptin is longer than that of AVP, so suppression of copeptin is not as readily achieved; complex osmoregulatory disorders that require a suppression test to define the osmotic threshold for AVP may still therefore depend on the more technically difficult measurement of plasma AVP.

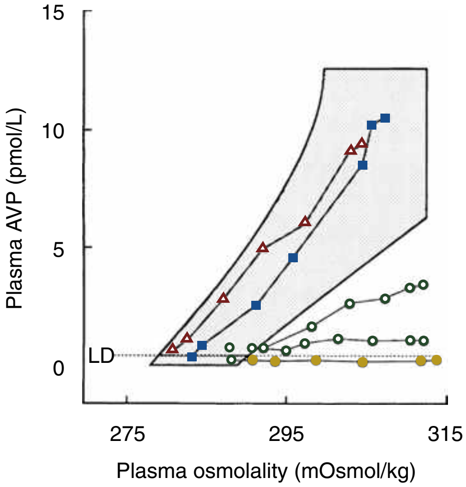


Fig. 12.8 Dynamic Tests of the Arginine Vasopressin (*AVP*) Axis in Patients with Polyuria. Plasma AVP (pAVP) and Plasma Osmolality (pOsm) Responses to Hypertonic (855 mmol/L) Saline Infusion. *Shaded Area,* Range of the Normal Response; *Red Triangles,* Normal Patients; *Blue Squares,* Patients with Nephrogenic Diabetes Insipidus; *Green and Yellow Circles,* Patients with Hypothalamic Diabetes Insipidus.

A seminal prospective multicenter study has shown that the diagnostic accuracy of hypertonic-stimulated copeptin measurement exceeds that of the traditional water deprivation test, with or without copeptin measurements. A stimulated copeptin cutoff of 4.9 pmol/L after hypertonic saline infusion had a 93% sensitivity and a 100% specificity to distinguish between primary polydipsia and HDI.[58] An elevated baseline copeptin concentration greater than 21.4pmol/L has been shown by the same group to distinguish NDI from other causes of polyuria, with sensitivity and specificity of 100%, mitigating the need for further dynamic testing in this circumstance.[61] More recent data suggest that intravenous infusion of arginine, to stimulate copeptin, can differentiate between polyuric states,[22] but there have been no studies reporting head-to-head comparisons with classical osmotic stimulation tests.

The use of a visual analog scale for thirst is helpful because patients with thirst excess show high baseline thirst readings, with exaggerated rises during the test; the most useful diagnostic parameter for primary polydipsia is the failure to suppress thirst ratings during water intake at the conclusion of the infusion.[26] On the other hand, the demonstration of absent osmoregulated thirst during osmotic stimulation is the gold standard for the diagnosis of adipsic DI (ADI).[26]

Once the diagnosis of HDI is confirmed, imaging of the hypothalamopituitary region with MRI scanning is mandatory to exclude structural lesions. Sellar tumors in patients with HDI are most often craniopharyngiomas. Identification of a stalk mass has a wider differential diagnosis that includes craniopharyngioma, germinoma, metastasis, lymphocytic hypophysitis, and stalk thickening in autoimmune HDI. A recent review of 152 patients with pituitary stalk lesions showed that 28% had DI; neoplastic (40%) and inflammatory (20%) conditions were the commonest etiologies.[62] Typical appearances may also be found in sarcoid, haemochromatosis, and other inflammatory conditions. The healthy posterior pituitary has as a characteristic bright signal on T1-weighted MRI, but this bright spot is absent in the majority of patients with HDI.[63] The intensity of the posterior lobe signal is also diminished in the elderly, in patients with anorexia nervosa, septic shock, or poorly controlled diabetes mellitus, and in hemodialysis patients. It is thought that this may represent AVP hypersecretion, and thus decreased VP-NPII complex content.

In idiopathic DI, consideration should be given to an autoimmune etiology. Measurement of thyroid function, vitamin B12, and autoantibodies is worthwhile routinely, whereas tests for primary adrenal failure should be considered where clinically indicated. Serum angiotensin-converting enzyme concentration and chest films should be considered if sarcoid is suspected. If initial hypothalamopituitary imaging is normal, it is important to repeat MRI imaging within 6 months, as clinical DI may precede the radiological manifestations of some parasellar tumors, such as germinomas.[64] It has been suggested that in younger patients, in whom germinomas are most likely to be present, repeated radiology should be performed every 3 months if initial scans are normal.[65] If stalk lesions are present, elevated serum (or cerebrospinal fluid) hCG or alpha-fetoprotein measurements may indicate the diagnosis of germinoma, and early repeat MRI to assess tumor growth is important. Biopsy should be considered early, as germinomas are best treated with chemotherapy and radiotherapy rather than surgical intervention. Dynamic tests of anterior pituitary function should be considered where structural pituitary lesions exist, though we have found that adult patients with idiopathic/autoimmune DI do not seem to have coexistent anterior pituitary hormone deficits.[36] In contrast, data in children suggest that growth hormone deficiency commonly accompanies autoimmune HDI.[66,67]

Patients with NDI should have renal investigations as appropriate, including urine microscopy, renal tract ultrasound to exclude obstruction, and tests of other aspects of tubular function. Genetic studies in suspected familial disease remain research tools.

KEY POINTS

- The first steps in the diagnostic workup of diabetes insipidus (DI) are to confirm the presence of hypotonic polyuria and to exclude other causes of polyuria, such as diabetes mellitus and hypercalcemia. In most cases of suspected DI, dynamic testing of the vasopressin-renal axis is required, of which both indirect (water deprivation test) and direct (hypertonic saline infusion with copeptin measurement) methods are possible, with the latter test showing superior diagnostic accuracy in differentiating between partial hypothalamic DI and primary polydipsia.

Adipsic Diabetes Insipidus

ADI is a rare disorder characterized by AVP deficiency that is complicated by failure to generate thirst sensation in response to hypernatremia. It is most commonly seen after clipping of anterior communicating artery aneurysms following subarachnoid hemorrhage.[68,69] The vascular supply to the osmoreceptors is derived from small arteries arising from the anterior communicating artery. It is assumed that these vessels are damaged during aneurysm clipping, with infarction of the circumventricular organs where the osmoreceptors are sited. The osmoregulatory defect is almost always permanent, although there are occasional reports of recovery.[70] Extensive hypothalamic surgery for craniopharyngioma,[69,71] and rarely pituitary macroadenoma,[71] have also been reported to cause ADI.[69]

In ADI secondary to clipping of the anterior communicating artery, there is a pure osmoreceptor defect. Osmotic stimulation with hypertonic saline causes no rise in plasma AVP or in thirst ratings, despite marked hyperosmolality. In contrast, nonosmotic baroregulated AVP secretion is completely normal, with the secretion of very high plasma AVP concentrations, often 100-fold higher than those needed for maximal urine concentration.[72] The AVP synthesis/secretion unit of the SON/PVN/posterior pituitary is thus intact, but not innervated by the osmoreceptors. In patients who have destructive surgery to the pituitary, all secretory function is lost.[72]

The diagnosis of ADI presents significant management challenges.[73] In the absence of thirst, patients are unable to perceive changes in plasma tonicity and are thus vulnerable to marked swings in plasma sodium concentration. Severe hypernatremia is a particular hazard, particularly during intercurrent illness. Management requires regular administration of desmopressin, fixed fluid intake, which may vary with climatic conditions, and regular review for measurement of plasma sodium. Hypothalamic abnormalities are often seen with ADI (Table 12.3), including hypothalamic obesity, seizure disorders, sleep apnea, and thermoregulatory disorders.[69] Episodes of dehydration are often complicated by an elevated hematocrit, with development of thrombosis, including pulmonary thromboembolism[69,74] and rhabdomyolysis.[75] Many patients die prematurely due to postsurgical complications, electrolyte abnormalities, or sleep apnea.[69,76]

Treatment of Diabetes Insipidus

Even in the absence of pharmacotherapy, the intact thirst mechanism in DI is sufficient to prevent dehydration, but the need to continually seek fluids and facilities for polyuria is a great inconvenience. The treatment of choice is desmopressin, a synthetic, long-acting AVP analog that possesses minimal pressor activity and has twice the antidiuretic potency of AVP. It can be administered as an intranasal spray, but more often is given orally in two to three daily doses. The main complication is dilutional hyponatremia,[77] which can produce

TABLE 12.3 **Comorbidities Associated With Adipsic Diabetes Insipidus**

Pt	Diagnosis	Baroreg AVP release	Sleep apnea	DVT	BMI kg/m^2	Other morbidity	Death
A	ACAA	Yes	No	No	24.8	Nil	
B	ACAA	Yes	No	No	24.2	Hemiparesis	
C	ACAA	Yes	No	No	27.6	Seizures, hypothermia	
D	ACAA	Yes	Yes	Yes	37.2	Nil	
E	ACAA	Yes	No	Yes	32.9	Nil	
F	TBI	Yes	No	No	25.5	Nil	
G	Toluene	Yes	Yes	No	35.7	Hypothermia, hypothalamic seizures, abnormal thermoregulation	
H	Cranio	No	No	No	56.9	Hypercholesterolemia	
I	Cranio	No	Yes	Yes	33.5	Seizures, hydrocephalus	24 yr
J	Cranio	No	No	No	34.7	Seizures	
K	Cranio	No	Yes	No	28.2	Hypernatremic seizure	
L	Cranio	No	Yes	Yes PTE	41.3	Abnormal thermoregulation	48 yr
M	PRLoma	No	Yes	Yes	53.0	Nil	
N	N/sarcoid	Yes	Yes	No	57.1	DM, seizures	36 yr
O	Congenital	Yes	Yes	Yes PTE	32.4	Behavioral disorder, acute pancreatitis, seizures	18 yr

ACAA, Post clipping of anterior communicating artery aneurysm; *Cranio*, craniopharyngioma (all postsurgery); *TBI,* traumatic brain injury; *DVT,* deep venous thrombosis; *PTE,* pulmonary thromboembolism; *DM,* diabetes mellitus; *PRLoma,* prolactinoma; *N/sarcoid*, neurosarcoidosis; *AVP,* arginine vasopressin; *BMI,* body mass index.

headaches, bloating, and occasionally seizures, and is more common with use of higher dosages and more frequent administration, such as three times daily. Dilutional hyponatremia can be prevented by a number of regimen adaptations:

1. Omit desmopressin once weekly. This allows an aquaresis that prevents water intoxication. This is highly effective in avoiding hyponatremia, but the day of dose omission can be severely disrupted by polyuria. This option is therefore unpopular with patients.
2. Delay the dose of desmopressin once or twice a week. This is also effective in preventing hyponatremia, but as desmopressin is administered as soon as polyuria develops, it is less disruptive of social life. However, as aquaresis is less marked, this usually has to be done twice weekly.
3. Perhaps the most effective, but most intrusive, strategy is to delay each dose of desmopressin until polyuria develops. This method allows little capacity for water retention, but many patients find it socially unacceptable.

Partial HDI (urine volume <4 L per 24 hours) can be managed with adequate fluids to quench thirst, with a nocturnal dose of desmopressin to prevent loss of sleep because of nocturia. Plasma electrolytes should be measured at regular intervals after the introduction of treatment to look for hyponatremia; a decrease in plasma sodium concentration should prompt a therapy review and introduction of one of the strategies documented above. In contrast to hyponatremia, outpatient hypernatremia almost never occurs in DI, where thirst is intact. However, hypernatremia is common during acute illness, particularly where vomiting or diarrhea increase nonrenal water losses.[77]

NDI due to an acquired metabolic problem is best managed by addressing the underlying electrolyte abnormality and maintaining adequate hydration while function recovers. For patients with congenital NDI or those with an acquired defect that is irreversible, a number of additional measures, including thiazide diuretics (hydrochlorothiazide, 25 mg/24 hours); prostaglandin inhibitors such as nonsteroidal antiinflammatory drugs (ibuprofen, 200 mg/24 hours), and dietary salt restriction can be used. All probably work through a combination of reducing glomerular filtration rate and interference with the diluting capacity of the distal nephron, and all are partially effective at best. Data in rats suggest that there may be therapeutic potential in the use of sildenafil for NDI, but there are no data in humans.[78] Fluid replacement remains the cornerstone of treatment of NDI.

As with other causes of DI, the approach to primary polydipsia is to address the underlying cause. This may be difficult, as the urge to drink in primary polydipsia may be powerful and distressing. Hyponatremia is unusual unless there is coprescription of a medicine that stimulates AVP secretion. Switching treatment to alternative agents that do not stimulate AVP may help those patients with chronic schizophrenia and a history of hyponatremia. Individuals with persistent primary polydipsia are at risk of hyponatremia if treated with desmopressin, so accurate diagnosis is important. Reduction in fluid intake is the only rational treatment.

KEY POINTS

- Long-term treatment with desmopressin can control the thirst and polyuria of hypothalamic diabetes insipidus (HDI). Hyponatremia is a relatively common complication of desmopressin treatment in patients with HDI and intact thirst; such patients should be counseled to allow regular periods of free water clearance to prevent this. On the other hand, patients with adipsic diabetes insipidus are at significant risk of hypernatremia, and management of this condition can be particularly challenging.

HYPONATREMIA

Morbidity and Mortality

Hyponatremia is the most common of the electrolyte disturbances seen in clinical practice, occurring in approximately 15% of hospitalized patients.[79] There is considerable evidence that shows that hyponatremia is associated with increased morbidity, mortality, and length of hospital stay.[80,81] Mild hyponatremia has been considered to be asymptomatic, but cognitive decline and gait instability[82] leading to increased falls,[83] bone fractures,[84] and osteoporosis[85] have all been reported. Hyponatremia has consistently been associated with increased mortality in almost every published study,[80,86,87] although some authors have suggested that hyponatremia is only a marker of severity of the underlying disease,[88] rather than the *de facto* cause of excess death. However, treatment of hyponatremia has been associated with a reduction in mortality in a number of studies.[89,90]

Differential Diagnosis of Hyponatremia

The causation of hyponatremia is myriad, and importantly, the presentation, behavior, and mortality vary according to etiology. The most widely used clinical approach is to classify the etiology of hyponatremia according to ECF status into hypo-, eu-, and hypervolemic hyponatremia (Table 12.4). This requires careful use of clinical and biochemical data, and the distinction between euvolemia and mild hypovolemia is particularly challenging. However, accurate categorization of patients according to ECF volume status and urine electrolyte excretion does allow the initiation of etiology-specific treatment. In addition, as there is good evidence that mortality in patients with hypervolemic and hypovolemic hyponatremia is substantially higher than in euvolemic hyponatremia,[91] careful categorization by blood volume allows more accurate prognostic predictions.

Hypovolemic Hyponatremia. Hypovolemic hyponatremia is characterized by ECF depletion due to loss of fluid through the skin, gastrointestinal tract, or kidney. Diminished blood volume causes off-loading of baroreceptors, with powerful stimulation of AVP release. Hyponatremia may develop, particularly in conjunction with

TABLE 12.4 Differential Diagnosis of Hyponatremia

	Clinical features	Urine sodium <20 mmol/L	Urine sodium >40 mmol/L
Hypovolemic	Dry mucus membranes, decreased skin turgor, tachycardia, hypotension (in particular orthostatic), low CVP, raised urea	Dehydration	CSW Diuretics Addison disease Salt-losing nephropathy
Euvolemic	Normal pulse and blood pressure	SIAD with FR	SIAD ACTH insufficiency Hypothyroidism
Hypervolemic	Peripheral edema, ascites, pulmonary edema	Inappropriate IV fluids Cardiac failure Cirrhosis	Renal failure

CSW, Cerebral salt wasting; *CVP*, central venous pressure; *FR*, fluid restriction; *IV*, intravenous; *SIAD,* syndrome of inappropriate antidiuresis; *ACTH*, adrenocorticotropic hormone.

inappropriate administration of hypotonic fluids. Demonstration of low central venous pressure is the gold standard diagnosis, but this is an invasive procedure that is impractical to perform in most cases. A clinical diagnosis of hypovolemia is based on orthostatic hypotension, tachycardia, dry mucous membranes, and decreased skin turgor, together with laboratory clues such as raised urea-to-creatinine ratio and elevated uric acid concentration. Urinary sodium concentration is nearly always low, often less than 10 mmol/L, in hypovolemia, due to activation of the renin-aldosterone system, and mineralocorticoid-induced sodium retention at the kidney. Hypovolemia with elevated urine sodium concentrations indicates that the site of volume/sodium loss is the kidneys, or that the patient is taking diuretics, particularly thiazides.

Primary adrenocortical failure is an important cause of hypovolemic hyponatremia. Combined glucocorticoid and mineralocorticoid deficiency leads to renal salt and water loss; hypotension causes baroreceptor-stimulated AVP release, which causes water retention, further lowering plasma sodium concentration. The biochemical picture differs from that of central adrenal insufficiency, where mineralocorticoid secretion is retained; severe hypovolemia and hyperkalemia are not seen.[92]

Hypervolemic Hyponatremia. Hypervolemic hyponatremia is characterized by peripheral or sacral edema, ascites, and elevated central and jugular venous pressures, all clinical parameters of fluid overload. Arterial underfilling leads to a baroreceptor mediated increase in AVP secretion and antidiuresis, activation of the renin-angiotensin system and secondary hyperaldosteronism, and increased sympathetic tone. Secondary hyperaldosteronism leads to sodium retention but the relative water excess, resulting from increased plasma AVP concentrations and upregulation of aquaporin-2 water channels, causes the development of hyponatremia.

Euvolemic Hyponatremia. Patients without clinical evidence of volume expansion or depletion should be considered euvolemic. Spot urinary sodium concentration is greater than 20 to 30 mmol/L in most euvolemic patients, though dietary sodium restriction can lead to misleadingly low urinary sodium levels. Other supporting laboratory parameters in euvolemia include low blood urea and low serum uric acid. The majority of cases of euvolemic hyponatremia are due to SIAD, and it is this variant of hyponatremia that is most often referred to endocrine services.

THE SYNDROME OF INAPPROPRIATE ANTIDIURESIS

Introduction

SIAD is the commonest cause of hyponatremia in clinical practice. The syndrome was first described in 1956 by Frederic Bartter and William Schwartz, in two patients with lung carcinoma who had severe hyponatremia at presentation.[93] In an elegant series of experiments in which fluid restriction was alternated with fluid loading, the investigators demonstrated that the patients were in a state of antidiuresis. They hypothesized that their patients manifested a clinical syndrome of inappropriate antidiuresis secondary to excess circulating antidiuretic hormone, well before the hormone itself was characterized. Subsequent development of sensitive radioimmunoassays allowed the measurement of plasma AVP, and their classic clinical description and pioneering pathophysiological hypothesis were substantiated by published reports of elevated plasma AVP concentrations in patients with the syndrome.[94] SIAD is now recognized to occur in wide spectrum of disease states; however, differentiation of SIAD from other causes of hyponatremia remains important to the prescription of appropriate treatment regimens.

Causes of Syndrome of Inappropriate Antidiuresis

SIAD is associated with a great number of illnesses and most often presents as a coincidental biochemical manifestation of the causative disease. The most common causes of SIAD are summarized in Table 12.5. The etiology of SIAD is divided into four main clinical categories: malignancy, pulmonary, pharmacological, and neurological causes. Bartter and Schwartz first described SIAD in association with bronchogenic lung carcinoma. In small cell carcinoma of the lung, SIAD is relatively common and has been associated with poor outcome. Most neoplasms have now been reported to cause SIAD, and the association between malignant disease and SIAD is so strong that a patient presenting with SIAD, general malaise, and unexplained weight loss should be considered to have an underlying malignancy until proven otherwise.

SIAD is also commonly associated with intracranial diseases, particularly traumatic brain injury (TBI),[37,38,40] where almost all cases resolve spontaneously with recovery from brain injury. Over 50% of patients with subarachnoid hemorrhage (SAH) develop hyponatremia in the first week following the bleed, and 80% of these are due to SIAD.[95,96] A significant proportion of patients presenting biochemically with SIAD after both TBI[39] and SAH[96] have hyponatremia due to acute glucocorticoid deficiency; such patients respond to treatment with hydrocortisone rather than fluid restriction. SIAD also commonly occurs after hypophysectomy and after surgery for primary brain tumors (see earlier). Cerebral salt wasting (CSW) is an extremely rare cause of neurosurgical hyponatremia; a prospective analysis of 100 cases of SAH that included careful clinical and biochemical assessment of volume status failed to reveal a single case.[96] Low plasma osmolality, high urine osmolality, and high urinary sodium concentration are features of both SIAD and CSW; a key to distinguishing between the two

TABLE 12.5 Etiology of Syndrome of Inappropriate Antidiuresis (SIAD)

Malignancy	Drugs	Pulmonary	Intracranial Pathology	Miscellaneous
SCC lung	DDAVP	Pneumonia (esp. mycoplasma)	Tumors	Multiple sclerosis
Mesothelioma	PGs	TB	Meningitis	Guillain–Barré syndrome
GI tract	SSRIs	Abscess	Encephalitis	Acute intermittent porphyria
Pancreas	TCAs	Vasculitis	Abscess	HIV
GU tract	Phenothiazines	Positive pressure ventilation	Vasculitis	Idiopathic
Nasopharyngeal	Haloperidol	Covid-19 infection	Subarachnoid hemorrhage	
Lymphoma	MDMA		Subdural hemorrhage	
Sarcoma	Quinolones Levetiracetam Carbamazepine Cyclophosphamide		Traumatic brain injury	
	Vincristine			

SSRI, Selective serotonin reuptake inhibitor; *TCA,* tricyclic antidepressant; *PGs,* prostaglandins; *MDMA,* 3,4-methylenedioxymethamphetamine; *SCC,* small cell lung; *GI,* gastrointestinal; *GU,* genitourinary; *DDAVP,* desmopressin; *TB,* tuberculosis; *HIV,* human immunodeficiency syndrome.

is the demonstration of an initial period of natriuresis and hypovolemia preceding the onset of hyponatremia in CSW.

Many drugs are important clinical causes of hyponatremia. The selective serotonin reuptake inhibitors (SSRIs) are thought to cause SIAD by direct stimulation of AVP secretion by serotonin.[97] SIAD usually occurs in the first few weeks after commencing SSRIs, particularly in elderly patients.[98] It has been estimated that 12% of hospitalized patients on SSRI therapy develop SIAD.[99] Psychotropic such as phenothiazines, haloperidol, and tricyclic antidepressants all cause SIAD. Many patients who have conditions treated with psychotropic drugs also have abnormal thirst, and if these patients develop drug-induced SIAD, they can develop very significant hyponatremia. For instance, up to 20% of patients with chronic schizophrenia have psychogenic polydipsia, and excess fluid intake can precipitate severe hyponatremia. MDMA ("ecstasy") is thought to stimulate both AVP release and thirst as a result of hyperthermia, and this has been implicated in the development of cases of life-threatening hyponatremia.[100]

SIAD needs to be distinguished from exercise-associated hyponatremia (EAH),[101] where the electrolyte profile mimics SIAD. The pathophysiology is different, however. EAH occurs in athletes who ingest excessive hypotonic fluid during exercise, with resultant dilutional hyponatremia. Women, marathon runners racing for more than 4 hours, and athletes of low body mass index are most at risk.[102] Although some athletes with EAH fail to suppress AVP, EAH should probably be considered to be distinct from SIAD.

Pathophysiology of Syndrome of Inappropriate Antidiuresis

SIAD occurs by definition when AVP secretion is not suppressed when plasma sodium concentration falls below the osmotic threshold for AVP release. Zerbe and colleagues identified four different types of SIAD, depending on the pattern of AVP secretion[94] (Fig. 12.9):

1. Type A is the most common form of SIAD (40%). Characteristically, type A patients exhibit excessive, random secretion of AVP, with loss of the close linear relationship between plasma osmolality and plasma AVP. Plasma AVP concentrations fluctuate widely in this variant, with no relationship to ambient plasma osmolality, but with consistent antidiuresis and inappropriate urine concentration despite hyponatremia. Type A is common in lung cancer; *in vitro* studies have demonstrated that some lung tumors synthesize AVP,[103] and tumor tissue stains positive for AVP mRNA.[104] However, type A SIAD also occurs also in nonneoplastic disease where ectopic AVP secretion could not be implicated. Plasma AVP concentrations in type A SIAD are not suppressed by drinking,[105] which makes patients vulnerable to the development of severe hyponatremia.
2. Type B is also common (40%). The osmotic threshold for AVP release is lowered—a "reset osmostat"—such that secretion of AVP occurs at lower plasma osmolalities than normal. At plasma osmolalities above this reset threshold, the linear relationship between plasma AVP and plasma osmolality is preserved. Equally, because AVP is suppressed at plasma osmolalities below the lower reset threshold, overhydration leads to suppression of AVP release. A hypotonic diuresis then develops, which protects against the progression to severe hyponatremia. Similar lowering of the osmotic threshold occurs in hypovolemia and hypotension, leading to speculation that the function of afferent baroregulatory pathways is disturbed in these patients.
3. Type C is a rare condition characterized by failure to suppress AVP secretion at plasma osmolalities below the osmotic threshold. Plasma AVP concentrations are thus inappropriately high at low plasma osmolalities, but there is a normal relationship between plasma osmolality and plasma AVP at physiological plasma osmolalities. This variant may be due to dysfunction of inhibitory neurons in the hypothalamus, leading to persistent low-grade basal AVP secretion.
4. Type D is a rare clinical picture of SIAD without detectable circulating AVP.[106] A nephrogenic syndrome of inappropriate antidiuresis (NSIAD) was postulated after two male infants were described in whom gain-of-function mutations in the V2 receptor led to a clinical picture of SIAD with undetectable AVP levels.[107] Various gain-of-function mutations have now been described.[108]

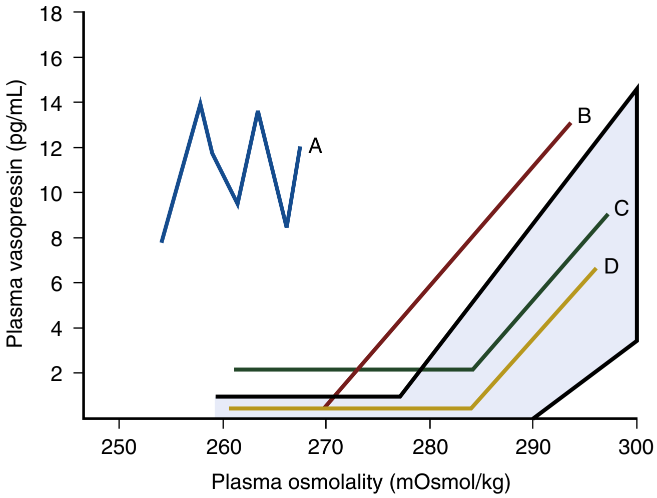


Fig. 12.9 Summary of the four different patterns of arginine vasopressin secretion in syndrome of inappropriate antidiuretic hormone secretion.

Studies in rat models of SIAD have demonstrated that the increase in water reabsorption is secondary to AVP-mediated expression of renal aquaporin-2,[109] with a consequent increase in aquaporin-2 protein excretion in the urine.[110] Aquaporin-2 recruitment has been shown to be reversible by administration of a V2 receptor antagonist,[109,111] which provides the scientific rationale for the use of AVP antagonists in the treatment of this condition. Prolonged SIAD is associated with "escape" from antidiuresis, with downregulation of AQP2 mRNA and protein expression.

Although patients with SIAD have ambient plasma osmolalities that are below the physiological osmotic threshold for thirst, they continue to drink apparently normal fluid volumes. In experiments where patients with SIAD of mixed etiology were infused intravenously with hypertonic sodium chloride solution, the osmotic threshold for thirst was lowered to a setting that was identical for that of AVP.[105] The parallel lowering of the thresholds for thirst and AVP release in SIAD ensured the maintenance of fluid intake, predisposing to persistent hyponatremia.

As fluid intake is maintained in the face of reduced free water clearance, patients with SIAD can develop severe hyponatremia. However, hyponatremia is often limited by "escape from antidiuresis." This protective homeostatic mechanism occurs when the kidney begins to increase free water clearance despite inappropriate plasma AVP concentrations.[112] Initial natriuresis is followed by an increase in urine flow,[113] with consequent water loss; this allows plasma sodium to stabilize and, occasionally, to rise. Although plasma sodium concentration does not usually rise into the normal physiological range during escape from antidiuresis, the development of severe hyponatremia is prevented.

Diagnosis of Syndrome of Inappropriate Antidiuresis

The essential criteria for diagnosis of SIAD are presented in Table 12.6. There are two additional supplemental criteria—raised plasma AVP concentration and abnormal water load test—but in clinical practice they are rarely used. Water-loading a patient with SIAD runs the risk of inducing symptomatic hyponatremia, and this procedure should therefore be restricted to centers with considerable experience in managing disorders of water balance. Plasma AVP and copeptin concentrations are elevated in almost all causes of hyponatremia and are therefore rarely of diagnostic benefit in hyponatremic patients.

The minimum data set for the diagnosis of SIAD is hyponatremia in a euvolemic patient with inappropriate concentrated urine (osmolality >100 mOsm/kg), elevated urine sodium (>30 mmol/L), and exclusion of cortisol and thyroid hormone insufficiency. Urine osmolalities in excess of 100 mOsm/kg indicate inappropriate urine concentration and are compatible with the diagnosis of SIAD. In clinical practice, the ascertainment of the minimum data set for the diagnosis is poor,[114] which raises concern that hyponatremic patients may be misclassified as having SIAD, and therefore be treated inappropriately, on the basis of inadequate investigation.

Differential Diagnosis of Syndrome of Inappropriate Antidiuresis

Glucocorticoid Deficiency. The major differential in euvolemic hyponatremia is ACTH deficiency, which causes hyponatremia with a biochemical profile identical to classical SIAD. Management with glucocorticoid replacement restores normonatremia,[115] whereas fluid restriction, which is first-line therapy for SIAD, is deleterious, in that it worsens hypovolemia.[116] Measurement of plasma cortisol is essential, as clinical signs do not reliably differentiate between glucocorticoid deficiency and SIAD.[117] Cortisol deficiency impairs free water clearance, and glucocorticoids suppress AVP release, such that elevation of plasma AVP concentrations has been reported in ACTH-deficient patients who develop hyponatremia.[118]

The differentiation of SIAD from glucocorticoid deficiency is particularly important in hyponatremic patients with neurosurgical conditions.[119] SIAD is the commonest cause of neurosurgical hyponatremia,[120] but acute pituitary dysfunction occurs commonly following TBI and subarachnoid hemorrhage. ACTH deficiency was reported in 16% of patients with acute[37] and 12.7% patients with chronic head injury[121]; recent data suggest that up to 80% of patients have inappropriately low plasma cortisol concentrations in the acute phase of brain injury, some of whom present as having SIAD.[39] Up to 10% of cases of SIAD presenting after SAH may be secondary to acute ACTH/cortisol deficiency.[96] In neurosurgical patients who present with biochemical manifestations of SIAD, coexistent hypoglycemia or hypotension suggest the diagnosis of hypopituitarism. In patients presenting with SIAD against a background of TBI, SAH, or intracranial hemorrhage, we consider that the possibility of acute hypopituitarism should be considered in the differential diagnosis. Published data have shown that, in euvolemic hyponatremia, a random plasma cortisol greater than 300 nmol/L (10.9 µg/dL) predicts a pass response to intramuscular *synacthen* (*cosyntropin*), so this seems to be a reasonable screening test. Where a random sample falls below this, the diagnosis should be confirmed with a *synacthen* test.

TABLE 12.6 Essential Criteria for the Diagnosis of Syndrome of Inappropriate Antidiuresis (SIAD)

Decreased plasma osmolality of the ECF (Posm <280 mOsm/kg H_2O)
Inappropriate urinary concentration during hyposmolality (Uosm > 100 mOsm/kg H_2O, normal renal function)
Clinical euvolemia
Elevated urinary sodium excretion (>30 mmol/L) on normal salt and water intake
Absence of other potential causes of euvolemic hypoosmolality, particularly hypothyroidism, hypocortisolism, and diuretic use

ECF, Extracellular fluid; *Posm*, plasma osmolality; *Uosm*, urine osmolality.

Hypothyroidism. Hyponatremia in hypothyroidism is rare, though life-threatening hyponatremia is occasionally seen in severely myxoedematous patients.[122,123] Hypothyroid patients have elevated AVP responses to subtle volume contraction, with increased recruitment of AQP2.[124] However, the hyponatremia is probably multifactorial. Hypothyroidism so rarely produces SIAD[125] that the need to check thyroid function tests has been questioned.

KEY POINTS

- The minimum data set for the diagnosis of syndrome of inappropriate antidiuresis (SIAD) is hyponatremia in a euvolemic patient with inappropriate concentrated urine (osmolality >100 mOsm/kg), elevated urine sodium (>30 mmol/L), and exclusion of cortisol and thyroid hormone insufficiency. The major differential diagnosis in euvolemic hyponatremia is adrenocorticotropic hormone deficiency, which causes hyponatremia with a biochemical profile identical to classical SIAD and should not be missed.

TREATMENT OF HYPONATREMIA

The key issues that determine optimum management of hyponatremia are accurate assessment of the underlying cause and awareness of the chronicity of low plasma sodium concentrations. While acute hyponatremia is a medical emergency that can lead to death if untreated, chronic hyponatremia has long been thought of as a benign condition, and is often left underinvestigated and untreated. However, emerging evidence of the role of hyponatremia in falls, osteoporosis, fractures, length of stay, and mortality now provides a rationale for its treatment also. Recent data have suggested that effective treatment of hyponatremia can improve cognitive function[126] and gait[82] and reduce mortality.[89]

Factors Affecting Therapy

Clinical guidelines have stressed that the urgency of treatment of hyponatremia is dependent upon the presence of symptomatic cerebral irritation. As patients with chronic hyponatremia adapt to the electrolyte derangement, it is patients with sudden-onset hyponatremia who are most likely to exhibit signs of cerebral irritation. Neurological signs of hyponatremia almost always reflect cerebral edema, leading to increased intracranial pressure, so the clinical imperative to treat hyponatremia is more urgent when the rate of fall of plasma sodium concentration has been rapid.

Acute hyponatremia is defined as a fall in plasma sodium concentration of less than 48 hours' duration and is generally regarded as a medical emergency,[127] with high mortality in the absence of intervention.[128] In contrast, patients with chronic hyponatremia (>48 hours' duration) are much less likely to develop symptoms, as they undergo cerebral adaptation by extruding solutes such as potassium and organic osmolytes into the plasma. This prevents osmotic movement of water into the brain, and as a result cerebral edema is avoided. In contrast, in acute hyponatremia this time-dependent process does not occur, with the subsequent development of brain edema and cerebral irritation.[129]

Treatment of Chronic Hyponatremia Owing to Syndrome of Inappropriate Antidiuresis

Clear guidelines for the management of hyponatremia due to SIAD have been published in both the United States[127] and Europe.[130] Although modest hyponatremia of greater than 130 mmol/L has traditionally been regarded as not in need of treatment, there is increasing evidence that comorbidities such as gait instability and falls,[131] fractures,[84] and osteoporosis[85] may occur even with modest hyponatremia. After careful exclusion of treatable underlying conditions such as glucocorticoid deficiency, there are a number of options for the treatment of SIAD.

Fluid Restriction. Clinical guidelines consistently recommend fluid restriction as first-line treatment, despite the absence of a strong evidence base. Fluid restriction of 1 L per day is occasionally sufficient to correct hyponatremia,[132] although the rate of rise of plasma sodium is not predictable. A recent randomized controlled trial of fluid restriction versus no treatment for chronic SIAD showed that fluid restriction produces a modest elevation of 3 mmol/L in plasma sodium, which is sustained over a month of therapy.[133] However, few patients normalized plasma sodium concentration over the study period. Fluid restriction, which is inclusive of all fluids, including intravenous fluids and liquid dietary supplements, is difficult to maintain in patients with cognitive difficulties, in patients with a history of polydipsia, or in patients receiving intravenous medications. In addition, in more severe hyponatremia, the degree of fluid restriction necessary to correct hypoosmolality may be too difficult to maintain. The US guidelines have stressed that, if urine osmolality is greater than 500 mOsm/kg, the Furst equation ratio is greater than 1, or 24-hour urine volume is less than 1.5 L, fluid restriction is unlikely to be effective.[127] In clinical practice 60% of patients have one of these predictors, and 40% have two,[134] and in a prospective study both urine osmolality and urine sodium were associated with nonresponse to fluid restriction.[135] In theory, therefore, it should be possible to select patients, based on these criteria, who will respond best to fluid restriction.

Pharmacological therapy tends to be reserved for cases that are unresponsive to fluid restriction, but it is clear from the disappointing results of prospective studies[133] that a substantial number of patients fall into this category.

Demeclocycline. Demeclocycline is a tetracycline antibiotic that induces NDI by a postreceptor effect. Demeclocycline can be effective but is not licensed for use in all countries. There is a variable response to this agent, and a systematic review of the literature in 2015 found a lack of clinical or economic evidence to support its use.[136] In addition, side effects, including photosensitivity and nephrotoxicity, particularly in patients with liver disease, can complicate chronic therapy. Occasionally, hypernatremic dehydration can develop if polyuria is marked. With the development of new therapies, there seems to be little justification for using this treatment.

Urea. Urea therapy has been advocated as second-line therapy in the European guidelines,[130] and in one small study was found to be as effective as vaptans.[137] The same group has reported data suggesting, from uncontrolled studies, that urea is a safe therapy for use in euvolemic hyponatremia in intensive care patients.[138] A recent retrospective review of nonrandomized patients found urea to be effective and safe,[139] but prospective randomized trials are awaited. Urea is contraindicated in renal failure, the taste is unpleasant, and it is difficult to source; its use is limited to a few countries.

Furosemide With Sodium Chloride Replacement. The combination of furosemide, which produces a natriuresis and a diuresis, with sodium chloride tablets to replace urine sodium loss, has been advocated for many years for the treatment of SIAD. A recent randomized controlled trial showed no benefit to the addition of furosemide and sodium chloride, with or without fluid restriction, compared with fluid restriction alone; there was, however, a higher incidence of acute kidney injury and hypokalemia in patients treated with furosemide and sodium chloride,[140] so better evidence is needed before recommending this treatment.

Vasopressin Receptor Antagonists. The specific V2 receptor antagonists (vaptans) offer a specific antidote to AVP-mediated hyponatremia.[141] These agents act as "aquaretics," in that they increase free water excretion without a natriuresis and offer a therapeutic effect regardless of the etiology of raised AVP. The selective V2 antagonist tolvaptan is licensed for use as an oral agent in Europe, whereas in the United States both tolvaptan and the combined V1a/V2 antagonist conivaptan[142] are available.

The efficacy and safety of oral tolvaptan in ambulatory patients was explored in two randomized, double-blind, placebo-controlled multicenter trials (SALT-1 and SALT-2) consisting of 448 patients with hyponatremia of multiple etiologies, including SIAD, heart failure, or cirrhosis.[143] The administration of tolvaptan resulted in a significant increase in serum sodium concentration compared with placebo, without the requirement of water restriction. The group treated with tolvaptan was found to have greater improvement in quality of life, as assessed by the Short Form–36 questionnaire. The incidence of overrapid correction of hyponatremia was low, and no patients developed central pontine myelinolysis. A subsequent extension study, the SALTWATER trial, showed that the safety and efficacy of tolvaptan were maintained over a mean of almost 2 years,[144] with no significant adverse effects. A subgroup analysis of the effects of tolvaptan on the SIAD cohort of the SALT studies confirmed the efficacy of the drug in SIAD.[145] The use of high-dose tolvaptan for the treatment of polycystic kidney disease has been associated with elevated serum transaminases,[146] so liver function tests should be monitored in patients on long-term treatment.

Summary of Second-Line Therapies. Although fluid restriction is the recommended first-line therapy in international guidelines, randomized prospective trials have shown modest effects, with relatively few patients attaining correction targets.[133] In addition, the "real world" effects of fluid restriction therapy have been even less impressive, as shown by the results of the Hyponatremia Registry, where the effect of fluid restriction was not significantly better than no treatment.[147] There is therefore a clear need for second-line therapy. The guidelines differ in their recommendations, in that the European Guidelines strongly endorsed urea therapy while advising against tolvaptan,[130] whereas the US guidelines endorsed physician choice over second-line therapy, while recognizing that the vaptans had the strongest supportive trial data.[127] Cost considerations mitigate against widespread use of vaptans, whereas urea is not universally available as a therapeutic compound and has a taste that is unpalatable. There is a clear need therefore for a cost-reasonable, effective alternative to fluid restriction.

Potential New Therapies. New agents are therefore being explored for the treatment of SIAD. A recent study found that empagliflozin, the sodium glucose cotransporter-2 inhibitor used as an oral hypoglycaemic agent, had significant effects on plasma sodium concentration in patients with hyponatremia due to SIAD. In a randomized comparison to fluid restriction, a short course of empagliflozin had more significant sodium-elevating effects than fluid restriction (10 vs. 7 mmol/L over 4 days) in hospitalized patients with SIAD; the effects were more marked in patients with severe hyponatremia.[148] Long-term effects were not assessed, nor was there time for the development of side effects, and the drug will certainly need further study before licensing for treatment of SIAD. However, the results were promising. The mechanism of action is almost certainly increased free water clearance secondary to glycosuria.

Apelin is a neurovasoactive peptide that is expressed in the SON and PVN, where it is thought to exert an inhibitory effect on AVP release.[149] It has been hypothesized that patients with SIAD have an imbalance in the apelin/AVP ratios, which contributes to the state of antidiuresis.[150] Apelin expression has also been demonstrated in the kidney, where apelin analogs have been shown in animal studies to reduce the generation of aquaporin-2 by AVP,[151] and the effects of apelin to reduce AVP antidiuresis may be an amalgamation of hypothalamic and renal effects.[152] Recent rat data has shown that an apelin-R agonist can increase urine osmolality and increase plasma sodium in hyponatremic animals.[153] Apelin clearly has profound effects on modulating AVP secretion and action, and its agonists have the potential for a therapeutic role in SIAD.

Treatment of Acute Hyponatremia

Untreated severe hyponatremia of acute onset, especially when associated with evidence of cerebral irritation, is potentially fatal.[154,155] However, treatment is potentially hazardous, as rapid correction of hyponatremia renders the patient vulnerable to the risk of osmotic demyelination, a devastating complication characterized by spastic quadriparesis and cranial nerve palsies. The risk of osmotic demyelination is highest when chronic hyponatremia is rapidly corrected; patients with chronic hyponatremia have often adapted, and leakage of inorganic ions equalizes the osmolality across the blood–brain barrier. This reduces brain swelling and intracranial pressure. As a result, patients with chronic hyponatremia tend to have fewer symptoms of cerebral irritation and are therefore less in need of heroic rescue. Conversely, osmotic demyelination during treatment is less of a risk when hyponatremia is of acute onset. Therefore, those patients at most risk of death from acute hyponatremia also have the least risk of developing osmotic demyelination.

Guidelines now advocate treatment paradigms based on the speed of onset of hyponatremia and the presence of symptoms of cerebral irritation, rather than how low the plasma sodium concentration is.[127,130] This is in contrast to earlier interpretations, which linked symptoms directly to the severity of hyponatremia.[156] Therefore, current recommendations are that rapid reversal of hyponatremia, using intravenous hypertonic sodium chloride solution, should be reserved for patients with hyponatremic encephalopathy and good evidence of the rapid onset of hyponatremia.

The traditional treatment of acute symptomatic hyponatremia has been the steady intravenous infusion of 3% sodium chloride solution, with the aim of correction of plasma sodium concentration by no more than 0.5 mmol/L/hour, or 12 mmol/L over 24 hours.[157] Recent expert panel recommendations have suggested some changes to this traditional guideline.[127] The first change is that the aim for the elevation in plasma sodium should not be for an evenly distributed increase across the first 24 hours. Observations in normonatremic neurosurgical patients, that acute elevation in plasma sodium concentration of 4 to 6 mmol/L reduces signs of cerebral herniation and lowered intracranial pressure by 50%,[158] prompted a review of available evidence in hyponatremic patients. This concluded that a similar rise in plasma sodium concentration of 4 to 6 mmol/L is sufficient to reverse the more dangerous effects of acute hyponatremia.[159] As the development of cerebral herniation due to acute hyponatremia occurs rapidly and is associated with significant mortality,[128] a new approach was recommended for hyponatremia with neurological symptoms. This entailed the acute elevation of the plasma sodium, initially by 3 to 5 mmol/L over 2 to 4 hours, by the intravenous injection of 50- to 100-mL boluses of 3% sodium chloride, supported by frequent checking of plasma sodium concentration; this strategy was considered more likely to reverse cerebral edema, reduce intracranial pressure, and prevent seizures.[127] The remainder of the target rise in plasma sodium can occur over 24 hours. This strategy was tested by prospective study of patients with acute symptomatic SIAD treated with bolus 3% saline, as compared with historical controls who had been treated with slow intravenous infusion.[160] A comparison of the rate of rise of plasma sodium concentration is shown (Fig. 12.10); bolus 3% saline led to a quicker rise in plasma sodium concentration in the first 6 hours, but similar rises at 24 hours. The important endpoint was a faster return to normal Glasgow Coma Scale (GCS) in the group treated with bolus injections. Although there were no cases of osmotic demyelination in either group, the need to relower plasma sodium because of overcorrection was slightly higher in the bolus group.

The second alteration in the US guidelines was the target rise in plasma sodium concentration over the first 24 hours.[127] As osmotic demyelination has occasionally been reported with rises in plasma sodium concentration within the 12 mmol/L/24 hours guidelines, the

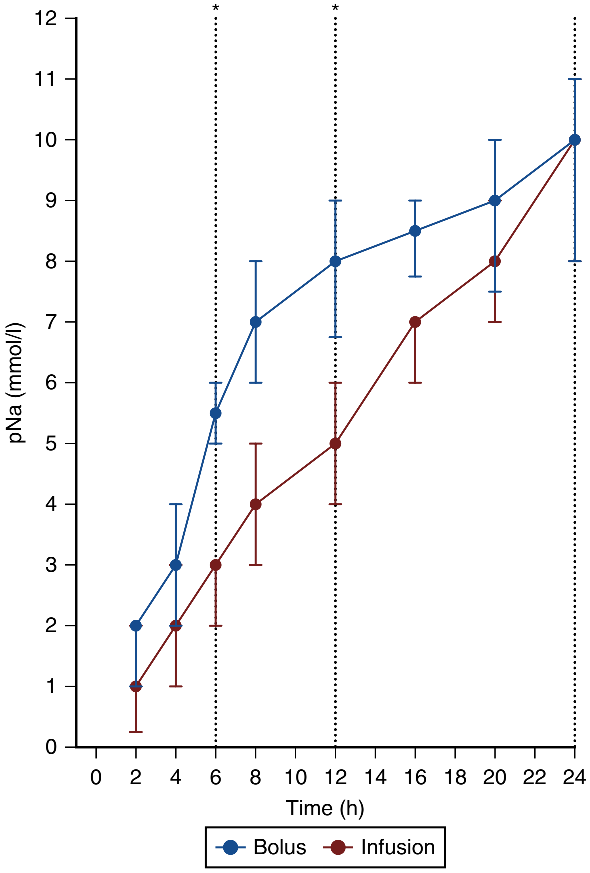


Fig. 12.10 Change in plasma sodium (*pNa*) concentration in patients with symptomatic hyponatremia due to syndrome of inappropriate antidiuresis (SIAD) treated with bolus versus slow infusion of hypertonic saline.

24-hour target was lowered to 8 mmol/L, with a maximum tolerable rise of 12 mmol/L. Lower targets still were recommended for patients with higher risk of osmotic demyelination, for instance those with malnutrition, liver disease, or alcohol abuse.[127]

There seems to be a greater awareness of the mortality associated with acute severe hyponatremia. Although early studies have shown high mortality associated with acute hyponatremia,[128,156] more recent studies have shown much lower rates,[90] which have been attributed to more specialized endocrine input and more careful attention to electrolyte monitoring.

KEY POINTS

- The key issues that determine optimum management of hyponatremia are accurate assessment of the underlying cause and awareness of the chronicity of low plasma sodium concentrations. Patients with chronic hyponatremia may have minimal symptoms, and cerebral herniation is uncommon. In this scenario, it is the risk of osmotic demyelination syndrome from overly rapid correction of plasma sodium concentration that guides treatment regimens. Fluid restriction is the recommended first-line therapy for chronic syndrome of inappropriate antidiuresis (SIAD) in international guidelines; however, randomized prospective trials have shown modest effects on plasma sodium concentration. Second-line therapies for SIAD are limited by cost, availability, adverse effects, and/or lack of prospective randomized data. Acute symptomatic hyponatremia is a medical emergency, and recent international guidelines emphasize the importance of rapid initial correction of plasma sodium correction, initially by 3 to 5 mmol/L over 2 to 4 hours, using boluses of hypertonic saline.

CONCLUSIONS

There have been significant developments in the areas of DI and hyponatremia in the past decade. The diagnostic approach to polyuria syndromes is evolving. Dynamic direct stimulation tests of AVP secretion, with measurement of the surrogate marker copeptin, look set to take the place of the traditional gold standard water deprivation test in differentiating partial HDI from primary polydipsia. In hyponatremia, prospective randomized trial data have shown fluid restriction produces only a modest improvement in plasma sodium concentration in patients with chronic SIAD. V2 receptor antagonists have been shown to be effective in randomized controlled trials, but their widespread use is limited by cost and concerns about risk of overcorrection. New affordable treatment options for SIAD are needed, both for clinical practice and to enable future studies to go one step further and demonstrate treatment effect on complications of hyponatremia. Future studies of hyponatremia treatment should focus not only on correction of hyponatremia as an endpoint, but also improvements in associated comorbidities such as unsteady gait, poor cognition, falls, and fractures. Finally, the debate as to whether treatment of hyponatremia results in improvements in mortality risk remains open, and a multinational prospective trial is underway to examine this.

REFERENCES

1. Schally AV. Hormones of the neurohypophysis. In: Lock W, Schally AV, eds. *The Hypothalamus and Pituitary in Health and Disease*. Springfield: Charles C. Thomas; 1972:154–171.
2. Russell JT, Brownstein MJ, Gainer H. Biosynthesis of vasopressin, oxytocin, and neurophysins: isolation and characterization of two common precursors (propressophysin and prooxyphysin). *Endocrinology*. 1980;107:1880–1891.
3. Dutton A, Dyball RE. Phasic firing enhances vasopressin release from the rat neurohypophysis. *J Physiol*. 1979;290:433–440.
4. Thibonnier M, Conarty DM, Preston JA, et al. Molecular pharmacology of human vasopressin receptors. *Adv Exp Med Biol*. 1998;449:251–276.
5. Gonen T, Walz T. The structure of aquaporins. *Q Rev Biophys*. 2006;39:361–396.
6. Robertson GL, Aycinena P, Zerbe RL. Neurogenic disorders of osmoregulation. *Am J Med*. 1982;72:339–353.
7. Zerbe RL, Robertson GL. Osmoregulation of thirst and vasopressin secretion in human subjects: effect of various solutes. *Am J Physiol*. 1983;244:E607–E614.
8. Thompson CJ, Davis SN, Butler PC, et al. Osmoregulation of thirst and vasopressin secretion in insulin-dependent diabetes mellitus. *Clin Sci*. 1988;74:599–606.
9. Robertson GL, Shelton RL, Athar S. The osmoregulation of vasopressin. *Kidney Int*. 1976;10:25–37.
10. Hammer M, Ladefoged J, Olgaard K. Relationship between plasma osmolality and plasma vasopressin in human subjects. *Am J Physiol*. 1980;238:E313–E317.
11. Baylis PH, Thompson CJ. Osmoregulation of vasopressin secretion and thirst in health and disease. *Clin Endocrinol*. 1988;29:549–576.
12. Thompson CJ, Bland J, Burd J, et al. The osmotic thresholds for thirst and vasopressin release are similar in healthy man. *Clin Sci*. 1986;71:651–656.
13. Baylis PH, Pippard C, Gill GV, et al. Development of a cytochemical assay for plasma vasopressin: application to studies on water loading normal man. *Clin Endocrinol*. 1986;24:383–393.
14. Baylis PH, Gill GV. The investigation of polyuria. *Clin Endocrinol Metab*. 1984;13:295–310.
15. Thompson CJ, Selby P, Baylis PH. Reproducibility of osmotic and nonosmotic tests of vasopressin secretion in men. *Am J Physiol*. 1991;260:R533–R539.
16. Thompson CJ, Burd JM, Baylis PH. Acute suppression of plasma vasopressin and thirst after drinking in hypernatremic humans. *Am J Physiol*. 1987;252:R1138–R1142.
17. Phillips PA, Rolls BJ, Ledingham JG, et al. Reduced thirst after water deprivation in healthy elderly men. *N Engl J Med*. 1984;311:753–759.
18. Crowe MJ, Forsling ML, Rolls BJ, et al. Altered water excretion in healthy elderly men. *Age Ageing*. 1987;16:285–293.
19. Thompson CJ, Davis SN, Baylis PH. Effect of blood glucose concentration on osmoregulation in diabetes mellitus. *Am J Physiol*. 1989;256:R597–R604.
20. McKenna K, Morris AD, Ryan M, et al. Renal resistance to vasopressin in poorly controlled type 1 diabetes mellitus. *Am J Physiol Endocrinol Metab*. 2000;279:E155–E160.
21. McKenna K, Morris AD, Azam H, et al. Exaggerated vasopressin secretion and attenuated osmoregulated thirst in human survivors of hyperosmolar coma. *Diabetologia*. 1999;42:534–538.
22. Winzeler B, Cesana-Nigro N, Refardt J, et al. Arginine-stimulated copeptin measurements in the differential diagnosis of diabetes insipidus: a prospective diagnostic study. *Lancet*. 2019;394:587–595.
23. Land H, Schütz G, Schmale H, et al. Nucleotide sequence of cloned cDNA encoding bovine arginine vasopressin-neurophysin II precursor. *Nature*. 1982;295:299–303.
24. Szinnai G, Morgenthaler NG, Berneis K, et al. Changes in plasma copeptin, the c-terminal portion of arginine vasopressin during water deprivation and excess in healthy subjects. *J Clin Endocrinol Metab*. 2007;92:3973–3978.
25. Thompson CJ. Polyuric states in man. *Baillieres Clin Endocrinol Metab*. 1989;3:473–497.
26. Thompson CJ, Baylis PH. Thirst in diabetes insipidus: clinical relevance of quantitative assessment. *Quartly J Med*. 1987;65:853–862.
27. Hensen J, Henig A, Fahlbusch R, et al. Prevalence, predictors and patterns of postoperative polyuria and hyponatraemia in the immediate course after transsphenoidal surgery for pituitary adenomas. *Clin Endocrinol (Oxf)*. 1999;50:431–439.
28. Smith D, Finucane F, Phillips J, et al. Abnormal regulation of thirst and vasopressin secretion following surgery for craniopharyngioma. *Clin Endocrinol*. 2004;61:273–279.

29. Nemergut EC, Zuo Z, Jane Jr JA, et al. Predictors of diabetes insipidus after transsphenoidal surgery: a review of 881 patients. *J Neurosurg*. 2005;103:448–454.
30. Adams JR, Blevins Jr LS, Allen GS, et al. Disorders of water metabolism following transsphenoidal pituitary surgery: a single institution's experience. *Pituitary*. 2006;9:93–99.
31. Sudhakar N, Ray A, Vafidis JA. Complications after trans-sphenoidal surgery: our experience and a review of the literature. *Br J Neurosurg*. 2004;18:507–512.
32. Schreckinger M, Walker B, Knepper J, et al. Post-operative diabetes insipidus after endoscopic transsphenoidal surgery. *Pituitary*. 2013;16:445–451.
33. Agha A, Sherlock M, Brennan S, et al. Hypothalamic-pituitary dysfunction after irradiation of nonpituitary brain tumors in adults. *J Clin Endocrinol Metab*. 2005;90:6355–6360.
34. Imura H, Nakao K, Shimatsu A, et al. Lymphocytic infundibuloneurohypophysitis as a cause of central diabetes insipidus. *N Engl J Med*. 1993;329:683–689.
35. Pivonello R, De Bellis A, Faggiano A, et al. Central diabetes insipidus and autoimmunity: relationship between the occurrence of antibodies to arginine vasopressin-secreting cells and clinical, immunological, and radiological features in a large cohort of patients with central diabetes insipidus of known and unknown etiology. *J Clin Endocrinol Metab*. 2003;88:1629–1636.
36. Hannon M, Orr C, Moran C, et al. Anterior hypopituitarism is rare and autoimmune disease is common in adults with idiopathic central diabetes insipidus. *Clin Endocrinol (Oxf)*. 2011;76:725–728.
37. Agha A, Rogers B, Mylotte D, et al. Neuroendocrine dysfunction in the acute phase of traumatic brain injury. *Clin Endocrinol (Oxf)*. 2004;60:584–591.
38. Agha A, Sherlock M, Phillips J, et al. The natural history of post-traumatic neurohypophysial dysfunction. *Eur J Endocrinol*. 2005;152:371–377.
39. Hannon MJ, Crowley RK, Behan LA, et al. Acute glucocorticoid deficiency and diabetes insipidus are common after acute traumatic brain injury and predict mortality. *J Clin Endocrinol Metab*. 2013;98:3229–3237.
40. Agha A, Thornton E, O'Kelly P, et al. Posterior pituitary dysfunction after traumatic brain injury. *J Clin Endocrinol Metab*. 2004;89:5987–5992.
41. Jialal I, Desai RK, Rajput MC. An assessment of posterior pituitary function in patients with Sheehan's syndrome. *Clin Endocrinol (Oxf)*. 1987;27:91–95.
42. Baylis PH, Thompson C, Burd J, et al. Recurrent pregnancy-induced polyuria and thirst due to hypothalamic diabetes insipidus: an investigation into possible mechanisms responsible for polyuria. *Clin Endocrinol (Oxf)*. 1986;24:459–466.
43. Ito M, Jameson JL. Molecular basis of autosomal dominant neurohypophyseal diabetes insipidus. Cellular toxicity caused by the accumulation of mutant vasopressin precursors within the endoplasmic reticulum. *J Clin Invest*. 1997;99:1897–1905.
44. Eubanks S, Nguyen TL, Deeb R, et al. Effects of diabetes insipidus mutations on neurophysin folding and function. *J Biol Chem*. 2001;276:29671–29680.
45. Elias PC, Elias LL, Torres N, et al. Progressive decline of vasopressin secretion in familial autosomal dominant neurohypophyseal diabetes insipidus presenting a novel mutation in the vasopressin-neurophysin II gene. *Clin Endocrinol (Oxf)*. 2003;59:511–518.
46. Barrett TG, Bundey SE. Wolfram (DIDMOAD) syndrome. *J Med Genet*. 1997;34:838–841.
47. Thompson CJ, Charlton J, Walford S, et al. Vasopressin secretion in the DIDMOAD (Wolfram) syndrome. *Q J Med*. 1989;71:333–345.
48. Dineen R, Bogdanet D, Thompson D, et al. Endocrinopathies and renal outcomes in lithium therapy: impact of lithium toxicity. *QJM: Monthly J Assoc Phys*. 2017;110:821–827.
49. Babey M, Kopp P, Robertson GL. Familial forms of diabetes insipidus: clinical and molecular characteristics. *Nat Rev Endocrinol*. 2011;7:701–714.
50. Barberis C, Mouillac B, Durroux T. Structural bases of vasopressin/oxytocin receptor function. *J Endocrinol*. 1998;156:223–229.
51. Asai T, Kuwahara M, Kurihara H, et al. Pathogenesis of nephrogenic diabetes insipidus by aquaporin-2 C-terminus mutations. *Kidney Int*. 2003;64:2–10.
52. Mulders SM, Bichet DG, Rijss JP, et al. An aquaporin-2 water channel mutant which causes autosomal dominant nephrogenic diabetes insipidus is retained in the Golgi complex. *J Clin Invest*. 1998;102:57–66.
53. McKenna K, Thompson C. Osmoregulation in clinical disorders of thirst appreciation. *Clin Endocrinol (Oxf)*. 1998;49:139–152.
54. Thompson CJ, Edwards CR, Baylis PH. Osmotic and non-osmotic regulation of thirst and vasopressin secretion in patients with compulsive water drinking. *Clin Endocrinol (Oxf)*. 1991;35:221–228.
55. Crowley RK, Hamnvik OP, O'Sullivan EP, et al. Morbidity and mortality in patients with craniopharyngioma after surgery. *Clin Endocrinol (Oxf)*. 2010;73:516–521.
56. Zerbe RL, Robertson GL. A comparison of plasma vasopressin measurements with a standard indirect test in the differential diagnosis of polyuria. *N Engl J Med*. 1981;305:1539–1546.
57. Robertson GL. Diagnosis of diabetes insipidus. In: Czernichow P, Robinson AG, eds. *Diabetes Insipidus in Man: Frontiers of Hormone Research*. Basel: S. Karger; 1985:176–189.
58. Fenske W, Refardt J, Chifu I, et al. A copeptin-based approach in the diagnosis of diabetes insipidus. *N Engl J Med*. 2018;379:428–439.
59. Rooke P, Baylis PH. A new sensitive radioimmunoassay for plasma arginine vasopressin. *J Immun*. 1982;3:115–131.
60. Balanescu S, Kopp P, Gaskill MB, et al. Correlation of plasma copeptin and vasopressin concentrations in hypo-, iso-, and hyperosmolar States. *J Clin Endocrinol Metab*. 2011;96:1046–1052.
61. Timper K, Fenske W, Kühn F, et al. Diagnostic accuracy of copeptin in the differential diagnosis of the polyuria-polydipsia syndrome: a prospective multicenter study. *J Clin Endocrinol Metab*. 2015;100:2268–2274.
62. Turcu AF, Erickson BJ, Lin E, et al. Pituitary stalk lesions: the Mayo Clinic experience. *J Clin Endocrinol Metab*. 2013;98:1812–1818.
63. Fujisawa I. Magnetic resonance imaging of the hypothalamic-neurohypophyseal system. *J Neuroendocrinol*. 2004;16:297–302.
64. Garrahy A, Moran C, Thompson CJ. Diagnosis and management of central diabetes insipidus in adults. *Clin Endocrinol*. 2019;90:23–30.
65. Maghnie M, Cosi G, Genovese E, et al. Central diabetes insipidus in children and young adults. *N Engl J Med*. 2000;343:998–1007.
66. Ghirardello S, Garre ML, Rossi A, et al. The diagnosis of children with central diabetes insipidus. *J Pediatr Endocrinol Metab*. 2007;20:359–375.
67. Di Iorgi N, Maria Allegri AE, Napoli F, et al. Central diabetes insipidus in children and young adults: etiological diagnosis and long-term outcome of idiopathic cases. *J Clin Endocrinol Metab*. 2013;99:1264–1272.
68. McIver B, Connacher A, Whittle I, et al. Adipsic hypothalamic diabetes insipidus after clipping of anterior communicating artery aneurysm. *BMJ (Clinical Research Ed)*. 1991;303:1465–1467.
69. Crowley RK, Sherlock M, Agha A, et al. Clinical insights into adipsic diabetes insipidus: a large case series. *Clin Endocrinol*. 2007;66:475–482.
70. Sinha A, Ball S, Jenkins A, et al. Objective assessment of thirst recovery in patients with adipsic diabetes insipidus. *Pituitary*. 2011;14:307–311.
71. Sherlock M, Agha A, Crowley R, et al. Adipsic diabetes insipidus following pituitary surgery for a macroprolactinoma. *Pituitary*. 2006;9:59–64.
72. Smith D, McKenna K, Moore K, et al. Baroregulation of vasopressin release in adipsic diabetes insipidus. *J Clin Endocrinol Metab*. 2002;87:4564–4568.
73. Ball SG, Vaidja B, Baylis PH. Hypothalamic adipsic syndrome: diagnosis and management. *Clin Endocrinol (Oxf)*. 1997;47:405–409.
74. Bergada I, Aversa L, Heinrich JJ. Peripheral venous thrombosis in children and adolescents with adipsic hypernatremia secondary to hypothalamic tumors. *Horm Res*. 2004;61:108–110.
75. Zantut-Wittmann DE, Garmes HM, Panzan AD, et al. Severe rhabdomyolysis due to adipsic hypernatremia after craniopharyngioma surgery. *Arq Bras Endocrinol Metabol*. 2007;51:1175–1179.
76. Arima H, Wakabayashi T, Nagatani T, et al. Adipsia increases risk of death in patients with central diabetes insipidus. *Endocr J*. 2014;61:143–148.
77. Behan LA, Sherlock M, Moyles P, et al. Abnormal plasma sodium concentrations in patients treated with desmopressin for cranial diabetes insipidus: results of a long-term retrospective study. *Eur J Endocrinol*. 2015;172:243–250.

78. Sanches TR, Volpini RA, Massola Shimizu MH, et al. Sildenafil reduces polyuria in rats with lithium-induced NDI. *Am J Physiol Renal Physiol*. 2012;302:F216–F225.
79. Upadhyay A, Jaber BL, Madias NE. Incidence and prevalence of hyponatremia. *Am J Med*. 2006;119:S30–S35.
80. Corona G, Giuliani C, Parenti G, et al. Moderate hyponatremia is associated with increased risk of mortality: evidence from a meta-analysis. *PloS One*. 2013;8:e80451.
81. Zilberberg MD, Exuzides A, Spalding J, et al. Epidemiology, clinical and economic outcomes of admission hyponatremia among hospitalized patients. *Curr Med Res Opin*. 2008;24:1601–1608.
82. Renneboog B, Musch W, Vandemergel X, et al. Mild chronic hyponatremia is associated with falls, unsteadiness, and attention deficits. *Am J Med*. 2006;119:71. e1–e8.
83. Fehlberg EA, Lucero RJ, Weaver MT, et al. Associations between hyponatraemia, volume depletion and the risk of falls in US hospitalised patients: a case-control study. *BMJ Open*. 2017;7:e017045.
84. Kinsella S, Moran S, Sullivan MO, et al. Hyponatremia independent of osteoporosis is associated with fracture occurrence. *CJASN*. 2010;5:275–280.
85. Verbalis JG, Barsony J, Sugimura Y, et al. Hyponatremia-induced osteoporosis. *J Bone Mineral Res*. 2010;25:554–563.
86. Sturdik I, Adamcova M, Kollerova J, et al. Hyponatraemia is an independent predictor of in-hospital mortality. *Eur J Int Med*. 2014;25:379–382.
87. Waikar SS, Mount DB, Curhan GC. Mortality after hospitalization with mild, moderate, and severe hyponatremia. *Am J Med*. 2009;122:857–865.
88. Chawla A, Sterns RH, Nigwekar SU, et al. Mortality and serum sodium: do patients die from or with hyponatremia? *CJASN*. 2011;6:960–965.
89. Corona G, Giuliani C, Verbalis JG, et al. Hyponatremia improvement is associated with a reduced risk of mortality: evidence from a meta-analysis. *PloS One*. 2015;10:e0124105.
90. Garrahy A, Cuesta M, Murphy B, et al. Active management of severe hyponatraemia is associated with improved mortality. *Eur J Endocrinol*. 2021;184:9–17.
91. Cuesta M, Garrahy A, Slattery D, et al. Mortality rates are lower in SIAD, than in hypervolaemic or hypovolaemic hyponatraemia: Results of a prospective observational study. *Clin Endocrinol*. 2017;87:400–406.
92. Oelkers W. Adrenal insufficiency. *N Engl J Med*. 1996;335:1206–1212.
93. Schwartz WB, Bennett W, Curelop S, et al. A syndrome of renal sodium loss and hyponatremia probably resulting from inappropriate secretion of antidiuretic hormone. *Am J Med*. 1957;23:529–542.
94. Zerbe R, Stropes L, Robertson G. Vasopressin function in the syndrome of inappropriate antidiuresis. *Ann Rev Med*. 1980;31:315–327.
95. Sherlock M, O'Sullivan E, Agha A, et al. The incidence and pathophysiology of hyponatraemia after subarachnoid haemorrhage. *Clin Endocrinol*. 2006;64:250–254.
96. Hannon MJ, Behan LA, O'Brien MM, et al. Hyponatremia following mild/moderate subarachnoid hemorrhage is due to SIAD and glucocorticoid deficiency and not cerebral salt wasting. *J Clin Endocrinol Metab*. 2014;99:291–298.
97. Matsumoto H. Hyponatremia associated with selective serotonin reuptake inhibitors. *Intern Med*. 2005;44:173–174.
98. Miehle K, Paschke R, Koch CA. Citalopram therapy as a risk factor for symptomatic hyponatremia caused by the syndrome of inappropriate secretion of antidiuretic hormone (SIADH): a case report. *Pharmacopsychiatry*. 2005;38:181–182.
99. Bouman WP, Pinner G, Johnson H. Incidence of selective serotonin reuptake inhibitor (SSRI) induced hyponatraemia due to the syndrome of inappropriate antidiuretic hormone (SIADH) secretion in the elderly. *Int J Geriatr Psychiatry*. 1998;13:12–15.
100. Budisavljevic MN, Stewart L, Sahn SA, et al. Hyponatremia associated with 3,4-methylenedioxymethylamphetamine ("Ecstasy") abuse. *Am J Med Sci*. 2003;326:89–93.
101. Hew-Butler T, Almond C, Ayus JC, Dugas J, et al. Consensus statement of the 1st International Exercise-Associated Hyponatremia Consensus Development Conference, Cape Town, South Africa 2005. *Clin J Sport Med*. 2005;15:208–213.
102. Almond CS, Shin AY, Fortescue EB, et al. Hyponatremia among runners in the Boston Marathon. *N Engl J Med*. 2005;352:1550–1556.
103. George JM, Capen CC, Phillips AS. Biosynthesis of vasopressin in vitro and ultrastructure of a bronchogenic carcinoma. Patient with the syndrome of inappropriate secretion of antidiuretic hormone. *J Clin Invest*. 1972;51:141–148.
104. Gross AJ, Steinberg SM, Reilly JG, et al. Atrial natriuretic factor and arginine vasopressin production in tumor cell lines from patients with lung cancer and their relationship to serum sodium. *Cancer res*. 1993;53:67–74.
105. Smith D, Moore K, Tormey W, et al. Downward resetting of the osmotic threshold for thirst in patients with SIADH. *Am J Physiol Endocrinol Metab*. 2004;287:E1019–E1023.
106. Baylis PH. The syndrome of inappropriate antidiuretic hormone secretion. *Int J Biochem Cell Biol*. 2003;35:1495–1499.
107. Feldman BJ, Rosenthal SM, Vargas GA, et al. Nephrogenic syndrome of inappropriate antidiuresis. *N Engl J Med*. 2005;352:1884–1890.
108. Miyado M, Fukami M, Takada S, et al. Germline-derived gain-of-function variants of Gsα coding GNAS gene identified in nephrogenic syndrome of inappropriate antidiuresis. *J Am Soc Nephrol*. 2019;30:877–889.
109. Ishikawa S, Saito T, Kasono K. Pathological role of aquaporin-2 in impaired water excretion and hyponatremia. *J Neuroendocrinol*. 2004;16:293–296.
110. Saito T, Higashiyama M, Nagasaka S, et al. Role of aquaporin-2 gene expression in hyponatremic rats with chronic vasopressin-induced antidiuresis. *Kidney Int*. 2001;60:1266–1276.
111. Ishikawa SE, Schrier RW. Pathophysiological roles of arginine vasopressin and aquaporin-2 in impaired water excretion. *Clin Endocrinol (Oxf)*. 2003;58:1–17.
112. Verbalis JG, Murase T, Ecelbarger CA, et al. Studies of renal aquaporin-2 expression during renal escape from vasopressin-induced antidiuresis. *Adv Exp Med Biol*. 1998;449:395–406.
113. Levinsky NG, Davidson DG, Berliner RW. Changes in urine concentration during prolonged administration of vasopressin and water. *Am J Physiol*. 1959;196:451–456.
114. Tzoulis P, Bouloux PM. Inpatient hyponatraemia: adequacy of investigation and prevalence of endocrine causes. *Clin Med (London, England)*. 2015;15:20–24.
115. Diederich S, Franzen NF, Bahr V, et al. Severe hyponatremia due to hypopituitarism with adrenal insufficiency: report on 28 cases. *Eur J Endocrinol*. 2003;148:609–617.
116. Chen YC, Cadnapaphornchai MA, Summer SN, et al. Molecular mechanisms of impaired urinary concentrating ability in glucocorticoid-deficient rats. *J Am Soc Nephrol*. 2005;16:2864–2871.
117. Olchovsky D, Ezra D, Vered I, et al. Symptomatic hyponatremia as a presenting sign of hypothalamic-pituitary disease: a syndrome of inappropriate secretion of antidiuretic hormone (SIADH)-like glucocorticosteroid responsive condition. *J Endocrinol Invest*. 2005;28:151–156.
118. Oelkers W. Hyponatremia and inappropriate secretion of vasopressin (antidiuretic hormone) in patients with hypopituitarism. *N Engl J Med*. 1989;321:492–496.
119. Fox JL, Falik JL, Shalhoub RJ. Neurosurgical hyponatremia: the role of inappropriate antidiuresis. *J Neurosurg*. 1971;34:506–514.
120. Hannon MJ, Finucane FM, Sherlock M, et al. Clinical review: disorders of water homeostasis in neurosurgical patients. *J Clin Endocrinol Metab*. 2012;97:1423–1433.
121. Agha A, Rogers B, Sherlock M, et al. Anterior pituitary dysfunction in survivors of traumatic brain injury. *J Clin Endocrinol Metab*. 2004;89:4929–4936.
122. Sari R, Sevinc A. Life-threatening hyponatremia due to cessation of L-thyroxine. *J Natl Med Assoc*. 2003;95:991–994.
123. Taskapan C, Sahin I, Taskapan H, et al. Possible malignant neuroleptic syndrome that associated with hypothyroidism. *Prog Neuropsychopharmacol Biol Psychiatry*. 2005;29:745–748.
124. Chen YC, Cadnapaphornchai MA, Yang J, et al. Nonosmotic release of vasopressin and renal aquaporins in impaired urinary dilution in hypothyroidism. *Am J Physiol Renal Physiol*. 2005;289:F672–F678.
125. Cuesta M, Garrahy A, Slattery D, et al. The contribution of undiagnosed adrenal insufficiency to euvolaemic hyponatraemia: results of a large prospective single-centre study. *Clin Endocrinol*. 2016;85:836–844.

126. Brinkkoetter PT, Grundmann F, Ghassabeh PJ, et al. Impact of resolution of hyponatremia on neurocognitive and motor performance in geriatric patients. *Sci Rep*. 2019;9:12526.
127. Verbalis JG, Goldsmith SR, Greenberg A, et al. Diagnosis, evaluation, and treatment of hyponatremia: expert panel recommendations. *Am J Med*. 2013;126:S1–S42.
128. Arieff AI, Llach F, Massry SG. Neurological manifestations and morbidity of hyponatremia: correlation with brain water and electrolytes. *Medicine*. 1976;55:121–129.
129. Verbalis JG. Disorders of body water homeostasis. *Best Pract Res Clin Endocrinol Metab*. 2003;17:471–503.
130. Spasovski G, Vanholder R, Allolio B, et al. Clinical practice guideline on diagnosis and treatment of hyponatraemia. *Eur J Endocrinol*. 2014;170:G1–G47.
131. Renneboog B, Musch W, Vandemergel X, et al. Mild chronic hyponatremia is associated with falls, unsteadiness, and attention deficits. *Am J Med*. 2006;119:71. e1–e8.
132. Yeates KE, Singer M, Morton AR. Salt and water: a simple approach to hyponatremia. *CMAJ*. 2004;170:365–369.
133. Garrahy A, Galloway I, Hannon AM, et al. Fluid restriction therapy for chronic SIAD; results of a prospective randomized controlled trial. *J Clin Endocrinol Metab*. 2020;105.
134. Cuesta M, Ortola A, Garrahy A, et al. Predictors of failure to respond to fluid restriction in SIAD in clinical practice; time to re-evaluate clinical guidelines? *QJM: Monthly J Assoc Phys.* 2017;110:489–492.
135. Winzeler B, Lengsfeld S, Nigro N, et al. Predictors of nonresponse to fluid restriction in hyponatraemia due to the syndrome of inappropriate antidiuresis. *J Intern Med*. 2016;280:609–617.
136. Miell J, Dhanjal P, Jamookeeah C. Evidence for the use of demeclocycline in the treatment of hyponatraemia secondary to SIADH: a systematic review. *Int J Clin Pract*. 2015;69:1396–1417.
137. Soupart A, Coffernils M, Couturier B, et al. Efficacy and tolerance of urea compared with vaptans for long-term treatment of patients with SIADH. *CJASN*. 2012;7:742–747.
138. Decaux G, Andres C, Gankam Kengne F, et al. Treatment of euvolemic hyponatremia in the intensive care unit by urea. *Crit care (London, England)*. 2010;14:R184.
139. Rondon-Berrios H, Tandukar S, Mor MK, et al. Urea for the treatment of hyponatremia. *CJASN*. 2018;13:1627–1632.
140. Krisanapan P, Vongsanim S, Pin-On P, et al. Efficacy of furosemide, oral sodium chloride, and fluid restriction for treatment of syndrome of inappropriate antidiuresis (SIAD): an open-label randomized controlled study (the EFFUSE-FLUID Trial). *Am J Kidney Dis*. 2020;76:203–212.
141. Verbalis JG. Vasopressin V2 receptor antagonists. *J Mol Endocrinol*. 2002;29:1–9.
142. Serradeil-Le Gal C, Wagnon J, Valette G, et al. Nonpeptide vasopressin receptor antagonists: development of selective and orally active V1a, V2 and V1b receptor ligands. *Prog Brain Res*. 2002;139:197–210.
143. Schrier RW, Gross P, Gheorghiade M, et al. Tolvaptan, a selective oral vasopressin V2-receptor antagonist, for hyponatremia. *N Engl J Med*. 2006;355:2099–2112.
144. Berl T, Quittnat-Pelletier F, Verbalis JG, et al. Oral tolvaptan is safe and effective in chronic hyponatremia. *JASN*. 2010;21:705–712.
145. Verbalis JG, Adler S, Schrier RW, et al. Efficacy and safety of oral tolvaptan therapy in patients with the syndrome of inappropriate antidiuretic hormone secretion. *Eur J Endocrinol*. 2011;164:725–732.
146. Higashihara E, Torres VE, Chapman AB, et al. Tolvaptan in autosomal dominant polycystic kidney disease: three years' experience. *Clin J Am Soc Nephrol*. 2011;6:2499–2507.
147. Greenberg A, Verbalis JG, Amin AN, et al. Current treatment practice and outcomes. Report of the hyponatremia registry. *Kidney Int*. 2015;88:167–177.
148. Refardt J, Imber C, Sailer CO, et al. A randomized trial of empagliflozin to increase plasma sodium levels in patients with the syndrome of inappropriate antidiuresis. *JASN*. 2020;31:615–624.
149. De Mota N, Reaux-Le Goazigo A, El Messari S, et al. Apelin, a potent diuretic neuropeptide counteracting vasopressin actions through inhibition of vasopressin neuron activity and vasopressin release. *Proc Natl Acad Sci U S A*. 2004;101:10464–10469.
150. Blanchard A, Steichen O, De Mota N, et al. An abnormal apelin/vasopressin balance may contribute to water retention in patients with the syndrome of inappropriate antidiuretic hormone (SIADH) and heart failure. *J Clin Endocrinol Metab*. 2013;98:2084–2089.
151. Boulkeroua C, Ayari H, Khalfaoui T, et al. Apelin-13 regulates vasopressin-induced aquaporin-2 expression and trafficking in kidney collecting duct cells. *Cell Physiol Biochem*. 2019;53:687–700.
152. Flahault A, Couvineau P, Alvear-Perez R, et al. Role of the vasopressin/apelin balance and potential use of metabolically stable apelin analogs in water metabolism disorders. *Front Endocrinol*. 2017;8:120.
153. Flahault A, Girault-Sotias PE, Keck M, et al. A metabolically stable apelin-17 analog decreases AVP-induced antidiuresis and improves hyponatremia. *Nat Comm*. 2021;12:305.
154. Arieff AI, Llach F, Massry SG. Neurological manifestations and morbidity of hyponatremia: correlation with brain water and electrolytes. *Medicine (Baltimore)*. 1976;55:121–129.
155. Ayus JC, Arieff AI. Chronic hyponatremic encephalopathy in postmenopausal women: association of therapies with morbidity and mortality. *JAMA*. 1999;281:2299–2304.
156. Gill G, Huda B, Boyd A, et al. Characteristics and mortality of severe hyponatraemia--a hospital-based study. *Clin Endocrinol*. 2006;65:246–249.
157. Cluitmans FH, Meinders AE. Management of severe hyponatremia: rapid or slow correction? *Am J Med*. 1990;88:161–166.
158. Koenig MA, Bryan M, Lewin 3rd JL, et al. Reversal of transtentorial herniation with hypertonic saline. *Neurology*. 2008;70:1023–1029.
159. Sterns RH, Nigwekar SU, Hix JK. The treatment of hyponatremia. *Sem Nephrol*. 2009;29:282–299.
160. Garrahy A, Dineen R, Hannon AM, et al. Continuous versus bolus infusion of hypertonic saline in the treatment of symptomatic hyponatremia due to SIAD. *J Clin Endocrinol Metab*. 2019;104:3595–3602.

13

Anatomy and Physiology of the Hypothalamus and Pituitary

Helen Clare Christian

OUTLINE

Anatomy of the Hypothalamus, 200
  Blood Supply of the Hypothalamus, 200
  The Median Eminence and Tanycytes, 200
Anatomy and Histology of the Pituitary, 202
  The Blood Supply of the Pituitary, 202
  The Pars Intermedia, 202
  The Pars Tuberalis, 202
  The Pars Distalis (Anterior Pituitary), 203
  Anterior Pituitary Cell Distribution and Organization in Networks, 203
  Coordinated Secretion: Relationship With the Pituitary Vasculature, 209
  Posterior Pituitary Gland (Neurohypophysis), 209
Conclusions and Future Directions, 210

ANATOMY OF THE HYPOTHALAMUS

The hypothalamus is located in the forebrain, positioned either side of the lower cerebral third ventricle. Its floor comprises the optic chiasm and tracts, pituitary stalk, mamillary bodies, and posterior perforated substance. Anteriorly, it borders the lamina terminalis and the anterior commissure, and posteriorly the midbrain. The upper superior surface of the hypothalamus is continuous with the thalamus and epithalamus, while its lateral sides border the zona incerta and internal capsule. Branches of the internal carotid enter and supply the ventral surface of the hypothalamus. Fig. 13.1 shows a sagittal view of the complex organization of the hypothalamus, its relationship to the pituitary gland, and its vasculature. The location of the hypothalamic nuclei, the hypothalamic neurohormones secreted, and their main functions are shown in Table 13.1. The median eminence is at the base of the hypothalamus and forms the interface at which the neuroendocrine parvocellular neurons make contact with a capillary plexus formed from superior hypophyseal branches of the internal carotid. The cell bodies of the parvocellular neurons are located in the paraventricular nucleus (corticotrophin-releasing hormone [CRH], thyrotropin-releasing hormone [TRH]), arcuate nucleus (dopamine, growth hormone–releasing hormone [GHRH]), medial basal hypothalamus (gonadotropin-releasing hormone [GnRH]), and periventricular zone (somatostatin). Magnocellular neurons producing and secreting oxytocin and vasopressin are located in the supraoptic nucleus and paraventricular nucleus and extend to the posterior pituitary.

Blood Supply of the Hypothalamus

The arterial supply to the hypothalamus arises from small perforating vessels that arise from different parts of the circle of Willis and pass through the anterior and posterior perforated substances. In addition, the superior hypophyseal arteries, which are branches of the internal carotid arteries, supply the optic chiasm and adjacent parts of the hypothalamus. The superior hypophysial arteries form a capillary network in the median eminence that recombine to form portal vessels that pass down the pituitary stalk, where they divide into another capillary network perfusing the anterior pituitary, which then drain into venous vessels via the cavernous sinus and, in turn, the jugular systemic circulation. The direction of blood flow in the portal vessels is thought to be predominantly hypothalamus to pituitary. The concept of neurohumoral control by the hypothalamus of the anterior pituitary by the secretion of releasing factors into portal vessels at the median eminence was championed by Geoffrey Harris, and his seminal work has been elegantly reviewed in the literature by those who worked with him in Oxford.[1,2]

The Median Eminence and Tanycytes

The median eminence is the functional interface between the hypothalamus and the pituitary, and comprises hypothalamic axons and nerve terminals, a dense network of capillaries and glial cells. It is organized into three zones: the innermost ependymal layer that lines the floor of the third ventricle, the internal zone, and the external zone. The internal zone lies beneath the ependymal layer and contains subependymal cells, astrocytes, and parvicellular and magnocellular fibers of the hypothalamo-hypophysial tract *en route* to the posterior pituitary. The external zone is the location where hypothalamic regulatory factors are released from parvocellular nerve terminals into the capillaries that will form the pituitary portal vessels. Ependymal cells in the ependymal layer form an interface between the cerebrospinal fluid (CSF) and the extracellular space of neural tissue of the median eminence, and are lined with microvilli, indicating a role in controlled transport of molecules, and complexes of desmosomes and tight junctions at their apical surface that prevent intercellular diffusion of larger molecular weight molecules. The ependymal layer also contains the soma of tanycytes, which send a short apical process to the ventricle surface and a long basal process across the three layers to the external zone of the median eminence. Here the tanycytes are in close association with parvocellular nerve terminals and fenestrated capillary beds that will form the pituitary stalk portal vessels. Tanycytes can be distinguished by the prominent cytoplasmic intermediate filaments along their length by electron microscopy (Fig. 13.2). Tanycytes also extend cytoplasmic processes to hypothalamic nuclei concerned with energy balance, such as the arcuate nucleus, and are considered important for their role in sensing energy and endocrine homeostasis through their strategic

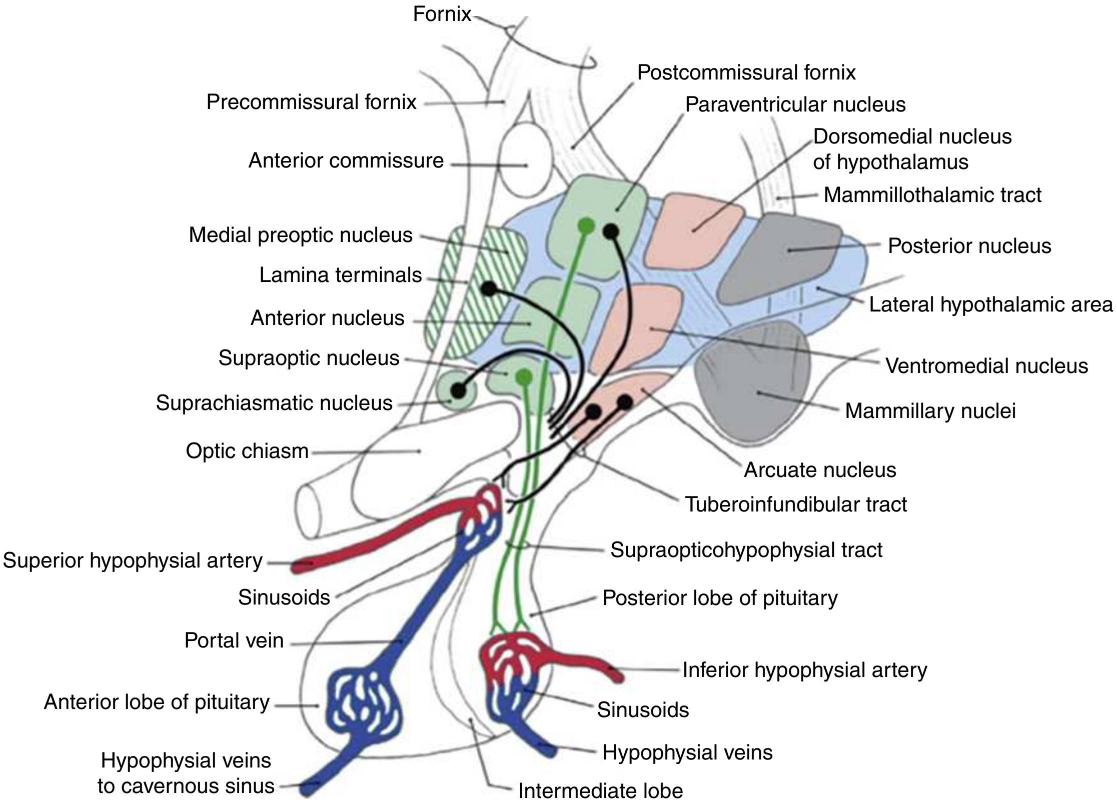

Fornix
Precommissural fornix
Anterior commissure
Medial preoptic nucleus
Lamina terminals
Anterior nucleus
Supraoptic nucleus
Suprachiasmatic nucleus
Optic chiasm
Superior hypophysial artery
Sinusoids
Portal vein
Anterior lobe of pituitary
Hypophysial veins
to cavernous sinus
Postcommissural fornix
Paraventricular nucleus
Dorsomedial nucleus
of hypothalamus
Mammillothalamic tract
Posterior nucleus
Lateral hypothalamic area
Ventromedial nucleus
Mammillary nuclei
Arcuate nucleus
Tuberoinfundibular tract
Supraopticohypophysial tract
Posterior lobe of pituitary
Inferior hypophysial artery
Sinusoids
Hypophysial veins
Intermediate lobe

Fig. 13.1 Sagittal view of the hypothalamus demonstrating the preoptic, supraoptic, tuberal, and mammillary groups of medial hypothalamic nuclei. (Reprinted with permission from Haines, D. *Fundamental Neuroscience for Basic and Clinical Applications*. 5th ed. San Diego: Elsevier; 2018.)

TABLE 13.1 Hypothalamic Nuclei and their Functions

Nucleus	Location	Major hormones/function
Supraoptic (SON)	Anterolateral, above the optic nerves	AVP; osmoregulation Oxytocin: regulation of milk ejection in lactation and uterine contractions
Paraventricular (PVN)	Dorsal anterior, adjacent to the third ventricle	Magnocellular PVN: AVP and oxytocin, same functions as SON above
Arcuate	Medial basal, close to the third ventricle	GHRH: stimulation of growth hormone GnRH: regulation of gonadotrophins FSH and LH (primates) Dopamine: inhibits lactotroph PRL secretion Somatostatin: inhibition of GHRH secretion Regulation of appetite: neuropeptide Y, alpha-MSH, cocaine- and amphetamine-related transcript
Periventricular	Anteroventral	Somatostatin: inhibition of pituitary somatotroph GH secretion
Ventromedial	Ventromedial	GHRH: stimulation of GH
Dorsomedial	Dorsomedial	Somatostatin: inhibition of GHRH release; appetite satiety center of the hypothalamus
Lateral hypothalamus	Lateral	Hunger center of the hypothalamus; orexins, melanin-concentrating hormone
Preoptic area	Preoptic area	GnRH in rodents (few in primates)
Anterior hypothalamus	Anterior	Thermoregulation cooling center and regulation of thirst
Posterior hypothalamus	Posterior	Thermoregulation heating center
Suprachiasmatic	Above the optic chiasm, anteroventral	Circadian rhythm regulator, AVP and vasoactive intestinal peptide, neurons project to the PVN

AVP, Arginine vasopression; *GHRH,* growth hormone–releasing hormone; *GnRH,* gonadatropin-releasing hormone; *FSH,* follicle-stimulating hormone; *LH,* luteinizing hormone; *PRL,* prolactin; *MSH,* α-melanocyte–stimulating hormone; *GH,* growth hormone.

location bridging hypothalamic neurons, CSF, and the blood, signaling the physiological and metabolic status of the animal.[3] Seasonal changes in neural-tanycyte interactions in the median eminence have been reported in seasonal breeding animals.[4] The extensive tanycyte network is a key component of the median eminence neurovascular interface, such that hypothalamic neurosecretory nerve terminals are distributed within the tanycyte network. In long photoperiod conditions (nonbreeding season), a marked tanycyte barrier between GnRH terminals and the basal lamina of the median eminence has been demonstrated, constraining GnRH release.[4] In contrast, in the short

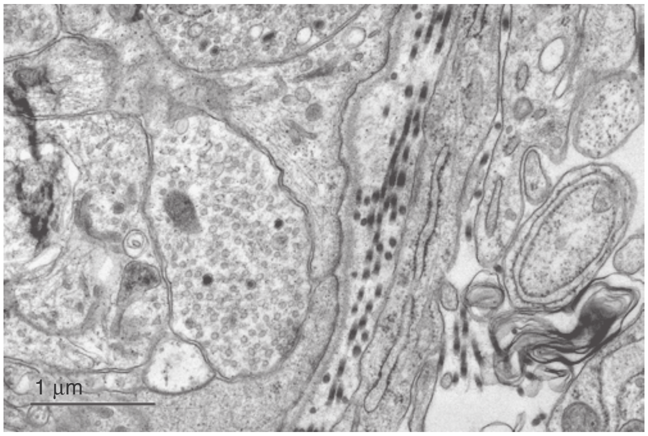


Fig. 13.2 Electron micrograph of the median eminence (ovine) illustrating a tanycyte end foot processes surrounding nerve terminals adjacent to the perivascular space containing collagen fibrils. The tanycytes can be distinguished by their intermediate filaments and the nerve terminals by secretory granules and vesicles.

photoperiod, GnRH nerve endings extend through to the basal lamina and perivascular space of the median eminence capillaries, and the tanycyte barrier is more dispersed.[5]

ANATOMY AND HISTOLOGY OF THE PITUITARY

The pituitary gland has a bilaterally symmetrical bean-shaped structure, approximately 13 mm in transverse, 8 mm in anterior-posterior diameter, and 6 mm deep. In adults it weighs on average 0.6 g (range 0.4–0.8 g), is typically smaller in men than in women, and enlarges in pregnancy.

The pituitary lies within a depression of the sphenoid bone at the base of the skull termed the pituitary fossa or sella turcica, which is lined by dura mater and covered by an extension of the dura mater called the diaphragma sellae. The pituitary is connected to the brain by the pituitary stalk (infundibulum), which penetrates through a small opening in the diaphragma sellae from the base of the hypothalamus above. Below the pituitary fossa is the sphenoid air sinus, and laterally on each side is the cavernous sinus and internal carotid artery. Superior and anterior to the pituitary and the pituitary stalk is the optic chiasm, which because of its close proximity can be compressed by growth of the pituitary, for example, adenomas resulting in "tunnel vision." The pituitary consists of two parts: the adenohypophysis (or anterior pituitary) and neurohypophysis (or posterior pituitary) (see also Chapters 4 and 15). The adenohypophysis comprises the pars distalis, pars intermedia, and pars tuberalis. The pituitary stalk comprises fenestrated portal vessels from the median eminence that supply the anterior pituitary and the large unmyelinated axons of magnocellular neurons originating in the supraoptic and paraventricular nuclei of the hypothalamus that terminate in the posterior pituitary.

The Blood Supply of the Pituitary

The anterior pituitary gland has a rich and intricate vasculature,[6] as beautifully revealed in scanning electron microscopy imaging of vascular casts of the rhesus monkey.[7] It is the most richly vascularized of all mammalian tissues, receiving approximately 0.8 mL/g/min of blood from the hypothalamo-hypophyseal portal system. Portal vessels pass down the pituitary stalk into the pars distalis and adjacent pars tuberalis and branch into a secondary plexus of fenestrated capillaries that facilitate drainage of endocrine secretions into the systemic circulation and the circulation of regulatory hormones from the hypothalamus and the periphery (thyroid, gonads, adrenal cortex). The capillaries comprise endothelial cells and basement membrane, with underlying mural cells termed pericytes. The capillaries reform themselves into efferent lateral hypophysial veins that drain into the cavernous sinus and subsequently the internal jugular venous system. The capillaries are not sinusoids, because there are no large spaces in their walls. The perivascular space between the vascular border and the secretory cells can be relatively large, and contains fibers of connective tissue such as collagen. Hormones released into the extracellular space diffuse into capillaries through fenestrations in the endothelial lining (Fig. 13.3). The posterior lobe receives its blood supply from the inferior hypophyseal branches of the internal carotid artery. Venous drainage of the posterior pituitary is to the cavernous sinus and circular sinus, and in turn the internal jugular vein. The intermediate lobe, in contrast, has little vasculature. For a more comprehensive review of the pituitary arterial supply and its clinical relevance, see Cironi et al.[6]

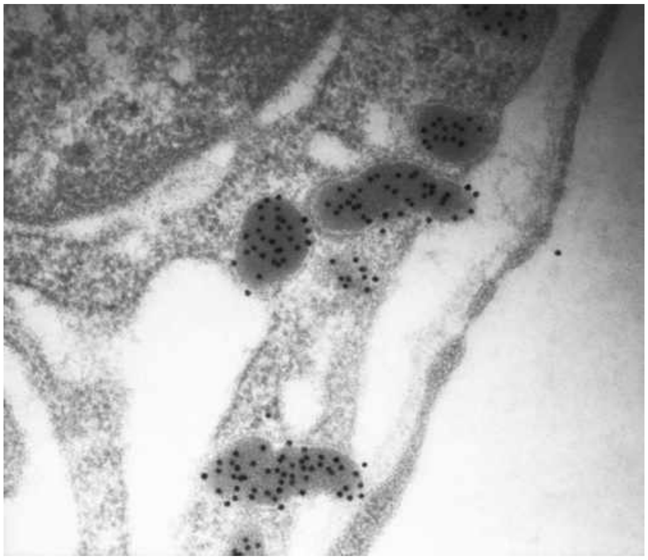


Fig. 13.3 High-power electron micrograph illustrating the perivascular space between a type 1 lactotroph and the fenestrated capillary endothelium (rat). Secretory granules are characteristically irregular in shape and immunogold-labeled for prolactin. Particles are 15 nm.

The Pars Intermedia

The pars intermedia (or intermediate lobe) of the anterior pituitary contains mostly α-melanocyte–stimulating hormone (MSH/proopiomelanocortin [POMC])-expressing cells surrounding a cleft between the anterior and posterior lobes known as the pituitary cleft or Rathke's cleft. In adult humans this is a thin layer of cells that often merge with the anterior pituitary, but is more prominent as a distinct zone of uniform cells in other mammals. The function of the mammalian intermediate lobe POMC-cells is not clear, but in lower vertebrates, amphibia, and fish, secretion of MSH, under the control of multiple hypothalamic nuclei, is involved in the control of pigmentation.[8] It is thought that the intermediate lobe is likely a developmental remnant of Rathke's pouch. The neural control and function of the intermediate lobe across vertebrates has been comprehensively reviewed by Low.[8]

The Pars Tuberalis

The pars tuberalis (PT) is composed of mainly follicular cells (similar in morphology to folliculo-stellate cells) and thyroid-stimulating hormone (TSH)-positive secretory endocrine cells (with sparse small secretory granules) that form a sheath surrounding the pituitary stalk pars nervosa neurons and hypophyseal stalk portal vessels extending

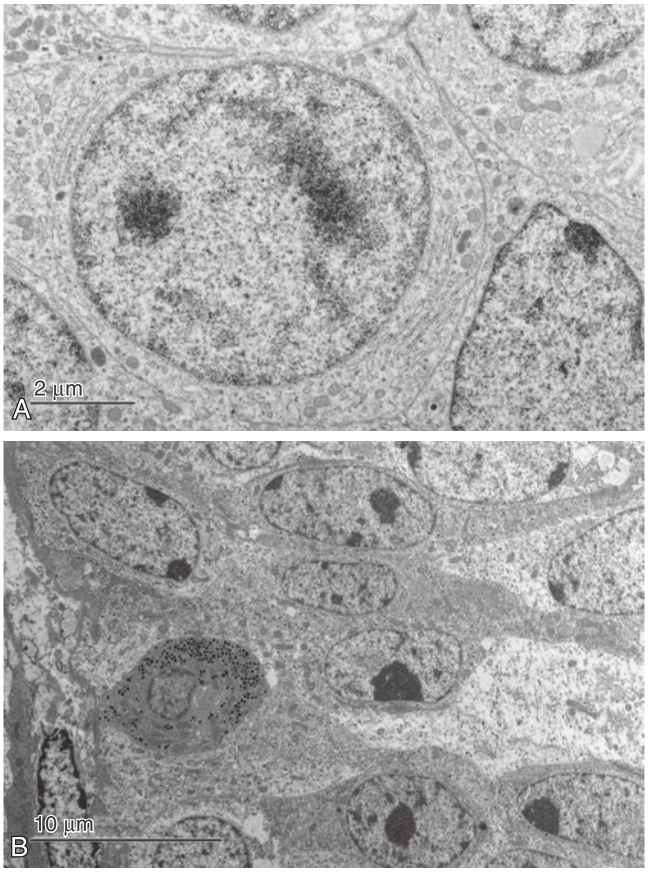


Fig. 13.4 Electron micrographs of a pars tuberalis thyrotroph within the ovine pars tuberalis (**A**)—note the very small sparse granules and extended rough endoplasmic reticulum—and a single pars tuberalis gonadotroph among pars tuberalis thyrotrophs (**B**).

up to the median eminence. A few gonadotrophs and somatotrophs can also be found in the tuberalis, usually close to the transition zone with the anterior pituitary (Fig. 13.4). The PT TSH-positive cells strongly express melatonin MT1 receptors. Melatonin, secreted in the dark (long photoperiod), controls seasonal fertility in sheep and other mammals via stimulation of the PT thyrotrophs to release factors, unidentified and known as tuberalins but proposed to be TSH. There may be retrograde flow in the portal vessels between the pituitary and hypothalamus, which would allow a means of circulating TSH to act on hypothalamic ependymal cells that, in turn, stimulate GnRH neurons and fertility in times of the year of short day-length. It has been proposed that the PT is the site of a circannual clock,[9] and that the PT thyrotrophs act as a seasonal calendar cell, driving hypothalamic and pituitary endocrine circuits.[10] In sheep, long photoperiods induce the circadian transcription factor BMAL2 in the PT, which activates summer physiology through the eyes absent/TSH pathway.[10]

The Pars Distalis (Anterior Pituitary)

The pars distalis forms the bulk of the pituitary gland (75%). Standard light two-dimensional microscopy methods show that the pars distalis consists of clusters (also known as acini) of secretory endocrine cells and nongranular cells surrounded by reticulin fibers and an extensive capillary microvasculature. The pars distalis contains five main secretory cell types, each named according to the hormone it principally releases: somatotrophs (growth hormone [GH]), lactotrophs (prolactin [PRL]), corticotrophs (adrenocorticotropic hormone [ACTH]), thyrotrophs (TSH), and gonadotrophs (luteinizing hormone [LH] and follicle-stimulating hormone [FSH]). The distinct "classical" secretory cell types can be identified immunocytochemically or on a basis of their ultrastructural characteristics, which are detailed in Table 13.1. Historically, histologists classified anterior pituitary cells as basophils, acidophils, or chromophobes on a basis of differential histochemical labeling with acidic and basic dyes such as periodic acid–Schiff orange G stain. These histological methods originally revealed the heterogeneity of secretory cells within the anterior lobe. In a normal adult human adenohypophysis, approximately 10% of endocrine cells are basophils, 40% acidophils, and 50% chromophobes. Experimentally, however, this classification has proved confusing, as the staining method is not able to distinguish hormones within specific cell types and is variable across species, and little is known of the basis of the chemical reactions between the dyes and components of the pituitary cells. Therefore, histochemical methods have been largely replaced by hormone antibody immunolabeling methods. There is no nervous innervation of the secretory cells of the anterior pituitary. Each of the secretory cell types has been well studied by transmission electron microscopy and can be distinguished by its secretory granule morphology, size, distribution, and number. Table 13.2 shows the ultrastructural characteristics associated with each of the secretory anterior pituitary cell types and folliculo-stellate cells. The physiology and control of each cell type is summarized later. The hypothalamus provides the major inputs for the control of pituitary hormone secretion, and this is balanced by feedback regulation exerted by circulating hormones signaling the status of peripheral endocrine glands. Superimposed on the control by blood-borne hormones is the intrapituitary, local paracrine and autocrine regulation by more than 100 compounds secreted locally by secretory and folliculo-stellate cells, detailed in Carl Denef's seminal review.[11]

Anterior Pituitary Cell Distribution and Organization in Networks

Conventional histological methods give the impression that anterior pituitary cells are randomly distributed within the acini across the gland. However, there is evidence in different species that the secretory cell types are concentrated in specific areas and organized in three-dimensional (3D) interconnected networks. In the human pituitary, corticotrophs are found to be more concentrated in the medial regions near to the marginal zone, somatotrophs and lactotrophs are found more in the lateral regions, thyrotroph clusters are found more anteriorly, and gonadotrophs are scattered throughout (Fig. 13.5). It is not known what the origin of this organization may be, but it may relate to sequences of cell determination.[12]

Whole-tissue 3D imaging of the mouse anterior pituitary using two-photon microscopy has shown that, in 3D, pituitary cells of the same type ("homotypic") are organized in extensive structured networks[13] similar to those first shown originally for folliculo-stellate cells.[14] Cell networks have also been visualized in the teleost fish. These approaches rely on transgenic models expressing fluorophore-tagged cells, and it is not clear whether similar cell network organization is also the case in larger mammals and the human pituitary. However, modifications in the organization of pituitary cell networks have been shown to occur in different physiological states. The pituitary gland is regarded as highly plastic, and dynamically adjusts throughout life to changing endocrine status, and as such numbers of each cell type vary with puberty, pregnancy, lactation, stress, the menopause, and aging. In most cases it is not clear if changes in pituitary size of networks are the result of increased cell number via trans differentiation of a different cell type, proliferation of existing cells, or differentiation of adult pituitary stem/progenitor cells and/or increased cell size.[15] Although the difference in responsiveness comparing *in vivo* pituitary hormone release to isolated cells *in vitro* has been attributed to the importance of coordination of secretion through an intact cell network, it is also

TABLE 13.2 Ultrastructural Characteristics Associated With Each of the Secretory Anterior Pituitary Cell Types

Cell Type	Hormone released and major role	% Anterior lobe	Secretory granule size (nm)	Granule shape	Distribution of granules
Somatotrophs	GH Postnatal growth and metabolism	45–60	350	Spherical and electron dense	Large numbers, occupy a large proportion of the cytoplasm
Lactotrophs	PRL Lactation	10–15	200–350	Larger granules are irregular in shape, smaller granules are spherical	Throughout cytoplasm
Gonadotrophs	LH and FSH Control of gonads and reproductive function	20	200, 350	Smaller electron-dense granules and larger electron-lucent	Throughout cytoplasm
Corticotrophs	ACTH Stress response, control of adrenal cortex	6–8	100–200	Moderately electron-dense, spherical	Line perimeter of inner plasma membrane
Thyrotrophs	TSH Control of thyroid T3/T4 production	5–6	100–200	Moderately electron-dense, spherical	Throughout cytoplasm
Folliculostellate	Paracrine control, release of mediators e.g., IL6, annexin 1, follistatin	4	-	Cells are agranular	Cells are agranular

GH, Growth hormone; *PRL,* prolactin; *LH,* luteinizing hormone; *FSH,* follicle-stimulating hormone; *ACTH,* adrenocorticotropic hormone; *TSH,* thyroid-stimulating hormone; *T3*, triiodothyronine; *T4*, thyroxine; *IL,* interleukin.

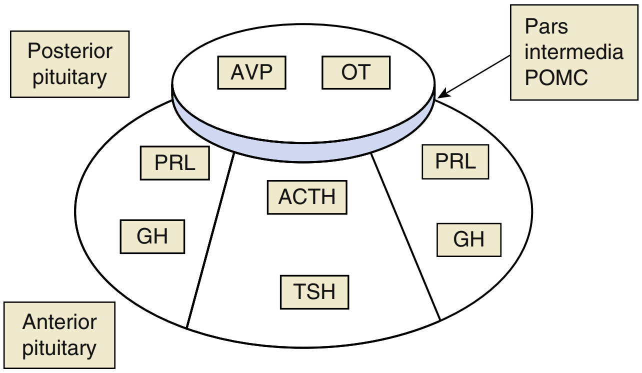


Fig. 13.5 Schematic illustrating the zonation of human anterior pituitary cells. *AVP*, Arginine vasopressin; *OT*, oxytocin; *PRL*, prolactin; *GH*, growth hormone; *ACTH*, adrenocorticotropic hormone; *TSH*, thyroid-stimulating hormone; *POMC,* proopiomelanocortin.

possible that proteolytic methods used in the preparation of single cells digest the cell surface receptors for stimulatory factors.

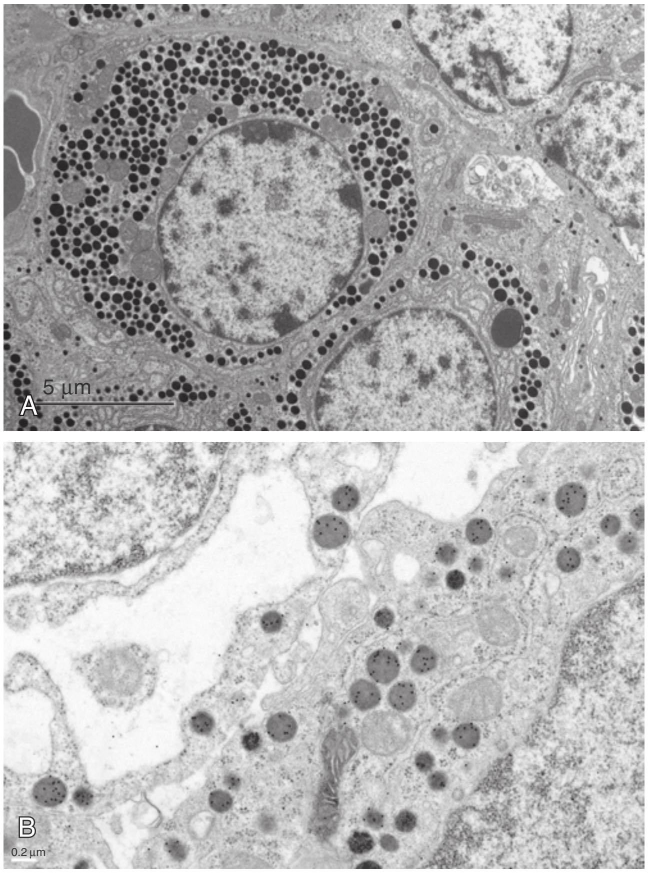


Fig. 13.6 A, Electron micrograph of a somatotroph (mouse); note the characteristically abundant large electron dense secretory granules. **B,** High-power micrograph of immunogold growth-hormone–labeled secretory granules.

Somatotrophs. Somatotrophs produce and secrete GH, a 191–amino acid polypeptide hormone. Somatotrophs are the most numerous cell type in the anterior pituitary and are readily identified by their ultrastructure of the cytoplasm filled with large-diameter electron-dense secretory granules (Fig. 13.6). Somatotrophs were the first of the pituitary cell types to be revealed as a 3D network extending throughout the pituitary by 3D *in vivo* imaging of GH–green fluorescent protein (GFP) mice. The somatotrophs are joined by adherens and gap junctions,[16] which are important for the rapid exchange of signals, such as cyclic adenosine monophosphate (cAMP) and calcium, between cells, and in turn coordinate synchronized secretory responses.[13] The pulsatile secretion of GH is controlled by GH-releasing hormone (GHRH) and somatostatin, which coordinate to generate the episodic circadian pattern of GH release with maximal release during early sleep. GHRH binds to G protein–coupled receptors (GPCRs) coupled to cAMP/protein kinase A/CREB signaling to stimulate exocytosis, whereas somatostatin, a potent inhibitor of GH secretion, binds to SST2 GPCRs that decrease cAMP production.[17] Long-acting analogs of somatostatin are used clinically in conditions of GH excess (see also Chapter 6). The gastrointestinal peptide ghrelin also directly stimulates somatotroph GH release by activation of the growth hormone secretagogue receptor

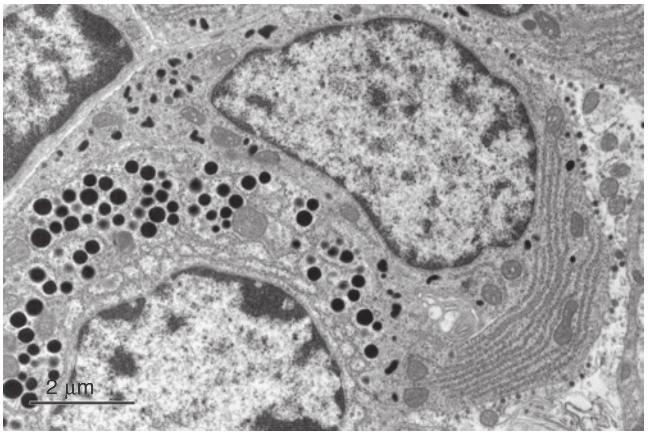


Fig. 13.7 Electron micrograph of a type 1 lactotroph (mouse); note the distinctive irregularly shaped secretory granules.

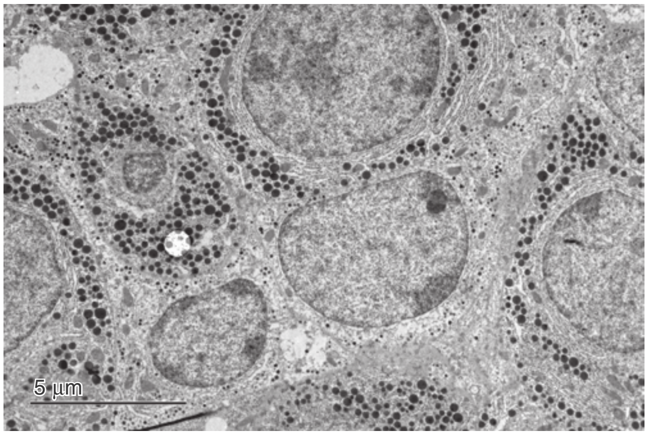


Fig. 13.8 Electron micrograph of a type II lactotroph (mouse).

GHS-R1a,[17] a GPCR that is highly expressed in somatotrophs coupled to the PLC/IP3/calcium signaling pathway. Ghrelin must be acylated by the enzyme ghrelin-O acyl transferase to be functionally active at GHS-R1a receptors. Elevated concentrations of GH in plasma exert "short-loop" negative feedback to inhibit GHRH neurons and stimulate somatostatin neurons, whereas liver-derived insulin-like growth factor (IGF)-1 exerts "long-loop" negative feedback via acting on somatotroph IGF tyrosine kinase receptors to inhibit GH mRNA transcription and GH secretion.[17] GH is important for body growth and metabolic homeostasis and is regulated by many other hormones and nutrients that influence energy expenditure at the hypothalamic and somatotroph level, including glucocorticoids, sex steroids, adiponectin, fatty acids, and leptin (see also Chapter 17). In general, GH concentrations rise in fasting and decline in obesity via the interplay of these regulatory molecules involved in the complex regulation of energy expenditure, and in different physiological contexts such as stress, exercise, and sleep. The circadian pattern of GH release diminishes with aging and is absent in postmenopausal women and men over 50 years.[18]

Lactotrophs. Lactotrophs produce PRL, a 198–amino acid polypeptide. Lactotrophs show a high degree of plasticity with respect to their number and size, particularly in females. Lactotroph number, and subsequent PRL secretion, increase during pregnancy to stimulate breast development in preparation for milk production at lactation. During pregnancy, the pituitary gland may increase in size by approximately 30% owing to lactotroph hyperplasia, which is sustained during lactation but rapidly declines by enhanced apoptosis of lactotrophs to their previous numbers after weaning. In adult rats and mice, there are sex differences in the number of lactotroph cells in the anterior pituitary. Female animals have more lactotroph cells than their male counterparts, and the mitotic activity of lactotroph cells is considerably higher in adult female animals and varies with the estrous cycle. The sex difference in the mouse becomes apparent between 21 days and 35 days of age, which corresponds to the period at which the female mouse ovary begins to secrete estrogens in increasing amounts. There is also an increase in lactotroph number at puberty associated with changes in cell morphology, such that, in adult rodents, the lactotroph population is heterogeneous based on the morphology of their PRL-containing secretory granules. Three types of lactotroph, have been described: type I, or "mature," lactotrophs are polygonal cells that contain irregularly-shaped, large secretory granules with a diameter of 300 to 700 nm (Fig. 13.7); type II, or "intermediate," lactotrophs are oval or polygonal cells that contain spherical granules with a diameter of 100 to 250 nm (Fig. 13.8); type III, or "immature," lactotrophs are oval or polygonal cells that contain small spherical granules with a diameter of 100 nm. Estrogen stimulates the development of the type I morphological subtype of lactotroph in the rat anterior pituitary, such that 80% of lactotrophs are type I in female rats. Estrogen effects on lactotroph cell growth are mediated via induction of several growth factors, such as IGF-1 and epidermal growth factor, but also by downregulation of dopaminergic control (see also Chapter 4).[18]

Lactotroph–lactotroph gap junctional contacts have been shown to underlie lactotroph network functions. Injection of a gap junction permeable dye into lactotrophs of lactating mice revealed passage of dye between lactotrophs by gap junction coupling that was blocked by gap junction blockers.[19] Augmented gap junction transfer of dye between lactotrophs was measured in primiparas and weaned dams compared with nulliparas, and electron microscopic identification of intercellular gap and adherens junctions revealed that the majority of intercellular junctions were between lactotrophs. An increase in homotypic lactotroph–lactotroph intercellular junctions in primiparas compared with nulliparas was observed, showing increased anatomical connectivity that paralleled an increase in the extent of dye transfer, representing functional connectivity. Furthermore, a decrease in junctions between lactotrophs and folliculo-stellate cells was measured in primiparas versus nulliparas, which may reflect a functional reduction in inhibitory mediator signaling via folliculo-stellate cell contacts. Long-lasting plasticity was found to be retained when females experienced further lactations, resulting in improved functional PRL secretion.

The coordination of lactotrophs into an interconnected network is also significant for the coordination of *PRL* gene transcription.[20,21] Studies combining real-time live-cell imaging of *PRL* gene activity in pituitary slices from adult hPRL-GFP rats, and mathematical modeling of the recorded expression patterns have determined how the spatial organization of lactotrophs affect the transcriptional dynamics of the *PRL* gene. Whereas in single lactotrophs the bursts of expression of the *PRL* gene are randomly timed, bursting patterns are coordinated with lactotroph neighbors. The coordination is dependent on gap junction connections and signaling. Mild trypsin treatment of slices reduced expression of the gap junction protein connexin 43 and adherens junction proteins, without tissue disaggregation, resulting in greater intracellular distances between lactotrophs and a decrease in the coordination of *PRL* transcription between closely located lactotrophs. Treatment of slices with gap junction inhibitors reduced the synchronization of transcription activity between neighboring lactotrophs. In developing pituitary glands, the low lactotroph density provides less opportunity for cell junction signaling, and reduced transcriptional coordination occurs between cells compared with in adults.[20] In the embryonic immature lactotroph network, lactotrophs are mainly isolated from each other, with only a small number of cell junctions detected and no transcriptional coordination. Overall, cellular communication between

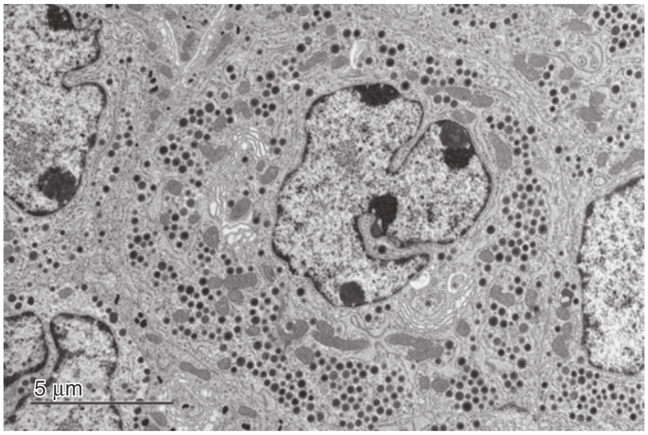


Fig. 13.9 Electron micrograph of a gonadotroph (mouse). The secretory granules have distinctive varying electron density.

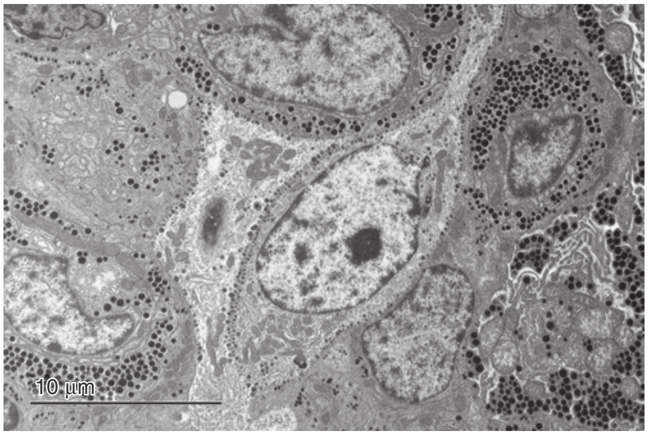


Fig. 13.10 Electron micrograph of two adjacent corticotrophs (mouse). The secretory granules are small and located along the plasma membrane.

lactotrophs is important for the coordination of *PRL* transcription activity, and it would be reasonable to predict that lactotroph transcriptional coordination would be increased during pregnancy and lactation because of the enlargement of the lactotroph cell population.

PRL secretion is pulsatile and episodic, and concentrations rise approximately 60 minutes after sleep onset, with peak plasma concentrations usually in the early hours of the morning at 4:00 AM to 7:00 AM. The tonic hypothalamic control of prolactin is mainly inhibitory, by dopamine[22] acting at the D2R GPCR, which signals to reduce cAMP levels, and therefore prolactin granule exocytosis. Dopamine inhibits PRL secretion from all lactotroph subtypes, but the PRL-releasing factors TRH and vasoactive intestinal peptide stimulate PRL release from type II and III lactotrophs.[23] Although the principal role of PRL is lactation, it has been implicated in a number of functions regarding metabolic homeostasis and roles in the skeleton, in the skin, and parental behaviors in males and females.[24]

Gonadotrophs. The gonadotrophins LH and FSH are synthesized within a single cell type, the gonadotroph. The gonadotrophins consist of a common alpha subunit and a specific beta subunit responsible for conferring biological activity on the heterodimer. All three subunits are encoded by separate genes localized on different chromosomes. Different secretory granule populations have been identified in gonadotrophs of the rat, mouse, and sheep, namely small LH- and secretogranin (SgII)-positive granules; large FSH- and chromogranin A (CgA)-positive granules; and intermediate-sized granules containing an electron-dense LH- and SgII-positive core and an electron-lucent CgA-positive outer region. Gonadotrophs can be distinguished from other anterior pituitary cell types by the variety of secretory granules of different sizes and electron density (Fig. 13.9). How this difference in secretory granule storage relates to different patterns or timing of hormone release is not known. Synthesis and secretion of the gonadotrophins is dependent on the decapeptide GnRH, which is produced within specific hypothalamic neurons. The hypothalamo-pituitary-gonadal (HPG) axis develops embryonically but, apart from short periods of activation immediately after birth, remains suppressed through a combination of gonadal and nongonadal mechanisms until puberty onset.[25] GnRH has long been regarded as the master hormone of the HPG axis, as, in the absence of this, hormone synthesis of LH and FSH is reduced such that mice carrying a mutation in this gene do not reach sexual maturity, but remain prepubertal throughout life (see also Chapters 102 and 103).[4] However, the discovery of kisspeptin stimulation of GnRH via KISS1 receptors on GnRH-expressing neurons revealed an additional level of control to the HPG axis.[25]

Corticotrophs. Corticotrophs secrete ACTH, a 39–amino acid peptide derived from the precursor prohormone POMC, which is released during stress on activation of the hypothalamic-pituitary-adrenal axis. Corticotrophs have a very distinctive morphology in rodents, appearing initially as small round cells and developing to become polygonal or stellate with long cytoplasmic processes known as cytonemes.[26] The secretory granules are small and lined up along the perimeter of the cell close to the plasma membrane (Fig. 13.10; Table 13.2). This organization would suggest that ACTH granules are primed for rapid fusion with the plasma membrane and secretion in response to stress. However, human corticotroph ultrastructure is very different, with round cells containing numerous large (250- to 500-nm diameter) secretory granules distributed throughout the cell. Conventional immunocytochemistry in several species shows that corticotrophs are concentrated around the medial central section of the anterior pituitary, but are also distributed sparsely across the gland. Whole-pituitary 3D imaging in a line of POMC-EGFP transgenic mice has revealed that corticotrophs organize in a connected 3D network that exists in close proximity to gonadotrophs and capillaries.[26]

The primary releasing factors for ACTH secretion are hypothalamic CRH and arginine vasopressin (AVP) released into the portal circulation in response to stress, while glucocorticoids exert negative feedback on ACTH secretion. CRH and AVP act on GPCRs to stimulate ACTH release. CRH stimulates both ACTH release and *POMC* transcription via CRH type 1 (CRHR1) receptors acting through cAMP/protein kinase A/CREB-dependent pathways. CRHR1 receptors are critical for control of ACTH release, as in transgenic mice in which CRHR1 is knocked out there is no ACTH response to stress. AVP acting through the V1b receptor is a weak stimulator of ACTH release administered alone, but potentiates the stimulatory effect of CRH on ACTH release.[27] The signaling mechanism involves V1b receptor activation of the PLC/IP3/DAG/PKC signaling cascade, which (1) increases intracellular calcium release from intracellular stores and (2) stimulates cAMP production in response to CRH via protein kinase C–mediated inhibition of phosphodiesterase activity, increasing the exocytosis of ACTH-containing granules. Pharmacological and receptor knockout approaches indicate that AVP is required for the full ACTH response to acute stress and to sustain responses during chronic stress, but in the absence of AVP the amount of ACTH released is sufficient for a full adrenal glucocorticoid response. During infection and inflammation, cytokines released by immune cells (e.g., interleukin [IL]-1, IL-6)[18] and intrapituitary folliculo-stellate cells (leukemia inhibitory factor, IL-6) also stimulate ACTH production to, in turn, amplify production of the antiinflammatory glucocorticoids.[28] Furthermore, CRH and AVP both act as mitogens and stimulate growth of corticotrophs. CRH and AVP both mediate the increase in corticotroph hyperplasia and hypertrophy that results after adrenalectomy. A lack of negative feedback by adrenal glucocorticoids results in an increase in CRH and AVP expression in paraventricular nucleus neurons, and treatment of rats with V1 receptor antagonists or deletion of V1 receptors in mice blocks the mitogenic response to adrenalectomy.[29] Experiments measuring deoxybromouridine uptake have shown that, after adrenalectomy, the newly-produced corticotrophs originate from undifferentiated cells, and not from the division of existing corticotrophs.[30]

Thyrotrophs. Thyrotrophs secrete the 28-kD glycoprotein TSH and have been the least studied of the anterior pituitary cell types. Thyrotroph ultrastructure is of polygonal cells with numerous small-diameter secretory granules (Table 13.2). TSH is composed of an alpha and beta subunit; the structure of the alpha subunit is the same as in LH and FSH (see also Chapter 65). Thyrotrophs appear sparsely distributed through the pituitary, and it is not known if the cells have a network organization. TSH release is pulsatile and has a diurnal pattern, with a peak at 2:00 AM to 4:00 AM and lower concentrations in plasma in the day.[31] TSH production and release are regulated by the hypothalamic hormones TRH and somatostatin, and by triiodothyronine (T3) and thyroxine (T4) negative feedback. TRH stimulates TSH release by stimulation of TRH type 1 GPCRs on thyrotrophs, which activates the PLC/IP3/calcium/PKC signaling pathway, whereas somatostatin inhibits TSH release by acting at SSTR2 receptors, which reduces intracellular calcium. Feedback inhibition by T3 is by stimulation of nuclear thyroid receptors, which downregulate transcription of TSH subunits. Glucocorticoids[32] and cytokines inhibit TSH production and release by both direct thyrotroph actions and indirect hypothalamic actions, and are responsible for low T3/T4 in sickness. The glucocorticoid diurnal pattern of secretion is thought to regulate in turn the circadian pattern of TSH release.[32] Local control of TSH secretion in the pituitary and of TRH within the median eminence also contributes to control of TSH secretion. TRH acts at TRHR1 receptors on tanycytes at the median eminence,[33] increasing tanycyte endfeet contact with TRH nerve terminals and stimulation of a TRH-degrading ectoenzyme, thereby inhibiting the potential to stimulate TSH. In the anterior pituitary, folliculo-stellate cells express TSH receptors, type 2 deiodinase, and the thyroid hormone transporter MCT8. TSH via folliculo-stellate cells stimulates type 2 deiodinase and local T3 production and induces the secretion of folliculo-stellate cell inhibitors of TSH secretion[34] as a paracrine mechanism that exerts ultra-short-loop feedback control of TSH secretion.

In thyroidectomized rodents in which T3 and T4 negative feedback is removed, dramatic changes in thyrotroph ultrastructure are observed; these altered cells are classically described as "thyroidectomy cells." The thyrotrophs are hypertrophied, and rough endoplasmic reticulum is vastly dilated, occupying most of the cytoplasm (Fig. 13.11), indicating hyperfunction and reduced numbers of secretory granules. Similar thyrotroph thyroidectomy morphology is observed in mice with transgenic deletion of the alpha subunit. As functional TSH is not produced, the lack of T3/T4 results in increased TSH beta subunit synthesis by the secretory pathway.[35]

Folliculo-Stellate Cells. Folliculo-stellate cells form a 3D network throughout the anterior pituitary, with gap and adherent junctions between each other and adjacent endocrine cells. Live imaging first revealed folliculo-stellate cell networks in rat anterior pituitary,[36] and more recently in tilapia fish,[37] using the fluorescent-tagged dipeptide β-Ala-Lys-Ne-AMCA. The dipeptide enters folliculo-stellate cells by peptide transporters and is distributed via gap junctions through the folliculo-stellate network. The connections between folliculo-stellate cells and with surrounding cells suggest functions that could efficiently coordinate changes in the pituitary in response to physiological demands, for example during pregnancy.

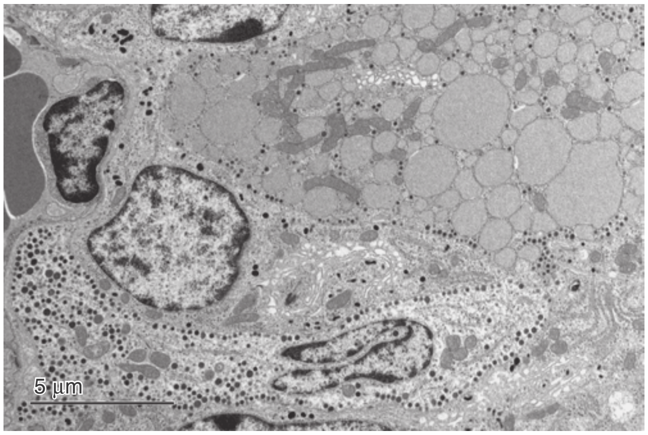


Fig. 13.11 Electron micrograph of a thyrotroph from a thyroidectomised mouse; note the dilated rough endoplasmic reticulum, Golgi apparatus, and mitochondria (a type I lactotroph is below the thyrotroph).

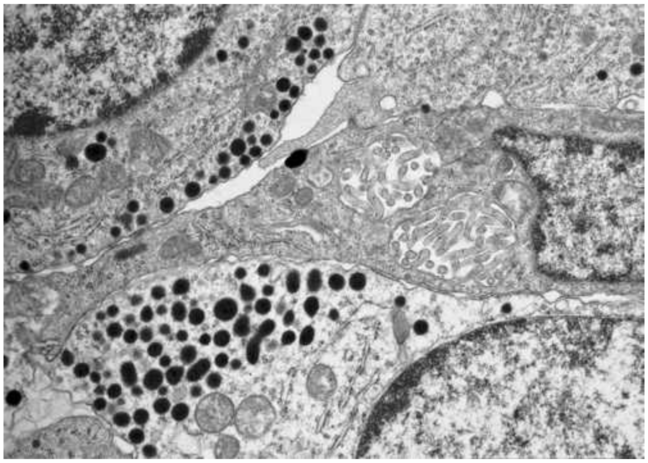


Fig. 13.12 Electron micrograph of a section of a folliculo-stellate cell (rat) forming microvilli-lined follicles.

Folliculo-stellate cells were named based on the fact that, on clustering, the cells formed small follicles (Fig. 13.12) and were highly stellate in structure, with no secretory granules and with sparse rough endoplasmic reticulum or Golgi.[28] Their long cytoplasmic processes facilitate contact via tight junctions with neighboring endocrine cells and interact extensively with perivascular spaces. Marker proteins used to identify folliculo-stellate cells include S100, a calcium-binding protein, and the intermediate filament proteins glial fibrillary acidic protein (GFAP), vimentin, and cytokeratin. Folliculo-stellate cells functionally provide support, protection, and paracrine control of secretion from adjacent endocrine cells via release of factors such as IL-6, annexin 1, and follistatin. Folliculo-stellate cells also regulate the pituitary vasculature, principally through the secretion of vascular endothelial growth factor, which stimulates angiogenesis.[38] Folliculo-stellate cells could therefore play an important part in the coordination of vascular-pituitary plasticity[39] and vascular development.

Pituitary Stem Cells. The presence and identity of pituitary stem cells and their therapeutic potential for regenerative medicine has been one of the most active areas of pituitary research focus in recent years, and is much debated.[40] A large number of studies have provided evidence

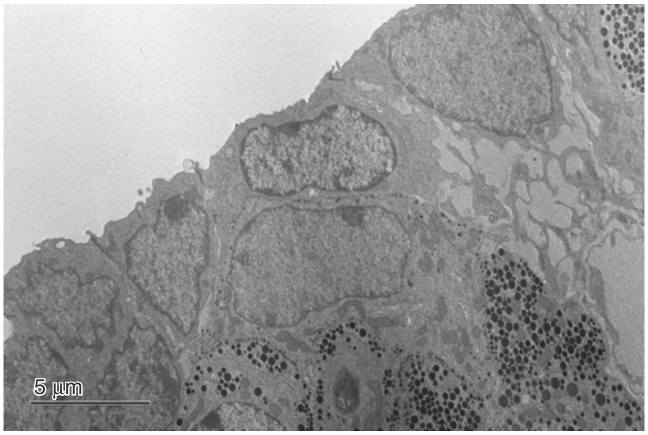


Fig. 13.13 Electron micrograph of marginal cells lining the Rathke's cleft. Note the apical tight junctional complexes and undifferentiated appearance containing few organelles and no secretory granules.

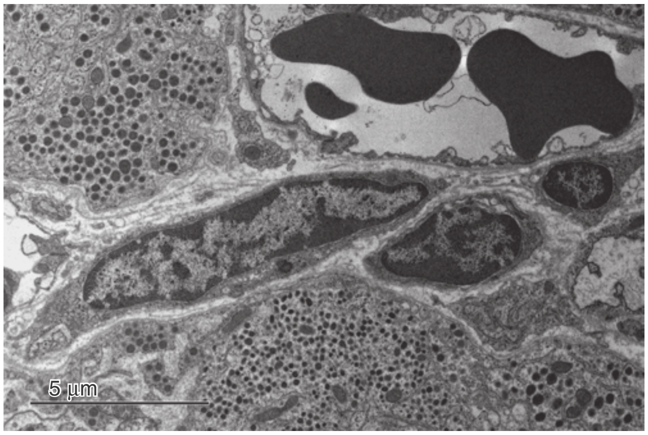


Fig. 13.14 Electron micrograph of three pericytes located in the perivascular space between two capillaries and surrounding secretory cells.

supporting the existence of multipotent stem cells and progenitors in the adult pituitary across species, but their characterization is challenging because of the small number of pituitary stem cells present in adults. Pituitary stem cells express the transcription factor SOX2 and the pituitary transcription factor PROP1, which is necessary to establish the stem cell population and promote an epithelial-to-mesenchymal–like transition, releasing progenitors from the niche.[13] During pituitary embryonic development, proliferating cells are primarily located around the remains of Rathke's cleft. The cleft is retained in adulthood, and the anterior pituitary surface of the cleft is lined with a layer of cells known as the marginal zone.[41] Electron microscopy imaging of the marginal cells (Fig. 13.13) has revealed that they appear undifferentiated, contain sparse organelles, contain no secretory granules, and are joined by intercellular adherens junctional complexes. It has been proposed that marginal cells are stem cells, as they express the classic stem cell markers Nestin, E-cadherin, Sox9, and Sox2,[40] with a similar expression profile and organization in mouse and human.[42] There is also a secondary Sox2- and Sox9-positive population of pituitary stem cells scattered in the anterior pituitary with similar functional features to the marginal cells that is proposed to be a subset of folliculo-stellate cells.[42] Individual pituitary progenitor cells were first shown to be able to produce all anterior pituitary cell types in pituisphere cultures,[43] and pituitary organoids have been successfully grown from Sox2-positive cells.[44] Mouse[45] and human[46] embryonic stem cells have been successfully used to grow 3D pituitary organoids, which are functional with respect to ACTH secretion. In addition, organoids have been grown from induced pluripotent stem cells derived from a patient with congenital hypopituitarism patient with a heterozygous mutation in the orthodenticle homeobox 2 (*OTX2*) gene, providing a powerful human model to be able to identify which step in the development process is not functioning and to study the underlying mechanisms of the disease.[47] Deregulation of pituitary stem cells may result in a tumor-initiating cell population[48] because tumor stem cells have been identified from expression of stem cell markers in pituitary adenomas and other pituitary neoplasms.[49]

KEY POINTS

- The anterior pituitary is situated at the base of the brain in a fold of dura within a boney depression, the sella turcica.
- There are 5 secretory cell types plus, plus folliculo-stellate cells, pericytes and microglia.
- It is connected to the hypothalamus via the portal blood system, which delivers hypothalamic regulators to the pituitary cells.

Finally, normal pituitary structure and function are dependent on the development of the supporting vasculature, yet the molecular detail of pituitary vascular development is not fully known. To date, regenerative pituitary models have focused on secretory cell differentiation and have not reported on blood vessels. In pituitary embryonic development it is not yet clear if development of the vasculature has a direct role in stimulating pituitary cell differentiation.[12] Further characterization of pituitary stem cells in the context of vascular development is required to realize therapeutic potential.

Pericytes. Pericytes are classically described as mural cells located around blood vessels in the perivascular space within the basement membrane and in close proximity to endothelial cells. They are referred to as mural cells because of the presence of contractile fibers that are similar to those found in vascular smooth muscle cells. Pericytes are typically found in the microvasculature, and not in larger vessels. Pericytes typically have a large nucleus with a small amount of cytoplasm and few organelles and extend long cytoplasmic processes within the perivascular space. In the brain, pericytes have been shown to be a critical component of the neurovascular unit, together with endothelial cells and astrocytes, which collectively play an essential role in the maturation and maintenance of the blood-brain barrier. Pericytes have also been well-characterized in the kidney, where they are referred to as mesangial cells, which surround the glomerular capillaries. It has been demonstrated that pericytes play an important role in the separation of vessels into the multiple glomerular capillaries of the renal glomerulus. Pericytes have been shown to be distributed along both the adenohypophysis and neurohypophysis vasculature, and display a phenotype between a fibroblast and a vascular smooth muscle cell. The presence of pericytes in the neurohypophysis was first determined in the 1980s, when pericytes were observed in the perivascular space, and transmission electron microscopy in both the adenohypophysis and neurohypophysis has revealed that pericyte morphology is similar to other organs, with limited cytoplasm and limited numbers of cell organelles. Marker proteins used to distinguish pericytes include desmin and NG2, but these markers are not unique to pericytes. Functional studies in the rat anterior pituitary have shown that pericytes produce components of extracellular matrix type 1 and type III collagen, the collagen-regulating protein fibromodulin, and the TGF-B receptor II. Folliculo-stellate cells produce TGF-B, suggesting that the paracrine control of pericyte collagen synthesis by folliculo-stellate cells is a regulated process. Although pituitary pericytes are found in close proximity to endothelial cells in the pituitary microvasculature, more intimate interactions with endothelial cells in the form of gap junctions and "peg and socket" arrangements have not been observed

(Fig. 13.14). A further network of nongranulated mural cells in the anterior pituitary that express the mural cell marker platelet-derived growth factor receptor beta has been characterized; these cells do not express folliculo-stellate cell markers and are not associated with a perivascular location.[50] The role of this mural cell network requires further characterization.

Pituitary Macrophages. Macrophages of both class M1 and M2 have been identified in the rat anterior pituitary by immunodetection of the respective marker proteins, inducible nitric oxide synthase and mannose receptor. In normal pituitary, both classes are few in number and are distinctively located to perivascular areas.[51] Structural characteristics of pituitary macrophages include heterochromatic nuclei, elongated mitochondria, and large intracellular vesicles lipid droplets and phagolysosomes. M2 macrophages are dramatically increased in number in a rat model of estrogen-induced prolactinoma and may play a role in promoting tumorigenesis.[51] Little is known as to the specific function of macrophages in the pituitary, although their location would render them well-positioned for paracrine communication between the vascular, neuroendocrine, and immune systems.

Coordinated Secretion: Relationship With the Pituitary Vasculature

Recent advances in imaging have revealed more functional details of the coordination of pituitary hormone release, relating particularly to the interrelationships of hypothalamic nerve endings and pituitary cells with the vasculature.[39] *In vivo* imaging of the mouse portal system and somatotroph network to study the delivery of the neurohormone GHRH to somatotrophs has shown that GHRH follows specific capillary routes and results in a specific regional regulation of GH release, and not homogeneous activation of the whole pituitary GH network.[16] Capillaries closely line the clusters and chains of cells that form the GH network. Furthermore, gonadotrophs have also been shown to have a dynamic association with capillary boundaries that varies throughout the ovarian cycle.[52] Electron microscopy imaging of somatotroph ultrastructure in pregnant rats has shown that secretory vesicles redistribute to the regions of the plasma membrane adjacent to a capillary[53] compared with nonpregnant control rats, consistent with more granules being primed to support increased basal GH secretion in pregnancy. Together, these general findings have been confirmed by 3D scanning electron microscopy, which has revealed that, where a cell type contacts a capillary wall, secretory granules in that cell are distributed toward the vessel, whereas if a cell is distant from any vasculature, the secretory granules are more evenly distributed throughout the cytoplasm[54]: 70% of anterior pituitary cells were located adjacent to a capillary, whereas 30% were at the center of acini and distant from perivascular regions. 3D scanning electron microscopy is not able to visualize immunolabeling to identify cell types, but the resolution was high enough that cell types could be differentiated on the basis of ultrastructure. The majority of thyrotrophs and lactotrophs were perivascular, but approximately one third of gonadotrophs, somatotrophs, and corticotrophs were isolated in the parenchyma. It is possible that the isolated secretory cells do not significantly contribute to secretory responses, or it may be the case that they do respond, but hormone that is secreted presumably will diffuse more slowly through the extracellular space to reach a capillary. Overall, the regulation of pituitary secretion is dynamic and involves the integration of the control of hypothalamic, vascular, and local pituitary regulators, and advances in imaging technologies will enable further investigation of these relationships.

Posterior Pituitary Gland (Neurohypophysis)

The posterior pituitary comprises the nerve endings of the oxytocin- and vasopressin-producing magnocellular neurons of the hypothalamus, capillaries, and pituicytes. Oxytocin and vasopressin are both 9–amino acid peptides produced from large precursor proteins that undergo extensive posttranslational processing. The hormones are produced in the cell bodies of the neurons, which are located in the supraoptic and paraventricular nuclei, and are rapidly transported in association with neurophysin along the unmyelinated axons within the hypothalamo-hypophysial tract comprising the pituitary stalk (Fig. 13.15). In the posterior pituitary, each axon branches into several thousand neurosecretory axon swellings (known as “Herring bodies”) and terminals that are packed with oxytocin- or vasopressin-containing dense-cored granules. It has been estimated that each magnocellular neuron contains 10 million secretory granules.[55] Vasopressin maintains body fluid balance and blood pressure by activation of renal V2 receptors (GPCRs coupled to cAMP/PKA/CREB) to increase water reabsorption via insertion of aquaporin water channels, and when blood volume/pressure is decreased, by activation of vascular smooth muscle V1 receptors (GPCRs) coupled to the PLC/IP3/calcium/PKC cascade to cause vasoconstriction (see also Chapter 12).[55] Oxytocin triggers uterine contractions during birth and milk ejection during lactation [55] via oxytocin GPCRs coupled to the PLC/IP3/calcium/PKC cascade, initiating, respectively uterine smooth muscle contractions and contraction of myoepithelial cells in the mammary ducts of the breast. Magnocellular cell bodies are depolarized by stimuli (e.g., dehydration, increased plasma osmolality, suckling) that propagate action potentials, which in turn depolarize the nerve terminals, at which point calcium channels open and intracellular calcium increases, promoting exocytosis of the granules. The axon terminals are closely associated with fenestrated capillaries, allowing the released oxytocin and vasopressin to diffuse across the perivascular space into the bloodstream. Pituicytes are a type of glial cell related to astrocytes that are distributed throughout the pituitary stalk and posterior pituitary (Fig. 13.15). Their role is to support the nerve terminals in clearing local metabolites and ions and in exerting local control of secretion.[56] Pituicytes express the marker protein GFAP, as well as receptors for several neuromodulatory factors such as atrial natriuretic peptide, opioids, and dopamine. The spatial relationship between pituicyte cytoplasmic processes and nerve terminals, including neuroglial synaptoid contacts, plays a modulatory role in the release of oxytocin and vasopressin.[57] When increased peptide secretion is required to meet physiological demands, (e.g., dehydration or lactation) the pituicytes rapidly withdraw their processes from around neurosecretory terminals to allow more ready access of secreted peptide to the circulation. This plasticity in the pituicyte neurovascular spatial relationships and regulation of secretion through morphological changes is similar to tanycyte plasticity in the median eminence and PT follicular cell remodeling with photoperiod.[9]

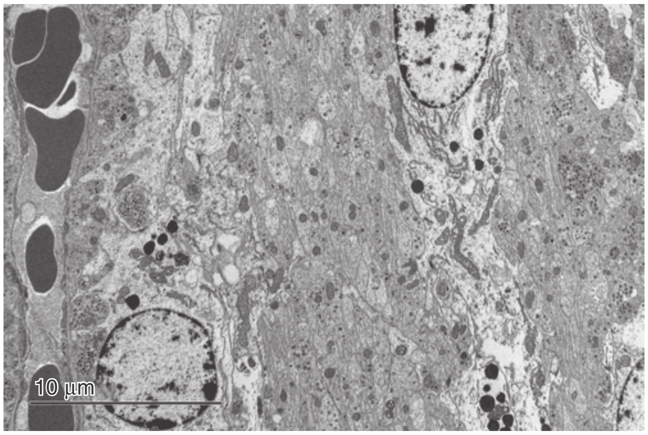


Fig. 13.15 Electron micrograph of magnocellular neuron axons (ovine) within the pituitary stalk; two nucleated pituicytes with lipid droplets are alongside, and a portal vessel is to the left of the image.

KEY POINTS

- The posterior pituitary is an outgrowth from the brain which makes close contact with the anterior pituitary.
- Its principal hormones, secreted from the nerve terminals, are vasopressin and oxytocin.
- Its blood supply is via the inferior hypophyseal artery, a branch of the internal carotid artery.

CONCLUSIONS AND FUTURE DIRECTIONS

The hypothalamus and pituitary gland unit has numerous regulatory and homeostatic roles and provides the pivotal means of communication between the brain and peripheral endocrine systems. There have been many advances in our understanding of the functional importance of the organization of pituitary cells in recent years. Ongoing areas of neuroendocrine research will provide a greater understanding of the plasticity between neuro-glial-vascular-pituitary cell networks in different physiological states. 3D-view serial block face scanning electron microscopy[58] is one technique that could resolve the question of whether a 3D pituitary cell network organization is evident in the human pituitary, and would require high-quality tissue preservation and ultrastructural resolution to allow morphological identification of cell types. With growing understanding of the importance of homotypic cell connections and functional interactions between pituitary cells, questions remain as to the integration of communication between secretory cells in networks via gap junctions and paracrine control via mediators released in the extracellular space. Finally, further knowledge of the factors required for pituitary organogenesis will be important for refining the use of pituitary stem cells as therapeutic tools for hypopituitarism and understanding stem cell involvement in pituitary adenoma development.

REFERENCES

1. Charlton H. Hypothalamic control of anterior pituitary function: a history. *J Neuroendocrinol*. 2008;20:641–646.
2. Raisman G. MEMOIR: Geoffrey Harris and my brush with his unit. *J Endocrinol*. 2015;226:T1–T11.
3. Langlet F. Tanycyte gene expression dynamics in the regulation of energy homeostasis. *Front Endocrinol (Lausanne)*. 2019;10.
4. Prevot V, Dehouck B, Sharif A, et al. The versatile tanycyte: a hypothalamic integrator of reproduction and energy metabolism. *Endocr Rev*. 2018;39:333–368.
5. Wood SH, Christian HC, Miedzinska K, et al. Binary switching of calendar cells in the pituitary defines the phase of the circannual cycle in mammals. *Curr Biol*. 2015;25:2651–2662.
6. Cironi KA, Decater T, Iwanaga J, et al. Arterial supply to the pituitary gland: a comprehensive review. *World Neurosurg*. 2020;142:206–211.
7. Bergland RM, Page RB. Can the pituitary secrete directly to the brain? (Affirmative anatomical evidence). *Endocrinology*. 1978;102:1325–1338.
8. Low MJ. Neural control of the intermediate lobe of the pituitary gland (Pars Intermedia) and proopiomelanocortin. In: Fink G, Pfaff DW, Levine JE, eds. *Handb Neuroendocrinol*. Amsterdam: Elsevier; 2012:157–173.
9. Wood S, Loudon A. The pars tuberalis: the site of the circannual clock in mammals? *Gen Comp Endocrinol*. 2018;258:222–235.
10. Wood SH, Hindle MM, Mizoro Y, et al. Circadian clock mechanism driving mammalian photoperiodism. *Nat Commun*. 2020;11:4291.
11. Denef C. Paracrinicity: the story of 30 years of cellular pituitary crosstalk. *J Neuroendocrinol*. 2008;20:1–70.
12. Davis SW, Ellsworth BS, Peréz Millan MI, et al. Pituitary gland development and disease: from stem cell to hormone production. *Curr Topic Develop Biol*. 2013;106:1–47.
13. Santiago-Andres Y, Golan M, Fiordelisio T. Functional pituitary networks in vertebrates. *Front Endocrinol (Lausanne)*. 2021;11:1–19.
14. Fauquier T, Guérineau NC, McKinney RA, et al. Folliculostellate cell network: a route for long-distance communication in the anterior pituitary. *Proc Natl Acad Sci U S A*. 2001;98:8891–8896.
15. Laporte E, Vennekens A, Vankelecom H. Pituitary remodeling throughout life: are resident stem cells involved? *Front Endocrinol (Lausanne)*. 2021;11:1–18.
16. Lafont C, Desarménien MG, Cassou M, et al. Cellular in vivo imaging reveals coordinated regulation of pituitary microcirculation and GH cell network function. *Proc Natl Acad Sci U S A*. 2010;107:4465–4470.
17. Steyn FJ, Tolle V, Chen C, et al. Neuroendocrine regulation of growth hormone secretion. *Compr Physiol*. 2016;6:687–735.
18. Perez-Castro C, Renner U, Haedo MR, et al. Cellular and molecular specificity of pituitary gland physiology. *Physiol Rev*. 2012;92:1–38.
19. Le Tissier PR, Hodson DJ, Martin AO, et al. Plasticity of the prolactin (PRL) axis: mechanisms underlying regulation of output in female mice. *Adv Exp Med Biol*. 2015;846:139–162.
20. Featherstone K, Hey K, Momiji H, et al. Spatially coordinated dynamic gene transcription in living pituitary tissue. *Elife*. 2016;5:1–25.
21. Momiji H, Hassall KL, Featherstone K, et al. Disentangling juxtacrine from paracrine signalling in dynamic tissue. *PLoS Comput Biol*. 2019;15:e1007030.
22. Grattan DR. The hypothalamo-prolactin axis. *J Endocrinol*. 2015;226:T101–T122.
23. Christian HC, Chapman LP, Morris JF. Thyrotrophin-releasing hormone, vasoactive intestinal peptide, prolactin-releasing peptide and dopamine regulation of prolactin secretion by different lactotroph morphological subtypes in the rat. *J Neuroendocrinol*. 2007;19:605–613.
24. Bernard V, Young J, Binart N. Prolactin — a pleiotropic factor in health and disease. *Nat Rev Endocrinol*. 2019;15:356–365.
25. Herbison AE. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. *Nat Rev Endocrinol*. 2016;12:452–466.
26. Budry L, Lafont C, El Yandouzi T, et al. Related pituitary cell lineages develop into interdigitated 3D cell networks. *Proc Natl Acad Sci U S A*. 2011;108:12515–12520.
27. Gillies GE, Linton EA, Lowry PJ. Corticotropin releasing activity of the new CRF is potentiated several times by vasopressin. *Nature*. 1982;299:355–357.
28. Morris J, Christian H. Folliculo-stellate cells: Paracrine communicators in the anterior pituitary. *Open Neuroendocrinol J*. 2011;4:77–89.
29. Subburaju S, Aguilera G. Vasopressin mediates mitogenic responses to adrenalectomy in the rat anterior pituitary. *Endocrinology*. 2007;148:3102–3110.
30. Aguilera G. The hypothalamic-pituitary-adrenal axis and neuroendocrine responses to stress. *Handb Neuroendocrinol*. 2012:175–196.
31. Russell W, Harrison RF, Smith N, et al. Free triiodothyronine has a distinct circadian rhythm that is delayed but parallels thyrotropin levels. *J Clin Endocrinol Metab*. 2008;93:2300–2306.
32. Haugen BR. Drugs that suppress TSH or cause central hypothyroidism. *Best Pract Res Clin Endocrinol Metab*. 2009;23:793–800.
33. Müller-Fielitz H, Stahr M, Bernau M, et al. Tanycytes control the hormonal output of the hypothalamic-pituitary-thyroid axis. *Nat Commun*. 2017;8:484.
34. Alkemade A, Friesema EC, Kuiper GG, et al. Novel neuroanatomical pathways for thyroid hormone action in the human anterior pituitary. *Eur J Endocrinol*. 2006;154:491–500.
35. Gergics P, Christian HC, Choo MS, et al. Gene expression in mouse thyrotrope adenoma: Transcription elongation factor stimulates proliferation. *Endocrinology*. 2016;157:3631–3646.
36. Fauquier T, Guérineau NC, McKinney RA, et al. Folliculostellate cell network: a route for long-distance communication in the anterior pituitary. *Proc Natl Acad Sci U S A*. 2001;98:8891–8896.
37. Golan M, Hollander-Cohen L, Levavi-Sivan B. Stellate cell networks in the teleost pituitary. *Sci Rep*. 2016;6:24426.
38. Budzik J, Omer S, Morris JF, et al. Vascular endothelial growth factor secretion from pituitary folliculostellate cells: role of KATP channels. *J Neuroendocrinol*. 2014;26:111–120.
39. Le Tissier P, Campos P, Lafont C, et al. An updated view of hypothalamic–vascular–pituitary unit function and plasticity. *Nat Rev Endocrinol*. 2016;13:257–267.

40. Camilletti MA, Martinez Mayer J, Vishnopolska SA, et al. From pituitary stem cell differentiation to regenerative medicine. *Front Endocrinol (Lausanne)*. 2021;11:1–7.
41. Alatzoglou KS, Gregory LC, Dattani MT. Development of the pituitary gland. *Compr Physiol*. 2020;10:389–413.
42. Cox B, Roose H, Vennekens A, et al. Pituitary stem cell regulation: who is pulling the strings? *J Endocrinol*. 2017;234:R135–R158.
43. Fauquier T, Rizzoti K, Dattani M, et al. SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland. *Proc Natl Acad Sci U S A [Internet]*. 2008;105:2907–2912.
44. Cox B, Laporte E, Vennekens A, et al. Organoids from pituitary as a novel research model toward pituitary stem cell exploration. *J Endocrinol*. 2019;240:287–308.
45. Suga H, Kadoshima T, Minaguchi M, et al. Self-formation of functional adenohypophysis in three-dimensional culture. *Nature*. 2011;480:57–62.
46. Ozone C, Suga H, Eiraku M, et al. Functional anterior pituitary generated in self-organizing culture of human embryonic stem cells. *Nat Commun*. 2016;7:1–10.
47. Matsumoto R, Suga H, Aoi T, et al. Congenital pituitary hypoplasia model demonstrates hypothalamic OTX2 regulation of pituitary progenitor cells. *J Clin Invest*. 2020;130:641–654.
48. Haston S, Manshaei S, Martinez-Barbera JP. Stem/progenitor cells in pituitary organ homeostasis and tumourigenesis. *J Endocrinol*. 2018;236:R1–R13.
49. Carreno G, Gonzalez-Meljem JM, Haston S, et al. Stem cells and their role in pituitary tumorigenesis. *Mol Cell Endocrinol*. 2017;445:27–34.
50. O'Hara L, Christian HC, Jeffery N, et al. Characterisation of a mural cell network in the murine pituitary gland. *J Neuroendocrinol*. 2020;32:1–10.
51. Fujiwara K, Yatabe M, Tofrizal A, et al. Identification of M2 macrophages in anterior pituitary glands of normal rats and rats with estrogen-induced prolactinoma. *Cell Tissue Res*. 2017;368:371–378.
52. Alim Z, Hartshorn C, Mai O, et al. Gonadotrope plasticity at cellular and population levels. *Endocrinology*. 2012;153:4729–4739.
53. El-Kasti MM, Christian HC, Huerta-Ocampo I, et al. The pregnancy-induced increase in baseline circulating growth hormone in rats is not induced by ghrelin. *J Neuroendocrinol*. 2008;20:309–322.
54. Yoshitomi M, Ohta K, Kanazawa T, et al. Three-dimensional ultrastructural analyses of anterior pituitary gland expose spatial relationships between endocrine cell secretory granule localization and capillary distribution. *Sci Rep*. 2016;6:1–10.
55. Brown CH. Magnocellular neurons and posterior pituitary function. *Compr Physiol*. 2016;6:1701–1741.
56. Miyata S. Advances in understanding of structural reorganization in the hypothalamic neurosecretory system. *Front Endocrinol (Lausanne)*. 2017;8:1–18.
57. Wittkowski W. Tanycytes and pituicytes: Morphological and functional aspects of neuroglial interaction. *Microsc Res Tech*. 1998;41:29–42.
58. Hughes L, Hawes C, Monteith S, et al. Serial block face scanning electron microscopy--the future of cell ultrastructure imaging. *Protoplasma*. 2014;251:395–401.

14

Genetic Disorders of the Pituitary

Donato Iacovazzo and Márta Korbonits

OUTLINE

Introduction, 212
Familial Isolated Pituitary Adenoma, 212
  AIP-Related Familial Isolated Pituitary Adenoma, 212
  X-Linked Acrogigantism, 214
  Genetically-Undetermined Familial Isolated Pituitary Adenoma, 216
Syndromic Pituitary Tumors, 216
  Multiple Endocrine Neoplasia Type 1, 216
  Multiple Endocrine Neoplasia Type 4, 216
  Carney Complex, 217
  McCune–Albright Syndrome, 217
  Pheochromocytoma/Paraganglioma With Pituitary Adenoma, 217
  DICER1 Syndrome, 218
  Lynch Syndrome, 218
  Neurofibromatosis 1 and Tuberous Sclerosis Complex, 218
  Other Germline Variants Linked With Pituitary Tumor Development, 219
Somatic Mutations in Pituitary Tumors, 219
Criteria for Genetic Screening and Clinical Monitoring, 220
Conclusion, 222

INTRODUCTION

The vast majority of pituitary tumors (PTs) are represented by neuroendocrine tumors (NETs) arising from the endocrine cells of the anterior pituitary, also referred to as pituitary adenomas (PAs) or pituitary NETs. Most PTs occur sporadically without identifiable predisposing genetic factors, whereas, in some 5% of cases, PTs arise in a familial setting. Familial PTs can develop as part of a syndromic condition, such as multiple endocrine neoplasia type 1 (MEN1), multiple endocrine neoplasia type 4 (MEN4), Carney complex (CNC), McCune-Albright syndrome (MAS), pheochromocytoma/paraganglioma with pituitary adenomas (*3P* associations, 3PAs), DICER1, and Lynch syndromes. However, as seen in the case of familial isolated pituitary adenoma (FIPA), pituitary adenomas can occur in the absence of other manifestations, as is the case for patients harboring mutations in the aryl hydrocarbon receptor-interacting protein (*AIP*) gene and in patients with X-linked acrogigantism (XLAG). Nevertheless, in the majority of FIPA families, the causative genetic mutations still remain to be identified.

Owing to either incomplete penetrance or the occurrence of *de novo* mutations, mutations in genes associated with PT development can also be identified in patients without a family history of PTs, especially in those with early-onset disease. A genetic diagnosis is particularly relevant in these patients, as it allows cascade genetic testing and the identification of seemingly unaffected carrier family members who can benefit from an early diagnosis and treatment or, in the case of syndromic diseases, allow screening for other manifestations in the proband.

In this chapter, we will discuss the genetic and clinical aspects of familial and sporadic PTs, focusing on germline and somatic mosaic mutations causing FIPA and syndromic conditions predisposing to PTs, including MEN1, MEN4, CNC, MAS, 3PAs, DICER1, and Lynch syndrome (summarized in Table 14.1).

FAMILIAL ISOLATED PITUITARY ADENOMA

FIPA is characterized by the occurrence of PTs in two or more members of the same family without other associated clinical manifestations.[1] It is estimated to account for around 2% to 4% of all PTs.[2] FIPA is a heterogeneous condition and includes families with the same PT subtype (homogeneous FIPA) or families whose members develop different subtypes of PTs (heterogeneous FIPA). Most homogeneous FIPA families harbor somatotrophinomas or prolactinomas, followed by nonfunctioning pituitary adenomas (NFPAs) and, rarely, corticotrophinomas, whereas in heterogeneous FIPA all possible combinations of different PT subtypes can be observed, the most common being the association of somatotrophinomas and prolactinomas. Most FIPA cases have no known genetic cause, whereas *AIP* mutations[3] can be identified in 10% of FIPA families.[4] As a consequence of incomplete penetrance, *AIP* mutations can also occur in subjects with apparently sporadic early-onset somatotrophinomas or prolactinomas.[5-7] The other well-established FIPA-related genetic abnormality is the duplication of the *GPR101* gene at Xq26.3, identified in patients with infant-onset gigantism.[8] Most of the described XLAG patients presented as isolated cases because of *de novo* germline or somatic mosaic mutations, and only three cases of familial XLAG have been reported so far.[8,9] Examples of different FIPA kindreds are shown in Fig. 14.1.

KEY POINTS

- FIPA is defined by the occurrence of PTs in two or more family members in the absence of other clinical manifestations. Two genes, *AIP* and *GPR101* are known to cause FIPA. Nonetheless, in the majority of FIPA families, the causative genes remain to be identified.

AIP-Related Familial Isolated Pituitary Adenoma

Mutations in the *AIP* gene (11q13.2) are identified in 10% of FIPA cases and 2% to 28% (depending on age at presentation and PT subtype) of

TABLE 14.1 **Germline and Somatic Mosaic Mutations Causing Pituitary Tumors**

Syndrome	Gene and Inheritance Pattern	Germline or Somatic Mosaic	Penetrance for Pituitary Disease	Main Clinical Characteristics
FIPA	*AIP* (AD)	Germline	15%–30%	Early-onset (typically in the second decade) somatotrophinomas or mixed somatotroph-lactotroph tumors and prolactinomas. Mutations in 10% of FIPA families and up to 20% of early-onset PTs (typically causing gigantism or early-onset acromegaly)
	GPR101 (X-linked)	Germline or somatic mosaic in males with sporadic disease	100%	Early-onset (<2 years) gigantism
MEN1	*MEN1* (AD)	Germline or rarely somatic mosaic	30%–40%	Hyperparathyroidism, PTs (mostly prolactinomas and NFPAs), GEP NETs, other neoplasms
MEN4	*CDKN1B* (AD)	Germline	Incomplete (% unknown)	MEN1-like phenotype
Carney complex	*PRKAR1A* (AD)	Germline	10% symptomatic acromegaly, 70% GH excess	Skin pigmented lesions, cardiac and cutaneous myxomas, multiple nonendocrine and endocrine neoplasms including pituitary hyperplasia and PTs (mostly somatotrophinomas and prolactinomas or mixed, very rarely corticotrophinomas)
	Unknown gene (2p16)	Germline	Unknown	Same as for PRKAR1A
	PRKACB	Germline	Unknown	Described in one patient with typical CNC phenotype[63]
McCune–Albright syndrome	*GNAS* (not inheritable)	Somatic mosaic	20%	Café-au-lait spots, polyostotic fibrous dysplasia, precocious puberty, GH excess in about 20% of patients
Pheochromocytoma/ paraganglioma with pituitary adenoma	*SDHA, SDHB, SDHC, SDHD, SDHAF2, MAX (AD)*	Germline	<1%	PPGL, RCC, rarely GIST; rarely PTs (prolactinomas, somatotrophinomas, and NFPAs); one case of pituitary carcinoma reported
DICER1 syndrome	*DICER1* (AD)	Germline or somatic mosaic	<1%	Multiple malignant and benign tumors. Rarely, early-onset pituitary blastomas (ACTH-secreting)
Lynch syndrome	*MSH2, MLH1, MHS6, PMS2, EPCAM (AD)*	Germline	<1%	Multiple cancers (more commonly colorectal and endometrial). Rarely, adult-onset PTs (including two reported pituitary carcinomas)

ACTH, adrenocorticotropic hormone; *AD*, autosomal dominant; *CNC*, Carney complex; *FIPA*, familial isolated pituitary adenoma; *GEP NET*, gastroenteropancreatic neuroendocrine tumor; *GH*, growth hormone; *GIST*, gastrointestinal stromal tumor; *MEN1*, multiple endocrine neoplasia type 1; *MEN4*, multiple endocrine neoplasia type 4; *NFPA*, nonfunctioning pituitary adenoma; *PPGL*, pheochromocytoma and paraganglioma; *PT*, pituitary tumor; *RCC*, renal cell carcinoma.

patients with early-onset sporadic PTs.[4,10] The disease penetrance in *AIP* mutation carriers is typically incomplete. Studies of large families show a penetrance of 15% to 30%. PTs in *AIP* mutation carriers arise at a younger age, presenting clinically in most cases between the second and third decades of life, are more often macroadenomas, and are frequently larger and more invasive compared to *AIP* mutation–negative PTs.[6,11-13] A male preponderance has been described in most studies,[4,12,14] although this could be potentially explained by an ascertainment bias because of the prevalent inclusion of patients with gigantism, a condition occurring more commonly in males.[15] The vast majority (80%) of *AIP*-related PTs are represented by growth hormone (GH)-secreting tumors (somatotrophinomas or mixed somatotroph-lactotroph PTs), followed by prolactinomas (10%–15%).[16] Clinically-presenting NFPAs are rare, accounting for approximately 5% of cases.[12,13] NFPAs are often found to express GH or prolactin on immunohistochemistry. A single case of a thyrotroph PT has been described, while no corticotrophinomas have been identified to date in patients with pathogenic *AIP* mutations. In our cohort, 44% of *AIP*-related PTs manifest clinically with pituitary gigantism, a manifestation of GH excess starting at an early age, before the closure of the growth plates[4] (Fig. 14.2A). Conversely, *AIP* mutations are identified in approximately 40% of patients with nonsyndromic gigantism (28% among those with sporadic disease and 65% of patients from FIPA families).[4] Typically, clinically-presenting *AIP*-related PTs are diagnosed as invasive macroadenomas (Fig. 14.2E–F). Pituitary apoplexy is relatively frequent (~8%–10% of all cases) and can be the presenting feature of the disease (Fig. 14.2E). *AIP*-related somatotrophinomas are often resistant to first-generation somatostatin analogs (SSAs),[12,14] and their management frequently requires a multimodal approach including surgery, radiotherapy, and/or use of the GH receptor antagonist pegvisomant to reduce insulin-like growth factor 1 (IGF-1) levels and prevent the detrimental consequences of GH excess.[17] Data on the use of the multireceptor ligand pasireotide are limited but suggest potential beneficial effects.[18] Prolactinomas in *AIP* mutation carriers are prevalently macroadenomas and are often resistant or partially resistant to medical treatment with dopamine agonists.[12,19]

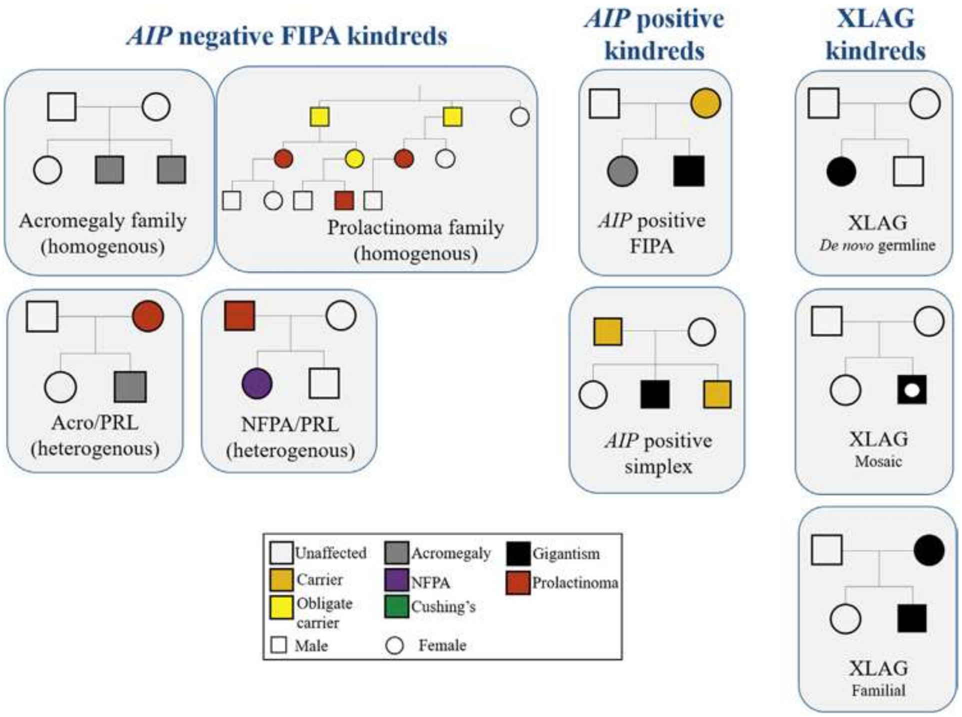


Fig. 14.1 Family trees showing typical examples of various types of familial isolated pituitary adenoma (*FIPA*) kindreds. Different combinations of pituitary tumors can be observed in *AIP* mutation–negative families. Owing to incomplete penetrance, *AIP* mutation–positive families include FIPA kindreds with two or more affected patients or simplex cases without a family history of pituitary tumors. Most X-linked acrogigantism (*XLAG*) patients are simplex cases with females carrying *de novo* germline mutations and males having somatic mosaic mutations. *NFPA,* Nonfunctioning pituitary adenoma; *PRL,* prolactinoma.

Following the identification of an *AIP* mutation, cascade genetic screening should be offered to all individuals at risk. This is particularly relevant because PTs can be diagnosed in approximately 10% of seemingly unaffected carriers, with about half of these patients having previously unrecognized GH excess.[4] Owing to an early diagnosis, treatment outcomes in these patients are generally more favorable compared with clinically presenting PTs. However, it should be noted that screening of clinically unaffected *AIP* mutation carriers can identify microadenomas in some cases, without apparent hormone excess.[4] These are somewhat similar to those identified in *MEN1* mutation carriers[20] and may represent incidentalomas occurring in the general population.

Over 100 different mutations have been identified in the *AIP* gene,[1,4] including nonsense, missense, frameshift, splicing, and promoter mutations, deletions, insertions, and segmental duplications. Approximately 70% of these mutations are truncating and lead to the loss of the C-terminal end of *AIP* as a consequence of either nonsense or frameshift mutations resulting in premature stop codons. The most common *AIP* pathogenic variant is the nonsense p.R304* mutation occurring at a CpG hotspot site: a founder effect secondary to this specific *AIP* mutation, as well as other hotspot mutations, has been described in several populations.[3,21,22] Large deletions account for approximately 5% of all mutations but can only be identified with dedicated techniques, such as multiplex ligation-dependent probe amplification. Most *AIP* mutations are inherited, and *de novo* mutations have been very rarely described.

AIP is a cochaperone with several interacting partners, including members of the heat shock protein family, growth factor receptors, nuclear receptors, and viral proteins. The protumorigenic effects of *AIP* mutations depend on the loss of its tumor suppressor function. *AIP* mutations, in fact, result in most cases in the premature truncation of the coding sequence (nonsense and frameshift mutation) or in unstable proteins with a reduced half-life (missense mutations or segmental duplications).[22,23] Moreover, loss of heterozygosity (LOH) at the AIP locus is often found in AIP-related PTs, confirming that tumorigenesis depends on the loss of functional copies of AIP within the tumor cells. The mechanisms linking loss of AIP with PT development remain to be fully elucidated but are likely to be multifactorial and include, among others, enhanced cyclic AMP (cAMP)-dependent protein kinase A (PKA) signaling[24] and downregulation of the inhibitory $G\alpha_{i\text{-}2}$ protein[24] and of the tumor suppressor ZAC1.[25] Alteration of the RET apoptotic pathway has been recently described as a tumorigenic mechanism.[26] Transcriptome analysis of *AIP*-related PTs revealed upregulation of specific cytokines with increased macrophage infiltration, highlighting changes in the tumor microenvironment as factors contributing to the invasive behavior of *AIP*-related PTs.[27]

KEY POINTS

- *AIP* mutations are identified in approximately 10% of FIPA cases and up to 28% of patients with sporadic gigantism or young-onset acromegaly. Penetrance is typically incomplete.

X-Linked Acrogigantism

XLAG is a condition of infant-onset pituitary gigantism secondary to the germline or somatic mosaic duplication of the *GPR101* gene (Xq26.3).[8,28,29] XLAG patients present significant GH excess, in most cases with concomitant hyperprolactinemia, caused by mixed somatotroph-lactotroph tumors associated, in some patients, with pituitary hyperplasia. A minority of patients present with hyperplasia in the absence of a PT. XLAG is very rare, with only 34 confirmed cases described so far.[8,9,28,30-34] Unlike other forms of gigantism, XLAG is characterized by a female preponderance. Some 25 of the 34 reported XLAG patients are females carrying germline duplications, whereas somatic mosaic mutations have been identified in the only four reported cases

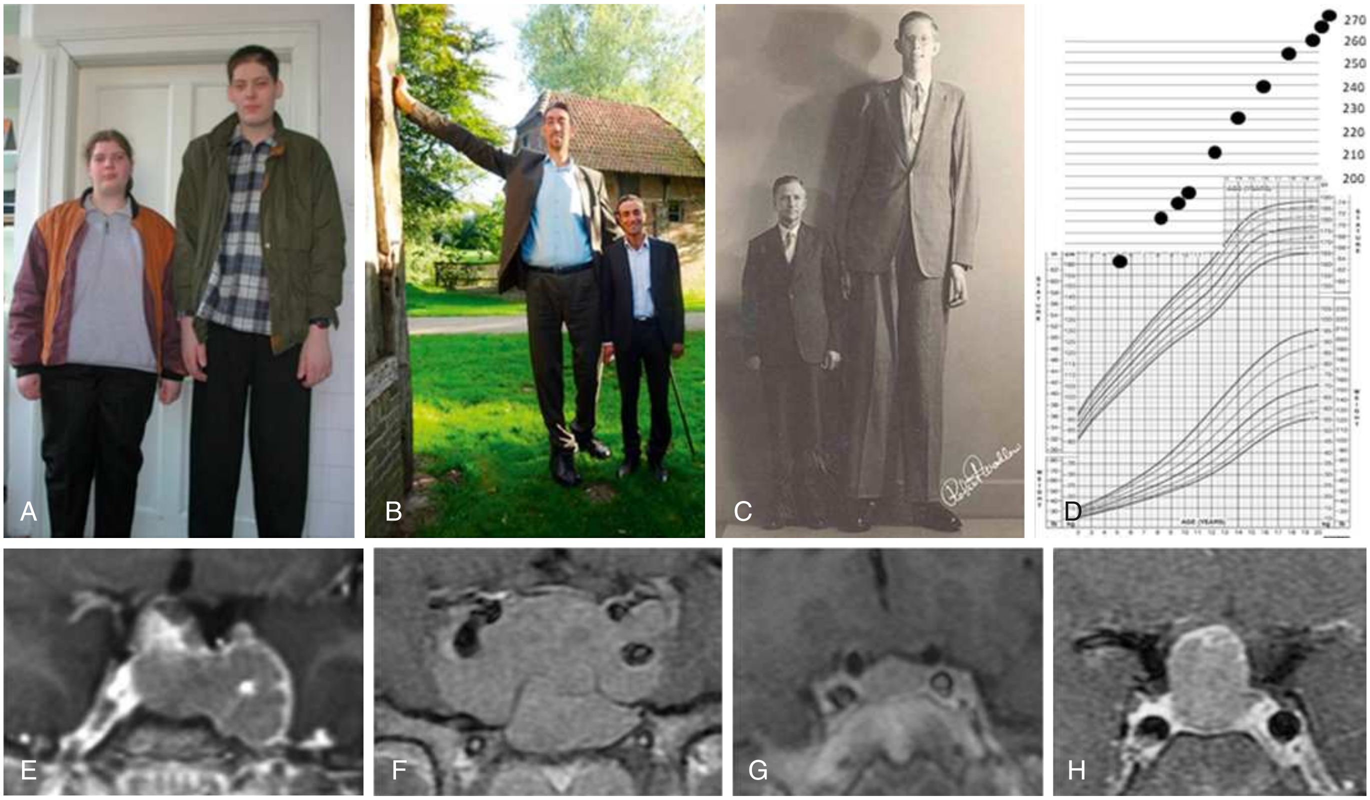


Fig. 14.2 Clinical and radiological features of pituitary gigantism secondary to *AIP* and *GPR101* mutations. **A,** Siblings with whole *AIP* gene deletion. The female patient was diagnosed and treated at the age of 10 years when she was 175 cm (5 ft 8.9 in), while her brother was diagnosed and treated at the age of 21 years when 210 cm (6 ft 10.7 in) (Prof. Vera Popovic's patients). **B–D,** X-linked acrogigantism (XLAG) can lead to extreme tall stature if not treated successfully. **B,** An affected patient (height 251 cm, 8 ft 2.8 in) is shown here next to an unaffected family member (picture kindly provided by Prof. Wouter de Herder, Erasmus MC, Rotterdam). **C–D,** Based on the clinical phenotype, Robert Wadlow, the Alton Giant, is likely to have suffered from XLAG. His birth weight was normal, but he immediately started growing at a rapid rate, and by 6 months of age his weight was 14 kg (30.9 lbs; 50th centile of a 2 year and 10 month old child). He received no treatment, reached a final height of 272 cm (8 ft 11.1 in), and died at the age of 28 years following an infection. Based on the detailed description provided by Behrens in 1932,[148] his sella turcica measured 2.5 cm, suggesting the presence of a pituitary tumor (**B–D** modified from Iacovazzo D & Korbonits M. Gigantism: X-linked acrogigantism and *GPR101* mutations. *Growth Horm IGF Res.* 2016;30-31:64-69). **E–F,** Clinically-presenting pituitary tumors in *AIP* mutation carriers are often macroadenomas with cavernous sinus invasion. Apoplexy (**E**, modified from Dutta P, Reddy KS, Rai A, et al. Surgery, octreotide, temozolomide, bevacizumab, radiotherapy, and pegvisomant treatment of an *AIP* mutation-positive child. *J Clin Endocrinol Metab.* 2019;104:3539-3544.) occurs in approximately 8% to 10% of cases. **G–H,** Examples of the radiological findings in patients with XLAG, including isolated pituitary hyperplasia (G) and a pituitary macroadenoma without evidence of cavernous sinus invasion (H).

of male patients with sporadic disease. In three independent families, mother-to-son transmission has been described, in all instances with full penetrance. As no other clinical manifestations have been described in these patients, XLAG is considered to be a rare cause of FIPA.

XLAG is characterized by the early onset of accelerated growth, in most cases during the first 2 years of life. Other frequently observed manifestations include acral enlargement, coarse facial features, increased appetite, and, less frequently, acanthosis nigricans, sleep apnea/snoring, and hyperhidrosis. Treatment of XLAG patients is often challenging, and, in the absence of prompt treatment, the disease results in markedly increased stature (Fig, 14.2B–D). The radiological features are variable and range from mild pituitary gland enlargement in patients with isolated hyperplasia (Fig. 14.2G) to macroadenomas (Fig. 14.2H), which only rarely show invasive features. Although radical surgery, often resulting in hypopituitarism, can provide control of the disease in some cases, several patients required a combination of surgery, radiotherapy, and medical treatment. SSAs, especially in treatment-naïve patients, are generally ineffective, whereas pegvisomant may prove effective in normalizing IGF-1 levels and growth velocity, allowing the achievement of a normal height in cases with timely diagnosis and prompt initiation of treatment. Hyperprolactinemia in XLAG patients is generally responsive to dopamine agonists.

Most XLAG patients harbor microduplications of Xq26.3 (spanning a region of about 500 kb) involving *GPR101* as well as three other neighboring genes. However, one patient with a typical clinical phenotype was found to have a complex genomic rearrangement with two duplicated regions separated by a normal copy number segment. The duplication in this patient allowed researchers to narrow down the genomic region shared by all patients to an area encompassing solely (in its entirety) the *GPR101* gene, confirming its pathogenic role.[28] Small duplications and low-level mosaicism can be missed by standard chromosome microarrays, and either gene dosage analysis by means of digital PCR or a high-resolution array should be employed when testing for XLAG. Testing of tumor-derived DNA should also be considered when suspecting a somatic mosaic mutation.

The *GPR101* gene encodes an orphan G-protein coupled receptor. In mice, *Gpr101* was found to be expressed primarily in the central nervous system, and particularly in the hypothalamus and amygdala.[35] In humans, *GPR101* is highly expressed at the mRNA level in the nucleus accumbens, as well as in the medulla and the occipital lobe.[36] Although *GPR101* is significantly overexpressed in the PTs of XLAG patients, it is not expressed in sporadic somatotrophinomas or in the normal adult pituitary gland.[8] The mechanisms underlying the pathogenesis of XLAG remain to be defined. GPR101 is coupled to the G-stimulatory protein and is constitutively active.[8,35] The activation of the cAMP-PKA pathway resulting from the overexpression of *GPR101* could potentially underlie the development of pituitary hyperplasia and PTs in XLAG patients. However, the high level of *GPR101* expression in the hypothalamus, and especially in the arcuate nucleus, suggests that this receptor could also potentially play a role in the hypothalamic regulation of GH secretion. Recently, GPR101 was identified as the receptor mediating the antiinflammatory actions of the lipid mediator Resolvin D5 in leukocytes.[37] However, the role of this lipid mediator in the GH axis has not been investigated as yet.

KEY POINTS

- XLAG is a rare condition of infant-onset gigantism caused by the duplication of the *GPR101* gene. Penetrance is complete and all patients develop accelerated growth starting during the first two years of life.

Genetically-Undetermined Familial Isolated Pituitary Adenoma

In most (90%) FIPA families, no causative genetic mutations are identified. Age at disease onset in these patients is comparable to that in patients with sporadic PTs, and both homogeneous (more commonly somatotrophinomas followed by prolactinomas and NFPAs) and heterogeneous (all possible combinations with somatotrophinomas/prolactinomas and somatotroph/NFPAs being the most common) combinations can be observed. Prospectively diagnosed PTs have been described in genetically-undetermined FIPA families, confirming the presence of inheritable predisposing genetic mutations. However, considering the relatively high frequency of PTs in the general population, it is possible that the occurrence of PTs in some of these families could be coincidental.

SYNDROMIC PITUITARY TUMORS

Multiple Endocrine Neoplasia Type 1

MEN1 is an autosomal dominant tumor predisposition syndrome characterized by the occurrence of primary hyperparathyroidism (in >90% of patients by the age of 50 years), gastroenteropancreatic NETs (in ~60% of patients) and PTs (30%–40% of cases) (see also Chapter 133). The prevalence is estimated between 1:10,000 and 1:100,000. Other endocrine and nonendocrine tumors may also occur, such as bronchial and thymic NETs, facial angiofibromas, lipomas, collagenomas, meningiomas, ependymomas, adrenal adenomas, breast cancer and, rarely, pheochromocytomas.[38-40] In 90% of cases, MEN1 is caused by heterozygous mutations in the *MEN1* gene (11q13). Most mutations are inherited, whereas *de novo* germline or mosaic mutations occur in approximately 10% of patients.

PTs occur in approximately 30% to 40% of MEN1 patients and represent the first manifestation of the disease in about 15% to 30% of cases. PTs are more commonly diagnosed in female patients. Prolactinomas are the most common clinically presenting PT subtype in MEN1 (40%–60%), followed by NFPAs (15%–40%), somatotrophinomas (5%–10%), and, rarely, corticotrophinomas or thyrotrophinomas. It should be noted that GH excess in the setting of MEN1 can be secondary to the ectopic secretion of GH-releasing hormone (GHRH) by NETs, most commonly pancreatic. PTs in MEN1 are frequently macroadenomas, they arise at a younger age compared with sporadic PTs, and, in rare instances, they can be multiple. With radiological screening, indolent, clinically nonfunctioning PTs have also been identified.[20] Responsiveness of prolactinomas to dopamine agonists is variable, and in some series they were found to be more often resistant to medical treatment compared with those arising in subjects without MEN1.[41] Considering that PTs may represent the first disease manifestation and the possibility of *de novo* mutations, screening for MEN1 should be considered in patients with childhood-onset pituitary macroadenomas, especially prolactinomas, even in the absence of a personal or family history of MEN1-related disorders, as a relatively high frequency (6%) of *MEN1* mutations has been found in patients with childhood-onset sporadic PTs.[7]

To date, more than 1500 *MEN1* mutations have been described.[42,43] Most of these are represented by frameshift, missense, and nonsense mutations distributed throughout the whole gene. A clear correlation between genotype and risk of PTs has not been demonstrated. *MEN1* is a tumor suppressor gene located on the long arm of chromosome 11 (11q13). LOH at this locus is identified in the vast majority of MEN1-related tumors. *MEN1* encodes a protein named menin, a scaffold protein involved in several cellular processes, including transcriptional regulation, genome stability, cell division, and proliferation.[38] Menin serves as a molecular adaptor to allow the interaction between the mixed lineage leukemia (MLL) protein and the transcriptional coactivator lens epithelium-derived growth factor, which is needed for the association of the MLL complex with chromatin and the expression of MLL target genes.[44] Menin recruits MLL to the promoters of the cyclin-dependent kinase inhibitor 1B (*CDKN1B*) and 1C (*CDKN1C*)[45,46] genes, promoting the transcription of these genes, which encode, respectively, p27 and p57. Loss of these tumor suppressors, which are predominantly expressed in endocrine tissues, might explain the high frequency of endocrine neoplasms in patients with MEN1.

KEY POINTS

- PTs occur in approximately 30-40% of MEN1 patients and include predominantly prolactinomas, NFPAs and somatotrophinomas.

Multiple Endocrine Neoplasia Type 4

MEN4 is a rare condition caused by mutations in the *CDKN1B* gene (12p13) encoding the tumor suppressor p27 (see also Chapter 134). The clinical manifestations generally overlap those of MEN1, with most reported subjects developing primary hyperparathyroidism and, less often, pituitary tumors or other NETs. MEN4 accounts for approximately 2% of *MEN1* mutation–negative MEN cases. To date, 47 cases have been reported.[47-49] Among these, 14 patients developed PTs, including five corticotrophinomas, four somatotrophinomas, four NFPAs, and one prolactinoma. It should be noted that, in the case of two *AIP* mutation–negative FIPA families harboring two distinct *CDKN1B* variants,[50] for one of the identified variants, segregation with the PT phenotype could not be assessed, whereas for the second, only one of the two affected family members carried the variant, therefore rendering it an unlikely cause of the familial PT. Most patients were diagnosed with PTs as adults. However, one patient presented with gigantism caused by a somatotrophinoma diagnosed at the age of five years,[51] and pathogenic *CDKN1B* variants have been recently identified in a small percentage of patients with sporadic childhood-onset Cushing disease,[49] suggesting that sequencing for *CDKN1B* should be considered in these patients. These data also suggest that, in families with known mutations, screening for PTs should start at an early age.

The *CDKN1B* mutations reported so far include frameshift, nonsense, and missense mutations, as well as 5’ untranslated region mutations leading to reduced p27 expression. Considering the rarity of the condition, the exact penetrance and potential genotype–phenotype correlations remain to be established.

KEY POINTS

- MEN4 is a rare autosomal dominant multiple endocrine neoplasia syndrome whose manifestations overlap those of MEN1. Rarely, PTs have been described in the setting of MEN4.

Carney Complex

CNC is a rare autosomal dominant condition characterized by pigmented lesions of the skin, cardiac and cutaneous myxomas, and multiple nonendocrine and endocrine neoplasms, including pituitary hyperplasia and PTs.[52] About two-thirds of patients present elevation of GH or IGF-1 levels, often with concomitant hyperprolactinemia, although symptomatic acromegaly occurs only in about 10% of cases, usually by the third decade of life, presenting clinically with acromegaly or, less frequently, gigantism.[53,54] Most patients with GH excess present with pituitary hyperplasia, typically affecting the somatolactotroph cells expressing both GH and prolactin, which can be accompanied, in some patients, by one or multiple tumors. Most PTs observed in CNC are somatotroph or mixed somatotroph-lactotroph microadenomas,[55,56] although large macroadenomas have also been described. Although most cases of Cushing syndrome in CNC patients are adrenocorticotropic hormone (ACTH)-independent, ACTH-dependent Cushing disease has been reported in two patients carrying a *PRKAR1A* mutation.[57,58] In both cases, LOH in the corticotrophinoma was shown, supporting a direct pathogenic role of the *PRKAR1A* mutation. Among the endocrine manifestations occurring in CNC patients, the most common is ACTH-independent Cushing syndrome owing to primary pigmented nodular adrenocortical disease, which is observed in about 25% of CNC patients.[59] Other endocrine manifestations include testicular tumors, especially large-cell calcifying Sertoli cell tumors, observed in about one third of affected males, thyroid nodules (mostly follicular adenomas), which occur in up to 75% of CNC patients, and, occasionally, differentiated thyroid cancer.

Approximately 70% of CNC cases are secondary to heterozygous inactivating mutations in the *PRKAR1A* gene (17q24.2), coding for the regulatory subunit type 1 alpha of PKA. To date, over 125 *PRKAR1A* mutations have been described,[53] most of which are represented by nonsense and frameshift mutations. In most cases, the predicted mutant proteins are *not* produced as a result of mRNA nonsense-mediated decay. Large deletions can be identified in about 20% of cases where *PRKAR1A* mutations cannot be detected by Sanger sequencing and are often associated with a more severe phenotype.[60] Approximately 70% of cases occur in a familial setting, whereas 30% present as sporadic cases because of *de novo* mutations. *PRKAR1A* acts as a tumor suppressor gene, as confirmed by the frequent finding of LOH in tumors associated with CNC. Loss of PRKAR1A leads to increased PKA activity resulting from the unopposed activation of the catalytic subunits of PKA. The mechanisms linking increased PKA activity with tumor formation remain to be fully established but are likely to involve multiple signaling pathways and include upregulation of D cyclins and activation of the mammalian target of rapamycin pathway.[61,62]

A single case of CNC was described in a patient with a chromosomal triplication affecting the *PRKACB* gene (1p31.1) encoding the catalytic subunit beta of PKA.[63] This patient presented at the age of 19 years with acromegaly, spotty skin hyperpigmentation, and multiple myxomas. A second genetic locus at 2p16 has been associated with CNC; however, the causative gene is still unknown.[64]

KEY POINTS

- Carney complex is a rare autosomal dominant syndrome characterized by multiple endocrine and nonendocrine neoplasms. Approximately 10% of patients present with symptomatic GH excess.

McCune–Albright Syndrome

Somatic mutations in the *GNAS* gene (20q13.32), encoding the stimulatory G protein Gsα, represent the most common recurrent mutation occurring in somatotrophinomas.[65,66] These activating missense mutations occur at either one of two residues: Arg201 or, more rarely, Gln227. Mutations at these sites determine loss of GTPase activity with constitutive activation of the cAMP-dependent PKA, resulting in increased proliferation in cAMP-responsive tissue, including the pituitary gland. When these mutations occur at an early postzygotic stage of development, the resulting somatic mosaicism underlies a rare syndromic condition known as MAS. MAS has an estimated prevalence between 1:100,000 and 1:1,000,000[67] and is characterized by polyostotic fibrous dysplasia, café-au-lait skin macules, and several endocrine manifestations, including precocious puberty (especially in females), hyperthyroidism, testicular lesions (Leydig and/or Sertoli cell hyperplasia), GH excess, and, more rarely, neonatal hypercortisolism. The chances of detecting a *GNAS* mutation by standard PCR are higher in affected tissues, while they are low in leukocyte-derived DNA, especially in patients with only one clinical manifestation. However, the detection rate in leukocyte-derived DNA can be improved by high-sensitivity sequencing using digital PCR.[68] The *GNAS* gene is paternally imprinted in the pituitary gland, and most PTs occur in patients harboring the mutation on the maternal allele.[69,70]

Approximately 10% to 20% of patients present with GH excess[71] associated, in about 80% of affected cases, with hyperprolactinemia. PTs (more commonly macroadenomas) can be identified in 30% to 50% of affected patients, with the remaining patients developing pituitary hyperplasia without a PT. In a group of over 100 MAS patients with GH excess, the mean age of onset was 24 years.[71] Although a significant proportion of patients present clinically with accelerated growth, the final stature is often normal, most likely as a consequence of the high prevalence of associated precocious puberty and increased sex steroid levels. Surgery is often challenging in MAS patients because of fibrous dysplasia affecting the skull base and is only rarely curative. The use of radiotherapy is controversial.[71,72] SSAs can induce control of the acromegaly in only 30% of cases, whereas pegvisomant has been shown to be effective in normalizing IGF-1 levels in most MAS patients in whom it has been employed.[71-73]

KEY POINTS

- MAS is a rare condition secondary to somatic mosaic *GNAS* mutations. Multiple endocrine conditions can occur, including GH excess in approximately 10% to 20% of affected patients.

Pheochromocytoma/Paraganglioma With Pituitary Adenoma

Germline heterozygous mutations in genes encoding succinate dehydrogenase subunits (*SDHx*) and the succinate dehydrogenase (SDH) complex assembly factor 2 protein (*SDHAF2*) represent a cause of hereditary pheochromocytoma and paraganglioma (PPGL) (see also Chapter 99). Several reports have described the occurrence of PTs in patients with *SDHx* mutations,[74-80] confirming an association of pheochromocytoma/paraganglioma with pituitary adenoma (3PAs). LOH has been shown in some of these cases, strengthening the causal association between *SDHx* mutations and PT development. The pituitary cells in *SDHx* mutation–positive PTs often show intracytoplasmic vacuoles, which could draw attention to a possible germline mutation in these cases.[77]

The penetrance of PTs in patients carrying SDHx mutations is low (<1%).[74-77] Prolactinomas, followed by somatotrophinomas and NFPAs, are the typical tumor types observed. Most patients presented with macroadenomas, often with aggressive behavior, including one case of a hormone-negative, steroidogenic factor 1–positive pituitary carcinoma in a patient carrying an *SDHB* mutation.[80] *SDHx* mutations in patients with isolated sporadic PTs are extremely rare,[76] and sequencing for *SDHx* genes should be reserved for patients with a personal or family history of paraganglioma or pheochromocytoma. It should be noted that, rarely, the association of PTs and pheochromocytomas can occur in the setting of MEN1, or the association of acromegaly or Cushing syndrome with pheochromocytoma could be because of ectopic GHRH, ACTH, or corticotropin-releasing hormone secretion by the pheochromocytoma.

The SDH enzymatic complex is composed of four subunits (SDHA, SDHB, SDHC, and SDHD) and is responsible for the reversible enzymatic conversion of succinate into fumarate within the citric acid cycle. The mechanisms linking loss of SDH function with tumor formation are likely to be multifactorial. Disruption of the SDH complex leads to the accumulation of succinate, with consequent inhibition of prolylhydroxylases leading to stabilization of the hypoxia-inducible factor 1α (HIF1α) and transcription of HIF-responsive genes, including growth factors with oncogenic properties, such as vascular endothelial growth factor and transforming growth factor.[81] Moreover, loss of SDH function is associated with genomic methylation and downregulation of several genes, including tumor suppressors and genes involved in epithelial-mesenchymal transition.[82]

Recently, PTs have been described in patients with pheochromocytomas carrying mutations in *MAX* (14q23.3), one of several genes causing familial PPGL.[83-86] These cases include prolactinomas, somatotrophinomas and a case of acromegaly secondary to ectopic GHRH secretion. Notably, one patient had a family history of acromegaly, although co-segregation could not be investigated. Other neuroendocrine and nonendocrine neoplasms have been described in these kindreds, suggesting that *MAX* might represent a novel multiple endocrine neoplasia gene.

KEY POINTS

- PTs can occur in patients carrying mutations in genes predisposing to the development of pheochromocytomas and paragangliomas. This association has been named pheochromocytoma/paraganglioma with pituitary adenoma (3PAs).

DICER1 Syndrome

The DICER1 syndrome, or pleuropulmonary blastoma–familial tumor and dysplasia syndrome, is a rare autosomal dominant condition caused by heterozygous mutations in the *DICER1* gene (14q32.13).

DICER1 syndrome is characterized by a variety of neoplasms, including pleuropulmonary blastoma, ovarian sex cord-stromal tumors (mostly Sertoli-Leydig cell tumor), cystic nephroma, renal sarcoma, genitourinary embryonal rhabdomyosarcoma, nodular hyperplasia of the thyroid, differentiated thyroid cancer, pituitary blastoma, nasal chondromesenchymal hamartoma, ciliary body medulloepithelioma, and pinealoblastoma.

PTs are rare in the setting of DICER1 syndrome (<1%). These neoplasms have been named pituitary blastomas because their features resemble pituitary embryonic tissue, and they are characterized by an aggressive clinical behavior.[87] These aggressive tumors usually arise in young children (median age at presentation is 8 months, with a range from 7 to 24 months), and they present clinically with severe ACTH-dependent hypercortisolism and, in some cases, ophthalmoplegia and visual disturbances. Pituitary blastomas are fatal in about 40% of cases.[87,88] Of 12 pituitary blastoma patients available for genetic testing, *DICER1* mutations were identified in 11 patients,[88] suggesting that these neoplasms represent a pathognomonic feature of DICER1 syndrome.

DICER1 encodes a cytoplasmic endoribonuclease that is responsible for processing miRNAs, which modulate mRNA expression at the posttranscriptional level. In DICER1 syndrome, a germline loss-of-function mutation (more often represented by a nonsense or frameshift mutation) is generally followed by a second somatic "hit," often involving the RNase IIIb catalytic domain of DICER1.[88,89] Somatic mosaic mutations, most commonly in the RNase IIIb catalytic domain, have been identified in some patients using high-sensitivity sequencing techniques, although no PTs were identified in these patients.[90,91]

KEY POINTS

- A subtype of aggressive, infant-onset ACTH-producing PTs named pituitary blastomas occur in the setting of DICER1 syndrome, a rare autosomal dominant condition caused by mutations in the *DICER1* gene.

Lynch Syndrome

Recently, six cases of PTs have been described in the context of Lynch syndrome,[92-95] an autosomal dominant cancer predisposition syndrome characterized by an increased risk of colon cancer as well as other malignancies. Lynch syndrome is inherited with an autosomal dominant pattern and is caused by heterozygous mutations in the DNA mismatch repair genes (*MSH2*, *MLH1*, *MSH6*, *PMS2*) or the *EPCAM* gene. Of the six reported cases, four were tumors with an aggressive clinical behavior, including three corticotrophinomas (one of which rapidly developed into a pituitary carcinoma) and a poorly-differentiated carcinoma. Notably, loss of expression of DNA mismatch repair proteins was described in three cases with available tissue and LOH at the *MSH2* locus was demonstrated in one case, supporting a pathogenic role for the DNA mismatch repair gene mutations. Although the occurrence of PTs in patients with Lynch syndrome appears to be a very rare event, these tumors seem to represent an aggressive subgroup, and the potential risk of pituitary malignancy should be considered in patients with this condition.

KEY POINTS

- Rarely, PTs including pituitary carcinomas have been described in the setting of Lynch syndrome, a tumor-predisposing conditions caused by mutations in a group of DNA mismatch repair genes.

Neurofibromatosis 1 and Tuberous Sclerosis Complex

Neurofibromatosis 1 (NF1) is an autosomal dominant condition characterized by café-au-lait spots, cutaneous and plexiform neurofibromas, and other manifestations, including brain neoplasms, especially childhood-onset optic gliomas.[96] NF1 occurs with a prevalence of 1:2000 to 1:4000 and is caused by mutations in the *NF1* gene (17q11.2). GH excess manifesting clinically with accelerated growth occurs in approximately 10% of NF1 children with optic gliomas.[97] GH excess in these patients is generally responsive to SSAs and can undergo spontaneous resolution with time. In some patients, GH excess is associated with precocious puberty. The underlying mechanisms are unknown, but, considering that usually no PTs are observed in these patients, hypothalamic dysregulation of somatostatin or GHRH signaling is likely to be involved. PTs, including prolactinomas, somatotrophinomas, silent corticotrophinomas, and NFPAs, have been reported in NF1 patients.[98-105] In one of these cases, LOH at the *NF1* gene locus was ruled out, and the PT was found to harbor a somatic *GNAS* mutation instead.[101] Considering the relatively high frequency of both PTs and NF1 in the general population, the occurrence of PTs in patients with NF1 is likely to represent a coincidental finding.

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder characterized by the development of multiple hamartomas in several organs, including the skin, central nervous system, kidneys, heart, eyes, gastrointestinal tract, lung, and bones.[106] It is caused by mutations in either of the two tumor suppressor genes *TSC1* (9q34) and *TSC2* (16p13.3). There are a few reports addressing PTs in TSC patients, including prolactinomas, corticotrophinomas, somatotrophinomas, and NFPAs.[107-112] In one patient carrying a variant of unknown significance in the *TSC2* gene, LOH was not detected, and TSC2 staining was preserved in the tumor cells.[107] It is not clear whether TSC patients are at increased risk of PTs, and, pending future studies, the role of *TSC1* and *TSC2* mutations in the pathogenesis of PTs remains uncertain.

Other Germline Variants Linked With Pituitary Tumor Development

A heterozygous variant in the prolactin receptor *PRLR* gene (5p13.2) was described in three sisters with hyperprolactinemia in the absence of a PT.[113] This variant resulted in loss of receptor signaling with consequent prolactin insensitivity. A more recent study looking at *PRLR* variants in patients with prolactinomas described six germline variants, including four rare ones, in a group of 46 patients. Out of the four rare variants, two were found to be significantly enriched compared to general population databases. Only one of these variants was found to increase prolactin receptor signaling, with upregulation of the PI3K-AKT pathway and increased cell proliferation *in vitro*.[114] However, pathogenic *PRLR* variants were not identified in an independent series of 88 prolactinoma patients.[115] Therefore, the role of *PRLR* variants in the pathogenesis of prolactinomas remains unclear. A further study described potentially pathogenic variants in the *RXRG* (1q23.3) and *TH* (11p15.5) genes, respectively encoding a retinoic acid receptor and tyrosine hydroxylase, in a single prolactinoma FIPA kindred.[116] The significance of these findings remains to be established.

Mutations in the immunoglobulin superfamily member 1 (*IGSF1*; Xq25) gene cause an X-linked syndrome of central hypothyroidism, prolactin deficiency, and macroorchidism.[117] Interestingly, male patients with IGSF1 deficiency present increased GH and IGF-1 levels.[118] Acromegalic facial features and increased head circumference are frequently observed in these patients. As these changes occur in the absence of PTs, it is likely that the loss of IGSF1, which is highly expressed in the hypothalamus and in pituitary cells of the POU1F1 lineage, affects the regulation of GH secretion at the hypothalamic or pituitary level.

Heterozygous missense *CDH23* gene (10q22.1) variants have been found to segregate with PTs in three FIPA families.[119] *CDH23* encodes a cadherin member previously implicated in the pathogenesis of a subtype of Usher syndrome, an autosomal recessive condition of hearing impairment, vestibular dysfunction, and retinitis pigmentosa. Potentially pathogenic *CDH23* variants were detected in 15 out of 125 patients with sporadic PTs of different subtypes. The mechanisms as to how *CDH23* mutations could lead to PTs remain unknown, and no functional studies have been performed. It should be noted that Usher syndrome patients or unaffected heterozygous carriers are not known to be at higher risk of developing PTs, and the polymorphic nature of *CDH23* can affect the interpretation of genetic variants in this gene. Hence, future studies will need to confirm the role of *CDH23* in PT development.

Potentially pathogenic missense *CABLES1* (18q11.2) variants were found in four out of 181 patients (2.2%) with Cushing disease.[120] All four patients had corticotroph macroadenomas with high proliferation index. CABLES1 is involved in the negative regulation of cell cycle progression in corticotroph cells in response to glucocorticoids.[121] *CABLES1* knockdown was found to stimulate the growth of a corticotroph cell line.[121] Moreover, CABLES1 protein expression was found to be frequently lost in corticotrophinomas, and this was associated with loss of p27.[121] Although these data support the role of CABLES1 in the pathogenesis of corticotrophinomas, further confirmatory studies will be necessary to assess the frequency of *CABLES1* mutations in patients with Cushing disease, as well as their potential role in the pathogenesis of other PT subtypes.

SOMATIC MUTATIONS IN PITUITARY TUMORS

Somatic mutations in “classical” oncogenes and tumor suppressor genes occur rarely in PTs. For instance, mutations in the *RAS* family of oncogenes have been very rarely described and seem to be limited to pituitary carcinomas and highly aggressive tumors. The most common somatic mutations occurring in PTs are represented by *GNAS* mutations in somatotrophinomas (10%–50%) and *USP8* (30%) mutations in corticotrophinomas. *GNAS*-mutated somatotrophinomas are often smaller, less invasive, and generally better responsive to SSAs compared with nonmutated tumors.[122-124]

Gain-of-function mutations in the deubiquitinase *USP8* (15q21.2) disrupt the interaction between USP8 and the 14-3-3 protein, allowing the cleavage of USP8 and promoting its enzymatic activity.[125,126] This results in increased expression of the epidermal growth factor receptor (EGFR), which promotes corticotroph cell proliferation and ACTH secretion. Somatic *USP8* mutations are associated with smaller tumors and, in some cohorts, they were detected more often in females.[127] Somatostatin receptor subtype 5 (SSTR5) was found to be more highly expressed in *USP8*-mutated tumors, suggesting that these tumors might respond better to pasireotide, a multireceptor SSA with high affinity for SSTR5.[128,129] EGFR inhibitors could potentially play a role in the treatment of *USP8*-mutated corticotrophinomas, and they have been shown to inhibit ACTH secretion *in vitro*.[126] Although virtually all *USP8* mutations are somatic, one patient with Cushing disease and a germline *USP8* mutation was recently described.[130] Notably, this patient had a complex phenotype with, among others, developmental delay, dysmorphic features, and dilated cardiomyopathy.

Mutations in the deubiquitinase gene *USP48* occur in 10% to 20% of *USP8*-wild type corticotrophinomas.[131,132] These mutations result in an increase in USP48 enzymatic activity, with activation of the sonic hedgehog pathway, and promote ACTH secretion in response to corticotrophin-releasing hormone *in vitro*.[132] Similarly to *USP8*, *USP48*-mutated tumors were smaller and had a tendency to occur more frequently in females.

The widespread use of pangenomic techniques has recently provided insights into the role of somatic mutations in the pathogenesis of PTs. Consistent with their low proliferation rate, PTs generally harbor a low number of somatic mutations, and, among these, only a few recurrent mutations have been identified.[65,66,133-136] Although *GNAS* mutations represent the only recurrent somatic mutation in somatotrophinomas, mutations in genes involved in calcium, cAMP, and adenosine triphosphate signaling have been described,[65,66] in keeping with the pivotal role of these pathways in the regulation of GH secretion. Recently, a recurrent hotspot mutation in the *SF3B1* gene has been identified in 20% of a series of prolactinomas investigated using whole-genome sequencing, and these findings were confirmed in an independent validation cohort.[137] The mutation was found to lead to the aberrant splicing of *ESRRG*, a member of the estrogen receptor-related receptor family, leading to increased pituitary-specific positive transcription factor 1 and prolactin expression. The mutation was found to occur more frequently in males and was associated with higher prolactin levels. Further data are needed to confirm these results.

Copy number alterations, including large chromosomal aberrations, occur frequently in PTs, especially in functioning tumors.[134] In corticotrophinomas, genomic instability was associated with larger tumors and a higher invasiveness rate.[138] In somatotrophinomas, chromosomal aberrations were found to occur more frequently in *GNAS* mutation–negative tumors, which were characterized by increased expression of PTTG1, a protein involved in sister chromatin separation and cell division.[139]

Significant advances have been made in the understanding of the driving mutations in craniopharyngiomas, brain tumors arising from remnants of the Rathke's pouch. The majority of adamantinomatous craniopharyngiomas harbor mutations in exon 3 of the *CTNNB1* gene causing activation of the β-catenin signaling pathway.[140] In contrast, the vast majority of papillary craniopharyngiomas harbor the *BRAF* V600E mutation.[141] These findings have therapeutic implications: case reports suggest that BRAF inhibitors can be effective in patients with recurrent *BRAF* V600E–positive craniopharyngiomas, and clinical trials are currently ongoing.[142-144]

CRITERIA FOR GENETIC SCREENING AND CLINICAL MONITORING

Although PTs are relatively common in the general population, only a small proportion of patients develop PTs as the consequence of a germline or mosaic mutation. Therefore, genetic testing is recommended in selected patients, including (1) patients with a family history of PTs, (2) patients with early-onset PTs, and (3) patients with a personal or family history of a syndromic condition known to predispose to PTs. It should be noted that guidelines only exist for the screening of MEN1 patients.[145] For other conditions, recommendations are based on expert opinion and the authors' review of the available evidence.

A detailed family history should be obtained from all PT patients, as it can allow a previously unrecognized history of familial PTs to be uncovered. Although it is possible that dedicated gene panels will become more available in the future, careful clinical phenotyping is highly important, as it can help select patients for genetic screening and guide testing strategies. A proposed algorithm is shown in Fig. 14.3.

Screening for *AIP* mutations should be considered in all FIPA families, as mutations are identified in about 10% of unselected families and 20% of those with familial acromegaly. Owing to the incomplete penetrance, screening for *AIP* mutations should also be considered in patients with early-onset somatotroph or lactotroph macroadenomas. As the vast majority of *AIP*-related PTs present during the second and third decades of life, it is generally recommended to screen all patients with age at onset of 18 years or younger and patients with macroadenomas and age at onset of between 19 and 30 years. The probabilities of identifying an *AIP* mutation vary based on the PT subtype and age at onset and are higher in patients with early-onset somatotrophinomas as about 30% of patients presenting clinically with gigantism harbor an *AIP* mutation. Cascade genetic testing should be offered to relatives at risk, and carriers should be monitored periodically. This is particularly relevant as, despite the low penetrance in these families, early diagnosis of *AIP*-related PTs is associated with improved treatment outcomes.[4] Modalities and the age when screening should start in *AIP* mutation carriers remain to be established, but it is reasonable to monitor growth starting in early childhood (≥3 years) and consider yearly biochemical testing (GH, IGF-1, and prolactin) starting at the age of 10 years.[146] In the absence of abnormalities, baseline pituitary magnetic resonance imaging (MRI) in *AIP* mutation carriers is advised around the age of 10 years and should be repeated every 5 years until the age of 30 years. Adults aged younger than 30 years should undergo a baseline pituitary MRI with yearly clinical and biochemical assessments. As no new diagnoses have been reported in *AIP* mutation carriers with normal biochemical and MRI assessment up to the age of 30 years, the screening frequency can be reduced after the age of 30 years, with consideration to stop screening at the age of 50 years. A proposed strategy for screening and monitoring relatives of *AIP* mutation carriers is summarized in Fig. 14.4.

Because most FIPA families do not have an identifiable genetic mutation, it would be difficult to establish recommendations for the follow-up of family members of *AIP* mutation-negative FIPA families. Penetrance in these families is low, and carriers cannot be identified by genetic screening. Therefore, the probability of diagnosing a PT in apparently unaffected family members is low.

XLAG owing to *GPR101* duplication presents with early-onset gigantism with significant overgrowth starting during the first 2 years of life. A careful evaluation of the timing of the first signs of the disease is important in patients with gigantism, as it can allow the identification of patients who should be tested for XLAG. It should be noted that standard chromosome microarrays might miss smaller or somatic mosaic duplications, and highly sensitive gene dosage techniques, such as digital droplet PCR, should be employed when testing for XLAG. Testing of tumor-derived DNA may be necessary in some cases to confirm a somatic mosaic mutation. XLAG should also be considered in *AIP* mutation-negative FIPA families with X-linked inheritance of early-onset gigantism.

Testing for genes responsible for syndromic conditions should be undertaken in patients with extrapituitary manifestations who fulfill the diagnostic criteria for a specific syndrome and in patients with a family history suggestive of an inheritable syndrome known to predispose to PTs.

Testing for *MEN1* should be considered in patients with a PT in case of a personal or family history of primary hyperparathyroidism and/or gastroenteropancreatic NETs. A negative family history does not rule out the possibility of *MEN1* mutations, as approximately 10% of cases occur as a result of *de novo* mutations. Genetic screening should be extended to relatives at risk, and lifelong clinical and biochemical follow-up should be offered to all mutation carriers. Screening for PTs is advised in children starting at the age of 5 years with annual clinical and biochemical assessments (prolactin and IGF-1). A pituitary MRI should be performed every 3 years.[145] In subjects with a clinical diagnosis of MEN1 who test negative for *MEN1* mutations, the possibility of MEN4 should be considered, and these patients should be tested for *CDKN1B* mutations.

Patients with clinical features in keeping with CNC should be tested for *PRKAR1A* mutations. Such mutations are identified in about 70% of CNC patients. Testing for *PRKACB* amplifications could be considered in the absence of *PRKAR1A* mutations. In patients with a clinical picture suggestive of MAS, a clinical diagnosis is generally made on the basis of the typical manifestations and can be confirmed by *GNAS* sequencing. Because these mutations occur in a mosaic state, the diagnostic yield is low when testing DNA isolated from peripheral blood, although highly sensitive techniques can improve the diagnostic rate. Patients with a personal or family history of paraganglioma or pheochromocytoma should be tested for *SDHx* mutations, and, considering the recently described association with *MAX* gene mutations, *MAX* gene sequencing, including testing for large deletions, could be considered in patients who do not harbor pathogenic mutations in the *SDHx* genes.

Pituitary blastomas represent a pathognomonic feature of DICER1 syndrome, therefore children with these rare tumors should be tested for *DICER1* mutations, and screening should be extended to first-degree relatives. Recommended surveillance strategies for DICER1 syndrome have been recently proposed.[147]

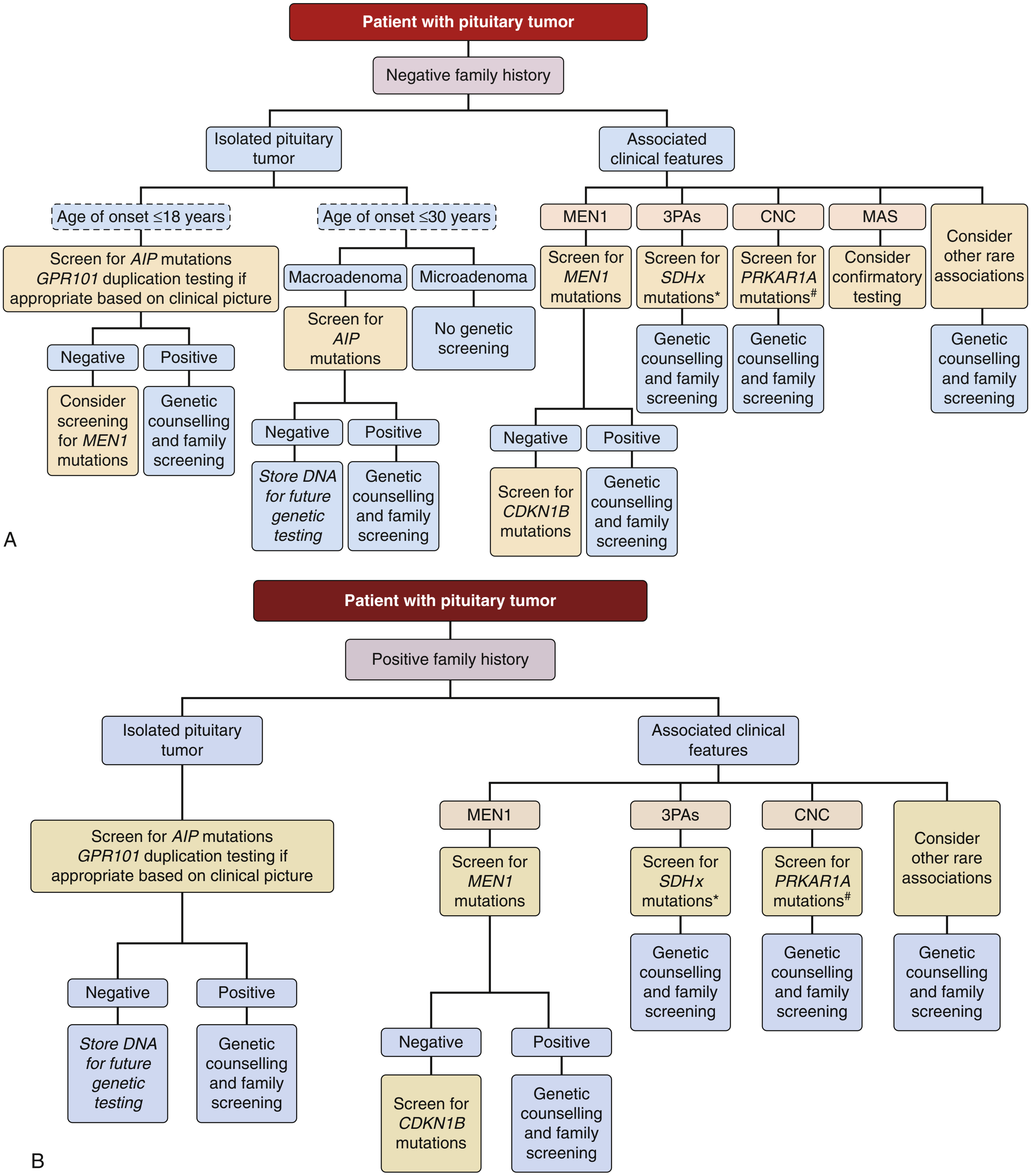

Patient with pituitary tumor
Negative family history
Isolated pituitary tumor
Associated clinical features
Age of onset ≤18 years
Age of onset ≤30 years
MEN1
3PAs
CNC
MAS
Consider other rare associations
Screen for AIP mutations GPR101 duplication testing if appropriate based on clinical picture
Macroadenoma
Microadenoma
Screen for MEN1 mutations
Screen for SDHx mutations*
Screen for PRKAR1A mutations#
Consider confirmatory testing
Screen for AIP mutations
No genetic screening
Genetic counselling and family screening
Genetic counselling and family screening
Genetic counselling and family screening
Negative
Positive
Consider screening for MEN1 mutations
Genetic counselling and family screening
Negative
Positive
Negative
Positive
Store DNA for future genetic testing
Genetic counselling and family screening
Screen for CDKN1B mutations
Genetic counselling and family screening
A
Patient with pituitary tumor
Positive family history
Isolated pituitary tumor
Associated clinical features
MEN1
3PAs
CNC
Screen for AIP mutations GPR101 duplication testing if appropriate based on clinical picture
Screen for MEN1 mutations
Screen for SDHx mutations*
Screen for PRKAR1A mutations#
Consider other rare associations
Genetic counselling and family screening
Genetic counselling and family screening
Genetic counselling and family screening
Negative
Positive
Store DNA for future genetic testing
Genetic counselling and family screening
Negative
Positive
Screen for CDKN1B mutations
Genetic counselling and family screening
B

Fig. 14.3 Algorithm for genetic screening in patients with a negative (**A**) or positive (**B**) family history of pituitary tumors. *CNC*, Carney complex; *MAS*, McCune–Albright syndrome; *MEN1*, multiple endocrine neoplasia type 1; *3PAs*, pheochromocytoma/paraganglioma with pituitary adenoma. *Screening for *MAX* mutations (including large deletions) could be considered in patients with 3PAs and no *SDHx* mutations; #screening for *PRKACB* mutations (including copy number variants) could be considered in patients with CNC and no *PRKAR1A* mutations.

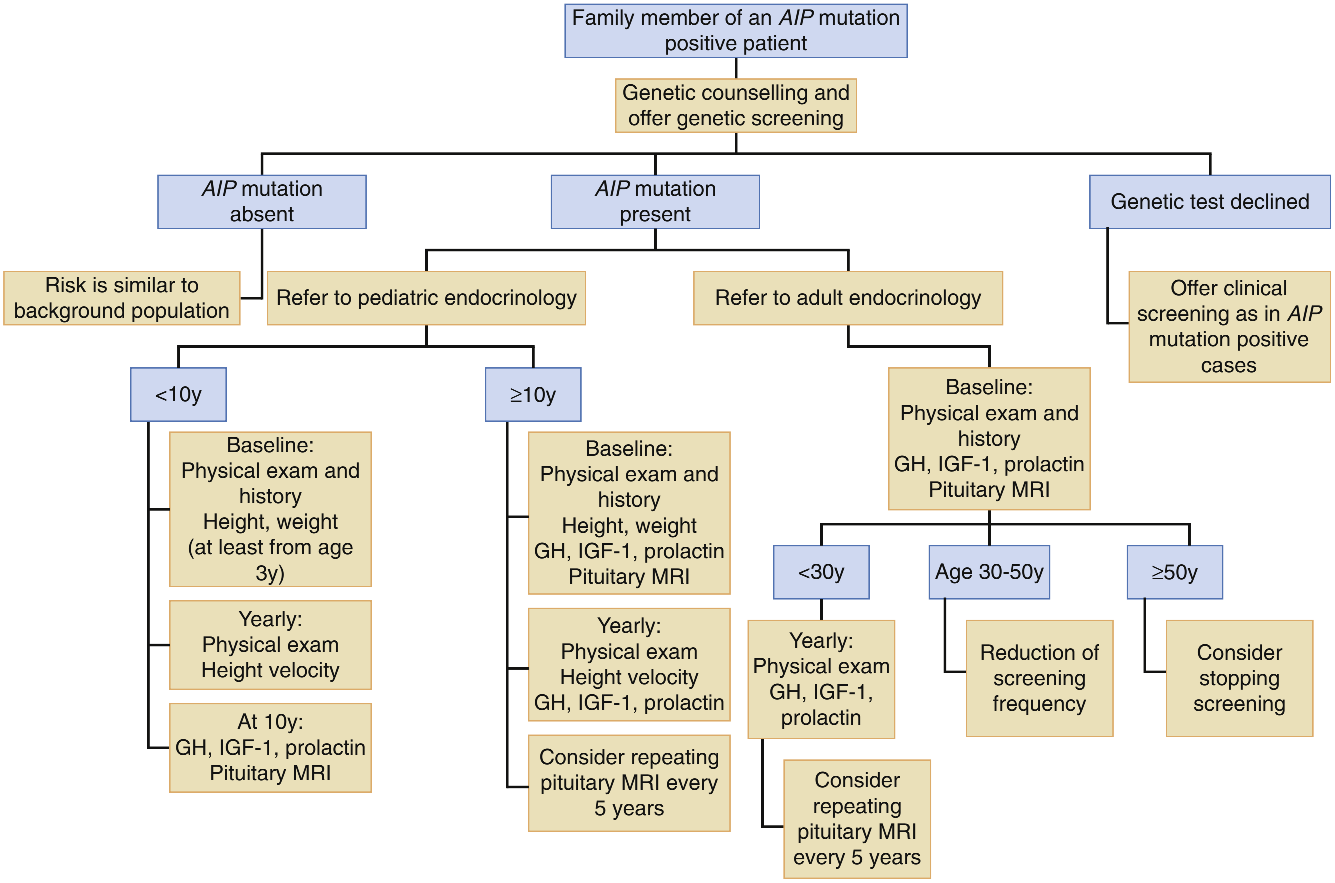

Family member of an *AIP* mutation positive patient
Genetic counselling and offer genetic screening
AIP mutation absent
AIP mutation present
Genetic test declined
Risk is similar to background population
Refer to pediatric endocrinology
Refer to adult endocrinology
Offer clinical screening as in *AIP* mutation positive cases
<10y
≥10y
Baseline: Physical exam and history Height, weight (at least from age 3y)
Yearly: Physical exam Height velocity
At 10y: GH, IGF-1, prolactin Pituitary MRI
Baseline: Physical exam and history Height, weight GH, IGF-1, prolactin Pituitary MRI
Yearly: Physical exam Height velocity GH, IGF-1, prolactin
Consider repeating pituitary MRI every 5 years
Baseline: Physical exam and history GH, IGF-1, prolactin Pituitary MRI
<30y
Age 30-50y
≥50y
Yearly: Physical exam GH, IGF-1, prolactin
Consider repeating pituitary MRI every 5 years
Reduction of screening frequency
Consider stopping screening

Fig. 14.4 Proposed strategy for genetic screening and monitoring of relatives of *AIP* mutation carriers. *GH,* Growth hormone; *IGF-1,* insulin-like growth factor 1; *MRI,* magnetic resonance imaging.

CONCLUSION

Significant advances have been made in the field of pituitary genetics in recent years, and a number of genetic mutations have been associated with an increased risk of developing PTs. Genetically-determined PTs can occur without other associated manifestations as FIPA: so far two genes, *AIP* and very rarely *GPR101*, have been identified to be responsible for this condition, whereas the causative gene(s) in the majority of FIPA families remain to be identified. PTs can also occur as part of syndromic conditions, and, in some cases, they can represent the first manifestation of the disease. Genetically-determined PTs often present at a younger age, can show aggressive behavior with rapid growth and are often resistant to treatment. Although the cost-effectiveness of genetic screening for some conditions is difficult to demonstrate owing to their rarity, the identification of the disease-causing mutations in patients with PTs offers several advantages. Following the identification of a causative mutation, cascade genetic testing allows the identification of carriers who are at increased risk of developing PTs as well as, in the case of syndromic condition, other neoplasms. These subjects will benefit from an early diagnosis as the result of focused monitoring strategies. Moreover, identifying a causative mutation helps inform clinicians of the clinical features of that specific condition, resulting in prompt initiation of effective treatments and earlier diagnosis of associated manifestations. Lastly, the discovery of causative genetic mutations offers the chance to broaden our understanding of the mechanisms underlying PT development.

REFERENCES

1. Beckers A, Aaltonen LA, Daly AF, et al. Familial isolated pituitary adenomas (FIPA) and the pituitary adenoma predisposition due to mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene. *Endocr Rev*. 2013;34:239–277.
2. Daly AF, Rixhon M, Adam C, et al. High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium. *J Clin Endocrinol Metab*. 2006;91:4769–4775.
3. Vierimaa O, Georgitsi M, Lehtonen R, et al. Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science*. 2006;312:1228–1230.
4. Marques P, Caimari F, Hernandez-Ramirez LC, et al. Significant benefits of AIP testing and clinical screening in familial isolated and young-onset pituitary tumors. *J Clin Endocrinol Metab*. 2020;105:e2247–e2260.
5. Tichomirowa MA, Barlier A, Daly AF, et al. High prevalence of AIP gene mutations following focused screening in young patients with sporadic pituitary macroadenomas. *Eur J Endocrinol*. 2011;165:509–515.
6. Hernandez-Ramirez LC, Gabrovska P, Denes J, et al. Landscape of familial isolated and young-onset pituitary adenomas: prospective diagnosis in AIP mutation carriers. *J Clin Endocrinol Metab*. 2015;100:E1242–E1254.
7. Cuny T, Pertuit M, Sahnoun-Fathallah M, et al. Genetic analysis in young patients with sporadic pituitary macroadenomas: besides AIP don't forget MEN1 genetic analysis. *Eur J Endocrinol*. 2013;168:533–541.
8. Trivellin G, Daly AF, Faucz FR, et al. Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. *N Engl J Med*. 2014;371:2363–2374.

9. Wise-Oringer BK, Zanazzi GJ, Gordon RJ, et al. Familial X-linked acrogigantism: postnatal outcomes and tumor pathology in a prenatally diagnosed infant and his mother. *J Clin Endocrinol Metab*. 2019;104: 4667–4675.
10. Korbonits M, Kumar AV. AIP familial isolated pituitary adenomas. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. Seattle: GeneReviews; 2020.
11. Daly AF, Vanbellinghen JF, Khoo SK, et al. Aryl hydrocarbon receptor-interacting protein gene mutations in familial isolated pituitary adenomas: analysis in 73 families. *J Clin Endocrinol Metab*. 2007;92:1891–1896.
12. Daly AF, Tichomirowa MA, Petrossians P, et al. Clinical characteristics and therapeutic responses in patients with germ-line AIP mutations and pituitary adenomas: an international collaborative study. *J Clin Endocrinol Metab*. 2010;95:E373–E383.
13. Igreja S, Chahal HS, King P, et al. Characterization of aryl hydrocarbon receptor interacting protein (AIP) mutations in familial isolated pituitary adenoma families. *Hum Mutat*. 2010;31:950–960.
14. Leontiou CA, Gueorguiev M, van der Spuy J, et al. The role of the aryl hydrocarbon receptor-interacting protein gene in familial and sporadic pituitary adenomas. *J Clin Endocrinol Metab*. 2008;93:2390–2401.
15. Rostomyan L, Daly AF, Petrossians P, et al. Clinical and genetic characterization of pituitary gigantism: an international collaborative study in 208 patients. *Endocr Relat Cancer*. 2015;22:745–757.
16. Stiles CE, Korbonits M. Familial isolated pituitary adenoma. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext*. South Dartmouth; 2020.
17. Mangupli R, Rostomyan L, Castermans E, et al. Combined treatment with octreotide LAR and pegvisomant in patients with pituitary gigantism: clinical evaluation and genetic screening. *Pituitary*. 2016;19:507–514.
18. Daly AF, Rostomyan L, Betea D, et al. AIP-mutated acromegaly resistant to first-generation somatostatin analogs: long-term control with pasireotide LAR in two patients. *Endocr Connect*. 2019;8:367–377.
19. Carty D, Harte R, Drummond RS, et al. *AIP* variant causing familial prolactinoma. *Pituitary*. 2021;24:48–52.
20. de Laat JM, Dekkers OM, Pieterman CR, et al. Long-term natural course of pituitary tumors in patients with MEN1: results from the DutchMEN1 Study Group (DMSG). *J Clin Endocrinol Metab*. 2015;100:3288–3296.
21. Radian S, Diekmann Y, Gabrovska P, et al. Increased population risk of AIP-related acromegaly and gigantism in Ireland. *Hum Mutat*. 2017;38:78–85.
22. Salvatori R, Radian S, Diekmann Y, et al. In-frame seven amino-acid duplication in AIP arose over the last 3000 years, disrupts protein interaction and stability and is associated with gigantism. *Eur J Endocrinol*. 2017;177:257–266.
23. Hernandez-Ramirez LC, Martucci F, Morgan RM, et al. Rapid proteasomal degradation of mutant proteins is the primary mechanism leading to tumorigenesis in patients with missense AIP mutations. *J Clin Endocrinol Metab*. 2016;101:3144–3154.
24. Tuominen I, Heliovaara E, Raitila A, et al. AIP inactivation leads to pituitary tumorigenesis through defective Galphai-cAMP signaling. *Oncogene*. 2015;34:1174–1184.
25. Chahal HS, Trivellin G, Leontiou CA, et al. Somatostatin analogs modulate AIP in somatotroph adenomas: the role of the ZAC1 pathway. *J Clin Endocrinol Metab*. 2012;97:E1411–E1420.
26. García-Rendueles AR, Chenlo M, Oroz-Gonjar F, et al. RET signalling provides tumorigenic mechanism and tissue specificity for AIP-related somatotrophinomas. *Oncogene*. 2021. https://doi.org/10.1038/s41388-021-02009-8.
27. Barry S, Carlsen E, Marques P, et al. Tumor microenvironment defines the invasive phenotype of AIP-mutation-positive pituitary tumors. *Oncogene*. 2019;38:5381–5395.
28. Iacovazzo D, Caswell R, Bunce B, et al. Germline or somatic GPR101 duplication leads to X-linked acrogigantism: a clinico-pathological and genetic study. *Acta Neuropathol Commun*. 2016;4:56.
29. Iacovazzo D, Korbonits M. X-linked acrogigantism. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, et al., eds. Seattle: GeneReviews; 2018.
30. Beckers A, Lodish MB, Trivellin G, et al. X-linked acrogigantism syndrome: clinical profile and therapeutic responses. *Endocr Relat Cancer*. 2015;22:353–367.
31. Rodd C, Millette M, Iacovazzo D, et al. Somatic GPR101 duplication causing X-linked acrogigantism (XLAG)-Diagnosis and management. *J Clin Endocrinol Metab*. 2016;101:1927–1930.
32. Gordon RJ, Bell J, Chung WK, et al. Childhood acromegaly due to X-linked acrogigantism: long term follow-up. *Pituitary*. 2016;19:560–564.
33. Beckers A, Fernandes D, Fina F, et al. Paleogenetic study of ancient DNA suggestive of X-linked acrogigantism. *Endocr Relat Cancer*. 2017;24:L17–L20.
34. Liang H, Gong F, Liu Z, et al. A Chinese case of X-linked acrogigantism and systematic review. *Neuroendocrinology*. 2020. https://doi.org/10.1159/000512240.
35. Bates B, Zhang L, Nawoschik S, et al. Characterization of Gpr101 expression and G-protein coupling selectivity. *Brain Res*. 2006;1087:1–14.
36. Trivellin G, Bjelobaba I, Daly AF, et al. Characterization of GPR101 transcript structure and expression patterns. *J Mol Endocrinol*. 2016;57:97–111.
37. Flak MB, Koenis DS, Sobrino A, et al. GPR101 mediates the pro-resolving actions of RvD5n-3 DPA in arthritis and infections. *J Clin Invest*. 2020;130:359–373.
38. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). *Mol Cell Endocrinol*. 2014;386:2–15.
39. Marini F, Falchetti A, Del Monte F, et al. Multiple endocrine neoplasia type 1. *Orphanet J Rare Dis*. 2006;1:38.
40. Dreijerink KM, Goudet P, Burgess JR, et al. Breast-cancer predisposition in multiple endocrine neoplasia type 1. *N Engl J Med*. 2014;371:583–584.
41. Verges B, Boureille F, Goudet P, et al. Pituitary disease in MEN type 1 (MEN1): data from the France-Belgium MEN1 multicenter study. *J Clin Endocrinol Metab*. 2002;87:457–465.
42. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat*. 2008;29:22–32.
43. Concolino P, Costella A, Capoluongo E. Multiple endocrine neoplasia type 1 (MEN1): an update of 208 new germline variants reported in the last nine years. *Cancer Genet*. 2016;209:36–41.
44. Yokoyama A, Cleary ML. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Canc Cell*. 2008;14:36–46.
45. Wu T, Hua X. Menin represses tumorigenesis via repressing cell proliferation. *Am J Cancer Res*. 2011;1:726–739.
46. Milne TA, Hughes CM, Lloyd R, et al. Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci USA*. 2005;102:749–754.
47. Alrezk R, Hannah-Shmouni F, Stratakis CA. MEN4 and CDKN1B mutations: the latest of the MEN syndromes. *Endocr Relat Cancer*. 2017;24:T195–T208.
48. Frederiksen A, Rossing M, Hermann P, et al. Clinical features of multiple endocrine neoplasia type 4: novel pathogenic variant and review of published cases. *J Clin Endocrinol Metab*. 2019;104:3637–3646.
49. Chasseloup F, Pankratz N, Lane J, et al. Germline CDKN1B loss-of-function variants cause pediatric Cushing's disease with or without an MEN4 phenotype. *J Clin Endocrinol Metab*. 2020;105:1983–2005.
50. Tichomirowa MA, Lee M, Barlier A, et al. Cyclin-dependent kinase inhibitor 1B (CDKN1B) gene variants in AIP mutation-negative familial isolated pituitary adenoma kindreds. *Endocr Relat Cancer*. 2012;19:233–241.
51. Sambugaro S, Di Ruvo M, Ambrosio MR, et al. Early onset acromegaly associated with a novel deletion in CDKN1B 5'UTR region. *Endocrine*. 2015;49:58–64.
52. Carney JA, Gordon H, Carpenter PC, et al. The complex of myxomas, spotty pigmentation, and endocrine overactivity. *Medicine (Baltim)*. 1985;64:270–283.
53. Correa R, Salpea P, Stratakis CA. Carney complex: an update. *Eur J Endocrinol*. 2015;173:M85–M97.
54. Bertherat J, Horvath A, Groussin L, et al. Mutations in regulatory subunit type 1A of cyclic adenosine 5'-monophosphate-dependent protein kinase (PRKAR1A): phenotype analysis in 353 patients and 80 different genotypes. *J Clin Endocrinol Metab*. 2009;94:2085–2091.
55. Stratakis CA, Matyakhina L, Courkoutsakis N, et al. Pathology and molecular genetics of the pituitary gland in patients with the 'complex of spotty skin pigmentation, myxomas, endocrine overactivity and schwannomas' (Carney complex). *Front Horm Res*. 2004;32:253v264.

56. Stergiopoulos SG, Abu-Asab MS, Tsokos M, et al. Pituitary pathology in Carney complex patients. *Pituitary*. 2004;7:73–82.
57. Kiefer FW, Winhofer Y, Iacovazzo D, et al. PRKAR1A mutation causing pituitary-dependent Cushing disease in a patient with Carney complex. *Eur J Endocrinol*. 2017;177:K7–K12.
58. Hernandez-Ramirez LC, Tatsi C, Lodish MB, et al. Corticotropinoma as a component of Carney complex. *J Endocr Soc*. 2017;1:918–925.
59. Stratakis CA, Raygada M. Carney complex. In: Pagon RA, Adam MP, Ardinger HH, et al., eds. Seattle: GeneReviews; 2018.
60. Salpea P, Horvath A, London E, et al. Deletions of the PRKAR1A locus at 17q24.2-q24.3 in Carney complex: genotype-phenotype correlations and implications for genetic testing. *J Clin Endocrinol Metab*. 2014;99:E183–E188.
61. Nadella KS, Kirschner LS. Disruption of protein kinase a regulation causes immortalization and dysregulation of D-type cyclins. *Cancer Res*. 2005;65:10307–10315.
62. Mavrakis M, Lippincott-Schwartz J, Stratakis CA, et al. Depletion of type IA regulatory subunit (RIalpha) of protein kinase A (PKA) in mammalian cells and tissues activates mTOR and causes autophagic deficiency. *Hum Mol Genet*. 2006;15:2962–2971.
63. Forlino A, Vetro A, Garavelli L, et al. PRKACB and Carney complex. *N Engl J Med*. 2014;370:1065–1067.
64. Stratakis CA. Carney complex: a familial lentiginosis predisposing to a variety of tumors. *Rev Endocr Metab Disord*. 2016;17:367–371.
65. Valimaki N, Demir H, Pitkanen E, et al. Whole-genome sequencing of growth hormone (GH)-secreting pituitary adenomas. *J Clin Endocrinol Metab*. 2015;100:3918–3927.
66. Ronchi CL, Peverelli E, Herterich S, et al. Landscape of somatic mutations in sporadic GH-secreting pituitary adenomas. *Eur J Endocrinol*. 2016;174:363–372.
67. Boyce AM, Florenzano P, de Castro LF, et al. Fibrous dysplasia/McCune-Albright syndrome. In: Pagon RA, Adam MP, Ardinger HH, et al., eds. Seattle: GeneReviews; 2019.
68. Elli FM, de Sanctis L, Bergallo M, et al. Improved molecular diagnosis of McCune-Albright syndrome and bone fibrous dysplasia by digital PCR. *Front Genet*. 2019;10:862.
69. Hayward BE, Barlier A, Korbonits M, et al. Imprinting of the G(s) alpha gene GNAS1 in the pathogenesis of acromegaly. *J Clin Invest*. 2001;107:R31–R36.
70. Mantovani G, Ballare E, Giammona E, et al. The gsalpha gene: predominant maternal origin of transcription in human thyroid gland and gonads. *J Clin Endocrinol Metab*. 2002;87:4736–4740.
71. Salenave S, Boyce AM, Collins MT, et al. Acromegaly and McCune-Albright syndrome. *J Clin Endocrinol Metab*. 2014;99:1955–1969.
72. Galland F, Kamenicky P, Affres H, et al. McCune-Albright syndrome and acromegaly: effects of hypothalamopituitary radiotherapy and/or pegvisomant in somatostatin analog-resistant patients. *J Clin Endocrinol Metab*. 2006;91:4957–4961.
73. Akintoye SO, Kelly MH, Brillante B, et al. Pegvisomant for the treatment of gsp-mediated growth hormone excess in patients with McCune-Albright syndrome. *J Clin Endocrinol Metab*. 2006;91:2960–2966.
74. Xekouki P, Pacak K, Almeida M, et al. Succinate dehydrogenase (SDH) D subunit (SDHD) inactivation in a growth-hormone-producing pituitary tumor: a new association for SDH? *J Clin Endocrinol Metab*. 2012;97:E357–E366.
75. Dwight T, Mann K, Benn DE, et al. Familial SDHA mutation associated with pituitary adenoma and pheochromocytoma/paraganglioma. *J Clin Endocrinol Metab*. 2013;98:E1103–E1108.
76. Xekouki P, Szarek E, Bullova P, et al. Pituitary adenoma with paraganglioma/pheochromocytoma (3PAs) and succinate dehydrogenase defects in humans and mice. *J Clin Endocrinol Metab*. 2015;100:E710–E719.
77. Denes J, Swords F, Rattenberry E, et al. Heterogeneous genetic background of the association of pheochromocytoma/paraganglioma and pituitary adenoma: results from a large patient cohort. *J Clin Endocrinol Metab*. 2015;100:E531–E541.
78. Maher M, Roncaroli F, Mendoza N, et al. A patient with a germline SDHB mutation presenting with an isolated pituitary macroprolactinoma. *Endocrinol Diabetes Metab Case Rep*. 2018;18-0078.
79. Lemelin A, Lapoirie M, Abeillon J, et al. Pheochromocytoma, paragangliomas, and pituitary adenoma: an unusual association in a patient with an SDHD mutation. *Case report. Medicine (Baltimore)*. 2019;98:e16594.
80. Tufton N, Roncaroli F, Hadjidemetriou I, et al. Pituitary carcinoma in a patient with an SDHB mutation. *Endocr Pathol*. 2017;28:320–325.
81. Zhao T, Mu X, You Q. Succinate: an initiator in tumorigenesis and progression. *Oncotarget*. 2017;8:53819–53828.
82. Aspuria PP, Lunt SY, Varemo L, et al. Succinate dehydrogenase inhibition leads to epithelial-mesenchymal transition and reprogrammed carbon metabolism. *Cancer Metab*. 2014;2:21.
83. Roszko KL, Blouch E, Blake M, et al. Case Report of a prolactinoma in a patient with a novel MAX mutation and bilateral pheochromocytomas. *J Endocr Soc*. 2017;1:1401–1407.
84. Daly AF, Castermans E, Oudijk L, et al. Pheochromocytomas and pituitary adenomas in three patients with MAX exon deletions. *Endocr Relat Cancer*. 2018;25:L37–L42.
85. Mamedova E, Vasilyev E, Petrov V, et al. Familial acromegaly and bilateral asynchronous pheochromocytomas in a female patient with a MAX mutation: a case report. *Front Endocrinol*. 2021;12:683492.
86. Seabrook AJ, Harris JE, Velosa SB, et al. Multiple endocrine tumors associated with germline MAX Mutations: multiple endocrine neoplasia type 5? *J Clin Endocrinol Metab*. 2021;106:1163–1182.
87. Scheithauer BW, Kovacs K, Horvath E, et al. Pituitary blastoma. *Acta Neuropathol*. 2008;116:657–666.
88. de Kock L, Sabbaghian N, Plourde F, et al. Pituitary blastoma: a pathognomonic feature of germ-line DICER1 mutations. *Acta Neuropathol*. 2014;128:111–122.
89. Heravi-Moussavi A, Anglesio MS, Cheng SW, et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. *N Engl J Med*. 2012;366:234–242.
90. Brenneman M, Field A, Yang J, et al. Temporal order of RNase IIIb and loss-of-function mutations during development determines phenotype in pleuropulmonary blastoma / DICER1 syndrome: a unique variant of the two-hit tumor suppression model. *F1000Res*. 2015;4:214.
91. de Kock L, Bah I, Brunet J, et al. Somatic DICER1 mutations in adult-onset pulmonary blastoma. *Eur Respir J*. 2016;47:1879–1882.
92. Bengtsson D, Joost P, Aravidis C, et al. Corticotroph pituitary carcinoma in a patient with Lynch syndrome (LS) and pituitary tumors in a nationwide LS cohort. *J Clin Endocrinol Metab*. 2017;102:3928–3932.
93. Uraki S, Ariyasu H, Doi A, et al. Atypical pituitary adenoma with MEN1 somatic mutation associated with abnormalities of DNA mismatch repair genes; MLH1 germline mutation and MSH6 somatic mutation. *Endocr J*. 2017;64:895–906.
94. Voisin MR, Almeida JP, Perez-Ordonez B, et al. Recurrent undifferentiated carcinoma of the sella in a patient with Lynch syndrome. *World Neurosurg*. 2019;132:219–222.
95. Loughrey PB, Baker G, Herron B, et al. Invasive ACTH-producing pituitary gland neoplasm secondary to MSH2 mutation. *Cancer Genet*. 2021:256–257. 36–39.
96. Friedman JM. Neurofibromatosis 1. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. Seattle: GeneReviews; 2019.
97. Cambiaso P, Galassi S, Palmiero M, et al. Growth hormone excess in children with neurofibromatosis type-1 and optic glioma. *Am J Med Genet*. 2017;173:2353–2358.
98. Smith N, Santoreneos S. Non-functioning pituitary macroadenoma in a child with neurofibromatosis type 1. *ANZ J Surg*. 2017;87:E220–E221.
99. Pinnamaneni K, Birge SJ, Avioli LV. Prolactin-secreting pituitary tumor associated with von Recklinghausen's disease. *Arch Intern Med*. 1980;140:397–399.
100. Adeloye A. Coexistence of acromegaly and neurofibromatosis in a Nigerian. *East Afr Med J*. 1979;56:38–39.
101. Hozumi K, Fukuoka H, Odake Y, et al. Acromegaly caused by a somatotroph adenoma in patient with neurofibromatosis type 1. *Endocr J*. 2019;66:853–857.
102. Barberis M, Gambacorta M, Versari P, et al. About a case of Recklinghausen's disease associated with pituitary adenoma. *Pathologica. 1979;* 71:265–272.

103. Boudin G, Pepin B, Vernant CL. Multiple tumours of the nervous system in Recklinghausen's disease. An anatomo-clinical case with chromophobe adenoma of the pituitary gland. *Presse Med*. 1970;78:1427–1430.
104. Nakajima M, Nakasu Y, Nakasu S, et al. Pituitary adenoma associated with neurofibromatosis: case report. *Nihon Geka Hokan*. 1990;59:278–282.
105. Kurozumi K, Tabuchi A, Ono Y, et al. Pituitary adenoma associated with neurofibromatosis type 1: case report. *Noshinkeigeka*. 2002;30:741–745.
106. Northrup H, Koenig MK, Pearson DA, et al. Tuberous sclerosis complex. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. Seattle: GeneReviews; 2020.
107. Regazzo D, Gardiman MP, Theodoropoulou M, et al. Silent gonadotroph pituitary neuroendocrine tumor in a patient with tuberous sclerosis complex: evaluation of a possible molecular link. *Endocrinol Diabetes Metab Case Rep*. 2018;18-0086.
108. Hoffman WH, Perrin JC, Halac E, et al. Acromegalic gigantism and tuberous sclerosis. *J Pediatr*. 1978;93:478–480.
109. Tigas S, Carroll PV, Jones R, et al. Simultaneous Cushing's disease and tuberous sclerosis; a potential role for TSC in pituitary ontogeny. *Clin Endocrinol*. 2005;63:694–695.
110. Nandagopal R, Vortmeyer A, Oldfield EH, et al. Cushing's syndrome due to a pituitary corticotropinoma in a child with tuberous sclerosis: an association or a coincidence? *Clin Endocrinol*. 2007;67:639–641.
111. Ilgren EB, Westmoreland D. Tuberous sclerosis: unusual associations in four cases. *J Clin Pathol*. 1984;37:272–278.
112. Galaction-Nitelea O, Dociu I, Murgu V. A case of tuberous sclerosis with acromegaly. *Rev Med Interna Neurol Psihiatr Neurochir Dermatovenerol Neurol Psihiatr Neurochir*. 1978;23:253–262.
113. Newey PJ, Gorvin CM, Cleland SJ, et al. Mutant prolactin receptor and familial hyperprolactinemia. *N Engl J Med*. 2013;369:2012–2020.
114. Gorvin CM, Newey PJ, Rogers A, et al. Association of prolactin receptor (PRLR) variants with prolactinomas. *Hum Mol Genet*. 2019;28:1023–1037.
115. Bernard V, Bouilly J, Beau I, et al. Germline prolactin receptor mutation is not a major cause of sporadic prolactinoma in humans. *Neuroendocrinology*. 2016;103:738–745.
116. Melo FM, Couto PP, Bale AE, et al. Whole-exome identifies RXRG and TH germline variants in familial isolated prolactinoma. *Cancer Genet*. 2016;209:251–257.
117. Joustra SD, Heinen CA, Schoenmakers N, et al. IGSF1 deficiency: lessons from an extensive case series and recommendations for clinical management. *J Clin Endocrinol Metab*. 2016;101:1627–1636.
118. Joustra SD, Roelfsema F, van Trotsenburg ASP, et al. IGSF1 deficiency results in human and murine somatotrope neurosecretory hyperfunction. *J Clin Endocrinol Metab*. 2020;105:e70–e84.
119. Zhang Q, Peng C, Song J, et al. Germline mutations in CDH23, encoding cadherin-related 23, are associated with both familial and sporadic pituitary adenomas. *Am J Hum Genet*. 2017;100:817–823.
120. Hernandez-Ramirez LC, Gam R, Valdes N, et al. Loss-of-function mutations in the CABLES1 gene are a novel cause of Cushing's disease. *Endocr Relat Cancer*. 2017;24:379–392.
121. Roussel-Gervais A, Couture C, Langlais D, et al. The Cables1 gene in glucocorticoid regulation of pituitary corticotrope growth and Cushing disease. *J Clin Endocrinol Metab*. 2016;101:513–522.
122. Efstathiadou ZA, Bargiota A, Chrisoulidou A, et al. Impact of gsp mutations in somatotroph pituitary adenomas on growth hormone response to somatostatin analogs: a meta-analysis. *Pituitary*. 2015;18:861–867.
123. Landis CA, Harsh G, Lyons J, et al. Clinical characteristics of acromegalic patients whose pituitary tumors contain mutant Gs protein. *J Clin Endocrinol Metab*. 1990;71:1416–1420.
124. Buchfelder M, Fahlbusch R, Merz T, et al. Clinical correlates in acromegalic patients with pituitary tumors expressing GSP oncogenes. *Pituitary*. 1999;1:181–185.
125. Reincke M, Sbiera S, Hayakawa A, et al. Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet*. 2015;47:31–38.
126. Ma ZY, Song ZJ, Chen JH, et al. Recurrent gain-of-function USP8 mutations in Cushing's disease. *Cell Res*. 2015;25:306–317.
127. Wanichi IQ, de Paula Mariani BM, Frassetto FP, et al. Cushing's disease due to somatic USP8 mutations: a systematic review and meta-analysis. *Pituitary*. 2019;22:435–442.
128. Castellnou S, Vasiljevic A, Lapras V, et al. SST5 expression and USP8 mutation in functioning and silent corticotroph pituitary tumors. *Endocr Connect*. 2020;9:243–253.
129. Hayashi K, Inoshita N, Kawaguchi K, et al. The USP8 mutational status may predict drug susceptibility in corticotroph adenomas of Cushing's disease. *Eur J Endocrinol*. 2016;174:213–226.
130. Cohen M, Persky R, Stegemann R, et al. Germline USP8 mutation associated with pediatric Cushing disease and other clinical features: a new syndrome. *J Clin Endocrinol Metab*. 2019;104:4676–4682.
131. Chen J, Jian X, Deng S, et al. Identification of recurrent USP48 and BRAF mutations in Cushing's disease. *Nat Commun*. 2018;9:3171.
132. Sbiera S, Perez-Rivas LG, Taranets L, et al. Driver mutations in USP8 wild-type Cushing's disease. *Neuro Oncol*. 2019;21:1273–1283.
133. Newey PJ, Nesbit MA, Rimmer AJ, et al. Whole-exome sequencing studies of nonfunctioning pituitary adenomas. *J Clin Endocrinol Metab*. 2013;98:E796–E800.
134. Bi WL, Horowitz P, Greenwald NF, et al. Landscape of genomic alterations in pituitary adenomas. *Clin Cancer Res*. 2017;23:1841–1851.
135. De Sousa SMC, Wang PPS, Santoreneos S, et al. The genomic landscape of sporadic prolactinomas. *Endocr Pathol*. 2019;30:318–328.
136. Salomon MP, Wang X, Marzese DM, et al. The epigenomic landscape of pituitary adenomas reveals specific alterations and differentiates among acromegaly, Cushing's disease and endocrine-inactive subtypes. *Clin Cancer Res*. 2018;24:4126–4136.
137. Li C, Xie W, Rosenblum JS, et al. Somatic SF3B1 hotspot mutation in prolactinomas. *Nat Commun*. 2020;11:2506.
138. Tatsi C, Pankratz N, Lane J, et al. Large genomic aberrations in corticotropinomas are associated with greater aggressiveness. *J Clin Endocrinol Metab*. 2019;104:1792–1801.
139. Valimaki N, Schalin-Jantti C, Karppinen A, et al. Genetic and epigenetic characterization of growth hormone-secreting pituitary tumors. *Mol Cancer Res*. 2019;17:2432–2443.
140. Sekine S, Shibata T, Kokubu A, et al. Craniopharyngiomas of adamantinomatous type harbor beta-catenin gene mutations. *Am J Pathol*. 2002;161:1997–2001.
141. Brastianos PK, Taylor-Weiner A, Manley PE, et al. Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nat Genet*. 2014;46:161–165.
142. Aylwin SJ, Bodi I, Beaney R. Pronounced response of papillary craniopharyngioma to treatment with vemurafenib, a BRAF inhibitor. *Pituitary*. 2016;19:544–546.
143. Roque A, Odia Y. BRAF-V600E mutant papillary craniopharyngioma dramatically responds to combination BRAF and MEK inhibitors. *CNS Oncol*. 2017;6:95–99.
144. Juratli TA, Jones PS, Wang N, et al. Targeted treatment of papillary craniopharyngiomas harboring BRAF V600E mutations. *Cancer*. 2019;125:2910–2914.
145. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990–3011.
146. Korbonits M, Storr H, Kumar AV. Familial pituitary adenomas - who should be tested for AIP mutations? *Clin Endocrinol*. 2012;77:351–356.
147. Schultz KAP, Williams GM, Kamihara J, et al. DICER1 and associated conditions: identification of at-risk individuals and recommended surveillance strategies. *Clin Cancer Res*. 2018;24:2251–2261.
148. Behrens LH, Barr DP. Hyperpituitarism beginning in infancy: the Alton giant. *Endocrinology*. 1932;16:120–128.

15

Aggressive Pituitary Tumors

Gérald Raverot, Alexandre Vasiljevic, and Mirela-Diana Ilie

OUTLINE

Definitions, 226
Pituitary Carcinomas, 226
Aggressive Pituitary Tumors, 226
Epidemiology, 226
Natural History of Pituitary Carcinomas, 227
Predictive Factors of Aggressiveness, 228
Pathological Markers of Aggressiveness, 228
Molecular Markers of Aggressiveness, 229
Treatment, 231
Temozolomide, 231
Therapeutic Perspectives, 232
Conclusion, 233

DEFINITIONS

Pituitary adenomas are tumors that arise from the hormone-expressing cells of the anterior pituitary and are generally considered benign. The majority of these tumors are not clinically relevant, are slow-growing, or are well controlled by surgery and/or medical treatment.

Only a minority of these tumors will give rise to distant metastases, although even in the absence of distant metastasis there are pituitary tumors that escape treatment using any therapeutic strategy. Over the past 10 years, more attention has been given to these not-so-uncommon pituitary tumors, leading to the definition of "aggressive pituitary tumors."

Pituitary Carcinomas

The definition of pituitary carcinoma is simple and well accepted because only a pituitary tumor with craniospinal or systemic metastasis is defined as a pituitary carcinoma. Despite this simple definition, it should be underlined that screening for metastatic disease is mandatory to either confirm or exclude the diagnosis of pituitary carcinoma.[1] Because this screening cannot be performed for all pituitary tumors, the decision to screen should be directed by clinical presentation.[2] The absence of systematic screening for metastasis may lead to an underestimation of the true incidence of pituitary carcinomas. Moreover, a recent case series of pituitary corticotroph carcinomas, which demonstrated the similarity between liver metastases and well-differentiated neuroendocrine tumors of visceral origin,[3] underlined another diagnostic challenge presented by pituitary carcinomas.

> **KEY POINTS**
>
> - Because the clinical definition of pituitary tumor aggressiveness is subject to interpretation, we define "clinically relevant tumor growth" based on signs and symptoms, whereas an "unusually rapid tumor growth rate" is defined as an increase in tumor size of 20% in 12 months or less.

Aggressive Pituitary Tumors

The term "aggressive pituitary tumor" has been used in the past to describe numerous clinical conditions, including giant tumors, invasive tumors, or tumors associated with high levels of hormone secretion. The absence of a clear definition has limited the identification of markers of aggressiveness and the evaluation of therapeutic strategies. In 2018, the European Society of Endocrinology (ESE) guidelines recommended that the diagnosis of an aggressive pituitary tumor be considered in patients with a radiologically-invasive tumor and unusually rapid tumor growth rate, or with clinically relevant tumor growth despite optimal standard therapies (surgery, radiotherapy, and conventional medical treatments).[2] This definition combined clinical and radiological criteria at presentation and tumor behavior during follow-up. It is important to highlight that an invasive pituitary tumor by itself is *not* synonymous with an aggressive tumor. Despite attempts to standardize the definition of an aggressive pituitary tumor, there are still no agreed-upon criteria for "clinically relevant tumor growth" or "unusually rapid tumor growth rate." Therefore, efforts should be made to better evaluate tumor growth and to expand the use of the response evaluation criteria in solid tumors (RECIST) that are used in other tumor types. Recently, Imber et al. demonstrated that the length of the tumor on its longest axis was, in general, adequately correlated to volumetric assessment, suggesting that the use of RECIST classification to evaluate pituitary tumor growth and response to treatment may indeed be appropriate.[4] In our practice, we define an increase in tumor volume of 20% in under 6 months (or 12 months when only annual magnetic resonance imaging [MRI] is available) as an unusually rapid tumor growth rate.

Better defining the criteria for aggressive pituitary tumors will aid in comparing the results of different treatment options but will still not permit the prediction of aggressiveness at diagnosis or enable management to be adapted to the prognosis. To this end, potential pathological and molecular markers associated with pituitary aggressiveness will be discussed in a later section.

EPIDEMIOLOGY

An excellent recent review has addressed the question of the epidemiology of aggressive pituitary tumors.[5] The high prevalence of pituitary tumors at autopsy or in radiological series (average prevalence rates of 14.4% [range 1%–35%] and 22.5% [range 1%–40%], respectively)[6] suggests that most pituitary tumors go unnoticed. According to cross-sectional community-based studies performed in different countries in Europe, the prevalence of clinically significant pituitary tumors ranged from 78 to 115 cases per 100,000 indivdiuals,[7,8] In the absence of strict criteria, the incidence of aggressive pituitary tumors can only be estimated based on the results of surgical series that evaluated the frequency of invasive tumors, postoperative recurrence, and response to standard treatment. Moreover, the incidence of pituitary carcinoma was obtained from large collections of pathology samples (0.13%–0.4%).[9,10] We can argue that, in the absence of systematic screening for metastasis, some carcinomas may have remained undiagnosed; however, this would not significantly change the number, which would still be low. Nevertheless, it is interesting to note that publications describing pituitary carcinomas dramatically increased after the first report of temozolomide (TMZ) treatment for pituitary carcinoma in 2006.[11] Despite these limitations, Dekker et al. estimated that less than 0.1% of all pituitary adenomas that are detected will become malignant, and most likely around 0.5% will display an aggressive course.[5] Although most pituitary tumors are benign and easily managed, we suggest that, as is the case for other rare tumors, aggressive pituitary tumors and pituitary carcinomas should be registered in an international cancer registry specific to pituitary tumors. The paucity of the data obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results database highlighted the need for a specific registry[12] to identify the true incidence of these tumors, exclude any publication bias, and allow a better characterization of mortality.

Natural History of Pituitary Carcinomas

Although no clinical features can predict the occurrence of pituitary carcinoma, the majority originate from macroadenomas, which are usually invasive into the cavernous sinus and resistant to medical treatment, requiring multiple surgical procedures and radiation therapy. A switch from a clinically nonfunctioning tumor to a functioning tumor seems to be a marker of evolution towards more aggressive behavior or towards metastasis.[13]

In a literature review of all the published cases of pituitary carcinoma from 1947 to 2017, Yoo et al. identified 72 individual cases in papers with full text available for review.[14] In 2017, we published the results of TMZ treatment in a large cohort of patients with tumors that included 14 pituitary carcinomas,[15] whereas the ESE survey, conducted in parallel, included 26 additional individual cases, leading to a total of 112 pituitary carcinoma cases: 65 in men and 47 in women. The median age at diagnosis was 45 years (range 9–75 years), and there was no difference in prevalence according to gender.

Details concerning metastases were available for 86 patients. The median time until identification of metastasis was 7 years, ranging from 0 to 31 years after diagnosis. *De novo* metastases were present in two cases, and the diagnosis was postmortem in three patients. The main locations of metastases were spinal and intracranial, but metastases could also be found in many locations, as seen in other types of cancer (liver, bone, lymph node, lung), and could present as either isolated single lesions or widespread dissemination to multiple organs (Fig. 15.1). Given the latency between primary diagnosis and metastasis, the ESE guidelines recommend screening for metastases in two instances: in the presence of site-specific symptoms (i.e., neurological complaints or neck/back pain) or in cases of discordance between radiological and biochemical findings (i.e., an increase in hormone levels without an increase in apparent tumor volume).[2]

KEY POINTS

- An extensive imaging work-up should be performed in the case of signs and symptoms that are atypical for pituitary tumors (cranial nerve palsies, neck/back pain, etc.) or in cases of discordant biochemical and radiological findings to identify metastases and confirm the diagnosis of pituitary carcinoma.

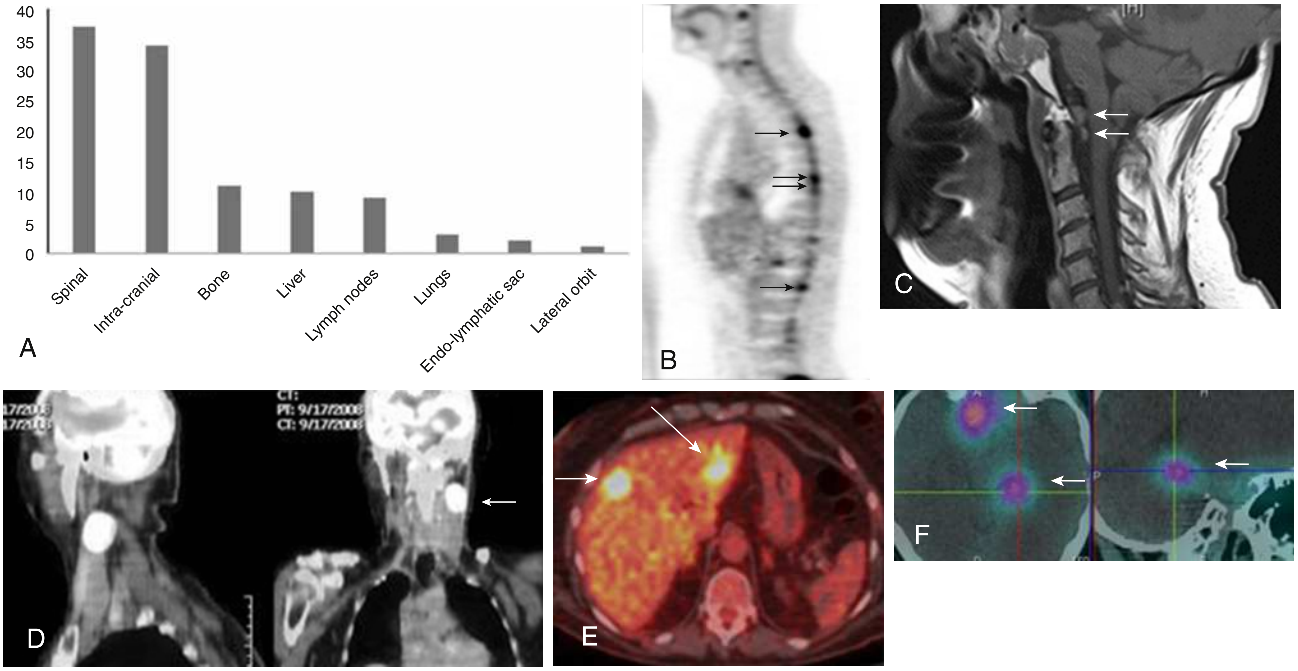


Fig. 15.1 Pituitary carcinoma metastases. **A,** Localization of pituitary carcinoma metastases (available data from 86 published cases). Bone scintigraphy demonstrating multiple vertebral metastases. **C,** T1-weighted sagittal magnetic resonance imaging showing two spinal cord metastases. **D,** Cervical lymph node metastasis on computed tomography (CT) scan. **E,** Liver metastases on (^{18}F)-fluorodeoxyglucose–positron emission tomography/CT. **F,** Brain metastases identified on ^{111}In-pentetreotide scintigraphy. All metastases are indicated by arrows.

Tumors of pituitary-specific transcription factor 1 (Pit1) lineage were found most frequently, representing 42% of all tumor types. Among these, prolactinomas ($n = 30$) were the main type, followed by growth hormone (GH; $n = 12$), and finally rare thyroid-stimulating hormone (TSH; $n = 3$) or TSH-prolactin ($n = 1$) tumors. The second most common lineage, with a similar number of cases, was the T-Pit lineage, with corticotroph tumors representing 40% of cases ($n = 45$). Gonadotroph tumors originating from the SF-1 lineage and hormone-negative tumors represented only 5% and 13% of the tumors, respectively (Fig. 15.2). Of the 112 cases, 52 deaths related to tumor progression were reported at the time of publication (survival status was unavailable for 10 patients).[13,14,15] The time until death was highly variable, from 10 months to 21 years after diagnosis.[14] However, it is noteworthy that, according to recent studies, TMZ treatment seems to improve the overall survival of responders compared with nonresponders.[13,15] For example, Lasolle et al. reported that median survival in patients treated with TMZ was 44 months (95% confidence interval [CI]: 42–infinity) among responders compared with 16 months (95% CI: 9–25) in nonresponders.[15]

PREDICTIVE FACTORS OF AGGRESSIVENESS

Pathological Markers of Aggressiveness

According to the 2017 World Health Organization (WHO) classification, the diagnosis of pituitary carcinoma cannot be made based solely on histopathological analysis; it also requires the identification of craniospinal or systemic metastases.[1] One of the major issues in pituitary pathology is the development of tools that will enable the early recognition of pituitary tumors that will evolve poorly and, most notably, allow the identification of pituitary carcinomas before they metastasize.

Ki-67 Proliferation Index, Mitotic Count, and P53. Taken individually, proliferation markers such as mitotic count and Ki-67–labeling index have failed to reliably identify those pituitary tumors that would later follow an aggressive course.[16,17] The results regarding anti-P53 immunodetection are similar in this respect. In one study, P53 immunoexpression was detected in 15% of invasive pituitary adenomas and in 100% of pituitary carcinomas but was absent in noninvasive pituitary adenomas.[18] Strong and diffuse P53 immunoexpression related to *TP53* mutations has also been described in a few cases of pituitary carcinoma.[19] Conversely, in other studies P53 expression was not found to be associated with recurrence[20] or invasiveness.[21] These inconsistent results regarding the Ki-67–labeling index and P53 immunoexpression may be related to different counting methods, or to technical variability (e.g., antibody clones used, tissue fixation). The "atypical adenoma" diagnostic category, which was first mentioned in 2004 in the WHO classification, does not appear in the more recent 2017 edition.[1] The lack of precise thresholds for assessment of mitotic counts or P53 immunodetection in the definition of atypical adenomas may account for difficulties in reliably identifying this entity in daily practice.[22]

The use of classifications including multiple clinical and pathological parameters has recently emerged as an alternative to a putative universal predictor of aggressiveness for pituitary tumors.[23] This approach is close to the definition of "high-risk adenomas" provided by the WHO classification.[1] High-risk adenomas are clinically aggressive neoplasms defined as rapidly growing, invasive tumors with a high Ki-67 proliferation index.

A clinicopathological classification taking into account pathological markers (mitotic count, Ki-67 proliferation index, P53 immunoexpression) and invasion has produced significant results in predicting the outcome of pituitary adenomas in several studies that included a total of 1470 patients[24-28] (Table 15.1). According to this classification, the risk of recurrence/progression is 3.72 to 12 times higher in grade 2b tumors (invasive and proliferative) (Fig. 15.3 A6, A9, A12) than in grade 1a tumors (noninvasive and nonproliferative).[24,25] Interestingly, pituitary carcinomas are mostly classified as grade 2b tumors.[24,28] In the 2016 ESE survey, it is noteworthy that aggressive pituitary tumors and pituitary carcinomas shared similar clinical and histopathological features.[29]

KEY POINTS

- We recommend the use of a pathological classification taking into account invasion and proliferative markers to identify pituitary tumors with a high risk of recurrence and resistance to medical treatment.

The experts who produced the WHO classification did not endorse any system of prognostic grading but strongly recommended the assessment of invasion and proliferation markers (mitotic count and Ki-67 proliferation index) for pituitary tumors.[1] This is in line with ESE guidelines that recommend at least a histopathological analysis, including hormone and Ki-67 immunodetection, as well as evaluation of P53 immunoexpression and mitotic count when the Ki-67 proliferation index is 3% or higher.[2]

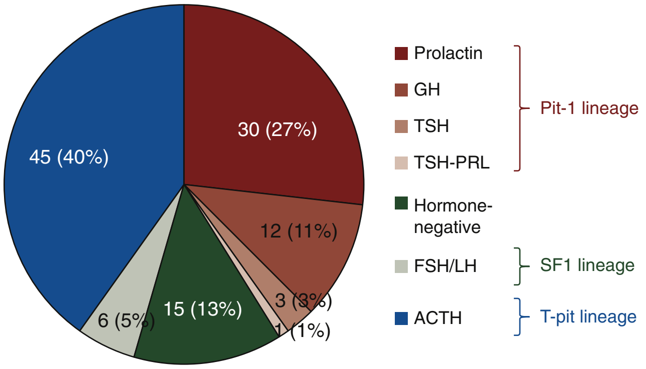


Fig. 15.2 Tumor type according to the lineage of the 112 pituitary carcinomas published in the literature. *ACTH*, adrenocorticotropic hormone; *FSH*, follicle-stimulating hormone; *GH*, growth hormone; *LH*, luteinizing hormone; *Pit-1*, pituitary-specific positive transcription factor 1; *PRL*, prolactin; *SF1*, steroidogenic factor 1; *T-pit*, T-box pituitary transcription factor 19; *TSH*, thyroid-stimulating hormone.

TABLE 15.1 French Five-Tiered Clinicopathological Prognostic Classification

Invasion is defined by histological and/or imaging and/or intraoperative evidence of invasion of the cavernous sinus and/or sphenoid sinus.

Proliferation is defined by at least two of the three following criteria:

- Mitotic count greater than two mitoses per 10 high-power fields (HPFs).
- Ki-67 proliferation index greater than or equal to 3%.
- P53 immunopositivity if more than 10 strongly positive nuclei per 10 HPFs.

Invasion	Proliferation		Metastasis +
	NO	YES	
NO	1a	1b	3
YES	2a	2b	

From Trouillas J, Roy P, Sturm N, et al. A new prognostic clinicopathological classification of pituitary adenomas: a multicentric case-control study of 410 patients with 8 years post-operative follow-up. *Acta Neuropathol.* 2013;126:123-135.

TABLE 15.2 Histological Subtypes of Aggressive Pituitary Adenomas According to the 2017 World Health Organization Classification System

Sparsely-granulated somatotroph adenoma
Silent corticotroph adenoma
Crooke cell adenoma
Lactotroph adenoma in men
Pit1-positive plurihormonal adenoma

From Lopes MBS. The 2017 World Health Organization classification of tumors of the pituitary gland: a summary. *Acta Neuropathol.* 2017;134:521-535.

Histotypes. The 2017 WHO classification of pituitary adenomas distinguishes some histopathological variants that may follow a more aggressive course. These include sparsely-granulated (SG) somatotroph adenomas, silent corticotroph adenomas (SCAs), Crooke cell adenomas, lactotroph adenomas in men, and Pit1-positive plurihormonal adenomas (Table 15.2).

Despite the recognition in the WHO classification of these subtypes as aggressive variants of pituitary adenomas, the true prognostic value of these histotypes is still debated in the literature.

Sparsely-Granulated Somatotroph Tumor. SG variants of somatotroph adenomas are defined as GH-producing pituitary neuroendocrine tumors that exhibit a characteristic cytoplasmic globular aggregate of cytokeratins termed a “fibrous body.” The 2017 WHO classification recognizes this variant as an aggressive subtype of pituitary adenoma.[1] However, studies comparing SG somatotroph adenomas with their densely-granulated (DG) counterparts show conflicting results in terms of invasion rate and postoperative outcome. In some studies, invasion was found to be more frequent in SG tumors,[30-32] whereas in others no significant difference was observed.[33,34] In one study, the postoperative remission rate was reported to be lower in SG tumors,[34] whereas no difference was observed in another study.[35] SG somatotroph tumors have also been found to express lower levels of somatostatin receptor (SSTR) 2a[33,36]. This may explain why SG tumors are usually less responsive to somatostatin analogs (SAs) than DG somatotroph adenomas.[33,34,37] This lower response of SG somatotroph tumors to SAs is certainly a contributing factor for their inclusion in the WHO group of aggressive pituitary adenomas.

Crooke Cell Tumor. Crooke cell adenoma is an exceedingly rare subtype of corticotroph tumor.[38] The very low number of cases available for prognostic analysis makes definite conclusions about their aggressiveness difficult. They are composed of neoplastic cells that harbor the typical hyaline ring-like cytoplasmic accumulation of cytokeratins seen in normal corticotrophs of patients with hypercortisolism. In comparison to macroadenomas in Cushing disease, invasion of the cavernous sinus is more frequent, but no significant difference regarding their rate of recurrence has been demonstrated.[38]

Silent Corticotroph Tumor. SCAs are clinically and biochemically nonfunctioning neuroendocrine tumors showing evidence of corticotroph differentiation.[39] Some authors have associated SCA with a poorer prognosis, and the 2017 WHO system classifies these tumors as a subtype of “aggressive” adenoma.[1,40] However, a metaanalysis of cases of SCAs reported up to 2017 did not find any evidence that supported a higher risk of recurrence compared with other nonfunctioning pituitary adenomas.[41] The poorer prognosis that has been suggested by some authors may be explained by the higher prevalence of SCA among the published pituitary carcinomas.[13]

Pituitary-specific positive transcription factor 1–Positive Plurihormonal Tumor. Previously referred to as “silent subtype 3 pituitary adenoma,” this neoplasm was initially defined as a rare ultrastructural entity. In the 2017 WHO classification, this tumor forms part of the Pit1 family.[42] Pit1 (which is encoded by *POU1F1*, or the POU class 1 homeobox 1 gene) is a transcription factor that is specific to the somatotroph, lactotroph, and thyrotroph lineages (Fig. 15.3A5). The clinical spectrum of Pit1-positive plurihormonal adenomas is broad as they may be ‘silent’ or associated with hyperthyroidism, acromegaly, and hyperprolactinemia (Fig. 15.3B). They lack diffuse and strong immunoexpression of Pit1-lineage hormones, showing negative, scattered, or focal immunopositivity for GH and/or prolactin and/or β-TSH (Fig. 15.3A4, A7, A10). The 2017 WHO system classifies this subtype in the group of “aggressive” adenomas.[1] These tumors are frequently large, invasive macroadenomas, with a high rate of residual disease and recurrence/progression (Fig. 15.3C).[42,43] The aggressiveness of this entity should now be confirmed in cohorts in which the tumors have been diagnosed according to the new definition of the 2017 WHO classification (anti-Pit1 immunohistochemistry).

Lactotroph Tumors in Men. In comparison with lactotroph tumors in women, lactotroph tumors in men are larger, more invasive, more often resistant to dopamine agonists, and show a lower level of estrogen receptor-alpha expression, which may be related to a lower degree of differentiation.[44]

Null Cell Adenoma. The 2017 WHO classification defines “null cell adenomas” (NCAs) as pituitary adenomas with neither hormonal expression nor transcription factor positivity.[1] In a study of 31 NCAs, defined according to the 2017 WHO criteria, they were more invasive at presentation and were associated with a shorter progression-free survival compared with gonadotroph adenomas.[45]

Molecular Markers of Aggressiveness

There is currently no independent molecular marker used in routine diagnostic practice that allows a pituitary tumor to be designated as aggressive.[46] The relevance of each molecular marker may also depend on the histological type of pituitary adenoma that is being considered.

Chromosomal alterations are more frequent in functional tumors than in silent tumors.[47] In one cohort of 134 pituitary adenomas, no correlation between chromosomal alterations and aggressiveness was found.[47] Conversely, in one study, loss of 11p was observed in aggressive lactotroph tumors, and an additional loss of 11q in malignant lactotroph tumors.[48]

Recurrent somatic mutations are rare in pituitary adenomas and are mostly restricted to *GNAS* (stimulatory G-protein alpha subunit) mutations in somatotroph tumors and *USP8* (ubiquitin-specific protease 8) mutations in corticotroph tumors. The prognostic value of these genetic alterations is unclear, and they likely do not work as independent factors.

GNAS mutations have not been reported to be more frequent in the aggressive SG subtype of somatotroph tumors than in the DG variant.[33,49,50] However, a metaanalysis suggested that *GNAS* mutational status may be a good predictive factor of response to SAs.[51]

In a study of 60 corticotroph adenomas in patients with clinically active Cushing disease, tumors carrying *USP8* mutations were significantly smaller and associated with a higher rate of surgical remission compared with wild-type *USP8* tumors.[52] However, the higher number of Crooke cell adenomas among the wild-type *USP8* tumors included in this study

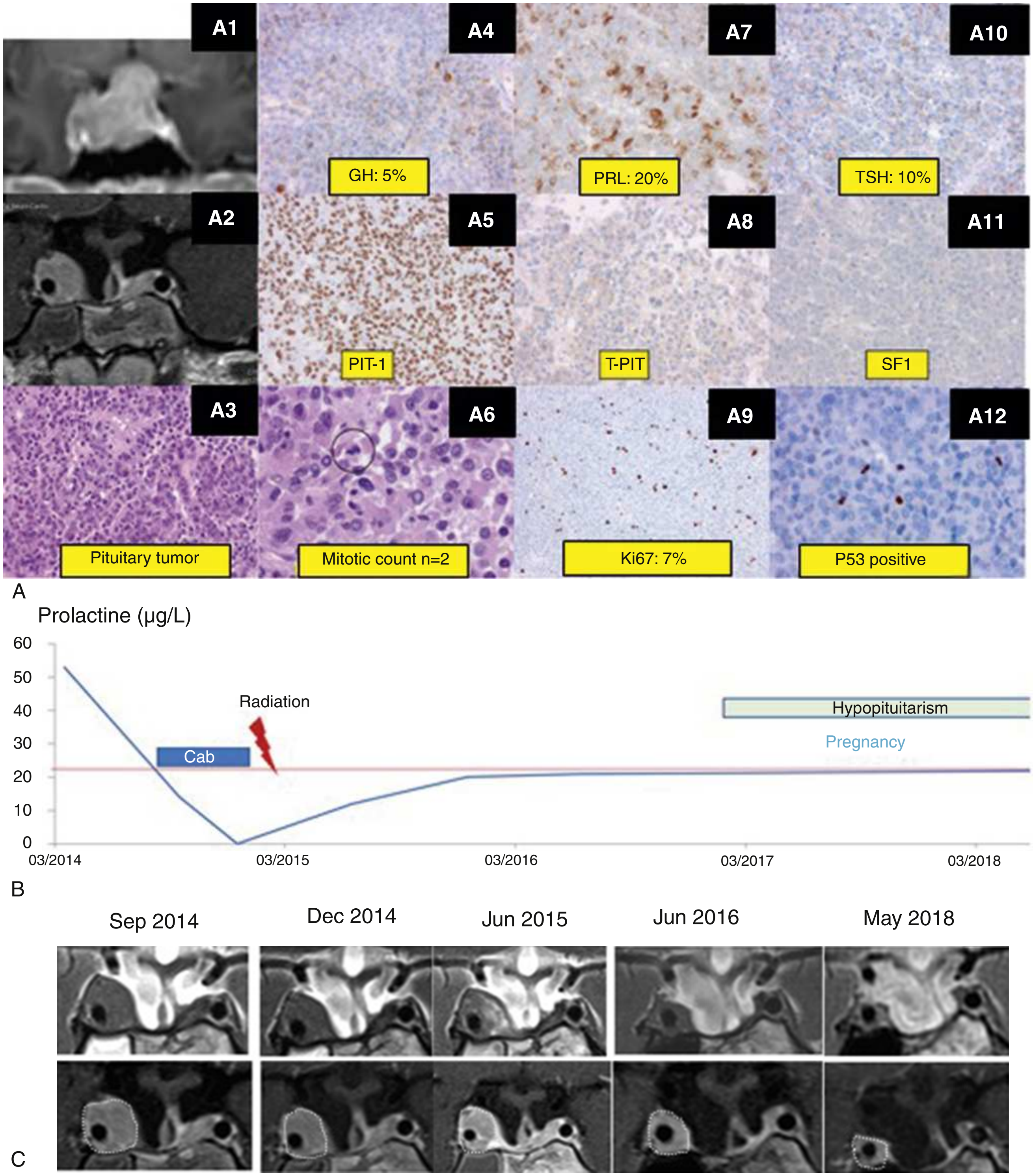

A1
A4
A7
A10
GH: 5%
PRL: 20%
TSH: 10%
A2
A5
A8
A11
PIT-1
T-PIT
SF1
A3
A6
A9
A12
Pituitary tumor
Mitotic count n=2
Ki67: 7%
P53 positive
A
Prolactine (µg/L)
60
50
40
30
20
10
0
03/2014
03/2015
03/2016
03/2017
03/2018
Radiation
Cab
Hypopituitarism
Pregnancy
B
Sep 2014
Dec 2014
Jun 2015
Jun 2016
May 2018
C

Fig. 15.3 Radiological and pathological illustration of a pituitary-specific positive transcription factor 1 (*Pit1*)-positive plurihormonal tumor. **A,** Preoperative (1) and postoperative (2) coronal T1 gadolinium magnetic resonance imaging (MRI) of the pituitary tumor. Hematoxylin and eosin (H&E) staining of the pituitary tumor (3). Immunohistochemistry of pituitary hormones: growth hormone (4), prolactin (7), thyroid-stimulating hormone (10); transcription factors: Pit-1 (5), t-Pit (8); SF1 (11), mitotic count on H&E (per 10 high power fields) (6). Immunohistochemistry of the Ki-67 proliferative marker (9) and p53 (12). **B,** Prolactin levels during follow-up of the 24-year-old patient before surgery and after cabergoline and radiation therapy. **C,** Radiological postoperative follow-up of the residual tumor on coronal T2 and T1 gadolinium MRI. The tumor is outlined by the dashed line. Note the tumor shrinkage after radiation therapy. *Cab,* Cabergoline; *GH*, growth hormone; *PRL,* prolactin; *SF1,* steroidogenic factor 1; *TSH*, thyroid-stimulating hormone.

may have contributed to these differences. Conversely, in another cohort of patients with Cushing disease, the prognosis for corticotroph tumors with *USP8* mutations showed earlier and more frequent recurrence.[53]

In their systematic review, Yang and colleagues highlighted the role of several factors in the molecular pathogenesis of pituitary carcinomas.[54] Factors that were upregulated included, notably, vascular endothelial growth factor (VEGF), a key factor in angiogenesis, matrix metalloproteinase 9, an enzyme involved in the degradation of the extracellular matrix and tumor cell migration, and cyclin D1, a protein that, in association with Cdk4 and Cdk6, promotes the G1-S cell cycle phase transition. Downregulated factors included cell cycle inhibitors such as P16ink4a and P27kip1.

TREATMENT

Temozolomide

TMZ is a second-generation alkylating agent that induces DNA damage through attachment of a methyl group to guanine bases. The O6-methylguanine (O6-MeG) lesions are responsible for the majority of TMZ cytotoxicity. The DNA mismatch repair (MMR) pathway that is activated by the mispairing cannot repair the O6-MeG lesion, which ultimately activates apoptosis.[55] An endogenous DNA repair enzyme, O6-MeG-DNA methyltransferase (MGMT), potentially counteracts the cytotoxic effects of TMZ by removing added methyl groups.

Standard therapy with TMZ begins with a dose of 150 mg/m^2 daily for 5 days every 28 days for one cycle; for subsequent cycles, the dose is increased to 200 mg/m^2, depending on patient tolerance, as recommended by the ESE guidelines. In some circumstances, when the maximal dose of radiation therapy has not been reached, a Stupp protocol combining radiotherapy and TMZ at a daily dose of 75 mg/m^2 for 6 weeks, followed by 6 to 12 months (or longer) of "standard therapy," may be used.[2,56]

Since the first reports in 2006,[11,57] numerous case reports and case series have been published, but the largest to date reported the results of a survey conducted by the ESE that included 40 pituitary carcinomas and 125 aggressive pituitary tumors.[13] Although initial publications only reported successful cases, thus introducing a positive reporting bias[58] the larger series[15,59-61] and the ESE survey[13] confirmed that approximately 40% of patients showed a partial or complete radiological response to TMZ, whereas one-third were stable, and the other third showed progression despite treatment. In addition to the radiological response, the biochemical response was also evaluated: 19% presented a complete biochemical response, 34% a partial response, 27% stable disease, and 21% progressive disease.[62] All of the studies were consistent, demonstrating that early response to TMZ (after 3 cycles) was the best predictor of long-term control (Fig. 15.4), whereas no delayed response has been reported to our knowledge. The data were sparse at the time of the publication of the ESE guidelines; however, the results from the ESE survey suggested that those patients who received concomitant radiotherapy (the Stupp protocol) showed an increased response rate (complete or partial radiological response in 71% of cases) compared with those on the standard TMZ treatment schedule. This difference could be attributed to a selection bias, but this possibility should be discussed in the future as an option in

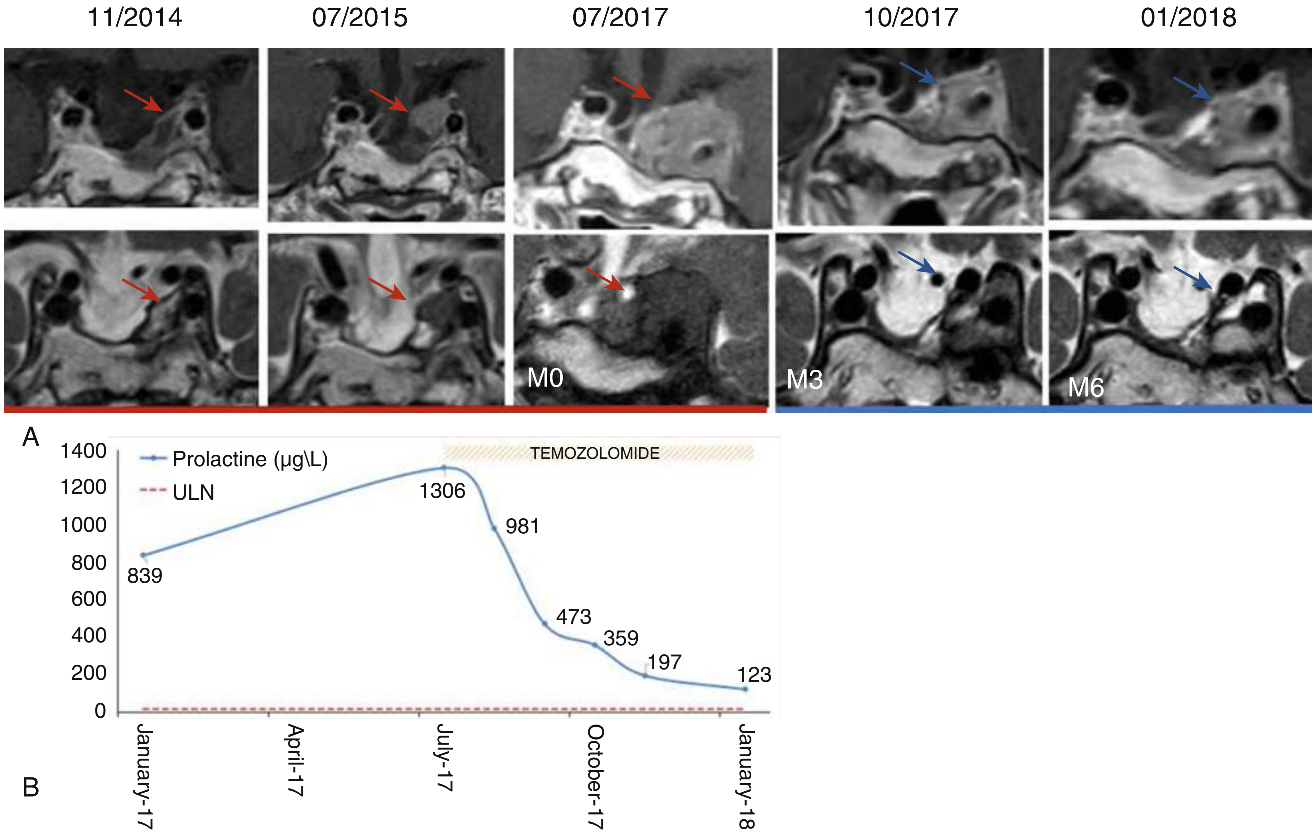


Fig. 15.4 Illustration of aggressive prolactinoma: radiological and hormonal follow-up before and after temozolomide treatment. **A,** Coronal T1 and T2 magnetic resonance imaging follow-up demonstrating tumor progression (*red arrow* indicates growing residual tumor) before temozolomide treatment (*red line*), and tumor response (*blue arrow* indicates shrinkage of the residual tumor) after 3 (M3) and 6 (M6) months of treatment (*blue line*). **B,** Graphical illustration of prolactin secretion before temozolomide and response after three and six cycles of temozolomide 200 mg/m^2 daily for 5 days every 28 days. *ULN,* Upper limit of normal.

selected patients with large residual tumors associated with clinical signs and pathology that point to a risk of aggressiveness.[63]

KEY POINTS

- Temozolomide is the first-line treatment for aggressive pituitary tumors and carcinomas. In some instances it could be considered for use as part of a Stupp protocol rather than only as salvage therapy.

Recent publications showed that a response to TMZ was associated with significant improvement in overall survival in responders compared with nonresponders (44 months vs. 16 months, respectively).[15] Progression-free survival after TMZ cessation was between 23 to 30 months,[15,61] and TMZ rechallenge in these patients was rarely efficient,[55] raising questions about the optimal duration of TMZ treatment. The ESE guidelines propose that, in responders, TMZ should be continued for 6 months or more when a continued therapeutic effect is noted[2]; however, data regarding the outcomes of long-term treatment are too limited to recommend its maintenance over time or for a specific duration. Some reports suggest that TMZ could be effective and well tolerated even after 2 years of treatment; however, to limit its potential cumulative toxicity, in our practice, we try to decrease the TMZ dose by half after 24 cycles.

KEY POINTS

- In responding patients, a longer duration of temozolomide treatment may be considered; however, despite its efficacy in about 40% of cases, there is still a need for alternative therapies. Among these alternatives, immunotherapy or antivascular endothelial growth factor should be considered.

The most frequent adverse effects are cytopenia (mostly thrombocytopenia), fatigue, nausea/vomiting, and abnormal liver function tests. A hematological profile should be obtained 7 days before starting a new cycle during standard treatment. In the presence of thrombocytopenia, a dose reduction or delay in treatment cycles can be proposed to allow continuation of the treatment.

Combination of TMZ with other therapeutic agents has also been described, including combination with capecitabine ($n = 4$) or bevacizumab ($n = 6$), but no conclusion could be drawn regarding the benefit of these combinations in light of the limited number of cases and their highly variable clinical presentation.[55,62]

Identification of markers that predict response to TMZ may assist in decision making in some circumstances. MGMT expression has been suggested as a potential predictive marker of tumor response to TMZ because MGMT-negative tumors typically exhibit a good response, and MGMT-positive tumors rarely achieve regression; however, the results have been inconsistent across a number of studies.[13,15,55,64] The inconsistency associated with these results may be caused by technical issues associated with MGMT immunostaining and the absence of an agreed cut-off to determine positive MGMT expression, but also to differences between the tumor tissue analyzed (obtained at the time of the surgery) and the same tumor at the time of TMZ treatment initiation.[15,65] MGMT methylation status has not been validated as a predictive marker of tumor response in pituitary tumors treated with TMZ.[2] Hirohata et al. studied the expression of DNA MMR proteins using immunohistochemistry and found that a lack of expression was associated with resistance to TMZ,[65] but this difference was not found in another study.[59]

Therapeutic Perspectives

Peptide Receptor Radionuclide Therapy. Peptide receptor radionuclide therapy (PRRT) is a site-directed targeted therapeutic strategy that specifically uses radiolabeled SSTR-binding molecules designed to deliver cytotoxic levels of radiation to tumor cells that express these receptors.[66] Most pituitary tumors express SSTRs, but so far PRRT has only been tested in a limited number of aggressive pituitary tumors or carcinomas. We recently reviewed the results of the 20 patients who have been treated using this approach.[62] Information regarding the initial presentation was available for 11 patients (three female and eight male patients): four nonfunctioning, two somatotroph, two lactotroph, one somatotroph-lactotroph, and two corticotroph (including one silent) tumors. None of these patients demonstrated a complete tumor response, three presented a partial response, three were stable, 11 progressed during treatment, and data were missing for three patients. These rare cases do not allow the identification of predictive markers of tumor response; however, five out of the six patients who exhibited stable disease or a partial response had not been previously treated with TMZ, suggesting that an earlier use of PRRT in potentially aggressive pituitary tumors, instead of as a salvage therapy, might work better. Potential side effects, mainly affecting bone marrow (transient anemia, thrombocytopenia, and/or leucopenia) and kidney, should be also taken into consideration.

Antivascular Endothelial Growth Factor Monoclonal Antibody. Although described in only 12 patients, bevacizumab is the alternative drug that has been most often used in the treatment of aggressive pituitary tumors and carcinomas, thus far being administered as monotherapy or in combination with TMZ. Results with this drug have shown a complete response in one corticotroph carcinoma,[67] whereas a partial response was seen in three patients, stable disease in five, and progressive disease in two.[62] These results are encouraging, but a reporting bias, with publication of only the successful cases, cannot be excluded.

Mammalian Target of Rapamycin Inhibitors. Rapidly accelerated fibrosarcoma/mitogen activated protein kinase/extracellular-signal-regulated kinase and phosphoinositide 3-kinase/protein kinase B/mammalian target of rapamycin are key signaling pathways that interact with each other to regulate cell growth and progression in pituitary tumors.[68] Everolimus, an mTOR inhibitor, has been shown to have antineoplastic activity *in vitro*, both in pituitary cell lines and in primary cultures.[69,70] Unfortunately, these encouraging preclinical results have not been confirmed by clinical experience. Of the six patients treated using this approach reported in the literature, only one, with an aggressive lactotroph tumor, presented stable disease (follow-up of 12 months).[62] The side effects with this treatment are generally mild, including fatigue, rash, stomatitis, diarrhea, myelosuppression, hyperglycemia, and hyperlipidemia.

Tyrosine Kinase Inhibitors. Following the demonstration that human prolactinomas express nuclear endothelial growth factor (EGFR), membranous EGFR, ErbB2, ErbB3, and ErbB4, and that their expression correlates with tumor invasion,[71] Cooper et al. conducted a pilot study examining the potential effect of lapatinib (a dual ErbB1/ErbB2 inhibitor) 1250 mg daily for 6 months in the treatment of aggressive prolactinomas. All six cases reported were dopamine-resistant lactotroph tumors not previously treated with TMZ. At 6 months, tumor volumes were stable ($n = 4$) or increased ($n = 2$), whereas hormone secretion demonstrated a partial response ($n = 3$) or increase ($n = 3$).[72,73] Four other cases of unknown tumor type that received lapatinib, erlotinib (an ErbB1 inhibitor), or sunitib (a multitargeted tyrosine kinase inhibitor) after TMZ showed progressive disease.[13,62] Tyrosine kinase inhibitors are associated with numerous side effects because of their actions on normal tissue. The most common adverse effects with lapatinib include diarrhea, moderate fatigue, nausea, and an acneiform rash.

Immune Checkpoint Inhibitors. Immune checkpoint inhibitors (ICIs) are monoclonal antibodies developed for the treatment of various cancers. ICIs target key points involved in immune system tolerance to tumors, including cytotoxic T lymphocyte antigen 4 (CTLA4), programmed cell death protein 1 (PD1), and its ligand, programmed cell death 1 ligand 1 (PDL1), thus enhancing the antitumor response.[74] The identification of tumor-infiltrating T lymphocytes and of PDL1 expression in pituitary tumors opened the door to this new therapeutic option for aggressive pituitary tumors and carcinomas.[75,76] Lin et al. published the first case of a corticotroph carcinoma showing a partial response on combined treatment with nivolumab (anti-PD1) and ipilimumab (anti-CTLA4).[77] Following this first publication, three other cases have since been published. Two patients were treated with a nivolumab and ipilimumab combination: one had a corticotroph carcinoma that responded to the treatment, and one had an aggressive prolactinoma that progressed during the treatment.[78] Finally, one patient with an aggressive corticotroph tumor presented rapid progression on pembrolizumab (anti-PD1).[79]

ICIs may be a good alternative for patients who show progression during or after TMZ treatment because TMZ induces somatic hypermutations, which may be associated with treatment response to immunotherapies.[77] However, the limited number of published cases does not allow any conclusions to be drawn, especially because ICIs, which were initially considered a promising option for numerous types of cancers, proved to be effective in only a minority of cases, and this efficacy may also be affected by gender.[80] Although the numbers are too limited to draw firm conclusions, the two responding patients were females, whereas the two males treated for pituitary tumors did not respond. It is noteworthy that microenvironments and genomic alterations are different among different pituitary tumor types; therefore, the tumor response to ICIs should be evaluated for each tumor type. For example, gonadotroph tumors, which present fewer tumor-infiltrating lymphocytes[81-83] and having fewer genomic alterations[47] may not be good candidates for treatment with ICIs.[84] The results from the two clinical trials on combined immunotherapy with ipilimumab and nivolumab that are currently underway (NCT04042753 and NCT02834013) may help physicians to better select patients to be treated with ICIs.

The potential downsides of therapy with ICIs are immune-related adverse events that can affect every organ, but particularly the skin (pruritus, rash, vitiligo), gastrointestinal tract (colitis, diarrhea), liver (hepatitis), and endocrine system (hypophysitis, primary adrenal insufficiency, dysthyroidism, and diabetes).[85]

CONCLUSION

The recent ESE guidelines on the management of aggressive pituitary tumors and carcinomas provided clinicians with the first signposts guiding the management of such patients. However, the definition of aggressive tumors is subject to individual interpretation, and a clearer definition of tumor growth velocity is needed. Although the majority of pituitary tumors are benign, we suggest that aggressive pituitary tumors and carcinomas should be included in cancer registries. Such a registry would allow clinical and pathological data to be amassed that can be used to improve our ability to identify tumors with a high risk of recurrence or of progression during medical treatment, and to evaluate responses to different therapeutic options. To date, TMZ is the best therapeutic option available, but whether it should be used earlier in the disease course and the optimal treatment duration remain open questions. Acknowledging that not all of these tumors are benign will help in the identification of predictive markers of aggressiveness and allow the development of both a personalized medicine approach and alternative therapies.

REFERENCES

1. Lopes MBS. The 2017 World Health Organization classification of tumors of the pituitary gland: a summary. *Acta Neuropathol*. 2017;134:521–535.
2. Raverot G, Burman P, McCormack A, et al. European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumours and carcinomas. *Eur J Endocrinol*. 2018;178:G1–G24.
3. Venable ER, Kerr SE, Lopes MBS, et al. Liver metastases from pituitary carcinomas mimicking visceral well-differentiated neuroendocrine tumors: a series of four cases. *Diagn Pathol*. 2020;15:81.
4. Imber BS, Lin AL, Zhang Z, et al. Comparison of radiographic approaches to assess treatment response in pituitary adenomas: is RECIST or RANO good enough? *J Endocr Soc*. 2019;3:1693–1706.
5. Dekkers OM, Karavitaki N, Pereira AM. The epidemiology of aggressive pituitary tumors (and its challenges). *Rev Endocr Metab Disord*. 2020;21:209–212.
6. Ezzat S, Asa SL, Couldwell WT, et al. The prevalence of pituitary adenomas: a systematic review. *Cancer*. 2004;101:613–619.
7. Daly AF, Rixhon M, Adam C, et al. High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium. *J Clin Endocrinol Metab*. 2006;91:4769–4775.
8. Fernandez A, Karavitaki N, Wass JA. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin Endocrinol (Oxf)*. 2010;72:377–382.
9. Saeger W, DK L, Buchfelder M, Fahlbusch R, et al. Pathohistological classification of pituitary tumors: 10 years of experience with the Germain pituitary tumors registry. *Eur J Endocrinol*. 2007;156:203–216.
10. Alshaikh OM, Asa SL, Mete O, Ezzat S. An institutional experience of tumor progression to pituitary carcinoma in a 15-year cohort of 1055 consecutive pituitary neuroendocrine tumors. *Endocr Pathol*. 2019;30:118–127.
11. Lim S, Shahinian H, Maya MM, et al. Temozolomide: a novel treatment for pituitary carcinoma. *Lancet Oncol*. 2006;7:518–520.
12. Hansen TM, Batra S, Lim M, et al. Invasive adenoma and pituitary carcinoma: a SEER database analysis. *Neurosurg Rev*. 2014;37:279–285; discussion 285-276.
13. McCormack A, Dekkers OM, Petersenn S, et al. Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*. 2018;178:265–276.
14. Yoo F, Kuan EC, Heaney AP, et al. Corticotrophic pituitary carcinoma with cervical metastases: case series and literature review. *Pituitary*. 2018;21:290–301.
15. Lasolle H, Cortet C, Castinetti F, et al. Temozolomide treatment can improve overall survival in aggressive pituitary tumors and pituitary carcinomas. *Eur J Endocrinol*. 2017;176:769–777.
16. Thapar K, Yamada Y, Scheithauer B, et al. Assessment of mitotic activity in pituitary adenomas and carcinomas. *Endocr Pathol*. 1996;7:215–221.
17. de Aguiar PH, Aires R, Laws ER, et al. Labeling index in pituitary adenomas evaluated by means of MIB-1: is there a prognostic role? A critical review. *Neurol Res*. 2010;32:1060–1071.
18. Thapar K, Scheithauer BW, Kovacs K, et al. p53 expression in pituitary adenomas and carcinomas: correlation with invasiveness and tumor growth fractions. *Neurosurgery*. 1996;38:765–770; discussion 770-771.
19. Tanizaki Y, Jin L, Scheithauer B, et al. P53 gene mutations in pituitary carcinomas. *Endocr Pathol*. 2007;18:217–222.
20. Hasanov R, Aydogan BI, Kiremitci S, et al. The prognostic roles of the Ki-67 proliferation index, P53 expression, mitotic index, and radiological tumor invasion in pituitary adenomas. *Endocr Pathol*. 2019;30:49–55.
21. Grimm F, Maurus R, Beschorner R, et al. Ki-67 labeling index and expression of p53 are non-predictive for invasiveness and tumor size in functional and nonfunctional pituitary adenomas. *Acta Neurochir (Wien)*. 2019;161:1149–1156.

22. Miermeister CP, Petersenn S, Buchfelder M, et al. Histological criteria for atypical pituitary adenomas—data from the German pituitary adenoma registry suggests modifications. *Acta Neuropathol Commun*. 2015;3:50.
23. Pappy 2nd AL, Savinkina A, Bicknese C, et al. Predictive modeling for pituitary adenomas: single center experience in 501 consecutive patients. *Pituitary*. 2019;22:520–531.
24. Trouillas J, Roy P, Sturm N, et al. A new prognostic clinicopathological classification of pituitary adenomas: a multicentric case-control study of 410 patients with 8 years post-operative follow-up. *Acta Neuropathol*. 2013;126:123–135.
25. Raverot G, Dantony E, Beauvy J, et al. Risk of recurrence in pituitary neuroendocrine tumors: a prospective study using a five-tiered classification. *J Clin Endocrinol Metab*. 2017;102:3368–3374.
26. Lelotte J, Mourin A, Fomekong E, et al. Both invasiveness and proliferation criteria predict recurrence of non-functioning pituitary macroadenomas after surgery: a retrospective analysis of a monocentric cohort of 120 patients. *Eur J Endocrinol*. 2018;178:237–246.
27. Asioli S, Righi A, Iommi M, et al. Validation of a clinicopathological score for the prediction of post-surgical evolution of pituitary adenoma: retrospective analysis on 566 patients from a tertiary care centre. *Eur J Endocrinol*. 2019;180:127–134.
28. Guaraldi F, Zoli M, Righi A, et al. A practical algorithm to predict post-surgical recurrence and progression of pituitary neuroendocrine tumours (PitNET)s. *Clin Endocrinol (Oxf)*. 2020;93:36–43.
29. Trouillas J, Burman P, McCormack A, et al. Aggressive pituitary tumours and carcinomas: two sides of the same coin? *Eur J Endocrinol*. 2018;178:C7-C9.
30. Mazal PR, Czech T, Sedivy R, et al. Prognostic relevance of intracytoplasmic cytokeratin pattern, hormone expression profile, and cell proliferation in pituitary adenomas of akromegalic patients. *Clin Neuropathol*. 2001;20:163–171.
31. Mori R, Inoshita N, Takahashi-Fujigasaki J, et al. Clinicopathological features of growth hormone-producing pituitary adenomas in 242 acromegaly patients: classification according to hormone production and cytokeratin distribution. *ISRN Endocrinol*. 2013:723432.
32. Obari A, Sano T, Ohyama K, et al. Clinicopathological features of growth hormone-producing pituitary adenomas: difference among various types defined by cytokeratin distribution pattern including a transitional form. *Endocr Pathol*. 2008;19:82–91.
33. Fougner SL, Casar-Borota O, Heck A, et al. Adenoma granulation pattern correlates with clinical variables and effect of somatostatin analogue treatment in a large series of patients with acromegaly. *Clin Endocrinol (Oxf)*. 2012;76:96–102.
34. Kiseljak-Vassiliades K, Carlson NE, Borges MT, et al. Growth hormone tumor histological subtypes predict response to surgical and medical therapy. *Endocrine*. 2015;49:231–241.
35. Sarkar S, Chacko AG, Chacko G. An analysis of granulation patterns, MIB-1 proliferation indices and p53 expression in 101 patients with acromegaly. *Acta Neurochir (Wien)*. 2014;156:2221–2230; discussion 2230.
36. Chinezu L, Vasiljevic A, Jouanneau E, et al. Expression of somatostatin receptors, SSTR2A and SSTR5, in 108 endocrine pituitary tumors using immunohistochemical detection with new specific monoclonal antibodies. *Hum Pathol*. 2014;45:71–77.
37. Bakhtiar Y, Hirano H, Arita K, et al. Relationship between cytokeratin staining patterns and clinico-pathological features in somatotropinomae. *Eur J Endocrinol*. 2010;163:531–539.
38. Di Ieva A, Davidson JM, Syro LV, et al. Crooke's cell tumors of the pituitary. *Neurosurgery*. 2015;76:616–622.
39. Ben-Shlomo A, Cooper O. Silent corticotroph adenomas. *Pituitary*. 2018;21:183–193.
40. Jahangiri A, Wagner JR, Pekmezci M, et al. A comprehensive long-term retrospective analysis of silent corticotrophic adenomas vs hormone-negative adenomas. *Neurosurgery*. 2013;73:8–17; discussion 17-18.
41. Fountas A, Lavrentaki A, Subramanian A, et al. Recurrence in silent corticotroph adenomas after primary treatment: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2018. https://doi.org/10.1210/jc.2018-01956.
42. Mete O, Gomez-Hernandez K, Kucharczyk W, et al. Silent subtype 3 pituitary adenomas are not always silent and represent poorly differentiated monomorphous plurihormonal Pit-1 lineage adenomas. *Mod Pathol*. 2016;29:131–142.
43. Erickson D, Scheithauer B, Atkinson J, et al. Silent subtype 3 pituitary adenoma: a clinicopathologic analysis of the Mayo Clinic experience. *Clin Endocrinol (Oxf)*. 2009;71:92–99.
44. Trouillas J, Delgrange E, Wierinckx A, et al. Clinical, pathological, and molecular factors of aggressiveness in lactotroph tumours. *Neuroendocrinology*. 2019;109:70–76.
45. Almeida JP, Stephens CC, Eschbacher JM, et al. Clinical, pathologic, and imaging characteristics of pituitary null cell adenomas as defined according to the 2017 World Health Organization criteria: a case series from two pituitary centers. *Pituitary*. 2019;22:514–519.
46. Mete O, Ezzat S, Asa SL. Biomarkers of aggressive pituitary adenomas. *J Mol Endocrinol*. 2012;49:R69-R78.
47. Neou M, Villa C, Armignacco R, et al. Pangenomic classification of pituitary neuroendocrine tumors. *Canc Cell*. 2020;37:123–134. e125.
48. Wierinckx A, Roche M, Raverot G, et al. Integrated genomic profiling identifies loss of chromosome 11p impacting transcriptomic activity in aggressive pituitary PRL tumors. *Brain Pathol*. 2011;21:533–543.
49. Mayr B, Buslei R, Theodoropoulou M, et al. Molecular and functional properties of densely and sparsely granulated GH-producing pituitary adenomas. *Eur J Endocrinol*. 2013;169:391–400.
50. Larkin S, Reddy R, Karavitaki N, et al. Granulation pattern, but not GSP or GHR mutation, is associated with clinical characteristics in somatostatin-naive patients with somatotroph adenomas. *Eur J Endocrinol*. 2013;168:491–499.
51. Efstathiadou ZA, Bargiota A, Chrisoulidou A, et al. Impact of gsp mutations in somatotroph pituitary adenomas on growth hormone response to somatostatin analogs: a meta-analysis. *Pituitary*. 2015;18:861–867.
52. Hayashi K, Inoshita N, Kawaguchi K, et al. The USP8 mutational status may predict drug susceptibility in corticotroph adenomas of Cushing's disease. *Eur J Endocrinol*. 2016;174:213–226.
53. Albani A, Perez-Rivas LG, Dimopoulou C, et al. The USP8 mutational status may predict long-term remission in patients with Cushing's disease. *Clin Endocrinol (Oxf)*. 2018. https://doi.org/10.1111/cen.13802.
54. Yang Z, Zhang T, Gao H. Genetic aspects of pituitary carcinoma: a systematic review. *Medicine (Baltim)*. 2016;95:e5268.
55. Burman P, Lamb L, McCormack A. Temozolomide therapy for aggressive pituitary tumours - current understanding and future perspectives. *Rev Endocr Metab Disord*. 2020;21:263–276.
56. Lasolle H, Ilie MD, Raverot G. Aggressive prolactinomas: how to manage? *Pituitary*. 2020;23:70–77.
57. Fadul CE, Kominsky AL, Meyer LP, et al. Long-term response of pituitary carcinoma to temozolomide. Report of two cases. *J Neurosurg*. 2006;105:621–626.
58. Raverot G, Castinetti F, Jouanneau E, et al. Pituitary carcinomas and aggressive pituitary tumours: merits and pitfalls of temozolomide treatment. *Clin Endocrinol (Oxf)*. 2012;76:769–775.
59. Bengtsson D, Schroder HD, Andersen M, et al. Long-term outcome and MGMT as a predictive marker in 24 patients with atypical pituitary adenomas and pituitary carcinomas given treatment with temozolomide. *J Clin Endocrinol Metab*. 2015;100:1689–1698.
60. Losa M, Bogazzi F, Cannavo S, et al. Temozolomide therapy in patients with aggressive pituitary adenomas or carcinomas. *J Neuro Oncol*. 2016;126:519–525.
61. Elbelt U, Schlaffer SM, Buchfelder M, et al. Efficacy of temozolomide therapy in patients with aggressive pituitary adenomas and carcinomas: a German survey. *J Clin Endocrinol Metab*. 2020;105.
62. Ilie MD, Lasolle H, Raverot G. Emerging and novel treatments for pituitary tumors. *J Clin Med*. 2019;8:1107.
63. Ilie MD, Jouanneau E, Raverot G. Aggressive pituitary adenomas and carcinomas. *Endocrinol Metab Clin North Am*. 2020;49:505–515.
64. McCormack AI, McDonald KL, Gill AJ, et al. Low O6-methylguanine-DNA methyltransferase (MGMT) expression and response to temozolomide in aggressive pituitary tumours. *Clin Endocrinol (Oxf)*. 2009;71:226–233.
65. Hirohata T, Asano K, Ogawa Y, et al. DNA mismatch repair protein (MSH6) correlated with the responses of atypical pituitary adenomas and pituitary carcinomas to temozolomide: the national cooperative study by the Japan Society for Hypothalamic and Pituitary Tumors. *J Clin Endocrinol Metab*. 2013;98:1130–1136.

66. Dash A, Chakraborty S, Pillai MR, et al. Peptide receptor radionuclide therapy: an overview. *Cancer Biother Radiopharm*. 2015;30:47–71.
67. Touma W, Hoostal S, Peterson RA, et al. Successful treatment of pituitary carcinoma with concurrent radiation, temozolomide, and bevacizumab after resection. *J Clin Neurosci*. 2017;41:75–77.
68. Dworakowska D, Wlodek E, Leontiou CA, et al. Activation of RAF/MEK/ERK and PI3K/AKT/mTOR pathways in pituitary adenomas and their effects on downstream effectors. *Endocr Relat Cancer*. 2009;16:1329–1338.
69. Cerovac V, Monteserin-Garcia J, Rubinfeld H, et al. The somatostatin analogue octreotide confers sensitivity to rapamycin treatment on pituitary tumor cells. *Cancer Res*. 2010;70:666–674.
70. Zatelli MC, Minoia M, Filieri C, et al. Effect of everolimus on cell viability in nonfunctioning pituitary adenomas. *J Clin Endocrinol Metab*. 2010;95:968–976.
71. Cooper O, Vlotides G, Fukuoka H, et al. Expression and function of ErbB receptors and ligands in the pituitary. *Endocr Relat Cancer*. 2011;18:R197-R211.
72. Cooper O, Mamelak A, Bannykh S, et al. Prolactinoma ErbB receptor expression and targeted therapy for aggressive tumors. *Endocrine*. 2014;46:318–327.
73. Cooper O, Bonert V, Rudnick J, et al. SUN-442 EGFR/ErbB2 targeted therapy for aggressive prolactinomas. *J Endocr Soc*. 2019;3:SUN-442.
74. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12:252–264.
75. Mei Y, Bi WL, Greenwald NF, et al. Increased expression of programmed death ligand 1 (PD-L1) in human pituitary tumors. *Oncotarget*. 2018;7:76565–76576.
76. Wang PF, Wang TJ, Yang YK, et al. The expression profile of PD-L1 and CD8(+) lymphocyte in pituitary adenomas indicating for immunotherapy. *J Neuro Oncol*. 2018;139:89–95.
77. Lin AL, Jonsson P, Tabar V, et al. Marked response of a hypermutated ACTH-secreting pituitary carcinoma to ipilimumab and nivolumab. *J Clin Endocrinol Metab*. 2018;103:3925–3930.
78. Duhamel C, Ilie MD, Salle H, et al. Immunotherapy in corticotroph and lactotroph aggressive tumors and carcinomas: two case reports and a review of the literature. *J Pers Med*. 2020;10:88.
79. Caccese M, Barbot M, Ceccato F, et al. Rapid disease progression in patient with mismatch-repair deficiency pituitary ACTH-secreting adenoma treated with checkpoint inhibitor pembrolizumab. *Anti Canc Drugs*. 2020;31:199–204.
80. Conforti F, Pala L, Bagnardi V, et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol*. 2018;19:737–746.
81. Ilie MD, Vasiljevic A, Raverot G, et al. The microenvironment of pituitary tumors-biological and therapeutic implications. *Cancers (Basel)*. 2019;11:1605.
82. Marques P, Grossman AB, Korbonits M. The tumour microenvironment of pituitary neuroendocrine tumours. *Front Neuroendocrinol*. 2020;58:100852.
83. Principe M, Chanal M, Ilie MD, et al. Immune landscape of pituitary neuroendocrine tumours reveals association between macrophages and gonadotroph-tumour invasion. *J Clin Endocrinol Metab*. 2020. https://doi.org/10.1210/clinem/dgaa520.
84. Ilie MD, Raverot G. Treatment options for gonadotroph tumors: current state and perspectives. *J Clin Endocrinol Metab*. 2020;105:dgaa497. https://doi.org/10.1210/clinem/dgaa497.
85. Castinetti F, Albarel F, Archambeaud F, et al. French Endocrine Society guidance on endocrine side effects of immunotherapy. *Endocr Relat Cancer*. 2019;26:G1–G18.

16

Radiotherapy in Pituitary Tumors

Nicola A. Rosenfelder, Francesca Solda, and Michael Brada

OUTLINE

Introduction, 236
Indications for Pituitary Radiotherapy, 236
Radiotherapy Techniques and Fractionation, 236
  Three-Dimensional Conformal Radiotherapy, 237
  Stereotactic Radiotherapy Techniques, 237
  Radiotherapy Fractionation, 238
  Linear Accelerator–Based Fractionated Stereotactic Radiotherapy, 239
  Linear Accelerator–Based Stereotactic Radiosurgery, 239
  Gamma Knife Stereotactic Radiosurgery, 239
  CyberKnife Stereotactic Radiosurgery, 239
  Proton Therapy, 239
Clinical Outcomes of Pituitary Radiotherapy, 239
  Evidence for Fractionated Radiotherapy and Changes in Practice Over the Years, 239
  Development of Radiosurgery, 240
Efficacy of Radiotherapy, 240
  Overall Survival, 240
  Radiological Tumor Control: Progression-Free Survival, 240
  Hormone Normalization, 240
  Visual Function, 243
Toxicity of Radiotherapy, 244
  Hypopituitarism, 244
  Visual Pathway Damage and other Cranial Nerve Injuries, 244
  Radionecrosis, 244
  Cerebrovascular Disease, 245
  Second Brain Tumors, 245
  Mortality, 245
  Proton Therapy, 246
Summary, 246
Reirradiation, 246
Conclusions and Future Directions, 246

INTRODUCTION

Although surgery represents the standard initial approach for the treatment of pituitary adenomas and provides long-term local and hormonal control in patients with completely excised lesions, radiotherapy is effective in patients with growing residual and recurrent adenomas and in tumors not amenable to surgery or medical treatment. It provides excellent rates of tumor control (>90% at 10–20 years) and control of hormonal hypersecretion in secreting adenomas, albeit with a delay.

Advances in radiotherapy techniques allow the delivery of an effective therapeutic dose to the tumor with a substantial reduction in the amount of dose to the surrounding brain, thereby potentially limiting the risk of treatment-related long term side-effects. We provide an overview of the role, effectiveness, and side-effect profile of current and evolving techniques of irradiation in secreting and nonsecreting pituitary adenomas.

INDICATIONS FOR PITUITARY RADIOTHERAPY

The optimal management of pituitary adenomas requires collaboration between surgeons, endocrinologists, radiologists, and clinical oncologists. The indication for radiotherapy has evolved in parallel with improvements in surgical techniques and the use of high-quality magnetic resonance imaging (MRI). The decision when to use radiotherapy requires striking a balance between its benefits and risks, especially in the context of a largely benign disease.

Radiotherapy had been considered for all patients with residual lesions after surgery, but is no longer *routinely* used as adjuvant treatment. In patients with residual nonfunctioning tumors following surgery, radiotherapy is generally withheld unless there are histological features suggestive of more aggressive behavior, or until there is evidence of tumor progression or particular concern about risk to vision. In progressive nonfunctioning and secreting adenomas, radiotherapy is highly effective in achieving tumor control in over 90% of patients at 10 years, and in the region of 90% at 20 years.[1-9]

In patients with functioning tumors, radiotherapy is considered where hormonal normalization is not achieved with surgery and/or medical therapy and in rapidly progressive or recurrent lesions. Hormone levels decline at a constant rate, with normalization reached over months or years after treatment, depending on pretreatment hormone values.[10,11]

KEY POINTS

- Radiotherapy is a recognised treatment option for non-secreting and secreting pituitary adenomas with residual progressive tumors after surgery, providing high rates of local tumor control and hormonal normalisation.

RADIOTHERAPY TECHNIQUES AND FRACTIONATION

Regardless of the technology adopted, the principles of pituitary radiotherapy are to:

- Accurately identify the treatment target within the sella and surrounding region,

- Accurately identify critical adjacent normal structures,
- Precisely deliver a predefined dose of radiation to the target, and
- Minimize the dose received by the surrounding neural structures to minimize the risk of normal tissue injury.

Modern radiotherapy integrates precise patient immobilization and high-quality imaging in the initial treatment preparation (planning) and during treatment. The absence of randomized controlled trials does not allow a direct comparison of different radiotherapy procedures and delivery techniques in terms of clinical outcome, and the advantages and disadvantages of the available technologies and their optimal indications remain a matter of debate.

KEY POINTS

- Radiotherapy techniques have evolved over time allowing better dose-conformality to the target, with less radiation reaching neighboring normal tissues. The standard of care remains highly-conformal fractionated radiotherapy with daily treatment over 4-5 weeks, while small pituitary lesions away from the optic apparatus may be amenable to single fraction radiosurgery.

Three-Dimensional Conformal Radiotherapy

Three-dimensional conformal radiotherapy (3D-CRT) given in multiple daily fractions (fractionated 3D-CRT) is the standard of care for pituitary radiotherapy. The treatment is given with photons (high-energy x-rays) using a linear accelerator (linac) and can be used to treat lesions of any size.

The patient is immobilized, either in a fixed or relocatable frame system, or more commonly in an individually molded relocatable plastic mask. The relocation accuracy of the immobilization mask is operator-dependent and is in the region of 3 to 5 mm,[12] and can be improved to 2 to 3 mm by using a tighter-fitting mask.[13]

The extent of the adenoma and the surrounding critical structures (optic nerves and chiasm) are identified on high-resolution MRI (thin-slice T1 MRI sequences with contrast), and these are accurately coregistered with computed tomography (CT) imaging acquired for radiotherapy planning.

The visible tumor is contoured on the MRI sequences in three planes to generate a volume encompassing the macroscopic extension of the pituitary lesion. A margin that accounts for any uncertainty in volume delineation and any estimated setup variation specific to the radiotherapy center is added to generate the treatment volume.

Radiotherapy beams are shaped to conform to the shape of the target with the use of a multileaf collimator (MLC), which is an array of computer-controlled parallel leaves that creates an adjustable aperture through which the beam of radiation is delivered. In modern MLC systems the leaf width is 5 mm or 3 mm (described as mini or micro MLCs, respectively).

MLCs also allow modulation of the intensity of the radiation beam across the target; this technique is described as intensity-modulated radiotherapy (IMRT). The technique of arcing IMRT (volumetric-modulated arc therapy [VMAT] or *RapidArc*) allows delivery of the radiation dose continuously while the linac rotates around the patient and offers a fast way of delivering complex, highly conformal treatments. VMAT/*RapidArc* is the currently preferred technique for high-precision treatment delivery.

The total dose prescribed in pituitary adenomas using 3D-CRT is 45 to 50 Gy given in daily fractions of 1.8 to 2 Gy 5 days per week. This dose is below the radiation tolerance of the surrounding critical neural structures, so additional dose constraints and specific avoidance of the adjacent organs are not generally employed. With increasing understanding of the sensitivity of functional structures, such as the hypothalamus and subregions of the temporal lobe, more refined avoidance of sensitive structures can be employed in the future. Most studies suggest that a dose of 45 Gy is sufficient for benign pituitary adenomas. See Fig. 16.1 for a radiotherapy plan using RapidArc.

Stereotactic Radiotherapy Techniques

Stereotactic radiotherapy (SRT) is a refinement of 3D-CRT in which the combination of more robust patient immobilization with accurate target identification provides a more precise method of treatment delivery that allows for a tighter margin around the treatment volume, and hence a reduction in the volume of normal brain tissue irradiated to high doses.[14] The principal aim of normal tissue sparing is to reduce the long-term toxicity of radiotherapy while maintaining its effectiveness.

Originally, the term "stereotactic" referred to the use of three-dimensional (3D) coordinates defined by external markers in relation to a neurosurgical type frame fixed to the head. While used in neurosurgical practice to define the site of a biopsy using the 3D x, y, and z coordinates, it was adopted in radiotherapy to determine the location of a lesion to be treated. For the purposes of high-precision radiotherapy, the traditional stereotactic localization has been superseded by the integration of modern imaging (CT and MRI) with less invasive forms of head immobilization, but the term "stereotactic" remains as a way of describing a precise method to deliver localized radiotherapy. As currently practiced, stereotactic radiotherapy should be appropriately termed "high-precision conformal radiotherapy."

All radiotherapy technologies use sophisticated on-treatment imaging systems (described as image guidance techniques) to ensure the treatment is delivered correctly. These include on-treatment CT imaging (cone beam CT), in-room x-ray imaging, and, more recently, MRI, and these are coupled with sophisticated automatic correction algorithms integrated with patient positioning devices with automatic adjustments to correct for potential errors.

Stereotactic, or high-precision, radiotherapy can be given using a number of radiotherapy delivery systems, which include a modern linac, multiheaded cobalt unit (Gamma Knife [GK]) and a small linac-mounted on a robotic arm coupled with an image-guided system (CyberKnife [CK]).

Linear Accelerator. Linacs are the "standard" radiotherapy machines used to deliver most radiotherapy treatments. Modern linacs fitted with image guidance systems are routinely used without specific modifications to perform SRT/high-precision conformal radiotherapy. The rotating, isocentrically-mounted linac allows for almost full spherical access to the target, resulting in homogeneous treatment dose delivery.

Gamma Knife. The GK, developed in the late 1960s, consists of multiple cobalt-60 sources arranged in a hemispherical distribution. The generated multiple beams of gamma radiation converge through a system of collimation at a single point (isocenter) to obtain a circumscribed spherical dose distribution within the treatment target. In the current version of the device, the cobalt sources are arranged in concentric rings, each one of which is able to move in a linear direction over an internal collimator, with several stopping positions. Each position corresponds to a different collimator size (from 4 to 16 mm) to produce beams of radiation of different spherical diameters or to selectively block specific groups of beams. The combination of several spherical doses of radiation allows the coverage of larger targets or nonspherical targets with a multiple isocentric arrangement.[15]

CyberKnife. Miniaturized linacs mounted on a robotic arm, commercially known as the CK, deliver many small photon beams under continual x-ray image guidance. The robotic arm has 6 degrees of freedom of movement, with a limited cone of access to create a dose distribution tightly conforming to the treatment target. The latest version of

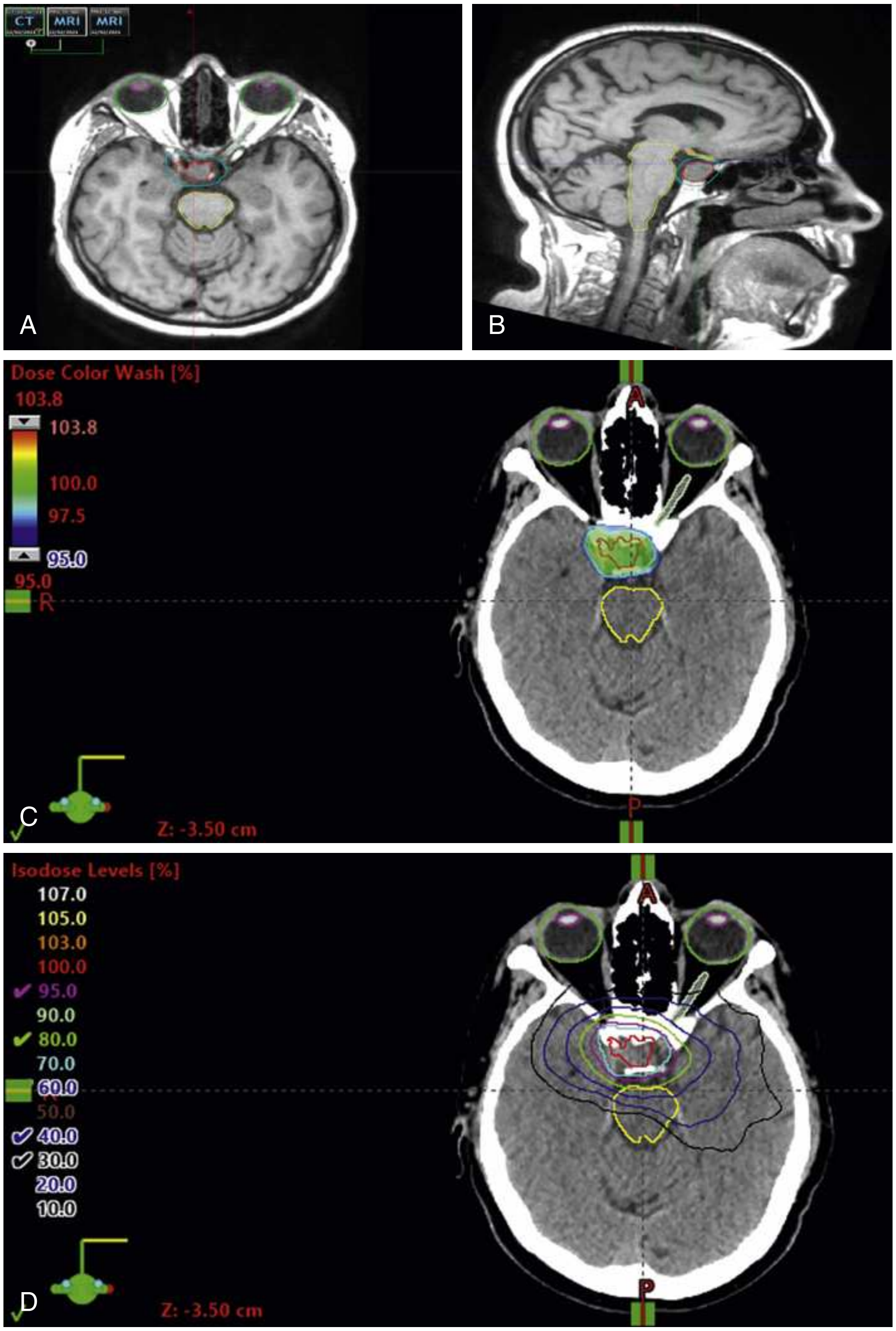


Fig. 16.1 Images from a pituitary radiotherapy volumetric arc plan. A and B, Axial and sagittal planning magnetic resonance imaging showing contour of tumor volume *(red)*, expanded to account for treatment uncertainties to create target treatment volume *(cyan)*. Organs at risk shown: brainstem *(yellow)*, optic chiasm *(orange)*, optic nerves *(light green)*, orbits *(apple green)*, and lenses *(purple)*. C, Axial planning computed tomography (CT) showing conformality of high-dose radiation region (displayed as *color wash*) to the target *(cyan contour)*, to ensure coverage of the target treatment volume by at least 95% of the prescribed dose and minimal dose outside target. D, Axial planning CT showing the dose distributions of higher and lower isodoses, demonstrating the dose fall-off outside the target. Dose falls off concentrically with distance from the high-dose target area.

the system is equipped with a multileaf MLC to shape the radiation beam and alter its intensity. The miniaturization results in small, low-output beams that require long treatment times.

Radiotherapy Fractionation

SRT can be delivered as fractionated treatment (fractionated stereotactic radiotherapy [FSRT]) delivering a treatment course over days or weeks, where the prescribed dose is given usually in small daily doses (fractions), or as a single high-dose treatment (described as stereotactic radiosurgery [SRS]).

The choice of radiotherapy fractionation is based in part on the size and location of the pituitary adenoma and the constraints demanded by the technology used. In the absence of randomized trials, the choice should be based on evidence of long-term outcomes of large cohorts of patients with data on long-term actuarial tumor and hormonal control, as well as the toxicity risks.

Most pituitary adenomas requiring radiotherapy lie in proximity to the optic apparatus and the nerves of the cavernous sinus. As the risk of radiation-induced optic neuropathy is dose- and fractionation-dependent, it is mandatory to spare the optic apparatus beyond single doses of 8 to 10 Gy to avoid unacceptable toxicity.[16,17] SRS is therefore reserved for small adenomas (≤3 cm) away from the optic apparatus (usually ≥3–5 mm distance between the target and the optic apparatus).

The rationale of dose fractionation is to lessen the risk of injury of the normal tissue surrounding the radiotherapy target, and therefore the risk of developing radiotherapy-related side effects. The total dose prescribed with FSRT for pituitary adenomas is 45 to 50 Gy delivered in fractions of 1.8 to 2 Gy. This dose range is below the tolerance of the central nervous system neural tissue, with a less than 1% to 2% risk of radiation-induced damage to normal cranial nerves.[16]

The use of a high single-dose of radiation, other than undoubted convenience, assumes that large doses per fraction may produce greater antitumor efficacy than predicted by classical radiobiological models for conventional fractionation.[18,19] The validity of these models at high doses per fraction is controversial, and they are not validated for single-fraction treatments or in the context of benign tumors. Currently, the available outcome data do *not* show superiority of SRS over fractionated irradiation in the treatment of pituitary adenomas.

Linear Accelerator–Based Fractionated Stereotactic Radiotherapy

FSRT is used to treat adenomas of all sizes, including lesions abutting or displacing the optic apparatus. Patients are immobilized in relocatable frames or in precisely fitting thermoplastic mask systems. On-treatment imaging carried out immediately before treatment (orthogonal x-rays or cone beam CT scans) enables instant, real-time correction for patient positioning deviations to ensure the precision of the daily treatment, which is generally of submillimeter accuracy.[20]

The treatment volume is irradiated with multiple fixed fields or, more frequently, with arcs (VMAT/*RapidArc*) at each daily session.

Linear Accelerator–Based Stereotactic Radiosurgery

Linac-based SRS can be used to treat adenomas less than or equal to 3 cm in maximum diameter. For larger lesions, treatment can be given in a few fractions, described as hypofractionated treatment ("hypofractionated SRT").

Gamma Knife Stereotactic Radiosurgery

Single-fraction GK treatment is used for small lesions not exceeding 3 cm in maximum diameter that are more than 3 to 5 mm from sensitive, critical structures.

Traditionally, the immobilization is with a fixed head frame attached to the outer layer of the skull under local anesthetic at four fixation points. The planning process (frame fixation and treatment planning scan) and the radiosurgical procedure are carried out on the same day.

Multiple isocenters are generally required to conform the treatment dose to the irregular shape of targets such as pituitary adenomas. This approach generates areas of inhomogeneous dose distribution within the target, where the contribution of different overlapping spheres of radiation can produce "hot spots" (areas of high doses of radiation).

Modern GK units can use frameless thermoplastic masks for immobilization to perform hypofractionated SRT, particularly for larger lesions. The system is equipped with a cone-beam CT for image guidance and a motion tracking device for monitoring the patient's position during treatment.

CyberKnife Stereotactic Radiosurgery

Patients are immobilized in closely fitting thermoplastic masks, and treatment for pituitary adenomas can be delivered as a single fraction or, more commonly, with a hypofractionated SRT regimen over two to five fractions. During treatment, the head position is monitored with real-time x-ray imaging, and the position of the beam in relation to the target is adjusted in real time.

Proton Therapy

Proton beam therapy (PBT) exploits the physical properties of heavy charged particles with dose deposition at depth over a narrow range (the so-called Bragg peak). Modern PBT is associated with some dose reduction beyond the target that can be theoretically associated with a better side-effect profile.[21] It is currently used as fractionated conformal radiotherapy for selected brain tumors, including pituitary adenomas, although the evidence of benefit remains unproven. Immobilization devices are similar to those used for photon-based techniques, and newer proton machines are equipped with cone-beam CT for image-guidance.

CLINICAL OUTCOMES OF PITUITARY RADIOTHERAPY

The management of pituitary tumors has changed considerably over the years. Although surgery remains the primary treatment for nonfunctioning pituitary adenomas (NFPA) and GH- and ACTH-secreting tumors, surveillance and medical treatment play a greater role pre- and postoperatively, and radiation is used less frequently.

The principal focus of the majority of the publications has been on relatively easily measurable endpoints of tumor control and hormonal control. When considering the efficacy of treatment for benign conditions, the outcome should be of clinical relevance to the patient. In pituitary radiotherapy, while tumor control remains of importance, endpoints such as preservation of vision, symptoms from excess/deficient hormone secretion, quality of life, and overall survival (which may not always be related to tumor control) are of principal clinical relevance.

Evidence for Fractionated Radiotherapy and Changes in Practice Over the Years

Reported recurrence rates after transphenoidal surgery alone for NFPA vary between 15% and 66%,[22,23] likely reflecting variations in patient population, surgical technique, and expertise, as well as in tumor biology.

Historically, postoperative radiotherapy was offered routinely to reduce the risk of tumor recurrence/progression. Nonrandomized studies suggested lower rates of tumor recurrence in patients treated with adjuvant radiotherapy; a summary of studies comparing the outcomes after adjuvant fractionated radiotherapy with no radiotherapy is shown in Table 16.1.[1,23-31]

However, the likelihood of recurrence is low following apparent complete excision, with recurrence risk largely confined to patients with subtotal excision.[27,32] The combination of excellent results following complete excision, the lack of randomized trials comparing adjuvant radiotherapy with surveillance, improvements in surgical techniques, and imaging and concern about potential late effects of radiotherapy have led to replacement of adjuvant radiotherapy with a policy of surveillance. Radiotherapy is currently reserved for patients with unresectable and progressive residual or recurrent tumors considered a threat to vision, or if the histology demonstrates atypical features suggesting a less benign tumor variant.

The recommended radiation doses for fractionated treatment, despite changes in delivery techniques, have remained largely unchanged, which is reasonable in the context of excellent long-term tumor control. The higher accuracy of dose delivery and better

TABLE 16.1 Studies Comparing the Outcomes After Adjuvant Fractionated Radiotherapy With no Radiotherapy

Authors	Year	Type of adenoma	Total number of patients (surgery +EBRT, surgery alone)	Median follow-up (years)	PFS 5-year (RT+)	PFS 5-year (RT-)	PFS 10-year (RT+)	PFS 10-year (RT-)
Jaffrain-Rea	1993	NFPA	57 (24, 33)	7.1	100	70	96	55
Gittoes	1998	NFPA	126 (63, 63)	9	-	-	93	68
Woollons	2000	NFPA	72 (50, 22)	5.3	72	34	-	-
Dekkers	2006	NFPA	109 (6, 103)	6	100	94	100	79
Van den Bergh	2007	NFPA	122 (76, 46)	8	95	49	95	22
Chang	2008	NFPA	352 (277, 75)	8.4	96	72	89	55
Olsson	2009	NFPA	235 (62, 173)	10	-	-	94	62
Brochier	2010	NFPA	142 (15, 127)	6.9	100	70	91	52
Vargas	2015	NFPA	112 (51, 61)	5	96	71	-	-
Deng	2019	NFPA	90 (26, 64)	7	96	76	96	64

Only studies with clinical follow-up >5 years were included.

EBRT, External beam radiotherapy; *PFS*, progression-free survival; *NFPA*, nonfunctioning pituitary adenoma; *RT*, radiotherapy.

(Modified from Chanson P, Dormoy A, Dekkers OM. Use of radiotherapy after pituitary surgery for non-functioning pituitary adenomas. *Eur J Endocrinol*. 2019;181:D1–D13.)

avoidance of adjacent normal tissue of stereotactic/high-precision techniques associated with lower doses to normal structures are assumed to translate into clinical benefit.

Development of Radiosurgery

The comparison of outcomes of single-fraction SRS with fractionated FSRT currently relies on comparisons of retrospective studies with inevitable selection bias, particularly based on tumor size and proximity to critical structures.

The recommended SRS dose for NFPAs is 12 Gy peripheral dose as the minimum effective dose, with no benefit to doses greater than 20 Gy,although some authors recommend a higher dose of 14 to 16 Gy.[33] There is no clear evidence whether dose escalation is beneficial for tumors with aggressive or atypical features. Higher doses are frequently used for functioning tumors, with doses of 15 to 3 5Gy (stereotactic-radiosurgery-radiotherapy-evidence-review.pdf) (england.nhs.uk), although this is not evidence based.

Fractionated radiosurgery schedules (hypofractionated SRT) with doses of 12 to 25 Gy delivered in three to five fractions are associated with good short-term outcomes, but here too longer-term follow-up data are required.[33]

KEY POINTS

- Outcome data for radiotherapy treatment of pituitary adenomas are largely derived from retrospective studies without prospective randomised trials comparing treatment techniques.

EFFICACY OF RADIOTHERAPY

Overall Survival

Overall survival is generally considered the most important clinical outcome in oncological studies, and its assessment includes excess mortality because of the disease and because of the treatment. While of relevance in patients with benign tumors such as pituitary adenomas, patients generally live for decades, and therefore the overall study time makes reliable survival outcome data collection less achievable, and results take decades to obtain. Overall survival, while a desirable outcome measure, is currently not an available endpoint for the assessment of NFPA outcomes.

Radiological Tumor Control: Progression-Free Survival

MRI tumor control, defined as progression-free survival (PFS), is a pragmatic objective endpoint allowing for comparison of treatment efficacy across studies. Although recurrence of a benign pituitary tumor is unlikely to directly affect survival, it is associated with increased morbidity and is therefore a useful outcome for assessment of treatment efficacy.

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy).

Studies assessing the outcome of fractionated radiotherapy demonstrate excellent tumor control, with total doses usually between 45 and 50.4 Gy delivered at 1.8 to 2 Gy per fraction, and are summarized in Table 16.2. The reported 10-year PFS rates for CRT and 3D-CRT are 86% to 100%, and appear to be maintained around 90% at 20 years.[1-5,7,8,34-48] Some patients treated in the era of conventional and conformal techniques will have received adjuvant radiotherapy regardless of the extent of surgical excision, and therefore these outcomes may represent a population with an overall lower risk of recurrence. After FSRT, even when mostly carried out for residual/recurrent tumors, the reported 5-year PFS is in the region of 95% to 99%, 10-year PFS is reported at 90%, and 20-year PFS is reported at 88%.[34,36,42,49-54]

The reported tumor control rates for NFPA treated with SRS are 95% to 97% at 5 years and 85% to 87% at 10 years,[55] when a marginal dose of 18 Gy is delivered.[56] These data are summarized in Table 16.3.[55-60] This is considered by some to be inferior compared with the results from fractionated treatment, as SRS is reserved for patients with smaller tumors, and the 10-year control rates appear similar to the 20-year rates reported after FSRT/3D-CRT.

Although systematic reviews and metaanalyses reported pooled outcome data from retrospective studies,[33,61] the heterogeneity of the studies, the variability in outcome reporting, and the short follow-up do not allow for reliable conclusions about comparative efficacy.

Hormone Normalization

For functioning pituitary tumors, normalization of hormone levels is an appropriate outcome, although this may also be considered a surrogate endpoint. The time taken to achieve hormone normalization varies with pretreatment levels. Time to normalization should be reported

TABLE 16.2 Outcome of Fractionated Radiotherapy in Pituitary Adenomas

FRACTIONATED STEREOTACTIC RADIOTHERAPY									
								TOXICITY (%)	
Author	**Year**	**Type of adenoma**	**Number of patients**	**RT dose**	**Median follow-up (years)**	**% PFS 5 years**	**% PFS 10 years**	**Visual**	**Hypopituitarism**
Wilson	2012	NFPA	67	50 Gy (median)	5.1	92.8	85.7	2	6
Kim	2013	54 NFPA, 22 SA	76	50.4Gy in $28^{\#}$	6.8	97.1 at 7 years		0	48
Reiken	2013	55 NFPA, 37 SA	92	52.2 Gy in $29^{\#}$	8.25	100 NFPA, 87 SA	100 NFPA, 81 SA	5.4	10.9
Minniti	2015	NFPA	68	45–50.4Gy in 25–$27^{\#}$	6.2	97	91	0	72 at 10 years
Puataweepong	2015	71 NFPA, 23 SA	94	45 Gy in $2^{\#}$	5.2	95 at 6 years		4	9.6
Three-dimensional conformal radiotherapy									
Author	Year	Type of adenoma	Number of patients	RT dose	Median follow-up (years)	% PFS 5 years	% PFS 10 years	Toxicity (%) Visual	Hypopituitarism
Grigsby	1989	NFPA, SA	121	30–50 Gy (range)	11.7		89.4	1.7	NA
McCollough	1991	NFPA, SA	105	42–55 Gy in 24–$32^{\#}$	7.8		95	NA	NA
Brada	1993	252 NFPA, 131 SA, 28 NK	411	45–50 Gy in 25–$30^{\#}$	10.8		94 at 10 years, 88 at 20 years	1.5	30 at 10 years
Tsang	1994	NFPA	128	45 Gy in $25^{\#}$	8.3		91 at 10 years	0	16
Rush	1997	46 NFPA, 24 SA	70	45 Gy in $25^{\#}$	8	NA	NA	NA	42
Breen	1998	NFPA	120	46.7 Gy (median)	9		87.5	0.8	NA
Gittoes	1998	NFPA	63	45–50 Gy in 25–$30^{\#}$	7.5		93 at 10 and 15 years	NA	NA
Barrande	2000	SA (GH)	128	52 Gy mean	11.5	39[a]	54 at 10 years and 61 at 15 years[a]	0	50
Epaminonda	2001	SA (GH)	67	53.6 Gy mean	11		65 at 15 years[a]	0	NA
Minniti	2005	SA (GH)	47	45–50 Gy in 25–$28^{\#}$	12	23[a]	42[a]	0	45 at 10 years
Langsenlehner	2007	61 NFPA, 26 SA	87	50.4 Gy	10.5	NA	NA	0	88 at 10 years
Snaed	2008	69 NFPA, 31 SA	100	43–50.4 Gy (mean 45 Gy)	6.7		98 NFPA and 73 for SA	1	35
Erridge	2009	189 NFPA, 196 SA	385	45 Gy	9.1		97	0.8	45
Rim	2011	25 NFPA, 35 SA	60	54 Gy	5.7		96 NFPA and 66 SA	0	76
Wilson	2012	NFPA	53	50 Gy (median)	5.1	86.9	76	2	6
Kim	2016	44 NFPA, 29 SA	73	45 Gy (median)	8		98	0	60
Scheick	2016	86 NFPA, 30 SA	116	45–50 Gy in 25–$28^{\#}$	9		96 NFPA and 62 SA	2	25

[a] Hormonal normalization. #, fraction

For fractionated stereotactic radiotherapy only studies with median follow-up >5 years and at least 50 patients were included.

For three-dimensional conformal radiotherapy only studies with follow-up >5 years and at least 50 patients were included.

PFS, progression-free survival; *NFPA*, nonfunctioning pituitary adenoma; *RT*, radiotherapy; *FSRT*, fractionated stereotactic radiotherapy; *3D-CRT*, three-dimensional conformal radiotherapy; *GH*, growth hormone; *SA*, secreting adenoma; *NK*, not known; *NA*, not available.

TABLE 16.3 Outcome of Pituitary Adenomas Treated With Stereotactic Radiosurgery

							TOXICITY (%)	
Author	**Year**	**Type of adenoma**	**Number of patients**	**RT dose (median)**	**Median follow-up (months)**	**TCR % at 5 years/10 years**	**Visual**	**hypopituitarism**
Iwai	2005	NFPA	34	14 Gy	60	93 at 5 years	0	6
Pollock	2008	NFPA	62	16 Gy	63	95 at 3 and 7 years	0	32 at 5 years
Park	2011	NFPA	125	13 Gy	62	94 at 5 years/76 at 10 years	0.8	24 at 2 years
Starke	2012	NFPA	140	18 Gy	50.4	97 at 5 years/87 at 10 years	12.8	30.3
Sheehan	2013	NFPA	512	16 Gy	36	95 at 5 years/85 at 10 years	6.6	21
Losa	2017	NFPA	272	15 Gy	79	95 at 5 years/79 at 10 years	NA	NA
						Hormone normalization %		
Vik-mo	2007	GH	53	26.5 Gy	67	58 at 5years/86 at 10 years	0	10 at 5 years
Pollock	2007	GH	46	20 Gy	63	11 at 2 years/60 at 5 years	0	10 at 2 years and 33 at 5 years
Losa	2008	GH	83	21.5	69	52 at 5 years/85 at 10 years	0	10 at 10 years
Ronchi	2009	GH	35	20	100	46 at 10 years	0	50
Iwai	2010	GH	26	20 Gy	84	16.9 at 5 years/47.4 at 10 years	0	8
Erdur	2011	GH	22	20 Gy	60	54.5 at 5 years	0	29 at 5 years
Franzin	2012	GH	112	22.5	75	56.9 at 5 years	0	7.8
Lee	2014	GH	136	25 Gy	61.5	64.5 at 4 years/82.6 at 8 years	3	33.1
Ding	2019	GH	371	24.2 Gy	79	59 at 10 years	4	26
Pollock	2008	ACTH	8	20 Gy	73	87 at 4 years	0	26 at 4 years
Sheehan	2013	ACTH	96	22 Gy	48	70 at 5 years	5	36

Only studies with median follow-up of >5 years or number of patients >100 were included.

TCR, tumour control rate; *RT*, radiotherapy; *GH*, growth hormone; *ACTH*, adrenocorticotropic hormone; *NFPA*, nonfunctioning pituitary adenoma.

together with the rate of hormone decline as the time to halving of pretreatment hormone levels to allow comparisons of treatment efficacy. The assessment of hormone level response is also complicated by the use of medical therapies, making it difficult to disentangle the effects of radiotherapy and medical treatment.

For patients with acromegaly, radiotherapy is usually reserved for those who failed surgical and medical therapy, although with improvements in radiotherapy delivery techniques, it may be recommended in selected patients as second-line treatment after surgery.[62,63] For ACTH-producing tumors, in the reported studies there is considerable heterogeneity in the measurement and definition of biochemical remission (as defined by serum cortisol and 24-hour urinary free cortisol) and a lack of consensus as to how best to define "normalization." Radiotherapy is recommended for patients with persistently raised cortisol levels after surgery, or those with known unresectable residual disease. For prolactinomas managed with medical therapy and, failing that, surgery, radiotherapy tends to be third-line treatment. Outcomes from radiotherapy for prolactinomas are therefore reported for heavily pretreated tumors resistant to other therapies.

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). For functioning tumors, normalization of hormonal levels is achieved in the majority of patients, although it can take months to years, and the rate and time taken to normalization vary with the hormone axis and the pretreatment levels. Publications reporting hormonal responses after 3D-CRT and FSRT are summarized in Table 16.4.[5,10,38-40,46,48,64-71]

In patients with GH-secreting tumors, the median time to halving of GH level is in the region of 2 years after CRT. The time to normalization of GH/IGF-1 levels related to the pretreatment GH level[40,64,72] varies widely between studies; following FSRT, GH normalization was reported in 84% of patients after a median follow-up of 26 months in one study, but in only 33% to 35% of patients at 39 months in two other studies,[50,65] and normalization can take up to 10 to 15 years.

In patients with ACTH-secreting adenomas, fractionated radiotherapy results in biochemical remission in 42% to 83% of patients,[72,73] with a median time to remission of 18 months. A 50% reduction in urinary free cortisol has been measured at 6 to 12 months after treatment, and in serum cortisol at approximately 1 year. Normalization of cortisol levels has been reported in 73%, 78%, and 84% of patients at 3, 5, and 10 years, respectively, after fractionated radiotherapy.[74] Similar outcomes are reported after FSRT, with tumor control rates of 99% at 5 years, and the results for ACTH response are likely to be similar to those achieved with CRT.[75]

Surgery is the primary treatment for gonadotroph pituitary tumors. Radiation treatment tends to be reserved for refractory tumors, although long-term outcomes are not clear.[75]

Despite being reserved for heavily-pretreated prolactinomas, radiotherapy has been reported to achieve 90% tumor control at 10 years and up to 20% to 60% hormone control.[72,76] In patients with prolactinomas treated with FSRT, at a median follow-up of 43 months (median 6–60 months), 33% of patients had normalization of serum PRL levels with reported times to normalization ranging from 5 to 40 months.

TSH-secreting adenomas are extremely rare and are usually controlled by surgery, with variable biochemical remission rates ranging from 35% to 84%.[75] Radiation therapy may be used in patients who do not achieve biochemical remission after surgery.

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). Endocrine outcomes following radiosurgery are mostly reported as hormone "normalization" affected by pretreatment hormone levels, which are likely to be lower for the

TABLE 16.4 **Reported Hormonal Response After Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy**

Author	Year	Type of adenoma	Number of patients	RT dose	Median follow-up (years)	% Hormone normalization 5 years	% Hormone normalization 10 years	Toxicities		RT Technique
								visual	hypopituitarism	
Tsang	1996	GH-, ACTH-, PRL-secreting	145	50 Gy	7.3		44	0	22%	CRT
Barrande	2000	GH-secreting	128	52 Gy (mean)	11		53	0	50% at 10 years	CRT
Biermasz	2000	GH-secreting	36	40–50 Gy	10		60	0	54 at 10 years	CRT
Epaminonda	2001	GH-secreting	67	53.6 Gy (mean)	10		65 at 15 years	0	NA	CRT
MilkerZabel	2004	GH-secreting	20	52.2 Gy	5	80		5	50%	FSRT
Minniti	2005	GH-secreting	47	45–50 Gy in 25–28#	12		52	0	45% at 10 years	CRT
Colin	2005	GH-secreting	33	50.4 Gy in 28#	6.6	20	50	0	37%	FSRT
Jenkins	2006	GH-secreting	645	45 Gy in 25#	7		60 at 10 years, 77 at 20 years	0	55 at 10 years	CRT
Minniti	2007	ACTH- secreting	40	45–50 Gy in 25–28#	9	78	84	0	62 at 10 years	CRT
Jallad	2007	GH-secreting	99	50 Gy	5.9	54		4	47%	CRT
Diallo	2015	GH-secreting	34	50 Gy in 27#	12.6		97	NR	39%	FSRT
Scheick	2016	GH, ACTH, PRL, TSH	30	54 Gy in 30#	9		62	NR	NA	3D-CRT
Knappe	2020	GH-secreting	233	NA	13.1		48	NR	46% at 9 years	CRT, 3D-CRT, FSRT

Only studies with median follow-up >5 years were included.

#, fraction

GH, Growth hormone; *ACTH*, adrenocorticotropic hormone; *PRL*, prolactin; *TSH*, thyroid-stimulating hormone; *CRT*, conformal radiotherapy; *FSRT*, fractionated stereotactic radiotherapy; *3D*, three-dimensional; *NA*, not available.

smaller tumors amenable to SRS.[77] In the absence of reporting the rate of hormone decline, comparison to other treatment techniques is difficult, if not impossible.

For GH-secreting tumors, a metaanalysis has shown that, at 5 years, estimated endocrine remission and endocrine control rates were 43.2% and 55.0%, respectively. At 10 years, these rose to 56.9% and 69.7%, respectively, and the estimated 10-year local control rate was 92.8%.[78] Better chances of remission have been reported in patients with lower pre-SRS IGF-1 levels.

For ACTH-producing tumors, approximately 50% of patients had biochemical remission at a corrected median follow-up of 42 months after SRS.[79] The time to hormonal response ranged from 3 months to 3 years, with no clear difference in the rate of hormonal decline compared with fractionated radiotherapy. Radiologic control was reported in up to 98% at a median follow-up of 4 years.[80] In other studies, biochemical remission was reported in 54% to 70% of patients at 5 to 10 years[80,81] and a median time to remission of 16.6 months, similar to fractionated treatment (18 months).[75] Although SRS for Cushing disease achieves high initial remission rates, approximately 20% of patients subsequently relapse.[82] [83] Comparisons between studies should be made with caution, as inevitably the treated populations will have wide variations in initial hormonal levels.

In prolactinomas, biochemical remission after SRS has been reported in 14% to 26% of patients who did not respond to medical or surgical approaches.[75,84,85] Remission is more likely to be achieved in patients with smaller adenomas and those that do not require dopamine agonist therapy at the time of radiotherapy.[84]

Visual Function

Visual function is one of the most relevant clinical outcomes for evaluation of treatment efficacy, particularly in patients with NFPA, and changes can be usefully assessed over a shorter time frame than that required for survival assessment. Effects of treatment can be measured objectively as improvement in visual fields and acuity following treatment. As radiotherapy is associated with a small risk of visual deterioration, the risk needs to be factored in when evaluating treatment outcome (see Toxicity section later).

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). Improvement in vision has been noted after surgery alone, and progressive improvement in visual acuity may continue for a year after treatment.[86] Improvements in visual impairment have also been reported in many studies of fractionated radiotherapy, with up to 55% to 78% experiencing improvement[4,87] after postoperative fractionated radiotherapy. In a separate study, following FSRT vision was stable in those with normal vision preradiotherapy and improved in 36% of those with visual impairment.[52]

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). There are limited data on improvements in visual function following SRS, most likely because of the contraindication to SRS of tumors close to the optic apparatus. However, improvement in visual function has been noted even for tumors close (<3 mm) to the optic apparatus.[88]

KEY POINTS

- Fractionated high-precision radiotherapy results in high rates of local control at 10 and 20 years (in the region of 90-95%), while the rate of hormone decline is specific for each hormone axis; normalisation occurs with a delay principally related to pre-treatment hormone levels.

TOXICITY OF RADIOTHERAPY

Acute side effects occur within 3 months of treatment, and long-term effects develop months or years after radiotherapy. Acute side effects include tiredness, temporary hair loss at the irradiation site, and minor skin reactions. They are transient and resolve within a few weeks, and none lead to long-term sequelae.

Radiotherapy to a pituitary adenoma leads to some inevitable irradiation of surrounding normal structures, which include cranial nerves, blood vessels, surrounding normal brain including the hypothalamus, and residual normal pituitary. Long-term side effects therefore include some risk to all these structures, although with conventional radiotherapy the dose delivered is below the radiation tolerance, and therefore the risk of damage is low, although it may be expressed late.

KEY POINTS

- The risk of radiation induced long-term toxicity following fractionated radiotherapy is low with hypopituitarism the most common side effect, while vascular toxicity and visual impairment are rare sequelae. Radiosurgery data is less mature with no clear difference between the toxicity outcomes for small tumors compared to conventionally-fractionated treatment.

Hypopituitarism

Radiation-induced hypopituitarism is a consequence of functional radiation damage to the hypothalamus and the pituitary, and the risk is affected by damage to the hypothalamo-pituitary axis from the direct effect of the tumor and from surgery. Because of the adverse impact of untreated hypopituitarism on survival,[89] patients treated with radiotherapy require long-term monitoring and hormone replacement.

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). After fractionated conformal radiotherapy, the risk of hypopituitarism is in the region of 30% at 5 years and 50% at 10 years.[4,5,38,40,43,44,90-92] The risk of radiation-induced hypopituitarism is dose-dependent, and it occurs more commonly after total doses of greater than 45 Gy or fraction size greater than 1.8 to 2 Gy for fractionated treatment. Higher rates of radiation-induced hypopituitarism have been reported in patients with functioning adenomas and may be related to the frequent use of higher radiation doses (50.4–54 Gy rather than 45 Gy).[40,75]

With increasing treatment accuracy and precision, and a reduction of planning margins, the risk of radiation-induced hypopituitarism after fractioned stereotactic treatment may be reduced, but longer-term outcome data are required because current reports are of short follow-up.

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). Early estimates of new hypopituitarism after SRS suggest a 5-year incidence of approximately 10% to 33% (see Table 16.3); assessment of the true long-term incidence requires longer follow-up combined with detailed reporting of hormonal status. As with fractionated radiotherapy, the risk of hypopituitarism after SRS is likely to be dose-dependent, with suggestion of improved outcomes with mean doses to the normal pituitary of less than 15 Gy,[93] or even less than 11 Gy.[94]

Many of the endocrine outcome data do not provide actuarial data, with no information on the length of follow-up available to be able to compare fractionated and single-fraction treatment, and claims of lower risk of hypopituitarism with SRS are currently not supported by the published data.

Visual Pathway Damage and other Cranial Nerve Injuries

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). Although the doses usually used for fractionated radiotherapy (45–50 Gy) are within the tolerance of the optic apparatus, late visual toxicity is reported in 0% to 2% of patients (see Table 16.2), with most studies reporting rates of less than 2%. Longstanding compressive effects on the optic apparatus by the primary tumor may contribute to the risk. The general perception is that the likelihood of developing visual impairment as a result of radiation damage is low in patients with normal vision before starting radiotherapy, and the risk is likely to be principally in patients with already compromised vision from compression of the optic nerve and/or chiasm before treatment. The risk is also considered higher in patients with a vasculopathy, such as that associated with diabetes, although hard data on this are not available.

Reported rates of other cranial neuropathies after fractionated radiotherapy are similarly less than 1% to 2%.

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). SRS is currently only offered to patients whose tumors are of adequate distance from the optic apparatus, and therefore the visual toxicity would be expected be lower than for fractionated treatment, where the tumor may abut the optic apparatus, which may already be compromised by prolonged compression.

The risk of optic neuropathy after SRS has been reported to affect less than 3% to 4% of patients (see Table 16.3). Although visual deterioration has been reported with doses less than 8 Gy,[59] the risk is thought to be low if the maximum point dose to the optic apparatus is below 8 to 10 Gy,[16,72,95] with significantly higher risks associated with doses greater than 12 Gy.[16,96,97] For fractionated SRS, suggested doses with limited risk of damage to the optic apparatus include less than 20 Gy in three fractions and less than 25 Gy in five fractions.[98] Other cranial neuropathies involving cranial nerve III and cranial nerves IV, V, and VI have also been also reported, with a risk of 0% to 3%[33] likely to be related to the extent of tumor infiltration and involvement of the cavernous sinus, and therefore potential preexisting damage. The rate of cranial neuropathy after SRS for Cushing disease was reported at 5%, probably a reflection of a higher dose of radiation.[80]

Radionecrosis

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). Radionecrosis is reported to occur in 0.2% to 3% of patients treated with fractionated conventional radiotherapy.[72,99,100] There is no additional concern of radionecrosis with fractionated SRT, as the doses used are within the tolerance of the brain, and the risk of this is likely to be as low as or lower than with conventional or 3D-CRT (<1%).[72]

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). The rates of radionecrosis after SRS may be higher than after fractionated treatment, with reported rates varying from 0% to 5% after SRS or hypofractionated radiotherapy.[72,101,102] Early use of SRS was associated with a high risk of radiation necrosis leading to significant concerns, but with increased understanding and greater expertise of SRS, particularly the move to restrict SRS to small lesions, the risks are reported to be low, although long-term outcomes from more recent SRS treatments need to be reevaluated.

Cerebrovascular Disease

The increased risk of stroke following pituitary radiotherapy is documented for NFPAs and functioning tumours.[89,103-106] However, the relative contribution by the radiotherapy to the excess risk is uncertain, as a number of factors are thought to contribute to the development of cerebrovascular disease. These include the pituitary tumor itself, which is associated with increased mortality mainly because of vascular disease, the cardiovascular effects of pituitary dysfunction (hormonal excess or deficiency), and the consequences of surgery.[107]

Fractionated treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). In a long-term follow-up study, the reported actuarial incidence of a stroke was 4% at 5 years, 11% at 10 years, and 21% at 20 years after radiotherapy, with a relative risk of 4.1 compared with the general population.[103,104] It has been suggested that the risk may be related to the dose delivered.[95]

It is unlikely that the increased risk is solely because of radiotherapy. Reviews of large cohorts suggest that the pretreatment duration of hypopituitary symptoms (indicative of hormone deficiency) may be a significant risk factor for stroke after pituitary adenoma treatment,[105] and both pituitary surgery and the tumor itself are likely to be contributory risk factors.[108]

The precise contribution of radiotherapy to the risk of cerebrovascular risk remains uncertain, and there are currently no clear modifiable avoidance strategies in the treatment of pituitary adenoma to reduce the risk. As patients with a history of a pituitary tumor are at higher risk of cerebrovascular disease, appropriate advice should be given for long-term monitoring and treatment of other cerebrovascular risk factors, such as hypertension, diabetes, and hypercholesterolaemia.

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). There is a paucity of data on the risk of cerebrovascular disease after SRS. Only two studies reported cerebrovascular outcomes after SRS, and no cases were reported in small cohorts of patients.[109,110] Longer follow-up of large cohorts is required to assess the risk of cerebrovascular events after SRS.

Second Brain Tumors

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). Reports of radiation-induced brain tumors following pituitary radiotherapy vary, ranging from no increased risk after radiotherapy[95,111-113] to a risk of 3% at 10 to 20 years.[43,113-118] Reported second brain tumors include meningiomas, sarcomas, neuroepithelial tumors, and gliomas.[111,112,119] How this incidence compares to a matched (unirradiated) control group of similar patients is difficult to ascertain.

In one of the largest series, the cumulative risk of second brain tumors was 2% at 10 years and 2.4% at 20 years. The relative risk of second brain tumors compared with the incidence in the normal population was 10.5 (95% confidence interval 4.3–16.7), and the relative risk was 24.3 for meningeal tumours.[115] In a further cohort, the 10-year and 20-year incidence of a second brain tumor was increased by 1.3%.[120]

Conversely, in a comparative study of over 462 patients, 236 of whom received postoperative radiotherapy, there was no significant difference in the incidence of posttreatment secondary tumors or mortality, even though the median follow-up was 14 years in those treated with radiotherapy compared with only 6 years in those treated with surgery alone.[113]

A comprehensive review of radiation-induced tumors after pituitary radiotherapy demonstrated that gliomas and sarcomas may arise from 5 years after radiotherapy and continue to rise at a constant rate until 20 to 25 years after treatment. Meningiomas have a longer latency and arise 7 to 10 years after radiotherapy,[117] and can continue to arise decades after treatment. The risk ratio in younger patients undergoing pituitary/sellar radiotherapy was reported as 3.34 for malignant brain tumors and 4.06 for meningiomas.[116]

Although there is compelling evidence for an increased incidence of tumors, it is likely that there are other contributing factors, such as the oncogenic effect of excess GH/IGF-1 in acromegaly and the immunocompromise associated with cortisol in patients with Cushing disease.[121,122]

The risk of developing a radiation-induced malignant brain tumor increased by 2.4-fold ($P = .005$) and meningioma by 1.6-fold with every 10 years of younger age at the time of radiotherapy ($P = .05$), with similar rates in patients treated with conventional radiotherapy or stereotactic radiotherapy.[116]

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). Secondary brain tumors have not been observed after SRS in published studies, although, as a late effect of treatment, this may be because of insufficient long-term follow-up data after SRS.[72,123] However, in a large, retrospective, multicenter cohort study of nearly 5000 patients who had undergone SRS for a variety of benign brain tumors, with a median follow-up of 8 years (and minimum of 5 years), there was no increased risk of secondary malignant central nervous system tumor compared with the general population.[124]

Although the risk of second brain tumors appears to be lower after SRS than following fractionated radiotherapy,[33] longer follow-up data are required. To date, the evidence is encouraging that SRS is not associated with a significantly increased risk.

Mortality

Fractionated Treatment (Three-Dimensional Conformal Radiotherapy and Fractionated Stereotactic Radiotherapy). Although treatment is aimed at prolonging life expectancy, the beneficial effect of treatment on survival is difficult to assess for patients with benign tumors, because of the need for long follow-up of large cohorts. It is therefore difficult, if not impossible, to assess the potential effects of radiotherapy on survival, whether improvement because of tumor control or detriment because of radiotherapy toxicity.

Increased mortality has been reported in patients with NFPAs and with acromegaly treated with conformal radiotherapy, with a relative risk of 1.6 to 4.1.[72,107,125,126] Excess mortality is attributed to cardiovascular disease, and a number of factors may contribute to this, such as hypopituitarism,[33,89] the effect of excess hormone production, and previous pituitary surgery, with uncertainty regarding the contribution from radiotherapy.[72,127-130]

Radiosurgery (Stereotactic Radiosurgery and Hypofractionated Stereotactic Radiosurgery). Based on the currently available information, there is no clear effect of SRS on life expectancy, although patient selection, the absence of randomized controlled trials, and limited follow-up of SRS cohorts mean that no definite information about a potential benefit or detriment is currently available. Similarly, it is not possible to assess the relative risk of SRS versus fractionated treatment.

KEY POINTS

- Retreatment of recurrent pituitary adenoma is feasible, but should be undertaken with caution and careful counseling in view of the increased risk of toxicity.

Proton Therapy

Early reports from very small series suggest that proton treatment results in reasonable pituitary tumor control rates.[131] However, the use of proton therapy for pituitary tumors has, to date, not demonstrated any superiority over fractionated or radiosurgical techniques. Outcomes from 30 patients treated with acromegaly included an 80% decrease in GH at 4.5 years, but at the expense of increased pituitary deficiencies and oculomotor nerve palsies.[132] Claims of efficacy of proton treatment must be validated with longer-term toxicity outcomes and in the context of well-designed studies. Currently, proton beam radiotherapy cannot be recommended as standard treatment for pituitary adenomas.

KEY POINTS

- The use of heavy-charged particles has not demonstrated any clear benefit over conventional photon (fractionated) therapy; long-term data are required to assess if proton therapy is associated with lower risk of sequelae.

SUMMARY

Radiotherapy remains an effective treatment in achieving tumor and hormonal control in patients with pituitary adenomas. Differences in the efficacy and toxicity of FSRT and SRS for treatment of pituitary adenomas are suggested by the comparison of reported outcomes in retrospective reviews, many of which are small. However, with differences in patient selection for SRS and fractionated treatment, comparisons between delivery methods are difficult, if not impossible, to make. Because fractionated treatment has been used for decades, information about the long-term effects of SRS is not yet as mature, and long-term data are largely missing.

In the absence of randomized data, technical developments in radiation delivery are assumed to be of benefit and are adopted into practice. However, when there are more substantial changes to treatment, such as changes in dose and fractionation, outcomes should continue to be monitored to ensure that the use of alternative treatment scheduling frequently dictated by technologies employed is as effective and safe as more traditional approaches. Early reports of novel technologies claiming superiority other than demonstrating feasibility should largely be ignored, as only long-term data on efficacy and toxicity should lead to changes in practice. The decision as to whether to use a fractionated versus a radiosurgical approach is generally based on the tumor characteristics and the relationship between the target and nearby normal tissue, but, when either option can be delivered within the dose constraints, available data currently suggest that either approach may be offered safely.

REIRRADIATION

Reirradiation for progressive or recurrent disease is considered with caution because of the higher risks of radiation-induced injury resulting from the late effects of treatment on the optic apparatus, cranial nerves, and vasculature, as well as on the normal brain tissue. However, reirradiation using fractionated conventional stereotactic methods is feasible and effective in achieving tumor control and recovery of tumor-related symptoms such as visual improvements, with acceptable toxicity when delivered after a minimum 3- to 4-year gap since the first irradiation and if the primary radiotherapy treatment was delivered to doses below the radiation tolerance of the central nervous system.[133] Risks of reirradiation include hypopituitarism, temporal lobe injury/radionecrosis, and an increased risk of radiation-related tumors, including pituitary carcinoma.[134-136] Although reirradiation can improve vision, the risk of optic neuropathy is higher after reirradiation, and careful counseling regarding the risks must be undertaken.

Radiosurgery has been used after previous radiotherapy, with some success in achieving hormonal control in acromegaly, although the duration of benefit is not known, and retreatment carries significant risk.[137]

CONCLUSIONS AND FUTURE DIRECTIONS

In the absence of randomized data, the comparison of radiotherapy treatment approaches relies on retrospective studies, with variation in populations and indications for treatment, nonstandardized follow-up protocols, varying duration of follow-up, and incomplete outcome reporting. Because of the heterogeneity of published studies, conclusions are often based on systematic reviews and quasi metaanalyses, which, despite combining data into larger cohorts, cannot improve on the limited quality of the initial datasets and therefore suffer from the same problems. The quality of the publications is also highly variable, and the impact of missing data, which are likely to have considerable impact on outcome, is rarely, if ever, discussed.

Despite the significant caveats, the published data show that radiotherapy is an effective treatment for functioning pituitary adenomas and NFPAs, with good long-term tumor and hormonal control. Long-term side effects of fractionated radiotherapy are acceptable, although improvements in late side effects continue to be sought.

The development of high-precision stereotactic treatment methods and hypofractionation with radiosurgery has coincided with the move away from routine postoperative radiotherapy to all patients, and the long-term outcomes of radiotherapy treatment using stereotactic methods would be expected to compare unfavorably with the results from fractionated conformal series. While the tumor control rates and hormonal control rates appear to be slightly lower, the results are comparable to those from studies where all patients were treated. The toxicity of radiosurgery for small tumors is comparable with that of fractionated treatments used for tumors of all sizes, with the current exception of second brain tumors. Currently, there is no clear evidence of superiority for one radiation method over another, if the tumor size and position allow the dosimetric constraints to be met.

REFERENCES

1. Gittoes NJ, Bates AS, Tse W, et al. Radiotherapy for non-function pituitary tumours. *Clin Endocrinol (Oxf)*. 1998;48:331–337.
2. Grigsby PW, Simpson JR, Emami BN, et al. Prognostic factors and results of surgery and postoperative irradiation in the management of pituitary adenomas. *Int J Radiat Oncol Biol Phys*. 1989;16:1411–1417.
3. McCollough WM, Marcus RB, Rhoton AL, et al. Long-term follow-up of radiotherapy for pituitary adenoma: the absence of late recurrence after greater than or equal to 4500 cGy. *Int J Radiat Oncol Biol Phys*. 1991;21:607–614.
4. Brada M, Rajan B, Traish D, et al. The long-term efficacy of conservative surgery and radiotherapy in the control of pituitary adenomas. *Clin Endocrinol (Oxf)*. 1993;38:571–578.
5. Tsang RW, Brierley JD, Panzarella T, et al. Radiation therapy for pituitary adenoma: treatment outcome and prognostic factors. *Int J Radiat Oncol Biol Phys*. 1994;30:557–565.
6. Zierhut D, Flentje M, Adolph J, et al. External radiotherapy of pituitary adenomas. *Int J Radiat Oncol Biol Phys*. 1995;33:307–314.
7. Rush S, Cooper PR. Symptom resolution, tumor control, and side effects following postoperative radiotherapy for pituitary macroadenomas. *Int J Radiat Oncol Biol Phys*. 1997;37:1031–1034.
8. Breen P, Flickinger JC, Kondziolka D, et al. Radiotherapy for nonfunctional pituitary adenoma: analysis of long-term tumor control. *J Neurosurg*. 1998;89:933–938.
9. Sasaki R, Murakami M, Okamoto Y, et al. The efficacy of conventional radiation therapy in the management of pituitary adenoma. *Int J Radiat Oncol Biol Phys*. 2000;47:1337–1345.

10. Minniti G, Brada M. Radiotherapy and radiosurgery for Cushing's disease. *Arq Bras Endocrinol Metabol*. 2007;51:1373–1380.
11. Minniti G, Clarke E, Scaringi C, et al. Stereotactic radiotherapy and radiosurgery for non-functioning and secreting pituitary adenomas. *Rep Pract Oncol Radiother*. 2016;21:370–378.
12. Khoo VS, Oldham M, Adams EJ, et al. Comparison of intensity-modulated tomotherapy with stereotactically guided conformal radiotherapy for brain tumors. *Int J Radiat Oncol Biol Phys*. 1999;45:415–425.
13. Karger CP, Jäkel O, Debus J, et al. Three-dimensional accuracy and interfractional reproducibility of patient fixation and positioning using a stereotactic head mask system. *Int J Radiat Oncol Biol Phys*. 2001;49:1493–1504.
14. Amichetti M, Amelio D, Minniti G. Radiosurgery with photons or protons for benign and malignant tumours of the skull base: a review. *Radiat Oncol*. 2012;7:210.
15. Lasak JM, Gorecki JP. The history of stereotactic radiosurgery and radiotherapy. *Otolaryngol Clin North Am*. 2009;42:593–599.
16. Mayo C, Martel MK, Marks LB, et al. Radiation dose-volume effects of optic nerves and chiasm. *Int J Radiat Oncol Biol Phys*. 2010;76:S28–S35.
17. Hiniker SM, Modlin LA, Choi CY, et al. Dose-response modeling of the visual pathway tolerance to single-fraction and hypofractionated stereotactic radiosurgery. *Semin Radiat Oncol*. 2016;26:97–104.
18. Larson DA, Flickinger JC, Loeffler JS. The radiobiology of radiosurgery. *Int J Radiat Oncol Biol Phys*. 1993;25:557–561.
19. Park C, Papiez L, Zhang S, et al. Universal survival curve and single fraction equivalent dose: useful tools in understanding potency of ablative radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;70:847–852.
20. Rosenfelder NA, Corsini L, McNair H, et al. Achieving the relocation accuracy of stereotactic frame-based cranial radiotherapy in a three-point thermoplastic shell. *Clin Oncol (R Coll Radiol)*. 2013;25:66–73.
21. Mohan R, Grosshans D. Proton therapy - present and future. *Adv Drug Deliv Rev*. 2017;109:26–44.
22. Reddy R, Cudlip S, Byrne JV, et al. Can we ever stop imaging in surgically treated and radiotherapy-naive patients with non-functioning pituitary adenoma? *Eur J Endocrinol*. 2011;165:739–744.
23. Dekkers OM, Pereira AM, Roelfsema F, et al. Observation alone after transsphenoidal surgery for nonfunctioning pituitary macroadenoma. *J Clin Endocrinol Metab*. 2006;91:1796–1801.
24. Jaffrain-Rea ML, Derome P, Bataini JP, et al. Influence of radiotherapy on long-term relapse in clinically non-secreting pituitary adenomas. A retrospective study (1970-1988). *Eur J Med*. 1993;2:398–403.
25. Woollons AC, Hunn MK, Rajapakse YR, et al. Non-functioning pituitary adenomas: indications for postoperative radiotherapy. *Clin Endocrinol (Oxf)*. 2000;53:713–717.
26. van den Bergh AC, van den Berg G, Schoorl MA, et al. Immediate postoperative radiotherapy in residual nonfunctioning pituitary adenoma: beneficial effect on local control without additional negative impact on pituitary function and life expectancy. *Int J Radiat Oncol Biol Phys*. 2007;67:863–869.
27. Chang EF, Zada G, Kim S, et al. Long-term recurrence and mortality after surgery and adjuvant radiotherapy for nonfunctional pituitary adenomas. *J Neurosurg*. 2008;108:736–745.
28. Olsson DS, Buchfelder M, Schlaffer S, et al. Comparing progression of non-functioning pituitary adenomas in hypopituitarism patients with and without long-term GH replacement therapy. *Eur J Endocrinol*. 2009;161:663–669.
29. Brochier S, Galland F, Kujas M, et al. Factors predicting relapse of non-functioning pituitary macroadenomas after neurosurgery: a study of 142 patients. *Eur J Endocrinol*. 2010;163:193–200.
30. Vargas G, Gonzalez B, Ramirez C, et al. Clinical characteristics and treatment outcome of 485 patients with nonfunctioning pituitary macroadenomas. *Int J Endocrinol*. 2015;2015:756069.
31. Deng WC, Yan JL, Chuang CC, et al. Adjuvant radiation therapy compared with observation alone for postoperative residual nonfunctional pituitary adenomas. *World Neurosurg*. 2019;128:e1024–e1033.
32. Molitch ME. Nonfunctioning pituitary tumors. *Handb Clin Neurol*. 2014;124:167–184.
33. Kotecha R, Sahgal A, Rubens M, et al. Stereotactic radiosurgery for non-functioning pituitary adenomas: meta-analysis and International Stereotactic Radiosurgery Society practice opinion. *Neuro Oncol*. 2020;22:318–332.
34. Wilson PJ, De-Loyde KJ, Williams JR, et al. A single centre's experience of stereotactic radiosurgery and radiotherapy for non-functioning pituitary adenomas with the Linear Accelerator (Linac). *J Clin Neurosci*. 2012;19:370–374.
35. Kim JO, Ma R, Akagami R, et al. Long-term outcomes of fractionated stereotactic radiation therapy for pituitary adenomas at the BC Cancer Agency. *Int J Radiat Oncol Biol Phys*. 2013;87:528–533.
36. Minniti G, Scaringi C, Poggi M, et al. Fractionated stereotactic radiotherapy for large and invasive non-functioning pituitary adenomas: long-term clinical outcomes and volumetric MRI assessment of tumor response. *Eur J Endocrinol*. 2015;172:433–441.
37. Puataweepong P, Dhanachai M, Hansasuta A, et al. Outcomes for pituitary adenoma patients treated with linac- based stereotactic radiosurgery and radiotherapy: a long term experience in Thailand. *Asian Pac J Cancer Prev*. 2015;16:5279–5284.
38. Barrande G, Pittino-Lungo M, Coste J, et al. Hormonal and metabolic effects of radiotherapy in acromegaly: long-term results in 128 patients followed in a single center. *J Clin Endocrinol Metab*. 2000;85:3779–3785.
39. Epaminonda P, Porretti S, Cappiello V, et al. Efficacy of radiotherapy in normalizing serum IGF-I, acid-labile subunit (ALS) and IGFBP-3 levels in acromegaly. *Clin Endocrinol (Oxf)*. 2001;55:183–189.
40. Minniti G, Jaffrain-Rea ML, Osti M, et al. The long-term efficacy of conventional radiotherapy in patients with GH-secreting pituitary adenomas. *Clin Endocrinol (Oxf)*. 2005;62:210–216.
41. Langsenlehner T, Stiegler C, Quehenberger F, et al. Long-term follow-up of patients with pituitary macroadenomas after postoperative radiation therapy: analysis of tumor control and functional outcome. *Strahlenther Onkol*. 2007;183:241–247.
42. Snead FE, Amdur RJ, Morris CG, et al. Long-term outcomes of radiotherapy for pituitary adenomas. *Int J Radiat Oncol Biol Phys*. 2008;71:994–998.
43. Erridge SC, Conkey DS, Stockton D, et al. Radiotherapy for pituitary adenomas: long-term efficacy and toxicity. *Radiother Oncol*. 2009;93:597–601.
44. Rim CH, Yang DS, Park YJ, et al. Radiotherapy for pituitary adenomas: long-term outcome and complications. *Radiat Oncol J*. 2011;29:156–163.
45. Kim MY, Kim JH, Oh YK, et al. Long-term outcomes of surgery and radiotherapy for secreting and non-secreting pituitary adenoma. *Radiat Oncol J*. 2016;34:121–127.
46. Scheick S, Amdur RJ, Kirwan JM, et al. Long-term outcome after fractionated radiotherapy for pituitary adenoma: the curse of the secretory tumor. *Am J Clin Oncol*. 2016;39:49–54.
47. Rieken S, Habermehl D, Welzel T, et al. Long term toxicity and prognostic factors of radiation therapy for secreting and non-secreting pituitary adenomas. *Radiat Oncol*. 2013;8:18.
48. Colin P, Jovenin N, Delemer B, et al. Treatment of pituitary adenomas by fractionated stereotactic radiotherapy: a prospective study of 110 patients. *Int J Radiat Oncol Biol Phys*. 2005;62:333–341.
49. Sheehan JP, Xu Z, Lobo MJ. External beam radiation therapy and stereotactic radiosurgery for pituitary adenomas. *Neurosurg Clin N Am*. 2012;23:571–586.
50. Minniti G, Traish D, Ashley S, et al. Fractionated stereotactic conformal radiotherapy for secreting and nonsecreting pituitary adenomas. *Clin Endocrinol*. 2006;64:542–548.
51. Kong DS, Lee JI, Lim DH, et al. The efficacy of fractionated radiotherapy and stereotactic radiosurgery for pituitary adenomas: long-term results of 125 consecutive patients treated in a single institution. *Cancer*. 2007;110:854–860.
52. Schalin-Jäntti C, Valanne L, Tenhunen M, et al. Outcome of fractionated stereotactic radiotherapy in patients with pituitary adenomas resistant to conventional treatments: a 5.25-year follow-up study. *Clin Endocrinol*. 2010;73:72–77.
53. Barber SM, Teh BS, Baskin DS. Fractionated stereotactic radiotherapy for pituitary adenomas: single-center experience in 75 consecutive patients. *Neurosurgery*. 2016;79:406–417.

54. Kopp C, Theodorou M, Poullos N, et al. Tumor shrinkage assessed by volumetric MRI in long-term follow-up after fractionated stereotactic radiotherapy of nonfunctioning pituitary adenoma. *Int J Radiat Oncol Biol Phys*. 2012;82:1262–1267.
55. Sheehan JP, Starke RM, Mathieu D, et al. Gamma Knife radiosurgery for the management of nonfunctioning pituitary adenomas: a multicenter study. *J Neurosurg*. 2013;119:446–456.
56. Starke RM, Williams BJ, Jane JA, et al. Gamma Knife surgery for patients with nonfunctioning pituitary macroadenomas: predictors of tumor control, neurological deficits, and hypopituitarism. *J Neurosurg*. 2012;117:129–135.
57. Iwai Y, Yamanaka K, Yoshioka K. Radiosurgery for nonfunctioning pituitary adenomas. *Neurosurgery*. 2005;56:699–705; discussion 699–705.
58. Pollock BE, Cochran J, Natt N, et al. Gamma knife radiosurgery for patients with nonfunctioning pituitary adenomas: results from a 15-year experience. *Int J Radiat Oncol Biol Phys*. 2008;70:1325–1329.
59. Park KJ, Kano H, Parry PV, et al. Long-term outcomes after gamma knife stereotactic radiosurgery for nonfunctional pituitary adenomas. *Neurosurgery*. 2011;69:1188–1199.
60. Losa M, Spatola G, Albano L, et al. Frequency, pattern, and outcome of recurrences after gamma knife radiosurgery for pituitary adenomas. *Endocrine*. 2017;56:595–602.
61. Heringer LC, Machado de Lima M, Rotta JM, et al. Effect of stereotactic radiosurgery on residual or relapsed pituitary adenoma: a systematic review and meta-analysis. *World Neurosurg*. 2020;136:374–381.e4.
62. Melmed S, Colao A, Barkan A, et al. Guidelines for acromegaly management: an update. *J Clin Endocrinol Metab*. 2009;94:1509–1517.
63. Colao A, Grasso LFS, Giustina A, et al. Acromegaly. *Nat Rev Dis Primers*. 2019;5:20.
64. Biermasz NR, van Dulken H, Roelfsema F. Long-term follow-up results of postoperative radiotherapy in 36 patients with acromegaly. *J Clin Endocrinol Metab*. 2000;85:2476–2482.
65. Milker-Zabel S, Debus J, Thilmann C, et al. Fractionated stereotactically guided radiotherapy and radiosurgery in the treatment of functional and nonfunctional adenomas of the pituitary gland. *Int J Radiat Oncol Biol Phys*. 2001;50:1279–1286.
66. Milker-Zabel S, Zabel A, Huber P, et al. Stereotactic conformal radiotherapy in patients with growth hormone-secreting pituitary adenoma. *Int J Radiat Oncol Biol Phys*. 2004;59:1088–1096.
67. Jenkins PJ, Bates P, Carson MN, et al. Conventional pituitary irradiation is effective in lowering serum growth hormone and insulin-like growth factor-I in patients with acromegaly. *J Clin Endocrinol Metab*. 2006;91:1239–1245.
68. Tsang RW, Brierley JD, Panzarella T, et al. Role of radiation therapy in clinical hormonally-active pituitary adenomas. *Radiother Oncol*. 1996;41:45–53.
69. Jallad RS, Musolino NR, Salgado LR, et al. Treatment of acromegaly: is there still a place for radiotherapy? *Pituitary*. 2007;10:53–59.
70. Diallo AM, Colin P, Litre CF, et al. Long-term results of fractionated stereotactic radiotherapy as third-line treatment in acromegaly. *Endocrine*. 2015;50:741–748.
71. Knappe UJ, Petroff D, Quinkler M, et al. Fractionated radiotherapy and radiosurgery in acromegaly: analysis of 352 patients from the German Acromegaly Registry. *Eur J Endocrinol*. 2020;182:275–284.
72. Gheorghiu ML, Fleseriu M. Stereotactic radiation therapy in pituitary adenomas, is it better than conventional radiation therapy? *Acta Endocrinol (Buchar)*. 2017;13:476–490.
73. Estrada J, Boronat M, Mielgo M, et al. The long-term outcome of pituitary irradiation after unsuccessful transsphenoidal surgery in Cushing's disease. *N Engl J Med*. 1997;336:172–177.
74. Minniti G, Osti M, Jaffrain-Rea ML, et al. Long-term follow-up results of postoperative radiation therapy for Cushing's disease. *J Neuro Oncol*. 2007;84:79–84.
75. Mehta GU, Lonser RR. Management of hormone-secreting pituitary adenomas. *Neuro Oncol*. 2017;19:762–773.
76. Mehta AE, Reyes FI, Faiman C. Primary radiotherapy of prolactinomas. Eight- to 15-year follow-up. *Am J Med*. 1987;83:49–58.
77. Zheng Q, Huang Y, Lin W, et al. Comparing Stereotactic Radiosurgery and Fractionated Stereotactic Radiotherapy in Treating patients with Growth Hormone-Secreting Adenomas: A Systematic Review and Meta-analysis. *Endocr Pract*. 2020. [Epub ahead of print].
78. Singh R, Didwania P, Lehrer EJ, et al. Stereotactic radiosurgery for acromegaly: an international systematic review and meta-analysis of clinical outcomes. *J Neuro Oncol*. 2020;148:401–418.
79. Ajithkumar TBM. Pituitary radiotherapy. In: Harris PE, Bouloux P-MG, eds. *Endocrinology in Practice.* 2nd ed. Boca Raton, Florida: CRC Press.
80. Sheehan JP, Xu Z, Salvetti DJ, et al. Results of gamma knife surgery for Cushing's disease. *J Neurosurg*. 2013;119:1486–1492.
81. Pivonello R, De Leo M, Cozzolino A, et al. The treatment of Cushing's disease. *Endocr Rev*. 2015;36:385–486.
82. Jagannathan J, Sheehan JP, Pouratian N, et al. Gamma Knife surgery for Cushing's disease. *J Neurosurg*. 2007;106:980–987.
83. Bunevicius A, Laws ER, Vance ML, et al. Surgical and radiosurgical treatment strategies for Cushing's disease. *J Neuro Oncol*. 2019;145:403–413.
84. Pouratian N, Sheehan J, Jagannathan J, et al. Gamma knife radiosurgery for medically and surgically refractory prolactinomas. *Neurosurgery*. 2006;59:255–266; discussion 255–266.
85. Tanaka S, Link MJ, Brown PD, et al. Gamma knife radiosurgery for patients with prolactin-secreting pituitary adenomas. *World Neurosurg*. 2010;74:147–152.
86. Dekkers OM, de Keizer RJ, Roelfsema F, et al. Progressive improvement of impaired visual acuity during the first year after transsphenoidal surgery for non-functioning pituitary macroadenoma. *Pituitary*. 2007;10:61–65.
87. Colao A, Cerbone G, Cappabianca P, et al. Effect of surgery and radiotherapy on visual and endocrine function in nonfunctioning pituitary adenomas. *J Endocrinol Invest*. 1998;21:284–290.
88. Liao HI, Wang CC, Wei KC, et al. Fractionated stereotactic radiosurgery using the Novalis system for the management of pituitary adenomas close to the optic apparatus. *J Clin Neurosci*. 2014;21:111–115.
89. Tomlinson JW, Holden N, Hills RK, et al. Association between premature mortality and hypopituitarism. West midlands prospective hypopituitary study group. *Lancet*. 2001;357:425–431.
90. Minniti G, Flickinger J, Tolu B, et al. Management of nonfunctioning pituitary tumors: radiotherapy. *Pituitary*. 2018;21:154–161.
91. Minniti G, Gilbert DC, Brada M. Modern techniques for pituitary radiotherapy. *Rev Endocr Metab Disord*. 2009;10:135–144.
92. Minniti G, Jaffrain-Rea ML, Osti M, et al. Radiotherapy for nonfunctioning pituitary adenomas: from conventional to modern stereotactic radiation techniques. *Neurosurg Rev*. 2007;30:167–175; discussion 175–176.
93. Marek J, Jezková J, Hána V, et al. Is it possible to avoid hypopituitarism after irradiation of pituitary adenomas by the Leksell gamma knife? *Eur J Endocrinol*. 2011;164:169–178.
94. Graffeo CS, Link MJ, Brown PD, et al. Hypopituitarism after single-fraction pituitary adenoma radiosurgery: dosimetric analysis based on patients treated using contemporary techniques. *Int J Radiat Oncol Biol Phys*. 2018;101:618–623.
95. Chanson P, Dormoy A, Dekkers OM. Use of radiotherapy after pituitary surgery for non-functioning pituitary adenomas. *Eur J Endocrinol*. 2019;181:D1–D13.
96. Stafford SL, Pollock BE, Leavitt JA, et al. A study on the radiation tolerance of the optic nerves and chiasm after stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys*. 2003;55:1177–1181.
97. Pollock BE, Link MJ, Leavitt JA, et al. Dose-volume analysis of radiation-induced optic neuropathy after single-fraction stereotactic radiosurgery. *Neurosurgery*. 2014;75:456–460; discussion 460.
98. Milano MT, Grimm J, Soltys SG, et al. Single- and multi-fraction stereotactic radiosurgery dose tolerances of the optic pathways. *Int J Radiat Oncol Biol Phys*. 2021;110:87–99.
99. Becker G, Kocher M, Kortmann RD, et al. Radiation therapy in the multimodal treatment approach of pituitary adenoma. *Strahlenther Onkol*. 2002;178:173–186.
100. Grigsby PW, Simpson JR, Stokes S, et al. Results of surgery and irradiation or irradiation alone for pituitary adenomas. *J Neuro Oncol*. 1988;6:129–134.
101. Narayan V, Mohammed N, Bir SC, et al. Long-term outcome of nonfunctioning and hormonal active pituitary adenoma after gamma knife radiosurgery. *World Neurosurg*. 2018;114:e824–e832.

102. Sherry AD, Khattab MH, Xu MC, et al. Outcomes of stereotactic radiosurgery and hypofractionated stereotactic radiotherapy for refractory Cushing's disease. *Pituitary*. 2019;22:607–613.
103. Brada M, Burchell L, Ashley S, et al. The incidence of cerebrovascular accidents in patients with pituitary adenoma. *Int J Radiat Oncol Biol Phys*. 1999;45:693–698.
104. Brada M, Ashley S, Ford D, et al. Cerebrovascular mortality in patients with pituitary adenoma. *Clin Endocrinol*. 2002;57:713–717.
105. Erfurth EM, Bülow B, Svahn-Tapper G, et al. Risk factors for cerebrovascular deaths in patients operated and irradiated for pituitary tumors. *J Clin Endocrinol Metab*. 2002;87:4892–4899.
106. Murphy ES, Xie H, Merchant TE, et al. Review of cranial radiotherapy-induced vasculopathy. *J Neuro Oncol*. 2015;122:421–429.
107. Olsson DS, Bryngelsson IL, Ragnarsson O. Higher incidence of morbidity in women than men with non-functioning pituitary adenoma: a Swedish nationwide study. *Eur J Endocrinol*. 2016;175:55–61.
108. Sattler MG, Vroomen PC, Sluiter WJ, et al. Incidence, causative mechanisms, and anatomic localization of stroke in pituitary adenoma patients treated with postoperative radiation therapy versus surgery alone. *Int J Radiat Oncol Biol Phys*. 2013;87:53–59.
109. van Westrhenen A, Muskens IS, Verhoeff JJC, et al. Ischemic stroke after radiation therapy for pituitary adenomas: a systematic review. *J Neuro Oncol*. 2017;135:1–11.
110. Inoue HK, Kohga H, Hirato M, et al. Pituitary adenomas treated by microsurgery with or without Gamma Knife surgery: experience in 122 cases. *Stereotact Funct Neurosurg*. 1999;72:125–131.
111. Bliss P, Kerr GR, Gregor A. Incidence of second brain tumours after pituitary irradiation in Edinburgh 1962-1990. *Clin Oncol*. 1994;6:361–363.
112. Erfurth EM, Bülow B, Mikoczy Z, et al. Is there an increase in second brain tumours after surgery and irradiation for a pituitary tumour? *Clin Endocrinol (Oxf)*. 2001;55:613–616.
113. Sattler MG, van Beek AP, Wolffenbuttel BH, et al. The incidence of second tumours and mortality in pituitary adenoma patients treated with postoperative radiotherapy versus surgery alone. *Radiother Oncol*. 2012;104:125–130.
114. Ecemis GC, Atmaca A, Meydan D. Radiation-associated secondary brain tumors after conventional radiotherapy and radiosurgery. *Expert Rev Neurother*. 2013;13:557–565.
115. Minniti G, Traish D, Ashley S, et al. Risk of second brain tumor after conservative surgery and radiotherapy for pituitary adenoma: update after an additional 10 years. *J Clin Endocrinol Metab*. 2005;90:800–804.
116. Burman P, van Beek AP, Biller BM, et al. Radiotherapy, especially at young age, increases the risk for de novo brain tumors in patients treated for pituitary/sellar lesions. *J Clin Endocrinol Metab*. 2017;102:1051–1058.
117. Yamanaka R, Abe E, Sato T, et al. Secondary intracranial tumors following radiotherapy for pituitary adenomas: a systematic review. *Cancers (Basel)*. 2017;9:103.
118. van Varsseveld NC, van Bunderen CC, Ubachs DH, et al. Cerebrovascular events, secondary intracranial tumors, and mortality after radiotherapy for nonfunctioning pituitary adenomas: a subanalysis from the Dutch National Registry of Growth Hormone Treatment in Adults. *J Clin Endocrinol Metab*. 2015;100:1104–1112.
119. Brada M, Ford D, Ashley S, et al. Risk of second brain tumour after conservative surgery and radiotherapy for pituitary adenoma. *BMJ*. 1992;304:1343–1346.
120. Norberg L, Johansson R, Rasmuson T. Intracranial tumours after external fractionated radiotherapy for pituitary adenomas in northern Sweden. *Acta Oncol*. 2010;49:1276–1282.
121. Bengtsson BA, Edén S, Ernest I, et al. Epidemiology and long-term survival in acromegaly. A study of 166 cases diagnosed between 1955 and 1984. *Acta Med Scand*. 1988;223:327–335.
122. Boguszewski CL, Ayuk J. Management of endocrine disease: acromegaly and cancer: an old debate revisited. *Eur J Endocrinol*. 2016;175:R147–R156.
123. Gheorghiu ML. Updates in outcomes of stereotactic radiation therapy in acromegaly. *Pituitary*. 2017;20:154–168.
124. Wolf A, Naylor K, Tam M, et al. Risk of radiation-associated intracranial malignancy after stereotactic radiosurgery: a retrospective, multicentre, cohort study. *Lancet Oncol*. 2019;20:159–164.
125. Colao A, Vandeva S, Pivonello R, et al. Could different treatment approaches in acromegaly influence life expectancy? A comparative study between Bulgaria and Campania (Italy). *Eur J Endocrinol*. 2014;171:263–273.
126. Mestron A, Webb SM, Astorga R, et al. Epidemiology, clinical characteristics, outcome, morbidity and mortality in acromegaly based on the Spanish Acromegaly Registry (Registro Espanol de Acromegalia, REA). *Eur J Endocrinol*. 2004;151:439–446.
127. Giustina A, Barkan A, Beckers A, et al. A consensus on the diagnosis and treatment of acromegaly comorbidities: an update. *J Clin Endocrinol Metab*. 2020;105:dgz096.
128. Brown PD, Blanchard M, Jethwa K, et al. The incidence of cerebrovascular accidents and second brain tumors in patients with pituitary adenoma: a population-based study. *Neurooncol Pract*. 2014;1:22–28.
129. Fleseriu M, Hashim IA, Karavitaki N, et al. Hormonal replacement in hypopituitarism in adults: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:3888–3921.
130. Ntali G, Capatina C, Fazal-Sanderson V, et al. Mortality in patients with non-functioning pituitary adenoma is increased: systematic analysis of 546 cases with long follow-up. *Eur J Endocrinol*. 2016;174:137–145.
131. Lesueur P, Calugaru V, Nauraye C, et al. Proton therapy for treatment of intracranial benign tumors in adults: a systematic review. *Cancer Treat Rev*. 2019;72:56–64.
132. Lüdecke DK, Lutz BS, Niedworok G. The choice of treatment after incomplete adenomectomy in acromegaly: proton--versus high voltage radiation. *Acta Neurochir (Wien)*. 1989;96:32–38.
133. Brada M, Jankowska P. Radiotherapy for pituitary adenomas. *Endocrinol Metab Clin North Am*. 2008;37:263–275, xi.
134. Schoenthaler R, Albright NW, Wara WM, et al. Re-irradiation of pituitary adenoma. *Int J Radiat Oncol Biol Phys*. 1992;24:307–314.
135. Edwards AA, Swords FM, Plowman PN. Focal radiation therapy for patients with persistent/recurrent pituitary adenoma, despite previous radiotherapy. *Pituitary*. 2009;12:30–34.
136. Verma J, McCutcheon IE, Waguespack SG, et al. Feasibility and outcome of re-irradiation in the treatment of multiply recurrent pituitary adenomas. *Pituitary*. 2014;17:539–545.
137. Alonso CE, Bunevicius A, Trifiletti DM, et al. Safety and efficacy of repeat radiosurgery for acromegaly: an International Multi-Institutional Study. *J Neuro Oncol*. 2019;145:301–307.

17

Somatic Growth and Maturation: Growth Hormone and Other Growth Factors

Alan D. Rogol and Robert M. Malina

OUTLINE

Introduction, 250
  Prenatal Considerations, 251
  Chronological Age, 252
  Growth Status, 252
  Growth Rate, 259
  Maturity Status, 259
  Skeletal Maturity Status, 259
  Secondary Sex Characteristics, 261
  Maturity Timing, 262
  Tempo of Maturation, 263
  Evaluation of Children for Disorders of Growth, 263
  Growth Charts, 264
Evaluation of Children With Short Stature, 264
  History, 265
  Physical Examination, 265
  Preliminary Laboratory Evaluation, 265
  Evaluation for Endocrine Disorders of Growth, 266
  Growth Hormone/Insulin Growth Factor-1 Axis, 266
  Growth Hormone Release and Action, 267
  Diagnostic Approach, 269
  Growth Hormone Stimulation Tests, 270
  Limitations of Growth Hormone Stimulation Tests, 270
  Growth Hormone Treatment (See Also Chapter 18), 270
  Dosing and Monitoring of Growth Hormone Therapy, 270
  Retesting for Growth Hormone Deficiency During the Transition to Emerging Adulthood, 270
Evaluation of Children With Tall Stature, 271
Endocrine Causes of Tall Stature, 271
  Treatment of Children/Adolescents With Tall Stature, 272
  Growth Evaluation of Children With Obesity, 272
Conclusions, 274

INTRODUCTION

One of the most common reasons for referral to a pediatric endocrine clinic is to evaluate an infant/child/adolescent for short stature in relation to their calendar or chronological age (CA). Allowing for the relatively wide range of normal variability in stature and rate of growth in stature, several clinical conditions suggest a need for appropriate therapy, even if the latter is just reassurance.

Nutritional status and the psychosocial concomitants of short stature or delayed maturation are inextricably entwined in such evaluations. These issues are seemingly more apparent during the interval of puberty, when the youngster stands out relative to same age peers who have already begun pubertal maturation. Of relevance, socialization and sports participation often track with maturational (biological) age to a greater degree than CA.

Evaluation of children for short stature and/or delayed maturation begins somewhat differently for boys and girls, and the differential diagnosis may be quite different after allowing for normal variation in growth and maturation. From the perspective of both children and parents, it appears that the psychosocial aspects of being short affect boys more so than girls, which often motivates earlier evaluation of boys by a pediatric endocrinologist.

The process of "growing up" among children and youth is often taken for granted. The process is in fact more complex and involves three interacting, though distinct, processes: physical growth, biological maturation, and behavioral development.[1] The latter is often subsumed in the term "socialization," which is specific to a culture. The three processes occur simultaneously and dominate the daily lives of children and adolescents (and their families) for the first two decades of life.

Growth refers to the increase in size of the body as a whole and of its parts. As children grow, height and weight increase. In addition, different segments of the body, e.g., extremities and trunk, grow at different rates and at different times, resulting in altered body proportions. Weight is a heterogeneous mass that includes a variety of organs, the skeleton, musculature, and fat or adipose tissue.

Maturation refers to progress towards the biologically mature state (i.e., maturity). Maturation occurs in all bodily organs and systems, but maturity varies with the individual biological system considered. Sexual maturity implies fully functional reproductive capability. Skeletal maturity is a fully ossified adult skeleton. Maturation of the nervous and endocrine systems are major factors that influence somatic growth and sexual and skeletal maturation. The processes of physical growth and biological maturation are the focus of this chapter. They occur concurrently and are related, and vary in *tempo* and *timing*. The former refers to the rate at which the processes progress from infancy through adolescence, while the latter refers to the ages when specific events or milestones occur. Indicators of growth may also be used in deriving estimates of maturation.

Although not a focus of this chapter, development refers to the acquisition and refinement of behaviors expected and in many instances set by society. As children experience life at home, neighborhood, school, church, sports, recreation, and other community activities, they develop cognitive, social, affective, moral, motor, and other behaviors expected by society, i.e., behavioral competence in a variety of domains. Proficiency in motor or movement behaviors, i.e.,

TABLE 17.1 **Comparison of features of the Silver-Russell and Beckwith-Wiedemann Syndromes**

Feature	Silver–Russell	Beckwith–Wiedemann
OMIM #	180860	130650
Epigenetic locus Imprinting center (control of fetal growth)	11p15.5	11p15.5
Major clinical findings (heterogeneous)	Severe intrauterine growth restriction; small for gestational age without catch-up growth; conserved head circumference (relative macrocephaly); triangular facies; skeletal asymmetry; feeding difficulties, especially early in life; some with early adrenarche or puberty; high probability of metabolic syndrome as adults	Macrosomia (90%); macroglossia; abdominal wall defects (omphalocoele); visceromegaly (kidneys, liver, pancreas); childhood embryonal tumors (Wilms tumor and hepatoblastoma most common)
Frequency	1/30,000–1/100,000 (may be higher given difficulty in diagnosis); diagnosis based on clinical findings—clinical scoring system	1/10,300–1/13,700; diagnosis based on clinical findings
Metabolic consequences	High probability of metabolic syndrome as adult	Neonatal hypoglycemia
Heritability		~85% sporadic

running, jumping, hopping, etc. is related in part to growth, including size and body composition. This association is especially apparent during pubertal maturation and the growth spurt.

Prenatal Considerations

Although this chapter focuses on the postnatal years and largely adolescence, it is important to understand the prenatal background of these processes. In this context, the potential role of maternal characteristics and behaviors related to energy intake and weight gain, alcohol consumption, smoking, and drug use have received attention. Evidence has accumulated over the past 30 to 40 years supporting the concept that events occurring prenatally condition postnatal growth and maturation, and perhaps adulthood and aging. This is often set in the context of the "fetal origins hypothesis" (Barker hypothesis), which emphasizes associations between birth weight, as a marker of fetal growth, and risk for several chronic, degenerative diseases in adulthood. The hypothesis highlights interactions among factors that influence prenatal growth and factors in the postnatal environment, especially diet and stress, which may have consequences of disease many decades after the original insult.

Normal fetal growth may be divided into two epochs. The first, roughly the first trimester of pregnancy, consists of proliferation, organization, and differentiation of the embryo. The second, comprising the remainder of pregnancy, involves continuing growth and functional maturation of the different tissues and organs of the fetus.[2] All of the following must be working simultaneously and in a coordinated fashion: genetic programming, placenta, integrity of the materno-placento-fetal unit, and an adequate supply of nutrients and oxygen, along with the proper hormonal milieu (fetal and maternal).

The hormonal requirements for growth differ markedly from the pre- to postnatal periods. *In utero* the Insulin-like growth factor (IGFs) and insulin predominate, without much effect of growth hormone (GH) or thyroid hormone. Congenitally GH- or thyroid hormone–deficient fetuses have a normal or near-normal birth length. The maternal hormone production and blood flow are regulated by the placenta and contribute to the differentiation and subsequent growth of the embryo/fetus. Predominant among the substrates are glucose and amino acids, which depend on their transport proteins. Fatty acid delivery meets the requirements for cell membrane synthesis, for other hormones, and as a source of total energy.

Rodent models that lack IGF-1, IGF-2, or the IGF-1 receptor have reduced birth weights, and deletion of the IGF-1 type 1 receptor gene results in greater growth retardation that deletion of *IGF-1* or *IGF-2* alone (in addition to being lethal), suggesting a role for both IGF-1 and IGF-2 in prenatal growth. IGF-1 and IGF-2 act through the IGF-1, insulin, and IGF-2 receptors, and IGF-2 is approximately equal in importance to IGF-1 for fetal growth, with each contributing approximately 40%.[3]

Six IGF-binding proteins (IGFBPs) regulate the amount of free IGF available and are regulated in turn by IGFBP proteases. High levels of IGFBPs have been associated with fetal growth inhibition, likely from sequestration of fetally derived IGF-1.[4,5] IGFBP-3 also prolongs the half-life of IGFs in circulation, and levels are reduced in the cord serum of small for gestational age (SGA) fetuses.[5]

Pregnancy-associated plasma protein A (PAPP-A), which cleaves IGFBP-4, increases free IGF availability,[6] and low first-trimester PAPP-A levels in pregnant women are associated with lower birth weight.[7]

Placental IGF-2 is an important factor for fetal growth and nutrient delivery. Disturbances in its production or action may lead to the asymmetric growth retardation of Silver–Russell syndrome, whereas overexpression of *IGF-2* leads to a large placenta and the Beckwith–Wiedemann syndrome of overgrowth (Table 17.1).

In addition to IGFs, insulin influences fetal growth through its anabolic effects on carbohydrate and protein metabolism. Infants of diabetic mothers and children with Beckwith–Weidemann syndrome (with hyperinsulinism) have excessive fetal growth, and those with pancreatic agenesis have poor fetal growth. Mutations in the insulin receptor and in IRS-1, a downstream molecule in the signaling pathways of both IGF-1 and insulin receptors, are associated with suboptimal fetal growth and insulin resistance.[8]

As noted, understanding the regulation of fetal growth has assumed particular importance because of potential links between prenatal growth and later disease. Strong experimental evidence in animal models indicates that an adverse fetal environment, as reflected in birth size, can lead to poor health outcomes in adults, likely through an epigenetic mechanism.[9,10] Diverse causes of poor fetal growth (including maternal undernutrition or glucocorticoid exposure) can have similar deleterious effects on postnatal health (including hypertension, cardiovascular disease, and glucose intolerance). Considerable human data support such a "developmental origins of health and disease" model.[9-12] The risks for insulin resistance, diabetes, and visceral adiposity are particularly high in SGA infants with rapid postnatal "catch-up" growth.[9,13] Girls with this condition may undergo slightly early (not precocious) but quite rapid pubertal maturation, losing a significant proportion of their pubertal growth (height) spurt.

KEY POINTS

- The prenatal environment is critical to the fetal (developmental) origins of adult diseases—developmental origins of health and disease. These conditions are one of the drivers of noncommunicable, especially cardiovascular, disease in older adults.

Chronological Age

Prior to discussing growth and maturation, it is essential to address CA, which is the point of reference in studies of growth and maturation *per se*, and also in the clinical context. CA is calculated as the difference between the date of measurement/observation and date of birth, and is ordinarily expressed as decimals of the whole year. Use of whole years, i.e., 6 years of age, can be misleading.

Growth Status

Growth status refers to the size attained at the date of observation. Height (length from birth to 2 years) and weight are the primary indicators of growth status. Height, or more appropriately standing height, is the distance from the standing surface to the top of the skull. Sitting height, the distance from the flat sitting surface to the top of the skull, is often measured and provides information on upper body segment length. Standing height minus sitting height provides an estimate of leg or lower body length (Fig. 17.1). Weight is a measure of body mass, which is heterogeneous in composition; it is generally partitioned into its fat and lean components. The latter includes two major components, lean tissue mass (LTM) or fat-free mass (FFM) and bone mineral content (BMC), while the former is expressed as fat mass (FM) or as a percentage of body weight (% FM).

Measurements of weight, height, and sitting height should be made by trained individuals using standard techniques (Fig. 17.1).[14] Quality control is essential; it refers to the accuracy and reliability of measurements, and to measurement variability within (intraobserver) and between (interobserver) technicians.[15]

Height and weight increase gradually through childhood, increase at an accelerated rate during adolescence (growth spurt), and then slowly increase into late adolescence. Growth in height stops in the late teens or early twenties, whereas weight often continues to increase. FFM, LTM, and BMC have a growth pattern like height and weight, and each has an adolescent spurt, while FM increases more gradually with CA. Relative fatness (% fat) increases during childhood, but declines during adolescence in males and continues to increase at a slower pace in females during adolescence. The decline in % fat in males is due to the rapid growth in FFM (Figs. 17.2 and 17.3).[1]

Growth Charts. The growth status of a child or group of children is commonly compared to growth charts, which are a reference of comparison. The World Health Organization (WHO)[16] defines a reference "...as a tool for grouping and analyzing data and provides a common basis for comparing populations." Several growth charts for height and weight are available. It is important to note that *reference values* are not *standards*. A standard is prescriptive and suggests the way things *ought* to be, and as such has an associated value judgment. The WHO has produced a set of standards for the growth of infants and children up to the age of 5 years; however, there are several countries in which infants and children grow in slight excess of them, lessening their utility (Fig. 17.4).[17]

Height and weight charts commonly used in the United States are based on surveys of nationally representative samples of children and youth between the 1960s and 1994, except for length at birth, which are from surveys in two states (Fig. 17.4).[18,19] Heights of American children and youth did not differ significantly between the 1960s and 1990s; however, body weights of children greater than or equal to 6 years increased significantly between surveys in 1976 to 1980 and 1988 to 1994.[18,19] The gain in body weight from the late 1970s through the mid-1990s was viewed as undesirable from a public health perspective, so that in developing the growth charts for body weight, the values from 1988 to 1994 were not used. The charts for weight are thus to some extent a standard, or criterion-referenced, in that public health authorities decided on levels of weight that are presumably conducive to better health. Although the US growth charts are seemingly somewhat dated, changes in the heights of American girls and boys have been relatively small across several national surveys spanning 2002 through 2014, although weights have increased over time.[20-23]

Sex-specific charts for recumbent length and weight are provided for children from birth to 3 years; the data for length at birth are from surveys in two states (Fig. 17.4).[18,19] Corresponding charts for standing height and weight are provided for children and youth from 2 to 20 years of age (Fig. 17.4).

Many other countries have designated "country-specific" growth curves to better represent their present status. These are reference charts and may deviate from the standards of the WHO. For example, in Indonesia the WHO charts significantly overestimate the true prevalence of undernutrition, because healthy children 0 to 5 years are short and slim when plotted on the WHO charts, with mean z-scores of approximately -2.0, indicating that approximately 50% of the children are stunted (Fig. 17.5).[24]

Weight for height, expressed as the body mass index (BMI, [weight (kg)/ height (m^2)]), is commonly used to evaluate the weight status of children and youth spanning the spectrum from low (often labeled as underweight or extreme thinness) through excess (labeled as obesity) weight for height. The BMI is often considered an indicator of adiposity or fatness, but it is not a measure of body composition. Moreover, the BMI is significantly correlated with both FFM and FM in normal-weight youth[25] and is perhaps more closely associated with LTM rather than FM among relatively thin youth (Table 17.2).[26]

Several charts of cutoff values of the BMI designed to define the weight status of an individual are available. The focus, however, is largely on overweight and obesity. The charts for American children and youth (see above)[18] and those recommended by the WHO[27] present specific percentiles for BMI. The WHO charts were developed from the WHO growth standards for children from birth to 5 years of age and the 1977 United States reference data at older ages using updated statistical methods.[27] The International Obesity Task Force (IOTF), in contrast, used adult criteria to define cut-points for overweight and obesity,[28] and also for several levels of thinness.[29] Data for nationally representative samples of children and adolescents aged 2 to 18 years from Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States were used to develop the IOTF criteria.[28,29] In establishing the cutoff points, curves were mathematically fitted to pooled BMI data from the six countries so that they passed through adult criteria to define overweight (25.0 kg/m^2) and obesity (30 kg/m^2) and three grades of thinness (16.0, 17.0, and 18.5 kg/m^2) at 18 years of age.[30]

Given the variety of criteria for evaluating weight status with the BMI, there is a need for a critical evaluation of the suggested cutoffs. For example, US and IOTF cut-points for BMI differ for obesity, but are reasonably similar for overweight. One can also inquire into the

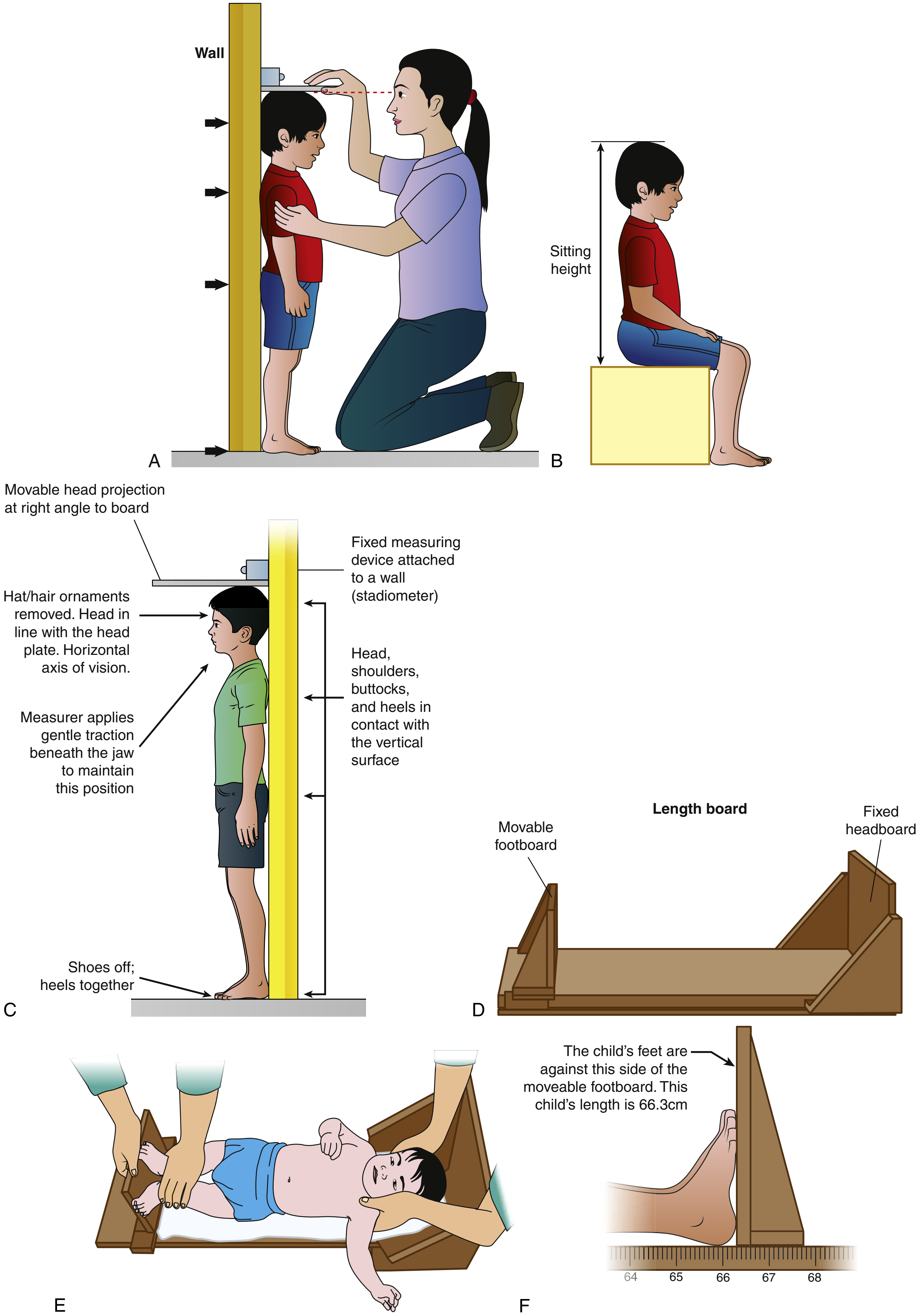

Wall
A
B
Sitting height
Movable head projection at right angle to board
Fixed measuring device attached to a wall (stadiometer)
Hat/hair ornaments removed. Head in line with the head plate. Horizontal axis of vision.
Head, shoulders, buttocks, and heels in contact with the vertical surface
Measurer applies gentle traction beneath the jaw to maintain this position
Shoes off; heels together
C
Length board
Movable footboard
Fixed headboard
D
The child's feet are against this side of the moveable footboard. This child's length is 66.3cm
64
65
66
67
68
E
F

Fig. 17.1 Measurement of Infants and Children. **A,** Child with vertical stadiometer. **B,** Sitting height. **C,** Vertical stadiometer with instructions for measurement. **D,** Length board. **E,** Length board with child. **F,** Scale for the length board.

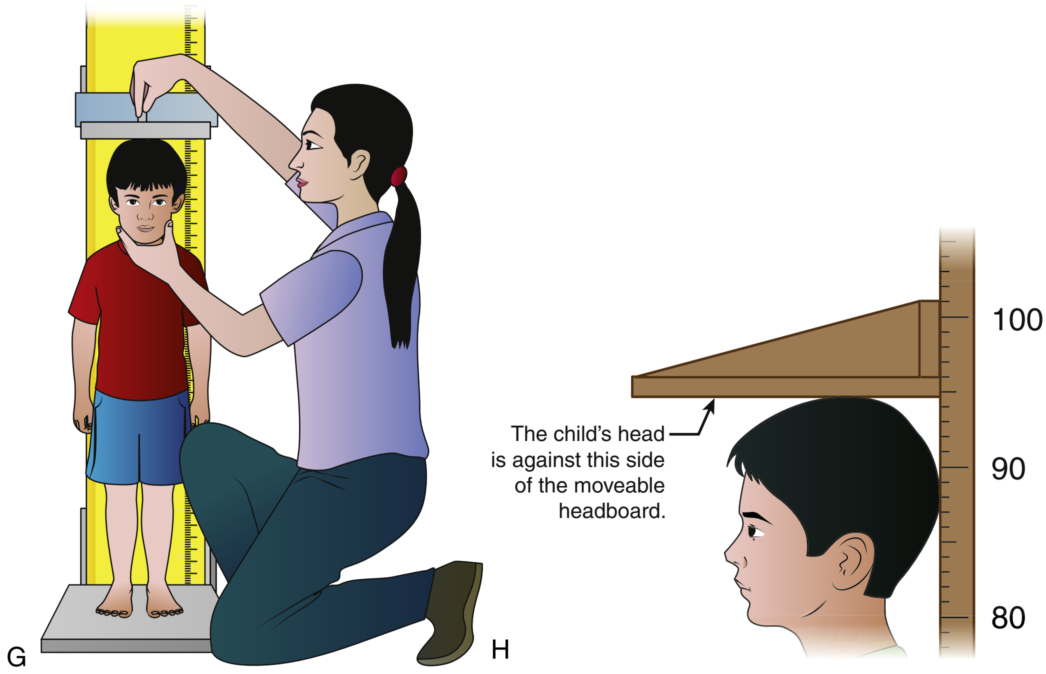

G
H
The child's head is against this side of the moveable headboard.
100
90
80

Fig. 17.1, cont'd G, Measurement of a child, standing. H, Scale for the height stadiometer.

A **Estimated gains in body composition from early to late adolescence**

	Boys	Girls	
(10-11 to 18-19 yrs)			(Literature composite, Db, TBW, Malina et al., 1988)
Fat-free mass	32.5 kg	17.3 kg	
Fat mass	3.2	7.1	
(10-12 to 18-20 yrs)			(Fels, Db, multicomponent model, Guo et al., 1997)
Fat-free mass	31.3 kg	14.0 kg	
Fat mass	3.4	7.4	
(8-11 to 16-19 yrs)			(NHANES 1999-2004, DXA, Borrud et al., 2010)
Fat-free mass	31.3 kg	14.9 kg	
Fat mass	7.2	10.3	

B **Components of body composition have growth spurts—estimated ages at peak velocity (yrs)**

	Girls (n=53)		Boys (n=60)	
	M	SD	M	SD
Stature	11.8	0.9	13.4	1.0
Weight	12.3	1.2	13.8	1.1
LM	12.1	1.0	13.7	0.9
BMC	12.5	0.9	14.0	1.0
FM	12.6	2.0	14.0	1.3

Based on DXA; LM = lean tissue mass excluding BMC
(adapted from Iuliano-Burns et al., 2001, cubic splines)

Fig. 17.2 Body Composition at Adolescence. A, Gains in body composition components from early to late adolescence. B, Growth spurts in body composition components.

growth and maturity characteristics of youth whose weight status is variably classified by different criteria, as well as potential behavioral implications of weight status *per se* and perhaps of weight status misclassifications.

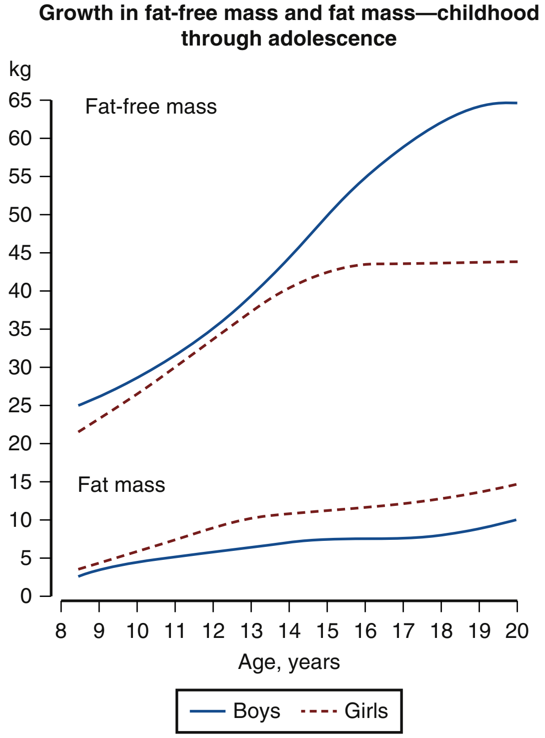
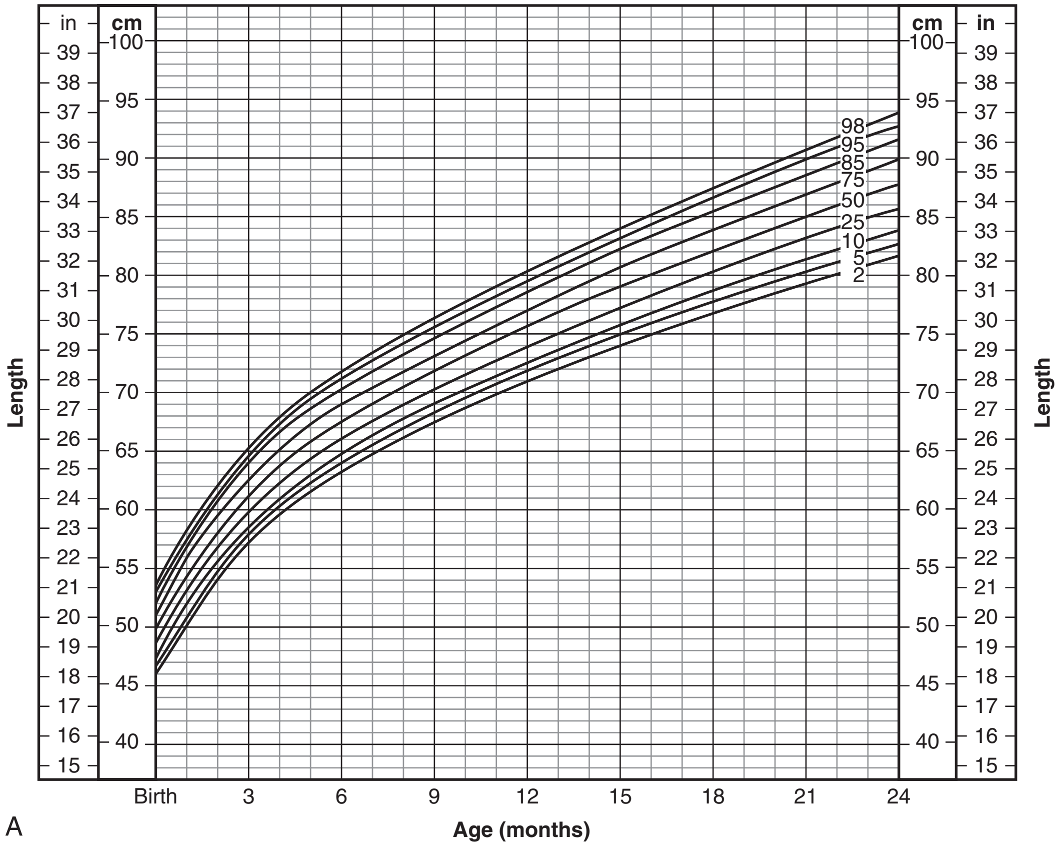


Fig. 17.3 Growth in Fat-Free and Fat Mass in Childhood Through Adolescence. (From Malina RM, Bouchard C, Beunen G. Human growth: Selected aspects of current research on well-nourished children. *Ann Rev Anthropol.* 1988;17:187–219.)

KEY POINTS

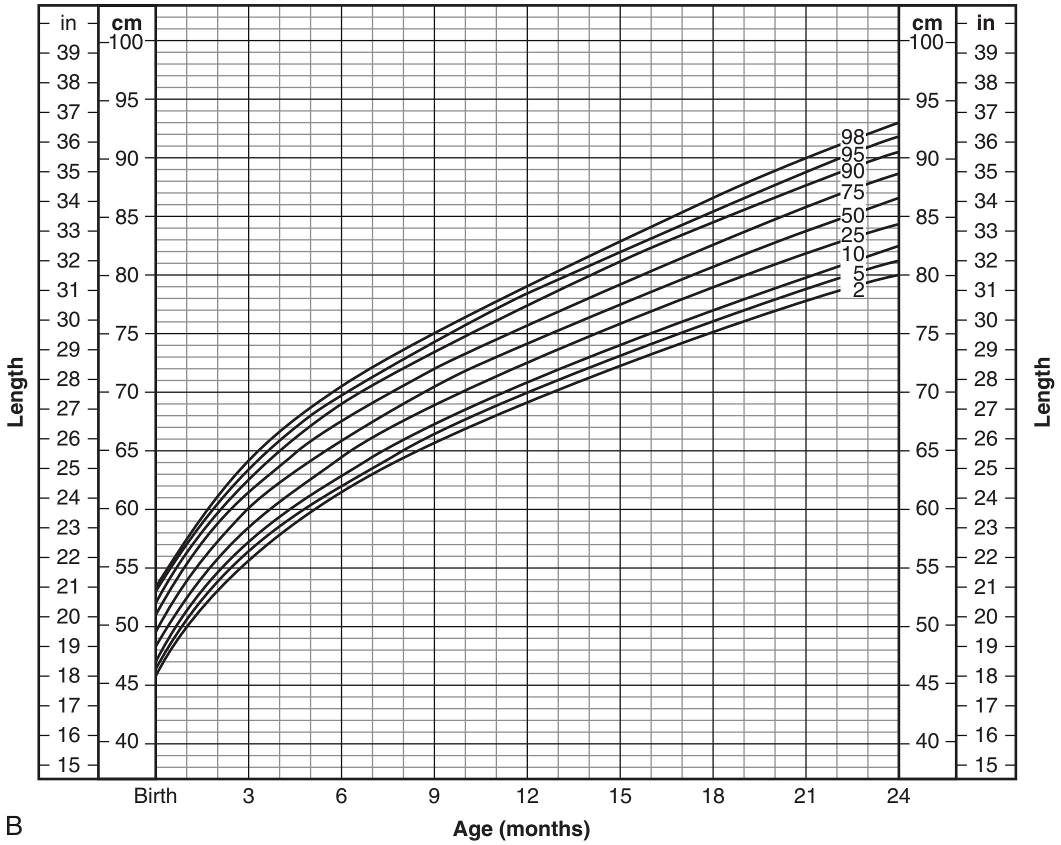
- Charts for height and weight are critical for the interpretation of linear growth and weight status. Many newer charts display the obesity status more accurately than the older Centers for Disease Control and Prevention charts.

Length-for-age Percentiles, Boys 0 to 24 Months, WHO Growth Standards



A

Length-for-age Percentiles, Girls 0 to 24 Months, WHO Growth Standards

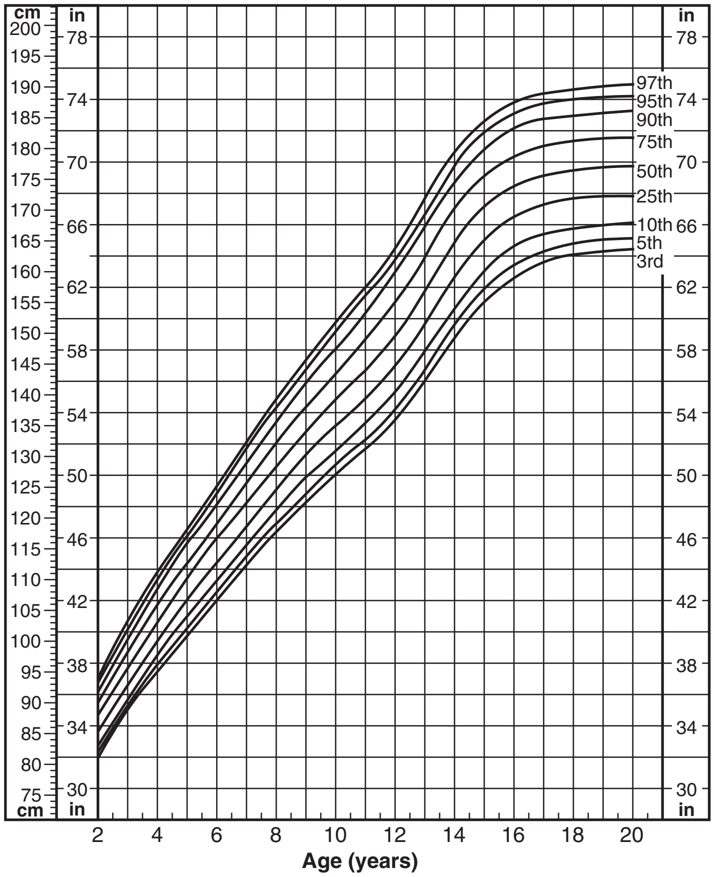


B

Fig. 17.4 Growth curves for children and adolescents. A, Length-for-age, boys, 0–24 months, World Health Organization (WHO) growth standard PP13. **B,** Length-for-age, girls, 0–24 months, WHO growth standard PP14.

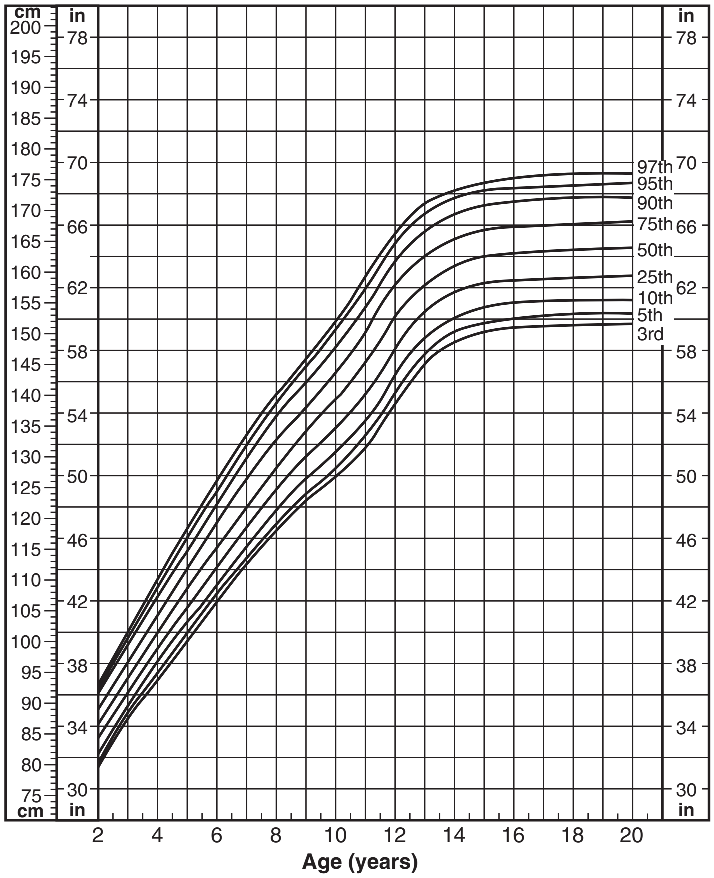
nued

Stature-for-age Percentiles, Boys, 2 to 20 years,
CDC Growth Charts: United States



C

Stature-for-age Percentiles, Girls, 2 to 20 Years,
CDC Growth Charts: United States



D

Fig. 17.4, cont'd **C,** Stature-for-age, boys, 2–20 years, US Centers for Disease Control and Prevention (CDC) growth charts, USA PP15. **D,** Stature-for-age, girls, 2–20 years, CDC growth charts, USA PP16.

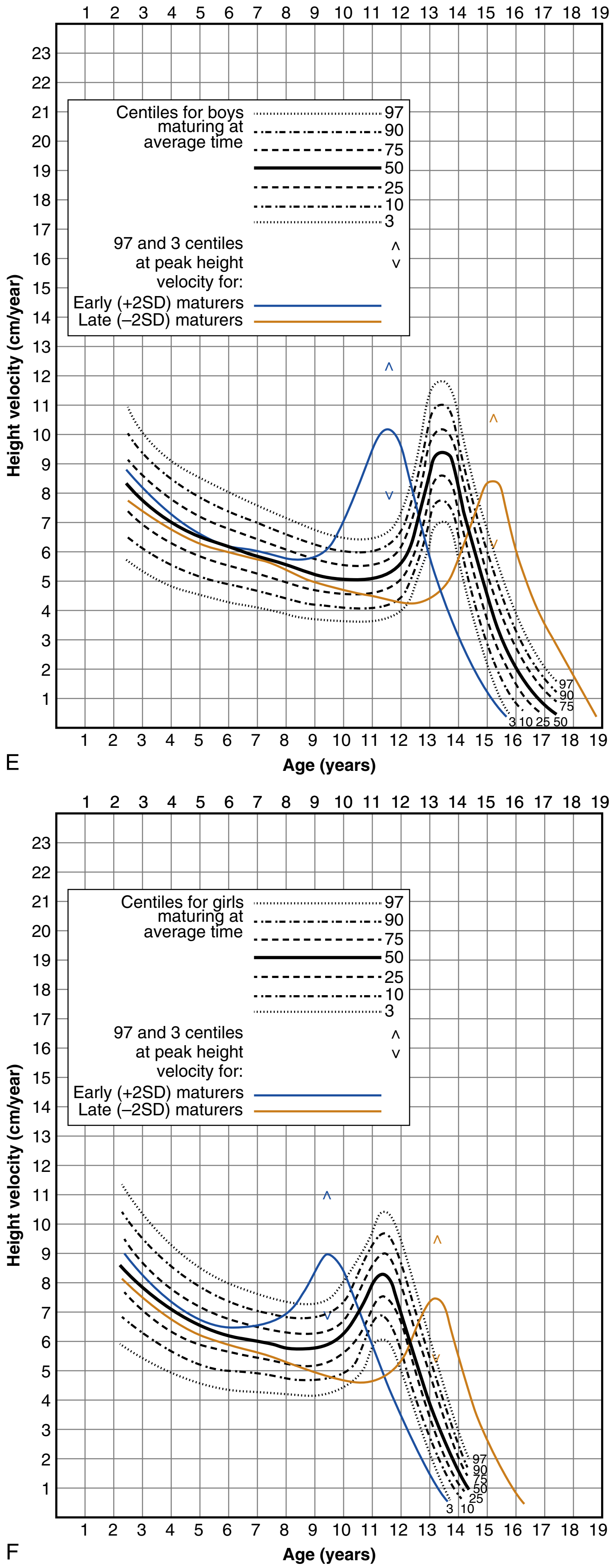


Fig. 17.4, cont'd **E,** Height velocity in American boys, CDC growth charts, USA PP18. **F,** Height velocity in American girls, CDC growth charts, USA PP19.

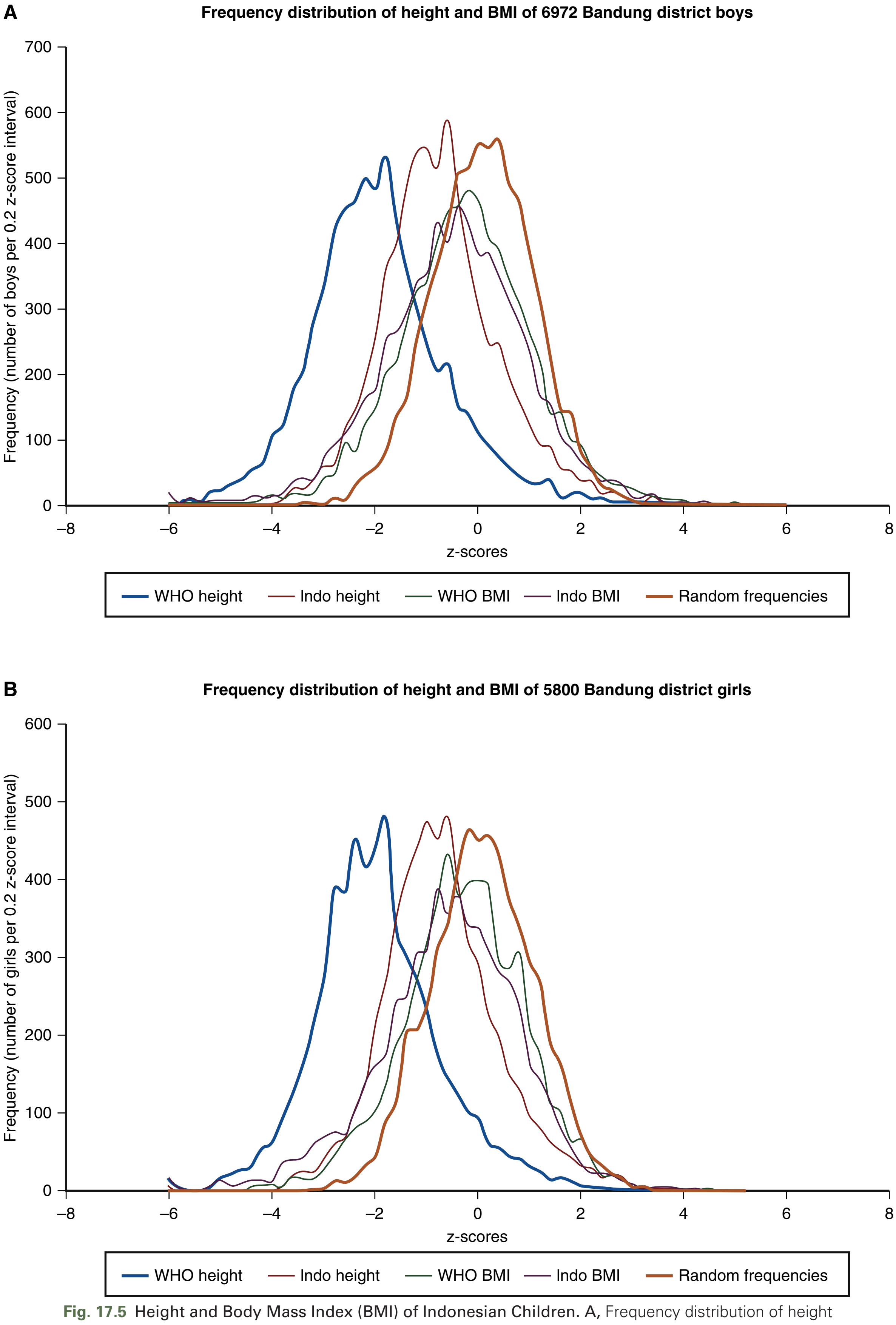


Fig. 17.5 Height and Body Mass Index (BMI) of Indonesian Children. A, Frequency distribution of height and BMI for boys. **B,** Frequency distribution of height and BMI for girls. (Novina N, Hermanussen M, Scheffler C, et al. Indonesian National Growth Reference Charts Better Reflect Height and Weight of Children in West Java, Indonesia, than WHO growth standards. *J Clin Res Pediatr Endocrinol.* 2020; 12:410–419.)

TABLE 17.2 **Comparison of the World Health Organization (WHO) Growth Standards and the US Centers for Disease Control and Prevention (CDC) Growth Reference**

Comparison	WHO Growth Chart	CDC Growth Chart
Studied population	Breastfed infants and toddlers	Breastfed and formula-fed infants and toddlers
Growth pattern	How healthy children SHOULD GROW in ideal conditions	How certain groups of children HAVE GROWN in the past
Concept of growth	A STANDARD by which all children should be compared	A REFERENCE does not imply that pattern of growth is optimal

Growth Rate

The increment in height or weight between two observations provides an estimate of growth rate or velocity or *tempo* of growth. Measurements, however, are not always taken at prescribed dates or intervals. As such, observed increments require adjustment for the actual interval between measurements.

Increments are influenced by technical errors of measurement at each observation, and also by diurnal variation (variation during the course of a day). Heights measured in the morning, shortly after arising, are generally taller than those taken later in the day. The "shrinkage" in height is attributed to the compression of the intervertebral discs associated with gravity and physical activity. Increments in height also show seasonal variation in some parts of the world. Evidence from several studies in North America and Europe indicates a greater childhood growth rate in height during the spring compared with the fall.[31] Allowing for measurement variability *per se* and diurnal and seasonal variation, estimates of growth rates over short durations, e.g., 3 or 6 months, must be interpreted with care.

Growth rates (cm/year, kg/year) are presented as velocity curves, which differ in shape from the curves of size attained at different ages described above (Fig. 17.2). The velocity curve for height (length among infants) indicates a brief acceleration in rate of growth after birth, which is followed by a constantly decelerating rate growth in stature during infancy and childhood, i.e., the child is getting taller, but at a constantly slower rate. The growth rate reaches its lowest point ("prepubertal dip") just prior to the initiation of the adolescent spurt. The lowest point is labeled as the age at takeoff of the adolescent spurt, because this is the point at which the velocity curve begins to accelerate. In contrast, after deceleration during infancy and the second year, rate of growth in weight occurs at a slight but constantly accelerating rate until the adolescent spurt (Fig. 17.4).

Rates of growth in both stature and weight accelerate during the adolescent spurt. Acceleration in height occurs prior to that in weight, and maximal velocity of growth in height (peak height velocity [PHV]) occurs earlier than that for weight (peak weight velocity). After peak velocity of growth in height is attained, the velocity of growth in height gradually declines, and growth in height eventually ceases in late adolescence. Estimated PHV occurs, on average, approximately 2 years earlier in girls than in boys. Girls stop growing in stature by approximately 16 years, on the average, whereas boys continue to grow for another 2 or 3 years. Velocity of growth in weight also declines after the peak, but often remains positive into the early twenties.

Reference values for annual or semiannual height increments are available for children and youth in the Fels Longitudinal Study from birth to 18 years,[32,33] and have been more recently reported for samples of Black and White American children and youth 6 to 18 years of age.[34] Reference values for short-term increments in heights of British children 7 to 10 years of age are also reported.[35]

In addition to the well-defined adolescent spurt, some children show a small growth spurt in stature and weight (increase in velocity of growth) several years before the onset of the adolescent growth spurt. The midgrowth spurt usually occurs between 6.5 and 8.5 years of age, occurs more frequently in boys than in girls, and does not appear to show a sex difference in timing.[36-38] The midgrowth spurt reflects biological variation among individuals, but the frequency of measurements in childhood is an additional factor. Children are usually measured annually during childhood "well-child" visits, and such an interval may not be sufficiently sensitive to detect the change in velocity of growth that defines the midgrowth spurt.

Maturity Status

Maturity status refers to the level or state of maturation at the time of observation. Indicators of skeletal and pubertal maturation are used most often.

Skeletal Maturity Status

Skeletal maturation is generally estimated as skeletal age (SA) based on the bones of the hand and wrist viewed on a standard radiograph. Each bone goes through a series of changes from initial ossification to adult morphology. The changes provide the basis for assessing skeletal maturation based on the assumption that specific features of each bone as observed on a radiograph occur regularly and in an irreversible order, and as such provide a record of the progress of each bone towards maturity. Other parts of the skeleton, e.g., knee and foot and ankle, as well as the hemiskeleton as an "integrated" determination by the time of appearance of various centers of ossification, have also been used to derive estimates of SA.[39]

Methods of Assessment. Three methods are commonly used to estimate SA of the hand-wrist. Each method calls for the hand-wrist radiograph of a child to be compared to specific criteria; ratings are subsequently converted to an SA *that is specific to the method*. Indicators of maturity defined for specific bones in each method suggest discrete steps in a continuous process.[1,40]

The Greulich–Pyle (GP) method[41] is an extension of the method initially described by Todd.[42] The method was developed on upper–socioeconomic status White American children from Cleveland, Ohio. Accordingly, each individual bone of the hand-wrist is rated relative to sex-specific standard plates representing specific SAs from infancy through adolescence. The method often requires interpolation between the standard plates. An SA is assigned to each bone, and the median of the SAs is the estimate of SA for the child. In practice, however, the GP method is often applied by comparing the radiograph as a whole to the pictorial standards and assigning the SA of the standard to which the radiograph most closely matches. As such, variation in level of maturity among individual bones may be large and is often overlooked. By convention, the left hand and wrist are used.

The Tanner–Whitehouse (TW) method[43] was developed on British children in the Harpenden Study. Specific criteria or stages spanning initial appearance (ossification) to maturity were described for each of 20 bones in the hand-wrist: the radius, ulna, and metacarpals and phalanges of the first, third, and fifth digits (long bones) and the carpals, except the pisiform. A maturity score was also assigned to each stage for each of the 20 bones. The scores assigned to each of the 20 bones are summed, with the seven carpals and 13 long bones each contributing

50% to the skeletal maturity score; the score for the 20 bones is then converted to an SA using sex-specific tables.

The first revision of the method, TW2,[44,45] did not modify the criteria for maturity indicators; however, the final stages for five carpals and the radius and ulna were eliminated. In addition to the original SA based on 20 bones (TW2 20 Bone SA), two other SAs were provided, an SA based on the seven carpals (TW2 Carpal SA) and an SA based on the radius, ulna, and short bones (TW2 RUS SA). The maturity scores assigned to the stages for each bone are sex-specific and vary for the 20-bone, carpal, and RUS protocols. The sum of the maturity scores for 20 bones, the seven carpals, and the 13 long bones is converted to an SA using sex-specific tables.

The most recent version, TW3,[46] retained the RUS (TW3 RUS) and carpal (TW3 Carpal) SAs but eliminated the 20 Bone SA. The tables for converting carpal maturity scores were not modified; the reference for TW3 Carpal SA was the original British series. In contrast, the tables for converting the RUS maturity scores to SAs were modified. Reference values for TW3 RUS SA were based on a composite of the original British series and samples of Belgian (Flemish), Italian, Spanish, Argentine, Japanese, and American children and adolescents surveyed in the late 1960s through the mid-1990s; the American sample was from an upper–socioeconomic status area in the Houston region (Texas). Ages at attaining skeletal maturity with the TW3 RUS protocol were also lowered from 16.0 to 15.0 years in girls and from 18.2 to 16.5 years in boys.

The Fels method[47] was based on participants in the Fels Longitudinal Growth Study of children from middle class families in south-central Ohio. The method specifies criteria for the radius, ulna, carpals, and metacarpals and phalanges of the first, third, and fifth rays. Grades are assigned to each bone depending on age and sex. Ratios of linear measurements of the widths of the epiphysis and metaphysis of the long bones are also used, and the presence (ossification) or absence of the pisiform and adductor sesamoid is noted. Grades assigned to the individual bones and width measurements are entered into a program that calculates SA and its standard error; the latter provides an estimate of the error of the assigned SA, which is not available with the GP and TW methods.

Skeletal Age. The SA assigned to the radiograph of an individual represents the CA at which a specific level of maturity of the hand-wrist bones was attained by the *reference sample* upon which the method of assessment was developed. An individual who has attained skeletal maturity is simply noted as mature. Of relevance, the individual is skeletally mature at the time of observation; when he/she attained the skeletally mature state is not known. As such, *an SA is not assigned when the hand-wrist radiograph indicates skeletal maturity*.

SA is ordinarily expressed relative to CA. The difference between SA and CA (SA minus CA) is often used as to indicate maturity status, i.e., early or advanced, average or "on time," and late or delayed.

SAs derived with the different methods, though related, are not equivalent, as criteria, methods, and reference samples differ. The GP, TW 20 Bone, and Fels methods each include the carpals, while the TW2 and TW3 RUS protocols are limited to the long bones (radius, ulna, metacarpals, and phalanges). SAs based on the GP and Fels methods, and three versions of the TW method (TW2 20 Bone, TW2 RUS, TW3 RUS) in a sample of German boys 6 to 16 years of age[48,49] are summarized in Table 17.2. Heights of the boys matched, on average, medians of current US reference data. The original study used the GP method, but the radiographs were made available and were assessed with the GP (bone by bone), Fels, TW 20 Bone and TW2 RUS methods.[50] The TW2 RUS scores were subsequently converted to TW3 SAs (Malina, unpublished).

Standard deviations for SAs were three to four times larger than those for CA, highlighting variation in skeletal maturity status within a *single-year* CA group (Table 17.3). Mean SAs with each method varied and overlapped within each CA group, except for consistently lower SAs with the most recent TW revision. Beginning at 9 to 10 years, TW3 RUS SAs were consistently lower than TW2 RUS SAs. This same trend was noted in a large series of youth soccer players.[51] The trends in comparisons among methods were generally consistent with observations of GP, Fels, and TW2 SAs in a clinically normal sample of 23 boys followed from 8 to 15 years.[52]

As noted, TW3 RUS SAs are systematically lower than TW2 RUS SAs beginning at about 10 years of age. The rationale for assigning lower SAs for the same RUS score with TW3 compared with TW2 was to accommodate secular change.[46] Although secular changes in height are evident in early childhood and continue through puberty,[1,53-54] modifications in SAs assigned to the same RUS maturity scores in boys (i.e., lower SA for the same maturity score with TW3) were only apparent beginning with SAs of about 10 years. Of relevance, secular increases in height were not necessarily accompanied by accelerated maturation between 1960 and 1980 in Belgium[55,56] and between 1980 and 1997 in the Netherlands.[57]

Although not indicated in Table 17.3, a number of boys were skeletally mature, especially with the TW method (one each at 14 and 15 years, and five at 16 years) compared with the GP (two at 16 years) and Fels (one at 16 years). Numbers of skeletally mature boys were larger at 17 years (GP 9, Fels 11, TW 17). The discrepancy between the GP and Fels methods and the TW method regarding skeletal maturity of the hand-wrist relates to the criterion for the final stage of the radius and ulna, respectively. With the TW method, the final stage for the radius and ulna is simply "fusion of the epiphysis and metaphysis has begun."[46] With this criterion, the interval between onset and completion of epiphyseal union of both bones is not considered. Many youth are thus classified as skeletally mature even though the process of fusion in each bone is still in progress. On the other hand, onset through *complete* fusion of the distal radius and ulna are used in the GP and Fels methods.

SA, or more colloquially "bone age," has been indicated for CA verification in medicolegal contexts for many years, and also in some youth sport competitions. Evidence from a review focused on male soccer players and female artistic gymnasts[58] indicates the following. Among adolescent male soccer players (and likely male athletes in other sports), a significant number will be identified as older than a CA cutoff due to advanced skeletal maturity status when they in fact have a valid CA. SA assessments of soccer players were generally comparable to magnetic resonance imaging (MRI) assessments of epiphyseal-diaphyseal union of the distal radius, which have been recommended for age verification for international competitions among U-17 soccer players.[59,60] Both protocols indicated a relatively large number of false negatives among youth players aged 15 to 17 years. Among adolescent female artistic gymnasts, a significant number of age-eligible gymnasts would be identified as younger than the CA cutoff due to later skeletal maturation, when in fact they have a valid CA. On the other hand, there is also the possibility of false positives—identifying athletes as younger than the CA cutoff due late skeletal maturation, when in fact their birth certificates or passports indicate a CA older than the cutoff.

Given the available data on the potential for false negatives and false positives, SA is therefore not a valid indicator of CA. The situation is more complicated for CA verification in the context of immigration regulations. Given the variation in SA within single-year CA groups, and also variation in SAs based on currently available methods, use of SAs as an indicator of CA for the purpose of immigration has major limitations.

TABLE 17.3 Skeletal Ages With Five Different Methods of Assessment in Boys 8–16 Years of Age

			SKELETAL AGES, YEARS									
	CA, yrs		GP		Fels		TW2 20 Bone		TW2 RUS		TW3 RUS	
N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
26	8.4	0.3	8.3	0.9	8.1	0.9	8.3	0.9	8.0	1.0	8.0	0.9
23	9.5	0.3	10.1	1.0	9.6	1.0	9.8	0.9	9.8	1.2	9.4	0.9
22	10.5	0.3	10.2	1.0	9.7	1.0	10.1	1.2	9.9	1.1	9.5	0.8
20	11.5	0.2	11.0	0.8	10.7	1.2	11.2	0.9	11.3	1.2	10.5	0.9
31	12.4	0.3	12.1	1.0	12.2	1.5	12.6	1.5	12.6	1.6	11.6	1.3
22	13.5	0.3	12.8	1.0	13.0	1.4	13.5	1.4	13.5	1.6	12.3	1.5
23	14.3	0.3	13.8	1.0	14.2	1.1	14.8	1.1	14.9	1.3	13.8	1.0
20	15.4	0.3	14.9	0.8	15.4	0.9	15.8	0.8	15.9	0.9	14.9	1.0
10	16.5	0.3	15.8	0.8	16.5	0.8	16.8	0.8	17.0	0.8	16.0	0.8

CA, Chronological age; *GP*, Greulich–Pyle; *TW*, Tanner–Whitehouse; *RUS*, radius, ulna, and short bones; *N*, number of subjects; *SD*, standard deviation; *M*, mean.

Other Protocols. Given advances in technology, several new protocols for the assessment of the skeletal maturity of the bones of the hand and wrist have been proposed. The procedures are generally based upon the GP and TW methods and are largely designed for clinical use. These include ultrasound assessment of the maturity status of the distal radius and ulna, with SA scaled relative to the GP method,[61,62] although its validity has been questioned.[63] Assessment of SA based on dual-energy x-ray absorptiometry scans of the hand-wrist have also been proposed.[64-66] Automated methods are also available.[67-69] The BoneXpert method[68] is unique in that it derives an "intrinsic" bone age based on bone borders (shapes) and wavelet texture of images of 15 bones: radius, ulna, and the five metacarpals and eight phalanges in the first, third, and fifth fingers. The "intrinsic" bone ages are subsequently calibrated to GP and TW RUS SAs.

Overview of Skeletal Age. SA can be used across approximately the first two decades postnatally; in contrast, other maturity assessments are limited to puberty and adolescence (see later). Estimates of SA by each method are reasonably precise and reliable, although intra- and interobserver variability in assessments is rarely reported. Use of SA is also criticized, because specific training is required to learn the protocol(s): this is a shallow criticism, as other protocols also need specific training. Major limitations of SA are the expense associated with the radiographs *per se* and radiation exposure, and also the limited availability of individuals familiar with the details of the different methods of assessment. With modern technology, exposure to radiation presents minimal risk at 0.001 millisievert, which is less than natural background radiation and radiation exposure associated with the equivalent of viewing 3 hours of television per day.[70,71]

The different methods for the assessment of SA are based largely on samples of European ancestry, although the TW3 RUS protocol included data from samples of Argentine and Japanese ancestry in converting maturity scores to SAs (see earlier). Nevertheless, applications of the GP and TW protocols show ethnic variation in SA.[72-78] Applications of the Fels method to youth of different ethnic groups are apparently not available, except for a study of indigenous school children in Oaxaca, southern Mexico.[79] Of potential relevance, identification of the ethnicity of youth in some countries is not permitted.

Secondary Sex Characteristics

Secondary sex characteristics in males include pubic hair (PH), genitalia (G; penis, scrotum, testes), testicular volume, voice change, and facial and axillary hair, while those in females include PH, breasts (B), axillary hair, and menarche.[1,80] Facial hair and voice change in boys and axillary hair in both sexes are also secondary sex characteristics, but they generally develop late during puberty and are not widely used.

Pubertal Stages. The five stages of PH, G, and B described by Tanner[81] are commonly the reference in assessing pubertal status (Fig. 17.6). The stages generally follow the criteria of earlier studies of Reynolds and Wines.[82,83] Stages are labeled PH1 through PH5, B1 through B5, and G1 through G5. Stage 1 of each characteristic indicates the prepubertal state, an absence of overt development, although hormonal changes that trigger puberty may already be underway. Stage 2 marks the overt development of each characteristic; B2 and G2 are typically the first overt signs of the transition into puberty, but PH2 may precede B2 and G2 in some youth. Stages 3 and 4 mark progress in pubertal maturation, and the respective stages are sometimes labeled as mid- and late-puberty. Stage 5 indicates the mature state (Fig. 17.6).

The stages are discrete categories unique to PH, B, and G. A youngster is either in a stage or not in a stage at the time of assessment; there are no intermediate stages. Moreover, stage at time of assessment provides no information on when the youngster entered the stage (timing) or how long he/she has been in the stage, related to the *tempo* of maturity.

It is important to emphasize that the stages of PH and B are unique to girls, and the stages of PH and G are unique to boys. *The stages are not equivalent*, i.e., B3 ≠ PH3, G3 ≠ PH3, B3 ≠ G3, PH3 in girls ≠ PH3 in boys, and so on. The term "Tanner stages," unfortunately, is often used without indicating the specific characteristic(s) that was (were) assessed.

Direct assessments of the stage of pubertal status are made at clinical examination. Self-assessments are often used in nonclinical settings; they require privacy, good-quality photographs of the stages, simplified descriptions, and a mirror to assist in the process. Some self-assessment scales include pictures or drawings of the stages, as well as questions regarding facial and axillary hair in males and axillary hair and menarcheal status in girls.[1]

There is a need for quality control, including intra- and interobserver reliability in assessment of stages, and concordance between self-assessments and those of experienced assessors. Overall reproducibility by experienced assessors is generally good, with approximately 80% agreement in assigning stages, but lower percentages have been reported.[1]

Accuracy of self-assessments is a concern, but opinions vary depending upon the purpose of the study. Based on self-assessments of pubertal status in three annual visits of girls between 11 and 14 years of age and assessments by trained examiners, it was concluded that "… self-assessment can substitute for examiner evaluation only when crude estimates of maturation are needed."[84] On the other hand, agreement to within one stage was suggested as potentially useful in epidemiological surveys of youth,[85] even though concordance between self- and physician assessments indicated limited accuracy. Concordance between and among self-assessment scales currently in use needs further evaluation.

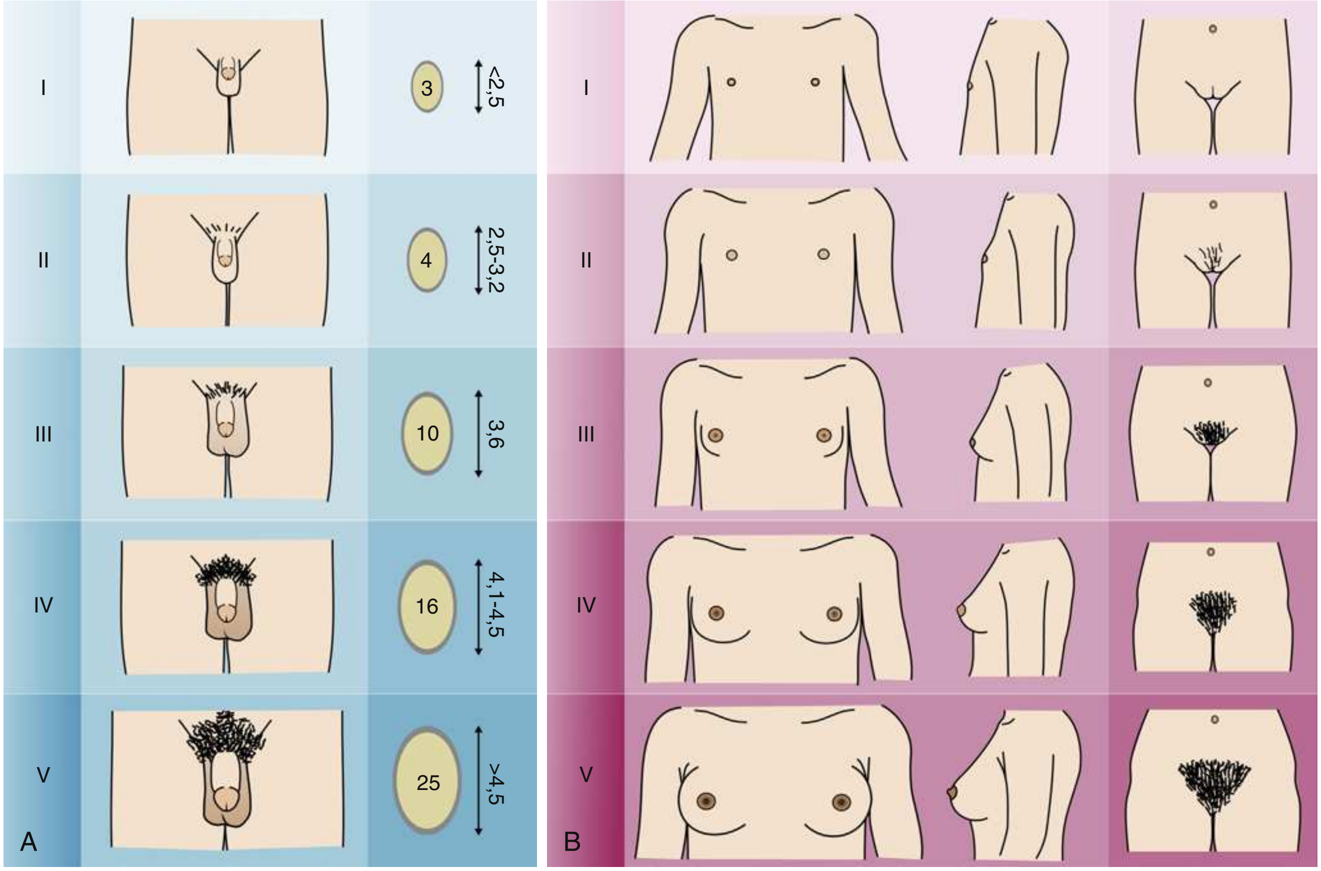


Fig. 17.6 Genital and Pubic Hair Stages at Puberty in Boys and Breast and Pubic Hair Stages in Girls at Puberty. (From Tenbergen G, Wittfoth M, Frieling H, et al. The neurobiology and psychology of pedophilia: recent advances and challenges. *Courtesy: Michal Komorniczak (Poland).*

Assessments of pubertal status are ordinarily done in a clinical context. However, in some cases, researchers are interested in whether a youngster is pre- or postpubertal, i.e., stages 1 and 5, respectively. Others may be interested in early-, mid-, or late- puberty, i.e., stages 2, 3, and 4, respectively. Note, however, as mentioned above, that stages are specific to B and PH in girls and G and PH in boys, and are not equivalent, i.e., B2 ≠ PH2, G3 ≠P H3, etc.

Analytical Concerns. Stages of PH, B, and G are variably reported. Ratings for individuals are periodically combined into a mean of B and PH or of G and PH; there is, however, no biological entity of a "mean stage." The stages are not equivalent and must be considered separately.

Stages are occasionally reported as 3+ or 4+. Note, a youngster is either in a stage or not in a stage; there are *no* intermediate stages. Although mean stages of PH, B, or G by CA at observation are often reported, distributions of stages within each CA group are more informative and relevant.

Youth are also commonly grouped by stage of puberty independent of CA. This is problematic from at least two perspectives. First, stages of puberty vary within a CA group, and second, also vary by CA within a stage. For example, within single-year CA groups of soccer players 11 to 14 years of age, boys in less advanced stages of PH tend to be younger, shorter, and lighter, on average, than players in more advanced stages who are older, taller, and heavier. And, among players grouped by stage of PH, younger boys tend to be, on average, shorter and lighter than older boys who are taller and heavier.[86] Corresponding classifications of girls would likely yield similar results.

Testicular Volume. Testicular volume provides a more direct estimate of genital maturity in boys. The method requires palpation of the testes in order to match their size with a series of models of known volume (Prader orchidometer).[87,88] The ellipsoid models have the shape of the testes and range from 1 to 25 mL; a volume above 4 mL marks the beginning of puberty. The method is used primarily in the clinical setting. Sonography can also be used to estimate testicular volume.[89] In a comparison study, Joustra and colleagues noted that testicular volumes above 5 mL, when evaluated by palpation, are reasonably accurate (compared with the ultrasound method), but may differ greatly below that, when it may be most relevant.[90] For clinical purposes, that may not be so important, just that the testes are prepubertal, but may take on greater importance if gonadotropin treatment is evaluated at minipuberty during infancy.

Menarcheal Status. Although age at menarche is an indicator of maturity timing, whether a girl is premenarcheal or postmenarcheal is an indicator of maturity status. Classifications of girls by menarcheal status are confounded by CA *per se*, i.e., an 11-year-old premenarcheal girl is quite different physically and behaviorally from a 14-year-old premenarcheal girl.

Overview of Secondary Sex Characteristics. Secondary sex characteristics (overt manifestations) are limited to the interval of pubertal maturation. Stages are discrete, but somewhat arbitrary. Direct assessment is often considered invasive, especially outside the clinical setting. Cultural sanctions may limit or prohibit assessment of secondary sex characteristics in some groups. Concordance of clinical and self-assessments is variable and needs further study. Stages of puberty are also variably reported and present analytical concerns.

Maturity Timing

Maturity timing refers to the CA at which specific maturational events occur. The two most commonly used indicators of timing are age at PHV and age at menarche. Both are limited to the adolescent period.

Age at Peak Height Velocity. Age at PHV is the estimated CA at the maximum rate of growth in height during the adolescent spurt. Age at PHV is estimated from serial height measurements of individuals taken annually or semiannually from late childhood through adolescence. Historically, determinations from individual height records were graphically plotted to identify takeoff, peak, and eventual cessation of growth. Mathematical modeling or fitting of individual height records is currently used, and a variety of methods are available.[91] Estimated ages at PHV vary somewhat among methods, but are generally more uniform than estimated peak velocity of growth in height (cm/yr).

Mean ages at PHV are reasonably similar among longitudinal samples of European and North American youth.[1,92] On the other hand, variation in ages at PHV *among individuals* is considerable. In longitudinal samples of British, Swiss, Polish, Belgian, Canadian, and American youth, estimated ages at PHV ranged from 9.0 to 15.0 years in individual girls and 11.1 to 17.3 years in individual boys.[1,93-96]

Age at Menarche. Menarche refers to the first menstrual flow; age at menarche is an indicator of the timing of this pubertal event. In longitudinal studies, girls and/or their mothers are interviewed at each regularly scheduled visit/observation regarding whether or not menarche has occurred. If menarche occurred between visits, further questions pinpoint the specific date/age of the first menstrual flow occurred. This is labeled the prospective method. Prospectively recorded ages at menarche in two longitudinal studies, one of American[94] and one of Polish[96] girls, ranged from 10.77 to 15.25 years and 10.49 to 16.30 years, respectively. Longitudinal studies generally follow subjects across adolescence so that early- and late-maturing girls are included. Depending on ages at which short-term longitudinal studies start and conclude, there is potential risk that early- and late-maturing girls may be excluded.

Ages at menarche based on the prospective method are sometimes confused with estimates based on the *status quo* method. The *status quo* method requires two pieces of information in a cross-sectional sample spanning 9 through 17 years: CA and whether or not menarche has occurred (yes/no). The data are subsequently analyzed with probits or logits to derive a median age at menarche and associated variance statistics for the sample. The *status quo* method is generally used in surveys.

In contrast to the prospective and *status quo* methods, ages at menarche are commonly obtained with the retrospective method, with late adolescents and/or adults asked to recall when they experienced their first menstruation. The method relies on memory, i.e., recall of the age when first menstrual flow occurred. In addition to potential errors with memory *per se*, reported ages are influenced by recall bias (the shorter the recall interval, the more accurate the recall, and *vice versa*) and a tendency to report whole years, typically age at the birthday before menarche.[1]

Estimates of age at menarche based on the retrospective method with samples of young adolescents are biased. Girls who have not yet attained menarche are excluded from the estimates. Some late-maturing girls may not attain menarche until 15 or 16 years, or perhaps later. In a nationally representative sample of US girls, 90% attained menarche by 13.75 years,[97] but 10% of girls attain menarche after this age.

Other Indicators of Timing and Interrelationships. Assuming longitudinal data are available, other potential maturity indicators can be estimated, e.g., age at takeoff of the growth spurt in height, ages at peak velocity for other body dimensions, and ages at attaining specific SAs, stages of pubertal maturation or specific percentages of adult height (see later).

Analyses of ages at attaining several different maturity indicators in two longitudinal series highlight interrelationships among indicators during adolescence.[98-100] Common indicators in the two longitudinal series included ages at PHV and menarche, ages at attaining stages of pubertal development, and ages at attaining specific SAs and percentages of adult height. The analyses indicated a general maturity factor in both sexes underlying the timing of maturity indicators during the interval of the adolescent spurt. The analyses for boys suggested a second factor that loaded on ages at attaining SAs of 11 and 12 years, 80% of adult height, and early stages of PH and genital development, which are characteristic of early puberty. The results also suggested a degree of independence of ages at attaining several maturity markers characteristic of the transition into puberty, i.e., late prepuberty or early puberty.[100] Longitudinal data for 30 boys also indicated considerable variation in SA at the time of pubertal onset (serum testosterone ≥30 ng/dL).[101]

Tempo of Maturation

Tempo refers to the rate at which maturation progresses. Evidence from the Zurich Longitudinal Study indicated the following trends. Intervals (means ± standard deviations) between B2 and B3 and between PH2 and PH3 in Swiss girls were, respectively, 1.4 ± 0.8 years and 1.8 ± 1.0 years, while intervals between G2 and G3 and between PH2 and PH3 in Swiss boys were, respectively, 1.7 ± 1.0 years and 1.3 ± 0.9 years.[102,103] The intervals between the transition into puberty (B2, G2, PH2) and the mature state (B5, G5, PH5) were, on average, 2.2 ± 1.1 years for B and 2.7 ± 1.1 years for PH development in girls, and 3.5 ± 1.1 years for genital and 2.7 ± 1.0 years for PH development in boys. The standard deviations for the transition through puberty for each characteristic approximated 1 year and highlight the variation in *tempo* of maturation of secondary sex characteristics within and among individuals.

Although estimated increments in GP SAs in a longitudinal sample of American children approximated 1 year, variation was considerable and was associated in part with maturity status, i.e., early versus late.[104,105] In a mixed longitudinal sample of White and Black American girls aged 6 to 12 years, mean single-year velocities for TW2 20 Bone SAs varied between 0.66 and 1.14 years/year, and standard deviations varied between 0.33 to 0.52; corresponding mean single-year velocities for boys varied between 0.75 and 1.27 years/year, and standard deviations ranged from 0.32 to 0.60.[73] Single-year rates of skeletal maturation expressed as maturity points per year of the American children[72] overlapped mean rates and ranges for British children.[106] Observations in a longitudinal series of 34 boys suggested that annual increments (years/year) in TW2 SAs (presumably 20 Bone) increased during the interval of puberty and the growth spurt, and that the increments appeared to reach a peak near PHV.[105] Allowing for limited data, it is legitimate to inquire whether a skeletal year equals a chronological year.

Evaluation of Children for Disorders of Growth

One of the most common reasons for referral to a pediatric endocrine clinic is to evaluate an infant/child/adolescent for short stature for his/her CA relative to a reference for the general population. In contrast, relatively few children are referred for tall stature, although the paradigm for evaluation is, in general, similar to that for short children.

Although the most common causes of short stature for calendar age are variations on the theme of normal growth, there are a number of conditions for which evaluation will yield a proper diagnosis and suggest an appropriate therapy; the latter, however, is often a matter of reassurance. Inextricably entwined in evaluations of short stature are nutritional status and the psychosocial concomitants of being short or maturing late. Nevertheless, evaluation of short stature must be viewed in the context of expected age and gender interactions that influence developmental, behavioral, metabolic, biochemical, and hormonal factors, in addition to genetic and environmental factors, including socioeconomic status. Interactions among the preceding are often brought into sharper focus at the time of puberty, especially when peers have begun pubertal maturation. Socialization during puberty, including sports participation, often tracks with biological maturation to a larger degree than CA.

Evaluation of children for short stature and/or delayed maturation begins somewhat differently for boys and girls, although the differential diagnosis may extend beyond the variation in normal growth and maturation. Perceptions and observations of children and parents suggest that the psychosocial aspects of being short affect boys to a greater extent than girls; as such, boys are more often presented to a pediatric endocrinologist for earlier evaluation than girls.

Normal linear growth in stature and weight is generally accepted as a sign of good health. The process of growth is continuous, but not linear, and shows a wide range of variation at different stages of a child's life. Three distinct periods have been formally noted: *infantile,* when the rapid rate of *in utero* growth in length declines during the first year of life; *childhood,* when growth rate is largely steady but declines somewhat just before the pubertal growth spurt begins; and *pubertal,* as the growth spurt occurs.[107,108] Estimated rates of growth in length/height decline from more than 25 cm/year shortly after birth to approximately 12 cm/year at the end of the first year and to approximately 7 cm/year at the end of the second year. Subsequently, growth rate in height varies between approximately 5 and 7 cm/year and declines, on average, slightly until the initiation of the pubertal growth spurt (age at takeoff), when the rate of growth in height accelerates and reaches approximately 9 cm/year in girls and 10.5 cm/year in boys at PHV. After PHV, growth rate declines until growth in height terminates in the late teens or early twenties. Individual children will often cross major percentiles on the growth chart during the first 24 months as they are no longer bound by the constraints of the intrauterine environment and move toward their genetic potential. Of interest, the correlation between birth length and adult height is approximately 0.1, while that between length (height) at 2 years and adult height is approximately 0.7.

Growth disorders may be divided into at least three categories: primary or intrinsic to the growth plate; secondary or due to changes in the milieu of the growth plate, as in hormonal deficiencies; and idiopathic. Most systemic diseases or their treatments will slow linear growth (at least transiently) such that velocity of growth in height is sufficiently reduced that height begins to fall from a previously defined percentile. Allowing for normal variability, the pediatrician becomes concerned when two major percentile lines on the standard growth chart for height (size attained) are crossed (Fig. 17.4). The preceding is more apparent using height velocity charts with due care in estimating growth velocities from sequential measurements (e.g., measurement variability *per se*, adjustment for the interval between observations) (Fig 17.4). The hallmark of this transient slowing of growth is the phenomenon of catch-up growth, i.e., growth in height at a higher than normal rate that allows a child to move upward on the growth chart, usually to the percentile along which he/she was previously growing, after the growth inhibiting factor(s) has (have) been removed.

Full evaluation includes noting chronological and biological age, measurements (and charting) of height (length) and weight, estimates of height velocity, as well as parental heights, personal and family history, physical examination, state of pubertal maturation, and review of different systems. Subsequent laboratory and imaging studies are based on these initial observations.

Growth Charts

Multiple growth charts for height, height velocity, and weight are available. The WHO charts are recommended for international use, while other charts are based on specific countries, regions or ethnic groups (Fig. 17.4). Growth charts can be integrated into the electronic health record and automatically plotted to permit immediate comparison with earlier observations. A secondary benefit of immediate comparison is a reduced risk of inaccurate measurements, as the child can be remeasured if necessary. Nevertheless, appropriately calibrated instruments and care in the measurement process are essential. In addition, a child's record can be evaluated by a growth "algorithm," which can also take into consideration deflection from the midparental target height and perhaps accelerate referral to specialists for children who may require such attention.

There are at least two types of charts: *reference charts* that provide a snapshot of how children are growing relative to the population, and *standards* that provide an estimate of how children *should* grow given optimal nutrition and care. These were noted in more detail earlier.

KEY POINTS

- Maturity status may be determined by physical examination of the secondary sex characteristics, often using the Tanner criteria or by assessing skeletal (bone) age by one of several standardized methods. The timing of certain events and the tempo of maturation are important milestones.

EVALUATION OF CHILDREN WITH SHORT STATURE

Short stature, defined as a height below –2 standard deviations (SD) for age, sex, and genetic background, including midparental height, is a statistical definition encompassing 2.3% of all children. It may, however, be the first presenting sign for an underlying condition. Short stature is not itself a disease, and the majority of children with short stature have a *physiologic* variant, i.e., familial short stature, constitutional delay of growth and then of puberty, or idiopathic short stature. The last category, which is also not a diagnosis, is ever-shrinking as genetic variants that affect the growth plate are identified.[109-111] Large-scale genome-wide association studies have noted that height is highly polygenic (hundreds of genes), with thousands of genetic variants distributed across the entire genome. These height-associated regions are enriched for genes in multiple metabolic pathways associated with growth.[111]

Guidelines (guidance) for the referral of a short child to a pediatric subspecialist, usually a pediatric endocrinologist, and for their detailed evaluation may vary widely by country or among regions of a country.[112,113]

The goals of the evaluation are several. The evaluation should initially differentiate variants of normal from pathological (and often treatable) causes of short stature. The former may account for 80% of referred children in many pediatric endocrine clinics.[114] Accurate height measurements following standardized procedures are essential, although children are often referred with few accurate measurements. Proper technique includes an appropriate device and a standardized protocol, but differs for infants and children. Infants are measured horizontally (recumbent length) on an examining table device with a flat board at the top and a moveable foot piece. The measurement is a two-person procedure, with one at the head being sure that it is firmly against the end, with the nose straight up. The second person applies pressure to the knees so that they are flat against the table and moves the foot piece to the soles, with the toes pointing vertically. The length is read from the scale along the table (Fig. 17.1).

A stadiometer or wall-mounted device (e.g., steel tape) is used for children 2 years of age and older. Standing height is measured with the child (shoes removed) positioned with heels, back, and shoulders against the device or wall, and the head in the Frankfurt horizontal plane with the eyes looking straight ahead. A head piece (or right-angle triangle) is then moved to the top of the head. The height of the child is read either from the scale on the device or from the steel tape (Fig. 17.1). In some cases, the head piece may have to be applied with pressure to compress the hair. This procedure should be repeated after the child steps away from the device. The two measurements can be averaged if they agree within 0.4 cm, or the process repeated if they do not. Sitting height (upper segment) is measured with a similar device, except that the child sits on the table with the head in the same plane as noted above and the head piece lowered until it reaches the top of the head. Sitting height is then read from the attached scale. Leg length (lower segment) is estimated as the difference between standing height and sitting height, and the ratio between the two (sitting height/leg length) is the upper-to-lower ratio (Fig. 17.1).

History

The history should start with the pregnancy with the child to determine if there were difficulties (including drugs the mother may have taken) that might have led to a small-for-date child. Birth weight and length are important to determine if the child was small-for-date. Other early historical notes should include feeding problems, medical diagnoses and treatments, motor and mental development, and behavioral issues in addition to height and weight records. Others in the family with a similar condition, as well as consanguinity, may offer clues for diagnosis. The heights of the parents, preferably measured, will provide an indication of the expected, sex-specific target height and range.

A child's target height is based on the sex-adjusted midparental height. It is a projected adult height (and range) based on the heights of the biological parents. The sex adjustment is important, because men are on average 13 cm (~5 inches) taller than women. The simplest way to predict the genetic potential for adult height is to add the parental heights (cm or in) and then add 13 cm for boys or subtract 13 cm for girls and then take the average. That places the other-sex parent on the same centile on the opposite sex chart. This is an estimate the genetic potential for *all* children of these parents.

For the individual child, adult height is predicted from his/her present height and bone age using, for example, the Bayley and Pinneau tables in the Greulich and Pyle atlas.[41] This may be done because bone age denotes approximately the percentage of adult height that one is at the time of the radiograph.

It should be noted that many conditions that present as short (or tall) stature do not begin at birth. As such, information that may identify a secondary growth disorder (due to nutrition, hormones, inflammatory cytokines, or extracellular fluid, as in renal tubular acidosis) should be sought. With emphasis on the growth curves for weight and height, a significant weight loss or gain, or a slowing of growth in height and when it began should be noted. Issues of relevance include growth faltering (in addition to short stature) or acceleration. Diseases of the gastrointestinal tract are often associated with both decreased weight and growth faltering, whereas several endocrine disorders may present a weight percentile greater than the height percentile (increased weight-for-height) and a growth trajectory that may be negative for height but positive for weight (as may be noted with Cushing syndrome or severe hypothyroidism). Child complaints should also be considered; for example, abdominal pain/discomfort and diarrhea may suggest inflammatory bowel disease, or severe headaches may suggest an intracranial mass. The history of medication is critical, even if certain medications have been stopped by the time the patient presents for evaluation, e.g., drugs for attention deficit disorder or glucocorticoids for inflammatory conditions. The family history may again be helpful concerning others with a similar condition and/or psychosocial, intellectual, or behavioral issues.

Physical Examination

The physical examination is likely more helpful in those with primary growth disorders (clinically defined syndromes and skeletal dysplasias caused by variants of genes associated with the growth plate), e.g., dysmorphic features, abnormal body proportions, and cardiac and other organ system abnormalities (see Fig. 17.7 for growth plate physiology). Nevertheless, most children with acquired growth faltering present a relatively normal physical examination and normal body proportions.

The degree of short stature may also offer clues to diagnosis. Children whose heights are slightly below the –2 SD cutoff are more likely to have a variant of normal growth than children with heights below –3 or –4 SD. Among more than 785 subjects with severe short stature (heights below –3 SD) after 3 years of age in the Helsinki University Hospital district, a pathologic cause for short stature, whether primary or secondary, was noted in approximately 70% of the children.[115] The conditions included various syndromes, disorders of organ systems, GH deficiency (GHD), small-for-gestational age *without* catch-up growth, and skeletal dysplasias. Normal variant, idiopathic short stature was far more prevalent at heights –3.0 to –3.5 SD than at heights more than –4.0 or –5.0 SD. More severe growth disorders (–4.0 SD and greater) were noted in children with syndromes and skeletal dysplasias (both primary growth disorders). Thus, the degree of short stature, especially at the extremes, separates the likelihood of a variant of normal short stature from the diagnosis of a pathological form.

Preliminary Laboratory Evaluation

The information in the preceding paragraphs focuses on the appropriate evaluation for children with asymptomatic short stature, i.e., those children who are short but growing at the lower end of the normal range for age. Many will have heights in the –2.0 to –3.0 SD range and a height velocity that follows along one of those trajectories. If they are truly asymptomatic after a careful history and a normal physical exam, it is overwhelmingly likely that they have a variant of normal growth and can be followed by the primary care physician or endocrinologist without laboratory evaluation, with the exception of a bone age determination. Evidence suggests a low incidence of pathology among children with asymptomatic short stature. Among 235 such patients, at most only three new pathologic diagnoses were noted; nevertheless, laboratory screening following the guidance of several Pediatric Endocrine Societies cost more than $100,000 per diagnosis.[116] The authors concluded that "… a clinician should use history, review of systems, and physical examination to guide the evaluation of individuals with short stature" [p 1050]. This information is quite helpful from several perspectives. Certain children with short stature do not require evaluation by an endocrinologist, although parents often request one. Perhaps more importantly, the history and physical examination, along with minimal laboratory testing, may lead the referring physician to a more appropriate pediatric subspecialist, for example, gastroenterologist, nephrologist, or pulmonologist.

KEY POINTS

- Is the child short?
- What is the height velocity?
- Detailed personal and family history
- Physical examination
- Bone age
- *Parsimonious* laboratory evaluation (at first)

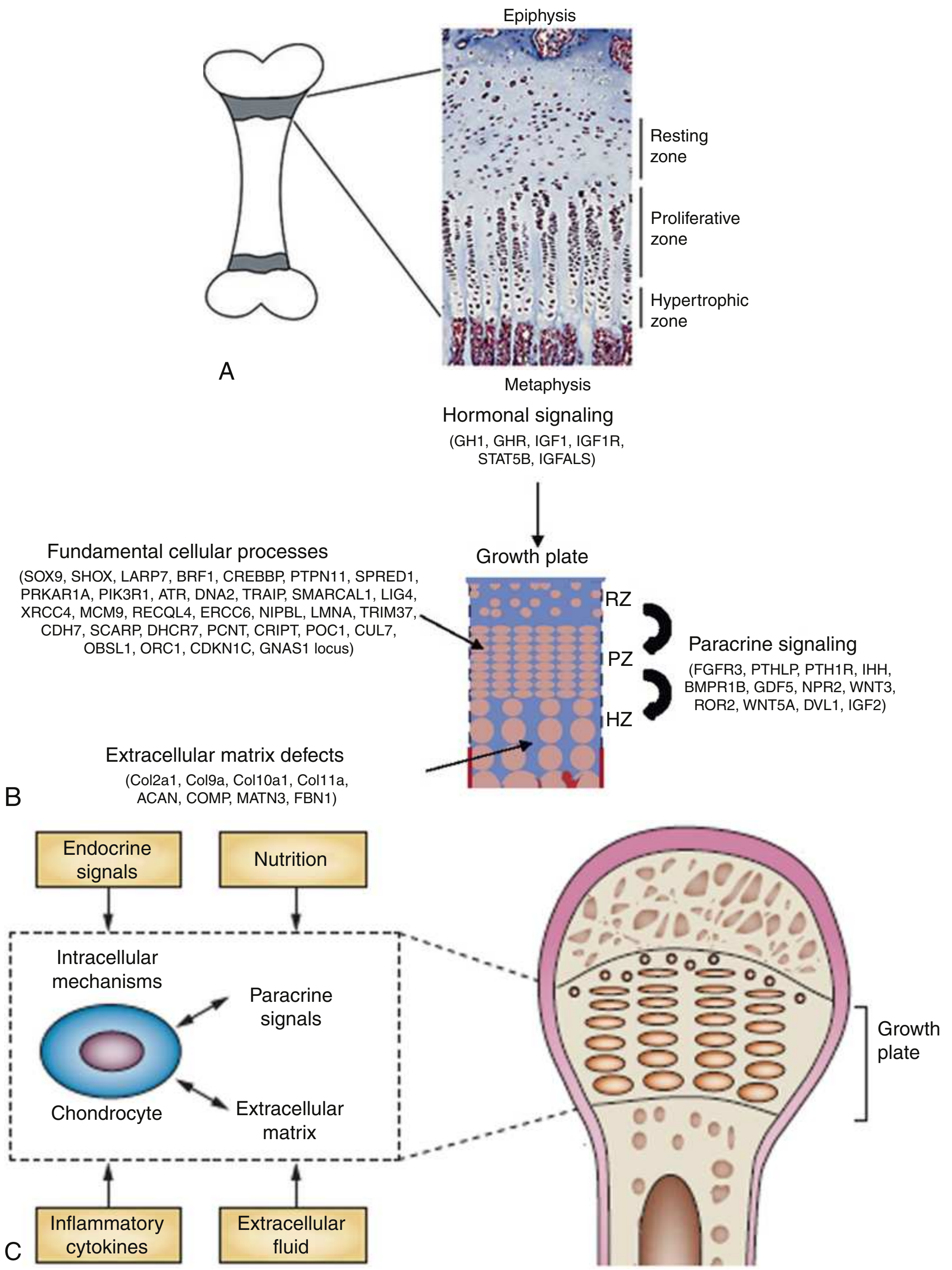

Epiphysis
Resting zone
Proliferative zone
Hypertrophic zone
Metaphysis
A
Hormonal signaling
(GH1, GHR, IGF1, IGF1R, STAT5B, IGFALS)
Growth plate
Fundamental cellular processes
(SOX9, SHOX, LARP7, BRF1, CREBBP, PTPN11, SPRED1, PRKAR1A, PIK3R1, ATR, DNA2, TRAIP, SMARCAL1, LIG4, XRCC4, MCM9, RECQL4, ERCC6, NIPBL, LMNA, TRIM37, CDH7, SCARP, DHCR7, PCNT, CRIPT, POC1, CUL7, OBSL1, ORC1, CDKN1C, GNAS1 locus)
RZ
PZ
HZ
Paracrine signaling
(FGFR3, PTHLP, PTH1R, IHH, BMPR1B, GDF5, NPR2, WNT3, ROR2, WNT5A, DVL1, IGF2)
Extracellular matrix defects
(Col2a1, Col9a, Col10a1, Col11a, ACAN, COMP, MATN3, FBN1)
B
Endocrine signals
Nutrition
Intracellular mechanisms
Paracrine signals
Chondrocyte
Extracellular matrix
Growth plate
Inflammatory cytokines
Extracellular fluid
C

Fig. 17.7 **Growth Plate Physiology.** **A,** Histology of the growth plate. The growth plate is a thin cartilage structure situated in the ends of tubular bones. It is commonly subdivided into three distinct zones; the resting, proliferative, and hypertrophic zones. **B,** Molecular mechanisms involved in longitudinal growth. **C,** Human growth plate histology from an 11-year-old boy. The growth plate comprises three histologically and functionally distinct zones; the resting, proliferative, and hypertrophic zones. Bar represents 100 μm. (**A** from Nilsson O, Marino R, De Luca F, et al. Endocrine regulation of the growth plate. *Horm Res.* 2005;64:157–165; **B** from Andrade AC, Jee YH, Nilsson O. New genetic diagnoses of short stature provide insights into local regulation of childhood growth. *Horm Res Pediatr.* 2017;88:22–37; **C** from Baron J, Sävendahl L, De Luca F, et al. Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol.* 2015;11:735–746.)

Evaluation for Endocrine Disorders of Growth

The hallmarks of endocrine disorders of growth are greater deflections in height velocity and height standard deviation score (SDS) than in weight velocity or weight status. An upward trajectory in weight may be noted as height SDS is declining. The most prevalent endocrine system–related conditions include hypothyroidism, Cushing syndrome, and GHD, and perhaps GH insensitivity. If poor nutrition, either malabsorption or insufficient intake, is prominent as in celiac disease, inflammatory bowel disease, or chronic caloric deficit, the loss of weight is generally in excess of that in height. Hypothyroidism and Cushing syndrome are described in specific chapters. The current discussion focuses on disorders of the GH/IGF-1 axis.

Growth Hormone/Insulin Growth Factor-1 Axis

The GH/IGF-1 system is the main regulator of postnatal human growth (Figs. 17.8 and 17.9). It regulates this complex process by integrating

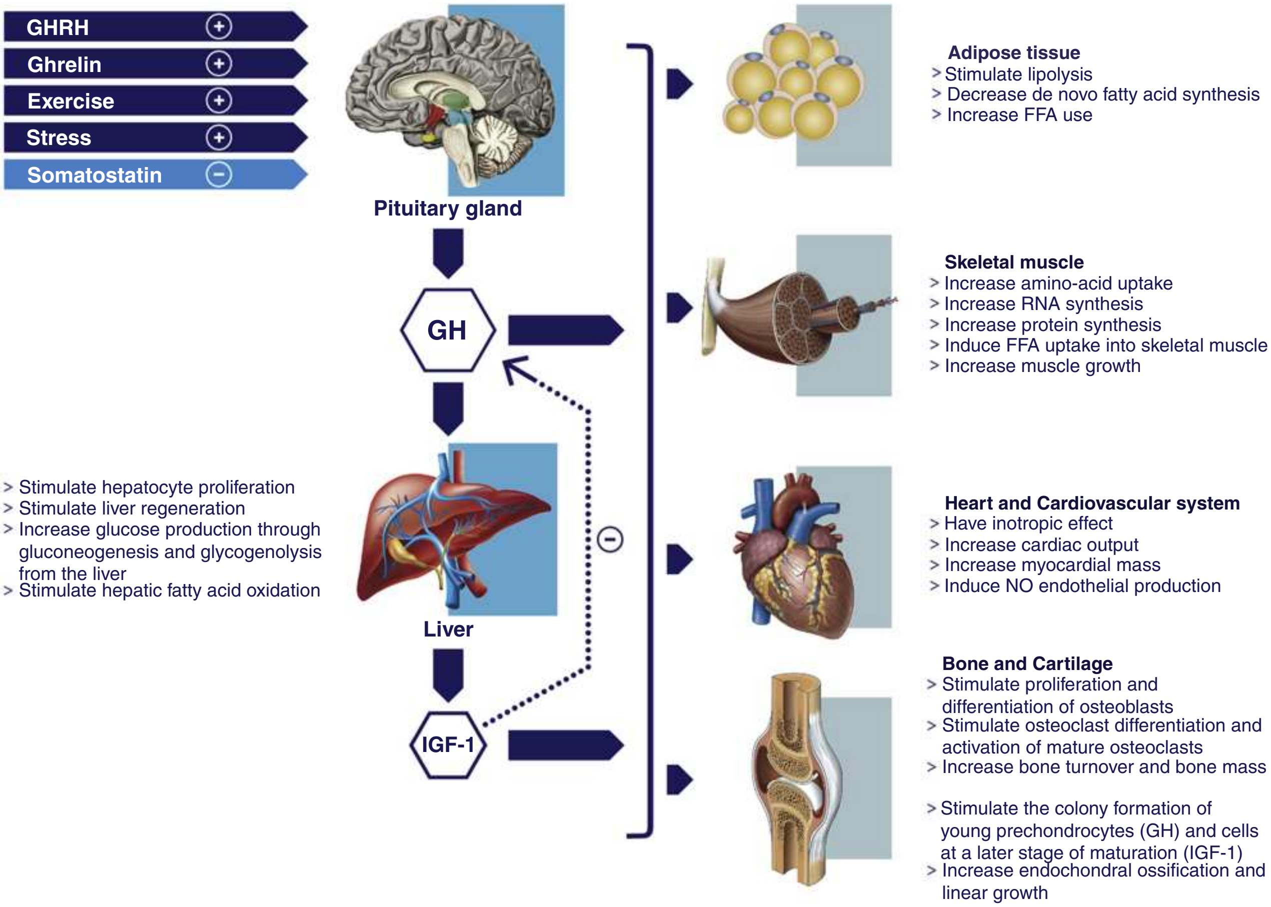

GHRH ⊕
Ghrelin ⊕
Exercise ⊕
Stress ⊕
Somatostatin ⊖
Pituitary gland
GH
Liver
IGF-1
⊖
> Stimulate hepatocyte proliferation
> Stimulate liver regeneration
> Increase glucose production through gluconeogenesis and glycogenolysis from the liver
> Stimulate hepatic fatty acid oxidation
Adipose tissue
> Stimulate lipolysis
> Decrease de novo fatty acid synthesis
> Increase FFA use
Skeletal muscle
> Increase amino-acid uptake
> Increase RNA synthesis
> Increase protein synthesis
> Induce FFA uptake into skeletal muscle
> Increase muscle growth
Heart and Cardiovascular system
> Have inotropic effect
> Increase cardiac output
> Increase myocardial mass
> Induce NO endothelial production
Bone and Cartilage
> Stimulate proliferation and differentiation of osteoblasts
> Stimulate osteoclast differentiation and activation of mature osteoclasts
> Increase bone turnover and bone mass
> Stimulate the colony formation of young prechondrocytes (GH) and cells at a later stage of maturation (IGF-1)
> Increase endochondral ossification and linear growth

Fig. 17.8 **Growth Hormone/Insulin Growth Factor-1 Main Actions on Liver, Adipose Tissue, Skeletal Muscle, Heart and Cardiovascular System, Bone, and Cartilage.** *FFA*, Free fatty acids; *NO*, nitric oxide. (Sbardella E, Pozza C, Isidori AM, et al. Dealing with transition in young patients with pituitary disorders. *Eur J Endocrinol*. 2019;181:R155–R171.)

genetic and nutritional factors, along with signals from thyroid hormone, insulin, and glucocorticoids. Psychosocial aspects may also be important.[117]

Human GH is a heterogeneous protein with several molecular isoforms. This heterogeneity is expressed at the level of the *GH* gene, mRNA splicing, posttranslational processing, and GH metabolism, and poses a challenge to complete understanding of GH bioactivity, accurate measurement, and assay standardization.[118,119] Two *GH* genes, *GH1* and *GH2*, are on the long arm of chromosome 17q23.3: the former encodes the predominant isoform, 22 kDa (191 amino acids), which contains two intramolecular disulfide bonds.

Growth Hormone Release and Action

The release of GH from the anterior pituitary is mediated by two peptides secreted by the hypothalamus, GH-releasing hormone (GHRH) and the inhibitory hormone somatostatin (somatotropin release–inhibiting hormone [SRIH]) (Fig 17.8). A pulse of GH is generated by the simultaneous rise in GHRH and decline in SRIH. The amount of GHRH released likely determines the amplitude of the GH peak, and the frequency and duration of the GH secretory event is primarily under SRIH control. The pulse is also stimulated by ghrelin, which is produced mainly in the stomach, but also in the hypothalamus. IGF-1 and GH itself exert negative-feedback control at the hypothalamus. GH is transported in the serum by binding proteins: the major one, GHBP, is identical to the receptor.[118]

GH acts by binding to the GH receptor (GHR), which is expressed in most tissues. The GHR is a transmembrane receptor belonging to the class 1 cytokine receptor family. Once this binding takes place, the GHR undergoes conformational changes that result in activation of Janus kinases (JAK2), followed by recruitment and phosphorylation of signal transducers and activators of transcription (STATs) that, among other actions, induce the synthesis of IGF-1, IGFBP3, and the acid-labile subunit (ALS) encoded by the IGFALS gene (Fig. 17.9).[120] This ternary complex circulates assembled as a 150-kDa protein, then IGF-1 (a 7.6-kDa protein, 70 amino acids), whose gene is located at chromosome 12q23, is liberated so it may bind with its receptor (IGF-IR) to promote growth. Circulating IGF-1, stimulated by GH action on the liver, is the main mediator of generalized GH actions; however, it is not the only route through which GH can exert its effects, for example, at the growth plate.[121,122] IGF-1 levels are age- and sex-dependent. Serum levels are high in the fetus, but drop shortly after birth and then increase slowly until late prepuberty or very early puberty. The levels subsequently increase approximately 3-fold, concomitant with the rise in the quantity of GH secreted, peaking near the time of PHV.[123] Peak values occur earlier in girls as their pubertal maturation increases.

Once free IGF-1 binds to its receptor, signaling cascades are triggered intracellularly, with two essential ones being the mitogen-activated protein kinase (MAPK) and the phosphatidylinositol 3-kinase (PI3K) cascades. The MAPK cascade is primarily responsible for the proliferation of muscle cells, and the PI3K cascade for cell differentiation.

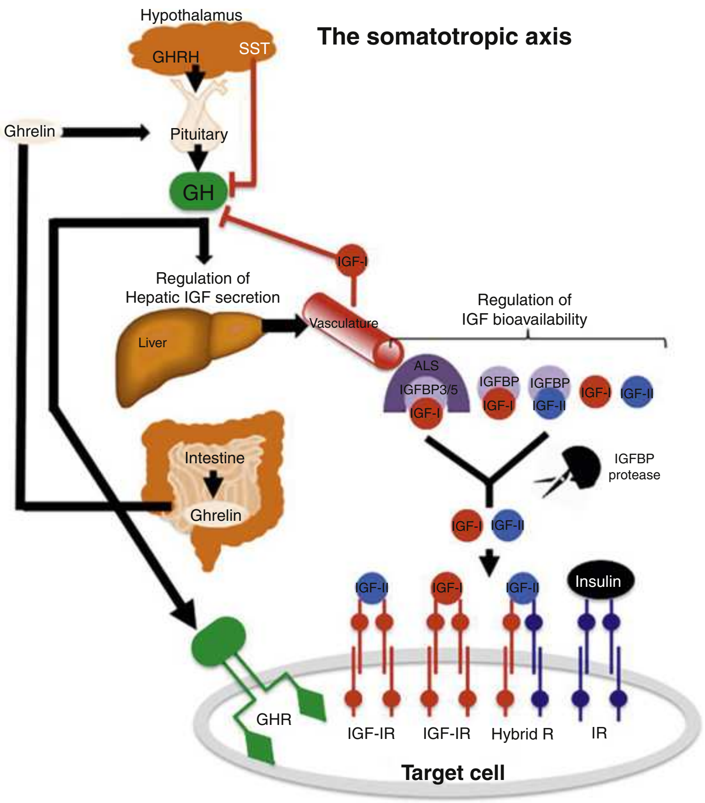


Fig. 17.9 **The Somatotropic Axis.** Hypothalamic growth hormone (GH)-releasing hormone and somatostatin, as well as intestine-secreted ghrelin, regulate pituitary secretion of GH. Once released to the circulation, GH stimulates liver production of insulin growth factors (IGFs) and a few of the IGF binding proteins (IGFBPs). In the vasculature, IGFs are found in binary and ternary complexes, which increase their half-lives. Serum and tissue proteases act upon the IGFBPs to liberate IGFs and increase their bioavailability. Target cells expressing the IGF-IR, insulin receptor, or hybrid receptors bind the IGFs and initiate phosphorylation cascades to enhance cellular proliferation, differentiation, and function. GH receptor is found on almost all cells. Upon binding to its receptor, GH initiates signaling cascades to promote cellular function that may be IGF-dependent or -independent. (Yakar, S, Werner H, Rosen CJ. Insulin-like growth factors: actions on the skeleton. *J Mol Endocrinol.* 2018;61:T115–T137.)

GH and IGF-1 (and other growth factors) work in concert to promote cartilage and bone growth. GH has a dual effect on the growth of epiphyseal cartilage and differentiation of cartilage cells, as well as the generation of IGF-1. The proximal zone of cartilage, which is close to the bony segment of the epiphysis, consists of a narrow band of germinal or stem cell chondrocytes. GH preferentially stimulates differentiation of these prechondrocytes, while IGF-1 stimulates the clonal expansion of the more differentiated cells in the distal proliferative zone. GH, but not IGF-1, stimulates lipolysis.

GHD may be congenital or acquired. It may also be isolated or in combination with other anterior hormone deficiencies (multiple pituitary hormone deficiencies [MPHD]) and dysmorphic findings or with involvement of other organ systems (see also Chapter 18). Children with congenital severe GHD have only a slightly reduced birth length and may not immediately show growth failure, which becomes prominent in the second half of the first year. The children also show a higher frequency of breech presentation and perinatal asphyxia. Neonatal morbidity may include hypoglycemia and prolonged jaundice. When GHD is combined with deficiency of adrenocorticotropic hormone (ACTH), hypoglycemia may be severe. The combination of GHD with gonadotropin deficiency can cause microphallus, cryptorchidism, and hypoplasia of the scrotum.

Children with acquired GHD present with severe growth failure, delayed bone age, and increased weight:height ratios. Causes of acquired GHD include intracranial tumors involving the hypothalamic-pituitary region (e.g., craniopharyngioma), cranial irradiation, and head trauma (Table 17.4).

Another cause of hypopituitarism in children more than 2 years of age is psychosocial short stature, also known as deprivational short stature. Children exposed to a severely traumatic home environment are characterized by bizarre behaviors including gorging and vomiting (hyperphagia); sleep disturbance, night wandering often in search of food; pain agnosia; abnormal and disturbed relationship with primary caregiver; temper tantrums; and poor peer relationships. Objectively, the children have subnormal height velocity, but often an appropriate weight for height.

Catch-up growth may be marked with removal from the disturbed environment, as may be amelioration of many of the signs and symptoms noted previously. The relationship to the GH/IGF-1 axis is that, if tested early after removal from the environment, whether measured by the spontaneous release of GH or by the GH response to a pharmacologic stimulus, the children respond as if GH-deficient. Within a few days, responses to these tests generally revert to normal. Growth rates become remarkably high within weeks following removal from the home environment, even during hospitalization. Given this high rate of growth following an extended period of subnormal growth and the reversion of the testing to normal, this condition may be considered a form of *reversible* hypopituitarism, at least for the GH/IGF-1 axis, and often for the corticotropin-releasing hormone/ACTH/adrenal axis.[117]

TABLE 17.4 Genetic Disorders of Growth Hormone (GH)

Gene	Phenotype	Inheritance	OMIM #
Isolated GH deficiency			
GH1	Isolated GH deficiency	AR, AD	139250, 612781, 173100, 262400
GHRHR	Isolated GH deficiency	AR	139197
Combined pituitary hormone deficiency			
POU1F1	GH, TSH, prolactin deficiencies	AR, AD	173110
PROP1	GH, TSH, LH, FSH, prolactin and evolving ACTH deficiencies	AR	262600
Specific syndromes			
HESX1	Septooptic dysplasia	AR, AD	182230
LHX3	GH, TSH, LH, FSH, prolactin deficiencies; limited neck rotation	AR	600577
LHX4	GH, TSH, ACTH deficiencies with cerebellar abnormalities	AD	262700
SOX3	Hypopituitarism and mental retardation	X-linked	312000
GLI2	Holoprosencephaly and multiple midline defects	AD	610829
SOX2	Anophthalmia, hypopituitarism, esophageal atresia	AD	206900
GLI3	Pallister–Hall syndrome	AD	146510
PITX2	Rieger syndrome	AD	180500

AD, autosomal dominant; *AR*, autosomal recessive; *TSH*, thyroid-stimulating hormone; *LH*, luteinizing hormone; *FSH*, follicle-stimulating hormone; *ACTH*, adrenocorticotropic hormone.

A number of conditions (syndromes) that involve linear growth and body composition have become of interest to internists because the children/adolescents who present with them survive into young adulthood. The largest group is likely the survivors of childhood cancer, many of whom have endocrine system–related morbidities, especially those related to growth and puberty.[124-126]

New to many internists are those adolescents with Prader–Willi syndrome or cystic fibrosis. Prader–Willi syndrome (OMIM 176270) is a multisystem, genetically heterogeneous condition caused by a lack of paternal gene expression in the chromosome 15q11-q13 PWS locus. Although the clinical phenotype may change over time, many adolescents and emerging adults have had the benefit of GH therapy, with increased adult height and a more normal body composition, indicated by an increased lean body mass and decreasing fat mass; although both remain abnormal in most subjects. The most common metabolic issue, in addition to hypogonadism, either secondary or mixed primary and secondary, is insulin resistance and disordered glucose metabolism. This often occurs in concert with the unresolved issue of adult treatment with recombinant human GH (rhGH), with effects on body composition and the regional distribution of body fat, as well as quality of life.[127]

Three and four decades ago it was uncommon for children and adolescents with the usual mutation causing cystic fibrosis to survive to emerging adulthood; many now survive into their fourth and fifth decades. Nutritional support is the cornerstone for metabolic function, but virtually half have a “new variety” of diabetes mellitus or cystic fibrosis–related diabetes, whose treatment is radically different from that for either type 1 or type 2 diabetes mellitus. The key deficiency is insulin itself and its strong anabolic action in muscle tissue. As this disease complication unfold during adolescence, there is a diminution of the anabolic action on muscle, and often a worsening of the pulmonary disease.[128] Patients with these and similar conditions now require an endocrinologist for part of their ongoing medical and psychological care.

Diagnostic Approach

GHD should be considered in the context of auxology, i.e., growth faltering and usually an increased weight:height ratio, but it is *confirmed* by biochemical and stimulation testing. The first step is to evaluate for other potential causes of growth failure, including chronic systemic disease, e.g., hypothyroidism, Turner syndrome (in girls), and skeletal disorders. This is accomplished through a thorough medical history, physical examination, and bone age determination. Additional laboratory evaluation should be performed when appropriate, including screens for systemic disease, undernutrition, inflammation, and thyroid function, and a karyotype in girls to rule out Turner syndrome. GHD is effectively excluded in children with normal height velocity and bone age.

Specific to GHD, in addition to the bone age radiograph obtained for most evaluations of short stature, an IGF-1 level should be measured; in children below age 3 years, an IGFBP-3 level should be measured. Although individual values will not permit the specific diagnosis of GHD, children with levels above 0 SDS for age and sex may be excluded. These findings are generally sufficient to exclude GHD without stimulation tests (see later).

When GHD is congenital and near-complete, the diagnosis is straightforward, as affected children present with severe growth failure, delayed bone age, and very low serum concentrations of GH, IGF-1, and IGFBP-3. For patients with these clinical characteristics, it is reasonable to make the diagnosis of GHD without performing GH stimulation testing (see later). Lesser degrees of growth faltering and decreased IGF-1 and IGFBP-3 levels are consistent with GHD, but are also consistent with a number of other causes of growth failure, including poor nutrition. If not explicable on the basis of undernutrition, low IGF-1 and/or IGFBP-3 levels are strongly suggestive of a diagnosis of GHD, but this must be confirmed by provocative GH testing (see later). MRI of the hypothalamo-pituitary region (with and without contrast) is recommended for children with suspected GHD and may be specifically useful in infants, where it is more likely that structural abnormalities will be found.[129]

A significant minority of children (3%–30%, depending on the study) have an affected parent or sibling, while several genetic causes, for example, transcription factor mutations or mutations of GH itself, the GHRH receptor $G_{s\alpha}$ or the GH secretagogue receptor, have been described in detail.[130,131]

Growth Hormone Stimulation Tests

Indications. Provocative (stimulation) GH testing is indicated for most patients to confirm a diagnosis of GHD. Because these tests have limitations, the results should not be used as the sole diagnostic criterion and should be interpreted in the context of auxologic findings, bone age, and IGF-1 and IGFBP-3 concentrations.[113] Provocative testing is *not* necessary for patients in whom other clinical criteria are sufficient to make the diagnosis of GHD:

- Pituitary abnormality (secondary to a congenital anomaly, tumor, or irradiation) and a known deficiency of at least one other pituitary hormone, in addition to auxologic criteria.[113,132]
- Newborn with a congenital pituitary abnormality (ectopic posterior pituitary and pituitary hypoplasia with abnormal stalk) or known deficiency of a pituitary hormone, along with hypoglycemia, at which time a simultaneous serum GH concentration is less than 5 mcg/L.[132]
- Infant or young child with extreme short stature (height <–3 SD), normal nutrition, significantly reduced IGF-1 (<–2 SD) and IGFBP-3, and delayed bone age. This is the classic presentation of congenital and severe GHD, including MPHD; most experts agree that provocative testing is not required to make the diagnosis.

Limitations of Growth Hormone Stimulation Tests

- Tests are pharmacologic and may not indicate the underlying physiology.
- Reproducibility in both normal and GHD patients is poor, even with the same tests done on different days.
- Age and sex steroid hormone status affect results, and use of sex-steroid "priming" does not have a consensus.
- Nutritional adequacy and body composition, particularly adipose tissue, can affect results.
- GH assays vary, with large interassay variance.
- All are expensive.

Growth Hormone Treatment (See Also Chapter 18)

The effects of GH treatment for GHD differ markedly depending upon the life phase of the individual. For infants and children, it is used mainly to increase linear growth, although in the former it may be important to prevent hypoglycemia as well. Relatively higher doses are necessary than those required in adults for promotion of lean body mass, lipolysis, and salutary effects on quality of life. The preceding pertains only to GHD, for children who receive GH therapy for any of the other seven FDA-approved indications will not require GH therapy as adults. Only a minority of those treated for GHD as children will require GH as adults, making it mandatory (with some exceptions) that stimulation testing be done at the transition of adolescent patients to emerging adulthood.

A special case may be made for those who receive cranial radiation therapy for central nervous system tumors. The "dose-response" curves for GHD depend on the type, photon or particle, with the latter having a sharper focus and causing less collateral destruction, as well as differences in how fractionated the radiation is. The particle-based therapies are often given just once. The dose should be related to that delivered to the hypothalamus/pituitary and not to the targeted tumor. Early on it is the hypothalamic area that is most sensitive, but later the pituitary itself must be considered. Generally speaking, a dose of less than 18 Gy does not cause hypopituitarism, but doses between 18 and 24 Gy increasingly are found in irradiated children. Importantly, those children who receive more than 18 Gy of radiation to the hypothalamus should be followed lifelong, at first for growth and then for the adult consequences of GHD.[126,133]

Although not completely agreed upon, there does not appear to be an increase in the recurrence rates for central nervous system tumors following irradiation, whether treated with rhGH or not. In a similar manner, the more recent studies note virtually no differences in secondary neoplasms (mainly meningiomas) following irradiation therapy based on treatment or not with rhGH.[133,134]

Most centers wait 1 year after the end of therapy to begin GH treatment, although the scientific basis for this practice is limited. For a "benign" tumor, craniopharyngioma, there seems little reason to wait that long if it is obvious that the child/adolescent is GH-deficient.

Dosing and Monitoring of Growth Hormone Therapy

The typical range of starting doses of rhGH is between 20 and 35 micrograms/kg/day, administered subcutaneously once daily (0.16–0.24 mg/kg/week); the lower end of the range is used for patients with severe GHD. Treatment should be started as early as possible to achieve the greatest growth response. This is likely to change soon, as multiple formulations of longer-acting GH are in phase 3 trials in children.

Monitoring is usually accomplished by measuring height and estimating height velocity and IGF-1 at 3- to 4-monthly intervals. After 6 to 12 months, more credence is given to the height velocity than the IGF-1 level, as some children will show catch-up growth even at relatively low IGF-1 SDS. The IGF-1 level serves as a safety marker, as one tries to keep it within +0.5 to +1.5 SDS for age and sex (stage of maturation during puberty). One may also use the IGF-1 level at 1 month for both a safety (too high) and efficacy (above 0 SDS) marker and increase or decrease the dose before the 3-month visit. The latter provides the physician an additional opportunity to evaluate how the patient and parents are reacting to the new medication.

GH therapy is generally continued until linear growth decreases to less than 1.5 to 2.0 cm/year. However, this is an individualized decision, and in some adolescents it may be appropriate to terminate therapy sooner. Patients should be retested (see later) for GHD using GH stimulation tests to determine if they remain GH-deficient. Treatment may then be continued for metabolic indications, i.e., to maintain healthy body composition, lipid profiles, and bone mass.

The more common adverse effects of rhGH therapy include unmet patient and parental expectations, idiopathic intracranial hypertension (especially in the first months), slipped capital femoral epiphysis, and worsening scoliosis (because of more rapid growth). Growth-inhibiting antibody production is exceedingly rare with the recombinant products. Longer-term risks, including new and recurring benign and malignant conditions, have recently been described in detail.[135]

In patients in whom rhGH treatment is not begun until adolescence, the patient may not acquire the desired height with GH therapy alone, as there already may be significant partial closure of the growth plates. To delay further epiphyseal maturation, adjunctive treatment with a gonadotropin-releasing hormone agonist or an aromatase inhibitor has been explored. These are not generally recommended, because their safety and efficacy have not been established.

Retesting for Growth Hormone Deficiency During the Transition to Emerging Adulthood

During the transition from adolescence to emerging and young adulthood after linear growth has essentially ceased, growth in bone mass and lean tissue mass (composition) continues and reflects the metabolic actions of GH. GH is a necessary element, in addition to quality of life.[136] Some adolescents will require retesting. Those with congenital, isolated GHD require retesting, because more than two thirds will have normal provocative tests near the end of the growth period. In those with genetic, organic, or structural causes of GHD, the deficiency is permanent in the vast majority. Individuals whose GHD is due to radiation therapy should be followed for a longer period of time, because the effect of radiation on the GH axis may not be manifest for 10 years.

How should the retesting be done? The first step is to measure an IGF-1 level after 4 weeks without receiving rhGH at the time of a significant decrement in height velocity, e.g., 2.0 cm/year. A normal test indicates sufficient GH, and these patients do not require additional testing or GH therapy. If the IGF-1 level is below –2 SD, one may proceed to stimulation testing. However, there are certain conditions for which retesting is not necessary. These include patients with three or more other anterior pituitary deficiencies, those with genetic causes of GHD, and those with structural causes of GHD. Although a plethora of tests have been described, insulin-induced hypoglycemia and glucagon are the most common. The cutoff values for the adult differ from those of children and adolescents, and are lower depending on the stimulus and laboratory normal range.

Macimorelin, which has high reproducibility, sensitivity, and specificity, is the newest FDA-approved agent for adults.[137] It is currently in clinical trials for patients aged less than 18 years. Patients remaining GH-deficient will require continued rhGH therapy, although the dose is very much smaller, varying from approximately 0.3 mg/day to 1.0 mg/day, with the higher doses for women. There is no consensus on how to gradually decrease the high doses for the promotion of growth in these adolescents (in the range of 1.5–3.5 mg/day) to the much smaller doses noted above for the adult. Common practice is to do so slowly over 1 to 2 years. The dose is often higher in girls, especially those taking oral estrogens. Rather than monitoring height velocity, one may monitor IGF-1 levels and perhaps perform an occasional dual-energy x-ray absorptiometry (DXA) scan for body composition and bone mineral density. Adequate therapy is indicated by an IGF-1 level between 0 and 1 SD and the absence of symptoms referable to GHD in the adult.

Retesting for GHD in late adolescence is limited to those who have GHD *per se*; the other seven indications for children are concerned with height velocity and near-adult height, but do not carry over to the adult. However, some recent studies point to an adult indication for those with the Prader–Willi syndrome; but it is indicated for its effects on body composition and quality of life.[138,139]

KEY POINTS

- Growth hormone (GH)/insulin growth factor (IGF)-1 axis
  - GH deficiency
  - GH insensitivity
- **Diagnostic approach**
  - Auxology
  - Skeletal age
  - IGF-1 level
  - GH stimulation tests
  - GH treatment and monitoring
    - Height velocity
    - Change in bone age

EVALUATION OF CHILDREN WITH TALL STATURE

Similar to the evaluation of children with short stature and growth faltering, a systematic diagnostic approach for children with tall stature or accelerated growth should be in place. Tall stature is indicated by height above +2 SD for children of the same age and sex, or above +2 SD for midparent height, while accelerated growth should be diagnosed or excluded in the context of primary (syndromic) or secondary growth disorders, e.g., excessive GH or premature secretion of sex steroid hormones. As with many children evaluated for short stature, those with tall stature are likely to have a variant of normal growth or constitutional tall stature, and to have tall parents.

The rationale and diagnostic approach to tall stature have been recently reviewed.[140] These conditions are considerably less common than those with short stature in most pediatric endocrine clinics. Nevertheless, several common syndromes present with tall stature and accelerated growth. The first includes several X and Y (sex) chromosome aneuploidies (SCAs); the most prevalent are Klinefelter syndrome (47,XXY) and 47,XXX, while the less common are 47,XYY and 48,XXYY. The genotypes are rarely diagnosed before puberty, except for increasingly routine prenatal diagnosis based on multiple types of genetic diagnoses, for example, array analysis, karyotype, and whole-exome sequence analysis, or diagnosis because of behavioral and learning phenotypes in these children. The hypothalamo-pituitary gonadal axis differs among those with SCAs:

- Males with a 47,XXY karyotype are tall with eunuchoid body proportions. They have hypergonadotropic hypogonadism and small, often firm, testes.
- Males with a 47,XYY karyotype have normal size testes and generally normal hypothalamic–pituitary–gonadal (HPG) axis hormones, i.e., slightly below-average luteinizing hormone (LH), testosterone, and inhibin B concentrations, and slightly above-average follicle-stimulating hormone (FSH) and antimullerian hormone concentrations.
- Girls with 47,XXX are tall and hypotonic. They have a mainly normal HPG axis, i.e., sexual maturation occurs largely within the physiologic range. However, psychological features including an increased prevalence of attention deficit, and mood disorders may be prominent at evaluation.
- Males with the very much less common 48,XXYY karyotype are tall with eunuchoid body proportions. They have hypergonadotropic hypogonadism (although often to a lesser degree than those with the 47,XXY karyotype), gynecomastia, small testes with germinal aplasia, and tubular hypoplasia.

A second common overgrowth syndrome is Marfan syndrome (OMIM 154700), for which a systemic clinical scoring system has been devised, followed by sequence analysis of the *FBN1* gene or of a panel of connective tissue disorders that includes the *FBN1* gene.[140] Thus, the sequence of evaluation includes diagnostic clues from personal and family histories of the patient; physical examination for dysmorphology and body proportions (height and sitting height), and of the genitalia; and evaluation of height growth curves for accelerated growth. Together, these analyses should permit a directed laboratory investigation.[140]

There are several other causes of syndromic tall stature, and use of genetic tests is increasingly more prominent in the diagnosis of these relatively rare and perhaps atypical conditions in which dysmorphic findings and comorbidities are consistent findings.[141]

KEY POINTS

- Is the child tall?
- What is the height velocity?
- Detailed personal and family history
- Physical examination
- Bone age
- *Parsimonious* laboratory evaluation (at first)

ENDOCRINE CAUSES OF TALL STATURE

Endocrine causes of tall stature include central precocious puberty and peripheral sexual precocity, mainly non–salt-losing congenital adrenal hyperplasia. (The salt-losing variety either leads to early diagnosis and treatment, or neonatal death in adrenal crisis.)

Constitutional activation of the LH receptor, which is also called "testotoxicosis," is less common. It is also uncommon for hyperthyroidism to cause significantly accelerated growth and tall stature, because the signs and symptoms of the disease are usually noted early, and treatment is well described. "Catch-down growth," that is, below the normal rate, is often noted with effective treatment of the hyperthyroidism.

Less commonly, a child (or family) with familial glucocorticoid deficiency or familial glucocorticoid resistance will present to a clinic; with this condition, ACTH drives adrenal androgen secretion and peripheral sexual precocity. Appropriate treatment with glucocorticoids slows the height velocity and diminishes genital and bone age advancement more in the former than in the latter.[142]

The McCune–Albright syndrome consists of polyosteotic fibrous dysplasia, precocious (and occasionally) peripheral sexual precocity, and café-au-lait skin pigmentation leading to tall stature in boys and girls, but is more common in girls. GH excess is considerably less common, but can occur along with other uncommon associated endocrinopathies: hyperthyroidism, Cushing disease, FGF23-mediated phosphate wasting, and hypercortisolism.[143]

Because estrogens accelerate epiphyseal closure, youth with syndromes of diminished estrogen synthesis (aromatase deficiency) or estrogen action (estrogen receptor α deficiency) will continue to grow well beyond adolescence. Both also have eunuchoid proportions and osteoporotic bones. However, those with the estrogen receptor α deficiency will have absent breast development despite elevated levels of estradiol and do not respond to exogenous estradiol with breast development. Those with aromatase deficiency will have virtually no circulating estradiol, but will respond to exogenous estradiol with breast development and closure of the epiphyses of the long bones.

GH excess causes both gigantism and acromegaly as the long bone epiphyses close. Its diagnosis is based on accelerated growth, acral enlargement, and a pituitary mass as noted on cranial MRI scan. It may often be due to a germline mutation (see Chapter 18). Very rarely, especially in children, accelerated growth and acral changes may be due to the ectopic secretion of GHRH. Treatment for the rare child with this condition differs little from that for adults—transsphenoidal tumor removal (if possible) and either somatostatin analogs, GH receptor antagonists, or radiation therapy.

KEY POINTS

- Growth hormone excess
- Precocious puberty
- McCune–Albright syndrome
- Congenital adrenal hyperplasia (inadequately treated)
- "Testotoxicosis"
- Familial glucocorticoid resistance

Treatment of Children/Adolescents With Tall Stature

Individuals with the constitutional (familial) tall stature do not ordinarily require treatment, as they do not have a pathological condition. Nevertheless, decades ago a number of constitutionally tall girls were treated with high doses of estrogens, perhaps 5-fold the normal replacement dose for a woman, to accelerate epiphyseal closure and diminish adult height. This likely led to diminished adult height, only if started in proximity to the initiation of pubertal maturation. However, this treatment led to rapid pubertal maturation out of synchrony with age peers. Of more clinical relevance, the treatment was associated with highly significant clinical signs—increased thromboembolic events, perhaps a heightened risk of breast cancer, and late negative effects on fertility.[144]

Some males with tall stature were treated with high-dose testosterone administration. The treatment was associated with rapid pubertal maturation, often aggressive behavior, frequent and occasionally painful erections, and very severe acne.

Growth Evaluation of Children With Obesity

Many syndromes and medical conditions associated with obesity are linked to growth disorders and may lead to short stature (majority), and occasionally tall stature. The subsequent emphasis is on obese children without syndromes and medications such as glucocorticoids that lead to both obesity and growth faltering.

The prevalence of obesity in children, adolescents, and adults is ever increasing, but methods for quantifying it have been relatively stable. Obesity, especially severe obesity, affects linear growth and pubertal maturation in youth and cardiometabolic risk in adults.

More specifically, the trends for the BMI indicate a small change at the medians (50th centile), but a major upward change especially above the 85th (overweight) and 95th centiles.[145] Plotted data aggregate between the 95th and 99th+ centiles, but it is difficult to evaluate the trajectory of the weight gain or the response to therapy for these obese subjects. Absolute values change during adolescence, and perhaps a more useful way to present the data is to use the BMI SDS. A single 99th percentile represents a large range of BMI values with a practical upper limit of $Z = +4.0$. However, at the very highest levels, the BMI SDS is weakly associated with other measures of body fatness and comorbidities.[146] These authors recommend that the BMI should be expressed relative to the US Centers for Disease Control and Prevention 95th centile (% BMI_{p95} scale), especially in trials that evaluate interventions.[147] This permits tracking of low- and normal-weight children in addition to those with extreme obesity (Fig. 17.10).

Investigators have used other constructs to aid the graphical presentation of the data. Although not used widely, the triponderal mass index (TMI; weight [kg]/height^3 [m^3]), in contrast to the BMI, shows greater stability during adolescence, so that it may be used as an absolute number rather than as a calculated score.[148] Moreover, the TMI provides a better estimate of body fat compared with DXA estimates and leads to a reduced misclassification of adolescents as overweight when they are of normal weight.

A large epidemiologic study in Korea with almost 10,000 subjects aged 10 to 20 years noted that virtually all factors *associated* with the metabolic syndrome (diastolic blood pressure, excepted) were more likely to be abnormal in those TMI scores in the overweight than in those within the normal range within the same BMI classification. Thus, tracking the TMI z-score may indicate which emerging and young adults should be more closely surveyed for impending cardiometabolic dysfunction.[149]

An additional comorbidity is apparently early (not necessarily precocious) puberty, with some growth and maturational consequences. In a longitudinal study of 9- and 10-year-old girls at baseline,[150] mean age at menarche was earlier in Black (11.7 years) than in White (12.0 years) participants, indicating a significant ethnic/racial gradient. A consequence was that early-maturing girls in both ethnic groups were taller and had a greater (often earlier) PHV than the later-maturing girls; however, at near-adult height, the later-maturing girls were more than 2 cm taller. The latter is a relatively small difference within the normal range, but highly significant given the large sample studied. The differences are likely attributable to estrogens, given the earlier thelarche and menarche and more rapid epiphyseal closure that led to the shorter stature.

Perhaps the easiest to visualize are charts derived from values above the 95th percentile for weight. It does not have an upper bound, and

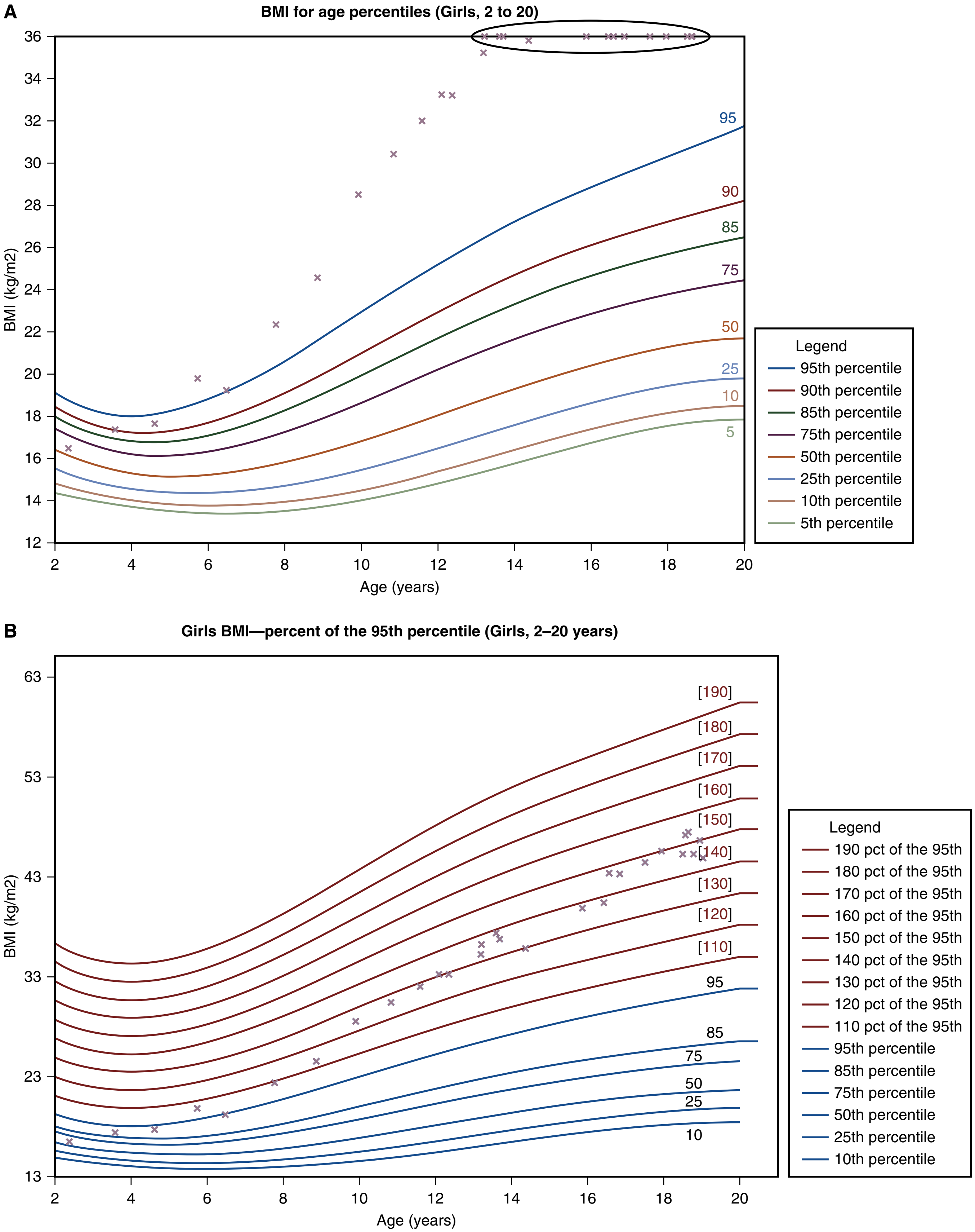


Fig. 17.10 **Clinical tracking of severely obese children. A,** US Centers for Disease Control and Prevention body mass index (BMI) (2000) growth chart, girls aged 2–20 years, percentiles. **B,** Obesity BMI growth chart, girls aged 2–20 years, percent of 95th percentile. *From Centers for Disease Control and Prevention (CDC)*

one can track trajectories over time, especially the effects of therapy. In concept, these differ little from some of the height charts for very short children that show –3 SD to at least –5 SD. All of the very short children are "far below" the –2 SD trajectory, but one can more easily denote response to therapy, e.g., in children with achondroplasia receiving one of the C-type natriuretic peptide analogs.[151]

In general, children with obesity are relatively tall for age, although their weight is in excess for height, leading to the increase in BMI above the 95th centile. Bone age determinations are often advanced, and the children may enter puberty earlier than children of the same age who are of average weight. The temporary increase in height gain in childhood may be compensated by subsequent subnormal height gain during adolescence, and in turn shorter adult stature.

In addition, those with the more severe forms of obesity are at greater risk for multiple cardiometabolic dysfunction, as noted for low high-density lipoprotein cholesterol levels, elevated systolic and diastolic blood pressure, and high triglyceride and glycated hemoglobin levels.[152,153] Although the relative risk is high, the absolute risk of cardiometabolic disease is low in the adolescent; however, these are the emerging and young adults who should be followed prospectively.

The trajectory of BMI throughout childhood and adolescence may also be predictive of risk of cardiometabolic dysfunction in adulthood. In a three-decade longitudinal study of Chinese children beginning at age 6 to 18 years, the high BMI (high trajectory) increasing group (~39%) had relative risk ratios above 3.0 for hypertension, type 2 diabetes, and low density lipoprotein–cholesterol concentration compared with the low BMI (low trajectory) increasing group. From these results one could derive preventive strategies, although caution is warranted. The dataset did not include younger children, especially the time of adiposity rebound, and the region of China in which the study was set had a high salt intake.[154]

KEY POINTS

- Continued weight gain with slowing height velocity is characteristic of children with endocrine causes of obesity: hypothyroidism, growth hormone deficiency, Cushing syndrome, and pseudohypoparathyroidism. On the other hand, high normal height velocity in the presence of excess weight gain is characteristic of the vast majority of children and adolescents with exogenous obesity.

CONCLUSIONS

Growth is a continuous process during childhood and adolescence. Its quantitation permits the clinician and investigator to note whether it is within or outside the broad range of normal variability. Growth disorders are heavily weighted toward individuals with various deficiencies or insufficiencies leading to short stature and falling further behind their peers.

We have discussed the methods of measuring and analyzing growth and maturation with emphasis on skeletal (bone) age; both set the stage for diagnosis, laboratory, and imaging determinations, as well as treatment considerations. Some of the variability in determining SA will diminish as automated radiograph analyses become more available and standardized, and perhaps in using nonradiation methods, such as ultrasound techniques.

The clinical section emphasizes disorders of the GH/IGF-1 axis, although there are likely many more that affect the growth plate directly, and the diagnoses of these will be materially aided by more modern genetic techniques that should become available in commercial laboratories with relatively low cost and quick turn-around time.

Obesity has become a major nutritional problem, especially over the past generation. More consideration is being given to the definition and responses to therapies as new charts are introduced to follow the effects of interventions on the weight trajectory. Most children and adolescents with nonsyndromic obesity are growing at the upper limits of normal with mildly advanced SA, and thus are often relatively tall as children but of more average height as late adolescents, because pubertal maturation is advanced as well.

REFERENCES

1. Malina RM, Bouchard C, Bar-Or O. *Growth, Maturation, and Physical Activity*. 2nd ed. Champaign: Human Kinetics; 2004.
2. Kanaka-Gantenbein C, Mastorakos G, Chrousos GP. Endocrine-related causes and consequences of intrauterine growth retardation. *Ann NY Acad Sci*. 2003;997:150–157.
3. Efstratiadis A. Genetics of mouse growth. *Int J Dev Biol*. 1998;42:955–976.
4. Randhawa R, Cohen P. The role of insulin-like growth factor system in prenatal growth. *Mol Genet Metab*. 2005;86:84–90.
5. Giudice LC, De Zegher F, Gargosky SE, et al. Insulin-like growth factors and their binding proteins in the term and preterm human fetus and neonate with normal and extremes of intrauterine growth. *J Clin Endocrinol Metab*. 1995;8:1548–1555.
6. Argente J, Chowen JA, Perez-Jurado L, et al. One level up: abnormal proteolytic regulation of IGF activity plays a role in human pathophysiology. *EMBO Mol Med*. 2017;9:1338–1345.
7. Fujimoto M, Hwa V, Dauber A. Novel modulators of the growth hormone-insulin-like growth factor axis: pregnancy-associated plasma protein-A2 and Stanniocalcin-2. *J Clin Res Pediatr Endocrinol*. 2017;9:1–8.
8. Araki E, Lipes MA, Patti M-E, et al. Alternative pathway of insulin signaling in mice with targeted disruption of the IRS-1 gene. *Nature*. 1994;372:186–190.
9. Gluckman PD, Cutfield W, Hofman P, et al. The fetal, neonata, and infant environments-the long-term consequences for disease risk. *Early Hum Develop*. 2005;81:51–59.
10. Block T, El-Osta A. Epigenetic programming, early life nutrition and the risk of metabolic disease. *Arteriosclerosis*. 2017;266:31–40.
11. Calkins K, Devaskar SU. Fetal origins of adult disease. *Curr Probl Pediatr Adolesc Health Care*. 2011;41:158–176.
12. Jazwiec P, Sloboda DM. Nutritional adversity, sex and reproduction: 30 years of DOHaD and what have we learned? *J Endocrinol*. 2019;242:T51–T68.
13. Miles HL, Hofman PL, Cutfield WS. Fetal origins of adult disease: a paediatric perspective. *J Endocrinol Metab Disord*. 2005;6:261–268.
14. Lohman TG, Roche AF, Martorell R, eds. *Anthropometric Standardization Reference Manual*. Champaign: Human Kinetics; 1988:1–177.
15. Malina RM. Anthropometry. In: Maud PJ, Foster C, eds. *Physiological Assessment of Human Fitness*. Champaign: Human Kinetics; 1995:205–219.
16. World Health Organization. *Physical Status: The Use and Interpretation of Anthropometry. Technical Report Series, No. 854*. Geneva: World Health Organization; 1995.
17. Klovgaard M, Nielsen NO, Sorensen TL, et al. Growth of children in Greenland exceeds the World Health Organization growth charts. *Acta Paediatr*. 2018;107:1953–1965.
18. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, et al. *CDC Growth Charts: United States. Advance Data from Vital and Health Statistics, No. 314*. Hyattsville: National Center for Health Statistics; 2000.
19. Roche AF. Postnatal physical growth assessment. *Clin Pediat Endoc*. 1999;8:1–12.
20. Ogden CL, Fryar CD, Carroll MD, et al. *Mean Body Weight, Height, and Body Mass index, United States 1962-2002. Advance Data from Vital and Health Statistics, No 347*. Hyattsville: National Center for Health Statistics; 2004.
21. McDowell MA, Fryar CD, Ogden CL, et al. *Anthropometric Reference Data for Children and Adults: United States, 2003-2006. National Health Statistics Reports, No 10*. Hyattsville: National Center for Health Statistics; 2008.

22. Fryar CD, Gu Q, Ogden CL, et al. *Anthropometric Reference Data for Children and Adults: United States, 2007-2010. Vital and Health Statistics, Series 11, No. 252*. Hyattsville: National Center for Health Statistics; 2012.
23. Fryar CD, Gu Q, Ogden CL, et al. *Anthropometric Reference Data for Children and Adults: United States, 2011-2014. Vital and Health Statistics, Series 3, No. 39*. Hyattsville: National Center for Health Statistics; 2016.
24. Novina N, Hermanussen M, Scheffler C, et al. Indonesian national growth reference charts better reflect height and weight of children in West Java Indonesian than WHO growth chart standards. *J Clin Res Pediatr Endocrinol*. 2020;12:410–419.
25. Malina RM, Katzmarzyk PT. Validity of the body mass index as an indicator of the risk and presence of overweight in adolescents. *Am J Clin Nutr*. 1999;70:131S–136S.
26. Freedman DS, Wang J, Maynard LM, et al. Relation of BMI to fat and fat-free mass among children and adolescents. *Int J Obes*. 2005;29:1–8.
27. De Onis M, Garza C, Onyango AW, et al. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *J Nutr*. 2007;137:144–148.
28. Cole TJ, Bellizzi MC, Flegal KM, et al. Establishing a standard definition for child overweight and obesity worldwide: international survey. *Br Med J*. 2000;320:1240–1245.
29. Cole TJ, Flegal KM, Nicholls D, et al. Body mass index cut offs to define thinness in children and adolescents: international survey. *Br Med J*. 2007;335:194–201.
30. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation on Obesity*. Geneva: World Health Organization; 1998.
31. Cole TJ. Seasonality of growth. In: Ulijaszek SJ, Johnston FE, Preece MA, eds. *The Cambridge Encyclopedia of Human Growth and Development*. Cambridge: Cambridge University Press; 1998:223.
32. Roche AF, Himes JH. Incremental growth charts. *Am J Clin Nutr*. 1980;33:2041–2052.
33. Baumgartner RN, Roche AF, Himes JH. Incremental growth tables: supplementary to previously published charts. *Am J Clin Nutr*. 1986;43:711–722.
34. Kelly A, Winer KK, Kalkwarf H, et al. Age-based reference ranges for annual height velocity in US children. *J Clin Endocrinol Metab*. 2014;99:2104–2112.
35. Marshall WA. Evaluation of growth rate in height over periods of less than one year. *Arch Dis Child*. 1971;46:414–420.
36. Tanner JM, Cameron N. Investigation of the mid-growth spurt in height, weight and limb circumferences in single year velocity data from the London 1966-67 growth survey. *Ann Hum Biol*. 1980;7:565–577.
37. Gasser T, Muller HG, Kohler W, et al. An analysis of the mid-growth and adolescent spurts of height based on acceleration. *Ann Hum Biol*. 1985;12:129–148.
38. Sheehy A, Gasser T, Molinari L, et al. An analysis of variance of the pubertal and midgrowth spurts for length and width. *Ann Hum Biol*. 1999;26:309–331.
39. Camp JD, Cilley EIL. Diagrammatic chart showing time of appearance of various centers of ossification and period of union. *AMA J Roentgenol*. 1931;26:905.
40. Acheson RM. Maturation of the skeleton. In: Falkner F, ed. *Human Development*. Philadelphia: Saunders; 1966:465–502.
41. Greulich WW, Pyle SI. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. 2nd ed. Stanford: Stanford University Press; 1959.
42. Todd TW. *Atlas of Skeletal Maturation. Part 1. The Hand*. St. Louis: Mosby; 1937.
43. Tanner JM, Whitehouse RH, Healy MJR. *A New System for Estimating Skeletal Maturity from the Hand and Wrist, with Standards Derived from a Study of 2,600 Healthy British Children*. Paris: International Children's Centre; 1962.
44. Tanner JM, Whitehouse RH, Marshall WA, et al. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. New York: Academic Press; 1975.
45. Tanner JM, Whitehouse RH, Cameron N, et al. *Assessment of Skeletal Maturity and Prediction of Adult Height*. 2nd ed. New York: Academic Press; 1983.
46. Tanner JM, Healy MJR, Goldstein H, et al. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*. 3rd ed. London: Saunders; 2001.
47. Roche AF, Chumlea WC, Thissen D. *Assessing the Skeletal Maturity of the Hand-Wrist: Fels Method*. Springfield: CC Thomas; 1988:1–339.
48. Bouchard C, Malina RM, Hollman W, et al. Relationships between skeletal maturity and submaximal working capacity in boys 8 to 18 years. *Med Sci Sports*. 1976;8:186–190.
49. Bouchard C, Leblanc C, Malina RM, et al. Skeletal age and submaximal working capacity in boys. *Ann Hum Biol*. 1978;5:75–78.
50. Kujawa KI. *Skeletal Maturation in Boys: Comparison of Methods and Relationships to Anthropometry and Strength. Doctoral Dissertation*. Austin: University of Texas at Austin; 1977.
51. Malina RM, Coelho-e-Silva MJ, Figueiredo AJ, et al. Tanner-Whitehouse skeletal ages in male youth soccer players: TW2 or TW3? *Sports Med*. 2018;48:991–1008.
52. Roemmich JN, Blizzard RM, Peddada SD, et al. Longitudinal assessment of hormonal and physical alterations during normal puberty in boys. IV: Predictions of adult height by the Bayley-Pinneau, Roche-Wainer-Thissen, and Tanner-Whitehouse methods compared. *Am J Hum Biol*. 1997;9:371–380.
53. Meredith HV. Change in stature and body weight of North American boys during the last 80 years. In: Lipsitt LP, Spiker CC, eds. *Advances in Child Development and Behavior*. New York: Academic Press; 1963:69–114.
54. Roche AF, Sun S. *Human Growth: Assessment and Interpretation*. Cambridge: Cambridge University Press; 2003.
55. Wellens R, Malina RM, Beunen G, et al. Age at menarche in Flemish girls: current status and secular change in the 20th century. *Ann Hum Biol*. 1990;17:145–152.
56. Hauspie RC, Vercauteren M, Susanne C. Secular changes in growth and maturation: an update. *Acta Paediatr*. 1997;424:20–27.
57. Mul D, Fredriks AM, van Buuren S, et al. Pubertal development in The Netherlands 1965-1997. *Pediatr Res*. 2001;50:479–486.
58. Malina RM. Skeletal age and age verification in youth sport. *Sports Med*. 2011;41:925–947.
59. Dvorak J, George J, Junge A, et al. Age determination by MRI of the wrist in adolescent male football players. *Br J Sports Med*. 2007;41:45–52.
60. Dvorak J, George J, Junge A, et al. Application of MRI of the wrist for age determination in international U-17 soccer competitions. *Br J Sports Med*. 2007;41:497–500.
61. Mentzel H-J, Vilser C, Eulenstein M, et al. Assessment of skeletal age at the wrist in children with a new ultrasound device. *Pediatr Radiol*. 2005;35:429–433.
62. Mentzel H-J, Vogt S, Vilser C, et al. [Abschätzung des Knochenalters mit einer neuen Ultraschallmethode. *Fortschr Röntgenstr*. 2005;177:1699–1705.
63. Khan KM, Miller BS, Hoggard E, et al. Application of ultrasound for bone age estimation in clinical practice. *J Pediatr*. 2009;154:243–247.
64. Płudowski P, Lebiedowski M, Lorenc RS. Evaluation of practical use of bone age assessments based on DXA-derived hand scans in diagnosis of skeletal status in healthy and diseased children. *J Clin Densitom*. 2005;8:48–60.
65. Płudowski P, Lebiedowski M, Lorenc RS. Evaluation of the possibility to assess bone age on the basis of DXA derived hand scans – preliminary results. *Osteoporos Int*. 2004;15:317–322.
66. Heppe DHM, Taal HR, Ernst GDS, et al. Bone age assessment by dual-energy X-ray absorptiometry in children: an alternative for X-ray? *Br J Radiol*. 2012;85:114–120.
67. Gertych A, Zhang A, Sayre J, et al. Bone age assessment using a digital hand atlas. *Computer Med Imag Graph*. 2007;31:322–331.
68. Thodberg HH, Kreilborg S, Juul A, et al. The boneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imag*. 2009;28:52–66.
69. van Rijn RR, Thodberg HH. Bone age assessment: automated techniques coming of age? *Acta Radiol*. 2013;54:1024–1029.
70. US Department of Energy. *Radiation Safety: Americans' Average Radiation ExposureOffice of Civilian Radioactive Waste Management*., DOE/YMP-0337 US Department of Energy; 2000. Available at: www.ymp.gov. Accessed January 2, 2010.

71. Radiological Society of North America. *Safety: radiation exposure in x-ray examinations*; 2009. Available at: www.radiologyinfo.org. Accessed January 2, 2010.
72. Malina RM. Skeletal maturation rate in North American Negro and white children. *Nature*. 1969;223:1075.
73. Malina RM. Skeletal maturation studied longitudinally over one year in American Whites and Negroes 6 through 13 years of age. *Hum Biol*. 1970;42:377–390.
74. Roche AF, Roberts J, Hamill PVV. *Skeletal Maturity of Youths 12-17 Years: Racial, Geographic Area, and Socioeconomic Differences. Vital and Health Statistics, Series 11, No 167. DHEW publication, (PHS) 79-1654*. Hyattsville: Department of Health Education and Welfare, National Center for Health Statistics; 1978.
75. Zhen O, Baolin L. Skeletal maturity of the hand and wrist in Chinese school children in Harbin assessed by the TW2 method. *Ann Hum Biol*. 1986;13:183–187.
76. Ontel FK, Ivanovic M, Ablin DS, et al. Bone age in children of diverse ethnicity. *Am J Roentgenol*. 1996;167:1395–1398.
77. Ashizawa K, Asami T, Anzo M, et al. Standard RUS skeletal maturation of Tokyo children. *Ann Hum Biol*. 1996;23:457–469.
78. Cole TJ, Rousham EK, Hawley NL, et al. Ethnic and sex differences in skeletal maturation among the Birth to Twenty cohort in South Africa. *Arch Dis Child*. 2015;100:138–143.
79. Peña Reyes ME. *[Aplication del metodo Fels para evaluar la maduracion del esqueleto en un grupo de niños de Oaxaca: Estudio comparativo de los métodos Fels y TW2*. Mexico City, Mexico: Instituto Nacional de Antropologia e Historia, Serie Antrolologica Fisica; 1992.
80. Beunen GP, Rogol AD, Malina RM. Indicators of biological maturation and secular changes in biological maturation. *Food Nutr Bull*. 2006;27:S244–S256.
81. Tanner JM. *Growth at Adolescence*. 2nd ed. Oxford: Blackwell; 1962.
82. Reynolds EL, Wines JV. Individual differences in physical changes associated with adolescence in girls. *Am J Dis Child*. 1948;75:329–350.
83. Reynolds EL, Wines JV. Physical changes associated with adolescence in boys. *Am J Dis Child*. 1851;82:529–547.
84. Wu Y, Schreiber GB, Klementowicz V, et al. Racial differences in accuracy of self-assessment of sexual maturation among young Black and White girls. *J Adol Hlth*. 2001;28:197–203.
85. Taylor SJC, Whincup PH, Hindmarsh PC, et al. Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatr Prenat Epidemiol*. 2001;15:88–94.
86. Malina RM, Cumming SP, Rogol AD, et al. Bio-banding in youth sports: background, concept, and application. *Sports Med*. 2019;49:1671–1685.
87. Prader A. Testicular size: assessment and clinical importance. *Triangle*. 1966;7:240–243.
88. Zachman M, Prader A, Kind HP, et al. Testicular volume during adolescence: cross-sectional and longitudinal studies. *Helveta Paediatr Acta*. 1974;29:61–72.
89. Goede J, Hack WWM, Sijstermans K, et al. Normative values for testicular volume measured by ultrasonography in a normal population from infancy to adolescence. *Horm Res Paediatr*. 2011;76:55–64.
90. Joustra SD, van der Plas EM, Goede J, et al. New reference charts for testicular valume in Dutch children and adolescents allow the calculation of standard deviation scores. *Acta Paediatr*. 2015;104:e271–e278.
91. Hauspie RC, Cameron N, Molinari L, eds. *Methods in Human Growth Research*. Cambridge: Cambridge University Press; 2004.
92. Beunen G, Malina RM. Growth and physical performance relative to the timing of the adolescent spurt. *Exerc Sport Sci Rev*. 1988;16:503–540.
93. Mirwald RL, Baxter-Jones ADG, Bailey DA, et al. An assessment of maturity from anthropometric measurements. *Med Sci Sports Exerc*. 2002;34:689–694.
94. Malina RM, Chow AC, Czerwinski SA, et al. Validation of maturity offset in the Fels longitudinal study. *Pediate Exerc Sci*. 2016;28:439–455.
95. Malina RM, Kozieł SM. Validation of maturity offset in a longitudinal sample of Polish boys. *J Sports Sci*. 2014;32:424–437.
96. Malina RM, Kozieł SM. Validation of maturity offset in a longitudinal sample of Polish girls. *J Sports Sci*. 2014;32:1374–1382.
97. Chumlea WC, Schubert CM, Roche AF, et al. Age at menarche and racial comparisons in US girls. *Pediatrics*. 2003;111:110–113.
98. Nicolson AB, Hanley C. Indices of physiological maturity: deviation and interrelationships. *Child Dev*. 1953;24:3–38.
99. Bielicki T. Interrelationships between various measures of maturation rate in girls during adolescence. Stud Phys Anthrop. *(Wrocław)*. 1975;1:51–64.
100. Bielicki T, Koniarek J, Malina RM. Interrelationships among certain measures of growth and maturation rate in boys during adolescence. *Ann Hum Biol*. 1984;11:201–210.
101. Flor-Cisneros A, Roemmich JN, Rogol AD, et al. Bone age and onset of puberty in normal boys. *Mol Cell Endocrinol*. 2006;254–255:202–206.
102. Largo RH, Prader A. Pubertal development in Swiss boys. *Helv Paediatr Acta*. 1983;38:211–228.
103. Largo RH, Prader A. Pubertal development in Swiss girls. *Helv Paediatr Acta*. 1983;38:229–243.
104. Johnston FE. Individual variation in the rate of skeletal maturation between five and eighteen years. *Child Dev*. 1964;35:75–80.
105. Johnston FE. The use of the Greulich-Pyle method in a longitudinal growth study. *Am J Phys Anthrop*. 1971;35:353–358.
106. Buckler JMH. Skeletal age changes in puberty. *Arch Dis Child*. 1984;59:115–119.
107. Marshall WA. Individual variation in rate of maturation. *Nature*. 1969;221:91.
108. Karlberg J. On the construction of the infancy-childhood-puberty growth standard. *Acta Paediatr Scand*. 1989;356:26–37.
109. Root AW. Genetic regulation of adult stature in humans. *J Clin Endocrinol Metab*. 2020;105:e2633–2635.
110. Jee YH, Baron J, Nilsson O. New developments in the genetic diagnosis of short stature. *Curr Opin Pediatr*. 2018;30:541–547.
111. Guo MH, Hirschhorn JN, Dauber A. Insights and implications of genome-wide association studies of height. *J Clin Endocrinol Metab*. 2018;103:3155–3168.
112. Wit JM, Kamp GA, Oostdijk W, et al. Towards a rational and efficient diagnostic approach in children referred for growth failure to the general paediatrician. *Horm Res Paediatr*. 2019;91:223–240.
113. Collett-Sollberg PF, Ambler G, Backeljauw PF, et al. Diagnosis, genetics, and therapy of short stature in children: a growth hormone research society international perspective. *Horm Res Paediatr*. 2019;92:1–14.
114. Sedlmeyer IL, Palmert MR. Delayed puberty: analysis of a large case series from an academic center. *J Clin Endocrinol Metab*. 2002;87:1613–1620.
115. Kärkinen J, Miettinen PJ, Raivio T, et al. Etiology of severe short stature below -3 SDS in a screened Finnish population. *Eur J Endocrinol*. 2020;183:481–488.
116. Sisley S, Trujillo MV, Khoury J, et al. Low incidence of pathology and high cost of screening in the evaluation of asymptomatic short children. *J Pediatr*. 2013;163:1045–1051.
117. Rogol AD. Emotional deprivation in children: growth faltering and reversible hypopituitarism. *Front Endocrinol*. 2020;11:596144.
118. Baumann G. Metabolism of growth hormone(GH) and different molecular forms of GH in biological fluids. *Horm Res*. 1991;36:5–10.
119. Baumann G. Growth hormone heterogeneity in human pituitary and plasma. *Horm Res*. 1999;51:2–6.
120. Dominé S, Dominé HM. The role of acid-labile subunit (ALS) in the modulation of GH-IGF-1 action. *Mol Cell Endocrinol*. 2020;518:111006.
121. Goldenberg N, Barkan A. Factors regulating growth hormone secretion in humans. *Endocrinol Metab Clin N Amer*. 2007;36:37–55.
122. Yakar S, Isaksson O. Regulation of skeletal growth and mineral acquisition by the GH/IGF-1 axis: lessons from the mouse. *GH IGF Res*. 2016;28:26–42.
123. Martha Jr PM, Rogol AD, Veldhuis JD, et al. Alterations in the pulsatile properties of circulating growth hormone concentrations during puberty in boys. *J Clin Endocrinol Metab*. 1989;69:563–570.
124. Chemaitilly W, Cohen LE, Mostoufi-Moab S, et al. Endocrine late effects in childhood cancer survivors. *J Clin Oncol*. 2018;36:2153–2159.
125. Sklar CA, Antal Z, Chemaitilly W, et al. Hypothalamic-pituitary and growth disorders in survivors of childhood cancer: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:2761–2784.

126. Sbardella E, Crocco M, Feola T, et al. GH deficiency in cancer survivors in the transition age: diagnosis and therapy. *Pituitary*. 2020;23:432–456.
127. Laurier V, Lapeyrade A, Copet P, et al. Medical, psychological and social features in a large cohort of adults with Prader-Willi syndrome: experience from a dedicated centre in France. *J Intell Disabil Res*. 2015;59:411–421.
128. Goldsweig B, Kaminski B, Sishaye A, et al. Puberty in cystic fibrosis. *J Cystic Fibro*. 2019;18:S88–S94.
129. Iorgi ND, Allegri JEM, Napoli F, et al. The use of neuroimaging for assessing disorders of pituitary development. *Clin Endocrinol*. 2012;76:161–176.
130. Murray PG, Dattani MT, Clayton PE. Controversies in the diagnosis and management of growth hormone deficiency in childhood and adolescence. *Arch Dis Child*. 2016;101:96–100.
131. Molitch ME, Clemmons DR, Malozowski S, et al. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1587–1609.
132. Grimberg A, DiVall SA, Polychronakos C, et al. Guidelines for growth hormone and insulin-like growth factor-1 treatment in children and adolescents with growth hormone deficiency, idiopathic short stature and primary insulin-like growth factor-1 deficiency. *Horm Res Paediatr*. 2016;86:361:397.
133. Chemaitilly W, Sklar CA. Childhood cancer treatments and associated endocrine late effects: a concise guide for the pediatric endocrinologist. *Horm Res Paediatr*. 2019;91:74–82.
134. Thomas-Teinturier C, Oliver-Petit I, Pacquement H, et al. Influence of growth hormone therapy on the occurrence of a second neoplasm in survivors of childhood cancer. *Eur J Endocrinol*. 2020;183:471–480.
135. Savendhal L, Cooke R, Tidblad A, et al. Long-term mortality after childhood growth hormone treatment: the SAGhE cohort study. *Lancet Diabetes Endocrinol*. 2020;8:683–692.
136. Webb SM, Crespo I, Santos A, et al. Quality of life tools for the management of pituitary disease. *Eur J Endocrinol*. 2017;177:R13–R26.
137. Garcia JM, Biller BMK, Korbonits M, et al. Macimorelin as a diagnostic test for adult GH deficiency. *J Clin Endocrinol Metab*. 2018;103:3083–3093.
138. Kuppens RJ, Bakker NE, Siemensma EPC, et al. Metabolic health profile in young adults with Prader-Willi syndrome: results of a 2-year randomized, placebo-controlled, crossover GH trial. *Clin Endocrinol*. 2017;86:297–304.
139. Höybye C. Growth hormone treatment of Prader-Willi syndrome has long-term, positive effects on body composition. *Acta Paeditra*. 2015;104:422–427.
140. Lauffer P, Kamp GA, Menke LA, et al. Towards a rational and efficient diagnostic approach in children referred for tall stature and/or accelerated growth to the general pediatrician. *Horm Res Paediatr*. 2019;91:293–310.
141. Meazza C, Gertosio C, Giacchero R, et al. Tall stature: a difficult diagnosis? *Ital J Pediatr*. 2017;43:66.
142. Charmandari E, Kino T. Chrousos syndrome: a seminal report, a phylogenetic enigma and the clinical implications of glucocorticoid signaling changes. *Eur J Clin Invest*. 2010;40:932–942.
143. Javaid MK, Boyce A, Appelman-Dijkstra N, et al. Best practice management guidelines for fibrous dysplasia/McCune-Albright syndrome: a consensus statement from the FD/MAS international consortium. *Orphanet J Rare Dis*. 2019;14:139–155.
144. Bruinsma FJ, Rayner J, Venn AJ, et al. Looking back in time: conducting a cohort study of the long-term effects of treatment of adolescent tall girls with synthetic hormones. *BMC Pub Health*. 2011;11:S7.
145. Ogden CL, Fryar CD, Hales CM, et al. Differences in obesity prevalence by demographics and urbanization in US children and adolescents, 2013-2016. *JAMA*. 2018;319:2410–2418.
146. Freedman DS, Butte NF, Taveras EM, et al. BMI z-scores are a poor indicator of adiposity among 2- to 19-year-olds with very high BMIs, NHANES 1999-2000 to 2013-2014. *Obesity*. 2017;25:739–746.
147. Gulati AK, Kaplan DW, Daniels SR. Clinical tracking of severely obese children: a new growth chart. *Pediatrics*. 2012;130:1136–1140.
148. Peterson CM, Su H, Thomas DM, et al. Tri-ponderal mass index vs Body mass index in estimating body fat during adolescence. *JAMA Pediatr*. 2017;171:629–636.
149. Park HK, Shim YS. Distribution of Tri-ponderal mass index and its relation to body mass index in children and adolescents aged 10 to 20 years. *J Clin Endocrinol Metab*. 2020;2020:e826–e834.
150. Biro FM, McMahon RP, Striegel-Moore R, et al. Impact of timing of pubertal maturation on growth in black and white female adolescents: the national heart, Lung, and blood Institute growth and health study J. *An Pediatr*. 2001;138:636–643.
151. Savarirayan R, Irving M, Bacino CA, et al. C-type natriuretic peptide analogue therapy in children with achondroplasia. *N Engl J Med*. 2019;381:25–35.
152. Skinner AC, Perrin EM, Moss LA, et al. Cardiometabolic risks and severity of obesity in children and young adults. *New Engl J Med*. 2015;373:1307–1317.
153. Velfer YA, Phillip M, Shalitin S. Increased prevalence of severe obesity and related comorbidities among patients referred to a Pediatric obesity clinic during the last decade. *Horm Res Paediatr*. 2019;92:169–178.
154. Yuan Y, Chu C, Zheng W-L, et al. Body mass index trajectories in early life is predictive of cardiometabolic risk. *J Pediatr*. 2020;219:31–37.

18

Growth Hormone Deficiency in Children

Mehul T. Dattani and Antonia Dastamani

OUTLINE

Introduction, 278
Epidemiology, 278
Pathogenesis, 279
  Genetic and Structural Abnormalities, 279
  Acquired Defects, 286
  The Hypothalamo-Pituitary-Somatotroph Axis, 289
Clinical Features, 289
  Neonatal Presentation, 289
  Infant And Childhood Years, 290
Diagnosis Of Growth Hormone Deficiency In Childhood, 290
  Guidance Derived From Clinical Assessment, 290
  Principles of Testing, 291
  Diagnosis of Growth Hormone Deficiency, 292
  A Practical Approach to Diagnostic Evaluation, 294
Treatment, 295
  Growth Hormone, 295
  Growth Hormone–Releasing Hormone and Growth Hormone Secretagogues, 298
  Transition to Adult Care, 298
Conclusions, 300

INTRODUCTION

The height of an individual is the culmination of a complex process that results from an interaction between genes, nutritional status, hormonal milieu, and environmental factors.[1] In terms of adult height, fetal growth is critical and has major implications for the ultimate stature of an individual. Birth length is approximately 30% of final height, and with a crown-rump length velocity of 50 to 60 cm/year; this period represents the fastest rate of growth achieved by an individual.[2] This growth is mediated by maternal nutrition and a number of growth factors such as insulin-like growth factor (IGF)-1, IGF2, fibroblast growth factor, epidermal growth factor, transforming growth factors α and β, and insulin. Any compromise in maternal nutrition or in the production of these growth factors is associated with intrauterine growth restriction.

Over the first 2 years, a period of "catch-up" or "catch-down" growth commonly occurs while the infant establishes his/her own growth trajectory, with a marked increase in the correlation between current height and final height ($r = 0.8$) by 3 years of age. As a result, growth along a predictable channel is the hallmark of the healthy child.

Postnatal growth is best described by the infancy-childhood-pubertal model of growth.[1] These three phases are regulated by different components of the endocrine system. During the infancy phase, growth is rapid but at a sharply decelerating rate. Growth at this stage is principally dependent on nutrition, although endocrine factors in the form of the growth hormone (GH)-IGF axis play an increasingly important role during the first year of life. Poor growth may be a manifestation of any underlying illness reflecting a wide variety of genetic, constitutional, and pathologic conditions, of which GH deficiency (GHD) is but one cause.[3] Stature itself is merely an indication of a potential abnormal physical state, and is not a diagnosis. The underlying explanation for an abnormal growth process needs to be sought.

By 4 years of age, average height velocity has declined to 7 cm/year, with a further decline to a rate of 5 to 5.5 cm/year at 8 years of age (Fig. 18.1). The onset of the childhood phase of growth is apparent from the age of 6 months, when there is overlap between the childhood and infancy phases of growth. This childhood growth is dependent mainly on endocrine factors such as GH and thyroxine.[4] The third phase of postnatal growth, the pubertal phase, is dependent upon the normal secretion of GH and sex steroids. It is extremely variable in terms of timing, with marked sexual dimorphism that gives rise to the average difference of 12.5 cm in adult height between the sexes.

Growth hormone is the main mediator of postnatal growth,[4] and virtually any chronic childhood illness will modify secretion.

As such, care needs to be exercised in the evaluation of GH secretion in these situations. Although GHD may be considered a form of IGF deficiency, this approach may be limited. Because GH receptor (GHR) and postreceptor issues and GHD in adults are considered elsewhere, this chapter will focus primarily on GHD as related to disorders of the hypothalamo-pituitary axis in children.

KEY POINTS

- The growth of a child is a dynamic process, and is impacted on by prenatal and postnatal factors. Postnatal growth is best described by the Infancy-Childhood-Pubertal (ICP) model, and is regulated by genetic, environmental and hormonal factors.

EPIDEMIOLOGY

The reported incidence of GHD is, to a large extent, dependent on the criteria employed to establish the diagnosis, and reflects the wide variation in the stringency of diagnostic testing. The incidence of congenital isolated GHD varies between 1:4,000 and 1:10,000 live births, depending on the population studied, while 3% to 30% of cases are familial, suggesting a genetic etiology.[5]

Several large surveys have indicated that approximately 25% of children diagnosed with GHD have an underlying "organic" cause for their condition, such as trauma, central nervous system (CNS) tumors, inflammation, irradiation, or anatomic abnormalities of the hypothalamus or pituitary.[6]

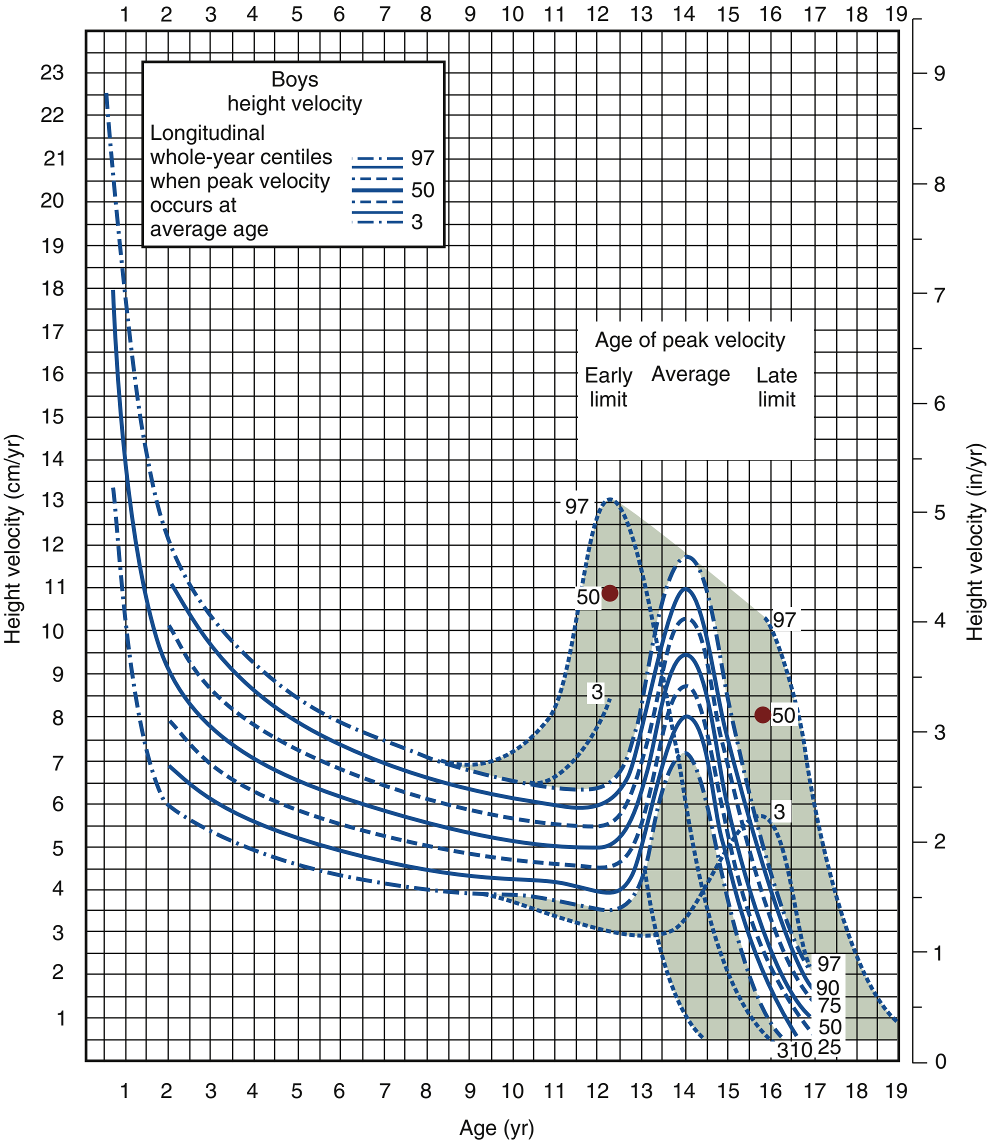


Fig. 18.1 Height velocity chart for boys aged 0–19 years. Centiles 3 to 97 illustrated, with 50th centile in bold. *Shaded zone,* Variation in timing of pubertal growth spurt. Visually, the chart depicts the rapid but rapidly decelerating growth during the first 4 years of life followed by a much slower decline until the onset of the pubertal growth spurt. (Copyright Castlemead Publications.)

The remainder of cases are labeled as "idiopathic" GHD. Such surveys are likely to overestimate the number of true cases of idiopathic GHD because of variation in the diagnosis of GHD. Recent advances in developmental endocrinology suggest that many patients labeled previously as having idiopathic GHD have genetic abnormalities or subtle anatomic abnormalities affecting the hypothalamus, pituitary, or both.

PATHOGENESIS

A list of causes of GHD is provided in Table 18.1. As mentioned already, idiopathic GHD constitutes by far the largest group of patients, although advances in developmental biology are forcing a rethink in this area.

Genetic and Structural Abnormalities

Pituitary Development. The pituitary gland, which consists of anterior, intermediate, and posterior lobes, is a central regulator of growth, metabolism, and development. Its complex functions are mediated via hormone signaling pathways that act to regulate the finely balanced homeostatic control in vertebrates by coordinating signals from the hypothalamus to peripheral endocrine organs (thyroid, adrenals, and gonads). The mature anterior pituitary gland is populated by five neuroendocrine cell types defined by the hormone produced: corticotropes (corticotropin, formerly known as adrenocorticotropic hormone [ACTH]), thyrotropes (thyroid-stimulating hormone [TSH]), gonadotropes (luteinizing hormone [LH], follicle-stimulating hormone), somatotropes (GH), and lactotropes (prolactin [PRL]).[78] The posterior gland secretes vasopressin and oxytocin. The origins of the anterior and posterior lobes of the pituitary gland are embryologically distinct. Rathke's pouch, the primordium of the anterior pituitary, arises from the oral ectoderm, whereas the posterior pituitary derives from neural ectoderm. Development of the anterior gland follows a similar pattern in a number of different species but has been best studied in rodents.[8,9]

In the mouse, anterior pituitary development occurs in four distinct stages: pituitary placode formation; the development of a

TABLE 18.1 Causes of Grwoth Hormone Deficiency

Congenital

Genetic:

- See Table 18.2

Associated with Structural Defects of the Brain:

- Agenesis of the corpus callosum
- Septooptic dysplasia
- Pituitary stalk interruption syndrome
- Holoprosencephaly
- Encephalocele
- Hydrocephalus

Associated with Midline Facial Defects:

- Cleft lip/palate, single central incisor

Idiopathic

Acquired

Trauma:

- Perinatal trauma
- Postnatal trauma

Infiltration:

- Hemochromatosis
- Thalassemia

Infection/inflammation:

- Meningitis/encephalitis
- Tuberculosis
- Sarcoidosis
- Toxoplasmosis

Central Nervous System Tumors:

- Craniopharyngioma
- Pituitary germinoma
- Histiocytosis
- Pituitary adenomas
- Dermoid or epidermoid cysts
- Lipomas
- Teratomas

Following Cranial Irradiation

Following Chemotherapy

Pituitary Infarction

Neurosecretory Dysfunction

Transient:

- Peripubertal
- Psychosocial deprivation
- Hypothyroidism

rudimentary Rathke's pouch; the formation of a definitive pouch; and finally, the terminal differentiation of the various cell types in a temporally and spatially regulated manner (Fig. 18.2). The apposition of Rathke's pouch and the diencephalon, which later develops into the hypothalamus, is maintained throughout the early stages of pituitary organogenesis[10] and appears to be critical for normal anterior pituitary development. A number of signaling molecules—fibroblast growth factor-8 (FGF8),[10,11] bone morphogenetic protein 4,[10] and NKX2.1,[11]—that are expressed in the neural ectoderm and not in Rathke's pouch are thought to play a significant role in normal anterior pituitary development, as illustrated by the phenotype of mouse mutants that are either null or hypomorphic for these alleles. These signaling molecules activate or repress key regulatory genes encoding transcription factors within the developing Rathke's pouch, such as *Hesx1*, LIM homeobox 3 (*Lhx3*), and LIM homeobox 4 (*Lhx4*), that are essential for the subsequent development of the pituitary.[7,8,10]

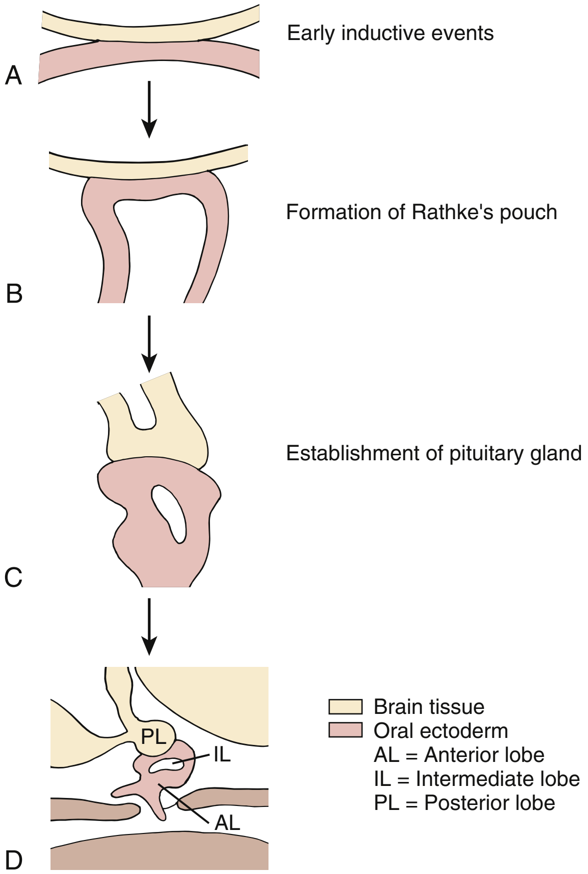


Fig. 18.2 Formation of the pituitary gland. Four-stage process commencing with early inductive events as the infundibulum of the diencephalon abuts the roof of the oral cavity. Pituitary established with signaling gradients generating spatial defined patterns of gene expression and specific cell lineages in the definitive gland. (Reproduced from Valette-Kasic and Enjalbert, *Topical Endocrinology*, February 2003.)

The final stage of pituitary gland development entails the terminal differentiation of the progenitor cells into the distinct cell types found within the mature pituitary gland. This process is tightly regulated by extrinsic factors (Fgf8, Bmp2, Bmp4, and Bmp7) that emanate from the surrounding infundibulum and the juxtapituitary mesenchyme. These then establish gradients of transcription factors (Lhx3, Six3, prophet of Pit1 [Prop1], Pou1f1 (Pit1), Nkx3.1, Islet-1, Lhx4, Six1, Brain-4, and pituitary forkhead).[11] These genetic gradients lead to a wave of cell differentiation. Each of the five anterior pituitary cell types differentiates in a temporally and spatially regulated manner (Fig. 18.3),[12] and this process is dependent upon a number of transcription factors such as Pou1f1, Tpit, and steroidogenic factor 1.[13]

In this multistep process, somatotropes appear by embryonic day (E)15.5 in the anterolateral wings of the developing gland, following the appearance of corticotropes and thyrotropes. The onset of expression of the activator Prop1 by E12.5 and its maintenance until E15.5 is a critical step, as it is required for the emergence of the Pou1f1 lineage

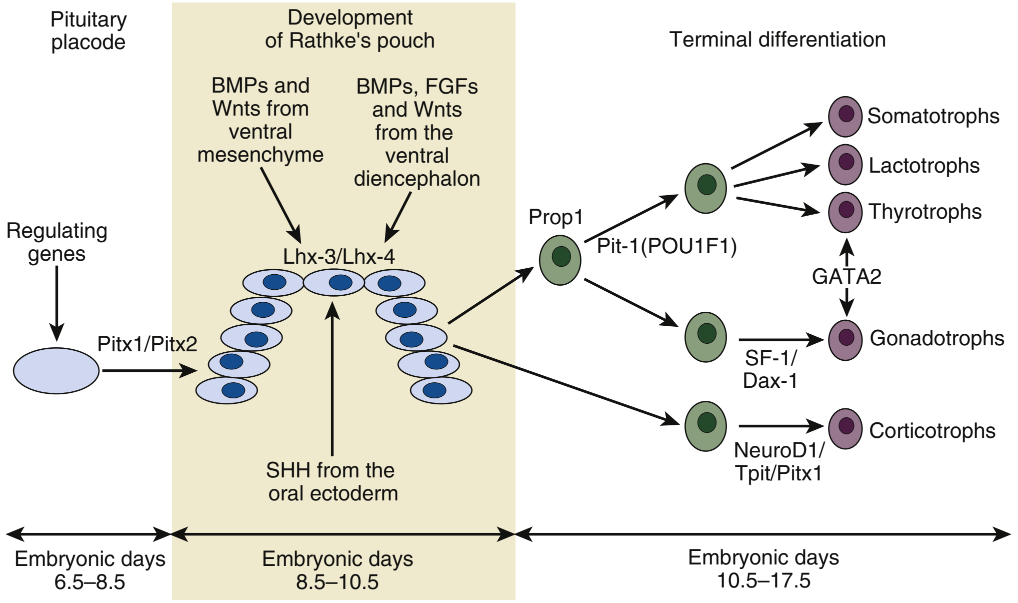


Fig. 18.3 Temporal sequence of events of pituitary development in the mouse. *BMP,* Bone morphogenetic protein; *SHH,* sonic hedgehog; *FGF,* fibroblast growth factor. (Reproduced from Valette-Kasic and Enjalbert, *Topical Endocrinology*, February 2003.)

(somatotropes, lactotropes, thyrotropes). In turn, Pou1f11 is detected by E13.5 and reaches its peak level of expression in the differentiated somatotrophs by E16, and its expression is maintained in adulthood. Activation of Pou1f1 is required for the production of GH, the expression of GH-releasing hormone receptor (GHRHR), and the postnatal expansion of GH-producing cells. The appearance of differentiated somatotropes is followed by a dramatic increase in their number and their migration by E18.5 throughout the central and lateral parts of the anterior lobe.[14] The preferential distribution of somatotropes in this location does not seem to depend on their time of exit from the cell cycle,[15] but is rather a dynamic process during which somatotropes migrate to form homotypic three-dimensional networks.[16] Recent elegant imaging studies reveal that this organization of somatotropes is not merely topographic, but that somatotropes are organized into a functional and structural network that facilitates their coordinated physiological response to different stimuli.[17] This organization into a cellular continuum is maintained postnatally, thus ensuring the plasticity of the system and its adaptation throughout life.[18] Although less is known about pituitary development in humans, it appears to mirror that in the rodent. Spontaneous or artificially induced mutations in the mouse have led to significant insights into human pituitary disease, and identification of mutations associated with human pituitary disease have in turn been invaluable in defining the genetic cascade responsible for the development of this complex structure.[8]

Syndromic Growth Hormone Deficiency in Humans. A number of genetic abnormalities have been identified in children who were previously thought to have idiopathic GHD or combined pituitary hormone deficiency (CPHD) (Table 18.2).[19]

Mutations in early transcription factors may result in GHD in association with other developmental abnormalities, such as septooptic dysplasia (SOD), skeletal defects, or intellectual impairment with or without the development of other pituitary hormone deficiencies.

For instance, mutations within the paired-like homeobox gene *HESX1* are associated with the phenotypes of GHD, CPHD, and SOD, a condition characterized by forebrain, pituitary, and eye abnormalities such as optic nerve hypoplasia.[20]

The inheritance and phenotypes are variable, with both dominant and recessive modes of inheritance described. *HESX1* mutations are an uncommon cause of hypopituitarism, representing less than 1% of cases,[21] and have been described in patients with variable phenotypes without obvious genotype-phenotype correlation. Intriguingly, *HESX1* mutations are classically associated with anterior pituitary hypoplasia with an undescended posterior pituitary and an absent or thin infundibulum.[22] On rare occasions, heterozygous *HESX1* mutations (p.E149K, p.S170L, p.T181A) may be associated with isolated GHD in patients with relatively milder phenotypes compared with the classic manifestations of SOD, who may or not have optic nerve hypoplasia, while an ectopic/undescended posterior pituitary and anterior pituitary hypoplasia are not always present (p.S170L)[21]

Orthodentic homeobox 2 (OTX2) is another early transcription factor that is important for the formation of anterior structures and the forebrain and has been implicated in the etiology of 2% to 3% of anophthalmia/microphthalmia syndromes in humans.[23]

The pituitary phenotype of patients with heterozygous *OTX2* mutations ranges from partial to complete GHD or hypopituitarism with or without an ectopic posterior pituitary on magnetic resonance imaging (MRI). There is no clear genotype-phenotype correlation, even among patients with the same mutation, and, in rare cases, patients may not even exhibit an ocular phenotype (p.N233S).[24,25,26] As there are no long-term follow-up data for patients with heterozygous *OTX2* mutations who present with isolated GHD, careful monitoring for the development of additional endocrinopathies is needed.

Mutations within the LIM-domain genes *LHX3* and *LHX4* are associated with CPHD with extrapituitary manifestations such as a rigid neck (70% of patients), vertebral anomalies, and more or less severe hearing loss (50% of cases) in the case of *LHX3*,[22] and an abnormal cerebellum in the case of *LHX4*.[27] The inheritance of *LHX3* mutations is recessive, whereas that associated with *LHX4* mutations is dominant, unlike the murine phenotype.

To date, all reported patients with heterozygous *LHX4* mutations have had GHD and short stature on presentation, in association with variable endocrine deficits or panhypopituitarism. However, there is remarkable variability in the phenotype even within the same family,

TABLE 18.2 List of Genes with Reported Pathogenic Variants Known to Cause Hypothalamo-Pituitary Disease[13]

Gene with Reported Variants	Phenotype	Mode of Inheritance
ARNT2	CPHD, congenital abnormalities of the kidneys and urinary tract	Recessive
CDON	PSIS	Dominant
EIF2S3	GHD, TSHD, glucose dysregulation, MEHMO syndrome	X-linked
FGF8	HH/KS; HPE	Dominant/recessive
FGFR1	HH/KS, SOD	Dominant
FOXA2	CPHD, HI, childhood-onset diabetes, choroidal coloboma, biliary atresia (cardiac/ endoderm-derived organ abnormalities)	Dominant
GH1	IGHD type IA	Recessive
	IGHD type IB	Recessive
	IGHD type II	Dominant
GHRHR	IGHD type IB	Recessive or dominant (rare)
GLI2	HPE, IGHD/CPHD, polydactyly, single central incisor	Dominant: haploinsufficiency
GPR161	PSIS	Recessive
HESX1	IGHD, CPHD, SOD	Dominant or recessive
IFT172	GHD, retinopathy, metaphyseal dysplasia, renal failure (ciliopathies)	Compound heterozygous
IGSF1	TSHD, hypoprolactinemia, transient GHD; usually with macroorchidism	X-linked
KAL1	HH/KS	X-linked
KCNQ1	GHD, maternally inherited gingival fibromatosis	Dominant
LHX3	CPHD, short neck with limited rotation	Recessive
LHX4	CPHD, Chiari malformation, cerebellar abnormalities, respiratory distress	Dominant or recessive
OTX2	IGHD, CPHD, SOD, anophthalmia/microphthalmia, retinal dystrophy	Dominant: haploinsufficiency or dominant negative
PCSK1	IAD, GHD, TSHD, DI, malabsorption	Dominant, compound heterozygous
PNPLA6	Oliver–McFarlane and Laurence–Moon syndrome; GH and gonadotrophin deficiencies	Recessive
POMC	IAD; early-onset obesity and red hair pigmentation	Recessive
POU1F1	GH, TSH, and ACTH deficiencies	Dominant or recessive
PROKR2	HH/KS	Recessive
PROP1	CPHD, pituitary tumors	Recessive
RAX	Anophthalmia/microphthalmia, CPHD, DI, and cleft palate	Recessive or compound heterozygous
RNPC3	IGHD	Recessive
ROBO1	PSIS	Dominant
SOX2	HH, anophthalmia/microphthalmia, learning difficulties, hypothalamo- pituitary tumors	Dominant
SOX3	GHD, CPHD, absent infundibulum, persistent craniopharyngeal canal	X-linked
TBL1X	TSHD, ASD	X-linked
TBX19	IAD	Recessive
TCF7L1	SOD	Dominant
TRHR	TSHD	Recessive
TSHB	TSHD	Recessive

>*CPHD*, Combined pituitary hormone deficiency; *PSIS*, pituitary stalk interruption syndrome; *GHD*, growth hormone deficiency; *TSHD*, thyroid-stimulating hormone deficiency; *MEHMO*, mental retardation, epileptic seizures, hypogonadism with hypogenitalism, microcephaly, and obesity; *HH*, hypogonadotropic hypogonadism; *KS*, Kallmann syndrome; *HPE*, holoprosencephaly; *SOD*, septooptic dysplasia; *HI*, congenital hyperinsulinism; *IGHD*, isolated growth hormone deficiency; *IAD*, isolated adrenocortical deficiency; *DI*, diabetes insipidus; *ASD*, autism spectrum disorder; *GH*, growth hormone; *TSH*, thyroid-stimulating hormone; *ACTH*, adrenocorticotropic hormone.

ranging from panhypopituitarism to GHD with partial TSH deficiency, or even partial GHD and short stature, diagnosed later in life.[27] Recently, a homozygous mutation in *LHX4* was identified in a pedigree of Pakistani origin with a phenotype characterized by severe anterior pituitary hypoplasia associated with severe early hypopituitarism and neonatal death in all affected siblings.[28]

Homozygous variants of *LHX3* lead to combined pituitary hormone deficiency, displaying consistent somatotrope, thyrotrope, and gonadotrope deficits, and inconsistent corticotrope deficits. On MRI, the pituitary was described as hypoplastic or aplastic in 60% of cases, normal in 10%, and hyperplastic in 30% of cases. There are also studies demonstrating the possibility that patients with a heterozygous variant of *LHX3* may have pituitary deficiencies, with a milder phenotype than patients with homozygous variants. It is thus vital to propose optimal follow-up of such patients who, until now, were considered as not being at risk of presenting pituitary deficiency.[29]

Mutations within the gene encoding the transcription factor Sox2 are associated with hypopituitarism in mouse and humans. *SOX2* is one of the earliest known genes to be expressed in embryonic stem cells and neural progenitors. Although mutations in the mouse are associated with a generalized reduction in all pituitary cell types, in humans the most frequent pituitary defect is hypogonadotropic hypogonadism, with GHD being less frequent. Other features include eye disorders (anophthalmia and microphthalmia), brain malformations (usually hippocampal), intellectual disabilities, esophageal atresia, genital abnormalities, sensorineural hearing loss, and a hypoplastic anterior pituitary.[30] Pituitary masses have been observed in affected patients, but they do not progress.[31]

Studies with *SOX3* suggest a possible explanation contributing to the predominance of males in many series of GHD, as it is now a recognized cause of X-linked or type III GHD. *SOX3* is located on the X chromosome (Xq26–27) and appears to not only be important for pituitary development but also to be associated with mental retardation.[32] These observations of GHD and brain developmental abnormalities are particularly important, because neurodevelopmental handicap has often been ascribed to untreated neonatal hypoglycemia, whereas structural developmental problems may be a more pertinent explanation. Patients with over- or underdosage of *SOX3* present with isolated GHD or combined pituitary hormone deficiencies or normosmic hypogonadotropic hypogonadism, with or without mental retardation or learning difficulties, and an ectopic/undescended posterior pituitary on MRI.[33]

In different pedigrees, large or submicroscopic duplications at the Xq26–27 locus encompassing *SOX3* have been associated with a phenotype. In addition, expansion of the first polyalanine tract of SOX3 by 7 or 11 residues or an in-frame deletion resulting in the loss of six alanine residues have been reported in association with variable phenotypes (isolated GHD or combined pituitary hormone deficiencies).[34] The variability in the size of the polyalanine tract in SOX3 is an uncommon cause of hypopituitarism or isolated GHD, and to date there have been no reported point mutations in *SOX3* leading to functional compromise.

Retina and anterior neural fold homeobox (RAX) is a transcription factor involved in eye and forebrain development. Bilateral microphthalmia or anophthalmia in some patients has arisen from compound heterozygous or recessive *RAX* mutations.[35] A severe clinical phenotype has recently been described in a patient with a homozygous truncating *RAX* mutation.[32] The patient showed anophthalmia, GH, TSH, and ACTH deficiencies with probable gonadotropin deficiency, diabetes insipidus (DI), bilateral cleft lip and palate, micropenis, and absent anterior and posterior pituitary with a rudimentary pituitary stalk.[36]

Roundoabout 1 (ROBO1) is one of the four mammalian genes encoding Roundabout receptors, and is involved in neurogenesis, angiogenesis, and cell proliferation. Heterozygous nonsense or missense *ROBO1* mutations have been described in subjects with pituitary stalk interruption syndrome (PSIS) with various hormone deficiencies and ocular anomalies such as strabismus or ptosis. Slit-Robo signaling is also implicated in mouse heart development, and loss-of-function variants in *ROBO1* have been reported in patients with ventricular septal defect and tetralogy of Fallot. Pituitary stalk duplication, hypoplastic anterior pituitary, and ectopic posterior pituitary have recently been described in two family members with a *ROBO1* deletion.[37]

Aryl-hydrocarbon receptor nuclear translocator 2 (ARNT2) is a member of the basic-helix-loop-helix-Per-Arnt-Sim superfamily of transcription factors. ARNT2 mutations are associated with CPHD involving congenital abnormalities of the kidneys and urinary tract. A homozygous frameshift *ARNT2* mutation has been described in six patients from a highly consanguineous pedigree with CPHD, postnatal microcephaly, frontotemporal lobe hypoplasia, and visual and renal abnormalities.[38]

De novo heterozygous mutations in *forkhead* box A2 (FOXA2) have been reported in syndromic forms of CPHD associated with congenital hyperinsulinism, craniofacial dysmorphism, choroidal coloboma, and lung and gastrointestinal malformations. In addition, cases with childhood-onset diabetes and cardiac malformations have recently been described.[39]

MAGE family member L2 (MAGEL2) is a member of the type II MAGE gene family involved in neurogenesis and brain function. MAGEL2 mutations have been recently described in pedigrees with CPHD associated with arthrogryposis.[40]

Hemizygous missense and frameshift mutations in eukaryotic translation initiation factor 2 subunit 3 (EIF2S3) have been described in patients with MEHMO syndrome, an X-linked disorder characterized by GHD, mental retardation, seizures, hypogonadism, microcephaly, and obesity.[41]

TBC1 domain family member 32 (TBC1D32) is a gene involved in ciliary function and sonic hedgehog (SHH) signaling. Biallelic loss-of-function variants in *TBC1D32* have recently been described in two Finnish siblings with CPHD, absent anterior pituitary, and mild craniofacial dysmorphism, and in a Pakistani child with GHD, anterior pituitary hypoplasia, and developmental delay.[42]

Potassium voltage-gated channel subfamily Q member 1 (KCNQ1*)* is known to be associated with cardiac arrhythmia syndromes. A clinical phenotype characterized by GHD and gonadotropin deficiencies associated with maternally inherited gingival fibromatosis has recently been described in subjects from three unrelated families with missense *KCNQ1* mutation.[43]

Two variants in *transcription factor 7-like 1* (TCF7L1) associated with forms of SOD in humans have been described.[25] The first variant, p.R92P, has been reported in a male infant with partial agenesis of the corpus callosum, thin anterior commissure, right optic tract hypoplasia, and low IGF1 concentrations. The second variant, p.R400Q, has been described in a male infant with bilateral optic nerve hypoplasia, anterior pituitary hypoplasia, absent septum pellucidum, and posterior pituitary bright spot.[44]

Holoprosencephaly (HPE) is a congenital defect of the brain characterized by incomplete cleavage of the embryonic forebrain into left and right hemispheres. It is characterized by variable degrees of defects in the separation of the cerebral hemispheres and ventricles, while failure of the frontal and parietal lobes to divide posteriorly results in an absent corpus callosum. HPE is commonly characterized by midline facial defects, ranging from cyclopia or anophthalmia to midfacial hypoplasia, cleft lip and palate, and a single central incisor.[45] DI is common, but anterior pituitary hormone deficiencies have also been described. An increasing number of genes are implicated in its etiology, including genes encoding members of the SHH signaling pathway such as *SHH*, *ZIC2*, *TGIF1*, *PTCH1*, *GLI2*, *DISP1*, *TDGF1*, *GAS1*, *EYA4,* and *FOXH1*. Heterozygous mutations in *GLI2* have been reported in patients with variable craniofacial abnormalities, isolated GH deficiency (IGHD) or CPHD, and pituitary hypoplasia. In addition, mutations in *FGF8* have been identified as the first genetic cause of autosomal recessive HPE.

More recently, a mutation in *CDON*, a SHH coreceptor and a component of the SHH signaling pathway previously implicated in the etiology of HPE, has been identified in a patient with PSIS with GH, TSH, and ACTH deficiencies, as well as neonatal hypoglycemia and cholestasis.[46] Interestingly, a murine Cdon mutation was associated with optic nerve hypoplasia, and the effect was phenocopied by ethanol administration.[47] A recessive variant in a further SHH component, GPR161, encoding the orphan G protein–coupled receptor 161, has also been reported in a consanguineous family with PSIS.[48] These data suggest that the SHH signaling pathway is critical for normal hypothalamo-pituitary (HP) development.

Compound heterozygosity for mutations in *IFT172* was identified in a patient with early growth retardation, anterior pituitary hypoplasia and an ectopic posterior pituitary, in addition to retinopathy, and hypertension with renal failure. The phenotype was consistent with a ciliopathy,[49] and was the first report of a mutation in this gene associated with a pituitary defect. *IFT172* encodes an intraflagellar transport subcomplex subunit that is essential for ciliary assembly and maintenance. Previously, mutations have been described in skeletal ciliopathies with variable polydactyly, as well as retinal, cerebellar, or hepatorenal malformations.[50] Consistent with these data, Alstro¨m syndrome, a rare autosomal recessive disease, consists of multiorgan dysfunction with GHD and is caused by mutated *ALMS1,* which encodes a protein localizing to centrosomes and basal bodies of ciliated cells.[51] These studies imply a critical emerging role for cilia in HP development.

It is clear that our understanding of the etiology of hypopituitarism and the mechanisms whereby mutations lead to a particular phenotype is rudimentary and evolving. Additionally, many cases of hypopituitarism may be due to changes in regulatory regions of known genes or perhaps within novel genes that have yet to be identified. The role of epigenetic factors cannot be discounted, and could be contributory to those cases of hypopituitarism that are unexplained.

Growth Hormone-Releasing–Hormone and Its Receptor. Because growth hormone–releasing hormone (GHRH) and its receptor (GHRHR; GHD type 1B) are critical to somatotrope population expansion, abnormalities in either are likely to be associated with severe GHD. No mutations of the human *GHRH* gene have been identified to date,[52] but mutations in *GHRHR* have been identified in a number of pedigrees.[53,54] In most cases, patients were of consanguineous pedigrees or from specific ethnic backgrounds.

With one exception,[55] all patients reported to date have been either homozygous or compound heterozygous for mutations of the *GHRHR* gene. Children have severe GHD and short stature but, compared with those with recessive *GH1* mutations, midfacial hypoplasia, neonatal hypoglycemia, and microphallus are not common. Serum GH concentrations fail to rise following standard provocative testing, as well as after GHRH administration. Because of the effect of GHRH on the proliferation of somatotropes, anterior pituitary hypoplasia on MRI is almost an invariable finding; the conflicting reports of apparently normal pituitary morphology in patients with the same *GHRHR* mutation[56] may be explained by the different age groups and lack of age-matched reference standards. However, recently a milder form of GHD was reported in association with a novel partial loss-of-function mutation in *GHRHR*.[57]

Mutations in Late Transcription Factors Affecting Somatotrope Development. A number of mouse models exist in which somatotrope development has been impaired. These include the Ames, the Jackson, and the Snell dwarf mice. A missense point mutation within the *PROP1* gene (S83P) has been shown to be responsible for the Ames dwarf mouse.[58] The phenotype results from a failure of initial determination of the Pit1 lineage required for production of GH, PRL, and TSH. The Ames pituitary gland contains less than 1% of the normal complement of somatotropes and decreased numbers of lactotropes and thyrotropes. In humans, mutations within the transcription factor Prop1 are associated with CPHD in the form of GH, PRL, TSH, and gonadotropin deficiency, while a proportion of individuals with *PROP1* mutations will develop cortisol deficiency.[59] Additionally, a number of individuals with mutations within *PROP1* develop transient pituitary masses that can wax and wane,[60] with subsequent involution (Fig. 18.4).[61] The exact mechanism underlying this phenomenon remains unclear, but it is clearly important to exclude mutations within *PROP1* in patients with pituitary "tumors," especially the nonfunctioning variety. There is considerable variability in the timing of the endocrinopathy: a number of patients will actually commence puberty but then arrest halfway through, others may present in childhood with growth hormone deficiency before the gradual development of TSH, PRL, and gonadotropin deficiency later in life, and in rare cases somatotrope function may be retained, as in the report of a patient who attained normal final height without GH replacement.[59,62] Pit1 (now known as Pou1f1) is a member of the POU family of homeodomain proteins and contains a highly conserved bipartite DNA-binding domain consisting of the POU homeodomain, required for low-affinity DNA binding, and a POU-specific domain, responsible for the specificity of DNA binding and potential interactions with other proteins.[59,63] The Snell dwarf mouse, characterized by pituitary hypoplasia and GH, PRL, and TSH deficiencies, has a point mutation (W261C) within the *Pit1* gene, affecting the third helix of the POU homeodomain. This abrogates binding of Pit1 to its target promoter sequences. Several mutations and deletions of the *POU1F1* gene have been identified in humans with CPHD, characterized by the combination of GH, PRL, and TSH deficiency.[63] Mutations have been described that separately affect the DNA-binding capacity of POU1F1 or its transactivation properties. Autosomal dominant transmission, resulting from a dominant negative effect, has been observed in mutations affecting POU1F1 dimerization or transactivation (p.P24L), or in the relatively common p.R271W mutation, which results in increased binding to promoter elements and disruption of transcriptional activation. Autosomal recessive transmission is found with other mutations, such as p.A172X, p.E250X, p.R143G, p.A158P, and p.P239S. Variability in phenotype has been reported, although most patients exhibit growth retardation during the first year of life.[64] GH and PRL deficiencies are complete; TSH secretion may be observed during infancy but declines progressively during the early months of life. MRI scanning revealed a marked variability in the size of the anterior pituitary, with some patients demonstrating a normal pituitary and others having a hypoplastic pituitary. After appropriate GH and thyroxine replacement, patients appear to enter puberty normally and have normal fertility. Lactation may be impaired. In some patients, TSH secretion may be normal.[64] A recent study has described the phenotype of isolated GHD in association with a heterozygous dominant mutation in *POU1F1*, p.Pro76Leu.[65]

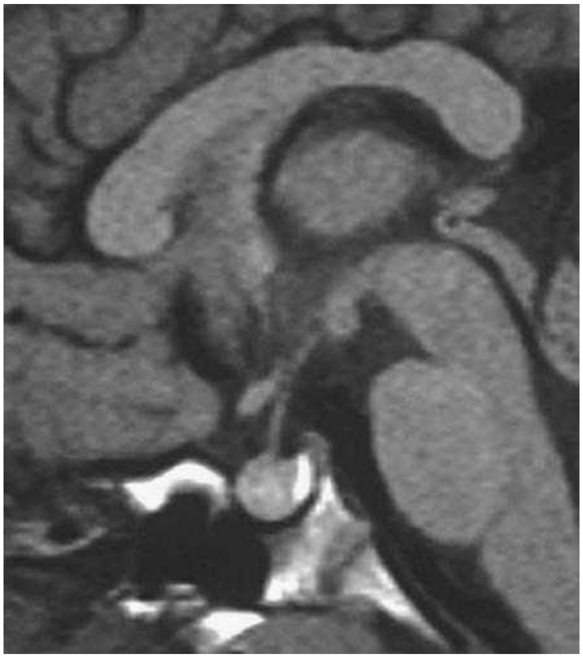


Fig. 18.4 **Pituitary "tumor" in patient with *PROP-1* mutation.** Saggital magnetic resonance imaging scan of the pituitary revealing a large, globular anterior pituitary with normal posterior pituitary enhancement. Subsequent scans revealed involution of the mass and resultant empty sella.

***GH1* Gene.** The *GH1* gene, located at chromosome 17q22-24, is part of a cluster of five structurally related genes: *GH1*, chorionic

TABLE 18.3 **Genetic Abnormalities in GH1**

Type	Inheritance	Gene(s)	Mutations	Phenotype
IA	Autosomal recessive	*GH1* (OMIM #139250)	Deletions (6.7–45 kilobases) Frameshift Nonsense	Undetectable GH, anti-GH antibodies on rhGH treatment
IB	Autosomal recessive	*GH1* (OMIM #139250)	Splice site Missense Nonsense Frameshift	Low, detectable GH; no antibodies on rhGH treatment
II	Autosomal dominant	*GH1* (OMIM #139250)	Splice site Splice site enhancers Missense	Variable short stature (severe to normal height); may have evolving MPHD
III	X-linked recessive	*SOX3* (OMIM #313430) *BTK* (OMIM #300300), unknown	PA tract deletions/expansions Gene deletions/duplications	IGHD or MPHD; ectopic posterior pituitary +/- mental retardation +/- agamma/hypogammaglobulinaemia
IV	Autosomal recessive	*GHRHR* (OMIM #139191)	Splice site Missense Nonsense	Extremely low GH concentrations, good response to rhGH treatment, can be milder phenotype (midfacial hypoplasia, neonatal hypoglycaemia, microphallus less common)
V	Autosomal recessive	*RNPC3* (OMIM #618016)	Missense	Severe postnatal growth failure, undetectable GH, low-normal PRL concentrations, may have multiple pituitary hormone deficiency, high total ghrelin concentrations (normal acylated ghrelin)

OMIM, Online Mendelian Inheritance in Man; *GH*, growth hormone; *rhGH*, recombinant human growth hormone; *IGHD*, isolated growth hormone deficiency; *MPHD*, multiple pituitary hormone deficiency; *PA*, polyalanine; *PRL*, prolactin.

somatomammotropin pseudogene (*CSHP*), chorionic somatomammotropin (*CSH*), *GH2* (or placental variant), and *CSH2*. Mutations within the *GH1* gene are associated with isolated GH deficiency (Table 18.3).[66] Large, recessive, inherited deletions are associated with an absence of GH-protein (type 1A, autosomal recessive GHD). Complete loss of pituitary GH secretion occurs secondary to deletions resulting from nonhomologous crossing over at different sites in the GH and chorionic somatomammotropin gene clusters. The most common deletion is 6.7 kb, but deletions of 7.0, 7.6, and greater than 45 kb have also been observed.[66] Patients typically present with severe growth failure (height standard deviation score [SDS] <–4.5) by the first 6 months of age and have an excellent initial response to GH therapy, but, because of the absence of a normal GH molecule in fetal life, an attenuation of the growth response to exogenous GH may result from the development of anti-GH antibodies, although this event has been described less frequently with newer GH preparations.[67]

Type IB GHD is due to homozygous splice-site, frameshift, and nonsense mutations within the *GH1* gene or homozygous or compound heterozygous mutations within *GHRHR*. Children with type IB GHD present with marked short stature, poor growth velocity, low but detectable GH concentrations, and an excellent response to GH treatment, with no formation of antibodies. In some families, unaffected carriers of the mutant allele may have significantly shorter stature than noncarriers.[67]

Type II GHD is autosomal dominant and associated with splice-site mutations. These mutations lead to the production of two alternatively spliced GH molecules, 20- and 17.5-kD human pituitary-derived GH (hGH). Mutations in an exon splice enhancer within exon 3 of the *GH1* gene have also been associated with autosomal dominant GHD.[68] The generation of the 17.5-kD form of hGH has a dominant negative effect and prevents the secretion of the normal wild-type 22-kD hGH, with a consequent deleterious effect on pituitary somatotropes. In a murine model of this dominant negative mutation, there is loss of somatotroph number[69] and progressive damage to adjacent pituitary cells (with later failure of PRL, TSH, and gonadotropin secretion).[70] In addition, missense mutations (p.R183H, p.P89L, and p.V110F) are also implicated in IGHD type II. These patients have a normal *GH1* allele but are unable to secrete the normal form of GH in appropriate concentrations. The mutant protein therefore exerts a dominant negative effect. As in the mouse, evolution of other hormonal deficiencies, including ACTH, TSH, and gonadotropin deficiencies, have been described in patients with some dominant *GH1* mutations.[71]

GHD type III, an X-linked form of IGHD, has been reported in patients with hypogammaglobulinemia. The role of the transcription factor SOX3 in the development of GHD with/without mental retardation has already been mentioned in a previous section of this chapter.

Bioinactive Growth Hormone Molecule. Because the GH molecule exists in multiple molecular forms resulting from alternative splicing or posttranslational processing, some cases of short stature have been hypothesized to be the consequence of abnormal ratios of the various GH forms. Heterozygous missense *GH1* mutations may impair growth by a complex mechanism that affects not only the secretion of GH, but also its bioactivity. Two individuals heterozygous for point mutations in *GH1* were described by Takahashi et al.,[72] who detailed the biochemistry and molecular genetics. The mutant GH molecules (p.R77C and p.D112G) were capable of binding to the GHR, perhaps even with increased affinity, but were unable to stimulate tyrosine phosphorylation of GH-activated intracellular signaling intermediates in a normal manner. The ability of the p.R77C mutant to behave in a dominant negative manner was demonstrated by its capacity to inhibit the *in vitro* actions of wild-type GH. Subsequently, the heterozygous p.R77C mutation has been reported in patients with growth retardation and delayed pubertal development, who showed normal catch-up growth with recombinant human GH (rhGH) replacement therapy.[73] However, there is no clear phenotype-genotype correlation, as family members of normal stature were also carriers. In this case, patients do not always have short stature, and they may have normal or slightly increased GH secretion and low IGF1 and GH binding protein (GHBP) concentrations. Recent *in vitro* studies did not show any difference between the mutant and wild-type GH molecule in terms of binding to the GHR and activation of the downstream Jak2/Stat5 pathway.[73] However, it is possible that the mutation results in reduced capability to induce the transcription of the genes encoding GHR and GHBP.[60] Missense *GH1* mutations (p.P59L and p.P59S)[74] may result in reduced binding affinity for the GHR, while in some cases (p.R178H) the mechanism of action is more complex, with the mutation affecting both GH and GH binding affinity for the GHR, resulting in reduced activation of the downstream signaling pathway.[75]

RNPC3 and GHD. Compound heterozygous mutations in *RNPC3*, encoding the RNA-binding region containing 3, a 65-kD protein that is a component of the U12-dependent spliceosome, have been associated with the IGHD phenotype. Both mutations, namely a missense mutation (p.P474T) and a nonsense mutation (p.R502X), were shown to impair the binding of the U11/U12-65K protein encoded by *RNPC3* to U12 and U6atac snRNAs. The two mutations therefore reduce the formation of the U12-type intron recognition complex, leading to defects of splicing in a subset of minor introns.[76] A more recent study described the presence of compound heterozygosity for two variants in *RNPC3*, namely c.443G>C, p.(G148A), and c.259C>T, p.(Q87X), in three siblings from a Caribbean pedigree. All three affected children manifested a phenotype characterized by CPHD (GH, TSH, and PRL deficiencies), as well as learning difficulties and congenital cataracts.[77]

Structural Abnormalities. In addition to the structural abnormalities associated with the genetic problems described above, GHD can occur in the setting of other cranial or midline abnormalities such as holoprosencephaly, nasal encephalocele, single central incisor, and cleft lip and palate.

As methods of radiologic evaluation of the CNS have improved, an increasing percentage of patients with idiopathic GHD have been identified to have structural abnormalities.[78] Many of these are associated with some of the genetic abnormalities described above, but the findings are worthy of separate consideration. In particular, earlier studies showed that the finding on MRI of an undescended (frequently called "ectopic") posterior pituitary (posterior pituitary ectopia [PPE]) was commoner in males than females (3:1 when PPE present vs. 1:1 if normal anatomy), in patients with CPHD as compared with IGHD (49% vs. 12%), breech delivery (32% vs. 7%), and associated congenital brain anomalies (12% vs. 7%).[79]

These findings appear to be best explained by a defect in induction of the mediobasal structure of the brain in the early embryo rather than the product of birth trauma, as previously suggested. Whether pituitary insufficiency is the result of hypothalamic or pituitary dysgenesis, or the product of hypoplasia or sectioning of the pituitary stalk, is not always clear. Perinatal problems, however, including breech presentation, may prove to be the *consequence* rather than the *cause* of underlying CNS abnormality. The concept that PPE, stalk section or hypoplasia, and pituitary hypoplasia may represent abnormal embryonic development rather than the consequences of birth trauma is supported by the finding of similar anatomic abnormalities in patients with SOD, type I Arnold-Chiari syndrome, and holoprosencephaly, and increasingly in patients with mutations in the genes controlling pituitary development.

In the empty sella syndrome, abnormalities of the sellar diaphragm allow herniation of the suprasellar subarachnoid space into the region of the sella turcica.[80] This may result in damage to the sella, including the pituitary. Empty sella syndrome may be the consequence of surgery or irradiation or may be idiopathic. It is often found in patients with mutations in *PROP1*, when it may have been preceded by a pituitary mass.

Acquired Defects

A wide range of destructive lesions involving the hypothalamus or pituitary may present with isolated GHD or CPHD. Birth trauma, associated with abrupt delivery, prolonged labor, or extensive use of forceps, has been associated frequently with subsequent hypothalamic or pituitary dysfunction.[81]

Destructive Lesions of the Hypothalamus and Pituitary. An increased incidence of GHD has been reported in breech deliveries, although it is still unclear whether such deliveries lead to the acquisition of pituitary dysfunction or, on the other hand, whether preexisting CNS abnormalities result in higher rates of abnormal birth presentations.

Tumors. CNS tumors are an important cause of isolated GHD and CPHD and must be excluded in every child with GHD who does not have an obvious alternative explanation for growth failure.

Midline brain tumors include germinomas, meningiomas, gliomas, colloid cysts of the third ventricle, ependymomas, and optic nerve gliomas. GHD or CPHD may also occur from local extension of tumors affecting the head or neck, such as craniopharyngeal carcinomas and lymphomas.

The major pediatric tumor involving the pituitary is the craniopharyngioma, which is probably an evolving congenital malformation that develops from remnants of Rathke's pouch.[82] It accounts for 5% to 15% of intracranial tumors in childhood and 80% of tumors in the HP region. Arising from rests of squamous cells at the embryonic junction of the adenohypophysis and neurohypophysis, it forms an enlarging cyst filled with degenerating cells, leading to cyst fluid or calcification but never to malignant degeneration (Fig. 18.5). These calcifications may be seen at times on skull films and constitute an important diagnostic sign. Although craniopharyngiomas represent the consequences of a congenital malformation, they may present clinically at any age. Significant growth failure may be observed in 30% to 50% of children at the time of diagnosis,[83] but patients most commonly present with complaints of increased intracranial pressure, such as headaches, vomiting, and oculomotor disturbances; visual field defects are frequently noted at the time of diagnosis. Deficiency of at least one pituitary hormone, most commonly GH or gonadotropin, is present in 50% to 80% of patients. DI is reported in 25% to 50% of patients at diagnosis.[82,83] The variation in the reported incidence of DI may be due to the fact that incidence of DI is either underestimated or may be masked by the simultaneous presence of ACTH deficiency.

The commonest type of craniopharyngiomas in childhood is adamantinomatous craniopharyngioma (ACP). Activating mutations in the gene encoding β-catenin (*CTNNB1*), a component of the Wnt signaling pathway, have long been identified in ACPs.[84] A murine model that expresses a degradation-resistant mutant form of β-catenin in early progenitors of Rathke's pouch has furthered our understanding.[85] Most mutant mice die perinatally, by 4 weeks of age, and those who survive exhibit pituitary hyperplasia and marked hypopituitarism with severe disruption of the differentiation of the POU1F1 lineage and extreme growth retardation. Ultimately, all animals develop lethal pituitary tumors that closely resemble human ACPs. The tumorigenic effect of the activated mutant β-catenin is observed only when it is expressed in undifferentiated progenitors, demonstrating that mutated β-catenin in pituitary progenitor/stem cells has a causative role in the etiology of murine tumors resembling human ACPs.[85]

Pituitary adenomas, dermoid or epidermoid cysts, lipomas ,and teratomas are less frequent causes of GHD in childhood.[86]

Langerhans cell histiocytosis (LCH) may also present at any age. LCH is characterized by clonal proliferation and accumulation of

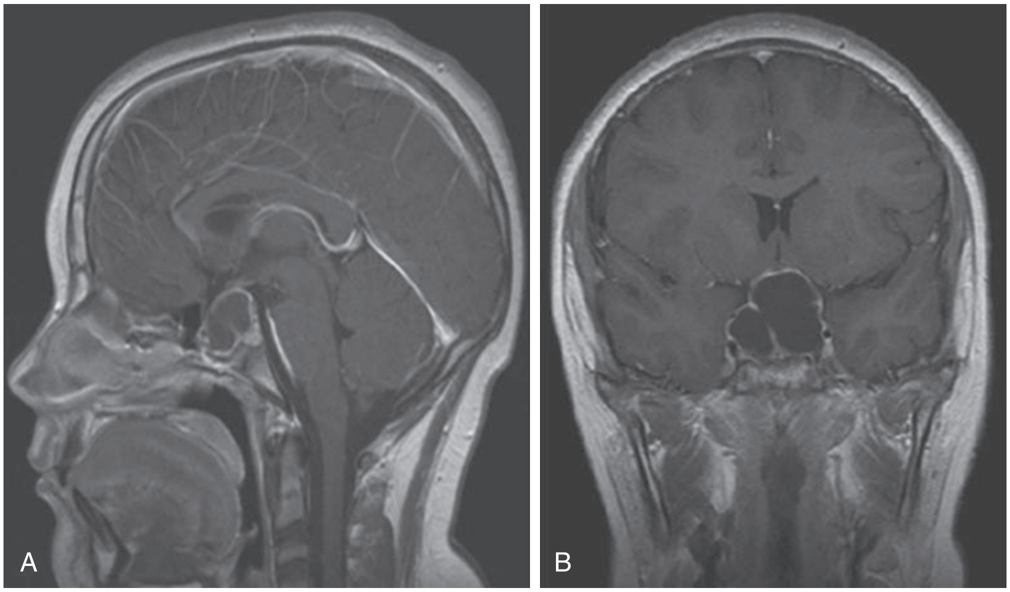


Fig. 18.5 **Magnetic resonance imaging (MRI) of cystic craniopharyngioma.** Saggital MRI scan (**A**) revealing a large, multicystic craniopharyngioma arising from the pituitary fossa and extending up to hypothalamus. Coronal section (**B**) of the same lesion delineating upward and lateral spread. Both images are T1-weighted, gadolinium-enhanced scans.

abnormal dendritic cells that can affect either a single site or many systems, causing multiorgan dysfunction. In children, the median age of diagnosis ranges between 1.8 and 3.4 years.

LCH infiltrates the HP area in 15% to 35% of patients, with subsequent development of at least one pituitary hormone deficiency.[87]

In a multicenter French national study of 589 pediatric patients with LCH, 145 patients (25%) had pituitary dysfunction. In 60 patients, pituitary involvement was already present at the time of diagnosis, and in 20 of them it was the first manifestation of the disease. Patients at high risk of pituitary involvement seem to be those with multisystem disease involving the skull and facial bones, mastoid, sinuses, and mucous membranes (i.e., gums, ear, nose, and throat region). Furthermore, compared with patients without pituitary involvement, patients with pituitary involvement have a higher rate of relapse (10% at 5 years vs. 4.8% at 5 years) and a higher incidence of neurodegenerative LCH.[88]

DI is the most frequently reported permanent consequence of LCH and the commonest endocrinopathy; almost all patients with pituitary involvement have DI. The second commonest endocrinopathy is GHD, which occurs in 14% of all patients with LCH and in more than 40% of patients who have pituitary involvement. In the vast majority of patients, GHD is associated with DI, with a median interval of 2.9 to 3.5 years between the diagnosis of DI and development of GHD. Isolated GHD, or the association of GHD with other anterior pituitary hormone deficiencies, occurs less commonly.

Pituitary MRI findings in patients with LCH include thickening of the pituitary stalk, suggestive of the infiltrative process enhancing changes in the pituitary gland and hypothalamus, and absence of the bright signal of the posterior pituitary in T1-weighted images, caused by the loss of the phospholipid-rich antidiuretic hormone secretory granules. The latter is an invariable feature of patients who develop DI.[89] Although at the time of diagnosis of DI 75% show a thickened pituitary stalk, only 24% have persistent stalk thickening after 5 years. These changes are variable and do not correlate with treatment or with clinical recovery; DI persists in all cases.

Long-term follow-up of patients with LCH has shown that the already established hormone deficiencies cannot be reversed by treatment.[89] Recently, however, isolated case reports have suggested that treatment with the purine analogue 2-chlorodeoxyadenosine (2-CDA) may reverse established DI. Subsequent studies of this form of therapy, used in refractory cases of LCH involving the CNS, showed that 2-CDA may result in partial or complete radiologic improvement of the mass lesion, but the endocrine consequences of the disease, including DI and panhypopituitarism, do not reverse.[90] Patients treated with the JLSG-96 protocol who have been followed up for 5 years developed DI with an incidence of 3.1% to 8.9%, depending on the extension of the disease (single system multisite vs. multisystem).[91]

Radiotherapy used for the treatment of LCH is within the dose range of 10 to 15 Gy, which is known to be unlikely to cause GH insensitivity. However, radiotherapy has been associated with an increased risk of GHD despite the use of a dose less than 15 Gy, a finding that may reflect the severity and extent of the disease rather than the direct effect of radiotherapy.

Irradiation of the Central Nervous System. Cranial irradiation used for the therapy of solid brain tumors and as prophylaxis for leukemia can lead to abnormal HP function. The sensitivity of the HP axis to radiation depends upon the dose, the fractionation, the tissue location, and the age of the patient.[92] Such damage is typically difficult to assess precisely, because the hypothalamus and pituitary may differ in the extent of involvement, and the loss of function may evolve with time.

Sensitivity to CNS radiation may differ among patients, although the majority of children will experience some degree of hypothalamic or pituitary dysfunction within 5 years of receiving 30 Gy. GHD also occurs with doses of 18 to 24 Gy,[93] and subtle dysfunction may be observed at even lower doses.

GH secretion generally appears to be the most sensitive to irradiation, followed by TSH, gonadotropins, and finally ACTH. This may relate to the unique position of the GHRH neurons on the surface of the hypothalamus and not deep within the structure, as previously thought.[94]

Pituitary dysfunction evolves over several years following irradiation, so such children should be monitored for growth deceleration.

Provocative GH testing may be within normal limits, but measures of spontaneous GH secretion frequently demonstrate abnormalities.

Serum concentrations of IGF1 or insulin-like growth factor-binding protein-3 (IGFBP-3) may not be reduced in the early years following cranial irradiation.[95]

Cranial irradiation may also result in precocious puberty, leading to an early pubertal growth spurt, advanced skeletal maturation, and ultimately reduced stature. This may be superimposed upon any growth restriction that results from the spinal irradiation for the primary problem.[96] Low-dose irradiation is frequently associated with a precocious onset of puberty; higher doses may result in gonadotropin deficiency and pubertal delay. In the irradiated child with early puberty, therapy with gonadotropin-releasing hormone analogs (GnRHas) should be considered, with or without GH treatment, to delay epiphyseal fusion.

Lower doses of radiation (24 Gy) are also associated with GHD in approximately 30% to 60% of cases. Craniospinal irradiation used in the treatment of posterior fossa tumors and total-body irradiation used in conditioning regimens for bone marrow transplant are also associated with damage to the epiphyses, with subsequent disproportionate short stature.

Traumatic Brain Injury. Traumatic brain injury (TBI) has been recognized as a cause of acquired hypopituitarism in a number of adult studies. Data on pediatric patients are sporadic, but TBI is probably underdiagnosed.[97]

These effects may be significant, considering the scale of the problem. In the United Kingdom, 180 children per 100,000 population per year sustain a head injury, with 5.6 per 100,000 requiring intensive care and almost one third of those admitted to the intensive care unit undergoing neurosurgery.

Although the pituitary gland is protected within the bony cavity of the sella turcica, the rich vascular network of the hypothalamus and pituitary and the structure of the pituitary stalk make it vulnerable to the effects of TBI. The hypothalamus and pituitary have a complex vascular supply consisting of an arterial supply via the superior and inferior hypophyseal arteries from the internal carotid artery, as well as long hypophyseal vessels and a rich network of portal capillaries that surround the pituitary and infundibulum. The pathophysiology of hypopituitarism related to TBI is not clearly defined, but it is thought that it is the result of direct trauma or of vascular injury resulting in ischemia and infarction,[98] an observation supported by the anatomical findings of autopsies following head trauma, which include anterior lobe necrosis, pituitary fibrosis, hemorrhage, infarction, or necrosis of the pituitary stalk.[99]

Hormone deficiencies may be identified in the first days to weeks posttrauma (acute phase) or may develop over time (late effect). Because there is overlap between the symptoms and signs of hypopituitarism and those of neurologic/psychologic sequelae of TBI, it is possible that late-evolving or partial deficiencies can remain undiagnosed for extended periods.

In the acute phase, alterations in the endocrine function may reflect an adaptive response to acute illness. The clinically significant alterations involve mainly the regulation of fluid and electrolyte balance (DI, syndrome of inappropriate secretion of antidiuretic hormone, cerebral salt wasting) and the hypothalamo-pituitary-adrenal axis. Most of the pituitary hormone changes observed in the acute phase are transient, and their development cannot predict the development of permanent hypopituitarism.[100]

Pituitary hormone deficiencies present in the acute phase are usually transient, but they may persist or appear and evolve over time. In adults, the incidence of permanent hypopituitarism ranges between 23% to 69%, depending on the study. The GH axis is the most frequently affected (10% to 33%), followed by the gonadal (8% to 23%), adrenal (5% to 23%), and thyroid (2% to 22%) axes. The prevalence of permanent DI varies between 0% and 6%. In children with a TBI diagnosis there was a 3.22 times higher risk of a subsequent central endocrine diagnosis compared with the general population (± 0.28).[101]

Until recently, there were only sporadic reports of hypopituitarism following TBI in children, but prospective studies designed to address the problem in the pediatric and adolescent population are in progress. The incidence of hypopituitarism is reported to range from 10% to 60%, and although this is lower in children as compared with adults, it is not uncommon.[102] In general, the long-term outcome of TBI seems to be more favorable in children, although quality-of-life issues and minor disability may persist. The extent to which endocrine dysfunction contributes to these outcomes has yet to be defined.

GHD appears to be the main endocrine manifestation, followed by gonadotropin deficiency. GHD can present as growth failure, whereas delayed or arrested puberty and secondary amenorrhea may present in adolescents and in patients in the transition phase. In a number of case reports, central precocious puberty has also been described in association with head injury, presenting 0.4 to 1.6 years after the event.[103]

Patients with hypopituitarism after head injury may have no clinical signs and symptoms suggestive of this disorder; its correct identification requires a high degree of suspicion. An international consensus guideline on the screening of adult patients post-TBI suggests that all patients who had TBI, regardless of its severity, should undergo baseline endocrine evaluation 3 and 12 months after the event or discharge from the intensive treatment unit.[104]

With reference to children, there are still not enough data from prospective studies to recommend a systematic approach for screening patients after TBI.[105] Even more controversial is the issue of recommending treatment with rhGH post-TBI to improve cognitive outcome and quality of life[106] or in patients with evidence of abnormal GH secretion on provocation testing who have normal growth.[105]

Infiltrative and Inflammatory Disorders. Infiltrative diseases are uncommon causes of GHD in the pediatric population, but pituitary insufficiency may be observed secondary to CNS involvement in tuberculosis, sarcoidosis, or toxoplasmosis. Inflammation associated with bacterial, viral, fungal, or parasitic disease may also result in HP dysfunction.[107] Lymphocytic hypophysitis has also been reported.[108]

A recently reported rare cause of acquired hypopituitarism in adults with GH, TSH, and PRL deficiency results from insults specific to the somatotrope cell lineage, associated with autoimmunity to Pit-1 (POU1F1), positive serum anti-Pit1 antibodies, and histologically a marked reduction of pituitary somatotropes with infiltration by lymphocytes and plasma cells.[109] It is as yet unknown if this may also be a cause of acquired GHD in children and adolescents with autoimmune disorders.

Thalassemia is a hereditary disorder characterized by quantitative defects in synthesis of globin chains that result in ineffective erythropoiesis and, in its more severe forms, transfusion dependence. The majority of complications are the consequence of the toxic effects of iron, which is deposited in organs of the reticuloendothelial system, the heart, and all target organs of the endocrine system, including the pituitary. The anterior pituitary is very sensitive to iron overload, resulting in defective GH secretion, reduced responsiveness of GH to GHRH, and hypogonadotropic hypogonadism. The gonadotrope cells seem to be particularly vulnerable to the toxic effects of iron deposition, which may be related to the way iron is transported in cells. Failure of pubertal development and growth impairment are the most prominent endocrine complications, and may occur despite early initiation of chelation therapy. It is estimated that 56% of thalassemic patients have at least one endocrinopathy; almost half have hypogonadism (40% to 59%), and 33% to 36% manifest growth failure.[110] Of

202 thalassemic patients below the age of 18 years, the reported endocrine complications included GHD in 4.5%, latent hypocortisolism in 4.4%, and central hypothyrodisim in 0.5%.[111]

The growth impairment is the result of a number of factors that include chronic anemia and tissue hypoxia, overchelation due to the toxic effects of desferrio-xamine on spinal cartilage, GH insufficiency, and possible GH insensitivity.

Vascular Lesions. Aneurysms may behave as space-occupying lesions and cause hypothalamic or pituitary destruction.[112]

Psychosocial Dwarfism. Psychosocial dwarfism is a form of poor growth associated with bizarre eating and drinking behavior, social withdrawal, delayed speech, and on occasion other evidence of developmental delay. Periodic hyperphagia is associated with decreased GH responsiveness to standard provocative stimuli but also with subnormal responses to exogenous GH therapy. Removal from the stressful environment, which usually involves removal from the home, is accompanied by a restoration of normal GH secretion, typically within weeks, and a period of catch-up growth.[113]

The mechanisms for this reversible form of GHD are unclear, but it is of note that a variety of psychiatric conditions in adults may be associated with decreased spontaneous and provocative GH secretion.

Establishing the diagnosis of psychosocial dwarfism requires documentation of catch-up growth and restoration of normal GH secretion following correction of the environmental situation.

The Hypothalamo-Pituitary-Somatotroph Axis

GH is secreted by somatotropes in the anterior pituitary gland. The secretory pattern is pulsatile, with discrete pulses of GH every 3 to 4 hours and virtually undetectable GH concentrations in between. Secretion of GH varies considerably with age and shows a sexually dimorphic pattern, with a greater average daily GH output in women. This pattern is the result of an interaction between the hypothalamic peptides GHRH and somatostatin (SS). The amplitude of the GH peak is determined by GHRH, which stimulates the pituitary somatotrophs to increase both the secretion of stored GH and GH gene transcription. SS determines trough levels of GH by inhibition of GHRH release from the hypothalamus and GH release from the pituitary. Withdrawal of SS, on the other hand, determines the timing of a GH pulse.

More recently, the use of synthetic GH-releasing peptides (GHRPs) has led to the identification of a GH secretagogue (GHS) receptor (GHS-R type 1a). The receptor is strongly expressed in the hypothalamus, but specific binding sites for GHRP have also been identified in other regions of the CNS and peripheral endocrine and nonendocrine tissues in both humans and other organisms.[114] The endogenous ligand for the GHS receptor, ghrelin, was isolated from the stomach and is an octynylated peptide consisting of 28 amino acids. It is expressed predominantly in the stomach, but lower amounts are present within the bowel, pancreas, kidney, immune system, placenta, pituitary, testis, ovary, and hypothalamus. Ghrelin not only leads to the secretion of GH but also stimulates PRL and ACTH secretion. Additionally, it influences endocrine pancreatic function and glucose metabolism, gonadal function, appetite, and behavior.[114] It can also control gastric motility and acid secretion and has cardiovascular and antiproliferative effects.[114] The role of endogenous ghrelin in normal growth during childhood remains unclear. Both ghrelin and GHRPs release GH synergistically with GHRH.[114]

The expression of the human *GH1* gene is regulated not only by a proximal promoter, but also by a locus control region (LCR) 15 to 32 kb upstream of the *GH1* gene. The LCR confers pituitary-specific, high-level expression of GH.[115] The full-length transcript from the *GH1* gene encodes a 191–amino acid, 22-kD protein that accounts for 85% to 90% of circulating GH. Alternative splicing of the mRNA transcript generates a 20-kD form of GH that accounts for the remaining 10% to 15%. Within both the proximal promoter and the LCR are located binding sites for the pituitary-specific transcription factor Pit1. Therefore, the LCR region may be considered as initial "entry" point for the recruitment of the tissue-specific Pou1f1 (Pit-1), which in turn triggers chromatin remodeling and recruitment of coactivators.[116] Additional binding sites for the transcription factor Zn15 are also located within the proximal promoter.

In the circulation, GH binds to two binding proteins, high-affinity GHBP and low-affinity GHBP.[117] Little is known about the low-affinity GHBP, which accounts for approximately 10% to 15% of GH binding, with a preference for binding to 20-kD hGH. The high-affinity GHBP is a 61-kD, glycosylated protein that represents a soluble form of the extracellular domain of the GHR that can bind to both 20- and 22-kD hGH and thereby prolong the half-life of GH. The GHR is present in a number of tissues. GH binding induces a conformation change of constitutively dimerized GHRs by rotation, activating a receptor-associated tyrosine kinase JAK2 that in turn is autophosphorylated and also phosphorylates the GHR. This then leads to signal transduction using the MAPK, STAT, and PI3 kinase pathways. The end result is activation of a number of genes that mediate the effects of GH. These include early-response genes encoding transcription factors such as c-Jun, c-Fos, and c-Myc— implicated in cell growth, proliferation, and differentiation—and IGF1, which mediates the growth-promoting effects of GH.[118] IGF1 and IGF2 are single-chain polypeptide hormones that are widely expressed. Together with a family of specific binding proteins, they are believed to mediate most of the actions of GH.[119]

KEY POINTS

- The development of the pituitary is a highly complex process dependent upon a number of genetic factors encoding transcription factors as well as signalling molecules. Mutations in several of these genes are associated with various congenital forms of GHD including syndromic and non-syndromic forms of hypopituitarism as well as isolated GHD. The inheritance, penetrance and phenotypes associated with these mutations are highly variable. Acquired forms of GHD are caused by tumours and their treatments, inflammation, infiltration, and rarely, trauma.

CLINICAL FEATURES

Neonatal Presentation

Recent studies in humans and in animal models have demonstrated marked similarities, but also critical differences, between the clinical features of GHD and various forms of IGF deficiency.[4,120] In GHD, prenatal growth is near normal, although mild reductions in birth length and weight have been observed.

GHD does not cause severe intrauterine growth restriction, whereas loss of placental GH does. However, loss of IGF1 *in utero* results in severe intrauterine growth restriction in both humans and mice, suggesting that IGF1 and the IGF1 receptor are critically involved in intrauterine growth.[121] IGF1 synthesis and secretion *in utero* are not regulated primarily by pituitary GH.

IGF1 production comes under GH regulation either in the last few months of fetal life or shortly after birth, and is well-established by 6 months of age. Growth failure is greater for skeletal growth than for body weight, so infants and young children have an appearance of relative adiposity. Neonates may present with hypoglycemia, and this suggests the possibility of other pituitary hormone deficiencies, especially ACTH. Normoglycemia is only maintained when cortisol replacement therapy is commenced, suggesting that ACTH (and consequently cortisol) secretion is critical for glucose homeostasis. However, the

GH-IGF1 axis also plays a role in maintaining glucose homeostasis, although IGHD is rarely associated with neonatal hypoglycemia. A diagnostic fast may be required to dissect CPHD from other causes of hypoglycemia, although the distinguishing feature from hyperinsulinism is the absence of ketone body formation in the latter.

The presence of concomitant gonadotropin deficiency is suggested by the presence of microphallus, cryptorchidism, and scrotal hypoplasia. Genital ambiguity would not be expected, owing to placental production of hCG. Prolonged jaundice with conjugated hyperbilirubinemia and cholestasis may also be observed, typically in patients with CPHD. The relative contributions of GH, ACTH, and TSH deficiency to this presentation are unclear. It is imperative that the diagnosis of pituitary insufficiency be considered in any infant (especially term) with hypoglycemia, cryptorchidism, and microphallus, or conjugated hyperbilirubinemia. Associated features that might indicate more widespread problems (midline defects of the face, a single central incisor, nystagmus, and/or optic nerve hypoplasia) should be looked for, and MRI undertaken.[22]

Infant And Childhood Years

After the perinatal period, the defining feature of GHD is growth failure. Reduced skeletal growth may be observed during the first 6 months of life in congenital GHD, but by 6 to 12 months of age early growth failure is almost inevitable.[3]

Height velocity is usually between –2 and –5 standard deviations (SDs) from the mean, leading to progressive height centile crossing. In patients with acquired GHD, the critical feature is a change in growth rate. Between the age of 2 years and the onset of puberty, children maintain their height percentile with remarkable integrity. Deviation from this channel (either acceleration or deceleration) needs investigation.

Thus, a child who has been growing along the 75th percentile but moves across to the 25th percentile warrants evaluation, even though his/her height may still be within the normal range.

Bone age is often delayed in patients with GHD, but this may not be so in acquired GHD. The close proximity of time to the growth failure or acquired GHD accompanied by accelerated puberty is occasionally seen in patients with intracranial tumors, when bone age may be accelerated. Delayed dentition may be observed, but in the absence of midline craniofacial abnormalities is otherwise normal. Other skeletal appearances include hypoplasia of facial bones, hypoplastic nasal bridge, frontal bossing, and delayed closure of sutures. Head circumference is usually at the lower limits of normal, indicating normal brain growth.[3,123]

An increase in adiposity, particularly central adiposity, can be detected by careful measurement of skinfold thickness. Genital growth prior to the onset of puberty is usually proportional to body size. Puberty may be delayed, but in the absence of other endocrine deficiencies is otherwise normal.

Limited data are available on the adult height of untreated GHD patients. These results are often difficult to interpret because of (1) heterogeneity in the timing of GHD, (2) heterogeneity in the severity of GHD, (3) the presence or absence of other pituitary deficiencies, and (4) delay in puberty, resulting in late epiphyseal fusion. Wit et al.[124] summarized the results from studies of 22 men and 14 women with untreated isolated GHD who underwent spontaneous puberty and reached a mean final height of 4.7 SDs (range 3.9–6.0) below the population mean.

DIAGNOSIS OF GROWTH HORMONE DEFICIENCY IN CHILDHOOD

The diagnostic evaluation of children with growth failure is complex, because there are multiple causes for short stature (Table 18.4). In the pursuit of the diagnosis of GHD, other causes for short stature need to be considered and excluded. This is because the diagnosis of GHD is a multistep process that requires meticulous evaluation of the clinical history, auxology, and biochemical data, with increasing contribution from pituitary MRI and results of genetic studies.[125] GH is the final common pathway for postnatal growth, and many causes of poor growth may secondarily affect GH secretion. There are a number of tests available for assessing GH status.[126] Considerable attention has been paid to the underlying mechanisms assessed by the tests, how the samples should be collected, and what type of measurement should be performed. Less attention has been paid to the statistical assumptions underlying the performance of diagnostic tests. The statistical theory behind many tests is complex, because the results do not follow an all-or-none law. Rather than being left with a clear-cut answer to the initial diagnostic question, the clinician is more likely to be left with a series of probabilities as to whether or not the patient is likely to have GHD.

TABLE 18.4 **Causes of Short Stature**

Nonpathogenic
Constitutional delay of growth and puberty
Familial short stature
Nutritional
Low Birth Weight
Systemic Disorders
Cardiovascular disease (e.g., congenital heart disease)
Renal (e.g., chronic renal failure, renal tubular disease)
Respiratory (e.g., cystic fibrosis, asthma)
Gastrointestinal disease (e.g., Crohn disease)
Neurologic (e.g., brain tumor)
Psychologic (e.g., anorexia nervosa, child abuse)
Endocrine Causes
GH (growth hormone)-related causes
GH deficiency: isolated or combined with other hormone deficiencies
Resistance to GH
Insulin-like growth factor-1 deficiency
Hypothyroidism
Pseudohypoparathyroidism
Glucocorticoid excess
Cushing syndrome
Congenital adrenal hyperplasia
Exogenous administration
Genetic Causes
Turner syndrome
Noonan syndrome
Down syndrome
Skeletal dysplasias: hypochondroplasia, achondroplasia, spondylo-epiphyseal dysplasia
Russell–Silver syndrome
Seckel syndrome
Prader–Willi syndrome
Miscellaneous other syndromes (e.g., Rothmund–Thompson syndrome, Leri–Weill syndrome, progeria, mucopolysaccharidoses)

Guidance Derived From Clinical Assessment

Neonatal Period. Several pointers to the diagnosis of GHD have already been considered in this discussion, but in the neonatal period GHD may be isolated or associated with other pituitary hormone deficiencies. Small genitalia may point to associated gonadotropin deficiency. Hypoglycemia in the newborn period is often a feature of ACTH deficiency, although, on an arbitrary basis, a serum GH of less than 10 ng/mL is considered consistent with a diagnosis of GHD under these circumstances. This is not universal, however, and caution needs to be exercised in interpreting the GH response to hypoglycemia under different circumstances.[127] Recently, Binder et al. reported that, using a highly sensitive hGH enzyme-linked immunosorbent assay, the median GH concentration in dry blood spot samples of healthy neonates was 16.4 µg/L, while the median serum GH concentration of nine newborns diagnosed with hypopituitarism was 2.1 µg/L (maximum 5.5 µg/L), with no significant overlap between the groups. Based on these groups, the authors concluded that a single cutoff value of 7 µg/L, when measured with the same assay, would have high sensitivity (100%) and specificity (98%) for the diagnosis of neonatal GHD.[128]

Prolonged neonatal jaundice raises the question of thyroxine (unconjugated) or cortisol (conjugated hyperbilirubinemia) deficiency. Given these features, it might be possible on the basis of pattern recognition to ascribe the diagnosis of GHD to a patient with a high degree of certainty. MRI of the brain should be obtained to look for an undescended posterior pituitary, anterior pituitary hypoplasia, hypoplasia or absence of the pituitary stalk, hypoplasia of the optic chiasm, and absence or hypoplasia of the corpus callosum and septum pellucidum.[129]

Infancy and Childhood. Diagnostic evaluation in children must be based upon auxology. Although there are a number of clinical features of GHD that are said to be classic, none is specific. For example, obesity is listed as a clinical feature of GHD, but if we simply restricted biochemical evaluation to patients with obesity as the main feature, testing the GH axis would yield a large number of individuals with a poor GH response, because obesity *per se* is associated with blunted GH responses to various stimuli. Individuals who exhibit GHD are often obese, but the converse is clearly not the case.

Little is known of the sensitivity and specificity of many of the clinical observations, either alone or in combination. The prevalence of many of the clinical features within the general population is unknown, which heightens the problem. Even the presence of specific features or a combination of features will only slightly increase the likelihood of disease if they are relatively insensitive.

The manifestation of GHD as a result of a *GH* gene deletion is early, and poor growth can be detected as early as the sixth month of postnatal life. With advancing age, more GH has to be secreted to maintain concentrations of GH sufficient for growth, so idiopathic isolated pituitary GHD may present at any time. It is the degree of deficiency that dictates when the individual comes to medical attention. Table 18.5 provides general clinical rules that are a useful aid when selecting patients for further study of the GH axis.

Principles of Testing

The aim of any diagnostic test is to progress the clinical history and examination to the point where patient care is altered. No test will ever benefit a patient. It is only when subsequent treatment has to differ depending on the test result that patients will be better off. There is a vast and bewildering body of literature on GH testing, but the clinician can be guided by asking the questions detailed in Table 18.6. It is important to remember that a diagnostic test is not just about whether a disease is present or not; it may also be important in determining severity and prognosis and responsiveness to and monitoring of therapy, and as a screening tool. As such, how the test performs under one circumstance may not be the same in another. Measuring serum IGF1 concentrations may be unhelpful in screening for GHD but may be excellent as a marker of response to therapy.

TABLE 18.5 Clinical Indicators for Further Evaluation of the Growth Hormone Axis[204]

1. Height SDS below −2. Although WHO growth charts can be utilized for children up to 2 years of age, local growth charts, when available, are more appropriate for older children.
2. Height that clearly deviates from the familial background.
3. A significant decrease in height SDS (i.e., a deflection of at least 0.3 SDS/year that is not explained by the normal channeling in infancy to adjust linear growth to target height trajectory, by the prepubertal growth dip or by pubertal delay).
4. Predisposing condition (tumor, radiation, etc.) or features suggestive of an underlying syndrome.
5. Neonatal signs consistent with pituitary hormone deficiencies.

SDS, Standard deviation score; *WHO,* World Health Organization.

TABLE 18.6 Underlying Principles of Assessing Tests

1. Has there been an independent blind comparison with the diagnostic “gold standard”?
2. Was the test conducted in a wide range of patients with and without the condition?
3. Is the test reproducible?
4. What was the definition of “normal” in the test situation?
5. How might the test interact with others in a diagnostic sequence?
6. Does the test entail risk or reduce risk for the patient?

Two points deserve special mention. First, it is unusual in endocrinology for there to be a diagnostic “gold standard.” The anterior pituitary is not accessible, and molecular biology is not sufficiently advanced to give definitive answers. Second, care needs to be taken in ascribing the role of a gold standard. It may change with time, and the test must be well validated by application to large numbers of individuals with and without the condition. The temptation is to use the extremes, but this may lead to a considerable overestimate of sensitivity and specificity, which may not be borne out in field studies.[130]

Two principles operate when using diagnostic tests.[131] First, probability is a useful marker of diagnostic uncertainty. This is when the sensitivity (ability to detect a target disorder when present or true positive rate) and specificity (ability to identify correctly the absence of the disorder or true negative rate) become important. If both were 85%, 15% of patients with disease would have a negative result (false negative), and 15% without disease would have a positive result (false positive). Abnormal results would occur in patients with and without disease. Whatever the result, new information has been generated that may or may not influence decision-making. Second, diagnostic tests should be obtained only when they can alter the management of the case—that is, if the test result alters the probability of the disease.

Pre- and Posttest Probability. The relationship between the probability of disease after the results of diagnostic tests are known (the posttest probability), and pretest probability of disease depends on the sensitivity and specificity of the test, as shown in Fig. 18.6. There are two important points to note: the first is that the more certain the clinician is of the diagnosis before the test is performed, the less effect the confirmatory test has on the probability of disease. The obverse is also true. The second point is that tests will have major effects on probability of disease in the intermediate zone. Testing is not likely to be beneficial if the pretest probability is very high or low. This is one reason why screening for GH problems in short children on the basis

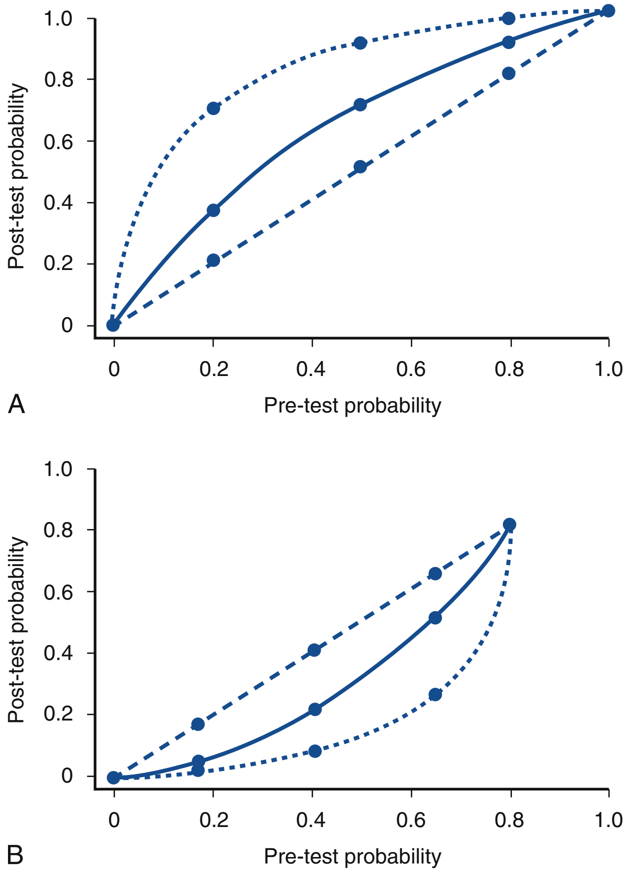


Fig. 18.6 The relation between pretest and posttest probability of disease. The data were constructed by using Bayes' theorem with a test sensitivity and specificity of either 70% *(solid line)* or 90% *(dashed line)*. **A**, The posttest probability if the test were positive. **B**, The posttest probability if the test were negative. If the posttest probability were the same as the pretest probability, then the relation would be given by the line of identity. (Reproduced from Brook CGD, Hindmarsh PC, Jacobs HS, eds. *Clinical Pediatric Endocrinology*. 4th ed. London: Blackwell Publishers, 2001.)

of biochemical tests is unhelpful; the pretest probability is 1 in 3000, or 0.03%.

Clinicians are often faced with the situation where they feel really sure the patient has the condition, but the test does not confirm this. Table 18.7 analyzes this concept. Here, specificity and sensitivity have been fixed, and the effects on posttest probability are considered. In the situation where there is a 90% pretest probability that the patient has GHD, then even if the test is negative in the individual, there is still a 67% probability (reduced by 23%) that they have the condition, so treatment would still be justified. When the pretest probability was 5% (very certain that the patient does not have GHD) and the test is positive, all the result says is that the patient has a 1 in 4 chance of having the condition, so we would probably not treat. In the middle ground, certainty in either direction is dramatically improved.

Multiple Tests. Table 18.7 could have been made much larger by introducing any number of pretreatment probabilities. There comes a point, however, when posttest probability changes to a level where a decision has to be made to stop and either accept or reject the proposal that the condition is present. The decision to stop investigation and to treat or not depends on how convinced the clinician is of the diagnosis, the benefits and risks of the therapy, and the potential yield and risks of further tests. There are two ways to assist this situation: conduct another test or use a more sophisticated analysis rather than a simple positive or negative.

TABLE 18.7 Effect on Posttest Probability of Differing Pretest Certainty, Assuming Constant Sensitivity and Specificity

	POSTTEST PROBABILITY	
Pretest Probability (%)	**Test Positive**	**Test Negative**
90	98 (+8)	67 (–23)
50	87 (+37)	19 (–31)
5	25 (+20)	1 (–4)

Change from pretest probability in parenthesis.

This is problematic in the GH field, because the methodology assumes that the results of the two tests are independent. In normal individuals undergoing repeat GHRH tests, dependence cannot be assumed.[132] Where repeat tests have been performed in children, concordance was observed 50% of the time, a value close to that calculated for independent events using a test with 70% to 85% efficiency. If all tests are treated as independent, there is a risk of over- or underestimating the presence of the condition. Another important issue is whether the test may change in individuals as they age. There is evidence that the clonidine test is less effective in releasing GH in young adults compared with children. Whether the magnitude of the response to other stimuli can be assumed to remain unchanged is unknown.

Assuming that the two tests are performed (on different days) and that they are dependent, then if both tests need to be positive for diagnosis, this maximizes specificity and avoids falsely labeling normal children, but it misses many treatable individuals. Insisting that both tests are negative maximizes sensitivity and minimizes misdiagnoses, but falsely labels many more normal child.

These challenges in the diagnosis of childhood GHD are highlighted by studies showing that up to 85% (28/33) of short prepubertal children who were diagnosed with GHD based on a peak GH of less than 10 μg/L in two provocation tests had a normal GH response when retested 1 to 6 months later.[133] In other series, almost 60% to 85% of patients diagnosed with GHD in childhood had an adequate GH secretion when retested in late adolescence or adulthood.[134]

Diagnosis of Growth Hormone Deficiency

Assessment of GH secretion is problematic, in part because of the pulsatile nature of GH secretion. The most consistent GH surges accompany slow-wave electroencephalographic rhythms during phases 3 and 4 of sleep. Although this rhythmicity is characteristic of GH secretion at all ages, the size of the amplitudes and the total integrated GH secretion varies with sex, age, pubertal status, and nutrition.[135] Between pulses, serum GH concentrations are extremely low, often less than 0.1 ng/mL. Consequently, measurement of random serum GH concentrations is of no value in the diagnosis of GHD. Measurement of spontaneous GH secretion requires multiple sampling, typically every 15 minutes over a 12- to 24-hour period. Such methodologies are inconvenient and expensive, and, while they allow identification of the patient with severe GHD, it is not clear that they can discriminate between partial GHD and normal secretory variation.[136] However, even the reproducibility of GH secretory patterns in children from day to day is uncertain. Rose and colleagues[136] reported that measurement of spontaneous GH secretion identified only 57% of children diagnosed as GHD by provocative testing. Lanes[137] reported that approximately 25% of normally growing children have low overnight GH concentrations. A longitudinal study of GH secretion in normal

TABLE 18.8 Growth Hormone Stimulation Tests

Stimulus	Dose	Sampling Protocol (min)	Notes
Exercise	Cycle for 10–15 min	15 min for 90 min	Variable response, highly dependent on degree of exercise
Levodopa	<15 kg, 125 mg 15–30 kg, 250 mg >30 kg, 500 mg	15 min for 90 min	Nausea
Clonidine	0.15 mg/m^2	30 min for 90 min	Tiredness; postural hypotension
Arginine HCl	0.5 g/kg (max, 30 g) IV given as 10% arginine HCl in 0.9% NaCl over 30 min	15 min for 90 min	May cause insulin release
Insulin	0.05–0.1 units/kg IV	15 min for 120 min	Hypoglycemia; requires supervision. Can also measure cortisol reliably
Glucagon	0.1 mg/kg IM (max, 1 mg)	30 min for 180 min	Nausea
GHRH	1 μg/kg IV	15 min for 120 min	Flushing. Only assesses pituitary reserve, not whole HP axis
GHRH-arginine			Needs further work to assess value in pediatrics
Tests should be performed after an overnight fast.			
Patients should be documented to be euthyroid.			
Prepubertal children should be primed with gonadal steroids.			

GHRH, Growth hormone-releasing hormone; IM, intramuscular; IV, intravenous.

boys during puberty indicated wide intersubject variation, and day-to-day variation has been noted among normal subjects. An alternative approach has been the measurement of urinary GH concentrations.[138] This methodology requires a timed urinary collection and a GH assay of high sensitivity, because urinary GH concentrations are low. The theoretical advantages of this approach include its relative ease of performance and noninvasive nature, as well as the requirement for only a single GH measurement. This must be balanced, however, by the need to assess the effects of renal function, the wide interindividual variation, and the lack of adequate age- and sex-related reference ranges.

Obesity is another aspect that complicates primary GHD diagnosis, as visceral obesity *per se* results in a *secondary* reduction in serum GH concentrations.[139] The reasons for the hyposomatotropinism in obesity and its mechanisms have yet to be clarified.

Reductions in spontaneous GH secretion (as much as 6% for each unit increase in body mass index [BMI]) and in the half-life of circulating GH have been reported. The GH response to pharmacological and physiological stimuli is impaired in obesity.[140]

Some of the theories on the cause of altered GH physiology in obesity involve the increased concentrations of leptin, insulin, free fatty acids, and IGF1.

As a result of these difficulties, the standard for the diagnosis of GHD has been provocative testing of GH "secretory reserve" (Table 18.8). Physiologic stimuli for such tests have included sleep and exercise, and pharmacologic stimuli have included a wide variety of agents.[141] None of these tests truly mimics normal GH secretory physiology, and none has been evaluated adequately in normal children and normal short children. The limitations of provocative GH testing in the diagnosis of GHD are described below and need to be considered in the light of statistical theory (see earlier):

1. Provocative testing by its nature is nonphysiologic. None of the commonly used stimuli truly mimics normal regulation of GH secretion.
2. The definition of a "normal" response to stimulation is arbitrary. Normal values are difficult to obtain in pediatric practice, and reference ranges would be needed for tall, normal, and short children, because their GH secretion differs. In addition, both age and pubertal stage influence GH secretion, as does body composition.[142] Values for these would also have to be included.
3. The classic approach of defining normal data in terms of a Gaussian distribution does not come without hazard. Endocrine testing rarely fits this distribution; even if it did, it would imply that the lowest and highest 2.5% of values are abnormal, and that all diseases have the same frequency—clearly unlikely. Creating upper and lower limits does not help either. It is more appropriate to identify a range of diagnostic test results, beyond which GHD is likely.
4. Most decisions on placing the value have been empirical rather than statistical. In practice, cutoff values could be chosen at an absolute extreme. If 100 short children were studied, and GH sufficiency or deficiency was defined by a peak response of less than 3 ng/mL, only 3% to 5% might have a response at this level. When testing the next 100 children, one or two normal individuals might have such a response. They will be outliers, but they are important, because the more patients studied, the greater the chance of finding outliers.
5. Moving the cutoff to more extreme values to exclude these patients restricts the population of treatable individuals. Relaxing the criteria interposes normal individuals into the diagnosis zone. Placing the cutoff is based partly on clinical judgment. Because there is no perceived disadvantage apart from financial cost in falsely labeling someone with GHD and treating him or her, a relaxed cutoff would be acceptable. However, this could lead to some children receiving unnecessary treatment with GH, which is not completely without side effects.
6. The question of cutoff points for use in tests becomes more important as individuals with differing severities of the disorder are considered. In constructing normal ranges, it is clearly best if large sample sizes are chosen. Choosing populations of disease-positive and disease-negative is unwise, either for this task or assessing test performance, because it is unlikely that the test will perform as well in the less severe cases.[143] Some useful points in the assessment of studies are outlined in Table 18.9. Referral bias remains a major issue in many studies. In studies emanating from referral centers, the strength of a factor such as short stature may appear to be less important, in that patients are already selected for this in the referral. Changes in the prevalence of a condition do not change test properties, whereas changes in the spectrum of the disorder do.[144]
7. The dependence of GH secretion on other factors needs to be taken into account. Marin and colleagues[145] demonstrated that, when exercise and arginine stimulation tests were performed on normal-stature children without sex steroid priming, the lower-limits-of-normal (–2 SD) peak serum GH concentration for

TABLE 18.9 Considerations to Give to Test and Control Populations

1. What is the population from which patients are drawn?
2. How have patients been filtered before joining the study?
3. Test properties may differ in subpopulations within the sample.
4. Recruit consecutive patients with the problem.
5. Can the findings be generalized?

prepubertal children was as low as 1.9 ng/mL and rose to 7.2 ng/mL on estrogen priming. Thyroxine and cortisol, which directly alter gene transcription, influence the results obtained, and these need to be controlled before undertaking a diagnostic study. Similarly, the presence of high concentrations of glucose or free fatty acids may influence the response obtained.

8. GH assays may measure a variety of immunoreactive molecular forms of GH, and individual standards need to be established for each laboratory. These should take into account factors intrinsic to the assay used, the heterogeneity of the GH molecule (different isoforms circulating in hetero- and homodimers), circulating GHBP, and the standard preparation used for calibration. At the time of the 2000 *Consensus of the GH Research Society* on the diagnosis of GHD in children,[125] the international standard (IS) available was the IS 88/624 (recombinant 22-kDa GH). Since then, it has been replaced by the IS 98/574, the implementation of which has been recommended by an international consensus statement on the standardization and evaluation of GH and IGF assays.[146] These recommendations and the current use of monoclonal assays make it imperative to review the accepted GH cutoff used for diagnosis of GHD in children. Assay precision, accuracy, and sensitivity play an important role in determining the success or failure of the diagnostic test. For example, assays for IGF1— despite having good standardization from the technical aspect (elimination of interference of IGFBPs and the use of high-affinity, high-specificity antisera)—may have different performance characteristics when concentrations are either above or below the normal range.[147]
9. Most endocrine tests are conducted over short periods of time, and results are extrapolated to longer time frames. GH provocation tests take 2 hours to perform, and the results are then compared with height velocity measurements obtained over a longer period of time, often 1 year. That there is a relationship is perhaps surprising; that there are high false-positive and false-negative rates is probably not.
10. Hormone pulsatility may also influence diagnostic tests if the test itself is influenced by oscillations (e.g., the stimulus applied) within the system under study. The GH response at any point in time is going to be heavily dependent on the interplay between the hypothalamic regulatory peptides involved in GH release, namely GHRH and SS. SS, in particular, is a key determinant of the amount of GH released as a result of GHRH stimulation. Attempts have been made to take control of this variable by the use of GHRH combined with arginine.[148] An alternative approach is the use of ghrelin.[149] However, in patients with congenital hypopituitarism the GH response to ghrelin depends on the degree of the anatomical abnormalities, as the main action of the peptide is at the hypothalamic level and requires the integrity of hypothalamic–pituitary connections.[150]
11. Endocrine systems are also subject to feedback from target tissues, and this is an issue not only in the interpretation of single provocation tests, but also where second tests are performed in rapid succession to the first. A diminished response to GHRH can be observed if the second stimulus is applied 1, 2, or 3 hours after the first. The implication of doing two tests on the same day, often following each other, are immense; the cutoff that might be implied to determine normality or not may not be the same for the second test as for the first, especially if the second stimulus is different from the first.
12. Provocative testing fails to give any consideration to the effect of negative feedback by serum IGF1. It probably makes more sense to interpret serum GH concentrations in the light of serum IGF concentrations, much as TSH concentrations are best assessed with a knowledge of circulating thyroxine concentrations.
13. In assessing the results of endocrine evaluations, it is generally assumed that the single or multiple samples measured are relatively stable, at least over short periods. When important changes are postulated to be taking place, for example in a disease process, some knowledge of the inherent variability within the measurement system is required. In the short term, a number of studies have demonstrated variability within and between individuals in terms of GH tests. Group data are usually reproducible, but problems can arise if it is assumed that individual oscillatory profiles are consistent from day to day.
14. In considering provocative tests, the situation may arise where no response is observed. A possible explanation is that the strength of the stimulus was insufficient to provoke hormone release. In such a situation, it is valuable to have an independent marker of stimulus application. In the insulin-induced hypoglycemia test, this marker is glucose and the attainment of adequate hypoglycemia. In the glucagon test, it may be the release of glucose. In other tests, there may be no independent markers, so doubt may be cast on the reliability of the nonresponse.
15. Careful consideration needs to be given to the age of the child under study. Not only may the cutoff point criteria differ at different ages, but the likelihood of disease presence will change with age. It is highly unlikely that GHD will manifest itself during puberty. It is possible, but it is more likely that any growth or GH secretory problems at this age relate more to delayed puberty rather than an abnormality in the GH axis. Even in childhood, the return in terms of diagnosis of GHD is not high if height screening is undertaken at school entry, so that with increasing age there is a diminishing diagnostic return.
16. Provocative GH testing is expensive and uncomfortable, and has risk. Insulin-induced hypoglycemia should only be performed in a supervised setting. Deaths have been documented in patients rendered hypoglycemic and corrected in an overly vigorous manner.
17. Of the provocative tests listed in Table 18.8, it should be noted that stimulation with GHRH is not designed to document whether a patient has GHD, but rather whether GHD, established by other methodologies, is the result of pituitary or hypothalamic dysfunction. Failure to respond to GHRH suggests that the abnormality is at the pituitary level. This test may be enhanced by the addition of arginine or pyridostigmine.

A Practical Approach to Diagnostic Evaluation

A practical approach to the diagnosis of a child with GHD is grounded on clinical assessment with allocation of pretest probability of disease presence. In the neonate and first few months of life, Binder and colleagues[151] proposed a GH cutoff of 7 μg/L, while most recent guidelines have suggested that a cutoff of 5 μg/L in a newborn with additional pituitary hormone deficiencies and/or structural abnormalities of the pituitary is more appropriate.[152] In the prepubertal child with abnormal growth, serum concentrations of IGF1, IGFBP3, and/or ALS provide a means for excluding a diagnosis of GHD. Provocative GH testing with appropriate sex steroid priming will provide information on GH secretory capability, and GHRH stimulation, with or without arginine or pyridostigmine, will allow determination of whether the defect is at the hypothalamic or pituitary level.

All data need to be interpreted together with known test performances and integrated with the pretest probability to generate a posttest probability that would then lead to a decision as to whether intervention is required.

The interpretation of second tests of GH secretion may be problematic from both the physiologic and statistical standpoints, and should be analyzed with extreme caution.

Documentation of GHD also requires that other pituitary functions be assessed periodically, including TSH, ACTH, and gonadotropin status. Other pituitary deficiencies may not be evident upon initial assessment but may develop over time. MRI of the hypothalamus and pituitary should be performed initially to determine if there is evidence of intracranial tumors, pituitary hypoplasia or PSIS, or midline defects. Even if the baseline MRI is normal, in the absence of an alternative explanation for GHD or CPHD, the possibility of tumors or structural defects should not be dismissed permanently. With increasing knowledge of the genetics of pituitary disorders, these should be looked for, because they have an impact on the likelihood of other pituitary hormone deficiencies evolving with time and allow genetic counseling to be undertaken.

KEY POINTS

- The diagnosis of GHD is by no means simple, and often involves careful evaluation of the clinical history, examination, basal biochemistry, GH response to stimulation, radiology including a bone age and MRI of the brain and pituitary, and molecular analysis. Confounding factors in the interpretation of GH responses to provocation include obesity.

TREATMENT

Growth Hormone

The first successful treatment of human GHD with hGH was in the 1950s, while rhGH became available in 1985. This allowed for potentially unlimited supplies, obviating the need for low-dose usage and interrupted therapeutic regimens. The initial rhGH preparation was an N-terminal methionine, met-rhGH, which was fully active biologically but was ultimately replaced by the mature 191–amino acid protein. Since then, information on the safety of rhGH treatment has mainly been obtained from large postmarketing databases including patients with GHD of variable etiology and non-GHD subjects, confirming an overall safe profile.[153,154]

Dose Studies. Investigations of optimal dosing of rhGH have been complicated by the use of heterogeneous study populations, as studies frequently include patients with unequivocal and complete GHD together with patients with partial GHD. Therefore, it is not surprising that, despite long experience in the use of rhGH, there still remains uncertainty regarding the best dosing regimen, its calculation based either on weight or surface area, and a satisfactory way to quantify and predict responsiveness to treatment. It is generally accepted that a daily dose of 0.025 to 0.035mg/kg/day is sufficient to increase growth velocity to more than 10 cm/year in children with severe GHD and the adult height in treated patients ranges from –1.5 to –0.8 SDS, depending on the study.[155,156] Several studies have demonstrated a dose-response relationship for hGH, but the slope of the response is relatively shallow. Cohen et al.[157] compared the growth responses of prepubertal, naive patients randomized to rhGH at a dosage of 0.175, 0.35, or 0.7 mg/kg/week for the first 2 years of treatment. Significantly greater height velocities and gains in height SD resulted from the 0.35 mg/kg/week versus 0.175 mg/kg/week, but no further significant improvement was observed with the 0.7-mg/kg/week dosage. Ultimately, the issues that should determine dosage in the child with GHD are (1) how best to return the GHD child to the normal growth curve, (2) how best to ensure that the child attains his or her genetic height potential, (3) risks, and (4) cost.

However, instead of fixed dose regimens, titration of the dose of rhGH according to a target IGF1 SDS near the age- and gender-adjusted mean (IGF1 SDS of −0.5 to +0.5) seems to be beneficial in terms of growth, at least in the short term, with great variability in the dose required.[178] Although dose titration to achieve higher IGF1 targets (IGF1 SDS of +1.5 to +2.5) resulted in a higher growth response over a 2-year term, it is not possible to conclude that this dosing strategy will result in improved adult height outcome.[158]

Response to GH therapy varies even when the diagnosis is homogeneous. This probably reflects differences in tissue responsivity, which may relate in part to the function of the GHR. A polymorphism in the *GHR* gene leading to retention (full-length) or deletion of exon 3 (d3), which encodes a 22–amino acid sequence in the extracellular domain,[159] has been associated with a greater height increase in response to GH replacement in children born small for gestational age (SGA), those with idiopathic short stature (ISS),[160] and in a GHD population.[161] Patients with at least one d3 allele had a significantly better first-year response leading to an improved adult height on GH treatment than patients with homozygosity for full-length GHR. However, not all reported studies are consistent, which may reflect differing populations and conditions.[162] In the current context, the use of the GHR polymorphism, alone or in association with other genetic variants within the GH-IGF1 axis, is of little value in prediction models to estimate the response to rhGH treatment in short children.[163]

False-positive findings are more likely with small sample sizes, and for quantitative trait loci, phenotypic variations tend to be overestimated with small sample sizes.

Frequency of Administration. Several studies have compared the short-term effects of administering hGH either daily or thrice weekly. Generally, daily injections are more effective, but increasing the frequency more than this makes little difference.

Multiple long-acting growth hormone (LAGH) preparations are currently at various stages of development, allowing for decreased GH injection frequency from daily to weekly, biweekly, or monthly, and thus improved adherence.[164,165]

Longer-acting agents (LAGHs) can be divided broadly into depot formulations, PEGylated formulations, prodrug formulations, noncovalent albumin-binding GH, and GH fusion proteins. Following administration of LAGH, the serum peak and trough GH and IGF1 concentrations vary depending upon the mechanism used to prolong GH action. Randomized, controlled clinical trials have not reported significant LAGH-related adverse events during short-term therapy. In order for LAGH to replace daily rhGH in the treatment of GHD, further studies are needed to clarify methods of dose adjustment, timing of monitoring of IGF1, safety, efficacy, and cost-effectiveness. Furthermore, long-term surveillance registries to evaluate efficacy and safety will be essential for understanding the impact of prolonged exposure to these compounds.[166]

Prediction Models. A series of models have been derived[167] that describe factors that may influence response, but none has gone on to be tested in formal randomized controlled trials. Further problems arise when large multicontributor databases are used, because of the

TABLE 18.10 Specific Issues with Prediction Models of Response to Growth Hormone Treatment Development

1. Regression to the mean
2. Prediction gives average effect; how to apply to individuals for prognosis
3. Predictors may not be independent of each other; assumptions made in linear and multiple regression modeling
4. Assumptions about uniformity of response
5. Spurious correlations involving time

accuracy of the data entered; when prognosis is considered, factors that may be important may not have been recorded. It has been reported that, in children with GHD, auxologic parameters such as chronologic age (the younger the patient, the better the response) and the difference between target height and actual height (the smaller the patient, the better the response) are better predictors of growth response than the cumulative weekly GH dosage. There are several problems with these types of models:

1. Although prediction models are useful to give an average effect,[168] they are not individualizable. These models explain only approximately 40% to 60% of the observed variability and are limited by their short-term endpoints, the lack of validation in long-term randomized trials, and the fact that they do not take into account the genetic makeup of the individual.
2. They often only focus on one outcome, usually short-term growth, whereas interest may be more centered on final height. The two need not necessarily be related, and the factors that influence response in the first year of treatment may differ totally from those that lead to prediction of the individual's final height.
3. Very few prediction models have been constructed from an *a priori* hypothesis, and care needs to be taken that there has been no interference from other factors accompanying the disease that might affect prognosis.
4. Rules derived from one data set may reflect associations that have occurred by chance and often result from overfitting of the data.
5. There is always the possibility that the predictors are idiosyncratic to the population, to the setting, to the clinicians, or to other aspects of the original study. Specific issues associated with growth-response models are summarized in Table 18.10.

Height and Other Outcomes. Early initiation of therapy, combined with careful attention to dosage adjustments and compliance, is the best predictor of cumulative growth response in patients with GHD. Final height correlates with height at the initiation of puberty, so it is important to maximize growth during the prepubertal period, within the limits of safety and economy. Analysis of data on final heights of rhGH-treated GHD is complicated by the heterogeneity of patient groups and dosage. In earlier reports a common observation was a general failure of children to reach their full genetic height potential, especially in the case of IGHD and particularly in females. Price and Ranke[69] reported final heights of –1.26 SD and –1.45 SD from the mean in males and females, respectively, with IGHD, and –0.22 SD and –0.52 SD from the mean in males and females, respectively, with CPHD. In patients with longer durations of treatment and higher dosages of rhGH, adult heights tended to be greater, although they still failed to achieve full genetic height potential. Recent reports of the final height of GHD children treated with rhGH show that, as a group, they achieve a normal final height within their genetic potential,[155] although there is still considerable variability in the response depending on the population, sex, and underlying diagnosis (IGHD compared with CPHD).[154,170] Males of European/North American and Japanese origins with IGHD achieved a near-adult height of –0.8 SDS, while females achieved a height of –1.0 SDS; differences between near-adult height and midparental height ranged between –0.6 and +0.2 SDS.[155]

The timing of GH treatment initiation increases the chance to achieve near-adult height. The NordiNet International Outcome Study, a noninterventional, multicenter study, identified that near-adult height SDS achieved by patients starting treatment early (n = 40 [boys, 70.0%)]; least squares mean [standard error] −0.76 [0.14]) exceeded that achieved by those starting later (intermediate, n = 42 [boys, 57.1%]; −1.14 (0.15); late, n = 90 [boys, 68.9%]; −1.21 [0.10]).

The study concluded that early initiation of GH treatment in children with isolated GHD improves their chance of achieving their genetic height potential.[171]

It is clear that the timing of puberty has a significant impact on adult height of individuals with rhGH-treated GHD. The duration of rhGH treatment and the height gained prepubertally are typically greater when puberty is induced rather than spontaneous. Final heights were greater after induced puberty compared with spontaneous puberty in boys (171.3 vs. 166.0 cm) and in girls (157.0 vs. 155.0 cm). Therapy designed to delay the onset of puberty (both normal and precocious) may augment the cumulative growth response to rhGH, but this is not accepted practice, and should only be considered in exceptional cases.

Final height gain can be particularly variable in children who have had treatment for malignancies. GHD is often complicated by skeletal damage following TBI or craniospinal irradiation, early puberty, hypothyroidism, gonadotropin deficiency, malnutrition, and concomitant chemotherapy. GnRHa therapy to arrest early puberty has been used in conjunction with GH treatment in this group of patients, with encouraging results. The use of GnRHa reduces the concentration of sex steroid, with a consequent delay in epiphyseal fusion. However, GH and GnRHa combination therapy in children with GHD is not widely used at present, and the long-term beneficial effects in terms of adult height are not well established.[172] It may be beneficial under certain circumstances—for example, where the diagnosis of GHD has been delayed. The effects of GnRHa in the long term are unknown, particularly on bone health; and the cost of this combination therapy would need to be weighed against the presumed benefit.[172]

Although considerable attention has been paid to growth, there is now recognition that the beneficial effects of rhGH in children extend beyond height. GH treatment in childhood can also normalize body composition, with a reduction in body fat, although effects on lean body mass are less evident.[173] It is also associated with reversible insulin insensitivity[154] and an increase in the ratio of high-density lipoprotein (HDL) to total cholesterol, improved cardiac function, and reduction of proinflammatory cytokines, along with an increase in glomerular filtration rate and acceleration of bone remodeling, with an increase in bone mineral mass.[174]

An increasing number of studies have focused on the role of GH in memory and neurocognition.[175] On the other hand, functional MRI has shown that, although children with IGHD do not have a global reduction in brain volume, specific structures are affected, with significantly lower volumes of the splenium of the corpus callosum ($P <$ 0.02), right pallidum ($P < 0.007$), right hippocampus ($P < 0.01$), and left thalamus ($P < 0.01$) compared with controls, while the development of the corticospinal tract is also affected.[176] They also exhibit specific impairments in cognition and motor skill function.[176] Well-designed studies assessing cognitive function and using advanced MRI are needed in order to clarify whether treatment with rhGH can reverse these effects and affect the neurocognitive outcome and development of GHD children.

Adverse Effects of Human Growth Hormone. Treatment with rhGH has had an excellent safety record.[153],[154] Reports of anti-GH antibodies in patients receiving rhGH have been few, with no untoward effect on the growth response. Fluid retention and carpal tunnel syndrome are observed in adults but are seldom significant in children. Whereas the incidence of type 1 diabetes mellitus is not higher in patients with idiopathic GHD than in the general population, the incidence of type 2 diabetes mellitus (T2D) is greater in those patients treated with GH, although this tends to be in those predisposed to developing the condition.[154] Idiopathic intracranial hypertension (pseudotumor cerebri) has been observed occasionally but resolves with cessation of treatment and then a gradual reintroduction of rhGH commencing at a lower dose. Slipped capital femoral epiphysis has been reported, but it is not clear whether the incidence is greater than that observed in normal children during rapid periods of growth. Other possible rare and still unproven complications include acute pancreatitis, increased growth of pigmented nevi, and prepubertal gynecomastia.

Although early reports suggested a link between GH treatment and leukemia, Blethen and colleagues[177] reported that an analysis of 47,000 patient-years of treatment in greater than 19,000 children indicated that, in children without known risk factors, rhGH therapy was *not* associated with an increased occurrence of tumors or recurrence rate of leukemia or CNS tumors. Subsequent reports with an increasing number of patient-years confirmed that rhGH treatment does not increase the risk of primary malignancy in children and adolescents without preceding risk factors.[153] In a further study, Bell et al.[178] reported on 54,996 patients enrolled in the National Cooperative Growth Study database, of whom 42% had idiopathic GHD (n = 23,393), and 15% organic GHD (n = 8351). Among the children with organic GHD, the authors observed recurrence of intracranial tumors (1.8%), second neoplasms (0.5%), and new-onset malignancies without identifiable risk factors (0.1%), and suggested that rhGH may increase the risk of second malignancy in previously treated patients.[178] However, in the entire cohort the overall risk of malignancy was comparable to the reference population, with a standardized risk ratio of 1.12 (95% confidence interval [CI]: 0.75–1.61). In reports from the Childhood Cancer Survivor Study, the adjusted relative risk for second neoplasm in patients treated with rhGH compared with nontreated patients ranged from 3.21 to 2.15 (95% CI: 1.88–5.46 and 1.33–3.47, respectively).[179]

Another metaanalysis of published studies to estimate the potential relation between GH replacement treatment (GHRT) and recurrence/progression of intracranial tumors, searched three databases (PUBMED, EMBASE, Cochrane Library) to the end of July 2014. They collected more than 2232 cases and 3606 controls. The results indicated that recurrence/progression of intracranial tumors was not associated with GHRT (risk ratio [RR] 0.48, 95% CI: 0.39–0.56; for children RR 0.44, 95% CI: 0.34–0.54). In subgroup analysis, risks of recurrence/progression were decreased for craniopharyngioma, medulloblastoma, astrocytoma, and glioma, but not for functioning pituitary adenomas, nonfunctioning pituitary adenomas, and ependymomas. However, fluctuations in serum concentrations of GH and IGF1 were observed in neoplastic disease, which may affect tumor progression. In conclusion, GHRT seemed safe with respect to the risk of tumor progression.[180]

It is nevertheless the accepted practice to delay rhGH treatment in children with treated brain tumors until they have been shown to be tumor-free for at least 1 to 2 years.

Long-term follow-up has revealed a higher-than-expected incidence and mortality of colon cancer and Hodgkin disease. However, these data need to be put in context, as affected patients had been treated with high doses of human pituitary-derived GH given two to three times a week. Hence, one could speculate that the IGF1 concentrations generated by this mode of GH therapy may have been excessive. IGF1 concentrations at the upper end of the normal range (top quartile for colon and prostate and top tertile for breast) have been associated with colon, breast, and prostatic cancer.[181] Given that lower doses of GH are given on a daily basis, it would be incorrect—at this stage at least—to extrapolate the data from the earlier studies to the present treatment regimens.

The study on the *Safety and Appropriateness of Growth hormone treatments in Europe* (SAGhE)[182],[183],[184] was set up to address the issue of the long-term safety of rhGH treatment, including 30,000 low-risk children treated with rhGH between 1980 and 1990 in eight European countries who had attained 18 years of age by the census date. Patients were classified as having a low mortality risk if they were diagnosed with isolated idiopathic GHD, SGA, or ISS without genetic syndromes or defects. An early report of the results on the analysis of mortality for the cohort treated in France[182] showed an increase in the overall standardized mortality ratio (SMR), calculated at 1.33 (95% CI: 1.08–1.64), with the majority of deaths observed in the group with idiopathic GHD. The increase in all-cause mortality was associated with a higher rhGH dose of greater than 50 μg/kg/day (SMR 2.94; 95% CI: 1.22–7.07) but not with the severity of GHD or the duration of total GH exposure. When looking at the cause of death, there was no increase in the overall risk of cancer-related deaths, although numbers were relatively small. However, there was an almost 5-fold increase in deaths due to bone tumors (SMR 5.0; 95% CI: 1.01–14.63), as well as an increase in mortality from cardiovascular disease (SMR 3.07; 95% CI: 1.40–5.83) and cerebrovascular hemorrhagic events (SMR 6.6; 95% CI: 1.79–17.05).[182]

There was no clear carcinogenic effect of rhGH in patients with growth failure without other major disease. However, bone and bladder cancer incidence were significantly higher in GH-treated patients without previous cancer. Cancer risk was unrelated to duration or cumulative dose of rhGH treatment, but for patients treated after previous cancer, cancer mortality risk increased significantly with increasing daily rhGH dose (*P* trend < 0.001). The incidence of Hodgkin lymphoma increased significantly with longer follow-up (*P* trend $= 0.001$ for patients overall and 0.002 for patients without previous cancer).[184]

These results were in contrast with those published concurrently in a joint report from Belgium, the Netherlands, and Sweden.[183] The increase in mortality rate was not confirmed, the deaths observed were mostly due to accidents, and none were attributable to cardiovascular disease, cancer, or high dose of rhGH in the range of 0.048 to 0.054 mg/kg/day.[183]

A recent report from the SAGhE study showed no significant increase in overall mortality in low-risk patients (IGHD, ISS).[185] Conversely, in patients with increased risk (multiple pituitary hormone deficiencies [MPHD], comorbidities), there was an increased cardiovascular and hematological mortality rate associated with the underlying conditions.[185] Mortality was not associated with mean daily or cumulative rhGH dose.[185] Similar findings have been reported in other studies.[186] The latter study examined mortality in 3847 patients with IGHD, ISS, or SGA treated with GH between 1985 and 2010. They developed a mortality model of the Swedish general population born from 1973 to 2010, adjusting for birth characteristics, sex, age intervals, congenital anomalies, and calendar year to estimate the SMR and to apply this model to assess expected deaths in Swedish rhGH-treated patients. The mortality ratio was 1.43 (95% CI: 0.89–2.19, $P = 0.14$), with an observed/expected death ratio of 21/14.68. Using the advanced mortality model, the SMR was 0.955 (95% CI: 0.591–1.456, $P = 0.95$). Hence, there was no difference in mortality once was the correction for birth characteristics and congenital abnormalities were applied.

A further study of the GENESIS cohort[187] examined mortality in 9504 GH-treated patients followed for more than 4 years (67,163 person-years of follow-up). Forty-two deaths were reported (SMR, 0.77; 95% CI: 0.56–1.05). The SMR was significantly elevated in patients with a history of malignant neoplasia (6.97; 95% CI: 3.81–11.69) and borderline elevated for those with other serious non-GHD conditions (2.47; 95% CI: 0.99–5.09). No significant increases in mortality were, however, observed for GH-treated children with idiopathic GHD, ISS, SGA, Turner syndrome, SHOX deficiency, or a history of benign neoplasia.

A position statement from European Society for Paediatric Endocrinology, Growth Hormone Research Society, and Pediatric Endocrine Society[188] suggested that many of the disorders treated with GH in children and adults have an inherently higher mortality risk related to the underlying disorder; in most studies, therefore, the potential impact of GH treatment on overall mortality is difficult to distinguish from the impact of the underlying disorder.

Available data in children do not indicate an increased risk of new primary cancers in GH recipients, but the numbers of subjects followed up in the long term are still relatively small, and the data are incomplete. The likelihood of developing glucose intolerance or overt T2D during GH treatment in pediatric patients with GH deficiency or ISS is generally very low. Although the lifetime risk of glucose intolerance and T2D in a number of conditions treated with GH, including Turner syndrome and short SGA children, is higher than in the background population, GH treatment does not increase the incidence of T2D in these conditions in the short term. The position statement recommends careful monitoring of IGF1, adrenal and thyroid function, glycated hemoglobin, and potential scoliosis in those on GH treatment.

Endpoints of Therapy. Given the discussion outlined above in terms of efficacy, safety, and clinical governance, Table 18.11 illustrates some suggested outcomes that can be monitored and used as a basis of short- and long-term audit, as well as safety monitoring.

Growth Hormone–Releasing Hormone and Growth Hormone Secretagogues

In children whose GHD is due to a hypothalamic abnormality, treatment with GHRH would appear to be an appropriate therapeutic option. Both GHRH$_{1\text{-}44}$ and GHRH$_{1\text{-}29}$ are biologically active in humans. Unfortunately, although GHRH can be absorbed nasally, this route of administration has not proved to be effective, and treatment must be via subcutaneous or intramuscular injection, as is the case with rhGH. Direct comparisons of rhGH and GHRH have not been performed, but a number of studies indicate that GHRH administered once or twice daily can increase the growth rates of children with GHD. No specific therapeutic advantage of GHRH over rhGH has been demonstrated to date, although further studies are warranted to investigate optimization of dosage and frequency of administration.[189]

Since the discovery of GHRPs, a variety of small GHRPs and nonpeptidyl small-molecule GH secretagogues have been manufactured. These molecules are potent stimulators of GH release, especially when administered together with GHRH, and may be active when administered by intravenous, intramuscular, subcutaneous, nasal, and oral routes. These potential advantages must be balanced by the likelihood that normal GHRH production is required to see maximal benefit from such agents. Clinical trials of their therapeutic use have proven disappointing.[189]

KEY POINTS

- Growth hormone deficiency is treated with GH doses in the order of 0.025-0.35 mg/kg/day. Monitoring of IGF1 concentrations may enable optimal titration of the dose. The response to GH treatment is highly variable and this is due to a number of factors, including adherence to treatment, and polymorphic variations in genes that dictate GH responsiveness. Optimal timing of GH treatment is critical in dictating final height outcomes. Whilst it is widely recognised that GH treatment is safe and effective in GHD, the SAGhE study has recently raised concerns, although these data were not confirmed in patients from other countries within the SAGhE cohort. The data nevertheless clearly underline the need for ongoing monitoring of patients on GH treatment.

Transition to Adult Care

During the transition period from childhood to adulthood, statural growth can be complete, but somatic development still progresses, and peak bone mass continues to change. Low or undetectable GH concentrations during the transitional and adult ages have been associated

TABLE 18.11 Short-, Medium-, and Long-Term Endpoints of Growth Hormone Therapy in Children with Growth Hormone Deficiency

	ENDPOINT	RATIONALE	MEASURE
Short-Term	1. Growth acceleration	1. Assess response	1. Minimum response of greater than 2 cm/yr
	2. Reduction in adipose mass	2. Assess response	2. Skinfold thickness
	3. Correct dose	3. Optimize therapy	3. IGF1 and growth
	4. Vision and headaches	4. Raised intracranial pressure	4. Fundoscopy
	5. Assessment of limp	5. Slipped femoral epiphysis	5. X-ray
Medium-Term	1. Bone maturation	1. Rate of skeletal maturation	1. Yearly bone age
	2. Pubertal status	2. Early puberty or rapid progression	2. Six-monthly Tanner staging
	3. Correct dose	3. Optimize therapy	3. IGF1 and growth response; return to target height within 6 years of therapy
	4. Thyroid status	4. Altered status or evolving endocrinopathy	4. Yearly thyroid function tests
	5. Other hormones	5. Evolving endocrinopathy	5. Gonadotroph and corticotroph function
	6. Metabolic status	6. Insulin insensitivity	6. Fasting glucose and insulin
Long-Term	1. Growth	1. Outcome	1. Final height within target height of parents
	2. Bone mineralization	2. GH effect on bone	2. DEXA scan
	3. Malignancy risk	3. ? GH cancer link	3. Cancer registry
	4. Cardiovascular risk	4. Insulin resistance or GH effect (? long-term GHD)	4. Fasting glucose and insulin, blood pressure, fasting lipids

DEXA, Dual-energy x-ray absorptiometry; *GHD*, growth hormone; *GHD*, growth hormone deficiency; *IGF*, insulin-like growth factor.

with alterations in the somatic composition, an increased fat-free mass with increased blood lipid concentrations, a compromised peak of bone mass, and a reduction in muscular strength.[190,191]

Furthermore, a negative effect of GH discontinuation during the transition period on quality of life, psychological well-being, and depression has been reported.[192]

Therefore, resuming rhGH replacement therapy in patients with confirmed persistent GHD during the transition period after achievement of final height is recommended, as most studies have reported long-term improvement in body composition, bone health, quality of life, and lipid metabolism in adulthood.[193]

While there is undoubtedly evidence in favor of GH continuation from a metabolic point of view, there is nevertheless the problem of establishing whether the state of GHD, diagnosed in childhood, still persists in adolescence and adulthood. Previous studies have shown that the vast majority of patients diagnosed with idiopathic GHD in childhood, who have no structural pituitary abnormality on MRI, demonstrated normalization of GH secretion when retested in late adolescence or adulthood (66%–85% depending on the series, the test, and cutoff used).[194] It is therefore recommended that patients with childhood-onset GHD be retested after a washout period of at least 1 month[259-261] to confirm the diagnosis of GHD when statural growth is completed (growth velocity <1.5–2 cm/year).[195]

Current guidelines suggest that the diagnosis of adult GHD can be made without the need for performing GH stimulation testing in certain patient cohorts, such as those patients with organic HP disease (e.g., suprasellar mass with previous surgery and cranial irradiation) and biochemical evidence of MPHD (greater than or equal to three pituitary hormone deficiencies [PHDs]), together with low serum IGF1 concentrations (<–2.0 SDS), genetic defects affecting the HP axis (mutation in a transcription factor, e.g., *POU1F1*, *PROP1*, *HESX1*, *LHX3*, or *LHX4* or mutation in *GH1* or *GHRHR*), and HP structural brain defects.[196] In patients with less than or equal to two PHDs, low serum IGF1 concentrations (<–2.0 SDS) alone are not sufficient to make a diagnosis of adult GHD, and clinicians should perform one GH stimulation test to confirm the diagnosis.

In transition, patients with isolated GHD and the presence of organic HP disease (e.g., craniopharyngioma, pituitary hypoplasia, ectopic posterior pituitary, or previous cranial irradiation), the number of GH stimulation tests to be undertaken should be guided by the degree of clinical suspicion for GHD. If clinical suspicion is high, one GH stimulation test is sufficient, but if clinical suspicion is low, then a second GH stimulation test should be performed.

Transition patients with idiopathic isolated GHD and those with low-normal (between 0 and –2 SDS) or low (<–2 SDS) serum IGF1 concentrations should be retested for GHD with GH stimulation tests after at least 1 month following discontinuation of rhGH therapy. The vast majority of patients diagnosed with GHD in childhood who have no structural pituitary abnormality on MRI will have adequate GH secretion when retested in late adolescence or adulthood (66%–85% depending on the series, the test, and cutoff used).

It should be noted that the thresholds for GH insufficiency differ between the various dynamic tests, and such cutoffs need to be critically appraised. For transition patients, a feasible and validated GH stimulation test has been less well studied. Nevertheless, the insulin tolerance test (ITT), glucagon, macimorelin, and arginine used alone or with GHRH have been used to confirm/reconfirm GHD in adolescence.[197,198]

In this patient population, the ITT (using a GH cutoff point of ≤5.0 mg/L) may be utilized, as it has a higher specificity (100%) and sensitivity (96%) in detecting permanent GHD.[197] If the test is contraindicated (seizure disorder, cardio-/cerebrovascular disease) or unfeasible to perform, the glucagon stimulation test (using a GH cutoff point of 3 μg/L for normal-weight [BMI <25 kg/m^2] patients and overweight [BMI 25–30 kg/m^2] patients with a high pretest probability and a lower GH cutoff point of 1 μg/L for overweight [BMI 25–30 kg/m^2] patients with a low pretest probability and obese [BMI >30 kg/m^2] patients) and the macimorelin test (using a GH cutoff point ≤2.8 mg/L) can be considered as alternative tests. Arginine and levodopa testing have not been systematically evaluated and validated. Although these tests have low sensitivity and specificity in adults and transition patients with suspected GHD, they are still used by some centers.[199]

For between 25% to 75% of patients, the GH response to provocation is in the normal range, as would be expected from the previous discussion on testing, and probability theory needs to be employed to determine whether therapy should be continued or not. In the remainder, continuation of GH therapy should be considered, and these patients should be followed up by an adult endocrinologist. In these individuals, adverse changes in body composition, quality of life, and bone mineral density may be an indication to recommence GH treatment, although it is less likely that these individuals will develop adult GHD syndrome. In those patients with MPHD, GHD due to a congenital lesion, and GHD secondary to radiotherapy, surgery, or a mass lesion, the GHD is highly unlikely to reverse.

Indeed, GHD in adults results in decreased quality of life, increased risk of bone fracture, increased low-density lipoprotein (LDL)-cholesterol, and decreased HDL-cholesterol concentrations.[200] While some question the efficacy of protecting against osteoporosis, most believe that replacement therapy protects against its development.

Throughout transition, rhGH treatment will allow patients to reach full skeletal maturation, an appropriate somatic development, and an increase in lean mass. Stopping treatment, while not recommended, should be at least accompanied by monitoring GH-dependent endpoints. Similarly, GH is needed for maintaining healthy body composition, as it has been shown that cessation of rhGH treatment leads to an increase in visceral body fat.[200] Multiple reports have suggested that untreated GHD may cause cognitive dysfunction, fatigue, reduced sense of well-being, and altered mood. Standard lipid profiles improve with rhGH, with an overall lowering of total cholesterol, decreased concentrations of LDL-cholesterol, and unchanged triglyceride concentrations.[199,201] While no major cardiac function abnormality has been observed after GH discontinuation, an improvement in markers of endothelial dysfunction and positive effects on left ventricular mass, intraventricular septum, diastolic function, and stroke volume index have been reported with rhGH therapy.[199,202,203] Long-term controlled trials evaluating the effects of GH replacement on cardiovascular mortality and morbidity are still lacking, yet no positive trend in cardiovascular risk has been determined to date.[204]

Treating patients during transition is via rhGH given subcutaneously daily, similar to the pediatric population. However, as GH secretion varies during a lifetime, the dosing should follow the pattern determined by age and gender, along with any comorbidities and estrogen status. Physicians should consider starting rhGH at low doses and gradually titrating upwards, with an aim to provide an optimal response with the least expected side effects without a specific weight-based regimen. For patients younger than 30 years, most guidelines recommend initiating a dose of 400 to 500 μg/day, with a mildly higher dose during transition, with an increase in daily dosing by 100 to 200 μg/day every 1 to 2 months based on the individual's response.[199] Serum IGF1 concentrations should be monitored every 4 to 6 weeks until achieving the optimal maintenance dose. A repeat follow-up should be measured every 6 to 12 months. There is no consensus on the optimal IGF1 concentration, but in general the goal is maintaining a normal concentration within age- and sex-specific ranges.

KEY POINTS

- GH secretion should be retested at the end of linear growth in those with isolated GHD and a structurally normal pituitary gland on MRI. However, it is not necessary in those with organic HP disease, and those with 3 or more hormone deficiencies or proven genetic defects in the HP axis. Continuation of rhGH in those with persistent GHD is indicated in order to improve body composition, bone health, lean body mass, lipid metabolism, and quality of life.

CONCLUSIONS

To conclude, the diagnosis of GHD can be far from straightforward, and involves a careful clinical and auxological evaluation, together with the use of biochemistry, neuroimaging, and possibly genetic testing, although the latter is often not widely available.

Significant advances have been made in the past decades in refining the diagnosis and determining the causes of childhood GHD while optimizing its treatment. Neuroimaging has become an integral component of the diagnosis and leads to identification of structural pituitary and brain abnormalities. This, in turn, can define patients who need to be screened for additional pituitary deficiencies, as well as those who may have permanent GHD and who require GH replacement therapy in adult life. Neuroimaging can also enable the targeting of appropriate molecular analysis, and this in turn may enable genetic counseling. The development of novel technologies such as next-generation and possibly whole-exome and -genome sequencing, is likely to identify further causes of GHD and hypopituitarism in the fullness of time. Various questions nevertheless remain to be answered, including the variable penetrance of genetic mutations, the considerable phenotypic variability, the role of environmental factors, and the interaction between candidate genes suggesting a significant role of digenicity/oligogenicity. The recent development of therapeutic long-acting GH preparations may offer another mode of therapy, and this may be more acceptable to those with other underlying conditions, as well as to adolescents and young adults. It will help in those families that struggle with adherence with daily injections. Nevertheless, a number of challenges remain, not least the poor reproducibility of GH stimulation tests and the complex underlying molecular basis in some cases.

REFEERENCES

1. Benonisdottir S, Oddsson A, Helgason A, et al. Epigenetic and genetic components of height regulation. *Nat Commun*. 2016;7:13490.
2. Jelenkovic Y, Yokoyama R, Sund, et al. Associations between birth size and later height from infancy through adulthood: an individual based pooled analysis of 28 twin cohorts participating in the CODATwins project. *Early Hum Dev*. 2018;120:53–60.
3. Wit JM, Kiess W, Mullis P. Genetic evaluation of short stature. *Best Pract Res Clin Endocrinol Metab*. 2011;25:1–17.
4. David A, Hwa V, Metherell LA, et al. Evidence of a continuum of genetic, phenotypic and biochemical abnormalities in children with growth hormone insensitivity. *Endocr Rev*. 2011;32:472–497.
5. Lindsay R, Feldkamp M, Harris D, et al. Utah Growth Study. Growth standards and the prevalence of growth hormone deficiency. *J Pediatr*. 1994;125:29–35.
6. Dasen JS, Rosenfeld MG. Signaling and transcriptional mechanisms in pituitary development. *Annu Rev Neurosci*. 2001;24:327–355.
7. Takuma N, Sheng HZ, Furuta Y, et al. Formation of Rathke's pouch requires dual induction from the diencephalon. *Development*. 1998;125:4835–4840.
8. Kelberman D, Rizzoti K, Lovell-Badge R, et al. Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev*. 2006;27:371–397.
9. Edwards W, Raetzman LT. Complex integration of intrinsic and peripheral signaling is required for pituitary gland development. *Biol Reprod*. 2018;99:504–513.
10. Ericson J, Norlin S, Jessell TM, et al. Integrated FGF and BMP signaling controls the progression of progenitor cell differentiation and the emergence of pattern in the embryonic anterior pituitary. *Development*. 1998;125:1005–1015.
11. Rosenfeld MG, Briata P, Dasen J, et al. Multistep signaling and transcriptional requirements for pituitary organogenesis in vivo. *Recent Prog Horm Res*. 2000;55:1–13.
12. Sheng HZ, Westphal H. Early steps in pituitary organogenesis. *Trends Genet*. 1999;15:236–240.
13. Gregory LC, Dattani MT. The molecular basis of congenital hypopituitarism and related disorders. *J Clin Endocrinol Metab*. 2020;105:dgz184.
14. Nogami H, Hisano S. Functional maturation of growth hormone cells in the anterior pituitary gland of the fetus. *Growth Horm IGF Res*. 2008;18:379–388.
15. Davis SW, Mortensen AH, Camper SA. Birthdating studies reshape models for pituitary gland cell specification. *Dev Biol*. 2011;352: 215–227.
16. Le Tissier PR, Hodson DJ, Lafont C, et al. Anterior pituitary cell networks. *Front Neuroendocrinol*. 2012;33:252–266.
17. Bonnefont X, Lacampagne A, Sanchez-Hormigo A, et al. Revealing the large-scale network organization of growth hormone-secreting cells. *Proc Natl Acad Sci USA*. 2012;102:16880–16885.
18. Fang Q, George AS, Brinkmeier ML, et al. Genetics of combined pituitary hormone deficiency: roadmap into the genome era. *Endocr Rev*. 2016;37:636–675.
19. Webb EA, Dattani MT. Septo-optic dysplasia. *Eur J Hum Genet*. 2010;18:393–397.
20. McNay DE, Turton JP, Kelberman D, et al. HESX1 mutations are an uncommon cause of septooptic dysplasia and hypopituitarism. *J Clin Endocrinol Metab*. 2007;92:691–697.
21. Carvalho LR, Woods KS, Mendonca BB, et al. A homozygous mutation in HESX1 is associated with evolving hypopituitarism due to impaired repressor-corepressor interaction. *J Clin Invest*. 2003;112:1192–1201.
22. Sobrier ML, Bitach T, Netchine I, et al. Pathophysiology of syndromic combined pituitary hormone deficiency due to a LHX3 defect in light of LHX3 and LHX4 expression during early human development. *Gene Expr Patterns*. 2004;5:279–284.
23. Mortensen AH, Schade V, Lamonerie T, et al. Deletion of OTX2 in neural ectoderm delays anterior pituitary development. *Hum Mol Genet*. 2015;24:939–953.
24. Dateki S, Kosaka K, Hasegawa K, et al. Heterozygous orthodenticle homeobox 2 mutations are associated with variable pituitary phenotype. *J Clin Endocrinol Metab*. 2010;95:756–764.
25. Ashkenazi-Hoffnung L, Lebenthal Y, Wyatt AW, et al. A novel loss-of-function mutation in OTX2 in a patient with anophthalmia and isolated growth hormone deficiency. *Hum Genet*. 2010;127:721–729.
26. Tajima T, Ohtake A, Hoshino M, et al. OTX2 loss of function mutation causes anophthalmia and combined pituitary hormone deficiency with a small anterior and ectopic posterior pituitary. *J Clin Endocrinol Metab*. 2009;94:314–319.
27. Pfaeffle RW, Hunter CS, Savage JJ, et al. Three novel missense mutations within the LHX4 gene are associated with variable pituitary hormone deficiencies. *J Clin Endocrinol Metab*. 2008;93:1062–1071.
28. Gregory LC, Humayun KN, Turton JPG, et al. Novel lethal form of congenital hypopituitarism associated with the first recessive *LHX4* mutation. *J Clin Endocrinol Metab*. 2015;100:2158–2164.
29. Jullien N, Romanet P, Philippon M, et al. Heterozygous LHX3 mutations may lead to a mild phenotype of combined pituitary hormone deficiency. *Eur J Hum Genet*. 2019;27:216–225.
30. Kelberman D, de Castro SC, Huang S, et al. SOX2 plays a critical role in the pituitary, forebrain and eye during human embryonic development. *J Clin Endocrinol Metab*. 2008;93:1865–1873.
31. Alatzoglou KS, Andoniadou CL, Kelberman D, et al. SOX2 haploinsufficiency is associated with slow progressing hypothalamo-pituitary tumour. *Hum Mutat*. 2011;32:1376–1380.
32. Laumonnier F, Ronce N, Hamel BC, et al. Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am J Hum Genet*. 2002;71:1450–1455.

33. Izumi Y, Suzuki E, Kanzaki S, et al. Genome-wide copy number analysis and systematic mutation screening in 58 patients with hypogonadotropic hypogonadism. *Fertil Steril*. 2014;102:1130–1136.e3.
34. Alatzoglou KS, Kelberman D, Cowell CT, et al. Increased transactivation associated with SOX3 polyalanine tract deletion in a patient with hypopituitarism. *J Clin Endocrinol Metab*. 2011;96:E685–E690.
35. Abouzeid H, Youssef MA, Bayoumi N, et al. RAX and anophthalmia in humans: evidence of brain anomalies. *Mol Vis*. 2012;18:1449–1456.
36. Brachet C, Kozhemyakina EA, Balikova I, et al. Truncating RAX mutations: anophthalmia, hypopituitarism, diabetes insipidus, and cleft palate in mice and men. *J Clin Endocrinol Metab*. 2019;104:2925–2930.
37. Liu Z, Chen X. A novel missense mutation in human receptor roundabout-1 *(ROBO1)* gene associated with pituitary stalk interruption syndrome. *J Clin Res Pediatr Endocrinol*. 2020;12:212–217.
38. Webb EA, AlMutair A, Kelberman, et al. ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. *Brain*. 2013;136:3096–3105.
39. Giri D, Vignola ML, Gualtieri A, et al. Novel *FOXA2* mutation causes hyperinsulinism, hypopituitarism with craniofacial and endoderm-derived organ abnormalities. *Hum Mol Genet*. 2017;26:4315–4326.
40. Gregory LC, Shah P, Sanner JRF, et al. Mutations in *MAGEL2* and *L1CAM* are associated with congenital hypopituitarism and arthrogryposis. *J Clin Endocrinol Metab*. 2019;104:5737–5750.
41. Gregory LC, Ferreira CB, Young-Baird SK, et al. Impaired EIF2S3 function associated with a novel phenotype of X-linked hypopituitarism with glucose dysregulation. *EBioMedicine*. 2019;42:470–480.
42. Hietamäki J, Gregory LC, Ayoub S, et al. Loss-of-function variants in TBC1D32 underlie syndromic hypopituitarism. *J Clin Endocrinol Metab*. 2020;105:1748–1758.
43. Tommiska J, Känsäkoski J, Skibsbye L, et al. Two missense mutations in KCNQ1 cause pituitary hormone deficiency and maternally inherited gingival fibromatosis. *Nat Commun*. 2017;8:1289.
44. Gaston-Massuet C, McCabe MJ, Scagliotti V, et al. Transcription factor 7-like 1 is involved in hypothalamo–pituitary axis development in mice and humans. *Proc Natl Acad Sci U S A*. 2016;113:E548–E557.
45. Cohen M. Holoprosencephaly: clinical, anatomic, and molecular dimensions. *Birth Defeects Res A Clin Mol Teratol*. 2006;76:658–673.
46. Bashamboo A, Bignon-Topalovic J, Rouba H, et al. A nonsense mutation in the hedgehog receptor CDON associated with pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*. 2016;101:12–15.
47. Kahn BM, Corman TS, Lovelace K, et al. Prenatal ethanol exposure in mice phenocopies Cdon mutation by impeding Shh function in the etiology of optic nerve hypoplasia. *Dis Model Mech*. 2017;10:29–37.
48. Karaca E, Buyukkaya R, Pehlivan D, et al. Whole- exome sequencing identifies homozygous GPR161 mutation in a family with pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*. 2015;100:E140–E147.
49. Lucas-Herald AK, Kinning E, Iida A, et al. A case of functional growth hormone deficiency and early growth retardation in a child with IFT172 mutations. *J Clin Endocrinol Metab*. 2015;100:1221–1224.
50. Bredrup C, Saunier S, Oud MM, et al. Ciliopathies with skeletal anomalies and renal insufficiency due to mutations in the IFT-A gene WDR19. *Am J Hum Genet*. 2011;89:634–643.
51. Romano S, Maffei P, Bettini V, et al. Alstrom syndrome is associated with short stature and reduced GH reserve. *Clin Endocrinol*. 2013;79:529–536.
52. Franca MM, Jorge AA, Alatzoglou KS, et al. Absence of GH-releasing hormone (GHRH) mutations in selected patients with isolated GH deficiency. *J Clin Endocrinol Metab*. 2011;96:E1457–E1460.
53. Carakushansky M, Whatmore AJ, Clayton PE, et al. A new missense mutation in the growth hormone-releasing hormone receptor gene in familial isolated GH deficiency. *Eur J Endocrinol*. 2003;148:25–30.
54. Alatzoglou KS, Turton JP, Kelberman D, et al. Expanding the spectrum of mutations in GH1 and GHRHR: genetic screening in a large cohort of patients with congenital isolated growth hormone deficiency. *J Clin Endocrinol Metab*. 2009;94:3191–3199.
55. Godi M, Mellone S, Petri A, et al. A recurrent signal peptide mutation in the growth hormone releasing hormone receptor with defective translocation to the cell surface and isolated growth hormone deficiency. *J Clin Endocrinol Metab*. 2009;94:3939–3947.
56. Alba M, Hall CM, Whatmore AJ, et al. Variability in anterior pituitary size within members of a family with GH deficiency due to a new splice mutation in the GHRH receptor gene. *Clin Endocrinol*. 2004;60:470–475.
57. Gregory LC, Alatzoglou KS, McCabe MJ, et al. Partial loss of function of the GHRH receptor leads to mild growth hormone deficiency. *J Clin Endocrinol Metab*. 2016:3608–3615.
58. Sornson MW, Wu W, Daser JS, et al. Pituitary lineage determination by the Prophet of Pit-1 homeodomain factor defective in Ames dwarfism. *Nature*. 1996;384:327–333.
59. Pernasetti F, Toledo SP, Vasilyev VV, et al. Impaired adrenocorticotropin-adrenal axis in combined pituitary hormone deficiency caused by a two-base pair deletion (301–302delAG) in the prophet of PIT1 gene. *J Clin Endocrinol Metab*. 2000;85:390–397.
60. Prince KL, Walvoord EC, Rhodes SJ. The role of homeodomain transcription factors in heritable pituitary disease. *Nat Rev Endocrinol*. 2011;7:727–737.
61. Turton J, Mehta AM, Raza J, et al. Mutations within the transcription factor PROP1 are rare in a cohort of patients with sporadic Combined Pituitary Hormone Deficiency (CPHD). *Clin Endocrinol (Oxf)*. 2005;63:10–18.
62. Arroyo A, Pernasetti F, Vasilyev VV, et al. A unique case of combined pituitary hormone deficiency caused by a PROP1 gene mutation (R120C) associated with normal height and absent puberty. *Clin Endocrinol*. 2002;57:283–291.
63. Pfaffle R, Klammt J. Pituitary transcription factors in the aetiology of combined pituitary hormone deficiency. *Best Pract Res Clin Endocrinol Metab*. 2011;25:43–60.
64. Mendonca BB, Osorio MG, Latronico AC, et al. Longitudinal hormonal and pituitary imaging changes in two females with combined pituitary hormone deficiency due to deletion of A301, G302 in the PROP1 gene. *J Clin Endocrinol Metab*. 1999;84:942–945.
65. Turton JP, Reynaud R, Mehta A, et al. Novel mutations within the POU1F1 gene associated with variable combined pituitary hormone deficiency (CPHD). *J Clin Endocrinol Metab*. 2005;90:4762–4770.
66. Mullis PE. Genetics of growth hormone deficiency. *Endocrinol Metab Clin North Am*. 2007;36:17–36.
67. Leiberman E, et al. Short stature in carriers of recessive mutation causing familial isolated growth hormone deficiency. *Am J Med Genet*. 2000;90:188–192.
68. Moseley C, Mullis P, Prince M, et al. An exon splice enhancer mutation causes autosomal dominant GH deficiency. *J Clin Endocrinol Metab*. 2002;87:847–852.
69. McGuiness L, Magoulas C, Mathers K, et al. Autosomal dominant growth hormone deficiency disrupts secretory vesicles: in vitro and in vivo studies in transgenic mice. *Endocrinology*. 2003;144:720–731.
70. Ryther RCC, McGuiness LM, Phillips III JA, et al. Disruption of exon definition produces a dominant-negative growth hormone isoform that causes somatotroph death and IGHD II. *Hum Genet*. 2003;113:140–148.
71. Mullis PE, Robinson IC, Salemi S, et al. Isolated autosomal dominant growth hormone deficiency: an evolving pituitary deficit? A multicenter follow-up study. *J Clin Endocrinol Metab*. 2005;90:2089–2096.
72. Takahashi Y, Shirono H, Arisaka O, et al. Biologically inactive growth hormone caused by an amino acid substitution. *J Clin Invest*. 1997;100:1159–1165.
73. Petkovic V, Thevis M, Lochmatter D, et al. GH mutant (R77C) in a pedigree presenting with the delay of growth and pubertal development: structural analysis of the mutant and evaluation of the biological activity. *Eur J Endocrinol*. 2007;157:S67–S74.
74. Petkovic V, Eble A, Pandey AV, et al. A novel GH-1 gene mutation (GH-P59L) causes partial GH deficiency type II combined with bioinactive GH syndrome. *Growth Horm IGF Res*. 2011;21:160–166.
75. Petkovic V, Godi M, Pandey AV, et al. Growth hormone (GH) deficiency type II: a novel GH-1 gene mutation (GH-R178H) affecting secretion and action. *J Clin Endocrinol Metab*. 2010;95:731–739.
76. Argente J, Flores R, Gutiérrez-Arumí A, et al. Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. *EMBO Mol Med*. 2014;6:299–306.
77. Verberne EA, Faries S, Mannens MMAM, et al. Expanding the phenotype of biallelic RNPC3 variants associated with growth hormone deficiency. *Am J Med Genet*. 2020;182:1952–1956.

78. Maghnie M, Lindberg A, Koltowska-Häggström M, et al. Magnetic resonance imaging of CNS in 15,043 children with GH deficiency in KIGS (Pfizer international growth database). *Eur J Endocrinol*. 2013;168:211–217.
79. Argyopoulou M, Perignon F, Brauner R, et al. Magnetic resonance imaging in the diagnosis of growth hormone deficiency. *J Pediatr*. 1992;120:886–891.
80. Lenz AM, Root AW. Empty sella syndrome. *Pediatr Endocrinol Rev*. 2012;9:710–715.
81. Albertsson-Wikland K, Niklasson A, Karlberg P. Birth data for patients who later develop growth hormone deficiency: preliminary analysis of a national register. *Acta Paediatr*. 1990;370:115–120.
82. Karavitaki N, Cudlip S, Adams CB, et al. Craniopharyngiomas. *Endocr Rev*. 2006;27:371–397.
83. Muller HL, Emser A, Faldum A, et al. Longitudinal study on growth and body mass index before and after diagnosis of childhood craniopharyngioma. *J Clin Endocrinol Metab*. 2004;89:3298–3305.
84. Sekine S, Shibata T, Kokubu A, et al. Craniopharyngiomas of adamantinomatous type harbor beta-catenin gene mutations. *Am J Pathol*. 2002;161:1997–2001.
85. Gaston-Massuet C, Andoniadou CL, Signore M, et al. Increased Wingless (Wnt) signaling in pituitary progenitor/stem cells gives rise to pituitary tumors in mice and humans. *Proc Natl Acad Sci U S A*. 2011;108:11482–11487.
86. Calao AM. *Pituitary Adenomas in Childhood*. MDText.com, Inc., South Dartmouth (MA), 2015.
87. Jezierska M, Stefanowicz J, Romanowicz G, et al. Langerhans cell histiocytosis in children – a disease with many faces. Recent advances in pathogenesis, diagnostic examinations and treatment. *Postepy Dermatol Alergol*. 2018;35:6–17.
88. Donadieu J, Rolon MA, Thomas C, et al. Endocrine involvement in pediatric-onset Langerhans' cell histiocytosis: a population-based study. *J Pediatr*. 2004;144(3):344–350.
89. Kaltsas GA, Powles TB, Evanson J, et al. Hypothalamo-pituitary abnormalities in adult patients with Langerhans cell histiocytosis: clinical, endocrinological, and radiological features and response to treatment. *J Clin Endocrinol Metab*. 2000;85:1370–1376.
90. Dhall G, Finlay JL, Dunkel IJ, et al. Analysis of outcome for patients with mass lesions of the central nervous system due to Langerhans cell histiocytosis treated with 2-chlorodeoxyadenosine. *Pediatr Blood Cancer*. 2008;50:72–79.
91. Morimoto A, Ikushima S, Kinugawa N, et al. Improved outcome in the treatment of pediatric multifocal Langerhans cell histiocytosis: results from the Japan Langerhans Cell Histiocytosis Study Group-96 protocol study. *Cancer*. 2006;107:613–619.
92. Darzy KH, Shalet SM. Hypopituitarism following radiotherapy. *Pituitary*. 2009;12:40–50.
93. Sklar C, Mertens A, Walter A, et al. Final height after treatment for childhood acute lymphoblastic leukemia: comparison of no cranial irradiation with 1800 and 2400 centigrays of cranial irradiation. *J Pediatr*. 1993;123:56–64.
94. Balthasar N, Mery PF, Magoulas CB, et al. Growth hormone-releasing hormone (GHRH) neurons in GHRH-enhanced green fluorescent protein transgenic mice: a ventral hypothalamic network. *Endocrinology*. 2003;144:2728–2740.
95. Sklar CA, Sarafoglou K, Whittam E. Effects of insulin-like growth factor binding protein 3 in predicting the growth hormone response to provocative testing in children treated with cranial irradiation. *Acta Endocrinol*. 1993;129:511–515.
96. Quigley C, Cowell C, Jimenez M, et al. Normal or early development of puberty despite gonadal damage in children treated for acute lymphoblastic leukemia. *N Engl J Med*. 1989;321:143–151.
97. Acerini CL, Tasker RC, Bellone S, et al. Hypopituitarism in childhood and adolescence following traumatic brain injury: the case for prospective endocrine investigation. *Eur J Endocrinol*. 2006;155:663–669.
98. Oertel M, Boscardin WJ, Obrist WD, et al. Posttraumatic vasospasm: the epidemiology, severity, and time course of an underestimated phenomenon: a prospective study performed in 299 patients. *J Neurosurg*. 2005;103:812–824.
99. Agha A, Thompson CJ. Anterior pituitary dysfunction following traumatic brain injury (TBI). *Clin Endocrinol*. 2006;64:481–488.
100. Tanriverdi F, Senyurek H, Unluhizarci K, et al. High risk of hypopituitarism after traumatic brain injury: a prospective investigation of anterior pituitary function in the acute phase and 12 months after trauma. *J Clin Endocrinol Metab*. 2006;91:2105–2111.
101. Ortiz JB, Sukhina A, Balkan B, et al. Epidemiology of pediatric traumatic brain injury and hypothalamic-pituitary disorders in Arizona. *Front Neurol*. 2019;10:1410.
102. Niederland T, Makovi H, Gal V, et al. Abnormalities of pituitary function after traumatic brain injury in children. *J Neurotrauma*. 2007;24:119–127.
103. Einaudi S, Bondone C. The effects of head trauma on hypothalamic-pituitary function in children and adolescents. *Curr Opin Pediatr*. 2007;19:465–470.
104. Tan CL, Alavi SA, Baldeweg SE, et al. The screening and management of pituitary dysfunction following traumatic brain injury in adults: British Neurotrauma Group guidance. *J Neurol Neurosurg Psychiatry*. 2017;88:971–981.
105. Casano-Sancho P, Suárez L, Ibáñez L, et al. Pituitary dysfunction after traumatic brain injury in children: is there a need for ongoing endocrine assessment? *Clin Endocrinol*. 2013.
106. Wamstad J, Norwood KW, Rogol AD, et al. Neuropsychological recovery and quality-of-life in children and adolescents with growth hormone deficiency following TBI: a preliminary study. *Brain Inj*. 2013;27:200–208.
107. Schaefer S, Boegershausen N, Meyer S, et al. Hypothalamic-pituitary insufficiency following infectious diseases of the central nervous system. *Eur J Endocrinol*. 2008;15:3–9.
108. Faje A. Hypophysitis: evaluation and management. *Clin Diabetes Endocrinol*. 2016;2:15.
109. Iguchi G, Takeno R, Okimura Y, et al. Adult combined GH, prolactin, and TSH deficiency associated with circulating PIT-1 antibody in humans. *J Clin Invest*. 2011;121:113–119.
110. Fung EB, Harmatz PR, Lee PD, et al. Increased prevalence of iron-overload associated endocrinopathy in thalassaemia versus sickle-cell disease. *Br J Haematol*. 2006;135:574–582.
111. De Sanctis V, Ashraf TS, Duran C, et al. An ICET-A survey on occult and emerging endocrine complications in patients with β-thalassemia major: conclusions and recommendations. *Acta Biomed*. 2018;89:481–489.
112. Tan LA, Sandler A, Todorova-Koteva K, et al. Recovery of pituitary function following treatment of an unruptured giant cavernous carotid aneurysm using Surpass flow-diverting stents. *BMJ Case Rep*. 2014;bcr2014011233.
113. Jagtap VS, Sarathi V, Lila AR, et al. Hyperphagic short stature: a case report and review of literature. *Indian J Endocrinol Metab*. 2012;16:624–626.
114. Kojima M, Kangawa K. Ghrelin. Structure and function. *Physiol Rev*. 2005;85:495–522.
115. Shewchuk BM, Asa SL, Cooke NE, et al. Pit-1 binding sites at the somatotrope-specific DNase-I hypersensitive sites I, II of the human growth hormone locus control region are essential for in-vivo hGH-N gene activation. *J Biol Chem*. 1999;274:35725–35733.
116. Yang X, Jin Y, Cattini PA. Appearance of the pituitary factor Pit-1 increases chromatin remodeling at hypersensitive site III in the human GH locus. *J Mol Endocrinol*. 2010;45:19–32.
117. Baumann G. Growth hormone heterogeneity: genes, isohormones, variants, and binding proteins. *Endocr Rev*. 1991;12:424–449.
118. Rosenfeld RG, Belgorosky A, Camacho-Hubner C, et al. Defects in growth hormone receptor signaling. *Trends Endocrinol Metab*. 2007;18:134–141.
119. Vottero A, Guzzetti C, Loche S. New aspects of the physiology of the GH-IGF-1 axis. *Endocr Dev*. 2013;24:96–105.
120. Savage MO. Phenotypes, investigation and treatment of primary IGF-1 deficiency. *Endocr Dev*. 2013;24:138–149.
121. Netchine I, Azzi S, Le Bouc Y, et al. IGF1 molecular anomalies demonstrate its critical role in fetal, postnatal growth and brain development. *Best Pract Res Clin Endocrinol Metab*. 2011;25:181–190.
122. Alatzoglou KS, Dattani MT. Genetic forms of hypopituitarism and their manifestation in the neonatal period. *Early Hum Dev*. 2009;85:705–712.
123. Wit JM, Oostdijk W. Evaluation of asymptomatic short children. *J Pediatr*. 2013;163:1534–1535.

124. Wit JM, Kamp G, Rikken B. Spontaneous growth and response to growth hormone treatment in children with growth hormone deficiency and idiopathic short stature. *Pediatr Res*. 1996;39:295–302.
125. Growth Hormone Research Society. Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. *J Clin Endocrinol Metab*. 2000;85:3990–3993.
126. Stanley T. Diagnosis of growth hormone deficiency in childhood. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:47–52.
127. Hussain K, Hindmarsh P, Aynsley Green A. Spontaneous hypoglycemia in childhood is accompanied by paradoxically low serum growth hormone and appropriate cortisol counterregulatory hormonal responses. *J Clin Endocrinol Metab*. 2003;88:3715–3723.
128. Binder G, Weidenkeller M, Blumenstock G, et al. Rational approach to the diagnosis of severe growth hormone deficiency in the newborn. *J Clin Endocrinol Metab*. 2010;95:2219–2226.
129. Di Iorgi N, Allegri AE, Napoli F, et al. The use of neuroimaging for assessing disorders of pituitary development. *Clin Endocrinol*. 2012;76:161–176.
130. Mitchell H, Dattani MT, Nanduri V, et al. Failure of IGF-1 and IGFBP-3 to diagnose growth hormone insufficiency. *Arch Dis Child*. 1999;80:443–447.
131. Cole TJ, Hindmarsh PC, Dunger DB. Growth hormone (GH) provocation tests and the response to GH treatment in GH deficiency. *Arch Dis Child*. 2004;89:1024–1027.
132. Suri D, Hindmarsh PC, Matthews DR, et al. The pituitary gland is capable of responding to two successive doses of growth hormone-releasing hormone (GHRH). *Clin Endocrinol*. 1991;34:13–17.
133. Loche S, Bizzarri C, Maghnie M, et al. Results of early reevaluation of growth hormone secretion in short children with apparent growth hormone deficiency. *J Pediatr*. 2002;140:445–449.
134. Gelwane G, Garel C, Chevenne D, et al. Subnormal serum insulin-like growth factor-I levels in young adults with childhood-onset nonacquired growth hormone (GH) deficiency who recover normal gh secretion may indicate less severe but persistent pituitary failure. *J Clin Endocrinol Metab*. 2007;92:3788–3795.
135. Misra M, Bredella MA, Tsai P, et al. Lower growth hormone and higher cortisol are associated with greater visceral adiposity, intramyocellular lipids, and insulin resistance in overweight girls. *Am J Physiol Endocrinol Metab*. 2008;295:E385–E392.
136. Rose SR, Ross JL, Uriarte M, et al. The advantage of measuring stimulated as compared with spontaneous growth hormone levels in the diagnosis of growth hormone deficiency. *N Engl J Med*. 1988;319:201–207.
137. Lanes R. Diagnostic limitations of spontaneous growth hormone measurements in normally growing prepubertal children. *Am J Dis Child*. 1989;143:1284–1286.
138. Skinner AM, Clayton PE, Price DA, et al. Urinary growth hormone excretion in the assessment of children with disorders of growth. *Clin Endocrinol (Oxf)*. 1993;39:201–206.
139. Scacchi M, Pincelli AI, Cavagnini F. Growth hormone in obesity. *Int J Obes Relat Metab Disord*. 1999;23:260–271.
140. Marcovecchio LM, Chiarelli F. Obesity and growth during childhood and puberty. *World Rev Nutr Diet*. 2013;106:135–141.
141. Richmond EJ, Rogol AD. Growth hormone deficiency in children. *Pituitary*. 2008;11:115–120.
142. Lee HS, Hwang JS. Influence of body mass index on growth hormone responses to classic provocative tests in children with short stature. *Neuroendocrinology*. 2011;93:259–264.
143. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *J Am Med Assoc*. 1999;282:1061–1066.
144. Ware J. The limitations of risk factors as prognostic tools. *N Engl J Med*. 2006;355:2615–2617.
145. Marin G, Domene HM, Barnes KM, et al. The effects of estrogen priming and puberty on the growth hormone response to standardized treadmill exercise and arginine-insulin in normal girls and boys. *J Clin Endocrinol Metab*. 1994;79:537–541.
146. Clemmons DR. Consensus statement on the standardization and evaluation of growth hormone and insulin-like growth factor assays. *Clin Chem*. 2011;57:555–559.
147. Clemmons DR. IGF-1 assays: current assay methodologies and their limitations. *Pituitary*. 2007;10:121–128.
148. Bernasconi S, Volta C, Cozzini A, et al. GH response to GHRH, insulin, clonidine and arginine after GHRH pretreatment in children. *Acta Endocrinol*. 1992;126:105–108.
149. Gasco V, Beccuti G, Baldini C, et al. Acylated ghrelin as a provocative test for the diagnosis of GH deficiency in adults. *Eur J Endocrinol*. 2012;168:23–30.
150. Maghnie M, Pennati MC, Civardi E, et al. GH response to ghrelin in subjects with congenital GH deficiency: evidence that ghrelin action requires hypothalamic-pituitary connections. *Eur J Endocrinol*. 2007;156:449–454.
151. Binder G, et al. Diagnosis of severe growth hormone deficiency in the newborn. *Clin Endocrinol*. 2020.
152. Grimberg A, et al. Guidelines for growth hormone and insulin-like growth factor-I treatment in children and adolescents: growth hormone deficiency, idiopathic short stature, and primary insulin-like growth factor-I deficiency. *Horm Res Paediatr*. 2016;86:361–397.
153. Wilton P, Mattsson AF, Darendeliler F. Growth hormone treatment in children is not associated with an increase in the incidence of cancer: experience from KIGS (Pfizer International Growth Database). *J Pediatr*. 2010;157:265–270.
154. Child CJ, Zimmermann AG, Scott RS, et al. Prevalence and incidence of diabetes mellitus in GH-treated children and adolescents: analysis from the GeNeSIS observational research program. *J Clin Endocrinol Metab*. 2011;96:E1025–E1034.
155. Reiter EO, Price DA, Wilton P, et al. Effect of growth hormone (GH) treatment on the near-final height of 1258 patients with idiopathic GH deficiency: analysis of a large international database. *J Clin Endocrinol Metab*. 2006;91:2047–2054.
156. Hindmarsh PC, Dattani MT. Use of growth hormone in children. *Nat Clin Pract Endocrinol Metab*. 2006;2:260–268.
157. Cohen P, Bright GM, Rogol AD, et al. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. *J Clin Endocrinol Metab*. 2002;87:90–98.
158. Cohen P, Weng W, Rogol AD, et al. Dose-sparing and safety-enhancing effects of an IGF-I-based dosing regimen in short children treated with growth hormone in a 2-year randomized controlled trial: therapeutic and pharmacoeconomic considerations. *Clin Endocrinol*. 2014;81:71–76.
159. Pantel J, Grulich-Henn J, Bettendorf M, et al. Heterozygous nonsense mutation in exon 3 of the growth hormone receptor (GHR) in severe GH insensitivity (Laron syndrome) and the issue of the origin and function of the GHRd3 isoform. *J Clin Endocrinol Metab*. 2003;88:1705–1710.
160. Dos Santos C, Essioux L, Teinturier C, et al. A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat Genet*. 2004;36:720–724.
161. Jorge AA, Marchisotti FG, Montenegro LR, et al. Growth hormone (GH) pharmacogenetics: influence of GH receptor exon 3 retention or deletion on first-year growth response and final height in patients with severe GH deficiency. *J Clin Endocrinol Metab*. 2006;91:1076–1080.
162. Räz B, Janner M, Petkovic V, et al. Influence of growth hormone (GH) receptor deletion of exon 3 and full-length isoforms on GH response and final height in patients with severe GH deficiency. *J Clin Endocrinol Metab*. 2008;93:974–980.
163. Boguszewski CL, Barbosa EJL, Svensson PA, et al. Mechanisms in endocrinology: clinical and pharmacogenetic aspects of the growth hormone receptor polymorphism. *Eur J Endocrinol*. 2017;177:R309–R321.
164. Bidlingmaier M, Kim J, Savoy C, et al. Comparative pharmacokinetics and pharmacodynamics of a new sustained-release growth hormone (GH), LB03002, versus daily GH in adults with GH deficiency. *J Clin Endocrinol Metab*. 2006;91:2926–2930.
165. Peter F, Savoy C, Ji HJ, et al. Pharmacokinetic and pharmacodynamic profile of a new sustained-release GH formulation, LB03002, in children with GH deficiency. *Eur J Endocrinol*. 2009;160:349–355.
166. Miller BS, Velazquez E, Yuen KCJ. Long-acting growth hormone preparations - current status and future considerations. *J Clin Endocrinol Metab*. 2019;pii:dgz149.
167. Ranke MB, Lindberg A. Observed and predicted growth responses in prepubertal children with growth disorders: guidance of growth

hormone treatment by empirical variables. *J Clin Endocrinol Metab*. 2010;95:1229–1237.

168. Ranke MB, Lindberg A, Mullis PE, et al. Towards optimal treatment with growth hormone in short children and adolescents: evidence and theses. *Horm Res Paediatr*. 2013;79:51–67.
169. Price DA, Ranke MB. Final height following growth hormone treatment. In: Ranke MB, Gunnarsson R, eds. *Progress in Growth Hormone Therapy: 5 Years of KIGS*. Mannheim: JJ Verlag; 1994:574–579.
170. Fujieda K, Tanaka T, Takano K, et al. Adult height after growth hormone treatment in Japanese children with idiopathic growth hormone deficiency: analysis from the KIGS Japan database. *J Pediatr Endocrinol Metab*. 2011;24:457–462.
171. Polak M, Blair J, Kotnik P, et al. Early growth hormone treatment start in childhood growth hormone deficiency improves near adult height: analysis from NordiNet® International Outcome Study. *Eur J Endocrinol*. 2017;177:421–429.
172. Colmenares A, González L, Gunczler P, et al. Is the growth outcome of children with idiopathic short stature and isolated growth hormone deficiency following treatment with growth hormone and a luteinizing hormone-releasing hormone agonist superior to that obtained by GH alone. *J Pediatr Endocrinol Metab*. 2012;25:651–657.
173. Coupaye M, Lorenzini F, Lloret-Linares C, et al. Growth hormone therapy for children and adolescents with Prader-Willi syndrome is associated with improved body composition and metabolic status in adulthood. *J Clin Endocrinol Metabol*. 2013;98:E328–E335.
174. Salerno M, Esposito V, Farina V, et al. Improvement of cardiac performance and cardiovascular risk factors in children with GH deficiency after two years of GH replacement therapy: an observational, open, prospective, case-control study. *J Clin Endocrinol Metab*. 2005;91:1288–1295.
175. Nyberg F, Hallberg M. Growth hormone and cognitive function. *Nat Rev Endocrinol*. 2013;9:357–365.
176. Webb EA, O'Reilly MA, Clayden JD, et al. Effect of growth hormone deficiency on brain structure, motor function and cognition. *Brain*. 2012;135:216–227.
177. Blethen SL, Alster DK, Graves D, et al. Safety of recombinant DNA–derived growth hormone (rhGH): the National Cooperative Growth Study experience. *J Clin Endocrinol Metab*. 1996;81:1704–1710.
178. Bell J, Parker KL, Swinford RD, et al. Long-term safety of recombinant human growth hormone in children. *J Clin Endocrinol Metab*. 2010.
179. Ergun-Longmire B, Mertens AC, Mitby P, et al. Growth hormone treatment and risk of second neoplasms in the childhood cancer survivor. *J Clin Endocrinol Metab*. 2006;91:3494–3498.
180. Shen L, Sun CM, Li XT, et al. Growth hormone therapy and risk of recurrence/progression in intracranial tumors: a meta-analysis. *Neurol Sci*. 2015;36:1859–1867.
181. Gallagher EJ, LeRoith D. Minireview. IGF, insulin, and cancer. *Endocrinology*. 2011;152:2546–2551.
182. Carel JC, Ecosse E, Landier F, et al. Long-term mortality after recombinant growth hormone treatment for isolated growth hormone deficiency or childhood short stature: preliminary report of the French SAGhE study. *J Clin Endocrinol Metab*. 2012;97:416–425.
183. Savendahl L, Maes M, Albertsson-Wikland K, et al. Long-term mortality and causes of death in isolated GHD, ISS, and SGA patients treated with recombinant growth hormone during childhood in Belgium, The Netherlands, and Sweden: preliminary report of 3 countries participating in the EU SAGhE study. *J Clin Endocrinol Metab*. 2012;97: E213–E217.
184. Swerdlow AJ, Cooke R, Beckers D, et al. Cancer risks in patients treated with growth hormone in childhood: the SAGhE European Cohort Study. *J Clin Endocrinol Metab*. 2017;102:1661–1672.
185. Sävendahl L, et al. Long-term mortality after childhood growth hormone treatment: the SAGhE cohort study. *Lancet Diabetes Endocrinol*. 2020;8:683–692.
186. Albertsson-Wikland K, et al. Mortality is not increased in recombinant human growth hormone-treated patients when adjusting for birth characteristics. *J Clin Endocrinol Metab*. 2015;101:2149–2159.
187. Quigley CA, et al. Mortality in children receiving growth hormone treatment of growth disorders: data from the Genetics and Neuroendocrinology of Short Stature International Study. *J Clin Endocrinol Metab*. 2017;102:3195–3205.
188. Allen, et al. GH safety workshop position paper: a critical appraisal of recombinant human GH therapy in children and adults. *Eur J Epidemiol*. 2016;174:1–9.
189. Aimaretti G, Baldelli R, Corneli G, et al. GHRH and GH secretagogues: clinical perspectives and safety. *Pediatr Endocrinol Rev*. 2004;2:86–92.
190. Ahmid M, Perry CG, Ahmed SF, et al. Growth hormone deficiency during young adulthood and the benefits of growth hormone replacement. *Endocr Connect*. 2016;5:R1–R11.
191. Gleeson H, Turner G. Transition to adult services. *Arch Dis Child Educ Pract Ed*. 2012;97:86–92.
192. Koltowska-Haggstrom M, Geffner ME, Jonsson P, et al. Discontinuation of growth hormone (GH) treatment during the transition phase is an important factor determining the phenotype of young adults with nonidiopathic childhood-onset GH deficiency. *J Clin Endocrinol Metab*. 2010:2646–2654.
193. Yuen KCJ, Biller BMK, Radovick S, et al. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for management of growth hormone deficiency in adults and patients transitioning from pediatric to adult care. *Endocr Pract*. 2019;25:1191–1232.
194. Çamtosun E, Şıklar Z, Berberoğlu M. Prospective follow-up of children with idiopathic growth hormone deficiency after termination of growth hormone treatment: is there really need for treatment at transition to adulthood? *J Clin Res Pediatr Endocrinol*. 2018;10:247–255.
195. Rosenfeld RG. Transitioning patients with childhood-onset growth hormone deficiency to treatment in adulthood. *J Pediatr Endocrinol Metab*. 2002;15:1361–1365.
196. Penta L, Cofini M, Lucchetti L, et al. Growth hormone (GH) therapy during the transition period: should we think about early retesting in patients with idiopathic and isolated GH deficiency? *Int J Environ Res Public Health*. 2019;16:E307.
197. Maghnie M, Aimaretti G, Bellone S, et al. Diagnosis of GH deficiency in the transition period: accuracy of insulin tolerance test and insulin-like growth factor-I measurement. *Eur J Endocrinol*. 2005;152:589–596.
198. Sbardella E, Pozza C, Isidori AM, et al. Endocrinology and adolescence: Dealing with transition in young patients with pituitary disorders. *Eur J Endocrinol*. 2019;181:R155–R171.
199. Yuen KCJ. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Growth Hormone Stimulation Tests in Assessing Adult Growth Hormone Deficiency*. South Dartmouth: Endotext; 2019.
200. Yang H, et al. Body composition and metabolic health of young male adults with childhood-onset multiple pituitary hormone deficiency after cessation of growth hormone treatment. *J Pediatr Endocrinol Metab*. 2018;31:533–537.
201. Elbornsson M, et al. Fifteen years of GH replacement improves body composition and cardiovascular risk factors. *Eur J Endocrinol*. 2013;168:745–753.
202. Setola E, et al. Effects of growth hormone treatment on arginine to asymmetric dimethylarginine ratio and endothelial function in patients with growth hormone deficiency. *Metabolism*. 2008;57:1685–1690.
203. Giagulli VA, et al. GH supplementation effects on cardiovascular risk in GH deficient adult patients: a systematic review and meta-analysis. *Endocr Metab Immune Disord Drug Targets*. 2017;17:285–296.
204. Collett-Solberg PF, Ambler G, Backeljauw PF, et al. Diagnosis, genetics, and therapy of short stature in children: a growth hormone research Society international Perspective. *Horm Res Paediatr*. 2019;92:1–14.

19

Basic and Clinical Appetite Regulation and Energy Expenditure

Janine Makaronidis, Rachel Batterham, Edward Melanson, and Sarah A. Purcell

OUTLINE

Introduction, 306
Components of Eating Behavior, 306
  Appetite Regulation, 306
  Leptin, Growth Differentiation Factor 15, and Endocannabinoids, 309
  The Vagus Nerve: The Role of Vagal Signaling in Appetite Regulation, 309
  Hypothalamus and the Central Nervous System: Integration of Energy Availability Signals, 309
  The Role of Genetics and Epigenetics, 310
Components of Energy Expenditure, 311
  Resting Metabolic Rate, 311
  Thermic Effect of Food, 313
  Physical Activity Energy Expenditure, 313
  Brown Adipose Tissue and Uncoupled Respiration, 313
Body Weight Regulation, 314
  Body Weight Regulation and Eating Behavior: Beyond Homeostasis, 314
  Contributions of Components of Total Daily Energy Expenditure With Prospective Weight Change, 314
  Physiological Responses to Weight Loss and Weight Regain, 315
  The Effects of Exercise on Energy Balance, 316
Conclusions, 317

INTRODUCTION

Both energy intake and expenditure are regulated by a complex set of physiological signals that impact energy balance, and thus body weight. Understanding the processes underpinning energy intake and expenditure regulation is crucial to understanding human behavior and potentially holds the key to finding effective treatment options for both underweight and obesity. This chapter will discuss the sophisticated orchestration of eating behavior and energy expenditure (EE) and how these determine body weight.

COMPONENTS OF EATING BEHAVIOR

Eating is essential to survival, and it is therefore not surprising that evolutionary adaptations have rendered it enjoyable. Historically, throughout most of the evolution of human metabolism, food availability has been sparse, resulting in powerful biological mechanisms that drive eating. Ensuring adequate energy intake to sustain basic physiological processes such as resting metabolic rate (RMR) and physical activity is the most fundamental homeostatic function for survival. The physiological processes regulating energy availability are therefore incredibly complex and encompass peripheral signals that reflect nutrient intake and energy stores, the autonomic nervous system, and brain regions that control energy homeostasis. In addition to these homeostatic mechanisms, eating behavior is also governed by reward-related pathways, as well as psychological and cognitive decision-making processes.

Appetite Regulation

The Gastrointestinal Tract and Gut Hormones. The gastrointestinal (GI) tract is the first point of contact with ingested food and acts as the key interface between nutrients and brain centers that regulate energy homeostasis. Gut-derived signals convey information about the composition and energy content of ingested food to the brain. In this section, we review endocrine and neuronal signals originating from the GI tract and their role in controlling appetite and eating behavior (Fig. 19.1).

The GI tract is one of the largest endocrine organs in the body and contains populations of enteroendocrine cells (EECs) along its entire length.[1] EECs detect the presence of ingested nutrients and in response synthesize and secrete a multitude of metabolically active peptides.[1] Gut hormones, through their autocrine, paracrine, and endocrine actions, regulate digestive function, eating behavior, EE, and glucose homeostasis.[2] Table 19.1 summarizes key gut hormones and their effect on energy and glucose homeostasis.

In conditions of energy deficit, the orexigenic hormone and growth-hormone secretagogue, ghrelin, is secreted from P/D1 cells located predominantly in the gastric fundus.[3] Ghrelin stimulates appetite and drives the desire to eat. Circulating ghrelin levels peak prior to nutrient ingestion and decrease rapidly postmeal. Ghrelin exerts its orexigenic effects by acting on the hypothalamus, an effect that is exaggerated by prolonged fasting.[3]

In contrast, following ingestion of a meal, EECs along the GI tract sense nutrients and secrete a number of gut hormones, which inform central appetite-controlling circuits and peripheral tissues about nutrient availability. The anorectic hormones peptide YY 36 (PYY), glucagon-like peptide 1 (GLP-1), and oxyntomodulin (OXM) are all secreted from enteroendocrine L-cells in the GI tract and suppress appetite.[2]

GLP-1 receptors are widely expressed on multiple organs and tissues. The hypothalamic GLP-1 receptor (GLP-1R), however, has a key role in the regulation of eating behavior. In mice, stimulation of this receptor leads to reduced food consumption, whereas inhibition causes increased food intake.[4] Furthermore, GLP-1R

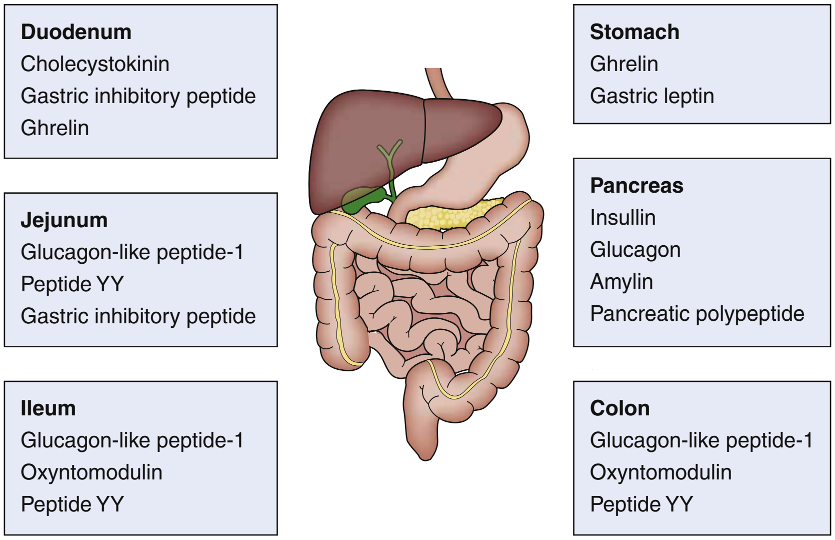

Duodenum
Cholecystokinin
Gastric inhibitory peptide
Ghrelin
Jejunum
Glucagon-like peptide-1
Peptide YY
Gastric inhibitory peptide
Ileum
Glucagon-like peptide-1
Oxyntomodulin
Peptide YY
Stomach
Ghrelin
Gastric leptin
Pancreas
Insullin
Glucagon
Amylin
Pancreatic polypeptide
Colon
Glucagon-like peptide-1
Oxyntomodulin
Peptide YY

Figure 19.1 **Several Metabolically Active Peptides are Secreted Along the Gastrointestinal (GI) Tract that Regulate Energy and Glucose Homeostasis and Body Weight.** Gut hormones are secreted by discrete populations of enteroendocrine cells located along the GI tract, from the stomach to the colon. These include the orexigenic hormone ghrelin, anorectic peptides (glucagon-like peptide-1, peptide YY, cholecystokinin, oxyntomodulin) and pancreatic hormones (insulin, pancreatic polypeptide, amylin).

TABLE 19.1 The Role of Gut Hormones in Appetite Regulation and Energy Expenditure

Gut Hormone	Source and Stimuli	Sites of Action	Mechanism of Action	Function in Appetite and Body Weight Regulation
Ghrelin[3,144]	P/D1-type cells in: • Gastric antrum • Gastric fundus • Duodenum In response to: • Fasting • Regular mealtimes • Food cues	Following acylation by ghrelin O-acyl-transferase into acyl-ghrelin: • Growth hormone secretagogue receptor type 1a (GHSR1a) • Hypothalamus • Ventral tegmental area and other CNS reward areas • Vagus nerve	• Hypothalamus: MC4R inhibition • GHSR1a activation: stimulates AgRP and NPY neurons, increasing the drive to eat • Stimulates endocannabinoid release	• Increases appetite and motivation to eat • Increases the reward value of food • Increases gastric emptying and gastric acid production • Decreases insulin secretion
Glucagon-like-peptide 1 (GLP-1)[5,144,170]	• Enteroendocrine L-cells • Brainstem neurons In response to: • Exposure to nutrients, mainly glucose and fatty acids • L-cell stimulation by bile acids	GLP-1 receptor (GLP-1R) is widely present on central and peripheral tissues, including • Hypothalamus • Liver • Pancreatic beta calls • Skeletal and cardiac muscle • Vagus nerve	• Pancreatic β cells: stimulates insulin secretion • Hypothalamus: stimulation of afferent fibers in the paraventricular nucleus, suppressing the drive to eat	• Reduces appetite and suppresses motivation to eat • Reduces the reward value of food • Delays gastric emptying
Peptide YY 3-36 (PYY)[6,7,144]	• Enteroendocrine L-cells • Brainstem neurons In response to: • Exposure to nutrients, mainly protein and fatty acids	Y2 receptors: • Throughout the CNS • Vagus nerve	• Hypothalamic Y2 receptors: inactivation of NPY/AGRP neurons, suppressing the drive to eat	• Reduces appetite and suppresses motivation to eat • Reduces the reward value of food
Oxyntomodulin (OXM)[171]	• Enteroendocrine L-cells • Cosecretion with GLP-1	• GLP-1R • Glucagon receptors • Hypothalamus	• GLP-1R coactivation	• Reduces appetite and inhibits energy intake • Delays gastric emptying • Drives glucose-dependent insulin secretion
Cholecystokinin (CCK)[11]	• Enteroendocrine L-cells and I-cells • CNS In response to: • Exposure to nutrients, mainly protein and fatty acids	• CCK-1 receptor: stomach, pancreas, gallbladder • CCK-2 receptor: CNS	• Stimulates vagal afferents terminating in the NTS • Activates paraventricular nucleus of the hypothalamus	• Reduces appetite and inhibits energy intake • Delays gastric emptying • Inhibits gastric acid secretion • Stimulates insulin secretion

Continued

TABLE 19.1 The Role of Gut Hormones in Appetite Regulation and Energy Expenditure—cont'd.

Gut Hormone	Source and Stimuli	Sites of Action	Mechanism of Action	Function in Appetite and Body Weight Regulation
Gastric inhibitory polypeptide (GIP)[172]	- Enteroendocrine K-cells In response to: - Exposure to nutrients, mainly glucose and fatty acids	GIP receptor: - Pancreatic islet cells - Hypothalamus - adipose tissue	- β cells: glucose-dependent insulin release - CNS: induces proliferation of hippocampal progenitor cells	- Stimulates insulin secretion - Reduces energy intake
Gastric leptin[24,173]	- Gastric chief cells - Gastric endocrine P cells In response to: - Nutrient ingestion - Insulin secretion - CCK secretion	Leptin receptors: - Vagus nerve	Vagal afferents: - Fasted state: inhibits vagal afferents, increases energy intake - Fed state: stimulates vagal afferents, inducing satiety	- Regulates energy intake in a nutritional status–dependent manner
Amylin[174,175]	- Pancreatic β cells - Cosecreted with insulin In response to: - Glucose and fatty acid ingestion	Amylin-specific receptors: - CNS - Gastric fundus - Bone	- CNS (area postrema): receptor activation is transmitted to the NTS, inhibiting eating	- Suppresses postprandial glucagon secretion - Suppresses appetite and inhibits energy intake - Slows gastric emptying
Uroguanylin[176]	- Enterochromaffin cells in duodenum and small intestine In response to: - Nutrient ingestion	Guanylyl cyclase 2C (GUCY2C) receptors: - Intestinal epithelium - Hypothalamus	- At intestinal level regulates fluid and electrolyte balance	- Hypothalamus: promotes satiety and reduces energy intake
Neurotensin (NT)[177]	- Enteroendocrine cells - CNS In response to: - Nutrient ingestion, mainly lipids	Neurotensin receptors NTR1, NTR2, NTR3: - CNS, particularly hypothalamus - Pancreas - GI tract	- Increases POMC expression - Activates dopaminergic system in the midbrain	- Suppresses appetite - Reduces GI motility and gastric secretions - Promotes glucose-dependent insulin secretion - Facilitates fat translocation

AgRP, agouti-related protein; *CNS*, Central nervous system; *NPY*, neuropeptide Y; *NTS*, nucleus of the solitary tract; *GI*, gastrointestinal.

knockdown in rodents results in hyperphagia and obesity,[4] highlighting the importance of this pathway in postmeal appetite suppression and weight regulation. GLP-1 also has a marked incretin effect, enhancing glucose-dependent insulin release and inhibiting glucagon secretion.[5] PYY has an important role in the postprandial suppression of eating. Peripheral administration of PYY to healthy human volunteers reduces appetite and energy intake.[6] However, when discussing the role of anorectic gut hormones in the regulation of energy balance and eating behavior, it is important to note that, in normal physiological conditions, their actions are synergistic, as highlighted by studies using administration of native peptides. For instance, GLP-1 and PYY administered in combination have a more potent effect on appetite suppression compared with either of the two hormones alone.[7,8] Furthermore, GLP-1, PYY, and OXM have an additive effect on promoting satiety and reducing energy intake, whereas gastric inhibitory polypeptide (GIP) and cholecystokinin (CCK) enhance the effects of GLP-1.[7,9-11]

Gut hormone secretion from EECs is regulated by a combination of their nutrient exposure, as well as neural and endocrine pathways.[12] Interestingly, the melanocortin-4 receptor (MC4R) is present on EECs, additionally regulating ghrelin and PYY secretion.[13] Furthermore, gut hormones influence energy homeostasis through interactions with the microbiome and bile acids. For instance, bile acids in humans stimulate EECs via TGR5 activation, which results in gut hormone secretion. Interestingly, administration of bile acids to humans leads to release of GLP-1 and improved glucose homeostasis.[14]

Pancreatic Hormones and Appetite Regulation. In addition to their role in glucose homeostasis, pancreatic hormones also contribute to appetite regulation. Insulin receptors are present in the arcuate nucleus of the hypothalamus (ARC), and their activation results in appetite suppression.[15] The pancreatic hormone amylin acts centrally as a satiety signal and reduces both food intake and body weight.[16] Both in animal and human studies, exogenous administration of amylin results in reduced energy intake and weight loss.[16] Amylin exerts its effects by acting upon receptors that are widely distributed across the central nervous system (CNS), including the hypothalamus. Amylin also directly acts on the area postrema of the hindbrain, offering an additional route for the regulation of energy homeostasis independent of hypothalamic circuits.

Taste, Visual, Gustatory, and Olfactory Cues and Appetite Regulation. The chemosensory properties of food are among the key determinants of food-related decision-making. Visual, gustatory, and olfactory cues can influence the motivation and decision to eat.[17] Taste and smell are key determinants of food choice and enjoyment, and gut hormones have a role in modulating the perceived palatability of the chemosensory properties of food. Ghrelin modulates olfactory responsiveness by increasing CNS responses to olfactory stimuli, and ghrelin levels directly correlate with olfactory activity.[18] The presence of gut hormones in saliva and their cognate receptors on taste buds implicates gut hormones in modulating gustatory function. Preclinical studies suggest that increasing PYY concentrations in saliva alters food preference, reduces energy intake, and reduces body weight.[19]

The oral cavity is the first point of contact with food, and chemoreception by taste receptors in the oral cavity is the first step in recognition of and metabolic responses to ingested nutrients.[20] *In vitro* and *in vivo* studies suggest that detection of nutrients by taste receptors modulates neuronal and endocrine signaling engendering downstream effects on energy homeostasis. Furthermore, a number of gut hormones, including PYY, ghrelin, insulin, and GLP-1, have been detected in saliva and are thought to have a role in modulating taste signaling and perception, particularly in response to sweet and fatty stimuli.[17,21] GLP-1 is expressed in sweet- and umami-sensing cells, and GLP-1 null mice have reduced taste sensitivity to sweet taste.[22,23] In addition, taste receptors are also present along the entire GI tract, including on EECs, and their activation has been linked to GLP-1 release in response to glucose ingestion.[23]

Leptin, Growth Differentiation Factor 15, and Endocannabinoids

Leptin. The adipokine leptin is secreted from adipose tissue and acts as a signal of long-term energy availability, signaling both the availability and stability of long-term energy stored in adipose tissue.[24] Leptin secretion promotes satiety by modulating orexigenic neurons in the ARC in an opposing manner to ghrelin.[24] Leptin deficiency in humans results in hyperphagia and severe obesity, which is reversed by leptin administration. A recent report of exogenous leptin treatment in a patient with leptin deficiency highlights key interactions between gut hormones. Supplementation resulted in significant rises in meal-stimulated insulin, GLP-1, and PYY levels and a reduction in ghrelin levels, highlighting the regulatory role of leptin.[25] Administration of leptin, however, in adults with obesity without leptin deficiency does not lead to significant weight loss.[26]

Growth Differentiation Factor 15. Growth differentiation factor 15 (GDF15), a metabolically active protein encoded by the macrophage inhibitory cytokine 1 gene, was originally described to be a stress- and inflammation-induced cytokine secreted in response to macrophage activation.[27] A growing body of evidence supports the role of GDF15 as an independent regulator of appetite and body weight. GDF15 knockout mice have increased adiposity and body weight associated with increased spontaneous food intake.[28] However, the phenotype can be reversed by infusion of recombinant GDF15 in concentrations equivalent to circulating levels in healthy humans, demonstrating that GDF15 has a role in energy homeostasis.[28] Circulating GDF15 acts on the ARC; however, it primarily mediates its anorectic effects via the area postrema and the nucleus of the solitary tract (NTS).[29]

Endocannabinoids. The chemosensory properties of fatty acids in the oral cavity during eating lead to secretion of lipid-derived endocannabinoid signals such as anadamide.[30] Endocannabinoids bind to cannabinoid receptors in the CNS and on peripheral tissues. Activation of the cannabinoid receptor 1 (CB1R) has orexigenic effects, whereas blockade in animal models reduces energy intake. Anadamide levels have been shown to increase with prolonged fasting, and circulating levels mirror plasma fatty acid levels.[30] Meal-stimulated changes in endocannabinoids have also been shown to depend on the palatability of ingested food and may relate to the perceived pleasantness of food, although the exact role of the endocannabinoid system in appetite regulation remains incompletely understood.

KEY POINTS

- The gastrointestinal (GI) tract plays a key role in energy regulation. Enteroendocrine cells across the entire length of the GI tract sense ingested nutrients and, through secretion of orexigenic (e.g., ghrelin) and anorectic (e.g., PYY, GLP-1, OXM) gut hormones, transmit signals of energy availability to the brain. Pancreatic hormones such as amylin (anorectic), the adipokine leptin, and meal-induced endocannabinoids also have a role in appetite regulation. However, these metabolic signals of appetite can be overridden through powerful hedonic responses to food cues.

The Vagus Nerve: The Role of Vagal Signaling in Appetite Regulation

The vagus nerve provides bidirectional communication between the brain and GI tract. Vagal afferent neurons act as early integrators of peripheral energy signals. Their firing is influenced by multiple signals and changes in the GI milieu, including gut hormones, leptin, gut-derived lipid signals, the intestinal microbiome, and GI inflammatory signals. Receptors for leptin, ghrelin, CCK, GLP-1, and PYY have been located on vagal afferents. Vagal afferents convey signals of energy availability to the NTS in the brainstem.[31] Furthermore, the vagus nerve receives direct synaptic signaling from EECs along the GI tract. EECs project synapse-like structures, called neuropods, that synapse directly with vagal neurons, allowing them to directly transmit signals relating to energy availability to neurons in one synapse to the brainstem.[32] Vagal efferent fibers emerge from the dorsal motor nucleus of the vagus nerve in the medulla that contain parasympathetic efferents and innervate the musculature and mucosal layer of GI organs.[33] Vagal efferent signals from the brain to the GI tract and together with metabolic and endocrine signals coregulate nutrient absorption and mobilization.[31] Vagal efferent signals also synapse onto enteric neurons, providing a link between the central and enteric nervous systems.[33]

Vagal afferents display inherent plasticity, responding to both short- and long-term changes in energy availability and altering their responsivity to gut-derived signals accordingly.[34] For instance, fasting increases CB1R expression and promotes orexigenic signaling, an effect that is quickly reversed following energy intake.[30] Up- and downregulation of gut hormone receptors is also a dynamic process. Both ghrelin and CCK have been implicated in regulating the phenotype of the vagus nerve, altering its sensitivity to gut-derived signals, possibly in an opposing manner.[31]

Hypothalamus and the Central Nervous System: Integration of Energy Availability Signals

Peripheral signals relating to energy and nutrient availability are conveyed to the CNS, where numerous neuronal circuits integrate these to control appetite and eating behavior.[35] NTS neurons project into the ARC. Within the CNS, the hypothalamus is the main center integrating peripheral homeostatic signals of energy availability into central networks.

Hypothalamic nuclei containing neurons expressing neuropeptide Y (NPY) and agouti-related protein (AgRP) become active in conditions of energy deficit and drive orexigenic responses; conversely, in conditions of energy surplus, proopiomelanocortin (POMC) neurons generate anorectic responses, in a mutually exclusive manner.[36] POMC neurons coexpress cocaine- and amphetamine related transcript (CART). Ghrelin activates NPY and AgRP neurons, generating a sensation of hunger and driving eating, an effect that is exacerbated by prolonged fasting.[3] In contrast, anorectic peptides including PYY, GLP-1, and OXM activate POMC neurons. In the paraventricular nucleus of the hypothalamus, the melanocortin melanocyte-stimulating hormone

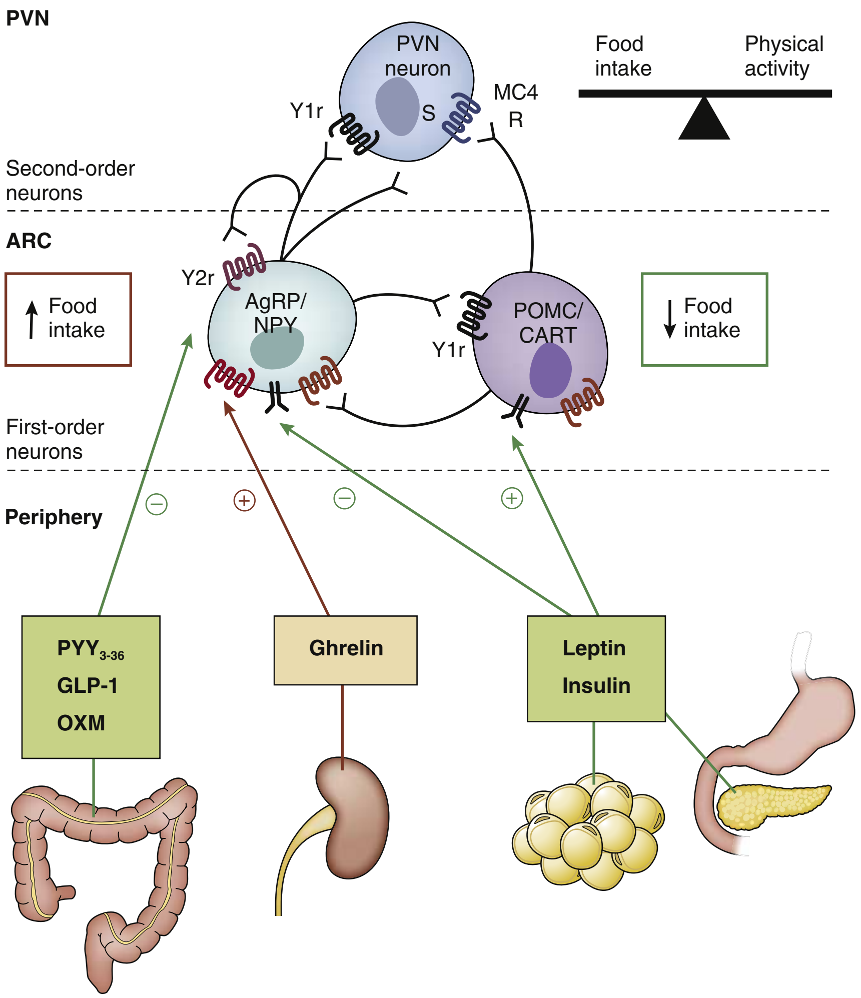


Figure 19.2 **Peripheral Signals of Energy Availability are Integrated in the Hypothalamus.** First-order neurons are located in the arcuate nucleus of the hypothalamus (ARC). Neurons expressing neuropeptide Y (NPY) and agouti-related protein (AgRP) become active in conditions of energy deficit and drive orexigenic responses. Ghrelin activates NPY and AgRP neurons.[36] In conditions of energy surplus proopiomelanocortin (POMC) neurons generate anorectic responses. POMC neurons coexpress cocaine- and amphetamine related transcript (CART). Anorectic peptides including PYY, GLP-1, and OXM activate POMC neurons and inhibit NPY/AgRP. Rises in insulin and leptin levels also have inhibitory effects on NPY/AgRP. In the paraventricular nucleus of the hypothalamus (PVN), melanocortin melanocyte-stimulating hormone a (a-MSH), subsequently activates the melanocortin-4 receptor (MC4R), which generates a feeling of satiety, improves insulin sensitivity, and increases energy expenditure.[37] Depending on energy needs versus availability, NPY/AgRP and POMC neurons are activated in a mutually exclusive manner.[36]

subsequently activates the MC4R, which generates a feeling of satiety, improves insulin sensitivity, and increases EE.[37] In conditions of energy deficit, the receptor is inhibited by AgRP (Fig. 19.2).[37] Loss-of-function MC4R mutations result in hyperphagia, severe obesity, rapid growth, and impaired glucose metabolism. MC4Rs have been localized on EECs and are thought to contribute to regulation of gut hormone secretion.[13] Ascending hypothalamic projections convey information on energy availability to higher CNS centers involved in the perception and control of eating, which are summarized in Table 19.2.

Research utilizing functional magnetic resonance imaging (fMRI) in combination with measurement or administration of gut hormones or during performance of food-related tasks has led to an increased understanding of how these incredibly complex pathways ultimately determine eating behavior.[38] Through measuring the degree of blood oxygen level–dependent contrast, a signal of activity-related change, these studies have demonstrated that gut hormones directly modulate neural activity in these brain centers.[39] Following a period of fasting, activity in the hypothalamus has been shown to correlate with subsequent energy intake.[40] Circulating ghrelin levels in the fasted state correlate with neural activity in the amygdala, orbitofrontal cortex (OFC), anterior insula, and striatum,[41] and administration of ghrelin increases food-cue reactivity. When ghrelin is administered in the fed state, it increases neural activity and food cue reactivity to a degree comparable to the fasting.[41,42] Following a meal, PYY and GLP-1 synergistically modulate activity in the brain's appetitive network and reduce the reward value of food.[39] A negative correlation has been demonstrated between PYY levels and activity in the brain's reward centers.[39] A study comparing the effects of a meal compared with the effects of PYY and GLP-1 infusions, administered separately and together, demonstrated that PYY and GLP-1 caused changes in brain activity and reduced energy intake comparably to a meal.[7]

The Role of Genetics and Epigenetics

Both homeostatic and hedonic drivers for eating behavior are subject to individual predisposition from genetic and epigenetic factors.

Population-wide studies have demonstrated that up to 85% of variability in body mass index is heritable. For example, the obesity risk variant rs9939609 in the fat-mass and obesity-associated (*FTO*) gene is linked to increased appetite levels and preferences for energy-dense foods. Normal-weight adults homozygous for the rs9939609 variant have higher postprandial circulating ghrelin levels, increased hunger, and altered responses to food cues in CNS reward regions compared with subjects with the low-risk variant.[43] In addition to genetic factors, epigenetic factors from the environment passed from one generation to the next through perinatal programming may contribute more acutely toward an individual's body weight.[44] Although the degree to which body weight is influenced by genetics remains unknown, it is evident that large, multifactorial variability exists between individuals in their predisposition to gain weight, appetite levels and food preference, and how their biology responds to attempts conditions or energy excess or restriction.

KEY POINTS

- The vagus nerve provides bidirectional communication between the brain/central nervous system and gastrointestinal tract. Vagal afferent signals are responsive to short- and long-term alterations in energy availability, and their responsiveness is highly adaptable based on nutritional state. Hypothalamic nuclei containing neurons expressing neurons expressing neuropeptide Y and agouti-related protein are active in conditions of energy deficit and drive orexigenic responses; conversely, proopiomelanocortin neurons (which coexpress cocaine- and amphetamine related transcript) generate anorectic responses when energy requirements are met. Importantly, eating behavior and body weight are influenced by individual genetic predisposition, with up to 85% of body mass index variability ascribed to genetic differences.

TABLE 19.2 Central Nervous System Areas Involved in the Control of Appetite and Eating

Brain Area	Role in Eating Behavior
Hypothalamus	Homeostatic control
Hippocampus	Learning and memory Links energy balance to incentive behavior
Amygdala	Emotional learning, assigning of value Encodes palatability of foods Link between homeostatic and hedonic interactions
Insula	"Ingestive" and "gustatory" cortex Encodes sensory information of taste, integrated with sensory cues to form flavor Responds to food cues and gut hormones
Nucleus of the solitary tract	Afferent target for vagus nerve
Ventral tegmental area	Encodes nutritional and reward values of food Generates value and motivational signals
Cerebellum	Integrates and coordinates somatic-visceral responses to food
Striatum: **Caudate nucleus** **Putamen** **Nucleus accumbens**	Encodes motivational and incentive properties of food Anticipatory reward Connected to all other brain regions of eating behavior Links motivation to action
Orbitofrontal cortex	Reward encoding Integration of sensory, cognitive, and reward aspects
Cingulate cortex	Decision making
Prefrontal cortex	Translation of internal and external cues to behavior, food-related decision-making

Adapted from Makaronidis JM, Batterham RL. Obesity, body weight regulation and the brain: insights from fMRI. *Br J Radiol.* 2018:20170910.

COMPONENTS OF ENERGY EXPENDITURE

Total daily EE (TDEE) is comprised of three main components: RMR, thermic effect of food (TEF), and physical activity EE (Fig. 19.3).

Resting Metabolic Rate

Basal metabolic rate (BMR) is the energy required to sustain basic physiological functions (e.g., cellular function, respiration, circulation, neurological function) by the body at complete rest and accounts for approximately 60% to 70% of TDEE in sedentary individuals. BMR is measured under strict, standard conditions that include being awake in the supine position after 10 to 12 hours of fasting and 8 hours of physical rest, with the measurement obtained under thermoneutral conditions.[45] Because these conditions are not always met, this is often referred to as RMR, which is approximately 5% higher than BMR. Although BMR and RMR are often used interchangeably, we will use the term RMR for consistency throughout the chapter, as that is what is more commonly measured in clinical settings.

RMR is measured in clinical and research settings using indirect calorimetry. These systems are open-circuit, meaning that subjects inspire ambient air, and the difference between inspired and expired concentrations of oxygen (O_2) and carbon dioxide (CO_2) is used to calculate EE[46] and substrate oxidation (e.g., fat and carbohydrate) (Fig. 19.4A). In clinical settings, subjects should be overnight fasted (at least 8 hours) and refrain from moderate-to-vigorous exercise for 24 to 48 hours. When these conditions are met, a coefficient of variation between repeated measurements is approximately 5%.[47]

Many research studies and some clinical settings measure EE with metabolic carts (Fig. 19.4B). Advances in technology have led to the development of "handheld" or "portable" indirect calorimeters (Fig. 19.4C). Many of these devices measure only O_2 consumption, and thus cannot measure substrate oxidation. Although these devices are accessible in clinical settings, they typically are less accurate compared with metabolic carts.[48,49]

Numerous predictive equations have been developed to estimate RMR. The most prevalent equations are the Harris–Benedict[50] and Mifflin–St. Jeor[51] formulas, which use sex, age, height, and weight to estimate RMR. Other equations include additional variables such as body composition and ethnicity; however, these equations often have been developed for use in specific populations. While most equations do well on a group level, individual-level accuracy is poor.[52] For example, the Harris–Benedict equation may under- or overpredict RMR by greater than 10% in approximately 30% to 40% of adults; similarly, the Mifflin–St. Jeor equation may under- or overpredict RMR in approximately 20% to 50% of adults.[53]

Determinants of Resting Metabolic Rate. RMR is primarily determined by body size and composition. Specifically, fat-free mass (FFM)—the sum of total body water, total body protein, carbohydrates, nonfat lipids, soft tissue minerals, and bone minerals—accounts for 50% to 70% of the variability in RMR between individuals. Although FFM is often considered to be synonymous with muscle mass, FFM is comprised of a number of energetically heterogeneous tissues. For example, the heart, liver, kidney, and brain constitute 5% to 6% of body weight but contribute to approximately 60% to RMR.[54,55] Skeletal muscle, adipose tissue, and bone constitute a much larger percentage of total body weight but contribute only approximately 20% to 30% to RMR (Table 19.3).[54,55] Although the metabolic rate of skeletal muscle per gram of tissue is lower than that of other organs, it is a strong predictor of RMR because of its larger mass.[54] Similarly, adipose tissue uses a small amount of energy per kilogram (4 kcal/kg/day) but may also substantially contribute to RMR in individuals with excess fat mass (FM).[56]

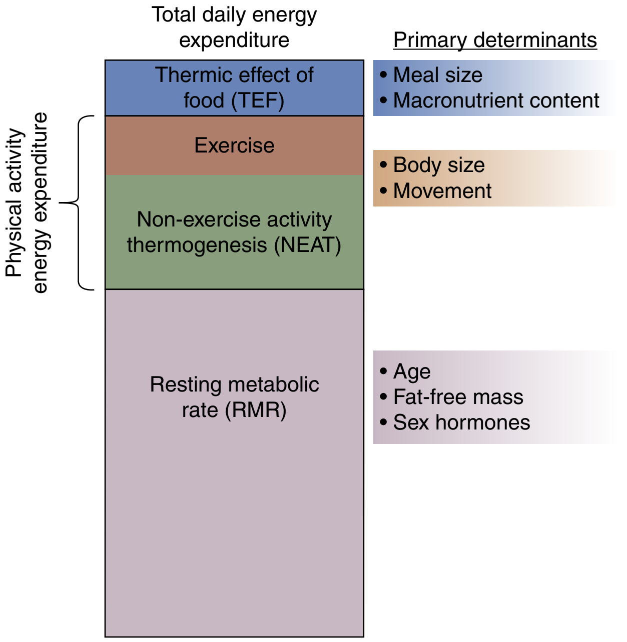

Total daily energy expenditure
Primary determinants
Thermic effect of food (TEF)
• Meal size
• Macronutrient content
Physical activity energy expenditure
Exercise
Non-exercise activity thermogenesis (NEAT)
• Body size
• Movement
Resting metabolic rate (RMR)
• Age
• Fat-free mass
• Sex hormones

Figure 19.3 Major Components and Determinants of Total Daily Energy Expenditure.

RMR decreases with advancing age, primarily due the age-associated decreases in FFM. However, organ- and tissue-specific metabolic rates are also lower in elderly individuals compared with young adults.[57] These differences are likely due to reductions in the cellular fraction of organs and tissues with age,[58] which are driven by reduced sodium-potassium adenosine triphosphatase activity, decreased norepinephrine from lower physical activity and dietary intake, and reduced protein synthesis.[59] Across the lifespan, absolute RMR is generally higher in males than females, but these differences are not apparent when accounting for inherent FFM differences between sexes.[60]

RMR fluctuates during the menstrual cycle, being approximately 6% to 7% higher in the luteal compared with the follicular phase,[61] which may relate to higher estrogen and progesterone in the luteal phase.[62] The impact of female sex steroids on RMR has been studied using pharmacological approaches to suppress gonadal function. In premenopausal women, suppressing gonadal function decreases RMR by 71 kcal/day compared with RMR measured in the follicular phase.[63] In a longer study that involved suppressing gonadal function for 5 months, RMR decreased 54 kcal/day, but in women in whom

TABLE 19.3 Contribution of Organs and Tissues to Resting Metabolic Rate (RMR)

Tissue	Tissue Weight (% Body Weight)	Tissue Metabolic Rate (%RMR)
Organs	<6	58
Skeletal muscle	40	22
Adipose tissue	21	<5
Miscellaneous (skin, bone, gut, etc.)	33	16

Values from a male without obesity. Organs and tissues make up <6% of body weight but contribute to >50% of RMR. Values from Dulloo AG, Jacquet J, Solinas G, et al. Body composition phenotypes in pathways to obesity and the metabolic syndrome. *Int J Obes*. 2010;34:S4–S17.

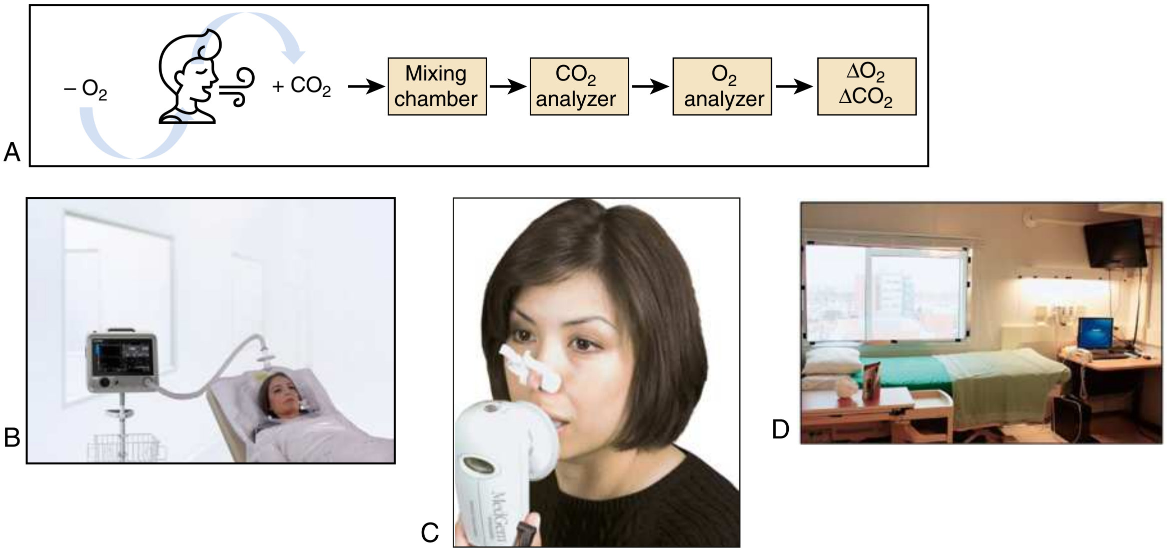


Figure 19.4 Basic Concepts and Examples of Indirect Calorimetry Systems. **A,** Simple schematic of metabolic system components. Expired oxygen (O_2) and carbon dioxide (CO_2) are transferred to a mixing chamber, where breath samples are homogenized. CO_2 and O_2 analyzers assess the concentrations of expired gases compared with gas concentrations in the room, which can then estimate O_2 consumption and CO_2 expiration. These values are then used in equations to estimate energy expenditure. **B,** Example of a metabolic cart with ventilated hood (Q-NRG, COSMED HQ, Italy – Test of Resting Energy Expenditure with Canopy hood © 2021, COSMED, Rome, Italy [www.cosmed.com]). **C,** Example of a handheld indirect calorimetry device (MedGem, Microlife, Clearwater, FL, USA). **D,** Example of whole-room indirect calorimetry, located at the University of Colorado, Anschutz Medical Campus.

estradiol was added back, this decrease in RMR was prevented.[64] The effect of oral contraceptives on RMR is unclear, with some studies showing higher RMR in pill-users versus non-pill-users, while others show no difference or even lower RMR among women on oral contraceptives.[62] In males, long-term changes in sex steroids may also contribute to decreasing RMR with age, but the evidence is conflicting. For example, 3 weeks of endogenous sex steroid suppression in older males (age range: 50–80 years) with add back of transdermal estradiol, testosterone, or both also did not affect RMR.[65] However, in another study testosterone and growth hormone replacement in hypopituitary males (age range: 33–77 years) increased RMR over 6 weeks.[66]

Thermic Effect of Food

TEF is the energy expended to digest, absorb, and store food. TEF accounts for approximately 10% of TDEE in adults in energy balance consuming a mixed diet.[67] TEF is proportional to the energy content of a meal and influenced by macronutrient content, with protein (20%–30% of the energy content of consumed food) and alcohol (10%–30%) eliciting a greater TEF response than fat (0%–3%) or carbohydrate (5%–10%).[67]

While it was initially thought that TEF does not decrease with age,[68] more recent investigations report that it is lower in older adults compared with younger individuals.[69,70] Although speculative, this age-associated decline in TEF may be attributed to differences in β-adrenergic sensitivity, which would reduce postprandial sympathoadrenal stimulation.[69] There is some controversy regarding TEF changes with weight loss or gain and in individuals with or without obesity, because some—but not all—investigations have reported differences in TEF.[71] Similarly, some studies have suggested that a low TEF predicts future weight gain,[72] which has not been corroborated by others,[73] as discussed later.

Physical Activity Energy Expenditure

PAEE is the most variable component of TDEE and is determined by body size and movement. PAEE is sometimes expressed as physical activity level (PAL), which is the ratio of TDEE to RMR. In free-living conditions, PAL is highly variable between individuals. For example, PAL of nonambulating chair-bound individuals is approximately 1.2, while PAL values of greater than 4 have been recorded in elite endurance athletes during competition, but are not sustainable (i.e., Tour de France cyclists, sled hauling across the Arctic, or Nordic skiers).[74] PAL values around 2.8 to 3.0 are sustainable for short periods of time among elite athletes during rigorous training, but likely do not represent year-round averages.[74] PAL within the general population lies between 1.4 and 1.69 in adults who are sedentary or perform light activity, 1.7 and 1.99 in adults with an active lifestyle, and 2.0 and 2.4 in adults with a vigorously active lifestyle.[45,74]

PAEE can be further categorized as exercise thermogenesis and nonexercise activity thermogenesis (NEAT), which encompasses EE associated with occupation, sitting, standing, and ambulatory activities. However, PAEE is challenging to measure. In research settings, PAEE can be measured using whole-room indirect calorimetry (Fig. 19.4D). These units resemble small hotel rooms and can assess distinct parts of EE such as exercise and sleep. However, because subjects are confined to a small space, often with restricted specific time schedules, this does not accurately measure free-living PAEE and TDEE. The doubly-labeled water (DLW) method is the most accurate way to measure free-living TDEE.[75,76] With this method, subjects ingest a dose of stable (nonradioactive) hydrogen (^2H, deuterium) and oxygen (^{18}O) isotopes. The labeled ^2H exits the body as water (primarily urine, with smaller amounts lost in breath and sweat), but the labeled ^{18}O exits the body through both water and expired CO_2. Thus, the difference in elimination rates of these two isotopes reflects CO_2 production over the period of measurement (typically 7–14 days), which is then used to estimate TDEE.[75,77] PAEE can then be calculated as TDEE – RMR, and PAL can be calculated as TDEE/RMR.

A limitation of the DLW approach is that it does not provide details about the nature or intensity of physical activity patterns. Thus, wearable activity monitors are better suited for this purpose. While most devices have reasonable accuracy in determining physical activity and sedentary behavior patterns, they *estimate* PAEE/TDEE using predictive equations, and there is substantial variability in the accuracy of both commercial and research devices.[78] Activity monitors that use additional biosignals (e.g., heart rate) to estimate PAEE and TDEE may improve accuracy.[78]

Brown Adipose Tissue and Uncoupled Respiration

Brown adipose tissue (BAT) is highly vascularized and richly innervated by the sympathetic nervous system (SNS) and generates heat through uncoupling of oxidation and phosphorylation by uncoupling protein 1 (UCP1) in a process known as "nonshivering thermogenesis" (NST).[79] Transient receptor potential (TRP) cation sensors on the skin sense the decrease in temperature and transmit signals to the hypothalamic preoptic area, evoking an increase in SNS activity, which stimulates both shivering and NST in several tissues, including BAT.[80] Until recently, BAT in humans was thought to be present in significant volume and activity only in infants.[81] However, evidence began to accumulate in the early 2000s that many humans have substantial BAT depots. In adult humans, BAT is most often observed in the cervical, clavicular, and paraspinal regions. Retrospective analysis of large cohorts of clinical positron emission tomography/computed tomography scans suggests that BAT is observable in approximately 5% to 10% of scans.[82] However, studies of cold-exposed individuals have shown metabolically active BAT is detectable in most adults.[83-86] The total volume of activated BAT is approximately 50 to 200 g in most humans studied, but has been reported to be as high as 1 kg.[87] Cold-induced BAT activity decreases with age[88-90] and is inversely associated with body fatness,[89,91,92] but whether a decline in BAT activity contributes to the development of human obesity is unclear.[93] However, during experimental cold exposure, skeletal muscle is the primary source of both shivering thermogenesis and NST.[84]

Although BAT contributes to EE and thermogenesis in rodents, the contribution of BAT to EE in humans is controversial.[79] Indirect evidence suggests a role for BAT in supporting RMR. Stimulated human BAT consumes approximately 150% more O_2 than white adipose tissue *in vivo*,[94,95] and the presence of active BAT is associated with higher RMR.[91,96] BAT may also contribute to TEF, because the postprandial increase in EE is higher in individuals with greater amounts of BAT,[97] and a variant in the *UCP1* gene is associated with lower postprandial EE.[98] Interestingly, it has recently been shown that meal consumption increases circulating levels of the BAT secreted hormone secretin. Chronic infusion of modified human secretin transiently elevates EE in diet-induced obese mice, and acute secretin infusion in humans increases BAT glucose uptake.[99]

Given the potential effect of BAT on EE in humans, there has been increased interest in understanding whether BAT activation can be used as a means to induce weight loss. Only a few studies have been performed in humans to determine the potential to increase BAT activity and EE and induce weight loss through cold exposure. Both acute and chronic interventions have reported increased BAT in response to intermittent cold exposure,[86,100-103] and others have reported concomitant small increases in RMR.[101,102,104] Theoretically, increased RMR may induce weight loss if sustained over time. However, most investigations of BAT-inducing interventions have reported no changes in body weight,[86,103] although one noted a small decrease in FM after 6 weeks of exposure to 62°F.[102] It should be noted that almost all of these studies were performed on lean adults; whether there would be an impact on body weight in people with overweight or obesity is not known. Additionally, few of these studies have measured the impact of cold exposure on energy intake. If energy intake increases in response to cold exposure, this could attenuate or even reverse the intended effects on body weight.

Because BAT is activated by SNS activity, it is possible that pharmacological or nutritional agents that stimulate SNS activity can increase BAT activity. One potential agent is mirabegron, a β3 agonist used for the treatment of hyperactive bladder. Acute treatment with 200 mg oral mirabegron increased BAT activity and RMR by approximately 200 kcal/day; however, this was accompanied by significant increases in resting heart rate and systolic blood pressure.[105] In a more recent study by the same group, 14 overweight women were treated with 100 mg mirabegron for 4 weeks. Chronic mirabegron therapy increased BAT metabolic activity and RMR, but there were no changes in body weight or composition.[106] However, recent evidence suggests that human BAT contains few β3 receptors, suggesting that the effects of mirabegron are due to off-target biding to β1 and β2 receptors.[107]

Due to the untoward effects of β-adrenergic receptor agonists, there is an interest in knowing whether food-derived stimulants activate BAT. The most heavily studied food-derived substances are capsaicin, a major component of red chili peppers that is responsible for its pungent taste, and capsinoids (capsaicin-like compound found in nonpungent red peppers). In rodents, oral administration of capsaicin activates the TRP vanilloid 1 channels located within sensory nerves of the GI tract, increasing SNS activity to BAT and whole-body O_2 consumption.[108,109] Chronic consumption of capsaicin or capsinoids upregulates BAT UCP1 protein and decreases body fat.[110] However, translating these studies to humans is difficult, because capsaicin in equivalent amounts provided to rodents would be greater than humans could tolerate. However, oral ingestion of capsinoids has been shown to increase whole-body EE in humans with high levels of BAT.[110] Cold-activated BAT is also increased by chronic consumption of capsinoids.[111] Another compound that has been studied in humans is catechin, a class of polyphenols found in green tea. Studies in rodents have demonstrated that supplementation with catechins increases UCP1 in both brown and white adipocytes.[112] Similar to studies with capsinoids, oral ingestion of tea containing catechins increases EE, but only in those with high levels of cold-stimulated BAT.[113] Although these studies are intriguing, there is limited evidence that dietary compounds activate BAT in humans and even less evidence that there is a stimulatory effect on EE or an effect on body weight.

KEY POINTS

- Total daily energy expenditure is comprised of three main components: 1) resting metabolic rate, which is the largest part of total daily energy expenditure (TDEE) and is primarily determined by body size and composition, 2) thermic effect of food, which comprises approximately 10% of TDEE and is mainly determined by energy and macronutrient intake, and 3) physical activity energy expenditure, which is the most variable component of TDEE, is determined by body size and movement, and can be subdivided into exercise and nonexercise activity thermogenesis. Recent evidence suggest that brown adipose tissue is present in a large proportion of adults, but its contribution to the development of obesity is unclear.

BODY WEIGHT REGULATION

Body Weight Regulation and Eating Behavior: Beyond Homeostasis

Eating is a behavior that is critical for survival. However, environmental changes in the past few decades have resulted in food surplus. Coupled with our metabolic systems, which developed on the back of hundreds of years of evolution in energy-restricted environments, this has led to habitual eating beyond energy needs and thereby have provided a gene-environment mismatch that led to the current obesity epidemic. In conditions of excess food availability, food choices are largely determined by cognitive, social, and emotional factors.[35] Food cues with a high reward value can themselves generate powerful physiological responses to override homeostatic signals of energy availability, which fail to suppress energy intake when the reward value of a certain stimulus is too high. CNS reward centers are responsive to the effects of PYY, GLP-1, and ghrelin. Interestingly, ghrelin levels have been shown to rise in anticipation of food consumption, and levels directly correlate with reward-center activity in response to viewing food images.[114,115] In contrast, fMRI studies have demonstrated that PYY and GLP-1 inversely correlate with reward responses to food cues, and their actions suppress reward responses.[40,116]

Contributions of Components of Total Daily Energy Expenditure With Prospective Weight Change

Resting Metabolic Rate. Whether a lower-than-expected RMR contributes to weight gain is controversial. Several studies have reported that low RMR may contribute a small amount to longitudinal weight gain. The Pima Native Americans of Southwestern Arizona are perhaps the most comprehensively described population for understanding EE and body weight regulation, because they one of the most obese populations in the world. In an often-cited study, the risk of weight gain over 2 years was increased in Pima Native Americans with a lower-than-predicted RMR.[117] However, RMR explained less than 15% of weight gain over this time span. This was corroborated in a separate but similar study in which deviation from expected TDEE was weakly inversely correlated to weight change at 6-year follow-up.[118] To put this in perspective, TDEE that was 100 kcal/day lower than predicted was associated with a 0.2-kg increase in body weight/year, of which approximately 0.1 kg was FM. Therefore, lower-than-anticipated RMR and TDEE in controlled settings likely contribute a small amount to free-living weight gain.

Thermic Effect of Food. As mentioned previously in this chapter, some, but not all, studies have suggested that TEF may contribute to weight gain. In a longitudinal analysis of 290 healthy subjects, TEF (measured in a whole-room indirect calorimeter) was inversely related to percent weight change during follow-up only in subjects with a body mass index greater than 29 kg/m^2.[72] These data contrast with an analysis by Weinsier et al.,[73] in which TEF was compared between previously obese and never obese postmenopausal women. TEF was similar between these groups; women with previous obesity gained a mean of 10.9 kg over a 4-year follow-up period, while women who never had obesity maintained body weight. TEF did not correlate with 4-year weight changes.

Physical Activity Energy Expenditure. On average, PAEE decreases with aging.[74] Observational data suggest that healthy adults who engage in the lowest quintile of physical activity gained the greatest FM (1.7 ± 0.3 kg) over 1 year compared with individuals with the highest levels of physical activity.[119] In fact, those in the lowest quintile of physical activity had a 3.8-fold higher risk of gaining greater than 3% FM.[119] Just as low PAEE predicts weight gain, high PAEE predicts weight maintenance. For example, women who successfully maintain body weight over 13 years engage in 60 minutes of self-reported moderate-intensity physical activity each day.[120] Similarly, improvements in maximal aerobic capacity (as an indirect indicator of exercise engagement) has also been shown to associate with reduced odds of weight gain over 7.5 years.[121] Each additional minute performed during a graded exercise test was associated with reduced odds of gaining greater than or equal to 5 kg body weight by 14% in men and 9% in women.[121] In a community-based intervention, individuals who successfully lost weight were more likely to engage in moderate- and high-intensity exercise compared with those who gained weight.[122] For perspective, an increase of one high-intensity exercise session per week was associated with a decrease in weight of 0.54 kg in men and 0.15 kg in women.[122]

In sum, RMR that is lower than predicted may modestly contribute to weight gain. It is unlikely that TEF contributes to weight gain, although current understanding may be limited by the small number of studies that longitudinally measure both TEF and weight change. It is likely that PAEE predicts weight change to a greater extent than RMR and TEF, although this is limited by self-reported PAEE, which is less accurate than objective measures of activity.

KEY POINTS

- Several environmental and physiological factors promote weight gain. The obesogenic environments that constitute many developed countries support food choices that are largely determined by cognitive, social, and emotional factors. Furthermore, food cues alter ghrelin, peptide YY 36, and glucagon-like peptide 1 in a manner promoting increased energy intake in environments of excess food availability. In terms of total daily energy expenditure, low physical activity energy expenditure likely contributes to the development of obesity.

Physiological Responses to Weight Loss and Weight Regain

Gut Hormone Secretion in Response to Weight Loss. Generating a negative energy balance has been the mainstay of lifestyle interventions to treat overweight and obesity. However, although restricting energy intake is successful in engendering weight loss in the short-term, this is followed by a number of very powerful physiological adaptations, aimed at defending the higher body weight, that commonly result in weight regain. Weight regain following dietary and lifestyle interventions is widely reported, with weight regain in the range of a third to half of the weight loss within the first year and weight returning to baseline within 3 to 5 years from the intervention.[123,124]

A key study by Sumithran et al. investigated the effects of a 10-week very low–calorie diet (VLCD) over a 52-week follow-up period. Significant reductions in PYY, CCK, insulin, leptin, and amylin levels, combined with increased ghrelin levels, were seen at 10 weeks and persisted at 52 weeks post-VLCD.[125] These changes will promote increases in hunger, and likely energy intake. A recent study with participants who sustained 15% weight loss 1 year following a VLCD demonstrated increased ghrelin levels and hunger ratings.[126] Reduced levels of postprandial anorectic hormones (circulating GLP-1, PYY, CCK, amylin) and leptin levels, coupled with reduced satiety have also been reported in the literature.[127-129]

Hypothalamic Responses to Weight Loss. In addition to changes in the periphery, weight loss also results in adaptive changes within the appetite- and weight-controlling circuits of the hypothalamus. Within the ARC, expression of NPY and AgRP increases, whereas expression of expression of POMC decreases, leading to increased drive to eat.[130] In addition, fMRI studies have shown altered hypothalamic activity in response to dietary weight loss.[131,132]

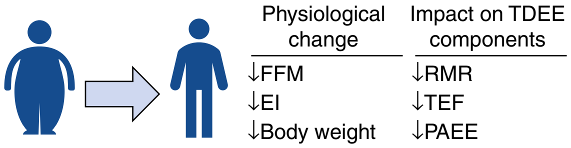


Figure 19.5 **Metabolic Alterations in Response to Weight Loss.** In a typical weight loss scenario, an individual with obesity will lose both fat mass and fat-free mass (FFM), which will lead to decreased absolute resting metabolic rate (RMR). Similarly, reduced energy intake results in lower thermic effect of food (TEF), and reduced body weight is related to decreased energy expended during activity (physical activity energy expenditure [PAEE]).

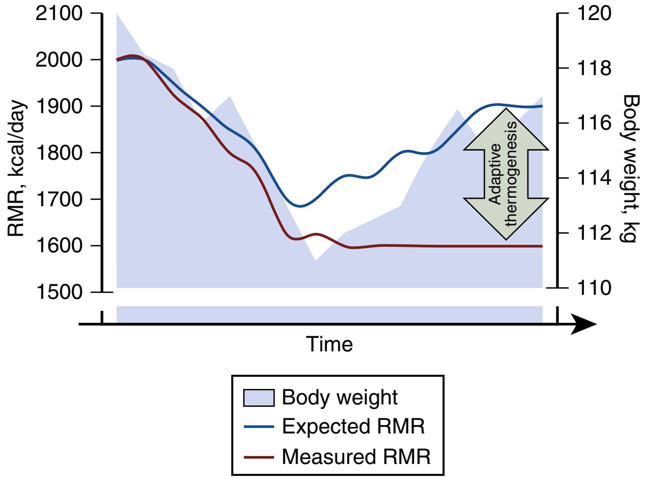


Figure 19.6 **Hypothetical Illustration of Resting Metabolic Rate (RMR) Adaptive Thermogenesis.** RMR decreases with weight loss; in some individuals, RMR may remain lower than expected after weight regain, called adaptive thermogenesis.

Changes in Components of Total Daily Energy Expenditure With Weight Loss. Weight loss through energy depletion results in a reduced sympathetic tone, as well as a reduction in thyroid hormone, which prime peripheral tissues to increase their capability for energy storage.[133] The reduction in overall body mass also results in a reduction of TDEE. Lower TDEE can be attributed to decreases in RMR due to loss of FFM and FM, TEF due to lower energy intake, and PAEE due to reduced energetic costs of activity, as summarized in Fig. 19.5.

Resting Metabolic Rate. Absolute RMR will decline as a result of lower body mass after weight loss. Some studies have suggested that RMR is lower than anticipated even after accounting for FM and FFM loss, called "adaptive thermogenesis" (Fig. 19.6).[134,135] In a highly publicized study, participants in a televised weight loss program ("The Biggest Loser") underwent RMR and body composition testing 6 years after the competition.[135] During the competition, mean weight loss was 58.3 ± 24.9 kg, and RMR decreased by approximately 600 kcal/day. At follow-up, the majority of weight lost had been regained, but RMR remained approximately 700 kcal/day below baseline and 500 kcal/day below predicted.[135] New evidence also suggests that early adaptive thermogenesis predicts weight loss during calorie restriction. Specifically, a 100 kcal/day decrease in 24-hour TDEE after 1 week of calorie restriction predicted 2.0 kg less weight loss over a 6-week inpatient weight-reduction diet among individuals with obesity.[136] Despite these rather striking results, it is important to note that adaptive thermogenesis may depend upon the amount and rate of weight loss and the duration of weight loss maintenance. For example, average RMR does not appear to differ among weight loss maintainers or normal-weight individuals who have never had obesity.[137] However, weight loss maintainers with the greatest overall weight loss were more likely to have lower-than-predicted RMR. Therefore, while some individuals may indeed have lower-than-predicted RMR, adaptive thermogenesis does not inevitably lead to weight regain.[137]

Thermic Effect of Food. Weight reduction often necessitates reduced energy intake; therefore, TEF will also decrease in accordance with this behavioral alteration. Independent of reduced energy intake, individuals with obesity may have slightly lower TEF before and after weight loss,[138] which may be due to greater suppression of hepatic gluconeogenesis, as well as lower glucose storage rate with weight loss and chronic energy intake reduction. However, TEF differences are still very small between obese and nonobese groups (8.2 ± 0.5 vs. 9.8 ± 0.5% of TDEE) and do not significantly change with weight loss (8.2 ± 0.5 vs. 8.7 ± 0.6% before and after weight loss, respectively).[138]

Physical Activity Energy Expenditure. Weight loss is associated with reduced energetic cost to move a smaller body and may therefore also relate to lower PAEE in the absence of behavioral modifications. High levels of PAEE appear to be important for maintaining a lower body weight. In a 2-year longitudinal analysis of healthy adults with no intention of weight loss, moderate-to-vigorous activity significantly increased in individuals who lost weight and decreased in individuals who gained weight.[139] Similarly, successful weight loss maintainers have higher PAEE (as measured by DLW and RMR) compared with normal-weight and obese control groups.[140] These findings have been previously corroborated by a systematic review that concluded an increased in PAEE of 1500 to 2000 kcal/week is necessary to support weight loss maintenance.[141] Therefore, high levels of PAEE are needed to maintain reduced body weight.

Insights from Bariatric Surgery. Roux-en-Y gastric bypass (RYGB) and sleeve gastrectomy (SG), the two most commonly performed bariatric operations worldwide, result in sustained weight loss and a marked change in eating behavior.[142] Following bariatric surgery, patients report reduced hunger and earlier satiety, a shift in food preference away from energy-dense foods, and changes in their sense of taste and smell.[132,143] Gut hormone secretion profiles markedly change to a favorable profile, characterized by significant rises in meal-stimulated levels of the anorectic peptides PYY, GLP-1, and OXM. Reductions in circulating ghrelin levels are also seen post–bariatric surgery.[143,144] These gut hormone secretion profiles are likely a consequence of increased EEC stimulation, which results from the anatomical changes caused by a RYGB or SG. The marked change in eating behavior following bariatric surgery, coupled with the change in gut hormone profiles, highlights the importance of these peptides in appetite and body weight regulation.

Because bariatric surgery often results in substantial weight loss, considerable changes in RMR occur and may be accompanied by adaptive thermogenesis. For example, RMR was reduced 27.3% over 12 months following SG and was approximately 200 kcal/day lower than predicted RMR using body composition.[145] However, several other investigations have not reported changes in body composition–adjusted RMR postoperatively, suggesting that adaptive thermogenesis does not occur after bariatric surgery.[146,147] Many patients also increase PAEE as a result of bariatric surgery, although approximately 25% are less active 12 months postoperatively. While there are large variations in PAEE after bariatric surgery, individuals engaging in exercise lose more weight than those who remain sedentary[148]; PAEE therefore likely plays a larger role in determining long-term success of bariatric surgery than RMR.

The Effects of Exercise on Energy Balance

The Effects of Exercise on Gut Hormone Secretion. An acute (single) bout of aerobic exercise reduces ghrelin and increases PYY, GLP-1, and pancreatic polypeptide.[149,150] The degree of this effect is likely dependent on exercise modality, as acute resistance exercise tends to elicit less pronounced alterations in gut hormone secretion.[149] Acute exercise also affects gut hormone secretion in a different manner than acute energy intake restriction. For example, a study comparing the metabolic effects of a similar energy deficit induced by either energy intake restriction or aerobic exercise reported that appetite and acyl-ghrelin levels were increased, while PYY was reduced with energy intake restriction; conversely, exercise led to a reduction in acyl-ghrelin levels and an increase in circulating PYY levels.[151]

To date, the effects of exercise on gut hormone secretion profiles in the longer term have not been characterized in depth. Some studies have suggested that acyl-ghrelin, PYY, and pancreatic polypeptide increase in response to chronic aerobic exercise interventions.[154-156] These changes theoretically support both higher hunger and satiety levels; exercise may therefore sensitize the appetite control system by balancing an increased drive to eat with improved meal-induced satiety.[149,155] These changes are particularly pronounced in response to chronic aerobic training compared with resistance training.[156] Notably, however, most of our current understanding of the appetitive effects of exercise arises from studies conducted in young, healthy, physically active males and in response to aerobic exercise. How different intensities, types, and durations of exercise regimes affect appetite in different populations such as women, older adults, or people living with obesity and other chronic health conditions remains unclear.

The Effects of Exercise on Energy Expenditure. Exercise is associated with lower risk for many chronic diseases and is important for weight loss maintenance. However, exercise may not increase TDEE in a dose-response manner. The relationship between structured exercise and components of TDEE largely depends on the type of exercise training and the population.

Resting Metabolic Rate. Exercise can be broadly categorized as resistance (strength), aerobic, or a combination of both. For the same amount of exercise, resistance exercise uses substantially less energy compared with aerobic exercise. Resistance exercise is typically more effective at increasing skeletal muscle and maintaining skeletal muscle during weight loss efforts. However, as discussed above, skeletal muscle has a low energetic cost per kilogram compared with other components of FFM (i.e., liver, heart, brain, etc.). Therefore, engaging in resistance exercise to simply increase RMR is largely ineffective. For context, healthy young males consuming sufficient calories and protein gain approximately 1.1 kg of skeletal muscle with resistance training, according to a systematic review of exercise interventions of varying duration.[157] However, this amount of skeletal muscle mass would equate to approximately 14 kcal/day to RMR (at rest).

In addition to the added caloric expenditure associated with exercise, greater SNS activity may contribute to higher RMR independently of FFM. For instance, RMR is higher in adults who regularly engage in aerobic endurance exercise compared with their sedentary counterparts.[158] This may be due in part to greater tonic SNS β-adrenergic receptor stimulation in high energy flux (i.e., high absolute levels of both energy intake and expenditure).[159] Indeed, in states of experimentally reduced energy flux among habitually

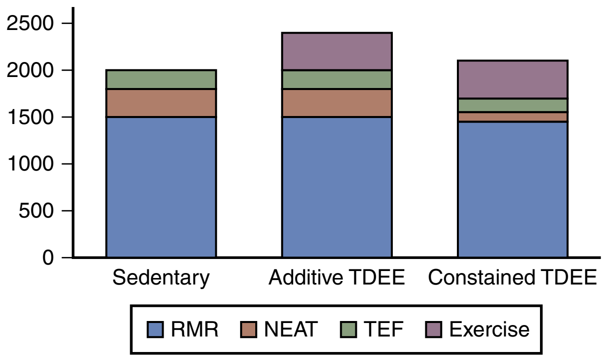


Figure 19.7 **Total Daily Energy Expenditure (TDEE) Without Exercise (Sedentary) and After Addition of Exercise Following an Additive TDEE Model and a Constrained TDEE Model.** In the additive TDEE model, TDEE increases with exercise to the same extent as the exercise. For example, an exercise session that uses 400 kcal would result in an increase of 400 kcal in TDEE in the additive model; nonexercise metabolic activity (which includes resting metabolic rate [RMR]) is not affected by exercise or physical activity variation. In the constrained TDEE model, exercise has a limited effect on TDEE and instead leads to a reduction in nonexercise activity thermogenesis (NEAT), RMR, or both. This keeps TDEE near its set point.

active older adults, there was a strong positive correlation between change in RMR and change of β-adrenergic receptor support of RMR.[159]

Physical Activity Energy Expenditure and Total Daily Energy Expenditure. A common assumption is that exercise will increase PAEE and TDEE in a linear manner, as part of an "additive" model of EE (Fig. 19.7). Although some guidelines of energy intake requirements are based on this model,[45] several exercise intervention studies have suggested that TDEE or PAEE does not increase in direct proportion to exercise EE. In fact, some studies have reported reduced NEAT in response to exercise interventions. In other words, individuals "compensate" for exercise EE by reducing other activities or increasing sedentary time outside of the exercise session. For example, an 8-week walking intervention resulted in lower NEAT among women with obesity.[160] Similar findings were reported in older men and women during 8 weeks of aerobic exercise.[161] However, it is important to note that these interventions consisted of small groups of individuals ($n \leq 13$) and may not translate to younger populations without obesity.

Compensation through reduced NEAT is not universally observed. For example, in the Dose Response to Exercise in Women study, 411 postmenopausal women with overweight or obesity were randomized to 6 months of nonexercise control or one of three exercise groups: 4, 8, or 12 kcal EE/kg body weight/week.[162,163] Individuals in the 12 kcal/kg/week group exhibited compensation (predicted weight loss: -2.7 kg; actual weight loss: –1.5 kg) that was significantly higher than groups with lower exercise loads.[163] Nevertheless, daily mean step count was not different among groups, suggesting that compensation was due to energy intake rather than NEAT reduction.[163] A more recent 6-month randomized controlled trial of 171 adults with overweight or obesity reported compensation during 8 and 12 kcal/kg/week exercise groups. While both exercise groups lost body weight, there was greater compensation in the 12 kcal/kg/week group (2.7 vs. 1.5 kg).[164] However, there were no differences in PAEE or activity patterns from accelerometry; therefore, compensation was likely due to increased energy intake related to higher appetite ratings, cravings for sweet foods, and compensatory health beliefs.[164]

Notably, early investigations observed greater weight loss in response to exercise intervention in men compared with women, which may have been due in part to compensation.[165,166] Newer investigations in which men and women are prescribed the same exercise expenditure suggest that alterations in NEAT do not differ between sexes,[164,167] although these studies are not specifically designed to investigate compensation.

Taken together, these studies suggest that PAEE does not increase in a linear fashion with the addition of structured exercise. There may be a threshold of activity where increased TDEE can be anticipated, although where the threshold lies is not clear. This phenomenon is in line with a "constrained" model of TDEE, which is informed by evolution-based principles of energy conservation. According to this model, TDEE increases linearly with PAEE at low activity levels but plateaus at higher levels to maintain TDEE within a narrow range (i.e., compensation occurs) (Fig. 19.7). This theory is supported by observations of PAEE and TDEE across five different populations from Africa and North America.[168] In an investigation of PAEE and TDEE in over 300 adults, a plateau in TDEE was observed in the upper four deciles of activity from accelerometers.[168] The constrained EE hypothesis posits that, in response to high levels of PAEE, energy allocation to essential physiological processes such as immunity, reproduction, and somatic maintenance may be reduced and contribute to lower PAEE.[169] However, because these downstream behavioral effects and subsequent compensation likely occur only at very high levels of exercise or in certain populations (i.e., older adults), the utility of the constrained model of TDEE in obesity unclear.

KEY POINTS

- Weight loss is often successful in the short term but is followed by several powerful physiological adaptations that defend body weight, which drive weight regain. Following weight loss, altered gut hormone secretion (e.g., reductions in peptide YY 36, cholecystokinin, insulin, leptin, and amylin levels and increased ghrelin levels) and hypothalamic activity (i.e., agouti-related protein and proopiomelanocortin neuronal activity), along with reduced resting metabolic rate and physical activity energy expenditure, contribute to weight regain. Exercise helps mitigate weight regain, although total daily energy expenditure may not increase in a linear fashion with structured exercise regimens due to reductions in nonexercise activity thermogenesis.

CONCLUSIONS

A set of complex physiological processes orchestrate energy intake and expenditure to regulate body weight. Key hormones such as ghrelin increase appetite and energy intake, while other hormones such as GLP-1, PYY, amylin, and CCK suppress appetite and energy intake. Together with adipose tissue–derived metabolites, these signals are integrated in the CNS to regulate eating behavior. Humans must eat to sustain TDEE, of which the largest component is RMR, and the most variable behaviorally-driven component is PAEE. BAT may also contribute to TDEE, but the effect of activated BAT on long-term energy balance is unknown. Obesogenic environments that are prevalent in developed countries promote excess energy intake and lower PAEE, and consequent long-term weight gain. However, weight loss is not as simple or straightforward as "eat less and exercise more," due to the robust and persistent adaptations in appetite and EE that cultivate weight regain.

KEY POINTS

- The gastrointestinal (GI) tract plays a key role in energy regulation. Enteroendocrine cells across the entire length of the GI tract sense ingested nutrients and through secretion of orexigenic and anorectic gut hormones transmit signals of energy availability to the brain.
- In the central nervous system, the hypothalamus plays a key role in integrating short- and long-term signals of energy availability.
- Homeostatic signals of energy availability can be overridden through powerful hedonic responses to food cues, which evolved during conditions of sparse food availability. These drive eating beyond energy requirement and in conditions of ample energy availability can result in weight gain and obesity.
- Dysregulation of fasting- and meal-stimulated gut hormone secretion is closely linked with the pathophysiology of obesity, as well as the compensatory biological responses driving weight regain following energy restriction.
- Total daily energy expenditure (TDEE) is comprised of resting metabolic rate (RMR; largely determined by body composition), the thermic effect of food (determined by meal size and macronutrient composition), and physical activity energy expenditure (PAEE; the most variable component, determined by body size and movement).
- Recent evidence suggest that brown adipose tissue (BAT) is present in a large proportion of adults. Although activated BAT generates heat increases energy expenditure, the contribution of BAT to obesity development or management is unclear.
- The obesogenic environments that constitute many developed countries support food choices that are largely determined by cognitive, social, and emotional factors (rather than physiological processes). Low PAEE (but not necessarily low RMR) also contributes to the development of obesity.
- With weight loss, altered gut hormone secretion and hypothalamic activity as well as reduced resting metabolic rate and PAEE contribute to weight regain. Exercise helps mitigate weight regain, although TDEE may not increase in a linear fashion with structured exercise regimens.

REFERENCES

1. Gribble FM, Reimann F. Enteroendocrine cells: chemosensors in the intestinal epithelium. *Annu Rev Physiol*. 2016;78:277–299.
2. Monteiro MP, Batterham RL. The importance of the gastrointestinal tract in controlling food intake and regulating energy balance. *Gastroenterology*. 2017;152:1707–1717.e2.
3. Muller TD, Nogueiras R, Andermann ML, et al. Ghrelin. *Mol Metab*. 2015;4:437–460.
4. Li C, Navarrete J, Liang-Guallpa J, et al. Defined paraventricular hypothalamic populations exhibit differential responses to food contingent on caloric state. *Cell Metab*. 2018;29:681–694.
5. Manning S, Pucci A, Batterham RL. GLP-1: a mediator of the beneficial metabolic effects of bariatric surgery? *Physiol*. 2015;30:50–62.
6. Batterham RL, Cowley MA, Small CJ, et al. Gut hormone PYY 3-36 physiologically inhibits food intake. *Nature*. 2002;418:650.
7. De Silva A, Salem V, Long CJ, et al. The gut hormones PYY 3-36 and GLP-1 7-36 amide reduce food intake and modulate brain activity in appetite centers in humans. *Cell Metab*. 2011;14:700–706.
8. Schmidt JB, Gregersen NT, Pedersen SD, et al. Effects of PYY3–36 and GLP-1 on energy intake, energy expenditure, and appetite in overweight men. *Am J Physiol Endocrinol Metab*. 2014;306:E1248–E1256.
9. Neary NM, Small CJ, Druce MR, et al. Peptide YY3-36 and glucagon-like peptide-17-36 inhibit food intake additively. *Endocrinology*. 2005;146:5120–5127.
10. Nauck MA, Bartels E, Orskov C, et al. Additive insulinotropic effects of exogenous synthetic human gastric inhibitory polypeptide and glucagon-like peptide-1-(7-36) amide infused at near-physiological insulinotropic hormone and glucose concentrations. *J Clin Endocrinol Metab*. 1993;76:912–917.
11. Dockray GJ. Cholecystokinin. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:8–12.
12. Spreckley E, Murphy KG. The L-cell in nutritional sensing and the regulation of appetite. *Front Nutr*. 2015;2:23.
13. Panaro BL, Tough IR, Engelstoft MS, et al. The melanocortin-4 receptor is expressed in enteroendocrine L cells and regulates the release of peptide YY and glucagon-like peptide 1 in vivo. *Cell Metab*. 2014;20:1018–1029.
14. Brighton CA, Rievaj J, Kuhre RE, et al. Bile acids trigger GLP-1 release predominantly by accessing basolaterally located G protein-coupled bile acid receptors. *Endocrinology*. 2015;156:3961–3970.
15. Kroemer NB, Krebs L, Kobiella A, et al. (Still) longing for food: insulin reactivity modulates response to food pictures. *Hum Brain Mapp*. 2013;34:2367–2380.
16. Mietlicki-Baase EG. Amylin-mediated control of glycemia, energy balance, and cognition. *Physiol Behav*. 2016;162:130–140.
17. Cummings DE. Taste and the regulation of food intake: it's not just about flavor. *Am J Clin Nutr*. 2015;102:717–718.
18. Loch D, Breer H, Strotmann J. Endocrine modulation of olfactory responsiveness: effects of the orexigenic hormone ghrelin. *Chem Senses*. 2015;40:469–479.
19. Acosta A, Hurtado MD, Gorbatyuk O, et al. Salivary PYY: a putative bypass to satiety. *PLoS One*. 2011;6:e26137.
20. Fabian TK, Beck A, Fejerdy P, et al. Molecular mechanisms of taste recognition: considerations about the role of saliva. *Int J Mol Sci*. 2015;16:5945–5974.
21. Martin C, Passilly-Degrace P, Chevrot M, et al. Lipid-mediated release of GLP-1 by mouse taste buds from circumvallate papillae: putative involvement of GPR120 and impact on taste sensitivity. *J Lipid Res*. 2012;53:2256–2265.
22. Bustos-Saldana R, Alfaro-Rodriguez M, Solis-Ruiz Mde L, et al. Taste sensitivity diminution in hyperglycemic type 2 diabetics patients. *Rev Med Inst Mex Seguro Soc*. 2009;47:483–488.
23. Shin YK, Martin B, Golden E, et al. Modulation of taste sensitivity by GLP-1 signaling. *J Neurochem*. 2008;106:455–463.
24. Rosenbaum M, Leibel RL. 20 years of leptin: role of leptin in energy homeostasis in humans. *J Endocrinol*. 2014;223:T83–T96.
25. Roth CL, von Schnurbein J, Elfers C, et al. Changes in satiety hormones in response to leptin treatment in a patient with leptin deficiency. *Horm Res Paediatr*. 2018;90:1–7.
26. Izquierdo AG, Crujeiras AB, Casanueva FF, et al. Leptin, obesity, and leptin resistance: where are we 25 years later? *Nutrients*. 2019;11:2704.
27. Tsai VWW, Lin S, Brown DA, et al. Anorexia–cachexia and obesity treatment may be two sides of the same coin: role of the TGF-b superfamily cytokine MIC-1/GDF15. *Int J Obes*. 2015;40:193.
28. Tsai VW, Macia L, Johnen H, et al. TGF-b superfamily cytokine MIC-1/GDF15 is a physiological appetite and body weight regulator. *PLoS One*. 2013;8:e55174.
29. Tsai VW, Manandhar R, Jorgensen SB, et al. The anorectic actions of the TGFbeta cytokine MIC-1/GDF15 require an intact brainstem area postrema and nucleus of the solitary tract. *PLoS One*. 2014;9:e100370.
30. Witkamp RF. The role of fatty acids and their endocannabinoid-like derivatives in the molecular regulation of appetite. *Mol Aspects Med*. 2018;64:45–67.
31. Cork SC. The role of the vagus nerve in appetite control: implications for the pathogenesis of obesity. *J Neuroendocrinol*. 2018;30:e12643.
32. Kaelberer MM, Buchanan KL, Klein ME, et al. A gut-brain neural circuit for nutrient sensory transduction. *Science*. 2018;361:eaat5236.
33. Berthoud H-R, Neuhuber WL. Functional and chemical anatomy of the afferent vagal system. *Auton Neurosci*. 2000;85:1–17.
34. de Lartigue G, Xu C. Mechanisms of vagal plasticity influencing feeding behavior. *Brain Res*. 2018;1693:146–150.
35. Berthoud HR. Metabolic and hedonic drives in the neural control of appetite: who is the boss? *Curr Opin Neurobiol*. 2011;21:888–896.
36. Schwartz MW, Morton GJ. Obesity: keeping hunger at bay. *Nature*. 2002;418:595.
37. Manning S, Batterham RL. Enteroendocrine MC4R and energy balance: linking the long and the short of it. *Cell Metab*. 2014;20:929–931.

38. Makaronidis JM, Batterham RL. Obesity, body weight regulation and the brain: insights from fMRI. *Br J Radiol*. 2018:20170910.
39. Zanchi D, Depoorter A, Egloff L, et al. The impact of gut hormones on the neural circuit of appetite and satiety: a systematic review. *Neurosci Biobehav Rev*. 2017;80:457–475.
40. Batterham RL, Ffytche DH, Rosenthal JM, et al. PYY modulation of cortical and hypothalamic brain areas predicts feeding behaviour in humans. *Nature*. 2007;450:106–109.
41. Malik S, McGlone F, Bedrossian D, et al. Ghrelin modulates brain activity in areas that control appetitive behavior. *Cell Metab*. 2008;7:400–409.
42. Goldstone AP, Prechtl CG, Scholtz S, et al. Ghrelin mimics fasting to enhance human hedonic, orbitofrontal cortex, and hippocampal responses to food. *Am J Clin Nutr*. 2014;99:1319–1330.
43. Karra E, Chandarana K, Batterham RL. The role of peptide YY in appetite regulation and obesity. *J Physiol*. 2009;587:19–25.
44. Dijk SJ, Molloy PL, Varinli H, et al. Epigenetics and human obesity. *Int J Obes*. 2015;39:85–97.
45. *Human energy requirements: Report of a Joint FAO/WHO/UNU Expert Consultation. Food and Agriculture Organization, FAO Food and Nutrition Technical Report Series No. 1*. 2004. Rome, Italy Available at: http://www.fao.org/3/a-y5686e.pdf.
46. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol*. 1949;109:1–9.
47. Bader N, Bosy-Westphal A, Dilba B, et al. Intra- and interindividual variability of resting energy expenditure in healthy male subjects – biological and methodological variability of resting energy expenditure. *Br J Nutr*. 2007;94:843–849.
48. Melanson EL, Coelho LB, Tran ZV, et al. Validation of the BodyGem hand-held calorimeter. *Int J Obes Relat Metab Disord*. 2004;28:1479–1484.
49. Hipskind P, Glass C, Charlton D, et al. Do handheld calorimeters have a role in assessment of nutrition needs in hospitalized patients? A systematic review of literature. *Nutr Clin Pract*. 2011;26:426–433.
50. Harris JA, Benedict FG. A biometric study of human basal metabolism. *Proc Natl Acad Sci U S A*. 1918;4:370–373.
51. Mifflin MD, St Jeor ST, Hill LA, et al. A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr*. 1990;51:241–247.
52. Weijs PJ, Vansant GA. Validity of predictive equations for resting energy expenditure in Belgian normal weight to morbid obese women. *Clin Nutr*. 2010;29:347–351.
53. Weijs PJ. Validity of predictive equations for resting energy expenditure in US and Dutch overweight and obese class I and II adults aged 18-65 y. *Am J Clin Nutr*. 2008;88:959–970.
54. Elia M. Organ and Tissue Contribution to Metabolic Rate. In: Kinney JM, Tucker HN, eds. *Energy Metabolism: Tissue Determinants and Cellular Coeollaries*. New York: Raven Press; 1992:61–80.
55. Gallagher D, Belmonte D, Deurenberg P, et al. Organ-tissue mass measurement allows modeling of REE and metabolically active tissue mass. *Am J Physiol*. 1998;275:E249–E258.
56. Bosy-Westphal A, Müller MJ, Boschmann M, et al. Grade of adiposity affects the impact of fat mass on resting energy expenditure in women. *Br J Nutr*. 2008;101:474–477.
57. Gallagher D, Allen A, Wang Z, et al. Smaller organ tissue mass in the elderly fails to explain lower resting metabolic rate. *Ann N Y Acad Sci*. 2006;904:449–455.
58. Wang Z, Heshka S, Heymsfield SB, et al. A cellular-level approach to predicting resting energy expenditure across the adult years. *Am J Clin Nutr*. 2005;81:799–806.
59. Manini TM. Energy expenditure and aging. *Ageing Res Rev*. 2010;9:1.
60. Buchholz AC, Rafii M, Pencharz PB. Is resting metabolic rate different between men and women? *Br J Nutr*. 2001;86:641–646.
61. Curtis V, Henry CJK, Birch E. Intraindividual variation in the basal metabolic rate of women. Effect of the Menstrual Cycle. *Am J Hum Biol*. 1996;8:631–639.
62. Davidsen L, Vistisen B, Astrup A. Impact of the menstrual cycle on determinants of energy balance: a putative role in weight loss attempts. *Int J Obes*. 2007;31:1777–1785.
63. Day DS, Gozansky WS, Van Pelt RE, et al. Sex hormone suppression reduces resting energy expenditure and {beta}-adrenergic support of resting energy expenditure. *J Clin Endocrinol Metab*. 2005;90:3312–3317.
64. Melanson EL, Gavin KM, Shea KL, et al. Regulation of energy expenditure by estradiol in premenopausal women. *J Appl Physiol*. 2015;119:975–981.
65. Santosa S, Khosla S, McCready LK, Jensen MD. Effects of estrogen and testosterone on resting energy expenditure in older men. *Obesity (Silver Spring)*. 2010;18:2392–2394.
66. Gibney J, Wolthers T, Johannsson G, et al. Growth hormone and testosterone interact positively to enhance protein and energy metabolism in hypopituitary men. *Am J Physiol Endocrinol Metab*. 2005;289:E266–E271.
67. Westerterp KR. Diet induced thermogenesis. *Nutr Metab (Lond)*. 2004;5 (1):1–5.
68. Melanson KJ, Saltzman E, Vinken AG, et al. The effects of age on postprandial thermogenesis at four graded energetic challenges: findings in young and older women. *J Gerontol A Biol Sci Med Sci*. 1998;53:B409–B414.
69. Jones PP, Van Pelt RE, Johnson DG, et al. Role of sympathetic neural activation in age- and habitual exercise-related differences in the thermic effect of food. *J Clin Endocrinol Metab*. 2004;89:5138–5144.
70. Du S, Rajjo T, Santosa S, Jensen MD. The thermic effect of food is reduced in older adults. *Horm Metab Res*. 2014;46:365–369.
71. Donahoo WT, Levine JA, Melanson EL. Variability in energy expenditure and its components. *Curr Opin Clin Nutr Metab Care*. 2004;7:599–605.
72. Piaggi P, Krakoff J, Bogardus C, et al. Lower "awake and fed thermogenesis" predicts future weight gain in subjects with abdominal adiposity. *Diabetes*. 2013;62:4043–4051.
73. Weinsier RL, Nelson KM, Hensrud DD, et al. Metabolic predictors of obesity. Contribution of resting energy expenditure, thermic effect of food, and fuel utilization to four-year weight gain of post-obese and never-obese women. *J Clin Invest*. 1995;95:980–985.
74. Black AE, Coward WA, Cole TJ, et al. Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *Eur J Clin Nutr*. 1996;50:72–92.
75. Schoeller DA, van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol*. 1982;53:955–959.
76. International Atomic Energy Agency. *Assessment of Body Composition and Total Energy Expenditure in Humans Using Stable Isotope Techniques*; Vienna, Austria 2009.
77. Speakman JR, Nair KS, Goran MI. Revised equations for calculating CO2 production from doubly labeled water in humans. *Am J Physiol*. 1993;264:E912–E917.
78. O'Driscoll R, Turicchi J, Beaulieu K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br J Sports Med*. 2020;54:332–340.
79. Carpentier AC, Blondin DP, Virtanen KA, et al. Brown adipose tissue energy metabolism in humans. *Front Endocrinol*. 2018;9:447.
80. Morrison SF, Madden CJ, Tupone D. Central neural regulation of brown adipose tissue thermogenesis and energy expenditure. *Cell Metab*. 2014;19:741–756.
81. Heaton JM. The distribution of brown adipose tissue in the human. *J Anat*. 1972;112:35–39.
82. Au-Yong ITH, Thorn N, Ganatra R, et al. Brown adipose tissue and seasonal variation in humans. *Diabetes*. 2009;58:2583–2587.
83. Ouellet V, Labbe SM, Blondin DP, et al. Brown adipose tissue oxidative metabolism contributes to energy expenditure during acute cold exposure in humans. *J Clin Invest*. 2012;122:545–552.
84. Blondin DP, Labbé SM, Noll C, et al. Selective impairment of glucose but not fatty acid or oxidative metabolism in brown adipose tissue of subjects with type 2 diabetes. *Diabetes*. 2015;64:2388–2397.
85. Blondin DP, Labbé SM, Phoenix S, et al. Contributions of white and brown adipose tissues and skeletal muscles to acute cold-induced metabolic responses in healthy men. *J Physiol*. 2015;593:701–714.
86. Blondin DP, Labbé SM, Tingelstad HC, et al. Increased brown adipose tissue oxidative capacity in cold-acclimated humans. *J Clin Endocrinol Metab*. 2014;99:E438–E446.
87. van der Lans AAJJ, Hoeks J, Brans B, et al. Cold acclimation recruits human brown fat and increases nonshivering thermogenesis. *J Clin Invest*. 2013;123:3395–3403.

88. Yoneshiro T, Aita S, Matsushita M, et al. Age-related decrease in cold-activated brown adipose tissue and accumulation of body fat in healthy humans. *Obesity (Silver Spring)*. 2011;19:1755–1760.
89. Matsushita M, Yoneshiro T, Aita S, et al. Impact of brown adipose tissue on body fatness and glucose metabolism in healthy humans. *Int J Obes (Lond)*. 2014;38:812–817.
90. Rogers NH. Brown adipose tissue during puberty and with aging. *Ann Med*. 2015;47:142–149.
91. van Marken Lichtenbelt WD, Vanhommerig JW, Smulders NM, et al. Cold-activated brown adipose tissue in healthy men. *N Engl J Med*. 2009;360:1500–1508.
92. Vijgen GHEJ, Bouvy ND, Teule GJJ, et al. Brown adipose tissue in morbidly obese subjects. *PLoS One*. 2011;6:e17247.
93. Bartelt A, Heeren J. Adipose tissue browning and metabolic health. *Nat Rev Endocrinol*. 2014;10:24–36.
94. U Din M, Saari T, Raiko J, et al. Postprandial oxidative metabolism of human brown fat indicates thermogenesis. *Cell Metab*. 2018;28:207–216.e3.
95. U Din M, Raiko J, Saari T, et al. Human brown adipose tissue [(15)O]O2 PET imaging in the presence and absence of cold stimulus. *Eur J Nucl Med Mol Imaging*. 2016;43:1878–1886.
96. Bakker LEH, Boon MR, van der Linden RAD, et al. Brown adipose tissue volume in healthy lean south Asian adults compared with white Caucasians: a prospective, case-controlled observational study. *Lancet Diabetes Endocrinol*. 2014;2:210–217.
97. Hibi M, Oishi S, Matsushita M, et al. Brown adipose tissue is involved in diet-induced thermogenesis and whole-body fat utilization in healthy humans. *Int J Obes (Lond)*. 2016;40:1655–1661.
98. Nagai N, Sakane N, Ueno LM, et al. The -3826 A-->G variant of the uncoupling protein-1 gene diminishes postprandial thermogenesis after a high fat meal in healthy boys. *J Clin Endocrinol Metab*. 2003;88:5661–5667.
99. Li Y, Schnabl K, Gabler S-M, et al. Secretin-activated brown fat mediates prandial thermogenesis to induce satiation. *Cell*. 2018;175:1561–1574.e12.
100. Chen KY, Brychta RJ, Linderman JD, et al. Brown fat activation mediates cold-induced thermogenesis in adult humans in response to a mild decrease in ambient temperature. *J Clin Endocrinol Metab*. 2013;98:E1218–E1223.
101. Lee P, Smith S, Linderman J, et al. Temperature-acclimated brown adipose tissue modulates insulin sensitivity in humans. *Diabetes*. 2014;63:3686–3698.
102. Yoneshiro T, Aita S, Matsushita M, et al. Recruited brown adipose tissue as an antiobesity agent in humans. *J Clin Invest*. 2013;123:3404–3408.
103. Blondin DP, Daoud A, Taylor T, et al. Four-week cold acclimation in adult humans shifts uncoupling thermogenesis from skeletal muscles to brown adipose tissue. *J Physiol*. 2017;595:2099–2113.
104. Romu T, Vavruch C, Dahlqvist-Leinhard O, et al. A randomized trial of cold-exposure on energy expenditure and supraclavicular brown adipose tissue volume in humans. *Metabolism*. 2016;65:926–934.
105. Cypess AM, Weiner LS, Roberts-Toler C, et al. Activation of human brown adipose tissue by a beta3-adrenergic receptor agonist. *Cell Metab*. 2015;21:33–38.
106. O'Mara AE, Johnson JW, Linderman JD, et al. Chronic mirabegron treatment increases human brown fat, HDL cholesterol, and insulin sensitivity. *J Clin Invest*. 2020;130:2209–2219.
107. Blondin DP, Nielsen S, Kuipers EN, et al. Human brown adipocyte thermogenesis is driven by β2-AR stimulation. *Cell Metab*. 2020;32:287–300.e7.
108. Ono K, Tsukamoto-Yasui M, Hara-Kimura Y, et al. Intragastric administration of capsiate, a transient receptor potential channel agonist, triggers thermogenic sympathetic responses. *J Appl Physiol*. 2011;110:789–798.
109. Kawada T, Watanabe T, Takaishi T, et al. Capsaicin-induced beta-adrenergic action on energy metabolism in rats: influence of capsaicin on oxygen consumption, the respiratory quotient, and substrate utilization. *Proc Soc Exp Biol Med Soc Exp Biol Med (New York, NY)*. 1986;183:250–256.
110. Kawabata F, Inoue N, Masamoto Y, et al. Non-pungent capsaicin analogs (capsinoids) increase metabolic rate and enhance thermogenesis via gastrointestinal TRPV1 in mice. *Biosci Biotechnol Biochem*. 2009;73:2690–2697.
111. Nirengi S, Homma T, Inoue N, et al. Assessment of human brown adipose tissue density during daily ingestion of thermogenic capsinoids using near-infrared time-resolved spectroscopy. *J Biomed Opt*. 2016;21:91305.
112. Nomura S, Ichinose T, Jinde M, et al. Tea catechins enhance the mRNA expression of uncoupling protein 1 in rat brown adipose tissue. *J Nutr Biochem*. 2008;19:840–847.
113. Yoneshiro T, Matsushita M, Hibi M, et al. Tea catechin and caffeine activate brown adipose tissue and increase cold-induced thermogenic capacity in humans. *Am J Clin Nutr*. 2017;105:873–881.
114. Simon JJ, Wetzel A, Sinno MH, et al. Integration of homeostatic signaling and food reward processing in the human brain. *JCI Insight*. 2017;2:e92970.
115. Kroemer NB, Krebs L, Kobiella A, et al. Fasting levels of ghrelin covary with the brain response to food pictures. *Addict Biol*. 2013;18:855–862.
116. Hayes MR, Schmidt HD. GLP-1 influences food and drug reward. *Curr Opin Behav Sci*. 2016;9:66–70.
117. Tataranni PA, Harper IT, Snitker S, et al. Body weight gain in free-living Pima Indians: effect of energy intake vs expenditure. *Int J Obes*. 2003;27:1578–1583.
118. Piaggi P, Thearle MS, Bogardus C, et al. Lower energy expenditure predicts long-term increases in weight and fat mass. *J Clin Endocrinol Metab*. 2013;98:E703–E707.
119. Shook RP, Hand GA, Drenowatz C, et al. Low levels of physical activity are associated with dysregulation of energy intake and fat mass gain over 1 year. *Am J Clin Nutr*. 2015;102:1332–1338.
120. Lee I-M, Djoussé L, Sesso HD, et al. Physical activity and weight gain prevention. *J Am Med Assoc*. 2010;303:1173–1179.
121. DiPietro L, Kohl 3rd HW, Barlow CE, et al. Improvements in cardiorespiratory fitness attenuate age-related weight gain in healthy men and women: the Aerobics Center Longitudinal Study. *Int J Obes Relat Metab Disord*. 1998;22:55–62.
122. Sherwood NE, Jeffery RW, French SA, et al. Predictors of weight gain in the Pound of Prevention study. *Int J Obes Relat Metab Disord J Int Assoc Study Obes*. 2000;24:395–403.
123. Dombrowski SU, Knittle K, Avenell A, et al. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. *BMJ*. 2014;348:g2646.
124. Ross Middleton KM, Patidar SM, Perri MG. The impact of extended care on the long-term maintenance of weight loss: a systematic review and meta-analysis. *Obes Rev*. 2012;13:509–517.
125. Sumithran P, Prendergast LA, Delbridge E, et al. Long-term persistence of hormonal adaptations to weight loss. *N Engl J Med*. 2011;365:1597–1604.
126. Nymo S, Coutinho SR, Eknes PH, et al. Investigation of the long-term sustainability of changes in appetite after weight loss. *Int J Obes*. 2018;42:1489–1499.
127. Zhao X, Han Q, Gang X, et al. The role of gut hormones in diet-induced weight change: a systematic review. *Horm Metab Res*. 2017;49:816–825.
128. Cummings DE, Weigle DS, Frayo RS, et al. Plasma ghrelin levels after diet-induced weight loss or gastric bypass surgery. *N Engl J Med*. 2002;346:1623–1630.
129. Crujeiras AB, Goyenechea E, Abete I, et al. Weight regain after a diet-induced loss is predicted by higher baseline leptin and lower ghrelin plasma levels. *J Clin Endocrinol Metab*. 2010;95:5037–5044.
130. Sainsbury A, Zhang L. Role of the arcuate nucleus of the hypothalamus in regulation of body weight during energy deficit. *Mol Cell Endocrinol*. 2010;316:109–119.
131. Neseliler S, Hu W, Larcher K, et al. Neurocognitive and hormonal correlates of voluntary weight loss in humans. *Cell Metab*. 2019;29:39–49.e4.

132. Makaronidis JM, Neilson S, Cheung WH, et al. Reported appetite, taste and smell changes following Roux-en-Y gastric bypass and sleeve gastrectomy: effect of gender, type 2 diabetes and relationship to post-operative weight loss. *Appetite*. 2016;107:93–105.
133. Greenway FL. Physiological adaptations to weight loss and factors favouring weight regain. *Int J Obes*. 2015;39:1188–1196.
134. Rosenbaum M, Hirsch J, Gallagher DA, et al. Long-term persistence of adaptive thermogenesis in subjects who have maintained a reduced body weight. *Am J Clin Nutr*. 2008;88:906–912.
135. Fothergill E, Guo J, Howard L, et al. Persistent metabolic adaptation 6 years after "The Biggest Loser" competition. *Obesity (Silver Spring)*. 2016;24:1612–1619.
136. Heinitz S, Hollstein T, Ando T, et al. Early adaptive thermogenesis is a determinant of weight loss after six weeks of caloric restriction in overweight subjects. *Metabolism*. 2020;110:154303.
137. Ostendorf DM, Melanson EL, Caldwell AE, et al. No consistent evidence of a disproportionately low resting energy expenditure in long-term successful weight-loss maintainers. *Am J Clin Nutr*. 2018;108:658–666.
138. Nelson KM, Weinsier RL, James LD, et al. Effect of weight reduction on resting energy expenditure, substrate utilization, and the thermic effect of food in moderately obese women. *Am J Clin Nutr*. 1992;55:924–933.
139. Drenowatz C, Hill JO, Peters JC, et al. The association of change in physical activity and body weight in the regulation of total energy expenditure. *Eur J Clin Nutr*. 2017;71:377–382.
140. Ostendorf DM, Lyden K, Pan Z, et al. Objectively measured physical activity and sedentary behavior in successful weight loss maintainers. *Obesity (Silver Spring)*. 2018;26:53–60.
141. Fogelholm M, Kukkonen-Harjula K. Does physical activity prevent weight gain – a systematic review. *Obes Rev*. 2000;1:95–111.
142. Sjostrom L. Review of the key results from the Swedish Obese Subjects (SOS) trial—a prospective controlled intervention study of bariatric surgery. *J Intern Med*. 2013;273:219–234.
143. Yousseif A, Emmanuel J, Karra E, et al. Differential effects of laparoscopic sleeve gastrectomy and laparoscopic gastric bypass on appetite, circulating acyl-ghrelin, peptide YY3-36 and active GLP-1 levels in non-diabetic humans. *Obes Surg*. 2014;24:241–252.
144. Steinert RE, Feinle-Bisset C, Asarian L, et al. Ghrelin, CCK, GLP-1, and PYY (3–36): secretory controls and physiological roles in eating and glycemia in health, obesity, and after RYGB. *Physiol Rev*. 2016;97:411–463.
145. Bettini S, Bordigato E, Fabris R, et al. Modifications of resting energy expenditure after sleeve gastrectomy. *Obes Surg*. 2018;28:2481–2486.
146. Das SK, Roberts SB, McCrory MA, et al. Long-term changes in energy expenditure and body composition after massive weight loss induced by gastric bypass surgery. *Am J Clin Nutr*. 2003;78:22–30.
147. Carey DG, Pliego GJ, Raymond RL. Body composition and metabolic changes following bariatric surgery: effects on fat mass, lean mass and basal metabolic rate: six months to one-year follow-up. *Obes Surg*. 2006;16:1602–1608.
148. Ren Z-Q, Lu G-D, Zhang T-Z, et al. Effect of physical exercise on weight loss and physical function following bariatric surgery: a meta-analysis of randomised controlled trials. *BMJ Open*. 2018;8:e023208.
149. Dorling J, Broom DR, Burns SF, et al. Acute and chronic effects of exercise on appetite, energy intake, and appetite-related hormones: the modulating effect of adiposity, sex, and habitual physical activity. *Nutrients*. 2018;10:1140.
150. Broom DR, Stensel DJ, Bishop NC, et al. Exercise-induced suppression of acylated ghrelin in humans. *J Appl Physiol*. 2007;102:2165–2171.
151. King JA, Wasse LK, Ewens J, et al. Differential acylated ghrelin, peptide YY3-36, appetite, and food intake responses to equivalent energy deficits created by exercise and food restriction. *J Clin Endocrinol Metab*. 2011;96:1114–1121.
152. Martins C, Kulseng B, King NA, et al. The effects of exercise-induced weight loss on appetite-related peptides and motivation to eat. *J Clin Endocrinol Metab*. 2010;95:1609–1616.
153. Morishima T, Kurihara T, Hamaoka T, et al. Whole body, regional fat accumulation, and appetite-related hormonal response after hypoxic training. *Clin Physiol Funct Imaging*. 2014;34:90–97.
154. Rosenkilde M, Reichkendler MH, Auerbach P, et al. Appetite regulation in overweight, sedentary men after different amounts of endurance exercise: a randomized controlled trial. *J Appl Physiol*. 2013;115:1599–1609.
155. King NA, Caudwell PP, Hopkins M, et al. Dual-process action of exercise on appetite control: increase in orexigenic drive but improvement in meal-induced satiety. *Am J Clin Nutr*. 2009;90:921–927.
156. Guelfi KJ, Donges CE, Duffield R. Beneficial effects of 12 weeks of aerobic compared with resistance exercise training on perceived appetite in previously sedentary overweight and obese men. *Metabolism*. 2013;62:235–243.
157. Benito PJ, Cupeiro R, Ramos-Campo DJ, et al. A systematic review with meta-analysis of the effect of resistance training on whole-body muscle growth in healthy adult males. *Int J Environ Res Public Health*. 2020;17:1285.
158. Gilliat-Wimberly M, Manore MM, Woolf K, et al. Effects of habitual physical activity on the resting metabolic rates and body compositions of women aged 35 to 50 years. *J Am Diet Assoc*. 2001;101:1181–1188.
159. Bell C, Day DS, Jones PP, et al. High energy flux mediates the tonically augmented beta-adrenergic support of resting metabolic rate in habitually exercising older adults. *J Clin Endocrinol Metab*. 2004;89:3573–3578.
160. Colley RC, Hills AP, King NA, et al. Exercise-induced energy expenditure: implications for exercise prescription and obesity. *Patient Educ Couns*. 79:327–332.
161. Goran MI, Poehlman ET. Endurance training does not enhance total energy expenditure in healthy elderly persons. *Am J Physiol*. 1992;263:E950–E957.
162. Church TS, Earnest CP, Skinner JS, et al. Effects of different doses of physical activity on cardiorespiratory fitness among sedentary, overweight or obese postmenopausal women with elevated blood pressure: a randomized controlled trial. *J Am Med Assoc*. 2007;297:2081–2091.
163. Church TS, Martin CK, Thompson AM, et al. Changes in weight, waist circumference and compensatory responses with different doses of exercise among sedentary, overweight postmenopausal women. *PLoS One*. 2009;4:e4515–e4515.
164. Martin CK, Johnson WD, Myers CA, et al. Effect of different doses of supervised exercise on food intake, metabolism, and non-exercise physical activity: the E-MECHANIC randomized controlled trial. *Am J Clin Nutr*. 2019;110:583–592.
165. Donnelly JE, Hill JO, Jacobsen DJ, et al. Effects of a 16-month randomized controlled exercise trial on body weight and composition in young, overweight men and women: the Midwest Exercise Trial. *Arch Intern Med*. 2003;163:1343–1350.
166. Westerterp KR, Meijer GA, Janssen EM, et al. Long-term effect of physical activity on energy balance and body composition. *Br J Nutr*. 1992;68:21–30.
167. Donnelly JE, Honas JJ, Smith BK, et al. Aerobic exercise alone results in clinically significant weight loss for men and women: midwest exercise trial 2. *Obesity*. 2013;21:E219–E228.
168. Pontzer H, Durazo-Arvizu R, Dugas LR, et al. Constrained total energy expenditure and metabolic adaptation to physical activity in adult humans. *Curr Biol*. 2016;26:410–417.
169. Pontzer H. Constrained total energy expenditure and the evolutionary biology of energy balance. *Exerc Sport Sci Rev*. 2015;43:110–116.
170. Liu J, Conde K, Zhang P, et al. Enhanced AMPA receptor trafficking mediates the anorexigenic effect of endogenous glucagon-like peptide-1 in the paraventricular hypothalamus. *Neuron*. 2017;96:897–909.e5.
171. Pocai A. Action and therapeutic potential of oxyntomodulin. *Mol Metab*. 2014;3:241–251.
172. Nauck MA, Meier JJ. Incretin hormones: their role in health and disease. *Diabetes Obes Metab*. 2018;20:5–21.
173. Kentish SJ, O'Donnell TA, Isaacs NJ, et al. Gastric vagal afferent modulation by leptin is influenced by food intake status. *J Physiol*. 2013;591:1921–1934.

174. Reda TK, Geliebter A, Pi-Sunyer FX. Amylin, food intake, and obesity. *Obes Res*. 2002;10:1087–1091.
175. Mietlicki-Baase EG, McGrath LE, Koch-Laskowski K, et al. Amylin receptor activation in the ventral tegmental area reduces motivated ingestive behavior. *Neuropharmacology*. 2017;123:67–79.
176. Folgueira C, Sanchez-Rebordelo E, Barja-Fernandez S, et al. Uroguanylin levels in intestine and plasma are regulated by nutritional status in a leptin-dependent manner. *Eur J Nutr*. 2016;55:529–536.
177. Grunddal KV, Ratner CF, Svendsen B, et al. Neurotensin is coexpressed, coreleased, and acts together with GLP-1 and PYY in enteroendocrine control of metabolism. *Endocrinology*. 2016;157:176–194.

20

Obesity: Risk, Risk Factors, and Medical Management

Diana Barb, William Troy Donahoo, and Robert V. Considine

OUTLINE

Prevalence of Overweight and Obesity, 323
Social and Environmental Contributions to Obesity, 323
Diseases Associated With Obesity, 324
  Cardiovascular and Cerebrovascular Disease, 324
  Type 2 Diabetes Mellitus, 325
  Nonalcoholic Fatty Liver Disease, 325
  Gastrointestinal Complications, 326
  Cancer, 326
  Reproductive Health, 326
  Obstructive Sleep Apnea, 326
  Osteoarthritis, 327
  Cognitive Dysfunction and Alzheimer Disease, 327
Therapy for Obesity, 327
  Lifestyle Intervention, 327
  Pharmacotherapy, 328
  Bariatric Surgery, 330

PREVALENCE OF OVERWEIGHT AND OBESITY

Due to ease of use, good correlation with body fat in a population, and association with adverse health consequences, overweight and obesity are best defined in the clinic setting using the body mass index (BMI, which is weight in kilograms divided by height in meters squared).[1] Healthy weight is a BMI of 18.5 to 24.9 kg/m^2, overweight is a BMI of 25.0 to 29.9 kg/m^2, class 1 obesity is 30.0 to 34.9 kg/m^2, class 2 obesity is 35.0 to 34.9 kg/m^2, and class 3 obesity is greater than or equal to 40.0 kg/m^2.[1] There are differences by race and ethnicity in the BMI value at which an increased risk for cardiometabolic abnormalities occurs, including high fasting glucose, low high density lipoprotein cholesterol, high triglyceride levels, and hypertension.[2] Specifically, compared with non-Hispanic Whites at a BMI of 25 kg/m^2, a similar prevalence of cardiometabolic risk factors occur in non-Hispanic Blacks at a BMI of 22.9 kg/m^2, at a BMI of 21.5 kg/m^2 for Hispanics, and a BMI of 19.6 kg/m^2 for South Asians.[2] Nonetheless, organizations including the National Institutes of Health, Centers for Disease Control, and World Health Organization have maintained the cutoffs of overweight as a BMI of 25.0 to 29.9 kg/m^2 and obesity as a BMI greater than or equal to 30.0 kg/m^2. The 2019 Behavioural Risk Factor Surveillance System (BRFSS) report indicated that 32.4% of the US population was obese, and 34.6% was overweight.[3] By race/ethnicity in the BRFSS, non-Hispanic White adults had a prevalence of obesity of 29.9%, Hispanic adults had a prevalence of 33.8%, and non-Hispanic Black adults had the highest prevalence, of 39.8%.[3] Obesity rates in adults vary by age, with those 45 to 54 years old having the highest rates, at 37.6%, and those in the 18- to 24-year age range having the lowest rate, of 18.9%.[3] Overall, obesity rates in the United States decrease with increasing education, ranging from 36.2% in those without a high school diploma or equivalent to 25% in college graduates.[3]

Although BMI correlates with body fat, its use to relate to disease is compromised by the fact that most studies classify obesity by BMI rather than body composition.[4] It is well known that BMI is not the best tool to quantify fat mass, as it does not distinguish between fat mass, fat-free mass, and lean body mass, and that individuals with similar BMIs may have different body composition. In fact, sarcopenic obesity, characterized by low lean mass with increased fat mass, could be missed using BMI alone. Those with sarcopenic obesity have a 24% increased risk for all-cause mortality compared with those with nonsarcopenic obesity.[4] Padwal et al.[5] found that both low BMI and high body fat percentage are independently associated with increased mortality.

In children, a healthy weight is defined as being between the 5th and the 85th percentile, with overweight between the 85th and 95th percentile, and obesity above the 95th percentile.[6] In youth 2 to 19 years old, in 2015 to 2016 the prevalence of obesity was 18.5%, with the lowest prevalence amongst preschoolers (13.9%) and the highest amongst adolescents (18.9%), although there was no significant difference in obesity prevalence between boys and girls.[7] However, these epidemiological data, especially when self-reported and in adults, must be interpreted with great caution. When measured objectively using data from electronic health records in the OneFlorida statewide registry, obesity rates in adults were overall 27.8% higher than those reported by BRFSS over the same time period.[8] A similar study of pediatric obesity using the OneFlorida consortium was much closer to national published data and found an overall prevalence of obesity of 16.9%, with preschoolers experiencing 9.9% obesity and adolescents 18.9%.[9]

SOCIAL AND ENVIRONMENTAL CONTRIBUTIONS TO OBESITY

In addition to individual characteristics, multiple social and environmental factors have been shown to influence obesity. A recent study examined the contribution of factors such as demographics, socioeconomics, environment, and health care on the county level variation on obesity prevalence in the United States, and is a nice summary of the contribution of each of these factors to the complex disease of obesity.[10] Demographic factors (rurality, higher percentage of females, higher percentage of individuals younger than 18 or older than 65, and a higher percentage of racial and ethnic minorities) explained approximately 45% of the variation, socioeconomic factors (e.g., household income, percentage unemployed, percentage with food insecurity) explained another third of the variation, environmental factors (access to exercise and food environment index) explained 15%, and healthcare factors (percentage uninsured and number of primary care providers) explained less than 10%.[10] With respect to the impact of the work environment on obesity, it remains to be seen to what extent the shift to working from home (and the closure of gyms) with the Covid-19 pandemic will have on obesity rates, although the shift from jobs requiring moderate to vigorous physical activity to those that are sedentary has been associated with weight gain over the last several decades.[11] Beyond the socioeconomic status evaluated above, experimentally manipulated lower social hierarchy can cause increased caloric consumption and increased anxiety, and it is plausible that this also occurs at the population level[12]

DISEASES ASSOCIATED WITH OBESITY

As reviewed elsewhere,[13,14] adipose tissue dysfunction in obesity leads to increased lipolysis and release of free fatty acids (FFAs), and perturbation of adipocytokines with high levels of leptin and proinflammatory cytokines and low adiponectin levels. In addition, impairment of adipocyte storage and lack of recruitment of new adipocytes in subcutaneous fat depots lead to expansion of the visceral fat compartment and "ectopic" fat deposition in the hepatic, pericardial, intramyocellular, and intramuscular space. This results in development of a number of obesity-related diseases, many of which are associated with higher mortality rates.[15]

KEY POINTS

- As adipose tissue expands in obesity, it becomes "dysfunctional," and resistant to insulin, resulting in impaired lipid storage in the subcutaneous adipose depot. The increased rate of lipolysis and higher serum free fatty acids promote expansion of the visceral fat compartment, with ectopic deposition of fat in various organs, contributing to development of obesity-related diseases.

Cardiovascular and Cerebrovascular Disease

Obesity is an independent risk factor for cardiovascular disease (CVD). Evidence comes from large prospective epidemiological studies like the Framingham Heart Study,[16] the Nurses' Health Study,[17] and the Manitoba study.[18] Obesity is not only associated with increased risk of cardiovascular (CV) morbidity and mortality, but also with shorter longevity, and a greater proportion of life lived with CVD, when compared with individuals with normal BMI.[19] In addition, obesity contributes to higher CV risk through hypertension, atherogenic dyslipidemia, type 2 diabetes, and nonalcoholic fatty liver disease (NAFLD), as detailed below.

Hypertension. In large cohort studies[20-23] the prevalence of hypertension in both men and women increased with elevated BMI. In one study, the prevalence of hypertension was five times higher in adults with a BMI greater than 30 kg/m^2 compared with those with a BMI less than 20 kg/m^2.[20] In the Framingham study population of 5209 men and women, systolic blood pressure (BP) increased by 4 mm Hg for each 4.5 kg of increased weight.[23] In the Nurses' Health Study, the strongest risk factor for developing hypertension was an elevated BMI, with women with obesity having a 4.7 times higher incidence of hypertension than did women with a BMI less than 23 kg/m^2.[22] Moreover, subjects who were overweight or obese as children and remained in this weight category as adults had a 2.7 higher relative risk of developing high BP[24] and 1.7 times higher risk for carotid artery atherosclerosis (increased thickness of the carotid artery intima media).[24]

Weight gain or loss of 5 to 10 kg increased or reduced, respectively, the risk of high BP in women by 74% or 15%.[21] In general, the magnitude of weight loss correlates with a reduction in BP. Body weight loss of 5.1 kg after diet and exercise program reduced systolic BP by 4.4 mm Hg and diastolic BP by 3.5 mm Hg in a metaanalysis of 25 studies.[25] Long-term remission of hypertension with no need for BP medications was reported in approximately 38.2% after gastric bypass surgery.[26]

Dyslipidemia. The hallmark of obesity associated dyslipidemia is low high-density lipoprotein (HDL) cholesterol (HDL-C) levels.[27] In addition to decreased HDL, obesity is associated with increased triglycerides, total cholesterol, total low-density lipoprotein (LDL), and small, atherogenic, dense LDL, which is one of the drivers of increased CVD. Individuals with obesity since childhood who remain in the obese category as adults have a two to threee times higher risk of reduced HDL-C, elevated LDL-C levels, and increased triglyceride levels.[24]

Several mechanisms may contribute to low plasma HDL-C in obesity: 1) a relative lack of lipoprotein lipase activity, resulting in an increase in triglyceride-rich lipoproteins and reduction in HDL-C formation, 2) increased hydrolysis of triglyceride-rich HDL by hepatic lipase, and 3) lower levels of adiponectin, which has been suggested to regulate cholesterol efflux capacity of HDL, particle size, and composition.[27,28]

Coronary Artery Disease. Numerous observational studies show increased CVD disease risk with higher BMI,[13,19] with some studies identifying BMI as independent of other cardiometabolic risk factors.[29] In a metaanalysis of ten studies, central adiposity, as reflected by a high waist circumference or high waist-to-hip ratio, was a better indicator of CV risk factors (including hypertension, diabetes, and dyslipidemia) than BMI alone.[30]

In two large prospective studies, excess weight (BMI >30 kg/m^2) doubled the risk of coronary artery disease (CAD; relative risk 2.1 in men and 2.5 in women).[31] Obesity has also been associated with occurrence at a younger age of first non–ST-segment elevation myocardial infarction compared with lean subjects.[32] This was noted to be 12 years earlier in patients with severe obesity (BMI greater than 40 kg/m^2).[32] A recent large study (n = 3884, 58% male) among patients undergoing coronary computed tomography angiography (CTA) found an increased burden of CAD for individuals with BMI greater than 25 kg/m^2 compared with those with normal BMI.[33]

Metabolically healthy obesity (MHO), defined as obesity with normal BP, glucose, and lipid metabolism, occurs in 10% to 20% of obese subjects and has been associated with better cardiorespiratory fitness and significantly less risk of CV events and all-cause CV mortality compared with individuals with metabolically unhealthy obesity (MUO).[29] Patients with MHO exhibit less CAD than those with MUO, as measured by coronary CTA.[33,34] However, CV risk in MHO is higher than in lean, metabolically healthy individuals.[29] Moreover, in long-term follow-up studies (>8–10 years) almost half of those with MHO will convert to a metabolically unhealthy phenotype, with at least one metabolic abnormality.[13,35,36]

Modest weight loss (5%–10%) has been associated with significant improvement in CV risk factors.[37] The greater weight loss achieved with bariatric surgery (>20%) is associated with a reduced number of CV deaths (adjusted hazard ratio of 0.47 after bariatric surgery).[38]

Heart Failure. Obesity is associated with structural and functional changes of the heart that promote risk of heart failure (HF). Framingham Heart Study data indicate that a 1-kg/m^2 rise in BMI increases risk of heart failure by approximately 5% to 7%.[39] Conversely, the risk of HF declines with weight loss, and the greater weight loss achieved with bariatric surgery reduces HF risk more than usual care.[40] The most common form of obesity-associated HF, which has increased in incidence in the past decade, is characterized by impaired diastolic function and preserved left ventricular ejection fraction (pEF) in the absence of major coronary artery, valvular, or arrhythmic disease. This is referred to as HFpEF, or diastolic HF.[4]

There are multiple mechanisms involved in development of obesity-related cardiomyopathy and HF.[41] Elevated inflammatory cytokines[42] (TNF-α, IL-6, C-reactive protein), and trophic hormones (insulin and leptin), combined with activation of the renin-angiotensin-aldosterone system, promote cardiac fibrosis and endothelial dysfunction. Accumulation of intramyocardial lipid and excess myocardial ceramide production[43] contribute to diastolic dysfunction, increased cardiac output, and increased BP. Atherogenic dyslipidemia and intertwined comorbidities including diabetes mellitus, sleep apnea, pulmonary hypertension,[44] and obesity hypoventilation syndrome also contribute to HF.

Despite an increased risk of HF with excess adiposity/higher BMI, a number of studies report a protective effect of higher BMI on survival and other clinical outcomes in individuals with already established HF. Whether this phenomenon, termed "the obesity paradox,"[4,13,41] reflects an actual protective effect of excess body fat on HF-associated mortality remains a matter of debate. Potential explanations for the obesity paradox include: 1) comparison to unhealthy thinner patients in end-stage disease states, 2) statistical confounding, 3) survival bias, and/or lead time bias, and 4) greater body energy stores for periods of time when food intake is compromised. The presence of other comorbidities and possible treatment differences between BMI groups may also contribute.[4,41] Despite the obesity paradox, a number of studies have shown benefits of weight loss on outcomes of HF, including improvement in left ventricular function[45] and a more than 40% reduction in the risk of HF after bariatric surgery.[46]

Stroke. Obesity, as a chronic proinflammatory, prothrombotic state, is associated with increased risk of stroke. While this may be attributable in part to the higher prevalence of hypertension, dyslipidemia, and type 2 diabetes mellitus in obese individuals, there is evidence for an independent association between excess adiposity and stroke. In a large prospective cohort of male physicians with 12.5 years' follow-up, each 1-unit increase in BMI was associated with a 4% increase in risk of ischemic stroke (independent of other variables such as hypertension, hyperlipidemia, and diabetes) and 6% for hemorrhagic stroke.[47] Central/abdominal adiposity with high waist circumference or waist-to-hip ratio, in comparison with BMI alone, has been more strongly associated with increased risk of stroke and transient ischemic attack, independent of other vascular risk factors.[48]

KEY POINTS

- Obesity is an independent risk factor for cardiovascular disease and is associated with a 4- to 5-fold increased risk of hypertension, a 2-fold higher risk of coronary artery disease, and higher risk of heart failure with preserved left ventricular ejection fraction. Central adiposity also increases the risk of ischemic stroke and transient ischemic attack.

Type 2 Diabetes Mellitus

In the last decades the increasing prevalence of obesity has been accompanied by an increased incidence of metabolic syndrome and type 2 diabetes mellitus (T2DM). Among American adults, 1 in 3 have prediabetes, and 1 in 10 have T2DM.[49] Central obesity predicts future development of T2DM, with insulin resistance being the earliest abnormality seen in both prediabetes and overt T2DM.[14] Adipose tissue plays an important role in maintenance of total body glucose homeostasis by regulating the release of FFAs from stored triglycerides and through the production of adipocytokines that influence insulin sensitivity in muscle and liver.[14] In the progression from normoglycemia to impaired glucose tolerance to T2DM, insulin secretion increases to compensate for insulin resistance; however, despite high concentrations, insulin secretion becomes insufficient due to presence of β cell dysfunction, and eventually β cell failure.[50,51] The imbalance between insulin sensitivity, insulin secretion, and insulin action leads to an increase in both fasting and postprandial glucose concentrations.[52] The high levels of peripheral insulin concentrations are also, in part, due to decreased hepatic insulin clearance in the setting of obesity, ectopic fat deposition and NAFLD.[51,53]

Obesity management and weight loss can prevent, delay, and even reverse diabetes.[54] A 5.5% weight loss reduces the risk of conversion from prediabetes to diabetes by 58%,[55] and loss of 20% to 25% of body weight with bariatric surgery can result in complete remission of diabetes.[56,57] See Chapters 40 and 27 for in-depth discussion of pathophysiology of diabetes and effects of bariatric surgery to improve outcomes.

Nonalcoholic Fatty Liver Disease

The hepatic manifestation of obesity and metabolic syndrome, NAFLD, encompasses a spectrum of disorders characterized by hepatic steatosis (liver fat content ≥5% on biopsy) in the absence of significant alcohol consumption (<20 g per day) or other causes of liver disease.[58] NAFLD is associated with a 2-fold higher risk for incident diabetes and increased CV morbidity and mortality.[51,59] Its more severe form, nonalcoholic steatohepatitis (NASH) is characterized by the presence of inflammation, hepatocyte injury (ballooning and hepatocyte necroinflammation), and often fibrosis, which can lead to liver cirrhosis and hepatocellular carcinoma.[60] NASH is currently the second leading cause of liver transplantation in the United States,[58,60] and has an estimated global prevalence of 3.5%.[61]

Recent metaanalyses estimate the global prevalence of NAFLD to be 25% in the general population[61] and 56% in patients with T2DM.[62] As assessed by liver biopsy at the time of bariatric surgery, NAFLD was present in 79% to 83% of patients with morbid obesity.[63] Using the gold standard magnetic resonance spectroscopy,[64] the prevalence of NAFLD is 34% in general population and 70% in adults with T2DM.[65-67] Given the increased risk of liver fibrosis in patients with T2DM,[67] the American Diabetes Association recommended in 2019 that patients with T2DM and elevated plasma alanine aminotransferase or steatosis on imaging be evaluated for NASH and presence of liver fibrosis.[68] The purpose of screening is to identify patients with liver fibrosis and advise interventions that can potentially prevent the progression to decompensated cirrhosis.

Hepatic lipid storage in NAFLD is driven by increased FFA availability from unrestrained lipolysis in dysfunctional adipose tissue (lipotoxicity).[59] In addition, hyperinsulinemia (through activation of SREBP1c) and hyperglycemia (through ChREBP) will further promote hepatic triglyceride accumulation. Genetic predisposition may also play a role in steatosis. In genome-wide association studies, the rs738409 polymorphism in the patatin-like phospholipase 3 (*PNPLA3*) gene, found in 30% to 50% of individuals,[69] and the rs58542926 polymorphism in the transmembrane 6 superfamily member 2 (*TM6SP2*) gene[70] have been associated with greater liver triglyceride accumulation and worse prognosis. The progression of NAFLD to NASH has been linked to a number of factors, including insulin resistance[53] and abnormal mitochondrial function, with an early compensatory increase in mitochondrial oxidative capacity in NAFLD and loss of this hepatic mitochondrial adaptation or "flexibility" in NASH.[71,72]

Substantial improvement in NAFLD and histologic outcomes of NASH, including fibrosis, are directly proportional with the magnitude of weight loss. Weight loss of 5% to 7% decreases steatosis, with 8% to 10% weight loss needed to reverse steatohepatitis,[73,74] and greater than or equal to 10% weight reduction may also improve fibrosis.[73] Bariatric surgery–induced weight loss of greater than or equal to 20% improves steatosis in 90% of patients, results in resolution of NASH in 70% to 90%, and improves fibrosis in 50% to 65% of patients.[75,76]

KEY POINTS

- The hallmark of obesity-associated dyslipidemia is low high-density lipoprotein cholesterol. The increasing prevalence of metabolic syndrome, prediabetes, and diabetes parallels that of obesity. Similarly, nonalcoholic fatty liver disease (NAFLD), the hepatic manifestation of metabolic syndrome, is associated with a 2-fold higher risk for incident diabetes and increased cardiovascular morbidity and mortality. The more severe form of NAFLD, nonalcoholic steatohepatitis, is currently the second leading cause of liver transplantation in the United States.

Gastrointestinal Complications

In addition to NAFLD, obesity is a risk factor for a number of other gastrointestinal complications, including gastroesophageal reflux disease (GERD), pancreatitis, and gallstones. The prevalence of GERD in the general population is 8% to 33%, but this increases to 73% in the severely obese.[77] Chronic esophageal acid exposure can progress to Barrett's esophagus and esophageal adenocarcinoma.[78]

Acute pancreatitis can range from mild to severe and have an acute onset. Multiple large prospective population-based studies and meta-analyses have demonstrated that BMI is an independent predictor of severity and mortality from acute pancreatitis, with visceral adipose tissue content an additional important prognostic indicator for severity.[79] Mechanisms linking adiposity to acute pancreatitis include increased release of proinflammatory cytokines and fatty acids, and reduced levels of adiponectin.[80]

Obesity has a strong association with gallstone diseases. The incidence of cholelithiasis, cholecystitis, and cholesterolosis are all greater in subjects with obesity compared with normal-weight individuals. Mechanisms include supersaturated bile, gall bladder dysmotility, dyslipidemia, and hyperinsulinemia.[78]

It has recently been appreciated that bile acids regulate glucose homeostasis by activating the farnesoid X receptor (FXR). In the liver, FXR increases hepatic triglyceride clearance and suppresses gluconeogenesis. Activation of FXR in pancreatic β-cells increases calcium influx (by inhibiting plasma membrane potassium channels), resulting in greater glucose-stimulated insulin secretion. The increase in circulating bile acids in response to meal intake in subjects with obesity intake appears to be suppressed in some, but not all, studies. There is also evidence that the composition of the bile acids is altered in obesity. Thus, as gall bladder disease progresses in obesity, and the composition of bile acids is altered, dysfunction in the gland may contribute to worsening of insulin resistance.[81,82]

Cancer

The association between excess body weight and death from cancer has been recognized for approximately five decades. There is strong evidence that being overweight or obese in adulthood increases the risk of cancer of the mouth, pharynx, larynx, esophagus, stomach, pancreas, gall bladder, liver, kidney, colon, rectum, ovary, and prostate.[83] In a recent population cohort study of over 3.5 million people with a mean follow-up time of approximately 13 years, obesity (BMI 30.0–34.9 kg/m^2) increased the hazard ratio for cancer of the liver (1.66 [1.37–2.02]), uterus (2.08 [1.66–2.61]), and kidney (1.77 [1.46–2.14]) compared with normal-weight individuals (BMI 22.5–24.9 kg/m^2) who never smoked.[84] Hazard ratios for a number of additional cancers were also significantly increased. One mechanism linking adiposity to cancer risk is insulin resistance, resulting in elevated insulin acting through the insulin-like growth factor receptor to promote tumor cell growth.[85] Increased release of adipokines, inflammation, and oxidative stress in obesity also play a role. There is evidence that tumor cells may actively alter the function of surrounding adipocytes to promote secretion of lipids and adipokines that support tumor growth.[86]

In postmenopausal women there is a strong causal relationship between obesity and estrogen receptor positive breast cancer, attributed to estrogen release from adipose tissue. In contrast, there is an inverse relationship between breast cancer incidence and BMI in premenopausal women, which has been partly attributed to higher rates of amenorrhea in obese premenopausal women resulting in lower decreased circulating estrogen.[87]

Significant weight loss can reduce cancer risk. Following bariatric surgery there is a 33% decrease in the incidence of cancer (particularly in obesity-related cancers, such as colon, endometrial, and postmenopausal breast cancer) compared with a nonsurgical control.[88]

Reproductive Health

Polycystic ovarian syndrome (PCOS) is the most common endocrinopathy affecting reproductive-age women (see Chapter 124). Approximately 50% of women with PCOS are overweight or obese, demonstrating that excess weight is not necessary for PCOS to occur. However, obesity augments the adverse metabolic and reproductive outcomes of PCOS.[89] Women with PCOS are more likely to have increased visceral adiposity and obesity-induced insulin resistance, resulting in elevated insulin levels that sensitize thecal cells to luteinizing hormone, which results in increased androgen production.[90]

Obesity reduces fertility in women.[91,92] Time to pregnancy is also increased in couples where both the man and woman are obese.[93] Oligo- and anovulation are observed more frequently in overweight and obese women. Changes in the endometrium, oocyte, and embryo have also been reported.[94] Overweight and obese men have higher rates of oligo- and azoospermia, erectile dysfunction, and altered hypothalamic-pituitary-gonadal axis function.[94]

Maternal obesity during pregnancy is associated with increased rates of cesarean section and medical complications for both the mother and fetus. Pregnant women with obesity are at increased risk for gestational diabetes (see Chapter 34), preeclampsia, and hypertension. The health risks for babies of mothers with obesity include an increase in preterm birth, large-for-gestational-age birth, and congenital malformations including neural tube defect and CV anomaly. Maternal obesity also increases the risk for stillbirth, miscarriage, and fetal death.[95]

Obstructive Sleep Apnea

The prevalence of obstructive sleep apnea in obese individuals is 30%, and there is a strong association with visceral adiposity.[96] Mechanisms include fat deposition in the lateral fat pads and base of the tongue leading to narrowing of the airway, compression of the chest and abdominal wall, and a more collapsible trachea. Lung volumes can be reduced with obesity, and respiratory control may be altered.[97] Conversely, insufficient sleep and poor sleep quality are risk factors for development of obesity and type 2 diabetes.[98]

TABLE 20.1 **Weight-Loss Treatment Guidelines**

Treatment	BODY MASS INDEX (kg/m^2)				
	25–26.9	27–29.9	30–34.9	35–39.9	≥40
Diet, physical activity, and behavior therapy	Yes	Yes	Yes	Yes	Yes
Pharmacotherapy		In patients with obesity-related disease	Yes	Yes	Yes
Surgery				In patients with obesity-related disease	Yes

(Modified from Eckel RH. Clinical practice. Nonsurgical management of obesity in adults. *N Engl J Med.* 358:1941–1950, 2008; and Wadden TA, Butryn ML, Wilson C. Lifestyle modification for the management of obesity. *Gastroenterology.* 2007;132:2226–2238.)

Osteoarthritis

Obesity is a primary risk factor for development of osteoarthritis of the knee. In a metaanalysis of new-onset knee pain, 24.6% of cases were related to overweight or obesity.[99] Obesity increases mechanical stress on load-bearing joints and contributes to a higher prevalence of osteoarthritis in non–weight-bearing areas through an increase in proinflammatory cytokines produced by excess adipose tissue.[100]

Cognitive Dysfunction and Alzheimer Disease

Overweight and obesity in midlife are strongly associated with increased risk for cognitive decline. Abdominal obesity may be a better predictor of cognitive decline than overall adiposity. Obesity in late life is not associated with dementia risk. This obesity paradox is attributed to weight loss in the preclinical stage of dementia, as BMI starts to decline 6 to 10 years prior to dementia diagnosis.[101] Total brain volume and gray matter volumes are reduced in middle-aged subjects with obesity.[102,103]

Oxidative stress, inflammation, and insulin resistance in the brain have all been implicated as mechanisms altering neuronal function to impact cognition.[101,104,105] Obesity-associated inflammation and oxidative stress in the brain may result in a vicious cycle that leads to an impaired ability to regulate subsequent caloric intake, promoting further weight gain.[104,105]

THERAPY FOR OBESITY

Many factors can influence body weight; therefore, a thorough medical history and evaluation are required before therapy is initiated. The assessment of overweight and obese patients should include: 1) personal and family medical history, 2) consideration of medications that contribute to weight gain, such as corticosteroids, antidiabetic agents (insulin, thiazoladinediones, and some sulfonylureas), antidepressants and antipsychotic agents, and antihistamines and progestin-based contraceptives, 3) previous approaches to weight reduction, 4) patterns of food intake, including binge eating, and (5) physical activity levels. Measurement of waist circumference and evaluation of BP, blood glucose, lipids, and liver function should be performed. Treatment of coexisting conditions should follow standard guidelines. The patient's readiness for weight loss can be addressed with instruments such as the S-Weight and P-Weight questionaires,[106] which are based on the transtheoretical model that includes stages in the process of change. The absence of readiness should not preclude communication between provider and patient concerning the importance of weight reduction, as evidence that readiness predicts successful weight loss is not strong.[107]

The goal of weight loss is to improve patient health and quality of life by reducing and maintaining a lower body weight. Prevention of additional weight gain in individuals who cannot lose weight also provides benefit. The treatment options for weight loss include comprehensive lifestyle intervention (diet, exercise, and behavioral modification), pharmacotherapy, endoscopic treatment, and bariatric surgery (Chapter 27). Table 20.1 presents a summary of the treatment guidelines originally put forth by the National Heart, Lung, and Blood Institute,[108] which remain consistent with current recommendations of medical associations.[1,109,110]

Lifestyle Intervention

The combination of reduced-calorie diet and increased physical activity is the primary recommendation to achieve weight loss. High-intensity behavioral counseling with 14 or more visits over 6 months is required for the best outcomes. Behavioral therapy provides the techniques needed to adopt dietary and physical activity recommendations, which should include recording food intake, activity, and weight on a regular basis. Weekly meetings with a trained interventionist should include a review of progress, goal setting, and instructions for problem solving. Referring patients to a high-intensity community program such as the YMCA, which offers a version of the Diabetes Prevention program, can be a viable option. Commercial weight loss programs such as Weight Watchers and Jenny Craig have also been shown in the peer review literature to be effective. Telephone- and web-based interventions that include personalized provider feedback can in some cases be as effective as in-person counseling, at reduced cost. Less intensive lifestyle counseling can be an option to prevent weight gain in patients at low risk for concomitant disease.

Weight regain is common following completion of lifestyle intervention programs. Continued support (every other week or monthly) is useful in preventing weight regain but is often difficult to implement and maintain.[111]

Diet. Common weight-loss prescriptions target a reduction in dietary intake of 500 kcal/day and increase energy expenditure through exercise (see later) to achieve a weight loss of one to two pounds per week. To achieve the reduction in energy intake, the subject can be instructed to target consumption at 1,200 to 1,500 kcal/day for women and 1,500 to 1,800 kcal/day for men (varying levels by the individual's body weight), which can be assumed to produce a deficit compared with their habitual intake. The intake recommendation can be more specifically tailored using a number of different formulas that incorporate sex, weight, and age. Alternatively, subjects can be advised to avoid specific types of foods, which can lead to voluntary reductions in caloric intake.

TABLE 20.2 **Intermittent Fasting Approaches**

DIET	FEEDING DAY Calories Allowed	FASTING DAY Calories Allowed	FAST DAYS PER WEEK
Alternate-day fasting	*Ad libitum*	None	3–4
Alternate-day modified fasting	*Ad libitum*	25% over a 2-hour period	3–4
5:2 diet	*Ad libitum*	25% at any time	2 (sometimes consecutive)
Time-restricted feeding	*Ad libitum*	100% over a 10-hour period	7

(Adapted from Hoddy KK, Marlatt KL, Cetinkaya H, et al. Intermittent fasting and metabolic health: from religious fast to time-restricted feeding. *Obesity (Silver Spring)*. 2020;28:S29–S37.)

A number of diets that differ by macronutrient content have been tested and proven efficacious for weight loss including low-fat, low-carbohydrate, high-protein, low–glycemic index, and Mediterranean-style prescriptions. These diets work by reducing the number of calories consumed. For example, low-fat and low–glycemic index diets reduce the caloric density of foods ingested. Low-carbohydrate diets reduce food choice to decrease consumption, and high-protein diets have prolonged satiating effects. Given the large and varied number of options the patient and provider can choose, a diet that will result in the best adherence should be chosen, as this is the most important factor in success.[112]

Over the past several years there has been increased interest in the use of very low calorie ketogenic diets (VLCKDs) for weight loss. Very low calorie diets (VLCDs) provide 400 to 800 kcal/day, often in a liquid or bar product containing large amounts of dietary protein. To be classified as ketogenic, the diet can provide no more than 30 to 50 g/day of carbohydrate. VLCDs and VLCKDs can be used prior to bariatric surgery to reduce liver volume, and there is also some evidence for a reduction in the risk of complications following surgery.[113] Trials in non–bariatric surgery populations suggest that VLCKDs can safely induce substantial weight loss in the short term (4–6 weeks), with subjects reporting less hunger.[114] However, over the longer term, weight loss with VLCKDs is similar to that of diets with higher quantities of carbohydrate, and long-term adherence is difficult to maintain.[112]

Intermittent fasting has recently emerged as a new approach to alter caloric intake. There are many variations of intermittent fasting, including alternate-day fasting (ADF), alternate-day modified fasting (ADMF), the 5:2 diet, and time-restricted feeding (TRF). All of these interventions utilize prolonged periods of fasting (Table 20.2), resulting in variable amounts of weight loss. ADF and TRF do not restrict calories but may indirectly result in reduced energy intake. ADMF and 5:2 diets utilize very low-calorie diets. Intermittent fasting appears to have additional health benefits beyond weight loss, including decreased fasting insulin. Compliance is comparable to standard diets, and there is heterogeneity in response, suggesting that a personalized approach may be beneficial.[115]

Exercise. Increased physical activity combined with dietary restriction is central to all weight-loss programs. Exercise provides CV benefits, prevents loss of muscle mass during weight loss, reduces visceral fat, and promotes better dietary adherence through improved mood.[116,117] Exercise in the absence of caloric restriction can result in modest weight loss, but requires high intensity at a frequency of three to five times per week. Such exercise is difficult for most people with obesity to maintain for the long periods of time necessary to achieve significant weight loss.[116]

Recommendations for increasing physical activity include programmed aerobic exercise activities such as fast walking, running, and swimming. For substantial health benefits, adults are recommended to complete at least 150 minutes a week of moderate-intensity physical activity, or 75 minutes of vigorous-intensity aerobic activity, in episodes of at least 10 minutes throughout the week. For more substantial health benefits, these recommendations should be doubled.[118] Lifestyle modifications that increase "steps," such as parking further away from store entrances and taking stairs instead of elevators, also have modest effects.[116] Resistance training (lifting weights or working with elastic bands) improves body composition, with modest weight loss, and can be prescribed for subjects who have difficulty participating in aerobic activities.[119] The combination of aerobic and resistance training provides both CV and body composition benefits during weight loss.

There is good evidence that increased amounts of sedentary time, measured as daily sitting time, television viewing, or low counts on an activity monitor such as an accelerometer, are associated with increased risk for incident CVD and T2DM, and for all-cause CVD mortality. The effect of interventions to decrease sedentary behavior and the confounding role of BMI remain to be fully elucidated.[120]

Pharmacotherapy

Antiobesity medications are recommended for patients who fail to achieve clinically significant weight loss (≥5% of baseline weight) after 6 months of lifestyle interventions.[1,109] These should be prescribed as part of a plan that continues to include lifestyle interventions. Use of weight-loss drugs should be considered long-term, and no completely risk-free formulations are available. Thus, a balance must be achieved between risk and efficacy in a highly heterogeneous group of individuals with a wide range of risk factors for significant disease.

Pharmacologic treatment of overweight and obesity has been practiced for at least 80 years, and many formulations have been withdrawn from the market for safety reasons. Early weight-loss pharmacotherapy was based on the short-term use of amphetamine or structurally similar compounds. Currently, phentermine and diethylpropion remain approved for short-term use (12 weeks) in the United States. These drugs act primarily by inhibiting reuptake of norepinephrine and dopamine at nerve endings, which enhances *satiation* (level of fullness during consumption of a meal) and *satiety* (level of hunger after consumption of a meal).[121]

Anorexiants that modulate serotonin levels in the brain include fenfluramine, its isomer dexfenfluramine, and sibutramine. Fenfluramine and dexfenfluramine were removed from the market in 1997 due to increased incidence of valvular heart disease. Sibutramine, a serotonin and norepinephrine uptake inhibitor approved in 1997, was withdrawn in 2010 due to increased risk of major CV events.[121] Lorcaserin, a selective serotonin 5-HT2c receptor agonist approved in the United States in 2012, was voluntarily recalled in 2020 due an increased risk of certain cancers.[122]

Four medications and one device are currently approved for long-term weight loss. These are discussed later and summarized in Table 20.3.

Orlistat. A synthetic derivative of lipstatin (from *Streptomyces toxytricini*), orlistat inhibits gastric and pancreatic lipases, enzymes critical

TABLE 20.3 **US Food and Drug Administration—Approved Weight Loss Treatments**

Name	Mechanism of Action	Term of Use
Phentermine	Sympathomimetic	Short-term (3 months)
Diethylpropion	Sympathomimetic	Short-term (3 months)
Orlistat	Intestinal lipase inhibitor	Long-term
Phentermine-topiramate	Combination sympathomimetic and carbonic anhydrase inhibitor	Long-term
Bupropion-naltexone	Combination dopamine/norepinephrine reuptake inhibitor and mu-opioid receptor antagonist	Long-term
Liraglutide	Glucagon-like peptide-1 receptor agonist	Long-term
Gelesis100	Absorbent hydrogel in stomach	Long-term

for the digestion and absorption of fat from the gastrointestinal tract. Inhibition of lipases causes a reduction in the absorption of fat, resulting in increased excretion of triglycerides in feces. In a 2-year randomized crossover study of 688 obese individuals (BMI ~36 kg/m^2, weight ~100 kg), orlistat, 120 mg three times daily, caused a 10.2% decrease in initial body weight after the first year of treatment compared with a 6.1% decrease in the placebo group.[123] Several metaanalyses have shown favorable effects of orlistat on weight loss and weight maintenance. In addition, pooled data show significant overall beneficial effects of orlistat on cholesterol, triglycerides, BP, and hemoglobin A1c. Orlistat is currently available as a prescription medication (120 mg, Xenical) and a US Food and Drug Administration (FDA)-approved over-the-counter preparation marketed under the trade name Alli (60 mg).

Side effects of orlistat, which is not absorbed from the intestinal tract, are limited to gastrointestinal symptoms (fatty/oily stool, fecal urgency and incontinence, flatulence). A small but significant reduction in fat-soluble vitamin uptake can occur, and multivitamin supplementation is recommended.

Phentermine-Topiramate. Four fixed-dose combinations of phentermine plus topiramate were approved in 2012 for prescription use in the United States. The dose combinations are: starting (3.75 mg phentermine plus 23 mg topiramate); recommended (7.5 mg phentermine plus 46 mg topiramate); titration (11.25 mg phentermine plus 69 mg topiramate); and top (15 mg phentermine plus 92 mg topiramate). Both phentermine and topiramate alone cause weight loss and are available as generics. Topiramate has several actions, including blocking sodium and L-type calcium channels and inhibiting carbonic anhydrase activity, although its precise mechanism for inhibiting appetite is not well understood. Similarly, the weight-loss mechanism of the combination is not understood but is attributed to a reduction in appetite. In clinical studies, phentermine-topiramate reduced body weight by 10.6%, 8.4%, and 5.1% with 15/92 mg, 7.5/46 mg, and 3.75/23 mg, respectively ($P < 0.0001$).[124]

Phentermine-topiramate side effects include tachycardia, suicidal ideation, vision changes, mood changes, insomnia, paraesthesia, dizziness, confusion, dysgusia, constipation, and dry mouth. Topiramate is a recognized teratogen, at least at the higher doses used for the treatment of epilepsy. The combination of phentermine-topiramate carries a warning of birth defects (cleft lip and palate) in the offspring of pregnant women taking the combination therapy. Pregnant women should not take the drug, nor should women who are trying to conceive or who are at risk for becoming pregnant.[125]

Bupropion-Naltrexone. Bupropion, an atypical antidepressant, is an inhibitor of dopamine and norepinephrine uptake. Naltrexone is an opioid receptor antagonist approved for treatment of alcoholism and opioid addition. The combination of bupropion and naltrexone was approved in 2014, with a titrated total daily dose of bupropion 360 mg/naltrexone 32 mg. Weight loss following 52 weeks of treatment ranges between 2.5% and 5.2% of initial body weight. The most common side effects are nausea, constipation, headache, dizziness, insomnia, dry mouth, and diarrhea. There is no potential for abuse with this treatment.[126,127]

Liraglutide. Originally approved in 2010 for treatment of T2DM, this glucagon-like peptide-1 (GLP-1) analog received approval as a weight-loss modality by the FDA in 2014. Dosing is titrated to 1.8 mg/day for glycemic control and to 3.0 mg/day for weight loss. Substantial work in rodent models has shown that physiologic release of GLP-1 from the gut in response to nutritive molecules signals to the brain via vagal afferents. In contrast, GLP-1 analogs act directly in the brain to inhibit appetite.[128] Three large clinical trials demonstrate weight loss of 4.0% to 6.1% over 56 weeks.[129] Liraglutide treatment produces significant reductions in BP and triglycerides. A significant 13% reduction in major adverse CV outcomes in obese adults with type 2 diabetes was achieved with 1.8 mg/day liraglutide in the LEADER trial, which was specifically designed to assess CV outcomes.[129] The most common side effects of liraglutide are nausea, vomiting, and abdominal pain.

Gelesis100. This device for weight loss was approved by the FDA in 2019 for adults with a BMI of 25 to 40 kg/m^2, with or without comorbidities. Gelesis100 is a hydrogel matrix that is ingested as three capsules (2.25 g/dose) with 500 mL of water 20 to 30 minutes prior to lunch or dinner. The hydrogel absorbs water and expands to fill approximately one fourth of stomach volume to promote fullness. In a 6-month randomized double-blind placebo-controlled trial in 436 adults, Gelesis100 induced a significant weight loss of 6.4% versus 4.4% in the placebo group. Of note, 58.6% of subjects taking Gelesis100 lost greater than 5% of baseline weight compared with 42.2% of controls. Side effects were abdominal distension, infrequent bowel movements, and dyspepsia.[130] Gelesis100 is contraindicated in pregnancy and individuals with gastrointestinal reflux diseases or other conditions that affect gastrointestinal transit and motility.

Diabetes Treatments that Promote Weight Loss. Metformin suppresses gluconeogenesis and increases peripheral insulin sensitivity. In the Diabetes Prevention Trial, metformin without lifestyle intervention resulted in a 2-kg weight loss that was maintained during follow-up.[131] Several mechanisms have been suggested for how metformin may induce weight loss, including its ability to activate adenosine monophosphate–activated protein kinase. Most recently it has been demonstrated that metformin increases circulating levels of GDF15, which acts in the hindbrain to suppress food intake.[132] Pramlintide, an analog of the pancreatic hormone amylin, regulates postprandial glucose control. Pramlintide reduces food intake through actions in the brain, resulting in weight loss of approximately 6% at 1-year follow-up.[133]

Sodium-glucose transport-2 (SGLT2) inhibitors act in the kidney to reduce glucose reabsorption, resulting in lower plasma glucose levels. In clinical trials in subjects with T2DM and phase II trials in obese subjects without diabetes, SGLT2 inhibitors produce approximately 2 to 3 kg of weight loss. The mechanism is thought to be the caloric loss resulting from enhanced glucose excretion.[134] See Chapter 40 for a detailed discussion of diabetes treatment, use of these medications, and their mechanism of action.

Bariatric Surgery

Surgery remains the most effective means of inducing significant weight loss in individuals with extreme obesity. In addition, bariatric surgery-induced weight loss is effective in reducing obesity-related comorbidities, and a benefit on overall and cause-specific mortality has also been demonstrated.[135]

The National Institutes of Health established guidelines for the surgical treatment of obesity in 1991.[136] Eligible patients with a BMI greater than or equal to 40 kg/m^2 should be well-informed and motivated. In addition, they should have demonstrated an inability to lose weight with conventional therapy and have acceptable operative risks. Adults with a BMI greater than or equal to 35 kg/m^2 with a serious comorbidity including T2DM, hypertension, cardiomyopathy, sleep apnea, or severe joint disease may also be surgical candidates. There is increasing evidence to suggest that surgery may be beneficial for patients with a BMI of 30 to 35 kg/m^2 who have T2DM.[135] Bariatric surgery should be performed in conjunction with a comprehensive follow-up plan consisting of nutritional, behavioral, and medical monitoring.

There are several mechanisms that have been proposed for how bariatric surgery induces weight loss and improves comorbidities beyond its effect to reduce caloric intake. See Chapter 27 for a detailed discussion of bariatric surgery techniques, mechanisms, and outcomes.

KEY POINTS

- Lifestyle modification is the first-line treatment for overweight/obesity. Pharmacologic treatment and bariatric surgery should be used in conjunction with lifestyle modification for best outcomes. Weight loss has a significant ability to decrease comorbidities associated with excess adiposity.

Acknowledgments

The authors would like to acknowledge Jose Caro, Michael Cowley, and Wendy A. Brown for their work on this chapter in previous editions.

REFERENCES

1. Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Obesity Society. *Circulation*. 2014;129:S102–S138.
2. Gujral UP, Vittinghoff E, Mongraw-Chaffin M, et al. Cardiometabolic abnormalities among normal-weight persons from five racial/ethnic groups in the United States: a cross-sectional analysis of two cohort studies. *Ann Intern Med*. 2017;166:628–636.
3. Centers for Disease Control and Prevention. National center for chronic disease prevention and health promotion, Division of population health. *BRFSS Prevalence & Trends Data [online]*. https://www.cdc.gov/brfss/brfssprevalence/. Published 2015. Accessed Mar 01, 2021.
4. Carbone S, Lavie CJ, Arena R. Obesity and heart failure: focus on the obesity paradox. *Mayo Clin Proc*. 2017;92:266–279.
5. Padwal R, Leslie WD, Lix LM, et al. Relationship among body fat percentage, body mass index, and all-cause mortality: a cohort study. *Ann Intern Med*. 2016;164:532–541.
6. Barlow SE, Expert C. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics*. 2007;120:S164–S192.
7. Hales CM, Carroll MD, Fryar CD, et al. *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016*. NCHS Data Brief; 2017:1–8.
8. Filipp SL, Cardel M, Hall J, et al. Characterization of adult obesity in Florida using the OneFlorida clinical research consortium. *Obes Sci Pract*. 2018;4:308–317.
9. Lemas DJ, Cardel MI, Filipp SL, et al. Objectively measured pediatric obesity prevalence using the OneFlorida Clinical Research Consortium. *Obes Res Clin Pract*. 2019;13:12–15.
10. Scheinker D, Valencia A, Rodriguez F. Identification of factors associated with variation in US county-level obesity prevalence rates using epidemiologic vs machine learning models. *JAMA Netw Open*. 2019;2:e192884.
11. Church TS, Thomas DM, Tudor-Locke C, et al. Trends over 5 decades in U.S. occupation-related physical activity and their associations with obesity. *PLoS One*. 2011;6:e19657.
12. Bratanova B, Loughnan S, Klein O, et al. Poverty, inequality, and increased consumption of high calorie food: experimental evidence for a causal link. *Appetite*. 2016;100:162–171.
13. Bray GA, Heisel WE, Afshin A, et al. The science of obesity management: an Endocrine Society scientific statement. *Endocr Rev*. 2018;39:79–132.
14. DeFronzo RA. Pathogenesis of type 2 diabetes mellitus. *Med Clin North Am*. 2004;88:787–835, ix.
15. Global BMIMC, Di Angelantonio E, Bhupathiraju S, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*. 2016;388:776–786.
16. Hubert HB, Feinleib M, McNamara PM, et al. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*. 1983;67:968–977.
17. Manson JE, Colditz GA, Stampfer MJ, et al. A prospective study of obesity and risk of coronary heart disease in women. *N Engl J Med*. 1990;322:882–889.
18. Rabkin SW, Mathewson FA, Hsu PH. Relation of body weight to development of ischemic heart disease in a cohort of young North American men after a 26 year observation period: the Manitoba Study. *Am J Cardiol*. 1977;39:452–458.
19. Khan SS, Ning H, Wilkins JT, et al. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA Cardiol*. 2018;3:280–287.
20. Rabkin SW, Chen Y, Leiter L, et al. Risk factor correlates of body mass index. Canadian Heart Health Surveys Research Group. *CMAJ*. 1997;157:S26–S31.
21. Huang Z, Willett WC, Manson JE, et al. Body weight, weight change, and risk for hypertension in women. *Ann Intern Med*. 1998;128:81–88.
22. Forman JP, Stampfer MJ, Curhan GC. Diet and lifestyle risk factors associated with incident hypertension in women. *J Am Med Assoc*. 2009;302:401–411.
23. Higgins M, Kannel W, Garrison R, et al. Hazards of obesity--the Framingham experience. *Acta Med Scand Suppl*. 1988;723:23–36.
24. Juonala M, Magnussen CG, Berenson GS, et al. Childhood adiposity, adult adiposity, and cardiovascular risk factors. *N Engl J Med*. 2011;365:1876–1885.
25. Neter JE, Stam BE, Kok FJ, et al. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. *Hypertension*. 2003;42:878–884.
26. Puzziferri N, Roshek 3rd TB, Mayo HG, et al. Long-term follow-up after bariatric surgery: a systematic review. *J Am Med Assoc*. 2014;312:934–942.
27. Stadler JT, Marsche G. Obesity-related changes in high-density lipoprotein metabolism and function. *Int J Mol Sci*. 2020:21:8985.
28. Woudberg NJ, Goedecke JH, Blackhurst D, et al. Association between ethnicity and obesity with high-density lipoprotein (HDL) function and subclass distribution. *Lipids Health Dis*. 2016;15:92.

29. Bluher M. Metabolically healthy obesity. *Endocr Rev*. 2020;41:405–20.
30. Lee CM, Huxley RR, Wildman RP, et al. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: a meta-analysis. *J Clin Epidemiol*. 2008;61:646–653.
31. Flint AJ, Hu FB, Glynn RJ, et al. Excess weight and the risk of incident coronary heart disease among men and women. *Obesity (Silver Spring)*. 2010;18:377–383.
32. Madala MC, Franklin BA, Chen AY, et al. Obesity and age of first non-ST-segment elevation myocardial infarction. *J Am Coll Cardiol*. 2008;52:979–985.
33. Hulten EA, Bittencourt MS, Preston R, et al. Obesity, metabolic syndrome and cardiovascular prognosis: from the Partners coronary computed tomography angiography registry. *Cardiovasc Diabetol*. 2017;16:14.
34. Hulten E, Bittencourt MS, O'Leary D, et al. Cardiometabolic risk is associated with atherosclerotic burden and prognosis: results from the partners coronary computed tomography angiography registry. *Diabetes Care*. 2014;37:555–564.
35. Eckel N, Li Y, Kuxhaus O, et al. Transition from metabolic healthy to unhealthy phenotypes and association with cardiovascular disease risk across BMI categories in 90 257 women (the Nurses' Health Study): 30 year follow-up from a prospective cohort study. *Lancet Diabetes Endocrinol*. 2018;6:714–724.
36. Hamer M, Bell JA, Sabia S, et al. Stability of metabolically healthy obesity over 8 years: the English longitudinal study of ageing. *Eur J Endocrinol*. 2015;173:703–708.
37. Wing RR, Lang W, Wadden TA, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. *Diabetes Care*. 2011;34:1481–1486.
38. Sjostrom L, Peltonen M, Jacobson P, et al. Bariatric surgery and long-term cardiovascular events. *J Am Med Assoc*. 2012;307:56–65.
39. Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med*. 2002;347:305–313.
40. Jamaly S, Carlsson L, Peltonen M, et al. Surgical obesity treatment and the risk of heart failure. *Eur Heart J*. 2019;40:2131–2138.
41. Ebong IA, Goff Jr DC, Rodriguez CJ, et al. Mechanisms of heart failure in obesity. *Obes Res Clin Pract*. 2014;8:e540–e548.
42. Bahrami H, Bluemke DA, Kronmal R, et al. Novel metabolic risk factors for incident heart failure and their relationship with obesity: the MESA (Multi-Ethnic Study of Atherosclerosis) study. *J Am Coll Cardiol*. 2008;51:1775–1783.
43. Park TS, Goldberg IJ. Sphingolipids, lipotoxic cardiomyopathy, and cardiac failure. *Heart Fail Clin*. 2012;8:633–641.
44. Friedman SE, Andrus BW. Obesity and pulmonary hypertension: a review of pathophysiologic mechanisms. *J Obes*. 2012;2012:505274.
45. Lavie CJ, Milani RV, Ventura HO. Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss. *J Am Coll Cardiol*. 2009;53:1925–1932.
46. Shimada YJ, Tsugawa Y, Brown DFM, et al. Bariatric surgery and emergency department visits and hospitalizations for heart failure exacerbation: population-based, self-controlled series. *J Am Coll Cardiol*. 2016;67:895–903.
47. Kurth T, Gaziano JM, Berger K, et al. Body mass index and the risk of stroke in men. *Arch Intern Med*. 2002;162:2557–2562.
48. Winter Y, Rohrmann S, Linseisen J, et al. Contribution of obesity and abdominal fat mass to risk of stroke and transient ischemic attacks. *Stroke*. 2008;39:3145–3151.
49. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2020*. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2020.
50. DeFronzo RA, Ferrannini E, Groop L, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers*. 2015;1:15019.
51. Gastaldelli A, Cusi K. From NASH to diabetes and from diabetes to NASH: mechanisms and treatment options. *JHEP Rep*. 2019;1:312–328.
52. Lillioja S, Mott DM, Howard BV, et al. Impaired glucose tolerance as a disorder of insulin action. Longitudinal and cross-sectional studies in Pima Indians. *N Engl J Med*. 1988;318:1217–1225.
53. Bril F, Barb D, Portillo-Sanchez P, et al. Metabolic and histological implications of intrahepatic triglyceride content in nonalcoholic fatty liver disease. *Hepatology*. 2017;65:1132–1144.
54. American Diabetes Association. 8. Obesity management for the treatment of type 2 diabetes: standards of medical care in diabetes-2021. *Diabetes Care*. 2021;44:S100–S110.
55. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*. 2009;374:1677–1686.
56. Sjostrom L, Lindroos AK, Peltonen M, et al. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med*. 2004;351:2683–2693.
57. Schauer PR, Bhatt DL, Kirwan JP, et al. Bariatric surgery versus intensive medical therapy for diabetes—5-year outcomes. *N Engl J Med*. 2017;376:641–651.
58. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67:328–357.
59. Cusi K. Role of obesity and lipotoxicity in the development of nonalcoholic steatohepatitis: pathophysiology and clinical implications. *Gastroenterology*. 2012;142:711–725.e716.
60. Stefan N, Haring HU, Cusi K. Non-alcoholic fatty liver disease: causes, diagnosis, cardiometabolic consequences, and treatment strategies. *Lancet Diabetes Endocrinol*. 2019;7:313–324.
61. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*. 2018;15:11–20.
62. Younossi ZM, Golabi P, de Avila L, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. *J Hepatol*. 2019;71:793–801.
63. Kleiner DE, Berk PD, Hsu JY, et al. Hepatic pathology among patients without known liver disease undergoing bariatric surgery: observations and a perspective from the longitudinal assessment of bariatric surgery (LABS) study. *Semin Liver Dis*. 2014;34:98–107.
64. Szczepaniak LS, Nurenberg P, Leonard D, et al. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. *Am J Physiol Endocrinol Metab*. 2005;288:E462–E468.
65. Bril F, Cusi K. Management of nonalcoholic fatty liver disease in patients with type 2 diabetes: a call to action. *Diabetes Care*. 2017;40:419–430.
66. Barb D, Portillo-Sanchez P, Cusi K. Pharmacological management of nonalcoholic fatty liver disease. *Metabolism*. 2016;65:1183–1195.
67. Lomonaco R, Godinez Leiva E, Bril F, et al. Advanced liver fibrosis is common in patients with type 2 diabetes followed in the outpatient setting: the need for systematic screening. *Diabetes Care*. 2020.
68. American Diabetes Association. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S37–S47.
69. He S, McPhaul C, Li JZ, et al. A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J Biol Chem*. 2010;285:6706–6715.
70. Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet*. 2014;46:352–356.
71. Koliaki C, Szendroedi J, Kaul K, et al. Adaptation of hepatic mitochondrial function in humans with non-alcoholic fatty liver is lost in steatohepatitis. *Cell Metab*. 2015;21:739–746.
72. Sunny NE, Parks EJ, Browning JD, et al. Excessive hepatic mitochondrial TCA cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. *Cell Metab*. 2011;14:804–810.
73. Vilar-Gomez E, Martinez-Perez Y, Calzadilla-Bertot L, et al. Weight loss through lifestyle modification significantly reduces features of nonalcoholic steatohepatitis. *Gastroenterology*. 2015;149:367–378. e365; quiz e314–e365.
74. Promrat K, Kleiner DE, Niemeier HM, et al. Randomized controlled trial testing the effects of weight loss on nonalcoholic steatohepatitis. *Hepatology*. 2010;51:121–129.
75. Mummadi RR, Kasturi KS, Chennareddygari S, et al. Effect of bariatric surgery on nonalcoholic fatty liver disease: systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2008;6:1396–1402.

76. Lassailly G, Caiazzo R, Buob D, et al. Bariatric surgery reduces features of nonalcoholic steatohepatitis in morbidly obese patients. *Gastroenterology*. 2015;149:379–388;quiz e315–e376.
77. Suter M. Gastroesophageal reflux disease, obesity, and Roux-en-Y gastric bypass: complex relationship—a narrative review. *Obes Surg*. 2020;30:3178–3187.
78. Camilleri M, Malhi H, Acosta A. Gastrointestinal complications of obesity. *Gastroenterology*. 2017;152:1656–1670.
79. Kuan LL, Dennison AR, Garcea G. Association of visceral adipose tissue on the incidence and severity of acute pancreatitis: a systematic review. *Pancreatology*. 2020;20:1056–1061.
80. Khatua B, El-Kurdi B, Singh VP. Obesity and pancreatitis. *Curr Opin Gastroenterol*. 2017;33:374–382.
81. Cortes VA, Barrera F, Nervi F. Pathophysiological connections between gallstone disease, insulin resistance, and obesity. *Obes Rev*. 2020;21: e12983.
82. McGlone ER, Bloom SR. Bile acids and the metabolic syndrome. *Ann Clin Biochem*. 2019;56:326–337.
83. World Cancer Research Fund/American Institute for Cancer Research. *Diet, Nutrition, Physical Activity and Cancer: A Global Perspective.* Continuous Update Project Expert Report; 2018. Available at: dietandcancerreport.org. Accessed April 2021.
84. Bhaskaran K, Dos-Santos-Silva I, Leon DA, et al. L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *Lancet Diabetes Endocrinol*. 2018;6:944–953.
85. Gallagher EJ, LeRoith D. Hyperinsulinaemia in cancer. *Nat Rev Cancer*. 2020;20:629–644.
86. Dumas JF, Brisson L. Interaction between adipose tissue and cancer cells: role for cancer progression. *Cancer Metastasis Rev*. 2021;40:31–46.
87. Scully T, Ettela A, LeRoith D, et al. Obesity, Type 2 diabetes, and cancer risk. *Front Oncol*. 2020;10:615375.
88. Schauer DP, Feigelson HS, Koebnick C, et al. Bariatric surgery and the risk of cancer in a large multisite cohort. *Ann Surg*. 2019;269:95–101.
89. Teede HJ, Misso ML, Costello MF, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertil Steril*. 2018;110:364–379.
90. Glueck CJ, Goldenberg N. Characteristics of obesity in polycystic ovary syndrome: Etiology, treatment, and genetics. *Metabolism*. 2019;92:108–120.
91. McKinnon CJ, Hatch EE, Rothman KJ, et al. Body mass index, physical activity and fecundability in a North American preconception cohort study. *Fertil Steril*. 2016;106:451–459.
92. Loy SL, Cheung YB, Soh SE, et al. Female adiposity and time-to-pregnancy: a multiethnic prospective cohort. *Hum Reprod*. 2018;33:2141–2149.
93. Sundaram R, Mumford SL, Buck Louis GM. Couples' body composition and time-to-pregnancy. *Hum Reprod*. 2017;32:662–668.
94. Mintziori G, Nigdelis MP, Mathew H, et al. The effect of excess body fat on female and male reproduction. *Metabolism*. 2020;107:154193.
95. Marchi J, Berg M, Dencker A, et al. Risks associated with obesity in pregnancy, for the mother and baby: a systematic review of reviews. *Obes Rev*. 2015;16:621–638.
96. Pillar G, Shehadeh N. Abdominal fat and sleep apnea: the chicken or the egg? *Diabetes Care*. 2008;31:S303–S309.
97. Joosten SA, Hamilton GS, Naughton MT. Impact of weight loss management in OSA. *Chest*. 2017;152:194–203.
98. Reutrakul S, Van Cauter E. Sleep influences on obesity, insulin resistance, and risk of type 2 diabetes. *Metabolism*. 2018;84:56–66.
99. Silverwood V, Blagojevic-Bucknall M, Jinks C, et al. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage*. 2015;23:507–515.
100. Wang T, He C. Pro-inflammatory cytokines: the link between obesity and osteoarthritis. *Cytokine Growth Factor Rev*. 2018;44:38–50.
101. Tanaka H, Gourley DD, Dekhtyar M, et al. Cognition, brain structure, and brain function in individuals with obesity and related disorders. *Curr Obes Rep*. 2020;9:544–549.
102. Hamer M, Batty GD. Association of body mass index and waist-to-hip ratio with brain structure: UK Biobank study. *Neurology*. 2019;92:e594–e600.
103. Dekkers IA, Jansen PR, Lamb HJ. Obesity, brain volume, and white matter microstructure at MRI: a cross-sectional UK Biobank study. *Radiology*. 2019;291:763–771.
104. Mullins CA, Gannaban RB, Khan MS, et al. Neural underpinnings of obesity: the role of oxidative stress and inflammation in the brain. *Antioxidants (Basel)*. 2020;9:1018.
105. Crispino M, Trinchese G, Penna E, et al. Interplay between peripheral and central inflammation in obesity-promoted disorders: the impact on synaptic mitochondrial functions. *Int J Mol Sci*. 2020:21:5964.
106. Ceccarini M, Borrello M, Pietrabissa G, et al. Assessing motivation and readiness to change for weight management and control: an in-depth evaluation of three sets of instruments. *Front Psychol*. 2015;6:511.
107. Casazza K, Brown A, Astrup A, et al. Weighing the evidence of common beliefs in obesity research. *Crit Rev Food Sci Nutr*. 2015;55:2014–2053.
108. The Practical Guide. Identification, evaluation, and treatment of overweight and obesity in adults. Available at: https://www.nhlbi.nih.gov/files/docs/guidelines/prctgd_c.pdf.
109. Garvey WT, Mechanick JI, Brett EM, et al. American Association of Clinical Endocrinologists and American College of Endocrinology comprehensive clinical practice guidelines for medical care of patients with obesity. *Endocr Pract*. 2016;22:1–203.
110. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults: US Preventive Services Task Force recommendation statement. *J Am Med Assoc*. 2018;320:1163–1171.
111. Heymsfield SB, Wadden TA. Mechanisms, pathophysiology, and management of obesity. *N Engl J Med*. 2017;376:254–266.
112. Bray GA, Fruhbeck G, Ryan DH, et al. Management of obesity. *Lancet*. 2016;387:1947–1956.
113. Bettini S, Belligoli A, Fabris R, et al. Diet approach before and after bariatric surgery. *Rev Endocr Metab Disord*. 2020;21:297–306.
114. Di Rosa C, Lattanzi G, Taylor SF, et al. Very low calorie ketogenic diets in overweight and obesity treatment: effects on anthropometric parameters, body composition, satiety, lipid profile and microbiota. *Obes Res Clin Pract*. 2020;14:491–503.
115. Hoddy KK, Marlatt KL, Cetinkaya H, et al. Intermittent fasting and metabolic health: from religious fast to time-restricted feeding. *Obesity (Silver Spring)*. 2020;28:S29–S37.
116. Chin SH, Kahathuduwa CN, Binks M. Physical activity and obesity: what we know and what we need to know. *Obes Rev*. 2016;17:1226–1244.
117. Foright RM, Presby DM, Sherk VD, et al. Is regular exercise an effective strategy for weight loss maintenance? *Physiol Behav*. 2018;188:86–93.
118. U.S. Department of Health and Human Services and U.S. Department of Agriculture. *2015-2020 Guidelines for Americans*. 8th ed. 2015. http://health.gov/dietaryguidelines/2015/guidelines/.
119. Zouhal H, Ben Abderrahman A, Khodamoradi A, et al. Effects of physical training on anthropometrics, physical and physiological capacities in individuals with obesity: a systematic review. *Obes Rev*. 2020;21:e13039.
120. Katzmarzyk PT, Powell KE, Jakicic JM, et al. Sedentary behavior and health: update from the 2018 physical activity guidelines advisory committee. *Med Sci Sports Exerc*. 2019;51:1227–1241.
121. Gadde KM, Apolzan JW, Berthoud HR. Pharmacotherapy for patients with obesity. *Clin Chem*. 2018;64:118–129.
122. Sharretts J, Galescu O, Gomatam S, et al. Cancer risk associated with lorcaserin - the FDA's review of the CAMELLIA-TIMI 61 trial. *N Engl J Med*. 2020;383:1000–1002.
123. Sjostrom L, Rissanen A, Andersen T, et al. Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. *Lancet*. 1998;352:167–172.
124. Smith SM, Meyer M, Trinkley KE. Phentermine/topiramate for the treatment of obesity. *Ann Pharmacother*. 2013;47:340–349.
125. FDA Medication Guide. https://www.fda.gov/media/83651/download.

126. Billes SK, Sinnayah P, Cowley MA. Naltrexone/bupropion for obesity: an investigational combination pharmacotherapy for weight loss. *Pharmacol Res*. 2014;84:1–11.
127. Sherman MM, Ungureanu S, Rey JA. Naltrexone/bupropion ER (Contrave): newly approved treatment option for chronic weight management in obese adults. *P T*. 2016;41:164–172.
128. Muller TD, Finan B, Bloom SR, et al. Glucagon-like peptide 1 (GLP-1). *Mol Metab*. 2019;30:72–130.
129. Vorsanger MH, Subramanyam P, Weintraub HS, et al. Cardiovascular effects of the new weight loss agents. *J Am Coll Cardiol*. 2016;68:849–859.
130. Greenway FL, Aronne LJ, Raben A, et al. A randomized, double-blind, placebo-controlled study of Gelesis100: a novel nonsystemic oral hydrogel for weight loss. *Obesity*. 2019;27:205–2216.
131. Igel LI, Sinha A, Saunders KH, et al. Metformin: an old therapy that deserves a new indication for the treatment of obesity. *Curr Atheroscler Rep*. 2016;18:16.
132. Coll AP, Chen M, Taskar P, et al. GDF15 mediates the effects of metformin on body weight and energy balance. *Nature*. 2020;578:444–448.
133. Boyle CN, Lutz TA, Le Foll C. Amylin—its role in the homeostatic and hedonic control of eating and recent developments of amylin analogs to treat obesity. *Mol Metab*. 2018;8:203–210.
134. Brown E, Wilding JPH, Barber TM, et al. Weight loss variability with SGLT2 inhibitors and GLP-1 receptor agonists in type 2 diabetes mellitus and obesity: Mechanistic possibilities. *Obes Rev*. 2019;20:816–828.
135. Arterburn DE, Telem DA, Kushner RF, et al. Benefits and risks of bariatric surgery in adults: a review. *J Am Med Assoc*. 2020;324:879–887.
136. NIH conference. Gastrointestinal surgery for severe obesity. Consensus Development Conference Panel. *Ann Intern Med*. 1991;115:956–961.

21

Genetic Syndromes Associated With Obesity

Stephen O'Rahilly and I. Sadaf Farooqi

OUTLINE

Introduction, 334
Monogenic Obesity Syndromes, 334
  Congenital Leptin and Leptin Receptor Deficiency, 334
  Proopiomelanocortin and PCSK1 Deficiency, 335
  MC4R Deficiency, 336
  SIM1 Deficiency, 337
  BDNF, TRKB, and SH2B1 Deficiency, 337
Pleiotropic Obesity Syndromes, 337
  Prader–Willi Syndrome, 337
  Albright's Hereditary Osteodystrophy, 338
  Bardet–Biedl Syndrome, 339
  Alstrom Syndrome, 339
  Cohen Syndrome, 339
Rare Variants Associated with Obesity, 339
Conclusions, 340

INTRODUCTION

While hyperphagia (increased food intake) and weight gain of short duration should prompt investigation for space-occupying lesions of the hypothalamus, sustained hyperphagia and obesity from early childhood are common features of the genetic obesity syndromes. Some disorders are associated with learning difficulties, developmental delay, and other overt clinical problems that bring children to medical attention at a young age. However, over the last 20 years there has been increasing recognition that there are a group of genetic disorders where severe obesity alone is the presenting feature. The molecular and clinical characterization of these monogenic obesity syndromes has informed our understanding of the function of the leptin-melanocortin pathway in the hypothalamus that regulates appetite and body weight, and as a result some of these disorders can now be treated.

Leptin was first discovered by studying severely obese *ob/ob* mice that were found to be lacking a hormone produced by adipose tissue.[1] Administration of leptin to these mice reduced their excessive food intake and corrected neuroendocrine and metabolic abnormalities, proving that leptin is a pivotal regulator of energy homeostasis.[2-4] Subsequent work by Ahima and Flier[5] showed that leptin's physiological role is to signal nutritional depletion, such that fasting or weight loss result in a fall in leptin levels, which then trigger an increase in energy intake and reduced energy expenditure to maintain energy homeostasis.

The physiological effects of leptin are mediated through the central nervous system, particularly the hypothalamus, where the long signaling isoform of the leptin receptor is highly expressed.[6] Leptin binding to its receptor phosphorylates the transcription factor STAT3, which dimerizes and translocates to the nucleus, where it stimulates the expression of proopiomelanocortin (POMC) (Fig. 21.1). POMC is posttranslationally processed to yield the melanocortin peptides (alpha, beta, and gamma melanocyte-stimulating hormone [MSH]), which are agonists with varying affinities for the melanocortin 3 receptor (MC3R) and melanocortin 4 receptor (MC4R) expressed on second-order neurons.[7] In the fed state, leptin stimulates the expression of POMC and the firing of POMC neurons; POMC-derived peptides activate MC4R to decrease food intake. Perhaps even more importantly, leptin suppresses the activity of neurons expressing AGRP, a peptide that antagonizes signaling at MC3R and MC4R. In addition to leptin, the POMC and AGRP neurons also respond to other signals of nutritional status such as insulin, ghrelin, and certain amino acids. Genetic disorders that directly or indirectly disrupt this pathway can cause hyperphagia and weight gain (Table 21.1). As these neurons project to and receive inputs from other brain regions, disruption of the leptin-melanocortin pathway can affect a range of behaviors, neuroendocrine function, and autonomic function, and thereby blood pressure and insulin secretion.

MONOGENIC OBESITY SYNDROMES

Congenital Leptin and Leptin Receptor Deficiency

Homozygous (and compound heterozygous) loss-of-function mutations in the genes encoding leptin (*LEP*) and the leptin receptor (*LEPR*) cause rapid weight gain in the first few months of life, resulting in severe obesity.[8,9] Children develop intense hyperphagia with food-seeking behavior and find all types of food very rewarding.[10] Although no changes in resting metabolic rate or total energy expenditure are detectable, abnormalities in sympathetic nerve function have been reported.[11] Leptin and leptin receptor deficiency are associated with hypothalamic hypothyroidism and hypogonadotropic hypogonadism.[12,13] However, some evidence for the delayed but spontaneous onset of menses has been reported in adults.[14] Children have impaired T-cell number and function, consistent with high rates of childhood infection and a high rate of childhood mortality.[13,15]

Congenital leptin deficiency can be diagnosed on the basis of undetectable serum leptin concentrations; a small number of mutations that result in detectable levels of leptin that is bioinactive have been reported.[16] Serum leptin levels are not elevated disproportionately in *LEPR* deficiency, although particular mutations located near

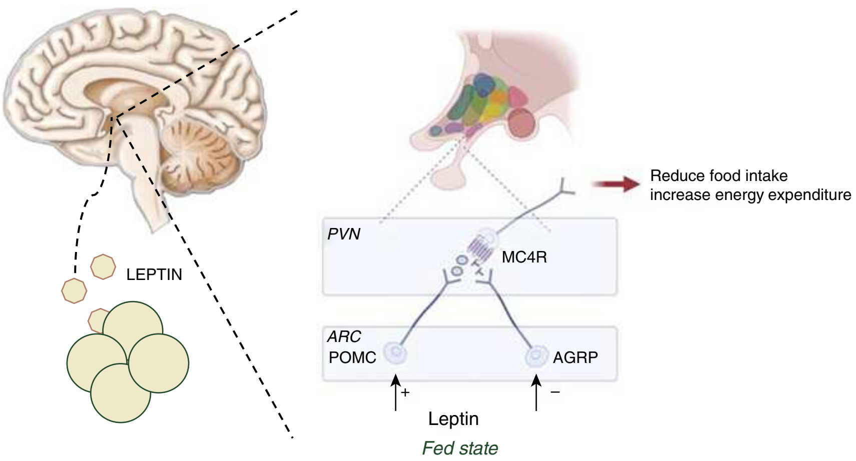


Fig. 21.1 Leptin regulates body weight through the melanocortin pathway. Leptin is secreted by adipocytes and acts on neurons in the hypothalamus. In the fed state, leptin stimulates primary neurons in the arcuate nucleus (ARC) of the hypothalamus that express proopiomelanocortin (POMC). POMC-derived peptides act on the melanocortin 4 receptor (MC4R) expressed on neurons in the paraventricular nucleus (PVN) to reduce energy intake. At the same time, leptin inhibits neurons expressing Agouti-related peptide (AGRP), which antagonizes melanocortin receptors.

TABLE 21.1 Monogenic Obesity Syndromes

Gene Affected	Inheritance	Additional Clinical Features
Leptin	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Leptin receptor	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Proopiomelanocortin	Autosomal Recessive	Hyperphagia, cholestatic jaundice or adrenal crisis due to adrenocorticotropic hormone deficiency, pale skin and red hair
Prohormone convertase 1	Autosomal recessive	Small bowel enteropathy, postprandial hypoglycemia, hypothyroidism, adrenocorticotropic hormone deficiency, hypogonadism, central diabetes insipidus
Carboxypeptidase E	Autosomal recessive	
Melanocortin 4 receptor	Autosomal dominant	Hyperphagia, accelerated linear growth
Single-minded 1	Autosomal dominant	Hyperphagia, accelerated linear growth, speech and language delay, autistic traits
Brain-derived neurotropic factor	Autosomal dominant	Hyperphagia, developmental delay, hyperactivity, behavioral problems including aggression
TRKB	Autosomal dominant	Hyperphagia, speech and language delay, variable developmental delay, hyperactivity, behavioral problems including aggression
SH2B1	Autosomal dominant	Hyperphagia, disproportionate hyperinsulinemia, early type 2 diabetes mellitus, behavioral problems including aggression

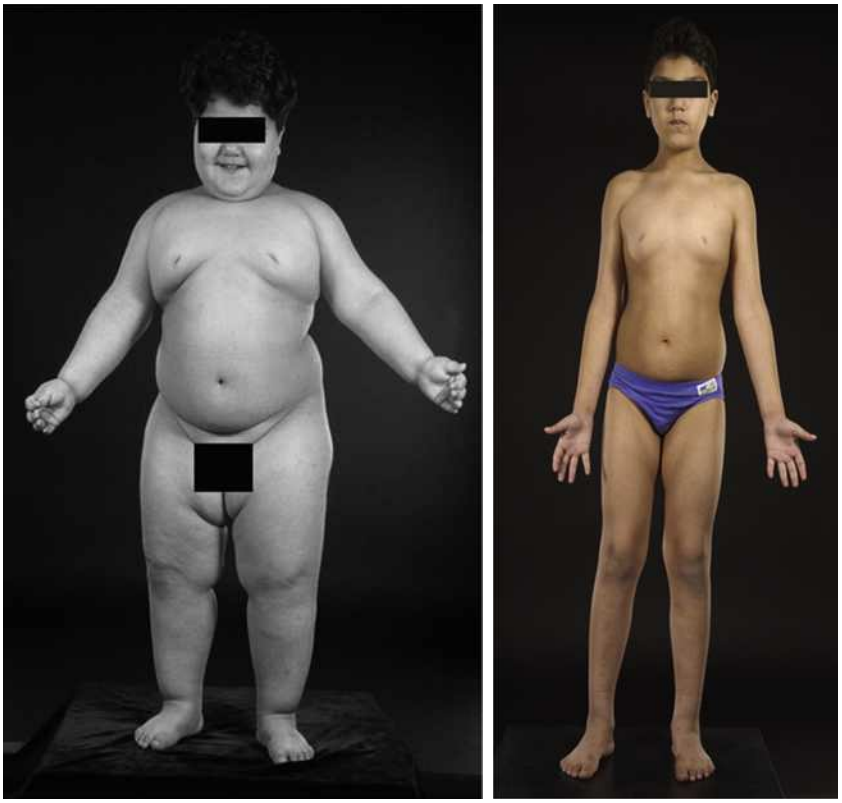
the transmembrane domain can result in a truncated extracellular domain that binds leptin, resulting in abnormally elevated leptin levels.[17] Sequencing of *LEP* and *LEPR* is therefore indicated in very severe obesity that begins in the first year of life, even when leptin levels are normal.

Treatment of leptin deficiency with recombinant leptin results in a dramatic improvement in hyperphagia, weight loss, and correction of thyroid and T-cell dysfunction[15,8] (Fig. 21.2). Leptin treatment permits progression of appropriately timed pubertal development, suggesting that leptin is a permissive factor for the development of puberty.[13,15] Leptin may exert these effects on the reproductive system through a number of molecules, including kisspeptin, which signals through GPR54 to modify the release of gonadotropin-releasing hormone. Linear growth is normal in childhood; however, due to the absence of a pubertal growth spurt, final height appears to be reduced if patients are not treated.[14] Until recently, LEPR deficiency was not treatable. However, trials of the MC4R agonist setmelanotide have shown promising results, with suppression of hunger and weight loss[19]; this medication has been licensed for the treatment of LEPR deficiency in the US.

Proopiomelanocortin and PCSK1 Deficiency

The POMC precursor peptide is posttranslationally processed into multiple peptides, in a tissue-specific manner. In the pituitary gland, POMC is cleaved to yield adrenocorticotropic hormone (ACTH). In the skin, the POMC-derived melanocortin peptides α- and β-MSH regulate pigmentation by signaling through the melanocortin 1 receptor (MC1R), and in the brain, α- and β-MSH activate signaling via MC4R to reduce food intake (Fig. 21.3). As a result, homozygous or compound heterozygous mutations in *POMC* lead to hyperphagia and early-onset obesity, isolated ACTH deficiency (presenting as neonatal hypoglycemia), and hypopigmentation of skin and hair, which is more noticeable in people of European ancestry.[20,21] Heterozygous null mutations in *POMC* and loss-of-function mutations in the part of the *POMC* gene that encodes α- and β-MSH significantly increase obesity risk but are not invariably associated with obesity.[22,23]

The POMC propeptide is processed into its active melanocortin products by a series of enzymes that are differentially expressed in the hypothalamus and the pituitary. Prohormone convertase 1 (PCSK1) is necessary for the production of ACTH and is highly expressed in the pituitary. The hypothalamus, but not the pituitary, expresses a related



3yr old weighing 42kg

7yr old weighing 32kg

Fig. 21.2 Treatment of leptin deficiency with recombinant leptin. Child with congenital leptin deficiency before *(left)* and after *(right)* treatment with recombinant leptin.

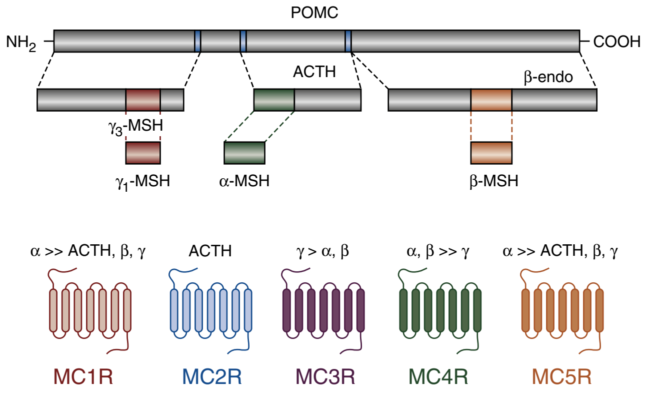


Fig. 21.3 Proopiomelanocortin (POMC)-derived peptides and melanocortin receptors. POMC is processed into adrenocorticotropic hormone (ACTH) and α, β, and γ-melanocyte stimulating hormone (MSH) peptides, which act on the melanocortin 1–5 receptors (MC1R–MC5R).

enzyme PCSK2, which acts on the products of the prior PCSK1 cleavage to produce the smaller melanocortin peptides. A further enzyme, carboxpeptidase E, is needed to produce the final mature peptides. Impaired processing of POMC contributes to the hyperphagic severe early-onset obesity and ACTH deficiency in people lacking PCSK1.[24] Hypogonadotropic hypogonadism due to the impaired processing of prepro–gonadotropin-releasing hormone and postprandial hypoglycemia due to impaired processing of proinsulin to insulin are additional clinical features. Where these assays are available, the finding of high proinsulin levels in the context of a low plasma insulin can point to this diagnosis. Impaired processing of gut-derived peptides in the enteroendocrine cells that express PCSK1 throughout the gut[25] may contribute to neonatal enteropathy; indeed, patients may first present to gastroenterologists with these symptoms. A single family with a homozygous mutation in carboxypeptidase E with overlapping clinical features has also been reported.[26] Setmelanotide has been shown to be effective in POMC and PCSK1 deficiency[27,28] and has been licensed for chronic weight management in these disorders in the United States.

MC4R Deficiency

POMC-derived peptides act as agonists at the MC4R. Targeted disruption of *Mc4r* in mice leads to increased food intake, weight gain, increased lean mass, and linear growth.[29] Heterozygous loss-of-function mutations in *MC4R* are found in 5% to 6% of patients with severe early-onset obesity[30,31] and at a frequency of approximately 1/330 in the general UK population, making this the commonest gene in which variants contribute to obesity. Most naturally occurring disease-causing MC4R mutations disrupt the expression and trafficking of the receptor to the cell surface.[32] The mechanism of G protein–coupled receptor (GPCR) dysfunction has potential interest, as pharmacological chaperones can increase the cell surface expression and signaling of mutant GPCRs, which represents a rational therapeutic approach for this condition.[33] Additionally, mutations can affect the production of cyclic adenosine monophosphate, the homodimerization of MC4R, the endocytosis of MC4R, and the recruitment of β-arrestins.[34]

MC4R mutations are inherited in a codominant manner, with variable penetrance and expression in heterozygous carriers[30]; less commonly, homozygous and double heterozygous (different mutations on both alleles) mutations have been reported in patients with severe obesity. The features of MC4R deficiency include hyperphagia, hyperinsulinaemia, increased lean mass, and increased linear growth.[30,35]

Complete loss-of-function mutations have a larger impact on phenotype than partial loss-of-function mutations.[36]

MC4R$^{-/-}$ mice maintain a normal blood pressure despite severe obesity.[37] Similarly, loss-of-function MC4R mutations in humans are associated with a reduced prevalence of hypertension, low systolic blood pressure, lower urinary noradrenaline excretion, and reduced peripheral nerve sympathetic activation compared with similarly obese people without MC4R mutations.[38,39] As MC4R neurons synapse with preganglionic sympathetic neurons, which regulate vascular tone, increased signaling through MC4R can lead to an increase in blood pressure, as seen with first-generation melanocortin receptor agonists given to obese volunteers. Setmelanotide, a second-generation MC4R agonist that causes weight loss without increasing blood pressure, has been shown to be very effective in POMC and PCSK1 deficiency and is currently being trialed in heterozygous MC4R deficiency. Liraglutide, the GLP-1 receptor agonist, is effective in some patients with MC4R deficiency,[40] and some patients with heterozygous (but not homozygous) mutations benefit from Roux-en-Y-bypass surgery.[41]

SIM1 Deficiency

SIM1 is a transcription factor that plays a key role in the development and function of the paraventricular nucleus of the hypothalamus, where MC4R is highly expressed. Dominantly inherited or *de novo* deletions of chromosome 6q14-21 that encompass the *SIM1* gene, as well as heterozygous loss-of-function *SIM1* mutations,[42,43] cause hyperphagia, severe obesity, and relatively low systolic blood pressure, features that closely overlap with those seen in MC4R deficiency. Many SIM1 mutation carriers have speech and language delay in childhood and exhibit neurobehavioral abnormalities, including autistic-type behaviors. These features show some overlap with the behavioral phenotypes seen in Prader–Willi syndrome (PWS). Expression of the neuropeptide oxytocin is reduced in mouse models of Sim1 deficiency, and oxytocin administration reduces food intake in *Sim1*-haploinsufficient animals.[44] As oxytocin is involved in energy homeostasis, emotional responses, and social interaction, reduced levels of oxytocin may contribute to the behavioral phenotype seen in SIM1 deficiency. A reduced number of oxytocin neurons and reduced oxytocin mRNA and protein levels have also been reported in PWS.[45,46] For this reason, some reports have referred to SIM1 deficiency as being PWS-like. A small number of mutations in OTP, another transcription factor involved in paraventricular nucleus development, have been reported in patients with obesity and autistic behaviors.[47]

BDNF, TRKB, and SH2B1 Deficiency

Chromosome 11p.12 deletions encompassing the brain-derived neurotrophic factor (*BDNF*) gene[48] or missense mutations that disrupt *BDNF* or its tyrosine kinase receptor, tropomycin-related kinase B (TrkB), are associated with hyperphagia and obesity.[49] Mutations often arise *de novo,* and as such should be considered where both parents are of normal weight and IQ. The neurotrophin BDNF is widely expressed in the brain and signals via TrkB to regulate neuronal differentiation and survival, synapse formation, and activity-dependent changes in synapse structure and function. BDNF and TrkB are expressed in areas of the brain involved in learning and memory, including the hippocampus. Obesity-associated mutations that impair TrkB signaling affect dendritic spine structure and function, which forms the neural substrate for learning and memory in hippocampal neurons.[49] In addition to severe obesity, carriers of these functional TrkB mutations have learning difficulties, impaired short-term memory, hyperactivity, repetitive behaviors often considered to be autistic-like, fearlessness, and in some cases aggression. Some of these neurobehavioral phenotypes, as well as increased locomotor activity, have been observed in animal models of *Bdnf/Trkb* disruption.[50]

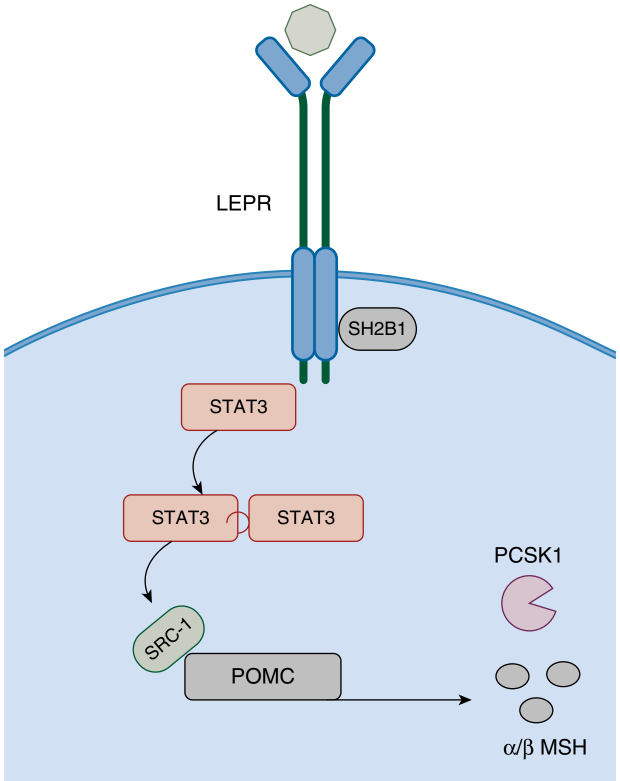


Fig. 21.4 Genes that modulate leptin signaling to cause obesity. Leptin signaling through its receptor (LEPR) is enhanced by SH2B1 and results in the activation of STAT3, which dimerizes and translocates into the nucleus to activate proopiomelanocortin (*POMC*) transcription. SRC-1 modulates *POMC* transcription by interacting with activated STAT3. POMC is processed by PCSK1 to yield the melanocortin peptides (α and β-MSH).

Src-homology-2 (SH2) B-adaptor protein-1 (SH2B1) is an intracellular adaptor protein that mediates signaling through a number of receptor tyrosine kinases (including TrkB) and cytokine receptors (including the leptin receptor) (Fig. 21.4); mutations impair signaling and neurite differentiation. Chromosomal deletions on 16p11.2 that include SH2B1[51] and heterozygous mutations in the gene itself[52] are associated with hyperphagia, dominantly inherited severe early-onset obesity, and disproportionate insulin resistance. Male mutation carriers often have behavioral problems, including social isolation and aggressive behavior from childhood.[52,53] Intriguingly, mice with brain-specific deletion of *Sh2b1* gain weight and develop reactive aggression[54]; brain-specific restoration of Sh2b1 completely reverses intermale aggression. Neural SH2B1 promotes brain development and growth, at least in part by enhancing the actions of BDNF and related neurotrophic factors.

PLEIOTROPIC OBESITY SYNDROMES

In a number of syndromes, obesity is one of a number of features (pleiotropic) alongside developmental/learning difficulties and dysmorphic features (Table 21.2). A number of these disorders arise from chromosomal abnormalities.

Prader–Willi Syndrome

PWS is relatively common, with an estimated prevalence of 1 in 25,000 births,[3] and is characterized by diminished fetal activity, hypotonia,

TABLE 21.2 Pleiotropic Genetic Obesity Syndromes

Syndrome	Inheritance	Additional Clinical Features
Prader–Willi	Autosomal dominant	Hypotonia, failure to thrive in infancy, developmental delay, short stature, hypogonadotropic hypogonadism, sleep disturbance, obsessive behavior
Albright's hereditary osteodystrophy	Autosomal dominant	Short stature in some, skeletal defects, developmental delay, shortened metacarpals; hormone resistance when mutation is on maternally inherited allele
Bardet–Beidl	Autosomal recessive	Syndactyly/brachydactyly/polydactyly, developmental delay, retinal dystrophy or pigmentary retinopathy, hypogonadism, renal abnormalities
Cohen	Autosomal recessive	Facial dysmorphism, microcephaly, hypotonia, developmental delay, retinopathy
Carpenter	Autosomal recessive	Acrocephaly, brachydactyly, developmental delay, congenital heart defects; growth retardation, hypogonadism
Alstrom	Autosomal recessive	Progressive cone-rod dystrophy, sensorineural hearing loss, hyperinsulinemia, early type 2 diabetes mellitus, dilated cardiomyopathy, pulmonary, hepatic and renal fibrosis
Tubby	Autosomal recessive	Progressive cone-rod dystrophy, hearing loss

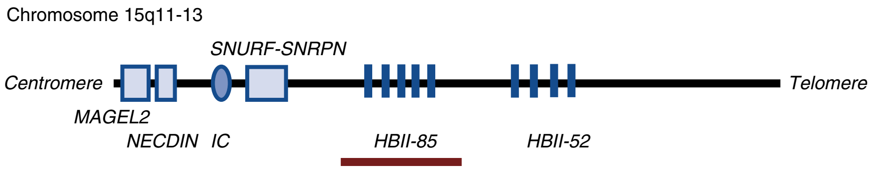


Fig. 21.5 **Chromosomal deletion in Prader–Willi syndrome (PWS).** Paternal chromosome 15q11-13 deletions cause PWS. A number of genes (*MAGEL2*, *NECDIN*), small nuclear ribonucleoprotein polypeptide N (*SNRPN*), and the imprinting center (IC) are deleted in most patients, but the minimal critical region involved in the phenotype involves the HBII-85 snoRNAs (noncoding small nucleolar RNAs).

developmental delay, short stature, hypogonadotropic hypogonadism, and obesity. PWS is caused by deletion of an imprinted region on the paternal chromosome 15q11-q13. Loss of the entire paternal chromosome 15 with the presence of two maternal homologs (uniparental maternal disomy) occurs in approximately 20% of patients. While a number of genes within the imprinted region have been studied, small deletions restricted to a family of noncoding RNAs (HBII-85 snoRNAs) recapitulate the classical features,[55,56] providing strong evidence that these noncoding RNAs play a causal role in PWS (Fig. 21.5). These noncoding RNAs affect the alternative splicing of a large number of target genes, with evidence of intron retention,[46] as is also seen in a number of neurodegenerative diseases. RNA sequencing of brain tissue from PWS patients has shown that expression of BDNF and its receptor TrkB is reduced, and downregulated genes affect neuronal differentiation, maintenance, and synaptic plasticity.[46] Neuroimaging studies in PWS have identified reduced gray matter volume in cortical areas and abnormal gyrification.[57] As such, it is likely that the complex clinical phenotype arises from the reduced expression of large number of genes in the hypothalamus and other brain regions.

Patients often have low birth weight at term (in keeping with other imprinted disorders) and poor feeding as neonates due to diminished swallowing and sucking reflexes, which may require assisted feeding for 3 to 4 months. Feeding difficulties generally improve by the age of 6 months. From 12 to 18 months, hyperphagia is a dominant feature in children with PWS, often associated with pica behavior. One suggested mediator of the obesity phenotype in PWS patients is the enteric hormone ghrelin, which is implicated in the regulation of mealtime hunger and also stimulates growth hormone (GH) secretion via the GH secretagogue receptor. Fasting plasma ghrelin levels are 4.5-fold higher in PWS subjects than in equally obese controls and patients with other obesity syndromes[58]; whether this finding is relevant in the pathogenesis of hyperphagia in these patients remain unclear. The mainstay of management has centered on low-calorie diet with regular exercise, rigorous supervision, restriction of food, and behavioral counseling for the patient and family, often in the context of group homes for PWS adolescents and adults.

Children with PWS have reduced linear growth, reduced muscle mass, and increased fat mass, body composition abnormalities that are consistent with GH deficiency. Treatment with GH significantly improves growth and final height, and GH is licensed for use in PWS. GH treatment also decreases body fat and increases muscle mass.[59] Gonadal maturation is delayed or incomplete as the result of hypogonadotropic hypogonadism.

Albright's Hereditary Osteodystrophy

Albright's hereditary osteodystrophy (AHO) is an autosomal dominant disorder caused by heterozygous loss-of-function mutations in *GNAS*, the gene encoding Gα_s (stimulatory G-protein alpha subunit), which mediates GPCR signaling. Patients present with developmental delay, brachydactyly (shortened metacarpals and/or metatarsals), subcutaneous ossifications, childhood-onset obesity with short adult stature (height below the 3rd percentile for chronological age), and in some cases hormone resistance syndromes.[60-62] Imprinting at the *GNAS* locus results in selective silencing of the paternally-inherited *GNAS* allele in some tissues. Patients with mutations on maternally-inherited alleles, which are preferentially expressed in the thyroid, pituitary, and renal proximal tubule, develop resistance to parathyroid hormone and other hormones that signal through Gα_s-coupled receptors (pseudohypoparathyroidism type 1A),[61] while patients with mutations on paternally-inherited alleles have AHO without hormone resistance (pseudopseudohypoparathyroidism [PPHP]).

Obesity is a well-recognized feature of PHP when mutations occur on the maternally inherited *GNAS* allele. Studies in patients presenting with severe childhood-onset obesity have revealed an unexpectedly high prevalence of *GNAS* mutations, indicating that screening for mutations in *GNAS* should be incorporated into the diagnostic workup of severe childhood-onset obesity.[63] Early diagnosis guides monitoring for hormone resistance syndromes, recognition of hypocalcemia as a cause of seizures, and treatment with thyroxine and with recombinant human GH before fusion of the growth plate.[62,64]

Brain-specific deletion of the maternal Gα_s allele impaired the ability of an MC4R agonist to reduce body weight, suggesting a role for centrally expressed MC4Rs in mediating obesity in *GNAS* deficiency.[65] Recent studies have shown that obesity-associated missense mutations in *GNAS* impair MC4R signaling.

Bardet–Biedl Syndrome

Bardet–Biedl syndrome (BBS) is an autosomal recessive syndrome characterized by obesity, developmental delay, syndactyly, brachydactyly or polydactyly, retinal dystrophy or pigmentary retinopathy, hypogonadism, and renal abnormalities. BBS is genetically very heterogeneous, with mutations in at least 20 genes reported to date and evidence of triallelic inheritance in some families (two mutations in one BBS gene plus an additional mutation in a second, unlinked BBS gene).[66] Most of the genes involved in BBS affect the structure and/or function of the basal body, a modified centriole that is essential for the function of nonmotile cilia.[67] There is some evidence that BBS proteins affect leptin signaling, and as such the MC4R agonist setmelanotide (which targets the MC4R pathway downstream of leptin) has been trialed in these patients, with some success.[68]

Alstrom Syndrome

Alstrom syndrome shows some overlap with BBS (retinitis pigmentosa, deafness, obesity, diabetes mellitus with recessive inheritance); however, classically, developmental delay, polydactyly, and hypogonadism are not features. Patients present with a severe visual defect in infancy, photophobia, nystagmus, and loss of central vision. Some affected individuals present with dilated cardiomyopathy (often diagnosed in infancy), hepatic dysfunction, renal dysfunction, hypothyroidism, male hypogonadism, short stature, and mild to moderate developmental delay.[69]

Although obesity is common in this syndrome, it is rarely severe. In contrast, insulin resistance due to impaired adipose tissue expandability[70] is frequently severe, and once diabetes develops, it may be very difficult to control. Hypertriglyceridemia may be severe and may result in acute pancreatitis. Mutations in *ALMS1*, which is localized to the centrosome and the ciliary basal bodies *in vitro*, suggest a role in the structure of the basal body or in the transport of proteins between the cytoplasm and the ciliary axoneme.

Cohen Syndrome

Cohen syndrome is an autosomal recessive disorder characterized by nonprogressive mild to severe psychomotor retardation, motor clumsiness, microcephaly, characteristic facial features, childhood hypotonia, intermittent isolated neutropenia, and a cheerful disposition. Progressive, often high-grade myopia and retinochoroidal dystrophy resembling retinitis pigmentosa are essential features in Cohen syndrome.[71] Vision starts to deteriorate early but generally is preserved until adulthood, and by the age of 40 years many patients are severely visually impaired. Mutations in *COH1* found in some Finnish patients with Cohen syndrome suggest a role in intracellular vesicle–mediated sorting and transport, although the link to obesity has not been established.

KEY POINTS

- Monogenic obesity syndromes are characterized by hyperphagia (increased drive to eat) and weight gain that begins before the age of 5 years. Disorders affect hypothalamic pathways involved in the regulation of hormone secretion, and as such patients sometimes have endocrine features such as hypogonadism. Genetic obesity syndromes that affect neural development can be associated with developmental delay and behavioral difficulties.

RARE VARIANTS ASSOCIATED WITH OBESITY

In addition to these dominant or recessively inherited monogenic obesity syndromes, there is increasing recognition that variants in a broader set of genes contribute to severe obesity. As these variants are rare, statistical evidence of enrichment in obese cases versus controls is often not achieved (power calculations predict that studies involving hundreds of thousands of patients would be needed[72]). While patients in whom these variants are described often share a set of clinical features, severe obesity is often not inherited in a classical Mendelian manner, and there can be incomplete penetrance and variable expressivity within and across families. Nonetheless, where variants cause a loss of function in cells, affect a mechanism known to regulate weight, and/or cause obesity when modeled in mice, there is sufficient evidence to indicate that rare obesity-associated variants contribute to the missing heritability in obesity, as seen for other complex traits.[73] Where the molecules encoded by these genes converge on pathways known to be involved in weight regulation, there is sometimes sufficient evidence to treat subsets of patients.

The semaphorin 3s are a large family of secreted proteins that have both repulsive and attractive effects on axons during development.[74] Recent evidence in mice has shown that SEMA3A guides vomeronasal neurons, which migrate into the hypothalamus to develop into gonadotropin-releasing hormone–expressing cells, and loss-of-function variants contribute to hypogonadotropic hypogonadism in a non-Mendelian manner. Semaphorin 3s acting via the neuropilin-2 receptor direct the development of the melanocortin circuit formed by POMC projections extending from the arcuate to the paraventricular nucleus of the hypothalamus.[75] Rare loss-of-function variants in semaphorin 3 ligands, receptors, and coreceptors are enriched in severely obese individuals compared with controls.[75] Some individuals had learning difficulties, behavioral abnormalities, and neurological disorders including epilepsy, as well as medication-resistant constipation in childhood. SEMA3 signaling is known to regulate the development of the enteric nervous system, and rare heterozygous loss-of-function variants in *SEMA3C* and *SEMA3D* have been associated with Hirschsprung disease,[76] a disorder characterized by failure of development of parasympathetic ganglion cells in the large intestine.

In the arcuate nucleus of the hypothalamus, leptin signaling leads to the phosphorylation of STAT3, which interacts with the transcription factor steroid receptor coactivator (SRC)-1 to modulate *POMC* transcription (Fig. 21.4). Disruption of the SRC-1-pSTAT3 interaction causes obesity in mice, and rare heterozygous variants in *SRC-1* identified in severely obese people decrease leptin-pSTAT3 mediated signaling and POMC expression in cells.[77] A mouse model of a human loss-of-function SRC-1 variant gains weight on a high-fat diet. As a result of this evidence, trials are ongoing to investigate whether patients with SRC-1 deficiency may lose weight with setmelanotide.

Nuclear pleckstrin homology domain–interacting protein (PHIP) directly affects the transcription of POMC, and obesity-associated PHIP mutants have been shown to decrease *POMC* transcription.[78] Heterozygous *PHIP* deletions and frameshift mutations have been reported in patients with developmental delay, intellectual disability, dysmorphic features, and in some cases obesity.[79] The absence of severe obesity in some family members carrying *PHIP* loss-of-function variants, as well as the presence of missense variants in control participants without obesity, suggests variable penetrance. In keeping with the growth phenotype seen in null mice, some, but not all, *PHIP* variant carriers exhibit low birth weight, reduced linear growth in childhood, hyperphagia, insulin resistance, and early type 2 diabetes.

Rare variants in the gene encoding MRAP2, an accessory protein that interacts with MC4R and leads to obesity when disrupted in mice, have been associated with severe obesity, although the detailed molecular mechanisms underlying this association are not known.[80]

To date, most of the genetic obesity syndromes are characterized by hyperphagia as the major driver of the obesity. Loss-of-function mutations in kinase suppressor of Ras2 (*KSR2*) increase food intake but are predominantly associated with reduced basal metabolic rate in the presence of normal thyroid function.[81] Some *KSR2* mutation carriers experience marked weight loss in childhood when prescribed the antidiabetic drug metformin (for severe insulin resistance). Further work will be needed to see if these observations can be replicated and to investigate the cellular mechanisms underlying these effects.

KEY POINTS

- Rare variants that disrupt the function of key genes involved in weight regulation can contribute to severe obesity. Patterns of inheritance may not be clearly Mendelian (e.g., dominant or recessive), with variable penetrance within and between families. Nonetheless, some of these disorders are treatable.

CONCLUSIONS

Cumulatively, up to 20% of children with severe obesity have rare chromosomal abnormalities and/or highly penetrant genetic mutations that drive their obesity.[78] This figure is likely to increase with wider accessibility to genetic testing and as new genes are identified from exome and genome sequencing. A genetic diagnosis can inform management (many such patients are relatively refractory to weight loss through changes in diet and exercise) and can inform clinical decision-making regarding the use of bariatric surgery (feasible in some; high-risk in others). Importantly, some genetic obesity syndromes are treatable. Recombinant leptin is dramatically effective at treating the severe obesity of congenital leptin deficiency. Setmelanotide, an MC4R agonist, has been used effectively in phase II/III clinical trials of POMC and LEPR deficiencies and is being explored for the treatment of other genetic obesity syndromes affecting the melanocortin pathway. Ultimately, understanding how these pathways are disrupted in people with weight problems may inform strategies to target these pathways for prevention and treatment.

REFERENCES

1. Zhang Y, Proenca R, Maffei M, et al. Positional cloning of the mouse obese gene and its human homologue. *Nature*. 1994;372:425–432.
2. Halaas JL, Gajiwala KS, Maffei M, et al. Weight-reducing effects of the plasma protein encoded by the obese gene. *Science*. 1995;269:543–546.
3. Campfield LA, Smith FJ, Guisez Y, et al. Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science*. 1995;269:546–549.
4. Pelleymounter MA, Cullen MJ, Baker MB, et al. Effects of the obese gene product on body weight regulation in ob/ob mice. *Science*. 1995;269:540–543.
5. Ahima RS, Prabakaran D, Mantzoros C, et al. Role of leptin in the neuroendocrine response to fasting. *Nature*. 1996;382:250–252.
6. Elmquist JK, Bjorbaek C, Ahima RS, et al. Distributions of leptin receptor mRNA isoforms in the rat brain. *J Comp Neurol*. 1998;395:535–547.
7. Gautron L, Elmquist JK, Williams KW. Neural control of energy balance: translating circuits to therapies. *Cell*. 2015;161:133–145.
8. Montague CT, Farooqi IS, Whitehead JP, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*. 1997;387:903–908.
9. Clement K, Vaisse C, Lahlou N, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*. 1998;392:398–401.
10. Farooqi IS, Bullmore E, Keogh J, et al. Leptin regulates striatal regions and human eating behavior. *Science*. 2007;317:1355.
11. Ozata M, Ozdemir IC, Licinio J. Human leptin deficiency caused by a missense mutation: multiple endocrine defects, decreased sympathetic tone, and immune system dysfunction indicate new targets for leptin action, greater central than peripheral resistance to the effects of leptin, and spontaneous correction of leptin-mediated defects. *J Clin Endocrinol Metab*. 1999;84:3686–3695.
12. Strobel A, Issad T, Camoin L, et al. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat Genet*. 1998;18:213–215.
13. Licinio J, Caglayan S, Ozata M, et al. Phenotypic effects of leptin replacement on morbid obesity, diabetes mellitus, hypogonadism, and behavior in leptin-deficient adults. *Proc Natl Acad Sci U S A*. 2004;101:4531–4536.
14. Farooqi IS, Wangensteen T, Collins S, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*. 2007;356:237–247.
15. Farooqi IS, Matarese G, Lord GM, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*. 2002;110:1093–1103.
16. Wabitsch M, Funcke JB, Lennerz B, et al. Biologically inactive leptin and early-onset extreme obesity. *N Engl J Med*. 2015;372:48–54.
17. Lahlou N, Clement K, Carel JC, et al. Soluble leptin receptor in serum of subjects with complete resistance to leptin: relation to fat mass. *Diabetes*. 2000;49(8):1347–1352.
18. Farooqi IS, Jebb SA, Langmack G, et al. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med*. 1999;341:879–884.
19. Clement K, Biebermann H, Farooqi IS, et al. MC4R agonism promotes durable weight loss in patients with leptin receptor deficiency. *Nat Med*. 2018;24:551–555.
20. Krude H, Biebermann H, Luck W, et al. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet*. 1998;19:155–157.
21. Krude H, Biebermann H, Schnabel D, et al. Obesity due to proopiomelanocortin deficiency: three new cases and treatment trials with thyroid hormone and ACTH4-10. *J Clin Endocrinol Metab*. 2003;88:4633–4640.
22. Farooqi IS, Drop S, Clements A, et al. Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes*. 2006;55:2549–2553.
23. Lee YS, Challis BG, Thompson DA, et al. A POMC variant implicates beta-melanocyte-stimulating hormone in the control of human energy balance. *Cell Metab*. 2006;3:135–140.
24. Jackson RS, Creemers JW, Ohagi S, et al. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*. 1997;16:303–306.
25. Jackson RS, Creemers JW, Farooqi IS, et al. Small-intestinal dysfunction accompanies the complex endocrinopathy of human proprotein convertase 1 deficiency. *J Clin Invest*. 2003;112:1550–1560.
26. Alsters SI, Goldstone AP, Buxton JL, et al. Truncating homozygous mutation of carboxypeptidase E (CPE) in a morbidly obese female with type 2 diabetes mellitus, intellectual disability and hypogonadotrophic hypogonadism. *PLoS One*. 2015;10:e0131417.
27. Kuhnen P, Clement K, Wiegand S, et al. Proopiomelanocortin deficiency treated with a melanocortin-4 receptor agonist. *N Engl J Med*. 2016;375:240–246.
28. Clement K, van den Akker E, Argente J, et al. Efficacy and safety of setmelanotide, an MC4R agonist, in individuals with severe obesity due to LEPR or POMC deficiency: single-arm, open-label, multicentre, phase 3 trials. *Lancet Diabetes Endocrinol*. 2020;8:960–970.
29. Huszar D, Lynch CA, Fairchild-Huntress V, et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*. 1997;88:131–141.
30. Farooqi IS, Keogh JM, Yeo GS, et al. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med*. 2003;348:1085–1095.
31. Stutzmann F, Tan K, Vatin V, et al. Prevalence of melanocortin-4 receptor deficiency in Europeans and their age-dependent penetrance in multigenerational pedigrees. *Diabetes*. 2008;57:2511–2518.

32. Lubrano-Berthelier C, Dubern B, Lacorte JM, et al. Melanocortin 4 receptor mutations in a large cohort of severely obese adults: prevalence, functional classification, genotype-phenotype relationship, and lack of association with binge eating. *J Clin Endocrinol Metab*. 2006;91:1811–1818.
33. Rene P, Le Gouill C, Pogozheva ID, et al. Pharmacological chaperones restore function to MC4R mutants responsible for severe early-onset obesity. *J Pharmacol Exp Ther*. 2010;335:520–532.
34. Brouwers B, de Oliveira EM, Marti-Solano M, et al. Human MC4R variants affect endocytosis, trafficking and dimerization revealing multiple cellular mechanisms involved in weight regulation. *Cell Rep*. 2021;34:108862.
35. Martinelli CE, Keogh JM, Greenfield JR, et al. Obesity due to melanocortin 4 receptor (MC4R) deficiency is associated with increased linear growth and final height, fasting hyperinsulinemia, and incompletely suppressed growth hormone secretion. *J Clin Endocrinol Metab*. 2011;96:E181–E188.
36. Collet TH, Dubern B, Mokrosinski J, et al. Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency. *Molecular Metabolism*. 2017;6:1321–1329.
37. Tallam LS, Stec DE, Willis MA, et al. Melanocortin-4 receptor-deficient mice are not hypertensive or salt-sensitive despite obesity, hyperinsulinemia, and hyperleptinemia. *Hypertension*. 2005;46:326–332.
38. Greenfield JR, Miller JW, Keogh JM, et al. Modulation of blood pressure by central melanocortinergic pathways. *N Engl J Med*. 2009;360:44–52.
39. Sayk F, Heutling D, Dodt C, et al. Sympathetic function in human carriers of melanocortin-4 receptor gene mutations. *J Clin Endocrinol Metab*. 2010;95:1998–2002.
40. Iepsen EW, Zhang J, Thomsen HS, et al. Patients with obesity caused by melanocortin-4 receptor mutations can be treated with a glucagon-like peptide-1 receptor agonist. *Cell Metab*. 2018;28:23–32.e3.
41. Hatoum IJ, Stylopoulos N, Vanhoose AM, et al. Melanocortin-4 receptor signaling is required for weight loss after gastric bypass surgery. *J Clin Endocrinol Metab*. 2012;97:E1023–E1031.
42. Ramachandrappa S, Raimondo A, Cali AM, et al. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest*. 2013;123:3042–3050.
43. Bonnefond A, Raimondo A, Stutzmann F, et al. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *J Clin Invest*. 2013;123:3037–3041.
44. Kublaoui BM, Gemelli T, Tolson KP, et al. Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice. *Mol Endocrinol*. 2008;22:1723–1734.
45. Swaab DF, Purba JS, Hofman MA. Alterations in the hypothalamic paraventricular nucleus and its oxytocin neurons (putative satiety cells) in Prader-Willi syndrome: a study of five cases. *J Clin Endocrinol Metab*. 1995;80:573–579.
46. Bochukova EG, Lawler K, Croizier S, et al. A transcriptomic signature of the hypothalamic response to fasting and BDNF deficiency in Prader-Willi syndrome. *Cell Rep*. 2018;22:3401–3408.
47. Moir L, Bochukova EG, Dumbell R, et al. Disruption of the homeodomain transcription factor orthopedia homeobox (Otp) is associated with obesity and anxiety. *Molecular Metabolism*. 2017;6:1419–1428.
48. Gray J, Yeo GS, Cox JJ, et al. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*. 2006;55:3366–3371.
49. Sonoyama T, Stadler LKJ, Zhu M, et al. Human BDNF/TrkB variants impair hippocampal synaptogenesis and associate with neurobehavioural abnormalities. *Sci Rep*. 2020;10:9028.
50. Rios M. Neurotrophins and the regulation of energy balance and body weight. *Handb Exp Pharmacol*. 2014;220:283–307.
51. Bochukova EG, Huang N, Keogh J, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 2010;463:666–670.
52. Doche ME, Bochukova EG, Su HW, et al. Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest*. 2012;122:4732–4736.
53. Flores A, Argetsinger LS, Stadler LKJ, et al. Crucial role of the SH2B1 PH domain for the control of energy balance. *Diabetes*. 2019;68:2049–2062.
54. Jiang L, Su H, Keogh JM, et al. Neural deletion of Sh2b1 results in brain growth retardation and reactive aggression. *FASEB J:*. 2018;32(4): 1830–1840.
55. Sahoo T, del Gaudio D, German JR, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet*. 2008;40:719–721.
56. de Smith AJ, Purmann C, Walters RG, et al. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet*. 2009;18:3257–3265.
57. Manning KE, Holland AJ. Puzzle pieces: neural structure and function in Prader-Willi syndrome. *Diseases*. 2015;3:382–415.
58. Cummings DE, Purnell JQ, Vaisse C, et al. Elevated plasma ghrelin levels in Prader Willi syndrome. *Nat Med*. 2002;8:643–644.
59. Tauber M, Diene G, Molinas C. Sequelae of GH treatment in children with PWS. *Pediatr Endocrinol Rev*. 2016;14:138–146.
60. Tafaj O, Juppner H. Pseudohypoparathyroidism: one gene, several syndromes. *J Endocrinol Invest*. 2017;40:347–356.
61. Weinstein LS, Liu J, Sakamoto A, et al. Minireview: GNAS: normal and abnormal functions. *Endocrinology*. 2004;145:5459–5464.
62. Germain-Lee EL. Management of pseudohypoparathyroidism. *Curr Opin Pediatr*. 2019;31:537–549.
63. Gruters-Kieslich A, Reyes M, Sharma A, et al. Early-onset obesity: unrecognized first evidence for GNAS mutations and methylation changes. *J Clin Endocrinol Metab*. 2017;102:2670–2677.
64. Mantovani G, Bastepe M, Monk D, et al. Diagnosis and management of pseudohypoparathyroidism and related disorders: first international Consensus Statement. *Nat Rev Endocrinol*. 2018;14:476–500.
65. Chen M, Wang J, Dickerson KE, et al. Central nervous system imprinting of the G protein G(s)alpha and its role in metabolic regulation. *Cell Metab*. 2009;9:548–555.
66. Forsythe E, Beales PL. Bardet-Biedl syndrome. *Eur J Hum Genet :* 2013;21:8–13.
67. Shoemark A, Dixon M, Beales PL, et al. Bardet Biedl syndrome: motile ciliary phenotype. *Chest*. 2015;147:764–770.
68. Forsythe E, Kenny J, Bacchelli C, et al. Managing Bardet-Biedl syndrome-now and in the future. *Front Pediatr*. 2018;6:23.
69. Tahani N, Maffei P, Dollfus H, et al. Consensus clinical management guidelines for Alstrom syndrome. *Orphanet J Rare Dis*. 2020;15:253.
70. Geberhiwot T, Baig S, Obringer C, et al. Relative adipose tissue failure in Alstrom syndrome drives obesity-induced insulin resistance. *Diabetes*. 2021;70:364–376.
71. Rodrigues JM, Fernandes HD, Caruthers C, et al. Cohen syndrome: review of the literature. *Cureus*. 2018;10:e3330.
72. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012;109:1193–1198.
73. Singh T, Kurki MI, Curtis D, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19:571–577.
74. Gurrapu S, Tamagnone L. Transmembrane semaphorins: Multimodal signaling cues in development and cancer. *Cell Adh Migr*. 2016;10:675–691.
75. van der Klaauw AA, Croizier S, Mendes de Oliveira E, et al. Human semaphorin 3 variants link melanocortin circuit development and energy balance. *Cell*. 2019;176:729–742.e18.
76. Jiang Q, Arnold S, Heanue T, et al. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet*. 2015;96:581–596.
77. Yang Y, van der Klaauw AA, Zhu L, et al. Steroid receptor coactivator-1 modulates the function of Pomc neurons and energy homeostasis. *Nat Commun*. 2019;10:1718.
78. Marenne G, Hendricks AE, Perdikari A, et al. Exome sequencing identifies genes and gene sets contributing to severe childhood obesity, linking PHIP variants to repressed POMC transcription. *Cell Metab*. 2020;31:1107–1119.e12.

79. Webster E, Cho MT, Alexander N, et al. De novo PHIP-predicted deleterious variants are associated with developmental delay, intellectual disability, obesity, and dysmorphic features. *Cold Spring Harb Mol Case Stud*. 2016;2:a001172.
80. Asai M, Ramachandrappa S, Joachim M, et al. Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science*. 2013;341:275–278.
81. Pearce LR, Atanassova N, Banton MC, et al. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell*. 2013;155:765–777.

22

Regulation of Intermediary Metabolism During Fasting, Feeding, and Exercise

David E. Kelley and Bret H. Goodpaster

OUTLINE

Postabsorptive Intermediary Metabolism, 343
- Introduction, 343
- Resting Energy Expenditure and Postabsorptive Respiratory Quotient, 344
- Postabsorptive Glucose Utilization, 345
- Postabsorptive Fatty Acid Utilization, 345
- Lipolysis and Control of Fatty Acid Availability During Postabsorptive Conditions, 347
- Fatty Acid Reesterification as a Component of Flux, 348
- Postabsorptive Glucose Production, 348
- Hormonal Governance of Glucose Production, 349
- Glucose Counterregulation, 350

Prandial and Postprandial Intermediary Metabolism, 351
- Introduction, 351
- Digestion, Absorption, and Enteroendocrine Cells, 351
- Insulin and Glucagon Responses, 353
- Effects on Energy Expenditure and Respiratory Quotient, 353
- Hepatic Glucose Metabolism, 353
- Postprandial Peripheral and Central Nervous System Glucose Utilization, 354
- Postprandial Adipose Tissue Metabolism, 355
- Hepatic Metabolism of Fatty Acids and Lipoproteins, 356
- Methodologic Approaches to Investigating Postprandial Conditions, 356

Protein and Amino Acid Metabolism, 357

Exercise Metabolism, 358
- Exercise Dramatically Increases Demand For Energy, 358
- Energy Metabolism During Exercise, 358
- Glucose Metabolism During Exercise, 358
- Fatty Acid Metabolism During Exercise, 360
- Effects of Exercise Training on Intermediary Metabolism, 360

Summary and Concluding Remarks, 360

POSTABSORPTIVE INTERMEDIARY METABOLISM

Introduction

All major body processes require energy, which is provided by intermediary metabolism, the conversion of nutrient-derived substances into forms that can be combusted to yield energy immediately or after a period of storage, and that collaterally can generate substrate for use in anabolic processes. A useful definition of *postabsorptive metabolism* is a time when energy needs are being met through mobilization of endogenous substrates, after digestion, absorption, tissue distribution, oxidation, and storage of ingested food has been completed (*postprandial metabolism*). A typical duration of postprandial metabolism is at least a few hours, as many as 6 or more, depending upon the size and composition of the meal. Accordingly, if one daily meal is consumed, postabsorptive metabolism would stretch for 18 or more hours of a day. More commonly though, two to three meals are consumed in a day, the successive postprandial phases often fuse one into the next (especially for postprandial lipoprotein metabolism), and the postabsorptive phase may comprise as few as 8 hours of the daily cycle of eating and fasting.

A striking characteristic of postabsorptive metabolism is that circulating metabolite concentrations remain quite stable, which is sharply different from the dynamic changes that characterize postprandial metabolism. The stability of metabolite concentration is particularly evident for plasma glucose, and somewhat less so for plasma free fatty acids (FFAs), which have a day-to-day variance of approximately 30%, as well as a greater variability for intermediary metabolites such as lactate and ketones. Intuitively, it is appreciated that a stability of circulating levels of substrates during postabsorptive metabolism does not infer metabolic stasis, even though metabolic rates and energy expenditure are at a relative nadir during resting postabsorptive conditions. Rather, constancy of circulating substrate concentrations reflects the underlying presence of tightly governed rates of substrate flux. Flux can be considered regarding a single enzymatic reaction or along an enzymatic pathway, but in the current context substrate flux refers to its rate of appearance into plasma (from endogenous stores and newly synthesized substrate) and rate of utilization, more commonly termed rate of disappearance, from plasma into tissues.[1] A persisting stability of substrate concentrations in plasma derives from a well-balanced equivalency between rates of appearance and disappearance. Briefly, determinations of substrate flux can be performed using an isotopically labeled metabolite and measurement of its dilution in the unlabeled (endogenous) pool. This is a largely a research method rather than a clinically pragmatic determination, but an understanding of the findings, qualitatively and, in select instances, quantitatively, provides mechanistic insights for understanding postabsorptive intermediary metabolism. Physiological governance of flux will be one focus of this section, with emphasis upon its endocrine regulation, and this understanding of flux and interorgan exchange begins with a review of the energy stores of fat, carbohydrate, and protein within the major tissues, as shown in Table 22.1.

A second key characteristic of postabsorptive metabolism is an approximate and sustained balance between the oxidation of carbohydrate versus that of lipid. This balance between glucose and fat oxidation is not a true equipoise, as approximately 60% of energy production derives from fat oxidation. But this sustained pattern too is a sharp difference from postprandial metabolism, where there is a pronounced

TABLE 22.1 Fuel Reserves in a Prototypical 70-kg Male

	AVAILABLE ENERGY (kcal)		
Organ	**Glucose or Glycogen**	**Free Fatty Acids or Triglycerides**	**Protein**
Blood	18	80	0
Liver	400	450	400
Brain	8	0	0
Adipose	80	135,000	40
Muscle	1200	350	24,000

early suppression of fat oxidation and stimulation of glucose oxidation, with a more gradual resumption of fat oxidation and coordinate decline of glucose oxidation as the transition toward postabsorptive metabolism begins to occur. Delineating the tissue responsible for glucose and fat oxidation will also be a focus of this section, and in relation to overall rates of energy expenditure. As a prelude, the concept to be developed is that endocrine governance of substrate oxidation during postabsorptive metabolism is mostly indirect, in the sense that rates of glucose uptake and oxidation are not directly insulin-stimulated (as opposed to postprandial conditions), nor is fat oxidation directly controlled by endocrine signals. Instead, glucose uptake and oxidation are largely "insulin-independent," and that of fatty acids is substrate-driven, with endocrine control exerted through governing rates of release of glucose and fatty acids. A schematic depicting the primary systems involved in postabsorptive energy metabolism is shown in Fig. 22.1.

Resting Energy Expenditure and Postabsorptive Respiratory Quotient

Energy expenditure is comprised of three principal components: resting energy expenditure (REE), physical activity, and the thermic effects of food ingestion.[2] The rate of REE is approximately 1 kcal/min, or approximately 1400 to 1500 kcal per day. REE accounts for approximately 60% of daily energy expenditure, and the thermic effect contributes approximately 10%, but considering that energy expenditure due to physical activity can vary greatly, these proportionalities are generalizations. Notably, REE is proportional to body size, revealing strong linear correlation with fat-free mass, which is comprised of the more metabolically active tissues. This is similar for fasting rates of hepatic glucose production (HGP; discussed subsequently). Thus, both REE and HGP reveal a vital communication between body composition, the tissues that mostly define energy requirements, and governance of postabsorptive metabolism.[3] However, as will also be addressed, the relationship between fat mass and rates of fatty acid flux is less rigorous than that for the relationships of HGP or REE and fat-free mass.

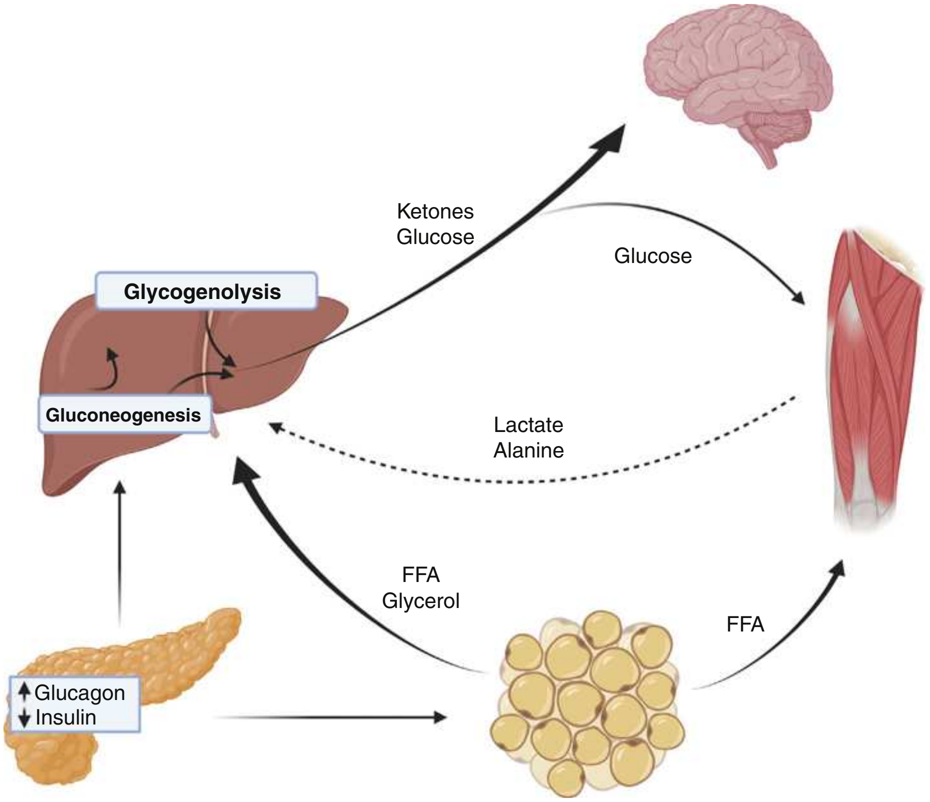


Fig. 22.1 **Postabsorptive (fasting) metabolism.** The driver of glucose flux during postabsorptive intermediary metabolism is hepatic glucose production, control over which is mediated by a relatively low ratio of insulin to glucagon. Because of low peripheral circulating insulin and in relation to the molecular physiology of tissue glucose transporter, glucose utilization in peripheral tissues is largely insulin-independent, and, apart from the central nervous system, is lower than the utilization of fatty acids. Low peripheral circulating insulin enables lipolysis in adipose tissue to proceed supporting a rate of free fatty acid (FFA) flux that is roughly double the circulating energy equivalence of that for glucose and underlying the 60% contribution of fat oxidation to resting rates of energy expenditure.

Measurement of REE (and the other components of energy expenditure) is most commonly performed by measurement of oxygen consumption and carbon dioxide production using the method of indirect calorimetry. An alternative is a doubly-labeled water isotopic method that can be used over short intervals in freely living individuals.[4] Each approach has advantages and limitations. One of the advantages of indirect calorimetry is that it can provide much more finite or precise temporal resolution of patterns of energy expenditure and substrate oxidation rates. Gas exchange can be measured in dedicated, specifically engineered calorimetry chambers, or less precisely but more accessibly at the bedside with a hood canopy or on a treadmill with a mouthpiece to collect respired air that feeds into gas analyzers. The equations of calculation of energy expenditure are referenced, together with a critique of some of the nuances and limitations of indirect calorimetry.[5] Beyond its use in metabolic research, one of the pragmatic clinical uses of indirect calorimetry is in assessing metabolic responses to enteral or parenteral nutrition in critically ill patients. Using these determinations, the quotient of carbon dioxide (CO_2) production to oxygen (O_2) consumption can be examined; this is termed the respiratory quotient (RQ). Values for RQ denote the relative reliance upon carbohydrate versus lipid oxidation, and with some additional determinations (e.g., urinary nitrogen), the contribution of protein oxidation, usually a modest clarifying adjustment, can also be estimated. Using rates of O_2 consumption, CO_2 production, and urinary nitrogen, actual rates of glucose and fat oxidation can be calculated. However, to examine concepts germane to integration of glucose and fat oxidation during postabsorptive and postprandial conditions (as well as during exercise), the RQ itself is a useful and straightforward parameter. The principles underlying the RQ stem from the chemical composition of oxidative substrates. Oxidation of glucose (a polymer of six CH_2O) to six CO_2 and six H_2O requires a consumption of an equimolar quantity of six O_2; hence, the corresponding RQ has a value of 1.0. Oxidation of fatty acid (a polymer of CH_2 units) requires a proportionately greater amount of oxygen for complete combustion and yields approximately in 7 moles of CO_2 production relative to a requirement for 10 moles of O_2 consumption; hence, the RQ for fat oxidation is 0.70. Intermediate RQ values reflect an admixture of glucose and fat oxidation. In a lean, healthy human, a typical fasting RQ value is in the range of 0.78 to 0.83. It can be estimated from this range of RQ that, during resting postabsorptive conditions, the oxidation of fat provides energy that accounts for approximately 60% of REE.

KEY POINTS: Resting Energy Expenditure

- Resting energy expenditure (REE) is approximately 1 kcal/min, 1400–1500 kcal/day, though proportional in amount to fat-free mass, which contains the metabolically active tissues that contribute to REE.
- REE generally accounts for approximately 60% of daily total energy expenditure, though the proportion will vary in relation to levels of energy expended in physical activity.
- Measurement of oxygen consumption and carbon dioxide production for estimation of energy expenditure also can be used to estimate respective rates of glucose and fat oxidation.

Postabsorptive Glucose Utilization

There is a well understood persistent and obligate requirement of the central nervous system (CNS) for glucose uptake from plasma and its oxidation to meet its bioenergetic needs. The brain does not store glycogen, and without continuous glucose delivery becomes deprived of its requisite substrate. This rate of CNS glucose uptake is approximately 70 to 80 mg/min or approximately 1 mg/min-kg body weight of glucose flux, which is about half of the postabsorptive rate of glucose flux (as further described). The predominant glucose transporter of the CNS is GLUT3, a high-affinity (low Km) glucose transporter, and as such GLUT3 maximizes glucose transport within the normal (i.e., euglycemic) range of fasting plasma glucose.[6] CNS glucose uptake is not further increased at higher plasma concentrations, as during postprandial conditions. As a further consequence of this molecular physiology of GLUT3, CNS glucose uptake can become critically reduced with a 25% to 30% reduction in fasting plasma glucose concentrations, as there is a sharp decline in CNS glucose uptake as plasma glucose concentration declines to levels lower than this.

Due to the properties of GLUT3, glucose uptake and oxidation by the brain is categorized as "insulin-independent," as is the great majority of the remaining glucose flux during postabsorptive conditions. Other than the brain, there is no single peripheral tissue that predominates in rates of postabsorptive glucose uptake; instead, glucose uptake is dispersed somewhat evenly across tissues, each at a relatively low rate and with an obligate need by red blood cells. Only a minor proportion of the peripheral tissues' (i.e., non-CNS) directed glucose flux, also approximately 1 mg/kg-min, undergoes oxidation; instead, it is mostly metabolized by glycolysis to the level of trioses (e.g., pyruvate and interconversion to lactate), while some undergoes a transamination reaction to form alanine, and these trioses reenter circulation as part of an interorgan exchange. A major fate is as a substrate for gluconeogenesis in the liver, completing an ongoing, recurring cycle within postabsorptive metabolism.

KEY POINTS: Molecular Physiology of Glucose Transporters as a Determinant of Glucose Utilization During Postabsorptive Conditions

- GLUT3, expressed in the CNS, is a high-affinity isoform that maximizes glucose uptake across the normal range of fasting glucose and does not further increase as blood glucose rises.
- GLUT4, expressed in skeletal muscle, heart, and adipose tissue, is sequestered in cytosolic vesicles under resting postabsorptive conditions, leading to low rates of glucose use by these tissues.
- GLUT2, expressed in liver and in pancreatic β cells, is a low-affinity isoform that enables blood glucose–dependent rates of transport, an attribute that further serves as a glucose sensor within these tissues that have key roles in responding to blood glucose changes.

Postabsorptive Fatty Acid Utilization

Nearly all tissues and organs have a capacity to oxidize fatty acids, with notable exceptions being red blood cells that lack mitochondria and the CNS, due to inability of fatty acids to traverse the blood–brain barrier. Kidney, heart, and liver are key tissues accounting for uptake and oxidation of fatty acids during postabsorptive metabolism, as is skeletal muscle. The persistent reliance of the CNS on glucose oxidation has already been stated, and, in an integrated physiology context, it is fat oxidation that accounts for a major proportion of energy production in peripheral tissues during postabsorptive conditions; as noted earlier, it accounts for approximately 60% of REE.

It is worthwhile to delve a bit deeper into skeletal muscle metabolism, as an example of a tissue that has a high postabsorptive utilization of fatty acids and yet also has a high capacity for glucose utilization, a capacity to transition or toggle between fuels that is manifest by other tissues, though perhaps not in as pronounced a manner as skeletal muscle. Thus, skeletal muscle has a robust capacity to "switch" fuels, and this metabolic flexibility is an important manifestation of metabolic health,[7,8] in appraising governance and response in transitions between postabsorptive and postprandial metabolism. The governing of this switching of fuels will be more fully addressed later.

During resting postabsorptive conditions, skeletal muscle has a quite low rate of glucose uptake and relies instead chiefly upon fatty acid oxidation. It has been estimated, based on arteriovenous differences across an arm or a leg (tissue beds that are predominated by, though not exclusively comprised of, skeletal muscle) that fractional extraction of plasma glucose is only 1% to 2%, whereas that for plasma fatty acids is approximately 30% to 35% during resting postabsorptive conditions.[9] The low rate of glucose usage by skeletal muscle during fasting conditions is attributable to low circulating insulin levels during postabsorptive conditions, inadequate to stimulate glucose uptake by skeletal muscle. Moreover, during fasting conditions there is a net efflux of pyruvate, lactate, and alanine from skeletal muscle, that as an aggregate molar amount of efflux approximates the uptake of glucose, leaving a net zero balance for glucose-derived carbons. The uptake and glycolytic metabolism of glucose by skeletal muscle, returning lactate to venous blood and for uptake in the liver as a substrate for gluconeogenesis, represents the well-described Cori cycle.[10] There is as well a low intrinsic rate of glycogenolysis in resting skeletal muscle, which contrasts with more active postabsorptive glycogenolysis in liver. Though muscle glycogen reserves, which in aggregate approximate 400 g (dry weight), are abundant and constitute approximately 80% of systemic glycogen reserves, the pool is held for mobilization in response to physical activity, especially strenuous exercise, as will be delineated later. Furthermore, skeletal muscle lacks glucose-6-phosphatase and cannot release free glucose.

A series of now classic human physiology studies identified that fatty acid oxidation is the principal energy source of resting skeletal muscle during postabsorptive conditions.[11] This finding was based upon the measurement of the RQ across the forearm, based on regional determinations of O_2 uptake and CO_2 production across the limb (rather than the systemic gas exchange earlier described). An RQ value in the range of 0.70 to 0.76 was observed in healthy, lean individuals, lower than systemic RQ (not surprising given the exclusion of CNS glucose oxidation from the regional measurement) and denoting a nearly exclusive reliance upon fat oxidation. Skeletal muscle, like most nonadipose tissues, does contain lipid droplets (LDs), comprised mostly of triglyceride (TG), and fatty acids can be mobilized from these droplets, as well as exchange with the TG in the droplets on uptake into myocytes, and uptake of plasma FFAs is the principal source for oxidation during resting postabsorptive conditions.

The uptake of FFAs into skeletal muscle and other tissues is not directly governed by hormonal actions at the sarcolemma, unlike the direct actions of insulin to stimulate glucose transport into muscle; rather, endocrine governance of fatty acid uptake into muscle and other tissues is indirect, mediated by endocrine control of lipolysis, and hence governance of plasma FFA concentrations. Plasma concentration of FFAs is the key factor governing rates of uptake into muscle and other tissues. During postabsorptive conditions, a typical circulating concentration of plasma FFAs is 0.5 mmol/L, or approximately one tenth the molar concentration of plasma glucose. Moreover, FFAs, being strongly hydrophobic molecules, circulate noncovalently bound to albumin; there are generally one to three fatty acids bound per albumin molecule at concentrations common in postabsorptive conditions. Fatty acids taken up by tissues are from the much smaller "free fraction" of unbound fatty acids that exists in equilibrium with albumin-bound fatty acids.

A comparison of the respective concentrations of glucose and FFAs in plasma (e.g., 5.0 vs. 0.5 mmol/L, respectively) as an index of relative availability during postabsorptive metabolism is somewhat misleading, because it overlooks the respective underlying rates of glucose and FFA flux. As earlier noted, a typical postabsorptive rate of plasma glucose flux is 2 mg/min-kg body weight (10 μmol/min-kg body weight; in aggregate approximately 700 μmol/min for a 70-kg individual). The rate of appearance of fatty acids into plasma is not too dissimilar, at approximately 1.7 mg/min-kg (6–8 μmol/min-kg body weight; in aggregate approximately 400–500 μmol/min in a 70-kg individual).[12,13] This rough equivalency in flux, somewhat lower for fatty acids than for glucose, reflects a more rapid turnover rate (shorter half-life) for FFAs than for glucose; the half-lives are 2 minutes and 15 minutes, respectively, for plasma FFAs and glucose. Also, the higher energy density of FFA relative to glucose (9 compared with 4 kcal/g, respectively) means that the potential circulating energy yield of FFA is nearly twice that of glucose. Taking each of these factors into consideration and recognizing the considerably more rapid turnover rate of the plasma FFA pool than that of glucose, it can be more clearly understood how it is that the oxidation of plasma FFAs accounts for approximately 60% of REE during postabsorptive conditions, despite an order of magnitude difference in plasma concentrations and an exclusive use of glucose by the CNS.

To delve a bit further into postabsorptive kinetics of glucose and fatty acid flux, it was earlier pointed out that rates of fasting HGP are linearly proportional to body weight, and more tightly so to fat-free mass. There is a linear relationship between postabsorptive rates of fatty acid rate of appearance and fat mass, the latter reflecting of course the tissue of origin from which fatty acids are released via lipolysis. This correlation accounts for roughly 40% of the interindividual variance in postabsorptive fatty acid flux, at least across a cohort of men and women ranging from lean to obese,[14] so it is an association of moderate strength only. Also, of interest, there is a rather strong negative, curvilinear correlation between postabsorptive fatty acid flux (normalized to per kg fat mass) and obesity, suggesting that the rate of release of fatty acid from adipose, per unit weight of fat, is reduced in obese relative to leaner individuals. Another way to express this is that, in lean individuals, the mobilization rate (i.e., lipolytic rate) per kilogram of fat mass is substantially accelerated compared with that in obese individuals. Yet, despite this, plasma fatty acid levels tend to be somewhat increased in obese compared with lean individuals, and thus, even though there appears to be a per unit of fat greater stringency in control of lipolysis, this does not appear to fully compensate for the overall expansion of fat mass. There are also findings indicating that body fat topography is associated with variances in rates of postabsorptive fatty acid flux, with higher rates in obese women who have an upper body fat distribution compared with obese women with a lower body fat distribution, the latter not differing in postabsorptive fatty acid flux from lean women.[15]

At the cellular level, unbound FFAs enter a cell down a concentration gradient (plasma to cytosol) by "flip-flop" diffusion across the cell membrane, and there is facilitation by tissue-specific families of fatty acid transporters. Once within the cytosol, fatty acids noncovalently bind to fatty acid–binding proteins, which protects the cell from detergent-like, damaging properties of unbound FFAs. Intracellular fatty acids are quickly "activated" by esterification to CoA, to form fatty acyl-CoA, an adenine triphosphate (ATP) requiring reaction catalyzed by the enzyme acyl-CoA-synthase, a reaction that is essential in maintaining a concentration gradient relative to plasma. Fatty acyl-CoA can be partitioned toward reesterification to form TG in LDs or other pathways of lipid synthesis (e.g., phospholipids, sphingolipids), but during postabsorptive conditions a considerable fraction are partitioned toward β-oxidation. β-oxidation is a repeating four-step catabolic set of enzymatic reactions that, with each cycle, progressively cleaves the full-length fatty acid into two-carbon fragments, acetyl-CoA moieties, which can then enter the TCA cycle to complete final oxidation. Both processes occur within the mitochondria matrix, yet fatty acyl-CoAs are unable to diffuse across the outer and inner mitochondrial membranes, which possess a phospholipid composition unique from that of the cell membrane. Instead of diffusion, a specific a three-component transport complex, the carnitine-palmitoyl transferase complex (CPT), mediates this entry of fatty acyl-CoA. Greater details of the intracellular catabolism of fatty acids are beyond the scope of this chapter, but the biochemistry is well described.[16]

There is one aspect of this molecular physiology that warrants highlighting, because it is a fulcrum for the governance of fat oxidation, and hence pivotal as well in the governance of intermediary metabolism. This aspect is the role of malonyl CoA in allosteric modulation of the activity of CPT-1, the enzyme that mediates transfer of an acyl moiety from acyl-CoA to carnitine, and it is the carnitine-acyl that is shuttled into the mitochondrial matrix, where acyl-CoA is reformed before entry into the β-oxidation cycle. The activity of CPT-1 is strongly inhibited by malonyl CoA, though there are some tissue differences in this sensitivity, with the isoform of CPT-1 in skeletal muscle being more sensitive to malonyl CoA inhibition than the isoform in liver. Malonyl CoA is a crucial intermediate in the process of *de novo* lipogenesis, derived in turn from a surfeit of acetyl-CoA, generated by glycolysis and hence conversion from pyruvate. Even in tissues like skeletal muscle that have a low overall capacity for *de novo* lipogenesis, the generation of malonyl CoA serves an important substrate-based signaling function in a reciprocal on-off toggling of glucose and fatty acid oxidation.[17]

During postabsorptive conditions, malonyl CoA concentrations in skeletal muscle (and other tissues) are low, and CPT-1 is unimpeded, and this, together with relatively avid uptake of plasma fatty acids, drives fat oxidation. Glucose oxidation is low, due primarily to the low rates of glucose uptake and glycolysis, and activity of the pyruvate dehydrogenase complex, which catalyzes formation of acetyl-CoA (for entry into the TCA cycle) from pyruvate upon the latter's entry into mitochondria, is correspondingly low as compared with insulin-stimulated conditions.

In the liver, the good availability of fatty acids during postabsorptive conditions, low concentrations of malonyl CoA, and, thus, relatively unimpeded CPT complex activity, create conditions wherein a surfeit of acetyl-CoA derived from β-oxidation of acyl-CoA can give rise to ketogenesis. Ketone bodies are the most water-soluble form of lipid-derived metabolites and are formed in the liver from the condensation of two acetyl-CoAs. Ketones (acetoacetate and β-hydroxybutyrate) can be utilized by numerous tissues and production rates, and blood concentrations rise with prolongation of fasting. With prolonged fasting, the CNS adapts for utilization of ketones, which provides an additional mechanism for sparing the use of glucose. Clinically, in severe insulin deficiency, accelerated rates of ketone production cause diabetic ketoacidosis.

KEY POINTS: Postabsorptive Fatty Acid Rate of Appearance and Utilization

- Usual concentrations of plasma fatty acids are one tenth of those of plasma glucose, but the rate of appearance (in mg/min-kg body weight) is roughly equivalent to glucose, and the energy yield is nearly 2-fold higher than for glucose.
- The half-life of plasma fatty acids is just 2 min, nearly an 8-fold more rapid turnover of plasma fatty acids than glucose.
- As a fraction of stored triglyceride in adipose tissue, the rate of lipolysis is a miniscule fraction (~10^{-5}), a thousandth of the rate of release of hepatic glycogen, reflecting tight, minute-to-minute governance of lipolysis.
- The endocrine control of lipolysis mostly represent the opposing actions of insulin (restraining) and catecholamines (stimulating) lipolysis.

Lipolysis and Control of Fatty Acid Availability During Postabsorptive Conditions

In the introduction to this section on postabsorptive intermediary metabolism, emphasis was placed upon the stability of fasting substrate concentrations, especially that of plasma glucose. In comparison, greater variability on a day-to-day basis is observed for plasma concentration of fatty acids, and this greater variability relative to glucose can also be observed within any given interval of postabsorptive metabolism, wherein fluctuations of plasma fatty acid concentrations are more evident than those for glucose. However, a conclusion that governance of fasting rates of fatty acid flux is less stringent than that of fasting rates of glucose flux is arguably incorrect, and the context to refute this notion is outlined here. A first context is to consider the size of the potential fatty acid pool (i.e., the reservoir of stored TG) and the extent to which a fasting rate of fatty acid flux represents a fractional turnover of this pool. Second, the rapid turnover rate of plasma fatty acids (half-life of approximately 2 minutes) is such that any modulation of rates of appearance of fatty acids quickly resettles into a near steady-state equilibrium and can accordingly manifest as a change in fatty acid concentration. Thus, a change in plasma fatty acids reveals acute changes in the level of control of lipolysis, and perhaps as well variance in rates of clearance by tissues of utilization, not *per se* a loss or absence of control. Third, and as will be delineated next, the specific steps controlling lipolysis are multifaceted and complex and, while still very much an area of new discoveries, denote intricate governance.[18]

Even in a lean individual there is a considerable reserve of stored energy in the form of TG in white adipose tissue. In a 70-kg lean adult, fat mass comprises approximately 10% of body weight in a man and somewhat more, approximately 30% in a healthy lean woman; certainly, these percentages are much greater in overweight and obese men and women. Each gram of fat, fully combusted, yields 9 kcal; thus, 7 kg of fat mass in a lean, 70-kg male or 20 kg of fat mass in a lean, 70-kg woman equates to a potential energy reserve (estimated at 7,700 kcal/kg adipose) of 77,000 kcal and 231,000 kcal, respectively, in a man or a woman. This far exceeds the caloric equivalent of 400 kcal corresponding to a replete hepatic glycogen content of 100 g and the 1600 kcal corresponding to a replete skeletal muscle glycogen content of 400 g. Thus, a fatty acid flux of approximately 120 mg/min in a 70-kg adult represents a small minute-to-minute fractional release (turnover) of the 10 to 15 kg of adipose tissue in a man or woman.

In white adipose tissue, adipocytes contain one large LD with a narrow perimeter of cytosol and organelles. The hydrophobic core of the LD is largely TG, its surface covered by a phospholipid monolayer coated with protein chaperones, notably perilipins, that can, in a regulated manner, constrain access by lipases to TG.[18] There are three key lipases that can carry out lipolysis of TG in a LD. These start with adipose triglyceride lipase (ATGL): TG is its preferred substrate, its activity is modulated by a protein cofactor, release of which is governed by and related to conformational changes in perilipins induced by endocrine stimulators of lipolysis, and its reaction yields a fatty acid and diacylglycerol. Upon the latter, hormone-sensitive lipase (HSL) next acts to yield a fatty acid and monoacylglycerol, and HSL activity, which has been well-studied, is increased by endocrine stimulators of lipolysis and inhibited by insulin, the endocrine suppressor of lipolysis. The third lipase is monoacylglycerol lipase, which cleaves the remaining fatty acid and thereby also yields glycerol.

The overall yield of lipolysis is three fatty acids and free glycerol. Free glycerol exits the adipocyte and serves as a gluconeogenic precursor in the liver. Increased cyclic adenosine monophosphate (cAMP) or cyclic guanosine monophosphate second messengers generated by catecholamines, growth hormone, atrial natriuretic peptide, and potentially glucagon (though the physiologic relevance of glucagon *in vivo* is unclear) signal transduction, and are the endocrine stimulators of lipolysis. These regulate reversible phosphorylation that induces conformational changes in perilipins and HSL activation. Insulin signaling induces the opposite set of responses. It is worth noting that, while catecholamines can stimulate lipolysis in adipose tissues and glucose production by the liver, the adipose response is more sensitive. This also applies to the suppressive effects of insulin: adipose is more responsive compared with liver.

In plasma, fat circulates in two forms, as FFAs and as TG contained in lipoproteins; most of this circulating TG during postprandial conditions is contained in very low density lipoproteins (VLDL) and with the addition of chylomicrons during postprandial conditions. Lipoprotein metabolism is addressed comprehensively in Chapter 25. TG catabolism certainly occurs during postabsorptive conditions and can accordingly supply fatty acids for oxidative metabolism, but a large majority of fat oxidation, perhaps as great as 90% during postabsorptive conditions, derives from circulating fatty acids. Brown adipose tissue too contains LDs, and LDs are found in numerous nonadipose tissues, but in aggregate these are minor reserves relative to white adipose tissue, and in general the fatty acids liberated from these LDs are consumed within the tissues in which these are present and not released into circulation, analogous therefore to the situation with muscle glycogen.

KEY POINTS: Lipolysis of Triglyceride Contained in the Adipocyte Lipid Droplet

- In addition to the phospholipid monolayer of lipid droplets (LDs), perilipin proteins are on the surface of LDs and can limit access of lipases to the TG contained in the core of LDs.
- In response to endocrine stimulators of lipolysis (catecholamines, growth hormone, atrial natriuretic peptide), the increase in second messenger cyclic adenosine monophosphate (or cyclic guanosine monophosphate) induces conformation changes in the perilipin coating of LD and makes accessible the protein cofactor for activation of adipose triglyceride lipase, the lipase that cleaves free one fatty acid to generate diacylglycerol.
- Stimulators of lipolysis also activate hormone-sensitive lipase (HSL), which cleaves a fatty acid from diacylglycerol; the activation of HSL and its inactivation by insulin (the sole endocrine suppressor of lipolysis) is a fulcrum of control over rates of lipolysis.
- The remaining fatty acid is cleaved by the action of monoacylglycerol lipase, and this too generates free glycerol. Adipose tissue lacks glycerol kinase, so glycerol liberated by lipolysis cannot be activated to form glycerol-3-phosphate, and thus cannot form a new backbone for TG synthesis. Instead, glycerol freed by lipolysis exits adipose to serve as a gluconeogenic precursor. Glycerol-3-phosphate for TG synthesis is derived from glycolysis in the adipocyte.

Fatty Acid Reesterification as a Component of Flux

There is another aspect of fatty acid flux that is germane to understanding postabsorptive intermediary metabolism, and this is fatty acid reesterification (to TG). Fatty acid esterification, an anabolic process, is the predominant pathway in adipocytes during postprandial conditions (when lipolysis is strongly suppressed), but some partitioning of newly liberated fatty acids to reesterification occurs in adipocytes during postabsorptive conditions, though mostly this takes place in other organs, notably the liver, to which fatty acids are taken up, and fatty acyl-CoAs are partitioned between pathways of esterification and oxidation. Theoretically, following lipolysis of TG, a molar 3:1 ratio of fatty acids to glycerol release from adipose tissue would be anticipated, but the observed ratio is closer to 2 to 2.5:1, denoting an incomplete release of fatty acids. Glycerol newly liberated from TG is released into circulation (and serves as a gluconeogenic precursor in liver), because adipocytes lack glycerol kinase to catalyze the formation of glycerol-3-phosphate (G-3P) from glycerol. Instead, G-3P needs to be generated from the glycolytic intermediate dihydroxyacetone, which in turn depends upon ongoing glucose uptake into adipocytes (which contain scant glycogen and do not carry out gluconeogenesis). When G-3P is available, a portion of newly released fatty acids undergo reesterification into TGs, constituting a futile cycle in this pathway.

After release into circulation, a substantial portion of circulating fatty acids is oxidized, and this fraction is estimated to represent approximately one half (somewhat less) of fatty acid flux under resting postabsorptive conditions. The balance undergoes reesterification; in the liver this contributes to TG packaged into VLDL as well as LD TG.[13] A portion contributes substrate to other pathways of lipid anabolic metabolism, such as phospholipid and sphingolipid synthesis. Considering the "excess" portion of plasma fatty acids, that portion otherwise destined to be partitioned toward reesterification, this portion in the plasma pool can be perceived as an immediately available "reserve" that can quickly "diverted" to meet an arising need for greater fat oxidation, if energy needs dictates, as for instance during the onset of physical activity. This ephemeral plasma pool is more immediately available than allowing for a time lag for mobilization of more fatty acids via an uptick in rates of adipose lipolysis and essentially buys time for this to manifest in increased fatty acid rate of appearance.

Postabsorptive Glucose Production

In a healthy 70-kg human, during postabsorptive conditions, there is at any moment approximately 4 g of glucose in circulation,[19] a reserve representing 16 kcal of available energy (~4 kcal/g glucose). Depletion of just a third of this amount leads to compromised CNS function and morbidity. Thus, the circulating 4 g of glucose is a deceptively thin reserve; consider that this is the glucose content of a "lifesaver" candy. Yet hypoglycemia rarely, if ever, occurs in an otherwise healthy individual, and circulating glucose is replenished continuously by HGP. HGP is governed judiciously, such that the fasting level of plasma glucose is maintained within a tight and stable range, and there are several pathways of metabolism that converge to constitute HGP.[20] In addition to HGP, there is a contribution from the kidneys, estimated to account for 10% to 15% of postabsorptive glucose production,[21] that is further addressed later. Together, the renal and hepatic contribution comprise endogenous glucose production, a term that somewhat erroneously but commonly is used interchangeably with HGP.

During postabsorptive conditions, the rate of HGP is approximately 2 mg/min-kg body weight (or 140 mg/min for the 70-kg human). At the transition into postabsorptive metabolism, if the preceding meals fulfilled daily caloric requirements, hepatic glycogen is in a replete state, at a concentration of approximately 100 g (dry weight glycogen, several fold higher due to the hydrated state of glycogen within tissues). Approximately 50% to 60% of HGP can be attributed to hepatic glycogenolysis as postabsorptive metabolism commences. Glycogenolysis is a tightly governed process, acutely regulated by endocrine signaling that coordinates an opposing reciprocal regulation of glycogen synthesis, so that there is little futile cycling of the glycogen pool. Glucagon and catecholamines, through cAMP, activate PKA, which phosphorylates (and activates) glycogen phosphorylase, while also phosphorylating (and inhibiting) glycogen synthase. Hyperglycemia, through allosteric actions of intracellular glucose in the hepatocyte, can inhibit glycogen phosphorylase. Insulin signaling can directly inhibit glycogenolysis, through an oppositely directed impact on reversible enzyme phosphorylation; thus, the fulcrum of control of glycogenolysis is the opposing actions (and relative concentrations) of glucagon and insulin, and with allosteric modulation by glucose. The action of glucose-6 phosphorylase yields free glucose, which is released into the venous circulation, and the glucose transporter isoform of the liver is GLUT2, a high-Km isoform that enables rapid transmembrane equilibration of glucose concentrations.[6] Together, GLUT2 with the hexokinase isoform of the liver (glucokinase) serve as a "glucose sensor" for the liver. This is also a pairing that acts as a glucose sensor in β cells of the pancreatic islet.

Approximately half of HGP can be ascribed to gluconeogenesis, the synthesis of new glucose molecules, using as substrate principally pyruvate, lactate, and alanine, as well as glycerol derived from lipolysis. The biochemistry of gluconeogenesis is largely a reversal of glycolysis, and the detailed biochemistry of this pathway, futile cycles within glycolysis and gluconeogenesis, and other aspects have been carefully delineated.[22] Control of gluconeogenesis is somewhat complex and entails components attributable to "substrate push" (i.e., availability of the above triose precursors), direct control by insulin, indirect control by insulin (through governing precursor availability), interaction with opposing effects of glucagon on hepatocytes, and longer-term transcriptional control over the expression of enzymes that catalyze gluconeogenesis. There is also support for the concept that insulin signaling in the CNS can exert importance governance over rates of HGP. Because of the importance of dysregulated control of gluconeogenesis to the pathogenesis of hyperglycemia in diabetes mellitus, and due to ongoing debate on the key site(s) of control of gluconeogenesis, the interested reader is provided several excellent references for deeper examination of the issues.[20,23-25] Hepatic oxidation of fatty acids, the principal oxidative substrate for the liver during postabsorptive conditions, provides the energy needed for gluconeogenesis, though fatty acid metabolites (i.e., acetyl-CoA) cannot directly provide substrate for gluconeogenesis. There is evidence that the peripheral action of insulin in inhibiting lipolysis is important in concomitantly controlling rates of gluconeogenesis, both in relation to limiting provision of glycerol as a gluconeogenic substrate and in relation to limiting provision of fatty acid, as an energy source, to the liver. This concept has been termed the "single gateway hypothesis,"[23] linking insulin control of lipolysis as a nexus by which control over gluconeogenesis is exerted. While glucagon has a strong acute effect to stimulate glycogenolysis, it also, through induction of cAMP and PKA activation, has a stimulatory effect on gluconeogenesis, mediated by reversible phosphorylation of enzymes in the pathway of gluconeogenesis. Both insulin and glucagon also mediate effects by influencing transcription control of rate-limiting enzymes in the pathway of gluconeogenesis.

Under usual daily conditions of feeding and fasting, postabsorptive metabolism prevails for 10 to 12 hours, and even this relatively short interval of fasting can result in a meaningful mobilization and depletion of hepatic glycogen (approximately 5 g per hour), especially if there is an interposed session of physical activity. Using the data provided earlier for HGP and fatty acid flux, it can be estimated that the fractional turnover (catalysis) of hepatic glycogen is approximately 0.07% per minute, a relatively small rate of breakdown, but still far greater than the estimated 0.001% per minute fractional turnover of the adipose TG pool. Nonetheless, if glycogenolysis was to proceed at a monotonic rate as postabsorptive metabolism extended, even a liver replete with glycogen would be markedly reduced within 24 hours of fasting. However, hepatic glycogenolysis does not proceed at a fixed, monotonic rate as fasting continues; instead, the proportionality between glycogenolysis and gluconeogenesis (approximately fifty-fifty proportionality at commencement of postabsorptive metabolism) modulates toward a lesser contribution of glycogenolysis and a larger contribution of gluconeogenesis as the duration of postabsorptive metabolism is extended. Regardless, hepatic glycogen is completely depleted within 48 hours of fasting. This adaptive shift, which serves to conserve remaining hepatic glycogen, occurs while largely maintaining the overall rate of endogenous glucose production.

Another aspect of adaptation to a lengthening of the usual period of fasting (e.g., 8–12 hours in a daily cycle) is an increased proportion of endogenous glucose production contributed by renal gluconeogenesis. The contribution of the kidneys to glucose homeostasis has three components: (1) renal gluconeogenesis; (2) renal utilization of glucose, especially in the renal medulla; and (3) renal resorption of filtered glucose, which is approximately 180 g per day in a healthy (nondiabetic) individual.[21] Therefore, it is the last of these, resorption of filtered glucose, that is the largest of the renal effects on glucose homeostasis. However, the focus here will be on renal gluconeogenesis. Like the liver, the kidney expresses glucose-6-phosphatase; these are the only two organs to do so and the only two organs that are capable of gluconeogenesis. Renal gluconeogenesis occurs within proximal tubular cells, which have high energy needs related to ion transport and glucose resorption and a high reliance upon fatty acid oxidation.[26] At the start of postabsorptive metabolism, the renal contribution to endogenous glucose production is estimated to comprise approximately 10% to 15%, all derived from gluconeogenesis. As the duration of fasting extends, especially past 48 hours, which causes complete depletion of hepatic glycogen, the relative contribution of renal gluconeogenesis rises to nearly 40% of endogenous glucoses production,[27] in essence replacing the now exhausted contribution earlier made from hepatic glycogen. To this is added the glucose-sparing effects of ketogenesis and increased rates of lipolysis with fasting. Parenthetically, the last citation is regarded as one of the classic physiologic papers addressing the adaptation of intermediary metabolism to prolonged fasting. Glutamine is an important substrate for renal gluconeogenesis, unlike the liver, which has a greater reliance upon alanine, though both liver and kidney use lactate as a precursor. Another difference is that glucagon is considered to have little effect on renal gluconeogenesis, and instead catecholamines, partly or mostly through influencing availability of substrate precursors, affects rates of renal gluconeogenesis.

KEY POINTS: Postabsorptive Glucose Rates of Appearance

- The rate of appearance for glucose is approximately 2 mg/min-kg body weight.
- 1 mg/min-kg is used by the central nervous system (CNS) and oxidized.
- The remaining half is distributed broadly across peripheral tissues and mostly metabolized only by glycolysis to trioses (pyruvate and lactate), which are recycled to liver as gluconeogenic precursors.
- The size of the blood glucose pool is approximately 4 g, a thin reserve equivalent to the glucose content of a "lifesaver" candy, and the turnover is relatively brisk, with a half-life for glucose of approximately 15 min.

Hormonal Governance of Glucose Production

The opposing actions of insulin and glucagon signaling on hepatocytes are the crucial endocrine regulators of HGP during postabsorptive conditions and have earlier been described with respect to specific pathways involved in HGP. During postabsorptive conditions, the action of insulin serves as a brake upon rates of glucose production, while glucagon serves as an accelerator of glucose production. As earlier noted, glucose utilization during postabsorptive conditions can properly be construed as "insulin-independent," and instead it is the action of insulin in governing HGP that is the way insulin governs rates of postabsorptive glucose flux. Thus, during postabsorptive conditions (and in sharp contrast to postprandial conditions or feeding), the physiological role of insulin is more accurately perceived as restraining or controlling rates of catabolism rather than as a stimulating anabolic influence, checking against excessively high release of stored glycogen and of excessive rates of lipolysis of TG in adipose tissue.

Pancreatic β cells are astute and sensitive glucose sensors, and even slight increases in circulating glucose evoke increased insulin secretion into the portal venous circulation. Conversely, subtle declines of plasma glucose induce corresponding declines of insulin secretion. GLUT2, a high-Km isoform that enables rapid transmembrane equilibration of glucose concentrations, together with glucokinase, serves as a nexus for "glucose sensing" in β cells, though downstream glucose metabolism is essential for triggering insulin secretion. It should also be noted that increased fatty acids can acutely stimulate insulin secretion (though long-term exposure to elevated fatty acids has untoward effects), as can amino acids, so β cells are tightly tuned to intermediary metabolism.[28] Throughout this chapter emphasis is given to how rates of substrate utilization in tissues are governed by underlying energy expenditure, but in this respect the β cell differs in that its rate of substrate uptake and metabolism is largely a function of substrate availability. This "push" aspect of substrate utilization underlies the "gluco-stat" sensing function of the β cell in its governance of intermediary metabolism.

Another important consideration is that approximately 50% of secreted insulin is extracted by the liver.[29] This fraction varies between people and has been estimated to range as high as 80% (though a value nearer 50% is more commonly observed) and to be manifest in part as a heritable factor. Clearly, hepatic first-pass insulin extraction is a key determinant of the amount of exposure of peripheral tissues to insulin at any given level of insulin secretion; the portal to peripheral gradient for insulin concentration is approximately 3:1. A second important consideration is that insulin secretion is pulsatile[30]; the pulses of insulin secretion have a periodicity of approximately 5 to 7 minutes and comprise most of the postabsorptive insulin secretion. It is thought that calcium oscillations, perhaps synchronized with glycolytic oscillations, are a mechanism underlying pulsatile insulin secretion. Although some debate continues, the pulsatile delivery of insulin is regarded as contributing to the effectiveness of insulin action in the liver; in the periphery, these pulses are greatly extinguished.

Characteristically in a metabolically healthy individual, there is a gradient in the insulin responsiveness of adipose, liver, and skeletal muscle, three key insulin-responsive tissues.[31] This cross-tissue gradient of insulin action shapes the pattern of postabsorptive metabolism (as well as governing postprandial metabolism, as will be addressed later). Adipose tissue manifests a high level of responsiveness to insulin, a greater "insulin sensitivity" than liver, as reflected in a left-shifted dose response for suppression of lipolysis relative to suppression of hepatic glucose production. Furthermore, liver is more insulin-sensitive than skeletal muscle, as reflected in a left-shifted dose-response for suppression of hepatic glucose production relative to stimulation of glucose uptake into muscle. Taking this gradient of insulin sensitivity into consideration (adipose > hepatic >> skeletal muscle), together with consideration of the portal delivery and large hepatic extraction of insulin, creating a sharp 3:1 portal to peripheral gradient of insulin concentration, a mechanistic basis can be appreciated, in that adipose and liver are the chief sites of insulin action during postabsorptive conditions, and that skeletal muscle metabolism during postabsorptive conditions is not strongly shaped by insulin action.

Pancreatic α cells are also astute and sensitive glucose sensors, but with opposite patterns of response to β cells, and even slight decreases in circulating glucose evoke proportionate, stepwise increases in glucagon secretion. There is a nearly linear response of increased hepatic glucose production to increases of portal venous glucagon.[32] And because changes in circulating glucose produce reciprocal responses of insulin and glucagon, these opposing changes in hormone concentration, typically quite subtle (at least as measurable in peripheral circulation), nonetheless achieve an amplifying physiological effect on hepatic glucose production.

KEY POINTS: Hepatic glucose production during postabsorptive metabolism:

- The rate of hepatic glucose production (HGP) after an overnight fast is approximately 2 mg/min-kg body weight, about 65% of which is contributed by glycogenolysis and the remaining 35% is contributed by gluconeogenesis.
- Insulin and glucagon are the key endocrine regulators of HGP. In fact, during postabsorptive conditions, the main action of insulin is as a brake upon HGP and lipolysis as glucose utilization is nearly entirely "insulin-independent".
- Glucagon, secreted into the portal circulation like insulin, is a potent stimulator of HGP and a crucial component of the endocrine response that comprises the hypoglycemic counter-regulatory response.
- With an extended period of postabsorptive conditions, the liver adapts to rely more upon gluconeogenesis and less upon glycogenolysis to provision HGP.

Glucose Counterregulation

Across a lifespan of 80 years or more, assuming mostly good general health, an individual will consume 80,000 meals (with corresponding ascents and descents of plasma glucose) and pass through 30,000 postabsorptive periods, yet never experience an episode of symptomatic hypoglycemia. In fact, an occurrence of hypoglycemia, especially during postabsorptive conditions, is never accepted as normal and instead heralds underlying disease, a dangerous adverse drug effect, or other discoverable extenuating circumstances. While the reciprocal actions of insulin and glucagon provide a foundational understanding of this extraordinary protection against hypoglycemia, there is a multifaceted hormonal counterregulatory response that is a vital additional buttress.[33,34] The normative counterregulatory response has been investigated by infusing exogenous insulin, permitting plasma glucose to decline until a symptomatic level of approximately 50 mg/dL is reached, and, after a relatively short interval at this level, halting the insulin infusion is and permitting the plasma glucose to recover. Monitoring of endogenous insulin (via measuring C-peptide levels in blood), glucagon, epinephrine, norepinephrine, cortisol, and growth hormone during the fall and recovery of plasma glucose delineate the counterregulatory hormone response. Based on this paradigm of insulin-induced hypoglycemia, the threshold for near-complete suppression of endogenous insulin secretion is approximately 70 mg/dL, the threshold for robust glucagon secretion is approximately 60 to 65 mg/dL, and that for eliciting epinephrine counterregulation is approximately 60 mg/dL. In response to hypoglycemia, the robust spike of glucagon secretion, much accentuated relative to the more subtle modulations to slight declines of fasting glucose, is only transient in duration, yet achieves prompt robust stimulation of glycogenolysis and a more persistent stimulation of gluconeogenesis. The catecholamine response also acts to increase HGP and accelerate lipolysis. The other key constituents of the counterregulatory hormone response are growth hormone and cortisol, and, relative to glucagon and catecholamines, occur later within the overall counterregulatory hormone response, have a slower onset of effect and a sustained effect. In diabetes mellitus, most especially in type 1 diabetes, there are notable impairments in the counterregulatory hormone response, with impaired glucagon response and, not infrequently, with impaired catecholamine response, and these contribute in a major way to the risk of insulin-induced hypoglycemia, which is one of the major iatrogenic complications of insulin therapy.[35,36]

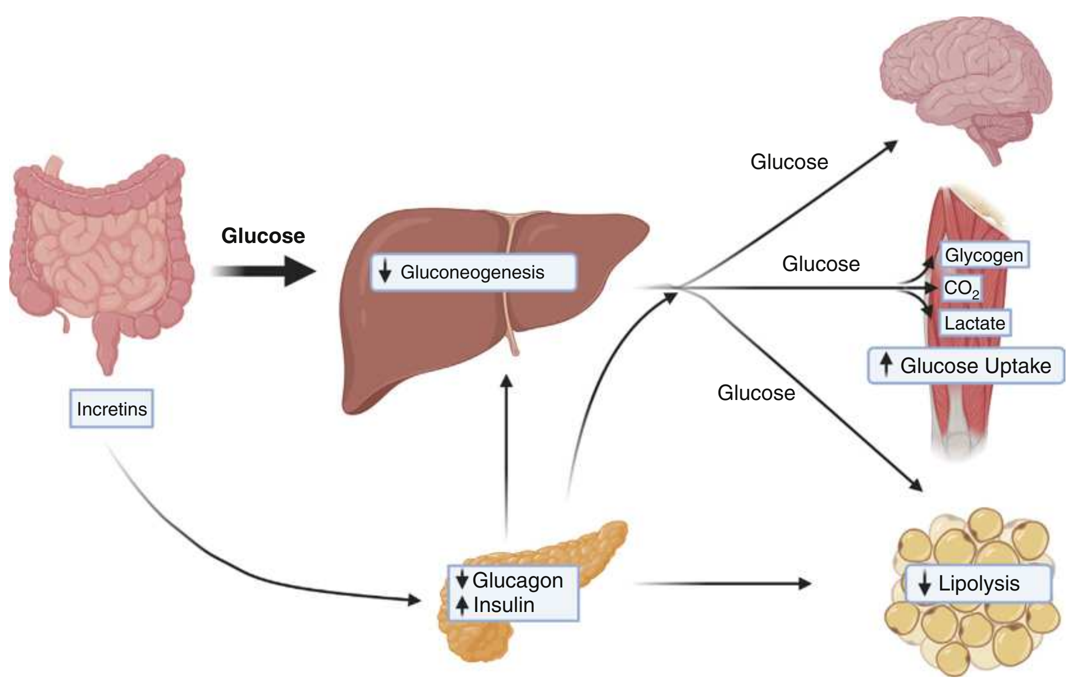


Fig. 22.2 Prandial and postprandial (fed state) metabolism. Glucose and proteins digested and absorbed in the gastrointestinal tract enter the portal venous circulation, enabling first-pass hepatic extraction, while absorbed fat instead enters the circulation via lymphatic entry and in essence undergo first-pass extraction in adipose tissue. One prominent effect of chemosensing of nutrients by enteroendocrine cells is secretion of incretins, which potentiates nutrient-stimulated insulin secretion (and suppression of glucagon secretion). Downstream actions of this pancreatic islet response include suppression of hepatic glucose production, stimulation of hepatic glucose uptake, suppression of lipolysis, and stimulation of glucose uptake into skeletal muscle. Glucose oxidation largely meets bioenergy needs during at least the first half of postprandial metabolism, and glucose availability is such that its storage as glycogen is a prominent pathway in the liver and muscle. *CO_2*, Carbon dioxide.

PRANDIAL AND POSTPRANDIAL INTERMEDIARY METABOLISM

Introduction

Ingestion of food is necessary for survival and brings pleasure (mostly), and there are complex neural circuits, gut hormone circuits, and diverse peripherally generated cues that signal hunger and elicit behaviors to seek food and respond to food ingestion.[37] Knowledge of prandial and postprandial intermediary metabolism is but one aspect of this complex physiology, yet it is foundational for understanding governance of appetite, energy balance, and weight regulation. This knowledge can also inform the dialog as to what comprises a healthy versus an unhealthy diet, though it is of course the longer-term outcomes rather than near-term changes in intermediary metabolism that are truly informative. In terms of definition, the term "prandial" refers to the period of eating, while the term "postprandial" refers to the subsequent period, when digestion and absorption are ongoing and exogenous nutrients are undergoing metabolic disposition in the liver and peripheral tissues. The duration of prandial conditions is approximately 0.5 hour, while postprandial metabolism, depending upon the size and macronutrient composition of the meal, can extend from several to 6 to 8 hours, before a transition is completed back into postabsorptive conditions (or until the next meal is ingested, starting the postprandial process anew).

The main areas of focus of this section will continue to be glucose and fatty acid metabolism within the context of endocrine regulators (Fig. 22.2). Admittedly, this focus on glucose and fatty acid metabolism can be faulted as a "rounding up of the usual suspects"; nontargeted metabolomic surveys indicate at least several thousand components, likely closer to 10,000 discrete small molecule metabolites, and thus comprise a broad spectrum of intermediary metabolism relevant to both postabsorptive and postprandial phases.[38] Considerable progress has been made in identifying biomarkers of metabolic disease and disease risk within the metabolome, one notable example being that of branched-chain amino acids, which cluster with insulin resistance and risk for type 2 diabetes mellitus. Another provocative finding from the metabolome surveys is that fatty acid length and number of double bonds can be a strong predicative risk factor for type 2 diabetes mellitus.[39] However, maturing an understanding of metabolome changes to yield mechanistic insights into the regulation of intermediary metabolism is still a work in progress. The focus here will be upon pathways of glucose and fatty acid metabolism and the context of responses to a prototypical meal, namely one that is a 7- to 10-kcal/kg caloric load (around 40% of daily calorie consumption), comprised of approximately 50% carbohydrate, 30% protein, and 20% fat. Clearly, there is considerable contemporary interest in how variations in macronutrient content and proportions may influence health, modify disease and disease risk, and influence weight regulation. However, an adequate consideration of how dietary variations further influence postprandial intermediary metabolism is beyond the scope of this section. Hopefully, a good grounding in the response to this putative standard meal will enable a balanced and critical appraisal of changes induced by diet modification.

Digestion, Absorption, and Enteroendocrine Cells

The aggregate surface area for absorption in the small intestine is estimated to be approximately 200 m^2, roughly the size of a tennis court, though more recent findings suggest it is only half this surface area.[40] Regardless, it is of a remarkable dimension, especially considering that it is folded in a fully functional capacity within the relatively tight confines of the abdominal cavity. This origami of folding of the surface area reflects the recurring pattern of millions of small intestine villi that amplify the surface area of the lumen, which is further expanded 10-fold by the brush border of the apical surface of enterocytes. A great majority of the cells of the small intestine epithelium are enterocytes, the workhorses of nutrient absorption, and these are supported by the protective action of a smaller population of mucous-secreting goblet cells. However, as a text focused upon endocrinology, it is important to call out the 1% of the gut epithelial population that are enteroendocrine cells (EECs). As a proportion of the cell population of villi, EECs are indeed a miniscule fraction, but in aggregate EECs constitute the largest endocrine organ of the body.

EECs are chemical sensors of nutrients and, in response to carbohydrate, lipid, protein, and other components of intestinal chyme, release peptide hormones and neurotransmitters.[41] What is conceptually unique about EECs relative to other endocrine glands is that, whereas the latter sense and respond to the internal milieu to maintain homeostasis, EECs respond to the external challenge presented by ingested food to prepare the gastrointestinal system for digestion and absorption, prepare peripheral tissues for metabolism of food, prepare pancreatic islets for insulin and glucagon responses, and carry out gut–brain signaling governing satiety.[42] Thus, the EECs are a complex endocrine system that is central in governing postprandial intermediary metabolism.

EECs are a highly diverse cell population. Formerly it was thought that there was one EEC cell type for each gut hormone, but recent advances in single-cell transcriptomics and other approaches now clearly demonstrate that an individual EEC can secrete several gut hormones in a coordinated and consistent manner, and that EEC subsets are distributed in an orchestrated anatomic pattern; the functional ramification of all this remains to be more fully comprehended. Several of the gut hormones secreted by EECs govern the satiety response to eating, including glucagon-like peptide-1 (GLP-1), peptide tyrosine tyrosine, and cholecystokinin, while ghrelin signals hunger. Some of these and other various gut hormones have a key role in regulating gut motility, bile acid release, pancreatic exocrine secretion, and intestinal water and electrolyte secretion. Here, we will focus upon the component of EEC response characterized as the incretin axis, and this is presented subsequently in connection with insulin and glucagon responses to meal ingestion.

Returning to the issues of digestion and absorption of food, the essential chemical reaction that renders food into its constituent components so that these are in a state that can be absorbed by small intestine enterocytes is hydrolysis. Hydrolysis cleaves starches and smaller carbohydrate polymers, cleaves acyl groups from the glycerol backbone of TGs (and from cholesterol esters), and breaks up ingested protein by cleaving amide linkages between amino acids. An array of enzymes, many secreted and others residing on the brush border surface of enterocytes, catalyze hydrolysis of ingested macromolecule nutrients. For the interested reader, more detailed information on the digestive and absorptive process is referenced.[43] The process is facilitated by and dependent upon bulk secretion of water and electrolytes that modulate pH and disperse nutrients within the gut lumen and by bile that emulsifies fat. Amylase is secreted by the exocrine pancreas in sufficient quantity to rapidly hydrolyze starch into disaccharides that are then efficiently cleaved by brush border enzymes into sugar monomers; mostly it is the monomer glucose that is absorbed and quickly released into the portal venous circulation. This abundant enzymatic capacity enables rapid entry of carbohydrate into circulation, a rate determinant being more gastric emptying than hydrolytic enzyme capacity for carbohydrate. A considerable proportion of glucose absorption is mediated by coupling with sodium absorption through the sodium-glucose transporter-1, and there is also contribution from GLUT2. Pepsin, secreted within the stomach and with optimal activity in the acidic environment of the stomach, dissembles collagen within proteins, providing pancreatic peptidases in the small intestine with more ready access to cleave protein into smaller fragments that are ultimately rendered to individual amino acids by enterocyte brush border peptidases. Amino acids, mostly similar to glucose but with some exceptions, appear relatively quickly in the portal circulation.

The situation with digestion and absorption of fat differs in several respects from that of carbohydrate and protein, nutrients that are more polar than fat and more water-soluble, and this physiology of intestinal lipid metabolism has recently been reviewed.[44] Briefly, the strong hydrophobicity of TG, the major component of ingested fat, is an impediment to its digestion. This hydrophobicity limits the surface area of fat globules and thus constrains access of pancreatic lipase and cholesterol esterase. Emulsification of fat by bile helps to form small lipid micelles, vastly expanding the surface area on which lipase can act. Parenthetically, bile acids are now recognized to have an important signaling role, clearly akin to motifs of endocrine signaling, and are the ligands for receptor-mediated actions in liver and peripheral tissues.[45] Accordingly, it is appropriate to include bile acid signaling as part of the multifaceted panoply of gut-derived messages that govern intermediary metabolism, intestinal physiology, systemic energy balance, and appetitive behavior.

Fatty acids and monoacyl-glycerol are absorbed into enterocytes, where long-chain acyl group are reesterified, and the newly formed TG is packaged with apoB48 into large chylomicron particles. Chylomicrons enter lacteals contained within small intestine villi and enter systemic circulation via the lymphatic thoracic duct, thereby foregoing first-pass hepatic extraction, unlike glucose and amino acids. Interestingly, large chylomicrons released by enterocytes have too large a diameter to permit passage through the fenestrations within hepatic capillaries, and, as will be addressed later, it is as chylomicron remnants, reduced in size by catabolism in peripheral tissues and notably adipose, that such passage becomes possible.

Another difference from carbohydrate and protein absorption is that TG entry into systemic circulation is significantly delayed, enough that ingested glucose and fat entries may, at least figuratively, be considered somewhat separate phases. It can be speculated that this temporal separation, together with the different ports of entry (portal vs. systemic) may act to diminish the fatty acid competition in the early prandial period that is predominated by glucose metabolism, and that later, with relatively prompt clearance of ingested glucose, the potential for glucose competition during metabolism of ingested fat. The notion of glucose–fatty acid substrate competition as a governing principle of intermediary metabolism is further addressed subsequently.

KEY POINTS: Digestion and Absorption of Nutrients

- The surface area of the small intestine is 100-fold greater than that of the body's skin (250 vs. 2 m^2), but food must be cleaved by hydrolysis into monomeric constituents to be absorbed by enterocytes.
- The more polar, water-soluble nutrients of glucose and amino acids are absorbed into enterocytes and released at their basal surface into the portal venous circulation for first-pass uptake and metabolism by the liver.
- The less polar, hydrophobic lipid nutrients are absorbed into enterocytes after emulsification with bile, after which lipase cleaves triglyceride (TG) to fatty acids and monoacylglycerol (MAG). On entering enterocytes, fatty acids and MAG undergo reesterification to form TG, and this is packaged into large chylomicrons.
- Chylomicrons are released into lacteals of the intestinal villi and enter systemic circulation at the level of the thoracic lymph duct, thus initially bypassing the liver, unlike carbohydrate and protein.

Insulin and Glucagon Responses

It is difficult to overstate the importance of insulin as a determinant of the metabolic response to feeding and in evoking the metabolic shift from fasting patterns of metabolism. Insulin secretion increases promptly upon eating, and by approximately 5- to 10-fold during prandial versus fasting conditions, dependent upon meal size and carbohydrate content. The sharp rise in insulin secretion has a mirror image in the suppression of glucagon secretion, the latter reflecting a regulatory role of the sharp rise in insulin to reduce glucagon secretion.

Certainly, the vigorous insulin secretion is in response to a rise in plasma glucose, and with potentiation by other nutrients (e.g., amino acids). Yet, a century ago, perceptive investigators postulated an additional impetus, at the time speculatively termed the incretin effect, the components of which were later identified.[46] One way to capsulize the effect of the incretin effect is to recount a crucial experimental finding: when a postprandial (peripheral) glucose concentration was simulated by a carefully modulated infusion of glucose, the insulin secretory response was clearly lower, by more than half, than when equivalent hyperglycemia occurred naturally in response to ingestion of glucose. This difference in the amount of insulin secreted suggested the action of a signal(s) additional to hyperglycemia itself; the components of the incretin axis were later identified as the incretin hormones (incretins) glucose-dependent insulinotropic polypeptide (GIP) and GLP-1. As noted earlier, GIP and GLP-1 are secreted by EECs. The EECs secreting GIP are predominately located in the small intestine, while EECs capable of secreting GLP-1 have a more extensive distribution in both the small and large intestines. In addition, there can be cosecretion of other gut hormones in conjunction with that of GIP and GLP-1. These EECs function as nutrient chemoreceptors responsive to ingested glucose, fatty acids and fatty acid metabolites, and amino acids. The secretion of GIP and GLP-1 is further modulated by G protein–coupled receptors (GPCRs) expressed on GIP- and GLP-1–secreting EECs. Prominent among these GPCRs are a family of lipid-sensing receptors (i.e., GPR40, GPR120, GPR119, and others) that respond to fatty acids of differing chain length and to fatty acid metabolites, the generation of which can also derive from metabolic activity in adjacent enterocytes, signifying operation of paracrine signaling.[47] In turn, many of these lipid-sensing receptors are also expressed on pancreatic β cells (and, for some, on pancreatic α cells). The receptors for GIP and GLP-1 are expressed on pancreatic β cells and α cells. Increases in GIP and GLP-1 secretion occur as early as 5 to 10 minutes after the start of eating and continue for most of the prandial period, and secretion of GIP is more sustained into the postprandial period. As alluded to earlier, GIP and GLP-1 strongly potentiate glucose-induced insulin secretion, which is also to state there is only a nominal effect to stimulate insulin secretion at euglycemia or hypoglycemia. There are emerging findings that GIP may amplify the glucagon response to hypoglycemia,[48] and that the GPCR GPR119 has a similar effect.[49] These latter findings raise the provocative notion that there is an aspect of incretin biology that helps to defend against hypoglycemia, perhaps acting in a physiological context in the descending arm of plasma glucose during postprandial metabolism by heightening the sensitivity of an α cell response.

It has been estimated that, in healthy individuals, the actions of these incretin hormone account for more than half of the prandial increase in insulin secretion. GLP-1 also has receptors in the CNS and a role in governing satiety and appetite; whether GIP has a similar role is less clear. GLP-1 also has effects on gut motility. It suffices to state that the incretin axis, and more broadly the EEC-derived hormones, have a major role in the overall endocrine governance of postprandial metabolism.

KEY POINTS: Entero-endocrine Cells and the Incretin Response

- Enteroendocrine cells (EECs) are distributed throughout the stomach, small intestine, and large intestine in a regulated distribution of pluripotential hormone secreting cells that act as chemosensors of nutrients and nutrient metabolites.
- EECs comprise just 1% of the epithelial cell population of the small intestine, but in aggregate constitute the largest endocrine organ in the body.
- One aspect of the EECs constitutes the incretin axis, composed of two gut hormones, glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like polypeptide-1 (GLP-1), that are secreted in response to meal ingestion.
- In a glucose-dependent manner, GIP and GLP-1 robustly amplify glucose-dependent insulin secretion, likely accounting for 50% of the meal-related increase in insulin secretion. The incretin effects are minor at fasting glucose levels due to low secretion of GIP and GLP-1 and because of their glucose-dependent action on insulin secretion.
- A number of lipid-sensing G protein–coupled receptors modulate GIP and GLP-1 secretion and can also modulate insulin secretion directly via receptor expression on β cells.

Effects on Energy Expenditure and Respiratory Quotient

During prandial conditions, extending into the postprandial period, there is a modest increase in systemic energy expenditure. This is termed the thermic effect of food and in aggregate comprises approximately 10% of daily energy expenditure. The processes of ingestion, digestion, and absorption, as well as downstream metabolism of nutrients, are considered to account for the increase in energy expenditure that comprises the thermic effect. The magnitude of this effect is partially dependent on the type of nutrient consumed, being less for carbohydrate than for protein or fat.

In addition to the thermic effect of food ingestion, there is characteristically a robust effect of meal consumption on the RQ, the quotient of CO_2 production to O_2 consumption, described earlier, and a parameter reflecting relative contributions of carbohydrate versus fat combustion. And, as earlier noted, a typical fasting RQ in a lean, healthy individual is approximately 0.80 ± 0.03, and this can characteristically increase to a value approaching or achieving 1.0 (signifying complete reliance upon glucose oxidation) within an hour of eating, especially if carbohydrate comprise 50% or more of the calories consumed.[7] This indicates that, in the early postprandial period, glucose oxidation nearly completely accounts for energy expenditure. As the postprandial period continues, the RQ value gradually recedes toward a resumption of postabsorptive values. Peak rates of glucose oxidation after a meal approach 3 mg/kg-min, or nearly 50% of the peak rate of systemic glucose appearance (~6–8 mg/kg-min) following meal ingestion, which in turn is a 3- to 4-fold elevation compared with fasting HGP rates (2 mg/kg-min).[50,51] In aggregate, across the prandial and postprandial periods, it has been estimated that oxidation of glucose accounts for 40% to 50% of the metabolic disposition of ingested glucose, with the remainder partitioned to "non-oxidative" metabolism, largely glycogen storage in liver and muscle, together with a smaller fraction that is metabolized glycolytically but not oxidized.

Hepatic Glucose Metabolism

In response to the ingestion of a mixed meal or a predominately glucose load, the liver promptly switches from its role during fasting as the primary producer of glucose to one of robust glucose uptake, accruing glycogen while concomitantly suppressing release of glucose. The activation of robust glucose uptake by the liver requires a combination of both hyperinsulinemia and hyperglycemia.[20,52] Experimentally raising insulin while maintaining euglycemic levels of plasma glucose (i.e., 80–100 mg/dL) does not substantially increase glucose uptake by the liver, though these conditions do suppress hepatic glucose production. Similarly, hyperglycemia by itself (e.g., experimentally controlling insulin concentration through concomitant somatostatin and exogenous insulin infusion) does not elicit strong glucose uptake by the liver, though this suppresses hepatic glycogenolysis. Yet, the combination of hyperinsulinemia and hyperglycemia is a strong stimulus for hepatic glucose uptake. It is important in this context to note that, during the absorptive phase of a meal, portal concentrations of glucose exceed those in the periphery, and this acts to potentiate the effect of the portal to peripheral gradient for insulin in stimulating hepatic glucose uptake. It has been further postulated that a portal to arterial glucose gradient generates a portal signal amplifying hepatic uptake of glucose; the exact nature of the putative portal signal has not been firmly identified.[53]

Following a 12-hour or longer postabsorptive phase, a single meal does not fully replenish hepatic glycogen, but, incrementally, over the course of two to three meals in a day, hepatic glycogen becomes fully replete, with a content of approximately 70 to 100 grams, dry weight (Table 22.1). Both increased intracellular glucose availability and insulin signaling activate the enzyme glycogen synthase in hepatocytes while reciprocally inhibiting glycogen phosphorylase; both enzymes are governed by phosphorylation and dephosphorylation cascades previously described. Because the hepatocyte glucose transporter, GLUT2, enables rapid exchange of glucose across the full range of glucose concentrations, glucose transport is not a rate-limiting step *per se* for rates of glycogen synthesis in the liver, but rather metabolic control is attributed to downstream steps, notably glucokinase and glycogen synthesis. This differs from skeletal muscle, as will be addressed later.

Overall, it is estimated that glucose uptake by the liver accounts for approximately one third of the disposition of ingested glucose. This alone is a sizeable contribution, but by itself is still an underestimation of how the prandial transitions of hepatic glucose metabolism shape glucose tolerance. Suppression of hepatic glucose production during prandial and postprandial conditions is the additional crucial determinant of glucose tolerance and indirectly governs disposition of the portion of ingested glucose that is not taken up by the liver and instead enters the peripheral circulation. Suppression of HGP occurs promptly during the prandial and postprandial phases; HGP reaches a nadir within an hour of eating, and there is a gradual return to fasting rates of HGP over 4 to 6 hours following meal ingestion.[50,51] Regarding governance of HGP suppression, hyperinsulinemia has a strong effect to suppress glycogenolysis, as noted earlier. Interestingly, gluconeogenesis is reduced but not nearly as completely suppressed as glycogenolysis. However, the output of newly synthesized glucose is redirected toward glycogen synthesis rather than as release from the hepatocyte into circulation. This contribution to glycogen synthesis, derived from glucose created by gluconeogenesis, is termed the "indirect pathway" of glycogen formation to distinguish it from the direct pathway based on hepatic glucose uptake and direct incorporation of glucose-6-phosphate into the glycogen synthesis pathway. The indirect pathway of glycogen formation is a meaningful component of glycogen repletion during postprandial conditions.

Postprandial Peripheral and Central Nervous System Glucose Utilization

The patterns of peripheral tissue utilization of glucose divide into those that are insulin-independent and those that are insulin-dependent. CNS consumption of glucose, as described earlier, is insulin-independent and continues unchecked during fed conditions, and, with a low-Km glucose transporter (i.e., GLUT3), this rate does not increase due to hyperglycemia. Nonetheless, the CNS does make an important contribution to the disposition of ingested glucose, accounting for an estimated 25% or more of the ingested amount. This partitioning of ingested glucose to the CNS is an "indirect effect" consequent to, and dependent upon, suppression of HGP and the replacement of circulating glucose with that absorbed from the meal. This principle of ingested glucose substituting for reduced HGP pertains to other tissues in accounting for a relatively rapid distribution and uptake of ingested glucose.

There is also a potentiating effect on glucose utilization caused by a reduced availability of plasma fatty acids. With the sharp increase in insulin during feeding, there is robust suppression of lipolysis in adipose tissue and a consequent sharp reduction in plasma fatty acids. Though ingested fat packaged into chylomicron delivers fatty acids to peripheral tissue (as will be discussed later), the temporal delay in entry of chylomicrons relative to ingested glucose creates a situation such that, in the early phase of postprandial metabolism (for approximately the first 2 hours), glucose is the more predominately available substrate. Thus, the many tissue and organs that have capacity to take up and utilize fatty acids have diminished access to these, and alternatively rely more fully upon circulating glucose. The increased glucose flux during postprandial conditions is associated with a postprandial rise in circulating lactate levels, an approximate doubling, which provides gluconeogenic substrate for the indirect pathway of hepatic glycogen synthesis described above.

In the preceding section on postabsorptive metabolism, skeletal muscle was highlighted as an exemplar of fatty acid uptake and oxidation and equally, but oppositely, during postprandial metabolism, skeletal muscle is an exemplar of insulin-dependent glucose uptake. Glucose transport across the sarcolemma is regarded as the rate-limiting step for insulin-stimulated glucose utilization by muscle[54,55]; physiological levels of hyperinsulinemia increase the kinetics of glucose transport in muscle by an order of magnitude compared with postabsorptive conditions. The postprandial increase in plasma insulin is sufficient to initiate the signal transduction cascade that greatly increases glucose transport capacity, and this stems from the molecular physiology of GLUT4, the glucose transporter predominant in skeletal muscle. Under postabsorptive conditions, GLUT4 is mostly contained within cytosolic vesicles, sequestered away from the sarcolemma. The insulin signaling cascade triggers translocation of GLUT4 to insertion within the sarcolemma, where it can manifest functionality to enable bidirectional facilitated transport of glucose.[6] Across a limb, the postprandial fractional extraction of glucose increases nearly 20-fold or more compared with postabsorptive conditions. Also, under resting conditions, not all capillaries in skeletal muscle are patent and receive equal perfusion: a sizeable fraction is closed at any instance during resting conditions, and in response to increased insulin a process of capillary recruitment and increased perfusion occurs.[56] Taken together, the enhanced delivery and distribution of insulin and glucose into the interstitial space of skeletal muscle potentiate glucose transport, the rate-limiting step for glucose metabolism in muscle. It is estimated that skeletal muscle accounts for the disposition of approximately one third of ingested glucose during postprandial conditions. Thus, the trilogy of brain, liver, and muscle accounts for much of the disposition, approximately 90%, of ingested glucose.

It is estimated that glycogen formation is the metabolic fate that accounts for approximately half of postprandial glucose uptake into muscle.[50,57] The role of insulin signaling in myocytes largely parallels that outlined for hepatocytes in activating glycogen synthesis, albeit without contribution from the indirect pathway utilizing gluconeogenesis that does not occur in muscle. Glucose oxidation accounts for approximately 40% of insulin-stimulated glucose uptake, and the remainder, less than 10%, is released from muscle as pyruvate, lactate, or alanine, thus continuing or increasing the Cori cycle. Like the rise in systemic RQ during feeding, in skeletal muscle regional indirect calorimetry reveals a sharp rise in RQ, and it is the rise in muscle RQ that in good measure underlies the rise of systemic RQ.[50] Cardiac muscle during postprandial conditions manifests a metabolic response similar as that of skeletal muscle, increasing glucose uptake and oxidation and increasing glycogen formation, while reducing reliance upon fatty acid oxidation.

The decrease in fat oxidation is also mediated by a suppression of lipolysis and reduced availability of plasma fatty acid, and intracellularly by generation of malonyl CoA from increased glycolytic flux, thus inhibiting CPT-1, crucial for translocation of acyl-CoA into mitochondria. The importance of substrate competition in governing glucose versus fatty acid metabolism, an important concept of intermediary metabolism, was initially articulated by Randle.[58] This substrate competition can be empirically demonstrated in humans as well as in *in vitro* laboratory studies. Experimentally maintaining fasting levels of plasma fatty acids (by lipid infusion) even while raising insulin into the upper physiological range impairs insulin-stimulated glucose uptake into muscle and lessens suppression of HGP.[9,59] In fact, the acutely induced pattern of insulin resistance caused by fatty acids under these experimental conditions bears a strong resemblance to that observed in obesity and type 2 diabetes mellitus and has been a strong impetus for deeper inquiry into the "lipotoxic" contributions to insulin resistance and disease pathogenesis.[60] Indeed, in a provocative and classic essay it was argued that, if not for the ease of the water-soluble chemistry of glucose relative to the complexity of lipid analyses, emphasis may not have been placed so strongly on perturbed glucose metabolism in studying diabetes, and instead more focus may have been allocated to perturbation of fatty acid metabolism in the pathogenesis of diabetes.[61]

The shift from a reliance upon fat oxidation during postabsorptive conditions to strong reliance upon glucose oxidation during postprandial conditions has been termed "metabolic flexibility,"[8,62] and is a prominent aspect or manifestation of healthy metabolic transition, especially for postabsorptive to postprandial transitions. Earlier, the substrate signaling role of malonyl-CoA was addressed in mediating the transition from fatty acid to glucose oxidation.[63] Further, hyperglycemia itself, independent of a concomitant rise in insulin, can evoke a rise in carbohydrate oxidation and a corresponding reduction in fat oxidation, systemically and within skeletal muscle.[64,65] Thus, the concept of substrate competition is bidirectional, with fatty acids impeding insulin-stimulated glucose metabolism and hyperglycemia perturbing postabsorptive patterns of fatty metabolism. This is an important concept within regulation of intermediary metabolism and is relevant to metabolic diseases like type 2 diabetes mellitus.

KEY POINTS: Postabsorptive Metabolism

- The relative stability and constancy of plasma glucose and fatty acid concentrations derives from tightly governed rates of appearance, mobilized from endogenous sources and rates of disappearance from plasma into tissues.
- Endocrine–substrate feedback loops are a primary mechanism for control of glucose and fatty acid rates of appearance, whereas rates of glucose and fatty acid utilization are concentration-driven and governed by energy requirements.
- There is a relative parity in oxidation of fatty acids and glucose during postabsorptive metabolism, the latter occurring mainly in the CNS, but the former broadly distributed across peripheral tissues.

Postprandial Adipose Tissue Metabolism

Glucose uptake into adipose tissues increases during postprandial conditions of hyperglycemia and elevated insulin. Like skeletal muscle, adipose tissue expresses the GLUT4 transporter isoform that undergoes translocation to the cell surface under insulin stimulation. However, it has been challenging to quantify rates of glucose uptake into adipose tissue, at least with the same clarity that has been achieved for skeletal muscle; arteriovenous balance studies are technically difficult due to the small vessel diameters in adipose, microdialysis has provided good insight, and, more recently, metabolic imaging approaches have yielded quantitative estimates.[66,67] The emerging consensus appears to be that adipose tissue may account for approximately 5% to 10% of glucose utilization during postprandial conditions. However, adipose tissue has a relatively low rate of oxygen consumption and contains minor amounts of glycogen, and the chief metabolic fate of glucose taken up by adipose is glycolysis, with the generation of G-3P to serve as a backbone for TG synthesis. Insulin-stimulated glucose uptake into adipose can properly be regarded as providing a requisite foundation for the crucial role of adipose in postprandial disposition of ingested TG. There is also efflux of lactate from adipose into venous circulation, which continues during postprandial metabolism, indeed at rates that exceed postabsorptive conditions. Special notation should be made of brown adipose tissue (BAT) and its capacity for substrate utilization, which considerably exceeds that of white adipose tissue when considered on a per cell or per unit weight basis. Indeed, while skepticism had once seemed to prevail about the existence of meaningful amounts of BAT depots in humans, metabolic imaging with glucose analogs convincingly demonstrated both the existence and heightened metabolic rates of BAT.[68] The postprandial response of BAT is a contributor to the aggregate thermic effect of food in keeping with the uncoupling of BAT mitochondria and this tissue's thermogenic function.

During the postprandial phase there is strong insulin-induced suppression of lipolysis and a shift toward TG storage. The drop in plasma fatty acids and glycerol is evident promptly in response to feeding, reaches a nadir at approximately 1 hour after the start of eating, and persists, only gradually returning to usual postabsorptive levels over the duration of the postprandial period. This is a pattern that parallels that for suppression of HGP and is evocative of the single-gateway hypothesis of insulin governance of lipolysis and HGP that was cited earlier. However, quantitatively, the most important aspect of adipose metabolism following meal ingestion concerns lipoprotein metabolism. Following fat ingestion, chylomicrons enter systemic circulation (via lymphatic drainage from splanchnic tissues), beginning to appear an hour or so after the meal, reaching a peak several hours after meal ingestion and remaining elevated for up to 6 to 8 hours if the amount of fat ingestion was relatively large. Chylomicrons are too large to exit capillaries, yet adipose tissue has a major role in the metabolism of chylomicrons, through its expression, synthesis, and release of lipoprotein lipase (LpL).[12] LpL released from adipocytes binds to the luminal surface of adjacent capillaries. The release of LpL from adipocytes is stimulated by insulin, and accordingly recedes during postabsorptive conditions. Thus, the molecular physiology of insulin action in adipose tissue may be construed as establishing a "first-pass" capacity for the partitioning of ingested fat toward this tissue by its catabolic effect on chylomicron TG. The armatures of LpL enable engagement with chylomicrons and catalyze hydrolysis of TGs into fatty acids that can be taken up by adipocytes, where, in conjunction with G-3P generated by glycolysis metabolism, provide substrate for esterification into TG and storage in the adipocyte LD.

Though there is a potential for spillover into systemic circulation of fatty acids liberated by LpL, the amount is relatively minor, denoting an efficient capacity for fatty acid uptake in insulin-stimulated adipocytes. Nonetheless, catabolism of chylomicrons is only partially completed during a passage through adipose capillaries, and the resulting remnant chylomicrons, now of diminished diameter and with a change in composition (decreased ratio of TG to phospholipid), are further acted upon in the liver (as discussed later) and in subsequent passages through adipose and other tissues. Skeletal muscle, heart, and other tissues express LpL. However, in skeletal muscle, LpL activity generally decreases in response to increased insulin and is higher during fasting conditions, thus directed toward VLDL rather than chylomicron catabolism.

KEY POINTS: Tissue Distribution of Ingested Glucose During Postprandial Metabolism

- The combination of hyperglycemia and hyperinsulinemia, accentuated within portal circulation relative to the peripheral circulation, triggers a shift in the liver from glucose export to robust glucose uptake, with a predominant metabolic fate being repletion of hepatic glycogen. Liver accounts for approximately one third of the disposition of ingested glucose.
- The strong, rapid, and sustained suppression of HGP after meal ingestion effectively substitutes ingested glucose for that of endogenous glucose as the source for CNS glucose oxidation, accounting for approximately one quarter of the disposition of ingested glucose.
- The prandial rise in insulin signals recruitment of GLUT4 to the cell surface in muscle and adipose. Glucose uptake by muscle accounts for a third or more of ingested glucose disposition, and the metabolic fate of this is directed toward glycogen formation in muscle and a marked shift from fat to glucose oxidation.
- Adipose glucose uptake may account for a tenth of glucose uptake during postprandial conditions and has a key role in providing generation of glycerol-3-phosphate to provide the backbone for triglyceride formation.
- Overall, roughly half of glucose metabolism during postprandial conditions is consumed in glucose oxidation, and the remainder is stored as glycogen, including a large portion of the gluconeogenic flux occurring in liver, the indirect pathway for glycogen synthesis.

Hepatic Metabolism of Fatty Acids and Lipoproteins

In considering the contributions of various organs and tissues to postprandial metabolism, we began with outlining hepatic glucose uptake and glycogen repletion, and, in beginning to conclude this section, a turn to another aspect of the postprandial hepatic response is warranted. The liver is the main metabolic "clearing house," and hepatic metabolism of remnant chylomicrons is an important aspect of postprandial intermediary metabolism and lipoprotein metabolism, one that is crucial for removing chylomicron remnants and restoring the predominance of VLDL particles as TG-rich lipoproteins during postabsorptive conditions. As noted earlier, newly formed chylomicrons are too large to traverse the 100-nm fenestrations of hepatic capillaries, but, in a reduced size after partial hydrolysis by adipose LpL, remnant chylomicrons (now approximately 75 nm in diameter) can be metabolized in the liver and undergo receptor-mediated uptake into hepatocytes for degradation. Some of the lipid components are recycled into newly formed VLDL, and a proportion enter other lipid metabolism pathways, including sphingolipid and phospholipid pathways and TG storage in hepatic LDs, and a portion undergo oxidation. Thus, there is a complex map of potential lipid metabolism pathways that operate in hepatocytes and that are important in completing the transition from postprandial to postabsorptive metabolism.[12]

The role of the liver in ketogenesis was addressed in the section on postabsorptive metabolism, as well as how ketogenesis increases as fasting is extended in duration. Conversely, during postprandial hepatic metabolism the pathway of *de novo* lipogenesis (DNL) can increase if the carbohydrate load is large and is consumed in conjunction with high caloric intake.[69] Attention has increasingly turned to the role of DNL in the pathogenesis of hepatic steatosis associated with obesity and insulin resistance, termed nonalcoholic fatty liver disease. When there is a surfeit of pyruvate-derived acetyl-CoA, sufficient to drive formation of mitochondrial citrate and its export from the mitochondria to the cytosol, then the enzyme acetyl-CoA carboxylase, under the influence of hyperinsulinemia, catalyzes the formation of malonyl-CoA. An important role for malonyl-CoA as an allosteric inhibitor of CPT-1 and fatty acid translocation into the mitochondria has been described earlier. In addition, the enzyme fatty acid synthase carries out DNL by repetitively using malonyl-CoA to construct fatty acids, most commonly C-16 palmitate. Yet another important contribution of malonyl-CoA is to provide substrate for fatty acid elongation to form very long chain fatty acids (i.e., chain length >22 carbons). Very long chain fatty acids serve specialized functions in membrane formation, and long chain fatty acids, particularly polyunsaturated fatty acids, can have a signaling function in transcriptional control of hepatic lipid metabolism.

Methodologic Approaches to Investigating Postprandial Conditions

Methods for assessing postabsorptive metabolism, notably isotope dilution flux studies and indirect calorimetry, were discussed in a previous section, and here attention is turned to investigations of postprandial intermediary metabolism. Descriptive studies are highly informative and of practical clinical utility, and are commonly done with measurements of metabolites and hormones following glucose challenges and mixed-composition meal ingestion. Flux studies of glucose metabolism can be integrated into these by including labeled glucose in the ingested load (to trace its entry into peripheral circulation and estimate its uptake in the splanchnic tissues, notably liver), while concomitantly infusing a separately labeled glucose isotope to enable determination of HGP and overall glucose rate of disappearance. Several citations that utilize this approach have been provided.[50,51] Also, indirect calorimetry can reveal patterns and rates of fatty acid and glucose oxidation. Measurements of arteriovenous differences provide further insights into the contributions of various tissue beds, notably for skeletal muscle, and with some application to the splanchnic bed and adipose tissues. However, there remains a main limitation to these approaches regarding measuring insulin sensitivity. The pronounced dynamic changes in glucose, fatty acids, and insulin concentrations, together with interindividual differences in respective levels, makes it quite challenging to "isolate" how efficiently (or inefficiently) insulin stimulates metabolism. Mathematical modeling approaches that have used the data from an oral glucose challenge or mixed meal challenge and consider insulin secretory rates have derived estimations of insulin action.[70] From these parameters a disposition index can be determined that capsulizes the integrated efficiency of glucose disposition.

A paradigm shift in quantifying response to insulin was the development of an experimental platform widely termed the "glucose clamp." Initially designed by Andres and colleagues,[71] and later refined, modified, and evolved by countless investigators, the core concept is to attempt to hold fixed (i.e., steady state) concentrations of insulin and plasma glucose, achieved by constant infusion of insulin and an adjustable infusion of glucose, adjusting the rate of glucose infusion based on frequent measurements of plasma glucose to hold steady or "clamped" blood glucose concentrations.[72] The key outcome is the steady-state rate of glucose infusion, which, in the context of stable insulin and glucose concentrations, provides a quantitative index of insulin sensitivity in stimulating glucose utilization. This allows comparison across individuals and groups of a parameter of insulin action. This platform can also employ isotope labelled glucose infusion to estimate the effect of insulin on HGP, can employ indirect calorimetry to study oxidative metabolism, and study effect on lipolysis and fatty acid metabolism, as well as other parameters. So, in many respects, a glucose clamp recapitulates a metabolic context analogous to postprandial conditions, but of course is not a full recapitulation, because it experimentally fixes glucose and insulin concentrations rather than allowing the dynamic changes that characterize postprandial intermediary metabolism. Moreover, the peripheral infusion of insulin (and of glucose) does not recreate the portal to peripheral gradients that are crucial factors in shaping postprandial hepatic metabolism.

Perhaps the greatest utility of the glucose clamp approach has been to enable studies of insulin-stimulated skeletal muscle metabolism. Under the conditions of hyperinsulinemia, and while maintaining euglycemia with the clamp method, skeletal muscle accounts for nearly 70% to 80% of overall glucose utilization. Also, the metabolic fate of glucose utilization during euglycemia clamp conditions fairly closely mirrors that observed under postprandial conditions: half of more of glucose taken up is stored as glycogen, a large fraction is immediately oxidized, and a smaller fraction undergoes glycolysis and release as lactate and related trioses. Arteriovenous limb balance studies can provide additional clarity or specificity regarding skeletal muscle metabolism.[9,57] Another methodological advance has been to use the glucose clamp as a physiologic platform upon which concomitant imaging can be done. One prominent example has been to use magnetic resonance spectroscopy to image the accretion of glycogen in skeletal muscle or liver.[73] Another approach, cited earlier, has been to use positron emission tomography in conjunction with glucose clamps and specific positron-labeled isotopes to measure the kinetic steps of glucose delivery, transport, and phosphorylation in skeletal muscle and to image glucose metabolism in adipose tissues.[66,67]

PROTEIN AND AMINO ACID METABOLISM

The 20 amino acids that are the building blocks underlying protein metabolism are an important aspect of intermediary metabolism. However, the details of the metabolic pathway of each amino acid (or class of amino acids) is a diverse topic, beyond the scope of this chapter, and these pathways are well described.[74] Instead, in this section, the interface or interdigitation of amino acids with carbohydrate and fatty acid metabolism will be addressed, notably when amino acid metabolism yields precursors for gluconeogenesis and under the metabolic conditions when amino acids are consumed as substrates for energy production. However, unlike carbohydrate and lipid intermediary metabolism, for which arguably the primary purpose is combustion for energy production, the main purpose of amino acid metabolism is to conserve the overall mass of proteins and support continued turnover or replacement of proteins to ensure intact functional capacity is maintained, and concepts on protein turnover (degradation and synthesis) will be addressed.

Collectively, amino acids circulate in plasma at an approximate concentration ranging between 35 to 65 mg/dL, divisible roughly evenly across the 20 constituents,[43] with some like glutamate and alanine trending toward higher concentrations, as these are integral to nitrogen exchange between tissues. There are active transport mechanisms involved both in the absorption of individual amino acids from the intestinal tract and from plasma into cells. To support protein synthesis the intracellular pool of amino acids must also reflect the entirety of the amino acid alphabet, including both essential and nonessential amino acids. Thus, part of the intermediary metabolism of amino acids, including crucial transamination pathways, is to ensure appropriate proportionality and availability across the amino acid pool, compensating or adjusting for dietary variances, and thereby enabling protein synthesis. The essential amino acids are the ten amino acids for which endogenous synthetic pathways are mostly insufficient to maintain adequate availability, and thus, there is an *essential* reliance on dietary availability. It is in this context of meeting the dietary requirements for essential amino acids, recommendations are for roughly 60 g of protein consumption daily. Though a daily consumption exceeding 100 g is quite common, and it is with these higher levels of consumption that partitioning amino acids toward oxidation and for gluconeogenesis (or even lipogenesis) tends to be a more prominent facet of postabsorptive and postprandial intermediary metabolism.

Upon absorption from the small intestine, amino acids enter the portal venous circulation, a process that tends to lag temporally behind that of glucose. There is a high splanchnic and hepatic fractional extraction of amino acids. These are both sites of relatively high protein turnover, in contrast to the large amount of protein in skeletal muscle, that also turns over but at a significantly slower rate, as does collagen in bone and other tissues. Enterocytes that cover the extensive intestinal epithelial surface themselves have a cell cycle of only several days, and this rapid rate of cellular turnover, together with their rich complement of transporters and enzymes, constitutes a high need for protein replacement; it has been estimated that splanchnic protein turnover is nearly 50% per day. The liver produces circulating plasma proteins (e.g., albumin, clotting factors, transporter proteins), as well as its own abundant enzymatic capacities, and these have a relatively rapid turnover. Albumin, as an example of a hepatic-synthesized protein, has a turnover of approximately 20% per day, while that of immunoglobins is not too dissimilar. In contrast, the turnover of skeletal muscle protein is estimated to be 1% to 2% per day. These are sharp contrasts, essentially dividing the patterns of protein turnover into categories of a fast-turnover (e.g., hepatic, and splanchnic) and a slow-turnover pool.[1]

There is not a dedicated storage form for amino acids, akin to TG for fatty acids or glycogen for glucose, that sequesters amino acids and can then later be mobilized. Instead, proteins are synthesized to fulfill specific essential functions and in accordance with turnover rates for that specific protein. In this sense, the overall mass of protein in the body, approximately 20% of body weight (Table 22.1), is a reserve of amino acids and indeed supplies nearly three quarters of the amino acids utilized in daily patterns of protein breakdown and synthesis, estimated at 300 to 400 g daily, with the remaining one quarter of amino acids utilized daily coming from food ingestion.[10] Excess amino acids beyond immediate needs for protein synthesis must be otherwise metabolized. To enter these pathways, the amino group is deaminated, and ammonia nitrogen is transferred to an α-keto acid, most typically α-ketogluturate. The resulting deaminated product is metabolized to acetyl-CoA, which can then enter oxidative metabolism to yield energy or serve as a substrate for conversion to newly synthesized glucose or to support *de novo* lipogenesis. This exchange of nitrogen carried out by transamination is quite vigorous.[74] In a seminal series of experiments, Schoenheimer and colleagues conducted labeling studies with the stable isotope ^{15}N and observed that, if the label were administered contained in only a single protein or compound, the isotope ^{15}N would nonetheless rapidly and extensively appear in numerous other proteins and amino acids, denoting rapid exchanges within the nitrogen pool of amino acids.[75] There is also an overriding tightly governed balance of nitrogen and an excess of amino acids, created either by higher rates of protein degradation or, more commonly, by ingestion of protein in surplus of that needed for protein balance, lead to higher nitrogen excretion and diversion of the resultant carbon backbone to other pathways. Urea is the pathway for disposal of the ammonia liberated by transamination that accounts for 80% of nitrogen excretion, the high solubility of urea in water making this efficient for nitrogen excretion. Urea synthesis occurs in the liver, and its excretion occurs of course in the kidney. In skeletal muscle, nitrogen liberated by transamination of amino acids can be carried to the liver via the previously cited Cahill cycle, as pyruvate can accept the nitrogen to form alanine, thereby carrying both a triose carbon backbone for gluconeogenesis and shuttling the nitrogen for urea synthesis. Another fate for the α-ketogluturate that is generated from an amino acid undergoing deamination is to be oxidized. An estimate of the contribution of protein to substrate oxidation can be based on the amount of urinary nitrogen. Approximately 1 g of urinary nitrogen arises from the oxidation of approximately 6.25 g of protein oxidation.[5] On a protein intake of approximately 1 g per kg body weight, typical of a modern diet, nitrogen excretion is approximately 10 g per day, representing approximately 60 g of protein oxidation daily. This compares with an average daily combustion by the brain alone of 120 g of glucose but is about on par with a combustion of about 60 to 70 g of fat daily, though the energy yield of fat is more than twice that of protein.

Altogether it is estimated that there is approximately 300 to 400 g of protein turnover daily, thus greatly exceeding the 60 to 100 g of daily protein ingestion.[1,10] Yet, despite this relatively large daily turnover of protein, the protein mass of the body remains highly conserved and is large, comprising 20% of body weight; viewed from another perspective, it comprises approximately three quarters of the solid mass of the body (this of course excludes the 60% of weight attributable to water and that due to anhydrous fat). For instance, skeletal muscle comprises approximately 40% of weight in a lean adult, and approximately 20% of muscle weight is represented by protein, and thus 5 to 6 kg of protein in muscle in a 70-kg adult. At the 1% turnover rate, it can be estimated that approximately 50 to 60 mg of skeletal muscle protein is degraded and synthesized in a balanced homeostasis each day. The role of exercise in influencing contractile protein turnover in muscle will be addressed later. Endocrine regulators of protein synthesis are insulin, growth hormone, and testosterone, which have an anabolic effect, whereas cortisol has a catabolic effect. Also of obvious importance for protein synthesis is a high availability of amino acids, and so in this sense, protein synthesis can be generally regarded within the panoply of postprandial anabolism. Exercise can stimulate protein synthesis in skeletal muscle (and in the heart), especially for several hours in the postexercise period when a previous bout of exercise can augment the effect of food and protein ingestion to stimulate muscle protein synthesis. As well, exercise can shape the remodeling adaptation of a specific muscle to adjust to repeated exposure to an exercise, and these effects are more pronounced in response to resistance types of exercise.

KEY POINTS: Protein and Amino Acid Metabolism

- Key function of amino acid metabolism is to ensure appropriate availability across the amino acid pool to facilitate protein synthesis, compensating or adjusting for dietary variances.
- Essential amino acids are designated as the ten amino acids for which there is an *essential* reliance on dietary availability.
- Protein stores in the body supply most of the 300-400 g of amino acids required for daily protein breakdown and synthesis, the remaining amino acids are supplied from ingested food.
- The liver produces circulating plasma proteins.
- Amino acids not utilized for protein synthesis are deaminated to yield products for oxidative metabolism, gluconeogenesis or de novo lipogenesis.

EXERCISE METABOLISM

Exercise Dramatically Increases Demand For Energy

Exercise and other forms of physical activity can dramatically increase energy expenditure, and these robust demands for energy mobilization and utilization involve unique metabolic regulation that require hormonal control as well as hormone-independent mechanisms (Fig. 22.3). The resting metabolic rate or energy expenditure is approximately 5 mL O_2 consumption per kg body weight per minute. During moderate exercise O_2 consumption can increase 10-fold, and in elite athletes O_2 consumption during maximal exercise can be as high as 90 mL per kg body weight per minute, an astounding 18-fold increase from rest. This increase is almost exclusively due to the increase in energy demand by contracting skeletal muscle, and to a lesser extent the rapidly beating heart.

KEY POINTS: Exercise Metabolism

- Increases in epinephrine during moderate to high intensity exercise stimulates glycogen breakdown to supply glucose to contracting skeletal muscle.
- Low circulating insulin and elevated glucagon during exercise help to increase hepatic glucose output to maintain blood glucose levels during exercise.
- Lactate and amino acid production by exercising muscle can be converted back to glucose in the liver where it is subsequently delivered back to exercising muscle via the circulation.
- The need for exogenous insulin and impairments in counterregulatory responses to hypoglycemia in Type 1 diabetes makes the regulation of blood glucose during exercise problematic for these patients.
- Low circulating insulin and elevated glucagon during exercise act to stimulate lipolysis of triglycerides within adipose tissue to increases fatty acids in the circulation for delivery to exercising muscle.
- Elevated epinephrine during moderate to high intensity exercise stimulates hydrolysis of triglycerides stored within skeletal muscle to supply fatty acids to contracting skeletal muscle.
- Exercise training increases the availability of carbohydrate and lipid fuels and the metabolic machinery to store, mobilize and utilize these fuels to support the energy demands of exercise. Exercise training also shifts fuel preference towards fatty acid oxidation in the postabsorptive (fasting) state and increases insulin sensitivity in the postprandial (fed) state.

Energy Metabolism During Exercise

The control of metabolism to supply energy during exercise involves most of the same systems of hormonal regulation described for postabsorptive metabolism, but often on a different level of magnitude and also involving unique mechanisms, some of which are only beginning to be understood. The increase in energy expenditure during exercise is supported by multiple systems in skeletal muscle. During high intensity exercise lasting up to several seconds, such as sprinting or weightlifting, stored phosphocreatine supplies energy in the form of ATP catalyzed by the creatine kinase (CK) enzymes. An additional function of CK is thought to be a buffer against increasing adenine diphosphate (ADP) concentrations, because ADP is a key regulator of many enzymatic reactions in the cell. CK is highly abundant in skeletal and cardiac muscle, and clinically, elevated CK levels in blood are used to diagnose damaged cardiac muscle (myocardial infarction) and skeletal muscle disease (muscular dystrophy).

Glucose Metabolism During Exercise

During exercise of lower to moderate intensity, energy can be provided by both glucose and fatty acids. Exercise of moderate to relatively high intensity relies predominantly on glycolytic or anaerobic metabolism, predominantly through the breakdown of glycogen stores in muscle.[76] This rapid increase in glycogen breakdown is regulated by mechanisms involving catecholamine stimulation of the adenylate cyclase reaction to ultimately activate glycogen phosphorylase, along with endogenous calcium activation released through muscle contraction.[77] Glycogen stores in skeletal muscle (Fig. 22.4) are limited to approxiamtely 400 g (1600 Kcal) (Table 22.1) and can be a limiting factor in marathon running or other long-duration endurance sports. Fortunately, other fuel sources can supply energy for exercise. The liver is remarkable in its ability to increase HGP to maintain blood glucose levels even during strenuous exercise. Even during moderate intensity exercise there are increases in glucagon and catecholamines, and a concomitant decrease in insulin secretion, that in concert enable increased HGP; HGP can increase many fold.[78] This effect on HGP and the accompanying endocrine response is more subtle during mild intensity exercise, such as walking, even when sustained. During a bout of moderate exercise

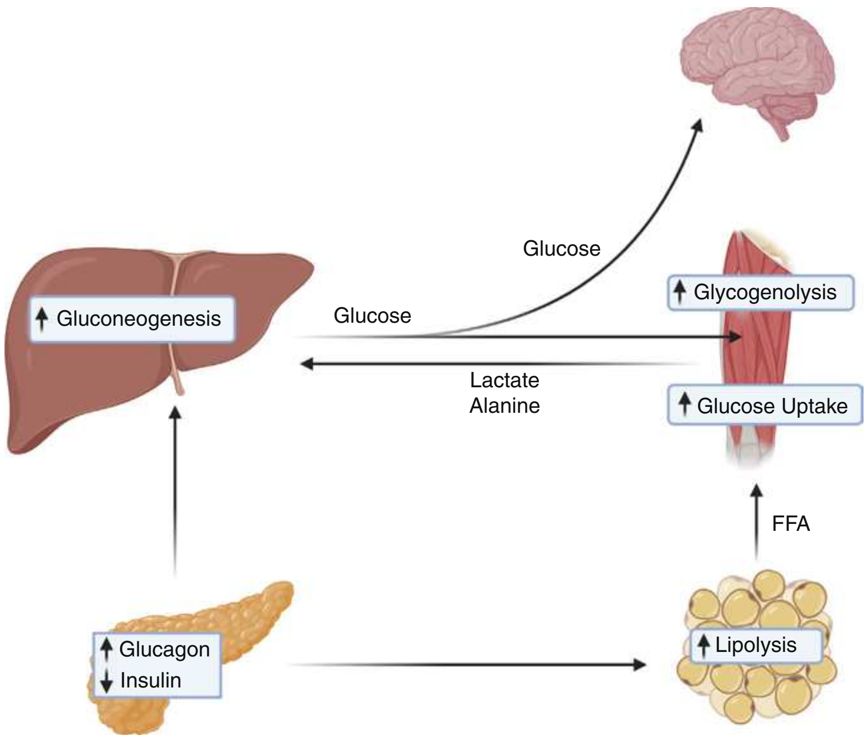


Fig. 22.3 **Exercise metabolism.** Glycogenolysis and insulin-independent glucose uptake are rapidly accelerated during exercise and compensate for the low circulating insulin levels, which also serve to release the suppression of lipolysis within adipose tissue and enable increased HGP. This synergizes with higher epinephrine and norepinephrine, which act to increase triglyceride hydrolysis in adipose tissue and skeletal muscle and increase glycogenolysis in skeletal muscle. The high rates of glucose metabolism result in the release of lactate and alanine, particularly during higher intensity exercise, which are then transported to the liver and converted back to glucose via the Cori and Cahill cycles, all of which markedly increase energy supply to skeletal muscle. A caveat to this description of intermediary metabolism during exercise is that the fuel admixture combusted during physical activity is strong influenced by the intensity of the exercise, the duration of exercise, and the underlying fitness of the individual. In general, lower to midlevel intensity exercise, a longer duration of such exercise, and a higher level of fitness are factors that act to increase reliance upon fatty acid oxidation rather than upon glucose metabolism. *FFA,* Free fatty acid.

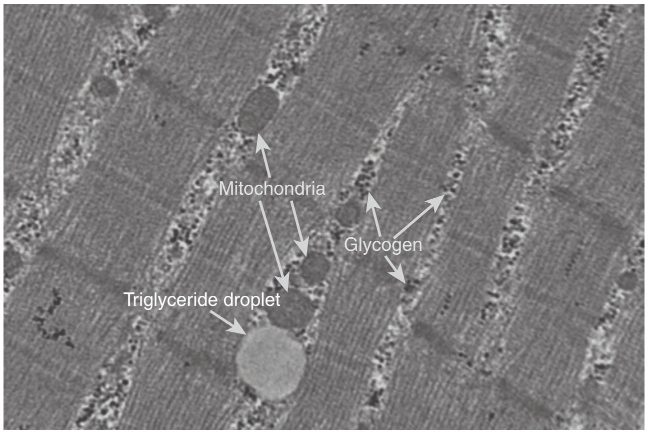


Fig. 22.4 Electron micrograph of human skeletal muscle (vastus lateralis) illustrating fuel stored as glycogen and as triglyceride in lipid droplets, and held in close proximity to mitochondria for efficient oxidative metabolism.

muscle cells take up and utilize glucose from the circulation supplied by the liver. This process is dependent on glucose transport via the GLUT4 glucose transporter, described in the sections on postabsorptive and postprandial metabolism, whereby these GLUT4-containing vesicles are translocated from the cytosol to the sarcoplasmic membrane and T-tubules. The mechanisms governing this GLUT4 translocation and glucose transport differ from that described above for insulin-stimulated glucose transport after feeding. Circulating insulin levels during exercise are quite low, necessitating a separate mechanism involving the signaling molecules AMPK, calcium, and nitric oxide.[79] The fate of this glucose once inside the cell can be complete mitochondrial oxidation or, during higher intensity exercise where oxygen is limiting, the glucose supplies energy through glycolysis but with pyruvate converting to lactate rather than continuing to oxidative pathways. If this high intensity exercise continues, the lactate in muscle is exported into the blood, where it can be produced at a faster rate than it can be cleared by the liver and converted back to glucose via the Cori cycle, thus resulting in an accumulation of lactate in the blood. In this scenario, lactate levels can increase from a resting level of approximately 1 mmol/L to upwards of 15 to 20 mmol/L. It is this acute increase in lactate during exercise with the increase in associated H^+ (acid) that reduces the blood pH significantly and results in muscle fatigue and the need to reduce exercise intensity or to completely stop exercise.

Patients with type 1 diabetes mellitus require insulin for survival, and this need makes the regulation of blood glucose during exercise problematic for these individuals. Because insulin and exercise itself have independent effects on glucose uptake, having an inappropriately higher amount of circulating insulin due to giving too much insulin prior to exercise will effectively add to the independent effects of muscle contraction on glucose uptake and cause excessive decreases in blood glucose, and hypoglycemia may ensue. Also, a relative excess of circulating insulin impedes the increase of HGP. Furthermore, the counterregulatory responses to hypoglycemia are impaired in type 1 diabetes mellitus, i.e., increases in catecholamines and glucagon.[80] Thus, this hormonal dysregulation and absolute lack of insulin in type 1 diabetes mellitus highlights the independent effects of exercise and insulin on glucose uptake and why having the right balance of exogenous insulin before, during, and after exercise is critical in type 1 diabetes mellitus.

Fatty Acid Metabolism During Exercise

The increase in the demand for energy during exercise can also be supported by fatty acids. The source of these fatty acids can vary according to availability, as well as the intensity and duration of exercise. A considerable portion of the fatty acids comes from the circulation, resulting from their release by adipose tissue TG hydrolysis. Low circulating insulin levels during exercise act to release the brake on lipolysis, thereby allowing more fatty acids to be released into the circulation and taken up by exercising muscle. In addition, epinephrine stimulates lipolysis and hydrolysis of adipose tissue TG by activating HSL as well as TG within circulating VLDL. The fatty acids that are free in circulation are taken up and oxidized by skeletal and cardiac muscle through beta oxidation and mitochondrial oxidative phosphorylation. During low-intensity exercise such as walking, plasma fatty acids comprise the major fuel source for exercising muscle and accordingly, given the enormous fuel reserve of TG in adipose, fuel availability is not a constraining factor, as contrasted with the much more finite reserve of muscle glycogen needed to support high intensity exercise.

Fatty acid oxidation during exercise can also be supported by the hydrolysis of TG stored inside muscle cells themselves (Fig. 22.4). This source of energy can supply approximately 50% of the fatty acid oxidation in exercising muscle.[81] Fig. 22.4 illustrates the location of these TG droplets adjacent to mitochondria. These LDs replete with TG are accompanied by LD proteins, perilipins, as well as lipases that are critical for TG hydrolysis and release of fatty acids close to mitochondria.[82] Given the need for fatty acid transport proteins in the aqueous cellular environment, this subcellular geography is important for the process of fatty acid storage, mobilization, and oxidation during exercise.

Effects of Exercise Training on Intermediary Metabolism

Chronic exercise training results in a plethora of changes and adaptations, which largely act to enhance both the availability of carbohydrate and lipid fuels and the metabolic machinery to store, mobilize, and utilize these fuels to support the energy demands of acute exercise bouts. Exercise training increases insulin sensitivity (see earlier section on postprandial metabolism), which serves to enhance rates of glycogen synthesis, and, along with increased enzymes regulating glycogen storage, culminates in an increase in the amount of glycogen storage. Because low muscle glycogen limits long-duration exercise performance, increasing glycogen storage in muscle can enhance exercise performance.[83] This limitation in muscle glycogen storage and in endogenous glucose production by the liver led to decades of research and billions of dollars spent on sports nutrition products to supplement glucose availability during bouts of exercise.

Exercise training also enhances storage of TGs in skeletal muscle, an important fuel source for exercising muscle. Interestingly, excess storage of muscle TGs in conditions of low physical activity and overnutrition, i.e., obesity, is associated with insulin resistance and increased risk for type 2 diabetes.[84] One of the most important adaptations that can occur with exercise training is the increase in the cellular machinery to mobilize, deliver, transport, and utilize both carbohydrate and fat fuels. Prominent among these adaptations are robust increases in mitochondria content and associated enzymes to enhance capacity for oxidative metabolism. These changes induced by training consequently enhance the capacity to perform exercise longer and at a higher intensity. Moreover, exercise training enhances regulation of both postabsorptive and postprandial metabolism, as discussed earlier, e.g., enhanced insulin sensitivity, fatty acid oxidation, and overall metabolic flexibility, and has a multitude of effects that reduce body fat and reduced risk for type 2 diabetes and cardiovascular diseases.

SUMMARY AND CONCLUDING REMARKS

Intermediary metabolism is a complex constellation of processes, principally enzymatic reactions, by which nutrition is converted into components that can be utilized to produce energy, stored in forms that can later be mobilized, and that can serve to support anabolic actions in the building of cellular and organ structure. In this chapter, emphasis was placed upon the roles of various organs, notably, brain, liver, adipose, and skeletal muscle, and how the integration of, and exchanges between, these individual organs comprise systemic intermediary metabolism. In particular, the concept of interorgan substrate flux was delineated as an integrated physiological aspect of metabolism, and it was emphasized how endocrine governance of intermediary metabolism can be largely understood by examining the effects on modulating rates of glucose and fatty acid flux. Another fulcrum of control in intermediary metabolism that was emphasized was the control of glucose and fat oxidation, how this partitioning is most fundamentally controlled by relative substrate availability and can manifest competition between carbohydrate and lipid metabolism, and how this substrate competition is kept salutary by endocrine actions. Hopefully, it has been useful to frame the delineation of intermediary metabolism in a context of three common daily experiences—fasting, feeding, and physical activity—and explore the deceptively, or seemingly silent, sharp changes that mark transitions from feeding to fasting and then back to fasting, and in response to exercise.

REFERENCES

1. Wolfe RR, Chinkes DL, eds Calculation of substrate kinetics: single-pool model. In *Isotope Tracers in Metabolic Research*. Hoboken: John Wiley & Sons, Inc.; 2005:21–50.
2. Westerterp KR. Control of energy expenditure in humans. In: Feingold KR, et al., ed. *South Dartmouth*. Endotext; 2000.
3. Heymsfield SB, et al. The anatomy of resting energy expenditure: body composition mechanisms. *Eur J Clin Nutr*. 2019;73:166–171.
4. Schutz Y. Respiration chamber calorimetry and doubly labeled water: two complementary aspects of energy expenditure? *Eur J Clin Nutr*. 2018;72:1310–1313.
5. Frayn K. Calculation of substrate oxidation rates in vivo from gaseous exchange. *J Appl Physiol Respirat Environment Exer Physiol*. 1983;55:628–634.
6. Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med*. 2013;34:121–138.
7. Goodpaster B, Sparks LM. Metabolic flexibility in health and disease. *Cell Metab*. 2017;25:1027–1036.
8. Kelley D, Mandarino LJ. Fuel selection in human skeletal muscle in insulin resistance: a reexamination. *Diabetes*. 2000;49:677–683.
9. Kelley D, Mokan M, Simoneau J-A, et al. Interaction between glucose and free fatty acid metabolism in human skeletal muscle. *J Clin Invest*. 1993;92:91–98.
10. Frayn K, Evans R. Integration of carbohydrate, fat and protein metabolism in normal daily life. In: Frayn K, Evans R, eds. *Human Metabolism: A Regulatory Perspective*. Hoboken, NJ: Wiley Blackwell; 2019:204–244.
11. Andres R, Cader G, Zierler KL. The quantitatively minor role of carbohydrate in oxidative metabolism by skeletal muscle in intact man in the basal state. Measurements of oxygen and glucose uptake and carbon dioxide

and lactate production in the forearm. *J Clin Invest*. 1956;35: 671–682.

12. Frayn K, Evans R. Lipoprotein metabolism and atherosclerosis. In: Frayn K, Evans R, eds. *Human Metabolism: A Regulatory Perspective*. Wiley Blackwell; 2019:302–324.
13. Jensen M, Ekberg K, Landau BR. Lipid metabolism during fasting. *Am J Physiol Endocrinol Metab*. 2001;281:E789–E793.
14. Mittendorfer B, Magkos F, Fabbrini E, et al. Relationship between body fat mass and free fatty acid kinetics in men and women. *Obesity*. 2009;17:1872–1877.
15. Jensen M, Haymond MW, Rizza RA, et al. Influence of body fat distribution on free fatty acid metabolism in obesity. *J Clin Invest*. 1989;83:1168–1173.
16. Snider M, McGarry JD, Hanson RW. Lipid metabolism I: synthesis, storage, and utilization of fatty acids and triacylglycerols. In: Devlin T, ed. *Textbook of Biochemistry with Clinical Correlations*. New York, NY: A. John Wiley & Sons, Inc.; 2010.
17. McGarry JD, Foster DW. In support of the roles of malonyl-CoA and carnitine acyltransferase I in the regulation of hepatic fatty acid oxidation and ketogenesis. *J Biol Chem*. 1979;254:8163–8168.
18. Yang A, Mottillo EP. Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. *Biochem J*. 2020;477:985–1008.
19. Wasserman DH. Four grams of glucose. *Am J Physiol Endocrinol Metab*. 2009;296:E11–E21.
20. Petersen M, Vatner DF, Shulman GI. Regulation of hepatic glucose metabolism in health and disease. *Nat Rev Endocrinol*. 2017;13:572–587.
21. Gerich J. The role of the kidney in normal glucose homeostasis and in the hyperglycemia of diabetes mellitus: therapeutic implications. *Diabetic Med*. 2010;27:136–142.
22. Harris R. Carbohydrate metabolism I: major metabolic pathways and their control. In: Devlin T, ed. *Textbook of Biochemistry with Clinical Correlations*. New York, NY: A. John Wiley & Sons, Inc.; 2010.
23. Bergman RN, Iyer MS. Indirect regulation of endogenous glucose production by insulin: the single gateway hypothesis revisited. *Diabetes*. 2017;66:1742–1747.
24. Edgerton D, Lautz M, Scott M, et al. Insulin's direct effects on the liver dominate the control of hepatic glucose production. *J Clin Invest*. 2006;116:521–527.
25. Hatting M, Tavares CDJ, Sharabi K, et al. Insulin regulation of gluconeogenesis. *Ann NY Acad Sci*. 2018;1411:21–35.
26. Sasaki M, Sasako T, Kubota N, et al. Dual regulation of gluconeogenesis by insulin and glucose in the proximal tubules of the kidney. *Diabetes*. 2017;66:2339–2350.
27. Owen O, Felig P, Morgan AP, et al. Liver and kidney metabolism during prolonged starvation. *J Clin Invest*. 1969;48:574–583.
28. Prentki M, Matschinsky FM, Madiraju SRM. Metabolic signaling in fuel-induced insulin secretion. *Cell Metab*. 2013;18:162–185.
29. Asare-Bediako I, et al. Variability of directly measured first-pass hepatic insulin extraction and its association with insulin sensitivity and plasma insulin. *Diabetes*. 2018;67:1495–1503.
30. Bertram R, Satin LS, Sherman AS. Closing in on the mechanisms of pulsatile insulin secretion. *Diabetes*. 2018;67:351–359.
31. Rizza R, Mandarino LJ, Gerich J. Dose-response characteristics for effects of insulin on production and utilization of glucose in man. *Am J Physiol Endocrinol Metab*. 1981;240:E630–E639.
32. Cherrington AD, Lacy WW, Chiasson J-L. Effect of glucagon on glucose production during insulin deficiency in the dog. *J Clin Invest*. 1978;62:664–677.
33. Gerich JE. Glucose counterregulation and its impact on diabetes mellitus. *Diabetes*. 1988;37:1608–1617.
34. Rizza RA, Cryer PE, Gerich JE. Role of glucagon, catecholamines, and growth hormone in human glucose counter-regulation. *J Clin Invest*. 1979;64:62–71.
35. Cryer P. Mechanisms of hypoglycemia-associated autonomic failure and its component syndromes in diabetes. *Diabetes*. 2005;54:3592–3601.
36. Gerich JLM, Noacco C, Karam J, et al. Lack of glucagon response to hypoglycemia in diabetes: evidence for an intrinsic pancreatic alpha-cell defect. *Science*. 1973;182:171–173.
37. MacLean P, Blundell JE, Mennella JA, et al. Biologic control of appetite: a daunting complexity. *Obesity*. 2017;25:S8–S16.
38. Newgard C. Metabolomics and metabolic diseases: where do we stand? *Cell Metab*. 2017;10:43–56.
39. Rhee E, Cheng S, Larson MG, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest*. 2011;121:1402–1411.
40. Helander HF, Fandriks L. Surface area of the digestive tract - revisited. *Scand J Gastroenterol*. 2014;49:681–689.
41. Psichas A, Reimann F, Gribble FM. Gut chemosensing mechanisms. *J Clin Invest*. 2015;125:908–917.
42. Latorre R, Sternini C, De Giorgio R, et al. Enteroendocrine cells: a review of their role in brain-gut communication. *Neurogastroenterol Motil*. 2016;28:620–630.
43. Guyton A, Hall JE. Digestion and absorption in the gastrointestinal tract. In: Guyton A, Hall JE, eds. *Textbook of Medical Physiology*. Philadelphia: Elsevier Saunders, Inc.; 2006:808–818.
44. Ko CW, Qu J, Black DD, et al. Regulation of intestinal lipid metabolism: current concepts and relevance to disease. *Nat Rev Gastroenterol Hepatol*. 2020;17:169–183.
45. Molinaro A, Wahistrom A, Marschall H-U. Role of bile acids in metabolic control. *Trend Endocrinol Metab*. 2018;29:31–41.
46. Pais R, Gribble FM, Reimann F. Stimulation of incretin secreting cells. *Therap Adv Endocrinol Metab*. 2016;7:24–42.
47. Miyauchi S, et al. New frontiers in gut nutrient sensor research: free fatty acid sensing in the gastrointestinal tract. *J Pharmacol Sci*. 2010;112:19–24.
48. Christensen MCS, Sparre-Ulrich AH, Kristensen PL, et al. Glucose-dependent insulinotropic polypeptide augments glucagon responses to hypoglycemia in type 1 diabetes. *Diabetes Care*. 2015;64:72–78.
49. Li NXBS, Kowalski T, Yang L, et al. GPR119 agonism increases glucagon secretion during insulin-induced hypoglycemia. *Diabetes*. 2018;67:1401–1413.
50. Kelley D, Mitrakou A, Marsh H, et al. Skeletal muscle glycolysis, oxidation, and storage of an oral glucose load. *J Clin Invest*. 1988;81:1563–1571.
51. Basu R, DiCamillo B, Toffolo G, et al. Use of a novel triple-tracer approach to assess postprandial glucose metabolism. *Am J Physiol Endocrinol Metab*. 2003;284:E55–E69.
52. Moore MC, et al. Regulation of hepatic glucose uptake and storage in vivo. *Adv Nutr*. 2012;3:286–294.
53. Dicostanzo CA, et al. Role of the hepatic sympathetic nerves in the regulation of net hepatic glucose uptake and the mediation of the portal glucose signal. *Am J Physiol Endocrinol Metab*. 2006;290:E9–E16.
54. Bertoldo A, Pencek RR, Azuma K, et al. Interactions between delivery, transport, and phosphorylation of glucose in governing uptake into human skeletal muscle. *Diabetes*. 2006;55:3028–3037.
55. Pencek R, Bertoldo A, Price J, et al. Dose-responsive insulin regulation of glucose transport in human skeletal muscle. *Am J Physiol Endocrinol Metab*. 2006;290:E1124–E1130.
56. Zhang L, et al. Insulin sensitivity of muscle capillary recruitment in vivo. *Diabetes*. 2004;53:447–453.
57. Kelley D, Reilly JP, Veneman T, et al. Effects of insulin on skeletal muscle glucose storage, oxidation, and glycolysis in humans. *Am J Physiol Endocrinol Metab*. 1990;258:E923–E929.
58. Randle PJ. Regulatory interactions between lipids and carbohydrates: the glucose fatty acid cycle after 35 years. *Diabetes Metab Rev*. 1998;14:263–283.
59. Boden G. Obesity, insulin resistance and free fatty acids. *Curr Opin Endocrinol Diabetes Obes*. 2011;18:139–143.
60. McGarry J. Banting Lecture 2001: Dysregulation of fatty acid metabolism in the etiology of type 2 diabetes. *Diabetes*. 2002;51:7–18.
61. McGarry J. What if Minkowski had been ageusic? An alternative angle on diabetes. *Science*. 1992;258:766–770.
62. Goodpaster BH, Sparks LM. Metabolic flexibility in health and disease. *Cell Metab*. 2017;25:1027–1036.
63. Saha A, Vavvas D, Kurowski TG, et al. Malonyl-CoA regulation in skeletal muscle: its link to cell citrate and the glucose-fatty acid cycle. *Am J Physiol Endocrinol Metab*. 1997;272:E641–E648.
64. Mandarino L, Consoli A, Kelley DE. Differential regulation of intracellular glucose metabolism by glucose and insulin in human muscle. *Am J Physiol (Endocrinol Metab)*. 1993;265:E898–E905.

65. Yki-Jarvinen H, Bogardus C, Howard B. Hyperglycemia stimulates carbohydrate oxidation in humans. *Am J Physiol Endocrinol Metab*. 1987;253:E376–E382.
66. Ng JM, et al. PET imaging reveals distinctive roles for different regional adipose tissue depots in systemic glucose metabolism in nonobese humans. *Am J Physiol Endocrinol Metab*. 2012;303:E1134–E1141.
67. Virtanen KA, et al. Glucose uptake and perfusion in subcutaneous and visceral adipose tissue during insulin stimulation in nonobese and obese humans. *J Clin Endocrinol Metab*. 2002;87:3902–3910.
68. Cypess AM, et al. Identification and importance of brown adipose tissue in adult humans. *N Engl J Med*. 2009;360:1509–1517.
69. Sanders FW, Griffin JL. De novo lipogenesis in the liver in health and disease: more than just a shunting yard for glucose. *Biol Rev Camb Philos Soc*. 2016;91:452–468.
70. Cobelli C, et al. The oral minimal model method. *Diabetes*. 2014;63:1203–1213.
71. Sherwin RS, et al. A model of the kinetics of insulin in man. *J Clin Invest*. 1974;53:1481–1492.
72. DeFronzo R, Tobin JD, Andres R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol Endocrinol Metab*. 1979;237:E214–E223.
73. Petersen KF, Shulman GI. Etiology of insulin resistance. *Am J Med*. 2006;119:S10–S16.
74. Mehler AH. Amino acid metabolism II: metabolism of the individual amino acids. In: Devlin TM, ed. *Textbook of Biochemistry: With Clinical Correlations*. New York: John Wiley & Sons Inc.; 1986:462–464.
75. Schoenheimer R, Ratner S, Rittenberg D. The process of continuous deamination and reamination of amino acids in the proteins of normal animals. *Science*. 1939;89:272–273.
76. Romijn JA, et al. Regulation of endogenous fat and carbohydrate metabolism in relation to exercise intensity and duration. *Am J Physiol*. 1993;265:E380–E391.
77. Gollnick PD, Hermansen L. Biochemical adaptations to exercise: anaerobic metabolism. *Exerc Sport Sci Rev*. 1973;1:1–43.
78. Trefts E, Williams AS, Wasserman DH. Exercise and the regulation of hepatic metabolism. *Prog Mol Biol Transl Sci*. 2015;135:203–225.
79. Richter EA, Hargreaves M. Exercise, GLUT4, and skeletal muscle glucose uptake. *Physiol Rev*. 2013;93:993–1017.
80. Camacho RC, et al. Glucoregulation during and after exercise in health and insulin-dependent diabetes. *Exerc Sport Sci Rev*. 2005;33:17–23.
81. Goodpaster BH, Wolfe RR, Kelley DE. Effects of obesity on substrate utilization during exercise. *Obes Res*. 2002;10:575–584.
82. Gemmink A, et al. Decoration of myocellular lipid droplets with perilipins as a marker for in vivo lipid droplet dynamics: a super-resolution microscopy study in trained athletes and insulin resistant individuals. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2021;1866:158852.
83. Hargreaves M, et al. Effect of carbohydrate feedings on muscle glycogen utilization and exercise performance. *Med Sci Sports Exerc*. 1984;16:219–222.
84. Kelley DE, Goodpaster BH. Skeletal muscle triglyceride. An aspect of regional adiposity and insulin resistance. *Diabetes Care*. 2001;24:933–941.

23

Adipose Tissue Function: Metabolic and Endocrine

Ulf Smith, Peter Arner, and Mikael Ryden

OUTLINE

Obesity and its Health Consequences, 363
  Different Types of Obesity and Associated Health Risks, 363
  Summary, 365
Adipogenesis and its Regulation, 365
  Adipose Cell Differentiation and Role of Genetic Factors, 365
  Role of Aging and Cell Senescence, 366
  Summary, 366
Immunometabolism: Its Regulation and Role in Insulin Resistance, Inflammation, and Disease, 366
  Linking Inflammation to Insulin Resistance, 367
  Causal Drivers of Insulin Resistance and Inflammation in Human Adipose Tissue?, 368
  Summary, 368
Lipid Metabolism in White Adipose Tissue, 368
  Lipid Turnover, 368
  Methodology, 369
  Lipid Uptake and Lipogenesis in Fat Cells, 369
  Regulation of Lipolysis, 370
  Glucose Metabolism, 371
  Summary, 371
Human Brown Adipose Tissue—Brown or Beige?, 372
  Brown versus Beige/Brown Adipose Cells, 372
  Beige/Brown Adipose Cells in the Subcutaneous White Adipose Tissue—A Possible Target in Obesity?, 372
  Aging Reduces Brown and Beige Adipose Cells, 373
  Summary, 373
Adipose Tissue Endocrine Functions and Changes in Obesity, 373
  Summary, 374
Adipose Tissue as a Therapeutic Target, 375
  Pharmacotherapy in Obesity, 376
  Summary, 376

OBESITY AND ITS HEALTH CONSEQUENCES

Our changed lifestyles with less physical activity and easy access to calorie-dense fast foods have caused a global epidemic of obesity. This has also led to an increase in obesity-associated chronic diseases, including cardiovascular disease, diabetes, fatty liver disease (nonalcoholic fatty liver disease [NAFLD]/nonalcoholic steatohepatitis [NASH] and nonalcoholic liver cirrhosis), osteoarthritis, and dementia, as well as liver, breast, and colon cancer. Fig. 23.1 shows a typical Caucasian individual with obesity and summarizes some common chronic disorders and characteristics.

Epidemiologically, the degree of obesity is usually expressed as body mass index (BMI), which is measured as body weight in kg/height in meters^2. Table 23.1 shows current BMI definitions of normal weight, overweight, and different degrees of obesity. Based upon current US trends, it is predicted that by 2030 around 50% of adults will be obese, and, of these, 25% will have severe obesity, with a BMI over 35 kg/m^2.[1] In addition to the rising prevalence of obesity, an additional concern is that obesity also manifests earlier in life, resulting in an earlier initiation and greater duration of the associated chronic diseases. The prevalence of obesity is increasing rapidly in several Asian countries, while it seems to have reached its peak in many European countries. Nevertheless, the current global obesity epidemic is a major challenge to population health and entails rising health costs. A large international age-adjusted study of global mortality in relation to relative degree of obesity measured as BMI showed that mortality in adults increased almost linearly with increasing BMI. Compared with a normal BMI of 18.5 to 25 kg/m^2, set to a hazard ratio (HR) of 1.0, overweight increased HR to 1.2, grade 1 obesity to 1.45, grade 2 obesity to 1.94, and grade 3 obesity to 2.76.[2] This documents the global health problem of obesity.

Although death from myocardial infarction has declined in several countries due to reduced smoking, better hypertension/lipid control, and technological advances in cardiac interventions, the prevalence of heart failure increases due to obesity. Other comorbidities associated with obesity have also become global epidemics, in particular diabetes and its different complications. The International Diabetes Federation[3] estimates that in 2019 the global prevalence of diagnosed diabetes was 463 million people, while an additional 232 million had undiagnosed diabetes. The global health expenditures due to diabetes are estimated to constitute around 10% of the total healthcare costs.[3] Thus, obesity is a major global health problem, and the healthcare system has to a large degree failed to establish treatments resulting in sustained body weight loss. While bariatric surgery is efficient in some of the most severe cases, it will never constitute a global treatment option. The pharmaceutical industry is therefore keen to developing new agents targeting appetite regulation and energy expenditure. In the United States, unlike in most other Western countries, life expectancy has been decreasing in recent years. Although many factors may contribute, the high incidence of obesity in this country is likely to be a causal factor.

Different Types of Obesity and Associated Health Risks

Although BMI is used as an indicator of relative obesity, it has long been recognized that the condition is not homogeneous, and that different types of obesity, related to body fat distribution, are also associated with a lower or higher risk of developing cardiovascular disease,

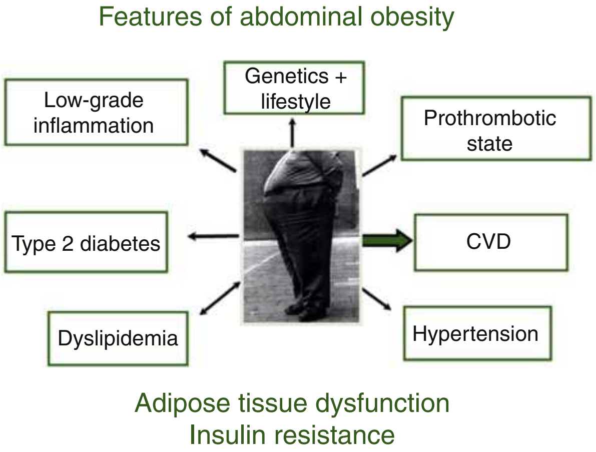

Features of abdominal obesity
Genetics + lifestyle
Low-grade inflammation
Prothrombotic state
Type 2 diabetes
CVD
Dyslipidemia
Hypertension
Adipose tissue dysfunction
Insulin resistance

Fig. 23.1 Features of Abdominal Obesity. *CVD,* Cardiovascular disease.

TABLE 23.1 Definitions of Obesity and Waist Circumference Thresholds

Relative Degree of Obesity, Bod Mass Index, kg/m^2	Waist Circumference Thresholds, cm	Waist Circumference Thresholds, cm
	Males	*Females*
Normal weight 18.5–24.9	>90	>80
Overweight 25–29.9	>100	>90
Obesity, Grade 1 30–34.9	>105	>110
Obesity, Grade 2 35–39.9	>125	>115
Obesity, Grade 3 40–60	>125	>115

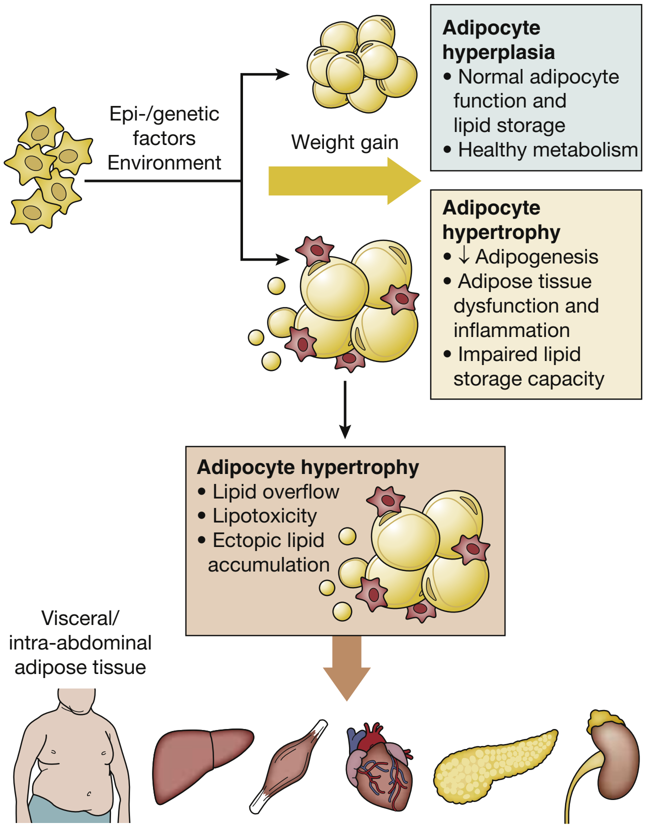

Epi-/genetic factors
Environment
Weight gain
Adipocyte hyperplasia
• Normal adipocyte function and lipid storage
• Healthy metabolism
Adipocyte hypertrophy
• ↓ Adipogenesis
• Adipose tissue dysfunction and inflammation
• Impaired lipid storage capacity
Adipocyte hypertrophy
• Lipid overflow
• Lipotoxicity
• Ectopic lipid accumulation
Visceral/ intra-abdominal adipose tissue

Fig. 23.2 Adipose Tissue Expansion Through Hypertrophy or Hyperplasia and Consequences for Ectopic Fat Accumulation.

diabetes, and other complications. In fact, adipose tissue is a good example of something of which both too much and too little can be harmful, as individuals with lipodystrophy, i.e., lacking areas of adipose tissue, show the same health risks and similar metabolic aberrations as individuals with obesity and expanded adipose tissue.

The subcutaneous adipose tissue (SAT) is the largest adipose depot of the body and also the most efficient and least harmful site to deposit excess lipids. The SAT expands by recruiting new adipose cells (hyperplasia) and/or by expanding the size of available cells (hypertrophy). Adipose cells are unique in their ability to expand and store lipids, as they can inflate in size by at least 10-fold! There is a continuous turnover of adipocytes in SAT of around 10%/year. Recruitment of new cells is essentially limited to before puberty,[4] although it seems that thigh/gluteal SAT can continue to recruit new cells in adult females.[5] Expansion of SAT by hyperplasia is less harmful, while hypertrophic growth is metabolically harmful and leads to a dysfunctional and inflamed adipose tissue, development of dyslipidemia, systemic inflammation, and local and systemic insulin resistance which, in turn, promote the risk of developing obesity-associated complications.[6] There are also other adipose tissue depots, including the visceral (intraabdominal) site, which has long been considered a marker of metabolically unhealthy obesity due to its close association with diabetes, cardiovascular disease, and NAFLD/NASH.[7] Over 50 years ago, this prompted the subdivision into “apple” versus “pear”-shaped obesity, where the former (also termed “abdominal type,” see Fig. 23.1) is more closely associated with health complications. The reason for each individual’s propensity to develop apple/abdominal or pear/peripheral obesity is unclear, but genetic factors play a role, in addition to gender, as women are more likely to have peripheral, and therefore less harmful, obesity, at least before the menopause. After menopause, women can also develop an abdominal type of obesity.[7]

The most simple and commonly used measure of abdominal versus peripheral obesity is waist circumference. There is ample evidence establishing the increased health risk of a large waist circumference for any given BMI. This has led to a recent Consensus Statement by the International Atherosclerosis Society that risk assessments of developing cardiovascular disease and other obesity-associated complications should include BMI *and* waist circumference.[8] Other institutions and governmental bodies, such as the World Health Organization and National Institutes of Health, have also issued recommendations on how waist circumference should be measured and stated that waist circumferences over 88 cm in women and 102 cm in men indicate an increased risk of complications, independent of BMI. However, it is also clear that increased obesity itself leads to expansion of the waist circumference, suggesting that the waist circumference should also be stratified according to BMI, and this is included in the Consensus Statement.[8] Importantly, there are also ethnic differences, and the recommended waist circumference thresholds in Japanese and Asian populations are generally lower than those in Caucasians.

Increased waist circumference is a marker of expanded visceral/intraabdominal adipose tissue, which reflects the available SAT cells becoming saturated with stored excess lipids, which, in turn, leads to increased ectopic lipid storage in other organs/tissues that can store lipids.[9] This is supported by the close association between the amounts of visceral fat and liver lipids (NAFLD/NASH) and also with the accumulation of ectopic lipids in skeletal muscles, pancreas, and heart (Fig. 23.2), with negative metabolic consequences and health risks. The ability of the subcutaneous adipocytes to expand makes their size an excellent marker of adipose tissue health and function, and also of systemic consequences, including degree of insulin resistance. This has been shown in many studies and is exemplified by the finding that the degree of reduction of subcutaneous cell size following bariatric surgery is a considerably better marker of improved insulin resistance than the amount of weight lost.[10] Extensive *in vivo* characterization has shown that subcutaneous adipose cell size is a marker of both the amount of ectopic fat accumulation and the degree of systemic insulin resistance.[11]

It should also be stressed that BMI is not a sufficiently sensitive measure of dysmetabolic state and health risk, given that around 20% to 30% of individuals with normal BMI exhibit insulin resistance and increased risk of developing diabetes and cardiovascular disease. Conversely, a similar proportion of people with obesity can be metabolically healthy; although epidemiological studies have shown that they still have an increased risk of developing diabetes and cardiovascular disease, this risk is considerably less than in weight-matched individuals who display a dysmetabolic state.[12] These data again show that BMI is not sufficiently sensitive at an individual level and indicates situations where additional markers such as waist circumference, or preferably subcutaneous adipose cell size (although not easy to measure in routine clinical settings), are helpful. It would be valuable to have early, sensitive, and clinically useful markers of an individual's that could predict the health risk associated with future weight increase and identify SAT dysfunction at an early stage. The only currently identified genetic marker of adipose cell size and health risk is related to the *KLF14* gene, primarily in females,[13] although another study[14] suggested that it was related to the risk of developing cardiovascular disease in both genders. Several other genes have also been shown to be associated with adipose cell differentiation and storage capacity, but with little individual impact.[15]

KEY POINTS

- Subcutaneous adipose tissue can expand through hyperplasia and, most commonly in adults, through hypertrophy. Expanded subcutaneous adipose cell size and increased waist circumference are indicators of the future risk of developing diabetes and cardiovascular disease.

Another condition of inherently dysfunctional adipose tissue has been identified in individuals with a genetic propensity for type 2 diabetes. Around 30% of first-degree relatives (FDRs) of individuals with type 2 diabetes exhibit an inappropriate increase in subcutaneous adipocyte size for any given BMI, indicating that they can be metabolically obese even at a normal BMI. It is also well-established that FDRs have a very high risk of developing both type 2 diabetes and cardiovascular disease.[16] This sensitivity to the environment is proposed to depend on an inherent (possibly genetic) reduction in the number of available SAT progenitor cells able to undergo differentiation, thus promoting inappropriate hypertrophic expansion of the adipocytes, resulting in dyslipidemia and insulin resistance. This is best seen in the nonobese state,[17,18] because adipose cells also have a finite expansion capability in obesity, and is also seen in nonobese individuals with type 2 diabetes.[19] However, much work needs to be done to better understand the interactions between the environment and adipose tissue genetics.

Summary

The current global epidemic of obesity is a major threat to both individual health and health expenditures. Obesity promotes metabolic dysfunctions including insulin resistance and dyslipidemia and is a major risk factor for developing cardiovascular disease, diabetes, liver disease, cognitive dysfunction and several forms of cancer. While the degree of obesity is commonly stratified by BMI, an individual risk score should also include waist circumference. The SAT has a limited ability to expand and, when exceeded, it leads to accumulation of ectopic fat in many organs/tissues enhancing health risks.

ADIPOGENESIS AND ITS REGULATION

Adipose Cell Differentiation and Role of Genetic Factors

White adipose tissue is a complex tissue consisting of several different cell types that exhibit close crosstalk with and balance among each other to ensure a healthy and functional tissue. The tissue contains uncommitted mesenchymal stem cells, committed preadipocytes at different stages of differentiation, immune cells of different types, and endothelial cells that regulate lipid transport to/from the blood and fibroblasts. As described above, adipose tissue expansion is dependent on the increase in size of differentiated and mature adipocytes (hypertrophy) and/or the induction of commitment and further differentiation of new adipocytes (hyperplasia) from existing progenitor cells, as well as further differentiation of the preadipocytes. In humans, the uncommitted progenitor cells are both of mesenchymal stem cell origin, as well as of hematopoietic (bone marrow) origin. Surprisingly, studies have shown that, in humans, around 10% of the progenitor cells are derived from the bone marrow.[20]

The undifferentiated mesenchymal stem cells are multipotent and can be differentiated *in vitro* to adipocytes, osteocytes, and chondrocytes, depending on environment and specific spatiotemporal differentiation signals. Commitment of both human and murine stem cells into the white adipogenic lineage is dependent on bone morphogenetic protein (BMP) 4,[21] while brown adipose cell (see later) commitment seems to be more dependent on BMP 7.[22] Following initial commitment into the adipose cell lineage, a complicated sequence of events involving a number of different temporally induced transcription factors promotes full differentiation. Among these, C/EBPα and PPARγ are essential to inducing and maintaining a differentiated adipocyte phenotype.

Many studies have shown that adipose tissue is a major regulator of whole-body insulin sensitivity, inflammation, and metabolism through its secretion of different adipokines and cytokines, as well as lipid storage and regulation of lipolysis. Dysfunctional adipose tissue is characterized by insulin resistance, elevated lipid levels, and increased systemic inflammation. Hypertrophic, rather than hyperplastic, expansion leads to dysfunctional adipose tissue and a dysmetabolic state, with increased risk of developing cardiovascular disease, diabetes, and other complications. It is clear from both *in vivo*[23] and *in vitro* studies[24] in humans that hypertrophic expansion of subcutaneous adipose cells is a consequence of a reduced ability to recruit and differentiate new adipose cells, and that there is also a negative correlation between estimated adipose cell number and cell size in humans.[25]

Several studies have identified a large number of genes related to an individual's ability to safely enhance SAT lipid storage and shown that these genes are also associated with a reduced risk of developing diabetes.[26] Genes related to adipose tissue distribution are also markers of adipose tissue cell size and metabolism.[25] As discussed earlier, individuals with a genetic predisposition for type 2 diabetes (FDRs) are also characterized by markers of reduced subcutaneous adipogenesis and inappropriately expanded cells for a given BMI. This is most clearly seen in nonobese individuals, because there is finite expandability of adipose cells in obesity. Together, these data support the concept that genetic factors regulate SAT expansion and distribution, and that these factors are associated with the metabolic state and insulin sensitivity.

There is much additional experimental support for this concept; for instance, it has been shown that functional mutations of PPARγ, the key transcriptional regulator of adipogenesis, are associated with an approximately 8-fold increase in the risk of developing diabetes.[27] Although these mutations are rare, they support the importance of a properly functioning SAT in reducing the risk of developing type 2 diabetes. As discussed later (in the Therapy section), this is also supported by the clinical use of the PPARγ ligands, the thiazolidinediones (TZDs), to treat type 2 diabetes. TZDs enhance differentiation of existing preadipocytes, thereby improving lipid storage capacity in the SAT and the release of beneficial adipokines such as adiponectin, while tissue inflammation is reduced. This leads to improved insulin sensitivity, lower plasma glucose levels, and reduced ectopic lipids in the liver and elsewhere while the SAT expands.[28]

KEY POINTS

- Adipose tissue is a major regulator of whole-body insulin sensitivity, and expanded adipose cell size is a marker of dysfunctional tissue. PPARγ ligands (thiazolidinediones) can increase recruitment of new adipose cells, and this is associated with enhanced insulin sensitivity.

Role of Aging and Cell Senescence

The most important general risk factor for many chronic diseases, including cardiovascular disease and diabetes, is aging. Aging (age >75 years) is associated with a reduced number of functional cells in many tissues, including heart, skeletal muscle, and adipose tissue, and an increased in fibrosis and tissue inflammation. The reason for this is that normally proliferating progenitor cells undergo proliferative senescence with aging that inhibits their ability to proliferate, differentiate, and replace dying cells. Senescent cells have left the cell cycle and become proinflammatory, and also secrete a number of different cytokines and chemokines, proteases, DNA, microRNA, and fibrotic factors: the so-called senescence-associated secretory phenotype (SASP).[29] Secretion of SASP-associated factors to neighboring cells, and systemically, also means that normal, nonsenescent cells become targets and inhibited from undergoing normal differentiation.

Many animal studies have shown that, with aging (defined as mice older than 12 months), the adipose tissue becomes smaller and displays fewer cells.[30]. In humans, it is more difficult to define a specific age for adipose-associated senescence, given that lifestyles are very variable; for instance, levels of physical activity, which prevents induction of cell senescence,[31] may remain high even in individuals of advanced age. However, pooling large databases showed a gradual increase in the relative amount of visceral fat with age, which probably also reflects changes in the amount and cell size of the SAT, even though this was not measured.[7] Insulin sensitivity is also reduced with aging in humans, but, again, defining specific ages is difficult, given that lifestyle has such an important impact on insulin sensitivity. Nevertheless, aging becomes gradually associated with reduced numbers of functional cells in most tissues and combined with enhanced inflammation in both animals and humans. However, aging is only one of several factors that promote cell senescence; other factors include DNA damage, inflammation, increased reactive oxygen species, and others.[32]

As discussed above, the number of subcutaneous adipose cells measured in the abdominal site apparently becomes fixed around the time of puberty.[4] This means that, when these cells have become fully differentiated to adipose cells and accumulated lipids, additional expansion is primarily achieved by hypertrophy, with the metabolic consequences discussed. However, there is also regular turnover of the adipose cells, with a half-time of around 10 years. With aging and increasing cell senescence there will be fewer functional progenitor cells to replace the dying adipose cells, thus leading to further expansion of the available cells. Hypertrophic expansion of the adipose cells is thus a marker of reduced availability of new progenitor cells to differentiate, as has been observed clinically.[33] Similarly, extensive studies with human adipose progenitor cells *in vitro* have shown that individuals with hypertrophic expansion of the adipose cells are also characterized by impaired differentiation of new cells.[24] Because the number of progenitor cells is apparently not reduced, these data suggest that hypertrophic expansion of the adipose cells in adult humans could be associated with increased cell senescence preventing normal differentiation of progenitor cells. In fact, this is exactly what was found with progenitor cells from individuals with hypertrophic obesity, and this was even further pronounced in cells from type 2 diabetic individuals.[34] Importantly, increased cell senescence was also seen in nondiabetic individuals with a genetic predisposition for type 2 diabetes (FDRs), who also exhibited markers of inappropriate expansion of the subcutaneous adipose cells, as discussed.

KEY POINTS

- Increased adipose cell senescence is associated with aging and chronic diseases and means that the adipose progenitor cells can not undergo proliferation as needed to replace dying cells. Senescent cells become proinflammatory and secrete inflammatory and other factors (senescence-associated secretory phenotype) that target and alter ambient cells.

Although senescent progenitor cells cannot undergo differentiation, the number of such cells in human adipose tissue is usually not more than around 10%, even in type 2 diabetic individuals. However, the consequences of cell senescence are amplified by their secretion of SASP-associated factors, which target neighboring, as well as peripheral cells, inhibiting their normal differentiation via para-and endocrine effects. Thus, conditioned tissue culture medium from senescent cells added to nonsenescent cells inhibits normal differentiation of the latter. Interestingly, it is not only the adipose progenitor cells that become senescent, but also the endothelial cells in adipose tissue, which has negative consequences for lipid transport in and out of the tissue[35,36] and contributes to systemic dyslipidemia. To what extent other cells in the adipose tissue also become senescent and what consequences this has needs to be studied.

Summary

Adipose tissue is a major regulator of whole-body insulin sensitivity, but when adipose cells expand (hypertrophic obesity), insulin sensitivity is reduced and inflammation and lipid levels increase, together with the risk of developing cardiovascular disease, type 2 diabetes, liver disease, and other conditions. This risk can be reduced by taking measures to reduce body weight and/or increase adipose cell storage capacity, including treatment with TZDs. Recent studies have identified cell senescence as a major inhibitor of normal cell differentiation and have found that aging, hypertrophic obesity, and type 2 diabetes involve increased cell senescence in the adipose tissue. Use of senolytic agents in animal models has shown that deletion of senescent cells markedly improves several chronic conditions, raising the expectations for future human treatments when potential safety issues have been clarified.

IMMUNOMETABOLISM: ITS REGULATION AND ROLE IN INSULIN RESISTANCE, INFLAMMATION, AND DISEASE

Inflammation is a physiological consequence that follows exposure to external and internal stressors and involves an acute response characterized by redness (rubor), heat (calor), pain (dolor), swelling (tumor), and dysregulated cellular/organ function (functio laesa).

That inflammation could be linked to metabolic diseases such as obesity and type 2 diabetes was known already in the nineteenth century, when salicylate was proposed as a treatment against obesity-associated diabetes. However, whether this effect was linked to adipose tissue was unknown. While the notion of inflammation as a culprit driver was subsequently established in a number of other diseases, its role in white adipose tissue was for a long time disregarded. It was not until the early 1990s that it was demonstrated in murine models that obesity and insulin resistance are linked to increased expression and local release of tumor necrosis factor alpha (TNFα) in white adipose tissue.[37] This sparked an intense line of research trying to understand the pathophysiological mechanisms and the role of this response. What soon became clear from these early studies in animals, and later also in human adipose tissue, was that obesity/insulin resistance is associated with a chronic low-grade inflammation that lacks most of the classical symptoms of acute inflammation. Instead, there is increased expression and release of cyto- and chemokines expressed and released from several different cell types, and some of these can act as chemoattractants to promote further infiltration of different immune cells. A plethora of studies have confirmed that adipocytes can indeed express and release several proinflammatory factors, including CCL2 (also termed monocyte chemoattractant protein-1) and interleukin-6 (IL6), while TNFα and IL1β are secreted locally, predominantly from adipose tissue–resident macrophages. The latter are central cell types in the innate immune system and are among the most abundant leukocytes in obese/insulin resistant adipose tissue. Typically, macrophages in the adipose tissue from subjects with obesity/insulin resistance are centered around what are considered to be "dying" adipocytes in crown-like structures. Traditionally, these macrophages display a phenotype more similar to a proinflammatory M1 subtype than to the tissue repair–associated M2 cells. Admittedly, the subdivision into M1/M2 is an oversimplification, and inflamed adipose tissue can display a large panorama of different macrophage subtypes,[38] the roles of which remain somewhat unclear. Moreover, while macrophages are the most abundant white blood cells in inflamed adipose tissue, most types of immune cells can be found therein, including T-cells, B-cells, and different forms of granulocytes.

The inflammatory phenotype described above is linked to a number of morphological changes in adipose tissue, including increased adipocyte size (see also earlier), collagen deposition, fibrosis, and possibly reduced vascularization and hypoxia,[39] although the latter is still debated regarding human white adipose tissue.[40] The functional changes in adipocytes linked to inflammation include increased basal but attenuated hormone-stimulated lipolysis (see also section on lipid turnover), reduced insulin-induced glucose uptake and lipogenesis, and lower adipogenesis. These are all processes where insulin plays a major regulatory role in promoting lipid accumulation, and thereby energy storage. Collectively, this reduction in adipocyte insulin sensitivity[41] will result in reduced lipid storage capacity, and it is therefore possible that the inflammatory response induced under caloric oversupply is an evolutionarily conserved response that helps limit adipose tissue expansion. This response becomes maladaptive under a constant hypercaloric state. It should also be mentioned that data in animal models suggest that a (possibly transient) proinflammatory response may be required for adipose tissue expansion, and that inhibiting this process may result in a pernicious metabolic profile.[42]

KEY POINTS

- White adipose tissue displays a proinflammatory response in the obese/insulin-resistant state that is characterized by an increase in the expression and secretion of chemoattractants in the adipocytes themselves, as well as an increased abundance of resident leukocytes.

KEY POINTS

- The proinflammatory response in white adipose tissue is linked to morphological and functional changes in tissue phenotype. While this chronic activation is detrimental, a transient inflammatory response may be important to allow white adipose tissue expansion.

Linking Inflammation to Insulin Resistance

A number of studies have identified the different mechanisms through which inflammation attenuates insulin signaling. These involve both transcriptional (mediated, e.g., by NF-kB activation) and posttranscriptional effects on different phosphorylation steps in the insulin signaling cascade.[43] Combined approaches, using both cyto-/chemokine blocking antibodies *in vivo* as well as genetically modified animal models (primarily tissue-specific knockouts and transgenic mice), have established the causal role that adipose tissue inflammation plays on insulin resistance. This has spurred substantial interest in targeting adipose inflammation in humans. Several clinical studies have been performed with blocking antibodies targeting cytokines. The first approach to be tested was to inhibit TNFα with a single intravenous injection, which proved unsuccessful in improving metabolic control.[44] Subsequent studies using more established TNFα inhibitors (e.g., infliximab, used routinely in the treatment of rheumatic and inflammatory bowel disease) have not demonstrated significant metabolic effects in subjects with type 2 diabetes. The reasons for this could be multiple and include the possibility that the antibody concentrations were insufficient, the fact that adipose tissue does not release TNFα into the circulation, and/or that antibodies penetrate poorly into the adipose tissue. Moreover, TNFα inhibition in other conditions can cause severe side effects such as serious/fatal infections (e.g., reactivation of tuberculosis). Instead, based on its central role in attenuating cell function in organs such as adipose tissue and the pancreas, several clinical studies have focused on inhibiting IL1β. The IL1 receptor antagonist anakinra and the IL1β-inhibitory antibody canakinumab, originally developed for use in rheumatic diseases, have been studied in randomized, placebo-controlled trials with glycemia and type 2 diabetes incidence as primary outcomes. Anakinra was shown to reduce systemic inflammation (measured as circulating C-reactive protein and IL6 levels) and improve glycated hemoglobin and insulin secretion, although the effects were less impressive than currently available antidiabetic drugs.[45] Canakinumab reduced the incidence of cardiovascular events[46] but had no effect in reducing the incidence of type 2 diabetes in patients with prediabetes at baseline.[47] Moreover, the rate of fatal infections was significantly higher in the canakinumab arm of the study. Altogether, while the effects on adipose tissue were not assessed in either of these studies, the results suggest that targeting inflammation is not a very efficient treatment for obesity—associated insulin resistance and/or type 2 diabetes and the rate of serious adverse events are not negligeble.[48] This indicates that inflammation in adipose tissue may be a consequence rather than a causal factor, and that efficient treatments targeting adipose tissue should focus on the mechanisms that drive inflammation in the first place. In line with this notion, data in genetically modified animals suggest that insulin resistance in adipocytes promotes the development of inflammation, not the other way around.[49]

KEY POINTS

- Therapies targeting white adipose tissue inflammation are not particularly efficient in improving insulin sensitivity/glucose control, suggesting that inflammation is a consequence rather than a causal factor underlying the development of insulin resistance/type 2 diabetes.

Causal Drivers of Insulin Resistance and Inflammation in Human Adipose Tissue?

These aspects have prompted the quest for identification of the drivers of insulin resistance and inflammation in adipose tissue. Much focus has been put on adipocyte size, as cell hypertrophy is closely linked to a proinflammatory adipose phenotype with increased secretion of, e.g., TNFα and CCL2 and reduced insulin sensitivity.[19] It is still unclear how adipocyte size could impact inflammation, but it has been proposed that hypertrophy is a result of reduced adipogenesis, fibrosis, and possibly local hypoxia.[50] Others have suggested that changes in adipocyte metabolism result in an epigenetically-driven change in gene transcription promoting a proinflammatory state. For example, obesity and adipocyte hypertrophy are characterized by lower glutamine levels in the tissue, which promotes an increase in O-GlcNAcylation, a posttranslational modification of proteins and chromatin that activates the transcription and secretion of IL6 and IL1β.[51] This, as well as other changes in adipocyte metabolism, could be part of an evolutionarily conserved response to increases in adipocyte size in order to prepare the tissue for tissue expansion, as discussed previously.

KEY POINTS

- The drivers of WAT inflammation remain unclear but increased fat cell size and disturbances in adipocyte metabolism are believed to be important.

Summary

While adipose tissue inflammation is closely linked to insulin resistance and a pernicious metabolic profile, the causal mechanisms remain unclear and its role as a therapeutic target are probably negligible. Instead, future efforts should aim at identifying the culprit pathways that enable this response which could possibly pave the wave for other treatment approaches.

LIPID METABOLISM IN WHITE ADIPOSE TISSUE

Adipose tissue was long thought to be an inert organ harboring lipids for energy storage and isolation purposes. In the middle of the twentieth century it became apparent that the fat cells within white adipose were very active, displaying marked metabolism of lipids, i.e., storage and release of fatty acids, to control energy homeostasis. These processes are relatively complex with regard to both regulation and physiological/pathophysiological impact, and have been reviewed in detail elsewhere.[52,53] Herein, the most important aspects are discussed, i.e., lipid turnover, lipid synthesis, lipid mobilization, and the impact of glucose metabolism.

Lipid Turnover

Fatty acids are the most energy-rich molecules in the body and are mainly stored as triglycerides in fat cells and liberated through hydrolysis of the triglycerides. Although the fatty acid turnover is mainly driven by storage/release in adipose tissue, other sources are also of importance. These include ingestion during meals, usage for very low density lipoprotein–triglyceride production by the liver, trapping (see later), and oxidation during physical or mental performance. Lipids are stored in mature white fat cells in a single lipid droplet mainly composed of triglycerides. This is different from all other cell types, which display multiple lipid droplets.[54] There is constant turnover of these lipids (Fig. 23.3), which constitutes a single metabolic pool within the lipid droplet in human fat cells.[55] During the 10-year lifespan of a fat cell, triglycerides are turned over approximately six times, irrespective of the size of the lipid droplet. Two adipocyte factors determine the turnover rate, namely the abilities to synthesize (K_{in}) and hydrolyze (K_{out}) triglycerides. The turnover is constant when $K_{in} = K_{out}$. In addition to these kinetic parameters, the turnover is governed by external factors, namely alimentation (mainly influencing K_{in}) and peripheral utilization (mainly influencing K_{out}). Changes in fat cell lipid turnover have clinical implications. Overweight and obesity are characterized by decreased turnover due to increased K_{in} and decreased K_{out}. Dyslipidemia, in particular a genetic form termed familial combined hyperlipidemia, is associated with decreased turnover that is linked to a combination of low K_{in} and K_{out}. It is speculated that this dual kinetic effect "uncouples" fat cells from fatty acid uptake/release and thereby increases the exposure of the liver to fatty acids, which are major substrates for the triglyceride component in lipoprotein synthesis. Differences in lipid turnover can also be observed between adipose depots. As discussed earlier, adipose tissue is distributed into different regions, which display very different sizes. The subcutaneous area is by far the largest (approximately 80% of all adipose tissue). It is believed that this region is a metabolic buffer with high capacity to retain lipids that otherwise would spill over to other tissues such as liver, skeletal muscle, and pancreatic β cells and thereby cause metabolic aberrations.[53,56] The "buffering"

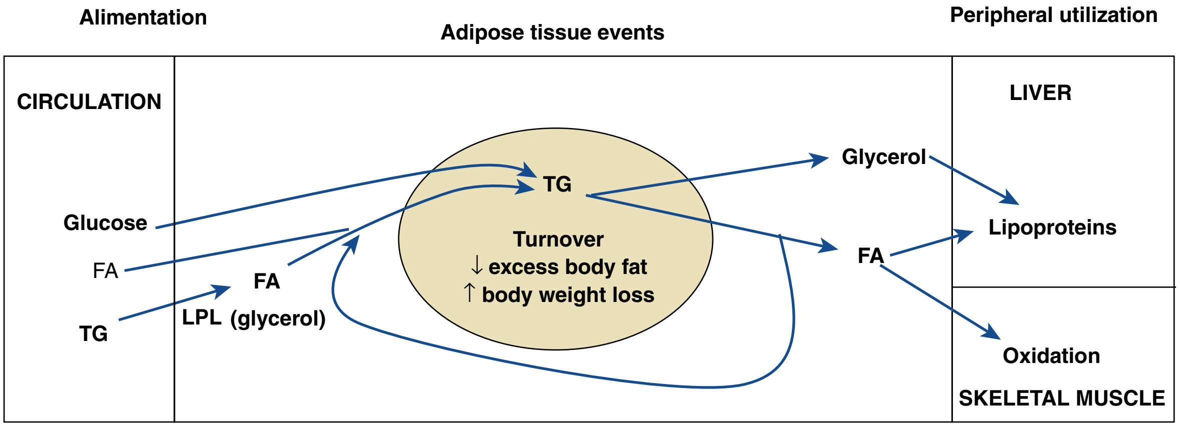


Fig. 23.3 Adipose Tissue Events. *FA,* Fatty acids; *TG,* triglycerides; *LDL,* low density lipoprotein.

capacity is most pronounced in the "lower body" (below the waist line) SAT,[57] which may explain, at least in part, why "upper body" (above the waist line) obesity is more pernicious from a cardiometabolic point of view.[58,59] The size and distribution of SAT is subject to sex dimorphism and differences in lipid kinetics between men and women.[60] In general the depot is larger in women than in men. In addition, women tend to expand lower SAT more than men, while men are prone to accumulate fat in the upper body region. There are no clear-cut differences between the sexes in subcutaneous adipose lipid turnover, although men appear to have less capacity to store and release fatty acids than women *in vivo*. From a pathophysiological point of view the smaller visceral adipose depot has a unique role.[58,61] It is drained by the portal vein to the liver, so fatty acids can directly be delivered to this organ from visceral fat cells and thereby influence lipoprotein and glucose metabolism. The lipid mobilization activity is greater in visceral adipose tissue than in the subcutaneous fat region. This is evident in upper body obesity, and in particular among men, who often have marked visceral adipose accumulation. The consequence is that large amounts of fatty acids are delivered to the liver, causing glucose intolerance and dyslipidemia according to mechanisms detailed elsewhere.[52,53,61] Enhanced peripheral delivery from upper SAT of fatty acids to the liver is also involved in the described pathophysiology. The regional differences in lipid turnover occur *in vivo*.[62] During the gradual transition from leanness to obesity, lipid turnover is lowered in SAT already in the overweight state. Visceral adipose tissue does not change in this way until marked obesity is present. This regional variation may be a factor responsible for higher lipid mobilization from visceral than SAT in less extreme conditions of excess body fat. Finally, lipid turnover is directly involved in changes of subcutaneous fat mass over time.[63] Aging is associated with decreased turnover, mainly due to a decrease in K_{out}, at least in part caused by an age-dependent decrease in catecholamine-stimulated lipolysis. If this is not compensated by a decrease in K_{in}, the subcutaneous fat mass will be expand. Conversely, an increase in K_{out}, such as occurs following bariatric surgery, is associated with long-term maintenance of reduced body weight.

Finally, a recently recognized and potentially important aspect of adipose lipid turnover is so-called fatty acid trapping.[64] These lipid species are not only stored in adipose tissue as triglycerides, they can also be retained within the extracellular space of adipose tissue bound to carriers such as albumin or specific transporter proteins. During esterification and lipolysis (see later) some fatty acids are trapped and not immediately taken up by fat cells or released into the circulation. This is most evident in obesity and dyslipidemia, where fatty acids are trapped in increased amounts and may cause postprandial dysmetabolism according to mechanisms discussed in detail elsewhere,[64,65]

KEY POINTS

- There is a rapid turnover of the triglycerides in the lipid droplet of human fat cells. During the 10-year lifespan of a fat cell, its lipid droplet triglycerides are renewed around six times. Changes in the rate of turnover are important for the regulation of adipose tissue. Decreased turnover is associated with enlargement of the fat mass and with dysmetabolic features.

Methodology

Human fat cell lipid (and glucose) metabolism has been intensely investigated thanks to advanced *in vivo* and *in vitro* methods summarized in Table 23.2. By determining the ^{14}C content in lipids it is possible to calculate their age and estimate the turnover. Fatty acid, glycerol, or glucose tracers can be infused intravenously and used in combination with catheterization or position emission tomography to determine uptake and release of relevant metabolites. Small water-soluble molecules (mainly glycerol, glucose, and lactate) can be measured *in situ* in the water space of SAT by means of microdialysis. Adipose tissue can be excised for ex *in vivo* analyses. Often freshly isolated fat cells are prepared for short-term (hours or a few days) metabolic studies. Finally, precursor cells such as uncommitted mesenchymal stem cells or committed so-called preadipocytes can be isolated from the stroma vascular compartment and cultured for a long time. These precursor cells have been instrumental for understanding the molecular regulation of fat cells, because they can be subjected to genetic transfer and editing.

Lipid Uptake and Lipogenesis in Fat Cells

Fat cells can synthesize triglycerides in several ways; the stepwise esterification starting with glucose and fatty acids is quantitatively the most important one (Fig. 23.3). Fatty acids and triglycerides in chylomicrons or very low density lipoprotein enter the extracellular adipose compartment from the bloodstream. The triglycerides are hydrolyzed by lipoprotein lipase.[66] This enzyme is synthesized by fat cells and excreted to the extracellular space of adipose tissue, where it catalyzes the breakdown of blood-derived triglycerides into fatty acids and glycerol. Glucose is transported into fat cells and metabolized to alpha glycerol phosphate, which is the initial glycerol backbone of triglycerides. The extracellular fatty acids derived from the lipoprotein lipase action, as well as the free fatty acids bound to albumin in the circulation, are transported into fat cells. They are esterified onto the glycerol backbone to form triglycerides. Fatty acids produced by lipolysis (see later) can also be used for triglyceride synthesis in the pathway mentioned earlier instead of leaving the fat cells. This is called reesterification and is quantitatively important; up to 30% of fatty acids produced by lipolysis may

TABLE 23.2 Methods for Studying White Adipose Tissue Metabolism in Humans

Method	Body Level	Main Use
Intravenous infusion of tracers, mainly labeled fatty acids and dexoyglucose in combination with positron emission tomography or catheterization	*In vivo* Whole body or regions	Lipolysis or uptake of glucose and fatty acids by adipose tissue
Catheterization of subcutaneous adipose tissue veins	*In vivo* Abdominal or gluteal regions	Uptake, release, or retention of fatty acids
Microdialysis of subcutaneous adipose tissue	*In situ* Abdominal or gluteal regions	Lipolysis (glycerol levels)
Determination of lipid age by ^{14}C dating	*Ex vivo* (in removed tissue samples)	Lipid turnover studies
Isolated mature fat cells	*In vitro* in living fat cells	Molecular and metabolic studies
Differentiated fat cells	*In vitro* using cell cultures of precursor cells	Molecular, genetic, and metabolic studies

TABLE 23.3 **Species Differences in White Fat Cell Lipid Metabolism**

Event	Humans	Rodents
Hormonal stimulation of lipolysis in adults	Only catecholamines and natriuretic peptides	Numerous hormones, but not natriuretic peptides
Catecholamine signaling	Mainly lipolytic beta 1 and 2 adrenoceptors and antilipolytic alpha-2 adrenoceptors	Lipolytic beta 1, 2, and 3 adrenoceptors; none of the antilipolytic alpha-2 type
Insulin signaling	Through insulin receptor substrate 1 and 2	Through insulin receptor substrate 1, 2, and 3
TNFα signaling to lipolysis	Involves cell death–inducing DNA fragmentation factor alpha-like effector A	Involves Gi proteins
De novo fatty acid synthesis	Little in mature fat cells	Marked in mature fat cells
Cell death–inducing DNA fragmentation factor alpha-like effector A	Important for lipid metabolism	Not expressed

be reesterified. There are also alternative pathways for triglyceride synthesis in fat cells, but they are quantitatively less important in humans. Glucose can be metabolized to fatty acids (*de novo* lipogenesis), and other metabolites such as pyruvate can be converted to glycerol (glyceroneogenesis). The glycerol molecules can be directly phosphorylated to alpha glycerol phosphate and used for fatty acid esterification. The major regulator of the adipocyte lipid synthesis is insulin. The hormone activates lipoprotein lipase and stimulates fat cell uptake and metabolism of glucose. Insulin may also directly stimulate the enzymes involved in fatty acid esterification. Following food intake, insulin levels are elevated, and the different regulatory factors described above are activated. During food deprivation, insulin levels are low and, consequently, the triglyceride accumulation in fat cells diminished. Obesity *per se* is associated with increased activity of lipoprotein lipase, which may facilitate net storage of triglycerides in fat cells.

Regulation of Lipolysis

KEY POINTS

- The quantitatively most important factor for lipid storage within fat cells is uptake and esterification of fatty acids derived from triglyceride-containing particles in the circulation. In addition, uptake of free fatty acids originating from the circulation or derived from fat cell lipolysis is also important. However, *de novo* lipid synthesis from glucose (lipogenesis) is only a minor contributor to total lipid storage in fat cells.

The signal and metabolic pathways regulating lipid mobilization are well recognized but subject to important species differences (Table 23.3). The regulation in human fat cells is depicted in Fig. 23.4. Triglycerides are stepwise broken down during a process termed lipolysis.[67] A specific lipase termed adipose triglyceride lipase (ATGL) and its cofactor truncated abhydrolase domain–containing protein-5 (CGI-58) catalyzes the breakdown of triglycerides into diglycerides and fatty acids. The diglycerides are hydrolyzed to monoglycerides and fatty acids by another enzyme, hormone-sensitive lipase (HSL). Finally, monoglyceride lipase catalyzes the final lipolysis step, namely formation of glycerol and fatty acids from the breakdown of monoglycerides. Despite being a relatively simple process, lipolysis is regulated in different manners by several hormonal and metabolic factors.[68,69] Five different signal transduction systems have been described, all using specific cell surface receptors. A lipolytic and an antilipolytic system are coupled to cell membrane–bound adenylyl cyclase via inhibitory G_i- and stimulatory G_s-coupled membrane proteins. Catecholamines are the most important hormones causing lipolysis stimulation, which occurs through beta-adrenergic receptors (in humans mainly type 1 and 2) and G_s-coupled proteins. However, they can also bind to alpha2A adrenergic antilipolytic receptors, which are coupled to G_i. Therefore, the net effect of catecholamines on lipolysis is determined by the balance between beta (usually dominant) and alpha effects. Other antilipolytic G_i-coupled hormones and metabolites are adenosine, prostaglandins, and lactate, which are produced locally and therefore act in a paracrine/autocrine fashion. The physiological role of these regulators is not well understood. G_s activation stimulates adenylyl cyclase, which increases the production of cyclic adenosine monophosphate (AMP) from adenosine triphosphate (ATP) and, in turn, activates an enzyme complex termed protein kinase A. Inhibition of adenylyl cyclase has the opposite effect. This mode of action also operates for two other endogenous substances, niacin and beta-hydroxybutyrate. They share the same G_i-coupled receptor, HCA2. Binding of the agonists to the fat cell surface receptors inhibits lipolysis through the described inhibitory pathway. Another class of prolipolytic hormones is the family of natriuretic peptides, which bind to cell surface receptors linked to cell membrane-associated guanylyl cyclase. This generates cyclic GMP which, in turn, activates the enzyme complex protein kinase G.

Insulin is the most important antilipolytic hormone. After receptor binding it activates receptor-associated proteins termed insulin receptor substrates. They activate the enzyme phosphoinositide 3-kinase and further downstream signaling, which ultimately stimulates the enzyme phosphodiesterase 3B. This hydrolyzes cyclic AMP into inactive 5’ AMP so that protein kinase A becomes less active. Protein kinase A and G have the same effects on the lipases discussed earlier. Thus, both kinases phosphorylate and activate HSL, as well as the ATGL–CGI-58 complex. In addition, a lipid droplet coating phosphoprotein, perilipin-1, is phosphorylated, which allows the lipases to access the lipid droplet surface for triglyceride hydrolysis.

In addition to the hormone-regulated pathways described earlier, a signal pathway through TNFα is important for controlling the chronic (basal) lipolytic activity. The cytokine is produced by fat cells and inflammatory cells within adipose tissue and is thus an autocrine/paracrine factor. The TNFα receptor signal pathway has two actions. Firstly, it inhibits phosphodiesterase 3B, which elevates the cyclic AMP level. Secondly, it activates another lipid droplet coating protein called cell death–inducing DNA fragmentation factor alpha-like effector A. Both cytokine effects increase basal lipolytic activity.

Although not definitively proven, changes in lipolysis are likely to have a clinical impact. Perhaps the most obvious factor is the relation to fat cell size.[70] The rates of spontaneous and catecholamine-stimulated lipolysis are increased, while the antilipolytic effect of insulin is decreased, in large fat cells. This may be because large fat cells contain more cyclic AMP than small ones. Obesity is often characterized by large fat cells, which would intuitively cause increased lipolysis rates. However, the impact of obesity on lipolysis is complex. The basal rate is elevated, probably as a consequence of inflammation and thereby enhanced local production of tumor necrosis factor alpha. The antilipolytic effect of

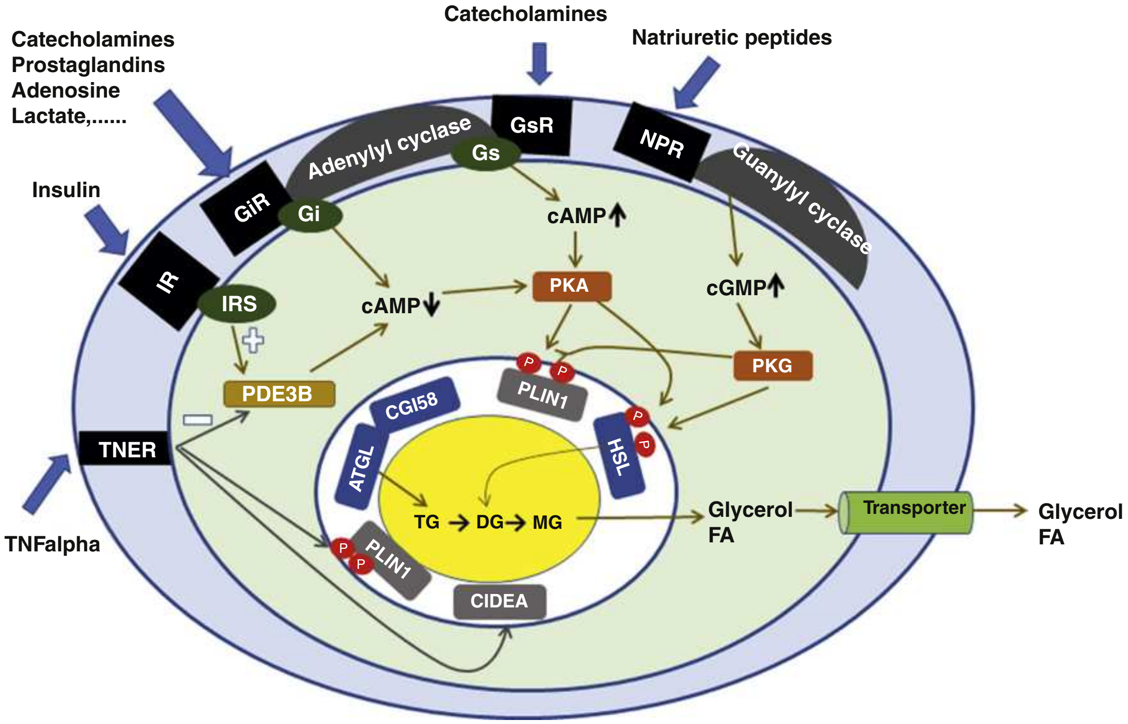


Fig. 23.4 Regulation in Human Fat Cells.

insulin is decreased, but the same is true for the prolipolytic effect of catecholamines and natriuretic peptides. The effects of excess body fat on lipolysis-regulating hormones may explain why fasting circulating fatty acid levels are only modestly elevated in the obese state. On the other hand, animal studies and short-term clinical studies show that inhibition of lipolysis improves insulin sensitivity.[71] It is therefore plausible that obesity leads to increased availability of fatty acids, which, in turn, contributes to insulin resistance, impaired insulin secretion, and ultimately type 2 diabetes according to mechanisms described in detail elsewhere.[71] On the other hand, prospective studies, which are needed to support a causal role of lipolysis in metabolic diseases, are rare. In one such study it was found that increased basal lipolysis and decreased catecholamine-induced lipolysis predict future glucose intolerance and body weight gain.[72] The failure of insulin to adequately suppress fatty acid levels in obesity may also lead to so-called metabolic inflexibility in skeletal muscle.[73] The utilization of glucose and fatty acids as fuels for skeletal muscle is switched towards fatty acid use, which may cause lipid toxicity, and thereby insulin resistance, as discussed in detail elsewhere.[73] Finally, an association between fat cell lipolysis *in vitro* and circulating lipid levels has been demonstrated in cross-sectional studies.

KEY POINTS

- In human fat cells, only a few hormone systems have a pronounced, acute effect on lipolysis: the stimulatory catecholamines and natriuretic peptides, plus the inhibitory insulin. In addition, the spontaneous (basal) lipolysis is regulated by local inflammatory factors, in particular TNFα, which increases the lipolytic rate. In obesity and other metabolic conditions, lipolysis regulation is altered. Humans with excess body fat have increased basal lipolysis, decreased ability to activate lipolysis by catecholamines, and resistance to the antilipolytic effect of insulin. Together, these alterations are involved in the metabolic complications of obesity, such as dyslipidemia and insulin resistance. Prospective studies suggest that altered fat cell lipolysis is a factor behind future body weight gain and glucose intolerance.

Glucose Metabolism

Fat cells use the same signal system and regulation as skeletal muscle for glucose metabolism. In both cell types, insulin is the major regulatory hormone (stimulation of uptake and further metabolism). However, in quantitative terms, white adipose tissue has a much smaller capacity than skeletal muscle to handle a glucose load. Approximately 10% to 15% of glucose dispersal after food intake is into adipose tissue. There is, on the other hand, one important exception as regards the quantitative role of fat cells in carbohydrate metabolism, namely lactate.[74] As a result of glucose metabolism, fat cells produce large amounts of lactate, and this is augmented in large fat cells in obesity. The latter might be due to decreased oxygenation in obese adipose tissue and/or changes in cellular metabolism leading to activated glycolysis (see earlier). Lactate serves as an autocrine inhibitor of lipolysis (Fig. 23.4). In addition, adipose tissue provides a source for circulating lactate to be used as three-carbon fuel.

KEY POINTS

- Adipose tissue plays a quantitatively minor role in the regulation of whole-body glucose uptake. However, glucose metabolism in fat cells has many indirect effects on carbohydrate metabolism, such as on lactate turnover and insulin sensitivity.

Summary

Adipocyte storage and release of energy-rich fatty acids is essential for survival during long periods of food deprivation and in connection with long-term physical exercise. In addition to being an energy storage organ, white adipose tissue is a reservoir for lipids that may be toxic in other places such as liver, pancreatic β cells, and skeletal muscle. Having too much or too little adipose tissue alters lipid handling and leads to disturbed metabolism, where insulin resistance is the most prominent and important dysfunction.[75] Enlarged fat cells and unfavorable adipose distribution into the upper body/visceral areas are major culprits in this scenario, and altered lipolysis probably the primary pathophysiological factor.

HUMAN BROWN ADIPOSE TISSUE—BROWN OR BEIGE?

Brown adipose tissue (BAT) is primarily found in small mammals and hibernating animals, and brown adipose cells in mice have the same origin as muscle cells.[76] BAT regulates body temperature through nonshivering thermogenesis and becomes activated when the animals are exposed to cold or temperatures below their thermoneutrality zone. In mice this is around 30°C, when metabolism is the lowest, and thus considerably higher than the usual room temperature at which mice are housed. Exposure to cold leads to activation of the sympathetic nervous system, which increases lipolysis and the release of fatty acids that are used by the mitochondria-dense brown adipose cells. Increased sympathetic activation and lipolysis in BAT increase nonshivering heat production through the UCP1 uncoupling protein in the mitochondria. UCP1 uncouples the mitochondria to produce heat rather than the normally generated ATP.[77] Many experiments in animals have also shown that sympathetic activation of BAT in the cold consumes enough calories to reduce the amount of adipose tissue, and thus could be a target to treat obesity.

BAT is not only present in small mammals, but also in the interscapular region in human infants, where it is critically important for thermogeneration during the first year of life. It then gradually disappears over the first decade. It was long considered to be completely absent in adult humans, with a few exceptions, including patients with pheochromocytomas secreting high amounts of catecholamines and outdoor workers chronically exposed to cold environments. However, in 2007 it was suggested that adult humans may also have some smaller BAT depots.[78] This provoked new scientific interest, and, in 2009, several clinical studies using advanced positron emission tomography/computed tomography scans documented the presence of distinct adipose depots in adult humans that were functional and increased glucose and fatty acid uptake, as well as thermogenesis, following stimulation with β3-adrenergic stimuli or exposure to cold. However, these putative BAT depots were small (around 50–100 g) and primarily found in the cervical, supraclavicular, axillary, mediastinal, paraspinal, and abdominal regions around the kidneys. These findings were followed by many studies suggesting that individuals with larger amounts of BAT also had lower body weights; i.e., it was hypothesized that increasing BAT could be a way of treating obesity. However, several studies performed since then have concluded that BAT in adult humans is not the same BAT observed in rodents, with small lipid droplets and high amounts of mitochondria and UCP1 protein. Instead, it was suggested that human BAT is more of a so-called “beige/brite” phenotype with large lipid droplets and low UCP1 expression, and is of unclear importance as a temperature regulator.

This has been an issue of long debate, but recent analysis suggests that human BAT is, in fact, similar to the animal tissue, but with a different morphology, because humans live in a thermoneutral temperature and are well insulated and, therefore, do not normally require sympathetic activation and BAT-induced regulation of thermogenesis.[79] Although this may be the case, it is still not likely that human BAT is large enough, contains sufficient UCP1, and/or can be activated to the extent required to actually be a target for body weight reduction. This is clearly in contrast to BAT in mice and other animals.

Brown versus Beige/Brown Adipose Cells

As discussed, human brown adipose cells have been characterized as intermediate between white and brown cells, and therefore named beige/brown cells. However, administration of lipolytically active β3-agonists was shown to increase metabolism in BAT in humans, although only at very high amounts, when β3-selectivity was unlikely and associated with increased heart rate and blood pressure.[80] This would not be acceptable for an agent to treat obesity. It is well-established that β3 receptors are important for activating BAT in mice, but further studies have shown that β2-adrenergic receptors responding to noradrenaline are mainly expressed in humans. In fact, β3 receptors do not seem to be expressed in human BAT cells at all. However, they are highly expressed in human urinary bladder and gall bladder, as well as in some areas in the heart, which can explain the increased heart rate and blood pressure seen with the administration of β3-selective agonist.[80]

The small amount and different morphology of human brown adipose cells (the beige/brown phenotype) have been taken as an indication that the tissue also is thermogenically unimportant. However, it has also been argued that the reason for this is that humans normally live in well-insulated, thermoneutral environments. In support of this, murine cells also become less active when the animals live in a thermoneutral environment for mice, i.e., around 30°C. Nevertheless, the amount of BAT in man is so small that it is unlikely that it could be activated to a sufficient extent to treat obesity, even if we had access to drugs that could specifically activate human BAT cells. Providing indirect support for this is that fact that the antidiabetic PPARγ agonists, the TZDs, induce beige/brown cells and UCP1 in human adipose cells.[81] However, treatment with TZDs is known to increase body weight, rather than decrease it, supporting the idea that human brown cells are a difficult obesity target.

Beige/Brown Adipose Cells in the Subcutaneous White Adipose Tissue—A Possible Target in Obesity?

Another scientific approach to find ways to increase oxidative capacity and thermogenesis in human adipose tissue is to increase mitochondria and UCP1 in the conventional white adipose cells through partial transdifferentiation of white cells to a beige/brown oxidative phenotype or to target available specific progenitor cells to become beige/brown and increase their oxidative capacity. It seems that both approaches may be possible, but a key question is if enough UCP1 can be induced to actually increase thermogenesis of sufficient magnitude to also reduce obesity in humans; this remains to be demonstrated. However, both in rodents and in humans, white adipose cells in the large SAT are susceptible to partial transdifferentiation to a beige/brown phenotype. TZDs can induce a beige/brown phenotype in white SAT cells. Mechanistically, this is due to sirtuin-dependent deacetylation of PPARγ promoting its interaction with PRDM 16 and EBF 2, leading to its binding to brown-specific enhancers. In addition to TZDs, thyroid hormones, FGF 21, and BMP4, 7, 8b, and 9 have been shown to increase beige/BAT, and TZDs and BMP4 also target human white adipose cells.[79,82]

Thus, the white subcutaneous adipose cells are quite susceptible to undergoing partial transdifferentiation from a white to a beige/brown phenotype, but the number of cells capable of being transdifferentiated and the overall consequences for thermogenesis and body weight reduction need to be demonstrated in humans. It has been shown that increased BMP4 levels in mice induce a beige/brown and oxidative phenotype of the white adipose tissue of sufficient magnitude to increase energy expenditure and prevent induction of obesity.[83] However, whether these findings can be translated to a possible therapeutic target in humans is currently unclear, considering the complex BMP signaling pathways and potential side effects.

It is clear that the quantity of induced beige/brown adipocytes required to increase energy expenditure is considerably higher than that required by normal and functional BAT. Considering that the UCP1 mRNA levels are around 1000-fold higher in brown than in white adipose cells, it has been calculated that a relevant browning agent needs to raise the amount of UCP1 in white adipose tissue by 10-fold in order to have around 1% of the amount in BAT.[84] Although the pharmaceutical industry is trying to develop browning agents suitable for humans, no such agent is currently available, and it is not clear if this will become a possibility.

KEY POINTS

- Brown adipose cells are present in small amounts in humans and are characterized by large lipid droplets and being more beige/brite than brown. Unlike murine cells, they only have small amounts of UCP1 and mitochondria. Although they respond to sympathetic activation, it is unlikely that they can increase metabolism to an extent needed to reduce body weight.

Aging Reduces Brown and Beige Adipose Cells

Aging is associated with profound changes in most organs and tissues, including BAT in both rodents and humans.[79] This reduction in BAT is more likely to have consequences for thermogenic regulation and propensity to increase body weight in animals than in humans.

It has also been shown that the number of beige/brown cells in the white adipose tissue in rodents and humans decrease with age, probably as a consequence of both cell senescence and reduced sympathetic activity. Thus, aging contributes to the loss of both BAT mass and number of beige/brown cells.

Summary

BAT is essential for energy homeostasis in rodents. The tissue is present in humans as well, but appears less important from a physiological and pathophysiological point of view. Thus, it is doubtful if human BAT can be used as a target for treating obesity and its comorbidities.

ADIPOSE TISSUE ENDOCRINE FUNCTIONS AND CHANGES IN OBESITY

It was already well-established in the 1960s that white adipose tissue releases glycerol, fatty acids, lactate, and perhaps some less abundant lipid species into the circulation. The pathophysiological role of these factors is described in detail under the lipid turnover section. In contrast, the endocrine function of adipose tissue and release of peptide factors and hormones, collectively termed “adipokines,” long remained unknown. Admittedly, parabiosis experiments in the 1950s using wild-type and severely obese *ob/ob* mice suggested that there was an adipose-derived satiety signal impacting on food intake. However, it was not until 1994, when leptin was cloned, that adipose tissue began to be considered a true endocrine organ.[76] Leptin is a peptide hormone belonging to the cytokine family and is primarily released from adipocytes in direct proportion to fat mass. One of its main actions is to attenuate hunger signals in the hypothalamus, and it therefore has a true endocrine function. Leptin is the gene that is disrupted in *ob/ob* animals, and leptin administration effectively normalizes food intake and body weight in animal models. Big clinical hopes were therefore put on leptin as a treatment modality in common forms of obesity. Unfortunately, leptin levels are very high in people with obesity, and the response to exogenous leptin is nonsignificant, possibly due to “leptin resistance.”[85] Leptin missense mutations are uncommon in humans but have been reported in a few families around the world. This results in massive obesity in early childhood, and in these cases recombinant leptin administration (metreleptin) is very effective. Leptin is also used in rare lipodystrophic conditions (with low circulating levels) and is believed to exert its action by reducing food and lipid intake, thereby normalizing the hypertriglyceridemia associated with these conditions.[86] More recently, metreleptin therapy has also been suggested for other leptin-deficient states, including anorexia nervosa.[87]

KEY POINTS

- Leptin remains, to date, the best characterized adipokine with a classical endocrine mechanism of action. While it does not display any relevant effects in the treatment of common forms of obesity, it has an important therapeutic role, particularly in rare forms of leptin deficiency.

The identification of leptin sparked an intense search for additional factors released by adipose tissue. Adipose tissue expresses up to one third of the proteins encoded by the genome, and this includes several hundred potentially secreted proteins. Proteomics screens of different human and murine adipocytes/adipose tissue systems have collectively identified greater than 600 released peptides/proteins.[88] This does not imply that all are actively secreted, as some factors may be released through cell necrosis/apoptosis or just under the culture conditions used *ex vivo*. Peptide factors released from adipose tissue include cyto-/chemokines (discussed in the immunometabolism section), growth factors (e.g., TGFβ), angiogenic factors (e.g., vascular endothelial growth factor), blood pressure regulators (e.g., angiotensinogen), and structural proteins (e.g., collagens). Importantly, with the exception of leptin and adiponectin (discussed later), very few of these “adipokines” are specific for adipose tissue, and most are expressed in several other organs. Furthermore, in addition to peptides/proteins, adipose tissue releases a large set of nonpolar and polar metabolites, apart from fatty acids, sphingolipids, and prostaglandins; these include, e.g., amino acids, lactate, and nitric oxide.[89] The role of some of these factors is briefly discussed later, but it is not possible to cover everything that is released/secreted from adipose tissue. This section will therefore highlight some of the most well-studied adipokines, which are also summarized in Table 23.4. Among these is adiponectin, a protein belonging to the complement factor family that is secreted almost exclusively from adipocytes. Adiponectin was cloned by several different groups shortly after the discovery of leptin and is secreted as a tri-, hexa-, or multimeric complex. It is now generally thought that the multimeric complex is the biologically active form and that it binds and activates two distinct receptors, AdipoR1 and AdipoR2, which display the highest expression in liver and skeletal muscle.[90] Adiponectin is released into the circulation, and is therefore considered an endocrine hormone, but in contrast to many other factors secreted from adipose tissue, the levels decrease with increased fat mass and insulin resistance. Adiponectin levels are therefore considered a good marker of metabolic risk. However, despite almost 25 years of research, the pathophysiological role of adiponectin in humans remains unclear, and most studies have been performed in murine models. In rodents, adiponectin improves insulin sensitivity, normalizes adipose tissue lipolysis, reduces hepatic gluconeogenesis, and increases skeletal muscle glucose uptake and fatty acid oxidation.[91] Despite these interesting effects, it has proven very difficult to translate this into efficient therapies in humans. Whether this is due to the complex structure of biologically active adiponectin, that it is more difficult to efficiently activate adiponectin receptors in humans, or that it simply plays another role in humans versus rodents is not clear.

Leptin and adiponectin remain to date the only true endocrine adipokines, i.e., factors released from the adipose tissue and exerting

TABLE 23.4 Adipokines

Factor, Main Function	Most Important Effects	Pathophysiology (Changes in Obesity)	Main Cellular Source in Adipose Tissue
Free fatty acids, endocrine factors	Increase very low density lipoprotein; gluconeogenesis; reduce glucose uptake in skeletal muscle; effects on insulin secretion?	Increased release primarily through non–hormone-stimulated lipolysis, reduced antilipolytic sensitivity to insulin contributes	Adipocytes
Leptin, endocrine factor	Reduced appetite; increased fatty acid oxidation; plays a role in reproduction	Increased release, primarily from subcutaneous depots; leptin resistance in obesity	Adipocytes
TNFα, autocrine or paracrine factor	Stimulates lipolysis and inhibits insulin action in fat cells	Increased	Macrophages and possibly other cells in the stroma vascular fraction
CCL2, paracrine factor	Chemoattractant factor that stimulates macrophage infiltration	Increased	Fat cells, immune cells, and other cells in the stroma vascular fraction
Adiponectin, endocrine factor	Protective against many pernicious metabolic effects	Reduced	Adipocytes
Interleukins, IL1-beta, IL6, and IL8, paracrine factors	Unclear role; promote an inflammatory role and may attenuate insulin signaling	Increased	Fat cells, immune cells, and other cells in the stroma vascular fraction
Acylstimulating protein, paracrine factor	Stimulates lipid deposition; attenuates fatty acid release	Increased	Fat cells, and other cells in the stroma vascular fraction
Plasminogenactivatorinhibitor-1, endo-, paracrine factor	Promotes thrombosis	Increased	Cells in the stroma vascular fraction
Visfatin, paracrine factor	Regulates inflammation	Increased	Macrophages
Osteopontin	Regulates inflammation	Increased	Fat cells, immune cells, and other cells in the stroma vascular fraction

KEY POINTS

- Adiponectin is well-studied, particularly in rodent models, where it plays important roles in regulating insulin sensitivity. However, its therapeutic role in humans remains unclear, and its druggability is hampered by the large size of the biologically active multimeric complexes.

effects through distinct receptors in another organ. Retinol-binding protein-4 (RBP4) is released from adipose tissue and is present at higher levels in the circulation of people with obesity/insulin resistance. RBP4 is a biomarker for insulin resistance and cardiometabolic risk, but to date an established mechanistic link has not been identified.[18] Virtually all other adipokines released from adipose tissue are believed to act locally in an auto- or paracrine fashion. Among these are, e.g., acylstimulating protein (ASP, also termed C3adesArg), plasminogen activator inhibitor-1 (PAI1), visfatin (also termed pre B-cell colony-enhancing factor [PBEF] or nicotinamide phosphoribosyltransferase [NAMPT]), and osteopontin, to name a few.

ASP is formed by enzymatic cleavage of the precursor protein C3 into C3a, initiated by the enzymes adipsin and factor B. ASP seems to play a role in triglyceride synthesis and hydrolysis, and increases fatty acid esterification and inhibits fatty acid release. The function of ASP is considerably attenuated in familial combined hyperlipidemia, although the mechanisms are not yet understood. A caveat in studying ASP is that rodents do not express relevant parts of the ASP system in adipose tissue, which has hampered further investigations.

PAI1 is a prothrombogenic factor released from several tissues, including adipose tissue, and the levels are increased in the obese state. As discussed elsewhere, it is also a marker of cellular senescence, and it has been speculated that it could play a role in linking obesity to cardiovascular complications.

Visfatin, which is identical to the previously described proteins PBEF and NAMPT, is an enzyme originally described as deriving from visceral fat macrophages. However, visfatin is also expressed in other organs, including muscle and liver. The expression of visfatin is associated with tissue inflammation and a pernicious metabolic profile, as well as atherosclerosis, but the casual links and the role of adipose tissue remain unclear.[92]

Osteopontin is a glycophosphoprotein expressed in several different cell types, including osteoblasts/osteocytes, where it is involved in bone metabolism and remodelling.[93] However, it is also expressed in adipose tissue, where it is believed to regulate the secretion of proinflammatory cytokines from different immune cells. Plasma concentrations of osteopontin and osteopontin mRNA in adipose tissue are upregulated in obesity and insulin resistance. Similar to many other adipokines, the causal role remains to be established.

Summary

White adipose tissue expresses at least two adipokines (leptin and adiponectin) with a clear endocrine function, in that they target other organs/tissue via specific receptors. Several other factors are released and present in the circulation, but these are not unique to adipose tissue *per se*, and the contribution to the circulating concentrations versus other peripheral organs is not yet clear. The latter can only be determined by comparing arteriovenous differences in vessels going to and from different tissues using clinical techniques that only very few groups in the world can perform. In some cases the receptors or signaling mechanisms of these adipokines remain to be elucidated. Most of the adipokines released from adipose tissue are believed to act locally and seem to play important roles in determining tissue function and lipid storage capacity. Obesity perturbs the expression of adipokines and in most cases leads to an increase in the expression of, e.g., leptin, proinflammatory proteins, and extracellular matrix genes.

TABLE 23.5 Adipose Tissue–Targeted Therapies for Human Obesity and its Comorbidities in Current Use

Treatment	Action
Restricted energy intake	1. Reduce adipose mass 2. Bariatric surgery long-term effective 3. Metabolic effect related to decrease fat cell size
Surgical removal of adipose tissue	1. Omentectomy has metabolic effects in connection with gastric banding, but not together with gastric bypass 2. Abdominal subcutaneous liposuction has no metabolic effects
Activation of brown adipose tissue	1. No important effect on body weight 2. Small metabolic effects
Glucagon-like peptides	1. Primarily used for diabetes treatment 2. Reduce body weight, sometimes effectively, and are currently used for this purpose. Probably have an indirect effect on adipose tissue
Thiazolidinediones	3. Increased subcutaneous adipose lipid storage causing decreased fatty acid levels 4. Antidiabetic, but risk of important side effects such as weight gain
Other antidiabetic agents	1. Biguanides may lead to some weight reduction 2. Sodium-glucose cotransporter 2 inhibitors may cause modest weight reduction 3. Probably indirect effects on adipose tissue
Nicotinic acids	1. Antilipolytic causing improved lipid profile but no change in body weight 2. May cause glucose intolerance
Endurance training	1. Enhances catecholamine-induced lipolysis 2. Long-term effects on body weight and metabolism are unknown
Leptin	1. Decreased appetite. Increased energy expenditure 2. Not effective in common obesity
Antibodies against TNFα	1. Neutralizing adipose inflammation 2. No clear effect on diabetes or insulin sensitivity 3. Body weight neutral

The exception to this rule is adiponectin, which is one of the few adipokines negatively regulated by increased fat mass. As mentioned, the role of different locally released metabolites is not discussed herein, but it is interesting to note that several of them (e.g., lactate, succinate, and beta-hydroxybutyrate) have been shown to bind and activate G-protein coupled receptors, which are expressed on both adipocytes and immune cells.[94] This implies that the number of “adipokines” could be substantially higher than what has so far been appreciated. Moreover, the levels of several amino acids (e.g., glutamine, branched-chain amino acids) and their derivatives (e.g., kynurenic acid) are altered in the obese state and impact adipose tissue function via both epigenetic (see immunometabolism section) and receptor-mediated pathways. This area of research is therefore still very active, and novel insights are reported continuously.

ADIPOSE TISSUE AS A THERAPEUTIC TARGET

Bearing in mind the multiple actions of adipose tissue, numerous site-directed therapeutic modalities are possible and have been tested in animal models. Those tried in human obesity and its comorbidities will be discussed and are summarized in Table 23.5. The most obvious approach is to decrease fat mass by means of reduced energy intake. Different diets have been publicly available since Banting published a formula based on reduced intake of carbohydrates and starch in 1863.[95] A weight reduction of 5% or more usually leads to metabolic improvements, which could be further boosted in combination with enhanced physical activity. However, no diet is known to exert long-term effects on obesity remission in the general population.[95] In fact, often the end result is higher body weight compared with the weight before starting the diet (“yo-yo” effect). In contrast, bariatric surgery leads to long-term reduction in fat mass and is accompanied by positive effects on diabetes incidence/remission, mortality, and several other obesity complications.[96] There are different surgical procedures. Earlier techniques are technically easier but result in less marked weight loss. The most common current operation procedures, such as gastric bypass and sleeve gastrectomy, are much more effective in terms of weight loss and are also sufficiently safe.[97] After an initial and marked weight loss, body weight increases gradually to a new long-term steady state, which is approximately 25% below the preoperative body weight. The mechanism behind the beneficial effects of body weight loss are not well understood but are likely to be multifaceted. Improved insulin sensitivity is linked more closely to diminished fat cell size than reduced fat mass *per se*.[98] In theory, surgical removal of adipose tissue is an attractive antiobesity therapy. However, large amounts adipose tissue cannot be safely removed, due to immediate and long-term risks of severe complications. Therefore, selective pernicious adipose areas that are relatively easy to remove have been taken out since the turn of the last century. The major omentum has been removed in connection with bariatric surgery.[99] Positive effects were reported in connection to gastric binding but with gastric bypass. The reason for this difference might be the stronger effect on weight reduction of the latter surgical procedure. Liposuction is often used for cosmetic surgery, but removal of approximately 10 kg of abdominal SAT in obesity had no effect on the cardiometabolic risk profile, on either a short- or long-term basis.[100]

KEY POINTS

- So far bariatric surgery is the only proven effective means of reducing body weight for a long time in subjects with common forms of obesity. Even modest reductions in body weight have a positive effect on metabolic obesity complications. The ameliorating effect of weight reduction is not likely to be due to diminished fat mass *per se*. Instead, the reduced fat cell volume is likely to play such a role.

Pharmacotherapy in Obesity

The approved antiobesity medications are covered in another chapter. Herein only adipose-specific therapies are discussed in some detail. Although many pharmacological agents have been tried, usually as a means to suppress appetite, most have been withdrawn from the market or are no longer used due to side effects.[101] These agents act centrally, and are thus not important for fat cell–mediated actions. Glucagon-like peptide 1 receptor agonists have been shown to induce long-term, and in some cases marked, body weight reduction. Although receptors for the hormone are present in human fat cells and cause metabolic effects, they cannot be linked to effects on body weight. Biguanides may also reduce body weight, but the effect is usually modest (circa 3%). Although metabolic/endocrine effects on human adipose tissue are described, they are not important for the effect on body weight, and they are not approved for weight reduction

Activation of BAT through stimulation with beta-3 adrenoceptor agonists has been tested clinically.[102] This issue was thoroughly discussed earlier.

TZDs are enhancers of insulin sensitivity and have been used to treated type 2 diabetes in obese/overweight patients.[103] Much of their antidiabetic effect is caused by activation of the nuclear receptor PPARγ in fat cells. This leads to retention of lipids, predominantly in SAT, and consequent lowering of circulating fatty acid levels, which, in turn, improves insulin action/secretion. However, these agents currently have limited use due to side effects such as weight gain (due to hyperplastic expansion of SAT), fluid retention, osteoporosis, and (at least for older agents) adverse cardiac events.

Nicotinic acid and its derivatives have long been used to treat hypertriglyceridemia, which often accompanies obesity.[104] One mechanism is through inhibition of fat cell lipolysis so that fewer fatty acids are delivered to the liver for synthesis of very low density lipoproteins. These drugs are body weight–neutral but may cause glucose intolerance, and they have been withdrawn from the market due to side effects.

The effects of exercise training on human adipose tissue function has been studied for many years.[105] If not compensated by increased energy intake, some body weight loss can be expected. Several exercise effects on adipose endocrine and metabolic functions, as well as on inflammation, have been reported. Most promising from a therapeutic point of view is that endurance training induces a long-term increase in the lipolytic effect of catecholamines. However, long-term effects of enhanced physical activity alone on body weight and metabolic profile are not documented.

As regards endocrine function and inflammation, only leptin and TNFα have been intensely investigated in humans to date.[90,106] These therapies were discussed thoroughly earlier.

KEY POINTS

- In theory, a number of fat cell–specific regulatory factors such as adipocyte differentiation, beta oxidation, lipolysis, and the endocrine functions are potential pharmaceutical targets for combatting obesity and dysregulated metabolism. So far, this therapeutic area is not very advanced; the only important exception hitherto are leptin and thiazolidinediones.

Summary

Other than bariatric surgery, no alternative adipose-targeted therapies have so far demonstrated clear-cut and long-term effectiveness in the treatment of obesity and its comorbidities in the general population. It is possible, though, that agents and their analogs derived from the gastrointestinal tract can be used as fruitful alternatives to bariatric surgery. Some important targets according to mouse models, such as adiponectin and hormone-sensitive lipase, have not yet been tested in humans. In the future, gene therapy might be useful. For example, several long noncoding RNAs are fat cell–specific and functional, and could be useful tools to improve fat cell function.[107]

REFERENCES

1. Ward ZJ, Bleich SN, Cradock AL, et al. Projected U.S. State-level prevalence of adult obesity and severe obesity. *N Engl J Med*. 2019;381:2440–2450.
2. Global BMIMC, Di Angelantonio E, Bhupathiraju Sh N, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*. 2016;388:776–786.
3. IDF Diabetes Atlas. 9th ed. Brussels, Belgium: International Diabetes Federation;2019.
4. Spalding KL, Arner E, Westermark PO, et al. Dynamics of fat cell turnover in humans. *Nature*. 2008;453:783–787.
5. Tchoukalova YD, Votruba SB, Tchkonia T, et al. Regional differences in cellular mechanisms of adipose tissue gain with overfeeding. *Proc Natl Acad Sci U S A*. 2010;107:18226–18231.
6. Hammarstedt A, Gogg S, Hedjazifar S, et al. Impaired adipogenesis and dysfunctional adipose tissue in human hypertrophic obesity. *Physiol Rev*. 2018;98:1911–1941.
7. Tchernof A, Despres JP. Pathophysiology of human visceral obesity: an update. *Physiol Rev*. 2013;93:359–404.
8. Ross R, Neeland IJ, Yamashita S, et al. Waist circumference as a vital sign in clinical practice: a Consensus statement from the IAS and ICCR working group on visceral obesity. *Nat Rev Endocrinol*. 2020;16:177–189.
9. Smith U. Abdominal obesity: a marker of ectopic fat accumulation. *J Clin Invest*. 2015;125:1790–1792.
10. Hoffstedt J, Andersson DP, Eriksson Hogling D, et al. Long-term protective changes in adipose tissue after gastric bypass. *Diabetes Care*. 2017;40:77–84.
11. Rawshani A, Eliasson B, Rawshani A, et al. Adipose tissue morphology, imaging and metabolomics predicting cardiometabolic risk and family history of type 2 diabetes in non-obese men. *Sci Rep*. 2020;10:9973.
12. Caleyachetty R, Thomas GN, Toulis KA, et al. Metabolically healthy obese and incident cardiovascular disease events among 3.5 million men and women. *J Am Coll Cardiol*. 2017;70:1429–1437.
13. Small KS, Todorcevic M, Civelek M, et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat Genet*. 2018;50:572–580.
14. Xie W, Li L, Zheng XL, et al. The role of Kruppel-like factor 14 in the pathogenesis of atherosclerosis. *Atherosclerosis*. 2017;263:352–360.
15. Lotta LA, Gulati P, Day FR, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*. 2017;49:17–26.
16. Cederberg H, Stancakova A, Kuusisto J, et al. Family history of type 2 diabetes increases the risk of both obesity and its complications: is type 2 diabetes a disease of inappropriate lipid storage? *J Intern Med*. 2015;277:540–551.
17. Arner P, Arner E, Hammarstedt A, et al. Genetic predisposition for Type 2 diabetes, but not for overweight/obesity, is associated with a restricted adipogenesis. *PLoS One*. 2011;6:e18284.
18. Smith U, Kahn BB. Adipose tissue regulates insulin sensitivity: role of adipogenesis, de novo lipogenesis and novel lipids. *J Intern Med*. 2016;280:465–475.
19. Acosta JR, Douagi I, Andersson DP, et al. Increased fat cell size: a major phenotype of subcutaneous white adipose tissue in non-obese individuals with type 2 diabetes. *Diabetologia*. 2016;59:560–570.
20. Ryden M, Uzunel M, Hard JL, et al. Transplanted bone marrow-derived cells contribute to human adipogenesis. *Cell Metab*. 2015;22:408–417.
21. Qian SW, Tang Y, Li X, et al. BMP4-mediated brown fat-like changes in white adipose tissue alter glucose and energy homeostasis. *Proc Natl Acad Sci U S A*. 2013;110:E798–E807.

22. Tseng YH, Kokkotou E, Schulz TJ, et al. New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature*. 2008;454:1000–1004.
23. Arner E, Mejhert N, Kulyte A, et al. Adipose tissue microRNAs as regulators of CCL2 production in human obesity. *Diabetes*. 2012;61:1986–1993.
24. Gustafson B, Hammarstedt A, Hedjazifar S, et al. Restricted adipogenesis in hypertrophic obesity: the role of WISP2, WNT, and BMP4. *Diabetes*. 2013;62:2997–3004.
25. Dahlman I, Ryden M, Brodin D, et al. Numerous genes in loci associated with body fat distribution are linked to adipose function. *Diabetes*. 2016;65:433–437.
26. Yaghootkar H, Lotta LA, Tyrrell J, et al. Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. *Diabetes*. 2016;65:2448–2460.
27. Majithia AR, Flannick J, Shahinian P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A*. 2014;111:13127–13132.
28. Balas B, Belfort R, Harrison SA, et al. Pioglitazone treatment increases whole body fat but not total body water in patients with non-alcoholic steatohepatitis. *J Hepatol*. 2007;47:565–570.
29. Campisi J, Kapahi P, Lithgow GJ, et al. From discoveries in ageing research to therapeutics for healthy ageing. *Nature*. 2019;571:183–192.
30. Baker DJ, Childs BG, Durik M, et al. Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. *Nature*. 2016;530:184–189.
31. Schafer MJ, White TA, Evans G, et al. Exercise prevents diet-induced cellular senescence in adipose tissue. *Diabetes*. 2016;65:1606–1615.
32. Khosla S, Farr JN, Tchkonia T, et al. The role of cellular senescence in ageing and endocrine disease. *Nat Rev Endocrinol*. 2020;16:263–275.
33. Arner E, Westermark PO, Spalding KL, et al. Adipocyte turnover: relevance to human adipose tissue morphology. *Diabetes*. 2010;59:105–109.
34. Gustafson B, Nerstedt A, Smith U. Reduced subcutaneous adipogenesis in human hypertrophic obesity is linked to senescent precursor cells. *Nat Commun*. 2019;10:2757.
35. Gogg S, Nerstedt A, Boren J, et al. Human adipose tissue microvascular endothelial cells secrete PPARgamma ligands and regulate adipose tissue lipid uptake. *JCI Insight*. 2019;4:1–14.
36. Briot A, Decaunes P, Volat F, et al. Senescence alters PPARgamma (peroxisome proliferator-activated receptor gamma)-dependent fatty acid handling in human adipose tissue microvascular endothelial cells and favors inflammation. *Arterioscler Thromb Vasc Biol*. 2018;38:1134–1146.
37. Reilly SM, Saltiel AR. Adapting to obesity with adipose tissue inflammation. *Nat Rev Endocrinol*. 2017;13:633–643.
38. Kratz M, Coats BR, Hisert KB, et al. Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. *Cell Metab*. 2014;20:614–625.
39. Halberg N, Khan T, Trujillo ME, et al. Hypoxia-inducible factor 1alpha induces fibrosis and insulin resistance in white adipose tissue. *Mol Cell Biol*. 2009;29:4467–4483.
40. Vogel MAA, Jocken JWE, Sell H, et al. Differences in upper and lower body adipose tissue oxygen tension contribute to the adipose tissue phenotype in humans. *J Clin Endocrinol Metab*. 2018;103:3688–3697.
41. Xu H, Barnes GT, Yang Q, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest*. 2003;112:1821–1830.
42. Wernstedt Asterholm I, Tao C, Morley TS, et al. Adipocyte inflammation is essential for healthy adipose tissue expansion and remodeling. *Cell Metab*. 2014;20:103–118.
43. Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. *Nature*. 2017;542:177–185.
44. Paquot N, Castillo MJ, Lefebvre PJ, et al. No increased insulin sensitivity after a single intravenous administration of a recombinant human tumor necrosis factor receptor: fc fusion protein in obese insulin-resistant patients. *J Clin Endocrinol Metab*. 2000;85:1316–1319.
45. Larsen CM, Faulenbach M, Vaag A, et al. Interleukin-1-receptor antagonist in type 2 diabetes mellitus. *N Engl J Med*. 2007;356:1517–1526.
46. Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med*. 2017;377:1119–1131.
47. Everett BM, Donath MY, Pradhan AD, et al. Anti-inflammatory therapy with canakinumab for the prevention and management of diabetes. *J Am Coll Cardiol*. 2018;71:2392–2401.
48. Donath MY, Dinarello CA, Mandrup-Poulsen T. Targeting innate immune mediators in type 1 and type 2 diabetes. *Nat Rev Immunol*. 2019;19:734–746.
49. Shimobayashi M, Albert V, Woelnerhanssen B, et al. Insulin resistance causes inflammation in adipose tissue. *J Clin Invest*. 2018;128:1538–1550.
50. Crewe C, An YA, Scherer PE. The ominous triad of adipose tissue dysfunction: inflammation, fibrosis, and impaired angiogenesis. *J Clin Invest*. 2017;127:74–82.
51. Petrus P, Lecoutre S, Dollet L, et al. Glutamine links obesity to inflammation in human white adipose tissue. *Cell Metab*. 2020;31:375–390 e11.
52. Morigny P, Boucher J, Arner P, et al. Lipid and glucose metabolism in white adipocytes: pathways, dysfunction and therapeutics. *Nat Rev Endocrinol*. 2021;17:276–295.
53. Luo L, Liu M. Adipose tissue in control of metabolism. *J Endocrinol*. 2016;231:R77–R99.
54. Thiam AR, Beller M. The why, when and how of lipid droplet diversity. *J Cell Sci*. 2017;130:315–324.
55. White U, Ravussin E. Dynamics of adipose tissue turnover in human metabolic health and disease. *Diabetologia*. 2019;62:17–23.
56. Chouchani ET, Kajimura S. Metabolic adaptation and maladaptation in adipose tissue. *Nat Metab*. 2019;1:189–200.
57. Karpe F, Pinnick KE. Biology of upper-body and lower-body adipose tissue–link to whole-body phenotypes. *Nat Rev Endocrinol*. 2015;11:90–100.
58. Oikonomou EK, Antoniades C. The role of adipose tissue in cardiovascular health and disease. *Nat Rev Cardiol*. 2019;16:83–99.
59. Walker GE, Marzullo P, Ricotti R, et al. The pathophysiology of abdominal adipose tissue depots in health and disease. *Horm Mol Biol Clin Investig*. 2014;19:57–74.
60. Santosa S, Jensen MD. The sexual dimorphism of lipid kinetics in humans. *Front Endocrinol (Lausanne)*. 2015;6:103.
61. Jensen MD. Visceral fat: culprit or canary? *Endocrinol Metab Clin North Am*. 2020;49:229–237.
62. Spalding KL, Bernard S, Naslund E, et al. Impact of fat mass and distribution on lipid turnover in human adipose tissue. *Nat Commun*. 2017;8:15253.
63. Arner P, Bernard S, Appelsved L, et al. Adipose lipid turnover and long-term changes in body weight. *Nat Med*. 2019;25:1385–1389.
64. McQuaid SE, Hodson L, Neville MJ, et al. Downregulation of adipose tissue fatty acid trafficking in obesity: a driver for ectopic fat deposition? *Diabetes*. 2011;60:47–55.
65. Hames KC, Koutsari C, Santosa S, et al. Adipose tissue fatty acid storage factors: effects of depot, sex and fat cell size. *Int J Obes*. 2015;39:884–887.
66. Andrade AC. Lipoprotein lipase: a general review. *Insights Enzyme Res*. 2018;2:1–14.
67. Zechner R, Madeo F, Kratky D. Cytosolic lipolysis and lipophagy: two sides of the same coin. *Nat Rev Mol Cell Biol*. 2017;18:671–684.
68. Arner P, Langin D. Lipolysis in lipid turnover, cancer cachexia, and obesity-induced insulin resistance. *Trends Endocrinol Metab*. 2014;25:255–262.
69. Nielsen TS, Jessen N, Jorgensen JO, et al. Dissecting adipose tissue lipolysis: molecular regulation and implications for metabolic disease. *J Mol Endocrinol*. 2014;52:R199–R222.
70. Stenkula KG, Erlanson-Albertsson C. Adipose cell size: importance in health and disease. *Am J Physiol Regul Integr Comp Physiol*. 2018;315:R284–R295.
71. Morigny P, Houssier M, Mouisel E, et al. Adipocyte lipolysis and insulin resistance. *Biochimie*. 2016;125:259–266.
72. Arner P, Andersson DP, Backdahl J, et al. Weight gain and impaired glucose metabolism in women are predicted by inefficient subcutaneous fat cell lipolysis. *Cell Metab*. 2018;28:45–54 e3.
73. Goodpaster BH, Sparks LM. Metabolic flexibility in health and disease. *Cell Metab*. 2017;25:1027–1036.
74. Rabinowitz JD, Enerback S. Lactate: the ugly duckling of energy metabolism. *Nat Metab*. 2020;2:566–571.

75. Mann JP, Savage DB. What lipodystrophies teach us about the metabolic syndrome. *J Clin Invest*. 2019;129:4009–4021.
76. Rosen ED, Spiegelman BM. What we talk about when we talk about fat. *Cell*. 2014;156:20–44.
77. Cinti S. UCP1 protein: the molecular hub of adipose organ plasticity. *Biochimie*. 2017;134:71–76.
78. Cannon B, de Jong JMA, Fischer AW, et al. Human brown adipose tissue: classical brown rather than brite/beige? *Exp Physiol*. 2020;105:1191–1200.
79. Zoico E, Rubele S, De Caro A, et al. Brown and beige adipose tissue and aging. *Front Endocrinol*. 2019;10:368.
80. Baskin AS, Linderman JD, Brychta RJ, et al. Regulation of human adipose tissue activation, gallbladder size, and bile acid metabolism by a beta3-adrenergic receptor agonist. *Diabetes*. 2018;67:2113–2125.
81. Choi SS, Kim ES, Jung JE, et al. PPARgamma antagonist gleevec improves insulin sensitivity and promotes the browning of white adipose tissue. *Diabetes.* 2016;65:829–839.
82. Gustafson B, Hammarstedt A, Hedjazifar S, et al. BMP4 and BMP antagonists regulate human white and beige adipogenesis. *Diabetes*. 2015;64:1670–1681.
83. Hoffmann JM, Grunberg JR, Church C, et al. BMP4 gene therapy in mature mice reduces bat activation but protects from obesity by browning subcutaneous adipose tissue. *Cell Rep*. 2017;20:1038–1049.
84. Maurer S, Harms M, Boucher J. The colorful versatility of adipocytes: white-to-brown transdifferentiation and its therapeutic potential in man. *FEBS J*. 2020;288:3628–3646.
85. Farr OM, Gavrieli A, Mantzoros CS. Leptin applications in 2015: what have we learned about leptin and obesity? *Curr Opin Endocrinol Diabetes Obes*. 2015;22:353–359.
86. Sollier C, Vatier C, Capel E, et al. Lipodystrophic syndromes: from diagnosis to treatment. *Ann Endocrinol (Paris)*. 2020;81:51–60.
87. Hebebrand J, Milos G, Wabitsch M, et al. Clinical trials required to assess potential benefits and side effects of treatment of patients with anorexia nervosa with recombinant human leptin. *Front Psychol*. 2019;10:769.
88. Kim EY, Kim WK, Oh KJ, et al. Recent advances in proteomic studies of adipose tissues and adipocytes. *Int J Mol Sci*. 2015;16:4581–4599.
89. Scheja L, Heeren J. The endocrine function of adipose tissues in health and cardiometabolic disease. *Nat Rev Endocrinol*. 2019;15:507–524.
90. van Andel M, Heijboer AC, Drent ML. Adiponectin and its isoforms in pathophysiology. *Adv Clin Chem*. 2018;85:115–147.
91. Funcke JB, Scherer PE. Beyond adiponectin and leptin: adipose tissue-derived mediators of inter-organ communication. *J Lipid Res*. 2019;60:1648–1684.
92. Audrito V, Messana VG, Deaglio S. NAMPT and NAPRT: two metabolic enzymes with key roles in inflammation. *Front Oncol*. 2020;10:358.
93. Icer MA, Gezmen-Karadag M. The multiple functions and mechanisms of osteopontin. *Clin Biochem*. 2018;59:17–24.
94. Husted AS, Trauelsen M, Rudenko O, et al. GPCR-mediated signaling of metabolites. *Cell Metab*. 2017;25:777–796.
95. Dombrowski SU, Knittle K, Avenell A, et al. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. *BMJ*. 2014;348:g2646.
96. Xia Q, Campbell JA, Ahmad H, et al. Bariatric surgery is a cost-saving treatment for obesity-A comprehensive meta-analysis and updated systematic review of health economic evaluations of bariatric surgery. *Obes Rev*. 2020;21:e12932.
97. Chang SH, Stoll CR, Song J, et al. The effectiveness and risks of bariatric surgery: an updated systematic review and meta-analysis, 2003-2012. *JAMA Surg*. 2014;149s:275–287.
98. Andersson DP, Eriksson Hogling D, Thorell A, et al. Changes in subcutaneous fat cell volume and insulin sensitivity after weight loss. *Diabetes Care*. 2014;37:1831–1836.
99. Lee Y, Pedziwiatr M, Major P, et al. The effect of omentectomy added to bariatric surgery on metabolic outcomes: a systematic review and meta-analysis of randomized controlled trials. *Surg Obes Relat Dis*. 2018;14:1766–1782.
100. Wu S, Coombs DM, Gurunian R. Liposuction: concepts, safety, and techniques in body-contouring surgery. *Cleve Clin J Med*. 2020;87:367–375.
101. May M, Schindler C, Engeli S. Modern pharmacological treatment of obese patients. *Ther Adv Endocrinol Metab*. 2020;11:2042018819897527.
102. Trayhurn P. Brown adipose tissue-a therapeutic target in obesity? *Front Physiol*. 2018;9:1672.
103. Soccio RE, Chen ER, Lazar MA. Thiazolidinediones and the promise of insulin sensitization in type 2 diabetes. *Cell Metab*. 2014;20:573–591.
104. Garg A, Sharma A, Krishnamoorthy P, et al. Role of niacin in current clinical practice: a systematic review. *Am J Med*. 2017;130:173–187.
105. Giolo De Carvalho F, Sparks LM. Targeting white adipose tissue with exercise or bariatric surgery as therapeutic strategies in obesity. *Biology (Basel)*. 2019;15:16.
106. Paz-Filho G, Mastronardi CA, Licinio J. Leptin treatment: facts and expectations. *Metabolism*. 2015;64:146–156.
107. Squillaro T, Peluso G, Galderisi U, et al. Long non-coding RNAs in regulation of adipogenesis and adipose tissue function. *Elife*. 2020;9:e, 1–15.

24

Lipodystrophy Syndromes

Angeliki M. Angelidi, Michael A. Tsoukas, and Christos S. Mantzoros

OUTLINE

Lipodystrophies: Definition and Diagnosis, 379
Pathophysiology of Lipodystrophy, 380
Classifications of Lipodystrophies and Their Clinical Manifestations, 381
  Generalized Lipodystrophies, 381
  Other Mutations Associated With a Congenital Generalized Lipodystrophy–Like Phenotype, 383
  Partial Lipodystrophies, 384
  Localized Lipodystrophies, 389
Mechanisms Responsible for Severe Insulin Resistance, 389
  Fat Redistribution and Fat Metabolism, 389
  Adipocytokines, 389
  Inflammation, 390
  Mitochondrial Stress, Oxidative Stress, and the Endoplasmic Reticulum, 390
  Other Mechanisms, 390
Treatment of Syndromes of Lipodystrophies, 390
  Lifestyle Modification, 390
  Management of Insulin Resistance, 391
  Management of Dyslipidemia, 392
  Management of Human Immunodeficiency Virus–Infected Patients With Highly Active Antiretroviral Therapy–Induced Metabolic Syndrome, 393
  Management of Cosmetic Appearance, 394
  Adipokines in Lipodystrophy, 394
Future Perspectives, 395
  Adiponectin, 395

KEY POINTS

- Lipodystrophy syndromes are a heterogeneous group of disorders characterized by complete or partial lack of adipose tissue in certain areas of the body, with excess of adipose tissue elsewhere.
- The etiology of lipodystrophy may be either a congenital or acquired and involves a loss of mature, functional adipocytes associated with failure of adipogenesis, adipocyte apoptosis, or storage of triglycerides.
- Congenital lipodystrophy is caused by abnormal gene activations and signaling that impair adipocyte differentiation, whereas acquired lipodystrophy is most commonly observed in patients with human immunodeficiency virus who are receiving highly active antiretroviral therapy.
- Metabolic complications, such as insulin resistance, dyslipidemia, hepatic steatosis, endovascular inflammation, and oxidative stress, are associated with lipodystrophy syndromes and constitute major comorbidities of these syndromes.
- Current treatments such as metformin, thiazolidinediones, growth hormone replacement, growth hormone–releasing hormone analogs, and US Food and Drug Administration–approved recombinant leptin therapy may help with clinical features and associated metabolic complications of lipodystrophy.

LIPODYSTROPHIES: DEFINITION AND DIAGNOSIS

Lipodystrophy syndromes are a diverse group of rare clinical disorders, the central feature of which is either a congenital or acquired complete or partial lack of adipose tissue (lipoatrophy) and/or a combination of lack of adipose tissue in certain body areas, with excess of adipose tissue (lipohypertrophy) elsewhere.[1] Despite progress in identifying the genetic basis and pathways of the disease, lipodystrophy is often underestimated or misdiagnosed because of its heterogeneity and rarity. According to the most recent multi-society practice guidelines, diagnosis of a lipodystrophy disorder is based on patient's history, clinical examination, metabolic state, and body composition.[2]

Metabolic disorders, and especially a severe form of the metabolic syndrome, are also implicated in this abnormal deposition of adipose tissue that cannot be stored in the appropriate subcutaneous depots.[3] Abnormalities such as insulin resistance and its associated clinical features are present in nearly all varieties of lipodystrophies.[4] In addition, patients with lipodystrophies also manifest a group of unique features such as hypertriglyceridemia and severe mixed dyslipidemia, progressive hepatic steatosis, progression to poorly controlled diabetes, and an increased metabolic rate (Fig. 24.1).[5] Severe dysmetabolic situations associated with lipodystrophy syndromes are responsible for the increased risk of severe comorbidities and mortality in these patients.[6]

No firm diagnostic criteria currently exist, and a lipodystrophy disorder is usually recognized clinically, often based on both patients' and doctors' perceptions and assessment. A meticulous physical examination is required and can be further supported by anthropometry, dual energy x-ray absorptiometry (DEXA), computed tomography (CT), and whole-body magnetic resonance imaging (MRI).

Anthropometry, or measurement of skin folds and limb circumference, is an easy, affordable, and practical way to estimate fat loss and fat redistribution. However, its reliability is heavily dependent on the consistency and the skill of the examiner(s), and it may have poor sensitivity, especially during early stages of the disease. Objective measurement of facial lipoatrophy also poses a challenge. Serial photographs (with the patient's consent) have been used to document and compare the facial wasting over time.

DEXA, CT, and MRI scans are additional modalities that have been developed to allow direct quantification of fat within specific tissues and/or body mass, as well as fat distribution. Although all these techniques are accurate and noninvasive, they are also expensive, have limited availability, and are not cost-effective for use in everyday clinical practice. Their use, therefore, has been limited to the research field.[7]

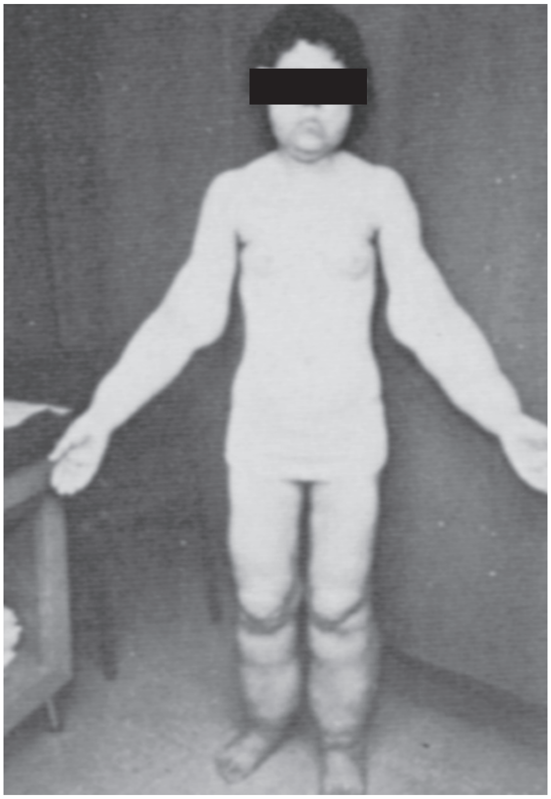


Fig. 24.1 Congenital lipodystrophy.

In the past few years, ultrasound has emerged as a promising alternative quantitative tool to assess body fat changes. Although it demonstrates good accuracy and accessibility, more studies are needed to elucidate its value and to comparatively evaluate ultrasound with other imaging modalities. Standardization of ultrasound techniques and analyses of cost-effectiveness will also be essential for its potential wider clinical applications.[8] Table 24.1 summarizes the main advantages and disadvantages of the aforementioned body composition and adipose tissue imaging techniques.

Laboratory testing and imaging are valuable tools for physicians to support their diagnosis and determine the type and severity of lipodystrophy. Patients with lipodystrophy tend to have low serum leptin levels. It should be noted that no defined serum leptin level threshold can be used to confirm or exclude the diagnosis of lipodystrophy.[2,9,10] Genetic testing is the gold standard and may assist in clinical diagnosis confirmation.[11] However, because of limited resources and DNA sequencing availability, the current recommendations suggest that genetic testing should only be considered in suspected familial lipodystrophies and at-risk family members.[2] It should be underlined that negative genetic tests cannot rule out the presence of a genetic lipodystrophy syndrome.

PATHOPHYSIOLOGY OF LIPODYSTROPHY

In previous decades, the study of mechanisms underlying lipodystrophy has attracted significant attention, to a certain extent owing to the interest of the scientific community in obesity research. It is now recognized that white adipose tissue is not an inert storage depot organ but an active endocrine organ that plays a critical role in regulating energy homeostasis and mediating inflammation.[12] Because there are common features in the etiopathogenesis of lipodystrophy and obesity, lessons learned from studies of the lipodystrophies may provide essential information not only for the management of these rather rare cases, but also for obesity research and management.

Much of our knowledge regarding the mechanisms underlying the pathogenesis and manifestations of lipodystrophies has been obtained

TABLE 24.1 **Advantages and Disadvantages of Different Body Composition and Adipose Tissue Imaging Techniques**

Ultrasound	Dual Energy X-Ray Absorptiometry	Computed Tomography	Magnetic Resonance Imaging
Advantages			
No radiation exposure Safe for repeated measurements Satisfactory reliability Easy to operate Portable and easily available Low cost	Low radiation exposure Well tolerated for repeated measures High accuracy and precision Information on total and regional percentage of fat Information on fat and fat-free mass Moderate cost	High accuracy and precision Information on specific body fat compartments Assessment of subcutaneous and visceral fat and fat-free mass	No radiation exposure Safe for repeated measurements High accuracy and precision Information on specific body fat compartments Body mass composition differentiation, assessment of subcutaneous and visceral fat
Disadvantages			
Low accuracy Interpretation is more difficult, subjective, and affected by excessive edema Experience-related No validated procedure or clinical protocols	Moderate cost Not portable Radiation exposure (even low) Hard to distinguish subcutaneous and visceral fat Interpretation is affected by fat-free mass hydration status and body thickness Experience-related Variability among different versions and manufacturers	High cost Not portable Limited availability Radiation exposure Need for specific software (for fat-free mass) and technical skill for image analysis	High cost Not portable Limited availability Requires specific software Long scan duration

through mouse studies and as a result of human genome sequencing. It is now understood that patients with lipodystrophy have primarily a loss of mature, functional adipocytes, as opposed to an absence of lipids in otherwise normal adipocytes.[13,14,15] The underlying defects could be associated with failure of adipogenesis, adipocyte apoptosis, or a failure to store triglycerides in existing adipocytes because of ineffective lipogenesis or excessive lipolysis.[16] Differentiation of adipocytes is controlled by a variety of mitogenic and adipogenic gene activations and signaling pathways. Abnormal development and differentiation of preadipocytes and adipocytes, as well as increased adipocyte apoptosis, may lead to the development of a particular lipodystrophy (Fig 24.2).[16] Furthermore, the overall alterations in adiposity may result in the metabolic derangements and subsequent complications observed in patients with lipodystrophies.[17,18,19]

In this chapter, we will discuss the classifications of different types of lipodystrophies, their distinctive clinical presentations, our current understanding of the underlying mechanisms, and the recommended treatment modalities.

CLASSIFICATIONS OF LIPODYSTROPHIES AND THEIR CLINICAL MANIFESTATIONS

The classification of lipodystrophies is based on distinct clinical presentation, unique patterns of adipose tissue deficiency (namely, generalized or partial lipodystrophy), and manner of acquisition (inherited and acquired forms). Thus, lipodystrophies are categorized into four major categories: congenital generalized lipodystrophy (CGL), acquired generalized lipodystrophy (AGL), familial partial lipodystrophy (FPL), and acquired partial lipodystrophy (APL). However, lipodystrophy syndromes include additional subtypes such as the progeroid syndromes, and it is likely that this classification may become inadequate as more genetic causes and pathophysiological mechanisms associated with lipodystrophy are identified.[11]

Generalized Lipodystrophies

Generalized lipodystrophy encompasses rare but clinically striking disorders that may be congenital (Berardinelli–Seip syndrome)[20,21] or acquired (Lawrence syndrome).

Congenital Generalized Lipodystrophy. CGL, or Berardinelli-Seip congenital lipodystrophy (BSCL), was initially described in 1954 by Berardinelli in Brazil in two children, and 5 years later by Seip in Norway in three other patients.[20,21] It is a rare syndrome that inherited in an autosomal recessive manner and is observed mainly in cases of parental consanguinity or in isolated communities, probably because of "founder effect."[22] It is characterized by near-complete absence of body fat that may spare the supportive adipose tissue in the palms, soles, joints, and the area under the scalp, depending on the underlying genetic cause.[23] CGL is also associated with the early onset and progressive development of several metabolic disorders and comorbidities.[6,24] To date, it has been reported in approximately 300 patients of various ethnic backgrounds.[24-29]

Babies with CGL are noted to have an abnormal appearance owing to absence of body fat that usually arises at birth or during the first year of life, whereas some specific subtypes of CGL may become prevalent during childhood or later.[11] Adipose tissue is absent from not only subcutaneous, but also from intraabdominal sites, and the resultant prominent musculature causes a striking recognizable phenotype. MRI of the abdomen shows complete absence of intraabdominal, retroperitoneal, and subcutaneous fat, but a prominent fatty liver, as well as the presence of fat in certain anatomic sites such as the orbits, palms, and soles. Thus, this genetic defect results in poor development of metabolically active but not mechanically important adipose tissue.[10,23,30,31] Other somatic abnormalities that contribute to the abnormal appearance are acanthosis nigricans, a protuberant abdomen associated with hepatomegaly, and/or splenomegaly, hernias, prominent musculature, and acromegaloid features.[24,27]

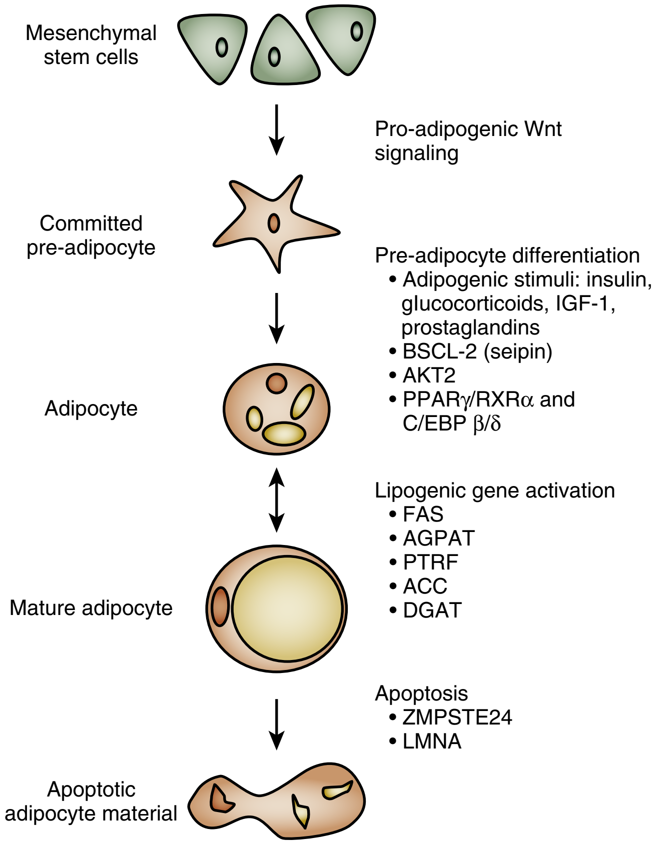

Mesenchymal stem cells
Pro-adipogenic Wnt signaling
Committed pre-adipocyte
Pre-adipocyte differentiation
• Adipogenic stimuli: insulin, glucocorticoids, IGF-1, prostaglandins
• BSCL-2 (seipin)
• AKT2
• PPARγ/RXRα and C/EBP β/δ
Adipocyte
Lipogenic gene activation
• FAS
• AGPAT
• PTRF
• ACC
• DGAT
Mature adipocyte
Apoptosis
• ZMPSTE24
• LMNA
Apoptotic adipocyte material

Fig. 24.2 Differentiation of adipocytes, demonstrating gene activation and signaling pathways, in relation to lipodystrophies: lipodystrophy genes and their associated protein products are shown in bold. Mesenchymal stem cells have the pluripotent ability to differentiate into osteocytes, myocytes, chondrocytes, stromal cells, or adipocytes. Wnt signal transduction pathways and other transcription factors lead to the development of a committed preadipocyte. Subsequently, adipogenic stimuli, such as insulin, glucocorticoids, IGF-1, and prostaglandins, initiate cell changes that lead to differentiation into an adipocyte. Current data suggest that the *BSCL-2* gene, coding for seipin, and the *AKT-2* gene play a role in adipocyte differentiation, and their mutations are involved in congenital generalized lipodystrophy (CGL) type 2 and familial partial lipodsytrophy (FPL) type 4, respectively. CCATT–enhancer-binding protein (C-EBP) β/δ transcription factors stimulate factors PPARγ, C-EBPα, and sterol regulatory element–binding protein (*SREBP*) 1c, which are upregulated in this process; PPARγ mutations are associated with FPL3, and intracellular lipid accumulation in an adipocyte is dependent on substrate availability. In response to a variety of lipogenic signals, the adipocyte matures as the size of the lipid droplets increases. Fatty acid synthase (*FAS*), acetyl coenzyme A carboxylase, and diacylglycerol acyltransferase (*DGAT*) expression are upregulated, which is necessary for the biosynthesis of phospholipids and triglycerides. *AGPAT2* is responsible for acylating phosphatidic acid to synthesize triglycerides and phospholipids, and its mutation is involved in CGL1. *PTRF* regulates caveolae intracellularly, and its functional loss results in CGL4. Apoptosis of adipocytes is influenced by *LMNA,* encoding for nuclear lamina proteins, and the *ZMPSTE24* gene product is required for posttranslational lamin A and C processing. Mutations in these genes lead to mandibuloacral dysplasia via premature nuclear disruption.

Congenital generalized muscular weakness and cervical spine instability have also been reported in occasional cases.[31] Many females have polycystic ovaries and may also present with enlarged clitoris, hirsutism, and absence of regular menstrual cycles or oligomenorrhea. Only a few affected women have had successful pregnancies, whereas affected men have normal fertility. Other rare manifestations of CGL include postpubertal focal lytic lesions in the long bones,[32-34] mild mental retardation,[28] and hypertrophic cardiomyopathy.[35-37] The basal metabolic rate may also be increased. Although patients may have accelerated linear growth and advanced bone age during their childhood, they normally have normal or reduced heights as adults.

Although both sexes are affected at similar rates, the multiple metabolic features associated with this syndrome tend to be more severe and develop earlier in girls. Children with CGL tend to have voracious appetites and accelerated growth; precocious menarche and/or pubarche, signaling early puberty, are rarely observed. Insulin resistance has been noted at an early age and may be present even at birth. Although diabetes is infrequently seen in infancy, its development is common during adolescence or young adulthood. It is rarely ketotic and is usually refractory to insulin therapy. Diabetic complications of nephropathy, retinopathy, acute pancreatitis, and hepatic steatosis are typically present and are a major cause of morbidity in affected individuals. Dyslipidemia poses a therapeutic dilemma as well. Hypertriglyceridemia is characterized by increased concentrations of very low-density lipoproteins and chylomicrons, whereas serum high-density lipoprotein (HDL) is usually low. Individuals with CGL may develop severe hypertriglyceridemia even from the first months of disease that may lead to acute pancreatitis.[38] Severely elevated triglycerides may not only provoke acute pancreatitis, but are also frequently related to the development of fatty liver disease. This commonly progresses to cirrhosis, which in many cases may be fatal. Serum adipocytokines, the hormones produced by adipose tissue (such as leptin and adiponectin), circulate in extremely low levels in CGL.[10] Patients with CGL usually demonstrate low leptin levels (~1.0 ng/mL), regardless of age or gender.[39] At least four molecularly distinct forms of congenital lipodystrophy have been defined, with mutations in *AGPAT2* and *BSCL2* responsible for 95% of reported cases of CGL.

Type 1 CGL (CGL1) is caused by *AGPAT2* gene mutation, as first described in 1999. This gene has been mapped to chromosome 9q34.[40,41] It encodes the enzyme 1-acylglycerol-3-phosphate O-acyltransferase 2 (AGPAT2), a lysophosphatidic acid acyltransferase isoform consisting of 278 amino acids localized within the endoplasmic reticulum (ER) membrane. It is highly expressed in white adipocytes[42] and catalyzes the acylation of lysophosphatidic acid to form phosphatidic acid, a key intermediate in the biosynthesis of triacylglyceride and glycerophospholipids.[43] AGPAT2 deficiency leads to a reduction of triglycerides inside the adipose organ and impaired signaling of critical elements engaged in adipogenesis such as PI3K/AKT and peroxisome proliferator-activated receptor gamma (PPARγ).[44] Only metabolically important adipose tissue (e.g., intraabdominal, intermuscular, subcutaneous, bone marrow) and not mechanically important adipose tissue (e.g., soles, palms, scalp, periarticular) are markedly reduced in patients with type 1 CGL. Forty-two mutations of the *AGPAT2* gene have been identified in 150 individuals, and the number of the related variants has increased over time.[45,46] Two more novel mutations of *AGPAT2* were recently observed in Italian individuals with CGL1.[46] Genetic testing of the first case revealed a point mutation (c.430C>T) in exon 3 of the *AGPAT2* gene. The patient, a 7-year-old boy, demonstrated clinical characteristics of CGL (namely lipoatrophy sparing the palms and soles, and muscle hypertrophy), as well as severe hypertriglyceridemia, elevated liver enzymes, and hepatic steatosis. The other case had a homozygous *AGPAT2* missense variant (c.475C>G, resulting in p. Arg159Cys) in exon 3. The 53-year-old female patient was diagnosed with lipoatrophic diabetes, bilateral proliferative retinopathy, diabetic and hypertensive nephropathy, hyperphagia, eruptive xanthomatosis, liver steatosis, and splenomegaly. The patient displayed signs of acromegaly (marked prognathism and enlarged hands and feet), muscle hypertrophy, phlebomegaly, acanthosis nigricans, and umbilical hernia.

Type 2 CGL (CGL2) is associated with *BSCL2* gene mutations including base substitutions, insertions, and deletions, events that can introduce a frameshift and/or a premature stop codon. This gene is located on chromosome 11q13 and encodes a 398–amino-acid protein named seipin, as a tribute to Martin Seip.[47] The *BSCL2* gene mutation has been found in patients of European and Middle Eastern origins, as well as those from South America and southern Asia. At least 36 genetic mutations of BSCL2 have been identified in 167 individuals.[45] Seipin is expressed diffusely in many tissues but predominantly in testis and brain[47]; its function in humans is largely unknown but includes adipocyte differentiation and lipid droplet formation.[48] According to a study performed in yeast, seipin is important for lipid droplet morphology and perhaps assembly.[26] Several studies indicate that *BSCL2* expression is critical for normal adipogenesis *in vitro*, as cells lacking *BSCL2* failed to induce expression of key lipogenic transcription factors (PPARγ and CCAAT/enhancer binding protein alpha [C/EBP-α]), as well as enzymes (AGPAT2, DGAT2, and lipin 1).[49-51] *BSCL2* mutations are usually related to a more predominant adipose tissue loss than in CGL1.[26] CGL attributed to *BSCL2* mutation appears to have a more severe disease phenotype than that owing to *AGPAT2* mutation, with a higher incidence of premature death and a lower prevalence of partial and/or delayed onset of lipodystrophy.[28] Furthermore, some patients with *BSCL2* mutations present with cardiomyopathy even from a young age[35] and have a higher prevalence of intellectual impairment.[27,28]

Compared with patients with CGL1, CGL2 patients have a more pronounced absence of adipose tissue. In addition to the loss of metabolically active fat (subcutaneous regions, intermuscular regions, bone marrow, intraabdominal and intrathoracic regions), CGL2 patients also lack mechanical fat (orbital regions, palms, soles, and joints).[23]

Both type 3 and type 4 are rarely observed and demonstrate some discrete clinical characteristics that are not present in type 1 or 2. Pathogenic variants of the caveolin 1 (*CAV1*) gene cause type 3 CGL, whereas variants in the polymerase I and transcript release factor (*PTRF*) gene are responsible for the development of type 4.[52,53]

Type 3 CGL (CGL3) is caused by *CAV1* gene mutations and was first described in 2008.[52] *CAV1* is located on chromosome 7q31, and its end product, caveolin 1, is a highly conserved 22-kD protein. The latter is a crucial component of plasma membrane microdomains known as *caveolae.* These domains have important roles in regulating signaling pathways and processes such as cell migration, polarization, and proliferation. Caveolae also promote insulin signaling and inhibit PKA signaling.[54] Caveolin 1 has been identified as a major binder of fatty acids on the plasma membranes, translocating them to lipid droplets. *CAV1* knockout mice are resistant to diet-induced obesity and exhibit severe hypertriglyceridemia, high postprandial free fatty acid (FFA) levels, reduced levels of both high and total molecular weight adiponectin, and increased risk of diabetes mellitus, atherosclerosis, and cardiovascular disease.[54-56] Mutated CAV1 may induce lipodystrophy by interfering with lipid handling, lipid droplet formation, and adipocyte differentiation[26,57] and is phenotypically classified as type CGL3. A mutation has been identified in one individual with CGL[24,52,58] who had a homozygous nonsense mutation of *CAV1,* probably as a result of a consanguineous union. The patient with *CAV1* mutation (CGL3) had clinical features similar to those of patients with CGL1 and CGL2, whereas the degree of her lipodystrophy was intermediate between these two phenotypes. The patient had also developed diabetes mellitus in adolescence and presented some distinctive features, including hypomagnesemia, hypocalcemia (probably owing to vitamin D resistance), functional megaoesophagus, short stature, and decreased bone density. The patient still preserved both mechanical adipose tissue and bone marrow fat, whereas the metabolically active adipose tissue was absent. Three distinct mutations of *CAV1* gene (one in exon 2 and two in exon 3) were identified in five individuals. Two of the three mutations were nonsense mutations (c.112G>T, p.Glu38 and c.424C>T, p.Gln142*), and the third was a frameshift mutation (c.479_480delTT, p.Phe160fs*).[45]

In 2009, Hayashi et al. reported for the first time a new type of CGL linked with mutations in the *PTRF*, gene, also known as *CAVIN1*.[53] Type 4 CGL (CGL4) is caused by homozygous or compound heterozygous mutations in the *PTRF* gene, which is located at 17q21.2, and its product Cavin-1 (PTRF) is involved in biogenesis of caveolae, regulating caveolin 1 and 3 expression.[53] It should be mentioned that caveolin 3 is most prevalent in muscle tissue, including cardiac myocytes.[59] It consists of 151 amino acids and is involved in signal transduction on the myocardial cell membrane and regulation of ion channel function in caveolae.[60]

Cavin-1 is a peripheral membrane protein localized on the inside surface of caveolae that supports the stabilization and assembly of the membrane structure, probably through interaction with the cellular cytoskeleton.[53,61] Cavin-1 is present in adipocytes and regulates their differentiation and the expandability of adipose tissue.[61,62] In type 4 CGL, adipose tissue loss may occur progressively during infancy and is significant in several areas of the body, including the face.

To date, 34 patients with 14 different mutations of the *CAVIN1* gene have been reported in the literature. Frameshift mutations were found to be most common, followed by nonsense and missense mutations.[45] Clinical features include generalized lipodystrophy in association with congenital distal myopathy, muscular hypertrophy, osteopenia, distal metaphyseal deformation, atlantoaxial instability, pyloric stenosis, gastrointestinal dysmotility, hepatosplenomegaly, and acromegaloid characteristics[31,53,63] Other findings are hypertriglyceridemia, insulin resistance, and elevated serum creatine kinase concentration. These patients exhibit a high risk of QT interval prolongation, cardiomyopathy, cardiac fibrosis, life-threatening cardiac arrhythmias such as exercise-induced ventricular tachycardia, catecholaminergic polymorphic ventricular arrhythmia, and sudden death.[59,63-66] One potential link between type 4 CGL and these serious cardiac complications is the impact of CAVIN1 on the expression pattern of caveolin-3 in cardiac myocytes.[60,67,68]

KEY POINTS

- Congenital generalized lipodystrophy (CGL) is a rare heterogeneous autosomal recessive syndrome observed mainly in cases of parental consanguinity. It is characterized by near-complete absence of body fat that may spare the supportive adipose tissue, as well as extreme muscularity, present at birth or soon thereafter. Patients with CGL develop profound hypoleptinaemia, and they are predisposed to metabolic disorders. Four distinct subtypes of congenital lipodystrophy have been defined, with *AGPAT2* and *BSCL2* mutations being the most frequent. Metreleptin is approved for use in patients with CGL.

EPIDEMIOLOGY: ETIOLOGY

Congenital generalized lipodystrophy (CGL) is a rare heterogeneous autosomal recessive syndrome observed mainly in cases of consanguinity and in isolated communities. Approximately 300 patients with CGL have been described. Four molecularly distinct subtypes of congenital lipodystrophy have been defined, with *AGPAT2* and *BSCL2* mutations being the most frequent.

PATHOPHYSIOLOGY

The pathophysiology of congenital generalized lipodystrophy may include impaired signaling of PI3K/AKT and PPARγ, affecting adipogenesis and lipid storage; defects in adipogenesis and expression of several enzymes (AGPAT2, DGAT2, and lipin 1) and lipogenic transcription factors (PPARγ, C/EBP-α); impaired lipid droplet biogenesis, accumulation, and metabolism; and defects in adipocyte differentiation, adipose tissue expandability, caveolae formation, and stability.

MAIN CLINICAL CHARACTERISTICS

Congenital generalized lipodystrophy (CGL) is characterized by near-complete absence of body fat that may spare the supportive adipose tissue, as well as extreme muscularity present at birth or soon thereafter. Patients with CGL develop profound hypoleptinaemia and are predisposed to metabolic disorders.

Other Mutations Associated With a Congenital Generalized Lipodystrophy–Like Phenotype

There are several patients with the CGL phenotype who do not have mutations in any of the four genes known to be mutated in CGL, which suggests that novel CGL-related genes have yet to be identified.

PPARγ plays an important role in adipogenesis and lipid synthesis. Mutations in the *PPARγ* gene are involved in the pathogenesis of FPL; however, biallelic *PPARγ* pathogenic variants may be associated with generalized lipodystrophy (recessive). A female patient has been reported with compound heterozygous mutations (c.413_416delAATG; p.Glu138Valfs*168 and c.490C>T; p.R164W) and a similar CGL phenotype. The individual showed the hallmark features of early onset of CGL with some facial adipose tissue. Moreover, the patient developed hypertriglyceridemia, eruptive xanthomata, acanthosis nigricans, insulin-resistant diabetes, end-stage renal disease, hyperparathyroidism, irregular menstruation, prominent veins, and marked hepatosplenomegaly. She has also experienced several episodes of pancreatitis and secondary bone fractures, and her plasma leptin levels are barely detectable.[69]

The *LMNA* gene encodes the nuclear envelope proteins lamin A and lamin C. Pathogenic variants of *LMNA* lead to the production of abnormal lamins and disrupt the interaction between chromatin and the nuclear lamina, which may result in apoptosis and premature adipocyte death. However, the pathogenic mechanisms by which these variants modify adipocyte function are still not fully understood.

Mutations in *LMNA* are associated with the development of FPL and progeroid syndrome.[70,71] However, a heterozygous lamin A/C (LMNA) p.T10I mutation was reported to be associated with near total loss of adipose tissue developing in early childhood, diabetes mellitus, insulin resistance, acanthosis nigricans, hypertriglyceridemia, hepatomegaly, and progeroid features.[72,73] Thus, it has been proposed that this should also be included as a potential CGL subtype.[73] Recently, a novel mutation in the *LMNA* gene and a subsequent autosomal recessive lipodystrophy syndrome was reported in two sisters aged 17 and 19 years old. Both individuals had the homozygous mutation p.Arg545his in *LMNA*, probably as a result of a consanguineous union. The patients had developed near-generalized loss of subcutaneous adipose, severe hypertriglyceridemia, hepatic steatosis, hepatomegaly, diabetes mellitus, short stature, intellectual disability, thin lips with perioral pigmentation, dry skin, bilateral clinodactyly of the toes, joint contractures, dystrophic nails, bilateral cataracts, and genitourinary disorders such as uterine fibroids, ovarian cysts, and hydrosalpinx.[74]

Another CGL-related gene variant involve mutation of the proto-oncogene encoding the cFOS protein. Sequence analyses performed in a White female patient with the typical appearance of CGL and insulin resistance identified a novel homozygous point mutation (c.–439, T>A) in the *C-FOS* promoter. No relevant mutations of the known pathogenic genes were identified. Her parents were healthy, not consanguineous, and did not carry this mutation. The patient died at the age of 8 years; the primary cause of death was hyper acute varicella infection. However, by the age of 5 years, the patient had also developed acanthosis nigricans, insulin resistance, prediabetes, severe hypercholesterolemia, mild hypertriglyceridemia, lipoatrophy and marked muscularity associated with complete loss of subcutaneous adipose tissue, hypertrichosis, and progressive hepatosplenomegaly. It has been proposed that diminished *C-FOS* expression potentially interferes with other genes essential for initiation of adipocyte differentiation or maturation of preadipocytes, and thus might play a role in CGL development.[75]

Acquired Generalized Lipodystrophy. The acquired syndrome of total lipoatrophy, also known as Lawrence syndrome, is similar to that of the congenital disorder, except that it develops in a previously healthy individual over days to weeks, often after a nonspecific febrile illness. The syndrome is very rare, has a later onset compared with CGL, and is more prevalent in females than males, with a ratio of 3:1. It commonly develops during childhood and adolescence, predominantly in White patients.[24,76]

In addition to the generalized loss of fat that has an active metabolic function, as seen in CGL, fat loss in AGL also occurs in palms, soles, and genital areas. In some cases, facial lipoatrophy is not initially present but develops gradually. However, retroorbital and bone marrow fat may be preserved.[58]

The median time to develop diabetes after loss of fat tissue is approximately 4 years.[58] Diabetic ketoacidosis has been reported, and severe hypertriglyceridemia, hepatic steatosis, acanthosis nigricans, menstrual irregularities, and polycystic ovary syndrome (PCOS) are also common findings.[58] Patients with AGL also have markedly reduced adiponectin levels and moderately reduced leptin levels.[10]

Several autoimmune diseases and inflammatory conditions have shown a temporal relationship with AGL. These include juvenile-onset dermatomyositis (JDM), rheumatoid arthritis, systemic sclerosis, systemic lupus erythematosus, Sjögren syndrome, and panniculitis.[58,77] JDM shows a particularly strong correlation with lipodystrophy; 8% to 40% of patients with JDM develop acquired lipodystrophy.[78-81] The chronicity and severity of JDM, as well as the high frequency of calcinosis, have been shown to predict the onset of lipodystrophy.[78] AGL following autoimmune diseases is also termed *AGL type 2* or the *autoimmune disease variety*. Patients with type 2 AGL tend to be older compared with patients with other AGL types.[58] Recently a new heterozygous missense mutation c.29C>T (p.T10I) in the *LMNA* gene was described in a 17-year-old African-American female patient with JDM and AGL. The patient had progressive subcutaneous adipose loss that started in the extremities and progressed to the entire body, including the face. Other characteristics include severe hypertriglyceridemia (>10,000 mg/dL), hepatosplenomegaly, nonalcoholic steatohepatitis, acute pancreatitis, diabetes, acute diabetic ketoacidosis, PCOS, ovarian cysts, bilateral renal cysts resulting in nephromegaly, micrognathia, bilateral parotid enlargement, atrial septal defect, and moderate aortic stenosis.[82]

Panniculitis is another inflammatory condition that frequently heralds the onset of AGL. It is estimated to be present in approximately 25% of affected patients.[58,83] Panniculitis manifests as subcutaneous inflammatory nodules that show a mixed infiltrate of lymphocytes and mononucleated macrophages in adipose tissue. The course of AGL is frequently protracted in patients with panniculitis and linked to less fat loss, lower prevalence of diabetes, and less severe metabolic disorders (e.g., mild hypertriglyceridemia).[76,79] The panniculitis variety is also known as *AGL type 1*. Up to 50% of AGL patients have no apparent history of autoimmune disease or panniculitis, however. These lipodystrophies are known as *AGL type 3* or the *idiopathic type*.

Of note, according to several recent studies, AGL is associated with antiprogrammed cell death 1 (anti-PD1) therapy and may result in severe metabolic complications. Evidence suggests that PLIN1 autoantibodies may be a possible cause of AGL development in these patients.[84] Anti-PD1 therapy is an advanced anticancer therapy with substantial and significant antitumor activity that targets several types of cancers. As a result, anti-PD1 has come to be used more frequently over time. Thus, physicians should be aware of this potential adverse event, as early recognition is of paramount importance.[85-87]

The pathogenesis of AGL is unknown. The autoimmune-mediated destruction of adipocytes or preadipocytes has been hypothesized to be the underlying mechanism. Autoantibodies against adipocyte membranes may also impair fat uptake and adipocyte differentiation.[77,78,88] Several antibodies have been found to be present in AGL, but no causative relationship has been established.[78] Cytokines, including tumor necrosis factor alpha (TNF-α) and interleukin 1 (IL-1), are also likely to play important roles in the immunopathogenesis of lipodystrophy. They can potentially lead to lipodystrophy by inhibiting adipogenesis[89] or increasing receptor-mediated apoptosis of adipocytes and preadipocytes.[90] Other studies in patients with AGL have documented an association between complement activation and adipocyte destruction.[91] Chronic hepatitis with autoimmune features and low serum complement 4 levels is also reported in some patients with AGL, suggesting involvement of the classical complement pathway causing loss of fat.[92] However, it is still not clear whether this association is characteristic of lipodystrophy or rather an outcome of the autoimmune process itself.

EPIDEMIOLOGY

Acquired generalized lipodystrophy is a rare syndrome (~80 patients) that develops in a previously healthy individual.

PATHOPHYSIOLOGY

The pathophysiology of acquired generalized lipodystrophy is not fully clarified, although several autoimmune diseases and inflammatory conditions have been shown to have a temporal relationship with this condition.

MAIN CLINICAL CHARACTERISTICS

Patients present with generalized loss of fat and marked metabolic disorders, including severe hypertriglyceridemia, nonalcoholic steatohepatitis, acute pancreatitis, acanthosis nigricans, diabetes mellitus, and polycystic ovary syndrome.

Partial Lipodystrophies

Partial lipodystrophy is characterized by selective regional fat loss. It is often associated with hypertrophy of adipose tissue in nonatrophic areas and is subclassified into inherited and acquired forms.

Familial Partial Lipodystrophies. Several syndromes have been described according to distinctive clinical features or underlying pathogenetic mechanisms. Most are inherited in an autosomal dominant fashion, and the patients are born with normal fat distribution but notice local fat loss, usually during puberty. FPLs exhibit phenotypic heterogeneity that has not yet been fully delineated.

Familial Partial Lipodystrophy Type 1 or Köbberling-Type Lipodystrophy. Also known as *Köbberling-type lipodystrophy,* FPL type 1 (FPL1) was first reported by Köbberling and colleagues in 1971. In comparison to other forms of FPL, the loss of adipose tissue is restricted to the extremities. The distribution of fat on the face and neck is normal or increased, frequently in association with significant central obesity. The hallmark anthropomorphic feature of this syndrome includes a palpable "ledge" formed between the normal and lipodystrophic areas and high triceps-to-forearm and abdomen-to-thigh skinfold ratios.[93] Metabolic syndrome, especially hypertriglyceridemia, is common in FPL1. This correlates with a high incidence of pancreatitis and premature coronary artery disease. Leptin concentrations are low and correspond to the body mass index (BMI) and the level of fat loss of individual patients.[93]

Mostly women have been diagnosed with FPL1 to date, as they are more severely affected than men. Diagnosis of affected men is difficult given the normal muscular physique and very mild clinical presentation that does not allow early detection. FPL1 tends to have a childhood onset, with metabolic complications occurring during adulthood. Acanthosis nigricans is present on the neck, axillae, and groin but is usually mild.

The genetic defect associated with FPL1 is currently unknown, and no *LMNA* or *PPARγ* mutations have been identified. It appears that this syndrome may be familial for some patients but may also occur spontaneously. However, according to recent studies, a significant number of cases of FPL1 follow polygenic inheritance.[93-95]

Familial Partial Lipodystrophy Type 2 or Dunnigan-Variety (Face-Sparing) Lipodystrophy. Also known as *Dunnigan variety (face-sparing) lipodystrophy,* FPL type 2 (FPL2) is an autosomal dominant condition found mostly, but not exclusively, in subjects of northern European descent. More than 300 cases have been reported, but the true prevalence of this syndrome is thought to be much higher. Currently known cases are associated with variants in exon 8 of the *LMNA* gene.[11] It is characterized by gradual loss of almost all subcutaneous fat from the extremities, commencing at puberty. This gives rise to the characteristic phenotype of "increased muscularity" in the arms and legs. Variable and progressive loss of fat from the anterior abdomen and chest occurs later. Excess fat may subsequently accumulate in the face and neck and in the intraabdominal region, resulting in a Cushingoid appearance.[24]

Affected females tend to have more recognizable phenotypes. Although questions of gender differences have been raised, anthropometric measures and MRI data demonstrated that both affected men and women have similar patterns of fat loss. In comparison to the affected men, women may have more severe hypoleptinemia and metabolic sequelae (insulin resistance), and they may also have higher prevalence of diabetes and atherosclerotic vascular disease, as well as higher serum triglycerides and lower high-density lipoproteins. The prevalence of hypertension and fasting serum insulin concentrations are similar in men and women.[58,83] The prevalence of diabetes is not related to age, menopausal status, or family history of type 2 diabetes.[88]

Patients with FPL2 are more prone to develop PCOS, infertility, and gestational diabetes. The prevalence of gestational diabetes and miscarriage is significantly higher in these patients than in women with similar BMI and PCOS.[96]

The first gene identified as containing mutations associated with FPL2, named *LMNA,* is located on chromosome 1q21-22. As previously mentioned, *LMNA* encodes lamins A and C, which are essential components of the nuclear lamina and provide structural integrity for the nuclear envelope. Most FPL2-associated mutations in *LMNA* are missense mutations within the 3′ end of the gene,[97] whereas the *LMNA* mutations leading to the R482W and R482Q amino acid substitutions seem to be the most common pathogenic variants.[98,99] The mutant gene products may disrupt interaction with chromatin or other nuclear lamina proteins, resulting in apoptosis and premature death of adipocytes.[100] The accumulation of prelamin A may also impair adipogenesis by interfering with the key adipocyte transcription factors/regulators, including sterol response element-binding protein 1 (SREBP-1) and PPARγ.[100-103] It is interesting to note that there is no difference in the levels of lamin A and lamin C expression in different adipose depots, even though the fat loss associated with FPL2 is regionally selective, suggesting that the downstream effects of *LMNA* mutations are differentially regulated in different areas of the body.[96]

According to recent findings, there is an anthropometric, clinical, and biochemical phenotypic heterogeneity among individuals with FPL2 based on *LMNA* mutation site and gender.[104] Females tend to have a more severe FPL2 phenotype, likely because of increased levels of proinflammatory cytokines or glucocorticoid receptor GRβ overexpression and *LMNA* mutations affecting both lamins A and C.[104] It has also been suggested that the percentage of fat in the lower limb is the most reliable anthropometric measure for FPL2 diagnosis in women. Females with coexisting metabolic disorders and a negligible lower limb fat content (<1st percentile) should undergo genetic testing for FPL2.[105]

Familial Partial Lipodystrophy Caused by PPARγ Mutations, or Familial Partial Lipodystrophy 3. Also referred to as FPL3, FPL caused by *PPARγ* mutations is associated with heterozygous mutations of the *PPARγ* gene on chromosome 3p25.[106] The phenotype is similar to that of FPL2, with the exception that fat accumulation in the head and neck may not occur.[106-109] Patients with FPL3 appear to have more severe metabolic abnormalities than those with FPL2.[103]

PPARγ encodes the peroxisome proliferator–activator receptor, which belongs to the superfamily of nuclear hormone receptors. It is expressed at high levels in adipocytes and plays an essential role in adipocyte differentiation and adipogenesis.[110] Heterozygous mutations may cause loss of function by directly interfering with normal gene function (dominant negative) or by reducing gene expression (haploinsufficiency).

Recently it was found that a novel missense heterozygous mutation,(c.1060A>G; p.Ile354Val) in *PPARγ2* is responsible for a decrease (i.e., ~ 30%–35%) in receptor transcriptional activity and for the development of FPL3 in an Italian family. This finding suggests that *PPARγ* mutations that affect transcriptional activity to a lesser extent than haploinsufficiency may be also responsible for FPL3 development.[111]

Familial Partial Lipodystrophy Caused by PLIN1 Mutation, or Familial Partial Lipodystrophy 4. Also known as *FPL4,* FPL caused by *PLIN1* mutation has been characterized phenotypically as loss of subcutaneous fat from the extremities. FPL4 is associated with mutation in the *PLIN1* gene coding for perilipin 1, which is a required component of lipid droplet membranes and is essential for lipolysis and lipid storage. Six patients with this mutation had histologically small adipocytes with increased macrophage infiltration and abundant fibrosis.[112] Recently, renal involvement has been described for the first time as a clinical manifestation in a patient with FPL4. In more detail, a novel heterozygous frameshift mutation (c.1201_1202insT) was described in a 15-year-old Chinese female with insulin-resistant diabetes, hypertriglyceridemia, nonalcoholic steatohepatitis, and proteinuria. [113]

Familial Partial Lipodystrophy Caused by CIDEC Mutation, or Familial Partial Lipodystrophy 5. FPL5 is a rare autosomal recessive condition that appears in early childhood and it is caused by pathogenic variants in the *CIDEC* gene. The *CIDEC* gene is located on chromosome 3 (3p25.3) and participates in adipocyte differentiation.

It also encodes the CIDEC protein, which is an essential regulator of glucose and lipid metabolism and insulin sensitivity. Pathogenic variants of *CIDEC* are hypothesized to prevent lipid droplets from storing fat and cause defects in adipocyte differentiation.[114,115] FPL5 is characterized by partial lipodystrophy and muscular hypertrophy, but no accumulation of adipose tissue, acanthosis nigricans, metabolic abnormalities, severe insulin-resistant diabetes, and hepatic steatosis. Recently, two single nucleotide polymorphisms (rs2479 or C allele at rs1053239) were associated with abnormal fasting glucose levels and hypertriglyceridemia. Patients with an A allele at rs2479 or a C allele at rs1053239 tend to rapidly develop high blood pressure and benefit from angiotensin II–targeted treatment.[116]

Familial Partial Lipodystrophy Caused by LIPE Mutation, or Familial Partial Lipodystrophy 6. FPL6 is an autosomal recessive syndrome, and only a few cases have been reported to date. It appears in adulthood and is caused by mutations in the *LIPE* gene (lipase E, hormone-sensitive type). It is associated with impaired lipolysis, which may result simultaneously in lipomatosis and partial adipose loss. It is also characterized by insulin resistance, diabetes, hypertriglyceridemia, and hepatic steatosis.[117] Individuals with FPL6 may exhibit multiple symmetric lipomatosis, accumulation of adipose tissue in the neck, axillae, supraclavicular area, and back, partial adipose tissue loss from the lower extremities, and in some cases distal symmetric muscular dystrophy.[117,118]

Familial Partial Lipodystrophy Caused by AKT2 Mutation. FPL caused by *AKT2* mutation has been reported in a single family by George and associates.[119] It is inherited in an autosomal dominant fashion and manifests as severe insulin resistance and partial lipodystrophy confined to the extremities.[119] Although the specific mutations in *AKT2* has not yet been identified, and this conditions is not included in earlier FPL classification systems,[98] recent studies consider it to be another FPL syndrome.[2,73]

AKT, also known as protein kinase B, is a serine/threonine protein kinase and plays multiple roles in cell signaling, cell growth, and glycogen synthesis, as well as in insulin-stimulated glucose transport.[120] Lipodystrophy in patients with *AKT2* mutations is thought to be attributed to reduced adipocyte differentiation and dysfunctional postreceptor insulin signaling.

Familial Partial Lipodystrophy Caused by CAV1 Mutation. Heterozygous *CAV1* mutation has also been identified as a rare cause of partial lipodystrophy.[121] Two cases with different frameshift *CAV1* mutations have been reported. Both patients were described to have partial lipodystrophy with subcutaneous fat loss in the face and upper body, micrognathia, and congenital cataracts. One case also exhibited abnormal neurologic findings. Diabetes, hypertriglyceridemia, and recurrent pancreatitis were reported in both cases.[121]

Familial Partial Lipodystrophy Caused by PCYT1A Mutation. Another gene that has been reported to be associated with FPL is the phosphate cytidylyltransferase 1 alpha (*PCYT1A*) gene. PCYT1A is the rate-limiting enzyme in the Kennedy pathway, participating in *de novo* phosphatidylcholine synthesis. Biallelic pathogenic loss-of-function mutations in *PCYT1A* have been reported to be associated with a severe form of partial lipodystrophy, significant insulin resistance and diabetes, severe nonalcoholic fatty liver disease, and low HDL cholesterol levels in two unrelated female patients.[122] Because of the phenotypes and characteristics of these patients, it has been suggested that this mutation could be considered a potential cause of CGL.[73]

Mandibuloacral Dysplasia–Associated Lipodystrophy. Mandibuloacral dysplasia (MAD) is an extremely rare autosomal recessive progeroid syndrome that has been reported in approximately 40 cases. MAD is characterized by postnatal growth retardation, craniofacial and skeletal abnormalities (mandibular and clavicular hypoplasia, delayed closure of the cranial sutures, acroosteolysis, joint contractures, birdlike face, dental abnormalities), cutaneous changes (restrictive dermatopathy, skin atrophy, alopecia, and mottled cutaneous pigmentation), and lipodystrophy. Although MAD is present at birth, dysmorphic manifestations and progeroid features become more prominent with time, and the full clinical phenotype is recognizable during the early school years. The patients have normal intelligence,[123] and their serum leptin concentration can be low or normal. Hyperinsulinemia, insulin resistance, impaired glucose tolerance, diabetes mellitus, and hyperlipidemia have been reported in some patients.

There are two distinctive phenotypes of MAD: type A involves the loss of subcutaneous fat from the extremities and trunk but normal or excessive deposition of fat in the face and neck, and type B is characterized by more generalized loss of subcutaneous fat and premature renal failure. MAD type A (MADA) is also considered to be as a result of mutations of the *LMNA* gene that result in accumulation of prelamin A and lead to alterations of nuclear architecture and chromatin defects. It remains unclear how different mutations in the same gene lead to a variety of phenotypes. Patients with MAD type B (MADB) have been reported to carry compound heterozygous mutations in *ZMPSTE24*, the gene encoding an endoprotease, zinc metalloprotease, which is located on chromosome 1q34. The enzyme is important in posttranslational processing of prelamin A to mature lamin A. Recently, eight related Surinamese individuals of African descent with MADB were described. All of these individuals, however, shared the same homozygous founder mutation in the *ZMPSTE24* gene, resulting in a missense variant c.1196A>G, p.(Tyr399Cys).[124] As in MADA, the accumulation of farnesylated prelamin A is proposed to be responsible for the phenotype.[123] Only eight other patients with *ZMPSTE24* mutations are known, the majority from Italy.[125] Focal segmental glomerulosclerosis and calcified skin nodules have been reported in patients with ZMPSTE24 deficiency.[126]

EPIDEMIOLOGY: ETIOLOGY

The majority of familial partial lipodystrophy (FPL) syndromes are inherited in an autosomal dominant pattern and are characterized by selective regional fat loss. More than 500 patients with FPL have been reported. Six main FPL types have been described that are caused by mutations in *LMNA*, *PPARγ*, *PLIN1*, *CIDEC,* and *LIPE*.

PATHOPHYSIOLOGY

The different types of familial partial lipodystrophy syndrome exhibit phenotypic heterogeneity and distinct genetic causes. The underlying pathophysiological mechanisms are not fully delineated and may involve disrupted adipogenesis, impaired differentiation of adipocytes, defects in lipid storage and lipolysis, increased adipocyte apoptosis, and premature adipocyte death.

MAIN CLINICAL CHARACTERISTICS

Patients are born with normal fat distribution, with the local fat loss occurring later in life, usually during puberty. Muscular hypertrophy and cardiometabolic disorders are common.

Autoinflammatory Syndromes. Several autoinflammatory syndromes may also be associated with lipodystrophic states. A rare autosomal recessive, autoinflammatory syndrome presenting with joint contractures, microcytic anemia, muscle atrophy, and panniculitis-induced lipodystrophy (JMP) in childhood has recently been reported in three individual patients from Japan and two families from Mexico and Portugal.[127] Additional clinical features of JMP include hepatosplenomegaly, intermittent fever, calcification of the basal ganglia, and hypergammaglobulinemia. Sequencing of candidate genes involved in immune system dysfunction led to the discovery of a loss-of-function mutation of the proteasome subunit beta-type 8 (*PSMB*-8) gene on chromosome 6. *PSMB8* encodes b5i, a catalytic subunit of immunoproteasomes, which mediates proteolysis and generates MHC class 1 molecules. Mutations may result in adipose tissue lymphocytic infiltration and loss of surrounding fat tissue.

CANDLE (chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature) is another newly described syndrome causing partial lipodystrophy that has so far been reported in six patients and is likely inherited in an autosomal recessive fashion.[128,129] Infants present with annular violaceous plaques and recurrent fevers, with eventual loss of adipose tissue from the upper limbs and face.[130] Other associated clinical characteristics include hepatosplenomegaly, anemia, eyelid swelling, and calcifications of the basal ganglia. The molecular mechanism of this syndrome is yet to be determined. It has been proposed that physicians should include this syndrome in their differential diagnosis when treating individuals with early-onset systemic inflammatory disease and skin manifestations.[131]

Other Syndromes and Progeroid Disorders With a Component of Lipodystrophy. Multiple other syndromes are also linked to lipodystrophy, and several have been identified as laminopathies, including Hutchinson–Gilford progeria syndrome (a very rare and uniformly fatal segmental progeroid syndrome with progressive and generalized fat loss), restrictive dermopathy, progeria-associated arthropathy, and atypical progeroid syndrome.[15] The majority of these syndromes are associated with *de novo LMNA* mutations, yielding a MADA-type lipodystrophy.

Werner syndrome (short stature, birdlike appearance of the face, early onset of aging processes, and progeroid features) has been linked to homozygous mutations in *RECQL2,* which encodes a DNA helicase. In contrast, the genetic basis and inheritance patterns have yet to be clarified for the following syndromes: Cockayne syndrome (short stature, photosensitivity, hearing loss, premature aging, and progeroid appearance), carbohydrate-deficient glycoprotein syndrome (nonprogressive ataxia associated with cerebellar hypoplasia, stable mental retardation, variable peripheral neuropathy, and strabismus), SHORT syndrome (S–short stature; H–hyperextensibility of joints and/or hernia (inguinal); O–ocular depression; R–Rieger anomaly; T–teething delay), and ectodermal dysplasia in association with generalized lipodystrophy acral renal ectodermal dysplasia lipoatrophic diabetes (AREDYLD) syndrome.

Acquired Partial Lipodystrophy (Barraquer–Simons Syndrome). First reported in 1885 by Mitchell, acquired partial lipodystrophy (APL) was further characterized by Barraquer Roviralta in 1907. There have been approximately 250 cases reported in the English-language literature.[24]

Patients with APL are primarily of European descent; however, cases have also been reported in Asian Indian, Vietnamese, and Samoan populations. The disease shows a female dominance, with a 4:1 ratio, and most patients have clinical manifestations in early puberty or early adulthood, usually before 15 years of age. Female patients normally have regular menses and intact fertility. The characteristic fat loss progresses in a “cephalocaudal” fashion, with fat loss appearing first in the face and spreading to the upper part of the body. Fat under the umbilicus is rarely affected. Many patients develop excess fat accumulation over the lower abdomen, gluteal region, thighs, and calves. Breasts may lose fat and consist of firm glandular tissue only.[58]

In contrast to other types of lipodystrophies, acanthosis nigricans, hirsutism, and hypertrichosis are rare. Insulin resistance is uncommon, and the prevalence of diabetes is much reduced compared to other types of lipodystrophies. The prevalence of the metabolic syndrome is also significantly lower in patients with APL. Hepatomegaly is common among patients with APL. In the series of case reports by Misra and Garg, 35% of APL patients had hypertriglyceridemia, and a third had low concentrations of HDL.[58] Serum leptin levels were normal in the majority of patients.[10]

A strong association between APL and membranoproliferative glomerulonephritis (MPGN) type 2 has been proposed. The spectrum of presentation ranges from acute glomerulonephritis, hematuria, nocturia, urinary casts, albuminuria, and nephritic syndrome to chronic glomerulonephritis and uremia.[121,132] The serum C3 complement levels are usually low, with the presence of C3 nephritic factor,[121] which blocks degradation of the enzyme C3 convertase. Patients with low C3 levels tend to have an earlier onset of lipodystrophy than those with normal serum C3 levels. The median time interval between the onset of lipodystrophy and the development of MPGN is approximately 5 to 10 years but could be as long as 20 years.[121,133]

Similar to AGL, APL is also frequently seen in the context of autoimmunity or infections.[77] The most frequently cited infection preceding APL is measles. The low C3 levels may also render APL patients susceptible to recurrent pyogenic infections, particularly *Neisseria.*[134] Systemic lupus erythematosus and dermatomyositis/polymyositis are the autoimmune diseases that are most frequently associated with APL.[58] The precise mechanisms leading to adipose-tissue atrophy in APL remain unclear. C3 nephritic factor has been shown to induce lysis of adipocytes expressing factor D (adipsin).[58]

According to recent findings, C3 hypocomplementemia (and to a lower extent low serum C4 complement levels) and autoimmunity are associated with Barraquer–Simons syndrome (BSS). Furthermore, C3NeF, as well as other autoantibodies (such as anti-C3, anti-P, and anti-FB) against components of the C3 convertase of the alternative pathway, may also be present in several patients with BSS.[135]

Rare variants in the *LMNB2* gene encoding lamin B were reported in five APL patients, but half of those variants were also reported in normal controls.[136] In a study by Guallar and coworkers, *PPARγ* gene downregulation and mitochondrial toxicity were observed in a patient with APL, suggesting that impaired adipogenesis and adipocyte metabolism may also underlie the pathogenesis of APL.[137]

KEY POINTS

- Acquired partial lipodystrophy (APL) is more frequent in females, and most patients have clinical manifestations in early puberty or early adulthood. The etiology of APL is still uncertain, although an association with autoimmune-mediated loss of adipose tissue has been suggested. Patients with APL may demonstrate circulating autoantibody (C3 nephritic factor) and low complement component 3 levels. The progressive loss of adipose tissue follows a “cephalocaudal” trend, starting from the face and spreading to the upper part of the body, while fat may be preserved or accumulated over the lower abdomen, gluteal region, thighs, and calves. Membranoproliferative glomerulonephritis or autoimmune disease may exist. Associated metabolic syndrome, insulin resistance, and related complications are not common.

EPIDEMIOLOGY

Acquired partial lipodystrophy is more common in females than in males; approximately 250 patients have been reported.

ETIOLOGY: PATHOPHYSIOLOGY

Although the etiology of acquired partial lipodystrophy (APL) is still uncertain, a mechanism involving autoimmune-mediated loss of adipose tissue has been suggested. Patients with APL may demonstrate circulating autoantibody (C3 nephritic factor) and low complement component 3 levels.

MAIN CLINICAL CHARACTERISTICS

Most patients have clinical manifestations in early puberty or early adulthood.

The progressive loss of adipose tissue follows a "cephalocaudal" trend, starting from the face and spreading to the upper part of the body, while fat may be preserved or accumulated over the lower abdomen, gluteal region, thighs, and calves. Membranoproliferative glomerulonephritis or autoimmune disease may exist. There is a low prevalence of metabolic syndrome, insulin resistance, and related complications.

Human Immunodeficiency Virus–Associated Lipodystrophy Syndrome. Human immunodeficiency virus (HIV)-associated lipodystrophy syndrome (HALS) is currently the most common form of partial lipodystrophy. First reported in 1998, HALS mainly develops in patients infected with HIV who are receiving highly active antiretroviral therapy (HAART). Importantly, it is independent of fat loss caused by either the HIV infection itself or by cancers or opportunistic infections. The pattern of fat loss is also different. HALS patients may experience lipoatrophy, lipohypertrophy, or a combination of both, namely mixed lipodystrophy.[138,139] Cases of lipodystrophy have also been reported in HAART-naïve patients. The prevalence of HALS increases with increased duration of exposure to HAART and is reported to be present in up to 50% of patients on antiviral treatment for more than 1 year, thus affecting more than 100,000 patients in the United States.

HALS usually manifests as peripheral fat wasting that involves the face, arms, legs, and buttocks. The generalized depletion of subcutaneous fat in HALS is distinct from HIV-related wasting, which is associated with advanced acquired immunodeficiency syndrome and loss of other tissues, such as muscle mass. Accumulation of fat is frequently seen in the dorsocervical area ("buffalo-hump" pads), abdomen, occasionally in the breasts of both men and women, in the suprapubic area, under the axillae, and over the anterior aspect of the neck. Lipomatosis manifests in a small percentage of patients. Facial fat loss can be quite severe, sometimes resulting in an emaciated appearance. Most studies have shown an increase in central fat over the first 6 months after the initiation of antiretroviral therapy that subsequently levels off.

Most HIV-infected patients with lipodystrophy are otherwise relatively healthy, but dyslipidemia, especially hypertriglyceridemia, is common among HIV-infected patients receiving HAART. HIV viremia has been linked to decreased plasma concentrations of total, low-density lipoprotein (LDL), and HDL cholesterol, and at later stages elevated triglyceride levels. HAART has been shown to cause a worsening lipid profile, with increased plasma triglycerides, increased total and LDL cholesterol, and decreased HDL cholesterol, which can be further accompanied by increases in small, dense LDL particles, lipoprotein (a), and apolipoproteins B, C-III, E, and H.[140] HAART-associated dyslipidemia is associated with accelerated atherosclerosis and signs of endothelial dysfunction.[141-144] Seen in 35% of patients, insulin resistance and frank diabetes mellitus are more prevalent in HIV subjects with lipodystrophy,[145] but acanthosis nigricans seems to be extremely rare. These metabolic disorders are predictive of the development of metabolic syndrome, whereas their prevalence may vary among HIV-infected individuals, according to findings from different studies.[146,147] The underlying mechanisms related to the development of the metabolic syndrome and the different subtypes of lipodystrophy in HIV-infected population have not been fully elucidated, and seem to involve a multifactorial process.[148] The pathophysiology of metabolic syndrome in HIV patients may include HIV *per se*, duration of antiretroviral drug use, and chronic inflammation.[149] Moreover, findings suggest that variants in candidate genes associated with lipid metabolism and adipogenesis may be predictors of these metabolic and anthropometric disorders.[150,151]

Hepatic steatosis may also develop. Both leptin and adiponectin levels are decreased in patients with HALS. The reduction of leptin levels correlates with decreased subcutaneous fat mass,[152] whereas decreased adiponectin levels are more closely associated with intraabdominal fat accumulation.[153]

Fat atrophy and fat deposition appear to be associated with different risk factors in HALS. Low baseline fat mass and increased disease severity are associated with a higher incidence of fat atrophy.[154] Epidemiology studies have also shown that coinfection with hepatitis C virus can increase the chance of fat atrophy in HIV-infected individuals.[155] On the other hand, older age, female sex, high baseline body fat, increased HIV viral load, low CD4 count, and longer duration of HAART are associated with a higher risk for fat accumulation in HIV patients.[156,157]

The frequency and manifestations of lipodystrophy also differ with respect to the drugs used. Nucleoside reverse transcriptase inhibitors (NRTIs), particularly zidovudine and stavudine, are commonly associated with morphologic changes, particularly fat loss from the extremities, whereas protease inhibitors (PIs) are more frequently linked to hypertriglyceridemia, insulin resistance, and localized fat accumulation.[158] Interestingly, fat loss worsens with ongoing HAART therapy but does not reverse on its discontinuation. Because PIs and NRTIs are usually given in combination as part of HAART, the individual effects of each drug on phenotype are not clear.

The mechanism behind HALS is complex and is currently not completely understood.[159] HAART has been widely accepted as playing a central role in the development of lipodystrophy, but accumulating evidence indicates that the HIV virus *per se*, as well as host immune responses, also contribute to the development of HALS. NRTIs have been shown to suppress adipogenesis either through mitochondrial toxicity (by inhibiting mitochondrial DNA polymerase gamma) or by induction of genes that inhibit adipogenesis. *In vitro* studies suggest that zalcitabine, didanosine, and stavudine have the strongest effects in a reducing order of magnitude, whereas tenofovir and lamivudine show minimal or no mitochondrial toxicity.[138] Combinations of drugs can act synergistically and lead to mitochondrial depletion. In addition, zidovudine, emtricitabine, and abacavir can also impair cell proliferation and increase lactate and lipid production. NRTIs may also contribute to insulin resistance by altering the levels of IL-6, TNF-α, and adiponectin.[158]

PIs can lead to adipose tissue changes through several potential mechanisms: (1) impairment of adipocyte differentiation by downregulation of the expression of master adipogenic transcription factors, such as C/EBP-α and C/EBP-β, PPARγ, and SREBP-1;[13] (2) increase of adipocyte apoptosis, leading to a reduction in cell numbers; and (3) decrease of lipid accumulation in adipocytes through reactive oxygen species (ROS) production and increased macrophage recruitment.[160] Therapy with PIs has also been implicated in the causation of metabolic abnormalities by inhibiting glucose transport 4 (GLUT4)-mediated glucose transport; by suppressing insulin signaling; and through activation of lipolysis, induction of IL-6 and TNF-α, reduction in gene expression, and secretion of adiponectin, as well as proteasome dysfunction. According to a study, ritonavir significantly affected the expression of 389 genes involved in adipocyte differentiation, glucose metabolism, adiponectin secretion, and triglyceride accumulation. In detail, genes responsible for adipocyte triglyceride accumulation (namely *CFD*, *CIDEC*, and *PPARγ*) and glucose transport genes (including *GLUT4* and *ADIPOQ*) were downregulated. Moreover, the expression of PPARγ regulatory genes *CEBPA* and liver-X-receptor α was also reduced, whereas IL-6 was increased.[161] Lopinavir, ritonavir, saquinavir, and nelfinavir are the worst offenders. The newer PI, atazanavir, has a much milder effect. Indinavir does not have much effect on cell viability or lipogenesis but inhibits glucose uptake to a greater extent than the other PIs.[158]

Nonnucleoside reverse transcriptase inhibitors (NNRTIs), including efavirenz and nevirapine, appear to have more favorable safety profiles in terms of lipodystrophy complications. Although an *in vitro* study showed that efavirenz may interfere with adipogenesis by reducing the expression of SREBP-1, a key adipogenic transcription factor,[162] and a prospective randomized trial suggested that efavirenz could have greater potential for causing lipoatrophy than the combination of lopinavir plus ritonavir,[163] the results from several clinical studies imply that the potential role of efavirenz in the development of lipodystrophy is minimal, and that it may depend on the NRTIs that form the backbone of the regimen.[164] Lipoatrophy is lowest with the NRTI-sparing regimen of lopinavir and efavirenz, but this combination led to worsening dyslipidemia.[165] Additionally, *in vitro* data have shown that efavirenz may increase production of ROS, reduce lipid content in mature adipocytes, and inhibit mitochondrial activity.[162,166]

There is also increasing evidence that HIV-1 infection itself, regardless of HAART, may induce inflammatory and proapoptotic pathways in adipose tissue and thus contribute to lipoatrophy. *In vitro* experiments have demonstrated that the HIV-1 viral protein R may act as a corepressor of PPARγ-mediated gene transcription, inhibiting adipocyte differentiation.[167] This could be either occurring through the direct HIV-1 infection of cells in adipose tissue or may be mediated by HIV-1–encoded proteins.[168] Genetic background may also influence the degree of HALS. A single nucleotide polymorphism in the resistin gene has been associated with increased risk for developing limb fat loss, dyslipidemia, and insulin resistance in patients on HAART therapy.[169] Additionally, two single nucleotide polymorphisms in the adiponectin receptor gene *ADIPOR2* (rs11061925 and rs929434) have been found to be associated with biochemical parameters, such as the levels of triglycerides, total cholesterol, glucose and adiponectin, in HIV-infected men receiving HAART.[170]

Inflammatory cytokines, including interferon alpha (IFN-α), TNF-α, IFN-γ, monocyte chemoattractant protein-1 (MCP-1), IL-1, IL-6, and IL-12, may also contribute to or mediate the clinical manifestations of this syndrome,[152,168] and this remains an active area of research.

EPIDEMIOLOGY: ETIOLOGY

Human immunodeficiency virus (HIV)-associated lipodystrophy syndrome (HALS) is currently the most common form of partial lipodystrophy and develops in patients infected with HIV who are receiving highly active antiretroviral therapy. Approximately 40% of HIV patients receiving protease inhibitor therapy for more than 1 year develop HALS.

PATHOPHYSIOLOGY

The pathophysiological mechanism behind human immunodeficiency virus–associated lipodystrophy syndrome is complex and not fully elucidated. It may involve impaired adipogenesis and expression of adipogenic transcription factors (such as PPARγ, SREBP-1, C/EBP-α, and C/EBP-β), as well as defects in adipocyte differentiation and apoptosis.

MAIN CLINICAL CHARACTERISTICS

Patients may experience lipoatrophy, lipohypertrophy, or a combination of both, namely mixed lipodystrophy. Human immunodeficiency virus–associated lipodystrophy usually manifests as peripheral fat wasting involving the face, arms, legs, and buttocks, and excess fat deposition in the neck ("buffalo-hump" pads) and abdomen. Of note, nucleoside reverse transcriptase inhibitors are commonly associated with morphologic changes, particularly fat loss from the extremities, whereas protease inhibitors are more frequently linked to hypertriglyceridemia, insulin resistance, and localized fat accumulation. Patients may experience metabolic disorders, such as severe insulin resistance, diabetes mellitus, dyslipidemia, and liver steatosis.

Localized Lipodystrophies

Localized lipodystrophies are characterized by loss of subcutaneous adipose tissue from small areas or from small parts of an extremity, but insulin resistance or metabolic abnormalities do not usually develop in these patients. Drug-induced lipoatrophy at the site of insulin injection was a frequent complication before the availability of purified human insulin but is rather uncommon today. Localized hypertrophy is still frequently seen in patients who use the same injection site too often. Other rare causes of localized lipodystrophy as a result of repeated pressure in specific areas, to panniculitis, or as part of a rare syndrome called *lipodystrophia centrifugalis abdominalis infantilis* have been reported.[171]

MECHANISMS RESPONSIBLE FOR SEVERE INSULIN RESISTANCE

As mentioned earlier, the mechanisms of insulin resistance in several syndromes associated with lipodystrophy are not fully understood, and it is likely that the etiology of lipodystrophies is multifactorial. Lipodystrophy may develop because of abnormal preadipocyte development, adipocyte dysfunction, or increased fat cell death. The study of lipodystrophy syndromes provides unique insights regarding the lack of adipocyte enlargement and adipose expansion and the inability of subcutaneous fat to store excess energy. The excess energy is deposited as ectopic fat with a subsequent decrease in adipokine levels, mechanisms that are associated with the development and severity of insulin resistance.[4,16,172]

Inflammation is thought to contribute to insulin resistance via impairment of lipolysis and adipocyte metabolism.[173] TNF-α reduces insulin receptor kinase activity, downregulating insulin receptor substrate (IRS)-1 and GLUT-4 phosphorylation. Furthermore, IL-6 at elevated levels may induce triglyceride secretion from the liver and promote hepatic gluconeogenesis. The impaired secretion of adipokines, such as adiponectin and leptin, leads to abnormalities of insulin sensitivity. Finally, mitochondrial stress plays a key role in metabolic dysregulation. These mechanisms are further discussed in the following sections.

Fat Redistribution and Fat Metabolism

Changes in fat distribution (reduced subcutaneous fat with or without increased visceral fat) may cause increased insulin resistance. The lack of adipose tissue can result in inadequate storage of and therefore increased levels of FFAs. Intracellular fatty acid accumulation can directly inhibit insulin-mediated glucose transport in skeletal muscle,[174,175] and excess FFAs can also lead to lipotoxicity by inducing ectopic fat accumulation in the liver and muscle, where the adipose tissue is considered to have more "pathogenic" potential. The deposition of fat in the pancreas can also impair the β cell response and further contribute to insulin resistance.[176] Moreover, because the fat tissue at different depots shows different degrees of metabolic activity, such as lipolysis and inflammation, lipodystrophic states that are associated with higher volume of visceral fat and abdominal fat are likely to display a higher degree of insulin resistance.[176]

Adipocytokines

Alterations of adipocytokine levels can affect metabolic homeostasis and insulin resistance. Leptin and adiponectin are two of the most abundant adipocytokines produced by adipocytes, and their levels decrease in the lipodystrophic states.[10] Serum adiponectin levels correlate positively with insulin sensitivity[177,178] and are upregulated by PPARγ agonists.[179] Adiponectin acts by reducing hepatic gluconeogenesis (mainly via adiponectin receptor 2 and activation of AMPK phosphorylation) and increasing fatty acid oxidation in muscle (mainly via adiponectin receptor 1).[180] An animal study also demonstrated that adiponectin may also act in the hypothalamus (via adiponectin receptor 1) to activate insulin and leptin signaling pathways, thus promoting reduction of food intake.[181] Antiinflammatory effects of adiponectin have also been shown in various animal models of liver inflammation[81] and suggested by several observational studies in humans.[182]

In earlier animal models, it was shown that increased expression of adiponectin correlates with improved insulin sensitivity, with adiponectin decreasing insulin resistance by decreasing the triglyceride content of liver and muscle.[183,184] Although adiponectin or leptin alone partially improves insulin resistance in mouse models of lipodystrophy,[185,186] the combined administration of physiologic doses of both fully normalizes insulin sensitivity.[183] In patients with lipodystrophy, adiponectin levels are low in certain subsets of the disease, particularly in many patients with CGL1 and HALS,[187-189] and implicate hypoadiponectemia in the development of associated insulin resistance, hypertriglyceridemia, and fat redistribution.[190,191]

Serum leptin levels reflect the overall amount of adipose tissue in the body and are positively correlated with adiposity.[176] In addition to regulating food intake and increasing energy expenditure, leptin also plays an important role in the regulation of glucose homeostasis, possibly independently of its weight-reducing effects.[192] Aside from its actions in the central nervous system (CNS), leptin may exert its insulin-sensitizing effects peripherally by decreasing gluconeogenesis in the liver and adipose tissue, by exerting its lipolytic activity, and/or by increasing glucose utilization in skeletal muscle.[192] Leptin may also prevent the "lipotoxic" effects of intramyocellular lipid accumulation by activating fatty acid oxidation in skeletal muscle.[192,193] It has been shown that leptin levels are decreased in a significant proportion of patients with HALS[175] and that leptin administration improves metabolic manifestations of HALS in humans in the short and long term.[194-196] In generalized lipodystrophy, leptin concentrations are very low to undetectable, while in partial lipodystrophy they may vary from normal to low.[10] Leptin levels depend on several factors such as age, sex hormones, and fat mass. Moreover, serum leptin concentration may assist physicians in identifying and treating lipodystrophy disorders.[5,197] Leptin levels *per se*, however, should not be considered as a reliable and sufficient index to confirm or exclude lipodystrophy diagnosis.[2]

Inflammation

Inflammation in adipose tissue is likely to contribute to increased insulin resistance in the lipodystrophic state. Altered innate immunity and chronic inflammation appear to be strongly associated with insulin resistance in obesity and type 2 diabetes.[198] Inflammatory adipocytokines, such as TNF-α, IL-6, IL-8, macrophage inflammatory protein (MIP)-1α and 1β, monocyte chemotactic protein-1 (MCP-1; also known as CCL-2), plasminogen activator inhibitor-1 (PAI-1), angiotensinogen, retinol-binding protein-4 (RBP-4), and others have been implicated as the key regulators of insulin sensitivity.[179,199] The expression of several of these proteins, including TNF-α, IL-6, and IL-8, as well as macrophage markers (CD68, ITGAM, EMR1, ADAM8) and chemokines (MCP-1 and CCL-3), is increased in the subcutaneous tissue of patients with HALS.[200] In a small study of HALS, plasma PAI-1 was also found to be elevated, although its expression level in adipose tissue was not.[201]

Accumulating evidence supports an association between inflammation and insulin resistance. TNF-α mediates insulin resistance via reduction of insulin receptor kinase activity, induction of lipolysis, and downregulation of GLUT-4.[202,203] It may also induce apoptosis of adipocytes.[203] The impact of IL-6 on insulin resistance is less clear, but the majority of the evidence indicates that chronic elevation of IL-6 promotes hepatic insulin resistance and impedes differentiation of adipose tissue,[204] whereas acute elevation of IL-6 levels after exercise may promote improved glucose and lipid metabolism.[204] MCP-1 has been shown to induce insulin resistance by downregulation of GLUT-4, beta-adrenergic receptors, and PPARγ in mice. It is also associated with increased levels of atherosclerosis. Two pathways, the NF-κB pathway and the c-Jun NH_2-terminal pathway, are essential for mediating insulin resistance. Pharmacologic inhibition of the pathways has resulted in improved insulin resistance.[179,199]

Mitochondrial Stress, Oxidative Stress, and the Endoplasmic Reticulum

Mitochondrial defects have been considered as a central factor in NRTI-induced lipodystrophy. Mitochondrial dysfunction will lead to oxidative phosphorylation defects and ROS accumulation. Observational studies have shown that clinical conditions associated with increased ROS levels are also associated with increased insulin resistance.[174] Oxidative stress may also trigger β cell apoptosis and may contribute further to insulin resistance.[205,206] Angiotensin receptor blockers can attenuate oxidative stress and prevent further progression of insulin resistance.[207]

The ER plays an important role in regulating lipid, glucose, cholesterol, and protein metabolism. Stress on the ER luminal environment may generate an increased load of unfolded or misfolded proteins and may lead to adipocyte apoptosis, inflammation, and insulin resistance.[174] Seipinopathy has recently been identified as an ER stress–associated disease[208] and may contribute to insulin resistance in patients with CGL2.

Other Mechanisms

In addition to the mechanisms mentioned earlier, HAART may also contribute to insulin resistance by directly blocking glucose uptake (PIs) or by reducing phosphorylation of a key step in postreceptor insulin signaling (indinavir).[209,210] As mentioned, CGL2/BSCL2 is a severe form of lipodystrophy that is caused by loss-of-function mutations of the *BSCL2*/seipin gene and is accompanied by several metabolic disorders, including extreme insulin resistance and hepatic steatosis. Recent findings support that seipin downregulates glycerol-3-phosphate acyltransferases (GPATs). The latter are probably also responsible for the defective adipogenesis. Furthermore, GPAT3 deficiency is associated with significant improvement in insulin sensitivity and hepatic steatosis.[211,212]

Finally, patients with severe lipodystrophic states who have voracious appetites, supposedly because of leptin insufficiency, are theoretically more prone to these insults because overnutrition may aggravate central obesity and has been associated with higher levels of inflammation and ROS load.

TREATMENT OF SYNDROMES OF LIPODYSTROPHIES

Lifestyle modification, hypoglycemics, and lipid-lowering agents are generally required to treat metabolic disturbances associated with lipodystrophies. However, they commonly yield limited results. The development of new drugs targeted to specifically reverse insulin resistance is expected with great anticipation, including the potential use of adipocytokines such as leptin in certain subsets of patients with lipodystrophic syndrome.

Lifestyle Modification

There are limited data on the effectiveness of diet and nutrition support on body composition and metabolic abnormalities in patients with lipodystrophy; however, a balanced macronutrient intake is recommended.[2,213] The general clinical recommendations have been to follow the standard dietary advice regarding management of dyslipidemia, obesity, insulin resistance, and impaired glucose tolerance, with the goal of attaining ideal body weight. Supplementation with dietary fiber and fish oil containing high doses of omega-3 fatty acids should be encouraged.[214,215] For patients with severe hypertriglyceridemia, an extremely low-fat diet (preferably <15% of daily caloric intake coming from fat) should be advised. Several lines of evidence suggest that the Mediterranean diet, consisting of intake of plentiful fruits, vegetables, whole grains, olive oil, fish, and dairy products, and avoidance of red meat and saturated fats, may be of benefit in patients with HALS; specifically, insulin resistance improved and HDL cholesterol was increased.[216] Patients should also avoid alcohol consumption to prevent chylomicronemia, acute pancreatitis, and hepatic steatosis. For patients with fat atrophy, nutrition and caloric intake need to be assessed in detail.[215] In infants and young children with lipodystrophy, hyperphagia as a result of severe leptin deficiency is a significant issue, as it is associated with metabolic dysregulation, diabetes, hyperlipidemia, and hepatic steatosis. Consulting a dietitian with experience with this condition is recommended for this subpopulation.[2,73]

Studies evaluating the effect of exercise regimens involving resistance training, aerobic exercise, or stretching and relaxation techniques have all shown only modest benefit in improving body composition and metabolic abnormalities in HIV-infected patients.[217,218] Although both strength and endurance training improve peripheral insulin sensitivity, only strength training reduces total body fat in HIV-infected patients with lipodystrophy.[219]

In general, exercise should be advised in the absence of specific contraindications. On the contrary, individuals with specific subtypes of lipodystrophy (i.e., CGL4), who are predisposed to cardiac disorders, should undergo medical evaluation before initiating an exercise program.[2]

Overall, diet and increased physical activity of any kind are essential factors in managing lipodystrophy comorbidities and improving or preventing the clinical manifestations of lipodystrophy. Patients with lipodystrophy should also be advised on smoking cessation and optimal blood pressure control to minimize cardiovascular risk factors.

Management of Insulin Resistance

The treatment strategy for the management of insulin resistance in patients with lipodystrophy is, in general, not different from that in other states of insulin resistance. Traditional insulin sensitizers such as metformin and/or thiazolidinediones can be considered, along with lifestyle modifications. There are no known pharmacologic interactions between antihyperglycemic agents and antiretroviral agents.[220]

Metformin. Metformin, which acts by inhibiting gluconeogenesis in the liver and increasing peripheral glucose utilization, has shown efficacy in improving insulin sensitivity in patients with lipodystrophies.[221] It may also potentially improve fat redistribution in HALS, as indicated by a randomized controlled trial.[222] Other studies have cast doubt on the potential usefulness of metformin, however, by suggesting that metformin may lead to no changes in waist-to-hip ratio and may possibly cause further loss of limb fat.[222-224] Moreover, metformin, particularly in combination with exercise training, may be useful in HIV-infected patients with significant lipohypertrophy and minimal lipoatrophy. Metformin, in combination with lifestyle modifications, is considered as first-line pharmacotherapy to treat individuals with insulin resistance and diabetes. Furthermore, the US Food and Drug Administration (FDA) has approved metformin usage in pediatric patients 10 years of age and older, noting its efficacy and safety in this subpopulation.[2,73]

Thiazolidinediones. The thiazolidinediones (TZDs), which act by stimulating nuclear transcription factor PPARγ, have shown promise in treating patients with lipodystrophies.[225] TZDs result in improved insulin sensitivity, but data on fat redistribution and lipid profiles have been variable and conflicting. Although some studies show that TZDs may increase subcutaneous fat in patients with lipodystrophy,[223,226-229] other studies in patients with HALS showed no statistically significant difference in limb fat gain between rosiglitazone- and placebo-treated groups.[230,231] Rosiglitazone may also be associated with increased LDL and triglyceride levels, which makes it a less desirable drug to treat lipodystrophies. Pioglitazone, with its more favorable lipid profile, may have a more central place in the control of metabolic abnormalities associated with HALS. Randomized studies have demonstrated that treatment with pioglitazone can improve limb fat in HIV-infected patients, although patients may not perceive the clinical benefits.[228] Both pioglitazone and rosiglitazone are of benefit to the metabolic syndrome of HALS. In one placebo-controlled randomized study, treatment with pioglitazone during a 1-year period increased serum adiponectin levels and was as efficacious as treatment with rosiglitazone in improving insulin resistance in patients with HAART-induced metabolic syndrome, in addition to having more favorable outcomes on blood pressure and lipid profiles.[232] A recent study found that, in HIV-positive subjects treated with pioglitazone, administration of recombinant leptin (metreleptin) further reduced fasting insulin concentration and insulin resistance and attenuated postprandial glycemia.[233] These findings indicate that the improvements in hyperinsulinemia and insulin resistance from leptin therapy in HALS may be successfully combined with the increases in total body and peripheral fat induced by pioglitazone. TZDs can also lower blood pressure and improve endothelial function, therefore reducing overall cardiovascular risk in the HALS population.[234] According to the results from several studies, TZDs may improve the metabolic profile in individuals with partial lipodystrophy. Thus, these drugs are suggested for the management of the metabolic complications related to partial lipodystrophy. However, limited evidence exists regarding their use in generalized lipodystrophy, and they should only be applied cautiously in this subgroup.[2,235] More, larger, and longer randomized studies are needed in this area.

Other Treatment Strategies. Individuals with lipodystrophy and diabetes usually require combination therapy with insulin and metformin. Most lipodystrophy patients demonstrate severe insulin resistance and may need large doses. In cases where patients require daily large insulin doses (i.e., >200 units daily), concentrated forms of insulin such as U-500 regular insulin should be considered. The pharmacokinetics of long-acting insulins may be affected when they are injected into lipodystrophic areas. Their long duration of action requires the presence of subcutaneous adipose tissue.[2,73,235,236]

The new antidiabetic drug classes, such as sodium-glucose transporter inhibitors and glucagon-like peptide-1 receptor agonists, display significant renal and cardiovascular benefits, with improvements in cardiovascular outcomes and all-cause mortality, especially in high-risk diabetes patients.[237,238] However, their efficacy and safety in patients with lipodystrophy syndromes have not been systematically confirmed.

Bariatric surgery is the most effective possible weight-loss therapeutic strategy and also demonstrates weight loss–independent effects on glycemic control and metabolism. [239,240] A number of studies have been conducted to explore the impact of Roux-en-Y gastric bypass (RYGB) in patients with FPL. RYGB led to significant weight loss and metabolic improvements.[241-243] However, long-term controlled studies are needed to examine the health outcomes, as well as the overall impact and safety of this intervention, in these specific subpopulations.

Management of Dyslipidemia

The treatment of dyslipidemia in lipodystrophy should follow the same guidelines as in the general population. Lifestyle modification should be emphasized and attempted first. If ineffective, a change of antiretroviral treatment may be considered initially, followed by starting lipid-lowering medications in high-risk HALS patients. Lipodystrophy-related dyslipidemia can be difficult to treat, and multiple agents may be necessary to lower lipids to target ranges.[220] Extra caution may be necessary, given the potential drug–drug interactions among lipid-lowering agents and HIV-specific treatments.

Statins. Statins, such as 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors, are normally used as first-line agents for hypercholesterolemia, especially for patients who have fasting triglyceride levels less than 500 mg/dL. Statins have antiinflammatory, antithrombotic, and endothelial effects that contribute to their overall beneficial effects on reducing mortality from cardiovascular disease. Extensive studies on the use of statins to treat the hyperlipidemia associated with HAART have demonstrated efficacy in lowering total and LDL cholesterol and triglycerides. Specifically, rosuvastatin and pravastatin have been shown to decrease total LDL cholesterol levels.[227,244] Among the statins, pravastatin may also increase subcutaneous and limb fat.[245] However, caution must be exercised in cases of coadministration of statins and HAART. Coadministration of PIs normally results in increased levels of statins, except for pravastatin. Therefore, simvastatin and lovastatin should be avoided in HIV-infected patients receiving PIs, whereas atorvastatin should be used with caution. A higher dose of pravastatin may be necessary to achieve optimal lipid-lowering activity in HALS,[246,247] and coadministration of statins with efavirenz reduces atorvastatin, simvastatin, and pravastatin serum concentrations, meaning that statins need to be administered at higher doses.[248] Tailoring statin therapy to the HAART therapy regimen is thus very important.

Fibrates and Omega-3 Fatty Acids. Fibrates and long-chain omega-3 fatty acids have been extensively used to treat severe hypertriglyceridemia, but they have not been formally studied in patients with lipodystrophy. However, both medications should be used in patients with elevated triglycerides, especially when triglycerides are higher than 500 mg/dL and may be considered for levels greater than 200 mg/dL.[2]

Combination therapy with fibrates and statins may be required to treat dyslipidemia to achieve lipid goals and should be used concomitantly with lifestyle modification.

However, caution regarding their use is advised, given the increased risk of hepatotoxicity and myopathy, especially in the presence of known myositis or muscular dystrophy, and in patients with renal insufficiency. Fibrates are normally well tolerated and efficacious in the context of HALS.[140] Head-to-head comparisons of the efficacy of fenofibrate versus gemfibrozil in HALS are lacking. Omega-3 fatty acids have also proved to be effective in reducing triglycerides in HALS.[249] They are generally well tolerated, but their use may be associated with increased LDL levels.[140,250]

In children and adolescents with lipodystrophy and very high levels of triglycerides, when lifestyle interventions are not sufficient, the use of lipid-lowering drugs should be considered. Fibrates are commonly used in children with severe hypertriglyceridemia who are at high risk of pancreatitis.[251] Omega-3 fatty acids may also prove beneficial.[251,252]. Fibrates alone or in combination with statins can effectively improve the lipid profile. However, data are limited in children, and physicians should pay attention to potential adverse effects.[252,253]

Nicotinic Acid. Nicotinic acid, or niacin, is effective for hypertriglyceridemia, but its use may be limited by adverse effects including flushing, rashes, pruritus, and exacerbation of insulin resistance and hyperuricemia. Extended-release niacin preparations are generally better tolerated in HIV lipodystrophy. The side effects can be controlled with daily aspirin intake.[140] Acipimox, a long-acting niacin analog, may also improve triglycerides and improve insulin sensitivity.[254]

In 2016, after reviewing the data of large cardiovascular outcome studies (HPS2-THRIVE, AIM-HIGH, and ACCORD), the FDA concluded that, when niacin extended-release tablets and fenofibric acid delayed-release capsules are coadministered with statins, the risks outweigh the benefits, and this combination is not recommended.[255-258]

Other Lipid-Lowering Agents and Novel Therapies. Ezetimibe, a cholesterol absorption inhibitor, might be useful in the treatment of statin-intolerant patients or in severe dyslipidemia associated with lipodystrophies.[140] Studies have shown that ezetimibe can provide incremental reduction in LDL cholesterol levels when combined with a statin, but a recent study suggested that the combination treatment may not provide additional reduction in cardiovascular risk.[259] Tetradecylthioacetic acid, cholestin, and L-carnitine have also shown efficacy in controlling dyslipidemia via unknown mechanisms,[140] but their use in HALS remains limited.

Treating extreme hypertriglyceridemia in lipodystrophy patients with conventional dietary and drug therapy may prove to be challenging. Novel lipid-lowering agents directed towards lowering triglyceride levels have been developed, and several potential therapeutic agents are under examination. Genome-wide association studies have uncovered potential targets in multiple circulating factors that may play a vital role in the regulation of lipid metabolism, hepatic steatosis, and insulin sensitivity. Furthermore, clinical pharmacology promotes the development of novel and promising therapeutic strategies, such as the antisense oligonucleotide-based therapies.[260]

Volanesorsen is a second-generation antisense oligonucleotide agent targeted to human apolipoprotein CIII (apoCIII) mRNA. ApoC-III is an important regulator of plasma triglyceride concentration. According to the results of phase III clinical trials, volanesorsen reduced triglycerides significantly compared to baseline.[261,262] Volanesorsen has been approved to treat patients with familial chylomicronemia syndrome.[263] The BROADEN study is an ongoing randomized, double-blind, placebo-controlled trial that enrolled patients with hypertriglyceridemia and FPL. The study aims to evaluate the safety and efficacy of volanesorsen in this subpopulation, and it is expected to be completed in the last quarter of 2021.

Moreover, inhibition of angiopoietin-like protein 3 and 4 affects lipid metabolism and is associated with low serum triglycerides.[264-267]

Trials involving ANPL-3 inhibition either in the form of monoclonal antibodies or antisense oligonucleotides are ongoing and may result in approval of novel treatments for hyperlipidemia. [268,269]

Patients with hypercholesterolemia who do not achieve sufficient LDL reduction despite receiving maximally tolerated first-line lipid-lowering treatment (namely statin drugs) may be prescribed additional therapy with ezetimibe or proprotein convertase subtilisin/kexin type 9 antibody inhibitor (PCSK9i).[270,271] Two PCSK9is have been approved in the United States and the European Union as adjuncts to dietary changes and maximal-tolerated statin therapy. Evolocumab (at a dose of 140 mg biweekly or 420 mg monthly) and alirocumab (75 mg/150 mg every 2 weeks or 300 mg monthly) are human monoclonal antibodies against PCSK9 that are administered subcutaneously. Both drugs are highly potent in reducing LDL (up to approximately 60% in addition to statin monotherapy) and the risk of major cardiovascular events, particularly in high-risk patients, in phase III studies.[272,273]

Another promising approach that is being examined for PCSK9 inhibition is inclisiran. Inclisiran is a chemically synthesized small interfering RNA molecule that reduces PCSK9 synthesis.[274] The ORION program includes several trials that examine the safety and efficacy of inclisiran. According to the findings of phase III placebo-controlled, double-blind, randomized clinical trials, ORION-10 and ORION-11, inclisiran at a dose of 300 mg every 6 months reduced LDL levels by approximately 50%.[275] It should be noted that inclisiran, because of its long duration of action, can be administered subcutaneously twice yearly, unlike conventional monoclonal PCSK9i. However, no available data exist regarding the efficacy and safety of these drugs in patients with lipodystrophy.

Bempedoic acid is a novel, once-daily, oral agent for treating hypercholesterolemia. It is an inhibitor of adenosine triphosphate citrate lyase, which is involved in the cholesterol synthesis pathway in the liver.[276] Bempedoic acid monotherapy was associated with LDL reduction up to 16.5% (vs. placebo) in patients with heterozygous familial hypercholesterolemia and/or atherosclerotic cardiovascular disease and up to 21.4% in patients with statin intolerance after 12 weeks of treatment.[277,278] Moreover, bempedoic acid administration, when added to maximally tolerated statin therapy, may reduce mean LDL by 18% (vs. placebo) in patients with heterozygous familial hypercholesterolemia and/or atherosclerotic cardiovascular disease and by 24% (vs. placebo) in patients with a history of statin intolerance.[279] A meta-analysis of seven randomized clinical trials has shown that bempedoic acid is an important option for patients at high cardiovascular risk and for those with statin intolerance, accompanied by an acceptable safety profile.[280] In 2020, bempedoic acid either as a monotherapy or as a combination with ezetimibe received approval (both in the United States and in the European Union) for use in patients with established atherosclerotic cardiovascular disease who require additional cholesterol lowering or in patients with heterozygous familial hypercholesterolemia.[281] Among the potential adverse events, elevated liver enzymes, tendon rupture, hyperuricemi,a and gouty attacks have been described.[281]

By the end of 2022, the randomized, double-blind, placebo-controlled, CLEAR Outcomes study is expected to be completed. The purpose of this phase III study is to examine the effects of bempedoic acid on major cardiovascular events in statin intolerant patients with or at high risk for cardiovascular disease (NCT02993406).

Finally, plasmapheresis is an option in patients with extreme hypertriglyceridemia associated with significant complications such as acute or relapsing pancreatitis who are unresponsive to conventional therapeutic agents. Plasmapheresis has been found to be effective; however, it must be repeated frequently, and long-term data are not available.[282,283]

Management of Human Immunodeficiency Virus–Infected Patients With Highly Active Antiretroviral Therapy–Induced Metabolic Syndrome

Growth Hormone and Growth Hormone–Releasing Hormone Analogs. Growth hormone (GH) replacement has been proposed to have a promising role in HALS treatment, in that this group of patients is prone to GH deficiency (GHD). It has been observed that basal GH concentrations, overnight GH secretion, and pulse amplitude are reduced in patients with HIV lipodystrophy,[284] although normal pulse frequency is maintained.[285] Relative GHD also appears to be common in HALS, as evidenced by decreased response to a standardized GH–releasing hormone (GHRH) arginine stimulation test.[286] Although the causative relationship between GHD and metabolic abnormalities is not well understood, interventions that normalize GH concentrations have demonstrated efficacy in improving metabolic abnormalities associated with visceral fat accumulation in the HIV-infected population.[285]

Treatment with high doses of GH (2–6 mg/day) has shown to effectively reduce visceral fat and improve lipid parameters.[287-289] However, its application is hampered by significant side effects, including fluid retention, arthralgia, myalgia, carpal tunnel syndrome, and worsening glucose control. There is also the theoretical concern that long-term GH therapy may increase the risk for cancer. In addition, high-dose GH replacement may lead to a further decrease in peripheral subcutaneous fat,[290] which is undesirable in this patient population from a cosmetic point of view. The effect of GH generally diminishes after discontinuation of treatment; however, improvement in facial lipoatrophy for up to 6 months after cessation of treatment has been reported.[291]

Increasing evidence demonstrates that physiologic GH replacement, with doses as low as 2 to 6 mcg/kg/day, can also effectively raise circulating IGF-1 levels and may provide similar benefits in visceral fat reduction but fewer adverse effects.[291-294] Although low-dose GH treatment did not result in worsened insulin sensitivity,[285] this potential side effect and the lack of durability with GH treatment limit its clinical application. GH therapy is not currently FDA-approved for treatment of HALS.

GHRH, the hormone that regulates GH secretion by the pituitary gland, has recently emerged as an alternative treatment option for HALS. GHRH augments endogenous GH pulsatility and may preserve the negative feedback of IGF-1 on the pituitary gland.[285] Initial recombinant forms used, such as Sermorelin or GHRH 1-29 (Geref), have been shown to increase lean body mass, decrease abdominal visceral fat, and decrease truncal fat without altering glucose levels.[286,295,296]

Tesamorelin (Egrifta), a GHRH analog, is the first FDA-approved medication for the treatment of HALS. Two large phase III clinical trials were designed to evaluate the efficacy and safety of tesamorelin, demonstrating evidence of benefit without affecting glycemic control; patients receiving tesamorelin experienced an average visceral fat reduction of 11% at 6 months, and 18% by 12 months, while preserving insulin sensitivity.[297] Trunk fat, waist circumference, and waist-to-hip ratio all significantly improved, and patients reported improved feelings about body image.[295] Treatment with tesamorelin is also associated with improvement of triglycerides and total cholesterol levels,[297] but its effect on HDL seems to be variable[297,298] and needs to be confirmed by long-term controlled studies. Adverse events are usually limited to arthralgia, myalgia, and local site irritation.[295] The current indication for tesamorelin, which is administered by subcutaneous injection of 1 to 2 mg/day in single or divided doses, is HIV-infected patients without active malignancy who are distressed by moderate to severe abdominal fat accumulation. Therapy should not be continued beyond 6 months if there is no treatment response. Similar to other interventions, such as lifestyle modification or metformin, the effect of GHRH analogs ceases to exist upon discontinuation, and while analogs are generally well-tolerated, long-term effects of treatment remain to be elucidated.

Human Immunodeficiency Virus–Associated Lipodystrophy Syndrome–Specific Treatment. In addition to the above-mentioned treatment options, HALS-specific options may also be valuable in the treatment of these patients.

Modification of Highly Active Antiretroviral Therapy. Interruption of antiretroviral therapy is associated with increased mortality and opportunistic infections. Therefore, it should be avoided. Instead, a careful evaluation for cardiovascular risk factors should be conducted before antiretroviral therapy is initiated, and the drug with the fewest metabolic implications and comparable efficacy should be selected. Once lipoatrophy or metabolic complications develop, switching from the offending drug to another agent may be an important strategy. However, the reversal is normally slow and gradual.

In the case of lipoatrophy, the treatment regimen with two thymidine analogs should generally be avoided. Switching to abacavir or tenofovir may partially restore subcutaneous fat.[299] Switching from a PI to an NNRTI or abacavir has not shown any beneficial effects in terms of improvement of lipoatrophy. Regarding metabolic abnormalities, switching from a PI to nevirapine or abacavir has generally resulted in improved total cholesterol and triglycerides, whereas switching to efavirenz has generated less consistent results.[300] Newer-generation PIs (e.g., atazanavir, darunavir, and saquinavir) are associated with favorable lipid profiles. However, because antiretroviral regimens containing PIs normally require low-dose ritonavir for its boosting effect, the impact of switching from one ritonavir-boosted PI to another may be modest.[220]

Uridine. Uridine is a pyridine precursor that reverses mitochondrial toxicity. It may have clinical value in treating HALS induced by pyrimidines such as zalcitabine and stavudine. Uridine can also reverse the cell depletion and lactic acidosis seen with zidovudine and lamivudine combination. It has no effect on lipoatrophy caused by didanosine, a purine analog. Uridine is generally well-tolerated.[158]

Management of Cosmetic Appearance

Autologous fat transplant and implantation of synthetic bulking agents have been used for the cosmetic correction of facial lipoatrophy and appear to be associated with improvements in quality of life. However, long-term, well-designed studies are needed to assess their efficacy and safety.[301] Occasionally, the fat harvested from HIV-infected patients with buffalo hump can cause hypertrophies in the transplanted sites (cheeks) and cause a disfiguring "hamster" appearance. The intradermal injections of synthetic agents (polylactic acid or New-Fill) have been shown to result in a more durable increase in total cutaneous thickness persisting up to 48 weeks.

The efficacy and the safety of surgical treatment of HIV-associated facial lipoatrophy using dermal fillers were assessed in a clinical trial. Four different dermal fillers (poly-L-lactic acid, calcium hydroxylapatite, polyacrylamide, and autologous fat) were examined. According to the results of the study, autologous fat transfer and permanent fillers achieved superior long-term durability for HIV-associated facial lipoatrophy, with high rates of facial volume restoration and patient satisfaction.[302]

Recently, a 10-year follow-up study demonstrated that highly purified liquid-injectable silicone was an effective long-term therapy for HIV-associated lipoatrophy. A microdroplet serial puncture technique was used, and small volumes of the product were injected at monthly or greater intervals to achieve the desired outcome. The most common adverse events that the patients experienced were excess fibroplasia of mild severity that could be treated and, to a lesser extent, severe acute inflammatory reaction presenting as facial edema without warmth or tenderness.[303]

Adipokines in Lipodystrophy

Leptin. A relatively large proportion of patients with lipodystrophy are found to have low leptin levels, including 20% to 30% of HALS patients[153] and the vast majority of patients with generalized lipodystrophy. Apart from its central and peripheral effects on satiety and glucose metabolism,[304,305] leptin has been associated with increased skeletal muscle lipoprotein lipase activity,[304] which may partially explain its ability to preserve lean tissue during weight loss.

The mechanisms through which leptin exerts its role remain under intensive investigation. Animal studies indicate that leptin acts mainly at the hypothalamus, particularly at neurons containing pro-opiomelanocortin and neuropeptide Y, to regulate food intake and fuel partitioning.[305] Functional MRI studies in humans confirm that leptin mediates its "adipostatic" effect through hypothalamic and other brain areas that are important in emotional and cognitive control.[306,307] Although animal and human studies indicate that leptin may also work peripherally (i.e., in the liver, muscle, and white adipose tissue) to affect lipid metabolism,[308,309] other studies suggest that these effects are largely mediated by the CNS.[310] Long-term leptin treatment may also attenuate β cell function and decrease glucose-induced insulin secretion,[305,311] but whether these effects are independent from reduction of insulin resistance remains to be seen.

Leptin Treatment in Generalized Lipodystrophy. The administration of recombinant leptin (r-metHuLeptin or metreleptin) has been tested in the treatment of congenital and acquired non–HIV-related lipodystrophies and has shown amelioration of metabolic abnormalities.[304,312] Several small, open-label studies show that subcutaneous injection of leptin (0.04–0.08 mg/kg/day) in patients with severe generalized lipodystrophy results in significant and sustained weight loss, with decreased fat and lean body mass. The weight loss is associated with a decrease in appetite, calorie intake, and resting energy expenditure. If excessive weight loss occurs, dosing should be decreased.[313-316] Metreleptin treatment has been associated with a significant reduction in hemoglobin A1c (HbA1c) levels, a reduction in fasting plasma glucose, and improvements in insulin sensitivity. Moreover, after metreleptin initiation, some patients were also able to discontinue insulin therapy.[317-319] Furthermore, metreleptin improved gonadotropin secretion, leading to the normal progression of female puberty and normalized menstrual cycles.[320-322]

Hypertriglyceridemia, usually refractory to traditional lipid-lowering agents, is commonly responsive to leptin treatment.[314] The liver volume decreases with leptin treatment, most likely as a result of decreased intrahepatic lipid content.[323,324] Transaminases and hepatocellular injuries that are associated with nonalcoholic steatohepatitis are also reduced.[323] In comparison with patients with generalized lipodystrophies, patients with familial partial lipodystrophy may have a less dramatic response to leptin treatment.[325,326] Because all these studies are uncontrolled and open label, it remains to be proven beyond any doubt whether the beneficial effects observed are r-metHuLeptin–specific.

Metreleptin has been approved in Japan as a leptin therapy for the treatment of lipodystrophy for several years.[327] Despite its declining use for metabolic disorders associated with partial lipodystrophy, the FDA advisory panel in 2014 suggested approval of metreleptin administration in adult and pediatric patients with generalized lipodystrophy. This announcement thus paved the way for a new modality in the treatment of this disorder, probably in the context of risk evaluation and mitigation strategy, given that the above uncontrolled studies have also demonstrated several potentially serious side effects. In individuals with generalized lipodystrophy, metreleptin in combination with dietary recommendations is the first-line treatment strategy for the disease-related metabolic and endocrine comorbidities. Metreleptin use may also be considered in children to prevent these disorders.[2]

Leptin Treatment in Partial Lipodystrophy. Some findings suggest that leptin replacement therapy may be beneficial in a subset of patients with severe partial lipodystrophy. Metreleptin appeared more effective in partial lipodystrophy patients with significant hypoleptinemia and severe metabolic disorders (such as HbA1c >8%, triglycerides >500 mg/dL).[5,197] Currently, metreleptin usage, both in CPL and APL, is off-label in the United States.

Leptin Treatment in Human Immunodeficiency Virus–Associated Lipodystrophy Syndrome. In addition to generalized lipodystrophies, we have reported a modest effect of leptin therapy on metabolic abnormalities in a randomized, placebo-controlled study in patients with HALS. Compared with placebo, r-metHuLeptin therapy administered at 0.02 mg/kg twice a day decreases body weight mass and truncal fat mass but not peripheral fat or lean body mass and improves fasting insulin levels, insulin resistance, and levels of HDL.[194]

Leptin treatment was associated with a 15% decrease in central fat mass, as well as significant improvements in glucose levels, insulin sensitivity, and fasting insulin, despite having no effects on LDL and triglycerides.[194] Other studies of longer duration confirmed these results, demonstrating a 32% decrease in visceral fat, increased hepatic insulin sensitivity, and improved dyslipidemia with metreleptin treatment.[195] The improvements in lipids and abdominal fat were comparable to those reported with metformin and thiazolidinediones in HALS patients and provide an advantage over GH replacement because leptin replacement has not been observed to induce glucose intolerance. Although no direct comparisons are available, the decrease of visceral adipose tissue and the lipid-lowering effect of leptin are either comparable or better than those reported for GHRH analogs.[293,297] Thus, recombinant human leptin holds promise as an agent that could improve HIV-associated lipoatrophy and features of the associated metabolic syndrome. Future randomized, placebo-controlled trials of adequate duration are needed to fully quantitate efficacy and clarify the side-effect profiles of metreleptin in HALS.

Clinical Use of Leptin: Benefits and Risks. Replacement of leptin may provide additional benefits beyond ameliorating metabolic abnormalities, which may include alleviation of the glomerular injury in humans with lipodystrophies[321,328] and improvement of pituitary-gonadal function in lipodystrophic patients with severe leptin deficiency.[322,329] It remains inconclusive, however, whether recombinant leptin therapy will affect such systems and/or bone density in the long term, given the uncontrolled nature of prior studies. Small human studies suggest that long-term treatment may not affect bone density in patients with generalized lipodystrophies with regular menses indicating normoleptinemia.[315,321,330]

Metreleptin combined with lifestyle modification is currently indicated for adults and pediatric patients with generalized lipodystrophy (congenital or acquired). The effectiveness and safety of metrelpetin for the treatment of partial lipodystrophy and liver disease, including nonalcoholic steatohepatitis, have not been established. Additionally, metreleptin is not indicated for patients with HIV-related lipodystrophy or metabolic disease (including hypertriglyceridemia and diabetes), without concurrent evidence of generalized lipodystrophy. The recommended daily dosage for patients with body weight 40 kg or under is 0.06 mg/kg/day at baseline, and may then be increased or decreased by 0.02 mg/kg to a maximum daily dose of 0.13 mg/kg. For patients with body weight more than 40 kg, the daily dosage for males is 2.5 mg/day, and then it may be increased or decreased by 1.25 to 2.5 mg/day to a maximum dose of 10 mg/day. The starting dose for females is 5 mg/day, which can be increased or decreased by 1.25 to 2.5 mg/day to a maximum amount of 10 mg/day. It should be noted that the effects of leptin do not appear to be sustained after the therapy is discontinued.[206,210,331,332]

Several patients experience side effects associated with metreleptin therapy. The most common and clinically significant adverse reactions are hypoglycemia (especially in insulin-treated patients), headache, abdominal pain, and decreased weight.[319] Other potential side effects of metreleptin treatment include deterioration of renal function[328] and possible occurrence of T-cell lymphomas. T-cell lymphoma has been reported in individuals with AGL, independent of treatment with metreleptin. Thus, further investigation is required. Physicians should evaluate the potential benefits and risks of metreleptin therapy in patients with significant hematologic abnormalities or AGL. Moreover, hypersensitivity reactions (including urticaria, anaphylaxis, generalized rash) and autoimmune disorder progression have been described in patients receiving metreleptin. *In vivo*, antimetreleptin antibodies with neutralizing activity have been identified. Individuals with loss of efficacy or severe infections during metreleptin treatment should be tested for neutralizing antibodies. The consequences of this neutralizing antibody response to leptin are not well characterized and may include inhibition of endogenous leptin activity and loss of metreleptin efficacy. However, potential risks are unclear and should be weighed against benefits.[320,332,334]

FUTURE PERSPECTIVES

Adiponectin

Adiponectin, as well as its receptors AdipoR1 and AdipoR2, are attractive future targets for drug development, given that levels of this hormone are lower in patients with HALS.[188] Based on the fact that adiponectin is decreased in lipodystrophic states and evidence from mouse studies showing that adiponectin administration can improve insulin sensitivity, dyslipidemia, sustained weight loss without reducing food intake, and production of proinflammatory cytokines,[335,336] adiponectin analog administration in replacement doses may prove to be an effective future treatment option. Because of its complex molecular structure, synthetic adiponectin is not yet available for therapy in humans. However, increasing adiponectin levels by pharmacologic means is likely to become a valuable addition to our armamentarium as a treatment option for lipodystrophy. Medications that increase endogenous levels of adiponectin in lipodystrophy, such as pioglitazone or INT-131, a selective PPARγ modulator currently in development, may prove useful as therapeutic possibilities, because they increase circulating adiponectin levels. Further research is needed to determine the efficacy and long-term durability of symptomatic improvement of lipodystrophy and its associated metabolic disorders, as well as short- and long-term morbidity and mortality.

For your free Expert Consult eBook with bibliographic citations, as well as the ability to take notes, highlight important content, search the full text, and more, visit http://www.ExpertConsult.Inkling.com.

REFERENCES

1. Fiorenza CG, Chou SH, Mantzoros CS. Lipodystrophy: pathophysiology and advances in treatment. *Nat Rev Endocrinχol*. 2011;7:137–150.
2. Brown RJ, Araujo-Vilar D, Cheung PT, et al. The diagnosis and management of lipodystrophy syndromes: a multi-society practice guideline. *J Clin Endocrinol Metab*. 2016;101:4500–4511.
3. Mann JP, Savage DB. What lipodystrophies teach us about the metabolic syndrome. *J Clin Invest*. 2019;130:4009–4021.
4. Bindlish S, Presswala LS, Schwartz F. Lipodystrophy: syndrome of severe insulin resistance. *Postgrad Med*. 2015;127:511–516.
5. Diker-Cohen T, Cochran E, Gorden P, et al. Partial and generalized lipodystrophy: comparison of baseline characteristics and response to metreleptin. *J Clin Endocrinol Metab*. 2015;100:1802–1810.
6. Akinci B, Meral R, Oral EA. Phenotypic and genetic characteristics of lipodystrophy: pathophysiology, metabolic abnormalities, and comorbidities. *Curr Diab Rep*. 2018;18:143.
7. Goodpaster BH. Measuring body fat distribution and content in humans. *Curr Opin Clin Nutr Metab Care*. 2002;5:481–487.
8. Gulizia R, Vercelli A, Gervasoni C, et al. Comparability of echographic and tomographic assessments of body fat changes related to the HIV associated adipose redistribution syndrome (HARS) in antiretroviral treated patients. *Ultrasound Med Biol*. 2008;34:1043–1048.

9. Handelsman Y, Oral EA, Bloomgarden ZT, et al. The clinical approach to the detection of lipodystrophy—an AACE consensus statement. *Endocr Pract*. 2013;19:107–116.
10. Haque WA, Shimomura I, Matsuzawa Y, et al. Serum adiponectin and leptin levels in patients with lipodystrophies. *J Clin Endocrinol Metab*. 2002;87:2395.
11. Araújo-Vilar D, Santini F. Diagnosis and treatment of lipodystrophy: a step-by-step approach. *J Endocrinol Invest*. 2019;42:61–73.
12. Grigoraş A, Amalinei C, Balan RA, et al. Adipocytes spectrum: from homeostasia to obesity and its associated pathology. *Ann Anat*. 2018;219:102–120.
13. Bastard JP, Caron M, Vidal H, et al. Association between altered expression of adipogenic factor SREBP1 in lipoatrophic adipose tissue from HIV-1-infected patients and abnormal adipocyte differentiation and insulin resistance. *Lancet*. 2002;359:1026–1031.
14. Domingo P, Matias-Guiu X, Pujol RM, et al. Subcutaneous adipocyte apoptosis in HIV-1 protease inhibitor-associated lipodystrophy. *AIDS*. 1999;13:2261–2267.
15. Reitman ML, Arioglu E, Gavrilova O, et al. Lipoatrophy revisited. *Trends Endocrinol Metab*. 2000;11:410–416.
16. Hafidi ME, Buelna-Chontal M, Sánchez-Muñoz F, et al. Adipogenesis: a necessary but harmful strategy. *Int J Mol Sci*. 2019;20:3675.
17. Cortés VA, Fernández-Galilea M. Lipodystrophies: adipose tissue disorders with severe metabolic implications. *J Physiol Biochem*. 2015;71:471–478.
18. Lagathu C, Béréziat V, Gorwood J, et al. Metabolic complications affecting adipose tissue, lipid and glucose metabolism associated with HIV antiretroviral treatment. *Expert Opin Drug Saf*. 2019;18:829–840.
19. Hussain I, Patni N, Garg A. Lipodystrophies, dyslipidaemias and atherosclerotic cardiovascular disease. *Pathology*. 2019;51:202–212.
20. Berardinelli W. An undiagnosed endocrinometabolic syndrome: report of 2 cases. *J Clin Endocrinol Metab*. 1954;14:193–204.
21. Seip M. Lipodystrophy and gigantism with associated endocrine manifestations. A new diencephalic syndrome? *Acta Paediatr*. 1959;48:555–574.
22. Magré J, Delépine M, Van Maldergem L, et al. Prevalence of mutations in AGPAT2 among human lipodystrophies. *Diabetes*. 2003;52:1573–1578.
23. Simha V, Garg A. Phenotypic heterogeneity in body fat distribution in patients with congenital generalized lipodystrophy caused by mutations in the AGPAT2 or seipin genes. *J Clin Endocrinol Metab*. 2003;88:5433–5437.
24. Garg A. Acquired and inherited lipodystrophies. *N Engl J Med*. 2004;350:1220–1234.
25. Garg A, Misra A. Lipodystrophies: rare disorders causing metabolic syndrome. *Endocrinol Metab Clin North Am*. 2004;33:305–331.
26. Szymanski KM, Binns D, Bartz R, et al. The lipodystrophy protein seipin is found at endoplasmic reticulum lipid droplet junctions and is important for droplet morphology. *Proc Natl Acad Sci U S A*. 2007;104:20890–20895.
27. Agarwal AK, Simha V, Oral EA, et al. Phenotypic and genetic heterogeneity in congenital generalized lipodystrophy. *J Clin Endocrinol Metab*. 2003;88:4840–4847.
28. Van Maldergem L, Magré J, Khallouf TE, et al. Genotype-phenotype relationships in Berardinelli-Seip congenital lipodystrophy. *J Med Genet*. 2002;39:722–733.
29. Patni N, Garg A. Congenital generalized lipodystrophies—new insights into metabolic dysfunction. *Nat Rev Endocrinol*. 2015;11:522–534.
30. Garg A, Fleckenstein JL, Peshock RM, et al. Peculiar distribution of adipose tissue in patients with congenital generalized lipodystrophy. *J Clin Endocrinol Metab*. 1992;75:358–361.
31. Simha V, Agarwal AK, Aronin PA, et al. Novel subtype of congenital generalized lipodystrophy associated with muscular weakness and cervical spine instability. *Am J Med Genet*. 2008;146A:2318–2326.
32. Brunzell JD, Shankle SW, Bethune JE. Congenital generalized lipodystrophy accompanied by cystic angiomatosis. *Ann Intern Med*. 1968;69:501–516.
33. Fleckenstein JL, Garg A, Bonte FJ, et al. The skeleton in congenital, generalized lipodystrophy: evaluation using whole-body radiographic surveys, magnetic resonance imaging and technetium-99m bone scintigraphy. *Skeletal Radiol*. 1992;21:381–386.
34. Guell-Gonzalez JR, Mateo DA, Alavez-Martin E, et al. Bone lesions in congenital generalised lipodystrophy. *Lancet*. 1971;2:104–105.
35. Bhayana S, Siu VM, Joubert GI, et al. Cardiomyopathy in congenital complete lipodystrophy. *Clin Genet*. 2002;61:283–287.
36. Bjornstad PG, Foerster A, Ihlen H. Cardiac findings in generalized lipodystrophy. *Acta Paediatr Suppl*. 1996;413:39–43.
37. Rheuban KS, Blizzard RM, Parker MA, et al. Hypertrophic cardiomyopathy in total lipodystrophy. *J Pediatr*. 1986;109:301–302.
38. Garg A. Clinical review: lipodystrophies: genetic and acquired body fat disorders. *J Clin Endocrinol Metab*. 2011;96:3313–3325.
39. Lima JG, Nobrega LH, de Lima NN, et al. Clinical and laboratory data of a large series of patients with congenital generalized lipodystrophy. *Diabetol Metab Syndr*. 2016;8:23.
40. Agarwal AK, Arioglu E, De Almeida S, et al. AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nat Genet*. 2002;31:21–23.
41. Garg A, Wilson R, Barnes R, et al. A gene for congenital generalized lipodystrophy maps to human chromosome 9q34. *J Clin Endocrinol Metab*. 1999;84:3390–3394.
42. Shindou H, Hishikawa D, Harayama T, et al. Recent progress on acyl CoA: lysophospholipid acyltransferase research. *J Lipid Res*. 2009;50:S46–S51.
43. Leung DW. The structure and functions of human lysophosphatidic acid acyl-transferases. *Front Biosci*. 2001;6:944–953.
44. Subauste AR, Das AK, Li X, et al. Alterations in lipid signaling underlie lipodystrophy secondary to AGPAT2 mutations. *Diabetes*. 2012;61:2922–2931.
45. Craveiro Sarmento AS, Ferreira LC, Lima JG, et al. The worldwide mutational landscape of Berardinelli-Seip congenital lipodystrophy. *Mutat Res*. 2019;781:30–52.
46. Ceccarini G, Magno S, Pelosini C, et al. Congenital generalized lipoatrophy (Berardinelli-Seip Syndrome) type 1: description of novel AGPAT2 homozygous variants showing the highly heterogeneous presentation of the disease. *Front Endocrinol (Lausanne)*. 2020;11:39.
47. Magré J, Delépine M, Khallouf E, et al. Identification of the gene altered in Berardinelli-Seip congenital lipodystrophy on chromosome 11q13. *Nat Genet*. 2001;28:365–370.
48. Cartwright BR, Binns DD, Hilton CL, et al. Seipin performs dissectible functions in promoting lipid droplet biogenesis and regulating droplet morphology. *Mol Biol Cell*. 2015;26:726–739.
49. Talukder MM, Sim MF, O'Rahilly S, et al. Seipin oligomers can interact directly with AGPAT2 and lipin 1, physically scaffolding critical regulators of adipogenesis. *Mol Metab*. 2015;4:199–209.
50. Payne VA, Grimsey N, Tuthill A, et al. The human lipodystrophy gene BSCL2/Seipin may be essential for normal adipocyte differentiation. *Diabetes*. 2008;57:2055–2060.
51. Chen W, Yechoor VK, Chang BH, et al. The human lipodystrophy gene product Berardinelli-Seip congenital lipodystrophy 2/seipin plays a key role in adipocyte differentiation. *Endocrinology*. 2008;150:4552–4561.
52. Kim CA, Delepine M, Boutet E, et al. Association of a homozygous nonsense caveolin-1 mutation with Berardinelli-Seip congenital lipodystrophy. *J Clin Endocrinol Metab*. 2008;93:1129–1134.
53. Hayashi YK, Matsuda C, Ogawa M, et al. Human PTRF mutations cause secondary deficiency of caveolins resulting in muscular dystrophy with gen- eralized lipodystrophy. *J Clin Invest*. 2009;119:2623–2633.
54. Razani B, Lisanti MP. Caveolin-deficient mice: insights into caveolar function human disease. *J Clin Invest*. 2001;108:1553–1561.
55. Williams TM, Lisanti MP. Caveolin-1 in oncogenic transformation, cancer, and metastasis. *Am J Physiol Cell Physiol*. 2005;288:C494–C506.
56. Asterholm IW, Mundy DI, Weng J, et al. Altered mitochondrial function and metabolic inflexibility associated with loss of caveolin-1. *Cell Metab*. 2012;15:171–185.
57. Garg A, Agarwal AK. Caveolin-1: a new locus for human lipodystrophy. *J Clin Endocrinol Metab*. 2008;93:1183–1185.
58. Misra A, Garg A. Clinical features and metabolic derangements in acquired generalized lipodystrophy: case reports and review of the literature. *Medicine (Baltimore)*. 2003;82:129–146.
59. Taniguchi T, Maruyama N, Ogata T, et al. PTRF/Cavin-1 deficiency causes cardiac dysfunction accompanied by cardiomyocyte hypertrophy and cardiac fibrosis. *PLoS One*. 2016;11:e0162513.

60. Vaidyanathan R, Reilly L, Eckhardt LL. Caveolin-3 microdomain: arrhythmia implications for potassium inward rectifier and cardiac sodium channel. *Front Physiol*. 2018;9:1548.
61. Liu L, Pilch PF. A critical role of cavin (polymerase I and transcript release factor) in caveolae formation and organization. *J Biol Chem*. 2008;283:4314–4322.
62. Perez-Diaz S, Johnson LA, DeKroon RM, et al. Polymerase I and transcript release factor (PTRF) regulates adipocyte differentiation and determines adipose tissue expandability. *FASEB J*. 2014;28:3769–3779.
63. Shastry S, Delgado MR, Dirik E, et al. Congenital generalized lipodystrophy, type 4 (CGL4) associated with myopathy due to novel PTRF mutations. *Am J Med Genet*. 2010;152A:2245–2253.
64. Rajab A, Straub V, McCann LJ, et al. Fatal cardiac arrhythmia and long-QT syndrome in a new form of congenital generalized lipodystrophy with muscle rippling (CGL4) due to PTRF-CAVIN mutations. *PLoS Genet*. 2010;6:e1000874.
65. Park DS, Woodman SE, Schubert W, et al. Caveolin-1/3 double-knockout mice are viable, but lack both muscle and non-muscle caveolae, and develop a severe cardiomyopathic phenotype. *Am J Pathol*. 2002;160:2207–2217.
66. Zhang L, Yin H, Jiao L, et al. Abnormal downregulation of caveolin-3 mediates the pro-fibrotic action of microRNA-22 in a model of myocardial infarction. *Cell Physiol Biochem*. 2018;45:1641–1653.
67. Vaidyanathan R, Reilly L, Eckhardt LL. Caveolin-3 microdomain: arrhythmia implications for potassium inward rectifier and cardiac sodium channel. *Front Physiol*. 2018;9:1548.
68. Kim SY, Kim KH, Schilling JM, et al. Protective role of cardiac-specific overexpression of caveolin-3 in cirrhotic cardiomyopathy. *Am J Physiol Gastrointest Liver Physiol*. 2020;318:G531–G541.
69. Dyment DA, Gibson WT, Huang L, et al. Biallelic mutations at PPARG cause a congenital, generalized lipodystrophy similar to the Berardinelli-Seip syndrome. *Eur J Med Genet*. 2014;57:524–526.
70. Broers JL, Ramaekers FC, Bonne G, et al. Nuclear lamins: laminopathies and their role in premature ageing. *Physiol Rev*. 2006;86:967–1008.
71. Jacob KN, Garg A. Laminopathies: multisystem dystrophy syndromes. *Mol Genet Metab*. 2006;87:289–302.
72. Hussain I, Patni N, Ueda M, et al. A novel generalized lipodystrophy-associated progeroid syndrome due to recurrent heterozygous LMNA p.T10I mutation. *J Clin Endocrinol Metab*. 2018;103:1005–1014.
73. Özen S, Akıncı B, Oral EA. Current diagnosis, treatment and clinical challenges in the management of lipodystrophy syndromes in children and young people. *J Clin Res Pediatr Endocrinol*. 2020;12:17–28.
74. Patni N, Hatab S, Xing C, et al. A novel autosomal recessive lipodystrophy syndrome due to homozygous LMNA variant. *J Med Genet*. 2020;57:422–426.
75. Knebel B, Kotzka J, Lehr S, et al. A mutation in the c-fos gene associated with congenital generalized lipodystrophy. *Orphanet J Rare Dis*. 2013;8:119.
76. Huemer C, Kitson H, Malleson PN, et al. Lipodystrophy in patients with juvenile dermatomyositis—evaluation of clinical and metabolic abnormalities. *J Rheumatol*. 2001;28:610–615.
77. Pope E, Janson A, Khambalia A, et al. Childhood acquired lipodystrophy: a retrospective study. *J Am Acad Dermatol*. 2006;55:947–950.
78. Bingham A, Mamyrova G, Rother KI, et al. Predictors of acquired lipodystrophy in juvenile-onset dermatomyositis and a gradient of severity. *Medicine (Baltimore)*. 2008;87:70–86.
79. McCann LJ, Juggins AD, Maillard SM, et al. The Juvenile Dermatomyositis National Registry and Repository (UK and Ireland)—clinical characteristics of children recruited within the first 5 yr. *Rheumatology (Oxford)*. 2006;45:1255–1260.
80. Singh S, Bansal A. Twelve years experience of juvenile dermatomyositis in North India. *Rheumatol Int*. 2006;26:510–515.
81. Verma S, Singh S, Bhalla AK, et al. Study of subcutaneous fat in children with juvenile dermatomyositis. *Arthritis Rheum*. 2006;55:564–568.
82. Sahinoz M, Khairi S, Cuttitta A, et al. Potential association of LMNA-associated generalized lipodystrophy withvjuvenile dermatomyositis. *Clin Diabetes Endocrinol*. 2018;4:6.
83. Billings JK, Milgraum SS, Gupta AK, et al. Lipoatrophic panniculitis: a possible autoimmune inflammatory disease of fat. Report of three cases. *Arch Dermatol*. 1987;123:1662–1666.
84. Corvillo F, Aparicio V, López-Lera A, et al. Autoantibodies against perilipin 1 as a cause of acquired generalized lipodystrophy. *Front Immunol*. 2018;9:2142.
85. Haddad N, Vidal-Trecan T, Baroudjian B, et al. Acquired generalized lipodystrophy under immune checkpoint inhibition. *Br J Dermatol*. 2020;182:477–480.
86. Falcao CK, Cabral MCS, Mota JM, et al. Acquired lipodystrophy associated with nivolumab in a patient with advanced renal cell carcinoma. *J Clin Endocrinol Metab*. 2019;104:3245–3248.
87. Jehl A, Cugnet-Anceau C, Vigouroux C, et al. Acquired generalized lipodystrophy: a new cause of anti-PD-1 immune-related diabetes. *Diabetes Care*. 2019;42:2008–2010.
88. Arioglu E, Andewelt A, Diabo C, et al. Clinical course of the syndrome of autoantibodies to the insulin receptor (type B insulin resistance): a 28-year perspective. *Medicine (Baltimore)*. 2002;81:87–100.
89. Feve B. Adipogenesis: cellular and molecular aspects. *Best Pract Res Clin Endocrinol Metab*. 2005;19:483–499.
90. Fischer-Posovszky P, Hebestreit H, Hofmann AK, et al. Role of CD95-mediated adipocyte loss in autoimmune lipodystrophy. *J Clin Endocrinol Metab*. 2006;91:1129–1135.
91. Corvillo F, Akinci B. An overview of lipodystrophy and the role of the complement system. *Mol Immunol*. 2019;112:223–232.
92. Savage DB, Semple RK, Clatworthy MR, et al. Complement abnormalities in acquired lipodystrophy revisited. *J Clin Endocrinol Metab*. 2009;94:10–16.
93. Herbst KL, Tannock LR, Deeb SS, et al. Köbberling type of familial partial lipodystrophy: an under-recognized syndrome. *Diabetes Care*. 2003;26:1819–1824.
94. Guillin-Amarelle C, Sanchez-Iglesias S, Castro-Pais A, et al. Type 1 familial partial lipodystrophy: understanding the Kobberling syndrome. *Endocrine*. 2016;54:411–421.
95. Lotta LA, Gulati P, Day FR, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*. 2017;49:17–26.
96. Vantyghem MC, Vincent-Desplanques D, Defrance-Faivre F, et al. Fertility and obstetrical complications in women with LMNA-related familial partial lipodystrophy. *J Clin Endocrinol Metab*. 2008;93:2223–2229.
97. Hegele RA, Joy TR, Al Attar SA, et al. Thematic review series: Adipocyte Biology. Lipodystrophies: windows on adipose biology and metabolism. *J Lipid Res*. 2007;48:1433–1444.
98. Jeru I, Vatier C, Araujo-Vilar D, et al. Clinical utility gene card for: familial partial lipodystrophy. *Eur J Hum Genet*. 2017:25.
99. Hegele RA, Cao H, Huff MW, et al. LMNA R482Q mutation in partial lipodystrophy associated with reduced plasma leptin concentration. *J Clin Endocrinol Metab*. 2000;85:3089–3093.
100. Agarwal AK, Barnes RI, Garg A. Genetic basis of congenital generalized lipodystrophy. *Int J Obes Relat Metab Disord*. 2004;28:336–339.
101. Araujo-Vilar D, Lattanzi G, Gonzalez-Mendez B, et al. Site-dependent differences in both prelamin A and adipogenic genes in subcutaneous adipose tissue of patients with type 2 familial partial lipodystrophy. *J Med Genet*. 2009;46:40–48.
102. Capanni C, Mattioli E, Columbaro M, et al. Altered pre-lamin A processing is a common mechanism leading to lipodystrophy. *Hum Mol Genet*. 2005;14:1489–1502.
103. Lloyd DJ, Trembath RC, Shackleton S. A novel interaction between lamin A and SREBP1: implications for partial lipodystrophy and other laminopathies. *Hum Mol Genet*. 2002;11:769–777.
104. Resende ATP, Martins CS, Bueno AC, et al. Phenotypic diversity and glucocorticoid sensitivity in patients with familial partial lipodystrophy type 2. *Clin Endocrinol (Oxf)*. 2019;91:94–103.
105. Vasandani C, Li X, Sekizkardes H, et al. Diagnostic value of anthropometric measurements for familial partial lipodystrophy, Dunnigan variety. *J Clin Endocrinol Metab*. 2020;105:2132–2141.
106. Agarwal AK, Garg A. A novel heterozygous mutation in peroxisome proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy. *J Clin Endocrinol Metab*. 2002;87:408–411.
107. Al Shali K, Cao H, Knoers N, et al. A single-base mutation in the peroxisome proliferator-activated receptor gamma4 promoter associated with

altered in vitro expression and partial lipodystrophy. *J Clin Endocrinol Metab*. 2004;89:5655–5660.

108. Ludtke A, Buettner J, Wu W, et al. Peroxisome proliferator-activated receptor-gamma C190S mutation causes partial lipodystrophy. *J Clin Endocrinol Metab*. 2007;92:2248–2255.
109. Monajemi H, Zhang L, Li G, et al. Familial partial lipodystrophy phenotype resulting from a single-base mutation in deoxyribonucleic acid-binding domain of peroxisome proliferator-activated receptor-gamma. *J Clin Endocrinol Metab*. 2007;92:1606–1612.
110. Barroso I, Gurnell M, Crowley VE, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999;402:880–883.
111. Padova G, Prudente S, Vinciguerra F, et al. The novel loss of function Ile354Val mutation in PPARG causes familial partial lipodystrophy. *Acta Diabetol*. 2020;57:589–596.
112. Gandotra S, Le Dour C, Bottomley W, et al. Perilipin deficiency and autosomal dominant partial lipodystrophy. *N Engl J Med*. 2011;364:740–748.
113. Chen RX, Zhang L, Ye W, et al. The renal manifestations of type 4 familial partial lipodystrophy: a case report and review of literature. *BMC Nephrol*. 2018;19:111.
114. Rubio-Cabezas O, Puri V, Murano I, et al. Partial lipodystrophy and insulin resistant diabetes in a patient with a homozygous nonsense mutation in CIDEC. *EMBO Mol Med*. 2009;1:280–287.
115. Li F, Gu Y, Dong W, et al. Cell death-inducing DFF45-like effector, a lipid droplet-associated protein, might be involved in the differentiation of human adipocytes. *FEBS J*. 2010;277:4173–4183.
116. Wang H, Ti Y, Zhang JB, et al. Single nucleotide polymorphisms in CIDEC gene are associated with metabolic syndrome components risks and antihypertensive drug efficacy. *Oncotarget*. 2017;8:27481–27488.
117. Zolotov S, Xing C, Mahamid R, et al. Homozygous LIPE mutation in siblings with multiple symmetric lipomatosis, partial lipodystrophy, and myopathy. *Am J Med Genet A*. 2017;173:190–194.
118. Farhan SM, Robinson JF, McIntyre AD, et al. A novel LIPE nonsense mutation found using exome sequencing in siblings with late-onset familial partial lipodystrophy. *Can J Cardiol*. 2014;30:1649–1654.
119. George S, Rochford JJ, Wolfrum C, et al. A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science*. 2004;304:1325–1328.
120. Chehab FF. Obesity and lipodystrophy—where do the circles intersect? *Endocrinology*. 2008;149:925–934.
121. Cao H, Alston L, Ruschman J, et al. Heterozygous CAV1 frameshift mutations (MIM 601047) in patients with atypical partial lipodystrophy and hypertriglyceridemia. *Lipids Health Dis*. 2008;7:3.
122. Payne F, Lim K, Girousse A, et al. Mutations disrupting the Kennedy phosphatidylcholine pathway in humans with congenital lipodystrophy and fatty liver disease. *Proc Natl Acad Sci U S A*. 2014;111:8901–8906.
123. Miyoshi Y, Akagi M, Agarwal AK, et al. Severe mandibuloacral dysplasia caused by novel compound heterozygous ZMPSTE24 mutations in two Japanese siblings. *Clin Genet*. 2008;73:535–544.
124. Hitzert MM, van der Crabben SN, Baldewsingh G, et al. Mandibuloacral dysplasia type B (MADB): a cohort of eight patients from Suriname with a homozygous founder mutation in ZMPSTE24 (FACE1), clinical diagnostic criteria and management guidelines. *Orphanet J Rare Dis*. 2019;14:294.
125. Novelli G, Muchir A, Sangiuolo F, et al. Mandibuloacral dysplasia is caused by a mutation in LMNA-encoding lamin A/C. *Am J Hum Genet*. 2002;71:426–431.
126. Agarwal AK, Zhou XJ, Hall RK, et al. Focal segmental glomerulosclerosis in patients with mandibuloacral dysplasia owing to ZMPSTE24 deficiency. *J Investig Med*. 2006;54:208–213.
127. Garg A, Hernandez MD, Sousa AB, et al. An autosomal recessive syndrome of joint contractures, muscular atrophy, microcytic anemia, and panniculitis-associated lipodystrophy. *J Clin Endocrinol Metab*. 2010;95:E58–E63.
128. Torrelo A, Patel S, Colmenero I, et al. Chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature (CANDLE) syndrome. *J Am Acad Dermatol*. 2010;62:489–495.
129. Ramot Y, Czarnowicki T, Maly A, et al. Chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature syndrome: a case report. *Pediatr Dermatol*. 2010;28:538–541.
130. Cardis MA, Montealegre Sanchez GA, Goldbach-Mansky R, et al. Recurrent fevers, progressive lipodystrophy, and annular plaques in a child. *J Am Acad Dermatol*. 2019;80:291–295.
131. Boyadzhiev M, Marinov L, Boyadzhiev V, et al. Disease course and treatment effects of a JAK inhibitor in a patient with CANDLE syndrome. *Pediatr Rheumatol Online J*. 2019;17:19.
132. Eisinger AJ, Shortland JR, Moorhead PJ. Renal disease in partial lipodystrophy. *Q J Med*. 1972;41:343–354.
133. Sissons JG, West RJ, Fallows J, et al. The complement abnormalities of lipodystrophy. *N Engl J Med*. 1976;294:461–465.
134. Allen BR. Partial lipodystrophy. *Br J Dermatol*. 1978;99:S48–49.
135. Corvillo F, Ceccarini G, Nozal P, et al. Immunological features of patients affected by Barraquer-Simons syndrome. *Orphanet J Rare Dis*. 2020;15:9.
136. Hegele RA, Cao H, Liu DM, et al. Sequencing of the reannotated LMNB2 gene reveals novel mutations in patients with acquired partial lipodystrophy. *Am J Hum Genet*. 2006;79:383–389.
137. Guallar JP, Rojas-Garcia R, Garcia-Arumi E, et al. Impaired expression of mitochondrial and adipogenic genes in adipose tissue from a patient with acquired partial lipodystrophy (Barraquer-Simons syndrome): a case report. *J Med Case Rep*. 2008;2:284.
138. Shlay JC, Sharma S, Peng G, et al. The effect of individual antiretroviral drugs on body composition in HIV-infected persons initiating highly active antiretroviral therapy. *J Acqui Immune Defic Syndr*. 2009;51: 298–304.
139. Guaraldi G, Stentarelli C, Zona S, et al. HIV-associated lipodystrophy: Impact of antiretroviral therapy. *Drugs*. 2013;73:1431–1450.
140. Oh J, Hegele RA. HIV-associated dyslipidaemia: pathogenesis and treatment. *Lancet Infect Dis*. 2007;7:787–796.
141. Depairon M, Chessex S, Sudre P, et al. Premature atherosclerosis in HIV-infected individuals—focus on protease inhibitor therapy. *AIDS*. 2001;15:329–334.
142. Maggi P, Serio G, Epifani G, et al. Premature lesions of the carotid vessels in HIV-1-infected patients treated with protease inhibitors. *AIDS*. 2000;14:F123–F128.
143. Seminari E, Pan A, Voltini G, et al. Assessment of atherosclerosis using carotid ultrasonography in a cohort of HIV-positive patients treated with protease inhibitors. *Atherosclerosis*. 2002;162:433–438.
144. Stein JH, Klein MA, Bellehumeur JL, et al. Use of human immunodeficiency virus-1 protease inhibitors is associated with atherogenic lipoprotein changes and endothelial dysfunction. *Circulation*. 2001;104:257–262.
145. Grinspoon S, Carr A. Cardiovascular risk and body-fat abnormalities in HIV-infected adults. *N Engl J Med*. 2005;352:48–62.
146. Swami A. Metabolic syndrome and HIV infection. *J HIV Retro Virus*. 2016;2:1–5.
147. Calza L, Colangeli V, Magistrelli E, et al. Prevalence of metabolic syndrome in HIV-infected patients naive to antiretroviral therapy or receiving a first-line treatment. *HIV Clin Trials*. 2017;18:110–117.
148. Sacilotto LB, Pereira PCM, Manechini JPV, et al. Body composition and metabolic syndrome components on lipodystrophy different subtypes associated with HIV. *J Clin Nutr Metab*. 2017;1:1–8.
149. Srinivasa S, Grinspoon SK. Metabolic and body com- position effects of newer antiretrovirals in HIV-infected pa- tients. *Eur J Endocrinol*. 2014;170:R185–R202.
150. Lazzaretti RK, Gasparotto AS, Sassi MGM, et al. Genetic markers associated to dyslipidemia in HIV-infected individuals on HAART. *Sci World*. 2013;J 26:1–10.
151. Gasparotto AS, Sprinz E, Lazzaretti RK, et al. Genetic polymorphisms in estrogen receptors and sexual dimorphism in fat redistribution in HIV-infected patients on HAART. *AIDS J*. 2012;26:19–26.
152. Sweeney LL, Brennan AM, Mantzoros CS. The role of adipokines in relation to HIV lipodystrophy. *AIDS*. 2007;21:895–904.
153. Nagy GS, Tsiodras S, Martin LD, et al. Human immunodeficiency virus type 1-related lipoatrophy and lipohypertrophy are associated with serum concentrations of leptin. *Clin Infect Dis*. 2003;36:795–802.

154. Jacobson DL, Knox T, Spiegelman D, et al. Prevalence of, evolution of, and risk factors for fat atrophy and fat deposition in a cohort of HIV-infected men and women. *Clin Infect Dis*. 2005;40:1837–1845.
155. Duong M, Petit JM, Piroth L, et al. Association between insulin resistance and hepatitis C virus chronic infection in HIV–hepatitis C virus–coinfected patients undergoing antiretroviral therapy. *J Acquir Immune Defic Syndr*. 2001;27:245–250.
156. Gavrila A, Tsiodras S, Doweiko J, et al. Exercise and vitamin E intake are independently associated with metabolic abnormalities in human immunodeficiency virus-positive subjects: a cross-sectional study. *Clin Infect Dis*. 2003;36:1593–1601.
157. Blumer RM, van Vonderen MG, Sutinen J, et al. Zidovudine/lamivudine contributes to insulin resistance within 3 months of starting combination antiretroviral therapy. *AIDS*. 2008;22:227–236.
158. Mallewa JE, Wilkins E, Vilar J, et al. HIV-associated lipodystrophy: a review of underlying mechanisms and therapeutic options. *J Antimicrob Chemother*. 2008;62:648–660.
159. Paruthi J, Gill N, Mantzoros CS. Adipokines in the HIV/HAART-associated lipodystrophy syndrome. *Metabolism*. 2013;62:1199–1205.
160. Lagathu C, Eustrace B, Prot M, et al. Some HIV antiretrovirals increase oxidative stress and alter chemokine, cytokine or adiponectin production in human adipocytes and macrophages. *Antivir Ther*. 2007;12:489–500.
161. Loonam CR, O'Dell SD, Sharp PA, et al. Microarray analysis reveals altered lipid and glucose metabolism genes in differentiated, ritonavir-treated 3T3-L1 adipocytes. *Curr HIV Res*. 2016;14:37–46.
162. El Hadri K, Glorian M, Monsempes C, et al. In vitro suppression of the lipogenic pathway by the nonnucleoside reverse transcriptase inhibitor efavirenz in 3T3 and human preadipocytes or adipocytes. *J Biol Chem*. 2004;279:15130–15141.
163. Haubrich RH, Kemper CA, Hellmann NS, et al. A randomized, prospective study of phenotype susceptibility testing versus standard of care to manage antiretroviral therapy: CCTG 575. *AIDS*. 2005;19:295–302.
164. Perez-Molina JA, Domingo P, Martinez E, et al. The role of efavirenz compared with protease inhibitors in the body fat changes associated with highly active antiretroviral therapy. *J Antimicrob Chemother*. 2008;62:234–245.
165. Haubrich RH, Riddler SA, DiRienzo AG, et al. Metabolic outcomes in a randomized trial of nucleoside, nonnucleoside and protease inhibitor-sparing regimens for initial HIV treatment. *AIDS*. 2009;23:1109–1118.
166. Blas-Garcia A, Apostolova N, Ballesteros D, et al. Inhibition of mitochondrial function by efavirenz increases lipid content in hepatic cells. *Hepatology*. 2010;52:115–125.
167. Shrivastav S, Kino T, Cunningham T, et al. Human immunodeficiency virus (HIV)-1 viral protein R suppresses transcriptional activity of peroxisome proliferator-activated receptor gamma and inhibits adipocyte differentiation: Implications for HIV-associated lipodystrophy. *Mol Endocrinol*. 2008;22:234–247.
168. Villarroya F, Domingo P, Giralt M. Lipodystrophy in HIV 1-infected patients: lessons for obesity research. *Int J Obes (Lond)*. 2007;31:1763–1776.
169. Ranade K, Geese WJ, Noor M, et al. Genetic analysis implicates resistin in HIV lipodystrophy. *AIDS*. 2008;22:1561–1568.
170. Castilhos JK, Sprinz E, Lazzaretti RK, et al. Polymorphisms in adiponectin receptor genes are associated with lipodystrophy-related phenotypes in HIV-infected patients receiving antiretroviral therapy. *HIV Med*. 2015;16:494–501.
171. Imamura S. Lipodystrophia centrifugalis abdominalis infantilis: statistical analysis of 168 cases. *Pediatr Dermatol*. 2012;29:437–441.
172. Polyzos SA, Perakakis N, Mantzoros CS. Fatty liver in lipodystrophy: a review with a focus on therapeutic perspectives of adiponectin and/or leptin replacement. *Metabolism*. 2019;96:66–82.
173. Ryden M, Arner P. Tumour necrosis factor-alpha in human adipose tissue—from signaling mechanisms to clinical implications. *J Intern Med*. 2007;262:431–438.
174. Mittra S, Bansal VS, Bhatnagar PK. From a glucocentric to a lipocentric approach towards metabolic syndrome. *Drug Discov Today*. 2008;13: 211–218.
175. Petersen KF, Shulman GI. Etiology of insulin resistance. *Am J Med*. 2006;119:S10–S16.
176. Bays HE, Gonzalez-Campoy JM, Bray GA, et al. Pathogenic potential of adipose tissue and metabolic consequences of adipocyte hypertrophy and increased visceral adiposity. *Expert Rev Cardiovasc Ther*. 2008;6:343–368.
177. Matsubara M, Maruoka S, Katayose S. Inverse relationship between plasma adiponectin and leptin concentrations in normal-weight and obese women. *Eur J Endocrinol*. 2002;147:173–180.
178. Weyer C, Funahashi T, Tanaka S, et al. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab*. 2001;86:1930–1935.
179. Tilg H, Moschen AR. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. *Nat Rev Immunol*. 2006;6:772–783.
180. Sheng T, Yang K. Adiponectin and its association with insulin resistance and type 2 diabetes. *J Genet Genomics*. 2008;35:321–326.
181. Coope A, Milanski M, Araujo EP, et al. AdipoR1 mediates the anorexigenic and insulin/leptin-like actions of adiponectin in the hypothalamus. *FEBS Lett*. 2008;582:1471–1476.
182. Mantzoros CS, Li T, Manson JE, et al. Circulating adiponectin levels are associated with better glycemic control, more favorable lipid profile, and reduced inflammation in women with type 2 diabetes. *J Clin Endocrinol Metab*. 2005;90:4542–4548.
183. Yamauchi T, Kamon J, Waki H, et al. Globular adiponectin protected ob/ob mice from diabetes and ApoE-deficient mice from atherosclerosis. *J Biol Chem*. 2003;278:2461–2468.
184. Hotta K, Funahashi T, Bodkin NL, et al. Circulating concentrations of the adipocyte protein adiponectin are decreased in parallel with reduced insulin sensitivity during the progression to type 2 diabetes in rhesus monkey. *Diabetes*. 2001;50:1126–1133.
185. Shimomura I, Hammer RE, Ikemoto S, et al. Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy. *Nature*. 1999;401:73–76.
186. Asilmaz E, Cohen P, Miyazaki M, et al. Site and mechanism of leptin action in a rodent form of congenital lipodystrophy. *J Clin Invest*. 2004;113:414–424.
187. Tong Q, Sankalé JL, Hadigan CM, et al. Regulation of adiponectin in human immunodeficiency virus-infected patients: relationship to body composition and metabolic indices. *J Clin Endocrinol Metab*. 2003;88:1559–1564.
188. Vigouroux C, Maachi M, Nguyen TH, et al. Serum adipocytokines are related to lipodystrophy and metabolic disorders in HIV-infected men under antiretroviral therapy. *AIDS*. 2003;17:1503–1511.
189. Antuna-Puente B, Boutet E, Vigouroux C, et al. Higher adiponectin levels in patients with Berardinelli-Seip congenital lipodystrophy due to seipin as compared with 1-acylglycerol-3-phosphate-o-acyltransferase-2 deficiency. *J Clin Endocrinol Metab*. 2010;95:1463–1468.
190. Sutinen J, Hakkinen AM, Westerbacka J, et al. Rosiglitazone in the treatment of HAART-associated lipodystrophy—a randomized double-blind placebo-controlled study. *Antivir Ther*. 2003;8:199–207.
191. Mynarcik DC, Combs T, McNurlan MA, et al. Adiponectin and leptin levels in HIV-infected subjects with insulin resistance and body fat redistribution. *J Acquir Immune Defic Syndr*. 2002;31:514–520.
192. Ceddia RB. Direct metabolic regulation in skeletal muscle and fat tissue by leptin: implications for glucose and fatty acids homeostasis. *Int J Obes (Lond)*. 2005;29:1175–1183.
193. Minokoshi Y, Kahn BB. Role of AMP-activated protein kinase in leptin-induced fatty acid oxidation in muscle. *Biochem Soc Trans*. 2003;31: 196–201.
194. Lee JH, Chan JL, Sourlas E, et al. Recombinant methionyl human leptin therapy in replacement doses improves insulin resistance and metabolic profile in patients with lipoatrophy and metabolic syndrome induced by the highly active antiretroviral therapy. *J Clin Endocrinol Metab*. 2006;91:2605–2611.
195. Mulligan K, Khatami H, Schwarz JM, et al. The effects of recombinant human leptin on visceral fat, dyslipidemia, and insulin resistance in patients with human immunodeficiency virus-associated lipoatrophy and hypoleptinemia. *J Clin Endocrinol Metab*. 2009;94:1137–1144.
196. Mantzoros CS. Whither recombinant human leptin treatment for HIV-associated lipoatrophy and the metabolic syndrome? *J Clin Endocrinol Metab*. 2009;94:1089–1091.

197. Ajluni N, Dar M, Xu J, et al. Efficacy and safety of metreleptin in patients with partial lipodystrophy: lessons from an expanded access program. *J Diabetes Metab*. 2016;7:659.
198. Guest CB, Park MJ, Johnson DR, et al. The implication of proinflammatory cytokines in type 2 diabetes. *Front Biosci*. 2008;13:5187–5194.
199. Shoelson SE, Herrero L, Naaz A. Obesity, inflammation, and insulin resistance. *Gastroenterology*. 2007;132:2169–2180.
200. Sevastianova K, Sutinen J, Kannisto K, et al. Adipose tissue inflammation and liver fat in patients with highly active antiretroviral therapy-associated lipodystrophy. *Am J Physiol Endocrinol Metab*. 2008;295:E85–E91.
201. He G, Andersen O, Haugaard SB, et al. Plasminogen activator inhibitor type 1 (PAI-1) in plasma and adipose tissue in HIV-associated lipodystrophy syndrome. Implications of adipokines. *Eur J Clin Invest*. 2005;35:583–590.
202. Cawthorn WP, Heyd F, Hegyi K, et al. Tumour necrosis factor-alpha inhibits adipogenesis via a beta-catenin/TCF4 (TCF7L2)-dependent pathway. *Cell Death Differ*. 2007;14:1361–1373.
203. Cawthorn WP, Sethi JK. TNF-alpha and adipocyte biology. *FEBS Lett*. 2008;582:117–131.
204. Glund S, Krook A. Role of interleukin-6 signalling in glucose and lipid metabolism. *Acta Physiol (Oxf)*. 2008;192:37–48.
205. Rhodes CJ. Type 2 diabetes: a matter of beta-cell life and death? *Science*. 2005;307:380–384.
206. Donath MY, Ehses JA, Maedler K, et al. Mechanisms of beta-cell death in type 2 diabetes. *Diabetes*. 2005;54:S108–S113.
207. White M, Lepage S, Lavoie J, et al. Effects of combined candesartan and ACE inhibitors on BNP, markers of inflammation and oxidative stress, and glucose regulation in patients with symptomatic heart failure. *J Card Fail*. 2007;13:86–94.
208. Ito D, Suzuki N. Seipinopathy: a novel endoplasmic reticulum stress-associated disease. *Brain*. 2009;132:8–15.
209. Samaras K. Metabolic consequences and therapeutic options in highly active antiretroviral therapy in human immunodeficiency virus-1 infection. *J Antimicrob Chemother*. 2008;61:238–245.
210. Noor MA, Seneviratne T, Aweeka FT, et al. Indinavir acutely inhibits insulin-stimulated glucose disposal in humans: a randomized, placebo-controlled study. *AIDS*. 2002;16:F1–F8.
211. Pagac M, Cooper DE, Qi Y, et al. SEIPIN regulates lipid droplet expansion and adipocyte development by modulating the activity of glycerol-3-phosphate acyltransferase. *Cell Rep*. 2016;17:1546–1559.
212. Gao M, Liu L, Wang X, et al. GPAT3 deficiency alleviates insulin resistance and hepatic steatosis in a mouse model of severe congenital generalized lipodystrophy. *Hum Mol Genet*. 2020;29:432–443.
213. Papendieck L, Araujo MB. Clinical outcome in a series of pediatric patients with congenital generalized lipodystrophies treated with dietary therapy. *J Pediatr Endocrinol Metab*. 2018;31:77–83.
214. Shah M, Tierney K, Adams-Huet B, et al. The role of diet, exercise and smoking in dyslipidaemia in HIV-infected patients with lipodystrophy. *HIV Med*. 2005;6:291–298.
215. Dong KR, Hendricks KM. The role of nutrition in fat deposition and fat atrophy in patients with HIV. *Nutr Clin Care*. 2005;8:31–36.
216. Tsiodras S, Koulia KA, Yannakoulia M, et al. Adherence to Mediterranean diet is favorably associated with metabolic parameters in HIV-positive patients with the highly active antiretroviral therapy-induced metabolic syndrome and lipodystrophy. *Metabolism*. 2009;58:854–859.
217. Terry L, Sprinz E, Stein R, et al. Exercise training in HIV-1-infected individuals with dyslipidemia and lipodystrophy. *Med Sci Sports Exerc*. 2006;38:411–417.
218. Dolan SE, Frontera W, Librizzi J, et al. Effects of a supervised home-based aerobic and progressive resistance training regimen in women infected with human immunodeficiency virus: a randomized trial. *Arch Intern Med*. 2006;166:1225–1231.
219. Lindegaard B, Hansen T, Hvid T, et al. The effect of strength and endurance training on insulin sensitivity and fat distribution in human immunodeficiency virus-infected patients with lipodystrophy. *J Clin Endocrinol Metab*. 2008;93:3860–3869.
220. Brown TT. Approach to the human immunodeficiency virus-infected patient with lipodystrophy. *J Clin Endocrinol Metab*. 2008;93:2937–2945.
221. Aboud M, Elgalib A, Kulasegaram R, et al. Insulin resistance and HIV infection: a review. *Int J Clin Pract*. 2007;61:463–472.
222. Kohli R, Shevitz A, Gorbach S, et al. A randomized placebo-controlled trial of metformin for the treatment of HIV lipodystrophy. *HIV Med*. 2007;8:420–426.
223. Mulligan K, Yang Y, Wininger DA, et al. Effects of metformin and rosiglitazone in HIV-infected patients with hyperinsulinemia and elevated waist/hip ratio. *AIDS*. 2007;21:47–57.
224. Driscoll SD, Meininger GE, Ljungquist K, et al. Differential effects of metformin and exercise on muscle adiposity and metabolic indices in human immunodeficiency virus-infected patients. *J Clin Endocrinol Metab*. 2004;89:2171–2178.
225. Arioglu E, Duncan-Morin J, Sebring N, et al. Efficacy and safety of troglitazone in the treatment of lipodystrophy syndromes. *Ann Intern Med*. 2000;133:263–274.
226. Hadigan C, Yawetz S, Thomas A, et al. Metabolic effects of rosiglitazone in HIV lipodystrophy: a randomized, controlled trial. *Ann Intern Med*. 2004;140:786–794.
227. Macallan DC, Baldwin C, Mandalia S, et al. Treatment of altered body composition in HIV-associated lipodystrophy: comparison of rosiglitazone, pravastatin, and recombinant human growth hormone. *HIV Clin Trials*. 2008;9:254–268.
228. Slama L, Lanoy E, Valantin MA, et al. Effect of pioglitazone on HIV-1-related lipodystrophy: a randomized double-blind placebo-controlled trial (ANRS 113). *Antivir Ther*. 2008;13:67–76.
229. van Wijk JP, de Koning EJ, Cabezas MC, et al. Comparison of rosiglitazone and metformin for treating HIV lipodystrophy: a randomized trial. *Ann Intern Med*. 2005;143:337–346.
230. Carr A, Workman C, Carey D, et al. No effect of rosiglitazone for treatment of HIV-1 lipoatrophy: randomised, double-blind, placebo-controlled trial. *Lancet*. 2004;363:429–438.
231. Cavalcanti RB, Raboud J, Shen S, et al. A randomized, placebo-controlled trial of rosiglitazone for HIV-related lipoatrophy. *J Infect Dis*. 2007;195:1754–1761.
232. Gavrila A, Hsu W, Tsiodras S, et al. Improvement in highly active antiretroviral therapy-induced metabolic syndrome by treatment with pioglitazone but not with fenofibrate: a 2 x 2 factorial, randomized, double-blinded, placebo-controlled trial. *Clin Infect Dis*. 2005;40: 745–749.
233. Magkos F, Brennan A, Sweeney L, et al. Leptin replacement improves postprandial glycemia and insulin sensitivity in human immunodeficiency virus-infected lipoatrophic men treated with pioglitazone: a pilot study. *Metabolism*. 2011;60:1045–1049.
234. Kovacic JC, Martin A, Carey D, et al. Influence of rosiglitazone on flow-mediated dilation and other markers of cardiovascular risk in HIV-infected patients with lipoatrophy. *Antivir Ther*. 2005;10:135–143.
235. Akinci B, Meral R, Oral EA. Update on therapeutic options in lipodystrophy. *Curr Diab Rep*. 2018;18:139.
236. Bolli GB, Owens DR. Insulin glargine. *Lancet*. 2000;356:443–445.
237. American Diabetes Association. 10. Cardiovascular disease and risk management: standards of medical care in diabetes-2021 [published correction appears in Diabetes Care. 2021 Sep;44(9):2183-2185]. *Diabetes Care*. 2021;44:S125–S150.
238. Pasternak B, Wintzell V, Eliasson B, et al. Use of glucagon-like peptide 1 receptor agonists and risk of serious renal events: Scandinavian cohort study. *Diabetes Care*. 2020;43:1326–1335.
239. Lager CJ, Esfandiari NH, Subauste AR, et al. Roux-en-Y gastric bypass vs. sleeve gastrectomy: balancing the risks of surgery with the benefits of weight loss. *Obes Surg*. 2017;27:154–161.
240. Chondronikola M, Harris LL, Klein S. Bariatric surgery and type 2 diabetes: are there weight loss-independent therapeutic effects of upper gastrointestinal bypass? *J Intern Med*. 2016;280:476–486.
241. Ciudin A, Baena-Fustegueras JA, Fort JM, et al. Successful treatment for the Dunnigan-type familial partial lipodystrophy with Roux-en-Y gastric bypass. *Clin Endocrinol*. 2011;75:403–404.
242. Melvin A, Adams C, Flanagan C, et al. Roux-en-Y gastric bypass surgery in the management of familial partial lipodystrophy type 1. *J Clin Endocrinol Metab*. 2017;102:3616–3620.

243. Grundfest-Broniatowski S, Yan J, Kroh M, et al. Successful treatment of an unusual case of FPLD2: the role of Roux-en-Y gastric bypass-case report and literature review. *J Gastrointest Surg*. 2017;21:739–743.
244. Johns KW, Bennett MT, Bondy G. Are HIV positive patients resistant to statin therapy? *Can J Cardiol*. 2007;23:49C–50C.
245. Mallon PW, Miller J, Kovacic JC, et al. Effect of pravastatin on body composition and markers of cardiovascular disease in HIV-infected men—a randomized, placebo-controlled study. *AIDS*. 2006;20:1003–1010.
246. Aberg JA, Rosenkranz SL, Fichtenbaum CJ, et al. Pharmacokinetic interaction between nelfinavir and pravastatin in HIV-seronegative volunteers: ACTG Study A5108. *AIDS*. 2006;20:725–729.
247. Fichtenbaum CJ, Gerber JG. Interactions between antiretroviral drugs and drugs used for the therapy of the metabolic complications encountered during HIV infection. *Clin Pharmacokinet*. 2002;41:1195–1211.
248. Gerber JG, Rosenkranz SL, Fichtenbaum CJ, et al. Effect of efavirenz on the pharmacokinetics of simvastatin, atorvastatin, and pravastatin: results of AIDS Clinical Trials Group 5108 Study. *J Acquir Immune Defic Syndr*. 2005;39:307–312.
249. Wohl DA, Tien HC, Busby M, et al. Randomized study of the safety and efficacy of fish oil (omega-3 fatty acid) supplementation with dietary and exercise counseling for the treatment of antiretroviral therapy-associated hypertriglyceridemia. *Clin Infect Dis*. 2005;41:1498–1504.
250. Andreev VP, Paz-Filho G, Wong ML, et al. Deconvolution of insulin secretion, insulin hepatic extraction post-hepatic delivery rates and sensitivity during 24-hour standardized meals: time course of glucose homeostasis in leptin replacement treatment. *Horm Metab Res*. 2009;41:142–151.
251. Shah AS, Wilson DP. Primary hypertriglyceridemia in children and adolescents. *J Clin Lipidol*. 2015;9:S20–S28.
252. Manlhiot C, Larsson P, Gurofsky RC, et al. Spectrum and management of hypertriglyceridemia among children in clinical practice. *Pediatrics*. 2009;123:458–465.
253. Gupta N, Asi N, Farah W, et al. Clinical features and management of non-HIV-related lipodystrophy in children: a systematic review. *J Clin Endocrinol Metab*. 2017;102:363–374.
254. Hadigan C, Liebau J, Torriani M, et al. Improved triglycerides and insulin sensitivity with 3 months of acipimox in human immunodeficiency virus-infected patients with hypertriglyceridemia. *J Clin Endocrinol Metab*. 2006;91:4438–4444.
255. HPS2-THRIVE Collaborative Group, Landray MJ, Haynes R, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med*. 2014;371:203–212.
256. AIM-HIGH Investigators, Boden WE, Probstfield JL, et al. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. *N Engl J Med*. 2011;365:2255–2267.
257. ACCORD Study Group, Ginsberg HN, Elam MB, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med*. 2010;362:1563–1574.
258. Food and Drug Administration. Withdrawal of approval of indications related to the coadministration with statins in applications for niacin extended-release tablets and fenofibric acid delayed-release capsules. *Federal Register*. April 18, 2016. https://s3.amazonaws.com/public-inspection.federalregister.gov/2016-08887.pdf.
259. Kastelein JJ, Akdim F, Stroes ES, et al. Simvastatin with or without ezetimibe in familial hypercholesterolemia. *N Engl J Med*. 2008;358:1431–1443.
260. Hegele RA, Tsimikas S. Lipid-lowering agents [published correction appears in Circ Res. 2020 Oct 23;127(10):e251]. *Circ Res*. 2019;124:386–404.
261. Gouni-Berthold I, Alexander V, Digenio A, et al. Apolipoprotein C-III inhibition with volanesorsen in patients with hypertriglyceridemia (COMPASS): a randomized, double-blind, placebo-controlled trial. *Atheroscler Suppl*. 2017;28:E1–E2.
262. Gaudet D, Digenio A, Alexander V, et al. The APPROACH study: a randomized, double-blind, placebocontrolled, phase 3 study of volanesorsen administered subcutaneously to patients with familial chylomicronemia syndrome (Fcs). *Clin Cardiol*. 2017;40:14.
263. Paik J, Duggan S. Volanesorsen: first global approval. *Drugs*. 2019;79:1349–1354.
264. Robciuc MR, Maranghi M, Lahikainen A, et al. Angptl3 deficiency is associated with increased insulin sensitivity, lipoprotein lipase activity, and decreased serum free fatty acids. *Arterioscler Thromb Vasc Biol*. 2013;33:1706–1713.
265. Ruscica M, Zimetti F, Adorni MP, et al. Pharmacological aspects of ANGPTL3 and ANGPTL4 inhibitors: new therapeutic approaches for the treatment of atherogenic dyslipidemia. *Pharmacol Res*. 2020;153:104653.
266. Kersten S. Angiopoietin-like 3 in lipoprotein metabolism. *Nat Rev Endocrinol*. 2017;13:731–739.
267. Geladari E, Tsamadia P, Vallianou NG. ANGPTL3 inhibitors - their role in cardiovascular disease through regulation of lipid metabolism. *Circ J*. 2019;83:267–273.
268. Ahmad Z, Banerjee P, Hamon S, et al. Inhibition of angiopoietin-like protein 3 with a monoclonal antibody reduces triglycerides in hypertriglyceridemia. *Circulation*. 2019;140:470–486.
269. Graham MJ, Lee RG, Brandt TA, et al. Cardiovascular and metabolic effects of angptl3 antisense oligonucleotides. *N Engl J Med*. 2017;377:222–232.
270. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: executive summary: a Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines [published correction appears in J Am Coll Cardiol. 2019 Jun 25;73(24):3234–3237]. *J Am Coll Cardiol*. 2019;73:3168–3209.
271. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J*. 2020;41:111–188.
272. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med*. 2017;376:1713–1722.
273. Schwartz GG, Steg PG, Szarek M, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *N Engl J Med*. 2018;379:2097–2107.
274. Fitzgerald K, White S, Borodovsky A, et al. A highly durable RNAi therapeutic inhibitor of PCSK9. *N Engl J Med*. 2017;376:41–51.
275. Ray KK, Wright RS, Kallend D, et al. Two phase 3 trials of inclisiran in patients with elevated LDL cholesterol. *N Engl J Med*. 2020;382:1507–1519.
276. Pinkosky SL, Newton RS, Day EA, et al. Liver-specific ATPcitrate lyase inhibition by bempedoic acid decreases LDL-C and attenuates atherosclerosis. *Nat Commun*. 2016;7:13457.
277. Ray KK, Bays HE, Catapano AL, et al. Safety and efficacy of bempedoic acid to reduce LDL cholesterol. *N Engl J Med*. 2019;380:1022–1032.
278. Laufs U, Banach M, Mancini GBJ, et al. Efficacy and safety of bempedoic acid in patients with hypercholesterolemia and statin intolerance. *J Am Heart Assoc*. 2019;8:e011662.
279. Banach M, Duell PB, Gotto AM, et al. Association of bempedoic acid administration with atherogenic lipid levels in phase 3 randomized clinical trials of patients with hypercholesterolemia. *JAMA Cardiol*. 2020;5:1124–1135.
280. Di Minno A, Lupoli R, Calcaterra I, et al. Efficacy and safety of bempedoic acid in patients with hypercholesterolemia: systematic review and meta-analysis of randomized controlled trials. *J Am Heart Assoc*. 2020;9:e016262.
281. Markham A. Bempedoic acid: first approval. *Drugs*. 2020;80:747–753.
282. Bolan C, Oral EA, Gorden P, et al. Intensive, long-term plasma exchange therapy for severe hypertriglyceridemia in acquired generalized lipoatrophy. *J Clin Endocrinol Metab*. 2002;87:380–384.
283. Grupp C, Beckermann J, Köster E, et al. Relapsing and progressive complications of severe hypertriglyceridemia: effective long-term treatment with double filtration plasmapheresis. *Blood Purif*. 2020;49:457–467.
284. Rietschel P, Hadigan C, Corcoran C, et al. Assessment of growth hormone dynamics in human immunodeficiency virus-related lipodystrophy. *J Clin Endocrinol Metab*. 2001;86:504–510.
285. Stanley TL, Grinspoon SK. GH/GHRH axis in HIV lipodystrophy. *Pituitary*. 2009;12:143–152.
286. Koutkia P, Canavan B, Breu J, et al. Growth hormone (GH) responses to GH-releasing hormone-arginine testing in human immunodeficiency virus lipodystrophy. *J Clin Endocrinol Metab*. 2005;90:32–38.

287. Bickel M, Zangos S, Jacobi V, et al. A randomized, open-label study to compare the effects of two different doses of recombinant human growth hormone on fat reduction and fasting metabolic parameters in HIV-1-infected patients with lipodystrophy. *HIV Med*. 2006;7:397–403.
288. Grunfeld C, Thompson M, Brown SJ, et al. Recombinant human growth hormone to treat HIV-associated adipose redistribution syndrome: 12 week induction and 24-week maintenance therapy. *J Acquir Immune Defic Syndr*. 2007;45:286–297.
289. He Q, Engelson ES, Albu JB, et al. Preferential loss of omental-mesenteric fat during growth hormone therapy of HIV-associated lipodystrophy. *J Appl Physiol*. 2003;94:2051–2057.
290. Engelson ES, Glesby MJ, Mendez D, et al. Effect of recombinant human growth hormone in the treatment of visceral fat accumulation in HIV infection. *J Acquir Immune Defic Syndr*. 2002;30:379–391.
291. Honda M, Yogi A, Ishizuka N, et al. Effectiveness of subcutaneous growth hormone in HIV-1 patients with moderate to severe facial lipoatrophy. *Intern Med*. 2007;46:357–360.
292. Andersen O, Haugaard SB, Flyvbjerg A, et al. Low-dose growth hormone and human immunodeficiency virus-associated lipodystrophy syndrome: a pilot study. *Eur J Clin Invest*. 2004;34:561–568.
293. Lo J, You SM, Canavan B, et al. Low-dose physiological growth hormone in patients with HIV and abdominal fat accumulation: a randomized controlled trial. *JAMA*. 2008;300:509–519.
294. Lo JC, Mulligan K, Noor MA, et al. The effects of low-dose growth hormone in HIV-infected men with fat accumulation: a pilot study. *Clin Infect Dis*. 2004;39:732–735.
295. Falutz J, Potvin D, Mamputu JC, et al. Effects of tesamorelin, a growth hormone–releasing factor, in HIV-infected patients with abdominal fat accumulation: a randomized placebo-controlled trial with a safety extension. *J Acquir Immune Defic Syndr*. 2010;53:311–322.
296. Luzi L, Meneghini E, Oggionni S, et al. GH treatment reduces trunkal adiposity in HIV-infected patients with lipodystrophy: a randomized placebo-controlled study. *Eur J Endocrinol*. 2005;153:781–789.
297. Falutz J, Allas S, Blot K, et al. Metabolic effects of a growth hormone-releasing factor in patients with HIV. *N Engl J Med*. 2007;357:2359–2370.
298. Falutz J, Allas S, Mamputu JC, et al. Long-term safety and effects of tesamorelin, a growth hormone-releasing factor analogue, in HIV patients with abdominal fat accumulation. *AIDS*. 2008;22:1719–1728.
299. Lundgren JD, Battegay M, Behrens G, et al. European AIDS Clinical Society (EACS) guidelines on the prevention and management of metabolic diseases in HIV. *HIV Med*. 2008;9:72–81.
300. Dube MP, Stein JH, Aberg JA, et al. Guidelines for the evaluation and management of dyslipidemia in human immunodeficiency virus (HIV)-infected adults receiving antiretroviral therapy: recommendations of the HIV Medical Association of the Infectious Disease Society of America and the Adult AIDS Clinical Trials Group. *Clin Infect Dis*. 2003;37: 613–627.
301. Carey D, Liew S, Emery S. Restorative interventions for HIV facial lipoatrophy. *AIDS Rev*. 2008;10:116–124.
302. Vallejo A, Garcia-Ruano AA, Pinilla C, et al. Comparing efficacy and costs of four facial fillers in human immunodeficiency virus-associated lipodystrophy: a clinical trial. *Plast Reconstr Surg*. 2018;141:613–623.
303. Jones DH, Carruthers A, Brody HJ, et al. Ten-year and beyond follow-up after treatment with highly purified liquid-injectable silicone for HIV-associated facial lipoatrophy: a report of 164 patients. *Dermatol Surg*. 2019;45:941–948.
304. Moon HS, Dalamaga M, Kim SY, et al. Leptin's role in lipodystrophic and nonlipodystrophic insulin-resistant and diabetic individuals. *Endocr Rev*. 2013;34:377–412.
305. Park S, Hong SM, Sung SR, et al. Long-term effects of central leptin and resistin on body weight, insulin resistance, and beta-cell function and mass by the modulation of hypothalamic leptin and insulin signaling. *Endocrinology*. 2008;149:445–454.
306. Farooqi IS, Bullmore E, Keogh J, et al. Leptin regulates striatal regions and human eating behavior. *Science*. 2007;317:1355.
307. Rosenbaum M, Sy M, Pavlovich K, et al. Leptin reverses weight-loss-induced changes in regional neural activity responses to visual food stimuli. *J Clin Invest*. 2008;118:2583–2591.
308. Farooqi IS, Matarese G, Lord GM, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*. 2002;110:1093–1103.
309. Gallardo N, Bonzon-Kulichenko E, Fernandez-Agullo T, et al. Tissue-specific effects of central leptin on the expression of genes involved in lipid metabolism in liver and white adipose tissue. *Endocrinology*. 2007;148:5604–5610.
310. Prieur X, Tung YC, Griffin JL, et al. Leptin regulates peripheral lipid metabolism primarily through central effects on food intake. *Endocrinology*. 2008;149:5432–5439.
311. Maedler K, Schulthess FT, Bielman C, et al. Glucose and leptin induce apoptosis in human beta-cells and impair glucose-stimulated insulin secretion through activation of c-Jun N-terminal kinases. *FASEB J*. 2008;22:1905–1913.
312. Beltrand J, Beregszaszi M, Chevenne D, et al. Metabolic correction induced by leptin replacement treatment in young children with Berardinelli-Seip congenital lipoatrophy. *Pediatrics*. 2007;120:e291–e296.
313. McDuffie JR, Riggs PA, Calis KA, et al. Effects of exogenous leptin on satiety and satiation in patients with lipodystrophy and leptin insufficiency. *J Clin Endocrinol Metab*. 2004;89:4258–4263.
314. Javor ED, Cochran EK, Musso C, et al. Long-term efficacy of leptin replacement in patients with generalized lipodystrophy. *Diabetes*. 2005;54:1994–2002.
315. Moran SA, Patten N, Young JR, et al. Changes in body composition in patients with severe lipodystrophy after leptin replacement therapy. *Metabolism*. 2004;53:513–519.
316. Oral EA, Simha V, Ruiz E, et al. Leptin-replacement therapy for lipodystrophy. *N Engl J Med*. 2002;346:570–578.
317. Chong AY, Lupsa BC, Cochran EK, et al. Efficacy of leptin therapy in the different forms of human lipodystrophy. *Diabetologia*. 2010;53:27–35.
318. Brown RJ, Meehan CA, Cochran E, et al. Effects of metreleptin in pediatric patients with lipodystrophy. *J Clin Endocrinol Metab*. 2017;102: 1511–1519.
319. Chan JL, Lutz K, Cochran E, et al. Clinical effects of long-term metreleptin treatment in patients with lipodystrophy. *Endocr Pract*. 2011;17:922–932.
320. Meehan CA, Cochran E, Kassai A, et al. Metreleptin for injection to treat the complications of leptin deficiency in patients with congenital or acquired generalized lipodystrophy. *Expert Rev Clin Pharmacol*. 2016;9:59–68.
321. Ebihara K, Kusakabe T, Hirata M, et al. Efficacy and safety of leptin-replacement therapy and possible mechanisms of leptin actions in patients with generalized lipodystrophy. *J Clin Endocrinol Metab*. 2007;92:532–541.
322. Musso C, Cochran E, Javor E, et al. The long-term effect of recombinant methionyl human leptin therapy on hyperandrogenism and menstrual function in female and pituitary function in male and female hypoleptinemic lipodystrophic patients. *Metabolism*. 2005;54:255–263.
323. Javor ED, Ghany MG, Cochran EK, et al. Leptin reverses nonalcoholic steatohepatitis in patients with severe lipodystrophy. *Hepatology*. 2005;41:753–760.
324. Simha V, Szczepaniak LS, Wagner AJ, et al. Effect of leptin replacement on intrahepatic and intramyocellular lipid content in patients with generalized lipodystrophy. *Diabetes Care*. 2003;26:30–35.
325. Guettier JM, Park JY, Cochran EK, et al. Leptin therapy for partial lipodystrophy linked to a PPAR-gamma mutation. *Clin Endocrinol (Oxf)*. 2008;68:547–554.
326. Park JY, Javor ED, Cochran EK, et al. Long-term efficacy of leptin replacement in patients with Dunnigan-type familial partial lipodystrophy. *Metabolism*. 2007;56:508–516.
327. Chou K, Perry CM. Metreleptin: first global approval. *Drugs*. 2013;73:989–997.
328. Javor ED, Moran SA, Young JR, et al. Proteinuric nephropathy in acquired and congenital generalized lipodystrophy: baseline characteristics and course during recombinant leptin therapy. *J Clin Endocrinol Metab*. 2004;89:3199–3207.

329. Oral EA, Ruiz E, Andewelt A, et al. Effect of leptin replacement on pituitary hormone regulation in patients with severe lipodystrophy. *J Clin Endocrinol Metab*. 2002;87:3110–3117.
330. Simha V, Zerwekh JE, Sakhaee K, et al. Effect of subcutaneous leptin replacement therapy on bone metabolism in patients with generalized lipodystrophy. *J Clin Endocrinol Metab*. 2002;87:4942–4945.
331. Myalept safely and effectively. https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/125390s004lbl.pdf
332. Drug label information. https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=c986f93b-855d-4ef0-b620-5d41a0513e48#S1
333. Chan JL, Koda J, Heilig JS, et al. Immunogenicity associated with metreleptin treatment in patients with obesity or lipodystrophy. *Clin Endocrinol (Oxf)*. 2016;85:137–149.
334. Beltrand J, Lahlou N, Le Charpentier T, et al. Resistance to leptinreplacement therapy in Berardinelli-Seip congenital lipodystrophy: an immunological origin. *Eur J Endocrinol*. 2010;162:1083–1091.
335. Duntas LH, Popovic V, Panotopoulos G. Adiponectin: novelties in metabolism and hormonal regulation. *Nutr Neurosci*. 2004;7:195–200.
336. Xu A, Yin S, Wong L, et al. Adiponectin ameliorates dyslipidemia induced by the human immunodeficiency virus protease inhibitor ritonavir in mice. *Endocrinology*. 2004;145:487–494.

25

Lipoprotein Metabolism and the Treatment of Lipid Disorders

Mason W. Freeman, Geoffrey Walford, and Janet Lo

OUTLINE

Lipoprotein Metabolism, 404
  Lipoproteins, 404
  Metabolism of Lipids and Lipoproteins, 408
Disorders of Lipid Metabolism in Patients with Diabetes, 409
  Type 1 Diabetes, 409
  Type 2 Diabetes, 410
Genetic Basis of Lipid Disorders, 410
  Monogenic Low-Density Lipoprotein Disorders, 410
  Monogenic High-Density Lipoprotein Disorders, 412
  Polygenic Lipid Disorders, 413
Diagnosis of Lipid Disorders, 413
Management/Treatment, 413
  Assessing the Need to Treat Patients With Hyperlipidemia, 414
  Dietary and Drug Treatment of Lipid Disorders, 415
  Outcome Studies of Pharmacologic Therapy for Hyperlipidemia, 416

Atherosclerosis is a chronic inflammatory disease of large- and medium-sized arteries that results from the deposition of lipids and lipoproteins in the intima of these vessels.[1,2] This process is a major cause of morbidity and mortality in the developed world, and the toll is particularly high in the diabetic population, accounting for approximately two thirds to three quarters of all diabetic deaths.[3] When combined with significant morbidity arising from ischemic injury and loss of limb perfusion, the impact of atherosclerosis on human health is enormous.[4] Improvements in the care of patients with and without diabetes, including the use of intensive lipid-lowering therapy, has led to significant reductions in mortality from cardiovascular (CV) disease in recent years.[5] This chapter will review basic principles of lipoprotein metabolism, focusing on the impact diabetes has on these lipid pathways. Data that address the value of lipid-altering treatments in reducing the burden of atherosclerosis in patients with and without diabetes will be presented. Finally, an approach to the management of patients with dyslipidemia will be outlined.

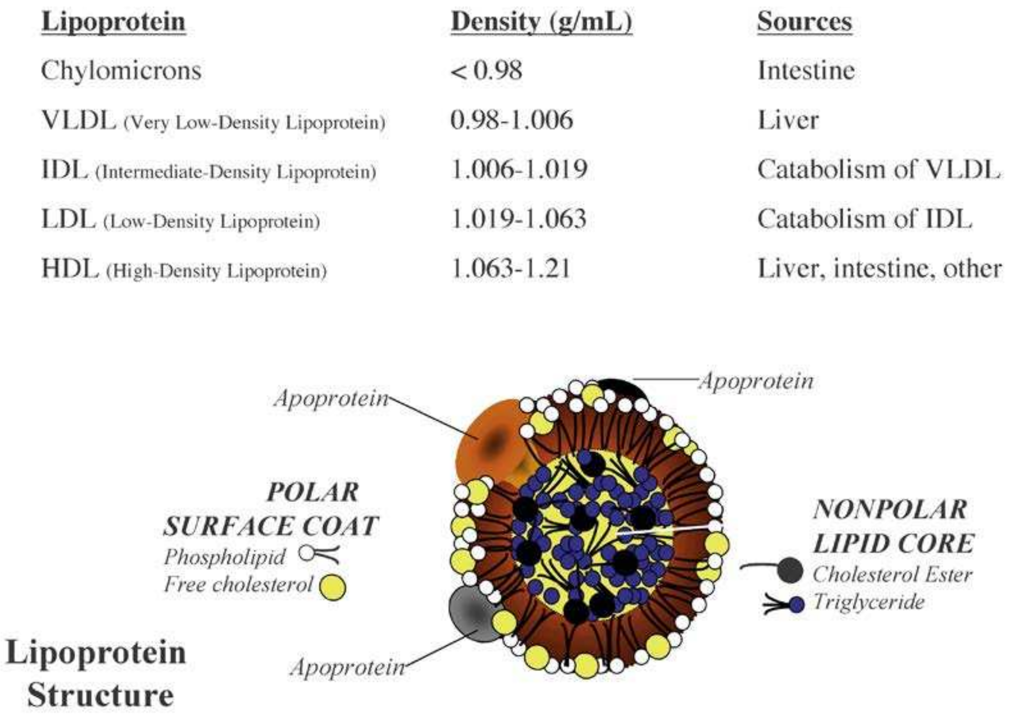
LIPOPROTEIN METABOLISM

Lipoproteins

The clinically reported serum triglyceride and cholesterol values represent the sum of the cholesterol and triglyceride contents of the circulating lipoproteins present in a volume of blood (expressed either as mg/dL or mmol/L). While knowledge of these total lipid values is very helpful in assessing CV risk, it is the distribution of cholesterol among the different lipoprotein fractions that is critical to any sophisticated understanding of atherosclerosis risk. Lipoproteins are structured to solve the problem of transporting highly hydrophobic lipids, such as cholesterol ester and triglyceride, in the aqueous environment of the blood. Because the bulk of lipoproteins are synthesized in the liver or gut, they must be moved via the bloodstream from these sites of synthesis to tissues where lipids will be taken up and utilized. By complexing neutral lipids with both proteins and more polar lipids such as unesterified cholesterol and phospholipids, the lipoprotein structure accomplishes this task. The more hydrophobic lipids are sequestered on the inside of the spherical lipoprotein, whereas more polar lipids and proteins decorate the outer surface (Fig. 25.1). In addition to solving the biophysical transport problem, the lipoprotein structure also enables particles to utilize a class of proteins known as *apoproteins* (Table 52.1) that serve as targeting molecules which interact with lipid-modifying enzymes (Table 25.2) or receptors and transporters (Table 25.3) critical for the normal functioning of the disparate lipoprotein classes. Complete loss of function mutations in the genes encoding apoproteins or the enzymes and receptors with which apoproteins interact (see Tables 25.1 to 25.3) are uncommon, but they can have a very significant influence on the circulating levels of lipoproteins in individuals with and without diabetes. As lipids play fundamental roles in a host of other cellular processes, mutations in lipid enzymes, receptors, transporters, and apoproteins can have profound effects on human biology beyond serum lipoproteins and CV disease, some of which are cited in Tables 25.1 to 25.3.

The nomenclature for most lipoprotein particles derives from the method by which they were originally identified and separated, that is, density gradient ultracentrifugation. For a brief period after the introduction of gel electrophoresis, lipoproteins were also classified by their migration properties on these gels, giving rise to alternative names such as *alpha* and *beta lipoproteins.* This period coincided with the identification of many of the clinical syndromes caused by abnormalities of lipid metabolism, so the names for some of these disorders still carry the signature of the electrophoresis era (e.g., *hypoalphalipoproteinemia* indicates low levels of the alpha-migrating or high-density lipoprotein [HDL]). Gel electrophoresis is rarely used in clinical laboratories today for standard lipoprotein measurements, though it may be used in specialty lipid labs to further subdivide classes of lipoproteins. As a consequence, the current medical literature has largely returned to the use of the density nomenclature to describe and characterize disorders of lipoprotein metabolism.

Although lipoproteins constitute a continuous spectrum of density within the plasma, they are commonly divided into five major classes (see Fig. 25.1). These are: (1) chylomicrons; (2) very low-density lipoproteins (VLDLs); (3) intermediate-density lipoproteins (IDLs); (4)



Lipoprotein	Density (g/mL)	Sources
Chylomicrons	< 0.98	Intestine
VLDL (Very Low-Density Lipoprotein)	0.98-1.006	Liver
IDL (Intermediate-Density Lipoprotein)	1.006-1.019	Catabolism of VLDL
LDL (Low-Density Lipoprotein)	1.019-1.063	Catabolism of IDL
HDL (High-Density Lipoprotein)	1.063-1.21	Liver, intestine, other

Fig. 25.1 The *upper panel* of the figure lists the density range that encompasses each lipoprotein class and the major tissue in which the lipoprotein is synthesized. The *bottom panel* is a graphic representation of a generic lipoprotein, emphasizing the spherical shape of the particle, the use of amphipathic apoproteins and polar lipids on the particle surface that interact with cells and the blood environment, and the sequestration of nonpolar lipids in the core of the lipoprotein. All lipoproteins share these general features, but they differ in size, lipid composition, and in the specific apoproteins embedded.

TABLE 25.1 Apolipoproteins of the Serum Lipoproteins

Apoprotein Ref	Major Lipoprotein	Functional Role (If Known)	Human Disease Association
AI[239]	HDL	LCAT activation	Cataracts, low HDL levels, CHD
AII[240]	HDL	Structural protein, ? function	No defect identified in a deficient subject
AIV[241]	HDL	Fat absorption, LCAT activation	? Altered plasma lipid and glucose levels
AV[242]	HDL, VLDL	TG metabolism	Hypertriglyceridemia
B_{100}[243]	LDL	VLDL and LDL synthesis	Hypobetalipoproteinemia (low LDL levels), familial defective apo B (high LDL levels)
B_{48}[243]	Chylomicron, chylo remnant	Intestinal fat absorption	Hypobetalipoproteinemia (low LDL levels)
CI[244]	HDL, VLDL, chylomicron	TG-rich lipoprotein uptake	Unknown
CII[245]	Chylomicron, VLDL	Lipoprotein lipase activation (LPL inhibition)	Chylomicron syndrome, pancreatitis
CIII[246]	Chylomicron, VLDL	Lipoprotein lipase inhibition	Hypertriglyceridemia
D[247]	HDL	? Activator of LCAT, neuronal function, cytoprotection	Unknown
E[248]	Chylomicron remnant, IDL, HDL	LDL receptor and LRP binding	Familial dysbetalipoproteinemia (type III), isoforms linked to Alzheimer risk
J (clusterin)[249]	HDL	Complement, chaperone, oxidative stress	Unknown
M[250]	HDL	? Cholesterol transport	Unknown

Modified from Genest J, Libby P, and Gotto AMJ. Lipoprotein disorders and cardiovascular disease. In: Zipes DP, Libby P, Bonow RO et al., eds, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine* 7th ed. Philadelphia: Elsevier Saunders; 2005; pp. 1013–1033.

HDL, High-density lipoprotein; *LCAT,* lecithin cholesterol acyltransferase; *CHD,* coronary heart disease; *VLDL,* very low-density lipoprotein; *LDL,* low-density lipoprotein; *LPL,* lipoprotein lipase; *TG,* triglyceride; *LRP,* low-density lipoprotein receptor–related protein; *IDL,* intermediate-density lipoproteins.

low-density lipoproteins (LDLs); and (5) HDLs. Within these classes, further subdivision of lipoproteins can be made. In diabetics, a higher proportion of LDL is carried in a denser, smaller diameter subfraction that is called *small dense LDL* (sdLDL). HDLs are commonly divided into HDL_2 and HDL_3 subclasses, but some investigators have adopted a gel separation method to segregate HDL into a much large number of subtypes.[6] In this chapter, the density classification terminology will be used for all lipoproteins other than chylomicrons. A standard lipid profile, consisting of measured total serum or plasma cholesterol and triglyceride levels, combined with a measured HDL-cholesterol

TABLE 25.2 Lipid Transporters, Receptors, and Regulatory Proteins

Protein Ref	Function	Major Tissue Location	Human Disease Association
ABCA1[43]	Transporter of phosphatidylcholine and cholesterol	Ubiquitous (macrophages, liver, and gut)	Tangier disease (low or absent HDL)
ABCA3[251]	Transporter of lung surfactant	Alveolar type II cells	Neonatal respiratory failure
ABCA4[252]	Transporter of *N*-retinylidene-phosphatidylethanolamine	Retinal pigment epithelial cells	Stargardt's macular degeneration
ABCA12[253,254]	Transporter of ceramide esters	Keratinocyte	Harlequin and lamellar ichthyosis
ABCG1[255]	Transporter of cholesterol	Macrophages	Unknown
ABCG5[256]	Transporter with G8 of sterols	Enterocyte	Sitosterolemia
ABCG8[256]	Transporter with G5 of sterols	Enterocyte	Sitosterolemia
Apo E-R2/(LRP8)[257]	Apo E receptor, reelin receptor	Brain, testis	Unknown
CD36[258]	Fatty acid translocase, scavenger receptor	Macrophages, skeletal muscle, platelets	? Hypertriglyceridemia
CETP[259]	Cholesterol transfer in plasma	Liver, adipose	Hyperalphalipoproteinemia (high HDL)
HM74/GPR109A[260]	Receptor for nicotinic acid	Liver	Unknown
LDL-R[36]	Receptor for LDL	Ubiquitous (liver, adrenal)	Familial hypercholesterolemia (high LDL)
LOX-1[261]	Scavenger receptor for oxidized LDL	Endothelium, smooth muscle	Unknown
LRP1[262]	Endocytosis of many ligands, including chylomicron remnants	Liver, brain, placenta	Unknown
LRP2/megalin[263]	Endocytosis receptor, multiple ligands, including lipoproteins	Kidney, brain	Donnai–Barrow and facio-oculoacoustico-renal syndromes
MTP[264]	Lipid transfer in assembly of VLDL	Liver	Abetalipoproteinemia
NPC1[265]	Cellular cholesterol trafficking	Brain, liver	Niemann–Pick C disease
NPC1L1[266]	Cellular cholesterol trafficking	Hepatocyte, enterocyte	Unknown
PLTP[267]	Phospholipid transfer protein	Ubiquitous, liver	Unknown
PCSK9[268]	LDL-R regulating protein	Liver	High and low levels of LDL
SR-A[269]	Scavenger receptor, multiple ligands	Macrophages	Unknown
SR-B1[270]	Scavenger receptor, CE transfer	Liver, adrenal	Unknown
VLDL-R[257]	Apo E receptor, reelin receptor	Brain, heart, adipose, muscle	Cerebellar hypoplasia

Modified from Genest J, Libby P, and Gotto AMJ. Lipoprotein disorders and cardiovascular disease In: Zipes DP, Libby P, Bonow RO et al., eds, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine* 7th ed. Philadelphia: Elsevier Saunders, 2005; pp. 1013–1033.

ABC, ATP-binding cassette; *ABCA,* a class of ABC transporter, *ABCG,* G class of transporter; *ACAT,* acyl cholesterol acyl transferase; *Apo E-R,* apo E receptor; *CETP,* cholesterol ester transfer protein; *LDL-R,* LDL receptor; *LOX-1,* lectin-like oxidized LDL receptors; *LRP,* LDL-R–related protein; *MTP,* microsomal triglyceride transfer protein, *NPCILI,* Niemann–Pick C1 like protein; *PLTP* phospholipid transfer protein; *PCSK9,* proprotein convertase, subtilisin/kexin type 9; *SR-A,* scavenger receptor A; *SR-BI,* scavenger receptor B1; *VLDL-R,* VLDL receptor; *HDL,* high-density lipoprotein; *LDL,* low-density lipoprotein; *VLDL,* very low-density lipoprotein; *CE*, cholesterol ester.

TABLE 25.3 Lipid Metabolism Enzymes

Protein Ref	Major Function	Major Tissue Location	Condition Caused by Altered Levels or Activity
ACAT1[44]	Cellular cholesterol esterification	Macrophages, steroidogenic tissues	Unknown
ACAT2[44]	Cellular cholesterol esterification	Liver, intestine	Unknown
Acid lysosomal lipase[272]	Hydrolysis of CE	Macrophage, adrenal, liver	CE storage disease, Wolman syndrome
Cholesterol 7-α hydroxylase[273]	Synthesis of bile acids	Liver	Cholestatic liver disease
Cholesterol 27-hydroxylase[273]	Synthesis of bile acids	Liver	Cerebrotendinous xanthomatosis
DGAT1[274]	TG synthesis	Intestine, liver, adipose, mammary	Unknown
DGAT2[274]	TG synthesis	Liver, adipose, mammary	Unknown
Endothelial lipase[275]	Hydrolysis of PL and TG	Liver, lung, kidney, endothelium	? Metabolic syndrome
Hepatic lipase[276]	Hydrolysis of TG	Liver	Elevated lipids, xanthomas, early atherosclerosis
HMG CoA reductase[277,278]	Cholesterol synthesis	Ubiquitous, liver	Unknown
Hormone-sensitive lipase[279]	Hydrolysis of intracellular TG	Adipose	Unknown
LCAT[280]	Cholesterol esterification (HDL)	Plasma, bound to HDL	Anemia, low HDL, renal failure, Fish eye disease (corneal opacities)
Lipoprotein lipase[281,282]	Hydrolysis of TG	Muscle, adipose	Chylomicronemia syndrome

ACAT, Acyl cholesterol acyl transferase; *CE*, cholesterol ester; *DGAT,* diacyl glycerol acyl transferase; *ER*, endoplasmic reticulum; *HMG*, 3-hydroxy-3-methylglutaryl; *LCAT,* lecithin cholesterol acyl transferase; *PL*, phospholipids; *TG*, triglyceride; *ACAT,* acyl cholesterol acyl transferase; *CE*, cholesterol ester; *DGAT,* diacyl glycerol acyl transferase; *HDL*, high-density lipoprotein.

level, is adequate to determine the relevant lipoprotein profile of most patients. Occasionally, specialty lipid assays are required to further characterize abnormal blood lipid values. A brief description of each of the five major lipoprotein classes is summarized in the following paragraphs. Readers interested in a more comprehensive introduction to lipoprotein metabolism should consult the lipid sections of *The Online Metabolic and Molecular Bases of Inherited Disease.* 8th ed.[7]

Chylomicrons. Chylomicrons derive from dietary fat and carry triglycerides throughout the body. Chylomicrons are generated in enterocytes in a process that appears to depend on lipin 2/3 phosphatidic acid phosphatase enzyme activity.[8] The major structural protein of chylomicrons is apolipoprotein (apo) B_{48}, a protein that is produced from apo B_{100} RNA by a unique editing process employed by gut cells that generates a protein that is 48% of the length of apo B 100.[9] Chylomicrons also contain other apoproteins, including members of the apo C and apo A family. These proteins can redistribute to other lipoprotein classes within the blood as chylomicrons transfer their lipid content to cells and interacting lipoproteins. Chylomicrons have the lowest density of all lipoproteins and will float to the top of a plasma specimen left in a refrigerator overnight, forming a cream-like layer. Owing to its large size, the chylomicron does not readily enter the artery wall and is therefore thought not to be atherogenic, but the atherosclerotic role of the triglyceride-depleted chylomicron remnant remains controversial. Triglyceride makes up most of the chylomicron and is removed by the action of an enzyme that is bound to the surface of endothelial cells, lipoprotein lipase (LPL). Patients deficient in this enzyme or its apoprotein cofactor (apo CII) have very high serum triglyceride levels and are at increased risk for developing acute pancreatitis. Inasmuch as insulin is also a critical cofactor for LPL activity, the majority of patients presenting with hypertriglyceridemic pancreatitis have poorly controlled diabetes as a major contributing factor to their delayed chylomicron clearance.

Very Low-Density Lipoproteins. VLDLs are also triglyceride-rich, but their triglyceride content is lower and their cholesterol content higher than those of chylomicrons. The protein composition of VLDLs also differs from chylomicrons in that the major structural protein is full-length apo B (apo B_{100}) as opposed to the truncated apo B_{48} form. Like chylomicrons, VLDLs are substrates for LPL-mediated triglyceride removal. The function of VLDL and chylomicrons is to carry triglycerides synthesized in the liver and intestines to capillary beds in adipose tissue and muscle, where they are hydrolyzed to provide fatty acids that can be oxidized to produce adenosine triphosphate (ATP) for energy production. Alternatively, if not needed for energy production, they can be reesterified to glycerol and stored as fat. After removal of their triglyceride, VLDL remnants (called *IDLs*) can be further metabolized to LDL. VLDLs serve as acceptors of cholesterol transferred from HDL, accounting in part for the inverse relation between HDL cholesterol and VLDL triglyceride. This transfer process is mediated by an enzyme called *cholesterol ester transfer protein* (CETP). The most common lipid abnormality seen in patients with type 2 diabetes is an elevated VLDL level, the causes of which will be discussed in the section on diabetic lipid abnormalities.

Low-Density Lipoproteins. LDLs are the major carriers of cholesterol in humans, responsible for supplying cholesterol to tissues with the highest sterol demands. LDLs are also the lipoproteins most clearly implicated in causing atherogenic plaque formation.[10] Circulating LDL levels can be increased in persons who consume large amounts of dietary saturated fat and/or cholesterol.[11] LDL levels are also elevated in those who have genetic defects that affect LDL receptor function (e.g., familial hypercholesterolemia [FH], proprotein convertase subtilisin kexin 9 (PCSK9) mutations, autosomal recessive hypercholesterolemia) or the structure of LDL's apoprotein, apo B, or who have polygenic disorders affecting LDL metabolism.[12] Circulating LDLs are taken up by the endothelium lining the artery wall, traverse it, and are deposited and trapped in the arterial intima. There, they may undergo oxidation or other biochemical modification, be taken up by macrophages, and stimulate atherogenesis. The association of total serum cholesterol with coronary heart disease (CHD) is predominantly a reflection of the strong correlation of LDL-cholesterol levels with total cholesterol levels in most humans. Studies in patients with diabetes have produced somewhat conflicting data on the prevalence of increased LDL levels, as defined by general population norms, but the preponderance of the evidence indicates that diabetics have LDL levels that are similar to those seen in well-matched, nondiabetic control populations. Given the greater risk of developing coronary artery disease (CAD) in individuals with diabetes, however, the National Cholesterol Education Program guidelines defined LDL levels above 100 mg/dL as being undesirable in those with diabetes. When judged by this stringent criterion, elevated LDL levels are extremely common in patients with diabetes. In addition, the LDL in individuals with diabetes is frequently smaller, denser, and more readily oxidized—all properties that are associated with a higher degree of atherogenicity.[13,14]

High-Density Lipoproteins. HDLs are the smallest of the lipoproteins; despite this, they carry a variety of apoproteins, including members of the apo A and C families. HDLs, unlike the other lipoproteins, are believed to function in humans primarily to return lipids from peripheral tissues to the liver and gut for excretion, rather than moving lipids from the gut organs to the periphery.[15] This perspective on HDL function is an oversimplification, as cell culture experiments indicate that HDL can donate lipids as well as pick them up; moreover, other components of blood in addition to HDL may be involved in the reverse transport process.[16-18] It has proven to be quite difficult to quantitate the net movement of cholesterol carried in HDL in and out of peripheral tissues in whole animals. Further complicating the issue is the fact that many of the standard laboratory animals used in metabolic studies carry most of their serum cholesterol in HDL rather than LDL, diminishing the relevance of studying their lipoprotein metabolism as a surrogate for understanding that of the LDL-rich human. Nevertheless, a considerable body of evidence suggests that HDL functions to protect tissues from unwanted accumulation of cholesterol. The mechanisms by which HDL may protect against atherosclerosis include its involvement in the reverse cholesterol transport pathway just described, as well as through contributions arising from its carriage of proteins that appear to have anti-inflammatory and antioxidant properties.[19,20] The unesterified cholesterol from tissues that is transferred to HDL is esterified by the action of lecithin cholesterol acyltransferase (LCAT) and stored in the central core of HDL. This esterified cholesterol can be transferred back to lower-density lipoproteins by the action of cholesterol ester transfer protein, or it may be removed at the liver by the action of a plasma membrane receptor called *scavenger receptor B-1.* A particularly effective reverse transport system is thought to explain, at least in part, the association of elevated HDL-cholesterol levels with a reduced risk of developing CHD. To date, however, therapies directed at raising HDL-cholesterol levels have failed to improve CV outcomes unless LDL cholesterol was also substantially lowered. In addition, studies of some genetic variants that appear to control HDL levels have been shown not to predict atherosclerosis risk, unlike genetic variants that control LDL levels, calling into question a causal role for HDL in impeding atherosclerosis.[21] Thus, the inverse association of HDL-cholesterol levels with atherosclerotic vascular disease risk, while epidemiologically well established, does not ensure that manipulations of HDL-cholesterol by therapeutic agents or lifestyle modifications will confer a CV benefit via that lipoprotein change. Exercise increases HDL levels; obesity, hypertriglyceridemia, and smoking lower them. Patients with diabetes, having higher body mass indices and hypertriglyceridemia, typically have lower HDL cholesterol levels than do nondiabetics, contributing further to the atherogenic lipid profile of the diabetic.

Metabolism of Lipids and Lipoproteins

The synthesis of lipoproteins in the liver and gut is a complex process regulated by hormonal and transcriptional networks that are affected by dietary intake of both lipid and carbohydrate. The disorders of carbohydrate and lipid metabolism caused by diabetes have both direct and indirect effects on lipoprotein metabolism. A brief summary of normal lipoprotein metabolism will be presented in this section, followed by a description of the impact of diabetes on these processes in the next section.

Following ingestion of a meal containing fat, the triacylglycerols (TAGs) are broken down by pancreatic triacylglycerol lipase to yield sn-2-monoacylglycerol and fatty acids. Ingested cholesterol is solubilized by bile salts into micellar structures, a step that is disrupted by bile acid resin binders. The lipids then transverse the enterocyte brush border membrane and are reassembled in the intestinal cell into a triglyceride-rich lipoprotein, the chylomicron. The role of transport proteins in the passage across the enterocyte membrane remains disputed, with evidence for both a passive diffusion and an active transport mechanism.[22,23] For cholesterol, a second transport step translocates the sterol into intracellular pools that can be accessed for lipoprotein packaging.[24] Niemann-Pick C1-like 1 protein appears to be required for this second step to occur and is the target of the cholesterol absorption inhibitor ezetimibe.[25,26] The internalized cholesterol is moved to the endoplasmic reticulum (ER), where the monoacylglycerol and fatty acids are also carried by transport proteins. The latter are recombined into TAG and bound by microsomal triglyceride transport protein (MTP). A second pathway of triglyceride synthesis uses acylation of glycerol-3-phosphate to phosphatidic acid, dephosphorylation to diacylglycerol (DAG), and then acylation of DAG to TAG. A high-density, protein-rich particle synthesized in the enterocyte, containing apo B_{48} and apo AIV, is then packaged with the absorbed lipids to generate the chylomicron.[27] This is then transported by vesicles to the basolateral membrane for exocytosis into the mesenteric lymph. The lipoproteins in the lymph enter the blood circulation via discharge from the thoracic duct.

Two ABCG half transporters, ABCG5 and G8, contribute to net cholesterol absorption in the gut through their ability to resecrete sterol that has been absorbed by the enterocyte back into the gut lumen, thereby reducing net sterol uptake.[28] The total amount of diet-derived TAG that is delivered to the circulation varies and is influenced by multiple factors, including the fat content of the diet, the amount of phosphatidylcholine in the intestinal lumen, and the expression level of apo AIV in the enterocyte. Overall, humans have a remarkable ability to absorb large quantities of both cholesterol and fat; 35% to 60% of ingested cholesterol is absorbed, with the amount inversely correlated with hepatic cholesterol synthesis. Up to 600 g of fat is absorbed with 95% efficiency. In individuals with normal lipid metabolism, the ingested fat is cleared from the enterocyte in less than 14 hours, forming the basis for measuring serum lipids after a fast of this duration. Circulating chylomicrons are subsequently cleared from the plasma by the action of endothelial-bound LPL, which cleaves the TAG, enabling the liberated glycerol and free fatty acids to move into the adjacent adipose or muscle tissue. The resulting chylomicron remnant particle can be further metabolized by lipases and appears to be retained in the hepatic space of Disse, where it can bind to heparan sulfate proteoglycans and acquire additional apo E. The acquisition of the apo E enables the remnant to be cleared efficiently by either the LRP1 or LDL receptor, but patients with diabetes have accumulation of remnants and slower lipoprotein hepatic clearance.[29,30]

The other major triglyceride-rich lipoprotein, VLDL, is synthesized in the liver and contains apo B_{100} rather than apo B_{48} as its major structural protein.[31] As with chylomicrons, VLDLs rely on MTP to combine lipids with the apo B protein to produce a nascent, lipid-poor lipoprotein.[32,33] This VLDL can be secreted from the liver or further lipidated to produce a mature VLDL. Individuals lacking MTP activity develop abetalipoproteinemia, a disorder characterized by erythrocyte acanthocytosis, anemia, steatorrhea, spinocerebellar degeneration, and fat-soluble vitamin deficiency. Acquisition of lipid by the mature VLDL appears to depend on triglycerides that are stored in the cytosol. Increased fatty acid delivery to the liver, from dietary sources or that released by lipase activity in peripheral tissues, stimulates VLDL production. After secretion of the mature VLDL from the liver, its triglycerides are initially removed by the action of LPL, and the resulting IDL can be further catabolized by hepatic lipase to produce mature, cholesterol-rich LDL.

Low-Density Lipoprotein Metabolism. There is still some dispute about whether mature LDL can be secreted from the liver rather than always being derived from the progressive removal of triglyceride from VLDL/IDL precursors.[34] However, the preponderance of data indicates that little if any LDL is directly secreted from the liver. Each LDL particle contains a single apo B molecule, which dictates that individuals heterozygous for apo B mutants will have two populations of LDL: one that carries a wild-type protein and one with a mutant protein. LDLs are utilized throughout the body as a source of exogenous cholesterol and are taken up by LDL receptors present at the plasma membrane. While LDL receptors are ubiquitous, the majority are expressed by the liver. The LDL receptor binds to both apo E– and apo B–containing lipoproteins. As the affinity of the receptor is higher for apo E than for apo B, apo E–containing lipoproteins, such as IDL, can compete with LDL for uptake by the LDL receptor. However, there are other receptors that can bind apo E–containing lipoproteins, permitting the clearance of the latter, but not that of LDL when LDL receptor activity is lost or diminished. Thus, mutations in the LDL receptor result in elevated LDL levels in the plasma, causing the autosomal codominant disorder FH, which is among the most common of Mendelian genetic disorders (see Genetics section later).[35] The elucidation of the cholesterol homeostatic mechanisms controlled by the LDL receptor constituted one of the major advances in our understanding of human physiology and disease.[36] The internalization of LDL by the LDL receptor leads to trafficking of the internalized lipoprotein to the endosome, where, after fusion with lysosomes, the sterol in the lipoprotein is released and shuttled to the ER. A sterol-sensing system in the ER regulates the movement of a membrane-bound transcription factor sterol response element binding protein (SREBP) from the ER to the nucleus.[37-42] SREBP works in concert with other transcriptional regulators to influence the expression of many of the cholesterol synthesis, metabolizing, and transport molecules, particularly in the liver. This elegant homeostatic system enables cells to reduce their *de novo* cholesterol synthesis when adequate sterol supplies are present and to increase synthesis and LDL receptor expression when cellular demands require more cholesterol.

High-Density Lipoprotein Metabolism. Although the relationship between elevated levels of HDL and lower rates of cardiovascular disease was recognized more than 65 years ago, a mechanism by which HDL causally protects against CHD remains to be elucidated. HDL are initially produced as lipid-poor phospholipid discs containing apolipoprotein AI. Following interaction with phospholipid/cholesterol transporters, primarily ABCA1, the nascent HDL particle acquires unesterified cholesterol from cellular stores.[43] For macrophages in particular, this cholesterol transport process constitutes a cholesterol efflux pathway that appears to be critical to the cell's ability to unload excess cholesterol acquired by phagocytosis of dead cells and by the unregulated uptake of lipoproteins by scavenger receptor pathways.[15] Cholesterol is stored in an intracellular lipid pool, having been esterified at its 3'OH position by the activity of acyl coenzyme A (CoA):cholesterol *O*-acyltransferase (ACAT).[44] A neutral cholesterol hydrolase cleaves the fatty acid from cholesterol ester to convert it back to free (unesterified) cholesterol when cellular signals call for the use or export of the sterol.[45] This free cholesterol can be shuttled to the plasma membrane, where it can be transported across by the action of ABCA1. The interaction of lipid-poor nascent HDL with ABCA1 appears to involve direct binding of apo AI to the transporter, but the actual lipid transfer process is poorly understood.[46-50] Once the unesterified cholesterol transfers to the nascent HDL, it can be reesterified by the action of LCAT and centralized in the lipid core of the growing HDL particle. This HDL particle can now interact with a second ABC transporter, ABCG1, and additional cholesterol can be acquired.[49] The role of ABCA1 is well established, as loss of its activity leads to Tangier disease (see Genetics section) and the absence or near absence of HDL in the plasma. Once the HDL has acquired sufficient cholesterol and esterified it, it can donate this lipid back to lower-density lipoproteins in the plasma, such as VLDL, through the activity of CETP. Alternatively, HDL can interact with a cellular receptor, SR-BI, to selectively transfer the cholesterol ester to cells, including the liver, where it could be excreted in the bile as cholesterol or one of its bile acid derivatives.[51] The protein component of HDL is cleared via the kidney, a process that appears to depend on the HDL being lipid depleted. The pathway just described, moving cholesterol from cellular storage pools to HDL and ultimately to the liver, is termed the *reverse cholesterol transport pathway.* The activity of this pathway has been postulated to explain at least in part the link between HDL levels and lower CHD risk. However, HDL also carries antioxidants and a host of proteins linked to complement and inflammation pathways that suggest it may be playing a more general role as a scavenger of molecules that stimulate inflammation. The complexity of its function has made targeting increases in HDL levels as a therapeutic objective very challenging. Both animal and human data have clearly established that simply raising HDL cholesterol levels does not ensure that a benefit on the atherosclerotic process will ensue.

KEY POINTS

- Lipoproteins vary in their lipid and protein composition forming a continuum of density from chylomicrons, which are the largest and least dense, to high density lipoproteins which are the smallest and most dense. They serve various biological functions but primarily are involved in moving cholesterol and triglycerides between tissues via the bloodstream. Both excess numbers and the inability to generate sufficient numbers of specific lipoproteins are associated with adverse health consequences.

DISORDERS OF LIPID METABOLISM IN PATIENTS WITH DIABETES

Type 1 Diabetes

Studies dating back more than three decades documented that individuals with type 1 diabetes have increased rates of CHD.[52-54] This risk was estimated to be 5- to 10-fold or greater than that seen in age-matched controls without diabetes. In the Diabetes UK cohort of 23,751 subjects who were diagnosed with insulin-dependent diabetes prior to the age of 30 years, for example, those individuals in the study who were 20 to 29 years of age at its onset had a standardized mortality ratio for ischemic heart disease of 11.8 for men and 44.8 for women. For those 30 to 39 years of age, the ratios were 8.0 and 41.6 for men and women, respectively.[55] More recent studies still show a significantly greater relative risk of atherosclerotic vascular disease in patients with type 1 diabetes, though the absolute risk has fallen substantially in that population in recent decades.[5,56,57] Despite the long and consistent reproducibility of this risk association, the pathophysiology underlying that relationship is still poorly understood. Unlike type 2 diabetes, where obesity and insulin resistance lead to changes in the standard lipid profile that are considered atherogenic, the type 1 diabetic who has not developed nephropathy and whose glucose is well controlled will typically have a normal serum lipid profile.[58] In the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications study, the mean LDL-cholesterol in the group that was given intensive glucose-lowering treatment was 111 ± 29 mg/dL in women and 119 ± 31 in men; HDL-cholesterol levels averaged 63 ± 16 and 51 ± 14 for women and men, respectively. Serum triglyceride levels were also in the normal range, with women averaging 76 ± 37 mg/dL, and male values running moderately higher at 98 ± 67 mg/dL. These lipid values, which would be considered normal in most Western countries, did not differ significantly from those of the patients enrolled in the conventional glucose-lowering treatment group of the study. These findings make clear that the standard lipid profile provides few insights into the increased risk of CHD in type 1 diabetics.[59]

Surprisingly, epidemiologic risk-factor assessment of the correlates of CHD in type 1 diabetes shows a weak association to the one metabolic risk that predominates in the type 1 population: i.e., hyperglycemia.[60,61] It was therefore somewhat unexpected when a 17-year follow-up of the DCCT cohorts showed that the intensive treatment group had a 57% reduction in the risk of nonfatal myocardial infarction, stroke, or death from CV disease compared with the conventional treatment group (95% confidence interval [CI]: 0.21–0.88; $P = 0.02$).[62] Other studies of the impact of glucose control on CV outcomes have produced results that suggest that strict control of HbA1c levels is less important in reducing atherosclerotic vascular events or, in some circumstances, may even be detrimental.[63] Independent of the mechanism by which CV benefit is conferred, recent studies do suggest that current therapies are making a clear impact on reducing disease in type I diabetics.[5] Miller et al. reported a significant improvement in overall life expectancy in type 1 diabetics who were diagnosed between 1965 and 1980 versus similarly aged type 1 diabetics whose diagnosis was made between 1950 and 1964.[64] Although the data presented do not permit delineation of what accounts for this improvement, better care of ketoacidosis, renal disease, blood pressure and lipids are all strong contenders for contributors to the longevity improvement. With increasing numbers of patients experiencing the onset of type 2 diabetes in childhood or adolescence, it is now possible to compare outcomes of type 1 and type 2 diabetics in individuals whose diabetes was diagnosed early in life. Those with early-onset type 1 diabetes fare significantly better than those with early-onset type 2 diabetes, and a clear difference in CV death has been observed.[65] Finally, a study of diabetes-related complications, not specific as to type 1 or type 2, but likely predominantly representing type 2 diabetics, did find a substantial reduction in major complications of diabetes between 1990 and 2010, with a reduction in acute myocardial infarction showing the largest improvement.[66]

Type 2 Diabetes

There are alterations in multiple lipoprotein metabolism pathways in type 2 diabetics that lead to the typical abnormal lipid profile commonly seen. This profile includes hypertriglyceridemia secondary to high VLDL levels, as well as increased amounts of sdLDL and lower levels of HDL cholesterol. LDL-cholesterol levels are typically not higher in diabetic populations when compared with well-matched nondiabetic subjects. In an elderly Finish population, for example, approximately 30% of diabetic subjects had serum triglyceride levels greater than 200 mg/dL compared with 13% of nondiabetic subjects. HDL cholesterol levels were less than 35 mg/dL in about 25% of diabetics compared with 12% of nondiabetics. In this study, elevated LDL levels as defined by an LDL greater than 130 mg/dL were actually more prevalent in the nondiabetic population than in those with diabetes.[67,68]

The causes of classic diabetic dyslipidemia are only partly understood. Individuals with insulin resistance and type 2 diabetes overproduce mature VLDL in the liver. This is accompanied by a slower clearance of triglyceride-rich lipoproteins via reduced activity of LPL. Coupled with decreased receptor uptake of the remnant and IDL particles that form after LPL partially depletes VLDL of its triglyceride, diabetes produces substantial elevations in serum triglyceride levels. Production of sdLDL appears to be tightly linked to the resulting hypertriglyceridemia insofar as CETP exchanges triglyceride from VLDL for cholesterol ester from LDL, producing a triglyceride-enriched LDL.[69] This lipoprotein appears to be a preferred substrate for hepatic lipase, which hydrolyzes the triglyceride to produce an LDL that is smaller, lower in cholesterol ester content, and higher in density than LDL produced in the absence of excess VLDL. The diabetic with increased sdLDL levels has a higher LDL particle number for an equivalent serum LDL-cholesterol concentration than does an individual with normal LDL. This increased particle number, whether measured by nuclear magnetic resonance assays or serum apo B levels, is associated with an increased risk of CAD.[14]

Thus, the central metabolic change that accounts for much of the diabetic dyslipidemia is the overproduction of VLDL by the liver. The molecular mechanisms accounting for VLDL overproduction are not fully elucidated, but there is evidence that insulin activates a phosphatidylinositol-3 kinase pathway that inhibits apo B secretion while activating a mitogen-activated protein kinase (MAPK) that downregulates MTP expression.[70,71] Insulin effects on transcriptional regulators of VLDL secretion may also be involved. *In vivo* metabolic studies suggest that the prevalence of hypertriglyceridemia and low HDL-cholesterol levels is higher in diabetics who are insulin-resistant as opposed to insulin-deficient, accounting for the higher prevalence of these disorders in patients with type 2 versus type 1 diabetes.

KEY POINTS

- Normal body mass individuals with type I Diabetes commonly have a normal lipid profile with neither high levels of LDL or VLDL nor low levels of HDL. Patients with type 2 diabetes, particularly those with significant elevations in BMI, often have increased levels of VLDL (serum triglycerides are elevated) and lower levels of HDL cholesterol. The overall CV risk of those with type 2 diabetes is greater than those with type 2 diabetes, but both groups are higher risk than those without diabetes

GENETIC BASIS OF LIPID DISORDERS

Although metabolic disorders associated with diabetes, hypothyroidism, and nephropathy can exacerbate or even cause hyperlipidemia (Table 25.4), a substantial number of dyslipidemic individuals have none of these problems. Individuals with the greatest deviation from normal levels of lipoproteins often have a single gene defect that is responsible for their lipid disorder. The vast majority of hyperlipidemic patients, however, do not have a monogenic defect. Rather, they are most likely to have a polygenic disorder and an additional contribution from environmental factors (e.g., excessive saturated fat intake for high LDL levels; obesity or smoking for lower HDL levels). Currently, a genetic diagnosis is rarely required to enable appropriate lipid care to be given. Nevertheless, the identification of the causes of monogenic lipid disorders has been critical to our understanding of both normal and abnormal lipoprotein physiology. The lipid disorders with the greatest impact on developing CAD risk are those that increase the serum LDL level or reduce the HDL level, and these are briefly presented next.

TABLE 25.4 Secondary Causes of Hyperlipidemia and Dyslipidemias

Metabolic/hormonal	Diabetes
	Hypothyroidism
	Lipodystrophy
	Polycystic ovarian disease
Hepatic	Primary biliary cirrhosis
	Other forms of cirrhosis
Renal	Chronic renal failure
	Nephrotic syndrome
Dietary	Alcohol
	Foods highly enriched in cholesterol and saturated fat
Medications	Estrogens
	Glucocorticoids
	Anti-HIV– treatments, especially protease inhibitors
	Oral androgens and anabolic steroids
	Thiazide diuretics
	Beta-blockers
	Retinoic acid
	Antipsychotics

Monogenic Low-Density Lipoprotein Disorders

The molecular basis of four monogenic disorders that primarily affect LDL levels have been characterized: (1) FH, (2) familial defective apo B, (3) autosomal recessive hypercholesterolemia, and (4) PCSK9 mutations. These disorders illustrate distinct mechanisms leading to elevated LDL levels.

Familial Hypercholesterolemia. FH is one of the most thoroughly studied and common genetic disorders of mankind. Recent analyses have suggested the prevalence of individuals carrying a mutation in one allele encoding the LDL receptor is approximately 1 in 300, considerably higher than the prior estimates of 1 in 500.[72] These mutations results in defective clearance of LDL from the blood and a rise in serum total and LDL cholesterol levels. The elucidation of this defect and its associated cell biology led to insights into the homeostatic control of cholesterol metabolism that transformed the lipid field.[35,36] Well over 1000 individual mutations in the LDL receptor have now been identified. The gene, located on chromosome 19 and spanning 45 kb, has 18 exons that encode a mature protein of 839 amino acids. The inheritance of the disorder is autosomal codominant. Heterozygous patients typically have LDL-cholesterol levels in the 200 to 500 mg/dL range, and the rare homozygous FH patient usually has an LDL-cholesterol well above 500 mg/dL. Heterozygous patients commonly have tendon xanthomas and premature CAD, whereas these are universal in the untreated homozygous individual. Mutations in the LDL receptor have

TABLE 25.5 **3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Inhibitors (Statins): Key Properties of Clinical Interest**

Statin	Atorvastatin	Fluvastatin	Lovastatin	Pitavastatin	Pravastatin	Rosuvastatin	Simvastatin
Dose (typical range for agent)	10–80 mg	20–80 mg	20–80 mg	1–4 mg	20–80 g	5–40 mg	10–80 mg*
LDL-C reduction	38%–54%	17%–33%	29%–48%	31%–41%	19%–36%	42%–63%	28%–48%
Cytochrome P450 metabolism	3A4	2C9	3A4	2C9, 2C8	None	Some 2C9	3A4, 3A5
Renal excretion of absorbed dose	2%	<6%	10%	15%	20%	10%	13%
Mortality or CHD event benefit shown	Yes	No	Yes	No	Yes	Yes	Yes
Generic available	Yes	Yes	Yes	No	Yes	Yes	Yes

*Though approved at 80 mg, the US Food and Drug Administration has issued a warning about the use of this dosage due to higher levels of myopathy.

LDL-C, Low-density lipoprotein–cholesterol; *CHD,* coronary heart disease.

been described that affect multiple aspects of receptor biology, ranging from defective LDL binding to abnormal trafficking of the receptor to the plasma membrane. The effect of all of these mutations is to reduce LDL and IDL clearance from the blood. The latter effect results in an enhanced conversion of IDL to LDL. The net result is markedly elevated LDL cholesterol levels in the blood and a strong predisposition to early CHD if the disorder is not treated. FH heterozygotes typically respond well to 3-hydroxy-3-methylglutaryl CoA (HMG-CoA) reductase inhibitors (statins), though additional therapies may be needed to reach desirable LDL levels, especially in those with the highest pretreatment LDL values. FH homozygotes respond poorly to reductase inhibitors alone and inevitably develop early atherosclerotic vascular disease unless much more aggressive treatment methods, such as LDL apheresis, are utilized. Recent advances in therapeutics, however, have offered hope for less arduous forms of treatment. The finding that many patients previously assumed to have homozygous FH in reality had compound heterozygous FH, with one partially functioning LDL receptor allele, led to the finding that PCSK9 inhibitors can produce substantially better results in those patients than statins, presumably due to better recycling of the functioning LDL receptors.[73-76] Two specific therapies for FH homozygotes have been approved by the US Food and Drug Administration (FDA): one an oral drug targeting microsomal triglyceride transport protein (lomitapide) and the other an antisense therapy directed against apo B (mipomersen).[77-80] LDL levels were reduced by 25% to 36% with mipomersen and approximately 50% with lomitapide, both outperforming the effect of statins on the LDL levels in these rare patients. However, both drugs appeared to lead to liver fat accumulation, which contributed to the withdrawal of mipomersen from the market. Orphan disease pricing made their use even in the rare FH homozygous patient population very difficult. A trial of a novel monoclonal antibody targeting ANGPTL3 inhibition (see section on ANGPTL3 later) has also recently shown promising effects in the FH homozygous population.[81]

Familial Defective Apolipoprotein B. Familial defective apo B (FDB) is an autosomal codominant disorder caused by mutations near the C-terminus of apo B.[82,83] Apo B is a 4536–amino acid protein encoded by a gene on chromosome 2 that is composed of 29 exons. The most commonly identified mutation in FDB affects an arginine at position 3500 that is mutated to either a glutamine or tryptophan. Most individuals with FDB are heterozygous for their mutation, and there is only one molecule of apo B per LDL particle, so the LDL in the serum of these individuals is a mixture of normal LDL and FDB LDL. The LDL containing the defective apo B does not bind normally to the LDL receptor and is therefore cleared more slowly. The explanation for the lower receptor affinity of LDL carrying the FDB mutations is not known with certainty. There is evidence, however, of a molecular interaction between the arginine at position 3500 and tryptophan at 4369 that appears to be necessary for the adoption of an apo B conformation recognized by the LDL receptor. Patients with FDB have elevations in their LDL levels that fall within the range of those seen in FH heterozygotes, but the mean is generally lower than the mean in FH cohorts. There is considerable variability in both genetic populations, however, making generalizations about the diagnostic value of this measurement not very useful. Tendon xanthomas are common but not universal, and premature CAD is prevalent. Treatment with reductase inhibitors or other standard LDL-lowering drugs is typically quite effective in FDB patients.

Autosomal Recessive Hypercholesterolemia. Studies of families in which LDL levels were inherited in an autosomal recessive pattern led to a successful genetic linkage analysis that mapped the genetic defect in these families to the short arm of chromosome 1. Candidate gene sequencing led to the identification of mutations in a gene now called *LDLRAP1*, also known as *autosomal recessive hypercholesterolemia* (*ARH*), comprising nine exons and eight introns. Patients diagnosed with ARH to date have been clinically similar to patients who are homozygous for LDL receptor mutations in that they have markedly elevated LDL cholesterol levels (typically >500 mg/dL), tendon xanthomas, and premature coronary atherosclerosis.[84] The autosomal recessive mode of inheritance is an important differentiating factor between ARH and FH, as is greater LDL receptor activity, as measured in cultured fibroblasts taken from the patients. The defect in ARH patients appears to affect liver cholesterol metabolism disproportionately, and current data indicate that the *ARH* gene product likely serves as an adaptor protein required for LDL receptor internalization via clathrin-coated pits.[85,86] ARH patients do respond to HMG CoA reductase inhibitors, but this treatment is usually inadequate to control their markedly elevated LDL-cholesterol levels, making them candidates for lomitapide and other combination lipid therapy.[87]

Proprotein Convertase Subtilisin Kexin 9 Mutations. PCSK9 is a 72-kD protease that is highly expressed in the liver and regulates the levels of LDL receptors acting at the plasma membrane.[88,89] Gain-of-function mutations in PCSK9 result in lower LDL receptor levels, whereas loss-of function mutations increase those levels. Serum LDL levels are inversely related to liver LDL receptor expression; the activity of the latter clears the former from the circulation. Thus, gain-of-function mutations are associated with higher serum LDL levels and increased risk of CHD, and loss-of-function mutations result in lower LDL levels and decreased CHD risk. The role of PCSK9 in lipid metabolism was originally recognized when it was discovered that specific heterozygous mutations in the protein were associated with autosomal dominant hypercholesterolemia. Approximately 1 in 50 Black Americans has one of two nonsense mutations in PCSK9, causing

LDL-cholesterol levels to fall approximately 30%. White Americans can carry a missense mutation that reduces LDL-cholesterol by approximately 15%. Individuals with complete loss of PCSK9 activity have extraordinarily low levels of LDL-cholesterol but appear to be otherwise normal. The mechanism by which PCSK9 regulates LDL receptor activity is still being elucidated, but it appears that the protease can bind to the receptor, targeting it for degradation in the lysosome. This action does not depend on the protein's enzymatic activity, and exogenously administered PCSK9 can reduce LDL receptor expression and increase serum LDL-cholesterol levels, while approaches that reduce the amount of circulating PCSK9 have the opposite effect. The mechanism by which PCSK9 works made it an extremely attractive target for pharmaceutical intervention, and multiple agents that inhibit this protein have been developed as LDL-cholesterol–lowering therapies (see Diet and Drug Treatment section of this chapter).

ANGPTL3 Mutations. *ANGPLT3* is hepatically-derived circulating protein that inhibits both LPL and endothelial lipase. Mutations in this gene were first discovered to confer low plasma lipid levels inherited in a recessive manner in animal models, and resequencing of *ANGPLT3, ANGPLT4,* and *ANGPLT5* in the Dallas Heart Study identified multiple rare synonymous variants associated with low plasma triglyceride concentrations.[90,91] Exome sequencing in 2010 identified mutations in *ANGPTL3* as the cause of low LDL, HDL, and triglyceride levels in two siblings.[92,93] Studies to date suggest that mutations in *ANGPTL3* may cause a significant percentage of hypolipoproteinemia not explained by *APOB* mutations, particularly in the context of low total cholesterol and HDL levels. Properties of this protein have made it an attractive target for novel lipid-lowering therapeutics, and a recent trial of a monoclonal antibody (evinacumab) targeting ANGPTL3 resulted in significant LDL lowering in an FH homozygous population.[81]

Monogenic High-Density Lipoprotein Disorders

Three distinct monogenic disorders cause markedly reduced levels of HDL in the plasma by different mechanisms. They are: (1) Tangier disease, (2) LCAT deficiency, and (3) apo AI mutations.

Tangier Disease. Tangier disease was first recognized in 1960 in a sibling pair living on Tangier Island in the Chesapeake Bay in Virginia. Enlarged, yellow-orange tonsils and little or no circulating HDL cholesterol are the classic findings in the disorder. Subsequently, cases have been identified in which the presenting symptom was a peripheral neuropathy. Patients may also have hepatosplenomegaly. Serum LDL and total cholesterol levels are usually quite low, while serum triglyceride values are moderately elevated. Tangier disease is an autosomal recessive disorder. The cause of Tangier disease was identified by several groups following mapping of the gene defect to chromosome 9. Candidate gene analysis in the appropriate genetic interval identified mutations in an ATP-binding cassette (ABC) transporter as the cause. The transporter, now called *ABCA1* (ATP-binding cassette, subfamily A, member 1), is a full-length ABC transporter transmembrane protein that is predicted to span the plasma membrane 12 times. ABCA1 is a 2261–amino acid protein encoded by a gene spanning 50 exons. Approximately 100 mutations in the gene have been identified to date.[50,94] ABCA1 mediates the efflux of cholesterol from cholesterol-enriched cells when stimulated by the major apoprotein of HDL, apo AI.[95] This activity is lost in Tangier patients and is reduced by approximately half in carriers of one abnormal *ABCA1* allele. The mechanism of the movement of cholesterol from inside the cell to outside the cell has not been established. Individuals with Tangier disease, as well as heterozygous carriers of ABCA1 mutations (a disorder called *familial hypoalphalipoproteinemia*), appear to have increased risk of premature coronary disease. There is no specific therapy for this disorder.

Lecithin Cholesterol Acyltransferase Deficiency. The esterification of free cholesterol in circulating lipoproteins is catalyzed by a plasma enzyme called *lecithin cholesterol acyltransferase.* Two clinically separable syndromes result from a deficiency of LCAT. Fish-eye disease is due to a partial deficiency of LCAT, with patients presenting with dense corneal opacities and very low HDL-cholesterol levels. Familial LCAT deficiency arises from a nearly completely absence of LCAT activity and produces a more severe syndrome characterized by corneal opacities, anemia, and proteinuric renal failure.[96] The serum lipid and lipoprotein profile in the more severe disorder is characterized by normal or increased triglyceride levels, reduced LDL-cholesterol values, and markedly diminished HDL levels. The gene encoding LCAT is located on chromosome 16 and is composed of six exons. Cleavage of a signal peptide of 24 amino acids converts the proenzyme from a 440–amino acid precursor to the final 416–amino acid glycoprotein that circulates in the plasma. When unesterified cholesterol from tissues is transferred to HDL, either by passive diffusion or by ABCA1-mediated lipid transport, LCAT's activity esterifies the transferred cholesterol, trapping it in the HDL core. Because cholesterol ester is more hydrophobic than unesterified cholesterol, it is energetically unfavorable for the cholesterol ester to transfer back to the cell of origin. These steps of cholesterol transfer and esterification are the initial events in the reverse cholesterol transport pathway whereby cholesterol is moved from peripheral tissues back to the liver. Apo AI is the major activator of LCAT activity, accounting for the predominant effect of the enzyme deficiency on HDL levels. LCAT does, however, contribute to esterification of cholesterol in lower-density lipoproteins as well. Despite very low HDL levels, patients with either fish-eye disease or familial LCAT deficiency do not seem to have a predilection for very early coronary atherosclerosis. The small number of patients with the disease, some of whom have been found to have CHD, makes it difficult to determine if the risk of coronary atherosclerosis is substantially altered by the enzyme deficiency. There is no specific treatment for LCAT deficiency. Corneal and kidney transplantation are performed in these patients to ameliorate their major clinical disabilities.

Apolipoprotein AI Mutations. Apo AI is the major structural protein of HDL. The gene encoding apo AI is located on chromosome 11 and comprises four exons. Following cleavage of the signal and prohormone sequences, a 243–amino acid mature protein is produced. The protein has multiple repeats of an amphipathic helical structure that enables it to interact with both lipid and aqueous environments. Mutations that cause profound alterations in apo AI structure or expression have been reported, though they are extremely rare.[97-100] The individuals carrying these mutations have virtually no circulating HDL and typically develop early CHD. There are, however, a number of case reports of apo AI mutations that are not associated with early CHD. Corneal opacities and xanthomas have been documented in many, but not all, individuals with major apo AI mutations. Most patients harboring these mutations have been found to be homozygous for the gene defect, usually as a result of consanguinity. Heterozygotes carrying these mutations commonly have half-normal HDL-cholesterol levels, although cases with greater reductions in HDL levels have been reported, suggesting that some mutations may exert a dominant negative effect. HDL typically contains four apo A molecules per particle (either four apo AI or two apo AI and two apo AII proteins), so a heterozygous individual carrying an expressed apo AI mutant would have at least one mutant apo AI on most HDL particles. The atherosclerosis of individuals with structural mutations in apo AI appears to be much more pronounced than that seen in patients with either LCAT deficiency or Tangier disease. No specific therapy is available for this disorder, but aggressive LDL lipid-lowering therapy is justified.

Polygenic Lipid Disorders

Approximately 50% of the total interindividual variation in serum lipid levels is attributed to heritable factors.[101] Several GWAS studies have been conducted to identify genetic variants that contribute to population-wide lipid variability. One of the early, large lipid GWA studies was conducted in approximately 10,000 individuals and identified a total of 31 genetic loci associated with HDL levels, 22 loci associated with LDL levels, 16 loci associated with triglyceride levels, and 39 loci associated with total cholesterol levels; these results accounted for approximately 25% to 30% of the genetic component of these lipid traits, suggesting many lipid associated genetic variants remain to be found.[102]

A main finding of lipid GWA studies has been that genes which confer monogenic lipid disorders also harbor commonly occurring genetic variants with more modest effects on lipid levels.[103,104] The implications of genetic variants discovered in gene regions without known effect on lipid biology are less clear. Novel biology has been revealed through the functional characterization of genetic variants discovered in the *SORT1* and *TRIB1* genes, but more work is required to understand the effect of many variants discovered through GWA studies.[105,106]

Findings from GWA studies have also been used to test hypotheses within lipid biology. If genetic variants contribute to serum lipid levels, and serum lipid levels contribute to CV risk, then genetic variants for lipid levels should influence CV risk. In 2012, Voight and colleagues found no relationship between genetic variants associated with HDL levels and myocardial infarction.[107] This experiment has contributed to the debate as to whether HDL directly contributes to, or is only a marker of, CV risk.

In the same publication, a clear association between genetic variants for LDL levels and myocardial infarction was identified, providing further support for the causal relationship between LDL levels and CV risk.

KEY POINTS

- Monogenic disorders of lipoprotein metabolism affect structural proteins in the lipoproteins or enzymes or receptors involved in their metabolism. They can cause major physiological disruptions (e.g. steatorrhea pr pancreatitis) or be associated with dramatically increased risks of atherosclerotic cardiovascular disease. Polygenic lipid disorders are much more common and the lipid phenotypes are usually less severe than those seen in the monogenic syndromes, but can overlap the latter.

DIAGNOSIS OF LIPID DISORDERS

The diagnosis of a lipid disorder should be based on more than one measurement of serum lipids, because combined analytic and biological variations in serum lipids range from 10% to 20%. The technology for measuring LDL-cholesterol levels directly has improved steadily over the years, but in most laboratories it remains a calculated value. To perform this calculation, the total and HDL-cholesterol levels, as well as the triglyceride value, are measured. The LDL-cholesterol concentration is then estimated using a formula first devised by Friedewald et al., which has subsequently undergone several modifications by later groups. The original formula was: LDL-cholesterol = total cholesterol – [HDL-cholesterol + triglyceride/5].

The triglyceride/5 factor represents an estimate of VLDL cholesterol. The validity of this formula for estimating LDL cholesterol has been confirmed by ultracentrifugal measurement of lipoprotein levels and remains reasonably accurate as long as the total triglyceride is less than 400 mg/dL. In order to obtain an accurate calculation, patients must fast for at least 12 hours to clear their blood of any chylomicrons; these lipoproteins distort the triglyceride ratio on which the Friedewald formula relies. If the triglyceride level is greater than 400 mg/dL, the LDL cholesterol must be determined by alternative methods, although a recent report suggested a modified calculation methodology may provide better LDL-cholesterol estimation at triglyceride values above 400 mg/dL than does the Friedewald formula.[108] Apo B_{100}, which is present in both LDL and VLDL, can be measured directly to get an assessment of lower-density lipoprotein particle numbers, though this entails additional expense. With increasing evidence for a greater atherogenicity of smaller, denser LDL particles, more sophisticated assessments of LDL number and composition are being introduced into clinical practice, such as nuclear magnetic resonance spectroscopy. The place of these more sophisticated and expensive assays in the routine diagnostic evaluation of most patients remains unsettled. It appears that most patients at high risk for CHD can still be identified using traditional lower-cost laboratory assays.

Before embarking on a treatment plan in a patient with hyperlipidemia, one must exclude other medical conditions that cause lipids to rise as a secondary consequence. The most common clinical conditions that cause this to occur are obesity, diabetes, and hypothyroidism. The latter two are best screened using a serum glucose (or HbA1c) measurement and a thyrotropin-stimulating hormone level, respectively. Many medications commonly cause a secondary hyperlipidemia, with antiretroviral therapy, a variety of psychotropic drugs, estrogens, and glucocorticoids heading the list.

Increasingly, the diagnosis of severe hyperlipidemic syndromes due to rare mutations in key proteins of the lipid metabolism system can be made by genetic testing. Using buccal smears or circulating white blood cells present in a blood sample, DNA can be extracted and mutations easily detected. At the present time, these assays are typically used to guide family counseling or to provide interested patients insights into the cause of their disorder, but the results rarely affect clinical decision-making or therapeutic choices. While many of the genetic assays have moved from research tools to readily obtainable clinical tests, few are covered by medical insurance in the United States. It seems likely that the declining costs of these tests and the potential for complex genetics to offer better predictive power in assessing CV risk will shift this dynamic toward greater utility in the near future.

KEY POINTS

- Although a wide array of specialty assays for lipoprotein numbers and size are available clinically, the use of the standard lipid profile which measures the total cholesterol, HDL cholesterol, and triglyceride levels in the serum and calculates the LDL cholesterol from those numbers remains the standard for routine evaluation of a patient's lipid status.

MANAGEMENT/TREATMENT

Apart from the patient at risk for hypertriglyceridemic pancreatitis, for most hyperlipidemic patients the primary goal of treatment is to reduce the risk of atherosclerotic vascular morbidity and mortality. The approach to the treatment of hyperlipidemia that dominated clinical practice in the United States for the 25 years following the approval of the first statin in 1987 was initially delineated by the National Cholesterol Education Program (NCEP) Adult Treatment Panel I in 1988. All of the Adult Treatment Panel guidelines focused on LDL-cholesterol as the major lipid risk promoting CHD, and therapy was intended to achieve specific target LDL cholesterol levels whose selections were based on an individual's overall CV risk profile. For a given degree of LDL-cholesterol elevation, the threshold for initiation of therapy decreased and the intensity of therapy increased with increasing global CHD risk. Dietary modification, complemented by exercise and weight reduction, was to be combined with appropriate pharmacologic therapy, when needed, to achieve the targets identified in the guidelines. In the fall of 2013, a joint American College of Cardiology/American Heart Association (ACC/AHA) expert panel issued the successor guidelines to the Adult Treatment Panel series.[109] These guidelines sparked considerable controversy in the lipid field, with many dissenting voices raised over the wisdom of moving away from specific LDL targets as the goal of therapy. Others criticized the accuracy of the specific risk calculator instrument favored by these guidelines to determine if the 10-year risk was greater than or equal to 7.5%, suggesting it led to overtreatment of many individuals whose LDL-cholesterol values were less than 100 mg/dL.[110]

An update to the ACC/AHA guidelines was subsequently published.[111] The 2018 ACC/AHA Guideline on the Management of Blood Cholesterol provides evidence-based recommendations for how to safely and effectively lower LDL-cholesterol levels and improve CV outcomes.[111] The Guideline's focus is 2-fold: first, it provides recommendations for how to estimate a patient's risk of developing atherosclerotic CV disease (ASCVD). Second, it suggests courses of treatment and management based on that estimated risk. The ACC/AHA continues to recommend the use of the pooled cohort equation (PCE), which takes into account cholesterol levels as well as age, smoking history, blood pressure, and diabetes status. The ACC/AHA delineates four levels of risk based on PCE results: low (under 5%), borderline (5%–7.49%), intermediate (7.5%–19.99%), and high (20% or higher).

Both medication- and lifestyle-based treatment options exist for patients with elevated cholesterol levels. Importantly, the ACC/AHA emphasizes that PCE predictions alone should not determine whether a patient receives medications to reduce cholesterol levels. Factors including the response of a patient's cholesterol levels to behavioral changes, as well as the financial burden of the medications, should be considered as patients work with their physicians to determine the best course of treatment. Among individuals who do not have ASCVD, the ACC/AHA highlights three populations for whom medications may be appropriate to reduce LDL-cholesterol levels: individuals whose LDL-cholesterol levels are greater than or equal to 190 mg/dL, patients with diabetes, and individuals between 40 and 75 years old whose LDL-cholesterol levels are at or above 70 mg/dL and whose ASCVD risk, as calculated using the PCE, is a minimum of 7.5%. In the ladder group, calculation of a coronary artery calcium (CAC) score may help confirm the chosen course of treatment. The ACC/AHA 2018 guidelines reference studies published between 2015 and 2017 to suggest that medications may not be necessary for patients whose CAC score is equal to 0, as these patients are unlikely to develop ASCVD within 10 years. The guidelines also recommend that medications (statins) be considered for patients with CAC scores of 100 Agaston units and above, lowered from the 300 Agaston unit minimum recommended in the 2013 guidelines.[109]

Statins are recommended as the starting point of treatment for all three populations without ASCVD. Patients on "moderate" and "low-intensity" treatment regimens are expected to experience LDL-cholesterol reductions of 30% to 49% and up to 30%, respectively. Patients for whom statin treatment is infeasible at the dosage corresponding to their risk level may consider cholesterol absorption inhibitors, such as ezetimibe, or bile acid sequestrants. For patients with ASCVD, the ACC/AHA suggests "high-intensity" statin treatment, particularly for those under 75 years old. Individuals at "very high risk" may benefit from the addition of ezetimibe, known to reduce LDL-cholesterol levels approximately 18% more than statins alone, an approach with documented outcomes benefit compared with monotherapy with the same statin regimen.[112] For "very high risk" patients whose LDL-cholesterol levels remain at or above 70 mg/dL while taking statins and ezetimibe, the 2018 guidelines recommend that PCSK9 inhibitors be considered. Two recent clinical trials, FOURIER and ODYSSEY, found that the PCSK9 inhibitors evolocumab and alirocumab, respectively, can lower LDL-cholesterol levels 43% to 64% more than statins alone and can improve outcomes in individuals with preexisting CV conditions.[113,114] These data are reflected in ACC/AHA's 2018 guideline recommendations that support the use of "combination therapy" for aggressive lowering LDL-cholesterol to reduce future ASCVD events. In addition, and in contrast with the 2013 guidelines, the new guidelines recommend the use of statins at "moderate intensity" in patients who are likely to live at least 3 to 5 years who have heart failure. This recommendation stems from the results of two smaller studies on the relationship between statin administration to individuals with heart failure and CV outcomes.[115]

While medications offer one useful approach to managing cholesterol levels and ASCVD outcomes, modifications to daily dietary and exercise habits are also often effective, either on their own or in conjunction with drug treatments. A metaanalysis reported in 2018 found an association between consumption of a diet rich in plant products and a fall in LDL-cholesterol levels of approximately 17%.[116] This decrease in LDL levels corresponded with a decline of approximately 13% in the likelihood of developing CHD within 10 years. The ACC/AHA recommends intake of a vegetable-rich, sugar- and saturated fat–poor diet (for example, the Dietary Approaches to Stop Hypertension diet), as well as frequent aerobic exercise (120–160 min weekly).[109,117] Lifestyle-based approaches to managing cholesterol levels are especially important in patients under 40 years, as fewer data are available regarding CV outcomes using drug therapies in this age group. The risk-benefit calculus for utilizing preventive lipid therapies early in adulthood for individuals with average risk for CV disease is a hotly debated topic that is not informed by actual outcome data. Extrapolating benefit using genetic risk markers and short-term studies in older individuals allows one to create a compelling argument for aggressive early preventive therapy, but the realities of medication and management costs, modest drug side effects, and poor compliance with long-term preventive medications have limited the uptake of this approach in most medical practices.

KEY POINTS

- Optimum treatment of lipid disorders typically requires a combination of lifestyle approaches involving diet and activity improvements that may require adjunctive medications. Statins are the mainstay of drug treatment for those with LDL cholesterol elevations and can work well even in the patient who does not modify their lifestyle very much. Disorders of triglyceride-rich lipoproteins frequently cannot be controlled by medications alone and require comprehensive metabolic improvements to achieve normal lipid levels.

Assessing the Need to Treat Patients With Hyperlipidemia

As the magnitude of the benefit of treating hyperlipidemia is positively correlated with the overall degree of CHD risk, it is important to assess that risk before initiating treatment. Assessment should be comprehensive, extending beyond lipid levels to include consideration of blood pressure, smoking, diabetes, family history of premature CHD, age, sex, and presence of established CHD or other atherosclerotic disease. Although still in its early days, it is likely that the growing list of polygenic risk factors associated with atherosclerotic vascular disease will be incorporated into future risk-stratification schema.[103,118-121]

Treatment recommendations follow directly from the degree of estimated CHD risk. Dietary modification is the sole mode of therapy for patients at the lower end of the CHD risk spectrum, while pharmacologic measures are reserved for patients at higher risk or for those who fail dietary intervention. The trend in preventive care over the past 20 years has been to push LDL levels ever lower, with no evidence to date that a value has been reached below which no further benefit is conferred. For the highest-risk patients, the current LDL-cholesterol target is to be under 70 mg/dL, although recent outcomes studies have demonstrated LDL-cholesterol values in the 50 mg/dL range or lower are associated with improved outcomes over those of approximately 70 mg/dL.[113,114,122]

The concept that diabetes is a risk equivalent for CHD derives from the high risk of CHD in middle-aged and older diabetic populations and was exemplified in the East-West study of Finnish patients.[123] In that study, patients with prior histories of myocardial infarction and no diabetes had a 7-year incidence of new CHD events of 18.8%, whereas diabetic patients without prior myocardial infarctions had an incidence of 20.2%. The hazard ratio for death from CHD comparing these two groups did not differ. A follow-up report by the same investigators, extending to 18 years of observation and with more CHD events recorded, reaffirmed this outcome and suggested that female diabetics were at particularly high risk.[124] A finding of similar degrees of carotid intimal medial thickness in US patients with diabetes and no CHD compared with patients with CHD but no diabetes further strengthened the concept of diabetes as a CHD equivalent.[125] Several other studies in different populations have supported this conclusion.[126-130] However, not all investigations of this topic have yielded similar conclusions, with some studies indicating lower rates of future CHD events in the diabetic population without CHD than are found in nondiabetics with CHD. Other studies have indicated that gender substantially influences this comparative outcome.[131-138] What all studies demonstrate, however, is that diabetes does confer a substantial increase in CHD risk, and that the combination of diabetes and CHD puts individuals at a very high risk of a future coronary event. Treating diabetes as a CHD risk equivalent in middle-aged and older patients does, therefore, appear to be clinically appropriate.

Other commonly assessed risk factors not included in most guideline calculations of risk include obesity, impaired fasting glucose, markers of inflammation (e.g., high sensitivity C-reactive protein), homocysteine, endothelial dysfunction, and a thrombosis predilection. The role of these other risk factors in improving prognosis remains controversial, but clinicians may want to modulate the standard risk-factor treatment guideline based on the presence or absence of these nontraditional risks in their individual patients.

Dietary and Drug Treatment of Lipid Disorders

Diet/Lifestyle Therapies. With the advent of very potent lipid-lowering therapies, the use of dietary approaches to controlling lipid levels has become deemphasized in much of the developed world. While drug treatments (especially for elevated LDL-cholesterol levels) are very effective, the role of diet in promoting hyperlipidemia should not be forgotten. The dietary approach to the treatment of elevated LDL-cholesterol levels focuses on reductions in saturated fat, partially hydrogenated unsaturated fatty acids, and dietary cholesterol. Substituting foods that provide polyunsaturated and monounsaturated fats in place of saturated and *trans* unsaturated fat is particularly important, whereas the value of reducing the total fat intake is less clear.[139] The 2013 clinical practice guidelines of the American Diabetes Association (ADA) recommend that saturated fat intake be restricted to less than 7% of calories, and that cholesterol intake not exceed 200 mg per day in patients with diabetes. These dietary recommendations mimic those proposed by the NCEP Adult Treatment Panel III guidelines for individuals without diabetes who have established coronary disease. For nondiabetics without coronary disease, the NCEP liberalizes saturated fat intake to less than 10% of calories and cholesterol to less than 300 mg/d. In both diets, it is recommended that *trans* fat intake be kept to a minimum. The adoption of a Mediterranean-style diet, with substitution of monounsaturated fats for saturated fat, has demonstrated benefits on LDL cholesterol reduction, insulin sensitivity, and endothelial reactivity.[11,140,141] The magnitude of the LDL reduction achieved with dietary interventions is quite variable and depends on the preintervention diet, as well as metabolic and genetic factors that influence diet-dependent lipid responses. LDL reductions of 5% to 20% are typical for most patients adopting the reduced fat and cholesterol diets recommended by the ADA and NCEP, but reductions of over 50% can occur in selected individuals.[142]

For patients with VLDL elevations, which include a large portion of the diabetic population, carbohydrate and alcohol consumption may be more important dietary factors to address than the intake of cholesterol or saturated fat. The current United States Department of Agriculture–recommended daily allowance of carbohydrate is 130 g/d, but studies of long-term carbohydrate-restricted diets are limited. Recently, investigations comparing low-carbohydrate and low-fat diets have demonstrated that low-carbohydrate diets can produce as good or even modestly better weight reduction than can low-fat diets in 6- to 12-month timeframes.[143-145] The lower-carbohydrate, higher-fat diets typically produce greater reductions in VLDL (serum triglycerides) and greater increases in HDL-cholesterol than do the fat-restricted, higher-carbohydrate diets. On the low-carbohydrate, higher-fat diets, the LDL-cholesterol level can vary significantly, and if a substantial amount of saturated fat is used in these diets, the rise in LDL could be substantial.[146] Therefore, the use of low-carbohydrate diets should be accompanied by the use of monounsaturated and polyunsaturated fats as the major sources of fat in the diet.

Weight loss (if obese), aerobic exercise, and smoking cessation can increase HDL levels and contribute to the dietary lowering of lower-density lipoproteins and CHD risk. They also reduce CHD risk by decreasing blood pressure. In the Diabetes Prevention Program (DPP), subjects in the lifestyle intervention group reduced their fat intake to 28% of calories after 1 year (down from 34%), and most were able to maintain the goal of 150 minutes per week of moderate physical activity.[147,148] On this program, in which a body weight loss of 7% was targeted, subjects experienced less progression to diabetes and reductions in multiple CV risk factors, including dyslipidemia and hypertension.[149,150] Although the DPP was conducted in an academic environment with substantial investigative resources devoted to ensuring compliance with the program, more recent efforts modeled on DPP suggest that the results can be translated to community-based outreach programs.[151,152]

The clinical trials data that demonstrate improvements in CAD outcomes as a result of dietary interventions alone have generally employed diets that restrict fat and/or cholesterol intake much more dramatically than the ADA and NCEP diets mentioned earlier. In the St. Thomas' Atherosclerosis Regression Study (STARS), cholesterol intake in men was reduced to 100 to 120 mg/d, and excess weight was addressed by prescribing a physical activity program. After 3 years, the group treated with a low-cholesterol diet had a slower rate of progression of coronary disease and a higher rate of regression as determined by angiography.[153] The Lifestyle Heart Trial employed an even more stringent diet, reducing cholesterol intake to less than 10 mg/d, combined with an exercise and behavioral modification program. This intervention led to regression of coronary lesions as measured by angiography and a reduction in symptoms of ischemia.[154] This study was conducted in a very small number of highly motivated volunteers; it would be challenging to replicate the same interventions in more typical clinical care settings. Nevertheless, a longer-term follow-up of a small number of participants in this study demonstrated that further regression of coronary atherosclerosis could be obtained.[155] A study with potentially greater applicability than either STARS or the Lifestyle Heart Trial is the Lyon Diet Heart Study.[156-158] Conducted using a Mediterranean diet rich in alpha-linolenic acid, with participants who had experienced a prior myocardial infarction, this trial demonstrated improvements in its primary endpoint (myocardial infarction and death) as well as several CV and secondary endpoints. These results have been validated in the Spanish PREDIMED study of 7447 individuals with either type 2 diabetes mellitus or at least three major CV risk factors who were randomized to a Mediterranean or control diet.[159] Participants in the Mediterranean diet groups experienced a 30% relative reduction in CV events, which was driven by a decrease in stroke rates. The Lyon Diet Heart Study and PREDIMED results, combined with other investigations of the impact of similar diets on lipids and lipoproteins, provide substantial support for the utility of Mediterranean-style diets in the prevention and treatment of CHD.

Outcome Studies of Pharmacologic Therapy for Hyperlipidemia

The addition of drug therapy to a diet and exercise program can greatly enhance lipid-lowering results and dramatically effect CHD outcomes. Numerous investigations over the past 30 years have demonstrated that the use of one class of drugs, the HMG CoA reductase inhibitors, or statins, leads to significant reductions in nonfatal and fatal cardiac events (i.e., myocardial infarction, revascularization, cardiac death) in appropriately selected populations. Reductions in all-cause mortality have also been demonstrated in lipid-lowering drug trials, particularly in higher-risk populations. With intensive drug therapy, the rate of plaque progression falls, and modest plaque regression can be demonstrated in major coronary vessels and systemic arteries. There is also evidence that lipid-lowering medication (statins) can reduce the risk of stroke in persons with atherosclerotic carotid disease. These benefits have been shown to accrue to men and women, diabetics and nondiabetics, and middle-aged and older individuals. A summary of the major outcomes studies using statins is provided in the coming paragraphs, followed by a brief review of the data available for outcomes using drugs that are not HMG CoA reductase inhibitors.

KEY POINTS

- The value of lipid-lowering in improving cardiovascular outcomes has been one of medicine's most intensely investigated topics. Statins and other more recently approved LDL cholesterol lowering drugs, when given to patients with clinically documented ASCVD and elevated LDL cholesterol levels, have proven benefits. For those with lower risks for future ASCVD events the benefit of preventive lipid therapy is not as great in the short-term but may still be valuable over longer timeframes.

Coronary Heart Disease Outcomes with 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Inhibitors (Statins).

The current era of lipid-lowering therapy began in 1984 with the publication of the Lipid Research Clinics Coronary Prevention Trial. A randomized, double-blind study employing cholestyramine compared with placebo, this trial demonstrated that an 8% reduction in total cholesterol yielded a 19% reduction in CHD death or myocardial infarction.[160,161] With the advent of statin therapy in 1987, much more effective LDL-cholesterol–lowering agents became available. During the 1990s, a series of landmark CV prevention trials were completed using these agents. In 1994, the Scandinavian Simvastatin Survival Study (4S) established for the first time that a lipid-lowering agent could prolong overall survival in individuals with preexisting CHD.[162] Randomizing 4444 patients with a prior myocardial infarction or angina to simvastatin or placebo, the drug treatment cohort of 4S had baseline LDL-cholesterol levels fall from approximately 190 mg/dL to 120 mg/dL. With a trial endpoint of death for 10% of the original study enrollees (444 study subjects), it took a little over 5 years to complete the study. The result was that 8% of the simvastatin patients died versus 12% of the placebo-treated individuals. This 30% reduction in mortality established statin therapy as an essential intervention in CHD patients with significant LDL elevations. In addition to the mortality benefit, 4S also produced substantial reductions in major coronary events, overall CHD death (19% vs. 28%, a relative risk reduction of 42%), and cerebrovascular events (2.7% vs. 4.3%). Unlike several of the pre–statin era cholesterol intervention studies, there were no increases in other causes of mortality to counterbalance the benefits on the CV system. When 4S was examined in greater detail, and after longer follow-up to determine if the therapy had been confined to any subset of patients, these analyses concluded that the benefit was widespread.[163-165] Individuals at the highest risk appeared to benefit the most, particularly those with diabetes.[166] Though only 202 of the 4444 patients in 4S were diagnosed with diabetes, these individuals experienced relative risk reductions in mortality (43%) and major CHD events (55%) that exceeded those in the nondiabetic population. An equally impressive outcome was noted in individuals who, in addition to their elevated LDL-cholesterol levels, had lipid phenotypes characteristic of the metabolic syndrome (low HDL-cholesterol levels and high triglycerides). Study subjects with this lipid triad had the highest event rates among those receiving placebo (35.9%) and the greatest reduction in relative risk with treatment (0.48; 95% CI: 0.33–0.69).[167]

Subsequent studies have further refined our understanding of the benefits of statins in patients with and without preexisting CHD. The West of Scotland Coronary Prevention Group study (WESCOPS) enrolled hyperlipidemic men who had not experienced a prior myocardial infarct but who had high levels of other CHD risk factors. Pravastatin reduced LDL-cholesterol levels from approximately 190 mg/dL to 140 mg/dL and, compared with the placebo-treated group, lowered CHD death and myocardial infarction by 30%.[168]

An improvement of similar magnitude in a younger population of lower-risk patients was observed in the Air Force/Texas Coronary Atherosclerosis Prevention Study when lovastatin was given to men and women with starting LDL-cholesterol levels of 150 mg/dL.[169] When compared with the placebo-treated patients, those receiving lovastatin reduced their LDL on average to 115 mg/dL and decreased their incidence of major coronary events by 30% to 40%. In the Cholesterol and Recurrent Events Trial (CARE), pravastatin was given to patients with myocardial infarctions whose LDL-cholesterol levels were substantially below (mean of 139 mg/dL) those of the 4S and WESCOPS patients. With a primary endpoint of a fatal coronary event or a nonfatal myocardial infarction, the pravastatin-treated group had a 24% relative risk reduction compared with placebo-treated patients. Stroke was reduced 31%.[170] A *post hoc* analysis of CARE by the investigators who performed it led to the conclusion that reduction of LDL-cholesterol in the range from 174 mg/dL to 125 mg/dL lowered the coronary event rate, but that further reductions below 125 mg/dL did not appear to confer additional benefit.[171] The value of LDL reductions to levels substantially below 125 mg/dL has been a topic of considerable controversy ever since CARE was published, but several subsequent studies have indicated that further reductions of LDL do confer benefit. For example, the Long-Term Intervention with Pravastatin in Ischaemic

Disease study group enrolled over 9000 patients with recent evidence of CHD and demonstrated benefits in overall mortality and CHD death using pravastatin.[172] When stratified by starting LDL-cholesterol, the magnitude of the benefit appeared to be maintained even in those with starting levels below 116 mg/dL.[173-175] The Treating to New Targets trial enrolled 10,000 patients with stable CHD and randomized treatment to high-dose (80 mg/d) or low-dose (10 mg/d) atorvastatin.[176] Pretreatment LDL-cholesterol levels were between 130 mg/dL and 250 mg/dL, and posttreatment LDL-cholesterol values were 77 mg/dL and 101 mg/dL, respectively, in the high- and low-dose groups. The primary endpoint of the study, which was major CV events, declined 22% more in the high-dose group than the low-dose cohort, as did rates of myocardial infarction and stroke. However, overall mortality was not lower in the high-dose group, owing to an increase in non-CV deaths. Cancer deaths accounted for most of these events, a finding not replicated in other studies using the same drug.

The Heart Protection Study (HPS), like several of its predecessor trials, attempted to examine the question of LDL-cholesterol lowering in patients with moderate hyperlipidemia. HPS was a much bigger trial, randomizing 20,536 patients to placebo or simvastatin (up to 40 mg/d). The inclusion criteria required evidence of prior CV disease and permitted a broad range of LDL-cholesterol levels, with a third of the patients having values less than 116 mg/dL. After an average follow-up of 5.5 years, all-cause mortality was reduced in the statin-treated cohort, as was CHD death and major CV events.[177] Notably, over 5000 diabetics were included in the study, and they also achieved a substantial relative risk reduction in myocardial infarction and stroke.[178,179] Particularly striking was the finding that the benefits of LDL-cholesterol lowering were produced in all three of the tertiles into which patients' baseline LDL-cholesterol were stratified. Even those with starting LDL-cholesterol levels approximating the current NCEP target goal of 100 mg/dL (for those with CHD) had a benefit when their value was reduced substantially below that number. Both men and women, as well as older and younger patients, derived similar benefits from the statin treatment. As with HPS, the Collaborative Atorvastatin Diabetes Study demonstrated a significant benefit in diabetic patients with lower LDL-cholesterol levels (<160 mg/dL) randomized to the statin arm of a placebo-controlled trial.[180] This benefit appeared in those whose starting LDL-cholesterol levels were both above and below 120 mg/dL. After 4 years of follow-up, the rate of CV events dropped 37% (95% CI: –52% to –17%).

The JUPITER study attempted to answer whether LDL lowering with rosuvastatin was beneficial for healthy individuals whose LDL levels were considered normal but who had elevated C-reactive protein (CRP) levels.[181] Over 17,000 men and women with LDL less than 130 mg/dL and high-sensitivity CRP values of 2.0 mg/L or higher were randomized to rosuvastatin 20 mg daily or placebo and followed for CV events. The trial was stopped after 1.9 years; at that time, LDL levels were reduced by 50% (to 55 mg/dL), high-sensitivity CRP levels by 37%, and CV events by 44% in the rosuvastatin group. This study supported the hypothesis that LDL lowering below the targets recommended at that time could be beneficial for a subset of apparently healthy individuals who have evidence of atheroinflammation, as indicated by the high-sensitivity CRP assay. In addition to studies of their effects on major clinical outcomes, many of the statins have been used in trials assessing their effects on atherosclerosis progression or regression in the coronary arteries or the carotids. The Familial Atherosclerosis Treatment Study compared three regimens that all contained a bile acid resin binder, colestipol, plus placebo, niacin, or lovastatin. The lovastatin-treated cohort lowered its LDL by 46% and had more regression and less progression of their CAD, as assessed by angiography, than did the placebo-treated group. The niacin group's response was equivalent to that seen with the lovastatin treatment.[182] The Regression Growth Evaluation Statin Study utilized pravastatin in over 800 men with total cholesterol levels between 155 mg/dL and 310 mg/dL to demonstrate delayed progression in CHD, as well as a lower rate of progression of femoral and carotid atherosclerosis.[183,184] The REVERSAL trial compared atorvastatin (80 mg/d) to pravastatin (40 mg/d) in patients with CHD by assessing plaque volume using intravascular ultrasound measurements (IVUS). Atheroma volume declined slightly in the atorvastatin-treated group, while increasing in the pravastatin cohort, and the difference between the two was statistically significant.[185] ASTEROID employed an open-label treatment with rosuvastatin (40 mg/d) in 507 patients with CHD to lower LDL-cholesterol by 53% (from a mean of 130 mg/dL to 61 mg/dL). With this intervention, atheroma volume, as measured by IVUS, was reduced. No control group was included in ASTEROID, making it difficult to assess what is otherwise an impressive reduction in coronary atherosclerosis.[186] Taken together, these statin studies provide impressive evidence that aggressive lipid lowering can reduce the rate of CAD progression in patients with established disease and may, with the most intensive therapy, even cause plaque regression.

In addition to their use in studies of patients with longstanding stable CHD, statins have been employed in the setting of acute or recent coronary syndromes to determine if more aggressive lipid lowering might provide additional benefit in these settings. The Prove-It study, for example, compared individuals with starting LDL-cholesterol levels of approximately 110 mg/dL who were randomized to receive pravastatin (40 mg/d) or atorvastatin (80 mg/d) after an acute coronary event.[112] The high-dose atorvastatin group lowered LDL-cholesterol levels to 62 mg/dL, whereas the middose pravastatin cohort only reduced their LDL-cholesterol values to 95 mg/dL. A 16% reduction in coronary events was achieved on the more aggressive therapy. It seems probable that this study established that an LDL-cholesterol of 62 mg/dL is better than one of 95 mg/dL in patients experiencing an acute coronary event, but the more rigorous conclusion is that 80 mg of atorvastatin works better for CHD outcomes in this population than does 40 mg of pravastatin. Similarly, the Incremental Decrease in End Points through Aggressive Lipid-lowering study used an open-label/blinded-endpoint study design to randomize nearly 9000 patients with a past history of myocardial infarction to receive either atorvastatin (80 mg/d) or simvastatin (20 mg/d).[187] After 24 weeks, the dose of simvastatin could be raised to 40 mg/d if the LDL-cholesterol remained greater than 190 mg/dL. The atorvastatin-treated cohort after nearly 5 years had a mean LDL-cholesterol that was lower than the simvastatin-treated cohort (81 mg/dL vs. 104 mg/dL), but no significant reduction in the primary endpoint of major coronary events was demonstrated. While there were reductions in some secondary endpoints on atorvastatin, no benefit on overall mortality was seen. Because several of the above studies employed different drugs and achieved different LDL endpoints using those agents, it is not always easy to interpret the outcomes. Overall, the data do support aggressive lipid-lowering therapy in the acute coronary setting, but it is much less clear whether one drug is superior to another when nearly comparable LDL-cholesterol lowering is achieved.

In summary, the outcomes studies presented in this section demonstrate that LDL-cholesterol reduction with multiple members of the statin class of drugs results in improvements in CHD events and in overall total mortality in patients with established coronary disease. These benefits are present or even magnified in individuals with diabetes or the metabolic syndrome. Lower rates of CHD events are also produced in patients without established CHD but who have substantial

CHD risk factors. In lower-risk populations, statins clearly reduce coronary events, but data are less convincing for benefits on CHD death or overall mortality. These conclusions have been confirmed by several metaanalyses of clinical trials that employed a variety of statins.[188-190] Although the precise goal of LDL-cholesterol reduction remains controversial, the NCEP guideline targets of progressively lower target LDLs for progressively higher levels of risk (i.e., <130, <100, and <70 mg/dL) are both reasonable and consistent with the findings of outcomes studies published to date.

Concerns about cholesterol-lowering producing increased risks of cancer and violent deaths, which emerged in some of the early, prestatin lipid-lowering trials, have largely been put to rest in the major prospective statin studies. In studies of over 25 randomized control trials, including five of the largest such trials conducted to date (using several different members of the statin class of drugs and collectively involving ~90,000–175,00 patients), no increase in the rates of cancer or noncardiac deaths have been reproducibly observed.[191,192] However, each class of cholesterol-lowering drug therapies has important adverse effects that need to be taken into account when considering pharmacologic intervention. With the trend towards more aggressive lowering of LDL-cholesterol, achieved through higher doses of statins and more complex multidrug regimens, comes an increase in the risk of drug side effects.

Another concern that has emerged recently pertains to the effect of statins on diabetes incidence. In addition to demonstrating clear evidence for CV benefit from LDL reduction, the JUPITER trial provided evidence for a 25% increase in new physician-diagnosed diabetes during statin treatment.[181] Following the publication of the JUPITER trial, two metaanalyses of large, placebo-controlled trials were rapidly performed.[193,194] Both of these studies reported a small but consistent increase in diabetes incidence (9%–13%) during statin treatment. These findings suggest that treating 255 patients with statins for 4 years would result in one additional case of diabetes. Although not directly comparable, a prior metaanalysis had concluded that five myocardial infarctions could be prevented for every case of diabetes occurred.[195] Thus, based on the present data, the benefits of statin therapy appear to clearly outweigh the small risk of increased diabetes incidence.

Whether statin therapy should be initiated in young individuals is a topic of debate, with little data available to inform the discussion. Young men (aged less than 35 years) and premenopausal women with no CHD risk factors other than hypercholesterolemia are typically at very low 10-year risks for developing CHD. An argument can be made to wait until the risk level escalates before initiating cholesterol-lowering therapies, inasmuch as the benefits of statin therapy are typically evident in less than 2 to 3 years. An alternative view cites the evidence that individuals with genetically-driven low LDL-cholesterol levels have extremely low rates of CHD, and that one should try to mimic this lifelong low-cholesterol state with earlier and more aggressive drug therapy.[88] Both perspectives have merit. For now, making a decision on intervention in younger patients will likely balance the magnitude of the individual risk with preferences for more or less aggressive therapy on the part of both physician and patient. Physicians prescribing statins for young women, however, must not forget the potential teratogenicity of this class of drugs, which remains contraindicated in pregnancy. Nonpharmacologic therapy, primarily focusing on improved diets and physical activity programs, should be adopted regardless of age. As widespread clinic use of statins moves into its fourth decade and the initially approved agents have become available in generic forms, the safety and cost issues associated with their use may recede in importance, and a greater consensus on earlier use of these remarkable drugs becomes possible. For now, the exercise of good clinical judgment is essential in these circumstances.

Proprotein Convertase Subtilisin Kexin 9 Inhibitors and Cardiovascular Benefit. The most effective LDL-lowering agents for patients who are either unable to achieve target LDL goals on statins or who are unable to tolerate statin use at all are the monoclonal antibody therapies directed against PCSK9, alirocumab and evolocumab. Preapproval studies showed excellent lipid-lowering efficacy as monotherapy and in combination with statins.[196-204] Evolocumab was found to lower LDL-cholesterol levels by 59% and the incidence of myocardial infarction, stroke, and coronary revascularization by 21% to 27% in a 20,564-participant clinical trial called FOURIER.[113] All study participants had been diagnosed previously with ASCVD and had LDL-cholesterol levels above guideline targets despite statin treatment. ODYSSEY, an 18,924-participant study investigating the use of alirocumab in patients with a history of recent acute coronary syndrome who were unable to achieve optimum LDL-cholesterol levels on statins, demonstrated an LDL-cholesterol reduction of approximately 55% to 60%. This reduction led to a significant benefit on a composite of CV outcomes, as well as a modest overall mortality benefit.[114]

Cardiovascular Outcomes Trials With Agents Other Than Statins and Proprotein Convertase Subtilisin Kexin 9 Inhibitors. As cited in the previous section, the current era of lipid lowering was successfully initiated with a bile acid resin binder, cholestyramine.[160,161] Though achieving an LDL-cholesterol reduction of only 8%, CHD events were reduced 19%. Thus, it is the belief of most experts in the lipid field that LDL-cholesterol lowering achieved by virtually any means will result in reduced CHD events, provided the therapeutic approach does not have off-target consequences that would countervail the benefit of the LDL lowering. This view is supported by a wealth of data on the pathogenesis of atherosclerosis, as well as clinical trials of multiple modalities that lower LDL-cholesterol, including nonpharmacologic approaches.[155,205-208] At the current time, statin use supersedes that of all other LDL-cholesterol–lowering agents as first-line monotherapy because of outstanding efficacy, low cost, and the wealth of outcomes data establishing its benefit. Many patients, however, either cannot meet the LDL targets set by guidelines or cannot achieve the levels obtained in the major outcomes trials using statins alone. Some patients have inadequate therapeutic responses to the drugs, while others cannot tolerate their adverse side effects. When that occurs, the use of other lipid-lowering agents becomes relevant. If the patient has documented ASCVD, PCSK9 antibodies can usually be added or used instead of statins, though cost remains a barrier to that approach for many patients due to drug insurance gaps or expensive copay payments. The following section addresses the current data on outcomes when therapeutics other than statins or PCSK9 inhibitors need to be considered.

Fibrates and Coronary Heart Disease Outcomes. Unlike statin therapy, none of the other oral LDL-cholesterol–lowering medications have produced an overall mortality benefit in a randomized, prospective, double-blind trial. For the most part, this is because studies employing those agents were not powered for a primary outcome focused on mortality, but instead targeted reductions in CHD events. With the latter endpoint, bile acid resins, gemfibrozil, and niacin have all been demonstrated in long-term prospective studies to reduce CHD and/or major CV events.[160,161,209-213] Although not available in the United States, bezafibrate also showed promising results in a secondary prevention study.[214] Because fibrates (gemfibrozil, fenofibrate, bezafibrate, clofibrate, and ciprofibrate) are more effective than statins at lowering serum triglyceride/VLDL levels, they would appear to be particularly attractive agents for the treatment of the dyslipidemia associated with diabetes and the metabolic syndrome. Subgroup analyses of many studies employing fibrates support this hypothesis, with greater benefit shown in those with high triglycerides and low HDL.[215] Insofar as neither bezafibrate nor ciprofibrate are available in the United States, and clofibrate is no longer used because of a significantly higher mortality rate due to non-CV adverse events produced by the drug, these agents will not be discussed in this section.

Of the fibrates, then, only fenofibrate and gemfibrozil are prescribed in the United States. Gemfibrozil was used in two major trials, the Helsinki Heart Study (HHS) and the Veterans Affairs High-Density Lipoprotein Intervention Trial (VA-HIT), with both reporting substantial benefit from the drug. The HHS was a double-blind, placebo-controlled 5-year trial that targeted HDL-cholesterol–raising and non–HDL-cholesterol–lowering.[209] A group of 4081 asymptomatic men with non-HDL-cholesterol levels greater than 200 mg/dL were studied over a 5-year period, with approximately half receiving 600 mg twice daily of the fibrate and half on placebo. HDL-cholesterol, non–HDL-cholesterol, LDL-cholesterol, and triglycerides all improved in the drug-treated cohort, which experienced a 34% reduction in CHD events. On drug treatment, HDL-cholesterol rose 11%, LDL-cholesterol fell 10%, and triglycerides dropped 35%. Subsequent subgroup analysis of this study showed that the greatest benefits were achieved in those with body mass indexes above 26 kg/m^2 or in individuals who had higher triglyceride levels or lower HDL-cholesterol values.[216] In the VA-HIT study, 2531 men with established CHD, LDL-cholesterol levels less than or equal to 140 mg/dL and HDL-cholesterol levels less than 40 mg/dL were treated with gemfibrozil (1200 mg/d) or placebo.[210] Those who received the fibrate experienced a 22% reduction in the primary outcome of nonfatal myocardial infarction or CHD death. These benefits were associated with a 6% higher HDL-cholesterol level, a 31% lower serum triglyceride value, and no significant change in LDL-cholesterol. Subgroup analysis in this study again revealed that those benefiting the most from the drug were individuals whose HDL-cholesterol rose. Somewhat surprisingly, the improvement in triglyceride was not associated with an improved CHD outcome.[217] The serious adverse events seen with clofibrate were not reproduced in the gemfibrozil studies, though overall mortality benefits were also not demonstrated in either study.

The results of the HHS and VA-HIT studies convinced many lipid experts that fibrate treatment was beneficial, and these agents were widely prescribed from 1990 to 2010 to treat hypertriglyceridemia and low HDL-cholesterol levels. As gemfibrozil was found to affect the metabolism of multiple statins, many lipid experts switched to using fenofibrate whenever a statin/fibrate combination therapy was deemed appropriate in order to reduce the likelihood of a serious myopathy.[218] Thus, the publication of the Fenofibrate Intervention and Event Lowering in Diabetes Study (FIELD) in 2005 was a much-anticipated event in defining lipid therapeutics.[219] This trial randomized 9795 middle-aged and older patients with type 2 diabetes to placebo or micronized fenofibrate (200 mg/d). Patients qualified by having total cholesterol levels between 116 mg/dL and 252 mg/dL and total/HDL-cholesterol ratios greater than 4.0 or plasma triglycerides between 88 mg/dL and 435 mg/dL. An unusual feature of the FIELD trial was that patients were permitted to be treated with other lipid-lowering agents at their own physician's discretion, so many received simultaneous statin therapy. The primary outcome of the study was CHD events. Triglycerides were reduced by the fibrate by 29%, but HDL-cholesterol rose by only 2%. While CHD events were reduced in the drug-treated group by 11%, this overall outcome was not statistically significant. More worrisome was the fact that the overall event rate represented the average of a positive and negative outcome, with nonfatal myocardial infarction falling 24%, but CHD mortality rising by 19%. When total mortality was examined, it was not significantly higher in the fenofibrate cohort. Overall, the FIELD study was widely perceived to have failed to deliver the expected benefits suggested by the earlier gemfibrozil studies. This outcome has led to vigorous and protracted debates about the utility of fibrate therapy in general, and fenofibrate in particular.

The Action to Control Cardiovascular Risk in Diabetes trial was designed to investigate fenofibrate's value in a diabetic population, randomizing 5518 patients with type 2 diabetes to placebo or fenofibrate.[220] All patients were also receiving simvastatin as background therapy. In patients receiving fenofibrate, triglycerides were reduced by 14%, and HDL increased by 2%. However, LDL levels were unaffected by fibrate therapy, and total mortality and the primary endpoint of CV events were similar in the placebo and fenofibrate groups. Despite the epidemiologic evidence linking high serum triglycerides to CHD events, the data from fibrate intervention studies have not consistently shown that reductions in triglycerides result in CHD benefit. As pointed out earlier in VA-HIT, the effect on HDL-cholesterol levels, though modest, correlated with the desirable CHD outcomes, while the marked triglyceride reduction of 31% did not. Thus, the triglyceride-lowering effect of fibrates may be most clinically valuable when treating severely hypertriglyceridemic patients (serum triglycerides >500 mg/dL) in whom the goal is to prevent triglyceride-induced pancreatitis. Future studies with fibrates may clarify their place in the prevention of CHD, but at present the currently available agents do not appear to have a great enough therapeutic benefit in most CHD patients to warrant widespread use.

Other Classes of Drugs to Treat Hyperlipidemia. Three other classes of drugs are commonly used to treat patients with hyperlipidemia: (1) nicotinic acid or niacin, (2) cholesterol absorption inhibitors (bile acid resins and ezetimibe), and (3) omega-3 fatty acids or fish oils. The decrease in CHD produced with niacin in the Coronary Drug Project, combined with long-term follow-up of those originally assigned the drug, provide substantial confidence in the therapeutic value of using this B vitamin.[212,221] While there is evidence for benefit when niacin is combined with a statin,[222,223] the AIM-HIGH study raised new questions about which patients are likely to obtain a benefit from niacin.[224] In AIM-HIGH, 3414 patients with CV disease who were already on statin therapy were randomized to niacin or placebo. The niacin arm failed to show a benefit of the drug on CV event rates despite raising HDL-cholesterol and lowering triglycerides and LDL-cholesterol levels substantially better than the placebo treatment. It is important to remember, however, that niacin was being added to a regimen that had already reduced LDL-cholesterol levels to approximately 70 mg/dL, so the lack of efficacy in this group may not be relevant to populations whose pretreatment LDL values are substantially higher. Niacin can lower triglycerides, raise HDL-cholesterol, and reduce LDL-cholesterol, so it can provide a very useful adjunct to statin therapy in treating patients with combined lipid disorders. Its side-effect profile and its potential for increasing insulin resistance are the major drawbacks to employing niacin more widely.

Bile acid resins were the mainstays of lipid therapy prior to the advent of statins; multiple CHD outcomes and atherosclerosis progression studies were performed with various members of the class. Overall, despite only modest LDL-cholesterol–lowering performance, these drugs yielded benefits on CHD and coronary atherosclerosis when given as monotherapy or in combination with niacin or sta tins.[153,160,161,182,225] Because bile acid resins can interfere with the absorption of other drugs, cause substantial gastrointestinal bloating and constipation, and often substantially increase serum triglycerides, the development of much more selective cholesterol absorption inhibitors was highly desirable. Ezetimibe, which inhibits cholesterol absorption from the gut via blockage of enterocyte cholesterol trafficking pathways, rapidly replaced bile acid resins in the lipid-therapeutic armamentarium because of its ease of use. In its first major outcome study, the addition of ezetimibe to simvastatin provided no additional benefit beyond simvastatin alone on carotid intimal medical thickness assessed by carotid sonography.[226] This controversial study demonstrated substantial additional LDL-cholesterol reduction with ezetimibe (16.5% greater than the simvastatin alone), indicating that the drug was effective in its primary action in the cohort that received it. The short duration of treatment, the mild degree of carotid disease at the outset of the study, and the use of patients previously treated with statins for many years may have all combined to make it extremely challenging for this study to have shown a benefit. Subsequently, the IMPROVE-IT trial showed that the combination of simvastatin (40 mg) and ezetimibe (10 mg) reduced LDL-cholesterol further than the simvastatin alone and lowered CV events.[122] This trial provided evidence that lowering LDL-cholesterol, even with a nonstatin medication, can lead to a reduction in CV events.

Another major category of hypolipidemic agents, omega-3 fatty acids, is moderately effective in lowering elevated serum triglycerides. At doses of 2 to 4 g/d, omega-3 fatty acids inhibit VLDL production, while having little or no effect on HDL-cholesterol. In hypertriglyceridemic subjects, there is typically a small rise in LDL-cholesterol when omega-3 preparations are used that contain both EPA (eicospentanoic acid) and DHA (docosahexaenoic acid). While there are many epidemiologic studies of fish consumption that suggest that fish oils produce a CV benefit, demonstrating that benefit in prospective trials of fish oil supplements has been difficult. Three large CV outcome studies were performed in the past decade testing the efficacy of fish oils.[227-229] These studies used differing amounts of the omega-3 fatty acids in different subsets of patients with varying degrees of CV risk, including patients with a history of prior myocardial infarction or diabetes. In all three studies, triglycerides were reduced, but none of the populations obtained a CV benefit from the tested preparations. The best available evidence suggested that omega-3 fatty acids may reduce lethal cardiac arrhythmic events but have a modest effect, if any, on coronary atherosclerosis.[230-232] Newer, EPA-only fish oil formulations appear to have a more favorable impact on LDL-cholesterol levels than preparations containing DHA, although this impact is usually lesser LDL-cholesterol elevation, not actual LDL-cholesterol lowering.[233] The REDUCE-IT trial compared icosapent ethyl 2 g twice daily to placebo in 8179 patients with CV disease (or who were 50 years old and had diabetes mellitus and at least one additional CV risk factor). Patients in the trial had fasting triglyceride levels between 135 mg/dL and 499 mg/dL, and almost all were receiving statin therapy. In contrast with other fish oil intervention trials utilizing combination fish oil formulations, REDUCE-IT demonstrated a statistically significant 25% reduction in the primary composite CV outcome.[234] The results of this study have been widely debated due to its benefits where other fish oil trials proved ineffective and because the mechanism through which the benefit was conferred remains uncertain. Again, as with the fibrate studies, the lowering of triglyceride levels did not appear to correlate well with the CV outcome benefit. Although there were higher rates of atrial fibrillation in the patients receiving the fish oil, as well as a modest and statistically nonsignificant increase in serious bleeding events, isocapent ethyl has become, as a result of the REDUCE-IT trial, the clearly preferred fish oil agent when used for the purpose of producing a CV outcome benefit.

Future Therapies. Several newer agents have been studied in the past few years that may prove useful in the near future for the treatment of lipoprotein disorders.

Bempedoic acid: This drug is an oral inhibitor of ATP citrate lyase, an enzyme linked to the cholesterol synthesis pathway, that reduces LDL-cholesterol. In a trial of 2230 patients on maximally tolerated statins whose LDL-cholesterol levels remained above 70 mg/dL, bempedoic acid was shown to reduce LDL values by approximately 18% compared with those measured in the group receiving placebo treatment. The drug did produce more gout adverse events. A greater incidence of tendon rupture has also been reported. In February 2020, the FDA approved this drug for use in adult FH heterozygotes or individuals with established ASCVD who require more LDL-cholesterol lowering. Its place in therapeutic algorithms is still uncertain, but clearly it would be utilized only if statin and/or ezetimibe therapy were unable to meet lipid-lowering goals.[235]

Inclisiran: This small interfering RNA molecule is a PCSK9 inhibitor that works via inhibition of the production of PCSK9 protein rather than by binding the secreted protein, as the antibodies alirocumab and evolocumab do. In the ORION study, 482 adults with heterozygous FH on maximal tolerated statin therapy were given subcutaneous injections of the inclisiran on days 1, 90, 279, and 450. At day 510, inclisiran reduced baseline LDL-cholesterol values of 153 mg/dL by a placebo-adjusted approximately 48%.[236] Inclisiran was approved in Europe in December 2020, but has not yet received approval in the United States. The twice-a-year dosing regimen of the drug could improve patient compliance with lipid medications, and, if priced substantially below the cost of the PCSK9 antibodies, it may provide greater access to this pathway for LDL-cholesterol lowering.

Antisense Rx of Lp(a): Another antisense therapy called AKCEA-APO(a)-LRx has demonstrated very significant reductions in lipoprotein(a) (also called Lp(a)). This lipoprotein represents a covalent linkage between a protein called apoprotein(a) and an LDL molecule. Higher levels of Lp(a) are positively correlated with higher ASCVD risk and are primarily determined by the genetic isoforms of apolipoprotein that a patient carries. No current therapy for lowering Lp(a) has been approved for use, and the CV benefit of that lowering has never been convincingly demonstrated, though niacin and PCSK9 antibodies have both been shown to lower Lp(a) by approximately 25%. With this new antisense technology, in early phase trials, reductions of 67% to 72% were achieved in trials of moderate doses, and at the highest dosing level, a 90% reduction was measured.[237] A more recent study reported dose-dependent Lp(a) reductions of up to 80%.[238] As the benefit of reducing Lp(a) by any means on CV risk is unknown, the drug will likely require a beneficial CV outcomes study to achieve regulatory approval. That study is currently being enrolled.

Acknowledgments

We express our gratitude to Caroline Diggins and Samuel Russo for their assistance in the revision of this chapter.

REFERENCES

1. Gistera A, Hansson GK. The immunology of atherosclerosis. *Nat Rev Nephrol*. 2017;13:368–380.
2. Libby P, Buring JE, Badimon L, et al. *Atherosclerosis. Nat Rev Dis Primers*. 2019;5:56.
3. Almdal T, Scharling H, Jensen JS, et al. The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13,000 men and women with 20 years of follow-up. *Archives Internal Med*. 2004;164:1422–1426.
4. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis: epidemiology, pathophysiology, and management. *Jama*. 2002;287:2570–2581.

5. Rawshani A, Rawshani A, Franzen S, et al. Mortality and cardiovascular disease in type 1 and type 2 diabetes. *New England J Med*. 2017;376:1407–1418.
6. Asztalos BF, Collins D, Cupples LA, et al. Value of high-density lipoprotein (HDL) subpopulations in predicting recurrent cardiovascular events in the Veterans Affairs HDL Intervention Trial. *Arteriosclerosis Thromb Vascular Biol*. 2005;25:2185–2191.
7. Havel RJ, Kane JP. Introduction: structure and metabolism of plasma lipoproteins. In: Valle DL, Antonarakis S, Ballabio A, Beaudet AL, et al., eds. *The Online Metabolic & Molecular Bases of Inherited Disease*. 8th ed. New York: McGraw-Hill; 2019.
8. Zhang P, Csaki LS, Ronquillo E, et al. Lipin 2/3 phosphatidic acid phosphatases maintain phospholipid homeostasis to regulate chylomicron synthesis. *J Clin Invest*. 2019;129:281–295.
9. Powell LM, Wallis SC, Pease RJ, et al. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell*. 1987;50:831–840.
10. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National cholesterol education program adult treatment panel III guidelines. *Circulation*. 2004;110:227–239.
11. Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins. A meta-analysis of 27 trials. *Arterioscler Thromb*. 1992;12:911–919.
12. Tosi I, Toledo-Leiva P, Neuwirth C, et al. Genetic defects causing familial hypercholesterolaemia: identification of deletions and duplications in the LDL-receptor gene and summary of all mutations found in patients attending the Hammersmith Hospital Lipid Clinic. *Atherosclerosis*. 2007;194:102–111.
13. Veniant MM, Beigneux AP, Bensadoun A, et al. Lipoprotein size and susceptibility to atherosclerosis—insights from genetically modified mouse models. *Curr Drug Targets*. 2008;9:174–189.
14. Packard CJ. Small dense low-density lipoprotein and its role as an independent predictor of cardiovascular disease. *Curr Opinion Lipidol*. 2006;17:412–417.
15. Cuchel M, Rader DJ. Macrophage reverse cholesterol transport: key to the regression of atherosclerosis? *Circulation*. 2006;113:2548–2555.
16. Holleboom AG, Jakulj L, Franssen R, et al. In vivo tissue cholesterol efflux is reduced in carriers of a mutation in APOA1. *J Lipid Res*. 2013;54:1964–1971.
17. Hung KT, Berisha SZ, Ritchey BM, et al. Red blood cells play a role in reverse cholesterol transport. *Arteriosclerosis Thromb Vascul Biol*. 2012;32:1460–1465.
18. Sankaranarayanan S, de la Llera-Moya M, Drazul-Schrader D, et al. Serum albumin acts as a shuttle to enhance cholesterol efflux from cells. *J Lipid Res*. 2013;54:671–676.
19. Vaisar T, Pennathur S, Green PS, et al. Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL. *J Clin Invest*. 2007;117:746–756.
20. Scanu AM, Edelstein C. HDL: bridging past and present with a look at the future. *FASEB J*. 2008;22:4044–4054. https://doi.org/10.1096/fj.08-117150.
21. Musunuru K, Kathiresan S. Surprises from genetic analyses of lipid risk factors for atherosclerosis. *Circul Res*. 2016;118:579–585.
22. Drover VA, Ajmal M, Nassir F, et al. CD36 deficiency impairs intestinal lipid secretion and clearance of chylomicrons from the blood. *J Clin Invest*. 2005;115:1290–1297.
23. Kamp F, Guo W, Souto R, et al. Rapid flip-flop of oleic acid across the plasma membrane of adipocytes. *J Biol Chem*. 2003;278:7988–7995.
24. Hui DY, Labonte ED, Howles PN. Development and physiological regulation of intestinal lipid absorption. III. Intestinal transporters and cholesterol absorption. *Am J Physiol. Gastrointestinal Liver Physiol*. 2008;294:G839–G843.
25. Altmann SW, Davis Jr HR, Zhu LJ, et al. Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol absorption. *Science*. 2004;303:1201–1204.
26. Garcia-Calvo M, Lisnock J, Bull HG, et al. The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1). *Proc Natl Acad Sci USA*. 2005;102:8132–8137.
27. Black DD. Development and physiological regulation of intestinal lipid absorption. I. Development of intestinal lipid absorption: cellular events in chylomicron assembly and secretion. *Am J Physiol Gastrointest Liver Physiol*. 2007;293:G519–G524.
28. Yu L, Li-Hawkins J, Hammer RE, et al. Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol. *J Clin Investigat*. 2002;110:671–680.
29. Dane-Stewart CA, Watts GF, Barrett PH, et al. Chylomicron remnant metabolism studied with a new breath test in postmenopausal women with and without type 2 diabetes mellitus. *Clin Endocrinol (Oxf)*. 2003;58:415–420.
30. Hussain MM, Innerarity TL, Brecht WJ, et al. Chylomicron metabolism in normal, cholesterol-fed, and Watanabe heritable hyperlipidemic rabbits. Saturation of the sequestration step of the remnant clearance pathway. *J Biol Chem*. 1995;270:8578–8587.
31. Kane JP, Havel RJ. Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins. In: Scriver CR, Beaudet AL, Sly WS, et al., eds. *The Metabolic and Molecular Bases of Inherited Disease*. 8th ed. New York: McGraw-Hill; 2001.
32. Kang S, Davis RA. Cholesterol and hepatic lipoprotein assembly and secretion. *Biochim Biophys Acta*. 2000;1529:223–230.
33. Olofsson SO, Asp L, Boren J. The assembly and secretion of apolipoprotein B-containing lipoproteins. *Curr Opinion Lipidol*. 1999;10:341–346.
34. Parhofer KG, Barrett PH. Thematic review series: patient-oriented research. What we have learned about VLDL and LDL metabolism from human kinetics studies. *J Lipid Res*. 2006;47:1620–1630.
35. Goldstein JL, Hobbs HH, Brown MS. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, et al., eds. *The Metabolic and Molecular Bases of Inherited Disease*. 8th ed. New York: McGraw-Hill; 2001.
36. Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986;232:34–47.
37. Radhakrishnan A, Ikeda Y, Kwon HJ, et al. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: oxysterols block transport by binding to Insig. *Proceed Nat Acad Sci United States Am*. 2007;104:6511–6518.
38. Sun LP, Seemann J, Goldstein JL, et al. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: Insig renders sorting signal in Scap inaccessible to COPII proteins. *Proceed Nat Acad Sci United States Am*. 2007;104:6519–6526.
39. Gong Y, Lee JN, Lee PC, et al. Sterol-regulated ubiquitination and degradation of Insig-1 creates a convergent mechanism for feedback control of cholesterol synthesis and uptake. *Cell Metabol*. 2006;3:15–24.
40. Flury I, Garza R, Shearer A, et al. INSIG: a broadly conserved transmembrane chaperone for sterol-sensing domain proteins. *EMBO J*. 2005;24:3917–3926.
41. Sun LP, Li L, Goldstein JL, et al. Insig required for sterol-mediated inhibition of Scap/SREBP binding to COPII proteins in vitro. *J Biol Chem*. 2005;280:26483–26490.
42. Adams CM, Reitz J, De Brabander JK, et al. Cholesterol and 25-hydroxycholesterol inhibit activation of SREBPs by different mechanisms, both involving SCAP and Insigs. *J Biol Chemis*. 2004;279:52772–52780.
43. Attie AD. ABCA1: at the nexus of cholesterol, HDL and atherosclerosis. *Trend Biochem Sci*. 2007;32:172–179.
44. Chang C, Dong R, Miyazaki A, et al. Human acyl-CoA:cholesterol acyltransferase (ACAT) and its potential as a target for pharmaceutical intervention against atherosclerosis. *Acta Biochim Biophys Sin (Shanghai)*. 2006;38:151–156.
45. Zhao B, Song J, Chow WN, et al. Macrophage-specific transgenic expression of cholesteryl ester hydrolase significantly reduces atherosclerosis and lesion necrosis in Ldlr mice. *J Clin Invest*. 2007;117:2983–2992.
46. Chroni A, Liu T, Fitzgerald ML, et al. Cross-linking and lipid efflux properties of apoA-I mutants suggest direct association between apoA-I helices and ABCA1. *Biochemistry*. 2004;43:2126–2139.
47. Fitzgerald ML, Mendez AJ, Moore KJ, et al. ATP-binding cassette transporter A1 contains an NH2-terminal signal anchor sequence that translocates the protein's first hydrophilic domain to the exoplasmic space. *J Biol Chem*. 2001;276:15137–15145.
48. Fitzgerald ML, Morris AL, Rhee JS, et al. Naturally occurring mutations in the largest extracellular loops of ABCA1 can disrupt its direct interaction with apolipoprotein A-I. *J Biol Chem*. 2002;277:33178–33187.

49. Tall AR, Yvan-Charvet L, Terasaka N, et al. HDL, ABC transporters, and cholesterol efflux: implications for the treatment of atherosclerosis. *Cell Metabol*. 2008;7:365–375.
50. Qian H, Zhao X, Cao P, et al. Structure of the human lipid exporter ABCA1. *Cell*. 2017;169:1228–1239.e10.
51. Trigatti BL, Krieger M, Rigotti A. Influence of the HDL receptor SR-BI on lipoprotein metabolism and atherosclerosis. *Arteriosclerosis Thrombosis Vascul Biol*. 2003;23:1732–1738.
52. Deckert T, Poulsen JE, Larsen M. Prognosis of diabetics with diabetes onset before the age of thirty-one. I. Survival, causes of death, and complications. *Diabetologia*. 1978;14:363–370.
53. Dorman JS, Laporte RE, Kuller LH, et al. The Pittsburgh insulin-dependent diabetes mellitus (IDDM) morbidity and mortality study. Mortality results. *Diabetes*. 1984;33:271–276.
54. Orchard TJ, Costacou T, Kretowski A, et al. Type 1 diabetes and coronary artery disease. *Diabetes Care*. 2006;29:2528–2538.
55. Laing SP, Swerdlow AJ, Slater SD, et al. Mortality from heart disease in a cohort of 23,000 patients with insulin-treated diabetes. *Diabetologia*. 2003;46:760–765.
56. Livingstone SJ, Looker HC, Hothersall EJ, et al. Risk of cardiovascular disease and total mortality in adults with type 1 diabetes: Scottish registry linkage study. *PLoS Med*. 2012;9:e1001321.
57. de Ferranti SD, de Boer IH, Fonseca V, et al. Type 1 diabetes mellitus and cardiovascular disease: a scientific statement from the American Heart Association and American Diabetes Association. *Diabetes Care*. 2014;37:2843–2863.
58. Jenkins AJ, Lyons TJ, Zheng D, et al. Serum lipoproteins in the diabetes control and complications trial/epidemiology of diabetes intervention and complications cohort: associations with gender and glycemia. *Diabetes Care*. 2003;26:810–818.
59. Tell S, Nadeau KJ, Eckel RH. Lipid management for cardiovascular risk reduction in type 1 diabetes. *Curr Opin Endocrinol Diabetes Obes*. 2020;27:207–214.
60. Orchard TJ, Olson JC, Erbey JR, et al. Insulin resistance-related factors, but not glycemia, predict coronary artery disease in type 1 diabetes: 10-year follow-up data from the Pittsburgh Epidemiology of Diabetes Complications Study. *Diabetes Care*. 2003;26:1374–1379.
61. Giunti S, Bruno G, Veglio M, et al. Electrocardiographic left ventricular hypertrophy in type 1 diabetes: prevalence and relation to coronary heart disease and cardiovascular risk factors: the Eurodiab IDDM Complications Study. *Diabetes Care*. 2005;28:2255–2257.
62. Nathan DM, Cleary PA, Backlund JY, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *New Engl J Med*. 2005;353:2643–2653.
63. Lind M, Svensson AM, Kosiborod M, et al. Glycemic control and excess mortality in type 1 diabetes. *New Engl J Med*. 2014;371:1972–1982.
64. Miller RG, Secrest AM, Sharma RK, et al. Improvements in the life expectancy of type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complications study cohort. *Diabetes*. 2012;61:2987–2992.
65. Constantino MI, Molyneaux L, Limacher-Gisler F, et al. Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. *Diabetes Care*. 2013;36:3863–3869.
66. Gregg EW, Li Y, Wang J, et al. Changes in diabetes-related complications in the United States, 1990-2010. *New Engl J Med*. 2014;370:1514–1523.
67. Mykkanen L, Laakso M, Penttila I, et al. Asymptomatic hyperglycemia and cardiovascular risk factors in the elderly. *Atherosclerosis*. 1991;88:153–161.
68. Haffner SM. Lipoprotein disorders associated with type 2 diabetes mellitus and insulin resistance. *Am J Cardiol*. 2002;90:55i–61i.
69. Packard CJ. Triacylglycerol-rich lipoproteins and the generation of small, dense low-density lipoprotein. *Biochem Society Transact*. 2003;31:1066–1069.
70. Sparks JD, Phung TL, Bolognino M, et al. Insulin-mediated inhibition of apolipoprotein B secretion requires an intracellular trafficking event and phosphatidylinositol 3-kinase activation: studies with brefeldin A and wortmannin in primary cultures of rat hepatocytes. *Biochem J*. 1996;313:567–574.
71. Au WS, Kung HF, Lin MC. Regulation of microsomal triglyceride transfer protein gene by insulin in HepG2 cells: roles of MAPKerk and MAPKp38. *Diabetes*. 2003;52:1073–1080.
72. Hu P, Dharmayat KI, Stevens CAT, et al. Prevalence of familial hypercholesterolemia among the general population and patients with atherosclerotic cardiovascular disease: a systematic review and meta-analysis. *Circulation*. 2020;141:1742–1759.
73. Raal FJ, Honarpour N, Blom DJ, et al. Inhibition of PCSK9 with evolocumab in homozygous familial hypercholesterolaemia (TESLA Part B): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;385:341–350.
74. Raal FJ, Hovingh GK, Blom D, et al. Long-term treatment with evolocumab added to conventional drug therapy, with or without apheresis, in patients with homozygous familial hypercholesterolaemia: an interim subset analysis of the open-label TAUSSIG study. *Lancet Diabetes Endocrinol*. 2017;5:280–290.
75. Stein EA, Honarpour N, Wasserman SM, et al. Effect of the proprotein convertase subtilisin/kexin 9 monoclonal antibody, AMG 145, in homozygous familial hypercholesterolemia. *Circulation*. 2013;128:2113–2120.
76. Blom DJ, Harada-Shiba M, Rubba P, et al. Efficacy and safety of alirocumab in adults with homozygous familial hypercholesterolemia: the ODYSSEY HoFH trial. *J Am Coll Cardiol*. 2020;76:131–142.
77. Raal FJ, Santos RD, Blom DJ, et al. Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2010;375:998–1006.
78. McGowan MP, Tardif JC, Ceska R, et al. Randomized, placebo-controlled trial of mipomersen in patients with severe hypercholesterolemia receiving maximally tolerated lipid-lowering therapy. *PLoS One*. 2012;7:e49006.
79. Cuchel M, Bloedon LT, Szapary PO, et al. Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia. *New England J Med*. 2007;356:148–156.
80. Cuchel M, Meagher EA, du Toit Theron H, et al. Efficacy and safety of a microsomal triglyceride transfer protein inhibitor in patients with homozygous familial hypercholesterolaemia: a single-arm, open-label, phase 3 study. *Lancet*. 2013;381:40–46.
81. Raal FJ, Rosenson RS, Reeskamp LF, et al. Evinacumab for homozygous familial hypercholesterolemia. *New Engl J Med*. 2020;383:711–720.
82. Myant NB. Familial defective apolipoprotein B-100: a review, including some comparisons with familial hypercholesterolaemia. *Atherosclerosis*. 1993;104:1–18.
83. Whitfield AJ, Barrett PH, van Bockxmeer FM, et al. Lipid disorders and mutations in the APOB gene. *Clin Chem*. 2004;50:1725–1732.
84. Soutar AK, Naoumova RP. Autosomal recessive hypercholesterolemia. *Seminars Vascul Med*. 2004;4:241–248.
85. He G, Gupta S, Yi M, et al. ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. *J Biol Chemis*. 2002;277:44044–44049.
86. Garcia CK, Wilund K, Arca M, et al. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science*. 2001;292:1394–1398.
87. D'Erasmo L, Di Costanzo A, Arca M. Autosomal recessive hypercholesterolemia: update for 2020. *Curr Opinion Lipidol*. 2020;31:56–61.
88. Cohen JC, Boerwinkle E, Mosley Jr TH, et al. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *New Engl J Med*. 2006;354:1264–1272.
89. Peterson AS, Fong LG, Young SG. PCSK9 function and physiology. *J Lipid Res*. 2008;49:1152–1156.
90. Koishi R, Ando Y, Ono M, et al. Angptl3 regulates lipid metabolism in mice. *Nature Genet*. 2002;30:151–157.
91. Romeo S, Yin W, Kozlitina J, et al. Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *The J Clin Investigat*. 2009;119:70–79.
92. Glazer NL. Exome sequencing links gene mutation in angiopoietin-like protein 3 with low-density lipoprotein cholesterol. *Circ Cardiovasc Genet*. 2011. PMID: 21325168.

93. Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med*. 2010;363:2220–2227.
94. Singaraja RR, Brunham LR, Visscher H, et al. Efflux and atherosclerosis: the clinical and biochemical impact of variations in the ABCA1 gene. *Arteriosclerosis Thrombosis Vascul Biol*. 2003;23:1322–1332.
95. Oram JF. ATP-binding cassette transporter A1 and cholesterol trafficking. *Curr Opinion Lipidol*. 2002;13:373–381.
96. Calabresi L, Pisciotta L, Costantin A, et al. The molecular basis of lecithin:cholesterol acyltransferase deficiency syndromes: a comprehensive study of molecular and biochemical findings in 13 unrelated Italian families. *Arteriosclerosis Thrombosis Vascul Biol*. 2005;25:1972–1978.
97. Sorci-Thomas MG, Thomas MJ. The effects of altered apolipoprotein A-I structure on plasma HDL concentration. *Trends Cardiovascul Med*. 2002;12:121–128.
98. Sasaki J, Matsunaga A, Huang W, et al. Structural and functional properties of apolipoprotein A-I mutants. *J Atherosclerosis Thrombosis*. 2000;7:67–70.
99. Schaefer EJ. Clinical, biochemical, and genetic features in familial disorders of high density lipoprotein deficiency. *Arteriosclerosis*. 1984;4:303–322.
100. Hovingh GK, de Groot E, van der Steeg W, et al. Inherited disorders of HDL metabolism and atherosclerosis. *Curr Opinion Lipidol*. 2005;16:139–145.
101. Kathiresan S, Manning AK, Demissie S, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet*. 2007;8:S17.
102. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713.
103. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*. 2009;41:56–65.
104. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*. 2009;41:47–55.
105. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714–719.
106. Burkhardt R, Toh SA, Lagor WR, et al. Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J Clin investigat*. 2010;120:4410–4414.
107. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380:572–580.
108. Sampson M, Ling C, Sun Q, et al. A new equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia. *JAMA Cardiol*. 2020;5:540–548.
109. 2013 Stone NJ, Robinson J, Lichtenstein AH, et al. ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Circulation*. 2014;129:S1–S45.
110. Keaney JF Jr, Curfman GD, Jarcho JA. A pragmatic view of the new cholesterol treatment guidelines. *N Engl J Med*. 2014;370:275–278.
111. 2018 Grundy SM, Stone NJ, Bailey AL, et al. AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Circulation*. 2019;139:e1082–e1143.
112. Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *New Engl J Med*. 2004;350:1495–1504.
113. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *New Engl J Med*. 2017;376:1713–1722.
114. Schwartz GG, Steg PG, Szarek M, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *New Engl J Med*. 2018;379:2097–2107.
115. Feinstein MJ, Jhund P, Kang J, et al. Do statins reduce the risk of myocardial infarction in patients with heart failure? A pooled individual-level reanalysis of CORONA and GISSI-HF. *Eur J Heart Fail*. 2015;17:434–441.
116. Chiavaroli L, Nishi SK, Khan TA, et al. Portfolio dietary pattern and cardiovascular disease: a systematic review and meta-analysis of controlled trials. *Prog Cardiovasc Dis*. 2018;61:43–53.
117. 2013 Eckel RH, Jakicic JM, Ard JD, et al. AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S76–S99.
118. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*. 2008;40:189–197.
119. Levin MG, Rader DJ. Polygenic risk scores and coronary artery disease: ready for prime time? *Circulation*. 2020;141:637–640.
120. Marston NA, Kamanu FK, Nordio F, et al. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER Trial. *Circulation*. 2020;141:616–623.
121. Damask A, Steg PG, Schwartz GG, et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES Trial. *Circulation*. 2020;141:624–636.
122. Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372:2387–2397.
123. Haffner SM, Lehto S, Ronnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *New Engl J Med*. 1998;339:229–234.
124. Juutilainen A, Lehto S, Ronnemaa T, et al. Type 2 diabetes as a "coronary heart disease equivalent": an 18-year prospective population-based study in Finnish subjects. *Diabetes Care*. 2005;28:2901–2907.
125. Haffner SM, Agostino Jr RD, Saad MF, et al. Carotid artery atherosclerosis in type-2 diabetic and nondiabetic subjects with and without symptomatic coronary artery disease (The Insulin Resistance Atherosclerosis Study). *Am J Cardiol*. 2000;85:1395–1400.
126. Malmberg K, Yusuf S, Gerstein HC, et al. Impact of diabetes on long-term prognosis in patients with unstable angina and non-Q-wave myocardial infarction: results of the OASIS (Organization to Assess Strategies for Ischemic Syndromes) Registry. *Circulation*. 2000;102:1014–1019.
127. Yusuf S, Sleight P, Pogue J, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *New Engl J Med*. 2000;342:145–153.
128. Mukamal KJ, Nesto RW, Cohen MC, et al. Impact of diabetes on long-term survival after acute myocardial infarction: comparability of risk with prior myocardial infarction. *Diabetes Care*. 2001;24:1422–1427.
129. Natarajan S, Liao Y, Sinha D, et al. Sex differences in the effect of diabetes duration on coronary heart disease mortality. *Archiv Internal Med*. 2005;165:430–435.
130. Hu FB, Stampfer MJ, Solomon CG, et al. The impact of diabetes mellitus on mortality from all causes and coronary heart disease in women: 20 years of follow-up. *Archiv Internal Med*. 2001;161:1717–1723.
131. Lotufo PA, Gaziano JM, Chae CU, et al. Diabetes and all-cause and coronary heart disease mortality among US male physicians. *Archiv Internal Med*. 2001;161:242–247.
132. Evans JM, Wang J, Morris AD. Comparison of cardiovascular risk between patients with type 2 diabetes and those who had had a myocardial infarction: cross sectional and cohort studies. *Bmj*. 2002;324:939–942.
133. Cho E, Rimm EB, Stampfer MJ, et al. The impact of diabetes mellitus and prior myocardial infarction on mortality from all causes and from coronary heart disease in men. *J Am Coll Cardiol*. 2002;40:954–960.
134. Becker A, Bos G, de Vegt F, et al. Cardiovascular events in type 2 diabetes: comparison with nondiabetic individuals without and with prior cardiovascular disease. 10-year follow-up of the Hoorn Study. *European Heart J*. 2003;24:1406–1413.

135. Natarajan S, Liao Y, Cao G, et al. Sex differences in risk for coronary heart disease mortality associated with diabetes and established coronary heart disease. *Archiv Internal Med*. 2003;163:1735–1740.
136. Lee CD, Folsom AR, Pankow JS, et al. Cardiovascular events in diabetic and nondiabetic adults with or without history of myocardial infarction. *Circulation*. 2004;109:855–860.
137. Vaccaro O, Eberly LE, Neaton JD, et al. Impact of diabetes and previous myocardial infarction on long-term survival: 25-year mortality follow-up of primary screenees of the Multiple Risk Factor Intervention Trial. *Archive Internal Med*. 2004;164:1438–1443.
138. Hu G, Jousilahti P, Qiao Q, et al. Sex differences in cardiovascular and total mortality among diabetic and non-diabetic individuals with or without history of myocardial infarction. *Diabetologia*. 2005;48:856–861.
139. Howard BV, Van Horn L, Hsia J, et al. Low-fat dietary pattern and risk of cardiovascular disease: the Women's health initiative randomized controlled dietary modification trial. *Jama*. 2006;295:655–666.
140. Mensink RP, Katan MB. Effect of a diet enriched with monounsaturated or polyunsaturated fatty acids on levels of low-density and high-density lipoprotein cholesterol in healthy women and men. *New Engl J Med*. 1989;321:436–441.
141. Ryan M, McInerney D, Owens D, et al. Diabetes and the Mediterranean diet: a beneficial effect of oleic acid on insulin sensitivity, adipocyte glucose transport and endothelium-dependent vasoreactivity. *QJM : Monthly journal of the Association of Physicians*. 2000;93:85–91.
142. Schaefer EJ, Lamon-Fava S, Ausman LM, et al. Individual variability in lipoprotein cholesterol response to National cholesterol education program step 2 diets. *Am J Clin Nutrit*. 1997;65:823–830.
143. Foster GD, Wyatt HR, Hill JO, et al. A randomized trial of a low-carbohydrate diet for obesity. *New Engl J Med*. 2003;348:2082–2090.
144. Stern L, Iqbal N, Seshadri P, et al. The effects of low-carbohydrate versus conventional weight loss diets in severely obese adults: one-year follow-up of a randomized trial. *Ann Internal Med*. 2004;140:778–785.
145. Gardner CD, Kiazand A, Alhassan S, et al. Comparison of the Atkins, Zone, Ornish, and LEARN diets for change in weight and related risk factors among overweight premenopausal women: the A TO Z Weight Loss Study: a randomized trial. *Jama*. 2007;297:969–977.
146. McAuley KA, Hopkins CM, Smith KJ, et al. Comparison of high-fat and high-protein diets with a high-carbohydrate diet in insulin-resistant obese women. *Diabetologia*. 2005;48:8–16.
147. Mayer-Davis EJ, Sparks KC, Hirst K, et al. Dietary intake in the diabetes prevention program cohort: baseline and 1-year post randomization. *Ann Epidemiol*. 2004;14:763–772.
148. Wing RR, Hamman RF, Bray GA, et al. Achieving weight and activity goals among diabetes prevention program lifestyle participants. *Obes Res*. 2004;12:1426–1434.
149. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New Engl J Med*. 2002;346:393–403.
150. Ratner R, Goldberg R, Haffner S, et al. Impact of intensive lifestyle and metformin therapy on cardiovascular disease risk factors in the diabetes prevention program. *Diabet Care*. 2005;28:888–894.
151. Balagopal P, Kamalamma N, Patel TG, et al. A community-based diabetes prevention and management education program in a rural village in India. *Diabet Care*. 2008;31:1097–1104.
152. Seidel MC, Powell RO, Zgibor JC, et al. Translating the Diabetes Prevention Program into an urban medically underserved community: a nonrandomized prospective intervention study. *Diabet care*. 2008;31:684–689.
153. Watts GF, Lewis B, Brunt JN, et al. Effects on coronary artery disease of lipid-lowering diet, or diet plus cholestyramine, in the St Thomas' Atherosclerosis Regression Study (STARS). *Lancet*. 1992;339:563–569.
154. Ornish D, Brown SE, Scherwitz LW, et al. Can lifestyle changes reverse coronary heart disease? The Lifestyle Heart Trial. *Lancet*. 1990;336:129–133.
155. Ornish D, Scherwitz LW, Billings JH, et al. Intensive lifestyle changes for reversal of coronary heart disease. *Jama*. 1998;280:2001–2007.
156. de Lorgeril M, Renaud S, Mamelle N, et al. Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. *Lancet*. 1994;343:1454–1459.
157. de Lorgeril M, Salen P, Martin JL, et al. Effect of a mediterranean type of diet on the rate of cardiovascular complications in patients with coronary artery disease. Insights into the cardioprotective effect of certain nutriments. *J Am Coll Cardiol*. 1996;28:1103–1108.
158. de Lorgeril M, Salen P, Martin JL, et al. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. *Circulation*. 1999;99:779–785.
159. Estruch R, Ros E, Salas-Salvado J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med*. 2013;368:1279–1290.
160. The Lipid Research Clinics Coronary Primary Prevention Trial results. II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *JAMA*. 1984;251:365–374.
161. The Lipid Research Clinics Coronary Primary Prevention Trial Results. I. Reduction in incidence of coronary heart disease. *JAMA*. 1984;251:351–364.
162. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994;344:1383–1389.
163. Miettinen TA, Pyorala K, Olsson AG, et al. Cholesterol-lowering therapy in women and elderly patients with myocardial infarction or angina pectoris: findings from the Scandinavian Simvastatin Survival Study (4S). *Circulation*. 1997;96:4211–4218.
164. Pedersen TR, Olsson AG, Faergeman O, et al. Lipoprotein changes and reduction in the incidence of major coronary heart disease events in the Scandinavian Simvastatin Survival Study (4S). *Circulation*. 1998;97:1453–1460.
165. Pedersen TR, Wilhelmsen L, Faergeman O, et al. Follow-up study of patients randomized in the Scandinavian Simvastatin Survival Study (4S) of cholesterol lowering. *Am J Cardiol*. 2000;86:257–262.
166. Pyorala K, Pedersen TR, Kjekshus J, et al. Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin Survival Study (4S). *Diabetes Care*. 1997;20:614–620.
167. Ballantyne CM, Olsson AG, Cook TJ, et al. Influence of low high-density lipoprotein cholesterol and elevated triglyceride on coronary heart disease events and response to simvastatin therapy in 4S. *Circulation*. 2001;104:3046–3051.
168. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *New Engl J Med*. 1995;333:1301–1307.
169. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *Jama*. 1998;279:1615–1622.
170. Sacks FM, Pfeffer MA, Moye LA, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. *New Engl J Med*. 1996;335:1001–1009.
171. Sacks FM, Moye LA, Davis BR, et al. Relationship between plasma LDL concentrations during treatment with pravastatin and recurrent coronary events in the Cholesterol and Recurrent Events trial. *Circulation*. 1998;97:1446–1452.
172. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. *N Engl J Med*. 1998;339:1349–1357.
173. Colquhoun D, Keech A, Hunt D, et al. Effects of pravastatin on coronary events in 2073 patients with low levels of both low-density lipoprotein cholesterol and high-density lipoprotein cholesterol: results from the LIPID study. *Euro Heart J*. 2004;25:771–777.
174. Simes RJ, Marschner IC, Hunt D, et al. Relationship between lipid levels and clinical outcomes in the Long-term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial: to what extent is the reduction in coronary events with pravastatin explained by on-study lipid levels? *Circulation*. 2002;105:1162–1169.

175. Tonkin AM, Colquhoun D, Emberson J, et al. Effects of pravastatin in 3260 patients with unstable angina: results from the LIPID study. *Lancet*. 2000;356:1871–1875.
176. LaRosa JC, Grundy SM, Waters DD, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. *New Engl J Med*. 2005;352:1425–1435.
177. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360:7–22.
178. Collins R, Armitage J, Parish S, et al. MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. *Lancet*. 2003;361:2005–2016.
179. Collins R, Armitage J, Parish S, et al. Effects of cholesterol-lowering with simvastatin on stroke and other major vascular events in 20536 people with cerebrovascular disease or other high-risk conditions. *Lancet*. 2004;363:757–767.
180. Colhoun HM, Betteridge DJ, Durrington PN, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet*. 2004;364:685–696.
181. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *New Engl J Med*. 2008;359:2195–2207.
182. Brown G, Albers JJ, Fisher LD, et al. Regression of coronary artery disease as a result of intensive lipid-lowering therapy in men with high levels of apolipoprotein B. *New Engl J Med*. 1990;323:1289–1298.
183. de Groot E, Jukema JW, Montauban van Swijndregt AD, et al. B-mode ultrasound assessment of pravastatin treatment effect on carotid and femoral artery walls and its correlations with coronary arteriographic findings: a report of the Regression Growth Evaluation Statin Study (REGRESS). *J Am Coll Cardiol*. 1998;31:1561–1567.
184. Jukema JW, Bruschke AV, van Boven AJ, et al. Effects of lipid lowering by pravastatin on progression and regression of coronary artery disease in symptomatic men with normal to moderately elevated serum cholesterol levels. The Regression Growth Evaluation Statin Study (REGRESS). *Circulation*. 1995;91:2528–2540.
185. Nissen SE, Tuzcu EM, Schoenhagen P, et al. Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. *Jama*. 2004;291:1071–1080.
186. Nissen SE, Nicholls SJ, Sipahi I, et al. Effect of very high-intensity statin therapy on regression of coronary atherosclerosis: the ASTEROID trial. *Jama*. 2006;295:1556–1565.
187. Pedersen TR, Faergeman O, Kastelein JJ, et al. High-dose atorvastatin vs usual-dose simvastatin for secondary prevention after myocardial infarction: the IDEAL study: a randomized controlled trial. *J Am Med Assoc*. 2005;294:2437–2445.
188. Cholesterol Treatment Trialists' (CTT) Collaborators, Baigent C, Blackwell L, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet*. 2010;376:1670–1681.
189. Cholesterol Treatment Trialists' (CTT) Collaborators, Kearney PM, Blackwell L, et al. Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. *Lancet*. 2008;371:117–125.
190. Cholesterol Treatment Trialists' (CTT) Collaborators, Mihaylova B, Emberson J, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet*. 2012;380:581–590.
191. Dale KM, Coleman CI, Henyan NN, et al. Statins and cancer risk: a meta-analysis. *J Am Med Assoc*. 2006;295:74–80.
192. Cholesterol Treatment Trialists' (CTT) Collaboration, Emberson JR, Kearney PM, et al. Lack of effect of lowering LDL cholesterol on cancer: meta-analysis of individual data from 175,000 people in 27 randomised trials of statin therapy. *PLoS One*. 2012;7:e29849.
193. Rajpathak SN, Kumbhani DJ, Crandall J, et al. Statin therapy and risk of developing type 2 diabetes: a meta-analysis. *Diabetes Care*. 2009;32:1924–1929.
194. Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet*. 2010;375:735–742.
195. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet*. 2005;366:1267–1278.
196. Stein EA, Gipe D, Bergeron J, et al. Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial. *Lancet*. 2012;380:29–36.
197. McKenney JM, Koren MJ, Kereiakes DJ, et al. Safety and efficacy of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 serine protease, SAR236553/REGN727, in patients with primary hypercholesterolemia receiving ongoing stable atorvastatin therapy. *J Am Coll Cardiol*. 2012;59:2344–2353.
198. Roth EM, McKenney JM, Hanotin C, et al. Atorvastatin with or without an antibody to PCSK9 in primary hypercholesterolemia. *New Engl J Med*. 2012;367:1891–1900.
199. Stein EA, Mellis S, Yancopoulos GD, et al. Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *New Engl J Med*. 2012;366:1108–1118.
200. Dias CS, Shaywitz AJ, Wasserman SM, et al. Effects of AMG 145 on low-density lipoprotein cholesterol levels: results from 2 randomized, double-blind, placebo-controlled, ascending-dose phase 1 studies in healthy volunteers and hypercholesterolemic subjects on statins. *J Am Coll Cardiol*. 2012;60:1888–1898.
201. Giugliano RP, Desai NR, Kohli P, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. *Lancet*. 2012;380:2007–2017.
202. Raal F, Scott R, Somaratne R, et al. Low-density lipoprotein cholesterol-lowering effects of AMG 145, a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 serine protease in patients with heterozygous familial hypercholesterolemia: the Reduction of LDL-C with PCSK9 Inhibition in Heterozygous Familial Hypercholesterolemia Disorder (RUTHERFORD) randomized trial. *Circulation*. 2012;126:2408–2417.
203. Koren MJ, Scott R, Kim JB, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 as monotherapy in patients with hypercholesterolaemia (MENDEL): a randomised, double-blind, placebo-controlled, phase 2 study. *Lancet*. 2012;380:1995–2006.
204. Sullivan D, Olsson AG, Scott R, et al. Effect of a monoclonal antibody to PCSK9 on low-density lipoprotein cholesterol levels in statin-intolerant patients: the GAUSS randomized trial. *J Am Med Assoc*. 2012;308:2497–2506.
205. Lusis AJ. *Atherosclerosis. Nature*. 2000;407:233–241.
206. Tabas I, Williams KJ, Boren J. Subendothelial lipoprotein retention as the initiating process in atherosclerosis: update and therapeutic implications. *Circulation*. 2007;116:1832–1844.
207. Buchwald H, Varco RL, Matts JP, et al. Effect of partial ileal bypass surgery on mortality and morbidity from coronary heart disease in patients with hypercholesterolemia. Report of the Program on the Surgical Control of the Hyperlipidemias (POSCH). *New Engl J Med*. 1990;323:946–955.
208. Sachais BS, Katz J, Ross J, et al. Long-term effects of LDL apheresis in patients with severe hypercholesterolemia. *J Clin Apher*. 2005;20:252–255.
209. Frick MH, Elo O, Haapa K, et al. Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment, changes in risk factors, and incidence of coronary heart disease. *New Engl J Med*. 1987;317:1237–1245.
210. Rubins HB, Robins SJ, Collins D, et al. Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial Study Group. *New Engl J Med*. 1999;341:410–418.

211. Berge KG, Canner PL. Coronary drug project: experience with niacin. Coronary drug project research group. *Euro J Clin Pharmacol*. 1991;40:S49–S51.
212. Canner PL, Berge KG, Wenger NK, et al. Fifteen year mortality in Coronary Drug Project patients: long-term benefit with niacin. *J Am Coll Cardiol*. 1986;8:1245–1255.
213. Stamler J. The coronary drug project--findings with regard to estrogen, dextrothyroxine, clofibrate and niacin. *Advan Experiment Med Biol*. 1977;82:52–75.
214. Bezafibrate Infarction Prevention (BIP). Secondary prevention by raising HDL cholesterol and reducing triglycerides in patients with coronary artery disease: the Bezafibrate Infarction Prevention (BIP) study. *Circulation*. 2000;102:21–27.
215. Barter PJ, Rye KA. Is there a role for fibrates in the management of dyslipidemia in the metabolic syndrome? *Arteriosclerosis Thrombosis Vascul Biol*. 2008;28:39–46.
216. Tenkanen L, Manttari M, Manninen V. Some coronary risk factors related to the insulin resistance syndrome and treatment with gemfibrozil. Experience from the Helsinki Heart Study. *Circulation*. 1995;92:1779–1785.
217. Robins SJ, Collins D, Wittes JT, et al. Relation of gemfibrozil treatment and lipid levels with major coronary events: VA-HIT: a randomized controlled trial. *Jama*. 2001;285:1585–1591.
218. Davidson MH. Statin/fibrate combination in patients with metabolic syndrome or diabetes: evaluating the risks of pharmacokinetic drug interactions. *Expet Opin Drug Saf*. 2006;5:145–156.
219. Keech A, Simes RJ, Barter P, et al. Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *Lancet*. 2005;366:1849–1861.
220. Group AS, Ginsberg HN, Elam MB, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. *New Engl J Med*. 2010;362:1563–1574.
221. The Coronary Drug Project Research Group. Clofibrate and niacin in coronary heart disease. *Jama*. 1975;231:360–381.
222. Brown BG, Zhao XQ, Chait A, et al. Simvastatin and niacin, antioxidant vitamins, or the combination for the prevention of coronary disease. *New Engl J Med*. 2001;345:1583–1592.
223. Taylor AJ, Sullenberger LE, Lee HJ, et al. Arterial Biology for the Investigation of the Treatment Effects of Reducing Cholesterol (ARBITER) 2: a double-blind, placebo-controlled study of extended-release niacin on atherosclerosis progression in secondary prevention patients treated with statins. *Circulation*. 2004;110:3512–3517.
224. AIM-HIGH Investigators, Boden WE, Probstfield JL, et al. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. *New Engl J Med*. 2011;365:2255–2267.
225. Cashin-Hemphill L, Mack WJ, Pogoda JM, et al. Beneficial effects of colestipol-niacin on coronary atherosclerosis. A 4-year follow-up. *Jama*. 1990;264:3013–3017.
226. Kastelein JJ, Akdim F, Stroes ES, et al. Simvastatin with or without ezetimibe in familial hypercholesterolemia. *New Engl J Med*. 2008;358:1431–1443.
227. Kromhout D, Giltay EJ, Geleijnse JM, et al. n-3 fatty acids and cardiovascular events after myocardial infarction. *New Engl J Med*. 2010;363:2015–2026.
228. ORIGIN Trial Investigators, Bosch J, Gerstein HC, et al. n-3 fatty acids and cardiovascular outcomes in patients with dysglycemia. *New Engl J Med*. 2012;367:309–318.
229. Risk and Prevention Study Collaborative Group, Roncaglioni MC, Tombesi M. n-3 fatty acids in patients with multiple cardiovascular risk factors. *New Engl J Med*. 2013;368:1800–1808.
230. Biscione F, Pignalberi C, Totteri A, et al. Cardiovascular effects of omega-3 free Fatty acids. *Curr Vascul Pharmacol*. 2007;5:163–172.
231. Ebbesson SO, Roman MJ, Devereux RB, et al. Consumption of omega-3 fatty acids is not associated with a reduction in carotid atherosclerosis: the Genetics of Coronary Artery Disease in Alaska Natives study. *Atherosclerosis*. 2008;199:346–353.
232. von Schacky C, Angerer P, Kothny W, et al. The effect of dietary omega-3 fatty acids on coronary atherosclerosis. A randomized, double-blind, placebo-controlled trial. *Ann Internal Med*. 1999;130:554–562.
233. Davidson MH. Omega-3 fatty acids: new insights into the pharmacology and biology of docosahexaenoic acid, docosapentaenoic acid, and eicosapentaenoic acid. *Curr Opin Lipidol*. 2013;24:467–474.
234. Bhatt DL, Steg PG, Miller M, et al. Cardiovascular risk reduction with icosapent ethyl for hypertriglyceridemia. *New Engl J Med*. 2019;380:11–22.
235. Ray KK, Bays HE, Catapano AL, et al. Safety and efficacy of bempedoic acid to reduce LDL cholesterol. *New Engl J Med*. 2019;380:1022–1032.
236. Raal FJ, Kallend D, Ray KK, et al. Inclisiran for the treatment of heterozygous familial hypercholesterolemia. *New Engl J Med*. 2020;382:1520–1530.
237. Viney NJ, van Capelleveen JC, Geary RS, et al. Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. *Lancet*. 2016;388:2239–2253.
238. Tsimikas S, Karwatowska-Prokopczuk E, Gouni-Berthold I, et al. Lipoprotein(a) reduction in persons with cardiovascular disease. *New Engl J Med*. 2020;382:244–255.
239. Davidson WS, Thompson TB. The structure of apolipoprotein A-I in high density lipoproteins. *J Biol Chemis*. 2007;282:22249–22253.
240. Martin-Campos JM, Escola-Gil JC, Ribas V, et al. Apolipoprotein A-II, genetic variation on chromosome 1q21-q24, and disease susceptibility. *Curr Opin Lipidol*. 2004;15:247–253.
241. Stan S, Delvin E, Lambert M, et al. Apo A-IV: an update on regulation and physiologic functions. *Biochim Biophys Acta*. 2003;1631:177–187.
242. Calandra S, Priore Oliva C, Tarugi P, et al. APOA5 and triglyceride metabolism, lesson from human APOA5 deficiency. *Curr Opinion Lipidol*. 2006;17:122–127.
243. Olofsson SO, Boren J. Apolipoprotein B: a clinically important apolipoprotein which assembles atherogenic lipoproteins and promotes the development of atherosclerosis. *J Internal Med*. 2005;258:395–410.
244. Gautier T, Tietge UJ, Boverhof R, et al. Hepatic lipid accumulation in apolipoprotein C-I-deficient mice is potentiated by cholesteryl ester transfer protein. *J Lipid Res*. 2007;48:30–40.
245. Pulawa LK, Jensen DR, Coates A, et al. Reduction of plasma triglycerides in apolipoprotein C-II transgenic mice overexpressing lipoprotein lipase in muscle. *J Lipid Res*. 2007;48:145–151.
246. Ooi EM, Barrett PH, Chan DC, et al. Apolipoprotein C-III: Understanding an emerging cardiovascular risk factor. *Clin Sci (Lond)*. 2008;114:611–624.
247. Rassart E, Bedirian A, Do Carmo S, et al. *Apolipoprotein D. Biochimica et biophysica acta*. 2000;1482:185–198.
248. Hatters DM, Peters-Libeu CA, Weisgraber KH. Apolipoprotein E structure: insights into function. *Trends Biochem Sci*. 2006;31:445–454.
249. Trougakos IP, Gonos ES. Regulation of clusterin/apolipoprotein J, a functional homologue to the small heat shock proteins, by oxidative stress in ageing and age-related diseases. *Free Radic Res*. 2006;40:1324–1334.
250. Ahnstrom J, Axler O, Jauhiainen M, et al. Levels of apolipoprotein M are not associated with the risk of coronary heart disease in two independent case-control studies. *J Lipid Res*. 2008;49:1912–1917.
251. Fitzgerald ML, Xavier R, Haley KJ, et al. ABCA3 inactivation in mice causes respiratory failure, loss of pulmonary surfactant, and depletion of lung phosphatidylglycerol. *J Lipid Res*. 2007;48:621–632.
252. Tsybovsky Y, Molday RS, Palczewski K. The ATP-binding cassette transporter ABCA4: structural and functional properties and role in retinal disease. *Advances in Experiment Med Biol*. 2010;703:105–125.
253. Haller JF, Cavallaro P, Hernandez NJ, et al. Endogenous β-glucocerebrosidase activity in Abca12$^{-/-}$epidermis elevates ceramide levels after topical lipid application but does not restore barrier function. *J Lipid Res*. 2014;55:493–503.
254. Zuo Y, Zhuang DZ, Han R, et al. ABCA12 maintains the epidermal lipid permeability barrier by facilitating formation of ceramide linoleic esters. *J Biol Chem*. 2008;283:36624–36635.
255. Tarling EJ. Expanding roles of ABCG1 and sterol transport. *Curr Opin Lipidol*. 2013;24:138–146.
256. Kidambi S, Patel SB. Sitosterolaemia: pathophysiology, clinical presentation and laboratory diagnosis. *J Clin Pathol*. 2008;61:588–594.

257. May P, Herz J, Bock HH. Molecular mechanisms of lipoprotein receptor signalling. *Cell Mol Life Sci*. 2005;62:2325–2338.
258. Rac ME, Safranow K, Poncyljusz W. Molecular basis of human CD36 gene mutations. *Molecul Med*. 2007;13:288–296.
259. Barter PJ, Brewer Jr HB, Chapman MJ, et al. Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2003;23:160–167.
260. Benyo Z, Gille A, Kero J, et al. GPR109A (PUMA-G/HM74A) mediates nicotinic acid-induced flushing. *J Clin Invest*. 2005;115:3634–3640.
261. Mehta JL, Sanada N, Hu CP, et al. Deletion of LOX-1 reduces atherogenesis in LDLR knockout mice fed high cholesterol diet. *Circulat Res*. 2007;100:1634–1642.
262. Lillis AP, Van Duyn LB, Murphy-Ullrich JE, et al. LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. *Physiol Rev*. 2008;88:887–918.
263. Fisher CE, Howie SE. The role of megalin (LRP-2/Gp330) during development. *Dev Biol*. 2006;296:279–297.
264. Smolenaars MM, Madsen O, Rodenburg KW, et al. Molecular diversity and evolution of the large lipid transfer protein superfamily. *J Lipid Res*. 2007;48:489–502.
265. Chang TY, Reid PC, Sugii S, et al. Niemann-Pick type C disease and intracellular cholesterol trafficking. *J Biol Chemis*. 2005;280:20917–20920.
266. Jia L, Betters JL, Yu L. Niemann-pick C1-like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport. *Ann Rev Physiol*. 2011;73:239–259.
267. Tzotzas T, Desrumaux C, Lagrost L. Plasma phospholipid transfer protein (PLTP): review of an emerging cardiometabolic risk factor. *Obes Rev*. https://pubmed.ncbi.nlm.nih.gov/19413703/.
268. Lambert G, Sjouke B, Choque B, et al. The PCSK9 decade. *J Lipid Res*. 2012;53:2515–2524.
269. Moore KJ, Freeman MW. Scavenger receptors in atherosclerosis: beyond lipid uptake. *Arterioscler Thromb Vasc Biol*. 2006;26:1702–1711.
270. Trigatti B, Covey S, Rizvi A. Scavenger receptor class B type I in high-density lipoprotein metabolism, atherosclerosis and heart disease: lessons from gene-targeted mice. *Biochem Society Transact*. 2004;32:116–120.
271. Genest J, Libby P, Gotto AMJ. Lipoprotein disorders and cardiovascular disease. In: Zipes DP, Libby P, Bonow RO, et al., eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. 7th ed. Philadelphia: Elsevier Saunders; 2005:1013–1033.
272. Bernstein DL, Hulkova H, Bialer MG, et al. Cholesteryl ester storage disease: review of the findings in 135 reported patients with an underdiagnosed disease. *J Hepatol*. 2013;58:1230–1243.
273. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem*. 2003;72:137–174.
274. Yen CL, Stone SJ, Koliwad S, et al. Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *J Lipid Res*. 2008;49:2283–2301.
275. Lamarche B, Paradis ME. Endothelial lipase and the metabolic syndrome. *Curr Opinion Lipidol*. 2007;18:298–303.
276. Santamarina-Fojo S, Gonzalez-Navarro H, Freeman L, et al. Hepatic lipase, lipoprotein metabolism, and atherogenesis. *Arteriosclerosis Thrombosis Vascu Biol*. 2004;24:1750–1754.
277. Burg JS, Espenshade PJ. Regulation of HMG-CoA reductase in mammals and yeast. *Progress Lipid Res*. 2011;50:403–410.
278. Goldstein JL, Brown MS. Progress in understanding the LDL receptor and HMG-CoA reductase, two membrane proteins that regulate the plasma cholesterol. *J Lipid Res*. 1984;25:1450–1461.
279. Yeaman SJ. Hormone-sensitive lipase--new roles for an old enzyme. *Biochem J*. 2004;379:11–22.
280. Peelman F, Vandekerckhove J, Rosseneu M. Structure and function of lecithin cholesterol acyl transferase: new insights from structural predictions and animal models. *Current Opinion Lipid*. 2000;11:155–160.
281. Otarod JK, Goldberg IJ. Lipoprotein lipase and its role in regulation of plasma lipoproteins and cardiac risk. *Curr Atheroscler Rep*. 2004;6:335–342.
282. Davies BS, Beigneux AP, Fong LG, et al. New wrinkles in lipoprotein lipase biology. *Curr Opinion Lipidol*. 2012;23:35–42.

26

Metabolic Syndrome

Neil B. Ruderman and Gerald I. Shulman

OUTLINE

Diagnosis and Prevalence, 428
History, 429
  Obesity, Type 2 Diabetes, Insulin Resistance, and Cardiovascular Disease, 429
  The Lipid Theory, 430
  Inflammation and Oxidative and Endoplasmic Reticulum Stress, 433
  Specific Tissues, 434
  AMPK and SIRT1, 436
  AMPK and Metabolic Syndrome in Humans, 436
Diagnosis, 438
Coronary Heart Disease and Type 2 Diabetes, 438
Linkage of Metabolic Syndrome to Other Disorders, 439
  Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis, 439
  Polycystic Ovarian Syndrome, 439
  Certain Cancers, 440
  Alzheimer Disease, 440
  Cushing Syndrome and Related Disorders, 440
  Lipodystrophy, 440
  Hyperalimentation, 440
Treatment of Metabolic Syndrome, 440
  Lifestyle Modification (Weight Loss and Physical Activity), 441
  Drug Therapy, 441
  Special Considerations, 441
Children and Adolescents, 442
Low Birth Weight Infants (Epigenetic Changes), 442
The Thrifty Genotype and the Increasing Incidence of Metabolic Syndrome, 442

KEY POINTS

- The metabolic syndrome is a very common and often heritable disorder associated with obesity and insulin resistance.
- It predisposes to type 2 diabetes, atherosclerostic cardiovascular disease, certain cancers, and other diseases.
- It is treatable early with diet, exercise, and pharmacologic agents, all of which may function by activating AMP-activated protein kinase.

For the purposes of this review, we will define *metabolic syndrome* as a state of metabolic dysregulation characterized by insulin resistance, hyperinsulinemia, and a predisposition to type 2 diabetes, atherosclerotic cardiovascular disease (ASCVD), hypertension, Alzheimer disease, certain cancers, and many other disorders.[1-3] Affected individuals are typically (although not always) obese (body mass index [BMI] >29 kg/m^2) or overweight (BMI >25–29 kg/m^2), or they manifest more subtle manifestations of excess adiposity, such as increases in triglyceride deposition in intraabdominal fat,[4,5] liver, and skeletal muscle (ectopic liquid). In addition, they may present with low-grade inflammation, oxidative and endoplasmic reticulum (ER) stress, evidence of mitochondrial dysfunction, and an impaired ability to exercise (deceased VO_2 max).[3,6-24] From a clinical perspective, metabolic syndrome is important for several reasons. First, it can antedate ASCVD, type 2 diabetes, and other diseases to which it predisposes by many years.[1,2,10] Second, it may be both a casual factor for these diseases and a target for their prevention and therapy. Third, it is extremely common (see the following section). A pathophysiologic depiction of metabolic syndrome and its three hypothetical clinical phases, as perceived by the authors, is presented in Fig. 26.1. The potential importance of dysregulation of many of the factors listed in phase 2, including AMP-activated protein kinase (AMPK) and the sirtuins, will be discussed later.

DIAGNOSIS AND PREVALENCE

As evident from the preceding section and Fig. 26.1, a multitude of factors are associated with metabolic syndrome. Furthermore, clinically useful tests for some of these factors are sufficiently complex or vary too much among laboratories to prove clinically useful (e.g., plasma insulin levels). Likewise, others may vary between, and to some extent even within, racial or ethnic groups (e.g., waist circumference). Thus, the diagnostic criteria shown in Table 26.1 (Adult Treatment Panel III [ATP III] guidelines) apply to White individuals of European ancestry.

Metabolic syndrome (phases 2 and 3) in adults is currently diagnosed clinically by the presence of at least three of the following criteria: fasting plasma glucose greater than 100 mg/dL; blood pressure greater than 130/80 mm Hg; dyslipidemia; plasma triglyceride greater than 150 mg/dL; high-density lipoprotein (HDL) cholesterol less than 40 mg/dL (men) or less than 50 mg/dL (women); and increased waist circumference, with precise values dependent on sex and ethnic group (see Table 26.1).[1] Based on ATP III guidelines, in 2002 in excess of 50 million individuals older than 20 years of age in the United States had this diagnosis, and the number in 2014 was probably at least 50% higher (Fig. 26.2). In addition, the prevalence of metabolic syndrome is increasing rapidly in children and adolescents, in parallel with the pandemic of obesity in these age groups.[24] Based on both its natural history[21] and studies of the offspring of patients with associated disorders, metabolic syndrome is preceded by a preclinical state characterized by

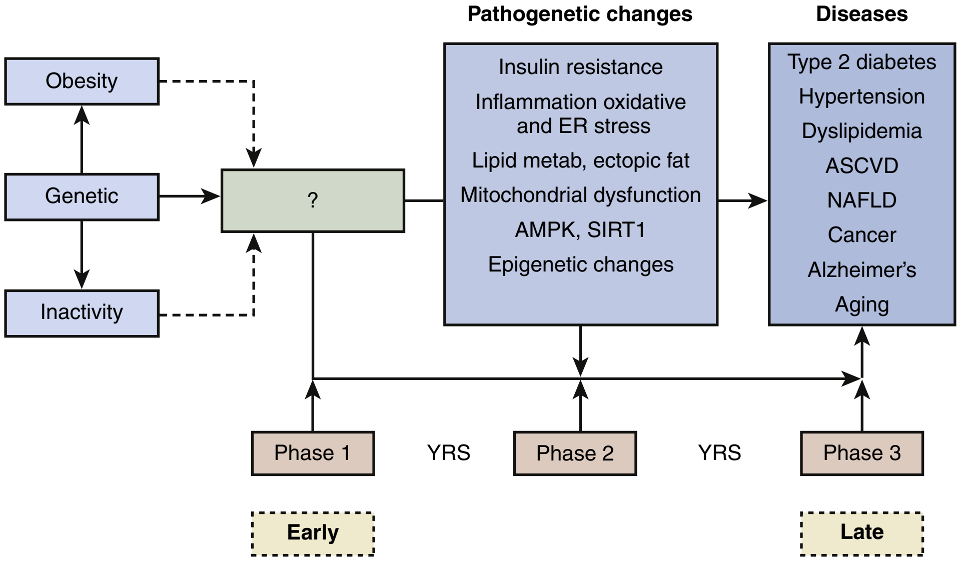


Figure 26.1 **Metabolic syndrome: a pathophysiologic overview and its three hypothetical clinical phases.** A combination of overnutrition, inactivity, and genetic and other factors interact to produce a state of metabolic dysregulation that leads to insulin resistance, ectopic lipid deposition with or without obesity, and, in many instances, inflammation, oxidative and endoplasmic reticulum (ER) stress, and other manifestations of cell dysfunction (phase 2). In individuals who are genetically predisposed, this in turn leads to one or more of the indicated disorders, many of which are associated with an increased risk for type 2 diabetes and atherosclerotic cardiovascular disease (ASCVD). Recent studies suggest that dysregulation of the AMP-activated protein kinase (AMPK) fuel-sensing and signaling network, and a closely related group of molecules, the sirtuins, may accompany these changes and could be both a causal factor and a target for their therapy. It is likely that one or more of the factors in phase 2 play a pathogenetic role in phase 1; however, this remains to be determined. Typically, systemic evidence of inflammation is not observed in phase 1, although abnormalities in mitochondrial gene expression, insulin resistance, and an altered capacity for exercise have been noted in young offspring of patients with type 2 diabetes. The transition from this preclinical phase to the clinically diagnosed metabolic syndrome (phase 2) to overt disease (phase 3) may take many years, with the rate of progression determined by environmental factors, such as diet and exercise, genetic predisposition, and pharmacotherapy for concurrent diseases.

TABLE 26.1 Adult Treatment Panel III Diagnostic Criteria for Metabolic Syndrome

Risk Factor	Defining Level
Abdominal Obesity (Waist Circumference)	
Men	>102 cm
Women	>88 cm
Triglycerides	>150 mg/dL
HDL Cholesterol	
Men	>40 mg/dL
Women	>50 mg/dL
Blood pressure	>130/80 mm Hg
Fasting glucose	>110 mg/dL*

*American Diabetes Association subsequently lowered fasting glucose level to 100 mg/dL.

HDL, High-density lipoprotein.

Data from Grundy SM, Brewer HB, Cleeman JI, et al. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association Conference on scientific issues related to definition. *Circulation.* 2004;109:433–438.

insulin resistance, hyperinsulinemia, and modest increases in plasma triglycerides, blood pressure, and glucose that by themselves may not be considered significant even in young normal-weight individuals (phase 1).

HISTORY

Obesity, Type 2 Diabetes, Insulin Resistance, and Cardiovascular Disease

The clustering of the major components of metabolic syndrome, such as obesity, type 2 diabetes, hypertension, and dyslipidemia, has long been recognized[25]; however, its delineation as a distinct entity took place only after its linkage to insulin resistance, hyperinsulinemia, and cardiovascular disease (CVD) became more apparent.[21,26,27] *Insulin resistance* has been defined as a state (of a cell, tissue, system, or body) in which greater than normal amounts of insulin are required to elicit a normal biological response.[28] In humans, it is currently diagnosed on the basis of increased plasma insulin concentrations, either fasting or during a glucose tolerance test, or by a decreased rate of glucose infusion required to maintain euglycemia during a hyperinsulinemic-euglycemic clamp.[8] The much greater prevalence of insulin resistance in patients with type 2 than type 1 diabetes was appreciated 70 years ago by Himsworth and colleagues, based on their substantially higher insulin requirement[29,30] and diminished response to exogenous insulin.[31] Shortly after the development of the insulin immunoassay by Yalow and Berson in 1960,[32] this suspicion was confirmed,[33,34] and other disorders associated with insulin resistance and hyperinsulinemia were identified, including coronary heart disease and several of its risk factors,[21,35-40] as well as obesity itself (see Fig. 26.1). In general, most adults with insulin resistance and hyperinsulinemia are obese (BMI >29 kg/m^2) or overweight (BMI 25–29 kg/m^2). However, a significant

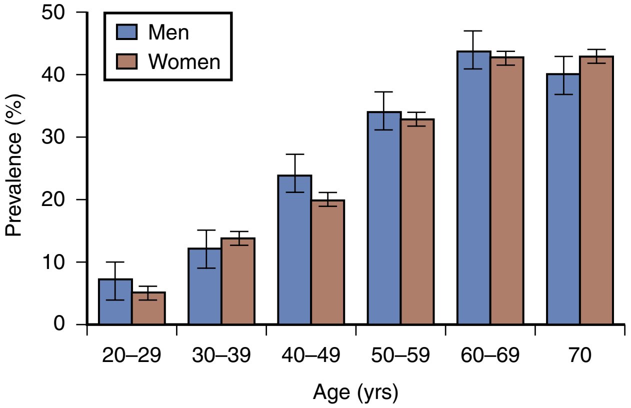


Figure 26.2 Prevalence of the metabolic syndrome in the United States according to age, based on National Health and Nutrition Examination Survey data and Adult Treatment Panel III criteria (see Table 26.1). Since these data were obtained, the criteria for abnormal glucose levels in patients with metabolic syndrome diminished from 110 mg to 100 mg. In addition, the prevalence of obesity and type 2 diabetes has increased in all of these age ranges and the general population (by as much as 50%). (Adapted from Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA.* 2002;287:356–359.)

percentage are normal-weight by BMI but show increases in visceral fat (central obesity),[4,8] increased ectopic lipid deposition in liver and skeletal muscle,[7,41] and/or enlarged fat cells.[5,42] The presence of central obesity has been shown to correlate strongly with an individual's predisposition to most of the diseases indicated in Fig. 26.1, including coronary heart disease,[43,44] and this is the reason waist circumference is one of the principal diagnostic criteria for metabolic syndrome, as discussed later.

Interestingly, hyperinsulinemia and insulin resistance are also present in normal-weight offspring of individuals with type 2 diabetes,[7,45-47] hypertension,[48] and hypertriglyceridemia[8,49,50] and in individuals at increased risk for coronary heart disease,[51-54] suggesting that they are early markers or pathogenetic factors for these disorders. Studies such as these, plus the presence of a high rate of ischemic heart disease in patients with type 2 diabetes at the time of diagnosis (20%–50%),[55-57] and to a somewhat lesser extent individuals with impaired glucose tolerance,[58] have led to the suggestion that treatment of metabolic syndrome at an early stage may be needed for preventing or at least delaying coronary heart disease.[8,22,39,59,60]

The presence of insulin resistance in otherwise normal offspring of patients with type 2 diabetes, hypertriglyceridemia, and hypertension has led to the notion that it is a causal factor for metabolic syndrome or an early pathogenetic event.[21] According to this widely held view, insulin resistance affects a number of organs (e.g., muscle, liver, adipose tissue, blood vessels), and hyperinsulinemia due to increased insulin secretion by the pancreatic β cell and decreased insulin degradation by the liver is a compensatory phenomenon.[14,21,35] The observation that therapies that increase insulin sensitivity and lower plasma insulin levels (e.g., lifestyle modification such as diet and exercise)[61-63] or treatment with metformin[61] and, to an even greater extent, the thiazolidinediones (TZDs)[64-65] prevent or delay the onset of diabetes in individuals with glucose intolerance is compatible with this notion, as is the efficacy of these therapies in some individuals with other disorders associated with metabolic syndrome, such as nonalcoholic fatty liver disease[66,67] and polycystic ovarian syndrome (PCOS).[68-70] Left unexplained by this hypothesis is the molecular mechanism by which insulin resistance develops initially and how it leads to hyperinsulinemia. Also, the possibility that hyperinsulinemia is the more primary of the two events or occurs simultaneously with the insulin resistance has not been ruled out.[70]

Hypothetically, metabolic syndrome could be related to genetic abnormalities in the insulin signaling cascade. In keeping with this possibility, mutations in *IRS1* and *IRS2,* which encode the initial targets of the insulin receptor tyrosine kinase, have been shown to lead to insulin resistance and diabetes in transgenic mice.[71,72] Evidence that these or other genetic defects in the insulin signaling cascade are common in humans with metabolic syndrome or type 2 diabetes and account for observed signaling defects[73] is still lacking, however.

KEY POINTS

- Metabolic syndrome is becoming increasingly common. As of 15 years ago, upward of 50 million individuals older than 20 years of age are affected in the United States, and the diagnosis is becoming increasingly common in children and adolescents.

The Lipid Theory

Insulin resistance and hyperinsulinemia in humans and experimental animals have been linked to obesity and dysregulation of cellular lipid metabolism in a wide variety of circumstances.[73-76] Early studies focused on fatty acid release from adipose tissue and assumed that insulin resistance occurs in skeletal muscle when plasma free fatty acid (FFA) levels are increased as a consequence of central obesity. More recently, it has become apparent that insulin resistance is associated with alterations in lipid metabolism in tissues other than skeletal muscle, and that a number of newly discovered hormones and intracellular regulatory mechanisms affect its appearance. In addition, it has been demonstrated that metabolic syndrome occurs in individuals who lack adipose tissue, as well as in those with excess adiposity, and that in both groups it is associated with triglyceride deposition in ectopic sites such as muscle, liver, and visceral fat. In this section, we will attempt to review the current status of this increasingly complex but intriguing area. We will discuss several distinct but often interrelated mechanisms that have been put forth to explain the link between altered lipid metabolism and components of metabolic syndrome.

Excess Free Fatty Acids. Over 50 years ago, Philip Randle and colleagues proposed the existence of a glucose–fatty acid cycle.[77] More specifically, they demonstrated that increased fatty acid levels diminish insulin-stimulated glucose utilization in a perfused rat heart preparation. They showed that this effect occurs within minutes, and that it is associated with enhanced mitochondrial fat oxidation that leads to both inhibition of glucose oxidation at the pyruvate dehydrogenase step and an increase in the cytosolic concentration of citrate. Randle and associates also demonstrated that the increase in citrate inhibited glycolysis at phosphofructokinase, and that the secondary increase in glucose 6-phosphate inhibited hexokinase and secondarily diminished insulin-stimulated glucose uptake[78] (Fig. 26.3). It was suggested that a similar mechanism in skeletal muscle may account for the insulin resistance observed in humans with obesity or type 2 diabetes, in both of whom plasma FFA levels were known to be elevated.[79] Over the next 25 years, most investigators were unable to reproduce these findings in skeletal muscle, however,[75,80-82] except in special circumstances.[83] For this reason, the acute contribution of elevated plasma FFA levels to the insulin resistance observed in most humans with obesity, type 2 diabetes, and other metabolic syndrome–associated disorders remained unclear.

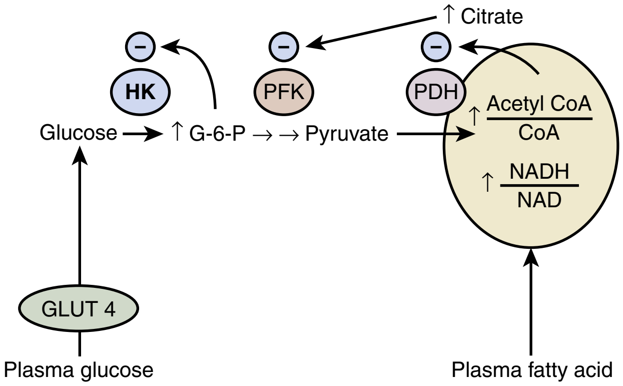


Figure 26.3 **Inhibition of glucose uptake and oxidation by fatty acids as described in heart muscle by Randle, Garland, Hales, and Newsholme (1964, 1965).** (See text for details.) (Adapted from Shulman GI: Cellular mechanisms of insulin resistance. *J Clin Invest.* 2000;106:171–176.)

This changed in 1991 when Boden and colleagues[84] definitively demonstrated that raising plasma FFAs (by infusing a lipid emulsion with heparin to activate lipoprotein lipase activity) in humans during a euglycemic-hyperinsulinemic clamp inhibits insulin-stimulated peripheral glucose uptake. Importantly, they found that this effect required 4 to 6 hours rather than a few minutes to become evident, and that it was not accompanied by an increase in citrate. Subsequent investigations by Shulman's laboratory in which ^{31}P magnetic resonance spectroscopy (MRS) was used to noninvasively measure intramyocellular glucose-6-phosphate found that its concentration was reduced under the same conditions, suggesting that fatty acids principally inhibit insulin-stimulated glucose transport or phosphorylation activity[85] and not the phosphofructokinase reaction, as suggested by Randle. Later, studies by the same group, in which ^{13}C MRS was used to assess intramyocellular glucose concentrations, revealed that insulin-stimulated glucose transport and not phosphorylation was the step inhibited by high plasma fatty acids, and that this defect was associated with lipid-induced defects in insulin-stimulated PI3 kinase activity.[86]

Other studies have demonstrated that insulin resistance in human muscle, caused by infusing lipids to increase plasma FFAs, is associated with impaired insulin signaling,[73] increases in the concentrations of muscle triglyceride, long-chain fatty acyl CoA[87] and diacylglycerol, and increases in protein kinase C (PKC) activity and the translocation of various PKCs from the cytosol to a membrane fraction.[76,88-90] Another key finding was a decrease in IKBα abundance, suggesting activation of NF-κB and proinflammatory events.[90] As discussed later, similar abnormalities have been found in rodent liver following a sustained exposure to fatty acids,[91-94] in rodent muscle in a wide variety of states associated with insulin resistance,[76,88,89] and in liver and muscle of massively obese insulin-resistant humans with type 2 diabetes.[95,96] Thus, the intracellular changes produced by an excess of fatty acids are associated with insulin resistance in many tissues.

A still unanswered question is whether an increase in plasma FFAs is an early pathogenetic event in metabolic syndrome. Elevated concentrations of plasma FFAs attributable to increased adipose tissue mass and the relative insensitivity of large fat cells and visceral fat to insulin[35,87] are present in individuals with obesity and type 2 diabetes, and they appear to contribute to insulin resistance when these disorders are established.[74,87,97,98] On the other hand, even some severely obese individuals remain insulin-sensitive (see later section entitled "Bariatric Surgery Patients: Insulin-Resistant and -Sensitive"), suggesting that other factors are involved.[99] It is also unclear whether plasma FFAs are elevated in individuals with metabolic syndrome in its very early stages. Only modest increases in plasma FFAs, if any, have been observed in normal-weight insulin-resistant individuals who are at risk for developing diabetes because of family history.[45,100]

Altered Fatty Acid Metabolism: Malonyl Coenzyme A, Mitochondrial Dysfunction, and Adenosine Monophosphate–Activated Protein Kinase. A second abnormality in lipid metabolism that could lead to insulin resistance is a disturbance in fatty acid metabolism in which the oxidation of cytosolic long-chain fatty acyl CoA (FACoA) by mitochondria is impaired, and its esterification and metabolism by other nonmitochondrial processes is enhanced.[76,101] This could occur if either the intrinsic ability of mitochondria to oxidize fatty acid is decreased or the activity of carnitine palmitoyltransferase 1, the enzyme that regulates the transfer of cytosolic long-chain fatty acyl CoA into mitochondria, is diminished.

Malonyl CoA. Altered fatty acid partitioning between the mitochondria and cytoplasm as a cause of insulin resistance was suggested by studies from the Ruderman laboratory in denervated rat muscle, in which enhanced diacylglycerol (DAG) synthesis and PKC activation were observed,[102] and later by similar findings in muscle of obese, insulin-resistant KKA^{y} mice.[103] In these and other instances, insulin resistance in rodent muscle correlated with an increase in the concentration of malonyl CoA,[76] an allosteric inhibitor of carnitine palmitoyltransferase. As shown in Fig. 26.4, an increase in malonyl CoA by decreasing the oxidation of cytosolic FACoA would increase its availability for the formation of DAG, triglycerides, ceramide, and possibly other factors linked to insulin resistance. In keeping with such a mechanism, McGarry concurrently demonstrated that treatment with etomoxir, a pharmacologic CPT1 inhibitor, both increases triglyceride accumulation and causes insulin resistance in rat skeletal muscle.[101,104] A number of other findings also support this notion: (1) mice lacking functional acetyl CoA carboxylase 2 (ACC$_2$, the principal isoform that generates the malonyl CoA that regulates CPT$_1$ in skeletal muscle) are more insulin-sensitive than control rats[105,106]; (2) the administration of an ACC inhibitor diminishes obesity and insulin resistance in fat-fed rats[107]; (3) antisense oligonucleotides directed at hepatic ACC1 and ACC2 diminish both hepatic steatosis and insulin resistance in mice fed a high-fat diet[108]; and (4) low rates of fatty acid oxidation have been reported in preobese humans,[109,110] Zucker diabetic rats,[111] and interleukin-6 (IL-6) knockout mice[112] prior to the onset of diabetes and obesity. Malonyl CoA was not assayed in muscle in any of these studies; however, in two of the rodent models[113,114] and in adipose tissue of obese insulin-resistant humans,[115,116] a decrease in the activity of AMPK was found, suggesting that its concentration was elevated (see later section entitled "AMP-Activated Protein Kinase").

Mitochondrial Dysfunction. Altered fatty acid partitioning in muscle and other tissues could also occur if fatty acid oxidation is depressed as a consequence of mitochondrial dysfunction. Decreases in mitochondrial function, and in some instances number and size, have been found in muscle of individuals with type 2 diabetes associated with obesity,[117] in lean insulin-resistant older adults,[118] and in lean insulin-resistant offspring of diabetic parents.[7] In the latter, the reduction in mitochondrial function was associated with a similar decrease in mitochondrial content.[119] Likewise, decreases in the mRNA for PGC1α, a transcriptional coactivator that enhances genes for mitochondrial biogenesis and function, have been observed in

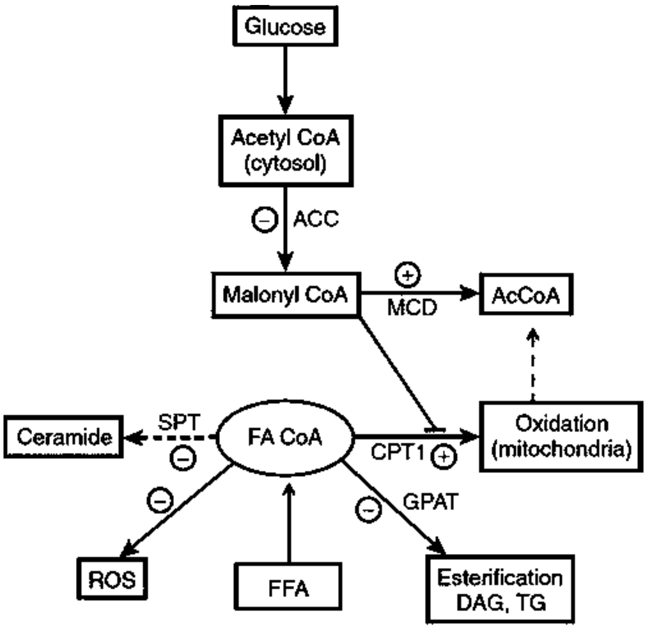


Figure 26.4 **AMP-activated protein kinase (AMPK) concurrently activates (+) or inhibits (–) multiple aspects of fatty acid and glucose metabolism.** By phosphorylating and inhibiting ACC and activating MCD, AMPK diminishes the concentration of malonyl CoA. This relieves the inhibition of CPT1 by malonyl CoA and results in both increased fatty acid oxidation and a decrease in the availability of cytosolic FACoA for TG, DAG, and ceramide synthesis, and possibly lipid peroxidation and protein acylation. Conversely, AMPK activation independently decreases the expression of glycerophosphate acyltransferase GPAT, fatty acid synthase *(not shown),* and SPT, decreasing the synthesis of glycerolipids, fatty acids (*de novo*), and ceramide (*de novo*), respectively. A decrease in AMPK produces the opposite effects. The basis for the ability of AMPK to inhibit oxidant stress (ROS generation) inflammation, and ER stress *(not shown)* in some settings is not completely understood. Whether AMPK activation enhances or inhibits a process or an enzyme in this scheme is denoted by *plus signs* and *minus signs,* respectively. Not shown in the diagram is that AMPK achieves these changes in some instances by phosphorylating the indicated enzyme and/or in other instances by regulating its expression at the level of the gene. For instance, AMPK can enhance mitochondrial biogenesis and function by enhancing the expression of the transcriptional coactivator PGC1α. *ACC,* Acetyl CoA carboxylase; *CPT1,* carnitine palmitoyltransferase 1; *DAG,* diacylglycerol; *FACoA,* cytosolic long-chain fatty acyl CoA; *GPAT,* glycerophosphate acyltransferase; *MCD,* malonyl CoA decarboxylase; *ROS,* reactive oxygen species; *SPT,* serine palmitoyl transferase. (Adapted from Ruderman N, Prentki M. AMP kinase and malonyl-CoA: targets for therapy of the metabolic syndrome. *Nat Rev Drug Discov.* 2004;3:340–351.)

some[120] but not all[119] studies of nondiabetic first-degree relatives of patients with type 2 diabetes, as well as in individuals with type 2 diabetes and impaired glucose tolerance.[121] Whether these mitochondrial changes are hereditary or secondary to metabolic events (e.g., lipotoxic changes due to abnormalities in intracellular lipid metabolism) or abnormalities in AMPK regulation (see the following section) remains to be determined. However, given the likely role for increased intracellular fatty acid metabolites in mediating insulin resistance, impaired mitochondrial function leading to decreased fatty acid oxidation will almost certainly exacerbate the problem. Also to be determined is whether the changes observed in the offspring of diabetic parents reflect a difference in muscle fiber type, because the ratio of mitochondrial-rich type 1 fibers to glycolytic type 2 fibers may be decreased in these individuals.[7,122]

BOX 26.1 AMP-Activated Protein Kinase

AMP-activated protein kinase (AMPK) is a heterotrimer containing α, β, and γ subunits, each of which has at least two isoforms. The α subunit contains the catalytic site; the β subunit contains a glycogen-binding domain; and the γ subunit contains two AMP-binding sites. All three subunits are necessary for full activity.[298,299] In general, AMPK is found in the cytosol of a cell; however, the A2 isoform of the enzyme is also present in the nucleus.[299] Decreases in the energy state of a cell that lead to increased binding of AMP and ADP in place of ATP on the γ subunit activate AMPK by a number of mechanisms, including allosteric activation and especially covalent modification due to phosphorylation of its catalytic subunit on Thr-172 as a result of the actions of the AMPK kinase LKB1. In addition, factors that increase cellular Ca^{2+} and CaMKKb also activate AMPK.[300,301] When activated, AMPK enhances a number of processes that increase ATP generation, including fatty acid oxidation and glucose transport (in skeletal muscle), and it decreases others that consume ATP, but are not acutely necessary for survival, including to varying extents fatty acid triglyceride and protein synthesis. In addition, AMPK can alter the expression of a wide variety of genes, including several that alter mitochondrial function (e.g., *PGC1a, UCP3*) and lipid synthesis *(SREBPIC).*

AMP-Activated Protein Kinase. AMPK (Box 26.1) is a fuel-sensing enzyme that appears to play a key role in regulating both cellular metabolism and mitochondrial function. In addition, an increasing body of evidence has suggested that its dysregulation could be a cause of metabolic syndrome (animal studies) as well as a target for its prevention and therapy (human and animal studies).[3,75] Furthermore, it has been shown definitively in rodents that exercise *in vivo* (treadmill running) activates AMPK not only in muscle but also in liver, adipose tissue,[123] and aortic endothelium.[124]

As shown in Fig. 26.4, when AMPK is activated (e.g., during exercise or in some tissues by caloric deprivation), it phosphorylates and inhibits acetyl CoA carboxylase, the enzyme that catalyzes the synthesis of malonyl CoA and (by a still undetermined mechanism) activates malonyl CoA decarboxylase, the enzyme that catalyzes malonyl CoA degradation.[123] In addition, in many tissues and cells, activation of AMPK concurrently inhibits the use of cytosolic FACoA for the synthesis of various factors that could lead to insulin resistance, including DAG and ceramides, and it diminishes the generation of lipid peroxides, ER stress, and activation of NF-κB caused by elevated concentrations of specific fatty acids and other factors.[75,125,126,127] Furthermore, it can do so in multiple tissues and by cells such as macrophages, which could contribute to inflammation in various tissues.[128] Thus, AMPK could protect cells and tissues against lipotoxicity by multiple mechanisms. For instance, in addition to modulating the above events acutely by phosphorylating specific proteins, AMPK chronically regulates the effects of transcriptional regulators (e.g., SREBP1C) that govern the synthesis of acetyl CoA carboxylase and other key enzymes that regulate fatty acid oxidation (see Fig. 26.4). Of specific relevance to this chapter is that decreases in AMPK activity are associated with insulin resistance in muscle (Table 26.2) and liver in many situations, and failure to activate AMPK in the fat cell during lipolysis is associated with increases in oxidative stress and inflammation.[126] In addition, hormones and other factors that decrease AMPK in peripheral tissues, such as glucocorticoids[129] and resistin,[130] cause insulin resistance, whereas hormones (e.g., adiponectin and leptin) and pharmacologic agents (e.g., metformin, TZDs, α-lipoic acid, salsalate, AICAR) that activate AMPK diminish it in experimental animals, as does exercise[75] (Table 26.3). Finally, as already noted, closely linked to mitochondrial theories of insulin resistance is PGC1α (PPARγ-coactivator 1α), a transcriptional coactivator of mitochondrial biogenesis whose expression is increased when AMPK is activated (e.g., by exercise or AICAR).[131-134]

TABLE 26.2 Abnormalities Associated With Insulin Resistance in Tissues of Humans and Experimental Animals

Model	TG	DAG	Malonyl CoA	PKC Activity	Activated IKK-NF-kB	AMPK Activity
fa/fa rat	(+)	(+)	(+)	(+)	ND	(–)
Glucose-infused rat	(+)	(+)	(+)	(+)	ND	(–)
Fat-fed rat	(+)	(+)	(+/–)	(+)	(+)	ND
Fat-infused humans	(+)	(+)	ND	(+)	(+)	ND
Obese insulin-resistant humans	(+)	ND	(+/–)	(+)	ND	(+/–)

AMPK, AMP-activated protein kinase; *DAG,* diacylglycerol; IKK, inhibitor of NF-κB kinase; *NF-kB,* nuclear factor kappa B; *PKC,* protein kinase C; *TG,* triglyceride; (+), increased; (–), decreased; *ND,* not determined.

Data are for skeletal muscle. They are taken from Ruderman N, Prentki M. AMP kinase and malonyl-CoA: targets for therapy of the metabolic syndrome. *Nat Rev Drug Discov.* 2004;3:340–351 and are from the laboratories of the authors and those of Turinsky, Kraegen, Caro, Boden, and Shoelson. Many of these changes have also been demonstrated in liver in insulin-resistant obese humans, fat-fed and glucose-infused rats, and fa/fa rats. Whether AMPK activity is depressed and the concentration of malonyl CoA elevated in skeletal muscle of obese insulin-resistant humans is controversial; however, it has been clearly observed in adipose tissue (see the section entitled "AMPK and the Metabolic Syndrome in Humans"). Studies, primarily *in vitro*, suggest that similar events occur in the liver, pancreatic β cell, and cultured vascular endothelium.

TABLE 26.3 Effect of Therapies That Activate AMP-Activated Protein Kinase in Humans (H) and/or Rodents on Various Manifestations of the Metabolic Syndrome

Factor	Insulin Resistance	Pancreatic β Cell Dysfunction	Endothelial Cell Dysfunction	Coronary Heart Disease Risk	NAFLD/NASH
Exercise (H)	(–)	ND	(–)	(–)	(–)
Calorie/weight reduction (H)	(–)	(–)	(–)	(–)	(–)
Adiponectin	(–)	(–)	(–)	(–)	(–)
AICAR	(–)	(–)	(–)	ND	ND
Leptin	(–)	(–)	(–)	ND	(–)
Metformin (H)	(–)	(–)	(–)	(–)	(–)
TZDs	(–)	(–)	(–)	(–)	(–)
Polyphenols	(–)	ND	(–)	(–)	(–)

AICAR, 5-Aminoimidazole-4-carboxamide riboside; *AMPK,* AMP-activated protein kinase; *NAFLD/NASH,* nonalcoholic fatty liver disease/nonalcoholic steatotic hepatitis; *TZDs,* thiazolidinediones; *(–),* decreased; *ND,* not determined.

Where studied, these factors also alter ectopic lipid deposition in keeping with their effects on AMPK and malonyl CoA. Inactivity, caloric excess (glucose), and deficiencies of leptin or adiponectin, where studied, have been shown to have opposite effects. Studies with AICAR and experimental polyphenols (resveratrol, SI-17834) have only been carried out in rodents. Polyphenols stimulate AMP-activated protein kinase, lower lipids, and inhibit accelerated atherosclerosis in diabetic LDL receptor-deficient mice. Resveratrol improves health and survival of mice on a high-calorie diet.

Inflammation and Oxidative and Endoplasmic Reticulum Stress

In recent years, considerable attention has been given to the theories in which the combination of oxidative and ER stress and NF-κB activation play a central role in the pathogenesis of metabolic syndrome. As noted by numerous investigators, all of these factors have been linked to insulin resistance in humans and experimental animals with a wide variety of disorders.[9,10,135-140] Because of this, it has been suggested that an abnormality of the innate immune system could be a proximal event in the pathogenesis of metabolic syndrome.[10] Although this possibility cannot be disproven, and it is unquestioned that proinflammatory events and oxidative and ER stress are often an integral component of metabolic syndrome, a number of observations including the following suggest that dysregulation of lipid metabolism that can lead to oxidative and ER stress and inflammation is likely to be a more primary factor: (1) the close correlation of metabolic syndrome in its early stages with weight gain, central obesity, ectopic lipid deposition, and elevated plasma FFA levels; (2) the presence of metabolic syndrome and ectopic lipid deposition in humans and experimental animals deficient in peripheral adipose tissue (lipodystrophy) and the reversal of these abnormalities in rodents by the implantation of fat[141-143] and in humans and/or rodents by the administration of leptin,[144,145] adiponectin,[146] and in some instances TZDs, all of which activate AMPK[75]; (3) the observation that elevating plasma FFA levels in humans and rodents by itself acutely (hours) leads to insulin resistance and proinflammatory changes[87,90,147]; and (4) the very early occurrence of alterations in cellular lipid metabolism and mitochondrial function in normal-weight normoglycemic, young offspring of diabetic parents.[7] To our knowledge, evidence of a proinflammatory state has not been reported in the latter group. These considerations aside, the possibility that proinflammatory changes leading to alterations in lipid metabolism are the cause of metabolic syndrome has not been ruled out; indeed, studies in rodents suggest that TNFα-induced inflammation can lead to a decrease in AMPK activity in skeletal muscle.[148] In addition, it has recently been demonstrated that treatment with the antiinflammatory drug salsalate improves glycemic status modestly (less than 0.24% reduction in hemoglobin $A1_c$) and reduces systemic inflammation in obese adults at risk for developing type 2 diabetes.[149] This agent has also been shown to activate AMPK, however.[150] Because therapies that target lipid metabolism, oxidative stress, and in some instances inflammatory events have all been shown to diminish insulin resistance in specific situations,[94,136-138] what is clear is that these factors are almost certainly interrelated.

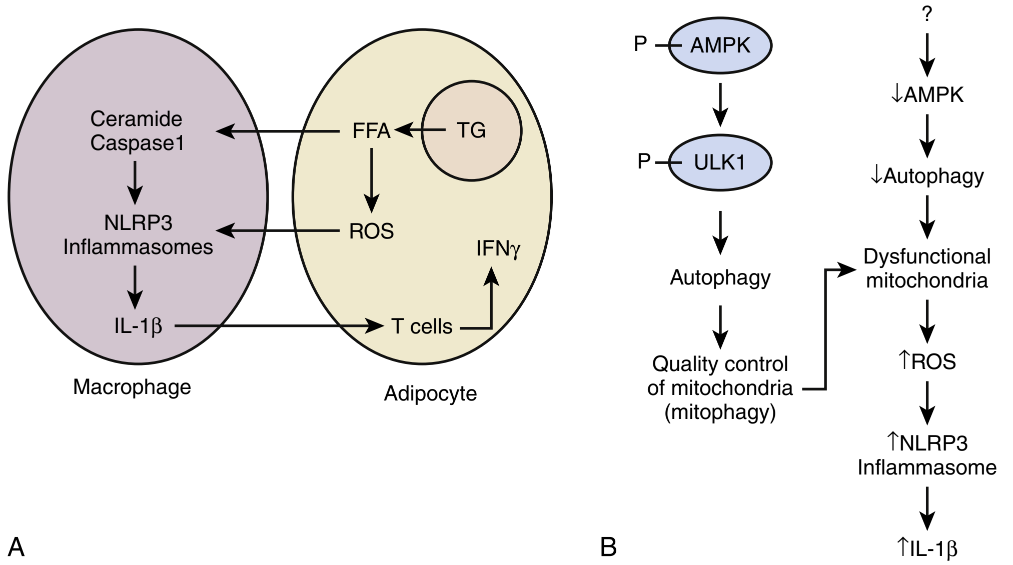

Ceramide
Caspase1
NLRP3
Inflammasomes
IL-1β
Macrophage
FFA
TG
ROS
IFNγ
T cells
Adipocyte
P
AMPK
P
ULK1
Autophagy
Quality control
of mitochondria
(mitophagy)
?
↓AMPK
↓Autophagy
Dysfunctional
mitochondria
↑ROS
↑NLRP3
Inflammasome
↑IL-1β
A
B

Figure 26.5 Hypothetical mechanisms for the activation of inflammasomes in adipose tissue of obese insulin-resistant subjects. A, It is proposed that increased free fatty acid release from lipid droplets in the adipocyte lead to changes in adjacent macrophages (ceramide, caspase 1 activation) that activate the inflammasome, which generates interleukin (IL)-1B and other inflammatory cytokines. **B,** Effects of AMP-activated protein kinase (AMPK) on events in the macrophage. The identity of the factor(s) that decrease AMPK activity in cells in which inflammasomes are found is uncertain, although oxidative stress and inflammatory cytokines such as IL-1B and TNFα could be factors. Not shown here is that activation of caspase 1 in the inflammasome complex cleaves and presumably inactivates SIRT1. ULK1, a regulator of the autophage lysosome is phosphorylated and activated by AMPK. (Adapted from Choi AM, Nakahira K. Dampening insulin signaling by an NLRP3 "meta-flammasome." *Nat Immunol.* 2001;12:379–380.)

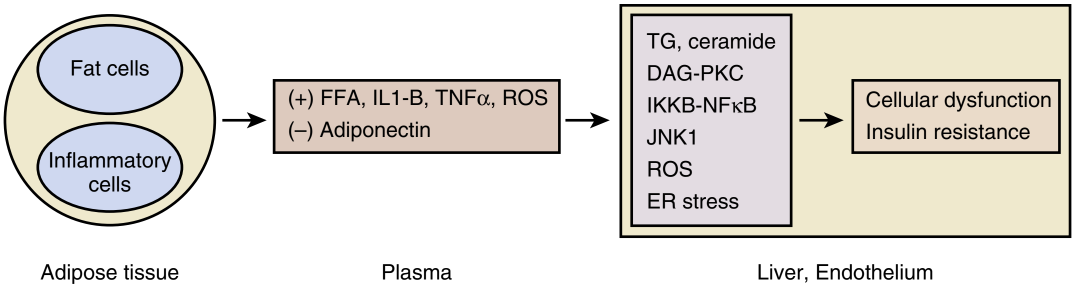

Fat cells
Inflammatory
cells
(+) FFA, IL1-B, TNFα, ROS
(–) Adiponectin
TG, ceramide
DAG-PKC
IKKB-NFκB
JNK1
ROS
ER stress
Cellular dysfunction
Insulin resistance
Adipose tissue
Plasma
Liver, Endothelium

Figure 26.6 Hypothetical mechanism by which dysregulation of adipose tissue in an obese insulin-resistant individual leads to insulin resistance and dysfunction in liver, aorta, and possibly other tissues. Adipose tissue of such patients releases increased amounts of free fatty acids, reactive oxygen species, inflammatory cytokines, and decreased amounts of adiponectin, all of which could contribute to insulin resistance and dysfunction in other tissues. AMP-activated protein kinase activity is decreased in adipose tissue of such individuals and in liver, aorta, and in some studies, muscle; however, it is uncertain whether it causes or is a result of the above changes.

Specific Tissues

Adipose Tissue. A number of lines of investigation have linked abnormalities in adipose tissue to the pathogenesis of metabolic syndrome. They also indicate that this likely reflects abnormalities both in adipocytes and in macrophages and other inflammatory cells in the stroma.[128,151] Thus, as already noted, elevated plasma FFA levels, attributable to an increase in their release from adipocytes in grossly or centrally obese individuals, correlate with the presence of insulin resistance in most patients.[87,152] In addition, when the function of the adipocyte as a store for lipid is impaired,[141,142] as it is in many insulin-resistant obese individuals,[99] it has been proposed that the excess fatty acid release could be a key event leading to the generation of the inflammatory cytokines IL-1, IL-8, TNFα, and others by stromal macrophages (Fig. 26.5). As shown in Fig. 26.6, an increase in the release of these cytokines, as well as of reactive oxygen species and FFA (from the adipocyte), are thought to be largely responsible for the cellular dysfunction and decreased AMPK observed in liver and other tissues of obese individuals with metabolic syndrome. As reviewed recently by Steinberg and Schertzer,[128] an overwhelming body of evidence indicates that AMPK is at the crossroad of metabolically driven macrophage inflammation, presumably by controlling its mitochondrial metabolism. In keeping with this conclusion, they point to multiple studies showing that decreasing AMPK activity by various means increases inflammation in the macrophage, whereas AMPK activation decreases it.[128]

Adiponectin. A deficiency of the adipokine adiponectin, also referred to as ACRP30, has been most closely linked to insulin resistance in humans. It is produced exclusively or at least predominantly in adipose tissue and is present in the circulation in trimeric, hexameric, and high–molecular-weight (HMW) forms. The biological relevance of the three oligomers[153,154] and still other forms of adiponectin is not completely understood. These considerations aside, there is abundant evidence that low immunoassayable adiponectin levels in plasma (accounted for mainly by the HMW form) are present in obese insulin-resistant individuals and in individuals at risk for type 2 diabetes[146,155] and coronary heart disease,[156] even in the absence of overt obesity. In addition, polymorphisms of the adiponectin gene have been associated with metabolic syndrome in some populations and a predisposition to type 2 diabetes in others.[146]

Adiponectin, like exercise, activates AMPK and stimulates AMPK-mediated events such as glucose transport and fatty acid oxidation in muscle[157,158] and inhibition of glucose production in liver (in rodents).[159] In addition, adiponectin has antiinflammatory actions.[160] Whether the insulin-sensitizing effect of adiponectin is AMPK-mediated has not been proven definitively; however, treatment with TZDs increases plasma adiponectin, and the insulin-sensitizing effect of these agents and their ability to activate AMPK are both markedly attenuated in adiponectin knockout mice.[161,162] Furthermore, decreased levels of adiponectin are one of the hallmarks of bariatric surgery patients who are insulin-resistant,[163] as is decreased AMPK activity in their adipose tissue.[115,116]

Leptin. Since its discovery by Friedman and colleagues,[164] interest in leptin has for the most part focused on its role as an appetite suppressant. However, it has also been recognized for some time that leptin increases oxidative metabolism and fatty acid oxidation in peripheral tissues, owing both to a direct action[165] and to an effect on the hypothalamus that appears to be mediated by the sympathetic nervous system.[166] As first reported by Minokoshi and Kahn and their coworkers, both the direct and centrally-mediated effects of leptin on peripheral tissues are associated with AMPK activation,[167] whereas its action on hypothalamic nuclei is associated with a decrease in AMPK activity.[168]

When leptin is lacking or its receptor is not functioning in peripheral tissues, lipid accumulates and cellular damage may result. Unger and associates[169,170] have reported that, in the Zucker diabetic fatty (ZDF) rat, which lacks the leptin receptor, such ectopic lipid accumulation occurs in liver, muscle, and the pancreatic β cell, and that it antedates the presence of diabetes and pancreatic β cell apoptosis. Furthermore, it has been found that AMPK is deficient in tissues of these rats,[114] and that treatment with AICAR[114,171] and exercise, which also activates AMPK in these animals,[171] prevents ectopic lipid accumulation, pancreatic β cell damage, and the development of diabetes. More recent studies have found that the antidiabetic effects of leptin may be attributed to its inhibition of the hypothalamic-pituitary-adrenal axis, leading to reduction in lipolysis and hepatic gluconeogenesis.[172]

Vascular Endothelial Cells and Atherogenesis. An impressive case has been made that atherogenesis is essentially an inflammatory response to a variety of risk factors, and that the consequences of this response include acute coronary artery and cerebrovascular syndromes.[173] An early site at which this inflammatory response appears to occur is the endothelial cell[173]; indeed, increases in NF-κB expression have been observed in endothelium at sites predisposed to atherosclerotic plaque formation[175] and *in vitro* in endothelial cells exposed to elevated concentrations of glucose[176] and fatty acids.[177] Likewise, impaired endothelium-dependent relaxation and increases in circulating adhesion molecules (VCAM1, ICAM, selectins), markers of cellular dysfunction and incipient ASCVD, have been observed in humans with type 2 diabetes and metabolic syndrome[135,178] and in normal individuals in whom plasma FFA levels are increased by a lipid infusion.[179] Conversely, endothelial cell dysfunction is diminished in humans by factors that diminish the proinflammatory state, including exercise and caloric restriction[180,181] and by treatment with TZDs.[178,182] As already noted, all of these interventions have been reported to activate AMPK in rodents. Studies with endothelial cells in culture also support such a protective role for AMPK. Thus, increases in oxidative stress and NF-κB–mediated gene expression observed in cultured human umbilical vein endothelial cells (HUVECs) incubated with palmitate are inhibited by AICAR and other AMPK activators.[125,183,184] Also, AICAR and (where studied) expression of a constitutively active AMPK have been shown to inhibit apoptosis, mitochondrial dysfunction, DAG synthesis, and the development of insulin resistance (diminished Akt activation) in HUVECs incubated in a high-glucose medium.[185] Finally, it has been demonstrated that the administration of atorvastatin, a cholesterol-lowering agent, by gavage activates AMPK and eNOS in the rat aorta,[186] and it has a similar effect when incubated with HUVECs. Whether this accounts for the antiatherogenic effect of statins attributed to its antiinflammatory action is an intriguing possibility.[187] Finally, it has long been appreciated that increases in macrophages are a hallmark of arteries undergoing atherosclerotic changes. As recently reviewed, an increasing body of evidence strongly suggests that treatments that activate AMPK combat atherosclerosis by diminishing proinflammatory events in macrophages, as they do in adipose tissue and presumably other sites.[128]

Liver. Changes similar to those in muscle and the endothelial cell occur in the liver in insulin-resistant states. Thus, as in muscle, an association between hepatic lipid deposition and insulin resistance has been clearly demonstrated in humans.[188,189] Also, in rats infused with a lipid emulsion to increase plasma FFA levels during a euglycemic-hyperinsulinemic clamp[91,147] or with short-term fat feeding,[93] the development of hepatic insulin resistance is associated with increases in DAG content, PKC activation, and a decrease in IKBα abundance—changes almost identical to those observed in human muscle.[90] Similar alterations in PKC have been noted in the livers of massively obese, insulin-resistant humans[95,190] and of fat-fed rats with hepatic steatosis.[93] Also, knockdown of PKC epsilon expression by antisense oligonucleotides protects rats from fat-induced hepatic insulin resistance.[191] As discussed later, bariatric surgery in such rodents is followed by increases in the activity of AMPK, SIRT1, and LKB1 in liver, as well as an increase in insulin sensitivity.[192]

Pancreatic β Cell. Insulin resistance in muscle and liver does not initially result in hyperglycemia, because it is accompanied by hyperinsulinemia. As noted previously, it is unclear whether the hyperinsulinemia is compensatory or results from the same factors that cause insulin resistance, in which event it may occur at the same time or even precede it.[70,193] In this context, it is noteworthy that increases in plasma FFA have been shown acutely to increase insulin secretion in certain settings, whereas chronic increases in the concentration of saturated fatty acids and glucose cause dysfunction and damage to the β cell and ultimately result in apoptosis.[75,101,194] Work from a number of laboratories[169,170,195,196] has both delineated the events that lead to these phenomena in the β cell and revealed their similarity to the changes observed in endothelium and other cells when exposed to high concentrations of FFAs or glucose. As already noted, in the ZDF rat, the leptin-receptor–deficient rodent characterized by Unger,[169] the activity of AMPK in multiple tissues is depressed, and treatment with troglitazone or AICAR or caloric deprivation prevents or at least markedly attenuates the development of β cell damage and dysfunction and hyperglycemia.[114,197] Likewise, it has been demonstrated that AMPK activation prevents the apoptosis and mitochondrial dysfunction observed in pancreatic β cells when incubated with saturated fatty acids at a high glucose concentration.[195] Finally, although theories linking triglyceride accumulation to β cell dysfunction are attractive and have led to interesting hypotheses,[22,182,198,199] to our knowledge there have been no definitive studies showing that triglyceride accumulates in human islets in patients with type 2 diabetes.

Molecular Mechanisms of Insulin Resistance and Cellular Dysfunction According to the Lipid Theory (Muscle and Liver). Based on studies reviewed in the preceding sections, a model can be proposed in which insulin resistance and cellular dysfunction in liver and muscle are due to an increase in intracellular fatty acid–derived molecules and metabolites such as fatty acyl CoA DAG[89,91,200] and ceramides, possibly secondary to dysregulation of AMPK (Fig. 26.7; see Table 26.2). According to this scheme, such changes activate a serine-threonine kinase cascade that includes conventional and/or novel protein kinase C isoforms,[89,91,201,202] IKKB,[203,204] and Jun-activated kinase, one or more of which phosphorylate serine residues on IRS-1 in muscle, and likely other tissues.[87] Similar changes in IRS1 in response to hyperglycemia may also result from activation of the mTOR/p70s6k signaling mechanism.[205] Serine phosphorylation of IRS-1 in turn impairs its ability to associate with P13-kinase, leading to diminished activation by insulin of Akt and PKC zeta, glucose transport, glycogen synthesis, and other downstream events. Studies demonstrating that transgenic mice with a muscle-specific alteration in IRS-1 Ser→Ala are protected from fat-induced insulin resistance in skeletal muscle strongly support this hypothesis.[206] Similar changes appear to occur in liver, except that the inhibitory actions of insulin on gluconeogenesis and glycogenolysis are impaired,[87,93,191] and IRS1 and IRS2 may be differentially affected.[207] Also possibly involved in this chain of events are increases in oxidative and ER stress, ceramide synthesis, NF-κB activation, and NF-κB–mediated gene expression that could explain, at least in part, the proinflammatory state associated with metabolic syndrome.[73,75,136] Interestingly, the hallmark of this insulin-resistant state is an increase in intracellular triglyceride in liver and muscle that can be quantified noninvasively with magnetic resonance imaging.[208,209] Triglyceride accumulation in muscle and liver is generally regarded as a marker of lipid-induced insulin resistance and cellular dysfunction, rather than a cause.[210] On the other hand, by providing an additional source of intracellular FFAs, it could play a more pathogenetic role.

The Hypothalamus, Food Intake, and Insulin Resistance. Leptin, which activates AMPK in peripheral tissues by a direct action, diminishes AMPK activity in the hypothalamus. This in turn leads to decreased food intake and activation of the sympathetic nervous system and, secondary to this, further activation of AMPK in peripheral tissues.[205] Conversely, glucocorticoids increase AMPK in the hypothalamus, leading to an increase in food intake, and they decrease AMPK and cause insulin resistance in peripheral tissues (see discussion of Cushing syndrome in the section entitled “AMPK and Metabolic Syndrome in Humans” later in the chapter and in Christ-Crain and associates[129]). Various antipsychotic drugs have also been found to activate AMPK in the hypothalamus, and, like glucocorticoids, they increase food intake and cause insulin resistance.[211] Why some agents that decrease or increase AMPK activity in the hypothalamus have opposite effects on peripheral tissues is not known. On the other hand, the possibility that a drug that activates or inhibits AMPK activity in peripheral tissues has the opposite effect in the hypothalamus or elsewhere in the central nervous system needs to be considered in evaluating its clinical efficacy and side effects.

KEY POINTS

- An increasing body of evidence suggests that a likely pathogenetic mechanism in most patients is an abnormality of cellular lipid metabolism that causes lipotoxic changes, including: insulin resistance, oxidant and ER stress, inflammation, and mitochondrial dysfunction in one or more tissues. Proposed causes of these abnormalities in cell lipid metabolism are elevated plasma FFA levels, dysregulation of AMPK, SIRT1, or malonyl CoA, and primary mitochondrial defects.

AMPK and SIRT1

Sirtuins are a group of histone 1 protein deacetylases that are regulated by changes in the cellular redox state ($NAD^+/NADH^+$ ratio) and increases in nicotinamide phosphoribosyltransferase (NAMPT), the rate-limiting enzyme for NAD synthesis. The seven sirtuins, and in particular SIRT1, the most studied member of the family, were evaluated initially because of their role in combatting aging.[212] As reviewed elsewhere,[213] SIRT1 responds to overfeeding, starvation, changes in energy expenditure, and exercise, as well as to adiponectin, much as does AMPK, although with somewhat different timing.[13,54,213] Interestingly, it has been shown that SIRT1 can activate AMPK by deacetylating its activator LKB1, which promotes its translocation from the nucleus to the cytosol, where LKB1 is activated, and it in turn phosphorylates and activates AMPK (which is already substantially located in the cytosol).[127,214,215] Conversely, AMPK can activate SIRT1 by increasing the $NAD^+/NADH$ ratio or the expression/activity of NAMPT.[216] These findings suggest the existence of an AMPK/SIRT1 cycle that links the cell’s energy and redox states.[217] In keeping with this notion, AMPK and SIRT1 (and probably other sirtuins) act on common transcriptional activators and coactivators, including members of the FOXO family and PGC1. In addition, both AMPK and SIRT1 activators have been shown to decrease atherosclerosis and prevent diabetes in experimental animals.[75,217,218] The latter could be related, at least in part, to the fact that changes in both AMPK and SIRT1 appear to exert effects on inflammation and fuel metabolism in M1 macrophages and T helper cells, both of which are more numerous in inflamed adipose tissue.[148] How AMPK and SIRT1 may interact with each other and with NF-κB and oxidative and ER stress in these cells to combat metabolic syndrome is illustrated in Fig. 26.8.

AMPK and Metabolic Syndrome in Humans

Although existing data in rodents with decreased AMPK activity and the beneficial effects of exercise and various pharmacologic agents that activate AMPK have linked it to metabolic syndrome in humans,[219] until recently, direct AMPK measurements in muscle of insulin-resistant humans were not strongly supportive, or yielded mixed results.[212] Since then, decreased AMPK activity has been clearly demonstrated in adipose tissue of two distinct groups of insulin-resistant individuals.

Cushing Syndrome. Korbonits and coworkers[220] observed decreased AMPK activity in the visceral fat of patients with Cushing syndrome, most of whom had elevated plasma cortisol levels due to a functioning adrenal adenoma. Such individuals are classically characterized by insulin resistance, visceral adiposity, hypertension, type 2 diabetes, and a predisposition to ASCVD and other diseases. Similar abnormalities also have been observed in patients treated with high doses of glucocorticoids for extended periods, although whether the incidence of ASCVD is increased is somewhat controversial.[221] Also of note, in contrast to most individuals with metabolic syndrome, patients with Cushing syndrome show little, if any, evidence of inflammation, and

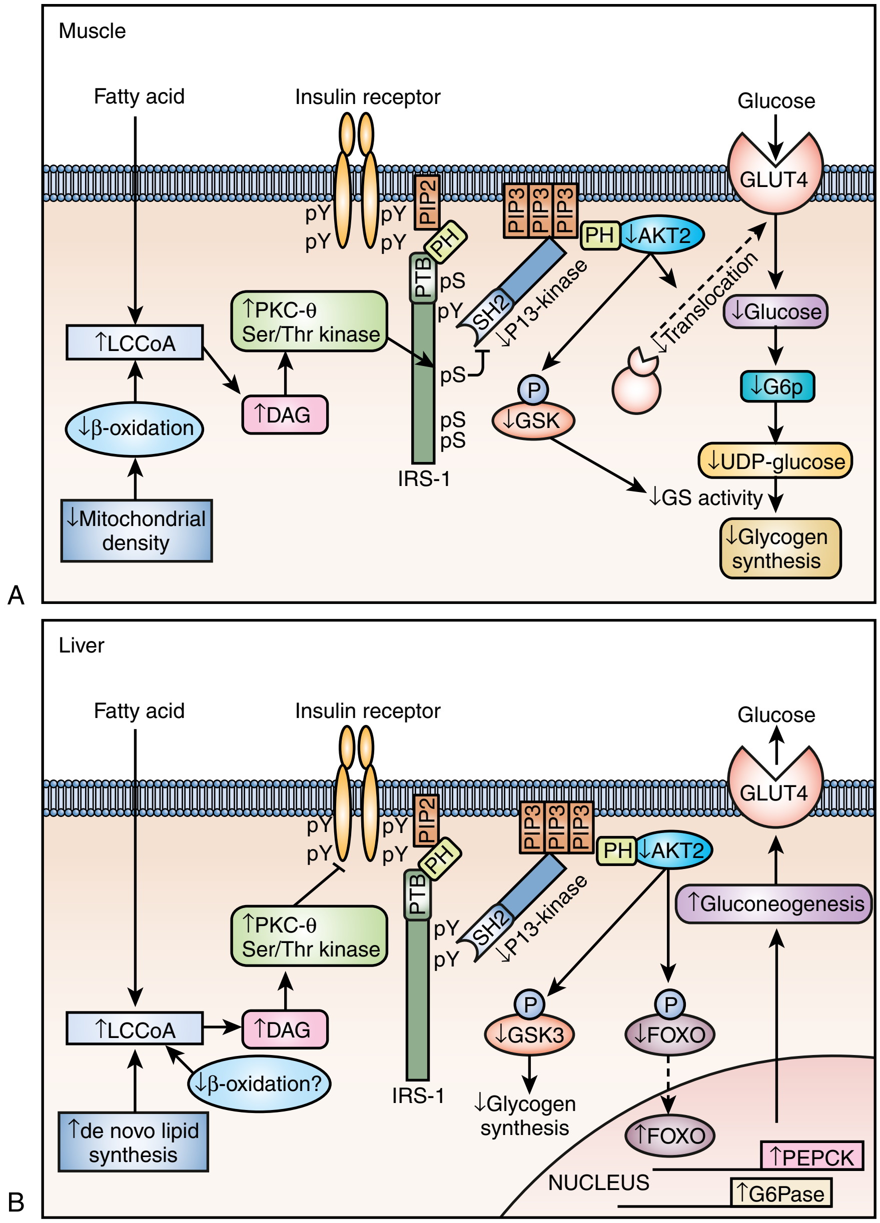

Muscle
Fatty acid
Insulin receptor
Glucose
GLUT4
pY
pY
pY
pY
PIP2
PH
PTB
PIP3
PIP3
PIP3
PH
↓AKT2
SH2
↓P13-kinase
pS
pY
pS
pS
pS
IRS-1
↑PKC-θ
Ser/Thr kinase
↑LCCoA
↑DAG
↓β-oxidation
↓Mitochondrial density
P
↓GSK
↓Translocation
↓Glucose
↓G6p
↓UDP-glucose
↓GS activity
↓Glycogen synthesis
A
Liver
Fatty acid
Insulin receptor
Glucose
GLUT4
pY
pY
pY
pY
PIP2
PH
PTB
PIP3
PIP3
PIP3
PH
↓AKT2
SH2
↓P13-kinase
pY
pY
IRS-1
↑PKC-θ
Ser/Thr kinase
↑LCCoA
↑DAG
↓β-oxidation?
↑de novo lipid synthesis
P
↓GSK3
↓Glycogen synthesis
P
↓FOXO
↑FOXO
↑Gluconeogenesis
↑PEPCK
↑G6Pase
NUCLEUS
B

Figure 26.7 **Proposed mechanisms for fatty acid induced insulin resistance in skeletal muscle (A) and liver (B).** **A,** Muscle. Increases is intramyocellular LCCoA and DAG, due to increased fatty acid delivery and/or decreased mitochondrial fatty acid oxidation, trigger a serine/threonine kinase (Ser/Thr) cascade initiated by nPKCs and possibly involving IKKB and/or JNK-1. This ultimately induces serine/threonine phosphorylation of critical IRS-1 sites in muscle, which inhibit IRS-1 tyrosine phosphorylation and activation of PI 3-kinase. This in turn results in reduced insulin-stimulated glucose transport and glycogen synthesis. **B,** Liver. Increases in intracellular DAG, due to enhanced lipogenesis and/or decreased mitochondrial fatty acid oxidation, activate PKCs (ε and/or possibly δ), which bind to and inactivate the insulin receptor kinase. This results in reduced insulin-stimulated IRS-1 and IRS-2 tyrosine phosphorylation, and decreased insulin activation of PI3-kinase and AKT2. This in turn leads to diminished GSK3 and FOXO phosphorylation, which decrease insulin-stimulated liver glycogen synthesis and hepatic gluconeogenesis, respectively. *DAG,* Diacylglycerol; *FOXO,* forkhead box protein O; *GLUT,* glucose transporter; *G6P,* glucose-6-phosphate; *GSK3,* glycogen synthase kinase 3; *IRS,* insulin receptor substrate; *IKKB,* IκB kinase β; *JNK-1,* Jun-activated kinase 1; *LCCoA,* long-chain acylcoenzyme A; *PEPCK,* phosphoenolpyruvate carboxykinase; *PI3-kinase,* phosphatidylinositol-3-kinase; *PTB,* phosphotyrosine-binding domain; *PH,* pleckstrin homology domain; *SH2,* src homology domain. (Adapted from Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiol Rev.* 2007;87:507–520.)

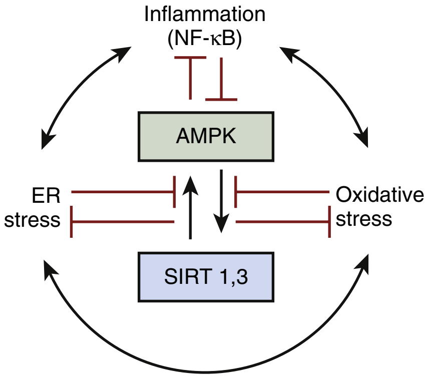


Figure 26.8 Proposed interrelations of AMP-activated protein kinase (AMPK) and sirtuins 1 and 3 (SIRTs) with oxidative and ER stress and inflammation. AMPK and SIRT1 both activate each other and diminish oxidative and endoplasmic reticulum (ER) stress and low-grade inflammation in various settings. Conversely, oxidative and ER stress and inflammation, which activate each other, appear to diminish AMPK and SIRT1. In principle, any of these factors could be targeted to combat insulin resistance and the development of metabolic syndrome–associated disorders; however, to date, the most success has been observed with therapies that target AMPK. (Adapted from Ruderman NB, Carling D, Prentki M, et al. AMPK, insulin resistance and the metabolic syndrome. *J Clin Invest.* 2013;123:2764–2772,)

to our knowledge ER stress in adipose tissue has not been studied. In what respects the lack of these abnormalities in patients with Cushing syndrome distinguishes them from most other individuals with metabolic syndrome in whom they are present has yet to be determined.

Bariatric Surgery Patients: Insulin-Resistant and -Sensitive.

Perhaps even more compelling evidence of a link between decreased AMPK in adipose tissue and metabolic syndrome–associated disorders has been provided by studies in severely obese patients undergoing bariatric surgery. Gauthier and colleagues[116] and Xu and associates[115] observed that AMPK activity is significantly diminished (30%–50%), and that chronic oxidative stress (protein carbonylation) increased in adipose tissue of the 75% of such individuals who were insulin-resistant compared with equally obese subjects who were insulin-sensitive (homeostasis model assessment insulin resistance 6.7 vs. 2.3). They also observed that AMPK activity and mitochondrial gene expression were lower in visceral fat than in subcutaneous fat in both groups, and that the expression of inflammatory genes was lower in both fat depots in the patients with insulin resistance.[115] In another study of bariatric surgery patients in which AMPK was not measured, infiltration of adipose tissue with macrophages and decreases in plasma adiponectin (an AMPK activator) were the strongest correlates of insulin resistance.[163] These studies in adipose tissue have interesting implications, because bariatric surgery reverses type 2 diabetes and other disorders associated with metabolic syndrome, including dyslipidemia, hypertension, and PCOS.[24,151] It also diminishes long-term mortality from coronary heart disease (30%–50% at 20 years),[222] and the prevalence of solid tumors (by 70% 5 years postsurgery).[151] Interestingly, the beneficial effect of the surgery on coronary heart disease was most prominent in patients in the two highest quintiles of plasma insulin concentration preoperatively.[222] As noted earlier, such individuals have been found to be insulin-resistant, with decreased AMPK activity and increased oxidative stress, inflammation, and presumably ER stress and decreased mitochondrial function in their adipose tissue.[213] Whether the decreases in AMPK and, when it occurs, SIRT1 in adipose tissue of these patients increases rapidly post–bariatric surgery is not known. Interestingly, in rats made obese by fat-feeding, substantial increases in AMPK and SIRT1 and LKB1 activity were observed in liver 9 weeks after bariatric surgery.[192] Possibly other factors contribute to differences between severely obese individuals who are insulin-sensitive and insulin-resistant, including differences in their microbiome, collagen VI deposition, and lipid droplet proteins,[213] all of which have yet to be intensively examined. With respect to the microbiome, it has been shown that germ-free mice are characterized by a substantial increase in AMPK activity,[223] suggesting that bacteria and possibly the inflammation they cause diminish AMPK activity in mammalian organisms.

DIAGNOSIS

No single definitive diagnostic test for metabolic syndrome is yet available. As discussed earlier, historically it has been diagnosed based on the presence of general or abdominal obesity, dyslipidemia, hypertension, and impaired fasting glucose or glucose intolerance in various combinations. In addition, the presence of premature coronary heart disease, type 2 diabetes in its early stages, and other disorders associated with insulin resistance have sometimes been considered diagnostic criteria. Several organizations have published standards for diagnosis, including the National Cholesterol Education Program, ATP III, and the World Health Organization (WHO).[224] They differ principally in that the WHO criteria place more emphasis on measures of insulin resistance, the presence of microalbuminuria, and the use of the glucose tolerance test, whereas the ATP III emphasizes abdominal obesity and risk factors for CVD such as dyslipidemia and hypertension. A more recent set of criteria proposed by the International Diabetes Federation is similar to that of the ATP III but with somewhat lower cutoff points for blood pressure and fasting glucose.[17] In part because of its relative simplicity, for clinical purposes the ATP guidelines (see Table 26.1) appear to be in widest use. On the other hand, they probably underestimate the prevalence of insulin resistance in the general population.[225] Also of note is that modified criteria have been developed for different ethnic groups, because metabolic syndrome is not as closely associated with obesity (based on BMI) in some groups (e.g., South Asians) as it is in Caucasians.[17,226]

CORONARY HEART DISEASE AND TYPE 2 DIABETES

The notion that metabolic syndrome, or its surrogate markers hyperinsulinemia and insulin resistance, antedate and contribute to the pathogenesis of coronary heart disease, diabetes, and at least some cases of hypertension was proposed many years ago.[21,35] Coronary heart disease in the setting of metabolic syndrome can to a great extent be attributed to dyslipidemia (increased dense low-density lipoprotein [LDL] cholesterol, diminished HDL cholesterol, and hypertriglyceridemia),[227] as well as to elevations in blood pressure and blood glucose and the presence of a procoagulant, proinflammatory state.[22,224] In addition, some studies suggest that hyperinsulinemia and insulin resistance, as well as hyperglycemia, may be independent risk factors.[51] Whether elevated FFA levels or a dysregulation of intracellular fatty acid metabolism contribute to atherosclerosis by directly altering the function of endothelium (see the section entitled "Vascular Endothelial Cells and Atherogenesis") or other cells in the vascular wall remains to be determined. Relevant to this discussion, low levels of adiponectin are associated with an increased risk for coronary heart disease in humans,[156] whereas, as noted earlier, overexpression of adiponectin or its globular subunit diminishes the severity of atherosclerosis in ApoE–/– mice.[228,229]

More definitive evidence that metabolic syndrome *per se* predisposes to coronary heart disease and cerebrovascular disease has been

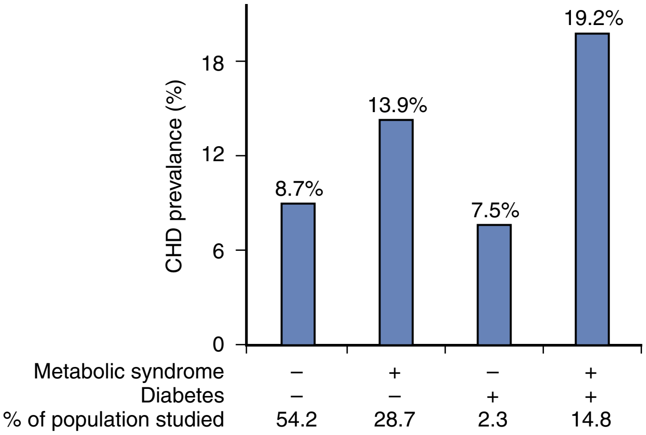


Figure 26.9 **Age-adjusted prevalence of CHD in the US population over 50 years of age, categorized by the presence of metabolic syndrome (Adult Treatment Panel III criteria) and diabetes.** Data are from the National Health and Nutrition Examination Survey study. The complete absence of an increase in CHD incidence in the diabetic patients without metabolic syndrome should be viewed with caution because of the small size of this group. *CHD,* Coronary heart disease. (Data from Alexander CM, Landsman PB, Teutsch SM, et al. NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. *Diabetes.* 2003;52:1210–1214.)

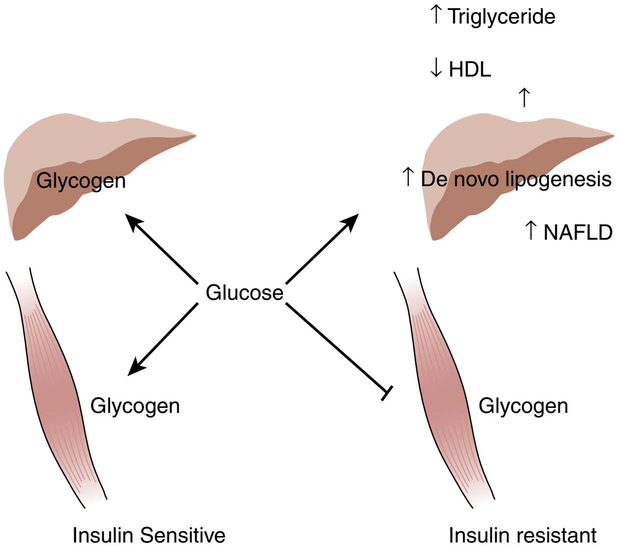


Figure 26.10 **Schematic of whole-body energy distribution after high-carbohydrate mixed meals in young insulin-sensitive and insulin-resistant individuals.** (Adapted from Petersen KF, Dufour S, Savage DB et al. The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc Natl Acad Sci USA.* 2007;104:12587–12594.)

reported. Thus, a 2-fold to 4-fold increase in subsequent cardiovascular events has been described in men and women with metabolic syndrome (modified WHO criteria) even in the absence of type 2 diabetes or impaired glucose tolerance.[230-232] Qualitatively, similar results have been obtained when metabolic syndrome is defined by ATP III criteria[23,73,234] (Fig. 26.9). In a compilation of multiple studies, the presence of metabolic syndrome had a greater impact on the risk for developing diabetes (5-fold) than ASCVD (2-fold).[22,182,199] In addition, where studied, the rate of cardiovascular events was higher in patients who had diabetes and metabolic syndrome than in individuals with only metabolic syndrome.[22,235]

KEY POINTS

- In both adolescents and adults, metabolic syndrome substantially increases the risk for type 2 diabetes, coronary heart disease, and other associated disorders.

LINKAGE OF METABOLIC SYNDROME TO OTHER DISORDERS

From a practical point of view, the ATP III and WHO guidelines focus on the relationship of metabolic syndrome to obesity and the risk for developing type 2 diabetes and ASCVD. However, insulin resistance and hyperinsulinemia are also associated with other disorders in individuals who, for genetic or other reasons (e.g., drug therapy), are more susceptible to them. A few of these disorders will be briefly discussed.

Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis

It was estimated over 10 years ago that nearly 20 million individuals in the United States have the diagnosis of nonalcoholic fatty liver disease (NAFLD), and approximately 10% of these individuals develop nonalcoholic steatohepatitis (NASH), a disorder characterized by mitochondrial dysfunction, increases in oxidative stress and cell cytokines, and a predisposition to cirrhosis (approximately 20% of patients with NASH), and less commonly hepatocellular carcinoma.[67,236-239] NAFLD is also seen with increasing frequency in children and adolescents in parallel with the increasing prevalence of obesity in these populations.[240] Presumably, mildly abnormal liver function tests in the presence of obesity or type 2 diabetes or other manifestations of insulin resistance would help to identify individuals with NAFLD at an early point in time. Recent studies in young, lean, insulin-resistant individuals have demonstrated that insulin resistance in skeletal muscle, as reflected by decreased glycogen synthesis, can promote atherogenic dyslipidemia by changing the fate of ingested carbohydrate away from skeletal muscle glycogen synthesis into hepatic *de novo* lipogenesis, resulting in an increase in plasma triglyceride concentrations and a reduction in plasma HDL concentrations (Fig. 26.10). Furthermore, insulin resistance in these subjects was independent of changes in the plasma concentrations of TNFα, IL-6, HMW adiponectin, resistin, retinol-binding protein 4, or intraabdominal obesity, suggesting that these factors do not play a primary role in causing insulin resistance in the early stages of metabolic syndrome.[241] Further evidence in support of this hypothesis comes from a recent study that found that a single bout of exercise was sufficient to promote increased muscle glycogen synthesis and decreased *de novo* lipogenesis and hepatic triglyceride synthesis following carbohydrate ingestion in young, lean individuals with muscle insulin resistance.[242]

Polycystic Ovarian Syndrome

Another common disorder that becomes more prevalent in the presence of metabolic syndrome is PCOS.[68,69] PCOS is characterized by genetically determined increases in ovarian androgen production and disordered gonadotropin secretion that may be exaggerated by hyperinsulinemia or insulin resistance. Perhaps because of its association with metabolic syndrome and often, although not always, with obesity, PCOS is a leading risk factor for glucose intolerance and type 2 diabetes in adolescent and premenopausal women.[243] It may also increase their risk for premature CVD.[244] PCOS is associated with increases in such inflammatory factors as plasminogen activator inhibitor, C-reactive protein, and TNFα.[68] Like NAFLD/NASH, type 2 diabetes, and other disorders associated with metabolic syndrome, PCOS often responds to treatments that activate AMPK and/

TABLE 26.4 Recommended Treatments for Metabolic Syndrome in Adults, Children, and Adolescents

	ADULTS		CHILDREN AND ADOLESCENTS	
Treatment	**Nondiabetic**	**IGT**	**Diabetic**	**Nondiabetic**
Diet and exercise	+	+	+	+
Metformin	ND	+	+	ND
TZDs	ND	+*	+*	ND

+, Currently recommended treatment; ND, as yet no definitive evidence for or against use.

*See text for discussion of controversy concerning TZD use.

For diabetic children and adolescents, diet and exercise are the recommended therapy. Where possible, however, metformin and insulin are often used. Data from Alberti G, Zimmet P, Shaw J, et al. Type 2 diabetes in the young: the evolving epidemic: the International Diabetes Federation Consensus Workshop. *Diabetes Care*. 2004;27:1798–1811.

or lower the concentration of malonyl CoA, such as diet and exercise, metformin, and TZDs[68] (see later section entitled “Treatment of Metabolic Syndrome”).

Certain Cancers

The reason for the link between metabolic syndrome, obesity, type 2 diabetes, and cancers of the colon, breast, liver, and other sites is less clear. It has generally been held that insulin-like growth factor 1 (IGF-1) could be one factor, because insulin both stimulates the synthesis of IGF-1 by liver and inhibits the synthesis of its binding protein IGFBP-1.[245] Another possibility relates to the fact that an upstream kinase that activates AMPK is LKB1,[246,247] a tumor suppressor that is deficient in individuals with Peutz–Jeghers syndrome and places them at increased risk for developing carcinomas of the colon, stomach, and pancreas and adenocarcinomas of the lung. The possibility that LKB1 and AMPK link metabolic syndrome to these cancers and are targets for their prevention or therapy warrants consideration.[248-250] Of note, agents that activate AMPK, such as metformin, have been shown to diminish the growth of cancer cells.[251]

Alzheimer Disease

Alzheimer disease and other disorders associated with dementia are more common in individuals with metabolic syndrome and type 2 diabetes. As reviewed elsewhere,[252] Alzheimer disease appears to be associated with insulin resistance as well as inflammation and mitochondrial dysfunction in the brain. Also of note, physical activity, which has been shown to activate AMPK in many tissues in rodents,[123] has demonstrated efficacy for preventing and treating Alzheimer disease in both humans and experimental animals.[3,252] Likewise, benefits from treatment with TZDs and metformin have been reported in some studies.[253]

Cushing Syndrome and Related Disorders

As already noted, patients with primary Cushing syndrome or Cushing syndrome due to therapy with glucocorticoids typically demonstrate central obesity, insulin resistance, hyperinsulinemia, and a predisposition to diabetes and hypertension.[254] In addition, like other patients with this clustering of events, they may be at increased risk for ASCVD.[255,256] The value of treatments aimed at AMPK and malonyl CoA in individuals with Cushing syndrome, when its primary cause cannot be corrected, has to our knowledge not been evaluated. It is noteworthy, however, that glucocorticoids diminish AMPK in peripheral tissues in rodents and, as discussed earlier, AMPK activity is decreased in visceral fat of patients with Cushing syndrome.[129] The possibility that an increase in cellular 11-β-dehydrogenase activity resulting in local increases in cortisol could be a cause of metabolic syndrome has been discussed elsewhere.[257]

Lipodystrophy

Patients with primary lipodystrophy and lipodystrophy secondary to drug therapy (e.g., protease inhibitors in HIV patients) appear to be subject to the same lipotoxicity observed in individuals with metabolic syndrome for other reasons. Two factors appear to contribute to an insulin-resistant state in these patients: (1) decreased peripheral fat cells cause more plasma FFA and lipoprotein triglyceride to be shunted to ectopic sites, and (2) plasma leptin and adiponectin levels are very low. The importance of the latter is suggested by successes in treating some patients with leptin[144,145] or with TZDs.[258] Interestingly, in the latter study, the beneficial effect of rosiglitazone was accompanied by an increase in adiponectin but not in subcutaneous fat mass.

Hyperalimentation

As suggested by Unger,[259] hyperalimentation can cause lipotoxic damage and, by inference, metabolic syndrome in some individuals. He hypothesized that at one extreme are normal individuals who eat an excess number of calories for an extended period and become hyperinsulinemic and insulin-resistant when the ability of their adipose tissue to deposit the excess lipid in their diet is exceeded.[260] At the other extreme are hyperalimented patients with an extensive loss of subcutaneous adipose tissue due to third-degree burns. Such burn patients are historically very insulin-resistant, which can be attributed at least in part to increased plasma concentrations of inflammatory cytokines, such as TNFα and IL-6. It remains to be determined if the burns contribute to their insulin resistance and whether the latter responds to pharmacologic agents that enhance insulin sensitivity.

Additional Disorders. Metabolic syndrome has also been associated with an increased prevalence of such disorders as gout, sleep apnea, gallstones, and chronic kidney disease.[199] Like other metabolic syndrome–related disorders, they are associated with obesity, type 2 diabetes, and a predisposition to ASCVD.

KEY POINTS

- The prevalence of metabolic syndrome is markedly increased in individuals who lack adipose tissue, as well as in individuals who are obese. In both populations, a common occurrence is ectopic lipid deposition in muscle, liver, and often visceral fat.

TREATMENT OF METABOLIC SYNDROME

The demonstration that metabolic syndrome increases the risk for developing other disorders in both otherwise normal individuals with type 2 diabetes strongly suggests that it warrants treatment (Table

26.4). Less clear are which therapies should be used in a given circumstance and when they should be started.

Lifestyle Modification (Weight Loss and Physical Activity)

The treatment of metabolic syndrome with the aim of preventing CVD has been reviewed by a clinical conference jointly sponsored by the American Heart Association, the National Heart, Lung and Blood Institute, and the American Diabetes Association (AHA/NHLBI/ADA). The consensus was that lifestyle modifications consisting of diet for the treatment of obesity and overweight and physical activity were the first line of therapy.[224,261,262] The efficacy of this approach for preventing disease has not been assessed specifically in patients with metabolic syndrome diagnosed by ATP III or WHO criteria; however, exercise has been shown to reverse the defects in insulin-stimulated glucose transport and muscle glycogen synthesis in young, lean, insulin-resistant offspring of parents with type 2 diabetes,[46] and in several prospective studies[61-63] the combination of diet and exercise has proven quite effective in delaying or preventing the onset of diabetes in patients with impaired glucose tolerance (most of whom probably had metabolic syndrome). Also, numerous epidemiologic studies have shown a 30% to 50% decrease in the risk for developing coronary heart disease, as well as type 2 diabetes, with the maintenance of a physically active compared with a sedentary lifestyle.[262] It must be emphasized, however, that the incidence of diabetes in the treated patients with impaired glucose tolerance in these studies was still higher than in the general population,[61,63] suggesting that, to be maximally effective, lifestyle changes may have to be introduced even earlier. Whether lifestyle changes have a similar effect on coronary heart disease is less certain. However, in follow-up studies of patients in the US and Finnish diabetes prevention program who were insulin-resistant, diet and exercise diminished nontraditional risk factors for coronary heart disease, such as C-reactive protein and fibrinogen, as well as some of the more classic risk factors associated with metabolic syndrome.[263,264]

Drug Therapy

When recommended therapeutic goals are not achieved with diet and exercise, pharmacologic therapy is necessary. The recommendations of the AHA/NHLBI/ADA at different stages of metabolic syndrome, based on Framingham risk scores for developing ASCVD, have been presented elsewhere.[22,235] Suffice it to say, they vary with an individual's risk for developing coronary heart disease and cerebrovascular disease over given time intervals. Thus, in an individual already at risk for these disorders because of diabetes, the use of statins and other agents to diminish plasma cholesterol has a lower goal (LDL cholesterol <100 mg/dL) than in an individual with isolated hypercholesterolemia. Likewise, certain antihypertensive agents such as angiotensin-converting enzyme inhibitors[265] and fibric-acid derivatives[266] may be of special benefit in patients with type 2 diabetes or at high risk for developing it.

In patients with diabetes, metformin and TZDs have been routinely used because of their blood glucose-lowering actions and, in the case of the TZDs, their insulin-sensitizing effect (see Table 26.4). In regard to metformin's mechanism of action, it is well established that metformin decreases fasting plasma glucose and insulin concentrations through inhibition of hepatic gluconeogenesis.[269] Early studies suggested this might be related to its ability to activate AMPK in the liver[268]; however, recently this effect was attributed to inhibition by metformin of mitochondrial glycerophosphate dehydrogenase activity resulting in an increase in the cytosolic redox state and decreased conversion of lactate and glycerol to glucose.[269] To what extent metformin and other oral agents should be used as second-line therapy for metabolic syndrome in the absence of diabetes or impaired glucose tolerance is still a matter of debate. In 2004, an AHA/NHLBI/ADA conference on metabolic syndrome[261] concluded that "there is insufficient evidence to recommend these drugs for anything other than their glucose-lowering action at this time." On the other hand, as already noted, both metformin and TZDs have shown some efficacy in treating NAFLD and PCOS, and in the United Kingdom prospective diabetes study[270] chronic metformin therapy was associated with a decrease in new cardiovascular events in obese individuals with type 2 diabetes that was still evident 10 years after this initial finding.[271] Likewise, effects of metformin on diminishing the growth of cancers[272] and slowing the progression of Alzheimer disease have been described, as have activation of AMPK by metformin in endothelial cells[273] and human adipose tissue.[274,275]

The picture with respect to the TZDs is less clear. TZDs have shown efficacy equal to that of lifestyle change and better than that of metformin in diminishing progression from impaired glucose tolerance to overt diabetes,[61,65] and they have shown utility in treating NAFLD.[239] However, a controversial metaanalysis[276] concluded that the TZD rosiglitazone increases the incidence of myocardial infarction in patients with established type 2 diabetes. In contrast, in a second metaanalysis by the same group[277] and in a large prospective study,[278] a different TZD, pioglitazone, was found to cause a modest decrease in cardiovascular events and/or death due to myocardial infarction. Pioglitazone has also been demonstrated to decrease the progression of atherosclerosis and diminish plaque size in patients with type 2 diabetes.[279] In part based on these findings, a special panel of the Joint American and European Diabetes Associations recommended that rosiglitazone not be used for the therapy of type 2 diabetes.[280] Whether pioglitazone should be used for this purpose is also uncertain, because it appears to predispose patients to osteoporosis[281] and, in a small percentage of individuals, cancer of the bladder. As recently reviewed, these findings have led to a concentrated effort to develop new TZDs that do not have these effects.[282]

The role of statin therapy in patients with metabolic syndrome is somewhat clearer. A subgroup analysis of statin trials revealed that this class of agents reduces the risk for cardiovascular events in individuals with metabolic syndrome.[282] To what extent this is because of their apolipoprotein B–containing lipoproteins and/or their antiinflammatory action[187,283] is uncertain. Recently it has been demonstrated that treatment with rosuvastatin (for 1.9 years) reduces the incidence of major cardiovascular events by 50% in apparently healthy 60- to 70-year-old individuals, and that it does so independently of hyperlipidemia and other indices of metabolic syndrome and of plasma cholesterol.[187] Presumably, statins would have the same or an even greater effect in individuals with metabolic syndrome and evidence of inflammation; however, this remains to be proven. On the other hand, the finding that statins increase the incidence of type 2 diabetes in a small subgroup of patients has raised some questions about its use.[284] The NHLBI/AHA/ADA panel[261] recommended that cigarette smoking be discouraged in all individuals with metabolic syndrome. It also suggested that low-dose aspirin for primary prevention of coronary heart disease in patients with metabolic syndrome is promising,[282] as may be therapy with salsalate, an aspirin derivative, which, as noted earlier, also activates AMPK.[150] The subject of treatment of metabolic syndrome, based on the risk for coronary heart disease by Framingham criteria, has recently been reviewed.[22,285]

Special Considerations

The diagnosis and treatment of metabolic syndrome presents some special obstacles. With upwards of 50 million individuals with this diagnosis in the United States, it is a massive public health problem. In some individuals, diabetes, coronary heart disease, liver disease, and hypertension are already present, and in others, the central problem is how to prevent these disorders from occurring. Finally, there are many therapeutic options at our disposal; however, some, such as diet and exercise, are often difficult to utilize and maintain successfully, and others, such as drug therapy, are expensive, and in some populations (e.g., children and adolescents) they are untested and/or available data are conflicting. Also, as already noted, the potential side effects of some of these agents need to be considered.

The magnitude of the problem is such that the prevention of diseases associated with metabolic syndrome will almost certainly require governmental, societal, and individual change. In such an effort, a primary target will be the prevention of obesity and central adiposity through increasing physical activity and promoting healthy eating.[18] It has recently been suggested that, for children at risk for developing type 2 diabetes, school-based programs and governmental actions that focus on lifestyle will be required.[18] It is highly likely that similar programs will be required for adults.

With respect to drug therapy, we already have an array of agents that are at least partially successful in treating insulin resistance and other components of metabolic syndrome (e.g., the proinflammatory state) and diminishing diseases that can emanate from it, including type 2 diabetes and coronary heart disease. As we gain a better understanding of the pathogenesis of metabolic syndrome, it is likely that newer agents will be developed and that we will discover how to use currently available drugs (e.g., metformin, TZDs, and statins) or drugs derived from them more effectively.

KEY POINTS

- Lifestyle modifications, consisting of diet to induce weight loss and exercise, have been shown to diminish many of the abnormalities of metabolic syndrome and to prevent or delay progression from impaired glucose tolerance to type 2 diabetes in prospective studies. Because of this and their safety and low cost, diet and exercise are the first line of therapy for children and adolescents with metabolic syndrome and for adults without associated diseases. The role of metformin, "insulin-sensitizing" agents such as TZDs in individuals who are not hyperglycemic remains to be determined.

CHILDREN AND ADOLESCENTS

Recent studies indicate that the prevalence of metabolic syndrome is increasing dramatically in children and adolescents in parallel with the increase in obesity and type 2 diabetes. In adolescents 12 to 19 years of age in the National Health and Nutrition Examination Survey study (1988–1994), metabolic syndrome, diagnosed by ATP III criteria, was present in 4% of the total population, 6.8% of those classified as overweight (BMI 85th to 95th percentile), and 28% of those categorized as obese (BMI >95th percentile).[286] Furthermore, because the incidence of obesity until recently has been increasing in the years since these data were obtained, these are likely underestimates of its prevalence at the present time.

Where studied, metabolic syndrome in children and adolescents has been associated with insulin resistance, central adiposity, dyslipidemia, elevations in blood pressure, and increases in intramyocellular and intrahepatic lipids, much as in adults, although with different absolute measurements.[19,20] In addition, as in adults, the prevalence of NAFLD[240] and PCOS[68] are increased, as are plasma levels of the proinflammatory markers C-reactive protein and IL-6, and immunoassayable adiponectin in plasma is diminished.[20] Also, the prevalence of CVD in these individuals later in life may be increased.[287-289] Current treatment recommendations for children with type 2 diabetes, most of whom have metabolic syndrome, have been discussed critically by an International Diabetes Federation Consensus Workshop.[18]

LOW BIRTH WEIGHT INFANTS (EPIGENETIC CHANGES)

It has been known for some time that a low birth weight, independent of gestational age, predicts the development of type 2 diabetes, coronary heart disease, central obesity, and other aspects of metabolic syndrome in middle age in some individuals.[290,291] Recent reports suggest that the greatest risk is in infants with low birth weight who gain weight rapidly in childhood.[292] It has been suggested that an altered intrauterine environment (e.g., poor nutrition) could be responsible for the low birth weight in these children; however, the precise nature of these alterations is not known. Because of the long period between the malnutrition and the development of disorders, many years later it was suggested that epigenetic changes may be involved. This question was subsequently investigated in individuals prenatally exposed to food deprivation during the Dutch famine at the end of World War II (1944–1945). Six decades later, it was found that those individuals who were malnourished during the first trimester of pregnancy showed less DNA methylation of the maternally imprinted insulin-like growth factor 2 (*IGF2*) gene compared with their unexposed same-sex sibilings.[293] Interestingly, in middle age, these individuals were predisposed to obesity, altered lipid profiles, CVD, and accelerated cognitive aging.[294]

THE THRIFTY GENOTYPE AND THE INCREASING INCIDENCE OF METABOLIC SYNDROME

The recent increase in the prevalence of metabolic syndrome and type 2 diabetes has been attributed to the epidemic of obesity that began in the second half of the twentieth century and continues unabated. As reviewed elsewhere,[295] this is almost certainly a reflection of environmental factors, most notably the increased availability of food (and possibly the higher carbohydrate content of our diet over the past 50 years) and the decrease in physical activity that occurred in many industrialized societies during this time period.

In 1962, Neel[296] raised the question of why diabetes (type 2), which has adverse effects on health, persisted in humans during the course of evolution. He suggested that it could be related to the existence of a "thrifty gene" that predisposes individuals to obesity, and secondarily diabetes, but had survival value for our hunter-gatherer ancestors. More specifically, he proposed that in the feast-famine environment of early humans "individuals exceptionally efficient in the uptake and utilization of food and its storage as fat" would have had a selective advantage. Neel further suggested that, in our modern environment of calorie surplus, such a gene (or genes) that protected us from death when food sources were scarce may make us prone to obesity and diabetes (and metabolic syndrome).

Over the years, many candidate thrifty genes have been proposed, including *UCP2, UCP3,* and those encoding B_3-adrenergic receptors. Genes governing the AMPK/malonyl CoA fuel-sensing network, which has been linked to the regulation of both food intake[105,168,297] and energy expenditure,[106,111] have also been suggested. This is an attractive notion, because a decreased ability to oxidize fatty acids appears to be characteristic of preobese humans.[76,110] For the same reason, genes that regulate mitochondrial biogenesis also need to be considered.[7] Decreased mitochondrial content/activity, be it acquired or inherited, may promote obesity by two mechanisms. First, by decreasing daily energy expenditure, even by a small amount, it could lead to progressive weight gain. For instance, a 50-calorie/day reduction in energy expenditure in a typical adult, if not accompanied by a comparable decrease in food intake, would lead to a weight gain of approximately 5 lb/year. Second, a decrease in mitochondrial activity in skeletal muscle, by predisposing an individual to altered cellular lipid metabolism (as reflected by an increase in intramyocellular triglyceride), like alterations in the AMPK/malonyl CoA network,[75] can lead to insulin resistance and hyperinsulinemia.[118] The latter, in turn, could promote obesity by increasing lipogenesis in liver and adipose tissue and by inhibiting lipolysis until a new, more obese, steady state is achieved.

ACKNOWLEDGMENTS

This work was supported in part by USPHS grants DK19514, DK49147, P01-HL-68758, R01 DK-40936, R01 DK-49230, R24 DK-085638, P30 DK-45735, U24 DK-059635, a grant from the Novo Nordisk Foundation for Basic Metabolic Research, and a grant from the Juvenile Diabetes Research Foundation. The authors thank Christina Nielsen-Campbell and Laura DelloStritto for their assistance in preparing this chapter.

- For your free Expert Consult eBook with bibliographic citations as well as the ability to take notes, highlight important content, search the full text, and more, visit http://www.ExpertConsult.Inkling.com.

REFERENCES

1. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet*. 2005;365:1415–1428.
2. Reaven G. Metabolic syndrome: pathophysiology and implications for management of cardiovascular disease. *Circulation*. 2002;106:286–288.
3. Richter EA, Ruderman NB. AMPK and the biochemistry of exercise: implications for human health and disease. *Biochem J*. 2009;418:261–275.
4. Kissebah AH, Vydelingum N, Murray R, et al. Relation of body fat distribution to metabolic complications of obesity. *J Clin Endocrinol Metab*. 1982;54:254–260.
5. Bernstein RS, Grant N, Kipnis DM. Hyperinsulinemia and enlarged adipocytes in patients with endogenous hyperlipoproteinemia without obesity or diabetes mellitus. *Diabetes*. 1975;24:207–213.
6. Ruderman NB, Schneider SH, Berchtold P. The "metabolically-obese," normal-weight individual. *Am J Clin Nutr*. 1981;34:1617–1621.
7. Petersen KF, Dufour S, Befroy D, et al. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med*. 2004;350:664–671.
8. Ruderman N, Chisholm D, Pi-Sunyer X, et al. The metabolically obese, normal-weight individual revisited. *Diabetes*. 1998;47:699–713.
9. Festa A, D'Agostino Jr R, Howard G, et al. Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS). *Circulation*. 2000;102:42–47.
10. Pickup JC. Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. *Diabetes Care*. 2004;27:813–823.
11. Schmidt MI, Duncan BB, Sharrett AR, et al. Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities Study): a cohort study. *Lancet*. 1999;353:1649–1652.
12. Juhan-Vague I, Alessi MC, Vague P. Increased plasma plasminogen activator inhibitor 1 levels. A possible link between insulin resistance and atherothrombosis. *Diabetologia*. 1991;34:457–462.
13. Sakkinen P, Geffken D, Cushman M, et al. Association between physical activity and markers of inflammation in a healthy elderly population. *Am J Epidemiol*. 2001;153:242–250.
14. Reaven GM. Insulin resistance/compensatory hyperinsulinemia, essential hypertension, and cardiovascular disease. *J Clin Endocrinol Metab*. 2003;88:2399–2403.
15. Grundy SM, Cleeman JI, Daniels SR, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation*. 2005;112:2735–2752.
16. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA*. 2002;287:356–359.
17. Grundy SM. Metabolic syndrome pandemic. *Arterioscler Thromb Vasc Biol*. 2008;28:629–636.
18. Alberti G, Zimmet P, Shaw J, et al. Type 2 diabetes in the young: the evolving epidemic: the International Diabetes Federation Consensus Workshop. *Diabetes Care*. 2004;27:1798–1811.
19. Weiss R, Dufour S, Taksali SE, et al. Prediabetes in obese youth: a syndrome of impaired glucose tolerance, severe insulin resistance, and altered myocellular and abdominal fat partitioning. *Lancet*. 2003;362:951–957.
20. Weiss R, Dziura J, Burgert TS, et al. Obesity and the metabolic syndrome in children and adolescents. *N Engl J Med*. 2004;350:2362–2374.
21. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*. 1988;37:1595–1607.
22. Grundy SM. Metabolic syndrome: a multiplex cardiovascular risk factor. *J Clin Endocrinol Metab*. 2007;92:399–404.
23. Kahn R, Buse J, Ferrannini E, et al. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the study of diabetes. *Diabetes Care*. 2005;28:2289–2304.
24. Beltrán-Sánchez H, Harhay MO, Harhay MM, Mcelligott S. Prevalence and trends of metabolic syndrome in the adult U.S. population, 1999-2010. *J Am Coll Cardiol*. 2013;62:697–703.
25. Dieterle P, Fehm H, Stroder W, et al. Asymptomatic diabetes mellitus in hypertensive patients of normal weight. Glucose tolerance and serum levels of insulin and nonesterified fatty acids in essential hypertension. *Ger Med Mon*. 1968;13:478–483.
26. Modan M, Halkin H, Almog S, et al. Hyperinsulinemia. A link between hypertension obesity and glucose intolerance. *J Clin Invest*. 1985;75:809–817.
27. Stern MP. The insulin resistance syndrome. In: Alberti KGMM, ed. *International Textbook of Diabetes Mellitus*. New York: John Wiley & Sons; 2004:255–283.
28. Flier JS. An overview of insulin resistance. In: Moller DE, ed. *Insulin resistance*. Chichester: Wiley;1993:1–8.
29. Himsworth H. Diabetes mellitus: its differentiation into insulin-sensitive and insulin-insensitive types. *Lancet*. 1936;1:127.
30. Himsworth HP, Kerr RB. Insulin sensitive and insulin-insensitive types of diabetes mellitus. *Clin Sci*. 1939;4:119–152.
31. Himsworth HP. The syndrome of diabetes mellitus and its causes. *Lancet*. 1949;1:465–473.
32. Yalow RS, Berson SA. Immunoassay of endogenous plasma insulin in man. *J Clin Invest*. 1960;39:1157–1175.
33. Berson SA, Yalow RS. Plasma insulin in health and disease. *Am J Med*. 1961;31:874–881.
34. Karam JH, Grodsky GM, Forsham PH. Excessive insulin response to glucose in obese subjects as measured by immunochemical assay. *Diabetes*. 1963;12:197–204.
35. Bjorntorp P. The relationship between obesity and diabetes. In: Alberti KGMM, ed. *International textbook of diabetes mellitus*. New York: John Wiley & Sons;2004:612–627.
36. Davidson PC, Albrink MJ. Insulin resistance in hyperglyceridemia. *Metabolism*. 1965;14:1059–1070.
37. DeFronzo RA, Ferrannini E. Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care*. 1991;14:173–194.
38. Malherbe C, Heller F, de Gasparo M, et al. Insulin response during prolonged glucose infusion. *J Clin Endocrinol Metab*. 1970;30:535–538.
39. Ruderman NB, Berchtold P, Schneider S. Obesity-associated disorders in normal-weight individuals: some speculations. *Int J Obes*. 1982;6:151–157.
40. Ruderman NB, Haudenschild C. Diabetes as an atherogenic factor. *Prog Cardiovasc Dis*. 1984;26:373–412.
41. Petersen KF, Dufour S, Feng J, et al. Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. *Proc Natl Acad Sci U S A*. 2006;103:18273–18277.
42. Bays H, Mandarino L, DeFronzo RA. Role of the adipocyte, free fatty acids, and ectopic fat in pathogenesis of type 2 diabetes mellitus: peroxisomal proliferator-activated receptor agonists provide a rational therapeutic approach. *J Clin Endocrinol Metab*. 2004;89:463–478.
43. Lapidus L, Bengtsson C, Larsson B, et al. Distribution of adipose tissue and risk of cardiovascular disease and death: a 12 year follow up of

participants in the population study of women in Gothenburg. *Sweden, Br Med J (Clin Res Ed)*. 1984;289:1257–1261.
44. Larsson B, Svardsudd K, Welin L, et al. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. *Br Med J (Clin Res Ed)*. 1984;288:1401–1404.
45. Beck-Nielsen H, Groop LC. Metabolic and genetic characterization of prediabetic states. Sequence of events leading to non-insulin-dependent diabetes mellitus. *J Clin Invest*. 1994;94:1714–1721.
46. Perseghin G, Price TB, Petersen KF, et al. Increased glucose transport-phosphorylation and muscle glycogen synthesis after exercise training in insulin-resistant subjects. *N Engl J Med*. 1996;335:1357–1362.
47. Rothman DL, Magnusson I, Cline G, et al. Decreased muscle glucose transport/phosphorylation is an early defect in the pathogenesis of non-insulin-dependent diabetes mellitus. *Proc Natl Acad Sci U S A*. 1995;92:983–987.
48. Ferrari P, Weidmann P, Shaw S, et al. Altered insulin sensitivity, hyperinsulinemia, and dyslipidemia in individuals with a hypertensive parent. *Am J Med*. 1991;91:589–596.
49. Ruderman NB, Schneider SH, Berchtold P. The "metabolically-obese," normal-weight individual. *Am J Clin Nutr*. 1981;34:1617–1621.
50. Werbin B, Tamir I, Heldenberg D, et al. Immunoreactive insulin and glucose response to oral glucose in offspring of patients with endogenous hypertriglyceridaemia. *Clin Chim Acta*. 1977;76:35–40.
51. Despres JP, Lamarche B, Mauriege P, et al. Hyperinsulinemia as an independent risk factor for ischemic heart disease. *N Engl J Med*. 1996;334:952–957.
52. Ducimetiere P, Eschwege E, Papoz L, et al. Relationship of plasma insulin levels to the incidence of myocardial infarction and coronary heart disease mortality in a middle-aged population. *Diabetologia*. 1980;19:205–210.
53. Pyorala K. Relationship of glucose tolerance and plasma insulin to the incidence of coronary heart disease: results from two population studies in Finland. *Diabetes Care*. 1979;2:131–141.
54. Welborn TA, Wearne K. Coronary heart disease incidence and cardiovascular mortality in Busselton with reference to glucose and insulin concentrations. *Diabetes Care*. 1979;2:154–160.
55. Nesto RW, Phillips RT, Kett KG, et al. Angina and exertional myocardial ischemia in diabetic and nondiabetic patients: assessment by exercise thallium scintigraphy. *Ann Intern Med*. 1988;108:170–175.
56. UKPDS. Complications of newly diagnosed type 2 diabetes patients and their association with different clinical and biochemical risk factors. *Diab Res*. 1990;13:1–11.
57. Uusitupa MI, Niskanen LK, Siitonen O, et al. Ten-year cardiovascular mortality in relation to risk factors and abnormalities in lipoprotein composition in type 2 (non-insulin-dependent) diabetic and non-diabetic subjects. *Diabetologia*. 1993;36:1175–1184.
58. Jarrett RJ, McCartney P, Keen H. The Bedford survey: ten year mortality rates in newly diagnosed diabetics, borderline diabetics and normoglycaemic controls and risk indices for coronary heart disease in borderline diabetics. *Diabetologia*. 1982;22:79–84.
59. Haffner SM, Stern MP, Hazuda HP, et al. Cardiovascular risk factors in confirmed prediabetic individuals. Does the clock for coronary heart disease start ticking before the onset of clinical diabetes? *JAMA*. 1990;263:2893–2898.
60. Panzer C, Brieke A, Ruderman NB. Prevention of type 2 diabetes and its macrovascular complications: whom, when and how should we treat? *Curr Opin Endocrinol Diabetes*. 2003;10:219–236.
61. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346:393–403.
62. Pan XR, Li GW, Hu YH, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance, The Da Qing IGT and Diabetes Study. *Diabetes Care*. 1997;20:537–544.
63. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*. 2001;344:1343–1350.
64. Buchanan TA, Xiang AH, Peters RK, et al. Preservation of pancreatic beta-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk Hispanic women. *Diabetes*. 2002;51:2796–2803.
65. Gerstein HC, Yusuf S, Bosch J, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet*. 2006;368:1096–1105.
66. Neuschwander-Tetri BA, Brunt EM, Wehmeier KR, et al. Improved nonalcoholic steatohepatitis after 48 weeks of treatment with the PPAR-gamma ligand rosiglitazone. *Hepatology*. 2003;38:1008–1017.
67. Neuschwander-Tetri BA, Caldwell SH. Nonalcoholic steatohepatitis: summary of an AASLD Single Topic Conference. *Hepatology*. 2003;37:1202–1219.
68. Dhindsa G, Bhatia R, Dhindsa M, et al. Insulin resistance, insulin sensitization and inflammation in polycystic ovarian syndrome. *J Postgrad Med*. 2004;50:140–144.
69. Sam S, Dunaif A. Polycystic ovary syndrome: syndrome XX? *Trends Endocrinol Metab*. 2003;14:365–370.
70. McGarry JD. What if Minkowski had been ageusic? An alternative angle on diabetes. *Science*. 1992;258:766–770.
71. Almind K, Accili D, Kahn CR. Knockout mice as a tool to the understanding of diabetes. In: LeRoith D, Taylor SI, Olefsky JM, eds. *Diabetes mellitus: a fundamental and clinical text*. Philadelphia: Lippincott;2003:245–254.
72. White MF. IRS proteins and the common path to diabetes. *Am J Physiol Endocrinol Metab*. 2002;283:E413–E422.
73. Shulman GI. Cellular mechanisms of insulin resistance. *J Clin Invest*. 2000;106:171–176.
74. Reaven GM. The fourth musketeer-from Alexandre Dumas to Claude Bernard. *Diabetologia*. 1995;38:3–13.
75. Ruderman N, Prentki M. AMP kinase and malonyl-CoA: targets for therapy of the metabolic syndrome. *Nat Rev Drug Discov*. 2004;3:340–351.
76. Ruderman NB, Saha AK, Vavvas D, et al. Malonyl-CoA, fuel sensing, and insulin resistance. *Am J Physiol*. 1999;276:E1–E18.
77. Randle PJ, Garland PB, Hales CN, et al. The glucose fatty-acid cycle. Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet*. 1963;1:785–789.
78. Randle PJ, Newsholme EA, Garland PB. Regulation of glucose uptake by muscle. 8. Effects of fatty acids, ketone bodies and pyruvate, and of alloxan-diabetes and starvation, on the uptake and metabolic fate of glucose in rat heart and diaphragm muscles. *Biochem J*. 1964;93:652–665.
79. Randle PJ, Garland PB, Newsholme EA, et al. The glucose fatty acid cycle in obesity and maturity onset diabetes mellitus. *Ann N Y Acad Sci*. 1965;131:324–333.
80. Randle PJ. Regulatory interactions between lipids and carbohydrates: the glucose fatty acid cycle after 35 years. *Diabetes Metab Rev*. 1998;14:263–283.
81. Ruderman NB, Toews CJ, Shafrir E. Role of free fatty acids in glucose homeostasis. *Arch Intern Med*. 1969;123:299–313.
82. Ruderman NB, Houghton CRS, Hems RA. Evaluation of the isolated perfused rat hindquarter for the study of muscle metabolism. *Biochem J*. 1971;124:639–651.
83. Ruderman NB, Saha AK, Vavvas D, et al. Malonyl CoA as a metabolic switch and a regulator of insulin sensitivity. In: Richter EA, Galbo H, Kiens B, et al., eds. *Skeletal muscle metabolism in exercise and diabetes*. Plenum;1998:263–270.
84. Boden G, Jadali F, White J, et al. Effects of fat on insulin-stimulated carbohydrate metabolism in normal men. *J Clin Invest*. 1991;88:960–966.
85. Roden M, Price TB, Perseghin G, et al. Mechanism of free fatty acid-induced insulin resistance in humans. *J Clin Invest*. 1996;97:2859–2865.
86. Dresner A, Laurent D, Marcucci M, et al. Effects of free fatty acids on glucose transport and IRS-1-associated phosphatidylinositol 3-kinase activity. *J Clin Invest*. 1999;103:253–259.
87. Boden G, Shulman GI. Free fatty acids in obesity and type 2 diabetes: defining their role in the development of insulin resistance and beta-cell dysfunction. *Eur J Clin Invest*. 2002;32:14–23.
88. Ellis BA, Poynten A, Lowy AJ, et al. Long-chain acyl-CoA esters as indicators of lipid metabolism and insulin sensitivity in rat and human muscle. *Am J Physiol Endocrinol Metab*. 2000;279:E554–E560.
89. Griffin ME, Marcucci MJ, Cline GW, et al. Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. *Diabetes*. 1999;48:1270–1274.

90. Itani SI, Ruderman NB, Schmieder F, et al. Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and IkappaB-alpha. *Diabetes*. 2002;51:2005–2011.
91. Boden G, She P, Mozzoli M, et al. Free fatty acids produce insulin resistance and activate the proinflammatory nuclear factor-kappaB pathway in rat liver. *Diabetes*. 2005;54:3458–3465.
92. Lam TK, van de Werve G, Giacca A. Free fatty acids increase basal hepatic glucose production and induce hepatic insulin resistance at different sites. *Am J Physiol*. 2003;284:E281–E290.
93. Samuel VT, Liu ZX, Qu X, et al. Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J Biol Chem*. 2004;279:32345–32353.
94. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. *J Clin Invest*. 2006;116:1793–1801.
95. Considine R, Nyce MR, Allen LE, et al. Protein kinase C is increased in the liver of humans and rats with non-insulin-dependent diabetes mellitus: an alteration not due to hyperglycemia. *J Clin Invest*. 1995;95: 2938–2944.
96. Cortright RN, Azevedo Jr JL, Zhou Q, et al. Protein kinase C modulates insulin action in human skeletal muscle. *Am J Physiol Endocrinol Metab*. 2000;278:E553–E562.
97. Boden G. Free fatty acids as a target for therapy. *Curr Opin Endocrinol Diab*. 2004;11:258–263.
98. Reaven G. Insulin resistance and its consequences type 2 diabetes mellitus and the insulin resistance syndrome. In: Leroith D, Taylor SI, Olefsky JM, eds. *Diabetes melliltus: a fundamental and clinical text*. Lippincott Williams & Wilkins;2003:899–915.
99. Puri V, Ranjit S, Konda S, et al. Cidea is associated with lipid droplets and insulin sensitivity in humans. *Proc Natl Acad Sci U S A*. 2008;105:7833–7838.
100. Perseghin G, Ghosh S, Gerow K, et al. Metabolic defects in lean nondiabetic offspring of NIDDM parents: a cross-sectional study. *Diabetes*. 1997;46:1001–1009.
101. McGarry JD. Banting lecture 2001: dysregulation of fatty acid metabolism in the etiology of type 2 diabetes. *Diabetes*. 2002;51:7–18.
102. Heydrick SJ, Ruderman NB, Kurowski TG, et al. Enhanced stimulation of diacylglycerol and lipid synthesis by insulin in denervated muscle. Altered protein kinase C activity and possible link to insulin resistance. *Diabetes*. 1991;40:1707–1711.
103. Saha AK, Kurowski TG, Colca JR, et al. Lipid abnormalities in tissues of the KKAy mouse: effects of pioglitazone on malonyl-CoA and diacylglycerol. *Am J Physiol*. 1994;267:E95–E101.
104. Dobbins RL, Szczepaniak LS, Bentley B, et al. Prolonged inhibition of muscle carnitine palmitoyltransferase-1 promotes intramyocellular lipid accumulation and insulin resistance in rats. *Diabetes*. 2001;50:123–130.
105. Abu-Elheiga L, Matzuk MM, Abo-Hashema KAH, et al. Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science*. 2001;291:2613–2616.
106. Choi CS, Savage DB, Abu-Elheiga L, et al. Continuous fat oxidation in acetyl-CoA carboxylase 2 knockout mice increases total energy expenditure, reduces fat mass, and improves insulin sensitivity. *Proc Natl Acad Sci U S A*. 2007;104:1648–16485.
107. Harwood Jr HJ, Petras SF, Shelly LD, et al. Isozyme-nonselective N-substituted bipiperidylcarboxamide acetyl-CoA carboxylase inhibitors reduce tissue malonyl-CoA concentrations, inhibit fatty acid synthesis, and increase fatty acid oxidation in cultured cells and in experimental animals. *J Biol Chem*. 2003;278:3709–37111.
108. Savage DB, Choi CS, Samuel VT, et al. Reversal of diet-induced hepatic steatosis and hepatic insulin resistance by antisense oligonucleotide inhibitors of acetyl-CoA carboxylases 1 and 2. *J Clin Invest*. 2006;116:817–824.
109. Ravussin E, Swinburn BA. Energy expenditure and obesity. *Diab Rev*. 1996;4:403–422.
110. Ruderman NB, Saha AK, Kraegen EW. Minireview: malonyl CoA, AMP-activated protein kinase, and adiposity. *Endocrinology*. 2003;144:5166–5171.
111. Etgen GJ, Oldham BA. Profiling of Zucker diabetic fatty rats in their progression to the overt diabetic state. *Metabolism*. 2000;49:684–688.
112. Faldt J, Wernstedt I, Fitzgerald SM, et al. Reduced exercise endurance in interleukin-6-deficient mice. *Endocrinology*. 2004;145:2680–2686.
113. Kelly M, Keller C, Avilucea PR, et al. AMPK activity is diminished in tissues of IL-6 knockout mice: the effect of exercise. *Biochem Biophys Res Commun*. 2004;320:449–454.
114. Yu X, McCorkle S, Wang M, et al. Leptinomimetic effects of the AMP kinase activator AICAR in leptin resistant rats: prevention of diabetes and ectopic lipid deposition. *Diabetologia*. 2004;47:2012–2021.
115. Xu XJ, Gauthier MS, Hess DT, et al. Insulin sensitive and resistant obesity in humans: AMPK activity, oxidative stress, and depot-specific changes in gene expression in adipose tissue. *J Lipid Res*. 2012;53:792–801.
116. Gauthier MS, O'Brien EL, Bigornia S, et al. Decreased AMP-activated protein kinase activity is associated with increased inflammation in visceral adipose tissue and with whole-body insulin resistance in morbidly obese humans. *Biochem Biophys Res Commun*. 2011;404:382–387.
117. Kelley DE, He J, Menshikova EV, et al. Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes*. 2002;51:2944–2950.
118. Petersen KF. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science*. 2003;300:1140–1142.
119. Morino K, Petersen KF, Dufour S, et al. Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. *J Clin Invest*. 2005;115:3587–3593.
120. Patti ME, Butte AJ, Crunkhorn S, et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A*. 2003;100:8466–8471.
121. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately down-regulated in human diabetes. *Nat Genet*. 2003;34:267–273.
122. Nyholm B, Mengel A, Nielsen S, et al. Insulin resistance in relatives of NIDDM patients: the role of physical fitness and muscle metabolism. *Diabetologia*. 1996;39:813–822.
123. Park H, Kaushik VK, Constant S, et al. Coordinate regulation of malonyl-CoA decarboxylase, sn-glycerol-3-phosphate acyltransferase, and acetyl-CoA carboxylase by AMP- activated protein kinase in rat tissues in response to exercise. *J Biol Chem*. 2002;277:32571–32577.
124. Cacicedo JM, Gauthier MS, Lebrasseur NK, et al. Acute exercise activates AMPK and eNOS in the mouse aorta. *Am J Physiol Heart Circ Physiol*. 2011;301:H1255–H1265.
125. Ruderman NB, Cacicedo JM, Itani S, et al. MalonylCoA and AMP-activated protein Kinase (AMPK): possible links between insulin resistance in muscle and early endothelial cell damage in diabetes. *Biochem Soc Trans*. 2003;31:202–206.
126. Gauthier MS, Miyoshi H, Souza SC, et al. AMP-activated protein kinase is activated as a consequence of lipolysis in the adipocyte: potential mechanism and physiological relevance. *J Biol Chem*. 2008;283:1651–16524.
127. Lan F, Cacicedo J, Ruderman N, Ido Y. SIRT1 modulation of the acetylation status, cytosolic localization, and activity of LKB1. Possible role in AMP-activated protein kinase activation. *J Biol Chem*. 2008;283:27628–27635.
128. Steinberg GR, Schertzer J. AMPK promotes macrophage fatty acid oxidative metabolism to mitigate inflammation: implications for diabetes and cardiovascular disease. *Immunol Cell Biol*. 2014;92:240–345.
129. Christ-Crain M, Kola B, Lolli F, et al. AMP-activated protein kinase mediates glucocorticoid-induced metabolic changes: a novel mechanism in Cushing's syndrome. *FASEB J*. 2008;22:1672–1683.
130. Banerjee RR, Rangwala SM, Shapiro JS, et al. Regulation of fasted blood glucose by resistin. *Science*. 2004;303:1195–1198.
131. Bergeron R, Ren JM, Cadman KS, et al. Chronic activation of AMP kinase results in NRF-1 activation and mitochondrial biogenesis. *Am J Physiol Endocrinol Metab*. 2001;281:E1340–E1346.
132. Jager S, Handschin C, St-Pierre J, et al. AMP-activated protein Kinase (AMPK) action in skeletal muscle via direct phosphorylation of PGC-1alpha. *Proc Natl Acad Sci U S A*. 2007;104:12017–12022.
133. Suwa M, Nakano H, Kumagai S. Effects of chronic AICAR treatment on fiber composition, enzyme activity, UCP3, and PGC-1 in rat muscles. *J Appl Physiol*. 2003;95:960–968.
134. Zong H, Ren JM, Young LH, et al. AMP kinase is required for mitochondrial biogenesis in skeletal muscle in response to chronic energy deprivation. *Proc Natl Acad Sci U S A*. 2002;99:15983–15987.

135. Meigs JB, Hu FB, Rifai N, et al. Biomarkers of endothelial dysfunction and risk of type 2 diabetes mellitus. *JAMA*. 2004;291:1978–1986.
136. Evans JL, Goldfine ID, Maddux BA, et al. Oxidative stress and stress-activated signaling pathways: a unifying hypothesis of type 2 diabetes. *Endocr Rev*. 2002;23:599–622.
137. Gregor ME, Yang L, Fabbrini E, et al. Endoplasmic reticulum stress is reduced in tissue of obese subjects after weight loss. *Diabetes*. 2009;58:693–700.
138. Salminen A, Hyttinen JM, Kaarniranta K. AMP-activated protein kinase inhibits NF-κB signaling and inflammation: impact on healthspan and lifespan. *J Mol Med*. 2011;89:667–676.
139. Odegaard JI, Chawla A. Pleiotropic actions of insulin resistance and inflammation in metabolic homeostasis. *Science*. 2013;339:172–177.
140. Tack CJ, Stienstra R, Joosten LA, et al. Inflammation links excess fat to insulin resistance: the role of the interleukin-1 family. *Immunol Rev*. 2012;249:239–252.
141. Gavrilova O, Marcus-Samuels B, Graham D, et al. Surgical implantation of adipose tissue reverses diabetes in lipoatrophic mice. *J Clin Invest*. 2000;105:271–278.
142. Kim JK, Gavrilova O, Chen Y, et al. Mechanism of insulin resistance in A-ZIP/F-1 fatless mice. *J Biol Chem*. 2000;275:8456–8460.
143. Garg A. Acquired and inherited lipodystrophies. *N Engl J Med*. 2004;350:1220–1234.
144. Oral EA, Simha V, Ruiz E, et al. Leptin-replacement therapy for lipodystrophy. *N Engl J Med*. 2002;346:570–578.
145. Petersen KF, Oral EA, Dufour S, et al. Leptin reverses insulin resistance and hepatic steatosis in patients with severe lipodystrophy. *J Clin Invest*. 2002;109:1345–1350.
146. Scherer PE. Adipose tissue: from lipid storage compartment to endocrine organ. *Diabetes*. 2006;55:1537–1545.
147. Lam TK, Yoshii H, Haber CA, et al. Free fatty acid-induced hepatic insulin resistance: a potential role for protein kinase C-delta. *Am J Physiol Endocrinol Metab*. 2002;283:E682–E691.
148. Steinberg GR, Michell BJ, van Denderen BJ, et al. Tumor necrosis factor alpha-induced skeletal muscle insulin resistance involves suppression of AMP-kinase signaling. *Cell Metab*. 2006;4:465–474.
149. Fleischman A, Shoelson SE, Bernier R, et al. Salsalate improves glycemia and inflammatory parameters in obese young adults. *Diabetes Care*. 2008;31:289–294.
150. Hawley SA, Fullerton MD, Ross FA, et al. The ancient drug salicylate directly activates AMP-activated protein kinase. *Science*. 2012;336:918–922.
151. Xu XJ, et al. What distinguishes adipose tissue of severely obese humans who are insulin sensitive and resistant? *Curr Opin Lipidol*. 2013;24:49–56.
152. Saloranta C, Groop L. Interactions between glucose and FFA metabolism in man. *Diab Metab Rev*. 1996;12:15–36.
153. Rajala MW, Scherer PE. Minireview: The adipocyte—at the crossroads of energy homeostasis, inflammation, and atherosclerosis. *Endocrinology*. 2003;144:3765–3773.
154. Tsao TS, Tomas E, Murrey HE, et al. Role of disulfide bonds in Acrp30/adiponectin structure and signaling specificity. Different oligomers activate different signal transduction pathways. *J Biol Chem*. 2003;278:50810–50817.
155. Lindsay RS, Funahashi T, Hanson RL, et al. Adiponectin and development of type 2 diabetes in the Pima Indian population. *Lancet*. 2002;360:57–58.
156. Pischon T, Girman CJ, Hotamisligil GS, et al. Plasma adiponectin levels and risk of myocardial infarction in men. *JAMA*. 2004;291:1730–1737.
157. Tomas E, Tsao TS, Saha AK, et al. Enhanced muscle fat oxidation and glucose transport by ACRP30 globular domain: acetyl-CoA carboxylase inhibition and AMP-activated protein kinase activation. *Proc Natl Acad Sci U S A*. 2002;99:16309–16313.
158. Yamauchi T, Kamon J, Minokoshi Y, et al. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nat Med*. 2002;8:1288–1295.
159. Combs TP, Berg AH, Obici S, et al. Endogenous glucose production is inhibited by the adipose-derived protein Acrp30. *J Clin Invest*. 2001;108:1875–1881.
160. Kadowaki T, Yamauchi T. Adiponectin and adiponectin receptors. *Endocr Rev*. 2005;26:439–451.
161. Kubota N, Terauchi Y, Kubota T, et al. Pioglitazone ameliorates insulin resistance and diabetes by both adiponectin-dependent and -independent pathways. *J Biol Chem*. 2006;281:8748–8755.
162. Nawrocki AR, Rajala MW, Tomas E, et al. Mice lacking adiponectin show decreased hepatic insulin sensitivity and reduced responsiveness to peroxisome proliferator-activated receptor gamma agonists. *J Biol Chem*. 2006;281:2654–2660.
163. Kloting N, et al. Insulin-sensitive obesity. *Am J Physiol Endocrinol Metab*. 2010;299:E506–E515.
164. Zhang Y, Proenca R, Maffei M, et al. Positional cloning of the mouse obese gene and its human homologue. *Nature*. 1994;372:425–432.
165. Muoio DM, Seefeld K, Witters LA, et al. AMP-activated kinase reciprocally regulates triacylglycerol synthesis and fatty acid oxidation in liver and muscle: evidence that sn-glycerol-3-phosphate acyltransferase is a novel target. *Biochem J*. 1999;338:783–791.
166. Friedman JM, Halaas JL. Leptin and the regulation of body weight in mammals. *Nature*. 1998;395:763–770.
167. Minokoshi Y, Kim YB, Peroni OD, et al. Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *Nature*. 2002;415:339–343.
168. Minokoshi Y, Alquier T, Furukawa N, et al. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature*. 2004;428:569–574.
169. Unger RH. Lipotoxic diseases. *Annu Rev Med*. 2002;53:319–336.
170. Unger RH, Orci L. Diseases of liporegulation: new perspective on obesity and related disorders. *FASEB J*. 2001;15:312–321.
171. Pold R, Jensen LS, Jessen N, et al. Long-term AICAR administration and exercise prevents diabetes in ZDF rats. *Diabetes*. 2005;54:928–934.
172. Perry RJ, Zhang XM, Zhang D, et al. Leptin reverses diabetes by suppression of the hypothalamic-pituitary-adrenal axis. *Nat Med*. 2014;20:759–763.
173. Pearson TA, Mensah GA, Alexander RW, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*. 2003;107:499–511.
174. Libby P. Inflammation in atherosclerosis. *Nature*. 2002;420:868–874.
175. Collins T, Cybulsky MI. NF-kappaB: pivotal mediator or innocent bystander in atherogenesis? *J Clin Invest*. 2001;107:255–264.
176. Nishikawa T, Edelstein D, Du XL, et al. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. *Nature*. 2000;404:787–790.
177. Hennig B, Meerarani P, Ramadass P, et al. Fatty acid-mediated activation of vascular endothelial cells. *Metabolism*. 2000;49:1006–1013.
178. Haffner SM, Greenberg AS, Weston WM, et al. Effect of rosiglitazone treatment on nontraditional markers of cardiovascular disease in patients with type 2 diabetes mellitus. *Circulation*. 2002;106:679–684.
179. Steinberg HO, Baron AD. Vascular function, insulin resistance and fatty acids. *Diabetologia*. 2002;45:623–634.
180. Esposito K, Pontillo A, Di Palo C, et al. Effect of weight loss and lifestyle changes on vascular inflammatory markers in obese women: a randomized trial. *JAMA*. 2003;289:1799–1804.
181. Hamdy O, Ledbury S, Mullooly C, et al. Lifestyle modification improves endothelial function in obese subjects with the insulin resistance syndrome. *Diabetes Care*. 2003;26:2119–2125.
182. Hu FB. *Obesity and Cardiovascular Disease Obesity, Epidemiology*. New York: Oxford University Press;2009.
183. Cacicedo JM, Yagihashi N, Keaney Jr JF, et al. AMPK inhibits fatty acid-induced increases in NF-kappaB transactivation in cultured human umbilical vein endothelial cells. *Biochem Biophys Res Commun*. 2004;324:1204–1209.
184. Kukidome D, Nishikawa T, Sonoda K, et al. Activation of AMP-activated protein kinase reduces hyperglycemia-induced mitochondrial reactive oxygen species production and promotes mitochondrial biogenesis in human umbilical vein endothelial cells. *Diabetes*. 2006;55:120–127.
185. Ido Y, Carling D, Ruderman N. Hyperglycemia-induced apoptosis in human umbilical vein endothelial cells: inhibition by the AMP-activated protein kinase activation. *Diabetes*. 2002;51:159–167.

186. Sun W, Lee TS, Zhu M, et al. Statins activate AMP-activated protein kinase in vitro and in vivo. *Circulation*. 2006;114:2655–2662.
187. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195–2207.
188. Petersen KF, Dufour S, Befroy D, et al. Reversal of nonalcoholic hepatic steatosis, hepatic insulin resistance, and hyperglycemia by moderate weight reduction in patients with type 2 diabetes. *Diabetes*. 2005;54:603–608.
189. Seppala-Lindroos A, Vehkavaara S, Hakkinen AM, et al. Fat accumulation in the liver is associated with defects in insulin suppression of glucose production and serum free fatty acids independent of obesity in normal men. *J Clin Endocrinol Metab*. 2002;87:3023–3028.
190. Kumashiro N, Erion D, Zhang D, et al. Cellular mechanism of insulin resistance in nonalcoholic fatty liver disease. *Proc Natl Acad Sci U S A*. 2011;108:16381–16385.
191. Samuel VT, Liu ZX, Wang A, et al. Inhibition of protein kinase Cepsilon prevents hepatic insulin resistance in nonalcoholic fatty liver disease. *J Clin Invest*. 2007;117:739–745.
192. Peng Y, Rideout DA, Rakita SS, et al. Does LKB1 mediate activation of hepatic AMP-protein kinase (AMPK) and sirtuin1 (SIRT1) after Roux-en-Y gastric bypass in obese rats? *J Gastrointest Surg*. 2010;14:221–228.
193. Poitout V, Robertson RP. Minireview: secondary beta-cell failure in type 2 diabetes—a convergence of glucotoxicity and lipotoxicity. *Endocrinology*. 2002;143:339–342.
194. El-Assaad W, Buteau J, Peyot ML, et al. Saturated fatty acids synergize with elevated glucose to cause pancreatic beta-cell death. *Endocrinology*. 2003;144:4154–4163.
195. Prentki M, Joly E, El-Assaad W, et al. Malonyl-CoA signaling, lipid partitioning, and glucolipotoxicity: role in beta-cell adaptation and failure in the etiology of diabetes. *Diabetes*. 2002;51:S405–S413.
196. Robertson RP, Harmon J, Tran PO, et al. Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*. 2003;52:581–587.
197. Higa M, Zhou YT, Ravazzola M, et al. Troglitazone prevents mitochondrial alterations, beta cell destruction, and diabetes in obese prediabetic rats. *Proc Natl Acad Sci U S A*. 1999;96:11513–11518.
198. Bakker S, Jzerman R, Teerlink T, et al. Cytosolic triglycerides and oxidative stress in central obesity: the missing link between excessive atherosclerosis, endothelial dysfunction, and beta-cell failure? *Atherosclerosis*. 2000;148:17–21.
199. Hu FB. *Metabolic Consequence of Obesity, Obesity, Epidemiology.* New York: Oxford University Press;2009.
200. Yu C, Chen Y, Cline GW, et al. Mechanism by which fatty acids inhibit insulin activation of Insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. *J Biol Chem*. 2002;277:50230–50236.
201. Laybutt DR, Schmitz-Peiffer C, Saha AK, et al. Muscle lipid accumulation and protein kinase C activation in the insulin-resistant chronically glucose-infused rat. *Am J Physiol*. 1999;277:E1070–E1076.
202. Schmitz-Peiffer C, Browne CL, Oakes ND, et al. Alterations in the expression and cellular localization of protein kinase C isozymes epsilon and theta are associated with insulin resistance in skeletal muscle of the high-fat-fed rat. *Diabetes*. 1997;46:169–178.
203. Hundal RS, Petersen KF, Mayerson AB, et al. Mechanism by which high-dose aspirin improves glucose metabolism in type 2 diabetes. *J Clin Invest*. 2002;109:1321–1326.
204. Yuan M, Konstantopoulos N, Lee J, et al. Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of IKKbeta. *Science*. 2001;293:1673–1677.
205. Kahn BB, Alquier T, Carling D, et al. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab*. 2005;1:15–25.
206. Morino K, Neschen S, Bilz S, et al. Muscle-specific IRS-1 Ser→Ala transgenic mice are protected from fat-induced insulin resistance in skeletal muscle. *Diabetes*. 2008;57:2644–2651.
207. Brown MS, Goldstein JL. Selective versus total insulin resistance: a pathogenic paradox. *Cell Metab*. 2008;7:95–96.
208. Jacob S, Machann J, Rett K, et al. Association of increased intramyocellular lipid content with insulin resistance in lean nondiabetic offspring of type 2 diabetic subjects. *Diabetes*. 1999;48:1113–1119.
209. Krssak M, Falk Petersen K, Dresner A, et al. Intramyocellular lipid concentrations are correlated with insulin sensitivity in humans: a 1H NMR spectroscopy study. *Diabetologia*. 1999;42:113–116.
210. Unger RH. Minireview: weapons of lean body mass destruction: the role of ectopic lipids in the metabolic syndrome. *Endocrinology*. 2003;144:5159–5165.
211. Kim SF, Huang AS, Snowman AM, et al. From the cover: antipsychotic drug-induced weight gain mediated by histamine H1 receptor-linked activation of hypothalamic AMP-kinase. *Proc Natl Acad Sci U S A*. 2007;104:3456–3459.
212. Sinclair DA, Guarente L. Unlocking the secrets of longevity genes. *Sci Am*. 2006;48:54–57.
213. Ruderman NB, Carling D, Prentki M, et al. AMPK, insulin resistance, and the metabolic syndrome. *J Clin Invest*. 2013;123:2764–2772.
214. Hou X, et al. SIRT1 regulates hepatocyte lipid metabolism through activating AMP-activated protein kinase. *J Biol Chem*. 2008;283:20015–20026.
215. Price NL, Gomes AP, Ling AJ, et al. SIRT1 is required for AMPK activation and the beneficial effects of resveratrol on mitochondrial function. *Cell Metab*. 2012;15:675–690.
216. Cantó C, Auwerx J. AMP-activated protein kinase and its downstream transcriptional pathways. *Cell Mol Life Sci*. 2010;67:3407–3423.
217. Ruderman NB, Xu XJ, Nelson L, et al. AMPK and SIRT1: a long-standing partnership? *Am J Physiol Endocrinol Metab*. 2010;298:E751–E760.
218. Zang M, Xu S, Maitland-Toolan KA, et al. Polyphenols stimulate AMP-activated protein kinase, lower lipids, and inhibit accelerated atherosclerosis in diabetic LDL receptor-deficient mice. *Diabetes*. 2006;55:2180–2191.
219. Ruderman N, Prentki M. AMP kinase and malonyl-CoA: targets for therapy of the metabolic syndrome. *Nat Rev Drug Discov*. 2004;3:340–351.
220. Kola B, Christ-Crain M, Lolli F, et al. Changes in adenosine 5'-monophosphate-activated protein kinase as a mechanism of visceral obesity in Cushing's syndrome. *J Clin Endocrinol Metab*. 2008;93:4969–4973.
221. La Montagna G, Cacciapuoti F, Buono R, et al. Insulin resistance is an independent risk factor for atherosclerosis in rheumatoid arthritis. *Diab Vasc Dis Res*. 2007;4:130–135.
222. Sjostrom L, Peltonen M, Jacobson P, et al. Bariatric surgery and long-term cardiovascular events. *JAMA*. 2012;307:56–65.
223. Bäckhed F, Manchester JK, Semenkovich CF, et al. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci U S A*. 2007;104:979–984. 2007.
224. Grundy SM, Brewer Jr HB, Cleeman JI, et al. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*. 2004;109:433–438.
225. Cheal KL, Abbasi F, Lamendola C, et al. Relationship to insulin resistance of the Adult Treatment Panel III diagnostic criteria for identification of the metabolic syndrome. *Diabetes*. 2004;53:1195–1200.
226. McKeigue PM, Pierpoint T, Ferrie JE, et al. Relationship of glucose intolerance and hyperinsulinaemia to body fat pattern in south Asians and Europeans. *Diabetologia*. 1992;35:785–791.
227. Brunzell JD, Ayyobi AF. Dyslipidemia in the metabolic syndrome and type 2 diabetes mellitus. *Am J Med*. 2003;115:24S–28S.
228. Okamoto Y, Kihara S, Ouchi N, et al. Adiponectin reduces atherosclerosis in apolipoprotein E-deficient mice. *Circulation*. 2002;106:2767–2770.
229. Yamauchi T, Kamon J, Waki H, et al. Globular adiponectin protected ob/ob mice from diabetes and ApoE-deficient mice from atherosclerosis. *J Biol Chem*. 2003;278:2461–2468.
230. Isomaa B, Almgren P, Tuomi T, et al. Cardiovascular morbidity and mortality associated with the metabolic syndrome. *Diabetes Care*. 2001;24:683–689.
231. Isomaa B, Henricsson M, Almgren P, et al. The metabolic syndrome influences the risk of chronic complications in patients with type II diabetes. *Diabetologia*. 2001;44:1148–1154.
232. Lakka HM, Laaksonen DE, Lakka TA, et al. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA*. 2002;288:2709–2716.

233. Alexander CM, Landsman PB, Teutsch SM, et al. NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. *Diabetes*. 2003;52:1210–1214.
234. Sattar N, Gaw A, Scherbakova O, et al. Metabolic syndrome with and without C-reactive protein as a predictor of coronary heart disease and diabetes in the West of Scotland Coronary Prevention Study. *Circulation*. 2003;108:414–419.
235. Grundy SM. Cardiovascular and metabolic risk factors: how can we improve outcomes in the high-risk patient? *Am J Med*. 2007;120:S3–S9.
236. Yu AS, Keeffe EB. Nonalcoholic fatty liver disease. *Rev Gastroenterol Disord*. 2002;2:11–19.
237. Green RM. NASH–hepatic metabolism and not simply the metabolic syndrome. *Hepatology*. 2003;38:14–17.
238. Marchesini G, Brizi M, Bianchi G, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. *Diabetes*. 2001;50:1844–1850.
239. Stefan N, Kantartzis K, Haring HU. Causes and metabolic consequences of fatty liver. *Endocr Rev*. 2008;29:939–960.
240. Roberts EA. Nonalcoholic steatohepatitis in children. *Curr Gastroenterol Rep*. 2003;5:253–259.
241. Petersen KF, Dufour S, Savage DB, et al. The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc Natl Acad Sci U S A*. 2007;104:12587–12594.
242. Rabøl R, Petersen KF, Dufour S, et al. Reversal of muscle insulin resistance with exercise reduces postprandial hepatic de novo lipogenesis in insulin resistant individuals. *Proc Natl Acad Sci U S A*. 2011;108:13705–13709.
243. Legro RS, Kunselman AR, Dodson WC, et al. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*. 1999;84:165–169.
244. Solomon CG, Hu FB, Dunaif A, et al. Menstrual cycle irregularity and risk for future cardiovascular disease. *J Clin Endocrinol Metab*. 2002;87:2013–2017.
245. Calle EE, Thun MJ. Obesity and cancer. *Oncogene*. 2004;23:6365–6378.
246. Hawley SA, Boudeau J, Reid JL, et al. Complexes between the LKB1 tumor suppressor, STRADalpha/beta and MO25alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. *J Biol*. 2003;2:28.
247. Woods A, Johnstone SR, Dickerson K, et al. LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. *Curr Biol*. 2003;13:2004–2008.
248. Luo Z, Saha AK, Xiang X, et al. AMPK, the metabolic syndrome and cancer. *Trends Pharmacol Sci*. 2005;26:69–76.
249. Motoshima H, Goldstein BJ, Igata M, et al. AMPK and cell proliferation—AMPK as a therapeutic target for atherosclerosis and cancer. *J Physiol*. 2006;574:63–71.
250. Xiang X, Saha AK, Wen R, et al. AMP-activated protein kinase activators can inhibit the growth of prostate cancer cells by multiple mechanisms. *Biochem Biophys Res Commun*. 2004;321:61–167.
251. Buzzai M, Jones RG, Amaravadi RK, et al. Systemic treatment with the antidiabetic drug metformin selectively impairs p53-deficient tumor cell growth. *Cancer Res*. 2007;67:6745–6752.
252. Kramer AF, Erickson KI. Capitalizing on cortical plasticity: influence of physical activity on cognition and brain function. *Trends Cogn Sci*. 2007;11:342–348.
253. Patrone C, Eriksson O, Lindholm D. Diabetes drugs and neurological disorders: new views and therapeutic possibilities. *Lancet Diabetes Endocrinol*. 2014;2(3):256–262.
254. Morris D. Cushing's syndrome. In: De Groot LJ, ed. *Endocrinology*. Philadelphia: Elsevier:429–464.
255. Seely WW, Williams GH. The cardiovascular system and endocrine disease. In: Becker KL, ed. *Principles and practice of endocrinology and metabolism.* Philadelphia: Lippincott Williams & Wilkins;2001:1857–1864.
256. De Leo M, Pivonello R, Auriemma RS, et al. Cardiovascular disease in Cushing's syndrome: heart versus vasculature. *Neuroendocrinology*. 2010;92:50–54.
257. Walker BR. Is "Cushing's disease of the omentum" an affliction of mouse and men? *Diabetologia*. 2004;47:767–769.
258. Sutinen J, Hakkinen AM, Westerbacka J, et al. Rosiglitazone in the treatment of HAART-associated lipodystrophy—a randomized double-blind placebo-controlled study. *Antivir Ther*. 2003;8:199–207.
259. Unger R. How adipocytes integrate surplus caloric intake with caloric storage: lessons from Morgan Spurlock and some French geese. *Curr Opin Endocrinol Diabetes Obes*. 2004;11:251–257.
260. Sims EA, Danforth Jr E, Horton ES, et al. Endocrine and metabolic effects of experimental obesity in man. *Recent Prog Horm Res*. 1973;29:457–496.
261. Grundy SM, Hansen B, Smith Jr SC, et al. Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes Association conference on scientific issues related to management. *Circulation*. 2004;109:551–556.
262. Skerrett PJ, Manson JE. Exercise and diabetes prevention: reduction in risk of coronary heart disease in diabetes. In: Ruderman N, Devlin JT, Schneider SH, et al., eds. *Handbook of diabetes in exercise.* Alexandria: American Diabetes Association:155–182.
263. Orchard TJ, Temprosa M, Goldberg R, et al. The effect of metformin and intensive lifestyle intervention on the metabolic syndrome: the Diabetes Prevention Program Randomized Trial. *Ann Intern Med*. 2005;142:611–619.
264. Lindstrom J, Ilanne-Parikka P, Peltonen M, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet*. 2006;368:1673–1679.
265. Yusuf S, Gerstein H, Hoogwerf B, et al. Ramipril and the development of diabetes. *JAMA*. 2001;286:1882–1885.
266. Robins SJ, Rubins HB, Faas FH, et al. Insulin resistance and cardiovascular events with low HDL cholesterol: the Veterans Affairs HDL Intervention Trial (VA-HIT). *Diabetes Care*. 2003;26:1513–1517.
267. Hundal RS, Krssak M, Dufour S, et al. Mechanism by which metformin reduces glucose production in type 2 diabetes. *Diabetes*. 2000;49:2063–2069.
268. Zang M, Zuccollo A, Hou X, et al. AMP-activated protein kinase is required for the lipid-lowering effect of metformin in insulin-resistant human HepG2 cells. *J Biol Chem*. 2004;279:47898–47905.
269. Madiraju AK, Erion DM, Rahimi Y, et al. Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. *Nature*. 2014;510:542–546.
270. UKPDS. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet*. 1998;352:854–865.
271. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577–1589.
272. Lin YC, Wu MH, Wei TT, et al. Metformin sensitizes anticancer effect of dasatinib in head and neck squamous cell carcinoma cells through AMPK-dependent ER stress. *Oncotarget*. 2014;5:298–308.
273. Cheang WS, Tian XY, Wong WT, et al. Metformin protects endothelial function in diet-induced obese mice by inhibition of endoplasmic reticulum stress through 5' adenosine monophosphate-activated protein kinase-peroxisome proliferator-activated receptor δ pathway. *Arterioscler Thromb Vasc Biol*. 2014;34:830–836.
274. Almabrouk TA, Ewart MA, Salt IP, et al. Perivascular fat, AMP-activated protein kinase and vascular diseases. *Br J Pharmacol*. 2014;171:595–617.
275. Boyle JG, Logan PJ, Jones GC, et al. AMP-activated protein kinase is activated in adipose tissue of individuals with type 2 diabetes treated with metformin: a randomised glycaemia-controlled crossover study. *Diabetologia*. 2011;54:1799–1809.
276. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007;356:2457–2471.
277. Lincoff AM, Wolski K, Nicholls SJ, et al. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: a meta-analysis of randomized trials. *JAMA*. 2007;298:1180–1188.
278. Dormandy JA, Charbonnel B, Eckland DJ, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROACTIVE Study (Prospective Pioglitazone Clinical Trial in Macrovascular Events): a randomised controlled trial. *Lancet*. 2005;366:1279–1289.
279. Nissen SE, Nicholls SJ, Wolski K, et al. Comparison of pioglitazone vs glimepiride on progression of coronary atherosclerosis in patients with

type 2 diabetes: the PERISCOPE randomized controlled trial. *JAMA*. 2008;299:1561–1573.

280. Nathan DM, Buse JB, Davidson MB, et al. Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy. *Diabetes Care*. 2008;29:1963–1972.
281. Ahmadian M, Suh JM, Hah N, et al. PPARγ signaling and metabolism: the good, the bad and the future. *Nat Med*. 2013;19:557–566.
282. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation*. 2004;110:227–239.
283. Ridker PM, Rifai N, Clearfield M, et al. Measurement of C-reactive protein for the targeting of statin therapy in the primary prevention of acute coronary events. *N Engl J Med*. 2001;344:1959–1965.
284. Shah RV, Goldfine AB. Statins and risk of new-onset diabetes mellitus. *Circulation*. 2012;126:e282–e284.
285. Rosenzweig JL, Ferrannini E, Grundy SM, et al. Primary prevention of cardiovascular disease and type 2 diabetes in patients at metabolic risk: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:3671–3689.
286. Cook S, Weitzman M, Auinger P, et al. Prevalence of a metabolic syndrome phenotype in adolescents: findings from the third National Health and Nutrition Examination Survey, 1988–1994. *Arch Pediatr Adolesc Med*. 2003;157:821–827.
287. Berenson GS, Srinivasan SR, Bao W, et al. Associations between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study. *N Engl J Med*. 1998;338:1650–1656.
288. Must A, Jacques PF, Dallal GE, et al. Long-term morbidity and mortality of overweight adolescents, a follow-up of the Harvard Growth Study of 1922 to 1935. *N Engl J Med*. 1992;327:1350–1355.
289. Steinberger J, Daniels SR. Obesity, insulin resistance, diabetes, and cardiovascular risk in children: an American Heart Association scientific statement from the Atherosclerosis, Hypertension, and Obesity in the Young Committee (Council on Cardiovascular Disease in the Young) and the Diabetes Committee (Council on Nutrition, Physical Activity, and Metabolism). *Circulation*. 2003;107:1448–1453.
290. Barker DJ, Hales CN, Fall CH, et al. Type 2 (noninsulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia*. 1993;36:62–67.
291. Phillips DI. Birth weight and the future development of diabetes. A review of the evidence. *Diabetes Care*. 1998;21:B150–B155.
292. Bhargava SK, Sachdev HS, Fall CH, et al. Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *N Engl J Med*. 2004;350:865–875.
293. Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008;105(44):17046–17049.
294. Schulz LC. The Dutch Hunger Winter and the developmental origins of health and disease. *Proc Natl Acad Sci U S A*. 2010;107:16757–16758.
295. Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. *Nature*. 2001;414:782–787.
296. Neel JV. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”. *Am J Hum Genet*. 1962;14:353–362.
297. Hu Z, Cha SH, Chohnan S, et al. Hypothalamic malonyl-CoA as a mediator of feeding behavior. *Proc Natl Acad Sci U S A*. 2003;100:1262–12629.
298. Kemp BE, Stapleton D, Campbell DJ, et al. AMP-activated protein kinase, super metabolic regulator. *Biochem Soc Trans*. 2003;31:62–168.
299. Hardie DG, Scott JW, Pan DA, et al. Management of cellular energy by the AMP-activated protein kinase system. *FEBS Lett*. 2003;546:113–120.
300. Sanders MJ, Grondin PO, Hegarty BD, et al. Investigating the mechanism for AMP activation of the AMP-activated protein kinase cascade. *Biochem J*. 2007;403:139–148.
301. Towler MC, Hardie DG. AMP-activated protein kinase in metabolic control and insulin signaling. *Circ Res*. 2007;100:328–341.

27

Bariatric Procedures and Operations

Todd Andrew Kellogg, Barham K. Abu Dayyeh, and Michael R. Rickels

OUTLINE

Introduction/Overview, 450
Commonly Performed Bariatric Operations, 450
  Vertical Sleeve Gastrectomy, 451
  Roux-en-Y Gastric Bypass, 451
  Biliopancreatic Diversion With Duodenal Switch, 453
  Single Anastomosis Duodenal Ileostomy Bypass With Sleeve, 453
  Laparoscopic Adjustable Gastric Banding, 453
  Endoscopic Bariatric and Metabolic Therapies, 453
Evolving Evidence Profile for the Use of Endoscopic Bariatric and Metabolic Therapies in Clinical Practice, 455
  Space-Occupying Devices, 455
  Aspiration Therapy, 456
  Small Bowel Intervention, 456
  Eligibility for Bariatric Operations and Procedures, 458
  Assessing Outcomes after Bariatric Interventions, 458
The Physiology of Weight Loss After Bariatric Operations and Procedures, 459
  Weight Loss Mechanisms, 459
  Caloric Restriction, 459
  Gastrointestinal Anatomy Changes, 459
  Gastrointestinal and Adipocyte Hormonal Changes, 459
  Central Nervous System Control, 459
  Vagal Signaling, 460
  Energy Expenditure, 460
Physiologic Effects of Bariatric Surgery on Cardiometabolic Health and Mortality, 460
  Diabetes Remission Following Bariatric Surgery, 460
  The Role of Insulin Sensitivity and Insulin Secretion in Diabetes Remission After Gastric Bypass, 460
  The Hindgut and the Foregut Hypotheses, 461
  Branched-Chain Amino Acids and Aromatic Amino Acids, 461
  Bile Acids, 461
  Gut Microbiota, 461
  Cardiovascular Disease and Mortality, 461
Impact of Bariatric Surgery on Other Obesity-Related Comorbid Conditions, 462
  Polycystic Ovarian Syndrome and Fertility, 462
  Pulmonary Disease, Including Obstructive Sleep Apnea, 463
  Gastroesophageal Reflux Disease, 463
  Pseudotumor Cerebri, 463
  Cancer, 463
Complications of Bariatric Operations and Procedures, 463
  Early Complications, 463
Late and Chronic Complications, 464
Weight Regain, 464
Dumping Syndrome, 464
  Alimentary Hypoglycemia With Neuroglycopenia, 464
Summary, 467

INTRODUCTION/OVERVIEW

Healthy food choices with portion control, as well as consistent exercise, are the mainstays of a healthy lifestyle and maintenance of a healthy weight. An imbalance of these factors can contribute to obesity with its associated diseases. Surgically-induced weight loss is the most potent form of weight loss intervention, particularly in individuals with class 2 and class 3 obesity, and can treat many obesity-associated diseases. The potential health benefits of surgically-induced weight loss include improved survival, improvement and remission of type 2 diabetes, and a reduced incidence of diabetes, cardiovascular disease, and cancers in women.

The epidemic of medically severe obesity coupled with the development and widespread utilization of less invasive laparoscopic techniques has led to a consistent increase in the annual number of bariatric operations performed, as well as the development of endoscopic bariatric interventions and together provide a multidisciplinary approach with escalating obesity interventions (Fig. 27.1).

We will begin by discussing the various bariatric operations and procedures, followed by discussion of the metabolic and physiologic consequences.

COMMONLY PERFORMED BARIATRIC OPERATIONS

With the development of endoscopic procedures for weight loss, it has become necessary to differentiate between bariatric surgical *operations* and bariatric endoscopic *procedures*. We will begin by discussing the anatomy and general mechanisms of the operations most commonly performed for weight management (Fig. 27.2).

The mechanisms behind bariatric operations have components of restriction, malabsorption, and neurohormonal changes. All operations have a varying combination of these three components as their mechanism, resulting in a unique anatomy and physiology that can affect caloric intake, nutrient handling, and nutrient sensing. These changes likely influence each operation's unique outcome, as discussed later in this chapter. Although much has been learned regarding these mechanisms, much remains to be understood.

All these operations are performed laparoscopically, which benefits the patient because there are fewer wound complications, less pain, and earlier return to activities compared with open operations.[1,2] Outcomes after bariatric surgery are categorized into the effect on weight and the effect on weight-related comorbid disease, particularly metabolic disease. Potential adverse effects to be considered include nutritional complications.

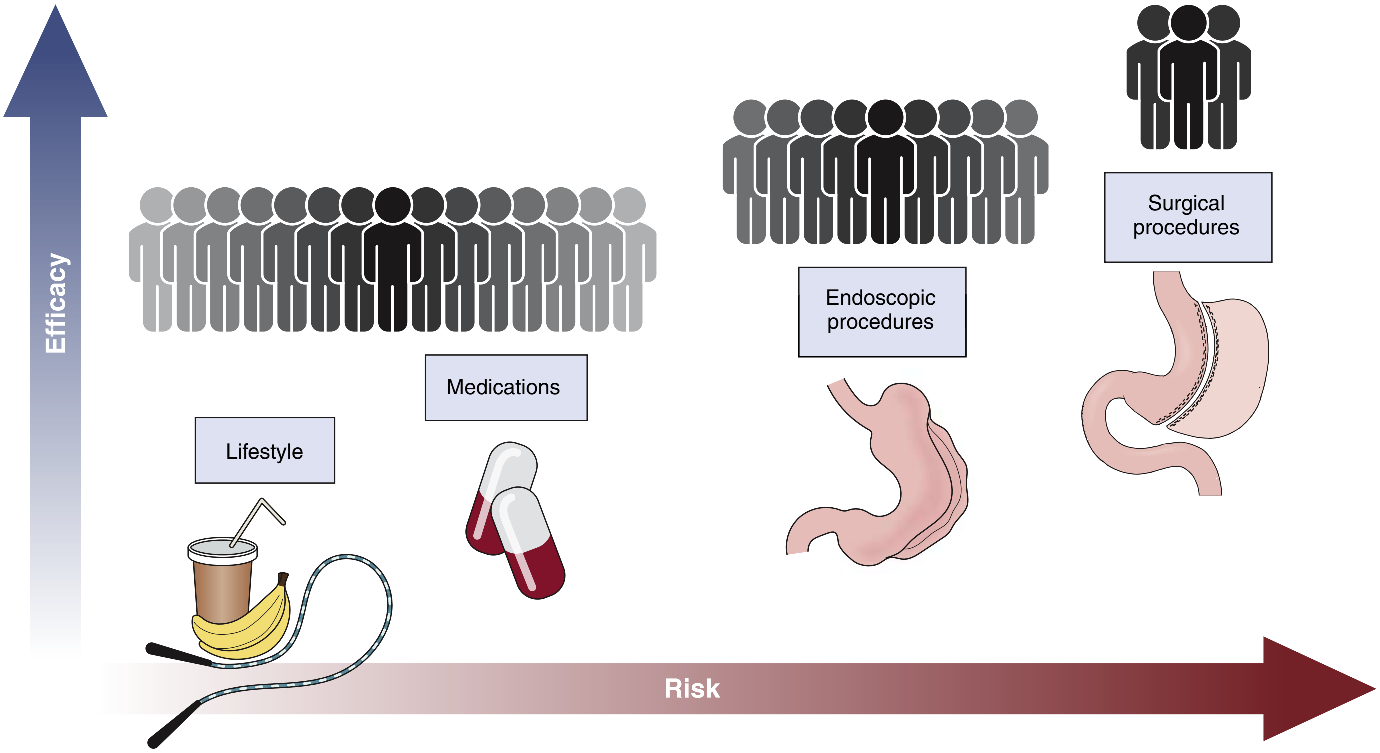


Figure 27.1 Escalating interventions in the treatment of obesity.

Vertical Sleeve Gastrectomy

Sleeve gastrectomy (SG) is currently the most commonly performed bariatric procedure in the United States and worldwide, comprising approximately 65% of all bariatric operations in the United States. Its rapid adoption can be attributed to it being a less technically demanding operation that has very good weight loss, as well as improvement of obesity-related comorbid diseases. Although the SG is a promising operation for treating obesity, the development of gastroesophageal reflux disease (GERD) postoperatively and long-term weight regain are growing concerns[3] and thus, as with any bariatric operation, careful patient selection is of paramount importance.

The SG is a restrictive operation with a neurohormonal component.[4] The restrictive component occurs by way of an approximately 75% to 80% reduction of the gastric volume via resection of the gastric greater curvature including the gastric fundus (see Fig. 27.2B). Its neurohormonal effects are not completely understood but are thought to be the result of the reduction of ghrelin-producing oxyntic gland endocrine P/D1 cells[5-7] located in the gastric fundus and parietal cells of the gastric greater curvature. A prospective, double-blind study that compared Roux-en-Y gastric bypass (RYGB) and laparoscopic SG demonstrated significant postprandial suppression of ghrelin after SG and no change after RYGB. Appetite reduction and excess weight loss (EWL) were greater after SG, suggesting an association with the suppression of ghrelin levels. In the same study, peptide YY (PYY) levels were increased similarly after either operation.[8]

An increase in gastric emptying with upregulation of glucagon-like peptide 1 (GLP-1) has also been proposed as a mechanism for the effects of SG. However, the evidence is mixed, which may be secondary to variations in surgical technique.

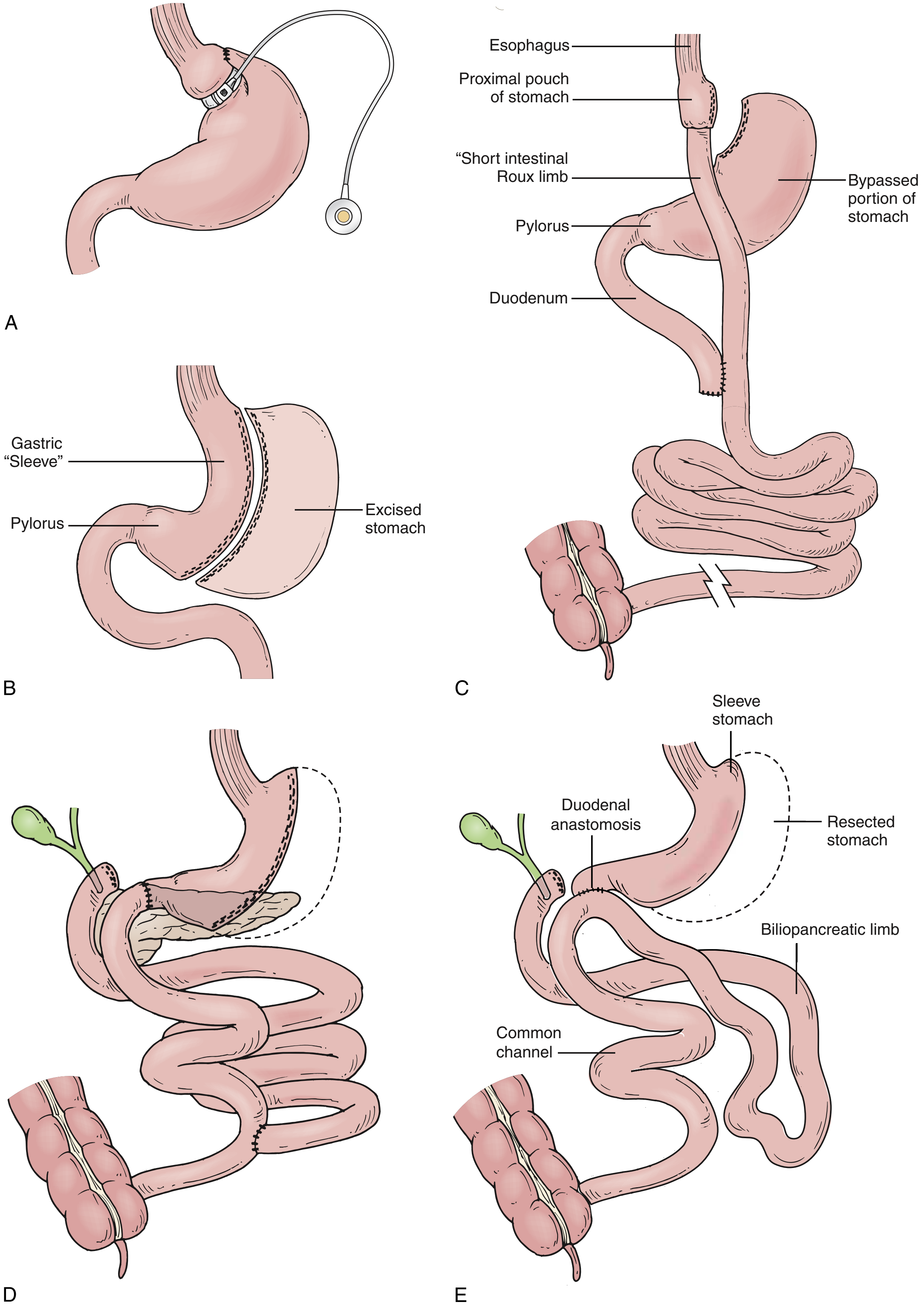
SG is a very safe operation, with a 30-day mortality rate of 0.2% to 0.5%.[9,10]

Roux-en-Y Gastric Bypass

Although the SG is currently the most commonly performed bariatric operation in the United States, in many ways the RYGB is still considered the gold standard bariatric operation. It has a solid track record of efficacy, durability, and safety over more than 20 years of follow-up, which the SG lacks. Although its effects are mostly due to restriction and neurohormonal changes, it also has a small malabsorptive component.

The RYGB creates a small approximately 15- to 30-mL gastric pouch that is constructed by dividing the upper aspect of the stomach, and this small pouch is then attached to a 75-cm to 150-cm jejunal Roux limb (see Fig. 27.2C). The so-called "biliopancreatic limb" is typically 40 cm to 100 cm in length and carries bile and digestive enzymes from the duodenum and from the remaining portion of the stomach (gastric remnant). The result of this anatomic configuration is that the ingested food and digestive juices are separated for the length of the Roux limb, after which they enter a long common channel of small bowel (the remainder of the jejunum and the entire ileum), where mixing of food with bile and digestive enzymes occurs. This anatomy results in the ingested food bypassing the majority of the stomach, the entire duodenum, and the first portion of the jejunum. In general, the relatively small length of bypassed intestine (roughly 30%) is not enough to cause protein-calorie malabsorption. However, it does impact the absorption of some micro- and macronutrients, and specifically calcium, iron, vitamin B12, and vitamin D.

The mechanism of action leading to weight loss and comorbidity improvement is multifactorial. Like the SG, the RYGB is mainly a restrictive procedure with additional neurohormonal effects. However, unlike the SG, bypass of the duodenum and proximal jejunum as well as rapid delivery of chime to the distal small bowel are thought to be important mechanisms resulting in alteration in of the entero-endocrine axis. The changes observed in gut peptides caused by rapid delivery of nutrients to the jejunum (glucose-dependent insulinotropic polypeptide and GLP-1) after bypass of the duodenum may be an important long-term mechanism for durable weight loss and comorbidity remittance.[11]



A
Esophagus
Proximal pouch of stomach
"Short intestinal Roux limb
Bypassed portion of stomach
Pylorus
Duodenum
Gastric "Sleeve"
Pylorus
Excised stomach
B
C
Sleeve stomach
Duodenal anastomosis
Resected stomach
Biliopancreatic limb
Common channel
D
E

Figure 27.2 **A,** Adjustable gastric band. A silicone band is looped around the proximal stomach to create a 15- to 20-mL pouch with an adjustable outlet. The stomach is wrapped around the band anteriorly to prevent the band from slipping out of position. The band consists of a rigid outer ring and an inner inflatable balloon reservoir connected by tubing to a subcutaneous port that can be accessed through the skin to adjust the tightness. **B,** Sleeve gastrectomy. A narrow gastric sleeve is created by stapling the stomach vertically. The fundus and greater curve of the stomach are surgically removed. **C,** Roux-en-Y gastric bypass (RYGB). A small gastric pouch (15–30 mL) is created by division of the upper stomach connected to a 100- to 150-cm limb of jejunum called the Roux limb. The small gastric pouch results in restriction of food intake. **D,** Biliopancreatic diversion with duodenal switch (BPD-DS). To avoid acid reflux, marginal ulcers, and dumping syndrome by maintaining the pylorus, the operation was modified with the pouch based on the lesser curve of the stomach and an anastomosis at the first portion of the duodenum. **E,** Single anastomosis duodenal ileal bypass with sleeve (SADI-S). A variation of the BPD-DS in which a single anastomosis is constructed 250–300 cm proximal to the ileocecal valve.

Gastric bypass is a very safe and extremely effective bariatric operation. The associated 30-day mortality is 0.2%, incidence of unplanned intensive care unit admissions is 1.3%, and incidence of 30-day reoperations is 2.5%.[12] This low perioperative risk is accompanied by a nearly 30% to 35% total body weight loss (TBWL). The weight loss outcomes are sometimes reported as "excess body weight loss," meaning the patient has lost an amount of weight above "ideal" body weight, which for RYGB would be approximately 70% excess body weight loss. In addition to the highly significant weight loss, there is a high rate of improvement or remission of obesity-related comorbidities,[13] which makes the RYGB a good choice for treatment of obesity and its complications.

Biliopancreatic Diversion with Duodenal Switch

Biliopancreatic diversion with duodenal switch (BPD-DS) is the most powerful bariatric procedure in terms of weight loss and the treatment of diabetes and other obesity-related comorbidities. It is a technically challenging operation and has a higher incidence of complications compared with RYGB and SG. Primary BPD-DS represents less than 1% of all bariatric operations in the Unites States [https://asmbs.org/resources/estimate-of-bariatric-surgery-numbers]. However, BPD-DS is beginning to gain popularity as a revision operation, especially in the setting of weight regain after SG.[14]

The effects of the BPD-DS can be attributed to components of restriction, malabsorption, and neurohormonal changes and, compared with the SG and RYGB, is the only bariatric operation that employs true protein-calorie malabsorption as a mechanism for weight loss (see Fig. 27.1D). The BPD-DS combines a modest SG with an extensive bypass of the small bowel while preserving the pylorus. Preservation of the pylorus theoretically decreases the risk of developing dumping syndrome and reactive hypoglycemia.[15,16] The BPD-DS anatomy is a combination of restriction (vertical SG) and malabsorption (extensive intestinal bypass). The malabsorptive component is profound, and its contribution is what creates the most amount of weight loss. The intestinal bypass has a metabolic effect thought to be due to early exposure of nutrients to the ileum leading to increased levels of PYY and GLP-1.[17] The length of the intervening bowel between the alimentary limb-biliopancreatic limb anastomosis and the cecum, termed the "common channel," is the main determinant of the degree of malabsorption. As the length of the common channel becomes shorter, the malabsorptive power of the operation increases, and with it the potential weight loss. However, the risks of the adverse effects associated with aggressive malabsorption also increase and can lead to severe macro- and micronutrient deficiencies, including sometimes fatal protein-calorie malnutrition.

The BPD-DS is the most effective bariatric operation currently performed, producing the most weight loss and the best rates of diabetes remission.[18-20] However, due mainly to its impact on nutrition, it is associated with higher morbidity compared with other procedures, with a reported 8% to 11% overall complication risk.[21] BPD-DS may be best reserved for patients with body mass index (BMI) greater than 50 kg/m^2 or as a revision procedure for patients with inadequate weight loss after SG. Because of the potential for malnutrition after BPD-DS, this operation should be avoided in patients who are unlikely to be compliant with postoperative dietary guidelines and nutrition supplements. In our experience, those with a heavy burden of mental health issues, lack of stability in their personal lives, and a poor social support system, as well as those who exhibit dubious compliance preoperatively, are not considered candidates in our program.

Single Anastomosis Duodenal Ileostomy Bypass With Sleeve

The single anastomosis duodenal ileostomy bypass with sleeve (SADI-S) is a modification of the classic BPD-DS and was initially described by Torres over 10 years ago.[22] The SG portion of the operation is the same. However, rather than a Roux-en-Y configuration of the bypassed small bowel, the SADI-S has a loop configuration with one anastomosis between the proximal duodenum and the ileum constructed 250 to 300 cm proximal to the ileocecal valve, which eliminates the biliopancreatic diversion portion (see Fig. 27.2E). This operation has recently been endorsed by the American Society for Metabolic and Bariatric Surgery (ASMBS) as effective and safe. It can be offered as a primary bariatric operation or for revision of SG in the setting of weight regain. Early studies have suggested that weight loss and comorbidity outcomes are similar to those of the BPD-DS. One retrospective study demonstrated 85.7% EWL at 4 years postoperatively, with 97.6% patients being able to maintain hemoglobin A1c (HbA1c) less than 6% (with or without diabetic medications).[23] Other retrospective studies have shown more weight loss in BPD-DS patients compared with SADI-S patients.[24] The postulated advantages over the traditional BPD-DS include a lower risk of internal hernias and lower risk of malnutrition and diarrhea.[25]

Laparoscopic Adjustable Gastric Banding

The adjustable gastric band (AGB) is a Silastic device that is comprised of a band with an inner balloon attached to tubing that leads to a subcutaneous port (see Fig. 27.2A). Its relatively simple concept, straightforward surgical procedure, low mortality, and reversibility drove its initial popularity to the point where it was the most popular weight loss operation in 2007. However, it has lost favor due to poor long-term weight loss and unacceptable long-term complication and reoperation rates, as well as the emergence of the SG as the restrictive operation of choice. The AGB currently comprises approximately 1% of all bariatric operations.

KEY POINTS

- The vertical sleeve gastrectomy (SG), a restrictive operation with a neurohormonal component, is currently the most commonly performed bariatric operation in the United States.
- The Roux-en-Y gastric bypass provides restriction as its main mechanism, coupled with a neurohormonal component. The limited malabsorptive component is not designed to provide protein-calorie malabsorption as a weight loss mechanism.
- The biliopancreatic diversion with duodenal switch and single anastomosis duodenal ileostomy with sleeve gastrectomy are primarily malabsorptive operations and can result in profound protein-calorie malabsorption. The SG portion of these operations also provides restriction, and there is an associated robust neurohormonal component.
- The adjustable gastric band is a purely restrictive operation that currently comprises less than 2% of the bariatric operations being performed.

Endoscopic Bariatric and Metabolic Therapies

The burden of obesity is compounded by an insufficient treatment arsenal. Behavioral changes are thwarted by low durability and patient compliance.[26,27] Pharmacologic therapies are associated with significant cost and potential adverse events, require lifelong compliance to maintain response, and, given their targeted nature of redundant evolutionary pathways defending body weight, are often inflicted with weight recidivism and attenuation of response.[28,29] Bariatric surgery, while associated with significant and sustained weight loss, is limited by its invasiveness, expense, long-term risks, and patient acceptance.[30-32] These barriers have prevented bariatric surgeries from reaching more than 2% of eligible patients, with the vast majority of patients with mild to moderate obesity not qualifying for surgery.[32,33] Furthermore, obesity equally inflicts younger patients, who will likely outlive the durability of bariatric surgery and require additional revision surgery, which significantly compounds the short- and long-term risks and cost.[34,35] This landscape has provided an opportunity for

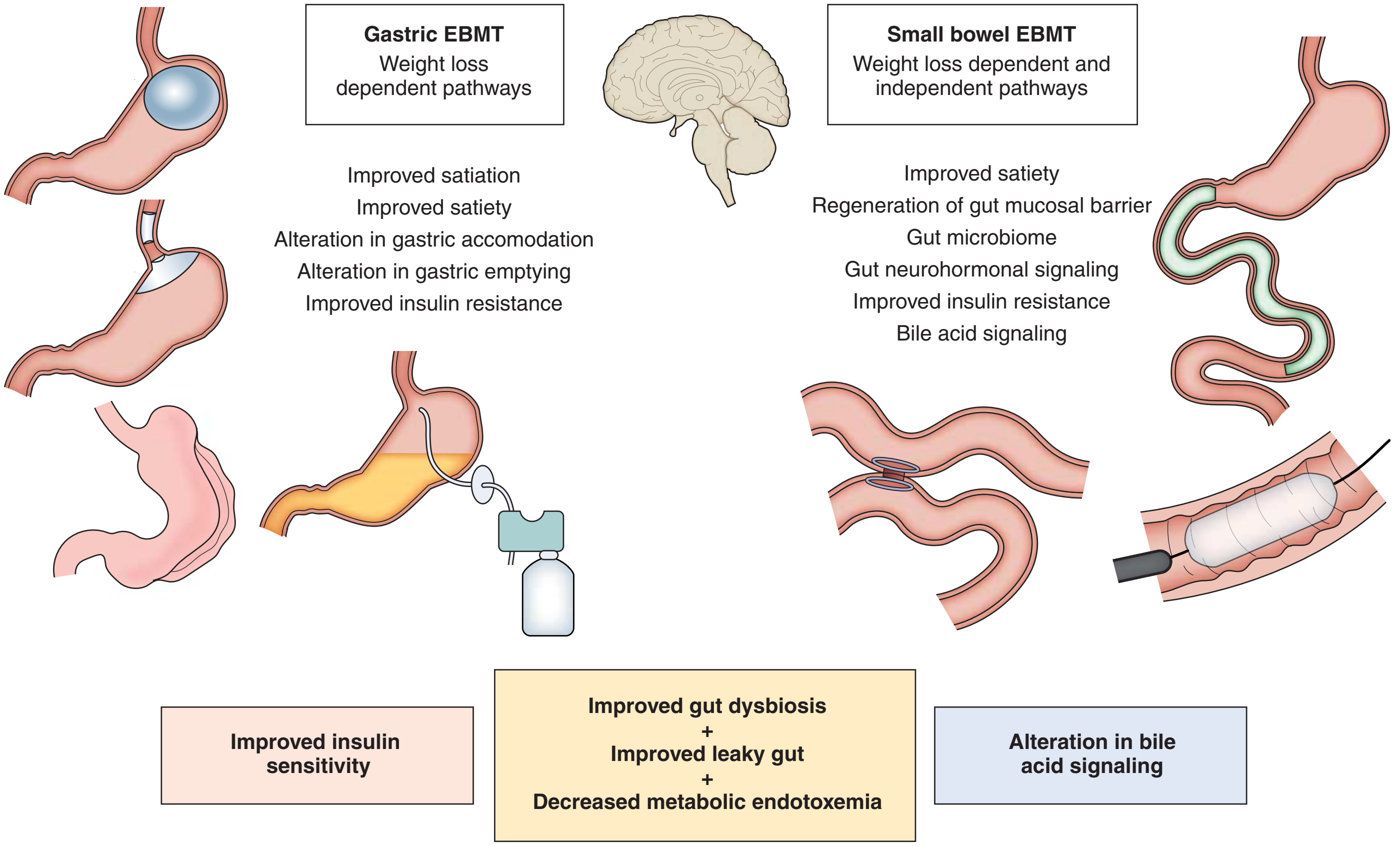

Gastric EBMT
Weight loss dependent pathways
Improved satiation
Improved satiety
Alteration in gastric accomodation
Alteration in gastric emptying
Improved insulin resistance
Small bowel EBMT
Weight loss dependent and independent pathways
Improved satiety
Regeneration of gut mucosal barrier
Gut microbiome
Gut neurohormonal signaling
Improved insulin resistance
Bile acid signaling
Improved insulin sensitivity
Improved gut dysbiosis
+
Improved leaky gut
+
Decreased metabolic endotoxemia
Alteration in bile acid signaling

Figure 27.3 Proposed pathways for effectiveness of endoscopic bariatric and metabolic therapies (EBMT).

minimally-invasive and anatomy-preserving endoscopic bariatric and metabolic therapies (EBMTs) to address the gaps in obesity care and increase access to effective therapies capable of producing significant weight loss and improvement in excess adiposity–related comorbidities.[36] Recognizing the noncurable, chronic, relapsing nature of obesity, an evolving management strategy pivoting on an intensive weight loss intervention utilizing EBMT for those with moderate forms of the disease is needed. The goal of such a strategy is to produce enough upfront weight loss that the patients will engage in lifelong weight stabilization and reintensification efforts, similar to any other chronic disease, where interventions are layered and sequenced for effective disease management.[37]

EBMTs are generally considered in patients who do not qualify for or who wish to pursue bariatric surgery, especially in those with class 1–2 obesity (BMI 30–40 kg/m^2) who have previously failed to lose weight using lifestyle methods. Furthermore, EBMTs can be utilized as a bridge to definitive bariatric surgery in patient with very high BMIs and prohibitive surgical risk.[38-40] Other indications include:

1. Prior to organ transplantation, allowing pretransplant candidates to meet transplant eligibility.[41-43]
2. Weight reduction prior to orthopedic procedures and for infertility treatment.[44]
3. For management of obesity-related comorbidities such as nonalcoholic steatohepatitis and type 2 diabetes.[45-48]

EBMT approaches are thought to target peripheral and central gastrointestinal pathways implicated in the regulation of energy intake and can be divided based on their anatomic manipulation into gastric or small intestinal interventions (Fig. 27.3).[49]

1. Gastric interventions include space-occupying devices that most commonly take the form of temporarily placed prostheses (intragastric balloons [IGBs]). The TransPyloric Shuttle (BAROnova Inc, Goleta, CA) intermittently seals the pyloric channel and delays gastric emptying in the fed state to induce early satiation and prolonged satiety. The Full Sense Device (BFKW, Grand Rapids, MI) is a modified, fully covered gastroesophageal stent with a cylindrical esophageal component and a gastric disc connected by struts that is thought to induce weight loss by exerting constant gentle, pressure on the gastric cardia, thereby triggering afferent vagal signaling to the central nervous system and a sensation of fullness. Additional approaches that target the stomach include gastric remodeling techniques that reduce the gastric reservoir by endoscopically creating a tubular sleeve along the greater curvature of the stomach through transoral suturing (Overstitch, Apollo Endosurgery, Austin, TX) or plication (POSE, USGI Medical, San Clemente, CA) and aspiration therapy that allows patients with obesity to dispose of a portion of their ingested meal via a specially designed percutaneous gastrostomy tube, known as the A-Tube (Aspire Bariatrics, King of Prussia, PA).
2. Small bowel interventions: it has been hypothesized that the proximal small intestine plays a role in the pathogenesis of metabolic disease, resulting in insulin resistance and chronic inflammation. The pathways that have been proposed include alteration of neurohormonal signaling, changes in the gut microbiome and mucosal barrier function, and activation of innate immune reactions caused by an excess of free fatty acids, bacterial lipopolysaccharides, chemokines, and cytokines.[50] The approaches that have been designed to address these pathways include the exclusion of the proximal small intestines by impermeable polymer duodenojejunal bypass liners (EndoBarrier, GI Dynamics, Lexington, MA; Metamodix, Minneapolis, MN) and ablative duodenal resurfacing techniques that regenerate the proximal small intestinal mucosal barrier by thermal or nonthermal ablation methods (Fractyl Laboratories, Cambridge, MA). Some of these therapies have shown promise in the

management of obesity and metabolic disease. Another theory is that primary and secondary bile acid concentrations and composition in the intestinal, portal, and systemic circulation play roles in insulin secretion and sensitivity, metabolic rate, liver lipogenesis and inflammation, and liver fibrosis progression through FXR/fibroblast growth factor (FGF)15 and FGF19/TGR5 signaling.[51] In hopes of favorably altering bile acid profiles, endoscopic treatments have been developed that deploy self-assembling magnets (GI Windows, Boston, MA) to create a dual-path enteral bypass between the proximal jejunum or duodenum and ileum in order to partially divert bile to the terminal ileum.

EVOLVING EVIDENCE PROFILE FOR THE USE OF ENDOSCOPIC BARIATRIC AND METABOLIC THERAPIES IN CLINICAL PRACTICE

Space-Occupying Devices

Space-occupying devices can be divided into balloon and nonballoon devices. At present time of this publication, there are four US Food and Drug Administration (FDA)-approved space-occupying devices (ReShape Duo, Orbera, Obalon, and the Transpyloric Shuttle). The ReShape Duo balloon is no longer being marketed. Two additional IGBs (Spatz3 and Elipse) have completed pivotal US regulatory trials and are awaiting approval in the United States.

1. **Orbera** is the most-studied IGB for obesity. The device is a single fluid-filled spherical balloon system that requires endoscopy to confirm placement. Once placed in the stomach, the balloon is filled with 450 to 700 mL of saline with optional methylene blue to detect spontaneous deflation. At 6 months, the balloon requires endoscopic removal. Balloon intolerance was reported to be 7.5%.[52] A prior metaanalysis of 55 studies that included over 6500 Orbera IGB placements demonstrated a pooled estimate of 13.2% TBWL at 6 months and 11.3% at 12 months.[53] In a pivotal US multicenter randomized controlled clinical trial, investigators compared subjects (n = 125) who received 6 months of balloon therapy plus lifestyle intervention with those who received lifestyle intervention alone (n = 130). At time of removal (6 months), the %TBWL in the Orbera group (n = 116) was 10.2 ± 6.6% versus 3.3 ± 5% in the control group (n = 99). At 12 months, patients in the balloon arm maintained more than 70% of their weight lost at 6 months.[52] In a metaanalysis of over 65 studies, accommodative symptoms after placement, such as abdominal pain and nausea, were frequent side effects, occurring in 29% to 34% of patients, but these symptoms mostly dissipated after a few days of therapy. Serious side effects such as migration and gastric perforation were 1.4% and 0.1%, with 50% of the perforations occurring in those with prior gastric surgery, which is a contraindication for placement in the United States.[53,54]
2. The **Obalon** balloon is a gas-filled balloon system packaged within a gelatin capsule. The capsule contains a self-sealing valve that is connected to a thin catheter extending from the stomach to the mouth when the capsule is ingested. Capsule entry into the stomach is verified by fluoroscopy or a percutaneous locating device that does not use radiation; upon entry into the stomach, the gelatin dissolves, freeing the balloon. The catheter is then used to inflate the balloon with a nitrogen gas mixture. The inflated balloon detaches from the catheter and remains in place until it is removed endoscopically after 12 to 26 weeks. Up to three capsules can be swallowed during the same session, or preferably sequentially throughout therapy as weight loss plateaus. In a US pivotal multicenter randomized sham-controlled clinical trial, investigators compared 198 patients receiving up to three consecutive balloon capsules plus lifestyle intervention to 189 patients receiving sham capsules in addition to lifestyle intervention. %TBWL in the IGB balloon arm was 6.6 ± 5.1% at 6 months from the first swallowed capsule compared with 3.4 ± 5.0% in the control group. The safety profile of the Obalon IGB system is favorable. In the pivotal trial, no unanticipated adverse device events occurred, and only one serious adverse event occurred, which was a gastric ulcer in the setting of nonsteroidal antiinflammatory medication use. Transient mild to moderate nausea and abdominal pain were common and self-limited, with the vast majority of patients completing therapy.[55]
3. **Spatz** is an adjustable fluid-filled IGB placed endoscopically. The balloon has an extractable thin, flexible tube that allows for endoscopic volume adjustment while the balloon is in place. The balloon volume may be decreased to improve tolerability or increased to aid in increased weight loss. Prior generations of the Spatz adjustable balloon resulted in up to 20% TBWL at 12 months but had device-related complications.[56] The new-generation device overcomes these design flaws and is approved for 12-month placement outside of the United States. A US multicenter pivotal trial randomized 288 participants (187 IGB and 101 open-label control). At 32 weeks the IGB cohort achieved 14.9 ± 7.2% TBWL compared with 3.6 ± 5.8% in the control group, for a difference of 11.3%([97.5% lower confidence bound 9.1%). The serious adverse events rate was 5.3%, and these events were mostly related to persistent accommodative gastrointestinal symptoms.[57] The device is currently awaiting regulatory approval.
4. The **Elipse** balloon is the first procedureless IGB that does not require an endoscopy to place or to remove. It is administered by swallowing a capsule with an attached microcatheter through which the balloon is filled once the balloon capsule is confirmed to be within the stomach on an abdominal x-ray. The balloon subsequently self-deflates after degradation of its valve mechanism, permitting release of balloon contents and the entire device and contents are expected to pass through the gastrointestinal tract and be excreted spontaneously in the stool approximately 4 months after administration. In a systematic review and metaanalysis including 2013 unique patients with mild to moderate obesity, the pooled early removal rate was 2.3% (95% confidence interval [CI]: 1.1%–3.5%; I2 31%) and pooled %TBWL after completion of treatment (4–6 months) was 12.8% (95% CI: 11.6%–13.9%; I2 83%) and at 12 months was 10.9% (95% CI: 5.0%–16.9%, I2 98%). Three patients in this study developed small bowel obstruction with balloon exit, and one patient had gastric perforation requiring surgery.[58] A pivotal US regulatory multicenter randomized sham-controlled trial has concluded, and the balloon is awaiting regulatory approval.
5. The **TransPyloric Shuttle** consists of a large spherical bulb attached to a smaller cylindrical bulb with a flexible tether. The device is delivered and removed through an overtube with endoscopic guidance. The large bulb remains in the stomach, while the smaller bulb can traverse to the duodenum during peristalsis. It creates an intermittent gastric outlet obstruction to delay the gastric emptying of ingested meals and induce satiety. The FDA approved this device in April 2019 for use in patients with a BMI of 30 to 40 kg/m^2 for 12 months. A US pivotal multicenter randomized double-blinded sham-controlled trial (ENDobesity II) involving 270 patients (181 TransPyloric Shuttle and 89 sham) led to this approval. At 12 months, the %TBWL was 9.5% for the TransPyloric Shuttle group, compared with 2.8% in the sham group. The incidence of serious adverse events in the TransPyloric Shuttle group was 2.8%, in which one patient experienced esophageal rupture that occurred

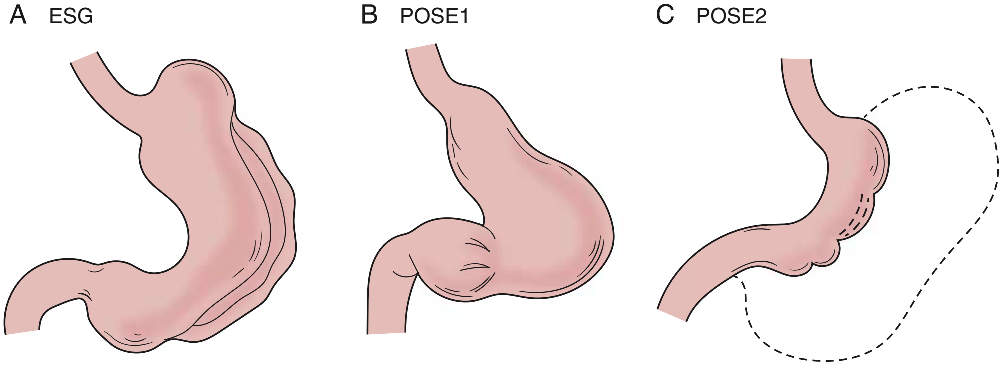


Figure 27.4 Gastric contour following different endoscopic gastric remodeling techniques. **A**, Endoscopic sleeve gastroplasty. **B**, Primary obesity surgery endoluminal 1. **C**, Primary obesity surgery endoluminal 2.

during an unsuccessful delivery attempt. The early removal rate due to adverse events was14.8%.(https://www.accessdata.fda.gov/cdrh_docs/pdf18/P180024B.pdf)

Gastric Remodeling Techniques

Endoscopic Sleeve Gastroplasty. Endoscopic sleeve gastroplasty (ESG) is an incisionless gastric volume reduction technique that creates a restrictive endoscopic sleeve. This endoscopic sleeve is achieved using a commercially available full-thickness suturing device (OverStitch; Apollo EndoSurgery, Austin, TX, USA) by placing a series of triangular full-thickness sutures through the gastric wall along the greater curvature of the stomach from the prepyloric antrum to the gastroesophageal junction, leaving a small pouch in the fundus (Fig. 27.4A).[59] In a systematic review and metaanalysis of eight studies involving 1776 patients with a follow-up of up to 2 years, the body weight loss at 6, 12, and 18 to 24 months was 15.1%, 16.5%, and 17.2%, respectively. The reported rate of serious adverse events was 2.2%.[60] Although most studies evaluated the impact of ESG on body weight, two evaluated its association with important metabolic parameters.[61,62] Sharaiha and colleagues observed a significant improvement in HbA1c, systolic blood pressure, triglycerides, and alanine aminotransferase 12 months post-ESG.[61] In another study, all patients with hypertension and 76.5% of patients with early type 2 diabetes mellitus achieved remission of these entities.[62] Limited longer-term studies have reported outcomes beyond 1 year.[63] However, beyond sustained weight loss, the postprocedural anatomy has not been routinely assessed beyond the immediate postprocedural phase. It is likely that considerable variability may be present, as the techniques, suture patterns, and follow-up protocols continue to evolve.

Primary Obesity Surgery Endoluminal. In its first iteration, primary obesity surgery endoluminal (POSE) used a peroral incisionless operating platform (USGI Medical, San Clemente, CA) to place transmural tissue anchor plications that reduce accommodation of the gastric fundus. Three additional plications are placed in the distal gastric body to delay gastric emptying (see Fig. 27.4B). POSE has been associated with both improved satiety, indicating a gut neurohormonal response.[64] A pivotal US multicenter randomized blinded clinical trial that compared 221 patients receiving POSE in addition to a low-intensity lifestyle intervention for 12 months with 111 patients receiving lifestyle intervention alone failed to meet its primary efficacy endpoints, with a reported weight loss at 12 months for the POSE group of 5 ± 7% compared with 1.4% ± 5.6% in the control group.[65]

An alternative plication method, POSE2.0, where plications are placed in the gastric body between the gastroesophageal junction and level of the incisura, with none placed in the fundus, thereby generating a small food reservoir in the fundus with a tubular configuration of the remainder of the stomach, is being investigated, with encouraging early results (see Fig. 27.4C).[66]

Aspiration Therapy

Aspiration therapy (AT) is another endoscopic treatment approach involving partial disposal of an ingested meal to reduce overall caloric intake through a modified percutaneous endoscopic gastrostomy (PEG) tube. This device is FDA-approved for use in the United States in patients with a BMI between 35 and 55 kg/m^2, a higher BMI range than for IGBs, for up to 5 years. It requires endoscopic placement and removal, unlike standard PEG tubes. The AT process uses a specifically designed percutaneous gastrostomy tube (A-Tube, Aspire Bariatrics, King of Prussia, PA) made of silicone. The tube is inserted in a similar fashion as standard PEG tubes. After 2 weeks, the A-tube is shortened to the level of the skin, and the skin port is attached. The skin port allows attachment to the aspiration device, which fits in a small zippered bag. Once the device is connected to the skin port, a water reservoir flushes tap water into the stomach to facilitate disposal of food. Patients are instructed to perform aspiration approximately 20 minutes after ingesting a meal, removing and discarding approximately one third of the meal. The process takes approximately 10 to 15 minutes to complete. The device has a built-in counter that locks the device to avoid overuse. In a US pivotal multicenter randomized controlled trial, 111 subjects underwent AT with lifestyle intervention for 12 months compared with 60 subjects who underwent lifestyle intervention alone. At 12 months, %TBWL for those who completed AT (n = 82) was 14.2% ± 9.8% compared with 4.9% ± 7% in the control group (n = 31). Because aspiration removes around 30% of ingested calories, it is estimated that AT accounts for 80% of the reported weight loss. AT requires patients to chew food thoroughly and eat slower, potentially leading to new eating behaviors. As far as safety, the adverse events included stoma granulation tissue (40.5%), stoma infection (14.4%), peritonitis (0.9%), and gastric ulcer (0.9%). Careful review of disordered eating behavior questionnaires found no evidence of worsening eating behaviors with therapy.[67]

Small Bowel Intervention

Duodenal-Jejunal Bypass Liner. The Endobarrier (GI Dynamics, Lexington, MA) is a 65-cm-long Teflon-coated duodenal jejunal bypass sleeve. It is anchored at the duodenal bulb by an anchoring crown and terminates at the proximal jejunum, allowing ingested nutrients to bypass the duodenum to the jejunum, thereby altering proximal duodenal interaction with ingested nutrients and subsequent

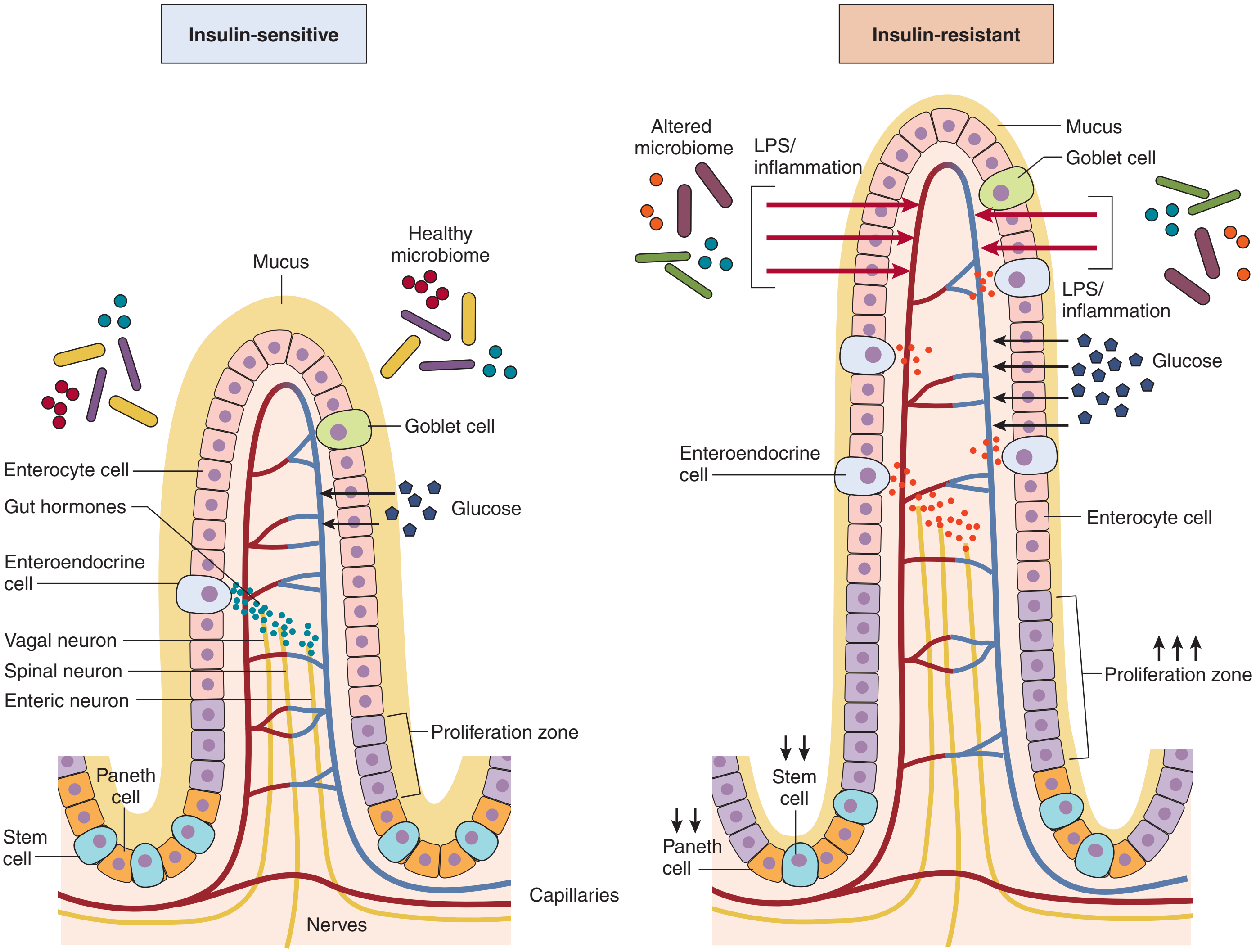

Insulin-sensitive
Insulin-resistant
Mucus
Healthy microbiome
Goblet cell
Enterocyte cell
Gut hormones
Glucose
Enteroendocrine cell
Vagal neuron
Spinal neuron
Enteric neuron
Proliferation zone
Paneth cell
Stem cell
Capillaries
Nerves
Altered microbiome
LPS/ inflammation
Mucus
Goblet cell
LPS/ inflammation
Glucose
Enteroendocrine cell
Enterocyte cell
↑↑↑
Proliferation zone
↓↓
Stem cell
↓↓
Paneth cell

Figure 27.5 **Intestinal adaptation to excess caloric intake.** These hypothesized pathophysiologic adaptations have been targeted by endoscopic technologies capable of resurfacing the mucosal layer of the duodenum in an attempt to reset the abnormal metabolic signaling and improve insulin sensitivity.

gut hormonal and enteric nervous system response. This device is placed and removed endoscopically and remains in the small intestine for 12 months. A US pivotal multicenter randomized sham-controlled trial (the ENDO trial) was terminated early after enrollment of 325 of 500 patients because of a 3.5% incidence of hepatic abscess. All were managed conservatively, but this was outside the predefined adverse event rate. An analysis of published trials of this device identified two studies that included a sham control arm with a total of 77 volunteers who received treatment for 12–24 weeks. The analysis was specifically aimed at assessing the effects of this approach on T2DM outcomes; for these two trials there was not a significant improvement in HbA1c between intervention and control studies. A redesigned clinical trial of the same device with a modified protocol designed to reduce the likelihood of adverse events has been approved by the FDA to begin enrolling patients. Furthermore, a newer atraumatic bypass liner system is being developed.

Duodenal Mucosal Resurfacing. Intestinal adaptation to excess caloric intake and states of metabolic disease has been hypothesized and is illustrated in Fig. 27.5.[50,68,69] These pathophysiologic adaptations have been targeted by endoscopic technologies capable of resurfacing the mucosal layer of the duodenum in an attempt to reset abnormal metabolic signaling and improve insulin resistance. Duodenal mucosal resurfacing (DMR; Fractyl, Lexington, MA), or the Revita procedure, is an endoscopic procedure that involves circumferential hydrothermal ablation of the duodenal mucosa. This procedure is primarily aimed at glycemic control. Though the precise mechanism is not yet defined, it is hypothesized that DMR induces remodeling of the diseased duodenum with an overgrowth of enteroendocrine cells and dysregulated secretion of gastrointestinal hormones, which have been observed in patients with diabetes. The first human study of 39 patients with type 2 diabetes demonstrated a 1.2% reduction of HbA1c and a 2.5-kg weight reduction at 6 months. A longer segment ablation resulted in more potent glycemic control. Three patients developed duodenal stenosis, which was successfully treated with endoscopic dilation.[70] In a double-blind, randomized controlled trial (56 DMR, 52 sham) of patients with poor glycemic control despite greater than or equal to one stable oral antidiabetic medications with 24 weeks' follow-up, the median HbA1c decreased by 10.4 mmol/mol in the DMR group versus 7.1 mmol/mol in sham group ($P = 0.147$). Because of the heterogeneity between the two cohorts participating in the trial (European vs. Brazilian) the groups were analyzed separately, as prespecified in the statistical analysis plan. In the European modified intention to treat cohort (39 DMR, 37 sham) the median HbA1c change was –6.6 mmol/mol versus –3.3 mmol/mol post-sham ($P = 0.03$). The Brazilian cohort showed a statistically nonsignificant trend towards DMR benefit in HbA1c. Patients with high baseline fasting plasma glucose, greater than or equal to 10 mmol/L, had significantly greater reductions in HbA1c post-DMR versus sham ($P = 0.002$). One jejunal perforation was observed in this study requiring surgical repair.[47] A US multicenter randomized sham-controlled trial of hydrothermal ablation is underway. A newer ablation technique, utilizing electroporation in hopes of reducing serious adverse events and to enable safe re-treatment of the duodenum, is currently in early human trials.

Incisionless Magnetic Anastomosis System. The Incisionless Magnetic Anastomosis System (GI Windows, West Bridgewater, MA) consists of two self-assembling magnets that are deployed, one each in the proximal jejunum by enteroscopy and in the ileum by colonoscopy, with ileoscopy creating jejunoileal anastomosis, which results in partial jejunoileal bypass. The coupled magnets then pass spontaneously after a week. The first human proof-of-concept study involving 10 patients demonstrated a %TBWL of 14.6% and a significant reduction of HbA1c of 1.9% at 12 months utilizing laparoscopic assistance and monitoring. All anastomoses were patent at a 12-month follow-up.[71]

Special Considerations and Future Directions. The magnitude and duration of weight loss varies by EBMT device and technique. Therefore, EBMTs should be administered with the support of a comprehensive team of primary care providers, gastroenterologists, nutritionists, dieticians, and behavioral therapy professionals. Moving forward, personalized EBMTs targeting clinical phenotype and responders' profiles will be imperative. The sequential and tandem use of EBMTs in conjunction with antiobesity pharmacotherapy and comprehensive lifestyle interventions is an ongoing area of research, which may yield a robust and durable weight loss response, achieving a lasting effect on obesity and its comorbidities. In a proof-of-concept propensity score-matched study of 52 patients, those receiving endoscopic sleeve gastroplasty with liraglutide achieved 25% TBWL at 12 months (7 months after the initiation of liraglutide) compared with 21% in the ESG-only group.[72]

KEY POINTS

- Endoscopic bariatric and metabolic therapies (EBMT) are minimally invasive peroral management options that could augment the treatment spectrum for obesity and metabolic disease.
- EBMTs are generally considered in patients who do not qualify for or wish to pursue bariatric surgery, especially in those with class 1–2 obesity who have previously failed to lose weight using lifestyle methods.
- EBMTs include gastric space–occupying devices, gastric remodeling techniques, and small bowel interventions targeting metabolic pathways.
- EBMTs should be administered with the support of a comprehensive multidisciplinary team. The sequential and tandem use of EBMTs in conjunction with antiobesity pharmacotherapy and comprehensive lifestyle interventions is an ongoing area of research.

Eligibility for Bariatric Operations and Procedures

According to the current National Heart, Lung, and Blood Institute guidelines, candidates for bariatric surgery in the United States include patients with a BMI greater than 40 kg/m^2 or with a BMI of 35 to 39.9 kg/m^2 and one or more obesity-associated comorbidity (type 2 diabetes, cardiovascular disease, sleep apnea, dyslipidemia, hypertension, GERD, or osteoarthritis).[72a] These indications, which were established more than 25 years ago, may expand as more societies adopt this safe and durable weight loss option. In 2018, the ASMBS published a statement on the role of metabolic surgery in class 1 obesity. The issued statement supported lowering the BMI cutoff of 35 kg/m^2 to 30 kg/m^2 in consideration of the high-quality supporting data available.[73] At the very least, prospective patients must undergo a thorough preoperative evaluation by a bariatric surgeon, a dietician, and a psychologist. An endocrinologist's expertise is often necessary to assist in managing the metabolic complications of obesity. This multidisciplinary approach ensures consideration of the "whole patient," thus leading to favorable and long-lasting outcomes that are not solely dependent on behavioral changes induced by the operation.

KEY POINTS

- To be eligible for bariatric surgery or endoscopic bariatric procedure, patients must have a body mass index (BMI) >40 kg/m^2 or a BMI of 35–39.9 kg/m^2 and one or more obesity-associated comorbidity (type 2 diabetes, cardiovascular disease, sleep apnea, dyslipidemia, hypertension, gastroesophageal reflux disease, or osteoarthritis).
- There is mounting evidence supporting the application of bariatric operations and procedures for class 1 obesity (BMI 30.0–34.9 kg/m^2), particularly for patients in this weight class with type 2 diabetes mellitus.

Assessing Outcomes after Bariatric Interventions

The two main outcomes important in determining the success or failure of bariatric operations and procedures are: 1) weight loss and 2) comorbidity resolution. Many previous bariatric surgery studies measured outcomes as percent EWL, which is defined as the percent change from excess to ideal body weight. However, %TBWL is currently preferable, because this standardization is more easily comparable to the medical literature. %TBWL has the lowest variation coefficient and best describes weight loss changes over time for patients in different weight categories.[56,74] However, the data obtained from %TBWL may not accurately reflect the clinical benefit. For example, someone with greater excess weight will need to lose more weight to achieve the clinical benefits. For this reason, %EWL is sometimes preferred.

Weight loss is greatest after BPD-DS, and is generally in the range of %TBBL 30% to 40%, %EWL 70% to 90%, followed by RYGB (%TBL 28%–32%, %EWL 65%–80%) and SG (%TBL 25%–32%, %EBL 50%–65%). Recent randomized studies comparing the RYGB with the SG have demonstrated equivalent weight loss at 5 years.[75,76] However, it is generally believed that the RYGB results in superior weight loss compared with the SG over the long term. It should be remembered that weight loss outcomes depend on multiple factors, perhaps most importantly long-term compliance to behavioral modifications, resulting in variable outcomes. Generally, the operations with the greatest weight loss also carry the greatest risk for surgical complications, including the development of nutritional deficiencies.

Determining which operation is most appropriate for a specific patient depends on several factors, most importantly which diseases need treatment, in addition to the degree of obesity. An analysis of the risks and benefits of each operation for each patient should be undertaken that incorporates medical comorbidities, weight loss needed to obtain health goals, likelihood of patient compliance, and patient preference.

THE PHYSIOLOGY OF WEIGHT LOSS AFTER BARIATRIC OPERATIONS AND PROCEDURES

Weight Loss Mechanisms

All commonly performed bariatric procedures involve a restrictive component that may or may not impact intestinal nutrient handling. Thus, in addition to caloric restriction, altered gastrointestinal anatomy may affect the entero-neuro-endocrine axis involving signaling from multiple gut hormones and metabolites that influence appetite and satiety through endocrine or neural pathways to a varying degree following different bariatric procedures (see Fig. 27.3).

Caloric Restriction

Caloric restriction creates an energy deficit that leads to mobilization of stored energy and weight reduction. The rate and magnitude of weight reduction are related to the degree of energy deficit produced, with very low-calorie diets (<800 kcal/day) leading to more rapid weight loss than low-calorie diets (800–1500 kcal/day) but being more challenging to sustain, and so resulting in a similar effect on long-term weight management.[77]

Diet-induced weight reduction of greater than 10% body weight can be achieved in 8 to 10 weeks but results in neurohormonal adaptation that increases hunger and may promote weight regain in the long-term.[78] Energy expenditure is reduced to a greater extent than that predicted by weight loss alone, an effect shown to persist for more than a year in individuals who have maintained a reduced body weight[79] and favoring weight regain during long-term weight management.

Short-term implementation of a very low–calorie diet, with or without the assistance of an IGB, may be considered for selected patients before bariatric surgery in order to reduce hepatomegaly and visceral adiposity, which can help to reduce surgical risk. Some studies suggest that even modest reduction of body weight by 5% to 10% in the immediate preoperative period may reduce the risk of operative complications[80] and postoperative death.[81]

Gastrointestinal Anatomy Changes

Eating behaviors differ in patients who have had bariatric surgery compared with those on low- calorie diets. Diet-induced weight loss tends to decrease satiety, increase hunger, and drive cravings for energy-dense foods.[82,83] In contrast, patients describe early satiety and less hunger, and the compensatory consumption of energy-rich foods is not observed following bariatric surgery.[84-87] In part, these behavioral changes are related to adaptive learning to avoid adverse gastrointestinal symptoms after surgery, such as nausea, vomiting, diarrhea, reflux, bloating, and dumping syndrome. The character of these symptoms will vary due to differences in the surgical procedures and postsurgical adaptation.

Gastrointestinal and Adipocyte Hormonal Changes

Ghrelin, a peptide hormone mainly produced by X/A cells in the oxyntic glands of the fundus or body of the stomach, acts on the G protein–coupled receptor known as the *growth hormone secretagogue receptor*[9] to stimulate neuropeptide Y–agouti-related protein neurons within the arcuate nucleus (ARC), thereby stimulating hunger and increasing food intake.[88] Most studies show reduced circulating ghrelin following RYGB[89,90] and SG,[91] but increased concentrations following adjustable gastric banding[84] and low-calorie diet,[90,78] suggesting that gastric exclusion may help to reduce hunger that is otherwise increased during weight reduction.

GLP-1, PYY, and oxyntomodulin are anorexic hormones released after meal ingestion from L-cells predominantly located in the mucosal epithelium of the ileum and colon. These peptides act at the hypothalamic ARC and brainstem to induce satiety and reduce food intake.[92] Postprandial levels of these anorexic hormones increase after RYGB[85,93-95] and SG,[82,96] but not with diet-induced weight loss[78,94] or adjustable gastric banding.[97] Similarly, cholecystokinin (CCK), another anorexic peptide released in response to food intake from I-cells located in the mucosal epithelium of the duodenum and jejunum[98] or released in response to GLP-1 or duodenal long-chain free fatty acids,[99] also increases following RYGB and SG.[100,101] In addition, FGF19, which is produced in the ileum in response to bile acid absorption, is increased following RYGB but not with adjustable gastric banding, and has been shown to be an anorexigenic hormone in rodents.[102] Thus, altered intestinal delivery and handling of nutrients and bile acids may act through gastrointestinal hormone changes to promote satiety and contribute to reduction in food intake.

The adipokine adiponectin functions as an insulin-sensitizing hormone regulating glucose and lipid metabolism,[103] with anti-inflammatory properties and beneficial effects on type 2 diabetes, atherosclerosis, obesity, and nonalcoholic fatty liver disease.[104,105] Adiponectin levels increase following bypass and restrictive surgeries, and the magnitude of increase may be larger following RYGB than band surgery, or compared with weight loss by caloric restriction.[83] Leptin, also secreted by adipocytes, acts primarily on the hypothalamus to increase energy expenditure and reduce food intake. Leptin functions as an indicator of energy stores, and so levels decrease in proportion to weight reduction.[106] It remains unclear to what extent changes in adiponectin, leptin, or other adipokine concentrations or activity contribute to weight loss durability and metabolic improvements beyond that seen with weight loss achieved through dietary intervention alone.

Central Nervous System Control

Regulation of food intake can be influenced by the central nervous system by affecting food-seeking behavior and satiety perception. The hypothalamus, frontal cortex, and caudal brainstem regulate weight homeostasis by detecting internal and external signals to stimulate appetite or satiety accordingly.[107] The hypothalamus contains two groups of neurons within the ARC with opposite effects: the proopiomelanocortin (POMC)-derived peptides, which act via melanocortin receptor 4 to reduce food intake and increase energy expenditure, and the orexigenic peptides neuropeptide Y (NPY) and Agouti-related protein (AgRP), which drive food intake.[87] The expression of these peptides (POMC, NPY, and AgRP) may change following bariatric surgery, as illustrated in studies of rats following vertical SG compared with pair-fed rats undergoing sham surgery.[108] AgRP levels were unchanged in the vertical SG group compared with pair-fed animals that had elevated AgRP. These findings suggest that calorically-restricted rats were hungry compared with those with vertical SG.

Neural systems of reward, cognition, and emotion may also contribute to regulation of eating behavior post–bariatric surgery, as shown in studies in humans demonstrating altered effects of food visual cues on brain activity in hedonistic centers assessed by functional magnetic resonance imaging (fMRI).[109] Post-RYGB patients had lower activation of brain reward systems, particularly to energy-rich foods, and chose healthier food options compared with BMI-matched controls or patients who had undergone adjustable gastric band.[109] Suppression of postprandial increases in the satiety hormones PYY, GLP-1, and FGF19 by octreotide increased brain-hedonic responses to anticipatory food reward as assessed using fMRI in individuals following RYGB, but not after adjustable gastric banding.[110] These findings are consistent with the hypothesis that increased gastrointestinal hormone effects resulting from altered gastrointestinal tract anatomy may contribute to long-term regulation of food intake.

Vagal Signaling

Vagal nerve afferent activity also contributes to food intake.[111] Stimulation of the vagal nerve by increased intraluminal pressures may contribute to reduced food intake and weight loss in patients who receive an adjustable gastric band. Patients with optimally filled bands are less hungry compared with patients with suboptimally filled or empty bands, as seen in a randomized, masked crossover study.[84] However, vagal nerve preservation does not influence clinical parameters (weight loss and gastrointestinal symptoms) and satiety scores compared with severing of the vagal nerve during RYGB procedures.[112] Similarly, in animal studies, no significant difference in body weight, body composition, meal patterns, and energy expenditure were found between rats undergoing RYGB alone and RYGB with common hepatic branch vagotomy.[113] Thus, further studies are required to explore the role of the vagus nerve in weight loss induced by bariatric surgical procedures.

Energy Expenditure

Exercise capacity can improve following bariatric surgery.[114] Additionally, human and animal studies demonstrate a lesser reduction in resting energy expenditure following RYGB compared with those with diet-induced weight loss.[115-117] Mechanisms underlying this phenomenon remain incompletely understood but could include increased fitness from increased physical activity after weight loss or mechanisms such as bile acid activation of brown adipose tissue or skeletal muscle oxidation.[118]

KEY POINTS

- Weight loss that follows bariatric procedures is dependent on caloric restriction. In addition, altered gastrointestinal anatomy may affect the entero-neuro-endocrine axis involving signaling from multiple gut hormones and metabolites that influence appetite and satiety.
- Gastric exclusion and altered intestinal delivery and handling of nutrients and bile acids may act through these gastrointestinal hormone changes to promote satiety and reduce hunger, which is otherwise increased during weight reduction.
- Neural systems of reward, cognition, and emotion may also contribute to regulation of eating behavior following bariatric procedures.

PHYSIOLOGIC EFFECTS OF BARIATRIC SURGERY ON CARDIOMETABOLIC HEALTH AND MORTALITY

Diabetes Remission Following Bariatric Surgery

Improvements are seen in glycemia, hypertension, and dyslipidemia in patients with type 2 diabetes following bariatric surgery. A metaanalysis of primarily observational cohort studies by Buchwald and colleagues reported that 78% of patients with type 2 diabetes undergoing bariatric surgery had complete remission of type 2 diabetes, defined as normoglycemia without medication, and 86% had either improvement or resolution of diabetes.[119] Patients who underwent biliopancreatic diversion had the highest remission rate, at 95%, followed by RYGB at 80%, and adjustable gastric banding at 57%. A review of randomized clinical trials by Arterburn and colleagues reported remission of type 2 diabetes in 22% to 90% of patients after RYGB, 19% to 73% after adjustable gastric banding, and 0% to 23% with diet intervention alone.[120] The wide variation in clinical response is thought to be due to a large variance of severity and duration of type 2 diabetes among the studies. Surgical approaches may also reduce the incidence of new diabetes.[121]

When considering patients with type 2 diabetes, greater weight loss and better glycemic outcomes are seen in the early years following bariatric surgery in randomized studies that compare surgical with medical diabetes and weight-management approaches.[122-125] In a randomized clinical trial comparing biliopancreatic diversion, RYGB, and medical therapy, diabetes remission rates of 95% following biliopancreatic diversion and 75% following RYGB compared to 0% with medical therapy at 2 years[124] were sustained at 68% and 42% compared with 0%, respectively, after 5 years of follow-up.[126] In a randomized clinical trial comparing RYGB, SG, and medical therapy, diabetes remission rates of 46% following RYGB and 29% following SG compared with 0% with medical therapy at 3 years[127] were sustained at 31% and 23% compared with 0%, respectively, after 5 years of follow-up.[128] In a randomized clinical trial comparing RYGB and medical therapy, diabetes remission rates of 42% following RYGB and compared with 0% with medical therapy at 2 years[129] were sustained at 16% compared with 0%, respectively, after 5 years of follow-up.[130] In a randomized clinical trial comparing RYGB, adjustable gastric banding, and medical therapy, diabetes remission rates at 5 years were 30% following RYGB and 19% following adjustable gastric banding with to 0% with medical therapy.[131] These studies demonstrate both durable glycemic improvement following bariatric surgical intervention when compared with intensive lifestyle modification in type 2 diabetes, but also the recurrence of diabetes in some individuals over time postoperatively.

While more durable weight reduction with surgical compared with medical intervention likely explains the high numbers of patients who may achieve diabetes remission following bariatric surgery, unsurprisingly, remission from diabetes may not be permanent.[132] Approximately 35% of patients who initially experience remission may relapse within 5 years, with a median duration of remission of approximately 8.3 years. Reoccurrence of diabetes is often associated with weight regain and dietary noncompliance. However, several factors predict lower initial remission and/or subsequent relapse, including poor preoperative glycemic control, longer duration since diagnosis of diabetes, insulin use,[132] and lower β cell glucose sensitivity,[133] together suggesting that adequate residual β cell function is important for initial remission, and that progressive disease may contribute to relapse. However, as many studies demonstrate important clinical benefits from a period of glycemic control on subsequent development of complications (metabolic memory),[134] even moderate-duration remission may be clinically important. Emerging evidence suggests that RYGB may lead to more durable weight reduction and improvement in glycemic control than SG.[120]

Caloric restriction and weight loss each have a substantial impact on improvement and remission of diabetes, but there is great interest in the hypothesis that additional mechanisms contribute to improvement in glucose homeostasis, and this remains heavily debated. While glycemic improvements occur within days following RYGB,[135,136] this effect may be explained by the perioperative caloric restriction,[137] as observed with other gastrointestinal surgery. Weight loss achieved by RYGB and SG is greater than that achieved by diet and exercise or conventional medical therapy, which most likely explains the greater improvement in glucose tolerance with surgical compared with medical intervention.[124] Neither liposuction nor surgical resection of omental fat corrects diabetes,[138,139] such that fat loss alone is insufficient for metabolic improvement.

The Role of Insulin Sensitivity and Insulin Secretion in Diabetes Remission after Gastric Bypass

Persons with type 2 diabetes are typically more insulin-resistant than persons without diabetes. There are parallel improvements in insulin sensitivity following RYGB in persons with or without type 2 diabetes, although those with diabetes continue to have lower insulin sensitivity at any given BMI.[133] Improvements in glucose utilization are greater early after RYGB in persons with type 2 diabetes compared with those without diabetes, presumably due to elimination of glucotoxic effects. However, by 1 year glucose utilization is similarly improved for those who did or did not have diabetes at the time of surgery. Early postoperative improvements in glycemia are mediated in part by improved insulin suppression of endogenous (primarily hepatic) glucose production, accompanied by decreased oxidative glucose disposal and increased lipid oxidation.[140] Effects of biliopancreatic diversion on insulin sensitivity appear greater in proportion to weight lost.[141] Following RYGB, hepatic insulin sensitivity is increased significantly in individuals with type 2 diabetes compared with nonsurgical controls with type 2 diabetes assessed at matched 10% weight loss; however, time to 10% weight reduction occurred more quickly with surgical versus nonsurgical intensive lifestyle intervention,[142] indicating that greater caloric restriction may explain the difference in hepatic glucose metabolism.[143,144] In contrast, improvement in insulin sensitivity following adjustable gastric banding appears proportionate to weight lost.[93] Similar improvements in insulin sensitivity have been reported in individuals who achieve 20% weight reduction, whether by RYGB, SG, or gastric banding.[145,146]

There is a robust early increase in insulin secretion, GLP-1, and β cell glucose sensitivity following a mixed meal after RYGB, but not adjustable gastric banding, at similar weight loss.[147] This is best explained by altered nutrient transit with accelerated delivery to the distal small intestine, leading to more rapid glucose absorption and exaggerated GLP-1 and insulin responses.[145,148] Although improved from baseline, β cell glucose sensitivity after RYGB remains lower in patients with type 2 diabetes compared with those without.[133] Importantly, despite an exaggerated postsurgical GLP-1 response to nutrient ingestion, postprandial glucose tolerance improved similarly in individuals with type 2 diabetes after RYGB or intensive lifestyle modification when assessed at a similar 10% weight reduction.[142] Furthermore, blockage of the GLP-1 receptor during meal ingestion resulted in similar deterioration of postprandial glucose tolerance in those treated by surgery or diet, providing evidence against a greater contribution of GLP-1 to postprandial glucose metabolism following surgical weight loss.[142]

The Hindgut and the Foregut Hypotheses

Two general leading hypotheses have been proposed to explain the effects of bariatric surgery on improved glycemia: the hindgut hypothesis and the foregut hypothesis. The hindgut hypothesis proposes that glycemic improvement is due to accelerated nutrient delivery to the distal intestine, which augments secretion of hormones that affect satiety and glucose homeostasis, including GLP-1, PYY, oxyntomodulin, and/or CCK.[101] However, increased delivery of nutrients to the proximal small intestine as occurs with RYGB and SG may stimulate distal enteroendocrine cell secretion via activation of an entero-enteric neural reflex. The foregut hypothesis suggests that exclusion of the proximal intestine from ingested food promotes antidiabetic effects by inhibiting an unidentified signal that promotes insulin resistance,[149] but it is more likely that altered nutrient digestion and accelerated distal delivery are responsible for these effects.

Branched-Chain Amino Acids and Aromatic Amino Acids

The branched-chain amino acids (BCAAs) valine, leucine, and isoleucine are among the nine essential amino acids (AAs) available to humans as dietary proteins. BCAAs and the aromatic AAs, especially phenylalanine and tyrosine, are associated with obesity and insulin resistance,[150-152] and BCAA concentrations decrease with weight loss.[150] Greater reductions in circulating BCAAs and AAs are seen in type 2 diabetes following gastric bypass compared with patients with nonsurgical, matched weight loss.[153] Reductions in BCAA correlate with improved insulin sensitivity estimated by homeostatic model assessment of insulin resistance or euglycemic clamp.[152,154] However, other studies showed BCAA decrease similarly after RYGB compared with matched weight loss following adjustable gastric banding in nondiabetic individuals,[152] so further studies are required to determine the causal significance of the relationship between AAs and modification of insulin sensitivity, and whether the magnitude of effects differ following surgically-induced weight loss.

Bile Acids

Bile acids are well recognized for their role in lipid metabolism. Recently they have been found to have additional metabolic effects affecting gut peptide secretion, energy expenditure, and glucose homeostasis.[155] RYGB leads to changes in the delivery of bile acid and nutrient flow through the gut. Diverting bile from the common bile duct to drain into distal jejunum with a catheter increases serum bile acids and postprandial GLP-1 response and improves glucose tolerance and hepatic steatosis compared with sham-operated rats.[156] Bile acids promote GLP-1 secretion through the G protein–coupled receptor TGR5.[157] Intrarectal infusion of the bile salt taurocholic acid increases plasma GLP-1, PYY, and insulin concentrations in obese men with type 2 diabetes,[158] and oral tauroursodeoxycholic acid improves liver and muscle insulin sensitivity in obese men.[159] Circulating bile acid concentrations increase following RYGB.[160,161] Increased circulating bile acids are associated with increased GLP-1 and improved glycemia,[160,162] but not with skeletal muscle insulin sensitivity, insulin response to mixed meal, or resting energy expenditure.[162] Changes in circulating bile acids are less pronounced after adjustable gastric banding compared with RYGB.[162] The relative importance of altered bile acids in improved metabolism also remains incompletely understood.

Gut Microbiota

The gut microbiota profile differs in individuals with obesity and metabolic syndrome compared with lean individuals. Recently, gut microbiota have been recognized to alter production or bioavailability of small molecules, proteins, and nutrients, and to contribute to inflammation in the host.[163,164] Changes in the dynamic symbiotic relationship between the microbiota and the host participate in the progression of obesity and diabetes. However, with a diversity of microorganisms present and a variety of factors that can affect the microbiome composition (such as dietary macronutrients, anatomic variance, antibiotic usage, pH, and bile flow), identifying the specific relationship between various gut microbiota and obesity remains a challenging field of investigation. RYGB causes shifts in the gut microbiome, including a change in ratio of Firmicutes to Bacteroidetes phyla and an increase in colonization with *Proteobacteria*.[165] Transfer of gut bacteria from mice that have undergone RYGB to unoperated germ-free mice promotes weight loss and reduced fat mass in recipient mice, suggesting that changes in gut microbiota contribute to weight loss following RYGB.[166] Changes in gut microbiota may be relevant to humans. Intriguingly, Vrieze and coworkers[167] infused microbiota from lean human donors into the duodenum of obese individuals and showed improved insulin sensitivity, independent of weight change.

Cardiovascular Disease and Mortality

The hallmark prospective Swedish Obese Subjects (SOS) study compared adjustable or nonadjustable gastric banding, the restrictive but now outdated vertical banded gastroplasty, and RYGB to medical management in well-matched but nonrandomized obese cohorts. After bariatric procedures, the combined surgical cohort lost an average of 16.1% of total body weight at 10 years compared with a slight weight gain in the nonsurgical management group.114 RYGB resulted in greater mean weight loss (–25%) than vertical banded gastroplasty (–16.5%) and adjustable gastric banding (–13.2%) at 2 years and 10 years. Overall, the surgical group had more favorable outcomes, with both improved resolution and lower incidence rates of type 2 diabetes, as well as reduced hypertriglyceridemia and hyperuricemia compared with medical management. Hypertension and dyslipidemia likewise improved,[168,169] which together would be anticipated to reduce major cardiovascular event rates and improve survival. Indeed, both the SOS study and a retrospective cohort study demonstrated improved total mortality by 30% to 40% in surgically treated groups.[170,171] Neither study was randomized; thus, referral biases could confound conclusions. Death attributed to cardiovascular disease was reduced by approximately 50%,[170,172] and deaths from diabetes listed on the death certificate were reduced by 90%.[45,170] Incident diabetes was reduced by approximately 80%.[48,121] Indeed, multiple obesity-related comorbidities have been shown to improve following bariatric surgery, including a reduction in cancers and an improvement in metabolic, cardiovascular, reproductive, musculoskeletal, respiratory, and some psychiatric conditions (Table 27.1).[168,171,173-176] With now over 20 years' follow-up in the SOS study, bariatric surgery has remained associated with a lower risk of death from both cardiovascular disease and cancer; however, survival following bariatric surgery remained lower than in the general population.[177]

Other studies that have examined the effect of bariatric surgery on hypertension and dyslipidemia include the GATEWAY study, a randomized trial of the effects of bariatric surgery in obese patients with hypertension, which determined that bariatric surgery represents an effective strategy for blood pressure control in a broad population of patients with obesity and hypertension.[178] A comparative systematic review and metaanalysis determined that obese patients treated with RYGB present a higher hypertension remission rate at 1 and 5 years compared with SG.[179] Other studies have noted that RYGB and SG obtain similar rates of remission of hypertension, but RYGB reaches significantly greater rates of dyslipidemia remission.[180]

KEY POINTS

- Patients with type 2 diabetes experience durable improvements in glycemia, hypertension, and dyslipidemia following bariatric surgery compared with intensive lifestyle modification, although recurrence of hyperglycemia occurs in some individuals over time postoperatively.
- Adequate residual β cell function is important for initial remission of type 2 diabetes, and progressive disease may contribute to relapse. Insulin sensitivity is improved by and proportionate to both caloric restriction and weight loss.
- Multiple additional obesity-related comorbidities have been shown to improve following bariatric surgery, including a reduction in cancers and an improvement in metabolic, cardiovascular, reproductive, musculoskeletal, respiratory, and some psychiatric conditions (Table 27.1).

TABLE 27.1 Improvements in Obesity Comorbid Conditions

Mortality

Multiple cohort matched studies suggest improved mortality.

Cardiometabolic

- Major cardiovascular event rates (myocardial infarction, stroke, and cardiovascular death)
- Type 2 diabetes

Nephropathy related to diabetes

- Type 2 diabetes prevention
- Hypertension
- Dyslipidemia
- Hyperuricemia

Respiratory

- Obstructive sleep apnea
- Obesity hypoventilation

Gastrointestinal

- Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis

Musculoskeletal

- Degenerative joint disease (osteoarthritis of weight-bearing joints)
- General physical functioning and exercise capacity
- Recovery and rehabilitation following joint surgery

Dermatological

- Hirsutism
- Intertrigo and candidiasis
- Wound healing
- Striae/stretch marks
- Acanthosis nigricans

Reproductive System

- Polycystic ovarian disease
- Improved fertility (female and male)
- Regulation of menstrual cycle
- Libido

Oncology

- Reduction in overall cancer risk and mortality

Psychosocial/Neurologic

- Quality-of-life health measures
- Cognitive function (improvement in learning and memory)
- Pseudotumor cerebri (intracranial hypertension)

Urology

- Urinary incontinence

IMPACT OF BARIATRIC SURGERY ON OTHER OBESITY-RELATED COMORBID CONDITIONS (TABLE 27.1)

Polycystic Ovarian Syndrome and Fertility

Obesity negatively impacts fertility in women.[181,182] In the United States, women in their reproductive years account for 83% of those undergoing bariatric surgery.[168,183]

Multiple studies have demonstrated improvements in polycystic ovarian syndrome (PCOS) and the associated clinical sequelae of menstrual irregularity, hirsutism, and infertility after bariatric surgery. One study showed that the incidence of PCOS decreased by nearly 40% at study endpoint, with nearly 50% improvement in menstrual irregularity and 30% improvement in hirsutism.[184] Other studies have confirmed these findings.[185,186] In one, pre- and postoperative anthropometric evaluations, menstrual cyclicity, and ovarian volume (OV), as well as markers of hyperandrogenism, dyslipidemia, and dysglycemia were evaluated and compared to a non-PCOS control group. The PCOS patients had a significant reduction in androgen levels and fewer met the criteria for hyperandrogenism and irregular menses. OV did not significantly decline in either group postoperatively.[187] A retrospective study examining 796 female patients after biliopancreatic diversion found that 47% of women previously unable to conceive preoperatively were successful postoperatively and 82% of these women had appropriate weight gain during pregnancy.[188] Miscarriage rates were not improved (26%).

Pulmonary Disease, Including Obstructive Sleep Apnea

Obesity has significant effects on respiratory function, contributes substantially to the burden of respiratory disease, and is linked with many respiratory symptoms and diseases, including exertional dyspnea, obstructive sleep apnea syndrome, obesity hypoventilation syndrome, chronic obstructive pulmonary disease (COPD), asthma, pulmonary embolism, and aspiration pneumonia. Not surprisingly, the weight loss associated with bariatric surgery has a favorable effect on pulmonary disease. Resolution of obstructive sleep apnea after bariatric surgery occurs in 66% of patients.[189,190] In one study, emergency department visits and hospitalizations for acute exacerbations of COPD decreased in obese patients after bariatric surgery.[191]

Gastroesophageal Reflux Disease

Because obesity is a risk factor for GERD, its prevalence in this population is elevated. GERD has become an important comorbidity to consider, because the two most common operations for obesity have diametrically opposed effects on GERD. *De novo* GERD develops in up to 50% of patients after SG, while those with preexisting GERD symptoms can expect a worsening of their symptoms after SG.[192] Consequently, most consider preexisting moderate or severe GERD and Barrett's esophagus to be a contraindication to SG. On the other hand, the RYGB is an excellent operation for the treatment of GERD. The anatomy of a RYGB diverts both acid (produced from parietal cells generally located in the bypassed stomach) and bile away from the distal esophagus. Seventy-eight to one hundred percent of patients with preexisting GERD experience either improvement or remission of GERD at 12 to 18 months postoperatively.[193,194]

Pseudotumor Cerebri

Idiopathic intracranial hypertension is a rare condition that typically affects young women with obesity. Elevated cerebral spinal fluid pressures in the brain can result in symptoms of headache and visual field loss and blindness due its detrimental effects on neural structures, such as excessive pressure on the ophthalmic nerve.[195] Surgical weight loss can correct this condition; a recent systematic review and comparison of metaanalyses suggested that weight loss surgery was superior to nonsurgical weight loss for the treatment of idiopathic intracranial hypertension.[196]

Cancer

There is a strong association between obesity and many difference types of cancers.[197,198] Although the exact mechanisms have not been established, increasing evidence demonstrates that bariatric surgery decreases this risk. In one multisite cohort of patients with severe obesity, bariatric surgery was associated with a lower risk of incident cancer, particularly obesity-associated cancers, such as postmenopausal breast cancer, endometrial cancer, and colon cancer.[199]

COMPLICATIONS OF BARIATRIC OPERATIONS AND PROCEDURES (TABLE 27.2)

Overall, operations for obesity are remarkably safe, particularly given the pervasive disease of obesity, which affects multiple organ systems and creates a challenging operative environment in terms of inflamed tissues and difficult intraabdominal exposure. Despite these issues, the risk of perioperative death is no more than that for a cholecystectomy, hip replacement, and hysterectomy, approximately 0.2%.

TABLE 27.2 Potential Complications of Bariatric Surgery

Early Complications (Within 30 Days of Surgery)

Hemorrhage
Infections
Anesthetic complications
Wound complications
Deep venous thromboembolism
Pulmonary embolism
Cardiorespiratory events
Technical failure
Anastomotic or staple line leaks (excludes adjustable gastric banding)
Perioperative mortality

Late Complications

Gastric esophageal reflux
Anastomotic ulcers
Fistulas and strictures
Reoperation

Life-Threatening Complications

Internal hernia
Bowel obstruction
Intussusception

Long-Term Metabolic Complications

Malnutrition
  Roux-en-Y gastric bypass—Iron, vitamin B_{12}, calcium, vitamin D, secondary hyperparathyroidism, folate
  Biliopancreatic diversion—Protein calorie malnutrition and fat-soluble vitamin deficiency
Nausea and vomiting (may lead to thiamine deficiency)
Inadequate weight loss/weight regain
Dumping syndrome
Hypoglycemia
Hernia
Gallstone formation (secondary to rapid weight loss)
Small intestinal bacterial overgrowth

Complications Specific to Gastric Band

Esophageal dilation and dysmotility
Gastric outlet obstruction
Band slippage
Band erosion
Subcutaneous port complications—pain, leak

Early Complications

Early surgical complications are defined as complications occurring within 30 days of operation. Fortunately, the most serious complications are the least common. These include gastrojejunal anastomosis leakage (<1%), pulmonary embolus (0.5%), bleeding (2%), bowel obstruction (2%), internal hernia (when small bowel anastomoses are performed) (2%), cardiorespiratory events, and wound complications. The incidence of wound complications such as infection and abdominal wall hernia is generally high in the obese but has decreased by 10-fold or more with the use of laparoscopy. One long-term multicenter consortium assessed a composite endpoint of death, thrombotic event, reintervention, or prolonged hospital stay (more than 30 days) and noted an incidence of 4.1% after bariatric surgery.[200]

LATE AND CHRONIC COMPLICATIONS

Complications occurring after 30 days postoperatively that are strictly related to surgery include gastrojejunostomy anastomotic stricture and marginal ulceration (5%–10%), small bowel obstruction (2%), and internal hernia (2%). As discussed above, GERD can develop in up to 50% of patients after vertical SG.[192] Long-term complications that can present a diagnostic and management challenge to the surgeon and nonsurgeon alike are discussed in detail below.

WEIGHT REGAIN

The most common late complication is failure to lose adequate weight or weight regain, which can occur in up to 20% of patients. Most often this is the result of lack of compliance with postoperative diet and exercise recommendations and patient attrition to follow-up, particularly with their bariatric program.[201] The initial rate of postoperative weight loss can be somewhat variable and is not generally an accurate predictor of ultimate weight loss outcome. Although this variability can be, in part, related to the operation itself, it may be more dependent on beginning weight/BMI and other patient factors, including behavioral compliance with respect to diet and exercise. However, the majority of weight loss typically occurs in the first 18 to 24 months.

Successful weight loss after bariatric surgery is generally defined as at least 50% EWL or a TBWL of at least 20%.[202] Patients with BMI above 50 kg/m^2 are at greater risk for weight regain.[203,204] Potential causes for weight regain can be patient-related, anatomy-related, or both. Patient-related etiologies may include poor dietary compliance, physical inactivity, inadequate follow-up, and/or mental health disorders.[205-208] Sometimes a new medication is the cause; it is well-established that antidepressants (particularly selective serotonin reuptake inhibitors and tricyclic antidepressants), mood stabilizers (particularly atypical anti-psychotic agents), and certain anti-hyperglycemic agents (particularly insulin, sulfonylureas, and thiazolidinediones) diabetic medications can all cause weight gain. Adjuvant use of weight loss medications can arrest weight regain and augment weight loss in patients after bariatric surgery.[209,210]

Anatomic causes of weight regain are sometimes present and need to be evaluated. After RYGB these can include dilation of the gastrojejunal stoma, which causes decreased satiety. Although very unlikely in the case of the contemporary RYGB due to modification of technique, a RYGB performed in the distant past may have a dilated gastric pouch or gastrogastric fistula as a cause for weight regain. Anatomic abnormalities that can cause weight regain after SG include dilated gastric sleeve and retained fundus. Many of these anatomic conditions can be endoscopically and surgically corrected.[210,211]

DUMPING SYNDROME

Dumping is a result of the anatomical configuration specifically associated with the RYGB and is a consequence of neurohormonal changes that occur, with rapid emptying of carbohydrates into the jejunal Roux limb. Whether this should be considered a true complication is debatable, because it is not dangerous, and those who develop episodes of dumping are typically straying from the prescribed postoperative diet for achieving the desired outcome of weight loss. Dumping has both early and late components. Early dumping is characterized by symptoms of bloating, abdominal discomfort, diarrhea, tachycardia, and lightheadedness and results from the delivery of hyperosmolar food into the small intestine causing a shift of fluid intraluminally from the intravascular compartment. Late dumping occurs within 2 to 3 hours of eating and is characterized by symptoms of diaphoresis, dizziness, and weakness that result from the rapid small intestine absorption of the ingested high glucose load, resulting in early hyperglycemia and consequent hyperinsulinism leading to late post-prandial or "reactive" hypoglycemia. Dumping after RYGB is considered an anticipated check on eating behavior and maintenance of dietary compliance. The threat of this unwanted side effect can be useful to encourage patients to avoid liquid carbohydrates or other foods with high sugar content. Those who experience episodes of dumping should undergo further dietary counseling. Generally, dumping can be avoided completely with dietary modification. However, there is growing evidence that at least some forms of dumping can be treated by endoscopic narrowing of the gastrojejunostomy stoma to reduce the rate of emptying of the gastric pouch, with improvement of symptoms.[212,213]

Alimentary Hypoglycemia with Neuroglycopenia

Alimentary hypoglycemia, also called post–gastric bypass hypoglycemia, is a rare condition that has some similarities with late dumping syndrome in terms of both symptoms and mechanism. Its prevalence is largely unknown, but it appears to be well below 1%,[214,215] although some studies have suggested a much higher incidence.[216,217] It can potentially occur with any upper gastrointestinal tract surgery that accelerates nutrient delivery to the small intestine, especially gastrectomy and RYGB with anatomical shortening of the upper gastrointestinal tract.[218,219] It has been described after operations that accelerate gastric emptying, such as fundoplication and SG, but is much less common after these operations.[220]

The general mechanism is thought to involve the release, in response to a high carbohydrate meal, of a high concentration of insulin due to the previous low insulin sensitivity of the obese state prior to surgery. However, after substantial weight loss this degree of insulin production is inappropriate, and hypoglycemia, sometimes severe with neuroglycopenia, can result. Peak insulin levels can be as high as 500 μU/l.[214] Mixed meal testing has demonstrated that, after a high-carbohydrate mixed meal, hyperglycemia associated with hyperinsulinemia is noted around 30 minutes. Hypoglycemia then develops as the serum insulin rapidly declines. A nadir is reached at approximately 120 minutes, and the plasma glucose level spontaneously corrects. After a low-carbohydrate high-protein mixed meal there is very little change in plasma glucose and only a modest increase in serum insulin.[214] Rapid nutrient delivery and absorption result in exaggerated early postprandial levels of glucose, GLP-1, and consequently insulin, with hypoglycemia typically occurring 1.5 to 3 hours after meals. Rarely, patients develop severe hypoglycemia with neuroglycopenia, manifested by confusion, loss of consciousness, seizures, and/or motor vehicle accidents.[221,222] When compared with postsurgical controls without hypoglycemia, individuals with alimentary hypoglycemia exhibit dysregulated islet function with increased insulin secretion and reduced insulin clearance, likely contributing to the development of late postprandial hypoglycemia.[223] The enhanced β cell sensitivity to glucose may be dependent, at least in part, on the exaggerated GLP-1 response, because blockage of the GLP-1 receptor prevents postprandial hypoglycemia in affected individuals that is associated with greater reduction in postprandial insulin secretion than that observed in postsurgical controls without alimentary hypoglycemia.[224] Counterregulatory responses to defend against the development of hypoglycemia are also impaired following RYGB, including marked reductions in glucagon and epinephrine secretion in response to insulin-induced hypoglycemia.[225] Because exposure to even mild hypoglycemia, as may occur frequently after meals, can blunt counterregulatory responses to hypoglycemia, it is possible that induction of hypoglycemia-associated autonomic failure may also contribute to the development of neuroglycopenic hypoglycemia long-term following bariatric surgery.

Initial management involves dietary modification, emphasizing controlled portions of low–glycemic index complex carbohydrates ingested together with protein and fat in order to slow the rate of glucose absorption into the circulation, as well as attention to correcting macro- and micronutrient deficiencies. Treatment of individual hypoglycemic episodes should include protein in addition to carbohydrate in order to prevent cyclic hypoglycemia. Patients should be advised to ingest low-carbohydrate, protein-rich foods, with sources of carbohydrate complex rather than simple. To prevent hypoglycemic episodes, alpha-glucosidase inhibition (i.e., acarbose) to slow carbohydrate absorption and minimize postprandial glycemia and thus reduce the stimulus for insulin secretion or use of somatostatin analogs or diazoxide to suppress insulin secretion may be effective.[226] Somatostatin has been used with varying degrees of success. Glucagon has recently emerged as potential therapy for treatment of postprandial hypoglycemia.[227] New liquid formulations may soon enable automated rescue dosing of pump-delivered glucagon informed by continuous glucose monitoring, which has been shown to reduce severe hypoglycemia in individuals.[228,229] Altering the route of nutrient delivery through gastrotomy tube placement[230] for short-term management and deconstruction of RYGB for long-term treatment have been tried, with general success. Outcomes after partial pancreatectomy are variable and appear to be no more effective than dietary modification. Moreover, pancreatectomy has high morbidity and thus is no longer considered an acceptable form of treatment.[221] As discussed above, reducing the rate of gastric emptying by decreasing the size of the gastrojejunostomy stoma via endoscopic transoral outlet reduction has been shown to be effective for treating dumping symptoms and is a promising strategy for alimentary limb hypoglycemia when the stoma is greater than 2 cm in diameter.[212,213]

KEY POINTS

- Alimentary hypoglycemia can occur after upper gastrointestinal tract rearrangement but is most commonly seen after Roux-en-Y gastric bypass. Synonyms include hyperinsulinemic hypoglycemia, and post–gastric bypass hypoglycemia.
- Symptoms can be similar to late dumping and include diaphoresis, tachycardia, mental fogginess, confusion, and, if severe, loss of consciousness and seizure. The most severe symptoms occur when there is associated neuroglycopenia.
- The mechanisms appear to involve exaggerated early postprandial levels of glucose, glucagon-like peptide 1, and consequently insulin, with hypoglycemia typically occurring 1.5–3 hours after rapid delivery and absorption of a high carbohydrate meal.
- The first line of treatment involves diet modification to a high-protein, low-carbohydrate diet. Medications such as acarbose, diazoxide, and somatostatin can be added as needed. Recently, glucagon is being considered as a potential therapy. Surgical options include reversal of anatomic changes (e.g., gastric bypass reversal) and endoscopic narrowing of the gastrojejunal stoma. Pancreatectomy is no longer considered a valid treatment.

Nutrient Deficiencies. Nutrient and vitamin deficiencies can occur following surgical manipulation of the gastrointestinal tract. After bariatric surgery, such deficiencies are seen in operations with a malabsorptive component such as BPD-DS and RYGB. Nutritional deficiencies can also occur to a lesser extent after restrictive operations due to poor food choices, decreased overall intake, or rarely from hyperemesis. Persistent hyperemesis with prolonged poor oral intake should prompt suspicion for low vitamin B1 (thiamine) levels, and supplementation should be given immediately. Gastrectomy can lead to iron, zinc, and vitamin B-12 deficiencies.[231] Preoperative screening and education, typically by a registered dietician, is vital in the assessment of candidates. All patients should be introduced to the Recommended Dietary Allowances as prescribed by the United States Department of Agriculture, should be made aware of the metabolic effects of their operation and its effects on their physiology, and should be educated that they will require lifelong vitamin and mineral supplementation to avoid deficiencies. A daily multivitamin and vitamin B12 supplementation for all bariatric patients, regardless of operation, is safe practice. Supplementation with iron, copper, and calcium is recommended after RYGB. Those with true malabsorptive anatomy (e.g., BPD-DS) may also require routine fat-soluble vitamin supplementation. Although uncommon, deficiencies can occasionally develop despite patient compliance, due to poor diet or extreme malabsorption.

Protein Deficiency. Protein status is often assessed by serum albumin, with a level of less than 3.5 g/dL indicating deficiency. Overall incidence has been estimated to range from 5% to 9% after RYGB and 10% to 15% after BPD-DS at 2 years' postoperative follow-up[232] and can present with muscle weakness/asthenia, alopecia, edema, anemia, or hypoalbuminemia of varying degrees. Management depends upon the severity of the deficiency and the available routes for replenishing protein levels. Over-the-counter low-carbohydrate, high-protein shakes can be useful in those who can tolerate oral intake. When the *per os* route is not available, enteral feeding via a percutaneously, endoscopically, or surgically placed feeding tube may be needed and is preferred over the parenteral route. After gastric bypass enteral feeding is typically accomplished via a remnant gastrostomy tube. Dietary counseling to encourage increased patient intake can correct less profound deficiencies when the oral route is available.

Iron Deficiency. Iron deficiency can be present in the obese population both before and after bariatric surgery. In a systematic review, Enani et al. reported an incidence of iron deficiency of 15.2% before bariatric surgery and 16.6% after bariatric surgery (36.6% before SG vs. 12.4% after SG and 12.9% before RYGB vs. 24.5% after RYGB).[213] The occurrence of iron deficiency anemia was 1.6% after SG and 16.7% after RYGB.[213] Risk factors for postoperative iron deficiency were preoperative iron deficiency, duration of follow-up, and premenopausal status. However, in a bariatric patient presenting with anemia, other common causes must be ruled out before attributing this deficiency solely to the bariatric operation (e.g., menorrhagia, marginal ulcers, or cancer). Ferrous rather than ferric preparations are favored because of relative achlorhydria after RYGB. Supplementation with oral ferrous sulfate, ferrous fumarate, or ferrous gluconate to provide up to 150 to 200 mg of elemental iron daily is recommended, and scheduled follow-ups and labs are important for prevention and management.

Vitamin Deficiency. *Vitamin B-12.* Vitamin B-12 is absorbed in the distal ileum. Intrinsic factor (IF), a glycoprotein made in the parietal cells of the stomach fundus, is required for absorption in the ileum, and gastric acidity greatly facilitates the binding of IF to vitamin B-12 in the stomach. In SG and BPD-DS, gastric IF and acid production is reduced; in RYGB, it is bypassed and defunctionalized.[233] Brolin et al. reported the incidence of vitamin B-12 deficiency at 8% to 37%.[234] Vitamin B-12 deficiency can lead to megaloblastic anemia. However, in folate-supplemented patients, neurological manifestations of vitamin B-12 deficiency without anemia predominate. Compliance with postoperative bariatric supplementation is sufficient to prevent and treat vitamin B-12 deficiency.

Vitamin B9/folate. Folic acid/folate are absorbed in the jejunum. Incidence of deficiency mirrors that of vitamin B-12. However, symptomatic folate deficiency is rare and can result in neural tube defects in a mother's fetus, macrocytic anemia, peripheral neuropathy, cognitive deficiencies, mouth ulcers, and sore/swollen tongue. Multivitamins usually contain the recommended dose of folate.

Vitamin B1/thiamine. Like folate, thiamine is a water-soluble vitamin absorbed in the jejunum. Incidence of thiamine deficiency is low and is seen in patients who are noncompliant with supplementation, those suffering from frequent vomiting, and those with significant malnutrition. The diagnosis of thiamine deficiency is based on fasting blood thiamine (deficiency defined as less than 78 nM). Thiamine deficiency can lead to Wernicke–Korsakoff encephalopathy, and thus a low index of suspicion for diagnosis and treatment is essential. Primavera et al.[235] showed a rate of Wernicke–Korsakoff encephalopathy of 0.18% in 1663 patients. Such patients respond well to intravenous/intramuscular thiamine.

Vitamin D deficiency, calcium deficiency, and bone health. Deficiencies of vitamin D and calcium occur together in bariatric patients. Vitamin D deficiency is the most common deficiency in pre–bariatric surgery obese individuals, and it is important to screen and treat deficient patients preoperatively.[212] Postoperatively, bariatric patients are at risk for metabolic bone disease. After RYGB, early catabolic bone remodeling, as evidenced by increased circulating levels of markers of bone turnover, have been found that persist, and subsequent decrease in bone mineral density can occur.[226,236-238] The mechanism is likely multifactorial. First, bone remodeling after bariatric surgery may represent a physiologic change in skeletal structure in response to skeletal unloading after rapid weight loss. Second, calcium and vitamin D deficiencies with or without secondary hyperparathyroidism can result in pathologic mechanisms that can lead to bone catabolism.[239] Finally, procedures that involve intestinal bypass can lead to deficiencies. Calcium is absorbed in the duodenum and proximal jejunum. Vitamin D, a fat-soluble vitamin, is absorbed in the ileum. A randomized study[128] comparing SG with RYGB versus intensive medical therapy in diabetic obese patients evaluated bone turnover by measuring bone formation marker osteocalcin, bone resorption marker serum C-telopeptide of type 1 collagen, and intact parathyroid hormone at baseline and at 5 years after surgery. Bone turnover was found to be highest for RYGB and was also elevated after SG, but to a lesser extent, and both RYGB and SG were significantly elevated compared with intensive medical management.[240]

Screening should include not only levels of calcium and 25-hydroxyvitamin D, but also parathyroid hormone and markers for bone turnover such as bone-specific alkaline phosphatase, serum C-telopeptide, urinary N-telopeptide and/or osteocalcin.[231] Higher-risk patients, stratified by age and gender, may require a baseline bone density scan and be marked for more stringent postoperative follow-up. Peterson et al. reported a vitamin D deficiency prevalence of 13% to 90% preoperatively.[241] The same study reported a similar prevalence post–bariatric surgery, but it was higher in RYGB patients. In fact, deficiency can occur despite patient compliance.

Vitamins A, E, and K. Deficiencies in these fat-soluble vitamins are mostly seen in patients after a primarily malabsorptive operation, such as the BPD-DS. Reported incidence is low, and clinical presentations are rare and limited to a few case reports typically seen in late follow-up (after 4 years). Vitamin A deficiency, which can cause night blindness and xerophthalmia, has a reported incidence of 5% to 12%. Vitamin E deficiency was reported in 5% in one small study, though no clinically significant complications were reported. Vitamin K deficiency was reported as prevalent post-BPD in one study of 170 patients (68% at 4 years).[231] It has not been shown to be accompanied by any increase in incidence of abnormal bleeding.

Zinc. Zinc deficiency has been reported at rates of 10.8% to as high as 50% in patients following the malabsorptive surgeries BPD and distal RYGB.[231] Blood zinc should be checked in long-term follow-up. Patients with zinc deficiency can present with persistent alopecia.

Selenium. Selenium deficiency occurred at rates of 14.5% in BPD patients in one study, with significant clinical complications.[231] Selenium is needed to convert T4 to the active form T3, and thus hypothyroidism can occur if it is absent.

Magnesium. Magnesium levels should be checked in postoperative patients, because a deficiency can have cardiovascular complications and cause hypocalcemia by impairing parathyroid hormone release. The reported incidence in BPD-DS patients was 5%, with no significant clinical complications.[231]

Copper. Hematologic abnormalities and neuropathies including a form of myelopathy have been described from copper deficiency after RYGB.[242] Although symptomatic deficiency is rare, there is a 10% incidence overall of copper deficiency after RYGB. Copper deficiency myelopathy is a condition indistinguishable from subacute combined degeneration of the spinal cord. Early diagnosis and treatment are crucial to avoid permanent neurologic sequelae. Prophylactic supplementation with at least 2 grams of copper daily is recommended.[243]

Cognitive and Psychological Issues. The National Institutes of Health recognized the importance of a patient's mental health by mandating preoperative psychological evaluation and clearance for all prospective bariatric patients. No patients with active untreated psychosis should be approved for surgery. Drug and alcohol addiction and eating disorders are all contraindications to surgery. Obese patients have a higher incidence of major depressive disorders, anxiety disorders, eating and binging disorders, and emotional eating than nonobese patients. Several health-related quality of life (HRQOL) surveys show that the majority of patients improve in their overall mental and physical health after bariatric surgery. In a study of bariatric patients matched with obese patients who were denied bariatric surgery, RYGB patients scored the highest on a HRQOL survey as far out as 6 years.[244] HRQOL scores correspond directly to weight loss, weight stability, and weight regain. A small subset of patients experience declining mental health postoperatively. Postoperative bariatric patients account for 6% of all inpatient treatment admissions for substance abuse.[245] The etiology for increased substance abuse has been attributed to increased peak blood levels with faster onset and stronger effect, addiction substitution from food to alcohol, new social status and attention received from being thin, and unresolved psychological problems. In fact, one third of patients experienced new-onset alcohol use or abuse after bariatric surgery.[246] In addition, the effect of alcohol consumption after RYGB is pronounced, in that peak alcohol levels occur quickly and disappear rapidly after consumption. This pattern may be a risk factor for alcohol addiction in these patients.[247] Thus, postoperative psychological counseling is indicated, but the cost can be prohibitive. The multidisciplinary team must identify at-risk patients during preoperative assessment.

KEY POINTS

- Nutritional deficiencies can occur after any bariatric operation or endoscopic procedure.
- The most profound nutrient deficiencies are associated with bariatric operations that provide protein-calorie malnutrition as the main mechanism of weight loss.
- Long-term nutritional supplementation and postoperative monitoring are necessary to prevent profound nutritional deficiencies.
- Nutritional deficiencies should be strongly considered in a patient with profound weight loss and adult "failure-to-thrive" even several years after a bariatric operation or endoscopic procedure.

SUMMARY

Bariatric surgery is a safe and effective treatment for obesity, with a proven track record, and it is the most effective obesity treatment available. More recently, the development of bariatric endoscopic therapies has provided an additional tool that serves as a useful adjunct. Both the bariatric operations and endoscopic procedures involve mechanisms that are restrictive, malabsorptive, and a combination thereof, with subsequent neurohormonal changes that provide both satiety and improvements in metabolic health. In the future, these operations and procedures, as well as their applications, will continue to evolve and be refined as we learn more about the physiology of obesity and its pathophysiologic consequences.

- For your free Expert Consult eBook with bibliographic citations as well as the ability to take notes, highlight important content, search the full text, and more, visit: http://www.ExpertConsult.Inkling.com.

REFERENCES

1. Soper NJ, Brunt LM, Kerbl K. Laparoscopic general surgery. *N Engl J Med*. 1994;330:409–419.
2. Ricciardi R, Town RJ, Kellogg TA, et al. Outcomes after open versus laparoscopic gastric bypass. *Surg Laparosc Endosc Percutan Tech*. 2006;16:317–320.
3. Genco A, Soricelli E, Casella G, et al. Gastroesophageal reflux disease and Barrett's esophagus after laparoscopic sleeve gastrectomy: a possible, underestimated long-term complication. *Surg Obes Relat Dis*. 2017;13:568–574.
4. Benaiges D, Mas-Lorenzo A, Goday A, et al. Laparoscopic sleeve gastrectomy: more than a restrictive bariatric surgery procedure? *World J Gastroenterol*. 2015;21:11804–11814.
5. Ariyasu H, Takaya K, Tagami T, et al. Stomach is a major source of circulating ghrelin, and feeding state determines plasma ghrelin-like immunoreactivity levels in humans. *J Clin Endocrinol Metab*. 2001;86:4753–4758.
6. Date Y, Kojima M, Hosoda H, et al. Ghrelin, a novel growth hormone-releasing acylated peptide, is synthesized in a distinct endocrine cell type in the gastrointestinal tracts of rats and humans. *Endocrinology*. 2000;141:4255–4261.
7. Arakawa R, Febres G, Cheng B, et al. Prospective study of gut hormone and metabolic changes after laparoscopic sleeve gastrectomy and Roux-en-Y gastric bypass. *PLoS One*. 2020;15:e0236133.
8. Karamanakos SN, Vagenas K, Kalfarentzos F, et al. Weight loss, appetite suppression, and changes in fasting and postprandial ghrelin and peptide-YY levels after Roux-en-Y gastric bypass and sleeve gastrectomy: a prospective, double blind study. *Ann Surg*. 2008;247:401–407.
9. Brethauer SA, Hammel JP, Schauer PR. Systematic review of sleeve gastrectomy as staging and primary bariatric procedure. *Surg Obes Relat Dis*. 2009;5:469–475.
10. Aminian A, Brethauer SA, Sharafkhah M, et al. Development of a sleeve gastrectomy risk calculator. *Surg Obes Relat Dis*. 2015;11:758–764.
11. Nergard BJ, Lindqvist A, Gislason HG, et al. Mucosal glucagon-like peptide-1 and glucose-dependent insulinotropic polypeptide cell numbers in the super-obese human foregut after gastric bypass. *Surg Obes Relat Dis*. 2015;11:1237–1246.
12. Nielsen AW, Helm MC, Kindel T, et al. Perioperative bleeding and blood transfusion are major risk factors for venous thromboembolism following bariatric surgery. *Surg Endosc*. 2018;32:2488–2495.
13. Adams TD, Pendleton RC, Strong MB, et al. Health outcomes of gastric bypass patients compared to nonsurgical, nonintervened severely obese. *Obesity*. 2010;18:121–130.
14. Merz AE, Blackstone RB, Gagner M, et al. Duodenal switch in revisional bariatric surgery: conclusions from an expert consensus panel. *Surg Obes Relat Dis*. 2019;15:894–899.
15. Weiner RA, Blanco-Engert R, Weiner S, et al. Laparoscopic biliopancreatic diversion with duodenal switch: three different duodeno-ileal anastomotic techniques and initial experience. *Obes Surg*. 2004;14:334–340.
16. Keshishian A, Zahriya K, Hartoonian T, et al. Duodenal switch is a safe operation for patients who have failed other bariatric operations. *Obes Surg*. 2004;14:1187–1192.
17. Hedberg J, Hedenstrom H, Karlsson FA, et al. Gastric emptying and postprandial PYY response after biliopancreatic diversion with duodenal switch. *Obes Surg*. 2011;21:609–615.
18. Buchwald H, Ikramuddin S, Dorman RB, et al. Management of the metabolic/bariatric surgery patient. *Am J Med*. 2011;124:1099–1105.
19. Buchwald H, Oien DM. Metabolic/bariatric surgery worldwide 2011. *Obes Surg*. 2013;23:427–436.
20. Marceau P, Biron S, Hould FS, et al. Duodenal switch: long-term results. *Obes Surg*. 2007;17:1421–1430.
21. Rezvani M, Sucandy I, Klar A, et al. Is laparoscopic single-stage biliopancreatic diversion with duodenal switch safe in super morbidly obese patients? *Surg Obes Relat Dis*. 2014;10:427–430.
22. Sanchez-Pernaute A, Herrera MA, Perez-Aguirre ME, et al. Single anastomosis duodeno-ileal bypass with sleeve gastrectomy (SADI-S). One to three-year follow-up. *Obes Surg*. 2010;20:1720–1726.
23. Zaveri H, Surve A, Cottam D, et al. Mid-term 4-year outcomes with single anastomosis duodenal-ileal bypass with sleeve gastrectomy surgery at a single US center. *Obes Surg*. 2018;28:3062–3072.
24. Surve A, Zaveri H, Cottam D, et al. A retrospective comparison of biliopancreatic diversion with duodenal switch with single anastomosis duodenal switch (SIPS-stomach intestinal pylorus sparing surgery) at a single institution with two year follow-up. *Surg Obes Relat Dis*. 2017;13:415–422.
25. Cottam S, Ng P, Sharp L, et al. Single-anastomosis duodenal ileostomy with sleeve is a safe and effective option for patients in an ambulatory surgical center. *Surg Obes Relat Dis*. 2019;15:1990–1993.
26. Gudzune KA, Doshi RS, Mehta AK, et al. Efficacy of commercial weight-loss programs: an updated systematic review. *Ann Intern Med*. 2015;162:501–512.
27. Dansinger ML, Gleason JA, Griffith JL, et al. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. *JAMA*. 2005;293:43–53.
28. Pilitsi E, Farr OM, Polyzos SA, et al. Pharmacotherapy of obesity: available medications and drugs under investigation. *Metabolism*. 2019;92:170–192.
29. Lee M, Lauren BN, Zhan T, et al. The cost-effectiveness of pharmacotherapy and lifestyle intervention in the treatment of obesity. *Obes Sci Pract*. 2020;6:162–170.
30. Chang SH, Stoll CR, Song J, et al. The effectiveness and risks of bariatric surgery: an updated systematic review and meta-analysis, 2003-2012. *JAMA Surg*. 2014;149:275–287.
31. Wolfe BM, Kvach E, Eckel RH. Treatment of obesity: weight loss and bariatric surgery. *Circ Res*. 2016;118:1844–1855.
32. English WJ, DeMaria EJ, Brethauer SA, et al. American Society for Metabolic and Bariatric Surgery estimation of metabolic and bariatric procedures performed in the United States in 2016. *Surg Obes Relat Dis*. 2018;14:259–263.
33. Santry HP, Gillen DL, Lauderdale DS. Trends in bariatric surgical procedures. *JAMA*. 2005;294:1909–1917.
34. Clapp B, Harper B, Dodoo C, et al. Trends in revisional bariatric surgery using the MBSAQIP database 2015-2017. *Surg Obes Relat Dis*. 2020;16:908–915.
35. Fulton C, Sheppard C, Birch D, et al. A comparison of revisional and primary bariatric surgery. *Can J Surg*. 2017;60:205–211.
36. Abu Dayyeh BK, Edmundowicz S, Thompson CC. Clinical practice update: expert review on endoscopic bariatric therapies. *Gastroenterol*. 2017;152:716–729.
37. Acosta A, Streett S, Kroh MD, et al. White Paper AGA: POWER - practice Guide on obesity and weight management, education, and Resources. *Clin Gastroenterol Hepatol*. 2017;15:631–649 e10.
38. Zorron R, Veltzke-Schlieker W, Adler A, et al. Endoscopic sleeve gastroplasty using Apollo Overstitch as a bridging procedure for superobese and high risk patients. *Endoscopy*. 2018;50:81–83.
39. Ball W, Raza SS, Loy J, et al. Effectiveness of intra-gastric balloon as a bridge to definitive surgery in the super obese. *Obes Surg*. 2019;29:1932–1936.

40. Li R, Veltzke-Schlieker W, Adler A, et al. Endoscopic sleeve gastroplasty (ESG) for high-risk patients, high body mass index (> 50 kg/m(2)) patients, and contraindication to abdominal surgery. *Obes Surg*. 2021.
41. Choudhary NS, Puri R, Saraf N, et al. Intragastric balloon as a novel modality for weight loss in patients with cirrhosis and morbid obesity awaiting liver transplantation. *Indian J Gastroenterol*. 2016;35:113–116.
42. Watt KD, Heimbach JK, Rizk M, et al. Efficacy and safety of endoscopic balloon placement for weight loss in cirrhotic patients awaiting liver transplantation. *Liver Transpl*. 2021;27:1239–1247.
43. Patel NJ, Gomez V, Steidley DE, et al. Successful use of intragastric balloon therapy as a bridge to heart transplantation. *Obes Surg*. 2020;30:3610–3614.
44. Musella M, Milone M, Bellini M, et al. The potential role of intragastric balloon in the treatment of obese-related infertility: personal experience. *Obes Surg*. 2011;21:426–430.
45. Bazerbachi F, Vargas EJ, Rizk M, et al. Intragastric balloon placement induces significant metabolic and histologic improvement in patients with nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol*. 2021;19:146–154 e4.
46. Salomone F, Currenti W, Magri G, et al. Effects of intragastric balloon in patients with nonalcoholic fatty liver disease and advanced fibrosis. *Liver Int*. 2021;41(9):2112–2116.
47. Mingrone G, van Baar AC, Deviere J, et al. Safety and efficacy of hydrothermal duodenal mucosal resurfacing in patients with type 2 diabetes: the randomised, double-blind, sham-controlled, multicentre REVITA-2 feasibility trial. *Gut*. 2022;71:254–264.
48. de Oliveira GHP, de Moura DTH, Funari MP, et al. Metabolic effects of endoscopic duodenal mucosal resurfacing: a systematic review and meta-analysis. *Obes Surg*. 2021;31:1304–1312.
49. Acosta A, Abu Dayyeh BK, Port JD, et al. Recent advances in clinical practice challenges and opportunities in the management of obesity. *Gut*. 2014;63:687–695.
50. van Baar ACG, Nieuwdorp M, Holleman F, et al. The duodenum harbors a broad untapped therapeutic potential. *Gastroenterology*. 2018;154:773–777.
51. Xie C, Huang W, Young RL, et al. Role of bile acids in the regulation of food intake, and their dysregulation in metabolic disease. *Nutrients*. 2021;13.
52. Courcoulas A, Abu Dayyeh BK, Eaton L, et al. Intragastric balloon as an adjunct to lifestyle intervention: a randomized controlled trial. *Int J Obes*. 2017;41:427–433.
53. Force ABET, Committee AT, Abu Dayyeh BK, et al. ASGE Bariatric Endoscopy Task Force systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting endoscopic bariatric therapies. *Gastrointest Endosc*. 2015;82:425–438.e5.
54. Shah R, Davitkov P, Abu Dayyeh BK, et al. AGA technical review on intragastric balloons in the management of obesity. *Gastroenterology*. 2021;160:1811–1830.
55. Sullivan S, Swain J, Woodman G, et al. Randomized sham-controlled trial of the 6-month swallowable gas-filled intragastric balloon system for weight loss. *Surg Obes Relat Dis*. 2018;14:1876–1889.
56. Brooks J, Srivastava ED, Mathus-Vliegen EM. One-year adjustable intragastric balloons: results in 73 consecutive patients in the U.K. *Obes Surg*. 2014;24:813–819.
57. AbuDayyeh BK, Noar MD, Lavin T, et al. 176 Pivotal randomized-controlled trial of the adjustable (SPATZ-3) intragastric balloon system for weight loss. *Gastrointest Endosc*. 2019;89:AB58–AB59.
58. Vantanasiri K, Matar R, Beran A, et al. The efficacy and safety of a procedureless gastric balloon for weight loss: a systematic review and meta-analysis. *Obes Surg*. 2020;30:3341–3346.
59. Abu Dayyeh BK, Rajan E, Gostout CJ. Endoscopic sleeve gastroplasty: a potential endoscopic alternative to surgical sleeve gastrectomy for treatment of obesity. *Gastrointest Endosc*. 2013;78:530–535.
60. Hedjoudje A, Abu Dayyeh BK, Cheskin LJ, et al. Efficacy and safety of endoscopic sleeve gastroplasty: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2020;18:1043–1053.e4.
61. Sharaiha RZ, Kumta NA, Saumoy M, et al. Endoscopic sleeve gastroplasty significantly reduces body mass index and metabolic complications in obese patients. *Clin Gastroenterol Hepatol*. 2017;15:504–510.
62. Alqahtani A, Al-Darwish A, Mahmoud AE, et al. Short-term outcomes of endoscopic sleeve gastroplasty in 1000 consecutive patients. *Gastrointest Endosc*. 2019;89:1132–1138.
63. Sharaiha RZ, Hajifathalian K, Kumar R, et al. Five-year outcomes of endoscopic sleeve gastroplasty for the treatment of obesity. *Clin Gastroenterol Hepatol*. 2021;19:1051–1057.e2.
64. Espinos JC, Turro R, Moragas G, et al. Gastrointestinal physiological changes and their relationship to weight loss following the POSE procedure. *Obes Surg*. 2016;26:1081–1089.
65. Sullivan S, Swain JM, Woodman G, et al. Randomized sham-controlled trial evaluating efficacy and safety of endoscopic gastric plication for primary obesity: the ESSENTIAL trial. *Obesity (Silver Spring)*. 2017;25:294–301.
66. Maselli DB, Mulki RH, Matar R, et al. Endoscopic gastric suturing for weight loss: techniques and outcomes. *Tech Innov Gastrointest Endosc*. 2020;22:136–144.
67. Thompson CC, Abu Dayyeh BK, Kushner R, et al. Percutaneous gastrostomy device for the treatment of class II and class III obesity: results of a randomized controlled trial. *Am J Gastroenterol*. 2017;112:447–457.
68. Duca FA, Bauer PV, Hamr SC, et al. Glucoregulatory relevance of small intestinal nutrient sensing in physiology, bariatric surgery, and pharmacology. *Cell Metab*. 2015;22:367–380.
69. West JA, Ghosh SS, Parkes DG, et al. Peptide-based GIP receptor inhibition exhibits modest metabolic changes in mice when administered either alone or combined with GLP-1 agonism. *bioRxiv*. 2019.
70. Rajagopalan H, Cherrington AD, Thompson CC, et al. Endoscopic duodenal mucosal resurfacing for the treatment of type 2 diabetes: 6-month interim analysis from the first-in-human proof-of-concept study. *Diabetes Care*. 2016;39:2254–2261.
71. Machytka E, Buzga M, Zonca P, et al. Partial jejunal diversion using an incisionless magnetic anastomosis system: 1-year interim results in patients with obesity and diabetes. *Gastrointest Endosc*. 2017;86:904–912.
72. Badurdeen D, Hoff AC, Hedjoudje A, et al. Endoscopic sleeve gastroplasty plus liraglutide versus endoscopic sleeve gastroplasty alone for weight loss. *Gastrointest Endosc*. 2021;93:1316–1324.e1.

72a. Consensus Development Conference Panel. NIH conference. Gastrointestinal surgery for severe obesity. *Ann Intern Med*. 1991;115:956–961. http://www.nhlbi.nih.gov/guidelines/ obesity/prctgd_b.pdf accessed 9/1/2013.

73. Aminian A, Chang J, Brethauer SA, et al. ASMBS updated position statement on bariatric surgery in class I obesity (BMI 30-35 kg/m(2)). *Surg Obes Relat Dis*. 2018;14:1071–1087.
74. Sczepaniak JP, Owens ML, Shukla H, et al. Comparability of weight loss reporting after gastric bypass and sleeve gastrectomy using BOLD data 2008-2011. *Obes Surg*. 2015;25:788–795.
75. Salminen P, Helmio M, Ovaska J, et al. Effect of laparoscopic sleeve gastrectomy vs laparoscopic Roux-en-Y gastric bypass on weight loss at 5 years among patients with morbid obesity: the SLEEVEPASS randomized clinical trial. *JAMA*. 2018;319:241–254.
76. Peterli R, Wolnerhanssen BK, Peters T, et al. Effect of laparoscopic sleeve gastrectomy vs laparoscopic Roux-en-Y gastric bypass on weight loss in patients with morbid obesity: the SM-BOSS randomized clinical trial. *JAMA*. 2018;319:255–265.
77. Eckel RH. Clinical practice. Nonsurgical management of obesity in adults. *N Engl J Med*. 2008;358:1941–1950.
78. Sumithran P, Prendergast LA, Delbridge E, et al. Long-term persistence of hormonal adaptations to weight loss. *N Engl J Med*. 2011;365:1597–1604.
79. Rosenbaum M, Hirsch J, Gallagher DA, et al. Long-term persistence of adaptive thermogenesis in subjects who have maintained a reduced body weight. *Am J Clin Nutr*. 2008;88:906–912.
80. Bettini S, Belligoli A, Fabris R, et al. Diet approach before and after bariatric surgery. *Rev Endocr Metab Disord*. 2020;21:297–306.
81. Sun Y, Liu B, Smith JK, et al. Association of preoperative body weight and weight loss with risk of death after bariatric surgery. *JAMA Netw Open*. 2020;3:e204803.
82. Chambers AP, Jessen L, Ryan KK, et al. Weight-independent changes in blood glucose homeostasis after gastric bypass or vertical sleeve gastrectomy in rats. *Gastroenterology*. 2011;141:950–958.
83. Butner KL, Nickols-Richardson SM, Clark SF, et al. A review of weight loss following Roux-en-Y gastric bypass vs restrictive bariatric surgery: impact on adiponectin and insulin. *Obes Surg*. 2010;20:559–568.
84. Dixon AF, Dixon JB, O'Brien PE. Laparoscopic adjustable gastric banding induces prolonged satiety: a randomized blind crossover study. *J Clin Endocrinol Metab*. 2005;90:813–819.

85. le Roux CW, Aylwin SJ, Batterham RL, et al. Gut hormone profiles following bariatric surgery favor an anorectic state, facilitate weight loss, and improve metabolic parameters. *Ann Surg*. 2006;243:108–114.
86. Hofmann W, van Koningsbruggen GM, Stroebe W, et al. As pleasure unfolds. Hedonic responses to tempting food. *Psychol Sci*. 2010;21:1863–1870.
87. Morton GJ, Cummings DE, Baskin DG, et al. Central nervous system control of food intake and body weight. *Nature*. 2006;443:289–295.
88. Lawrence CB, Snape AC, Baudoin FM, et al. Acute central ghrelin and GH secretagogues induce feeding and activate brain appetite centers. *Endocrinology*. 2002;143:155–162.
89. Barazzoni R, Zanetti M, Nagliati C, et al. Gastric bypass does not normalize obesity-related changes in ghrelin profile and leads to higher acylated ghrelin fraction. *Obesity (Silver Spring)*. 2013;21:718–722.
90. Cummings DE, Weigle DS, Frayo RS, et al. Plasma ghrelin levels after diet-induced weight loss or gastric bypass surgery. *N Engl J Med*. 2002;346:1623–1630.
91. Dimitriadis E, Daskalakis M, Kampa M, et al. Alterations in gut hormones after laparoscopic sleeve gastrectomy: a prospective clinical and laboratory investigational study. *Ann Surg*. 2013;257:647–654.
92. Batterham RL, Ffytche DH, Rosenthal JM, et al. PYY modulation of cortical and hypothalamic brain areas predicts feeding behaviour in humans. *Nature*. 2007;450:106–109.
93. Korner J, Inabnet W, Febres G, et al. Prospective study of gut hormone and metabolic changes after adjustable gastric banding and Roux-en-Y gastric bypass. *Int J Obes (Lond)*. 2009;33:786–795.
94. Laferrere B, Swerdlow N, Bawa B, et al. Rise of oxyntomodulin in response to oral glucose after gastric bypass surgery in patients with type 2 diabetes. *J Clin Endocrinol Metab*. 2010;95:4072–4076.
95. Shin AC, Zheng H, Townsend RL, et al. Meal-induced hormone responses in a rat model of Roux-en-Y gastric bypass surgery. *Endocrinology*. 2010;151:1588–1597.
96. Nannipieri M, Baldi S, Mari A, et al. Roux-en-Y gastric bypass and sleeve gastrectomy: mechanisms of diabetes remission and role of gut hormones. *J Clin Endocrinol Metab*. 2013;98:4391–4399.
97. Korner J, Inabnet W, Conwell IM, et al. Differential effects of gastric bypass and banding on circulating gut hormone and leptin levels. *Obesity (Silver Spring)*. 2006;14:1553–1561.
98. Lieverse RJ, Jansen JB, Masclee AA, et al. Satiety effects of a physiological dose of cholecystokinin in humans. *Gut*. 1995;36:176–179.
99. Beglinger S, Drewe J, Schirra J, et al. Role of fat hydrolysis in regulating glucagon-like Peptide-1 secretion. *J Clin Endocrinol Metab*. 2010;95:879–886.
100. Peterli R, Wolnerhanssen B, Peters T, et al. Improvement in glucose metabolism after bariatric surgery: comparison of laparoscopic Roux-en-Y gastric bypass and laparoscopic sleeve gastrectomy: a prospective randomized trial. *Ann Surg*. 2009;250:234–241.
101. Rubino F, Gagner M, Gentileschi P, et al. The early effect of the Roux-en-Y gastric bypass on hormones involved in body weight regulation and glucose metabolism. *Ann Surg*. 2004;240:236–242.
102. Ryan KK, Kohli R, Gutierrez-Aguilar R, et al. Fibroblast growth factor-19 action in the brain reduces food intake and body weight and improves glucose tolerance in male rats. *Endocrinology*. 2013;154:9–15.
103. Hotta K, Funahashi T, Bodkin NL, et al. Circulating concentrations of the adipocyte protein adiponectin are decreased in parallel with reduced insulin sensitivity during the progression to type 2 diabetes in rhesus monkeys. *Diabetes*. 2001;50:1126–1133.
104. Ukkola O, Santaniemi M. Adiponectin: a link between excess adiposity and associated comorbidities? *J Mol Med (Berl)*. 2002;80:696–702.
105. Vilarrasa N, Vendrell J, Sanchez-Santos R, et al. Effect of weight loss induced by gastric bypass on proinflammatory interleukin-18, soluble tumour necrosis factor-alpha receptors, C-reactive protein and adiponectin in morbidly obese patients. *Clin Endocrinol (Oxf)*. 2007;67:679–686.
106. Rosenbaum M, Nicolson M, Hirsch J, et al. Effects of weight change on plasma leptin concentrations and energy expenditure. *J Clin Endocrinol Metab*. 1997;82:3647–3654.
107. Berthoud HR, Morrison C. The brain, appetite, and obesity. *Annu Rev Psychol*. 2008;59:55–92.
108. Stefater MA, Perez-Tilve D, Chambers AP, et al. Sleeve gastrectomy induces loss of weight and fat mass in obese rats, but does not affect leptin sensitivity. *Gastroenterology*. 2010;138:2426–2436, 2436 e1–e3.
109. Scholtz S, Miras AD, Chhina N, et al. Obese patients after gastric bypass surgery have lower brain-hedonic responses to food than after gastric banding. *Gut*. 2014;63:891–902.
110. Goldstone AP, Miras AD, Scholtz S, et al. Link between increased satiety gut hormones and reduced food reward after gastric bypass surgery for obesity. *J Clin Endocrinol Metab*. 2016;101:599–609.
111. Gribble FM. The gut endocrine system as a coordinator of postprandial nutrient homoeostasis. *Proc Nutr Soc*. 2012;71:456–462.
112. Perathoner A, Weiss H, Santner W, et al. Vagal nerve dissection during pouch formation in laparoscopic Roux-Y-gastric bypass for technical simplification: does it matter? *Obes Surg*. 2009;19:412–417.
113. Shin AC, Zheng H, Berthoud HR. Vagal innervation of the hepatic portal vein and liver is not necessary for Roux-en-Y gastric bypass surgery-induced hypophagia, weight loss, and hypermetabolism. *Ann Surg*. 2012;255:294–301.
114. Sjostrom L, Lindroos AK, Peltonen M, et al. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med*. 2004;351:2683–2693.
115. Faria SL, Faria OP, Cardeal Mde A, et al. Diet-induced thermogenesis and respiratory quotient after Roux-en-Y gastric bypass. *Surg Obes Relat Dis*. 2012;8:797–802.
116. Werling M, Olbers T, Fandriks L, et al. Increased postprandial energy expenditure may explain superior long term weight loss after Roux-en-Y gastric bypass compared to vertical banded gastroplasty. *PLoS One*. 2013;8:e60280.
117. Stylopoulos N, Hoppin AG, Kaplan LM. Roux-en-Y gastric bypass enhances energy expenditure and extends lifespan in diet-induced obese rats. *Obesity (Silver Spring)*. 2009;17:1839–1847.
118. Watanabe M, Houten SM, Mataki C, et al. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature*. 2006;439:484–489.
119. Buchwald H, Estok R, Fahrbach K, et al. Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. *Am J Med*. 2009;122:248–256.e5.
120. Arterburn DE, Telem DA, Kushner RF, et al. Benefits and risks of bariatric surgery in adults: a review. *JAMA*. 2020;324:879–887.
121. Carlsson LM, Peltonen M, Ahlin S, et al. Bariatric surgery and prevention of type 2 diabetes in Swedish obese subjects. *N Engl J Med*. 2012;367:695–704.
122. Dixon JB, O'Brien PE, Playfair J, et al. Adjustable gastric banding and conventional therapy for type 2 diabetes: a randomized controlled trial. *JAMA*. 2008;299:316–323.
123. Ikramuddin S, Korner J, Lee WJ, et al. Roux-en-Y gastric bypass vs intensive medical management for the control of type 2 diabetes, hypertension, and hyperlipidemia: the Diabetes Surgery Study randomized clinical trial. *JAMA*. 2013;309:2240–2249.
124. Mingrone G, Panunzi S, De Gaetano A, et al. Bariatric surgery versus conventional medical therapy for type 2 diabetes. *N Engl J Med*. 2012;366:1577–1585.
125. Schauer PR, Kashyap SR, Wolski K, et al. Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N Engl J Med*. 2012;366:1567–1576.
126. Mingrone G, Panunzi S, De Gaetano A, et al. Bariatric-metabolic surgery versus conventional medical treatment in obese patients with type 2 diabetes: 5 year follow-up of an open-label, single-centre, randomised controlled trial. *Lancet*. 2015;386:964–973.
127. Schauer PR, Bhatt DL, Kirwan JP, et al. Bariatric surgery versus intensive medical therapy for diabetes—3-year outcomes. *N Engl J Med*. 2014;370:2002–2013.
128. Schauer PR, Bhatt DL, Kirwan JP, et al. Bariatric surgery versus intensive medical therapy for diabetes—5-year outcomes. *N Engl J Med*. 2017;376:641–651.
129. Ikramuddin S, Billington CJ, Lee WJ, et al. Roux-en-Y gastric bypass for diabetes (the Diabetes Surgery Study): 2-year outcomes of a 5-year, randomised, controlled trial. *Lancet Diabetes Endocrinol*. 2015;3:413–422.

130. Ikramuddin S, Korner J, Lee WJ, et al. Lifestyle intervention and medical management with vs without Roux-en-Y gastric bypass and control of hemoglobin A1c, LDL cholesterol, and systolic blood pressure at 5 years in the diabetes surgery study. *JAMA*. 2018;319:266–278.
131. Courcoulas AP, Gallagher JW, Neiberg RH, et al. Bariatric surgery vs lifestyle intervention for diabetes treatment: 5-year outcomes from a randomized trial. *J Clin Endocrinol Metab*. 2020;105:866–876.
132. Arterburn DE, Bogart A, Sherwood NE, et al. A multisite study of long-term remission and relapse of type 2 diabetes mellitus following gastric bypass. *Obes Surg*. 2013;23:93–102.
133. Nannipieri M, Mari A, Anselmino M, et al. The role of beta-cell function and insulin sensitivity in the remission of type 2 diabetes after gastric bypass surgery. *J Clin Endocrinol Metab*. 2011;96:E1372–E1379.
134. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577–1589.
135. Pories WJ, Swanson MS, MacDonald KG, et al. Who would have thought it? An operation proves to be the most effective therapy for adult-onset diabetes mellitus. *Ann Surg*. 1995;222:339–350;discussion 350–352.
136. Schauer PR, Burguera B, Ikramuddin S, et al. Effect of laparoscopic Roux-en Y gastric bypass on type 2 diabetes mellitus. *Ann Surg*. 2003;238:467–484;discussion 84–85.
137. Isbell JM, Tamboli RA, Hansen EN, et al. The importance of caloric restriction in the early improvements in insulin sensitivity after Roux-en-Y gastric bypass surgery. *Diabetes Care*. 2010;33:1438–1442.
138. Fabbrini E, Tamboli RA, Magkos F, et al. Surgical removal of omental fat does not improve insulin sensitivity and cardiovascular risk factors in obese adults. *Gastroenterology*. 2010;139:448–455.
139. Klein S, Fontana L, Young VL, et al. Absence of an effect of liposuction on insulin action and risk factors for coronary heart disease. *N Engl J Med*. 2004;350:2549–2557.
140. Camastra S, Gastaldelli A, Mari A, et al. Early and longer term effects of gastric bypass surgery on tissue-specific insulin sensitivity and beta cell function in morbidly obese patients with and without type 2 diabetes. *Diabetologia*. 2011;54:2093–2102.
141. Muscelli E, Mingrone G, Camastra S, et al. Differential effect of weight loss on insulin resistance in surgically treated obese patients. *Am J Med*. 2005;118:51–57.
142. Vetter ML, Wadden TA, Teff KL, et al. GLP-1 plays a limited role in improved glycemia shortly after Roux-en-Y gastric bypass: a comparison with intensive lifestyle modification. *Diabetes*. 2015;64:434–446.
143. Dunn JP, Abumrad NN, Breitman I, et al. Hepatic and peripheral insulin sensitivity and diabetes remission at 1 month after Roux-en-Y gastric bypass surgery in patients randomized to omentectomy. *Diabetes Care*. 2012;35:137–142.
144. Bojsen-Moller KN, Dirksen C, Jorgensen NB, et al. Early enhancements of hepatic and later of peripheral insulin sensitivity combined with increased postprandial insulin secretion contribute to improved glycemic control after Roux-en-Y gastric bypass. *Diabetes*. 2014;63:1725–1737.
145. Bradley D, Conte C, Mittendorfer B, et al. Gastric bypass and banding equally improve insulin sensitivity and beta cell function. *J Clin Invest*. 2012;122:4667–4674.
146. Bradley D, Magkos F, Eagon JC, et al. Matched weight loss induced by sleeve gastrectomy or gastric bypass similarly improves metabolic function in obese subjects. *Obesity (Silver Spring)*. 2014;22:2026–2031.
147. Kashyap SR, Daud S, Kelly KR, et al. Acute effects of gastric bypass versus gastric restrictive surgery on beta-cell function and insulinotropic hormones in severely obese patients with type 2 diabetes. *Int J Obes (Lond)*. 2010;34:462–471.
148. Salehi M, Prigeon RL, D'Alessio DA. Gastric bypass surgery enhances glucagon-like peptide 1-stimulated postprandial insulin secretion in humans. *Diabetes*. 2011;60:2308–2314.
149. Rubino F, Forgione A, Cummings DE, et al. The mechanism of diabetes control after gastrointestinal bypass surgery reveals a role of the proximal small intestine in the pathophysiology of type 2 diabetes. *Ann Surg*. 2006;244:741–749.
150. Shah SH, Crosslin DR, Haynes CS, et al. Branched-chain amino acid levels are associated with improvement in insulin resistance with weight loss. *Diabetologia*. 2012;55:321–330.
151. Felig P, Marliss E, Cahill Jr GF. Plasma amino acid levels and insulin secretion in obesity. *N Engl J Med*. 1969;281:811–816.
152. Magkos F, Bradley D, Schweitzer GG, et al. Effect of Roux-en-Y gastric bypass and laparoscopic adjustable gastric banding on branched-chain amino acid metabolism. *Diabetes*. 2013;62:2757–2761.
153. Laferrere B, Reilly D, Arias S, et al. Differential metabolic impact of gastric bypass surgery versus dietary intervention in obese diabetic subjects despite identical weight loss. *Sci Transl Med*. 2011;3:80re2.
154. Laferrere B. Do we really know why diabetes remits after gastric bypass surgery? *Endocrine*. 2011;40:162–167.
155. Thomas C, Pellicciari R, Pruzanski M, et al. Targeting bile-acid signalling for metabolic diseases. *Nat Rev Drug Discov*. 2008;7:678–693.
156. Kohli R, Setchell KD, Kirby M, et al. A surgical model in male obese rats uncovers protective effects of bile acids post-bariatric surgery. *Endocrinology*. 2013;154:2341–2351.
157. Katsuma S, Hirasawa A, Tsujimoto G. Bile acids promote glucagon-like peptide-1 secretion through TGR5 in a murine enteroendocrine cell line STC-1. *Biochem Biophys Res Commun*. 2005;329:386–390.
158. Adrian TE, Gariballa S, Parekh KA, et al. Rectal taurocholate increases L cell and insulin secretion, and decreases blood glucose and food intake in obese type 2 diabetic volunteers. *Diabetologia*. 2012;55:2343–2347.
159. Kars M, Yang L, Gregor MF, et al. Tauroursodeoxycholic acid may improve liver and muscle but not adipose tissue insulin sensitivity in obese men and women. *Diabetes*. 2010;59:1899–1905.
160. Patti ME, Houten SM, Bianco AC, et al. Serum bile acids are higher in humans with prior gastric bypass: potential contribution to improved glucose and lipid metabolism. *Obesity (Silver Spring)*. 2009;17:1671–1677.
161. Pournaras DJ, Glicksman C, Vincent RP, et al. The role of bile after Roux-en-Y gastric bypass in promoting weight loss and improving glycaemic control. *Endocrinology*. 2012;153:3613–3619.
162. Kohli R, Bradley D, Setchell KD, et al. Weight loss induced by Roux-en-Y gastric bypass but not laparoscopic adjustable gastric banding increases circulating bile acids. *J Clin Endocrinol Metab*. 2013;98:E708–E712.
163. Sweeney TE, Morton JM. The human gut microbiome: a review of the effect of obesity and surgically induced weight loss. *JAMA Surg*. 2013;148:563–569.
164. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444:1027–1031.
165. Li JV, Ashrafian H, Bueter M, et al. Metabolic surgery profoundly influences gut microbial-host metabolic cross-talk. *Gut*. 2011;60:1214–1223.
166. Liou AP, Paziuk M, Luevano Jr JM, et al. Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci Transl Med*. 2013;5:178ra41.
167. Vrieze A, Van Nood E, Holleman F, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*. 2012;143:913–916.e7.
168. Buchwald H, Avidor Y, Braunwald E, et al. Bariatric surgery: a systematic review and meta-analysis. *JAMA*. 2004;292:1724–1737.
169. Vetter ML, Cardillo S, Rickels MR, et al. Narrative review: effect of bariatric surgery on type 2 diabetes mellitus. *Ann Intern Med*. 2009;150:94–103.
170. Adams TD, Gress RE, Smith SC, et al. Long-term mortality after gastric bypass surgery. *N Engl J Med*. 2007;357:753–761.
171. Sjostrom L, Narbro K, Sjostrom CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med*. 2007;357:741–752.
172. Sjostrom L, Peltonen M, Jacobson P, et al. Bariatric surgery and long-term cardiovascular events. *JAMA*. 2012;307:56–65.
173. Christou NV, Sampalis JS, Liberman M, et al. Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg*. 2004;240:416–423;discussion 423–424.
174. Pontiroli AE, Morabito A. Long-term prevention of mortality in morbid obesity through bariatric surgery. a systematic review and meta-analysis of trials performed with gastric banding and gastric bypass. *Ann Surg*. 2011;253:484–487.
175. Vest AR, Heneghan HM, Schauer PR, et al. Surgical management of obesity and the relationship to cardiovascular disease. *Circulation*. 2013;127:945–959.

176. Karlsson J, Taft C, Ryden A, et al. Ten-year trends in health-related quality of life after surgical and conventional treatment for severe obesity: the SOS intervention study. *Int J Obes (Lond)*. 2007;31:1248–1261.
177. Carlsson LMS, Sjoholm K, Jacobson P, et al. Life Expectancy after bariatric surgery in the Swedish obese subjects study. *N Engl J Med*. 2020;383:1535–1543.
178. Schiavon CA, Bersch-Ferreira AC, Santucci EV, et al. Effects of bariatric surgery in obese patients with hypertension: the GATEWAY randomized trial (gastric bypass to treat obese patients with steady hypertension). *Circulation*. 2018;137:1132–1142.
179. Climent E, Goday A, Pedro-Botet J, et al. Laparoscopic Roux-en-Y gastric bypass versus laparoscopic sleeve gastrectomy for 5-year hypertension remission in obese patients: a systematic review and meta-analysis. *J Hypertens*. 2020;38:185–195.
180. Ruiz-Tovar J, Carbajo MA, Jimenez JM, et al. Long-term follow-up after sleeve gastrectomy versus Roux-en-Y gastric bypass versus one-anastomosis gastric bypass: a prospective randomized comparative study of weight loss and remission of comorbidities. *Surg Endosc*. 2019;33:401–410.
181. Jensen TK, Scheike T, Keiding N, et al. Fecundability in relation to body mass and menstrual cycle patterns. *Epidemiology*. 1999;10:422–428.
182. Zaadstra BM, Seidell JC, Van Noord PA, et al. Fat and female fecundity: prospective study of effect of body fat distribution on conception rates. *BMJ*. 1993;306:484–487.
183. Ogden CL, Carroll MD, Curtin LR, et al. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*. 2006;295:1549–1555.
184. Skubleny D, Switzer NJ, Gill RS, et al. The impact of bariatric surgery on polycystic ovary syndrome: a systematic review and meta-analysis. *Obes Surg*. 2016;26:169–176.
185. Teitelman M, Grotegut CA, Williams NN, et al. The impact of bariatric surgery on menstrual patterns. *Obes Surg*. 2006;16:1457–1463.
186. Eid GM, Cottam DR, Velcu LM, et al. Effective treatment of polycystic ovarian syndrome with Roux-en-Y gastric bypass. *Surg Obes Relat Dis*. 2005;1:77–80.
187. Christ JP, Falcone T. Bariatric surgery improves hyperandrogenism, menstrual Irregularities, and metabolic Dysfunction among women with Polycystic Ovary Syndrome (PCOS). *Obes Surg*. 2018;28:2171–2177.
188. Marceau P, Kaufman D, Biron S, et al. Outcome of pregnancies after biliopancreatic diversion. *Obes Surg*. 2004;14:318–324.
189. Fredheim JM, Rollheim J, Sandbu R, et al. Obstructive sleep apnea after weight loss: a clinical trial comparing gastric bypass and intensive lifestyle intervention. *J Clin Sleep Med*. 2013;9:427–432.
190. Sarkhosh K, Switzer NJ, El-Hadi M, et al. The impact of bariatric surgery on obstructive sleep apnea: a systematic review. *Obes Surg*. 2013;23:414–423.
191. Goto T, Tsugawa Y, Faridi MK, et al. Reduced risk of acute exacerbation of COPD after bariatric surgery: a self-controlled case series study. *Chest*. 2018;153:611–617.
192. Borbely Y, Schaffner E, Zimmermann L, et al. De novo gastroesophageal reflux disease after sleeve gastrectomy: role of preoperative silent reflux. *Surg Endosc*. 2019;33:789–793.
193. Perry Y, Courcoulas AP, Fernando HC, et al. Laparoscopic Roux-en-Y gastric bypass for recalcitrant gastroesophageal reflux disease in morbidly obese patients. *J Soc Laparoendosc Surg*. 2004;8:19–23.
194. Frezza EE, Ikramuddin S, Gourash W, et al. Symptomatic improvement in gastroesophageal reflux disease (GERD) following laparoscopic Roux-en-Y gastric bypass. *Surg Endosc*. 2002;16:1027–1031.
195. Leslie DB, Kellogg TA, Boutelle KN, et al. Preserved vision without growth retardation after laparoscopic Roux-en-Y gastric bypass in a morbidly obese child with pseudotumor cerebri: 36-month follow-up. *J Pediatr Surg*. 2008;43:e27–e30.
196. Sun WYL, Switzer NJ, Dang JT, et al. Idiopathic intracranial hypertension and bariatric surgery: a systematic review. *Can J Surg*. 2020;63:E123–E128.
197. Avgerinos KI, Spyrou N, Mantzoros CS, et al. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism*. 2019;92:121–135.
198. Lega IC, Lipscombe LL. Review: diabetes, obesity, and cancer-pathophysiology and clinical implications. *Endocr Rev*. 2020;41:33–52.
199. Schauer DP, Feigelson HS, Koebnick C, et al. Bariatric surgery and the risk of cancer in a large multisite cohort. *Ann Surg*. 2019;269:95–101.
200. Longitudinal Assessment of Bariatric Surgery (LABS) Consortium, Flum DR, Belle SH, et al. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med*. 2009;361:445–454.
201. Jurgensen JA, Reidt W, Kellogg T, et al. Impact of patient attrition from bariatric surgery practice on clinical outcomes. *Obes Surg*. 2019;29:579–584.
202. Grover BT, Morell MC, Kothari SN, et al. Defining weight loss after bariatric surgery: a call for standardization. *Obes Surg*. 2019;29:3493–3499.
203. Gracia JA, Martinez M, Elia M, et al. Obesity surgery results depending on technique performed: long-term outcome. *Obes Surg*. 2009;19:432–438.
204. Magro DO, Geloneze B, Delfini R, et al. Long-term weight regain after gastric bypass: a 5-year prospective study. *Obes Surg*. 2008;18:648–651.
205. Odom J, Zalesin KC, Washington TL, et al. Behavioral predictors of weight regain after bariatric surgery. *Obes Surg*. 2010;20:349–356.
206. Colles SL, Dixon JB, O'Brien PE. Grazing and loss of control related to eating: two high-risk factors following bariatric surgery. *Obesity (Silver Spring)*. 2008;16:615–622.
207. Kinzl JF, Schrattenecker M, Traweger C, et al. Psychosocial predictors of weight loss after bariatric surgery. *Obes Surg*. 2006;16:1609–1614.
208. Freire RH, Borges MC, Alvarez-Leite JI, et al. Food quality, physical activity, and nutritional follow-up as determinant of weight regain after Roux-en-Y gastric bypass. *Nutrition*. 2012;28:53–58.
209. Nor Hanipah Z, Nasr EC, Bucak E, et al. Efficacy of adjuvant weight loss medication after bariatric surgery. *Surg Obes Relat Dis*. 2018;14:93–98.
210. Velapati SR, Shah M, Kuchkuntla AR, et al. Weight regain after bariatric surgery: prevalence, etiology, and treatment. *Curr Nutr Rep*. 2018;7:329–334.
211. Kellogg TA. Revisional bariatric surgery. *Surg Clin North Am*. 2011;91:1353–1371, x.
212. Vargas EJ, Abu Dayyeh BK, Storm AC, et al. Endoscopic management of dumping syndrome after Roux-en-Y gastric bypass: a large international series and proposed management strategy. *Gastrointest Endosc*. 2020;92:91–96.
213. Enani G, Bilgic E, Lebedeva E, et al. The incidence of iron deficiency anemia post-Roux-en-Y gastric bypass and sleeve gastrectomy: a systematic review. *Surg Endosc*. 2020;34:3002–3010.
214. Kellogg TA, Bantle JP, Leslie DB, et al. Postgastric bypass hyperinsulinemic hypoglycemia syndrome: characterization and response to a modified diet. *Surg Obes Relat Dis*. 2008;4:492–499.
215. Sjostrom L. Review of the key results from the Swedish Obese Subjects (SOS) trial - a prospective controlled intervention study of bariatric surgery. *J Intern Med*. 2013;273:219–234.
216. Brix JM, Kopp HP, Hollerl F, et al. Frequency of hypoglycaemia after different bariatric surgical procedures. *Obes Facts*. 2019;12:397–406.
217. Lee CJ, Wood GC, Lazo M, et al. Risk of post-gastric bypass surgery hypoglycemia in nondiabetic individuals: a single center experience. *Obesity (Silver Spring)*. 2016;24:1342–1348.
218. Gebhard B, Holst JJ, Biegelmayer C, et al. Postprandial GLP-1, norepinephrine, and reactive hypoglycemia in dumping syndrome. *Dig Dis Sci*. 2001;46:1915–1923.
219. Goldfine AB, Mun EC, Devine E, et al. Patients with neuroglycopenia after gastric bypass surgery have exaggerated incretin and insulin secretory responses to a mixed meal. *J Clin Endocrinol Metab*. 2007;92:4678–4685.
220. Palladino AA, Sayed S, Levitt Katz LE, et al. Increased glucagon-like peptide-1 secretion and postprandial hypoglycemia in children after Nissen fundoplication. *J Clin Endocrinol Metab*. 2009;94:39–44.
221. Patti ME, McMahon G, Mun EC, et al. Severe hypoglycaemia post-gastric bypass requiring partial pancreatectomy: evidence for inappropriate insulin secretion and pancreatic islet hyperplasia. *Diabetologia*. 2005;48:2236–2240.
222. Service GJ, Thompson GB, Service FJ, et al. Hyperinsulinemic hypoglycemia with nesidioblastosis after gastric-bypass surgery. *N Engl J Med*. 2005;353:249–254.

223. Salehi M, Gastaldelli A, D'Alessio DA. Altered islet function and insulin clearance cause hyperinsulinemia in gastric bypass patients with symptoms of postprandial hypoglycemia. *J Clin Endocrinol Metab*. 2014;99:2008–2017.
224. Salehi M, Gastaldelli A, D'Alessio DA. Blockade of glucagon-like peptide 1 receptor corrects postprandial hypoglycemia after gastric bypass. *Gastroenterology*. 2014;146:669–680.e2.
225. Abrahamsson N, Borjesson JL, Sundbom M, et al. Gastric bypass reduces symptoms and hormonal responses in hypoglycemia. *Diabetes*. 2016;65:2667–2675.
226. Coates PS, Fernstrom JD, Fernstrom MH, et al. Gastric bypass surgery for morbid obesity leads to an increase in bone turnover and a decrease in bone mass. *J Clin Endocrinol Metab*. 2004;89:1061–1065.
227. Lobato CB, Pereira SS, Guimaraes M, et al. A potential role for endogenous glucagon in preventing post-bariatric hypoglycemia. *Front Endocrinol (Lausanne)*. 2020;11:608248.
228. Mulla CM, Zavitsanou S, Laguna Sanz AJ, et al. A randomized, placebo-controlled double-blind trial of a closed-loop glucagon system for post-bariatric hypoglycemia. *J Clin Endocrinol Metab*. 2020;105:e1260–e1271.
229. Wilson LM, Castle JR. Stable liquid glucagon: beyond emergency hypoglycemia rescue. *J Diabetes Sci Technol*. 2018;12:847–853.
230. McLaughlin T, Peck M, Holst J, et al. Reversible hyperinsulinemic hypoglycemia after gastric bypass: a consequence of altered nutrient delivery. *J Clin Endocrinol Metab*. 2010;95:1851–1855.
231. Bloomberg RD, Fleishman A, Nalle JE, et al. Nutritional deficiencies following bariatric surgery: what have we learned? *Obes Surg*. 2005;15:145–154.
232. Suarez Llanos JP, Fuentes Ferrer M, Alvarez-Sala-Walther L, et al. Protein malnutrition incidence comparison after gastric bypass versus biliopancreatic diversion. *Nutr Hosp*. 2015;32:80–86.
233. Koppe U, Nitsch D, Mansfield KE, et al. Long-term effects of bariatric surgery on acute kidney injury: a propensity-matched cohort in the UK Clinical Practice Research Datalink. *BMJ Open*. 2018;8:e020371.
234. Brolin RE, LaMarca LB, Kenler HA, et al. Malabsorptive gastric bypass in patients with superobesity. *J Gastrointest Surg*. 2002;6:195–203; discussion 204–205.
235. Primavera A, Brusa G, Novello P, et al. Wernicke-Korsakoff encephalopathy following biliopancreatic diversion. *Obes Surg*. 1993;3:175–177.
236. Reinert RR. Clinical efficacy of ketolides in the treatment of respiratory tract infections. *J Antimicrob Chemother*. 2004;53:918–927.
237. Bruno C, Fulford AD, Potts JR, et al. Serum markers of bone turnover are increased at six and 18 months after Roux-en-Y bariatric surgery: correlation with the reduction in leptin. *J Clin Endocrinol Metab*. 2010;95:159–166.
238. Fleischer J, Stein EM, Bessler M, et al. The decline in hip bone density after gastric bypass surgery is associated with extent of weight loss. *J Clin Endocrinol Metab*. 2008;93:3735–3740.
239. Stein EM, Carrelli A, Young P, et al. Bariatric surgery results in cortical bone loss. *J Clin Endocrinol Metab*. 2013;98:541–549.
240. Crawford MR, Pham N, Khan L, et al. Increased bone turnover in type 2 diabetes patients randomized to bariatric surgery versus medical therapy at 5 years. *Endocr Pract*. 2018;24:256–264.
241. Peterson LA, Zeng X, Caufield-Noll CP, et al. Vitamin D status and supplementation before and after bariatric surgery: a comprehensive literature review. *Surg Obes Relat Dis*. 2016;12:693–702.
242. King D, Siau K, Senthil L, et al. Copper deficiency myelopathy after upper gastrointestinal surgery. *Nutr Clin Pract*. 2018;33:515–519.
243. Kumar P, Hamza N, Madhok B, et al. Copper deficiency after gastric bypass for morbid obesity: a systematic review. *Obes Surg*. 2016;26:1335–1342.
244. Kolotkin RL, Davidson LE, Crosby RD, et al. Six-year changes in health-related quality of life in gastric bypass patients versus obese comparison groups. *Surg Obes Relat Dis*. 2012;8:625–633.
245. Ivezaj V, Saules KK, Wiedemann AA. "I didn't see this coming.": why are postbariatric patients in substance abuse treatment? Patients' perceptions of etiology and future recommendations. *Obes Surg*. 2012;22:1308–1314.
246. Smith KE, Engel SG, Steffen KJ, et al. Problematic alcohol use and associated characteristics following bariatric surgery. *Obes Surg*. 2018;28:1248–1254.
247. Steffen KJ, Engel SG, Pollert GA, et al. Blood alcohol concentrations rise rapidly and dramatically after Roux-en-Y gastric bypass. *Surg Obes Relat Dis*. 2013;9:470–473.

28

Development of the Endocrine Pancreas

Matthias Hebrok and Michael S. German

OUTLINE

Pancreas Morphogenesis, 474
Early Organ Specification and Bud Formation, 475
  Tissue Interaction and Signaling Pathways, 475
  Induction of the Pancreatic Gene Expression Program, 476
Cell Type Differentiation, 478
  Signaling Pathways, 478
  Transcription Factors, 479
  Islet Formation, 480
The Role of Pancreatic Development in Human Diabetes, 481

The pancreas provides both the enzymes that digest the food in the gut and the hormones that control utilization of the nutrients supplied by that digested food. Two distinct components of the pancreas, the exocrine and endocrine compartments, accomplish these related tasks. The ductal and acinar cells of the exocrine compartment comprise the majority of the organ, while the islets of Langerhans of the endocrine portion are scattered throughout the exocrine matrix.

During pancreatic development, a common pool of progenitor cells in the early gut endoderm differentiates into these distinct cell types. Positional and temporal cues provided by extracellular signals direct sequential changes in the gene expression program of individual cells and ultimately determine the phenotype of the differentiated cells and the organization of the mature organ. Any breakdown in the orchestration of this complex process of growth, differentiation, migration and organization can damage the mature organ and can cause diabetes by impairing the function of the insulin-producing β cells. An understanding of pancreatic development has provided us with new insights into the causes of diabetes and new strategies for intervening in the disease.

PANCREAS MORPHOGENESIS

During gastrulation, mesendodermal cells that migrate through the node region of the early vertebrate embryo start to segregate to form distinct mesodermal and endodermal germ layers. Endodermal cells aggregate to form a contiguous epithelium that spans the whole length of the forming gastrointestinal tract of the developing embryo. Digestive organs develop along the anterior-posterior and dorsal-ventral axes of this sheet of cells in a precise and predetermined pattern.[1-3]

The pancreas forms from clumps of cells that bud from the dorsal and ventral aspects of the gut endoderm near the foregut/midgut junction[4,5] (Fig. 28.1). Well before any morphologic evidence of the pancreas becomes apparent, interactions between the mesodermally-derived notochord and the endodermal epithelium initiate dorsal pancreas formation.[6] Similarly, interactions between the lateral plate mesoderm and the lateral endoderm ensure proper induction of the ventral pancreas.[7] The first morphological sign of pancreas formation is a thickening of the dorsal endodermal sheet caudal to the stomach anlage. This bud structure continues to grow and initiates contact with the dorsal aorta that has fused in the midline and thus separated the notochord and the endoderm.[8] Subsequently, dorsal mesenchyme replaces the aorta and provides essential signals that stimulate pancreatic epithelial branching and cell differentiation.

Similar tissue interactions guide the development of the two ventral buds. In contrast to the dorsal bud that forms in the midline of the epithelium, the ventral buds are derived from the ventro-lateral edges of the endodermal sheet. Instructive signals provided by the lateral plate mesenchyme initiate ventral pancreas formation, and interactions with smaller blood vessels further support organ growth.[8] One of these ventral buds degenerates, while the other one fuses with the dorsal bud when stomach and gut rotate to form the mature organ. The regulation of this rotation is not well understood, but perturbations lead to malformations in humans, including annular pancreas, wherein a part of the ventral pancreas is mislocalized and can constrict the adjacent duodenum.[9]

Mesenchymal-epithelial interactions are required for proper development of the mature tissue.[10,11] Recent studies indicate that the pancreatic mesenchyme governs not only the expansion of early pancreatic progenitors,[12] but also regulate the proliferation of already differentiated endocrine and exocrine cells towards the end of gestation.[13] However, the epithelial cells start to outgrow the mesenchyme during later stages of development, and few remnants of this tissue remain in the adult pancreas. Epithelial cells branch into the surrounding mesenchyme to form an elaborate duct system that allows transport of exocrine enzymes into the duodenum. Distal cells of the branching epithelium differentiate into clusters of acinar cells that are connected to the duct system via centroacinar cells. Acinar cells are organized as small glands (acini) that produce a variety of digestive enzymes, including amylases, peptidases, nucleases, and lipases.

In contrast to the continuous acinar-duct system, endocrine precursors delaminate from the immature ducts during development and migrate into the surrounding mesenchyme, where they aggregate and organize into islets of Langerhans. These distinct islet cells produce a variety of hormones, including insulin, glucagon, pancreatic polypeptide, somatostatin, and ghrelin, which are secreted directly into the bloodstream and regulate gastrointestinal function and nutrient intake, storage, and utilization. Due to intimate interactions between endocrine cells and blood vessels, islets are highly vascularized, with a highly fenestrated network of blood vessels providing ready access to

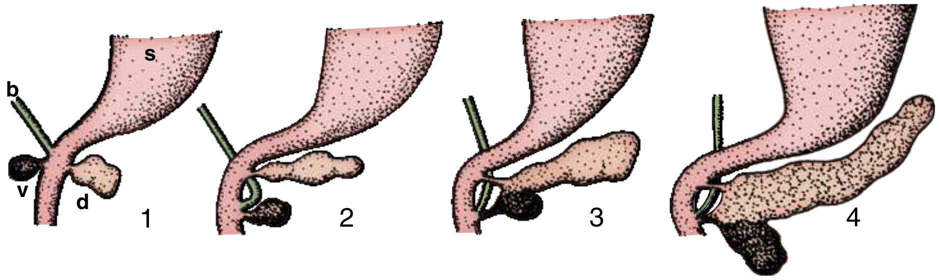


Figure 28.1 Schematic outline of the development of the mammalian pancreas. 1, Dorsal and ventral buds arise from the duodenal part of the primitive gut endoderm *(blackened).* **2,** The dominant ventral bud, following the bile duct, turns and approaches the dorsal bud. **3,** The two buds fuse together. **4,** The ventral bud becomes part of the head of the pancreas (the pancreatic polypeptide–rich duodenal lobe), whereas the dorsal bud gives rise to the body and tail (the glucagon-rich splenic lobe); the pancreatic juice is drained off by the main and accessory pancreatic ducts. *b*, Bile duct; *d*, dorsal bud; *s*, stomach; *v*, ventral bud.

circulating nutrients, signaling molecules, particles, and cells. Similarly, hormone secretion is controlled at least in part by the nervous system; the sympathetic, parasympathetic, and sensory neurons that innervate the islets derive from cells that migrate from the neural crest into the pancreas just as the two buds fuse.[14]

KEY POINTS

- Very early interactions within and between the forming germ layers of the developing embryo pattern the cells that will eventually form the pancreas and initiate the formation of the pancreatic buds.

EARLY ORGAN SPECIFICATION AND BUD FORMATION

Tissue Interaction and Signaling Pathways

With the exception of supporting mesenchyme, blood vessels, and innervating neurons, all mature pancreatic cell types derive from epithelial cells of endodermal origin.[15-17] Several studies have addressed the question of how the pancreas anlage becomes specified within the epithelial sheet that forms the endoderm, and increasing evidence points to a series of tissue interactions as important steps during early stages of organogenesis. During gastrulation, the mesectodermal portion of the embryo signals to the endodermal sheet to establish a prepattern along the anterior-posterior axis.[3]

Although the exact signals required for patterning of the fore-midgut area are not completely understood, studies in mice have implicated the fibroblast growth factor (FGF) signaling pathway in this process.[18] Before any morphologic signs of pancreas formation are apparent, the midline of the endodermal epithelium comes in close contact with the overlying notochord, a mesodermal structure known to regulate organogenesis and cell differentiation in adjacent structures. The notochord provides secreted signaling peptides that are required for initiation of dorsal pancreas organogenesis. Members of the transforming growth factor β (TGF-β) and FGF signaling pathways have been implicated in the notochord-mediated induction of pancreas development. A critical aspect of their activity involves repression of the expression of Sonic hedgehog (Shh), a member of the Hedgehog signaling pathway within the pancreas anlage.[19] Ectopic elevation of Hedgehog signaling at the onset of pancreas formation results in pancreas agenesis, indicating that tight regulation of the activity of this pathway is essential for proper organ formation. However, the notochord provides only permissive signals, because it cannot induce expression of pancreatic markers in nonpancreatic endoderm.[6]

In contrast to the singular dorsal pancreas, two distinct pancreatic buds form within the ventro-lateral endoderm, a region that is not contacted by the notochord. Tissue recombination experiments in chicks and mice demonstrate that similar but distinct signals from the lateral plate mesoderm (LPM) ensure the correct temporal-spatial induction of the ventral pancreas.[7,18] Before it contacts the LPM, no pancreatic markers are detectable in the presumptive ventral pancreas, but LPM or the TGF-β superfamily members activin and morphogenetic proteins (BMPs), as well as retinoic acid (RA), have been shown to induce expression of pancreatic genes in underlying endoderm (see following section).[7] Of note, mesenchyme isolated from the pancreatic region carries the potential to induce pancreatic gene expression in anterior endoderm normally fated to develop into stomach and esophagus. Thus, in contrast to the notochord–endoderm interaction that induces dorsal pancreas development in a permissive fashion, the LPM actively instructs uncommitted endoderm to differentiate into pancreatic tissue. Activin, BMP, and RA signals that mimic the pancreas instructive activity of the LPM are potentially useful in designing a cocktail of growth factors designed to regulate differentiation of uncommitted progenitor cells towards a pancreatic fate.

Other studies have shown that interactions between heart mesenchyme and ventral foregut endoderm regulate differentiation of ventral pancreas and liver progenitor cells.[20] In an analogy to the notochord-dorsal pancreas bud interactions, the septum transversum and heart mesenchyme produce BMP and FGF signals, respectively, that regulate organ differentiation of the ventral endodermal organs, including liver and pancreas. Because of its close proximity, the concentration of FGF ligands produced by the heart mesenchyme is higher at the area fated to develop into liver and lower at the more distal region that develops into ventral pancreas.[20] This is also the case for dorsal pancreatic epithelium, in which low levels of FGF signals have been shown to initiate pancreatic gene expression and block Shh expression, while higher levels of FGF signaling block pancreas induction via increased Hedgehog signaling. Thus, formation of dorsal and ventral pancreatic buds depends on low FGF signals and inhibition of Hedgehog signaling. It is important to note that Hedgehog signaling is active in areas immediately adjacent to the dorsal and ventral buds, thereby establishing a molecular boundary that prevents ectopic expansion of pancreatic tissue beyond its normal borders.

During subsequent stages, the flattened endodermal sheet expands and folds itself to form a tube-like structure. The first morphologic sign of pancreas formation is an epithelial thickening at the dorsal side of the forming gut tube caudal to the stomach anlage, shortly followed by the appearance of two ventral thickenings next to the liver diverticulum. Outgrowth and tissue-specific gene expression continue to depend on cues derived from adjacent tissues. In contrast to earlier stages, mesenchymal signals are now produced from forming blood vessels: the dorsal aorta and smaller vitelline veins that contact the dorsal and ventral buds, respectively.[21] Depletion of blood vessels via explantation in *Xenopus* or via homozygous recombination in transgenic mice that lack the *Flk-1* gene, which encodes a receptor for vascular epithelial growth factor (VEGF), impairs the differentiation of the dorsal pancreatic epithelium.[21,22] Vice versa, ectopic expression of VEGF leads to hypervascularization and islet hyperplasia in pancreatic tissues, as well as ectopic insulin expression in nonpancreatic stomach epithelium, suggesting that VEGF or other molecules secreted by endothelial cells support endocrine cell differentiation. Detailed analysis of the interaction between blood vessels and pancreas buds has revealed physical contact between the aorta and the dorsal bud epithelium, while vitelline veins and ventral bud epithelium are separated by a fine band of mesenchymal cells, although endothelial cells are less critical for ventral bud formation.[22]

Recent results have provided further evidence for the role of signaling molecules in setting up organ boundaries in the pancreas anlage. Studies in zebrafish suggest that Fgf10 expression is critical for separating pancreatic and liver cells in the area of the hepatopancreatic ductal system. Fgf10 is expressed in the mesenchyme surrounding the developing hepatopancreatic duct and functions to restrain inappropriate differentiation of adjacent pancreas into liver cells, as well as liver into pancreatic cells.[23] Thus, secreted signaling molecules are essential in defining the boundaries of the pancreas.

Upon specification of the dorsal and ventral pancreatic anlagen, additional mesenchymal-epithelial interactions contribute to later steps of pancreas organogenesis.[10,11] Splanchnic mesoderm expands medially and surrounds the pancreatic epithelial buds. Signals from the mesenchyme promote epithelial expansion that results in the formation of a branching structure composed of undifferentiated ductal cells. Based on morphologic criteria, Rutter and colleagues argued that the early endodermal cells are "protodifferentiated," and that mature exocrine and endocrine cells appear only after the secondary transition, a process during which the numbers of differentiated acinar and β cells increase significantly.[4,24] More recent studies have shown that early pancreatic mesenchyme generally induces growth and proliferation of epithelial cells, while at later stages mesenchymal signals promote expansion of exocrine and endocrine cells.[13]

KEY POINTS

- A series of signals from adjacent mesoderm and then from overlying mesenchyme first specify the prepancreatic endoderm and then direct the formation, growth, and morphogenesis of the dorsal and lateral pancreatic buds near the foregut/midgut junction.

Induction of the Pancreatic Gene Expression Program

The morphologic changes that occur as pancreas differentiates from gut endoderm depend on sequential changes in gene expression. Extracellular signals provided by tissue interactions between the developing pancreatic endoderm and adjacent tissues ultimately affect cell phenotype by altering gene expression within the nucleus. Recent studies have outlined the underlying molecular events that control these changes in gene expression and have focused in particular on transcription factors (nuclear proteins that control gene transcription) and the epigenetic changes that these factors and other signals induce.

Transcription factors expressed broadly in the endoderm provide competence to respond to the endoderm patterning signals and include early endoderm factors SOX17[25] and RFX6[26]; several transcription factors originally described in the adult liver, such as hepatic nuclear factor HNF1B (a POU-homeodomain factor),[27] forkhead factors FOXA1 and FOXA2,[28] HNF4A (a fatty-acid-binding nuclear receptor),[29-31] and ONECUT1[32,33]; and the zinc finger transcription factors GATA4 and GATA6.[34,35] These genes function in transcriptional hierarchies that not only control endoderm patterning, but also persist in endoderm-derived organs and play roles in the mature pancreas.[36-41] Several of these endoderm factors have also been directly implicated in the control of mature pancreatic gene expression (see Fig. 28.5 and Table 28.1).

The earliest genes selectively expressed in the region destined to become pancreas, the prepancreatic endoderm, are two transcription factors, the parahox homeodomain factor PDX1[42-44] and the basic-helix-loop-helix (bHLH) transcription factor PTF1A.[45,46] PDX1 expression first appears in prepancreatic endoderm more than a day before the initial formation of the dorsal pancreatic bud, and is preceded immediately by the appearance of another parahox factor, MNX1,[47,48] which is expressed more broadly than PDX1 in anterior endoderm. Expression of both MNX1 and PDX1 persists in the initial pancreatic buds, although MNX1 expression is extinguished quickly thereafter, while PDX1 expression lasts a few more days. Expression of both factors is reactivated in mature β cells.

Dorsal, but not ventral, expression of PDX1 depends on MNX1.[47,48] Because MNX1 is also expressed in the notochord during the same period, it is formally possible that MNX1 could control dorsal PDX1 expression via signals from the notochord. PDX1 expression in the dorsal prepancreatic endoderm does not require signals from the notochord, however, suggesting that MNX1 functions cell-intrinsically in inducing PDX1. Extrinsic signals that induce PDX1 expression in the dorsal prepancreatic endoderm have not been identified.[6] On the other hand, as described in the preceding section, signals from the LPM including activin, BMPs, and retinoic acid can induce PDX1 expression in the ventral prepancreatic endoderm.[7]

Studies of the *Pdx1* gene promoter have identified several additional endodermal transcription factors as potential intrinsic regulators of *Pdx1* expression.[49-56] These include members of the HNF1 and FOXA families of transcription factors, Onecut1, the paired homeodomain factor PAX6, and PDX1 itself[49,50,55,57,58] (MNX1 also may act through these same PDX1 binding sites, given its similarity to PDX1 in the DNA-binding homeodomain). HNF1A and PAX6 are not expressed early enough or broadly enough to initiate the early expression of PDX1 in the embryonic gut and pancreatic buds, but the expression patterns of Onecut1, HNF1B, FOXA1, and FOXA2 suggest that they could play this role (Fig. 28.2). Mice lacking Onecut1 have reduced expression of PDX1,[58] and embryoid bodies lacking FOX2 fail to activate the *Pdx1* gene,[55] suggesting that these two factors may be *bona fide* activators of PDX1 expression in the prepancreatic endoderm and pancreatic buds and therefore lie directly upstream of PDX1 in the hierarchy of factors involved in initiating pancreas development (see Fig. 28.5).

PDX1 is required for outgrowth of the pancreatic buds. In mice with targeted disruptions of the *Pdx1* gene, the pancreatic buds fail to branch and expand after initial formation, yielding an animal that lacks a pancreas at birth.[43,44,59] Confirming its role upstream of PDX1 in the endoderm, lack of Onecut1 results in reduced pancreatic size[58], and removal of MNX1 specifically arrests the growth of the dorsal pancreatic bud.[47,48] Homozygous null mutations in the *PDX1* gene, which encodes human PDX1, have been identified in humans born with pancreatic agenesis,[60] and mutations in the zebrafish gene encoding PDX1 also affect pancreas development,[61] reflecting an evolutionary conservation of PDX1 function.

Expression of PTF1A follows shortly after PDX1 and is restricted within the endoderm to the dorsal and ventral pancreatic buds.[62,63]

TABLE 28.1 Pancreas Transcription Factors

Factor	Family	Expression	Mouse Mutations*	Human Mutations	References
NEUROG3	bHLH	Fetal pancreas (endocrine progenitor cells) and CNS	Diabetes, no islet cells	Hom: malabsorption, early onset diabetes	94,95,105,115,147,148
NEUROD1	bHLH	Islet, gut endocrine cells, CNS	Diabetes, decreased islet cells	Het: MODY6, late onset diabetes Hom: PND, DD, ataxia, deafness, and visual impairment	114,141,149-153
PTF1A	bHLH	Exocrine pancreas, CNS	Exocrine pancreatic agenesis, islets in spleen	Hom: PND, pancreatic agenesis, cerebellar agenesis	46,62,63,154
PDX1	Parahox homeodomain	α and β cells, duodenum, stomach, CNS	Pancreatic agenesis	Het: MODY4 Hom: PND, pancreatic agenesis	42-44,59,60,139,155-161
MNX1	Parahox homeodomain	β cells, gut, lymphoid, CNS	Dorsal pancreatic agenesis	Het: Sacral agenesis Hom: PND, sacral agenesis, DD, pulmonary defects	47,48,144,162
PAX4	Paired homeodomain	fetal pancreas and CNS	Decreased β and δ cells	Het: late onset diabetes, MODY, KPD Hom: early diabetes, KPD	122,163-165
PAX6	Paired homeodomain	Islet, gut endocrine cells, CNS	Decrease in all islet cells, decreased glucagon	Het: Ocular anomalies, IGT Compound Het: PND, hypopituitarism, microcephaly, microophthalmia	128,129,166,167
ARX	Paired-related homeodomain	α and δ cells and CNS	Decreased α cells	X-linked mental retardation	130,168
NKX2.2	NK homeodomain	β,α, and PP cells and CNS	Diabetes, no insulin	Hom: PND, DD, cortical blindness, hearing impairment	125,126,144,169,170
NKX6.1	NK homeodomain	β cells and CNS	Decreased β cells, postnatal lethal		125,126,169,171
ISL1	LIM homeodomain	Islet and CNS	No islet cells, embryonic lethal	Het: late onset diabetes	127,142,172-176
HNF1A	Pou homeodomain	Islet, liver, kidney	Diabetes, impaired β cell glucose sensing	Het: MODY3	40,107,122,138,177-179
HNF1B	Pou homeodomain	Islet, pancreatic duct, liver, kidney	Embryonic lethal	Het: MODY5, renal defects	122,140,180,181
ONECUT1	Cut homeodomain	Pancreatic duct, liver	IGT, small islets		32,33,107,108,182,183
GATA4	GATA-binding zinc finger domain	Heart, liver, gut, exocrine pancreas	Early embryonic lethal	Het: congenital cardiac and intestinal defects, early-onset diabetes	34,184
GATA6	GATA-binding zinc finger domain	Endoderm, heart, lung, gut, endocrine pancreas	Early embryonic lethal	Het: congenital cardiac, intestinal defects, pancreatic hypoplasia or agenesis, early-onset diabetes	34,184
RFX6	RFX	Endoderm, islet, gut endocrine	No α, β, δ, or ε cells, increased PP cells; intestinal defects	Hom: PND, intestinal defects	26
HNF4A	Nuclear receptor	Liver, islet, kidney	Embryonic lethal	Het: MODY1	30,122,137,185-187
MAFA	bZip	β cells, eye and CNS	Diabetes, normal islet development	Het: Insulinomatosis (multiple insulinomas) or diabetes	188-192
MAFB	bZip	α cells and immature β cells in mice, mature α and β cells in humans	Reduced α and β cells		192-195

CNS, Central nervous system; *PP,* pancreatic polypeptide; *IGT*, impaired glucose tolerance; *MODY,* maturity-onset diabetes of the young; *PND,* permanent neonatal diabetes; *DD,* double diabetes; *KPD,* ketosis-prone diabetes.

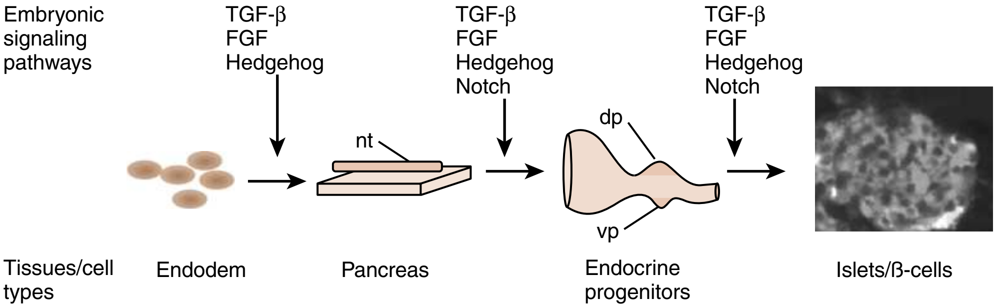


Figure 28.2 Embryonic signaling pathways regulate different steps of pancreas and endocrine islet formation. *dp*, Dorsal pancreas; *nt*, notochord; *vp*, ventral pancreas; *TGF*, transforming growth factor; *FGF*, fibroblast growth factor.

The aorta induces expression of PTF1A in the dorsal pancreatic bud, but vascular-derived signals do not appear to play as critical a role in its expression in the ventral buds.[22] Although PTF1A was initially described as an exocrine-specific transcription factor,[46,63] in fact it plays an essential role in determining pancreatic cell fate. Although in the mature pancreas its expression is restricted to the acinar cells, in mouse embryos lacking PTF1A, cells normally fated to contribute to both exocrine and endocrine lineages in the pancreas instead revert to duodenal epithelium.[45]

KEY POINTS

- The formation and growth of the pancreatic buds depends on changes in gene expression directed by a set transcription factors initially expressed broadly in the epithelial layer of buds.

CELL TYPE DIFFERENTIATION

Signaling Pathways

Several embryonic signaling pathways regulate distinct aspects of pancreas organogenesis and endocrine cell differentiation. As discussed earlier, the TGF-β, FGF, and Hedgehog signaling pathways coordinate initiation of pancreas formation. After bud formation, these pathways play additional roles in the differentiation of the distinct cell types that form the mature pancreas.

FGF7 and FGF10 are expressed transiently in pancreatic mesenchyme and have been shown to promote proliferation of pancreatic epithelial cells.[12,64] Transgenic mice that carry a homozygous deletion in the *FGF10* gene display severe defects in pancreas morphogenesis that result from loss of PD1-positive pancreatic epithelial progenitor cells.[12] FGF signaling also has been shown to regulate proliferation of endocrine and exocrine cells,[65,66] and expression of a dominant active form of the FGF receptor 2 (FGFR2) inhibits endocrine and exocrine cells.[65-67]

TGF-β/activin signaling has been shown to affect differentially exocrine and endocrine cell differentiation. Treatment of isolated pancreatic buds with soluble TGF-β1 stimulates differentiation of endocrine cells, particularly β cells and pancreatic polypeptide cells.[68] Signaling via activins, a subgroup of TGF-β signaling members, also has been shown to affect preferentially endocrine cell differentiation. Treatment of pancreatic buds with follistatin, an antagonist that physically binds to activins, inhibits endocrine cell differentiation while promoting exocrine cell formation.[69] Transgenic mice ectopically expressing dominant-negative activin type II receptor (dnActRII) develop islet hypoplasia, a phenotype also observed in mice carrying targeted mutations in both ActRIIA and ActRIIB.[70-72] More recent work has focused on the downstream TGF-β/activin signaling cascade of. Ligand binding to their respective receptors on receiving cells results in the phosphorylation of receptor associated Sma- and Mad-related proteins 1, 2, 3, 5, and 8 (R-Smads). Another Smad, Smad4, acts as a universal partner for the R-Smads, and the R-Smad/Smad4 complex interacts with distinct DNA-binding factors to regulate target gene transcription. Surprisingly, loss of Smad4 has no appreciable effects on pancreas formation,[73,74] suggesting Smad4-independent TGF-β/activin functions. Additional support for the TGF-β/activin requirement during pancreas development comes from studies in which Smad7, a negative factor that interferes with Smad–receptor and Smad–Smad interactions, is ectopically expressed in pancreatic tissue. Embryonic expression of Smad7 in pancreatic epithelium results in pancreas and β cell hypoplasia.[75] Of note, ectopic expression of Smad7 specifically in adult pancreatic PDX1+-cells resulted in hypoinsulinemia followed by diabetes.

The role of Hedgehog signaling during endocrine cell differentiation is less clear. At early stages, elevation of Hedgehog signaling in the pancreas anlage results in transformation of the pancreatic mesenchyme into duodenal mesoderm.[19,76] Increased signaling impairs pancreatic organogenesis, at least in part, because signals normally provided by the pancreatic mesenchyme are reduced. Also, as has been shown for other organs, ectopic activation of Hedgehog signaling reduces FGF10 expression, which is known to promote pancreatic epithelial and endocrine cell expansion.[77,78] By contrast, a low level of Hedgehog signaling is detected throughout pancreas organogenesis and in mature islets.[79,80] In addition, cell culture experiments have revealed that Hedgehog signaling activates the *Pdx1* and insulin gene promoters in cultured insulinoma cells,[81,82] suggesting a different and potentially important role for the pathway in maintaining mature β cell function. Elimination of Smoothened, a critical component of the Hedgehog signaling cascade, in pancreatic epithelial cells revealed a requirement of the pathway during embryonic β cell formation. While β cell numbers were restored by birth, studies on adult transgenic animals showed defects in insulin production and glucose intolerance.[79] Intriguingly, similar defects in β cell function were observed in transgenic mice with overactive Hedgehog signaling,[83] suggesting that perfectly balancing the output of this signaling cascade is critical for maintenance of normal β cell function.

Hedgehog signaling shares many features with another embryonic pathway, the canonical Wnt cascade. Several Wnt ligands and their respective Frizzled receptors have been identified in embryonic and adult pancreas.[84,85] Functional studies in which dominant-negative receptors were expressed ectopically, or β-catenin, an essential component of the canonical Wnt signaling cascade, was eliminated, have pointed to an important role of this pathway during embryonic pancreas formation; however, its exact functions in specific pancreas cell types remains under debate. Although several studies found a requirement for canonical Wnt signaling during exocrine development,[86-88] others have noted a role for β-catenin in endocrine function.[89] Ectopic activation of β-catenin signaling results in different phenotypes, depending on the time of expression. Activation at the onset of pancreas formation blocks progenitor cell expansion, while later expression results in a profound increase in acinar cell numbers.[90] A recent study analyzing the requirement for the mouse homolog of Pygopus, Pygo2, a necessary nuclear component of the Wnt cascade, revealed a novel role for Wnt signaling in early pancreatic mesenchyme.[91] According to this study, mesenchymal mPygo2 is essential in regulating epithelial progenitor proliferation, including endocrine progenitors. Thus, Wnt signaling regulates pancreas formation at various stages in both mesenchyme and epithelium. Furthermore, studies in adult animals indicate a role for canonical Wnt signaling in adult β cell proliferation.[92]

In addition to the TGF-β, FGF, Hedgehog, and Wnt signaling pathways, Notch signaling contributes to cell fate choices and differentiation in the pancreas, as it does in many other organs during embryogenesis. Notch signaling commonly regulates cell fate decision via a process known as lateral inhibition, in which a given cell within a homogenous field of cells becomes less susceptible to ligand-activated Notch signaling. As a consequence, this cell initiates a specific differentiation program while continuing to provide Notch ligands to adjacent cells, thereby blocking the differentiation of its neighbors (Fig. 28.3).[93] A similar mechanism regulates the differentiation of endocrine cells during pancreas development. Reduction of Notch pathway activity in transgenic mice lacking essential Notch signaling components results in upregulation of expression of Neurogenin3 (NEUROG3), a bHLH transcription factor required for endocrine formation[94] (see following section), and precocious endocrine differentiation.[95,96]

More recent studies have provided an even more nuanced view of Notch activity during endocrine progenitor formation. Studies by Shih and colleagues describe how intermediate levels of Notch signaling activate expression of Sox9, which in turn activates Ngn3 expression. In contrast, high levels of Notch activate both Sox9 and Hes1 expression and thereby promote formation of duct progenitors within the trunk region of the forming epithelial network.[97] The key findings from this work establish Notch as a critical determinant in both endocrine and duct fate development.

Other studies have revealed roles for Notch signaling in other steps of pancreatic formation, both before and after its well-recognized role in regulating endocrine specification. Even before pancreas is specified, Notch signaling in the foregut endoderm helps to define the area of the pancreas domain.[98] Ectopic expression of a constitutively-active, truncated form of the Notch1 receptor at the onset of pancreas formation blocks both exocrine and endocrine cell differentiation, suggesting that Notch signaling normally may allow the expansion of undifferentiated precursor cells.[99-101] After endocrine specification, constitutive activation of Notch signaling also blocks endocrine cell differentiation in endocrine precursor cells and can change the differentiation potential of some α/β cells away from the endocrine phenotype toward the ductal fate.[102] By contrast, Notch activation in fully matured endocrine cells is not sufficient to change their differentiation status.

Evidence from mouse development suggests that FGF signaling moderates Notch activity during pancreas formation. Ectopic expression in the pancreas epithelium of FGF10, an FGF ligand normally found in pancreatic mesenchyme, activates expression of Notch ligands and receptors and the Notch target gene *HES1*.[103,104] Given the evidence of interactions between embryonic signaling pathways during the formation of other organs, it will be critical to determine how these exchanges regulate pancreas development and pancreas cell differentiation.

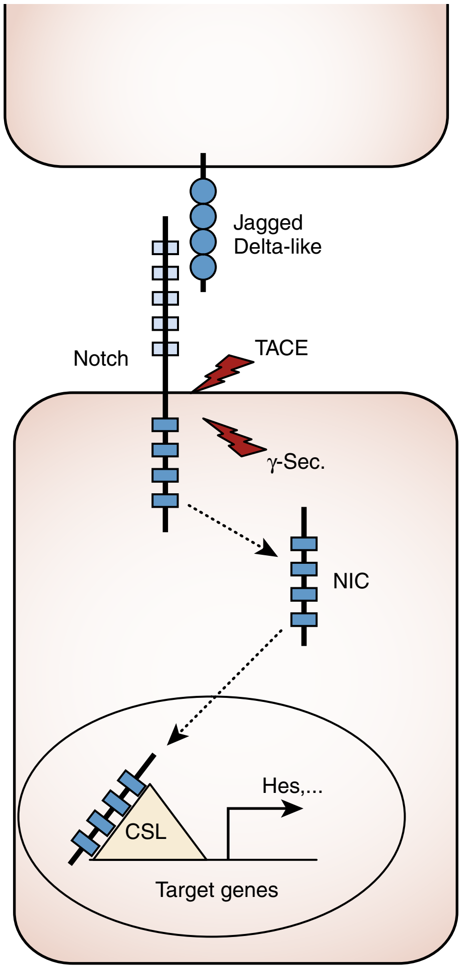


Figure 28.3 **Notch signaling pathway.** Binding of Jagged/Delta-like ligands to Notch receptors activates the Notch signaling cascade. Two proteolytic events mediated by the tumor necrosis factor-α-converting enzyme (TACE) and γ-secretase (γ-Sec.) enzymes result in the generation of an intracellular Notch fragment (NIC) that translocates from the membrane to the nucleus. Interaction with transcription factors from the CSL (CBF1/suppressor of hairless/Lag-1) family activates transcription of Notch target genes, including *HES* genes known to block the transcription of Neurog3.

KEY POINTS

- As the pancreatic buds start to grow, local signaling events direct the initially uniform clusters of epithelial cells to differentiate toward the endocrine and exocrine cell lineages that will eventually populate the mature pancreas.

Transcription Factors

The target of Notch inhibition in the embryonic pancreas, bHLH transcription factor NEUROG3, functions as a proendocrine factor: it is sufficient by itself to drive precursor cells to an endocrine fate.[95,105] When expressed broadly in the developing pancreatic bud, NEUROG3 can force all of the cells in the developing pancreas to differentiate prematurely into islet cells. NEUROG3 is expressed abundantly but transiently in the developing pancreas in scattered ductal cells and occasional periductal cells, but it is never found at similar levels in mature, hormone-producing cells. Together, these data support a model in which NEUROG3 acts upstream of other islet differentiation

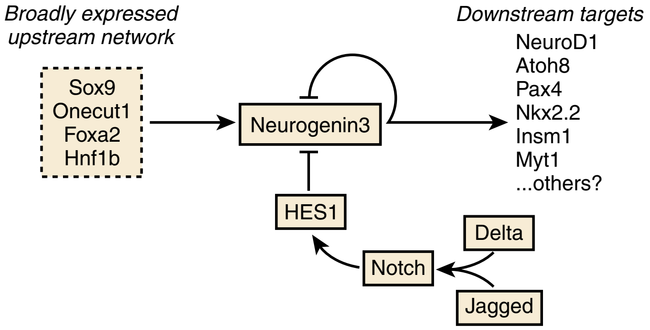


Figure 28.4 **Control of Neurog3 expression.** A model is shown for the positive and negative regulation of Neurog3 expression in pancreatic endocrine progenitor cells.

factors, initiating the differentiation of endocrine cells but switching off prior to final differentiation. This conclusion is supported by the observation that mice homozygous for a targeted disruption of the *Neurog3* gene fail to form any α, β, or δ cells in the pancreas and do not express most of the other islet differentiation factors,[94] as well as by lineage tracing experiments.[106]

Although Notch signaling restricts NEUROG3 to scattered cells within the pancreatic epithelium, additional positive signals are required to initiate NEUROG3 expression in the absence of Notch signaling. Studies of the *NEUROG3* gene promoter implicate several transcription factors in this role, including HNF1, FOXA, ONECUT1, and SOX9,[107-110] which function together as an interacting network[109] (Fig. 28.4). Roles for ONECUT1 and SOX9 have been confirmed *in vivo*.[108,110]

Because NEUROG3 initiates, but does not complete, islet differentiation, it must induce a set of factors that complete the tasks of islet cell differentiation, that is, determination and differentiation of islet cell subtype, formation and organization of islets, and function of the mature islet cells. Several direct targets of NEUROG3 have been identified, including the genes encoding the bHLH factors NEUROD1 and ATOH8, the zinc finger factors MYT1 and INSM1, the paired-homeodomain factor PAX4, and the homeodomain factor NKX2.2.

NEUROD1 expression initiates slightly later than NEUROG3 during pancreatic development; but, unlike NEUROG3, it persists in the mature islet cells, where it plays a role in the expression of a number of differentiated endocrine cell products including insulin.[111-114] NEUROG3 can bind to and activate the *NeuroD1* gene promoter, and it can induce NEUROD1 expression in *Xenopus* embryos[115] and pancreatic ductal cells.[116,117] Conversely, NEUROG3-null embryos completely lack NEUROD1 expression in the pancreas,[94] while NEUROG3 expression is unchanged in NEUROD1-null embryos[105]. Expression of the distantly related bHLH factor ATOH8 and the zinc finger factors MYT1 and INSM1 also is activated by NEUROG3, and ATOH8 and INSM1 appear to play a role in feedback regulation of NEUROG3 expression.[118-120] In addition, MYT1 expression persists in the differentiated endocrine cells and may play a role in their maturation.[121]

Just as it can with the *Neurod1* gene promoter, NEUROG3 binds to and activates the *Pax4* gene promoter[122]; it can activate the intact gene in cultured cells[117,123]; and in its absence PAX4 expression is lost in the pancreas.[94] In addition, studies of the *Pax4* gene promoter implicate the more broadly expressed endoderm factors HNF1B and HNF4A along with NEUROG3 in cooperative activation of PAX4 expression.[122,123] In the case of the gene encoding NKX2.2, NEUROG3 binds to and activates one of three alternate promoters in the gene and cooperates with FOXA transcription factors in activating the gene.[124]

Unlike NEUROD1, ATOH8, and MYT1, however, PAX4 and NKX2.2 expression are not induced in all islet lineages. PAX4, therefore, falls into a class of factors that play specific roles in the differentiation of the different islet cell subtypes. These factors can be grouped into the early group, factors like PAX4 and the NK homeodomain factors NKX2.2[125] and NKX6.1[126] that are coexpressed with NEUROG3 in endocrine progenitor cells,[105] and the late group, genes like the basic leucine zipper transcription factor *MAFA*,[127] *PDX1* and *RFX6* (both in a second phase of their expression), *PAX6*,[128,129] and the paired-related homeodomain transcription factor *ARX*,[130] that are expressed at the final stage of differentiation and are largely restricted to differentiated islet cells (Table 28.1). It is not clear whether any of these factors can actually determine the cell-type fate of the cell in which they are expressed, or if they simply complete the differentiation process initiated by other gene products and signals.

Once NEUROG3 is expressed in an appropriate progenitor, that cell is destined to become an islet cell, but the decision as to which of the four islet cell types it will become apparently is controlled by other factors. Ectopic broad expression of NEUROG3 in the developing pancreatic bud causes uniform and precocious endocrine differentiation, but the type of endocrine cell depends on when NEUROG3 is induced: early expression of NEUROG3 produces α cells,[95,105] but later expression of NEUROG3 produces other cell types.[127] It can be hypothesized, therefore, that the cell-intrinsic competence of the progenitor cells to generate different islet cell types changes over time, or that extracellular signals that change with time guide the differentiation of the NEUROG3-expressing cells into the different islet cell types. The identities of the signals or factors that determine individual islet subtype fates remain to be determined.

KEY POINTS

- The differentiation of exocrine and endocrine cells and their subsequent differentiation into sublineages depend on the action of a cascade of lineage-specific transcription factors.

Islet Formation

Endocrine cell differentiation, including insulin- and glucagon-positive cells, can be detected shortly after the pancreatic buds form, but these cells are immature and do not contribute to the mature islets. Mature β cells form during the secondary transition at approximately embryonic day 13 to 14 in the mouse. At this time endocrine precursor cells leave the endodermal epithelium and migrate through the adjacent extracellular matrix (ECM) into the surrounding mesenchyme.

Cell culture assays have implicated matrix metalloproteinases in this process, although experimental evidence from *in vivo* studies supporting these results is currently missing.[131] Cell attachment to and migration via the ECM is mediated by integrins, heterodimeric transmembrane proteins that serve as extracellular matrix receptors.[132] Functional evidence for an essential role of integrins during migration of islet precursors comes from *in vivo* studies in which islet formation in human fetal pancreas transplanted under the kidney capsule of recipient mice was inhibited by injection of integrin-blocking peptides.[133] The integrin family comprises a large number of different members with distinct substrate specificities, and fetal and mature human β cells express different repertoires of integrins.[134] Changes in integrin composition may explain why fetal β cells are more motile than their adult counterparts, a difference that also may provide insight into the molecular mechanisms that prevent dispersion of endocrine cells once mature islet structures have formed.

Nonetheless, a number of open questions remain with regard to the regulation of islet progenitor migration and islet formation. For example, studies using transgenic animals have shown that a given

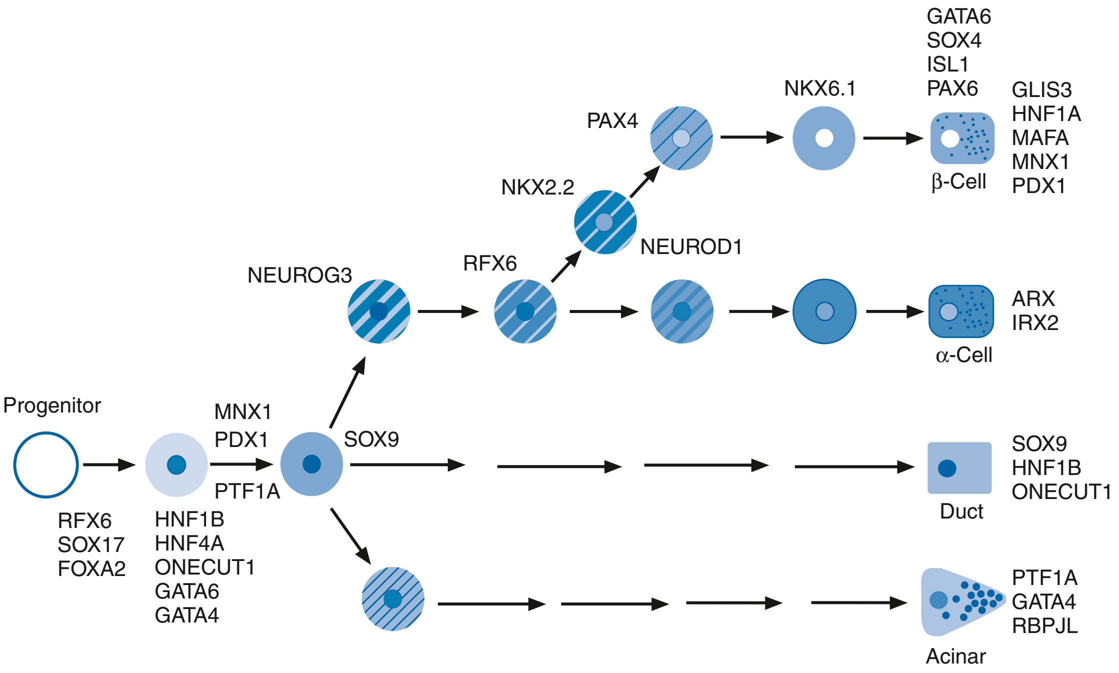


Figure 28.5 A simplified model for the role of islet transcription factors in endocrine differentiation in the developing pancreas. The proposed position for each transcription factor is based on its timing of expression, timing of predominant functional role, or both. Clearly, some factors function at multiple steps, but only key steps are shown for simplicity.

islet originates from several independent endocrine precursors.[135] Thus, endocrine cells delaminating from the ductal structures during embryogenesis must follow a specific guiding mechanism that coordinates their migration path toward a specific location. Although cell migration has been well studied in neural cells that respond to attractive and repulsive guidance cues, a putative, localized source of chemoattractant molecules that stimulate islet precursor migration has not been identified. Furthermore, evidence from studies in human patients with diabetes suggests neogenesis of endocrine cells throughout life.[136] Although still controversial, some endocrine cells appear to form away from already existing islets, raising the question whether these cells migrate toward already formed islets or cluster with other newly formed endocrine cells to initiate aggregation of a new islet. In either case, it is likely that specific, as yet unidentified, guidance mechanisms control these processes.

KEY POINTS

- As cells of the endocrine lineage start to differentiate in the forming pancreas, the new endocrine cells delineate from the forming ducts and begin to form clusters that eventually develop into the islets of Langerhans, which comprise the endocrine compartment of the mature pancreas.

THE ROLE OF PANCREATIC DEVELOPMENT IN HUMAN DIABETES

One beneficial consequence of our improving understanding of pancreatic islet development has been new insights into human diabetes. Short-term energy balance and the control of plasma glucose levels depend on the ability of the β cells to sense changes in nutrient status and respond with appropriate insulin production and secretion. This fine-tuned capability develops along with morphologic development of the islets, and depends on an adequate number of β cells with precisely functioning gene expression programs placed in the appropriate local environment. Any flaws in pancreatic islet development, therefore, can potentially lead to impaired insulin secretion and diabetes.

As a consequence, it is not surprising that mutations in a number of human genes involved in pancreatic islet development, especially genes encoding transcription factors, can lead to diabetes (see Table 28.1). Heterozygous mutations in transcription factors shown in Fig. 28.5 account for five of the first six genes identified that cause maturity onset diabetes of the young,[41,137-141] an autosomal recessive form of diabetes that usually first presents in children and young adults. In addition, mutations in the genes encoding ISL1 and PAX4 have been implicated in families with later-onset diabetes,[141,142] and heterozygous mutations in *GATA4* and *GATA6* and homozygous mutations in *RFX6, MNX1, PDX1, PTF1A, NEUROD1, NKX2.2*, and *PAX6* have been identified in patients with neonatal diabetes.[143-145] Interestingly, although homozygous mutations of the gene encoding NEUROG3 cause neonatal diabetes in mice, humans with homozygous null mutations of the *NEUROG3* gene still generate a reduced number of islet cells and have a later-onset form of diabetes, showing that NEUROG3 is not absolutely required for the islet cell differentiation in humans.[146] Analysis of these mutations has provided insight into the process by which β cell dysfunction may lead to the development of diabetes in humans.

KEY POINTS

- Insights into the process of pancreas and islet development have led to the identification of multiple developmental genes, especially genes encoding transcription factors, that can cause human diabetes when mutated, leading to new insights both into the development of the human pancreas and into the causes of human diabetes.

REFERENCES

1. Tam PP, Kanai-Azuma M, Kanai Y. Early endoderm development in vertebrates: lineage differentiation and morphogenetic function. *Curr Opin Genet Dev*. 2003;13:393–400.
2. Ober EA, Field HA, Stainier DY. From endoderm formation to liver and pancreas development in zebrafish. *Mech Dev*. 2003;120:5–18.
3. Wells JM, Melton DA. Vertebrate endoderm development. *Annu Rev Cell Dev Biol*. 1999;15:393–410.
4. Pictet R, Rutter WJ. Development of the embryonic endocrine pancreas. In: Society AP, Steiner DF, Frenkel N, eds. *Handbook of Physiology*. Washington, DC: Williams and Wilkins; 1972:25–66.
5. Slack JM. Developmental biology of the pancreas. *Development*. 1995;121:1569–1580.
6. Kim SK, Hebrok M, Melton DA. Notochord to endoderm signaling is required for pancreas development. *Development*. 1997;124:4243–4252.
7. Kumar M, Jordan N, Melton D, et al. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev Biol*. 2003;259:109–122.
8. Lammert E, Cleaver O, Melton D. Role of endothelial cells in early pancreas and liver development. *Mech Dev*. 2003;120:59–64.
9. Hill D, Lebenthal E. Congenital abnormalities of the exocrine pancreas. In: Go VLW, Dimagno EP, Gardner JD, et al., eds. *The Pancreas: Biology, Pathobiology, and Disease*. 2nd ed. New York: Raven Press Ltd.; 1993:1029–1040.
10. Golosow N, Grobstein C. Epitheliomesenchymal interaction in pancreatic morphogenesis. *Dev Biol*. 1962;4:242–255.
11. Wessells NK, Cohen JH. Early pancreas organogenesis: morphogenesis, tissue interactions and mass effects. *Dev Biol*. 1967;15:237–270.
12. Bhushan A, Itoh N, Kato S, et al. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development*. 2001;128:5109–5117.
13. Landsman L, Nijagal A, Whitchurch TJ, et al. Pancreatic mesenchyme regulates epithelial organogenesis throughout development. *PLoS Biol*. 2011;9:e1001143.
14. Nekrep N, Wang J, Miyatsuka T, et al. Signals from the neural crest regulate beta-cell mass in the pancreas. *Development*. 2008;135:2151–2160.
15. Pictet RL, Rall LB, Phelps P, et al. The neural crest and the origin of the insulin-producing and other gastrointestinal hormone-producing cells. *Science*. 1976;191:191–192.
16. Fontaine J, Le Douarin NM. Analysis of endoderm formation in the avian blastoderm by the use of quail-chick chimaeras. The problem of the neurectodermal origin of the cells of the apud series. *J Embryol Exp Morphol*. 1977;41:209–222.
17. Le Douarin NM. On the origin of pancreatic endocrine cells. *Cell*. 1988;53:169–171.
18. Wells JM, Melton DA. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development*. 2000;127:1563–1572.
19. Hebrok M, Kim SK, Melton DA. Notochord repression of endodermal sonic hedgehog permits pancreas development. *Genes Dev*. 1998;12:1705–1713.
20. Deutsch G, Jung J, Zheng M, et al. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development*. 2001;128:871–881.
21. Lammert E, Cleaver O, Melton D. Induction of pancreatic differentiation by signals from blood vessels. *Science*. 2001;294:564–567.
22. Yoshitomi H, Zaret KS. Endothelial cell interactions initiate dorsal pancreas development by selectively inducing the transcription factor ptf1a. *Development*. 2004;131:807–817.
23. Dong PD, Munson CA, Norton W, et al. Fgf10 regulates hepatopancreatic ductal system patterning and differentiation. *Nat Genet*. 2007;39:397–402.
24. Pictet RL, Clark WR, Williams RH, et al. An ultrastructural analysis of the developing embryonic pancreas. *Dev Biol*. 1972;29:436–467.
25. Kanai-Azuma M, Kanai Y, Gad JM, et al. Depletion of definitive gut endoderm in sox17-null mutant mice. *Development*. 2002;129:2367–2379.
26. Smith SB, Qu HQ, Taleb N, et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature*. 2010;463:775–780.
27. Ott MO, Rey-Campos J, Cereghini S, et al. Vhnf1 is expressed in epithelial cells of distinct embryonic origin during development and precedes hnf1 expression. *Mech Dev*. 1991;36:47–58.
28. Monaghan AP, Kaestner KH, Grau E, et al. Postimplantation expression patterns indicate a role for the mouse forkhead/hnf-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development*. 1993;119:567–578.
29. Yuan X, Ta TC, Lin M, et al. Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One*. 2009;4:e5609.
30. Duncan SA, Manova K, Chen WS, et al. Expression of transcription factor hnf-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: hnf-4 is a marker for primary endoderm in the implanting blastocyst. *Proc Natl Acad Sci U S A*. 1994;91:7598–7602.
31. Taraviras S, Monaghan AP, Schutz G, et al. Characterization of the mouse hnf-4 gene and its expression during mouse embryogenesis. *Mech Dev*. 1994;48:67–79.
32. Rausa F, Samadani U, Ye H, et al. The cut-homeodomain transcriptional activator hnf-6 is coexpressed with its target gene hnf-3 beta in the developing murine liver and pancreas. *Dev Biol*. 1997;192:228–246.
33. Landry C, Clotman F, Hioki T, et al. Hnf-6 is expressed in endoderm derivatives and nervous system of the mouse embryo and participates to the cross-regulatory network of liver-enriched transcription factors. *Dev Biol*. 1997192:247–257.
34. Decker K, Goldman DC, Grasch CL, et al. Gata6 is an important regulator of mouse pancreas development. *Dev Biol*. 2006;298:415–429.
35. Ketola I, Otonkoski T, Pulkkinen MA, et al. Transcription factor gata-6 is expressed in the endocrine and gata-4 in the exocrine pancreas. *Mol Cell Endocrinol*. 2004;226:51–57.
36. Zaret KS. Molecular genetics of early liver development. *Annu Rev Physiol*. 1996;58:231–251.
37. Cereghini S. Liver-enriched transcription factors and hepatocyte differentiation. *Faseb J*. 1996;10:267–282.
38. Zaret K. Developmental competence of the gut endoderm: Genetic potentiation by gata and hnf3/fork head proteins. *Dev Biol*. 1999;209:1–10.
39. Boj SF, Parrizas M, Maestro MA, et al. A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc Natl Acad Sci U S A*. 2001;98:14481–14486.
40. Shih DQ, Screenan S, Munoz KN, et al. Loss of hnf-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes*. 2001;50:2472–2480.
41. Shih DQ, Stoffel M. Dissecting the transcriptional network of pancreatic islets during development and differentiation. *Proc Natl Acad Sci U S A*. 2001;98:14189–14191.
42. Guz Y, Montminy MR, Stein R, et al. Expression of murine stf-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development*. 1995;121:11–18.
43. Ahlgren U, Jonsson J, Edlund H. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in ipf1/pdx1-deficient mice. *Development*. 1996;122:1409–1416.
44. Offield MF, Jetton TL, Labosky PA, et al. Pdx-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development*. 1996;122:983–995.
45. Kawaguchi Y, Cooper B, Gannon M, et al. The role of the transcriptional regulator ptf1a in converting intestinal to pancreatic progenitors. *Nat Genet*. 2002;32:128–134.
46. Krapp A, Knofler M, Frutiger S, et al. The p48 DNA-binding subunit of transcription factor ptf1 is a new exocrine pancreas-specific basic helix-loop-helix protein. *Embo J*. 1996;15:4317–4329.
47. Li H, Arber S, Jessell TM, et al. Selective agenesis of the dorsal pancreas in mice lacking homeobox gene hlxb9. *Nat Genet*. 1999;23:67–70.
48. Harrison KA, Thaler J, Pfaff SL, et al. Pancreas dorsal lobe agenesis and abnormal islets of langerhans in hlxb9-deficient mice. *Nat Genet*. 1999;23:71–75.
49. Sharma S, Jhala US, Johnson T, et al. Hormonal regulation of an islet-specific enhancer in the pancreatic homeobox gene stf-1. *Mol Cell Biol*. 1997;17:2598–2604.

50. Wu KL, Gannon M, Peshavaria M, et al. Hepatocyte nuclear factor 3beta is involved in pancreatic beta-cell-specific transcription of the pdx-1 gene. *Mol Cell Biol*. 1997;17:6002–6013.
51. Stoffers DA, Heller RS, Miller CP, et al. Developmental expression of the homeodomain protein idx-1 in mice transgenic for an idx-1 promoter/lacz transcriptional reporter. *Endocrinology*. 1999;140:5374–5381.
52. Ben-Shushan E, Marshak S, Shoshkes M, et al. A pancreatic beta -cell-specific enhancer in the human pdx-1 gene is regulated by hepatocyte nuclear factor 3beta (hnf-3beta), hnf-1alpha, and sps transcription factors. *J Biol Chem*. 2001;276:17533–17540.
53. Marshak S, Ben-Shushan E, Shoshkes M, et al. Regulatory elements involved in human pdx-1 gene expression. *Diabetes*. 2001;50:S37–S38.
54. Gerrish K, Cissell MA, Stein R. The role of hepatic nuclear factor 1 alpha and pdx-1 in transcriptional regulation of the pdx-1 gene. *J Biol Chem*. 2001;276:47775–47784.
55. Gerrish K, Gannon M, Shih D, et al. Pancreatic beta cell-specific transcription of the pdx-1 gene. The role of conserved upstream control regions and their hepatic nuclear factor 3beta sites. *J Biol Chem*. 2000;275:3485–3492.
56. Samaras SE, Cissell MA, Gerrish K, et al. Conserved sequences in a tissue-specific regulatory region of the pdx-1 gene mediate transcription in pancreatic beta cells: role for hepatocyte nuclear factor 3beta and pax6. *Mol Cell Biol*. 2002;22:4702–4713.
57. Marshak S, Benshushan E, Shoshkes M, et al. Functional conservation of regulatory elements in the pdx-1 gene: pdx-1 and hepatocyte nuclear factor 3beta transcription factors mediate beta- cell-specific expression. *Mol Cell Biol*. 2000;20:7583–7590.
58. Jacquemin P, Lemaigre FP, Rousseau GG. The onecut transcription factor hnf-6 (oc-1) is required for timely specification of the pancreas and acts upstream of pdx-1 in the specification cascade. *Dev Biol*. 2003;258:105–116.
59. Jonsson J, Carlsson L, Edlund T, et al. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature*. 1994;371:606–609.
60. Stoffers DA, Zinkin NT, Stanojevic V, et al. Pancreatic agenesis attributable to a single nucleotide deletion in the human ipf1 gene coding sequence. *Nat Genet*. 1997;15:106–110.
61. Yee NS, Yusuff S, Pack M. Zebrafish pdx1 morphant displays defects in pancreas development and digestive organ chirality, and potentially identifies a multipotent pancreas progenitor cell. *Genesis*. 2001;30:137–140.
62. Obata J, Yano M, Mimura H, et al. P48 subunit of mouse ptf1 binds to rbp-jkappa/cbf-1, the intracellular mediator of notch signalling, and is expressed in the neural tube of early stage embryos. *Gene Cell*. 2001;6:345–360.
63. Krapp A, Knofler M, Ledermann B, et al. The bhlh protein ptf1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev*. 1998;12:3752–3763.
64. Ye F, Duvillie B, Scharfmann R. Fibroblast growth factors 7 and 10 are expressed in the human embryonic pancreatic mesenchyme and promote the proliferation of embryonic pancreatic epithelial cells. *Diabetologia*. 2005;48:277–281.
65. Le Bras S, Miralles F, Basmaciogullari A, et al. Fibroblast growth factor 2 promotes pancreatic epithelial cell proliferation via functional fibroblast growth factor receptors during embryonic life. *Diabetes*. 1998;47:1236–1242.
66. Miralles F, Czernichow P, Ozaki K, et al. Signaling through fibroblast growth factor receptor 2b plays a key role in the development of the exocrine pancreas. *Proc Natl Acad Sci U S A*. 1999;96:6267–6272.
67. Celli G, LaRochelle WJ, Mackem S, et al. Soluble dominant-negative receptor uncovers essential roles for fibroblast growth factors in multi-organ induction and patterning. *Embo J*. 1998;17:1642–1655.
68. Sanvito F, Herrera PL, Huarte J, et al. Tgf-beta 1 influences the relative development of the exocrine and endocrine pancreas in vitro. *Development*. 1994;120:3451–3462.
69. Miralles F, Czernichow P, Scharfmann R. Follistatin regulates the relative proportions of endocrine versus exocrine tissue during pancreatic development. *Development*. 1998;125:1017–1024.
70. Shiozaki S, Tajima T, Zhang YQ, et al. Impaired differentiation of endocrine and exocrine cells of the pancreas in transgenic mouse expressing the truncated type ii activin receptor. *Biochim Biophys Acta*. 1999;1450:1–11.
71. Yamaoka T, Idehara C, Yano M, et al. Hypoplasia of pancreatic islets in transgenic mice expressing activin receptor mutants. *J Clin Invest*. 1998;102:294–301.
72. Kim SK, Hebrok M, Li E, et al. Activin receptor patterning of foregut organogenesis. *Genes Dev*. 2000;14:1866–1871.
73. Izeradjene K, Combs C, Best M, et al. Kras(g12d) and smad4/dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. *Cancer Cell*. 2007;11:229–243.
74. Bardeesy N, Cheng KH, Berger JH, et al. Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev*. 2006;20:3130–3146.
75. Smart NG, Apelqvist AA, Gu X, et al. Conditional expression of smad7 in pancreatic beta cells disrupts tgf-beta signaling and induces reversible diabetes mellitus. *PLoS Biol*. 2006;4:e39.
76. Apelqvist A, Ahlgren U, Edlund H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr Biol*. 1997;7:801–804.
77. Kawahira H, Ma NH, Tzanakakis ES, et al. Combined activities of hedgehog signaling inhibitors regulate pancreas development, *Development* 130:4871-4879, 2003.
78. Chuang PT, Kawcak T, McMahon AP. Feedback control of mammalian hedgehog signaling by the hedgehog-binding protein, hip1, modulates fgf signaling during branching morphogenesis of the lung. *Genes Dev*. 2003;17:342–347.
79. Lau J, Hebrok M. Hedgehog signaling in pancreas epithelium regulates embryonic organ formation and adult beta-cell function. *Diabetes*. 2010;59:1211–1221.
80. Hebrok M, Kim SK, St Jacques B, et al. Regulation of pancreas development by hedgehog signaling. *Development*. 2000;127:4905–4913.
81. Thomas MK, Lee JH, Rastalsky N, et al. Hedgehog signaling regulation of homeodomain protein islet duodenum homeobox-1 expression in pancreatic beta-cells. *Endocrinology*. 2001;142:1033–1040.
82. Thomas MK, Rastalsky N, Lee JH, et al. Hedgehog signaling regulation of insulin production by pancreatic beta- cells. *Diabetes*. 2000;49:2039–2047.
83. Landsman L, Parent A, Hebrok M. Elevated hedgehog/gli signaling causes beta-cell dedifferentiation in mice. *Proc Natl Acad Sci U S A*. 2011;108:17010–17015.
84. Heller RS, Dichmann DS, Jensen J, et al. Expression patterns of wnts, frizzleds, sfrps, and misexpression in transgenic mice suggesting a role for wnts in pancreas and foregut pattern formation. *Dev Dyn*. 2002;225:260–270.
85. Heller RS, Klein T, Ling Z, et al. Expression of wnt, frizzled, sfrp, and dkk genes in adult human pancreas. *Gene Expr*. 2003;11:141–147.
86. Papadopoulou S, Edlund H. Attenuated wnt signaling perturbs pancreatic growth but not pancreatic function. *Diabetes*. 2005;54:2844–2851.
87. Murtaugh LC, Law AC, Dor Y, et al. Beta-catenin is essential for pancreatic acinar but not islet development, *Development* 132:4663-4674, 2005.
88. Wells JM, Esni F, Boivin GP, et al. Wnt/beta-catenin signaling is required for development of the exocrine pancreas. *BMC Dev Biol*. 2007;7:4.
89. Dessimoz J, Bonnard C, Huelsken J, et al. Pancreas-specific deletion of beta-catenin reveals wnt-dependent and wnt-independent functions during development. *Curr Biol*. 2005;15:1677–1683.
90. Heiser PW, Lau J, Taketo MM, et al. Stabilization of beta-catenin impacts pancreas growth. *Development*. 2006;133:2023–2032.
91. Jonckheere N, Mayes E, Shih HP, et al. Analysis of mpygo2 mutant mice suggests a requirement for mesenchymal wnt signaling in pancreatic growth and differentiation. *Dev Biol*. 2008;318:224–235.
92. Rulifson IC, Karnik SK, Heiser PW, et al. Wnt signaling regulates pancreatic beta cell proliferation. *Proc Natl Acad Sci U S A*. 2007;104:6247–6252.
93. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science*. 1999;284:770–776.
94. Gradwohl G, Dierich A, LeMeur M, et al. Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc Natl Acad Sci U S A*. 2000;97:1607–1611.
95. Apelqvist A, Li H, Sommer L, et al. Notch signaling controls pancreatic cell differentiation. *Nature*. 1999;400:877–881.

96. Jensen J, Pedersen EE, Galante P, et al. Control of endodermal endocrine development by hes-1. *Nat Genet*. 2000;24:36–44.
97. Shih HP, Kopp JL, Sandhu M, et al. A notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. *Development*. 2012;139:2488–2499.
98. Fukuda A, Kawaguchi Y, Furuyama K, et al. Ectopic pancreas formation in hes1 -knockout mice reveals plasticity of endodermal progenitors of the gut, bile duct, and pancreas. *J Clin Invest*. 2006;116:1484–1493.
99. Hald J, Hjorth JP, German MS, et al. Activated notch1 prevents differentiation of pancreatic acinar cells and attenuate endocrine development. *Dev Biol*. 2003;260:426–437.
100. Murtaugh LC, Stanger BZ, Kwan KM, et al. Notch signaling controls multiple steps of pancreatic differentiation. *Proc Natl Acad Sci U S A*. 2003;100:14920–14925.
101. Esni F, Ghosh B, Biankin AV, et al. Notch inhibits ptf1 function and acinar cell differentiation in developing mouse and zebrafish pancreas. *Development*. 2004;131:4213–4224.
102. Greenwood AL, Li S, Jones K, et al. Notch signaling reveals developmental plasticity of pax4(+) pancreatic endocrine progenitors and shunts them to a duct fate. *Mech Dev*. 2007;124:97–107.
103. Hart A, Papadopoulou S, Edlund H. Fgf10 maintains notch activation, stimulates proliferation, and blocks differentiation of pancreatic epithelial cells. *Dev Dyn*. 2003;228:185–193.
104. Norgaard GA, Jensen JN, Jensen J. Fgf10 signaling maintains the pancreatic progenitor cell state revealing a novel role of notch in organ development. *Dev Biol*. 2003;264:323–338.
105. Schwitzgebel VM, Scheel DW, Conners JR, et al. Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development*. 2000;127:3533–3542.
106. Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: ngn3+ cells are islet progenitors and are distinct from duct progenitors. *Development*. 2002;129:2447–2457.
107. Lee JC, Smith SB, Watada H, et al. Regulation of the pancreatic proendocrine gene neurogenin3. *Diabetes*. 2001;50:928–936.
108. Jacquemin P, Durviaux SM, Jensen J, et al. Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3. *Mol Cell Biol*. 2000;20:4445–4454.
109. Lynn FC, Smith SB, Wilson ME, et al. Sox9 coordinates a transcriptional network in pancreatic progenitor cells. *Proc Natl Acad Sci U S A*. 2007;104:10500–10505.
110. Seymour PA, Freude KK, Dubois CL, et al. A dosage-dependent requirement for sox9 in pancreatic endocrine cell formation. *Dev Biol*. 2008;323:19–30.
111. Ohneda K, Mirmira RG, Wang J, et al. The homeodomain of pdx-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter. *Mol Cell Biol*. 2000;20:900–911.
112. Glick E, Leshkowitz D, Walker MD. Transcription factor beta2 acts cooperatively with e2a and pdx1 to activate the insulin gene promoter. *J Biol Chem*. 2000;275:2199–2204.
113. Qiu Y, Guo M, Huang S, et al. Insulin gene transcription is mediated by interactions between the p300 coactivator and pdx-1, beta2, and e47. *Mol Cell Biol*. 2002;22:412–420.
114. Naya FJ, Stellrecht CM, Tsai MJ. Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes Dev*. 1995;9:1009–1019.
115. Huang HP, Liu M, El-Hodiri HM, et al. Regulation of the pancreatic islet-specific gene beta2 (neurod) by neurogenin 3. *Mol Cell Biol*. 2000;20:3292–3307.
116. Heremans Y, Van De Casteele M, in't Veld P, et al. Recapitulation of embryonic neuroendocrine differentiation in adult human pancreatic duct cells expressing neurogenin 3. *J Cell Biol*. 2002;159:303–312.
117. Gasa R, Mrejen C, Leachman N, et al. Proendocrine genes coordinate the pancreatic islet differentiation program in vitro. *Proc Natl Acad Sci U S A*. 2004;101:13245–13250.
118. Wang S, Hecksher-Sorensen J, Xu Y, et al. Myt1 and ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. *Dev Biol*. 2008;317:531–540.
119. Lynn FC, Sanchez L, Gomis R, et al. Identification of the bhlh factor math6 as a novel component of the embryonic pancreas transcriptional network. *PLoS One*. 2008;3:e2430.
120. Mellitzer G, Bonne S, Luco RF, et al. Ia1 is ngn3-dependent and essential for differentiation of the endocrine pancreas. *Embo J*. 2006;25:1344–1352.
121. Wang S, Zhang J, Zhao A, et al. Loss of myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. *Mech Dev*. 2007;124:898–910.
122. Smith SB, Watada H, Scheel DW, et al. Autoregulation and maturity onset diabetes of the young transcription factors control the human pax4 promoter. *J Biol Chem*. 2000;275:36910–36919.
123. Smith SB, Gasa R, Watada H, et al. Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of pax4. *J Biol Chem*. 2003;278:38254–38259.
124. Watada H, Scheel DW, Leung J, et al. Distinct gene expression programs function in progenitor and mature islet cells. *J Biol Chem*. 2003;278:17130–17140.
125. Sussel L, Kalamaras J, Hartigan-O'Connor DJ, et al. Mice lacking the homeodomain transcription factor nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. *Development*. 1998;125:2213–2221.
126. Sander M, Sussel L, Conners J, et al. Homeobox gene nkx6.1 lies downstream of nkx2.2 in the major pathway of beta-cell formation in the pancreas. *Development*. 2000;127:5533–5540.
127. Ahlgren U, Pfaff SL, Jessell TM, et al. Independent requirement for isl1 in formation of pancreatic mesenchyme and islet cells. *Nature*. 1997;385:257–260.
128. Sander M, Neubuser A, Kalamaras J, et al. Genetic analysis reveals that pax6 is required for normal transcription of pancreatic hormone genes and islet development. *Genes Dev*. 1997;11:1662–1673.
129. St-Onge L, Sosa-Pineda B, Chowdhury K, et al. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature*. 1997;387:406–409.
130. Collombat P, Mansouri A, Hecksher-Sorensen J, et al. Opposing actions of arx and pax4 in endocrine pancreas development. *Genes Dev*. 2003;17:2591–2603.
131. Miralles F, Battelino T, Czernichow P, et al. Tgf-beta plays a key role in morphogenesis of the pancreatic islets of langerhans by controlling the activity of the matrix metalloproteinase mmp-2. *J Cell Biol*. 1998;143:827–836.
132. Hynes R. Integrins. Bidirectional, allosteric signaling machines. *Cell*. 2002;110:673.
133. Cirulli V, Beattie GM, Klier G, et al. Expression and function of alpha(v) beta(3) and alpha(v)beta(5) integrins in the developing pancreas. Roles in the adhesion and migration of putative endocrine progenitor cells. *J Cell Biol*. 2000;150:1445–1460.
134. Kaido T, Perez B, Yebra M, et al. Alphav-integrin utilization in human beta-cell adhesion, spreading, and motility. *J Biol Chem*. 2004;279:17731–17737.
135. Deltour L, Leduque P, Paldi A, et al. Polyclonal origin of pancreatic islets in aggregation mouse chimaeras, *Development* 112:1115-1121, 1991.
136. Butler AE, Janson J, Bonner-Weir S, et al. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes*. 2003;52:102–110.
137. Yamagata K, Furuta H, Oda N, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (mody1). *Nature*. 1996;384:458–460.
138. Yamagata K, Oda N, Kaisaki P, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (mody3). *Nature*. 1996;384:455–458.
139. Stoffers DA, Ferrer J, Clarke WL, et al. Early-onset type-ii diabetes mellitus (mody4) linked to ipf1 [letter]. *Nat Genet*. 1997;17:138–139.
140. Horikawa Y, Iwasaki N, Hara M, et al. Mutation in hepatocyte nuclear factor-1 beta gene (tcf2) associated with mody [letter]. *Nat Genet*. 1997;17:384–385.
141. Malecki MT, Jhala US, Antonellis A, et al. Mutations in neurod1 are associated with the development of type 2 diabetes mellitus. *Nat Genet*. 1999;23:323–328.

142. Shimomura H, Sanke T, Hanabusa T, et al. Nonsense mutation of islet-1 gene (q310x) found in a type 2 diabetic patient with a strong family history. *Diabetes*. 2000;49:1597–1600.
143. Schwitzgebel VM. Many faces of monogenic diabetes. *J Diabetes Investig*. 2014;5:121–133.
144. Flanagan SE, De Franco E, Lango AH, et al. Analysis of transcription factors key for mouse pancreatic development establishes nkx2-2 and mnx1 mutations as causes of neonatal diabetes in man. *Cell Metabol*. 2014;19:146–154.
145. Naylor RN, Greeley SAW, Bell GI, et al. Genetics and pathophysiology of neonatal diabetes mellitus. *J Diabetes Investig*. 2011;2:158–169.
146. Solorzano-Vargas RS, Bjerknes M, Wang J, et al. Null mutations of neurog3 are associated with delayed-onset diabetes mellitus. *JCI Insight*. 2020;5.
147. Jensen J, Heller RS, Funder-Nielsen T, et al. Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. *Diabetes*. 2000;49:163–176.
148. Wang J, Cortina G, Wu SV, et al. Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N Engl J Med*. 2006;355:270–280.
149. Lee JE, Hollenberg SM, Snider L, et al. Conversion of xenopus ectoderm into neurons by neurod, a basic helix-loop-helix protein. *Science*. 1995;268:836–844.
150. Naya FJ, Huang HP, Qiu Y, et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in beta2/neurod-deficient mice. *Genes Dev*. 1997;11:2323–2334.
151. Mutoh H, Naya FJ, Tsai MJ, et al. The basic helix-loop-helix protein beta2 interacts with p300 to coordinate differentiation of secretin-expressing enteroendocrine cells. *Genes Dev*. 1998;12:820–830.
152. Qiu Y, Sharma A, Stein R. P300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. *Mol Cell Biol*. 1998;18:2957–2964.
153. Rubio-Cabezas O, Minton JA, Kantor I, et al. Homozygous mutations in neurod1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities. *Diabetes*. 2010;59:2326–2331.
154. Sellick GS, Barker KT, Stolte-Dijkstra I, et al. Mutations in ptf1a cause pancreatic and cerebellar agenesis. *Nat Genet*. 2004;36:1301–1305.
155. Waeber G, Thompson N, Nicod P, et al. Transcriptional activation of the glut2 gene by the ipf-1/stf-1/idx-1 homeobox factor. *Mol Endocrinol*. 1996;10:1327–1334.
156. Ahlgren U, Jonsson J, Jonsson L, et al. Beta-cell-specific inactivation of the mouse ipf1/pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev*. 1998;12:1763–1768.
157. Watada H, Kajimoto Y, Umayahara Y, et al. The human glucokinase gene beta-cell-type promoter: an essential role of insulin promoter factor 1/pdx-1 in its activation in hit-t15 cells. *Diabetes*. 1996;45:1478–1488.
158. Watada H, Kajimoto Y, Miyagawa J, et al. Pdx-1 induces insulin and glucokinase gene expressions in alphatc1 clone 6 cells in the presence of betacellulin. *Diabetes*. 1996;45:1826–1831.
159. Watada H, Kajimoto Y, Kaneto H, et al. Involvement of the homeodomain-containing transcription factor pdx-1 in islet amyloid polypeptide gene transcription. *Biochem Biophys Res Commun*. 1996;229:746–751.
160. Schwartz PT, Perez-Villamil B, Rivera A, et al. Pancreatic homeodomain transcription factor idx1/ipf1 expressed in developing brain regulates somatostatin gene transcription in embryonic neural cells. *J Biol Chem*. 2000;275:19106–19114.
161. Perez-Villamil B, Schwartz PT, Vallejo M. The pancreatic homeodomain transcription factor idx1/ipf1 is expressed in neural cells during brain development. *Endocrinology*. 1999;140:3857–3860.
162. Harrison KA, Druey KM, Deguchi Y, et al. A novel human homeobox gene distantly related to proboscipedia is expressed in lymphoid and pancreatic tissues. *J Biol Chem*. 1994;269:19968–19975.
163. Smith SB, Ee HC, Conners JR, et al. Paired-homeodomain transcription factor pax4 acts as a transcriptional repressor in early pancreatic development. *Mol Cell Biol*. 1999;19:8272–8280.
164. Sosa-Pineda B, Chowdhury K, Torres M, et al. The pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature*. 1997;386:399–402.
165. Shimajiri Y, Sanke T, Furuta H, et al. A missense mutation of pax4 gene (r121w) is associated with type 2 diabetes in Japanese. *Diabetes*. 2001;50:2864–2869.
166. Hill ME, Asa SL, Drucker DJ. Essential requirement for pax6 in control of enteroendocrine proglucagon gene transcription. *Mol Endocrinol*. 1999;13:1474–1486.
167. Solomon BD, Pineda-Alvarez DE, Balog JZ, et al. Compound heterozygosity for mutations in pax6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia. *Am J Med Genet*. 2009;149A:2543–2546.
168. Bienvenu T, Poirier K, Friocourt G, et al. Arx, a novel prd-class-homeobox gene highly expressed in the telencephalon, is mutated in x-linked mental retardation. *Hum Mol Genet*. 2002;11:981–991.
169. Rudnick A, Ling TY, Odagiri H, et al. Pancreatic beta cells express a diverse set of homeobox genes. *Proc Natl Acad Sci U S A*. 1994;91:12203–12207.
170. Watada H, Mirmira RG, Leung J, et al. Transcriptional and translational regulation of beta-cell differentiation factor nkx6.1. *J Biol Chem*. 2000;275:34224–34230.
171. Jensen J, Serup P, Karlsen C, et al. Mrna profiling of rat islet tumors reveals nkx 6.1 as a beta-cell-specific homeodomain transcription factor. *J Biol Chem*. 1996;271:18749–18758.
172. Vallejo M, Penchuk L, Habener JF. Somatostatin gene upstream enhancer element activated by a protein complex consisting of creb, isl-1-like, and alpha-cbf-like transcription factors. *J Biol Chem*. 1992;267:12876–12884.
173. Leonard J, Serup P, Gonzalez G, et al. The lim family transcription factor isl-1 requires camp response element binding protein to promote somatostatin expression in pancreatic islet cells. *Proc Natl Acad Sci USA*. 1992;89:6247–6251.
174. Thor S, Ericson J, Brannstrom T, et al. The homeodomain lim protein isl-i is expressed in subsets of neurons and endocrine cells in the adult rat. *Neuron*. 1991;7:1–9.
175. Dong J, Asa SL, Drucker DJ. Islet cell and extrapancreatic expression of the lim domain homeobox gene isl-1. *Mol Endocrinol*. 1991;5:1633–1641.
176. Karlsson O, Thor S, Norberg T, et al. Insulin gene enhancer binding protein isl-1 is a member of a novel class of proteins containing both a homeo- and a cys-his domain. *Nature*. 1990;344:879–882.
177. Emens LA, Landers DW, Moss LG. Hepatocyte nuclear factor 1a is expressed in a hamster insulinoma line and transactivates the rat insulin i gene. *Proc Natl Acad Sci USA*. 1992;89:7300–7304.
178. Noguchi T, Yamada K, Yamagata K, et al. Expression of liver type pyruvate kinase in insulinoma cells: involvement of lf-b1 (hnf1). *Biochem Biophys Res Commun*. 1991;181:259–264.
179. Pontoglio M, Sreenan S, Roe M, et al. Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice. *J Clin Invest*. 1998;101:2215–2222.
180. Coffinier C, Thepot D, Babinet C, et al. Essential role for the homeoprotein vhnf1/hnf1beta in visceral endoderm differentiation. *Development*. 1999;126:4785–4794.
181. Coffinier C, Barra J, Babinet C, et al. Expression of the vhnf1/hnf-1beta homeoprotein gene during mouse organogenesis. *Mech Dev*. 1999;89:211–213.
182. Lemaigre FP, Durviaux SM, Truong O, et al. Hepatocyte nuclear factor 6, a transcription factor that contains a novel type of homeodomain and a single cut domain. *Proc Natl Acad Sci U S A*. 1996;93:9460–9464.
183. Rausa FM, Ye H, Lim L, et al. In situ hybridization with 33p-labeled rna probes for determination of cellular expression patterns of liver transcription factors in mouse embryos [published erratum appears in *Methods* 1998;16:359–360]. *Methods*. 1998;16:29–41.
184. Villamayor L, Cano DA, Rojas A. Gata factors in pancreas development and disease. *IUBMB Life*. 2020;72:80–88.
185. Sladek FM, Zhong WM, Lai E, et al. Liver-enriched transcription factor hnf-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev*. 1990;4:2353–2365.
186. Miquerol L, Lopez S, Cartier N, et al. Expression of the l-type pyruvate kinase gene and the hepatocyte nuclear factor 4 transcription factor in exocrine and endocrine pancreas. *J Biol Chem*. 1994;269:8944–8951.

187. Duncan SA, Navas MA, Dufort D, et al. Regulation of a transcription factor network required for differentiation and metabolism. *Science*. 1998;281:692–695.
188. Planque N, Leconte L, Coquelle FM, et al. Interaction of maf transcription factors with pax-6 results in synergistic activation of the glucagon promoter. *J Biol Chem*. 2001;276:35751–35760.
189. Olbrot M, Rud J, Moss LG, et al. Identification of beta-cell-specific insulin gene transcription factor ripe3b1 as mammalian mafa. *Proc Natl Acad Sci U S A*. 2002;99:6737–6742.
190. Kataoka K, Han SI, Shioda S, et al. Mafa is a glucose-regulated and pancreatic beta-cell-specific transcriptional activator for the insulin gene. *J Biol Chem*. 2002;277:49903–49910.
191. Zhang C, Moriguchi T, Kajihara M, et al. Mafa is a key regulator of glucose-stimulated insulin secretion. *Mol Cell Biol*. 2005;25:4969–4976.
192. Vanhoose AM, Samaras S, Artner I, et al. Mafa and mafb regulate pdx1 transcription through the area ii control region in pancreatic beta cells. *J Biol Chem*. 2008;283:22612–22619.
193. Nishimura W, Rowan S, Salameh T, et al. Preferential reduction of beta cells derived from pax6-mafb pathway in mafb deficient mice. *Dev Biol*. 2008;314:443–456.
194. Artner I, Blanchi B, Raum JC, et al. Mafb is required for islet beta cell maturation. *Proc Natl Acad Sci U S A*. 2007;104:3853–3858.
195. Artner I, Le Lay J, Hang Y, et al. Mafb: an activator of the glucagon gene expressed in developing islet alpha- and beta-cells. *Diabetes*. 2006;55:297–304.

29

Regulation of Islet Hormone Synthesis and Secretion

Adrian Vella and R. Paul Robertson

OUTLINE

Introduction, 487
Hormonal Synthesis, 487
  Insulin, 487
  Glucagon, 488
Hormonal Secretion and Action, 489
  Insulin, 489
  Glucagon, 491
  Glucagon-Like Peptide-1 Secretion and Signaling, 494

INTRODUCTION

Glucose concentrations are tightly regulated in the fasting and postprandial state by the actions of two hormones with opposing actions. Insulin secretion is stimulated by rising glucose concentrations and suppressed by low glucose concentrations. Signaling through its cognate receptor, it stimulates glucose, fat, and amino acid uptake by insulin-sensitive organs while suppressing endogenous glucose production, release of free fatty acids, and protein catabolism.

On the other hand, glucagon secretion is stimulated by hypoglycemia and results in stimulation of endogenous glucose production. As such it is a key counterregulatory hormone. Abnormalities in the secretion and signaling of both hormones play important roles in the pathogenesis of type 2 diabetes. Understanding their regulation, synthesis, and secretion, as well as their interactions, is key to understanding diabetes.

In addition, there are other novel aspects that may be important in health and disease. Glucagon is an important insulin secretagogue and may help potentiate insulin secretion *in vivo*. More importantly, this signaling may occur through a receptor for glucagon-like peptide (GLP)-1 present on β cells. GLP-1 is a product of proglucagon processing that also may have actions of relevance to the regulation of metabolism. These topics will be considered in turn in this chapter.

HORMONAL SYNTHESIS

Insulin

Insulin is synthesized in the β cell as preproinsulin, a polypeptide comprising (in sequential order) a signal peptide, the B domain, a C domain flanked by dibasic cleavage sites, and the A domain (Fig. 29.1). During and after translation, preproinsulin is transported to the endoplasmic reticulum, where the signal peptidase cleaves the signal peptide, transforming preproinsulin to proinsulin. Proinsulin then folds, forming three highly conserved disulfide bonds into a heterodimer linked by the C domain, which is cleaved off during the transition from endoplasmic reticulum to the Golgi complex and secretory granules. The mature secretory granules contain insulin and C-peptide, which are secreted in a 1:1 ratio.

The proteolytic cleavages required for the conversion of proinsulin to insulin are catalyzed by the prohormone convertases (PCs)—both PC-1/3 and PC-2 are required, although PC-2 preferentially cleaves the C-peptide/A chain junction, while PC-1/3 cleaves the B chain/C-peptide junction.[1] Work by Rhodes et al. has suggested that PC-2 follows the action of PC-1/3 and facilitates formation of mature insulin granules.[2] The accumulation and secretion of proinsulin in diabetes may reflect acquired defects in these enzymes, or a relative deficiency, although their expression rises in the presence of hyperglycemia.[3] At present there is no evidence that common genetic variation in the prohormone convertases is associated with predisposition to diabetes.[4]

Mutations that affect protein folding (for example loss of the cysteine residues forming the disulfide bonds B7-A7, B19-A20, and A6-A11) lead to the accumulation of misfolded insulin and resultant endoplasmic reticulum stress, which ultimately lead to β cell dysfunction and loss. Common genetic variation in the enzymes mediating insulin synthesis or regulating the environment within the endoplasmic reticulum will also increase misfolding and predispose to secretory dysfunction.[5-7] Zinc accumulates within the β cell, specifically in the insulin secretory granule, where it may stabilize insulin into hexamers. A genetic variant in the zinc transporter is associated with type 2 diabetes.[8]

Insulin secretion is triggered by glucose concentrations above a threshold, but synthesis occurs at lower concentrations. This implies that even at normal physiologic concentrations, insulin stores are being replenished. Insulin synthesis occupies 10% to 50% of the protein synthetic activity of the β cell. Consequently, the complex sequence of transcription, translation, and posttranslational modification is very susceptible to stressors such as caloric excess or impaired insulin signaling.[9,10] Pancreas/duodenum homeobox protein 1 (PDX-1) and V-maf musculoaponeurotic fibrosarcoma oncogene homolog A (MafA) are two of the most prominently studied regulators of insulin gene expression, both of which are essential for normal insulin gene expression.[11,12] Both of these transcription factors are in activated by hyperglycemia, but they both can be deactivated when hyperglycemia is chronic and become integral parts of the manifestations of glucose toxicity and accompanying oxidative stress (see Chapter 38). The mechanism for loss of PDX-1 is posttranscriptional,[13,14] whereas the loss of MafA is posttranslational.[15,16]

β cell stress (*in vitro*),[5] type 2 diabetes,[17] and impaired glucose tolerance increase proinsulin concentrations in the circulation, together with an accompanying decrease in insulin secretion. The proinsulin/insulin ratio in the peripheral circulation has been proposed as a marker of β cell health,[18,19] but this has significant limitations:

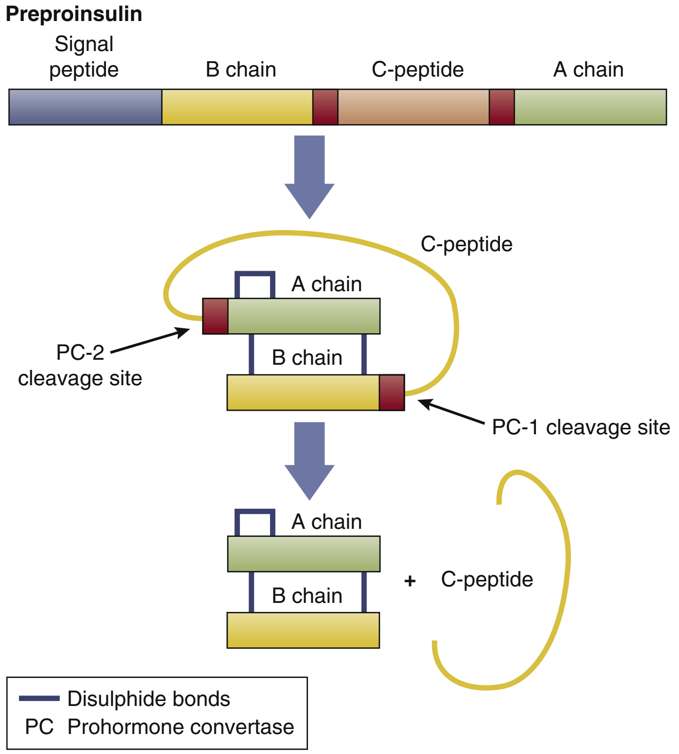

Preproinsulin
Signal peptide
B chain
C-peptide
A chain
C-peptide
A chain
PC-2 cleavage site
B chain
PC-1 cleavage site
A chain
B chain
+
C-peptide
Disulphide bonds
PC Prohormone convertase

Fig. 29.1 Insulin Synthesis from Preproinsulin. (Modified from Sun J, Cui J, He Q. Proinsulin misfolding and endoplasmic reticulum stress during the development and progression of diabetes. *Mol Aspects Med*. 2015;42:105–118.)

proinsulin has a longer half-life (20–30 min)[20] than insulin (~5 min) and, unlike insulin, is not extracted by the liver.[21,22] This limitation can be overcome only by direct and simultaneous measurement of insulin and proinsulin secretion.[19,23] It remains to be ascertained if increased proinsulin concentrations in the peripheral circulation reflect endoplasmic reticulum stress and can be used to predict future decline in β cell function and glucose tolerance.

KEY POINTS

- Insulin is synthesized as preproinsulin, a polypeptide that, after translation, is processed by the endoplasmic reticulum to the mature folded protein. This process is subject to misfolding and under conditions of high demand may produce endoplasmic reticulum stress and secretory dysfunction.

Glucagon

Glucagon is a product of the *GCG* (glucagon) gene, which is strongly conserved across vertebrate species and is expressed in pancreatic α cells, in intestinal enteroendocrine cells, and in some neurons (chiefly within the brainstem). The gene is transcribed and translated to proglucagon, which is then cleaved by the two prohormone convertases, described above, that are differentially expressed in specific tissues (Fig. 29.2). PC-1/3 is mainly expressed in intestinal enteroendocrine cells and cleaves proglucagon to form oxyntomodulin, GLP-1, and GLP-2, amongst other peptide products. On the other hand, PC-2 is expressed in the α cell, and the predominant product of proglucagon cleavage is glucagon.[24]

The selective expression of PC-1/3 and PC-2 may not be absolute, and there are circumstances where PC-2 may be expressed outside of the islet, leading to intestinal synthesis and secretion of glucagon. In multiple animal models exposed to stressors such as overfeeding, there is evidence of inducible expression of PC-1/3 within the islet and subsequent intraislet secretion of GLP-1. For example, interleukin 6,[25] hyperglycemia,[26] and free fatty acids[27] seem to increase PC-1/3 expression in islets. Whether this is of any relevance to humans in health and disease remains unclear, but will be discussed.

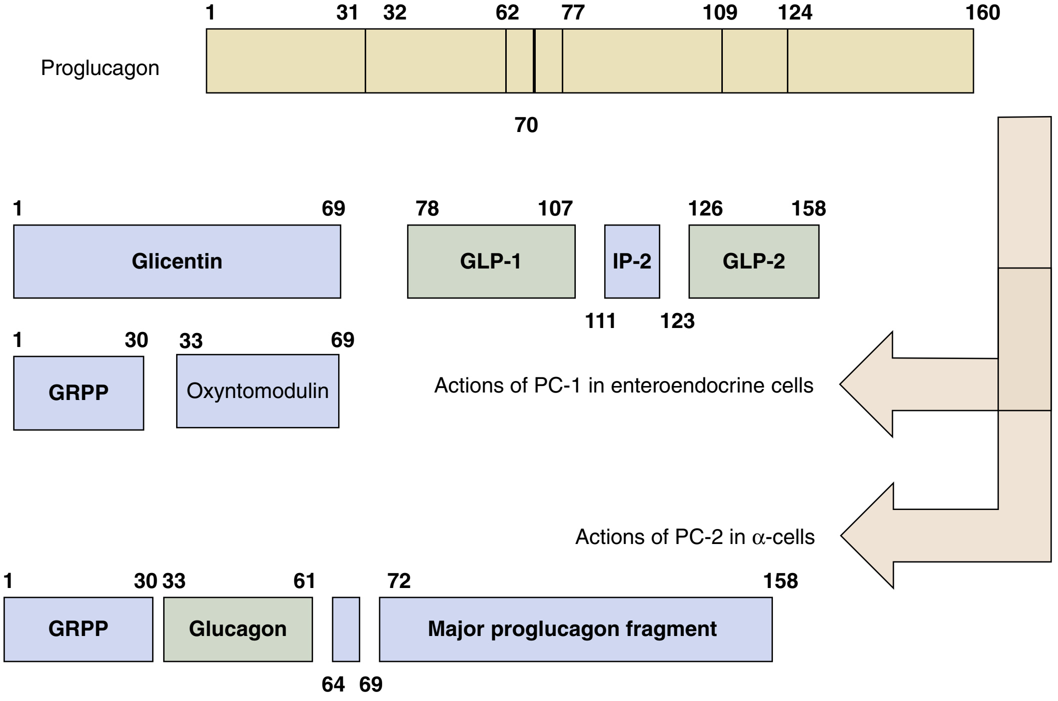

1
31
32
62
77
109
124
160
Proglucagon
70
1
69
78
107
126
158
Glicentin
GLP-1
IP-2
GLP-2
111
123
1
30
33
69
GRPP
Oxyntomodulin
Actions of PC-1 in enteroendocrine cells
Actions of PC-2 in α-cells
1
30 33
61
72
158
GRPP
Glucagon
Major proglucagon fragment
64
69

Fig. 29.2 Posttranslational Fate of Proglucagon. *GLP,* Glucagon-like peptide; *PC,* prohormone convertase; *IP,* Intervening Peptide; *GRPP,* Glicentin-related polypeptide. (Modified from Holst, J. J., et al. "Proglucagon processing in porcine and human pancreas." *J Biol Chem* 269(29): 18827–18833. Copyright © 1994 American Society for Biochemistry and Molecular Biology. Published by Elsevier Inc. on behalf of American Society for Biochemistry and Molecular Biology. All rights reserved.)

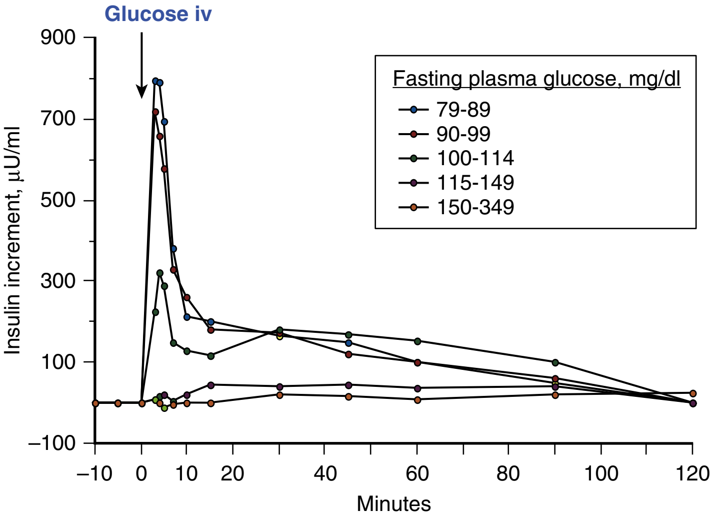


Fig. 29.3 **First and Second Phases of Insulin Secretion.** Insulin secretion, most especially the first phase, is attenuated as fasting glucose increases. *iv,* Intravenous. (Modified from Brunzell JD, Robertson RP, Lerner RL. Relationships between fasting plasma glucose levels and insulin secretion during intravenous glucose tolerance tests. *J Clin Endocrinol Metab.* 1976;42:222–229.)

Glucagon Synthesis in the Enteroendocrine Cell. The suggestion that extrapancreatic synthesis and secretion of glucagon might occur is not novel,[28] with multiple conflicting reports in the literature.[29] Definitive conclusions have been hampered by the sensitivity and specificity of the assays used to measure glucagon. PC-2 may be expressed in some enteroendocrine cells, and Lund et al. demonstrated the presence of intact glucagon in the sera of patients who had undergone total pancreatectomy.[30] In this instance glucagon was measured using both a mass spectrometry–based method as well as a sandwich enzyme-linked immunosorbent assay targeting both the N- and C-terminal regions of glucagon. Whether this intestinal production of glucagon is significant in humans who exhibit "paradoxical" postprandial rises in glucagon after intestinal surgery and retain their pancreas[31] remains to be ascertained.

KEY POINTS

- Glucagon arises from the posttranslational processing of proglucagon by prohormone convertase enzymes. The selective expression of prohormone convertase-2 vs. prohormone convertase-1/3 determines the final cleavage product of proglucagon.

HORMONAL SECRETION AND ACTION

Insulin

Stimulators. When an intravenous glucose challenge is used to stimulate β cell function, insulin consisting of preformed and docked insulin granules is released into the circulation and peaks 3 to 5 minutes after the injection. This is referred to as the acute insulin response or first phase of insulin secretion (Fig. 29.3).[32-34] The second phase of insulin secretion occurs over the next hour (depending on whether stimulation is transient[25] or sustained[26]). Other agents used to stimulate insulin secretion include glucagon,[35] secretin,[36] and the beta adrenergic stimulator isoproterenol.[37,38] Probably the commonest function test utilized is the oral glucose tolerance test—usually the 75-g glucose test, which has been used to characterize glucose tolerance status in prediabetic individuals (Fig. 29.4). With appropriate sampling frequency, this test can be used to reliably measure insulin secretion.[39] Mixed meal tests also use an oral challenge containing carbohydrate, fat, and protein. However, these are less standardized and may differ in terms of whether a liquid or a solid challenge is used. Their advantage is that they can be more easily utilized for tracer-based measurements of glucose metabolism.[40]

Somewhat counterintuitively, insulin concentrations should not be used to measure insulin secretion *per se*. This is because insulin concentrations represent the end result of insulin secretion by the β cells into the portal circulation, where variable hepatic extraction of insulin occurs prior to insulin reaching the systemic circulation.[41] Whether hepatic extraction is a passive or active process remains to be ascertained.[42] However, the magnitude of extraction seems to be related to insulin secretion, so that in health as much as 50% of insulin in the portal circulation does not reach the systemic circulation. To date, with one exception,[43] there is no clear evidence that differences in hepatic extraction can contribute to how individuals respond to a glucose challenge. Given the confounding effects of variable hepatic insulin extraction on measures of β cell function using insulin concentrations, C-peptide (which does not undergo hepatic extraction and is secreted in a 1:1 ratio with insulin) has been utilized. Because C-peptide has a longer half-life (~35 min) than insulin and will therefore accumulate in the circulation, knowledge of its kinetics will enable deconvolution of the insulin secretion rate. The use of a population model for C-peptide kinetics has been validated and avoids the need for individually determining C-peptide clearance in subjects with intact renal function.[44]

A research tool termed *glucose potentiation of arginine-induced insulin secretion* uses insulin responses to both glucose and arginine to estimate insulin secretory reserve.[45] This consists of measuring a baseline insulin response to an intravenous injection of 5 g arginine that is followed by infusion of intravenous glucose for 60 minutes and then with a repeat of the arginine injection (Fig. 29.5). The calculation of *insulin secretory reserve* is made by subtracting the magnitude of the insulin response to the first arginine injection from the response to the second arginine injection. The validation of this tool lies in the close correlation of insulin secretory reserve data to the known quantity of pancreatic islets transplanted into humans,[46] thus allowing this metabolic test to be used as a functional measure of beta cell mass (see Chapter 43).

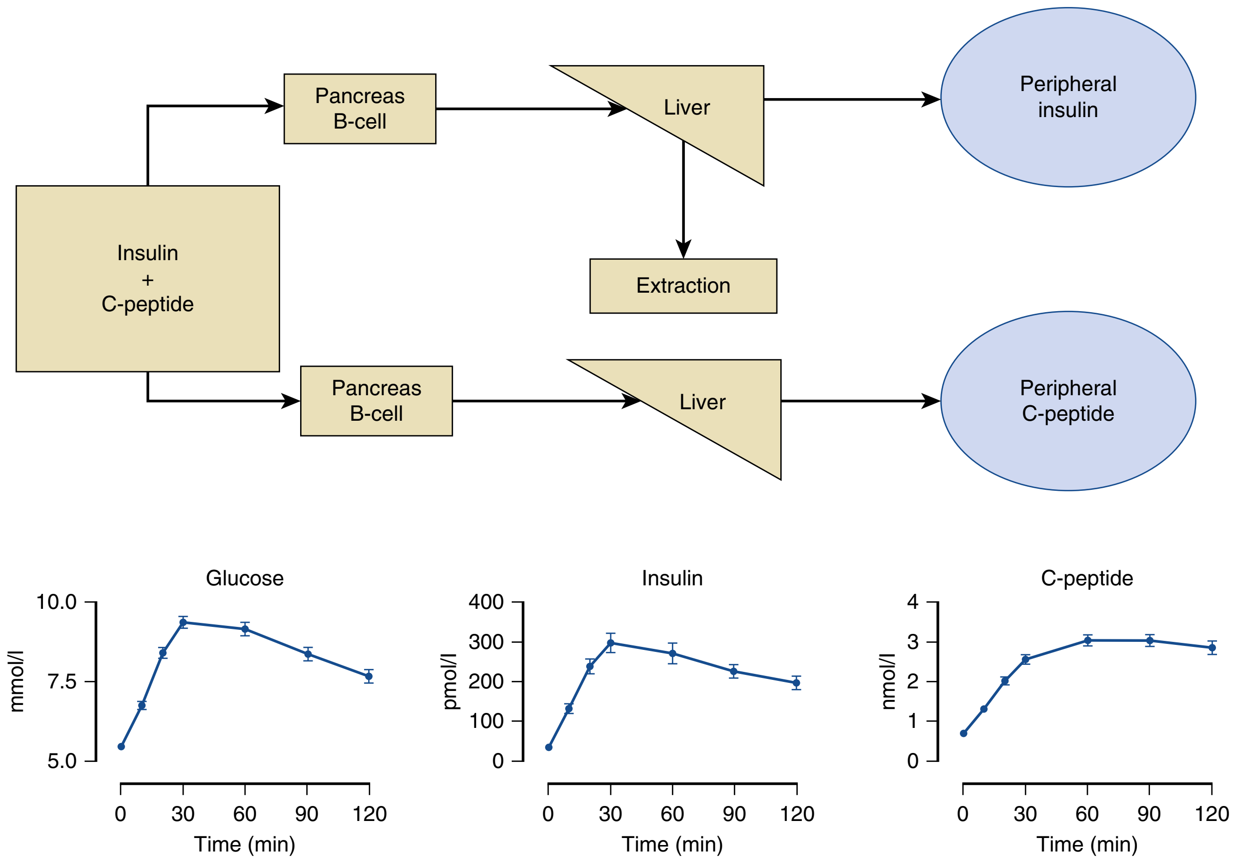


Fig. 29.4 Use of an Oral Glucose Tolerance Test to Measure Insulin Secretion and Action. (Modified from Cobelli C, Dalla Man C, Toffolo G. The oral minimal model method. *Diabetes*. 2014;63:1203–1213.)

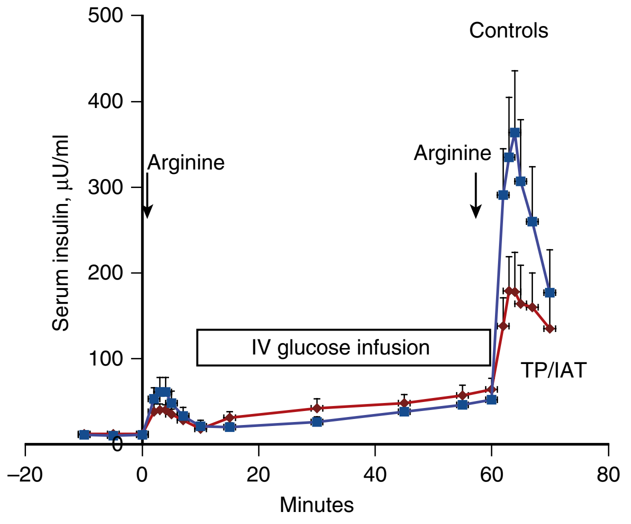


Fig. 29.5 Insulin responses to arginine alone and in the presence of an intravenous glucose infusion in total pancreatectomy and autoislet transplantation (TP/IAT) recipients and healthy control subjects. *IV*, Intravenous. (Modified from Robertson RP, Bogachus LD, Oseid E. Assessment of beta-Cell Mass and alpha- and beta-Cell Survival and Function by Arginine Stimulation in Human Autologous Islet Recipients. *Diabetes*. 2015;64:565–572.)

Mechanisms regulating β cell secretion. High glucose concentrations enter β cells via glucokinase activity[47] and elicit electrical activity by closing adenosine triphosphate (ATP)-sensitive potassium (K^+) channels (Fig. 29.6). These channels consist of four pore-forming subunits (Kir6.2, encoded by *KCNJ11*) and four regulatory subunits (the sulfonylurea receptor SUR1, encoded by *ABCC8*). Intracellular ATP binds Kir6.2 to cause channel closure and depolarization. Drugs that bind to SUR1 can impair insulin secretion (e.g., diazoxide) by opening the channel, while drugs that cause channel closure (e.g., sulfonylureas such as glipizide) stimulate insulin secretion.[9] Gain-of-function mutations in *KCNJ11* or in *ABCC8* will cause neonatal diabetes, while loss-of-function mutations result in congenital hyperinsulinism. Common genetic variation in *KCNJ11* is associated with risk of type 2 diabetes.[48,49]

Membrane depolarization triggers the activation of voltage-gated calcium (Ca^{2+}) channels, and the subsequent influx of Ca^{2+} initiates Ca^{2+}-dependent exocytosis of insulin granules.[9] When exposed to glucose concentrations in the physiologic range, the electrical activity of β cells is characterized by a pulsatile pattern of secretion in which the pulse interval decreases and the pulse amplitude increases as glucose concentrations rise.[50] A rise in Ca^{2+} concentrations through L-type channels drives closure of Ca^{2+}-activated K^+ channels. ATP consumption by Ca^{2+}-ATPases partially reopen K^+ channels for repolarization, which causes cessation of insulin secretion.[51]

It is important to note that the response to a change in glucose concentrations occurs on a background of pulsatile insulin secretion with isolated oscillations having a periodicity of around 5 minutes. They are likely driven by fluctuations in intracellular calcium concentrations and metabolism.[52] Glycolytic oscillations can themselves drive fluctuations in the intracellular ATP/adenosine diphosphate (ADP) ratio, and change in calcium concentrations *per se* can also drive changes in metabolism and energy generation.

Ideally, insulin secretion is expressed as a function of the prevailing insulin action. This requires the use of a minimal model that uses the relationship between changing glucose and C-peptide concentrations and the relationship of changing glucose and insulin concentrations

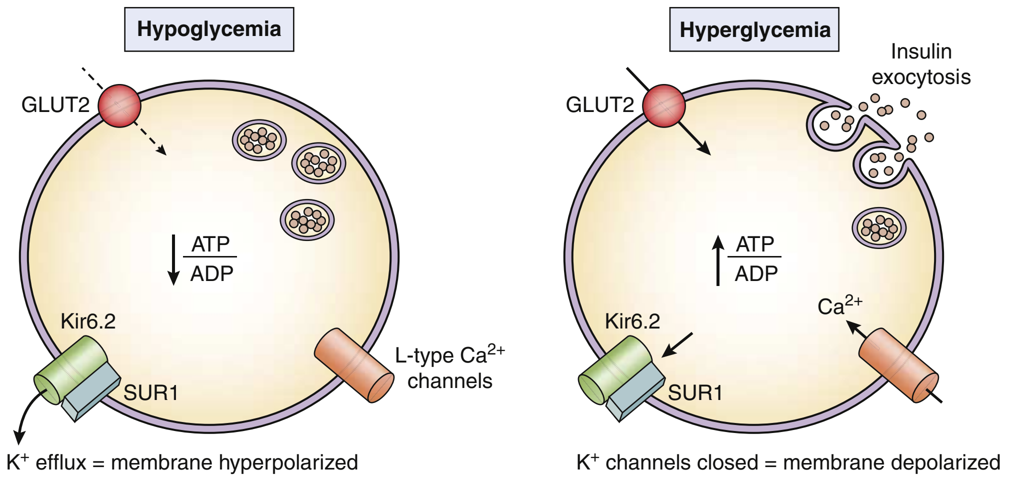

Hypoglycemia
GLUT2
$\downarrow \frac{ATP}{ADP}$
Kir6.2
SUR1
L-type Ca^{2+} channels
K^+ efflux = membrane hyperpolarized
Hyperglycemia
Insulin exocytosis
GLUT2
$\uparrow \frac{ATP}{ADP}$
Ca^{2+}
Kir6.2
SUR1
K^+ channels closed = membrane depolarized

Fig. 29.6 **The Potassium Channel and its Central Role in Insulin Secretion.** Insulin secretion is inhibited at low glucose concentrations *(left panel)* and stimulated at high glucose levels *(right panel)*. At 5 mM glucose, K_{ATP} channels are active because of the low intracellular ATP/ADP ratio, which keeps the membrane hyperpolarized and prevents electrical activity. At 10 mM glucose, via stimulation of glucose metabolism, the intracellular ATP/ADP ratio increases, resulting in a reduced K^+ current. When K^+ channel activity is sufficiently low, the membrane depolarizes, initiating electrical activity and Ca^{2+} influx. Elevation of intracellular Ca^{2+} activates ATP-consuming Ca^{2+} pumps, causing a fall in the ATP/ADP ratio, reactivation of K^+ channels, membrane repolarization, and temporary termination of electrical activity. Elevated intracellular Ca^{2+} also opens Ca^{2+}-activated K^+ channels, which contributes to the membrane repolarization (not shown). In the hyperpolarized state, Ca^{2+} influx ceases, intracellular Ca^{2+} returns to basal levels, ATP consumption falls, the ATP/ADP ratio is restored, K^+ channels close, and electrical activity is reinitiated. *ATP,* Adenosine triphosphate; *ADP,* adenosine diphosphate; *Ca^{2+}*, calcium; *K^+*, potassium. (Modified from Ashcroft FM, Rorsman P. K(ATP) channels and islet hormone secretion: new insights and controversies. *Nat Rev Endocrinol.* 2013;9:660–669.)

to calculate these indices. The product—the disposition index—represents the hyperbolic relationship between insulin secretion and action and can be used to measure response to intervention or weight loss or to predict incident diabetes.[53]

Insulin signaling. See Chapter 33.

KEY POINTS

- Insulin is secreted in a 1:1 ratio with C-peptide into the portal circulation. However, insulin undergoes hepatic extraction so that the insulin appearing in the systemic circulation represents the sum of two processes: insulin secretion and hepatic extraction. Because of this, the rate of appearance of C-peptide (which does not undergo hepatic extraction) in the systemic circulation is a more reliable measure of insulin secretion. The primary regulator of insulin secretion is glucose, but other substances such as amino acids and beta adrenergic stimulation can also stimulate insulin release. Glucose also potentiates insulin secretory responses to arginine and glucagon.

Inhibitors. Stimulation of insulin secretion is negatively modulated by the actions of a group of three substances whose function is mediated by negative regulatory G proteins: somatostatin, epinephrine, and prostaglandin E (Figs. 29.7, 29.8, and 29.9). Somatostatin is produced by delta cells within the islet and thought to have the paracrine effect of negatively modulating glucose-induced insulin release.[54] Epinephrine and norepinephrine have a similar negatively modulatory effects whether they are released from islet sympathetic nerve terminals (norepinephrine) or arrive to the islet via the general circulation from the adrenal medulla (epinephrine).[55] It has been well documented that, during clinical stress states such as surgical procedures and extensive body burns, adrenergic inputs to the islet diminish insulin secretion, stimulating hepatic glucose production and thereby causing hyperglycemia. Prostaglandin E is synthesized by the islet via the arachidonic acid cascade and has specific binding sites on islet plasma membranes. It is ideally located to negatively modulate glucose-induced insulin secretion.[56-58] Sodium salicylate, an inhibitor of cyclooxygenase and prostaglandin E synthesis, has been shown to augment glucose-induced insulin secretion *in vitro* and *in vivo* in diabetic humans (Fig. 29.10).[58]

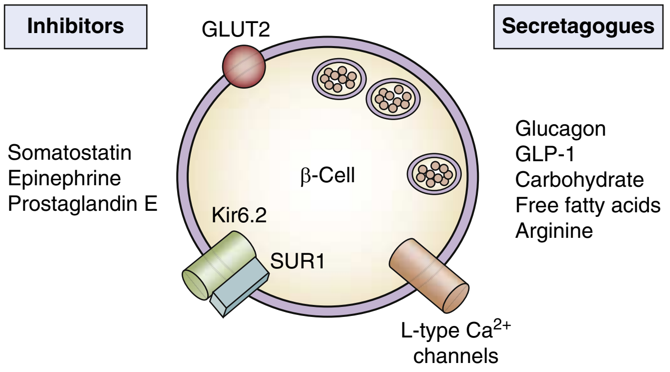


Fig. 29.7 **Regulation of Insulin Secretion.** *GLP,* Glucagon-like peptide; *Ca^{2+}*, calcium.

Glucagon

Stimulators: Glucagon provides the first and most important line of defense against hypoglycemia. When blood glucose levels reach the 65 to 68 mg/dL range,[59] usually because of exogenous insulin injections or purposefully during insulin tolerance tests or hypoglycemic clamps, glucagon is immediately released from α cells into the hepatic portal system (Figs. 29.11 and 29.12). The mechanism of the glucagon release is related to a switching off the physiological suppressive effect that insulin and/or zinc bound to insulin has on glucagon secretion.[60-62] When glucagon reaches the hepatocytes, it stimulates glycogenolysis, and glucose is released into the systemic circulation to correct systemic

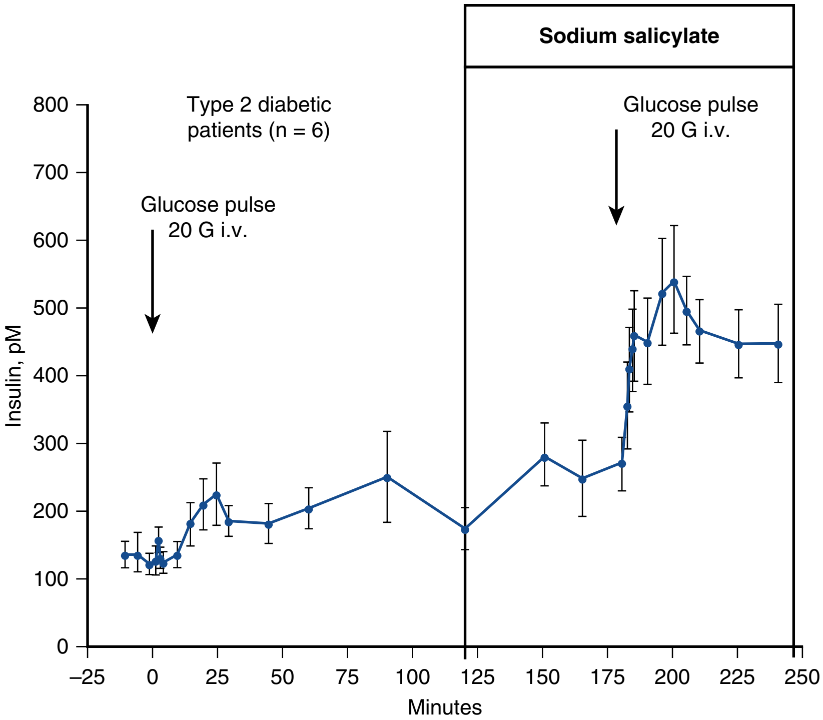


Fig. 29.8 Inhibition of prostaglandin synthesis by salicylate increases the secretion of insulin in response to an intravenous glucose challenge in subjects with type 2 diabetes. (Modified from Robertson RP, Chen M. A role for prostaglandin E in defective insulin secretion and carbohydrate intolerance in diabetes mellitus. *J Clin Invest*. 1977;60:747–753.)

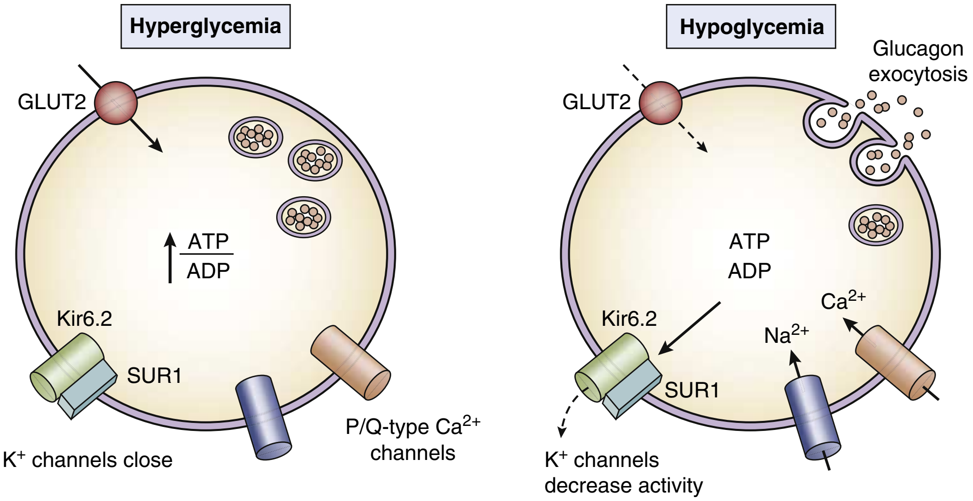


Fig. 29.9 Unlike the β cell, glucagon secretion is stimulated at low glucose levels *(left panel)* and inhibited *(right panel)* by hyperglycemia. Potassium (K^+) channels are under strong tonic inhibition (99%) at low glucose levels, so that the membrane is depolarized sufficiently to elicit electrical activity, which consists of large-amplitude action potentials. Electrical activity is associated with activation of voltage-gated sodium (Na^+) channels and voltage-gated calcium (Ca^{2+}) channels, and Ca^{2+} influx through the latter triggers glucagon secretion. Increasing glucose levels to 6 mM closes K^+ channels completely, further depolarizing the membrane. (Modified from Ashcroft FM, Rorsman P. K(ATP) channels and islet hormone secretion: new insights and controversies. *Nat Rev Endocrinol.* 2013;9:660–669.)

hypoglycemia. Other substances that stimulate glucagon release include amino acids, epinephrine, and secretin.[10]

KEY POINTS

- Key inhibitory substances for insulin release are mediated mechanistically through negative regulatory G proteins that diminish glucose-induced insulin secretion. These include somatostatin, epinephrine, and prostaglandin E.

Glucagon is stored in granules prior to exocytosis and the secretory machinery is similar to that found in β cells. However, exocytosis is slightly faster than in β cells because of the higher number of granules per cell in the readily releasable pool. The rate of granule refilling after depletion of this pool is also higher in α cells, perhaps due to larger Ca^{2+} currents in relatively smaller cells.[9] Glucagon secretion exhibits a somewhat reciprocal relation to insulin secretion and has been characterized as a counterregulatory hormone protecting

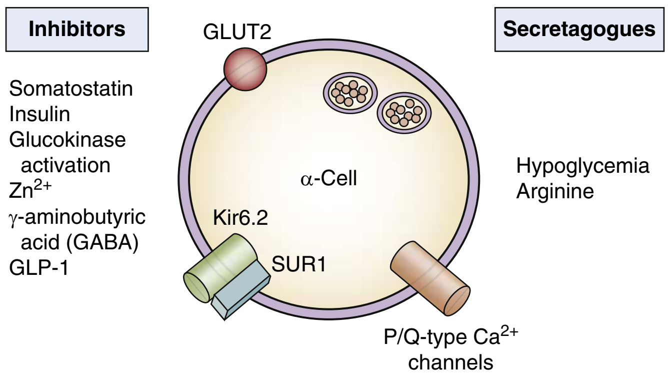


Fig. 29.10 Regulation of glucagon secretion.

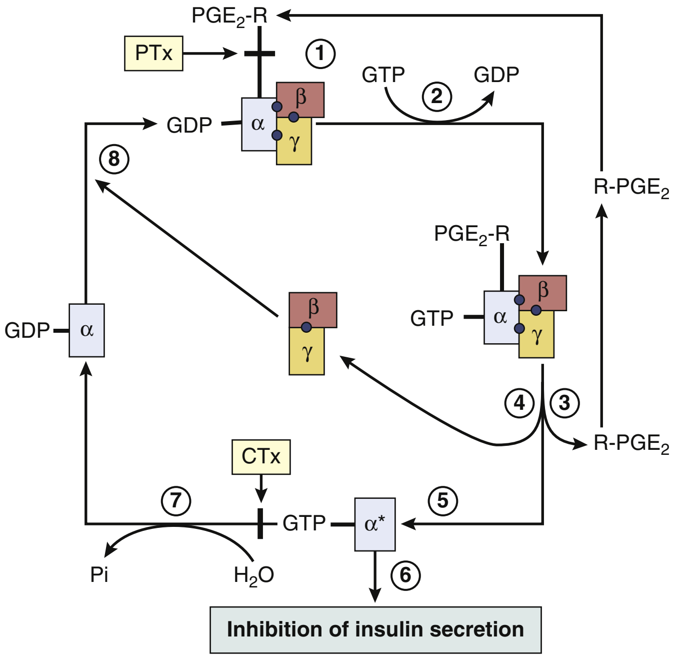


Fig. 29.11 Insulin Secretion is Modulated by G Proteins Within the β Cell. After formation of hormone–receptor complex, this associates with guanosine diphosphate (GDP)-binding G protein (comprised of α-, β-, and γ-subunits). This union promotes exchange of guanosine triphosphate (GTP) for GDP on the α-subunit (step 2). This forms an unstable complex that quickly dissociates into free HR complex (step 3), free β-γ-subunits (step 4), and activated α-subunit binding GTP (step 5). This activated species modulates islet effector systems, e.g., adenylylcyclase, ion channels, phospholipases, and distal sites in exocytosis (step 6). Activated species are inactivated by intrinsic GTPase activity of the α-subunit, which hydrolyzes bound GTP to GDP (step 7) and allows reassociation of α-GDP with the β-γ-subunit (step 8). This completes the cycle and returns G protein to its quiescent state. (Modified from Robertson RP, Seaquist ER, Walseth TF. G proteins and modulation of insulin secretion. *Diabetes.* 1991;40:1–6.)

against hypoglycemia. It is stimulated by low systemic glucose concentrations, enters the islet's periportal circulation, and travels to the liver via the portal vein, where it acutely mobilizes hepatic glycogen (glycogenolysis) to restore normoglycemia. In the presence of hyperglycemia, glucagon secretion is suppressed. However, beyond counteracting insulin-induced hypoglycemia, glucagon is also secreted tonically in the fasting state and consequently plays an important role in metabolic regulation during both the fasting and postprandial state.

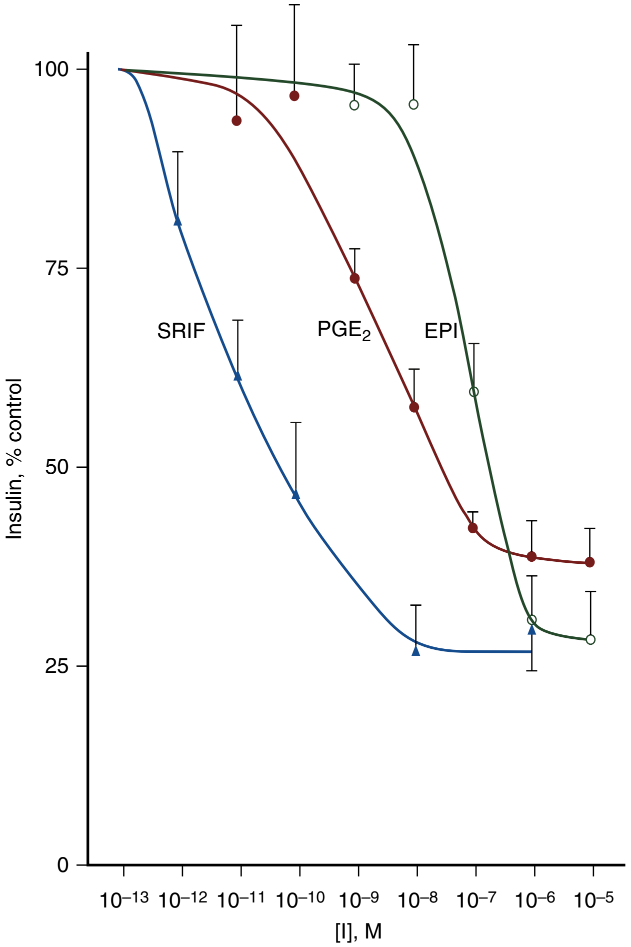


Fig. 29.12 Inhibition of insulin secretion during static 60-min incubations with 11.1 mM glucose (KRP buffer, pH 7.4) by increasing concentrations of somatostatin (SRIF), PGE2, and epinephrine (EPI) from four, five, and six separate experiments, respectively, performed in triplicate. PGE2 (IC^ 1 × 10”’ M) was more potent than epinephrine but less so than somatostatin. (Modified from Robertson RP, Tsai P, Little SA. Receptor-mediated adenylate cyclase-coupled mechanism for PGE2 inhibition of insulin secretion in HIT cells. *Diabetes.* 1987;36:1047–1053.)

KEY POINTS

- The secretion and action of insulin and glucagon in the regulation of glucose concentrations is reciprocal. The most important stimulator of glucagon secretion is hypoglycemia. At low glucose concentrations, β cells are hyperpolarized because of their open adenosine triphosphate (ATP)-sensitive potassium (K^+) channels. In contrast, α cells have a relatively elevated ATP/adenosine diphosphate ratio resulting in closure of most, but not all, ATP-sensitive K^+ channels, leading to depolarization and subsequent glucagon exocytosis. Secreted glucagon travels directly to the liver to stimulate glycogenolysis for release of glucose into the systemic circulation to correct hypoglycemia. The presence of hyperglycemia results in a further ATP increase within α cells. This results in closure of all K^+ channels. Key negative modulators of α cell function are insulin, glucose, and somatostatin.

Glucagon Signaling. Glucagon signals through a G protein–coupled receptor with seven transmembrane helices.[63] Hormone-receptor binding activates $G_{\alpha s}$-coupled proteins which increases cyclic adenosine monophosphate (cAMP) and activates protein kinase A (PKA) and cAMP response element–binding (CREB) protein. PKA activates phosphorylase kinase which, in turn, activates phosphorylase, resulting in mobilization of glycogen to glucose-1-phosphate while inhibiting glycogen synthase. Other actions of PKA impair glycolysis and increase gluconeogenesis. CREB directly induces transcription of glucose-6-phosphatase and phosphoenolpyruvate carboxykinase, which increases gluconeogenesis.[64]

The initial response to glucagon (increased release of glucose into the bloodstream from the liver) is the result of glycogenolysis rather than gluconeogenesis, because the latter requires gene transcription. The effects on glycogenolysis require phosphorylation, which can occur in minutes. Gluconeogenesis is substrate-driven, which requires activation of other mechanisms such as those mediated by cortisol or epinephrine.[65,66] Glucagon is most effective at increasing endogenous glucose production acutely. In situations where hepatic glycogen is depleted, as for example at the end of a 72-hour fast, no glucose response to glucagon is expected.[67]

Glucagon stimulates hepatic amino acid metabolism but does not directly promote amino acid release from skeletal muscle. To date there is no evidence of glucagon receptors on skeletal muscle in humans.[64] The role of glucagon in lipid metabolism is unclear.[64] Human adipocytes do not appear to express the glucagon receptor. It is possible that, by binding its hepatic receptor and activating CREB, it sets in chain a train of events that result in increased fatty acid catabolism. PKA activation also triggers a cascade of metabolic changes that produce increasing β-oxidation of free fatty acids rather than reesterification to triglycerides.[68]

Glucagon is an insulin secretagogue[45,69] in the concentrations found within the islet[70] during α cell dysfunction, as occurs in type 2 diabetes[71,72] and in prediabetes.[73,74] In these conditions, glucagon signals through the GLP-1 receptor present on β cells,[75] but its net contribution to physiologic insulin secretion is unknown.

Intrinsic Mechanisms Regulating α Cell Secretion. There are significant similarities between β cells and α cells with regards to their secretory apparatus. Glucagon secretory granule exocytosis is triggered by an influx of Ca^{2+} through voltage-dependent Ca^{2+} channels. An increase in extracellular glucose will lead to a decrease in intracellular Ca^{2+}. While α cells also express ATP-sensitive K^+ channels, voltage-dependent K^+ channels, and voltage-dependent Ca^{2+} channels, the isoforms of the latter group differ so that there is a greater contribution of P/Q channels in α cells compared with β cells.[76] Tetrodotoxin inhibits glucagon release, implying that voltage-dependent Na^+ channels also play a role in glucagon secretion.[10]

At low glucose concentrations, β cells are hyperpolarized because of their open ATP-sensitive K^+ channels (Fig. 29.6). In contrast, α cells have a relatively elevated ATP/ADP ratio, resulting in closure of most, but not all, ATP-sensitive K^+ channels, leading to depolarization and subsequent glucagon exocytosis (Fig. 29.11). The presence of hyperglycemia results in a further ATP increase within α cells. This results in closure of all K^+ channels. Taken together, these findings imply that small changes in K^+ channel conductance could regulate the glucagon secretory response to changing glucose concentrations.[9] This is supported by the observation that mice lacking functional K^+ channels exhibit impaired glucagon secretion during hypoglycemia.[77]

AMP-activated protein kinase (AMPK) may act as the sensing mechanism for the metabolic state that drives glucagon secretion. Hyperglycemia, which is associated with inhibition of glucagon secretion, inhibits AMPK activity. In addition, direct activation of AMPK stimulates glucagon secretion.[78]

Extrinsic Mechanisms Regulating α-Cell Secretion. Glucokinase, an enzyme that is expressed in α cells, may play a role in glucose-mediated regulation of glucagon secretion.[79] In addition to direct cellular responses to changes in glucose concentrations, factors extrinsic to the α cell also regulate glucagon secretion. Insulin reduces K^+ channel sensitivity to ATP, thereby restraining glucagon secretion. Studies examining the pulsatile secretion of insulin have suggested that, at least in healthy individuals, glucagon pulses are antiphasic to insulin pulses.[80] Type 2 diabetes is clearly a bihormonal disease characterized by delayed and decreased insulin secretion coupled with impaired postprandial suppression of glucagon.[81] The latter, a manifestation of α cell dysfunction, was thought to occur at a relatively late stage in the evolution of diabetes, perhaps because of disrupted insulin secretion. However, cross-sectional studies in people with prediabetes have suggested that defects in α cell secretion do not correlate with indices of insulin secretion. Indeed, impaired insulin signaling is a better marker of abnormal α cell regulation.[73,82] This may suggest that, beyond a minimum threshold of insulin secretion, it is the efficacy of insulin signaling that contributes to individual variation in glucagon secretion.[83] In addition, the association of genetic variants that predispose to type 2 diabetes with impaired glucagon suppression imply that these defects occur early in the evolution of diabetes.[84,85]

The β cell also secretes other inhibitors of glucagon secretion including zinc and γ-aminobutyric acid. Delta cell–secreted somatostatin is likely to inhibit glucagon secretion by the α cell. This assumes greater importance in insulin-deficient states. It has been demonstrated that antagonism of the somatostatin receptor increases α cell responses to arginine. Somatostatin receptor antagonists do not demonstrate significant effects on the response of glucagon to insulin-induced hypoglycemia in nondiabetic rodents. This is not the case in diabetic animals, where somatostatin receptor antagonists restore the glucagon responses to hypoglycemia.[76]

KEY POINTS

- Glucagon secretion is regulated by electrophysiological mechanisms intrinsic to the α cell. In addition, glucokinase expression in the α cell confers direct glucose sensing. The β cell also regulates glucagon secretion via paracrine effects of insulin and other inhibitors of glucagon secretion, including zinc and γ-aminobutyric acid. The delta cell, through the secretion of somatostatin, also inhibits glucagon secretion by the α cell.

Measurement of α-Cell Function *In Vivo*. The peptide sequence of glucagon overlaps considerably with that of other proglucagon-derived peptides. Many commercially available immunoassays are subject to significant crossreactivity with peptides such as oxyntomodulin and glicentin and may be insufficiently sensitive to adequately measure glucagon suppression by glucose. This issue may be a particular problem in patients with end-stage renal disease, where crossreacting glucagon fragments accumulate.[86,87] Currently, there are no specific tests of α cell function, and usually glucagon is measured as part of standardized challenges to measure β cell function.[88] A notable exception to this is glucagon-induced β cell secretion, where glucagon is used as a secretagogue. Of note, arginine-stimulation testing, which produces β cell stimulation similar to that produced by glucagon (with a lower incidence of nausea), potently stimulates glucagon secretion.[45] Postchallenge suppression of glucagon depends to a great extent on the nature of the caloric challenge, so that protein-containing meals can stimulate rather than suppress glucagon secretion.[65]

Glucagon-Like Peptide-1 Secretion and Signaling

As discussed previously, posttranslational processing of proglucagon in enteroendocrine cells yields a highly conserved peptide hormone, GLP-1.[89,90] GLP-1 signals through its cognate receptor, which shares its architecture with the glucagon receptor. Although GLP-1–based therapy is now used to treat type 2 diabetes and obesity,[91] important questions about its physiological role remain.[92] It has a $t_{1/2}$ of approximately 1 minute in the circulation due to rapid degradation by a ubiquitous peptidase, di-peptidyl peptidase-4.[93] A significant proportion of endogenous GLP-1 is degraded close to the site of production.[94] Competitive antagonism of the GLP-1 receptor with exendin-9,39 has little effect on postprandial insulin secretion, unless postprandial GLP-1 concentrations are markedly elevated.[31] These conditions are typically only fulfilled after Roux-en-Y gastric bypass[95] and, to a lesser extent, sleeve gastrectomy.[96] On the other hand, when infused at pharmacological concentrations, GLP-1 stimulates insulin secretion and suppresses glucagon secretion in a glucose-dependent manner.[97] These effects are accompanied by delayed gastric emptying of solids and an increase in gastric accommodation with accompanying changes in compliance (increased) and wall tension (decreased).[98,99] Many of these gastrointestinal effects are mediated through vagal signaling[99] but are subject to tachyphylaxis even after relatively brief periods of GLP-1 receptor agonism.[100] This is also observed with GLP-1 analogs used clinically to treat type 2 diabetes and obesity.[101]

KEY POINTS

- Glucagon-like peptide-1 (GLP-1) also arises from the posttranslational processing of proglucagon by prohormone convertase-1/3. Endogenous GLP-1 has a short half-life in the circulation because of the ubiquity of di-peptidyl peptidase-4, which degrades GLP-1 to an inactive form. Although debate as to the physiologic role of GLP-1 continues, in pharmacologic concentrations it is a powerful insulin secretagogue.

Other Proglucagon-Derived Peptides and Their Potential Role(s).

The oxyntomodulin sequence consists of glucagon with an additional eight amino acids at its C-terminus. It binds to both the glucagon and GLP-1 receptors with low affinity. It has elicited interest because at pharmacologic concentrations in rodents and humans it decreases caloric intake and increases energy expenditure.[102] Substitution of a single amino acid that prevents binding to the glucagon receptor abolishes most of its effects on weight, although there is further decrease in efficacy in GLP-1 receptor knockout mice.[103] These observations have underpinned the impetus to develop GLP-1 and glucagon receptor coagonists for the treatment of type 2 diabetes and obesity.

REFERENCES

1. Smeekens SP, et al. Proinsulin processing by the subtilisin-related proprotein convertases furin, PC2, and PC3. *Proc Natl Acad Sci U S A*. 1992;89:8822–8826.
2. Rhodes CJ, Lincoln B, Shoelson SE. Preferential cleavage of des-31,32-proinsulin over intact proinsulin by the insulin secretory granule type II endopeptidase. Implication of a favored route for prohormone processing. *J Biol Chem*. 1992;267:22719–22727.
3. Skelly RH, Schuppin GT, Ishihara H, et al. Glucose-regulated translational control of proinsulin biosynthesis with that of the proinsulin endopeptidases PC2 and PC3 in the insulin-producing MIN6 cell line. *Diabetes*. 1996;45:37–43.
4. Mahajan A, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50:1505–1513.
5. Alarcon C, et al. Pancreatic beta-cell adaptive plasticity in obesity increases insulin production but adversely affects secretory function. *Diabetes*. 2016;65:438–450.
6. Haataja L, et al. Disulfide mispairing during proinsulin folding in the endoplasmic reticulum. *Diabetes*. 2016;65(4):1050–1060.
7. Sun J, et al. Proinsulin misfolding and endoplasmic reticulum stress during the development and progression of diabetes. *Mol Aspects Med*. 2015;42:105–118.
8. Tamaki M, et al. The diabetes-susceptible gene SLC30A8/ZnT8 regulates hepatic insulin clearance. *J Clin Invest*. 2013;123(10):4513–4524.
9. Ashcroft FM, Rorsman P. K(ATP) channels and islet hormone secretion: new insights and controversies. *Nat Rev*. 2013;9:660–669.
10. Barg S. Mechanisms of exocytosis in insulin-secreting B-cells and glucagon-secreting A-cells. *Pharmacol Toxicol*. 2003;92:3–13.
11. Rutter GA, Pullen TJ, Hodson DJ, et al. Pancreatic beta-cell identity, glucose sensing and the control of insulin secretion. *Biochem J*. 2015;466:203–218.
12. Talchai C, Xuan S, Lin HV, et al. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. *Cell*. 2012;150:1223–1234.
13. Robertson RP, Harmon J, Tran PO, et al. Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*. 2012;52:581–587.
14. Olson LK, Redmon JB, Towle HC, et al. Chronic exposure of HIT cells to high glucose concentrations paradoxically decreases insulin gene transcription and alters binding of insulin gene regulatory protein. *J Clin Invest*. 1993;92:514–519.
15. Pino MF, et al. Elevated glucose attenuates human insulin gene promoter activity in INS-1 pancreatic beta-cells via reduced nuclear factor binding to the A5/core and Z element. *Mol Endocrinol*. 2005;19:1343–1360.
16. Harmon JS, Stein R, Robertson RP. Oxidative stress-mediated, post-translational loss of MafA protein as a contributing mechanism to loss of insulin gene expression in glucotoxic beta cells. *J Biol Chem*. 2005;280:11107–11113.
17. DeFronzo RA, et al. Effects of exenatide (exendin-4) on glycemic control and weight over 30 weeks in metformin-treated patients with type 2 diabetes. *Diabetes Care*. 2005;28:1092–1100.
18. Ward WK, et al. Disproportionate elevation of immunoreactive proinsulin in type 2 (non-insulin-dependent) diabetes mellitus and in experimental insulin resistance. *Diabetologia*. 1987;30:698–702.
19. Tura A, et al. Basal and dynamic proinsulin-insulin relationship to assess beta-cell function during OGTT in metabolic disorders. *Am J Physiol Endocrinol Metab*. 2003;285:E155–E162.
20. Stoll RW, Touber JL, Menahan LA, et al. Clearance of porcine insulin, proinsulin, and connecting peptide by the isolated rat liver. *Proc Soc Exp Biol Med*. 2003;133:894–896.
21. Katz AI, Rubenstein AH. Metabolism of proinsulin, insulin, and C-peptide in the rat. *J Clin Invest*. 1973;52:1113–1121.
22. Sonksen PH, Tompkins CV, Srivastava MC, et al. A comparative study on the metabolism of human insulin and porcine proinsulin in man. *Clin Sci Mol Med*. 1973;45:633–654.
23. Varghese RT, et al. Performance of individually measured vs population-based C-peptide kinetics to assess beta-cell function in the presence and absence of acute insulin resistance. *Diabetes Obes Metab*. 2018;20:549–555.
24. Holst JJ, et al. Proglucagon processing in porcine and human pancreas. *J Biol Chem*. 1994;269:18827–18833.
25. Ellingsgaard H, et al. Interleukin-6 enhances insulin secretion by increasing glucagon-like peptide-1 secretion from L cells and alpha cells. *Nat Med*. 2011;17:1481–1489.
26. Marchetti P, et al. A local glucagon-like peptide 1 (GLP-1) system in human pancreatic islets. *Diabetologia*. 2012;55:3262–3272.
27. Ellingsgaard H, et al. Interleukin-6 regulates pancreatic alpha-cell mass expansion. *Proc Natl Acad Sci U S A*. 2008;105:13163–13168.
28. Unger RH, Ketterer H, Eisentraut AM. Distribution of immunoassayable glucagon in gastrointestinal tissues. *Metabolism*. 1966;15:865–867.
29. Barnes AJ, Bloom SR. Pancreatectomised man: a model for diabetes without glucagon. *Lancet*. 1976;1:219–221.
30. Lund A, et al. Evidence of extrapancreatic glucagon secretion in man. *Diabetes*. 2016;65:585–597.
31. Shah M, et al. Contribution of endogenous glucagon-like peptide 1 to glucose metabolism after Roux-en-Y gastric bypass. *Diabetes*. 2014;63:483–493.

32. Cobelli C, et al. Diabetes: models, signals, and control. *IEEE Rev Biomed Eng*. 2009;2:54–96.
33. Lerner RL, Porte Jr D. Acute and steady-state insulin responses to glucose in nonobese diabetic subjects. *J Clin Invest*. 1972;51:1624–1631.
34. Simpson RG, Benedetti A, Grodsky GM, et al. Early phase of insulin release. *Diabetes*. 1968;17:684–692.
35. Hazzard WR, Crockford PM, Williams RH. The insulin response to glucagon: effects of diabetes and obesity on nonglucose-dependent insulin secretion. *Diabetes*. 1968;17:297.
36. Lerner RL, Porte Jr D. Studies of secretin-stimulated insulin responses in man. *J Clin Invest*. 1972;51:2205–2210.
37. McRae JR, Chen M, Robertson RP. Improvement of defective insulin responses to glucose, arginine, and beta-adrenergic stimulation in diabetics by sodium salicylate. *Adv Prostaglandin Thromboxane Res*. 1980;8:1287–1289.
38. Robertson RP, Porte Jr D. The glucose receptor. A defective mechanism in diabetes mellitus distinct from the beta adrenergic receptor. *J Clin Invest*. 1973;52:870–876.
39. Dalla Man C, et al. Two-hour seven-sample oral glucose tolerance test and meal protocol: minimal model assessment of beta-cell responsivity and insulin sensitivity in nondiabetic individuals. *Diabetes*. 2005;54:3265–3273.
40. Basu R, et al. Use of a novel triple-tracer approach to assess postprandial glucose metabolism. *Am J Physiol Endocrinol Metab*. 2003;284:E55–E69.
41. Sathananthan A, et al. A concerted decline in insulin secretion and action occurs across the spectrum of fasting and postchallenge glucose concentrations. *Clin Endocrinol (Oxf)*. 2012;76:212–219.
42. Meier JJ, Veldhuis JD, Butler PC. Pulsatile insulin secretion dictates systemic insulin delivery by regulating hepatic insulin extraction in humans. *Diabetes*. 2005;54:1649–1656.
43. Piccinini F, Polidori DC, Gower BA, et al. Hepatic but not extrahepatic insulin clearance is lower in African American than in European American Women. *Diabetes*. 2017;66:2564–2570.
44. Van Cauter E, Mestrez F, Sturis J, et al. Estimation of insulin secretion rates from C-peptide levels. Comparison of individual and standard kinetic parameters for C-peptide clearance. *Diabetes*. 1992;41:368–377.
45. Robertson RP, et al. Arginine is preferred to glucagon for stimulation testing of beta-cell function. *Am J Physiol Endocrinol Metab*. 2014;307:E720–E727.
46. Robertson RP, et al. Assessment of beta-cell mass and alpha- and beta-cell survival and function by arginine stimulation in human autologous islet recipients. *Diabetes*. 2015;64:565–572.
47. Matschinsky FM, Glaser B, Magnuson MA. Pancreatic beta-cell glucokinase: closing the gap between theoretical concepts and experimental realities. *Diabetes*. 1998;47:307–315.
48. Stanley CA. Perspective on the genetics and diagnosis of congenital hyperinsulinism disorders. *J Clin Endocrinol Metab*. 2016;101:815–826.
49. Hattersley AT, Pearson ER. Minireview: pharmacogenetics and beyond: the interaction of therapeutic response, beta-cell physiology, and genetics in diabetes. *Endocrinology*. 2006;147:2657–2663.
50. Porksen N. Early changes in beta-cell function and insulin pulsatility as predictors for type 2 diabetes. *Diabetes Nutr Metab Clin Exp*. 2002;15:9–14.
51. Rolland JF, Henquin JC, Gilon P. Feedback control of the ATP-sensitive K(+) current by cytosolic Ca(2+) contributes to oscillations of the membrane potential in pancreatic beta-cells. *Diabetes*. 2002;51:376–384.
52. Bertram R, Satin LS, Sherman AS. Closing in on the mechanisms of pulsatile insulin secretion. *Diabetes*. 2018;67:351–359.
53. Cobelli C, et al. The oral minimal model method. *Diabetes*. 2014;63:1203–1213.
54. Huising MO, van der Meulen T, Huang JL, et al. The difference delta-cells make in glucose control. *Physiology (Bethesda)*. 2018;33:403–411.
55. Straub SG, Sharp GW. Evolving insights regarding mechanisms for the inhibition of insulin release by norepinephrine and heterotrimeric G proteins. *Am J Physiol Cell Physiol*. 2018;302:C1687–C1698.
56. Robertson RP. Arachidonic acid metabolism, the endocrine pancreas, and diabetes mellitus. *Pharmacol Ther*. 1984;24:91–106.
57. Robertson RP. Prostaglandins as modulators of pancreatic islet function. *Diabetes*. 1979;28:942–948.
58. Robertson RP, Chen M. A role for prostaglandin E in defective insulin secretion and carbohydrate intolerance in diabetes mellitus. *J Clin Invest*. 1977;60:747–753.
59. Mitrakou A, et al. Hierarchy of glycemic thresholds for counterregulatory hormone secretion, symptoms, and cerebral dysfunction. *Am J Physiol*. 1991;260:E67–E74.
60. Hardy AB, Serino AS, Wijesekara N, et al. Regulation of glucagon secretion by zinc: lessons from the beta cell-specific Znt8 knockout mouse model. *Diabetes Obes Metab*. 2011;13:112–117.
61. Zhou H, Zhang T, Harmon JS, et al. Zinc, not insulin, regulates the rat alpha-cell response to hypoglycemia in vivo. *Diabetes*. 2007;56:1107–1112.
62. Raju B, Cryer PE. Loss of the decrement in intraislet insulin plausibly explains loss of the glucagon response to hypoglycemia in insulin-deficient diabetes: documentation of the intraislet insulin hypothesis in humans. *Diabetes*. 2005;54:757–764.
63. Lee Y, Wang MY, Du XQ, et al. Glucagon receptor knockout prevents insulin-deficient type 1 diabetes in mice. *Diabetes*. 2011;60:391–397.
64. Janah L, et al. Glucagon receptor signaling and glucagon resistance. *Int J Mol Sci*. 2019;20:1–27.
65. Vella A, Camilleri M. The gastrointestinal tract as an integrator of mechanical and hormonal response to nutrient ingestion. *Diabetes*. 2017;66:2729–2737.
66. Vella A, Service FJ, O'Brien PC. Glucose counterregulatory hormones in the 72-hour fast. *Endocr Pract*. 2003;9:115–118.
67. O'Brien T, O'Brien PC, Service FJ. Insulin surrogates in insulinoma. *J Clin Endocrinol Metab*. 1993;77:448–451.
68. Longuet C, et al. The glucagon receptor is required for the adaptive metabolic response to fasting. *Cell Metab*. 2008;8:359–371.
69. Service FJ, et al. The classification of diabetes by clinical and C-peptide criteria. A prospective population-based study. *Diabetes Care*. 1997;20:198–201.
70. Finan B, Capozzi ME, Campbell JE. Repositioning glucagon action in the physiology and pharmacology of diabetes. *Diabetes*. 2020;69:532–541.
71. Muller WA, Faloona GR, Aguilar-Parada E, et al. Abnormal alpha-cell function in diabetes. Response to carbohydrate and protein ingestion. *N Engl J Med*. 1970;283:109–115.
72. Shah P, Basu A, Basu R, et al. Impact of lack of suppression of glucagon on glucose tolerance in humans. *Am J Physiol*. 1999;277:E283–E290.
73. Sharma A, et al. Impaired insulin action is associated with increased glucagon concentrations in non-diabetic humans. *J Clin Endocrinol Metab*. 2018;103:314–319.
74. Varghese RT, et al. Mechanisms underlying the pathogenesis of isolated impaired glucose tolerance in humans. *J Clin Endocrinol Metab*. 2016;101:4816–4824.
75. Moens K, et al. Dual glucagon recognition by pancreatic beta-cells via glucagon and glucagon-like peptide 1 receptors. *Diabetes*. 1998;47:66–72.
76. Gaisano HY, Macdonald PE, Vranic M. Glucagon secretion and signaling in the development of diabetes. *Front Physiol*. 2012;3:349.
77. Gromada J, et al. ATP-sensitive K+ channel-dependent regulation of glucagon release and electrical activity by glucose in wild-type and SUR1-/- mouse alpha-cells. *Diabetes*. 2004;53:S181–S189.
78. Leclerc I, et al. AMP-activated protein kinase regulates glucagon secretion from mouse pancreatic alpha cells. *Diabetologia*. 2011;54:125–134.
79. Vella A, et al. Targeting hepatic glucokinase to treat diabetes with TTP399, a hepatoselective glucokinase activator. *Sci Transl Med*. 2019;11:1–11.
80. Meier JJ, Kjems LL, Veldhuis JD, et al. Postprandial suppression of glucagon secretion depends on intact pulsatile insulin secretion: further evidence for the intraislet insulin hypothesis. *Diabetes*. 2006;55:1051–1056.
81. Unger RH, Orci L. The essential role of glucagon in the pathogenesis of diabetes mellitus. *Lancet*. 1975;1:14–16.
82. Faerch K, et al. Insulin resistance is accompanied by increased fasting glucagon and delayed glucagon suppression in individuals with normal and impaired glucose regulation. *Diabetes*. 2016;65:3473–3481.

83. Lee Y, et al. Hyperglycemia in rodent models of type 2 diabetes requires insulin-resistant alpha cells. *Proc Natl Acad Sci U S A*. 2014;111:13217–13222.
84. Shah M, et al. TCF7L2 genotype and alpha-cell function in humans without diabetes. *Diabetes*. 2016;65:371–380.
85. Zhou H, Zhang T, Harmon JS, et al. Zinc, not insulin, regulates the rat alpha cell response to hypoglycemia in vivo. *Diabetes*. 2007;56:1107–1112.
86. Wewer Albrechtsen NJ, et al. Hyperglucagonaemia analysed by glucagon sandwich ELISA: nonspecific interference or truly elevated levels? *Diabetologia*. 2014;57:1919–1926.
87. Wewer Albrechtsen NJ, et al. Inability of some commercial assays to measure suppression of glucagon secretion. *J Diabetes Res*. 2016;2016:8352957.
88. Shankar SS, et al. Standardized mixed-meal tolerance and arginine stimulation tests provide reproducible and complementary measures of beta-cell function: results from the Foundation for the National Institutes of Health Biomarkers Consortium Investigative Series. *Diabetes Care*. 2016;39:1602–1613.
89. Kieffer TJ, Habener JF. The glucagon-like peptides. *Endocr Rev*. 1999;20:876–913.
90. Vella A, Rizza RA. Extrapancreatic effects of GIP and GLP-1. *Horm Metab Res*. 2004;36:830–836.
91. Drucker DJ, Nauck MA. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet*. 2006;368:1696–1705.
92. Drucker DJ. Evolving concepts and translational relevance of enteroendocrine cell biology. *J Clin Endocrinol Metab*. 2015:jc20153449.
93. Smushkin G, et al. Defects in GLP-1 response to an oral challenge do not play a significant role in the pathogenesis of prediabetes. *J Clin Endocrinol Metab*. 2012;97:589–598.
94. Orskov C, Holst JJ, Poulsen SS, et al. Pancreatic and intestinal processing of proglucagon in man. *Diabetologia*. 1987;30:874–881.
95. Laferrere B, et al. Effect of weight loss by gastric bypass surgery versus hypocaloric diet on glucose and incretin levels in patients with type 2 diabetes. *J Clin Endocrinol Metab*. 2008;93:2479–2485.
96. Jimenez A, et al. GLP-1 and glucose tolerance after sleeve gastrectomy in morbidly obese subjects with type 2 diabetes. *Diabetes*. 2014;63:3372–3377.
97. Nauck MA, El-Ouaghlidi A. The therapeutic actions of DPP-IV inhibition are not mediated by glucagon-like peptide-1. *Diabetologia*. 2005;48:608–611.
98. Delgado-Aros S, et al. Effect of GLP-1 on gastric volume, emptying, maximum volume ingested, and postprandial symptoms in humans. *Am J Physiol Gastrointest Liver Physiol*. 2002;282:G424–G431.
99. Delgado-Aros S, et al. Effects of glucagon-like peptide-1 and feeding on gastric volumes in diabetes mellitus with cardio-vagal dysfunction. *Neuro Gastroenterol Motil*. 2003;15:435–443.
100. Nauck MA, Kemmeries G, Holst JJ, et al. Rapid tachyphylaxis of the glucagon-like peptide 1-induced deceleration of gastric emptying in humans. *Diabetes*. 2011;60:1561–1565.
101. Halawi H, et al. Effects of liraglutide on weight, satiation, and gastric functions in obesity: a randomised, placebo-controlled pilot trial. *Lancet Gastroenterol Hepatol*. 2017;2:890–899.
102. Wynne K, et al. Oxyntomodulin increases energy expenditure in addition to decreasing energy intake in overweight and obese humans: a randomised controlled trial. *Int J Obes (Lond)*. 2006;30:1729–1736.
103. Kosinski JR, et al. The glucagon receptor is involved in mediating the body weight-lowering effects of oxyntomodulin. *Obesity (Silver Spring)*. 2012;20:1566–1571.

30

Mechanisms of Insulin Action

Morris F. White and Kyle D. Copps

OUTLINE

Introduction, 498
Insulin, 499
  Introduction, 499
  Structure and Function of the Extracellular Domain of the Insulin Receptor, 499
  Regulation of the Insulin Receptor Tyrosine Kinase, 503
  The Insulin Receptor–Related Receptor, 503
The Insulin Signaling Cascade, 503
  Introduction, 503
  Insulin Receptor Tyrosine Protein Scaffolds, 504
  The IRS→PI3K→AKT Cascade, 505
  Insulin-Regulated Proteostasis, 506
  Insulin-Regulated Glucose Transport, 507
Heterologous Regulation of Insulin Signaling Cascades, 507
  Proximal Insulin Signals, 507
  Transcriptional Control of Insulin Receptor Substrate 1, 508
  Transcriptional Control of Insulin Receptor Substrate 2, 508
  Regulation of Insulin Receptor Substrate Signaling by Ser/Thr Phosphorylation, 509
  Regulation of Insulin Receptor Substrate Degradation, 509
  Regulation by Protein and Lipid Phosphatases, 510
  miRNA-Mediated Posttranscriptional Regulation, 510
Mouse Genetics Reveal Tissue-Integrated Insulin Signaling, 511
  Introduction, 511
  Systemic Insulin Receptor Signaling Cascade, 511
  Hepatic InsR—┤ FoxO1 Signaling, 511
  Insulin and Insulin-Like Growth Factor Signaling and Glucose Homeostasis In Skeletal Muscle, 512
  Insulin and Insulin-Like Growth Factor Signaling in Cardiac and Vascular Smooth Muscle, 512
  Adipose Insulin Signaling, 512
  Insulin Receptor Substrate 2 as a Gateway to β Cell Function, 513
  Insulin/IGF Signaling in Neurodegenerative Disease, 513
Insulin Resistance and Metabolic Disease, 513
  Introduction, 513
  Chronic Hyperinsulinemia, 514
  Nutrient Excess, 515
  Proinflammatory Cytokines, 515
  Hepatic Follistatin Links Liver and White Adipose Tissue Insulin Resistance, 515
  Endoplasmic Reticulum Stress and the Unfolded Protein Response, 516

KEY POINTS

- Diabetes is a metabolic disease of persistent hyperglycemia owing to various mechanisms that cause deficiency of insulin and/or insulin action in various target tissues. Understanding the mechanisms that dysregulate insulin action is important because type 2 diabetes mellitus is associated with chronic life-threatening complications, including nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, cardiovascular disease, kidney disease, and age-related neurodegeneration.

INTRODUCTION

Insulin and insulin-like growth factor signaling (IIS) integrates metazoan metabolism, growth reproduction, and lifespan with environmental signals.[1] Lower animals have a wide array of insulin-like peptides—thirty-eight in *Caenorhabditis elegans* and seven in fruit flies—that bind to a single insulin-like receptor tyrosine kinase.[2] The human genome encodes a smaller family of structurally related insulin-like peptides—including insulin itself, insulin-like growth factor (IGF) 1, and IGF2—which activate five similar receptor tyrosine kinases assembled from two genes. This chapter focuses mainly upon the insulin branch of mammalian IIS and its heterologous regulatory mechanisms.

Circulating glucose enters pancreatic islet β cells, where it promotes insulin gene expression and insulin secretion.[3] By contrast, endocrine IGF1 is secreted largely from hepatocytes during stimulation by nutrients, insulin, and growth hormone; however, IGF1 and IGF2 are also produced locally in many tissues and cells, including the central nervous system and various tumors.[4] IGF1 and IGF2 coordinate with insulin to regulate nutrient homeostasis, insulin sensitivity, pancreatic β cell function, and cell growth and survival.[4,5] Dysregulated insulin signaling is associated mainly with the “metabolic syndrome” and type 2 diabetes mellitus (T2DM), whereas dysregulated IGF1 and IGF2 signaling can promote cancer by inhibiting apoptosis and stimulating cell proliferation.[6]

Diabetes is a metabolic disease of persistent hyperglycemia owing to various mechanisms which cause deficiency of insulin and/or insulin action—including impaired glucose sensing or insulin secretion by pancreatic β cells (maturity onset diabetes of youth [MODY]); autoimmune-mediated β cell destruction (type 1 diabetes mellitus [T1DM]); or insufficient β cell insulin secretory capacity to compensate for peripheral insulin resistance (T2DM).[7] MODY is caused by mutations in at least 14 genes associated with β cell function—especially hepatocyte nuclear factor-4α (MODY1), glucokinase (GCK, MODY2), and hepatocyte nuclear factor-1α (MODY3).[8] Circulating autoantibodies against a variety of islet antigens, including insulin itself, are the hallmarks of T1DM.[9]

T2DM is the most prevalent form of diabetes that usually manifests at middle age, but can also develop in obese children and adolescents.[10] Physiologic stress owing to chronic excess nutrients and systemic inflammation impairs insulin action to promote T2DM.[11] The microbiome may also affect the onset and progression of obesity and T2DM owing to the production of metabolites that act as hormones, neurotransmitters, and inflammatory cytokines that can enter the circulation.[12,13] Genetic variations can attenuate or exacerbate the progression of metabolic disease and T2DM by directly or indirectly modulating the response to insulin.[14] Which genetic variations might converge upon a short list of novel "druggable" targets to prevent T2DM remains to be established.[15,16]

Understanding insulin signaling is important because T2DM is associated with chronic life-threatening complications—including nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH)[17], cardiovascular and kidney disease[18], and age-related neurodegeneration.[19] Drugs that reduce metabolic stress can have beneficial effects, including metformin[20] and sodium-glucose cotransporter 2 inhibitors.[21] Calorie restriction or intermittent fasting can promote nutrient homeostasis and reduce the demand for increased insulin signaling to promote health and longevity across all metazoans—including people.[22]

KEY POINTS

- The discovery of insulin 100 years ago and its application to the treatment of human disease marked a major turning point in the history of medicine. Our understanding of how insulin works has progressed over the last 50 years, starting with the identification of the insulin receptor and its tyrosine kinase activity. Understanding the insulin and insulin-like growth factor signaling systems as important integrators of metabolism, growth, and lifespan has expanded our understanding of insulin resistant states and diabetes.

INSULIN

Insulin was purified 100 years ago and used to treat T1DM. Some 50 years later the structure of insulin was solved.[23] During the past 50 years, the insulin receptor (InsR) has been identified, and its signaling mechanisms have been investigated.[1,24-26]

Insulin is synthesized in pancreatic β cells as a single polypeptide (proinsulin) that is processed by prohormone convertase-1/3 to excise the "C-peptide" and generate the bioactive disulfide-linked A- and B-chains composed of three α-helices (Fig. 30.1A).[1] Human IGF1 and IGF2 display high sequence similarity with the A- and B-chains of insulin while retaining the C-peptide (Fig. 30.1B).[27] Insulin dimerizes through a β-sheet between the two C-terminal portions of the insulin B-chain, including residues Phe^{B24}, Phe^{B25}, and Tyr^{B26} (see Fig. 30.1A). Three insulin dimers form a hexamer in the presence of zinc; however, IGFs do not aggregate because the connecting peptide blocks the dimerization surface.[25]

Insulin possesses two asymmetric receptor binding surfaces called "S1" (the classical site) and "S2" (the novel site) (see Fig. 30.1A).[28] Natural mutations of insulin, together with alanine-scanning mutagenesis and structural refinements, show that the high-affinity receptor binding site (S1, $K_d = 6$ nM) is composed of residues from the A- and B-chains (see Fig. 30.1A).[29,30] Phe^{B24}, Phe^{B25}, and Tyr^{B26} of the dimerization surface undergo a conformational change to facilitate high-affinity insulin binding (Fig. 30.1C).[1,31] A second receptor binding site (S2) is on the opposite side of insulin (see Fig. 30.1A).[29] Most of these contacts have been validated by cryo–electron microscopy (EM) to interact with specific regions in the α-subunit of the InsR.[32]

KEY POINTS

- The insulin receptor (InsR) precursor is encoded by a 150-kb gene composed of 22 exons on human chromosome 19p13.3-p13.2. The *INSR* gene encodes the extracellular domain that binds insulin and the intracellular tyrosine kinase that is activated during insulin binding. Feedback and heterologous signaling cascades modulate the activity of the InsR to integrate metabolism throughout the body.

Introduction

Most receptor tyrosine kinases (RTKs) exist as monomers that span the plasma membrane once and dimerize during ligand binding to bring two intracellular kinase domains into proximity for tyrosine transphosphorylation, which initiates downstream signaling (Fig. 30.2).[26] Bivalent or dimerized ligands are common strategies to homodimerize receptors through their extracellular immunoglobulin-like domains—including receptors for platelet-derived growth factor, vascular endothelial growth factor (VEGF), fibroblast growth factor, nerve growth factor, and others (see Fig. 30.2). By comparison, epidermal growth factor binding exposes a dimerization surface on the epidermal growth factor receptor (EGFR) that facilitates transphosphorylation to activate the dimers.[26] Insulin and IGFs are bivalent ligands that activate the preexisting catalytically inactive receptor dimer by inducing a receptor-wide conformational change that brings the catalytic domains into close proximity (see Fig. 30.2).[24,33]

Structure and Function of the Extracellular Domain of the Insulin Receptor

The InsR proreceptor is encoded by a 150-kb gene composed of 22 exons on human chromosome 19p13.3-p13.2, which encodes the extracellular domain (ECD) and the intracellular RTK (Fig. 30.3A). During synthesis, the proreceptor enters the endoplasmic reticulum, where it is glycosylated and stabilized by disulfide bonds, and then moves into the Golgi, where furin cleaves it at the Arg-Lys-Arg-Arg motif to generate the αβ·αβ dimer (Fig. 30.3A and B).[25,34] The ECD is composed of the entire α-subunit and an N-terminal portion of the β-subunit. The α-subunit begins with two N-terminal leucine-rich repeats (L1, amino acids [aa] 1–155; L2, aa 315–470) flanking a cysteine-rich (aa 195–310) region containing eight modular cysteine-linked repeats (see Fig. 30.3B). These modules are followed by three fibronectin-III motifs—including $\text{Fn}_{\text{iii}}1$ (aa 473–591) and $\text{Fn}_{\text{iii}}3$ (aa 808–904) flanking $\text{Fn}_{\text{iii}}2$ (aa597–717)—that are interrupted by a 120-aa insert domain containing the furin cleavage site (aa 636–690) (see Fig. 30.3A and B).

The disulfide-linked α-subunits and extracellular portion of the β-subunits fold into an inverted V configuration, with the apex formed by the L2 and $\text{Fn}_{\text{iii}}1$ domains (Fig. 30.3C).[35] Site-directed mutagenesis and photo-affinity labeling experiments with wild-type and mutant InsR originally predicted the location of two insulin-binding sites in the α-subunit. The high-affinity site-1 is composed of L1 from one α-subunit and CTa' (aa 693–710) from the other (Fig. 30.3D).[36,37] Cryo-EM resolves several points of contact of insulin with $\text{Fn}_{\text{IiI}}1$ and CTa, which increase binding affinity. When the InsR is saturated at high insulin concentrations, cryo-EM reveals the location of a distinct low-affinity binding site-2 near the $\text{Fn}_{\text{iii}}1$–$\text{Fn}_{\text{iii}}2$ interface (see Fig. 30.3D).[26,32,35,38] Insulin binding to site-1 and site-2 can stabilize the ECD in the T conformation, promoting convergence of the adjacent membrane-proximal $\text{Fn}_{\text{iii}}3$ domains toward each other, which transmits a conformational change through the hydrophobic transmembrane domain (TMD) to mediate tyrosine transphosphorylation of the RTK.[39]

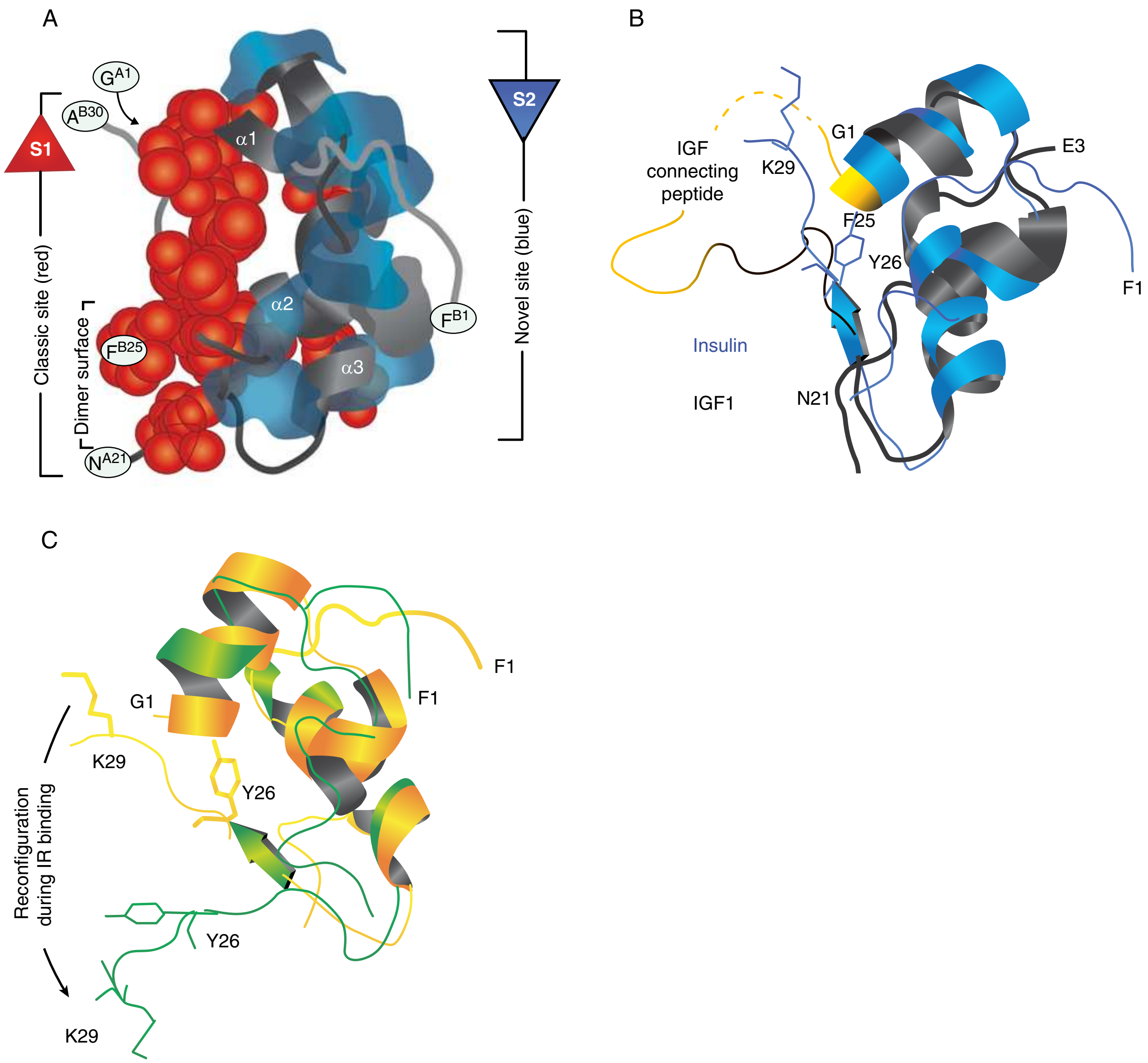


Fig. 30.1 Structural features of insulin and insulin-like growth factor 1 (IGF1). A, Pig insulin structure (pdb id 4 ins) showing the position of amino acids that compose the two binding surfaces (S1 and S2) that interact with the L1·CR·CT and $Fn_{iii}1 \cdot Fn_{iii}2$ regions of the insulin receptor, respectively (see Fig. 30.3). The insulin A- chain is shown in red, the B-chain is shown in blue, and the amino- and carboxyl-terminal residues of each chain are labeled, together with Phe25 of the B chain. B, Structural alignment of human insulin (in blue, pdb id 4ins) and insulin-like growth factor 1 (in black, pdb id 1imx), highlighting (in yellow) the position of the connecting peptide in IGF1 that is absent in the mature insulin polypeptide. C, Structural comparison of unbound insulin (in yellow, pdb id 4ins) versus that of insulin bound to the high affinity site on the insulin receptor (in green, pdb id 6sof). Cartoons were created in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

Exon 11 of the *INSR* gene is alternatively spliced depending upon the tissue and developmental stage to produce two InsR isoforms—including IRA that omits exon 11 and IRB that includes exon 11, adding 12 aa to the CTα (C-terminus of the α-subunit) (Fig. 30.3F).[28] Insulin binds with high affinity ($K_d < 0.5$ nM) to the homodimeric IRB ($\alpha\beta^{IR} \cdot \alpha\beta^{IR}$), which predominates in adult liver, muscle, and adipose tissues. IRA binds insulin with slightly lower affinity, but unlike IRB also binds IGF1 and IGF2 in the physiologic range (see Fig. 30.3F).[4,40] The homologous IGF1R homodimer ($\alpha\beta^{IGF1R} \cdot \alpha\beta^{IGF1R}$) is assembled without alternative splicing from its 19-exon gene located on human chromosome 15. IGF1R binds IGF1 and IGF2 with high affinity but binds insulin weakly (Fig. 30.3F). Binding affinity of the hybrid receptors ($\alpha\beta^{IR} \cdot \alpha\beta^{IGF1R}$) resemble IGF1R, which bind insulin with low affinity (see Fig. 30.3F).[41] Because the InsR occurs in two isoforms, a total of five receptors are generated from the two receptor genes (see Fig. 30.3F). Under ordinary conditions, insulin never activates the IGF1R tyrosine kinase, whereas IGF1 and IGF2 can activate the InsR tyrosine kinase while dimerized with the IGF1R.[42] IRA mediates the effects of

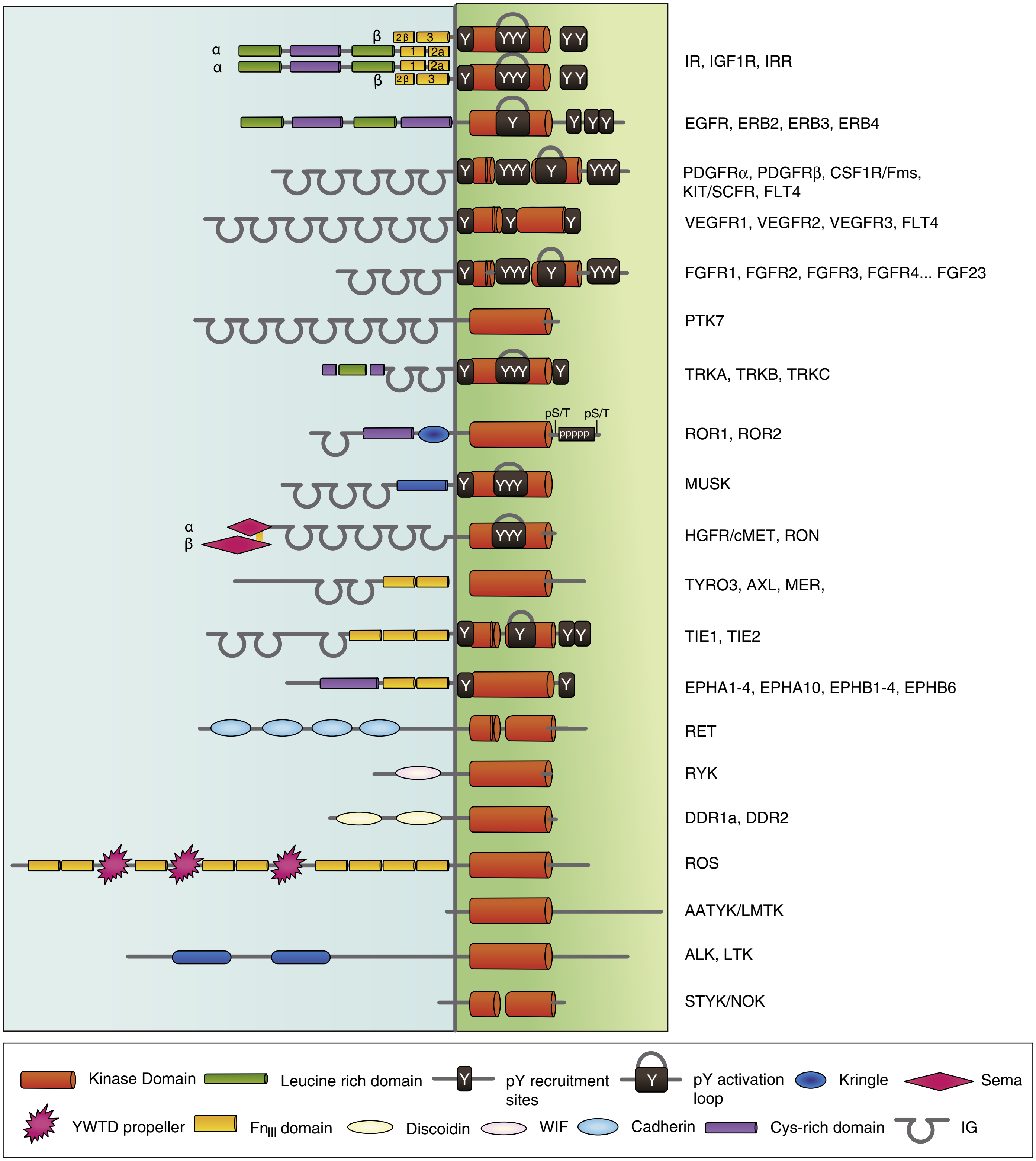

α
α
β
β
2β
3
1
2a
Y
YYY
YY
IR, IGF1R, IRR
EGFR, ERB2, ERB3, ERB4
PDGFRα, PDGFRβ, CSF1R/Fms, KIT/SCFR, FLT4
VEGFR1, VEGFR2, VEGFR3, FLT4
FGFR1, FGFR2, FGFR3, FGFR4... FGF23
PTK7
TRKA, TRKB, TRKC
pS/T
pS/T
ppppp
ROR1, ROR2
MUSK
HGFR/cMET, RON
TYRO3, AXL, MER,
TIE1, TIE2
EPHA1-4, EPHA10, EPHB1-4, EPHB6
RET
RYK
DDR1a, DDR2
ROS
AATYK/LMTK
ALK, LTK
STYK/NOK
Kinase Domain
Leucine rich domain
pY recruitment sites
pY activation loop
Kringle
Sema
YWTD propeller
Fn$_{III}$ domain
Discoidin
WIF
Cadherin
Cys-rich domain
IG

Fig. 30.2 **Twenty families of human RTKs (receptor tyrosine kinases).** The extracellular region is illustrated to the left on the blue-shaded background; the intracellular tyrosine kinase is illustrated to the right on the green-shaded background. Functional domains are defined in the gray box.

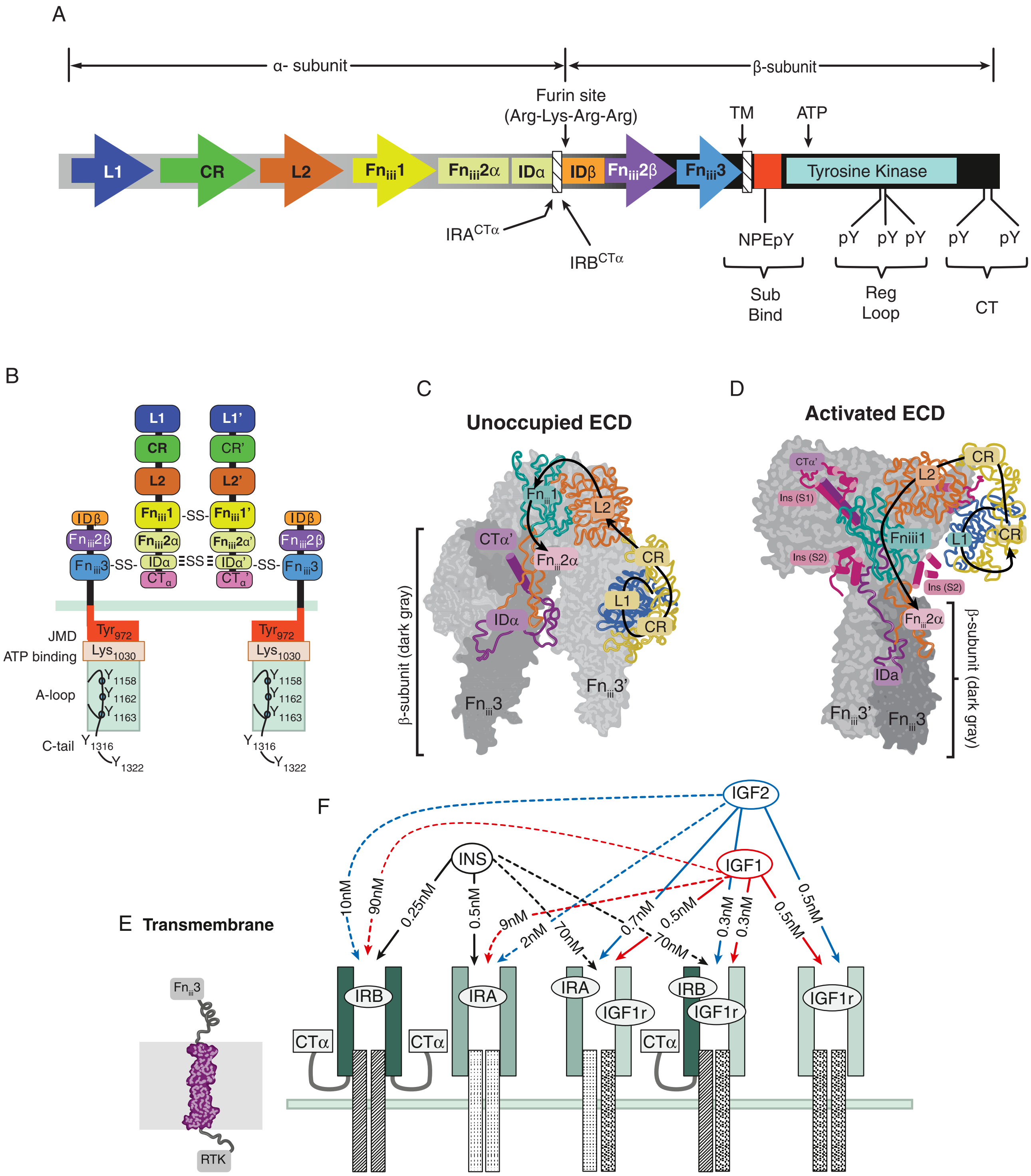

A
α- subunit
β-subunit
Furin site
(Arg-Lys-Arg-Arg)
TM
ATP
L1
CR
L2
Fniii1
Fniii2α
IDα
IDβ
Fniii2β
Fniii3
Tyrosine Kinase
IRACTα
IRBCTα
NPEpY
pY pY pY
pY pY
Sub Bind
Reg Loop
CT
B
L1
CR
L2
Fniii1
Fniii2α
IDα
CTα
L1'
CR'
L2'
Fniii1'
Fniii2α'
IDα'
CTα'
-SS-
≡SS≡
IDβ
Fniii2β
Fniii3
JMD
ATP binding
A-loop
C-tail
Tyr972
Lys1030
Y1158
Y1162
Y1163
Y1316
Y1322
C
Unoccupied ECD
β-subunit (dark gray)
Fniii1
L2
CTα'
Fniii2α
CR
L1
CR
IDα
Fniii3
Fniii3'
D
Activated ECD
CTα'
Ins (S1)
L2
CR
Fniii1
L1
CR
Ins (S2)
Ins (S2)
Fniii2α
IDa
Fniii3'
Fniii3
β-subunit (dark gray)
E Transmembrane
Fniii3
RTK
F
IGF2
INS
IGF1
10nM
90nM
0.25nM
0.5nM
9nM
2nM
70nM
0.7nM
0.5nM
70nM
0.3nM
0.3nM
0.5nM
0.5nM
IRB
CTα
CTα
IRA
IRA
IGF1r
IRB
IGF1r
CTα
IGF1r

Fig. 30.3 **Insulin receptor and insulin-like growth factor (IGF) 1 receptor hybrids.** A, Diagram of the insulin receptor (InsR) precursor showing the position of important modules in the α- and β-subunits—including leucine-rich regions (L1 and L2), a cysteine-rich region (CR); fibronectin iii domains (Fn$_{iii}$1, Fn$_{iii}$2a, Fn$_{iii}$2b, Fn$_{iii}$3); the insert domains IDa and IDb flanking the Arg-Lys-Arg-Arg furin cleavage site; the alternative IRA/IRB splice site that generates the short CTα^{IRA} or long CTβ^{IRB}; a transmembrane region (TM); the IRS binding motif (NPEpY); the activation loop autophosphorylation sites; and C-terminal tyrosine phosphorylation sites (CT). B, Diagram of the mature insulin receptor composed of two extracellular α-subunits and two transmembrane β-subunits. Contiguous modules of the two α-subunits are indicated with the relative location of disulfide bonds (S-S) between cysteine residues of the α and β subunits. Two regions within the α-subunit contribute to insulin binding—including L1·CR and the CTα that binds S1 of insulin, and the junction between Fn$_{iii}$1 and Fn$_{iii}$2 that binds S2 of insulin. The β-subunit contains the tyrosine kinase catalytic domain with an ATP binding site (Lys$_{1030}$) and several tyrosine phosphorylation sites. C, Cartoon of the inactive conformation of the extracellular domain (ECD) of the InsR (pdb id 6pxv). One ECD is shown in color, with the various modules of the a-subunit traced with a black line; the β-subunit is shown in dark grey; the adjacent ECD is shown in light grey. D, Cartoon of the active conformation of the ECD of the InsR (pdb id 6SOF).[39] The conformational change of the insulin-occupied ECD is illustrated by the black line tracing the relative location of various modules of the α-subunit, which brings the Fn$_{iii}$3 domain together with the adjacent Fn$_{iii}$3′ domain to facilitate transphosphorylation of the intracellular RTK. E, The transmembrane domain linking the Fn$_{iii}$3 domain of the ECD to the juxtamembrane region of the receptor tyrosine kinase (pdb id 2MFR). F, The insulin and IGF family of receptors. The insulin/IGF family consists of three peptide hormones—including insulin, IGF1, and IGF2—which bind as indicated to five distinct homo- or heterodimeric receptor isoforms—including IRA, IRB, IGF1R, and two hybrid receptors, IRA·IGF1R and IRB·IGF1R. The approximate binding affinities (K$_d$, nM) are shown in the figure.

IGF2 during embryogenesis and fetal development and in the adult brain.[4]

Regulation of the Insulin Receptor Tyrosine Kinase

Hard-won biochemical results obtained using [^{32}P]-labeled cells and proteins revealed the tyrosine kinase activity of the InsR[43]; however, until the InsR cDNA was isolated and sequenced,[44,45] other mechanisms for signal transduction were proposed—and some still persist.[46,47] Regardless, the discovery that rare cases of severe insulin resistance in humans are associated with *INSR* mutations that inactivate the tyrosine kinase—without altering insulin binding—supports the central hypothesis that tyrosyl phosphorylation initiates the intracellular signaling cascade.[43,48]

The intracellular portion of the InsR β-subunit begins with a 30-residue juxtamembrane domain (JMD) between the TMD and the tyrosine kinase domain that terminates with the 70-residue C-terminus (see Fig. 30.3A). Each intracellular region contains tyrosyl phosphorylation sites (numbered as in IRB): Tyr^{965} and Tyr^{972} in the JMD; Tyr^{1158}, Tyr^{1162}, and Tyr^{1163} in the activation loop (A-loop) of the kinase domain; and Tyr^{1328} and Tyr^{1334} in the C-terminus (see Fig. 30.3B).[49,50] Before insulin binds, the adjacent kinase domains of the $(\alpha\beta)_2$ dimer are widely separated to suppress transphosphorylation.[51] Conformational changes in the ECD during insulin binding reduce the distance between the TMD to facilitate transphosphorylation (compare Fig. 30.3C and D). In addition to kinase proximity, other mechanisms inhibit InsR kinase activity: (1) the unphosphorylated Tyr^{1162}—the second of the three A-loop tyrosine residues—blocks access of peptide substrates to the active site; (2) Asp^{1161} preceding Tyr^{1162} stabilizes the closed A-loop to inhibit unstimulated transphosphorylation[52]; (3) the N-terminal end of the A-loop (Asp-Phe-Gly motif) competes to elevate the $\mathrm{K_m}$ for adenosine triphosphate (ATP).[24] Early studies suggest that multisite transphosphorylation in the A-loop generates the insulin signal.[49] Intramolecular transphosphorylation begins at Tyr^{1158} and progresses to Tyr^{1162} and slowly to Tyr^{1163} until tris-phosphorylation stabilizes the open A-loop to fully activate the kinase (Fig. 30.4A–C).[53,54] This model of kinase regulation is confirmed by an Asp to Ala substitution at residue 1161 that shifts the A-loop toward the open configuration.[52] Substitution of Tyr^{1162} with Phe also increases basal transphosphorylation, consistent with its role to stabilize the closed conformation or block ATP and protein substrate entry into the active site (see Fig. 30.4C).[55,56]

The Insulin Receptor–Related Receptor

A third gene encodes the insulin receptor–related receptor (IRR), which has 50% identity (65% similarity) to the receptors for insulin and IGF1.[57] High-affinity peptide ligands are unknown for the IRR because the structure of the extracellular domain is rather different than that of the InsR.[58] The IRR is activated by exposure to alkaline environments, which generates insulin-like signaling in several tissues—including kidney, pancreas, stomach, ovary, and nerves.[58] The structure of the homologous IRR is different, as it assumes a unique conformation owing to a shorter distance between the fibronectin domains within the ECD.[58] Importantly, pH changes fail to induce conformational changes in the ECD, suggesting that the IRR may sense alkaline conditions inside the cell by changes in the JMD of the β-subunit. The IRR, together with the InsR and IGF1R, promotes male sexual development in mice.[57]

THE INSULIN SIGNALING CASCADE

KEY POINTS

- The insulin receptor (InsR) signal is transmitted into the cell through tyrosine phosphorylation of a family of insulin receptor substrates called IRS1, IRS2, IRS3 (rodents), and IRS4 that bind to various signaling proteins to control glucose uptake and metabolism and protein and lipid synthesis and degradation. The IRS complex can dissociate from the InsR to carry the signaling complex into the cell.

Introduction

Following discovery of the InsR tyrosine kinase, many groups searched for cellular proteins that may mediate downstream signals.[33,59] The first evidence for a substrate of any RTK came from antiphosphotyrosine antibody immunoprecipitates that revealed a 185-kDa phosphoprotein (pp185) in insulin-stimulated hepatoma cells.[60] Purification and molecular cloning of pp185 revealed the first member of a family of signaling scaffolds engaged by RTKs and the first insulin receptor substrate, called IRS1.[61] Three homologous *IRS* genes are expressed in human tissues—including *IRS1*, *IRS2*, and *IRS4*—whereas rodents also express IRS3 in adipose tissue.[62] IRS1 and IRS2 are broadly expressed in mammalian tissues, whereas IRS4 is mainly restricted to the hypothalamus.[63,64] Work with transgenic mice reveals that systemic IIS is largely mediated through IRS1, IRS2, or both.

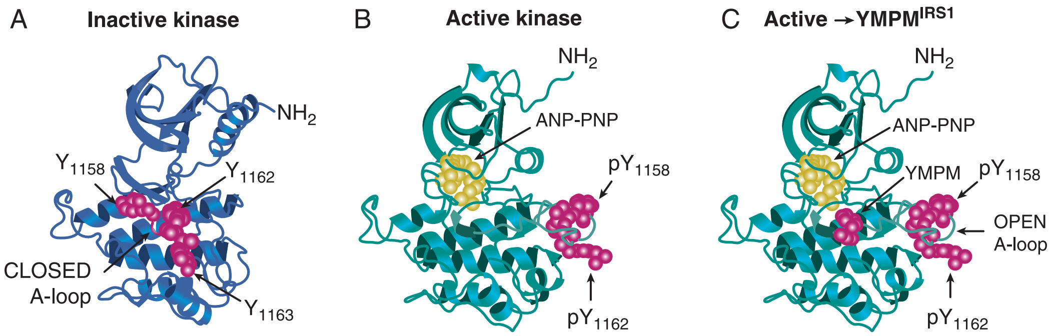


Fig. 30.4 **Structure and functional intracellular domains of the insulin receptor (InsR).** A, The inactive structure of the InsR kinase domain below the juxtamembrane domain is shown (pdb id 1IRK). The tyrosine residues of the A-loop are labeled and shown as red space-filling residues in the closed conformation. B, Following autophosphorylation of A-loop tyrosine residues (red space-filling), the A-loop moves out of the catalytic site to allow entry of adenosine triphosphate (yellow space-filling) to enter the active site (pdb id 1IR3). C, Peptide substrates such as YMXM-peptides of the IRS proteins enter the catalytic sites when the A-loop is open.

Insulin Receptor Tyrosine Protein Scaffolds

IRS proteins are composed of tandem, structurally similar, pleckstrin homology (PH) and phosphotyrosine binding (PTB) domains, followed by a long unstructured tail of tyrosine phosphorylation sites that coordinate downstream signals. IRS proteins also contain serine/threonine phosphorylation sites, which modulate IIS through feedback or heterologous mechanisms (Fig. 30.5). IRS proteins are strongly conserved in *Drosophila* to humans. Upon insulin stimulation, the IRS proteins are recruited to the JMD of the activated tyrosine kinase, where several tyrosine residues in the unstructured tail are phosphorylated and bind to the Src homology 2 (SH2) domains in various signaling proteins—especially the 85-kDa regulatory subunit (p85) of the phosphatidylinositol 3-kinase (PI3K) (Fig. 30.6). The binding between phosphorylated YMPM-motifs in IRS1 and both SH2 domains in PI3K was the first insulin signaling cascade to be reconstituted *in vivo* and *in vitro* (see Fig. 30.6).[65] Tyrosine phosphorylation of other sites in IRS1 or SHC (SHC adaptor protein 1) can recruit Grb2·SOS, which activates the mitogen-activated protein kinase cascade composed of RAS→MEK→ERK (Fig. 30.6).

Tyrosine phosphorylation of specific residues in IRS is accomplished through at least two mechanisms: (1) specific recruitment of the PTB domain of IRS to the phosphorylated NPEpY motif in the JMD of the InsR; and (2) recognition of specific phosphorylation motifs in the tail by the activated InsR catalytic domain. Based upon phosphopeptide mapping, tandem mass spectroscopy analysis, and phosphorylation of synthetic peptides, several motifs in IRS1 have been identified as optimal InsR substrates—including several YMXM-motifs and the YVNI-, YIDL-, and YASI-motifs.[66-68] There are other tyrosine phosphorylation motifs conserved in IRS1 and IRS2, but their functional specificity is poorly characterized (see Fig. 30.5). The phosphorylated motifs are recognized by the InsR catalytic domain as antiparallel β-strands relative to the C-terminal end of the open A-loop, positioning the site into two hydrophobic pockets on the activated kinase.[67]

The selective and regulated recruitment of IRS to the activated InsR and IGF1R adds an essential level of signaling specificity. The IRS proteins or SHC bind through their PTB domain to the NPEpY^{972} motif in the juxtamembrane region of InsR, which fills an L-shaped cleft on the PTB-domain[69,70]; however, the

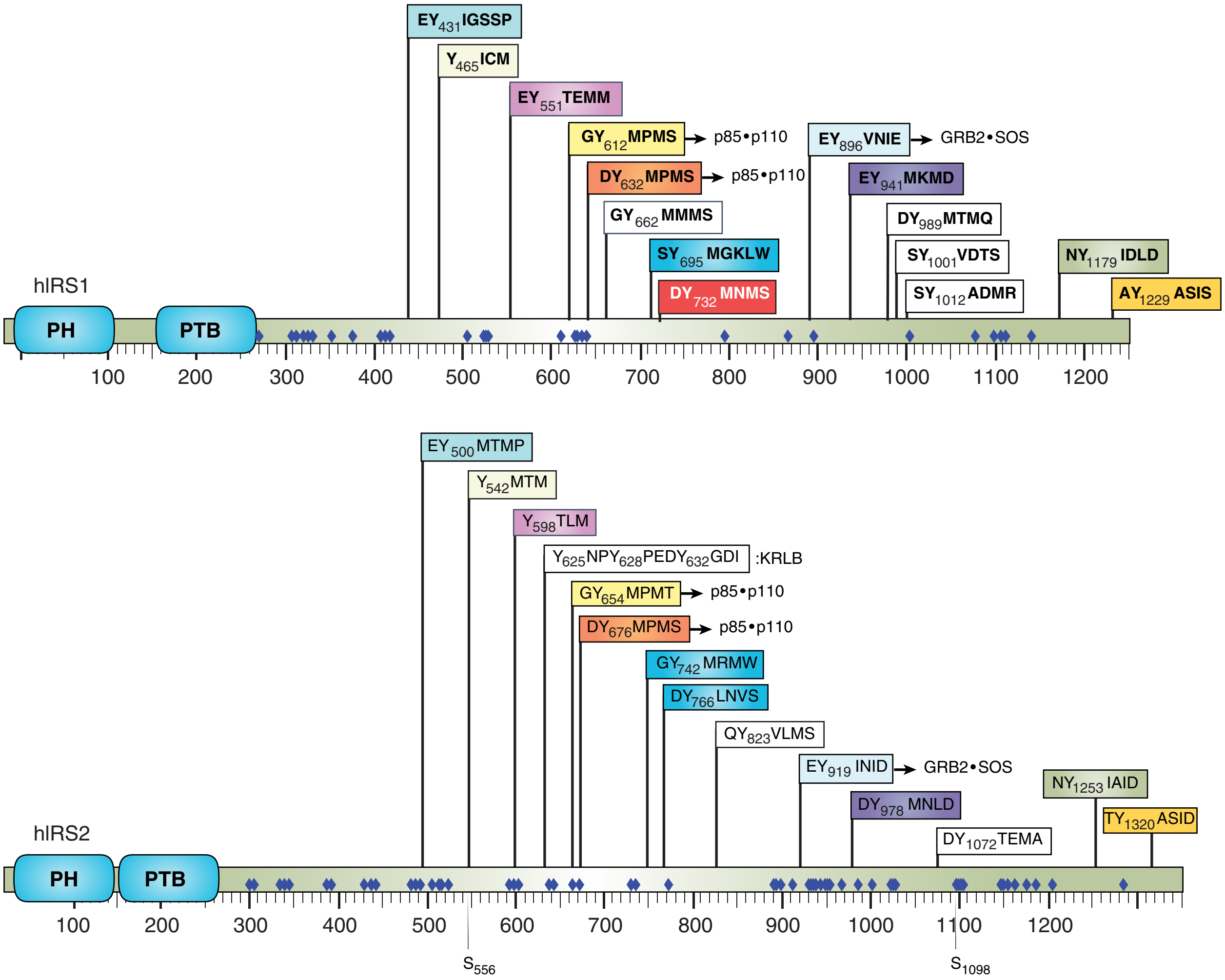

EY431IGSSP
Y465ICM
EY551TEMM
GY612MPMS → p85•p110
DY632MPMS → p85•p110
GY662MMMS
SY695MGKLW
DY732MNMS
EY896VNIE → GRB2•SOS
EY941MKMD
DY989MTMQ
SY1001VDTS
SY1012ADMR
NY1179IDLD
AY1229ASIS
hIRS1
PH
PTB
100
200
300
400
500
600
700
800
900
1000
1100
1200
EY500MTMP
Y542MTM
Y598TLM
Y625NPY628PEDY632GDI :KRLB
GY654MPMT → p85•p110
DY676MPMS → p85•p110
GY742MRMW
DY766LNVS
QY823VLMS
EY919INID → GRB2•SOS
DY978MNLD
DY1072TEMA
NY1253IAID
TY1320ASID
hIRS2
PH
PTB
100
200
300
400
500
600
700
800
900
1000
1100
1200
S556
S1098

Fig. 30.5 Alignments of human IRS1 and IRS2 tyrosine phosphorylation sites relative to the N-terminal pleckstrin homology (PH) and phosphotyrosine binding (PTB) domains. Conserved tyrosine phosphorylation motifs—including their number in the human protein and the surrounding amino acid sequences—are color-coded to highlight alignments; white boxes indicate unique sites in IRS1 or IRS2. The relative position of S/T-phosphorylation sites in IRS1 or IRS2 revealed by MS/MS are indicated with blue diamonds (◆). The unique kinase regulatory loop binding (KRLB) domain in IRS2 is centered about Y^{624}, which is not conserved in IRS1.

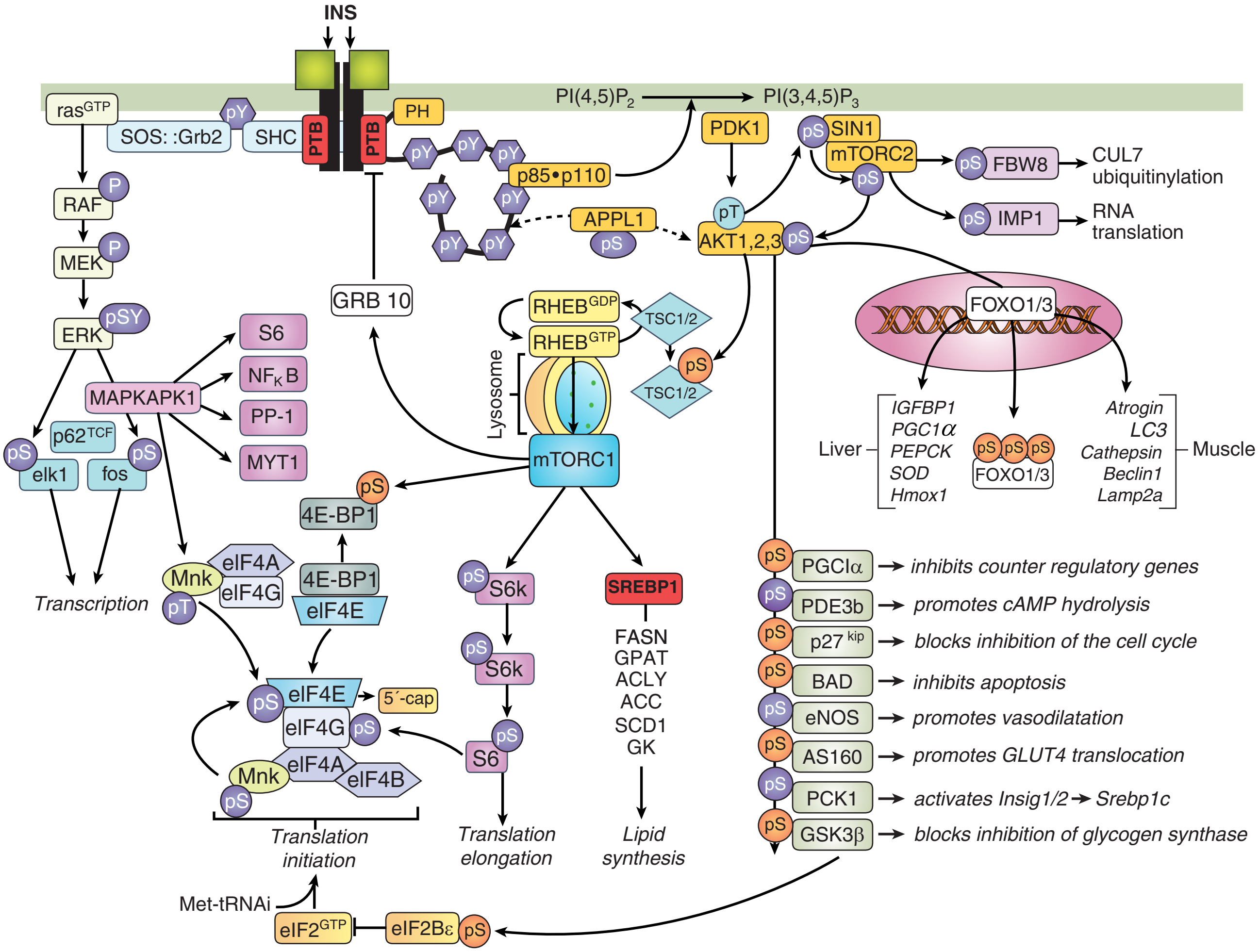

INS
ras^{GTP}
pY
SOS: :Grb2
SHC
PTB
PTB
PH
PI(4,5)P$_2$
PI(3,4,5)P$_3$
pY
pY
pY
p85•p110
pY
pY
pY
pY
PDK1
pS
SIN1
mTORC2
pS
FBW8
CUL7 ubiquitinylation
pS
IMP1
RNA translation
APPL1
pS
pT
AKT1,2,3
pS
P
RAF
P
MEK
pSY
ERK
GRB 10
S6
NF$_\kappa$B
PP-1
MYT1
MAPKAPK1
p62^{TCF}
pS
elk1
pS
fos
RHEB^{GDP}
RHEB^{GTP}
TSC1/2
pS
TSC1/2
Lysosome
mTORC1
FOXO1/3
IGFBP1
PGC1α
PEPCK
SOD
Hmox1
Liver
pS pS pS
FOXO1/3
Atrogin
LC3
Cathepsin
Beclin1
Lamp2a
Muscle
pS
4E-BP1
eIF4A
Mnk
eIF4G
4E-BP1
eIF4E
Transcription
pT
pS
S6k
pS
S6k
SREBP1
FASN
GPAT
ACLY
ACC
SCD1
GK
pS
eIF4E
5´-cap
eIF4G
pS
pS
S6
eIF4A
Mnk
eIF4B
pS
Translation initiation
Translation elongation
Lipid synthesis
pS PGCIα → inhibits counter regulatory genes
pS PDE3b → promotes cAMP hydrolysis
pS p27 kip → blocks inhibition of the cell cycle
pS BAD → inhibits apoptosis
pS eNOS → promotes vasodilatation
pS AS160 → promotes GLUT4 translocation
pS PCK1 → activates Insig1/2 → Srebp1c
pS GSK3β → blocks inhibition of glycogen synthase
Met-tRNAi
eIF2^{GTP}
eIF2Bε
pS

Fig. 30.6 **A canonical insulin/insulin-like growth factor (IGF) signaling cascade.** Two main branches propagate insulin signals generated via tyrosine phosphorylation of the IRS proteins, which bind and activate the PI3K and GRB2·SOS cascades. The Grb2/SOS complex promotes GDP/GTP exchange on p21^{Ras}, which activates the RAS→RAF→MEK→ERK1/2 cascade. Activated ERK stimulates transcriptional activity by direct phosphorylation of ETS domain-containing protein (ELK1) and by indirect phosphorylation of cFOS through MAPK-activated protein kinase-1 (MAPKAPK1). MAPKAPK1 also phosphorylates other proteins including S6 (ribosomal protein S6), NFkB, PP1 and MYT1. The activation of PI3K by recruitment to IRS produces PI3,4P$_2$ and PI3,4,5P$_3$ (antagonized by the action of PTEN or SHIP2), which recruits PDK1 and AKT to the plasma membrane. AKT is activated upon phosphorylation at T308 by PDK1 and at S473 by mTORC2. mTORC1 is activated by Rheb^{GTP}, which accumulates upon inhibition of the GAP activity of the TSC1·TSC2 complex following AKT-mediated phosphorylation of TSC2. mTORC1-mediates phosphorylation of S6K and SREBP1, which promotes protein or lipid synthesis, respectively. AKT phosphorylates many cellular proteins inactivating PGC1α, p21^{kip}, GSK3β, BAD and AS160, and activating PDE3b, PCK1 and eNOS. AKT-mediated phosphorylation of FOXO1 causes their sequestration in the cytoplasm, which inhibits their influence upon transcriptional activity. Insulin stimulates protein synthesis by altering the intrinsic activity or binding properties of key translation initiation and elongation factors (eIFs and eEFs, respectively), as well as critical ribosomal proteins. Components of the translational machinery that are targets of insulin regulation include eIF2B, eIF4E, eEF1, eEF2, and the S6 ribosomal protein.[114]

NPEpY^{972}-motif is a low-affinity binding site for the PTB domain of IRS1 (Kd ~87 μM), owing to a destabilizing effect of E^{971} that is essential for autophosphorylation of Y^{972} by the InsR.[24,71] By comparison, the PTB domain of SHC binds to NPEpY^{972} with a much higher affinity (K_d ~4 μM) (see Fig. 30.6). Regardless, the structurally similar but functionally distinct PH domain immediately upstream of its PTB domain helps recruit IRS1 to the InsR (see Fig. 30.6).[72,73] The molecular specificity of the IRS PH domain remains to be established, as heterologous PH domains substituted from other proteins reduce InsR signaling.[74,75]

IRS2 uses an additional mechanism—the kinase regulatory-loop binding (KRLB) domain—to interact with the InsR (see Fig. 30.5).[76,77] Structural analysis reveals that Tyr^{621} of IRS2 inserts into the receptor ATP binding pocket, whereas Tyr^{628} aligns for transphosphorylation.[78] This interaction may promote IR→IRS2 signaling, because the KRLB interaction is apparently absent in the IGF1R.[78]

The IRS→PI3K→AKT Cascade

PI3K. The PI3K→AKT cascade begins when insulin stimulates tyrosyl phosphorylation of YMPM-motifs in IRS proteins, which directly bind and activate the class 1A PI3K (see Fig. 30.6).[79] The PI3K is composed of a catalytic and a regulatory subunit. The catalytic subunit—including the widely expressed p110α (PIK3CA) and p110β (PIK3CB), and the leukocyte-restricted p110δ (PIK3CD) and p110γ(PIK3CG)—is inhibited and stabilized upon association with an 85-kDa regulatory subunit. Several regulatory subunits include p85α (PIK3R1) and p85β (PIK3R2). Alternative splicing of PIK3R1 produces p55α or p50α, and PIK3R3 encodes p55γ, both of which lack some of the N-terminal regulatory features of p85 while retaining affinity toward the catalytic subunits.[80,81] All the regulatory subunits contain two SH2 domains that bind phosphorylated YXXM motifs, which disinhibit the catalytic domain to produce phosphatidylinositol 3,4,5-trisphosphate (PI(3,4,5)P3).[82-84] Inhibition of PI3K production by chemical or genetic means blocks almost all metabolic responses stimulated by insulin—including glucose influx, glycogen and lipid synthesis, and adipocyte differentiation—confirming that PI3K is a critical node coordinating insulin action.

Upon binding of p85 to phosphorylated YMPM motifs in IRS, the activated PI3K converts PI(3,4)P2 into PI(3,4,5)P3, which binds to the PH domains in various signaling proteins to recruit and activate them at the plasma membrane—including 3'-phosphoinosotide-dependent protein kinase-1 (PDK1), v-akt murine thymoma viral oncogene (AKT), and target of rapamycin complex 2 subunit MAPKAP1 (SIN1). AKT is activated by phosphorylation at Thr^{308} in the A-loop by the juxtaposed PDK1 (see Fig. 30.6). SIN1 is recruited to the plasma membrane by PI(3,4,5)P3, where it is phosphorylated at Thr^{86} by $\text{pT308}^{\text{AKT}}$. SIN1 complexes with mTORC2, which phosphorylates Akt at Ser473.[85] Bis-phosphorylated AKT can phosphorylate a wider array of substrates, including FOXO transcription factors (see Fig. 30.6). In adipocytes and muscle, SIN1 can recruit atypical protein kinase C to PI(3,4,5)P3, where it can be activated by PDK1 and promote insulin-stimulated glucose uptake.[81,86,87] Finally, SIN1 can be phosphorylated by ribosomal protein S6 kinase (S6K1), which creates a feedback mechanism to inhibit mTORC2.[85,88,89]

AKT. Mammalian AKT1, AKT2, and AKT3 phosphorylate more than 100 substrates that control cell survival, growth, proliferation, angiogenesis, metabolism, and migration (see Fig. 30.6). Although overlap exists, AKT1 mainly regulates development, growth, and survival, whereas AKT2 regulates metabolism through insulin-stimulated GLUT4 translocation and liver glucose and lipid metabolism.[90] Human subjects with a dominant negative mutation in AKT2 display features of T2DM.[91] AKT3 promotes neuronal development, and homologous glucocorticoid-regulated kinase 3 (Sgk3) synergizes with AKT2 in pancreatic β cells to stimulate proliferation and insulin release.[92,93]

mTORC1. Mechanistic target of rapamycin (mTOR) is a Ser/Thr kinase that belongs to the PI3K-related kinase family and forms two large functionally distinct protein complexes—mTORC1 and mTORC2—owing to the assembly of common and unique subunits.[94] Both complexes are controlled by insulin, IGF1, and other growth factors through the PI3K→AKT cascade; however, each complex is recruited to different compartments and responds distinctly to nutrients, stress, hypoxia/energy status, and other factors.[94] mTORC1 is targeted to the lysosome, where it is inhibited by the TSC1·TSC2 complex until AKT phosphorylates and inhibits TSC2 (tuberin)—a GTPase-activating protein (GAP)—which promotes GTP·Ras homolog enriched in brain (RHEB) accumulation (see Fig. 30.6). FK506 binding protein 8 (FKBP8) inhibits mTORC1 until RHEB·GTP promotes its dissociation from mTORC1.[95] Regulation by TSC1·TSC2→RHEB is also modulated by AKT-mediated phosphorylation of AKT substrate 1, which promotes its dissociation from RAPTOR to activate mTORC1.[92] RHEB resides on endomembranes, including lysosomes, where it can interact with mTORC1 only if enough amino acids are available (see Fig. 30.6).[96,97] The dependence upon amino acids ensures that the cellular environment is sufficient to support growth factor (insulin) action. Activated mTORC1 promotes many cellular anabolic processes needed for growth and proliferation—including the stimulation of glycolytic flux and mitochondrial function, protein and lipid synthesis, and the inhibition of autophagy and lysosomal biogenesis (see Fig. 30.6).[89] Thus, the IRS→PI3K→AKT→mTORC1 cascade controls many key cellular processes that balance cellular integrity, long-term growth, and survival.

mTORC2. In addition to fully activating AKT, mTORC2 plays a role in mRNA processing when it phosphorylates insulin-like growth factor 2 mRNA-binding protein 1, which promotes IGF2-leader 3'-mRNA translational initiation by internal ribosomal entry.[98] Thus, mTORC2 can promote growth by regulating IGF2 production, which can activate IRA in the mouse embryo. mTORC2 also regulates protein ubiquitination by phosphorylating F-box and WD repeat domain–containing 8 (FBW8), a Cullin 7 E3 ubiquitin ligase recognition subunit (see Fig. 30.6). The phosphorylation of FBW8 stabilizes and promotes ubiquitination of targeted substrates, including IRS1, which contributes to insulin resistance in certain tissues and cells (see Fig. 30.6).[99] Thus, mTORC2 signaling can modulate IIS through its direct effect upon AKT activity and its indirect control of IIS regulatory nodes.[98]

FOXO. Forkhead box O subfamily (FOXO) proteins—FOXO1, FOXO3a, FOXO4, and FOXO6—are integrative transcription factors that control target gene expression to coordinate many biological process, including metabolism, autophagy, apoptosis, reactive oxygen species detoxification, DNA repair, cell cycle, stem cell maintenance, and longevity.[100] FOXOs contain five highly conserved domains, including the N-terminal region with several AKT phosphorylation sites[101]; a highly conserved forkhead DNA binding domain (DBD); a nuclear localization signal located just downstream of the DBD; a nuclear export sequence; and a C-terminal transactivation domain.[102] Retention of FOXO in the nucleus promotes or suppresses expression of genes by binding to TTGTTTAC motifs.[103] FOXO nuclear localization is regulated by various posttranslational modifications, including acetylation, methylation, glycosylation, and ubiquitylation[104]; however, one of the most critical is phosphorylation by AKT or other kinases, including IκB kinase and SGK, which promotes nuclear exclusion.[101,103] Deacetylation of FOXO by silent mating type information regulation NAD-dependent deacetylase (SIRT1) can suppress transcriptional activity directly.[105]

FOXOs regulate metabolism by inducing genes that promote hepatic glucose production (HGP); insulin secretion in β cells and β cell growth and differentiation; and fat and muscle mass.[106-108] In mammalian liver, decreased circulating insulin during fasting promotes the nuclear localization of FOXO, where it interacts with peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC1α) and cAMP response element-binding protein (CREB) complexed with CREB regulated transcription coactivator 2 (CRTC2) to increase the expression of glucose-6-phosphatase (G6PC) and phosphoenolpyruvate carboxykinase 1 (PCK1) for HGP.[109] FOXO also regulates hundreds of other hepatic genes directly or indirectly to coordinate glucose and lipid homeostasis and systemic insulin sensitivity. Severe and persistent starvation can trigger FOXO1-mediated hepatic insulin-like growth factor-binding protein 1 expression to reduce circulating IGF1 bioavailability and attenuate somatic growth.[110] In muscle, nuclear FOXO1 promotes autophagy that breaks down muscle protein for use by the liver to make glucose. Thus, reduced IIS in skeletal muscle can underlie the loss of muscle mass and glycemic control during chronic insulin resistance.[111]

Insulin-Regulated Proteostasis

Protein synthesis is one of the best understood mTORC1-regulated processes that is controlled, at least in part, through phosphorylation/

activation of S6K1/2 (the ribosomal protein S6 kinases) and phosphorylation/inhibition of eukaryotic translation initiation factor 4E binding protein 1 (4E-BP1) (see Fig. 30.6). Insulin mainly stimulates protein synthesis in liver, while it suppresses protein degradation and amino acid release by muscle; however, hepatic protein synthesis can occur during starvation owing to influx of amino acids from muscle. Muscle expresses receptors for insulin and IGF1, but deletion of the InsR reduces muscle size by 20%, while deletion of the IGF1R has no effect; however, deletion of both receptors is usually fatal, as it decreases muscle size by at least 60%.[112] The deletion of muscle IRS1 or IRS2 barely influences skeletal muscle mass; however, deletion of both reduces muscle mass by approximately 50%.[113] Thus, InsR is more important than IGF1R, and both IRS1 and IRS2 contribute to muscle proteostasis. In the absence of InsR and IGF1R, the compound deletion of FOXO1, FOXO3, and FOXO4 can prevent muscle loss.[112]

Insulin coordinates protein synthesis through various translation factors—including eIF2B (multisubunit guanine nucleotide exchange factor for translation initiation factor eIF2), eIF4E (mRNA cap-binding protein), eEF1 (eukaryotic elongation factor 1), eEF2 (eukaryotic elongation factor 2), and the S6 ribosomal protein[114,115] (see Fig. 30.6). During insulin stimulation, AKT—┤ GSK3β reduces phosphorylation of eIF2Bε, which promotes the formation of $eIF2^{GTP}$ that recruits $tRNAi^{Met}$ to the ribosome to promote translation initiation.[114,116] Translation initiation also involves eukaryotic initiation factor 4F (eIF4F) that directs ribosomes to the cap structure of mRNAs. Translation initiation is strongly inhibited by 4E-BP1, which directly interacts with eIF4E to block the eIF4F complex from repressing translation (see Fig. 30.6). mTORC1-mediated phosphorylation of 4E-BP1 dissociates the 4E-BP1·eIF4E complex to allow interaction between eIF4E and the eIF4G scaffold protein in the eIF4F complex (see Fig. 30.6).[117] MAP kinase-interacting serine/threonine-protein kinase 1, an insulin-stimulated kinase activated through the RAS→ERK cascade, also resides in the eIF4F complex, where it phosphorylates eIF4E (see Fig. 30.6).[118] Phosphorylation of eIF4E increases the binding affinity for mRNA caps to enhance translation initiation. Moreover, eIF4G is phosphorylated by S6 kinase, which enhances binding of mRNA to the 43S initiation complex (see Fig. 30.6).[119]

Although IIS enhances muscle protein synthesis, inhibition of protein degradation by IIS also controls proteostasis.[112,120] Proteins are degraded by the ubiquitin/proteasome or the autophagy/lysosome systems.[121] The proteasome is a multisubunit enzyme complex that coordinates degradation of ubiquinated proteins that control cell cycle, transcription, signaling, trafficking, and protein quality control.[122] The autophagy system, which is a less specific but highly regulated, engulfs portions of cytoplasm, organelles, glycogen, and protein aggregates in double-membrane vesicles that fuse with lysosomes.[100,122] Inhibition of IIS in skeletal muscle by gene deletion reduces muscle mass owing to more proteasomal and autophagy/lysosomal mediated degradation. Inactivation of FOXOs—FOXO1, FOXO3, and FOXO4—in skeletal muscles prevents muscle loss in response to prolonged fasting or inactivated IIS, because these transcription factors are required for the induction of several atrophy-related genes. Regardless, inhibition of autophagy might promote feedback insulin resistance through ubiquitination and proteasome-mediated degradation of IRS-1.[123]

Insulin-Regulated Glucose Transport

Glucose transport is the prototype insulin response that is rate limiting for dietary glucose utilization by muscle and adipose.[124,125] Members of the sodium-independent, facilitated-diffusion glucose/hexose transporter (GLUT) family are distinguished by substrate specificity and affinity, tissue distribution, and regulation. GLUT4 is the principal insulin-responsive glucose transporter that is expressed in white adipose tissue (WAT), brown adipose tissue (BAT), skeletal muscle, and heart, where it plays a direct role in glucose homeostasis. The molecular mechanisms linking the insulin signal to increased glucose influx involves complex multistep coupling between the signaling cascade and vesicle trafficking.[124,125]

Under basal conditions, GLUT4 accumulates in a general recycling compartment that, upon sorting, generates GLUT4 storage vesicles (GSVs) that form insulin-responsive vesicles (IRVs) that fuse with the plasma membrane in response to insulin[124-126] (Fig. 30.7). In addition to GLUT4, the GSVs contain other components, including sortilin, which mediates GLUT4 retrieval from trans-Golgi network and sorting endosomes to form the GSVs.[124]

Insulin acts at several sites to regulate GLUT4 translocation and recycling—including GSV assembly, release from intracellular retention sites, transport and fusion at the plasma membrane, and return. Insulin appears to promote dissociation of DGKζ·PIP5K1α (a complex of diacylglycerol kinase-ζ and phosphatidylinositol 4-phosphate 5-kinase 1a) from IRS1, which facilitates formation of the IRS1·PI3K complex to activate AKT and Rac family small GTPase 1 (RAC1) (see Fig. 30.7).[124,127] RAC1 may promote AKT membrane compartmentalization or actin remodeling to enhance insulin-stimulated GLUT4 translocation.[124]

Akt substrate of 160 kDa (AS160) and TBC1 domain family member 4 (TBC1D4) are GTPase-activating proteins that maintain their target RABs—including RAB8A, RAB10, RAB14, and others that regulate intracellular membrane flux—in an inactive GDP-bound state under basal conditions.[124,125] AKT phosphorylates and inhibits AS160 in adipose and TBC1Ds in muscle, which promotes RAB·GTP accumulation.[128,129] GTP-loaded RAB8A and RAB14 (muscle) or RAB10 (adipose) facilitates translocation of GSVs to the PM (see Fig. 30.7).[124,128] AKT2 phosphorylates and inactivates RAL·GAP, which leads to the accumulation and activation of RAL·GTP that binds to exocyst components—Sec5 and Exo70—to tether IRVs to the plasma membrane (see Fig. 30.7). Thus, insulin-stimulated AKT regulates translocation, docking, and fusion of GLUT4 vesicles with the plasma membrane.[124,128]

Insulin may regulate GLUT4 recycling back from the plasma membrane into an endocytic recycling compartment (ERC).[130] Formation of GLUT4 vesicles from the ERC relies upon a coat complex containing clathrin and ACAP1—an ARF6 GAP—that is recruited to ERC membranes by GRP1, an ARF6-GEF.[131] Budding of GSVs from the ERC is stimulated through the AKT-mediated phosphorylation of GRP1 (see Fig. 30.7).[132] Phosphomimetic mutations in GRP1 increase plasma membrane GLUT4 localization in 3T3-L1 white adipocytes, to a large degree bypassing the GLUT4 vesicle retention mechanisms.[132]

HETEROLOGOUS REGULATION OF INSULIN SIGNALING CASCADES

KEY POINTS

- Many mechanisms based upon heterologous regulation of gene expression and protein function exist to integrate and balance the InsR signal with other biological response. When these balanced signals are coopted by chronic inflammation or endoplasmic reticulum stress, insulin resistance can develop, which can progress to diabetes.

Proximal Insulin Signals

A couple of decades of genetic experiments in mice have established that changes in the relative function of a broad array of insulin signaling

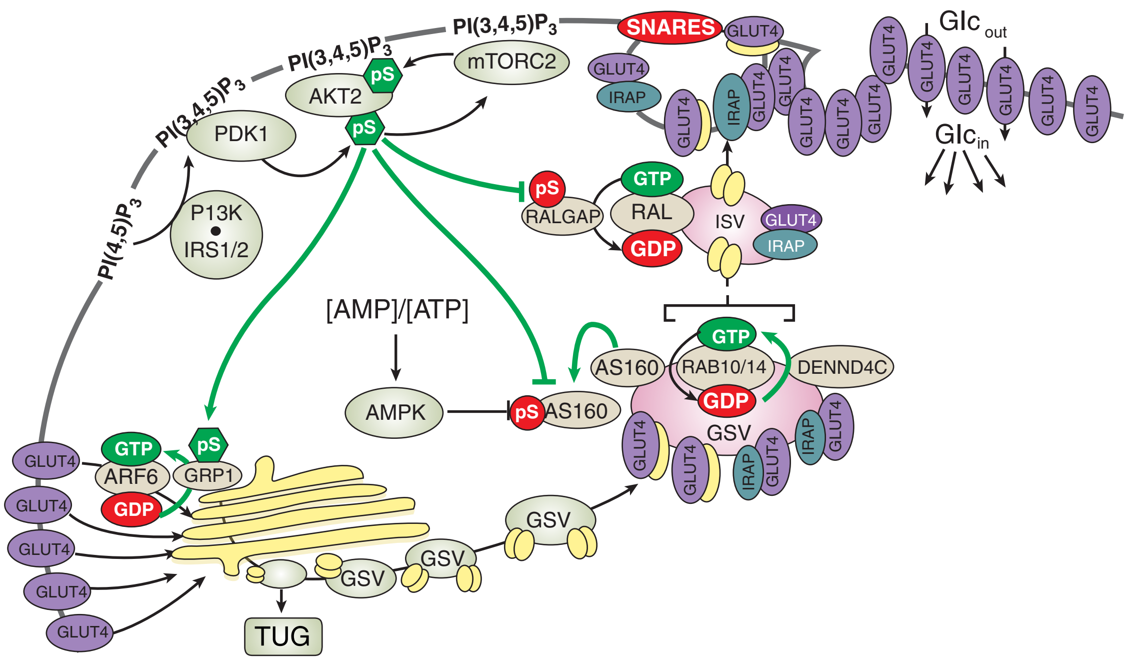


Fig. 30.7 **A mechanism of insulin-stimulated glucose transport by the PI3K→AKT→AS160/RALGAP/GRP1 branch of the insulin signaling pathway in adipose.** These AKT substrates mediate various steps of the assembly and trafficking of GLUT4 storage vesicles (GSVs). Endocytosis of GLUT4 cycles through the trans-Golgi network and is stimulated by AKT→GRP1·ARF6 signaling. AKT→AS160/RALGAP promotes translocation of the GSVs to the plasma membrane. Activation of these pathways by insulin promotes the accumulation of GLUT4-containing vesicles in the plasma membrane, which increases glucose influx.

components, nutrient sensors, and their downstream metabolic effectors can have profound effects upon insulin sensitivity, pancreatic β cell function, and nutrient homeostasis. Indeed, a 50% reduction in the concentration of the InsR IRS1, and IRS2 achieved by genetic means causes growth deficits and diabetes in mice[133]; however, physiologic mechanisms that regulate the concentration and function of proximal insulin signaling components are not understood well enough between tissues to guide the development of efficacious and safe treatments. For example, proximal IIS components can be regulated by yin yang 1 (YY1) in skeletal muscle or liver—including IGF1 and IGF2, IRS1 and IRS2, and AKT1, AKT2, and AKT3.[134,135] YY1 is a ubiquitous homeobox transcription factor that has diabetic (liver) or antidiabetic (skeletal muscle) effects, making it a challenging therapeutic target.

Transcriptional Control of Insulin Receptor Substrate 1

Recent work suggests that IRS1 expression may be regulated by transcriptional repressors, including AP2β or the p160 family of nuclear receptor coactivators p/CIP (p300/CBP/cointegrator-associated protein) and SRC1 (steroid receptor coactivator-1).[136,137] AP2β is expressed in adipose tissue where it promotes adipocyte hypertrophy, inhibits adiponectin expression, and enhances the expression of inflammatory adipokines such as IL6 (interleukin 6) and MCP1 (C-C motif chemokine ligand 2)[136] AP2β decreases IRS1 mRNA and protein concentration in adipocyte cell lines. Interestingly, GWAS reveals AP2β as a candidate gene for the risk of obesity and T2DM, which may involve negative regulation of IRS1 expression.[138] By contrast, p/CIP and SRC1 serve as transcriptional coactivators for nuclear hormone receptors and certain other transcription factors.[139] IRS1 expression increases upon inactivation of p/CIP and SRC1 in mice, which increases glucose uptake and enhanced insulin sensitivity in WAT and skeletal muscle. Finally, muscle-specific transcriptional coactivator with PDZ-binding motif (Taz) knockout mice display decreased IRS1 expression and insulin resistance.[140] Statins can reduce TAZ levels, which may reveal a mechanism of insulin resistance owing to decreased IRS1 expression; however, effects of statins upon protein prenylation might also be involved.[140]

Transcriptional Control of Insulin Receptor Substrate 2

IRS2 transcription is regulated by multiple factors, including CREB and its coactivator CRTC2, FOXO isoforms, HIF2α (hypoxia-inducible factor-2α encoded by EPAS1), NFAT (nuclear factor of activated T-cells), PGC1α, sterol regulatory element binding protein 1 (SREBP1c), signal transducer and activator of transcription 3 (STAT3), and transcription factor E3 (TFE3).[141-144] Multiple transcriptional mechanisms induce hepatic IRS2 expression during fasting, including nuclear FOXO that creates a direct feedback loop to augment insulin signaling during fasting and sensitively suppress it after a meal through AKT⊣ FOXO. Fasting or exercise also induce the CREB·CRTC2 transcriptional complex through glucagon signaling that increases expression of gluconeogenic genes along with IRS2.[142] Moreover, glucagon acting through PGC1α can increase hepatic IRS2 expression while suppressing IRS1, which can fine-tune gluconeogenesis and HGP during the fasting→feeding transition, with a minimal effect upon lipogenesis.[144] An E-box overlapping the FOXO site binds basic helix-loop-helix transcription factor E3 (TFE3), which converges with FOXO to promote IRS2 expression[145]; however, these elements also overlap with an SRE that binds sterol regulatory element binding transcription factor 1 (SREBF1), an important transcriptional activator of lipid synthesis.[146] Active hepatic SREBF1 increases during nutrient excess and chronic insulin stimulation, which decreases IRS2 expression to promote insulin resistance while promoting lipogenesis.[147,148] HIF2α is expressed in the hypoxic perivenous zone of the liver, where it restrains gluconeogenesis compared with other liver zones.[149,150] Thus, IRS2 expression is highly integrated through multiple metabolic sensors to promote insulin sensitivity and metabolic homeostasis.

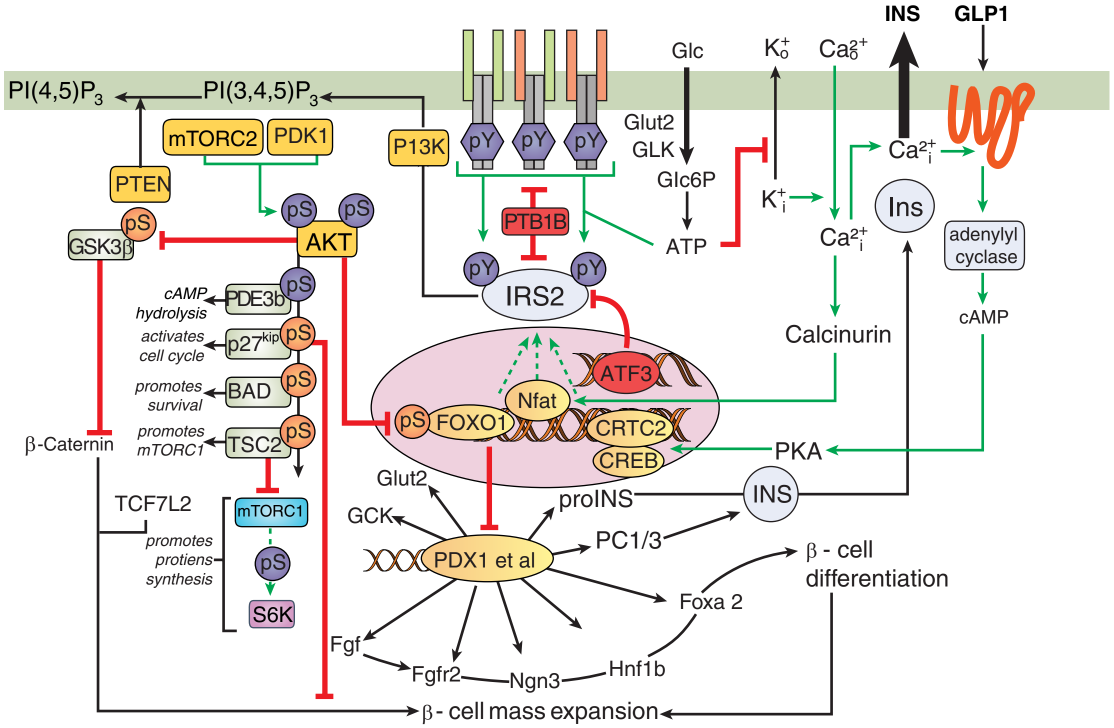


Fig. 30.8 **The integrative role of IRS2 signaling in pancreatic β-cell function.** The diagram shows the relation between the IRS2-branch of the insulin signaling pathway and upstream and downstream mechanisms regulating β-cell growth and function. Since InsR and IGF1R are constitutively active in β-cells, activation of GLP1→cAMP→PKA→CREB, glucose→Ca^{2+}→CRTC2, and calcineurin→NFAT induce IRS2 expression to regulate PI3K→AKT cascade, which places β-cell growth, function and survival under the control of glucose and incretins.

Regulation of Insulin Receptor Substrate Signaling by Ser/Thr Phosphorylation

IRS1 and IRS2 can be regulated through a complex mechanism involving phosphorylation of more than 50 serine/threonine residues (phospho-S/Ts) located in the long unstructured tail (see Fig. 30.5). Understanding how phospho-S/Ts regulate insulin signaling is challenging, because so many sites and mechanisms are involved.[151] Heterologous signaling cascades initiated by proinflammatory cytokines or metabolic excess—including TNFα (tumor necrosis factor-α), endothelin-1, angiotensin II, excess nutrients (free fatty acids, ceramides, amino acids and glucose) or endoplasmic reticulum stress—are implicated in IRS1/2 phospho-S/Ts, which associate with less insulin-stimulated tyrosine phosphorylation[151] These chronic heterologous mechanisms might contribute to glucose intolerance and pancreatic β-cell failure.[133] Regardless, most IRS1/2 phospho-S/Ts are stimulated by the PI3K→Akt→mTOR cascade during insulin stimulation.[151,152] Thus, IRS1/2 phospho-S/T is first and foremost a feedback mechanism that develops during chronic insulin stimulation, which can be coopted by metabolic stress to inhibit insulin signaling and dysregulate metabolic homeostasis (Fig. 30.8).[151] Thus, hyperinsulinemia may be the important physiologic mediator of insulin resistance in animals, and there is some experimental evidence that this is so.[153,154]

Mouse S307^{IRS1} (human S312^{IRS1}) is one of the best-studied phospho-S/Ts in IRS1, and is often used as barometer of insulin resistance.[151] Insulin can promote pS307^{Irs1} phosphorylation through the PI3K→AKT→mTORC1→S6K1 pathway (see Fig. 30.8).[152] In mice, free fatty acids promote pS307^{IRS1}, but insulin resistance and hyperinsulinemia are not excluded as the cause.[151] c-Jun N-terminal kinase (JNK1) or mTORC1 activated in obese mice promotes pS307^{IRS1} and other sites.[155,156] Although most, if not all, cell-based investigations support an inhibitory role for pS307^{IRS1}, mouse-based experiments are less convincing. Genetic knockin to replace S307^{IRS1} with alanine (A307^{IRS1}) increases fasting insulin and glucose levels while decreasing p110^{PI3K} binding to IRS1 PI3K.[157] During consumption of a high-fat diet, A307^{IRS1} mice exhibit more severe glucose intolerance and higher fasting insulin than control S307^{IRS1} mice. In mice, pS307^{IRS1} and other phospho-S/Ts may attenuate the detrimental effects of compensatory hyperinsulinemia to maintain a sensitive insulin response.[157] Careful analysis using molecular genetics is needed to establish the mechanisms involved.

Regulation of Insulin Receptor Substrate Degradation

Proteasome-mediated degradation regulates many biological processes including signal transduction, gene transcription, and cell cycle progression. Proteins targeted for destruction by the 26S proteasome are polyubiquitinated by various complexes composed of a ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin-protein ligase (E3).[158] IRS1 and IRS2 can be polyubiquitinated during chronic inflammatory states, nutrient excess, and hyperinsulinemia through various tissue-specific mechanisms.[159] The first pathway discovered involves suppressor of cytokine signaling 1 and 3 (SOCS1/3), which binds to IRS1 through its SH2 domain and to a ubiquitin E3 ligase through its SOCS box (see Fig. 30.9).[160,161] Mutations in the SOCS box prevent ubiquitinylation and degradation of IRS1 or IRS2,[160] and inhibition of SOCS1/3 expression by antisense oligonucleotides improves insulin sensitivity in obese and diabetic mice.[162] Cytokines like interferon gamma (INFγ) or IL6 activate Janus kinases (JAKs), which promote phosphorylation and dimerization of signal transducer and activator of transcription (STAT) factors that migrate into the nucleus to induce SOCS1/3 expression (see Fig.

30.9).[163] Angiopoietin-like 7 (Angptl7) also increases SOCS3 in obesity.[164] Thus, SOCS-mediated polyubiquitination of IRS1 or IRS2 can promote insulin resistance and glucose intolerance during infection, inflammation, or metabolic stress (see Fig. 30.9).

Other mechanisms of IRS degradation are coordinated by phospho-S/Ts. Cullin-RING E3 ubiquitin ligase 7 (CRL7) mediate IRS1 degradation downstream of phospho-S/T generated by the PI3K→AKT→mTORC1 cascade.[165] CRL7 complex contains cullin 7 (CUL7), a molecular scaffold that assembles F-box/WD repeat-containing protein 8 (FBW8) to recruit phosphorylated substrates, and RING-box protein 1 (Rbx1) associated with an E2-conjugating enzyme (see Fig. 30.9). FBW8 apparently binds to IRS1 through phospho-S/T residues generated by the mTORC1→S6K cascade—including human $pS307^{IRS1}$, $pS312^{IRS1}$, and $pS527^{IRS1}$, but possibly others—to mediate polyubiquitination of IRS1 that progresses to degradation (see Fig. 30.8).[165,166] This regulatory mechanism may be engaged during nutrient excess or hyperinsulinemia because it depends upon chronic hyperactivated mTORC1→SK6.[166]

Chronic consumption of high-calorie diets upregulates Cbl proto-oncogene B (CBLB), a RING-type E3 ubiquitin ligase that belongs to the casitas B-lineage lymphoma family of proteins.[111] CBL proteins share a conserved N-terminal region containing a tyrosine kinase–binding domain and a RING-finger domain to facilitate E3 ubiquitin ligase activity. Calorie excess induces carbohydrate-responsive element-binding protein (ChREBP) and SREBP1c, which upregulates myostatin (MSTN) in murine muscle to induce CBLB expression that drives insulin resistance through the polyubiquitination and degradation of IRS1.[167]

Regulation by Protein and Lipid Phosphatases

Phosphatases modulate IIS by dephosphorylating key proteins or lipids in the signaling cascade—including tyrosine-protein phosphatase nonreceptor type 1 (PTP1B, PTPN1), tyrosine-protein phosphatase nonreceptor type 2 (PTPN2, TCPTP), protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1), phosphatase and tensin homolog (pTEN), tensin-like C1 domain-containing phosphatase (C1-TEN), and Src homology 2 domain–containing inositol 5′–phosphatase 2 (SHIP2). PTP1B and PTPN2 are related phosphotyrosine phosphatases that dephosphorylate various RTKs, including the bisphosphorylated regulatory loop of the InsR[168]; however, their biological effects can be distinct, owing to a different time courses, tissue expression, and substrate selection.[169] *Ptp1b*$^{-/-}$ mice display increased insulin sensitivity, lower circulating insulin concentrations, and decreased pancreatic β-cell mass.[170] Physiologically, both PTP1B and PTPN2 can be inactivated by reactive oxygen species generated during insulin stimulation.[47,171] In pancreatic β-cells, the IRS2→PI3K→AKT cascade is important for growth, function, and survival.[172] Thus, inactivation PTP1B can maintain β-cells mass in mice lacking IRS2, which prevents the early onset of diabetes.[172] Regardless, without IRS2 even the *Irs2*$^{-/-}$·*Ptp1b*$^{-/-}$ mice lose β-cell mass between 8 and 9 months of age as IRS1 signaling fails to compensate. PTP1B is a selective inhibitor of leptin signaling (LepRb→JAK2), as it dephosphorylates JAK2, whereas PTPN2 dephosphorylates JAK1⁄3 but not JAK2.[168] Central nervous system (CNS) inhibition of PTP1B can protect against obesity, whereas peripheral inhibition of PTP1B promotes glucose tolerance.[173] Inhibition of PTP1B or TCPTP in the brain, or specific hypothalamic neurons, promotes insulin and leptin signaling and prevents diet-induced obesity, T2D, and fatty liver disease.[174] Regardless, targeting PTP1B and TCPTP for treatment has been problematic because of challenges in targeting inhibitor to specific tissues; however, intranasal targeting of PTP1B and TCPTP can increase leptin and insulin sensitivity and promotes weight loss by repressing feeding and increasing energy expenditure.[175]

PTEN is a potent negative regulator of insulin action and cellular proliferation, and is a frequently mutated gene in human cancers.[176,177] PTEN attenuates IIS by dephosphorylating $PI(3,4)P_2$ and $PI(3,4,5)P_3$ at the 3-position, which reduces the recruitment and activation of PDK1, AKT, and others (see Fig. 30.8).[177] PTEN heterozygosity can increase peripheral insulin sensitivity in *Irs2*$^{-/-}$·*Pten*$^{+/-}$ mice and normalizes glucose tolerance, as the small islets in these mice produce enough insulin until death between 10 and 12 months age. These experiments highlight the complex relation between nutrient homeostasis, insulin sensitivity and secretion, and cancer that can emerge in rodents and humans.

Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase (SHIP2), encoded by the *INPPL1* gene, attenuates insulin signaling by dephosphorylating the 5′–position of PtdIns(3,4,5) (see Fig. 30.8). Several genetic studies link SHIP2 to metabolic disorders, showing that polymorphisms in *INPPL1* may contribute to the pathogenesis of the metabolic syndrome, hypertension, and T2DM.[178] Despite this complexity, inhibition of SHIP2 may have therapeutic value. Metformin may increase insulin sensitivity by inhibiting SHIP2, which enhances glucose uptake while protecting renal podocytes from apoptosis in diabetic rodent models.[179]

Phosphoserine-directed phosphatases—including PP2A and PP1—have complex effects depending upon substrate targeting. PP2A is a widely expressed Ser/Thr protein phosphatase that forms a heterotrimeric complex with scaffold and regulatory subunits to target various subcellular locations and substrates.[180] PP2A can be a negative regulator of insulin's metabolic signaling by dephosphorylating and inactivating AKT or dephosphorylating and activating GSK3b.[181,182] PP2A also targets phospho-S/Ts in IRS1, which stabilizes IRS1 for tyrosine phosphorylation to enhance IIS.[183,184] PP1 also targets and dephosphorylates IRS1, which can promote IIS. Phosphoproteomic analysis of L6 skeletal muscle cells reveals the interaction of myosin phosphatase targeting subunit 1 (MYPT1)—a targeting subunit of protein phosphatase 1catalytic subunit beta (PP1cβ) with IRS1.[185] Activation of the MYPT1·PP1 complex by PKA promotes dephosphorylation of serine residues in IRS1 to increase tyrosine phosphorylation and stimulate PI3K→AKT cascade.[186]

miRNA-Mediated Posttranscriptional Regulation

miRNAs are short (~20 nucleotides) noncoding RNA molecules that negatively modulate gene expression through their specific binding within the 3′ untranslated region sequence of mRNA to inhibit translation or destabilize the target mRNA. Most of the proximal components of IIS can be regulated in a tissue-specific way by miRNAs.[187] LET7 miRNA interferes with many proximal IIS components—including IGF1R, InsR, IRS2, PIK3IP1, AKT2, TSC1, and RICTOR.[188] LET7 interference can be inhibited by the RNA-binding proteins Lin28a and Lin28b, which block production of mature LET7 to increase translation of the insulin signaling components. LET7 associates with gigantism, puberty delay, and glucose homeostasis by modulating insulin→PI3K→mTOR signaling and insulin sensitivity.[188]

Many other miRNAs display specificity against proximal IIS components during metabolic challenge in various tissues. Hepatic miR-424-5p, miR-15b, miR-195, and miR-96 increase in mouse liver on a high-fat diet, which associates with less InsR expression.[189-192] IRS1 and IRS2 are targeted by miRNAs in multiple peripheral insulin target tissues: miR-222 can suppress IRS1 in liver and adipose during high-fat diet and sucrose feeding[193]; resistin upregulates miR-145 to inhibit IRS1 levels[194]; miR-29a and miR-29c suppress IRS1 expression in muscle of T2DM patients or mice fed a high-fat diet[195]; and IRS1 can be downregulated in mouse endothelial cells by miR-126, which dysregulates angiogenesis.[196] Chronic angiotensin-II induced hypertension can increase the expression of miR-487b in rat aorta, which suppresses IRS1 expression and promotes hypertension-induced cardiovascular disease and formation of aortic aneurysms owing to the loss of medial smooth muscle.[197]

IRS2 expression is also suppressed by various miRNAs. The miR-126 targets IRS2 in pancreatic β cells, which can suppress islet growth.[198] Upregulation of miR-33b reduces hepatic IRS2→AKT signaling, whereas miR-135a reduces IRS2→AKT signaling in muscle of T2DM patients and *db/db* mice.[199] miR-26b appears to suppress PTEN to promote GLUT4 translocation in adipose and β cell function during obesity.[200] Other proximal IIS components are also modulated by various miRNAs—including SHIP2 (miR-205-5p), PDK1 (miR-375), AKT (miR-143—| ORP8), PP2A (miR-19b, miR-429, miR-29, and miR-155), and IGF1R.[187] Animal models suggest that modulating miRNA expression *in vivo* may be a beneficial strategy to restore insulin sensitivity and prevent the progression of T2DM.[201]

MOUSE GENETICS REVEAL TISSUE-INTEGRATED INSULIN SIGNALING

Introduction

KEY POINTS

- Genetic experiments in mice provide a robust and rigorous experimental platform to understand the role of insulin signaling throughout the body.

During the past 25 years, investigators used genetically manipulated mice to establish the molecular and integrated physiology of insulin action. Whole-body deletion of InsR, Irs proteins, or Akt isoforms confirmed the importance of IIS for metabolic and growth regulation. While informative, metabolic interactions between the various tissues complicates analysis of whole-body gene deletion. Thus, Cre-loxP technology was employed to investigate—albeit not perfectly—the role of InsR→Irs→Akt—| FoxO1 signaling in specific tissues or organs.[202,203] Even tissue-specific knockouts have limitations, as the elimination of IIS is rarely involved in common metabolic disease. Moreover, dysregulated insulin action in specific target tissues can influence metabolic regulation throughout the organism owing to crosstalk by secreted peptides, metabolites, or neuronal interconnections.[204,205] Regardless, Cre-loxP technology continues to be valuable to dissect IIS cascades.

Systemic Insulin Receptor Signaling Cascade

InsR and IGF1R have overlapping and distinct effects during development and adult life. Mice without InsR or the IGF1R die for different reasons shortly after birth.[206] IGF1R-deficient mice are born 50% smaller than normal littermates and die after a few days owing to developmental defects.[207] By contrast, mice lacking the InsR have nearly normal size but less adipose tissue mass at birth. InsR-deficient mice die a few days after birth with severe hyperglycemia, severe pancreatic failure, and ketoacidosis.[208] Thus, IGF1/2→IGF1R signaling is the principle growth regulatory pathway in fetal mice, whereas insulin→InsR signaling is essential for postnatal nutrient homeostasis.[209] InsR-deficient mice can be rescued genetically by transgenic expression of the InsR in the brain, liver, and pancreatic β cells.[210] Unlike mice, humans lacking functional InsR display intrauterine growth retardation, failure to thrive, and hypoglycemia for weeks until hyperglycemia develops.[209] This paradox may involve limited availability of gluconeogenic substrates or extreme hyperinsulinemia that may activate enough IGF1 receptors to generate some insulin-like actions.[211]

Work with transgenic and whole-body knockout mice confirms that the vast majority, if not all, of IIS is mediated through the IRS proteins.[212] IRS1 and IRS2 can display unique signaling properties in various tissues, apparently owing to different expression, receptor coupling, or regulation.[213] Systemic deletion of Irs1 produces small, insulin-resistant mice with nearly normal glucose homeostasis owing to β cell expansion and lifelong compensatory hyperinsulinemia.[214] These results suggest that mouse Irs1 is not essential for β cell growth during insulin resistance. By contrast, mice in which Irs2 is deleted display nearly normal body growth and even gain fat mass; however, male *Irs2*$^{-/-}$ mice develop life-threatening diabetes after 8 weeks of age owing to the progressive loss of β cell mass that disrupts compensatory hyperinsulinemia.[215] Female mice develop diabetes more slowly, and many, but not all, develop severe hyperglycemia by 6 months.[216] Compound deletion of Irs1 and Irs2 appears to be embryonic lethal; however, littermates retaining one allele of *Irs1* (*Irs1*$^{+/-}$•*Irs2*$^{-/-}$) or one allele of *Irs2* (*Irs1*$^{-/-}$•*Irs2*$^{+/-}$) can be born alive.[217] *Irs1*$^{+/-}$•*Irs2*$^{-/-}$ mice develop severe fasting hyperglycemia and die by 4 weeks of age as β cells fail to grow or survive. By contrast, *Irs1*$^{-/-}$•*Irs2*$^{+/-}$ mice are very small but have enough circulating insulin for glucose tolerance.[217] Thus, mouse Irs1 and Irs2 are essential for development and nutrient homeostasis, and some Irs2 is essential for β cell growth function and survival.

Hepatic InsR—| FoxO1 Signaling

The liver in an important site of insulin action that plays a role in systemic glucose and lipid homeostasis.[202] InsR$^{L/L}$·Cre^{Alb} (LIRKO; hepatic albumin Cre driver) mice display moderately elevated fasting glucose and severe postprandial hyperglycemia and glucose intolerance.[218] LIRKO mice also develop severe hyperinsulinemia owing to a combination of increased insulin secretion and reduced hepatic insulin degradation.[219] LIRKO mice display reduced levels of circulating free fatty acids and triglycerides on ordinary chow diets[202,218]; however, when fed an atherogenic diet, LIRKO mice develop dyslipidemia that can progress to atherosclerosis.[220] InsR deletion dysregulates hundreds of hepatic genes, including reduced glucokinase, hexokinase 4 (GCK) and elevated phosphoenolpyruvate carboxykinase 1 (PCK1), G6PC, and pyruvate kinase (PK1).[202,218] Chronic hyperinsulinemia in LIRKO mice exacerbates peripheral insulin resistance, whereas streptozotocin injections to reduce insulin secretion improve peripheral insulin resistance but fail to suppress HGP.[221] Thus, heterologous mechanisms dysregulated by hepatic insulin resistance—including the systemic effects of hyperinsulinemia—may exacerbate HGP through indirect mechanisms that may promote adipose insulin resistance and delivery of excess glycerol and nonesterified fatty acids (NEFAs) to the liver for HGP.[205,222,223]

For over a decade, we have modeled hepatic insulin resistance by genetic inactivation of hepatic Irs1 and Irs2 in LDKO (Irs1$^{L/L}$·Irs2$^{L/L}$·Cre^{Alb}) mice[205,224-226]. LDKO mice display unsuppressed HGP, hyper insulinemia, glucose intolerance, and diabetes.[205,224-226] At first, these results suggest that hepatic Irs1/2 mediate systemic nutrient homeostasis through hepatic insulin signaling; however, WAT, BAT, and skeletal muscle are also insulin-resistant in LDKO mice, suggesting that dysregulated hepatic metabolism manifests systemically.[205,227] Remarkably, metabolic health is restored upon hepatic inactivation of FoxO1 in LTKO (LDKO·FoxO1$^{L/L}$) mice, even though the liver is mechanistically unresponsive to insulin.[205,224-227] These findings are confirmed and extended with compound hepatic-specific InsR·FoxO1- or Akt1·Akt2·FoxO1-knockout mice.[90,228-230] Clearly, genetic disruption of hepatic FoxO1 substantially normalizes the expression of hundreds of dysregulated hepatic genes.[90,224] Within such mice, FoxO1-dependent gene expression or metabolites generated in hepatocytes and circulating systemically could promote peripheral insulin resistance and the delivery of metabolic intermediates to the liver—including excess glycerol and free fatty acids from adipose tissue.[231] Comprehensive analyses of the LDKO- and LTKO-mice revealed that follistatin (Fst)—best known for its modulation of TGF-β superfamily members[232,233]—is a key FoxO1-dependent hepatokine that can promote WAT lipolysis during hepatic insulin resistance.[205] Thus, while failing to restore hepatic insulin signaling *per se*, disruption of hepatic FoxO1 in LTKO mice and similar models may restore glucose tolerance by normalizing hepatokine secretion.[224,228,229,234]

Insulin and Insulin-Like Growth Factor Signaling and Glucose Homeostasis In Skeletal Muscle

IIS in skeletal muscle is initiated by the activation of the insulin and/or IGF1 receptor tyrosine kinases, which can exist as hybrids linked to the downstream pathways through the IRS1 and IRS2 branches of the cascade.[235] Some, but not all, aspects of metabolic disease emerge upon deletion of skeletal muscle InsR in MIRKO (InsR$^{L/L}$·Cre^{Mck}; muscle creatine kinase Cre driver) mice—including mild obesity and elevated circulating free fatty acids and triglycerides; however, glucose and hyperinsulinemia never develop.[202,236] Moreover, deletion of InsR and IGF1R in MIGIRKO (InsR$^{L/L}$·IGF1R$^{L/L}$·Cre^{ACTA1}; human skeletal muscle actin Cre driver) mice results in more than 60% less muscle mass, while glucose and insulin tolerance remain normal, owing, at least in part, to increased basal glucose uptake.[237] Consistent with these results, MDKO (Irs1$^{L/L}$·Irs2$^{L/L}$·Cre^{Mck}) mice also fail to develop hyperglycemia and glucose intolerance during progressive and fatal skeletal and cardiac muscle autophagy.[113] Akt phosphorylation is completely lost in MDKO mice, suggesting that InsR/IGF1R signaling is its major agonist.[113] Isolated skeletal muscles from MDKO mice show elevated basal but absent insulin-stimulated glucose uptake, while glucose metabolism shifts to lactate production—which elevates the AMP/ATP ratio to stimulate AMP-activated protein kinase (AMPK) phosphorylation of AS160/Tcb1D4 and TBC1D! and insulin-independent glucose uptake (see Fig. 30.7).[113]

By comparison, muscle-specific deletion of GLUT4 *(Slc2a4)* dysregulates glucose homeostasis and promotes systemic insulin resistance to a greater extent than deletion of proximal insulin signaling components.[238] MG4KO (GLUT4$^{L/L}$·Cre^{Mck}) mice develop hyperglycemia, glucose intolerance, and insulin resistance by 8 weeks of age.[239] These mice also display dysregulated glucose metabolism in adipose and liver. Hyperglycemia owing to diminished muscle glucose utilization appears to drive heterologous insulin resistance, because normalization of circulating glucose by kidney excretion can restore insulin action in adipose and liver.[238,240]

Insulin and Insulin-Like Growth Factor Signaling in Cardiac and Vascular Smooth Muscle

MDKO (Irs1$^{L/L}$·Irs2$^{L/L}$·Cre^{Mck}) mice display cardiac arrest between 3 and 4 weeks of age owing to severe left ventricular failure.[113] Because a single allele of either *Irs1* or *Irs2* in cardiac muscle can prevent sudden death, we conclude that either Irs1 or Irs2 mediates enough IIS to prevent excess autophagy and fatal cardiac muscle loss.[113] Regardless, pressure overload–induced cardiac hypertrophy and heart failure are associated with insulin resistance and hyperinsulinemia, which can exacerbate left ventricular remodeling.[241] Reducing cardiac insulin signaling by lowering circulating insulin or reducing InsR→AKT1 signaling can attenuate heart failure.[242] Experiments with CIrs1KO (Irs1$^{L/L}$·Cre$^{\alpha rer}$; myosin, heavy polypeptide 6, cardiac muscle, alpha Cre driver) or CIrs2KO mice suggest that Irs1 may promote a proinflammatory response following pressure overload.[241] CIrs1KO hearts display less Akt1 activity and mild cardiac hypertrophy during pressure overload; however, CIrs2KO hearts display severe cardiac dysfunction that can be attenuated by *Akt1* haploinsufficiency.[241] Thus, IIS→IRS1 signaling can exacerbate heart failure in mice, which may explain how hyperinsulinemia during T2DM aggravates cardiovascular disease.

Vascular disease is a major cause of morbidity associated with insulin resistance and T2DM owing to a greater risk of myocardial infarction compared against normal individuals.[243] Endothelial cells (ECs) and vascular smooth muscle cells (VSMCs) are the major constituents of medium- and large-sized arteries with essential functions in vascular homeostasis. The progression of atherogenesis can involve dysregulated crosstalk between immune cells, ECs, and VSMCs.[244] Specific deletion of InsR or IGF1R in VSMCs show that InsR—but not IGF1R—mediates VSMC proliferation and intimal hyperplasia following intimal injury.[245] The PI3K→AKT cascade phosphorylates eNOS, which promotes vascular relaxation to optimize blood flow (see Fig. 30.6). Without the endothelial InsR, atherosclerotic lesion size increases more than 2-fold in atherosclerosis-prone apolipoprotein E–deficient (*Apoe*$^{-/-}$) mice.[246] Moreover, the extent of the aortic intima occupied by the lesion increases in *Irs2*$^{+/-}$·*ApoE*$^{-/-}$ mice compared with *ApoE*$^{-/-}$ mice.[247] Deletion of myocardial InsR can decrease VEGF expression and capillary density and reduce reactive angiogenesis during ischemia reperfusion damage or myocardial infarction.[248] Because Irs2 is more abundant than Irs1 in endothelial cells, endothelial cell–specific Irs2 knockout (ETIrs2KO) impairs eNOS phosphorylation, muscle insulin delivery, and glucose uptake.[203,249] Taken together, managing InsR→IRS2 signaling in the endothelium may benefit cardiovascular complications.

Adipose Insulin Signaling

Insulin signaling promotes adipogenesis and stimulates glucose influx, lipid synthesis, and antilipolysis in mature adipose.[250] WAT is important for the storage of postprandial glucose as triglyceride and the secretion of signaling factors that regulate appetite and energy homeostasis.[154,251] WAT mainly expresses IRB without detectable IGF1R; however, both receptors play a role in adipogenesis. Interpretation of genetic deletion of adipose Insr$^{L/L}$ is complicated by the Cre drivers used to generate the various experimental mice—including Cre^{aP2} (FIRKO mice; fatty acid binding protein 4 Cre driver) verses Cre^{Adipo} (F-IRKO mice; adiponectin Cre driver).[252,253] FIRKO mice display variable recombination efficiency in fat depots with some off-target events, whereas F-IRKO mice have more efficient InsR deletion across fat depots with few or no off-target effects.[254] With respect to these caveats, FIRKO mice display beneficial metabolic effects, including systemic insulin sensitivity, normal glucose tolerance, and less fat mass on ordinary chow diets.[252] FIRKO mice also have a longer lifespan, suggesting that leanness and insulin sensitivity are associated with longevity.[255] By contrast, F-IRKO mice display lipodystrophy with hepatomegaly, steatosis, and increased enzymes of *de novo* lipogenesis.[253] F-IRKO mice develop a full spectrum of NAFLD that can progress to NASH, fibrosis, and liver dysplasia—which is more in line with the expected effects of adipose insulin resistance.[154] These phenotypes are exacerbated upon compound deletion of InsR and IGF1R by Cre^{Adipo}, which promotes lipodystrophy of WAT and BAT accompanied by diabetes, insulin resistance, increased β cell mass, and ectopic lipid accumulation.[254] Thus, the InsR appears to be essential for the formation and maintenance of WAT, whereas both InsR and IGF1R contribute to BAT mass development and thermogenesis.

Insulin-regulated GLUT4-mediated glucose transport is rate-limiting for glucose influx into adipose tissues under normal glucose and insulin concentrations. GLUT4 translocation to the plasma membrane is extremely sensitive to insulin in adipocytes, as submaximal activation of AKT2 is enough to maximally stimulate glucose transport. Before high-fat diet–induced insulin resistance inhibits AKT phosphorylation, GLUT4 translocation is strongly inhibited, suggesting that decreased glucose utilization may reside downstream of proximal IIS and AKT activity.[154] Increased adipose GLUT4 expression can rescue systemic insulin resistance of mice fed the high-fat diet.[256]

Adipose triacylglycerol (TAG) storage is regulated by circulating factors with opposing effects on lipolysis—β-adrenergic receptors (βAR1/2/3) and insulin. Activation of βAR1/2/3 stimulates cAMP production to promote PKA-mediated phosphorylation of perilipin 1 (PLIN1), which releases abhydrolase domain–containing 5 (ABHD5) to fully activate patatin-like phospholipase domain–containing 2 (ATGL/PNPLA2), hydrolyzing one fatty acid from TAG.[257] PKA also phosphorylates hormone-sensitive lipase (HSL) to promote its translocation to PLIN1 to hydrolyze FA from diacylglycerol (DAG) and produce monoacylglycerol, which is hydrolyzed by monoacylglycerol lipase (MGL) to release the final FA and glycerol.[258] Insulin inhibits lipolysis to promote fatty acid storage as TAG in adipose lipid droplets[258,259]; however, the mechanism is only partly understood, as deletion of AKT2 only partly impairs the ability of insulin to suppress lipolysis.[258,260] Regulated lipolysis is critical, as it controls the delivery of NEFA and glycerol to the liver, which stimulates HGP.[154] Insulin inhibits lipolysis by promoting hydrolysis of cAMP via an uncertain mechanism. Originally, AKT was thought to activate PDE3b by phosphorylation; however, the direct phosphorylation of PDE3B by AKT might not be required.[260] Instead, AKT phosphorylates ab hydrolase domain-containing protein 15 (ABHD15), which binds and stabilizes PDE3B for the antilipolytic action of insulin.[258,261,262]

Insulin Receptor Substrate 2 as a Gateway to β Cell Function

The capacity of pancreatic β cells to maintain glucose homeostasis during chronic physiologic and immunologic stress is important for cellular and metabolic homeostasis.[263] Pancreatic β-cells have a special place in nutrient homeostasis as the unique source of insulin secretion, and like other cells require IIS for growth, function, and survival (see Fig. 30.8). β cells appear to be regulated by multifactor transcriptional control of IRS2 through the action of FOXO1/3, NFAT, and CREB•CRTC2 (see Fig. 30.8).[264] In β-cells, FOXO can account for as much as 80% of the basal IRS2 expression[265]. Because the IRS2→PI3K→AKT cascade phosphorylates and inhibits FOXO, insulin or IGF1 have inhibitory effects upon FOXO-mediated transcription of IRS2. Because β-cell mass and function must be protected during chronic nutrient excess that provokes excess insulin secretion, other mechanisms—especially glucose-stimulated Ca^{2+} influx and cAMP production— promote IRS2 expression to avoid progressive β-cell failure (see Fig. 30.8). In addition to its immediate role in insulin secretion, glucose→calcium influx activates calcineurin, which dephosphorylates NFAT to facilitate its entry into the nucleus, where it induces expression of *IRS2* and other genes (see Fig. 30.8).[266] Glucose, GLP1 (glucagon-like peptide-1) and other GPCR agonists also increase the cAMP in β-cells, which has many important effects, including the activation of CREB·CRTC2 that promotes *IRS2* transcription.[267,268] Through this mechanism, the role of insulin/IGF1 to trigger downstream IIS in β-cells is replaced by indirect control through glucose, incretins, or neuronal signals—the physiologically relevant regulators of pancreatic β-cell function (Fig. 30.8). IRS2 expression through these mechanisms maintains PDX1 action, which is essential for β cell growth, function, and survival.[269] Compounds that augment IRS2 expression may provide treatments for β cell failure during insulin resistance and the progression of T2DM.[263]

Insulin/IGF Signaling in Neurodegenerative Disease

Insulin resistance associated with obesity promotes T2DM, cardiovascular disease, certain cancers and—with aging—cognitive decline and Alzheimer's disease and related dementia (ADRD).[19,270,271] How dysregulated brain IIS or peripheral insulin resistance and its metabolic sequelae promote cognitive decline is unclear and difficult to study.[19,272,273] In the brain, IRS1, IRS2, and IRS4 link IRA, insulin-like growth factor-1 receptor (IGF1R), and their hybrids (IRA•IGF1R) to downstream pathways that mediate IIS. Neurotrophic tyrosine kinase (TrkA) and TrkB—which regulate synaptic strength and plasticity—also engage IRS in neurons.[274,275] Regardless, mechanisms linking peripheral insulin resistance and central IIS to cognitive decline and ADRD are elusive. Numerous factors regulate the strength or duration of IIS, including variable expression, stability, and localization of the proximal downstream components—IRS1/2/4, PI3K, and AKT1/2.[134,276,277]

Central IIS, modified by peripheral insulin sensitivity, provides an important link between metabolic health and sustained hippocampal function.[19,270] Impaired central IIS in Alzheimer's disease brain samples,[278,279] as well as mouse models of Alzheimer's disease,[280] suggests that enhanced central IIS could be therapeutic—a hypothesis that has prompted nasal delivery of insulin to treat ADRD.[270] Clinical studies associate neuronal pS/T^{IRS1} in humans with ADRD[278,280,281]—but the mechanistic basis and import of these observations for central insulin resistance remain unclear.[152,157,282]

The role of central IIS is unsettled because reducing IIS→Irs2 may increase lifespan and promote memory formation in mice[283-286]—even while producing obesity, peripheral insulin resistance, and hyperinsulinemia.[215,283,287-291] Similarly, decreased *Irs2* gene dose improves motor performance and extends lifespan in R6/2 Huntington disease model mice, associated with reduced neuronal oxidative stress and enhanced autophagy and mitochondrial function.[292] Less Irs2 also increases nuclear FoxO1 in R6/2-mice,[292] which may prevent age-related axonal degeneration.[293]

An important question is whether IIS→IRS modulates adult hippocampal neurogenesis directly or through peripheral metabolic dysregulation. New neurons are continuously generated in the dentate gyrus (DG) of the hippocampus.[294-296] DG neurogenesis in humans can persist into the ninth decade of life—but declines in ADRD.[297,298] Newly formed neurons may sustain and enhance memory or facilitate further memory acquisition—including better "pattern separation" and spatial memory.[295,299,300] DG neurogenesis is augmented by physical activity and extrinsic (e.g. TNFα, TGFβ, INFγ, insulin/IGF[296,301]) or intrinsic factors (Sox2, Hes1/5, REST, FoxOs, NeuroD1, CREB, Prox1). DG neurogenesis measurably improves rodent learning in the Morris water maze,[302] so it will be important to understand the role of IIS in these processes. Thus, a fuller investigation of hippocampal IIS is warranted, along with behavioral experiments that reveal effects of peripheral IIS on learning and memory.

INSULIN RESISTANCE AND METABOLIC DISEASE

Introduction

KEY POINTS

- It is challenging to understand the molecular basis of insulin resistance, because physiologic mechanisms like hyperinsulinemia and hyperglycemia are tightly integrated with multiple signaling cascades, dyslipidemia, inflammatory cytokines, endoplasmic reticulum stress, and hepatokine secretion.

Insulin resistance is considered the fundamental trigger of T2DM; however, transient insulin resistance may be a feedback mechanism to protect cells and tissues from acute effects of overnutrition.[154,303] Regardless, significant gaps exist in our knowledge of the molecular mechanisms that cause insulin resistance and the progression of T2DM. Several pathologic mechanisms associated closely with overnutrition and inactivity promote insulin resistance and T2DM—including compensatory hyperinsulinemia and β cell dysfunction; InsR and

IRS1/2 degradation; elevated circulating glucose, circulating NEFA, and dysregulated proteostasis; local or circulating proinflammatory cytokines, adipokines, myokines, or hepatokines; oxidative and endoplasmic reticulum (ER) stress; and endothelial dysfunction.[11,304,305] Regardless, in humans and animals it is challenging to understand the mechanistic hierarchy and their integration for disease progression.

Clinical staging suggests that early insulin resistance in skeletal muscle—the primary site of glucose disposal in humans—may initiate the progression of T2DM.[111] Compatible with this hypothesis, healthy human subjects who display skeletal muscle insulin resistance exhibit decreased incorporation of glucose into muscle glycogen and increased hepatic triglyceride synthesis[306]; however, animal models of pure skeletal muscle insulin resistance—including mice that lack muscle InsR, IGF1R or Irs1/2—show surprisingly mild glycemic changes.[113,236] Thus, uncoupled effector systems such as insulin-stimulated glucose influx and oxidation—rather than dysregulated proximal InsR signaling—may emerge as the cause of metabolic disease.[154] Moreover, hepatic insulin resistance plays an important role in uncontrolled HGP, but how cell-autonomous mechanisms—including impaired IRS1/2 or AKT1/2 phosphorylation—and dysregulated interorgan metabolic flux and communication contribute are unresolved.[90,204,205,218,307]

Chronic Hyperinsulinemia

Chronic overnutrition and inactivity are usually associated with obesity and insulin resistance that can progress to T2DM when sufficient insulin is unavailable to promote nutrient storage and suppress HGP. The mechanisms are difficult to resolve, because insulin resistance triggers hyperinsulinemia, which in turn promotes insulin resistance and obesity.[154] Chronic compensatory hyperinsulinemia overactivates the IR→IRS cascade that can promote feedback mechanisms to downregulate the cell-surface InsR concentration and uncouple IRS→PI3K→AKT from its effector systems (Fig. 30.9). As obesity progresses, hepatic IRS2 expression decreases due to hyperinsulinemia, which impairs insulin signaling in the liver.[308] As the IRS2:IRS1 ratio falls, IRS1 continues

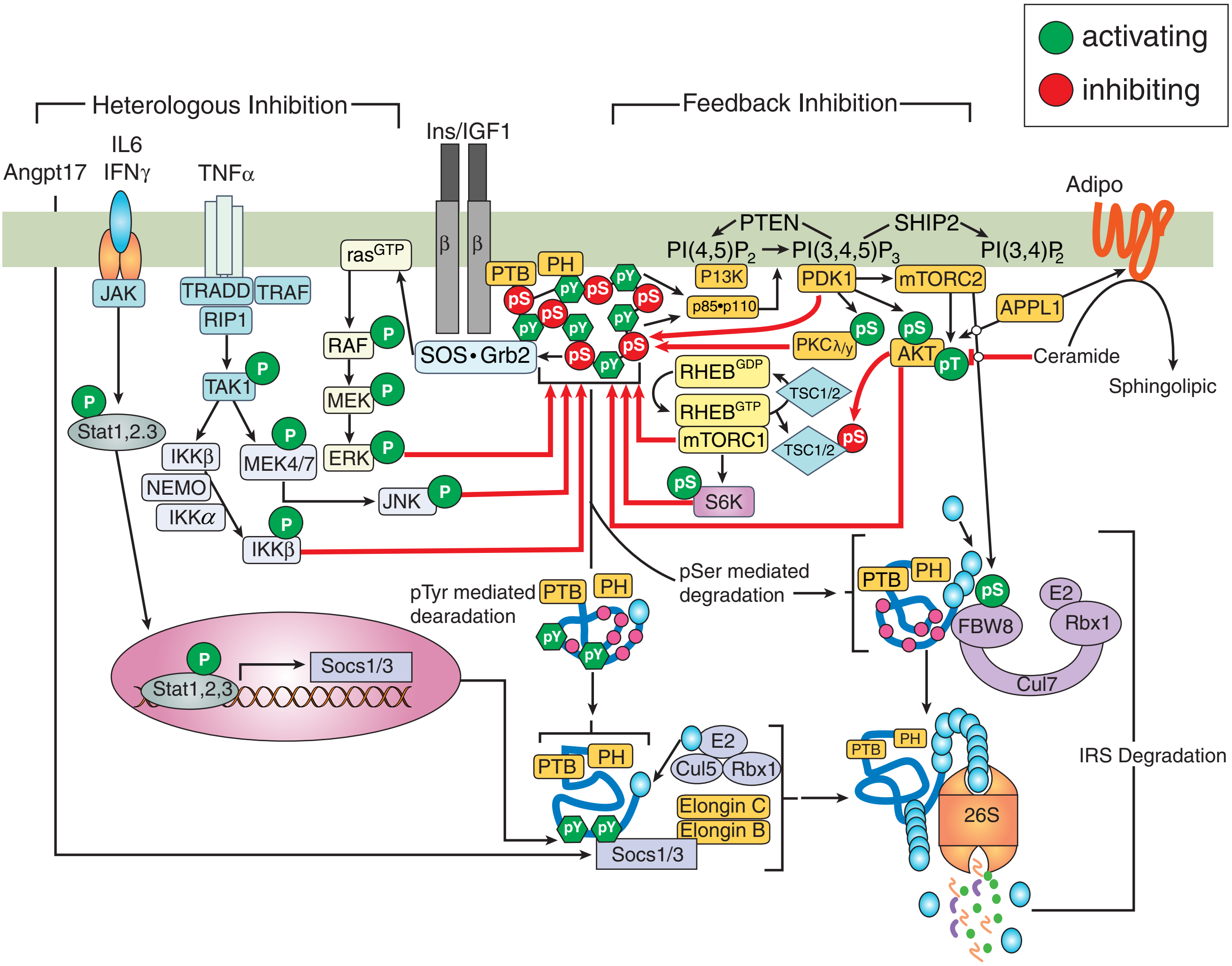

activating
inhibiting
Heterologous Inhibition
Feedback Inhibition
Angpt17
IL6
IFNγ
TNFα
Ins/IGF1
Adipo
β
β
PTEN
SHIP2
PI(4,5)P2
PI(3,4,5)P3
PI(3,4)P2
rasGTP
PTB
PH
P13K
p85•p110
PDK1
mTORC2
APPL1
JAK
TRADD
TRAF
RIP1
TAK1
RAF
MEK
ERK
SOS•Grb2
PKCλ/γ
AKT
pT
Ceramide
Sphingolipic
RHEBGDP
RHEBGTP
TSC1/2
mTORC1
TSC1/2
S6K
Stat1,2,3
IKKβ
NEMO
IKKα
MEK4/7
JNK
IKKβ
P
pY
pS
pTyr mediated dearadation
pSer mediated degradation
PTB
PH
E2
Rbx1
FBW8
Cul7
Stat1,2,3
Socs1/3
PTB
PH
E2
Cul5
Rbx1
Elongin C
Elongin B
Socs1/3
26S
IRS Degradation

Fig. 30.9 Schematic diagram of heterologous and feedback regulation of insulin signaling. Various kinases in the insulin signaling cascade mediate feedback inhibition of IRS, including PKB, mTOR, S6K, ERK, AKT, and atypical PKC isoforms.[152] Other kinases activated by heterologous signals, including lipids such as ceramide, are also involved. Adiponectin signaling may promote insulin signaling by driving the conversion of ceramide to nontoxic sphingosine. Serine phosphorylation of IRS1 can recruit CRL7, which can promote ubiquitinylation and degradation of IRS1 through the 26S proteasome. Many proinflammatory cytokines cause insulin resistance through SOCS1 or SOCS3, which target phosphotyrosine-containing proteins like IRS1 or IRS2 for ubiquitination by a BC-containing ubiquitin ligase (E3) and degradation.[347-349]

to drive hepatic lipogenesis, unless feedback mechanisms further inhibit IR→IRS1 signaling.[226] Hepatic insulin resistance can accelerate hyperinsulinemia to impair insulin signaling in other tissues as well. Although various mechanisms are probably involved, hyperglycemia can provoke pancreatic β cell mass expansion by glucose→calcium→NFAT–mediated IRS2 expression.[309] The important role of hyperinsulinemia to promote insulin resistance is validated by reduced weight gain and better glucose tolerance in mice with genetically attenuated hyperinsulinemia during chronic high-fat diet feeding.[153,154,310]

Nutrient Excess

The struggle to store excess nutrients by nonoxidative disposal as muscle glycogen or adipose triglyceride can promote metabolic disease and T2DM.[11] Even while elevated average insulin may promote enough muscle glycogen storage to defend against hyperglycemia, it can increase hepatic lipogenesis, WAT expansion, and hepatic steatosis.[11] The accumulation of nutrient-derived toxic metabolites—including NEFA, DAG, or ceramides—within insulin-sensitive tissues can activate novel PKC isoforms that mediate Ser/Thr phosphorylation of INSR or IRS1, inhibiting the proximal signaling cascade.[305] Moreover, while insulin action decreases, NEFA and glycerol flux from insulin-resistant adipose can promote liver gluconeogenesis by mass action and allosteric mechanisms that exacerbate average circulating glucose and insulin secretion.

Nutrient excess and progressive insulin resistance can dysregulate muscle proteostasis to exaggerate circulating branched-chain amino acids levels, which stimulate mTORC1→S6K/GRB10 activity that can further inhibit IR→IRS1/IRS2 signaling.[311,312] Chronic mTORC1 activity also promotes synthetic processes that exacerbate ER stress, activating an unfolded protein response that can cause insulin resistance.[17,313-315] Thus, chronic mTORC1 activity by amino acid excess can feedback inhibit the InsR→IRS1/IRS2 cascade to impair glycogen storage (muscle) and antilipolysis (WAT) while stimulating hepatic lipogenesis to promote T2DM.

Proinflammatory Cytokines

Low-grade chronic inflammation occurs in adipose tissue, liver, pancreatic islets, vasculature, and circulating leukocytes during obesity and T2DM.[17,304,316] The expanding adipose compartment creates a microenvironment conducive to inflammation owing to hypoxia, adipose cell death, and dysregulated adipokines, including decreased adiponectin and increased leptin, resistin, and RBP4.[303,317] Proinflammatory M1-like adipose-tissue macrophages (ATMs) accumulate in expanding WAT[318] and contribute to local and circulating concentrations of proinflammatory cytokines—including monocyte chemoattractant protein 1 (MCP1), interleukin 6 (IL6), and TNFα (tumor necrosis factor-α.[319,320] TNFα promotes glucose intolerance, and both TNFα and IL6 can induce lipolysis.[320]

TNFα can promote insulin resistance during nutrient excess and obesity as it activates JNK1 (Jun N-terminal Kinase-1) and IKKβ (inhibitor of nuclear factor kappa B kinase subunit beta) in insulin-sensitive tissues.[304,305] JNK appears to co-opt feedback mechanisms ordinarily activated by chronic hyperinsulinemia, especially S/T-phosphorylation of IRS1.[151,152] Inhibition of JNK activity in obese mice using kinase inhibitors increases insulin sensitivity and improves glucose homeostasis; moreover, Jnk1-deficient mice are protected from obesity and insulin resistance caused by high-fat diet feeding.[155,321] Regardless, tissue-specific Jnk1 ablation in various peripheral tissues—including adipose, liver, muscle, and myeloid—does not reduce obesity; however, genetic deletion of TNFα in the CNS reduces energy expenditure that associates with increased iodothyronine 5'-deiodinase (DIO2).[322] High-fat diets cause JNK-mediated DIO2 expression, which decreases the hypothalamic–pituitary–thyroid (HPT) axis and promotes obesity.[323] Regardless, administration of TNFα inhibitors to insulin resistant humans does not improve glucose homeostasis, despite reduction in some parameters of systemic inflammation.[324] Thus, the physiologic relevance of the TNFα→JNK cascade as a cause of insulin resistance in humans remains unclear.[305]

Cytokines can have contradictory effects upon metabolic regulation. During acute injury, IL6 secreted from T-cells and macrophages is beneficial, as it mobilizes stored nutrients from muscle and adipose tissues to provide substrates and energy to fight infection and repair tissue.[320,325] Moreover, physical activity promotes the secretion IL6 from skeletal muscle, which can mediate anti-inflammatory and immunoregulatory effects to reduce circulating TNFα.[325] Infusion of recombinant human IL6 into healthy individuals enhances whole-body insulin sensitivity, fat oxidation, and reesterification of NEFAs.[325,326] Chronic genetic overexpression of IL6 in mice fed a high-fat diet reduces daily food consumption, adiposity, and circulating leptin, while increasing energy expenditure.[327] Although IL6 stimulates the STAT3→SOCS3 cascade—which can inhibit IR→IRS1 signaling—it protects hepatic IR→IRS2 signaling in mice fed a high-fat diet to improve hepatic insulin sensitivity.[327] Thus, although circulating IL6 increases in genetically diabetic LDKO mice ($Irs1^{L/L}\cdot Irs2^{L/L}\cdot Cre^{Alb}$), it has no beneficial effect without Irs2.[205]

Hepatic Follistatin Links Liver and White Adipose Tissue Insulin Resistance

Proteins exclusively or predominantly produced and secreted from a specific tissue can have paracrine or endocrine actions. Adipokines are perhaps the best-known, although other classes of tissue-specific secreted proteins have been identified—including myokines, lipokines, and hepatokines.[350] Over the past decade, evidence points to a dysregulated pattern of hepatokine secretion, which directly affects metabolism.{Stefan, 2013, 23337953} Recent work suggests that FOXO1-induced hepatic follistatin (FST) can upregulate HGP by promoting WAT insulin resistance and lipolysis.[205] Viral suppression of hepatic Fst in LDKO ($Irs1^{L/L}\cdot Irs2^{L/L}\cdot Cre^{Alb}$) mice restores WAT insulin sensitivity and glucose tolerance, whereas its over-expression in livers of wild-type or healthy LTKO ($LDKO\cdot FoxO1^{L/L}$) mice has opposite effects.[205] Thus, excess hepatic Fst impairs glucose homeostasis regardless of the state of hepatic insulin signaling. These data add to the growing body of evidence that hepatokines can drive peripheral insulin resistance, which promotes T2DM.[329-331]

FST may be the first hepatokine identified to induce adipose tissue lipolysis, thereby shifting NEFA and glycerol from WAT to the liver.[332] Subjects in the Tübingen Family Study/Tübingen Lifestyle Program cohort were interrogated to establish the relation between FST and nutrient homeostasis in patients.[332,333] Magnetic resonance was used to estimate body fat distribution, and ^1H–magnetic resonance to estimate liver fat content.[330] FST, glucose, insulin, and NEFA were measured in serum, and an oral glucose tolerance test (OGTT) was used to measure whole-body insulin sensitivity (IS_{body}) and adipose insulin sensitivity ($IS_{adipose}$).[333] Univariate analysis shows that circulating FST is independent of sex and age but positively associated with body mass index ($r = 0.27$, $P = .0001$), total body fat mass ($r = 0.28$, $P < .0001$), visceral fat mass ($r = 0.18$, $P = .008$), liver fat content ($r = 0.23$, $P = .0008$), and circulating NEFA during the OGTT at 0 minutes ($r = 0.22$, $P = .001$), 60 minutes ($r = 0.28$, $P < .0001$), and 120 minutes ($r = 0.28$, $P < .0001$).[332] After adjusting for age, sex, and total body fat mass, $IS_{adipose}$ correlates negatively with circulating FST. Moreover, liver fat content correlates positively with circulating FST (standard $\beta = 0.17 \pm 0.07$, $P = .009$); however, this correlation disappeared after adjusting for leg fat mass ($r = 0.09 \pm 0.06$, $P = .14$).[332] Thus, FST may impair IS_{adipo} in leg fat—an important compartment of benign fat storage.[334,335] FST also increases significantly in subjects with NAFLD, suggesting that impaired IS_{adipo} may shift NEFA to the liver to promote NAFLD.[336,337]

Endoplasmic Reticulum Stress and the Unfolded Protein Response

The ER is a network of interconnected membrane-enclosed tubes that are continuous with the outer membrane of the nuclear envelope. Smooth ER lacks ribosomes and functions in lipid and cholesterol biosynthesis, carbohydrate metabolism, and detoxification; rough ER is coated with ribosomes, where protein synthesis is especially prominent.[338,339] In many cases, disulfide bond formation contributes to protein folding, which requires nutrient homeostasis to support an oxidizing environment.[340] The unfolded protein response is activated when circumstances disrupt protein folding, including glucose and energy deprivation, cholesterol accumulation, viral infection, and other factors that dysregulate protein synthesis.[338,339] Distinct branches of the unfolded protein response are initiated by type-I transmembrane kinases—PKR-like endoplasmic reticulum kinase (PERK) and inositol requiring enzyme-1 (IRE1)—and activating transcription factor-6 (ATF6).[338,339] In the liver, PERK reduces global protein synthesis by phosphorylating eIF2a, while selective translation can lead to obesity and insulin resistance.[341,342] IRE1 signaling activates JNK in the liver and adipose tissues to increase Ser/Thr phosphorylation of IRS1.[343] ATF6 upregulates expression of the X-box binding protein 1 (XBP1) transcription factor to increase the protein folding capacity that compensates during protein overload. ATF6→XBP1 also interacts with FOXO1 to direct it toward proteasome-mediated degradation, which can contribute to systemic insulin sensitivity.[344] Attenuation of ER stress in obese and diabetic mice through the use of chemical chaperones—agents that promote ER folding capacity—improves insulin sensitivity and glucose tolerance.[345,346]

REFERENCES

1. Weiss MA, Lawrence MC. A thing of beauty: structure and function of insulin's "aromatic triplet". *Diabetes Obes Metab*. 2018;20:51–63.
2. Kenyon C. The plasticity of aging: insights from long-lived mutants. *Cell*. 2005;120:449–460.
3. Deng D, Sun P, Yan C, et al. Molecular basis of ligand recognition and transport by glucose transporters. *Nature*. 2015;526:391–396.
4. Belfiore A, Malaguarnera R, Vella V, et al. Insulin receptor isoforms in physiology and disease: an updated view. *Endo Rev*. 2017;38:379–431.
5. Clemmons DR. Involvement of insulin-like growth factor-I in the control of glucose homeostasis. *Curr Opin Pharmacol*. 2006;6:620–625.
6. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? *Nat Rev Clin Oncol*. 2017;14:85–99.
7. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet*. 2005;365:1333–1346.
8. Owen KR. Monogenic diabetes in adults: what are the new developments? *Curr Opin Genet Dev*. 2018;50:103–110.
9. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet*. 2018;391:2449–2462.
10. Monzavi R, Dreimane D, Geffner ME, et al. Improvement in risk factors for metabolic syndrome and insulin resistance in overweight youth who are treated with lifestyle intervention. *Pediatrics*. 2006;117:e1111–e1118.
11. Roden M, Shulman GI. The integrative biology of type 2 diabetes. *Nature*. 2019;576:51–60.
12. Perry RJ, Peng L, Barry NA, et al. Acetate mediates a microbiome-brain-beta-cell axis to promote metabolic syndrome. *Nature*. 2016;534:213–217.
13. Komaroff AL. The microbiome and risk for obesity and diabetes. *JAMA*. 2017;317:355–356.
14. Flannick J, Mercader JM, Fuchsberger C, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019;570:71–76.
15. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981–990.
16. Chen Z, Yu H, Shi X, et al. Functional screening of candidate causal genes for insulin resistance in human preadipocytes and adipocytes. *Circ Res*. 2020;126:330–346.
17. Samuel VT, Shulman GI. Nonalcoholic fatty liver disease, insulin resistance, and ceramides. *NEJM*. 2019;381:1866–1869.
18. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*. 2005;54:1615–1625.
19. Arnold SE, Arvanitakis Z, Macauley-Rambach SL, et al. Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. *Nat Rev Neurol*. 2018;14:168–181.
20. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet*. 2017;389:2239–2251.
21. Beitelshees AL, Leslie BR, Taylor SI. Sodium-glucose cotransporter 2 inhibitors: a case study in translational research. *Diabetes*. 2019;68:1109–1120.
22. Mattson MP, Longo VD, Harvie M. Impact of intermittent fasting on health and disease processes. *Ageing Res Rev*. 2017;39:46–58.
23. Adams MJ, Blundell TL, Dodson EJ, et al. Structure of rhombohedral 2 zinc insulin crystals. *Nature*. 1969;224:491–495.
24. Hubbard SR. The insulin receptor: both a prototypical and atypical receptor tyrosine kinase. *CSH Perspect*. 2013;5:1–12.
25. De Meyts P. Insulin/receptor binding: the last piece of the puzzle? What recent progress on the structure of the insulin/receptor complex tells us (or not) about negative cooperativity and activation. *Bioessays*. 2015;37:389–397.
26. Ferguson KM, Hu C, Lemmon MA. Insulin and epidermal growth factor receptor family members share parallel activation mechanisms. *Protein Sci*. 2020;29:1331–1344.
27. Brzozowski AM, Dodson EJ, Dodson GG, et al. Structural origins of the functional divergence of human insulin-like growth factor-I and insulin. *Biochem*. 2002;41:9389–9397.
28. De Meyts P. The insulin receptor: a prototype for dimeric, allosteric membrane receptors? *Trends Biochem Sci*. 2008;33:376–384.
29. De Meyts P. Insulin and its receptor: structure, function and evolution. *Bioessays*. 2004;26:1351–1362.
30. Menting JG, Whittaker J, Margetts MB, et al. How insulin engages its primary binding site on the insulin receptor. *Nature*. 2013;493:241–245.
31. Menting JG, Yang Y, Chan SJ, et al. Protective hinge in insulin opens to enable its receptor engagement. *Prod Natl Acad Sci U S A*. 2014;111:E3395–E3404.
32. Scapin G, Dandey VP, Zhang Z, et al. Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis. *Nature*. 2018;556:122–125.
33. Kasuga M, Karlsson FA, Kahn CR. Insulin stimulates the phosphorylation of the 95,000-dalton subunit of its own receptor. *Science*. 1982;215:185–187.
34. Bravo DA, Gleason JB, Sanchez RI, et al. Accurate and efficient cleavage of the human insulin proreceptor by the human proprotein-processing protease furin. Characterization and kinetic parameters using the purified, secreted soluble protease expressed by a recombinant baculovirus. *J Biol Chem*. 1994;269:25830–25837.
35. Whittaker J. Structure and function of the insulin receptor. In: Post TW, ed. *Up to Date*. Waltham: Uptodate; 2020.
36. De Meyts P, Whittaker J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nat Rev Drug Discov*. 2002;1:769–783.
37. Whittaker J, Sorensen H, Gadsboll VL, et al. Comparison of the functional insulin binding epitopes of the A and B isoforms of the insulin receptor. *J Biol Chem*. 2002;277:47380–47384.
38. Gutmann T, Kim KH, Grzybek M, et al. Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *J Cell Biol*. 2018;217:1643–1649.
39. Gutmann T, Schafer IB, Poojari C, et al. Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *J Cell Biol*. 2020;219:e201907210.

40. Xu Y, Kong GK, Menting JG, et al. How ligand binds to the type 1 insulin-like growth factor receptor. *Nat Commun*. 2018;9:821.
41. Slaaby R. Specific insulin/IGF1 hybrid receptor activation assay reveals IGF1 as a more potent ligand than insulin. *Sci Rep*. 2015;5:7911.
42. Siddle K. Molecular basis of signaling specificity of insulin and IGF receptors: neglected corners and recent advances. *Front Endocrinol*. 2012;3:34.
43. Kasuga M. Structure and function of the insulin receptor-a personal perspective. *Proc Jpn Acad Ser B Phys Biol Sci*. 2019;95:581–589.
44. Ullrich A, Bell JR, Chen EY, et al. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature*. 1985;313:756–761.
45. Ebina Y, Ellis L, Jarnagin K, et al. The human insulin receptor cDNA: the structural basis for hormone activated transmembrane signalling. *Cell*. 1985;40:747–758.
46. Czech MP. Molecular basis of insulin action. *Annu Rev Biochem*. 1977;46:359–384.
47. Cheng Z, White MF. kNOXing on the door of selective insulin resistance. *Arterioscler Thromb Vasc Biol*. 2012;32:1063–1065.
48. Taylor SI. Deconstructing type 2 diabetes. *Cell*. 1999;97:9–12.
49. White MF, Shoelson SE, Keutmann H, et al. A cascade of tyrosine autophosphorylation in the beta-subunit activates the phosphotransferase of the insulin receptor. *J Biol Chem*. 1988;263:2969–2980.
50. Rajagopalan M, Neidigh JL, McClain DA. Amino acid sequences Gly-Pro-Leu-Tyr and Asn-Pro-Glu-Tyr in the submembranous domain of the insulin receptor are required for normal endocytosis. *J Biol Chem*. 1991;266:23068–23073.
51. Hubbard SR, Miller WT. Closing in on a mechanism for activation. *Elife*. 2014;3:e04909.
52. Till JH, Ablooglu AJ, Frankel M, et al. Crystallographic and solution studies of an activation loop mutant of the insulin receptor tyrosine kinase: insights into kinase mechanism. *J Biol Chem*. 2001;276:10049–10055.
53. White MF, Haring HU, Kasuga M, et al. Kinetic properties and sites of autophosphorylation of the partially purified insulin receptor from hepatoma cells. *J Biol Chem*. 1984;259:255–264.
54. Wei L, Hubbard SR, Hendrickson WA, et al. Expression, characterization, and crystallization of the catalytic core of the human insulin receptor protein-tyrosine kinase domain. *J Biol Chem*. 1995;270:8122–8130.
55. Ellis L, Clauser E, Morgan DO, et al. Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. *Cell*. 1986;45:721–732.
56. Wilden PA, Kahn CR, Siddle K, et al. Insulin receptor kinase domain autophosphorylation regulates receptor enzymatic function. *J Biol Chem*. 1992;267:16660–16668.
57. Nef S, Verma-Kurvari S, Merenmies J, et al. Testis determination requires insulin receptor family function in mice. *Nature*. 2003;426:291–295.
58. Shtykova EV, Petoukhov MV, Mozhaev AA, et al. The dimeric ectodomain of the alkali-sensing insulin receptor–related receptor (ectoIRR) has a droplike shape. *J Biol Chem*. 2019;294:17790–17798.
59. Kasuga M, Zick Y, Blithe DL, et al. Insulin stimulates tyrosine phosphorylation of the insulin receptor in a cell-free system. *Nature*. 1982;298:667–669.
60. White MF, Maron R, Kahn CR. Insulin rapidly stimulates tyrosine phosphorylation of a Mr 185,000 protein in intact cells. *Nature*. 1985;318:183–186.
61. Sun XJ, Rothenberg P, Kahn CR, et al. Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature*. 1991;352:73–77.
62. Bjornholm M, He AR, Attersand A, et al. Absence of functional insulin receptor substrate-3 (IRS-3) gene in humans. *Diabetologia*. 2002;45:1697–1702.
63. Numan S, Russell DS. Discrete expression of insulin receptor substrate-4 mRNA in adult rat brain. *Brain Res Mol Brain Res*. 1999;72:97–102.
64. Sadagurski M, Dong XC, Myers Jr MG, et al. Irs2 and Irs4 synergize in non-LepRb neurons to control energy balance and glucose homeostasis. *Mol Metab*. 2014;3:55–63.
65. Backer JM, Myers Jr MG, Shoelson SE, et al. Phosphatidylinositol 3'-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO*. 1992;11:3469–3479.
66. Songyang Z, Cantley LC. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem Sci*. 1995;20:470–475.
67. Hubbard SR. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO*. 1997;16:5572–5581.
68. Humphrey SJ, Azimifar SB, Mann M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nature Biotech*. 2015;33:990–995.
69. White MF, Livingston JN, Backer JM, et al. Mutation of the insulin receptor at tyrosine 960 inhibits signal transmission but does not affect its tyrosine kinase activity. *Cell*. 1988;54:641–649.
70. Eck MJ, Dhe-Paganon S, Trub T, et al. Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. *Cell*. 1996;85:695–705.
71. Farooq A, Plotnikova O, Zeng L, et al. Phosphotyrosine binding domains of Shc and insulin receptor substrate 1 recognize the NPXpY motif in a thermodynamically distinct manner. *J Biol Chem*. 1999;274:6114–6121.
72. Yenush L, Makati KJ, Smith-Hall J, et al. The pleckstrin homology domain is the principal link between the insulin receptor and IRS-1. *J Biol Chem*. 1996;271:24300–24306.
73. Dhe-Paganon S, Ottinger EA, Nolte RT, et al. Crystal structure of the pleckstrin homology-phosphotyrosine binding (PH-PTB) targeting region of insulin receptor substrate 1. *Proc Natl Acad Sci U S A*. 1999;96:8378–8383.
74. Burks DJ, Pons S, Towery H, et al. Heterologous pleckstrin homology domains do not couple IRS-1 to the insulin receptor. *J Biol Chem*. 1997;272:27716–27721.
75. Burks DJ, Wang J, Towery H, et al. IRS pleckstrin homology domains bind to acidic motifs in proteins. *J Biol Chem*. 1998;273:31061–31067.
76. Sawka-Verhelle D, Baron V, Mothe I, et al. Tyr624 and Tyr628 in insulin receptor substrate-2 mediate its association with the insulin receptor. *J Biol Chem*. 1997;272:16414–16420.
77. Sawka-Verhelle D, Tartare-Deckert S, White MF, et al. Insulin receptor substrate-2 binds to the insulin receptor through its phosphotyrosine-binding domain and through a newly identified domain comprising amino acids 591-786. *J Biol Chem*. 1996;271:5980–5983.
78. Wu J, Tseng YD, Xu CF, et al. Structural and biochemical characterization of the KRLB region in insulin receptor substrate-2. *Nature Struct Mol Bio*. 2008;15:251–258.
79. Thorpe LM, Yuzugullu H, Zhao JJ. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat Rev Cancer*. 2015;15:7–24.
80. Antonetti DA, Algenstaedt P, Kahn CR. Insulin receptor substrate 1 binds two novel splice variants of the regulatory subunit of phosphatidylinositol 3-kinase in muscle and brain. *Mol Cell Biol*. 1996;16:2195–2203.
81. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol*. 2006;7:85–96.
82. Cantley LC. The phosphoinositide 3-kinase pathway. *Science*. 2002;296:1655–1657.
83. Vanhaesebroeck B, Stephens L, Hawkins P. PI3K signalling: the path to discovery and understanding. *Nat Rev Mol Cell Biol*. 2012;13:195–203.
84. Xu F, Na L, Li Y, et al. Roles of the PI3K/AKT/mTOR signalling pathways in neurodegenerative diseases and tumours. *Cell Biosci*. 2020;10:54.
85. Yang G, Murashige DS, Humphrey SJ, et al. A positive feedback loop between Akt and mTORC2 via SIN1 phosphorylation. *Cell Rep*. 2015;12:937–943.
86. Humphrey SJ, James DE. Uncaging akt. *Science Sig*. 2012;5:pe20.
87. Yingying X, Caijuan W, Yenan Y, et al. The effect of SIN1 and microtubules on insulin induced PKC zeta activation. *Med Sci Monit*. 2017;23:3666–3672.
88. Liu P, Gan W, Inuzuka H, et al. Sin1 phosphorylation impairs mTORC2 complex integrity and inhibits downstream Akt signalling to suppress tumorigenesis. *Nat Cell Biol*. 2013;15:1340–1350.

89. Xie J, Proud CG. Crosstalk between mTOR complexes. *Nat Cell Biol*. 2013;15:1263–1265.
90. Lu M, Wan M, Leavens KF, et al. Insulin regulates liver metabolism in vivo in the absence of hepatic Akt and Foxo1. *Nature Med*. 2012;18:388–395.
91. George S, Rochford JJ, Wolfrum C, et al. A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science*. 2004;304:1325–1328.
92. Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. *Cell*. 2007;129:1261–1274.
93. Yao LJ, McCormick JA, Wang J, et al. Novel role for SGK3 in glucose homeostasis revealed in SGK3/Akt2 double-null mice. *Mol Endocrinol*. 2011;25:2106–2118.
94. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*. 2012;149:274–293.
95. Proud CG. Cell signaling. mTOR, unleashed. *Science*. 2007;318:926–927.
96. Sancak Y, Bar-Peled L, Zoncu R, et al. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell*. 2010;141:290–303.
97. Su MY, Morris KL, Kim DJ, et al. Hybrid structure of the Raga/C-ragulator mTORC1 activation complex. *Mol Cell*. 2017;68:835–846.e3.
98. Dai N, Christiansen J, Nielsen FC, et al. mTOR complex 2 phosphorylates IMP1 cotranslationally to promote IGF2 production and the proliferation of mouse embryonic fibroblasts. *Genes Dev*. 2013;27:301–312.
99. Kim SJ, DeStefano MA, Oh WJ, et al. mTOR complex 2 regulates proper turnover of insulin receptor substrate-1 via the ubiquitin ligase subunit Fbw8. *Mol Cell*. 2012;48:875–887.
100. Sandri M. Autophagy in skeletal muscle. *FEBS Lett*. 2010;584:1411–1416.
101. Zhang K, Guo X, Yan H, et al. Phosphorylation of forkhead protein FoxO1 at S253 regulates glucose homeostasis in mice. *Endocrinol*. 2019;160:1333–1347.
102. Obsil T, Obsilova V. Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene*. 2008;27:2263–2275.
103. Brown AK, Webb AE. Regulation of FOXO factors in mammalian cells. *Curr Top Dev Biol*. 2018;127:165–192.
104. Link W. Introduction to FOXO biology. *Methods Mol Biol*. 2019;1890:1–9.
105. Brunet A, Sweeney LB, Sturgill JF, et al. Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science*. 2004;303:2011–2015.
106. Lee S, Dong HH. FoxO integration of insulin signaling with glucose and lipid metabolism. *J Endocrinol*. 2017;233:R67–R79.
107. Lundell LS, Massart J, Altintas A, et al. Regulation of glucose uptake and inflammation markers by FOXO1 and FOXO3 in skeletal muscle. *Mol Metab*. 2019;20:79–88.
108. O'Neill BT, Bhardwaj G, Penniman CM, et al. FoxO transcription factors are critical regulators of diabetes-related muscle atrophy. *Diabetes*. 2019;68:556–570.
109. Matsumoto M, Pocai A, Rossetti L, et al. Impaired regulation of hepatic glucose production in mice lacking the forkhead transcription factor foxo1 in liver. *Cell Metab*. 2007;6:208–216.
110. Barthel A, Schmoll D, Unterman TG. FoxO proteins in insulin action and metabolism. *Trends Endocrinol Metab*. 2005;16:183–189.
111. Samuel VT, Shulman GI. Mechanisms for insulin resistance: common threads and missing links. *Cell*. 2012;148:852–871.
112. O'Neill BT, Lee KY, Klaus K, et al. Insulin and IGF-1 receptors regulate FoxO-mediated signaling in muscle proteostasis. *J Clin Invest*. 2016;126:3433–3446.
113. Long YC, Cheng Z, Copps KD, et al. Insulin receptor substrates Irs1 and Irs2 coordinate skeletal muscle growth and metabolism via the Akt and AMPK pathways. *Mol Cell Biol*. 2011;31:430–441.
114. Rhoads RE. Signal transduction pathways that regulate eukaryotic protein synthesis. *J Biol Chem*. 1999;274:30337–30340.
115. Pavitt GD. eIF2B, a mediator of general and gene-specific translational control. *Biochem Soc Trans*. 2005;33:1487–1492.
116. Proud CG, Denton RM. Molecular mechanisms for the control of translation by insulin. *Biochem J*. 1997;328:329–341.
117. Brunn GJ, Hudson CC, Sekulic A, et al. Phosphorylation of the translational repressor PHAS-I by the mammalian target of rapamycin. *Science*. 1997;277:99–101.
118. Waskiewicz AJ, Johnson JC, Penn B, et al. Phosphorylation of the cap-binding protein eukaryotic translation initiation factor 4E by protein kinase Mnk1 in vivo. *Mol Cell Biol*. 1999;19:1871–1880.
119. Morley SJ, Dever TE, Etchison D, et al. Phosphorylation of eIF-4F by protein kinase C or multipotential S6 kinase stimulates protein synthesis at initiation. *J Biol Chem*. 1991;266:4669–4672.
120. Milan G, Romanello V, Pescatore F, et al. Regulation of autophagy and the ubiquitin-proteasome system by the FoxO transcriptional network during muscle atrophy. *Nat Commun*. 2015;6:6670.
121. James HA, O'Neill BT, Nair KS. Insulin regulation of proteostasis and clinical implications. *Cell Metab*. 2017;26:310–323.
122. Rousseau A, Bertolotti A. Regulation of proteasome assembly and activity in health and disease. *Nat Rev Mol Cell Biol*. 2018;19:697–712.
123. Frendo-Cumbo S, Jaldin-Fincati JR, Coyaud E, et al. Deficiency of the autophagy gene ATG16L1 induces insulin resistance through KLHL9/KLHL13/CUL3-mediated IRS1 degradation. *J Biol Chem*. 2019;294:16172–16185.
124. Jaldin-Fincati JR, Pavarotti M, Frendo-Cumbo S, et al. Update on GLUT4 vesicle traffic: a cornerstone of insulin action. *Trends Endocrinol Metab*. 2017;28:597–611.
125. Klip A, McGraw TE, James DE. Thirty sweet years of GLUT4. *J Biol Chem*. 2019;294:11369–11381.
126. Bogan JS, Kandror KV. Biogenesis and regulation of insulin-responsive vesicles containing GLUT4. *Curr Opin Cell Biol*. 2010;22:506–512.
127. Liu T, Yu B, Kakino M, et al. A novel IRS-1-associated protein, DGKzeta regulates GLUT4 translocation in 3T3-L1 adipocytes. *Sci Rep*. 2016;6:35438.
128. Bogan JS. Regulation of glucose transporter translocation in health and diabetes. *Annu Rev Biochem*. 2012;81:507–532.
129. Huang S, Czech MP. The GLUT4 glucose transporter. *Cell Metab*. 2007;5:237–252.
130. Foley K, Boguslavsky S, Klip A. Endocytosis, recycling, and regulated exocytosis of glucose transporter 4. *Biochemy*. 2011;50:3048–3061.
131. Li J, Peters PJ, Bai M, et al. An ACAP1-containing clathrin coat complex for endocytic recycling. *J Cell Biol*. 2007;178:453–464.
132. Li J, Malaby AW, Famulok M, et al. Grp1 plays a key role in linking insulin signaling to glut4 recycling. *Develop Cell*. 2012;22:1286–1298.
133. Kido Y, Burks DJ, Withers D, et al. Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1, and IRS-2. *J Clin Invest*. 2000;105:199–205.
134. Blattler SM, Cunningham JT, Verdeguer F, et al. Yin Yang 1 deficiency in skeletal muscle protects against rapamycin-induced diabetic-like symptoms through activation of insulin/IGF signaling. *Cell Metab*. 2012;15:505–517.
135. Verdeguer F, Blattler SM, Cunningham JT, et al. Decreased genetic dosage of hepatic Yin Yang 1 causes diabetic-like symptoms. *Mol Endocrinol*. 2014;28:308–316.
136. Meng X, Kondo M, Morino K, et al. Transcription factor AP-2beta: a negative regulator of IRS-1 gene expression. *Biochem Biophys Res Commun*. 2010;392:526–532.
137. Wang Z, Shah OJ, Hunter T. The transcriptional coactivators p/CIP and SRC-1 control insulin resistance through IRS1 in obesity models. *PLoS One*. 2012;7:e36961.
138. Maeda S, Tsukada S, Kanazawa A, et al. Genetic variations in the gene encoding TFAP2B are associated with type 2 diabetes mellitus. *J Human Genet*. 2005;50:283–292.
139. Xu J, Li Q. Review of the in vivo functions of the p160 steroid receptor coactivator family. *Mol Endocrinol*. 2003;17:1681–1692.
140. Hwang JH, Kim AR, Kim KM, et al. TAZ couples Hippo/Wnt signalling and insulin sensitivity through Irs1 expression. *Nat Commun*. 2019;10:421.
141. Jhala US, Canettieri G, Screaton RA, et al. cAMP promotes pancreatic beta-cell survival via CREB-mediated induction of IRS2. *Genes Dev*. 2003;17:1575–1580.
142. Canettieri G, Koo SH, Berdeaux R, et al. Dual role of the coactivator TORC2 in modulating hepatic glucose output and insulin signaling. *Cell Metab*. 2005;2:331–338.
143. Awazawa M, Ueki K, Inabe K, et al. Adiponectin enhances insulin sensitivity by increasing hepatic IRS-2 expression via a macrophage-derived IL-6-dependent pathway. *Cell Metab*. 2011;13:401–412.

144. Besse-Patin A, Jeromson S, Levesque-Damphousse P, et al. PGC1A regulates the IRS1:IRS2 ratio during fasting to influence hepatic metabolism downstream of insulin. *Prod Natl Acad Sci U S A*. 2019;116:4285–4290.
145. Nakagawa Y, Shimano H, Yoshikawa T, et al. TFE3 transcriptionally activates hepatic IRS-2, participates in insulin signaling and ameliorates diabetes. *Nat Med*. 2006;12:107–113.
146. Foretz M, Guichard C, Ferre P, et al. Sterol regulatory element binding protein-1c is a major mediator of insulin action on the hepatic expression of glucokinase and lipogenesis-related genes. *Prod Natl Acad Sci U S A*. 1999;96:12737–12742.
147. Shimomura I, Bashmakov Y, Ikemoto S, et al. Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes. *Prod Natl Acad Sci U S A*. 1999;96:13656–13661.
148. Ide T, Shimano H, Yahagi N, et al. SREBPs suppress IRS-2-mediated insulin signalling in the liver. *Nat Cell Biol*. 2004;6:351–357.
149. Taniguchi CM, Finger EC, Krieg AJ, et al. Cross-talk between hypoxia and insulin signaling through Phd3 regulates hepatic glucose and lipid metabolism and ameliorates diabetes. *Nature Med*. 2013;19:1325–1330.
150. Kubota N, Kubota T, Kajiwara E, et al. Differential hepatic distribution of insulin receptor substrates causes selective insulin resistance in diabetes and obesity. *Nat Commun*. 2016;7:12977.
151. Copps KD, White MF. Regulation of insulin sensitivity by serine/threonine phosphorylation of insulin receptor substrate proteins IRS1 and IRS2. *Diabetologia*. 2012;55:2565–2582.
152. Hancer NJ, Qiu W, Cherella C, et al. Insulin and metabolic stress stimulate multisite serine/threonine phosphorylation of insulin receptor substrate 1 and inhibit tyrosine phosphorylation. *J Biol Chem*. 2014;289:12467–12484.
153. Mehran AE, Templeman NM, Brigidi GS, et al. Hyperinsulinemia drives diet-induced obesity independently of brain insulin production. *Cell Metab*. 2012;16:723–737.
154. Czech MP. Insulin action and resistance in obesity and type 2 diabetes. *Nature Med*. 2017;23:804–814.
155. Hirosumi J, Tuncman G, Chang L, et al. A central role for JNK in obesity and insulin resistance. *Nature*. 2002;420:333–336.
156. Um SH, Frigerio F, Watanabe M, et al. Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature*. 2004;431:200–205.
157. Copps KD, Hancer NJ, Opare-Ado L, et al. Irs1 serine 307 promotes insulin sensitivity in mice. *Cell Metab*. 2010;11:84–92.
158. Varshavsky A. The ubiquitin system, autophagy, and regulated protein degradation. *Annu Rev Biochem*. 2017;86:123–128.
159. Rui L, Fisher TL, Thomas J, et al. Regulation of insulin/insulin-like growth factor-1 signaling by proteasome-mediated degradation of insulin receptor substrate-2. *J Biol Chem*. 2001;276:40362–40367.
160. Rui L, Yuan M, Frantz D, et al. SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *J Biol Chem*. 2002;277:42394–42398.
161. Liu F, Yang T, Wang B, et al. Resistin induces insulin resistance, but does not affect glucose output in rat-derived hepatocytes. *Acta Pharmacol Sin*. 2008;29:98–104.
162. Ueki K, Kondo T, Kahn CR. Suppressor of cytokine signaling 1 (SOCS-1) and SOCS-3 cause insulin resistance through inhibition of tyrosine phosphorylation of insulin receptor substrate proteins by discrete mechanisms. *Mol Cell Biol*. 2004;24:5434–5446.
163. Aaronson DS, Horvath CM. A road map for those who don't know JAK-STAT. *Science*. 2002;296:1653–1655.
164. Xu T, Xu L, Meng P, et al. Angptl7 promotes insulin resistance and type 2 diabetes mellitus by multiple mechanisms including SOCS3-mediated IRS1 degradation. *FASEB J*. 2020;34:13548–13560.
165. Xu X, Sarikas A, Dias-Santagata DC, et al. The CUL7 E3 ubiquitin ligase targets insulin receptor substrate 1 for ubiquitin-dependent degradation. *Mol Cell*. 2008;30:403–414.
166. Xu X, Keshwani M, Meyer K, et al. Identification of the degradation determinants of insulin receptor substrate 1 for signaling cullin-RING E3 ubiquitin ligase 7-mediated ubiquitination. *J Biol Chem*. 2012;287:40758–40766.
167. Bonala S, Lokireddy S, McFarlane C, et al. Myostatin induces insulin resistance via Casitas B-lineage lymphoma b (Cblb)-mediated degradation of insulin receptor substrate 1 (IRS1) protein in response to high calorie diet intake. *J Biol Chem*. 2014;289:7654–7670.
168. Tiganis T. PTP1B and TCPTP--nonredundant phosphatases in insulin signaling and glucose homeostasis. *FEBS*. 2013;280:445–458.
169. Galic S, Hauser C, Kahn BB, et al. Coordinated regulation of insulin signaling by the protein tyrosine phosphatases PTP1B and TCPTP. *Mol Cell Biol*. 2005;25:819–829.
170. Elchebly M, Payette P, Michaliszyn E, et al. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene [see comments]. *Science*. 1999;283:1544–1548.
171. Meng TC, Buckley DA, Galic S, et al. Regulation of insulin signaling through reversible oxidation of the protein-tyrosine phosphatases TC45 and PTP1B. *J Biol Chem*. 2004;279:37716–37725.
172. Kushner JA, Haj FG, Klaman LD, et al. Islet-sparing effects of protein tyrosine phosphatase-1b deficiency delays onset of diabetes in IRS2 knockout mice. *Diabetes*. 2004;53:61–66.
173. Cho H. Protein tyrosine phosphatase 1B (PTP1B) and obesity. *Vitam Horm*. 2013;91:405–424.
174. Zhang ZY, Dodd GT, Tiganis T. Protein tyrosine phosphatases in hypothalamic insulin and leptin signaling. *Trends Pharmacol Sci*. 2015;36:661–674.
175. Dodd GT, Xirouchaki CE, Eramo M, et al. Intranasal targeting of hypothalamic PTP1B and TCPTP reinstates leptin and insulin sensitivity and promotes weight loss in obesity. *Cell Rep*. 2019;28:2905–2922.e5.
176. Li J, Yen C, Liaw D, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*. 1997;275:1943–1947.
177. Parsons R. Human cancer, PTEN and the PI-3 kinase pathway. *Semin Cell Dev Biol*. 2004;15:171–176.
178. Lehtonen S. SHIPping out diabetes-Metformin, an old friend among new SHIP2 inhibitors. *Acta Physiol (Oxf)*. 2020;228:e13349.
179. Polianskyte-Prause Z, Tolvanen TA, Lindfors S, et al. Metformin increases glucose uptake and acts renoprotectively by reducing SHIP2 activity. *FASEB J*. 2019;33:2858–2869.
180. Fowle H, Zhao Z, Grana X. PP2A holoenzymes, substrate specificity driving cellular functions and deregulation in cancer. *Adv Cancer Res*. 2019;144:55–93.
181. Ugi S, Imamura T, Maegawa H, et al. Protein phosphatase 2A negatively regulates insulin's metabolic signaling pathway by inhibiting Akt (protein kinase B) activity in 3T3-L1 adipocytes. *Mol Cell Biol*. 2004;24:8778–8789.
182. Javadpour P, Dargahi L, Ahmadiani A, et al. To be or not to be: PP2A as a dual player in CNS functions, its role in neurodegeneration, and its interaction with brain insulin signaling. *Cell Mol Life Sci*. 2019;76:2277–2297.
183. Hartley D, Cooper GM. Role of mTOR in the degradation of IRS-1: regulation of PP2A activity. *J Cell Biochem*. 2002;85:304–314.
184. Carlson CJ, White MF, Rondinone CM. Mammalian target of rapamycin regulates IRS-1 serine 307 phosphorylation. *Biochem Biophys Res Commun*. 2004;316:533–539.
185. Geetha T, Langlais P, Luo M, et al. Label-free proteomic identification of endogenous, insulin-stimulated interaction partners of insulin receptor substrate-1. *J Am Soc Mass Spect*. 2011;22:457–466.
186. Law NC, White MF, Hunzicker-Dunn ME. G protein-coupled receptors (GPCRs) that signal via protein kinase A (PKA) cross-talk at insulin receptor substrate 1 (IRS1) to activate the PI3K/AKT pathway. *J Biol Chem*. 2016;291:27160–27169.
187. Nigi L, Grieco GE, Ventriglia G, et al. MicroRNAs as regulators of insulin signaling: research updates and potential therapeutic perspectives in type 2 diabetes. *Int J Mol Sci*. 2018;19:3705.
188. Zhu H, Shyh-Chang N, Segre AV, et al. The Lin28/let-7 axis regulates glucose metabolism. *Cell*. 2011;147:81–94.
189. Min KH, Yang WM, Lee W. Saturated fatty acids-induced miR-424-5p aggravates insulin resistance via targeting insulin receptor in hepatocytes. *Biochem Biophys Res Commun*. 2018;503:1587–1593.
190. Yang WM, Jeong HJ, Park SW, et al. Obesity-induced miR-15b is linked causally to the development of insulin resistance through the repression of the insulin receptor in hepatocytes. *Mol Nutr Food Res*. 2015;59:2303–2314.

191. Yang WM, Jeong HJ, Park SY, et al. Saturated fatty acid-induced miR-195 impairs insulin signaling and glycogen metabolism in HepG2 cells. *FEBS Lett*. 2014;588:3939–3946.
192. Yang WM, Min KH, Lee W. Induction of miR-96 by dietary saturated fatty acids exacerbates hepatic insulin resistance through the suppression of INSR and IRS-1. *PLoS One*. 2016;11:e0169039.
193. Ono K, Igata M, Kondo T, et al. Identification of microRNA that represses IRS-1 expression in liver. *PLoS One*. 2018;13:e0191553.
194. Wen F, Yang Y, Jin D, et al. MiRNA-145 is involved in the development of resistin-induced insulin resistance in HepG2 cells. *Biochem Biophys Res Commun*. 2014;445:517–523.
195. Massart J, Sjogren RJO, Lundell LS, et al. Altered miR-29 expression in type 2 diabetes influences glucose and lipid metabolism in skeletal muscle. *Diabetes*. 2017;66:1807–1818.
196. Fang S, Ma X, Guo S, et al. MicroRNA-126 inhibits cell viability and invasion in a diabetic retinopathy model via targeting IRS-1. *Oncol Lett*. 2017;14:4311–4318.
197. Nossent AY, Eskildsen TV, Andersen LB, et al. The 14q32 microRNA-487b targets the antiapoptotic insulin receptor substrate 1 in hypertension-induced remodeling of the aorta. *Annal Surg*. 2013;258:743–751; discussion 52–53.
198. Tao H, Wang MM, Zhang M, et al. MiR-126 suppresses the glucose-stimulated proliferation via IRS-2 in INS-1 beta cells. *PLoS One*. 2016;11:e0149954.
199. Agarwal P, Srivastava R, Srivastava AK, et al. miR-135a targets IRS2 and regulates insulin signaling and glucose uptake in the diabetic gastrocnemius skeletal muscle. *Biochim Biophys Acta*. 2013;1832:1294–1303.
200. Xu G, Ji C, Song G, et al. MiR-26b modulates insulin sensitivity in adipocytes by interrupting the PTEN/PI3K/AKT pathway. *Int J Obes (Lond)*. 2015;39:1523–1530.
201. Regazzi R. MicroRNAs as therapeutic targets for the treatment of diabetes mellitus and its complications. *Expert Opin Ther Targets*. 2018;22:153–160.
202. Biddinger SB, Kahn CR. From mice to men: insights into the insulin resistance syndromes. *Annu Rev Physiol*. 2006;68:123–158.
203. Kubota T, Kubota N, Kadowaki T. Imbalanced insulin actions in obesity and type 2 diabetes: key mouse models of insulin signaling pathway. *Cell Metab*. 2017;25:797–810.
204. Stefan N, Haring HU. The role of hepatokines in metabolism. *Nat Rev Endocrinol*. 2013;9:144–152.
205. Tao R, Wang C, Stohr O, et al. Inactivating hepatic follistatin alleviates hyperglycemia. *Nature Med*. 2018;24:1058–1069.
206. Nandi A, Kitamura Y, Kahn CR, et al. Mouse models of insulin resistance. *Physiol Rev*. 2004;84:623–647.
207. Sell C, Dumenil G, Deveaud C, et al. Effect of a null mutation of the insulin-like growth factor I receptor gene on growth and transformation of mouse embryo fibroblasts. *Mol Cell Biol*. 1994;14:3604–3612.
208. Accili D, Drago J, Lee EJ, et al. Early neonatal death in mice homozygous for a null allele of the insulin receptor gene. *Nat Genet*. 1996;12:106–109.
209. Kitamura T, Kahn CR, Accili D. Insulin receptor knockout mice. *Annu Rev Physiol*. 2003;65:313–332.
210. Okamoto H, Nakae J, Kitamura T, et al. Transgenic rescue of insulin receptor-deficient mice. *J Clin Invest*. 2004;114:214–223.
211. Krook A, Brueton L, O'Rahilly S. Homozygous nonsense mutation in the insulin receptor gene in infant with leprechaunism. *Lancet*. 1993;342:277–278.
212. White MF. Insulin signaling in health and disease. *Science*. 2003;302:1710–1711.
213. Huang C, Wu M, Du J, et al. Systematic modeling for the insulin signaling network mediated by IRS(1) and IRS(2). *J Theor Biol*. 2014;355:40–52.
214. Araki E, Lipes MA, Patti ME, et al. Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature*. 1994;372:186–190.
215. Withers DJ, Gutierrez JS, Towery H, et al. Disruption of IRS-2 causes type 2 diabetes in mice. *Nature*. 1998;391:900–904.
216. Burks DJ, Font de Mora J, Schubert M, et al. IRS-2 pathways integrate female reproduction and energy homeostasis. *Nature*. 2000;407:377–382.
217. Withers DJ, Burks DJ, Towery HH, et al. Irs-2 coordinates Igf-1 receptor-mediated beta-cell development and peripheral insulin signalling. *Nat Genet*. 1999;23:32–40.
218. Michael MD, Kulkarni RN, Postic C, et al. Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell*. 2000;6:87–97.
219. Titchenell PM, Lazar MA, Birnbaum MJ. Unraveling the regulation of hepatic metabolism by insulin. *Trends Endocrinol Metab*. 2017;28:497–505.
220. Biddinger SB, Hernandez-Ono A, Rask-Madsen C, et al. Hepatic insulin resistance is sufficient to produce dyslipidemia and susceptibility to atherosclerosis. *Cell Metab*. 2008;7:125–134.
221. Fisher SJ, Kahn CR. Insulin signaling is required for insulin's direct and indirect action on hepatic glucose production. *J Clin Invest*. 2003;111:463–468.
222. Hu C, Hoene M, Plomgaard P, et al. Muscle-liver substrate fluxes in exercising humans and potential effects on hepatic metabolism. *J Clin Endo Metab*. 2020;105:1196–1209.
223. Staiger H, Keuper M, Berti L, et al. Fibroblast growth factor 21-metabolic role in mice and men. *Endo Rev*. 2017;38:468–488.
224. Dong XC, Copps KD, Guo S, et al. Inactivation of hepatic Foxo1 by insulin signaling is required for adaptive nutrient homeostasis and endocrine growth regulation. *Cell Metab*. 2008;8:65–76.
225. Cheng Z, Guo S, Copps K, et al. Foxo1 integrates insulin signaling with mitochondrial function in the liver. *Nat Med*. 2009;15:1307–1311.
226. Guo S, Copps KD, Dong X, et al. The Irs1 branch of the insulin signaling cascade plays a dominant role in hepatic nutrient homeostasis. *Mol Cell Biol*. 2009;29:5070–5083.
227. Stohr O, Tao R, Miao J, et al. FoxO1 suppresses Fgf21 during hepatic insulin resistance to impair peripheral glucose utilization and acute cold tolerance. *Cell Rep*. 2021;34:16.
228. O'Sullivan I, Zhang W, Wasserman DH, et al. FoxO1 integrates direct and indirect effects of insulin on hepatic glucose production and glucose utilization. *Nat Commun*. 2015;6:7079.
229. Titchenell PM, Quinn WJ, Lu M, et al. Direct hepatocyte insulin signaling is required for lipogenesis but is dispensable for the suppression of glucose production. *Cell Metab*. 2016;23:1154–1166.
230. Perry RJ, Camporez JP, Kursawe R, et al. Hepatic acetyl CoA links adipose tissue inflammation to hepatic insulin resistance and type 2 diabetes. *Cell*. 2015;160:745–758.
231. Samuel VT, Shulman GI. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *J Clin Invest*. 2016;126:12–22.
232. Han HQ, Zhou X, Mitch WE, et al. Myostatin/activin pathway antagonism: molecular basis and therapeutic potential. *Int J Biochem Cell Biol*. 2013;45:2333–2347.
233. Hansen JS, Plomgaard P. Circulating follistatin in relation to energy metabolism. *Mol Cell Endocrinol*. 2016;433:87–93.
234. Titchenell PM, Chu Q, Monks BR, et al. Hepatic insulin signalling is dispensable for suppression of glucose output by insulin in vivo. *Nat Commun*. 2015;6:7078.
235. Laustsen PG, Russell SJ, Cui L, et al. Essential role of insulin and insulin-like growth factor 1 receptor signaling in cardiac development and function. *Mol Cell Biol*. 2007;27:1649–1664.
236. Bruning JC, Michael MD, Winnay JN, et al. A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. *Mol Cell*. 1998;2:559–569.
237. O'Neill BT, Lauritzen HP, Hirshman MF, et al. Differential role of insulin/IGF-1 receptor signaling in muscle growth and glucose homeostasis. *Cell Rep*. 2015;11:1220–1235.
238. Herman MA, Kahn BB. Glucose transport and sensing in the maintenance of glucose homeostasis and metabolic harmony. *J Clin Invest*. 2006;116:1767–1775.
239. Zisman A, Peroni OD, Abel ED, et al. Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nat Med*. 2000;6:924–928.
240. Kim JK, Zisman A, Fillmore JJ, et al. Glucose toxicity and the development of diabetes in mice with muscle-specific inactivation of GLUT4. *J Clin Invest*. 2001;108:153–160.
241. Riehle C, Weatherford ET, Wende AR, et al. Insulin receptor substrates differentially exacerbate insulin-mediated left ventricular remodeling. *JCI Insight*. 2020;5:e134920.
242. Shimizu I, Minamino T, Toko H, et al. Excessive cardiac insulin signaling exacerbates systolic dysfunction induced by pressure overload in rodents. *J Clin Invest*. 2010;120:1506–1514.

243. Riehle C, Abel ED. Insulin signaling and heart failure. *Circ Res*. 2016;118:1151–1169.
244. Hu D, Yin C, Luo S, et al. Vascular smooth muscle cells contribute to atherosclerosis immunity. *Front Immunol*. 2019;10:1101.
245. Li Q, Fu J, Xia Y, et al. Homozygous receptors for insulin and not IGF-1 accelerate intimal hyperplasia in insulin resistance and diabetes. *Nat Commun*. 2019;10:4427.
246. Rask-Madsen C, Li Q, Freund B, et al. Loss of insulin signaling in vascular endothelial cells accelerates atherosclerosis in apolipoprotein E null mice. *Cell Metab*. 2010;11:379–389.
247. Clough MH, Schneider DJ, Sobel BE, et al. Attenuation of accumulation of neointimal lipid by pioglitazone in mice genetically deficient in insulin receptor substrate-2 and apolipoprotein E. *J Histochem Cytochem*. 2005;53:603–610.
248. He Z, Opland DM, Way KJ, et al. Regulation of vascular endothelial growth factor expression and vascularization in the myocardium by insulin receptor and PI3K/Akt pathways in insulin resistance and ischemia. *Arterioscler Thromb Vasc Biol*. 2006;26:787–793.
249. Kubota T, Kubota N, Kumagai H, et al. Impaired insulin signaling in endothelial cells reduces insulin-induced glucose uptake by skeletal muscle. *Cell Metab*. 2011;13:294–307.
250. Rosen ED, Walkey CJ, Puigserver P, et al. Transcriptional regulation of adipogenesis. *Genes Dev*. 2000;14:1293–1307.
251. Fasshauer M, Bluher M. Adipokines in health and disease. *Trends Pharmacol Sci*. 2015;36:461–470.
252. Bluher M, Michael MD, Peroni OD, et al. Adipose tissue selective insulin receptor knockout protects against obesity and obesity-related glucose intolerance. *Dev Cell*. 2002;3:25–38.
253. Softic S, Boucher J, Solheim MH, et al. Lipodystrophy due to adipose tissue-specific insulin receptor knockout results in progressive NAFLD. *Diabetes*. 2016;65:2187–2200.
254. Boucher J, Softic S, El Ouaamari A, et al. Differential roles of insulin and IGF-1 receptors in adipose tissue development and function. *Diabetes*. 2016;65:2201–2213.
255. Bluher M, Kahn BB, Kahn CR. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science*. 2003;299:572–574.
256. Kahn BB. Lilly lecture 1995. Glucose transport: pivotal step in insulin action. *Diabetes*. 1996;45:1644–1654.
257. Zimmermann R, Strauss JG, Haemmerle G, et al. Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science*. 2004;306:1383–1386.
258. Yang A, Mottillo EP. Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. *Biochem*. 2020;477:985–1008.
259. Unger RH, Scherer PE. Gluttony, sloth and the metabolic syndrome: a roadmap to lipotoxicity. *Trends Endocrinol Metab*. 2010;21:345–352.
260. DiPilato LM, Ahmad F, Harms M, et al. The role of PDE3B phosphorylation in the inhibition of lipolysis by insulin. *Mol Cell Biol*. 2015;35:2752–2760.
261. Walenta E, Pessentheiner AR, Pelzmann HJ, et al. alpha/beta-hydrolase domain containing protein 15 (ABHD15)--an adipogenic protein protecting from apoptosis. *PLoS One*. 2013;8:e79134.
262. Xia W, Pessentheiner AR, Hofer DC, et al. Loss of ABHD15 Impairs the anti-lipolytic action of insulin by altering PDE3B stability and contributes to insulin resistance. *Cell Rep*. 2018;23:1948–1961.
263. Kuznetsova A, Yu Y, Hollister-Lock J, et al. Trimeprazine increases IRS2 in human islets and promotes pancreatic beta cell growth and function in mice. *JCI Insight*. 2016;1:e80749.
264. Rhodes CJ, White MF, Leahy JL, et al. Direct autocrine action of insulin on beta-cells: does it make physiological sense? *Diabetes*. 2013;62:2157–2163.
265. Tsunekawa S, Demozay D, Briaud I, et al. FoxO feedback control of basal IRS-2 expression in pancreatic beta-cells is distinct from that in hepatocytes. *Diabetes*. 2011;60:2883–2891.
266. Demozay D, Tsunekawa S, Briaud I, et al. Specific glucose-induced control of insulin receptor substrate-2 expression is mediated via Ca2+-dependent calcineurin/NFAT signaling in primary pancreatic islet beta-cells. *Diabetes*. 2011;60:2892–2902.
267. Assmann A, Ueki K, Winnay JN, et al. Glucose effects on beta-cell growth and survival require activation of insulin receptors and insulin receptor substrate 2. *Mol Cell Biol*. 2009;29:3219–3228.
268. Park S, Dong X, Fisher TL, et al. Exendin-4 uses Irs2 signaling to mediate pancreatic beta cell growth and function. *J Biol Chem*. 2006;281:1159–1168.
269. Kushner JA, Flint CL, Dow MA, et al. Insulin receptor substrate-2 and PDX-1 act convergently to regulate beta cell mass in vivo. *Diabetes*. 2001;50:A338.
270. Kullmann S, Heni M, Hallschmid M, et al. Brain insulin resistance at the crossroads of metabolic and cognitive disorders in humans. *Physiolog Rev*. 2016;96:1169–1209.
271. Bischof GN, Park DC. Obesity and aging: consequences for cognition, brain structure, and brain function. *Psychosom Med*. 2015;77:697–709.
272. Cholerton B, Baker LD, Craft S. Insulin, cognition, and dementia. *Eur J Pharmacol*. 2013;719:170–179.
273. Verdile G, Fuller SJ, Martins RN. The role of type 2 diabetes in neurodegeneration. *Neurobiol Dis*. 2015;84:22–38.
274. Geetha T, Rege SD, Mathews SE, et al. Nerve growth factor receptor TrkA, a new receptor in insulin signaling pathway in PC12 cells. *J Biol Chem*. 2013;288:23807–23813.
275. Huang EJ, Reichardt LF. Trk receptors: roles in neuronal signal transduction. *Annu Rev Biochem*. 2003;72:609–642.
276. Yoneyama Y, Lanzerstorfer P, Niwa H, et al. IRS-1 acts as an endocytic regulator of IGF-I receptor to facilitate sustained IGF signaling. *Elife*. 2018;7:e32893.
277. Yoneyama Y, Matsuo M, Take K, et al. The AP-1 complex regulates intracellular localization of insulin receptor substrate 1, which is required for insulin-like growth factor I-dependent cell proliferation. *Mol Cell Biol*. 2013;33:1991–2003.
278. Talbot K, Wang HY, Kazi H, et al. Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J Clin Invest*. 2012;122:1316–1338.
279. Moloney AM, Griffin RJ, Timmons S, et al. Defects in IGF-1 receptor, insulin receptor and IRS-1/2 in Alzheimer's disease indicate possible resistance to IGF-1 and insulin signalling. *Neurobiol Aging*. 2010;31:224–243.
280. Bomfim TR, Forny-Germano L, Sathler LB, et al. An anti-diabetes agent protects the mouse brain from defective insulin signaling caused by Alzheimer's disease- associated Abeta oligomers. *J Clin Invest*. 2012;122:1339–1353.
281. Lourenco MV, Clarke JR, Frozza RL, et al. TNF-alpha mediates PKR-dependent memory impairment and brain IRS-1 inhibition induced by Alzheimer's beta-amyloid oligomers in mice and monkeys. *Cell Metab*. 2013;18:831–843.
282. Copps KD, Hancer NJ, Qiu W, et al. Serine 302 phosphorylation of mouse insulin receptor substrate 1 (Irs1) Is dispensable for normal insulin signaling and feedback regulation by hepatic S6 kinase. *J Biol Chem*. 2016;291:8602–8617.
283. Irvine EE, Drinkwater L, Radwanska K, et al. Insulin receptor substrate 2 is a negative regulator of memory formation. *Learn Memory*. 2011;18:375–383.
284. Freude S, Hettich MM, Schumann C, et al. Neuronal IGF-1 resistance reduces Abeta accumulation and protects against premature death in a model of Alzheimer's disease. *FASEB J*. 2009;23:3315–3324.
285. Killick R, Scales G, Leroy K, et al. Deletion of Irs2 reduces amyloid deposition and rescues behavioural deficits in APP transgenic mice. *Biochem Biophys Res Commun*. 2009;386:257–262.
286. Cohen E, Paulsson JF, Blinder P, et al. Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell*. 2009;139:1157–1169.
287. Taguchi A, Wartschow LM, White MF. Brain IRS2 signaling coordinates life span and nutrient homeostasis. *Science*. 2007;317:369–372.
288. Sadagurski M, White MF. Integrating metabolism and longevity through insulin and IGF1 signaling. *Endocrinol Metab Clin North Am*. 2013;42:127–148.
289. Kleinridders A, Ferris HA, Cai W, et al. Insulin action in brain regulates systemic metabolism and brain function. *Diabetes*. 2014;63:2232–2243.
290. White MF. IRS2 integrates insulin/IGF1 signalling with metabolism, neurodegeneration and longevity. *Diabetes Obes Metab*. 2014;16:4–15.
291. Tanokashira D, Wang W, Maruyama M, et al. Irs2 deficiency alters hippocampus-associated behaviors during young adulthood. *Biochem Biophys Res Commun*. 2021;559:148–154.

292. Sadagurski M, Cheng Z, Rozzo A, et al. IRS2 increases mitochondrial dysfunction and oxidative stress in a mouse model of Huntington disease. *J Clin Invest*. 2011;121:4070–4081.
293. Hwang I, Oh H, Santo E, et al. FOXO protects against age-progressive axonal degeneration. *Aging Cell*. 2018;17:e12701.
294. Kempermann G, Song H, Gage FH. Neurogenesis in the adult hippocampus. *CSH Perspect*. 2015;7:a018812.
295. Toda T, Gage FH. Review: adult neurogenesis contributes to hippocampal plasticity. *Cell Tissue Res*. 2018;373:693–709.
296. Bond AM, Song H. Infiltration of old brains by T cells causes dysfunction of neural stem cells. *Nature*. 2019;571:178–179.
297. Moreno-Jimenez EP, Flor-Garcia M, Terreros-Roncal J, et al. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nature Med*. 2019;25:554–560.
298. Gage FH. Adult neurogenesis in mammals. *Science*. 2019;364:827–828.
299. Bond AM, Ming GL, Song H. Adult mammalian neural stem cells and neurogenesis: five decades later. *Cell Stem Cell*. 2015;17:385–395.
300. DeCarolis NA, Kirby ED, Wyss-Coray T, et al. The role of the microenvironmental niche in declining stem-cell functions associated with biological aging. *CSH Perspect*. 2015;5:a025874.
301. Dulken BW, Buckley MT, Navarro Negredo P, et al. Single-cell analysis reveals T cell infiltration in old neurogenic niches. *Nature*. 2019;571:205–210.
302. Garthe A, Kempermann G. An old test for new neurons: refining the Morris water maze to study the functional relevance of adult hippocampal neurogenesis. *Front Neurosci*. 2013;7:63.
303. Saltiel AR. Insulin resistance in the defense against obesity. *Cell Metab*. 2012;15:798–804.
304. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol*. 2011;11:98–107.
305. Petersen MC, Shulman GI. Mechanisms of insulin action and insulin resistance. *Physiolog Revs*. 2018;98:2133–2223.
306. Petersen KF, Dufour S, Savage DB, et al. The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Prod Natl Acad Sci U S A*. 2007;104:12587–12594.
307. Meex RCR, Watt MJ. Hepatokines: linking nonalcoholic fatty liver disease and insulin resistance. *Nat Rev Endocrinol*. 2017;13:509–520.
308. Kubota N, Kubota T, Itoh S, et al. Dynamic functional relay between insulin receptor substrate 1 and 2 in hepatic insulin signaling during fasting and feeding. *Cell Metab*. 2008;8:49–64.
309. Terauchi Y, Takamoto I, Kubota N, et al. Glucokinase and IRS-2 are required for compensatory beta cell hyperplasia in response to high-fat diet-induced insulin resistance. *J Clin Invest*. 2007;117:246–257.
310. Page MM, Skovso S, Cen H, et al. Reducing insulin via conditional partial gene ablation in adults reverses diet-induced weight gain. *FASEB J*. 2018;32:1196–1206.
311. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009;9:311–326.
312. Yu Y, Yoon SO, Poulogiannis G, et al. Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science*. 2011;332:1322–1326.
313. Lo KA, Labadorf A, Kennedy NJ, et al. Analysis of in vitro insulin-resistance models and their physiological relevance to in vivo diet-induced adipose insulin resistance. *Cell Rep*. 2013;5:259–270.
314. Holland WL, Miller RA, Wang ZV, et al. Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nature Med*. 2011;17:55–63.
315. Levy JR, Campbell KP, Glass DJ. MG53's new identity. *Skelet Musc*. 2013;3:25.
316. Reilly SM, Saltiel AR. Obesity: a complex role for adipose tissue macrophages. *Nat Rev Endocrinol*. 2014;10:193–194.
317. Wedell-Neergaard AS, Lang Lehrskov L, Christensen RH, et al. Exercise-induced changes in visceral adipose tissue mass are regulated by IL-6 signaling: a randomized controlled trial. *Cell Metab*. 2019;29:844–855.e3.
318. Lumeng CN, Bodzin JL, Saltiel AR. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J Clin Invest*. 2007;117:175–184.
319. Shoelson SE, Herrero L, Naaz A. Obesity, inflammation, and insulin resistance. *Gastroenterol*. 2007;132:2169–2180.
320. Benatti FB, Pedersen BK. Exercise as an anti-inflammatory therapy for rheumatic diseases-myokine regulation. *Nat Rev Rheumatol*. 2015;11:86–97.
321. Kaneto H, Nakatani Y, Miyatsuka T, et al. Possible novel therapy for diabetes with cell-permeable JNK-inhibitory peptide. *Nat Med*. 2004;10:1128–1132.
322. Sabio G, Davis RJ. cJun NH2-terminal kinase 1 (JNK1): roles in metabolic regulation of insulin resistance. *Trends Biochem Sci*. 2010;35:490–496.
323. Vernia S, Cavanagh-Kyros J, Barrett T, et al. Diet-induced obesity mediated by the JNK/DIO2 signal transduction pathway. *Genes Dev*. 2013;27:2345–2355.
324. Lo J, Bernstein LE, Canavan B, et al. Effects of TNF-alpha neutralization on adipocytokines and skeletal muscle adiposity in the metabolic syndrome. *Am J Physiol Endocrinol Metab*. 2007;293:E102–E109.
325. Pedersen BK, Febbraio MA. Muscles, exercise and obesity: skeletal muscle as a secretory organ. *Nat Rev Endocrinol*. 2012;8:457–465.
326. van Hall G, Steensberg A, Sacchetti M, et al. Interleukin-6 stimulates lipolysis and fat oxidation in humans. *J Clin Endocrinol Metab*. 2003;88:3005–3010.
327. Sadagurski M, Norquay L, Farhang J, et al. Human IL6 enhances leptin action in mice. *Diabetologia*. 2010;53:525–535.
328. Petersen MC, Vatner DF, Shulman GI. Regulation of hepatic glucose metabolism in health and disease. *Nat Rev Endocrinol*. 2017;13:572–587.
329. Pal D, Dasgupta S, Kundu R, et al. Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *Nat Med*. 2012;18:1279–1285.
330. Stefan N, Haring HU. Circulating fetuin-A and free fatty acids interact to predict insulin resistance in humans. *Nature Med*. 2013;19:394–395.
331. Meex RC, Hoy AJ, Morris A, et al. Fetuin B is a secreted hepatocyte factor linking steatosis to impaired glucose metabolism. *Cell Metabolism*. 2015;22:1078–1089.
332. Wu C, Borné Y, Gao R, et al. Circulating follistatin predicts type 2 diabetes risk. *Nat Commun*. 2021 (in press).
333. Wagner R, Heni M, Tabak AG, et al. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. *Nature Med*. 2021;27:49–57.
334. Karpe F, Pinnick KE. Biology of upper-body and lower-body adipose tissue--link to whole-body phenotypes. *Nat Rev Endocrinol*. 2015;11:90–100.
335. Stefan N, Schick F, Haring HU. Causes, characteristics, and consequences of metabolically unhealthy normal weight in humans. *Cell Metab*. 2017;26:292–300.
336. Azzu V, Vacca M, Virtue S, et al. Adipose tissue-liver cross talk in the control of whole-body metabolism: implications in nonalcoholic fatty liver disease. *Gastroenterol*. 2020;158:1899–1912.
337. Polyzos SA, Kountouras J, Anastasilakis AD, et al. Activin A and follistatin in patients with nonalcoholic fatty liver disease. *Metabolism*. 2016;65:1550–1558.
338. Marciniak SJ, Ron D. Endoplasmic reticulum stress signaling in disease. *Physiol Rev*. 2006;86:113–1149.
339. Schroder M, Kaufman RJ. The mammalian unfolded protein response. *Annu Rev Biochem*. 2005;74:739–789.
340. Bernales S, Papa FR, Walter P. Intracellular signaling by the unfolded protein response. *Annu Rev Cell Development Biol*. 2006;22:487–508.
341. Oyadomari S, Harding HP, Zhang Y, et al. Dephosphorylation of translation initiation factor 2alpha enhances glucose tolerance and attenuates hepatosteatosis in mice. *Cell Metab*. 2008;7:520–532.
342. Rozpedek W, Pytel D, Mucha B, et al. The role of the PERK/eIF2alpha/ATF4/CHOP signaling pathway in tumor progression during endoplasmic reticulum stress. *Curr Mol Med*. 2016;16:533–544.
343. Ozcan U, Cao Q, Yilmaz E, et al. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science*. 2004;306:457–461.
344. Zhou Y, Lee J, Reno CM, et al. Regulation of glucose homeostasis through a XBP-1-FoxO1 interaction. *Nature Med*. 2011;17:356–365.
345. Liu J, Lee J, Salazar Hernandez MA, et al. Treatment of obesity with celastrol. *Cell*. 2015;161:999–1011.

346. Ozcan U, Yilmaz E, Ozcan L, et al. Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science*. 2006;313:1137–1140.
347. Kamura T, Sato S, Haque D, et al. The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. *Genes Dev*. 1998;12:3872–3881.
348. Tyers M, Willems AR. One ring to rule a superfamily of E3 ubiquitin ligases [comment]. *Science*. 1999;284(601):603–604.
349. Zhang JG, Farley A, Nicholson SE, et al. The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proc Natl Acad Sci U S A*. 1999;96:2071–2076.
350. Stefan N, Haring HU. The role of hepatokines in metabolism. *Nat Rev Endocrinol*. 2013;9:144–152.

31

Classification and Diagnosis of Diabetes Mellitus

Aoife M. Egan and Sean F. Dinneen

OUTLINE

Definition, 524
Classification, 524
- Type 1 Diabetes, 525
- Type 2 Diabetes, 526
- Other Specific Types of Diabetes, 527
- Genetic Defects in **β** Cell Function, 527
- Genetic Defects in Insulin Action, 529
- Diseases of the Exocrine Pancreas, 529
- Endocrinopathies, 529
- Drug- or Chemical-Induced Diabetes, 529
- Infections, 530
- Uncommon Forms of Immune-Mediated Diabetes, 530
- Other Genetic Syndromes Sometimes Associated with Diabetes, 530
- Gestational Diabetes, 530

Diagnosis, 531
- Diagnostic Criteria, 531
- What Level of Fasting Plasma Glucose Constitutes Diabetes?, 532
- What Level of Fasting Plasma Glucose Constitutes Normality?, 532

What's New?, 534

DEFINITION

The term *diabetes mellitus* does not represent a single disease entity, but rather a set of disease states that share certain characteristics. Foremost among these is the presence of elevated plasma glucose levels. As is discussed in the following section, the presence of hyperglycemia in a patient is used both to diagnose diabetes and to guide management decisions, which are directed largely toward avoiding hyperglycemia. The hyperglycemia itself results from a combination of defects in insulin secretion, insulin action, or both.[1] An important characteristic of the various disease states that are labeled as diabetes is the development of end-organ damage in vital organs of the body, including the retina, the renal glomerulus, and peripheral nerves. Damage results, at least in part, from the long-term effects of hyperglycemia and is mediated through glycation of tissue proteins, increased activity of the polyol pathway, or other, as yet unrecognized, mechanisms.[2] Individual patients vary in their predisposition to develop these so-called microvascular complications. Because of this and because of the length of time they take to develop (frequently decades), the complications of diabetes cannot be used to classify or diagnose the disease. People with diabetes are at considerably greater risk of developing atherosclerotic disease affecting the coronary, cerebrovascular, peripheral arterial, or other parts of the circulation. A cause-and-effect relationship between chronic hyperglycemia and these so-called macrovascular complications of diabetes has not been as clearly established, although evidence linking the two is accumulating.[3] Any definition of diabetes that refers only to carbohydrate metabolism is incomplete. Oskar Minkowski is reputed to have first identified the association between the insulin-deficient pancreatectomized state of his laboratory dogs and the sweet taste of their urine. It has been suggested that if Minkowski had lacked a sense of taste but possessed a keen sense of smell, he might have smelled the ketones on the breath of his animals and thereby directed diabetes research toward the study of fat metabolism.[4] Disordered fat and protein metabolism must be included in a complete definition of the disease, although an emphasis on the pathogenesis of hyperglycemia continues to this day. To define a disease purely in biochemical terms is to diminish the component of the disease that leads to much physical, mental, and psychosocial distress for the many millions of people around the world who live with it every day. Chronic rheumatic diseases such as rheumatoid arthritis are not associated with any biochemical hallmark, and their definition is based largely on patient-reported symptoms and physical signs. Therefore, it is important not to forget the patient's perspective in any "definition" of the chronic disease referred to as diabetes.

KEY POINTS

- Diabetes mellitus represents a heterogeneous set of disorders that share raised plasma glucose levels as their defining characteristic.

CLASSIFICATION

Before 1979, a classification system for diabetes was not well established, and many different terms were used to describe essentially the same clinical entity. Following publication that year of the report of the National Diabetes Data Group (NDDG),[5] some order was brought to bear in this area. The recommendations of the NDDG were subsequently endorsed by the World Health Organization (WHO) in a publication in 1980, and minor modifications were later made in a document published in 1985.[6] This classification in large part was based on pharmacologic treatment of the disease. Insulin-dependent diabetes mellitus (IDDM) and non–insulin-dependent diabetes mellitus (NIDDM) were the two major forms of diabetes that had been identified. The term *IDDM* was used to describe patients who typically were lean at presentation, were prone to ketosis, and required insulin for survival. The term *NIDDM* was used to describe patients who typically were overweight or obese at presentation, were not prone to ketosis, and did not require insulin for survival. The NDDG also had categories for gestational diabetes and malnutrition-related diabetes mellitus (MRDM) and a category labeled “other types,” which included certain forms of diabetes for which a cause had been suggested at that time. As the terms *IDDM* and *NIDDM* became widely used during the 1980s and 1990s, several problems became apparent. The main problem arose from the fact that many patients with NIDDM ended up at some point in the course of their disease being treated with insulin and being misclassified as having IDDM or having the rather confusing term *insulin-requiring NIDDM* applied to them. In addition, as more information became available on the causes of the various forms of diabetes, it became apparent that a classification based on therapy was not always consistent with new insights into the pathogenesis of the various forms of diabetes. For this reason, the American Diabetes Association (ADA) convened an expert panel in 1995 to address the issue of classification. This panel published its recommendations in 1997,[7] and these were subsequently endorsed by a WHO consultation group in a 1998 report.[8] The main thrust of this proposal was to move away from a classification based on therapy and toward one based on pathogenesis. Four major categories were proposed: type 1 diabetes, type 2 diabetes, other specific types of diabetes (including categories for which a cause has been established), and gestational diabetes. The details of this classification system, which are outlined in Table 31.1 (with contemporary updates), are discussed in the following sections.[9]

KEY POINTS

- The major categories of diabetes mellitus include type 1 diabetes, type 2 diabetes, gestational diabetes, and other specific types of diabetes. Accurate classification informs the best therapeutic interventions.

Type 1 Diabetes

Type 1 diabetes is characterized by the development of a state of complete insulin deficiency arising as a consequence of β cell destruction. In its fully developed form, people with this condition, if deprived of exogenous insulin, will develop ketoacidosis, coma, and death. Biochemical testing reveals a marked reduction or complete absence of circulating C-peptide (a marker of insulin secretion) despite hyperglycemia. In 2009, approximately 167,000 youth were living with type 1 diabetes in the United States, with 18,000 new cases diagnosed each year.[10] While prevalence and incidence data among adults in the

TABLE 31.1 **Classification of Diabetes Mellitus**

1. Type 1 diabetes
    A. Immune-mediated
    B. Idiopathic
2. Type 2 diabetes
3. Other specific types
    A. Genetic defects in β cell function
        1. Maturity onset diabetes of the young (MODY) 1–14
        2. Neonatal diabetes
        3. Mitochondrial DNA
        4. Others
    B. Genetic defects in insulin action
        1. Donohue syndrome
        2. Rabson Mendenhall syndrome
        3. Type A insulin resistance
        4. Lipodystrophic diabetes
    C. Diseases of the exocrine pancreas
        1. Pancreatitis
        2. Trauma/pancreatectomy
        3. Neoplasia
        4. Cystic fibrosis
        5. Hemochromatosis
        6. Fibrocalculous pancreatic diabetes
        7. Others
    D. Endocrinopathies
        1. Cushing’s syndrome
        2. Acromegaly
        3. Glucagonoma
        4. Pheochromocytoma
        5. Somatostatinoma
        6. Aldosteronoma
        7. Hyperthyroidism
        8. Others
    E. Drug- or chemical-induced
        1. Immune checkpoint inhibitors
        2. Vacor
        3. Pentamidine
        4. Nicotinic acid
        5. Glucocorticoids
        6. Thyroid hormone
        7. Diazoxide
        8. β-Adrenergic agonists
        9. Thiazides
        10. Second-generation antipsychotics
        11. Protease inhibitors
        12. Pasireotide
        13. New-onset diabetes after transplantation (NODAT)
        14. Others
    F. Infections
        1. Congenital rubella
        2. Cytomegalovirus
        3. Others
    G. Uncommon forms of immune-mediated diabetes
        1. “Stiff-person” syndrome
        2. Anti–insulin receptor antibodies
        3. Others
    H. Other genetic syndromes sometimes associated with diabetes
        1. Down syndrome
        2. Klinefelter syndrome
        3. Turner syndrome
        4. Wolfram syndrome
        5. Friedreich ataxia
        6. Huntington chorea
        7. Lawrence–Moon–Biedel syndrome
        8. Myotonic dystrophy
        9. Porphyria
        10. Prader–Willi syndrome
        11. Others
4. Gestational diabetes mellitus

Adapted from the Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 1997;20:1183–1197.

United States are limited, information collected by the 1999–2010 National Health and Nutrition Examination Survey estimates that 740,000 to 970,000 Americans are affected.[11] Although the peak incidence is in childhood and early adolescence, this form of diabetes can occur at any age. The incidence of the disease shows marked regional variation, with the highest worldwide incidence reported in parts of Scandinavia.[10] Epidemiologic and immunologic research have led to the recognition of two major forms of type 1 diabetes based on the presence or absence of certain immunologic markers.

Autoimmune Type 1 Diabetes. Autoimmune type 1 diabetes is a prototypic organ-specific autoimmune disorder. Individuals who develop this form of diabetes are born with a genetic predisposition to autoimmune dysfunction, which may manifest in the development of other autoimmune conditions such as Addison disease, vitiligo, pernicious anemia, Hashimoto thyroiditis, or celiac disease. The genetic predisposition is not well understood but is known to be linked to the major histocompatibility locus on chromosome 6.[12] The presence of certain human lymphocyte antigen (HLA) haplotypes appears to predispose the individual to the disease, while other HLA haplotypes appear to be protective. In predisposed individuals, a poorly understood environmental trigger sets off a series of immunologic events that culminate in selective T-cell–mediated destruction of the β cells of the pancreatic islet. Many antigens have been investigated as potential triggers for the disease with inconclusive results. These include certain viral antigens,[13] as well as an antigen contained in cow's milk protein[14] and gluten.[15] A large series of investigations has been carried out for the "Diabetes Autoimmunity Study in the Young." These studies initially investigated the early detection of antiislet antibodies in children[16] and their role in the prediction of future type 1 diabetes.[17,18] The rate at which β cell destruction occurs varies from individual to individual and may be very brief, as is seen when type 1 diabetes presents in children or young adults,[18,19] or may be prolonged, as is seen in what has been called *latent autoimmune diabetes* in adults.[20] The rate of progression also depends on the number of autoantibodies, antibody specificity, and autoantibody titer.[21,22] Antibodies appear in the circulation early in the process of β cell destruction.[23] In type 1 diabetes, autoantibodies are believed to be markers (rather than true instigators) of the immune response, and, as already alluded to, our understanding of the precise immunology is limited. An ongoing clinical study will examine blood samples from those who develop type 1 diabetes at a very young age in the hope that studying those with extreme forms of the disease will facilitate discovery of specific pathologic mechanisms.[24]

The presence of autoantibodies can help to classify a newly diagnosed patient with diabetes, and screening for these autoantibodies has led to recognition of autoimmune type 1 diabetes in individuals who otherwise may have been labeled as having type 2 diabetes[25] (see later). Islet cell antibodies, the first autoantibodies to be discovered, are directed against a range of islet antigens. The best characterized autoantibodies are those directed against glutamic acid decarboxylase (GAD), an enzyme that is involved in γ-aminobutyric acid synthesis.[26] Isoforms of GAD are found in the central nervous system and in β cells of the pancreatic islet. Other autoantibodies include those directed against tyrosine phosphatases, namely IA-2 and IA-2β, as well as antibodies directed against zinc (ZnT8) and insulin itself (antiinsulin antibodies). Autoantibodies are present in 85% to 90% of individuals with type 1 diabetes at the time of presentation with fasting hyperglycemia.[16] Childhood-onset type 1 diabetes is associated with higher levels of autoantibody in serum. Testing for these autoantibodies is still restricted to a limited number of laboratories. The ADA states that autoantibodies are not required for the routine diagnosis of diabetes, but agree that standardized autoantibody tests can be used for classification of diabetes.[27] They also advise that screening for type 1 diabetes risk with a panel of islet autoantibodies is an option for a first-degree family member of a person with type 1 diabetes. In this situation, persistence of autoantibodies is a risk factor for progression to diabetes and may facilitate intervention in the setting of a clinical trial.[27]

Idiopathic Type 1 Diabetes. The term *idiopathic type 1 diabetes* is used to describe a small subset of individuals with type 1 diabetes who appear not to have an autoimmune basis for their β cell destruction.[28] Other features of this subtype include its occurrence predominantly in individuals of Black American or Asian ethnicity, its lack of HLA association, and its intermittent proneness to ketosis. A number of different terms, including *atypical diabetes, Flatbush diabetes,* and *type 1.5 diabetes,* have been used to describe this form of diabetes. The preferred term in the recent literature appears to be *ketosis-prone type 2 diabetes.*[29]

KEY POINTS

- Type 1 diabetes is characterized by total insulin deficiency. While this is typically caused by autoimmune destruction of the β cells in predisposed individuals, a minority will have idiopathic type 1 diabetes.

Type 2 Diabetes

Type 2 diabetes represents the most common form of diabetes. The number of adults with type 2 diabetes has doubled over the past three decades. Current estimates for the US population indicate that over 30 million people have type 2 diabetes,[30] while worldwide estimates for type 2 diabetes reach over 400 million.[31] The condition is characterized by hyperglycemia that arises from defects in insulin secretion and action, impaired glucose effectiveness, and impaired postprandial suppression of glucagon.[32,33] In any given individual, the degree to which these defects contribute to hyperglycemia may vary. The disease usually has its onset after 40 years of age, although type 2 diabetes is being seen increasingly in young adults and adolescents.[34] Although progressive β cell failure is believed by many to be an important part of the natural history of this form of diabetes,[35] the β cell destruction is not autoimmune-mediated[1] and typically does not progress to the point at which the patient becomes dependent on insulin for survival. Ketoacidosis is unusual in this form of diabetes, and, when it occurs, it usually does so in the setting of a major intercurrent illness such as myocardial infarction or stroke, or when the patient is undergoing treatment with certain drugs such as glucocorticoids or sodium-glucose cotransporter 2 inhibitors.[36] Individuals with type 2 diabetes are not at increased risk for autoimmune disease, but have a higher prevalence of metabolic abnormalities, including obesity, hypertension, and a dyslipidemia that is characterized by hypertriglyceridemia and low levels of high-density lipoprotein cholesterol. This combination of metabolic derangements is associated with a marked increase in risk for atherosclerotic disease. In fact, the prevalence of atherosclerotic disease in individuals with type 2 diabetes has led to the suggestion that, rather than one leading to the other, the two conditions may share common antecedents.[3] Insulin resistance may be an important predisposing factor for both conditions. An inflammatory paradigm of innate immune cell activation has been proposed as an underlying factor for both conditions.[37]

The cause of type 2 diabetes remains to be determined. Any pathogenetic model of the disease must include both genetic and environmental factors. Challenges in establishing the cause of type 2 diabetes include the following: (1) The disease lacks an easy-to-define phenotype and instead is characterized by considerable heterogeneity across different ethnic groups; this heterogeneity often is represented by a spectrum ranging from a predominant defect in insulin secretion on the one hand to a predominant defect in insulin action on the other. (2) The relatively late age of onset makes it difficult to establish large kindreds, and therefore limits genetic studies. (3) No easy-to-apply methods are available for screening populations for insulin resistance and defective insulin secretion. (4) The pathways that are involved in mediating insulin action are complex and are not fully understood; most authors believe that a single genetic defect will explain only a subset of the disease; it is much more likely that type 2 diabetes represents a set of disorders. Evidence to support a genetic component of the disease comes from the strong concordance for the disease that is seen among monozygotic twins.[38] On the other hand, the dramatic increase in incidence and prevalence of type 2 diabetes that accompanies the change to a so-called Westernized lifestyle strongly supports an environmental component as well.[39]

KEY POINTS

- Type 2 diabetes is the most common form of diabetes and is associated with defects in insulin secretion and action, impaired glucose effectiveness, and impaired postprandial glucagon suppression. Although progressive β cell failure is a feature of this condition, most patients are not dependent on insulin for survival.

Other Specific Types of Diabetes

Other specific types of diabetes are included in both the 1979 NDDG and the 1997 ADA classification systems.[5,7] Several changes occurred in the subtypes of diabetes listed between the two eras. In particular, monogenetic diabetes syndromes, including neonatal diabetes and maturity-onset diabetes of the young (MODY), have been better defined genetically; and the classification of monogenic diabetes has evolved from one based on clinical characteristics to one based on molecular genetics. While updates are reflected annually in the ADA Standards of Medical Care in Diabetes, this is a rapidly evolving area, and further changes to existing classification systems are anticipated.[27] Another change from the original 1979 classification system resulted from removal of MRDM and inclusion of fibrocalculous pancreatic diabetes (which was previously a subtype of MRDM) as a disease of the exocrine pancreas. The decision to remove MRDM as a separate entity resulted from an international conference on this subject.[40,41] The findings of the conference did not support a direct cause-and-effect relationship between protein-calorie malnutrition and the development of diabetes. Rather, it was thought that the presence of malnutrition could influence the manner in which diabetes might present in an otherwise predisposed individual.

Genetic Defects in β Cell Function

In the past, the term *MODY* was used to describe a subset of patients with a form of diabetes characterized by early age of onset of hyperglycemia with an autosomal dominant mode of inheritance. Mutations in certain genes that are involved in regulating insulin secretion have now been shown to be responsible for the hyperglycemia seen in MODY kindreds.[42] A karyogram of the various MODY mutations is shown in Fig 31.1. To date, 14 forms of MODY have been described,[43] and their clinical and genetic features are outlined in Table 31.2 (see Chapter 49). All forms of MODY are associated with defective insulin secretion without any significant degree of insulin resistance.[44,45] In the case of the most common subtype, GCK-MODY (MODY2), the

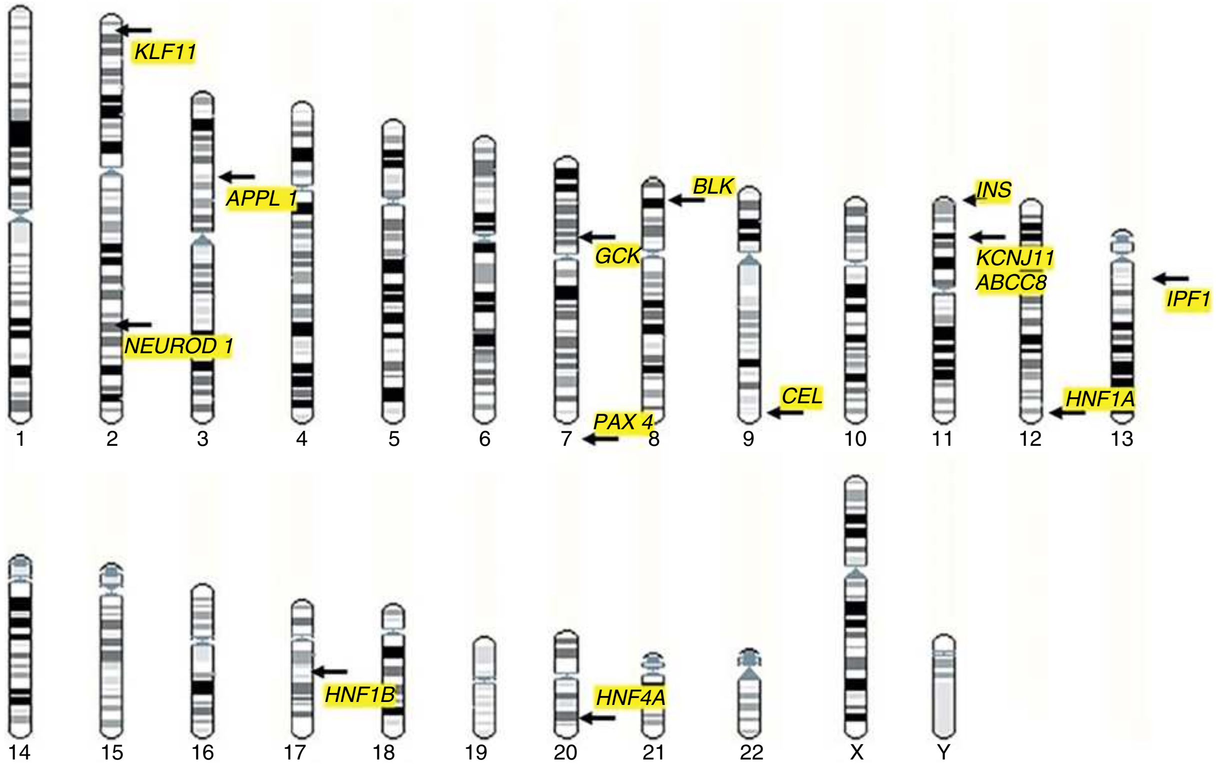


Figure 31.1 **Karyogram of the common and rare forms of maturity-onset diabetes of the young (MODY) mutations.** (From Aarthy R, Aston-Mourney K, Mikocka-Walus, et al. Clinical features, complications and treatment of rarer forms of maturity-onset diabetes of the young (MODY)—a review. *J Diabetes Complications*. 2021;35:107640.)

TABLE 31.2 Maturity-Onset Diabetes of the Young (MODY)-Related Genes and the Clinical Phenotypes Associated With Mutations in the Genes

MODY Type	Gene	Clinical Features of Heterozygous State	Most Common Treatment	Molecular Basis	Clinical Features of Homozygous State
MODY 1	*HNF-4α*	May have high birth weight and transient neonatal hypoglycemia; diabetes; microvascular complications (in many cases); reductions in serum concentrations of triglycerides, apolipoproteins AII and CIII, and Lp(a) lipoprotein	Sensitive to sulfonylurea drugs	Abnormal regulation of gene transcription in β cells, leading to a progressive defect in metabolic signaling of insulin secretion, β cell mass, or both	
MODY 2	Glucokinase	Impaired fasting glucose, small rise in 2-hour plasma glucose on oral glucose tolerance test	Diet and exercise, insulin therapy may be required during pregnancy	Nonprogressive defect in sensitivity of β cells to glucose due to reduced glucose phosphorylation; defect in hepatic storage of glucose as glycogen	Permanent neonatal diabetes requiring insulin treatment
MODY 3	*HNF-1α*	Diabetes, microvascular complications (in many cases), renal glycosuria, increased sensitivity to sulfonylurea drugs	Low-dose sulfonylureas, insulin	Abnormal regulation of gene transcription in β cells, leading to a progressive defect in metabolic signaling of insulin secretion, β cell mass, or both	
MODY 4	*IPF-1*	Diabetes	Oral hypoglycemic agent, insulin	Abnormal transcriptional regulation of β cell development and function	Pancreatic agenesis and neonatal diabetes requiring insulin treatment
MODY 5	*HNF-1β*	Diabetes; renal cysts and other abnormalities of renal development; progressive nondiabetic renal dysfunction, leading to chronic renal insufficiency and failure; internal genital abnormalities (in female carriers)	A minority respond to sulfonylurea drugs, Insulin	Abnormal regulation of gene transcription in β cells, leading to a defect in metabolic signaling of insulin secretion, β cell mass, or both	
MODY 6	NeuroD1 or *BETA2*	Diabetes, neonatal diabetes	Insulin	Abnormal transcriptional regulation of β cell development and function	Neurological abnormalities
MODY 7	*KLF11*	Diabetes	Oral hypoglycemic agent, insulin	Abnormal *PDX1* transcription in the β cell	
MODY 8	*CEL*	Diabetes +/-exocrine pancreas dysfunction, associated with increased risk of diabetic kidney disease	Insulin	Abnormal endocrine and exocrine pancreas regulation	
MODY 9	*PAX4*	Diabetes with increased risk of microvascular complications	Oral hypoglycemic agent, insulin required for the majority	Abnormal regulation of gene transcription in β cells	
MODY 10	*INS*	Diabetes	Oral hypoglycemic agent, but usually requires insulin	Insulin gene mutation, defective insulin secretion	
MODY 11	*BLK*	Diabetes and obesity	Oral hypoglycemic agent, insulin	Defect in insulin production	
MODY 12	*ADCC8*	Diabetes	Sensitive to sulfonylurea drugs, insulin may be required	Defect in insulin release	
MODY 13	*KCNJ11*	Diabetes	Sensitive to sulfonylurea drugs	Defect in insulin release	Neonatal diabetes
MODY 14	*APPL1*	Diabetes, may have developmental delay	Oral hypoglycemic agent, insulin	Defect in insulin secretion	

BETA2, β cell Ebox transactivator 2; *HNF,* hepatocyte nuclear factor; *IPF,* insulin promoter factor; *MODY,* maturity-onset diabetes of the young; *NeuroD1,* neurogenic differentiation factor 1.

From American Diabetes Association: Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2021;44:S15-S33; and Aarthy R, Aston-Mourney K, Mikocka-Walus, et al. Clinical features, complications and treatment of rarer forms of maturity-onset diabetes of the young (MODY) - A review. *J Diabetes Complications*. 2021;35:107640.

pathophysiologic mechanism is clear, because the affected enzyme is involved in phosphorylating glucose, one of the first steps in glucose metabolism. A glucokinase mutation therefore decreases the ability of the β cell to sense glucose. MODY2 is associated with relatively mild hyperglycemia that is best left untreated. MODY1 and MODY3, on the other hand, can be associated with more severe hyperglycemia, a greater likelihood that insulin will be required for management, and a higher propensity to develop complications.[42] The link between diabetes and mutations in the genes for hepatocyte nuclear factor (HNF)-1a, HNF-1β (hepatocyte transcription factors also expressed in β cells), and HNF-4a (a member of the steroid-thyroid hormone superfamily and an upstream regulator of HNF-1a), appears to involve regulation of expression of the insulin gene.[46] As screening for MODY is typically physician-based rather than population-based, the majority of cases remain misclassified, and it has been suggested that MODY may account for between 2% and 5% of cases of type 2 diabetes.[47] In the United Kingdom, the marked regional variation in the prevalence of confirmed MODY directly results from differences in referral rates and likely reflects access to specialized centers with genetic testing.[48]

Neonatal diabetes occurs in children under 6 months of age. It is rare and may be transient or permanent.[49] Several genetic defects have been identified in these patients. Transient neonatal diabetes is typically due to overexpression of genes on chromosome 6q24, and, most notably, mutations in the adenosine triphosphate (ATP)-sensitive potassium channel (KATP) of the β cell cause permanent diabetes.[50] The KATP channel comprises two subunits (Kir6.2 and SUR1), and mutations in these can result in loss of function (leading to channel closure and neonatal hyperinsulinemic hypoglycemia) or gain of function (leading to channel opening and neonatal diabetes). Although rare, elucidation of these conditions and the molecular mechanisms underlying them has improved our understanding of this area of human biology. Moreover, diabetes arising from KATP channelopathies can retain sensitivity to the insulin secretagogue effects of sulfonylureas.[51]

Other genetic disorders that are associated with impaired β cell function include certain maternally inherited forms of diabetes with a mutation in mitochondrial DNA,[52] disorders that lead to impaired conversion of proinsulin to insulin,[53] and disorders that lead to synthesis of an aberrant form of the insulin molecule.[54] Of the former conditions, diabetes associated with an A-to-G transition at nucleotide 3243 in mitochondrial transfer RNA has been best characterized and appears to have a wide phenotypic expression from type 2 through to type 1 diabetes. The latter two conditions are inherited in an autosomal dominant manner and are associated with relatively mild glucose intolerance.

As genetic testing has become more accessible, the ADA now recommends that all children diagnosed with diabetes in the first 6 months of life have genetic testing for neonatal diabetes. They further advise that children and young adults who have diabetes not characteristic of type 1 or 2 diabetes that occurs in successive generations should have genetic testing for MODY.[27]

Genetic Defects in Insulin Action

Insulin resistance is closely associated with type 2 diabetes and, while commonly found in those with obesity, may rarely occur in a severe form in people with genetic variations. Semple et al. propose two major categories of syndromes of severe insulin resistance: (1) primary insulin-signaling defects, and (2) insulin resistance secondary to adipose tissue abnormalities (including syndromes of severe obesity and lipodystrophy, the latter of which is frequently associated with diabetes mellitus).[55] Abnormal glucose homeostasis, ovarian dysfunction, and acanthosis nigricans are features that occur in nearly all forms of insulin resistance; however, other manifestations including dyslipidemia and hepatic steatosis are limited to specific subtypes.[55]

Insulin receptor defects include Donohue syndrome (formerly "leprechaunism") and Rabson Mendenhall syndrome.[56,57] Both syndromes are autosomal recessive, present within the first decade of life, and are associated with diabetes, hyperinsulinemia, and altered growth. More commonly, insulin receptor defects occur later in life with an autosomal dominant inheritance pattern and features of oligomenorrhea, hyperandrogenism, and acanthosis nigricans.[55] This includes syndromes that were previously named HAIR-AN (hyperandrogenism, insulin resistance, and acanthosis nigricans in females) and type A insulin resistance (severe insulin resistance in nonobese patients).[58,59] Of note, type B insulin resistance is a term used to describe acquired insulin resistance in the presence of anti–insulin receptor antibodies (see later).

Lipodystrophy is characterized by pathological adipose tissue deficiency with multiple subtypes. Broadly speaking, lipodystrophy may be partial or generalized, and inherited or acquired (see Chapter 24).[60] Lipodystrophy should be included in the differential diagnosis in severely insulin-resistant patients with dyslipidemia and/or nonalcoholic fatty liver disease.[55] The use of genetic testing in the classification of genetic syndromes of insulin action remains an area for evaluation within subspecialized services at this time.

Diseases of the Exocrine Pancreas

Hyperglycemia can occur during an episode of acute pancreatitis and is associated with a poor prognosis. It is unusual for permanent diabetes to develop following a single episode of acute pancreatitis. Furthermore, removal of up to 90% of the pancreas does not always cause diabetes. On the other hand, diabetes has been reported in association with very small pancreatic adenocarcinomas, leading some investigators to speculate that these tumors produce some diabetogenic factor or factors.[61] Five to fifteen percent of patients with cystic fibrosis develop diabetes.[62] Insulin replacement in cystic fibrosis–related diabetes (CFRD) is often challenging, as these individuals retain insulin sensitivity and are prone to hypoglycemic episodes. Moreover, insulin requirement in the setting of CFRD is complicated by malabsorption and the associated need for enteral feeding, as well as by infectious exacerbations associated with underlying bronchiectasis.[63] Hemochromatosis can cause diabetes.[64] Because glucose tolerance may improve with venesection, this condition represents an important, potentially reversible cause of diabetes. Fibrocalculous pancreatic diabetes is seen mainly in the tropics and is associated with abdominal pain and calcification of the pancreas on abdominal imaging.[65] The natural history of this form of diabetes has been established,[66] along with a genetic marker of the disease.[67]

Endocrinopathies

Cortisol, growth hormone, glucagon, and the catecholamines (epinephrine and norepinephrine) can antagonize insulin action. Tumors that produce these hormones in excess lead to Cushing syndrome, acromegaly, glucagonoma, and pheochromocytoma, respectively. Although all of these conditions are associated with a degree of glucose intolerance, overt diabetes develops only in a subset of patients. The importance of recognizing these secondary forms of diabetes lies in the fact that resection of the underlying tumor can cure diabetes. The hyperglycemia that is seen in the setting of aldosterone-producing adenomas and somatostatinomas results from alteration of insulin secretion. An added challenge to diabetes mellitus arising from endocrinopathy is that somatostatin analogs often used to control conditions such as growth hormone excess and, more recently, glucocorticoid excess can induce diabetes in their own right (described in the following section).

Drug- or Chemical-Induced Diabetes

While the development of immune checkpoint inhibitors (ICIs) has improved the clinical outcomes for many people with cancer, immune-related adverse effects can involve various organs, including the pancreas.[68] This class of drugs include cytotoxic T cell–associated protein 4 inhibitors, programmed cell death protein-1 inhibitors, and programmed death-ligand 1 inhibitors. ICI-associated diabetes mellitus reportedly occurs in 0.2% to 1.27% of people treated with these medications following a median duration of therapy of approximately 4 months.[68,69] The condition is characterized by pancreatic atrophy, and the associated hyperglycemia is irreversible, necessitating lifelong insulin therapy. Presentation is typically fulminant, often with diabetic ketoacidosis. Those who have positive GAD65 antibodies prior to therapy and those treated with combinations of these medications appear to be at greatest risk.[70]

Additional compounds that are cytotoxic to β cells include the rat poison vacor and the antipneumocystis drug pentamidine. The hyperglycemia that results from these agents is usually not reversible. Thiazide diuretics can inhibit insulin secretion by causing hypokalemia, and this effect may be picked up on an oral glucose tolerance test as early as 4 weeks following the initiation of therapy. However, these agents do not induce diabetes mellitus in the absence of underlying susceptibility.[71] Glucocorticoids and nicotinic acid cause hyperglycemia by impairing insulin action. Glucocorticoid excess, whether endogenous or exogenous, is also independently associated with many so-called secondary complications of diabetes such as atherosclerosis and cardiovascular disease.[72] "New-onset diabetes after transplantation" is a term used to describe individuals who develop diabetes after transplant and excludes posttransplant hyperglycemia that resolves by the time of discharge from hospital.[27] While this cohort of patients have many risk factors for diabetes development (including severe physiological stress), it is believed that immunosuppressive therapies are a major contributor, particularly in those with preexisting risk factors for type 2 diabetes. Protease inhibitors (PIs) commonly used in the treatment of individuals with human immunodeficiency virus infection can cause hyperglycemia. This effect of PIs probably results from inducing a form of partial lipodystrophy that selects the face and limbs, while causing fat accumulation in the region around the waist.[73] The so-called *atypical* or *second-generation* antipsychotics (for example, olanzapine, clozapine, and quetiapine) are associated with metabolic disorders, including overt diabetes mellitus. The ADA published a consensus statement highlighting this association and identifying the need for clinical research in this area.[74] Although it was initially assumed that the risk of diabetes was directly linked to medication-associated weight gain, there is accumulating evidence that these medications independently decrease insulin sensitivity and insulin secretory capacity.[75] Finally, somatostatin analogs, often used in the treatment of endocrinopathy and neuroendocrine tumors, also inhibit insulin release. This is particularly true for pasireotide, an agent that has a broader selectivity across the somatostatin receptor subtypes. In endocrinopathies associated with insulin resistance such as Cushing disease or acromegaly, use of these agents may result in overt diabetes.[76]

Infections

Certain viral infections, including rubella[77] and coxsackie B virus,[78] have been associated with diabetes. Some studies suggest that a viral infection can trigger autoimmune destruction of β cells in genetically predisposed individuals, leading to autoimmune type 1 diabetes. During the coronavirus disease 2019 (COVID-19) pandemic, a relationship between COVID-19 and diabetes was noted. Like many viral illnesses, the risk of severe COVID-19 infection is higher in those with preexisting diabetes.[79] However, new-onset diabetes is observed in patients with COVID-19 pneumonia more frequently than among those with non–COVID-19 pneumonia, providing support for a potential diabetogenic effect of COVID-19. Viral infection of pancreatic β cells is one proposed mechanism, with significant work ongoing in this area.[80]

Uncommon Forms of Immune-Mediated Diabetes

The stiff-person syndrome is a rare neurologic syndrome characterized by spasticity of the axial muscles. It is associated with very high titers of anti-GAD antibodies, and up to one third of patients develop diabetes.[81] Autoantibodies directed against the insulin receptor represent another rare cause of diabetes (referred to as *type B insulin resistance*).[82] These antibodies have the potential to change from being receptor antagonists (causing insulin resistance) to being receptor agonists (leading to potentially life-threatening hypoglycemia). Spontaneous remission of antibody production can occur. The syndrome is typically seen in Black American females in association with other autoimmune diseases (most commonly systemic lupus erythematosus).

Other Genetic Syndromes Sometimes Associated With Diabetes

Many of the genetic syndromes listed under H in Table 31.1 are known to be associated with diabetes. Wolfram syndrome, also known as DIDMOAD (reflecting the components of the disorder: **d**iabetes **i**nsipidus, **d**iabetes **m**ellitus, **o**ptic **a**trophy, and **d**eafness), is a rare autosomal recessive disorder caused by mutations in the *WFS1* gene on the short arm of chromosome 4. Polymorphisms in the *WFS1* gene have been linked to an increased risk for type 2 diabetes,[83] providing an example of how the study of rare disorders can improve our understanding of common conditions.

KEY POINTS

- "Other forms of diabetes" is a broad category that includes genetic forms of diabetes, diseases of the exocrine pancreas, diabetes related to endocrinopathies, drug-induced diabetes, and diabetes due to specific infections.

Gestational Diabetes

Gestational diabetes mellitus (GDM) is a common medical complication of pregnancy and is defined as diabetes with onset or first recognition during pregnancy, excluding diabetes that was clearly present prior to gestation. Risk factors for GDM include age (it is more common among older women), ethnicity (higher rates are seen among women from ethnic groups with a high incidence of type 2 diabetes), prepregnancy body mass index (the risk increases with degree of obesity), parity (the risk increases with the number of previous pregnancies), and family history of diabetes. A previous pregnancy that was complicated by GDM or a history of delivery of a macrosomic infant also represents a strong risk factor for future GDM.[84] Dysglycemia during pregnancy is associated with complications for both mother and fetus. Fetal/neonatal complications include large for gestational age birth weight, macrosomia, and neonatal hypoglycemia. Maternal complications include higher rates of cesarean section.[84,85] Appropriate diagnosis and management of gestational diabetes potentially reduces the incidence of these related complications. In the long term, women with GDM are at greater risk of type 2 diabetes, and postpartum testing of glycemia is necessary in all women with a diagnosis of GDM.[86]

Table 31.3 outlines commonly accepted approaches to diagnosing GDM. The traditional two-step approach uses diagnostic cutoffs that are predominantly based on the woman's risk of type 2 diabetes following the pregnancy, and outcomes relating to the index pregnancy were not considered at the time of development. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study aimed to clarify the relationship between adverse pregnancy outcomes and maternal glucose levels (below the threshold for overt diabetes).[87] The results from this study suggested that updated threshold values were necessary to develop an "outcomes-based" system for diagnosis of GDM.

TABLE 31.3 Screening for and Diagnosis of Gestational Diabetes Mellitus[*]

One-step strategy:

In women not previously diagnosed with overt diabetes, perform a 75-g OGTT (with measurement of plasma glucose fasting and at 1- and 2-hours post–glucose load) at 24 to 28 weeks' gestation when the patient is fasting.

The diagnosis of gestational diabetes mellitus is made when any of the following plasma glucose values are met or exceeded:

Fasting: 92 mg/dL (5.1 mmol/L)
1-hour: 180 mg/dL (10.0 mmol/L)
2-hour: 153 mg/dL (8.5 mmol/L)

Two-step strategy:

Step 1: Perform a 50-g glucose challenge test (nonfasting), with plasma glucose measurement at 1 hour, at 24–28 weeks of gestation in women not previously diagnosed with diabetes.

If the plasma glucose level measured 1 h after the load is ≥130, 135, or 140 mg/dL (7.2, 7.5, or 7.8 mmol/L, respectively)[†], proceed to a 100-g OGTT.

Step 2: A 100-g OGTT should be performed when the patient is fasting.

The diagnosis of GDM is made when at least two of the following four plasma glucose levels are met or exceeded:

Fasting: 95 mg/dL (5.3 mmol/L)
1-hour: 180 mg/dL (10.0 mmol/L)
2-hour: 155 mg/dL (8.6 mmol/L)
3-hour: 140 mg/dL (7.8 mmol/L)

OGTT, Oral glucose tolerance test; *GDM,* gestational diabetes mellitus.Diagnosis based on International Association of Diabetes and Pregnancy Study Groups recommendations and endorsed by the American Diabetes Association.

[*] There is no consensus regarding universal screening versus screening in "at-risk" individuals.

[†] Each of these thresholds may be used for the glucose challenge test. The choice is based upon the trade-off between sensitivity and specificity, and individual clinical practices vary in their approach.

(From International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy, *Diabetes Care.* 2010;33:676–682; and American Diabetes Association: Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2021;44:S15–S33).

Consequently, the International Association of Diabetes and Pregnancy Study Groups (IADPSG) published a consensus statement that sets out updated recommendations for the diagnosis of GDM using the 75-g oral glucose tolerance test (OGTT), performed between 24 and 28 weeks (Table 31.3).[86] The diagnostic cutoffs were chosen based on an adjusted odds ratio of 1.75 for three neonatal outcome variables. These variables had a strong, predominantly linear association with maternal glucose and included birth weight greater than the ninetieth percentile, percent body fat greater than the ninetieth percentile, and cord C-peptide greater than the ninetieth centile.

Worldwide, uptake of the IADPSG guidelines is variable. For example, the International Federation of Gynecology and Obstetrics and the WHO have endorsed the IADPSG criteria,[88,89] and the ADA also recommends these criteria but acknowledges that a two-step strategy can be used.[27] In other jurisdictions such as the United Kingdom, guidelines reflect an adaptation of the IADPSG guidelines based on local prevalence and cost analyses.[90] The prevalence of gestational diabetes increases in parallel with the prevalence of type 2 diabetes in a population. Previous data from a large US healthcare organization described a prevalence of 7.4 cases per 100 pregnancies.[91] Using diagnostic thresholds as set out by the IADPSG, the prevalence of GDM is far higher than that found using a two-step approach. For example, the average prevalence for all HAPO study participants was 17.8%, but this varied from 9.3% at an Israeli center to 25.5% at a US center.[92] Concerns have been raised that the IADPSG criteria impose a diagnosis of GDM on many women who have very modest dysglycemia, without a clear treatment benefit. A recent pragmatic randomized controlled trial supports this argument and reported that, despite more diagnosis of GDM with the IADPSG criteria versus a two-step approach (16.5% vs. 8.5%), there was no significant between-group differences in the risk of primary outcomes relating to pregnancy complications.[93] These observations have led to the suggestion that adoption of diagnostic criteria defining GDM should be based on response to treatment at different diagnostic thresholds of maternal glucose.[94] This would require a randomized controlled trial that would be challenging to design and execute.

Screening for GDM presents a further challenge, and currently there is no consensus on who should be screened.[86,95] In some countries (e.g., the United States), universal screening is undertaken routinely, and this is the recommendation of the IADPSG.[86] However, in other regions (e.g., parts of Europe), only women who are believed to be at high risk are screened.[90] Timing of GDM screening and diagnosis is another area of controversy. The majority of guidelines recommend screening at 24 to 28 weeks' gestation, which coincides with a sharp increase in insulin resistance and risk of hyperglycemia.[27,84] However, the rising tide of obesity and type 2 diabetes in pregnancy requires screening for undiagnosed diabetes as near to conception as possible. Screening for undiagnosed, pregestational diabetes in early pregnancy also identifies women with hyperglycemia less than overt diabetes, but at increased risk of adverse outcomes. There is no agreed approach to diagnosis and management of these women, and while some institutions apply standard GDM criteria to at-risk women in early pregnancy, this practice is not evidence-based. It is hoped that an ongoing randomized controlled trial will clarify if diagnosis and treatment of GDM earlier in pregnancy will result in improved outcomes.

KEY POINTS

- Gestational diabetes mellitus (GDM) is associated with an increased risk of pregnancy complications. While the associated hyperglycemia typically resolves postpartum, women with GDM have a lifelong increased risk of type 2 diabetes.

DIAGNOSIS

Diagnostic Criteria

The modern-day classification of diabetes is based on clinical presentation combined with (in some cases) genetic and autoantibody testing. The diagnosis of diabetes is based on measurement of glucose and/or glycated hemoglobin (HbA_{1c}). A historical perspective is necessary to understand how we got to this point. The Expert Committee of the ADA, in their 1997 report, recommended changes in the diagnostic criteria used for diabetes, in addition to recommending changes in the classification system.[7] These recommendations were subsequently endorsed by the WHO[8] (Table 31.4). Major changes included a reduction in the fasting plasma glucose cut point used to diagnose diabetes from 140 mg/dL to 126 mg/dL and a recommendation to use fasting plasma glucose rather than an OGTT for diagnosis. The ability to use casual (i.e., random) plasma glucose levels in patients with hyperglycemic symptoms was retained, as was the requirement that, in asymptomatic individuals, a diagnosis should be based on testing carried out on more than one occasion. The ADA report based its criteria on use of the fasting plasma glucose level, and the WHO report included equivalent cut points for whole-blood venous and capillary glucose.[8] The

TABLE 31.4 Diagnostic Thresholds for Diabetes and Lesser Degrees of Impaired Glucose Regulation*

Category	Fasting Plasma Glucose	2-Hour Plasma Glucose	HbA$_{1c}$¶
Normal	<100 mg/dL (<5.6 mmol/L)	<140 mg/dL (<7.8 mmol/L)	≤5.7%
IFG	100–125 mg/dL* (5.6–6.9 mmol/L)	—	❖
IGT	—	140–199 mg/dL* (7.8–11.0 mmol/L)	
Diabetes	≥126 mg/dL† (≥7.0 mmol/L)	≥200 mg/dL (≥11.1 mmol/L)	≥6.5%

IFG, Impaired fasting glucose; *IGT* impaired glucose tolerance; HbA$_{1c}$, glycated hemoglobin.

*When both tests are performed, IFG or IGT should be diagnosed only if diabetes is not diagnosed by the other test.

†A diagnosis of diabetes needs to be confirmed on a separate day.

¶The test should be performed in a laboratory using a method that is National Glycohemoglobin Standardization Program (or equivalent) certified and standardized to the Diabetes Control and Complications Trial assay.

❖The American Diabetes Association recommends a range of 5.7%–6.4% as identifying individuals with prediabetes. The use of HbA$_{1c}$ for the diagnosis of prediabetes has not been endorsed by the World Health Organization.

Expert Committee of the ADA published two further reports in 2003 and 2009.[96,97] While recommendations regarding diagnostic criteria for diabetes on the basis of fasting glucose, OGTT, and random glucose have remained unchanged, the 2009 report also added recommendations for use of HbA$_{1c}$ in the diagnosis of diabetes mellitus (see Table 31.4). These updated recommendations for HbA$_{1c}$ have subsequently been included in ADA Clinical Practice Recommendations.[9]

In the 1997 report, the ADA also introduced a new system for classifying the intermediate zone between normal glucose tolerance and overt diabetes.[7] A new category called *impaired fasting glucose* (IFG) was introduced, along with the preexisting category of impaired glucose tolerance (IGT). The presence of one or both of these conditions is often referred to as *prediabetes.*

While threshold values for the diagnosis of diabetes were originally formulated based on the risk for developing retinopathy, the basis for the "diagnosis" of these intermediate categories of hyperglycemia is the risk for developing overt diabetes, hence the terminology "Categories of Increased Risk for Diabetes."[97] The ADA diagnosis of prediabetes uses threshold values of fasting glucose (IFG: 100–125 mg/dL) or 2-hour glucose following a 75-g glucose load (IGT: 140–199 mg/dL).[27] However, there is no overall consensus between the ADA and the WHO for threshold values of fasting glucose, with the latter using an IFG cutoff of 110 mg/dL.[8] The ADA recommends a HbA$_{1c}$ range of 5.7% to 6.4% as identifying individuals with prediabetes.[9] The use of HbA$_{1c}$ for the diagnosis of prediabetes has not been endorsed by the WHO. The main questions of clinical relevance are: (1) what constitutes diabetes? and (2) what constitutes "normal"? We will address both questions now.

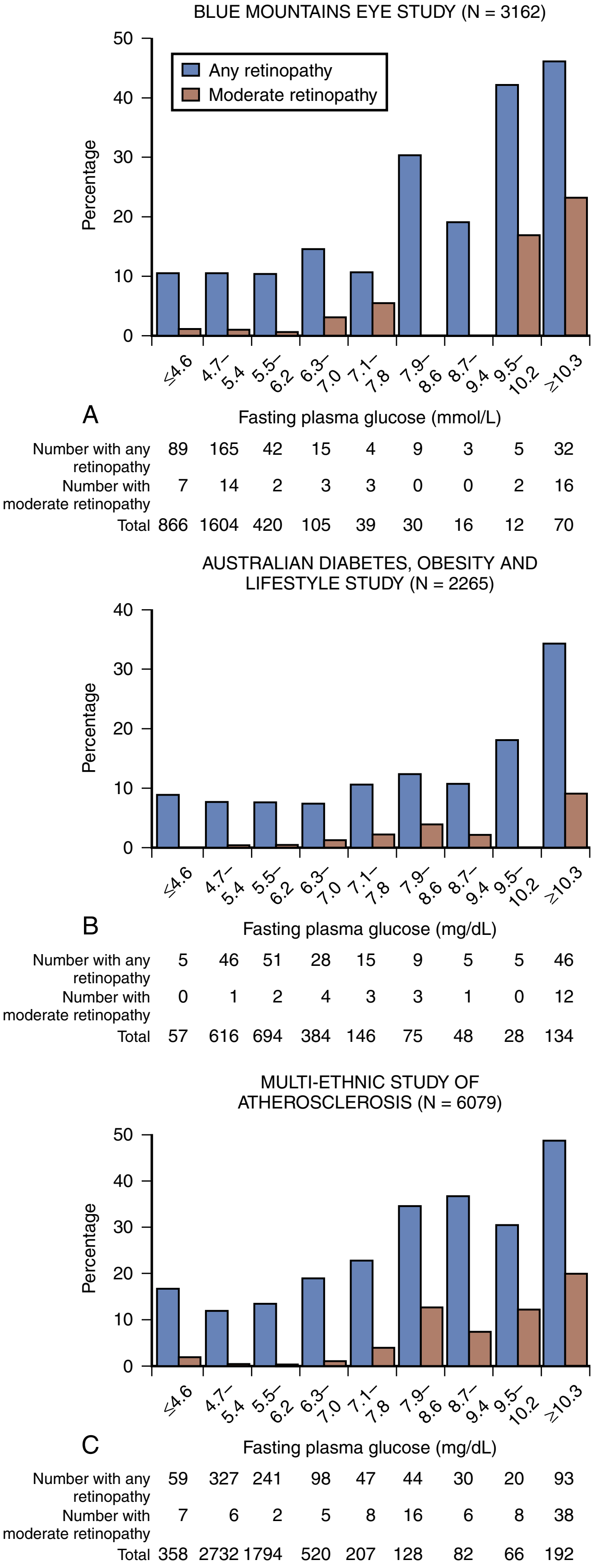
What Level of Fasting Plasma Glucose Constitutes Diabetes?

The development of microvascular changes, particularly in the retina, is one of the hallmarks of diabetes mellitus. Because these complications can take many years to develop, it is not possible to use their presence to diagnose the condition. Instead, the level of plasma glucose that confers a risk for occurrence of these changes is used for diagnostic purposes. At the time the 1997 report was written, the best estimate of this level was fasting plasma glucose of 126 mg/dL. This value was based on data from three epidemiologic studies involving Pima Indians, Egyptians, and a US national sample.[7] In all three studies, there appeared to be a threshold of risk (i.e., a level below which the risk for retinopathy was negligible, and above which the risk rose steeply). In the 2003 report of the expert committee, reference was made to a separate analysis of Pima Indian data, suggesting that the threshold at which retinopathy risk increased may be lower than that previously reported.[98] However, the presence of a threshold was not disputed, and the 126 mg/dL cut point for diagnosis remained in place. An analysis of data from two cross-sectional Australian cohort studies and one longitudinal US cohort study has raised doubt about the presence of a retinopathy threshold (Fig. 31.2).[99] All three studies used multiple-field digital retinal photography to identify retinopathy, a more sophisticated method than that used in earlier studies that informed the ADA expert committee decision. The risk for retinopathy in all three studies appeared to be continuous across the entire range of plasma glucose, with up to 10% of individuals demonstrating some degree of retinopathy at levels of fasting plasma glucose well within the range considered normal. These data are consistent with an earlier report from the Diabetes Prevention Program, which also identified retinopathy in individuals with IGT.[100] Taken together, these data suggest that current diagnostic criteria may need further revision. Nonetheless, the current threshold values for glucose in the diagnosis of diabetes have remained unchanged in the most recent recommendations.[8,97]

The most significant new recommendation from the 2009 Expert Committee of the ADA was the inclusion of an HbA$_{1c}$ value of greater than or equal to 6.5% as a criterion for the diagnosis of diabetes.[97] The measurement of HbA$_{1c}$ has been the mainstay of the assessment of glycemic control in clinical practice for many years. However, because of poor standardization of the assay in the past, HbA$_{1c}$ results could not previously be compared across populations. Standardization of the HbA$_{1c}$ assay has been completed in the United States. A further complication of the use of HbA$_{1c}$ in the diagnosis of diabetes is the recommendation by the International Federation of Clinical Chemists that the units used for reporting results be changed from percentage (as used in Diabetes Control and Complications Trial [DCCT]-aligned assays in the United States) to millimoles per mole (mmol/mol). This recommendation has been adopted by many (but not all) European countries. The diagnostic value of greater than or equal to 6.5% (for DCCT-aligned assays) was chosen based on its association with the inflection point for retinopathy, upon which the fasting and 2-hour glucose values have been chosen. HbA$_{1c}$ offers a useful so-called "chronic measurement" upon which to base the diagnosis of diabetes. Measurement does not require the patient to fast or consume a glucose load. However, these benefits must be balanced with consideration of the higher cost of the assay as well as errors associated with high red blood cell turnover states and hemoglobinopathies.

What Level of Fasting Plasma Glucose Constitutes Normality?

Less straightforward than diagnosing overt diabetes has been the characterization of intermediate levels of (what has been termed) *dysglycemia.* At what level does one cross the boundary between normal and abnormal? Glucose is a continuous variable, and therefore choosing a cut point to define the upper limit of normal fasting glucose (and the lower limit of IFG), while informed by clinical risk and observation, is ultimately arbitrary. The Expert Committee, in its 1997 report, chose a value of 110 mg/dL but revised this value to 100 mg/dL in 2003.[96] The updated lower recommendation was based on receiver operator



	≤4.6	4.7–5.4	5.5–6.2	6.3–7.0	7.1–7.8	7.9–8.6	8.7–9.4	9.5–10.2	≥10.3
Number with any retinopathy	89	165	42	15	4	9	3	5	32
Number with moderate retinopathy	7	14	2	3	3	0	0	2	16
Total	866	1604	420	105	39	30	16	12	70

	≤4.6	4.7–5.4	5.5–6.2	6.3–7.0	7.1–7.8	7.9–8.6	8.7–9.4	9.5–10.2	≥10.3
Number with any retinopathy	5	46	51	28	15	9	5	5	46
Number with moderate retinopathy	0	1	2	4	3	3	1	0	12
Total	57	616	694	384	146	75	48	28	134

	≤4.6	4.7–5.4	5.5–6.2	6.3–7.0	7.1–7.8	7.9–8.6	8.7–9.4	9.5–10.2	≥10.3
Number with any retinopathy	59	327	241	98	47	44	30	20	93
Number with moderate retinopathy	7	6	2	5	8	16	6	8	38
Total	358	2732	1794	520	207	128	82	66	192

Figure 31.2 **Prevalence of retinopathy across a wide range of fasting plasma glucose levels in three epidemiologic studies using digital retinal photography to screen the retina.** (Data from Wong TY, Liew G, Tapp RJ, et al. Relation between fasting glucose and retinopathy for diagnosis of diabetes: three population-based cross-sectional studies. *Lancet.* 2008;371:736–743.)

characteristic (ROC) curve analyses of the ability of various baseline levels of fasting plasma glucose to predict future diabetes. ROC curves were derived from cohort studies available to the committee at the time but were not published in the report.

A factor that undoubtedly influenced the committee was a desire to achieve "equivalence" (in terms of predicting future diabetes) between IFG and IGT. The term *equivalence* can be considered as cross-sectional or prospective. In cross-sectional terms, complete equivalence would amount to those patients with IFG also having IGT. This was not the case when the cut point was at 110 mg/dL and still is not the case with the cut point at 100 mg/dL. In fact, it has been suggested that these two states of "prediabetes" may be different biological entities, with altered insulin action and altered β cell function contributing in different ways to their pathogenesis.[101] A lesser degree of cross-sectional equivalence would be a situation in which population prevalences of IFG and IGT are the same. Lowering the cut point to 100 mg/dL did help move things toward equivalence by this definition. In prospective terms, equivalence amounts to an equal ability of these two intermediate states of hyperglycemia to predict future events, such as overt diabetes and cardiovascular disease or mortality. This is the point about which most of the controversy has arisen. A large European epidemiologic consortium published data clearly demonstrating that 2-hour plasma glucose following an OGTT was superior to fasting plasma glucose in predicting future cardiovascular events.[102] Members argued that the OGTT should not be discarded from clinical practice or from epidemiologic research. Many studies have looked at the ability of fasting plasma glucose to predict future diabetes. One of these, the Baltimore Longitudinal Study of Aging (BLSA), followed a cohort of more than 800 subjects with serial OGTTs for up to 20 years.[103] Investigators documented the natural history of progression from normal glucose tolerance to intermediate states of hyperglycemia to overt diabetes. Many subjects reverted to normal from intermediate states of hyperglycemia, but in general, a slow, steady progression to diabetes was seen, with male gender and obesity increasing risk. By altering the threshold that was used to define the lower limit of IFG, BLSA investigators showed that much of the "discrepancy" between rates of progression from IFG versus IGT to overt diabetes was a function of the threshold used rather than of any inherent biological difference between the two states of hyperglycemia. Using the definition of a threshold that we discussed previously, one would expect that the cut point between normal and IFG would be associated with a significant increase in risk. This is not the case. Several investigators have demonstrated that risk for future diabetes and risk for cardiovascular events increase across the continuum of glucose levels from normal through IFG or IGT.[104-107] There does not appear to be a threshold of risk.

Has HbA_{1c} added clarity to this area? The Expert Committee position statement from 2009 declared that HbA_{1c} values of 6.0% to 6.4% represented a category that is described as a "*sub-diabetic, high risk state.*"[97] The ADA, in their 2013 position statement, set out that "*it is reasonable to consider an A_{1c} range between 5.7 and 6.4% as identifying individuals with high risk for future diabetes, to whom the term pre-diabetes may be applied.*"[8,97] These HbA_{1c} thresholds are based upon the future risk of developing diabetes, rather than cardiovascular complications relating to diabetes. The choice of threshold HbA_{1c} values are informed by data from the National Health and Nutrition Examination Surveys (NHANES) as well as the Diabetes Prevention Program (DPP), in which study participants with "prediabetes" had HbA_{1c} values between 5.5% and 6.0% (NHANES) or a mean HbA_{1c} of 5.9% (DPP).[8,100] Several prospective studies were systematically reviewed in a large metaanalysis that included a combined sample of 44,203 individuals. The results of this metaanalysis were published since the last meeting of the Expert Committee, and they have been used in resolving the updated HbA_{1c} threshold values for the identification of a so-called "subdiabetic, high-risk state."[8,108] The HbA_{1c} may actually represent a more powerful predictor of type 2 diabetes than fasting glucose. Ultimately, HbA_{1c}, like fasting glucose, is a continuous variable, and its association with future risk for diabetes or macrovascular or microvascular outcomes is curvilinear. Hence, the choice of the appropriate inflection point at which to "draw a line in the sand" is likely to remain an area of controversy and debate.

KEY POINTS

- Diagnosis of diabetes mellitus is based on fasting glucose, glucose measurements following an oral glucose load, and glycated hemoglobin measurements. Diagnostic values are common to most forms of the disease except for gestational diabetes mellitus.

WHAT'S NEW?

Classification and diagnosis of diabetes has become ever more complicated as our understanding of this heterogeneous set of conditions, underpinned by chronic hyperglycemia, increases. Recently, attempts were made to use clustering methods to create a data-driven classification for patients with diabetes, using six parameters that could be measured in a single clinic visit, including: age at diagnosis, body mass index, glutamate decarboxylase antibodies, HbA_{1c}, homeostatic model assessment 2 (HOMA2)-B (to assess β cell function), and HOMA2-insulin resistance (to assess insulin sensitivity).[109] The study, based in Sweden, identified five replicable clusters of patients with diabetes that had significantly different patient characteristics and risk of complications. Although it was hoped that this would represent a significant step towards precision medicine in diabetes, a validation study did not reveal sufficient predictive accuracy or cluster stability to consider implementation in practice.[110] For the moment, therefore, the appropriate diagnosis, classification, and management of diabetes, while guided by various position statements, require clinical acumen. This is even truer considering the gray area that represents so-called *prediabetes,* a categorization applicable principally to individuals at risk for type 2 diabetes. It remains most sensible to address the diagnosis of prediabetes by taking a holistic approach and considering body mass index, lipid profile, and blood pressure, as well as various glycemic thresholds. In light of the global obesity epidemic, perhaps we should be more liberal with our advice (and interventions) regarding healthy lifestyle, rather than waiting for members of the population to cross arbitrary thresholds.

- For your free Expert Consult eBook with bibliographic citations as well as the ability to take notes, highlight important content, search the full text, and more, visit http://www.ExpertConsult.Inkling.com.

REFERENCES

1. Kahn SE. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia*. 2003;46:3–19.
2. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature*. 2001;414:813–820.
3. Nathan DM, Cleary PA, Backlund J-YC, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353:2643–2653.
4. McGarry J. What if Minkowski had been ageusic? An alternative angle on diabetes. *Science*. 1992;258:766–770.
5. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes*. 1979;28:1039–1057.
6. World Health Organization. *Diabetes Mellitus: Report of a WHO Study Group*. Geneva: World Health Organization; 1985.
7. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*. 1997;20:1183–1197.
8. Alberti KGMM, Zimmet PZ. Definition, diagnosis, and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabet Med*. 1998;15:539–553.
9. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2013;36:S67–S74.
10. Imperatore G, Mayer-Davis EJ, Orchard TJ, et al. Prevalence and incidence of type 1 diabetes among children and adults in the United States and comparison with non-U.S. countries. In: National Diabetes Data Group, ed. *Diabetes in America.* 3rd ed. Bethesda: National Institutes of Health;2018:37–46.
11. Menke A, Orchard TJ, Imperatore G, et al. The prevalence of type 1 diabetes in the United States. *Epidemiology*. 2013;24:773–774.
12. Buzzetti R, Quattrocchi C, Nistico L. Dissecting the genetics of type 1 diabetes: relevance for familial clustering and differences in incidence. *Diabetes Metab Rev*. 1998;14:111–128.
13. Dahlquist G. The aetiology of type 1 diabetes: an epidemiological perspective. *Acta Paediatr*. 1998;Suppl 425:5–10.
14. Scott F, Norris J, Hubert K. Milk and type 1 diabetes: examining the evidence and broadening the focus. *Diabetes Care*. 1996;19:379–383.
15. Lund-Blix NA, Dong F, Marild K, et al. Gluten intake and risk of islet autoimmunity and progression to type 1 diabetes in children at increased risk of the disease: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetes Care*. 2019;42:789–796.
16. Robles DT, Eisenbarth GS, Wang T, et al. Identification of children with early onset and high incidence of anti-islet autoantibodies. *Clin Immunol*. 2002;102:217–224.
17. Barker JM. Prediction of autoantibody positivity and progression to type 1 diabetes: diabetes Autoimmunity Study in the Young (DAISY). *J Clin Endocrinol Metab*. 2004;89:3896–3902.
18. Steck AK, Johnson K, Barriga KJ, et al. Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. *Diabetes Care*. 2011;34:1397–1399.
19. Ludvigsson J, Carlsson A, Deli A, et al. Decline of C-peptide during the first year after diagnosis of type 1 diabetes in children and adolescents. *Diabetes Res Clin Pract*. 2013;100:203–209.
20. Zimmet P, Tuomi T, Mackay R, et al. Latent autoimmune diabetes mellitus in adults (LADA): the role of antibodies to glutamic acid decarboxylase in diagnosis and prediction of insulin dependency. *Diabet Med*. 1994;11:299–303.
21. Ziegler AN, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *J Am Med Assoc*. 2013;309:2473–2479.
22. Achenbach P, Warncke K, Reiter J, et al. Stratification of type 1 diabetes risk on the basis of islet autoantibody characteristics. *Diabetes*. 2004;53:384–392.
23. Bingley PJ, Williams AJ, Gale EA. Optimized autoantibody-based risk assessment in family members: implications for future intervention trials. *Diabetes Care*. 1999;22:1796–1801.
24. ClinicalTrials.gov. EXtremely early-onset type 1 diabetes (A Musketeers' Memorandum study) (EXE-T1D). Available at: https://clinicaltrials.gov/ct2/show/NCT03369821. Accessed on: July 2, 2021.
25. Humphrey A, McCarty D, Mackay I, et al. Autoantibodies to glutamic acid decarboxylase and phenotypic features associated with early insulin treatment in individuals with adult-onset diabetes mellitus. *Diabet Med*. 1998;15:113–119.
26. Willis J, Scott R, Brown L, et al. Islet cell antibodies and antibodies against glutamic acid decarboxylase in newly diagnosed adult-onset diabetes mellitus. *Diabetes Res Clin Pract*. 1996;33:89–97.
27. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021. *Diabetes Care*. 2021;44(suppl 1):S15–S33.

28. Maldonado M, Hampe CS, Gaur LK, et al. Ketosis-prone diabetes: dissection of a heterogeneous syndrome using an immunogenetic and beta-cell functional classification, prospective analysis and clinical outcomes. *J Clin Endocrinol Metab*. 2003;88:5090–5098.
29. Umpierrez GE, Kitabchi AE. Ketosis-prone type 2 diabetes mellitus. *Ann Intern Med*. 2006;144:350–357.
30. Centers for Disease Control and Prevention. Type 2 diabetes. Available at: CDC.gov/diabetes/basics/type2.html. Accessed on: July 2, 2021.
31. World Health Organization. Global report on diabetes. Available at: www.who.int/diabetes/global-report/en/ Accessed on: July 2, 2021.
32. Shah P, Vella A, Basu A, et al. Lack of suppression of glucagon contributes to postprandial hyperglycemia in subjects with type 2 diabetes mellitus. *J Clin Endocrinol Metab*. 2000;85:4053–4059.
33. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *J Clin Invest*. 1981;68:1456–1567.
34. D'Adamo E, Caprio S. Type 2 diabetes in youth: epidemiology and pathophysiology. *Diabetes Care*. 2011;34:S161–S165.
35. Rudenski A, Hadden D, Atkinson A, et al. Natural history of pancreatic islet β-cell function in type 2 diabetes mellitus studied over six years by homeostasis model assessment. *Diabet Med*. 1988;5:36–41.
36. Umpierrez G, Korytkowski M. Diabetic emergencies — ketoacidosis, hyperglycaemic hyperosmolar state and hypoglycaemia. *Nat Rev Endocrinol*. 2016;12:222–232.
37. Shapiro H, Pecht T, Shaco-Levy R, et al. Adipose tissue foam cells are present in human obesity. *J Clin Endocrinol*. 2013;98:1173–1181.
38. Barnett A, Eff C, Leslie R, et al. Diabetes in identical twins: a study of 200 pairs. *Diabetologia*. 1981;20:87–93.
39. Knowler W, Saad M, Pettitt D, et al. Determinants of diabetes mellitus in the Pima Indians. *Diabetes Care*. 1993;16:216–227.
40. Hoet J, Tripathy B, Rao R, et al. Malnutrition and diabetes in the tropics. *Diabetes Care*. 1996;19:1014–1017.
41. Tripathy B, Samal K. Overview and consensus statement on diabetes in tropical areas. *Diabetes Metab Rev*. 1997;13:63–76.
42. Fajans SS, Bell GI, Polonsky KS. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med*. 2001;345:971–980.
43. Aarthy R, Aston-Mourney K, Mikocka-Walus, et al. Clinical features, complications and treatment of rarer forms of maturity-onset diabetes of the young (MODY)—a review. *J Diabetes Complications*. 2021;35:107640.
44. Byrne M, Sturis J, Menzel S, et al. Altered insulin secretory response to glucose in diabetic and nondiabetic subjects with mutations in the diabetes susceptibility gene MODY 3 on chromosome 20. *Diabetes*. 1996;45:1503–1510.
45. Clement K, Pueyo M, Vaxillaire M, et al. Assessment of insulin sensitivity in glucokinase-deficient subjects. *Diabetologia*. 1996;39:82–90.
46. Habener J, Stoffers D. A newly discovered role of transcription factors involved in pancreas development and the pathogenesis of diabetes mellitus. *Proc Assoc Am Physicians*. 1998;110:12–21.
47. Velho G, Froguel P. Maturity-onset diabetes of the young (MODY), MODY genes and non-insulin-dependent diabetes mellitus. *Diabetes Metab*. 1997;23:34–37.
48. Shields BM, Hicks S, Shepherd MH, et al. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia*. 2010;12:2504–2508.
49. Sperling MA. Neonatal diabetes mellitus: from understudy to center stage. *Curr Opin Pediatr*. 2005;17:512–518.
50. Du X, Ounissi-Benkalha H, Loder MK, et al. Overexpression of ZAC impairs glucose-stimulated insulin translation and secretion in clonal pancreatic beta-cells. *Diabetes Metab Res Rev*. 2012;28:645–653.
51. Sperling MA. ATP-sensitive potassium channels—neonatal diabetes and beyond. *N Engl J Med*. 2006;355:507–510.
52. Walker M, Turnbull D. Mitochondrial related diabetes: a clinical perspective. *Diabet Med*. 1997;14:1007–1009.
53. O'Rahilly S, Gray H, Humphreys P, et al. Impaired processing of prohormones associated with abnormalities of glucose homeostasis and adrenal function. *N Engl J Med*. 1995;333:1386–1390.
54. Haneda M, Polonsky K, Bergenstal R, et al. Familial hyperinsulinemia due to a structurally abnormal insulin: definition of an emerging new clinical syndrome. *N Engl J Med*. 1984;310:1288–1294.
55. Semple RK, Savage DB, Cochran EK, et al. Genetic syndromes of severe insulin resistance. *Endocr Rev*. 2011;32:498–514.
56. Rabson SM, Mendenhall EN. Familial hypertrophy of pineal body, hyperplasia of adrenal cortex and diabetes mellitus; report of 3 cases. *Am J Clin Pathol*. 1956;26:283–290.
57. Donohue WL, Uchida I. Leprechaunism: a euphemism for a rare familial disorder. *J Pediatr*. 1954;45:505–519.
58. Barbieri RL, Ryan KJ. Hyperandrogenism, insulin resistance, and acanthosis nigricans syndrome: a common endocrinopathy with distinct pathophysiologic features. *Am J Obstet Gynecol*. 1983;147:90–101.
59. Kahn CR, Flier JS, Bar RS, et al. The syndromes of insulin resistance and acanthosis nigricans — insulin-receptor disorders in man. *N Engl J Med*. 1976;294:739–745.
60. Garg A. Acquired and inherited lipodystrophies. *N Engl J Med*. 2004;350:1220–1234.
61. Gullo L, Pezzilli R, Morselli-Labate A, et al. Diabetes and the risk of pancreatic cancer. *N Engl J Med*. 1994;331:81–84.
62. Moran A, Doherty L, Wang X, et al. Abnormal glucose metabolism in cystic fibrosis. *Pediatrics*. 1998;133:10–17.
63. Kelly A, Moran A. Update on cystic fibrosis. *J Cyst Fibros*. 2013;12:318–331.
64. Phelps G, Chapman I, Hall P, et al. Prevalence of genetic haemochromatosis among diabetic patients. *Lancet*. 1989;2:233–234.
65. Yajnik C, Shelgikar K, Naik S, et al. The ketoacidosis-resistance in fibrocalculous-pancreatic-diabetes. *Diabetes Res Clin Pract*. 1993;15:149–156.
66. Mohan V, Barman KK, Rajan VS, et al. Natural history of endocrine failure in tropical chronic pancreatitis: a longitudinal follow-up study. *J Gastroenterol Hepatol*. 2005;20:1927–1934.
67. Hassan Z, Mohan V, Ali L et al. SPINK1 is a susceptibility gene for fibrocalculous pancreatic diabetes in subjects from the Indian subcontinent. *Am J Hum Genet*. 2002;71:964–968.
68. Liu J, Zhou H, Zhang Y, et al. Reporting of immune checkpoint inhibitor therapy-associated diabetes, 2015-2019. *Diabetes Care*. 2020;43:e79–e80.
69. Barroso-Sousa R, Barry WT, Garrido-Castro AC, et al. Incidence of endocrine dysfunction following the use of different immune checkpoint inhibitor regimens: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4:173–182.
70. Byun DJ, Braunstein R, Flynn J, et al. Immune checkpoint inhibitor-associated diabetes: a single-institution experience. *Diabetes Care*. 2020;43:3106–3109.
71. Stears AJ, Woods SH, Watts MM, et al. A double-blind, placebo-controlled, crossover trial comparing the effects of amiloride and hydrochlorothiazide on glucose tolerance in patients with essential hypertension. *Hypertension*. 2012;59:934–942.
72. Neary NM, Booker OJ, Abel BS, et al. Hypercortisolism is associated with increased coronary arterial atherosclerosis: analysis of noninvasive coronary angiography using multidetector computerized tomography. *J Clin Endocrinol Metab*. 2013;98:2045–2052.
73. Margolis AM, Heverling H, Pham PA, et al. A review of the toxicity of HIV medications. *J Med Toxicol*. 2014;10:26–39.
74. American Diabetes Association. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care*. 2004;27:596–601.
75. Holt RIG. Association between antipsychotic medication use and diabetes. *Curr Diab Rep*. 2019;19:96.
76. Colao A, Petersenn S, Newell-Price J, et al. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med*. 2012;366:914–924.
77. Forrest J, Menser M, Burgess JA. High frequency of diabetes mellitus in young patients with congenital rubella. *Lancet*. 1971;2:332–334.
78. King M, Bidwell D, Shikh A, et al. Coxsackie-B-virus-specific IgM responses in children with insulin-dependent (juvenile-onset; type 1) diabetes mellitus. *Lancet*. 1983;1:1397–1399.

79. Gregory JM, Slaughter JC, Duffus SH, et al. COVID-19 severity is tripled in the diabetes community: a prospective analysis of the pandemic's impact in type 1 and type 2 diabetes. *Diabetes Care*. 2021;44:526–532.
80. Steenblock C, Richter S, Berger I, et al. Viral infiltration of pancreatic islets in patients with COVID-19. *Nat Commun*. 2021;12:3534.
81. Solimena M, De Camilli P. Autoimmunity to glutamic acid decarboxylase (GAD) in stiff-man syndrome and insulin-dependent diabetes mellitus. *Trends Neurosci*. 1991;14:452–457.
82. Arioglu E, Andewelt A, Diabo C, et al. Clinical course of the syndrome of autoantibodies to the insulin receptor (Type B insulin resistance): a 28-year perspective. *Medicine*. 2002;8:87–100.
83. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet*. 2007;39:951–953.
84. ACOG practice Bulletin No. 190: Gestational Diabetes. *Obstet Gynecol.* 2018;131:e49–e64.
85. Persson B, Hanson U. Neonatal morbidities in gestational diabetes mellitus. *Diabetes Care*. 1998;21:B79–B84.
86. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*. 2010;33:676–682.
87. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. 2008;358:1991–2002.
88. Hod M, Kapur A, Sacks DA, et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on gestational diabetes mellitus: a pragmatic guide for diagnosis, management, and care. *Int J Gynaecol Obstet*. 2015;131:S173–S211.
89. World Health Organization. *Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy*. London: WHO/NMH/MND/13.2; 2013. Available at: https://apps.who.int/iris/bitstream/handle/10665/85975/WHO_NMH_MND_13.2_eng.pdf. Accessed on: July 2, 2021.
90. National Institute for Health and Care Excellence (NICE). *Diabetes in Pregnancy: Management from Preconception to the Postnatal Period*. Clinical guideline NG3. Paris: WHO Press; 2015. Available at: www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-ofdiabetes-and-itscomplications-from-preconceptionto-the-postnatal-period-51038446021. Accessed on: July 2, 2021.
91. Lawrence JM, Contreras R, Chen W, et al. Trends in the prevalence of preexisting diabetes and gestational diabetes mellitus among a racially/ethnically diverse population of pregnant women, 1999–2005. *Diabetes Care*. 2008;31:899–904.
92. Sacks DA, Hadden DR, Maresh M, et al. Frequency of gestational diabetes mellitus at collaborating centers based on IADPSG consensus panel-recommended criteria: the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. *Diabetes Care*. 2012;35:526–528.
93. Hillier TA, Pedula KL, Ogasawara KK, et al. A pragmatic, randomized clinical trial of gestational diabetes screening. *N Engl J Med*. 2021;384:895–904.
94. Bilous RW, Jacklin PB, Maresh MJ, et al. Resolving the gestational diabetes diagnosis conundrum: the need for a randomized controlled trial of treatment. *Diabetes Care*. 2021;44:858–864.
95. Simmons D, Moses RG. Gestational diabetes mellitus: to screen or not to screen?: is this really still a question? *Diabetes Care*. 2013;36:2877–2878.
96. The expert committee on the diagnosis and classification of diabetes mellitus. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*. 2003;26:3160–3167.
97. The International Expert Committee. International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32:1327–1334.
98. Gabir MM, Hanson RL, Dabelea D, et al. Plasma glucose and prediction of microvascular disease and mortality. *Diabetes Care*. 2000;23:1113–1118.
99. Wong TY, Liew G, Tapp RJ, et al. Relation between fasting glucose and retinopathy for diagnosis of diabetes: three population-based cross-sectional studies. *Lancet*. 2008;371:736–743.
100. Diabetes Prevention Program Research Group. The prevalence of retinopathy in impaired glucose tolerance and recent-onset diabetes in the Diabetes Prevention Program. *Diabet Med*. 2007;24:137–144.
101. Abdul-Ghani MA, Triphany D, DeFronzo RA. Contributions of β-cell dysfunction and insulin resistance to the pathogenesis of impaired glucose tolerance and impaired fasting glucose. *Diabetes Care*. 2006;29:1130–1139.
102. The DECODE Study Group, the European Epidemiology Study Group. Glucose tolerance and cardiovascular mortality: comparison of fasting and 2-hour diagnostic criteria. *Arch Intern Med*. 2001;161:397–404.
103. Meigs JB, Muller DC, Nathan DM, et al. The natural history of progression from normal glucose tolerance to type 2 diabetes in the Baltimore Longitudinal Study of Aging. *Diabetes*. 2003;52:1475–1484.
104. Dinneen S, Maldonado D, Leibson C, et al. Effects of changing diagnostic criteria on the risk of developing diabetes. *Diabetes Care*. 1998;21:1408–1413.
105. Nichols GA, Hillier TA, Brown JB. Progression from newly acquired impaired fasting glucose to type 2 diabetes. *Diabetes Care*. 2007;30:228–233.
106. Tirosh A, Shai I, Tekes-Manova D, et al. Normal fasting plasma glucose levels and type 2 diabetes in young men. *N Engl J Med*. 2005;353:1454–1462.
107. Khaw KT, Wareham N, Bingham S, et al. Association of hemoglobin A1c with cardiovascular disease and mortality in adults: the European prospective investigation into cancer in Norfolk. *Ann Intern Med*. 2004;141:413–420.
108. Zhang X, Gregg EW, Williamson DF, et al. A1c level and future risk of diabetes: a systematic review. *Diabetes Care*. 2010;33:1665–1673.
109. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6:361–369.
110. Lugner M, Gudbjörnsdottir S, Sattar N, et al. *Comparison between Data-Driven Clusters and Models Based on Clinical Features to Predict Outcomes in Type 2 Diabetes: Nationwide Observational Study*. Diabetologia; 2021;64:1973–1981.

32

Autoimmune (Type 1) Diabetes Mellitus: Etiology, Pathogenesis, Prediction, and Prevention

Markus Lundgren and Åke Lernmark

OUTLINE

Definition, 537
Epidemiology, 537
  Prevalence, 537
  Incidence, 538
  Geographical Distribution, 538
  Variation with Age, 539
  Variation With Gender, 540
  Seasonal Variation, 540
Inheritance, 541
Natural History and Disease Staging, 541
Etiology, 541
  Genetic Etiology, 541
  Environmental Etiology, 543
Pathogenesis, 546
  Humoral Immunopathophysiology, 546
  Cellular Immunopathophysiology, 548
Diagnostic Criteria, 549
Prediction, 550
Prevention, 550
Primary Prevention, 550
Secondary Prevention, 550
Intervention, 551
Summary and Future Directions, 551

DEFINITION

Autoimmune type 1 (insulin-dependent) diabetes (T1D) is associated with several immune abnormalities. This form, previously called "juvenile-onset diabetes" or "insulin-dependent diabetes," accounts for 5% to 10% of diabetes patients. T1D results from a selective destruction of the pancreatic β cells, eventually leading to an absolute lack of insulin and hyperglycemia. T1D is frequently recognized because of symptoms of polyuria, polydipsia, weight loss, and fatigue. These symptoms develop due to the failing β cell function, gradually worsening over time. Diabetes as such is diagnosed based on plasma glucose criteria, by fasting plasma glucose value, random plasma glucose, or the 2-hour plasma glucose value during a 75-g oral glucose tolerance test (OGTT)[1] concurrent with diabetes symptoms. In the absence of unequivocal hyperglycemia or diabetes symptoms, the elevated glucose values need to be repeated for a diagnosis to be made. Typically, when age at onset of diabetes is young, classification of the disease is uncomplicated.[1] T1D may occur globally and at any age.[2] The fact that T1D in adults may fulfill clinical criteria of type 2 diabetes (T2D) demonstrates the limitation of a disease classification based on clinical symptoms rather than on etiology and pathogenesis. Pancreatic islet autoantibodies that both predict and differentiate T1D from other forms of diabetes are included among the diagnostic criteria for T1D.[1]

The T1D autoimmune biomarkers include GAD65 autoantibody (GADA), insulin autoantibody (IAA), islet antigen-2 autoantibody (IA-2A and IA-2βA), and zinc transporter 8 autoantibody (ZnT8A).[3,4] These autoantibodies are either endpoints in ongoing primary prevention studies[5] or inclusion criteria in secondary prevention studies aimed at preventing the onset of diabetes.[6] Autoantibody biomarkers are additionally used to stage T1D (Fig. 32.1). Stage 1 diabetes is defined by the presence of two or more autoantibodies but normoglycemia and an absence of symptoms. Stage 2 is characterized by two or more biomarkers with dysglycemia but no symptoms. Stage 3 is diagnosed diabetes. Staging T1D should prove useful to improve the design of both primary and secondary prevention trials.

Classification of diabetes has become complicated during the current obesity epidemic among the young, making differentiation between T1D and T2D more difficult.[7,8] It has been known for decades that diabetes mellitus can be diagnosed at different degrees of severity, and that the diagnostic criteria used to differentiate between T1D (insulin-dependent) and T2D (non–insulin-dependent) are important for understanding the etiology and pathogenesis of these two disease entities. The etiology of T1D has not been clarified, while the pathogenesis is somewhat better understood (Fig. 32.1). Using molecular genetics, it has also been possible to define a number of monogenic forms of diabetes[9,10] involving mutations in the genes for insulin, the insulin receptor, glucokinase (maturity-onset diabetes in the young; MODY-2), HNF-4α (MODY-1), and HNF-1a (MODY-3).[9] The clarification of the etiology of these diabetes phenotypes is a major advancement that has increased the diagnostic precision of diabetes syndromes. Our understanding of the more complex and multifactorial T1D syndrome has also seen significant advances.

EPIDEMIOLOGY

Prevalence

The prevalence of T1D is low compared with that of T2D. Among individuals aged 30 years or younger, the prevalence of T1D does not usually exceed approximately 0.3%, compared with prevalence rates for T2D of 4.2% worldwide and nearly 25% in certain high-risk

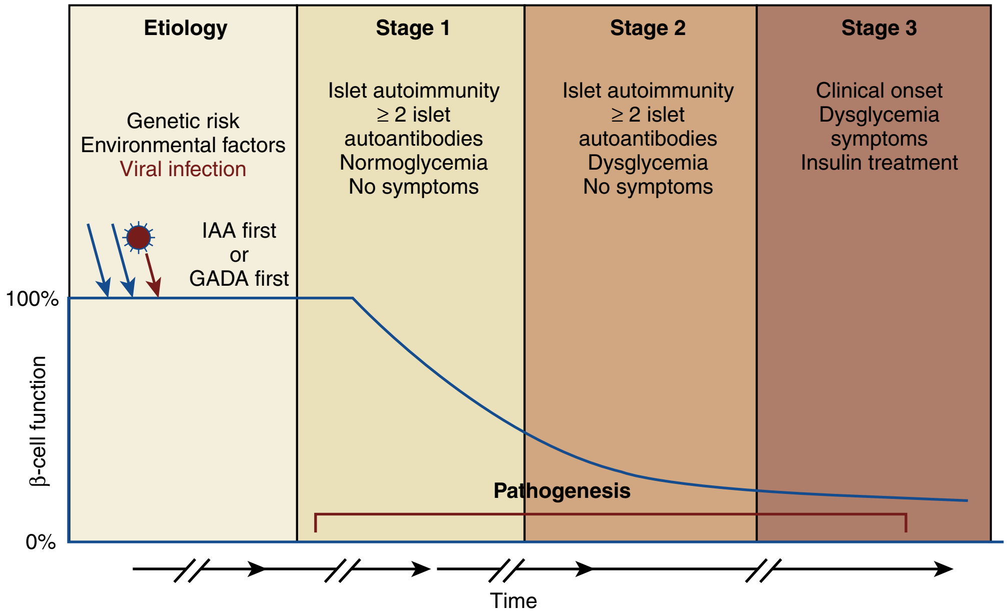


Fig. 32.1 Etiology and pathogenesis of autoimmune type 1 diabetes (T1D). The etiology is genetic and environmental, in yet unknown combinations. Biomarkers of two different endotypes are autoantibodies appearing either against insulin autoantibody (IAA) or glutamic acid decarboxylase autoantibody (GADA) first. The pathogenesis is initiated by islet autoimmunity to two or more autoantibodies (IAA, GADA, or autoantibodies against islet antigen-2, or Zn transporter 8) but normoglycemia and no symptoms (stage 1). Two or more autoantibodies remain, but dysglycemia develops (stage 2). The clinical onset is defined by dysglycemia (hyperglycemia), diabetes symptoms, and insulin requirement (stage 3). The loss of functional β cell mass is illustrated as a continuous variable, but the rate of loss varies between individuals dependent on the age at the first appearing islet autoantibody in combination with genetic and environmental (e.g., additional viral infections) factors. Dependent on the endotype of first appearing autoantibody, time represent months to years.

populations.[11] While current attention is directed toward the trends of the "epidemic" or "tsunami" of T2D, recent data suggest that there is a parallel rise in the prevalence of both T1D and T2D.[11] Both geographic and ethnic variations are seen in the prevalence rates. The International Diabetes Federation estimated that in 2015 there were 415 million (uncertainty interval: 340–536 million) people with diabetes aged 20 to 79 years.[11] It has been pointed out that the results of prevalence studies should be viewed with caution. Different age groups were studied, and, although geographical variation may be related to variations in genetic predisposition, environmental risk factors, or both, it may also be due to some research methodology problems.[11] In addition, it is unclear whether recently observed increases in incidence rates worldwide and in specific regions reflect a change in the age at onset of diabetes or a true increase in prevalence of T1D.[2,8,12]

Incidence

The incidence rate is the frequency with which new patients with T1D are detected during a defined period. The rate is expressed as an annual number of patients per 100,000 age-corrected individuals. A determination of incidence rate therefore requires a precise knowledge of the total number of individuals in each age group and the number of new patients diagnosed during one year in the particular area. Determinations of secular trends constitute an important part of population-based epidemiologic studies. However, such analyses are rare because they require careful follow-up investigations for several subsequent years.

Globally, the current rise in incidence rates by 3% (2%–5%) is expected to be 40% higher in 2020 compared with figures from 1998[12,13] (Fig. 32.2). Currently it is estimated that more than 70,000 new patients are diagnosed each year worldwide.[12-14] There are suggestions that the increasing incidence rates and changing trends are seen not only in populations with high incidence but also in populations with previously lower incidence rates.[12,14] Over the last three decades, T1D has shown temporal trends in most of the regions of the world except Central America and the West Indies, while European data showed that the main recent rise in incidence rates is in Central-Eastern countries.[2,11] EURODIAB[15] and DiaMond[16] have attempted to better define the incidence rates in various populations by using prospective, population-based, geographically defined registries. Overall, the average annual incidence rate varies widely, from 0.1 per 100,000 children in parts of Asia and South America to 64.2 per 100,000 in Finland.[11,16] Pooled data from all sites in the EURODIAB study demonstrated an annual rate of increase in incidence of 3.4%, with more rapid rates of increase occurring in certain regions.[15] Approximately 1% of all children born in Scandinavian countries will manifest T1D during their lifetimes. The causes of the rapid rise in incidence are not understood, and further studies are needed to define to what extent the rise may be due to environmental factors, genetic admixture, or both.

Geographical Distribution

T1D can occur in any region of the world, and recently it was observed that the incidence is increasing in countries with both high and low incidence rates; in the latter group, the rise is steeper.[11] The annual incidence of T1D is higher in northern Europe than in the Mediterranean area, with the exception of Sardinia (Fig. 32.2). Furthermore, the incidence in Iceland is lower than that in Sweden or Finland.[17] Surprisingly, the T1D incidence rate in Estonia is approximately 25% of the rate in Finland, in spite of their close geographic proximity.[17,18] The incidence

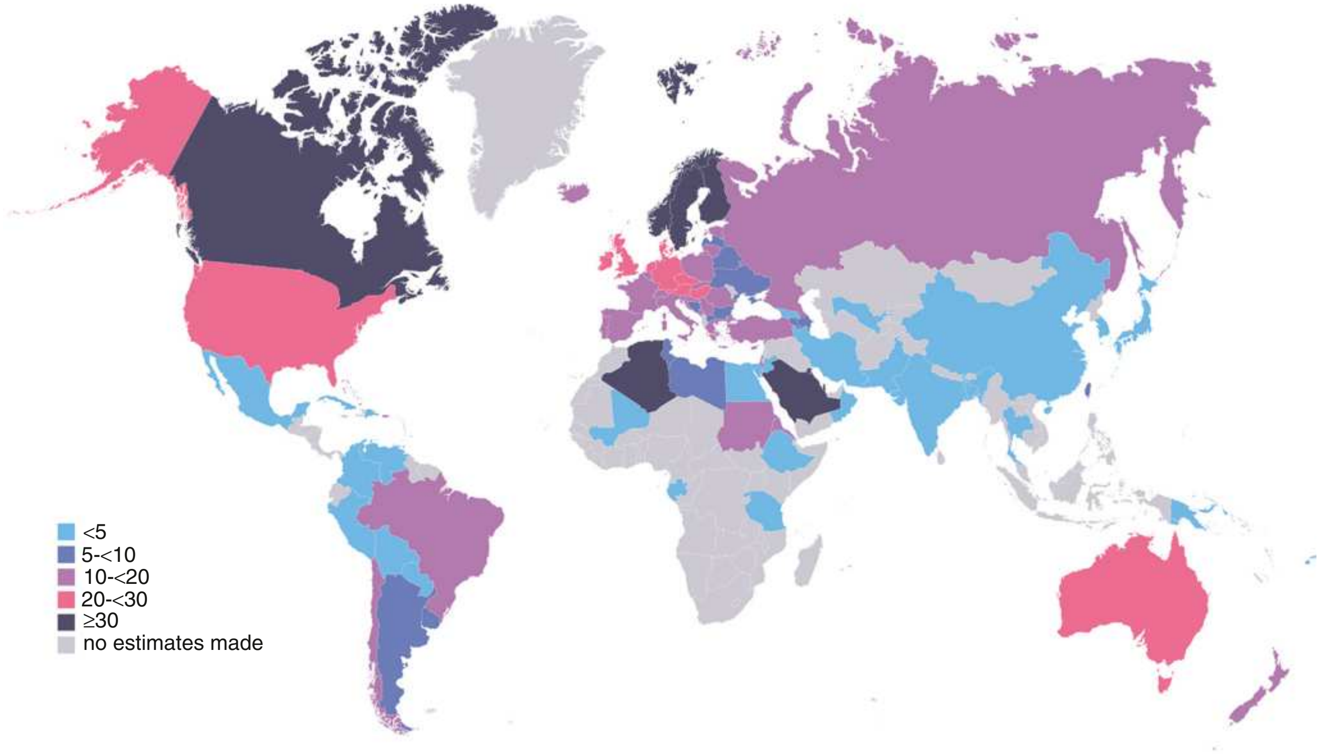


Fig. 32.2 Incidence rates standardized for age and sex per 100,000 population per year of type 1 diabetes in children and adolescents aged 0–14 years. (From International Diabetes Federation. IDF Diabetes Atlas, 10th edn. Brussels, Belgium: International Diabetes Federation, 2021. http://www.diabetesatlas.org.)

rates within countries also show stable differences. The eastern and southern parts of Finland,[19] as well as the central and southern parts of Sweden,[17] have had higher incidence rates than the northern parts of those countries.

The cause of geographic variation remains unknown, but it has been speculated to be related to genetic factors primarily associated with different human leukocyte antigen (HLA) DR or DQ genotypes of the major histocompatibility complex (MHC) on chromosome 6. HLA class II molecules from the DQ and DR loci that are necessary (but not sufficient) for disease vary greatly between countries, and therefore affect disease incidence.[20] In addition, environmental factors are important in understanding both the etiology and the pathogenesis of T1D and may also help to explain differences in geographic distribution. This is supported by the observation that monozygotic twins show less than 20% to 30% concordance rates for T1D.[21]

In the *European region,* incidence rates show correlation with the frequency of HLA susceptibility genes in the general population. In this region, there is also wide variation in incidence rates, ranging from 6 to 62 cases/100,000/year among children 0 to 14 years of age, with higher rates observed in Scandinavian countries, especially Finland and Sweden. In countries with populations of European origins, there is less variation: *Canada* reports 21/100,000/year, the *United States* 16/100,000/year, and *New Zealand and Australia* have incidence rates ranging from 16 to 20 cases/100,000/year. In the *Middle East* and *North Africa,* few reliable data are available. The incidence rates range from 1/100,000/year in Pakistan to 8 to 12/100,000/year in Egypt, Sudan, and Libya; however, reports from Kuwait showed a rise in incidence rates from incidence rose from 17.7 in 1992 to 1994 to 40.9 per 100,000 per year in 2011 to 2013.[22,23] There are few published data from the *Southeast Asian region,* and data from the *African region* are mostly lacking. In *South* and *Central America,* the incidence rates are low, except for Argentina and Uruguay, with 6.8 cases/100,000/year and 8.3 cases/100,000/year, respectively. The *Western Pacific* countries have low incidence rates, especially China, which has one of the lowest incidence rates worldwide, ranging from 0.1 to 0.6 patients/100,000/year (Fig. 32.2).[11,23]

It has been described that migration of people from relatively low-incidence countries into countries with a “diabetogenic” environment (i.e., with high incidence rates) increases the risk among the migrant groups. Migrants, within a short period of time, assume comparative risks of the native population, and several studies confirm the importance of country of origin in the risk of T1D.[24] Improved epidemiology and better diagnostic criteria to distinguish different forms of diabetes are critical in obtaining reliable incidence rates for various countries and states.

Variation with Age

Epidemiologic studies with rigorous diagnostic criteria suggest that the clinical onset of T1D may occur at any age. The incidence rate varies with both age and gender (Fig. 32.3). The peak for both girls and boys, at age 11 to 14 years, has been discernible in most studies and seems to be present irrespective of the country or area studied. This peak is associated with puberty and the maximal velocity of pubertal growth, and it may be associated with reduced insulin sensitivity due to hormonal factors related to growth.

Previous studies have suggested that the annual rate of increase in incidence is higher in younger age groups, with rates of 6.4% in 0- to 4-year-olds, 3.1% in 5- to 9-year-olds, and 2.4% in 10- to 14-year-olds.[15]

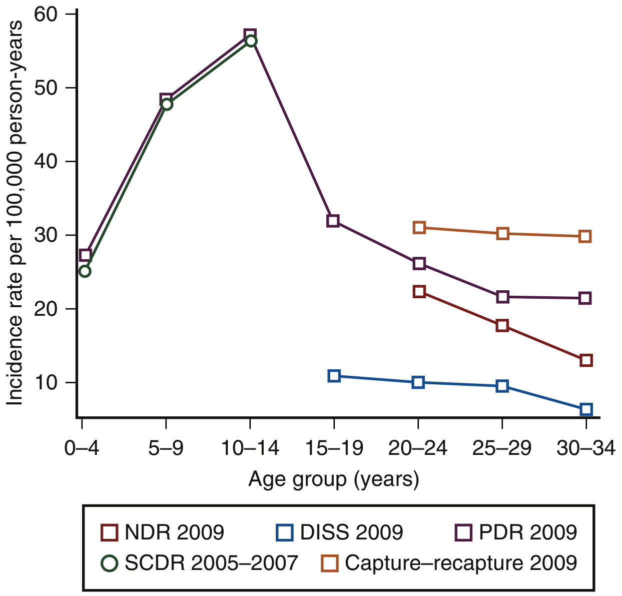


Fig. 32.3 Age-adjusted incidence rates per 100,000 person-years of type 1 diabetes by age group and register. *NDR*, National Diabetes Registry; *DISS*, Diabetes Incidence Study in Sweden; *PDR*, Prescribed Drug Register; *SCDR*, Swedish Childhood Diabetes Register. (Data from Rawshani A, Landin-Olsson M, Svensson AM, et al. The incidence of diabetes among 0-34 year olds in Sweden: new data and better methods. *Diabetologia*. 2014;57:1375–1381. Used with permission.)

In children, minor incidence peaks also occur at 4 to 6 years and 7 to 8 years, which have been associated with preschool or school programs.[15] Recent data suggest that around 50% to 60% of cases of T1D in Western countries occur before 15 years of age, and that more than 90% of childhood and adolescent diabetes is T1D.[12,25] Epidemiologic investigation has indicated that the incidence rate of T1D in patients older than 20 years is lower than that seen in children, except for a possible peak at around 50 to 65 years. In addition to adults having what is considered a more classic T1D clinical characteristics, some adult patients initially classified and treated as having T2D may require insulin after 1 to 5 years of therapy with diet, exercise, or oral hypoglycemic agents. This type of diabetes is referred to as *type 1.5 diabetes, latent autoimmune diabetes in the adult* (LADA), or *slowly progressive insulin-dependent diabetes mellitus.*[26-28] These patients are positive for islet autoantibodies and have lower body mass index and higher frequencies of high-risk HLA genotypes compared with other “type 2” patients. It is speculated that these patients in fact have T1D in addition to genetic factors that may inhibit rapid progression of β cell killing.[29]

Although rare, some forms of diabetes mellitus can occur at birth or during the first months of life. *Neonatal diabetes* (NDM) is diagnosed during the first months of life (up to 6 months) at a rate of 1 in 500,000 live births. It is characterized by insulin-sensitive hyperglycemia. Approximately half of patients with NDM are transient (TNDM), but the remaining progress into a permanent (PNDM) type, with lifelong dependence on exogenous insulin. Clinically the infant presents with polydipsia, polyuria, hyperglycemia, and/or glycosuria, but the condition may progress into ketoacidosis. PNDM usually proceeds after the second year, and differentiation between the two forms can only be done through genetic testing. TNDM occurs at birth up to 3 to 6 months and is correlated with defects in specific genes: ZAC pleomorphic adenoma gene-like 1 (also known as *PLAG1*)/hydatidiform mole-associated and imprinted transcript (*ZAC/HYMAI*), *KCNJ11*, *Kir6.2* gene, hepatocyte nuclear factor 1β (*HNF1β*), or sulfonylurea receptor 1 (*SUR1*).[30,31] PNDM, on the other hand, is related to several gene defects, such as activating mutations in the *KCNJ11/Kir6.2* gene that encodes for the ATP potassium channel (K_{ATP}), mutations in the *ABCC8* gene that encodes SUR1, complete deficiency of the glucokinase *IPF1/PDX1* gene that encodes insulin promoter factor 1, mutations in the *PTF1A* gene that encodes pancreas transcription factor 1A, mutations in the *EIF2AK3* gene encoding protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK),[32] and mutations in the forehead box p3 (*FOXP3*) gene (also known as the *T-cell regulatory gene*).[33]

Variation With Gender

It has been reported that the peak incidence in girls occurs earlier than in boys. If the clinical onset of T1D is linked to pubertal growth, this difference in incidence rate may be explained by the fact that pubertal growth occurs earlier in girls. In children followed from birth due to T1D genetic risk, the male-to-female ratio was 1.4:1.0 before 6 years of age at T1D diagnosis.[34] This ratio was increased (1.7:1.0) among children younger than 18 months of age. Prepubertal boys were found to be taller at the clinical onset of T1D.[35] In addition, newly diagnosed children of both genders showed advanced skeletal maturity. Even if boys tend to show an increased height compared with controls, their growth seems to cease approximately 35 weeks before the clinical onset of T1D.[35] Therefore, it is possible that processes affecting the pancreatic β cell mass and the ability to produce insulin may have profound effects on body growth and function at a young age. Because these processes differ slightly between boys and girls, growth characteristics may offer a simple explanation for the differences seen in incidence rates between the genders.

A study that examined sex differences in T1D incidence before the age of 15 years showed a slight male preponderance in many, but not all, European countries, whereas a female preponderance was found in most African and Asian countries.[36] Interestingly, the male excess was seen in all countries with the highest incidence rates ($>20/10^5$/yr), whereas a female excess was seen in all countries with low incidence rates ($<4.5/10^5$/yr). Studies of older age groups have consistently shown a male preponderance for new patients with T1D, with male/female ratios ranging from 1.3:1 to 2.5:1. One registry of T1D among 15- to 34-year-olds (Fig. 32.3) demonstrated that T1D was 1.5 times more common among men than women.[4,37] Further studies are necessary to document gender-dependent incidence rates and to explain their mechanisms.

Seasonal Variation

T1D shows cyclic, sinusoidal appearance of seasonal variation in date of diagnosis, with a peak during the winter months. This variation is proposed to be related to the timing of occurrences of certain precipitating factors such as viral infection and cold climate.[38] It is of interest that islet cell autoantibodies (ICAs) appearing during the prediabetic state follow similar seasonal variations, being more prevalent in colder months and rare in summer and spring. Additionally, the triggering of β cell autoimmunity appears to have variable trends over subsequent years and does not equally affect genetically susceptible siblings.[24] These observations suggest that triggering of immunologic markers may be related to factors (most likely viruses) that prevail mostly in colder months with variable frequency of occurrence.

KEY POINTS

- The global rise in incidence rates by 3% and is expected to be 40% higher in 2020 compared with 1998. Currently it is estimated that more than 70,000 new patients are diagnosed each year worldwide.

INHERITANCE

The mode of inheritance is complex, and around 80% to 85% of cases of T1D occur sporadically without familial aggregation. Among HLA-identical siblings of T1D-affected patients, approximately 20% eventually manifest the disease. The overall lifetime risk for first-degree relatives has been estimated at approximately 8% for siblings—15 times higher than the general population—and 5% for children of parents with T1D. The risk of T1D among offspring ranges from 2% to 4% if the mother was affected, compared with 6% to 9% if the father was affected, and the risk increases to 30% when both parents are affected.[39,40]

NATURAL HISTORY AND DISEASE STAGING

Early histologic studies of pancreatic tissue of diabetic patients who died shortly after clinical onset revealed that the pancreatic islets were altered by fibrosis, hyalinosis, atrophy, and infiltration of inflammatory cells.[41] The inflammatory lesion of the islets of Langerhans was described as *insulitis,* and quantitative studies of the pancreatic islets showed a specific loss of insulin-producing cells in association with clinical onset of T1D.[41] It was therefore suggested that the pathogenesis of T1D involved autoimmune reactions directed toward the endocrine pancreas. This notion was supported by leukocyte migration inhibition to pancreatic islet antigens.[42] Numerous studies have confirmed the presence of insulitis[43-45]; however, in spite of the assumption that T1D is a T-cell–mediated disease, reproducible and standardized tests of blood T-cell reactivity against islet autoantigens are yet to be established.

Long-sought-for ICAs were described in 1974,[46,47] islet cell surface antibodies (ICSAs) in 1978,[48] and complement-dependent antibody-mediated islet cell cytotoxicity in 1980.[49] The first autoantigen recognized by islet antibodies, an islet protein of 64,000 relative molecular mass (*Mr*), or 64K, was described in 1982.[50] The 64K protein was found to have glutamic acid decarboxylase (GAD) activity,[51] but molecular cloning showed that the human islet GAD was a novel isoform, GAD65.[52] Autoantibodies to insulin were demonstrated in 1983,[53] to islet antigen-2 (IA-2), a tyrosine phosphatase-like protein, in 1994,[54] and to the zinc transporter 8 (ZnT8) in 2007.[55] In contrast to T-cell analyses, GADAs and IA-2As may be determined in standardized assays.[56] Taken together, ample evidence exists that islet cell autoimmunity is of major importance in the etiopathogenesis of T1D. However, these autoantibodies caused a paradigm shift that rewrote the natural history of the disease. It was possible for the first time to ask the when the first of these T1D biomarkers appeared.

Because T1D is detected in the majority of patients after a relatively short period of symptoms, such as excessive thirst, polyuria, and unexplained weight loss, the natural history of the disease has been poorly defined until recently. With the current ability to define individuals at high risk for T1D based on high-risk HLA and autoantibody positivity, our understanding of the prediabetic period is improving. Prospective studies from birth of children with increased risk for T1D (TEDDY, DIPP, DAISY, DiPiS, BABYDIAB, PANDA) have revealed that autoantibodies to GAD65, IA-2, or insulin and decreased ability to release insulin in response to glucose may develop several years before the clinical diagnosis. The sequence of events preceding the diagnosis of overt T1D would include the following: (1) genetic predisposition; (2) overt immunologic abnormality reflected in a first appearing islet autoantibody but with normal glucose levels and no symptoms; (3) development of a second or third islet autoantibody still with normal blood glucose and no symptoms; (4) development of β cell dysfunction and dysglycemia but no symptoms; (5) development of overt hyperglycemia with detectable C-peptide; and (6) the final stage of insulin dependency, with loss of detectable C-peptide. The fact that T1D develops in persons of all ages must be taken into account when studying the natural history in children and adults. Based on the natural history and using islet autoantibodies as biomarkers for islet autoimmunity, type 1 diabetes has been staged as follows: *stage 1*, two or more islet autoantibodies with normoglycemia; *stage 2*, two or more islet autoantibodies with dysglycemia; *stage 3*, diagnosis of diabetes[6] (Fig. 32.1).

KEY POINTS

- Using islet autoantibodies as biomarkers for islet autoimmunity, type 1 diabetes has been classified into the following stages: *stage 1,* two or more islet autoantibodies with normoglycemia; *stage 2,* two or more islet autoantibodies with dysglycemia; *stage 3,* diagnosis of diabetes.

ETIOLOGY

The absence of an unambiguous mode of inheritance, the presence of a period of subclinical islet autoimmunity preceding clinical onset of disease, HLA genes that control the immune response, and age and seasonal variation must be taken into account in attempts to explain the cause of T1D. A defined etiologic factor, endogenous or exogenous, capable of causing T1D remains to be identified. Because evidence exists of genetic heterogeneity in T1D, it is possible that different causative factors are responsible. The approach to dissecting the etiology of T1D has been to identify children at risk at birth and follow them prospectively until the appearance of a first islet autoantibody and subsequent follow-up for a second or third autoantibody until the clinical diagnosis of diabetes. In a first attempt, children born to mothers or fathers with T1D were followed from 9 months of age in BABYDIAB.[57] Subsequently, children born with high-risk HLAs for T1D were followed from 3 months of age in DIPP[58,59] and DAISY.[60,61] While IAA showed an early appearance in BABYDIAB,[57,62] it was not until the DIPP[63] and TEDDY[64,65] studies that two endotypes were discovered.[66] The first was children who develop IAA as the first appearing autoantibody at 1 to 4 years of age (Fig. 32.4A), primarily in HLA DR4-DQ8 children.[66,67] The second was children who developed GADA at 2 to 3 years of age (Fig. 32.4B), primarily in HLA DR3-DQ2 children.[66,67] While the incidence of IAA first decreased with increasing age, the incidence of GADA first remained elevated. Further data suggest that the two endotypes—IAA first and GADA first—may also, prior to seroconversion, be related to reported virus infection,[68] analysis of the virome to indicate enterovirus B prolonged shedding,[69] and metabolomics differentiating the two endotypes.[70,71]

Genetic Etiology

Human Leukocyte Antigen Genes. T1D is both genetically associated with and linked to certain HLA genetic factors of the MHC on chromosome 6. DNA sequence analysis has revealed that more than 90% of all patients in whom T1D onset occurs before age 30 years are positive for the HLA haplotypes *DRB1*04-DQAI*03:01-BI*03:02(DR4-DQ8)*, *DRBI*03-DQA1*05:01-BI*02:01*, or both (Table 32.2). Although some 40% of the background population carry these HLA factors, they represent necessary but insufficient prerequisites for T1D. It has been estimated that HLA contributes approximately 50% of T1D risk among first-degree relatives of T1D patients.[72] Hence, it is not surprising that less than 10% of genetically susceptible subjects

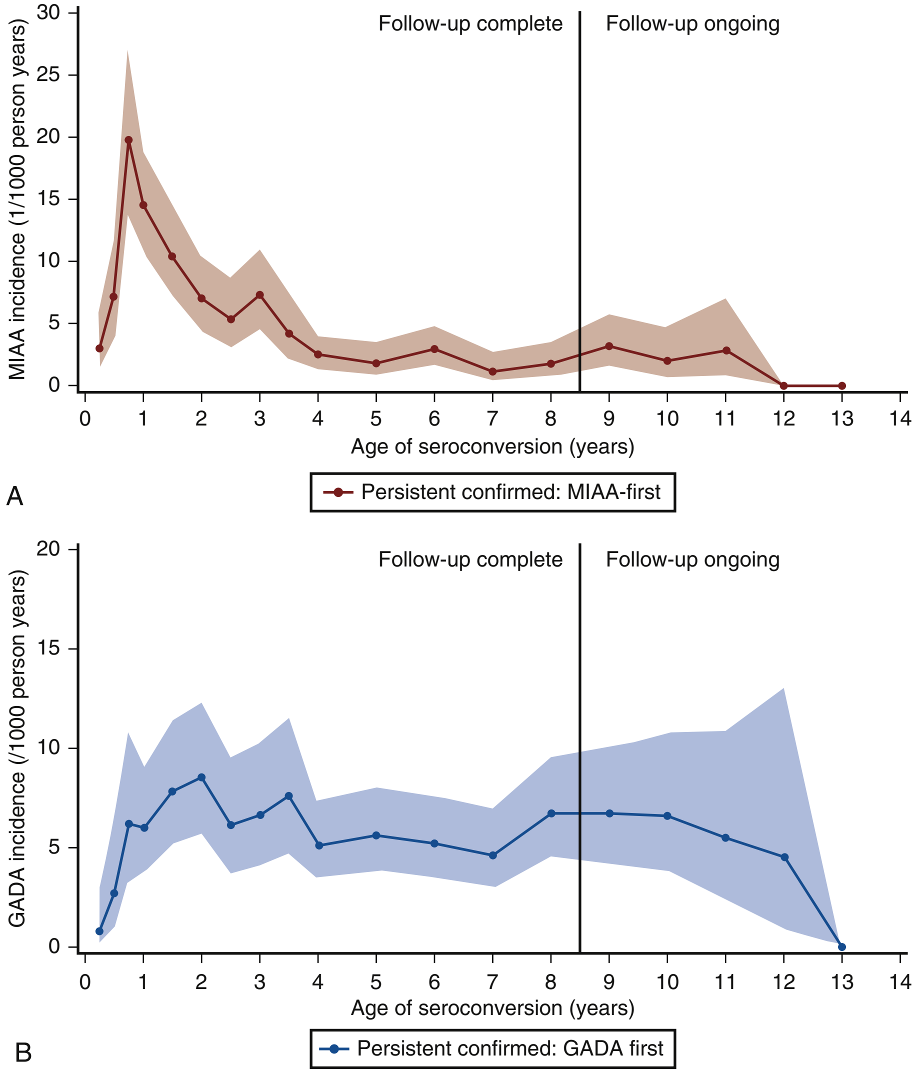


Fig. 32.4 **A,** Incidence rate of insulin autoantibodies (IAA) as the first appearing islet autoantibody in the The Environmental Determinants of Diabetes in the Young (TEDDY) study. **B,** Incidence rate of autoantibodies against glutamic acid decarboxylase (GADA) as the first appearing islet autoantibody in the TEDDY study. The difference in incidence rate suggests that the triggering of islet autoimmunity represents two different endotypes. *MIAA,* Microassay insulin autoantibodies. (Data from the TEDDY study. Used with permission.)

develop the disease. Other genetic factors have indeed been identified, but none has shown a level of importance comparable to that of HLA.[39,73]

The detailed mechanisms by which HLA confers either risk or resistance is not fully understood. The function of these molecules is to display peptide antigens to be recognized by T cell receptors. The disease association may therefore be related either to an inability to induce immunologic tolerance to certain autoantigens or to antigen presentation of an endogenous autoantigen such as enterovirus.[69] In the first case, the subject may be exposed to infectious agents that mimic autoantigens. Reactivity to the infectious agent sets off an immune response that crossreacts with self. In the second case, the immune reaction may result in a direct attack on the individual's own cells.

Non–Human Leukocyte Antigen Genes. The genetics of T1D have been studied extensively, because it represents a paradigm for genetically complex diseases. Genome screens and studies of candidate genes have provided evidence for genetic linkage between T1D and more than 60 polymorphic genetic markers (single nucleotide polymorphisms).[9,72,74] Currently, the genetic markers associated with T1D risk need to be related to possible mechanistic contributions.[72] The clinical diagnosis of T1D is based on specific criteria, not whether IAA or GADA is the first islet autoantibody to appear (as a marker of etiology) or whether a genetic factor representing pathogenesis is present. The latter would raise the question of whether a genetic marker would indicate T1D pathogenesis (stage 1 and 2) or definitive end-stage T1D (Table 32.3). *PTPN22* (chr.1p13), *CTLA-4* (chr.2q33), *IFH1* (chr.2q24), *IL2* (chr.4q27), *ITPR3* (chr.6p21), *IL2RA* (chr.10p15), *INS-VNTR* (chr.11p15), *TH* (chr.11p15), *ERBB3* (chr.12p13), *C12orf30* (chr.12q24), *CLEC16A/KIAA0350* (chr.16p13), and *PTPN2* (chr.18p11), as well as *BACH2* (chr.6q15) and *UBASH3A* (chr.21q22), are all associated with increased T1D risk (Table 32.3). Taken together, these and other genetic factors have limited effects on the relative risk of T1D.

The genetic factors need to be analyzed in relation to the etiology and pathogenesis. The etiology is reflected by the first appearing islet autoantibody. IAA-first was associated with the *INS* gene and GADA-first with *BACH2-T* allele.[67,75,76] Once two or more islet autoantibodies have

TABLE 32.1 Criteria for the Diagnosis of Diabetes

FPG ≥126 mg/dL (7.0 mmol/L). Fasting is defined as no caloric intake for at least 8 h.*

OR

2-h PG ≥200 mg/dL (11.1 mmol/L) during OGTT. The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.*

OR

HbA1c ≥6.5% (48 mmol/mol). The test should be performed in a laboratory using a method that is NGSP-certified and standardized to the DCCT assay.*

OR

In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥200 mg/dL (11.1 mmol/L).

DCCT, Diabetes Control and Complications Trial; *FPG*, fasting plasma glucose; *OGTT*, oral glucose tolerance test; *WHO*, World Health Organization; *2-h PG, 2-h* plasma glucose; *NGSP*, National Glycohemoglobin Standardization Program.

*In the absence of unequivocal hyperglycemia, diagnosis requires two abnormal test results from the same sample or in two separate test samples.

(Data from American Diabetes Association. 2. Classification and Diagnosis of Diabetes. *Diabetes Care*. 2020;43:S14–S31.)

TABLE 32.2 Class II Human Leukocyte Antigen (HLA) Haplotype Risk Genes of Type 1 Diabetes in White Americans

Effect	HLA Locus			
Susceptibility*	**DRBI**	**DQA1**	**DQB1**	**Abbreviation**
S1	04:05	03:01	03:02	DQ8
S2	04:01	03:01	03:02	DQ8.1
S3	03:01	05:01	02:01	DQ2.5
S4	04:02	03:01	03:02	DQ8
S5	04:04	03:01	03:02	DQ8
Protection†				
P1	07:01	02:01	02:01	DQ2.1
P2	14:01	01:04	05:03	DQ5.3
P3	15:01	01:02	06:02	DQ6.2
P4	04:03	03:01	03:02	DQ8.2

Modified from Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk. Analysis of the Type 1 Diabetes Genetics Consortium families. *Diabetes*. 2008;57:1084–1092.

*The five most susceptible (S).

†The four most protective (P).

The commonest high-risk haplotype is: HLA DR4-DQ8 (DQBI*03:02-A1*03:01).

The commonest moderate-risk haplotype is: HLA DR3-DQ2 (DQB1*02:01-A1*05:01).

developed without (stage 1) or with (stage 2) dysglycemia, other genetic factors contribute to the pathogenesis. It needs to be determined if these genetic factors contribute to the clinical onset of T1D (stage 3) by being protective or if they have an additive or potentiating effect on the risk provided by HLA. It is noted that most of the genetic factors linked to T1D risk are involved in the function of the immune system.

Environmental Etiology

Viral Etiology. There are numerous reports of individual patients in whom T1D onset followed an acute viral infection, borne out by subsequent animal studies in which viruses have been shown to have diabetogenic activity.[77] The first case linking T1D to an acute viral infection was reported in the late nineteenth century; the onset of T1D appeared to be precipitated by a mumps infection in a child.[78] Many similar reports have followed since, and, taken together, suggest a relation between the clinical onset of T1D, and several viruses, including rubella virus, mumps virus, coxsackie B virus, rotavirus, cytomegalovirus, and Epstein–Barr virus, have been implicated.[77] The true relation between these viral diseases and the clinical onset of T1D remains conjectural (Table 32.3).

Viruses are thought to increase the risk of T1D through different proposed mechanisms. Whereas some viruses induce a rapid and direct effects, causing β cell destruction and subsequent insulin deficiency, other viruses induce slower, longstanding effects through activation of autoreactive T cells (Table 32.4). Viruses may also inhibit insulin production by inducing interferons and HLA antigen expression, or they may mimic β cell autoantigens.[77,79]

Maternal enteroviral infection during pregnancy also appears to be a risk factor for childhood T1D. It is therefore possible that gestational infections by many types of viruses affect the maturation of the immune system, causing certain children to be more predisposed to autoimmunity, thereby increasing the risk for T1D. Prior exposure to measles, mumps, and rubella, but not vaccination, decreased the prevalence of pancreatic and thyroid autoantibodies and also the prevalence of appearance of IAA first.[69,80]

Coxsackie B virus was isolated from the pancreas of a child who died at presentation with T1D. It was propagated in *in vitro* cultures of endocrine pancreatic cells and was shown to have diabetic activity in certain mouse strains. Two important hypotheses follow from these experiments. One hypothesis is that the coxsackie B virus induces β cell neoantigens, which initiate an (auto)immune reaction. This hypothesis can be tested by analyzing the appearance of such neoantigens. The neoantigen may initiate a devastating reaction if its structure mimics a self protein. Another hypothesis is that coxsackie B virus replication in β cells results in β cell necrosis and the formation of antibodies against β cell constituents or "hidden antigens" not normally surveyed by the immune system. An autoimmune reaction is initiated, reflected by IAA as the first appearing autoantibody. It was reported that GADA first may appear in relation to a gastrointestinal infection, possibly by mastadenovirus, also related to prolonged shedding.[69] The etiology of T1D may therefore be related to an early initiation where infection by virus is involved but may not be sufficient, as indicated by the possible contribution of gestational infections[68] or maternal life events during pregnancy.[75]

Psychosocial Factors. Case control studies have reported that negative life events occurring during the first 2 years of life, difficult adaptation, or child behavioral deviances may increase the risk of T1D.[81] It has also been reported that psychological stress, measured as psychosocial strain during pregnancy, seems to be involved in the induction or progression of β cell–related autoimmunity in cord blood and during the first year of life.[82] Additionally, high parenting stress, experiences of serious life events, foreign origin of the mother, and low socioeconomic status were all associated with β cell–related autoimmunity in young children.[82-84] It was hypothesized that the parent–infant relationship may be a "psychobiological regulator," and that early disruption of this relationship may cause stress and affect the hypothalamic-pituitary-adrenal axis (Table 32.4). Such changes may affect the immature autonomic nervous system of the child and subsequently impair the regulation of autoimmunity.[75]

Maternal Factors. Several maternal factors were reported to increase the risk of T1D in children. Higher maternal age (and to a lesser extent,

TABLE 32.3 Non–Human Leukocyte Antigen Genetic Factors Contributing to Type 1 Diabetes Risk

Markers	P-Value[a]	Odds Ratio[b]	Causal Gene Candidate[c]	Gene function[d]	Regulome_ score[e]
rs2476601, rs6679677	1.10E-122	1.89	*PTPN22*	Regulation of innate immune response, T cell activation, NK cell proliferation	2b
rs3024493, rs3024505	1.18E-07 6.40E-08	0.86	*IL10*	Cytokines and inflammatory response	2b
rs9653442	3.39E-07	1.11	*AFF3*	Regulation of transcription	6
rs1990760, rs2111485, rs35667974	5.84E-17 3.80E-18 9.30E-09	0.85 0.85 0.59	*IFIH1*	Innate immune system NF-kB activation	7
rs7574865	4.11E-04	1.1	*STAT4*	Cytokine-mediated signaling pathway	7
rs11571316, rs3087243	2.44E-15 7.40E-21	0.82 0.84	*CTLA4*	T-cell activation	3a
rs113010081, rs333	4.60E-08 5.87E-06	0.85	*CCR5*	Th1 development and chemokine-mediated signaling pathway	4
rs17388568, rs4505848, rs6827756, rs75793288	1.06E-09 3.77E-10 5.40E-12 5.60E-13	1.13 1.12 1.14 1.15	*IL21, IL2*	Cytokines and inflammatory response, Th1/Th2 Differentiation	7
rs11954020	4.40E-08	1.11	*IL7R*	T-cell–mediated cytotoxicity, immunoglobulin production, antigen binding	5
rs11755527, rs597325, rs72928038	4.66E-12 1.12E-11 6.40E-14	1.1 0.88 1.20	*BACH2*	Transcription	5
rs6920220	7.26E-06	1.12	*TNFAIP3*	Inflammatory response	6
rs1738074	7.59E-09	0.92	*TAGAP*	Signal transduction	4
rs10272724, rs62447205	3.32E-08 2.50E-08	0.89	*IKZF1*	Immune cell regulation	5
rs10758593, rs6476839, rs7020673	1.42E-09 1.00E-09 1.79E-08	1.12 1.12 0.9	*GLIS3*	Regulation of transcription	4
rs10795791, rs12251307, rs41295121, rs61839660, rs7090530	2.60E-23 7.93E-24 5.53E-09 2.80E-39 2.93E-15	1.2 0.73 0.52 0.62 0.82	*IL2RA*	Alternative mRNA splicing Th1/Th2 Differentiation	3a
rs11258747	9.84E-09	0.69	*PRKCQ*	Apoptotic process, inflammatory response, innate immune response, T-cell receptor signaling pathway	1f
rs722988	4.88E-08	1.11	*NRP1*	Signal transduction	4
rs689, rs7111341, rs72853903	1.11e-233 3.14E-88 3.13E-88	0.42 0.63 0.63	*INS*	Insulin signaling pathway	5
rs694739	7.69E-06	0.92	*BAD*	Apoptosis	1b
rs10492166, rs4763879	6.00E-09 2.50E-07	0.87 1.1	*CD69*	Signal transduction	1d
rs11170466	7.86E-09	1.19	*ITGB7*	Response to virus, regulation of immune response	4
rs11171710, rs2292239, rs705704, rs705705	7.23E-22 6.01E-30 8.25E-32 4.40E-32	1.25	*ERBB3,*	Regulation of transcription, innate immune response, lipid metabolism	1f
rs10877012	3.80E-06	0.82	*CYP27B1*	Metabolism of lipids, lipoproteins, steroid hormones, and vitamin D	6
rs17696736, rs3184504, rs653178	3.62E-29 1.56E-44 1.56E-44	1.24 0.76 0.76	*SH2B3*	Signal transduction	3a
rs9585056	3.30E-08	1.12	*GPR183*	Humoral immune response	7
rs56994090, rs941576	1.10E-11 1.62E-10	0.88 0.90	*DLK1*	Regulation of gene expression	5

TABLE 32.3 **Non–Human Leukocyte Antigen Genetic Factors Contributing to Type 1 Diabetes Risk—cont'd.**

Markers	P-Value[a]	Odds Ratio[b]	Causal Gene Candidate[c]	Gene function[d]	Regulome_ score[e]
rs12908309, rs72727394	4.31E-08 3.60E-10	0.85 1.15	*RASGRP1*	Inflammatory response to antigenic stimulus, cytokine production	4
rs12148472, rs34593439, rs3825932	4.18E-12 9.00E-14 1.16E-07	0.81 0.78 0.9	*CTSH*	Immune response-regulating signaling pathway, T-cell–mediated cytotoxicity, adaptive immune response	1f
rs12708716, rs12927355, rs193778	3.39E-21 3.00E-22 4.40E-10	0.83 0.82 1.14	*CLEC16A*		1f
rs151234, rs4788084, rs9924471	4.80E-11 1.03E-09 1.21E-11	1.19 0.9 1.24	*IL27*	Inflammatory response, regulation of defense response to virus	1a
rs12453507, rs2290400	1.00E-08 1.31E-08	0.9	*ORMDL3*	Protein binding	6
rs1893217, rs2542151	1.20E-15 4.75E-15	1.2	*PTPN2*	Cytokine signaling, B- and T- cell differentiation	3a
rs1615504, rs763361	1.80E-11 8.03E-11	1.13	*CD226*	Immunoregulatory, adaptive immune system	6
rs12720356, rs2304256, rs34536443	3.70E-07 3.61E-11 4.40E-15	0.82 0.87 0.67	*TYK2*	Cytokine-mediated signaling pathway, intracellular signal transduction, type I interferon signaling pathway	5
rs516246, rs601338, rs602662	5.20E-14 4.30E-18 7.80E-14	0.87 0.75 0.87	*FUT2*	Metabolic pathways	5
rs11203202, rs11203203	1.20E-15 2.91E-15	1.16	*UBASH3A*	Regulation of cytokine production, regulation of T-cell receptor signaling pathway	3a
rs229533, rs229541	1.80E-08 7.94E-08	1.11	*C1QTNF6*	B-cell receptor signaling pathway, chemokine signaling pathway, natural killer cell mediated cytotoxicity	6

paternal age) at the time of delivery increased the risk of T1D in the offspring. On the other hand, birth order seems to have positive association with the risk of T1D, as the first child carries the highest risk, and this risk declines by 15% with each new child. Exposure of the mother to enteroviral infection, colder season of delivery, ABO blood group incompatibility, maternal islet autoantibodies, or eclampsia all increased the risk of T1D.[85,86] An analysis of nondiabetic mothers with islet autoantibodies at delivery indicated that an association between HLA and islet autoimmunity may depend on enterovirus exposure during pregnancy.[87 88] Cord blood from children who developed islet autoimmunity or T1D before 8 years of age showed changes in phospholipids that would mark their risk for T1D.[89,90] These and other investigations in humans underscore the importance of clarifying whether transmission of maternal autoimmunity, including placental transfer of autoantibodies, affects the incidence of T1D in susceptible offspring.

Dietary Factors. The dietary hypothesis is based on earlier observations that breastfed children have a lesser risk of T1D. It was also noted that early exposure of infants younger than 6 months to cow's milk proteins may double the T1D risk, particularly in HLA high-risk children. Other studies have described an association between islet autoimmunity and shorter duration of breastfeeding, as well as early exposure of infants to bovine protein (Table 32.4). The use of a hydrolyzed formula, when compared with a conventional formula, did not reduce the incidence of diabetes-associated autoantibodies after 7 years in the TRIGR study.[91] Similarly, the EURODIAB Substudy-2 Group suggested that rapid growth, rather than cow's milk or early introduction of solid food, would explain the increased incidence rate of T1D.[38] Vitamin D or n-3 fatty acid supplementation during pregnancy showed no benefit regarding islet autoantibody risk.[92] Early probiotic supplementation may reduce the risk of islet autoimmunity in children at the highest genetic risk of T1D,[93] but not the intake of dietary soluble fiber in early life.[94] Later introduction of gluten was associated with increased risk of any islet autoantibody.[95]

Toxic Substances. Alloxan and streptozotocin are widely used to induce β cell destruction and diabetes in experimental animals. Some species, including humans, are more resistant to these drugs than others (Table 32.4). Streptozotocin is commonly used in treating certain gastroenteric tumors, including glucagonomas; however, T1D rarely develops in these patients. Several compounds structurally related to streptozotocin and alloxan have been implicated as possible environmental agents contributing to T1D. Pyriminil (Vacor), an effective rodenticide, is highly diabetogenic in humans.[96] The TEDDY study is an international effort to substantiate the possibility that environmental chemicals are causative factors in the development of T1D.[97]

In summary, viruses and chemical agents have direct effects on pancreatic β cells and may therefore represent the causative factors that initiate the autoimmune process against these cells. The alternative hypothesis is that these agents potentiate a β cell–destructive process that is genetically determined or initiated by environmental factors.

Birth Weight and Growth Rate. Several studies suggest a positive association between T1D HLA risk and an increase in perinatal and postnatal growth.[34,98] HLA genotypes may be associated with intrauterine growth, independent of T1D risk.[99] Greater weight in the first years of life was associated with an increased risk for a first appearing islet autoantibody.[100] A higher rate of weight gain was associated with increased risk of progression from GADA first to T1D.[101] Furthermore, being overweight at 2to 10 years of age was associated with increased risk of progression from multiple islet autoantibodies to clinical onset of T1D.[102] Children with higher weight-for-height or body mass index during infancy or childhood have greater T1D risk.[103] T1D boys were significantly taller than healthy boys from 1 to 7 years before clinical onset, regardless of age at onset.[35]

KEY POINTS

- Two endotypes were discovered. The first is children who develop insulin autoantibodies (IAAs) as the first appearing autoantibody at 1–4 years of age, primarily in human leukocyte antigen (HLA) DR4-DQ8 children. IAA first may be related to enterovirus B prolonged shedding. The second is children who develop GAD65 autoantibodies (GADA) at 2–3 years of age, primarily in HLA DR3-DQ2 children. GADA first may be related to mastadenovirus infection. While the incidence of IAA first decreases with increasing age, the incidence of GADA first remains elevated.

TABLE 32.4 Viruses and Other Environmental Factors Implicated In Type 1 Diabetes Etiology and Pathogenesis

	COMMENT	
Factor	**Host**	**Diabetes in Animals**
Viruses		
Enterovirus B incl. Coxsackie B (RNA)	Human	Mice
Rubella virus (RNA)	Human	Hamsters
Mumps virus (RNA)	Human	None
Cytomegalovirus (DNA)	Human	Cytomegalovirus in rats
Echovirus (RNA)	Human	None
Ljunganvirus (RNA)	Human	Bank voles
Rotavirus (RNA)	Human	None
Diet and Nutrition		
Cow's milk protein/bovine insulin	Early introduction before 4 months of age	
Cereals/gluten	Early introduction before 4 months of age	
High-calorie carbohydrates		
Nitrosamine-containing foods		
Vitamin D deficiency		
Growth, Body Size, and Weight Gain	Excess stress on β cells (accelerator hypothesis)	
Psychological Stress	Possibly through changing the hypothalamic-pituitary-adrenal regulatory mechanism	
Hygiene Theory	Inadequate maturation of immune system and subsequent autoimmune reaction	
Toxins		
Pyriminyl (Vacor)		
Pentamidine isothiocyanate		
N-3-Pyridylmethyl-N'p'-nitrophenylurea (PNU)		

PATHOGENESIS

The pathogenesis of T1D is strongly associated with several immune abnormalities. These immune abnormalities, in particular autoantibodies directed against insulin, GAD65, IA-2, and ZnT8, are dynamic markers of an ongoing disease process. These pathogenic markers are useful for predicting and classifying T1D. HLA does not seem to influence the tempo of the disease process when progressing from a first to a second islet autoantibody.[76,104] The role of the HLA class II molecules in the immune response is to convey cell–cell interactions between T-helper (CD4+) lymphocytes and antigen-presenting cells (APCs) or B lymphocytes. The CD4+ T cell provides help to cytotoxic T lymphocytes (CD8+) as well as to B lymphocytes. The importance of the immune system to T1D pathogenesis is underscored by recent genome-wide association studies, which suggest that variants of genetic factors important to the immune response increase the risk of T1D.[72]

What are the factors that contribute to progression from multiple autoantibodies without dysglycemia (stage 1) and multiple autoantibodies with dysglycemia (stage 2) to clinical onset (stage 3)? First, the number of islet autoantibodies accelerates progression to stage 3.[105,106] In persistently autoantibody-positive children, higher IAA and IA-2A levels, but not GADA levels, increased the risk for stage 3.[107] Different risk scores have been developed to predict the time to stage 3 among subjects with multiple islet autoantibodies.[108-111] Islet autoantibodies are becoming widespread as biomarkers of autoimmune T1D, and an increasing number of subjects are detected in stage 1 or stage 2 T1D. The need for additional biomarkers to better predict the time to clinical onset will be needed in anticipation of treatment that will halt the pathogenesis at stage 1 or stage 2.

Humoral Immunopathophysiology

Islet Cell Antibodies. ICAs detected by indirect immunofluorescence analysis of frozen sections of human blood group O pancreas were the first indication of autoantibodies in T1D.[46] ICSAs (see Table 32.5) in newly diagnosed T1D patients were demonstrated in dispersed-cell preparations of rat or mouse pancreatic islets.[48] The observation that antibodies are capable of binding to living β cells is important, because it allows an answer to the question of whether islet autoantibodies either mediate immune effector mechanisms or directly affect the function of the β cells. The former phenomena include complement-mediated cytotoxicity and antibody-dependent cellular cytotoxicity.[49] Either mechanism could contribute to killing pancreatic β cells, provided that the *in vitro* phenomenon is also occurring in vivo. However, knowledge of the autoantigen has improved our understanding and detection of islet autoimmunity. Several assay systems are now in use to determine the presence of antibodies reactive with pancreatic islet cell autoantigens (Table 32.5). Autoantibodies to GAD65, IA-2, insulin, and the zinc transporter ZnT8 have replaced ICA (Fig. 32.5). One or more of these autoantibodies may be present in more than 95% of newly diagnosed T1D patients,[112,113] and they present variably during the prodrome of islet autoimmunity long before the clinical onset.

In order to provide unified comparative assays, standardization of IAA, GADA, IA-2A, and ZnT8A has been carried out since 1985[114] and continued through the Diabetes Antibody Standardization Program (DASP)[115] and Islet Autoantibody Standardization Program (IASP).[56,116] These results imply the value of GADA, IAA, IA-2A, and ZnT8A for screening high-risk populations, as well as classifying patients with diabetes.

Glutamic Acid Decarboxylase Autoantibodies. Glutamic acid decarboxylase is present mainly in pancreatic β cells and neuron cells. This enzyme produces γ-aminobutyric acid, which is stored in

TABLE 32.5 Methods to Detect Autoantigen-Specific Islet Cell Autoantibodies

	METHOD OF DETECTION			
Preparation	**RBA**	**ELISA**	**Autoantibody Abbreviation**	**References**
Human recombinant glutamic acid decarboxylase	+	+	GADA	204,205
Human recombinant insulin	+	To be developed	IAA	53,123
Human recombinant IA-2	+	+	IA-2A	206
Human recombinant phogrin	+	To be developed	IA-2βA	115,207
Human recombinant zinc transporter ZnT8	+	+	ZnT8A	55,116
Other Non–Autoantigen-Specific Assays				
Frozen sections of human pancreas	Indirect immunofluorescence		ICA	46
Dispersed rat islet cells	Indirect immunofluorescence		ICSA	48
Monolayers of rat islet cells	^{51}Cr release		C′AMC	49

ELISA, Enzyme-linked immunosorbent assay; *RBA*, radioligand binding assay; *GADA*, glutamic acid decarboxylase autoantibody; *IAA*, insulin autoantibody; *IA-2*, islet antigen-2 autoantibody; *ICA*, islet cell autoantibody; *ICSA*, islet cell surface autoantibody; *C′AMC*, complement-dependent antibody-mediated cytotoxicity; ZnT8A, zinc transporter 8 autoantibody.

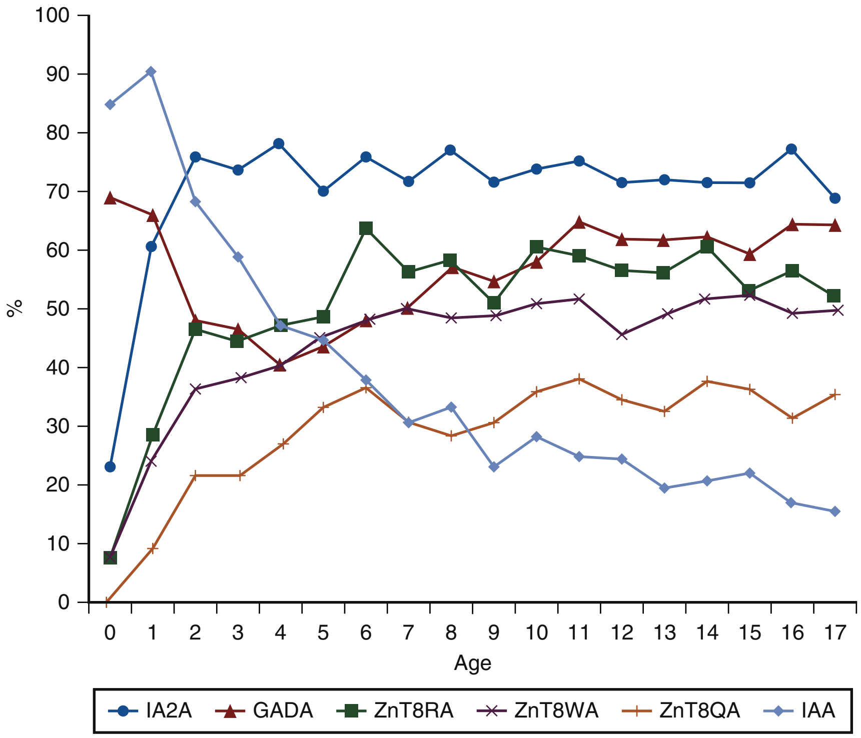


Fig. 32.5 Frequency of autoantibodies against glutamic acid decarboxylase (GADA) and insulinoma-associated antigen 2 (IA2A), as well as islet cell antibodies (ICAs), in relation to the age group in which the patient was diagnosed. The number of patients in each age group is shown beneath the age on the x-axis. *IAA,* Insulin autoantibodies *ZnT8RA,* ZnT8 arginine autoantibodies; *ZnT8WA,* ZnT8 tryptophane autoantibodies; *ZnT8QA,* ZnT8 glutamine autoantibodies. (Data from Andersson C, Kolmodin M, Ivarsson SA, et al. Islet cell antibodies (ICA) identify autoimmunity in children with new onset diabetes mellitus negative for other islet cell antibodies. *Pediatr Diabetes.* 2014;15:336–344. Used with permission.)

small neurotransmitter vesicles. Two isoforms are known: GAD65 and GAD67. GAD65, detected in and cloned from human islets after immunoprecipitation, is coded for by a gene on human chromosome 10.[52] The previously known GAD67, encoded by a gene on chromosome 2, shares 65% of the amino acids, but this isoform is not expressed in human β cells.

The frequency of GADAs in children with new-onset T1D is 70% to 80% (Fig. 32.5). Patients with GAD67 autoantibodies are almost invariably GADA-positive. GADA frequency in new-onset patients is little affected by the age at onset; however, in children younger than 10 years, more girls than boys have GADA. GADAs are evanescent, but less so than ICA. GADAs tend to be the most prevalent and persist for longer duration after the diagnosis of T1D when compared with other autoantigens. Almost 50% of patients with disease duration of 10 years may still be GADA-positive.

Any assessment of an individual's risk of developing T1D depends on the ability to accurately determine the presence of autoantibodies in a prospective analysis. This is particularly important, because GADA may appear temporarily in healthy individuals. Persistent GADAs are more often detected in subjects with *DRBI*03-DQA1*05:01-B1*02:01* than in those with *DRB1*04-DQ A1*03:01-B1*03:02.*[117] Truncated GADA may be present in both HLA haplotypes.[118] GADA may be detected in approximately 1% of the general population, but the frequency increases to approximately 8% among first-degree relatives of patients with T1D. The positive predictive value of GADA for T1D is approximately 50% or higher among first-degree relatives. GADAs also predict stiff person syndrome, although the GADA epitopes are different from those in T1D.[119] T1D is best predicted when GADA is combined with IAA, IA-2A, and ZnT8A.[25,120]

Insulin Autoantibodies. Insulin is produced after the C-peptide has been cleaved from proinsulin. The molecular weight of insulin is 5.8 kD, and the molecule is composed of the A and B chains. It was demonstrated in a radioligand binding assay that approximately 50% of patients with newly diagnosed but untreated T1D had IAAs.[53] Among children at T1D genetic risk and younger than 5 years of age, IAAs were reported to be the first appearing islet autoantibodies.[57,63,66] IAA is therefore the most common islet autoantibody in young children, but the frequency at diagnosis (stage 3) decreases with increasing age (Fig. 32.5). Proinsulin is viewed as a primary specific autoantigen, especially because it is unique to the β cells. IAAs react with both insulin and proinsulin, indicating that these proteins possess shared epitopes as targets of IAA.

IAAs were positively associated with DR4, perhaps because they are in linkage disequilibrium with *DQA1*03:01-DQB1*03:02.*[66,117] A first analysis of putative IAA epitopes indicates that the amino acids B1 to B3 and A8 to A13 are important.[121] IAAs and insulin antibodies that appear after insulin therapy have similar binding characteristics.[122]

IAA assays have been standardized, and only the fluid-phase radioimmunoassay (but not the enzyme linked immunosorbent assay [ELISA]) was found to have high diagnostic sensitivity and specificity for T1D. The fluid-phase radioimmunoassay for IAA is still lacking an international standard, and the IAA test is therefore yet to be standardized.[123]

Islet Antigen-2 Autoantibodies. Trypsin treatment of the 64-kD immunoprecipitate revealed 40-kD and 37-kD fragments of islet antigen, yielding two isoforms of IA-2A: IA-2 (ICA512) and IA-2β (phogrin).[54,124-126] IA-2 and IA-2β are both associated with the secretion granule membrane. These isoforms share high structural similarity but lack enzymatic activity, and their functions in neuroendocrine tissues such as the hypothalamus and pancreatic islets are not fully understood. IA-2As are detected in 60% to 70% of patients with new-onset T1D (Fig. 32.5), but in older patients IA-2As contribute less. Longitudinal studies of first-degree relatives suggest that IA-2As tend to appear closer to the clinical onset of T1D compared with GADAs. IA-2A is rarely a first appearing islet autoantibody.[63,66]

IA-2As were associated primarily with *DRB1*04:01* and not with *DQA1*03:01-DQBI*03:02,*[127] and the risk for this antibody is increased in males. Surprisingly, IA-2A in T1D was negatively associated with *DQA1 *05:01-DQBI*02:01.*[117] The mechanisms by which the IA-2 proteins become recognized as autoantigens are therefore not yet understood. The frequency of IA-2A in the general population is less than 1%. In young children with genetic susceptibility, progression to clinical T1D onset was associated with broad epitope response and juxtamembrane reactivity to the IA-2 antigen.[127]

The predictive value of IA-2A for T1D is best estimated in combination with GADA and IAA. ELISA tests for IA-2Ab are available, with similar diagnostic sensitivity and specificity to radioligand assays, and combination ELISA tests using biotin-labeled preparations of both IA-2A and GADA have been developed. IA-2Ab performance with ELISA was improved in DASP and IASP to achieve levels of equivalent to in-house results for radioimmunoassay (see Table 32.7).

ZnT8 TransporterAutoantibodies. The β cell–specific ZnT8 transporter was found not only to be the target of islet autoimmunity but also a genetic marker for T2D.[128] The ZnT8 transporter is important to the formation of insulin crystals in that it facilitates zinc transportation into the β cell insulin granules. This antigen is encoded by the *SLC30A8* gene, which is one of nine human genes for multispanning transmembrane proteins. Epitope mapping of these autoantibodies showed that ZnT8A primarily react with the C-terminal end of the protein.[129] Polymorphic variants of ZnT8 at amino acid position 325 in the cytosolic tail were found to be either arginine (ZnT8-R) or tryptophan (ZnT8-W), both representing markers of ZnT8A susceptibility with potential diagnostic, therapeutic, and prognostic uses.

ZnT8As were detected in 26% of subjects who were reported to be negative for GADA, IA-2A, and IAA. On the other hand, ZnT8As were found to be positive in 60% to 80% of newly diagnosed T1D patients (Fig. 32.5) but in only 2% of controls and less than 3% of T2D patients.[130]

ZnT8As were also found during the islet autoimmune prodrome, often after 2 years of age and independently of other autoantibodies, but mainly later than GADA and IAA.[63,66] The high specificity and independent appearance of ZnT8As highlight their usefulness for prediction, especially in older children and adolescents, as ZnT8Ab prevalence and titer levels increase with age.[25,112]

Following similar principles of GADA and IA-2A standardization, fluid-phase radioassays of ZnT8Ab (C-terminal) were standardized and validated in the DASP and IASP workshops (see Table 32.7).[116]

In summary, currently available methods for detecting autoantigen-specific autoantibodies (GADA, IA-2A, IAA, and ZnT8A) are of value in predicting T1D. Standardized assays for these autoantibodies are important when selecting participants for prevention and intervention trials. How the uptake, processing, and presentation of these autoantigens by APCs that initiate the formation of autoantigen-reactive T and B lymphocytes are accomplished are important questions that remain to be answered.

Candidate (Minor) Autoantigens. Several proteins or other substances implicated in T1D autoimmunity may represent candidate or minor autoantigens. Autoantigens reported to be associated with T1D include imogen38,[131,132] vesicle-associated membrane protein-2 (VAMP2),[133] neuropeptide-Y (NPY),[133,134], glima-38[135] representing TSPAN7,[136,137] heat shock protein 60 (HSP60),[138,139] ICA69,[140,141] carboxypeptidase H (CPH),[142] DNA topoisomerase II,[143] CD38 (ADP ribosyl cyclase/cyclic ADP-ribose hydrolase),[144] type-2 glucose transporter (GLUT-2),[145,146] and ICA12/SOX13.[147,148] The role of these candidate autoantigens in islet autoimmunity, T1D, or both, remains to be determined.

Cellular Immunopathophysiology

Pancreas Pathology. Studies of the pancreas in newly diagnosed T1D patients have indicated that the gland is diminished in size compared with that in matched controls.[149] An atrophic pancreas with little or negligible residual insulin is typical of a patient with longstanding T1D. The atrophy affects primarily the tail of the pancreas. In this part, the composition of the endocrine pancreas is dominated by β and α cells, whereas in the pancreatic head, β and pancreatic polypeptide (PP) cells predominate. In addition to insulin deficiency, the pancreatic atrophy appears to be reflected in lower serum levels of pancreatic trypsin and isoamylase.[150,151] It is speculated that the pancreatic atrophy is a result of an intrapancreatic insulin deficiency affecting the growth of the pancreas.

The insulin deficiency is the result of a specific loss of the β cells, which affects the total mass of the endocrine pancreas in T1D

TABLE 32.6 **Morphometric Analysis of the Endocrine Pancreas in Control Individuals and Patients with Type 1 Diabetes Mellitus (T1DM)**

	ENDOCRINE PANCREAS (μg)	
Cell T`ype	Controls	T1DM
β cells	850	0
α cells	230	150
δ cells	125	97
PP cells	190	166
All cells	1395	413

PP, Pancreatic polypeptide.
Adapted from Rachier J, Goebbles RM, Henquin JC. Cellular composition of the human diabetic pancreas. *Diabetologia.* 1983;24:336–371. Used with permission.

(Table 32.6). The islets are small and appear pseudoatrophic. In contrast to the normal pancreas, in which the β cells predominate, the endocrine cells in the diabetic pancreas are primarily α, δ, and PP cells. Lobular variation has been described, and areas with β cell–containing islets without pancreatic atrophy also have been detected, suggesting that parts of the pancreas may be spared in individuals with newly diagnosed T1D.[41] Sometimes β cells are seen in close proximity to duct cells, suggesting neoformation of cells. The understanding of the developing endocrine pancreas has improved, and a number of transcription factors controlling the process have been identified.[72] However, one of the characteristics of the pancreas in many children with newly diagnosed T1D is the presence of inflammatory cells adjacent to islets or cords of newly formed β cells.[152]

Immunocytochemical investigation in rare specimens of pancreas from patients who died shortly after the clinical onset of T1D indicates that all cell types considered part of the immune system populate the islets to form the insulitis.[41,152,153] T lymphocytes, B lymphocytes, macrophages, and occasional natural killer cells may be seen. In agreement with earlier studies, examination of pancreatic islets in laparoscopically obtained biopsy specimens from newly diagnosed T1D patients[43] demonstrated insulitis in approximately 60% of patients, with a T cell–predominant infiltration and increased expression of HLA class I.[154,155] Notably, the presence of insulitis was strongly correlated with the presence of other autoimmune features, specifically positivity for GADAs or IA-2As.[44] In contrast, examination of specimens from 1507 pancreatic donors aged 25 to 60 years revealed 62 (4%) islet autoantibody–positive subjects.[156] Insulitis was found in only two (3%) of the 62 subjects; the two subjects were positive for multiple islet autoantibodies and susceptible HLA. None of the 62 matched controls had insulitis. These results suggest that positivity for islet autoantibodies does not immediately mean that insulitis is present. Islet autoantibodies seem to precede insulitis and seem to require not only one but a concurrent presence of multiple islet autoantibodies.

Taken together, current observations support the hypothesis that T1D is an immune-mediated disease and that immunocytes play a pivotal role in the loss of β cells. It cannot be excluded that a diminished β cell mass alters the rate of insulin biosynthesis and secretion in remaining β cells, as well as insulin resistance in peripheral tissues.

Peripheral Blood Mononuclear Cells. It is possible that APC activity is altered in patients or in individuals susceptible to T1D. Blood monocytes in T1D patients may be defective in prostaglandin synthetase and dendritic cells from patients with T1D have a maturational and functional defect.[157] APCs from T1D patients secrete markedly higher levels of proinflammatory cytokines in response to a nonantigenic stimulus.[158] Also, healthy HLA DR3–positive subjects have an abnormally prolonged Fc receptor-mediated mononuclear phagocyte system.[159]

T lymphocytes are also important to the pathogenesis of T1D. The earliest demonstration involved tests for leukocyte migration inhibition or blast formation, suggesting that T1D patients may be sensitized to pancreatic antigens.[42] Several disorders of autoimmune character have been found to have an imbalance in the peripheral blood between T-helper (CD4+), cytotoxic (CD8+), and regulatory (CD4+CD25+) T-cells. The use of HLA tetramers makes it possible to detect and clone autoantigen-specific human T-cells.[160-162] Other approaches to detect T-cell reactivity against islet cell antigens, such as enzyme-linked immune absorbent spot, which measures cytokine secretion from individual T-cells, are being explored.[163] Studies on peripheral blood cells are rare in subjects later shown to seroconvert to either IAA first or GADA first.[63,66] Patterns of RNA extracted from whole blood indicated that T1D is characterized by early and longitudinal changes in gene expression to be associated with a natural killer cell signature with progression to clinical onset.[164] The timing of blood samples in children at increased genetic risk for T1D, during seroconversion to a first appearing autoantibody, during stage 1 and 2 as well as at stage 3, will be critical to identify peripheral blood cells involved in both the etiology and the pathogenesis of T1D. In conclusion, antigen-specific tests in HLA DQ– and HLA DR–matched T1D subjects and control individuals should eventually allow a proper test of the hypothesis that a specific immunoregulation abnormality involving β cell antigens is associated with the development of T1D.

DIAGNOSTIC CRITERIA

T1D was often thought to be a disorder of acute onset, and the clinical onset may be dramatic. Over the years, however, numerous reports have noted that signs of subclinical diabetes precede the clinical onset. In addition, in adult diabetes patients classified with T2D, a change sometimes occurs from an insulin-independent to insulin-dependent state.

T1D is said to have four major clinical phases: preclinical diabetes, overt diabetes, partial remission phase (honeymoon), and the chronic phase of lifelong dependency on injected insulin. Autoantibodies against GAD65, insulin, IA-2, or ZnT8, alone or in combination, may be present up to several years before the clinical onset, and predict T1D.[6,105] The clinical onset is not likely to occur until a major loss (80%–90%) of the islet β cells has occurred. Currently it is not possible to determine the human β cell mass or volume. β cell function tests are affected by both β cell mass and insulin sensitivity. It is therefore not possible to relate clinical onset to β cell mass, especially since some reports suggest that around 40% to 50% of β cells may be viable, owing to β cell regeneration at the time of overt hyperglycemia.

The occurrence of diabetic ketoacidosis (DKA) is more commonly seen with new-onset disease, especially in children younger than 4 years of age, but it is not specific to T1D and may also occur in T2D. DKA occurs invariably in different populations, ranging from 15% to 67% of patients in Europe and North America, whereas in the developing world, around 80% of children with T1D present initially with DKA.[165,166] Typically the child is acidotic with acetone fruity odor, air hunger and Kussmaul respiration, abdominal pain, nausea and vomiting, and polyuria and polydipsia. The child has hyperglycemia, glucosuria, ketonemia, and ketonuria. Without timely management, severe fluid and electrolyte depletion soon develop, with signs of hypoperfusion and deterioration in the level of consciousness due to cerebral edema that may lead to coma and death.

The basis for distinction between T1D and T2D is islet autoimmunity and the patient's dependence on insulin. Most patients have a history of polyuria, polydipsia, and unexplained weight loss. The American Diabetes Association and the International Society for Pediatric and Adolescent Diabetes have recommended diagnostic criteria for T1D (Table 32.1).

Glucose tolerance tests, both oral and intravenous, are being used to evaluate diabetic states. The OGTT is a diagnostic criterion (Table 32.1). Although glucose tolerance tests are more important in disorders of impaired glucose tolerance and T2D, they may also find an increased use in the prediction and diagnosis of T1D.

KEY POINTS

- The number of islet autoantibodies are associated with an accelerated progression to stage 3. In persistently autoantibody-positive children, higher insulin autoantibody and islet antigen-2 autoantibody levels, but not GAD65 autoantibody levels, increased the risk for stage 3.

PREDICTION

Highly effective, precise, and reproducible screening assays for HLA genotypes, as well as for GADA, lA-2A, ZnT8A, and IAA, are in use to detect individuals with autoimmune islet disease, often referred to as islet autoimmunity. Early detection of T1D serves several purposes. Among them is the distinct possibility that ketoacidosis associated with a dramatic and traumatic onset of T1D, in some cases leading to death, would be forestalled. Prediction and prevention of DKA is a rationale for islet autoantibody screening in the general pediatric population.[120] In addition, immunotherapeutic modalities for prevention of β cell destruction are under active investigation with both oral insulin[5,167,168] and anti-CD3,[169] with the hope that T1D may someday be a preventable disease.

The fact that clinical onset of T1D is preceded by a long "prediabetes" phase marked by islet autoimmunity biomarkers makes T1D a candidate for prediction and prevention. The predictive value of these autoantibodies varies with age, and combining more than one autoantibody increases the positive predictive value. The improved ability to measure autoantibodies against GAD65, IA-2, ZnT8, and insulin made it possible to predict T1D with up to 98% accuracy when ZnT8A was added to the arsenal of islet autoantibody tests.[105,170,171] A high predictive value of screening tests is critical to designing interventional studies. The levels of autoantibodies reflect the degree of aggressiveness of the autoimmune reaction and are related to persistence of positivity and subsequent progression into clinical disease.

HLA testing alone is not a primary indicator for the prediction of T1D, but it has a significant role when used with autoimmune markers to increase sensitivity of prediction among subgroups with moderate risk. However, adding HLA DQ did not affect the diagnostic sensitivity of islet autoantibodies.[25,172,173] Several other genetic loci have also been reported to be associated with T1D genetic risk, but their contribution was less than HLA (see Table 32.4).

PREVENTION

The ability to prevent or reverse autoimmune diabetes in laboratory animals has led to several prevention and intervention studies in humans. Studies in the NOD mouse and the BB rat showed that treatment with insulin or nicotinamide reduced spontaneous diabetes. These studies initiated large clinical trials, such as the Diabetes Prevention Trial-1, the Diabetes Prediction and Prevention Project, and the European Nicotinamide Diabetes Intervention Trial, all without effects (see following discussion). Preclinical studies in rodents may therefore not always be suitable to guide human trials.

Whether to choose a preventive or intervention strategy is a matter of debate, and is related to several factors. First, it has been suggested that intervention trials should be reserved for therapies with greater risk that need to be used on a small number of newly diagnosed patients within limited follow-up periods. Second, prevention trials may be implemented with safer therapies that can be used in the preclinical phase with longer periods of follow-up.

KEY POINTS

- Risk scores have been developed to predict the time to stage 3 among subjects with multiple islet autoantibodies.

PRIMARY PREVENTION

Primary prevention trials would target genetic high-risk groups such as infants with HLA high risk,[174] such as the POInT study,[5] or a positive family history, such as Trial to Reduce IDDM in the Genetically At Risk (TRIGR) study.[175] While the POInT study is ongoing, the randomized, placebo-controlled, double-blind trial testing the effects of cow's milk or hydrolyzed casein formulae following 6 to 8 months of breastfeeding failed to show a difference in the risk for a first islet autoantibody.[175] The NIP study is using oral docosahexaenoic acid, a fatty acid thought to prevent islet autoimmunity, given to the mothers as a supplement. Infants will be followed up until 9 years of age. The gluten-free diet trial, PREVFIN, tested the possible effect of eliminating gluten from the diet on the risk of T1D. It ended in 2002 with no observed effects on islet autoantibodies or disease progression.

KEY POINTS

- Islet autoantibodies are endpoints in ongoing primary prevention studies such as oral insulin to prevent the appearance of insulin autoantibody first or GAD65 autoantibody first in children with increased genetic risk for type 1 diabetes.

SECONDARY PREVENTION

Secondary prevention strategies recruit children with high HLA genetic risk with positive autoantibodies and children and young adults with multiple autoantibodies in either stage 1 or 2. The aim is to delay and suppress β cell destruction by aggressive immunity. Antigen-based secondary prevention trials using insulin as a specific antigen have been carried out. The Diabetes Prevention Trial-1 (DPT-1) was a large randomized trial of treatment efficacy that had two arms: parenteral insulin (high-risk subject, no placebo, completed in 2002) and oral insulin (moderate-risk subject, placebo-controlled, completed in 2005) showed no effect in either arm.[176,177] A subanalysis of the oral arm gave strong evidence on reducing the incidence of T1D in relatives with insulin autoimmunity (IAA ≥80 U). A repeat study in autoantibody-positive relatives of patients with T1D, oral insulin at a dose of 7.5 mg/d, compared with placebo, did not delay or prevent the development of T1D over 2.7 years.[167] Nicotinamide (vitamin B_3) was used orally as a non–antigen-specific agent in the European Nicotinamide Diabetes Intervention Trial (ENDIT) but showed no effect on T1D outcome.[178] The Diabetes Prediction and Prevention Project (DIPP) was a randomized, placebo-controlled trial implemented in Finland;

it showed no effect in high–genetic risk children of intranasal insulin on the prevention of T1D.[179] GAD-alum as a subcutaneous prime and boost injection was safe in stage 1 and 2 children but did not affect progression to T1D.[180] Anti-CD3 (teplizumab) was used in a secondary prevention trial, and fewer participants in the teplizumab group compared with the placebo group were diagnosed with T1D.[169] The protective effect was predominantly observed in HLA DR3–negative, HLA DR4–positive subjects and among those negative for ZnT8A. Other trials that used cyclosporine, vitamin E, or intensive insulin treatment in combination with nicotinamide were without effects.

KEY POINTS

- Secondary prevention strategies recruit children and young adults with multiple autoantibodies in either stage 1 or stage 2. Anti-CD3 (teplizumab) had fewer participants with a type 1 diabetes diagnosis than the placebo group. The protective effect was predominantly observed in human leukocyte antigen (HLA) DR3–negative, HLA DR4-positive subjects and among those negative for zinc transporter 8 autoantibody.

INTERVENTION

Intervention or tertiary prevention trials target newly diagnosed T1D patients, aiming at preserving the residual β cell mass and enhancing β cell functionality and regeneration. Like prevention trials, intervention trials can also be either antigen-specific or non–antigen-specific. Immunomodulating injections with specific islet autoantigens have been adopted by several trials.

- The altered peptide ligand (APL) insulin B:9-23, NBI-6024, at repeated doses of 0.1, 0.5, or 1.0 mg did not improve or maintain β cell function.[181]
- DiaPep277 was an immunomodulatory HSP60 peptide given subcutaneously to subjects with new-onset T1D. All publications were retracted.[182]
- GAD65 formulated with alum has been used in randomized, placebo-controlled immunomodulation trials in LADA[183] and 10- to 18-year-old T1D patients in stage 2[184,185] and stage 3,[186] with possible preservation of residual β cell function. Recent analyses suggested that GAD-alum responders are to be found among HLA DR3–positive patients.[187,188] Intranodal injections as an alternative to subcutaneous injections are being attempted in controlled clinical trials.[189]
- A DNA plasmid encoding proinsulin (BHT-3021) was tested in 80 T1D patients to test if the treatment would reduce CD8+ T cells in the peripheral blood. C-peptide levels were preserved, and CD8+ T-cells declined.[190]

The non–antigen-specific approach has been used in numerous interventional studies. Randomized, placebo-controlled trials with oral cyclosporin-A showed reduced loss of residual C-peptide and insulin dosages; however, the effects were transient, and side effects outweighed the benefits.[191-193] Although azathioprine trials showed no effects when used alone, its combination with prednisone showed partial, temporary remission of new-onset T1D.[194] Antithymocyte globulin, used in organ transplantation, showed reduction in insulin requirements when combined with prednisone, but severe, transient thrombocytopenia emerged as a major side effect.[195] The CD3ε monoclonal antibody otelixizumab preserved β cell function but showed dose-dependent Epstein–Barr virus activation.[196] The CD3 monoclonal antibody teplizumab, which suppress T-cell activity by blocking T-cell proliferation and differentiation, preserved insulin production and reduced exogenous insulin in some patients with new-onset T1D.[197,198] Costimulation modulation with abatacept preserved β cell function and improved HbA1c in recent-onset T1D. The effect was sustained for at least 1 year, or 3 years from T1D diagnosis.[199]

The anti-CD20 agent rituximab is a monoclonal antibody that reduces blood B-cells by blocking the CD20 receptors and showed transient preservation of residual β cell function.[200,201] Combination therapies have been rare. However, anti–IL-21 and liraglutide in combination only could preserve β cell function in recently diagnosed type 1 diabetes.[202]

Tertiary intervention includes both whole-organ pancreatic transplantation and isolated islet transplantation. Both approaches have proved successful in terms of restoring glycemic levels and insulin independence, but isolated islet transplantation had lesser glycemic control, transient effects, increased morbidity, and higher costs.[203]

SUMMARY AND FUTURE DIRECTIONS

The islet β cells are the specific targets in an autoimmune process that eventually leads to the clinical onset of T1D. The etiology is related to HLA DR-DQ genetic factors, which are necessary but not sufficient for disease development. The environmental etiology is related to enterovirus infection in early childhood, resulting in two endotypes. In HLA DR4-DQ8 subjects, prolonged shedding of enterovirus B may induce autoantibody biomarkers against insulin (IAA). In HLA DR3-DQ2 subjects, other gastrointestinal viruses may initiate autoantibodies against glutamic acid decarboxylase (GADA). In 60% of children, initiation of islet autoimmune pathogenesis is followed within 1 year by a second autoantibody, including IAA, GADA, IA-2A, or ZnT8A. Two or more islet autoantibodies predict clinical onset in 70% of subjects within 10 years. The pathogenesis may be related to several non-HLA genetic factors, as well as to environmental factors. The β cell destruction by cellular immunity (insulitis) appears late and is related to prediabetic metabolic disturbances prior to overt diabetes. The aim of future prevention and intervention studies in humans is to test whether antigen-specific immunosuppression or other immunomodulatory therapies in T1D-susceptible individuals will prevent the loss of pancreatic β cells.

Acknowledgements

Studies in the authors' laboratories were supported by the National Institutes of Health (1UO1 DK063861), the Swedish Research Council (2020-01537), Diabetesfonden (DIA2017-262), Juvenile Diabetes Research Foundation, the Leona M. and Harry B. Helmsley Charitable Trust, the Ingabritt and Arne Lundbergs Scientific Foundation, the European Union Interreg Foundation, the SUS Fund, the Swedish Research Council (Strategic Research Area Exodiab, Dnr 2009-1039), and the Swedish Foundation for Strategic Research (Dnr IRC15-0067).

REFERENCES

1. American Diabetes Association. Classification and diagnosis of diabetes. *Diabetes Care*. 2020;43:S14–S31.
2. Diaz-Valencia PA, Bougneres P, Valleron AJ. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. *BMC Publ Health*. 2015;15:255.
3. Winter WE, Schatz DA. Autoimmune markers in diabetes. *Clin Chem*. 2011;57(2):168–175.
4. Katsarou A, Gudbjornsdottir S, Rawshani A, et al. Type 1 diabetes mellitus. *Nat Rev Dis Primers*. 2017;3:17016.
5. Ziegler AG, Achenbach P, Berner R, et al. Oral insulin therapy for primary prevention of type 1 diabetes in infants with high genetic risk: the GPPAD-POInT (global platform for the prevention of autoimmune diabetes primary oral insulin trial) study protocol. *BMJ Open*. 2019;9:e028578.

6. Insel RA, Dunne JL, Atkinson MA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*. 2015;38:1964–1974.
7. Shields BM, Peters JL, Cooper C, et al. Can clinical features be used to differentiate type 1 from type 2 diabetes? A systematic review of the literature. *BMJ Open*. 2015;5:e009088.
8. Dabelea D, Sauder KA, Jensen ET, et al. Twenty years of pediatric diabetes surveillance: what do we know and why it matters. *Ann N Y Acad Sci*. 2021;1495:99–120.
9. Torres JM, Cox NJ, Philipson LH. Genome wide association studies for diabetes: perspective on results and challenges. *Pediatr Diabetes*. 2013;14:90–96.
10. Zhang H, Colclough K, Gloyn AL, et al. Monogenic diabetes: a gateway to precision medicine in diabetes. *J Clin Invest*. 2021;131:e142244.
11. Ogurtsova K, da Rocha Fernandes JD. Huang yet al. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract*. 2017;128:40–50.
12. Imperatore G, Mayer-Davis EJ, Orchard TJ, et al. Prevalence and incidence of type 1 diabetes among children and adults in the United States and comparison with non-U.S. countries. In: Cowie CC, Casagrande SS, Menke A, et al, eds. Diabetes in America. 3rd ed. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (US); 2018.
13. Patterson C, Guariguata L, Dahlquist G, et al. Diabetes in the young - a global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Res Clin Pract*. 2014;103:161–175.
14. Mobasseri M, Shirmohammadi M, Amiri T, et al. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health Promot Perspect*. 2020;10:98–115.
15. Patterson CC, Gyürüs E, Rosenbauer J, et al. Trends in childhood type 1 diabetes incidence in Europe during 1989-2008: evidence of non-uniformity over time in rates of increase. *Diabetologia*. 2012;55:2142–2147.
16. Fsadni P, Fsadni C, Fava S, et al. Correlation of worldwide incidence of type 1 diabetes (DiaMond) with prevalence of asthma and atopic eczema (ISAAC). *Clin Respir J*. 2012;6:18–25.
17. Samuelsson U, Westerberg L, Aakesson K, et al. Geographical variation in the incidence of type 1 diabetes in the Nordic countries: a study within NordicDiabKids. *Pediatr Diabetes*. 2020;21:259–265.
18. Peet A, Hämäläinen AM, Kool P, et al. Early postnatal growth in children with HLA-conferred susceptibility to type 1 diabetes. *Diabetes Metab Res Rev*. 2014;30:60–68.
19. Harjutsalo V, Sjöberg L, Tuomilehto J. Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet*. 2008;371:1777–1782.
20. Tokunaga K, Ohashi J, Bannai M, et al. Genetic link between Asians and native Americans: evidence from HLA genes and haplotypes. *Hum Immunol*. 2001;62:1001–1008.
21. Triolo TM, Fouts A, Pyle L, et al. Identical and nonidentical twins: risk and factors involved in development of islet autoimmunity and type 1 diabetes. *Diabetes Care*. 2019;42:192–199.
22. Shaltout AA, Wake D, Thanaraj TA, et al. Incidence of type 1 diabetes has doubled in Kuwaiti children 0-14 years over the last 20 years. *Pediatr Diabetes*. 2017;18:761–766.
23. Tuomilehto J, Ogle GD, Lund-Blix NA, et al. Update on worldwide trends in occurrence of childhood type 1 diabetes in 2020. *Pediatr Endocrinol Rev*. 2020;17:198–209.
24. Knip M, Veijola R, Virtanen SM, et al. Environmental triggers and determinants of type 1 diabetes. *Diabetes*. 2005;54:S125–S136.
25. Delli AJ, Vaziri-Sani F, Lindblad B, et al. Zinc transporter 8 autoantibodies and their association with SLC30A8 and HLA-DQ genes differ between immigrant and Swedish patients with newly diagnosed type 1 diabetes in the Better Diabetes Diagnosis study. *Diabetes*. 2012;61:2556–2564.
26. Zimmet PZ, Tuomi T, Mackay IR, et al. Latent autoimmune diabetes mellitus in adults (LADA): the role of antibodies to glutamic acid decarboxylase in diagnosis and prediction of insulin dependency. *Diabet Med*. 1994;11:299–303.
27. Kobayashi T, Tanaka S, Okubo M, et al. Unique epitopes of glutamic acid decarboxylase autoantibodies in slowly progressive type 1 diabetes. *J Clin Endocrinol Metab*. 2003;88:4768–4775.
28. Kobayashi T, Nishida Y, Tanaka S, et al. Pathological changes in the pancreas of fulminant type 1 diabetes and slowly progressive insulin-dependent diabetes mellitus (SPIDDM): innate immunity in fulminant type 1 diabetes and SPIDDM. *Diabetes Metab Res Rev*. 2011;27:965–970.
29. Rolandsson O, Palmer JP. Latent autoimmune diabetes in adults (LADA) is dead: long live autoimmune diabetes. *Diabetologia*. 2010;53:1250–1253.
30. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab*. 2008;4:200–213.
31. Slingerland AS, Hattersley AT. Mutations in the Kir6.2 subunit of the KATP channel and permanent neonatal diabetes: new insights and new treatment. *Ann Med*. 2005;37:186–195.
32. Rubio-Cabezas O, Patch AM, Minton JA, et al. Wolcott-Rallison syndrome is the most common genetic cause of permanent neonatal diabetes in consanguineous families. *J Clin Endocrinol Metab*. 2009;94:4162–4170.
33. Hattersley AT, Greeley SAW, Polak M, et al. ISPAD Clinical Practice Consensus Guidelines 2018: the diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes*. 2018;19:47–63.
34. Larsson HE, Hansson G, Carlsson A, et al. Children developing type 1 diabetes before 6 years of age have increased linear growth independent of HLA genotypes. *Diabetologia*. 2008;51:1623–1630.
35. Blom L, Persson LA, Dahlquist G. A high linear growth is associated with an increased risk of childhood diabetes mellitus. *Diabetologia*. 1992;35:528–533.
36. Karvonen M, Pitkäniemi M, Pitkäniemi J, et al. Sex difference in the incidence of insulin-dependent diabetes mellitus: an analysis of the recent epidemiological data. World Health Organization DIAMOND Project Group. *Diabetes Metab Rev*. 1997;13:275–291.
37. Rawshani A, Landin-Olsson M, Svensson AM, et al. The incidence of diabetes among 0-34 year olds in Sweden: new data and better methods. *Diabetologia*. 2014;57:1375–1381.
38. Soltesz G, Patterson CC, Dahlquist G, et al. Worldwide childhood type 1 diabetes incidence—what can we learn from epidemiology? *Pediatr Diabetes*. 2007;8:6–14.
39. Todd JA, Farrall M. Panning for gold: genome-wide scanning for linkage in type 1 diabetes. *Hum Mol Genet*. 1996;5:1443–1448.
40. Akesson K, Nyström L, Färnkvist L, et al. Increased risk of diabetes among relatives of female insulin-treated patients diagnosed at 15-34 years of age. *Diabet Med*. 2005;22:1551–1557.
41. Gepts W. Pathologic anatomy of the pancreas in juvenile diabetes mellitus. *Diabetes*. 1965;14:619–633.
42. Nerup J, Andersen OO, Bendixen G, et al. Anti-pancreatic, cellular hypersensitivity in diabetes mellitus. Antigenic activity of fetal calf pancreas and correlation with clinical type of diabetes. *Acta Allergol*. 1973;28:223–230.
43. Krogvold L, Wiberg A, Edwin B, et al. Insulitis and characterisation of infiltrating T cells in surgical pancreatic tail resections from patients at onset of type 1 diabetes. *Diabetologia*. 2016;59:492–501.
44. Rodriguez-Calvo T, Krogvold L, Amirian N, et al. One in ten CD8. *Diabetes*. 2021;70:752–758.
45. Lernmark A, Klöppel G, Stenger D, et al. Heterogeneity of islet pathology in two infants with recent onset diabetes mellitus. *Virchows Arch*. 1995;425:631–640.
46. Bottazzo GF, Florin-Christensen A, Doniach D. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. *Lancet*. 1974;2:1279–1283.
47. MacCuish AC, Irvine WJ, Barnes EW, et al. Antibodies to pancreatic islet cells in insulin-dependent diabetics with coexistent autoimmune disease. *Lancet*. 1974;2:1529–1531.
48. Lernmark A, Freedman ZR, Hofmann C, et al. Islet-cell-surface antibodies in juvenile diabetes mellitus. *N Engl J Med*. 1978;299:375–380.
49. Dobersen MJ, Scharff JE, Ginsberg-Fellner F, et al. Cytotoxic autoantibodies to beta cells in the serum of patients with insulin-dependent diabetes mellitus. *N Engl J Med*. 1980;303:1493–1498.
50. Baekkeskov S, Nielsen JH, Marner B, et al. Autoantibodies in newly diagnosed diabetic children immunoprecipitate human pancreatic islet cell proteins. *Nature*. 1982;298:167–169.

51. Baekkeskov S, Aanstoot HJ, Christgau S, et al. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature*. 1990;347:151–156.
52. Karlsen AE, Hagopian WA, Grubin CE, et al. Cloning and primary structure of a human islet isoform of glutamic acid decarboxylase from chromosome 10. *Proc Natl Acad Sci U S A*. 1991;88:8337–8341.
53. Palmer JP, Asplin CM, Clemons P, et al. Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science*. 1983;222:1337–1339.
54. Lan MS, Lu J, Goto Y, Notkins AL. Molecular cloning and identification of a receptor-type protein tyrosine phosphatase, IA-2, from human insulinoma. *DNA Cell Biol*. 1994;13:505–514.
55. Wenzlau JM, Juhl K, Yu L, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A*. 2007;104:17040–17045.
56. Lampasona V, Pittman DL, Williams AJ, et al. Islet Autoantibody Standardization Program 2018 Workshop: interlaboratory comparison of glutamic acid decarboxylase autoantibody assay performance. *Clin Chem*. 2019;65:1141–1152.
57. Ziegler AG, Hummel M, Schenker M, et al. Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. *Diabetes*. 1999;48:460–468.
58. Hahl J, Simell T, Ilonen J, et al. Costs of predicting IDDM. *Diabetologia*. 1998;41:79–85.
59. Nejentsev S, Sjöroos M, Soukka T, et al. Population-based genetic screening for the estimation of Type 1 diabetes mellitus risk in Finland: selective genotyping of markers in the HLA-DQB1, HLA-DQA1 and HLA-DRB1 loci. *Diabet Med*. 1999;16:985–992.
60. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia*. 1996;39:807–812.
61. Rewers M, Norris JM, Eisenbarth GS, et al. Beta-cell autoantibodies in infants and toddlers without IDDM relatives: diabetes autoimmunity study in the young (DAISY). *J Autoimmun*. 1996;9:405–410.
62. Ziegler AG, Bonifacio E; BABYDIAB-BABYDIET Study Group. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia*. 2012;55:1937–1943.
63. Ilonen J, Hammais A, Laine AP, et al. Patterns of β-cell autoantibody appearance and genetic associations during the first years of life. *Diabetes*. 2013;62:3636–3640.
64. Hagopian WA, Lernmark A, Rewers MJ, et al. TEDDY--The environmental determinants of diabetes in the young: an observational clinical trial. *Ann N Y Acad Sci*. 2006;1079:320–326.
65. Hagopian WA, Erlich H, Lernmark A, et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. *Pediatr Diabetes*. 2011;12:733–743.
66. Krischer JP, Lynch KF, Schatz DA, et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia*. 2015;58:980–987.
67. Krischer JP, Lynch KF, Lernmark Å, et al. Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. *Diabetes Care*. 2017;40:1194–1202.
68. Lynch KF, Lee HS, Törn C, et al. Gestational respiratory infections interacting with offspring HLA and CTLA-4 modifies incident β-cell autoantibodies. *J Autoimmun*. 2018;86:93–103.
69. Vehik K, Lynch KF, Wong MC, et al. Prospective virome analyses in young children at increased genetic risk for type 1 diabetes. *Nat Med*. 2019;25:1865–1872.
70. Li Q, Parikh H, Butterworth MD, et al. Longitudinal metabolome-wide signals prior to the appearance of a first islet autoantibody in children participating in the TEDDY study. *Diabetes*. 2020;69:465–476.
71. Li Q, Liu X, Yang J, et al. Plasma metabolome and circulating vitamins stratified onset age of an initial islet autoantibody and progression to type 1 diabetes: the TEDDY study. *Diabetes*. 2021;70:282–292.
72. Pociot F, Lernmark Å. Genetic risk factors for type 1 diabetes. *Lancet*. 2016;387:2331–2339.
73. Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. *N Engl J Med*. 2009;360:1646–1654.
74. Grant SFA, Wells AD, Rich SS. Next steps in the identification of gene targets for type 1 diabetes. *Diabetologia*. 2020;63:2260–2269.
75. Johnson SB, Lynch KF, Roth R, et al. First-appearing islet autoantibodies for type 1 diabetes in young children: maternal life events during pregnancy and the child's genetic risk. *Diabetologia*. 2021;64:591–602.
76. Krischer JP, Liu X, Vehik K, et al. Predicting islet cell autoimmunity and type 1 diabetes: an 8-year TEDDY study progress report. *Diabetes Care*. 2019;42:1051–1060.
77. Hyöty H. Viruses in type 1 diabetes. *Pediatr Diabetes*. 2016;17:56–64.
78. Harris HF. A case of diabetes mellitus quickly following mumps on the pathological alterations of salivary glands, closely resembling those found in pancreas, in a case of diabetes mellitus. *Boston Med Surg J*. 1899;140:465–469.
79. van der Werf N, Kroese FG, Rozing J, et al. Viral infections as potential triggers of type 1 diabetes. *Diabetes Metab Res Rev*. 2007;23:169–183.
80. Sioofy-Khojine AB, Lehtonen J, Nurminen N, et al. Coxsackievirus B1 infections are associated with the initiation of insulin-driven autoimmunity that progresses to type 1 diabetes. *Diabetologia*. 2018;61:1193–1202.
81. Thernlund GM, Dahlquist G, Hansson K, et al. Psychological stress and the onset of IDDM in children. *Diabetes Care*. 1995;18:1323–1329.
82. Lernmark B, Lynch K, Lernmark A. Cord blood islet autoantibodies are related to stress in the mother during pregnancy. *Ann N Y Acad Sci*. 2006;1079:345–349.
83. Lundgren M, Ellström K, Elding Larsson H, et al. Influence of early-life parental severe life events on the risk of type 1 diabetes in children: the DiPiS study. *Acta Diabetol*. 2018;55:797–804.
84. Sepa A, Frodi A, Ludvigsson J. Mothers' experiences of serious life events increase the risk of diabetes-related autoimmunity in their children. *Diabetes Care*. 2005;28:2394–2399.
85. Patterson CC, Dahlquist G, Soltész G, et al. Maternal age and risk of type 1 diabetes in children. Relative risks by maternal age are biased. *BMJ*. 2001;322:1489–1490;author reply 90–91.
86. Waernbaum I, Dahlquist G, Lind T. Perinatal risk factors for type 1 diabetes revisited: a population-based register study. *Diabetologia*. 2019;62:1173–1184.
87. Rešić Lindehammer S, Honkanen H, Nix WA, et al. Seroconversion to islet autoantibodies after enterovirus infection in early pregnancy. *Viral Immunol*. 2012;25:254–261.
88. Lindehammer SR, Hansson I, Midberg B, et al. Seroconversion to islet autoantibodies between early pregnancy and delivery in non-diabetic mothers. *J Reprod Immunol*. 2011;88:72–79.
89. La Torre D, Seppänen-Laakso T, Larsson HE, et al. Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. *Diabetes*. 2013;62:3951–3956.
90. Oresic M, Gopalacharyulu P, Mykkänen J, et al. Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. *Diabetes*. 2013;62:3268–3274.
91. Knip M, Åkerblom HK, Becker D, et al. Hydrolyzed infant formula and early β-cell autoimmunity: a randomized clinical trial. *J Am Med Assoc*. 2014;311:2279–2287.
92. Uusitalo U, Lee HS, Andrén Aronsson C, et al. Early infant diet and islet autoimmunity in the TEDDY study. *Diabetes Care*. 2018;41:522–530.
93. Uusitalo U, Liu X, Yang J, et al. Association of early exposure of probiotics and islet autoimmunity in the TEDDY study. *JAMA Pediatr*. 2016;170:20–28.
94. Beyerlein A, Liu X, Uusitalo UM, et al. Dietary intake of soluble fiber and risk of islet autoimmunity by 5 y of age: results from the TEDDY study. *Am J Clin Nutr*. 2015;102:345–352.
95. Silvis K, Aronsson CA, Liu X, et al. Maternal dietary supplement use and development of islet autoimmunity in the offspring: TEDDY study. *Pediatr Diabetes*. 2019;20:86–92.
96. Karam JH, Lewitt PA, Young CW, et al. Insulinopenic diabetes after rodenticide (Vacor) ingestion: a unique model of acquired diabetes in man. *Diabetes*. 1980;29:971–978.
97. Vehik K, Fiske SW, Logan CA, et al. Methods, quality control and specimen management in an international multicentre investigation of type 1 diabetes: TEDDY. *Diabetes Metab Res Rev*. 2013;29:557–567.

98. Hansen B, Lernmark A, Nielsen JH, et al. New approaches to therapy and diagnosis of diabetes. *Diabetologia*. 1982;22:61–67.
99. Larsson HE, Lynch K, Lernmark B, et al. Diabetes-associated HLA genotypes affect birthweight in the general population. *Diabetologia*. 2005;48:1484–1491.
100. Elding Larsson H, Vehik K, Haller MJ, et al. Growth and risk for islet autoimmunity and progression to type 1 diabetes in early childhood: the environmental determinants of diabetes in the young study. *Diabetes*. 2016;65:1988–1995.
101. Liu X, Vehik K, Huang Y, et al. Distinct growth phases in early life associated with the risk of type 1 diabetes: the TEDDY study. *Diabetes Care*. 2020;43:556–562.
102. Nucci AM, Virtanen SM, Cuthbertson D, et al. Growth and development of islet autoimmunity and type 1 diabetes in children genetically at risk. *Diabetologia*. 2021;64:826–835.
103. Hyppönen E, Kenward MG, Virtanen SM, et al. Infant feeding, early weight gain, and risk of type 1 diabetes. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes Care*. 1999;22:1961–1965.
104. Vehik K, Bonifacio E, Lernmark Å, et al. Hierarchical order of distinct autoantibody spreading and progression to type 1 diabetes in the TEDDY study. *Diabetes Care*. 2020;43:2066–2073.
105. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *J Am Med Assoc*. 2013;309:2473–2479.
106. Elding Larsson H, Vehik K, Gesualdo P, et al. Children followed in the TEDDY study are diagnosed with type 1 diabetes at an early stage of disease. *Pediatr Diabetes*. 2014;15:118–126.
107. Steck AK, Vehik K, Bonifacio E, et al. Predictors of progression from the appearance of islet autoantibodies to early childhood diabetes: the Environmental Determinants of Diabetes in the Young (TEDDY). *Diabetes Care*. 2015;38:808–813.
108. Sosenko JM, Skyler JS, Palmer JP, et al. The prediction of type 1 diabetes by multiple autoantibody levels and their incorporation into an autoantibody risk score in relatives of type 1 diabetic patients. *Diabetes Care*. 2013;36:2615–2620.
109. Redondo MJ, Geyer S, Steck AK, et al. A type 1 diabetes genetic risk score predicts progression of islet autoimmunity and development of type 1 diabetes in individuals at risk. *Diabetes Care*. 2018;41:1887–1894.
110. Beyerlein A, Bonifacio E, Vehik K, et al. Progression from islet autoimmunity to clinical type 1 diabetes is influenced by genetic factors: results from the prospective TEDDY study. *J Med Genet*. 2019;56:602–605.
111. Ferrat LA, Vehik K, Sharp SA, et al. A combined risk score enhances prediction of type 1 diabetes among susceptible children. *Nat Med*. 2020;26:1247–1255.
112. Andersson C, Vaziri-Sani F, Delli A, et al. Triple specificity of ZnT8 autoantibodies in relation to HLA and other islet autoantibodies in childhood and adolescent type 1 diabetes. *Pediatr Diabetes*. 2013;14:97–105.
113. Andersson C, Kolmodin M, Ivarsson SA, et al. Islet cell antibodies (ICA) identify autoimmunity in children with new onset diabetes mellitus negative for other islet cell antibodies. *Pediatr Diabetes*. 2014;15:336–344.
114. Gleichmann H, Bottazzo GF. Progress toward standardization of cytoplasmic islet cell-antibody assay. *Diabetes*. 1987;36:578–584.
115. Bingley PJ, Bonifacio E, Mueller PW. Diabetes Antibody Standardization Program: first assay proficiency evaluation. *Diabetes*. 2003;52:1128–1136.
116. Lampasona V, Schlosser M, Mueller PW, et al. Diabetes antibody standardization program: first proficiency evaluation of assays for autoantibodies to zinc transporter 8. *Clin Chem*. 2011;57:1693–1702.
117. Graham J, Hagopian WA, Kockum I, et al. Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. *Diabetes*. 2002;51:1346–1355.
118. Wester A, Skärstrand H, Lind A, et al. An increased diagnostic sensitivity of truncated GAD65 autoantibodies in type 1 diabetes may be related to HLA-DQ8. *Diabetes*. 2017;66:735–740.
119. Li L, Hagopian WA, Brashear HR, et al. Identification of autoantibody epitopes of glutamic acid decarboxylase in stiff-man syndrome patients. *J Immunol*. 1994;152:930–934.
120. Ziegler AG, Kick K, Bonifacio E, et al. Yield of a public health screening of children for islet autoantibodies in Bavaria, Germany. *J Am Med Assoc*. 2020;323:339–351.
121. Castaño L, Ziegler AG, Ziegler R, et al. Characterization of insulin autoantibodies in relatives of patients with type I diabetes. *Diabetes*. 1993;42:1202–1209.
122. Brooks-Worrell BM, Nielson D, Palmer JP. Insulin autoantibodies and insulin antibodies have similar binding characteristics. *Proc Assoc Am Physicians*. 1999;111:92–96.
123. Schlosser M, Mueller PW, Törn C, et al. Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies. *Diabetologia*. 2010;53:2611–2620.
124. Payton MA, Hawkes CJ, Christie MR. Relationship of the 37,000- and 40,000-M(r) tryptic fragments of islet antigens in insulin-dependent diabetes to the protein tyrosine phosphatase-like molecule IA-2 (ICA512). *J Clin Invest*. 1995;96:1506–1511.
125. Hawkes CJ, Wasmeier C, Christie MR, et al. Identification of the 37-kDa antigen in IDDM as a tyrosine phosphatase-like protein (phogrin) related to IA-2. *Diabetes*. 1996;45:1187–1192.
126. Hatfield EC, Hawkes CJ, Payton MA, et al. Cross reactivity between IA-2 and phogrin/IA-2beta in binding of autoantibodies in IDDM. *Diabetologia*. 1997;40:1327–1333.
127. Richardson CC, McLaughlin KA, Morgan D, et al. Influence of HLA-DR and -DQ alleles on autoantibody recognition of distinct epitopes within the juxtamembrane domain of the IA-2 autoantigen in type 1 diabetes. *Diabetologia*. 2016;59:334–340.
128. Wijesekara N, Chimienti F, Wheeler MB. Zinc, a regulator of islet function and glucose homeostasis. *Diabetes Obes Metab*. 2009;11:202–214.
129. Wenzlau JM, Liu Y, Yu L, et al. A common nonsynonymous single nucleotide polymorphism in the SLC30A8 gene determines ZnT8 autoantibody specificity in type 1 diabetes. *Diabetes*. 2008;57:2693–2697.
130. Yu L, Boulware DC, Beam CA, et al. Zinc transporter-8 autoantibodies improve prediction of type 1 diabetes in relatives positive for the standard biochemical autoantibodies. *Diabetes Care*. 2012;35:1213–1218.
131. Arden SD, Roep BO, Neophytou PI, et al. Imogen 38: a novel 38-kD islet mitochondrial autoantigen recognized by T cells from a newly diagnosed type 1 diabetic patient. *J Clin Invest*. 1996;97:551–561.
132. Geluk A, van Meijgaarden KE, Roep BO, et al. Altered peptide ligands of islet autoantigen Imogen 38 inhibit antigen specific T cell reactivity in human type-1 diabetes. *J Autoimmun*. 1998;11:353–361.
133. Hirai H, Miura J, Hu Y, et al. Selective screening of secretory vesicle-associated proteins for autoantigens in type 1 diabetes: VAMP2 and NPY are new minor autoantigens. *Clin Immunol*. 2008;127:366–374.
134. Skärstrand H, Vaziri-Sani F, Delli AJ, et al. Neuropeptide Y is a minor autoantigen in newly diagnosed type 1 diabetes patients. *Pediatr Diabetes*. 2015;16:621–628.
135. Aanstoot HJ, Kang SM, Kim J, et al. Identification and characterization of glima 38, a glycosylated islet cell membrane antigen, which together with GAD65 and IA2 marks the early phases of autoimmune response in type 1 diabetes. *J Clin Invest*. 1996;97:2772–2783.
136. McLaughlin KA, Richardson CC, Ravishankar A, et al. Identification of tetraspanin-7 as a target of autoantibodies in type 1 diabetes. *Diabetes*. 2016;65:1690–1698.
137. Eugster A, Kraus G, Lidzba V, et al. Cytoplasmic ends of tetraspanin 7 harbour epitopes recognised by autoantibodies in type 1 diabetes. *Diabetologia*. 2019;62:805–810.
138. Sims TJ, Lernmark A, Mancl LA, et al. Serum IgG to heat shock proteins and Porphyromonas gingivalis antigens in diabetic patients with periodontitis. *J Clin Periodontol*. 2002;29:551–562.
139. Huurman VA, Decochez K, Mathieu C, et al. Therapy with the hsp60 peptide DiaPep277 in C-peptide positive type 1 diabetes patients. *Diabetes Metab Res Rev*. 2007;23:269–275.
140. Geluk A, van Meijgaarden KE, Schloot NC, et al. HLA-DR binding analysis of peptides from islet antigens in IDDM. *Diabetes*. 1998;47:1594–1601.
141. Tree TI, O'Byrne D, Tremble JM, et al. Evidence for recognition of novel islet T cell antigens by granule-specific T cell lines from new onset type 1 diabetic patients. *Clin Exp Immunol*. 2000;121:100–105.

142. Castaño L, Russo E, Zhou L, et al. Identification and cloning of a granule autoantigen (carboxypeptidase-H) associated with type I diabetes. *J Clin Endocrinol Metab*. 1991;73:1197–1201.
143. Chang YH, Hwang J, Shang HF, et al. Characterization of human DNA topoisomerase II as an autoantigen recognized by patients with IDDM. *Diabetes*. 1996;45:408–414.
144. Pupilli C, Giannini S, Marchetti P, et al. Autoantibodies to CD38 (ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase) in Caucasian patients with diabetes: effects on insulin release from human islets. *Diabetes*. 1999;48:2309–2315.
145. Inman LR, McAllister CT, Chen L, et al. Autoantibodies to the GLUT-2 glucose transporter of beta cells in insulin-dependent diabetes mellitus of recent onset. *Proc Natl Acad Sci U S A*. 1993;90:1281–1284.
146. Pehuet-Figoni M, Alvarez F, Bach JF, et al. Autoantibodies in recent onset type-1 diabetic patients to a Mr 60K microsomal hepatic protein: new evidence for autoantibodies to the type-2 glucose transporter. *Clin Exp Immunol*. 2000;122:164–169.
147. Steinbrenner H, Lohmann T, Ostendorf B, et al. Autoantibodies to ICA12 (SOX-13) are not specific for Type I diabetes. *Diabetologia*. 2000;43:1381–1384.
148. Lampasona V, Scirpoli M, Bosi E, et al. ICA12(SOX13) autoantibodies are unlikely to be a useful marker for pre-clinical type I diabetes. *Diabetologia*. 2001;44:267.
149. Williams AJ, Thrower SL, Sequeiros IM, et al. Pancreatic volume is reduced in adult patients with recently diagnosed type 1 diabetes. *J Clin Endocrinol Metab*. 2012;97:E2109–E2113.
150. Landin-Olsson M, Borgström A, Blom L, et al. Immunoreactive trypsin(ogen) in the sera of children with recent-onset insulin-dependent diabetes and matched controls. The Swedish Childhood Diabetes Group. *Pancreas*. 1990;5:241–247.
151. Moles KW, Kerr JI, Armstrong E, et al. Serum concentrations of trypsin-like immunoreactivity and pancreatic isoamylase in insulin dependent diabetic patients. *Pancreas*. 1988;3:135–139.
152. Gepts W, Lecompte PM. The pancreatic islets in diabetes. *Am J Med*. 1981;70:105–115.
153. Pipeleers D, Ling Z. Pancreatic beta cells in insulin-dependent diabetes. *Diabetes Metab Rev*. 1992;8:209–227.
154. Richardson SJ, Rodriguez-Calvo T, Gerling IC, et al. Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. *Diabetologia*. 2016;59:2448–2458.
155. Lundberg M, Krogvold L, Kuric E, et al. Expression of interferon-stimulated genes in insulitic pancreatic islets of patients recently diagnosed with type 1 diabetes. *Diabetes*. 2016;65:3104–3110.
156. In't Veld P, Lievens D, De Grijse J, et al. Screening for insulitis in adult autoantibody-positive organ donors. *Diabetes*. 2007;56:2400–2404.
157. Jansen A, van Hagen M, Drexhage HA. Defective maturation and function of antigen-presenting cells in type 1 diabetes. *Lancet*. 1995;345:491–492.
158. Plesner A, Greenbaum CJ, Gaur LK, et al. Macrophages from high-risk HLA-DQB1*0201/*0302 type 1 diabetes mellitus patients are hypersensitive to lipopolysaccharide stimulation. *Scand J Immunol*. 2002;56:522–529.
159. Lawley TJ, Hall RP, Fauci AS, et al. Defective Fc-receptor functions associated with the HLA-B8/DRw3 haplotype: studies in patients with dermatitis herpetiformis and normal subjects. *N Engl J Med*. 1981;304:185–192.
160. Graham KL, Sutherland RM, Mannering SI, et al. Pathogenic mechanisms in type 1 diabetes: the islet is both target and driver of disease. *Rev Diabet Stud*. 2012;9:148–168.
161. Novak EJ, Masewicz SA, Liu AW, et al. Activated human epitope-specific T cells identified by class II tetramers reside within a CD4high, proliferating subset. *Int Immunol*. 2001;13:799–806.
162. Uibo R, Lernmark A. GAD65 autoimmunity-clinical studies. *Adv Immunol*. 2008;100:39–78.
163. Meierhoff G, Ott PA, Lehmann PV, et al. Cytokine detection by ELISPOT: relevance for immunological studies in type 1 diabetes. *Diabetes Metab Res Rev*. 2002;18:367–380.
164. Xhonneux LP, Knight O, Lernmark Å, et al. Transcriptional networks in at-risk individuals identify signatures of type 1 diabetes progression. *Sci Transl Med*. 2021;13:eabd5666.
165. Dhatariya KK, Glaser NS, Codner E, et al. Diabetic ketoacidosis. *Nat Rev Dis Primers*. 2020;6:40.
166. Cherubini V, Marino M, Carle F, et al. Effectiveness of ketoacidosis prevention campaigns at diagnosis of type 1 diabetes in children: a systematic review and meta-analysis. *Diabetes Res Clin Pract*. 2021:108838.
167. Krischer JP, Schatz DA, Bundy B, et al. Effect of oral insulin on prevention of diabetes in relatives of patients with type 1 diabetes: a randomized clinical trial. *J Am Med Assoc*. 2017;318:1891–1902.
168. Sosenko JM, Skyler JS, Herold KC, et al. Slowed metabolic decline after 1 year of oral insulin treatment among individuals at high risk for type 1 diabetes in the Diabetes Prevention Trial-Type 1 (DPT-1) and TrialNet Oral Insulin Prevention trials. *Diabetes*. 2020;69:1827–1832.
169. Herold KC, Bundy BN, Long SA, et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N Engl J Med*. 2019;381:603–613.
170. Howson JM, Krause S, Stevens H, et al. Genetic association of zinc transporter 8 (ZnT8) autoantibodies in type 1 diabetes cases. *Diabetologia*. 2012;55:1978–1984.
171. Long AE, Gooneratne AT, Rokni S, et al. The role of autoantibodies to zinc transporter 8 in prediction of type 1 diabetes in relatives: lessons from the European Nicotinamide Diabetes Intervention Trial (ENDIT) cohort. *J Clin Endocrinol Metab*. 2012;97:632–637.
172. Carlsson A, Shepherd M, Ellard S, et al. Absence of islet autoantibodies and modestly raised glucose values at diabetes diagnosis should lead to testing for MODY: lessons from a 5-year pediatric Swedish national cohort study. *Diabetes Care*. 2020;43:82–89.
173. Andersson C, Larsson K, Vaziri-Sani F, et al. The three ZNT8 autoantibody variants together improve the diagnostic sensitivity of childhood and adolescent type 1 diabetes. *Autoimmunity*. 2011;44:394–405.
174. Bonifacio E, Beyerlein A, Hippich M, et al. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. *PLoS Med*. 2018;15:e1002548.
175. Knip M, Åkerblom HK, Al Taji E, et al. Effect of hydrolyzed infant formula vs conventional formula on risk of type 1 diabetes: the TRIGR randomized clinical trial. *J Am Med Assoc*. 2018;319:38–48.
176. Diabetes Prevention Trial–Type 1 Diabetes Study Group. Effects of insulin in relatives of patients with type 1 diabetes mellitus. *N Engl J Med*. 2002;346:1685–1691.
177. Skyler JS, Krischer JP, Wolfsdorf J, et al. Effects of oral insulin in relatives of patients with type 1 diabetes: the Diabetes Prevention Trial--Type 1. *Diabetes Care*. 2005;28:1068–1076.
178. Gale EA, Bingley PJ, Emmett CL, et al. European Nicotinamide Diabetes Intervention Trial (ENDIT): a randomised controlled trial of intervention before the onset of type 1 diabetes. *Lancet*. 2004;363:925–931.
179. Näntö-Salonen K, Kupila A, Simell S, et al. Nasal insulin to prevent type 1 diabetes in children with HLA genotypes and autoantibodies conferring increased risk of disease: a double-blind, randomised controlled trial. *Lancet*. 2008;372:1746–1755.
180. Elding Larsson H, Lundgren M, Jonsdottir B, et al. Safety and efficacy of autoantigen-specific therapy with 2 doses of alum-formulated glutamate decarboxylase in children with multiple islet autoantibodies and risk for type 1 diabetes: a randomized clinical trial. *Pediatr Diabetes*. 2018;19:410–419.
181. Walter M, Philotheou A, Bonnici F, et al. No effect of the altered peptide ligand NBI-6024 on beta-cell residual function and insulin needs in new-onset type 1 diabetes. *Diabetes Care*. 2009;32:2036–2040.
182. Raz I, Ziegler AG, Linn T, et al. Treatment of recent-onset type 1 diabetic patients with DiaPep277: results of a double-blind, placebo-controlled, randomized phase 3 trial. *Diabetes Care*. 2014;37:1392–1400. *Diabetes Care*. 2015;38:178.
183. Agardh CD, Cilio CM, Lethagen A, et al. Clinical evidence for the safety of GAD65 immunomodulation in adult-onset autoimmune diabetes. *J Diabetes Complications*. 2005;19:238–246.
184. Ludvigsson J, Faresjö M, Hjorth M, et al. GAD treatment and insulin secretion in recent-onset type 1 diabetes. *N Engl J Med*. 2008;359:1909–1920.
185. Wherrett DK, Bundy B, Becker DJ, et al. Antigen-based therapy with glutamic acid decarboxylase (GAD) vaccine in patients with recent-onset type 1 diabetes: a randomised double-blind trial. *Lancet*. 2011;378:319–327.

186. Ludvigsson J, Krisky D, Casas R, et al. GAD65 antigen therapy in recently diagnosed type 1 diabetes mellitus. *N Engl J Med*. 2012;366:433–442.
187. Beam CA, MacCallum C, Herold KC, et al. GAD vaccine reduces insulin loss in recently diagnosed type 1 diabetes: findings from a Bayesian meta-analysis. *Diabetologia*. 2017;60:43–49.
188. Hannelius U, Beam CA, Ludvigsson J. Efficacy of GAD-alum immunotherapy associated with HLA-DR3-DQ2 in recently diagnosed type 1 diabetes. *Diabetologia*. 2020;63:2177–2181.
189. Casas R, Dietrich F, Barcenilla H, et al. Glutamic acid decarboxylase injection into lymph nodes: beta cell function and immune responses in recent onset type 1 diabetes patients. *Front Immunol*. 2020;11:564921.
190. Roep BO, Solvason N, Gottlieb PA, et al. Plasmid-encoded proinsulin preserves C-peptide while specifically reducing proinsulin-specific $CD8^+$ T cells in type 1 diabetes. *Sci Transl Med*. 2013;5:191ra82.
191. Cyclosporin-induced remission of IDDM after early intervention. Association of 1 yr of cyclosporin treatment with enhanced insulin secretion. The Canadian-European Randomized Control Trial Group. *Diabetes*. 1988;37:1574–1582.
192. Feldt-Rasmussen B, Jensen T, Dieperink H, et al. Nephrotoxicity of cyclosporin A in patients with newly diagnosed type 1 diabetes mellitus. *Diabet Med*. 1990;7:429–433.
193. Assan R, Feutren G, Sirmai J. Cyclosporine trials in diabetes: updated results of the French experience. *Transplant Proc*. 1988;20:178–183.
194. Leslie RD, Pyke DA, Denman AM. Immunosuppressive therapy in diabetes. *Lancet*. 1985;1:516.
195. Schatz DA, Riley WJ, Silverstein JH, et al. Long-term immunoregulatory effects of therapy with corticosteroids and anti-thymocyte globulin. *Immunopharmacol Immunotoxicol*. 1989;11:269–287.
196. Keymeulen B, van Maurik A, Inman D, et al. A randomised, single-blind, placebo-controlled, dose-finding safety and tolerability study of the anti-CD3 monoclonal antibody otelixizumab in new-onset type 1 diabetes. *Diabetologia*. 2021;64:313–324.
197. Herold KC, Gitelman S, Greenbaum C, et al. Treatment of patients with new onset Type 1 diabetes with a single course of anti-CD3 mAb Teplizumab preserves insulin production for up to 5 years. *Clin Immunol*. 2009;132:166–173.
198. Herold KC, Gitelman SE, Ehlers MR, et al. Teplizumab (anti-CD3 mAb) treatment preserves C-peptide responses in patients with new-onset type 1 diabetes in a randomized controlled trial: metabolic and immunologic features at baseline identify a subgroup of responders. *Diabetes*. 2013;62:3766–3774.
199. Orban T, Bundy B, Becker DJ, et al. Costimulation modulation with abatacept in patients with recent-onset type 1 diabetes: follow-up 1 year after cessation of treatment. *Diabetes Care*. 2014;37:1069–1075.
200. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, et al. Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. *N Engl J Med*. 2009;361:2143–2152.
201. Pescovitz MD, Torgerson TR, Ochs HD, et al. Effect of rituximab on human in vivo antibody immune responses. *J Allergy Clin Immunol*. 2011;128:1295–1302,e5.
202. von Herrath M, Bain SC, Bode B, et al. Anti-interleukin-21 antibody and liraglutide for the preservation of β-cell function in adults with recent-onset type 1 diabetes: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Diabetes Endocrinol*. 2021;9:212–224.
203. Triolo TM, Bellin MD. Lessons from human islet transplantation inform stem cell-based approaches in the treatment of diabetes. *Front Endocrinol (Lausanne)*. 2021;12:636824.
204. Grubin CE, Daniels T, Toivola B, et al. A novel radioligand binding assay to determine diagnostic accuracy of isoform-specific glutamic acid decarboxylase antibodies in childhood IDDM. *Diabetologia*. 1994;37:344–350.
205. Rahmati K, Lernmark A, Becker C, et al. A comparison of serum and EDTA plasma in the measurement of glutamic acid decarboxylase autoantibodies (GADA) and autoantibodies to islet antigen-2 (IA-2A) using the RSR radioimmunoassay (RIA) and enzyme linked immunosorbent assay (ELISA) kits. *Clin Lab*. 2008;54:227–235.
206. Törn C, Mueller PW, Schlosser M, et al. Diabetes Antibody Standardization Program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2. *Diabetologia*. 2008;51:846–852.
207. Kawasaki E, Eisenbarth GS, Wasmeier C, et al. Autoantibodies to protein tyrosine phosphatase-like proteins in type I diabetes. Overlapping specificities to phogrin and ICA512/IA-2. *Diabetes*. 1996;45:1344–1349.

33

Type 2 Diabetes Mellitus: Etiology, Pathogenesis, and Natural History

W. Timothy Garvey and R. Paul Robertson

OUTLINE

Introduction, 557
Epidemiology, 557
Origins of the Disease: Etiologic Factors, 558
  Region, Race, Ethnicity, and Genetic Factors, 558
  Environment and the Social Determinants of Health, 560
  Behavior and Culture, 561
  Biological Factors, 561
Natural History of Diabetes as an Evolving Glucoregulatory Defect, 561
  Diagnosis of Type 2 Diabetes, 561
  Overview of the Evolving Glucoregulatory Defect, 563
  Progression of Abnormal Glucose Tolerance, 564
  Mechanisms Causing Pancreatic β Cell Dysfunction, 568
The Pathophysiology of Insulin Resistance, 570
  Measurement of Insulin Sensitivity in Humans: Defect in Muscle Glucose Uptake, 570
  Glucose Transport, Oxidative Metabolism, and Nonoxidative Metabolism, 570
  Hepatic Insulin Resistance, 571
The Molecular Basis of Human Insulin Resistance in Target Tissues, 572
  General Principles, 572
  Insulin Resistance at the Level of the Insulin Receptor, 573
  Insulin Resistance at the Level of Signal Transduction, 574
  Insulin Resistance at the Level of Biological Effector Systems, 575
Cell Stress Responses, Nutrient Sensing, and Insulin Resistance, 577
  Cell Stress Responses and Insulin Action, 577
  Nutrient Sensors and Nutrient Stress, 578
  The Hexosamine Biosynthetic Pathway, TRIB3, and Glucose-Induced Insulin Resistance, 579
Syndromes of Severe Insulin Resistance, 579
  Heterogeneity of Type 2 Diabetes, 579
  Severe Insulin Resistance, 580
Natural History of Type 2 Diabetes as an End-Stage Manifestation of Cardiometabolic Disease, 580
  The Spectrum of Cardiometabolic Disease, 580
  Role of Obesity, 580
  Implications for Type 2 Diabetes Prevention and Treatment, 581
  Disclosures, 582

INTRODUCTION

The term "type 2 diabetes mellitus" (T2D) is often erroneously thought to refer to a discrete disease entity that has a specific cause and therapeutic approaches that can be used to target and reverse its pathophysiologic consequences. However, this term more accurately describes a heterogeneous, polygenic syndrome. The most prominent features are fasting and postprandial hyperglycemia without an unmitigated progression to ketoacidosis, often combined with hyperlipidemia, and with onset in middle life, although the disease can initiate in childhood or adolescence, as well as in the elderly. In addition, T2D lies within the broader context of cardiometabolic disease (CMD), with progression to cardiovascular disease (CVD) and nonalcoholic fatty liver disease (NAFLD). Both a previous term (non–insulin-dependent diabetes mellitus) and the current term (T2D) were created to distinguish this syndrome from type 1 diabetes (T1D). The latter disease (see Chapter 32) is a more clearly definable disease entity for which laboratory-based measurements implicate both autoimmune and genetic causations.

EPIDEMIOLOGY

T2D accounts for more than 90% of patients with diabetes and is an increasing cause of morbidity and mortality. In 2019, approximately half a billion individuals had T2D, which amounted to 9.3% of adults worldwide.[1] Prevalence rates rise as a function of age, with 4.4% of individuals affected aged 15 to 49 years, 15% aged 50 to 69 years, and 22% aged 70 years and above.[2] Incidence rates rise progressively up to age 55 to 60 years, and the gender distribution is approximately equal. Over 1 million deaths per year internationally can be attributed to diabetes alone, making it the ninth leading cause of mortality, although diabetes heavily contributes to mortality ascribed to other disease processes such as CVD, NAFLD, dementia, kidney disease, and cancer. In the United States,[3] 13% of adults had diabetes in 2018, and 21% of these individuals were unaware that they had the disease. Eighty-nine percent of individuals with T2D have overweight or obesity, and 38% were physically inactive. One in three US adults has prediabetes, which is of concern because conversion to overt diabetes occurs at a rate of greater than 5% per year,[4] although this varies as a function of associated risk factors.[5] The burden of diabetes mellitus is rising globally. Prevalence in adults is projected to increase by 25% in 2030 and 51% by 2045,[1] and the increasing number of children and adolescents who are developing T2D is particularly problematic. These data point to the need for concerted efforts at disease prevention.

KEY POINTS

- Type 2 diabetes accounts for 90% of diabetes and is a polygenic complex syndrome rather than a specific disease. It affects 9.3% of adults worldwide, and directly accounts for over 1 million deaths/year worldwide.

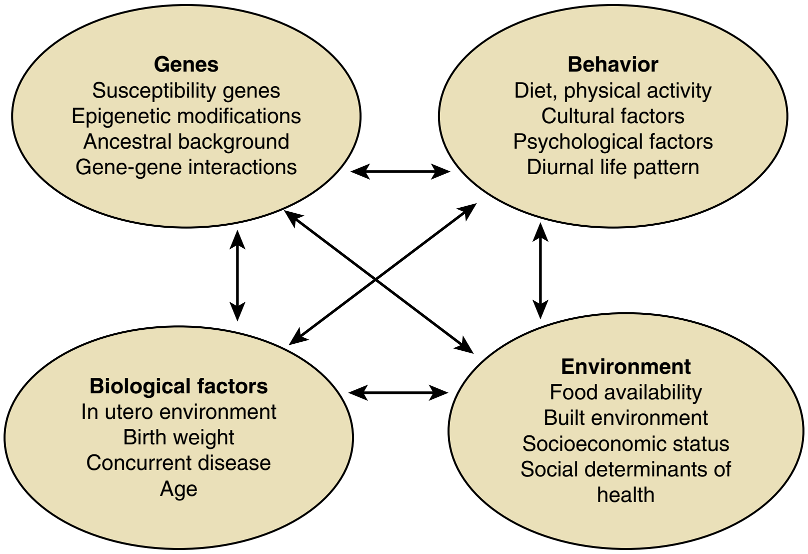

Genes
Susceptibility genes
Epigenetic modifications
Ancestral background
Gene-gene interactions
Behavior
Diet, physical activity
Cultural factors
Psychological factors
Diurnal life pattern
Biological factors
In utero environment
Birth weight
Concurrent disease
Age
Environment
Food availability
Built environment
Socioeconomic status
Social determinants of health

Fig. 33.1 Interacting genetic, biological, behavioral, and environmental factors involved in the pathogenesis of type 2 diabetes.

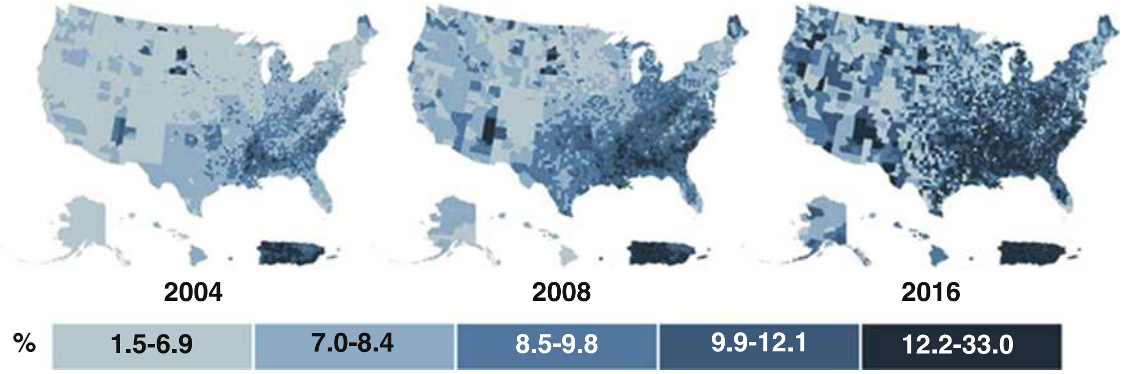

2004
2008
2016
%
1.5-6.9
7.0-8.4
8.5-9.8
9.9-12.1
12.2-33.0

Fig. 33.2 The US Diabetes Density Map. County-by-county diabetes prevalence map showing percent of US adults with diabetes in 2004, 2008, and 2016. (Data from the Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020.)

ORIGINS OF THE DISEASE: ETIOLOGIC FACTORS

T2D is a complex chronic disease that is the result of variable interactions among genetic, behavioral, biological, and environmental factors, as illustrated in Fig. 33.1. Because of this complexity, T2D presents as a heterogeneous disorder. While all patients are characterized by hyperglycemia, the physical and metabolic phenotype of the disease can vary as a function of age, region and ethnicity, body mass index (BMI), distribution of adipose tissue depots, syndromic features in patients with mono/oligogenetic disease, propensity for vascular complications, and the degree to which insulin resistance and impaired insulin secretion contribute to hyperglycemia.

Region, Race, Ethnicity, and Genetic Factors

T2D variably impacts global regions, countries, and ethnic and racial groups. Globally, Middle East and North Africa have the highest diabetes prevalence rates, followed by the Western Pacific and Southeast Asia.[1] Prevalence rates are particularly high in West Pacific Island and Arabic countries. The country with the highest number of diabetes patients is China, followed by India, the United States, and Pakistan. In countries with diverse populations, diabetes disproportionately affects specific ethnic and racial groups. In the United States,[3] prevalence rates are increased 2-fold in Native Americans, 1.6-fold in non-Hispanic Black Americans and people of Hispanic origin, and 1.2-fold in non-Hispanic Asian Americans compared with non-Hispanic White Americans. In addition, T2D prevalence rates among youth are increasing most dramatically in non-Hispanic Black Americans and Hispanics. Fig. 33.2 shows county-by-county diabetes prevalence rates across the United States and the increments in prevalence from 2004 to 2016. Due to high concentrations of the disease in Appalachia and the southern tier of states, the Centers for Disease Control has termed this region the “Diabetes Belt.”[6]

There is a strong genetic contribution to the development of T2D, as evidenced by five lines of investigation: (i) racial aggregation, (ii) familial aggregation and segregation, (iii) twin studies, (iv) relatively uncommon monogenic forms of diabetes such as maturity onset diabetes of the young (MODY), and (v) the identification of susceptibility alleles in candidate gene and genome-wide association studies (GWAS).

Evidence for the Genetic Basis of Type 2 Diabetes: Race, Family, and Twins. Racial aggregation is exemplified by racial or ethnic groups with defined ancestral origins having higher prevalence rates of diabetes than other subpopulations living in a comparable environment. Examples include higher diabetes rates in Black, Hispanic, and Japanese Americans compared with the general US population.

T2D also has a strong family history. Sibling recurrence risk ratio (λ_s) is a measure of familial aggregation that is operationally defined as the proportion of affected siblings of affected individuals compared with the general population. This ratio varies among populations, but generally the relative risk is approximately 1.5 to 2.0 for T2D.[7] The risk of developing T2D in a child with a parent who has diabetes is one in seven if the affected parent was diagnosed under 50 years of age, and one in 13 if diagnosed when older than 50 years. Some studies indicate that risk of T2D in offspring may be somewhat higher if the affected parent is the mother. If both patients have T2D, the risk in a child increases to approximately 50%.

Some of the strongest evidence for the genetic basis of T2D derives from twin studies.[8-10] These studies have demonstrated higher rates of concordance for T2D in monozygotic twins (80%–95%) who share 100% of their DNA compared with lower concordance rates (30%–40%) in dizygotic twins who share only 50% of their DNA. Despite strong family history and concordance rates in twins, the inheritance pattern of common T2D does not conform to any recognizable Mendelian pattern. Thus, T2D in the vast majority of patients is not the result of a single gene defect, but rather a polygenic disorder that results from the inheritance of multiple genetic mutations, each with a small effect size. This can account for phenotypic heterogeneity within T2D, because individuals inherit different subsets of susceptibility genes that interact in varying ways with different environmental factors.

Genetic Studies: Candidate Gene and Whole-Genome Approaches. The candidate gene approach begins with identifying key proteins highly relevant to glucose homeostasis and then assessing whether corresponding genes display polymorphisms that are associated with diabetes. A refined candidate gene approach conducts similar analyses on metabolic subphenotypes, such as proteins relevant to insulin secretion or insulin action, with the presumption that these traits are less polygenic and more proximal manifestations of genetic defects causing diabetes. Genetic studies using candidate gene approaches and linkage analyses have identified several T2D-associated loci, including peroxisome proliferator activated receptor γ (PPARG), inwardly rectifying Kir6.2 component of the pancreatic β cell adenosine triphosphate–dependent potassium channel (KCNJ11), transcription factor 7–like 2 (TCF7L2), and wolframin (WFS1).[11]

GWAS examine single nucleotide polymorphisms (SNPs) or other types of genetic markers interspersed throughout the genome and identify which polymorphisms may be associated with a disease in a large case-control cohort. Specific disease-associated polymorphisms serve as markers of potential causal genes at their respective chromosomal loci. Nucleotide sequencing or positional cloning techniques are then used to identify specific genes with causal mutations within each associated locus. GWAS have employed increasing sample sizes as consortia have been formed to combine genomic and phenotypic data from multiple studies, and over 400 SNPs at loci throughout the genome have been associated with T2D.[12-14] For example, a metaanalysis combined GWAS data from the Diabetes Genetics Replication and Meta-analysis, Genetic Epidemiology Research on Aging, and the full-cohort release of the UK Biobank to establish a database of 62,892 patients with T2D and 596,424 unaffected controls.[13] This study identified 139 common and four rare SNPs independently associated with T2D, including 39 variants reaching genome-wide significance that had not previously been reported. Pathway analyses showed that loci associated with T2D contained genes that were most highly enriched in molecular pathways involved in glucose homeostasis and MODY. Further functional enrichment incorporating protein–protein interactions pointed to four genes involved in insulin secretion (*ADCY5*, *CAMK2G*, *KCNJ11*, and *KCNU1*) and another four relevant to MODY (*HHEX*, *HNF1A*, *HNF1B*, and *FOXA2*). This study and others have demonstrated that identified susceptibility alleles individually confer only a small relative risk for diabetes (generally <1%) and in aggregate explain only a small proportion (<10%) of T2D heritability. Of these polymorphic loci, the preponderance of genes associated with T2D and abnormal glucose tolerance are involved in pancreatic islet development and regulation of insulin secretion, including *CDKAL1, KCNQ1, IDE, HHEX, ABCA1, ABCC8,* BCL11A, TCF7L2, SLC30A8, HHEX/IDE, and KCNJ11.[12,14,15] The strongest genome-wide association signals were observed for SNPs in *TCF7L2*,[16] a gene involved in the WNT signaling pathway and affecting pancreatic β cell function.[17] This is consistent with the heritability factor for the insulin secretory defect, estimated to be as high as 70%.[18-20] The heritably scores (h^2) for insulin resistance are generally more modest, in the range of approximately 30%.[18,19] Nevertheless, several chromosomal regions have been found to associate with various measures of insulin resistance in GWAS, such as *PPARG, IRS1, NAT2, IGF1, BCL2,* and *FAM19A2*.[21-24] Again, each of these genes individually confers only a small relative risk for impaired insulin secretion and insulin resistance, which are considered key mechanistic subphenotypes of T2D. The relatively low heritability scores for insulin resistance indicate that factors such as physical inactivity and diets enriched in saturated fats, refined grains, and low fiber could be responsible for a substantial amount of individual variation in insulin sensitivity.[25]

There is a strong heritability component to BMI that varies between 40% to 70%,[26,27] and over 100 susceptibility genes for elevated BMI have been identified in GWAS studies.[28,29] The vast preponderance of susceptibility genes for obesity do not overlap with those associated with insulin resistance or diabetes,[30,31] suggesting that, while interactions are operative in the expression of these traits, obesity and insulin resistance largely segregate independently regarding genetic risk. The fat mass and obesity–associated (*FTO*) and melanocortin-4 receptor (*MC4R*) genes are exceptions, in that mutations in these genes are associated with both elevated BMI and diabetes.

With advances in nucleotide sequencing and throughput, large-scale sequencing has been conducted of entire genomes or of all exomes that contain the protein-coding component, estimated to comprise less than 2% of the genome. Exome sequencing promises to enable the identification of causal genes within GWAS loci, and also of rare variants not detectable by GWAS.[32-34] The T2D-Genes Consortium has endeavored to conduct exome sequencing in patients with T2D and controls across five ethnicities: Black, South Asian, East Asian, Hispanic, and European. Investigators have reported analyses of 20,791 individuals with T2D and 24,440 nondiabetic control participants in aggregate from the T2D-Genes Consortium and five additional consortia.[32] These authors found 2.26 million variants associated with T2D leading to a change in amino acid sequence (i.e., nonsynonymous), and 93.5% of these were rare, having an allele frequency of less than 0.5%. Rare variants reaching genome-wide significance included a novel mutation in a known gene (*MC4R*), a novel gene (*UBE2NL*), and several other genes known to be associated with T2D risk (e.g., *SLC30A8, PAM*). Interestingly, the rare alleles together could explain far less risk than that conferred by common variants, indicating that a significant component of T2D risk cannot be explained by unknown rare variants with high effect size. Of 595 genes within known GWAS loci, 40 were associated with T2D in the exome data, and only 14 of these genes represented the named gene given to this locus, which underscores the hazards of assuming an attractive candidate within a GWAS locus is actually responsible for the association. Other genes not quite reaching whole-exome significance were grouped into functional sets and found to be of sub–exome-wide significance, including multiple drug targets (*GLP1R, IGF1R, PPARG, INSR, SLC5A2, DPP4,* and *ABCC8*). This landmark study[32] still leaves unexplained the question of missing genetic factors needed to fully explain heritability; this perhaps will require more sophisticated analyses of multiple gene interactions and interactions with environmental factors.

Epigenetic Factors. There is increasing interest in the epigenetics of diabetes as a mechanism by which changes in DNA or chromatin structure can alter gene expression without involving alterations in nucleotide sequence. Epigenomics refers to genome-wide studies that focus on three processes: (i) DNA methylation, primarily on cytosine residues present at CpG sites, that can result in suppression of gene expression when occurring in promoter or enhancer regions; (ii) modification of histone proteins by acetylation, methylation, sumoylation, or ubiquitination, which alters chromatin remodeling required for gene transcription; and (iii) noncoding RNAs (e.g., microRNAs) that can interfere with mRNA processing or translation. Each of these mechanisms has been shown to alter cellular biology in ways that are relevant to diabetes pathogenesis.

The alterations in DNA and histone structure can be inherited from maternal or paternal germ cells or arise *de novo* in a tissue specific manner. Importantly, epigenetic changes can be induced by nutritional and environmental factors,[35,36] providing a mechanism by which environmental exposures can produce persisting changes in metabolism over the lifespan. In humans, circulating white blood cells provide a convenient source of material for epigenetic studies; however, these changes may not reflect epigenetic modifications and effects on gene expression in metabolically relevant tissues.[37] Even so, investigators have used genomic DNA from white blood cells in T1D patients enrolled in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study and have shown that persisting methylation at key CpG sites was associated with hemoglobin A1c (HbA1c).[38] This observation provided a potential explanation for "metabolic memory," by which a previous period of intensive glycemic control could provide benefits with continued follow-up long after the intervention. Animal studies allow for the study of epigenetic changes and corresponding effects on gene expression within relevant tissues. For example, studies of genome-wide methylation patterns in pancreatic β cells from young and old mice found that methylation of proliferative gene loci contributes to a decline in insulin secretory capacity that is recapitulated by epigenetic changes identified in human islets obtained from patients with T2D.[37,39,40]

Accumulating evidence indicates that differential CpG methylation can occur as a function of the *in utero* environment and participate in the programming of body weight and metabolism in offspring into adulthood, allowing for the intergenerational transmission of metabolic phenotypes. Epigenetic modifications have been implicated that alter body composition and metabolism in offspring and that confer insulin resistance and increased risk for T2DM, obesity, and CVD persisting into adulthood.[41-44] These studies assess epigenetic modifications using cord blood or placenta at birth and have identified changes in the methylation of genes involved in inflammation, insulin signaling, and leptin.[45-49] These studies have demonstrated that gestational diabetes is associated with differential methylation of *MEST, NR3C1,* and *LEP* in offspring compared with children born to glucose-tolerant mothers, associated with greater insulin and leptin concentrations. Fetal hyperinsulinemia due to intrauterine exposure to gestational diabetes or obesity persists into childhood and adolescence, consistent with the hypothesis that insulin sensitivity and secretion are programmed *in utero.*[50]

Methylation of CpGs is mediated by DNA methyltransferases, but these changes are reversible via the action of a family of ten-eleven translocation enzymes. Demethylation of key CpG islands can be induced by weight loss interventions involving exercise, diet, and weight loss surgery.[51-53] Interestingly, genes involved in obesity, insulin receptor signaling, T2D, and leptin signaling were differently methylated in children born before versus after their mothers' bariatric surgery–associated weight loss.[52,53] Clearly, epigenetics provides mechanisms by which environment and behavior lead to persisting alterations in gene expression that can contribute to pathogenesis or reduce the risk of T2D.

Noncoding RNA species also play a role in T2D. MicroRNAs are short noncoding RNA molecules usually consisting of a maximum of 22 nucleotides that bind complementary target mRNAs in the RNA-induced silencing complex and either degrade the target mRNA or block its translation. MicroRNAs can act locally within tissues and cells to regulate gene expression or be released into the circulation, protected from degradation in exosomes or by adherence to circulating proteins. Circulating miRNAs can be targeted for uptake by specific cell types and alter gene expression in the host cell,[54] thus constituting a mechanism for intertissue communication.[55,56] MicroRNAs can play a regulatory role in normal metabolic homeostasis, but also contribute to disease processes, as described in more detail in Chapter 30. Work in humans has most often focused on blood levels of microRNAs, and multiple species have been identified that correlate with diabetes, obesity, and other metabolic traits.[57-60] As an example, circulating miR-150 levels in humans are elevated in insulin resistance (i.e., negatively correlated with clamp measures of insulin sensitivity).[57] In cultured cell systems and rodent models, miR-150 has been shown to regulate lipid metabolism and inflammatory responses[61] and attenuate the browning process in adipocytes.[62] One of the targets of miR-150 is pgc-1α,[62] which is known to be involved in lipid metabolism and adipocyte browning, and pgc-1α in turn upregulates peroxisome proliferator–activated receptor α (PPARα), a key transcription factor affecting fatty acid (FA) oxidation.[63] An increase in miR-150 could reverse these processes by suppressing expression of pgc-1α. In intact mice, total body knockout of miR-150 reduced whole-body fat mass, improved systemic glucose tolerance and insulin sensitivity, and reduced production of proinflammatory cytokines in adipose tissue.[64] More definitive work is required to address mechanisms by which changes in circulating microRNAs contribute to T2D. Nevertheless, this field is promising, because microRNAs can potentially serve as biomarkers, or even as therapeutic agents by intravenous administration.[65-67]

Environment and the Social Determinants of Health

It is well documented that social and environmental factors are associated with higher rates of diabetes and help explain differences in prevalence and health disparities in populations.[68] Low socioeconomic status and reduced economic viability are associated with greater diabetes risk. This is often accompanied by a lower education level affecting capacity for income generation, health literacy, and access to care determined by lack of health insurance, ability to afford care, availability of health care professionals, and transportation. The built environment can contribute to higher diabetes prevalence when there is lack of access to unprocessed food and resources enabling physical activity. Social support systems and public health messaging are additional factors that have been well documented to impact outcomes of chronic disease. Sociopolitical factors can generate discrimination based on age, ethnicity, social class, and other attributes, which produces disenfranchisement and both actual and perceived stress. These determinants not only dictate access to resources and the capacity to prevent and/or manage diabetes, but also contribute to psychological stress and depression, which have been linked to poor diabetes outcomes.

Behavior and Culture

Lifestyle patterns based on personal or cultural preferences have a substantial impact in the pathogenesis of diabetes, particularly in relation to nutrition and physical activity. Recommendations for a "healthy meal pattern" generally include: (i) limitations on fat intake; (ii) emphasis on poly/monounsaturated fats over saturated fats; (iii) no *trans* fats; (iv) avoidance of added sugar and sugar-sweetened beverages; (v) complex over simple carbohydrates; (vi) whole grains; (vii) fruits and vegetables; (viii) dietary fiber, (ix) limited sodium intake; (x) moderation in alcohol; (xi) reduced consumption of processed food.[69] These recommendations are largely supported by epidemiological data, while long-term controlled clinical trials specifying progression to T2D and CVD events as primary outcomes are lacking. Fortunately, there is a large body of data indicating that isocaloric substitution of specific macronutrients can influence insulin sensitivity and CVD risk factors. Specifically, isocaloric enrichments or substitutions of macronutrients have been demonstrated to increase insulin sensitivity (mono- and polyunsaturated fats, whole grains, high fiber, low glycemic index) or worsen insulin resistance (saturated and trans fats, refined grains, low fiber, high glycemic index) in studies employing glucose clamps for measurement of insulin sensitivity.[25] Similarly, physical inactivity and a sedentary lifestyle augment diabetes risk, and this can occur independent of obesity.[25] Importantly, lifestyle interventions have been shown to prevent progression to diabetes among patients at high risk and to reduce glycemia or even produce disease remission among patients with overt T2D. This includes exercise programs or meal patterns such as the Mediterranean diet[70,71] that can reduce T2D risk even without weight loss, and lifestyle modifications producing weight loss in patients with overweight or obesity.[72] As an example of the latter, the Diabetes Prevention Program demonstrated that weight loss as a result of a reduced calorie diet and exercise in patients with impaired glucose tolerance prevented progression to diabetes by 58% compared with placebo.[73]

KEY POINTS

- Type 2 diabetes is highly inheritable and highly polygenic. Genome-wide association studies have identified over 400 susceptibility loci, each conferring only a small relative risk for the disease and, in aggregate, explaining <10% of diabetes risk. More research is warranted addressing gene–gene and gene–environment interactions and epigenetic factors.

Cultural identification and acculturation regarding diet and food preferences and emphasis on physical activity are important determinants and can affect T2D risk in individuals as well as populations. The influence of culture and lifestyle is underscored by the change in diabetes prevalence when ethnic groups migrate from a rural to an urban environment or from an agrarian society to a more sedentary, developed, or "Westernized" society. This is exemplified by Gullah-speaking Black Americans living in the South Carolina low country compared with the genetically-related ancestral population in Sierra Leone, West Africa.[74] Gullah-speaking Black Americans have high rates of T2D characterized by relatively high rates of diabetic complications, early age of onset, and a high relative risk to siblings of T2D (λ_S=3.3) and consume a diet rich in animal fats.[75,76] This is compared with a lower diabetes prevalence in Sierra Leone, where diets are high in carbohydrates.[77] Another example is the group of Pima Indians who moved from northern Mexico to southern Arizona approximately 2,000 years ago and since the early 1900s have adopted a more sedentary and "Westernized" lifestyle.[78-80] The Pima Indians who continue to live in northern Mexico and have retained an agrarian lifestyle have a diabetes prevalence of 6.9%, which is less than one-fifth the value of 38% in the Arizona Pima who also have higher rates of obesity. These lifestyle changes in Gullahs and Pimas predispose to the development of obesity, which is a major factor in the development of diabetes, as is discussed below.

Biological Factors

Biological factors unrelated to diabetes also affect diabetes pathogenesis, including aging, menopause, concurrent diseases, birth weight, and effects of the *in utero* environment on offspring. The prevalence of T2D rises progressively with age, although the incident rate increases up to approximately 60 years of age before slowly declining over the next several decades. The increase in T2D risk can at least partially be explained by an age-related decline in both insulin sensitivity and insulin secretion leading to a worsening of glucose tolerance,[81] together with a reduction in physical activity, decrease in muscle mass, and increase in percent body fat. Menopause in females can also be associated with an increase in total and intraabdominal fat, a reduction in physical activity, dyslipidemia, and insulin resistance.[82]

Low birth weight is a risk factor for the development of insulin resistance and diabetes mellitus later in life.[83] The mechanisms responsible for this association are unknown but may be related to epigenetic fetal adaptation to nutritional stimuli or excess fetal glucocorticoid exposure. For example, among a group of 60 adults with low birth weight and prenatal exposure to famine during the Dutch Hunger Winter in 1944 to 1945, there was less DNA methylation of the imprinted insulin-like growth factor 2 (*IGF2*) gene 60 years later compared with unexposed, same-sex siblings.[84] Even in normal-weight neonates, the *in utero* environment in mothers with gestational diabetes or obesity can produce an insulin-resistant phenotype in offspring.[41-44] Maternal health has become recognized as an important factor in decreasing rates of diabetes, obesity, and CVD.

KEY POINTS

- Type 2 diabetes (T2D) pathogenesis involves complex interactions among genetic and epigenetic factors, ethnicity, behavior, cultural practices, environment, biological factors, and social determinants of health. These varied interactions underlie the clinical heterogeneity of T2D.

Finally, a host of comorbid diseases can predispose to T2D. Many of these diseases worsen insulin resistance or impair insulin secretion, and include hormonal and metabolic disorders, lipoatrophy, pancreatic disease, iatrogenic medications, and genetic syndromes. The reader is referred to Table 33.1 and Chapter 35 for more extensive information.

NATURAL HISTORY OF DIABETES AS AN EVOLVING GLUCOREGULATORY DEFECT

Diagnosis of Type 2 Diabetes

The onset of T2D is most often insidious, with unexplained weight loss, lethargy, blurred vision, and polyuria; however, up to a third of patients remain undiagnosed and are unaware of the disease. Prediabetes can be identified[85] on the basis of: (i) impaired fasting glucose (IFG) with fasting glucose levels of less than or equal to 100 mg/dL (5.6 mmol/L); (ii) impaired glucose tolerance (IGT) diagnosed when 2-hour blood glucose levels are greater than 140 mg/dL (7.8 mmol/L) but less than 200 mg/dL (11.1 mmol/L) during a 75-gram oral glucose

TABLE 33.1 Unusual Phenotypes of Type 2 Diabetes

Category	Disease	Chapter/Reference
Genetic defects of β cell function	Maturity onset diabetes of the young (MODY) 1–6	Chapter 31
	Neonatal diabetes	Chapter 31
	Mitochondrial diabetes	*(i)*
Genetic defects in insulin action	Severe insulin resistance type A	Current chapter
	Leprechaunism, Donohue Syndrome	*(ii)*, *(iii)*
	Rabson–Mendenhall Syndrome	*(iv)*
Immune mechanisms	Severe insulin resistance type B	Current chapter
	Latent autoimmune diabetes in adults	Chapter 31
	"Stiff person" syndrome	*(v)*
Phenotypical features transitioning between type 2 diabetes and type 2 diabetes, depending on stage of disease	Latent autoimmune diabetes in adults (LADA)	Chapter 31
	Atypical diabetes	*(vi)*
Pancreatic disease	Cystic fibrosis	Chapter 35
	Chronic pancreatitis	Chapter 35
	Pancreatectomy	Chapter 35
	Tropical fibrocalculous diabetes	Chapter 35
Lipoatrophy	Generalized lipodystrophy	Chapter 24
	Partial lipodystrophy	Chapter 24
	Nonclassical patterns of lipodystrophy, e.g., type 2 diabetes and partial lipodystrophy of the limbs	*(vii)*
Metabolic disease	Hemochromatosis	Chapter 35
	Pseudoacromegaly	Chapter 35
Endocrinopathies	Hypercortisolism, acromegaly, hyperthyroidism, aldosteronoma	Chapter 35
	Polycystic ovary syndrome	Chapter ?
Drug-induced	Vacor, glucocorticoids, diazoxide pentamidine, nicotinic acid, thiazides, β-adrenergic agonists, dilantin, interferon, antiretrovirals	Chapter 35
Infections	Congenital rubella, cytomegaovirus	*(viii)*
Secretory neuroendocrine tumors	Somatostatinoma, glucagonoma, pheochromocytoma	Chapter 35
Genetic syndromes with increased incidence of diabetes	Down syndrome, Klinefelter syndrome, Turner syndrome, Friedreich ataxia, Wolfram syndrome, Huntington chorea, Laurence–Moon–Biedl syndrome, porphyria, Prader–Willi syndrome, progeria, Werner syndrome, Alström syndrome, myotonic dystrophy, mandibuloacral dysplasia	*(ix)*

[i] Chow J, Rahman J, Achermann JC, et al. Mitochondrial disease and endocrine dysfunction. *Nat Rev Endocrinol*. 2017;13:92–104.

[ii] Sethu-Kumar Reddy S, Lauris V, Kahn C. Insulin receptor function in fibroblasts from patients with Leprechaunism: differential alterations in binding, autophosphorylation, kinase activity, and receptor-mediated internalization. *J Clin Invest*. 1988;82:1359–1365.

[iii] Kirkwood A, Stuart G, Harding L. Donohue syndrome: A review of literature, case series, and anesthetic considerations. *Paediatr Anaesth*. 2018;28:23–27.

[iv] Yoshimasa Y, Seino S, Whittaker J, et al. Insulin-resistant diabetes due to a point mutation that prevents insulin proreceptor processing. *Science*. 1988;240:784–787.

[v] McKeon A, Robinson MT, McEvoy KM, et al. Stiff-man syndrome and variants: clinical course, treatments, and outcomes. *Arch Neurol*. 2012;69:230–238.

[vi] Steenkamp DW, Alexanian SM, Sternthal E. Approach to the patient with atypical diabetes. *Canadian Med Assoc J*. 2014;186:678–684.

[vii] Strickland LR, Guo F, Lok K, Garvey WT. Type 2 diabetes with partial lipodystrophy of the limbs: a new lipodystrophy phenotype. *Diabetes Care*. 2013;36:2247–2253.

[viii] Rogers MAM, Kim C. Congenital infections as contributors to the onset of diabetes in children: A longitudinal study in the United States, 2001-2017. *Pediatr Diabetes*. 2020;21:456–459.

[ix] Robinson S, Kessling A. Diabetes secondary to genetic disorders. *Baillieres Clin Endocrinol Metab*. 1992;6:867–898.

tolerance test (OGTT); and/or (iii) with HbA1c values of 5.7% to 6.4% (39–47 mmol/mol). The diagnosis of T2D requires fasting glucoses greater than or equal to 126 mg/dL (7.0 mmol/L), 2-hour OGTT values of greater than or equal to 200 mg/dL (11.1 mmol/L), and/or HbA1c greater than or equal to 6.5% (48 mmol/mol). Unlike T1D, exogenous insulin is generally not required to sustain life in T2D, and patients do not develop diabetic ketoacidosis except under conditions of high biological stress. Nonetheless, many T2D patients use insulin to improve their glycemic management. Patients with T2D experience declining insulin secretory function over time, and with longstanding disease retain minimal secretion reflected by C-peptide levels greater than 2.0 ng/mL.

The are several important considerations in making the clinical diagnosis of T2D. First, the categories of normoglycemia, prediabetes, and mild diabetes based on fasting or 2-hour OGTT glucose values can vary in individual patients over time, and glycemic status should be confirmed by repeated testing on a separate day. Second, HbA1c has high specificity but relatively low sensitivity for diagnosing prediabetes or mild diabetes.[86] Clinicians should confirm glycemic status using fasting and/or 2-hour OGTT glucose rather than reliance on HbA1c alone. Third, with advancing age above 50 years of age, an increasing majority of patients will have prediabetes on the basis of elevated 2-hour OGTT alone,[87] and the diagnosis will be missed if an OGTT is not performed.

HbA1c levels reflect the average of the ambient blood glucose levels developed throughout each day for the preceding 3 months. This measurement has been adapted clinically as a reliable measure of successful glucose control during treatment for T1D and T2D. It is closely correlated to the development of retinopathy and other vascular complications due to elevated glycemic levels. Maintenance of HbA1c levels at 6.5% to 7.0% with no recurring hypoglycemia during treatment with oral agents or insulin is generally considered acceptable glycemic regulation (see Chapter 40). Levels between 7.0% and 8.0% are recommended for people recurrently at risk for hypoglycemia during treatment.

Overview of the Evolving Glucoregulatory Defect

Glucose homeostasis reflects a balance between insulin secretion and tissue insulin sensitivity in both the fed and fasted state. Hyperglycemia in most cases of T2D is a consequence of both impaired insulin secretion and insulin resistance.[88-90] There are four key tissues of particular significance in the development of abnormal glucose tolerance. They are the pancreas, liver, skeletal muscle, and adipose tissue, as illustrated in Fig. 33.2. Insulin resistance in liver and skeletal muscle contribute directly to the systemic manifestation of insulin resistance, namely hyperinsulinemia in the face on normo- or hyperglycemia. In insulin-resistant individuals, normal fasting glucose levels are maintained by compensatory increases in rates of insulin secretion that inhibit hepatic glucose output (HGO) in balance with counterregulatory hormones such as glucagon, glucocorticoids, and growth hormone that increase HGO. When fasting glucose levels become elevated in IFG and T2D, the HGO rate is increased, because insulin secretion is no longer sufficient to compensate for hepatic insulin resistance and is highly correlated with the severity of the fasting hyperglycemia. Elevated HGO rates are primarily the result of accelerated gluconeogenesis, which are facilitated by increased flux of gluconeogenic precursors and elevated free FAs (FFAs) and glucagon levels. The other key tissue contributing to insulin resistance is skeletal muscle, which is responsible for 70% to 80% of all insulin-mediated glucose uptake *in vivo*. Peripheral insulin resistance is largely due to insulin action defects in skeletal muscle that impair insulin's ability to stimulate glucose transport. Essentially all patients with CMD who progress to T2D exhibit a relative degree of insulin resistance in skeletal muscle.

Not all patients with CMD and prediabetes progress to overt T2D. Those patients who do develop T2D and frank hyperglycemia have defective pancreatic β cells that undergo exhaustion under conditions of the chronic metabolic stress of insulin resistance and continuous hyperglycemia. In some people with longer duration of diabetes, this leads to accelerated apoptosis and reduction in β cell mass. Thus, although causal mechanisms and clinical phenotypes can be heterogeneous in different T2D patient groups, the ultimate development of chronic hyperglycemia involves some combination of impaired insulin secretion, insulin resistance, and increased hepatic glucose production. The relative magnitude and contribution of these three metabolic abnormalities can vary depending upon specific subsets of inherited susceptibility genes, as well as other biological, behavioral, and/or environmental factors operative in individual patients or subgroups of patients. For example, T2D can develop in some lean people who do not clinically appear to have insulin resistance because they use only small amounts of insulin (e.g., 20 units/day), yet maintain acceptable levels of HbA1c.

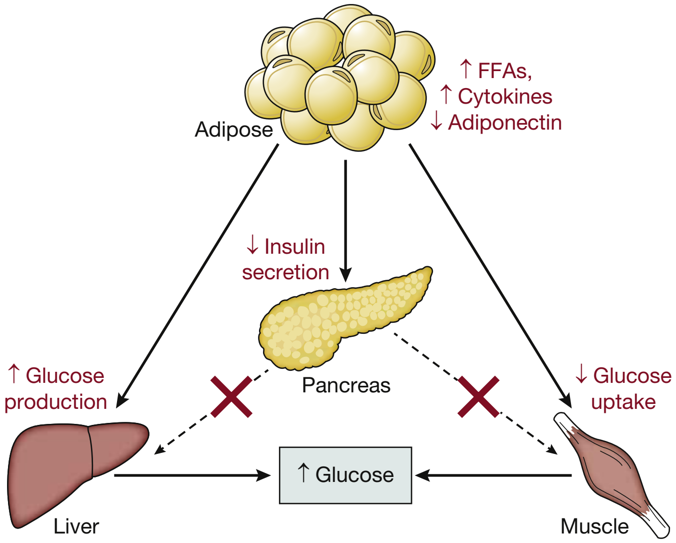


Fig. 33.3 Abnormalities in four key tissues contribute to hyperglycemia in type 2 diabetes. Dysfunctional and inflamed adipose tissue results in dysregulated secretion of adipocytokines and defects in fat storage, which leads to increased release of free fatty acids (FFAs) and ectopic accumulation of lipid in myocytes and hepatocytes. Insulin resistance develops in skeletal muscle with impaired stimulation of glucose uptake and in liver with decreased suppression of hepatic glucose output. A critical defect involves pancreatic β cell exhaustion with an inability to secrete sufficient insulin to compensate for insulin resistance in muscle and liver.

Adipose tissue is an additional tissue contributing to T2D pathogenesis and is increasingly thought to play a primary role in CMD and in the development of associated abnormalities in insulin secretion and action.[91] This tissue is a complex organ that is a central regulator of metabolism via the secretion of hormones and is also a depot for fuel storage as triglycerides and FA release to meet energy needs. In the insulin-resistant state, there is a relative redistribution of fat to the intraabdominal compartment, and adipose tissue contains increased numbers of inflammatory macrophages. Crossstalk involving factors from both adipocytes and macrophages result in dysregulated secretion of adipocytokines such as adiponectin, leptin, resistin, PAI-1, and proinflammatory cytokines.[92-95] Adipocytes themselves become insulin-resistant regarding stimulation of glucose transport, lipogenesis, and antilipolysis, and, as a consequence, dysfunctional triglyceride storage leads to increased release of FFAs. These factors can produce oxidative stress and inflammation in other tissues, a systemic clotting diathesis, and accumulation of intracellular fat that increases insulin resistance in muscle and liver and predisposes pancreatic β cells to secretory exhaustion. The superimposition of obesity further expands intraabdominal adipose tissue under these conditions and further worsens these pathophysiological processes. While adipocytes develop defects in insulin-stimulated glucose transport, the contribution of adipose tissue to total body glucose disposal is minor in comparison to muscle. Therefore, the adverse impact of dysfunctional adipose tissue on glucose homeostatic mechanisms occurs primarily through the abnormal secretion of adipocytokines and FFAs. Adipose tissue may also affect metabolism via additional mechanisms including the release of microsomes containing microRNAs that can influence gene expression in other tissues; local nerve fiber networks that can signal to the central nervous system (CNS), altering sympathetic tone; and release of other biofactors.[91]

In addition to abnormalities illustrated in Fig. 33.3, other defects and dysfunctional tissues contribute to the development of

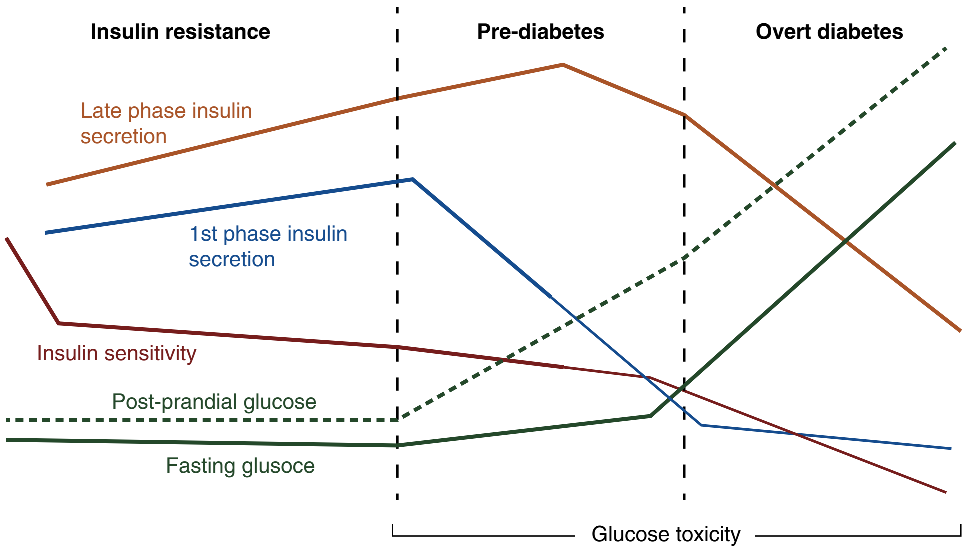


Fig. 33.4 The metabolic evolution of type 2 diabetes (T2D). The development of T2D begins with the insulin resistance phase, where increased first- and second-phase insulin secretion compensates for a decrease in insulin sensitivity in muscle and liver, sufficient to maintain normoglycemia. In the prediabetes phase, insulin secretion becomes inadequate to maintain normoglycemia. First-phase insulin secretion is abrogated, with a rise in postprandial glucoses followed by a decline in second-phase insulin secretion and the appearance of impaired fasting glucose. The diabetes phase features overt hyperglycemia accompanied by progressive β cell exhaustion. In both the prediabetes and diabetes phases, glucose toxicity at the level of the pancreatic β cell contributes to reductions in glucose-induced insulin secretion (also see Chapter 36).

hyperglycemia in T2D,[96] including: (i) increased glucagon secretion by pancreatic α cells, which augments hepatic glucose production[97]; (ii) decreased incretin production contributing to reduced insulin secretory responses[98]; (iii) increased reabsorption of filtered glucose in the renal proximal tubules[99]; (iv) altered neurotransmitter functions (e.g., dopamine) in the CNS affecting the sympathetic nervous system and circadian activities[100]; and (v) dysregulated interaction of satiety hormones acting on central feeding systems that augment caloric intake and cause obesity.[101]

Progression of Abnormal Glucose Tolerance

Metabolic Evolution of Type 2 Diabetes. Metabolic progression to overt T2D represents a continuous process but can be described in terms of transitions involving three pathophysiological phases.

Insulin Resistance Phase. The schematic in Fig. 33.4 illustrates the progression of changes in insulin secretion, insulin resistance, and glycemia towards overt diabetes. In most patients, the progression to abnormal glucose tolerance begins with insulin resistance,[89,102] although there are smaller subsets of lean patients in whom the defect in insulin secretion predominates, and insulin resistance is minimal or nonexistent. Relative differences in insulin sensitivity are evident in children, a trait that is inherited but is also impacted significantly by environmental factors, early life stress, and *in utero* exposures to maternal obesity, gestational diabetes, low birth weight, and other factors. In those individuals who are on the insulin-resistant side of the spectrum, pancreatic β cell secretion is increased over periods of time that can last well into adulthood, resulting in elevated circulating insulin and C-peptide levels. To the degree that hyperinsulinemia compensates for the prevailing degree of insulin resistance, normal values of fasting and postprandial glucose are maintained.[90] While hyperinsulinemia is sufficient to normalize HGO and muscle glucose uptake in the insulin resistance phase, pathophysiological processes inherent in the insulin resistance state are ongoing (see Natural History of Type 2 Diabetes as an End-Stage Manifestation of Cardiometabolic Disease section and Fig. 33.18). Therefore, these patients remain at increased risk for development of T2D, NAFLD, and CVD.

Prediabetes Phase. A subpopulation of patients will progress to prediabetes when insulin secretory responses become insufficient to compensate for progressive insulin resistance.[90] Patients with declining insulin secretion harbor a primary genetic predisposition to β cell exhaustion under the chronic stress of insulin resistance, as well as to the toxic effects of the insulin resistant state caused by elevated FFAs (lipotoxicity), chronically elevated glucose levels (glucotoxicity), oxidative stress, and inflammation (also see Chapter 36). The first detectable abnormality in prediabetes is a rise in postprandial glucose levels, which meet criteria for IGT. This is when the 2-hour glucose value during a 75-gram OGTT rises above the normal value of less than 140 mg/dL (<7.8 mmol/L) to 140 to 199 mg/dL (7.8–11.0 mmol/L). This is followed by IFG with rising fasting glucose values into the range of 100 to 125 mg/dL (6.9 mmol/L).

The most sensitive test for defective β cell function is the intravenous glucose tolerance test (IVGTT), which has been used to study the decline in insulin secretion and to differentiate and quantify early or first-phase and late, or second-phase, glucose-induced insulin secretion. First-phase insulin secretion is a sharp peak of rising then falling insulin levels from 3 to 5 minutes from the time of the glucose injection, while second-phase insulin secretion has a lower and more sustained elevation over baseline that generally lasts approximately 2 hours or more until normoglycemia is restored.[103,104]

Fig. 33.5 demonstrates the loss of first-phase and eventual decline in second-phase insulin secretion with the progressive rise in fasting plasma glucose. As patients progress to IFG, first-phase insulin secretion begins to decline and is completely lost as the fasting glucose rises above 115 mg/dL.[105] Second-phase insulin secretion is initially augmented above normal but then undergoes a progressive decline as fasting glucose values rise above 115 mg/dL. Defective first-phase insulin response is a specific abnormality relative to glucose as a secretagogue, because first-phase responses to nonglucose stimuli, such as isoproterenol, secretin, and arginine, remain intact in hyperglycemic patients[106-108]; Fig. 33.6). Thus, in this sense, the defect in glucose-stimulated first-phase insulin secretion mechanistically represents a problem with glucose sensing rather than a global secretory defect in β cells. This is further illustrated by impaired "glucose potentiation" of responses to nonglucose secretagogues. For example, the acute insulin response to intravenous arginine increases as the glucose concentration is experimentally raised. Investigators have plotted the increase in insulin response as a function of the increasing glucose level and quantified the slope of this relationship as the glucose potentiation slope.[109] This measure of β cell function is reduced in patients with T2D (the reader is referred to Chapter 29 for a full discussion of regulation of insulin, glucagon, and GLP-1 secretion).

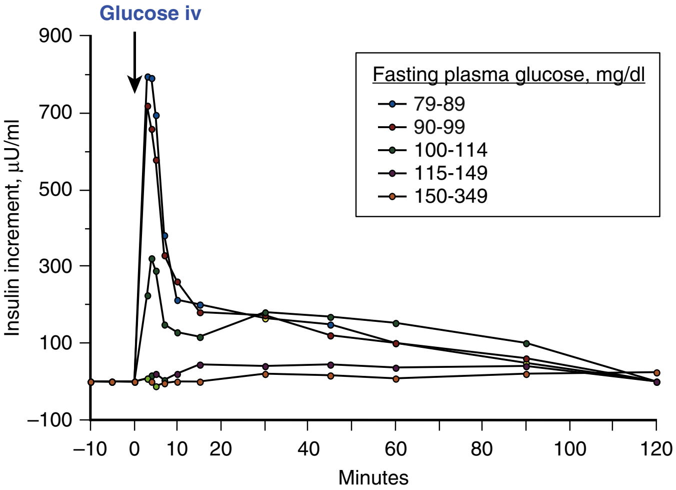


Fig. 33.5 **Intravenous glucose tolerance tests performed in subsets of patients with increasingly elevated fasting plasma glucose.** First-phase insulin secretion occurring 3–5 minutes after an intravenous (iv) glucose challenge is abrogated as fasting glucose levels rise from 90 to 115 mg/dL. Second-phase insulin secretion is initially augmented above normal but then undergoes a progressive decline as fasting glucose values rise above 115 mg/dL. (Data from Brunzell JD, Robertson RP, Lerner RL, et al. Relationships between fasting plasma glucose levels and insulin secretion during intravenous glucose tolerance tests. *J Clin Endocrinol Metab.* 1976;42:222–229.)

Diabetes Phase. With progressive deterioration, prediabetes transitions to diabetes as fasting glucose rises to greater than or equal to 126 mg/dL (greater than or equal to 7.0 mmol/L) and 2-hour OGTT glucose levels are greater than or equal to 200 mg/dL (greater than or equal to 11.1 mmol/L). This is associated with ongoing progressive deterioration of β cell function, with the further decline in second-phase insulin secretory responses often in relationship to the degree of hyperglycemia. There are also further increases in HGO. The continued worsening of these defects continues over the duration of T2D[110] and is responsible for progressive hyperglycemia and the ongoing need for intensification of diabetes treatments. Basal or fasting insulin levels may be normal or elevated in T2D; however, it is important to consider that patients are hyperglycemic. When nondiabetic people are made hyperglycemic by intravenous infusion of glucose, circulating insulin levels are much higher than those found in T2D,[109,111] indicative of the underlying impairment in insulin secretion.

Responses to Oral Glucose and Meals.

Fig. 33.7 illustrates the spectrum of glucose and insulin responses during OGTTs in groups of individuals with progressive glucose intolerance.[112] Normoglycemic individuals have the lowest glucose values and intermediate insulin levels. With progression to IGT, postingestion glucose levels are elevated despite being accompanied by the highest levels of insulin with a delay of the peak insulin response. However, once fasting glucose levels rise above normal in IFG, all glucose levels become more markedly increased, and insulin levels fall with a further delay in the insulin peak. The slower rise in insulin levels and delay in the peak response indicate that there is a defect in early-phase insulin secretion and more sluggish responses to glucose, analogous to the loss of first phase secretion with the IVGTT. Patients with diabetes have the highest level of hyperglycemia and profoundly diminished stimulated insulin levels, with minimal stimulation by oral glucose. The responses to mixed meals largely recapitulate the OGTT data, with one exception. While β cell responses to glucose are impaired, secretion stimulated by nonglucose secretagogues such as amino acids is preserved,[106-108] particularly in patients with mild or shorter-duration diabetes. These patients have more robust late-phase insulin responses to mixed meals from 2 to 4 hours after the meal, even if peak responses continue to be delayed, and the relative degree of glucose intolerance with mixed meals is less severe than that observed during OGTTs. In patients with longer

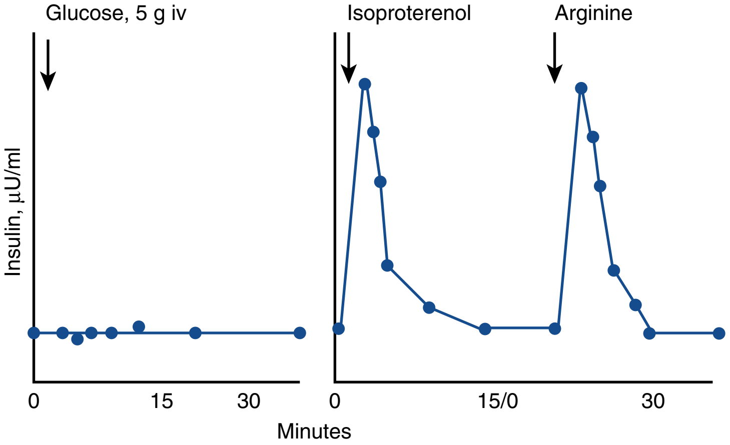


Fig. 33.6 **First-phase insulin secretory responses to glucose, isoproterenol, and arginine in patients with hyperglycemia.** First-phase secretion is absent in response to intravenous (iv) glucose but intact following administration of iv isoproterenol or arginine, indicating that there is a predominate defect in glucose sensing by the pancreatic β cell. (Data from Robertson RP, Porte Jr D. The glucose receptor. A defective mechanism in diabetes mellitus distinct from the beta adrenergic receptor. *J Clin Invest.* 1973;52:870–876.)

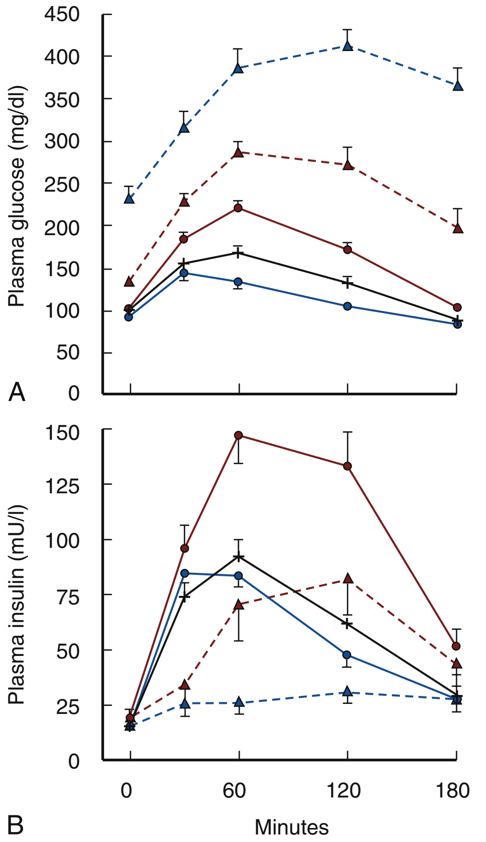


Fig. 33.7 **Glucose and insulin responses during oral glucose tolerance tests.** The data show mean ± standard error values for plasma glucose (**A**) and insulin (**B**) in normal controls *(blue circles)*, borderline glucose intolerance *(cross, black line)*, impaired glucose tolerance *(red circles)*, fasting hyperglycemia 110–150 mg/dL *(red triangles, broken line)*, and fasting hyperglycemia >150 mg/dL *(blue triangles, broken line)*. (Data from Reaven GM, Olefsky JM. Relationship between heterogeneity of insulin responses and insulin resistance in normal subjects. *Diabetologia*. 1977;13:201–206.)

duration of diabetes and more severe hyperglycemia, insulin responses to both mixed meals and glucose remain diminutive.

Progression of Prediabetes to Type 2 Diabetes. The proportion of patients with prediabetes (IFG, IGT, or both) who progress to T2D can vary depending on ethnicity, environmental factors, biological factors, and the methods of assessment. In the Diabetes Prevention Program, patients with IGT developed T2D at a rate of 10% per year.[113] In a metaanalysis, annualized incidence rates of diabetes were 4% to 6% for isolated IGT, 6% to 9% for isolated IFG, and 15% to 19% in those with combined IFG and IGT.[114] Over time, up to 70% of patients with prediabetes will progress to T2D, while others will remain with prediabetes or revert to normal glucose tolerance. The presence of metabolic syndrome and its composite traits also impact T2D risk. Risk of incident diabetes rises progressively over 40-fold from the lowest risk category comprised of obese individuals with no traits of metabolic syndrome, to isolated IFG or IGT, up to the highest risk in obese patients who meet criteria for both metabolic syndrome and prediabetes.[5] In this way, risk stratification can identify those at highest risk for application of more intensive interventions for diabetes prevention.

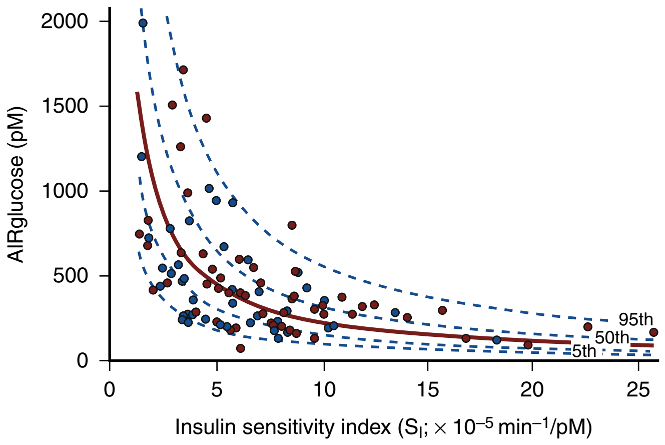


Fig. 33.8 **Relationship between insulin secretion and insulin sensitivity in healthy normoglycemic individuals: the disposition index.** The hyperbolic plot relates the acute insulin response (AIR) to intravenous glucose against the insulin sensitivity index and shows the best fit relationship (50th percentile) as the solid curve, and dashed curves for the 5th, 25th, 75th, and 95th percentiles. Normoglycemia is maintained, because changes in insulin sensitivity are balanced by a reciprocal and proportionate change in β cell function. (Data from Kahn SE, Prigeon RL, McCulloch DK, et al. Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects: evidence for a hyperbolic function. *Diabetes*. 1993;42:1663–1672.)

Relationship Between Insulin Secretion and Insulin Resistance. There have been academic arguments about whether insulin resistance or impaired insulin secretion is the primary defect in T2D. Some investigators have contended that insulin resistance is primary, because insulin resistance occurs early and can predict the future development of T2D. However, this may constitute a false argument. Insulin resistance must be enjoined by impaired β cell secretion for the development of T2D, which does not occur in most obese or insulin-resistant individuals.[115] Rather, individuals who progress to diabetes are those who inherit subsets of susceptibility genes that adversely affect β cell function under the chronic stressor conditions of the insulin-resistant state. These latter patients fail to secrete sufficient insulin to maintain normoglycemia. This is best illustrated conceptually by the disposition index shown in Fig. 33.8. In normoglycemic individuals, the relationship between measures of β cell function and insulin sensitivity is hyperbolic as insulin secretion increases to match progressive declines in insulin sensitivity.[116-118] In patients with prediabetes or T2D, insulin secretion falls below that required to maintain normoglycemia at any level of insulin resistance. The product of the measures of insulin sensitivity and secretion is referred to as the disposition index, and low values that fall below the hyperbolic curve regardless of the degree of insulin resistance are required for development of hyperglycemia. The disposition index predicts progression to overt diabetes among insulin-resistant individuals[119] and has a high heritability index of 0.67.[120] The reduced compensatory capacity of the β cell in insulin-resistant patients with IGT is also manifest in patients undergoing graded glucose infusions. Patients with IGT secrete less insulin at each increasing level of induced hyperglycemia than do normoglycemic individuals matched for similar degrees of insulin resistance and obesity.[121]

In addition to diminished glucose-induced insulin secretion, there are other manifestations of abnormal β cell function under conditions of chronic stress from insulin resistance. First, the proportion of proinsulin that is secreted relative to insulin is increased in hyperinsulinemic states, perhaps reflecting incomplete processing of the proinsulin molecule prior to secretion.[90] Second, insulin secretion is normally characterized by ultradian oscillations occurring every 80 to 150 minutes, with superimposed rapid insulin pulses occurring every

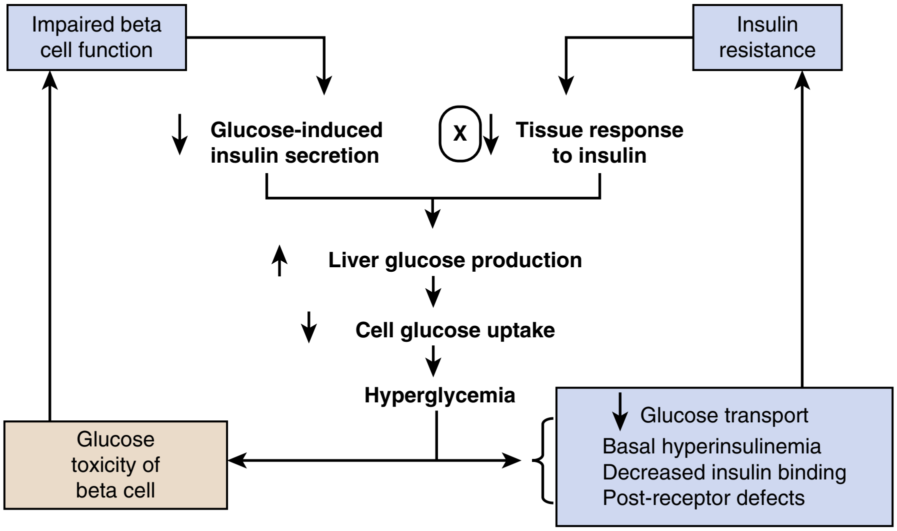

Impaired beta cell function
Insulin resistance
↓ Glucose-induced insulin secretion
X
↓ Tissue response to insulin
↑ Liver glucose production
↓ Cell glucose uptake
Hyperglycemia
Glucose toxicity of beta cell
↓ Glucose transport
Basal hyperinsulinemia
Decreased insulin binding
Post-receptor defects

Fig. 33.9 **The "two hit" formulation for the development of type 2 diabetes.** Both insulin resistance and impaired β cell function are required for the development of hyperglycemia.

5 to 15 minutes.[122] These pulses become less frequent and irregular in patients with T2D and in their nondiabetic relatives.[123,124] Finally, amylin is cosecreted with insulin from the β cell. In T2D, amylin can aggregate as amyloid fibrils in pancreatic islets, and accumulations of islet amyloid peptide (IAPP) are found in most patients with T2D at autopsy. While the clinical significance is controversial, IAPP is toxic to β cells *in vitro* and could contribute to a reduction in β cell mass via apoptosis.[125]

In summary, there are several salient observations regarding glucoregulatory abnormalities leading to T2D. First, the progression of abnormal glucose tolerance requires two basic defects, the presence of insulin resistance and the predisposition to β cell failure over time, consistent with the "two hit" formulation for diabetes pathogenesis (Fig. 33.9). If insulin secretion remains robust and sufficient to compensate for insulin resistance, then normoglycemia will be sustained. Second, the contribution of HGO in driving fasting glucose values to higher levels is critical for progressive hyperglycemia. Thirdly, diabetes is itself a progressive disease with further worsening of insulin resistance and deterioration of β cell secretory capacity, which is attributable in part to glucose toxicity and chronic oxidative stress[126] (see Fig. 36.5).

Reversibility of Defects in Insulin Secretion and Insulin Sensitivity in Type 2 Diabetes. Insulin resistance and impaired insulin secretion in T2D are at least partially reversible following an overnight infusion of insulin to lower glucose levels to the normoglycemic range[127] or a period (~2–3 weeks) of intensive therapy that produces normalization or near normalization of glycemia.[128] The improvements with intensive treatment are achieved whether euglycemia is achieved by insulin therapy, oral agents, or weight loss. The reversibility of defects in insulin secretion and sensitivity are consistent with the concept of glucose toxicity.[128-133] In poorly controlled T2D, restoration of euglycemia enhances β cell function primarily due to improved glucose sensing with restoration of insulin secretion. Regarding insulin resistance, exposure to high glucose decreases activity of the insulin-responsive glucose transport system both in cultured calls and *in vivo* in skeletal muscle, and these defects are reversible in a euglycemic environment. Therefore, in T2D, there are two components to overall insulin resistance; the first is the insulin resistance of CMD that may antedate the development of T2D, and the second is the additive effect of glucose toxicity that can be reversed by euglycemic therapy.[128]

Weight loss also leads to improvements in insulin secretion and sensitivity and can effectively be used to prevent progression to T2D in patients with metabolic syndrome or prediabetes, or to treat diabetes once the disease is present. In all instances, weight loss in T2D leads to better glycemic control with less need for diabetes medications. In the Look AHEAD Study, there were progressive improvements in HbA1c, fasting glucose, blood pressure, triglycerides, and high-density lipoprotein–cholesterol with increasing degrees of weight loss above 5%.[134] Substantial rates of diabetes remission have been observed with greater than or equal to 7% to 10% weight loss, whether achieved by lifestyle interventions including a very low calorie diet,[135] with obesity medications,[136] or following bariatric surgery.[137] Following bariatric surgery, the majority of patients with T2D require reductions or cessation of diabetes medications within days after surgery even before weight loss occurs. This is due to an increase in insulin sensitivity associated with marked reductions in intracellular lipid in muscle and liver, as has been observed within days of instituting a very low calorie diet.[138,139] Insulin secretion is also enhanced following surgery, and longevity of these improvements is the major determinant of the period of time T2D remains in remission (see Chapter 27 for further information about bariatric surgery for T2D).

KEY POINTS

- Four major tissues contribute to progression from normoglycemia, to prediabetes, to type 2 diabetes: impaired insulin-stimulated glucose uptake into skeletal muscle, impaired suppression of hepatic glucose output by insulin, diminished glucose-induced insulin secretion from the pancreas, and dysfunctional adipose tissue characterized by inflammation and abnormal release of adipocytokines and free fatty acids.

KEY POINTS

- The two essential causes of hyperglycemia in type 2 diabetes are islet β cell dysfunction and insulin resistance. First-phase insulin responses deteriorate when fasting glucose is greater than 100 mg/dL and disappear when it exceeds 115 mg/dL. Muscle, liver, and adipose tissue are the dominant sites for insulin resistance.

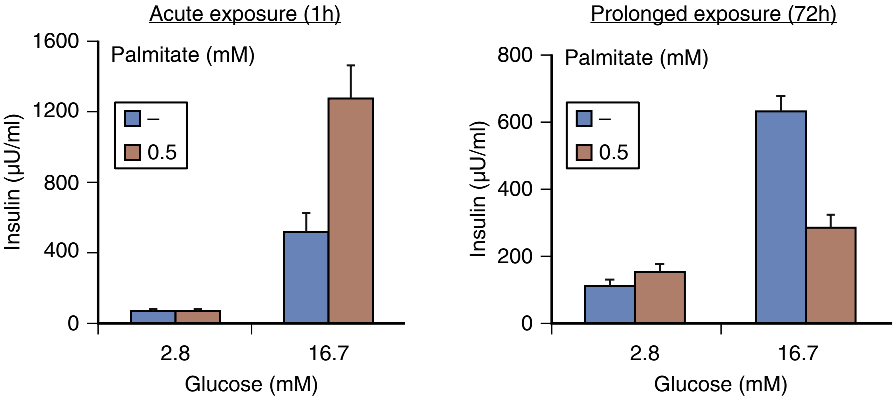


Fig. 33.10 **Effects of acute and chronic exposure to palmitate on glucose-induced insulin secretion in cultured β cells.** Acute (1 hour) exposure of cells to palmitate amplifies the insulin secretory response to glucose (16.7 mM), while chronic exposure (72 hours) impairs glucose-induced insulin secretion compared with controls. (Data from Briaud I, Harmon JS, Kelpe CL, et al. Lipotoxicity of the pancreatic beta-cell is associated with glucose-dependent esterification of fatty acids into neutral lipids. *Diabetes*. 2001;50:315–321.)

Remission of diabetes and its duration, as well as defects in insulin secretion and action, are greater in patients with shorter duration of diabetes (e.g., <5 years) as opposed to longer periods of sustained hyperglycemia (e.g., >10–15 years).[137] While mechanisms are not clear, it appears that the adverse effects of glucose toxicity and lipotoxicity become less reversible with time.

Mechanisms Causing Pancreatic β Cell Dysfunction

A general and essential focus for endocrinologists is to identify mechanisms of disease, which in turn paves the way to new therapies for treatment and prevention. However, because T2D is a clinical diagnosis based primarily on hyperglycemia and its secondary complications, it is not possible to develop a list of single mechanisms for T2D, with the exception of less common monogenic forms, such as patients with MODY (see Chapter 31). Nevertheless, glucose toxicity is an important consequence (not a cause) of chronic hyperglycemia that is responsible for oxidative damage in body tissues, including the pancreatic β cell. In the β cell, there are metabolic processes that normally form reactive oxygen species (ROS), which in physiologic concentrations are known to contribute to health in many ways, such as supporting gene transcription and killing dangerous bacteria. However, studies using *in vitro* islets and intact rodents have shown that excessive ROS concentrations in the pancreatic islet cause β cell apoptosis and adversely affect β cell replication.[140,141] The preferred metabolic pathway for metabolism of glucose is via the Krebs cycle and oxidative phosphorylation. However, when the β cell is presented with hyperglycemia, excessive glucose entry into the cell results in glucose catabolism via multiple alternative pathways in which ROS levels rise to deleterious levels. This is particularly dangerous for β cells because they do not contain the normal complement of endogenous antioxidant enzymes compared to other tissues (the reader is referred to Chapter 36 for a full discussion of the molecular mechanisms of β cell glucose toxicity and chronic oxidative stress).

Importantly, the dimension of time is highly essential to the development of glucotoxicity and lipotoxicity. While brief exposure of β cells *in vitro* to elevated concentrations of glucose or FFAs actually augment glucose-induced insulin secretion, it is only with prolonged exposure of β cells to glucose and FFA that deleterious effects on glucose-induced insulin secretion are seen[142] (Fig. 33.10). Prolonged, time-related exposure of β cells to glucose and FFAs is accompanied by elevated intracellular markers of ROS, which are mechanistic in causing cell damage and functional failure.[142] Such failure, however, is reversible if the high concentrations of glucose and FFA are removed. Unger originally popularized the term lipotoxicity,[144] and Prentki and Corkey[145] have proposed the term glucolipotoxicity to allow for the possibility that prolonged elevated glucose and lipid levels together account for deterioration of β cell function in T2D (Fig. 33.11). This relationship was examined in studies of Zucker diabetic fatty (ZDF) rats that develop hyperglycemia and hyperlipidemia when fed a high-fat diet[146] (Fig. 33.12). One group of animals was treated with bezafibrate to lower lipid levels, and another group was treated with phlorizin (a sodium-glucose cotransporter inhibitor) to lower glucose levels during ingestion of the high-fat diet. Adverse effects were assessed by measurements of β cell insulin mRNA levels. Phlorizin prevented the development of hyperglycemia and prevented the decrease in insulin mRNA levels. Bezafibrate prevented the elevation of lipid levels but did not prevent the decrease in insulin mRNA levels. Hence, in this animal model of T2D, the functional damage to the β cells caused by the high-fat diet were more likely to be due to glucose toxicity rather than lipotoxicity.

KEY POINTS

- As hyperglycemia becomes chronic, resultant glucose toxicity impacts the β cell via oxidative stress, causing defects in *PDX-1* and *MAFA* gene expression, which are required for normal insulin synthesis.

Studies have not been performed in humans to ascertain whether glucose toxicity, lipotoxicity, or a combination of the two is necessarily the dominant means by which hyperglycemia-induced ROS damages β cells. However, it is known that supraphysiological levels of blood glucose have adverse effects on the nuclear binding of two very important insulin gene regulatory proteins, Pdx-1 and MafA. These problems arise by a posttranscriptional defect in *PDX-1* gene expression[147] and a posttranslational defect in MafA gene expression[148] (see Chapter 36), as shown in β cell lines and *in vivo* in rodents. In other studies, it has been shown that β cells of ZDF rats undergo damage as a consequence of a very brief exposure to a high-fat diet. During this time blood glucose levels rise, as do β cell ROS concentrations. These adverse changes are accompanied by damage to β cell structure and function, which are recovered after discontinuation of the high-fat diet for 3 weeks when blood glucose levels return to normal.[149] This recovery is mediated by translocation of cytoplasmic inactive nuclear factor erythroid 2–related factor 2 (Nrf2) to the nucleus, where it then becomes activated

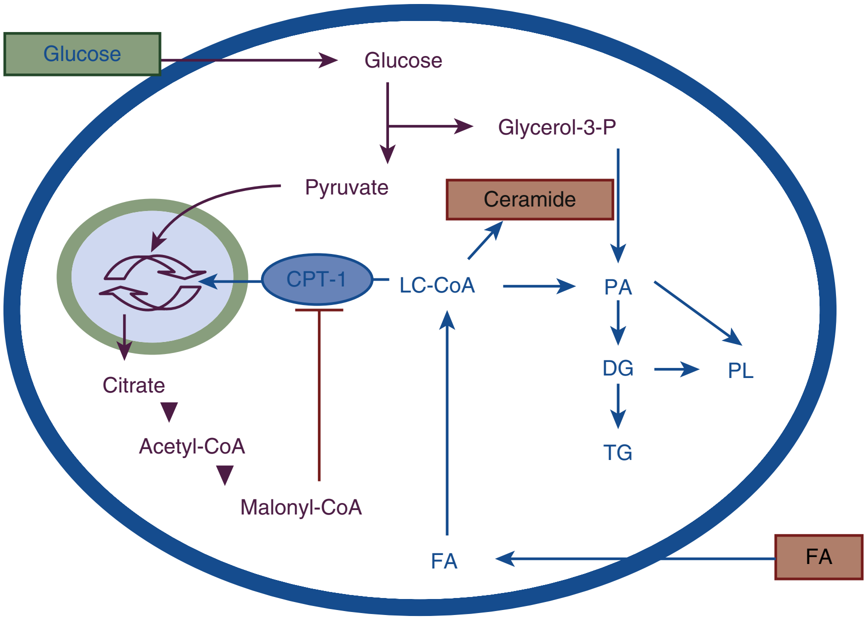


Fig. 33.11 **Potential mechanisms supporting the contention that the combination of elevated glucose and lipids ("glucolipotoxicity") accounts for β cell dysfunction in type 2 diabetes.** *PA*, Phosphatidic acid; *PL*, phospholipid; *DG*, diacylglycerol; *TG*, triacylglycerol; *FA*, fatty acid. Excessive metabolism of glucose leads to increased formation of malonyl CoA, which inhibits entry of long-chain fatty acids into the mitochondrion via carnitine palmitoyltransferase 1 (CPT-1) and subsequent oxidation. High circulating levels of long-chain fatty acids (LC) enter the cell; however, the LC-CoAs are unable to enter the mitochondrion, and instead lead to an increase in ceramide levels, which can cause formation of reactive oxygen species and oxidative stress. (Modified from Prentki M, Corkey BE. Are the beta-cell signaling molecules malonyl-CoA and cystolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM? *Diabetes*. 1996;45:273–283.)

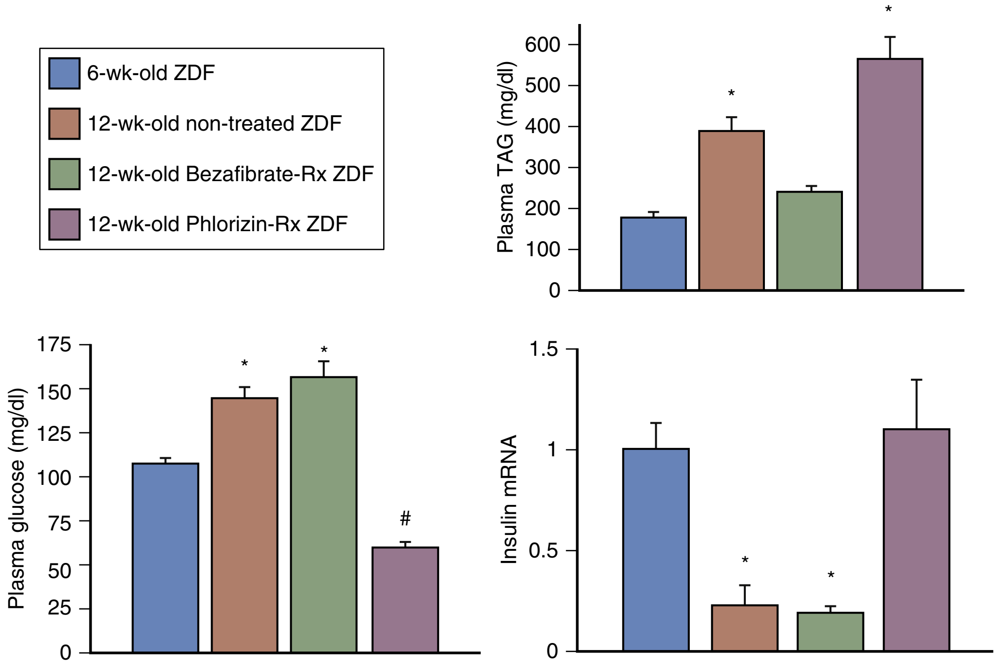


Fig. 33.12 **Effects of phlorizin and bezafibrate on plasma glucose, triglyceride levels, and insulin mRNA expression in β cells in Zucker diabetic fatty (ZDF) rats fed a high-fat diet.** The decline in insulin mRNA levels was only prevented when phlorizin was used to avoid hyperglycemia, not when bezafibrate was used to prevent the rise in plasma triacylglycerol (TAG) levels. The data are consistent with the conclusion that β cell injury is due to hyperglycemia (glucotoxicity) but not due to lipid elevations (lipotoxicity). (Data from Harmon JS, Gleason CE, Tanaka Y, et al. Antecedent hyperglycemia, not hyperlipidemia, is associated with increased islet triacylglycerol content and decreased insulin gene mRNA level in Zucker Diabetic Fatty rats. *Diabetes*. 2001;50:2481–2486.)

and stimulates production of hemoxygenase. Because no changes were seen in rates of apoptosis or cell proliferation throughout these studies, it was concluded that endogenous self-repair of oxidative damage by existing β cells via increased generation of endogenous antioxidants had occurred (see Chapter 36).

THE PATHOPHYSIOLOGY OF INSULIN RESISTANCE

Measurement of Insulin Sensitivity in Humans: Defect in Muscle Glucose Uptake

Insulin resistance is a trait that develops early in life, if not *in utero*. The interaction of genetic, epigenetic, environmental, and behavioral factors produces a wide range of individual variability in systemic insulin sensitivity. In research, the gold standard for measuring insulin sensitivity is the hyperinsulinemic euglycemic clamp. In this technique, an insulin infusion establishes steady-state hyperinsulinemia, and glucose is infused at a variable rate to maintain a target level of euglycemia. At high insulin levels when HGO is completely suppressed, the rate of glucose infusion equals the glucose disposal rate in tissues. At submaximal steady-state serum insulin concentrations, when HGO is not completely suppressed, it is necessary to coinfuse a glucose tracer (i.e., a radioactive or stable isotope–labeled glucose). Dilution of the tracer reflects the amount of unlabeled glucose being infused during the clamp plus the appearance of glucose from the liver. Subtraction of the known amount of infused glucose from the total glucose appearance calculated from dilution of tracer provides a measure of HGO. This allows investigators to calculate both the degree of hepatic suppression and total glucose uptake or disposal. The preponderance of glucose uptake under these conditions is into skeletal muscle, and, when normalized for the amount of lean body mass, is a measure of peripheral insulin sensitivity.

The clamp technique allows for the measurement of insulin sensitivity at multiple steady-state serum insulin concentrations and can be used to construct dose-response curves for insulin-mediated glucose uptake, as well as suppression of HGO. As shown in Fig. 33.13, patients with IGT have a glucose disposal curve that is shifted to the right without a pronounced decrease in the maximal response, in this study reflecting diminished sensitivity.[150] Patients with T2D have profound reductions in glucose uptake at all insulin concentrations, including a decrease in maximally stimulated glucose uptake indicative of a decrease in maximal responsiveness. The insulin resistance in patients with obesity and T2D is only slightly worse than that observed in lean patients with diabetes. This same approach for assessing the dose-response of insulin action can be employed before and after intensive therapy to illustrate reversibility of insulin resistance following a period of euglycemic therapy. In Fig. 33.14, patients with T2D were studied using multiple insulin dose clamps in the untreated state, managed with intensive insulin therapy by continuous subcutaneous infusion for 3 weeks, withdrawn from treatment, and then restudied.[128] The period of euglycemia resulted in marked increments in both sensitivity (shift to the right of the dose-response curve) and maximal insulin responsiveness, although patients did not achieve the higher glucose uptake rates observed in normoglycemic controls.

Insulin sensitivity as measured by hyperinsulinemic clamps varies by over 5-fold among individuals without diabetes.[25] Those individuals at the more insulin-resistant range of the spectrum will preferentially develop other properties of the insulin-resistant state and CMD (see Natural History of Type 2 Diabetes as an End-Stage Manifestation of Cardiometabolic Disease section and Fig. 33.18) and are at increased risk of progressing to T2D. Obesity is not necessary or sufficient as a cause of diabetes, as lean individuals develop T2D, and other patients with obesity are insulin-sensitive and not at increased risk of CMD

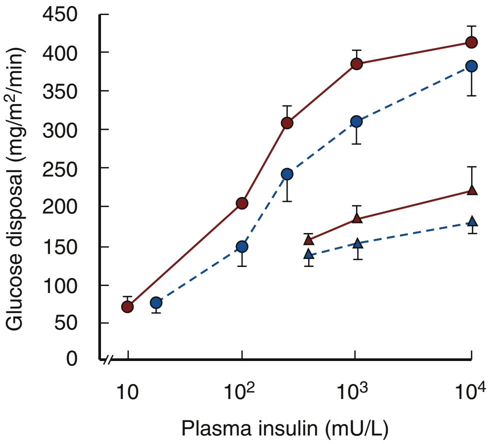


Fig. 33.13 **Dose-response curves for insulin-stimulated glucose disposal using the hyperinsulinemic euglycemic clamp method.** Mean values of glucose disposal or uptake rates expressed as mg glucose per meter squared body surface area per minute are shown at increasing steady-state serum insulin levels. The group with impaired glucose tolerance *(blue circles)* has a rightward shift in the dose-response curve (i.e., decreased sensitivity) compared with controls *(red circles)*. Lean *(red triangles)* and obese *(blue triangles)* patients with type 2 diabetes have both a rightward shift and a reduction in the maximal insulin response (i.e., decreased sensitivity and responsiveness). (Data from Kolterman OG, Gray RS, Griffin J, et al. Receptor and post-receptor defects contribute to the insulin resistance in noninsulin-dependent diabetes mellitus, *J Clin Invest.* 1981;68:957–969.)

(i.e., the metabolically healthy obese).[5] However, obesity and overweight can worsen insulin resistance and are present in the majority of patients with T2D.

Other methods for quantifying insulin sensitivity in humans include the insulin tolerance test, where a fixed amount of insulin is infused, and the rate of decline in plasma glucose is assessed as a measure of insulin sensitivity. Bergman and colleagues have modeled plasma glucose and insulin levels during an intravenous glucose tolerance test to generate an index of insulin sensitivity referred to as the minimal model.[151] Another method involves the suppression of endogenous insulin secretion using somatostatin followed by fixed rate infusions of insulin and glucose.[152] In this scenario, the steady-state plasma glucose concentration reflects the action of insulin at the serum level generated by the insulin infusion, and is used as a measure of *in vivo* insulin sensitivity. All these methods represent research methodologies and are not feasibly performed in clinical practice. To address this need, investigators have established various indices based on ambient glucose and insulin levels either under fasting conditions, such as the homeostatic model assessment,[153] or during an OGTT. However, these indices are variably predictive of clamp measurements of insulin sensitivity due to the fact that there are multiple physiological and pathophysiological determinants of circulating glucose and insulin independent of insulin sensitivity.[154] The development of metabolic syndrome traits as a clinical manifestation of the insulin-resistant state is perhaps the most immediate way to detect the present of insulin resistance in the clinic. Approaches such as CMD staging, which is based on the presence and severity of metabolic syndrome traits, can be used to quantify diabetes risk and is proportional to the degree of insulin resistance.[5,155]

Glucose Transport, Oxidative Metabolism, and Nonoxidative Metabolism

The reduction in insulin-stimulated glucose uptake into muscle during hyperinsulinemic clamp studies could be due to defects in glucose

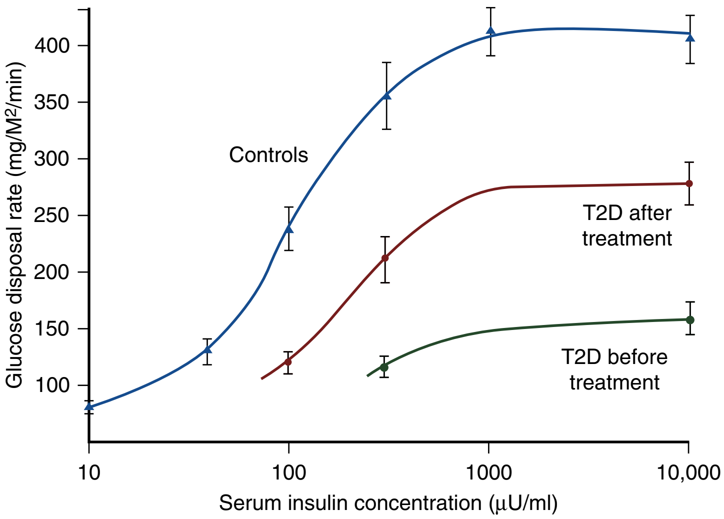


Fig. 33.14 **Insulin dose-response curves for stimulation of glucose disposal in normoglycemic controls and patients with type 2 diabetes (T2D) before and after intense euglycemic therapy.** Patients with T2D were studied in the untreated state and after a 3-week period of intensive glycemic control achieved using continuous subcutaneous inulin infusion. Euglycemic clamp procedures were used as described for Fig. 33.13. Insulin resistance in the untreated patients with T2D, manifested as a rightward shift in the dose-response curve and a decrease in maximal responsiveness, was partially reversed by the period of euglycemic therapy; however, patients did not achieve the degree of insulin sensitivity observed in lean controls. (Data from Garvey WT, Olefsky JM, Griffin RF, et al. The effect of insulin treatment on insulin secretion and insulin action in Type II diabetes mellitus. *Diabetes*. 1985;34:222–234.)

transport, formation of glucose-6P by hexokinase, or subsequent intracellular glucose metabolism. In cultured muscle cells and adipocytes, the glucose transport step is rate-limiting for overall insulin-stimulated glucose uptake and metabolism, and the same appears to be true for skeletal muscle *in vivo*. Investigators have used nuclear magnetic resonance spectroscopy to quantify intracellular glucose-6P[156] and biochemical methods to measure intracellular free glucose in muscle, and found both to be low under conditions of high glucose flux into muscle during hyperinsulinemic euglycemic clamp studies. This demonstrated that reduced stimulation of muscle glucose transport is directly responsible for lower glucose uptake rates and insulin resistance, as opposed to intracellular defects in glucose metabolism causing accumulation of unmetabolized glucose or glucose-6P. Nevertheless, defects in cellular glucose metabolism are also present in human insulin resistance.

KEY POINTS

- Skeletal muscle accounts for ~80% of insulin-mediated glucose uptake *in vivo*, and impaired stimulation of glucose transport into muscle is the predominant defect responsible for peripheral insulin resistance.

Once phosphorylated within the cell, glucose can be metabolized to glycogen for storage (i.e., nonoxidative metabolism) or undergo oxidation (i.e., oxidative metabolism). During clamp studies, the application of indirect calorimetry can assess the amount of glucose being oxidized, and nonoxidative glucose metabolism calculated as total glucose uptake minus oxidized glucose. Insulin's ability to stimulate glycogen synthase (GS), which is essential for nonoxidative glucose metabolism as well as oxidative metabolism, are both impaired in human insulin resistance.[157] In skeletal muscle during clamp studies the reductions in nonoxidative metabolism and rates of glycogen formation are explained by reduced glucose-6P availability as a result of the reduction in glucose transport, while the rate of insulin-stimulated glucose oxidation is impaired beyond that attributable to reduced glucose uptake. However, nonoxidative metabolism to glycogen accounts for the greater component of glucose disposal,[158] and the smaller contribution of oxidative metabolism does not impact overall rates of glucose uptake beyond that attributable to the transport defect. From this perspective, reduced rates of insulin-stimulated glucose transport into muscle are fundamentally responsible for insulin resistance.

Hepatic Insulin Resistance

The liver plays a key role in glucose homeostasis by releasing glucose derived from glycogenolysis or gluconeogenesis into the hepatic vein and out to the systemic circulation. Hepatic glucose production is fueled by uptake of FFAs, gluconeogenic substrates, and carbohydrates from the portal vein and hepatic artery. After an overnight fast, 90% of glucose released into the circulation comes from the liver, which is critical for maintaining euglycemia and energy required by the brain and other organs. After ingestion of carbohydrates, HGO falls rapidly, primarily due to the suppressive effects of insulin, and this helps limit the postprandial rise in plasma glucose levels. However, following the digestion and absorption of nutrients, glucose production is restored to meet systemic needs for glucose as insulin levels fall, and driven by the actions of counterregulatory hormones, the most important being glucagon, but also includin epinephrine, cortisol, and growth hormone.

The rate of hepatic glucose production after an overnight fast is increased in patients with IFG and T2D and is highly correlated with the elevation in fasting glucose levels. Fig. 33.15 shows correlation between fasting glucose and HGO in patients with T2D in the untreated state and after 3 weeks of euglycemic therapy achieved by continuous subcutaneous insulin infusion.[128] It is clear that, under fasting conditions, hepatic glucose production is a powerful determinant of the degree of fasting hyperglycemia. A second point is that elevated HGO rates are reduced following intensive treatment, and the reversal of this defect provides for the maintenance of lower fasting glucose levels following a period of euglycemia.

The increments in HGO are enabled by enhanced rates of gluconeogenesis, which in turn is driven by several pathophysiological processes in

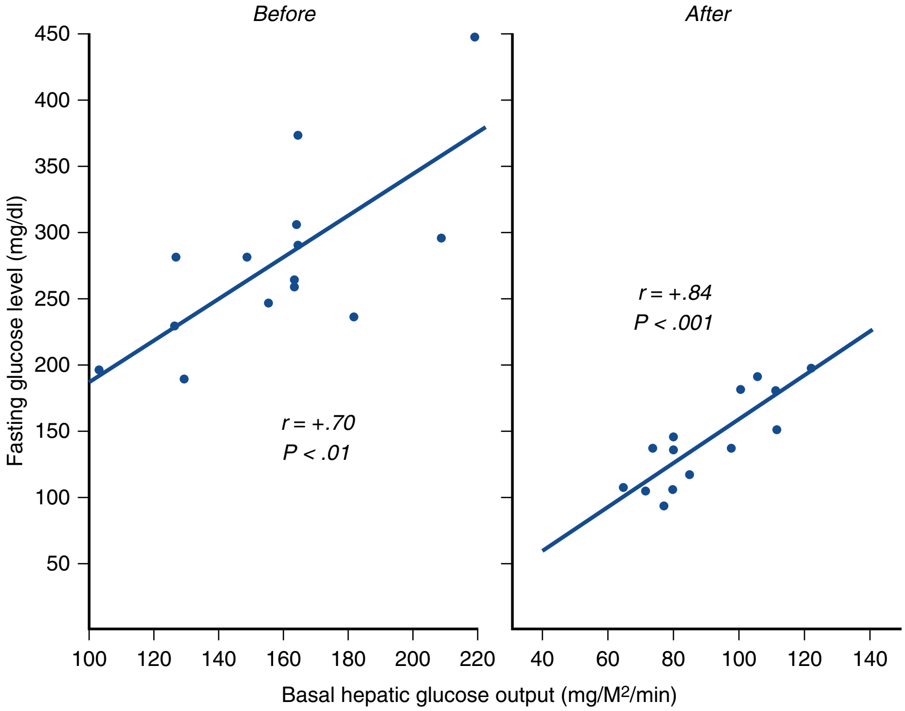


Fig. 33.15 The relationship between fasting glucose levels and basal hepatic glucose output rates in patients with type 2 diabetes (T2D) before and after euglycemic therapy. Patients with T2D were studied in the untreated state and after a 3-week period of intensive glycemic control achieved using continuous subcutaneous inulin infusion. Hepatic glucose output rates are expressed as mg glucose per meter squared body surface area per minute and were determined based on the dilution of an infused glucose tracer. Hepatic glucose production was highly correlated with the high glucose values in untreated patients and the lower glucose values after treatment. (Data from Garvey WT, Olefsky JM, Griffin RF, et al. The effect of insulin treatment on insulin secretion and insulin action in Type II diabetes mellitus. *Diabetes*. 1985;34:222–234.)

diabetes. First, glucagon levels are increased in T2D.[97] Glucagon release from pancreatic α cells is normally suppressed by rising glucose levels, and the presence of hyperglucagonemia in the face of hyperglycemia is indicative of α cell resistance to the inhibitory effects of glucose.[159] The importance of elevated glucagon in T2D is underscored by the observation that inhibition of glucagon secretion by infusion of somatostatin lowers hepatic glucose production and fasting glucose levels. Second, elevated plasma FFA levels, due to the impaired antilipolytic action of insulin in adipose tissue, provide an energy source that helps promote gluconeogenesis. Third, increased flux of gluconeogenic substrates to the liver, including alanine, lactate, and glycerol, act to sustain ongoing high rates of gluconeogenesis. Perhaps the most important cause of increased HGO in T2D is hepatic insulin resistance. In normal individuals, the effective dose for half-maximal suppression (ED50) of HGO is lower (~26 μU/mL) than for stimulation of peripheral glucose uptake (~58 μU/mL).[160] In T2D, the insulin dose-response curve for hepatic suppression is shifted to the right, with an approximately 2-fold increase in the ED50, while hepatic glucose production can still be completely suppressed at higher serum insulin concentrations. This reduction in hepatic insulin sensitivity further promotes the overall increase in glucose production in patients with T2D.

An addition hepatic defect contributes to hyperglycemia in T2D involving uptake of glucose from the portal vein. While skeletal muscle is quantitatively the most important tissue for disposal of an oral glucose load, approximately 20% to 25% of oral glucose is taken up by the liver.[161] Unlike muscle, where insulin directly stimulates glucose transport into the tissue, insulin plays only a permissive role in promoting hepatic glucose uptake. The main determinant of glucose transport into and out of the liver is the glucose concentration gradient between sinusoids and hepatocytes,[162] and reduced expression of hepatic glucose kinase may impair intracellular glucose trapping and glucose flux across this differential.[163] In T2D, uptake of glucose by the liver is impaired, and this contributes to postprandial hyperglycemia.[164]

In summary, in T2D the liver overproduces glucose in the basal state. This is an important cause of fasting hyperglycemia and indicates that the therapeutic suppression of hepatic glucose production is critical for achieving glucose control. High glucose output rates are due to accelerated gluconeogenesis, which is enabled by multiple properties of the diabetic state, including elevated glucagon and FFA levels, increased flux of gluconeogenic substrates to the liver, and a decrease in insulin sensitivity with respect to suppressing hepatic glucose production. Additionally, hepatic glucose uptake is impaired, and this contributes to postprandial hyperglycemia.

KEY POINTS

- Hepatic glucose output rates are elevated in type 2 diabetes and are highly correlated with the degree of fasting hyperglycemia. Increased glucose production is a consequence of hepatic insulin resistance and an inability to suppress augmented rates of gluconeogenesis.

THE MOLECULAR BASIS OF HUMAN INSULIN RESISTANCE IN TARGET TISSUES

General Principles

As a backdrop to discussing molecular events causing insulin resistance, the reader is referred to Chapter 30 for a discussion of general aspects of normal insulin action.

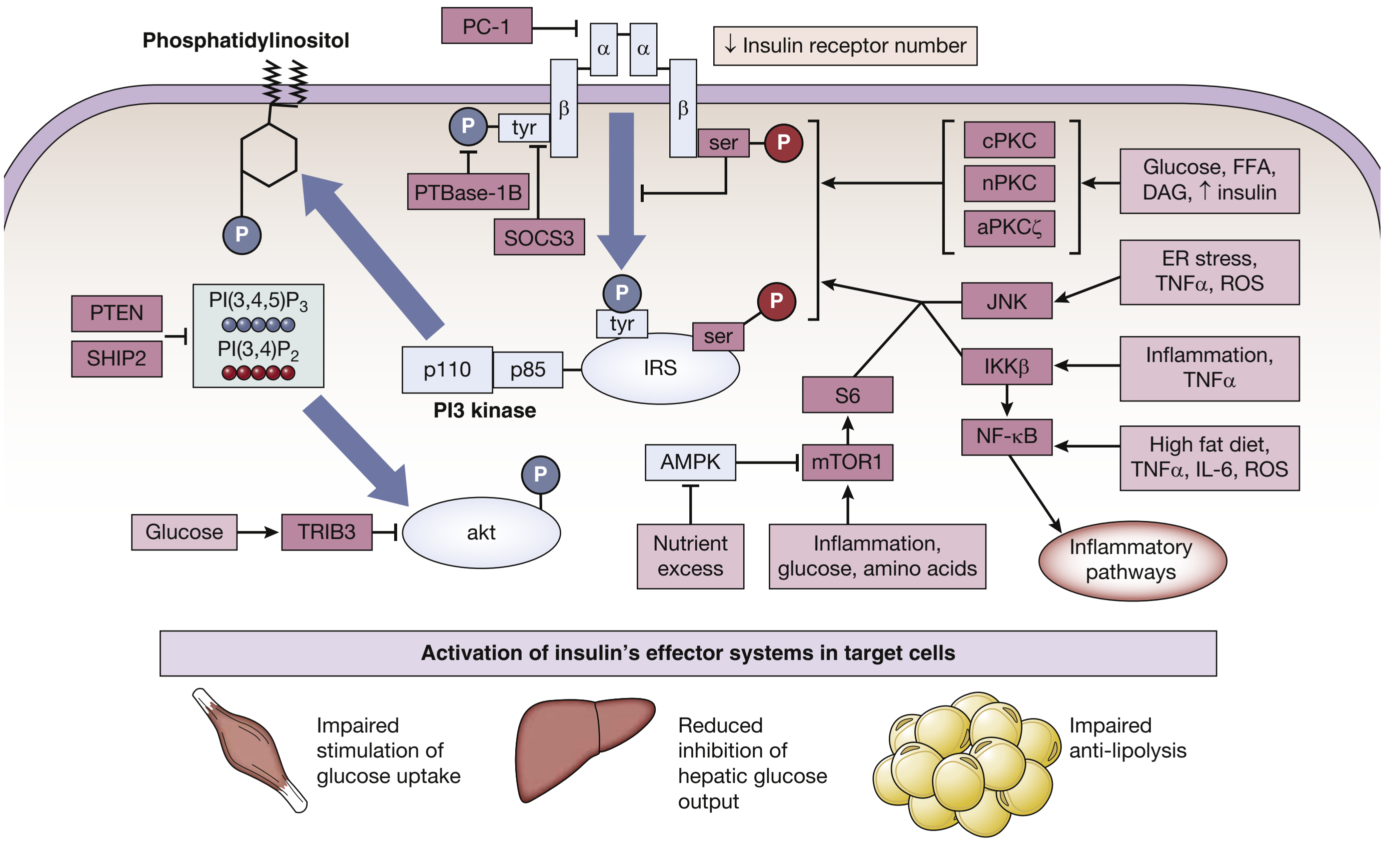

Phosphatidylinositol
PC-1
α
α
β
β
↓ Insulin receptor number
P
tyr
ser
P
cPKC
nPKC
aPKCζ
Glucose, FFA, DAG, ↑ insulin
PTBase-1B
SOCS3
P
ER stress, TNFα, ROS
JNK
P
tyr
P
ser
PTEN
SHIP2
PI(3,4,5)P3
PI(3,4)P2
p110
p85
IRS
PI3 kinase
IKKβ
Inflammation, TNFα
S6
NF-κB
High fat diet, TNFα, IL-6, ROS
AMPK
mTOR1
P
Glucose
TRIB3
akt
Nutrient excess
Inflammation, glucose, amino acids
Inflammatory pathways
Activation of insulin's effector systems in target cells
Impaired stimulation of glucose uptake
Reduced inhibition of hepatic glucose output
Impaired anti-lipolysis

Fig. 33.16 **Mechanisms of human insulin resistance.** *Ser,* Serine; *tyr*, tyrosine; *PC-1*, plasma differentiation factor 1; α and β,subunits of insulin receptor; *IRS*, insulin receptor substrate protein; *akt*, AKT kinase or protein kinase B; *cPKC*, classical protein kinase C isoforms; *nPKC*, novel PKCs; *aPKC*, atypical PKCs; *JNK*, c-Jun N-terminal kinase; *IKKβ*, IκB kinase complex catalytic subunit; *NF-κB*, nuclear factor kappa-light-chain-enhancer of activated B cells; *DAG*, diacylglycerol; *ER*, endoplasmic reticulum; *ROS*, reactive oxygen species; TNFα,tumor necrosis factor α; *IL-6*, interleukin 6; *S6*, S6 kinase; *mTOR1*, mammalian target of rapamycin; *AMPK*, 5′AMP-activated protein kinase; *PTEN*, phosphatase and tensin homolog; *SHIP2*, SH2-containing inositol phosphatase; *TRIB3*, Tribbles homolog 3. The figure shows normal insulin signal transduction in blue from the insulin receptor to activation of Akt and specific biological effector systems in target tissues. Molecules causing insulin resistance are indicated in filled red rectangles, together with environmental influences and processes that promulgate these defects.

Insulin resistance is central to the pathophysiology of T2D and CMD.[165] This term could apply to any of insulin's pleiotropic effects involving substrate metabolism, gene expression, ion flux, vasoregulation, and growth and differentiation. In this context, it applies to abnormalities causing the glucoregulatory defects leading to T2D, in particular molecular events impairing insulin's ability to stimulate glucose uptake and metabolism, reduce hepatic glucose production, and inhibit lipolysis. Approximately 80% of systemic insulin-mediated glucose uptake occurs in skeletal muscle, and insulin resistance at the level of muscle has been demonstrated in T2D using forearm and leg perfusion studies.[166,167] Investigators using *in vivo* techniques such as hyperinsulinemic clamps have demonstrated that systemic measurements of insulin resistance are correlated with reductions in insulin-stimulated glucose transport assessed in fat and muscle tissues removed from insulin-resistant subjects and studied *ex vivo.*[168] The conclusion is that systemic insulin resistance is due to insulin action defects at the level of target cells.

Once insulin enters tissues and arrives at the cell surface, insulin resistance could occur due to defects involving the insulin receptor, signal transduction, or effector systems that mediate insulin's biological effects. These biological processes can mediate normal dynamic functioning of insulin action as organisms adapt to changing physiological conditions, or, when unbalanced, can lead to pathophysiological consequences and the development of insulin resistance. In T2D, there is not one single lesion that predominates as a general cause of insulin resistance. Rather, there is an extensive array of mechanisms that dampen or inhibit signal transduction and impair the biological activity of effector systems, as illustrated in Fig. 33.16. Thus, variable but overlapping subsets of processes and susceptibility genes operate in individual patients to impair insulin action in a manner that contributes to the heterogeneity of T2D.

Insulin Resistance at the Level of the Insulin Receptor

Downregulation of Insulin Receptor Number. The first critical step in insulin signaling is binding of the hormone to cell surface receptors. In most patients with insulin resistance, including T2D and obesity, insulin target cells are observed to have fewer numbers of cell surface insulin receptors. Insulin receptor downregulation can be attributed to chronic exposure to high insulin concentrations, because this increases ligand-mediated internalization of receptors followed by lysosomal degradation, as well as diminished gene expression. The resulting decrement in cell surface insulin receptors without a change in binding affinity has been observed in isolated adipocytes, hepatocytes, and skeletal muscle from patients with obesity and T2D.[169,170] The decrease in receptor number did not approach the minimum complement of 10% to 20% of receptors needed to obtain maximal insulin responses. As a consequence, there remained "spare" receptors for maximal insulin action even after receptor downregulation. In this scenario, receptor downregulation results in a rightward shift of the insulin dose response curve, with decreased effects at submaximal insulin concentrations (e.g., the ED50), while maintenance of maximally stimulated responses is preserved. A decrease in maximal responsiveness would need to be attributed to a post–insulin binding defect, which was also observed in cells from these patients, indicative of downstream defects in insulin action.

Modulators of Insulin Receptor Function. Plasma differentiation factor 1 (PC-1), also referred to as ectonucleotide pyrophosphatase phosphodiesterase 1 (ENPP1), is a membrane glycoprotein with pyrophosphatase activity that appears to act as an intrinsic inhibitor of the insulin receptor tyrosine kinase. PC-1 has been shown to bind to the receptor and prevent the conformational change in receptor β subunits required for insulin binding. Muscle expression of PC-1 is elevated in diabetic patients and in obesity, and it correlates with diminished insulin receptor tyrosine phosphorylation and muscle glucose uptake.[171] Growth factor receptor–bound proteins (Grb10 and Grb14) can also bind the insulin receptor on phosphotyrosine residues, alter receptor tyrosine kinase activity, and disrupt receptor signaling.[172] The physiological role of Grb proteins in humans has not been fully clarified.

Insulin Resistance at the Level of Signal Transduction

Serine-Threonine Phosphorylation of Insulin Receptors and Receptor Substrate

Proteins. In target tissues obtained from patients with T2D, there is a reduction in insulin receptor tyrosine kinase activity and signaling through insulin receptor substrate docking proteins (e.g., IRS-1).[173,174] To probe a causal relationship with human insulin resistance, investigators have studied perturbations that either enhance or diminish *in vivo* insulin sensitivity to determine if there were commensurate changes in phosphorylation/dephosphorylation of the insulin receptor and IRS-1. Using muscle biopsies and euglycemic clamps, investigators have shown beneficial effects of treatment with thiazolidinediones,[175] exercise,[176] caloric restriction,[177] and tight glucose control.[178] In contrast, insulin resistance induced by infusion of a triglyceride emulsion worsened both insulin sensitivity and signaling defects.[170]

Reductions in the activity of the insulin receptor tyrosine kinase and IRS docking proteins is due in large part to serine-threonine phosphorylation of these proteins, which constitutes a major mechanism for negative modulation of insulin signal transduction. Serine phosphorylation diminishes insulin receptor tyrosine kinase activity and decreases receptor–IRS coupling by inhibiting insulin-mediated tyrosine phosphorylation of IRS-1, which is required for subsequent binding and activation of PI-3 kinase and stimulation of glucose transport.[180-183] There are consensus sequences in IRS-1 that make it susceptible to a wide variety of serine-threonine kinases. Several of these kinases have been shown to function as physiological modulators, causing desensitization of insulin signaling pathways under conditions of nutrient excess, inflammation, and cell stress responses. A second abnormality applies to the cell content of IRS-1, which can be diminished due to increased proteasomal degradation. Even so, the defect in tyrosine phosphorylation predominates in impaired signaling through IRS-1 to the degree that the decrease in IRS-1 protein may not be physiologically relevant. Key serine-threonine kinases participating in human insulin resistance include:

Protein Kinase C. PKCs are serine-threonine kinases that can phosphorylate IRS-1 and impair its ability to associate with the insulin receptor and with PI-3 kinase, resulting in desensitization of the PI-3 kinase pathway.[181,183-185] Hyperinsulinemia, hyperglycemia, and elevated circulating FFAs (e.g., nutrient excess) leads to increased intracellular diacylglycerol (DAG), which in turn activates conventional and novel PKC isoforms (e.g., PKCθ), principally via recruitment to the plasma membrane. In addition, insulin activates atypical PKCs, such as PKCζ, via the PI-3 kinase pathway, which is also capable of phosphorylating and impairing IRS function.[186]

Tumor necrosis factor-α (TNFα). TNFα is a cytokine produced by immune cells and also by adipocytes and muscle tissue that can mediate serine phosphorylation of IRS-1. TNFα expression is increased in adipose and muscle tissues as a function of insulin resistance, where TNFα appears to operate largely as an autocrine/paracrine factor.[187,188] In adipose tissue, TNFα and other proinflammatory cytokines are produced by adipocytes and macrophages under conditions of obesity and insulin resistance. TNFα activates other pathways that directly mediate increased serine phosphorylation of IRS-1, including PKC isoforms, the stress-induced kinase JNK, and IKKβ.[189-191]

JNK. JNK isoform proteins belong to the extended family of MAPKs and control many cellular functions through regulation of activator protein-1 (AP-1). Insulin has also been shown to activate JNK1 and JNK2, which then display increased serine kinase activity against multiple intracellular substrates, including IRS-1, IRS-2, and Shc.[192] In this way, JNK represents a negative feedback mechanism by which insulin inhibits its own signaling. However, JNK activity is also augmented during cellular stress responses, such as endoplasmic reticulum (ER) stress, and has been found to be increased in insulin-resistant states.[193]

NF-κB. NF-κB is a transcription factor that functions as a proinflammatory "master switch" during inflammation, upregulating the transcription of a wide range of inflammatory mediators.[194] NF-κB is normally retained in the cytoplasm by binding to members of the inhibitor of κB (IκB) protein family. During inflammatory or metabolic stress, NF-κB is activated by the IκB kinase (IKK) complex, which consists of two catalytic subunits, IKKα and IKKβ, and a regulatory subunit, IKKγ. The IKKβ subunit phosphorylates IκB, resulting in its ubiquitination and subsequent proteasomal degradation and in the release of NF-κB. The free NFκB is then able to translocate into the nucleus and activate the transcription of over 100 genes, most of which are proinflammatory.[195] Activation of NF-κB has been shown to accompany insulin resistance following high-fat feeding, in obesity, and in the metabolic syndrome.[196,197] Proinflammatory cytokines such as TNFα and IL-6 can also activate NF-κB. Weight loss, caloric restriction, and exercise training can lead to a reduction in transcription of the proinflammatory genes regulated by NF-κB. Drugs with antiinflammatory properties, such as thiazolidinediones, statins, and salicylates, can inhibit proinflammatory cytokine secretion by interfering with the NF-κB pathway.

IκB kinase-β (IKKβ). In addition to activating NF-κB, IKKβ is a serine kinase that has the ability to directly desensitize insulin signaling through serine phosphorylation of IRS or the insulin receptor.[194,198] The ability of TNFα to desensitize insulin action as a consequence of IRS serine phosphorylation may in part be mediated through IKKβ. An important biological effect of salicylates is the inhibition of the IKKβ, and treatment with high doses of salicylates improves glucose tolerance and enhances insulin sensitivity and secretion in human and rodents.[199-201]

Additional Factors that Inhibit Signaling by Affecting Tyrosine Phosphorylation. Endogenous phosphotyrosine phosphatases (PTPase) are able to dephosphorylate tyrosine residues on the insulin receptor β subunit and IRS docking molecules, resulting in a dampening of insulin signal transduction. Membrane-associated PTPase activity is increased in skeletal muscle from patients with T2D,[202] principally due to increments in cytosolic PTPase-1B and membrane-associated leukocyte antigen-related phosphatase.[203,204]

The suppressor of cytokine signaling-3 (SOCS) family of proteins was originally described as a negative feedback loop for cytokine receptors involving Janus kinase (JaK). The ability of cytokines and growth factors to augment expression of SOCS genes (*CIS* and *SOCS1–7*) via the JaK and signal transducer and activator of transcription (STAT) pathway is described in Chapter 30. SOCS1, SOCS3, and SOCS6 can bind to the C-terminus of the insulin receptor β subunit and block interaction between the insulin receptor and IRS.[204-207] Several factors that induce cellular insulin resistance also induce SOCS3 expression, including TNFα, growth hormone, and leptin.

Mechanisms Causing Downstream Signaling Defects. The reduction in IRS-1 tyrosine phosphorylation decreases the ability of the PI-3 kinase regulatory subunit to bind IRS-1, thus preventing activation of the PI-3 kinase catalytic subunit. Accordingly, insulin-stimulated PI-3 kinase activity in skeletal muscle is reduced in both obesity and T2D in a manner that correlates with the decrease in whole-body glucose disposal. PI-3 kinase acts to increase production of PI(3,4,5)P_3 and the downstream activation of PDK1 and Akt. The phosphatase and tensin homolog (PTEN) can dephosphorylate PI(3,4,5)P_3 on the 3 position, and SH2-containing inositol phosphatases (SHIPs) can dephosphorylate PI(3,4,5)P_3 on the 5 position of the inositol ring; in either instance, the ability of PI(3,4,5)P_3 to stimulate PDK1 is lost. Therefore, PTEN and SHIP can exert negative regulatory influences on insulin signaling with respect to its metabolic actions.[208] The degree to which these phosphatases play a role in human insulin resistance has yet to be fully defined.

The Tribbles (TRB) gene family in mammals is comprised of three proteins that have a truncated kinase domain lacking an adenosine triphosphate (ATP) binding site. Accordingly, TRBs are "psuedokinases" that lack detectable kinase activity but can bind to kinase substrates in phosphorylation cascades and inhibit their phosphorylation. TRB proteins, for example, can bind to Akt/PKB and inhibit its phosphorylation and activation in response to insulin,[209] while downregulation of TRB3 improves insulin sensitivity.[210] TRB3 is a glucose-responsive gene that is upregulated in human muscle as a function of rising glycemia. Importantly, it may mediate glucose-induced insulin resistance ("glucose toxicity") under conditions of sustained hyperglycemia.[211-213]

Insulin Resistance at the Level of Biological Effector Systems

Insulin regulates whole-body fuel homeostasis via specific effects in multiple target tissues. The nature of these biological actions varies dramatically from tissue to tissue, and these variations, for the most part, are not due to differences in insulin signal transmission (described earlier). Rather, tissue-specific insulin effects are principally explained by effector systems that are uniquely expressed in skeletal muscle, adipose tissue, and liver.

Insulin Effector System Defects in Skeletal Muscle. Skeletal muscle accounts for the bulk of insulin-stimulated glucose uptake *in vivo*,[166] and the hallmark of insulin resistance in humans is the impaired ability of insulin to stimulate the glucose transport effector system in skeletal muscle.[214] The major transporter isoforms that mediate glucose transport in cells with an insulin-responsive glucose transport system are GLUT1 and GLUT4. In unstimulated cells, GLUT1 predominates at the cell surface and facilitates glucose diffusion across the plasma membrane into the cytosol, where glucose is rapidly phosphorylated by hexokinase and metabolized. In contrast, over 90% of GLUT4 resides in intracellular vesicles in the basal state. The mechanism by which insulin augments glucose transport activity is by recruiting intracellular GLUT4 vesicles to the plasma membrane, followed by fusion and exposure of the glucose pore to the exofacial surface of the plasma membrane. The process of insulin-mediated GLUT4 translocation is described in Chapter 30.

In lean and obese patients with T2D or metabolic syndrome, insulin's ability to stimulate glucose transport into muscle is impaired due to a decrease in the cell surface appearance of GLUT4 following insulin stimulation.[214] Such an abnormality could be due to either impaired insulin signaling or defects intrinsic to the glucose transport effector system, and, in fact, both abnormalities are operative. Insulin signaling through the receptor/IRS/PI-3 kinase pathway is defective in human muscle. Activation of Akt is downstream of PI-3 kinase, and the data addressing whether Akt phosphorylation is defective in muscle from T2D patients are controversial.[215] Different authors report normal or decreased Akt activation or reduced activation at high, but not low, insulin concentrations, and the role of Akt is further complicated by the fact that only a small fraction of Akt phosphorylation is needed for maximal stimulation of glucose transport.[216] PKC isoforms, in particular PKCλ, are also downstream of PI-3 kinase and upstream of glucose transport stimulation, and there are reports of defects in cellular content and subcellular distribution of atypical PKC isoforms in T2D.[217,218] In any event, downstream insulin-stimulated phosphorylation of AS160 (TBC1D4) and TBC1D1 is impaired, and these Rab GTPase-activating proteins are final or near-final steps in insulin signaling and most proximal to engagement of the glucose transport effector system.[219]

Defects intrinsic to the glucose transport system have also been shown to be an important cause of insulin resistance in muscle and fat tissues from insulin-resistant and diabetic patients. Cellular depletion of GLUT4 transporters is a major mechanism of insulin resistance in adipocytes in obesity and T2D,[220] while GLUT4 expression is relatively normal in skeletal muscle.[221] However, in both muscle and fat, GLUT4 accumulates or is redistributed to a dense membrane compartment under basal conditions, and this abnormality is linked to impaired GLUT4 translocation to the plasma membrane in response to insulin.[221-223] These data are indicative of a trafficking or targeting abnormality impairing GLUT4 translocation in tissues derived from insulin-resistant subjects. Trafficking and translocation of GLUT4 is a complex process, and a long list of proteins involved in vesicle trafficking, membrane fusion, and the cytoskeleton could all be contributory. Additional study of pathways and factors involved in abnormal GLUT4 traffic may elucidate mechanisms causing human insulin resistance and establish targets for drug development.

Muscle contraction results in GLUT4 translocation and stimulates muscle glucose transport activity.[224] This effect occurs without any change in serum insulin concentration and does not involved activation of insulin receptors, PI-3 kinase, or Akt/PKB. In muscle from insulin-resistant humans and rodents, GLUT4 translocation is impaired in response to insulin but normal in response to acute exercise. This response appears to be at least partially dependent on increments in intracellular 5'AMP, and subsequent activation of 5'AMP-activated protein kinase (AMPK).[225] The maintenance of normal glucose uptake with muscle contraction highlights the therapeutic utility of regular exercise in diabetes.

Insulin Effector System Defects in Adipose Tissue. Fat accumulation in adipocytes is determined by the balance between triglyceride synthesis (glucose transport, FA uptake, and lipogenesis) and breakdown (lipolysis and FA oxidation). In regulating both functions, insulin alters the activity of relevant proteins through posttranslational modifications and alterations of gene expression. These proteins constitute the effector system for these biological effects of insulin in adipose tissue. In insulin-resistant individuals, adipose tissue is dysfunctional as an organ responsible for fuel storage and triglyceride turnover, and these defects adversely affect systemic metabolism.

a) **Lipogenesis.** Regarding lipogenesis, insulin augments availability of both glycerol and FAs for triglyceride synthesis by increasing the uptake of glucose in the adipose cell, as well as by activating lipogenic and glycolytic enzymes. One defect compromising lipogenesis is insulin's reduced ability to stimulate glucose transport into adipocytes needed for glycolytic flux to glycerol-3P required for synthesis of triglycerides. Stimulation of glucose transport in adipocytes is diminished due to defects in insulin signaling and GLUT4 translocation, as observed in skeletal muscle, but is further compromised by reductions in the expression and cellular content of GLUT4.[220-223] *De novo* synthesis of FAs for glycerol esterification and triglyceride synthesis may also be impaired. Some studies report that the adipocyte expression of genes involved in lipogenesis and FA esterification (e.g., diacylglycerol acyl transferase) is diminished in patients with obesity and insulin resistance. However, most of the FFAs used for triglyceride storage derive from circulating FFAs or from very low-density lipoprotein (VLDL) acted upon by lipoprotein lipase, and the contribution of *de novo* lipogenesis to adipocyte triglyceride content is small. For this reason, any abnormalities in *de novo* lipogenesis in adipocytes may not contribute heavily to abnormal lipid turnover.

b) **Antilipolysis.** In addition to abnormalities in lipogenesis, adipocytes display a general defect in their ability to store and retain triglycerides in patients with insulin resistance and T2D. Of key importance is the impaired ability of insulin to suppress lipolysis and the release of FFAs.[226] Lipolysis in normal subjects is exquisitely sensitive to inhibition by insulin, such that half-maximal suppression of lipolysis occurs at insulin concentrations well below those needed for significant stimulation of glucose uptake by skeletal muscle. Hormones regulate lipolysis in adipocytes via coordinated action involving two major effector systems, namely, hormone-sensitive lipase (HSL) and perilipins localized to the surface of lipid droplets. Lipolytic and antilipolytic (i.e., insulin) hormones exert opposite effects on HSL and perilipins by determining cyclic adenosine monophosphate (cAMP) availability and protein kinase A (PKA) activity. Catecholamines induce lipolysis and release FFAs from adipocytes by binding to β-adrenergic receptors coupled by heterotrimeric G proteins to adenylate cyclase, which increases production of cAMP and activates PKA. The two main targets for PKA phosphorylation are HSL and the perilipins.[227] The ability of insulin to antagonize hormone-induced lipolysis is to a large extent accounted for by its ability to lower cAMP levels, and thereby reduce PKA activity. The decrease in cAMP is mainly the result of an insulin-mediated phosphorylation and activation of phosphodiesterase 3B (PDE3B) via Akt/PKB.[228] Perilipins are localized at the surface of the lipid droplet in adipocytes and inhibit lipolysis by acting as a barrier against hydrolysis of the triacylglycerol by lipases. When PKA is activated, perilipin becomes phosphorylated and translocates away from the lipid droplet, which allows HSL to hydrolyze the adipocyte triglyceride core.[229] Insulin blocks lipolysis by inhibiting PKA-mediated phosphorylation of HSL and perilipin, thus reducing both HSL activity and its access to triglycerides in the fat droplet.

In insulin-resistant individuals, the dose-response curve for serum insulin versus reductions in FFAs is right-shifted such that FFAs are increased at any given insulin concentration over the broad range of physiological concentrations.[226] The impaired antilipolytic effects of insulin are due to defects in insulin signaling resulting in reduced activation of PDE3B, together with decreased levels of the lipid droplet proteins that protect stored triglycerides from lipolysis. The content of key perilipins and related proteins (e.g., CIDEA) in adipocytes has been shown to have an inverse correlation with lipolytic rates and a positive correlation with plasma glycerol in humans, and these proteins are reduced in patients with obesity and insulin resistance.[230-233] The mechanisms that downregulate adipocyte lipid storage proteins in obesity are not fully understood but may reflect decreased production and action of adiponectin as an autocrine/paracrine factor in adipose tissue[233-235] and attenuated activities of transcription factors that regulate expression of these proteins, such as CHREBPα/CHREBPβ and PPARγ. For example, insulin-sensitizing thiazolidinediones lower FFA levels, augment adiponectin secretion, and induce perilipin via a PPARγ-responsive element in its promoter region.[236]

Dysfunctional white adipose tissue is characterized by a decreased capacity to synthesize, esterify, and store triglycerides in a central lipid droplet, which results in higher levels of circulating FFAs. High FFAs contribute to toxic lipid accumulation of lipid in the liver and skeletal muscle and the exacerbation of insulin resistance in these organs. Cellular levels of lipid metabolites, such as DAG, ceramide, and long-chain fatty acyl-CoAs, are increased and act to impair insulin action and activate transcription factors such NF-κB that augment inflammatory processes. When the abnormalities in lipid turnover are considered together with the dysregulated secretion of adipocytokines in insulin-resistant states, adipose tissue can be viewed both as a central regulator of systemic glucose homeostasis and, when tissue function is impaired, as fundamental in the propagation of insulin resistance and CMD.[237]

Insulin Effector System Defects in Liver. Hepatic insulin resistance refers to impaired suppression of glucose production by insulin. The increase in HGO in T2D is due to a failure of insulin to inhibit both glycogenolysis and gluconeogenesis, and largely determines the severity of fasting hyperglycemia. Insulin regulates hepatic metabolism through acute posttranslational modifications of enzymes (e.g., phosphorylation) and through changes in gene expression. The metabolic enzymatic pathways responsible for glycogenesis/glycogenolysis and glycolysis/gluconeogenesis constitute effector systems for insulin in liver.

a) **Physiological Regulation of Hepatic Glucose Output.** Hepatic glucose production rates represent a balance between the inhibitory effects of insulin that predominate in the fed state and the stimulatory effects of counterregulatory hormones such glucagon, catecholamines, and glucocorticoids that prevail under fasting conditions. With insulin stimulation, the major portion of glucose uptake in liver is stored as glycogen. Insulin promotes glycogen synthesis by activating GS, which adds activated glucosyl groups to growing polysaccharide chains in the final step of glycogen synthesis.[238] Insulin's ability to stimulate GS in the liver requires proximal signaling through tyrosine phosphorylation of the docking protein IRS-2, activation of PI-3 kinase and Akt/PKB, followed by phosphorylation and inactivation of GS kinase-3 (GSK3). This results in reduced phosphorylation of GS, which augments its activity. In contrast, glucagon promotes glycogenolysis via formation of cAMP and activation of PKA, which in turn phosphorylates and activates phosphorylase kinase. Phosphorylase kinase converts inactive glycogen phosphorylase b into its active form phosphorylase a, which is responsible for the release of glucose 1-phosphate from glycogen polymers. Glycogenolysis is also regulated by the activity glycogen-associated protein phosphatases, and in liver the PP1 catalytic subunit is complexed with the G_L glycogen-binding subunit. Glycogen phosphorylase itself acts as a glucose sensor, because it can bind glucose, which renders it a better substrate for the $PP1/G_L$ phosphatase, thereby inactivating phosphorylase (i.e., converting it to its inactive b form) and diminishing glycogenolysis. Because the phosphorylated phosphorylase a acts as an allosteric inhibitor of the GS phosphatase activity of $PP1/G_L$, activation of GS is also enhanced. Thus, high glucose suppresses glycogenolysis and promotes glycogen synthesis in the liver and, conversely, low glucose favors glycogenolysis.

Gluconeogenesis is predominantly regulated through changes in gene expression for two key enzymes: (i) phosphoenolpyruvate carboxykinase (PEPCK), which catalyzes the rate-limiting step in glycogenolysis, and (ii) glucose-6-phosphatase (G-6-Pase), which catalyzes the final step producing free glucose for transport out of liver via GLUT2.[239] Transcription of the *PEPCK* gene is tightly regulated by cAMP. Glucagon increases cAMP, PKA activity, and phosphorylation of the transcription factor CREB, leading to induction of PEPCK, while this pathway is suppressed by insulin. Suppression of HGO by insulin also relies heavily on suppression of *PEPCK* and G-6-Pase gene expression through a complex process involving multiple transcription factors, including FOXO, SREBP-1c, LXR, Sp1, and others.[239-244] Under basal conditions, FOXO1 and FOXO3 proteins reside within the nucleus and actively regulate transcription of the G-6-Pase and *PEPCK* genes. Upon stimulation by insulin, Akt/PKB phosphorylates FOXO family members at conserved serine and threonine residues, which is followed by the dissociation of FOXO proteins from their DNA binding sites and exclusion from the nucleus. The relocation of FOXO results in reduced expression of G-6-Pase and PEPCK. On the other hand, glucagon increases expression of PGC1α via a cAMP-dependent process, and PGC1α participates as a coactivator with FOXO1 to increase the FOXO1-dependent transcriptional activity of G-6-Pase and PEPCK expression.[245] In this role, both FOXO1 and PGC1α cooperate to fully promote the induction of gluconeogenic genes in liver.

The suppressive effects of insulin on gluconeogenesis extend to additional transcription factors, as well as to systemic effects involving gluconeogenic substrates. For example, insulin stimulates SREBP1c, which contributes to the suppression of PEPCK and also promotes glycolysis and lipogenesis via regulation of glucokinase, pyruvate kinase, FA synthase, and acetyl CoA carboxylase.[246,247] In addition to direct hormonal effects on hepatocytes, HGO is diminished by insulin's ability to reduce delivery of gluconeogenic substrates to the liver, such as lactate, amino acids, and FFA. Insulin also minimizes counterregulatory effects of glucagon by inhibiting its secretion from the pancreatic α cell.

b) **Mechanisms of Hepatic Insulin Resistance.** Many of the defects underlying insulin resistance in the liver parallel those described in skeletal muscle and other tissues. These include decrements in insulin receptor tyrosine kinase activity, tyrosine phosphorylation of IRS-2, which serves as the principal IRS docking protein in liver, PI-3 kinase phosphorylation, and activation of AKT. Defects in insulin signal transduction have been attributed to serine phosphorylation of the insulin receptor and IRS-2 as a result of increased intracellular FAs and metabolites that activate PKC and JNK. The signaling defects impair insulin's ability to stimulate glycogenesis and inhibit gluconeogenesis, resulting in unrestrained HGO. The reduced ability of insulin to stimulate GS and promote glycogenesis is due to an inability to phosphorylate and inactivate GSK3. Inhibition of glycogenolysis is similarly impaired via a reduced ability to activate phosphodiesterase, which limits conversion of cAMP to AMP. The resulting higher levels of cAMP activate PKA, causing ongoing phosphorylation and activation of phosphorylase kinase. This allows glycogen phosphorylase a to promote the ongoing release of glucose-1P from glycogen, conversion to glucose-6P, and export of glucose from the liver following the action of glucose-6 phosphatase. In addition, glycogen phosphorylase a retains PP1 in its bound and inactive state, which prevents it from dephosphorylating and activating GS.

KEY POINTS

- Insulin resistance in target tissues involves defects in insulin signal transduction including downregulation of insulin receptor (IR) number, decreased receptor tyrosine kinase activity, and reduced tyrosine phosphorylation of IR substrate (IRS) docking proteins. This is due in part to serine phosphorylation of IR and IRS molecules by inflammation and stress-activated serine kinases. Additional mechanisms of insulin resistance involve defects in tissue-specific biological effector systems.

The defects in insulin signaling also affect the complex transcriptional network that regulates glycolysis and gluconeogenesis. Insulin is less able to inhibit gluconeogenesis in T2D because insulin-induced phosphorylation and nuclear exclusion of FOXO-1, stimulation of SREBP1, and effects on other nuclear factors promoting gluconeogenesis are impaired. At the same time, the effects of counterregulatory hormones to increase hepatic glucose production are unabated. Glucagon levels are increased in T2D and stimulate adenyl-cyclase activity in hepatocytes, leading to increased intracellular cAMP levels. This activates PKA and subsequent phosphorylation and activation of the nuclear transcription factor CREB. CREB is a master positive regulator of gluconeogenic gene expression in conjunction with other nuclear coactivators, including CBP (CREB-binding protein), TORC2 (transducer of regulated CREB activity 2), and other transcription factors, including PGC1α, FOX01, and HNF-4α that positively activate the gluconeogenic program.[248] Once the liver is insulin-resistant, phosphorylation of these cofactors is impaired, resulting in continued high levels of FOX01, TORC2, CBP, and SREBP1 in the nucleus that sustain ongoing gluconeogenesis.

CELL STRESS RESPONSES, NUTRIENT SENSING, AND INSULIN RESISTANCE

The described defects involving the insulin receptor, signal transduction, and hormone effector systems are not self-created, but rather arise as a consequence of cell stress responses that are integral to the insulin resistant state, as well as nutrient stress responses involving nutrient sensor pathways. Fuel metabolism, in particular, can interact with hormone signaling pathways to regulate a broad range of cell functions and link nutrient availability with insulin's ability to regulate metabolism. Chronic nutrient excess can activate cell stress responses and impair signal transduction via transcriptional and posttranscriptional mechanisms. Three stress responses (oxidative stress, inflammation, and ER stress) and three predominant pathways for nutrient sensing will be discussed (mammalian target of rapamycin [mTOR], AMP kinase, and hexosamine biosynthetic pathway).

Cell Stress Responses and Insulin Action

Oxidative Stress. In addition to mediating β cell dysfunction in T2D[126] (see Mechanisms Causing Pancreatic β Cell Dysfunction section), oxidative stress represents an important factor in the development of insulin resistance.[249] The mediator of oxidative stress is an increase in the formation of ROS within cells due to increased generation and/or compromised antioxidant systems. The predominant portion of ROS is produced in cells during mitochondrial respiration and is augmented by nutrient excess, physical inactivity, and hyperglycemia. High levels of ROS activate NF-κB and upregulate expression of proinflammatory genes such as TNFα, IL-6, and C-reactive protein.[250] Oxidative stress activates serine kinases that impair insulin signal transduction by phosphorylating insulin receptor and IRS molecules, and also promotes proteosomal degradation of IRS-1.

Inflammation. Low-grade systemic inflammation in patients with insulin resistance is marked by elevated levels of circulating proinflammatory cytokines and other biomarkers of inflammation. Various factors including elevated saturated FAs, oxidative stress, ER stress, and cytokines can all initiate proinflammatory responses by activating the JNK/AP-1 and IKKb/NF-κB signaling pathways.[197] JNK-1 and IKKb serine kinases activate the transcription factors AP-1 (C-Jun/Fos) and NF-κB, which then activate serine kinases that phosphorylate the insulin receptor and IRS proteins[180-183] and increase transcription of *SOCS* genes.

Endoplasmic Reticulum Stress. ER stress can be an adaptive cellular response to an accumulation of unfolded or misfolded proteins in its lumen, but can also be triggered by other factors, including chronic nutrient excess, high FA concentrations, and oxidative stress.[251] ER stress activates the JNK pathway, with subsequent serine phosphorylation of IRS-1 and the development of insulin resistance. The mechanism for JNK activation involves transmembrane ER proteins with luminal domains responsive to ER stress signals and cytoplasmic effector domains that influence intracellular pathways. One of these stress transducers is inositol-requiring enzyme 1, a serine-threonine kinase that links ER stress to JNK activation. Obesity and extended periods of high-fat feeding can induce the ER stress response in adipose tissue and liver, as indicated by ER stress markers such as PERK, eIF2α phosphorylation, and JNK activation.

Nutrient Sensors and Nutrient Stress

The Mammalian Target of Rapamycin Signaling Pathway. mTOR is a central signal integrator that receives signals arising from growth factors, nutrients, and cellular energy metabolism, and then activates pathways that control cell growth as well as insulin action.[252] The C-terminal region of mTOR is a serine-threonine protein kinase and exists in two functionally distinct complexes dubbed mTOR complex 1 (mTORC1) and mTORC2. mTORC1 can impair insulin signaling via its ability to activate S6K, which then can phosphorylate serine residues of IRS-1, resulting in desensitization of the PI-3 kinase/Akt pathway.[249,250] High levels of glucose and amino acids, especially leucine, lead to an increase in mTOR activity, while depletion of these metabolic substrates or mitochondrial (e.g., oligomycin) and glycolytic (e.g., 2-deoxyglucose) inhibitors attenuate mTOR activity. Thus, activation of S6K by mTOR serves as a mechanism by which nutrient excess can desensitize insulin action through serine phosphorylation of IRS-1.[255] Rapamycin in a complex with its intracellular receptor FKBP12 can bind and inhibit in mTOR, which results in enhanced insulin-mediated glucose uptake by inhibiting phosphorylation of S6K. The proximal signaling component immediately responsible for inhibiting the mTOR pathway is the tuberous sclerosis complex (TSC). TSC can integrate information from the insulin signaling cascade and the AMPK signaling pathway. These multiple steps involved in nutrient sensing by the TSC and mTOR represent attractive targets for treating insulin resistance in metabolic diseases caused by nutrient excess, such as T2D and obesity.

AMPK Signaling Pathway. The AMPK signaling pathway is activated by elevation of the AMP/ATP ratio. The pathway is switched on by nutrient deprivation leading to accumulation of AMP, or perturbations that interfere with ATP synthesis, such as glucose deprivation, hypoxia, metabolic inhibitors (2-deoxyglucose), and disruption of oxidative phosphorylation. Once activated, it switches on fuel metabolism pathways that generate ATP, while inhibiting anabolic processes that consume ATP.[256] To conserve ATP, AMPK activation suppresses protein synthesis by inhibiting the mTOR pathway and by inhibiting translational elongation of proteins via activation of elongation factor-2 kinase. To generate ATP, AMPK activates processes such as glucose uptake, glycolysis, FA oxidation, and mitochondrial biogenesis, while switching off anabolic processes that consume ATP, including FA and cholesterol biosynthesis, gluconeogenesis, and glycogenesis. To enhance the capacity for ATP generation, AMPK appears to promote mitochondrial biogenesis through concerted action with another metabolic sensor, the NAD^+-dependent deacetylase SIRT1, and the PGC-1α transcriptional co-factor. In skeletal muscle, AMPK stimulates lipid oxidation, increases insulin sensitivity, and mediates contraction-induced increments in glucose uptake during exercise.[225,257] In liver, AMPK activation suppresses gluconeogenic enzyme expression and leads to decreased blood glucose values, while FA oxidation is stimulated, and cholesterol and triglyceride biosynthesis are blocked.[259] These biological actions are the result of phosphorylation and inactivation of both acetyl-CoA carboxylase (ACC) and 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase) by AMPK. The phosphorylation of ACC is central to AMPK's ability to promote FA oxidation, as this reduces its activity and results in decreased production of malonyl-CoA, an inhibitor of carnitine palmitoyltransferase I (CPT1). The reduction in malonyl-CoA relieves the inhibition of CPT1 and promotes CPT1-mediated transport of FAs into the mitochondria for oxidation. At the transcriptional level, AMPK blocks lipogenesis by limiting endogenous LXR ligand production, which inhibits LXR-dependent transcription of SREBP-1.[259]

AICAR has been used experimentally to study the effects of AMPK activation on metabolism and insulin signaling pathways.[260] AICAR is an adenosine analog that mimics AMP and activates AMPK signaling through direct binding. AICAR has been shown to stimulate glucose uptake, increase adiponectin secretion, and inhibit production of cytokines such as TNFα and IL-6. AMPK is activated by the insulin-sensitizing hormone adiponectin, and two classes of antidiabetic drugs, biguanides and thiazolidinediones, both act in part by activation of AMPK.

Branched-Chain Amino Acids. Metabolomic studies have shown that serine and glycine levels are positively correlated with hyperinsulinemic clamp measures of insulin sensitivity, while branched-chain amino acids (BCAAs), aromatic amino acids, and C3 and C5 acylcarnitines are associated with insulin resistance.[261,262] Elevated BCAA levels confer a higher risk of developing T2D, underscoring the association with insulin resistance.[263] Increased BCAA concentrations could result from dietary sources or due to greater protein breakdown and release into the circulation as a result of insulin's decreased ability to inhibit proteolysis. The mechanistic link with insulin resistance has not been fully elucidated but may involve activation of the mTOR complex and S6 kinase pathway.[264]

The Gut Microbiome. Metagenomic studies have revealed that approximately 90% of the bacterial species present in the gut of adults belongs to two phyla, *Bacteroides* (gram-negative) and *Firmicutes* (gram-positive). In patients with obesity and metabolic diseases, investigators have reported alterations in gut microbiota that include a reduction in bacterial diversity together with a relative enrichment in *Firmicutes* and reduction in *Bacteroides* species. The question as to whether changes in these bacterial populations can influence health has been an active area of research, and several mechanisms have been proposed.[265,266] First, the bacteria secrete metabolites that can act locally in the gut to influence immune cells in the intestinal wall or bind gut epithelial receptors. An example of the latter would include bacterial-derived medium-chain FAs that can activate G protein–coupled receptors (GRP40 and GRP120) on the surface of enteroendocrine L-cells in the intestine and influence secretion of the incretin GLP-1. The second possibility is these bacterial products increase intestinal permeability, allowing metabolites to be absorbed into the bloodstream, where they can alter metabolism in distal tissues. Bacterial products that can reach the systemic circulation include bacterial lipopolysaccharides, bacterial DNA, short-chain FAs, and bile acid derivatives, which potentially can influence insulin sensitivity, inflammation, and energy metabolism.[267]

Most of this work has been performed in rodents, in particular gnotobiotic and germ-free mice receiving fecal transplants from mice with metabolic disease. For example, recolonization of the intestine of germ-free mice with gut microbiota obtained from obese mice induces obesity in the host.[268] Far fewer data are available in humans, and the majority of studies are descriptive, assessing differences in microbiota in cases versus controls or changes over time following an intervention. One randomized study conducted fecal transplants from lean or obese donors to obese human hosts showed that transplants from lean donors were associated with improved insulin sensitivity without any effect on body weight over a 6-week period.[269] However, within 1 year, the original microbiota had reestablished itself, and insulin resistance returned to pretransplant levels. Of note, the metabolic responses to different diets have been shown to vary between individuals depending on the composition of their gut microbiota. For example, a randomized clinical study reported that participants with high Prevotella-to-Bacteriodes (P/B) ratio were more susceptible to lose body fat on diets high in fiber than subjects with a low P/B ratio.[270]

The impact of the gut microbiota on T2D has not been elucidated, and the therapeutic potential of manipulating gut bacterial populations will require additional study in humans. Bacterial populations are altered in the short term by diet and other factors, and this plasticity makes it more difficulty to ascribe a specific bacterial mix as a causal factor in chronic disease. Fecal transplantation is being used for treatment of some human diseases (*Clostridium difficile* colitis, inflammatory bowel disease), so there is hope that eventually gut microbiota can be targeted for therapeutic benefits in metabolic diseases.

The Hexosamine Biosynthetic Pathway, TRIB3, and Glucose-Induced Insulin Resistance

Hyperglycemia-induced insulin resistance in T2D is due to the ability of glucose to directly impair the activity of the glucose transport system in target cells. Marshall and others have demonstrated that one important mechanism underlying glucose-induced insulin resistance involves glucose metabolism through a minor intracellular pathway, the hexosamine biosynthetic pathway.[271] Glutamine:fructose-6-phosphate amidotransferase is the rate-limiting enzyme of this pathway that leads to the production of glucosamine-6-phosphate. Glucosamine-6-phosphate is then metabolized to UDP-*N*-acetylglucosamine (UDP-GlcNAc), which serves as the substrate for attachment of *O*-GlcNAc to proteins catalyzed by O-linked N-acetylglucosamine transferase. Thus, increased glucose flux through the hexosamine biosynthetic pathway leads to increased posttranslational O-linked glycosylation of cytoplasmic and nuclear proteins,[272] which can modulate their enzyme activities, proteasomal degradation, and binding with DNA and other proteins. UDP-GlcNAc levels can fluctuate with the availability of nonesterified FAs, uridine, glutamine, as well as glucose, and in this way the hexosamine biosynthetic pathway acts as a nutrient sensor and regulator.

Increased glucose flux through the hexosamine biosynthetic pathway results in reduced rates of basal and insulin-stimulated glucose transport in adipocytes and muscle cells.[212,213,273] However, the mechanisms by which products of the hexosamine biosynthetic pathway induce insulin resistance have not been fully elucidated. Increased metabolic flux through this pathway results in increased O-glycosylation of the insulin receptor, IRS1, Akt/PKB, GLUT4, GSK3β, and the transcription factor Sp1, and this can impair the functional capabilities of these proteins or accelerate their proteasomal degradation. For example, increased O-glycosylation of the insulin receptor is accompanied by a decrease in insulin-stimulated receptor autophosphorylation.[272] O-glycosylation can impair phosphorylation and DNA binding of C/EBPβ and decrease NK-κB binding, with IκB freeing NF-κB for relocation to the nucleus and induction of proinflammatory gene transcription. However, a key mechanism for glucose-induced insulin resistance appears to involve Tribbles homolog 3 (TRIB3). TRIB3 is a glucose-induced gene that encodes a pseudokinase capable of binding and blocking protein tyrosine phosphorylation sites. Specifically, TRIB3 can block activation of AKT by occupying its phosphorylation site and inhibit insulin signal transduction.[274] TRIB3 is significantly upregulated in skeletal muscle of T2D patients and hyperglycemic rodent models in a manner that is correlated with circulating glucose.[274] The ability of glucose to induce *TRIB3* gene expression is dependent on its metabolism via the hexosamine biosynthetic pathway and O-glycosylation of Sp1. Moreover, glucose-induced insulin resistance in diabetic streptozotocin-treated mice was found to be dependent upon upregulation of TRIB3 in experiments involving muscle-specific *TRIB3* deletion.[213]

KEY POINTS

- Insulin action defects in type 2 diabetes are induced by cell stress responses that are integral to the insulin resistant state, as well as nutrient stress responses involving nutrient sensor pathways. Cell stress is generated an excess of glucose and lipids inducing oxidative stress, inflammation, and endoplasmic reticulum stress. These responses are driven by nutrient sensor pathways involving mammalian target of rapamycin signaling, adenosine monophosphate kinase, and the hexosamine biosynthetic pathway.

SYNDROMES OF SEVERE INSULIN RESISTANCE

Heterogeneity of Type 2 Diabetes

T2D is a heterogeneous disease across spectrums of age, BMI, and the degree to which defects in insulin secretion versus insulin action contribute to hyperglycemia. In addition, there are distinctive phenotypes that occur relatively uncommonly associated with identifiable genetic defects, unique syndromic features, or other disease processes. These are listed in Table 33.1. Many of these are described in other chapters or referenced publications as indicated in the table, and severe insulin resistance types A and B will be addressed in this chapter. In aggregate, metabolic and phenotypic heterogeneity in T2D is indicative of disease complexity involving varying molecular defects, genetic abnormalities, behaviors, and gene–environment interactions.

Severe Insulin Resistance

Severe Insulin Resistance Type A. These patients carry different mutations in the insulin receptor gene.[275,276] The syndrome may be characterized by autosomal dominant or recessive inheritance and involve mutations that can be heterozygous or homozygous when mutations cause relatively mild defects in receptor function. These mutations impair either insulin binding or signal transduction through receptor tyrosine kinase activity.[277,278] Other patients who conform with this clinical phenotype do not have mutations in the insulin receptor gene and may have as yet unidentified mutations in genes encoding downstream signaling molecules (e.g., *AKT2*, *AS160*). Clinically, these patients have extreme hyperinsulinemia and most often present as adolescent or young females with acanthosis nigricans, abnormal glucose tolerance, hyperandrogenism, and oligomenorrhea. In males, the disease has less distinctive features, and patients may only present with acanthosis nigricans. In some patients, prominent musculature has been described that may result from hyperandrogenism in females or crossover stimulation of the insulin-like growth factor-1 receptor as a consequence of elevated serum insulin concentrations.

Severe Insulin Resistance Type B. Type B severe insulin resistance is the result of circulating antibodies directed against the insulin receptor that block insulin binding and receptor activation.[279] The clinical presentation features hyperinsulinemia, hyperglycemia, acanthosis nigricans, and evidence of autoimmunity, if not the presence of overt autoimmune disorders such as rheumatoid arthritis, Raynaud's phenomenon, or systemic lupus erythematosus. Consistent with other autoimmune disorders, the disease most often occurs in middle-aged females, but can also occur in adolescents and the elderly, and has a higher incidence in Blacks. Laboratory findings include serum insulin receptor blocking antibodies, proteinuria, and elevated antinuclear antibody, erythrocyte sedimentation rate, and serum immunoglobulin G levels. On examination, patients can be found to have arthritis, vitiligo, alopecia, and enlarged salivary glands.

Patients can require high doses of insulin to control hyperglycemia, sometimes greater than 1,000 units/day, when use of concentrated insulin preparations (e.g., U500) is warranted. Other patients may have more mild degrees of abnormal glucose tolerance but still display extreme endogenous hyperinsulinemia. The clinical course can vary with changing titers of antireceptor antibodies or even undergo spontaneous remission for periods of months to years. Treatment involves therapy for the underlying autoimmune process, glucocorticoids, and/or plasmapheresis in severe cases.

Paradoxically, some patients may present with or develop hypoglycemia.[280] While directed against the insulin-binding α subunit, the receptor antibodies can have assorted biological effects. In addition to blocking insulin binding, some antibodies can have agonist properties and stimulate glucose transport and metabolism, which can explain hypoglycemia in some patients over the course of their disease.[281] Antibodies isolated from hyperglycemic or hypoglycemic patients may acutely stimulate insulin receptor activity *in vitro*; however, cells likely become refractory to these effects, which can explain the changing clinical course and heterogeneity of clinical presentations. In hypoglycemic patients, elevated insulin levels may lead to a false diagnosis of insulinoma, and it is important to test for the presence of insulin receptor antibodies in these patients.[282]

NATURAL HISTORY OF TYPE 2 DIABETES AS AN END-STAGE MANIFESTATION OF CARDIOMETABOLIC DISEASE

The Spectrum of Cardiometabolic Disease

In addition to examining the evolution of T2D as a disease of glucoregulation, it is appropriate to include its pathophysiology in the larger context of CMD. In this light, T2D is an end-stage manifestation of a single disease process affecting both metabolism and the vasculature, which explains why T2D, CVD, NAFLD, and obesity are epidemiologically linked and often present as comorbidities in single patients. The spectrum of CMD is shown in Fig. 33.17. The core abnormality in the progression of CMD is insulin resistance.[165] The classic cardinal manifestations of insulin resistance include normoglycemia or hyperglycemia in the face of hyperinsulinemia due to defects in insulin-stimulated glucose uptake in muscle and suppression of hepatic glucose production. However, the glucoregulatory defect is associated with additional pathophysiological processes that adversely impact both metabolism and the vasculature, including: (i) impaired mitochondrial function and substrate oxidation; (ii) increased inflammation and oxidative stress; (iii) dysfunctional and inflamed adipose tissue with altered secretion of adipocytokines; (iv) impaired lipid storage in adipocytes associated with accumulation of "ectopic" lipid within muscle and liver cells; (v) alterations in lipids and lipoproteins contributing to CVD risk, including increased large triglyceride-rich VLDL, lower HDL, and increased concentrations of small dense low-density lipoprotein molecules; (vi) abnormalities in vasoregulation due to a reduction in endothelial nitric oxide synthase activity and nitric oxide production; (vii) a clotting diathesis featuring elevated PAI-1 and fibrinogen levels; and (viii) functional and structural abnormalities affecting the myocardium predisposing to heart failure with preserved ejection fraction, congestive heart failure, and atrial fibrillation. It is not clear whether the glucoregulatory defect precedes the development of these latter processes or *vice versa*. For this reason, in the context of what initiates and promulgates CMD, the combination of these multiple pathophysiological processes, together with defects in glucoregulation, are referred to as the insulin-resistant state.

The insulin-resistant state can remain subclinical for decades; however, the progression of CMD gives rise to clinically identifiable states of high risk in the form of metabolic syndrome, prediabetes, prehypertension, dyslipidemia, ventral adiposity, and hepatic steatosis.[165] While there is no practical method of directly measuring insulin resistance in the clinic, its presence can be denoted in patients by the presence of these traits.[5,283] These clinical manifestations indicate that the pathophysiological processes integral to the insulin-resistant state are operational, placing patients at risk of end-stage metabolic and vascular disease outcomes, including T2D, nonalcoholic steatohepatitis (NASH), hypertension, myocardial infarction, stroke, myocardial dysfunction, and chronic kidney disease.

Role of Obesity

Obesity is a chronic disease of energy balance driven by dysregulated interactions involving satiety factors and the CNS resulting in increased caloric intake and an excess in adipose tissue mass (see Obesity and Nutrition). The increase in adiposity causes chronic complications that confer increased morbidity and mortality.[136] The relationship between body weight gain and CMD is complex. The correlation between generalized adiposity and insulin sensitivity indicates that only approximately 11% of individual variability in insulin sensitivity can be explained by BMI.[25] Many individuals with CMD who progress to T2D and/or CVD are lean. In addition, a certain proportion of patients with BMIs in the overweight or obesity range are insulin-sensitive, do not develop metabolic syndrome traits, and are not at increased risk of T2D and CVD.[5,284] These patients are referred to as the metabolically healthy obese, and represented approximately 15% to 20% of US adults with overweight or obesity in the National Health and Nutrition Examination Survey cohort.[285] Thus, obesity is not necessary nor sufficient as a cause of CMD and T2D.

However, if weight gain occurs on an insulin-resistant background, there is preferential accumulation of fat in the intraabdominal

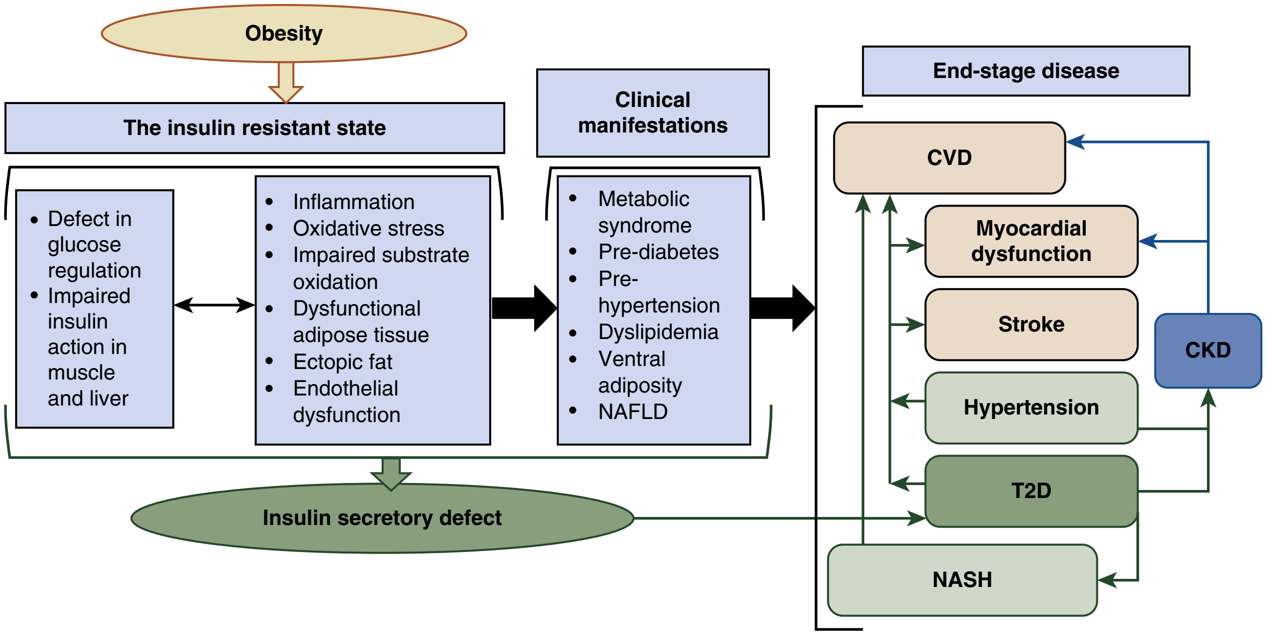


Fig. 33.17 **The spectrum of cardiometabolic disease.** Pathophysiological properties of the insulin-resistant state *(gray)* produce clinical manifestations that mark patients as insulin-resistant and indicate the presence of cardiometabolic disease. End-stage manifestations of these disease process include the vascular diseases in red (cardiovascular disease [CVD], myocardial dysfunction, stroke) and metabolic diseases in green (type 2 diabetes [T2D], nonalcoholic steatohepatitis [NASH], hypertension). The presence of the metabolic diseases further exacerbates vascular pathogenesis and contributes to chronic kidney disease (CKD). Obesity can exacerbate insulin resistance and drive the progression of cardiometabolic disease.

compartment, which contains an increased number of resident macrophages[92,286] predominantly involving classically activated M1 and a lesser proportion of alternatively activated M2 macrophages. These resident macrophages secrete cytokines and other factors that crosstalk with adipocytes and alter the systemic release of adipocytokines from adipose tissue.[237] Altered release of adipose tissue factors such as adiponectin, resistin, free FAs, leptin, angiotensinogen, and PAI-1, and increased production of inflammatory cytokines including IL-6, IL-1, and TNFα alter metabolism in multiple organs. Adipocytes residing in the inflamed adipose tissue environment are dysfunctional, exhibiting insulin resistance with respect to simulation of glucose transport and antilipogenesis, and reach a limit where they are unable to store further calories as fat. Serum levels of FFAs become elevated, and ectopic lipid accumulates within liver and muscle cells, leading to NAFLD and worsening insulin resistance. With the development of obesity, there is a greater need to dispose of increased calories, which further augments lipid deposited within muscle and liver cells and intensifies processes such as inflammation, oxidative stress, and glucose intolerance. The systemic inflammation and oxidative stress activate NF-κB, further propagating inflammatory pathways and production of cytokines. Intracellular levels of DAGs and ceramides also increase and activate both serine/threonine kinases, which desensitize insulin action via phosphorylation of signaling molecules (i.e., insulin receptor substrate-1 and insulin receptor).[283] Adiponectin is an insulin-sensitizing hormone produced only by adipocytes, and reduced plasma levels are associated with insulin resistance, obesity, and CVD.[288,289] Diminished production of adiponectin is a mechanism that can explain metabolic disease both by adding to defects in adipocyte function as an autocrine/paracrine factor,[95,234] and in the vascular wall by promoting foam cell formation and fatty streak development via reduced inhibition of cholesterol accumulation in macrophages.[95,290] These molecular events exacerbate insulin action defects in muscle and liver, worsen systemic inflammation and metabolism, and impel progression towards end-stage developments in CMD.

Thus, there is an interaction between excess adiposity and insulin resistance to worsen pathophysiological processes associated with the insulin-resistant state, and this explains why obesity is a major risk factor for the development of metabolic syndrome and prediabetes, as well as for progression to overt T2D, NASH, and CVD. Clinically, the critical point is that, in those patients with elevated BMI and insulin resistance, weight loss becomes a highly effective therapeutic modality to prevent and treat CMD.[165]

Implications for Type 2 Diabetes Prevention and Treatment

Within the context of CMD, T2D can be seen as an end-stage development of a larger chronic disease process. Patients who remain normoglycemic continue to exist in a state of accelerated atherosclerosis and are at increased risk of CVD events. Patients who are predisposed to pancreatic β cell exhaustion develop prediabetes, followed by progression to overt T2D. The perspective of T2D as an end-stage development of a chronic progressive disease has two salient clinical implications. First, this

KEY POINTS

- In most patients, type 2 diabetes (T2D) is an end-stage manifestation of cardiometabolic disease (CMD), with insulin resistance as the core mechanism responsible for disease progression to prediabetes and the metabolic syndrome to overt T2M. Other end-stage manifestations include cardiovascular disease, functional and structural myocardial disease, stroke, hypertension, nonalcoholic fatty liver disease and steatohepatitis, and chronic kidney disease. Obesity can exacerbate insulin resistance and impel CMD progression. This perspective emphasizes opportunities for primary, secondary, and tertiary disease prevention, with weight loss being an effective intervention, and points to the need for comprehensive vascular and metabolic care in patients with T2D.

assimilates the framework for designing rational strategies for primary, secondary, and tertiary prevention. It is no longer sufficient to wait until diabetes becomes overt to begin treatment. Given the increasing burden of patient suffering and social costs exacted by T2D, it becomes imperative to intervene early to halt CMD progression. Primary prevention is appropriate for individuals at risk of the insulin-resistant state. While the insulin-resistant state is largely subclinical, individuals at risk can be identified based on risk stratification paradigms,[5,15], family history, *in utero* exposure to gestational diabetes, genetic factors, ethnicity, behavioral factors, and social and environmental determinants of health. Primary preventive strategies are warranted, such as public health messaging, the built environment, access to unprocessed foods, healthy gestation, economic viability, stress reduction, weight control, and access to preventive care. Once patients develop the clinical manifestations of the insulin resistant state (e.g., metabolic syndrome, prediabetes, nonalcoholic fatty liver, prehypertension), more aggressive and targeted secondary preventive interventions are needed, such as 7% to 10% weight loss,[136,291] physical activity and healthy meal plan prescriptions, and control of dyslipidemia, abnormal glucose tolerance, and elevated blood pressure.

The second implication applies to patients who have progressed to overt T2D. In addition to T2D, other end-stage manifestations of CMD include coronary artery disease, stroke, functional myocardial impairment, hypertension, and NASH. Once CMD progresses to end-stage disease outcomes, aggressive tertiary management and comprehensive care is required, because these diseases adversely impact each other. Development of T2D worsens NASH and further accelerates atherogenesis and myocardial dysfunction, predisposing to myocardial infarction and chronic heart failure. T2D and hypertension lead to a decline in renal function, which exacerbates both CVD and metabolic pathophysiology. From this perspective, treatment of diabetes encompasses much more than glycemic control and must concomitantly encompass the prevention and treatment of all end-stage manifestations of CMD.

The neglect for early intervention has failed our patients and our societies, given the rising rates of diabetes and the suffering and social costs this entails internationally. In particular, because the clear majority of patients with CMD have overweight and obesity, weight loss therapy using the evidence-based tools of obesity medicine are too often neglected, despite documentation that they are highly effective in preventing and treating T2D.[136]

Disclosures

W.T.G. has served on *ad hoc* advisory boards for Novo Nordisk, Boehringer-Ingelheim, Jazz Pharmaceuticals, Pfizer, BOYDSense, the Milken Foundation, and the American Medical Group Association, and he has conducted research sponsored by the University of Alabama at Birmingham funded by Lexicon, Pfizer, and Novo Nordisk.

REFERENCES

1. Saeedi P, Petersohn I, Salpea P, the IDF Diabetes Atlas Committee, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes Federation diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019;157:107843.
2. Khan MAB, Hashim MJ, King JK, et al. Epidemiology of type 2 diabetes–global burden of disease and forecasted trends. *J Epidemiol Glob Health*. 2020;10:107–111.
3. Centers for disease control and prevention. *National Diabetes Statistics Report, 2020*. Atlanta: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2020.
4. Tabak AG, Herder C, Rathmann W, et al. Pre-diabetes: a high-risk state for developing diabetes. *Lancet*. 2012;379:2279–2290.
5. Guo F, Moellering DR, Garvey WT. The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging system applicable to obesity. *Obesity*. 2014;22:110–118.
6. Barker LE, Kirtland KA, Gregg EW, et al. Geographic distribution of diagnosed diabetes in the US: a diabetes belt. *Am J Prev Med*. 2011;40:434–439.
7. Weijnen CF, Rich SS, Meigs JB, et al. Risk of diabetes in siblings of index cases with Type 2 diabetes: implications for genetic studies. *Diabet Med*. 2002;19:41–50.
8. Barnett A, Eff C, Leslie R, et al. Diabetes in identical twins: a study of 200 pairs. *Diabetologia*. 1981;20:87–93.
9. Willemsen G, Ward KJ, Bell CG, et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the Discordant Twin (DISCOTWIN) Consortium. *Twin Res Hum Genet*. 2015;18:762–771.
10. Poulsen P, Kyvik KO, Vaag A, et al. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—a population-based twin study. *Diabetologia*. 1999;42:139–145.
11. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet*. 2007;8:657–662.
12. Scott RA, Scott LJ, Mägi R, et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes*. 2017;66:2888–2902.
13. Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun*. 2018;9:2941.
14. Mahajan A, Taliun D, Thurner M, et al. Fine mapping of an expanded set of type 2 diabetes loci to single variant resolution using high density imputation and islet specific epigenome maps. *Nat Genet*. 2018;50:1505–1513.
15. Fuchsberger C, Flannick J, Teslovich TM. The genetic architecture of type 2 diabetes. *Nature*. 2016;536:41–47.
16. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2diabetes. *Nat Genet*. 2006;38:320–323.
17. Munoz J, Lok KH, Gower BA, et al. A polymorphism in the transcription factor 7-like 2 (TCF7L2) gene is associated with reduced insulin secretion in non-diabetic women. *Diabetes*. 2006;55:3630–3634.
18. Almgren P, Lehtovirta M, Isomaa B, et al. Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia*. 2011;54:2811–2819.
19. Lehtovirta M, Kaprio J, Forsblom C, et al. Insulin sensitivity and insulin secretion in monozygotic and dizygotic twins. *Diabetologia*. 2000;4:285–293.
20. Elbein SC, Hasstedt SJ, Wegner K, et al. Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. *J Clin Endocrinol Metab*. 1999;84:1398–1403.
21. Walford GA, Gustafsson S, Rybin D, et al. Genome-wide association study of the modified stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes*. 2016;65:3200–3211.
22. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105–116.
23. Ingelsson E, Langenberg C, Hivert MF, et al. Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. *Diabetes*. 2010;59:1266–1275.
24. Dimas AS, Lagou V, Barker A, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. 2014;63:2158–2171.
25. Garvey WT, Lara-Castro C. Diet, insulin resistance, and obesity: zoning in on data for Atkins dieters living in South Beach. *J Clin Endocrinol Metab*. 2004;89:4197–4205.
26. Magnusson PK, Rasmussen F. Familial resemblance of body mass index and familial risk of high and low body mass index. A study of young men in Sweden. *Int J Obes Relat Metab Disord*. 2002;26:1225–1231.

27. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997;27:325–351.
28. Haqq S, Ganna A, van der Laan SW, et al. Gene-based meta-analysis of genome wide association studies implicates new loci involved in obesity. *Hum Mol Genet*. 2015;24:6849–6860.
29. Locke AE, Kahali B, Berndt SI. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206.
30. Karaderi T, Drong AW, Lindgren CM. Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of obesity-related traits. *Curr Diab Rep*. 2015;15:83.
31. Grarup N, Sandholt CH, Hansen T, et al. Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia*. 2014;57:1528–1541.
32. Flannick J, Mercader JM, Fuchsberger C, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019;570:71–76.
33. Albrechtsen A, Grarup N, Li Y, et al. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia*. 2013;56:298–310.
34. Tanaka D, Nagashima K, Sasaki M, et al. Exome sequencing identifies a new candidate mutation for susceptibility to diabetes in a family with highly aggregated type 2 diabetes. *Mol Gen Metab*. 2013;109:112–117.
35. Rosen ED, Kaestner KH, Natarajan R, et al. Epigenetics and epigenomics: implications for diabetes and obesity. *Diabetes*. 2018;67:1923–1931.
36. Kaelin WG, McKnight SL. Influence of metabolism on epigenetics and disease. *Cell*. 2013;153:56–69.
37. Ling C, Rönn T. Epigenetics in human obesity and type 2 diabetes. *Cell Metab*. 2019;29:1028–1044.
38. Chen Z, Miao F, Paterson AD, et al. Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc Natl Acad Sci USA*. 2016;113:E3002–E3011.
39. Bacos K, Gillberg L, Volkov P, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun*. 2016;7:11089.
40. Ouni M, Saussenthaler S, Eichelmann F, et al. Epigenetic changes in islets of Langerhans preceding the onset of diabetes. *Diabetes*. 2020;69:2503–2517.
41. Yang IV, Zhang W, Davidson EJ, et al. Epigenetic marks of *in utero* exposure to gestational diabetes and childhood adiposity outcomes: the EPOCH study. *Diabet Med*. 2018;35:612–620.
42. Ruchat SM, Houde AA, Voisin G, et al. Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. *Epigenetics*. 2013;8:935–943.
43. Kang J, Lee C-N, Li H-Y, et al. Genome-wide DNA methylation variation in maternal and cord blood of gestational diabetes population. *Diabetes Res Clin Pract*. 2017;132:127–136.
44. Sobngwi E, Boudou P, Mauvais-Jarvis F, et al. Effect of a diabetic environment *in utero* on predisposition to type 2 diabetes. *Lancet*. 2003;361:1861–1865.
45. Nicholas LM, Rattanatray L, MacLaughlin SM, et al. Differential effects of maternal obesity and weight loss in the periconceptional period on the epigenetic regulation of hepatic insulin-signaling pathways in the offspring. *FASEB J*. 2013;27:3786–3796.
46. Liu X, Chen Q, Tsai HJ, et al. Maternal preconception body mass index and offspring cord blood DNA methylation: exploration of early life origins of disease. *Environ Mol Mutagen*. 2014;55:223–230.
47. Sharp GC, Lawlor DA, Richmond RC, et al. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2015;44:1288–1304.
48. Thakali KM, Saben J, Faske JB, et al. Maternal pregravid obesity changes gene expression profiles toward greater inflammation and reduced insulin sensitivity in umbilical cord. *Pediatr Res*. 2014;76:202–210.
49. Lesseur C, Armstrong DA, Paquette AG, et al. Maternal obesity and gestational diabetes are associated with placental leptin DNA methylation. *Am J Obstet Gynecol*. 2014;211:654. e1–e9.
50. Bush NC, Chandler-Laney PC, Rouse DJ, et al. Higher maternal gestational glucose concentration is associated with lower offspring insulin sensitivity and altered {beta}-cell function. *J Clin Endocrinol Metab*. 2011;96:E803–E809.
51. Barrón-Cabrera E, Ramos-Lopez O, González-Becerra K, et al. Epigenetic modifications as outcomes of exercise interventions related to specific metabolic alterations: a systematic review. *Lifestyle Genomics*. 2019;12:25–44.
52. Berglind D, Muller P, Willmer M, et al. Differential methylation in inflammation and type 2 diabetes genes in siblings born before and after maternal bariatric surgery. *Obesity*. 2016;24:250–261.
53. Guenard F, Deshaies Y, Cianflone K, et al. Differential methylation in glucoregulatory genes of offspring born before vs after maternal gastrointestinal bypass surgery. *Proc Natl Acad Sci USA*. 2013;110:11439–11444.
54. Thomou T, Mori MA, Dreyfuss JM, et al. Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature*. 2017;542:450–455.
55. Dumortier O, Hinault C, Van Obberghen E. MicroRNAs and metabolism crosstalk in energy homeostasis. *Cell Metab*. 2013;18:312–324.
56. Trajkovski M, Hausser J, Soutschek J, et al. MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature*. 2011;474(7353):649–653.
57. Ma E, Fu Y, Garvey WT. Relationship of circulating miRNAs with insulin sensitivity and associated metabolic risk factors in humans. *Metab Syndr Relat Disord*. 2018;16:82–89.
58. Shah R, Murthy V, Pacold M, et al. Extracellular RNAs are associated with insulin resistance and metabolic phenotypes. *Diabetes Care*. 2017;40:546–553.
59. Wang R, Hong J, Cao Y, et al. Elevated circulating microRNA-122 is associated with obesity and insulin resistance in young adults. *Eur J Endocrinol*. 2015;172:291–300.
60. Willeit P, Skroblin P, Moschen AR, et al. Circulating microRNA-122 is associated with the risk of new-onset metabolic syndrome and type 2 diabetes. *Diabetes*. 2017;66:347–357.
61. Luo N, Garvey WT, Wang D-Z, et al. MicroRNA-150 regulates lipid metabolism and inflammatory response. *J Metab Syndr*. 2013;2:131.
62. Chou CF, Lin YY, Wang HK, et al. KSRP ablation enhances brown fat gene program in white adipose tissue through reduced miR-150 expression. *Diabetes*. 2014;63:2949–2961.
63. Mazzucotelli A, Viguerie N, Tiraby C, et al. The transcriptional coactivator peroxisome proliferator activated receptor (PPAR)gamma coactivator-1 alpha and the nuclear receptor PPAR alpha control the expression of glycerol kinase and metabolism genes independently of PPAR gamma activation in human white adipocytes. *Diabetes*. 2007;56:2467–2475.
64. Kang M, Liu X, Fu Y, Garvey WT. Improved systemic metabolism and adipocyte biology in miR-150 knockout mice. *Metabolism*. 2018;83:139–148.
65. Desgagne V, Bouchard L, Guerin R. microRNAs in lipoprotein and lipid metabolism: from biological function to clinical application. *Clin Chem Lab Med*. 2017;55:667–686.
66. Brandao BB, Guerra BA, Mori MA. Shortcuts to a functional adipose tissue: the role of small non-coding RNAs. *Redox Biol*. 2017;12:82–102.
67. Li J, Zhang Y, Liu Y, et al. Microvesicle-mediated transfer of microRNA-150 from monocytes to endothelial cells promotes angiogenesis. *J Biol Chem*. 2013;288:23586–23596.
68. Haire-Joshu D, Hill-Briggs F. The next generation of diabetes translation: a path to health equity. *Annu Rev Public Health*. 2019;40:391–410.
69. Eckel RH, Jakicic JM, Ard JD, et al. AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J Amer Coll Cardiol*. 2014;63:2960–2984.
70. Estruch R, Ros E, Salas-Salvadó J, et al. Retraction and republication: primary prevention of cardiovascular disease with a Mediterranean diet, *N Engl J Med*. 2013;368:1279–1290. *N Engl J Med*. 2018;378:2441–2442.
71. de Lorgeril M, Salen P, Martin JL, et al. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. *Circulation*. 1999;99:779–785.

72. Salas-Salvadó J, Bulló M, Babio N, et al. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care*. 2011;34:14–19.
73. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, et al. 10-year follow- up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*. 2009;374:1677–1686.
74. Sale MM, Lu L, Spruill IJ, et al. Genome-wide linkage scan in Gullah-speaking African American families with type 2 diabetes: the Sea Islands Genetic African American Registry (Project SuGAR). *Diabetes*. 2009;58:260–267.
75. Hunt KJ, Kistner-Griffin E, Spruill I, et al. Cardiovascular risk in Gullah African Americans with high familial risk of type 2 diabetes mellitus: Project SuGAR. *South Med J*. 2014;107:607–614.
76. Garvey WT, McClean DC, Spruill I. The search for obesity genes in isolated populations: Gullah-speaking African Americans and the role of uncoupling protein 3 as a thrifty gene. In: Medeiros G, Halpern A, Bouchard C, eds. *Progress in Obesity Research*. Paris: John Libbey Eurotext; 2003:373–380.
77. Sundufu AJ, Bockarie CN, Jacobsen KH. The prevalence of type 2 diabetes in urban Bo, Sierra Leone, and in the 16 countries of the West Africa region. *Diab Metab Res Rev.* 2017;33. e2904.
78. Schulz LO, Bennett PH, Ravussin E, et al. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care*. 2006;29:1866–1871.
79. Esparza-Romero J, Valencia ME, Urquidez-Romero R, et al. Environmentally driven increases in type 2 diabetes and obesity in Pima Indians and Non-Pimas in Mexico over a 15-year period: the Maycoba Project. *Diabetes Care*. 2015;38:2075–2082.
80. Esparza-Romero J, Valencia ME, Martinez ME, et al. Differences in insulin resistance in Mexican and U.S. Pima Indians with normal glucose tolerance. *J Clin Endocrinol Metab*. 2010;95:E358–E362.
81. Chen M, Bergman R, Pacini G, et al. Pathogenesis of age-related glucose intolerance in man: insulin resistance and decreased β-cell function. *J Clin Endocrinol Metab*. 1985;60:13–20.
82. Gutierrez C, Jannausch M, Li X, et al. Changes in body composition in women over six years at midlife: ovarian and chronological aging. *J Clin Endocrinol Metab*. 2007;92:895–901.
83. Barker D. The developmental origins of insulin resistance. *Horm Res.* 2005;64:2–7.
84. Tobi EW, Slagboom PE, van Dongen J, et al. Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci within IGF2/H19. *PLoS One*. 2012;7:e37933.
85. American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetes—2021. *Diabetes Care*. 2021;44:S15–S33.
86. Guo F, Moellering DR, Garvey WT. Use of HbA1c for diagnoses of diabetes and prediabetes: comparison with diagnoses based on fasting and 2-hour glucose values and effects of gender, race, and age. *Metabol Syndr Relat Dis*. 2014;12:258–268.
87. Consequences of the new diagnostic criteria for diabetes in older men and women. DECODE study (diabetes Epidemiology: Collaborative analysis of Diagnostic criteria in Europe). *Diabetes Care*. 1999;22:1667–1671.
88. DeFronzo R. The triumvirate: β-cell, muscle, liver: a collusion responsible for NIDDM. *Diabetes*. 1988;37:667–687.
89. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*. 1988;37:1595–1607.
90. Porte D. Banting lecture 1990: β-cells in type II diabetes mellitus. *Diabetes*. 1991;40:166–190.
91. Czech MP. Mechanisms of insulin resistance related to white, beige, and brown adipocytes. *Mol Metab*. 2020;34:27–42.
92. Olefsky JM, Glass CK. Macrophages, inflammation, and insulin resistance. *Annu Rev Physiol*. 2010;72:219–246.
93. Xu H, Barnes GT, Yang QT, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest*. 2003;112:1821–1830.
94. Festa A, D'Agostino R, Mykkanen GHL, et al. Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS). *Circulation*. 2000;102:42–47.
95. Luo N, Chung BH, Wang X, et al. Enhanced adiponectin actions by overexpression of adiponectin receptor 1 in macrophages. *Atherosclerosis*. 2013;228:124–135.
96. Defronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes*. 2009;58:773–795.
97. Baron AD, Schaeffer L, Shragg P, et al. Role of hyperglucagonemia in increased rates of hepatic glucose output in type II diabetics. *Diabetes*. 1987;36:274–283.
98. Meier JJ, Nauck MA. Incretins and the development of type 2 diabetes. *Curr Diab Rep*. 2006;6:194–201.
99. Abdul-Ghani MA, DeFronzo RA. Inhibition of renal glucose reabsorption: a novel strategy for achieving glucose control in type 2 diabetes mellitus. *Endocr Pract*. 2008;14:782–790.
100. Pijl H, Ohashi S, Matsuda M, et al. Bromocriptine: a novel approach to the treatment of type 2 diabetes. *Diabetes Care*. 2000;23:1154–1161.
101. Sumithran P, Proietto J. The defense of body weight: a physiological basis for weight regain after weight loss. *Clin Sci (Lond)*. 2013;124:231–241.
102. Warram JH, Martin BC, Krolewski AS, et al. Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann Intern Med*. 1990;13:909–915.
103. Grodsky GM, Bolaffi JL. Desensitization of the insulin-secreting beta cell. *J Cell Biochem*. 1992;48:3–11.
104. Porte Jr D, Pupo AA. Insulin responses to glucose: evidence for a two pool system in man. *J Clin Invest*. 1969;48:2309–2319.
105. Brunzell JD, Robertson RP, Lerner RL, et al. Relationships between fasting plasma glucose levels and insulin secretion during intravenous glucose tolerance tests. *J Clin Endocrinol Metab*. 1976;42:222–229.
106. Robertson RP, Porte Jr D. The glucose receptor. A defective mechanism in diabetes mellitus distinct from the beta adrenergic receptor. *J Clin Invest*. 1973;52:870–876.
107. Lerner RL, Porte Jr D. Studies of secretin-stimulated insulin responses in man. *J Clin Invest*. 1972;51:2205–2210.
108. Crockford PM, Hazzard WR, Williams RH. Insulin response to glucagon. The opposing effects of diabetes and obesity. *Diabetes*. 1969;18:216–224.
109. Halter JB, Graf RJ, Porte Jr D. Potentiation of insulin secretory responses by plasma glucose levels in man: evidence that hyperglycemia in diabetes compensates for impaired glucose potentiation. *J Clin Endocrinol Metab*. 1979;48:946–954.
110. U.K. Prospective Diabetes Study Group. U.K. prospective diabetes study 16: overview of 6 years' therapy of type II diabetes: a progressive disease. *Diabetes*. 1995;44:1249–1258.
111. Ward W, Bolgiano D, McKnight B, et al. Diminished β cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. *J Clin Invest*. 1984;74:1318–1328.
112. Reaven GM, Olefsky JM. Relationship between heterogeneity of insulin responses and insulin resistance in normal subjects. *Diabetologia*. 1977;13:201–206.
113. Orchard TJ, Temprosa M, Goldberg R, et al. The effect of metformin and intensive lifestyle intervention on the metabolic syndrome: the Diabetes Prevention Program randomized trial. *Ann Intern Med*. 2005;142:611–619.
114. Gerstein HC, Santaguida P, Raina P, et al. Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: a systematic overview and meta-analysis of prospective studies. *Diabetes Res Clin Pract*. 2007;78:305–312.
115. Cook J, Page R, Levy J, et al. Hyperglycaemic progression in subjects with impaired glucose tolerance: association with decline in beta cell function. *Diabet Med*. 1993;10:321–326.
116. Kahn SE, Prigeon RL, McCulloch DK, et al. Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects: evidence for a hyperbolic function. *Diabetes*. 1993;42:1663–1672.
117. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity

and B-cell glucose sensitivity from the response to intravenous glucose. *J Clin Invest*. 1981;68:1456–1467.

118. Bergman RN, Ader M, Huecking K, et al. Accurate assessment of β-cell function: the hyperbolic correction. *Diabetes*. 2002;51:S212–S220.
119. Lorenzo C, Wagenknecht LE, Rewers MJ, et al. Disposition index, glucose effectiveness, and conversion to type 2 diabetes: the Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes Care*. 2011;33:2098–2103.
120. Elbein SC, Hasstedt SJ, Wegner K, et al. Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. *J Clin Endocrinol Metab*. 1999;84:1398–1403.
121. Polonsky K. The beta-cell in diabetes: from molecular genetics to clinical research. *Diabetes*. 1995;44:705–717.
122. Lang D, Matthews D, Peto J, et al. Cyclic oscillations of basal plasma glucose and insulin concentrations in human beings. *N Engl J Med*. 1979;301:1023–1027.
123. O'Rahilly S, Turner R, Matthews D. Impaired pulsatile secretion of insulin in relatives of patients with noninsulin-dependent diabetes. *N Engl J Med*. 1988;318:1225–1230.
124. Hunter S, Atkinson A, Ennis C, et al. Association between insulin secretory pulse frequency and peripheral insulin action in NIDDM and normal subjects. *Diabetes*. 1996;45:683–686.
125. Kanatsuka A, Kou S, Makino H. IAPP/amylin and β-cell failure: implication of the risk factors of type 2 diabetes. *Diabetol Int*. 2018;9:143–157.
126. Robertson RP. Chronic oxidative stress as a central mechanism for glucose toxicity in pancreatic islet beta cells in diabetes. *J Biol Chem*. 2004;279:42351–42354.
127. Vague P, Moulin JP. The defective glucose sensitivity of the B-cell in non-insulin-dependent diabetes: improvement after twenty hours of normoglycaemia. *Metabolism*. 1982;31:1139–1142.
128. Garvey WT, Olefsky JM, Griffin RF, et al. The effect of insulin treatment on insulin secretion and insulin action in Type II diabetes mellitus. *Diabetes*. 1985;34:222–234.
129. Garvey WT, Olefsky JM, Matthaei S, et al. Glucose and insulin co-regulate the glucose transport system in primary cultured adipocytes: a new mechanism of insulin resistance. *J Biol Chem*. 1987;262:189–197.
130. Rossetti L, Giaccari A, DeFronzo RA. Glucose toxicity. *Diabetes Care*. 1990;13:610–630.
131. Wolff SP, Dean RT. Glucose autoxidation and protein modification. The potential role of 'autoxidative glycosylation' in diabetes. *Biochem J*. 1987;245:243–250.
132. Unger RH. *Diabetes Mellitus: A Fundamental and Clinical Text*. Philadelphia: Lippencott Williams and Wilkins, Inc; 2004.
133. Robertson RP, Zhang HJ, Pyzdrowski KL, et al. Preservation of insulin mRNA levels and insulin secretion in HIT cells by avoidance of chronic exposure to high glucose concentrations. *J Clin Invest*. 1992;90:320–325.
134. Wing RR, Lang W, Wadden TA, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. *Diabetes Care*. 2011;34:1481–1486.
135. Lean MEJ, Leslie WS, Barnes AC, et al. Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster- randomized trial. *Lancet Diabetes Endocrinol*. 2019;7:344–355.
136. Garvey WT, Mechanick JI, Brett EM, et al. American association of clinical endocrinologists and American College of Endocrinology clinical practice guidelines for comprehensive medical care of patients with obesity. *Endocr Pract*. 2016;22:1–203.
137. Sjöström L, Peltonen M, Jacobson P, et al. Association of bariatric surgery with long-term remission of type 2 diabetes and with microvascular and macrovascular complications. *J Amer Med Assoc*. 2014;311:2297–2304.
138. Lara-Castro C, Newcomer BR, Rowell J, et al. Effects of short-term very low calorie diet on intramyocellular lipid and insulin sensitivity in non-diabetic and type 2 diabetic subjects. *Metabolism*. 2008;57:1–8.
139. Vitola BE, Deivanayagam S, Stein RI, et al. Weight loss reduces liver fat and improves hepatic and skeletal muscle insulin sensitivity in obese adolescents. *Obesity*. 2009;17:1744–1748.
140. Mahadevan J, Parazzoli S, Oseid E, et al. Ebselen treatment prevents islet apoptosis, maintains intranuclear Pdx-1 and MafA levels, and preserves β-cell mass and function in ZDF rats. *Diabetes*. 2013;62:3582–3588.
141. Harmon JS, Bogdani M, Parazzoli SD, et al. ß-cell specific overexpression of glutathione peroxidase preserves intranuclear MafA and reverses diabetes in db/db mice. *Endocrinology*. 2009;150:4855–4862.
142. Briaud I, Harmon JS, Kelpe CL, et al. Lipotoxicity of the pancreatic beta-cell is associated with glucose-dependent esterification of fatty acids into neutral lipids. *Diabetes*. 2001;50:315–321.
143. Robertson RP, Harmon J, Tran PO, et al. Glucose toxicity in beta cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*. 2003;52:581–587.
144. Unger RH. *Diabetes Mellitus: A Fundamental and Clinical Text*. Philadelphia: Lippencott Williams and Wilkins, Inc; 2004.
145. Prentki M, Corkey BE. Are the beta-cell signaling molecules malonyl-CoA and cystolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM? *Diabetes*. 1996;45:273–283.
146. Harmon JS, Gleason CE, Tanaka Y, et al. Antecedent hyperglycemia, not hyperlipidemia, is associated with increased islet triacylglycerol content and decreased insulin gene mRNA level in Zucker Diabetic Fatty rats. *Diabetes*. 2001;50:2481–2486.
147. Olson LK, Redmon JB, Towle HC, et al. Chronic exposure of HIT cells to high glucose concentrations paradoxically decreases insulin gene transcription and alters binding of insulin gene regulatory protein. *J Clin Invest*. 1993;92:514–519.
148. Harmon JS, Stein R. Robertson, RP. Oxidative stress-mediated, post-translational loss of MafA protein as a contributing mechanism to loss of insulin gene expression in glucotoxic beta cells. *J Biol Chem*. 2005;280:11107–11113.
149. Abebe T, Mahadevan J, Bogachus L, et al. Nrf2/antioxidant pathway mediates β cell self-repair after damage by high-fat diet-induced oxidative stress. *JCI Insight.* 2017;2:e92854.
150. Kolterman OG, Gray RS, Griffin J, et al. Receptor and post-receptor defects contribute to the insulin resistance in noninsulin-dependent diabetes mellitus. *J Clin Invest*. 1981;68:957–969.
151. Welch S, Gebhart S, Bergman R, et al. Minimal model analysis of intravenous glucose tolerance test-derived insulin sensitivity in diabetic subjects. *J Clin Endocrinol Metab*. 1990;71:1508–1518.
152. Pei D, Jones CN, Bhargava R, et al. Evaluation of octreotide to assess insulin-mediated glucose disposal by the insulin suppression test. *Diabetologia*. 1994;37:843–845.
153. Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28:412–419.
154. Pisprasert V, Ingram KH, Lopez-Davila MF, et al. Limitations in the use of indices using glucose and insulin levels to predict insulin sensitivity: impact of race and gender and superiority of the indices derived from oral glucose tolerance test in African Americans. *Diabetes Care*. 2013;36:845–853.
155. Wilkinson L, Yi N, Mehta T, et al. Development and validation of a model for predicting incident type 2 diabetes using quantitative clinical data and a Bayesian logistic model: a nationwide cohort and modeling study. *PLoS Med*. 2020;17:e1003232.
156. Rothman D, Shulman R, Shulman G. 31P nuclear magnetic resonance measurements of muscle glucose-6-phosphate. *J Clin Invest*. 1992;89:1069–1075.
157. Golay A, DeFronzo R, Ferrannini E, et al. Oxidative and non-oxidative glucose metabolism in non-obese type 2 (non-insulin dependent) diabetic patients. *Diabetologia*. 1988;31:585–591.
158. Shulman G, Rothman D, Jue T, et al. Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy. *N Engl J Med*. 1990;322:223–228.
159. Abdul-Ghani M, DeFronzo RA. Fasting hyperglycemia impairs glucose- but not insulin-mediated suppression of glucagon secretion. *J Clin Endocroinol Metab*. 2007;92:1778–1784.
160. Campbell PJ, Mandarino LJ, Gerich JE. Quantification of the relative impairment in actions of insulin on hepatic glucose production and peripheral glucose uptake in non-insulin-dependent diabetes mellitus. *Metabolism*. 1988;37:15–21.

161. Ferrannini E, Bjorkman O, Reichard G. The disposal of an oral glucose load in healthy subjects: a quantitative study. *Diabetes*. 1985;34:580–588.
162. Niewoehner C, Nuttall F. Relationship of hepatic glucose uptake to intrahepatic glucose concentration in fasted rats after glucose load. *Diabetes*. 1988;37:1559–1566.
163. Basu A, Basu R, Shah P, et al. Type 2 diabetes impairs splanchnic uptake of glucose but does not alter intestinal glucose absorption during enteral glucose feeding: additional evidence for a defect in hepatic glucokinase activity. *Diabetes*. 2001;50:1351–1362.
164. Ludvik B, Nolan J, Roberts A, et al. Evidence for decreased splanchnic glucose uptake after oral glucose administration in non-insulin-dependent diabetes mellitus. *J Clin Invest*. 1997;100:2354–2361.
165. Garvey WT, Mechanick JI. Chapter 6: cardiometabolic disease: insulin resistance, obesity, and the metabolic syndrome. In: Fuster V, Harrington RA, Narula J, et al., eds. *Hurst's the Heart*. 15th ed. New York: McGraw-Hill; 2021.
166. DeFronzo R, Jacot E, Jequier E, et al. The effect of insulin on the disposal of intravenous glucose: results from indirect calorimetry and hepatic and femoral venous catheterization. *Diabetes*. 1981;30:1000–1007.
167. Edelman S, Laakso M, Wallace P, et al. Kinetics of insulin mediated and non-insulin mediated glucose uptake in man. *Diabetes*. 1990;39:955–964.
168. Garvey WT, Kolterman OG. Correlation of in vivo and in vitro actions of insulin in obesity and noninsulin-dependent diabetes mellitus: role of the glucose transport system. *Diabetes Metab Rev*. 1988;4:543–569.
169. Freidenberg G, Reichart D, Olefsky J, et al. Reversibility of defective adipocyte insulin receptor kinase activity in non-insulin dependent diabetes mellitus: effect of weight loss. *J Clin Invest*. 1990;82:1398–1406.
170. Maegawa H, Shigeta Y, Egawa K, et al. Impaired autophosphorylation of insulin receptors from abdominal skeletal muscles in nonobese subjects with NIDDM. *Diabetes*. 1991;40:815–819.
171. Youngren JF, Maddux BA, Sasson S, et al. Skeletal muscle content of membrane glycoprotein PC-1 in obesity. Relationship to muscle glucose transport. *Diabetes*. 1996;45:1324–1328.
172. Wick KR, Werner ED, Langlais P, et al. Grb10 inhibits insulin-stimulated insulin receptor substrate (IRS)-phosphatidylinositol 3-kinase/Akt signaling pathway by disrupting the association of IRS-1/IRS-2 with the insulin receptor. *J Biol Chem*. 2003;278:8460–8467.
173. Cusi K, Maezono K, Osman A, et al. Insulin resistance differentially affects the PI 3-kinase- and MAP kinase-mediated signaling in human muscle. *J Clin Invest*. 2000;105:311–320.
174. Zierath JR, Krook A, Wallberg-Henriksson H. Insulin action and insulin resistance in human skeletal muscle. *Diabetologia*. 2000;43:821–835.
175. Nolan JJ, Ludvik B, Beerdsen P, et al. Improvement in glucose tolerance and insulin resistance in obese subjects treated with troglitazone. *New Engl J Med*. 1994;331:1188–1189.
176. Turcotte LP, Fisher JS. Skeletal muscle insulin resistance: roles of fatty acid metabolism and exercise. *J Physical Therapy*. 2008;88:1279–1296.
177. Yassine HN, Marchetti CM, Krishnan RK, et al. Effects of exercise and caloric restriction on insulin resistance and cardiometabolic risk factors in older obese adults: a randomized clinical trial. *J Gerontol A Biol Sci Med Sci*. 2009;64:90–95.
178. Pratipanawatr T, Cusi K, Ngo P, et al. Normalization of plasma glucose concentration by insulin therapy improves insulin-stimulated glycogen synthesis in type 2 diabetes. *Diabetes*. 2002;51:462–468.
179. Andersen DK, Hundal RS, Rothman DL, et al. Effects of free fatty acids on glucose transport and IRS-1-associated phosphatidylinositol 3-kinase activity. *J Clin Invest*. 1999;103:253–259.
180. Tanti JF, Gremeaux T, van Obberghen E, et al. Serine/threonine phosphorylation of insulin receptor substrate 1 modulates insulin receptor signaling. *J Biol Chem*. 1994;269:6051–6057.
181. Mothe I, Van Obberghe E. Phosphorylation of insulin receptor substrate-1 on multiple serine residues, 612, 632, 662, and 731, modulates insulin action. *J Biol Chem*. 1996;271:11222–11227.
182. Paz K, Hemi R, LeRoith D, et al. A molecular basis for insulin resistance. elevated serine/threonine phosphorylation of IRS-1 and IRS-2 inhibits their binding to the juxtamembrane region of the insulin receptor and impairs their ability to undergo insulin-induced tyrosine phosphorylation. *J Biol Chem*. 1997;272:29911–29918.
183. Tanti JF, Jager J. Cellular mechanisms of insulin resistance: role of stress-regulated serine kinases and insulin receptor substrates (IRS) serine phosphorylation. *Curr Opin Pharmacol*. 2009;9:753–762.
184. Danielsen AG, Liu F, Hosomi Y, et al. Activation of protein kinase C alpha inhibits signaling by members of the insulin receptor family. *J Biol Chem*. 1995;270:21600–21605.
185. Tanti JF, Jager J. Cellular mechanisms of insulin resistance: role of stress-regulated serine kinases and insulin receptor substrates (IRS) serine phosphorylation. *Curr Opin Pharmacol*. 2009;9:753–762.
186. Ravichandran LV, Esposito DL, Chen J, et al. Protein kinase C-zeta phosphorylates insulin receptor substrate-1 and impairs its ability to activate phosphatidylinositol 3-kinase in response to insulin. *J Biol Chem*. 2001;276:3543–3549.
187. Hotamisligil GS, Shargill NS, Spiegelman BM. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science*. 1993;259:87–91.
188. Saghizadeh M, Ong JM, Garvey WT, et al. The expression of TNF alpha by human muscle. Relationship to insulin resistance. *J Clin Invest*. 1996;97:1111–1116.
189. Hotamisligil GS, Peraldi P, Budavari A, et al. IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF-alpha- and obesity-induced insulin resistance. *Science*. 1996;271:665–668.
190. Kanety H, Feinstein R, Papa MZ, et al. Tumor necrosis factor alpha-induced phosphorylation of insulin receptor substrate-1 (IRS-1). Possible mechanism for suppression of insulin-stimulated tyrosine phosphorylation of IRS-1. *J Biol Chem*. 1995;270:23780–23784.
191. Aguirre V, Werner ED, Giraud J, et al. Phosphorylation of Ser307 in insulin receptor substrate-1 blocks interactions with the insulin receptor and inhibits insulin action. *J Biol Chem*. 2002;277:1531–1537.
192. Miller BS, Shankavaram UT, Horney MJ, et al. Activation of cJun NH2-terminal kinase/stress-activated protein kinase by insulin. *Biochemistry*. 1996;35:8769–8775.
193. Hirosumi J, Tuncman G, Chang L, et al. A central role for JNK in obesity and insulin resistance. *Nature*. 2002;420:333–336.
194. Cai D, Yuan M, Frantz DF, et al. Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med*. 2005;11:183–190.
195. Verma IM, Stevenson JK, Schwarz EM, et al. Rel/NF-kappa B/I kappa B family: intimate tales of association and dissociation. *Genes Dev*. 1995;9:2723–2735.
196. Chung HY, Cesari M, Anton S, et al. Molecular inflammation: underpinnings of aging and age-related diseases. *Ageing Res Rev*. 2009;8:18–30.
197. Osborn O, Olefsky JM. The cellular and signaling networks linking the immune system and metabolism in disease. *Nat Med*. 2012;18:363–374.
198. Gao Z, Hwang D, Bataille F, et al. Serine phosphorylation of insulin receptor substrate 1 by inhibitor kappa B kinase complex. *J Biol Chem*. 2002;277:48115–48121.
199. Hundal RS, Petersen KF, Mayerson AB, et al. Mechanism by which high-dose aspirin improves glucose metabolism in type 2 diabetes. *J Clin Invest*. 2002;109:1321–1326.
200. Yuan M, Konstantopoulos N, Lee J, et al. Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of Ikk beta. *Science*. 2001;293:1673–1677.
201. Shoelson SE, Lee J, Yuan M. Inflammation and the IKK beta/I kappa B/NF-kappa B axis in obesity- and diet-induced insulin resistance. *Int J Obes Relat Metab Disord*. 2003;27:S49–S52.
202. Asante-Appiah E, Kennedy BP. Protein tyrosine phosphatases: the quest for negative regulators of insulin action. *Am J Physiol Endocrinol Metab*. 2003;284:E663–E670.
203. Ahmad F, Azevedo JL, Cortright R, et al. Alterations in skeletal muscle protein-tyrosine phosphatase activity and expression in insulin-resistant human obesity and diabetes. *J Clin Invest*. 1997;100:449–458.
204. Goldstein BJ, Ahmad F, Ding W, et al. Regulation of the insulin signaling pathway by cellular protein-tyrosine phosphatases. *Mol Cell Biochem*. 1998;182:91–99.
205. Mooney RA, Senn J, Cameron S, et al. Suppressors of cytokine signaling-1 and -6 associate with and inhibit the insulin receptor. A potential

mechanism for cytokine-mediated insulin resistance. *J Biol Chem*. 2001;276:25889–25893.

206. Emanuelli B, Peraldi P, Filloux C, et al. SOCS-3 is an insulin-induced negative regulator of insulin signaling. *J Biol Chem*. 2000;275:15985–15991.
207. Starr R, Willson TA, Viney EM, et al. A family of cytokine-inducible inhibitors of signaling. *Nature*. 1997;387:917–921.
208. Cantley LC. The phosphoinositide 3-kinase pathway. *Science*. 2002;296:1655–1657.
209. Du K, Herzig S, Kulkarni RN, et al. TRB3: a tribbles homolog that inhibits Akt/PKB activation by insulin in liver. *Science*. 2003;300:1574–1577.
210. Koo SH, Satoh H, Herzig S, et al. PGC-1 promotes insulin resistance in liver through PPAR-alpha-dependent induction of TRB-3. *Nat Med*. 2004;10:530–534.
211. Liu J, Wu X, Franklin JL, et al. Mammalian *tribbles* homolog 3 impairs insulin action in skeletal muscle: role in glucose-induced insulin resistance. *Am J Physiol*. 2010;298:E565–E576.
212. Zhang W, Liu J, Tian L, et al. TRIB3 mediates glucose-induced insulin resistance via a mechanism that requires the hexosamine biosynthetic pathway. *Diabetes*. 2013;62:4192–4200.
213. Zhang W, Wu M, Kim T, et al. Skeletal muscle TRIB3 mediates glucose toxicity in diabetes and high fat diet-induced insulin resistance. *Diabetes*. 2016;65:2380–2391.
214. Hunter SJ, Garvey WT. Insulin action and insulin resistance: diseases involving defects in insulin receptors, signal transduction, and the glucose transport effector system. *Am J Med*. 1998;105:331–345.
215. Kim YB, Nikoulina SE, Ciaraldi TP, et al. Normal insulin- dependent activation of Akt/protein kinase B, with diminished activation of phosphoinositide 3-kinase, in muscle in type 2 diabetes. *J Clin Invest*. 1999;104:733–741.
216. Leng Y, Karlsson HK, Zierath JR. Insulin signaling defects in type 2 diabetes. *Rev Endocr Metab Disord*. 2004;5:111–117.
217. Kotani K, Ogawa W, Matsumoto M, et al. Requirement of atypical protein kinase C lambda for insulin stimulation of glucose uptake but not for Akt activation in 3T3-L1 adipocytes. *Mol Cell Biol*. 1998;18:6971–6982.
218. Gonzalez E, McGraw TE. Insulin signaling diverges into Akt-dependent and -independent signals to regulate the recruitment/docking and the fusion of GLUT4 vesicles to the plasma membrane. *Mol Biol Cell*. 2006;17:4484–4493.
219. Håkan KR, Karlsson J, Zierath R, et al. Insulin-stimulated phosphorylation of the Akt substrate AS160 is impaired in skeletal muscle of type 2 diabetic subjects. *Diabetes*. 2005;54:1692–1697.
220. Garvey WT, Maianu L, Huecksteadt TP, et al. Pretranslational suppression of a glucose transporter protein causes cellular insulin resistance in non-insulin-dependent diabetes mellitus and obesity. *J Clin Invest*. 1991;87:1072–1081.
221. Garvey WT, Maianu L, Zhu J-H, et al. Evidence for defects in the trafficking and translocation of GLUT4 glucose transporters in skeletal muscle as a cause of human insulin resistance. *J Clin Invest*. 1998;101:2377–2386.
222. Maianu L, Keller SR, Garvey WT. Adipocytes express abnormalities in translocation and trafficking of vesicles containing GLUT4 and insulin-regulated aminopeptidase in type 2 diabetes. *J Clin Endocrinol Metab*. 2001;86:5450–5456.
223. Garvey WT, Maianu L, Zhu J-H, et al. Multiple defects in the adipocyte glucose transport system cause cellular insulin resistance in gestational diabetes: heterogeneity in the number and a novel abnormality in subcellular localization of GLUT 4 glucose transporters. *Diabetes*. 1993;42:1773–1785.
224. Hayashi T, Wojtaszewski JF, Goodyear LJ. Exercise regulation of glucose transport in skeletal muscle. *Am J Physiol*. 1997;273:E1039–E1051.
225. Mu J, Brozinick Jr JT, Valladares O, et al. A role for AMP-activated protein kinase in contraction- and hypoxia-regulated glucose transport in skeletal muscle. *Mol Cell*. 2001;7:1085–1094.
226. Coppack SW, Jensen MD, Miles J. In vivo regulation of lipolysis in humans. *J Lipid Res*. 1994;35:177–193.
227. Frühbeck G, Méndez-Giménez L, Fernández-Formoso J, et al. Regulation of adipocyte lipolysis. *Nutr Res Rev*. 2014;27:63–93.
228. Holm C. Molecular mechanisms regulating hormone-sensitive lipase and lipolysis. *Biochem Soc Trans*. 2003;31:1120–1124.
229. Fricke K, Heitland A, Maronde E. Cooperative activation of lipolysis by protein kinase A and protein kinase C pathways in 3T3-L1 adipocytes. *Endocrinology*. 2004;145:4940–4947.
230. Yang A, Mottillo EP. Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. *Biochem J*. 2020;477:985–1008.
231. Wang Y, Sullivan S, Trujillo M, et al. Perilipin expression in human adipose tissues: effects of severe obesity, gender, and depot. *Obes Res*. 2003;11:930–936.
232. Yeaman SJ. Hormone-sensitive lipase–new roles for an old enzyme. *Biochem J*. 2004;379:11–22.
233. Wedellova Z, Dietrich J, Siklova-Vitkova M, et al. Adiponectin inhibits spontaneous and catecholamine-induced lipolysis in human adipocytes of non-obese subjects through AMPK-dependent mechanisms. *Physiol Res*. 2011;60:139–148.
234. Fu Y, Luo N, Klein RL, et al. Adiponectin promotes adipocyte differentiation, insulin sensitivity, and lipid accumulation potential role in auto-regulation of adipocyte metabolism and adipose mass. *J Lipid Res*. 2005;46:1369–1379.
235. Fruebis J, Tsao TS, Javorschi S, et al. Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice. *Proc Natl Acad Sci USA*. 2001;98:2005–2010.
236. Miyazaki Y, Mahankali A, Wajcberg E, et al. Effect of pioglitazone on circulating adipocytokine levels and insulin sensitivity in type 2 diabetic patients. *J Clin Endocrinol Metab*. 2004;89:4312–4319.
237. Guiherme A, Virbasius JV, Puri V, et al. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat Rev Mol Cell Biol*. 2008;9:367–377.
238. Bollen M, Keppens S, Stalmans W. Specific features of glycogen metabolism in liver. *Biochem J*. 1998;336:19–31.
239. O'Brien RM, Streeper RS, Ayala JE, et al. Insulin-regulated gene expression. *Biochem Soc Trans*. 2001;29:552–558.
240. Puigserver P, Rhee J, Donovan J, et al. Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. *Nature*. 2003;423:550–555.
241. Samson SL, Wong NC. Role of Sp1 in insulin regulation of gene expression. *J Mol Endocrinol*. 2002;29:265–279.
242. Tang ED, Nunez G, Barr FG, et al. Negative regulation of the forkhead transcription factor FKHR by Akt. *J Biol Chem*. 1999;274:16741–16746.
243. Hall RK, Yamasaki T, Kucera T, et al. Regulation of phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein-1 gene expression by insulin. The role of winged helix/forkhead proteins. *J Biol Chem*. 2000;275:30169–30175.
244. Cao G, Liang Y, Broderick CL, et al. Antidiabetic action of a liver x receptor agonist mediated by inhibition of hepatic gluconeogenesis. *J Biol Chem*. 2003;278:1131–1136.
245. Yoon JC, Puigserver P, Chen G, et al. Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature*. 2001;413:131–138.
246. Matsumoto M, Ogawa W, Teshigawara K, et al. Role of the insulin receptor substrate 1 and phosphatidylinositol 3-kinase signaling pathway in insulin-induced expression of sterol regulatory element binding protein 1c and glucokinase genes in rat hepatocytes. *Diabetes*. 2002;51:1672–1680.
247. Raghow R, Yellaturu C, Deng X, et al. SREBPs: the crossroads of physiological and pathological lipid homeostasis. *Trends Endocrinol Metab*. 2008;19:65–73.
248. Herzig S, Long F, Jhala US, et al. Creb regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature*. 2001;413:179–183.
249. Nourooz-Zadeh J, Rahimi A, Tajaddini-Sarmadi J, et al. Relationships between plasma measures of oxidative stress and metabolic control in NIDDM. *Diabetologia*. 1997;40:647–653.
250. Nathan C. Specificity of a third kind: reactive oxygen and nitrogen intermediates in cell signaling. *J Clin Invest*. 2003;111:769–778.
251. Kaufman R, Scheuner D, Schröder M, et al. The unfolded protein response in nutrient sensing and differentiation. *Nat Rev Mol Cell Biol*. 2002;3:411–421.

252. Krebs M, Brunmair B, Brehm A, et al. The mammalian target of rapamycin pathway regulates nutrient-sensitive glucose uptake in man. *Diabetes*. 2007;56:1600–1607.
253. Shah OJ, Wang Z, Hunter T. Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. *Curr Biol*. 2004;14:1650–1656.
254. Sarbassov DD, Guertin DA, Ali SM, et al. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science*. 2005;307:1098–1101.
255. Um SH, Frigerio F, Watanabe M, et al. Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature*. 2004;431:200–205.
256. Hardie DG, Carling D, Carlson M. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Ann Rev Biochem*. 1998;67:821–855.
257. Kramer HF, Witczak CA, Fujii N, et al. Distinct signals regulate AS160 phosphorylation in response to insulin, AICAR, and contraction in mouse skeletal muscle. *Diabetes*. 2006;55:2067–2076.
258. Leclerc I, Kahn A, Doiron B. The 5'-AMP-activated protein kinase inhibits the transcriptional stimulation by glucose in liver cells, acting through the glucose response complex. *FEBS Lett*. 1998;431:180–184.
259. Yang J, Craddock L, Hong S, et al. AMP-activated protein kinase suppresses LXR-dependent sterol regulatory element-binding protein-1c transcription in rat hepatoma McA-RH7777 cells. *J Cell Biochem*. 2009;106:414–426.
260. Merrill GF, Kurth EJ, Hardie DG, et al. AICA riboside increases AMP-activated protein kinase, fatty acid oxidation, and glucose uptake in rat muscle. *Am J Physiol*. 1997;273:E1107–E1112.
261. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009;9:311–326.
262. Thalacker-Mercer AE, Ingram KH, Guo F, et al. BMI, RQ, diabetes, and gender affect the relationships between amino acids and clamp measures of insulin action in humans. *Diabetes*. 2014;63:791–800.
263. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17:448–453.
264. Tremblay F, Krebs M, Dombrowski L, et al. Overactivation of S6 kinase 1 as a cause of human insulin resistance during increased amino acid availability. *Diabetes*. 2005;54:2674–2684.
265. Nicholson JK, Holmes E, Kinross J, et al. Host-gut microbiota metabolic interactions. *Science*. 2012;336:1262–1267.
266. Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. *Nature*. 2012;489:242–249.
267. Campbell C, Grapov D, Fiehn O, et al. Improved metabolic health alters host metabolism in parallel with changes in systemic xeno-metabolites of gut origin. *PLoS One*. 2014;9. e84260.
268. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444:1027–1031.
269. Vrieze A, Van Nood E, Holleman F, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*. 2012;143:913–916.
270. Hjorth MF, Roager HM, Larsen TM, et al. Pre-treatment microbial Prevotella-to-Bacteroides ratio, determines body fat loss success during a 6-month randomized controlled diet intervention. *Int J Obesity*. 2017;42:580–583.
271. Marshall S. Role of insulin, adipocyte hormones, and nutrient-sensing pathways in regulating fuel metabolism and energy homeostasis: a nutritional perspective of diabetes, obesity, and cancer. *Sci STKE*. 2006; 2006:re7.
272. D'Alessandris C, Andreozzi F, Federici M, et al. Increased O-glycosylation of insulin signaling proteins results in their impaired activation and enhanced susceptibility to apoptosis in pancreatic beta-cells. *FASEB J*. 2004;18:959–961.
273. Cooksey RC, Hebert Jr LF, Zhu JH, et al. Mechanism of hexosamine-induced insulin resistance in transgenic mice overexpressing glutamine:fructose-6-phosphate amidotransferase: decreased glucose transporter GLUT4 translocation and reversal by treatment with thiazolidinedione. *Endocrinology*. 1999;140:1151–1157.
274. Liu J, Wu X, Franklin JL, et al. Mammalian *tribbles* homolog 3 impairs insulin action in skeletal muscle: role in glucose-induced insulin resistance. *Am J Physiol*. 2010;298:E565–E576.
275. Bar R, Muggeo M, Kahn C, et al. Characterization of insulin receptors in patients with the syndrome of insulin resistance and acanthosis nigricans. *Diabetologia*. 1980;18:209–216.
276. O'Rahilly S, Moller D. Mutant insulin receptors in syndromes of insulin resistance. *Clin Endocrinol*. 1992;36:121–132.
277. Podskalny J, Kahn C. Cell culture studies on patients with extreme insulin resistance. I. Receptor defects on cultured fibroblasts. *J Clin Endocrinol Metab*. 1982;54:261–268.
278. Grigorescu F, Flier JS, Kahn CR. Defect in insulin receptor phosphorylation in erythrocytes and fibroblasts associated with severe insulin resistance. *J Biol Chem*. 1984;259:15003–15006.
279. Arioglu E, Andewelt A, Diablo C, et al. Clinical course of the syndrome of autoantibodies to the insulin receptor (type B insulin resistance): a 28-year perspective. *Medicine (Baltim)*. 2002;81:87–100.
280. Taylor S, Grunberger G, Marcus-Samuels B, et al. Hypoglycemia associated with antibodies to the insulin receptor. *N Engl J Med*. 1982;307:1422–1426.
281. Kahn C, Baird K, Flier J, et al. Effects of autoantibodies to the insulin receptor on isolated adipocytes. *J Clin Invest*. 1977;60:1094–1106.
282. Taylor S, Barbetti F, Accili D, et al. Syndromes of autoimmunity and hypoglycemia: autoantibodies directed against insulin and its receptor. *Endocrinol Metab Clin North Am*. 1989;18:123–143.
283. Liao Y, Kwon S, Shaughnessy S, et al. Critical Evaluation of ATP III criteria in identifying insulin resistance with dyslipidemia. *Diabetes Care*. 2004;27:978–983.
284. Guo F, Garvey WT. Cardiometabolic disease risk in metabolically healthy and unhealthy obesity: stability of metabolic health status in adults. *Obesity*. 2016;24:516–525.
285. Guo F, Garvey WT. Trends in cardiovascular health metrics in obese adults: National Health and Nutrition Examination Survey (NHANES), 1988-2014. *J Am Heart Assoc*. 2016;5:e003619.
286. Weisberg SP, McCann D, Desai M, et al. Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest*. 2003;112:1796–1808.
287. Petersen MC, Shulman GI. Roles of diacylglycerols and ceramides in hepatic insulin resistance. *Trends Pharmacol Sci*. 2017;38:649–665.
288. Kadowaki T, Yamauchi T. Adiponectin and adiponectin receptors. *Endocr Rev*. 2005;26:439–451.
289. Lara-Castro C, Fu Y, Chung BH, et al. Adiponectin and the metabolic syndrome: mechanisms mediating the risk of metabolic and cardiovascular disease. *Curr Opinion Lipidol*. 2007;18:263–270.
290. Tian L, Luo N, Klein RL, et al. Adiponectin reduces lipid accumulation in macrophage foam cells. *Atherosclerosis*. 2009;202:152–161.
291. Evert AB, Dennison M, Gardner CD, et al. Nutrition therapy for adults with diabetes or prediabetes: a consensus report. *Diabetes Care*. 2019;42:731–754.

34

Diabetes Mellitus and Pregnancy

David Simmons and N. Wah Cheung

OUTLINE

Diabetes and Pregnancy and the Mother, 589
  Metabolic Effects of Pregnancy, 589
  Classification of Diabetes in Pregnancy, 590
  Pathogenesis of Gestational Diabetes Mellitus and Diagnosis, 590
  Diagnostic Criteria for Gestational Diabetes Mellitus, 591
  Glucokinase Monogenic Diabetes and Pregnancy, 592
  Bariatric Surgery and Gestational Diabetes, 592
  The Diagnosis of Gestational Diabetes After Bariatric Surgery, 592
Diabetes and Pregnancy and the Offspring, 592
  Congenital Malformations and Early Fetal Loss, 592
  Disturbances of Metabolism and Fetal Growth, 593
  Maternal Obesity and the Offspring, 594
Management of Diabetes in Pregnancy, 594
  Contraception and Preconception Management, 594
  Management of Type 1 Diabetes Mellitus in Pregnancy, 596
  Glycemia Monitoring, 597
  Diet and Physical Activity, 598
  Insulin Management, 598
  Management of Type 2 Diabetes and Overt Diabetes in Pregnancy, 599
  Management of Diabetes Complications in Pregnancy, 600
  Management of Gestational Diabetes Mellitus, 600
  Prevention of Gestational Diabetes, 601
Obstetric Management of Diabetes in Pregnancy, 602
  Obstetric Surveillance, 602
  Antenatal Steroids for Preterm delivery, 602
  Hypertensive Disorders of Pregnancy, 602
  Decrease in Insulin Requirements, 603
  Peripartum Management and Birth, 603
Long-Term Outcomes and Management of Gestational Diabetes, 603
  Long-Term Risks of Gestational Diabetes to the Mother, 603
  Benefits of Breastfeeding for the Mother, 604
  Postpartum Follow-up of Mothers, 604
  Postpartum Follow-Up of Offspring, 604
  Risk of Obesity to the Offspring, 604
  Risk of Neurocognitive Disorders and Autism in the Offspring, 604
  Potential Mechanisms for Long-term Effects of Preexisting Maternal Diabetes and Gestational Diabetes on the Offspring, 604
Summary and Future Directions, 605

DIABETES AND PREGNANCY AND THE MOTHER

Metabolic Effects of Pregnancy

Pregnancy is characterized by hormonal changes that influence maternal glucose, protein, and fat metabolism. In preparation for the demands of later pregnancy, early pregnancy is an anabolic phase in which there is an accumulation of maternal fat stores. Glucose is the major and obligate source of fuel for the fetoplacental unit and is mainly transported by glucose transporter-1. In early pregnancy there is a small decrease in insulin sensitivity, accompanied by modest rises in insulin secretion, followed by an increase in insulin sensitivity from approximately gestational weeks 7 to 18.

Late pregnancy is a catabolic state that ensures that energy stores are directed to the fetus, even at the expense of the mother, and has been described by Freinkel as a state of accelerated starvation.[1] This is characterized by marked increases in maternal insulin resistance, with insulin sensitivity dropping by 60%.

KEY POINTS

- Pregnancy is associated with a small decrease in insulin sensitivity up to week 7, then an increase in insulin sensitivity up to week 18, and then a 2- to 3-fold increase in insulin resistance up to term. This has major implications for diabetes in pregnancy management.

The increase in insulin resistance diverts glucose to the fetus. The increased insulin resistance is accompanied by a 3-fold increase in insulin secretion. This may not be entirely as a result of a compensatory response of individual β cells, as both increased β cell hypertrophy and hyperplasia have been observed in animal studies. Despite there also being an increase in basal hepatic gluconeogenesis, maternal fasting glucose levels fall during normal pregnancy.

The increase in insulin resistance of pregnancy is largely mediated by placental hormones and cytokines, in particular human placental lactogen (HPL), placental growth hormone (GH), tumor necrosis factor-α (TNF-α), corticotropin-releasing hormone (CRH), and progesterone. There is a marked rise in HPL, placental GH, TNF-α, and CRH toward the end of the second trimester. There is also an increase in maternal insulin-like growth factor-I, cortisol, prolactin, and free fatty acids (FFAs), with the former two driven by placental GH and CRH, respectively.

Lipid metabolism is altered later in pregnancy, with a marked rise in triglyceride levels and, to a lesser extent, cholesterol. Increased lipolysis mobilizes fatty acids, which serve as an alternative fuel source for the mother. Placental lipoprotein lipase (LPL) also hydrolyses dietary triglycerides to FFAs, which become available for fetal uptake.[2] The increase in FFAs predisposes to higher plasma levels of ketone bodies during periods of fasting. The metabolic adaptations

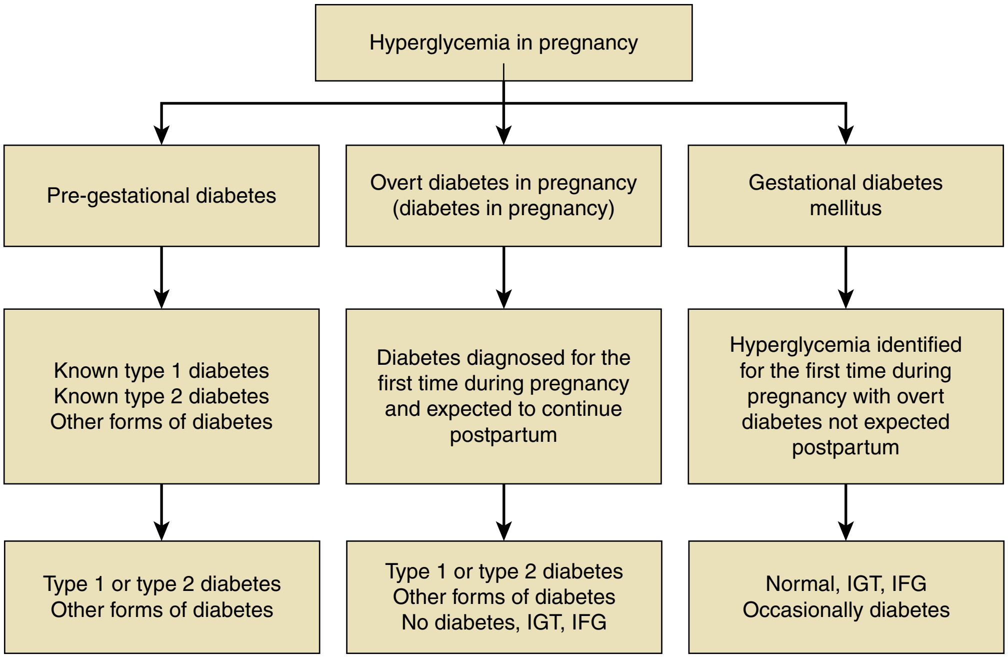

Hyperglycemia in pregnancy
Pre-gestational diabetes
Overt diabetes in pregnancy (diabetes in pregnancy)
Gestational diabetes mellitus
Known type 1 diabetes
Known type 2 diabetes
Other forms of diabetes
Diabetes diagnosed for the first time during pregnancy and expected to continue postpartum
Hyperglycemia identified for the first time during pregnancy with overt diabetes not expected postpartum
Type 1 or type 2 diabetes
Other forms of diabetes
Type 1 or type 2 diabetes
Other forms of diabetes
No diabetes, IGT, IFG
Normal, IGT, IFG
Occasionally diabetes

Fig. 34.1 Current classification of hyperglycemia in pregnancy and postpartum state. *IFG*, Impaired fasting glucose; *IGT*, impaired glucose tolerance.

in pregnancy are geared toward protein accretion by the maternal-fetal unit. Maternal protein turnover appears to be increased, with protein synthesis increasing by 25%, against a 10% decrease in amino acid oxidation. Circulating levels of amino acids are lower during pregnancy, but it is unclear if this is related to increased fetal uptake of amino acids, accelerated maternal uptake of amino acids, or splanchnic pooling.

The net result of the changes in maternal macronutrient metabolism is to prioritize the supply of glucose to the fetus, provide an adequate alternative energy source for the mother in the form of ketones and FFAs, and protect the mother against uncontrolled protein catabolism.

Classification of Diabetes in Pregnancy

Diabetes in pregnancy, also known as hyperglycemia in pregnancy (HIP), is generally divided into diabetes that predated the pregnancy (type 1 diabetes mellitus [T1DM], type 2 diabetes mellitus [T2DM], secondary forms of diabetes, and monogenic diabetes) and hyperglycemia that is first detected in pregnancy (Fig. 34.1). Hyperglycemia which is first detected in pregnancy includes those with likely pre-existing diabetes that was undiagnosed. Hence, hyperglycemia which is first detected in pregnancy into "overt diabetes in pregnancy" (ODIP) and "gestational diabetes mellitus" (GDM). In this classification, ODIP refers to women who had an oral glucose tolerance test (OGTT) result meeting the non-pregnant criteria for diabetes, and GDM refers to the milder form of HIP.[3] This classification assumes that women with ODIP are at higher risk of adverse pregnancy outcomes and are more likely to have diabetes post-partum, but this is not always the case.

In reality, it is increasingly recognized that diabetes in pregnancy represents a spectrum, from preexisting T1DM and T2DM to GDM, which classically was thought to develop toward the end of the second trimester. There is growing appreciation that many women have prediabetes or mild hyperglycemia before pregnancy, leading to GDM early in pregnancy. Criteria for early GDM are under active debate.

Pathogenesis of Gestational Diabetes Mellitus and Diagnosis

GDM has been defined as carbohydrate intolerance that is first discovered during pregnancy. However, it is now recognized that it is a heterogeneous disorder: some women diagnosed with GDM have preexisting unrecognized glucose intolerance, and many have subtle metabolic disturbances. The diagnosis of GDM is controversial, with varying criteria used in different parts of the world, and limited data are available regarding what level of glucose is harmful in early pregnancy.

GDM is classically considered to develop toward the end of the second trimester of pregnancy, coinciding with the rise in insulin resistance of late pregnancy. The increase in insulin resistance is similar between women with GDM and those with normal glucose tolerance.[4,5] The major difference resulting in hyperglycemia is that women with GDM fail to adequately compensate with increased insulin secretion, with a most notable decrease in first-phase insulin response.[4]

It is now recognized that more subtle degrees of hyperglycemia may develop early in pregnancy, although the significance remains unclear. It has also been proposed that an exaggerated fetal glucose phenomenon occurs when fetal hyperinsulinemia is established early in pregnancy.[6] This increases glucose flux across the placenta and continues to expose the fetus to glucose excess, without the mother necessarily becoming hyperglycemic. This highlights the active role that the placenta plays in determining fetal growth during HIP, although the mechanisms are not fully understood.

In GDM the maternal metabolic derangement is not limited to that of glucose metabolism. In the fasting state, FFAs and triglycerides are higher than in normal pregnancy.[2] Insulin-sensitive amino acids such as leucine, isoleucine, and valine are increased. Transplacental transfer of these additional nutrients to the hyperglycemic fetal milieu further fuels excessive fat accretion and growth. There is some evidence that maternal FFA and triglycerides may be a stronger predictor of fetal fat accretion and large for gestational age (LGA) babies in GDM than glucose.[2]

TABLE 34.1 Diagnostic Criteria for Gestational Diabetes

Criterion	Glucose Load	Abnormal Values Required	Fasting mmol/L (mg/dL)	1-Hour mmol/L (mg/dL)	2-Hour mmol/L (mg/dL)	3-Hour mmol/L (mg/dL)	Comment
O'Sullivan and Mahan, 1964	100g	2	5 (90)	9.2 (166)	8.1 (146)	6.9 (124)	Based on whole blood, no longer used
NDDG, 1979	100 g	2	5.8 (105)	10.6 (191)	9.2 (166)	8.1 (146)	
Carpenter and Coustan, 1982	100 g	2	5.3 (95)	10 (180)	8.6 (155)	7.8 (140)	When following a two-step strategy, one abnormal value can be used for diagnosis
ADA, 2000	75 g	1	5.3 (95)	10 (180)	8.6 (155)		No longer used
IADPSG/WHO, 2010	75 g	1	5.1 (92)	10 (180)	8.5 (155)		

AD, American Diabetes Association; *IADPSG*, International Association of Diabetes and Pregnancy Study Group; *NDDG*, National Diabetes Data Group; *WHO*, World Health Organization.

Diagnostic Criteria for Gestational Diabetes Mellitus

The diagnosis of GDM on the basis of an OGTT was first defined by O'Sullivan and Mahan in 1964.[7] In their seminal study of 752 women given a 3-hour 100-g OGTT during pregnancy and a further 1013 women followed with periodic OGTTs for up to 8 years, the glucose criteria for the diagnosis of GDM were determined on the basis of a high likelihood (23%) of the mother developing diabetes.

Since that time there have been a number of modifications to the O'Sullivan and Mahan (Table 34.1) criteria, with the National Diabetes Data Group and then the Carpenter and Coustan criteria being modifications of the O'Sullivan and Mahan criteria. The American Diabetes Association uses the same thresholds as Carpenter and Coustan but shortened the test to a 2-hour 75-g OGTT.[8]. More recently, the International Association of Diabetes in Pregnancy Study Groups (IADPSG) has developed new criteria based on the outcomes of the Hyperglycemia and Pregnancy Outcomes (HAPO) study.[9]

The HAPO study evaluated 25,505 women who had a 75-g OGTT at 24 to 32 weeks' gestation.[10] The results of the OGTT were blinded to both the women and caregivers, and therefore there was no treatment for GDM. As a safety net, unblinding and treatment only occurred for women who had a fasting plasma glucose of greater than or equal to 5.9 mmol/L (106 mg/dL) or a 2-hour result greater than or equal to 11.1 mmol/L (200 mg/dL). The investigators examined pregnancy outcomes for women whose OGTTs remained blinded, so that the relationship between the glucose levels and pregnancy outcomes could be established. There was a continuous relationship between the fasting, 1-hour, and 2-hour glucose levels on the OGTT and the four primary outcomes of the study (birth weight >90th percentile, primary cesarean section, clinically detected neonatal hypoglycemia, and cord blood c-peptide level >90th percentile) and no evidence of a threshold effect. Consensus criteria were subsequently developed, with the diagnostic cutoffs determined on the basis of glucose levels at which there was a 1.75 times increase in the odds ratio risk of the four primary endpoints, relative to the mean, at the three time points of the OGTT.[9] With these IADPSG criteria, one abnormality out of a fasting glucose level greater than or equal to 5.1 mmol/L (92 mg/dL), 1-hour glucose level greater than or equal to 10 mmol/L (180 mg/dL), and 2-hour glucose level greater than or equal to 8.5 mmol/L (153 mg/dL) establishes a diagnosis of GDM. These criteria are supported by the WHO and the International Diabetes Federation and are among the criteria accepted by the American Diabetes Association.[3,8] However, they remain contentious, with a National Institutes of Health panel indicating that there is insufficient evidence to support this approach.

Screening for hyperglycemia early in pregnancy is commonly recommended for women at high risk of GDM, and early GDM is associated with worse pregnancy outcomes.[11] If the glycated hemoglobin (HbA1c) or OGTT results performed early in pregnancy are consistent with diagnostic criteria for diabetes outside of pregnancy (i.e., HbA1c ≥6.5% [47 mmol/mol], and/or fasting glucose ≥7.0 mmol/L [126 mg/dL], 2-hour glucose ≥11.1 mmol/L [200 mg/dL]), then it is likely that the woman had preexisting diabetes and would be classified as having ODIP. The diagnosis of early GDM on the basis of the IADPSG criteria, however, is problematic. There are limited data relating the OGTT results at this stage of pregnancy to outcomes, and many centers empirically apply the IADPSG criteria. Moreover, a fasting glucose level of greater than or equal to 5.1 mmol/L (92 mg/dL) in early pregnancy is a poor predictor of the woman meeting diagnostic criteria for GDM at 24 to 38 weeks. The IADPSG has therefore recommended against using a 5.1-mmol/L (92-mg/dL) threshold for the diagnosis of early GDM,[9] and new data have led to dropping of this threshold.[11] There is no consensus as to how to diagnose GDM in early pregnancy. The Treatment of Booking Gestational Diabetes Mellitus trial, which is currently underway, will help to answer this question, as well as the value of early treatment.[12] The American Diabetes Association recommends that women with either impaired fasting glucose (fasting blood glucose ≥5.6 mmol/L [100 mg/dL]) and/or impaired glucose tolerance (2-hour post 75-g glucose ≥7.8 mmol/L [140 mg/dl]) are treated as having hyperglycemia in early pregnancy.

KEY POINTS

- Hyperglycemia in pregnancy includes women with type 1 diabetes mellitus, type 2 diabetes mellitus (T2DM), rare forms of diabetes, diabetes in pregnancy (possible undiagnosed T2DM), and gestational diabetes mellitus (GDM). GDM includes women with mild hyperglycemia from conception and those who develop mild hyperglycemia in the second and third trimesters.

The Incidence of Gestational Diabetes. The incidence of GDM varies depending on the criteria used for diagnosis, the timing of the OGTT, and the population in question. Factors that affect the likelihood of GDM include ethnicity, maternal age and obesity, previous history of GDM or polycystic ovarian syndrome, family history of diabetes, previous macrosomia, and multiple pregnancy. The incidence of GDM has been increasing, rising from 0.3% of pregnancies in the United States on the basis of hospital discharge data in 1979 to 1980 to 5.8% in 2008 to 2010.[13] The incidence in most other countries is higher.

With the lower fasting glucose threshold and the removal of a two-step process for the diagnosis of GDM, the new IADPSG criteria have resulted in more women being diagnosed with GDM. In the multinational population included in the HAPO study,[10] the overall incidence of GDM, diagnosed between 24 and 32 weeks' gestation by the IADPSG criteria, was 17.8%. The WHO estimates that, in 2019, 20 million women worldwide had some form of diabetes in pregnancy, with 84% of these being GDM.

Glucokinase Monogenic Diabetes and Pregnancy

With the low fasting glucose threshold of greater than or equal to 5.1 mmol/L (92 mg/dL) in the IADPSG criteria, glucokinase–maturity onset diabetes of the young (GCK-MODY), which is characterized by a high fasting glucose level, generally between 5.5 and 8.0 mmol/L (99–144 mg/dL), and typically a rise in glucose level on the OGTT of less than 4.6 mmol/L (83 mg/dL), is more likely to be misdiagnosed as GDM upon OGTT. It has been estimated that 0.9% of cases of GDM in fact have GCK-MODY.[14] The detection of GCK-MODY is important because the mother requires treatment only if the fetus does not inherit the genetic mutation. The combination of a fasting glucose greater than or equal to 5.5 mmol/L (99 mg/dL) and a normal body mass index (BMI) has been proposed as a clinical screening criterion to consider genetic testing of the mother for GCK-MODY.[15]

Bariatric Surgery and Gestational Diabetes

The high prevalence of obesity in the community has led to increasing uptake of bariatric surgery. In 2016, it is estimated that 162,969 bariatric surgery operations were performed in the United States.[16] Some 80% of operations were for women, and most women undergoing bariatric surgery are of reproductive age. Both obesity and bariatric surgery have significant ramifications for the incidence, diagnosis, and management of GDM.

Obesity is associated with increased pregnancy risks, including perinatal death, preeclampsia, GDM, and delivery by cesarean section. The literature is discordant as to whether women who have had bariatric surgery have a higher or lower risk of serious adverse pregnancy outcomes, which in part may depend upon the comparator group.[16-18] One metaanalysis showed that, compared with controls matched for presurgery BMI, there are reductions in gestational hypertension, preeclampsia, macrosomia, and cesarean section.[17] The odds ratio for developing GDM was 0.21 (95% confidence interval [CI]: 0.12-0.36). However, when women who had bariatric surgery were compared with control women matched for prepregnancy BMI, there was no difference in GDM. This suggests that the likelihood of adverse outcomes is primarily related to prepregnancy BMI, rather than being an independent effect of the surgery.

Of concern, and relevant to the management of GDM, bariatric surgery is associated with an increase in small for gestational age (SGA) babies and prematurity.[17-19] Although LGA is more commonly associated with GDM, close attention also needs to be paid to intrauterine growth restriction, especially for women who have had bariatric surgery. There are no data regarding how aggressively treatment to glucose targets should be pursued in pregnancies where there is SGA, therefore it would be prudent to take care to avoid overtreatment in this situation.

The greater incidence of SGA may be related to nutritional deficiencies, which have been observed independent of the type of surgery.[18] Micronutrient deficiencies are common following bariatric surgery, and supplementation is necessary during pregnancy.[20] Women with diabetes in pregnancy should generally avoid simple carbohydrates, but an additional reason to do so following bariatric surgery is the risk of dumping syndrome, which can occur in up to 15% of people who undergo bypass operations. Achievement of maternal protein requirements may be more challenging following bypass surgery, and a minimum of 60 g of protein per day has been recommended.[20] Energy requirements need to be individualized on the basis of BMI, gestational weight gain, and physical activity, with close dietetic monitoring being essential.[20]

The Diagnosis of Gestational Diabetes After Bariatric Surgery

The interpretation of the OGTT during pregnancy is difficult for women who have had bariatric surgery. The results are affected by the more rapid transit and absorption of a glucose load, as well as the heightened insulin and incretin response that may occur with late dumping syndrome. It may not be well tolerated, and hypoglycemia can occur. One study found that pregnant women who had gastric bypass surgery had lower fasting glucose levels, higher 1-hour glucose levels, and lower 2-hour glucose levels than BMI-matched controls (4.2 ± 0.4 vs. 4.7 ± 0.5 mmol/L [76 ± 7.2 vs. 85 ± 9.0 mg/dL], 9.5 ± 2.4 vs. 8.2 ± 1.8 mmol/L [171 ± 43 vs. 148 ± 32 mg/dL], and 3.7 ± 1.6 vs. 6.2 ± 1.7 mmol/L [67 ± 29 vs. 112 ± 31 mg/dL], respectively).[21] Half the women who had bypass surgery had hypoglycemia, defined as a serum glucose less than 3.3 mmol/L (60 mg/dL) at 2 hours. Another study using 100-g OGTTs found that hypoglycemia at 2 hours (defined as plasma glucose <3.1 mmol/L [56 mg/dL]) occurred at 2 or 3 hours among 83% of women who had Roux-en-Y gastric bypass, 55% of those who had a sleeve gastrectomy, and even 12% of those who had a gastric band procedure.[22]

Therefore, an OGTT cannot be reliably used to diagnose GDM in women who have had bariatric surgery, although it may be more acceptable for women who have had a gastric band procedure.[20] Recommendations range from undertaking a fasting glucose level and an HbA1c using existing thresholds, to performing capillary glucose day profiles or continuous glucose monitoring.[14,20,23] Although the latter two options appear sensible, as treatment is determined on the basis of glucose profiles, the altered glucose kinetics arising from bariatric surgery may invalidate the usual glucose targets used in pregnancy, and there are no data relating these to pregnancy outcomes.

DIABETES AND PREGNANCY AND THE OFFSPRING

The new IADPSG/WHO criteria for GDM have been developed based upon the potential for perturbations of glycemia to increase fetal growth, particularly adiposity, and fetal insulin secretion, with its effect on fetal growth and the risk of neonatal hypoglycemia. Higher levels of hyperglycemia are associated with greater risks of adverse metabolic outcomes, with additional impacts upon organogenesis.

Congenital Malformations and Early Fetal Loss

Organogenesis occurs in the first 8 weeks of pregnancy, and disruption of this process can lead to congenital malformations. The background rate of congenital malformations is approximately 2% to 3%. Many malformations are lethal, contributing to the background early fetal

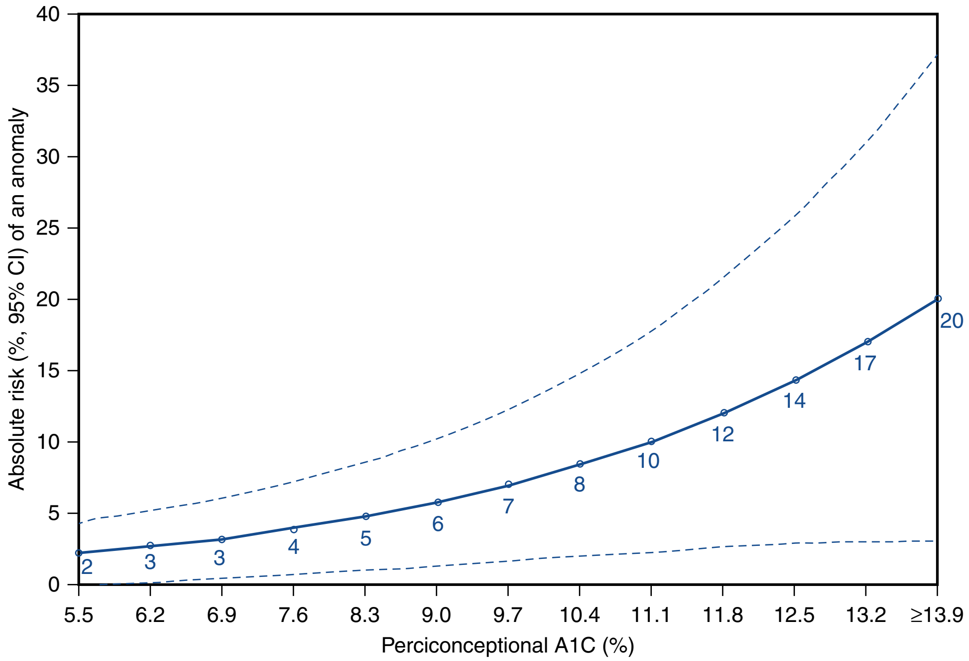


Fig. 34.2 Relationship between glycated hemoglobin in percent and absolute risk of a fetal anomaly. *A1C,* Glycated hemoglobin test; *CI,* confidence interval.

loss. Malformation rates are higher with increasing hyperglycemia as assessed by periconception HbA1c ,as shown in Fig. 34.2.[24]

It is currently unclear whether there is a threshold for increased risk for malformations and spontaneous abortions, previously reported to be around 6.7 to 7.2 mmol/L (120–130 mg/dL). However, malformation rates are also higher among obese women and women with mild hyperglycemia at conception (e.g., with MODY 2).

The more common major malformations include congenital heart defects and neural tube defects (NTDs; including anencephaly and spina bifida). Even when periconception HbA1c is near normal, malformation rates are higher than the background rate. Rates are similar between women with T1DM and T2DM in spite of the generally lower HbA1c at conception among the latter. Other factors likely to contribute to the high malformation rates are obesity, the extent of uptake of prepregnancy folic acid, and possibly antihyperglycemic, antihypertensive, and lipid-lowering medications.

The biology of malformations is largely separate from the biology of macrosomia and the mid to late pregnancy fetal complications. Animal and embryonic stem cell research has revealed several embryonic developmental processes that are affected by maternal hyperglycemia, leading to "diabetic embryopathy."[25] Expression of the glucose transporter 2 (GLUT-2) confers susceptibility to malformations and is important for transporting both glucose and the amino sugar glucosamine. Transient hyperglycemia inhibits Pax3 expression (involved in cell signaling), increases neural tube apoptosis, and increases NTDs. Other mechanisms include oxidative stress, increased generation of free radicals, reduced glucosamine, and disruption of other signaling pathways.

Disturbances of Metabolism and Fetal Growth

Fetal growth is determined by interactions between genetic, maternal, placental, and fetal factors. Maternal size, reflecting some of these factors, and as a composite of both lean mass and adiposity, is a major correlate of overall growth. Nutrient transfer and nutrient use by the fetus are regulated by a variety of maternal, placental, and fetal endocrine mechanisms, including insulin.

Whereas diabetes is defined by random, fasting, and/or postprandial hyperglycemia, other fuels (amino acids, lipids, ketone bodies) are also disturbed. Maternal insulin levels are insufficient in relation to need, ranging from minimal to absent insulin secretory capacity in T1DM to sometimes high levels in, for example, GDM. The original work by Pederson[26] proposed that "maternal hyperglycemia results in fetal hyperglycemia and, hence, hypertrophy of fetal islet tissue with insulin-hypersecretion." Freinkel[1] built upon the Pederson hypothesis, including observations beyond the short-term effects of fetal hyperinsulinemia, with associated macrosomia, neonatal hypoglycemia, polycythemia, and neonatal jaundice, but also long-term effects on behavior (e.g., autism) and long-term metabolic risk (e.g., obesity, dysglycemia). The results from the HAPO study showed a linear relationship between maternal glycemia (from normal to overt diabetes), fetal insulin, and fetal growth.[10] This relationship continues beyond excess adiposity at birth to measures of body composition up to 14 years. Fetal hyperinsulinemia can be great enough to create a maternal-fetal gradient that drives maternal glucose toward the fetus, known as the "fetal-steal" phenomenon.[6] This is postulated to increase fetal adiposity while lowering maternal glucose to levels that falsely reassure treating clinicians.

KEY POINTS

- The Pederson hypothesis proposes that excess maternal glucose crossing the placenta stimulates fetal hyperinsulinemia, leading to macrosomia, neonatal hypoglycemia, polycythemia, and neonatal jaundice. The Hyperglycemia and Adverse Pregnancy Outcomes Study showed a linear relationship between maternal glycemia and fetal insulin and growth.

However, disturbance of the other fuels is also likely to contribute to fetal growth. For example, the elevated maternal FFA and triglyceride concentrations correlate with macrosomia. Maternal triglycerides are hydrolyzed at the placental membrane by placental lipases (e.g., placental LPL) releasing FFAs, which then either diffuse passively or via facilitated placental transport. FFAs can then be directly transported

TABLE 34.2 Institute of Medicine Recommendations for Gestational Weight Gain

Prepregnancy Body Mass Index	Total Weight Gain (kg)	Rate of Weight Gain in 2nd and 3rd Trimester (mean (range) kg/week)
Underweight (<18.5 kg/m^2)	12.5–18	0.51 (0.44–0.58)
Normal weight (18.5–24.9 kg/m^2)	11.5–16	0.42 (0.35–0.50)
Overweight (25.0–29.9 kg/m^2)	7–11.5	0.28 (0.23–0.33)
Obese (≥30.0 kg/m^2)	5–9	0.22 (0.17–0.27)

Calculations assume 0.5–2 kg weight gain in the 1st trimester and presume a linear weight gain throughout the 2nd and 3rd trimesters.

to the fetal circulation from the placenta and metabolised through β-oxidation, with conversion of fatty acyl CoA to acylcarnitine being the rate-limiting step. Alternatively, FFAs can be recoupled with glycerol and stored as triglycerides in the placenta for future transfer to the fetus. However, monitoring and managing triglycerides has not been shown to affect pregnancy outcomes.[27]

More recently, and following the demonstration of heterogeneity in T2DM, there have been a number of publications supporting GDM as also being heterogenous.[28,29] Besides those with overt diabetes in pregnancy, developing T1DM (with antibodies) and monogenic diabetes (including glucokinase mutations), GDM appears to include three other phenotypes: "predominantly insulin-resistant," "predominantly reduced insulin secretion," and mixed insulin resistance/reduced insulin secretion. Women with the maternal insulin-resistant phenotype appear to be more likely to experience complications and adverse pregnancy outcomes (e.g., macrosomia).

Maternal Obesity and the Offspring

The prevalence of obesity and overweight during pregnancy has been growing over the last 30 years. Increasing BMI on entry into pregnancy is associated with increasing risks of GDM, and, even if euglycemic, with greater adiposity and more macrosomia in the offspring. Maternal periconception obesity and GDM development appear to be additive risks for babies being LGA at birth. Excess gestational weight gain is also associated with an increased risk of LGA babies. The Institute of Medicine (IOM)[30] has defined the degree of gestational weight gain expected in each trimester according to maternal BMI, as shown in Table 34.2.

Although excess gestational weight gain is associated with fetal fat accumulation, below the recommended gestational weight gain is associated with reduced fetal growth and a risk of SGA birth weight, with infant catch-up growth and a future risk of obesity, dysglycemia, and cardiovascular disease (the thrifty phenotype).[31] One area of uncertainty is the recommended gestational weight gain for women with a BMI greater than or equal to 30 kg/m^2: whereas the IOM guidelines advise 5 to 9 kg, some clinicians have shown fewer LGA babies and other complications with caloric restriction for a weight gain of 0 to 5 kg. However, this might also be associated with more SGA babies.[32]

MANAGEMENT OF DIABETES IN PREGNANCY

The goal of managing diabetes in pregnancy is to achieve a pregnancy outcome comparable to women without diabetes. Unfortunately, women with all forms of diabetes in pregnancy continue to experience adverse outcomes more frequently than the background population. Women with T1DM are usually aware of at least some of these risks, either before or after conception, and often with a negative "spin" from health professionals or others. Pregnancy is a special time for all women, including women with diabetes, with a range of both hopes and fears. However, women with diabetes have additional pressures, needing to add their diabetes regimen requirements to regular pregnancy-associated actions, all with the knowledge that their pregnancy is at an increased risk of an adverse outcome. As a result, communication of pregnancy risks (with their negative connotations) always needs to be balanced with a positive perspective regarding the potential for a healthy baby, as well as the benefits of pregnancy and of having a family.

Best practice for management is to take a team approach, including an endocrinologist (or internist) experienced in diabetes in pregnancy management, a certified diabetes educator, a dietitian, and an obstetrician or perinatologist. A neonatologist is likely to be needed at birth. Other team members, depending on need, include a social worker and/or a psychologist. Some countries also include a midwife, who may also have certified diabetes educator training, to support women during the pregnancy.

Contraception and Preconception Management

Conception with hyperglycemia is not only associated with an increased risk of malformations and early fetal loss, but also with an increased risk of perinatal mortality and of adverse pregnancy outcomes overall. Women who become pregnant with a high HbA1c usually also exhibit greater hyperglycemia through the pregnancy, which maps to the adverse pregnancy outcomes associated with LGA babies (e.g., shoulder dystocia, birth trauma, and cesarean section) and neonatal hyperinsulinism (neonatal hypoglycemia, neonatal hyperbilirubinemia (jaundice), polycythemia, stillbirth). Beyond managing hyperglycemia, preconception management also includes identifying and replacing any potentially fetoxic medications.

As a result of the need to optimize hyperglycemia before conception, the American Diabetes Association has recommended since 2009 that preconception counseling be given at each clinic visit for all women with diabetes of childbearing age, starting at puberty. This allows information and guidance to be provided regarding contraception choices if sexually active and commencement of prepregnancy or preconception management if pregnancy is actively being considered. On the whole, the preferred contraception choice is one with the highest effectiveness that is least likely to promote weight gain and impact on glycemia. These are essentially the long-acting reversible contraception (LARC) devices, that is an intrauterine contraception system (e.g., Mirena), an intrauterine hormonal contraception device, or a contraception implant (e.g., Implanon; Etonorgestrel). Two injectable preparations are currently marketed, depot-medroxy-progesterone acetate and norethisterone enantate, which are given every 3 and 2 months, respectively. Weight gain and delays in fertility return make injectable contraceptives less preferred options. Combination oral contraceptives containing estrogen and progestin, and progestin-only oral contraceptives are commonly used but are less effective than LARC devices. Any of the hormonal contraceptives can impact on glycemia. Barrier methods assist with sexually transmitted disease control and do reduce pregnancy risk but have a high contraception failure rate. Pregnancy timing methods again have a high failure rate and are not recommended but may be the only choice in those refusing other forms of contraception. Such women (and any pregnancy) would benefit from optimal glycemia management, and potentially folic acid supplementation. In spite of the range and accessibility of different contraception choices, at least 50% of women still report their pregnancies to be unplanned.

Management of glycemia in T1DM, including continuous subcutaneous insulin infusion (CSII) and glucose monitoring technologies, is discussed in detail in Chapter 39. The Continuous Glucose Monitoring in Type 1 Diabetes Pregnancy Trial (CONCEPTT)[33] showed no significant difference among women planning pregnancy between those using continuous glucose monitoring and those using self-blood glucose monitoring. Although human insulins are considered safe at conception and during pregnancy, analog insulins are considered different molecules requiring testing.[34] Insulin lispro, aspart, and detemir have all undergone clinical trials in pregnancy. Fast-acting aspart is considered identical to insulin aspart from a fetotoxicity point of view. Insulin degludec has a trial currently underway in pregnancy. Insulins glargine and glulisine have not yet undergone trials in pregnancy. Insulin glargine does have a number of case series reported that do not suggest teratogenicity or an impact on the fetus.

Although management of T2DM is discussed in Chapter 40, during the prepregnancy period all noninsulin diabetes medications have historically been stopped and an insulin regimen commenced tailored to the glucose profile of the woman. Whereas the noninsulin therapies are not necessarily teratogenic, fetotoxicity is a concern: the newer agents (e.g., SGLT2 inhibitors and GLP1 receptor agonists) have not been tested in pregnancy, and few case series have been reported of conception while using these drugs. One exception is the use of metformin. Using metformin is now known to be safe at the time of conception and through the first trimester.[35] However, some practitioners prefer to avoid metformin, as it is known to cross the placenta. If metformin is used prepregnancy and periconceptually, it is prudent to test for vitamin B_{12} deficiency because this can be associated with metformin use.

Cardiovascular and renal risk factors such as hypertension, dyslipidemia, and microalbuminuria and other comorbidities such as depression are more common among women with T2DM (and T1DM). Several of the drugs used for these conditions are also fetotoxic at different stages during pregnancy, and generally each is suspended during the prepregnancy and antenatal periods if alternatives exist that are as effective. As a result, lipid-lowering agents are usually stopped, and antihypertensive agents are replaced with labetalol, methyldopa, and calcium channel blockers acceptable for use in pregnancy. Antidepressants are decided on a case-by-case basis. The use of angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers is associated with an increased risk of fetal defects in treated women compared with the normotensive population, but not compared with hypertensive women. The use of ACE inhibitors and angiotensin receptor blockers for albuminuria therefore also needs to be discussed on a case-by-case basis, balancing the known benefits from renoprotection with the putative risks from first-trimester exposure to the drugs.[36]

Preconception management includes commencing 5 mg folic acid for use up to 12 weeks' gestation to reduce the risk of neurodevelopmental and cardiac malformations. Women who are not taking folic acid supplements before conception have a 2.2-fold (95% CI: 1.3–3.9) risk of adverse pregnancy outcomes.[37] Although 5 mg daily is recommended, the evidence that a high dose is required to reduce the risk of neurodevelopmental malformations is mixed. There is some evidence to suggest that doses of folic acid above 5 mg should not be used, as they may adversely influence psychomotor development in the future.[38] The recommended maximum daily dose of 5 mg folic acid should therefore include that contained in any other pregnancy supplements being taken (where possible).

The mother herself is at increased risk from changes in insulin sensitivity through the first trimester. Insulin sensitivity increases from approximately gestational week 6 to 7, resulting in an increased risk of severe hypoglycemia in insulin-treated women (mainly T1DM: up to 45% of women).[39] This can lead to loss of consciousness, seizures, and death if no support is available. If driving, there is a risk of a road crash and injury or death. Reducing insulin needs along with severe hypoglycemia may be the first indication of pregnancy in some women. As a result, preconception care needs to optimize diabetes self-management (particularly insulin management and hypoglycemia management) to allow the woman to adjust her own insulin in real time. Training partners to be able to administer glucagon, strictly adhering to driving requirements, and carrying diabetes identification are particularly important during this time. The earliest possible attendance at the diabetes in pregnancy clinic is recommended to help support insulin management during this period. Optimizing glycemia and insulin management often takes time, and contraception is recommended to avoid pregnancy during the preconception glucose stabilization period.

TABLE 34.3 Components of Preconception Care

- Multidisciplinary specialist team
- Screen for complications
- Review medications, ACE inhibitors/A2RB, statins—switch to safe diabetes medications
- Individual counseling re: pregnancy risks/prevention
- Risks of hypoglycemia during pregnancy
- Effective contraception until glycemia and management optimized
- Smoking cessation
- Start 5 mg folic acid
- Self–blood glucose monitoring 4–7 times/day or sensor
- Negotiate targets for HbA1c, blood glucose, time in range—intensification of glycemic management
- Dietitian input
- Immunization and other usual pregnancy preparation
- Promote early booking into diabetes in pregnancy clinic

ACE, Angiotensin-converting enzyme; *A2RB*, angiotensin 2 receptor blockers; *HbA1c*, glycated hemoglobin.

Optimizing glycemia presents further risks for women with preexisting diabetes, particularly if they are relatively hyperglycemic initially. Many women often wish for pregnancy "as soon as possible" for a range of personal and family reasons, leading to rapid improvement in glycemia. This can be associated with acceleration in the development of diabetic eye disease (maculopathy, proliferative retinopathy). Pregnancy itself can be associated with progression in diabetic eye disease,[39] and early attendance at the diabetes in pregnancy clinic also allows early retinal screening to be organized. Screening for microalbuminuria and renal function before pregnancy also allows discussion of pregnancy-related risks. In general, women with chronic kidney disease 3a or worse or significant proteinuria are counseled about antenatal deterioration in renal function and their increased risk of preeclampsia and extreme premature birth.[40] Table 34.3 summarizes the components of preconception care.

The high rates of unplanned pregnancy and pregnancy with suboptimal diabetes management has led to the development of specialized prepregnancy clinics. Attendance at a prepregnancy clinic is associated with a 71% reduction in risk of malformations (risk ratio [RR] 0.29; 95% CI: 0.21–0.40)[41] and a 54% reduction in perinatal mortality (RR 0.46; 95% CI: 0.30–0.73), suggesting that such clinics should be widely available. However, these clinics are more likely to be accessed by women who are already motivated, aware, and accepting that diabetes-related pregnancy risks can be reduced. As a result, a small number of population-based prepregnancy programs have been tested that generally include educational materials provided/sent to the women, educational programs for non–diabetes specialist healthcare providers (e.g., primary care), and prepregnancy clinics.[42] These programs have improved periconceptual glycemia, higher dose folic acid uptake, teratogenic medication avoidance, and early pregnancy clinic attendance.

KEY POINTS

- Maternal hyperglycemia at conception correlates with increasing fetal malformations. Attending prepregnancy clinics with counseling and education, optimizing glucose control, avoiding fetotoxic drugs, and taking 5 mg folate daily are associated with a 71% reduction in malformations and a 54% reduction in perinatal mortality.

Management of Type 1 Diabetes Mellitus in Pregnancy

Once a woman with T1DM has conceived, there is an urgency warranting intensive management, including evaluation of the existing maternal and fetal risks. Risk assessment for the endocrinologist includes a general medical review including weight, blood pressure, full foot examination and ensuring that ophthalmological examination has occurred. Renal function assessment (serum creatinine, estimated glomerular filtration rate, and microalbuminuria/proteinuria screening) and, when appropriate, cardiovascular assessment should occur. Glycemia history including diabetic ketoacidosis, weight management and particularly hypoglycemia history need to be considered. Hypoglycemia assessment includes the assessment of hypoglycemia awareness using a standard approach (e.g., Gold score) and potentially whether autonomic neuropathy is present (e.g., lying/standing blood pressure). Understanding the hypoglycemia risk is of particular importance in the second half of the first trimester, when insulin sensitivity increases, and the rate of hypoglycemia, and particularly life-threatening severe hypoglycemia, can increase. Counseling relating to the personal risks of pregnancy, including miscarriage, need to be undertaken sensitively within a positive and optimistic outlook. Commencing intensive glucose management and ensuring that folic acid (5 mg) is in place up to gestational week 12 (if not already in place) are part of this initial review.

Maximizing self-management skills is clearly particularly important during pregnancy, and this will have hopefully occurred before conception, including carbohydrate counting, quantitative approaches to mild hypoglycemia, glucagon use (for the partner), driving requirements, sick-day planning with blood ketone meter use, insulin site rotation, and fresh needle use. Self–blood glucose monitoring has been revolutionized by the development of continuous glucose monitoring systems (CGMSs) and flash glucose monitoring systems (FGMSs), also known as "sensors." Insulin management has similarly been evolving following the introduction of CSII technology, "smart" meters (with their increasing replacement by "apps"), and "closed loop" technology (where the CSII talks to the sensor).

Many of those with T1DM have a range of psychological, financial, social, and medical barriers to care and self-care, and these need to be carefully assessed at this point and throughout pregnancy. Pregnancy, and the need to manage glycemia tightly on a day-by-day basis to protect the growing fetus, places a major personal burden on many women. Some women may need to adapt their working life to adjust to the time demands of managing their diabetes in pregnancy. A small number of women may need to be hospitalized at some point during their pregnancy if it becomes clear that glycemia management needs a major correction, if there are major psychosocial barriers to self-management (e.g., family or work), or if there is a risk of severe hypoglycemia (or diabetic ketoacidosis). Naturally, such hospitalization and inpatient stabilization occurs in a dramatically different nondomestic, less mobile environment, with hypoglycemia risks (or reversion to prior self-management approaches) on discharge and reverting back to life at home.

Following conception, insulin needs vary weekly, as shown in Fig. 34.3.[43] Besides maximizing self-management skills, addressing this variation requires regular clinical reviews to ensure that proactive insulin regimen changes are made. This is usually achieved by a mix of face-to-face, online, and telephone consultations to be able to interpret the glycemic pattern, including any aberrations as a result of lifestyle events.

Fig. 34.3, using data from women with T1DM in pregnancy, shows the initial increase in insulin requirements as the glucose increases up to gestational week 7, then a decrease up to gestational week 12, slowly rising glucose to gestational week 30, and then a decrease after this point.

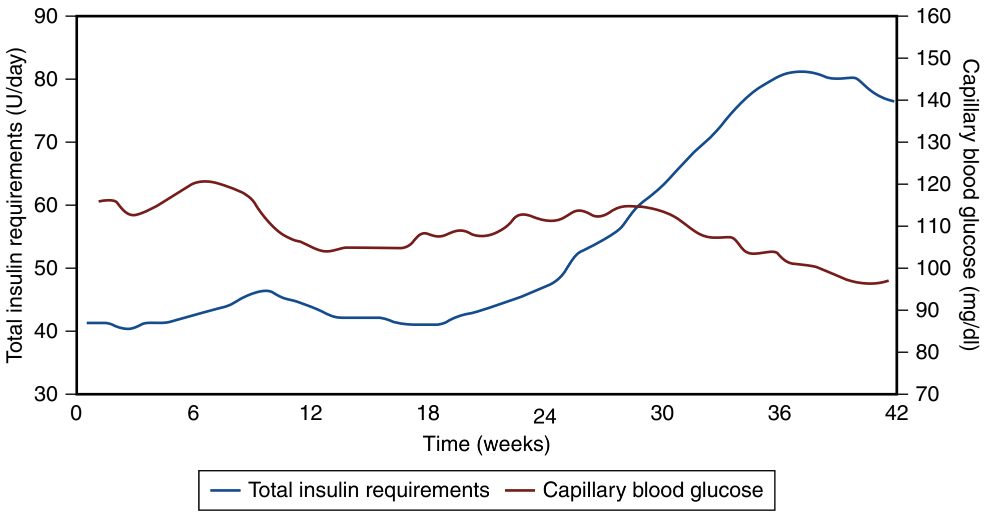


Fig. 34.3 Changes in insulin and glucose during pregnancy among women with type 1 diabetes. (Data from García-Patterson A, Gich I, Amini SB, et al. Insulin requirements throughout pregnancy in women with type 1 diabetes mellitus: three changes of direction. *Diabetologia*. 2010;53:446–451.)

Insulin management is tailored to glucose monitoring, although gestational weight gain above the IOM recommendations can indicate overinsulinization, along with the associated hypoglycemia and/or additional food consumption. Weight management needs to be monitored closely and, where necessary, discussed. Psychological factors, including fear of hypoglycemia, can impact on achievement of glycemia goals associated with optimal pregnancy outcomes.

Glycemia Monitoring

Traditionally (after the era of testing for glycosuria), self–blood glucose monitoring has been undertaken using a fingerprick to access capillary blood and a strip with a reagent to metabolize glucose from the sample and create an electric current measured by a blood glucose meter. This provided much more useful, and real-time, data than urine testing. Daily self-blood glucose monitoring during pregnancy for women with T1DM is recommended, as for nonpregnant adults, fasting (or at breakfast time for those who do not eat breakfast) to identify the effect of overnight insulin treatment, before meals to assess hypoglycemia risk, and before bed. However, during pregnancy, postprandial glycemic spikes can stimulate fetal hyperinsulinism and associated fetal adiposity, with their associated adverse pregnancy outcomes. As a result, postprandial monitoring is particularly important during pregnancy. Recommended glycemic targets are shown in Table 34.4 but need to be individualized with regards to hypoglycemia patterns, hypoglycemia unawareness, and physical activity and with an understanding of mothers' perspectives on their hypoglycemia risk.

In the first trimester, the rapidly changing glucose can pose challenges, and the initial upward glucose direction can be the first indication of pregnancy. However, as this drops after gestational weeks 6 to 7, and the hypoglycemia risk increases, it can be prudent to set higher preprandial targets (e.g., 5.5 mmol/L [99 mg/dL] or even up to 6.0 mmol/L [108 mg/dL]). There is little evidence that a slightly higher target in the first trimester is harmful and some evidence that hypoglycemia during the first trimester may be associated with an increased risk of macrosomia.[44] Another important glycemic pattern to actively seek is the overnight glucose. Besides its own intrinsic life-threatening risk, repeated nocturnal hypoglycemia can blunt hypoglycemia awareness and is possibly associated with fetal fuel undersupply and SGA babies. As symptoms are unreliable, nocturnal hypoglycemia may warrant self–blood glucose monitoring at 3:00 AM, the glucose nadir. The reasons for the hypoglycemia can be identified by reviewing the glucose pattern and insulin therapy among the evening meal, before bed, and 3:00 AM. The 3:00-AM test also allows the dawn phenomenon to be identified where this occurs. This is a critical piece of information, as it prevents an excess increase in overnight background/basal insulin based upon the before-breakfast glucose test results. Such women often require a small bolus quick-acting dose (without food) at approximately 4:00 AM (tailored to the individual) to prevent the rise. "Testing" at 3:00 AM is easily achievable using a sensor, of course.

TABLE 34.4 Recommended Glucose Targets During Pregnancy

	Preconception[a]	Pregnancy[a]
Fasting and premeal: mmol/L (mg/dL)	<6.0 (≤108)	<5.3 (≤95)
1 hour postmeal: mmol/L (mg/dL)	<8.5 (≤153)	<7.8 (≤140)
2 hours postmeal: mmol/L (mg/dL)	<7.5 (≤135)	<6.7 (≤120)

[a] If these targets can be safely achieved without causing significant hypoglycemia.

During pregnancy, when glucose targets are much tighter than outside pregnancy to protect the fetus, and normal fasting glucose is lower, meter accuracy and precision are of particular importance. However, many glucose meters in current use have a range of systematic and nonsystematic factors (e.g., hematocrit) that can lead to under- and overreading and greater variation.[45] Table 34.5 shows the factors that should be considered when deciding on which meter to use during pregnancy. Some meters ("smart meters") have inbuilt algorithms to provide advice on how much insulin to give based upon the amount of carbohydrate entered and the glucose level, with preentered carbohydrate ratio and correction factor along with insulin action. Such meters are now being replaced by apps that can undertake these tasks from a mobile phone.

There are several criticisms of self–blood glucose monitoring using fingersticks. In pregnancy, the ongoing variation in glycemia, tight glucose targets, and importance of postprandial hyperglycemia make the intermittent nature of self–blood glucose monitoring a suboptimal approach to managing T1DM. CGMS and FGMS glucose monitoring measure the glucose in the interstitial fluid, so there is a lag of 10 to 15 minutes compared with capillary blood glucose monitoring. Although CGMS measures the glucose frequently and continuously transmits the data to a receiver, FGMS measures the glucose every minute and stores the data for up to 8 hours. FGMS then picks up the data when a reader is passed over the sensor. CGMS may require calibrating, depending on the manufacturer. FGMSs are currently worn on the posterior aspect of the upper arm, and CGMSs usually on the abdomen (but there are other sites). CGMS and FGMS can provide falsely low readings, and the lag may mean that hypoglycemia is not detected. Thus, it is recommended that a confirmatory fingerprick test take place if hypoglycemia is reported or if there is discordance between the expected and reported glucose (e.g., with hypoglycemia symptoms). A confirmatory blood glucose is recommended before changing the insulin regimen. Several systems recommend a test before driving. CGMS and FGMS can provide alarms when the glucose is low or high, allowing rectifying action to be initiated, thereby avoiding extreme glycemic values.

CGMS and FGMS provide a great deal of data, the interpretation of which can be challenging. There may be an overall pattern to guide changes in insulin therapy, although confirmatory fingerprick testing may still be needed, particularly at night. The reports can provide "wear time," which should be greater than 70% to be considered reliable. This can be a problem during pregnancy, with the growing abdominal circumference. The glucose management index provides an estimated HbA1c, and the coefficient of variation is one measure of glycemic variability (should be <36%). Aggregate and daily patterns will show times of hypo- and hyperglycemia. The target time in range (TIR) in pregnancy[46] is greater than 70%, but at a lower target range than for nonpregnant adults at between 3.5 and 7.8 mmol/L (63–140 mg/dL), with less than 4% below 3.5 mmol/L (63 mg/dL) and less than 25% above 7.8 mmol/L (140 mg/dL). For women who are pregnant, TIR is insufficient alone, as the preprandial and postprandial glucose targets still need to be followed to minimize excess fetal fuel supply (e.g., a fasting glucose of 5.3 mmol/L [95 mg/dL] at

TABLE 34.5 Characteristics of Preferred Blood Glucose Meters During Pregnancy

- Lowest deviation possible (i.e., a total error <5% with an imprecision (coefficient of variation) <2%).
- Takes into account the potential influence of hematocrit changes on the meter values.
- Automatic measurement and correction for hematocrit changes.

TABLE 34.6 **Current Safety Status of the Different Types of Insulins During Pregnancy**

Insulin	Insulin type	FDA Pregnancy Category	RCT Findings Versus Human
Regular U-100	Short-acting	B	-
Regular U-500	Short-acting, concentrated	B	-
Lispro U-100 and U-200	Rapid-acting	B	Vs. regular—no difference
Aspart	Rapid-acting	B	Vs. regular—no difference
Fast-acting Aspart	Very fast-acting	B	As per Aspart
Glulisine	Rapid-acting	C	No RCT in pregnancy
NPH	Intermediate-acting	B	-
Detemir	Long-acting	B	RCT vs. NPH—no difference
Glargine U-100	Long-acting	C	No RCT in pregnancy
Glargine U-200	Long-acting	C	No RCT in pregnancy
Degludec U-100 and U-200	Long-acting	C	RCT underway
Inhaled human insulin	Inhaled	C	No data

FDA, US Food and Drug Administration; *NPH*, neutral protamine Hagedorn; *RCT*, randomized controlled trial.

Category A: Adequate and well-controlled studies have failed to demonstrate a risk to the fetus in the first trimester of pregnancy (and there is no evidence of risk in later trimesters).

Category B: Animal reproduction studies have failed to demonstrate a risk to the fetus, and there are no adequate and well-controlled studies in pregnant women.

Category C: Animal reproduction studies have shown an adverse effect on the fetus, and there are no adequate and well-controlled studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks.

24 to 28 weeks' gestation was associated with a 2-fold risk of adverse outcomes[10]). Data from CGMS/FGMS can also be used in an app to guide insulin bolus calculations, although a fingerprick test is recommended where there is discordance with expectations or, for example, to change dosage.

Although one Danish randomized controlled trial (RCT) did not show benefit from CGMS,[47] a multicenter RCT, CONCEPTT, showed reduced rates of LGA infants (odds ratio [OR] 0.51; 95% CI: 0.28–0.90), neonatal intensive care admissions lasting 24 hours or more (OR 0.48; 95% CI: 0.26–0.86), and neonatal hypoglycemia (OR 0.45; 95% CI: 0.22–0.89).[33] CONCEPTT also showed increased TIR. Pregnancy outcomes, while improved, were not normalized with CGMS.

Whereas HbA1c can provide an overall guide to glycemia, as in nonpregnant adults, increased erythrocyte turnover lowers the red cell glycation time, leading to a lower target (often taken as 6.0% (42 mmol/mol), but suggested to be 5.6% (37 mmol/mol) in at least one study[48]). HbA1c does not provide the granularity of glycemic variation required to manage diabetes in pregnancy, but a measurement in each trimester is useful for an overall longer-term benchmark for use in consultations.

Diet and Physical Activity

In general, women with T1DM follow the dietary, physical activity, and weight gain recommendations for women without diabetes, with a number of caveats relating to their personal glycemic response to food or to physical activity. Foods with a lower glycemic index (GI) are helpful.[49] In the first trimester, and sometimes into the early second trimester, additional snacks may be required to avoid hypoglycemia between meals, depending on the insulin regimen in place. Weight gain in the first trimester is to be avoided, though, as per the IOM guidelines. As insulin resistance increases, the carbohydrate content of meals may need to be limited (e.g., to 50 g) to avoid a postprandial spike. Mornings are often associated with particularly high postprandial spikes, and some services reduce the morning carbohydrate content to 20 to 30 g in some individuals. The postprandial spike can be mitigated to some extent by increasing the time between insulin administration and the meal, as well as by postprandial physical activity,[50] but caution is needed to avoid hypoglycemia during this time. It is thought to be important to maintain carbohydrate intake and minimize ketogenesis following the demonstration that maternal antenatal 3-hydroxybutyrate concentrations correlate negatively with cognitive function in the offspring.[51]

Hypoglycemia risk in relation to physical activity may be heightened during the first trimester but reduced during the third trimester, when insulin resistance is at its greatest. Reduction in postprandial hyperglycemia can be effected through exercise (e.g., walking after meals). Foot care and footwear remain particularly important for those women with neuropathy.

Insulin Management

Women with T1DM will enter pregnancy receiving either a multiple dose insulin (MDI) regimen, involving background and prandial insulin, or CSII. Human insulin is considered safe during pregnancy, but insulin analogs vary in the level of safety data available. Table 34.6 summarizes the current status of the use of the different types of insulins during pregnancy.

Although insulin management is covered in detail in Chapter 39, during pregnancy, insulin dosing is not only different for each individual, but also varies in the setting of changing physiological insulin needs (periods of insulin sensitivity and increasing insulin resistance), varying degrees of nausea (nil to hyperemesis gravidarum over different periods of time), different degrees of physical activity (generally decreasing in the third trimester), gestational weight gain, and individual responses and choices of food and physical activity. This is further affected by factors that influence insulin absorption outside of pregnancy, including lipohypertrophy, injection technique, injection site, and prandial insulin dose (particularly as the dose can, for example, triple by the third trimester).[52] One further factor of importance during pregnancy is the potential for worsening dawn phenomenon. Although the dawn phenomenon (a transient rise in blood glucose between 4:00 AM and 6:00 AM) is not important among pregnant women with normal glucose tolerance,[53] it creates additional challenges for many pregnant women with T1DM.[54] As an example, in one Danish study, although the carbohydrate-to-insulin ratio was reduced from median 12 (range 4–20) at 8 weeks' gestation to 3 (range 2–10) grams carbohydrate per unit insulin at 33 weeks' gestation, ratios only declined to 6 (range 2–13) at lunchtime and 4 (range 2–10) at dinner.[55] Over the same period, basal insulin increased by from 0.8 (range 0.5–2.2) to 1.2 (range 0.6–2.5) IU/h at 5:00 AM and from 1.0 (range 0.6–1.5) to 1.3 (range 0.2–2.3) IU/h at 5:00 PM.

Dealing with this variation and increased insulin resistance in the morning requires substantial work by the woman. CGMS and FGMS can assist with identifying the times when glucose is high and low. A consistently high morning glucose may reflect the dawn phenomenon, nocturnal hypoglycemia, and/or insufficient nocturnal insulin dosing. Where CGMS or FGMS are unavailable, it is usually wise for the woman to at least intermittently test at 3:00 AM before the usual dawn rise to assess the changing pattern. Dealing with the dawn phenomenon, if sufficient to necessitate treatment, requires a bolus without carbohydrate cover (sometimes a difficult ask of women with, for example, hypoglycemia fear). This involves identifying the time that the glucose rises and injecting quick-acting insulin 0 to 10 minutes beforehand. When women are using CSII, this can be preprogrammed with an increase in the hourly basal rate. If MDI is being used, then this requires waking up and injecting (e.g., at 4:00 AM) and hopefully returning to sleep.

Although CGMS has been shown to improve glycemia and improve pregnancy outcomes in one large study, findings from studies of CSII suggest little benefit during pregnancy. Theoretically, CSII should provide many benefits for self-management, including the Bolus Wizard to avoid guessing bolus doses, the incorporation of active insulin time to reduce the risk of stacking (where several insulin boluses are acting at the same time), and the use of temporary basal rates to adjust infusion rates with exercise or when it is unclear whether needs have changed or not (a major issue during pregnancy!). However, a systematic review of RCTs of CSII versus MDI performed up to 2016 showed no clear differences, with a suggestion that mean birth weight might be higher with CSII.[56] A systematic review and metaanalysis of cohort studies showed no difference in pregnancy outcomes and overall no difference in HbA1c.[57] However, technology continues to develop, and CSII continues to improve, with additional functionalities. CSII can be controlled through a handset or through a mobile phone. Tubing is either required separately or as part of a discardable device (a patch pump). The rapid changes in pump technology pose problems for trialists, as the technologies will often be out of date by the time a trial is completed and analyzed. Furthermore, trials involving CSII alone may already be redundant. CGMS can be linked with CSII to create a hybrid closed-loop system that allows real-time communication and the use of algorithms, with the glucose level and its trends automatically adjusting basal inulin delivery without user input.[58] Overnight glycemia can be stabilized, insulin delivery can be suspended to avoid impending hypoglycemia, and alarms allow user action at high or low glucose thresholds. Glucose-level predictive algorithms can allow user action well before hypoglycemia becomes severe. There remain a number of challenges with these systems, including dependency on the sensor, which can read low, and user exhaustion resulting in the user ignoring or switching off alarms. The tight glucose targets during pregnancy currently remain a challenge for these systems, with hyperglycemic rebound after basal insulin suspension and automated glycemic targets well above the pregnancy glucose targets. However, new models are emerging to address these issues.

Management of Type 2 Diabetes and Overt Diabetes in Pregnancy

Women with T2DM who become pregnant and those considered to have had undiagnosed (likely T2DM) diabetes at conception (overt diabetes in pregnancy) face overlapping but different challenges to those with T1DM. Women with overt diabetes in pregnancy are managed as if they have T2DM in pregnancy. Key actions are:

- Urgent commencement of intensive management
- Risk assessment including risk factor/complication review, (e.g., eyes, feet, renal, cardiovascular disease)
- Glycemia assessment including hypoglycemia risk (potentially suggestive of misclassified T1DM, monogenic or secondary diabetes)
- Medication history, particularly antidiabetes medications besides metformin, antihypertensive, and lipid treatment
- Counseling regarding the personal risks of pregnancy, including miscarriage
- If treated with metformin, discussion of safety in the first trimester, transplacental passage, antenatal balance between reduced cesarean section and weight gain, but increased risk of SGA and uncertain long-term risks
- Initiation of self–blood glucose monitoring if not yet in place (CGMS/FGMS if possible and affordable)
- Folic acid (5 mg) if not in place up to gestational week 12
- Aspirin 100 to150 mg from week 8 to 16
- Cessation of oral antidiabetes medications/GLP1 receptor agonist, commencement of insulin therapy (including education regarding driving), but continuation of metformin unless there is concern about it crossing the placenta
- Switching from antihypertensives to pregnancy-safe drugs (methyldopa, labetalol, nifedipine)
- Stopping lipid-lowering medications

The management of women with T2DM includes women commencing pregnancy well-managed with diet and physical activity alone and no comorbidities, all the way to women with severe insulin resistance, level 3 obesity, hypertension, dyslipidemia, and other comorbidities (e.g., microvascular disease). Beyond lifestyle, glucose monitoring, regular review, and counseling, treatment includes metformin and/or insulin therapy. Metformin safety during pregnancies among women with T2DM was demonstrated by the Metformin in Women with Type 2 Diabetes in Pregnancy RCT of adding metformin treatment up to 2 g/day (if tolerated) among insulin-treated pregnant women with T2DM.[59] Metformin (vs. placebo)–treated women achieved lower HbA1c at 34 weeks' gestation (5.9% vs. 6.1% [41.0 mmol/mol vs. 43.2 mmol/mol]) and mean glucose (109 mg/dL vs. 113 mg/dL [6.05 mmol/L vs. 6.27 mmol/L]), required less insulin (1.1 units/kg/day vs. 1.5 units/kg/day) and gained less weight (7.2 kg vs. 9.0 kg). Although most women birthed by cesarean section, the proportion was lower among metformin-treated women (53% vs. 63%). Babies were less fat, with fewer above the 97th centile (9% vs. 15%) and over 4000 g at birth (12% vs. 19%). However, the frequency of SGA babies was greater (13% vs. 7%), albeit with no excess neonatal intensive care unit admission, highlighting a potential future risk of metabolic syndrome components in the offspring.[31] As a result of this putative long-term risk, in association with metformin crossing the placenta and no long-term studies in T2DM, there remains caution over metformin use. However, among women with T2DM in pregnancy, these theoretical risks need to be weighed against the actual benefits of reduced cesarean section and less macrosomia.

Where metformin is not used, and glucose targets are exceeded, insulin management follows a similar pattern to that in T1DM, although generally a fixed dose insulin regimen rather than carbohydrate ratios is used. Women with T2DM usually have sufficient insulin secretory capacity to buffer some variations in glucose. Hypoglycemia is less of a risk and, when it occurs, is less likely to be severe. However, some women may have less insulin secretory reserve and may be more insulin-sensitive than others. It may be helpful to assess diabetes type among these women (e.g., T1DM), as the usual phenotype of T2DM is one of underlying increased insulin resistance accompanying the insulin insufficiency.

Insulin resistance can become an increasing problem, necessitating large amounts of insulin. Although metformin can help reduce insulin requirements in T2DM[59] (e.g., from 155.29 ± 134.01 units/day to 109.76±105.1 units/day), the insulin dose remains sizeable, and the broad variance reflects some women needing a substantially higher dose. CSII has been used to overcome insulin resistance among women requiring more than 100 units insulin/day. In one study, the median insulin dose was 246 (range 116–501) units/day or 2.68 units/kg booking (first antenatal) weight.[60] Glycemia improved among 79% of the women within 1 to 2 weeks. Another option is to use U-500 insulin for the meal boluses, which reduces the surface area and increases the speed of absorption.

Management of Diabetes Complications in Pregnancy

The White classification of diabetes in pregnancy,[61] based on diabetes duration and the presence of different diabetes complications, is no longer useful in guiding assessment of diabetes in pregnancy risk or management. The focus is now on assessing glycemia and blood pressure before and during pregnancy, alongside detailed assessment of acute (hypoglycemia, diabetic ketoacidosis) and chronic (diabetic nephropathy, retinopathy/maculopathy, ischemic heart disease, neuropathy including autonomic neuropathy, and diabetic foot disease) complications.

Nephropathy. Pregnancy is a time when glomerular filtration rate normally increases by 50%, and, among women with preexisting diabetes and renal disease, both the kidneys and the fetus can be compromised. Women with existing known diabetic nephropathy will usually have been treated before pregnancy with renin-angiotensin system (RAS) blockade agents (ACE inhibitors or angiotensin receptor blockers). There is some evidence that RAS blockade agents are teratogenic, with retrospective cohort studies reporting an increased risk of fetal defects in women on RAS blockade compared with the normotensive population (RR 3.7; 95% CI: 1.8–7.3[62] and OR 1.5; 95% CI: 1.1–2.6[63]). However, no difference in rates was shown compared with other hypertensive women (RR 0.8; 95% CI: 0.7–1.1[64] and OR 1.1; 95% CI: 0.6–1.9[63]). Hence currently, clinical judgment is required to determine whether the benefits of RAS blockade outweigh the possible harm in the first trimester. Otherwise, RAS blockade agents are contraindicated in the second and third trimester, as they increase the risk of fetal anuria, oligohydramnios, and renal tubule dysplasia. For those without known nephropathy, urinalysis is usually performed at the initial visit (and at each subsequent visit to screen for preeclampsia). Where proteinuria is detected (urine albumin-to-creatinine ratio >35 mg/mmol), a monthly urinary protein:creatinine ratio can be used to monitor worsening renal function. Women in the nephrotic range might benefit from thromboprophylaxis.

Pregnant women with diabetes, microalbuminuria, and normal renal function are unlikely to experience loss of kidney function.[65] Conversely, women with chronic kidney disease 3a (glomerular filtration rate <60 mL/min) and/or proteinuria great than or equal to 3 g/24 hours at the beginning of pregnancy are at approximately 40% risk of permanent kidney damage. Renal function can deteriorate significantly during pregnancy, and a nephrologist review is likely to be required among these women or if renal function deteriorates. Some women do require dialysis. Blood pressure management is important, but the optimal blood pressure is unclear, with some recommending 135/85 mm Hg.[66]

Pregnancy outcomes have improved but remain a concern, with fetal survival rates of 89% to 100%. There are increased rates of preeclampsia (15%–55%), cesarean section (70%–82%), intrauterine growth restriction (11%–21%), and premature delivery (16%–31% before gestational week 35).[65]

Retinopathy. Diabetic retinopathy (DR) has been reported to be present in 34% to 72% of pregnant women with T1DM and 14% of women with T2DM.[39] Treatment for proliferative DR and diabetic macular edema (DME) will preferably occur preconception, and glycemia should be optimized before and during pregnancy. Progression of DR during pregnancy occurs at about double the rate as in nonpregnant women.[66] Progression is more likely with longer diabetes duration, existing DR at the beginning of pregnancy, poor glycemic control, and elevated blood pressure. Patients with T2DM might have a lower risk of DR progression, but for both T1DM and T2DM, rapid improvement in glycemia is associated with accelerated development of DR. As a result of the increased risk, women with preexisting diabetes (previously diagnosed or otherwise) undergo retinal review in the first trimester, at the end of the second trimester, and, if considered at risk of progression, during the third trimester. Fluorescein does cross the placenta and is not recommended. If retinopathy is present, monitoring is likely to be more frequent. Proliferative DR is treated with panretinal photocoagulation. DME can be sight-threatening, and of course is not generally amenable to panretinal photocoagulation, but the new anti–vascular endothelial growth factor drugs remain contraindicated during pregnancy (and probably breastfeeding). Intravitreal triamcinolone and systemic steroids (e.g., dexamethasone) have been used previously with success and no side effects. DR is no longer considered a contraindication to vaginal birth, although there remains a theoretical risk if severe DR is present, and this should be discussed with the woman. Retinopathy can regress postpartum.

Coronary Artery Disease. Preexisting coronary artery disease remains rare among women with diabetes who become pregnant. Nevertheless, enquiry regarding cardiac disease risk factors and symptoms remains important both prepregnancy and when pregnant. In terms of treatment for acute events, pregnant women have usually been excluded from trials, and hence the safety of the current acute coronary artery disease management regimens during pregnancy among women with diabetes is unclear. In a 2019 review of 125 women with manifestations of coronary artery disease, 18% had diabetes (including GDM).[67] Maternal mortality was 2%, mean gestation at birth was 37 weeks, premature birth occurred in 32%, 50% had a cesarean delivery, and neonatal death occurred in 4%.

Neuropathy. Peripheral neuropathy and diabetic feet remain a risk during pregnancy and may be exacerbated by edema and the increasing weight. As a result, usual foot care recommendations should be followed. Of more concern is autonomic neuropathy including orthostatic hypotension, gastroparesis, and impaired hypoglycemia awareness. Orthostatic hypotension, aggravated by the pregnancy-related fall in systemic vascular resistance, can almost immobilize a small number of women, particularly in association with hyperemesis and dehydration. Hyperemesis can be worse with gastroparesis and, if severe, gastroparesis can be a relative contraindication to pregnancy. Use of prokinetic and other drugs needs to be balanced with their potential risks to the fetus if they cross the placenta.

Management of Gestational Diabetes Mellitus

The management of GDM is based upon the two major RCTs performed to date: (1) the Australian Carbohydrate Intolerance Study in Pregnant Women RCT, which showed a reduction in the primary perinatal composite outcome consisting of neonatal death, shoulder dystocia, bone fracture, and nerve palsy and (2) the Maternal-Fetal

Medicine Units Network RCT, which showed that GDM treatment was associated with a significantly decreased risk of LGA, shoulder dystocia, cesarean section, gestational hypertension, and preeclampsia.[68,69]

KEY POINTS

- The two major randomized controlled trials showing reductions in adverse pregnancy outcomes were the Australian Carbohydrate Intolerance Study in Pregnant Women study and the Maternal-Fetal Medicine Units Network study.

Once GDM has been diagnosed, its management involves counseling/education regarding the pregnancy risks, physical activity advice, medical nutrition therapy, weight management, self–blood glucose monitoring, and obstetric monitoring, and, if the glucose targets in Table 34.4 are exceeded, pharmacotherapy is commenced. Weight management follows IOM guidelines (Table 34.2). At least 30 minutes per day of moderate-intensity exercise is recommended,[70] with several goals beyond the holistic needs of pregnancy, including improving fasting and postprandial glucose and reducing the need for insulin or reducing the dose of insulin. Some studies have shown that physical activity intervention among women with GDM decreases gestational weight gain, and births are less likely to be complicated by macrosomia, cesarean section, preterm birth, and SGA. Recent studies show that sedentary behavior (any waking behavior characterized by an energy expenditure ≤1.5 metabolic equivalents while in a sitting or reclining posture) is associated with worse neonatal outcomes, and advice should therefore include approaches to increase movement overall.

RCTs to date have shown that dietary modification among women with GDM is associated with lower fasting (−4.07 mg/dL [95% CI: −7.58, −0.57]) (−0.2 mmol/L [95% CI: −0.4, −0.03]; *P* = .02) and postprandial (−7.78 mg/dL [95% CI: −12.27, −3.29] −0.4 mmol/l [95% CI: −0.7, −0.2]; *P* = .0007) glucose, as well as a 35% (95% CI: 12-53; *P* = .006) lower need for medication.[71] Dietary modification was also associated with less macrosomia (RR 0.49 [95% CI: 0.27−0.88]; *P* = .02). Different diets have been used, including the Mediterranean diet, low-GI diet, a diet with increased or decreased carbohydrate content, and diets with modifications of fat or protein quality or quantity. Energy restriction and low-carbohydrate diets associated with gestational weight loss are a concern in view of the association with lower fetal lean mass and, potentially, long-term metabolic disease.[31] A metaanalysis of the low-GI diet showed no benefits in terms of neonatal outcomes.[72]

Insulin therapy has traditionally been the main pharmacotherapy for treating GDM when glucose thresholds are exceeded. The usual approach is to use self–blood glucose monitoring to identify where the glycemic thresholds are exceeded and then to use quick-acting insulin analogs such as aspart or lispro (see Table 34.6), injected 15 to 20 minutes before a meal to reduce each postprandial glucose peak, and/or neutral protamine Hagedorn (NPH) insulin injected before bed to reduce the overnight/fasting glucose. Detemir in pregnancy is noninferior to NPH insulin and is particularly useful for women who are more susceptible to nocturnal hypoglycemia using NPH insulin. Reducing glucose among some women with GDM can be challenging, requiring large amounts of insulin. Strategies can include increasing the time between quick-acting insulin administration and meals, spreading the carbohydrate load (reducing carbohydrates in each meal, particularly breakfast), adding NPH in the morning, using concentrated insulins (e.g., U-500), and using CSII. CGMS/FGMS are also being used to assist with management.

Many women would prefer not to inject insulin, with its risks of hypoglycemia, weight gain, discomfort, and complexity. As a result, oral therapies have been trialed as more convenient and cheaper approaches to GDM management. The medications tested, metformin and glyburide (glibenclamide), can both lower blood glucose and allow insulin to be avoided in 50% or more women. Dosage is not as flexible, but for many women that dosage does not need to be adjusted to adapt to varying insulin secretory function. Beyond the side effects of these drugs described in Chapter 40, the major concern is that both cross the placenta and can potentially impact on the fetus. Glyburide is generally no longer recommended, owing to its association with higher fetal birth weight with more frequent macrosomia and neonatal hypoglycemia compared with insulin.[73] This may be the result of a direct effect on the fetal β cells inducing hyperinsulinemia. There are also no long-term follow-up studies of offspring with intrauterine exposure to glyburide. Compared with insulin use, metformin has been shown to result in less maternal gestational weight gain, lower postprandial glucose, less pregnancy-induced hypertension and less severe neonatal hypoglycemia.[73] Metformin is also preferred by women. However, a high proportion of metformin crosses the placenta, and follow-up studies of trial participants treated with metformin from the end of the second trimester have variably shown associations with adiposity and possibly hyperglycemia.[35] As a result, it is important for women to be given information regarding, and have explained, the uncertainties about any long-term consequences of metformin use, balancing this with the benefits shown in trials. If metformin is used, then it can be started with 500 mg once per day and increased by 500 mg every 3 days to achieve a dose of 2500 mg across the three usual meals. If gastrointestinal or other side effects occur, the prior dose without side effects can be used. About half of women treated with metformin will require supplemental insulin.

Prevention of Gestational Diabetes

A prevailing view has been that because T2DM can be prevented by lifestyle interventions and/or insulin sensitizer use, the incidence of GDM should also be able to be reduced. However, only metformin is currently acceptable for use during pregnancy (with the caveats discussed earlier) and has not been shown to be effective for preventing GDM. While outside of pregnancy weight loss is a key factor in preventing progression to T2DM, gestational weight loss is not recommended,[32] and even significant gestational weight gain limitation is not associated with reductions in hyperglycemia development.[74] Indeed, lifestyle interventions have largely been negative, with a Cochrane review yielding a RR of 0.85 (95% CI: 0.71–1.01) of GDM with nutritional and/or exercise interventions, associated with heterogeneity between studies and often insufficient descriptions of interventions.[75] Another metaanalysis of 29 lifestyle RCTs yielded a RR of 0.78 (95% CI: 0.64–0.96) with lifestyle interventions beginning before 15 weeks' gestation but a RR 0.97 (95% CI: 0.82–1.13) in those starting later.[76] Most studies did not exclude women with hyperglycemia at the beginning of pregnancy, among whom GDM is already present, also explaining the heterogeneity that has been shown.

KEY POINTS

- A metaanalysis of 29 lifestyle RCTs yielded a RR of 0.78 (95% CI: 0.64–0.96) with lifestyle interventions, but no significant GDM risk reduction, in studies beginning before 15 weeks' gestation and a RR 0.97 (95% CI: 0.82–1.13) in those starting later.

OBSTETRIC MANAGEMENT OF DIABETES IN PREGNANCY

Obstetric Surveillance

Preexisting Diabetes in Pregnancy. Ideally, the obstetric team should become involved in the management of women with preexisting diabetes before pregnancy, as part of the multidisciplinary preconception planning team. At the first antenatal visit, it is also important to reiterate the rationale for good diabetes management to optimize pregnancy outcomes. Between the obstetrician and the endocrinologist, complications of diabetes (such as nephropathy, cardiac disease, retinopathy) and associated conditions (such as hypertension, thyroid disease, and coeliac disease) should be assessed.

Ultrasound. As for pregnancies in general, a first-trimester ultrasound should be performed to assess the viability of the pregnancy and confirm the gestational age. For women with pregestational diabetes, in the second trimester a fetal anomaly scan should be undertaken at 18 to 20 weeks. Inclusion of a four-chamber cardiac view and the outflow tracts is important because of the increased risk of diabetes-related congenital heart disease.[77]

In the third trimester for both pregestational diabetes and GDM, the focus of the ultrasounds is on fetal growth and well-being. These are commonly performed for growth monitoring, although ultrasound estimation of LGA has not been shown to be more accurate than clinical estimation, and the sensitivity of either for predicting birth weight of 4 kg or more is only 50% to 70%. However sonographic estimation and monitoring of SGA is important, as this is superior to clinical estimation.

Fetal abdominal circumference is a better predictor of birth weight than standard equations for estimating fetal size based on fetal biometry. There is a more rapid increase in fetal abdominal circumference with higher levels of HbA1c. An increase in fetal ultrasound abdominal circumference can also be a useful guide for an insulin treatment strategy in the management of GDM. Asymmetrical growth retardation may be an indication of impaired placental perfusion, and ultrasound monitoring is particularly useful in this situation.

Other Obstetric Screening and Monitoring. Screening for aneuploidy should also proceed as per other pregnancies. First-trimester maternal serum pregnancy-associated plasma protein A levels, and second-trimester alpha fetoprotein and unconjugated estriol levels are decreased in diabetes, so adjustments to the results need to be made. Nuchal translucency scans are unaffected by diabetes.

Maternal assessment of fetal movements is the simplest form of monitoring for fetal well-being. For both pregestational diabetes and GDM, regular fetal monitoring by the nonstress test (cardiotocography) for fetal well-being is usually also recommended from 32 weeks' gestation.[77] A nonreactive nonstress test may indicate fetal compromise. For women with GDM who are medically treated, have poor glycemic control, or have other factors associated with adverse pregnancy outcome, fetal surveillance is also usually recommended. However, the evidence suggests that fetal monitoring is of little benefit for well-managed diet-controlled GDM, and there is no consensus regarding its use among these women.[78] The nonstress test may be combined with an ultrasound to form the biophysical profile, which includes an evaluation of fetal breathing, gross body movements, fetal tone, and amniotic fluid volume. There is a lack of RCT evidence that the biophysical profile improves pregnancy outcomes compared with the nonstress test alone for high-risk pregnancies.

Some units also perform serial uterine artery Doppler studies for women with pregestational diabetes. An abnormal uterine artery systolic to diastolic velocity ratio and the presence of a systolic notch have been shown to have moderate predictive value for preeclampsia and intrauterine growth retardation in high-risk pregnancies.[79]

Timing of Delivery. The timing of delivery is important with the risk of stillbirth weighed up against the risk of neonatal and infant morbidity and mortality. A study of 138,917 women with GDM found that induction of labor or elective cesarean section was associated with a higher risk of neonatal morbidity and mortality compared with expectant management at 36 and 37 weeks, but the risk decreased from 38 weeks.[80] For women with T1DM there was increased risk at 37 weeks, and for T2DM at 36 weeks. One cohort study of 4 million deliveries showed an increase in stillbirth with gestational age beyond 36 weeks, but the nadir for neonatal and infant death was at 39 to 40 weeks.[81] At 39 weeks the stillbirth rate was 5.7/10,000 pregnancies, with a neonatal death rate of 2.6/10,000 and an infant death rate of 8.7/10,000 live births. In general, induction of labor before 40 weeks' gestation has been associated with a lower risk of cesarean section compared with expectant management.[82]

However, in the GINEXMAL RCT, which compared induction of labor at 38 to 39 weeks to expectant management for women with low-risk GDM, no difference was found in the incidence of cesarean section, shoulder dystocia, or admission to neonatal intensive care.[83] In another RCT of women with uncomplicated insulin-requiring gestational or pregestational diabetes, expectant management did not reduce the incidence of cesarean section compared with induction of labor at 38 weeks,[84] but there was an increase in LGA infants and shoulder dystocia.

Most organizations, including the American College of Obstetricians and Gynecologists (ACOG), recommend delivery by 39 to 40 weeks for women with GDM, but not earlier if the pregnancy is uncomplicated.[85] For women with higher-risk GDM or preexisting diabetes, earlier delivery is often necessary to minimize risk to the fetus. The risk of adverse outcomes is affected by factors such as maternal age, maternal weight, estimated fetal weight, and comorbidities or complications, hence these need to be considered in the decision regarding the need for early delivery.

Antenatal Steroids for Preterm delivery

Antenatal steroids, betamethasone or dexamethasone in large doses, are administered to promote fetal lung maturity where premature delivery is threatened or necessary. This may result in women with diet-managed GDM requiring insulin, and women who already are on insulin will require an increase in dosage. Insulin requirements increase by 40% to 45%, peaking at days 2 to 3 after the initial dose of steroid. Even after 5 days, there can be a modest increase in insulin requirement. Local protocols should be in place to enable a rapid response to counter the hyperglycemia arising from antenatal steroids. After 37 weeks' gestation, antenatal steroids are not recommended for women with diabetes because of their association with an increased risk of neonatal intensive care unit attendance with neonatal hypoglycemia (with no excess risk of respiratory issues).[86]

Hypertensive Disorders of Pregnancy

Pregestational diabetes is associated with the development of gestational hypertension and preeclampsia. A recent metaanalysis that included 19 studies with women who had diabetes found that pregestational diabetes was associated with a RR of 3.7 (95% CI: 3.1–4.3) for the development of preeclampsia compared with to the general pregnant population.[87] A systematic review of studies with women who had T1DM determined the prevalence of preeclampsia to be 17%.[88] The presence of diabetic nephropathy, microalbuminuria, retinopathy, and preexisting hypertension were also associated with increased risk.

Some, but not all, studies have shown an association between HbA1c and the development of preeclampsia.

A metaanalysis of studies of prophylactic antiplatelet therapy determined that aspirin therapy reduced the likelihood of preeclampsia by 10%, with this being equally effective among women with diabetes.[89] In a recent RCT with 1776 high-risk women, aspirin 150 mg daily was commenced between 11 and 14 weeks' gestation among intervention subjects.[90] The incidence of preterm preeclampsia was 1.6% in the intervention group and 4.3% in the placebo group, with an OR of 0.38 (95% CI: 0.2–0.74; P = .004). Although neither this study nor any other significant study has specifically focused on women with diabetes, it is widely accepted that the data suggest that low-dose aspirin is just as effective for women with diabetes as for any other high-risk group for the prevention of preeclampsia. The US Preventative Services Task Force and the ACOG recommend low-dose aspirin to optimally be commenced between 12 and 16 weeks' gestation,[77] and the American Diabetes Association recommends that it be commenced by the end of the first trimester for women with pregestational diabetes.[8]

There is an increased risk of gestational hypertension and preeclampsia with GDM, and the risk increases with higher glucose levels, even when below overt diabetes range.[10] There are no studies that have examined the utility of aspirin prophylaxis for preeclampsia in GDM, and there are no recommendations for this. However, in both of the large RCTs of glucose management for GDM, glucose control has been shown to reduce the incidence of preeclampsia.[68,69]

ACE inhibitors and angiotensin 2 receptor blockers are contraindicated in pregnancy, although in some situations they may be used in the first trimester where the benefits outweigh the risks. Antihypertensive agents that are considered safest in pregnancy include labetalol, nifedipine, and methyldopa, although the latter is less favored by the ACOG, as it is less potent.[91] Other agents that have been used are oxyprenolol, hydralazine, prazosin, hydrochlorothiazide, and clonidine.

Decrease in Insulin Requirements

Some pregnant women with T1DM or T2DM experience large reductions in insulin requirement in late pregnancy.[78] This has now been associated with adverse pregnancy outcomes including preeclampsia, preterm delivery, and admission to neonatal intensive care.[78] A decrease in insulin requirements of 15% from the peak dose is also associated with derangement of placental angiogenic factors. A large decrease in insulin requirement is therefore an indicator of potential placental dysfunction and should prompt increased fetal monitoring and a discussion with the obstetrics team.

Peripartum Management and Birth

According to the Pederson hypothesis, maternal hyperglycemia results in fetal hyperinsulinism. At birth, the maternal supply of glucose is terminated, and the fetus becomes dependent on gluconeogenesis and glycogenolysis to meet its glucose requirements. Neonatal hypoglycemia is common, and the incidence is higher if the mother had preexisting diabetes, required insulin, or had higher HbA1c. It is less clear if tight intrapartum glucose control reduces the likelihood of neonatal hypoglycemia, with conflicting findings in the literature.[92] However, it is commonplace for mothers who required larger doses of antepartum insulin to receive an insulin glucose infusion for the delivery to facilitate tight glycemic control.

An insulin glucose infusion is important for women with T1DM to avoid ketoacidosis, although an alternative is for insulin pump therapy to be continued through labor. There is evidence of better glycemic control among women who continue insulin pump therapy through labor and delivery, and having staff who are familiar with their use available minimizes any concerns.[93] Women with T1DM who receive an insulin glucose infusion need to be transitioned back to subcutaneous insulin. Women with GDM can generally have their insulin ceased immediately upon delivery.

An insulin management plan for delivery and postpartum should be developed for each woman who requires insulin during pregnancy. For women who need to continue insulin postpartum (all women with T1DM, and some with T2DM), the requirement is dramatically less than antepartum, with a dose requirement in the order of 50% of the prepregnancy dosage for women with T1DM.[94] Breastfeeding should be promoted, but owing to its increased caloric demands, this may further lower insulin requirements. A proactive reduction in insulin dose is required to avoid hypoglycemia, particularly for women who breastfeed.

LONG-TERM OUTCOMES AND MANAGEMENT OF GESTATIONAL DIABETES

Long-Term Risks of Gestational Diabetes to the Mother

The development of GDM represents an unmasking of an underlying predisposition to diabetes and the metabolic syndrome in general. The majority of women with GDM return to normal glucose tolerance following pregnancy. However, there is a high risk of subsequent diabetes, and this is generally T2DM. The original study by O'Sullivan and Mahan remains the largest study, with the longest duration of follow-up of women with GDM, and a 50% long-term risk of diabetes was observed.[7] A metaanalysis of 20 observational studies found that the risk for T2DM among women who had GDM was almost 10 times that of women who did not have GDM during pregnancy.[95] It has been estimated that up to 31% of parous women who develop T2DM would have had a previous GDM pregnancy.[96]

More recently, the follow-up of 4697 women in the HAPO study has provided data regarding the development of glucose intolerance following GDM as determined by the current IADPSG criteria.[97] After a median follow-up of 11 years, 11% of GDM mothers had developed T2DM, compared with 2% of mothers who did not have GDM (5-fold risk in the adjusted model). For prediabetes it was 42% compared with 18% (3-fold risk in the adjusted model). However, for mothers meeting the more stringent Carpenter–Coustan criteria for the definition of GDM, the incidence of diabetes and prediabetes was considerably higher. As would be expected, this indicates that less stringent criteria for the diagnosis of GDM will define more women as having GDM, but the risk to these women for the development of future diabetes is lower.

There is also evidence that GDM may be a harbinger of T1DM, especially in countries with a higher prevalence of T1DM. In a Finnish 23-year follow-up study, 5.7% of women with GDM developed T1DM, and there were no cases among control women.[98] The onset is usually rapid, with all cases in this study being diagnosed within 7 years of pregnancy.

Women who have had GDM are also more prone to cardiovascular disease. A metaanalysis of nine observational studies found that women with GDM had twice the incident risk of cardiovascular events as women who did not have GDM in pregnancy.[99] Among women who had not yet developed T2D, the increase in risk was 56%. A metaanalysis of 93 studies found that, compared with women without previous GDM, those who had GDM had a 2 mm Hg higher systolic and diastolic blood pressure, were 1.5 kg/m more obese, had 0.3 mmol/L higher total cholesterol, 0.2 mmol/L higher low density lipoprotein cholesterol, 0.6 mmol/L higher triglycerides, and 0.3 mmol/L lower high density lipoprotein cholesterol.[100]

KEY POINTS

- Women with prior gestational diabetes mellitus have a 5- to 10-fold risk of developing type 2 diabetes mellitus over time and double the risk of developing cardiovascular disease.

Benefits of Breastfeeding for the Mother

It is strongly recommended that women who have had GDM or preexisting diabetes in pregnancy breastfeed their newborns. In addition to the health benefits of breastfeeding for the child, there is evidence that breastfeeding may have positive effects for maternal metabolism as well. Breastfeeding increases maternal energy expenditure by 15% to 25% per 24 hours. This is associated with lower maternal glucose, insulin, and lipid levels, as well as better weight control. In addition, it improves insulin sensitivity.

In the SWIFT Study, Latina women who had GDM were prospectively followed for 2 years with annual OGTTs.[101] There were graded associations between lactation intensity and duration with the development of T2DM. Women who exclusively breastfed or breastfed for more than 10 months had half the incidence of diabetes compared with women who exclusively formula-fed,[101]

Postpartum Follow-up of Mothers

Because GDM may reflect underlying preexisting diabetes, it is routinely recommended that women with GDM be tested by a 75-g OGTT in the first 3 months postpartum.[15] An OGTT is preferred over an HbA1c test for diabetes screening, as the HbA1c may be affected by red cell changes during pregnancy or blood loss at delivery.

The increased risk of T2DM and cardiovascular disease following a GDM pregnancy highlights the need for ongoing follow-up, screening, and cardiovascular risk factor management. Ongoing diabetes screening every 1 to 3 years is recommended, by HbA1c, fasting glucose, or 75-g OGTT.[8].The OGTT is more sensitive in the short-term detection of diabetes in this cohort of women, as GDM diagnosed by an increased glucose level in response to a glucose load is more likely to be associated with diabetes diagnosed in response to a glucose load.[102] However, because it is less convenient, this may be a barrier to testing.

When caring for women who have had GDM, particular attention should be paid to the long-term management of cardiovascular risk factors. This includes healthy eating, undertaking regular physical activity, healthy weight management, and lipid and blood pressure screening and management. In prospective observational studies, women who follow healthier eating patterns have a reduced risk of developing T2DM. Both intensive dietary and physical interventions, as well as metformin, as demonstrated by the Diabetes Prevention Program, reduce the progression from impaired glucose tolerance to T2DM among women with GDM . In the GEM Study, conducted among women who had recent GDM, also with lifestyle coaches and dietitians supporting a target of achieving the pregravid BMI if this was less than 25 kg/m^2 or losing 5% of body weight if the pregravid BMI was greater than or equal to 25 kg/m^2 and engaging in 150 minutes of moderate to intensive physical activity per week was effective in increasing physical activity and the achievement of weight goals after 6 months.[103] However, this study was unable to demonstrate a reduction of diabetes within this short time frame.

Postpartum Follow-Up of Offspring

Long-Term Risks of Gestational Diabetes and Preexisting Diabetes in Pregnancy to the Offspring

Risk of Diabetes and Glucose Intolerance to the Offspring. The concept that hyperglycemia *in utero* results in fetal programming that predisposes the offspring to future obesity and diabetes began with the landmark studies of Pettit et al., who followed the offspring of Pima Indian women who had an OGTT during pregnancy. The worse the maternal glucose status during pregnancy, from normal to impaired glucose tolerance to diabetes, the greater was the likelihood of the offspring developing diabetes by the age of 24 years.[104] A dose-response relationship was also demonstrated between the maternal 2-hour glucose levels on the OGTT in pregnancy, with the prevalence of the offspring having impaired glucose tolerance or diabetes during their own pregnancy. In further sibship studies, it was demonstrated that this was not caused by genetic predisposition, and is therefore more likely to be directly related to the level of maternal glycemia in pregnancy.

More recently, there are data from the HAPO study demonstrating a similar relationship for the offspring of women with untreated hyperglycemia in the GDM range.[105] At age 10 to 14 years, 4160 children from the original cohort were evaluated with an OGTT. There was a positive association between maternal glucose levels on the OGTT during pregnancy and child glucose levels and HbA1c.[106] The prevalence of impaired glucose tolerance among the offspring of mothers who met criteria for GDM (but who were not treated) was 11% compared with 5% for mothers who did not have GDM. For impaired fasting glucose by American Diabetes Association criteria, the prevalence was 9% compared with 7%, respectively.[105] There was an inverse association between maternal glucose levels and child insulin sensitivity and disposition index, as determined by the Matsuda index.[107]

Risk of Obesity to the Offspring

Older studies have demonstrated that the severity of maternal glucose intolerance in pregnancy is not just associated with offspring glucose intolerance, but also with obesity in the offspring. In the follow-up of offspring in the HAPO study, 40% were overweight or obese if the mother met criteria for GDM, compared with 29% where the mother did not have GDM.[97] However, in a fully adjusted model (including maternal weight), the difference was not significant. These findings highlight the difficulty of teasing out the effect of intrauterine glucose exposure from the effects of genetics and family environment. Although multiple studies have now demonstrated a relationship between maternal HIP and offspring glucose intolerance and obesity, there remains as yet no good evidence that treatment of maternal glycemia alters this risk.

Risk of Neurocognitive Disorders and Autism in the Offspring

The suggestion that there is a link between hyperglycemia in pregnancy and neurocognitive disorders arose from correlations between measures of neurobehavioral development and HbA1c during pregnancy. A recent metaanalysis found that offspring exposed to maternal preexisting diabetes scored a mean of 3 intelligence quotient points lower than normal control subjects.[108]

An association between maternal preexisting T2DM and early GDM with the development of autism spectrum disorder in the offspring has been observed, although the relationship with maternal T2DM lost significance after adjustment.[109] At this point, the data regarding neurocognition and autism need to be viewed with caution, given the difficulty in controlling for potential confounders and the lack of prospective data.

Potential Mechanisms for Long-term Effects of Preexisting Maternal Diabetes and Gestational Diabetes on the Offspring

Freinkel modified the Pederson hypothesis to arrive at the concept of "fuel-mediated teratogenesis."[1] He postulated that the excessive transplacental passage of glucose, amino acids, and lipids in diabetic pregnancy leads to fetal programming changes that may affect adult health. There is now some evidence that epigenetic changes contribute to the fetal programming effect.

Epigenetic changes are heritable molecular mechanisms that alter gene expression without changing DNA sequences. These can result from environmental factors, and the preconception period, fertilization, gestation, and early years of life are periods of higher susceptibility. DNA methylation, histone modifications, and noncoding RNAs are potential epigenetic mechanisms. Epigenetic changes, including those that develop *in utero*, can be passed to the next generation, resulting in a transgenerational effect.

In human studies, differential DNA methylation has been demonstrated between cord blood from GDM pregnancies and controls.[110] In particular, there are differentially methylated regions associated with diabetes, metabolic disease pathways, obesity, insulin secretion, and coronary artery disease. Differential DNA methylation in genes associated with schizophrenia and schizoaffective disorders has also been found in placental samples. In an epigenome-wide association study that examined cord blood, GDM was associated with lower methylation of genes associated with autism spectrum disorder and T1DM and T2DM.[111] The study of epigenetics is likely to reveal much more about the fetal programming effects of the intrauterine environment in years to come.

SUMMARY AND FUTURE DIRECTIONS

The management of HIP should commence before conception, whether through prepregnancy planning or pregnancy avoidance (preferably using LARC), among those with known preexisting diabetes, or through weight management and diabetes prevention among women at risk of diabetes. Antenatal management is complicated and challenging for many women and requires ongoing supervision and early counseling regarding the risks and need for frequent visits to review and, where necessary, adjust medication. Postpartum management means a major downward reduction in medications and promotion of breastfeeding and maintenance of weight management and physical activity. Women and their partners need to be aware of the potential long-term risks for the offspring and the need to adopt a family approach to reducing this risk through healthy lifestyles.

HIP is experiencing rapid change and future directions, including:

- Ongoing development of technology and testing, whether or not these benefit the mother and offspring. The required tight glycemic targets are not yet possible for many women.
- Identifying the optimal glycemic targets during each trimester of pregnancy requires the performance of trials that include the use of TIR alongside pre- and postprandial targets.
- Obtaining evidence for the optimal criteria for the diagnosis of GDM in the first trimester and before 20 weeks' gestation and whether treatment is beneficial or harmful.
- Obtaining evidence for the optimal criteria for the diagnosis of GDM at 24 to 28 weeks; these may change with dividing GDM into that present early in pregnancy (i.e., already present or prevalent GDM) and that which develops *de novo,* usually by 24 to 28 weeks' gestation (i.e., incident GDM).
- Further research into models of care to normalize as many pregnancies as possible, to ensure that pregnancy is as positive an experience for women with diabetes as for women without diabetes.
- Further research into the heterogeneity of GDM and the potential for personalized care (including cost-effective approaches to detecting monogenic diabetes).
- Further research into preventing GDM before pregnancy or in the first trimester and into optimizing lifestyle choices during pregnancy among women with diabetes.
- Further research into ways to prevent progression to T2DM and cardiovascular disease among women with prior GDM, including strategies for widespread implementation.
- Further research into the long-term harm to the offspring of exposure to antenatal excess fuel supply, as well as the long-term benefits and possible harms from treatment of HIP.

REFERENCES

1. Freinkel N. Banting Lecture 1980. Of pregnancy and progeny. *Diabetes*. 1980;9:1023–1035.
2. Barbour LA, Hernandez TL. Maternal contributors to fetal growth in obesity and gestational diabetes: spotlight on lipids. *Curr Diabetes Rep*. 2018;8:37.
3. World Health Organization. Diagnostic criteria and classification of hyperglycemia first detected in pregnancy: a World Health Organization Guideline. *Diabetes Res Clin Pract*. 2014;103:341–363.
4. Buchanan TA, Metzger B, Freinkel N, et al. Insulin sensitivity and B-cell responsiveness to glucose during late pregnancy in lean and moderately obese women with normal glucose tolerance or mild gestational diabetes. *Am J Obstet Gynecol*. 1990;162:1008–1014.
5. Catalano PM, Tyzbir ED, Wolfe RR, et al. Carbohydrate metabolism during pregnancy in control subjects and women with gestational diabetes. *Am J Physiol Endocrinol Metab*. 1993;244:E60–E67.
6. Desoye G, Nolan CJ. The fetal glucose steal: an underappreciated phenomenon in diabetic pregnancy. *Diabetologia*. 2016;59:1089–1094.
7. O'Sullivan JB, Mahan CM. Criteria for the oral glucose tolerance test in pregnancy. *Diabetes*. 1964;13:278–285.
8. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes – 2020. *Diabetes Care*. 2020;43(suppl 1):S14–S31.
9. International Association of Diabetes and Pregnancy Study Groups. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*. 2010;33:676–682.
10. The HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. 2008;358:1991–2002.
11. McIntyre HD, Sacks DA, Barbour LA, et al. Issues with the diagnosis and classification of hyperglycemia in early pregnancy. *Diabetes Care*. 2016;39:53–54.
12. Simmons D, Hague WM, Teede H, et al. Protocol of the Treatment of BOoking Gestational diabetes Mellitus Study (TOBOGM): the effect of treating gestational diabetes mellitus early in pregnancy: a randomised controlled trial. *Med J Aust*. 2018;209:405–406.
13. Lavery JA, Friedman AM, Keyes KM, et al. Gestational diabetes in the United States: temporal changes in prevalence rates between 1979 and 2010. *BJOG*. 2017;124:804–813.
14. Adam S, Ammori B, Soran H, et al. Pregnancy after bariatric surgery: screening for gestational diabetes. *Br Med J*. 2017;356:j533.
15. Chakera AJ, Spyer G, Vincent N, et al. The 0.1% of the population with glucokinase monogenic diabetes can be recognized by clinical characteristics in pregnancy. The Atlantic diabetes in pregnancy cohort. *Diabetes Care*. 2014;37:1230–1236.
16. Campos GM, Khoraki J, Browning MG, et al. Changes in utilization of bariatric surgery in the United States from 1993–2016. *Ann Surg*. 2020;271:201–209.
17. Kwong W, Tomlinson G, Feig DS. Maternal and neonatal outcomes after bariatric surgery; a systematic review and meta-analysis: do the benefits outweigh the risks? *Am J Obstet Gynecol*. 2018:573–580.
18. Akhter Z, Rankin J, Ceulemens D, et al. Pregnancy outcomes after bariatric surgery and adverse perinatal outcomes: a systematic review and meta-analysis. *PLoS Med*. 2019;16:e1002866.
19. Johansson K, Cnattingius S, Naslund I, et al. Outcomes of pregnancy after bariatric surgery. *N Engl J Med*. 2015;372:814–824.

20. Shawe J, Ceulemens D, Akhter Z, et al. Pregnancy after bariatric surgery: consensus recommendations for periconception, antenatal and postnatal care. *Obes Rev*. 2019;20:1507–1522.
21. Feichtinger M, Stopp T, Hoffman S, et al. Altered glucose profiles and risk for hypoglycemia during oral glucose tolerance testing in pregnancies after gastric bypass surgery. *Diabetologia*. 2017;60:153–157.
22. Rottenstreich A, Elazary R, Ezra Y, et al. Hypoglycemia during oral glucose tolerance test among post-bariatric surgery pregnancy patients: incidence and perinatal significance. *Surg Obes Relat Dis*. 2018;14:347–353.
23. Benhalima K, Minschart C, Ceulemens D, et al. Screening and management of gestational diabetes after bariatric surgery. *Nutrients*. 2018;10:1479.
24. Guerin A, Nisenbaum R, Ray JG. Use of maternal GHb concentration to estimate the risk of congenital anomalies in the offspring of women with prepregnancy diabetes. *Diabetes Care*. 2007;30:1920–1925.
25. Loeken MR. Mechanisms of congenital malformations in pregnancies with pre-existing diabetes. *Curr Diab Rep*. 2020;20:54.
26. Pedersen J. *Diabetes and Pregnancy: Blood Sugar of Newborn Infants (Ph.D. Thesis)*. Copenhagen: Danish Science Press; 1952.
27. Barrett HL, McIntyre HD, D'Emden M, et al. Home monitoring of fasting and postprandial triglycerides in late pregnancy: a pilot study. *Diabetes Care*. 2017;40:e1–e2.
28. Powe CE, Allard C, Battista MC, et al. Heterogeneous contribution of insulin sensitivity and secretion defects to gestational diabetes mellitus. *Diabetes Care*. 2016;39:1052–1055.
29. Immanuel J, Simmons D, Harreiter J, et al. The metabolic phenotypes of early gestational diabetes mellitus and their association with adverse pregnancy outcomes. *Diabet Med*. 2020:e14413.
30. Rasmussen KM, Yaktine AL, eds. *Weight Gain during Pregnancy: Reexamining the Guidelines. Institute of Medicine (US) and National Research Council (US) Committee to Reexamine IOM Pregnancy Weight Guideliens*. Washington (DC): National Academies Press (US); 2009.
31. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia*. 1992;35:595–601.
32. Catalano PM, Mele L, Landon MB, et al. Inadequate weight gain in overweight and obese pregnant women: what is the effect on fetal growth?. *Am J Obstet Gynecol*. 2014;211:137.e1. e7.
33. Feig DS, Donovan LE, Corcoy R, et al. Continuous glucose monitoring in pregnant women with type 1 diabetes (CONCEPTT): a multicentre international randomised controlled trial. *Lancet*. 2017;390:2347–2359. Erratum in: *Lancet*. 2017;390:2346.
34. Simmons D. Safety considerations with pharmacological treatment of gestational diabetes mellitus. *Drug Saf*. 2015;38:65–78.
35. Lindsay RS, Loeken MR. Metformin use in pregnancy: promises and uncertainties. *Diabetologia*. 2017;60:1612–1619.
36. Lewis G, Maxwell AP. Should women with diabetic nephropathy considering pregnancy continue ACE inhibitor or angiotensin II receptor blocker therapy until pregnancy is confirmed? *Diabetologia*. 2014;57:1082–1083.
37. Wald NJ, Law MR, Morris JK, et al. Quantifying the effect of folic acid. *Lancet*. 2001;358:2069–2073.
38. Valera-Gran D, Garcia de la Hera M, Navarrete-Munoz EM, et al. Folic acid supplements during pregnancy and child psychomotor development after the first year of life. *JAMA Pediatr*. 2014;168:e142611.
39. Amoaku WM, Ghanchi F, Bailey C, et al. Diabetic retinopathy and diabetic macular oedema pathways and management: UK Consensus Working Group. *Eye*. 2020;34:1–51.
40. Spotti D. Pregnancy in women with diabetic nephropathy. *J Nephrol*. 2019;32:379–388.
41. Wahabi HA, Fayed A, Esmeil S, et al. Systematic review and meta-analysis of the effectiveness of pre-pregnancy care for women with diabetes for improving maternal and perinatal outcomes. *PloS One*. 2020;15:e0237571.
42. Sina M, MacMillan F, Dune T, et al. Development of an integrated, district-wide approach to pre-pregnancy management for women with pre-existing diabetes in a multi-ethnic population. *BMC Pregnancy Childbirth*. 2018;18:402.
43. García-Patterson A, Gich I, Amini SB, et al. Insulin requirements throughout pregnancy in women with type 1 diabetes mellitus: three changes of direction. *Diabetologia*. 2010;53:446–451.
44. Law GR, Ellison GTH, Secher AL, et al. Analysis of continuous glucose monitoring in pregnant women with diabetes: distinct temporal patterns of glucose associated with large-for-gestational-age infants. *Diabetes Care*. 2015;38:1319–1325.
45. Immanuel J, Simmons D. A perspective on the accuracy of blood glucose meters during pregnancy. *Diabetes Care*. 2018;41:2053–2058.
46. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. 2019;1:1593–1603.
47. Secher AL, Ringholm L, Andersen HU, et al. The effect of real-time continuous glucose monitoring in pregnant women with diabetes: a randomized controlled trial. *Diabetes Care*. 2013;36:1877–1883.
48. Nielsen LR, Ekbom P, Damm P, et al. HbA1c levels are significantly lower in early and late pregnancy. *Diabetes Care*. 2004;27:1200–1201.
49. Zhang R, Han S, Chen GC, et al. Effects of low-glycemic-index diets in pregnancy on maternal and newborn outcomes in pregnant women: a meta-analysis of randomized controlled trials. *Eur J Nutr*. 2018;57:167–177.
50. Goudie RJB, Lunn D, Roman Hovorka R, et al. Pharmacokinetics of insulin aspart in pregnant women with type 1 diabetes: every day is different. *Diabetes Care*. 2014;37:e121–e122.
51. Rizzo T, Metzger BE, Burns WJ, et al. Correlations between antepartum maternal metabolism and intelligence of offspring. *N Engl J Med*. 1991;325:911–916.
52. Gradel AKJ, Porsgaard T, Lykkesfeldt J, et al. Factors affecting the absorption of subcutaneously administered insulin: effect on variability. *J Diabetes Res*. 2018;2018:1205121.
53. Mandujano A, Thomas A, Huston Presley L, et al. Does the dawn phenomenon have clinical relevance in normal pregnancy?. *Am J Obstet Gynecol*. 2013;209:116.e1-116.e5.
54. Roeder HA, Moore TR, Ramos GA. Insulin pump dosing across gestation in women with well-controlled type 1 diabetes mellitus. *Am J Obstet Gynecol*. 2012;207:324.e1-324.e5.
55. Mathiesen JM, Secher AL, et al. Changes in basal rates and bolus calculator settings in insulin pumps during pregnancy in women with type 1 diabetes. *J Matern Fetal Neonatal Med*. 2014;27:724–728.
56. Farrar D, Tuffnell DJ, West J, et al. Continuous subcutaneous insulin infusion versus multiple daily injections of insulin for pregnant women with diabetes. *Cochrane Database Syst Rev*. 2016;I6:CD005542.
57. Ranasinghe PD, Maruthur NM, Nicholson WK, et al. Comparative effectiveness of continuous subcutaneous insulin infusion using insulin analogs and multiple daily injections in pregnant women with diabetes mellitus: a systematic review and meta-analysis. *J Womens Health (Larchmt)*. 2015;24:237–249.
58. Nosova EV, O'Malley G, Dassau E, et al. Leveraging technology for the treatment of type 1 diabetes in pregnancy: a review of past, current, and future therapeutic tools. *J Diabetes*. 2020;12:714–732.
59. Feig DS, Donovan LE, Zinman B, et al. Metformin in women with type 2 diabetes in pregnancy (MiTy): a multicentre, international, randomised, placebo-controlled trial. *Lancet Diabetes Endocrinol*. 2020;8:834–844.
60. Simmons D, Thompson CF, Conroy C, et al. Use of insulin pumps in pregnancies complicated by type 2 diabetes and gestational diabetes in a multiethnic community. *Diabetes Care*. 2001;24:2078–2082.
61. White P. Pregnancy complicating diabetes. *Am J Med*. 1949;7:609–616.
62. Li DK, Yang C, Andrade S, et al. Maternal exposure to angiotensin converting enzyme inhibitors in the first trimester and risk of malformations in offspring: a retrospective cohort study. *BMJ*. 2011;343:d5931.
63. Bateman BT, Patorno E, Desai RJ, et al. Angiotensin-converting enzyme inhibitors and the risk of congenital malformations. *Obstet Gynecol*. 2017;129:174–184.
64. Spotti D. Pregnancy in women with diabetic nephropathy. *J Nephrol*. 2019;32:379–388.
65. Nielsen LR, Damm P, Mathiesen ER. Improved pregnancy outcome in type 1 diabetic women with microalbuminuria or diabetic nephropathy: effect of intensified antihypertensive therapy? *Diabetes Care*. 2019;32:38–44.

66. Kitzmiller JL, Block JM, Brown FM, et al. Managing preexisting diabetes for pregnancy: summary of evidence and consensus recommendations for care. *Diabetes Care*. 2008;31:1060–1079.
67. Lameijer H, Burchill LJ, Baris L, et al. Pregnancy in women with pre-existent ischemic heart disease: a systematic review with individualised patient data. *Heart*. 2019;105:873–880.
68. Crowther CA, Hiller JE, Moss JR, et al. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med*. 2005;352:2477–2486.
69. Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med*. 2009;361:1339–1348.
70. American Diabetes Association. Management of diabetes in pregnancy: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S183–S192.
71. Yamamoto JM, Kellett JE, Balsells M, et al. Gestational diabetes mellitus and diet: a systematic review and meta-analysis of randomized controlled trials examining the impact of modified dietary interventions on maternal glucose control and neonatal birth weight. *Diabetes Care*. 2018;41:1346–1361.
72. Han S, Middleton P, Shepherd E, et al. Different types of dietary advice for women with gestational diabetes mellitus. *Cochrane Database Syst Rev*. 2017;2:CD009275.
73. Balsells M, García-Patterson A, Solà I, et al. Glibenclamide, metformin, and insulin for the treatment of gestational diabetes: a systematic review and meta-analysis. *BMJ*. 2015;350:h102.
74. Simmons D, Devlieger R, van Assche A, et al. Association between gestational weight gain, gestational diabetes risk, and obstetric outcomes: a randomized controlled trial post hoc analysis. *Nutrients*. 2018;10:1568.
75. Shepherd E, Gomersall JC, Tieu J, et al. Combined diet and exercise interventions for preventing gestational diabetes mellitus. *Cochrane Database Syst Rev*. 2017;11:CD010443.
76. Song C, Li J, Leng J, et al. Lifestyle intervention can reduce the risk of gestational diabetes: a meta-analysis of randomized controlled trials. *Obes Rev*. 2016;17:960–969.
77. ACOG practice bulletin No 201: pregestational diabetes mellitus. *Obstet Gynecol*. 2018;132:e228–e248.
78. Padmanabhan S, Lee VW, Mclean M, et al. The association of falling insulin requirements with maternal biomarkers and placental dysfunction: a prospective study of women with pre-existing diabetes in pregnancy. *Diabetes Care*. 2017;40:1323–1330.
79. Chien PFW, Arnott N, Owen P, et al. How useful is uterine artery Doppler flow velocimetry in the prediction of pre-eclampsia, intrauterine growth retardation and perinatal death? An overview. *BJOG*. 2000;107:196–208.
80. Metcalfe A, Hutcheon JA, Sabr Y, et al. Timing of delivery in women with diabetes: a population-based study. *Acta Obstet Gynecol Scand*. 2020;99:341–349.
81. Rosenstein MG, Cheng YW, Snowden JM, et al. The risk of stillbirth and infant death stratified by gestational age in women with gestational diabetes. *Am J Obstet Gynecol*. 2012;206:309.e1-309.e7.
82. Melamed N, Ray JG, Geary M, et al. Induction of labor before 40 weeks is associated with lower rate of cesarean delivery in women with gestational diabetes mellitus. *Am J Obstet Gynecol*. 2016;214:364.e1. e8.
83. Alberico S, Erenbourg A, Hod M, et al. Immediate delivery or expectant management in gestational diabetes at term: the GINEMAXAL randomised controlled trial. *BJOG*. 2017;124:669–677.
84. Kjos SL, Henry OA, Montoro M, et al. Insulin requiring diabetes in pregnancy: a randomised trial of active induction of labor and expectant management. *Am J Obstet Gynecol*. 1993;169:611–615.
85. ACOG practice bulletin No 190: gestational diabetes mellitus. *Obstet Gynecol*. 2018;131:e49–e64.
86. Gupta K, Rajagopal R, King F, et al. Complications of antenatal corticosteroids in infants born by early term scheduled cesarean. *Diabetes Care*. 2020;43:906–908.
87. Bartsch E, Medcalf KE, Park AL, et al. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *Br Med J*. 2016;353:i1753.
88. Vestgaard M, Sommer MC, Ringholm M, et al. Prediction of pre-eclampsia in type 1 diabetes in early pregnancy by clinical predictors: a systematic review. *J Matern Fetal Neonatal Med*. 2018;31:1933–1939.
89. Askie LM, Duley L, Henderson-Smart DJ, et al. Anti-platelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet*. 2007;369:1791–1798.
90. Rolnik DL, Wright D, Poon LC, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. *N Engl J Med*. 2017;377:613–622.
91. ACOG Practice Bulletin 203. Chronic hypertension in pregnancy. *Obstet Gynecol*. 2019;133:e26–e50.
92. Yamamoto JM, Benham J, Mohammad K, et al. Intrapartum glycemic control and neonatal hypoglycemia in pregnancies complicated by diabetes: a systematic review. *Diabet Med*. 2018;35:173–183.
93. Drever E, Tomlinson G, Bai AD, et al. Insulin pump use compared with intravenous insulin during labour and delivery: the INSPIRED observational cohort study. *Diabet Med*. 2016;33:1253–1259.
94. Achong N, Duncan EL, McIntyre HD, et al. Peripartum management of glycemia in women with type 1 diabetes. *Diabetes*. 2014;37:364–371.
95. Vounzoulaki E, Khunti K, Abner SC, et al. Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. *BMJ*. 2020;369:m1361.
96. Cheung NW, Byth K. The population health significance of gestational diabetes. *Diabetes Care*. 2003;26:2005–2009.
97. Lowe WL, Scholtens DM, Lowe LP, et al. Association of gestational diabetes with maternal disorders of glucose metabolism and childhood adiposity. *J Am Med Assoc*. 2018;320:1005–1016.
98. Auvinen AM, Luiro K, Jokelainen J, et al. Type 1 and type 2 diabetes after gestational diabetes: a 23 year cohort study. *Diabetologia*. 2020;63:2123–2128.
99. Kramer CK, Campbell S, Retnakaran R. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. *Diabetologia*. 2019;62:905–914.
100. Pathirana MM, Lassi Z, Ali A, et al. Cardiovascular risk factors in women with previous gestational diabetes mellituys: a systematic review and meta-analysis. *Rev Endocr Metab Disord*. 2020. [Epub ahead of print.].
101. Gunderson EP, Hurston SR, Ning X, et al. Lactation and progression to type 2 diabetes mellitus after gestational diabetes mellitus: a prospective cohort study. *Ann Intern Med*. 2015;163:889–898.
102. Duke A, Yap C, Bradbury R, et al. The discordance between HbA1c and glucose tolerance testing for the postpartum exclusion of diabetes following gestational diabetes. *Diabetes Res Clin Pract*. 2015;108:72–77.
103. Ferrara A, Hedderson M, Brown SD, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes melliitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial. *Diabetes Care*. 2016;39:65–74.
104. Pettit DJ, Aleck KA, Baird R, et al. Congenital susceptibility to NIDDM. Role of intrauterine environment. *Diabetes*. 1988;37:622–628.
105. Lowe WL, Scholtens DM, Kuang A, et al. Hyperglycemia and adverse pregnancy outcome follow-up Study (HAPO FUS): maternal gestational diabetes mellitus and childhood glucose metabolism. *Diabetes Care*. 2019;42:372–380.
106. Scholtens DM, Kuang A, Lowe LP, et al. Hyperglycemia and adverse pregnancy outcome follow-up study (HAPO FUS): maternal glycemia and childhood glucose metabolism. *Diabetes Care*. 2019;42:381–392.
107. Yamamoto JM, Benham JL, Dewey D, et al. Neurocognitive and behavioural outcomes in offspring exposed to maternal pre-existing diabetes: a systematic review and meta-analysis. *Diabetologia*. 2019;62:1561–1574.
108. Xiang AH, Wang X, Martinez MP, et al. Association of maternal diabetes with autism in offspring. *J Am Med Assoc*. 2015;131:1425–1434.
109. Quilter CR, Cooper WN, Cliffe KM, et al. Impact on offspring methylation patterns of maternal gestational diabetes mellitus and intrauterine growth restrain suggest common genes and pathways linked to subsequent type 2 diabetes risk. *FASEB J*. 2014;28:4868–4879.
110. Howe CG, Cox B, Fore R, et al. Maternal gestational diabetes mellitus and newborn DNA methylation: findings from the pregnancy and childhood epigenetics consortium. *Diabetes Care*. 2020;43:98–105.
111. Immanuel J, Simmons D. Screening and treatment for early-onset gestational diabetes mellitus: a systematic review and meta-analysis. *Curr Diab Rep*. 2017;17:115.

35

Hyperglycemia Secondary to Non-Traditional Diabetic Conditions

Harold E. Lebovitz

OUTLINE

Hyperglycemia and Disorders of the Pancreas, 609
- Pancreatectomy, 609
- Acute Pancreatitis, 611
- Chronic Pancreatitis, 611
- Pancreatic Cancer, 612
- Autoimmune Pancreatitis, 613
- Hemochromatosis, 614
- Hemosiderosis, 615
- Cystic Fibrosis, 615

Hyperglycemia Associated With Endocrinopathies, 617
- Acromegaly, 617
- Growth Hormone Treatment, 618
- Glucocorticoid Effects on Glucose Metabolism and Corticosteroid-Mediated Diabetes Mellitus, 618
- Cushing Syndrome, 619
- Glucagonoma Syndrome, 619
- Somatostatinoma, 620
- Pheochromocytoma, 620
- Hyperglycemia in Critical Care Patients, 621

Drugs That Can Cause Hyperglycemia, 622
- Drugs That Directly Adversely Affect **β** Cell Function, 622
- Drugs That Inhibit Increases in **β** Cell Cytosolic Calcium, 622
- Drugs That Cause Potassium Depletion, 622
- Antiimmune and Cancer-Treatment Drugs, 622
- Drugs That Cause Insulin Resistance, 624
- Oral Contraceptives and Sex Hormones, 624
- Nicotinic Acid, 624

Lipid-Lowering Agents and New-Onset Diabetes Mellitus, 624
- Statins, 624
- Proprotein Convertase Subtilisin/Kexin Type 9 Inhibitors, 625
- Cholesterol Ester Transfer Protein Inhibitors, 625

Acquired Immunodeficiency Syndrome and Antiretroviral Drugs, 625

Atypical Antipsychotic Agents, 626

Conclusion, 627

Plasma glucose levels in humans are normally maintained between 70 and 180 mg/dL. Plasma glucose levels outside of this range lead to acute and chronic complications. Central nervous system function is exquisitely sensitive to blood glucose levels. Acute hypoglycemia impairs cognitive function and, if severe (<30 mg/dL) and prolonged (>6 hours), may lead to irreversible brain damage. Severe acute hyperglycemia causes hyperosmolar coma, which can cause irreparable brain damage. Chronic hyperglycemia above 180 to 200 mg/dL impairs physiologic processes such as protein synthesis and phagocytic function and can result in both microvascular and macrovascular diseases. Because maintaining normal glucose homeostasis is essential, many physiologic processes regulate plasma glucose levels. Insulin and glucagon secretion are the major regulators, whereas catecholamine and gastrointestinal incretin hormone secretions modify regulation in specific circumstances.

Insulin promotes hepatic glucose uptake and glycogenesis, stimulates muscle and adipose tissue glucose uptake and metabolism, and inhibits adipose tissue lipolysis and muscle proteolysis. Glucagon stimulates hepatic gluconeogenic precursor uptake and increases hepatic glycogenolysis, gluconeogenesis, and ketogenesis. Because insulin and glucagon have opposite effects, the ratio of the two hormones mitigates normal plasma glucose levels. A fall in plasma glucose suppresses insulin secretion and stimulates glucagon secretion. A rise in plasma glucose stimulates insulin secretion and suppresses glucagon secretion.

Insulin effects on cells are mediated by a series of intracellular events that occur following insulin binding to plasma membrane insulin receptors. These intracellular events that mediate insulin's actions may be modified by other intracellular events and thus can make the effect of a plasma insulin concentration greater or less. An increase in fatty acid products within a muscle or adipose tissue cell diminishes insulin's effects in increasing glucose transport into the cell. The cell's ability to respond to insulin is referred to as insulin sensitivity. Because insulin secretion is determined by the plasma glucose level, a decrease in insulin sensitivity normally causes a compensatory increase in insulin secretion, which restores plasma glucose homeostasis. Hyperglycemia occurs when either there is an absolute deficiency in insulin secretion or when β cell insulin secretion cannot increase sufficiently to overcome a decrease in insulin sensitivity (insulin resistance).

Except for maturity onset of diabetes in youths, most types of diabetes mellitus (DM) are heterogeneous disorders in which gene polymorphisms provide the predispositions, and environmental factors serve as the precipitating causes of hyperglycemia. These gene polymorphisms are multiple and can affect β cell function or insulin sensitivity. Many individuals with some genetic predispositions to develop hyperglycemia may not manifest impaired glucose tolerance (IGT) or fasting hyperglycemia throughout their lifetime unless they are exposed to some environmental factor that exaggerates the effects of these polymorphisms. If a pathologic condition develops in such individuals, or if they take a medication that disturbs their compensated state, hyperglycemia will develop. Thus, hyperglycemic states resulting from nondiabetic conditions or therapies can be subdivided into those that can cause hyperglycemia in any individual because they radically interfere with a major regulatory pathway (Table 35.1) and those that precipitate diabetes only in genetically predisposed individuals because they alter the compensated state (Table 35.2). Because of the high prevalence of a genetic predisposition to type 2 diabetes in various populations, it is not always possible to make the distinction with certainty.

KEY POINTS

- Primary regulators of glucose homeostasis are insulin and glucagon secretion. Hyperglycemia occurs when a pathologic process or drug causes a primary defect in pancreatic β cell insulin secretion or when the β cell cannot compensate with enough insulin secretion to overcome an induced insulin resistance.

HYPERGLYCEMIA AND DISORDERS OF THE PANCREAS

Hyperglycemia and DM caused by disorders of the pancreas are classified as type 3c diabetes. An analysis of 1868 patients with DM evaluated during a 2-year period at University Hospital in Giessen, Germany determined that 23% had type 1, 67% had type 2, and 9.2% had type 3c diabetes.[1] The specific diagnoses of the 172 patients classified as type 3c diabetes were: chronic pancreatitis (CP), 79%; pancreatic cancer 8%; post–pancreatic resection 2%; cystic fibrosis (CF) 4%; and hereditary hemochromatosis 7%.

The characteristics of the hyperglycemia of the various disorders of the pancreas will depend on whether the pathologic process is uniform through the pancreas or whether it is restricted to a lobe of the pancreas as in pancreatic cancer or surgical pancreatectomy.

TABLE 35.1 Conditions That Can Cause Hyperglycemia in the Absence of Known β Cell Genetic Abnormalities

- Diseases of the Pancreas
  - Pancreatectomy
  - Trauma
  - Acute pancreatitis
  - Chronic pancreatitis
  - Autoimmune pancreatitis
  - Fibrocalcific pancreatitis
  - Pancreatic carcinoma
  - Hemochromatosis
  - Hemosiderosis
  - Amyloidosis
  - Cystic fibrosis
- Overproduction of Other Islet Hormones
  - Glucagonoma
  - Somatostatinoma
- Drugs and Toxins
  - Pyriminil (Vacor)
  - Pentamidine
  - Interferon α
  - K_{ATP} channel openers
  - Phenytoin (Dilantin)
  - Immune checkpoint inhibitors
  - mTOR and phosphoinositide 3-Kinase inhibitors
  - L-Asparaginase

Pancreatectomy

The pancreas is formed by fusion of the ventral and dorsal pancreatic protrusions from primitive gut endoderm.[2] The islets of the anterior head of the pancreas and the tail arise from the ventral pancreatic protrusion and contain 80% β (insulin-producing) and 15% α (glucagon-producing) cells. The islets of the posterior protrusion (uncinate lobe) contains 80% γ (pancreatic polypeptide–producing) cells. The tail of the pancreas contains twice the islet cell mass as the head or body. Human pancreatic islets contain 30% α cells, 60% β cells, and 10% δ (somatostatin-producing) and γ cells.

DM occurring after surgical removal of the pancreas is insulin-dependent. It is characterized by insulin and glucagon deficiency.[3] The magnitude and characteristics of the hyperglycemia depend on the quantity of pancreas removed (Table 35.3). Total or near-total pancreatectomy results in severe hyperglycemia, decreased plasma insulin, virtually absent plasma glucagon, and elevated plasma levels of

TABLE 35.2 Conditions That Precipitate Hyperglycemia in Individuals With a Genetic Predisposition to Type 2 Diabetes

- Endocrinopathies
  - Acromegaly
  - Cushing syndrome
  - Pheochromocytoma
  - Hyperthyroidism
- Drugs
  - Drugs that interfere with insulin secretion
    - β-adrenergic antagonists
    - Diuretics
  - Drugs that impair insulin action
    - Glucocorticoids
    - Oral Contraceptives
    - Nicotinic Acid
    - Statins

TABLE 35.3 Estimated Frequency of Diabetes Reported in Pancreatic Diseases

95% pancreatectomy	100%
50% pancreatectomy	0% in 2-year follow-up 26% in 9.6- to 16-year follow-up
Pancreatitis	
Acute after 1 year	15%
Acute long-term	30%–40%
Chronic calcifying	
(increases with duration of disease)	40%–70%
Chronic noncalcifying	15%–30%
Cystic fibrosis	
Childhood	2%
Adolescence	20%
Adult	40%–50%
Pancreatic carcinoma	
At diagnosis	68%
Diabetes mellitus onset ≤2 years from diagnosis	55%
Hemochromatosis	
Diagnosed by clinical criteria	35.6%
Diagnosed by genetic testing	17.7%

gluconeogenic precursors (alanine, lactate, glycerol). When exocrine function is significantly reduced, the ensuing malabsorption results in diminished meal-mediated incretin release, which contributes further to the diminished insulin secretion.[4]

The effect of a 50% pancreatectomy was evaluated in 28 normal transplant donors 1 year after surgery.[5] The donors lost a mean of 3.4 kg of body weight, and their mean fasting plasma glucose (FPG) level had risen by 9 mg/dL (88 ± 7 to 97 ± 16 mg/dL). Similarly, their mean serum glucose concentration 2 hours after an oral glucose load was higher (117 ± 18 to 156 ± 53 mg/dL), and the area under the 5-hour plasma glucose curve after oral glucose was 19.5% higher. Both the mean fasting plasma insulin concentration and the area under the 5-hour plasma insulin curve were significantly lower than preoperatively (–14% and –31%, respectively). None of the donors had any evidence of deficient pancreatic exocrine function. Further analysis noted that 21 of the donors had no significant postoperative change in plasma glucose or insulin, whereas seven showed a marked increase in the entire 5-hour plasma glucose curve (either IGT or DM) with no concomitant increase in 5-hour plasma insulin curves. The seven donors in whom some degree of hypoinsulinemia and hyperglycemia had developed did not have fasting hyperglycemia 1 year postoperatively. Two of the seven were studied from 2 to 7 years after surgery and had not had a further increase in FPG. Evaluation of eight donor/recipient pairs evaluated 9 to 18 years after their original surgery indicated that residual pancreatic mass is a significant determinant of long-term glucose homeostasis.[6] However, other variables were implicated in the discordancy for the development of DM. Obesity seemed to be a major factor, in that all individuals (four donors and two recipients) who had developed DM were among the eight patients who were obese. The data were interpreted as “suggesting that obesity should be a contraindication to donation of pancreatic segments and that donors should assiduously avoid becoming obese.”

In a 3- to 10-year follow-up of 21 normal individuals who had donated 50% of their pancreas between 1997 and 2003, two donors were taking oral diabetic medications, two had impaired FPG (IFG), one had IGT, three had both IFG and IGT, and one met the criteria for the diagnosis of diabetes.[7] Six of the patients were lost to follow-up. Only six of 15 patients with follow-up data had normal glucose values. Despite the use of stringent criteria to exclude those at risk for developing abnormalities in glucose metabolism, 43% of the total population of healthy individuals who underwent a 50% pancreatectomy developed some abnormality of glucose metabolism.

Follow-up of 46 living donors who underwent segmental pancreas donation from 1994 through 2013 found that seven had developed DM requiring oral hypoglycemic agents after a mean of 9.2 ± 3 years, and five had developed insulin-dependent diabetes after a mean of 7 ± 5.4 years.[8] Risk factors for developing DM were initial FPG of 100 mg/dL or more, 2-hour plasma glucose during an oral glucose tolerance of 120 mg/dL or more, basal insulin of 9 μU/mL or more and an increase in body mass index (BMI) after surgery greater than 15%. Data from a USA Scientific Registry of Transplant Recipients (1987–2015) were matched with records from a nationwide pharmacy claims warehouse (2005–2015) for diabetes medications and supplies as a measure of postdonation DM.[9] Among 73 pancreas donors, 45 were identified in the pharmacy database. Over a mean follow-up of 16.3 years, 26.7% fulfilled the criteria for DM, compared with 5.9% of matched kidney donors. The prevalence of insulin use was 11.1% in pancreas donors versus 0% in kidney donors, and the prevalence of oral antidiabetic use 20.0% versus 5.9%.

Studies in patients with a 50% partial pancreatectomy because of CP ($n = 14$), pancreatic carcinoma ($n = 10$), or extrapancreatic tumors impinging on the pancreas ($n = 13$) showed improvement in glucose tolerance compared with the baseline abnormalities several weeks following surgery.[10] This was greatest in those with CP. Resection of the head of the pancreas improved glucose tolerance, whereas resection of the tail of the pancreas caused worsening of glucose tolerance. Insulin and C-peptide secretion were decreased equally by both surgical procedures. Repeat studies 3 years after surgery in both the CP and extrapancreatic tumor patients showed restoration of glucose tolerance to initial baseline values.[11] Postchallenge insulin and C-peptide concentrations that had reduced after surgery were partly normalized at the time of follow-up. These data suggest that some regeneration can occur in the adult pancreatic islets. Sun and colleagues studied the metabolic effects of removing 20% to 88% of the pancreas in dogs and found that no significant metabolic changes occurred until approximately 50% had been removed.[12] From the available data, it appears that metabolic abnormalities are clinically relevant after removal of 50% of the pancreas and increase progressively as the extent of pancreatectomy increases. The concomitant presence of insulin resistance further increases the likelihood of metabolic abnormalities.

Characteristics of DM following total or extensive pancreatectomy include an absence of glucagon secretion and marked impairment in insulin secretion. Absence of glucagon slows, but does not interfere with, the development of hyperglycemia and ketonemia after insulin withdrawal. Although not necessary for development of the metabolic abnormalities of insulin- dependent DM, the presence of glucagon increases the metabolic abnormalities of insulin deficiency, and the absence of glucagon secretion does leave a pancreatectomized individual with diabetes at high risk for severe hypoglycemia during insulin treatment.[13,14] This situation is exaggerated by the associated nutritional deficiencies and weight loss that ordinarily accompany exocrine pancreatic insufficiency. A concomitant deficiency of pancreatic polypeptide may contribute to persistent hyperglycemia caused by impaired hepatic insulin action.[15] Treatment of an individual with pancreatectomy-induced diabetes requires insulin, is associated with marked lability in glucose regulation, and is linked with an increased rate of both ketoacidosis and death from hypoglycemia. The development of autonomic neuropathy in such a patient greatly adds to the risk for severe hypoglycemia with insulin treatment.

Patients with total pancreatectomy can maintain reasonably good glycemic control with few serious hypoglycemic or ketoacidosis events provided they use self–blood glucose monitoring or continuous glucose monitoring (CGM) and are followed closely by their physicians.[16] Among 33 patients available for a median follow-up of 50 months, the median hemoglobin A1c (HbA1c) level was 8.2%, which was comparable to that in type 1 diabetic patients followed at the same institution. Patients resected for cancer maintained better glycemic control than those treated for CP. The median insulin dose for the entire group was 46 units per day.

A feasible method for minimizing the development or severity of DM following major or total resection of the pancreas is isolation and autotransplantation of the individual’s islets. Among 20 of 37 nondiabetic patients who had undergone islet autotransplantation after a 50% to 60% distal partial pancreatectomy for benign pancreatic tumors, the 7-year diabetes-free survival rate was 51%.[17] The response was a function of the islet yield per gram of pancreas weight and was greatest when there were more than 5154 islet equivalents per gram (IEQ/g). The proportion of patients with IGT 2 years after distal pancreatectomy was 12/16 in the control group, 6/7 in patients with less than 5154 IEQ/g and 3/11 in those with greater than 5154 IEQ/g.

KEY POINTS

- Resection of 50% of the pancreas will result in a progressive deterioration of glucose metabolism that is minimal after 2 years but results in frank diabetes mellitus in approximately 26% of pancreatic donors after 10 to 16 years. Those who are most likely to develop glucose intolerance are those who gain excessive weight and those with the initial FPG greater than 100 mg/dL. Resection of the head of the pancreas is associated with preserved or improved glucose metabolism, whereas resection of the tail of the pancreas results in deterioration of glucose metabolism. Total pancreatectomy is characterized by severe insulin, glucagon, and exocrine pancreatic deficiency. Patients with total pancreatectomy are insulin-dependent, require pancreatic enzymes, and are at risk for hypoglycemia or ketoacidosis. Autologous pancreatic islet transplantation is a good therapeutic option if possible.

Acute Pancreatitis

Acute pancreatitis (AP) is an inflammatory disease of the pancreas and is the leading cause of hospitalization for gastrointestinal disorders in the United States. Risk factors for AP include gallstones, blockage or manipulation of the pancreatic ducts, or alcohol misuse. Hyperglycemia is a common early feature of AP and was generally considered to be a transient phenomenon that completely resolved in almost all patients.[18,19] However, a metaanalysis of 24 prospective clinical studies involving 1102 patients with a first episode of AP, published in 2014, showed that prediabetes and/or DM occurred in 37% of individuals after AP.[20] The pooled prevalences of prediabetes, DM, and insulin-treated DM after AP were 16%, 23%, and 15%, respectively. Newly diagnosed DM developed in 15% of individuals within 12 months after the first episode of AP, and the risk increased to nearly 40% by 5 years (relative risk 2.7). A similar trend was observed with insulin treatment. The severity of AP, its etiology, age, and gender had minimal effect on the glycemic outcomes.

Subsequently, individual centers have reported mean follow-ups of 2.7 and 3.5 years in 113 and 100 patients with AP and found the prevalence of new-onset DM to be 14% and 30% and the prevalence of exocrine pancreatic insufficiency to be 21% and 29%, respectively.[21,22] A metaanalysis of 234 patients with AP followed for 12 to 179 months reported a prevalence of prediabetes and new-onset DM of 43% and prevalence of pancreatic exocrine insufficiency of 29%. Pancreatic exocrine insufficiency occurred in 40% of those who developed prediabetes or DM.[23]

A retrospective analyses of case records from 2.36 million patients from primary care practices in England reported in 2017 that diabetes occurring after a median follow-up of 4.5 years from the onset of pancreatic disease was 2.59/100,000 person-years.[24] This was greater than the onset of new type 1 diabetes. Diabetes following AP was more common than that following CP (361 vs. 91 patients). The majority of the 559 cases of pancreatic diabetes had been diagnosed as type 2 diabetes (87.8%). Patients with pancreatic diabetes at diagnosis had a median age of 59 years and a median BMI of 29.2 kg/m^2. Patients with pancreatic diabetes had poorer glycemic control than type 2 diabetic patients, and insulin use within 5 years was 20.9% in diabetes following AP and 45.8% in patients with CP, as compared with 4.1% in patients with type 2 diabetes. A population-based study of 3 million residents of New Zealand over a 10-year period reported a prevalence of diabetes associated with exocrine pancreas disease to be 1.13/1000 individuals, with AP contributing 61% to the overall prevalence.[25]

The high prevalence of insulin-dependent DM following a single bout of AP, regardless of severity, suggests that a significant percentage may be autoimmune in origin. Several studies have reported the presence of glutamic acid decarboxylase or other antiislet antibodies in some patients with AP. An increase in serum GIP and peptide YY that correlated with increases in interleukin-6, tumor necrosis factor α, and monocyte chemoattractant peptide was reported in 83 individuals following an episode of AP.[26] A hypothesis to be tested is that destruction of exocrine pancreatic tissue stimulates an intensive inflammatory response that triggers an islet cell immune response in genetically susceptible individuals.

Chronic Pancreatitis

CP is characterized by fibrosis and inflammation of the pancreas. In the United States the prevalence of CP is between 42 and 73/100,000 adults.[27] Specific pathologic and clinical findings are pancreatic atrophy, fibrosis, ductal strictures and distortion, calcifications, exocrine and endocrine deficiency, and chronic pain.[27-31] Excessive alcohol consumption is the most common etiology for CP (relative risk in US and European populations 2.7–3.3).[27,28] Idiopathic CP is the second most common etiology, accounting for 28% to 80% of cases, depending on the population assessed.[27,31] A number of susceptibility genes for CP have been recognized. These include variants of genes involved in trypsinogen activation (cationic trypsinogen 1, anionic trypsinogen, serine protease inhibitor Kazal type 1, and chymotrypsin C) and genes involved in intrapancreatic trypsin regulation (cystic fibrosis transmembrane conductance regulator [*CFTR*] and calcium-sensing receptor).[31] These genes not only are associated with idiopathic CP but also increase the susceptibility to other risk factors for CP. These genetic variants occur in 47.8% of idiopathic CP patients in a US population and 57.1% in a Chinese population (compared with 5.9% of control patients).[31] Smoking is associated with CP in a dose-dependent manner.[10,11] Hypertriglyceridemia, hypercalcemia, and autoimmune disease are less common risk factors for CP.[27-30] The underlying pathologic process in CP is destruction of acinar cells with release of activated enzymes, leading to necrosis and inflammation.

CP has been thought to account for less than 1% of DM in Western countries and Japan. Several recent publications dispute this and have reported a prevalence of 7.2% to 8%.[29,32] The presumed reasons for failure to recognize the higher prevalence of pancreatic diabetes (type 3c diabetes) have been misdiagnosis and lack of clear diagnostic criteria. In tropical countries, fibrocalculous pancreatic diabetes occurs with a self-reported prevalence of 0.36%.[33] The incidence may be somewhat higher, but comprehensive data are not available. Diabetes resulting from CP can be differentiated from type 1 and type 2 diabetes by an absent pancreatic polypeptide response to a mixed meal, insulin-induced hypoglycemia, or secretin infusion.[34] The development of DM in patients with CP is most frequent in those with calcific disease (55%–70%) and occurs less often in those with noncalcific disease (30%).[35] The prevalence of diabetes in patients with CP increases with increasing duration of pancreatitis and with increasing exocrine deficiency.

The inflammatory response causes loss of exocrine tissue, extensive fibrosis, and distorted and blocked ducts. The islets of Langerhans are relatively resistant and undergo pathologic changes only late in the disease. CP is associated with decrease in pancreatic mass, a loss of functioning β cells of approximately 29%, and a lesser loss of α cells.[36,37] Hormonal alterations include a decrease in insulin secretion in response to nutrients, followed later by a decrease in fasting C-peptide levels. Plasma C-peptide levels rather than insulin levels are a better means of assessment of insulin secretion in CP, because associated liver disease may change hepatic extraction rates of insulin. With progressive CP, insulin secretion falls even lower. Glucagon secretion is impaired in moderate to severe CP.[37-39] Insulin resistance frequently develops in such individuals.

DM is seen after several years of CP. In an unselected series of patients with CP, 35% of cases resembled type 1 diabetes, 31% resembled type 2 diabetes or IGT, and 34% exhibited normal glucose tolerance.[39] In a series of 2011 patients with CP followed for a mean of 22 years, the duration of the CP determined the prevalence of DM: the cumulative rate at 29 years was 45.8%, and at 50 years was 90%.[40] Mild pancreatitis may be associated only with IGT, whereas severe pancreatitis is associated with insulin dependency. Patients with CP and DM fail to secrete glucagon in response to hypoglycemia. Incretin secretion is frequently impaired. If patients have concomitant autonomic neuropathy, they are extremely susceptible to severe and prolonged hypoglycemia.

Pancreatic exocrine function has been reported to be reduced in some type 1 and type 2 diabetic patients. Although this has been explained as a complication of diabetes, the relationship between exocrine and endocrine function of the pancreas makes it difficult to distinguish patients with DM who have developed secondary exocrine deficiency from patients with primary pancreatic exocrine disease who have developed islet cell deficiency, and secondarily DM. A retrospective analysis of pancreatograms of patients with known diabetes has suggested that CP may be much more common as a cause of diabetes than was previously thought.[41] A total of 38 type 1 and 118 type 2 diabetic patients underwent endoscopic retrograde cholangiopancreatography studies for various reasons. Pancreatic ducts were classified as normal in 23.3% and as exhibiting CP degree I, II, and III in 22.7%, 32.7%, and 21.3%, respectively. The investigators suggested that a substantial number of patients with primary DM may have a concomitant CP, or perhaps many cases of primary DM may be related to CP.

Treatment of DM in patients with CP should entail the use of small doses of short-acting or rapid-acting insulins to manage the hyperglycemia, replacement enzymes for the malnutrition and malabsorption, and elimination of the use of alcoholic beverages.[27,29,31,38,39] Surgery with subtotal resection or near-total resection may be necessary to relieve severe pain. Following pancreatic surgery for pain or other complications, there is a significant incidence of DM; however, whether diabetes develops is dependent on the type of surgical intervention used. Whipple procedures usually are associated with an incidence of 25% to 40%, a Frey or Berne procedure 8% to 22%, and a distal pancreatectomy approximately 60%.[42,43] In such patients, successful islet allotransplants and autotransplants in the liver have been able to maintain near-normoglycemia. The hepatic islet cell transplants secrete insulin in response to nutrients but are unable to secrete glucagon in response to hypoglycemia. A recent report of the results of pancreas allotransplants in patients who had undergone total pancreatectomy in the past showed a 70% survival rate and successful correction of both exocrine and endocrine deficiencies.[44]

Sutherland and coworkers have summarized their experience with 409 patients with CP and severe, unrelenting pain who underwent a total pancreatectomy with islet autotransplantation.[45] The etiology of the CP in most of the patients was idiopathic, familial, or pancreas divisum. The 5-year survival rate was 90%. The status of glucose control following the surgery was directly proportional to the IEQ/kg transplanted.[45] The patients were classified as receiving less than 2500 IEQ/kg, 2500 to 5000 IEQ/kg, or greater than 5000 IEQ/kg. Glucose control status was defined as insulin-independent, partial function, or insulin-dependent. At 3 years after surgery, 72% of patients who received greater than 5000 IEQ/kg were insulin-independent, whereas 24% had partial function, and 4% were insulin-dependent. In contrast, 12% of those receiving less than 2500 IEQ/kg were insulin-independent, 33% had partial function, and 55% were insulin-dependent. Some 22% of those receiving 2500 to 5000 IEQ/kg had insulin independence, whereas16% had insulin dependence. The percentage of patients with mean HbA1c less than 7% at 3 years was 71% for those who received less than 2500 IEQ/kg 86% for those who received 2500 to 5000 IEQ/kg, and 94% for those who received greater than 5000 IEQ/kg.

KEY POINTS

- Acute pancreatitis is the commonest gastrointestinal disorder requiring hospitalization. Etiology is most often caused by gallstones, obstruction of the pancreatic duct, or excess alcohol ingestion. Acute pancreatitis leads to diabetes mellitus (DM) in 40% of patients within 5 years. A significant proportion of the diabetes is insulin-requiring and is thought to represent the result of autoimmune activation secondary to pancreatic inflammation. Chronic pancreatitis is a recurring inflammatory disease of the exocrine pancreas that causes progressively more severe abdominal pain and exocrine deficiency, followed by islet destruction, and secondarily DM. The commonest etiologies are chronic alcohol use or idiopathic. Variants of several genes regulating pancreatic exocrine secretions have been associated with the development of chronic pancreatitis. The prevalence of DM increases with duration of chronic pancreatitis and is close to 70% after 20 years. DM owing to pancreatitis is often confused with type 2 diabetes and may account for as much as 5% to 8% of total DM cases. Severe recurrent abdominal pain in patients with chronic pancreatitis may require total pancreatectomy, in which case autologous islet transplantation is recommended to control glucose metabolism.

Pancreatic Cancer

Pancreatic ductal adenocarcinoma is the fourth leading cause of cancer deaths in the United States. Because most patients have extensive disease spread at the time of diagnosis, the reported 5-year survival ranges from 1% to 8%.[46,47] DM is greatly increased in patients with invasive ductal pancreatic cancer. The relationship between DM and pancreatic cancer is bidirectional, in that longstanding DM is associated with an increased risk for the development of pancreatic cancer, and pancreatic cancer is associated with an increased risk for the development of new-onset DM (occurring ≤2–3 years before the diagnosis of pancreatic cancer).[48-51] In a comparative study, DM occurred in 68% of patients with pancreatic cancer compared with 19.8% of age-matched patients with lung cancer, 19.4% with breast cancer, 14.8% with prostate cancer, 20.7% with colon cancer, and 23.5% of noncancer controls.[52] Longstanding DM has a relative risk for the development of pancreatic cancer of approximately 1.6 compared with nondiabetic individuals, which is similar to the risk for several other cancers. This risk is increased by poor glycemic control. Pancreatic cancer is associated with a 5- to 8-fold increased risk for newly diagnosed DM. This increased risk can be identified by increasing FPG, worsening oral glucose tolerance, or the diagnosis of DM starting at 36 months and becoming progressively higher up to the time that pancreatic cancer is diagnosed.[53] These data suggest that pancreatic cancer itself is the cause of the new-onset DM. The observations that longstanding DM is unchanged or slightly worse after surgical removal of the tumor, whereas new-onset diabetes is markedly diminished after tumor resection, provide substantial proof for this hypothesis. Arising fasting plasma glucose (FPG) level, a deteriorating oral glucose tolerance test (OGTT), or the development of DM in a previously normal glucose-tolerant individual can be an early indication of pancreatic cancer. An additional factor that heightens the suspicion for pancreatic carcinoma is the progressive weight loss accompanying new-onset deterioration of glucose control. The development of ordinary type 2 diabetes is generally accompanied by weight gain, with weight loss ensuing only after the development of severe hyperglycemia. In a study from the Mayo Clinic, it was observed that the increase in FPG at the time of diagnosis, as well as up to 24 months before diagnosis of pancreatic cancer, was positively correlated with the volume of the resected tumor from greater than 1.0 mL to greater than 16.0 mL.[3] Wakasugi and colleagues reported that 53.1% of patients with invasive ductal pancreatic carcinoma had DM, and in 45.9% it was thought to be secondary to the carcinoma.[54] Hyperglycemia or DM is present in approximately 85% and 50%, respectively, of patients with pancreatic carcinoma at the time

TABLE 35.4 **Interval Between Diagnosis of Diabetes Mellitus and Diagnosis of Pancreatic Cancer or Date of Examination of Control Population**

Interval Between Diagnosis of Diabetes Mellitus and Pancreatic Cancer	Pancreatic Patients (*n* = 164)	Control Patients (*n* = 60)	P-Value
0	66 (40.2%)	2 (3.3%)	<0.001
1	15 (9.2%)	2 (3.3%)	0.17
1–14	81 (49.4%)	35 (58.3%)	NS
≥15	17 (10.4%)	23 (38.3%)	<0.001

NS, Not significant.

(From Gullo L, Pezzilli R, Morselli-Labate AM, et al. Diabetes and the risk of pancreatic cancer. *N Engl J Med.* 1994;331:81–84.)

of diagnosis. The reasons for this association have been the subject of much speculation. In another study, a diagnosis of diabetes was made in 22.8% of the patients with pancreatic cancer and 8.3% of the control population. In 40.2% of patients with pancreatic cancer the diagnosis of diabetes was made concurrently with the diagnosis of the cancer, and in 15.9% within 2 years before the diagnosis[55] (Table 35.4). In the control population, only 13.3% of the diabetes was of 2 years' duration or less. The greatest percentage of diabetes in the control population was of 10 years' duration or greater (53.3%) (Table 35.4). These data further support the idea that DM predisposes to the development of pancreatic cancer in a small number of patients, but that pancreatic cancer causes the development of DM in most patients. Possible mechanisms by which pancreatic cancer could contribute to the development of type 2 diabetes include: (1) destruction of islets, (2) impairment of the insulin secretory mechanism, (3) development of insulin resistance, and (4) tumor-related pancreatitis. Insulin and C-peptide measurements during OGTTs in patients with pancreatic cancer have shown abnormal β cell function with reduced plasma C-peptide responses in 50% of patients, as well as increased plasma proinsulin–to–C-peptide ratios.[56,57] Insulin resistance has been demonstrated in most patients with pancreatic carcinoma.[58-60] Morphometric studies of tumor-free regions of the pancreas have shown reduced β cell populations.[61] An inverse correlation was noted between the number of β cells and the FPG concentration. These data indicate that pancreatic cancers produce substances or responses that destroy normal β cells.[59-62] The presence of diabetes in patients with pancreatic cancer predicts that the tumor is less likely to be resectable,[63] and the patient has a poorer prognosis for disease-free and overall survival than if diabetes is not present.[63,64] Patients with pancreatic cancer typically have type 2 diabetes, and most have been treated with oral antihyperglycemic agents.[55,64] Reasons for the poorer survival of patients with diabetes may relate to the mitogenic activity of the higher insulin levels, as well as increases in insulin and insulin-like growth factor-1 receptors in the tumors.

Several intriguing observations that were reported recently make this relationship between pancreatic cancer and DM even more complex and confusing. Li and associates reported that metformin treatment of diabetic patients was associated with a decreased risk for developing pancreatic carcinoma, whereas insulin or insulin secretagogue treatment was associated with an increased risk.[65] Metaanalyses of survival outcomes in metformin-treated patients with pancreatic cancer present conflicting data.[66] One metaanalysis of two studies involving 181 patients showed no benefit of metformin treatment. In contrast, a metaanalysis of eight cohorts involving 2805 patients showed an improvement in overall survival at 6 months.[66] In a large multicenter clinical trial that compared the treatment of type 2 diabetic patients treated with rosiglitazone-based therapy versus metformin plus sulfonylurea therapy, the development of pancreatic cancer during the study was statistically less frequent in the rosiglitazone arm than in the active comparator arm (two cases vs. 13 cases; $P < .007$).[67]

Islet autotransplantation in pancreatic resection for pancreatic carcinoma has not been a generally accepted procedure because of the concern of introducing tumor cells carried along with the islets during the purification process. Balzano and coworkers recently reported the results of islet autotransplant in 14 patients with pancreatic or periampullary adenocarcinoma.[68] Eleven were disease-free at a median follow-up of 725 days, and none had tumor recurrence at the transplantation site. Further studies and better methods of islet purification will be necessary for this procedure to become more acceptable.

KEY POINTS

- Pancreatic ductal carcinoma is increasing worldwide and is the third or fourth most common cause of cancer deaths in the United States. Because it is diagnosed late in its course, 5-year survival rates range from 1% to 8%. Type 2 diabetes occurs in 68% of patients with pancreatic ductal carcinoma compared with 20% in age- and sex-matched patients with other carcinomas. Patients with long duration of type 2 diabetes have an increase in pancreatic cancer (1.6-fold), comparable to that of other cancers. However, type 2 diabetes occurring in the 2- to 3-year period before the diagnosis of pancreatic cancer is 5- to 8-fold higher than diabetes diagnosed during a comparable 2- to 3-year period in matched control patients. These data, plus the observations that diabetes frequently improves after tumor resection, indicate the involvement of pancreatic cancer in the pathogenesis of new-onset diabetes mellitus. The development of new-onset diabetes mellitus associated with weight loss should be an alert to evaluate the patient for pancreatic ductal cancer and make the diagnosis several years before symptoms from the cancer itself.

Autoimmune Pancreatitis

Autoimmune pancreatitis (AIP) is a steroid-responsive pancreatic fibroinflammatory disease that has been recognized with increasing frequency since 1995. The disease is associated with infiltration of immune cells into pancreatic tissue, and the disease responds dramatically to corticosteroids. Two subtypes have been identified.

Type 1 AIP is an IgG-4–related disease and is characterized by elevated serum IgG-4 levels and IgG-4–producing plasma cells infiltrating the pancreas.[69] It is the most frequent subtype and affects patients in their 60 and 70 years of life. Obstructive jaundice was the presenting symptom in 75% of the patients in a series of 731 AIP patients. IgG-4 accounts for less than 5% of the total serum IgG in healthy individuals and is ordinarily present at less than 140 mg/dL.[70] The median IgG-4 level in one series of patients with AIP was 663 mg/dL compared with 51 mg/dL in healthy age- and sex-matched controls. Type 1 AIP is frequently associated with involvement of other organs such as proximal biliary tract, kidney, salivary gland, lacrimal gland, and retroperitoneum. Although it responds to steroid therapy, the risk of recurrence on treatment is high. Patients with type 1 AIP have a variety of circulating autoantibodies. Late stages of type 1 AIP lead to exocrine and endocrine failure resulting in steatorrhea and DM.

Type 2 AIP is a pancreatic specific disorder not associated with IgG-4.[71] It accounts for approximately 10% to 20% of AIP patients and affects younger patients (third to fourth decade). Its only extrapancreatic organ involvement is inflammatory bowel disease, which occurs in approximately 30% of patients.[71] In the previously mentioned cohort of 731 patients with AIP, abdominal pain was the presenting symptom in 68% of patients with type 2 AIM.[70] The response rate to steroid therapy is very high, with a low risk of recurrence.

AIP is common in old age and is often associated with DM (33%–78%). In one series from Japan, 30 of 69 patients (46%) with AIP were diagnosed as having DM.[72] In 17, the diagnoses of DM and AIP were made simultaneously, and in 13 the diabetes was preexisting. Twenty-four of the patients were treated with prednisone 0.6 mg/kg body weight/day and gradually tapered over 3 months to a maintenance dose of 5 mg/day, which was maintained for 1 to 3 years. All of them responded well, with improvement in radiologic and serologic parameters. After 3 months of steroid therapy, 6/11 with simultaneous onset improved, and 7/13 with preexisting diabetes improved. Improvement was defined as a decrease in HbA1c greater than 0.5%. The improvement continued through 3 years of therapy in 66% and 60% of the new-onset and preexisting cases of diabetes, respectively. In another study in Japan, 47 patients with AIP were treated with steroids.[73] Of those, 33 (70%) had DM. Thirteen patients had improved glucose tolerance, nine had worsening glucose tolerance, and 32 were unchanged during a mean follow-up of 60 months. Some 17 patients had C-peptide response to a 1-mg intravenous bolus of glucagon at 6 minutes measured before and 1 month after starting steroid therapy. The C-peptide response was improved by the steroid treatment. A poor C-peptide response at baseline predicted the need for insulin even during steroid therapy. A good C-peptide response predicted successful diabetes treatment without the need for insulin. Steroid therapy in AIP is effective in treating the diabetes, as well as the other components of the disease.

First-line therapy for AIP is a loading dose of prednisolone (0.4–0.6 mg/kg) for 2 to 4 weeks and tapered by 5 mg every 1 to 2 weeks until the maintenance dose of 2.5 to 5.0 mg daily is reached (usually 6 months). Maintenance therapy is continued up to 3 years. With low-dose steroid treatment or treatment withdrawal the incidence of AIP recurrence has been reported to range from 24% to 52%. Steroid-sparing agents such as immunomodulators or rituximab are effective treatment modalities.

KEY POINTS

- Autoimmune pancreatitis and fibrocalculous pancreatic diabetes are two rare causes of diabetes mellitus that are important to recognize because of their specific treatment issues and their focus on the vulnerability of the exocrine pancreas to pathologic processes, which then secondarily involve the pancreatic islets. It has been recently recognized that genetic variants of genes regulating exocrine secretory processes predispose the exocrine pancreas to a variety of inflammatory and autoimmune triggers. Both variants of autoimmune pancreatitis respond to glucocorticoid treatment and perhaps other immunosuppressive therapies. Little is known about fibrocalculous pancreatic diabetes other than that its clinical features are changing, and its prevalence is decreasing. It is likely that it is caused by an unknown environmental trigger that activates a susceptible exocrine pancreas to an inflammatory state. The high risk for pancreatic cancer is important to recognize.

Hemochromatosis

Hemochromatosis is an autosomal recessive genetic disorder that results in excessive deposition of iron in parenchymal cells of the liver, pancreas, muscle, heart, anterior pituitary gland, and other organs. The clinical diagnosis in the past was made by the findings of DM, hepatomegaly, and skin pigmentation. More recently, it has been recognized through biochemical and genetic testing.[74]

Dietary iron is absorbed only to compensate for loss and, to a limited extent, storage in the liver.[74,75] When dietary iron is taken up by enterocytes, tissue ferritin binds and stores iron in the enterocyte, preventing it from directly entering the bloodstream. Ferroportin allows iron to be transported out of the enterocytes for release into plasma. Once in the plasma, iron is bound to transferrin and delivered to bone marrow for haem synthesis and to the liver for storage. A hepatic peptide, hepcidin, is the key hormone involved in the regulation of iron absorption.[75] Hepcidin interacts with and inactivates the iron export protein ferroportin, which limits the amount of dietary iron entering the plasma from enterocytes, recycled iron from macrophages, and stored iron from hepatocytes. Hepcidin production is regulated by the bioavailability of iron; hepcidin synthesis increases with increased iron storage in the liver, limiting further absorption of iron. Iron deficiency results in downregulation of hepcidin production by the liver, allowing iron absorption to increase.

Hemochromatosis is commonly attributed to the autosomal recessive inheritance of a mutation in the *HFE* gene (type 1 hemochromatosis) encoding a C282Y substitution that results in a deficient or inactive hepcidin.[75] Although this accounts for 80% to 90% of cases of hemochromatosis, several mutations in genes other than *HFE* (including *HIV* and *HAMP, TFR2*, and *ferroportin 1*) have been recently discovered and account for types 2, 3, and 4 hemochromatosis, respectively.[74] The carrier rate for the C282Y-encoding mutation among populations of Northern European ancestry is between 10% and 15%. Heterozygotes may have mildly elevated iron studies but do not develop clinical disease. Even homozygotes may have a variable penetrance of the disease. Additional metabolic abnormalities such as obesity with steatohepatitis or excessive alcohol consumption increase the penetrance and severity of the disease.

The first phenotypic expression of the disease is an elevation in serum transferrin saturation. This abnormality is followed by iron accumulation in the tissues and an elevation in the serum ferritin concentration. Early clinical findings are related to hepatic dysfunction and joint symptoms. Clinical DM and skin pigmentation occur relatively late during the disease. DM has been reported in 50% to 60% of patients with hemochromatosis.[74,76] Another 20% to 30% had glucose intolerance. These figures represent data from older series in which the diagnosis was made late in the disease. DM is more frequent in patients who have a family history of DM.

Follow-up of 237 patients who were given a diagnosis of hemochromatosis and treated at a single center from 1983 through 2005 has provided insights into the natural history of the cirrhosis and DM associated with the disease.[77] Before 1996, hemochromatosis was diagnosed by the classic clinical and laboratory features of the disease. Since 1996, genetic testing, which is now available commercially, made it possible to diagnose hemochromatosis in asymptomatic patients.[78] Of 45 patients diagnosed before 1996, 30.2% had cirrhosis of the liver, and 35.6% had DM. After 1996, 192 patients were given the diagnosis of hemochromatosis (34% by family screening and 40% by clinical suspicion), only 7.5% had cirrhosis of the liver, and 17.7% had DM at the time of diagnosis. It is obvious that early diagnosis through genetic testing and effective treatment can significantly reduce development of the complications of hemochromatosis. The established treatment for hemochromatosis is frequent phlebotomies to reduce the iron stores. Of note is the observation that performing phlebotomy and reducing tissue iron in well-established clinical disease did not improve diabetes or lessen its management requirements. Thus, early diagnosis and treatment is primarily preventative.[78]

The metabolic studies that have been done show that patients with hemochromatosis have marked insulin resistance.[76] Histologic study of the pancreas shows iron deposits that are greatest in the acinar cells but do involve islet cells.[74] Insulin secretion in response to glucose or arginine is decreased; however, glucagon secretory responses to arginine are increased and unaffected by glucose.[76] The data are compatible with a marked reduction in β cell function and no disturbance in α cell function. The hyperglycemia is a result of insulin resistance and decreased β cell function.[76] The prevalence of DM probably could be reduced by early diagnosis of hemochromatosis and initiation of phlebotomy therapy.[74] Therapy for patients with hemochromatosis and clinical diabetes frequently requires insulin (40%–50% of patients), although no systematic studies of therapy have been done.[76] Reduction of tissue iron stores, although most beneficial in the early stages of disease, nonetheless can help improve glycemic control in 35% to 45% of patients.

Hemosiderosis

Excessive iron deposition occurs in a variety of conditions other than primary hemochromatosis. In thalassemia major, frequent blood transfusions are necessary and may lead to massive iron overload. The reported prevalence of DM in treated thalassemia major is approximately 16%. This figure is highly correlated with the number of blood transfusions given and the duration of the disease. The incidence of IGT is reported to be 60%.[79] Further evidence that deposits of excess tissue iron themselves are responsible for many of the metabolic abnormalities seen in hemochromatosis and thalassemia major comes from studies in rural male Bantus.[80] Many Bantus drink alcoholic beverages that are brewed in iron containers and ingest in excess of 100 mg of iron per day. In those individuals, the prevalence of DM is 10-fold higher than in nonalcoholic beverage–consuming rural Bantu males.

Mechanistic studies in patients with thalassemia major with normal, impaired, and diabetic glucose tolerance tests show that increased iron stores are associated with the development of insulin resistance and a delay in early insulin secretion.[81] A correlation between increased iron stores in normal women and the development of type 2 diabetes was demonstrated recently in the Nurses' Health Study.[82] Several large epidemiologic studies have shown that high body iron stores are associated with an increased risk for developing type 2 diabetes. In a prospective cohort of 1613 males in the Kuopio Ischemic Heart Disease Risk Factor study, type 2 diabetes developed in 331 cases during a 16.8-year follow-up.[83] Baseline serum ferritin was 191 μg/L in those who developed type 2 diabetes compared with 151 μg/L in those who remained nondiabetic. This represented a 14% increase in diabetes for each 100-μg/L increase in serum ferritin. Similar findings were found in 27,548 individuals in the European Prospective Investigation into Cancer and Nutrition–Potsdam study,[84] where after adjustment for all known risk factors, those in the highest quintile of serum ferritin concentrations had a relative risk for type 2 diabetes of 1.73 compared with those in the lowest quintile. No significant association was found for the soluble transferrin receptor.

KEY POINTS

- Diabetes mellitus (DM) associated with hemochromatosis and hemosiderosis is caused by deposition of iron into the pancreatic islet cells. The iron causes death of the islet β cells which lead to progressive decreases in insulin secretion. Glucagon secretion seems to be little-affected. Iron deposition in liver, muscle, and other tissues causes insulin resistance. Hemochromatosis is a genetic disorder involving the hepcidin gene (*HEF1*) or non-*HEF1* genes that regulate iron absorption. The clinical manifestations are bronze coloration of the skin, hepatic cirrhosis, and diabetes. The diagnosis now is made by genetic testing of susceptible individuals. The prevalence of DM depends on the time of diagnosis. Those diagnosed by genetic testing have a prevalence of DM of 18%, whereas in those diagnosed by clinical criteria it is 36%. The treatment of hemochromatosis is frequent phlebotomies. Early treatment can reduce the development of DM, but once it develops it cannot decrease, although the management of hyperglycemia is easier. Hemosiderosis is caused by frequent blood transfusions such as occur in thalassemia major. The prevalence of DM in hemosiderosis is approximately 16%. Treatment of DM in hemosiderosis or thalassemia is insulin.

Cystic Fibrosis

Cystic fibrosis (CF) is a monogenetic disorder with abnormal cyclic adenosine monophosphate-regulated chloride channel activity. It is an autosomal recessive genetic disease with an incidence of 1 in 2500 live births in White populations. CF affects over 90,000 individuals, and they are heterogeneously distributed worldwide. Over 2000 mutations have been identified in the *CFTR* gene. Approximately 440 are disease-causing. *CFTR* mutations are categorized into five groups: class I mutations are associated with the more severe phenotypes ,and class V the milder phenotypes. The F508del is the most prevalent CF-causing mutation, affecting approximately 82% of the CF population. It is associated with deletion of phenylalanine at position 508. There is a strong association between the *TCF7L2* genotype and the risk for CF-related diabetes (CFRD).[85,86] The disease-causing mutations result in thick, sticky mucous secretions in mucous-producing organs such as lung, exocrine pancreas, large and small intestine, hepatobiliary system, and sweat glands. Failure to secrete sodium, bicarbonate, and water leads to retention of enzymes in the pancreas, and ultimately to destruction of pancreatic tissue. Histologic examination of the pancreas in patients with CF shows fatty infiltration, necrosis, and fibrosis of the exocrine pancreas. Islet cell architecture is disrupted, and the absolute number of pancreatic islets is diminished. Islets that are present show significant decreases in β cells, α cells, and pancreatic polypeptide-producing cells and increases in δ (somatostatin-producing) cells. Islet amyloid deposits have been found in 69% of diabetic CF cases examined.[87-89]

CFRD is a severe complication of CF and has increased in prevalence as patient survival has increased from 8 years of age in 1970 to 40 to 50 years of age at the present time.[87-89] The prevalence of CFRD increases from 2% in childhood, to 20% in adolescence, to 40% to 50% in adults. The presence of CFRD impacts survival in patients with CF: fewer than 25% of patients with CFRD survive to an age of 30 years compared with 60% of CF patients with normal glucose tolerance.[87-89]

DM requiring medical therapy (usually insulin) has been reported in 4.9% of patients of all ages with CF in a large European study of 1348 patients and in 5.1% of 18,627 patients of all ages monitored at CF centers in the United States and Canada. Diabetes occurs more often in individuals who are homozygous for the most common CF mutation, F508del. Diabetes has been reported in 32% of Danish patients who were older than 25 years of age. Routine OGTTs suggests that, of the total CF population 5 years of age or older, 35% have normal glucose tolerance, 37% have IGT, 17% have CFRD without fasting hyperglycemia, and 11% have CFRD with fasting hyperglycemia.[87,88,90]

The primary defect in CFRD is a reduced and delayed insulin secretory response to carbohydrates that starts in childhood, progresses over time, and results in a 50% prevalence of DM in CF patients over 35 years of age. The decrease in islet cell function is secondary to the exocrine pancreatic insufficiency, as the glucose and insulin responses to oral glucose in CF patients are normal until there is exocrine deficiency.[91] Recent data indicate that defective β cell CFTR protein also plays a role in reducing insulin secretory function. This pattern of insulin insufficiency spares basal insulin secretion and results in a glucose abnormality characterized by normal FPG and postprandial hyperglycemia. Late in CFRD, total insulin, glucagon, and pancreatic polypeptide secretions are decreased. First-phase insulin secretion after intravenous glucose is markedly reduced in patients with CF compared with matched controls. The pattern of abnormal glycemia is normal plasma glucose before breakfast with abnormal postprandial glucose rises that are highest after the evening meal.[85,87,88]

Peripheral insulin sensitivity is usually normal or increased in CF patients, but serious infections and systemic corticosteroid treatment increase insulin resistance. Insulin resistance occurs later as IGT and diabetes develop, and patients experience additional complications related to their CF.

The delay in insulin secretion commonly results in hypoglycemia, even in individuals who do not have CFRD. Among children and adolescents with CF, 13% experience hypoglycemia during routine OGTTs, and this represents a mismatch between the prompt absorption of an oral glucose solution and delayed first-phase insulin secretion in CF. Ketosis is rare.[90]

Registry studies from the 1980s and 1990s showed significant increases in mortality in CF patients who developed CFRD, even after adjustment for other risk factors. In adults, overall mortality was double in patients with CFRD compared with CF patients without DM. Earlier diagnosis and treatment have been shown to decrease mortality in patients with CFRD: comparing cohorts for 1992 to 1997 and 2008 to 2012, age-adjusted mortality for individuals with CFRD fell from 4.9 to 1.8 per 100 person-years. Infection rates are increased, and lung function declines more quickly in people with CFRD. Although the risk of microvascular complications occurs as in other forms of diabetes, the main reason for treatment is to prevent deterioration in lung function and weight loss. Several studies have shown that weight loss and increasing pulmonary function deterioration are observed several years before the diagnosis of CFRD, indicating that early changes in insulin secretion and glucose metabolism hasten the clinical course of CF. The evidence that glucose abnormalities that do not fulfill the diagnostic criteria for diabetes can significantly impact clinical status influences the decision to treat CFRD. Treatment may therefore be required at an earlier stage in CFRD than for other types of diabetes.[85,88]

The recommended diagnostic criterion for CFRD is a 2-hour plasma glucose of 11.1 mmol/L or more in an OGTT. Neither FPG nor a HbA1c is reliable to diagnose CFRD. Screening is recommended to start at 10 years of age. More recently, it has been proposed to use CGM for diagnosis. In people with CFRD undergoing CGM, glucose was detected in airway secretions when blood glucose levels rose above 8 mmol/L. This level was exceeded for 45% of the day in those with CFRD, compared with 6% of the day in people with CF and normal glucose tolerance.[85,87,88]

Hyperglycemia in patients with CF can be intermittent or permanent. Intermittent hyperglycemia occurs with glucocorticoid therapy, infection, or stress and must be treated with insulin until it resolves. Permanent hyperglycemia is always treated with insulin. It is likely that patients with CF progress from intermittent hyperglycemia to permanent hyperglycemia as pancreatic destruction continues to occur.

Treatment of CFRD consists of a high-calorie, low–simple carbohydrate diet. Many people with CF take pancreatic enzymes to improve food- and fat-soluble vitamin absorption and to prevent steatorrhea. Active, but not total, glucagon-like peptide-1 (GLP-1) plasma levels are decreased in CF patients and tend to be even lower in patients with CFRD. Pancreatic enzyme replacement improves GLP-1 secretion. People with CF have rapid gastric emptying.[86-89]

Oral sulfonylureas and other antidiabetic agents have been tried for treatment of CFRD; however, the treatment of choice for the postprandial hyperglycemia is short-acting insulin before meals. Treatment with insulin has been associated with up to 22-mmol/mol (2%) improvement in HbA1c after 1 year's treatment, and 11-mmol/mol (1%) improvement after 5 years.[87,90] Improved glycemic control is associated with improvements in body weight (BMI improvement of 1–3 kg/m^2) and lung function (6%–35% increased forced expiratory volume in 1 second).[88-90] Insulin therapy may also lead to a reduction in the frequency of chest infections, although this has not been demonstrated in all studies. There is no convincing evidence to support the use of any insulin regimen. Of the 10,810 people recorded in the CF Foundation Patient Registry for 2015, 1982 (18.3%) were receiving diabetes treatment.[85,88-90]

Specific treatment for CF requires replacing the specific mutant gene by modifying it to a non–disease-causing mutation. This would be a formidable task, as the gene is expressed in so many tissues. A more feasible approach is to determine the process by which the mutant gene acts to create the abnormally functional CFTR protein and create small molecules that can be administered to patients that can improve the defect responsible for the poor function of CFTR. This is the process that has been and is being pursued. The new therapeutic agents being used to treat CF are classified by their action on specific abnormalities of CFTR generated by the gene mutations.[92] They include:

1. Potentiators: restoring the channel gating and conductance. Around 5% of CF-causing mutations lead to impaired CFTR channel gating or conductance as primary defects (classes III and IV).
2. Correctors: rescuing the protein folding, processing, and trafficking. Most CF patients carry a mistrafficking CFTR mutation such as F508del, which is the most prevalent CF-causing mutation (class II).
3. Stabilizers: rescuing the protein stability at the plasma membrane. Although certain CFTR mutants are functional and present at the plasma membrane, the protein may still display a significant reduction in half-life (class VI).
4. Read-through agents and nonsense-mediated mRNA decay inhibitors: rescuing the protein synthesis. A significant fraction of the CF-causing mutations are in-frame nonsense, frameshift, and splicing variants that introduce a premature termination codon into the *CFTR* mRNA, thus abrogating CFTR protein synthesis or resulting in translation of shortened, truncated forms (class I).
5. Amplifiers: increasing the abundance of protein substrate. Some CF-causing mutations lead to reduction in the synthesis or maturation of the CFTR protein.
6. Antisense oligonucleotides (ASOs): correcting the aberrant splicing and more. ASOs are chemically modified synthetic RNA-like molecules that act by complementary base pairing to the target sequence. ASOs have been demonstrating promising results to correct nonsense and splicing mutations, or even to replace missing bases caused by deletion mutations such as F508del.

Ivacaftor is one of the new modifiers that have been developed to treat some forms of CF. It partially reverses the effects of the mildest mutations in CF (gating mutations); there are newer agents in development with effects on the more severe mutations. Ivacaftor is being evaluated for its effects on glucose metabolism and insulin secretion in patients with CF to determine its potential for the prevention and/or treatment of CFRD. A study of five people with CF (three with normal glucose tolerance, one with recent diagnosis of CFRD, and one with CFRD for 16 years) showed that, after 1 month's treatment with ivacaftor, the insulin response to oral glucose improved by 66% to 178%, except for the individual with longstanding diabetes, who showed no improvement. A recent study of 12 CF patients with a median age of 13.8 years, seven of whom had a normal OGTT, four of whom had IGT, and one with a 1-hour OGTT glucose greater than 155 mg/dL evaluated the effects of 4 months' treatment with ivacaftor on glucose, insulin, and incretin plasma levels following oral glucose and meal stimulation and an intravenous arginine test. Insulin secretion improved; however, plasma glucose and incretin levels were not improved.[93] The mechanism for the improvement in plasma insulin is not understood but points to a role of the CFTR modulators in the regulation of insulin secretion by the β cell. The duration and clinical benefit of this effect on insulin secretion in this very young CF population needs to be investigated.

Microvascular complications have been observed in patients who had CFRD with fasting hyperglycemia for longer than 10 years. Fourteen percent had microalbuminuria, and 16% had retinopathy. Macrovascular disease appears to be quite uncommon.

KEY POINTS

- Cystic fibrosis–related diabetes (CFRD) is a form of diabetes that occurs in individuals with mutations of the cystic fibrosis transmembrane conductance regulator gene. The result of these mutations is an abnormal cyclic adenosine monophosphate–regulated chloride channel, which causes mucous secretions to be abnormal and very viscous, such that they block pulmonary bronchi, pancreatic ducts, and other secretory channels. The results are chronic and progressive pulmonary disease, as well as pancreatitis and DM. The destruction of islets and deficient β cell insulin secretion start at birth and increase with the duration of the disease. The prevalence of CFRD increases from 2% in childhood, to 20% in adolescence, to 40% to 50% in adults. The presence of CFRD impacts survival in patients with cystic fibrosis: fewer than 25% of patients with CFRD survive to an age of 30 years, compared with 60% of CF patients with normal glucose tolerance. CFRD is characterized by relatively minimal fasting hyperglycemia with marked postprandial hyperglycemia. Treatment is directed toward treating the postprandial hyperglycemia. Treatment of hyperglycemia should commence as early as possible, as hyperglycemia hastens pulmonary decline.

HYPERGLYCEMIA ASSOCIATED WITH ENDOCRINOPATHIES

In the complex regulation of fuel homeostasis, many hormones other than insulin play a complementary role. Growth hormone (GH), by itself and through stimulating synthesis of IGF-1, controls many aspects of amino acid transport, protein synthesis, and lipid metabolism. Glucagon and catecholamines are counterregulatory hormones that protect against hypoglycemia and provide extra glucose when needed during stress states. Glucocorticoids have both a permissive role in the normal physiologic regulation of gluconeogenesis and a pharmacologic role in providing increased glucose availability during stress. Somatostatin is a paracrine hormone that appears to act locally to help regulate the normal secretory patterns of GH, insulin, glucagon, and several gastrointestinal hormones.

The mechanisms responsible for the action of these various hormones are described in detail in other chapters.

This section will address the unique characteristics of hyperglycemia as it relates to each endocrinopathy, as well as its treatment.

Acromegaly

Acromegaly is characterized by excessive and autonomous secretion of GH and IGF-1. The disease has usually been present for a mean of 7 years at the time of diagnosis. According to the current consensus guidelines, acromegaly is diagnosed if IGF-1 levels are above the age-adjusted normal range and the GH nadir during an OGTT is higher than 0.4 μg/L.[94-96] The prevalence of overt DM reported in different series of acromegalic patients ranges from 10.5% to 52.6%. IGT may be present in as many as 36% of acromegalic patients. Less than 30% of patients with acromegaly are normoglycemic.[97,98] The prevalence of DM in acromegalic patients is a function of increasing age, higher BMI, longer duration of untreated acromegaly, and the genetic predisposition of the corresponding nonacromegalic population to develop DM. Approximately 20% of newly diagnosed patients with acromegaly have DM, and undiagnosed acromegaly is present in 0.5% to 3% of patients being followed for type 2 DM. The onset of DM can precede the diagnosis of acromegaly by 1 to 3 years, as well as following its clinical diagnosis. Successful treatment of acromegaly has been reported to result in a remission of the diabetes in 23% to 58% of the patients. The longer the acromegaly is untreated, the less likely it is that remission will occur with treatment.

Elevated GH and IGF-1 levels cause increased adipose tissue lipolysis and insulin resistance, which results in excessive hepatic glucose production and impaired insulin-mediated muscle and adipose tissue glucose uptake. Insulin resistance is correlated with circulating IGF-1 levels and has been demonstrated by the euglycemic hyperinsulinemic clamp and the minimal model techniques.[94,97]

The primary treatment for acromegaly is transsphenoidal surgical removal of the tumor producing GH or GH-releasing factor. Surgical remission is defined as a random or a nadir GH level in an OGTT less than 1.0 μg/L and a normal age-matched IGF-1 level 3 months after surgery. This is achieved in approximately 50% of patients, with total tumor resection being more common in microadenomas (75%–95%) than in macroadenomas (40%–69%). Thus, in general, half of the patients will require adjuvant treatment to achieve remission. Although not supported by clinical follow-up, some consensus groups suggest that a nadir GH level during an OGTT using newer, higher-quality assays of 0.4 μg/L should be the value used to define remission.[99]

Medical therapy is recommended for patients with persistent disease after surgical resection of the adenoma, as well as for patients in whom surgery is not appropriate. Adjuvant medical treatment use consists of three drug classes: somatostatin receptor ligands (SRLs), dopamine agonists (DAs), and GH receptor (GHR) antagonists.[96,99]

The SRLs (octreotide LAR and lanreotide autogel) have been considered the first-choice medical treatment.[94,95,99] These drugs act through targeting somatostatin receptor subtype 2 and, to a much lesser extent, receptor subtype 5, leading to a decrease in GH secretion and tumor shrinkage. Approximately 40% to 50% of patients treated with these drugs achieve full biochemical control of acromegaly. For patients receiving medical treatment, disease control is defined as a normal age-matched IGF-1 level. The GH nadir in the OGTT has a low accuracy and should not be used, as SRLs interfere with the feedback of glucose. The multireceptor-targeted SRL pasireotide differs from the other SRLs owing to its higher affinity for somatostatin receptors 3 and 5. It is more effective than other SRLs and has a broader range of activities (suppression of incretin secretion, greater inhibition of insulin secretion, and some suppression of glucagon secretion). In a randomized prospective study, disease control with pasireotide was achieved in 20% of patients whose disease was uncontrolled after treatment with octreotide LAR or lanreotide autogel.

The DA cabergoline acts by binding to dopamine-2 receptors. It has a lower efficacy than other drug classes (~20%–30% as a monotherapy and 30%–35% in combination therapy with SRLs), and some patients present treatment escape during long-term therapy. Cabergoline can be used as an adjuvant treatment, particularly for patients with mild disease (IGF-1 levels up to 1.5-fold the upper limit of the normal range).[95,99]

Pegvisomant (PEG) is a GHR antagonist that binds to peripheral GHRs without activating the intracellular signaling pathways. PEG has no action on somatotropin-producing pituitary tumors. It is the drug with the highest efficacy in controlling acromegaly, with success rates ranging from 63% to 97%, depending on the study. PEG has been increasingly used as first-line monotherapy or in combination with SRLs or cabergoline.[99]

Radiotherapy is effective at controlling tumor volume in 90% of cases and leads to biochemical control in approximately 60% of cases, but GH normalization takes several years. Although highly effective, radiotherapy is associated with many significant long-term side effects. Radiotherapy is considered a third-line therapy and is reserved for patients with aggressive tumors that are resistant to surgical and medical treatment.[94,95,99]

Surgical removal of GH-secreting pituitary tumors improves both glucose tolerance and diabetes. Insulin secretion and sensitivity improve within 1 to 2 months after surgery, and normalization of glucose homeostasis occurred in 23% to 58% of patients with preexisting diabetes, depending on the cohort studied. Long-term response data are not available, particularly in subjects in whom cure of acromegaly and/or improved β cell function was not achieved.[95,99]

Use of medical treatments (SRL, GHR antagonist, DA) as primary therapy or to supplement prior inadequate surgical treatment has allowed attainment of greater and more consistent normalization in circulating GH and IGF-1 levels (77%). Normalization of IFG-1 in acromegalic patients with DM results in a 30% remission of the diabetes and easier management of the remainder. Treatment of acromegalic patients with SRLs presents several issues with respect to glucose metabolism. Reductions in circulating GH and IGF-1 levels will decrease insulin resistance and should lead to improvement in glycemic control in subjects with DM-related IGT. However, pharmacologic doses of SRLs also reduce insulin secretion (decreased insulinogenic index), and such a reduction should cause a deterioration in glucose tolerance. In a specific patient, SRL therapy will modify glucose metabolism as an interaction of these competing effects.[99]

Approximately two-thirds of acromegalic patients with DM are treated with insulin, and one-third are treated with oral hypoglycemic agents. SRL treatment in patients with DM and acromegaly frequently leads to improvement in glycemic control, as measured by a reduction in the insulin dose, conversion from insulin therapy to oral hypoglycemic agent therapy, or conversion from oral hypoglycemic agent therapy to dietary management. Some patients (those with more severe insulin deficiency), however, will have significant deterioration in glycemic control. When high doses of SRLs are given to acromegalic patients with normal glucose tolerance, IGT and even frank DM (at rates as high as 20% and 29%, respectively) may develop.[95,99]

An alternative to treatment with SRLs is treatment with the GHR blocker PEG. This agent decreases IGF-1 production and improves glycemic control. In contrast to SRLs, it increases GH secretion and has no inhibitory effects on insulin, glucagon, or incretin secretion. However, it causes an increase in visceral fat and occasionally is associated with increased liver enzymes.[95,99]

Appropriate treatment for acromegaly is necessary to reduce the increased mortality that has been seen in the past. This increased mortality is attributed to cardiovascular, cerebrovascular, and neoplastic diseases. The best determinants of outcome in acromegalic patients are age at diagnosis, interval between symptoms and diagnosis, and mean long-term circulating GH and IGF-1 levels. Because insulin resistance and DM contribute significantly to cardiovascular risk, aggressive diagnosis and management of the DM associated with acromegaly are essential.

Growth Hormone Treatment

The availability of recombinant DNA technology to make human GH has provided an opportunity to treat many GH-deficient individuals. One consideration in treatment with recombinant human GH is the question of whether long-term treatment leads to the development of DM. A recent study of glucose metabolism in 22,311 children and adolescents treated with human GH and followed for a mean of 4.2 years found a 3.8-fold greater frequency of type 2 diabetes (16.8 cases per 100,000 years of GH treatment) than predicted.[100] Thirteen of the 18 patients had risk factors for diabetes. The frequency of type 1 diabetes was the same as expected. A French study of 5100 children treated with GH and evaluated at 30 years of age (mean follow-up 15 years) had the same prevalence of treated diabetes (oral drugs or insulin) as the general population of France.[101] The data suggest that GH treatment accelerates the development of type 2 diabetes only in those individuals who have risk factors that predispose to the development of type 2 diabetes.

Glucocorticoid Effects on Glucose Metabolism and Corticosteroid-Mediated Diabetes Mellitus

Glucocorticoids are insulin antagonistic hormones. When administered in pharmacologic doses, they increase hepatic gluconeogenesis, decrease muscle glucose uptake and glycogen synthesis, increase muscle proteolysis, and increase adipose tissue lipolysis.[102,103] Compensatory insulin secretion that should increase as a consequence of insulin resistance is blunted by glucocorticoid modulation of gene expression in β cells, which causes a decrease in cytoplasmic calcium stimulation of insulin secretory granule exocytosis.[102,103] Additionally, short-term exposure to glucocorticoids reduces the insulinotropic effects of GLP-1. A consequence of these pharmacologic effects is glucocorticoid-induced glucose intolerance, the magnitude of which is dependent on the dose and duration of glucocorticoid exposure. When sustained pharmacologic concentrations of glucocorticoids occur in normal individuals, the compensatory increase in insulin secretion maintains FPG within the normal range, but the postprandial plasma glucose concentration is elevated above normal in 25% to 90% of such individuals.[102-104] Fasting hyperglycemia would be expected in individuals with limited baseline β cell insulin secretory reserve, as occurs in individuals genetically susceptible to type 2 or other forms of β cell–deficient diabetes.

Exogenous steroid use involves as much as 1% of the general population; 40% of this is for respiratory diseases, and the remainder is split between antiinflammatory and immunosuppressive purposes.[104] Two types of exogenous steroid treatment regimens prevail: high-dose administration for a limited period followed by a slow downward titration and prolonged low dose/discontinuation, or high prolonged maintenance dose for severe diseases. The odds ratio for development of steroid-induced new-onset DM is reported to be 1.36 to 2.31.[104] Among renal transplant recipients receiving long-term corticosteroid therapy, steroid-induced DM has been reported to develop in as little as 5.5% to as much as 46% of patients.[103,104] Factors that influence the development of DM during corticosteroid therapy include a family history of DM, increasing age, obesity, and both average daily and total cumulative corticosteroid dose. In individuals with a previous onset of DM, administration of glucocorticoids significantly worsens glycemic control and requires modification of diabetes management.

Some insight into the mechanisms by which glucocorticoids cause DM was obtained by investigating the effects of administration of dexamethasone on oral glucose tolerance, on glucose turnover under basal conditions and during glucose infusion, and on the insulin response during hyperglycemic clamp studies in normal individuals who had been characterized previously as low insulin responders or high insulin responders.[105] Dexamethasone caused a higher FPG concentration, a greater rise in plasma glucose, and a lesser rise in plasma insulin during the OGTT in the low insulin responders than in the high insulin responders. A diabetic OGTT result was obtained in three of the six low and none of the six high insulin responders. Dexamethasone increased hepatic glucose production only in the low insulin responders and increased insulin secretion during the hyperglycemic clamp study only in the high insulin responders. The conclusion drawn from these studies is that type 2 diabetes develops when plasma glucocorticoids are elevated in individuals with limited β cell secretory function.

Steroid-induced diabetes may be permanent or transient. In general, insulin is required for treatment if the FPG concentration exceeds 180 mg/dL. At lesser levels of fasting hyperglycemia, many physicians treat the hyperglycemia with oral hypoglycemic agents. Few studies have adequately evaluated the efficacy of pharmacologic treatment of steroid-induced DM. Because insulin resistance is a major abnormality, perhaps a combination of insulin sensitizers and insulin would be most effective.[106,107] A recent 14-day study in men with metabolic syndrome assessed the benefit of coadministration of a DPP-4 inhibitor (sitagliptin) with prednisone in preventing the diabetogenic effects of glucocorticoid administration. The concomitant administration of the DPP-4 inhibitor improved various aspects of pancreatic islet-cell function but did not prevent the prednisone-induced increment in postprandial glucose concentrations.[107] A 4-week randomized study of patients receiving glucocorticoids showed that metformin prevented the deterioration of glucose effects during OGTT as compared with placebo treatment.[107] A reduction in corticosteroid dose improves glycemic control and, in some individuals, may even reverse the diabetes. The more severely elevated the FPG concentration, the less likely it is that reducing corticosteroid levels will reverse the diabetes. Ketoacidosis is very uncommon with steroid-induced diabetes or Cushing syndrome. Hyperosmolar nonketotic coma, however, is not uncommon.

Cushing Syndrome

Depending on the reported patient population, 20% to 45% of patients with Cushing syndrome have DM, and another 10% to 30% have impaired glucose metabolism, meaning that the overall prevalence of abnormal glucose metabolism can be as high as 70%.[108,109] Because FPG is a late manifestation of glucocorticoid-mediated glucose abnormalities, FPG may be insufficient to diagnosis diabetes in more than 50% of Cushing syndrome patients. Age, genetic predisposition, lifestyle, cortisol secretory rate, and duration of cortisol excess are major risk factors for development of DM in Cushing syndrome. Treatment of the Cushing syndrome results in remission of diabetes in some patients and improvement in the remainder.

The recommended first-line therapy for pituitary-mediated Cushing syndrome (Cushing disease) is resection of the pituitary adenoma. The remission rate ranges from 65% to 90%, with a risk of recurrence at 10 years of between 10% and 20%.[108,110] When surgery is not possible or has only been partially successful, pharmacologic agents are used to control the hypercortisolism and the diabetes. The medical therapies that have been used are somatostatin analogues, DAs, and adrenal steroidogenesis inhibitors (ketoconazole, metyrapone, mitotane, and etomidate).[108,111,112] Each has its own issues with effectiveness and safety concerns. Carbergoline, a DA, decreases the prevalence of diabetes and glucose intolerance by 60% and 46%, respectively, in patients with pituitary Cushing syndrome. The long-acting somatostatin analogu octreotide LAR and lanreotide autogel are not effective in normalizing hypercortisolism in patients with pituitary Cushing syndrome and can worsen diabetes by inhibiting β cell insulin secretion. Pasireotide, a multireceptor somatostatin analog, may be effective in treating the endogenous hypercortisolism, but it has been shown to cause hyperglycemia and worsen diabetes by its effect in inhibiting insulin secretion. Mifepristone, a glucocorticoid receptor antagonist, produces clinical benefits in patients with Cushing syndrome and improves FPG (from 149 to 105 mg/dL) and HbA1c (from 7.43% to 6.29%) in those with diabetes.[109,112]

Adrenalectomy is the therapeutic gold standard for patients with adrenal CS caused by functioning adrenal adenomas. Bilateral adrenalectomy could be performed as a rescue treatment in patients with persistent and severe disease, but this approach causes permanent adrenal insufficiency and carries the risk of corticotroph tumor progression (Nelson syndrome).[108,110] Obviously, bilateral adrenalectomy results in adrenal insufficiency and will require lifelong corticosteroid replacement therapy.

Adrenal-directed drugs reduce cortisol secretion by the adrenal glands.[108,109,112] The therapies most often used or under investigation are ketoconazole, metyrapone, mitotane, etomidate, and LCI699. They can be used in every form of Cushing syndrome and may be highly effective in controlling cortisol excess. Ketoconazole, metyrapone, and LCI699 are dose-dependent reversible inhibitors of adrenal cortisol synthesis, acting on one or more enzymes responsible for adrenal steroidogenesis. Ketoconazole (200–1200 mg daily) improves glycemic control in 30% to 56.3% of patients with Cushing syndrome and DM. At doses higher than 4 g/day, mitotane also has an adrenolytic effect, inducing cellular necrosis, so it is indicated mainly for the treatment of adrenal cancer.

Recent studies have documented that steroid-induced DM occurs in adrenal disorders other than those associated with frank Cushing syndrome.[113] Measurement of oral glucose tolerance in 64 consecutive patients with “nonfunctioning” adrenal adenomas identified normal glucose tolerance in 25, glucose intolerance in 17, and DM in 22, including six patients with previously diagnosed DM. Autonomous cortisol secretion without the clinical stigmata of Cushing syndrome has been recognized recently as preclinical Cushing syndrome. A retrospective study of 63 such individuals found that 17.5% had DM. A cross-sectional study of 90 obese, poorly controlled type 2 diabetic patients found that three had the preclinical Cushing syndrome abnormality.[114] Screening 813 patients without clinical characteristics of hypercortisolism being followed for type 2 diabetes found six patients who failed to suppress cortisol secretion during a 2-day standard 2-mg/day dexamethasone suppression test.[114] Excess IGT and DM have been reported in hypopituitary adults receiving conventional replacement therapy, and may be related to the intermittently higher plasma levels that occur after dosing than would occur under normal hypothalamic-pituitary-adrenal axis function.

The DM associated with Cushing syndrome exaggerates the detrimental effects of the other complications of Cushing syndrome such as cardiovascular disease, osteoporosis, hyperlipidemia, and central obesity, and the recognition and treatment of the hyperglycemia is necessary to improve clinical outcomes.

Glucagonoma Syndrome

Proglucagon is synthesized in the pancreatic islet α cells and the intestinal L cells. Proglucagon is processed in the α cell by prohormone convertase 2 into glucagon and several inactive peptides.[115] In the L cell, proglucagon is processed by prohormone convertase 1 and 3 into oxyntomodulin, GLP-1, and GLP-2.[115] Glucagon plays a primary role in facilitating the uptake of amino acids by the liver and converting them into glucose by gluconeogenesis and urea by ureagenesis. GLP-1 is an incretin hormone, and GLP-2 promotes nutrient absorption through activation of crypt cell proliferation and expansion of small bowel epithelium and, perhaps more importantly, inhibits epithelial apoptosis.[115]

Excess and unregulated proglucagon synthesis occurs in some islet cell tumors. A classic syndrome has been described in individuals whose tumors process this proglucagon into glucagon, resulting in severe hyperglucagonemia. This syndrome was recognized in 1974 and is referred to as glucagonoma syndrome.[116] Features of this syndrome include necrolytic migratory erythema, mild non–insulin-requiring type 2 diabetes, glossitis, angular cheilitis, weight loss, anemia, and psychiatric symptoms. Laboratory studies show markedly elevated plasma glucagon levels and severe hypoaminoacidemia (<25% normal). Glucagonomas are rare, with a reported incidence of 1 per 20 to 200 million population. Several reviews of the literature indicate that most tumors occur in the tail of the pancreas; they have an average tumor diameter of 3.6 cm (but as many as one-third are <2 cm), are malignant in approximately two-thirds of cases and have metastases in other organs in 51% to 54% of patients at the time of diagnosis.[117] The DM is characterized by mild hyperglycemia and is nonketotic. The tumors are relatively slow growing, and the 10-year survival rate is 52% in those with metastases and 64% in those without metastases.

The hyperglycemia is owing to excess glucose production by the liver. The hypoaminoacidemia results from an increase in amino acid clearance by the liver. Necrolytic migratory erythema, glossitis, weight loss, and anemia are a consequence of protein malnutrition.[118] Deep venous thrombosis not associated with coagulation disorders is common.

All components of the syndrome are improved if the hyperglucagonemia can be reduced. Treatment consists of surgical removal of the tumor, followed by hepatic artery embolization if necessary for liver metastases. The tumor cells have abundant somatostatin receptors, and somatostatin analogs are very effective in reducing residual plasma glucagon levels.[117] Cytotoxic agents such as streptozotocin and fluorouracil may be valuable as additional modes of therapy. Zinc and amino acid supplementation have been used to treat the rash but are relatively ineffective if plasma glucagon levels remain very high. Antiplatelet therapy should be used to prevent venous thrombosis. The hyperglycemia, although mild, generally requires treatment with an antihyperglycemic agent. Insulin would appear to be the most appropriate agent, although few or no clinical outcome data are available to support this hypothesis. Treatment of the hyperglucagonemia will ameliorate the hyperglycemia in most cases.

If the proglucagon synthesized by the tumor is defective, or if it is cleaved by prohormone convertase 1 or 3 into intestinal glucagon, GLP-1, or GLP-2, the tumor will be associated with either no hormonal abnormalities or with syndromes related to the active products secreted (hyperinsulinemic hypoglycemia or intestinal symptoms with giant villae), and not with the glucagonoma syndrome.[115]

Somatostatinoma

Case reports of patients with hyperglycemia and a pancreatic tumor containing large quantities of somatostatin first appeared in 1977. One of those patients became euglycemic after complete resection of the tumor. Since that time, it has been recognized that large somatostatin-producing pancreatic tumors may be associated with a clinical syndrome consisting of hyperglycemia, cholelithiasis, steatorrhea, and hypochlorhydria.[119]

Somatostatin-producing tumors arise with approximately equal frequency from the gastrointestinal tract and the pancreatic islets. Duodenal somatostatinomas with and without von Recklinghausen disease are seldom associated with recognizable somatostatinoma syndrome, often contain psammoma bodies, and may be associated with demonstrable metastases at the time of surgery. The clinical features of pancreatic somatostatin-producing tumors are variable, and this variation is related to differences in the quantity and qualitative features of somatostatin variants that are synthesized and secreted by these tumors. Marked differences in the degree to which insulin, glucagon, and GH secretion are affected in various patients with pancreatic somaostatin-producing tumors probably account for some of the variation in the clinical syndrome. Hyperglycemia in patients with pancreatic somatostatinomas can vary from mild to modest hyperglycemia to severe diabetic ketoacidosis (DKA).[119-121]

Somatostatin infusions in humans are associated with a pronounced decrease in bile flow and bile acid secretion and an increase in bile cholesterol saturation. *In vitro*, somatostatin has a direct inhibitory effect on cholecystokinin stimulation of gallbladder contraction. These observations provide a basis for understanding the cholelithiasis and steatorrhea commonly seen with pancreatic somatostatinomas. Additionally, they explain the development of gallstones in 23.5% of acromegalic patients treated with octreotide during the first year of treatment.

The hyperglycemia seen as part of somatostatinoma syndrome most likely is related to suppression of insulin secretion. In some patients, a relative insulin deficiency leads to reduced peripheral glucose utilization without impairing suppression of hepatic glucose production.[122] In more severe suppression of insulin secretion, both features of insulin action are reduced.

Somatostatin-producing tumors are rare, usually asymptomatic or only mildly symptomatic, and frequently remain undiagnosed for many years. The diagnosis frequently is made late, and the prognosis is poor because of extensive metastases. The use of somatostatin receptor scintigraphy improves the ability to detect somatostatinomas earlier. Early diagnosis and surgical removal can lead to cure, but medical treatment has produced questionable results. A recent study suggests that somatostatinomas possess functioning somatostatin receptors, and that octreotide therapy (0.5 mg/day subcutaneously) or other somatostatin analogs can effectively decrease somatostatin production by the tumor and improve diabetes and diarrhea.

Pheochromocytoma

Glucose intolerance occurs in approximately 30% of patients with pheochromocytoma, but overt DM is uncommon.[123] Mechanisms responsible for glucose intolerance include suppression of insulin by α-adrenergic receptor stimulation of β cells; an increase in insulin resistance, probably related to elevated plasma free fatty acid levels; and increased hepatic glucose output as a result of β-adrenergic stimulation of hepatocytes. α-adrenergic receptor blockade improves glucose tolerance and insulin secretion. Removal of the pheochromocytoma restores glucose tolerance to normal in most cases. Treatment of glucose intolerance with antihyperglycemic agents is rarely required.

KEY POINTS

- Excess hormone secretion as occurs in many endocrine diseases will lead to hyperglycemia and an increase in diabetes mellitus (DM). The pathogenesis and dynamics of the hyperglycemia depend on the hormone in excess and the drugs that may be used to treat the excess hormone secretion. Excess growth hormone (GH) and insulin-like growth factor-1 (IGF-1) cause severe insulin resistance and, in patients with limited insulin secretory reserve, cause hyperglycemia and DM. Excess cortisol secretion results in increased glucose production by the liver, peripheral insulin resistance, and some impairment of β cell insulin secretion. Individuals with increased cortisol exposure and normal β cell function will have normal fasting plasma glucose levels but postprandial hyperglycemia. Those with limited β cell reserve will develop DM. The percentage of patients with growth hormone or cortisol excess who develop DM depends on the duration and magnitude of hormone excess. Treatment of the hormone excess will resolve the DM in some patients and ease the management of hyperglycemia in other patients. The use of somatostatin receptor ligands (SRLs) to treat GH excess may improve or worsen hyperglycemia, depending on the competitive effects in reducing GH and IGF-1 versus the reduction in SRL-mediated insulin secretion. Tumors secreting unregulated quantities of glucagon or somatostatin present with unique syndromes, including DM. Tumors secreting catecholamines have minimal effects on glucose metabolism.

Hyperglycemia in Critical Care Patients

The ability of humans to adapt to stress is critical for survival. Stress changes the disposition of energy from an anabolic to a catabolic process to provide the substrates necessary for the individual to combat the stress and activates the inflammatory and immune systems.[124] This is accomplished through activation of the sympathetic nervous system, the release of pituitary hormones, the release of cytokines, and central nervous system activation of recognition and behavioral centers. The shift in energy expenditure results in stress-induced hyperglycemia, utilization of muscle protein for gluconeogenesis, and an increase in adipose tissue lipolysis.[124] Although this metabolic and immunologic activation is beneficial in the short term, an excessive and prolonged response can be detrimental and even fatal.

Critical illness is a life-threatening situation that is generally cared for in an intensive care unit (ICU). Patients who require intensive care for more than 5 days have a 20% risk of death and substantial morbidity. Hyperglycemia is present in almost all patients admitted to an ICU and has been shown to have a predictable effect on morbidity and mortality in these patients.[125] Most patients admitted to an ICU have stress-induced hyperglycemia (no previous history of DM). This hyperglycemia is the result of increased secretion of catecholamines, cortisol, glucagon, and GH in the presence of insulin resistance and impaired insulin secretion. The net effect is an uncontrolled production and release of hepatic glucose, a decrease in glucose uptake by insulin sensitive tissues, and an increased glucose uptake by insulin-independent tissues.[124,125]

In 2001, it was reported that patients who had been admitted to a surgical ICU and treated with intensive insulin therapy (n = 765) reduced mortality during intensive care from 8.0% with conventional treatment (n = 783) to 4.6% ($P < .04$).[126] Conventional insulin therapy consisted of starting a continuous infusion of insulin only if the blood glucose exceeded 215 mg/dL and, if started, maintaining the blood glucose between 180 and 200 mg/dL. Intensive insulin therapy consisted of a continuous infusion of insulin when the blood glucose exceeded 110 mg/dL and maintaining it between 80 and 110 mg/dL. The clinical benefits were seen in those receiving ICU care for greater than 5 days (mortality 20.2% in conventional treatment and 10.6% in intensive treatment). Mortality in those treated in the ICU for 1 to 5 days was no different (1.8% vs. 1.7%). Intensive insulin therapy reduced overall in-hospital mortality by 34%, bloodstream infections by 46%, acute renal failure requiring dialysis or hemofiltration by 41%, the median number of red-cell transfusions by 50%, and critical-illness polyneuropathy by 44%, and patients receiving intensive therapy were less likely to require prolonged mechanical ventilation and intensive care.

In contrast to the initial study in surgical ICU patients, a similar study of intensive insulin therapy in patients in a medical ICU showed that intensive insulin treatment reduced blood glucose levels, but did not significantly reduce in-hospital mortality (40.0% in the conventional-treatment group vs. 37.3% in the intensive-treatment group, P = .33).[127] However, morbidity was significantly reduced by the prevention of newly acquired kidney damage, more rapid weaning from mechanical ventilation, and discharge from the ICU. Among 433 patients who stayed in the ICU for less than 3 days, mortality was greater among those receiving intensive insulin therapy. In contrast, among 767 patients who stayed in the ICU for three or more days, in-hospital mortality in the 386 who received intensive insulin therapy was reduced from 52.5% to 43.0% (P = .009).

The benefit of intensive insulin treatment in ICU patients was not supported by the large multicenter NICE-SUGAR study.[128] Within 24 hours of admission to an ICU, 6304 patients were randomized to intensive glucose control with a target blood glucose range of 81 to 108 mg/dL or conventional glucose control with a target of 180 mg/dL or less. The primary endpoint was death from any cause within 90 days of randomization. A total of 829 patients (27.5%) in the intensive-control group and 751 (24.9%) in the conventional-control group died (odds ratio for intensive control 1.14; 95% confidence interval [CI] 1.02–1.28; P =.02). Severe hypoglycemia (blood glucose level ≤40 mg/dL) was reported in 206 of 3016 patients (6.8%) in the intensive-control group and 15 of 3014 (0.5%) in the conventional-control group ($P <$ .001). In this study, intensive glucose control increased mortality in ICU patients.

There is a marked difference in the design of the multicenter study versus the individual center studies, such as the routes of insulin administration, nutritional support, and glycemic targets in the control group. Some investigators have suggested that the target blood glucose control for ICU patients should be 140 to 180 mg/dL. This minimizes hypoglycemia, yet prevents excessive hyperglycemia.[126,129]

KEY POINTS

- Hyperglycemia is a common clinical issue in severely ill patients. It may represent previously unknown diabetes mellitus or a metabolic response to severe stress. Blood glucose levels that remain above 200 mg/dL are associated with increased morbidity and mortality. Maintaining blood glucose levels between 80 and 120 mg/dL with intensive insulin therapy increases the risk of severe hypoglycemia and does not reduce morbidity and mortality. The general consensus is that severe hyperglycemia in critically ill patients benefit if the hyperglycemia is treated to and maintained at a level of 140 to 180 mg/dL.

DRUGS THAT CAN CAUSE HYPERGLYCEMIA

Blood glucose is regulated by the balance between insulin secretion and insulin action. A drug that destroys β cells or blocks their insulin secretory function will cause hyperglycemia in any individual (Table 35.5). A drug that directly or indirectly increases insulin resistance can cause hyperglycemia only in individuals with β cells that have limited insulin secretory reserve (individuals with a predisposition to type 2 diabetes).

Drugs That Directly Adversely Affect β Cell Function

Pentamidine and pyriminil (Vacor) are substances that resemble streptozotocin and alloxan chemically. Pentamidine is an antiprotozoal agent that is used extensively to treat *Pneumocystis carinii* infection. Pyriminyl is a nitrosourea-derived rodenticide that can be accidentally or intentionally ingested by humans. These agents cause necrosis of β cells, leading initially to hyperinsulinemia and hypoglycemia, followed by permanent hyperglycemia and an insulin-dependent DM.[130-132] In a series of 128 patients with acquired immunodeficiency syndrome (AIDS) who were treated with pentamidine for *P. carinii* pneumonia, severe glucose homeostasis disorders developed in 48 patients (37.5%), hypoglycemia in seven, hypoglycemia and then diabetes in 18, and diabetes alone in 23.[131] Of the 41 patients in whom diabetes developed, 26 required insulin therapy. Risk factors for the development of dysglycemia include higher pentamidine doses, higher plasma creatinine, and more severe anoxia. Whereas most of the dysglycemic patients received parenteral pentamidine, six were treated exclusively with pentamidine aerosols.

Pyriminyl ingestion has been followed by severe insulinopenic DM in numerous instances.[132] Pyriminyl specifically inhibits the nicotinamide adenine dinucleotide–ubiquinone reductase activity of complex 1 in mammalian mitochondria.[133] This inhibition of mitochondrial respiration of NAD-linked substrates appears to be the mechanism by which pyrminyl destroys β cells and causes a form of insulin-dependent DM.

TABLE 35.5 Drugs That Interfere With Insulin Secretion

- Destroy β Cells
  - Pentamidine
  - Pyriminal
- Decrease Calcium Entry
  - Diazoxide–adenosine triphosphate–dependent potassium channel opener
  - Phenytoin (Dilantin)
- Decrease Potassium
  - Thiazides
  - Loop diuretics
- Activate Immune System Against β Cells
  - Immune checkpoint inhibitors
- Block Metabolism of β Cell or Insulin Action Cascade
  - L-Asparaginase
  - Tacrolimus
  - mTOR inhibitors
  - Phosphoinositide 3-Kinase Inhibitors
- Mechanism Unknown
  - β-adrenergic antagonists
  - Cyclosporin
  - Opiates
  - L-Asparaginase

Fluoroquinolones are a class of antimicrobial agents that can cause hypoglycemia or hyperglycemia. Although most of these complications occur in patients with known DM, they have also been observed in nondiabetic individuals. The fluoroquinolones that have been used clinically and for which glycemic complication data have been examined are gatifloxacin, levofloxacin, monifloxicin, and ciprofloxacin. In a review of patients who were being treated at hospitals in Ontario, Canada for dysglycemia occurring after antimicrobial therapy, 788 had hypoglycemia, and 470 had hyperglycemia during a 2-year period. Of those patients, gatifloxacin had an adjusted odds ratio for causing hypoglycemia of 4.3 and for causing hyperglycemia of 16.7 compared with macrolides or other nonfluoroquinolone antibiotics.[134] A retrospective analysis of a cohort of patients treated from October 1, 2000 through September 30, 2005 examined the crude incidence rates (cases per 1000 patients) for severe hypoglycemia and hyperglycemia in those treated with gatifloxacin (0.35 and 0.45), levofloxacin (0.19 and 0.18), ciprofloxacin (0.10 and 0.12), or azithromycin (0.07 and 0.10).[134,135] The odds ratios for hypoglycemia in diabetic patients compared with azithromycin were gatifloxacin 4.3, levofloxacin 2.1, and ciprofloxacin 1.1. The odds ratios for hyperglycemia were gatifloxicin 4.5, levofloxacin 1.8, and ciprofloxacin 1.0. In another study, dysglycemia (serum glucose above 200 mg/dL or below 50 mg/dL) was assessed within 72 hours of receiving fluroquinolones or ceftriaxone in 17,108 patients.[136] Dysglycemia rates were 1.01% for gatifloxacin, 0.93% for levofloxacin, 0.18% for ceftriaxone, and 0% for ciprofloxacin. Of the 101 patients with dysglycemia, hyperglycemia occurred in 92 and hypoglycemia in nine. The data show that the dysglycemic effect among individual fluoroquinolones varies greatly.

The fluoroquinolones can either acutely increase insulin secretion causing hypoglycemia or decrease insulin secretion causing hyperglycemia. *In vitro* studies with gatifloxacin have shown that it stimulates insulin release by blocking the Kir 6.2 subunit of the β cell adenosine triphosphate–dependent potassium (K_{ATP}) channel.[137] More chronic treatment with gatifloxicin inhibits insulin biosynthesis, thereby decreasing β cell insulin content and secretion. The predominant clinical effect will depend on the baseline status of the β cell at the time of fluoroquinolone treatment.[138]

Drugs That Inhibit Increases in β Cell Cytosolic Calcium

An increase in the β cell cytosolic calcium concentration is the major mechanism responsible for insulin secretion. Calcium ion entry into the β cell is controlled by several calcium ion channels. A voltage-dependent L-type calcium channel has its activity linked to a K_{ATP} channel. The K_{ATP} channel is closed by increases in plasma glucose. Drugs that keep the K_{ATP} channel open, such as diazoxide, block glucose-mediated insulin secretion and lead to hyperglycemia.[139]

Phenytoin (Dilantin) and other phenylhydantoins can interfere with sodium, potassium, and calcium ion transport. Dilantin administration has been reported to increase plasma glucose and decrease plasma insulin.[140,141] Several studies indicate that it interferes with calcium ion entry into the β cell by a process different from the K_{ATP} channel–related mechanism. Dilantin has been reported to cause hyperosmolar nonketotic hyperglycemia or new-onset DM. This complication occurs infrequently, and probably in individuals with some preceding genetic susceptibility. Currently used calcium channel blockers do not affect glucose metabolism. Confirmatory data come from a reanalysis of the data from the NAVIGATOR study in which calcium channel blockers were started in 1171 patients with IGT who were treated for a median of 5 years, and no increase in new-onset diabetes was noted.[142]

Drugs That Cause Potassium Depletion

Numerous studies have shown that diuretics can cause a deterioration in glucose tolerance in nondiabetic individuals and worsening of glycemic control in patients with type 2 diabetes.[143,144] The effect appears to be dose-related and is either absent or markedly reduced with low-dose diuretic therapy.[144,145] The Atherosclerosis Risk in Communities (ARIC) study compared 458 hypertensive patients treated with a thiazide diuretic for 6 years versus a control hypertensive cohort on no medications and found that the rates of development of new DM were no different.[145] In contrast, the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), which randomized 15,255 hypertensive patients to 12.5 to 25 mg/day of a diuretic, found that 11.6% of patients developed new-onset diabetes over 4 years of treatment.[14]. In contrast, the rate for patients randomized to a calcium channel blocker was 9.8%, and the rate for patients randomized to an angiotensin-converting enzyme inhibitor was 8.1%. In humans, diuretics appear to worsen glycemia primarily by inhibiting insulin secretion, although they also have a modest effect in increasing insulin resistance. Several studies have shown that the decreased insulin secretory response is attributed to intracellular potassium depletion and can be restored toward normal with potassium repletion.[147,148] In the ALLHAT study, 8.5% of the cohort randomized to the diuretic had a serum potassium less than 3.5 mEq/L at 4 years of treatment, but only 1.9% of the cohort randomized to the calcium channel blocker had this degree of hypokalemia.[146] Diabetogenic effects are seen with most diuretics. Therefore, if diuretics are used in diabetic or prediabetic patients to control blood pressure, doses should be restricted to that equivalent to 12.5 to 25.0 mg hydrochlorothiazide, and potassium supplements should be prescribed as needed.

Antiimmune and Cancer-Treatment Drugs

Cyclosporin and Tacrolimus. Cyclosporine and tacrolimus treatment for immunosuppression in renal transplant recipients is associated with an increased incidence of DM. Direct inhibitory effects of cyclosporine on β cells *in vitro* have been associated with decreased insulin secretion.[149] Although it is likely that cyclosporine has direct effects in causing deterioration in glucose tolerance in renal transplant patients, it is not possible to exclude a contributory role of the associated corticosteroid therapy.[150] The onset of DM in cyclosporine-treated patients occurs within the first several months of treatment and often requires insulin treatment. The reported incidence of DM and glucose intolerance in cyclosporine-treated transplant patients ranges from 13% to 47%.

Tacrolimus has been associated with development of DM and, in a few cases, with DKA.[151] During a 3-year period, 7.1% of patients treated with tacrolimus for interstitial pneumonia secondary to polymyositis and dermatomyositis developed newly diagnosed DM. Tacrolimus-induced DKA has been reported in 15 patients after transplantation. Evidence suggests that tacrolimus suppresses insulin secretion by inhibiting the transcription of the insulin gene because of association between tacrolimus and the FK506-binding protein 12. The diabetogenic effect of tacrolimus is dos- dependent and reversible.

mTOR Inhibitors. mTOR inhibitors are anticancer treatments, and hyperglycemia has been reported in 12% to 50% and hyperlipidemia in 7% to 73% of patients in phase III trials of everolimus and temsirolimus.[152] Severe hyperglycemia (>13.9 mmol/L) is less frequent (4%–12%). DKA has not been reported. In a metaanalysis of 24 clinical trials of everolimus or temsirolimus in 4261 patients with various types of tumors, hyperglycemia and severe hyperglycemia were 2.95 and 5.25 more likely with mTOR inhibitor treatments than with control treatment.[153] A median time of 58 days to maximal hyperglycemia was noted in 75 patients treated by everolimus or temsirolimus. TOR inhibition by rapamycin leads to a 33% reduction in glucose-induced insulin secretion and a 50% reduction in β cell mass by increased apoptosis. Rapamycin also reduces insulin sensitivity at the level of the liver and adipose tissue. The development of hyperglycemia and hyperlipidemia during therapy requires standard treatments and often continuation of the mTOR inhibitor, albeit at a lower dose.

Immune Checkpoint Inhibitors. Immune checkpoint inhibitors (ICPIs) are immunotherapy agents that work primarily by using the immune system to target and eradicate cancer cells. Currently, ICPIs include three drug classes of monoclonal antibodies named after their targets: anti–cytotoxic T-lymphocyte 4 antigen antibody (anti-CTLA-4 Ab), programmed cell death-1 inhibitors (PD-1 inhibitors), and programmed cell death ligand-1 inhibitors (PDL-1 inhibitors). Examples of drugs in these categories are ipilimumab and tremelimumab (anti-CTLA-4 Ab), nivolumab and pembrolizumab (PD-1 inhibitors), and atezolizumab and durvalumab (PDL-1 inhibitors). ICPIs have undesired off-target effects, collectively referred to as immune-related adverse events (irAEs). Overall, irAEs are seen in approximately 50% to 60% of patients treated with ICPIs. Endocrinopathies reported in clinical trials include lymphocytic hypophysitis (0%–17%), Grave disease and thyroiditis (6%–20%), and, less commonly, primary adrenal insufficiency and type 1 diabetes. PD-1 inhibitor treatments (nivolumab or pembrolizumab) have caused hyperglycemia within days to weeks after the first dose.[154,155] These patients may even present with DKA.[155] Of five hyperglycemic PD-1 inhibitor–treated patients, three had DKA, with undetectable C-peptide levels and positive glutamic acid decarboxylase 65 antibodies resembling the findings of classic type 1 diabetes.

Tyrosine Kinase Inhibitors. Tyrosine kinase activates a family of enzymes involved in cell proliferation, cellular metabolism, and angiogenesis. Blocking these pathways with tyrosine kinase inhibitors (TKIs) is a key component for antiangiogenic therapies. Examples of these drugs, which have been used to treat a variety of cancers, include sorafenib, sunitinib, pazopanib, axitinib, vandetanib, lenvatinib, erlotinib, gefitinib, and regorafenib. Tyrosine kinase activation is a vital component for activation of the insulin signaling cascade, and it might be expected that TKI would increase insulin resistance. However, in patients with preexisting diabetes receiving TKIs, mean blood glucose concentrations declined by 9 to 53 mg/dL compared with normal controls. In one study involving TKI treatment, eight of the 17 patients with diabetes were able to discontinue their glucose-lowering medications, including insulin. Although the mechanism of glucose lowering is not understood, it appears that TKI treatment increases β cell mass and function.

Phosphoinositide 3-Kinase Inhibitors. Phosphoinositide 3-kinase (PI3K) activation mediates intracellular insulin action, as well as that of other growth regulators like epidermal growth factors, IGFs, and vascular endothelial growth factors, making it a key regulator of cell cycle checkpoints affecting apoptosis, proliferation, and growth. There are three classes of PI3K pathway inhibitors: pan-PI3K inhibitors, isoform-specific PI3K inhibitors, and pan–protein kinase B (Akt) inhibitors. Alterations of the PI3K/Akt pathways are common in breast and other cancers. Blocking PI3K interferes with insulin signaling and results in insulin resistance.[156,157] There are many PI3K inhibitor drugs, each of which is specific for its class of inhibitor. Hyperglycemia, which typically appears within the first two cycles of treatment, has been reported in the range of 51% to 65% of patients taking alpelisib with endocrine therapy in clinical trials.[156] Hyperosmolar and ketoacidosis states are extremely rare but may occur in patients with preexisting diabetes. Once hyperglycemia is detected, mild increases in FPG (160 mg/dL) do not require dose modifications of the PI3K inhibitor, but rather initiating metformin 500 mg once daily and increasing each dose by an additional 500 mg as needed up to a maximum dose of 2000 mg daily. For FPGs between 160 and 250 mg/dL, the PI3K pathway inhibitor can be maintained at the same dose, and the metformin dose increased. If the metformin dose has been maximized, and the FPG has not improved, an insulin-sensitizing agent should be added. If hyperglycemia does not improve within 21 days to FPG less than 160 mg/dL, the patient should restart alpelisib at a lower dose. In patients who develop FPG greater than 250 mg/dL, treatment with alpelisib should be interrupted for 1 to 2 days until blood glucose improves.

L-Asparaginase. L-asparaginase is a chemotherapeutic agent for management of acute lymphoblastic leukemia and other hematopoietic malignancies. The antineoplastic activity results from cleavage of L-asparagine to aspartic acid and ammonia and depletion of the L-asparagine circulating pool, which in turn inhibits protein synthesis, causes cell cycle arrest in the G1-phase, and ultimately causes apoptosis in susceptible leukemic cells. Unlike normal cells, leukemic cells and other cancer cells have little or no asparagine synthetase. L-asparaginase treatment is associated with reduced insulin synthesis and secretion and possibly a decrease in the expression of insulin receptors.[158,159] L-asparaginase–associated hyperglycemia has been reported in 4% to 20% of pediatric patients receiving *Escherichia coli* L-asparaginase and in 4% to 17% of patients receiving *Erwinia chrysanthemi* asparaginase.[158,159] Hyperglycemia is more common during phases of therapy when L-asparaginase and corticosteroids are administered together and in relatively higher doses. The L-asparaginase preparation and the type of corticosteroid (prednisone or dexamethasone) used in treatment do not seem to influence the risk of hyperglycemia. Hyperglycemia tends to resolve within 12 days after the last dose. More severe cases and DKA can occur in patients with other risk factors for DM. Insulin therapy may be required in severe cases; however, L-asparaginase therapy can continue if the patient shows glucose levels of 200 mg/dL or less with insulin treatment.

KEY POINTS

- Drugs that are toxic to β cells and cause their deaths, interfere with the insulin secretory processes in the β cell, or block the intracellular action of insulin cause hyperglycemia. If they cause destruction of β cells, such as occurs with pentamidine or fluroquinolones, the clinical picture will be initial hypoglycemia followed later by permanent diabetes mellitus. Checkpoint inhibitors cause autoimmune destruction of β cells and a type 1 diabetes–like syndrome. If drugs interfere with insulin secretion by blocking calcium entry or potassium flux in the β cell, the hyperglycemia is reversible by discontinuing the drug or adding a drug to treat the hyperglycemia. The same is true for drugs that block the intracellular insulin action pathway (tacrolimus, mTOR inhibitors, phosphoinostide 3-kinase inhibitors, L-asparaginase).

Mechanisms Unknown. β-adrenergic antagonists have been shown to worsen glycemic control in type 2 diabetic patients and to impair glucose tolerance or even precipitate type 2 diabetes in nondiabetic patients.[160] A 6-year follow-up of hypertensive patients treated with β-adrenergic receptor antagonists in the ARIC study showed a 28% higher risk for development of diabetes than that in hypertensive patients not undergoing drug therapy.[145] The deleterious effect of β-adrenergic blockade on glucose tolerance is worse with nonspecific β-adrenergic receptor blockade than with specific β_2-adrenergic antagonists.[160,161] Combination therapy with β-adrenergic antagonists and diuretics has an additive effect of worsening glucose tolerance. Among 19,257 hypertensive patients randomized in the Anglo-Scandinavian Cardiac Outcome Trial Blood Pressure Lowering Arm, baseline FPG greater than 5 mmol/L, BMI, and the use of atenolol were among the highest risk factors for the development of new-onset DM.[162] β-adrenergic antagonists impair glucose tolerance and worsen hyperglycemia by blocking nutrient-mediated insulin secretion.[163] β_2-adrenergic receptor activation stimulates, and β_2-adrenergic receptor blockade inhibits, insulin secretion. Drugs with both α- and β-adrenergic receptor antagonist activity have no effect or may slightly improve glucose metabolism.

Drugs That Cause Insulin Resistance

Drugs that increase insulin resistance will not affect glucose metabolism in individuals with normal β cell function. However, when administered to individuals with limited β cell reserve, they will cause glucose intolerance or overt DM. Agents that cause an increase in body weight, and particularly those associated with an increase in central obesity, will cause insulin resistance. Drugs or hormones that elevate plasma free fatty acid levels will lead to insulin resistance. Counterregulatory hormones or drugs that raise circulating levels of counterregulatory hormones will cause impairment of insulin action. The most used agents that can increase insulin resistance are glucocorticoids, estrogens, progestogens, and nicotinic acid. Hyperglycemia associated with glucocorticoid therapy has been discussed in the section on Cushing syndrome.

Oral Contraceptives and Sex Hormones

Older studies investigating the effects of oral contraceptives, estrogens, and progestogens showed that deterioration of glucose tolerance and development of type 2 diabetes were occasional complications of long-term therapy.[164] Most studies attributed the diabetogenic effects of these steroids to an increase in insulin resistance. Recent studies with lower-dose natural estrogens (ethinyl-estradiol lower than 35 mcg) combined with progestogens indicate that these regimens produce little or no increase in glucose intolerance, even though they may cause a small increase in insulin resistance.[165] Progesterone-only contraceptives have weak effects on glucose and lipid metabolism that vary depending on the specific molecules employed.

Nicotinic Acid

Nicotinic acid is used extensively to treat mixed hyperlipidemias and has been shown to decrease morbidity and mortality from cardiovascular disease. Short-term administration of nicotinic acid reduces plasma free fatty acid levels by 30% to 40%, but as the drug effect wears off, the plasma free fatty acids rebound to 50% to 100% above baseline concentrations. Long-term administration of nicotinic acid at 1 to 4.5 g/day causes severe insulin resistance, presumably as the result of elevated plasma free fatty acid levels. Normal individuals have a compensatory rise in insulin secretion, and glucose tolerance remains normal. In individuals with diminished insulin secretory reserve, nicotinic acid causes glucose intolerance and type 2 diabetes.[166] In type 2 diabetic patients, nicotinic acid administration results in marked deterioration in glycemic control.[167,168] Newer extended-release niacin preparations at doses of 1000 and 1500 mg/day have been shown to be very effective in raising plasma high-density lipoprotein (HDL) cholesterol and lowering plasma triglyceride levels with markedly reduced flushing and hepatotoxicity. The Assessment of Diabetes Control and Evaluation of the Efficacy of Niaspan Trial, which evaluated the effect of extended-release niacin in type 2 diabetic patients, showed no effect on glycemic control at 1000 mg/day and only a minor increase in HbA1c (+0.29%) over 16 weeks of therapy.[168] These data indicate that the advantages of extended-release niacin in improving diabetic dyslipidemia far outweigh any detrimental effects on glycemic control.

Two large clinical trials have studied the effects of adding extended-release niacin to statin therapy to increase cardiovascular benefits in high-risk individuals. The AIM-HIGH study failed to show an incremental cardiovascular benefit when extended-release niacin was administered to patients with atherosclerotic heart disease and low-density lipoprotein (LDL) cholesterol levels less than 1.81 mmol/L. Patients who had DM had a modest increase in their hyperglycemia.[169] Extended-release niacin caused an increase in those with normal glucose tolerance progressing to IGT (58.6% vs. 41.5% of controls) but no increase in new-onset DM.[168] In the Heart Protection Study 2-Treatment of HDL to Reduce the Incidence of Vascular Events study, niacin–laropiprant, as compared with placebo, was associated with a 55% proportional increase in disturbances in diabetes control(11.1% vs. 7.5%, $P < .001$) most of which led to hospitalization.[170] In an analysis of the 17,374 participants who did not have diabetes at the time of randomization, assignment to niacin–laropiprant, as compared with assignment to placebo, was associated with a 32% proportional increase in the diagnosis of diabetes (5.7% vs. 4.3%, $P < 0.001$) [170].

LIPID-LOWERING AGENTS AND NEW-ONSET DIABETES MELLITUS

Statins

Multiple clinical studies show that statin therapy is associated with an increase in the development of type 2 diabetes and a worsening of glycemic control in patients who have type 2 diabetes. In a collaborative metaanalysis of 13 randomized statin trials involving 91,140 participants, 4278 developed diabetes during a mean of 4 years.[171] Of the patients who developed diabetes, 2236 had been assigned to statin therapy and 2032 to control treatment. Statin therapy was associated with a 9% increase in diabetes. Treatment of 255 individuals with statin therapy for 4 years resulted in one extra new case of diabetes.

The benefits of statin therapy in lowering cardiovascular events and death have been assessed in patients at high to low risk for cardiovascular disease. Overall statin therapy shows that, for each 1-mmol/L decrease in LDL cholesterol, there is a 10% proportional reduction in all-cause mortality, a 20% reduction in deaths from coronary heart disease, and a 22% reduction in major vascular events. Even in individuals with a 5-year risk of major vascular events lower than 10%, each 1-mmol/L reduction in LDL cholesterol produces an absolute reduction of major vascular events of approximately 11/1000 over 5 years.[172]

A detailed analysis of the cardiovascular benefits and diabetes risks of statin therapy in primary prevention was carried out for the 17,603 participants in the JUPITER trial.[173] Patients were classified into those with one or more major risk factors for diabetes at baseline (metabolic syndrome, impaired fasting glucose, BMI ≥30 kg/m^2, or HbA1c ≥6%; n = 11,508) and those with no major risk factors (n = 6095). In individuals with one or more major risk factors for diabetes at baseline, statin therapy was associated with a 39% reduction in the cardiovascular endpoint, a 36% reduction in venous thromboembolism, a 17% reduction in total mortality, and a 28% increase in DM. In trial participants with no major diabetes risk factors, statin treatment was associated with a 52% reduction in the primary cardiovascular endpoint, a 53% reduction in venous thromboembolism, a 22% reduction in all-cause mortality, and no increase in DM.

All data indicate that the benefits of statin therapy in reducing cardiovascular events far exceed the risk for statin-induced acceleration of the onset of type 2 diabetes in both primary prevention and secondary treatment.[174,175]

Proprotein Convertase Subtilisin/Kexin Type 9 Inhibitors

The proprotein convertase subtilisin/kexin type 9 (PCSK9) modulates cholesterol homeostasis by increasing the endosomal and lysosomal cleavage of hepatic LDL receptor, which increases serum LDL cholesterol levels. Inhibition of PCSK9 dramatically lowers circulating LDL cholesterol. Several monoclonal antibodies that inhibit PSCK9 are available and have been shown to lower LDL cholesterol. Evolocumab, when added to patients on statin therapy with a median LDL cholesterol of 92 mg/dL, decreased it to a median of 30 mg/dL (59%).

Multiple studies and metaanalyses show that PCSK9 inhibitors do not alter glucose metabolism and do not increase new-onset DM. These data prove that statin-induced increases in new-onset DM are not related to LDL cholesterol lowering.[176]

Cholesterol Ester Transfer Protein Inhibitors

A metaanalysis of four cholesterol ester transfer protein (CETP) inhibitor trials involving a total of 73,479 patients, including 960 newly diagnosed cases of diabetes in the CETP inhibitor group versus 1086 in the placebo group, showed a significant (12%) reduction in the incidence of DM (odds ratio 0.88; 95% CI: 0.81–0.96; P = .005).[177] A statistically nonsignificant trend (Z = –1.13, P = .26) was noted for an increase in HDL cholesterol and the inhibition of CETP. In contrast to the cholesterol-lowering agents, CETP inhibitors reduced the incidence of new-onset DM. The role of the increase in HDL cholesterol in the improvement in glucose metabolism is unclear.

KEY POINTS

- Nicotinamide causes insulin resistance and increases hyperglycemia and the prevalence of new-onset diabetes mellitus (DM) in patients with reduced insulin secretory capacity. Statin therapy is associated with an increase in new-onset DM or worsening of glycemia in patients with DM. The mechanism by which statins cause deterioration of glucose metabolism is multifactorial and includes an increase in insulin resistance, as well as modification of β cell insulin secretion. Individuals at risk for statin-induced DM are those who previously had known risk factors for DM. Analyses have demonstrated that the cardiovascular benefits of statins greatly exceed the detrimental effects of the associated new-onset DM. Proprotein convertase subtilisin/kexin type 9 inhibitor treatment of hypercholesterolemia has no increased risk of DM, indicating that the statin-induced DM is independent of its cholesterol-lowering activity. Clinical trials have shown that inhibitors of cholesterol ester transfer protein decrease the development of new-onset DM.

ACQUIRED IMMUNODEFICIENCY SYNDROME AND ANTIRETROVIRAL DRUGS

Human immunodeficiency virus (HIV) has infected 76 million people, 33 million of whom have died since the epidemic began in the 1980s. Initially, AIDS was a rapidly progressive disease culminating in death within several years. Antiretroviral drugs (protease inhibitors [PIs], nucleoside reverse-transcriptase inhibitors [NRTIs], nonnucleotide reverse-transcriptase inhibitors [NNRTIs], and, more recently, integrase inhibitors) were developed that, when administered in various combinations, suppress viral replication to levels that are undetectable in the peripheral circulation and improve immune function.[178] Consequently, AIDS has gone from being an acute fatal disease to a chronic illness that, with proper management, can be associated with a life expectancy several years shy of normal. Globally, 38 million people (an estimated 0.7% of those aged 15–49 years worldwide) were living with HIV at the end of 2019. The burden of the epidemic continues to vary considerably between countries and regions. Africa remains most severely affected, with nearly 1 in every 25 adults in this region living with AIDS.

Antiretroviral drugs (PIs, NRTIs, and NNRTIs) are associated with the development of lipodystrophies, dyslipidemia, and glucose intolerance.[178] HIV/AIDS itself is associated with increased plasma triglycerides and decreased plasma total, LDL, and HDL cholesterol levels.[179] The PIs indinavir, nelfinavir, ritonavir, and saquinavir were early antiretroviral agents, and their use was reported to cause an unusual lipodystrophy. The main clinical features of this lipodystrophy are peripheral lipoatrophy of the face, limbs, and buttocks and central fat accumulation, including over the dorsocervical spine.[180] Patients with HIV who were newly exposed to antiretroviral therapy (ART) were found to have prevalence of HIV lipodystrophy between 20% and 35% after 12 to 18 months of therapy.[180] In addition to lipodystrophy, ART has been associated with the development of new hyperglycemia; hypercholesterolemia including elevated LDL cholesterol, very low-density lipoprotein, and ApoB levels; low HDL cholesterol; and hypertriglyceridemia.[180-185] Newer antiretroviral agents cause less peripheral lipodystrophy but still increase central fat, insulin resistance, and the metabolic syndrome. The dyslipidemia occurs within 3 months of initiation of ART therapy and plateaus within 6 to 9 months.[179] The prevalence of new-onset hyperglycemia with ART has been reported to vary from 1% to approximately 8%. In the Multicenter AIDS Cohort Study, the incidence of type 2 DM (14%) was more than four times greater in HIV-infected men receiving ART than in HIV-seronegative controls.[185] Tsiodras and coworkers presented data from a 5-year cohort study of 221 HIV-infected patients treated with PIs showing that the cumulative incidences of new-onset hyperglycemia, hypercholesterolemia, hypertriglyceridemia, and lipodystrophy were 5%, 24%, 19%, and 13%, respectively.[186] The prevalence and incidence of hyperglycemia in HIV patients treated with ART are influenced by concomitant hepatitis C virus infection. Retrospective analysis of a cohort of 1230 persons on their first ART regimen revealed a prevalence of hyperglycemia in hepatitis C–coinfected patients of 5.9% versus 3.3% in uninfected patients.[187] The mechanism responsible for the protease-induced hyperglycemia appears to be the development of peripheral insulin resistance in skeletal muscle and adipose tissue through inhibition of GLUT-4 translocation.[188] Pancreatic β cell function is also impaired, so it cannot compensate for the insulin resistance.[188,189] The thymidine analog NRTIs induce mitochondrial dysfunction.[178] At the present time, it is unclear how ART causes lipodystrophy and other related metabolic side effects. The data indicate that ART-mediated metabolic side effects are not related to virologic suppression, CD4 cell count, or changes in weight. Dube and colleagues reported seven new cases of diabetes occurring in a population of 1050 patients treated with PIs over a 9-month period, for a rate of less than 1%.[182] Two patients were treated successfully with insulin and two with sulfonylureas. Two patients had resolution of their hyperglycemia by stopping indinavir therapy. Kilby and Tabereaux determined the frequency of severe hyperglycemia in a university clinic for HIV-1–infected patients and found a prevalence of less than 2%.[183] Preexisting diabetes was present in 12 of 1392 adults, and new cases of hyperglycemia occurred in 13 of these 1392 adults.[184] Most of the incident cases could be attributed to megestrol or corticosteroid treatment. All HIV ART series show a much lower prevalence of hyperglycemia than of dyslipidemia or lipodystrophy.

The use of HbA1c measurements for the diagnosis or for following treatment for type 2 diabetes is confounded. In one study, HbA1c underestimated average blood glucose by 1.7 mmol/L (equivalent to 1% HbA1c).[188] This was not confirmed in a study of HIV-infected women.[178,190] Factors associated with this discordance between actual and predicted HbA1c include higher mean corpuscular volume, NRTI use, and abacivir use. Additional studies are required to clarify the validity of HbA1c measurements in HIV/AIDS patients on ART.

The therapy of type 2 diabetes in HIV/AIDS patients on ART must consider drug interactions. Antidiabetic agents that are metabolized by the cytochrome P-450 3A4, such as sitagliptin, should be avoided, as their use will interfere with the metabolism of some of the antiretroviral agents such as ritonavir.[178]

ATYPICAL ANTIPSYCHOTIC AGENTS

The atypical antipsychotic drugs represent several chemical classes of drugs that have been shown to have beneficial effects in patients with schizophrenia and manic-depressive disorder. They differ from older agents in that they are associated with significantly fewer extrapyramidal side effects and provide superior beneficial effects. They bind to a variety of serotonin, dopamine, and histamine receptors.[191] Atypical antipsychotic drugs include clozapine, olanzapine, risperidone, quetiapine, ziprasidone, aripiprazole, and amisulpride.

Shortly after their introduction for the treatment of schizophrenia, case reports appeared documenting the development of new cases of DM during therapy with clozapine and olanzapine.[192] Because clozapine and olanzapine cause considerable weight gain,[193] the question was posed as to whether the weight gain might be responsible for the development of diabetes. It also was speculated that schizophrenia itself is associated with an increased prevalence of type 2 diabetes. A metaanalysis of healthy volunteers who received atypical antipsychotics during clinical trials evaluated their effects on insulin sensitivity with or without weight gain and with treatment length. The data showed that atypical antipsychotics decrease insulin sensitivity (standardized mean difference = −0.437, $P < .001$).[194] It was found that decreases in insulin sensitivity were potentially dependent on treatment length and not on weight gain. Decreases in insulin sensitivity occurred in less than 13 days, whereas weight gain occurred in studies that lasted 14 days and longer (maximum 28 days). These findings indicated that atypical antipsychotics cause insulin resistance and weight gain directly, independent of psychiatric disease, and may be associated with length of treatment.[194]

Subsequently, numerous reports, large healthcare databases, and reviews have appeared in the literature citing new cases of diabetes and hypertriglyceridemia as complications of therapy with clozapine and olanzapine.[191,195] Clozapine and olanzapine treatment are associated with a mean weight gain of approximately 10 kg over the first year of treatment. Risperidone and quetiapine cause weight gain of approximately 2 to 3 kg. In contrast, ziprasidone and aripiprazole treatment cause minor amounts of weight gain (0.5–1.0 kg over 1 year). The amount of weight gain caused by olanzapine and clozapine varies considerably among patients. In a US study of 474 adults treated for 24 months, 23.2% gained no weight, 47.5% gained up to 10 kg, 20.8% gained between 10 and 20 kg, and 8.6% gained more than 20 kg.[195] Clozapine and olanzapine treatment have been shown in several studies to cause insulin resistance.[196] Ziprasidone and aripiprazole did not appear to cause insulin resistance. The data on risperidone and quetiapine were not enough to confirm, but these agents appeared to have little effect.

In one long-term study of clozapine treatment of patients with schizophrenia, the 5-year cumulative prevalence of DM was 36%, and in another with a mean follow-up of 12.3 years, the rate of newly detected diabetes was 22.3%.[197,198] Published reports and Med-Watch data from the US Food and Drug Administration (FDA) have been reviewed by Koller et al.,[199,200] who found new-onset cases of diabetes in 188 patients on olanzapine, 132 on risperidone, and 242 on clozapine. Approximately 40% of those cases presented as DKA. Some 70% of patients developed diabetes within 6 months of starting olanzapine. When olanzapine was discontinued, 76% of patients exhibited improvement or remission of their diabetes. Among 10 patients who had had a remission and were rechallenged with olanzapine, diabetes recurred in 80%. Although the development of new-onset diabetes was frequently associated with significant weight gain, this was not always the case. In a 9-day treatment of normal individuals, olanzapine was shown to increase insulin resistance and increase meal-mediated secretion of insulin, GLP-1, and glucagon in the absence of weight gain.[201] In contrast, aripiprazole increased insulin resistance but had no effect on postprandial hormone secretions. Ziprasidone and aripiprazole appear to have significantly fewer metabolic side effects.

The mechanism by which atypical antipsychotics cause diabetes is unclear. A literature search and metaanalysis published in 2017 [202] identified 83 psychiatric patients who presented with DKA while taking atypical antipsychotic drugs: olanzapine 32, clozapine 19, risperidone nine, quetiapine eight, aripiprazole six, and two atypical antipsychotics nine. The DKA patients were more likely to be male and younger, and a greater proportion of them were not obese, compared with patients who develop type 2 diabetes. The development of DKA in so many patients taking atypical antipsychotics indicates that these drugs significantly impair β cell insulin secretory function, as well as impairing insulin action.

In a metaanalysis involving 23,799 patients with schizophrenia and related diseases, metabolic syndrome was found in 51.9% of patients on clozapine and 28.2% of those on olanzapine.[203] Nineteen percent of the pooled patients had hyperglycemia (plasma glucose >100 mg/dL), and 39.3% had hypertriglyceridemia. In an analysis of a database of 345,937 patients treated with an antipsychotic drug and 1,426,488 unexposed controls, the risk for developing DM was 1.3-fold greater with second-generation antipsychotic drugs compared with treatment with first-generation drugs.[204]

In addition to diabetes, clozapine and olanzapine cause hypertriglyceridemia in many patients.

Because no randomized, controlled comparative studies have investigated the effects of the different atypical antipsychotic agents on the development of diabetes, current conclusions must be viewed with some reservations. The FDA has taken the position that all atypical antipsychotic agents should be viewed as creating a risk for new-onset diabetes. A joint consensus conference sponsored by the American Diabetes Association and the American Psychiatric Society came to a different conclusion.[205] Their conclusions can be summarized as follows: the prevalence of the three adverse conditions (obesity, diabetes, dyslipidemia) differs among the second-generation antipsychotics:

1. Clozapine and olanzapine are associated with the greatest weight gain and with diabetes and dyslipidemia.
2. Risperidone and quetiapine appear to have intermediate effects.

3. Aripiprazole and ziprasidone are associated with little or no significant weight gain, diabetes, or dyslipidemia but have less usage exposure.

The consensus panel made the following recommendations:

1. If a patient gains 5% of his or her initial weight during therapy, one should consider switching the specific antipsychotic agent.
2. For patients who develop worsening glycemia or dyslipidemia while on antipsychotic therapy, the panel recommends considering switching to an agent that has not been associated with weight gain or diabetes.

Unfortunately, little progress has been made in further defining the risks or the mechanisms of the metabolic abnormalities associated with the different atypical antipsychotics. A large, randomized, initially double-blind, controlled clinical trial comparing the effectiveness and safety of olanzapine, risperidone, quetiapine (a first-generation antipsychotic), perphenazine, and ziprasidone was carried out in 1493 patients from January 2001 through December 2004 under the auspices of the National Institute of Mental Health; it was called the Clinical Antipsychotic Trial of Intervention Effectiveness.[206] The first phase was projected to last 18 months; however, 74.5% of patients discontinued their medications before the end of the first phase. The results suggested that olanzapine was somewhat more effective in controlling the symptoms of schizophrenia; however, its use was associated with a mean weight gain of 9.4 kg, an increase in plasma glucose of 15 mg/dL, an increase in HbA1c of 0.41%, an increase in plasma cholesterol of 9.7 mg/dL, and an increase in plasma triglyceride of 42.9 mg/dL. Quetiapine and risperidone showed minimal increases compared with the first-generation agent, and ziprasidone caused no weight, glycemic, or lipid effects. Because of the high rate of discontinuation of initial medications, the validity of the data must be interpreted with caution. Scheen and De Hert performed a prospective, open, nonrandomized study comparing the metabolic effects of six atypical antipsychotics in 238 schizophrenic patients who underwent a fasting analysis and an OGTT at baseline and after 3 months of treatment.[207] Weight gain, glucose alterations, incidence of new diabetes, and development of metabolic syndrome were greatest with clozapine, followed by olanzapine, then quetiapine, risperidone, and amisulpride. Aripiprazole was not associated with weight gain, improved glucose tolerance, and reduced the occurrence of metabolic syndrome. Thus, the newer data, although not ideal, confirm the conclusions derived from previous reports.

Another 6-month multicenter, randomized, double-blind trial involving 59 patients attempted to determine the mechanisms by which olanzapine and risperidone cause their metabolic effects.[208] Patients had baseline anthropomorphic measurements and measurement of total body fat by dual-energy x-ray absorptiometry analysis, visceral fat by abdominal computed tomography, and insulin sensitivity and disposition index by frequently sampled intravenous glucose tolerance using the Bergman technique. Patients were randomized to olanzapine or risperidone, and the studies were repeated at 6 months. Both drugs induced weight gain and increased total body and visceral fat. Insulin sensitivity did not deteriorate, nor was a significant change noted in disposition index (a composite measure of changes in insulin sensitivity and insulin secretion). A subset analysis of Black American and Hispanic subjects suggested that they had greater susceptibility to the detrimental effects of olanzapine in increasing insulin resistance and decreasing pancreatic compensation to insulin resistance.

It seems fair to conclude that weight gain plays some role in the development of the metabolic effects of atypical antipsychotic agents. However, other unknown factors also contribute, and the factors contributing to β cell failure are still unknown. Clozapine and olanzapine are the worst offenders in causing metabolic side effects, quetiapine and risperidone cause significantly fewer metabolic alterations, and ziprasidone and aripiprazole seem to produce minimal metabolic side effects. Metformin has been suggested as a treatment option for atypical antipsychotic–induced DM; however, many of these patients will require insulin. Vitamin D has been suggested as an agent that may prevent the atypical antipsychotic–mediated hyperglycemia.[209]

KEY POINTS

- Both the antiretroviral agents developed to treat acquired immunodeficiency syndrome and the atypical antipsychotic drugs used to treat schizophrenia have been shown to cause dramatic disturbances in lipid and glucose metabolism. The antiretroviral agents cause an unusual lipodystrophy, hypertriglyceridemia, and new-onset diabetes mellitus (DM). The lipid and adipose tissue effects are most prevalent and severe, whereas the effects on glucose metabolism are more modest and variable. The reported increase in DM varies from 1% to as high as 12%. The mechanism responsible for the metabolic effects are unclear, but the drugs do cause an increase in insulin resistance. The metabolic effects of the atypical antipsychotics clozapine and olanzapine are characterized by striking weight gain, a prevalence of DM of 22%–36%, and a significant incidence of diabetic ketoacidosis. Although the marked weight gain causes insulin resistance, which would increase DM because of limited insulin reserve, the development of ketoacidosis in many treated patients indicate some impairment of β cell function.

CONCLUSION

Hyperglycemia is a frequent complication of many diseases and medications that interfere with either insulin secretion or insulin action. Hyperglycemia decreases physiologic responses, such that defense against infections, metabolic adaptation, and protein synthesis are impaired. The consequences are that hyperglycemia increases the complications of the associated illness or the therapy and, in many instances, decreases survival. The recognition of the potential for hyperglycemia should alert the healthcare provider to undertake preventive measures or, in the case of advent of hyperglycemia, immediate and effective therapy.

REFERENCES

1. Ewald N, Kaufmann C, Raspe A, et al. Prevalence of diabetes mellitus secondary to pancreatic diseases (type 3c). *Diabetes Metab Res Rev*. 2012;28:338–342.
2. Wynne K, Devereaux B, Dornhorst A. Diseases of the exocrine pancreas. *J Gastroenterol Hepatol*. 2019;34:346–354.
3. Barnes AJ, Bloom SR. Pancreatectomized man: a model for diabetes without glucagon. *Lancet*. 1976;1:219–221.
4. Pezzilli R. Commentary—diabetic control after pancreatectomy. *Dig Liver Dis*. 2006;38:420–422.
5. Kendall DM, Sutherland DER, Najarian JS, et al. Effects of hemipancreatectomy on insulin secretion and glucose tolerance in healthy humans. *N Engl J Med*. 1990;322:898–903.
6. Robertson RP, Lanz KJ, Sutherland DE, et al. Relationship between diabetes and obesity 9 to 18 years after hemipancreatectomy and transplantation in donors and recipients. *Transplantation*. 2002;73:736–741.
7. Kumar AF, Gruessner RWG, Seaquist ER. Risk of glucose intolerance and diabetes in hemipancreatectomized donors selected for normal preoperative glucose metabolism. *Diabetes Care*. 2008;31:1639–1643.
8. Kirchner VA, Finger EB, Bellin Md, et al. Long-term outcomes for living pancreas donors in the modern era. *Transplantation*. 2016;100:1322–1328.
9. Lam NN, Schnitzler Ma, Seveg DL, et al. Diabetes mellitus in living pancreas donors: use of integrated national registry and pharmacy claims data to characterize donor-related health outcomes. *Transplantation*. 2017;101:1276–1281.
10. Menge BA, Schrader H, Breuer TGK, et al. Metabolic consequences of a 50% partial pancreatectomy in humans. *Diabetologia*. 2009;52:306–317.

11. Menge BA, Breuer TGK, Ritter PR, et al. Long-term recovery of β-cell function after partial pancreatectomy in humans. *Metab Clin Exper*. 2012;61:620–624.
12. Sun AM, Codding JA, Haist RE. A study of glucose tolerance and insulin response in partially depancreatectomized dogs. *Diabetes*. 1974;23:424–432.
13. Barnes AJ, Bloom SR, George K, et al. Ketoacidosis in pancreatectomized man. *N Engl J Med*. 1977;296:1250–1253.
14. Barnes AJ, Bloom SR, Mashiter K, et al. Persistent metabolic abnormalities in diabetes in the absence of glucagon. *Diabetologia*. 1977;13:71–75.
15. Slezak LA, Andersen DK. Pancreatic resection: effects on glucose metabolism. *World J Surg*. 2001;25:452–460.
16. Jethwa P, Sodergren M, Lala A, et al. Diabetic control after total pancreatectomy. *Dig Liver Dis*. 2006;38:415–419.
17. Jin SM, Oh SH, Kim SK, et al. Diabetes-free survival in patients who underwent islet autotransplantation after 50 to 60% distal partial pancreatectomy for benign pancreatic tumors. *Transplantation*. 2013;95:1396–1403.
18. Lankisch PG, Apte M, Banks PA. Acute pancreatitis. *Lancet*. 2015;386:85–96.
19. van Dijk SM, Hallensleben NDL, van Santvoort HC, et al. Acute pancreatitis: recent advances through clinical trials. *Gut*. 2017;66:2024–2032.
20. Das SLM, Singh PP, Phillips RJ, et al. Newly diagnosed diabetes mellitus after acute pancreatitis: a systemic review and meta-analysis. *Gut*. 2014;63:818–831.
21. Tu J, Zhang J, Ke L. Endocrine and exocrine pancreatic insufficiency after acute pancreatitis: long-term follow-up study. *BMC Gastroenterol*. 2017;17:114–122.
22. Vujasinovic M, Tepes B, Makuc J, et al. Pancreatic exocrine insufficiency, diabetes mellitus and serum nutritional markers after acute pancreatitis. *World J Gastroenterol*. 2014;20:18432–18438.
23. Das SLM, Kennedy JI, Murphy R, et al. Relationship between the exocrine and endocrine pancreas after acute pancreatitis. *World J Gastroenterol*. 2014;20:17196–17205.
24. Woodmansey C, McGovern AP, McCollough KA, et al. Incidence, demographics, and clinical characteristics of diabetes of the exocrine pancreas (type 3c): a retrospective cohort study. *Diabetes Care*. 2017;40:1486–1493.
25. Pendharkar SA, Mathew J, Petrov MS. Age- and sex-specific prevalence of diabetes associated with diseases of the exocrine pancreas: a population-based study. *Dig Liver Dis*. 2017;49:540–544.
26. Pendharkar SA, Singh RG, Chand SK, et al. Pro-inflammatory cytokines after an episode of acute pancreatitis: associations with fasting gut hormone profile. *Inflamm Res*. 2018;67:339–350.
27. Singh VK, Yadav D, Garg P. Diagnosis and management of chronic pancreatitis a review. *JAMA*. 2019;322:2422–2434.
28. Majunder S, Chari S. Chronic pancreatitis. *Lancet*. 2016;387:1957–1966.
29. Hart PA, Bellin M, Andersen DK, et al. Type 3c (pancreatogenic) diabetes mellitus secondary to chronic pancreatitis and pancreatic cancer. *Lancet Gastroenterol Hepatol*. 2016;1:226–237.
30. Ewald N, Hardt PD. Diagnosis and treatment of diabetes mellitus in chronic pancreatitis. *World J Gastroenterol*. 2013;19:7276–7281.
31. Witt H, Apte M, Keim V, et al. Chronic pancreatitis: challenges and advances in pathogenesis, genetics, diagnosis and therapy. *Gastroenterology*. 2007;132:1557–1573.
32. Ewald N, Bretzel RG. Diabetes secondary to pancreatic disease (Type 3c)-Are we neglecting an important disease? *Eur J Intern Med*. 2013;24:203–206.
33. Mohan V, Farooq S, Deepa M. Prevalence of fibrocalculous pancreatic diabetes in Chennai in South India. *J Pancreas*. 2008;9:489–492.
34. Sliwinska-Mosson M, Marek G, Milnerowicz H. The role of pancreatic polypeptide in pancreatic disease. *Adv Clin Exp Med*. 2017;26:1447–1455.
35. DelPrato S, Tiengo A. Diabetes secondary to acquired diseases of the pancreas. In: Alberti KGMM, Zimmet P, DeFronzo RA, et al., eds. *International Textbook of Diabetes Mellitus*. 2nd ed. New York: John Wiley & Sons; 1997:189–212.
36. Schrader H, Menge BA, Schneider S, et al. Reduced beta cell volume and beta cell area in patients with chronic pancreatitis. *Gastroenterology*. 2009;136:513–522.
37. Meier JJ, Giese A. Diabetes associated with pancreatic diseases. *Curr Opin Gastroenterol*. 2015;31:400–406.
38. Wynne K, Devereaux B, Dornhorst A. Diabetes of the exocrine pancreas. *J Gastroenterol and Hepatol*. 2019;34:346–354.
39. Larsen S, Hilsted J, Tronier B, et al. Metabolic control and β cell function in patients with insulin-dependent diabetes mellitus secondary to chronic pancreatitis. *Metabolism*. 1987;36:964–967.
40. Pan J, Xin L, Wang D, et al. Risk factors for diabetes mellitus in chronic pancreatitis. A cohort of 2011 patients. *Medicine*. 2016;95:e3251.
41. Hardt PD, Killinger A, Nalop J, et al. Chronic pancreatitis and diabetes mellitus: a retrospective analysis of 156 ERCP investigations in patients with insulin-dependent and non-insulin-dependent diabetes mellitus. *Pancreatology*. 2002;2:30–33.
42. Gaia E, Salacone P. Medical complications of pancreatic resections: conference report. *J Pancreas*. 2007;8(suppl 1):114–117.
43. Muller MW, Friess H, Leitzbach S, et al. Perioperative and follow-up results after central pancreatic head resection (Berne technique) in a consecutive series of patients with chronic pancreatitis. *Am J Surg*. 2008;196:364–372.
44. Gruessner RWG, Sutherland DER, Drangstveit MB, et al. Pancreas allotransplants in patients with a previous total pancreatectomy for chronic pancreatitis. *J Am Coll Surg*. 2008;206:458–465.
45. Sutherland DER, Radesevich DM, Bellin MD, et al. Total pancreatectomy and islet autotransplantation for chronic pancreatitis. *J Am Coll Surg*. 2012;214:409–426.
46. Khadka R, Tian W, Hao X, et al. Risk factor, early diagnosis and overall survival on outcome of association between pancreatic cancer and diabetes mellitus. Changes and advances, a review. *Int J Surg*. 2018;52:342–346.
47. Wolpin BM, Bao Y, Oian ZR, et al. Hyperglycemia, insulin resistance, impaired pancreatic β-cell function, and risk of pancreatic cancer. *J Natl Cancer Inst*. 2013;105:1027–1035.
48. Er K-C, Hsu C-Y, Lee Y-K, et al. Effect of glycemic control on the risk of pancreatic cancer. *Medicine*. 2016;95:24(e3021).
49. Song S, Wang B, Zhang X, et al. Long-term diabetes mellitus is associated with an increased risk of pancreatic cancer: a meta-analysis. *PLOS One*. 2015;10:e0134321.
50. Huang BZ, Pandol SJ, Jeon CY, et al. New-onset diabetes, longitudinal trends in metabolic markers, and risk of pancreatic cancer in a heterogeneous population. *Clin Gastroenterol Hepatol*. 2020;18:1812–1821.
51. Danker R, Boker LK, Boffetta P, et al. A historical cohort study on glycemic control and cancer risk among patients with diabetes. *Cancer Epidemiol*. 2018;57:104–109.
52. Aggawal G, Kamada P, Chari ST. Prevalence of diabetes mellitus in pancreatic cancer compared to common cancers. *Pancreas*. 2013;42:198–201.
53. Shama A, Smyrk TC, Levy MJ, et al. Fasting blood glucose levels provide an estimate of duration and progression of pancreatic cancer before diagnosis. *Gastroenterology*. 2018;155:490–500.
54. Wakasugi H, Funakoshi A, Iguchi H. Clinical observations of pancreatic diabetes caused by pancreatic carcinoma, and survival period. *Int J Clin Oncol*. 2001;6:50–54.
55. Gullo L, Pezzilli R, Morselli-Labate AM, et al. Diabetes and the risk of pancreatic cancer. *N Engl J Med*. 1994;331:81–84.
56. Basso D, Plebani M, Fogar P, et al. Beta-cell function in pancreatic adenocarcinoma. *Pancreas*. 1994;9:332–335.
57. Nakamori S, Ishikawa O, Ohigashi H, et al. Increased blood proinsulin and decreased C-peptide levels in patients with pancreatic cancer. *Hepatogastroenterology*. 1999;46:16–24.
58. Schwartz SS, Zeidler A, Moossa AR, et al. A prospective study of glucose intolerance, insulin, C-peptide and glucagon response in patients with pancreatic carcinoma. *Am J Dig Dis*. 1978;23:1107–1114.
59. Ahren B, Andren-Sandberg A. Capacity to secrete islet hormones after subtotal pancreatectomy for pancreatic cancer. *Eur J Surg*. 1993;159:223–227.
60. Perhert J, Ihse I, Jorfeldt L, et al. Improved glucose metabolism after subtotal pancreatectomy for pancreatic cancer. *Br J Surg*. 1993;80:1047–1050.
61. Hayashida CY, Suzuki K, Fujiya H, et al. Morphometrical quantitation of pancreatic endocrine cells in patients with carcinoma of the pancreas. *Tohoku J Exp Med*. 1983;141:311–322.

62. Noy A, Bilezikian JP. Clinical review 63: diabetes and pancreatic cancer: clues to early diagnosis of pancreatic malignancy. *J Clin Endocrinol Metab*. 1994;79:1223–1231.
63. Rosewicz S, Wiedenmann B. Pancreatic carcinoma. *Lancet*. 1997;349:485–489.
64. Cannon RM, LeGrand R, Chagpar RB, et al. Multi-institutional analysis of pancreatic adenocarcinoma demonstrating the effect of diabetes status on survival after resection. *HPB*. 2012;14:228–235.
65. Li D, Yeung S-CJ, Hassan MM, et al. Antidiabetic therapies affect risk of pancreatic cancer. *Gastroenterology*. 2009;137:482–488.
66. Dong Y-W, Shi YQ, He LW, et al. Effects of metformin on survival outcomes of pancreatic carcinoma: a meta-analysis. *Oncotarget*. 2017;8:55478–55488.
67. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomized, open-label trial. *Lancet*. 2009;373:2125–2135.
68. Balzano G, Maffi P, Nano R, et al. Extending indications for islet autotransplantation in pancreatic surgery. *Ann Surg*. 2013;258:210–218.
69. Watanabe T, Minaga K, Kamata K, et al. Mechanistic insights into autoimmune pancreatitis and IgG4-related disease. *Trends Immunol*. 2018;39:874–889.
70. Madhani K, Farrell JJ. Management of autoimmune pancreatitis. *Gastrointestin Endoscopy Clin N Am*. 2018;28:493–519.
71. Oh D, Song TJ, Moon SH, et al. Type 2 autoimmune pancreatitis (idiopathic duct-centric pancreatitis) highlighting patients presenting as clinical acute pancreatitis: a single center experience. *Gut Liver*. 2019;13:461–470.
72. Miyamoto Y, Kamisawa T, Tabata T, et al. Short and long-term outcomes of diabetes mellitus in patients with autoimmune pancreatitis after steroid therapy. *Gut Liver*. 2012;6:501–504.
73. Hirano K, Isogawa A, Tada M, et al. Long-term prognosis of autoimmune pancreatitis in terms of glucose tolerance. *Pancreas*. 2012;41:691–695.
74. Powell LW, Seckington RC, Deugnier Y. Hemochromatosis. *Lancet*. 2016;388:706–716.
75. Sangkhae V, Nemeth E. Regulation of the iron homeostatic hormone hepcidin. *Adv Nutr*. 2017;8:126–136.
76. Barton JC, Acton RT. Diabetes in HFE Hemochromatosis. *Diab Res*. 2017:ID9826930.
77. O'Brien T, Barrett B, Murray DM, et al. Usefulness of biochemical screening of diabetic patients for hemochromatosis. *Diabetes Care*. 1990;13:532–534.
78. O'Sullivan EP, McDermott JH, Murphy MS, et al. Declining prevalence of diabetes mellitus in hereditary haemochromatosis—the results of earlier diagnosis. *Diabetes Res Clin Pract*. 2008;81:316–320.
79. Saudek CD, Hemm RM, Peterson CM. Abnormal glucose tolerance in β-thalassemia major. *Metabolism*. 1977;26:43–52.
80. Isaacson C, Seftel SC, Keeley KJ, et al. Siderosis in the Bantu: the relationship between iron overload and cirrhosis. *J Lab Clin Med*. 1961;58:845–853.
81. Cario H, Holl RW, Debatin KM, et al. Insulin sensitivity and beta-cell secretion in thalassemia major with secondary haemochromatosis: assessment by oral glucose tolerance test. *Eur J Pediatr*. 2003;162:139–146.
82. Jiang R, Manson JE, Meigs JB, et al. Body iron stores in relation to risk of type 2 diabetes in apparently healthy women. *JAMA*. 2004;291:711–717.
83. Arebesola A, Voutilainen S, Virtanen JK, et al. Body iron stores and risk of type 2 diabetes. *Eur J Endocrinol*. 2013;169:247–253.
84. Montonen J, Boeing H, Steffen A, et al. Body iron stores and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *Diabetologia.* 2012; 55:2613–2621.
85. Kelsey R, Manderson-Koivula, McClenahan NH. Cystic fibrosis-related diabetes: pathophysiology and therapeutic challenges. *Clin Med Insights Endocrinol Diabetes*. 2019;12:1–7.
86. Elborn JS. Cystic fibrosis. *Lancet*. 2016;388:2519–2531.
87. Bridges N, Rowe R, Holt RIG. Unique challenges of cystic fibrosis-related diabetes. *Diabet Med*. 2018;35:1181–1188.
88. Kayanl K, Mohammed R, Mohiaddin H. Cystic fibrosis-related diabetes. *Front Endocrinol*. 2018;9:20.
89. Kwong E, Desai S, Chong L, et al. The impact of cystic fibrosis-related diabetes on health on health-related quality of life. *J Cyst Fibos*. 2019;18:734–736.
90. Raffeq MM, Murad HAS. Cystic fibrosis: current therapeutic targets and future approaches. *J Transl Med*. 2017;15:84.
91. Moran A, Diem P, Klein DJ, et al. Pancreatic endocrine function in cystic fibrosis. *J Pediat*. 1991;118:715–723.
92. Lopez-Pacheco M. CFTR modulators: the changing face of cystic fibrosis in the era of precision medicine. *Front Pharmacol*. 2020;10:1662.
93. Konstan MW, McKone EF, Moss RB, et al. Assessment of safety and efficacy of long-term treatment with combination lumacaftor and ivacaftor therapy in patients with cystic fibrosis homozygous for the F508del-CFTR mutation (PROGRESS): a phase 3, extension study. *Lancet Respir Med*. 2017;5:107–118.
94. Dal J, Feldt-Rassmussen U, Andersen M, et al. Acromegaly incidence, prevalence, complications and long-term prognosis: a nationwide cohort study. *Eur J Endocrinol*. 2016;175:181–190.
95. Gadelha MR, Kasuki L, Lim DST, et al. Systemic complications of acromegaly and the impact of the current treatment landscape: an update. *Endocr Rev*. 2019;40:268–332.
96. Abreu A, Tovar AP, Castellanos R, et al. Challenges in the diagnosis and management of acromegaly: a focus on comorbidities. *Pituitary*. 2016;19:448–457.
97. Frara S, Maffezzoni F, Mazziotti G, et al. Current and emerging aspects of diabetes mellitus in acromegaly. *Trends Endocrinol and Metabol*. 2016;27:470–483.
98. Dreval A, Trigolosova IV, Misnikova IV, et al. Prevalence of diabetes mellitus in acromegaly. *Endocr Connect*. 2014;3:93–98.
99. Melmed S, Bronstein MD, Chanson P, et al. A consensus statement on acromegaly therapeutic outcomes. *Natr Rev Endocrinol*. 2018;14:552–561.
100. Child CJ, Zimmermann AZ, Chrousos GP, et al. Safety outcomes during pediatric GH therapy: final results from prospective GeNeSIS observational program. *J Clin Endocrinol Metab*. 2019;104:379–389.
101. Poidvin A, Weill A, Ecosse E, et al. Risk of diabetes treated in early adulthood after growth hormone treatment of short stature in childhood. *J Clin Endocrin Metab*. 2017;102:1291–1298.
102. Hwang JL, Weiss RE. Steroid induced diabetes: a clinical and molecular approach to understanding and treatment. *Diabetes Metab Res Rev*. 2014;30:96–102.
103. Radhakutty A, Burt MG. Critical review of the evidence underlying management of glucocorticoid-induced hyperglycaemia. *Eur Assoc Endocrinol*. 2018;179:R207–R218.
104. Tamez-Perez HE, Quintanilla-Flores DL, Rodriguez-Gutierrez R, et al. Steroid hyperglycemia: Prevalence, early detection and therapeutic recommendations: a narrative review. *World J Diabetes*. 2015;6:1073–1081.
105. Wajngot A, Giacca A, Grill V, et al. The diabetogenic effects of glucocorticoids are more pronounced in low than in high insulin responders. *Proc Natl Acad Sci U S A*. 1992;89:6035–6039.
106. Wallace MD, Metzger NL. Optimizing the treatment of steroid-induced hyperglycemia. *Ann Pharmacother*. 2018;52:86–90.
107. Roberts A, James J, Dhatariya K. Management of hyperglycaemia and steroid (glucocorticoid) therapy: a guideline from the joint British Diabetes Societies (JBDS) for inpatient care group. *Diabet Med*. 2018;35:1011–1018.
108. Pivonello R, Deleo M, Cozzolino A, et al. The treatment of Cushing's disease. *Endocr Rev*. 2015;36:385–486.
109. Feelders RA, Newell-Price J, Pivonello R, et al. Advances in the treatment of Cushing's syndrome. *Lancet Diabetes Endocrinol*. 2019;7:300–312.
110. Pivonello R, Isidori AM, De Martino MC, et al. Complications of Cushing's syndrome: state of the art. *Lancet Diabetes Endocrinol*. 2016;4:611–629.
111. Scaroni C, Zilo M, Foti M, et al. Glucose metabolism abnormalities in Cushing's syndrome from molecular basis to clinical management. *Endocr Rev*. 2017;38:189–219.
112. Baroni MG, Giorgino F, Pezzino V, et al. Italian Society for Study of Diabetes (SID)/Italian Endocrinological Society (SIE) guidelines for the

treatment of hyperglycemias in Cushing's syndrome and acromegaly. *Nutr Metab Cardiovasc Dis*. 2016;26:85–102.

113. Di Dalmazi G, Pasquali R, Beuschlein F, et al. Subclinical hypercortisolism: a state, a syndrome, or a disease. *Eur J Endocrinol*. 2015;173:M61–M71.
114. Steffensen C, Pereira AM, Dekkers OM, et al. Prevalence of hypercortisolism in type 2 diabetes patients: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175:R247–R253.
115. Albrechtsen NJW, Challis BG, Damjanov I, et al. Do glucagonomas always produce glucagon? *Bosn J Basic Med Sci*. 2016;16:1–7.
116. Mallison CN, Bloom SR, Warin AP, et al. A glucagonoma syndrome. *Lancet*. 1974;2:1–5.
117. Eldor R, Glaser B, Fraenkel M, et al. Glucagonoma and glucagonoma syndrome—cumulative experience with an elusive tumour. *Clin Endocrinol*. 2011;74:593–598.
118. Barazzoni R, Zanetti M, Tiengo A, et al. Protein metabolism in glucagonoma. *Diabetologia*. 1999;42:326–329.
119. Krejs GJ, Orci L, Conlon JM, et al. Somatostatin syndrome, biochemical, morphologic and clinical features. *N Engl J Med*. 1979;301:285–292.
120. Willcox PA, Immelman EJ, Barron JL, et al. Pancreatic somatostatinoma: presentation with recurrent episodes of severe hyperglycemia and ketoacidosis. *Q J Med*. 1988;68:559–571.
121. Jackson JA, Raju BU, Fachnie JD, et al. Malignant somatostatinoma presenting with diabetic ketoacidosis. *Clin Endocrinol*. 1987;26:609–621.
122. Lowry SF, Burt ME, Brennan MF. Glucose turnover and gluconeogenesis in a patient with somatostatinoma. *Surgery*. 1981;89:309–313.
123. Stenstrom G, Sjostrom I, Smith U. Diabetes mellitus in pheochromocytoma: fasting blood glucose level before and after surgery in 60 patients with pheochromocytoma. *Acta Endocrinol*. 1984;106:511–515.
124. Preiser JC, Ichai C, Orban JC, et al. Metabolic response to the stress of critical illness. *Br J Anesth*. 2014;113:945–954.
125. Egi M, Furushima N, Makino S, et al. Glycemic control in acute illness. *Korean J Anesthesiol*. 2017;70:591–595.
126. Van den Berghe G, Wouters P, Weekers P, et al. Intensive insulin therapy in critically ill patients. *New Engl J Med*. 2001;345:1359–1367.
127. Van der Berghe G, Wilmer A, Hermans G, et al. Intensive insulin in the medical ICU. *New Engl J Med*. 2006;354:449–461.
128. The NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *New Engl J Med*. 2009;360:1281–1297.
129. Yatabe T, Inoue S, Sakaguchi M, et al. The optimal target for acute glycemic control in critically ill patients: a network meta-analysis. *Intensive Care Med*. 2017;43:16–28.
130. Bouchard P, Sai P, Reach G, et al. Diabetes mellitus following pentamidine-induced hypoglycemia in humans. *Diabetes*. 1982;31:40–45.
131. Assan R, Mayaud C, Perronne C, et al. Pentamidine-induced derangements of glucose homeostasis. *Diabetes Care*. 1995;18:47–55.
132. Karam JH, Lewitt PA, Young CW, et al. Insulinopenic diabetes after rodenticide (Vacor) ingestion: a unique model of acquired diabetes in man. *Diabetes*. 1980;29:971–978.
133. Esposti MD, Ngo A, Myers MA. Inhibition of mitochondrial complex I may account for IDDM induced by intoxication with the rodenticide Vacor. *Diabetes*. 1996;45:1531–1534.
134. Park-Wyllie LY, Juurlink DN, Kopp A, et al. Outpatient gatifloxacin therapy and dysglycemia in older adults. *N Engl J Med*. 2006;354:1352–1361.
135. Aspinall SL, Good CB, Jiang R, et al. Severe dysglycemia with fluoroquinolones: a class effect. *Clin Infect Dis*. 2009;49:402–408.
136. Mohr JF, McKinnon PS, Peymann PJ, et al. A retrospective, comparative evaluation of dysglycemias in hospitalized patients receiving gatifloxicin, levofloxacin, ciprofloxacin or ceftriaxone. *Pharmacotherapy*. 2005;25:1303–1309.
137. Saraya A, Yokokura M, Gonoi T, et al. Effects of fluoroquinolones on insulin secretion and beta cell ATP-sensitive potassium channels. *Eur J Pharmacol*. 2004;497:111–117.
138. Yamada C, Nagashima K, Takahashi A, et al. Gatifloxacin acutely stimulates insulin secretion and chronically suppresses insulin biosynthesis. *Eur J Pharmacol*. 2006;553:67–72.
139. Danforth Jr E. Hyperglycemia after diazoxide. *N Engl J Med*. 1971;285:1487.
140. Fariss BL, Lutcher CL. Diphenylhydantoin-induced hyperglycemia and impaired insulin release: effect of dosage. *Diabetes*. 1971;20:177–181.
141. Malherbe C, Burrill KC, Levin SR, et al. Effect of diphenylhydantoin on insulin secretion in man. *N Engl J Med*. 1972;286:339–342.
142. Shen L, Shah BR, Reyes EM, et al. Role of diuretics, β blockers, and statins in increasing the risk of diabetes in patients with impaired glucose tolerance: reanalysis of data from the NAVIGATOR study. *BMJ*. 2013;347:16745.
143. Murphy MB, Kohner E, Lewis PJ, et al. Glucose intolerance in hypertensive patients treated with diuretics: a fourteen year follow-up. *Lancet*. 1982;2:1293–1295.
144. Donahue R, Abbott R, Wilson P. Effect of diuretics on the development of diabetes mellitus: the Framingham study. *Horm Metab Res*. 1990;22(suppl 1):46–48.
145. Gress TW, Nieto FJ, Shahar E, et al. Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. *N Engl J Med*. 2000;342:905–912.
146. ALLHAT Officers. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic. *JAMA*. 2002;288:2981–2997.
147. Helderman JH, Elahi D, Anderson DK, et al. Prevention of the glucose intolerance of thiazide diuretics by maintenance of body potassium. *Diabetes*. 1983;32:106–111.
148. Shafi T, Appel LJ, Miller ER, et al. Changes in serum potassium mediate thiazide-induced diabetes. *Hypertension*. 2008;52:1022–1029.
149. Bending JJ, Ogg CS, Viberti GC. Diabetogenic effect of cyclosporin. *BMJ*. 1987;294:401–402.
150. Ost L, Tyden G, Fehrman I. Impaired glucose tolerance in cylcosporin-prednisolone–treated renal graft recipients. *Transplantation*. 1988;46:370–372.
151. Maruyama K, Chujo D. Tacrolimus-induced diabetic ketoacidosis with subsequent rapid recovery of endogenous insulin secretion after cessation of tacrolimus. A case report with review of literature. *Medicine*. 2019;98:36.
152. Bouillet B, Buffier P, Smati S, et al. Expert opinion on the metabolic complications of mTOR inhibitors. *Ann Endocrinol (Paris)*. 2018;79:583–590.
153. Sivendran S, Agarwal N, Gartrell B, et al. Metabolic complications with the use of mTOR inhibitors for cancer. *Cancer Treat Rev*. 2014;40:190–196.
154. Stamatouli AM, Quandt Z, Perdigoto AL, et al. Collateral damage: insulin-dependent diabetes induced with checkpoint inhibitors. *Diabetes*. 2018;67:1471–1480.
155. Byun BJ, Braunstein R, Flynn J, et al. Immune checkpoint inhibitor-associated diabetes: a single institution experience. *Diabetes Care*. 2020;43:3106–3109.
156. Nunnery SE, Mayer IA. Management of toxicity to isoform α-specific PI3K inhibitors. *Ann of Oncol*. 2019;30(Suppl 10):x21–x26.
157. Cheson BD, O'Brien S, Ewer MS, et al. Optimal management of adverse events from Copanlisib in the treatment of patients with non-Hodgkin Lymphomas. *Clin Lymphoma Myeloma Leuk*. 2019;19:135–141.
158. Hwangbo Y, Lee EK. Acute hyperglycemia associated with anti-cancer medications. *Endocrino Metab(Seoul)*. 2017;32:23–29.
159. Grimes A, Mohamed A, Sopfe J, et al. Hyperglycemia during childhood cancer therapy: incidence, implications, and impact on outcomes. *J Nat Cancer Inst Monogr*. 2019;54:Igz022.
160. Micossi P, Pollavini G, Piaggi U, et al. Effects of metoprolol and propranolol on glucose metabolism and insulin secretion in diabetes mellitus. *Horm Metab Res*. 1984;16:59–63.
161. Whitcroft I, Wilkinson N, Ranthorne A, et al. Beta-adrenoreceptor antagonist impairs long-term glucose control in hypertensive diabetes: role of beta-adrenoreceptor selectivity and lipid solubility. *Br J Clin Pharm*. 1986;22:236–237.
162. Gupta AK, Dahlof B, Dobson J, et al. Determinants of new-onset diabetes among 19,257 hypertensive patients randomized in the Anglo-Scandinavian Cardiac Outcomes Trial-blood pressure lowering arm and the relative influence of antihypertensive medication. *Diabetes Care*. 2008;31:982–988.

163. Totterman K, Groop L, Groop PH, et al. Effect of beta blocking drugs on beta cell function and insulin sensitivity in hypertensive non-diabetic patients. *Eur J Clin Pharmacol*. 1984;26:13.
164. Perlman JA, Russell-Briefel R, Ezzati T, et al. Oral glucose tolerance and the potency of contraceptive progestins. *J Chronic Dis*. 1985;38:857–864.
165. Gourdy P. Diabetes and oral contraceptives. *Best Pract Res Clin Endocrinol Metab*. 2013;27:67–76.
166. Lithell H, Vessby B, Hellsing K. Changes in glucose tolerance and plasma insulin during lipid-lowering treatment with diet, clofibrate and niceritrol. *Atherosclerosis*. 1982;43:177–184.
167. Garg A, Grundy SM. Nicotinic acid as therapy for dyslipidemia in non–insulin-dependent diabetes. *JAMA*. 1990;264:723–726.
168. Grundy SM, Vega GL, McGovern ME, et al. Efficacy, safety, and tolerability of once-daily niacin for the treatment of dyslipidemia associated with type 2 diabetes. *Arch Intern Med*. 2002;162:1568–1576.
169. Goldberg RB, Bittner VJ, Dunbar RL, et al. Effects of extended-release niacin added to simvastin/ezetimibe on glucose and insulin values in AIM-HIGH. *Amer J Med*. 2016;129:753e13–753e22.
170. HPS-2-Thrive Collaborative Group. Effects of extended release niacin with laropiprant in high risk patients. *New Engl J Med*. 2014;371:203–212.
171. Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomized statin trials. *Lancet*. 2010;375:735–742.
172. Cholesterol Treatment Trialists' (CTT) Collaborators, Mihaylova B, Emberson J, Blackwell L, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet*. 2012;380:581–590.
173. Ridker PM, Pradhan A, MacFadyen, et al. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the Jupiter trial. *Lancet*. 2012;380:565–571.
174. Wang KL, Liu CJ, Chao TF, et al. Statins, risk of diabetes, and implications on outcomes in the general population. *J Am Coll Cardiol*. 2012;60:1231–1238.
175. Huupponen R, Viikari J. Statins and the risk of developing diabetes. *BMJ*. 2013;346:f3156.
176. Khan SU, Rahman H, Okunrintemi V, et al. Association of lowering low-density lipoprotein cholesterol with contemporary lipid lowering therapies and risk of diabetes mellitus: a systemic review and meta-analysis. *J Am Heart Assoc*. 2019;8:e011581.
177. Masson W, Lobo M, Siniawski D, et al. Therapy with cholesterol ester transfer protein (CEPT) inhibitors and diabetes risk. *Diabetes Metab*. 2018;44:508–513.
178. Brown TT, Glesby MJ. Management of the metabolic effects of HIV and HIV drugs. *Nat Rev Endocrinol*. 2012;8:11–21.
179. Souza SJ, Luzia LA, Santos SS, et al. Lipid profile of HIV-infected patients in relation to antiretroviral therapy: a review. *Rev Assoc Med Bras*. 2013;59:186–198.
180. Carr A. HIV lipodystrophy: risk factors, pathogenesis, diagnosis and management. *AIDS*. 2003;17(suppl 1):S141–S148.
181. Tershakovec AM, Frank I, Rader D. HIV-related lipodystrophy and related factors. *Atherosclerosis*. 2004;174:1–10.
182. Dube MP, Johnson DL, Currier JS, et al. Protease inhibitor–associated hyperglycaemia. *Lancet*. 1997;350:713–714.
183. Kilby JM, Tabereaux PB. Severe hyperglycemia in an HIV clinic: preexisting versus drug associated diabetes mellitus. *J Acquir Immune Syndr Human Retrovirol*. 1998;17:46–50.
184. Kaufman MB, Simionatta C. A review of protease inhibitor–induced hyperglycemia. *Pharmacotherapy*. 1999;19:114–117.
185. Gutierrez AD, Balasubramanyam A. Dysregulation of glucose metabolism in HIV patients: epidemiology, mechanisms and management. *Endocrine*. 2012;41:1–10.
186. Tsiodras S, Mantzoros C, Hammer S, et al. Effects of protease inhibitors on hyperglycemia, hyperlipidemia, and lipodystrophy. *Arch Intern Med*. 2000;160:2050–2056.
187. Mehta SH, Moore RD, Thomas DL, et al. The effect of HAART and HCV infection on the development of hyperglycemia among HIV-infected persons. *J Acquir Immune Defic Syndr*. 2003;33:577–584.
188. Woerle HJ, Mariuz PR, Meyer C, et al. Mechanisms for the deterioration in glucose tolerance associated with HIV protease inhibitor regimens. *Diabetes*. 2003;52:918–925.
189. Yarasheki KE, Tebas P, Sigmund C, et al. Insulin resistance in HIV protease-inhibitor-associated diabetes. *J Acquir Immun Defic Synd Hum Retrovirol*. 1999;21:209–216.
190. Kim PS, Woods C, Georgoff P, et al. A1C underestimates glycemia in HIV infection. *Diabetes Care*. 2009;32:1591–1593.
191. Kowalchuk C, Castellani LN, Chintoh A, et al. Antipsychotics and glucose metabolism: how brain and body collide. *Am J Physiol Endocrinol Metab*. 2019;316:E1–E15.
192. Jin H, Meyer JM, Jeste DV. Phenomenology of and risk factors for new-onset diabetes mellitus and diabetic ketoacidosis associated with atypical antipsychotics: an analysis of 45 published cases. *Ann Clin Psychiatry*. 2002;14:59–64.
193. Allison DB, Mentore JL, Heo M, et al. Antipsychotic-induced weight gain: a comprehensive research synthesis. *Am J Psychiatry*. 1999;156:1686–1696.
194. Burghardt KJ, Seyoum B, Mallisho A, et al. Atypical antipsychotics, insulin resistance and weight; a meta-analysis of healthy volunteer studies. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;83:55–63.
195. DeHert M, Detraux J, van Winkel R, et al. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. *Nat Rev Endocrinol*. 2012;8:114–126.
196. Newcomer JW, Haupt DW, Fucetola R, et al. Abnormalities in glucose regulation during antipsychotic treatment of schizophrenia. *Arch Gen Psychiatry*. 2002;59:337–345.
197. Henderson DC, Cagliero E, Gray C, et al. Clozapine, diabetes mellitus, weight gain, and lipid abnormalities: A five-year naturalistic study. *Am J Psychiatry*. 2000;157:975–981.
198. Schulte PFJ, Bocxe JTH, Dodeman HJ, et al. Risk of new-onset diabetes after long-term treatment with clozapine in comparison to other antipsychotics in patients with schizophrenia. *J Clin Psychopharm*. 2016;36:115–119.
199. Koller E, Schneider B, Bennett K, et al. Clozapine-associated diabetes. *Am J Med*. 2001;111:716–723.
200. Koller EA, Doraiswamy PM. Olanzapine-associated diabetes mellitus. *Pharmacotherapy*. 2002;22:841–852.
201. Teff KL, Rickels MR, Grudziak J, et al. Antipsychotic-induced insulin resistance and postprandial hormonal dysregulation independent of weight gain or psychiatric disease. *Diabetes*. 2013;62:3232–3240.
202. Vuk A, Kuzman MR, Baretic M, et al. Diabetic ketoacidosis associated with antipsychotic drugs: case reports and a review of literature. *Psychiatria Daubia*. 2017;29:121–135.
203. Mitchell AJ, Vancampfort D, Sweers K, et al. Prevalence of metabolic syndrome and metabolic abnormalities in schizophrenia and related disorders—a Prisma meta-analysis. *Schizoph Bull*. 2013;39:306–313.
204. Kewssing LV, Thomsen AF, Mogensen UB, et al. Treatment with antipsychotics and the risk of diabetes mellitus in practice. *Br J Psychiatry*. 2010;197:266–272.
205. American Diabetes Association. American Psychiatric Association; American Association of Clinical Endocrinologists; North American Association for the Study of Obesity: Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care*. 2004;27:596–601.
206. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209–1223.
207. Scheen AJ, De Hert MA. Abnormal glucose metabolism in patients treated with antipsychotics. *Diabetes Metab*. 2007;33:169–175.
208. Ader M, Garvey T, Phillips LS, et al. Ethnic heterogeneity in glucoregulatory function during treatment with atypical antipsychotics in patients with schizophrenia. *J Psychiatr Res*. 2008;42:1076–1085.
209. Nagashima T, Shirakawa H, Nakagawa T, et al. Prevention of antipsychotic-induced hyperglycaemic by Vitamin D: a data mining prediction followed by experimental exploration of the molecular mechanism. *Scient Rep*. 2016;6:26375.

36

β Cell Glucose Toxicity and Oxidative Stress

Anath Shalev, Jamie S. Harmon and R. Paul Robertson

OUTLINE

Introduction, 632
Normal Carbohydrate Biochemistry and Abnormal Glucose Metabolism Secondary to Chronic Hyperglycemia With Excess Generation of Reactive Oxygen Species, 632
General Overview of Free Radical Biology and Antioxidants as Defense Mechanisms Against Elevated Cellular Generation of Reactive Oxygen Species, 632
Evaluation of Oxidative Stress and Defective Endogenous Antioxidant Response in β Cells From the Points of View of the Glutathione and Thioredoxin Pathways, 634
Brief Overview of the Contributions of Oxidative Stress to Secondary Complications of Diabetes, 637

INTRODUCTION

This chapter considers mechanisms by which chronic elevations of blood glucose over an extended period can contribute to the deterioration of β cell function observed in type 2 diabetes mellitus (T2DM) and to the secondary complications that accompany both type 1 diabetes mellitus (T1DM) and T2DM. At the center of this information is the hypothesis that chronically elevated cellular levels of glucose and glucose-derived reactive oxygen species (ROS) cause tissue damage, hence the term *glucose toxicity*. This hypothesis recognizes that physiologic levels of ROS are favorable for many biochemical processes but emphasizes that prolonged elevations of ROS cause tissue damage. This hypothesis allows for possible adverse contributions from fatty acids and other lipids, which gives rise to the term *lipotoxicity*. Often the combined term, *glucolipotoxicty,* is used in the scientific literature.[1,2] This chapter will emphasize an evaluation of two important endogenous antioxidant systems designed to regulate the level of ROS in β cells to provide critical defense mechanisms against oxidative stress, the glutathione (Gsh)and thioredoxin (Trx) pathways.

KEY POINTS

- This chapter will emphasize an evaluation of two important endogenous antioxidant systems designed to regulate the level of reactive oxygen species in β cells to provide critical defense mechanisms against oxidative stress, the glutathione and thioredoxin pathways.

NORMAL CARBOHYDRATE BIOCHEMISTRY AND ABNORMAL GLUCOSE METABOLISM SECONDARY TO CHRONIC HYPERGLYCEMIA WITH EXCESS GENERATION OF REACTIVE OXYGEN SPECIES

The preferred biochemical pathway for glucose metabolism is glycolysis and oxidative phosphorylation. High glucose concentrations increase the mitochondrial proton gradient because of overproduction of electron donors by the tricarboxylic acid cycle. Excess glucose increases production of mitochondrial superoxide, a ROS. With prolonged exposure to high glucose levels, oxidative phosphorylation is swamped, and glucose is diverted to other pathways (Fig. 36.1) that independently increase ROS levels.[3] This creates a toxic state for the islet, because excess ROS has adverse consequences for insulin gene expression, insulin content, and insulin secretion.[4-8] The defect in insulin gene expression is caused by the loss of at least two essential transcription factors, Pdx-1 and MafA (Fig. 36.2 and 36.3). The defect in Pdx-1 synthesis is posttranscriptional,[9] whereas the defect in MafA is posttranslational.[10]

The clinical state of diabetes mellitus is often accompanied by elevated levels of blood cholesterol, triglyceride, and free fatty acids, which gave rise to the concept of *lipotoxicity*.[11,12] The adverse effects of high glucose levels converge with independent adverse consequences of lipotoxicity, both of which can cause increased β cell apoptosis.[1,13] Prolonged exposure of β cells to fatty acids has been reported to inhibit insulin gene expression.[11,12] A prominent hypothesis is that in the presence of high glucose and fatty acid concentrations there is an accumulation of cytosolic citrate, the precursor of malonyl-CoA, which inhibits carnitine palmitoyltransferase-1 synthesis. This enzyme is responsible for fatty acid transport into the mitochondrion, and in its absence free fatty acids are shunted toward esterification pathways, where they can increase production of lipids, including ceramide, a proapoptotic substance.[13,15]

GENERAL OVERVIEW OF FREE RADICAL BIOLOGY AND ANTIOXIDANTS AS DEFENSE MECHANISMS AGAINST ELEVATED CELLULAR GENERATION OF REACTIVE OXYGEN SPECIES

Oxygen is the most abundant element in the human body and accounts for approximately 65% of body mass.[16] Carl Wilhelm Sheele, Joseph Priestly, and Antoine-Laurent Lavoisier are credited with the discovery of oxygen.[16] Although essential for living creatures, the metabolism of oxygen gives rise to ROS, which can be a double-edged sword. ROS mediate killing of pathogenic bacteria by phagocytic cells and also act as second messengers for physiologic cellular signal transduction.[16] However, high oxygen levels can lead to excessive levels of oxidizing free radicals. In 1969 McCord and Fridovich reported the seminal observation that a substance, now referred to as superoxide dismutase (SOD), was catalyzing the dismutation of superoxide anion radicals to produce hydrogen peroxide and molecular oxygen.[17] Shortly thereafter, Babior and colleagues reported that one of the primary bactericidal actions of leukocytes was the enzymatic generation of superoxide.[18] Since then research into free radical mechanisms involved in the pathogenesis of a variety of human diseases has proliferated.

The term free radical refers to a chemical species that contains one or more unpaired electrons. ROS is a collective term that includes superoxide, hydrogen peroxide (H_2O_2), and hydroxyl radicals, among others. The term antioxidant is used to refer to any substance that can prevent or repair ROS-induced damage to proteins, lipids, and nucleic acids. Adverse events are collectively referred to as oxidative stress. Common antioxidant systems include (1) SOD and catalase, (2) the Gsh system, and (3) the Trx system.

1. SOD and catalase are considered together because they are the two enzymes responsible for converting superoxide to water. SOD converts superoxide to H_2O_2; catalase converts H_2O_2 to water. The two main species of SOD are Sod-1 and Sod-2. Sod-1 resides in the cytosol, nuclei, and mitochondrial intermembrane space; Sod-2 resides in the mitochondrial matrix. Catalase resides primarily in peroxisomes.
2. The Gsh system includes the reduced form of Gsh and the enzymes gamma-glutamylcysteine ligase and Gsh synthetase (Fig. 36.4). Glutathione reductase regenerates Gsh from Gsh disulfide. Gsh is synthesized from three amino acids: L-cysteine, L-glutamate, and glycine. Gsh can react directly with electrophilic compounds, such as reactive aldehydes, and lipid peroxidation products. Gsh is also the cofactor for glyoxalase-1, which is involved in the detoxification of glycolysis-derived alpha-oxoaldehyde and methylglyoxal, and thereby provides protection against hyperglycemia-induced oxidative stress (Fig. 36.4). Glutathione peroxidase (GPx) comprises a family of enzymes,

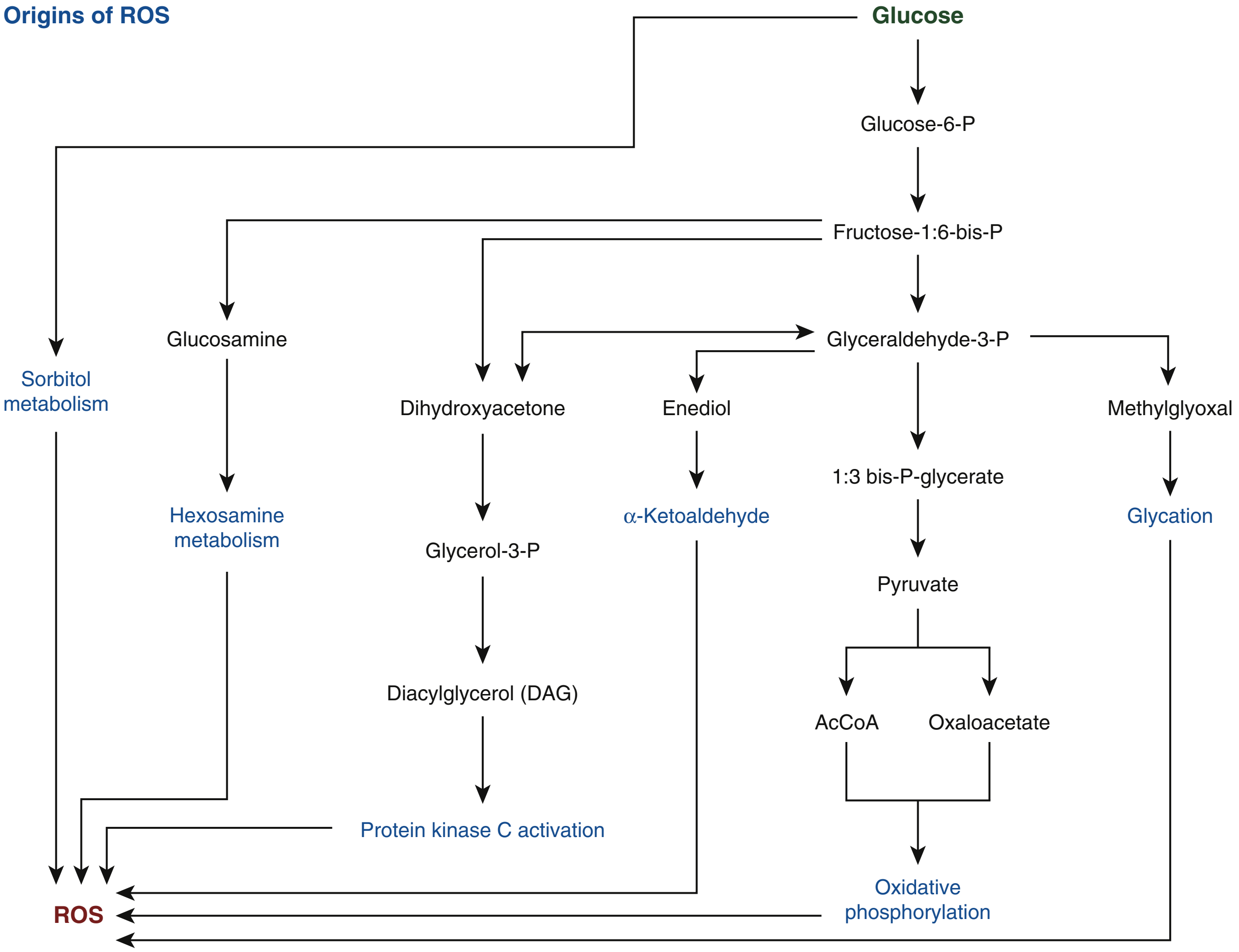


Fig. 36.1 Origins of reactive oxygen species. Six biochemical pathways by which glucose metabolism can form reactive oxygen species (*ROS*) are shown. Under physiologic conditions, glucose primarily undergoes glycolysis and oxidative phosphorylation. Under the pathophysiologic condition of hyperglycemia, excessive glucose levels can swamp the glycolytic process and be shunted to the other pathways, with the outcome of excessive ROS levels. *AcCoA,* Acetyl coenzyme A. (From Robertson RP. Chronic oxidative stress as a central mechanism for glucose toxicity in pancreatic islet beta cells in diabetes. J Biol Chem. 2004;279:42351-42354. Copyright © 2004 American Society for Biochemistry and Molecular Biology. Published by Elsevier Inc. on behalf of American Society for Biochemistry and Molecular Biology. All rights reserved.)

including GPx1, 2, 3, and 4, which are selenoproteins. These enzymes decompose H_2O_2 and organic peroxides, which are among the major ROS that cause oxidative tissue injury. Consequently, these are among the most important antioxidant defenses against stress-associated disease pathology.

3. The Trx system includes Trx (also known as Txn), peroxiredoxin, Trx reductase, and sulfiredoxin. Trxs directly scavenge ROS; act as cofactors for peroxiredoxins, which are important enzymes for the detoxification of H_2O_2 and organic hydroperoxides; and are cofactors for methionine sulfoxide reductases, enzymes involved in protecting against oxidative protein damage and oxidative injury.

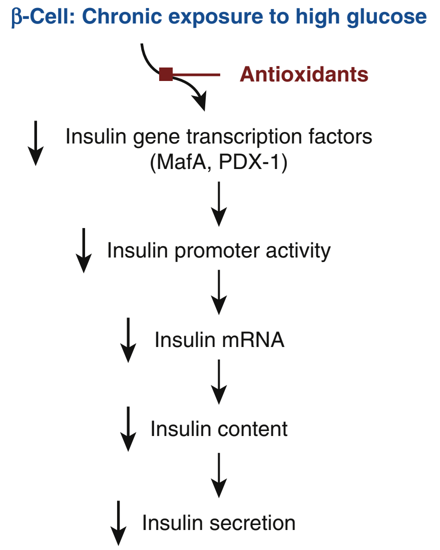

β-Cell: Chronic exposure to high glucose
Antioxidants
↓ Insulin gene transcription factors (MafA, PDX-1)
↓ Insulin promoter activity
↓ Insulin mRNA
↓ Insulin content
↓ Insulin secretion

Fig. 36.2 Adverse effects on insulin synthesis and secretion from chronic exposure of β cells to high glucose and elevated levels of reactive oxygen species. Antioxidant treatment strategies in multiple animal models of diabetes have been shown to prevent these adverse effects on β cell function.

EVALUATION OF OXIDATIVE STRESS AND DEFECTIVE ENDOGENOUS ANTIOXIDANT RESPONSE IN β CELLS FROM THE POINTS OF VIEW OF THE GLUTATHIONE AND THIOREDOXIN PATHWAYS

Glutathione pathway. As with all cells, pancreatic β cells are dependent upon exogenous antioxidants to combat damage caused by endogenous ROS. However, the β cell is not among the cell types that are well-endowed with antioxidant systems, such as those found in the liver and kidneys. β cells have been reported to contain virtually no Gsh peroxidase or catalase and only modest amounts of superoxides dehydrogenases 1 and 2 and hemoxygenase.[19-21] This is a case of double jeopardy, because it is the dysfunction of synthesis and secretion of insulin by β cells that is primarily responsible for hyperglycemia, and yet it is that very defect that leads to chronic hyperglycemia and increased levels of ROS production in β cells as well as other tissues.[22] As such, this is an example of a feed-forward pathogenic mechanism wherein a functional defect in a cell causes an event that goes on to do further damage to the same cell (Fig. 36.5).

Studies have been designed to evaluate whether Gpx-1 enrichment of Gpx-1–deficient β cells may protect them against the adverse effects of hyperglycemia and increased levels of ROS. Beneficial effects were observed of transgenic overexpression of Gpx-1 specifically in β cells in diabetic rodents. The rodents developed increased levels of intranuclear MafA and Gpx-1, increased β cell mass, and normal levels of circulating glucose.[22-24] In other studies, the oral drug ebselen, which is a selenorganic Gpx mimetic, was fed to diabetic rodents. Beneficial effects were observed on β cell mass and islet function despite the presence of continuing hyperglycemia.[25] The importance of this observation derives from the fact that it is a very potent antioxidant that can be given orally and has been shown to be safe and effective in humans.[26-28]

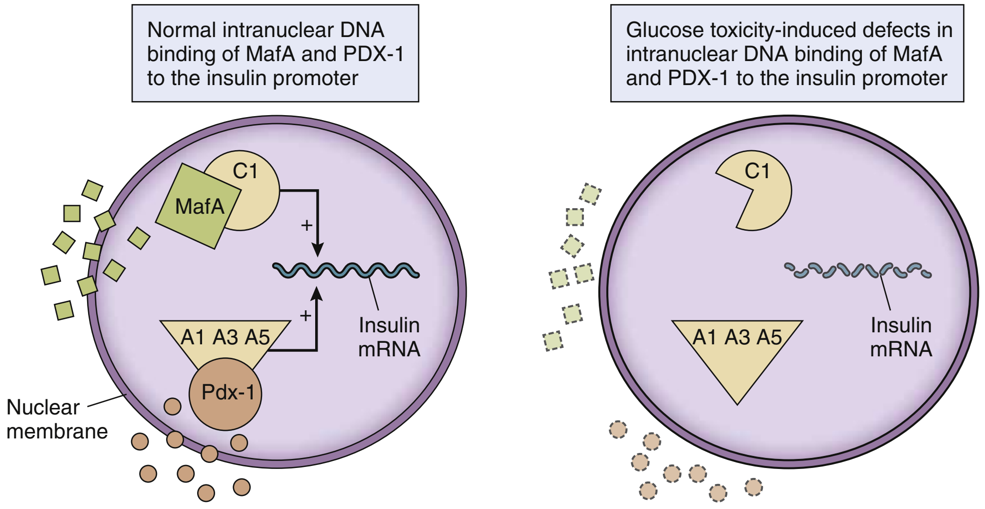

Normal intranuclear DNA binding of MafA and PDX-1 to the insulin promoter
C1
MafA
+
+
Insulin mRNA
A1 A3 A5
Pdx-1
Nuclear membrane
Glucose toxicity-induced defects in intranuclear DNA binding of MafA and PDX-1 to the insulin promoter
C1
Insulin mRNA
A1 A3 A5

Fig. 36.3 Transcription factor binding to the insulin promoter under normal and glucotoxic conditions. Under physiologic conditions, two major regulators of the insulin gene, Pdx-1 and MafA, are activated by rising glucose levels and leave the cytoplasm to enter the nucleus, where they bind to the insulin promoter to increase insulin mRNA levels. This process has been shown to be inhibited by chronic supraphysiologic levels of glucose and reactive oxygen species formation in multiple animal models of type 2 diabetes.

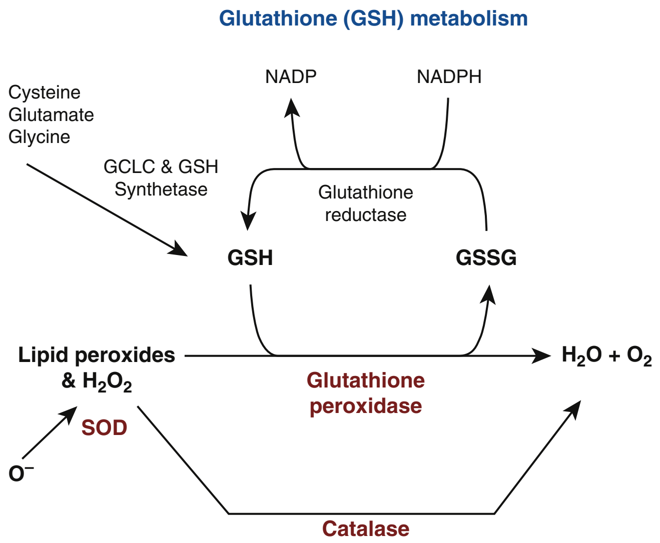

Glutathione (GSH) metabolism
Cysteine
Glutamate
Glycine
GCLC & GSH Synthetase
NADP
NADPH
Glutathione reductase
GSH
GSSG
Lipid peroxides & H_2O_2
$H_2O + O_2$
Glutathione peroxidase
SOD
O^-
Catalase

Fig. 36.4 Glutathione (Gsh) metabolism. Gsh synthesis is regulated by the enzymes glutamate–cysteine ligase catalytic subunit (*GCLC*) and Gsh synthetase, which use the substrates cysteine, glutamate, and glycine. Reduced Gsh is used by Gsh peroxidase to catabolize lipid peroxides and hydrogen peroxidase (*H2O2*) to form water and oxygen. Independently, superoxide ions are catabolized by superoxide dismutase (SOD) to form lipid peroxides and H_2O_2. Catalase independently also catabolizes H_2O_2 to form water and oxygen. Uniquely, the β cell is relatively deficient in both Gsh peroxidase and catalase, and therefore is reliant solely on SOD to catabolize reactive oxygen species. *GSSG*, Glutathione disulfide; *H2O*, water; *NADP*, nicotinamide adenine dinucleotide phosphate radical; *NADPH*, nicotinamide adenine dinucleotide phosphate; *O^-*, oxygen radical; *O_2*, oxygen.

KEY POINTS

- Pancreatic islet β cells are different than most body tissues in that they do not contain substantial levels of glutathione peroxidase or catalase, two essential endogenous antioxidants. This places β cells at heightened risk for damage by the excessive levels of reactive oxygen species caused by chronic hyperglycemia.

Nuclear factor erythroid-derived-2–related factor (Nrf2) upregulates the Kelch-like ECH-associated protein 1/Nrf2/antioxidant pathway, which is known as the master switch for gene expression of antioxidants (Fig. 36.6). The role of Nrf2 in the repair of oxidative stress-induced damage of β cells was examined in male ZDF rats. The rats were studied in the early phase of diabetes caused by eating a high-fat diet for short periods of time followed by 3 weeks' return to regular chow. While eating the high-fat diet the animals developed increased nuclear endogenous Nrf2 and hemoxygenase synthesis, reversal of hyperglycemia, and reversal of abnormal β cell structure in existing β cells[29] (Fig. 36.7). These studies suggested an innate ability for β cells that had undergone oxidative stress and structural damage to undergo self-repair via endogenous antioxidant synthesis. The conclusion for self-repair is strengthened by the fact that no evidence for increased apoptosis or increased β cell replication rates were observed during the short exposure to a high fat diet.

KEY POINTS

- If the adverse effects caused by hyperglycemia are reversed early in the course of developing chronic hyperglycemia, the β cell is capable of mounting an antioxidant defense via activation of cytoplasmic Nrf2 for translocation to the nucleus, where it stimulates genetic production of hemoxygenase.

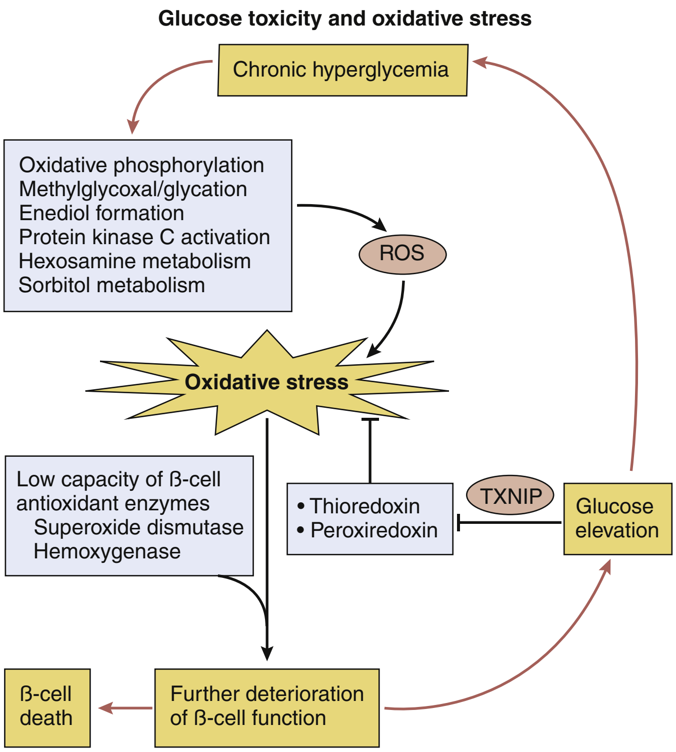

Glucose toxicity and oxidative stress
Chronic hyperglycemia
Oxidative phosphorylation
Methylglycoxal/glycation
Enediol formation
Protein kinase C activation
Hexosamine metabolism
Sorbitol metabolism
ROS
Oxidative stress
Low capacity of ß-cell antioxidant enzymes
Superoxide dismutase
Hemoxygenase
• Thioredoxin
• Peroxiredoxin
TXNIP
Glucose elevation
ß-cell death
Further deterioration of ß-cell function

Fig. 36.5 Glucose toxicity and oxidative stress. Chronic elevation of glucose levels causes increased oxidative phosphorylation, methylglycoxal/glycation, alpha-ketoaldehyde formation, protein kinase C activation, and hexosamine and sorbitol metabolism, resulting in production of reactive oxygen species (*ROS*), which in turn causes oxidative stress on cells. In the case of the β cell this is made worse by a low capacity of antioxidant enzymes, especially catalase and glutathione peroxidase, which leads to further deterioration of β cell function, and ultimately β cell death. This results in additional glucose elevation and increased expression of thioredoxin-interacting protein (*Txnip*), which inhibits thioredoxin function and thereby prevents the thioredoxin system from counteracting the growing oxidative stress the β cell is exposed to, leading to even more damage and worsening of the hyperglycemia.

Thioredoxin pathways. Members of the Trx system are ubiquitously expressed in all cells, including in pancreatic β cells, where they play a particularly important role[30] because of the low levels of other antioxidant systems.[19-21] The six known mammalian peroxiredoxins (Prdx1–6) detoxify H_2O_2 and reduce organic hydroperoxides, primarily by being oxidized at their cysteine residues. This hyperoxidation of the active-site cysteine leads to inactivation of the peroxiredoxins, which then depend upon reactivation by the oxidoreductase sulfiredoxin[31] (Fig. 36.8). Recently, Prdx3 has also been shown to be acetylated in response to high glucose, which adds to the β cell glucose toxicity and revealed yet another level of regulation.[32] Interestingly, there is also crosstalk between the Gsh and Trx systems. For example, sulfiredoxin not only reactivates peroxiredoxin, but also reverses cysteine glutathionylation.[33] In addition, ROS can lead to oxidation of free and protein-bound methionine residues, which impairs protein function. Methionine sulfoxide reductases catalyze the reduction of these residues back to methionine, and Trxs provide the reducing equivalents for this reaction.[34] Trxs also act as electron donors and modulate the activity of a variety of enzymes, including peroxidases and ribonucleotide reductases.[35]

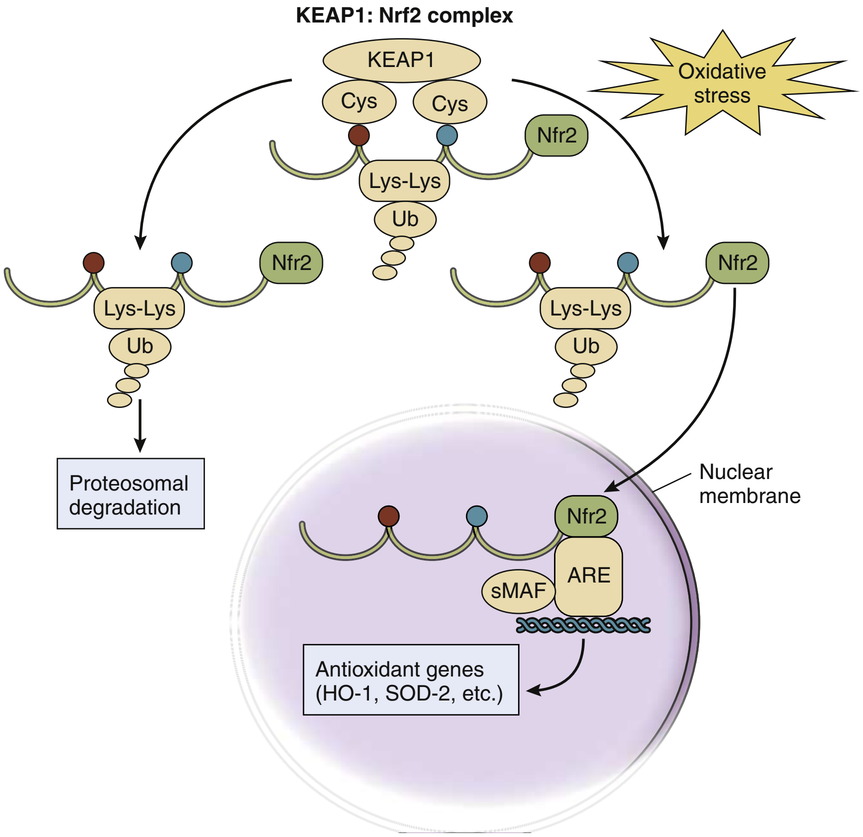


Fig. 36.6 **The Kelch-like ECH-associated protein 1 (*KEAP1*): nuclear factor erythroid-derived-2-related factor (*Nrf2*) complex.** Nrf2 is generally considered to be the master regulator of antioxidant gene transcription. Under quiescent conditions, Nrf2 is bound by KEAP1 in the cytoplasm, which ushers Nrf2 to proteosomes for degradation. However, in the face of oxidative stress, Nrf2 is released from KEAP1 and enters the nucleus, where it serves as a key activator of the promoter for antioxidant genes. *Cys,* Cysteine; *Lys,* lysine; *Ub,* ubiquitin.

However, Trxs primarily function as thioloxidoreductases that protect cells against oxidative stress by reducing oxidized proteins.[36] This cysteine thioldisulfide exchange reaction results in oxidation of the two cysteine residues of Trx. Reduction by Trx reductase returns Trx back to an active state[37,38] (Fig. 36.9). Taking advantage of the disulfide bonds present in the insulin molecule (see Chapter 32), reduction of insulin is being used as the classical test for the activity of Trx.[39] Trxs are highly conserved across species and essential for survival in mammals.[40] In humans there are two Trxs, a major cytosolic form (Trx1) and a mitochondrial form (Trx2), and loss-of-function mutation of either is embryonically lethal.

KEY POINTS

- The thioredoxin system is critical for survival and especially important in pancreatic β cells, where it helps to compensate for the relative lack of other antioxidant mechanisms. Thioredoxins function by reducing oxidized proteins and undergoing oxidation of their cysteines in the process.

Overall, the Trx system has been shown to be involved in multiple cellular processes, including cell proliferation and apoptosis.[37,41-43] Moreover, transgenic mouse studies have further established the critical role that this system plays in β cell biology and protection against diabetes. One of the first examples of this was the observation that β cell–specific overexpression of Trx prevented autoimmune and streptozotocin-induced diabetes in mice.[44]

In addition to their function as antioxidants, Trxs have been shown to interact with and inhibit the activity of apoptosis signal-regulating kinase 1 and thereby to protect cells from apoptosis[45,46] (Fig. 36.10). Most notably, Trxs interact with Trx-interacting protein (Txnip), a member of the arrestin family of proteins. Txnip inhibits the ability of Trxs to reduce oxidized proteins and thereby increases oxidative stress in the cell.[37,47-50] Consistent with this notion, increased Txnip expression has been linked to dysfunction and apoptosis, especially in β cells.[51,52] Txnip overexpression has also been found to lead to increased expression of the microRNA mir-204, which in turn targets the 3’ untranslated region of MafA, leading to decreased expression of this important insulin transcription factor and inhibition of insulin production.[53] Moreover, Txnip has been shown to represent a critical link between glucose toxicity and β cell death.[54] Interestingly, Txnip has been demonstrated to be upregulated in islets from donors with T1DM or T2DM,[55,56] in islets of different mouse models of diabetes, and in isolated islets or β cells exposed to high glucose.[51,54,57] Moreover, Txnip expression is induced by a variety of cellular stresses common in T1DM and T2DM, including endoplasmic reticulum stress and inflammatory cytokines.[58,59] Together this leads to a relentless cycle of oxidative stress, glucose toxicity and β cell damage (Fig. 36.5). In contrast, Txnip inhibition has emerged as an attractive therapeutic approach

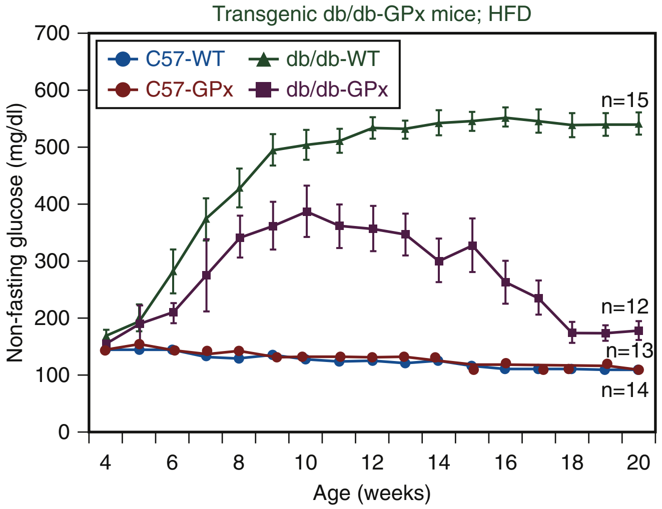


Fig. 36.7 Protection of mouse β cells by the *GPX-1* transgene against the glucotoxic effects of chronic hyperglycemia. Both normoglycemic, wild-type C57BL/6J (*C57-WT*) mice and db/db mice with a type 2 diabetic phenotype fed a high-fat diet (*HFD*) were used for these experiments. Transgenic expression of *GPX-1* had no effect in the C57-WT mice compared to C57-WT mice that did not express the transgene. However, transgenic expression of *GPX-1* initially dampened the onset of hyperglycemia, and with time reversed hyperglycemia, in the db/db mice, whereas the control db/db mice that did not express the transgene developed and maintained severe hyperglycemia. (From Harmon JS, Bogdani M, Parazzoli SD, et al. Beta-cell-specific overexpression of glutathione peroxidase preserves intranuclear mafa and reverses diabetes in db/db mice. *Endocrinology*. 2009;150:4855-4862.)

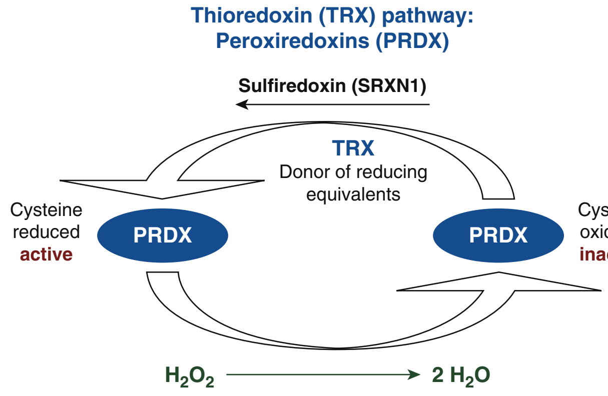

Thioredoxin (TRX) pathway:
Peroxiredoxins (PRDX)
Sulfiredoxin (SRXN1)
TRX
Donor of reducing equivalents
Cysteine reduced active
PRDX
PRDX
Cysteine oxidized inactive
H_2O_2
$2\ H_2O$

Fig. 36.8 Thioredoxin (*Trx*) pathway: peroxiredoxins (*Prdx*). When Prdx reduces hydrogenperoxide to water, its active cysteine residues are being oxidized and rendered inactive. Sulfiredoxin catalyzes the reaction to reduce and reactivate them, whereby Trx provides the reducing equivalents. *H2O,* Water; *H2O2,* hydrogen peroxide.

for T1DM and T2DM, as demonstrated by genetic deletion experiments[60,61] and pharmacological inhibition studies in mice[62,63] and humans.[64]

KEY POINTS

- Diabetes, glucose, and cellular stresses induce thioredoxin-interacting protein (Txnip) expression, and Txnip interaction with thioredoxin inhibits the antioxidative and antiapoptotic effects.

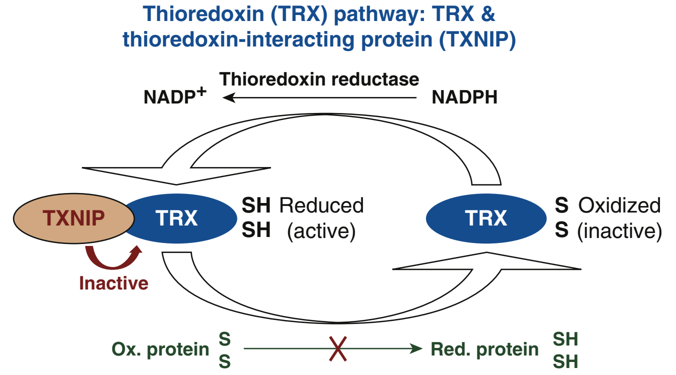

Thioredoxin (TRX) pathway: TRX &
thioredoxin-interacting protein (TXNIP)
Thioredoxin reductase
$NADP^+$
NADPH
TXNIP
TRX
SH Reduced
SH (active)
TRX
S Oxidized
S (inactive)
Inactive
Ox. protein
S
S
Red. protein
SH
SH

Fig. 36.9 Thioredoxin (*Trx*) pathway: Trx and thioredoxin-interacting protein (*Txnip*). Trx helps maintain the cellular redox status by reducing oxidized proteins. In the process, its cysteine residues are being oxidized and rendered inactive. Binding of Txnip to Trx also renders it inactive and unable to reduce oxidized proteins, thereby promoting oxidative stress. Trx reductase reactivates Trx by reversing the oxidation of its cysteines.

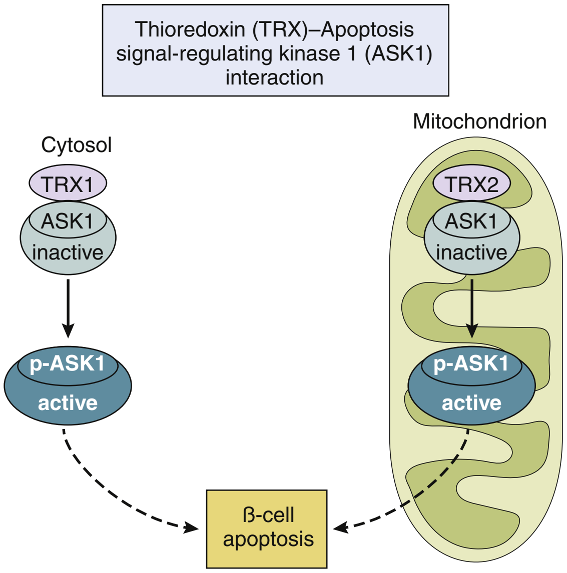

Thioredoxin (TRX)–Apoptosis signal-regulating kinase 1 (ASK1) interaction
Cytosol
Mitochondrion
TRX1
ASK1
inactive
p-ASK1
active
TRX2
ASK1
inactive
p-ASK1
active
ß-cell apoptosis

Fig. 36.10 Thioredoxin (*Trx*)–apoptosis signal-regulating kinase 1 (*Ask1*) interaction. Cytosolic Trx1 and mitochondrial Trx2 bind to Ask1 and thereby prevent its phosphorylation and activation, protecting cells from apoptosis.

BRIEF OVERVIEW OF THE CONTRIBUTIONS OF OXIDATIVE STRESS TO SECONDARY COMPLICATIONS OF DIABETES

Although β cells may be particularly sensitive, other tissues including those affected by secondary diabetes complications are also negatively impacted by glucolipotoxicity, inflammation, and oxidative stress (see Chapter 32). The mechanisms involved are similar, including excess glucose resulting in increased mitochondrial production of ROS in the heart (cardiomyopathy), kidney (nephropathy), eye (retinopathy), and peripheral nervous system (neuropathy), and SOD and catalase as well as the Gsh and Trx regulatory pathways serving as the major defense mechanisms against the ensuing cell dysfunction, damage, and apoptosis.[65,66] (Fig. 36.11). In fact, upregulation of these antioxidant pathways by transgenic overexpression or, more frequently, indirectly by downregulation of Txnip, the natural Trx inhibitor, has been shown

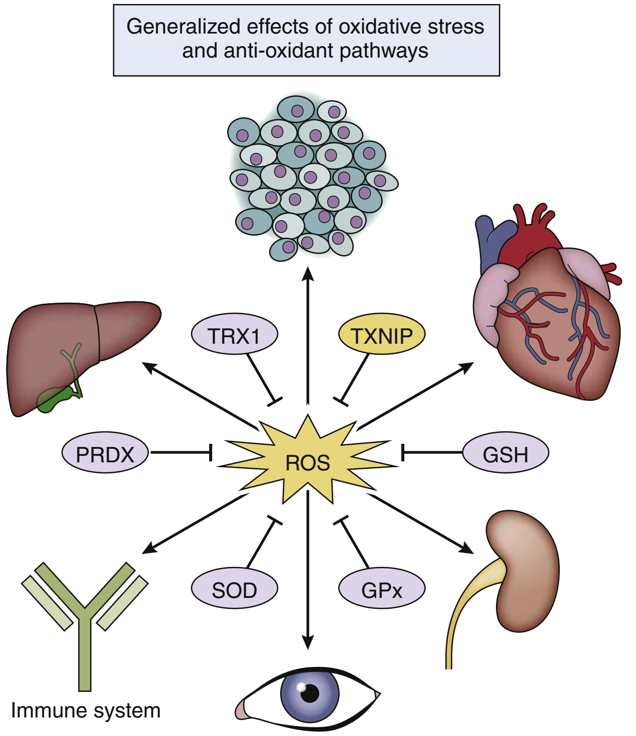

Generalized effects of oxidative stress and anti-oxidant pathways
TRX1
TXNIP
PRDX
ROS
GSH
SOD
GPx
Immune system

Fig. 36.11 **Generalized effects of oxidative stress and antioxidant pathways.** Aside from severe detrimental effects on islet β cells, reactive oxygen species (*ROS*) also impact other organ systems affected by diabetes, including the heart, the kidney, the eye, the immune system, and the liver. There they contribute to secondary diabetes complications such as cardiomyopathy, nephropathy, and retinopathy. The antioxidant glutathione (*GSH, GPx*), superoxide dismutase (*SOD*) and thioredoxin (*TRX, PRDX*) regulatory pathways oppose and protect against these damaging effects.

to result in protective effects. These beneficial effects were found in the context of inflammation in retinal injury,[67-69] vascular oxidative stress, and myocardial ischemia reperfusion,[70,71] as well as mesangial cell and podocyte apoptosis and dysregulation of tubular autophagy in diabetic nephropathy.[72-74] In addition, Txnip has been shown to be involved in peripheral glucose metabolism[63,75] and play a critical role in diet-induced insulin resistance of T2DM.[61,76] Moreover, oxidative stress has recently also emerged as an important factor in the autoimmunity of T1DM,[77] as well as in diabetes-associated metabolic liver disease.[78] Taken together, oxidative stress represents an integral aspect of metabolic disease, diabetes, and pancreatic β cell dysfunction.

KEY POINTS

- Oxidative stress and the systems controlling it not only play a major role in β cell pathobiology, but are also critical in the context of diabetes complications affecting many organ systems.

REFERENCES

1. Poitout V, Robertson RP. Minireview: secondary beta-cell failure in type 2 diabetes--a convergence of glucotoxicity and lipotoxicity. *Endocrinology*. 2002;143:339–342.
2. Poitout V, Robertson RP. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev*. 2008;29:351–366.
3. Wolff SP, Dean RT. Glucose autoxidation and protein modification. The potential role of 'autoxidative glycosylation' in diabetes. *Biochem J*. 1987;245:243–250.
4. Robertson RP, Zhang HJ, Pyzdrowski KL, et al. Preservation of insulin mRNA levels and insulin secretion in hit cells by avoidance of chronic exposure to high glucose concentrations. *J Clin Invest*. 1992;90:320–325.
5. Sharma A, Olson LK, Robertson RP, et al. The reduction of insulin gene transcription in hit-t15 beta cells chronically exposed to high glucose concentration is associated with the loss of ripe3b1 and stf-1 transcription factor expression. *Mol Endocrinol*. 1995;9:1127–1134.
6. Tanaka Y, Gleason CE, Tran PO, et al. Prevention of glucose toxicity in hit-t15 cells and zucker diabetic fatty rats by antioxidants. *Proc Natl Acad Sci U S A*. 1999;96:10857–10862.
7. Robertson RP. Chronic oxidative stress as a central mechanism for glucose toxicity in pancreatic islet beta cells in diabetes. *J Biol Chem*. 2004;279:42351–42354.
8. Kaneto H, Xu G, Fujii N, et al. Involvement of c-Jun N-terminal kinase in oxidative stress-mediated suppression of insulin gene expression. *J Biol Chem*. 2002;277:30010–30018.
9. Olson LK, Redmon JB, Towle HC, et al. Chronic exposure of hit cells to high glucose concentrations paradoxically decreases insulin gene transcription and alters binding of insulin gene regulatory protein. *J Clin Invest*. 1993;92:514–519.
10. Harmon JS, Stein R, Robertson RP. Oxidative stress-mediated, post-translational loss of mafa protein as a contributing mechanism to loss of insulin gene expression in glucotoxic beta cells. *J Biol Chem*. 2005;280:11107–11113.
11. Unger RH. *Diabetes Mellitus: A Fundamental and Clinical Text*. Philadelphia: Lippencott Williams and Wilkins, Inc.; 2004.
12. Prentki M, Corkey BE. Are the beta-cell signaling molecules malonyl-coa and cystolic long-chain acyl-coa implicated in multiple tissue defects of obesity and niddm? *Diabetes*. 1996;45:273–283.
13. Briaud I, Harmon JS, Kelpe CL, et al. Lipotoxicity of the pancreatic beta-cell is associated with glucose-dependent esterification of fatty acids into neutral lipids. *Diabetes*. 2001;50:315–321.
14. Federici M, Hribal M, Perego L, et al. High glucose causes apoptosis in cultured human pancreatic islets of langerhans: a potential role for regulation of specific bcl family genes toward an apoptotic cell death program. *Diabetes*. 2001;50:1290–1301.
15. Kelpe CL, Moore PC, Parazzoli SD, et al. Palmitate inhibition of insulin gene expression is mediated at the transcriptional level via ceramide synthesis. *J Biol Chem*. 2003;278:30015–30021.
16. Hopkins RZ, Li YR. *Essentials of Free Radical Biology and Medicine*. Cell Med Press AIMSCI, Inc; 2017.
17. McCord JM, Fridovich I. Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *J Biol Chem*. 1969;244:6049–6055.
18. Babior BM, Kipnes RS, Curnutte JT. Biological defense mechanisms. The production by leukocytes of superoxide, a potential bactericidal agent. *J Clin Invest*. 1973;52:741–744.
19. Grankvist K, Marklund S, Taljedal IB. Superoxide dismutase is a prophylactic against alloxan diabetes. *Nature*. 1981;294:158–160.
20. Tiedge M, Lortz S, Munday R, et al. Complementary action of antioxidant enzymes in the protection of bioengineered insulin-producing rinm5f cells against the toxicity of reactive oxygen species. *Diabetes*. 1998;47:1578–1585.
21. Tonooka N, Oseid E, Zhou H, et al. Glutathione peroxidase protein expression and activity in human islets isolated for transplantation. *Clin Transpl*. 2007;21:767–772.
22. Robertson RP, Harmon J, Tran PO, et al. Glucose toxicity in beta cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*. 2003;52:581–587.
23. Tanaka Y, Tran PO, Harmon J, et al. A role for glutathione peroxidase in protecting pancreatic beta cells against oxidative stress in a model of glucose toxicity. *Proc Natl Acad Sci U S A*. 2002;99:12363–12368.
24. Harmon JS, Bogdani M, Parazzoli SD, et al. Beta-cell-specific overexpression of glutathione peroxidase preserves intranuclear mafa and reverses diabetes in db/db mice. *Endocrinology*. 2009;150:4855–4862.
25. Mahadevan J, Parazzoli S, Oseid E, et al. Ebselen treatment prevents islet apoptosis, maintains intranuclear pdx-1 and mafa levels, and preserves beta-cell mass and function in zdf rats. *Diabetes*. 2013;62:3582–3588.
26. Yamaguchi T, Sano K, Takakura K, et al. Ebselen in acute ischemic stroke: a placebo-controlled, double-blind clinical trial. Ebselen study group. *Stroke*. 1998;29:12–17.

27. Ogawa A, Yoshimoto T, Kikuchi H, et al. Ebselen in acute middle cerebral artery occlusion: a placebo-controlled, double-blind clinical trial. *Cerebrovasc Dis*. 1999;9:112–118.
28. Saito I, Asano T, Sano K, et al. Neuroprotective effect of an antioxidant, ebselen, in patients with delayed neurological deficits after aneurysmal subarachnoid hemorrhage. *Neurosurgery*. 1998;42:269–277; discussion 277–268.
29. Abebe T, Mahadevan J, Bogachus L, et al. Nrf2/antioxidant pathway mediates beta cell self-repair after damage by high-fat diet-induced oxidative stress. *JCI Insight*. 2017;2:e92854.
30. Stancill JS, Broniowska KA, Oleson BJ, et al. Pancreatic beta-cells detoxify h2o2 through the peroxiredoxin/thioredoxin antioxidant system. *J Biol Chem*. 2019;294:4843–4853.
31. Jeong W, Bae SH, Toledano MB, et al. Role of sulfiredoxin as a regulator of peroxiredoxin function and regulation of its expression. *Free Radic Biol Med*. 2012;53:447–456.
32. Elumalai S, Karunakaran U, Moon JS, et al. High glucose-induced prdx3 acetylation contributes to glucotoxicity in pancreatic beta-cells: prevention by teneligliptin. *Free Radic Biol Med*. 2020;160:618–629.
33. Findlay VJ, Townsend DM, Morris TE, et al. A novel role for human sulfiredoxin in the reversal of glutathionylation. *Cancer Res*. 2006;66:6800–6806.
34. Lowther WT, Brot N, Weissbach H, et al. Thiol-disulfide exchange is involved in the catalytic mechanism of peptide methionine sulfoxide reductase. *Proc Natl Acad Sci U S A*. 2000;97:6463–6468.
35. Arner ES, Holmgren A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem*. 2000;267:6102–6109.
36. Holmgren A. Thioredoxin and glutaredoxin systems. *J Biol Chem*. 1989;264:13963–13966.
37. Nishiyama A, Masutani H, Nakamura H, et al. Redox regulation by thioredoxin and thioredoxin-binding proteins. *IUBMB Life*. 2001;52:29–33.
38. Holmgren A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J Biol Chem*. 1979;254:9627–9632.
39. Wang Y, De Keulenaer GW, Lee RT. Vitamin d(3)-up-regulated protein-1 is a stress-responsive gene that regulates cardiomyocyte viability through interaction with thioredoxin. *J Biol Chem*. 2002;277:26496–26500.
40. Nordberg J, Arner ES. Reactive oxygen species, antioxidants, and the mammalian thioredoxin system. *Free Radic Biol Med*. 2001;31:1287–1312.
41. Saitoh M, Nishitoh H, Fujii M, et al. Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ask) 1. *EMBO J*. 1998;17:2596–2606.
42. Noguchi T, Takeda K, Matsuzawa A, et al. Recruitment of traf family proteins to the ask1 signalosome is essential for oxidative stress-induced cell death. *J Biol Chem*. 2005;280:37033–37040.
43. Liu Y, Min W. Thioredoxin promotes ask1 ubiquitination and degradation to inhibit ask1-mediated apoptosis in a redox activity-independent manner. *Circ Res*. 2002;90:1259–1266.
44. Hotta M, Tashiro F, Ikegami H, et al. Pancreatic beta cell-specific expression of thioredoxin, an antioxidative and antiapoptotic protein, prevents autoimmune and streptozotocin-induced diabetes. *J Exp Med*. 1998;188:1445–1451.
45. Zhang R, Al-Lamki R, Bai L, et al. Thioredoxin-2 inhibits mitochondria-located ask1-mediated apoptosis in a jnk-independent manner. *Circ Res*. 2004;94:1483–1491.
46. Saxena G, Chen J, Shalev A. Intracellular shuttling and mitochondrial function of thioredoxin-interacting protein. *J Biol Chem*. 2010;285:3997–4005.
47. Junn E, Han SH, Im JY, Choi I, et al. Vitamin d3 up-regulated protein 1 mediates oxidative stress via suppressing the thioredoxin function. *J Immunol*. 2000;164:6287–6295.
48. Nishiyama A, Matsui M, Iwata S, et al. Identification of thioredoxin-binding protein-2/vitamin d(3) up-regulated protein 1 as a negative regulator of thioredoxin function and expression. *J Biol Chem*. 1999;274:21645–21650.
49. Yamanaka H, Maehira F, Oshiro M, et al. A possible interaction of thioredoxin with vdup1 in hela cells detected in a yeast two-hybrid system. *Biochem Biophys Res Commun*. 2000;271:796–800.
50. Patwari P, Higgins LJ, Chutkow WA, et al. The interaction of thioredoxin with txnip: evidence for formation of a mixed disulfide by disulfide exchange. *J Biol Chem*. 2006;281:21884–21891.
51. Minn AH, Hafele C, Shalev A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. *Endocrinology*. 2005;146:2397–2405.
52. Shalev A. Minireview: thioredoxin-interacting protein: regulation and function in the pancreatic beta-cell. *Mol Endocrinol*. 2014;28:1211–1220.
53. Xu G, Chen J, Jing G, et al. Thioredoxin-interacting protein regulates insulin transcription through microrna-204. *Nat Med*. 2013;19:1141–1146.
54. Chen J, Saxena G, Mungrue IN, et al. Thioredoxin-interacting protein: a critical link between glucose toxicity and beta cell apoptosis. *Diabetes*. 2008;57:938–944.
55. Russell MA, Redick SD, Blodgett DM, et al. Hla class ii antigen processing and presentation pathway components demonstrated by transcriptome and protein analyses of islet beta-cells from donors with type 1 diabetes. *Diabetes*. 2019;68:988–1001.
56. Bompada P, Atac D, Luan C, et al. Histone acetylation of glucose-induced thioredoxin-interacting protein gene expression in pancreatic islets. *Int J Biochem Cell Biol*. 2016;81:82–91.
57. Shalev A, Pise-Masison CA, Radonovich M, et al. Oligonucleotide microarray analysis of intact human pancreatic islets: identification of glucose-responsive genes and a highly regulated tgfbeta signaling pathway. *Endocrinology*. 2002;143:3695–3698.
58. Lerner AG, Upton JP, Praveen PV, et al. Ire1alpha induces thioredoxin-interacting protein to activate the nlrp3 inflammasome and promote programmed cell death under irremediable er stress. *Cell Metab*. 2012;16:250–264.
59. Hong K, Xu G, Grayson TB, et al. Cytokines regulate beta-cell thioredoxin-interacting protein (txnip) via distinct mechanisms and pathways. *J Biol Chem*. 2016;291:8428–8439.
60. Chen J, Hui ST, Couto FM, et al. Thioredoxin-interacting protein deficiency induces akt/bcl-xl signaling and pancreatic beta cell mass and protects against diabetes. *FASEB J*. 2008;22:3581–3594.
61. Chutkow WA, Birkenfeld AL, Brown JD, et al. Deletion of the alpha-arrestin protein txnip in mice promotes adiposity and adipogenesis while preserving insulin sensitivity. *Diabetes*. 2010;59:1424–1434.
62. Xu G, Chen J, Jing G, et al. Preventing beta-cell loss and diabetes with calcium channel blockers. *Diabetes*. 2012;61:848–856.
63. Thielen LA, Chen J, Jing G, et al. Identification of an anti-diabetic, orally available small molecule that regulates txnip expression and glucagon action. *Cell Metab*. 2020;32:353–365 e8.
64. Ovalle F, Grimes T, Xu G, et al. Verapamil and beta cell function in adults with recent-onset type 1 diabetes. *Nat Med*. 2018;24:1108–1112.
65. Yamawaki H, Berk BC. Thioredoxin: a multifunctional antioxidant enzyme in kidney, heart and vessels. *Curr Opin Nephrol Hypertens*. 2005;14:149–153.
66. Aon MA, Tocchetti CG, Bhatt N, et al. Protective mechanisms of mitochondria and heart function in diabetes. *Antioxid Redox Signal*. 2015;22:1563–1586.
67. Perrone L, Devi TS, Hosoya KI, et al. Inhibition of txnip expression in vivo blocks early pathologies of diabetic retinopathy. *Cell Death Dis*. 2010;1:e65.
68. Devi TS, Lee I, Huttemann M, et al. Txnip links innate host defense mechanisms to oxidative stress and inflammation in retinal muller glia under chronic hyperglycemia: implications for diabetic retinopathy. *Exp Diabetes Res*. 2012;2012:438238.
69. Wang S, Du S, Lv Y, et al. Elevated microrna-20b-3p and reduced thioredoxin-interacting protein ameliorate diabetic retinopathy progression by suppressing the nlrp3 inflammasomes. *IUBMB Life*. 2020;72:1433–1448.
70. Schulze PC, Yoshioka J, Takahashi T, et al. Hyperglycemia promotes oxidative stress through inhibition of thioredoxin function by thioredoxin-interacting protein. *J Biol Chem*. 2004;279:30369–30374.
71. Yoshioka J, Chutkow WA, Lee S, et al. Deletion of thioredoxin-interacting protein in mice impairs mitochondrial function but protects the myocardium from ischemia-reperfusion injury. *J Clin Invest*. 2012;122:267–279.

72. Shi Y, Ren Y, Zhao L, et al. Knockdown of thioredoxin interacting protein attenuates high glucose-induced apoptosis and activation of ask1 in mouse mesangial cells. *FEBS Lett*. 2011;585:1789–1795.
73. Song S, Qiu D, Wang Y, et al. Txnip deficiency mitigates podocyte apoptosis via restraining the activation of mtor or p38 mapk signaling in diabetic nephropathy. *Exp Cell Res*. 2020;388:111862.
74. Huang C, Zhang Y, Kelly DJ, et al. Thioredoxin interacting protein (txnip) regulates tubular autophagy and mitophagy in diabetic nephropathy through the mTOR signaling pathway. *Sci Rep*. 2016;6:29196.
75. Parikh H, Carlsson E, Chutkow WA, et al. Txnip regulates peripheral glucose metabolism in humans. *PLoS Med*. 2007;4:e158.
76. Blouet C, Schwartz GJ. Nutrient-sensing hypothalamic txnip links nutrient excess to energy imbalance in mice. *J Neurosci*. 2011;31:6019–6027.
77. Burg AR, Tse HM. Redox-sensitive innate immune pathways during macrophage activation in type 1 diabetes. *Antioxid Redox Signal*. 2018;29:1373–1398.
78. Dewidar B, Kahl S, Pafili K, Roden M. Metabolic liver disease in diabetes – from mechanisms to clinical trials. *Metabolism*. 2020;111S:154299.

Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State

Ketan Dhatariya, Francisco J. Pasquel, and Guillermo Umpierrez

OUTLINE

Diabetic Ketoacidosis, 641
- Epidemiology, 641
- Pathogenesis of Diabetic Ketoacidosis, 642
- Precipitating Factors, 643
- Diagnosis of Diabetic Ketoacidosis, 644
- Treatment, 646
- Complications of Therapy, 647
- Prevention, 648

Hyperglycemic Hyperosmolar State, 648
- Epidemiology, 648
- Pathophysiology, 648
- Precipitating Events, 648
- Diagnosis, 649
- Treatment, 650
- Complications, 650
- Prevention, 650

Diabetic ketoacidosis (DKA) and hyperglycemic hyperosmolar state (HHS) are the most serious hyperglycemic emergencies in patients with type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM). Although DKA and HHS are often discussed as separate entities, they represent points along a spectrum of hyperglycemic emergencies because of poorly controlled diabetes. Both are characterized by insulinopenia and severe hyperglycemia. However, clinically they differ only by the degree of dehydration and the severity of metabolic acidosis.[1-4] The overall mortality in children and adult subjects with DKA is reported to be less than 1%, but a mortality rate higher than 5% is reported in the elderly and in patients with concomitant life-threatening illnesses.[3,5-7] The rate is also high in less developed nations, with mortality rates reported at over 40%.[8-10] In more developed nations, mortality in patients with HHS is higher than in those with DKA and has been reported as being as high as 20%.[11-13] Successful treatment of both conditions requires early recognition and diagnosis, frequent monitoring, replacement of fluid and electrolyte deficits to restore circulatory volume and tissue perfusion, correction of hyperglycemia, and a careful search for the precipitating cause. This chapter reviews recent advances on the pathophysiology, epidemiology, diagnosis, clinical manifestations, management recommendations, and prevention of DKA and HHS.

DIABETIC KETOACIDOSIS

Epidemiology

KEY POINTS

- DKA can occur in type 1 and type 2 diabetes and can occur at any age. Whilst it may be the presenting feature of new onset diabetes, this is uncommon, and most cases are in people with an existing diagnosis of diabetes.

DKA is the most common and serious acute complication of diabetes. DKA is responsible for more than 188,000 hospital admissions per year in the United States, at substantial cost related to direct medical expenses and indirect costs.[3,14-16] Recent epidemiological studies indicate that hospitalizations for DKA are increasing, with the majority of cases being recurrent cases in the same subjects.[17-21] Along with this, treatment of DKA uses many resources. In the United States an individual episode was estimated to cost $26,566, accounting for an estimated annual total cost in 2014 of over $5 billion.[15] In the United Kingdom the cost of an episode of DKA has been estimated at £2064 for adults and £1387 for children.[22,23]

Observational studies from the United States have reported that, in 2016, DKA accounted for 2.4% of all hospital admissions where diabetes was listed as a diagnosis (equal to a crude rate of 8.1 per 1000 people with diabetes).[24] Worldwide incidence rates are higher in populations with limited access to medical care for social or economic reasons. A systematic review of incidence around the world suggested a rate of 8.0 to 51.3 cases per 1000 people with T1DM, with the highest rates being in those aged between 15 and 39 years.[25] Data from China suggested the rates were higher, up to 263 per 1000 patient-years.[26] In the United States, the SEARCH for Diabetes in Youth Study found that, between 2002 and 2010, the rates of DKA as the presenting feature of T1DM amongst people younger than 20 years remained steady, at around 30%.[27] In addition, previous data from the same study showed that 29.4% of participants less than 20 years of age with type 1 diabetes presented with DKA, as compared with 9.7% of youth with T2DM.[28] In community-based studies, more than 40% of adult patients with DKA are older than 40 years, and more than 20% are older than 55 years.[29] Mortality increases with age and with the presence of comorbidities.[6,7] Patients with T2DM may develop DKA under stressful conditions such as trauma, surgery, or infection.[4,29] However, an increasing number of unprovoked ketoacidosis cases without precipitating cause have been reported in children and adult subjects with T2DM.[28] Unprovoked DKA cases are more commonly reported in African Americans and Latinos,[30-33] but have been reported in all ethnic groups.[34-37] A potentially helpful classification divides patients with DKA into four categories based on the presence or absence of biomarkers of autoimmunity (A+ or –) and the presence or absence of residual insulin secretory capacity as evidenced by C-peptide release (β+ or –).[38] Classic type 1 disease would be A+β–, while classic type 2 would be A–β+.[38,39]

In children and adult subjects with DKA, the overall mortality is less than 1%, but a mortality rate higher than 5% is reported in the elderly and in patients with concomitant life-threatening illnesses.[3]

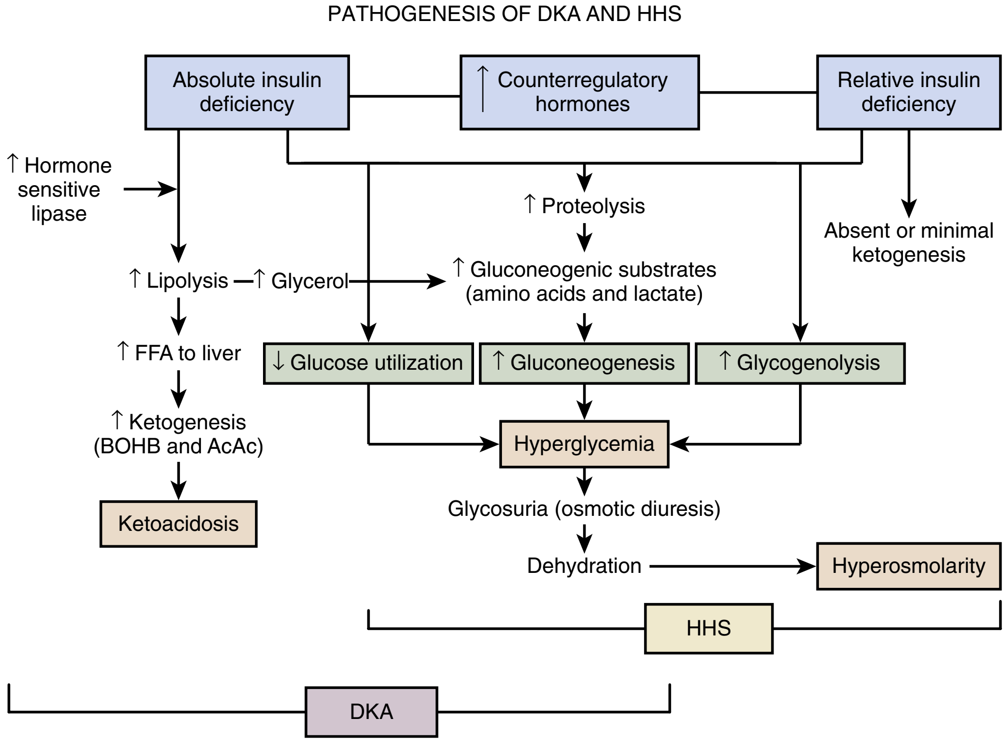

PATHOGENESIS OF DKA AND HHS
Absolute insulin deficiency
↑ Counterregulatory hormones
Relative insulin deficiency
↑ Hormone sensitive lipase
↑ Proteolysis
Absent or minimal ketogenesis
↑ Lipolysis
↑ Glycerol
↑ Gluconeogenic substrates (amino acids and lactate)
↑ FFA to liver
↓ Glucose utilization
↑ Gluconeogenesis
↑ Glycogenolysis
↑ Ketogenesis (BOHB and AcAc)
Hyperglycemia
Ketoacidosis
Glycosuria (osmotic diuresis)
Dehydration
Hyperosmolarity
HHS
DKA

Fig. 37.1 **Pathogenesis of diabetic ketoacidosis.** Hyperglycemia and ketogenesis result from a relative or absolute insulin deficiency combined with excess levels of counterregulatory hormones including glucagon, cortisol, catecholamines, and growth hormone. Hyperglycemia develops as a result of increased gluconeogenesis, generation of gluconeogenic precursors (alanine, amino acids, and glycerol), accelerated glycogenolysis, and impaired glucose utilization by peripheral tissues. Accumulation of ketone bodies is caused by increased lipolysis and circulating free fatty acids (*FFAs*). In the liver, FFAs are oxidized to ketone bodies, a process predominantly stimulated by glucagon. The two major ketone bodies are β-hydroxybutyrate (*bOHB*) and acetoacetic acid (*AcAc*). Accumulation of ketone bodies leads to a decrease in serum bicarbonate concentration and metabolic acidosis. The ketosis and acidosis in diabetic ketoacidosis (*DKA*) contribute to electrolyte disturbances, vomiting and dehydration. *HHS,* Hyperglycemic hyperosmolar state.

In adult patients with diabetes, mortality increases substantially with age, with mortality rates for those over 65 to 75 years reaching 10% to 20%.[29] In the older age groups, the major cause of death relates to the underlying medical illness (i.e., trauma, infection) that precipitated the ketoacidosis, but in the younger patient, mortality is more likely to be because of the metabolic disarray.

Pathogenesis of Diabetic Ketoacidosis

The mechanisms that trigger DKA are multifactorial and include a combination of reduced secretion and action of insulin and raised levels of counterregulatory hormones (glucagon, catecholamines, cortisol, and growth hormone).[1,4,40] The insulin deficiency of DKA can be absolute in T1DM or relative as in T2DM in the presence of increased release in counterregulatory hormones that causes worsening of insulin resistance and further impairment of insulin secretion.[4,40,41]

Ketogenesis and Ketoacidosis. The combination of insulin deficiency and glucagon excess (elevated glucagon/insulin ratio) is critical for the development of hyperglycemia and ketoacidosis (see Chapters 30 and 36).[42] The increase in counterregulatory hormones leads to the release of free fatty acids (FFAs) into the circulation from adipose tissue (lipolysis) and to unrestrained hepatic fatty acid oxidation in the liver to ketone bodies (Fig. 37.1).[4,43,44] Insulin deficiency causes the activation of hormone-sensitive lipase in adipose tissue. The increased activity of tissue lipase causes breakdown of triglyceride into glycerol and FFAs. Although glycerol becomes an important substrate for gluconeogenesis in the liver, the massive release of FFAs assumes pathophysiologic predominance, as they are the hepatic precursors of the ketoacids. In the liver, FFAs are oxidized to ketone bodies, a process predominantly stimulated by glucagon-induced generation of cyclic adenosine monophosphate (cAMP). Increased concentration of glucagon lowers the hepatic levels of malonyl coenzyme A (CoA) by blocking the conversion of pyruvate to acetyl CoA through inhibition of acetyl CoA carboxylase, the first rate-limiting enzyme in *de novo* fatty acid synthesis. Malonyl CoA, inhibits carnitine palmitoyltransferase 1 (CPT-1), the rate-limiting enzyme for transesterification of fatty acyl CoA to fatty acyl carnitine,[45] allowing oxidation of fatty acid to ketone bodies. CPT-1 is required for movement of FFA into the mitochondria, where fatty acid oxidation takes place.[45] The increased fatty acyl CoA and CPT-1 activity in DKA lead to increased ketogenesis in DKA. Increased production ketone bodies (acetoacetate and β-hydroxybutyrate) lead to ketonemia and metabolic acidosis.[46,47]

Hyperglycemia. In normal subjects during fasting state, plasma glucose is maintained between 3.9 and 5.6 mmol/L (70 - 100 mg/dl) by a finely regulated balance between hepatic glucose production and glucose utilization in peripheral tissues (see Chapter 30). Insulin controls hepatic glucose production by suppressing hepatic gluconeogenesis and glycogenolysis. In insulin-sensitive tissues such as muscle, insulin promotes protein anabolism, glucose uptake, and glycogen synthesis and inhibits glycogenolysis and protein breakdown. In addition, insulin is a powerful inhibitor of lipolysis, FFA oxidation, and ketogenesis.

TABLE 37.1 **Diagnostic Criteria for Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State**

DIAGNOSTIC CRITERIA AND CLASSIFICATION	DIABETIC KETOACIDOSIS			
	Mild	Moderate	Severe	Hyperglycemic Hyperosmolar State
Plasma glucose (mg/dL)	>250	>250	>250	>600
Arterial pH	7.25–7.30	7.00–≤7.24	<7.00	>7.30
Serum bicarbonate (mEq/L)	15–18	10–≤15	<10	>15
Urine ketone	Positive	Positive	Positive	Small
Serum ketone	Positive	Positive	Positive	Small
Effective serum osmolality	Variable	Variable	Variable	>320 mOsm/kg
Anion gap	>10	>12	>12	<12
Alteration in sensorium or mental obtundation	Alert	Alert/drowsy	Stupor/coma	Stupor/coma

The counterregulatory hormones (glucagon, catecholamines, cortisol, and growth hormone) promote metabolic pathways opposite to insulin action, both in the liver and peripheral tissues.[31]

Hyperglycemia in patients with DKA develops as a result of three processes: increased gluconeogenesis, accelerated glycogenolysis, and impaired glucose utilization by peripheral tissues (Fig. 37.1).[40] From a quantitative standpoint, increased hepatic glucose production represents the major pathogenic disturbance in DKA.[48] Renal gluconeogenesis also contributes approximately 20% of circulating glucose in healthy individuals, but a metabolic acidosis increases renal glucose output, while at the same time inhibiting hepatic glucose output.[49,50] The increased hepatic glucose production results from the high availability of gluconeogenic precursors, such as amino acids (alanine and glutamine, as a result of accelerated proteolysis and decreased protein synthesis), lactate (as a result of increased muscle glycogenolysis), and glycerol (as a result of increased lipolysis) and from the increased activity of gluconeogenic enzymes (phosphoenol pyruvate carboxykinase, fructose-1,6-bisphosphatase, and pyruvate carboxylase).[51,52] The activation of gluconeogenic enzymes appears to be predominantly stimulated by an increased synthesis of peroxisome proliferator-activated receptor γ-coactivator-1α mediated by glucagon-induced production of cAMP.[53,54] Elevation of cAMP also causes a fall in fructose 2,6-bisphosphate leading to inhibition of phosphofructokinase, a key regulatory enzyme in the glycolytic pathway.[55]

In addition to insulin deficiency, increased counterregulatory hormones play an important role in glucose overproduction in DKA. Elevated glucagon and catecholamine levels leads to increased gluconeogenesis and glycogenolysis.[56-58] Epinephrine stimulates glucagon secretion and inhibits insulin release by pancreatic β cells.[40,41] High cortisol levels stimulate protein catabolism and increase circulating amino acid concentrations, providing precursors for gluconeogenesis.[59,60] In addition, increased glucose levels causes an osmotic diuresis, which contributes to dehydration, hypovolemia, and decreased glomerular filtration rate, which further increase the severity of hyperglycemia.[4]

The development of hyperglycemia and ketoacidosis results in an inflammatory state characterized by an elevation of proinflammatory cytokines and increased oxidative stress markers.[61,62] Circulating levels of tumor necrosis factor-α (TNF-α), interleukin (IL)-6, IL-6β, IL-8, and C-reactive protein have been found to be increased 2- to 4-fold on admission in patients with hyperglycemic crises compared with control subjects, with levels returning to normal after insulin treatment and resolution of hyperglycemic crises.[62] The increased cytokine concentrations probably contribute to the insulin resistance associated with DKA through altered regulation of the insulin-signaling pathway and interaction at the level of the insulin receptor itself.[63] TNF-α causes a moderate decrease in the insulin-stimulated autophosphorylation of the insulin receptor and a dramatic decrease in tyrosine phosphorylation of insulin receptor substrate-1 (IRS-1). Activation of c-Jun NH2-terminal kinase by TNF-α mediates serine phosphorylation of IRS-1, preventing insulin-mediated activation of phosphatidylinositol 3-kinase, a key regulator of tissue glucose uptake. Similarly, elevated fatty acids during DKA could stimulate phosphorylation of IRS-1 on its serine residues, leading to worsening insulin resistance.[64,65]

Increased oxidative stress and generation of reactive oxygen species during acute illness can lead to cellular damage of lipids, membranes, proteins, and DNA. Oxidative stress decreases the expression and activity of key transcription factors that regulate genes involved in β cell function. The increased inflammatory state in DKA can lead to capillary perturbation through activation of vasoactive peptides like endothelin 1, vascular endothelial growth factor, and intercellular adhesion molecule-1, whose expression levels are influenced by ketones bodies,[66,67] and may explain why hyperglycemia is associated with poor vascular outcomes in hyperglycemic patients.[61]

Precipitating Factors

DKA is the initial manifestation of diabetes in up to 20% of adult patients and in 30% to 40% of children with T1DM.[4,27] Recurrent episodes have been reported to occur in between 21.6% and 66% of people with either T1DM or T2DM.[18,68] In known diabetic patients, precipitating factors for DKA, particularly for recurrent episodes, include infections, intercurrent illnesses, psychological stress, and poor compliance with therapy.[68,69] Infection is the most common precipitating factor for DKA, occurring in 14% to 58% of cases (Table 37.1).[4] Urinary tract infection and pneumonia account for the majority of infections. Other acute conditions that may precipitate DKA include cerebrovascular accident, alcohol abuse, pancreatitis, pulmonary embolism, myocardial infarction, and trauma. Drugs that affect carbohydrate metabolism such as corticosteroids, thiazides, sympathomimetic agents, and pentamidine may also precipitate the development of DKA, as well as the use of the atypical antipsychotic agents.[6,70] The sodium-glucose cotransporter inhibitors, when used in people with T2DM, have been shown to be associated with an increased risk of developing DKA when compared to those not on these agents. US Food and Drug Administration data show that, among patients who met criteria for T2DM, DKA rates ranged from 0.32 (95% confidence interval [CI]: 0.26–0.40) for dapagliflozin to 0.44 (95% CI: 0.40–0.48) per 100 person-years for canagliflozin.[71] In people with T1DM, before the widespread use of these agents, the rate of DKA in the United Kingdom were reported at between 3.0% and 8.7% (depending on the definition of DKA used).[72] With the use of these agents in people with T1DM, in clinical trials the relative risk has risen to between 1.76- to 18.64-fold higher than with placebo.[73] The exception to this increased risk was the use of 2.5 mg of empagliflozin, which was associated with a 72% relative risk reduction in the incidence of DKA.[74]

Pregnant women with diabetes are also at higher risk compared with nonpregnant diabetic women. Risk factors that can predispose pregnant women include accelerated starvation, dehydration, nausea or hyperemesis gravidarum, decreased buffering capacity as a result of increased pulmonary ventilation and compensatory decrease in bicarbonate, stress, and increased production of human placental lactogen, prolactin, and cortisol.[75,76]

The importance of noncompliance and psychological factors in the incidence of DKA has been particularly apparent in younger patients. In adult patients with T1DM, poor adherence to diabetes treatment program is also the major precipitating cause of DKA.[4,6,70,77] Several behavioral, socioeconomic, and psychosocial factors contribute to poor treatment adherence. Psychiatric disorders such as depression, anxiety, and eating disorders are common among patients with T1DM, and the history of a psychiatric disorder can double the risk of having a DKA episode.[4,68,78] Studies have reported eating disorders in up to 20% of recurrent episodes of ketoacidosis in young women.[79]

Diagnosis of Diabetic Ketoacidosis

Symptoms and Signs. The clinical manifestations are polyuria, polydipsia, signs of dehydration, and progressive deterioration of mental status leading to coma. Abdominal pain, sometimes mimicking an acute abdomen, is reported in 40% to 75% of cases of DKA.[80] In addition, nausea and vomiting are commonly seen in DKA and are likely to be caused largely by the combined effects of dehydration, ketonemia, and delayed gastric emptying. The presence of abdominal pain is associated with more severe metabolic acidosis, but not with the severity of hyperglycemia.[80] The osmotic diuresis leads to a high urine output until extreme volume depletion leads to a critical decrease in renal blood flow and glomerular filtration.

Physical examination reveals signs of dehydration, including dry mucous membranes, tachycardia, and hypotension. Acetone on the breath and labored Kussmaul respiration characterized by rapid, deep respirations in response to metabolic acidosis are frequently found in children and adults with severe metabolic acidosis. Mental status can vary from full alertness to profound lethargy; however, less than 20% of patients are hospitalized with loss of consciousness.[40,81] Hypothermia, potentially owning to a failure of heat production as a result of impaired oxygen consumption, is not uncommon, and can potentially aggravate ketoacidosis.[82,83] The presence of fever is highly suggestive of a concomitant infection.

KEY POINTS

- DKA should only be diagnosed with glucose is >200mg/dl OR there is a known history of diabetes (the "D"), there is a ketonemia of >3.0mmol/l OR ketonuria of ≥2+ (the "K"), and a high anion gap metabolic acidosis with a pH of <7.3 or a bicarbonate of <18mmol/l (the "A").

Laboratory Findings. The syndrome of DKA consists of the triad of hyperglycemia (the "D"), ketonemia (the "K"), and metabolic acidosis (the "A").[3,84] DKA can be classified as mild, moderate, or severe, depending on the extent of metabolic acidosis and level of sensorium (Table 37.1).[3] The severity of metabolic acidosis often bears little relationship to the degree of hyperglycemia, and cases of ketoacidosis with normal or only modestly elevated glucose levels (<13.8 mmol/L, 250 mg/dL) have been reported.[72,85] This phenomenon has been reported during pregnancy, in patients with prolonged starvation, and in those who present after receiving insulin.[86] Similarly, relatively low glucose concentrations may occur in the presence of impaired hepatic glucose output, such as in patients with alcohol abuse or liver failure.[4,87,88]

The key diagnostic feature is the elevation in circulating total blood ketone concentration. Assessment of ketonemia is usually performed by the nitroprusside reaction, which provides a semiquantitative estimation of acetoacetate and acetone levels. Although the nitroprusside test (urine and serum) is highly sensitive, it can underestimate the severity of ketoacidosis, because this assay does not recognize the presence of β-hydroxybutyrate, the main metabolic product in ketoacidosis.[51,89] If available, measurement of serum β-hydroxybutyrate should be used for the diagnosis and monitoring of progress.[47,90] Previous work has shown that home capillary blood β-hydroxybutyrate testing is more effective than urine acetoacetate testing in reducing emergency visits, hospitalization, and time to recovery from DKA.[91,92] Bedside capillary β-hydroxybutyrate testing is advocated as one of the cornerstones of management in the UK DKA guidelines.[93]

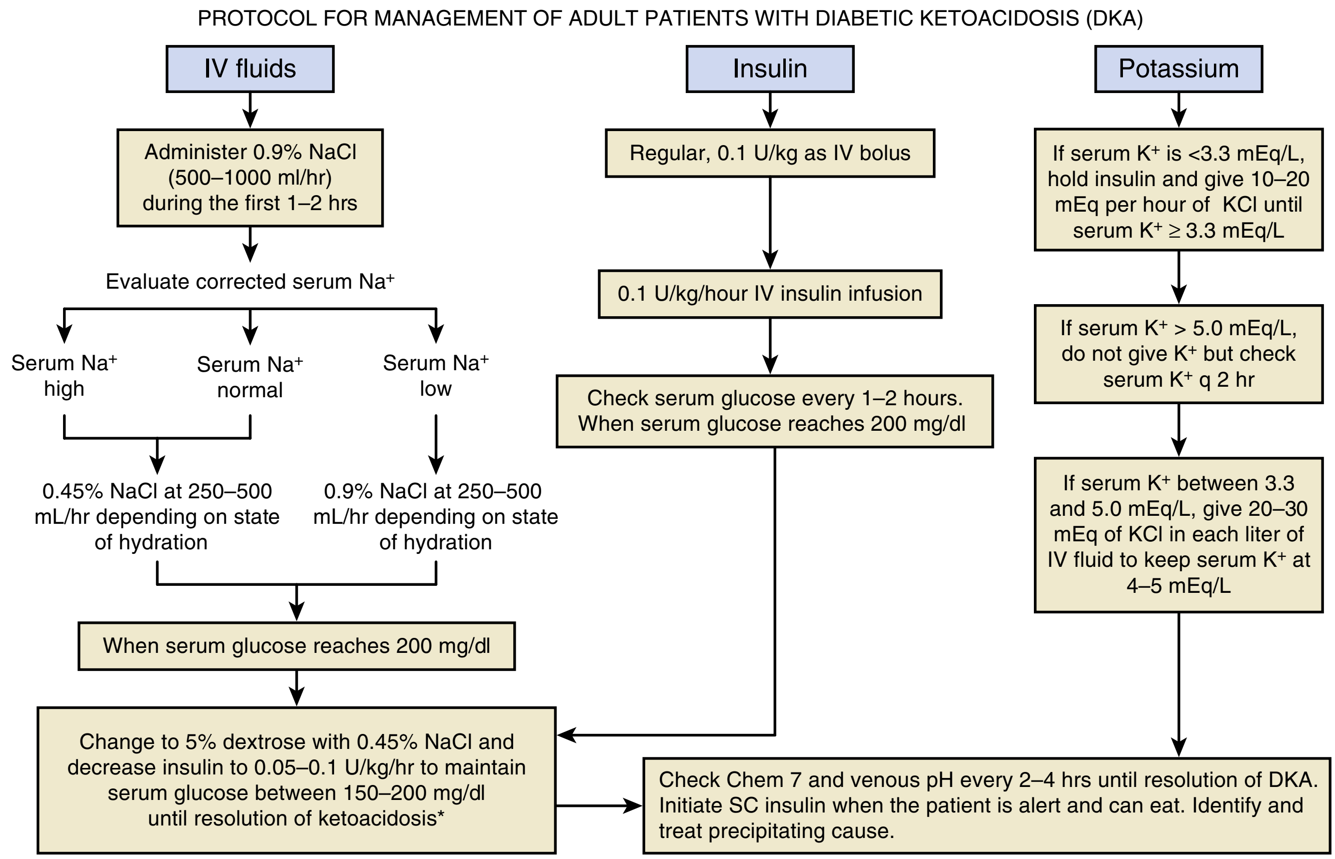
Accumulation of ketoacids results in an increased anion gap metabolic acidosis. The anion gap is calculated by subtracting the sum of chloride and bicarbonate concentration from the sodium concentration ($Na^+ - [Cl^- + HCO_3^-]$) (Fig. 37.2). A normal anion gap is between 7 and 9 mEq/L, and an anion gap greater than 12 mEq/L indicates the presence of increased anion gap metabolic acidosis.[40] Although most subjects with DKA present with a high anion gap acidosis, it is important to keep in mind that patients may present with mixed acid-base disorders.[94] More importantly, during treatment and resolution of DKA, most patients develop a transient hyperchloremic non–anion gap metabolic acidosis.[88,95]

Patients with DKA have an extracellular fluid volume deficit, usually in the range of 4% to 10%.[96-99] Clinical estimates of the volume deficit are subjective and inaccurate.[99-101] The magnitude of dehydration cannot be assessed accurately by either clinical or biochemical parameters. Calculation of effective osmolality [sodium ion (mEq/L) × 2 + glucose (mg/dL)/18] is the best marker of volume depletion, and levels greater than 320 mmol/L are frequently associated with impaired mental status (Table 37.2).

The admission serum sodium is often low because of the osmotic flux of water from the intracellular to the extracellular space in the presence of hyperglycemia. To assess the severity of sodium and water deficit, serum sodium may be corrected by adding 1.6 mg/dL to the measured serum sodium for each 100 mg/dL of glucose above 100 mg/dL (Table 37.2).[40] An increase in serum sodium concentration in the presence of hyperglycemia indicates a rather profound degree of water loss.

The admission serum potassium concentration is usually elevated in patients with DKA.[4,40] These high levels occur because of a shift of potassium from the intracellular to the extracellular space because of acidemia, insulin deficiency, and hypertonicity. In addition, as the circulating volume decreases owing to the hyperglycemia-induced osmotic diuresis, aldosterone concentrations rise in an attempt to retain sodium, but at the expense of further potassium losses. Similarly, the admission serum phosphate level may be normal or elevated because of metabolic acidosis.

Patients with DKA frequently present with leukocytosis in the absence of infection.[102] However, a leukocyte count greater than 25,000 per mm^3 or the presence of greater than 10% neutrophil bands is seldom seen in the absence of bacterial infection.[3] Because most patients with DKA present with abdominal pain, nausea, or vomiting, the differential diagnosis of acute pancreatitis should be considered.


PROTOCOL FOR MANAGEMENT OF ADULT PATIENTS WITH DIABETIC KETOACIDOSIS (DKA)
IV fluids
Insulin
Potassium
Administer 0.9% NaCl (500–1000 ml/hr) during the first 1–2 hrs
Evaluate corrected serum Na+
Serum Na+ high
Serum Na+ normal
Serum Na+ low
0.45% NaCl at 250–500 mL/hr depending on state of hydration
0.9% NaCl at 250–500 mL/hr depending on state of hydration
When serum glucose reaches 200 mg/dl
Change to 5% dextrose with 0.45% NaCl and decrease insulin to 0.05–0.1 U/kg/hr to maintain serum glucose between 150–200 mg/dl until resolution of ketoacidosis*
Regular, 0.1 U/kg as IV bolus
0.1 U/kg/hour IV insulin infusion
Check serum glucose every 1–2 hours. When serum glucose reaches 200 mg/dl
If serum K+ is <3.3 mEq/L, hold insulin and give 10–20 mEq per hour of KCl until serum K+ ≥ 3.3 mEq/L
If serum K+ > 5.0 mEq/L, do not give K+ but check serum K+ q 2 hr
If serum K+ between 3.3 and 5.0 mEq/L, give 20–30 mEq of KCl in each liter of IV fluid to keep serum K+ at 4–5 mEq/L
Check Chem 7 and venous pH every 2–4 hrs until resolution of DKA. Initiate SC insulin when the patient is alert and can eat. Identify and treat precipitating cause.

*Resolution of ketoacidosis = blood glucose < 250 mg/dl, bicarbonate > 18 mEq/L, and pH > 7.30.

Fig. 37.2 Protocol for management of patients with diabetic ketoacidosis (*DKA*). *IV,* Intravenous; *K+,* potassium ion; *KCl,* potassium chloride; *Na+,* sodium ion; *NaCl,* sodium chloride.

TABLE 37.2 Useful Formulas for the Evaluation of Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State

1	Calculation of anion gap (AG): $\text{AG} = [\text{Na}^+] - [\text{Cl}^- + \text{HCO}_3^-]$
2	Total and effective serum osmolality: $\frac{\text{Total} = 2[\text{Na}^+] + \text{glucose (mg/dL)}}{18} + \frac{\text{BUN (mg/dL)}}{2.8}$
3	$\frac{\text{Effective} = 2[\text{Na}^+] + \text{glucose (mg/dL)}}{18}$
4	Corrected serum sodium: Corrected $[\text{Na}^+] = \frac{1.6 \times (\text{mg/dL}) - 100}{100} + [\text{measured Na}^+]$
	Total body water (TBW) deficit: TBW deficit $= [\text{weight (kg)} \times 0.6] - \frac{\text{corrected Na}^+ - 1}{140}$

BUN, Blood urea nitrogen.

Increased amylase levels are reported in 21% to 79% of patients with DKA. There is little correlation between the presence, degree, or isoenzyme type of hyperamylasemia and the presence of symptoms or pancreatic imaging studies; therefore, an increased serum amylase is not definitive for the diagnosis of pancreatitis in DKA.[80] Nonspecific serum lipase elevation has also been reported in one third of patients with DKA in the absence of clinical and radiological evidence of acute pancreatitis.[80] Even though the patient may present with features of an "acute abdomen," the DKA must be treated before any attempt at surgical intervention.

Not all patients who present with anion gap acidosis or ketoacidosis have DKA.[2] Patients with chronic ethanol abuse with a recent binge culminating in nausea, vomiting, and acute starvation may present with alcoholic ketoacidosis. The key diagnostic feature that differentiates diabetic and alcohol-induced ketoacidosis—in individuals not known to have diabetes—is the concentration of blood glucose.[88,103] Although DKA is characterized by severe hyperglycemia, the presence of ketoacidosis without hyperglycemia in an alcoholic patient is virtually diagnostic of alcoholic ketoacidosis. Of course, as mentioned, with the advent of sodium-glucose cotransporter inhibitor use, euglycemic DKA is within the differential diagnosis, and so (as with everything) taking a good history is paramount. Some patients with decreased food intake (<500 calories/day) for several days may present with starvation ketosis.[104] However, a healthy subject is able to adapt to prolonged fasting by increasing ketone clearance by peripheral tissue (brain and muscle) and by enhancing the kidney's ability to excrete ammonia to compensate for the increased acid production. In starvation, ketone concentrations rise slowly, reaching 5 to 6 mmol/L after a few days of absolute starvation or 4 to 5 mmol/L after 10 days.[105-107]

Treatment

KEY POINTS

- Whilst there is no true physiological replacement fluid for DKA, most people use 0.9% sodium chloride solution as the fluid of choice. Fluid replacement (with potassium supplementation) is paramount, followed by insulin. The routine use of bicarbonate, phosphate or magnesium replacement is discouraged.

Fig. 37.2 shows the American Diabetes Association (ADA) algorithm for the treatment of DKA.[3] The goals of DKA therapy are: (1) to improve circulatory volume and ultimately tissue perfusion, (2) to correct hyperglycemia by gradually decreasing serum glucose and plasma osmolarity, (3) to correct electrolyte imbalance and reduce serum ketone body concentrations, and (4) to identify and treat precipitating events. Frequent monitoring of vital signs, volume and rate of fluid administration, insulin dosage, and urine output and assessment of response to medical treatment are essential for the successful management of DKA. Serial laboratory measurements include glucose, electrolytes, venous pH, bicarbonate, and anion gap values until resolution of ketoacidosis.

There are recent data to show that guidelines used in treating DKA are safe and effective when treating adult patients with DKA in an intensive care unit (ICU) or in non-ICU settings.[5,108-112] Regular feedback given to healthcare staff also helps to further improve care.[113] Even though guidelines do have parameters to suggest when ICU teams should be involved, several observational and prospective studies have indicated no clear benefits in treating DKA patients in the ICU compared with step-down units or general medicine wards.[114-117] The mortality rate, length of hospital stay, or time to resolve ketoacidosis are similar between patients treated in ICU and non-ICU settings. In addition, ICU admission has been shown to be associated with more testing and significantly higher hospitalization cost in patients with DKA.[114,118] Thus, patients with mild to moderate DKA can be safely managed in the emergency department, in step-down units, or on general medical wards with only patients with severe DKA or those with a critical illness as precipitating cause (e.g., myocardial infarction, gastrointestinal bleeding, sepsis) be treated in the ICU.[3,119]

Fluid Therapy. Patients with DKA are volume depleted, with an estimated water deficit of approximately 100 mL/kg of body weight.[40] The water deficit can be estimated, based on corrected serum sodium concentration, using the following equation: water deficit = (0.6)(body weight in kilograms) × (1- [corrected sodium/140]).[40] The goal is to replace half the estimated water deficit over a period of 24 hours. Aggressive fluid therapy expands intravascular volume, restores renal perfusion, and improves insulin sensitivity by reducing circulating counterregulatory hormone concentrations. Isotonic saline (0.9% sodium chloride [NaCl]) at a rate of 500 to 1000 mL/hour during the first 1 to 2 hours is usually sufficient to restore blood pressure and renal perfusion. Subsequent choice and rate of fluid replacement depend on hemodynamics, the state of hydration, serum electrolyte levels, and urinary output. In general, 0.45% NaCl infused at 250 to 500 mL/hour is appropriate if the corrected serum sodium is normal or elevated; 0.9% NaCl at a similar rate is appropriate if corrected serum sodium is low (Fig. 37.2).[40] The protocol for a systematic review of the fluids used to treat DKA has recently been published.[120]

Once the plasma glucose is approximately 250 mg/dL, 5% or 10% dextrose should be added to replacement fluids to allow continued insulin administration until ketonemia is controlled, while at the same time avoiding hypoglycemia.[3] The administration of large amounts of isotonic saline, which contains a high content of chloride, can lead to hyperchloremic metabolic acidosis.[121] Preliminary studies with the use of Ringer's lactate solution have reported a reduction of hyperchloremic acidosis, but this approach has not been shown to improve outcome or to reduce hospital length of stay compared with saline administration.[121,122]

Insulin Therapy. Insulin administration is the cornerstone of DKA therapy. Insulin increases peripheral glucose utilization and decreases hepatic glucose production, as well as inhibiting the release of FFA from adipose tissue and decreasing ketogenesis. Randomized controlled studies in patients with DKA have shown that lower doses of insulin therapy, compared with the very high doses used decades ago, are effective regardless of the route of administration.[123-127] A common and effective practice (particularly when the intravenous [IV] infusion may be delayed) includes an initial IV bolus of regular insulin of 0.1 units/kg of body weight followed by a continuous infusion of regular insulin at a dose of 0.1 units/kg per hour.[3,93] Previous work has shown that the initial bolus is not necessary if the hourly insulin infusion rate is started at 0.14 U/kg body weight.[128] Once the serum glucose has declined to about 250 mg/dL, the insulin infusion rate should be reduced to 0.05 units/kg per hour. Thereafter, the rate of insulin administration may need to be adjusted to maintain glucose levels at approximately 200 mg/dL and continued until ketoacidosis is resolved.

A patient with uncomplicated DKA can be treated with subcutaneous rapid-acting insulin lispro or insulin aspart.[116,117] These studies showed that, after an initial subcutaneous dose of 0.2 to 0.3 U/kg, the administration of 0.1 or 0.2 units/kg every 1 or 2 hours, respectively, was as effective as the continuous infusion of regular insulin. This is done in the emergency room or in non-ICU wards if adequate personnel are available for frequent glucose monitoring. More recently, the advent of the SARS-CoV-2 (COVID-19) viral pandemic has meant that many people who present with DKA may also need insulin given—either subcutaneously if they have mild biochemical derangement, in situations of low resource availability, or when health systems are overwhelmed and infusion pumps to deliver IV insulin are not available.[129,130] Several examples of such protocols are available to download at www.covidindiabetes.org/.

Potassium. Patients with DKA have an estimated total potassium deficit of approximately 3 to 5 mEq/kg of body weight. Despite this deficit, most patients with DKA have a serum potassium level at or above the upper limits of normal.[3,5] These high levels occur because of a shift of potassium from the intracellular to the extracellular space because of acidemia, insulin deficiency, and hypertonicity. Both insulin therapy and correction of acidosis decrease potassium levels by stimulating cellular potassium uptake in peripheral tissues and by increasing urinary excretion owing to high aldosterone concentrations. IV potassium should be initiated when serum concentration is below 5.0 mEq/L. The treatment goal is to maintain serum potassium levels within the normal range of 4 to 5 mEq/L.[3,93] Generally, 20 to 30 mEq of potassium chloride in each liter of infusion fluid is sufficient to maintain a potassium concentration within the normal range. Despite widespread use of protocols with an aggressive potassium replacement regimen, hypokalemia still occurs.[5,94] This raises the argument that insulin infusion rates may need to drop to 0.05 U/kg per hour once glucose concentrations drop. To prevent severe hypokalemia and risk of cardiac arrhythmias, patients with a serum potassium lower than 3.3 mEq/L should receive potassium replacement at a rate of 10 to 20 mEq per hour, with the insulin infusion delayed until potassium is 3 mEq/L or higher.

Bicarbonate. Several studies have evaluated the effect of alkalinization in patients with DKA and reported no advantage of bicarbonate therapy in improving cardiac and neurologic functions or in the rate of recovery of hyperglycemia and ketoacidosis.[131-139] Moreover, bicarbonate therapy can lead to higher risk of hypokalemia, decreased tissue oxygen uptake, and cerebral edema.[3] However, because severe acidosis may result in impaired myocardial contractility, cerebral vasodilatation, and coma, treatment guidelines recommend that, in patients with severe metabolic acidosis (pH <6.9), 50 to 100 mmol of sodium bicarbonate should be given as isotonic solution (in 200 mL of water) every 2 hours until pH rises to approximately 6.9 to 7.0. In patients with arterial pH greater than or equal to 7.0, no bicarbonate therapy is necessary.

Phosphate and Magnesium. Severe hypophosphatemia may lead to rhabdomyolysis, hemolytic uremia, and muscle weakness; however, several studies have failed to show any beneficial effect of phosphate replacement on clinical outcome. Furthermore, aggressive phosphate therapy is potentially hazardous, as indicated in case reports of children with DKA who developed hypocalcemia and seizures secondary to IV phosphate administration. Phosphate replacement should be reserved for patients cardiac dysfunction, anemia, respiratory depression, and a serum concentration lower than 1.0 to 1.5 mg/dL.[3]

Patients with DKA frequently have large magnesium deficits, but there are no data to determine whether replacement of magnesium is beneficial in improving clinical outcome. Replacement of magnesium should be considered in the occasional patient with severe hypomagnesemia and hypocalcemia. The recommended dose is 25 to 50 mg/kg for two to four doses given every 4 to 6 hours, with a maximum infusion rate of 2 g per hour.

Transition to Subcutaneous Insulin. IV insulin therapy should continue until the hyperglycemic crisis has resolved and the patient is eating and drinking. Criteria for resolution of DKA include a blood glucose lower than 200 mg/dL (11.1 mmol/L), a serum bicarbonate level equal to or greater than 18 mEq/L, and a venous pH greater than 7.3.[3] When this occurs, subcutaneous insulin therapy can be started.

It is important to recognize that the half-life of IV insulin is approximately 5 to 7 minutes; therefore, if insulin infusion is interrupted suddenly, the insulin concentration in blood could reach undetectable levels, with relapse of the hyperglycemia and ketoacidosis. Therefore, insulin infusion should be continued for 2 to 4 hours after subcutaneous insulin is started. Previous work showed that basal insulin given a few hours after starting insulin infusion facilitates treatment and prevents rebound hyperglycemia after discontinuation of IV insulin therapy.[140]

The best time to transition to subcutaneous insulin is once the patient is alert and able to take food by mouth. Patients previously treated with insulin can resume their previous regimen. Insulin-naïve adult patients can be started on 0.5 to 0.7 U/kg/day, in divided doses, to achieve adequate glucose control.[3,141] A basal prandial approach with insulin basal (e.g., glargine detemir, or degludec) and rapid-insulin (lispro, aspart, glulisine) analogs is the preferred treatment regimen after resolution of DKA. An alternative is the use of split-mixed insulin regimen with neutral protamine Hagedorn (NPH) and regular insulin twice daily or a multidose regimen of short- or rapid-acting and intermediate- or long-acting insulins.[3,141] A previous study compared the safety and efficacy of insulin analogs and human insulin during the transition from IV to subcutaneous insulin in patients with DKA. There were no differences in mean daily glucose levels, but 41% patients treated with NPH and regular insulin experienced one or more hypoglycemic events, compared with 15% of patients treated with insulin analogs. Thus, a basal bolus regimen with insulin analogs is safer and should be preferred over NPH and regular insulin following the resolution of DKA.[142]

Complications of Therapy

The two most common complications associated with the treatment of DKA in adult subjects are hypoglycemia and hypokalemia. Hypoglycemia is reported in 10% to 25% of patients during insulin therapy. Hypoglycemia is reported in 10% to 30% of patients with DKA, with the majority of hypoglycemic events occurring after several hours of insulin infusion (between 8 and 16 hours) or during the transition phase.[5] Failure to reduce the insulin infusion rate or to use dextrose-containing solutions when glucose levels reach 250 mg/dL are the two most common causes of hypoglycemia during insulin therapy. Frequent blood glucose monitoring during insulin infusion (every 1–2 hours) is mandatory to recognize hypoglycemia, because many patients with DKA do not experience adrenergic manifestations of sweating, nervousness, hunger, or tachycardia.

Other complications of diabetes include hyperchloremic acidosis with an excessive use of NaCl or potassium chloride, resulting in a nonanion gap metabolic acidosis. This acidosis has no adverse clinical effects and is gradually corrected over the subsequent 24 to 48 hours by enhanced renal acid excretion.[143] The development of hyperchloremia can be prevented with reduction of the chloride load by judicious use of hydration solutions.[40]

Cerebral edema is the most serious complication in children and has been reported in two studies to occur in approximately 1% or 0.3% to 0.9% of DKA episodes in children, respectively[144,145]; however, it is seldom reported in adult patients with DKA.[146] It is a serious complication in children, with a reported mortality of approximately 20% and permanent neurological morbidity. Symptoms and signs of cerebral edema are variable and include onset of headache, gradual deterioration in level of consciousness, seizures, sphincter incontinence, pupillary changes, papilledema, bradycardia, elevation in blood pressure, and respiratory arrest. Cerebral edema typically occurs 4 to 12 hours after treatment is activated, but it can be present before treatment has begun or may develop any time during treatment for DKA.[144] Although no single factor has been identified that can be used to predict the development of cerebral edema, a number of mechanisms have been proposed, including the role of cerebral ischemia and hypoxia, the generation of various inflammatory mediators, increased cerebral blood flow, disruption of cell membrane ion transport, and rapid shift in extracellular and intracellular fluids resulting in changes in osmolality. A large randomized controlled trial assessing 1389 episodes of DKA in 1255 children compared the use of rapid versus slow administration of either 0.9% or 0.45% NaCl solution. No differences in the rates of cerebral injury were seen, suggesting that the development of this complication may be idiosyncratic.[147]

Rhabdomyolysis may occur in patients with DKA, and more commonly with HHS, resulting in acute kidney failure, severe hyperkalemia, hypocalcemia, and muscle swelling causing compartment syndrome. The classic symptom triad of rhabdomyolysis includes myalgia, weakness, and dark urine, and monitoring creatine kinase concentrations every 2 to 3 hours is recommended for early detection.[148] In pediatric patients, rhabdomyolysis has been associated with a malignant hyperthermia–like syndrome that is not well understood.[149,150]

Prevention

KEY POINTS

- Education on 'sick day rules' and the provision of blood glucose and ketone testing equipment remain the most important interventions in preventing admission.

The most common precipitating causes of DKA include infection, intercurrent illness, psychological stress, and noncompliance with therapy. With improved outpatient treatment programs and better adherence to self-care, approximately 50% to 75% of DKA admissions may be preventable. In addition, provision of "sick day rules" can help when individuals are unwell. These state that insulin must never be stopped, even if the patient is not eating or drinking, and that frequent small sips of fluids should be continued to maintain hydration. Home capillary glucose should be done every 2 to 3 hours, and urine or capillary ketone measurements every few hours. Correction doses of insulin may be required. Outpatient management is more cost effective and can minimize missed days of school and work for patients with diabetes and their family members. The frequency of hospitalizations for DKA has been reduced following diabetes education programs, improved follow-up care, and access to medical advice from diabetes nurses or educators.[151] Many patients with recurrent DKA are unaware of sick-day management or the consequences of skipping or discontinuing insulin therapy.[152] It has been shown that quarterly visits to the endocrine clinic will reduce the number of emergency department admissions for DKA. An important feature of patient education is how to deal with illness. This includes: (1) communicating the importance of insulin therapy during illness and emphasizing that insulin should never be discontinued, (2) early contact with healthcare providers, (3) initiation of early management of fevers and infections, and (4) adequate fluid intake.[40] Diabetes education and sick-day management should be reviewed periodically in patients with T1DM and should include specific information on when to contact the healthcare provider, use of supplemental short- or rapid-acting insulin during illness, and, most importantly, the importance of never discontinuing insulin. If available, patients with T1DM should be instructed on the use of home blood ketone monitoring during illness and persistent hyperglycemia, which may allow for early recognition of impending ketoacidosis. β-hydroxybutyrate testing (capillary blood test) seems to be more effective than urine acetoacetate testing in reducing emergency department visits and hospitalizations.[47,92]

HYPERGLYCEMIC HYPEROSMOLAR STATE

Epidemiology

KEY POINTS

- HHS has a higher mortality rate that DKA because it most often occurs in an elderly population who have more co-morbidities. It is most frequent in Type 2 diabetes, but can occur in type 1 diabetes.

The term "hyperglycemic hyperosmolar state" was first coined in 1994.[1] This condition had previously been called several other things, including hyperglycemic hyperosmolar nonketotic coma and hyperosmolar nonacidotic diabetes. HHS is a serious and potentially lethal acute complication of diabetes.[94,153] The different forms of coma associated with diabetes were first described in 1886, with the distinction being made between the coma that comes on quickly, and all of the features of DKA, and the other associated with people who were overweight in whom symptoms came on slowly.[154] However, the importance of hypersomolality as an essential component of the syndrome was not recognized until the late 1950s.[155,156] Mortality in patients with HHS has decreased in recent decades, but remains higher than in DKA.[9,12,81,157-163] Previous work has reported a mortality rate of up to 20%.[11-13] An increased number of pediatric cases are being diagnosed, with a mortality rate as high as 37%.[164,165] Furthermore, HHS is not limited to people with T2DM. Data from a large European diabetes register showed that 31.6% of cases of HHS occurred in people with T1DM.[166]

Pathophysiology

KEY POINTS

- Infection, intercurrent illness, psychological factors and intentional or unintentional lack of compliance remain the most common precipitating factors for HHS. Other causes include drugs affecting carbohydrate metabolism and antipsychotic agents.

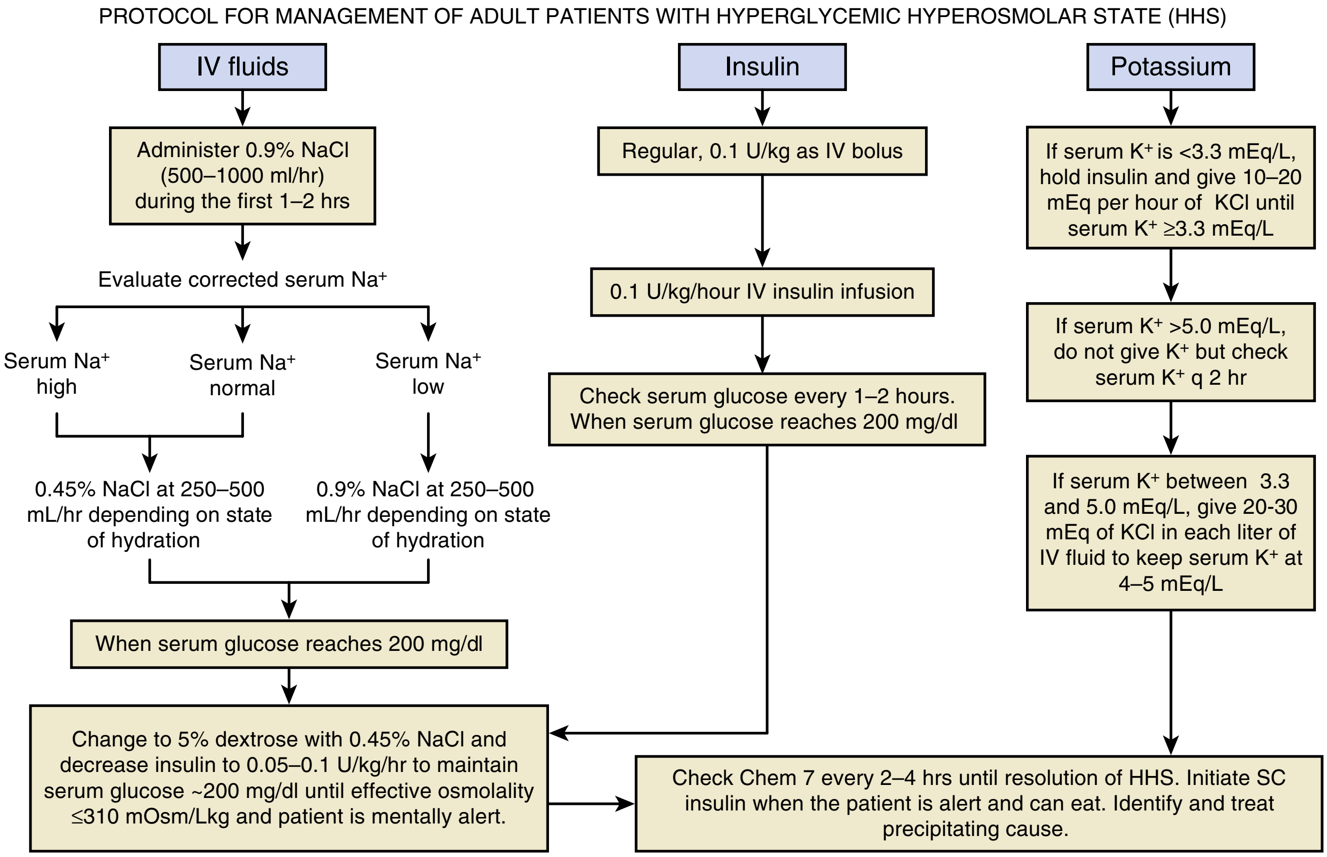
The pathophysiology of HHS is similar to that reported in DKA. Relative insulin deficiency and increased counterregulatory hormone concentrations leads to increase gluconeogenesis and decreased glucose uptake in peripheral tissues. The long duration and severity of hyperglycemia in HHS result in severe dehydration, hyperosmolality, and impaired renal function, which in turn lead to decreased excretion of glucose. It is speculated that higher levels of circulating insulin and lower levels of counterregulatory hormones in patients with HHS can explain the absent or minimal ketosis, which is the key difference from DKA. In general, three major mechanisms have been proposed for the lack of ketoacidosis in HHS: (1) higher levels of endogenous insulin in HHS (i.e., adequate insulin concentration to prevent lipolysis but inadequate to inhibit hepatic glucose production and/or stimulate glucose use), (2) lower levels of counterregulatory hormones and FFAs, and (3) inhibition of lipolysis by the hyperosmolar state, thereby decreasing ketogenesis.[1,40,158,167-169] However, HHS and DKA can occur in a mixed picture, with 27% of people presenting with a hyperglycemic emergency in one study having features of both at presentation.[94] Mortality was significantly high in this group, at 8% versus 3% for those with DKA and 5% for those with HHS ($P = .003$).[94]

Precipitating Events

KEY POINTS

- Infection, intercurrent illness, psychological factors and intentional or unintentional lack of compliance remain the most common precipitating factors for DKA. Other causes include drugs affecting carbohydrate metabolism and antipsychotic agents.

HHS occurs most commonly in older subjects with T2DM, but may be seen in younger and T1DM patients as well.[170] Up to 20% of patients admitted with HHS do not have a previous diagnosis of diabetes.[81] The most common precipitating causes are pneumonia and urinary tract infection, accounting for 30% to 50% of cases.[160,171,172] Other acute medical illnesses as precipitating causes include acute coronary syndromes, trauma, surgery, and cerebrovascular accident that provoke the release of counterregulatory hormones and/or compromise access to water. Certain medications that cause DKA may also precipitate the development of HHS, including glucocorticoids, thiazide diuretics, phenytoin, and β-blockers.[173] During the last few years, several case reports and retrospective studies have suggested an increased risk of developing diabetes mellitus in patients treated with atypical antipsychotics compared with schizophrenic patients treated with conventional antipsychotics or those without treatment.[174-178] Most cases of hyperglycemic crises associated with antipsychotics have been reported during treatment with clozapine and olanzapine.[179-181]


PROTOCOL FOR MANAGEMENT OF ADULT PATIENTS WITH HYPERGLYCEMIC HYPEROSMOLAR STATE (HHS)
IV fluids
Insulin
Potassium
Administer 0.9% NaCl (500–1000 ml/hr) during the first 1–2 hrs
Regular, 0.1 U/kg as IV bolus
If serum K^+ is <3.3 mEq/L, hold insulin and give 10–20 mEq per hour of KCl until serum K^+ ≥3.3 mEq/L
Evaluate corrected serum Na^+
0.1 U/kg/hour IV insulin infusion
If serum K^+ >5.0 mEq/L, do not give K^+ but check serum K^+ q 2 hr
Serum Na^+ high
Serum Na^+ normal
Serum Na^+ low
Check serum glucose every 1–2 hours. When serum glucose reaches 200 mg/dl
0.45% NaCl at 250–500 mL/hr depending on state of hydration
0.9% NaCl at 250–500 mL/hr depending on state of hydration
If serum K^+ between 3.3 and 5.0 mEq/L, give 20-30 mEq of KCl in each liter of IV fluid to keep serum K^+ at 4–5 mEq/L
When serum glucose reaches 200 mg/dl
Change to 5% dextrose with 0.45% NaCl and decrease insulin to 0.05–0.1 U/kg/hr to maintain serum glucose ~200 mg/dl until effective osmolality ≤310 mOsm/Lkg and patient is mentally alert.
Check Chem 7 every 2–4 hrs until resolution of HHS. Initiate SC insulin when the patient is alert and can eat. Identify and treat precipitating cause.

*Resolution of HHS = serum glucose < 250 mg/dl and effective plasma osmolality < 310 mOsm/kg

Fig. 37.3 Protocol for management of patients with hyperglycemic hyperosmolar state (*HHS*). *IV,* Intravenous; *K+*, potassium ion; *KCl*, potassium chloride; *Na+*, sodium ion; *NaCl*, sodium chloride; *SC,* subcutaneous.

Diagnosis

Symptoms and Signs. Although it is felt that the typical patient with HHS has undiagnosed diabetes, is between 55 and 70 years of age, and is frequently a nursing home resident, more recent data have suggested that there are few differences with respect to age and gender when compared with people presenting with DKA.[166] Indeed, approximately 1 in 230 people with T1DM presented with HHS in this large cohort.[166] However, most patients who develop HHS do so over days to weeks, during which they experience polyuria, dehydration, and progressive decline in the level of consciousness. Physical examination reveals signs of volume depletion. Fever because of underlying infection is common, and signs of acidosis (Kussmaul respiration, acetone breath) are usually absent. Gastrointestinal manifestations (abdominal pain, vomiting) frequently reported in patients with DKA are not part of HHS; thus, the presence of abdominal pain in patients without significant metabolic acidosis needs to be investigated.[80] In some patients, focal neurologic signs (hemiparesis, hemianopsia) and seizures (partial motor seizures more common than generalized) may be the dominant clinical features, resulting in a common misdiagnosis of stroke. Despite the focal nature of neurological findings, the neurologic manifestations often reverse completely after correction of the metabolic disorder. Neurologic symptoms of encephalopathy are commonly seen when serum sodium levels are in excess of 160 mEq/L or with calculated total and effective osmolalities higher than 340 and 320 mOsm/kg H_2O, respectively.

KEY POINTS

- HHS should be diagnosed when the glucose concentrations are >600mg/dl, without significant ketonuria or ketonemia, and no acidosis (pH>7.3). The calculated osmolality should be >320 mOsmol/Kg. The combination of HHS and DKA often occur together.

Laboratory Findings. The initial diagnostic criteria were reported by Gerich[158] and Arieff and Carroll[182] and included a blood glucose concentration greater than 600 mg/dL, a serum osmolality greater than 350 mOsm/L, and a serum acetone reaction no greater than 2+. In the original case series by Gerich[158] and Arieff and Carroll,[182] the mean plasma osmolarity in comatose patients was approximately 380 mOsml/L compared with approximately 320 to 330 mOsm/L in conscious subjects.[158,182,183] The importance of increased serum osmolality in the clinical presentation and prognosis of patients with HHS is well established. Altered sensorium (lethargy, stupor, coma) correlates better with hyperosmolality than with patient's age or severity of acid-base disturbance. The original formula to calculate serum osmolarity was Sosm = 2 [sodium (mEq/L)] + glucose (mg/dL)/18 + blood urea nitrogen (mg/dL)/2.8). In recent years, the total serum osmolality formula has been replaced by "effective" serum osmolality (Table 37.2).[40]

The ADA position statement diagnostic criteria for HHS are a plasma glucose concentration greater than 600 mg/dL, a serum osmolality greater than 320 mOsm/kg, and the absence of ketoacidosis[3] (Table 37.1). The UK definition is very similar.[184] Although by definition patients with HHS have a serum pH greater than 7.3, a serum bicarbonate greater than 18 mEq/L, and negative ketone bodies in urine and plasma, mild ketonemia may be present. Approximately 50% of patients with HHS have an increased anion gap metabolic acidosis as the result of concomitant ketoacidosis and/or an increase in serum lactate levels.[94]

Treatment

KEY POINTS

- Fluid replacement is the mainstay of treatment. Insulin should only be introduced once the glucose has stopped falling with fluids alone, to prevent rapid changes in osmolality.

At the same time as treating any underlying precipitating condition, the therapeutic measures for HHS are similar to those recommended for patients with DKA (Fig. 37.3). In general, the treatment of HHS should be directed at replacing volume deficit, correcting hyperosmolality and electrolyte disturbances, and managing the underlying illness that may have precipitated metabolic decompensation.[3,184] Patients with HHS may be severely dehydrated, with an average fluid deficit of 8 to 10 liters. Aggressive fluid replacement with normal saline at a rate of 1000 mL/hour for the first 2 to 3 hours is the usual recommendation, followed by 0.45% saline at a rate of 200 to 500 mL/hour. Insulin treatment does not need to be aggressive if fluid replacement is vigorously pursued. The UK guideline suggests that insulin should only be started once glucose concentrations have stopped falling with fluid resuscitation and then be started at 0.05 U/kg/hour.[184] This is different from the ADA guideline, which suggests that insulin be administered by an initial bolus of 0.1 U/kg followed by a continuous IV infusion calculated to deliver 0.1 U/kg/hour and continued at this rate until blood glucose has decreased to approximately 250 to 300 mg/dL. At this time, IV fluids should be changed to dextrose-containing solutions (5% dextrose), and the insulin dose should be decreased by 50% (0.05 U/kg/hour), or to 2 to 3 U/hour. Thereafter, the rate of insulin administration is adjusted to maintain a blood glucose level of approximately 200 mg/dL. Using this protocol, it has been reported that the mean duration of treatment in HHS for serum glucose levels to decrease into target range is approximately 11 hours.[81]

IV insulin infusion is usually continued until the patient is hemodynamically stable, the level of consciousness is improved, and the patient is able to tolerate food intake. Although most patients require insulin therapy after recovery from HHS, some patients could be managed with diet alone and/or diet plus an oral hypoglycemic agent during the initial admission or shortly after presentation.[1,185]

Increased serum potassium is commonly found in patients with HHS despite severe total body potassium depletion. Hyperglycemia and hyperosmolality cause a shift of potassium from the intracellular compartment into plasma and may contribute to a false estimate of total body potassium.[1,185] The serum potassium deficit in patients with HHS is estimated to be approximately 3 to 5 mEq/L of body weight.[1,185] The principles of potassium replacement in HHS are the same as in DKA. We recommend that potassium replacement be initiated after serum levels fall below 5.5 mEq/L, with the goal of maintaining a serum potassium concentration within the normal range of 4 to 5 mEq/L.

Complications

Venous thromboembolic events: Arterial and venous thrombosis have been described in patients with HHS.[186] Patients with T2DM are hypercoagulable compared with normal subjects.[187] However, uncomplicated diabetes is not an independent risk factor for incident venous thromboembolism (VTE).[188] The risk of thromboembolic events is higher in patients with HHS.[12,182,186,189-191] In a retrospective review of 426,831 cases of VTE, the overall incidence of thromboembolism among patients with hyperosmolarity was 1.7%.[189] Severe dehydration, along with hyperviscosity and hypotension, may predispose to intravascular clotting, and heparinization during the initial phase of treatment has been considered by some authors.[153,186,192] However, this has not been evaluated prospectively.

Rhabdomyolysis: Hyperosmolality alone could itself be a risk factor for rhabdomyolysis.[193,194] In 1981 Rumpf et al. described the association of acute rhabdomyolysis with myoglobinuric renal failure in a patient with HHS.[195] Since then it has been reported in adult and pediatric patients.[150] Subclinical rhabdomyolysis appears to be a common finding in HHS, and there seems to be a linear relationship between serum osmolarity, serum sodium, and the level of creatine kinase increase.[196] Rhabdomyolysis is a potentially life-threatening complication associated with acute kidney failure, multiorgan failure, severe hyperkalemia, and hypocalcemia leading to cardiac arrest, as well as muscle swelling causing compartment syndrome.[149,197]

Prevention

The most common precipitating causes of HHS are infection and intercurrent illness. Improved outpatient education and treatment programs, home capillary glucose monitoring, and better adherence to self-care should result in reduced admission of patients with hyperglycemic crises.

REFERENCES

1. Ennis ED, Stahl EJ, Kreisberg RA. The hyperosmolar hyperglycemic syndrome. *Diabetes Rev*. 1994;2:115.
2. Wilson JF. Diabetic ketoacidosis. *Ann Int Med*. 2010;152(1):ITC1–ITC16.
3. Kitabchi AE, Umpierrez GE, Miles JM, et al. Hyperglycemic crises in adult patients with diabetes. *Diabetes Care*. 2009;32:1335–1343.
4. Dhatariya KK, Glaser NS, Codner E, et al. Diabetic ketoacidosis. *Nat Rev Dis Primers*. 2020;6:40.
5. Dhatariya KK, Nunney I, Higgins K, et al. A national survey of the management of diabetic ketoacidosis in the UK in 2014. *Diabet Med*. 2016;33:252–260.
6. Umpierrez G, Korytkowski M. Diabetic emergencies - ketoacidosis, hyperglycaemic hyperosmolar state and hypoglycaemia. *Nat Rev Endocrinol*. 2016;12:222–232.
7. Azevedo LC, Choi H, Simmonds K, et al. Incidence and long-term outcomes of critically ill adult patients with moderate-to-severe diabetic ketoacidosis: retrospective matched cohort study. *J Crit Care*. 2014;29:971–977.
8. Agarwal A, Yadav A, Gutch M, et al. Prognostic factors in patients hospitalized with diabetic ketoacidosis. *Endocrinol Metab*. 2016;31:424–432.
9. Chung ST, Perue GG, Johnson A, et al. Predictors of hyperglycaemic crises and their associated mortality in Jamaica. *Diabet Res Clin Pract*. 2006;73:184–190.
10. Otieno CF, Kayima JK, Omonge EO, et al. Diabetic ketoacidosis: risk factors, mechanisms and management strategies in sub-Saharan Africa: a review. *East Afr Med J*. 2005;82:S197–S203.
11. Karslioglu French E, Donihi AC, et al. Diabetic ketoacidosis and hyperosmolar hyperglycemic syndrome: review of acute decompensated diabetes in adult patients. *BMJ*. 2019;365:l1114.
12. Fadini GP, De Kreutzenberg SV, Rigato M, et al. Characteristics and outcomes of the hyperglycemic hyperosmolar non-ketotic syndrome in a cohort of 51 consecutive cases at a single center. *Diabet Res Clin Pract*. 2011;94:172–179.
13. Pasquel FJ, Umpierrez GE. Hyperosmolar hyperglycemic state: a historic review of the clinical presentation, diagnosis, and treatment. *Diabet Care*. 2014;37:3124–3131.

14. Benoit SR, Zhang Y, Geiss LS, et al. Trends in diabetic ketoacidosis hospitalizations and in-hospital mortality - United States, 2000-2014. *Morb Mort Wkly Rep*. 2018;67:362–365.
15. Desai D, Mehta D, Mathias P, et al. Health care utilization and burden of diabetic ketoacidosis in the U.S. over the past decade: a nationwide analysis. *Diabet Care*. 2018;41:1631–1638.
16. Centers for Disease Control and Prevention. National diabetes statistics report, 2017. Estimates of diabetes and its burden in the United States. Accessed at: https://dev.diabetes.org/sites/default/files/2019-06/cdc-statistics-report-2017.pdf.
17. Venkatesh B, Pilcher D, Prins J, et al. Incidence and outcome of adults with diabetic ketoacidosis admitted to ICUs in Australia and New Zealand. *Crit Care*. 2016;19:451.
18. Zhong VW, Juhaeri J, Mayer-Davis EJ. Trends in hospital admission for diabetic ketoacidosis in adults with type 1 and type 2 diabetes in England, 1998-2013: a retrospective cohort study. *Diabet Care*. 2018;41: 1870–1877.
19. Vellanki P, Umpierrez GE. Increasing hospitalizations for DKA: a need for prevention programs. *Diabet Care*. 2018;41:1839–1841.
20. Benoit SR, Hora I, Pasquel FJ, et al. Trends in emergency department visits and inpatient admissions for hyperglycemic crises in adults with diabetes in the U.S., 2006 - 2015. *Diabet Care*. 2020;43:1057–1064.
21. Farsani SF, Brodovicz K, Soleymanlou N, et al. Incidence and prevalence of diabetic ketoacidosis (DKA) among adults with type 1 diabetes mellitus (T1D): a systematic literature review. *BMJ Open*. 2017;7:e016587.
22. Dhatariya KK, Skedgel C, Fordham R. The cost of treating diabetic ketoacidosis in the UK: a national survey of hospital resource use. *Diabet Med*. 2017;34:1361–1366.
23. Dhatariya KK, Parsekar K, Skedgel C, et al. The cost of treating diabetic ketoacidosis in an adolescent population in the UK: a national survey of hospital resource use. *Diabet Med*. 2019;36:982–987.
24. Prevention CfDCa. Centers for Disease Control and Prevention National diabetes statistics report 2020. Estimates of diabetes and its burden in the United States. Accessed at: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf.
25. Diaz-Valencia PA, Bougneres P, Valleron AJ. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. *BMC Public Health*. 2015;15:255.
26. Li J, Yang D, Yan J, et al. Secondary diabetic ketoacidosis and severe hypoglycaemia in patients with established type 1 diabetes mellitus in China: a multicentre registration study. *Diabet Metab Res Rev*. 2014;30:497–504.
27. Dabelea D, Rewers A, Stafford JM, et al. Trends in the prevalence of ketoacidosis at diabetes diagnosis: the SEARCH for Diabetes in Youth Study. *Pediatrics*. 2014;133:e938–e945.
28. Rewers A, Klingensmith G, Davis C, et al. Presence of diabetic ketoacidosis at diagnosis of diabetes mellitus in youth: the Search for Diabetes in Youth Study. *Pediatrics*. 2008;121:e1258–e1266.
29. Davis SN, Umpierrez GE. Diabetic ketoacidosis in type 2 diabetes mellitus—pathophysiology and clinical presentation. *Nat Clin Pract Endocrinol Metab*. 2007;3:730–731.
30. Balasubramanyam A, Yajnik CS, Tandon N. Non-traditional forms of diabetes worldwide: Implications for translational investigation. *Transl Endocrinol Metab*. 2011;2:43–67.
31. Umpierrez GE, Woo W, Hagopian WA, et al. Immunogenetic analysis suggests different pathogenesis for obese and lean African-Americans with diabetic ketoacidosis. *Diabet Care*. 1999;22:1517–1523.
32. Umpierrez GE, Casals MM, Gebhart SP, et al. Diabetic ketoacidosis in obese African-Americans. *Diabetes*. 1995;44:790–795.
33. Banerji MA, Chaiken RL, Huey H, et al. GAD antibody negative NIDDM in adult black subjects with diabetic ketoacidosis and increased frequency of human leukocyte antigen DR3 and DR4. Flatbush diabetes. *Diabetes*. 1994;43:741–745.
34. Likitmaskul S, Santiprabhob J, Sawathiparnich P, et al. Clinical pictures of type 2 diabetes in Thai children and adolescents is highly related to features of metabolic syndrome. *J Med Assoc Thai*. 2005;88:S169–S175.
35. Yamada K, Nonaka K. Diabetic ketoacidosis in young obese Japanese men. *Diabet Care*. 1996;19:671.
36. Maldonado MR, Otiniano ME, Lee R, et al. Ethnic differences in beta-cell functional reserve and clinical features in patients with ketosis-prone diabetes. *Diabet Care*. 2003;26:2469.
37. Tan KC, Mackay IR, Zimmet PZ, et al. Metabolic and immunologic features of Chinese patients with atypical diabetes mellitus. *Diabet Care*. 2000;23:335–338.
38. Balasubramanyam A, Nalini R, Hampe CS, et al. Syndromes of ketosis-prone diabetes mellitus. *Endocr Rev*. 2008;29:292–302.
39. Umpierrez GE. Ketosis-prone type 2 diabetes. *Diabet Care*. 2006;29:2755–2757.
40. Kitabchi AE, Umpierrez GE, Murphy MB, et al. Management of hyperglycemic crises in patients with diabetes. *Diabet Care*. 2001;24:131–153.
41. White NH. Diabetic ketoacidosis in children. *Endocrinol Metab Clin North Am*. 2000;29:657–682.
42. Unger RH. Role of glucagon in the pathogenesis of diabetes: the status of the controversy. *Metabolism*. 1978;27:1691–1709.
43. Miles JM, Haymond MW, Nissen SL, et al. Effects of free fatty acid availability, glucagon excess, and insulin deficiency on ketone body production in postabsorptive man. *J Clin Investig*. 1983;71:1554–1561.
44. McGarry JD, Foster DW. Regulation of hepatic fatty acid oxidation and ketone body production. *Annu Rev Biochem*. 1980;49:395–420.
45. McGarry JD, Woeltje KF, Kuwajima M, et al. Regulation of ketogenesis and the renaissance of carnitine palmitoyltransferase. *Diabet Metab Rev*. 1989;5:271–284.
46. Laffel L. Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. *Diabet Metab Res Rev*. 1999;15:412–426.
47. Dhatariya K. Blood ketones—measurement, interpretation, limitations and utility in the management of diabetic ketoacidosis. *Rev Diabet Stud*. 2016;13:217–225.
48. Miles JM, Rizza RA, Haymond MW, et al. Effects of acute insulin deficiency on glucose and ketone body turnover in man: evidence for the primacy of overproduction of glucose and ketone bodies in the genesis of diabetic ketoacidosis. *Diabetes*. 1980;29:926–930.
49. Gerich JE, Meyer C, Woerle HJ, et al. Renal gluconeogenesis. *Diabetes Care*. 2001;24:382–391.
50. Exton JH. Gluconeogenesis. *Metab Clin Exp*. 1972;21:945–990.
51. Foster DW, McGarry JD. The metabolic derangements and treatment of diabetic ketoacidosis. *New Engl J Med*. 1983;309:159–169.
52. van de Werve G, Jeanrenaud B. Liver glycogen metabolism: an overview. *Diabet Metab Rev*. 1987;3:47–78.
53. Herzig S, Long F, Jhala US, et al. CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature*. 2001;413:179–183.
54. Yoon JC, Puigserver P, Chen G, et al. Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature*. 2001;413:131–138.
55. Pilkis SJ, Granner DK. Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis. *Annu Rev Physiol*. 1992;54:885–909.
56. Gerich JE, Lorenzi M, Bier DM, et al. Prevention of human diabetic ketoacidosis by somatostatin. Evidence for an essential role of glucagon. *New Engl J Med*. 1975;292:985–989.
57. Gerich JE, Lorenzi M, Bier DM, et al. Effects of physiologic levels of glucagon and growth hormone on human carbohydrate and lipid metabolism. Studies involving administration of exogenous hormone during suppression of endogenous hormone secretion with somatostatin. *J Clin Investig*. 1976;57:875–884.
58. Exton JH. Mechanisms of hormonal regulation of hepatic glucose metabolism. *Diabet Metab Rev*. 1987;3:163–183.
59. Felig P, Sherwin RS, Soman V, et al. Hormonal interactions in the regulation of blood glucose. *Recent Prog Horm Res*. 1979;35:501–532.
60. Schade DS, Eaton RP. The temporal relationship between endogenously secreted stress hormones and metabolic decompensation in diabetic man. *J Clin Endocrinol Metab*. 1980;50:131–136.
61. Chaudhuri A, Umpierrez GE. Oxidative stress and inflammation in hyperglycemic crises and resolution with insulin: implications for the acute and chronic complications of hyperglycemia. *J Diabetes Complicat*. 2012;26:257–258.

62. Stentz FB, Umpierrez GE, Cuervo R, et al. Proinflammatory cytokines, markers of cardiovascular risks, oxidative stress, and lipid peroxidation in patients with hyperglycemic crises. *Diabetes*. 2004;53:2079–2086.
63. Li J, Huang M, Shen X. The association of oxidative stress and pro-inflammatory cytokines in diabetic patients with hyperglycemic crisis. *J Diabetes Complicat*. 2014;28:662–666.
64. Hotamisligil GS. Inflammation and metabolic disorders. *Nature*. 2006;444:860–867.
65. Hotamisligil GS, Peraldi P, Budavari A, et al. IRS –1 mediated inhibition of insulin receptor tyrosine kinase activity in TNF alpha and obesity induced insulin resistance. *Science*. 1996;271:665–668.
66. Isales CM, Min L, Hoffman WH. Acetoacetate and beta-hydroxybutyrate differentially regulate endothelin-1 and vascular endothelial growth factor in mouse brain microvascular endothelial cells. *J Diabetes Complicat*. 1999;13:91–97.
67. Hoffman WH, Cheng C, Passmore GG, et al. Acetoacetate increases expression of intercellular adhesion molecule-1 (ICAM-1) in human brain microvascular endothelial cells. *Neurosci Lett*. 2002;334:71–74.
68. Mays JA, Jackson KL, Derby TA, et al. An evaluation of recurrent diabetic ketoacidosis, fragmentation of care, and mortality across Chicago, Illinois. *Diabetes Care*. 2016;39:1671–1676.
69. Skinner TC. Recurrent diabetic ketoacidosis: causes, prevention and management. *Hormone Res*. 2002;57:78–80.
70. Randall L, Begovic J, Hudson M, et al. Recurrent diabetic ketoacidosis in inner-city minority patients. *Diabetes Care*. 2011;34:1891–1896.
71. United States Food and Drug Administration. FDA briefing document. Endocrinologic and metabolic drugs advisory committee meeting 17th January 2019. Accessed at: https://www.fda.gov/media/121284/download.
72. Macfarlane J, Dhatariya K. The incidence of euglycemic diabetic ketoacidosis in adults with type 1 diabetes in the UK prior to the widespread use of sodium glucose co-transporter 2 inhibitors. *Mayo Clin Proc*. 2019;94:1909–1910.
73. Janssens B, Caerels S, Mathieu C. SGLT inhibitors in type 1 diabetes: weighing efficacy and side effects. *Ther Adv Endocrinol Metab*. 2020;11: 2042018820938545.
74. Rosenstock J, Marquard J, Laffel LM, et al. Empagliflozin as adjunctive to insulin therapy in type 1 diabetes: The EASE trials. *Diabetes Care*. 2018;41:2560–2569.
75. de Veciana M. Diabetes ketoacidosis in pregnancy. *Semin Perinatol*. 2013;37:267–273.
76. Prowse CM, Gaensler EA. Respiratory and acid-base changes during pregnancy. *Anesthesiology*. 1965;26:381–392.
77. Morris AD, Boyle DI, McMahon AD, et al. Adherence to insulin treatment, glycaemic control, and ketoacidosis in insulin-dependent diabetes mellitus. *Lancet*. 1997;350:1505–1510.
78. Butalia S, Johnson JA, Ghali WA, et al. Clinical and socio-demographic factors associated with diabetic ketoacidosis hospitalization in adults with type 1 diabetes. *Diabet Med*. 2013;30:567–573.
79. Torjesen I. Diabulimia: the world's most dangerous eating disorder. *BMJ*. 2019;364:l982.
80. Umpierrez G, Freire AX. Abdominal pain in patients with hyperglycemic crises. *J Crit Care*. 2002;17:63–67.
81. Umpierrez GE, Kelly JP, Navarrete JE, et al. Hyperglycemic crises in urban blacks. *Arch Int Med*. 1997;157:669–675.
82. Nambu T, Mori K, Shinoto Y, et al. Diabetic ketoacidosis accompanied by hypothermia: a case report. *Diabetes Res Clin Pract*. 2012;96:326–330.
83. Gale EA, Tattersall RB. Hypothermia: a complication of diabetic ketoacidosis. *Br Med J*. 1978;2:1387–1389.
84. Dhatariya KK, Umpierrez GE. Guidelines for management of diabetic ketoacidosis: time to revise? *Lancet Diabetes Endocrinol*. 2017;5:321–323.
85. Jenkins D, Close CF, Krentz AJ, et al. Euglycaemic diabetic ketoacidosis: does it exist? *Acta Diabetologica*. 1993;30:251–253.
86. Modi A, Agrawal A, Morgan F. Euglycemic diabetic ketoacidosis: a review. *Curr Diabetes Rev*. 2017;13:315–321.
87. Munro JF, Campbell IW, McCuish AC, et al. Euglycaemic diabetic ketoacidosis. *Br Med J*. 1973;2:578–580.
88. Umpierrez GE, DiGirolamo M, Tuvlin JA, et al. Differences in metabolic and hormonal milieu in diabetic- and alcohol-induced ketoacidosis. *J Crit Care*. 2000;15:52–59.
89. Stephens JM, Sulway MJ, Watkins PJ. Relationship of blood acetoacetate and 3-hydroxybutyrate in diabetes. *Diabetes*. 1971;20:485–489.
90. Sheikh-Ali M, Karon BS, Basu A, et al. Can serum beta-hydroxybutyrate be used to diagnose diabetic ketoacidosis? *Diabetes Care*. 2008;31:643–647.
91. Laffel LM, Wentzell K, Loughlin C, et al. Sick day management using blood 3-hydroxybutyrate (3-OHB) compared with urine ketone monitoring reduces hospital visits in young people with T1DM: a randomized clinical trial. *Diabet Med*. 2006;23:278–284.
92. Klocker AA, Phelan H, Twigg SM, et al. Blood ß-hydroxybutyrate vs. urine acetoacetate testing for the prevention and management of ketoacidosis in type 1 diabetes: a systematic review. *Diabet Med*. 2013;30:818–824.
93. Savage MW, Dhatariya KK, Kilvert A, et al. Joint British Diabetes Societies guideline for the management of diabetic ketoacidosis. *Diabetic Med*. 2011;28:508–515.
94. Pasquel FJ, Tsegka K, Wang H, et al. Clinical outcomes in patients with isolated or combined diabetic ketoacidosis and hyperosmolar hyperglycemic state: a retrospective, hospital-based cohort study. *Diabetes Care*. 2020;43:349–357.
95. Adrogue HJ, Eknoyan G, Suki WK. Diabetic ketoacidosis: role of the kidney in the acid-base homeostasis re-evaluated. *Kidney Int*. 1984;25: 591–598.
96. Sottosanti M, Morrison GC, Singh RN, et al. Dehydration in children with diabetic ketoacidosis: a prospective study. *Arch Dis Child*. 2012;97:96–100.
97. Nabarro JD, Spencer AG, Stowers JM. Metabolic studies in severe diabetic ketosis. *Q J Med*. 1952;21:225–248.
98. Atchley DW, Loeb RF, Richards DW, et al. On diabetic acidosis: a detailed study of electrolyte balances following the withdrawal and reestablishment of insulin therapy. *J Clin Investig*. 1933;12:297–326.
99. Fagan MJ, Avner J, Khine H. Initial fluid resuscitation for patients with diabetic ketoacidosis: how dry are they? *Clin Pediatr (Phila)*. 2008;47:851–855.
100. Mackenzie A, Barnes G, Shann F. Clinical signs of dehydration in children. *Lancet*. 1989;334:605–607.
101. Koves IH, Neutze J, Donath S, et al. The accuracy of clinical assessment of dehydration during diabetic ketoacidosis in childhood. *Diabetes Care*. 2004;27:2485–2487.
102. Xu W, Wu HF, Ma SG, et al. Correlation between peripheral white blood cell counts and hyperglycemic emergencies. *Int J Med Sci*. 2013;10:758–765.
103. Palmer BF, Clegg DJ. Electrolyte disturbances in patients with chronic alcohol-use disorder. *New Eng J Med*. 2017;377:1368–1377.
104. Cahill GF. Fuel metabolism in starvation. *Annu Rev Nutr*. 2006;26: 1–22.
105. Owen OE. Ketone bodies as a fuel for the brain during starvation. *Biochem Mol Biol Educ*. 2005;33:246–251.
106. Wildenhoff KE, Ladefoged K, Sorensen NS. Clinical physiology: the concentration of ketone bodies, free fatty acids, and glycerol in the blood of obese persons after injection of insulin and glucose studies before and during absolute fasting. *Scandinavian J Clin Lab Investig*. 1975;35:129–133.
107. Cahill GF. Starvation in man. *New Eng J Med*. 1970;282:668–675.
108. Clark A, Kraut E, Yen HH, et al. Evaluation of a diabetic ketoacidosis order set in adults with type 1 and type 2 diabetes at a tertiary academic medical centre: a retrospective chart audit. *Can J Diabetes*. 2019;43:304–308.e303.
109. Khan A, Stewart V, Meakin O, et al. Comparison of fixed vs variable rate infusion in management of diabetic keto-acidosis (DKA): experience from three UK centres. *Diabet Med*. 2019;36:P334.
110. Galm BP, Bagshaw SM, Senior PA. Acute management of diabetic ketoacidosis in adults at 3 teaching hospitals in Canada: a multicentre, retrospective cohort study. *Can J Diabetes*. 2019;43:309–315.e302.
111. Balmier A, Dib F, Serret-Larmande A, et al. Initial management of diabetic ketoacidosis and prognosis according to diabetes type: a French multicentre observational retrospective study. *Ann Intensive Care*. 2019;9:91.

112. Mohamed M, Assal A, Boyle L, et al. Development and implementation of a DKA protocol for adults with type 1 and type 2 diabetes at a tertiary care multi-campus hospital. *Can J Diabetes*. 2018;43:256–260.e.253.
113. Kempegowda P, Chandan JS, Coombs B, et al. Regular performance feedback may be key to maintain good quality DKA management: results from a five-year study. *BMJ Open Diabetes Res Care*. 2019;7:e000695.
114. May ME, Young C, King J. Resource utilization in treatment of diabetic ketoacidosis in adults. *Am J Med Sci*. 1993;306:287–294.
115. Moss JM. Diabetic ketoacidosis: effective low-cost treatment in a community hospital. *South Med J*. 1987;80:875–881.
116. Umpierrez GE, Latif K, Stoever J, et al. Efficacy of subcutaneous insulin lispro versus continuous intravenous regular insulin for the treatment of patients with diabetic ketoacidosis. *Am J Med*. 2004;117:291–296.
117. Umpierrez GE, Cuervo R, Karabell A, et al. Treatment of diabetic ketoacidosis with subcutaneous insulin aspart. *Diabetes Care*. 2004;27:1873–1878.
118. Javor KA, Kotsanos JG, McDonald RC, et al. Diabetic ketoacidosis charges relative to medical charges of adult patients with type I diabetes. *Diabetes Care*. 1997;20:349–354.
119. Glaser NS, Ghetti S, Casper TC, et al. Pediatric diabetic ketoacidosis, fluid therapy, and cerebral injury: the design of a factorial randomized controlled trial. *Pediatr Diabetes*. 2013;14:435–446.
120. Gershkovich B, English SW, Doyle MA, et al. Choice of crystalloid fluid in the treatment of hyperglycemic emergencies: a systematic review protocol. *Syst Rev*. 2019;8:228.
121. Van Zyl DG, Rheeder P, Delport E. Fluid management in diabetic-acidosis—Ringer's lactate versus normal saline: a randomized controlled trial. *QJM*. 2012;105:337–343.
122. Chua HR, Venkatesh B, Stachowski E, et al. Plasma-Lyte 148 vs 0.9% saline for fluid resuscitation in diabetic ketoacidosis. *J Crit Care*. 2012;27:138–145.
123. Kitabchi AE, Umpierrez GE, Fisher JN, et al. Thirty years of personal experience in hyperglycemic crises: diabetic ketoacidosis and hyperglycemic hyperosmolar state. *J Clin Endocrinol Metab*. 2008;93:1541–1552.
124. Fisher JN, Shahshahani MN, Kitabchi AE. Diabetic ketoacidosis: low-dose insulin therapy by various routes. *New Eng J Med*. 1977;297:238–241.
125. Kidson W, Casey J, Kraegen E, et al. Treatment of severe diabetes mellitus by insulin infusion. *Br Med J*. 1974;2:691–694.
126. Page MM, Alberti KG, Greenwood R, et al. Treatment of diabetic coma with continuous low-dose infusion of insulin. *Br Med J*. 1974;2:687–690.
127. Semple PF, White C, Manderson WG. Continuous intravenous infusion of small doses of insulin in treatment of diabetic ketoacidosis. *Br Med J*. 1974;2:694–698.
128. Kitabchi AE, Murphy MB, Spencer J, et al. Is a priming dose of insulin necessary in a low-dose insulin protocol for the treatment of diabetic ketoacidosis? *Diabetes Care*. 2008;31:2081–2085.
129. Palermo NE, Sadhu AR, McDonnell ME. Diabetic ketoacidosis in COVID-19: unique concerns and considerations. *J Clin Endocrinol Metab*. 2020;105:2819–2829.
130. Rayman G, Lumb A, Kennon B, et al. Guidance on the management of diabetic ketoacidosis in the exceptional circumstances of the COVID-19 pandemic. *Diabetic Med*. 2020;37:1214–1216.
131. Viallon A, Zeni F, Lafond P, et al. Does bicarbonate therapy improve the management of severe diabetic ketoacidosis? *Crit Care Med*. 1999;27:2690–2693.
132. Green SM, Rothrock SG, Ho JD, et al. Failure of adjunctive bicarbonate to improve outcome in severe pediatric diabetic ketoacidosis. *Ann Emerg Med*. 1998;31:41–48.
133. Okuda Y, Adrogue HJ, Field JB, et al. Counterproductive effects of sodium bicarbonate in diabetic ketoacidosis. *J Clin Endocrinol Metab*. 1996;81:314–320.
134. Morris LR, Murphy MB, Kitabchi AE. Bicarbonate therapy in severe diabetic ketoacidosis. *Ann Int Med*. 1986;105:836–840.
135. Gamba G, Osequera J, Casterjon M, et al. Bicarbonate therapy in severe diabetic ketoacidosis. A double blind, randomized, placebo controlled trial. *Rev Invest Clin*. 1991;43:234–238.
136. Hale PJ, Crase JE, Nattrass M. Metabolic effects of bicarbonate in the treatment of diabetic ketoacidosis. *Br Med J*. 1984;290:1035–1038.
137. Lever E, Jaspan JB. Sodium bicarbonate therapy in severe diabetic ketoacidosis. *Am J Med*. 1983;75:263–268.
138. Latif KA, Freire AX, Kitabchi AE, et al. The use of alkali therapy in severe diabetic ketoacidosis. *Diabetes Care*. 2002;25:2113–2114.
139. Assal JP, Aoki TT, Manzano FM, et al. Metabolic effects of sodium bicarbonate in management of diabetic ketoacidosis. *Diabetes*. 1974;23:405–411.
140. Hsia E, Seggelke S, Gibbs J, et al. Subcutaneous administration of glargine to diabetic patients receiving insulin infusion prevents rebound hyperglycemia. *J Clin Endocrinol Metab*. 2012;97:3132–3137.
141. George S, Dale J, Stanisstreet D. A guideline for the use of variable rate intravenous insulin infusion in medical inpatients. *Diabet Med*. 2015;32:706–713.
142. Umpierrez GE, Jones S, Smiley D, et al. Insulin analogs versus human insulin in the treatment of patients with diabetic ketoacidosis. *Diabetes Care*. 2009;32:1164–1169.
143. Dhatariya KK. Diabetic ketoacidosis. *Br Med J*. 2007;334:1284–1285.
144. Glaser N, Barnett P, McCaslin I, et al. Risk factors for cerebral edema in children with diabetic ketoacidosis. *New Eng J Med*. 2001;344:264–269.
145. Edge JA, Jakes RW, Roy Y, et al. The UK case–control study of cerebral oedema complicating diabetic ketoacidosis in children. *Diabetologia*. 2006;49:2002–2009.
146. Wolfsdorf J, Glaser N, Sperling MA. Diabetic ketoacidosis in infants, children, and adolescents: a consensus statement from the American Diabetes Association. *Diabetes Care*. 2006;29:1150–1159.
147. Kuppermann N, Ghetti S, Schunk JE, et al. Clinical trial of fluid infusion rates for pediatric diabetic ketoacidosis. *New Eng J Med*. 2018;378:2275–2287.
148. Kutlu AO, Kara C, Cetinkaya S. Rhabdomyolysis without detectable myoglobulinuria due to severe hypophosphatemia in diabetic ketoacidosis. *Pediatr Emerg Care*. 2011;27:537–538.
149. Hollander AS, Olney RC, Blackett PR, et al. Fatal malignant hyperthermia-like syndrome with rhabdomyolysis complicating the presentation of diabetes mellitus in adolescent males. *Pediatrics*. 2003;111:1447–1452.
150. Casteels K, Beckers D, Wouters C, et al. Rhabdomyolysis in diabetic ketoacidosis. *Pediatr Diabetes*. 2003;4:29–31.
151. Evans NR, Richardson L, Dhatariya KK, et al. Diabetes specialist nurse telemedicine: admissions avoidance, costs and casemix. *Pract Diabetes*. 2012;29:25–28.
152. Laffel L. Sick-day management in type 1 diabetes. *Endocrinol Metab Clin North Am*. 2000;29:707–723.
153. Nyenwe EA, Kitabchi AE. Evidence-based management of hyperglycemic emergencies in diabetes mellitus. *Diabet Res Clin Pract*. 2011;94:340–351.
154. Dreschfeld J. The Bradshawe lecture on diabetic coma. *Br Med J*. 1886;2:358–363.
155. De Graeff J, Lips JB. Hypernatraemia in diabetes mellitus. *Acta Medica Scandinavica*. 1957;157:71–75.
156. Sament S, Schwartz MB. Severe diabetic stupor without ketosis. *South Afr Med J*. 1957;31:893–894.
157. Jackson WP, Forman R. Hyperosmolar nonketotic diabetic coma. *Diabetes*. 1966;15:714–722.
158. Gerich JE, Martin MM, Recant L. Clinical and metabolic characteristics of hyperosmolar nonketotic coma. *Diabetes*. 1971;20:228–238.
159. Arieff AI, Carroll HJ. Nonketotic Hyperosmolar Coma with hyperglycaemia: clinical features, pathophysiology, renal function, acid base balance, plasma cerebrospinal fluid equilibria and the effects of therapy in 37 cases. *Medicine*. 1972;51:73–94.
160. Wachtel TJ, Tetu-Mouradijian LM, Goldman DL, et al. Hyperosmolarity and acidosis in diabetes mellitus: a three-year experience in Rhode Island. *J Gen Int Med*. 1991;6:495–502.
161. Pinies JA, Cairo G, Gaztambide S, et al. Course and prognosis of 132 patients with diabetic non ketotic hyperosmolar state. *Diabete et Metab*. 1994;20:43–48.
162. Macisaac RJ, Lee LY, McNeil KJ, et al. Influence of age on the presentation and outcome of acidotic and hyperosmolar diabetic emergencies. *Int Med J*. 2002;32:379–385.

163. Zouvanis M, Pieterse AC, Seftel HC, et al. Clinical characteristics and outcome of hyperglycaemic emergencies in Johannesburg Africans. *Diabet Med*. 1997;14:603–606.
164. Rosenbloom AL. Hyperglycemic hyperosmolar state: An emerging pediatric problem. *J Pediatr*. 2010;156:180–184.
165. Canarie MF, Bogue CW, Banasiak KJ, et al. Decompensated hyperglycemic hyperosmolarity without significant ketoacidosis in the adolescent and young adult population. *J Pediatr Endocrinol Metab*. 2007;20:1115–1124.
166. Tittel SR, Sondern KM, Weyer M, et al. Multicentre analysis of hyperglycaemic hyperosmolar state and diabetic ketoacidosis in type 1 and type 2 diabetes. *Acta Diabetol*. 2020;57:1245–1253.
167. Delaney MF, Zisman A, Kettyle WM. Diabetic ketoacidosis and hyperosmolar nonketotic syndrome. *Endocrinol Metab Clin North Am*. 2000;39:683–705.
168. Umpierrez GE, Khajavi M, Kitabchi AE. Review: diabetic ketoacidosis and hyperglycemic hyperosmolar nonketotic syndrome. *Am J Med Sci*. 1996;311:225–233.
169. Chupin M, Charbonnel B, Chupin F. C-peptide blood levels in ketoacidosis and in hyperosmolar non-ketotic diabetic coma. *Acta Diabetol Latina*. 1981;18:123–128.
170. Klingensmith GJ, Connor CG, Ruedy KJ, et al. Presentation of youth with type 2 diabetes in the Pediatric Diabetes Consortium. *Pediatr Diabetes*. 2016;17:266–273.
171. Wachtel TJ. The diabetic hyperosmolar state. *Clin Geriatr Med*. 1990;6:797–806.
172. Kitabchi AE, Wall BM. Diabetic ketoacidosis. *Med Clin North Am*. 1995;79:9–37.
173. Lorber D. Nonketotic hypertonicity in diabetes mellitus. *Med Clin North Am*. 1995;79:39–52.
174. Marder SR, Cannon TD. *Schizophrenia. N Engl J Med.* 2019;381:1753–1761.
175. Holt RI. Association between antipsychotic medication use and diabetes. *Curr Diabetes Rep*. 2019;19:96.
176. Tavakoli SA, Arguisola MS. Diabetic ketoacidosis in a patient treated with olanzapine, valproic acid, and venlafaxine. *South Med J*. 2003;96:729–730.
177. Ananth J, Parameswaran S, Gunatilake S. Side effects of atypical antipsychotic drugs. *Curr Pharm Des*. 2004;10:2219–2229.
178. Wilson DR, D'Souza L, Sarkar N, et al. New-onset diabetes and ketoacidosis with atypical antipsychotics. *Schizophr Res*. 2003;59:1–6.
179. Nihalani ND, Tu X, Lamberti JS, et al. Diabetic ketoacidosis among patients receiving clozapine: a case series and review of socio-demographic risk factors. *Ann Clin Psychiatry*. 2007;19:105–112.
180. De Hert M, Detraux J, van Winkel R, et al. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. *Nature Rev Endocrinol*. 2011;8:114–126.
181. Henderson DC, Cagliero E, Copeland PM, et al. Elevated hemoglobin A1c as a possible indicator of diabetes mellitus and diabetic ketoacidosis in schizophrenia patients receiving atypical antipsychotics. *J Clin Psychiatry*. 2007;68:533–541.
182. Arieff AI, Carroll HJ. Hyperosmolar nonketotic coma with hyperglycemia: Abnormalities of lipid and carbohydrate metabolism. *Metab Clin Exp*. 1971;20:529–538.
183. Fulop M, Tannenbaum H, Dreyer N. Ketotic hyperosmolar coma. *Lancet*. 1973;302:635–639.
184. Scott A. Joint British Diabetes Societies (JBDS) for Inpatient Care, JBDS hyperosmolar hyperglycaemic guidelines group. Management of hyperosmolar hyperglycaemic state in adults with diabetes. *Diabetic Med*. 2015;32:714–724.
185. Cruz-Caudillo JC, Sabatini S. Diabetic hyperosmolar syndrome. *Nephron*. 1995;69:201–210.
186. Halmos PB, Nelson JK, Lowry RC. Hyperosmolar non-ketoacidotic coma in diabetes. *Lancet*. 1966;287:675–679.
187. Tripodi A, Branchi A, Chantarangkul V, et al. Hypercoagulability in patients with type 2 diabetes mellitus detected by a thrombin generation assay. *J Thromb Thrombolysis*. 2011;31:165–172.
188. Heit JA, Leibson CL, Ashrani AA, et al. Is diabetes mellitus an independent risk factor for venous thromboembolism? *Arterioscler Thromb Vasc Biol*. 2009;29:1399–1405.
189. Keenan CR, Murin S, White RH. High risk for venous thromboembolism in diabetics with hyperosmolar state: comparison with other acute medical illnesses. *J Thromb Haemost*. 2007;5:1185–1190.
190. Scharf Y, Nahir M, Tatarsky I, et al. Fatal venous thrombosis in hyperosmolar coma. *Diabetes*. 1971;20:308–309.
191. Small M, Alzaid A, MacCuish AC. Diabetic hyperosmolar non-ketotic decompensation. *QJM: Int J Med*. 1988;66:251–257.
192. Milionis HJ, Elisaf MS. Therapeutic management of hyperglycaemic hyperosmolar syndrome. *Exp Opin Pharmacother*. 2005;6:1841–1849.
193. Kung AW, Pun KK, Lam KS, et al. Rhabdomyolysis associated with cranial diabetes insipidus. *Postgrad Med J*. 1991;67:912–913.
194. Rosa EC, Lopes AC, Liberatori Filho AW, et al. Rhabdomyolysis due to hyperosmolarity leading to acute renal failure. *Renal Failure*. 1997;19:295–301.
195. Rumpf KW, Kaiser H, Gröne HJ, et al. [Myoglobinuric renal failure in hyperosmolar diabetic coma (author's transl)]. *Dtsch Med Wochenschr*. 1981;106:708–711.
196. Singhal PC, Abramovici M, Venkatesan J. Rhabdomyolysis in the hyperosmolal state. *Am J Med Sci*. 1990;88:9–12.
197. Carchman RM, Dechert-Zeger M, Calikoglu AS, et al. A new challenge in pediatric obesity: pediatric hyperglycemic hyperosmolar syndrome. *Pediatr Crit Care Med*. 2005;6:20–24.

38

Hypoglycemia and Hypoglycemic Syndromes

Stephen N. Davis and Elizabeth M. Lamos

OUTLINE

Hypoglycemia, 655
  Accurate Measurement of Blood or Plasma Glucose, 655
Physiology of Hypoglycemia, 655
  Brain Fuel Metabolism, 655
Counterregulatory Responses During Hypoglycemia, 656
  Hypoglycemia-Associated Neuroendocrine and Autonomic Failure, 658
General Approach to the Patient, 659
  Acute Treatment of Hypoglycemia, 659
  Evaluation of the Hypoglycemic Patient, 660
Hypoglycemia in Diabetes, 660
  Mortality and Cardiovascular Disease, 661
  Strategies to Reduce Hypoglycemia, 661
Hypoglycemia in the Absence of Diabetes, 663
Hypoglycemia in Medicated or Ill Individuals, 663
Hypoglycemia in Apparently Well Individuals, 665
  Endogenous Hyperinsulinism, 665
Neonatal Hypoglycemia, 667
Conclusion and Future Directions, 669

HYPOGLYCEMIA

Hypoglycemia, a condition of low plasma glucose, has many varied causes, which are discussed in the following sections. Hypoglycemia is uncommon in individuals without a diagnosis of diabetes. This chapter will review the pathophysiology of hypoglycemia, the defense mechanisms to safeguard against a low blood glucose, and the etiologies of hypoglycemia in the individual with and without diabetes (Table 38.1).

The definition of hypoglycemia is often debated, but in this chapter, the glucose values adopted by the Steering Committee[1] of the major stakeholders in international diabetes care will be used. Level 1 hypoglycemia is defined as a plasma glucose below 70 mg/dL (3.9 mmol/L) but greater than or equal to 54 mg/dL (3.0 mmol/L). Level 2 hypoglycemia is a plasma glucose less than 54 mg/dL (3.0 mmol/L), and level 3 hypoglycemia is a severe hypoglycemic event associated with significant neurologic deficits requiring assistance from a third party.

Accurate Measurement of Blood or Plasma Glucose

Before a thorough workup for possible hypoglycemia, certain nuances should be considered. Whole blood, plasma, or serum can be used to measure glucose. Most laboratories report plasma glucose values. Whole blood glucose values are 10% to 15% lower than plasma glucose levels because of plasma's higher water content. Thus, 60 mg/dL (3.3 mmol/L) would be the lower limit of normal for a whole blood glucose sample. Hence, it is essential to establish whether the glucose sample is blood or plasma. Additionally, this principle will pertain to glucose meters that can give blood or plasma glucose values. Another important consideration is that mixed venous blood samples can be dramatically lower than arterial glucose values. This will vary depending upon the insulin sensitivity of the individual and the prevailing insulinemia. In addition, fasting glucose samples of capillary and venous samples are similar, but glucose concentrations during an oral glucose tolerance test (OGTT) in capillary blood are significantly higher than those in venous blood. Therefore, hypoglycemic values obtained from mixed venous blood should be interpreted with care during an OGTT.[2]

Artifactual hypoglycemia can also occur if a blood glucose sample is not collected in a tube containing citrate buffer, for example, which will inhibit glycolysis. Without appropriate sample collection, glucose values can decrease by 10 to 20 mg/dL (0.5–1.0 mmol/L) per hour at room temperature. In addition, even if glucose samples are collected in tubes containing glycolytic inhibitors, artifactually low readings can be obtained if the sample remains unmeasured for many hours, if the sample contains large quantities of white blood cells, if the sample is heavily lipemic with triglycerides, or if the patient has polycythemia vera, Waldentrom macroglobulinemia, or a condition in which there is decreased capillary flow.[3]

KEY POINTS

- Level 1 hypoglycemia is defined as a plasma glucose below 70 mg/dL (3.9 mmol/L) but 54 mg/dL (3.0 mmol/L) or above. Level 2 hypoglycemia is a plasma glucose below 54 mg/dL (3.0 mmol/L), and level 3 hypoglycemia is a severe hypoglycemic event associated with significant neurologic deficits requiring assistance from a second party.

PHYSIOLOGY OF HYPOGLYCEMIA

Brain Fuel Metabolism

During typical physiologic conditions, the brain requires a constant and adequate supply of glucose. Under normal postabsorptive conditions (e.g., an overnight fast), the brain accounts for approximately 65% of whole-body glucose uptake. Following feeding, the amount of glucose taken up by the brain can increase, but insulin does not influence the kinetics of glucose in the brain as it does in other organs, such as liver or muscle. Although the brain was classically considered an insulin-insensitive organ, recent work has challenged this concept.[4] Several studies have elegantly demonstrated that insulin can regulate feeding mechanisms and energy expenditure in rodent models.[4] Additionally, insulin administration into areas of the hypothalamus has been demonstrated to regulate hepatic glucose output by decreasing glucose uptake and glycogen synthesis. Furthermore, studies in dogs[5] and mice[6] have demonstrated direct central nervous system effects of insulin to amplify autonomic nervous system (ANS) counterregulatory responses to hypoglycemia. Thus, accumulating data indicate that insulin can modulate several components of energy homeostasis via direct neuronal signaling in certain areas of the brain.

Although typically dependent upon glucose as a fuel, the brain can adapt and use other substrates. Thus, during periods of fasting, ketone bodies, lactate, and alanine can be used as alternative brain fuels.[7] Several studies have demonstrated experimentally that high levels of alternative substrates (β-hydroxybutyrate and lactate) can be infused during acute hypoglycemia with a concomitant reduction in neuroendocrine and ANS responses. This indicates that the brain has the capacity to switch from glucose to alternative substrates in a matter of hours. However, it should be noted that the concentrations of substrates infused experimentally are far higher than levels observed during most physiologic conditions (the exception is levels of ketone bodies that occur during prolonged fasting).

TABLE 38.1 **Causes of Hypoglycemia in Adults**

Type 1 and type 2 diabetes
Hypoglycemia in medicated or ill individuals
Drug-induced Hypoglycemia (see Table 38.3)
Critical illness (hepatic, renal and heart disease, malnutrition)
Hormonal deficiency
Non–islet cell tumors
Hypoglycemia in apparently well individuals
Hyperinsulinism (insulinoma, nesidioblastosis, post–gastric bypass)
Autoimmune hypoglycemia (insulin autoimmune and antiinsulin receptor antibody syndromes)
Other (factitious, reactive hypoglycemia, pregnancy)

During hypoglycemia, brain glucose uptake falls. The exact glycemic value for the start of decreased blood-to-brain glucose transport is debated but is thought to be around 65 to 70 mg/dL (3.6–3.9 mmol/L) in humans. As hypoglycemia deepens (approximately 54 mg/dL [3 mmol/L]), blood-to-brain glucose transport becomes rate-limiting for brain glucose metabolism. Glycolytic-derived lactate and a small amount of stored astrocytic glycogen can provide a short duration of fuel supply. The presence of detectable brain glycogen levels, even after hypoglycemia, support that brain glycogen is an important reservoir of glucose equivalents.[8] Brain glycogen depletion likely occurs after about 120 minutes of hypoglycemia, thus emphasizing the need for a continuous and adequate supply of glucose to the brain from the circulation. Recent studies have demonstrated that that the sympathoadrenal response to hypoglycemia is not in response to a decrease in whole-brain glucose metabolism.[9]

COUNTERREGULATORY RESPONSES DURING HYPOGLYCEMIA

As plasma glucose falls, a well-orchestrated response of multiple physiologic mechanisms is activated (Fig. 38.1). The initial defense is a reduction in endogenous insulin secretion. This occurs as plasma glucose levels fall to less than 80 mg/dL (4.4 mmol/L). The reduction in endogenous insulin in response to falling plasma glucose is often overlooked in the hierarchy of defenses against hypoglycemia. Extended follow-up data from the Diabetes Control and Complications Trial (DCCT) demonstrated that a stimulated C-peptide (i.e., endogenous insulin) was associated with a decreased risk of hypoglycemia.[10] Similarly, experience from islet cell transplantations clearly demonstrates that the ability to modulate endogenous insulin levels is also protective against episodes of hypoglycemia.[11] As glucose levels continue to fall at or around

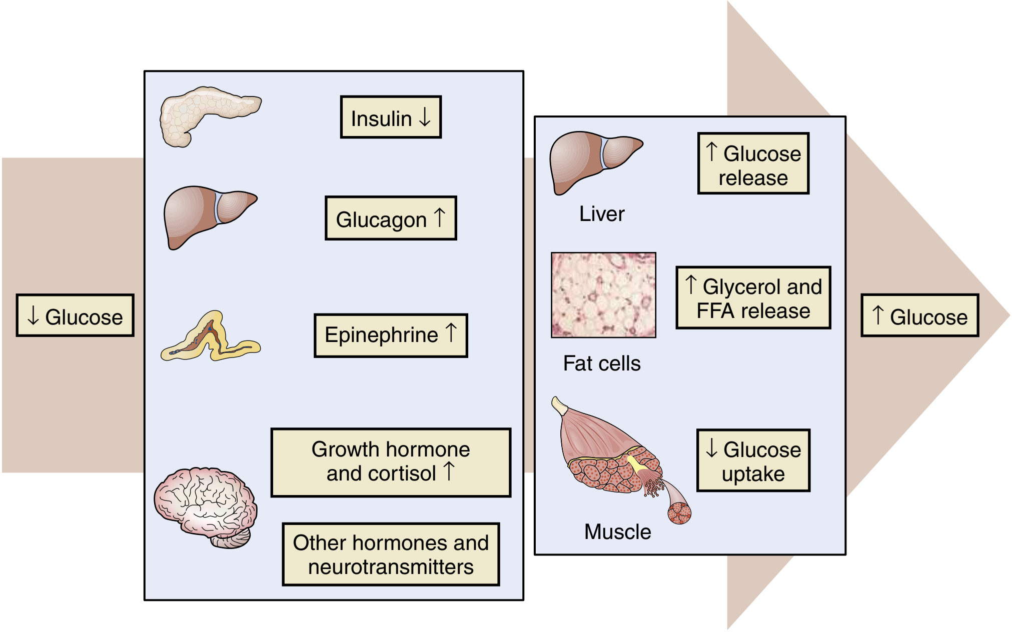


Fig. 38.1 Normal or healthy counterregulatory responses to hypoglycemia. As plasma glucose declines, powerful hormone regulation at the level of the pancreas, liver, and adrenal glands is activated. The liver, adipose tissue, and peripheral tissue respond to ultimately defend against a falling blood glucose. *FFA,* Free fatty acids. (From McDonnell ME, Umpierrez GE: Insulin therapy for the management of hyperglycemia in hospitalized patients. *Endocrinol Metab Clin North Am.* 2012; 41:175–201.)

70 mg/dL (3.9 mmol/L), a coordinated release of antiinsulin (counterregulatory hormones) occurs. Epinephrine (from the adrenal medulla), glucagon (pancreatic α cells), norepinephrine (sympathetic nerve endings and adrenal medulla), growth hormone (anterior pituitary gland), and cortisol (adrenal cortex) all have been demonstrated to have protective metabolic effects during acute or prolonged hypoglycemia. It should be noted that the release of neuroendocrine counterregulatory hormones and the inhibition of endogenous insulin secretion occur before a healthy adult can feel any symptoms of hypoglycemia. If plasma glucose continues to fall, at approximately 60 mg/dL (3.3 mmol/L), a series of autonomic (sometimes called neurogenic) signs and symptoms is activated. Autonomic warning responses to hypoglycemia include adrenergic and cholinergic symptoms. Adrenergic symptoms include palpitations, tremor, dry mouth, warmth, and anxiety. Cholinergic symptoms include sweating, hunger, and paresthesias. Signs of adrenergic activation include sweating and pallor. If the glucose level continues to fall, neuroglycopenic symptoms are activated at approximately 54 mg/dL (3.0 mmol/L). These include blurred vision, drowsiness, slurred speech, confusion, and difficulty concentrating. Defects in cognitive function are also apparent at this level of glycemia. If plasma glucose continues to fall, individuals can become drowsy, enter into a coma, and suffer seizures. Alternatively, individuals can become aggressive, which can be difficult to control and can be distinct from their usual personality. If severe hypoglycemia is prolonged, life-threatening events such as arrhythmias, myocardial infarction, and stroke, can occur.[12] Long-term cognitive damage and even death can occur if very severe hypoglycemia continues.[13] Although these represent typical responses to falling glucose levels, it should be appreciated that many patients have idiosyncratic neurologic and symptomatic responses to hypoglycemia that fall outside the classical description. Thus, it is worthwhile to measure the glucose level in anyone who presents with neurologic deficits and/or strange and uncharacteristic behavior.

The typical physiologic responses to hypoglycemia can be modified by a number of factors, including antecedent hypoglycemia, long-duration diabetes, age, gender, pregnancy, autonomic neuropathy, and use of certain drugs. These altered physiologic and pathophysiologic responses are discussed in detail in the following sections.

Insulin. Insulin is the principal physiologic factor that lowers plasma glucose. Insulin is secreted primarily in response to glucose, but amino acids, nonesterified fatty acids (NEFAs), β_2-adrenergic stimulation, and acetylcholine can also activate secretion of the hormone. Insulin secretion can be inhibited by hypoglycemia, insulin itself, somatostatin, and α_2-adrenergic activity.[14]

Insulin is released into the portal vein following secretion from pancreatic β cells. As the result of hepatic extraction, portal vein levels of insulin are approximately 2-fold higher than peripheral levels. Insulin lowers plasma glucose through several different mechanisms. Endogenous glucose production is inhibited directly first by suppressing hepatic glycogenolysis and then by restraining hepatic and renal gluconeogenesis. Simultaneously, insulin increases glucose uptake into insulin-sensitive tissues (primarily muscle but also liver and adipose tissue). As soon as the rate of exit of glucose exceeds the rate of entry of the substrate into the circulation, then the plasma glucose level starts to fall. Additionally, insulin can suppress glucose production indirectly by restraining lipolysis. (This restricts the flow of glycerol, an important gluconeogenic precursor, and NEFAs, which provide energy for gluconeogenesis.) Insulin also restricts proteolysis, which suppresses the flow of amino acids as gluconeogenic precursors.

Insulin secretion wanes as blood glucose returns to postabsorptive levels, removing inhibition of hepatic glucose production that will sustain fasting glucose levels until the next meal. After a typical overnight fast, the liver produces at least 85% of the glucose that enters the circulation. At this point, about 50% to 66% of glucose production is driven by glycogenolysis, and the rest is gluconeogenesis. If the fast were to continue, then most glucose production would be driven by gluconeogenesis, and glycogen stores would be depleted. After 48 hours of fasting, gluconeogenesis is responsible for 80% of glucose output. By 72 hours, gluconeogenesis becomes the only source of glucose production in the body.

Glucagon. Glucagon is released from the α cells in the islet of Langerhans. Similar to insulin, a number of physiologic factors can regulate secretion. These include hypoglycemia, amino acids, catecholamines (epinephrine and norepinephrine via β_2-adrenergic mechanisms), and free fatty acids.[15] Inhibitors of glucagon release include insulin and somatostatin. The regulation of glucagon release during hypoglycemia in humans is still undecided.[16] Normal glucagon secretion is observed in the setting of islet cell denervation and in *in vitro* pancreas preparations, suggesting that loss of these extrapancreatic influences does not explain the loss of glucagon response observed in severe β cell failure (type 1 diabetes mellitus [T1DM] or long-duration diabetes). Direct α cell stimulation of glucagon is signaled by a decrease in insulin secretion from the β cell in the setting of a low plasma glucose. Additionally, convincing data demonstrate that autonomic input (both sympathetic and parasympathetic) into the pancreas can also result in glucagon secretion.

Glucagon's physiologic actions are restricted almost exclusively to the liver, stimulating a rapid increase in hepatic glucose production over a period of 10 to 15 minutes. The initial rise in glucose output is provided by an increase in hepatic glycogenolysis. If hypoglycemia continues, glucagon can stimulate hepatic gluconeogenesis, but only if three carbon precursors such as glycerol, lactate, and amino acids are present.

Epinephrine and Norepinephrine. Epinephrine (adrenaline) is released from the adrenal medulla. Similar to glucagon, the hormone can act rapidly to increase hepatic glucose output by stimulating hepatic glycogenolysis. If hypoglycemia continues, and three-carbon precursors are present, epinephrine will stimulate gluconeogenesis in the liver and kidneys.[17] Unlike glucagon, epinephrine has important effects on peripheral tissues. Epinephrine can restrict insulin-stimulated glucose uptake in skeletal muscle, which, when quantified in terms of maintaining glucose in the circulation, is greater than the contribution made by any increase in endogenous glucose production. This latter property of epinephrine is especially important in the defense against hypoglycemia that typically is encountered in clinical practice. Unlike the model of rapid hypoglycemia produced during an insulin tolerance test, the usual clinical course involves a slower decline into hypoglycemia and a more protracted duration of low glycemia, which, during the night, can last up to several hours. With the acute induction of hypoglycemia caused by a large bolus of rapid-acting insulin, it is the activation of endogenous glucose production that is the primary physiologic defense against hypoglycemia, but in a model of more prolonged hypoglycemia, the restriction to insulin-mediated glucose uptake is paramount.[18] Epinephrine's important metabolic effects, which are mediated via β_2-adrenoreceptors, also include stimulation of lipolysis to provide substrate (glycerol) and energy (NEFA) for gluconeogenesis. Additional effects on muscle provide lactate, pyruvate, and amino acids for gluconeogenic precursors.

Norepinephrine has similar metabolic actions to those of epinephrine. However, because 90% of norepinephrine is taken up at the level of sympathetic clefts, and a further 90% of the remaining catecholamine is taken up by the gut and then the liver, the increase in circulating norepinephrine during hypoglycemia is relatively modest (~50% as compared with the 30-fold elevations that can occur with epinephrine).

Cortisol and Growth Hormone. The acute metabolic effects of cortisol and growth hormone are similar.[19,20] Both can increase glucose production through increases in gluconeogenesis. Both hormones also can inhibit insulin-stimulated peripheral glucose uptake and can increase lipolysis and proteolysis. However, prolonged hypoglycemia (3–5 hours) is needed before the metabolic effects of growth hormone or cortisol are measurable, and even at that time represents only approximately 20% to 25% of the action of epinephrine. Thus, although chronic deficiencies of cortisol and/or growth hormone can cause hypoglycemia, these hormones serve no demonstrable function in the defense against acute hypoglycemia.

Hypoglycemia-Associated Neuroendocrine and Autonomic Failure

As discussed earlier, the four primary physiologic defenses against falling plasma glucose include inhibition of endogenous insulin secretion, release of glucagon and epinephrine, and symptomatic cues to ingest carbohydrate (Fig. 38.1). Unfortunately, in patients with diabetes, all of these physiologic defenses can become defective and/or deficient. In T1DM and in long-duration type 2 diabetes mellitus (T2DM), the individual becomes critically insulin deficient, and thus the first physiologic line of defense (modulation of endogenous insulin) becomes lost (Fig. 38.2). After varying periods of duration of disease (~5 years), the ability of individuals with T1DM to release glucagon in response to hypoglycemia is lost (Fig. 38.2).[21,22] This defect also occurs to a similar extent in long-duration T2DM but is more gradual.[23] Thus, individuals are left with an attenuated epinephrine response. Sympathoadrenal activation is shifted to a lower glycemic threshold by recent antecedent hypoglycemia, prior exercise, and sleep. The lone attenuated epinephrine response causes a defective counterregulatory response, increasing the risk of severe hypoglycemia by 25 times, and hypoglycemic unawareness, the diminished or loss of the warning symptoms of hypoglycemia, the absence of which delays ingestion of carbohydrates. Hypoglycemic unawareness also confers an increased risk of hypoglycemia. Thus, the term hypoglycemia-associated autonomic failure (HAAF) describes the condition of insulin and glucagon deficiency in which antecedent hypoglycemia (or exercise or sleep) causes an attenuated epinephrine response and hypoglycemic unawareness that results in a vicious cycle of hypoglycemia[24] (Fig. 38.3).

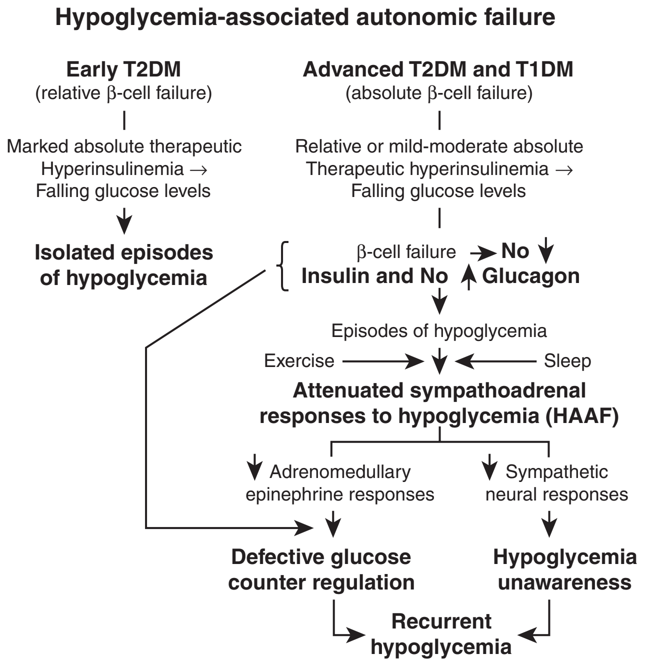

Hypoglycemia-associated autonomic failure
Early T2DM
(relative β-cell failure)
Advanced T2DM and T1DM
(absolute β-cell failure)
Marked absolute therapeutic Hyperinsulinemia → Falling glucose levels
Relative or mild-moderate absolute Therapeutic hyperinsulinemia → Falling glucose levels
Isolated episodes of hypoglycemia
β-cell failure → No ↓ Insulin and No ↑ Glucagon
Episodes of hypoglycemia
Exercise
Sleep
Attenuated sympathoadrenal responses to hypoglycemia (HAAF)
↓ Adrenomedullary epinephrine responses
↓ Sympathetic neural responses
Defective glucose counter regulation
Hypoglycemia unawareness
Recurrent hypoglycemia

Fig. 38.3 Schematic diagram of defective glucose counterregulation and hypoglycemia unawareness and the unifying concept of hypoglycemia-associated autonomic failure in type 1 diabetes mellitus. *HAAF,* hypoglycemia-associated autonomic failure; *T1DM,* type 1 diabetes mellitus; *T2DM,* type 1 diabetes mellitus. (From Cryer PE. Exercise-related hypoglycemia-associated autonomic failure. Diabetes 2009;58(9):1952; with permission of the American Diabetes Association.)

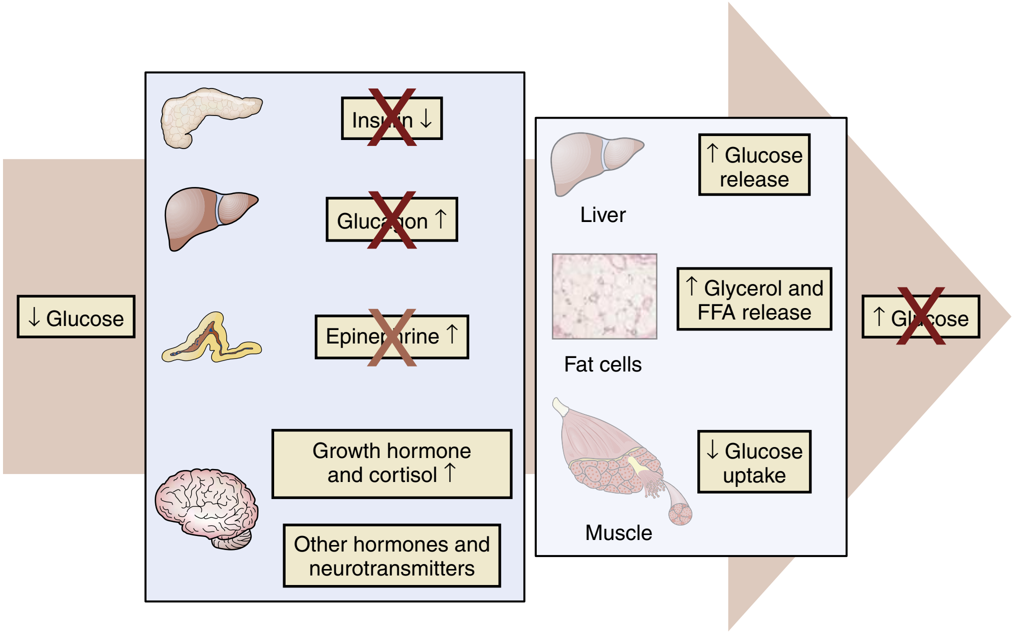

↓ Glucose
Insulin ↓
Glucagon ↑
Epinephrine ↑
Growth hormone and cortisol ↑
Other hormones and neurotransmitters
Liver
↑ Glucose release
Fat cells
↑ Glycerol and FFA release
Muscle
↓ Glucose uptake
↑ Glucose

Fig. 38.2 Counterregulatory response to hypoglycemia in type 1 diabetes mellitus and long-duration type 2 diabetes mellitus. As plasma glucose declines, impaired insulin and glucagon counterregulatory responses leave on the epinephrine response to defend against a falling glucose. In some cases, the epinephrine response is also impaired. This can result in an increased frequency of moderate and severe hypoglycemia. *FFA,* Free fatty acids.

Pathophysiology. Although the concept of HAAF is supported by an improvement in the epinephrine response to hypoglycemia with 2 to 3 weeks of strict avoidance of hypoglycemia, the mechanisms responsible for the syndrome of HAAF are not fully elucidated.[22,25] β cell failure in T1DM and long-duration T2DM explains the inability of insulin levels to decline with falling glucose levels. The glucagon defect appears to reside at the level of the islet. In diabetes with defective β cell function, α cells have a persistent inhibitory signal resulting in a diminished, if not absent, glucagon response. This is supported by evidence that glucagon response is maintained in scenarios in which denervation to the islet has been investigated (i.e., pancreas transplant, spinal cord transections) and in the setting of amino acid administration, a known glucagon-releasing stimulant. Lastly, the loss of insulin and glucagon secretion are correlated.

HAAF appears to only result when the sympathoadrenal response to hypoglycemia are affected. This is felt to reside in the central nervous system or the pathways by which signals are carried to and from the brain. The current hypotheses include the systemic-mediator hypothesis, the brain-fuel transport hypothesis, the brain metabolism hypothesis, and the cerebral-network hypothesis, which are described in detail by Cryer.[22]

Risk Factors. The risk factors that increase HAAF include: the extent (and duration) of β cell insulin deficiency, history of severe hypoglycemia, history of hypoglycemic unawareness, recent hypoglycemia before exercise or sleep, and intensive glucose control.[26] Additionally, it also appears that, in healthy persons, estrogen is a major mechanism responsible for sexual dimorphic counterregulatory responses during hypoglycemia.[27] Women appear to be more resistant than men to the blunting effects of prior hypoglycemia. Thus, antecedent hypoglycemia can have up to a 2-fold greater suppressive effect on subsequent counterregulatory responses in men than in women.[28]

Strategies to Improve Counterregulatory Responses During Hypoglycemia. Strategies for improving ANS and neuroendocrine responses during hypoglycemia are ongoing and are often carried out in parallel with investigation into the mechanism of HAAF. As mentioned previously, 2 to 3 weeks of strict avoidance of hypoglycemia with treatment changes or deescalation of glycemic goals results in improvement in the epinephrine response to hypoglycemia. Further strategies to reduce and monitor for hypoglycemia (i.e., continuous glucose monitoring) are discussed in "Strategies to Reduce Hypoglycemia."

Preclinical research approaches to increase neuroendocrine responses during hypoglycemia are many, and a few are presented here. The ventral medial nucleus (VMH) is a key region of the brain for detecting hypoglycemia. Activation of adenosine monophosphate–activated protein kinase (AMPK) appears to play a role in VMH sensing of hypoglycemia, and increases in AMPK action (a possible therapeutic target) can increase glucagon and epinephrine responses during hypoglycemia.[29] Activation of the VMH adrenergic receptors acutely appears to enhance the counterregulatory response to hypoglycemia; however, frequent activation may contribute to counterregulatory failure. Peroxisome-proliferator–activated receptor-γ agonists are known to activate AMPK and fructose, which under certain conditions can inhibit glucokinase, and increase counterregulatory responses in both healthy men and those with T1DM.[30] Treatment with low-dose carvedilol was found to improve VMH lactate levels, epinephrine, and feeding response in nondiabetic rodents with recurrent hypoglycemia.[31] Terbutaline, an epinephrine-stimulating β_2-adrenergic agonist, or alanine, a glucagon-releasing amino acid, before bed increases plasma glucose levels during the night and prevents nocturnal hypoglycemia.[32] Opioid receptor blockade, via treatment with naloxone, during hypoglycemia and antecedent exercise was shown to prevent blunting of responses (epinephrine and endogenous glucose production) to next-day hypoglycemia in individuals with T1DM.[33,34] Prolonged (i.e., weeks) use of two different selective serotonin reuptake inhibitors (sertraline and fluoxetine) has led to dramatic (30% to 60%) increases in ANS (epinephrine) responses during hypoglycemia, highlighting the possible role of serotonin in modulating counterregulatory responses during hypoglycemia.[35]

KEY POINTS

- Regulation of endogenous insulin, glucagon, and epinephrine is the primary acute hormonal response against an acutely falling blood glucose. Norepinephrine, cortisol, and growth hormone play little or no role in the acute defense against hypoglycemia.
- The term "hypoglycemia-associated autonomic failure" describes the condition in which antecedent hypoglycemia (or exercise or sleep) causes an attenuated epinephrine response and hypoglycemic unawareness that results in a vicious cycle of hypoglycemia.

GENERAL APPROACH TO THE PATIENT

Acute Treatment of Hypoglycemia

Treatment is aimed at restoring euglycemia (regardless of the cause), preventing recurrences, and, if possible, alleviating the underlying etiology.[36] Most cases of hypoglycemia in patients with diabetes are self-diagnosed and treated without medical emergency services. Initial treatment generally involves consumption of 15 to 20 g of simple carbohydrate. In severe hypoglycemia, with a measured glucose of less than 54 mg/dL (3.0 mmol/L), 20 to 30 g of carbohydrate may be necessary. Blood glucose should be remeasured within 15 minutes of oral carbohydrate, and the process should be repeated if the individual remains symptomatic, or in the setting of persistent hypoglycemia. These measures may be less effective if acarbose is prescribed because of the inhibition of carbohydrate absorption from the gut. In these cases, pure glucose gel or tablets are recommended. Additionally, physiologic or medication-induced delayed gastric emptying may result in unreliable absorption of glucose. Fat-containing foods will impair the absorption of glucose and should be avoided.

With more severe hypoglycemia preventing oral administration of carbohydrate, glucagon given subcutaneously, intramuscularly, or intranasally should be sufficient to raise the blood glucose concentration and revive the patient so that oral carbohydrate can be given. Intranasal glucagon has been shown to be as effective as intramuscular and subcutaneous glucagon in the treatment of hypoglycemia.[37] This newer glucagon preparation, which is needle free, is significantly easier to use than the current glucagon emergency kits, particularly in the pediatric population or in those with limited dexterity.

Comatose patients should receive intravenous glucose, followed by an infusion to maintain glucose at around 80 to 100 mg/dL (4.4–5.6 mmol/L). Protocols include a 25-g bolus of 50% dextrose in water (D50), but some use 10% dextrose or a combination of dextrose concentrations with a continuous infusion of 25 g of D50 or 1 g/kg of body weight in D50 in a large-gauge port and vein. A 10% or 20% dextrose infusion can be provided in a peripheral vein. A sulfonylurea or meglitinide overdose or exogenous long-acting insulin can result in prolonged hypoglycemia as long as several days requiring sustained intravenous glucose infusion aimed at keeping the blood glucose level at approximately 80 to 100 mg/dL (4.4–5.6 mmol/L) to avoid hyperglycemia, causing further stimulation of insulin secretion and setting in motion a vicious cycle. When other drugs are involved, their use should be discontinued if possible. Occasionally, diazoxide or a somatostatin analog may be needed to inhibit insulin secretion.

Blood glucose levels should be monitored initially every 15 to 30 minutes and subsequently at 1- to 2-hour intervals. Once euglycemia is achieved, a complex meal should be provided to restore any lost glycogen stores. In other conditions, the underlying disorder (e.g., sepsis, heart failure, endocrine deficiency) should be treated, and the blood glucose level should be supported.

Evaluation of the Hypoglycemic Patient

In an individual with diabetes, a careful review of the duration of diabetes, treatment plan, adherence, and timing of hypoglycemia to mealtimes, medication administration, and exercise are critical (Table 38.2). Strategies to reduce hypoglycemia in diabetes are discussed later. Without a diagnosis of diabetes, the first step in the evaluation of hypoglycemia is to establish Whipple's triad (low plasma glucose, typical hypoglycemic symptoms, and resolution of symptoms with euglycemia). It is important to determine what the patient's symptoms were. When did they occur in relation to the patient's last meal, and what was the patient doing when the episode occurred? Was it an isolated event, or had it occurred before? How frequently do these episodes occur? Is there any pattern to the occurrences? How long have these events been occurring? Did weight gain or weight loss occur during this period? Is the patient taking any medications? If so, he or she should be checked to make sure that no mistake has been made. Did the patient lose consciousness? If so, were premonitory signs present? Was hypoglycemia documented? Did the patient recover spontaneously? What did the patient do to prevent recurrences or relieve symptoms? What is the patient's occupation? Does the patient or any immediate relative have diabetes? If so, how is it being treated? Does the patient have any medical conditions? Do family members have any endocrine disorders? For example, a family history of autoimmune disorders should raise the suspicion of Addison disease.

In a retrospective study, diabetes, infection/sepsis, renal and liver disease, and malignancies accounted for the majority of emergency room cases of hypoglycemia.[38] Virtually all patients had decreased level of consciousness, giddiness, or other nonspecific symptoms. The incidence of death was approximately 16 per 1000 emergency department visits. Among diabetic patients, 60% had intensive control of blood sugars with oral hypoglycemic agents, insulin, or both.

Most outpatients who have a disorder that is causing hypoglycemia should describe discrete episodes. Classically, premonitory symptoms such as a feeling of anxiety, hunger, blurred vision, difficulty thinking, weakness and sweating, and palpitations will be noted. Loss of consciousness can occur. With repetitive episodes, patients learn that they can abort these symptoms by eating; therefore, patients often have a history of weight gain. Spells similar to those caused by hypoglycemia can represent a vasovagal reflex, cardiac arrhythmia, a convulsive disorder, orthostatic hypotension, or a transient ischemic attack. In mild cases of insulinoma, spells might occur only after vigorous exercise or prolonged fasting. On the other hand, skipping lunch and having several cocktails before dinner can precipitate an isolated hypoglycemic episode in an otherwise healthy person. The history (e.g., prior pituitary surgery, symptoms of galactorrhea, amenorrhea, impotence, or endocrine insufficiency, as with thyroid replacement therapy) can provide clues to the possibility of hypopituitarism. The patient's occupation or family diseases can be important with regard to access to hypoglycemia-producing drugs (e.g., insulin, sulfonylureas) or multiple endocrine neoplasia type 1 (MEN1).

The physical examination can be useful in excluding carotid disease, orthostatic hypotension, severe cardiac and hepatic disease, hypothyroidism, cachexia, and cancer, which should be obvious at the stage at which they would cause hypoglycemia. Routine laboratory testing can exclude renal insufficiency or can raise the possibility of adrenal insufficiency (hyponatremia). If all of the discussed are negative, and prior hypoglycemia has not been documented unequivocally, this must be done before one proceeds with other diagnostic tests.

KEY POINTS

- The general approach should be to identify hypoglycemic quickly. Treat hypoglycemia with oral simple carbohydrate (15–30 g). If severe or comatose, the patient should be treated with glucagon and may require intravenous glucose. Assessment of the underlying etiology may require additional/alternate treatment. Individuals should be monitored for initial improvement at short intervals (15–30 min). The last steps are to address the underlying condition.

HYPOGLYCEMIA IN DIABETES

Hypoglycemia is associated with significant morbidity and can be fatal.[39] Hypoglycemia and the fear of hypoglycemia are significant barriers to achieving adequate glucose control. Almost immediately after the discovery of insulin in the early 1920s, hypoglycemia was recognized as an unpleasant and dangerous side effect of therapy. More than 90% of all patients with T1DM have suffered an episode of hypoglycemia. Typically, a patient with T1DM will experience two to 10 episodes of subjective hypoglycemia per week.[40] Nocturnal hypoglycemia is frequent and prolonged with conventional therapy. Episodes of major hypoglycemia requiring assistance or hypoglycemic events resulting in seizures or coma increase with intensive management.[41] Glycosylated hemoglobin (HbA1c) is inversely associated with risk for

TABLE 38.2 Diagnostic Approach to an Adult With Documented Fasting Hypoglycemia

1. Consider the most likely disorders (drugs, critical illness, endocrine deficiency, non–β cell tumor, and insulinoma) while supporting the plasma glucose concentration if necessary.
2. Examine the history, physical examination, and available laboratory data for clinical clues to include or exclude the above categories.

Cause	Response
Insulin- or sulfonylurea-treated diabetes	Adjust the therapeutic regimen
Use of other drugs known or suspected to cause hypoglycemia	Discontinue use of the drug
Hepatic, renal, or cardiac failure; sepsis; or inanition	Treat the underlying disorder
Anorexia, weight loss, change in skin pigmentation, known pituitary or adrenocortical disease, hypotension, hyponatremia, hyperkalemia	Evaluate for adrenocortical/pituitary insufficiency
Known non–β cell tumor, mass on examination or imaging studies	Check for high IGF-2 or IGF-2:IGF-1 ratio

3. In the absence of clinical clues, consider medication error, endogenous hyperinsulinism (see Tables 38.1 and 38.3), and surreptitious or malicious insulin secretagogue or insulin administration.

IGF, Insulin-like growth factor.

episodes of severe hypoglycemia, but the extent of risk depends on the defined threshold.[41,42] The duration of T1DM also plays an important role in increasing the frequency of severe hypoglycemia. Recent data from a 30-year follow-up of the DCCT/Epidemiology of Diabetes Interventions and Complications trial[42] demonstrate that the frequency of severe hypoglycemia remains at a rate of 36 to 41 episodes per 100 patient-years, even among those with longer duration diabetes and similar HbA1c levels. Lower HbA1c and younger age at study entry were risks for severe hypoglycemia.

Because of the 10-fold greater incidence of T2DM, the total episodes of hypoglycemia are greater than what occurs in T1DM. However, the risk and frequency of hypoglycemia are considered to be much lower in T2DM as compared with T1DM. This is undoubtedly true in patients with T2DM treated with a noninsulin secretagogue or combination therapy of metformin, thiazolidinedione, dipeptidyl peptidase-4 inhibitor, sodium-glucose cotransporter-2 inhibitor, α-glucosidase inhibitor, and/or a glucagon-like peptide-1 (GLP1) agonist. However, the incidence of hypoglycemia increases sharply in longer duration (insulin-requiring) T2DM. Recent clinical trials investigating insulin replacement strategies in T2DM report hypoglycemia five times more often compared with metformin.[43] Hypoglycemia requiring any assistance was more common with intensive glucose control compared with standard care; however, as intensive control was relaxed, rates of hypoglycemia equalized.[44] Rates of severe hypoglycemia have improved for patients over 75 years of age with T2DM.[45] Unlike observations in T1DM, individuals with T2DM appear to be at an increased risk for severe hypoglycemia at both near-normal (<6%) and poorly controlled (≥9%) HbA1c values.[46]

Mortality and Cardiovascular Disease

Death from hypoglycemia can be considered rare, given the very large number of episodes of hypoglycemia that occur in clinical practice. Mortality owing to hypoglycemia is estimated to be less than 8% in individuals with T1DM under 56 years of age and 4% in all people with diabetes under 50 years of age.[12]

Initially, it was thought that the rate of cardiovascular disease in T2DM may be reduced with more intensive glycemic treatment compared with standard control; however, subsequent large trials failed to show this benefit.[12] Additionally, intensive control was associated with an increased risk for severe hypoglycemia compared with standard control and an increase in mortality. Risks for hypoglycemia included long-duration diabetes, advanced age or frailty, a history of hypoglycemia, and known established atherosclerotic or microvascular disease. However, low HbA1c levels do not appear to be correlated with increased cardiovascular events if treated with medications that do not cause hypoglycemia. In T1DM and T2DM studies, individuals with hypoglycemia have a 1.5- to 6-fold increased risk of cardiovascular outcomes and mortality compared with those with no hypoglycemia. This increased risk can extend up to 1 year after the hypoglycemic event. Severe hypoglycemia appears to be an indicator of increased risk of cardiovascular events, cardiovascular mortality, and all-cause mortality, regardless of treatment goals, particularly if the individuals had an elevated risk of cardiovascular disease at baseline.[39] Experts continue to debate whether these results are causative or confounding. Hypoglycemia has been associated with electrocardiogram changes, arrhythmias, and hypercoagulable and prothrombotic states.

Strategies to Reduce Hypoglycemia

Glycemic Targets. Glycemic targets and therapies in diabetic patients should be individualized.[47] Although a HbA1c goal for many nonpregnant adults of less than 7% is reasonable, a lower HbA1c goal of less than 6.5% may be appropriate, if this can be achieved without significant hypoglycemia. However, a relaxed glycemic goal of less than 8% are appropriate in those with a history of severe hypoglycemia, multiple comorbid diseases, including advanced diabetic complications, or in those with long-duration diabetes who are unable to achieve glycemic control despite maximized treatment, education, and resources.

In hospitalized patients, hypoglycemia is associated with short- and long-term mortality.[48] Use of electronic treatment order sets and hypoglycemia protocols increases the likelihood of time in target range without increasing hypoglycemia and reduces severe hypoglycemic events.[49] Additionally, remote management of inpatients with diabetes by specialists was found to reduce hypoglycemia.[50] For most hospitalized patients, a glucose target of 140 to 180 mg/dL (7.8–10.0 mmol/L) is recommended.[51] However, selected patients may benefit from more intensive glucoses targets (110–140 mg/dL [6.1-7.8 mmol/L]) if hypoglycemia can be avoided. These recommendations were based on the Normoglycemia in Intensive Care evaluation and Survival Using Glucose Algorithm Regulation study, which demonstrated that moderate inpatient glucose targets showed no treatment advantage and slightly improved mortality compared with intensive control.[52] Additionally, hypoglycemia was 10 to 15 times higher in the intensively treated group.

Treatment Regimens and Technology. By far the most common cause of hypoglycemia in diabetes is medication (Table 38.3). Hypoglycemia induced by oral insulin secretagogues is much less frequent than that caused by insulin but in certain instances can still be common. For example, hypoglycemia can occur in up to approximately 35% of patients receiving glyburide or repaglinide, which are glucose-independent. Hypoglycemia rates are lower in agents with glucose-dependent insulin secretion (i.e., stimulation of insulin release is reduced during periods of hypoglycemia). Thus, the percentage of patients who experience hypoglycemia when receiving glimepiride, glipizide XL, gliclazide, or nateglinide is lower, in the range of 5% to 10%. However, combination of even glucose-dependent insulin secretagogues with agents such as insulin or GLP1 agonists can result in a much higher frequency of hypoglycemia (>35%). Generally, newer sulfonylureas produce less hypoglycemia than do older first-generation sulfonylureas such as chlorpropamide. Severe hypoglycemia remains relatively uncommon with sulfonylureas, at 0.8%.[53]

Over the last generation, interest has increased in replacing insulin in the most physiologic manner for patients with diabetes. In the 1980s, the introduction of recombinant human insulin reduced the formation of antibodies and provided more predictable pharmacokinetic profiles. The next decade produced analog insulins that initially were designed to provide a quicker onset and a shorter duration of action. These insulins (lispro, aspart, glulisine) were designed to reproduce more closely the typical physiologic prandial spikes of insulin observed following meals. The second wave produced long-acting basal types of insulin (glargine, detemir, degludec) designed to mimic background constitutive insulin release. Studies in T1DM have demonstrated that hypoglycemia (particularly nocturnal) can be reduced when short-acting analogs rather than regular (soluble) insulin are used.[54] Similarly, long-acting analogs have been demonstrated to reduce daytime and severe hypoglycemia by 26% and 36%, respectively, in patients with T2DM when compared with neutral protamine Hagedorn–based regimens.[55] Thus, current recommendations are to use analog-based insulin replacement regimens whenever possible. Inappropriate dosing or timing of insulin administration to food intake is a common cause of hypoglycemia in an individual with diabetes.

Continuous Subcutaneous Insulin Infusion. Continuous subcutaneous insulin infusion (CSII), or insulin pump, development began in the 1970s and over the past 20 years has become a major method of insulin replacement in the United States. Although

TABLE 38.3 Drug-Induced Hypoglycemia

Drugs capable of causing hypoglycemia by themselves
Antidiabetic drugs
Insulin
Sulfonylureas
Benzoic acid derivatives (meglitinide)
Other
Alcohol
Salicylates
Propranolol
Pentamidine
Sulfonamides
Vacor rodenticide
Quinine/quinidine
Propoxyphene
Para-aminobenzoic acid
Perhexiline
Drugs that probably cause hypoglycemia only in combination with insulin/sulfonylurea/benzoic acid derivatives or under special circumstances (e.g., malnutrition, infection, renal insufficiency)
Biguanides
Angiotensin-converting enzyme
Phenylbutazone
Lidocaine
Warfarin (Coumadin)
Ranitidine, cimetidine
Doxepin
Danazol
Azopropazone
Oxytetracycline
Clofibrate, benzofibrate
Colchicine
Ketoconazole
Chloramphenicol
Haloperidol
Monoamine oxidase inhibitors
Thalidomide
Orphenadrine
Selegiline
Abenzolene
Flecainide
Fluoxetine
Clomipramine
Indomethacin
Chloroquine

widely acknowledged for its usefulness, the cost of treatment (pump and supplies) has restricted a more widespread acceptance of this technology. However, CSII in individuals with T1DM does not appear to consistently or significantly reduce the risk of severe hypoglycemia compared with traditional insulin injection therapy.[56,57]

Continuous Glucose Monitors. Continuous glucose monitoring (CGM) and flash real-time glucose monitoring is now an established technology in clinical practice. A variety of devices can be worn in combination with an insulin pump or independently. In studies of CGM in individuals with T1DM using reduction of hypoglycemia as an endpoint, time below target (defined as <70 and <54 mg/dL [3.9 and 3.0 mmol/L]) is reduced.[47] CGM is superior to self-monitoring of blood glucose in terms of time spent in target and reduction of severe hypoglycemia.[58] The results are less robust in CGM use in T2DM.[59] CGM compared with flash glucose monitoring was more successful in reducing level 2 hypoglycemia in individuals with T1DM and a history of hypoglycemic unawareness.[60] Continuous glucose monitors are not approved for inpatient use. However, studies have demonstrated a reduction in hypoglycemia in various inpatient populations, particularly those in intensive care units (ICUs).[61] CGMs may be particularly useful, given that ICU patients may be unable to report hypoglycemia symptoms, for example during altered mental status or intubation, and may have blunted counterregulatory responses to hypoglycemia.

Integration of Continuous Subcutaneous Insulin Infusion and Continuous Glucose Monitoring. Certain insulin pumps with sensor-augmented features suspend basal insulin when low blood glucose is detected. Furthermore, an additional predictive feature can suspend insulin delivery based on a sensor-predicted low blood glucose algorithm. Use of these features has been shown to reduce the time in hypoglycemia and daytime and nocturnal severe hypoglycemia compared with sensor augmented treatment.[62,63]

"Hybrid" closed-loop systems automatically modify basal insulin delivery using real-time sensor glucose trends but do not independently dispense bolus insulin. There is a dedicated hypoglycemia safety algorithm. Compared with sensor-augmented insulin pump treatment, the Medtronic MiniMed 670G demonstrated less time in hypoglycemia, and the Tandem Control-IQ showed no increase in time in hypoglycemia, when compared with sensor-augmented therapy.[64,65]

Pancreas and Islet Cell Transplantation. One of the major criteria for consideration of pancreas or islet cell transplantation (see Chapter 41) is complete insulin deficiency with a history of recurrent severe hypoglycemia. Pancreas and islet cell transplantation may improve hormonal response to insulin-induced hypoglycemia.[66] Initially, after pancreatic transplantation, both basal and glucose-stimulated insulin secretion are increased compared with normal controls. This occurs in the absence of meaningful innervation of the transplanted pancreas. In response to hypoglycemia, individuals with T1DM having received a whole pancreas transplant demonstrate a partial suppression of endogenous insulin that is sufficient to result in appropriate glucagon secretion. Whole pancreas transplant results in normalization of the glycemic thresholds for counterregulatory hormone response (epinephrine and growth hormone) to hypoglycemia; however, in whole pancreas transplantation, the epinephrine response is only partially restored.

In individuals with T1DM who receive islet allotransplantation, endogenous insulin secretion is suppressed with hypoglycemia, and glucagon secretion is partially restored. The transplanted islet cell mass is critical to the extent of recovery of the glucagon response. Additionally, normalization of the glycemic thresholds for counterregulatory hormone response to hypoglycemia are observed within 6 months; however, the extent of the epinephrine response may be partial and/or delayed for months to years. In individuals receiving islet autotransplantation, the extent of the glucagon response recovery may depend on the transplant location (hepatic or nonhepatic site). For reference, nonhepatic site islet cell autotransplantation results in a normal glucagon response to hypoglycemia.

Exercise. Individuals with T1DM or longstanding T2DM who exercise are at increased risk for hypoglycemia during and many hours after exercise as a result of imperfect insulin replacement, increased insulin action, reduced glucagon replacement, and impaired counterregulatory responses.[67] In general, if an individual is experiencing exercise-related hypoglycemia, it is appropriate to consider alterations to his or her insulin regimen (both basal and bolus) and carbohydrate intake. Short high-intensity sprints, resistance exercise before aerobic exercise, and activity timing are all strategies to reduce the risk of hypoglycemia. The risk of nocturnal or delayed hypoglycemia (out to 48 hours) can be reduced with temporary basal reductions, bedtime snacks, and use of continuous glucose monitoring.

It should be noted that hypoglycemia has been documented after prolonged strenuous exercise (i.e., marathon runners), sometimes in the setting of medications like β blockers. However, most instances of exercise-associated hypoglycemia are asymptomatic, self-limited, and readily reversed by carbohydrates.

General Considerations. All visits with individuals with diabetes should include an assessment of hypoglycemia and hypoglycemia symptoms.[47] Additionally, in individuals who are prescribed medications that are more likely to cause hypoglycemia, heightened awareness and consideration of hypoglycemic unawareness is important. Recurrent level 3 episodes of hypoglycemia or hypoglycemic unawareness should prompt renewed hypoglycemia education and reassessment of the treatment plan and goals. For individuals who are prescribed insulin, a pattern of hypoglycemia or one level 3 hypoglycemic episode should result in adjustment of the treatment plan and strict avoidance of hypoglycemia for at least 2 weeks. Ongoing evaluation of cognitive function is important, as level 3 hypoglycemia is associated with established or declining cognitive function. Vulnerable groups, including children and older adults, may be unable to convey their symptoms or needs effectively. For those with diabetes who use intermittent fasting or are of Islamic faith and observe the month-long Ramadan fast, patient education, continuous glucose monitoring, and diet modification, in conjunction with adjustment of insulin timing and dosing, should be considered to prevent hypoglycemia.[68] Specific situations, including driving and recreational exercise or sports, may require relaxed glucose control to avoid hypoglycemia. Frequency of testing should be increased, and a glucose source should be readily available.

KEY POINTS

- In type 1 and type 2 diabetes mellitus studies, individuals with hypoglycemia have a 1.5- to 6-fold increased risk of severe adverse cardiovascular outcomes and mortality compared with those with no hypoglycemia.
- Glycemic targets and therapies in diabetic patients should be individualized. Although a HbA1c goal of less than 7% is reasonable for many nonpregnant adults, a lower HbA1c goals of under 6.5% may be appropriate in the absence of hypoglycemia. A relaxed glycemic goal of below 8% is appropriate in some individuals.

HYPOGLYCEMIA IN THE ABSENCE OF DIABETES

In nondiabetic individuals, hypoglycemia is rare, with blood glucose levels of less than or equal to 70 mg/dL (≤3.9 mmol/L) and less than or equal to 54 mg/dL (≤3.0 mmol/L) reported in as few as 1082/100,000 and 910/100,000 emergency medical service encounters, respectively.[69] In hospitalized patients, nondiabetic hypoglycemia accounts for 20% to 30% of the hospitalizations with hypoglycemia. Mortality in hospitalized patients who experienced hypoglycemia was 15%. Older age, low body mass index, renal and heart failure, chronic liver disease, sepsis, and malignancies are risk factors for hypoglycemia.[70]

Establishment of Whipple's triad is integral to confirming a hypoglycemic disorder.[21] The threshold for symptomatic hypoglycemia shifts to lower plasma glucose levels in individuals with recurrent hypoglycemia. Fingerstick or CGM low glucose values should be confirmed with a low plasma glucose level. Well-documented severely low blood glucose (<40 mg/dL [2.2 mmol/L]) in the absence of symptoms should still be investigated. Typical hypoglycemic symptoms with normal plasma blood glucose (>70 mg/dL [3.9 mmol/L]) are not consistent with a hypoglycemic disorder. Underlying causes of hypoglycemia are diverse in people without diabetes, and diagnosis can be complex and requires careful consideration of patient history, and potentially further workup. Causes of hypoglycemia are generally divided into those in medicated or ill individuals and those in apparently well individuals.

HYPOGLYCEMIA IN MEDICATED OR ILL INDIVIDUALS

Drug-Induced Hypoglycemia. Some 23% of hospital admissions for adverse drug events are attributed to drug-induced hypoglycemia.[71] Insulin and insulin-secreting medications (i.e., sulfonylurea or meglitinides) are common causes of drug-induced hypoglycemia.[21] This is discussed in "Hypoglycemia in Diabetes." Alcohol is a common cause of hypoglycemia; however, hypoglycemia is an uncommon complication of alcohol abuse.[72] In a series of deaths caused by hypoglycemia, alcohol was the most common causative agent.[73] The most common situation is a glycogen-depleted state, including chronic liver disease or circumstances in which an individual drinks after a considerable fast or drinks and then fasts. In the latter situation, blood alcohol levels can be low, but cellular metabolic effects at the liver persist. Alcohol induces hypoglycemia by inhibiting gluconeogenesis, with evidence of impaired counterregulatory hormone responses. Thus, in a patient with suspected alcohol-induced hypoglycemia, oral or intravenous glucose is the treatment of choice. If there is a clinical concern for malnutrition, thiamine supplementation should be provided to avoid Wernicke's encephalopathy.

Many of the drugs listed in Table 38.3 have been reported to cause hypoglycemia, generally the subject of isolated case reports, and their etiologic significance remains to be established.[21,71] Often, dosing was correct, and risk factors for associated hypoglycemia were older age, renal, liver or systemic disease, or coadministration of insulin or sulfonylurea. However, their use in a patient with otherwise unexplained hypoglycemia should be discontinued whenever possible.

Critical Illness. In a retrospective study of critically ill patients, mortality was 56% among patients with severe hypoglycemia (<40 mg/dL [<2.2 mmol/L]) compared with 40% among controls.[74] Risk factors included insulin administration within 12 hours and no nutritional therapy. Patients were more likely to have diabetes, to be under a tight glucose control protocol, and to have septic shock. A single episode of hypoglycemia conferred an increased risk of mortality.

Initially, the response to the stress of infection is an increase in glucose turnover, with glucose production often exceeding glucose utilization and resulting in mild hyperglycemia.[75] This early response appears to be mediated by an increase in cortisol, epinephrine, and glucagon. As the infection worsens, the rate of glucose utilization exceeds that of glucose production, and hypoglycemia ensues. This hypoglycemia is secondary to inhibition of gluconeogenesis. Muscle and fat glucose uptake are blocked and may represent a beneficial effect of redirecting glucose to more central tissues. Appropriate management entails (1) treatment of the underlying infection, (2) restoration of normal peripheral perfusion, and (3) infusion of glucose to satisfy tissue demands.

Hepatic and Renal Disease. The liver and kidneys are the only organs that are capable of releasing glucose into the circulation, inasmuch as other tissues generally lack or have minimal amounts of the enzyme glucose-6-phosphatase. Consequently, it is not surprising that patients with hepatic or renal disease are prone to hypoglycemia.[76] Additionally, clearance of insulin is reduced in chronic kidney disease, leading to an increased risk of hypoglycemia. Nevertheless, it is uncommon for hypoglycemia to occur simply as a result of loss of mass or function of these organs, and when it does occur, the cause is usually multifactorial.[76] The large capacities of these organs to release glucose into the circulation and their ability to compensate for each other's shortcomings appear to provide an explanation for this phenomenon.

Normally, the liver accounts for 80% to 85% of all glucose production; it can increase its output (initially mainly by glycogenolysis, later by gluconeogenesis) over a sustained period by 2- to 3-fold. The kidney can also increase its output over a prolonged period by 2- to 3-fold. Studies indicate that the kidney can acutely increase its output to compensate for decreased hepatic glucose production and vice versa, a phenomenon that is referred to as *hepatorenal reciprocity*.[76] Thus, hypoglycemia would be unlikely to develop in patients with hepatic disease until the liver's capacity to produce glucose was reduced beyond the ability of the kidney to compensate.

Liver Disease. Although hypoglycemia has been associated with a wide range of liver diseases (hepatocellular carcinoma, cirrhosis, fatty metamorphosis, toxic and infectious hepatitis, cholangitis, and biliary obstruction), its occurrence is actually uncommon in the absence of other complicating factors (i.e., sepsis or heart failure).[76] The incidence and extent of hypoglycemia in liver disease is variable based on acuity of dysfunction and extent of parenchyma loss. In chronic liver disease associated with hypoglycemia, additional factors, such as malnutrition and infection, are usually involved. A 50% incidence of hypoglycemia was found in patients with liver disease, associated with sepsis and circulatory collapse.[77] In such situations, liver function tests may not parallel the severity of hypoglycemia.

Hyperinsulinemia often accompanies hepatic disease as a consequence of decreased insulin degradation by the liver, but the hypoglycemia of hepatic disease is almost always accompanied by appropriate suppression of plasma insulin concentrations. Only case reports have implicated overproduction of insulin-like growth factors (IGFs) by the liver in primary hepatocellular carcinoma, the neoplasm most frequently associated with fasting hypoglycemia.[78]

Renal Disease. Except in infants, hypoglycemia rarely occurs with acute renal failure. However, with chronic renal failure, hypoglycemia in adult patients without diabetes has an incidence of 1% to 3%.[79,80] It can occur as an isolated event or can be repetitive. In general, neuroglycopenia rather than autonomic symptoms predominates.

The origin of renal hypoglycemia is complex.[79] Many factors, including altered drug metabolism, delayed gastric emptying, malnutrition, infection, dialysis, altered insulin sensitivity and clearance, associated hepatic and cardiac disease, and impaired renal and hepatic glucose production predispose uremic patients to hypoglycemia. Drugs are probably the most common immediate cause. Any drug that can cause hypoglycemia is more likely to do so in a uremic patient because of a prolonged half-life (e.g., insulin, certain sulfonylureas, especially chlorpropamide) or decreased protein binding secondary to hypoalbuminuria. Although hypoglycemia occurs in nondiabetic as well as diabetic patients with renal failure, it is more likely to occur in the latter because of the use of hypoglycemic agents, and because patients with longstanding diabetes have autonomic neuropathy and defects in glucose counterregulation. Malnutrition secondary to anorexia or vomiting, which can reduce hepatic glycogen stores and the availability of gluconeogenic precursors, is a common feature that increases the risk for hypoglycemia.[76]

Several reports suggest that renal failure *per se* predisposes to the development of hypoglycemia.[80] Most evidence points to diminished glucose production rather than increased glucose utilization as a cause of the hypoglycemia. Accompanying malnutrition and acidosis would also contribute to diminished hepatic glycogenolytic and gluconeogenic potential. The expected compensatory increase in renal gluconeogenesis in response to acidosis would be compromised by loss of renal mass and exacerbated by inappropriately high plasma insulin levels caused by reduced renal insulin degradation.

At the present time, no evidence indicates that abnormal counterregulatory hormone responses play a primary role in the pathogenesis of the hypoglycemia associated with renal or liver dysfunction.

Cardiac Failure. Spontaneous hypoglycemia can occur with severe heart failure.[81] The condition has been attributed to a variety of mechanisms, including reduce gluconeogenesis, poor dietary intake, gastrointestinal malabsorption secondary to gut edema, and increased glucose utilization in poorly perfused tissues.

Malnutrition. The prevalence of malnutrition is high among hospitalized patients, and as much as 50% of all patients admitted are malnourished or at risk for malnutrition.[82] The risk of hypoglycemia increases with reduced dietary intake and/or weight loss preceding hospital admission. Patients are more likely to have hypoalbuminemia and comorbid renal failure and anemia. Additionally, hypoglycemia has been reported in eating disorders, particularly anorexia nervosa–restricting type, with a prevalence of 7%.[83]

Hormonal Deficiency. Although cortisol and growth hormone have been demonstrated to contribute independently to glucose counterregulation via their actions to promote glucose production and limit glucose uptake,[19,20] hypoglycemia does not develop in most adults who lack these hormones. Hypoglycemia in adults is more common in secondary adrenal insufficiency (i.e., panhypopituitarism and acute Sheehan syndrome) and patients with T1DM and adrenal insufficiency. However, serious hypoglycemia often develops in infants and children with primary adrenal insufficiency, especially after a period of fasting or during an intercurrent illness.

Patients with longstanding T1DM, adrenalectomized individuals, and those with autonomic neuropathy have impaired catecholamine responses during insulin-induced hypoglycemia,[84] but their increased risk for hypoglycemia can be compensated for by increases in the secretion of other counterregulatory hormones, in particular glucagon. No case of hypoglycemia secondary to isolated catecholamine deficiency have been reported in an adult. It is important to note that acute hypoglycemia can occur during surgical removal of a pheochromocytoma, presumably because of disinhibition of insulin release and abrupt withdrawal of the antiinsulin actions of catecholamines.

Pancreatectomized individuals, patients with longstanding T1DM, and those in whom insulin-requiring diabetes develops as a result of chronic pancreatitis or cystic fibrosis are glucagon-deficient and prone to severe hypoglycemia during treatment for their diabetes. An imbalance between insulin and glucagon secretion (relative glucagon deficiency) has been reported in reactive postprandial hypoglycemia. However, the condition is extremely rare.

Non–Islet Cell Tumors. A large variety of non–islet cell tumors are associated with hypoglycemia.[78] Tumors of mesenchymal origin are generally large and slow-growing, but often malignant. About one third are retroperitoneal, one third are intraabdominal, and one third are intrathoracic. Hypoglycemia may be the presenting sign in about half of patients with IGF-2–secreting tumors, and the treatment is surgical excision. Characteristically, circulating plasma insulin and C-peptide levels are suppressed. Excessive release of IGF-like molecules by the tumor can increase glucose utilization in tissues such as muscle, can suppress endogenous glucose production, and can reduce or overcome the secretion of counterregulatory hormones. Rarely, hypoglycemia results mainly from increased glucose utilization by the tumor.

KEY POINTS

- In nondiabetic individuals, hypoglycemia is rare, with blood glucose levels of 70 mg/dl (≤3.9 mmol/l) or less and 54 mg/dl (≤3.0 mmol/l) or less reported in as few as 1082/100,000 and 910/100,000 emergency medical service encounters, respectively. In hospitalized patients, nondiabetic hypoglycemia accounts for 20% to 30% of hospitalizations with hypoglycemia.

HYPOGLYCEMIA IN APPARENTLY WELL INDIVIDUALS

Endogenous Hyperinsulinism

Failure of insulin secretion to fall in the setting of falling and low blood glucose is abnormal. When the cause of hypoglycemia in an individual without diabetes is not apparent from a detailed history and physical examination, use of diagnostic laboratory studies can prove useful to guide diagnosis.

It is best to obtain laboratory studies during an episode of symptoms.[21] Laboratory evaluation should include: plasma glucose, insulin, C-peptide, β-hydroxybutyrate, proinsulin, and a sulfonylurea and meglitinide screen. If there is concern for endogenous hyperinsulinism, insulin antibodies and/or IGF-2 levels can be obtained but do not require the patient being hypoglycemic at the time of sampling.

If the symptoms occur mostly in the fasting state, the studies should be performed during a fast. This can often be achieved with an observed fast in the clinic or after fasting overnight with subsequent observation in the clinic. If this fails to result in symptomatic hypoglycemia (<55 mg/dL [3 mmol/L]) and there is significant clinical concern, a 72-hour fast should be performed.

The 72-hour fast (Table 38.4) should be supervised to prevent inadvertent caloric consumption or surreptitious drug administration. The patient should be encouraged to be active to simulate real-life situations and prevent sedentary decreases in glucose utilization. Blood samples for plasma glucose, insulin, C-peptide, β-hydroxybutyrate, and proinsulin are generally obtained every 6 hours, and more frequently as the blood glucose drops below 60 mg/dL (3.3 mmol/L). The evaluation is stopped if the blood glucose is less than 55 mg/dL with a history of Whipple's triad, the blood glucose is less than 45 mg/dL (2.5 mmol/L) with symptoms, or 72 hours have elapsed. At this point, a last set of blood samples is obtained, 1 mg of glucagon is administered intravenously, and glucose is measured at 10, 20, and 30 minutes. Finally, the patient is fed. Interpretation of spontaneous or fasted results is given in Table 38.5.

Insulin-Producing Islet Cell Tumors. Insulinoma is a very rare disorder with an incidence of 1 in 250,000 patient-years.[21,85] There is no ethnic predominance, and it is slightly more predominant in women than men. Although insulinomas occur at any age, most occur between 30 and 70 years of age. Approximately 80% of insulinomas are attributed to benign single adenomas, which average less than 2 cm in diameter; 10% are attributed to multiple benign adenomas. Malignant insulinomas are rare (<10% of insulinomas) and are more likely to occur with older age. Insulinoma associated with MEN1 is also rare and is more likely to occur with younger age.

Hypoglycemia in patients with insulinoma is the result of dysregulated insulin release. Insulin release is chaotic, often not appropriately increased by hyperglycemia, and some patients occasionally have impaired glucose tolerance. There is an increased proinsulin-to-insulin ratio and a decreased insulin concentration relative to normal islet tissue. Hypoglycemia occurs as the result of suppression of glucose production rather than increased glucose utilization, inasmuch as plasma insulin concentrations are usually only 50% to 150% above

TABLE 38.4 Sample Protocol for a 72-Hour Fast

Admit before the evening meal. Discontinue all nonessential medications. Insert an intravenous line for blood sampling.

Begin blood sampling (plasma glucose, insulin, C-peptide) just before the meal and continue every 30 minutes for 6 hours and thereafter every 2 to 3 hours.

Sampling should also include sulfonylureas/meglitinides. Additional studies could include morning cortisol, insulin antibodies, and/or IGF-2 based on clinical suspicion and do not necessarily require hypoglycemia.

Patient may consume calorie- and caffeine-lacking liquids and should ambulate.

The fast is ended at 72 hours or earlier if the patient has a plasma glucose level <45 mg/dL (2.5 mmol/L) with symptoms or <55 mg/dL (3.1 mmol/L) with prior Whipple's triad or if 72 hours have elapsed.

At the end of the fast, draw samples for all the above measurements and inject 1 mg of glucagon intravenously and measure plasma glucose at 10, 20, and 30 minutes.

IGF-1, Insulin-like growth factor-1.

TABLE 38.5 Interpretation of a 72-Hour Fast

Diagnosis	Symptoms or Signs	Plasma Glucose (mg/dL)	Plasma Insulin (μU/mL)	Plasma C-Peptide (nmol/L)	β-Hydroxybutyrate	Autoimmune Antibodies	IGF-mediated	Oral Agent Screen
Normal	No	<55	<3	<0.2	>2.7	–	N	–
Endogenous hyperinsulinism	Yes	<55	>3	>0.2	<2.7	–	N	–
Factitious insulin	Yes	<55	>3	<0.2	<2.7	–	N	–
Factitious oral agent	Yes	<55	>3	>0.2	<2.7	–	N	+
Non–islet cell tumor	Yes	<55	<3	<0.2	<2.7	–	↑	–
Autoimmune disorder	Yes	<55	>>3	>>0.2	<2.7	+	N	–

IGF, Insulin-like growth factor; N, normal.

normal levels (but of course inappropriate for the plasma glucose concentration).

Except in late-diagnosed malignant insulinoma cases in which an abdominal mass and signs of metastasis may be present, the physical examination is usually normal. Symptoms are the result of hypoglycemia. Patients with MEN1 also can exhibit symptoms as a result of hypercalcemia and accompanying hyperparathyroidism or other excessive hormone secretions (islet production of adrenocorticotropic hormone, gastrin, and vasoactive intestinal peptide). Frequency and/or severity of hypoglycemia increases over time. Patients with insulinoma can acquire the syndrome of hypoglycemia unawareness.

Although insulinoma has long been classified as one of the so-called fasting hypoglycemias, it is important to remember that such patients can experience hypoglycemia at any time of day, even postprandially. Patients often learn that eating frequently reduces episodes; therefore, weight gain has been observed in about 50% of patients.

There is often a delay in diagnosis. The "gold standard" for diagnosis of insulinoma remains the 72-hour fast. Various tests have been proposed over the years to rule in or rule out the presence of an insulinoma, but all suffer from frequent false positives and false negatives, and none are recommended. Hypoglycemia will develop in essentially all patients with insulinoma during the 72-hour fast; in fact, 75% will become hypoglycemic within 24 hours, and up to 90% over 72 hours.

Only after the diagnosis of hypoglycemia associated with inappropriate hyperinsulinemia has been established biochemically should localization studies be performed. Ninety-nine percent of insulinomas occur within the pancreas. Of the rare insulinomas that are not found in the pancreas, most are found in the wall of the duodenum or gastrosplenic omentum. Pancreatic insulinomas average about 2 cm in diameter and appear with equal frequency in the head, body, and tail.[85] Although approximately 75% to 90% can be identified correctly by palpation at surgery, preoperative localization generally is recommended to minimize manipulation of the pancreas, shorten operative time, avoid blind partial pancreatectomy when the tumor is not palpable, and assist in reoperations when scarring and fibrosis from prior surgery make palpation difficult.[306] Preoperative computed tomography, magnetic romance imaging endoscopic ultrasound, celiac arteriography, and transsplenic portal venous sampling can be used for localization. Currently, preoperative transabdominal ultrasonography followed by intraoperative ultrasonography is considered the most sensitive method.

Surgery is the treatment of choice for insulinoma. Without a diagnosis of MEN1, recurrence after surgery is 7%; with a diagnosis of MEN1, recurrence is as high as 21%. Multiple adenomas, hyperplasia, and malignancy require more extensive surgery. Debulking of a malignant tumor is worthwhile in helping to control hypoglycemia.

Medical therapy is reserved for operated patients with recurrence who refuse another exploration and for those with inoperable malignant tumors. Diazoxide, which inhibits insulin secretion, has been used most widely. The somatostatin analog octreotide has been found to be effective in some patients but must be injected. Continuous glucose monitoring may be useful.

Malignant insulinomas respond poorly to chemotherapy. Survival can be prolonged with multimodal therapy with chemoembolization, radiofrequency ablation, further liver resection, and/or liver transplantation in those without extrahepatic metastases.

Nesidioblastosis. Diffuse islet hyperplasia and/or nesidioblastosis is uncommon in adults (i.e., one to two cases per 100 million patient-years), and congenital hyperinsulinism refers to nesidioblastosis in infants.[21,86] The condition is benign and characterized by diffuse, although not necessarily uniform, hypertrophy and hyperplasia of islet cells (insulin- as well as glucagon- and somatostatin-producing cells), usually associated with differentiation of ductal cells into insulin-producing cells. Adult-onset nesidioblastosis typically occurs around 50 years of age, generally involves the entire pancreas, and is mostly associated with a history of gastric bypass surgery (see "Gastric Bypass Surgery").

It is worth emphasizing that many patients with nesidioblastosis will have a normal 72-hour fast, and their only biochemical abnormality will be postprandial hypoglycemia associated with abnormal plasma insulin and C-peptide levels, which is why it may be preferable to alter the 72-hour protocol and begin the 72-hour fast with a standard meal.

In patients who are suspected of having nesidioblastosis, localization with (^{18}F)-fluoro-3,4-dihydroxyphenylalnine positron emission tomography (^{18}F-DOPA PET) scans and/or the selective calcium infusion procedure, because these tests will confirm the diagnosis of pancreatic hyperinsulinism as the probable cause of the hypoglycemia when reliable localization procedures are negative.

Surgery is the treatment of choice for nesidioblastosis. In patients with nesidioblastosis, gradient-directed resection can result in cure of hypoglycemia. Treatment generally consists of 60% to 90% distal pancreatectomy. Half of patients undergoing this procedure are cured of hypoglycemia and are normoglycemic.

Gastric Bypass Surgery. Accompanying the rapid increase in obesity is an increase in the number of bariatric surgical procedures performed. Recently, accumulating reports are demonstrating an increase in severe postprandial hypoglycemia following gastric bypass surgery. The etiology is not well understood. Although the disorder was initially thought to represent a version of the adult nesidioblastosis syndrome, subsequent analysis of the pancreata have revealed that this may be caused by a combination of gastric dumping and inappropriately increased insulin secretion.[21,87] Patients with this syndrome have shown exaggerated GLP1 and insulin responses to a mixed meal. There is a strong female preponderance, and this condition can develop months or even years after bypass surgery is performed.

Treatment for hypoglycemia has proved to be problematic. Dietary changes, acarbose, diazoxide, and octreotide may alleviate symptoms. If a patient fails to respond to these measures, partial pancreatectomy is recommended.

Autoimmune Hypoglycemia. Autoimmune hypoglycemia is a very rare diagnosis caused by autoantibodies against insulin itself or against the insulin receptor. Both are rare and can produce fasting, as well as postprandial, reactive hypoglycemia—primarily the former.[21,88]

Insulin Autoimmune Syndrome. Insulin autoimmune syndrome (Hirata syndrome) is described as hyperinsulinemic hypoglycemia with high concentrations of insulin autoantibodies. It is classically described in individuals without exposure to exogenous insulin but has also been described in individuals treated with insulin. It has been best reported in patients of Japanese or Korean ethnicity. It can be described in isolation and has also been included in type 3A or type 4 autoimmune polyendocrine syndromes. The use of certain drugs (i.e., methimazole, carbimazole, alpha-lipoic acid) has been implicated in initiating the syndrome.

Hypoglycemia occurs because dissociation of insulin from the antibodies causes prolonged hyperinsulinemia. Postprandial hypoglycemia is more common with this syndrome than is fasting hypoglycemia. Circulating insulin levels are increased, and C-peptide levels may not be suppressed because they are in patients who are taking insulin surreptitiously. Unlike the other syndrome of autoimmune hypoglycemia, the course of this condition is benign and self-limited, with remission usually occurring within 1 year. Simple interventions such as frequent small meals with cornstarch and a low content of simple sugars often suffice. Rarely, glucocorticoids, somatostatin analogs, diazoxide, and immunosuppressive agents are used. CGM may be useful.

Antiinsulin Receptor Antibodies. This is an even more rare condition.[89] Hypoglycemia that develops as a result of antibodies directed against the insulin receptor has been reported. These antibodies act as agonists and produce hypoglycemia similar to insulin. Most patients have evidence of other conditions associated with altered autoimmunity. Some patients have developed severe insulin resistance as a result of the antibodies blocking the insulin receptor before the antibodies became agonists. Low circulating insulin and C-peptide levels, normal IGF-1, and appropriate counterregulatory hormone responses are characteristic in this condition.

With time, antibody titers appear to decline, and most patients have remission. Aggressive therapy is appropriate because of the severity of the hypoglycemia. Treatment with high-dose glucocorticoids, plasmapheresis, and alkylating agents has been attempted, with some success.

Factitious Hypoglycemia. Factitious hypoglycemia refers to a situation in which an individual intentionally attempts to create the impression of the presence of a hypoglycemic disorder, typically by using sulfonylureas or insulin.[21,90] This includes diabetic individuals who falsify their self–glucose monitoring or who overdose or underdose themselves to intentionally create the impression of better than actual glycemic control or brittle diabetes, as well as cases of child abuse. Generally excluded are inadvertent sulfonylurea ingestion (see "Drug-Induced Hypoglycemia"), homicides, suicide attempts, and substance abuse in which the additional use of insulin or sulfonylureas is intended to "obtain a high," rather than to create the impression of a hypoglycemic disorder. These situations nevertheless can present diagnostic challenges.

The condition is probably more common than this figure would indicate, because most cases go unreported.[91] In nearly all instances, the individuals had diabetes; had been a relative, spouse, or friend of a diabetic patient; or were in a medical or paramedical profession. Thus, factitious hypoglycemia should be suspected in individuals or their relatives who have unexplained hypoglycemia and knowledge of and/or access to insulin, sulfonylureas, or meglitinides.

Those in whom hypoglycemia develops as a result of surreptitious insulin injection will have inappropriate plasma insulin levels and a suppressed plasma C-peptide level. This finding is diagnostic. Patients taking sulfonylureas or a meglitinide will have inappropriate plasma insulin and C-peptide levels because of stimulation of endogenous insulin secretion, and this pattern will mimic that seen in patients with insulinoma. A urine or plasma assay for sulfonylureas and meglitinides is diagnostic. Many patients who are documented biochemically to have self-induced hypoglycemia will adamantly deny inducing the condition when confronted with the evidence. Nevertheless, psychological counseling with other physicians is warranted to prevent subsequent episodes or substitution of other potentially self-destructive behaviors.

Reactive Hypoglycemia. Reactive hypoglycemia refers to hypoglycemia that occurs after meals.[92] An OGTT is a test for diabetes mellitus and is inappropriate to use to evaluate the presence of reactive hypoglycemia. Generally, a standardized mixed meal is used for diagnosis, but some contend that hypoglycemia should be assessed during "normal" life. If the diagnosis is confirmed, one should exclude impaired glucose tolerance or mild diabetes, renal glycosuria, peptic ulcer disease, and other gastrointestinal disorders. Alimentary hypoglycemia can occur 2 to 4 hours after meal ingestion in patients who have undergone various abdominal procedures (i.e., gastrectomy or vagotomy) or with altered gastric motility, for example. Rapid gastric emptying and glucose absorption, which cause hyperglycemia and release of gut insulin secretagogues, result in excessive secretion of insulin; the biological actions of insulin to suppress endogenous glucose production and to stimulate tissue glucose uptake persist after the carbohydrate in the meal has been absorbed. This imbalance leads to postprandial hypoglycemia. GLP1, high insulin sensitivity, defects in counterregulation, and abnormal glucagon response are possible mechanisms. Treatment involves prevention of rapid absorption of large amounts of carbohydrate, frequent small meals, avoidance of large amounts of simple sugars, addition of fiber to the diet, and use of α-glucosidase inhibitors.

Pregnancy. Despite a decrease in insulin sensitivity and an increase in glucose turnover to accommodate the needs of the fetus, mean plasma glucose levels are normally 10% to 20% lower during the third and second trimesters of pregnancy, respectively.[93] Nevertheless, a great number of metabolic changes occur during pregnancy that make a woman more vulnerable to hypoglycemia. Any condition that can cause hypoglycemia in a nonpregnant woman can do so in a pregnant woman. First-trimester hypoglycemia is frequent in women treated with insulin, and one in four maternal deaths have been attributed to hypoglycemia ("dead in bed syndrome").[13,94] Balancing the priorities of glucose control in pregnancy for maternal and fetal outcomes, without excessive hypoglycemia, remains a challenge.

KEY POINTS

- Insulinoma is a very rare disorder occurring with an incidence of 1 in 250,000 patient-years. Approximately 80% of insulinomas are attributed to benign single adenomas; 10% are attributed to multiple benign adenomas. Malignant insulinomas are rare.

NEONATAL HYPOGLYCEMIA

Hypoglycemia is an important cause of morbidity in children.[95] There is a high incidence of neurodevelopmental delays, making the diagnosis critical.

Children's brains have a three times higher glucose utilization rate per kilogram body weight than adults. Thus, in situations such as hyperinsulinism or disorders of fatty acid oxidation, which are hypoketotic, there are not sufficient levels of alternate energy fuels for the brain, and the risk of brain failure increase. Two- to three-day-old neonates, infants, and children have similar metabolic defenses against hypoglycemia as adults. However, because of their larger brains, blood glucose drops more quickly, and hyperketonemia develops more rapidly.

In neonates aged older than 48 hours, infants, and children, blood glucose levels are generally the same as those observed in adults. In the neonate who is less than 48 hours old, glucose often decreases just after birth below levels observed in older infants and children. Neonates at risk of hypoglycemia include those who are large for gestational age, experienced perinatal stress, had premature or postmature delivery, have a diabetic mother, or have a family history of genetic hypoglycemia. Neonates with severe hypoglycemia, which is defined as an inability to consistently maintain blood glucose greater than 50 mg/dL for the first 48 hours of life and greater than 60 mg/dL thereafter, family history of genetic hypoglycemia and congenital syndromes should also be evaluated.

Numerous pathologic conditions can cause neonatal hypoglycemia (Table 38.6). It should be noted that, although inborn errors of metabolism and hormonal deficiencies can be responsible for hypoglycemia in neonates, adolescents and adults may present with hypoglycemia arising from such defects without prior diagnosis. However, these cases are less frequent and severe than in neonates.

TABLE 38.6 Causes of Neonatal Hypoglycemia and Inborn Errors of Metabolism

Acquired hyperinsulinism
- Maternal diabetes
- Intravenous glucose during labor and delivery
- Maternal oral hypoglycemic agents, propranolol, labetalol
- Low birth weight, prematurity, preeclampsia
- Neonatal sepsis, perinatal asphyxia

Counterregulatory hormone deficiency
- Hypopituitarism
- Growth hormone deficiency
- Adrenal insufficiency

Congenital hyperinsulinism
- Beckwith–Wiedemann syndrome
- Adenosine triphosphate–sensitive channel hyperinsulinism
- Sur-1 encoded by *ABCC8*
- Kir6.2 encoded by *KCNJ11*
- Glucokinase encoded by *GCK*
- Glutamate dehydrogenase encoded by *GLUD-1*
- Mitochondrial enzyme short-chain-3-hydroxyacety1-cort dehydrogenase encoded by *HAD1-1S*
- Disorders of glycosylation

Mitochondrial fatty acid oxidation disorders

Gluconeogenic disorders—fructose-1-6-biphosphate deficiency

Glycogen storage disorders
- Glucose-6-phosphate deficiency
- Glycogen synthase deficiency

Galactosemia

Maple syrup urine disease

Propionic acidemia

TABLE 38.7 Blood and Urine Tests for Metabolic and Endocrine Causes of Neonatal Hypoglycemia and Inborn Errors of Metabolism

Blood Glucose
- Lactate, pyruvate, free fatty acids, glycerol, ketones
- Amino acids
- Carnitine
- Acyl carnitine
- Insulin, C-peptide, proinsulin:insulin ratio, growth hormone, cortisol, thyroid hormones, thyroid-stimulating hormone (TSH)
- Growth hormone, cortisol, thyroid hormones, TSH
- Galactosemia screen

Urine
- Ketones
- Organic acids (3-hydroxybutyrate)
- Carnitine

Diagnostic Investigation of an Infant with Persistent Hypoglycemia. Because of the wide range of conditions that can cause hypoglycemia in an infant, a variety of blood and urine tests can be helpful in identifying metabolic and endocrine causes of the disorder (Table 38.7).[95]

It is surprising that plasma insulin levels often are not strikingly elevated in the congenital hyperinsulinism syndromes. Similar to insulinomas, plasma insulin levels often are modestly increased but are inappropriately raised in relation to the prevailing hypoglycemia. However, measurement of free fatty acids and β-hydroxybutyrate levels reveals dramatic suppression and evidence of significant insulin action. An increase in plasma glucose greater than 30 mg/dL (1.7 mmol/L) following a glucagon challenge test is also supportive of the diagnosis of relative hyperinsulinism. Genetic testing is now widely available for identification of genes associated with congenital hyperinsulinism, both monogenic and syndromic congenital forms.

Congenital Hyperinsulinism. The most common causes of persistent neonatal hypoglycemia are the congenital hyperinsulinism syndromes. Although they are rare overall, the incidence is increased in consanguineous families. These syndromes have been estimated to occur at a rate of approximately 1:50,000 births worldwide. The pathophysiology and the molecular and genetic bases for congenital hyperinsulinism have been reviewed by De Leon and Stanley.[96] Of these, the most frequently occurring involve loss-of-function mutations in the pancreatic β cell adenosine triphosphate–dependent potassium (K_{ATP}) channel, Kir6.2 channel, and Sur-1 receptor.

The clinical presentation of infants with congenital hyperinsulinism may be varied but typically involves prolonged and severe hypoglycemia with lethargy, seizures, and apnea, and babies are often large for gestational age. Typically, hypoglycemia occurs in the fasting or postprandial state, and these infants require large rates of glucose infusion greater than 10 mg/kg/min to prevent severe hypoglycemia.

Initially, treatment consists of diazoxide, which is a K_{ATP} channel agonist and has a suppressive effect on insulin secretion. If diazoxide does not produce a clinical response, octreotide has been used to suppress insulin release. However, in children who are unresponsive to medical therapy, the next treatment is surgery. This can comprise a local resection or a near-total (98%) pancreatectomy, depending upon the presence of focal or diffuse disease. The application of ^{18}F-DOPA PET scans to differentiate focal from diffuse pancreatic involvement in the congenital hyperinsulinism syndromes has provided a clinical breakthrough in the management of these children.

Mitochondrial Fatty Acid Oxidation Disorders. Eleven different mitochondrial fatty acid oxidation disorders can present with fasting (hypoketotic) or postprandial hypoglycemia.[97] Hepatic oxidation of fatty acids produces energy and acetyl-CoA, both of which are essential for gluconeogenesis. Additionally, fatty acid oxidation is required for hepatic ketone production. However, because of specific defects in mitochondrial fatty acid oxidation, gluconeogenesis is reduced, and this results in hypoglycemia and a lack of ketone bodies. The combination creates a scenario whereby the brain becomes starved of glucose and its primary alternative substrate (ketones), which creates conditions for severe neurologic consequences. Plasma membrane carnitine transporter deficiency is an autosomal recessive disorder that occurs in approximately 1:40,000 infants. Low plasma carnitine and acyl carnitine are indicative of this disorder, and treatment with dietary carnitine prevents hypoglycemia. Carnitine palmitoyl transferase 1 and 2 (CPT-1 and CPT-2) deficiencies can present as severe, life-threatening events that require frequent feeding and supplementation with dietary medium-chain triglycerides and L-carnitine (CPT-2 deficiency). Unfortunately, the most severe form of CPT-2 deficiency can present soon after birth and can be fatal. Medium-chain acetyl-CoA dehydrogenase deficiency is the most common fatty acid oxidation defect and is inherited as an autosomal recessive trait. Hypoglycemia usually occurs during fasting and conditions of stress such as viral illness. The presentation can be serious, with vomiting, apnea, coma, and encephalopathy. It has been estimated that, if left undiagnosed, up to 25% of children may die during the presenting event. Blood tests demonstrate elevated medium-chain acylcarnitine. As with other mitochondrial fatty acid oxidation disorders, the enzyme deficiency can be demonstrated diagnostically in fibroblasts. Treatment consists of avoidance of fasting and precipitation of attacks that could lead to neurologic defects and even death. Thus, frequent meals, bedtime snacks, uncooked cornstarch (to avoid nocturnal hypoglycemia), and carnitine supplementation have been used successfully in these children.

Very long chain acyl-CoA dehydrogenase (VLCAD) and short-chain acyl-CoA dehydrogenase (SCAD) deficiencies can present with serious, sometimes fatal, episodes of hypoglycemia and metabolic acidosis. These defects are also associated with cardiomyopathy (VLCAD), hepatomegaly, and neurologic defects (SCAD). Treatment involves avoidance of fasting and maintenance of glucose levels during stress, with frequent meals enriched with carbohydrate and medium-chain triglycerides. Carnitine supplementation in these conditions is not recommended (VLCAD) or has limited applicability (SCAD).

Long-chain 3-hydroxyacyl-CoA dehydrogenase (LCHAD) and mitochondrial trifunctional protein deficiency are also autosomal recessive disorders. Both conditions can present with severe hypoglycemia, encephalopathy, neurologic complications, and death. Both conditions can result in peripheral neuropathy, and LCHAD deficiency can result in pigmentary retinopathy and blindness. Treatment is similar to that provided for the above fatty acid oxidation defects. Long-chain ketothilase and 2-4-dienoyl-CoA reductase deficiency are very rare conditions.

Glycogen Storage Diseases. All glycogen storage diseases are caused by an inability to use or store glycogen.[98] Glycogen synthase deficiency, glucose-6-phosphatase deficiency, and glycogen debrancher deficiency often show in the neonatal period with hypoglycemia. Additionally, glycogen-phosphorylase deficiency can present later in childhood. Of the disorders, glucose-6-phosphate deficiency is associated with the most severe hypoglycemic phenotype. Prevention of hypoglycemia is the same for all of the glycogen storage disorders. Prevention of hypoglycemia during stress, exercise, and during the nighttime hours includes consuming frequent carbohydrate-containing meals, as well as bedtime snacks. Intravenous glucose may be necessary to maintain plasma glucose levels above 70 mg/dL (3.9 mmol/L) under intense stress. Lysosomal acid α-glucosidase deficiency can be treated with enzyme replacement therapy using recombinant alglucosidase alfa (Lumizyme), which degrades lysosomal glycogen.

CONCLUSION AND FUTURE DIRECTIONS

Hypoglycemia remains a significant problem for those with diabetes. Future advancements in technology, including increasingly less burdensome CGM and progressively more complex integration of CGM with insulin delivery (perhaps combined with other hormone therapy), are likely to result in significant improvements in quality of life for those who suffer from hypoglycemia. There remains much to learn about the concept of HAAF and how the underlying physiology can be used to direct therapies against this challenge. Lastly, elucidating the relationship between hypoglycemia and cardiovascular risk is critical, as is an understanding of which populations are vulnerable to this association.

REFERENCES

1. International Hypoglycaemia Study Group. Glucose concentrations of less than 3.0 mmol/l (54 mg/dl) should be reported in clinical trials: a joint position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*. 2017;40:155–157.
2. Sacks DB, Arnold M, Bakris GL, et al. Position statement executive summary: guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care*. 2011;34:1419–1423.
3. Tarasova VD, Zena M, Rendell M. Artifactual hypoglycemia: an old term for a new classification. *Diabetes Care*. 2014;37:e85–e86.
4. Vogt MC, Brüning JC. CNS insulin signaling in the control of energy homeostasis and glucose metabolism—from embryo to old age. *Trends Endocrinol Metab*. 2013;24:76–84.
5. Davis SN, Colburn C, Dobbins R, et al. Evidence that the brain of the conscious dog is insulin sensitive. *J Clin Invest*. 1995;95:593–602.
6. Diggs-Andrews KA, Zhang X, Song Z, et al. Brain insulin action regulates hypothalamic glucose sensing and the counterregulatory response to hypoglycemia. *Diabetes*. 2010;59:2271–2280.
7. Stanley S, Moheet A, Seaquist ER. Central mechanisms of glucose sensing and counterregulation in defense of hypoglycemia. *Endocr Rev*. 2019;40:768–788.
8. Gruetter R. Glycogen: the forgotten cerebral energy store. *J Neurosci Res*. 2003;74:179–183.
9. Lee JJ, Khoury N, Shackleford AM, et al. Dissociation between hormonal counterregulatory responses and cerebral glucose metabolism during hypoglycemia. *Diabetes*. 2017;66:2964–2972.
10. Lachin JM, McGee P, Palmer JP. Impact of C-peptide preservation on metabolic and clinical outcomes in the diabetes control and complications trial. *Diabetes*. 2014;63:739–748.
11. Ryan EA, Paty BW, Senior PA, et al. Five-year follow-up after clinical islet transplantation. *Diabetes*. 2005;54:2060–2069.
12. Hypoglycaemia, cardiovascular disease, and mortality in diabetes: epidemiology, pathogenesis, and management. *Lancet Diabetes Endocrinol*. 2019;7:385–396.
13. Hsieh A, Twigg SM. The enigma of the dead-in-bed syndrome: challenges in predicting and preventing this devastating complication of type 1 diabetes. *J Diabetes Complications*. 2014;28:585–587.
14. Fu Z, Gilbert ER, Liu D. Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes. *Curr Diabetes Rev*. 2013;9:25–53.
15. Adeva-Andany MM, Funcasta-Calderón R, Fernández-Fernández C, et al. Metabolic effects of glucagon in humans. *J Clin Transl Endocrinol*. 2019;15:45–53.
16. Cryer PE. Minireview: glucagon in the pathogenesis of hypoglycemia and hyperglycemia in diabetes. *Endocrinology*. 2012;153:1039–1048.
17. Meyer C, Stumvoll M, Welle S, et al. Relative importance of liver, kidney, and substrates in epinephrine-induced increased gluconeogenesis in humans. *Am J Physiol Endocrinol Metab*. 2003;285:E819–E826.
18. De Feo P, Perriello G, De Cosmo S, et al. Comparison of glucose counterregulation during short-term and prolonged hypoglycemia in normal humans. *Diabetes*. 1986 May;35:563–569.
19. De Feo P, Perriello G, Torlone E, et al. Contribution of cortisol to glucose counterregulation in humans. *Am J Physiol*. 1989;257:E35–E42.
20. De Feo P, Perriello G, Torlone E, et al. Demonstration of a role for growth hormone in glucose counterregulation. *Am J Physiol*. 1989;256:E835–E843.
21. Cryer PE, Axelrod L, Grossman AB, et al. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2009;94:709–728.
22. Cryer PE. Mechanisms of hypoglycemia-associated autonomic failure in diabetes. *N Engl J Med*. 2013;369:362–372.
23. Davis SN, Mann S, Briscoe VJ, et al. Effects of intensive therapy and antecedent hypoglycemia on counterregulatory responses to hypoglycemia in type 2 diabetes. *Diabetes*. 2009;58:701–709.
24. Ertl AC, Davis SN. Evidence for a vicious cycle of exercise and hypoglycemia in type 1 diabetes mellitus. *Diabetes Metab Res Rev*. 2004;20:124–130.
25. American Diabetes Association. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2018. *Diabetes Care*. 2018;41:S73–S85.
26. Cryer PE. Hypoglycemia in type 1 diabetes mellitus. *Endocrinol Metab Clin North Am*. 2010;39:641–654.
27. Sandoval DA, Ertl AC, Richardson MA, et al. Estrogen blunts neuroendocrine and metabolic responses to hypoglycemia. *Diabetes*. 2003;52:1749–1755.
28. Davis SN, Shavers C, Costa F. Gender-related differences in counterregulatory responses to antecedent hypoglycemia in normal humans. *J Clin Endocrinol Metab*. 2000;85:2148–2157.

29. Fan X, Ding Y, Brown S, et al. Hypothalamic AMP-activated protein kinase activation with AICAR amplifies counterregulatory responses to hypoglycemia in a rodent model of type 1 diabetes. *Am J Physiol Regul Integr Comp Physiol*. 2009;296:R1702–R1708.
30. Gabriely I, Wozniak R, Hawkins M, et al. Troglitazone amplifies counterregulatory responses to hypoglycemia in nondiabetic subjects. *J Clin Endocrinol Metab*. 2001;86:521–528.
31. Farhat R, Su G, Sejling A-S, et al. Carvedilol prevents counterregulatory failure and impaired hypoglycaemia awareness in non-diabetic recurrently hypoglycaemic rats. *Diabetologia*. 2019;62:676–686.
32. Saleh TY, Cryer PE. Alanine and terbutaline in the prevention of nocturnal hypoglycemia in IDDM. *Diabetes Care*. 1997;20:1231–1236.
33. Vele S, Milman S, Shamoon H, et al. Opioid receptor blockade improves hypoglycemia-associated autonomic failure in type 1 diabetes mellitus. *J Clin Endocrinol Metab*. 2011;96:3424–3431.
34. Milman S, Leu J, Shamoon H, et al. Opioid receptor blockade prevents exercise-associated autonomic failure in humans. *Diabetes*. 2012;61:1609–1615.
35. Briscoe VJ, Ertl AC, Tate DB, et al. Effects of the selective serotonin reuptake inhibitor fluoxetine on counterregulatory responses to hypoglycemia in individuals with type 1 diabetes. *Diabetes*. 2008;57:3315–3322.
36. Lamos EM, Younk LM, Davis SN. Hypoglycemia. In: *Endocrine and Metabolic Medical Emergencies [Internet]*. John Wiley & Sons, Ltd; 2018:506–530. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119374800.ch31.
37. Pontiroli AE, Tagliabue E. Intranasal versus injectable glucagon for hypoglycemia in type 1 diabetes: systematic review and meta-analysis. *Acta Diabetol*. 2020;57:743–749.
38. Kumar JG, Abhilash KPP, Saya RP, et al. A retrospective study on epidemiology of hypoglycemia in Emergency Department. *Indian J Endocrinol Metab*. 2017;21:119–124.
39. Davis SN, Duckworth W, Emanuele N, et al. Effects of severe hypoglycemia on cardiovascular outcomes and death in the Veterans Affairs Diabetes Trial. *Diabetes Care*. 2019;42:157–163.
40. Ólafsdóttir AF, Polonsky W, Bolinder J, et al. A randomized clinical trial of the effect of continuous glucose monitoring on nocturnal hypoglycemia, daytime hypoglycemia, glycemic variability, and hypoglycemia confidence in persons with type 1 diabetes treated with multiple daily insulin injections. *Diabetes Technol Ther*. 2018;20:274–284.
41. Hypoglycemia in the Diabetes Control and Complications Trial. The diabetes control and complications trial research group. *Diabetes*. 1997;46:271–286.
42. Gubitosi-Klug RA, Braffett BH, White NH, et al. Risk of severe hypoglycemia in type 1 diabetes over 30 years of follow-up in the DCCT/EDIC Study. *Diabetes Care*. 2017;40:1010–1016.
43. Dunkley AJ, Fitzpatrick C, Gray LJ, et al. Incidence and severity of hypoglycaemia in type 2 diabetes by treatment regimen: a UK multisite 12-month prospective observational study. *Diabetes Obes Metab*. 2019;21:1585–1595.
44. Gerstein HC, Miller ME, Genuth S, et al. Long-term effects of intensive glucose lowering on cardiovascular outcomes. *N Engl J Med*. 2011;364:818–828.
45. Lipska KJ, Yao X, Herrin J, et al. Trends in drug utilization, glycemic control, and rates of severe hypoglycemia, 2006-2013. *Diabetes Care*. 2017;40:468–475.
46. Lipska KJ, Warton EM, Huang ES, et al. HbA1c and risk of severe hypoglycemia in type 2 diabetes: the Diabetes and Aging Study. *Diabetes Care*. 2013;36:3535–3542.
47. 6. Glycemic targets: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S66–S76.
48. Akirov A, Grossman A, Shochat T, et al. Mortality among hospitalized patients with hypoglycemia: insulin related and noninsulin related. *J Clin Endocrinol Metab*. 2017;102:416–424.
49. Araque KA, Kadayakkara DK, Gigauri N, et al. Reducing severe hypoglycaemia in hospitalised patients with diabetes: early outcomes of standardised reporting and management. *BMJ Open Qual*. 2018;7:e000120.
50. Rushakoff RJ, Sullivan MM, MacMaster HW, et al. Association between a virtual glucose management service and glycemic control in hospitalized adult patients: an observational study. *Ann Intern Med*. 2017;166:621–627.
51. 15. Diabetes care in the hospital: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S193–S202.
52. Finfer S, Chittock DR, Su SY-S, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360:1283–1297.
53. Schopman JE, Simon ACR, Hoefnagel SJM, et al. The incidence of mild and severe hypoglycaemia in patients with type 2 diabetes mellitus treated with sulfonylureas: a systematic review and meta-analysis. *Diabetes Metab Res Rev*. 2014;30:11–22.
54. Melo KFS, Bahia LR, Pasinato B, et al. Short-acting insulin analogues versus regular human insulin on postprandial glucose and hypoglycemia in type 1 diabetes mellitus: a systematic review and meta-analysis. *Diabetol Metab Syndr*. 2019;11:2.
55. Rosenstock J, Fonseca V, Schinzel S, et al. Reduced risk of hypoglycemia with once-daily glargine versus twice-daily NPH and number needed to harm with NPH to demonstrate the risk of one additional hypoglycemic event in type 2 diabetes: evidence from a long-term controlled trial. *J Diabetes Complications*. 2014;28:742–749.
56. Pala L, Dicembrini I, Mannucci E. Continuous subcutaneous insulin infusion vs modern multiple injection regimens in type 1 diabetes: an updated meta-analysis of randomized clinical trials. *Acta Diabetol*. 2019;56:973–980.
57. Qin Y, Yang LH, Huang XL, et al. Efficacy and safety of continuous subcutaneous insulin infusion vs. multiple daily injections on type 1 diabetes children: a meta-analysis of randomized control trials. *J Clin Res Pediatr Endocrinol*. 2018;10:316–323.
58. van Beers CAJ, DeVries JH, Kleijer SJ, et al. Continuous glucose monitoring for patients with type 1 diabetes and impaired awareness of hypoglycaemia (IN CONTROL): a randomised, open-label, crossover trial. *Lancet Diabetes Endocrinol*. 2016;4:893–902.
59. Dicembrini I, Mannucci E, Monami M, et al. Impact of technology on glycaemic control in type 2 diabetes: a meta-analysis of randomized trials on continuous glucose monitoring and continuous subcutaneous insulin infusion. *Diabetes Obes Metab*. 2019;21:2619–2625.
60. Reddy M, Jugnee N, El Laboudi A, et al. A randomized controlled pilot study of continuous glucose monitoring and flash glucose monitoring in people with type 1 diabetes and impaired awareness of hypoglycaemia. *Diabet Med*. 2018;35:483–490.
61. Wallia A, Umpierrez GE, Rushakoff RJ, et al. Consensus statement on inpatient use of continuous glucose monitoring. *J Diabetes Sci Technol*. 2017;11:1036–1044.
62. Abraham MB, Nicholas JA, Smith GJ, et al. Reduction in hypoglycemia with the predictive low-glucose management system: a long-term randomized controlled trial in adolescents with type 1 diabetes. *Diabetes Care*. 2018;41:303–310.
63. Ly TT, Nicholas JA, Retterath A, et al. Effect of sensor-augmented insulin pump therapy and automated insulin suspension vs standard insulin pump therapy on hypoglycemia in patients with type 1 diabetes: a randomized clinical trial. *J Am Med Assoc*. 2013;310:1240–1247.
64. Brown SA, Kovatchev BP, Raghinaru D, et al. Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes. *N Engl J Med*. 2019;381:1707–1717.
65. Forlenza GP, Ekhlaspour L, Breton M, et al. Successful at-home use of the tandem control-iq artificial pancreas system in young children during a randomized controlled trial. *Diabetes Technol Ther*. 2019;21:159–169.
66. Rickels MR, Robertson RP. Pancreatic islet transplantation in humans: recent progress and future directions. *Endocr Rev*. 2019;40:631–668.
67. Colberg SR, Sigal RJ, Yardley JE, et al. Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. *Diabetes Care*. 2016;39:2065–2079.
68. Lamos EM, Malek R, Munir KM. Effects of intermittent fasting on health, aging, and disease. *N Engl J Med*. 2020;382:1771.
69. Vihonen H, Kuisma M, Nurmi J. Hypoglycaemia without diabetes encountered by emergency medical services: a retrospective cohort study. *Scand J Trauma Resusc Emerg Med*. 2018;26:12.
70. Sako A, Yasunaga H, Matsui H, et al. Hospitalization with hypoglycemia in patients without diabetes mellitus: a retrospective study using a national inpatient database in Japan, 2008-2012. *Medicine (Baltim)*. 2017;96:e7271.

71. Murad MH, Coto-Yglesias F, Wang AT, et al. Clinical review: drug-induced hypoglycemia: a systematic review. *J Clin Endocrinol Metab*. 2009;94:741–745.
72. Marks V, Teale JD. Drug-induced hypoglycemia. *Endocrinol Metab Clin North Am*. 1999;28:555–577.
73. Klatt EC, Beatie C, Noguchi TT. Evaluation of death from hypoglycemia. *Am J Forensic Med Pathol*. 1988;9:122–125.
74. Vriesendorp TM, DeVries JH, Rosendaal F, et al. Severe hypoglycemia in critically ill: risk and outcomes. *Crit Care Med*. 2008;36:1390; author reply 1390–1391.
75. Maitra SR, Wojnar MM, Lang CH. Alterations in tissue glucose uptake during the hyperglycemic and hypoglycemic phases of sepsis. *Shock*. 2000;13:379–385.
76. Alsahli M, Gerich JE. Renal glucose metabolism in normal physiological conditions and in diabetes. *Diabetes Res Clin Pract*. 2017;133:1–9.
77. Nouel O, Bernuau J, Rueff B, et al. Hypoglycemia. A common complication of septicemia in cirrhosis. *Arch Intern Med*. 1981;141:1477–1478.
78. Garla V, Sonani H, Palabindala V, et al. Non-islet cell hypoglycemia: case series and review of the literature. *Front Endocrinol*. 2019;10:316.
79. Arem R. Hypoglycemia associated with renal failure. *Endocrinol Metab Clin North Am*. 1989;18:103–121.
80. Alsahli M, Gerich JE. Hypoglycemia in patients with diabetes and renal disease. *J Clin Med*. 2015;4:948–964.
81. Hedayati HA, Beheshti M. Profound spontaneous hypoglycaemia in congestive heart failure. *Curr Med Res Opin*. 1977;4:501–504.
82. Leibovitz E, Adler H, Giryes S, et al. Malnutrition risk is associated with hypoglycemia among general population admitted to internal medicine units. Results from the MENU study. *Eur J Clin Nutr*. 2018;72:888–893.
83. Mehler PS, Blalock DV, Walden K, et al. Medical findings in 1,026 consecutive adult inpatient-residential eating disordered patients. *Int J Eat Disord*. 2018;51:305–313.
84. De Feo P, Perriello G, Torlone E, et al. Contribution of adrenergic mechanisms to glucose counterregulation in humans. *Am J Physiol*. 1991;261:E725–E736.
85. Okabayashi T, Shima Y, Sumiyoshi T, et al. Diagnosis and management of insulinoma. *World J Gastroenterol*. 2013;19:829–837.
86. Kowalewski AM, Szylberg Ł, Kasperska A, et al. The diagnosis and management of congenital and adult-onset hyperinsulinism (nesidioblastosis)—literature review. *Pol J Pathol*. 2017;68:97–101.
87. Meier JJ, Butler AE, Galasso R, et al. Hyperinsulinemic hypoglycemia after gastric bypass surgery is not accompanied by islet hyperplasia or increased beta-cell turnover. *Diabetes Care*. 2006;29:1554–1559.
88. Cappellani D, Macchia E, Falorni A, et al. Insulin autoimmune syndrome (Hirata disease): a comprehensive review fifty years after its first description. *Diabetes Metab Syndr Obes*. 2020;13:963–978.
89. Taylor SI, Barbetti F, Accili D, et al. Syndromes of autoimmunity and hypoglycemia. Autoantibodies directed against insulin and its receptor. *Endocrinol Metab Clin North Am*. 1989;18:123–143.
90. Awad DH, Gokarakonda SB, Ilahi M. Factitious hypoglycemia. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2021. Available from: https://www.ncbi.nlm.nih.gov/books/NBK542310/.
91. Marks V, Teale JD. Hypoglycemia: factitious and felonious. *Endocrinol Metab Clin North Am*. 1999;28:579–601.
92. Brun JF, Fedou C, Mercier J. Postprandial reactive hypoglycemia. *Diabetes Metab*. 2000;26:337–351.
93. Nigam A, Varun N, Sharma S, et al. Glycaemic profile in the second and third trimesters of normal pregnancy compared to non-pregnant adult females. *Obstet Med*. 2020;13:30–36.
94. Leinonen PJ, Hiilesmaa VK, Kaaja RJ, et al. Maternal mortality in type 1 diabetes. *Diabetes Care*. 2001;24:1501–1502.
95. Thornton PS, Stanley CA, De Leon DD, et al. Recommendations from the Pediatric Endocrine Society for evaluation and management of persistent hypoglycemia in neonates, infants, and children. *J Pediatr*. 2015;167:238–245.
96. De León DD, Stanley CA. Mechanisms of Disease: advances in diagnosis and treatment of hyperinsulinism in neonates. *Nat Clin Pract Endocrinol Metab*. 2007;3:57–68.
97. Kompare M, Rizzo WB. Mitochondrial fatty-acid oxidation disorders. *Semin Pediatr Neurol*. 2008;15:140–149.
98. Stone WL, Basit H, Adil A. Glycogens storage disease. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2021. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459277/.

39

Treatment of Type 1 Diabetes Mellitus in Adults

Ravi Retnakaran and Anne L. Peters

OUTLINE

Goals of Management, 672
  Glycemic Goals, 674
Acute Complications, 675
Chronic Complications, 675
Quality of Life, 675
Team Approach to Management, 675
Monitoring, 677
  Glycemic Control, 677
  Ketone Testing, 679
  Complication Surveillance, 679
Insulin Therapy, 679
  Principles of Insulin Replacement, 679
  Insulin Preparations, 680
  Intensive Insulin Therapy Regimens, 683
  Adjustment of Insulin Therapy in Special Situations, 685
Nutrition, 685
Other Therapies For Type 1 Diabetes, 686
  Metformin, DPP-IV Inhibitors, and GLP-1 RAs, 686
Future Perspectives, 686

The discovery of insulin in the summer of 1921 by Frederick Banting, Charles Best, James Collip, and J. J. R. Macleod at the University of Toronto and its first therapeutic use on January 11, 1922, at the Toronto General Hospital, stand as one of the greatest achievements of contemporary medicine.[1] By providing life-sustaining therapy for a previously fatal wasting condition, the discovery of insulin dramatically shifted the prognosis of type 1 diabetes to that of a chronic disease that could be treated but not cured. Research since then has shown that achieving and maintaining near-euglycemia can prevent and/or slow the progression to microvascular complications, and to some degree macrovascular disease. The Diabetes Control and Complications Trial (DCCT) and other studies evaluating optimization of glycemic control have provided compelling evidence for the benefits of tight glycemic control.[2-5] However, outside of a research setting, normalizing and maintaining near-euglycemia can be a daunting task, and targets are not achieved by the majority of people with type 1 diabetes.[6] Newer insulins and technologies have made this goal increasingly possible. But the clinical practice of type 1 diabetes management must be able to provide the support, knowledge, and education patients need to adopt and implement these new tools. An interdisciplinary healthcare team consisting of individuals such as physicians, diabetes educators, dietitians, mental health providers, pharmacists, and anyone else the patient needs for support is key to providing the resources a patient may need. In this chapter, we review the essential elements of this approach to comprehensive care of patients with type 1 diabetes.

GOALS OF MANAGEMENT

The overriding goal in the management of type 1 diabetes is the optimization of the patient's quality of life through the prevention and amelioration of the acute and chronic complications that are associated with this disease. In the modern era, a substantial body of epidemiologic and experimental evidence has established the relationship between diabetic complications and hyperglycemia (see Chapter 42). In particular, the findings of the DCCT have had an enormous impact on the management of type 1 diabetes and warrant review before discussion of specific therapeutic goals.

The DCCT was a multicenter, National Institutes of Health–funded randomized, controlled clinical trial initiated in 1982 that was designed to assess whether intensive treatment with the goal of maintaining near-normal blood glucose concentrations could reduce the frequency and severity of diabetic complications.[2] A total of 1441 patients with type 1 diabetes from the United States and Canada participated in the study and made up two cohorts: a primary prevention cohort of 726 subjects with no retinopathy at baseline and a duration of diabetes of 1 to 5 years, and a secondary intervention cohort of 715 individuals with mild retinopathy and duration of diabetes of up to 15 years. Participants were randomly assigned to receive either the conventional therapy of the day (one or two injections of insulin per day) or intensive therapy (either three or more insulin injections or continuous subcutaneous insulin infusion (CSII) by external pump, with frequent blood glucose monitoring). After a mean of 6.5 years of follow-up, mean glycosylated hemoglobin (A_{1c}) was 7.2% in the intensive therapy arm and 9.0% in the conventional treatment arm. The intensive therapy group showed a relative risk reduction of 76% (95% confidence

KEY POINTS

- The goal of type 1 diabetes management is the achievement of sustained euglycemia, with avoidance of hypoglycemia, through the physiologic replacement of insulin.
- Intensive insulin therapy, assisted by technology, is recommended for all patients with type 1 diabetes.
- Intensive insulin therapy can reduce the risk for microvascular and macrovascular complications in type 1 diabetes.

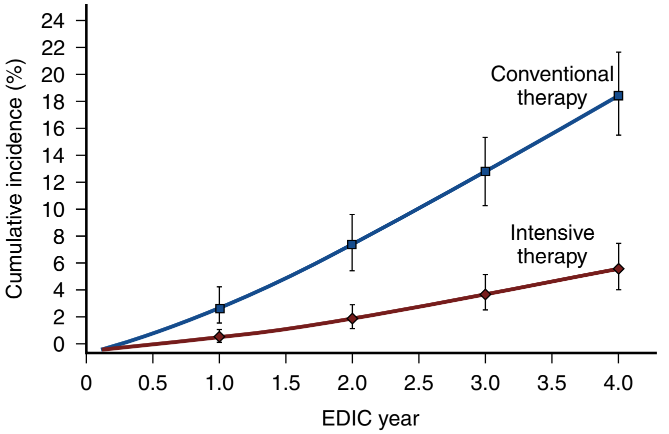


Fig. 39.1 Cumulative incidence of further progression of retinopathy during the Epidemiology of Diabetes Interventions and Complications study in the former conventional therapy and intensive therapy cohorts from the Diabetes Control and Complications Trial (DCCT). Data are based on regression analysis adjusted for the level of retinopathy at the end of the DCCT, whether patients received therapy as primary prevention or as secondary intervention, and for the duration of diabetes and the glycosylated hemoglobin value on enrollment in the DCCT. (From Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med.* 2000;342:387.)

interval [CI]: 62%–85%) for the development of retinopathy and a 54% relative risk reduction (95% CI: 39%–66%) for progression of retinopathy. Similarly, intensive therapy was associated with relative risk reductions of 39% (95% CI: 21%–52%) for the development of microalbuminuria (defined as urinary albumin excretion >40 mg per 24 hours) and 54% (95% CI: 19%–74%) for progression to albuminuria (urinary albumin excretion >300 mg per 24 hours). The incidence of clinical neuropathy was also decreased in the intensive group, with a relative risk reduction of 60% (95% CI: 38%–74%). Thus, the DCCT provided conclusive evidence that improved glycemic control with intensive therapy can reduce the incidence of microvascular complications (primary prevention) and can slow the progression of established microvascular disease (secondary intervention).

The benefits of near-normal glycemic control were found to persist for many years after the DCCT, according to data from the follow-up study, the Epidemiology of Diabetes Interventions and Complications (EDIC).[7] In the ongoing EDIC observational study, a large cohort consisting of more than 95% of the DCCT participants (n = 1375) has been followed since the completion of the original trial and has been studied by using an intention-to-treat approach based on the original randomization in the DCCT. The mean A_{1c} levels of the former intensive treatment cohort and the former conventional therapy group have converged ($A_{1c} \approx 8\%$) over the course of the EDIC study and were no longer significantly different at 5 years post-DCCT. Remarkably, however, despite similar glycemic control, the former intensive therapy group continued to display reduced risk for both new retinopathy and new nephropathy at 4 years and 7 to 8 years post-DCCT, respectively, compared with the group of patients who were randomized originally to conventional therapy (Fig. 39.1).[7,8] With continued follow-up in the EDIC study, the former intensive therapy group has exhibited further benefits, including risk reduction for the development of impaired glomerular filtration and hypertension.[9,10] These persistent beneficial effects of previous near-normal glycemic control have extended the results of the DCCT and demonstrate that intensive treatment should be initiated as early as is safely possible in the management of type 1 diabetes.

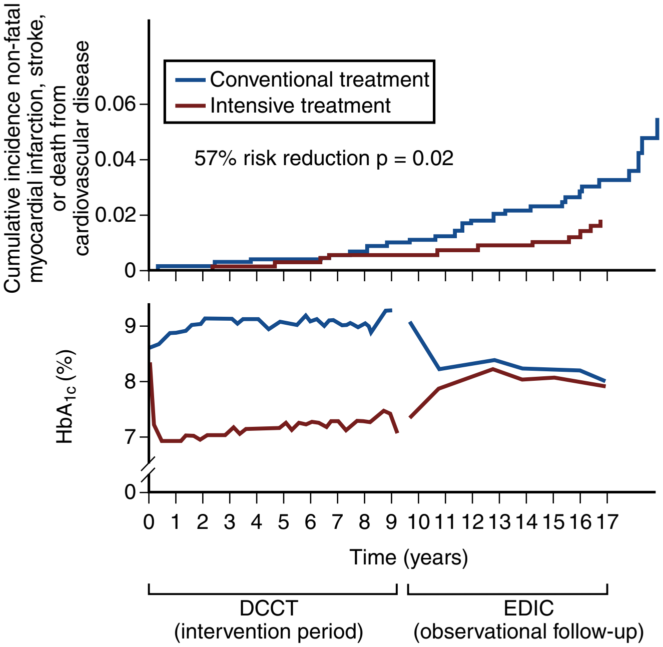


Fig. 39.2 Differences in glycosylated hemoglobin *(lower panel)* and cardiovascular event rates *(upper panel)* over time between the intensive and conventional diabetes treatment arms of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study cohort. (From Retnakaran R, Zinman B. Type 1 diabetes, hyperglycaemia and the heart. *Lancet.* 2008;371:1790–1799.)

Although the DCCT conclusively documented that intensive therapy can reduce the risk for microvascular complications, the effect on macrovascular disease was initially less clear. By the end of the DCCT, the intensive therapy arm displayed a relative risk reduction of 34% (95% CI: 7%–54%) for the development of elevated low-density lipoprotein (LDL) cholesterol levels. Nevertheless, even though intensive therapy was associated with a 41% relative risk reduction (95% CI: –10% to 68%) for macrovascular complications, this difference was not statistically significant.[2] It is important to note, however, that the macrovascular benefits derived from near-normal glycemic control in the former intensive treatment arm of the DCCT have emerged over time in the EDIC study. Indeed, after a mean follow-up of 17 years, it was observed that intensive therapy during the DCCT had reduced the subsequent risk for any cardiovascular disease (CVD) event by 42% (95% CI: 9%–63%; $P = 0.02$) and the risk for nonfatal myocardial infarction, stroke, or cardiovascular death by 57% (95% CI: 12%–79%; $P = 0.02$) (Fig. 39.2).[11] This remarkable risk reduction was largely explained by differences in the updated mean A_{1c} during the DCCT. Furthermore, the reduction in CVD events was consistent with earlier observations that intensive therapy had reduced the progression of atherosclerosis (as measured by carotid artery intima media thickness [IMT]) and the prevalence of coronary artery calcification (CAC) during the DCCT/EDIC study, with both effects associated with the reduction in mean A_{1c} achieved with intensive therapy.[12,13] Thus, intensive therapy (with associated reduced glycemic exposure) during the mean 6.5 years of the DCCT first significantly reduced surrogate measures of atherosclerotic disease (carotid IMT, CAC) and later reduced clinical CVD events over the course of the EDIC study (Fig. 39.3). These data demonstrate that the long-term risk associated with past glycemic exposure extends to both microvascular and macrovascular disease.[14] Indeed, it has been hypothesized that glycemic exposure early in the course of the disease may have an effect of metabolic imprinting (referred to as "metabolic memory") that contributes to future risk for vascular complications.[11] Taken together, these data

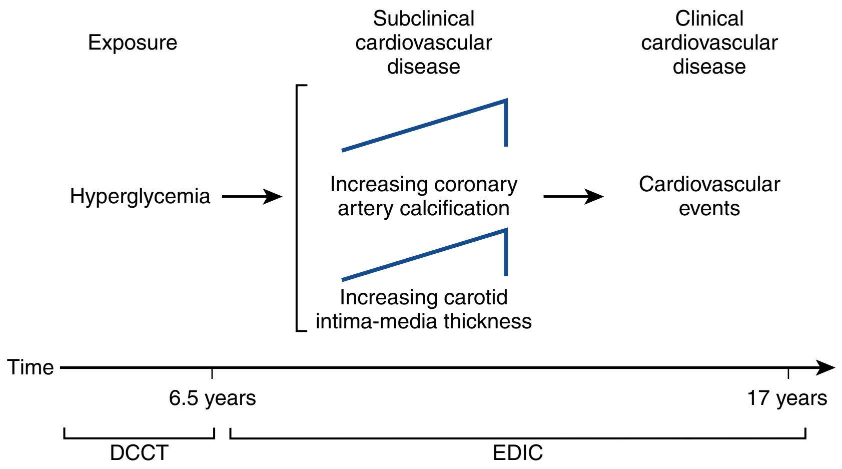

Exposure
Subclinical cardiovascular disease
Clinical cardiovascular disease
Hyperglycemia
Increasing coronary artery calcification
Cardiovascular events
Increasing carotid intima-media thickness
Time
6.5 years
17 years
DCCT
EDIC

Fig. 39.3 **Pathophysiologic model of cardiovascular disease in type 1 diabetes based on observations from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study.** (From Retnakaran R, Zinman B. Type 1 diabetes, hyperglycaemia and the heart. *Lancet.* 2008;371:1790–1799.)

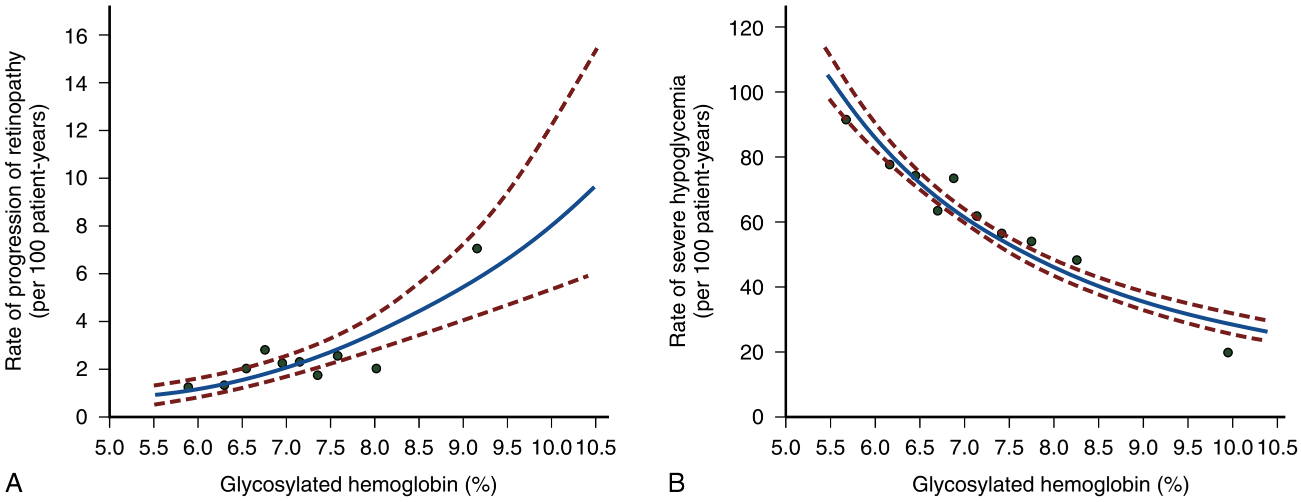


Fig. 39.4 **Risk for sustained progression of retinopathy (A) and rate of severe hypoglycemia (B) in patients receiving intensive therapy, in relation to mean glycosylated hemoglobin value during the Diabetes Control and Complications Trial.** (From the Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med.* 1993;329:984.)

further underscore the importance of initiating intensive therapy as early as safely possible in the management of type 1 diabetes.

Overall, the DCCT helped to clarify the relationship between A_{1c} and diabetic complications. For every 10% reduction in A_{1c}, the risk for developing retinopathy was reduced by approximately 45%.[15] Moreover, a continuous curvilinear relationship was observed between A_{1c} and the incidence of retinopathy, the risk for progressive retinopathy being decreased substantially at A_{1c} levels below 7%. Although the risk for progressive retinopathy was further reduced at A_{1c} concentrations below 7% (i.e., no threshold effect), this benefit was achieved at the cost of a significantly increased risk for hypoglycemia, given an observed continuous and inverse relationship between hypoglycemic risk and A_{1c} (Fig. 39.4). However, the DCCT was performed before the advent of some of the newer analog insulins, continuous glucose monitoring (CGM), or hybrid closed loop (HCL) systems, which are associated with lower rates of hypoglycemia than were seen in the DCCT.

Weight gain was an additional undesired impact of intensive therapy in a subset of individuals.[16] Thus, the DCCT firmly established the importance of glycemic control in the management of type 1 diabetes, demonstrated the feasibility of intensive therapy in improving glycemic control, and identified the limitations imposed on such control by hypoglycemia and weight gain. In doing so, this study has provided a framework for current treatment goals.

KEY POINTS

- Intensive insulin therapy can reduce the risk for microvascular and macrovascular complications in type 1 diabetes.

Glycemic Goals

On the basis of the findings of the DCCT/EDIC and other studies, the ultimate objective in the treatment of type 1 diabetes is metabolic normalization, that is, the achievement of sustained glycemic control as close to normal as possible and the avoidance of hypoglycemia. The glycemic target should be individualized based on the duration of diabetes, age/life expectancy, comorbid conditions, known CVD or advanced microvascular complications, hypoglycemia unawareness,

TABLE 39.1 American Diabetes Association Recommendations for Glycemia Treatment Targets in Nonpregnant Adults With Diabetes

Glycemic Parameter	Target
A_{1c}	<7.0%
Preprandial plasma glucose	70–130 mg/dL
Postprandial plasma glucose	<180 mg/dL
Time in range	>70%
GMI	<7%
CV	<36%

Goals are individualized for a given patient.
A_{1c}, Glycosylated hemoglobin; *GMI*, glucose management index; *CV*, coefficient of variation.
(Adapted from American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care*. 2021;44(suppl 1):S73–S84.)

and individual patient considerations and may change over time. Goals should be achieved in conjunction with an understanding of the person's psychosocial needs and a reduction in diabetes distress if elevated. In this context, the 2021 American Diabetes Association (ADA) treatment guidelines suggest the following therapeutic targets for nonpregnant adults with type 1 or type 2 diabetes: (1) A_{1c} less than 7.0% (referenced to nondiabetic range 4.0%–6.0%); (2) preprandial plasma glucose 70 to 130 mg/dL; and (3) peak postprandial plasma glucose less than 180 mg/dL (Table 39.1).[17] Achievement of lower A_{1c} levels than the goal of 7% (53 mmol/mol) may be acceptable, and even beneficial, following discussion between the person with diabetes and his or her healthcare professional, if it can be achieved safely without significant hypoglycemia or other adverse effects of treatment. Less stringent A_{1c} goals (such as <8% [64 mmol/mol]) may be appropriate for individuals with limited life expectancy, or where the harms of treatment are greater than the benefits.[18]

An alternative to the measurement of A_{1c} is an assessment of the Glucose Management Index (GMI) and Time in Range (TIR) from CGM data. In this setting, as well as the GMI goal of 7% (53 mmol/mol), a parallel goal would be a TIR of greater than 70% with time below range less than 4%. Fig. 39.5 shows the ambulatory glucose profile, which is the preferred method for evaluating data from continuous glucose monitors. In addition to TIR and GMI, it also provides an assessment of time above and below ranges, as well as glycemic variability.[19]

KEY POINTS

- The ultimate objective in the treatment of type 1 diabetes is the achievement of sustained glycemic control as close to normal as possible and the avoidance of hypoglycemia.

ACUTE COMPLICATIONS

The major acute complications of type 1 diabetes are ketoacidosis and hypoglycemia. These topics are discussed in detail in Chapters 37 and 38, respectively. The risk for these acute complications can be reduced through the judicious use of insulin therapy. Specifically, the risk for ketoacidosis can be reduced through appreciation of the concept that patients with type 1 diabetes require insulin therapy at all times. Appropriate management of special situations, such as intercurrent illness, requires adherence to this principle (see the section entitled "Adjustment of Insulin Therapy in Special Situations").

As was observed in the DCCT (see Fig. 39.2B), tight glycemic control typically is associated with an increased risk for hypoglycemia, although this may be mitigated through use of current technology. Regardless, patients therefore must be educated about the recognition of hypoglycemic symptoms and the use of appropriate treatment measures when such symptoms arise. Patients are advised to keep rapidly absorbable sources of carbohydrate, such as glucose tablets, close at hand at all times. In general, 15 g of carbohydrate is needed to increase blood glucose levels by approximately 38 mg/dL within 20 minutes.[20,21] Besides glucose tablets, other treatments that provide a similar amount of glucose include (1) 15 mL (1 tablespoon) or three packets of table sugar dissolved in water, (2) 175 mL (4 ounces) of juice, and (3) 15 mL (1 tablespoon) of honey. In addition, patients with type 1 diabetes should have an emergency glucagon kit that can be administered intranasally, subcutaneously, or intramuscularly. Both the patient and his or her family members should be trained in the use of glucagon as a means of rapidly increasing blood glucose concentration if the patient is severely hypoglycemic and unable to safely consume oral carbohydrate.

CHRONIC COMPLICATIONS

The chronic microvascular complications of type 1 diabetes include retinopathy, neuropathy, and nephropathy (discussed in Chapters 43, 44, and 45, respectively). To reduce the impact of these complications, management strategies include regular surveillance for early detection of complications and intensive diabetes management for both primary prevention and secondary intervention, as demonstrated in the DCCT.

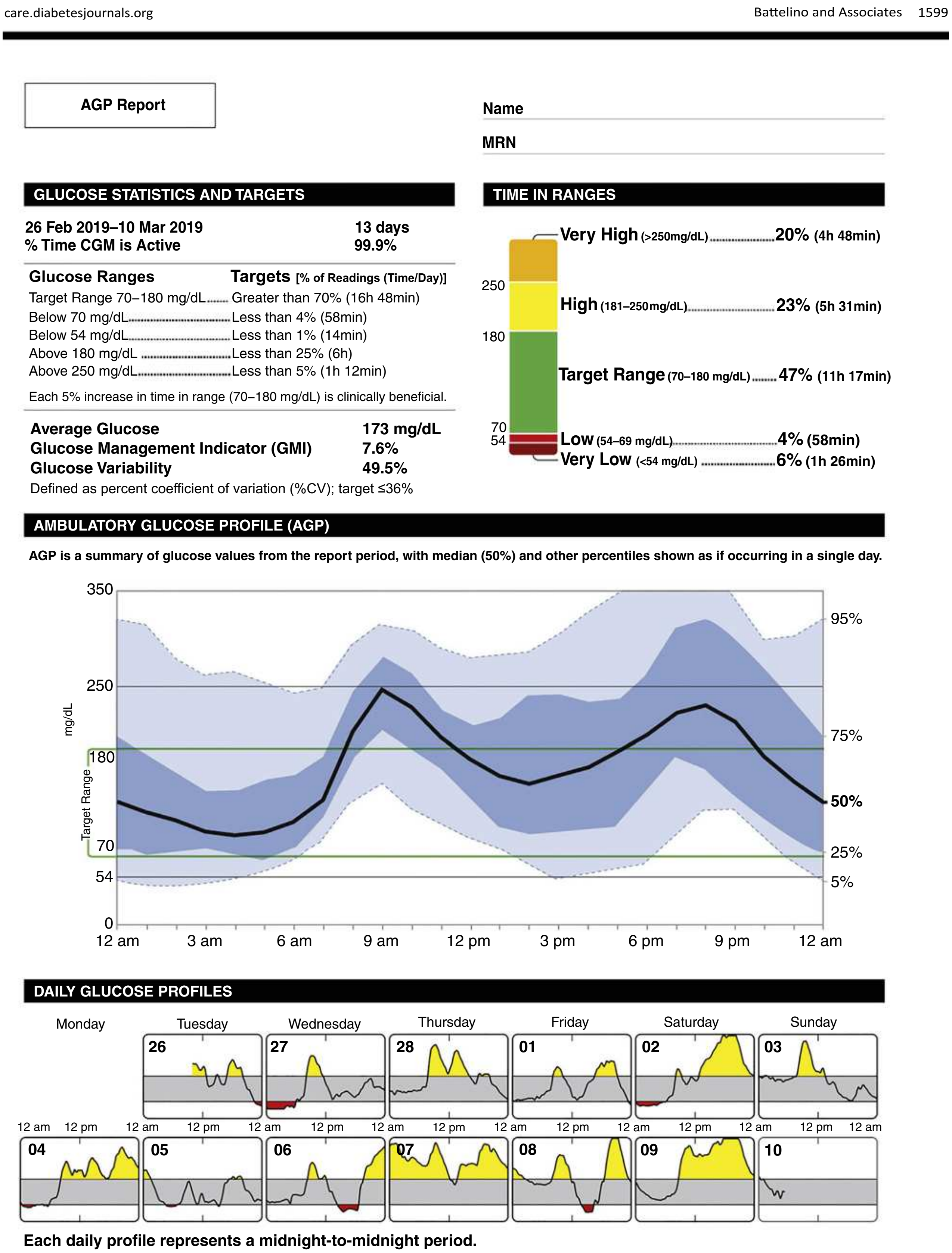
Macrovascular disease, as manifested by cerebrovascular, peripheral vascular, and coronary artery disease, is a major chronic complication of type 1 diabetes (discussed in Chapter 46). Regular surveillance for vascular disease and management of cardiovascular risk factors is essential. In particular, modification of reversible cardiovascular risk factors, such as hypertension, dyslipidemia, and cigarette smoking, is an important management goal. The 2021 ADA treatment guidelines suggest a target blood pressure of less than 140/90 mm Hg in patients with diabetes and less than 130/80 mm Hg in certain higher-risk individuals. In addition, the following treatment targets for lipids are recommended: LDL less than 100 mg/dL (or lower if known CVD), triglycerides less than 150 mg/dL, and high-density lipoprotein (HDL) greater than 40 mg/dL in men and HDL greater than 50 mg/dL in women.[17]

QUALITY OF LIFE

Type 1 diabetes is a chronic condition that requires a complex healthcare strategy that demands significant patient effort through frequent insulin injections or insulin pump therapy, regular self-monitoring of blood glucose (SMBG) or CGM, and constant attention to nutrition and physical activity. As such, effective management requires that the patient take responsibility for his or her care. In this context, it should be noted that psychosocial factors as well as social determinants of health all play a large role in terms of a patient's ability to manage his or her diabetes.[22] Recognition of these concepts at all times by the healthcare team will facilitate the success of treatment initiatives, as well as the need to involve mental health providers in patient care. Additionally, the concept of diabetes distress[23] is one that healthcare providers should be aware of, and reducing distress should be part of any treatment plan.

TEAM APPROACH TO MANAGEMENT

Given the complexity of modern treatment regimens, comprehensive care for patients with type 1 diabetes is best accomplished through a team approach. The core members of the diabetes healthcare team include the patient, the primary physician, a diabetes nurse educator, and a registered dietitian. Other individuals participating in the



care.diabetesjournals.org Battelino and Associates 1599

AGP Report

Name

MRN

GLUCOSE STATISTICS AND TARGETS

26 Feb 2019–10 Mar 2019 **13 days**
% Time CGM is Active **99.9%**

Glucose Ranges	**Targets** [% of Readings (Time/Day)]
Target Range 70–180 mg/dL	Greater than 70% (16h 48min)
Below 70 mg/dL	Less than 4% (58min)
Below 54 mg/dL	Less than 1% (14min)
Above 180 mg/dL	Less than 25% (6h)
Above 250 mg/dL	Less than 5% (1h 12min)

Each 5% increase in time in range (70–180 mg/dL) is clinically beneficial.

Average Glucose	**173 mg/dL**
Glucose Management Indicator (GMI)	**7.6%**
Glucose Variability	**49.5%**

Defined as percent coefficient of variation (%CV); target ≤36%

TIME IN RANGES

Very High (>250mg/dL)	**20%** (4h 48min)
High (181–250mg/dL)	**23%** (5h 31min)
Target Range (70–180 mg/dL)	**47%** (11h 17min)
Low (54–69 mg/dL)	**4%** (58min)
Very Low (<54 mg/dL)	**6%** (1h 26min)

250
180
70
54

AMBULATORY GLUCOSE PROFILE (AGP)

AGP is a summary of glucose values from the report period, with median (50%) and other percentiles shown as if occurring in a single day.

350
250
180
70
54
0
mg/dL
Target Range
95%
75%
50%
25%
5%
12 am
3 am
6 am
9 am
12 pm
3 pm
6 pm
9 pm
12 am

DAILY GLUCOSE PROFILES

Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
26
27
28
01
02
03
12 am
12 pm
04
05
06
07
08
09
10

Each daily profile represents a midnight-to-midnight period.

Fig. 39.5 Sample of an ambulatory glucose profile from continuous glucose monitoring system. From Battelino, T et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the International Consensus on Time in Range. *Diabetes Care* 2019;42:1593–1603.

patient's care may include a pharmacist, a medical social worker, a mental health specialist, a foot care specialist, an ophthalmologist, and other medical specialists such as cardiologists, nephrologists, and neurologists. At all times, the central role of the patient as a member of the healthcare team must be recognized. Accordingly, the efforts of the healthcare team should always be pursued from a patient-centered perspective. Inherent in this approach is the central importance of patient education, as the patient's active participation in his or her care demands an understanding of the issues associated with the treatment plan. As will be discussed in later sections of this chapter, the patient needs to be equipped with the knowledge and skills required for activities such as SMBG (with fingersticks or a continuous glucose monitor), glycemic pattern management, estimation of the carbohydrate content of meals, insulin dose adjustment, and appropriate insulin administration.

MONITORING

Type 1 diabetes is a chronic condition involving complex treatment regimens and multiple management goals. As such, regular monitoring of patient status is an essential component of comprehensive care. Patient monitoring in type 1 diabetes addresses two fundamental issues. First of all, frequent acute assessment of metabolic status guides adjustment of the treatment regimen, which, by nature, must be dynamic and able to accommodate day-to-day physiologic variability. Second, regular surveillance of both important clinical outcomes and associated surrogate markers allows for the detection and treatment of chronic complications and attendant comorbidities.

Glycemic Control

Regular monitoring of glycemic status by both patients and healthcare providers is fundamental to diabetes care. Glycemic monitoring involves frequent, acute measurement of blood glucose levels by the patient and periodic evaluation of chronic glycemic control through laboratory measurement of markers of glucose concentration such as A_{1c}, as well as assessments of TIR and GMI for those on CGM.

Acute Measurement of Glycemic Control. SMBG by patients is an invaluable tool that has empowered patient-driven management of type 1 diabetes, although it is increasingly being supplanted by use of CGM. SMBG involves using a small lancet to draw a drop of capillary blood from the fingertip. The patient applies a small drop of blood (usually 3–5 μL) to a testing strip, from which a portable glucose meter can determine the glucose concentration of the blood by enzymatic means within 5 to 30 seconds. In this way, patients can immediately determine their capillary blood glucose concentration at any time.

Patients with type 1 diabetes should monitor blood glucose frequently. In the DCCT, insulin doses for patients in the intensive treatment arm were adjusted according to the results of SMBG performed at least four times a day.[24] In other studies there was a linear relationship between A_{1c} and frequency of SMBG up to 10 measurements per day.[25]

Given the value of blood glucose data, it is important to consider potential problems that are inherent in the process of self-monitoring. The patient's use of the meter and self-monitoring technique should be reviewed periodically in the clinic. There can be clinical variability in the accuracy of US Food and Drug Administration (FDA)-approved meters,[26,27] so patients need to use well-validated devices with unexpired test strips that are kept in the provided sealed bottle or foil wrapper. Glucose monitors utilize an enzymatic reaction linked to an electrochemical reaction, either glucose oxidase or glucose dehydrogenase,[28] although most now use the latter method. Glucose oxidase monitors are sensitive to the oxygen available and should only be used with capillary blood in patients with normal oxygen saturation. Higher oxygen tensions (i.e., arterial blood or oxygen therapy) may result in false low glucose readings, and low oxygen tensions (i.e., high altitude, hypoxia, or venous blood readings) may lead to false high glucose readings. Glucose dehydrogenase monitors are not sensitive to oxygen.

Interference with SMBG results can occur from a variety of substances. For glucose oxidase monitors, this includes uric acid, galactose, xylose, acetaminophen, L-dopa, and ascorbic acid. Glucose dehydrogenase monitors have had interference issues with icodextrin, which is used in peritoneal dialysis.[29]

Continuous Glucose Monitoring. Because blood glucose concentrations are dynamic and fluctuate constantly, it is readily apparent that the limited data provided by even frequent SMBG might not be truly representative of an individual's glucose excursions. In this context, CGM devices represent an important development. This technology utilizes the fact that glucose in interstitial fluid can reflect blood glucose levels.[30] CGM systems typically consist of a subcutaneous sensor that continuously measures the glucose concentration in interstitial fluid and sends a signal to a smart phone or a reader.[28] Using this input, the monitor determines a blood glucose concentration every 1 to 5 minutes that can be seen in real time by the patients and downloaded to a computer to provide a graphical retrospective display of the patient's glycemic profile over the period of time monitored.[31] There are two classes of CGM devices available: "professional" or blinded CGM, in which the devices are owned by the clinic and worn by the patient episodically, and "personal" CGM, which are devices owned by the patient and worn nearly continuously. Personal CGM devices are either real time (RT)-CGM, which provides patients with alarms and alerts on a continuous basis, or intermittently scanned (IS)-CGM, which requires that the sensor is scanned with the reader/smart phone in order to display the data. IS-CGM does have the capacity to provide an optional low and high alarm.

Randomized clinical trials have been primarily done with RT-CGM devices. Depending on the trial and primary endpoint, these studies have shown reductions in A_{1c} levels[32] and reduction in rates of hypoglycemia.[33,34] This is true for patients on insulin pumps[32] as well as on multiple daily injections (MDI) of insulin,[35,36] and has been shown in older patients as well as younger individuals.[37] IS-CGM results have been more mixed, with fewer clinical trials and reductions in A_{1c} and hypoglycemia in some, but not all, studies.[38,39] However, multiple observational studies have shown benefit, particularly if the device is used in conjunction with a structured education program and is scanned frequently.[40,41] These studies have not used the most recently available IS-CGM systems that have alarms for high and low blood glucose levels, a feature not found on the previously studied devices.

Overall, individuals with type 1 diabetes should have access to CGM devices. Some require calibration with SMBG, and others do not. They vary in terms of how long they last, and all can be uploaded to cloud based programs for analysis. Patient preference and coverage determinations often dictate which device is used. Based on the available literature, for individuals with type 1 diabetes and episodes of hypoglycemia, particularly with hypoglycemia unawareness, RT-CGM is preferred.

In addition, the technology exists wherein CGM systems have been combined with insulin pumps in order to optimize insulin delivery. These are called HCL systems. The sensor data is fed into an algorithm that controls the delivery of insulin by the pump. These systems have been shown to improve A_{1c} and TIR, with reductions in rates of hypoglycemia.[42]

Glucose levels, either with SMBG or CGM, should be assessed before administration of insulin prior to a meal and at bedtime. Patterns of nocturnal hypoglycemia resulting from excessive bedtime insulin can be detected by blood glucose measurement in the early hours of the morning or reviewing overnight CGM data. Similarly, postprandial values 2 hours after a meal can assess the adequacy of the preprandial insulin dose. Patients should look at their blood glucose level before and during driving a motor vehicle, as well as before, during, and after exercise.

Monitoring of glucose levels is valuable only to the extent that the information it provides is used to guide treatment, and should be coupled with diabetes education. Glucose data can be used in real time, to determine insulin doses as well as to assess for hypoglycemia. The data can also be evaluated retrospectively for assessment of patterns and adjustments in dosing and food intake.

KEY POINTS

- Acute monitoring of glucose levels by either capillary testing, or now CGM systems, is an essential element of self-management of type 1 diabetes.

Measurement of Chronic Glycemic Control. Glucose can attach to proteins in the blood through an irreversible, nonenzymatic process, resulting in the formation of glycated proteins.[43] Because this process is irreversible, glucose remains attached to the protein until the latter is metabolized. The degree of glycation is a function of integrated blood glucose concentration over time. Thus, measurement of glycated proteins provides a means of estimating chronic blood glucose concentration over a time period proportional to the half-life of the protein in question.

Erythrocytes are freely permeable to glucose. In the circulation, glucose attaches irreversibly to hemoglobin in red blood cells, leading to the formation of glycated hemoglobins called hemoglobin A_{1a}, A_{1b}, and A_{1c}. Measurement of hemoglobin A_{1c} (also known as *A_{1c}, glycohemoglobin,* or *glycosylated hemoglobin*) has emerged as the most widely used test of chronic glycemic control. Specifically, the proportion of glycosylated hemoglobin to the total number of hemoglobin molecules reflects overall blood glucose concentration over the preceding 2 to 3 months.[44] Therefore, regular measurement of A_{1c} every 3 months is recommended as a method of determining overall glycemic control and evaluating the adequacy of the treatment regimen. For instance, a significantly elevated A_{1c} value, consistent with suboptimal glycemic control, would suggest the need for a change in the treatment regimen. Analysis of daily glycemic patterns from the patient's glucose records can indicate the specific changes to be made.

Many different types of assays can be used for measurement of glycated hemoglobins.[45] Some assays report A_{1c} as a percentage of total hemoglobin, while other methods measure total glycated hemoglobin. Thus, nonstandardization of assays can complicate comparison of A_{1c} measurements between laboratories. The National Glycohemoglobin Standardization Program, initiated in 1996, is an initiative that is designed to standardize laboratory A_{1c} assays to the DCCT reference method.[46] The International Federation of Clinical Chemistry developed reference methods for A_{1c} analysis, with A_{1c} reported in units of mmol/mol.[47]

The DCCT provided insight into the correlation between A_{1c} level and mean plasma glucose concentration.[48] This relationship has since been studied with the use of CGM data (Fig 39.6).[49]

In practice, in some situations the measured A_{1c} value may be discordant with the level that would be expected on the basis of the patient's SMBG or CGM results. This can be due to a number of factors. For instance, rapid red blood cell turnover leads to a disproportionate representation of younger erythrocytes, whose collective exposure to ambient blood glucose levels has been shorter than that of other red blood cells. Accordingly, A_{1c} levels tend to be reduced under such conditions.[50] Pregnancy and hemolytic anemias are examples of settings characterized by rapid red blood turnover, and hence lowered A_{1c} levels. Chronic kidney disease and anemia can cause alterations in

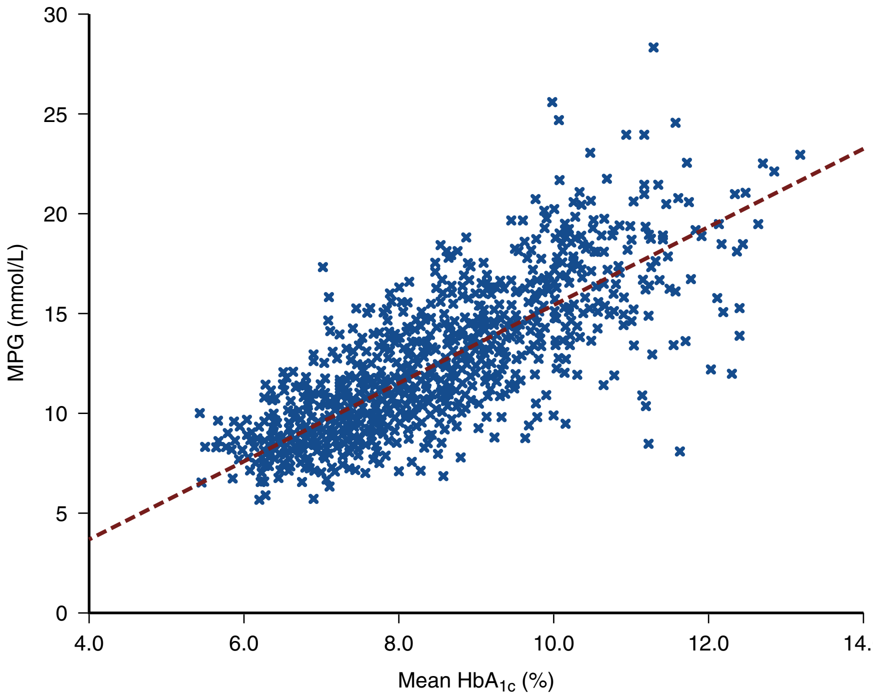


Fig. 39.6 The relationship between mean plasma glucose and mean glycosylated hemoglobin in the Diabetes Control and Complications Trial (n = 1429; r = 0.82). (From Rohlfing CL, Wiedmeyer HM, Little RR, et al. Defining the relationship between plasma glucose and HbA_{1c}: Analysis of glucose profiles and HbA_{1c} in the Diabetes Control and Complication Trial. *Diabetes Care.* 2002;25:276.)

red blood cell lifespan and impact A_{1c} levels.[51] Similarly, A_{1c} concentration may be lowered in hemoglobinopathies, such as thalassemia and sickle cell disease, because of a diminished propensity for glycation of the abnormal hemoglobin molecule. Conversely, depending on the assay used, A_{1c} values may be falsely elevated in the presence of hemoglobin F, the carbamylated hemoglobin formed in uremia, or acetaldehyde-bound hemoglobin (seen in alcohol abuse).[52] Finally, ethnicity can impact A_{1c} levels.[53]

In situations in which measurement of A_{1c} might be unreliable, an alternative approach to the estimation of glycemic control is the measurement of glycated serum proteins or albumin. The term *fructosamine* has been applied to the glycated proteins measured by such assays.[54] Given the 14- to 21-day half-life of albumin, a serum fructosamine measurement reflects mean blood glucose concentration over the preceding 1 to 2 weeks. Generally, serum fructosamine values correlate well with A_{1c} measurements.[55] However, certain caveats should be noted in measuring fructosamine. First of all, fructosamine values can be affected by conditions that alter the synthesis or clearance of serum proteins. Indeed, the question of whether or not to correct fructosamine assays for serum protein or albumin concentration remains controversial. Second, serum fructosamine levels show greater intrasubject variability than do A_{1c} measurements.[56] Third, unlike A_{1c}, fructosamine has not yet been shown to correlate with the risk for diabetic complications. Finally, use of CGM and calculation of the GMI can provide a reasonable estimation of overall glucose control over the past 2 to 4 weeks and may be more clinically relevant and accessible.

Ketone Testing

The presence of detectable levels of ketones in urine or blood may indicate impending or established ketoacidosis in patients with type 1 diabetes. Accordingly, patients should test for ketones in the following situations: 1) when blood glucose concentrations are persistently elevated, 2) during periods of acute illness or stress, or 3) when the patient is experiencing symptoms compatible with ketoacidosis, such as nausea, vomiting, and abdominal pain.[57] Traditionally, patients have used urine dipsticks to test for ketonuria. However, home monitors have been developed that can measure the capillary blood concentration of β-hydroxybutyrate, the most prevalent intermediate molecule in ketone formation.[58] Blood ketone testing avoids some of the limitations of traditional urinary ketone testing, such as false-positive readings, the inability to detect β-hydroxybutyrate (urinary dipsticks measure different ketone bodies), and patient reluctance to perform urinary testing. In studies to date, home capillary blood ketone measurement appears to be more sensitive than urinary testing in the detection of ketosis.[59,60] The use of blood ketone monitoring by patients is expected to increase in the future. In the clinical setting, serum ketone measurement is recommended over urinary testing.

Complication Surveillance

An essential component of ongoing care for patients with type 1 diabetes is surveillance for complications and comorbidities, as follows:

1. Retinopathy: Initial dilated and comprehensive eye examination should be performed at the time of diagnosis of type 1 diabetes and at regular intervals thereafter.[17] In patients without retinopathy who are well controlled, the interval may be every 2 years; for others receiving treatment, it could be monthly. Retinal photographs can suffice for screening if used appropriately.[61]
2. Nephropathy: Patients with type 1 diabetes should undergo an annual screening test for the presence of albuminuria.[17] The preferred screening test is a measurement of the albumin-to-creatinine ratio in a random, spot urine sample.[62] A serum creatinine and estimated glomerular filtration rate should also be obtained annually.
3. Neuropathy and foot care: A comprehensive foot examination should be performed annually.[17] This examination includes evaluation of sensation by 10-g Semmes–Weinstein monofilament, testing of vibration sense by tuning fork, assessment of superficial pain sensation, palpation, and visual inspection. In screening for neuropathy, superficial pain sensation testing, monofilament examination, and vibration testing all show similar operating characteristics and may be used interchangeably.[63] In patients with high-risk feet, an exam should be done at each visit, and patients should be encouraged to perform home foot exams daily.
4. Cardiovascular disease: Regular evaluation of cardiovascular risk factors is recommended. Blood pressure should be measured at every diabetic clinic appointment. Lipid profile should be tested annually. Screening exercise stress testing may be considered in patients with typical or atypical cardiac symptoms; abnormal resting electrocardiogram; a history of peripheral or cerebrovascular disease; or sedentary lifestyle, age over 35 years, and plans to start a vigorous exercise program. Weight should be measured at office visits, and weight loss discussed with patients who are obese or overweight.
5. Other: Other potential associated problems that warrant periodic screening include erectile dysfunction, depression, diabetes distress, celiac, and thyroid disease. Cognitive function screening may be needed in seniors.

KEY POINTS

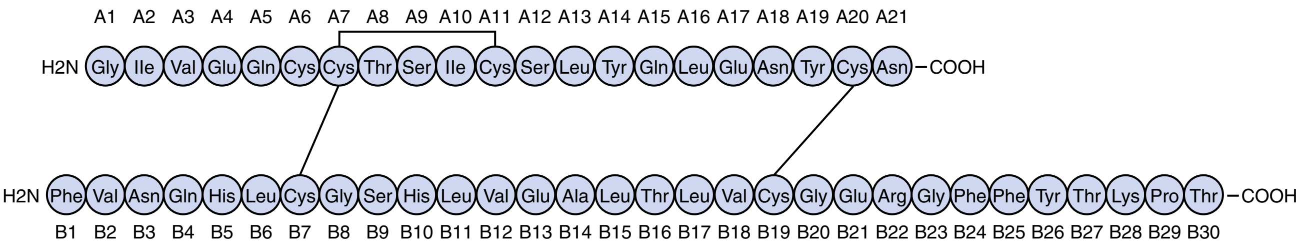
- Measurement of hemoglobin A_{1c} for monitoring chronic glycemic control and surveillance for complications of diabetes are standard elements in the clinical management of type 1 diabetes.

INSULIN THERAPY

Principles of Insulin Replacement

Insulin replacement strategies in type 1 diabetes ideally aim to mimic the normal physiologic secretion of insulin by the pancreas.[6] Therefore, to appreciate the rationale underlying the design of current insulin replacement regimens, an understanding of the basic structure and biochemistry of endogenous insulin is necessary (Fig. 39.7). This topic, discussed in detail in Chapter 29, is reviewed briefly here.

The pancreatic β cells secrete endogenous insulin into the portal venous system in response to physiologic demand for glucose homeostasis, as determined by nutrient intake and energy expenditure. Despite broad fluctuations in these determinants, healthy individuals are able to maintain plasma glucose concentration within a narrow range of 3.5 to 7.0 mmol/L throughout the day (Fig. 39.8).[64] This normal glucose homeostasis requires tightly regulated insulin secretion that consists of two components: a basal secretion rate and surges of markedly increased secretion following ingestion of a mixed meal. Accounting for approximately 40% of total 24-hour pancreatic insulin output, the basal secretion of approximately 1 μ/hour serves to limit hepatic glucose production and adipocyte lipolysis in the postabsorptive (fasting) state, such as between meals and overnight.[65] Conversely, with ingestion of a mixed meal, dietary secretagogues (e.g., glucose, amino acids) and gastrointestinal hormones (e.g., glucagon-like peptide 1) stimulate robust pulses of insulin secretion up to five times the basal rate. These pulses regulate postprandial glycemia by inhibiting hepatic glucose production and increasing peripheral glucose uptake. Insulin secretion is also affected by energy expenditure. For instance, with moderate exercise, insulin secretion rapidly decreases to prevent hypoglycemia.[65] Conversely, with strenuous exercise, catecholamine-mediated hyperglycemia may occur, leading to increased insulin secretion in the postexercise period.[66] Thus, the tight coupling of insulin secretion to plasma glucose concentration in healthy individuals (see Fig. 39.8) requires the complex integration of multiple signals of nutrient availability and energy expenditure.


A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21
H2N Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn COOH
H2N Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr COOH
B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 B13 B14 B15 B16 B17 B18 B19 B20 B21 B22 B23 B24 B25 B26 B27 B28 B29 B30

Rapid-acting analogues

- Lispro (Lys B28, Pro B29)
- Aspart (Asp B28)
- Glulisine (Lys B3, Glu B29)

Long-acting analogues

- Glargine (Gly A21, Arg B31, Arg B32)
- Detemir (Lys-tetradecanoyl B29, des-B30)

Fig. 39.7 The amino acid sequence and structure of human insulin. The A and B chains are linked by two disulfide bonds. A third disulfide bridge is present within the A chain. (Adapted from Cheng AY, Zinman B: Insulin analogues and the treatment of diabetes. In: Raz I, Skyler JS, Shafrir E, eds. *Diabetes: from research to diagnosis and treatment.* London: Martin Dunitz Ltd; 2002:331–346.)

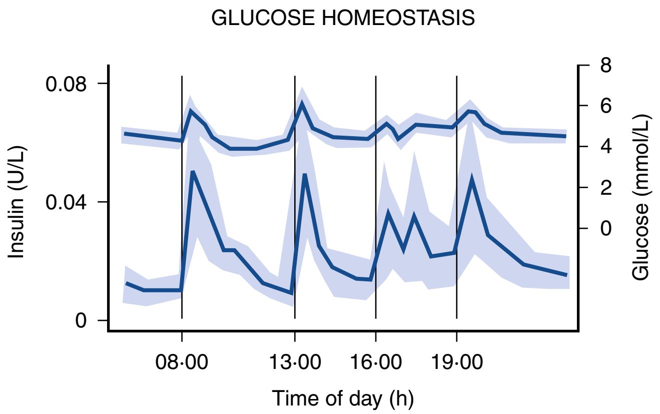


Fig. 39.8 Twenty-four–hour plasma glucose and insulin profiles in healthy individuals (n = 12) (shown as mean values with 95% confidence interval). (From Owens DR, Zinman B, Bolli GB. Insulins today and beyond. *Lancet.* 2001;358:739.)

In providing insulin replacement for patients with type 1 diabetes, the task at hand is to recreate this complex physiology. Insulin replacement strategies in type 1 diabetes aim to faithfully imitate both basal secretion and the appropriately integrated prandial surges that are seen in nondiabetic individuals. Generally, one component of the treatment regimen serves as basal insulin (i.e., simulates basal insulin secretion by the pancreas between meals, overnight, and in the fasting state), while a second component is considered meal insulin (i.e., simulates the normal prandial insulin surge). In this context, important factors to consider in the design of the treatment regimen include the types of insulin to be used for basal and meal coverage and the method of delivery.

KEY POINTS

- The goal of insulin replacement therapy in type 1 diabetes is to imitate both basal secretion and the appropriately integrated prandial surges that characterize insulin secretion by the pancreas in individuals without diabetes.

Insulin Preparations

For the first 60 years after the introduction of insulin therapy in 1922, the only commercially available preparations were animal insulins derived from bovine or porcine pancreatic extracts. Porcine and bovine insulin preparations differ from human insulin by one and three amino acids, respectively. The use of animal insulins, however, was inherently complicated by several issues, including potential supply limitations, incomplete purification, and a propensity to induce the formation of antiinsulin antibodies (which may affect the activity and absorption of the exogenous animal insulin and lead to unpredictable pharmacodynamics, although this relationship is controversial). These issues were reconciled with the introduction of recombinant human insulin in the 1980s.[67] The past three decades have seen further advances, with the introduction of insulin analogs bearing molecular modifications that confer advantageous pharmacokinetics. Recombinant human insulin and the newer analogs are now the main preparations used in the treatment of type 1 diabetes.

In selecting types of insulin for use in a treatment regimen, the most important parameters to consider are the onset, peak, and duration of action of the given insulin preparation (Table 39.2). Commercially available human insulins and analogs include ultrarapid-, rapid-, short-, intermediate-, and long-acting preparations. The pharmacokinetic properties will determine whether a preparation should be used as either meal insulin or basal insulin. Specifically, ultrarapid-, rapid-, and short-acting preparations are used as meal (or bolus) insulins, and intermediate- and long-acting preparations provide basal insulins.

Human Insulins. Human insulin was the first medication to be commercially manufactured using recombinant DNA technology. Currently available human insulins include short-, intermediate-, and long-acting preparations. Each preparation has advantages and disadvantages, which will be considered in turn.

Short-Acting Human Insulin: Regular. Regular recombinant human insulin is identical to the endogenous insulin polypeptide. In solution, however, regular insulin tends to self-associate, first forming dimers and subsequently hexamers. After subcutaneous injection, the absorption of hexameric insulin molecules is delayed, pending dissociation into monomers and dimers, which can diffuse rapidly from subcutaneous tissue into the systemic circulation.[68] This slowed absorption has important pharmacokinetic implications, leading to 1) a modestly delayed onset of biologic action *in vivo* (30–60 minutes after injection), 2) a relatively late peak effect (2–4 hours after injection), and 3) a prolonged duration of action (6–8 hours) (see Table 39.2).[69] Clearly, this pharmacokinetic profile is quite different from the rapid and short-lived endogenous insulin response to a mixed meal that is seen in individuals without diabetes. Therefore, although regular insulin was widely used in the past, its role as meal insulin has declined since the introduction of rapid-acting analogs. However, in the United States, where the cost of insulin is a major issue, regular insulin may need to be used as the prandial insulin, because it is much less expensive than analog insulin.[70]

Intermediate-Acting Human Insulin: Neutral Protamine Hagedorn and Lente. The development of intermediate-acting

TABLE 39.2 **Approximate Pharmacokinetic Properties of Human Insulin and Insulin Analogs Following Subcutaneous Injection**

Insulin Preparation	Onset of Action	Peak of Action	Duration of Action
Meal Insulin			
Fast-acting aspart	~3–4 minutes	1 hour	3–5 hours
Ultrarapid lispro	~3–4 minutes	1 hour	3–5 hours
Lispro	10–15 minutes	1–1.5 hours	3–5 hours
Aspart	10–15 minutes	1–2 hours	3–5 hours
Glulisine	10–15 minutes	1–2 hours	3–5 hours
Regular	30–60 minutes	2–4 hours	6–8 hours
Basal Insulin			
Neutral protamine Hagedorn	2.5–3 hours	5–7 hours	13–16 hours
Lente	2.5–3 hours	7–12 hours	Up to 18 hours
Detemir	2–3 hours	6–8 hours	≈24 hours
Glargine U100	2–3 hours	No peak	≈24 hours
Glargine U300	6 hours	No peak	24–36 hours
Degludec	2–4 hours	No peak	>42 hours

human insulins reflected the need for an exogenous insulin preparation that could mimic the basal insulin secretion that is observed in individuals who do not have diabetes. The principle underlying the development of these insulins is that the rate of absorption of exogenous insulin from subcutaneous tissue can be significantly reduced by manipulating its suspension. First introduced in 1946, neutral protamine Hagedorn (NPH) insulin is a suspension of insulin complexed with protamine and zinc.[71] NPH is poorly absorbed from subcutaneous tissue, leading to delayed onset of action (2.5–3 hours), late peak effect (5–7 hours), and prolonged duration of action (13–16 hours) (see Table 39.2). A second intermediate-acting insulin preparation is lente, a crystalline suspension of insulin with zinc and acetate. It exhibits a pharmacokinetic profile similar to that of NPH, but with a slightly later peak and longer duration of action.

The intermediate-acting insulins show substantial variation in subcutaneous absorption, both within an individual patient and between patients, resulting in variable glycemic excursions.[72] Indeed, variability in the absorption of intermediate-acting insulin may account for as much as 80% of day-to-day variation in blood glucose concentrations.[73] Other factors that contribute to the unpredictability of NPH and lente are dose-dependent changes in pharmacokinetics and the practical variability associated with resuspension of these insulin preparations by the patient before subcutaneous injection.

Given their underlying pharmacokinetic profiles (duration of action of 13–18 hours) and clinical variability in action, the intermediate-acting insulins generally do not provide adequate basal insulin coverage. To increase their utility as basal insulins, these preparations are given twice daily in at least 20% of patients with type 1 diabetes.[74] Moreover, their peak effect is an undesirable quality in a basal insulin, insofar as it is clearly dissimilar to the relatively peakless profile of basal insulin secretion seen in an individual without diabetes.

Insulin Analogs. The pharmacokinetic shortcomings of human insulin preparations have provided a strong impetus for the development of insulin analogues. Through specific modifications of the insulin molecule (see Fig. 39.7), researchers have designed analogs with tailored pharmacokinetic profiles that allow for better meal or basal coverage. Currently available analogs include ultrarapid-, rapid-, and long-acting preparations.

Rapid-Acting Analogs: Lispro, Aspart, and Glulisine. Because the tendency to self-associate and the resultant delayed absorption of regular insulin compromise its effectiveness in mimicking the normal prandial insulin surge, it follows that an analog with enhanced absorption may provide a superior meal insulin. In this context, it was previously noted with interest that human insulin-like growth factor-1 (IGF-1), despite significant homology to insulin, exhibited a reduced tendency to form multimers. Identification of the salient structural differences between IGF-1 and insulin led to the development of lispro insulin, the first commercially available analog.[75] Lispro is identical to human insulin except for a reversal of the 28th and 29th amino acid residues (proline and lispro, respectively) of the normal insulin B chain (see Fig. 39.7). The resultant conformational change in the C-terminus of the B chain introduces steric hindrance at interfaces involved in dimerization. Thus, dimer formation with lispro is reduced by a factor of 300 compared with regular human insulin.[76] It is important to note that, apart from reduced self-aggregation, lispro displays biologic activity similar to that of regular insulin, with both comparable affinity for the insulin receptor and equivalent hypoglycemic potency.[75] Thus, clinically, lispro acts similarly to monomeric human insulin. It is rapidly absorbed in 10 to 15 minutes, reaches peak activity in 60 to 90 minutes, and has a duration of action of 3 to 5 hours (see Table 39.2).[77] With this pharmacokinetic profile, lispro mimics the normal prandial insulin surge in response to carbohydrate ingestion better than any of the exogenous meal insulins that preceded it do.

A second rapid-acting analog is insulin aspart, engineered through replacement of the proline residue at position 28 of the B chain with aspartic acid. The negative charge of the aspartic acid residue causes repulsion from other negatively charged amino acids and thus leads to decreased self-association of insulin aspart monomers.[78] Accordingly, insulin aspart has a similar pharmacokinetic profile to that of lispro and also is well suited for use as meal insulin.[79] Finally, a third, recently introduced, rapid-acting analog is insulin glulisine, in which the asparagine residue at position B3 of human insulin has been replaced by lysine and the lysine at position B29 has been replaced by glutamic acid.[80]

Given their pharmacokinetics, the rapid-acting insulin analogs have clear advantages over regular human insulin. Although patients must inject these analogs 15 or more minutes before meals, their use is associated with greater mealtime flexibility, easier application of carbohydrate estimation techniques (see the section entitled “Nutrition”), and improved quality of life.[81] Both lispro and aspart provide improved postprandial glycemic control with less postprandial and nocturnal hypoglycemia than occurs with regular insulin.[82-85] Despite these advantages, however, use of these rapid-acting analogs in intensive insulin regimens has not been uniformly associated with substantial improvement in A_{1c} compared with regular insulin.[81] Nevertheless, even in the absence of a consistent A_{1c} benefit, the combination of improved postprandial glycemia with less hypoglycemia suggests an overall reduction in glycemic excursion, and hence better glycemic control, with rapid-acting analogs.[83]

Ultrarapid-Acting Insulin. The requirement that first-generation rapid-acting analogs be administered 15 or more minutes before eating can be inconvenient for patients in practice. This limitation has led to the development of ultrarapid-acting insulin analogs that are absorbed even more rapidly than their predecessors. These formulations include fast-acting insulin aspart (Fiasp) and ultrarapid lispro.

Fiasp is a formulation of insulin aspart that contains the incipients nicotinamide and L-arginine, which contribute to faster initial absorption. Compared with aspart, Fiasp demonstrates a more physiologic profile, with faster onset and greater early glucose-lowering effect.[86,87] In clinical trial comparison of Fiasp versus aspart as part of basal-bolus therapy in type 1 diabetes, these properties have translated into better postprandial glycemic control with Fiasp.[88,89]

Ultrarapid lispro is a formulation of lispro in which the locally active excipients citrate and the prostacyclin analog treprostinil have been added to enhance insulin absorption. Specifically, treprostinil increases local vasodilation, while citrate enhances vascular permeability at the injection site.[90] Accordingly, ultrarapid lispro demonstrates faster onset of action and greater early glucose-lowering compared with lispro.[90] As observed with Fiasp, these features have yielded better postprandial glycemic control when comparing ultrarapid lispro to lispro in basal-bolus therapy for type 1 diabetes.[91]

Long-Acting Analogs: Glargine (U100), Glargine (U300), Detemir, Degludec (U100, U200). An ideal basal insulin would have a peakless, 24-hour time-action profile.[92] Attempts to develop long-acting analogs have focused on methods of achieving slow, prolonged absorption following subcutaneous injection. Insulin glargine (U100), the first commercially available long-acting analog, was introduced to the US market in 2001. It has two modifications compared with human insulin: two arginines have been added to the C-terminus of the B chain, and a glycine residue replaces an acid-sensitive asparagine at position A21.[93] These changes have shifted the isoelectric point of glargine toward neutrality, such that this analog is completely soluble in its acidic injection solution at pH 4.0 but is much less soluble at the neutral physiologic pH of subcutaneous tissue. After injection into subcutaneous tissue, the acidic injection solution is neutralized, causing glargine to microprecipitate.[93] Glargine subsequently is slowly absorbed into the systemic circulation from the subcutaneous microprecipitates, resulting in a smooth and gradual rise in serum concentration. Therefore, the pharmacokinetic profile of glargine shows an onset of action 2 to 3 hours after injection with a relatively peakless, 24-hour duration of action (see Table 39.2).[94] Moreover, glargine is associated with less variability in absorption than is either NPH or lente.[95,96] Furthermore, its *in vivo* hypoglycemic potency is equivalent to that of human insulin. The promise associated with these properties has been further supported by clinical trial evidence demonstrating lower fasting plasma glucose levels and less hypoglycemia (including nocturnal hypoglycemia) with the use of glargine, compared with NPH, as basal insulin in intensive therapy.[97-99]

Glargine U300 is a three times more concentrated formulation of glargine that provides a long-acting, less variable basal insulin. Owing to its greater concentration and smaller volume, the microprecipitates that form following the subcutaneous injection of glargine U300 take longer to be absorbed.[100] The resultant slower release of glargine U300 leads to a longer duration of action, a flatter 24-hour profile, and less variability compared with glargine U100.[101,102] In head-to-head comparison in type 1 diabetes, glargine U300 provided similar glycemic control, with possible reduction in the risk of nocturnal hypoglycemia.[103,104]

The design of insulin detemir, a second long-acting analog, reflects a different approach to achieve long-acting basal coverage. With detemir, the B30 amino acid has been removed, and a 14-carbon aliphatic fatty acid has been acylated to the B29 amino acid.[105] This modification allows for reversible binding between albumin and the added fatty acid. After injection, an equilibrium develops between free and bound detemir, in which 98% of the analog is bound to albumin. Because only the free analog binds to the insulin receptor, the duration of action of detemir is prolonged, owing to sustained release of the analog that has been bound to circulating albumin.[106] The time-action profile of insulin detemir is characterized by peak activity at 6 to 8 hours after injection and a prolonged 24-hour duration of action (see Table 39.2).[106] Clinically, significant reductions in hypoglycemic risk have been demonstrated with detemir as basal insulin in intensive regimens compared with NPH.[107] In addition, detemir has been associated with more predictable glycemic control than NPH, with significantly less within-subject variation in blood glucose levels.[108,109] However, it is not as long-acting as the other long-acting insulins, and twice-daily administration is often required in people with type 1 diabetes.

Insulin degludec, available in U100 and U200 formulations, is a new-generation ultralong-acting basal insulin designed to provide a smooth, peakless pharmacokinetic profile and a prolonged duration of action of more than 40 hours.[110,111] Its structure maintains the amino acid sequence of human insulin, but with the deletion of threonine in position B30 and the addition of a fatty acid side chain attached to lysine in position B29 of the insulin B chain through a glutamic acid linker. In solution, degludec exists in a dihexameric state. Following subcutaneous injection, additives in the formulation disperse (primarily phenol) and degludec dihexamers self-associate to form large multihexamers. The resultant depot provides a slow, continuous release of degludec monomers that are absorbed into the circulation, yielding an ultralong duration of action. In clinical studies, degludec has provided effective glucose lowering, with low rates of hypoglycemia, particularly overnight.[112-114] In patients with type 1 diabetes on intensive insulin regimens, degludec has shown similar glycemic control as glargine U100 but with reduced risk for nocturnal hypoglycemia.[113,114] Overall, it is similar to glargine U300 in clinical practice.

KEY POINTS

- The pharmacokinetic shortcomings of human insulin preparations have led to the development of basal and meal insulin analogs that have been designed to more closely replicate the features of physiologic insulin secretion.

Inhaled Insulin. Despite their improved performance, modern subcutaneous insulin preparations have not been able to replicate the rapid on/off features of endogenous insulin secretion. In this context, pulmonary delivery of inhaled insulin has been proposed as an option of interest. Technosphere insulin is a dry powder formulation of regular human insulin adsorbed onto microparticles containing the excipient fumaryl diketopiperazine, which is highly soluble in water and at basic pH.[115,116] When these inhaled microparticles reach the lungs, they dissolve in the setting of physiologic pH and release the insulin for systemic absorption. The rapid absorption of insulin is further facilitated by the enormous surface area of well-perfused lungs. Inhaled insulin can thus provide meal insulin coverage. In patients with type 1 diabetes receiving basal insulin therapy, technosphere insulin achieved noninferiority in A_{1c} reduction compared with insulin aspart, coupled with less hypoglycemia and less weight gain but also an increased incidence of cough.[116] Monitoring of pulmonary function is recommended in patients receiving inhaled insulin.

Intensive Insulin Therapy Regimens

The DCCT clearly established that the use of intensive insulin therapy can achieve lower A_{1c} concentrations and can reduce the risk for microvascular complications in patients with type 1 diabetes. Therefore, intensive insulin therapy currently is recommended for all patients with type 1 diabetes. Two main regimens can be utilized to provide intensive insulin therapy: MDI and CSII. Both regimens attempt to mimic physiologic insulin secretion through appropriate meal and basal insulin replacement. Ultimately, the choice of insulin regimen generally is determined by patient-driven factors, including lifestyle issues, finances, and personal preference. As noted above, glycemic control is improved in individuals using CGM in addition to either MDI or CSII therapy.

Multiple Daily Injection Regimen. An MDI regimen involves at least four injections of insulin per day, consisting of boluses of meal insulin before each meal and at least one injection of basal insulin, usually at bedtime. If needed, a second injection of basal insulin may be added at breakfast or before lunch. The meal insulin of choice is an ultrarapid- or rapid-acting analog. Use of rapid-acting analogs in MDI therapy has been associated with better postprandial glycemic control, a reduced incidence of hypoglycemia, and improvements in quality of life, compared with regular insulin.[69] On the other hand, although some studies have shown a significant decrease in A_{1c} with the use of rapid-acting analogs as compared with regular insulin in MDI therapy, other studies have not shown a significant difference in overall glycemic control.[69] As noted earlier, however, this phenomenon likely reflects suboptimal basal insulin replacement in the studies in question. Moreover, with similar A_{1c}, the combination of reduced postprandial glycemic excursion with less hypoglycemia suggests reduced overall glycemic variability (and hence better control) with the use of rapid-acting analogs compared with regular insulin in MDI therapy.

In the past, basal insulin replacement in MDI usually has consisted of bedtime injection of NPH, lente, or ultralente. In older clinical studies of MDI regimens using rapid-acting analogs as meal insulin, all three of these basal preparations have shown similar efficacy with regard to glycemic control and incidence of hypoglycemia. Given their pharmacokinetic advantages, however, the long-acting analogs have become the basal insulins of choice in MDI therapy. In studies comparing glargine and NPH as basal insulins in MDI regimens, glargine (U100 and U300) and degludec have been associated with improved fasting glucose levels and reduced rates of hypoglycemia. A metaanalysis has shown that, compared with NPH, long-acting basal insulin analogs (glargine and detemir) have a small but significant effect on A_{1c}, coupled with reduced risk for nocturnal and severe hypoglycemia in patients with type 1 diabetes.[117] In addition, a second, prebreakfast injection of basal insulin is far less likely to be required with glargine than with NPH, lente, or ultralente. Indeed, in studies using lispro as meal insulin in MDI regimens, glycemic control was similar with a single injection of glargine at breakfast, dinner, or bedtime.[118] Similarly, degludec can be given at any time of day.[119] Studies evaluating the efficacy of MDI regimens that use rapid-acting analogs for meal insulin and long-acting analogs for basal coverage have been performed,[120] proving that this combination will provide the most physiologic insulin replacement among MDI regimens.

Continuous Subcutaneous Insulin Infusion. The external insulin infusion pump was first developed in the late 1970s, providing the basis for modern CSII therapy.[121] With CSII, an external infusion pump delivers a continuous infusion of rapid-acting insulin through a catheter inserted into the subcutaneous tissue of the abdominal wall. The pump is preprogrammed by the patient to deliver insulin continuously at a specified rate that is designed to meet the individual's basal insulin demands. For meal coverage, patients use the pump to deliver a specified bolus of insulin before eating. The rapid-acting or ultrarapid-acting insulin analogs are the insulins of choice in CSII, as both lispro and aspart have been shown to reduce postprandial glycemia, A_{1c} concentration, and the incidence of hypoglycemia compared with regular insulin in this setting.[122-124]

CSII offers advantages over MDI therapy in terms of convenience and flexibility.[125] Because the continuous infusion provides basal insulin at all times, the timing of meals with CSII, unlike with MDI, is completely flexible. At any given time, the rate of basal infusion can be adjusted immediately—an option that is not available with MDI. Moreover, current insulin pumps have multiphasic basal settings, allowing the user to set varying rates of basal insulin replacement at different times of the day depending on requirements. For instance, the pump can be preprogrammed to increase the basal rate of insulin infusion in the early morning hours in patients in whom the physiologic early morning secretion of growth hormone otherwise would lead to hyperglycemia (the "dawn phenomenon"). In addition, many insulin pumps have programming capabilities that facilitate dose selection for carbohydrate counting and insulin-correction boluses.

More recently, HCL technology has become available. These systems use a traditional insulin pump that is enabled to response to a control algorithm, which uses data from a CGM to adjust insulin delivery. These systems still require that the user enter his or her carbohydrates and give premeal insulin bolus doses, but they adjust the basal rates to avoid hypo- and hyperglycemia, and in some cases can provide "autocorrection" boluses.[126,127] Automating insulin delivery in this way reduces A_{1c} levels and rates of hypoglycemia and improves TIR.[126,127]

The major limitation with CSII is the significant cost associated with the pump and necessary supplies, such as the tubing, which must be changed every 48 to 72 hours. Another disadvantage is the risk for infection at the insertion site of the catheter. Catheter site infections occur at an estimated rate of 7.3 to 11.3 events per 100 years of patient follow-up.[128] These infections usually are readily treatable with antibiotics and a change of insertion site. In rare cases, a subcutaneous abscess requiring surgical drainage can develop. Finally, interruption of basal insulin delivery due to pump malfunction or catheter disruption can rapidly lead to hyperglycemia or even ketoacidosis, as patients will quickly become markedly insulinopenic owing to the use of only rapid- or short-acting insulin in CSII.[129]

Clinical Efficacy of Continuous Subcutaneous Insulin Infusion Versus Multiple Daily Injections. In clinical trials comparing optimized MDI therapy with CSII, glycemic control has been found to be similar with both regimens, especially when using insulin analogs.[125] Both are simply ways to give insulin and depend on user input and adjustments, especially for meals and corrections. However, both are improved by the addition of CGM, and combining CGM with an HCL system does provide additional benefit.[126,127]

Implementation of Intensive Insulin Therapy. Implementation of an intensive insulin regimen using MDI or CSII can be accomplished in many ways. In general, starting doses are estimated by using an algorithm such as the method shown in Fig. 39.9. The initial total daily insulin (TDI) requirement is estimated at 0.5 unit/kg of body weight, although most adult patients with type 1 diabetes ultimately require 0.6 to 0.7 unit/kg. Alternatively, for patients who are already on an insulin regimen, TDI is simply the sum of all current insulin doses. When CSII is implemented in such a patient, this TDI dose must be reduced by 20% to determine the total daily amount of insulin to be delivered by the pump.

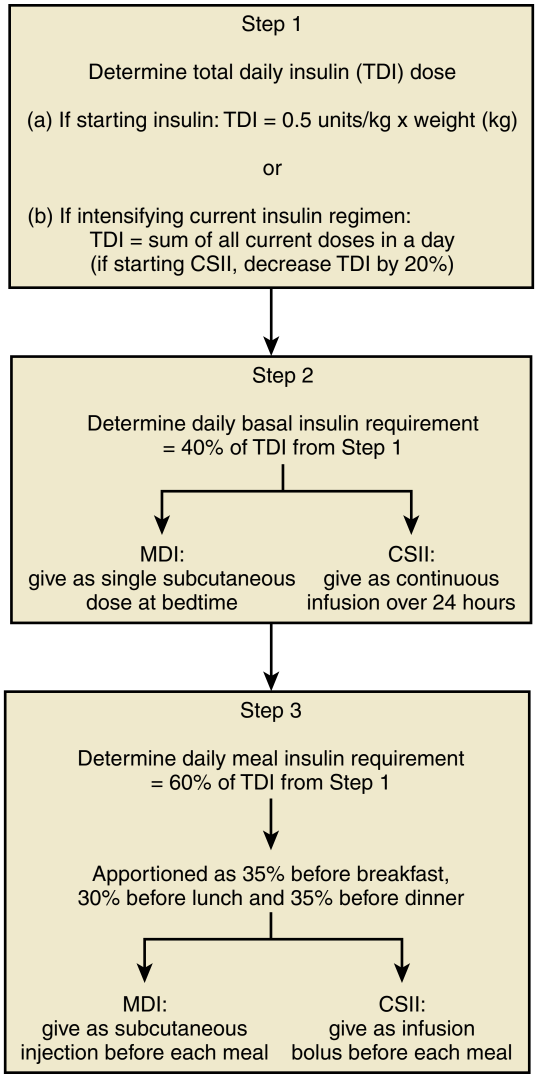

Step 1
Determine total daily insulin (TDI) dose
(a) If starting insulin: TDI = 0.5 units/kg x weight (kg)
or
(b) If intensifying current insulin regimen:
TDI = sum of all current doses in a day
(if starting CSII, decrease TDI by 20%)
Step 2
Determine daily basal insulin requirement
= 40% of TDI from Step 1
MDI:
give as single subcutaneous
dose at bedtime
CSII:
give as continuous
infusion over 24 hours
Step 3
Determine daily meal insulin requirement
= 60% of TDI from Step 1
Apportioned as 35% before breakfast,
30% before lunch and 35% before dinner
MDI:
give as subcutaneous
injection before each meal
CSII:
give as infusion
bolus before each meal

Fig. 39.9 One algorithm for the determination of initial insulin doses when intensive insulin therapy using multiple daily injection or continuous subcutaneous insulin infusion is implemented. (Adapted from Shah BR, Zinman B. Insulin regimens for type 1 diabetes. In: Sperling MA, ed. *Contemporary endocrinology: type 1 diabetes: etiology and treatment.* Totowa: Humana Press; 2003:205.)

Basal insulin makes up approximately 40% to 50% of the total daily dose, depending on the carbohydrate consumption of the patient. With MDI, this basal insulin initially is provided with a single bedtime injection, although postabsorptive daytime hyperglycemia may subsequently dictate the need for a second dose of basal insulin before breakfast. With CSII, the total daily basal insulin is provided as a continuous infusion over 24 hours, with hourly basal rates typically ranging from 0.6 to 1.2 units/hour. The remaining 50% to 60% of the TDI is provided in the form of meal or bolus insulin, with lunch generally requiring less insulin than breakfast and dinner.

As an example, consider the following case of a patient who is currently on a nonintensified regimen of 12 units of NPH at both breakfast and bedtime and six units of regular insulin at breakfast and dinner. To convert this patient's insulin therapy to an intensified MDI regimen, the algorithm in Fig. 39.9 would be applied as follows:

- The TDI would be 12 + 12 + 6 + 6 = 36 units of insulin per day.
- The basal requirement would be estimated at 40% × 36 units = 14 units per day.
- The daily meal insulin requirement would be estimated at 60% × 36 units = 22 units per day. This meal insulin could be divided such that the patient receives 35% × 22 units = 8 units at breakfast, 30% × 22 units = 6 units at lunch, and 35% × 22 units = 8 units at dinner. These meal insulin doses ideally would be provided in the form of preprandial subcutaneous injections of a rapid-acting analog.

TABLE 39.3 Example of a Variable Insulin Dose Scale

Blood Glucose (BG), mg/dL	UNITS OF MEAL INSULIN Breakfast	Lunch	Dinner
<90	–2	–2	–2
Goal BG 90–130	x	y	z
131–180	+2	+2	+2
181–230	+4	+4	+4
231–280	+6	+6	+6
281–330	+8	+8	+8
>330	+10	+10	+10

Sample variable insulin dose scale to guide adjustment of meal insulin based on preprandial glucose levels. Baseline doses of x, y, and z units of meal insulin at breakfast, lunch, and dinner, respectively, are adjusted by the indicated amounts based on the preprandial glucose level. (Example for illustrative purposes only.)
(Adapted from Leadership Sinai Centre for Diabetes, Mount Sinai Hospital, Toronto, Ontario, Canada.)

In practice, the insulin doses calculated in Fig. 39.9 are adjusted continually by the patient on the basis of the following three considerations: 1) preprandial glucose concentration, 2) anticipated carbohydrate intake, and 3) upcoming physical activity. By determining the preprandial blood glucose level using capillary monitoring, the patient can adjust prandial insulin dosage using a variable insulin dose scale (Table 39.3). Using this approach, the patient is directed to take a higher dose of prandial bolus insulin when the blood glucose concentration is above a specified range. The second consideration in dose adjustment is anticipated carbohydrate intake. By estimating the amount of carbohydrate to be ingested in the upcoming meal (a practice that is referred to as *carbohydrate counting*), the patient can adjust the prandial insulin dose accordingly. Although most patients require an additional 1 unit of insulin per 10 to 15 grams of carbohydrate ingested at a meal, the actual ratio must be determined individually for each meal in any given patient. Finally, patients must consider the effects of upcoming physical activity when adjusting insulin dosage. For instance, with MDI therapy, prandial insulin administered at the meal before moderate exercise might have to be reduced, depending on the individual's glycemic response to exercise.[129,130] Patients using CSII have the additional flexibility of being able to adjust their basal insulin rate to accommodate for exercise and further reduce their insulin on board.

KEY POINTS

- The recommended treatment for type 1 diabetes is intensive insulin therapy, which can be delivered through either multiple daily insulin injections (MDI) or CSII with an insulin pump.

Complications of Insulin Therapy. Insulin therapy is lifesaving in patients with type 1 diabetes. Nevertheless, complications of insulin treatment can pose obstacles to the implementation of intensive therapy.

The most feared and serious complication of intensive insulin therapy is severe hypoglycemia (see Chapter 38). In the DCCT, the risk for hypoglycemia was approximately three fold higher in the intensive treatment group than in the conventional therapy arm.[131] Because insulin analogs were not available at the time of the DCCT, it is not clear whether this finding reflects current practice, although more recent data suggest that glycemia variability, rather than A_{1c} level, is the biggest risk for episodes of hypoglycemia. It is surprising that, despite the high rates of hypoglycemia with intensive therapy in the DCCT, no evidence of detectable long-term decline in cognitive function was noted in study participants after a mean 18 years of follow-up in the EDIC study.[132] Nevertheless, hypoglycemia is an important factor that may limit the ability to achieve optimal glycemic control. Hypoglycemia generally reflects an imbalance among carbohydrate intake, physical activity, and the dose of exogenous insulin. In addition, patients with type 1 diabetes have a deficient counterregulatory response to hypoglycemia.[133] Previous episodes of severe hypoglycemia and physiologic inability to detect the attendant symptoms (hypoglycemia unawareness) are other factors that are associated with increased hypoglycemic risk. Thus, a patient's propensity for hypoglycemia must be considered in establishing therapeutic goals, as excessive hypoglycemia is best managed by raising glycemic targets. Similarly, in patients with hypoglycemia unawareness, raising glycemic targets is utilized as a means of restoring physiologic awareness of hypoglycemic symptoms.[134] Specifically, with increased overall glycemia, the careful avoidance of further hypoglycemia can help to restore awareness and improve defective counterregulation.

Weight gain is another adverse effect that is associated with insulin therapy. In the DCCT, patients receiving intensive therapy gained 4.75 kg more than participants in the conventional treatment arm over the course of the study.[135] This effect reflects both insulin's anabolic properties and the decrease in glycosuria observed with improved glycemic control. Specifically, for patients with poor glycemic control, the elimination of glycosuria with insulin therapy will lead to a positive caloric balance and weight gain. In addition, intermittent hypoglycemia associated with intensive insulin therapy can lead to hunger and increased caloric intake. Concerns regarding weight gain may cause some patients to be reluctant to initiate intensive therapy and should be addressed by the healthcare team.

Rapid improvement in glycemic control following the initiation of intensive insulin therapy can cause transient exacerbation of underlying retinopathy.[137] Patients with proliferative retinopathy and baseline A_{1c} greater than 10% are at highest risk for this complication.[137] Baseline ophthalmologic evaluation with frequent surveillance should be considered for such patients when intensive therapy is started.[136] In addition, more gradual reduction in A_{1c} may be prudent in such cases.

Allergic reactions to insulin are rare, particularly because human insulin is much less immunogenic than were earlier animal insulins. Most allergic reactions to insulin reflect local hypersensitivity at the site of injection. Generalized allergic reactions can occur rarely. Protocols for desensitization have been developed for both local and systemic hypersensitivity to insulin.[138]

Lipohypertrophy refers to localized swelling at a site of repeated insulin injection. This complication is related to the lipogenic effects of insulin. Continued injection of insulin into sites of lipohypertrophy is discouraged, as absorption of the insulin can be erratic, which may lead to unpredictable glycemic effects. Lipoatrophy reflects a different pathophysiology, in which an immune-mediated response to exogenous insulin leads to atrophy of subcutaneous tissue at the site of injection.[139] With the current use of recombinant human insulin preparations and analogues, the problem of insulin-induced lipoatrophy is infrequent.

Adjustment of Insulin Therapy in Special Situations

The absence of endogenous insulin secretion in patients with type 1 diabetes dictates an absolute reliance at all times on exogenous insulin for both prandial and basal metabolic needs. Thus, even when not eating, patients with type 1 diabetes continue to require insulin therapy to meet basal requirements and to avoid diabetic ketoacidosis. Recognition of this principle underlies the appropriate management of their insulin therapy in special situations such as at times of intercurrent illness and during surgery.

Intercurrent Illness. The physiologic stress associated with illness stimulates the release of counterregulatory hormones that antagonize insulin action. Thus, during times of illness, patients are at risk for both hyperglycemia and diabetic ketoacidosis. During such times, regular monitoring of blood glucose concentrations at least every 4 hours is recommended. If hyperglycemia is detected, then supplemental insulin is indicated. Conversely, if blood glucose levels are low, then the amount of insulin taken at that time should be reduced significantly. As noted earlier, patients must not withhold insulin therapy completely. Instead, under such circumstances, patients are advised to drink sugar-rich fluids such as fruit juices to maintain adequate carbohydrate and fluid intake. If patients are unable to tolerate even this limited intake, however, then presentation to the hospital is necessary for administration of intravenous fluids.

Even with continued insulin therapy, patients with type 1 diabetes are at risk for developing ketoacidosis during times of illness owing to the effects of counterregulatory hormones in antagonizing insulin action and promoting ketogenesis. Frequent testing of blood ketone levels using a home capillary blood ketone monitor is recommended at such times. Alternatively, if such testing is unavailable at home, then traditional urinary ketone monitoring should be performed (see the section entitled "Ketone Testing" earlier in the chapter).

Surgery. The goals of perioperative management of patients with type 1 diabetes are avoidance of marked hyperglycemia and avoidance of hypoglycemia. As with periods of illness, surgery is associated with both physiologic stress that promotes hyperglycemia and reduced caloric intake that may predispose to hypoglycemia (because most surgical procedures are performed on fasting patients). As before, the principles of management include the need for continued insulin therapy and close metabolic monitoring. Typically, major surgical procedures are managed with intravenous insulin therapy and frequent blood glucose monitoring.

NUTRITION

The nutritional recommendations associated with healthy eating habits for the general public are also applicable to patients with type 1 diabetes.[17] The important difference in patients with type 1 diabetes is that dietary intake must be coordinated appropriately with insulin therapy and physical activity to achieve regulation of blood glucose levels. In particular, attention must be paid to the sources and amount of carbohydrate in the diet. Carbohydrate is the main dietary constituent that affects postprandial glycemic excursion, and hence prandial insulin requirements.[140] This concept has led to the practice of "carbohydrate counting."

Carbohydrate counting is a meal-planning approach in which patients are taught to estimate the carbohydrate content of a meal preprandially and to adjust their dose of meal insulin accordingly.[141] Most patients require 1 extra unit of insulin for each 10 to 15 g of carbohydrate ingested, although this ratio must be determined individually on the basis of personal food preferences and individual physiologic response. As such, carbohydrate counting is an integral component of patient education, and it is a skill that patients generally develop with the assistance of a clinical dietitian. Although the concept of carbohydrate counting was described shortly after the discovery of insulin,[142] its recent resurgence is linked in part to its inclusion as one of the dietary interventions in the intensive treatment arm of the DCCT. In the DCCT, the practice of adjusting food and/or insulin dosage in response to hyperglycemia emerged as a dietary behavior associated with improved glycemic control.[143] Moreover, carbohydrate counting allows for significant flexibility in food choices. Alternatively, for patients on fixed insulin doses who are unwilling to adjust prandial insulin dosage on the basis of dietary content, consistency of carbohydrate intake is advised.

Although carbohydrate content is the key modifier of prandial insulin dosage, total energy intake from protein and fat cannot be neglected. Total energy intake must be regulated to avoid unnecessary weight gain, particularly because intensive insulin therapy is frequently associated with weight gain as well.

Another important consideration that affects food choices is hypoglycemia. For instance, patients must recognize that ingested alcohol may reduce hepatic glucose production and mask the symptoms of hypoglycemia. Indeed, moderate alcohol consumption (one to two standard drinks) 2 to 3 hours after the evening meal can lead to delayed hypoglycemia the next morning after breakfast.[144] Conversely, patients are advised to avoid the overtreatment of hypoglycemia by excessive food intake, as this practice may lead to weight gain and significant glycemic excursion.

OTHER THERAPIES FOR TYPE 1 DIABETES

Islet cell and pancreas transplantation are therapies, in certain cases, for type 1 diabetes. These treatment modalities are discussed in detail in Chapter 41. However, due to multiple constraints, these are rarely done in otherwise healthy individuals with type 1 diabetes.

Premeal subcutaneous injection of pramlintide, an analogue of the pancreatic β cell hormone amylin that is otherwise deficient in patients with type 1 diabetes, is an FDA-approved therapy for adjunctive treatment of type 1 diabetes. Amylin affects glucose homeostasis through multiple mechanisms, including slowing of gastric emptying, regulation of postprandial glucagon, and reduction in food intake.[145] Pramlintide can be used in the treatment of type 1 diabetes in conjunction with insulin therapy (although it should be noted that pramlintide and insulin cannot be mixed in the same syringe). Overall, its effect on glycemic control is modest, but pramlintide may help to limit weight gain and reduce meal insulin requirements.[146]

Metformin, DPP-IV Inhibitors, and GLP-1 RAs

Off-label use of medications approved for individuals with type 2 diabetes has shown varying results.[147,148] Metformin is commonly added in individuals with type 1 diabetes and metabolic syndrome and/or polycystic ovarian syndrome (PCOS). Although studies have not shown improvements in A_{1c},[149,150] there may be a slight reduction in CVD risk,[150,151] as well as improvement in PCOS.[152]

DPP-IV inhibitors have shown no benefit, but use of GLP-1 RAs has been associated with weight loss[153,154] and, in some studies, improvements in glycemic control. The weight loss is not surprising, given that these agents are increasingly being used as weight loss medications in people without diabetes. SGLT-2 inhibitors, which act independently of β cell function, have been shown to provide some benefit in people with type I diabetes.[155-158] However, they are associated with an increased risk for diabetic ketoacidosis, including euglycemic diabetic ketoacidosis.[159] that limits their use. They are approved in Europe for individuals with type 1 diabetes and a body mass index greater than or equal to 27[160] who can follow guidelines for ketone testing.

FUTURE PERSPECTIVES

Despite remarkable advances in the more than 90 years since the discovery of insulin, the goal of sustained euglycemia in patients with type 1 diabetes remains largely unrealized. Even with the best treatments available, many patients continue to have A_{1c} levels that are above target[161]. When treating patients, consideration needs to be given to their psychosocial needs as well as to social determinants of health. Many underserved patients fail to realize the benefits seen by their more affluent counterparts.[162] Nevertheless, many different approaches to dealing with type 1 diabetes are being explored, from early detection and prevention to β cell replacement techniques and, perhaps most robustly, to ever newer and better diabetes technologies. Although diabetes is not curable, our treatment options continue to grow and must be adopted by as many as possible in order for people with type 1 diabetes to safely achieve and maintain their glycemic targets.

REFERENCES

1. Banting FG, Best CH, Collip JB, et al. Pancreatic extracts in the treatment of diabetes mellitus. *Can Med Assoc J*. 1922;12:141–146.
2. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977–986.
3. Reichard P, Nilsson B-Y, Rosenqvist U. The effect of long-term intensified insulin treatment on the development of microvascular complications of diabetes mellitus. *N Engl J Med*. 1993;329:304–309.
4. Lauritzen T, Rost-Larsen K, Larsen H-W, et al. Two-year experience with continuous subcutaneous insulin infusion in relation to retinopathy and neuropathy. *Diabetes*. 1985;34:74–79.
5. Brinchmann-Hansen O, Dahl-Jorgensen K, Hanssen KF, et al. The response of diabetic retinopathy to 41 months of multiple insulin injectins, insulin pumps, and conventional insulin therapy. *Arch Ophthalmol*. 1988;106:1242–1246.
6. Foster NC, Beck RW, Miller KM, et al. State of type 1 diabetes management and outcomes from the T1D Exchange in 2016-2018. *Diabetes Technol Ther*. 2019;21:66–72. Erratum in: *Diabetes Technol Ther*. 2019;21:230.
7. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med*. 2000;342:381–389.
8. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy. *JAMA*. 2003;290:2159–2167.
9. DCCT/EDIC Research Group, de Boer IH, Sun W, Cleary PA, et al. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *N Engl J Med*. 2011;365:2366–2376.
10. de Boer IH, Kestenbaum B, Rue TC, et al. Diabetes Control and Complications Trial (Dcct)/Epidemiology of Diabetes Interventions And Complications (Edic) Study Research Group. Insulin therapy, hyperglycemia, and hypertension in type 1 diabetes mellitus. *Arch Intern Med*. 2008;168:1867–1873.
11. Nathan DM, Cleary PA, Backlund JY, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353:2643–2653.
12. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Intensive therapy and carotid intimamedia thickness in type 1 diabetes mellitus. *N Engl J Med*. 2003;348:2294–2303.
13. Cleary PA, Orchard TJ, Genuth S, et al. The effect of intensive glycemic treatment on coronary artery calcification in type 1 diabetic participants of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study. *Diabetes*. 2006;55:3556–3565.

14. Retnakaran R, Zinman B. Type 1 diabetes, hyperglycaemia, and the heart. *Lancet*. 2008;371:1790–1799.
15. The Diabetes Control and Complications Trial Research Group. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. *Diabetes*. 1995;44:968–983.
16. Purnell JQ, Zinman B, Brunzell JD. DCCT/EDIC Research Group. The effect of excess weight gain with intensive diabetes mellitus treatment on cardiovascular disease risk factors and atherosclerosis in type 1 diabetes mellitus: results from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study (DCCT/EDIC) study. *Circulation*. 2013;127:180–187.
17. American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care*. 2021;44(suppl 1):S1–S232.
18. Holt RIG, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, et al. The management of type 1 diabetes in adults: a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2021;44(11):2589–2625.
19. Battelino T, Danne T, Bergenstal RM, et al. et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: recommendations From the International Consensus on Time in Range. *Diabetes Care*. 2019;42(8):1593–1603.
20. Canadian diabetes association clinical practice guidelines expert committee: hypoglycemia. *Can J Diabetes*. 2018;42:S104–S115.
21. Slama G, Traynard P-Y, Desplanque N, et al. The search for an optimized treatment of hypoglycemia. Carbohydrates in tablets, solution, or gel for the correction of insulin reactions. *Arch Intern Med*. 1990;150:589–593.
22. Hill-Briggs F, Adler NE, Berkowitz SA, Chin MH, Gary-Webb TL, et al. Social determinants of health and diabetes: a scientific review. *Diabetes Care* 2021;44(1): 258–279.
23. van Duinkerken E, Snoek FJ, de Wit M. The cognitive and psychological effects of living with type 1 diabetes: a narrative review. *Diabet Med*. 2020;37:555–563.
24. The Diabetes Control and Complications Trial Research Group. Implementation of treatment protocols in the Diabetes Control and Complications Trial. *Diabetes Care*. 1995;18:361–376.
25. Miller KM, Beck RW, Bergenstal RM, et al. Evidence of a strong association between frequency of self-monitoring of blood glucose and hemoglobin A1c levels in T1D exchange clinic registry participants. *Diabetes Care*. 2013;36:2009–2014.
26. Klonoff DC, Parkes JL, Kovatchev BP, et al. Investigation of the accuracy of 18 marketed blood glucose monitors. *Diabetes Care*. 2018;41:1681–1688.
27. Freckmann G, Pleus S, Grady M, et al. Measures of accuracy for continuous glucose monitoring and blood glucose monitoring devices. *J Diabetes Sci Technol*. 2019;13:575–583.
28. Reynolds LR, Karounos DG. Emerging technology in diabetes mellitus: glucose monitoring and new insulins. *South Med J*. 2002;95:914–918.
29. Ginsberg BH. Factors affecting blood glucose monitoring: sources of errors in measurement. *J Diabetes Sci Technol*. 2009;3:903–1326.
30. Rebrin K, Steil GM, van Antwerp WP, et al. Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring. *Am J Physiol*. 1999;277:E561–E571.
31. Sherwood JS, Russell SJ, Putman MS. New and emerging technologies in type 1 diabetes. *Endocrinol Metab Clin North Am*. 2020;49:667–678.
32. Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group, Tamborlane WV, Beck RW, et al. Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med*. 2008;359:1464–1476.
33. Riddlesworth T, Price D, Cohen N, et al. Hypoglycemic event frequency and the effect of continuous glucose monitoring in adults with type 1 diabetes using multiple daily insulin injections. *Diabetes Ther*. 2017;8:947–951.
34. Heinemann L, Freckmann G, Ehrmann D, et al. Real-time continuous glucose monitoring in adults with type 1 diabetes and impaired hypoglycaemia awareness or severe hypoglycaemia treated with multiple daily insulin injections (HypoDE): a multicentre, randomised controlled trial. *Lancet*. 2018;391:1367–1377.
35. Beck RW, Riddlesworth T, Ruedy K, et al. Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: the DIAMOND randomized clinical trial. *JAMA*. 2017;24:371–378.
36. Lind M, Polonsky W, Hirsch IB, et al. Continuous glucose monitoring vs conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections: the GOLD randomized clinical trial. *JAMA*. 2017;24:379–387.
37. Pratley RE, Kanapka LG, Rickels MR, et al. Wireless Innovation for Seniors with Diabetes Mellitus (WISDM) Study Group. Effect of continuous glucose monitoring on hypoglycemia in older adults with type 1 diabetes: a randomized clinical trial. *JAMA*. 2020;323:2397–2406.
38. Cowart K, Updike W, Bullers K. Systematic review of randomized controlled trials evaluating glycemic efficacy and patient satisfaction of intermittent-scanned continuous glucose monitoring in patients with diabetes. *Diabetes Technol Ther*. 2020;22:337–345.
39. Bidonde J, Fagerlund BC, Frønsdal KB, et al. *FreeStyle Libre Flash Glucose Self-Monitoring System: A Single-Technology Assessment [Internet]*. Oslo, Norway: Knowledge Centre for the Health Services at The Norwegian Institute of Public Health (NIPH); 2017 Aug 21. Report from the Norwegian Institute of Public Health No. 2017-07. PMID: 29553668.
40. Paris I, Henry C, Pirard F, et al. The new FreeStyle libre flash glucose monitoring system improves the glycaemic control in a cohort of people with type 1 diabetes followed in real-life conditions over a period of one year. *Endocrinol Diabetes Metab*. 2018;1:e00023.
41. Leelarathna L, Wilmot EG. Flash forward: a review of flash glucose monitoring. *Diabet Med*. 2018;35:472–482.
42. Brown SA, Kovatchev BP, Raghinaru D, et al. Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes. *N Engl J Med*. 2019;31:1707–1717.
43. Bunn HF, Haney DN, Gabbay KH, et al. Further identification of the nature and linkage of the carbohydrate in hemoglobin A1c. *Biochem Biophys Res Commun*. 1975;67:103–109.
44. Nathan DM, Singer DE, Hurxthal K, et al. The clinical information value of the glycosylated hemoglobin assay. *N Engl J Med*. 1984;310:341–346.
45. Little RR, Wiedmeyer HM, England JD, et al. Interlaboratory comparison of glycohemoglobin results: college of American Pathologists survey data. *Clin Chem*. 1991;37:1725–1729.
46. Little RR, Rohlfing C, Sacks DB. The national glycohemoglobin standardization program: over 20 years of improving hemoglobin a1c measurement. *Clin Chem*. 2019;65:839–848.
47. Hanas R, John G. International HbA Consensus Committee. 2010 consensus statement on the worldwide standardization of the hemoglobin A1c measurement. *Diabet Med*. 2010;27:737–738.
48. Rohlfing CL, Wiedmeyer HM, Little RR, et al. Defining the relationship between plasma glucose and HbA1c: analysis of glucose profiles and HbA1c in the diabetes control and complication trial. *Diabetes Care*. 2002;25:275–278.
49. Nathan DM, Kuenen J, Borg R, et al. A1c-Derived Average Glucose Study Group. Translating the A1c assay into estimated average glucose values. *Diabetes Care*. 2008;31:1473–1478.
50. Panzer S, Kronik G, Lechner K, et al. Glycosylated hemoglobins (GHb): an index of red cell survival. *Blood*. 1982;59:1348–1350.
51. Paisey R, Banks R, Holton R, et al. Glycosylated haemoglobin in uremia. *Diabet Med*. 1986;3:445–448.
52. Rhea JM, Molinaro R. Pathology consultation on HbA(1c) methods and interferences. *Am J Clin Pathol*. 2014;141:5–16.
53. Cavagnolli G, Pimentel AL, Freitas PA, et al. Effect of ethnicity on HbA1c levels in individuals without diabetes: systematic review and meta-analysis. *PLoS One*. 2017;12:e0171315.
54. Armbruster DA. Fructosamine: structure, analysis and clinical usefulness. *Clin Chem*. 1987;33:2153–2163.
55. Baker JR, Metcalf PA, Holdaway IM, et al. Serum fructosamine concentration as a measure of blood glucose control in type 1 (insulin-dependent) diabetes mellitus. *Br Med J*. 1985;290:352–355.
56. Howey JE, Bennet WM, Browning MC, et al. Clinical utility of assays of glycosylated hemoglobin and serum fructosamine compared: use of data on biologic variation. *Diabet Med*. 1989;6:793–796.

57. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Monitoring glycemic control. *Can J Diabetes*. 2018;42:S47–S53.
58. Byrne HA, Tieszen KL, Hollis S, et al. Evaluation of an electrochemical sensor for measuring blood ketones. *Diabetes Care*. 2000;23:500–503.
59. Guerci B, Benichout M, Floriot M, et al. Accuracy of an electrochemical sensor for measuring capillary blood ketones by fingerstick samples during metabolic deterioration after continuous subcutaneous insulin infusion interruption in type 1 diabetic patients. *Diabetes Care*. 2003;26:1137–1141.
60. Wallace TM, Meston NM, Gardner SG, et al. The hospital and home-use of a 30 second hand-held blood ketone meter: guidelines for clinical practice. *Diabet Med*. 2001;18:640–645.
61. Huemer J, Wagner SK, Sim DA. The evolution of diabetic retinopathy screening programmes: a chronology of retinal photography from 35 mm slides to artificial intelligence. *Clin Ophthalmol*. 2020;14:2021–2035.
62. Bakris GL, Williams M, Dworkin L, et al. Preserving renal function in adults with hypertension and diabetes: a consensus approach. *Am J Kid Dis*. 2000;36:646–661.
63. Perkins BA, Olaleye D, Zinman B, et al. Simple screening tests for peripheral neuropathy in the diabetes clinic. *Diabetes Care*. 2001;24:250–256.
64. Owens DR, Zinman B, Bolli GB. Insulins today and beyond. *Lancet*. 2001;358:739–746.
65. Zinman B, Murray FT, Vranic M, et al. Glucoregulation during moderate exercise in insulin treated diabetes. *J Clin Endocrinol Metab*. 1977;45:641–652.
66. Mitchell TH, Abraham G, Schiffrin A, et al. Hyperglycemia after intense exercise in IDDM subjects during continuous subcutaneous insulin infusion. *Diabetes Care*. 1988;11:311–317.
67. Chance RE, Kroeff EP, Hoffmann JA, et al. Chemical, physical and biologic properties of biosynthetic human insulin. *Diabetes Care*. 1981;4:147–154.
68. Kang S, Brange J, Burch A, et al. Subcutaneous insulin absorption explained by insulin's physicochemical properties: evidence from absorption studies of soluble human insulin and insulin analogues in humans. *Diabetes Care*. 1991;14:942–948.
69. Hirsch IB, Juneja R, Beals JM, et al. The evolution of insulin and how it informs therapy and treatment choices. *Endocr Rev*. 2020;41:733–755.
70. Hirsch IB. Ranting in 2020: reflecting about the past and the future (with concerns about the present). *Diabetes Technol Ther*. 2020;22:69–71.
71. Krayenbuhl C, Rosenberg T. Crystalline protamine insulin. *Rep Steno Hosp*. 1946;1:60–73.
72. Barnett AH. A review of basal insulins. *Diabet Med*. 2003;20:873–885.
73. Lauritzen T, Faber OK, Binder C. Variation in 125 I-insulin absorption and blood glucose conception. *Diabetologia*. 1979;17:291–295.
74. Zinman B, Ross S, Campos R, et al. Effectiveness of human ultralente versus NPH insulin in providing basal insulin replacement for an insulin lispro multiple daily injection regimen. *Diabetes Care*. 1999;22:603–608.
75. Holleman F, Hoekstra JBL. Insulin lispro. *New Engl J Med*. 1997;337:176–183.
76. Brems DN, Alter LA, Beckage MJ, et al. Altering the association properties of insulin by amino acid replacement. *Protein Eng*. 1992;5:527–532.
77. Howey DC, Bowsher RR, Brunelle RF, et al. [Lys(B28), Pro(B29)]-human insulin: a rapidly absorbed analogue of human insulin. *Diabetes*. 1994;43:396–402.
78. Lee W, Zinman B. From insulin to insulin analogs: progress in the treatment of type 1 diabetes. *Diabetes Rev*. 1998;6:73–88.
79. Mudaliar SR, Lindberg FA, Joyce M, et al. Insulin aspart (B28 Asp-insulin): a fast-acting analog of human insulin. *Diabetes Care*. 1999;22:1501–1506.
80. Becker RH, Frick AD. Clinical pharmacokinetics and pharmacodynamics of insulin glulisine. *Clin Pharmacokinet*. 2008;47:7–20.
81. Owens DR, Bolli GB. The continuing quest for better subcutaneously administered prandial insulins: a review of recent developments and potential clinical implications. *Diabetes Obes Metab*. 2020;22:743–754.
82. Brunelle RL, Llewelyn J, Anderson JH, et al. Meta-analysis of the effect of insulin lispro on severe hypoglycemia in paitents with type 1 diabetes. *Diabetes Care*. 1998;21:1726–1731.
83. Nicolucci A, Ceriello A, Di Bartolo P, et al. Rapid-acting insulin analogues versus regular human insulin: a meta-analysis of effects on glycemic control in patients with diabetes. *Diabetes Ther*. 2020;11:573–584.
84. Raskin P, Guthrie RA, Leiter L, et al. Use of insulin aspart, a fast-acting insulin analog, as the mealtime insulin in the management of patients with type 1 diabetes. *Diabetes Care*. 2000;23:583–588.
85. Rodbard HW, Rodbard D. Biosynthetic human insulin and insulin analogs. *Am J Ther*. 2020;27:e42–e51.
86. Heise T, Hovelmann U, Brondsted L, et al. Faster-acting insulin aspart: earlier onset of appearance and greater early pharmacokinetic and pharmacodynamic effects than insulin aspart. *Diabetes Obes Metab*. 2015;17:682–688.
87. Heise T, Stender-Petersen K, Hovelmann U, et al. Pharmacokinetic and pharmacodynamic properties of faster-acting insulin aspart versus insulin aspart across a clinically relevant dose range in subjects with type 1 diabetes mellitus. *Clin Pharmacokinet*. 2017;56:649–660.
88. Russell-Jones D, Bode BW, De Block C, et al. Fast-acting insulin aspart improves glycemic control in basal-bolus treatment for type 1 diabetes: results of a 26-week multicenter, active-controlled, treat-to-target, randomized, parallel-group trial (onset 1). *Diabetes Care*. 2017;40:943–950.
89. Mathieu C, Bode BW, Franek E, et al. Efficacy and safety of fast-acting insulin aspart in comparison with insulin aspart in type 1 diabetes (onset 1): a 52-week, randomized, treat-to-target, phase III trial. *Diabetes Obes Metab*. 2018;20:1148–1155.
90. Linnebjerg H, Zhang Q, LaBell E, et al. Pharmacokinetics and glucodynamics of Ultra Rapid Lispro (URLi) versus Humalog (Lispro) in younger adults and elderly patients with type 1 diabetes mellitus: a randomised controlled trial. *Clin Pharmacokinet*. 2020.
91. Klaff L, Cao D, Dellva MA, et al. Ultra rapid lispro improves postprandial glucose control compared with lispro in patients with type 1 diabetes: results from the 26-week PRONTO-T1D study. *Diabetes Obes Metab*. 2020;22:1799–1807.
92. Cheng AYY, Patel DK, Reid TS, et al. Differentiating basal insulin preparations: understanding how they work explains why they are different. *Adv Ther*. 2019;36:1018–1030.
93. Bolli GB, Owens DR. Insulin glargine. *Lancet*. 2000;356:443–444.
94. Heinemann L, Linkeschova R, Rave K, et al. Time-action profile of the long-acting insulin analog insulin glargine (HOE901) in comparison with those of NPH insulin and placebo. *Diabetes Care*. 2000;23:644–649.
95. Owens DR, Coates PA, Luzio SD, et al. Pharmacokinetics of 125I-labelled insulin glargine (HOE901) in healthy men: comparison with NPH insulin and the influence of different subcutaneous injection sites. *Diabetes Care*. 2000;23:813–819.
96. Lepore M, Pampanelli S, Fanelli C, et al. Pharmacokinetics and pharmacodynamics of subcutaneous injection of long-acting insulin analog glargine, NPH insulin, and ultralente human insulin and continuous subcutaneous infusion of insulin lispro. *Diabetes*. 2000;49:2142–2148.
97. Ratner RE, Hirsch IB, Neifing JL, et al. Less hypoglycemia with insulin glargine in intensive therapy for type 1 diabetes. *Diabetes Care*. 2000;23:639–643.
98. Raskin P, Klaff L, Berenstal R, et al. A 16-week comparison of the novel insulin analog insulin glargine (HOE 901) and NPH human insulin used with lispro in patients with type 1 diabetes. *Diabetes Care*. 2000;23:1666–1671.
99. Rosenstock J, Park G, Zimmerman J. Basal insulin glargine (HOE 901) versus NPH insulin in patients with type 1 diabetes on multiple daily insulin regimens. *Diabetes Care*. 2000;23:1137–1142.
100. Misra S, Mathieu C. Are newer insulin analogues better for people with Type 1 diabetes? *Diabet Med*. 2020;37:522–531.
101. Becker RH, Dahmen R, Bergmann K, et al. New insulin glargine 300 units/mL provides a more even activity profile and prolonged glycemic control at steady state compared with insulin glargine 100 units/mL. *Diabetes Care*. 2015;38:637–643.
102. Porcellati F, Lucidi P, Candeloro P, et al. Pharmacokinetics, pharmacodynamics, and modulation of hepatic glucose production with insulin glargine U300 and glargine U100 at steady state with individualized clinical doses in type 1 diabetes. *Diabetes Care*. 2019;42:85–92.

103. Home PD, Bergenstal RM, Bolli GB, et al. New insulin glargine 300 units/mL versus glargine 100 units/mL in people with type 1 diabetes: a randomized, phase 3a, open-label clinical trial (EDITION 4). *Diabetes Care*. 2015;38:2217–2225.
104. Home PD, Bergenstal RM, Bolli GB, et al. Glycaemic control and hypoglycaemia during 12 months of randomized treatment with insulin glargine 300 U/mL versus glargine 100 U/mL in people with type 1 diabetes (EDITION 4). *Diabetes Obes Metab*. 2018;20:121–128.
105. Markussen J, Havelund S, Kurtzhals P, et al. Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. *Diabetologia*. 1996;39:281–288.
106. Heinemann L, Sinha K, Weyer C, et al. Time-action profile of the soluble, fatty acid acylated, long-acting insulin analogue NN304. *Diabet Med*. 1999;16:332–338.
107. Kramer CK, Retnakaran R, Zinman B. Insulin and insulin analogs as antidiabetic therapy: a perspective from clinical trials. *Cell Metab* 2021;33(4):740–747.
108. Hermansen K, Madsbad S, Perrild H, et al. Comparison of the soluble basal insulin analog insulin detemir with NPH insulin. *Diabetes Care*. 2001;24:296–301.
109. Vague P, Selam JL, Skeie S, et al. Insulin detemir is associated with more predictable glycemic control and reduced risk of hypoglycemia than NPH insulin in patients with type 1 diabetes on a basal-bolus regimen with premeal insulin aspart. *Diabetes Care*. 2003;26:590–596.
110. Wang F, Surh J, Kaur M. Insulin degludec as an ultralong-acting basal insulin once a day: a systematic review. *Diabetes Metab Syndr Obes*. 2012;5:191–204.
111. Zinman B. Newer insulin analogs: advances in basal insulin replacement. *Diabetes Obes Metab*. 2013;15:6–10.
112. Ratner RE, Gough SC, Mathieu C, et al. Hypoglycaemia risk with insulin degludec compared with insulin glargine in type 2 and type 1 diabetes: a pre-planned meta-analysis of phase 3 trials. *Diabetes Obes Metab*. 2013;15:175–184.
113. Birkeland KI, Home PD, Wendisch U, et al. Insulin degludec in type 1 diabetes: a randomized controlled trial of a new-generation ultra-long-acting insulin compared with insulin glargine. *Diabetes Care*. 2011;34:661–665.
114. Heller S, Buse J, Fisher M, et al. Insulin degludec, an ultra-longacting basal insulin, versus insulin glargine in basal-bolus treatment with mealtime insulin aspart in type 1 diabetes (BEGIN Basal-Bolus Type 1): a phase 3, randomised, open-label, treat-to-target non-inferiority trial. *Lancet*. 2012;379:1489–1497.
115. Nally LM, Sherr JL, Van Name MA, et al. Pharmacologic treatment options for type 1 diabetes: what's new? *Expert Rev Clin Pharmacol*. 2019;12:471–479.
116. Bode BW, McGill JB, Lorber DL, et al. Inhaled technosphere insulin compared with injected prandial insulin in type 1 diabetes: a randomized 24-week trial. *Diabetes Care*. 2015;38:2266–2273.
117. Monami M, Marchionni N, Mannucci E. Long-acting insulin analogues vs. NPH human in type 1 diabetes. A meta-analysis. *Diabetes Obes Metab*. 2009;11:372–378.
118. Hamann A, Matthaei S, Rosak C, et al. HOE901/4007 Study Group: a randomized clinical trial comparing breakfast, dinner or bedtime administration of insulin glargine in patients with type 1 diabetes. *Diabetes Care*. 2003;26:1738–1744.
119. Goldman J, Kapitza C, Pettus J, et al. Understanding how pharmacokinetic and pharmacodynamic differences of basal analog insulins influence clinical practice. *Curr Med Res Opin*. 2017;33:1821–1831.
120. Vliebergh J, Lefever E, Mathieu C. Advances in newer basal and bolus insulins: impact on type 1 diabetes. *Curr Opin Endocrinol Diabetes Obes*. 2021;28(1):1–7.
121. Pickup J, Keen H. Continuous subcutaneous insulin infusion at 25 years: evidence base for the expanding use of insulin pump therapy in type 1 diabetes. *Diabetes Care*. 2002;25:593–598.
122. Zinman B, Tildesley H, Chiasson JL, et al. Insulin lispro in CSII: results of a double-blind crossover study. *Diabetes*. 1997;46:440–443.
123. Melki V, Renard E, Lassmann-Vague V, et al. Improvement of HbA1c and blood glucose stability in IDDM patients treated with lispro insulin analog in external pumps. *Diabetes Care*. 1998;21:977–982.
124. Renner R, Pfutzner A, Trautmann M, et al. Use of insulin lispro in continuous subcutaneous insulin infusion treatment: results of a multicenter trial: German Humalog-CSII Study Group. *Diabetes Care*. 1999;22:784–788.
125. Pickup JC. Is insulin pump therapy effective in Type 1 diabetes? *Diabet Med*. 2019;36:269–278.
126. Fuchs J, Hovorka R. Closed-loop control in insulin pumps for type-1 diabetes mellitus: safety and efficacy. *Expert Rev Med Devices*. 2020;17:707–720.
127. Pease A, Lo C, Earnest A, et al. The efficacy of technology in type 1 diabetes: a systematic review, network meta-analysis, and narrative synthesis. *Diabetes Technol Ther*. 2020;22:411–421.
128. Bonato L, Taleb N, Gingras V, et al. Duration of catheter use in patients with diabetes using continuous subcutaneous insulin infusion: a review. *Diabetes Technol Ther*. 2018;20:506–515.
129. McGaugh SM, Zaharieva DP, Pooni R, et al. Carbohydrate requirements for prolonged, fasted exercise with and without basal rate reductions in adults with type 1 diabetes on continuous subcutaneous insulin infusion. *Diabetes Care*. 2021;44(2):610–613.
130. Moser O, Riddell MC, Eckstein ML, et al. Glucose management for exercise using continuous glucose monitoring (CGM) and intermittently scanned CGM (isCGM) systems in type 1 diabetes: position statement of the European Association for the Study of Diabetes (EASD) and of the International Society for Pediatric and Adolescent Diabetes (ISPAD) endorsed by JDRF and supported by the American Diabetes Association (ADA). *Pediatr Diabetes*. 2020;21:1375–1393.
131. The Diabetes Control and Complications Trial Research Group. Adverse events and their association with treatment regimens in the Diabetes Control and Complications Trial. *Diabetes Care*. 1995;18:1415–1427.
132. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study Research Group, Jacobson AM, Musen G, Ryan CM, et al. Long-term effect of diabetes and its treatment on cognitive function. *N Engl J Med*. 2007;356:1842–1852.
133. Cryer PE. Managing diabetes: lessons from type 1 diabetes mellitus. *Diabet Med*. 1998;15:S8–S12.
134. Bolli GB, Pampanelli S, Porcellati F, et al. Recovery and prevention of hypoglycemia unawareness in type 1 diabetes mellitus. *Diabetes Nutr Metab*. 2002;15:402–409.
135. The Diabetes Control and Complications Trial Research Group. Influence of intensive diabetes treatment on body weight and composition of adults with type 1 diabetes in the Diabetes Control and Complications Trial. *Diabetes Care*. 2001;24:1711–1721.
136. DeWitt DE, Hirsch IB. Outpatient insulin therapy in type 1 and type 2 diabetes mellitus. *JAMA*. 2003;289:2254–2264.
137. Chantelau E, Kohmer EM. Why some cases of retinopathy worsen when diabetic control improves. *Br Med J*. 1997;315:1105–1106.
138. Jacquier J, Chik CL, Senior PA. A practical, clinical approach to the assessment and management of suspected insulin allergy. *Diabet Med*. 2013;30:977–985.
139. Reeves W, Allen B, Tatersall R. Insulin-induced lipoatrophy: evidence for an immune pathogenesis. *Br Med J*. 1980;280:1500.
140. Nuttal FQ. Carbohydrate and dietary management of clients with insulin-requiring diabetes. *Diabetes Care*. 1993;16:1039–1042.
141. Bell KJ, Barclay AW, Petocz P, et al. Efficacy of carbohydrate counting in type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*. 2014;2:133–140.
142. Joslin EP. The diabetic diet. *J Am Diet Assoc*. 1927;3:89–92.
143. Delahanty LM, Halford BN. The role of diet behaviours in achieving improved glycemic control in intensively treated patients in the Diabetic Control and Complications Trial. *Diabetes Care*. 1993;16: 1453–1458.
144. Turner BC, Jenkins E, Kerr D, et al. The effect of evening alcohol consumption on next-morning glucose control in type 1 diabetes. *Diabetes Care*. 2001;24:1888–1893.
145. Ling W, Huang YM, Qiao YC, et al. Human amylin: from pathology to physiology and pharmacology. *Curr Protein Pept Sci*. 2019;20:944–957.
146. Marrero DG, Crean J, Zhang B, et al. Effect of adjunctive pramlintide treatment on treatment satisfaction in patients with type 1 diabetes. *Diabetes Care*. 2007;30:210–216.
147. Lane K, Freeby M. Adjunctive therapies in type 1 diabetes mellitus. *Curr Opin Endocrinol Diabetes Obes*. 2021;28(1):8–13.

148. Wright LA, Hirsch IB. Non-insulin treatments for Type 1 diabetes: critical appraisal of the available evidence and insight into future directions. *Diabet Med*. 2019;36:665–678.
149. Beysel S, Unsal IO, Kizilgul M, et al. The effects of metformin in type 1 diabetes mellitus. *BMC Endocr Disord*. 2018;18:1.
150. Petrie JR, Chaturvedi N, Ford I, et al. REMOVAL Study Group. Cardiovascular and metabolic effects of metformin in patients with type 1 diabetes (REMOVAL): a double-blind, randomised, placebo-controlled trial. *Lancet Diabetes Endocrinol*. 2017;5:597–609. Erratum in: *Lancet Diabetes Endocrinol*. 2017;5:e7.
151. Libman IM, Miller KM, DiMeglio LA, et al. Effect of metformin added to insulin on glycemic control among overweight/obese adolescents with type 1 diabetes: a randomized clinical trial. *JAMA*. 2015;314:2241–2250.
152. Hafez M, Musa N, Elbehairy S, et al. Effect of metformin on clinical and biochemical hyperandrogenism in adolescent girls with type 1 diabetes. *J Pediatr Endocrinol Metab*. 2019;32:461–470.
153. Guyton J, Jeon M, Brooks A. Glucagon-like peptide 1 receptor agonists in type 1 diabetes mellitus. *Am J Health Syst Pharm*. 2019;76:1739–1748.
154. Ghanim H, Batra M, Green K, et al. Liraglutide treatment in overweight and obese patients with type 1 diabetes: a 26-week randomized controlled trial; mechanisms of weight loss. *Diabetes Obes Metab*. 2020;22:1742–1752.
155. Phillip M, Mathieu C, Lind M, et al. Long-term efficacy and safety of dapagliflozin in patients with inadequately controlled type 1 diabetes: pooled 52-week outcomes from the DEPICT-1 and -2 studies. *Diabetes Obes Metab*. 2021;23(2):549–560.
156. Boeder S, Edelman SV. Sodium-glucose co-transporter inhibitors as adjunctive treatment to insulin in type 1 diabetes: a review of randomized controlled trials. *Diabetes Obes Metab*. 2019;21:62–77.
157. Mathieu C, Van Den Mooter L, Eeckhout B. Empagliflozin in type 1 diabetes. *Diabetes Metab Syndr Obes*. 2019;12:1555–1561.
158. Danne T, Garg S, Peters AL, et al. International consensus on risk management of diabetic ketoacidosis in patients with type 1 diabetes treated with sodium-glucose cotransporter (SGLT) inhibitors. *Diabetes Care*. 2019;42:1147–1154.
159. Peters AL, Buschur EO, Buse JB, et al. Euglycemic diabetic ketoacidosis: a potential complication of treatment with sodium-glucose cotransporter 2 inhibition. *Diabetes Care*. 2015;38:1687–1693.
160. Danne T, Edelman S, Frias JP, et al. Efficacy and safety of adding sotagliflozin, a dual SGLT1 and SGLT2 inhibitor, to optimized insulin therapy in adults with type 1 diabetes and baseline BMI ≥27 kg/m2. *Diabetes Obes Metab*. 2021;23(3):854–860.
161. Albanese-O'Neill A, Grimsmann JM, Svensson AM, Miller KM, Raile K, et al. Changes in HbA1c between 2011 and 2017 in Germany/Austria, Sweden, and the United States: a lifespan perspective. *Diabetes Technol Ther*. 2022;24(1):32–41.
162. Agarwal S, Kanapka LG, Raymond JK, et al. Racial-ethnic inequity in young adults with type 1 diabetes. *J Clin Endocrinol Metab*. 2020;105:e2960–e2969.

40

Clinical Management of Type 2 Diabetes Mellitus

Irl B. Hirsch, David D'Alessio, Nicole Ehrhardt, Alison B. Evert, Lisa K. Gilliam, and Michael Riddell

OUTLINE

Introduction, 691
Diabetes Education, 691
  Diabetes Self-Management Education and Support—the Evidence, 692
  Individualization of the Diabetes Education Care Plan, 692
Nutrition, 692
  Nutrition Therapy—the Evidence, 692
  Individualization of the Nutrition Prescription, 695
Summary, 695
Physical Activity, 695
Pharmacotherapy of Type 2 Diabetes, 696
  Metformin, 696
  Thiazolidinediones, 697
  Insulin Secretagogues, 697
Incretin-Based Therapies, 698
  Glucagon-Like Peptide-1 Receptor Agonists, 698
  Dipeptidyl Peptidase Inhibitors, 699
Sodium-Glucose Cotransporter-2 Inhibitors, 699
  α-Glucosidase Inhibitors, 699
Insulin Therapy in Type 2 Diabetes, 699
  Short-Acting Insulin Analogs, 700
  Long-Acting Insulin Analogs, 700
  Insulin Dosing and Regimens, 700
Practical Aspects for Managing Type 2 Diabetes, 701
  Initial Therapy (Fig. 40.1), 701
  Initial Combination Therapy, 701
  Addition of Medications for Patients not Achieving Glycemic Targets, 701
Diabetes Technology, 702
  Self-Monitoring of Blood Glucose, 703
  Continuous Subcutaneous Insulin Infusion Systems, "Smart" Pens, and Continuous Glucose Monitoring, 703
  Digital Health Technology, 703
  Telemedicine:, 704
Population Management, 705
Conclusion, 706

INTRODUCTION

Currently more than 1 in 10 Americans have diabetes, yet worldwide we can expect a 51% increase by 2045.[1,2] Fortunately, over the past three decades we have gained a better understanding of both islet function and insulin resistance, due to intense research of both nutrition and the impact of physical activity on metabolism. The introduction of a spectrum of new drugs with novel mechanisms of action provided clinicians with new tools. Later came large, randomized controlled trials showing that these agents could improve not only glycemia, but also cardiovascular and renal outcomes. Landmark studies have shown reductions in cardiovascular mortality, heart failure admissions, and need for renal replacement therapy with drugs originally developed to induce a modest reduction in HbA1c. This explosion of clinical trial data, especially in the past 5 years, has been extraordinary. Adding to this the increased use of diabetes-specific technology, treatment of type 2 diabetes is vastly different now than it was a few years ago. Nevertheless, the emphasis on nutrition and physical activity has not changed.

While this chapter is glucocentric in nature, control of blood pressure and hypercholesterolemia are both critical to the overall management of type 2 diabetes. There have also been many advances in the treatment of obesity that have powerful beneficial effects on diabetes outcomes.

The bulk of this chapter will deal with the glycemic approaches that will inform clinical decision-making. A comprehensive review of all the subtleties of diabetes management is beyond the scope of any chapter. An excellent source of information on these issues, which is updated annually, is the Clinical Practice Recommendations from the American Diabetes Association (ADA), which are published every January in Diabetes Care.

DIABETES EDUCATION

Diabetes is a complex condition that requires individuals with diabetes to make decisions about their self-care on a daily basis. Medical nutrition therapy (MNT) and diabetes self-management education and support (DSME/S) provided by qualified healthcare professionals using evidence-based approaches can result in improved clinical, psychological, and behavioral outcomes.[3,4] DSME/S is important for all people with diabetes and necessary to improve outcomes in a disease that is largely self-managed. DSME/S includes the ongoing process that facilitates the knowledge, skills, and abilities necessary for diabetes self-care throughout the life span. The support component includes activities outside of formal training that help individuals implement and sustain the behaviors needed to cope with and manage diabetes.[5] The National Standards for DSME/S are designed to define quality DSME/S and assist those who provide DSME/S services to implement evidence-based DSME/S.[6] The core curriculum topics of DSME/S,[6] also commonly known as the AADE-7 Self-Care Behaviors,[7] demonstrate successful outcomes and must be reviewed to determine which are applicable to the participant, and include the following: diabetes pathophysiology and treatment options; healthy eating; physical activity; medication usage; monitoring and using patient-generated health data; preventing, detecting, and treating acute and chronic complications; healthy coping with psychosocial issues and concerns and problem-solving.

Diabetes Self-Management Education and Support—the Evidence

Numerous studies have shown the benefits of DSME/S, which include improved clinical outcomes and quality of life while reducing hospitalizations and health care costs.[3,8,9] Key takeaways of DSME/S include:

- Average HbA1c reductions of 0.45% to 0.57% compared with usual care for people with a variety of modalities (lifestyle alone, oral and injected antihyperglycemic medication)[4,10]
- Reduction in the onset and/or worsening of diabetes-related complications[11,12]
- Participation in DSME/S reveals enhancement of empowerment and self-efficacy[13]
- Better health outcomes are associated with an increased amount of time spent with a diabetes educator[4,8,14]

Individualization of the Diabetes Education Care Plan

In 2020, a consensus report was authored by a diverse group of healthcare professionals from the ADA, the Association of Diabetes Care and Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Physician Assistants, the American Association of Nurse Practitioners, the American Academy of Family Physicians, and the American Pharmacists Association on DSME/S in adults with type 2 diabetes. The consensus report provides specific recommendations for both clinicians and the health system to increase access to and participation in DSME/S services.[3] The joint effort also provides details on the four key times to implement referrals to DSME/S: 1) at time of diagnosis, 2) on an annual basis, 3) when new complicating factors influence self-management, and 4) at times of transitions in care. The joint statement includes a DSME/S *Algorithm of Care* for healthcare professionals and integrates with MNT delivered by a registered dietician nutritionist (RDN), as well as emotional health provider by a mental health professional, if needed. Another helpful resource included in the consensus report is a checklist that identifies the areas of focus and action steps that can be considered (Table 40.1).

KEY POINTS

- Successful diabetes self-management education and support is evidenced-based, will lower HbA1c levels, and needs to be individualized.

NUTRITION

In addition to DSME/S, MNT provided by an RDN is an essential component of the diabetes treatment plan. The Institute of Medicine broadly defines nutrition therapy as the treatment of a disease or condition through the modification of nutrient or whole-food intake.[15] MNT is an evidence-based application of the Nutrition Care Process provided by the RDN and is the legal definition of nutrition counseling in the United States.[16] Individualized MNT is recommended by the National Academy of Medicine, provided by a RDN upon physician referral.[15] In the 2020 consensus report on the topic of DSME/S, the authors recommended that, in addition to DSME/S, adults with type 2 diabetes should also receive individualized or group diabetes MNT at diagnosis and as needed throughout the lifespan and during times of changing health status to achieve treatment goals.[3] Whenever possible, the unique academic preparation, skills, expertise, and training make the RDN the preferred member of the healthcare team to provide diabetes nutrition therapy. While advanced certification (such as Certified Diabetes Care and Education Specialist, formerly known as Certified Diabetes Educator) is not required, ideally the RDN will have comprehensive experience in and knowledge of diabetes care.[17]

Nutrition Therapy—the Evidence

Clinical trials and outcome studies of diabetes MNT provided by an RDN can improve HbA1c, with absolute reductions of up to 2% in type 2 diabetes at 3 to 6 months, depending on the duration of diabetes and level of glycemia.[17,18] Studies have demonstrated that ongoing RDN-provided follow-up encounters ranging from three to six sessions, with a minimum of one follow-up session annually, can be helpful in maintaining glycemic improvements. During this time, it can be determined whether target goals can be achieved by implementation of MNT in combination with physical activity or whether medication(s) will need to be combined with MNT. In addition, MNT helps prevent, delay, or treat other comorbidities commonly found with diabetes.[12,18] Not all studies of cardiovascular disease (CVD) risk factors report benefits, but in some cases this is likely due to a limited opportunity to show benefit, when, for example, lipid and blood pressure levels were already controlled or only mildly abnormal, or lipid-lowering or antihypertensive medications are present as cointerventions and mask the effect. MNT provided by an RDN is also cost-effective, and people who have received MNT show improved quality of life and clinical outcomes.[17]

The 2019 ADA Nutrition Therapy for Adults with Diabetes and Prediabetes: A Consensus Report reviewed research studies of persons with diabetes following different types of eating patterns, including Mediterranean-style, vegetarian and vegan, low-fat, low-carbohydrate (26%–45% total kcal) and very low carbohydrate (<26% total kcal/ketogenic diet), Dietary Approaches to Stop Hypertension) Paleo, and intermittent fasting.[19] The ADA nutrition consensus report concluded that a variety of eating patterns are acceptable for the management of diabetes. It should be noted that an eating pattern represents the totality of all foods and beverages consumed, whereas an eating plan (or diet) is a guide to help individuals plan what, how much, and when to eat on a daily basis of the foods emphasized in the individual's selected eating pattern.[20] Personal preferences (e.g., tradition, culture, religion, health beliefs and goals, economics) and metabolic goals should be considered when implementing one eating pattern over another. The eating pattern that the individual with type 2 diabetes chooses to follow will likely be the one that they will be able to follow long-term. Refer to Table 40.2 for the eating patterns reviewed for ADA nutrition consensus report, including a description of each eating pattern and list of potential benefits.

The ADA nutrition consensus report also includes recommendations for adults with type 2 diabetes who are overweight or obese. The authors recommend that it is important to encourage reduced energy intake as part of an individualized diabetes eating plan, along with enhanced physical activity.[19] In people with type 2 diabetes, a 5% weight loss is recommended to achieve clinical benefit, and the benefits are progressive. The report also recommends that the goal for optimal outcomes is 15% or more when needed and can be feasibly and safely accomplished. A recent systematic review of the effectiveness of diabetes MNT revealed mixed weight loss outcomes in participants with type 2 diabetes.[17] In addition to an eating plan that results in energy deficit, behavioral strategies and regular physical activity are critical components of any long-term weight loss and maintenance plan.[21,22] Weight loss interventions can be provided in usual care settings and telehealth programs.[19]

TABLE 40.1 Checklist for Providing and Modifying Diabetes Self-Management Education and Support at Four Critical Times

Four Critical Times	Primary Care Provider/ Endocrinologist/Clinical Care Team's Role in Diabetes Education	Diabetes Care and Education Specialist's Role in Diabetes Education
At diagnosis (series of visits)	• Answer questions and provide emotional support regarding diagnosis	• Assess cultural influences, social determinants of health, health beliefs, current knowledge, physical limitations, family support, financial and work status, medical history, learning preferences and barriers, literacy, and numeracy to determine which content to provide and how
	• Shared decision-making of treatment and treatment targets	• Medication—choices, access, action, titration, side effects
	• Teach survival skills to address immediate requirements (safe use of medication, hypoglycemia treatment if needed, introduction of eating guidelines)	• Monitoring blood glucose—when to check, interpreting and using glucose pattern management for feedback
	• Identify and discuss resources for education and ongoing support	• Physical activity—safety, short-term vs. long-term goals/recommendations
	• Make referrals for DSME/S and MNT	• Preventing, detecting, and treating acute and chronic complications
		• Nutrition—food plan, planning meals, purchasing food, preparing meals, portioning food
		• Risk reduction—smoking cessation, foot care, cardiac risk
		• Developing personal strategies to address psychosocial issues and concerns; adjusting to a life with diabetes
		• Developing personal strategies to promote health and behavior change
		• Problem identification and solutions
		• Identifying and accessing resources
Annually and/or when not meeting treatment targets	• Refer for new techniques, technology, and updated information	• Review and reinforce treatment goals and self-management needs
	• Assess and refer if self-management targets not met to address barriers to self-care	• Review barriers to treatment effectiveness
		• Emphasize reducing risk for complications and promoting quality of life
		• Discuss how to adjust diabetes treatment and self-management to life situations and competing demands
		• Support efforts to sustain initial behavior changes and cope with the ongoing burden of diabetes
When complicating factors develop	• Identify presence of factors that inhibit or facilitate achievement of treatment targets and personal goals	• Provide support for the provision of self-management skills in an effort to delay progression of the disease and prevent new complications
	• Discuss impact of complications and successes with treatment and self-management	• Provide/refer for emotional support for diabetes-related distress and depression
		• Develop and support personal strategies for behavior change and healthy coping
		• Develop personal strategies to accommodate sensory or physical limitation(s), adapt to new self-management demands, and promote health and behavior change
When transitions in life and care occur	• Develop diabetes transition plan	• Adjust diabetes self-management plan as needed
	• Communicate transition plan to new health care team members	• Provide support for independent self-management skills and self-efficacy
	• Establish DSME/S regular follow-up care	• Identify level of significant other involvement and facilitate education and support
		• Assist with facing challenges affecting usual level of activity, ability to function, health benefits, and feelings of well-being
		• Maximize quality of life and emotional support for the person with diabetes (and family members)
		• Provide education for others now involved in care
		• Establish communication and follow-up plans with the provider, family, and others
		• Develop goals and personal strategies to promote health and behavioral change and improve quality of life

MNT, Medical nutrition therapy; *DSME/S,* diabetes self-management education and support.

From Margaret A. Powers, et. al., Diabetes Self-management Education and Support in Adults With Type 2 Diabetes: A Consensus Report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. *Diabetes Care*. 2020;43(7):1636–1649. https://doi.org/10.2337/dci20-0023.

TABLE 40.2 **Eating Patterns Reviewed for this Report**

Type of Eating Pattern	Description	Potential Benefits Reported
USDA Dietary Guidelines For Americans (DGA)	Emphasizes a variety of vegetables from all of the subgroups; fruits, especially whole fruits; grains, at least half of which are whole intact grains; lower-fat dairy; a variety of protein foods; and oils. This eating pattern limits saturated fats and *trans* fats, added sugars, and sodium.	DGA added to the table for reference; not reviewed as part of this Consensus Report
Mediterranean-style	Emphasizes plant-based food (vegetables, beans, nuts and seeds, fruits, and whole intact grains); fish and other seafood; olive oil as the principal source of dietary fat; dairy products (mainly yogurt and cheese) in low to moderate amounts; typically fewer than four eggs/week; red meat in low frequency and amounts; wine in low to moderate amounts; and concentrated sugars or honey rarely.	• Reduced risk of diabetes • HbA1c reduction • Lowered triglycerides • Reduced risk of major cardiovascular events
Vegetarian or vegan	The two most common approaches found in the literature emphasize plant-based vegetarian eating devoid of all flesh foods but including egg (ovo) and/or dairy (lacto) products, or vegan eating devoid of all flesh foods and animal-derived products.	• Reduced risk of diabetes • HbA1c reduction • Weight loss • Lowered LDL-C and non–HDL-C
Low-fat	Emphasizes vegetables, fruits, starches (e.g., breads/crackers, pasta, whole intact grains, starchy vegetables), lean protein sources (including beans), and low-fat dairy products. In this review, defined as total fat intake ≤30% of total calories and saturated fat intake ≤10%.	• Reduced risk of diabetes • Weight loss
Very low-fat	Emphasizes fiber-rich vegetables, beans, fruits, whole intact grains, nonfat dairy, fish, and egg whites and comprises 70%–77% carbohydrate (including 30–60 g fiber), 10% fat, 13%–20% protein.	• Weight loss • Lowered blood pressure
Low-carbohydrate	Emphasizes vegetables low in carbohydrate (such as salad greens, broccoli, cauliflower, cucumber, cabbage, and others); fat from animal foods, oils, butter, and avocado; and protein in the form of meat, poultry, fish, shellfish, eggs, cheese, nuts, and seeds. Some plans include fruit (e.g., berries) and a greater array of nonstarchy vegetables. Avoids starchy and sugary foods such as pasta, rice, potatoes, bread, and sweets. There is no consistent definition of "low" carbohydrate. In this review, a low-carbohydrate eating pattern is defined as reducing carbohydrates to 26%–45% of total calories.	• HbA1c reduction • Weight loss • Lowered blood pressure • Increased HDL-C and lowered triglycerides
Very low-carbohydrate (VLC)	Similar to low-carbohydrate pattern but further limits carbohydrate-containing foods, and meals typically derive more than half of calories from fat. Often has a goal of 20–50 g of nonfiber carbohydrate per day to induce nutritional ketosis. In this review a VLC eating pattern is defined as reducing carbohydrate to <26% of total calories.	• HbA1c reduction • Weight loss • Lowered blood pressure • Increased HDL-C and lowered triglycerides
Dietary Approaches to Stop Hypertension	Emphasizes vegetables, fruits, and low-fat dairy products; includes whole intact grains, poultry, fish, and nuts; reduced in saturated fat, red meat, sweets, and sugar-containing beverages. May also be reduced in sodium.	• Reduced risk of diabetes • Weight loss • Lowered blood pressure
Paleo	Emphasizes foods theoretically eaten regularly during early human evolution, such as lean meat, fish, shellfish, vegetables, eggs, nuts, and berries. Avoids grains, dairy, salt, refined fats, and sugar.	• Mixed results • Inconclusive evidence

Source: Evert AB, Dennison M, Gardner C, et al. Nutrition therapy for adults with diabetes and prediabetes: a consensus report. *Diabetes Care*. 2019;42:731–754.

USDA, US Department of Agriculture; *LDL-C,* low-density lipoprotein cholesterol; *HDL-C,* high-density lipoprotein cholesterol.

From Evert, et al. Eating patterns reviewed for ADA Nutrition Consensus Report

Although numerous factors influence glycemic response to foods, monitoring intake of carbohydrates has demonstrated the most evidence for improving glycemia and may be applied in a variety of eating patterns that meet individual needs and preferences.[19] Evidence exists that both the quantity and type of carbohydrate eaten influence blood glucose levels; however, the total amount of carbohydrate eaten is the primary predictor of glycemic response.

For select adults with type 2 diabetes not meeting glycemic targets, or where reducing glucose-lowering medications is a priority, reducing overall carbohydrate intake with low- or very low-carbohydrate eating plans is a viable approach.[19] Day-to-day consistency in the amount of carbohydrate eaten at meals and snacks is reported to improve glycemic control, especially in persons on either MNT alone, glucose-lowering medications, or fixed insulin regimens.

Individualization of the Nutrition Prescription

The first step in developing a meal plan for the person with diabetes is to base it on an individualized assessment.[19] The initial visit with a RDN takes approximately 1 hour. The RDN focuses on the preferences and needs of the individual with diabetes to collaboratively develop a diabetes eating plan that can be followed long-term.[17,23] The RDN will take a detailed history, including an analysis of usual eating habits, past diets, comorbidities that affect nutrition, socioeconomic factors, cultural influences, and readiness to change in order to, collaboratively with the patient, identify the best approach to meal planning. This patient-centered approach sets the stage for the adoption of small, incremental lifestyle changes that include modifications in food and beverage choices as well as physical activity to improve overall health, reduce the risk of diabetes complications, and manage complications if indicated.[23]

Common meal planning approaches from the simple to complex include some of the following: reducing portion sizes, using the plate method, carbohydrate or calorie counting, using meal replacements, as well as incorporating behavioral strategies such as mindful eating, all while working within the eating preferences described by the person with diabetes.

Research studies and practice papers report the role of the RDN in organization-approved medication adjustment protocols.[24,25] The experienced RDN who has demonstrated competency in medication adjustment and practicing within their scopes of practice can help to reduce clinical inertia and/or reduce risk of hypoglycemia and hyperglycemia.

SUMMARY

MNT and DSME/S referrals should be made at the diagnosis of type 2 diabetes, annually, when complications may arise, and when transitions in life occur. Initial treatment recommendations typically include lifestyle modification and possibly medication. Due to the progressive nature of type 2 diabetes, the treatment plan will likely evolve over time to include intensification of medications to optimally manage glucose levels. However, ongoing lifestyle modification and behavioral support continue to be critical components of care and should be integrated with the overall treatment plan through MNT and DSME/S.

KEY POINTS

- Medical nutrition therapy (MNT) is an essential component of treatment for type 2 diabetes. Evidenced-based MNT can improve HbA1c, in addition to preventing, delaying, or treating other comorbidities commonly found with diabetes. Like most aspects of diabetes therapy, care needs to be individualized.

PHYSICAL ACTIVITY

Physical activity is characterized as any body movement that increases energy expenditure above rest, whereas exercise is a planned, purposeful, and often structured form of physical activity. Regular physical activity and structured exercise sessions can improve daily metabolic control, enhance well-being, and reduce cardiovascular risk factors in type 2 diabetes.[26] Aerobic physical activity involves repeated and continuous movement, typically for more than 5 to 10 minutes, of large muscle groups (e.g., walking, cycling, jogging, swimming). Although walking may be the most convenient activity mode, some individuals, because of peripheral neuropathy and/or other physical impairment, may need to perform non weight-bearing activities. Acute aerobic sessions are typically associated with a drop in baseline glucose level and/or a reduction in the postmeal glucose excursion.[26] Resistance or strength-based activities are defined as activities associated with contracting muscles against a load, often using one's own body weight, or by using free weights or resistance bands. These activities can be associated with a transient rise in glycemia,[27,28] perhaps associated with the rise in counterregulatory hormones.[27] Flexibility activities include stretching and/or moving joints through their range of motion, while balance exercises typically include a transfer of body weight to a single limb with or without another touch point. These activities tend not to change glucose levels significantly but are important to help with flexibility and maintaining joint range of motion.[26]

The importance of regular physical activity and structured exercise sessions for people living with type 2 diabetes is often underappreciated. In this population, extended sedentary time and prolonged sitting are associated with poor glycemic control and an enhanced cardiovascular risk profile.[29] Interrupting prolonged siting with some form of physical activity improves postprandial glucose control, insulin sensitivity, and plasma triglyceride levels.[29] Beyond breaking up sedentary time with light activity, exercise training sessions can further improve cardiometabolic outcomes.[26] While consensus guidelines exist for the optimal volume, mode, intensity, and duration of exercise sessions for people living with type 2 diabetes, it should be acknowledged that these recommendations should vary depending on the individual's characteristics, motivations, and current health status.[26] Because individuals with type 2 diabetes typically have a reduced aerobic capacity and a higher rating of perceived exertion,[30] the physical activity prescription may need to be gradually increased from a modest volume at a light intensity (e.g., 10–15 minutes of light walking per day) to a higher volume and a more vigorous intensity (e.g., 45 minutes of brisk walking, jogging, swimming, or cycling), with the addition of other training modes (e.g., resistance training, high intensity interval training [HIIT]) as the tolerance for exercise steadily improves (Table 40.3).

Performing aerobic activities regularly (i.e., three times per week for 20–40 minutes) for 8 or more weeks increase skeletal muscle oxidative capacity, insulin sensitivity, blood vessel reactivity and compliance, capillary density, lung function, immune function, and cardiac performance.[31] Daily engagement in some form of sustained aerobic activity is ideal, because the favorable changes in glucose tolerance and insulin sensitivity typically deteriorate in approximately 24 to 48 hours, depending on the length and intensity of the session.[30] Resistance exercise (i.e., weightlifting) has additive benefits over aerobic exercise and should be performed at least two times per week, using all the major muscle groups (two to three sets of 13 repetitions each).[26] Resistance exercise helps maintain muscle mass, strength, balance, and insulin sensitivity[26] and promotes favorable effects on ambulatory blood pressure and glycemia in the 24-hour period after the session.[32] With weekly aerobic and/or resistance exercise sessions, a lower HbA1c level by 0.5% to 1.0% is typically achieved, along with reductions in

TABLE 40.3 Exercise

	Phase 1	Phase 2	Phase 3	Phase 4
Aerobic				
Duration (min)	10–15	15–20	20–30	30–45
Intensity	Light	Moderate	Somewhat hard	Hard
Frequency (days/week)	3	4	4	5
Resistance				
# of sets per muscle group	1	2	3	3
# of repetitions	10	10	10	8
Intensity (% of 1 repetition maximal	50%	50%	70%	80%
Frequency (days/week)	2	2	3	3
HIIT (optional)*				
Frequency (days/week)			2	3
Min low, minute high, number of cycles			1, 2, 3	2, 2, 5
Intensity (% maximal aerobic capacity during low and high intervals)			40%, 70%	50%, 90%

Exampled phase approach to ramp up the physical activity prescription in individuals living with type 2 diabetes. Each phase may last 2–3 weeks before progressing, if desired. Note: Baseline assessment of physical activity patterns should be performed to determine what phase to initiate. * Note: HIIT can replace one or more aerobic or resistance exercise sessions weekly up to three HIIT sessions per week. *HIIT*, High-intensity interval training.

visceral fat mass, but body weight may not change significantly because of increases in lean mass.[26,30]

The impact that regular physical activity has on cardiovascular mortality in type 2 diabetes is somewhat controversial. A large randomized controlled clinical trial of intensive lifestyle therapy on cardiovascular morbidity and mortality in overweight/obese adults with type 2 demonstrated that lifestyle intervention, which recommended regular physical activity (of at least 175 min/week of unsupervised aerobic activity, along with a low-calorie diet) for a sustained period of time (median 9.6 years) improves glycemic outcomes and cardiovascular risk profile, among several other beneficial effects,[33] including reduced sleep apnea, reduced disability incidence,[34] and enhanced cognitive functioning.[35] However, it is important to note that the incidence of the primary outcome, a composite of fatal or nonfatal myocardial infarction, stroke, hospitalization for angina, or cardiovascular death, did not differ between the intensive lifestyle group and the standard of care group.[33] The intensive lifestyle intervention appeared to reduce cardiovascular events in those who achieved the desired weight loss goal of 7%[36] and in those who had no baseline CVD.[37] Overall, exercise interventions are largely deemed safe and effective with multiple comorbidities associated with diabetes.[38]

HIIT, which is a time-restricted exercise session composed of high- and low-intensity activities lasting anywhere between 6 and 20 minutes, has been demonstrated to have numerous health and fitness benefits in people living with type 2 diabetes.[39-42] However, there is a lack of clarity on whether HIIT is more or less effective (or as safe) as traditional aerobic exercise sessions.[41]

For individuals on insulin or insulin secretagogues, increased physical activity increases the risk for hypoglycemia, including severe hypoglycemia, particularly in settings of weight loss and caloric restriction.[43] As such, lifestyle interventions with exercise should be accompanied with sulfonylurea and/or insulin dose reductions. As noted above, higher intensity effort, as is associated with HIIT, likely reduces the likelihood of exercise-associated hypoglycemia but may result in sustained hyperglycemia if engaged in in the early morning.

KEY POINTS

- Exercise is a key component of treatment for type 2 diabetes and needs to be individualized based on many factors. While daily engagement in some form of sustained aerobic activity is ideal, resistance exercise has added benefits. Exercise interventions are largely deemed safe and effective with multiple comorbidities associated with diabetes.

PHARMACOTHERAPY OF TYPE 2 DIABETES

There are now a range of medications available that effectively reduce blood glucose in persons with type 2 diabetes. Although the expansion of treatment options over the past two decades raises the possibility of widespread improvements to glycemic control, this has not yet been borne out across the population.[44] Moreover, while the differing mechanisms of action among the many agents now on the market suggest potential synergistic combinations, this has not been conclusively demonstrated in general practice or in specific groups of patients. Nonetheless, the variety of options now available to providers is a major advance in diabetes care. The major glucose-lowering medications are reviewed along with their cardiovascular and renal effects (Table 40.4).

Metformin

Metformin has been the consensus choice for initial treatment of type 2 diabetes[45] and should be started at the time of diagnosis in most patients. Metformin has potent glucose-lowering actions, is inexpensive, and may reduce risk of cardiovascular events and death.[46] Although not studied in large randomized trials, observational studies have consistently reported safety of the drug in a wide range of type 2 diabetes patient groups, inclusive of those with major comorbidities.[47] Metformin has little effect on blood glucose in normoglycemic states, does not stimulate the release of insulin or other islet hormones, and rarely causes hypoglycemia. However, even in persons with only mild hyperglycemia, metformin lowers blood glucose. Metformin is effective in combination with other glucose-lowering medications for type 2 diabetes, and fixed-dose combinations are available for many of these.

The immediate release form should be titrated to an initial dose of 500 mg twice a day over several weeks to prevent gastrointestinal complaints; there is little therapeutic benefit to using more than 2000 mg daily.

Metformin is cleared by renal filtration, and very high circulating levels (e.g., as a result of overdose or acute renal failure) have been associated with lactic acidosis. However, the occurrence of this complication is now known to be rare, and metformin may be safely used in patients with reduced estimated glomerular filtration rates (eGFR). The US Food and Drug Administration has revised the label for metformin to reflect its safety in patients with eGFR greater than or equal to 30 mL/min/1.73 m^2; dose reduction starting at eGFR of 40 to 50 mL/min/1.73 m^2 is also recommended. Hospitalized patients are at increased risk for acute kidney injury, and metformin should be used cautiously in the inpatient setting.[45] Metformin use is associated with vitamin B12 deficiency, likely the result of abnormal gastrointestinal absorption, and this can compound existing diabetic neuropathy.

TABLE 40.4 **Glucose-Lowering Medications**

Class	HbA1c Reduction	Effect on Weight	Primary Side Effects	Contraindications
Metformin	1%–2%	neutral	Gastrointestinal	eGFR <30 mL/min/1.73 m^2
Sulfonylurea	1%–2%	1- to 3-kg gain	Hypoglycemia	Type 1 diabetes
Thiazolidinedione	0.5%–1%	2- to 4-kg gain	Fluid retention, bone fracture	Heart failure
Glucagon-like peptide 1 receptor agonist	0.5%–1.5%	Weight loss variable—up to 7 kg (semaglutide)	Gastrointestinal in nature	Gastroparesis; some still concerned about pancreatitis
Sodium glucose cotransporter 2 inhibitor	0.5%–0.7%	2- to 4-kg loss	Genital mycotic infection, urinary tract infection	Type 1 diabetes
Dipeptidyl peptidase 4 Inhibitors	0.5%–0.8%	Neutral	Nothing consistent	Heart failure with saxagliptin

eGFR, Estimated glomerular filtration rate.

Thiazolidinediones

Thiazolidinediones are ligands for the PPARγ receptor, a nuclear hormone receptor involved in the regulation of genes related to glucose and lipid metabolism. Only pioglitazone is broadly available to treat type 2 diabetes. Pioglitazone is an insulin sensitizer and increases insulin-mediated glucose uptake by 30% to 50% in patients with type 2 diabetes, primarily in adipose and skeletal muscle. In addition to promoting peripheral glucose disposal, this drug reduces hepatic glucose output. Beyond effects on glucose metabolism, pioglitazone reduces plasma triglycerides by 10% to 15% and raises high-density lipoprotein– and low-density lipoprotein–cholesterol. Pioglitazone has demonstrated consistent effects to reduce hepatic steatosis and nonalcoholic steatohepatitis in affected patients with type 2 diabetes.[48] Finally, pioglitazone significantly reduces the progression of at-risk individuals to type 2 diabetes[49,50] and decreases rates of macrovascular events in high-risk insulin-resistant people at risk for type 2 diabetes.[50]

Pioglitazone is inexpensive and is started at a dose of 15 to 30 mg. It is metabolized by the liver and may be administered to patients with renal insufficiency, but should not be used if there is active hepatic disease. Pioglitazone causes weight gain and fluid retention, dose-dependent side effects that can be limiting. Beyond average gains of 2 to 4 kg of fat mass, there is plasma volume expansion; the use of thiazolidinediones requires caution and careful monitoring, and they should not be used in patients with moderate to severe heart failure. Treatment with pioglitazone has been associated with a consistent increased risk of bone fracture in women, with some studies also showing effects in men. Therefore, the presence of osteoporosis and other risks for fracture should be considered before starting this drug. While preclinical studies and observations in treated patients raised the possibility that pioglitazone increased the risk of bladder cancer, recent observational[51] and pharmacoepidemiologic[52] studies suggest that this is not a substantial risk.

Insulin Secretagogues

Agents in this class lower blood glucose entirely through stimulation of insulin release islet β cells. The two subclasses of drugs, sulfonylureas and meglitinides, act by binding the sulfonylurea receptor to trigger ion fluxes in the β cell that promote insulin exocytosis.

Sulfonylureas. Sulfonylureas (SUs) in common usage in the United States include glimepiride, glipizide, and glyburide. These drugs have comparable effectiveness and cause significant glucose lowering in 50% to 80% of people with type 2 diabetes. They are inexpensive and have well-established safety profiles. Chronic administration of SUs leads to an apparent shift in β cell sensitivity to glucose. While secondary failure, i.e., gradual loss of efficacy over months to years of treatment, is a frequently cited concern with SU use, this occurs with most diabetes drugs and is most likely due to the progressive β cell failure that is at the core of type 2 diabetes.[53] Based on the results of prospective treatment studies, treatment with SU reduces microvascular complications due to diabetes.[54]

SUs are generally well tolerated and are available in once-daily formulations that increase compliance. SUs are effective as monotherapy and when combined with other glucose-lowering drugs, including insulin. The SUs have wide dose ranges, but the maximally effective doses are generally lower than the maximum marketed doses. To minimize adverse drug effects, use of low doses of SUs is prudent, with addition of other medications if a therapeutic effect is not realized at a dose 50% of the suggested maximum.

The most common side effects of SUs are hypoglycemia and weight gain. Hypoglycemia due to SUs is the result of β cell stimulation independent of blood glucose, with continued insulin secretion at near euglycemic levels causing continued glucose disposal and impaired counterregulation. Fortunately, severe hypoglycemia is not common,[54] but care is required when treating elderly people or those with significant comorbidities or irregular access to food. Hypoglycemia is dose-dependent, but there is also evidence that glyburide causes higher rates of events than other SUs.

There has long been controversy as to the cardiovascular safety of SU, with an early trial and some metaanalyses suggesting that these drugs increase rates of atherosclerotic CVD (ASCVD) events. However, two recent randomized clinical trials including subjects at high and low risk for CVD demonstrated no difference in major adverse cardiovascular events (MACEs) in subjects treated with SUs or comparators, either linagliptin[55] or pioglitazone,[56] shown in other studies to have no deleterious effect on CVD risk. Thus, at present it is reasonable to conclude that SUs confer neither risk or benefit on cardiovascular risk over a period of 5 years.

Meglitinides. This class of secretagogues is distinct from the SUs in that they do not bind directly to SU receptors, but rather modulate the membrane electrical activity and ion fluxes that are the downstream effectors of SUs. There are two meglitinides: nateglinide and repaglinide. These agents are rapidly absorbed and produce a more rapid but shorter stimulus for insulin secretion. Because of the short duration of action, these medications must be given with each meal, and thus are useful primarily for prandial glucose control. The pharmacokinetics of the meglitinides may also explain the lower rates of hypoglycemia and weight gain than those associated with SUs. Similar to the SUs, there is a minimal glucose-lowering advantage of high doses versus moderate doses of repaglinide.

INCRETIN-BASED THERAPIES

The term *incretin* was originally coined almost 100 years ago to describe factors released by the gut that augment glucose clearance after meals. In recent decades incretin has come to mean insulinotropic gastrointestinal hormones that augment glucose-stimulated insulin secretion. There are two established incretins, glucose-dependent insulinotropic peptide (GIP) and glucagon-like peptide 1 (GLP-1), that account for 50% to 70% of prandial insulin secretion and are necessary for normal glucose tolerance.[57] GLP-1 is a potent insulin secretagogue that also affects blood glucose by delaying gastric emptying, as well as causing satiety. GIP also stimulates insulin secretion, and like GLP-1 does so only when blood glucose levels are elevated. GLP-1 and GIP are rapidly inactivated by the protease dipeptidyl peptidase-4 (DPP-4) that is widely distributed on capillary endothelium; GIP and GLP-1 have plasma half-lives of 7 and 2 minutes, respectively. Currently two approaches have been applied to harness the endogenous incretin system to treat type 2 diabetes[58]: injectable peptides modified for extended residence in the circulation (GLP-1 receptor agonists [GLP-1RAs]), and small molecules that inhibit DPP-4 activity and extend the activity of GLP-1 and GIP (DPP-4 inhibitors [DPP-4i]).

Because GLP-1RAs and DPP4is act through the incretin system, they are frequently discussed as parallel treatment options. In fact, these two subclasses of incretin-based therapies have several major differences. DPP-4is are orally available, increase incretin levels by approximately 2-fold, have only moderate glucose lowering potency, and do not affect body weight, but have very favorable safety and tolerability profiles. GLP-1RAs are administered by injection, raise GLP-1 activity to supraphysiologic levels, are among the most potent glucose-lowering drugs, and cause weight loss, but have common symptomatic side effects.

Glucagon-Like Peptide-1 Receptor Agonists

There are six GLP-1RAs available for use in the United States. Four of these agents are based on the sequence of native GLP-1, modified to reduce inactivation by DPP-4 and extend circulating half-life, and two use exendin-4, an analog of a reptilian peptide that binds the GLP-1 receptor. The most notable differences among these agents are their duration of action, which contributes to both their activity and their dosing frequency. However, because all act through the same specific signaling system, with minimal off-target effects, their pharmacology is broadly shared. All GLP-1RAs stimulate insulin secretion and reduce plasma glucagon, delay gastric emptying and intestinal motility, induce satiety and weight loss, cause modest reductions in blood pressure and increases in heart rate, and are associated with nausea and other gastrointestinal side effects. While there may be some differences among agents in the magnitude of these effects, in any given patient some version this pharmacologic profile can be expected. Recent evidence indicates that GLP-1RAs reduce the risk of MACEs and improve parameters of diabetic nephropathy[59]; it is unclear whether these effects are shared across the class.

The hallmark side effect of GLP-1RAs is nausea, affecting approximately 40% of subjects, with lesser rates of vomiting and diarrhea. Gastrointestinal side effects are dose-dependent, and titration from low to high doses is generally effective in limiting them. Moreover, most patients tend to develop tolerance to the drugs over 2 to 3 weeks of treatment. Treatment with GLP-1RAs does not commonly cause hypoglycemia, but can potentiate this effect when given with SUs or insulin. The exendin-based drugs are renally excreted and should be avoided in persons with advanced renal insufficiency; the GLP-1–based drugs are not renally excreted, but there is limited experience with use in people who have eGFRs below 30 mL/min/1.73 m^2. There is also limited published experience on treating patients with significant hepatic dysfunction with GLP-1RAs, but no consistent adverse effects have been reported. Previous concerns about links between GLP-1RA use and pancreatitis, pancreatic cancer, and metabolic bone disease have been mostly dispelled.[60]

The currently prescribed GLP-1RAs are:

Exenatide is a synthetic replica of the 39–amino acid reptilian peptide exendin-4. Exenatide is a potent GLP-1RA that shares many of the effects of GLP-1. It is not metabolized by DPP-4, and so has extended activity following injection. In acute studies, exenatide reduces prandial blood glucose by augmenting insulin secretion and delaying gastric emptying. Exenatide is approved for use as monotherapy and in combination with other glucose-lowering drugs. Exenatide comes in a short-acting formulation with a plasma half-life of 2 to 3 hours that is injected twice daily in doses of 5 or 10 mcg. There is also a long-acting formulation injected as a suspension of embedded exenatide on a polymeric microsphere that delays absorption, allowing weekly dosing at 2 mg.

In clinical trials, exenatide, alone or in combination with metformin, SU, thiazolidinedione, or basal insulin was associated with an approximately 1% decrease in HbA1c, and weight loss that averaged 2.5 to 4 kg. The extended-release form of exenatide conferred no risk of CVD when evaluated in a randomized cardiovascular outcome trial.[61]

Lixisenatide is the other exendin-4 based GLP-1RA, with short extension that distinguishes it from exenatide. Lixisenatide has similar pharmacokinetics to exenatide but is used as a once-daily injection. It is available for coadministration with insulin glargine in a single-injection pen. Data from a recent cardiovascular outcome trial (CVOT) of diabetic subjects with a prior history of CVD indicated that lixisenatide was safe but did not have beneficial effects to reduce rates of subsequent MACEs.

Liraglutide is a long-acting, DPP-4–resistant form of GLP-1, in which the native sequence is modified by addition of an α-glutamic acid spacer coupled to a C16 fatty acyl group at Lys 26.[61] The fatty acid side chain permits binding to albumin and other plasma proteins and accounts for a half-life of 12 hours, permitting once-daily administration. The pharmacodynamic profile of liraglutide mimics GLP-1 and exenatide, and chronic use leads to reductions in HbA1c of 1.0% to 1.5% and weight loss of 4 kg. Liraglutide can be added to oral agents or basal insulin. It is administered in 1.2-mg or 1.8-mg doses and is also available for fixed-dose combination administration with insulin degludec.

Liraglutide was the first of the GLP-1RAs to show a significant benefit to reduce MACEs (nonfatal myocardial infarction, or nonfatal stroke, and cardiovascular death) in subjects with established CVD and type 2 diabetes.[62] This finding, and the results of CVOT for other GLP-1RAs, has led to new recommendations favoring the use of these drugs in persons with, or at high risk for, CVD.[63] Liraglutide also reduces the progression of diabetic nephropathy, as marked by proteinuria, in affected patients.[64]

Dulaglutide is a fusion protein consisting of two molecules of a DPP-4–resistant version of GLP-1 linked to the Fc portion of a human immunoglobulin. Attachment to a large carrier protein delays clearance of dulaglutide, and even though it is rapidly absorbed after subcutaneous injection, a protracted circulating half-life allows for once-weekly dosing at 0.75 or 1.5 mg. Average reduction of HbA1c is 1.0% to 1.5%, with weight loss of 2 to 3 kg. Dulaglutide has similar benefits in terms of decreasing MACE and proteinuria as liraglutide.[61]

Semaglutide has a similar structure to liraglutide but differs in two amino acids and the fatty acyl group and linker. These changes increase binding affinity to albumin, giving an extended circulating half-life that allows once-weekly dosing. Semaglutide is the most potent GLP-1RA for glucose lowering, with mean reductions in HbA1c of 1.5% to 2.0% that are greater than those seen with liraglutide and dulaglutide. Semaglutide is also more effective for weight loss, with chronic treatment causing 5- to 7-kg reductions. Semaglutide has similar benefits in terms of decreasing MACE and proteinuria as liraglutide.[61]

Oral semaglutide is now available for clinical use. The active compound is identical to the injectable formulation, but includes a modified fatty acid absorption enhancer to allow passage across the gastrointestinal tract (likely the gastric mucosa). Only approximately 1% of oral semaglutide crosses into the circulation, and the drug must be ingested daily in the absence of food. The oral form of semaglutide has a potency for glucose-lowering and weight loss more comparable to liraglutide than injectable semaglutide, and it shares the beneficial effects on cardiovascular and nephropathic risk reduction.[61]

Dipeptidyl Peptidase Inhibitors

DPP-4 is a widely distributed serine protease expressed as an ectoenzyme on vascular endothelial cells and T-lymphocytes. DPP-4 inactivates a range of peptides, including GLP-1 and GIP, through specific cleavage. DPP-4is increase circulating levels of active, unmetabolized GLP-1 and GIP after meals, and enhanced incretin action is the mechanism of their glucose lowering effect. There are four DPP4is available in the United States: sitagliptin, saxagliptin, linagliptin, and alogliptin. Vildagliptin is available in the European Union and other countries.

Most clinical data suggest that the agents in this class are comparable. Administration causes elevation of active GIP and GLP-1 in the circulation, increases insulin secretion, reduces glucagon levels, and improves both fasting and postprandial hyperglycemia.[63] Inhibition of DPP-4 does not appear to have direct effects on insulin sensitivity, gastric motility, or satiety, nor does chronic treatment with a DPP-4i affect body weight. DPP-4is, used as monotherapy in type 2 diabetic patients, reduce HbA1c levels by an average of 0.5% to 1.0%. These compounds are also effective for chronic glucose control when added to the treatment of diabetic patients receiving metformin, thiazolidinediones, SUs, and insulin, with a further reduction in HbA1c of 0.5%.

There are no consistent adverse effects that have been noted in clinical trials with any of the DPP-4is. CVOTs have been completed for alogliptin, saxagliptin, linagliptin, and sitagliptin. There was no impact of these drugs on the incidence of cardiovascular events in diabetic patients; patients treated with saxagliptin had an increase in hospitalization for heart failure. The DPP-4is are renally cleared except for linagliptin, and for the other agents dose reductions are recommended for patients with renal insufficiency.

SODIUM-GLUCOSE COTRANSPORTER-2 INHIBITORS

Sodium-glucose cotransporter-2 (SGLT2) is a sodium-glucose cotransporter located almost exclusively in the proximal portion of the renal tubule. On an average day most humans reabsorb all of the 150 to 200 g of glucose filtered through their kidneys, and SGLT2 accounts for 80% to 90% of this reclamation. The SGLT-2 inhibitors (SGLT2is) were developed to use renal glucose elimination to lower blood glucose. There are currently four SGLT2is inhibitors available for clinical use: canagliflozin, dapagliflozin, empagliflozin, and ertugliflozin. The pharmacologic actions of SGLT2is are targeted and specific, and most clinical data suggests that the agents in this class are comparable. They have relatively long plasma half-lives and can be dosed once daily. SGLT2is can be used as monotherapy and in combination with other oral agents and insulin. Glucose-lowering potency is moderate, with average reductions in HbA1c of 0.7% to 1.0%, and the glycemic effect is dependent on renal function, with the amount of glucose that can be excreted a function of the amount filtered. SGLT2is also decrease blood pressure by 2 to 4 mm Hg and cause an approximately 4-kg weight loss.

The side effects of SGLT2is are predictable from their mechanism of action. There is a small increase in lower urinary tract infections, a 3-fold increase in genital mycotic infections, and mild volume contraction with rare episodes of acute kidney injury.[63] SGLT2is do not cause hypoglycemia but can increase the likelihood when combined with drugs that do. Canagliflozin was associated with a small increase in the risk of fractures and amputations of the lower extremity in a CVOT, but these results have not been observed with other SGLT2is and are not well-established side effects. SGLT2is cause a small but significant increase in diabetic ketoacidosis when given to patients with type 1 diabetes. Diabetic ketoacidosis in persons with type 2 diabetes is reported but rare, and occurs mostly in those treated with insulin, raising the possibility that some of these patients have insulinopenic diabetes.

Recent clinical trials have demonstrated important benefits of SGLT2is in persons with major comorbidities of type 2 diabetes. Empagliflozin, canagliflozin, and dapagliflozin have all demonstrated beneficial effects in lowering rates of MACE.[65] The effects seem to be independent of glucose lowering, and are most pronounced for patients with established CVD, heart failure with reduced ejection fraction, and diabetic nephropathy. In fact, dapagliflozin reduced symptoms, hospitalization, and mortality in patients with reduced ejection fraction who did not have diabetes.[66] Dapagliflozin also reduced progression of renal disease and MACE in a cohort of patients with proteinuria and reduced eGFR.[67] The results of these studies have led to revisions of treatment recommendations from expert panels to favor the use of SGLT2is in patients with established CVD, heart failure, and nephropathy.[68]

α-Glucosidase Inhibitors

α-Glucosidase inhibitors (AGIs) reduce intestinal absorption of starch, dextrin, and disaccharides by inhibiting the action of α-glucosidase in the intestinal brush border. These drugs also increase the release of the glucoregulatory hormone GLP-1 into the circulation, which may contribute to their glucose-lowering effects. The AGIs available in the United States are acarbose and miglitol. AGIs slow the rise of blood glucose by delaying absorption and are indicated as adjuncts to diet and exercise in patients with type 2 diabetes. They can also be used in combination with other oral antidiabetic agents and/or insulin. AGIs do not cause weight gain or have significant effects on plasma lipids. Acarbose reduces the conversion to type 2 diabetes of high-risk persons with impaired glucose tolerance.[69]

Both acarbose and miglitol are given before meals, e.g., two to three times daily. The most prominent adverse effects are malabsorption, flatulence, diarrhea, and abdominal bloating due to the passage of undigested carbohydrate to the lower intestine. Hypoglycemia has been described when AGIs are added to insulin or an insulin secretagogue.

INSULIN THERAPY IN TYPE 2 DIABETES

For the majority of people with type 2 diabetes, the disease is progressive, with a need for increasing intensity of therapy over time.[70] This natural history is thought to be the result of a gradual loss of β cell function and steady decline in insulin production. As other medications lose efficacy with the progression of diabetes, insulin becomes the treatment of last resort, an agent that can replace and compensate for the loss of endogenous insulin production.[71]

Over the last 30 years there have been several major advances in insulin chemistry and formulation that have greatly expanded the options for glycemic management. Porcine and bovine insulins, the standard of care for the first seven decades of insulin use, were replaced by human insulins, regular and neutral protamine Hagedorn (NPH), made by recombinant genetic methods. Subsequently, a number of insulin analogs were developed,[72] each with distinct physicochemical properties that altered their pharmacokinetics and provided a range of onset and duration of activities.

Human regular insulin. Native, or regular, insulin molecules associate as hexamers in aqueous solution at a neutral pH, and this aggregation slows absorption following subcutaneous injection. Regular insulin is generally used to promote glucose control during meals but is slower than the rapid-acting analogs (Table 40.5). The optimal use of regular insulin for prandial coverage is to inject 30 minutes before a meal.

Human NPH insulin. NPH insulin is a suspension of native insulin complexed with zinc and protamine, causing a cloudy solution that contrasts with the clear appearance of other insulins. NPH insulin dissolves more gradually when injected subcutaneously, and thus its duration of action is prolonged (Table 40.5). NPH insulin is usually given either once a day (at bedtime) or twice a day in combination with a prandial insulin before breakfast and supper.

Short-Acting Insulin Analogs

There are three of these insulins available, insulin lispro, insulin aspart, and insulin glulisine, which have similar characteristics and clinical efficacy (Table 40.5). More recently, somewhat faster analogs of insulin aspart and lispro have been released, but the clinical impact of this is unclear. Each has one or more amino acid substitutions that limit the natural aggregation into hexamers. As a result, all are absorbed more rapidly from subcutaneous sites than regular insulin and can be injected approximately 10 minutes before a meal.

Long-Acting Insulin Analogs (Table 40.5)

Insulin glargine is a long-acting analog of human insulin. Three substitutions in the amino acid sequence shift the pH for solubility of glargine to 4.0, which stabilizes the insulin hexamer. When injected into the neutral pH of the subcutaneous space, insulin glargine aggregates, with slow dissolution and steady, predictable absorption from the injection site. Clinical trial data suggest that glargine has a modestly lower risk of nocturnal hypoglycemia compared with NPH insulin. Glargine may be administered at any time during the day with equivalent efficacy.

Insulin detemir is an insulin analog modified by the addition of a saturated fatty acid to the molecule that allows binding to albumin and a prolonged half-life in the circulation.

Insulin degludec is a modified insulin with one amino acid deleted and a long-chain fatty acid bound to the amino acid chain. Degludec forms multihexamers after injection subcutaneously and dissolves slowly.

Concentrated insulins. Several concentrated insulin preparations are currently available. U-500 (500 units/mL) regular insulin is, by definition, five times more concentrated than the standard U-100 (100 units/mL) formulation. Regular U-500 has distinct pharmacokinetics that are closer to NPH insulin, in duration of action, but can still be used as a prandial insulin. U-300 glargine and U-200 degludec are three and two times as concentrated as their U-100 formulations, and a U-200 formulation of insulin lispro is available. These concentrated preparations may provide more consistent absorption and less discomfort for patients who require large amounts of insulin, because the volume of injection is smaller. All concentrated insulins are now available in prefilled pens to minimize the risk of dosing errors.

Premixed insulins. Premixed insulin formulations include a 70/30 mixture of regular and NPH insulin, a 75/25 mixture of lispro and NPH, and a 70/30 mixture of aspart and NPH. These insulins have advantages in convenience because they can deliver prandial and basal insulins with a minimum of injections. The activity profiles of these insulins are the sum of the constituents, with superimposed faster and longer-acting components. These insulins are typically given twice daily before the morning and evening meals.

Insulin Dosing and Regimens. While insulin may be the first drug initiated to treat people with type 2 diabetes, particularly those who present with very high blood glucose, symptoms of glucosuria, and weight loss, for most patients it is added later in the treatment regimen. A common strategy is to add basal insulin to oral agents in patients who need more glucose lowering. Starting doses can be estimated based on body weight (0.1–0.2 units/kg/day) and the degree of hyperglycemia. The principal action of basal insulin is to restrain hepatic glucose production and limit hyperglycemia overnight and between meals, so titration of dose is usually based on the fasting glucose level.

Many individuals with type 2 diabetes will also require doses of insulin before meals. While this typically becomes necessary for

TABLE 40.5 Insulin: Pharmacokinetics of Insulin Preparations

Insulin Type	Approximate Onset of Action	Peak Effect	Approximate Duration of Action[*]
Regular	30–60 min	2–4 hours	5–8 hours
NPH	1–2 hours	2–8 hours[*]	14–24 hours[*][†]
Lispro, lispro-aabc, aspart, faster aspart, glulisine	15–30 min	45–75 min[‡]	3–5 hours[‡]
Glargine	2–4 hours	Minimal	20–24 hours[†]
Detemir	2–3 hours	3–8 hours[†]	8–24 hours[‖]
Degludec	Not applicable	None	>40 hours

[*] May vary considerably

[†] Extremely dose-dependent

[‡] Lispro-aabc and faster aspart generally faster and shorter duration

[‖] Duration is dose-dependent. At higher doses (≥0.8 units/kg), mean duration of action is longer and less variable (22 to 23 hours)

NPH, Neutral protamine Hagedorn.

(Adapted from Hirsch IB, Juneja R, Beals JM, et al. The evolution of insulin and how it informs therapy and treatment choices. *Endo Rev.* 2020;41:733–755.)

patients with long duration of diabetes, it may be needed early in the course of treatment in some, often relatively lean, people whose type 2 diabetes is due to significant loss of β cell function. For the more typical patients with obesity as a comorbidity, insulin resistance is a central aspect of the pathophysiology of type 2 diabetes, and the average dose of insulin required for these patients is often above 1 u/kg. Prandial dosing can be safely started at 10% of the amount of basal insulin at each meal; optimized regimens usually include 40% to 50% of the total daily dose as prandial insulin. Premeal insulin can be titrated to blood glucose levels before the next meal. Doses can be adjusted based on carbohydrate content, or simply meal size. It is also common to add insulin to the prandial dose based on the premeal blood glucose value, a correction dose.

Adverse events. Hypoglycemia is the major risk of insulin use that must be weighed against the benefits of improved glucose control. Risk of hypoglycemia varies widely among persons with type 2 diabetes, with patients who have lower body weights, are more insulin sensitive, or have comorbidities such as renal and hepatic insufficiency or impaired cognition being more susceptible. In patients with higher body mass index and insulin resistance, hypoglycemia is less common. Insulin treatment is associated with modest weight gain, usually 2 to 4 kg. It is not uncommon to see patients who have been poorly controlled, with elevated HbA1c, glucosuria, and weight loss, gain 8 to 10 kg as their blood glucose is brought under control with insulin.

PRACTICAL ASPECTS FOR MANAGING TYPE 2 DIABETES

Glucose lowering is the specific challenge of diabetes management, and the steady expansion of drug classes providing distinct mechanisms to treat hyperglycemia provides a range of options that can be fit to particular patient needs. The first step in glycemic management is setting a target that is appropriate for individual patients, incorporating their capacity and resources for care, health goals, disease duration, established diabetes complications, comorbidities, and specific predilection for drug side effects.[73] There is now abundant evidence that hyperglycemia is the major modifiable risk for microvascular disease and neuropathy.[74] Based on clinical trials, a goal of reducing HbA1c to 7% or less is a reasonable standard for patients in whom microvascular disease prevention is a priority. For this group of generally younger people without established diabetes comorbidities, the benefit of intensifying treatment to hit a lower glycemic target outweighs the burden of drug side effects, medication costs, and the greater patient effort for self-care. For older patients with longer duration of diabetes, established microvascular disease, or serious comorbidities, the risk-to-benefit ratio for intensive treatment may be different. Setting the glycemic goal (usually an HbA1c target) is an important and useful point of discussion with patients and should be revisited over their course of care.

Although the number of medications available to treat diabetes has grown rapidly in recent years, clear evidence as to their optimal use has lagged. There is reasonable evidence on which to base initial therapy and for the stepwise addition of medications to maintain glycemic targets. There is also increasing evidence for the benefits of using some drugs, SGLT2is and GLP-1RAs preferentially, in patients with increased risk for CVD and renal disease. However, there are not yet clear algorithms for glycemic management that can be broadly applied. Each class of drugs has advantages and drawbacks, and the clinical issues for patients with type 2 diabetes are heterogeneous, so treatment must be individualized. Recent statements from expert panels provides some guidance and order to glycemic treatment.[63]

Initial Therapy (Fig. 40.1)

Metformin remains the preferred initial glucose-lowering medication for most people with type 2 diabetes and should be added to lifestyle measures in newly diagnosed patients. This recommendation is based on efficacy, safety, tolerability, low cost, and extensive clinical experience. Results from a substudy of the UK Prospective Diabetes Study showed benefits of initial treatment with metformin on clinical outcomes related to diabetes, with less hypoglycemia and weight gain than with insulin or SUs.[75] In type 2 diabetes patients started on metformin, maintenance of blood glucose in the target range wanes over the course of several years.[53] Thus, it is common for additional medications to be added to metformin as this occurs; all other available medications for glucose lowering have been studied in combination with metformin.

An important consideration when initiating therapy is the approach to patients with severe hyperglycemia (glucose >300 mg/dL; HbA1c >10 %), with symptoms of hyperglycemia, and/or or with evidence of ongoing catabolism.[71] This constellation of symptoms can occur in type 2 diabetes, but raise the possibility of significant insulin deficiency and the possibility that the patient has autoimmune (type 1) or pancreatic diabetes (type 3c). In these patients, using insulin as the initial treatment is often the most prudent approach. Often people who have type 2 diabetes can be transitioned to oral medications once their initial hyperglycemia is treated, but using insulin when some question exists about the diagnosis and physiology of an uncontrolled patient prevents further decompensation.

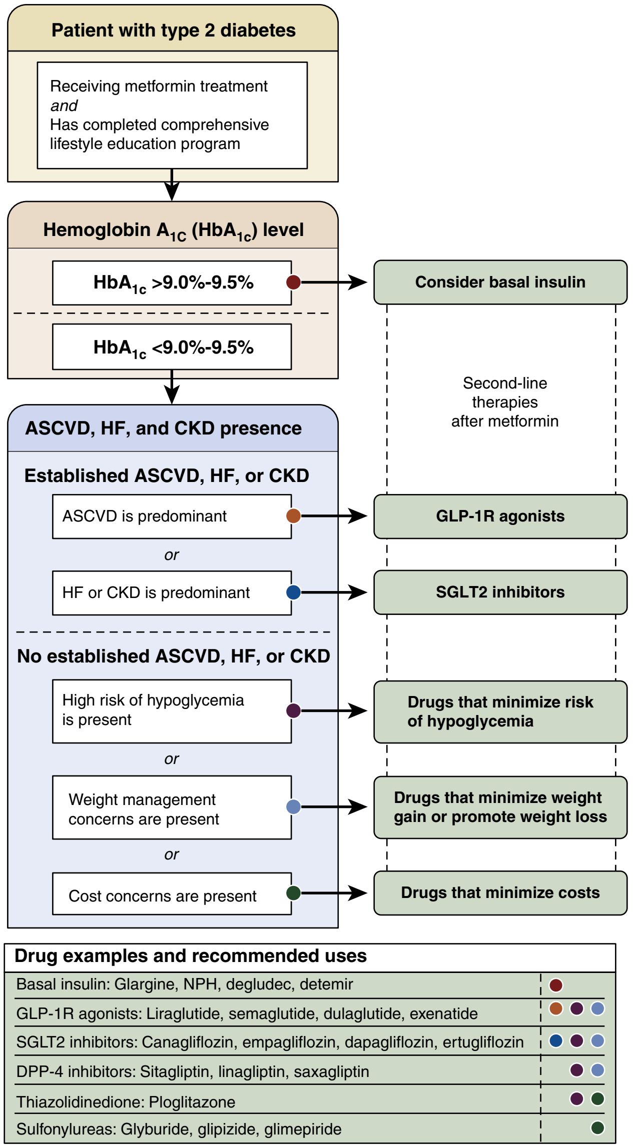
Initial Combination Therapy

The stepwise addition of glucose-lowering medication has been favored over initial combination therapy for most patients. Stepwise therapy (i.e., adding medications to metformin to maintain HbA1c at target) is supported by clinical trials.[45] While there is some support for initial combination therapy due to the greater initial reduction of HbA1c than can be provided by metformin alone, there is only limited evidence that this approach is superior to sequential addition of medications for maintaining glycemic control or slowing the progression of diabetes. However, because the absolute effectiveness of most oral medications rarely exceeds a 1% HbA1c reduction, initial combination therapy may be considered in patients presenting with HbA1c levels more than 1.5% above their target. Potential benefits of combination therapy need to be weighed against the exposure of patients to multiple medications and potential side effects, increased cost, and, in the case of fixed combination medications, less flexibility in dosing.

Addition of Medications for Patients Not Achieving Glycemic Targets

The selection of medication added to metformin will vary among patients based their preferences and clinical characteristics (Fig. 40.1). Important considerations include the presence of established ASCVD and other comorbidities such as heart failure or chronic kidney disease; the risk for specific adverse medication effects, particularly hypoglycemia and weight gain; and safety, tolerability, and cost.

1. *Patients with cardiovascular or renal disease.* Evidence from clinical trials supports the use of SGLT2is and GLP-1RAs as early add-on therapy for people with type 2 diabetes and at high risk for CVD, heart failure, or diabetic nephropathy.[68] GLP-1RAs reduce cardiovascular events in people with established coronary disease or multiple risk factors for coronary artery disease. There is similar data for SGLT2is, and these drugs appear to be particularly effective in people with heart failure and proteinuric renal insufficiency. Both GLP-1RAs and SGLT2is decrease the progression of diabetic kidney disease. The mechanisms by which these two classes of drugs mediate


Patient with type 2 diabetes
Receiving metformin treatment
and
Has completed comprehensive lifestyle education program
Hemoglobin A1C (HbA1c) level
HbA1c >9.0%-9.5%
Consider basal insulin
HbA1c <9.0%-9.5%
Second-line therapies after metformin
ASCVD, HF, and CKD presence
Established ASCVD, HF, or CKD
ASCVD is predominant
GLP-1R agonists
or
HF or CKD is predominant
SGLT2 inhibitors
No established ASCVD, HF, or CKD
High risk of hypoglycemia is present
Drugs that minimize risk of hypoglycemia
or
Weight management concerns are present
Drugs that minimize weight gain or promote weight loss
or
Cost concerns are present
Drugs that minimize costs
Drug examples and recommended uses
Basal insulin: Glargine, NPH, degludec, detemir ●
GLP-1R agonists: Liraglutide, semaglutide, dulaglutide, exenatide ● ● ●
SGLT2 inhibitors: Canagliflozin, empagliflozin, dapagliflozin, ertugliflozin ● ● ●
DPP-4 inhibitors: Sitagliptin, linagliptin, saxagliptin ● ●
Thiazolidinedione: Pioglitazone ● ●
Sulfonylureas: Glyburide, glipizide, glimepiride ●

ASCVD indicates atherosclerotic cardiovascular disease; CKD, chronic kidney disease; DPP-4, dipeptidyl peptidase 4; GLP-1R, glucagon-like peptide 1 receptor; HF, heart failure; NPH, neutral protamine Hagedorn; SGLT2, sodium glucose contransporter 2.

Fig. 40.1 Proposed approach for earlier initiation of insulin therapy for patients with type 2 diabetes.

beneficial effects on diabetic complications are not yet established, but appear to be independent of glucose lowering. There are ongoing debates about whether SGLT2is and GLP-1RAs should be used more broadly or targeted to the patient groups that benefit most, but the effects of these agents on the consequences of macrovascular and microvascular disease adds a new dimension to diabetes care.

2. *Patients at risk for common side effects of glucose-lowering drugs.* SUs and insulin are associated with an increased risk for causing hypoglycemia and are often avoided for persons in whom this is an especially serious complication, i.e., the aged, infirm, those with renal insufficiency, or those with impairment of cognition and mental status. Insulin and SUs, as well as pioglitazone, are associated with weight gain; because it is common for people with type 2 diabetes to struggle with obesity, this side effect is problematic for many. In subjects trying to reduce body weight, the weight loss provided by SGLT2is and GLP-1RAs can be useful.
3. *Cost.* Even in countries with advanced healthcare systems, the cost of drug acquisition is increasingly passed to patients. Thus, drug costs are a major concern for a substantial portion of the population of people with diabetes. Drug costs are especially relevant for diabetic patients because of the many medications they usually use and their generally long duration of treatment. Failure to recognize this issue in practice can lead to patient stress, financial hardship, nonadherence, and poor outcomes. The least expensive glucose-lowering drugs are SUs, metformin, pioglitazone, and human insulins; brand-name agents (GLP-1RAs, SGLT2is, and DPP4is) are considerably more expensive.

Intensification of treatment beyond dual therapy to maintain glycemic targets requires consideration of the impact of medication side effects on comorbidities, as well as the burden of treatment and cost. The lack of a substantial response to one or more noninsulin therapies should raise the issue of adherence and, in those with weight loss, the possibility that the patient has autoimmune (type 1) or insulinopenic diabetes. However, it is common for people with longstanding diabetes to require more than two glucose-lowering agents, often including insulin. Compared with the knowledge base guiding dual therapy of type 2 diabetes, there is even less evidence guiding these choices.[63] In general, intensification of treatment beyond two medications follows the same general principles as the addition of a second medication, with the assumption that the efficacy of third and fourth medications will be somewhat less than additive.

No specific glucose-lowering drug combinations have demonstrated consistent superiority over others, with the exception of insulin plus a GLP-1RA. This combination is particularly potent, and the weight loss effects of the GLP-1RA tend to blunt weight gain from insulin. As more medications are added, there is an increased risk of adverse effects.

While most patients require intensification of glucose-lowering medications, some may require medication reduction or discontinuation, particularly if the therapy is ineffective or is exposing patients to a higher risk of side effects, or when glycemic goals have changed, e.g., due to new or worse comorbidities or even healthy aging. A guiding principle is that, for all therapies, the response should be reviewed at regular intervals, including the impact on efficacy (HbA1c, weight) and safety; the therapy should be stopped or the dose reduced if there are minimal benefits or if harm outweighs any benefit.

KEY POINTS

- Pharmacotherapy for type 2 diabetes has dramatically evolved into a more personalized approach from past eras. The decision to use a particular agent is based on glycemic control, the presence or absence of atherosclerotic cardiovascular disease, heart failure, or renal disease, and concerns about weight, hypoglycemia, and cost.

DIABETES TECHNOLOGY

Diabetes technology continues to rapidly expand. Traditionally, this term has encompassed mechanisms of insulin administration, including continuous subcutaneous insulin infusion (CSII)/pump therapy and devices that monitor blood glucose either by meter or continuous glucose monitoring (CGM). However, as pumps and CGM integrate and digital technology increases, diabetes technology now also includes the software that carries out this integration, programs/applications that help facilitate patient self-management, patient and provider engagement, and even virtual access /direct care through telemedicine. Similar to the concept of personalized HbA1c, diabetes technology should be individualized for each patient. As well, it is important to highlight that diabetes technology is advancing so quickly that the literature is often obsolete as new technology comes to the market and published research lags behind innovation.[76,77]

Self-Monitoring of Blood Glucose

In clinical practice, self-monitoring of blood glucose (SMBG) is considered a key component of self-management and treatment in patients with type 2 diabetes irrespective of insulin use. This is especially true when SMBG is prescribed as part of a DSME/S program. However, based on the evidence, there is controversy about the benefit of SMBG in type 2 diabetes. Structured self-monitoring has been showed to modestly reduce HbA1c only in some studies.[78] In patients not receiving insulin, a recent randomized study showed no benefit with SMBG,[79] but a more current metaanalysis revealed improvement in both HbA1c and body mass index when SMBG frequency was greater than eight times per week.[80] The potential cost and burden to patients with frequent glucose monitoring without clear advantage highlights the need to tailor SMBG to each individual. SMBG testing may be most helpful when altering lifestyle factors such as diet and physical activity. SMBG is also particularly helpful with initiation/titration of medications that can cause hypoglycemia. For example, although studies have not showed benefit for SMBG testing in those taking basal insulin, there is evidence that measuring fasting glucose in order to titrate insulin dosage lowers HbA1c.[81] Overall, SMBG use should be reevaluated at each visit, and patients should be empowered by their clinician to understand SMBG data in order to optimize lifestyle and pharmacologic therapy.

Continuous Subcutaneous Insulin Infusion Systems, "Smart" Pens, and Continuous Glucose Monitoring

CSII delivers hourly rapid-acting insulin (basal insulin) and allows the user to deliver mealtime insulin (bolus dose) according to oral intake and current blood glucose levels. There are now also hybrid/partially closed looped systems that both monitor glucose and deliver insulin based on glucose and predicted glucose. To date, there are no studies assessing these more sophisticated pumps in type 2 diabetes.

V-Go is a simple CSII system that is reapplied every 24 hours and gives prefixed hourly basal insulin the ability to mechanically administer bolus insulin.[82] InPen is an insulin pen with an app that helps patients calculate mealtime insulin doses and records insulin use for providers to review with their patients.[83] Although clinically useful and theoretically promising, there have been no studies looking at glycemic improvement compared with standard insulin delivery systems in patients with either type 1 or type 2 diabetes for either the V-Go or this first "smart" pen.

For type 2 diabetes and CSII, there are only data with older technology. A metaanalysis of patient-level data for participants with suboptimally controlled type 2 diabetes and CSII or multiple daily doses of insulin (MDI) showed a small but statistically significant reduction in HbA1c (by 0.4%) with a more impressive 26% reduction of insulin dose in those on CSII.[84] A more recent metaanalysis failed to show a difference between CSII and MDI and reported that pump therapy was only possibly associated with a lower hypoglycemia risk.[85] Limited use of CSII pump in this population is expected, especially as there is more use of the newer insulin-sparing agents.

Although utilization of pumps is uncommon in type 2 diabetes, CGM use is expanding. CGMs are devices that measure interstitial glucose. CGMs correlate well with plasma glucose and provide up to 288 glucose readings every 24 hours. There are two basic types of CGM devices: retrospective/professional and real-time. Retrospective/professional CGM devices blind patients to blood glucose data. Conversely, real-time CGMs can have automated alarms based both on current and predicted changes in glucose. Flash/intermittent scanned CGM (isCGM) technology is a type of CGM that measures levels continuously, but the patient must chose to scan to reveal glucose levels and their trend.

CGM now has implantable technology where the user can have a small device implanted in their arm. There is also a standard report called the ambulatory glucose profile, and CGM metrics goals have been established[76] (Fig. 40.2).

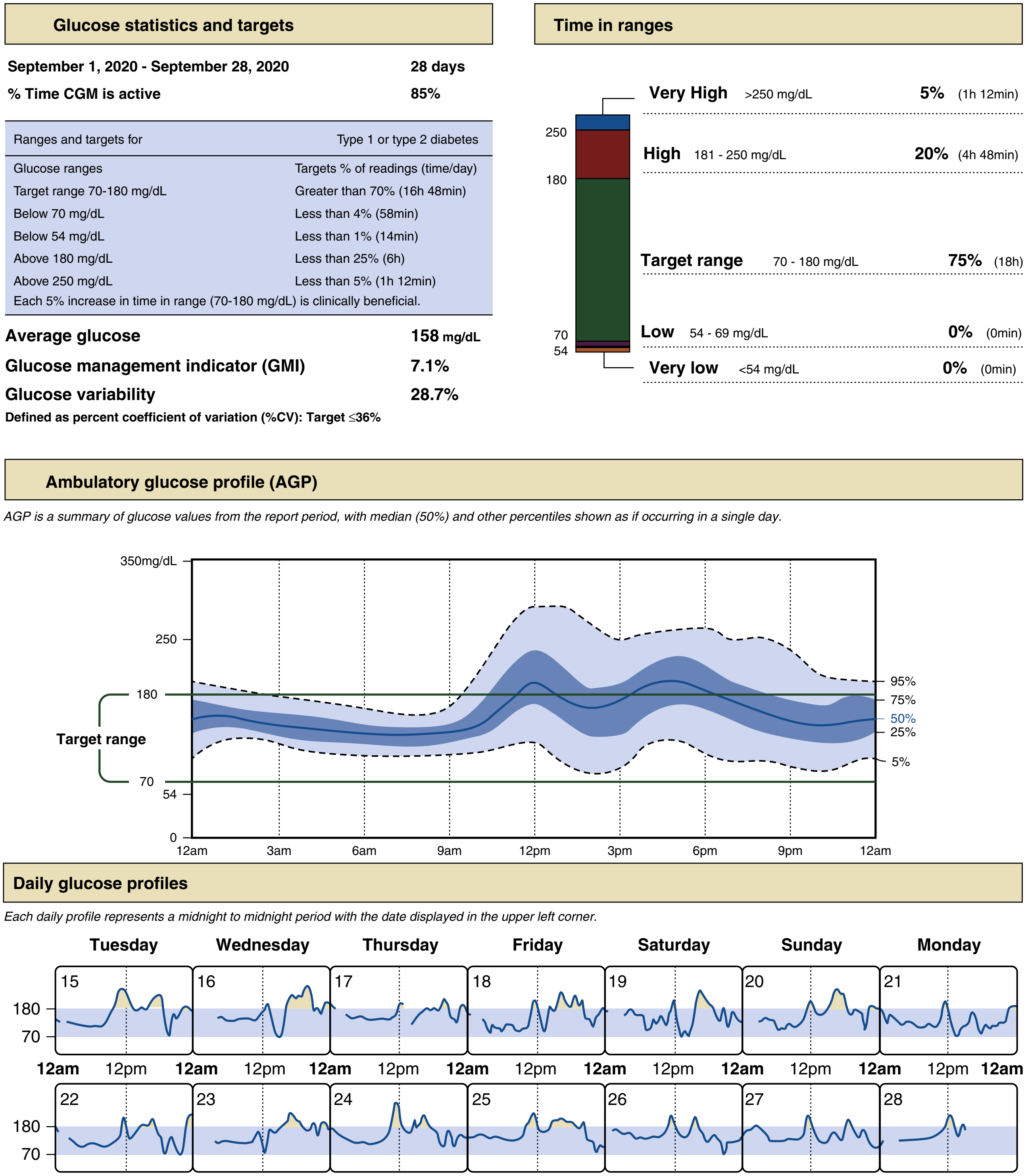
CGM studies in patients with type 2 diabetes are limited. There is only one completed study which used the most updated CGM technology and whose participants were not on MDI. This recent RCT was in an ethnically diverse population only on basal insulin and other insulin-sparing agents and did show a benefit from RT-CGM use. Mean HbA1c level decreased from 9.1% at baseline to 8.0% at 8 months in the CGM group and from 9.0% to 8.4% in the BGM group (adjusted difference, −0.4% [95% CI, −0.8% to −0.1%]; P = .02).[120] The three largest and most recent studies have primarily assessed those receiving intensive insulin therapy with type 2 diabetes. The Multiple Daily Injections and Continuous Glucose Monitoring in Diabetes (DIAMOND) study using real-time CGM showed a statistically significant 0.3% reduction in HbA1c ($P = 0.022$) compared with control, and importantly there were no significant changes in hypoglycemia.[86] Conversely, isCGM demonstrated a decreased duration of hypoglycemia ($P = 0.0014$) but did not result in a significant change in HbA1c.[87] Finally, in the most recent isCGM study in individuals with type 2 diabetes on MDI, there was a 0.49% greater HbA1c reduction in the CGM group ($P = 0.005$), with no significant change in rates of hypoglycemia.[88] Overall, a recent metaanalysis comparing SMBG and CGM showed only marginal significance favoring CGM (−0.24%; $P = 0.05$).[85]

There is also paucity of data in patients using CGM who are not receiving MDI therapy and are only receiving basal insulin and/or oral agents. Two studies that excluded patients receiving prandial insulin did show meaningful reduction in HbA1c levels[89,90]; the larger study used cyclical real-time CGM for 12 weeks (2 weeks on with 2 weeks off) and found a 0.5% greater reduction in HbA1c at 12 weeks ($P = 0.006$).[90] At 52 weeks there was a sustained glycemic effect with no intensification of medical management.[91] There are only a few small studies that have assessed how CGM influences nutritional and physical activity, with some positive findings, but the data are limited.[92]

CGM use in type 2 diabetes continues to expand with rapidly changing technology and aggressive marketing. Hopefully with patients' growing awareness of CGM and continued technology innovation, more type 2 diabetes patients will have access to CGM, and more research will better define the role of this technology for type 2 diabetes.

Digital Health Technology

Digital health technology for diabetes management is also rapidly expanding and continuously evolving. Literature has shown some benefit for everything from automated telephone, web, and internet delivery of education, mobile/digital applications ("apps"), and even videogames, gamification, and virtual environments in the self-management of diabetes.[93] The technology that has grown the most rapidly is probably mobile health apps for diabetes. For example, apps that allow glucose meters to pair with a phone are now common, as are apps that combine meter data with active diabetes and lifestyle coaching. These apps help with self- management and target nutrition, physical activity, glucose monitoring, insulin titration, and delivery.[94,95] A recent survey showed that one third of patients with type 2 diabetes queried were using apps and also found that, after adjusting for confounding factors, diabetes app use increased confidence in diabetes self-management (cumulative self-care score 1.18 [95% confidence interval: 0.26–2.09]).[96] However, across the United States and Europe, mobile apps intended to manage diabetes are largely unregulated. There is inadequate evidence on app accuracy and clinical validity, with



Glucose statistics and targets

September 1, 2020 - September 28, 2020	**28 days**
% Time CGM is active	**85%**

Ranges and targets for	Type 1 or type 2 diabetes
Glucose ranges	Targets % of readings (time/day)
Target range 70-180 mg/dL	Greater than 70% (16h 48min)
Below 70 mg/dL	Less than 4% (58min)
Below 54 mg/dL	Less than 1% (14min)
Above 180 mg/dL	Less than 25% (6h)
Above 250 mg/dL	Less than 5% (1h 12min)

Each 5% increase in time in range (70-180 mg/dL) is clinically beneficial.

Average glucose	**158 mg/dL**
Glucose management indicator (GMI)	**7.1%**
Glucose variability	**28.7%**

Defined as percent coefficient of variation (%CV): Target ≤36%

Time in ranges

Very High	>250 mg/dL	**5%**	(1h 12min)
High	181 - 250 mg/dL	**20%**	(4h 48min)
Target range	70 - 180 mg/dL	**75%**	(18h)
Low	54 - 69 mg/dL	**0%**	(0min)
Very low	<54 mg/dL	**0%**	(0min)

Ambulatory glucose profile (AGP)

AGP is a summary of glucose values from the report period, with median (50%) and other percentiles shown as if occurring in a single day.

Daily glucose profiles

Each daily profile represents a midnight to midnight period with the date displayed in the upper left corner.

250
180
70
54
350mg/dL
250
180
Target range
70
54
0
12am
3am
6am
9am
12pm
3pm
6pm
9pm
12am
95%
75%
50%
25%
5%
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Monday
15
16
17
18
19
20
21
180
70
12am
12pm
12am
12pm
12am
12pm
12am
12pm
12am
12pm
12am
12pm
12am
12pm
12am
22
23
24
25
26
27
28
180
70

Fig. 40.2 Ambulatory glucose profile. (From Battelino T et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the International Consensus on Time in Range. Diabetes Care, 2019 Jun. https://doi.org/10.2337/dci19-0028.)

concern over insufficient data security and poor interoperability and standardization. These issues lead to the ongoing questions regarding how to regulate this digital technology and if clinicians should recommend this potential helpful tool in diabetes management.[94]

Telemedicine:

Telemedicine has a broad definition that encompasses both virtual direct care by specialists (teleconsultation) and remote monitoring and feedback through digital technology (telemonitoring). Telemedicine has demonstrated good patient satisfaction scores, specifically for education and management, and has been shown to reduce cost and improve access to diabetes care.[97] A recent metaanalysis specific to remote monitoring of diabetes and self-care found modest reductions in weight and HbA1c,[98] as did another metaanalysis that included both teleconsultation and telemonitoring, concluding that it was especially effective in older patients with type 2 diabetes and with longer program duration.[99] “Parity” laws in 2016 mandating reimbursement for telemedicine passed in 32 US states and created significant growth in primary care and tele–mental health but not tele–endocrine health.[100] Currently, everything from retinal screening[101] to The Diabetes Prevention Program Classes[102] are available remotely. Until recently, telehealth visits for diabetes were not evolving at the same rate as available technology.[100] Typical care for diabetes has drastically changed secondary to the SARS-Cov-2 virus. Telemedicine for diabetes is now increasing rapidly, given the needs of the patients and reimbursement available to providers.[103-104] In this new “landscape” for healthcare systems, telemedicine will continue to evolve to support diabetes patient care.

KEY POINTS

- Diabetes technology has had tremendous growth for the treatment of type 2 diabetes. While self-monitoring of blood glucose is still considered a key component of self-management and treatment, there is growing interest in the use of continuous glucose monitoring, particularly for those receiving insulin therapy. Digital diabetes health is growing rapidly, assisting the rise in the use of telemedicine.

POPULATION MANAGEMENT

With the rising prevalence of diabetes in the population,[105] the skyrocketing costs of diabetes medications,[106] and the shrinking number of diabetes specialists,[107] a new paradigm for the clinical care of diabetes patients has evolved in the form of population management. Population management is a tool for managing chronic diseases at the health system level that focuses on health outcomes of a group of individuals, including the consideration of how the outcomes are distributed within the group.[108] Population management utilizes the following key components: 1) disease registries that allow patients to be identified proactively, 2) electronic health records utilized for both individual and population care, 3) evidence-based treatment protocols and decision support tools, 4) self-management support, 5) team-based diabetes care, and 6) a focus on patients at risk for poorer outcomes.[108,109]

The concept of population management originated in the 1990s when Ed Wagner and colleagues proposed the “Chronic Care Model,” a model that aimed to improve health outcomes by shifting care processes from acute and reactive to proactive and population-based.[110] In an acute care, or “sick care,” system healthcare providers rely on patients to present to the clinic with symptoms, the provider recommends a treatment, and there is very little emphasis placed on the patient’s overall health, prevention of disease or complications, or continuity of care across providers, because these care aspects are not incentivized for the busy provider. Effective chronic disease management necessitates a different approach, in which individual health care is planned, ongoing, integrated, and evidence-based. In this setting, health promotion and prevention strategies target the entire population, not just the patient who walks into the office, with a focus on higher risk groups. This approach, termed “population management,” requires moving from a reactive to a proactive care delivery system and enables the system to achieve two important goals: improved patient outcomes and cost-effectiveness.

The first central feature of a population management program is a means of identifying patients who need care, by way of a comprehensive disease registry.[111] This is a list of all patients within the health system (at the individual level) with a diagnosis of diabetes, derived from diagnosis codes, prescription data, and/or lab results. The registry can be greatly enhanced with the inclusion of electronic medical record (EMR) data in the registry output such as HbA1c values, last eye exam, blood pressure, smoking status, and statin use. This output can highlight care gaps, and it can also permit sorting and prioritizing to identify patients with care gaps for proactive outreach (e.g., retinal exam is due) or track population-level outcomes (e.g., what proportion of diabetes patients have uncontrolled blood pressure).[112]

Another key component of population management that is enabled by a disease registry and an EMR includes the use of a simple evidence-based treatment protocol that provides primary doctors with evidence-based, guideline-concordant support at the point of care.[111] This protocol can be used to design and implement technology tools, such as best practice alerts (e.g., electronic prompts to the provider that an HbA1c is due), or decision-support tools in the form of clinical calculators that pull data from the EMR (e.g., risk calculators for CVD).

Team-based care is a fourth key feature of a population management program.[111] The multidisciplinary care team may include a primary care physician, medical assistant, nutritionist, nurse, pharmacist, health educator, behavioral medicine specialist, endocrinologist, ophthalmologist, nephrologist, cardiologist, and/or podiatrist. Not only does team-based care allow for providers to work at the top of their licenses, but it also has been shown to improve patient outcomes[113] and reduce costs.[114] Team-based care, in conjunction with technology tools, can also be used to ensure that patients are effectively managed between visits with their primary doctor. For example, remote glucose monitoring systems that provide inbox notifications when patients are having glycemic issues at home can enable midlevel providers to reach out by phone, video, or secure message to intensify treatment and avoid significant glycemic deterioration, rather than waiting 3 to 6 months for a HbA1c result to indicate a problem. This proves a particularly effective strategy for “virtually” helping diabetes patients achieve good outcomes in the setting of a pandemic like COVID-19.

One facet of population care that poses both a challenge and an opportunity is the ability to address health disparities, i.e., to target groups that are at higher risk for disease complications to receive earlier or more aggressive interventions. Diabetes and diabetes complications disproportionately affect vulnerable subgroups, including economically disadvantaged persons and some racial minorities.[115,116] While population care strategies can have dramatic effects on reducing diabetes complications, morbidity, and mortality in a population as a whole, this type of systemic, standardized approach has the potential to leave out those who are not engaged with the healthcare system, leading to increases in racial/ethnic disparities in health outcomes.[117] Thus, the health system needs to strike a careful balance between individual needs and population needs. On the flip side, population care also presents opportunities to reduce health disparities by using systems-level mechanisms to screen for social determinants of health (e.g., potentially vulnerable groups can be identified by demographic, socioeconomic, and language information available in the EMR) and providing this information to care teams to outreach to at-risk subpopulations and provide culturally and linguistically appropriate care[117] in order to both reduce disparities and improve overall quality of care.[118]

One major challenge in implementing a population management strategy in a fee-for-service practice environment, such as the one present in many parts of the United States, is the disconnect between payments for “sick care” management versus those for preventative care or chronic disease management. Implementing effective population management requires that healthcare resources be shifted upstream, to fund registries and EMRs, invest in technology tools, enable team-based care, and support outreach efforts—endeavors that may not be fully reimbursed in the fee-for-service care model.

External incentives have helped to move some systems in the direction of population management,[119] but reimbursement concerns remain a reality for many healthcare providers in fee-for-service practices. Ideally, resources are shifted even further upstream, to support and strengthen the existing public health infrastructure and to address social determinants of health at a societal level. As we think towards the future, we need to ensure that population-based approaches to diabetes care are well-aligned with current healthcare policies and that best practices in diabetes population care are disseminated in ways that increase both the accountability and patient-centeredness of the healthcare system, as well as reducing racial and ethnic disparities in healthcare outcomes.[109]

KEY POINTS

- Population management is becoming an increasingly important tool for the management of type 2 diabetes. Requirements for success includes disease registries, electronic health records, evidence-based treatment protocols and decision support tools, self-management support, team-based diabetes care, and a focus on patients at risk for poorer outcomes

CONCLUSION

While the science of both nonpharmacologic and medications to treat type 2 diabetes and its complications has exploded, the technologies to assist in improving glycemic control are being developed at a similarly impressive pace. Importantly, all of these advances are resulting in better clinical outcomes. The challenge now is to translate this wealth of new information to the millions of people who can benefit, ideally in a cost-effective manner.

REFERENCES

1. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2020*. Atlanta: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2020.
2. International Diabetes Federation. *IDF Diabetes Atlas*. 9th ed. Brussels: International Diabetes Federation; 2019.
3. Powers MA, Bardsley JK, Cypress M, et al. Diabetes self-management education and support in adults with type 2 diabetes: a consensus report of the American Diabetes Association, the Association of Diabetes Care and Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Practice Physicians, the American Academy of Pas, the American Association of Nurse Practitioners, and the American Pharmacists Association. *Diabetes Care*. 2020;43:1636–1649.
4. Steinsbekk A, Rygg LØ, Lisulo M, et al. Group based diabetes self-management education compared to routine treatment for people with type 2 diabetes mellitus. A systematic review with meta-analysis. *BMC Health Serv Res*. 2012;12:213.
5. Coulter A, Entwistle VA, Eccles A, et al. Personalized care planning for adults with chronic or long-term health conditions. *Cochrane Database Syst Rev*. 2015;3:CD010523.
6. Beck J, Greenwood DA, Blanton L, et al. National standards for diabetes self-management education and support. 2017 *Diabetes Spectr*. 2017;30:301–314.
7. American Association of Diabetes Educators. An effective model of diabetes care and education: revising the AADE7 Self-Care Behaviors. *Diabetes Educ*. 2020;46:139–160.
8. Duncan I, Ahmed T, Li QE, et al. Assessing the value of the diabetes educator. *Diabetes Educ*. 2011;37:638–657.
9. Robbins JM, Thatcher GE, Webb DA, et al. Nutritionist visits, diabetes classes, and hospitalization rates and charges: the Urban Diabetes Study. *Diabetes Care*. 2008;31:655–660.
10. Tshiananga JKT, Kocher S, Weber C, et al. The effect of nurse-led diabetes self-management education on glycosylated hemoglobin and cardiovascular risk factors: a meta-analysis. *Diabetes Educ*. 2012;38:108–123.
11. Diabetes Control, Complications Trial Research Group, Nathan DM, Genuth S, Lachin J, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977–986.
12. Stratton IM, Adler AI, Neil HAW, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ*. 2000;321:405–412.
13. Tang TS, Funnell MM, Oh M. Lasting effects of a 2-year diabetes self-management support intervention: outcomes at 1-year follow-up. *Prev Chronic Dis*. 2012;9:E109.
14. Pillay J, Armstrong MJ, Butalia S, et al. Behavioral programs for type 2 diabetes mellitus: a systematic review and network meta-analysis. *Ann Intern Med*. 2015;163:848–860.
15. Institute of Medicine. *The Role of Nutrition in Maintaining Health in the Nation's Elderly: Evaluating Coverage of Nutrition Services for the Medicare Population [Internet]*. Available from: 1999. Accessed 22, August 2020 https://www.nap.edu/catalog/9741/therole-of-nutrition-in-maintaining-health-in-thenations-elderly.
16. Lacey K, Pritchett E. Nutrition Care Process and Model: ADA adopts road map to quality care and outcomes management [published correction appears in *J Am Diet Assoc*. 2003;103:1293]. *J Am Diet Assoc*. 2003;103:1061–1072.
17. Franz MJ, MacLeod J, Evert A, et al. Academy of Nutrition and Dietetics Nutrition practice guideline for type 1 and type 2 diabetes in adults: systematic review of evidence for medical nutrition therapy effectiveness and recommendations for integration into the nutrition care process. *J Acad Nutr Diet*. 2017;117:1659–1679.
18. Academy of Nutrition and Dietetics. Diabetes mellitus types 1 and 2 systematic review and guideline 2015. Available at: http://www.andeal.org/topic.cfm?menu=5305. Accessed August 22, 2020.
19. Evert AB, Dennison M, Gardner C, et al. Nutrition therapy for adults with diabetes and prediabetes: a consensus report. *Diabetes Care*. 2019;42:731–754.
20. U.S. Department of Health and Human Service; U.S. Department of Agriculture. *2015–2020 Dietary Guidelines for Americans* [Internet]. 8th ed. 2015. Available from: https://health.gov/dietaryguidelines/2015/guidelines/. Accessed August 23, 2020.
21. Wadden TA, Neiberg RH, Wing RR, et al. Look AHEAD Research Group. Four-year weight losses in the Look AHEAD study: factors associated with long-term success. *Obesity*. 2011;19:1987–1998.
22. Franz MJ, Boucher JL, Rutten-Ramos S, et al. Lifestyle weight-loss intervention outcomes in overweight and obese adults with type 2 diabetes: a systematic review and meta-analysis of randomized clinical trials. *J Acad Nutr Diet*. 2015;115:1447–1463.
23. McLeod J, Franz MJ, Handu D, et al. Academy of Nutrition and Dietetics nutrition practice guideline for type 1 and type 2 diabetes in adults: nutrition intervention evidence reviews and recommendations. *J Acad Nutr Diet*. 2017;117:1637–1658.
24. Briggs Early K, Stanley K. Position of the Academy of Nutrition and Dietetics: the role of medical nutrition therapy and registered dietitian nutritionists in the prevention and treatment of prediabetes and type 2 diabetes. *J Acad Nutr Diet*. 2018;118:343–353.
25. Dickinson JK, Guzman SJ, Maryniuk MD, et al. The use of language in diabetes care and education. *Diabetes Care*. 2017;40:1790–1799.
26. Colberg SR, Sigal RJ, Yardley JE, et al. Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. *Diabetes Care*. 2016;39:2065–2079.
27. French DN, Kraemer WJ, Volek JS, et al. Anticipatory responses of catecholamines on muscle force production. *J Appl Physiol*. 2007;102:94–102.
28. Turner D, Gray BJ, Luzio S, et al. Similar magnitude of post-exercise hyperglycemia despite manipulating resistance exercise intensity in type 1 diabetes individuals. *Scand J Med Sci Sports*. 2016;26:404–412.
29. Loh R, Stamatakis E, Folkerts D, et al. Effects of interrupting prolonged sitting with physical activity breaks on blood glucose, insulin and tria-

cylglycerol measures: a systematic review and meta-analysis. *Sports Med*. 2020;50:295–330.

30. Albright A, Franz M, Hornsby G, et al. American College of Sports Medicine position stand. Exercise and type 2 diabetes. *Med Sci Sports Exerc*. 2000;32:1345–1360.
31. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43:1334–1359.
32. Brown EC, Franklin BA, Regensteiner JG, et al. Effects of single bout resistance exercise on glucose levels, insulin action, and cardiovascular risk in type 2 diabetes: a narrative review. *J Diabetes Complications*. 2020;34:107610.
33. Greenway FL, Wing RR, Bolin P, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013;369:145–154.
34. Gregg EW, Lin J, Bardenheier B, et al. Impact of intensive lifestyle intervention on disability-free life Expectancy: the Look AHEAD study. *Diabetes Care*. 2018;41:1040–1048.
35. Carmichael OT, Neiberg RH, Dutton GR, et al. Long-term change in physiological markers and cognitive performance in type 2 diabetes: The Look AHEAD Study. *J Clin Endocrinol Metab*. 2020;105:e4778–e4791.
36. Look AHEAD Research Group, Gregg EW, Jakicic JM, et al. Association of the magnitude of weight loss and changes in physical fitness with long-term cardiovascular disease outcomes in overweight or obese people with type 2 diabetes: a post-hoc analysis of the Look AHEAD randomised clinical trial. *Lancet Diabetes Endocrinol*. 2016;4:913–921.
37. de Vries TI, Dorresteijn JAN, van der Graaf Y, et al. Heterogeneity of treatment effects from an intensive lifestyle weight loss intervention on cardiovascular events in patients with type 2 diabetes: data from the Look AHEAD Trial. *Diabetes Care*. 2019;42:1988–1994.
38. Bricca A, Harris LK, Jäger M, et al. Benefits and harms of exercise therapy in people with multimorbidity: a systematic review and meta-analysis of randomised controlled trials. *Ageing Res Rev*. 2020;63:101166.
39. Munan M, Oliveira CLP, Marcotte-Chénard A, et al. Acute and chronic effects of exercise on continuous glucose monitoring outcomes in type 2 diabetes: a meta-analysis. *Front Endocrinol*. 2020;11:495.
40. Winding KM, Munch GW, Iepsen UW, et al. The effect on glycaemic control of low-volume high-intensity interval training versus endurance training in individuals with type 2 diabetes. *Diabetes Obes Metab*. 2018;20:1131–1139.
41. Martland R, Mondelli V, Gaughran F, et al. Can high-intensity interval training improve physical and mental health outcomes? A meta-review of 33 systematic reviews across the lifespan. *J Sports Sci*. 2020;38:430–469.
42. Savikj M, Gabriel BM, Alm PS, et al. Afternoon exercise is more efficacious than morning exercise at improving blood glucose levels in individuals with type 2 diabetes: a randomised crossover trial. *Diabetologia*. 2019;62:233–237.
43. Look AHEAD Research Group (Appendix A), Greenway FL. Severe hypoglycemia in the Look AHEAD trial. *J Diabetes Complications*. 2016;30:935–943.
44. Lipska KJ, Yao X, Herrin J, et al. Trends in drug utilization, glycemic control, and rates of severe hypoglycemia, 2006-2013. *Diabetes Care*. 2017;40:468–475.
45. American Diabetes Association. 9. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S98–S110.
46. Holman RR, Paul SK, Bethel MA, et al. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med*. 2008;359:1565–1576.
47. Maruthur NM, Tseng E, Hutfless S, et al. Diabetes medications as monotherapy or metformin-based combination therapy for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med*. 2016;164:740–751.
48. Budd J, Cusi K. Nonalcoholic fatty liver disease: what does the primary care physician need to know? *Am J Med*. 2020;133:536–543.
49. DeFronzo FL, Mudaliar S, Ratner RE, et al. Pioglitazone for diabetes prevention in impaired glucose tolerance. *N Engl J Med*. 2011;364:1104–1115.
50. Kernan WN, Viscoli CM, Furie KL, et al. Pioglitazone after ischemic stroke or transient ischemic attack. *N Engl J Med*. 2016;374:1321–1331.
51. Lewis JD, Habel LA, Quesenberry CP, et al. Pioglitazone use and risk of bladder cancer and other common cancers in persons with diabetes. *J Am Med Assoc*. 2015;314:265–277.
52. Levin D, Bell S, Sund R, et al. Pioglitazone and bladder cancer risk: a multipopulation pooled, cumulative exposure analysis. *Diabetologia*. 2015;58:493–504.
53. Kahn SE, Haffner SM, Heise MA, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med*. 2006;355:2427–2443.
54. Khunti K, Chatterjee S, Gerstein HC, et al. Do sulphonylureas still have a place in clinical practice? *Lancet Diabetes Endocrinol*. 2018;6:821–832.
55. Rosenstock J, Kahn SE, Johansen OE, et al. Effect of linagliptin vs glimepiride on major adverse cardiovascular outcomes in patients with type 2 diabetes: the CAROLINA randomized clinical trial. *J Am Med Assoc*. 2019;322:1155–1166.
56. Vaccaro O, Masulli M, Nicolucci A, et al. Effects on the incidence of cardiovascular events of the addition of pioglitazone versus sulfonylureas in patients with type 2 diabetes inadequately controlled with metformin (TOSCA.IT): a randomised, multicentre trial. *Lancet Diabetes Endocrinol*. 2017;5:887–897.
57. Drucker DJ. Mechanisms of action and therapeutic application of glucagon-like peptide-1. *Cell Metab*. 2018;27:740–756.
58. Nauck MA, Meier JJ, Cavender MA, et al. Cardiovascular actions and clinical outcomes with glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. *Circulation*. 2017;136:849–870.
59. Drucker DJ. The ascending GLP-1 road from clinical safety to reduction of cardiovascular complications. *Diabetes*. 2018;67:1710–1719.
60. Storgaard H, Cold F, Gluud LL, et al. Glucagon-like peptide-1 receptor agonists and risk of acute pancreatitis in patients with type 2 diabetes. *Diabetes Obes Metab*. 2017;19:906–908.
61. Nauck MA, Meier JJ. Management of endocrine disease: are all GLP-1 agonists equal in the treatment of type 2 diabetes? *Eur J Endocrinol*. 2019;181:R211–R234.
62. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016;375:31–322.
63. Davies MJ, D'Alessio DA, Fradkin J, et al. A consensus report by the American diabetes association (ADA) and the European association for the study of diabetes (EASD). *Diabetes Care*. 2018;41:2669–2701.
64. Mosenzon O, Bain SC, Heerspink HJL, et al. Cardiovascular and renal outcomes by baseline albuminuria status and renal function: results from the LEADER randomized trial. *Diabetes Obes Metab*. 2020;22:2077–2088.
65. Zelniker TA, Wiviott SD, Raz I, et al. Comparison of the effects of glucagon-like peptide receptor agonists and sodium-glucose cotransporter 2 inhibitors for prevention of major adverse cardiovascular and renal outcomes in type 2 diabetes mellitus. *Circulation*. 2019;139:2022–2031.
66. McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381:1995–2008.
67. Mahaffey KW, Jardine MJ, Bompoint S, et al. Canagliflozin and cardiovascular and renal outcomes in type 2 diabetes mellitus and chronic kidney disease in primary and secondary cardiovascular prevention groups. *Circulation*. 2019;140:739–750.
68. Buse JB, Wexler DJ, Tsapas A, et al. 2019 update to: management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia*. 2020;63:221–228. Erratum in: Diabetologia. 2020;63:1667.
69. Chiasson JL, Josse RG, Gomis R, et al. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *Lancet*. 2002;359:2072–2077.
70. U.K. prospective. Diabetes study 16. Overview of 6 years' therapy of type II diabetes: a progressive disease. U.K. Prospective Diabetes Study Group. *Diabetes*. 1995;44:1249–1258. Erratum in: Diabetes. 1996;45:1655.
71. Hirsch IB, Gaudiani LM. Using insulin to treat poorly controlled type 2 diabetes in 2020. *J Am Med Assoc*. 2020;323:2419–2420.

72. Heise T, Mathieu C. Impact of the mode of protraction of basal insulin therapies on their pharmacokinetic and pharmacodynamic properties and resulting clinical outcomes. *Diabetes Obes Metab*. 2017;19:3–12.
73. Buse JB, Wexler DJ, Tsapas A, et al. Update to management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association and European Association for the Study of Diabetes. *Diabetes Care*. 2020;43:487–493.
74. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2019. *Diabetes Care*. 2019;42:S61–S70.
75. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*. 1998;352:854–865. Erratum in: Lancet. 1998;352:1558.
76. Diabetes Technology. Standards of medical care in diabetes—2020 American diabetes association. *Diabetes Care*. 2020;43:S77–S88.
77. Peters AL, Ahmann AJ, Battelino T, et al. Diabetes technology-continuous subcutaneous insulin infusion therapy and continuous glucose monitoring in adults: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:3922–3937.
78. Polansky WH, Fisher L, Schikman CH, et al. Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. *Diabetes Care*. 2011;34:262–267.
79. Young LA, Buse JB, Weaver MA, et al. Glucose self-monitoring in non-insulin-treated patients with type 2 diabetes in primary care settings: a randomized trial. *JAMA Intern Med*. 2017;177:920–929.
80. Xu Y, Tan DHY, Lee JY. Evaluating the impact of self-monitoring of blood glucose frequencies on glucose control in patients with type 2 diabetes who do not use insulin: a systematic review and meta-analysis. *Int J Clin Pract*. 2019;73:e1335.
81. Garber AJ Treat-to-target trials: uses, interpretation and review of concepts. *Diabetes Obes Metab*. 2014;16:193–205.
82. V-Go. Available at: https://www.go-vgo.com. Last accessed September 6, 2020.
83. Inpen. Available at: https://www.companionmedical.com/why-inpen./ Last accessed September 6, 2020.
84. Pickup JC, Reznik Y, Sutton AJ. Glycemic control during continuous subcutaneous insulin infusion versus multiple daily insulin injections in type 2 diabetes: individual patient data meta-analysis and meta-regression of randomized controlled trials. *Diabetes Care*. 2017;40:715–722.
85. Dicembrini I, Mannucci E, Monami M, et al. Impact of technology on glycaemic control in type 2 diabetes: a meta-analysis of randomized trials on continuous glucose monitoring and continuous subcutaneous insulin infusion. *Diabetes Obes Metab*. 2019;21:2619–2625.
86. Beck RW, Riddlesworth TD, Ruedy K, et al. Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. *Ann Intern Med*. 2017;167:365–374.
87. Haak T, Hanaire H, Ajjan R, et al. Flash glucose-sensing technology as a replacement for blood glucose monitoring for the management of insulin-treated type 2 diabetes: a multicenter, open-label randomized controlled trial. *Diabetes Ther*. 2017;8:55–73.
88. Yaron M, Roitman E, Aharon-Hananel G, et al. Effect of flash glucose monitoring technology on glycemic control and treatment satisfaction in patients with type 2 diabetes. *Diabetes Care*. 2019;42:1178–1184.
89. Yoo HJ, An HG, Park SY, et al. Use of a real time continuous glucose monitoring system as a motivational device for poorly controlled type 2 diabetes. *Diabetes Res Clin Pract*. 2008;82:73–79.
90. Ehrhardt NM, Chellappa M, Walker MS, et al. The effect of real-time continuous glucose monitoring on glycemic control in patients with type 2 diabetes mellitus. *J Diabetes Sci Technol*. 2011;5:668–675.
91. Vigersky RA, Fonda SJ, Chellappa M, et al. Short and long-term effects of real-time continuous glucose monitoring in patients with type 2 diabetes. *Diabetes Care*. 2012;35:32–38.
92. Ehrhardt N, Al Zaghal E. Behavior modification in prediabetes and diabetes: potential use of real-time continuous glucose monitoring. *J Diabetes Sci Technol*. 2019;13:271–275.
93. Kaufman N, Khurana I. Using digital health technology to prevent and treat diabetes. *Diabetes Technol Ther*. 2016;18:S56–S68.
94. Fleming GA, Petrie JR, Bergenstal RM, et al. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) diabetes technology working group. *Diabetes Care*. 2020;43(1):250–260.
95. Digital Diabetes. Available at: https://research2guidance.com/product/digital-diabetes-care-market-2018-2022-ready-to-take-off/accessed. Last accessed March 1, 2020.
96. Kebede MM, Pischke CR. Popular diabetes apps and the impact of diabetes app use on self-care behaviour: a survey among the digital community of persons with diabetes on social media. *Front Endocrinol*. 2019;10:135.
97. Levin K, Madsen JR, Petersen I, et al. Telemedicine diabetes consultations are cost-effective, and effects on essential diabetes treatment parameters are similar to conventional treatment: 7-year results from the Svendborg Telemedicine Diabetes Project. *J Diabetes Sci Technol*. 2013;7:587–595.
98. Michaud TL, Ern J, Scoggins D, et al. Assessing the impact of telemonitoring-facilitated lifestyle modifications on diabetes outcomes: a systematic review and meta-analysis. *Telemed J e Health*. 2021;27:124–136.
99. Tchero H, Kangambega P, Briatte C, et al. Clinical effectiveness of telemedicine in diabetes mellitus: a meta-analysis of 42 randomized controlled trials. *Telemed J e Health*. 2019;25:569–583.
100. Barnet ML, Ray KN, Souza J. Trends in telemedicine use in a large commercially insured population, 2005-2017. *J Am Med Assoc*. 2018;320:2147–2214.
101. Khan FN, Micheletti JM, Hendrick AM, Pasquel, et al. Current and next generation portable screening devices for diabetic retinopathy. *J Diabetes Sci Technol*. 2016;10:295–300.
102. Vadheim LM, Patch K, Brokaw SM, et al. Telehealth delivery of the diabetes prevention program to rural communities. *Transl Behav Med*. 2017;7:286–291.
103. Caballero AE, Ceriello A, Misra A, et al. COVID-19 in people living with diabetes: an international consensus. *J Diabetes Complications*. 2020;34:107671.
104. COVID-19 Emergency Declaration. Available at: https://www.cms.gov/files/document/summary-covid-19-emergency-declaration-waivers.pdf. Last accessed September 7, 2020.
105. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2020*. Atlanta: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2020.
106. Cefalu WT, Dawes DE, Gavlak G, et al. Insulin access and affordability working group: conclusions and recommendations. *Diabetes Care*. 2018;41:1299–1311.
107. Vigersky RA, Fish L, Hogan P, et al. The clinical endocrinology workforce: current status and future projections of supply and demand. *J Clin Endocrinol Metab*. 2014;99:3112–3121.
108. American Diabetes Association. 1. Improving care and promoting health in populations: standards of medical care in diabetes—2019. *Diabetes Care*. 2019;42:S7–S12.
109. Schmittdiel JA, Gopalan A, Lin MW, et al. Population health management for diabetes: health care system-level approaches for improving quality and addressing disparities. *Curr Diab Rep*. 2017;17:31.
110. Coleman K, Austin BT, Brach C, et al. Evidence on the chronic care model in the new millennium. *Health Aff*. 2009;28:75–85.
111. Rana JS, Karter AJ, Liu JY, et al. Improved cardiovascular risk factors control associated with a large–scale population management program among diabetes patients. *Am J Med*. 2018;131:661–668.
112. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *J Am Med Assoc*. 2002;288:1775–1779.
113. Shojania KG, Ranji SR, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *J Am Med Assoc*. 2006;296:427–434.
114. Joret MO, Osman K, Dean A, et al. Multidisciplinary clinics reduce treatment costs and improve patient outcomes in diabetic foot disease. *J Vasc Surg*. 2019;70:806–814.

115. Agency for Healthcare Research and Quality. *2015. National Healthcare Quality and Disparities Report and 5th Anniversary Update on the National Quality Strategy. AHRQ Pub. No. 16-0015*. Rockville: Agency for Healthcare Research and Quality; 2016.
116. Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am J Med Sci*. 2016;351:366–373.
117. Patel MR, Piette JD, Resnicow K, et al. Social determinants of health, cost-related nonadherence, and cost reducing behaviors among adults with diabetes: findings from the national health interview survey. *Med Care*. 2016;54:796–803.
118. Bartolome RE, Chen A, Handler J, et al. Population care management and team-based approach to reduce racial disparities among African Americans/blacks with hypertension. *Perm J*. 2016;20:53–59.
119. Kaczorowski J, Goldberg O, Mai V. Pay-for-performance incentives for preventive care. *Can Fam Physician*. 2011;57:690–696.
120. Martens T, Beck RW, Bailey R, Ruedy KJ, Calhoun P, et al. Effect of continuous glucose monitoring on glycemic control in patients with type 2 diabetes treated with basal insulin: a randomized clinical trial. *JAMA*. 2021;325(22):2262–2272.

41

Pancreas and Islet Transplantation

Melena D. Bellin, Michael R. Rickels, and R. Paul Robertson

OUTLINE

Introduction, 710
  Historical Perspective, 710
Pancreas Transplantation, 711
  Patient Selection for Pancreas Transplantation, 711
  Surgical Approaches to Pancreas Transplantation, 711
  Patient and Graft Survival After Pancreas Transplant, 713
  Metabolic Outcomes, 713
Allogeneic Islet (Alloislet) Transplantation, 714
  Patient Selection for Alloislet Transplantation, 714
  Surgical Approaches to Alloislet Transplantation, 715
  Patient and Graft Survival After Alloislet Transplant, 715
  Metabolic Outcomes, 716
Immunosuppressive Drugs, 717
Autologous Islet (Autoislet) Transplantation, 719
  Patient Selection for Autoislet Transplantation, 719
  Surgical Approaches to Autoislet Transplantation, 720
  Patient and Graft Survival After Autoislet Transplant, 720
  Metabolic Outcomes, 722
Effect of Successful Pancreas or Islet Transplantation on Complications of Diabetes, 723
Future Directions and Conclusions, 723

INTRODUCTION

Allogeneic whole organ pancreas and isolated islet transplantation are established approaches to β cell replacement therapy for the treatment of insulin-dependent forms of diabetes that restore normoglycemia with minimal glucose variability, may reverse microvascular complications, and protect against life-threatening metabolic disturbances.[1-6] Because β cell replacement therapy currently requires a deceased pancreas donor (for pancreas transplant or islet isolation) and immunosuppression, transplantation is limited to selected high-risk patients with diabetes, usually with recurrent severe hypoglycemia and hypoglycemia unawareness or with renal failure requiring kidney transplantation.[2,7-9] Most pancreas and islet transplant recipients have type 1 diabetes mellitus (T1DM), but approximately 10% to 15% of simultaneous pancreas-kidney (SPK) transplants and an increasing number of pancreas-after-kidney transplants are performed in patients with type 2 diabetes mellitus (T2DM), and rarely pancreas or islet transplantation is performed for other forms of diabetes, including post-pancreatectomy or cystic fibrosis–related diabetes.[2,10] Development of new renewable cell sources and advances in immunoprotection of islets are promising avenues to bring islet replacement therapy to a larger group of patients.[7,8,10]

Unlike pancreas and allogeneic islet transplantation, *auto*logous islet transplantation is performed to preserve a patient's own islets when performing total pancreatectomy to treat exocrine pancreatic disease (usually chronic or recurrent acute pancreatitis).[11] Most patients who undergo this procedure do not have diabetes before pancreatectomy and autoislet transplantation. Immunosuppression is not required because there is no risk for alloimmune rejection, or recurrent autoimmunity. In addition to being an important therapy to minimize or prevent diabetes in patients who require total pancreatectomy, autoislet transplantation has served as a unique model for islet transplantation in the absence of immunity and provided first proof-of-concept to support work in islet allotransplantation.[12]

KEY POINTS

- Allogeneic pancreas and islet transplantation can restore endogenous β cell function for a select group of patients with diabetes mellitus (most often type 1 diabetes) complicated by renal failure or recurrent severe hypoglycemia. Autologous islet transplantation is performed in patients without diabetes undergoing total pancreatectomy to treat chronic pancreatitis; because patients receive their own islets, no immunosuppression is needed for autotransplantation.

Historical Perspective

The first human pancreas transplant was performed by Dr. Kelly and Dr. Lillehei in 1966 at the University of Minnesota, as a duct-ligated segmental pancreas graft along with a simultaneous kidney transplant.[13,14] This first procedure resulted in only transient insulin production and was followed shortly thereafter by a second pancreas transplant with only a short period (<2 months) of insulin independence.[14] Thereafter, Dr. Najarian and Dr. Sutherland led the pancreas transplant program and reported a greater number of procedures over the following two decades.[1] Their efforts and those of others led to advances in the technical procedures, including drainage of pancreatic duct secretions, development of improved immunosuppression, cyclosporine, and T-cell depleting induction (antilymphocyte globulin), which led to more robust success. Beginning in the 1980s, utilization of pancreas transplantation as a treatment for T1DM gained greater attention and traction.[1] As of 2019, more than 33,000 pancreas transplant procedures had been performed in the United States, and more than 53,000 worldwide, the majority of these being SPK transplant.[2,15]

Around the same time as whole-organ pancreas transplant was introduced in clinical practice, researchers were exploring the concept of isolating and transplanting pancreatic islets. Isolated islets were shown to be successful in reversing diabetes in rodent models in work by Paul Lacy at Washington University and Clyde Barker at the University of Pennsylvania in the early 1970s.[16,17] Based on experience with large animal models of islet transplantation in the 1970s, the intraportal site was selected as the most promising site for islet infusion.[18] Subsequently, the first isolated human islet transplants were performed in the 1970s, with the first islet allograft to treat type 1 diabetes performed in 1974 and the first islet autograft to treat surgical diabetes (pancreatectomy) in 1977.[19-22] Initial allogeneic islet transplants for T1DM were unsuccessful, largely because of inadequate immunosuppression, β cell toxicity of glucocorticoids, and use of inadequate β cell mass. In contrast, the first patient to receive a total pancreatectomy with autoislet transplant was insulin-independent for 2 years, thus providing proof of concept for the procedure of islet isolation and transplantation.[22]

Subsequent efforts in islet transplant focused on both developing the procedures for isolating viable islets more effectively and tailoring transplant immunosuppression protocols to enhance alloislet graft survival. The semiautomated method of islet isolation was introduced by Camillo Ricordi in 1988, and shortly thereafter, in 1990, Dr. Lacy and Dr. Scharp reported the first case of insulin independence for 13 months in an islet allograft recipient.[23,24] Prolonged insulin independence remained rare (~10% of islet allograft cases) until 2000, when James Shapiro and colleagues introduced the Edmonton protocol, which restored insulin independence in seven of seven recipients by increasing the transplanted islet mass (β cell mass) by using at least two pancreas donors for each recipient and removing corticosteroids from the transplant immunosuppression regimen.[25] Since that time, modification of induction immunosuppression to include T-cell–depleting agents and tumor necrosis factor (TNF)-α inhibitors has allowed for increased ability to restore insulin independence with a single islet graft and better longevity of insulin independence.[4,26]

Access to islet transplantation varies by geographical region. In many areas of Europe, Canada, and Australia this procedure is considered standard of care, whereas in the United States isolated islets can only be administered under an Investigational New Drug protocol. The National Institutes of Health–-sponsored multicenter Clinical Islet Transplant (CIT) consortium was established to conduct multicenter trials to support future US Food and Drug Administration (FDA) licensure of islets as a clinical therapy for selected patients with type 1 diabetes. The resulting CIT-07 trial, published in 2016, demonstrated efficacy of islet allotransplantation to eliminate severe hypoglycemia and reduce glucose variability, while meeting glycated hemoglobin (HbA1c) targets for near-normal glycemic control (≤6.5%), in patients with longstanding type 1 diabetes and hypoglycemic unawareness.[3] The subsequent CIT-06 trial, published in 2020, demonstrated similar efficacy in patients who had a previous kidney transplant and also, importantly, the safety of preserved renal allograft function.[27] In the modern era, from 2000 to 2015, more than 1000 islet allotransplant recipients and more than 2100 islet transplant procedures were reported to the Collaborative Islet Transplant Registry (CITR).[28]

Autoislet transplant is often discussed separately from the alloislet transplant because the patient population, indication for transplant, and postoperative management (no immunosuppression) are different.[29,30] Since the first successful autoislet transplant procedure was reported in 1977, total pancreatectomy with autoislet transplantation slowly gained traction as a therapy for chronic pancreatitis. Early reports in the 1990s established the insulin secretory capacity of autoislets and ability to achieve insulin independence when sufficient numbers of islets could be isolated from the diseased and fibrotic pancreas, and, in the absence of autoimmunity, early autoislets exhibited better durability than alloislet transplants.[31,32] As more data emerged to support the success of total pancreatectomy with autoislet transplant to relieve the unrelenting pain of pancreatitis and restore quality of life, acceptance of this procedure as a treatment for pancreatitis among pancreatologists grew, and the number of centers performing the procedure increased in the United States over the past 10 to 15 years.[11,33,34] Although the number of autoislet transplants performed is not as systematically tracked as pancreas and islet allotransplants, based on data from the CITR and the published literature it is likely at least 1500 autoislet transplants have been performed worldwide with high success rates, with most occurring in the United States.[34]

KEY POINTS

- Pancreas transplantation was first performed in 1966; more than 53,000 procedures have been performed worldwide, with the majority performed as a simultaneous pancreas-kidney transplant. Islet transplants were first performed in the 1970s, with proof of concept for diabetes remission established with the first autoislet recipient in 1977. Refinements in immunosuppression and use of multiple pancreas donors have improved islet allotransplant outcomes in the modern era.

PANCREAS TRANSPLANTATION

Patient Selection for Pancreas Transplantation

Pancreas transplantation is considered for restoration of normoglycemia in patients with insulin-dependent diabetes who have renal failure requiring kidney transplantation or who suffer acute metabolic complications of diabetes, particularly recurrent severe hypoglycemia. The majority of transplant recipients have type 1 diabetes. However, an increasing number of SPK transplant procedures have been performed in those with insulin-dependent T2DM, now comprising approximately 15% of those on the waitlist for pancreas transplantation in the United States.[2] In the United States, the Pancreas Allocation System introduced by the United Network of Organ Sharing in 2014 established a national policy for pancreas allocation that included restricting SPK for T2DM to those with low C-peptide levels (≤2 ng/mL fasting). If C-peptide was greater than 2 ng/mL, the recipient needed to be less than a maximum allowed body mass index (BMI; initially 28 kg/m^2 and later 30 kg/m^2). However, in 2018 to 2019 these qualifying criteria were removed, and selection of candidates with T2DM is now at the discretion of the local transplant center.[35] Rarely, pancreas transplant has been performed for other forms of diabetes. In patients who have a total pancreatectomy alone or with a failed autoislet transplant, pancreas transplant has been performed in those with severe glycemic lability or recurrent severe hypoglycemia and can have a secondary benefit of reversing exocrine insufficiency.[36,37] Patients with insulin-dependent cystic fibrosis–related diabetes who require a liver transplant or lung transplant and thus are already treated with immunosuppression may also be candidates for pancreas transplant.[38]

Surgical Approaches to Pancreas Transplantation

Most pancreas transplants are performed in patients with T1DM who also require kidney transplantation. Over 80% of pancreas transplants are performed as an SPK, and approximately 10% as pancreas-after-kidney transplants.[2] The SPK approach allows for the possibility of restoring normoglycemia and insulin independence with a single surgical procedure.[39] Because these patients already require major surgery and immunosuppression for kidney transplant, and because the addition of a pancreas graft can improve patient and kidney graft survival, the risk-benefit ratio is easy to justify. Although it requires a second transplant surgery, pancreas-after-kidney is considered for patients experiencing poor or unstable glycemic control with prior living related donor kidney transplant, deceased donor kidney without pancreas transplant, or a prior simultaneous pancreas and kidney but with a failed pancreas graft. Only approximately 10% of pancreas transplants are pancreas transplants alone, without simultaneous kidney (or other organ) transplants, usually used for severe glycemic lability and hypoglycemia unawareness and/or recurrent severe hypoglycemia warranting risks of major surgery and immunosuppression.[2,40]

Nearly all pancreata transplanted are procured from deceased donors. Transplant of a segmental pancreas graft (hemipancreas) from a living donor is feasible and was performed in the past, most often for living donor SPK transplant, less often as a solitary hemipancreas for pancreas-after-kidney or hemipancreas transplant alone.[41] Between 1979 and 2008, 155 living donor pancreas transplants were reported to the International Pancreas Transplant Registry, with the vast majority (92%) performed in the United States.[41] An historically important observation came from a twin study in which a diabetic twin received a hemipancreas from her nondiabetic twin. Because they were identical twins, no immunosuppressive drugs were given to the recipient. However, soon after the transplant the recipient again became hyperglycemic. Biopsy of the donated hemipancreas did not show signs of rejection but did demonstrate necrotic β cells. This indicated that the recipient's autoimmune response was activated by her identical twin's β cells, illustrating the autoimmune nature of T1DM.[42] Moreover, anti–T-cell immunosuppression proved highly effective at abrogating β cell autoimmunity. As a living related donor often shares human leukocyte antigens with the recipient, in the early era of pancreas transplantation a living related donor offered an immunologic benefit with lower risk for alloimmune pancreas rejection. However, with current immunosuppression, use of a deceased donor pancreas offers similar pancreas graft survival, and thus it is difficult to justify the risk of hemipancreatectomy for a living donor. Hemipancreatectomy for the donor is associated with surgical risks including splenectomy (4%–20%), pancreatic fluid collection (<5%–11%), and pancreatitis (4%) and metabolic risks associated with reduced pancreatic islet mass.[41,43] Metabolic study of living pancreas donors showed increased fasting glucose and 25% with impaired glucose tolerance at 1 year after hemipancreatectomy, reduced insulin secretion to intravenous glucose or arginine, reduced glucagon secretory response to arginine but preserved response to hypoglycemia, and long-term increased risk for diabetes mellitus, especially with higher BMI.[44-47] Despite refined protocols restricting pancreas donors to nonobese individuals with adequate insulin secretory responses lacking insulin resistance, long-term risk for diabetes mellitus remained relatively high, with one in every four patients developing diabetes by an average of 7 to 9 years after hemipancreatectomy, with one out of 10 on insulin treatment.[43,48]

For current pancreas transplants, the entire pancreas and a small segment of duodenum (for anastomoses to bowel or bladder for drainage of exocrine secretion) is procured from an appropriate ABO-compatible deceased donor without history of diabetes mellitus. Historically, most donor pancreases were procured from "ideal" nonobese, brain dead donors of relatively younger age. Donor selection protocols have evolved over time to accept organs from deceased donors previously considered nonideal, including donors after circulatory death or donors with higher BMI (>30 kg/m^2), thus expanding the donor pool.[49] Emerging data suggest no detrimental outcomes in 1-year graft survival when SPK or pancreas transplant alone were performed with organs from donors after cardiac death versus brain death.[50] In experienced transplant centers with high transplant volumes, more aggressive donor selection protocols do not appear to adversely impact outcomes.[51]

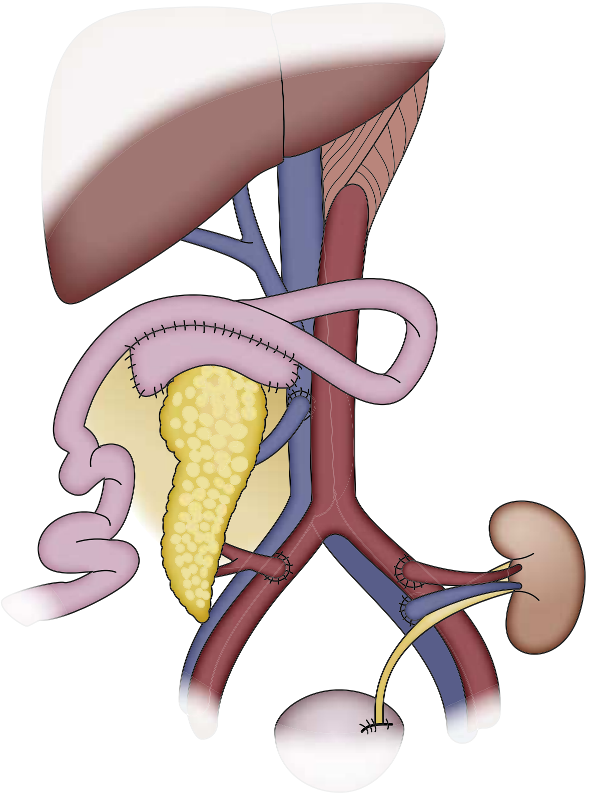


Fig. 41.1 Simultaneous Pancreas-Kidney Transplant. The majority of procedures involve enteric drainage of exocrine secretions from the pancreaticoduodenal graft, arterial reconstruction with a Y-graft of donor ileac arteries anastomosed to the recipient ileac artery, and venous anastomosis of the graft portal vein to the common ileac vein or inferior vena cava (systemic venous drainage) or less commonly the superior mesenteric vein (portal venous drainage). (Reproduced with permission from Larsen JL. Pancreas transplantation: indications and consequences. *Endocr Rev.* 2004;25:919–946.)

The procured pancreas is transplanted in the right ileac fossa of the transplant recipient, with a similar approach for SPK, pancreas-after-kidney, or pancreas transplant alone (Fig. 41.1).[52] Systemic venous drainage is most often established by anastomosing the donor pancreatic vein to the recipient ileac vein or inferior vena cava.[53] Thus, in contrast to the physiology of the native pancreas, in which insulin is secreted into the hepatic portal circulation, after pancreas transplantation insulin from the transplanted pancreas is secreted into the systemic circulation, resulting in mild peripheral hyperinsulinemia.[53-55] An alternate but less common approach is to anastomose the donor portal vein to the recipient superior mesenteric vein to achieve portal venous drainage, but this approach has a greater technical burden and unclear long-term metabolic benefit over systemic drainage for the recipient.[56,57] For management of the exocrine secretions of the pancreas, the donor duodenum can be anastomosed to the bladder (bladder drainage) or anastomosed to the recipient jejunum (enteric drainage) by creating a roux loop from recipient jejunum or by a side-by-side anastomosis to the recipient jejunum.[53,55] Bladder drainage offered the historical advantage of closer monitoring for rejection via measurement of urinary lipase and amylase levels. However, bladder drainage was also complicated

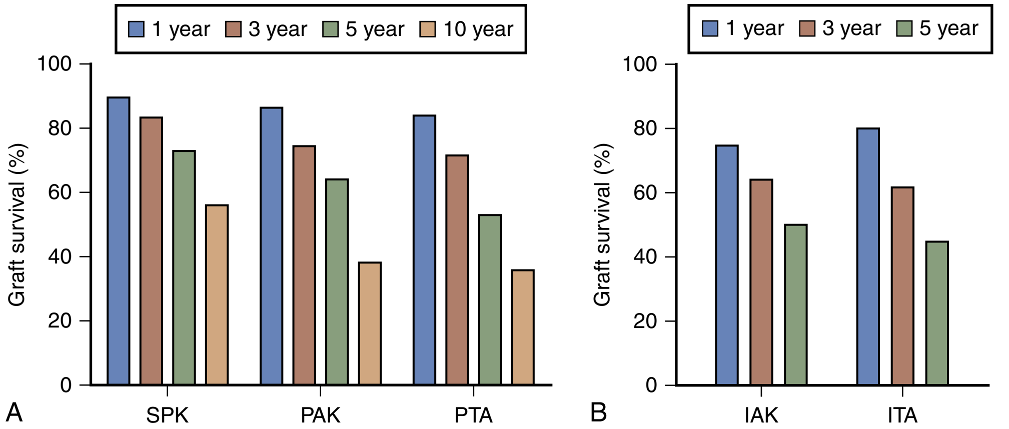


Fig. 41.2 **A,** Pancreas and **B,** islet graft survival defined by registry definitions of insulin independence after pancreas transplant and C-peptide ≥0.3 ng/mL for islet transplant. IAK, islet-after-kidney, ITA, islet transplant alone, PAK, pancreas-after-kidney, PTA, pancreas transplant alone, SPK, simultaneous pancreas-kidney. (Data from Collaborative Islet Transplant Registry: 10th Annual Report. https://citregistry.org/system/files/10th_AR.pdf. Accessed 2/20/2020.)

by metabolic acidosis, hematuria, recurrent urinary tract infections, urethritis, and urethral strictures, making this approach difficult to tolerate for many transplant recipients, and a surgical conversion to enteric drainage was required in 9% to 19% of patients.[1] With improved immunosuppression, and now procedures for radiologic-guided pancreas biopsies, the disadvantages of bladder drainage now outweigh the benefit for most pancreas transplant recipients, and enteric drainage is more often favored (used in more than 90% of recent cases).

Patient and Graft Survival After Pancreas Transplant

Patient survival after pancreas transplant is approximately 97% at 1 year and 65% to 72% at 10 years for SPK, pancreas-after-kidney, or pancreas transplant alone.[58] Pancreas graft survival (as traditionally defined by insulin independence) is highest with a SPK transplant, at 90% at 1 year and 56% at 10 years; in pancreas-after-kidney or pancreas transplant alone, pancreas graft survival is around 84% to 87% at 1 year and decreases to 36% to 38% at 10 years (Fig. 41.2).[58,59] Survival of the pancreas graft depends on technical and immunologic factors. Early pancreas graft loss may occur secondary to technical complications (most commonly graft thrombosis), pancreatitis, or immunologic rejection, while late pancreas allograft loss is often immunologically mediated, because of alloimmune rejection or autoimmune recurrence.[58,60] Several studies have shown superior pancreas graft survival when procedures are performed at high-volume pancreas transplant centers (defined by >11–14 procedures per year), even despite a tendency of higher volume centers to accept higher risk organs for transplant.[61-63]

Receiving a pancreas transplant with a kidney transplant appears to convey a survival benefit when compared with patients with T1DM who undergo renal transplant alone. In a 30-year follow-up cohort in The Netherlands, SPK transplant reduced mortality by 44% at 10 years (hazard ratio 0.56) and by 31% at 20 years compared with deceased or living donor kidney transplant alone, and these benefits were particularly strengthened by having a functioning pancreas graft.[39] Additional studies from the United States and the United Kingdom similarly found a benefit of SPK and/or pancreas-after-kidney transplant on patient survival, as well as significantly better survival of kidney graft, likely related to the lack of chronic hyperglycemia exposure to the newly transplanted kidney.[64-66]

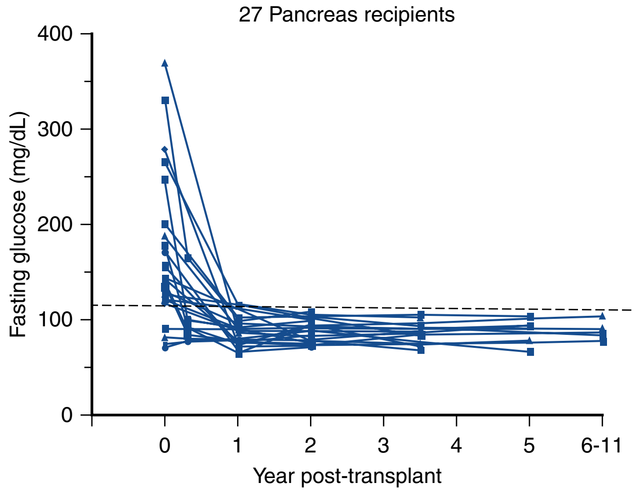


Fig. 41.3 Fasting blood glucose in 27 insulin-independent pancreas transplant recipients demonstrating restoration of normoglycemia for >5 years posttransplant.

Metabolic Outcomes

Successful pancreas transplantation almost invariably results in fasting blood glucose levels that are below 110 mg/dL (Fig. 41.3), and HbA1c levels are well below those achieved by intensive insulin treatment reported in the DCCT trial (Fig. 41.4). Insulin and C-peptide secretion are usually assessed during oral and intravenous glucose tolerance tests. Insulin levels in the basal and glucose-stimulated state after pancreas transplantation are slightly elevated (Fig. 41.5) because the venous return from the transplanted pancreas is emptied into a vein (usually the iliac vein) that leads directly to the inferior vena cava.[53-55] The latter situation is very different from the blood return from the native pancreas, which delivers insulin via the portal vein directly into the liver, where insulin is metabolized at least by 50% in a single passage. The postoperative complication of hypoglycemia rarely occurs after pancreas transplantation, probably because the transplanted pancreas, unlike the native pancreas in people with type 1 diabetes, secretes glucagon normally in response to hypoglycemia, and consequently restores hepatic glucose production during hypoglycemia that is

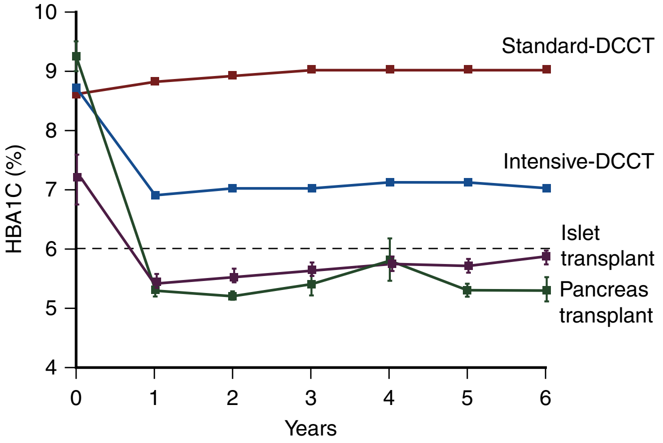


Fig. 41.4 Glycated hemoglobin (*HbA1c*) levels normalize after pancreas (authors' data) and islet transplant (updated data from Rickels MR, Liu C, Shlansky-Goldberg RD, et al. Improvement in β-cell secretory capacity after human islet transplantation according to the CIT07 protocol. *Diabetes*. 2013;62:2890–2897), contrasted with medical insulin therapy in the Diabetes Control and Complications Trial (DCCT).

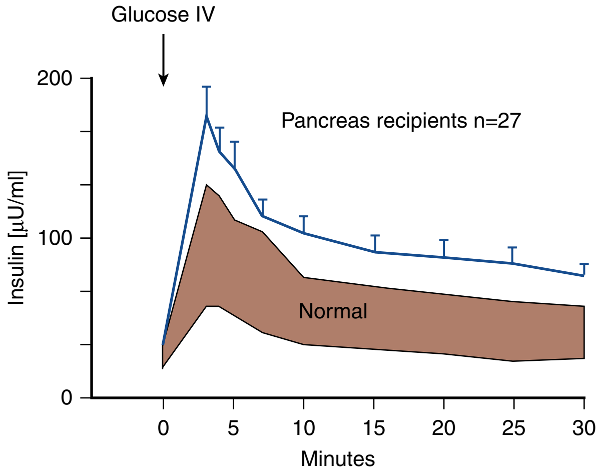


Fig. 41.5 Pancreas transplant recipients exhibit normal restoration of first phase insulin secretion in response to an intravenous glucose bolus, with higher insulin levels versus controls because of the systemic drainage of the transplanted pancreas. *IV,* Intravenous.

necessary to prevent or correct low blood glucose (Fig. 41.6).[67-69] This is of interest, because the intrapelvic transplanted pancreas is unlikely to receive meaningful intracranial sympathetic or parasympathetic nerve inputs, which have been shown to contribute to glucagon release during hypoglycemia in normal, nontransplanted subjects. This speaks to the efficiency of the transplanted pancreas alone, independent of the central nervous system as a provider of glucagon secretion during hypoglycemia.

KEY POINTS

- Pancreas transplantation is performed for type 1 (85%) and type 2 (15%) diabetes. Pancreas graft survival is 85% to 90% at 1 year and 36% to 87% at 10 years, with the highest pancreas survival in simultaneous pancreas-kidney transplant. Because the venous drainage of the pancreas graft is usually systemic, insulin levels are mildly elevated after successful pancreas transplant, but with normal fasting glucose and HbA1c and restoration of glucagon responses to hypoglycemia.

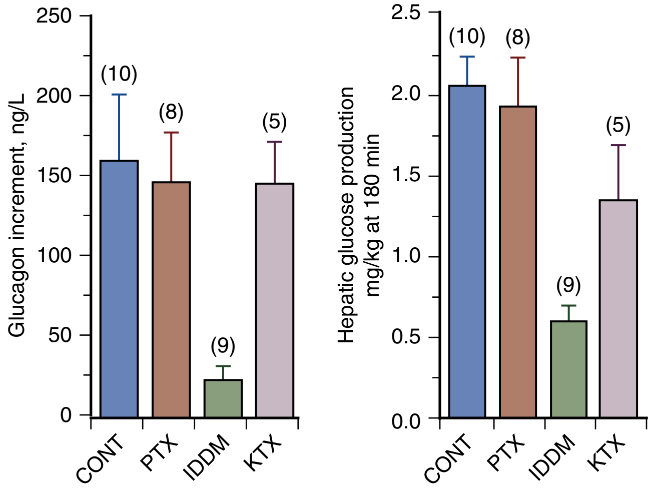


Fig. 41.6 Glucagon and hepatic glucose production during hypoglycemic clamp studies are normalized in pancreas transplant (*PTX*) recipients when compared with controls (*CONT*). In contrast, type 1 diabetes mellitus (T1DM) without transplant show significantly impaired glucagon and hepatic glucose output in response to hypoglycemia. Nondiabetes kidney transplant (*KTX*) are displayed as a transplant control group. *IDDM*, insulin-dependent diabetes mellitus or T1DM. (From Barrou Z, Seaquist ER, Robertson RP. Pancreas transplantation in diabetic humans normalizes hepatic glucose production during hypoglycemia. *Diabetes*. 1994;43(5):661–666.)

ALLOGENEIC ISLET (ALLOISLET) TRANSPLANTATION

Patient Selection for Alloislet Transplantation

The appeal of an islet transplant alone compared with pancreas transplantation alone is that the transplant procedure is relatively minor compared with the major surgery required for a whole pancreas, with very low risk for procedural complications.[10] Similar to pancreas transplants, islet allografts can be performed as islet transplant alone, islet-after-kidney transplant, or simultaneous islet-kidney transplant.[28,70,71] However, in contrast to pancreas transplants, the vast majority of procedures performed to date are islet transplant alone, followed by islet-after-kidney, and seldom simultaneous islet-kidney. In the CITR, which tracks all islet allografts performed in the United States and the many of those performed worldwide, 81% of reported transplants were islet transplant alone, 17% islet-after-kidney, and only 2% simultaneous islet-kidney.[28]

Nearly all islet allotransplants to date have been performed for patients with T1DM, most often for treatment of excessive glycemic lability, hypoglycemia unawareness, and severe hypoglycemia despite optimal medical therapy, or, in the case of islet-after-kidney, inability to maintain adequate glycemic control using exogenous insulin.[3,71-76] In the phase 3 multicenter CIT-07 trial, which was designed to collect clinical data to support licensure of islets as a cellular therapy in the United States, adult patients with C-peptide–negative type 1 diabetes of 5 years or more duration were considered candidates for islet transplant alone if they had impaired awareness of hypoglycemia (Clarke score ≥4) or marked glycemic lability, and a history of severe hypoglycemic episodes in the prior year despite optimal medical therapy.[3] Similar criteria have been used to determine candidacy across other studies and centers, including in the European-based phase 3 multicenter, two-arm, open-label TRIMECO study comparing islet transplant alone to insulin therapy.[70,72,77] For these patients, the risk-benefit ratio of taking immunosuppression is considered reasonable in comparison to the high risks for recurrent life-threatening severe hypoglycemia with medical management with insulin. For patients with a prior history of kidney transplant and T1DM, islet-after-kidney transplant is considered, including in the phase 3 multicenter CIT-06 trial, if patients had impaired hypoglycemia awareness and experienced severe hypoglycemia (similar to islet transplant alone) or if unable to achieve a HbA1c level less than 7.5% despite optimal medical management with insulin therapy.[27] For islet-after-kidney, the relative risks associated with islet allotransplant are less, as the patient is already receiving immunosuppression, and the surgical procedure is minor. The theoretical risk that an islet allotransplant could create an antigenic stimulus that may increase risk for kidney rejection has not been observed in the course of phase 3 investigation.[27] Less commonly, islet allotransplantation (as islet transplant alone) has been performed in patients with type 1 diabetes and early microvascular complications, with the goal of halting progression of microvascular disease.[6,78,79]

Under most protocols, patients with type 1 diabetes who also have high insulin needs (>1 U/kg/day) or high BMI (>30 kg/m^2) are ineligible for islet allotransplantation, as the transplanted islet mass may be insufficient to provide clinical benefit in the setting of insulin resistance.[3,27] In part for this reason, unlike whole organ pancreas transplantation, islet transplant is not currently considered for patients with T2DM. Only rarely has islet transplantation been performed for patients with other types of diabetes mellitus. There are a few case reports from Canada and Europe of islet allotransplantation performed for cystic fibrosis–related diabetes in patients also receiving lung and/or liver transplants or for postpancreatectomy diabetes mellitus.[80-84]

The number of patients with T1DM that can receive islet allotransplants is currently limited by both the risks of immunosuppression and the limited supply of islets. However, the approach of islet transplantation carries the potential to develop renewable sources of cells (discussed later) and for engineering of islets or the transplant site to reduce the need for immunosuppression in the future.

Surgical Approaches to Alloislet Transplantation

Pancreatic islets for transplant are procured from an ABO blood type–compatible deceased donor pancreas. Selection of deceased donors for islet transplantation includes consideration of both organ viability and the likelihood that sufficient numbers of islets will be obtained for transplant. Islet allotransplant is conventionally performed only if greater than 4000 to 5000 islet equivalents per kilogram recipient body weight (IEQ/kg) are obtained from the isolation procedure, as the islet mass transplanted needs to be sufficient to justify the risk of administering immunosuppression, particularly in islet transplant alone.[3,85] The majority of pancreases are obtained from brain dead donors owing to a desire to avoid pancreatic ischemia associated with cardiac death, although some islet transplants from cardiac death donors have been reported.[86,87] In contrast to pancreas transplant, in which high-BMI donors are avoided, in islet transplantation donors with high BMI or high body surface area are desirable, as these organs also yield a greater islet mass.[88] Thus, islet transplants provide opportunity to use donor organs that would otherwise have been discarded for whole-organ pancreas transplant. Islet allotransplant can also use pancreases from older donors that are not accepted for whole-organ pancreas transplant. However, when available, younger donors (age <45 years) provide more functional and viable islets.[89,90]

Once the donor pancreas is procured, the pancreas must be transported to the islet isolation facility. Islet processing is often performed locally (at the same institution performing the transplant), but some countries have developed regional islet isolation networks where islets are isolated at designated centers and then distributed to the transplanting facility to improve patient accessibility to islet transplantation, with similar metabolic outcomes for the recipient.[91-95] A cold ischemia time of less than 8 hours (maximum 12 hours) is necessary to optimize viability of the transplanted islets.[96,97] Allogeneic islets are considered an investigational biologic therapy in the United States, and thus islet manufacturing process are regulated under FDA current Good Manufacturing Practices.

Islet isolation is performed by enzymatic and mechanical digestion of the pancreas. In the islet lab, the pancreatic duct is cannulated, and a collagenase and neutral protease enzyme cocktail is infused intraductally under pressure.[96,98,99] Following completion of enzymatic distention, the pancreas is mechanically disrupted by use of the semiautomated Ricordi method, and the islets are further purified by density centrifugation, targeting a total tissue volume of less than 7.5 to 10 mL.[96,100] Isolated islets are maintained in a culture media for 36 to 72 hours while the transplant recipient begins induction immunosuppression. This culture period may also reduce immunogenicity of the graft by decreasing class I human leukocyte antigen (HLA) expression, passenger leukocytes, and islet production of chemokines and tissue factor before infusion.[101-105]

For the transplant procedure, the islets are infused intraportally into the liver either through direct access of the portal vein via a transhepatic percutaneous approach with interventional radiology guidance (Fig. 41.7) or via infusion in a mesenteric or omental vein accessed by minilaparotomy.[97,106-108] The expression of tissue factor on the surface of the islets to the portal blood triggers an instant blood-mediated inflammatory reaction. This inflammatory reaction can cause a partial loss of transplanted islets, as suggested by preclinical studies and biomarkers in autoislet recipients, and causes risk for thrombosis around the islet graft.[109-113] Portal branch vein thrombosis has been reported in 3% to 4% of transplant recipients historically, but less frequently in the more recent era, and typically resolves with anticoagulation therapy.[114] Heparin is administered with the islets and subsequently by intravenous infusion to minimize this risk for portal vein thrombosis[114] and transitioned to subcutaneous administration for at least the first week posttransplant to prevent microthrombi from affecting islet revascularization. Unlike whole-organ pancreas transplants, islets are devascularized at the time of transplantation and infused into a relatively hypoxic environment, and will require several weeks to engraft.[115] Thus, islet allotransplant recipients are placed back on insulin therapy after the procedure, and insulin is weaned as tolerated over the next few months as islets engraft and become more functional.[3,27]

Patient and Graft Survival After Alloislet Transplant

Patient survival after alloislet transplant is high. Of 1086 alloislet recipients in the CITR, the mortality rate was 3% over a mean follow-up duration of 4.2 years (± 3.5 years standard deviation [SD]), 57% of which occurred in transplant recipients transplanted in the earliest registry era (1999–2002). Over half of deaths occurred in simultaneous islet-kidney transplant recipients, even though this procedure comprised only 2% of all islet transplants. Islet transplant alone recipients had a 98% rate of survival at 5 years posttransplant. This compares to a reported mortality rate of 4% over a mean follow-up duration of 2.7 years (± 2.8 years SD) among individuals with T1DM referred for islet transplantation who are either found to be unsuitable for transplant or remained active on the waiting list.[116]

Islet graft survival is most commonly defined by detectable C-peptide production. Insulin independence and resolution of severe hypoglycemia while meeting HbA1c targets are common endpoints reported in clinical trials of islet transplantation.[3] Data from the CITR suggest improvements in insulin independence and islet function as greater experience and refinement of islet transplant protocols have evolved. Recipients of islet transplants performed in the more recent

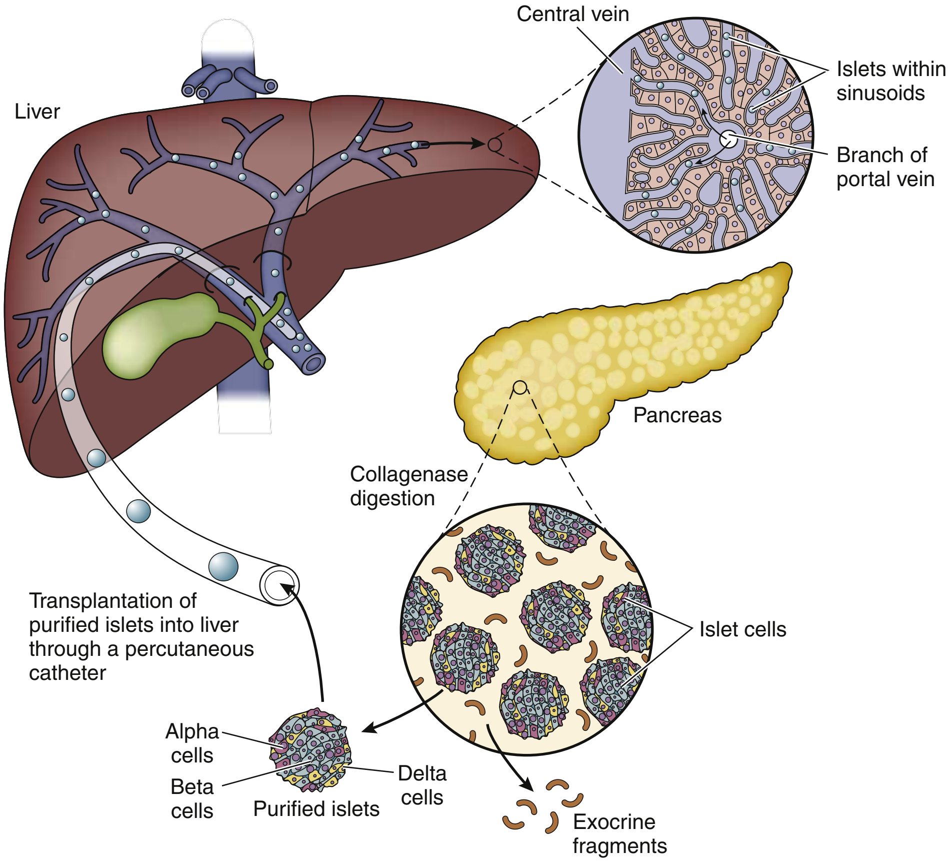


Fig. 41.7 Allogeneic Islet Transplantation. The procedure involves procurement of a deceased donor pancreas for collagenase digestion and centrifuge purification in a current good manufacturing practice islet isolation facility. The final islet product may be cultured for up to 72 hours before intraportal delivery to the main portal vein using percutaneous, transhepatic access, as shown here, or via mesenteric vein access achieved by minilaparotomy. (Reproduced with permission from Robertson RP. Islet transplantation as a treatment for diabetes - a work in progress. *N Engl J Med.* 2004;350:694–705.)

era were insulin-independent in 66% of cases at 1 year, 55% at 2 years, and 44% at 3 years. Nearly 90% were C-peptide–positive through 3 years.[75] When current standard immunosuppression regimens was given, including T-cell–depleting induction therapy and TNF-α inhibition, the 5-year insulin independence rate was 50%, similar to pancreas transplant alone.[4] Recipient characteristics may also affect outcomes, with age 35 years or over associated with more favorable outcomes.[28] More than one pancreas donor may be required to achieve sufficient islet mass for insulin independence. CIT-07, CIT-06, and the CITR data indicate that 28% to 46% of islet transplant alone recipients received a single islet graft, while others received two or more infusions to reach reported metabolic outcomes.[3,27,28,75] However, high rates of prolonged insulin independence can be achieved with a single pancreas donor providing greater than 6000 IEQ/kg under appropriate immunosuppression.[26,117] Achievement of insulin independence is associated with significantly longer islet graft survival,[75] suggesting a long-term benefit to targeting initially high islet engraftment.

Because islet transplant alone is primarily performed to treat type 1 diabetes complicated by hypoglycemia unawareness and recurrent severe hypoglycemic episodes, the primary therapeutic goal of islet transplantation is often to resolve severe hypoglycemia, regardless of whether partial exogenous insulin supplementation is still needed. In the phase 3 CIT-07 trial, the primary endpoint of islet transplant alone was complete resolution of hypoglycemia (days 28–365 posttransplant) with HbA1c less than 7%, an endpoint that was determined in consultation with the FDA. At 1 year, 87.5% met the primary endpoint, and at 2 years the median HbA1c was maintained at 5.6%.[3] Secondary analyses from this trial as well as the phase 3 CIT-06 trial further show a benefit to health-related quality of life, particularly reduced diabetes distress scores, reduced fear of hypoglycemia, and improved visual analog quality of life reported on the EuroQol-5D.[27,118]

Metabolic Outcomes

Successful alloislet transplantation rivals pancreas transplantation in terms of establishing levels of fasting glucose and HbA1c in the normal or nearly normal ranges (Figs. 41.8 and 41.4). Unlike pancreas transplantation, where systemic venous drainage bypasses first-pass hepatic insulin extraction and results in hyperinsulinemia, the intraportal delivery of transplanted islets for intrahepatic engraftment results in normal hepatic insulin extraction,[119] and fasting insulin levels are consequently normal (Fig. 41.9).[120] Like pancreas transplantation, first-phase insulin secretion follows normal secretory dynamics; however, peak insulin secretory responses are often lower than in controls or recipients of whole pancreas transplant (Fig. 41.9). This reduction in first-phase insulin secretion in response to glucose is best explained by a reduced islet β cell mass ultimately surviving engraftment.[117] Consistent with other models of reduced β cell mass, first-phase insulin secretion in response to arginine is preserved in islet transplant recipients, and the reduced functional β cell mass is most accurately measured from glucose potentiation of arginine-induced insulin secretion (GPAIS) that gives the β cell secretory capacity (Fig. 41.10). With a decreased β cell secretory capacity there is a limited insulin secretory reserve, such that

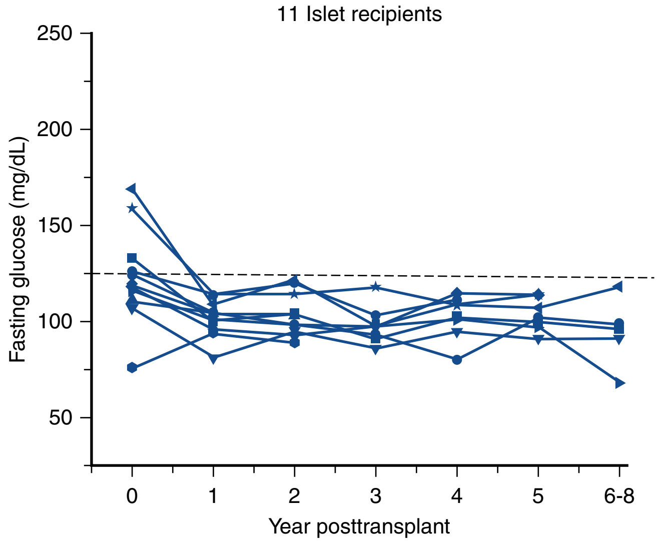


Fig. 41.8 Fasting blood glucose in 11 insulin independent islet transplant recipients demonstrating restoration of near-normoglycemia for >5 years posttransplant. (Adapted from Rickels MR, Liu C, Shlansky-Goldberg RD, et al. Improvement in β-cell secretory capacity after human islet transplantation according to the CIT07 protocol. *Diabetes*. 2013;62(8):2890–2897.)

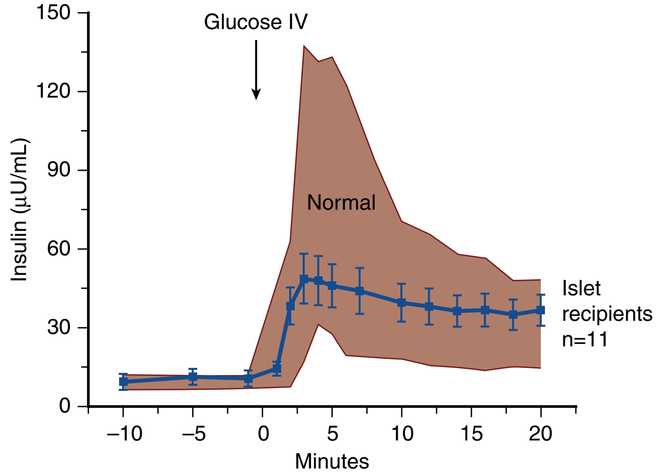


Fig. 41.9 Insulin secretion in response to an intravenous glucose bolus in islet transplant recipients shows a similar time course of first phase insulin secretion, with lower peak than control insulin values because of the reduced β cell secretory capacity of the transplanted islets surviving engraftment in the liver. *IV,* Intravenous. (Modified from data from Rickels MR, Kong SM, Fuller C, et al. Insulin sensitivity index in type 1 diabetes and following human islet transplantation: comparison of the minimal model to euglycemic clamp measures. *Am J Physiol Endocrinol Metab.* 2014;306:E1217–1224.)

even insulin-independent islet transplant recipients may require temporary periods of exogenous insulin administration during periods of intercurrent illness or pregnancy[121] that increase insulin demand. Most importantly, approximately 90% of the incidence of severe hypoglycemia and accompanying hypoglycemia unawareness is eliminated (Fig. 41.11). This is explained by a recovery of appropriate islet cell responses to hypoglycemia, with suppression of endogenous insulin secretion and partially increased glucagon secretion that, together with an improvement in adrenomedullary epinephrine secretion, restores the endogenous (primarily hepatic) glucose production response required to defend against the development of low blood glucose (Fig. 41.12).[10,122-124] This is of utmost importance, because frequent hypoglycemia and symptom unawareness is the major clinical indication for alloislet transplantation.

KEY POINTS

- In islet allotransplantation, islets are isolated from the pancreas by enzymatic and mechanical digestion and infused into the liver via the portal vein using a minor surgical procedure or interventional radiologic approach. Alloislet transplants are nearly exclusively performed for treatment of T1DM, most often for management of hypoglycemia unawareness and life-threatening episodes of severe hypoglycemia. Successful islet transplant restores sufficient insulin and glucagon secretion to maintain normal glucose homeostasis.

IMMUNOSUPPRESSIVE DRUGS

Pancreas and islet transplantation both require immunosuppression to prevent alloimmune rejection and autoimmune recurrence. In general, modern immunosuppressive therapy for either pancreas or islet transplant consists of an induction regimen that is given over a short time frame in the peritransplant period and a maintenance regimen of daily immunosuppression that is taken for as long as the transplanted organ(s) continue to function. Administration of any immunosuppression must balance the need to prevent rejection, while at the same time keeping the risks of immunosuppressive toxicities and infection as minimal as possible.[8] A unique consideration for patients with type 1 diabetes receiving a pancreas or islet transplant is that immunosuppression must also be adequate to prevent recurrence of the β cell autoimmunity of type 1 diabetes. Evidence that transplanted islet β cells are susceptible to destruction by recurrent autoimmunity was first provided by Ali Naji and colleagues using an experimental animal model.[125] As mentioned earlier, in the initial experience with living related donor pancreas transplantation, when hemipancreas transplants were performed using an identical twin or HLA-matched sibling donor in the absence of immunosuppression, recurrence of β cell autoimmunity was observed in the transplanted pancreas.[126] T-cell autoreactivity against β cell antigens has also been associated with functional graft loss in islet transplantation.[127]

Interleukin (IL)-2 receptor antagonists (daclizumab and basiliximab) or T-cell–depleting agents (antithymocyte globulin or alemtuzumab) are given for induction immunosuppression for pancreas or islet transplants, with the T-cell–depleting agents largely gaining favor in the recent era.[1,3,15,75,77] T-cell depletion is advantageous over IL-2 receptor antagonists for protection against acute cellular rejection in pancreas and islet transplants, and antithymocyte globulin may offer the additional benefit of relative expansion of the regulatory T-cell pool.[128-131] For islet transplantation specifically, TNF-α inhibition is often coadministered with T-cell depletion in the induction period to protect intraportally infused islets from the damaging effect of the instant blood-mediated inflammatory reaction.[4]

Maintenance immunosuppression in pancreas transplantation most commonly consists of the calcineurin-inhibitor tacrolimus with mycophenolate mofetil with or without a low dose of corticosteroid.[15] Sirolimus or everolimus (mTOR inhibitor) are occasionally used in place of mycophenolate mofetil. For islet transplantation, initial maintenance therapy usually includes tacrolimus with either mycophenolate mofetil or sirolimus.[3,28] Corticosteroids are generally avoided for islet transplantation because of the inherent β cell toxicity of the drug and because of the insulin resistance corticosteroids cause. Calcineurin inhibitors are a concern for their known potential

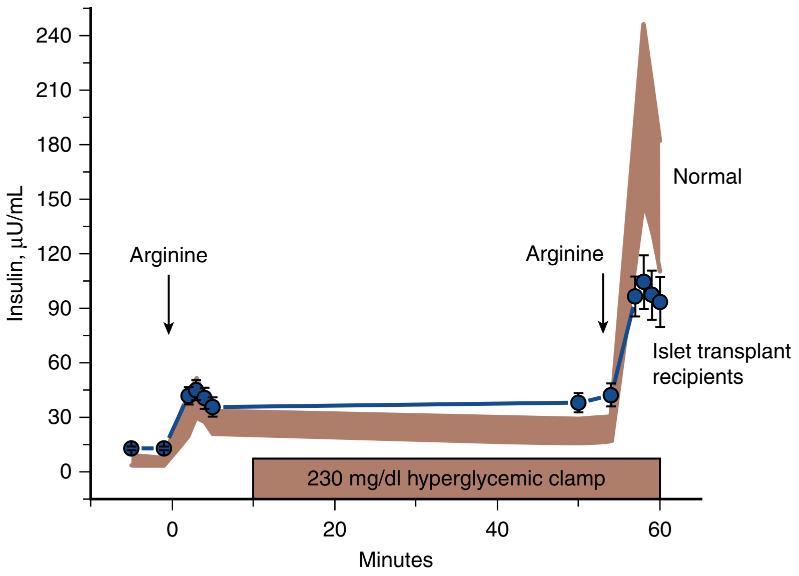


Fig. 41.10 Data demonstrating the acute insulin response to arginine and glucose-potentiated arginine stimulation in patients with type 1 diabetes mellitus after successful islet transplant in the Clinical Islet Transplant–07 study. (Reproduced with permission from Rickels MR, Liu C, Shlansky-Goldberg RD, et al. Improvement in β-cell secretory capacity after human islet transplantation according to the CIT07 protocol. *Diabetes.* 2013;62:2890–2897.)

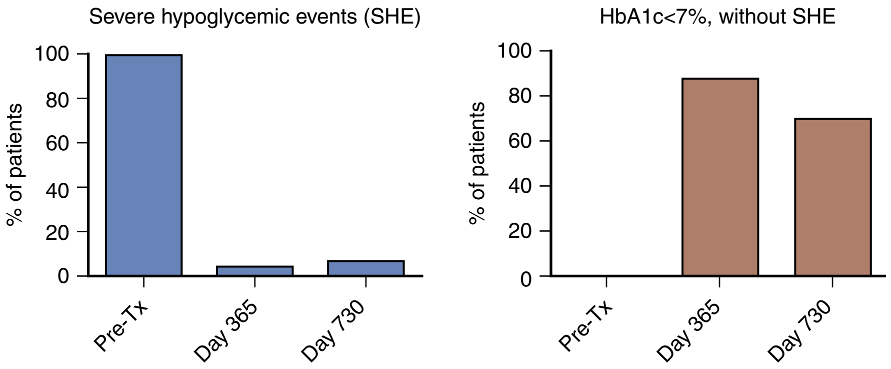


Fig. 41.11 Data from the multicenter Clinical Islet Transplant–07 islet transplant alone trials demonstrating resolution of severe hypoglycemia **A**, with glycated hemoglobin (HbA1c) <7% **B**, after islet transplant in patients with longstanding labile type 1 diabetes mellitus. *Pre-Tx*, Pretransplantation. (Modified from data from Hering BJ, Clarke WR, Bridges ND, et al. Phase 3 trial of transplantation of human islets in type 1 diabetes complicated by severe hypoglycemia. *Diabetes Care.* 2016;39:1230–1240.)

to inhibit glucose-dependent insulin secretion and potential for nephrotoxicity, but to date have been a key component of the immunosuppression cocktail to prevent rejection of a pancreas or islet graft.[132-135] Recent introduction of T-cell costimulatory blockade has offered some hope for calcineurin inhibitor–free maintenance immunosuppression. Small studies show favorable islet transplant outcomes for recipients treated with the T-cell costimulatory inhibitors efalizumab (now off the market) or belatacept in the absence of a calcineurin inhibitor.[136,137] However, in a phase 2 multicenter open-label trial in SPK transplants in which patients were randomized to conventional immunosuppression (tacrolimus plus mycophenolate mofetil) or an investigational arm of belatacept, mycophenolate mofetil, and low-dose tacrolimus with planned withdrawal of tacrolimus, more pancreas rejection events occurred in the belatacept arm, most after discontinuation of tacrolimus, and the trial was ultimately stopped early.[138]

Common side effects of immunosuppression for the transplant recipient include nausea, diarrhea, electrolyte abnormalities, oral ulcerations, hyperlipidemia, cytopenias, and increased serum creatinine. Sirolimus can be associated with painful oral ulcers, and many islet transplant recipients initially treated with sirolimus-tacrolimus require transition to a mycophenolate- tacrolimus combination for better tolerance.[139] Major concerns of immunosuppression are the risks of severe or opportunistic infection, nephrotoxicity, and cancers. The risk for skin cancer varies by geographical region and immunosuppression regimen, but overall risk for skin cancer is around 20% by 10 years after solid organ transplantation.[140] The use of calcineurin inhibitor therapy (tacrolimus, cyclosporine), alone or in combination with sirolimus, presents risk for nephrotoxicity posttransplant in patients with type 1 diabetes. In islet transplant recipients, low dose tacrolimus plus sirolimus maintenance therapy was associated with an approximately 5 mL/min/year/1.73 m^2 decline in glomerular filtration rate, but there was significant heterogeneity in between individuals.[141] Sirolimus presents an additional risk of albuminuria, present in approximately 5% of treated patients.[142] Careful selection of transplant

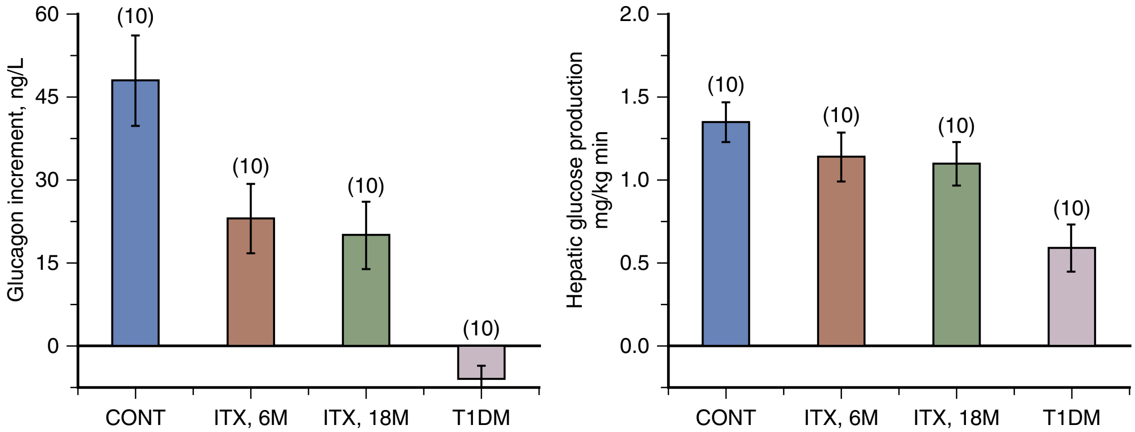


Fig. 41.12 Glucagon levels during hypoglycemic clamp studies in islet transplant (*ITX*) recipients at 6 and 18 months (*M*) posttransplant remains lower than in controls (*CONT*), but are significantly improved when compared with pretransplant levels in patients with type 1 diabetes mellitus (*T1DM*). Importantly, hepatic glucose production during hypoglycemic clamp studies is normalized in ITX recipients when compared with CONT. In contrast, T1DM patients pretransplant show significantly impaired hepatic glucose output in response to hypoglycemia. (Data from Rickels MR, Peleckis AJ, Markmann E, et al. Long-term improvement in glucose control and counterregulation by islet transplantation for type 1 diabetes. *J Clin Endocrinol Metab.* 2016;101:4421–4430.)

recipients with normal kidney function before transplant and careful monitoring of renal function and immunosuppression dosing can minimize risk for adverse renal outcomes.[143]

Epstein–Barr virus (EBV) and cytomegalovirus (CMV) are of particular concern posttransplant and are more common with T-cell–depleting induction therapies.[144-146] Valgancyclovir is commonly administered for several months posttransplant as prophylaxis against CMV. Both pancreas and islet transplant recipients are at risk for CMV, although risk is lower in islet transplant because of low tissue volume, small number of passenger donor leukocytes, and lack of lymphoid tissue in islet grafts.[145,147] Quantitative assays (polymerase chain reaction) for EBV and CMV viremia are often performed at protocol-defined intervals, even in asymptomatic patients. Posttransplant lymphoproliferative disease has been reported after pancreas transplant, but rarely after islet transplant, with transplant recipients who are EBV-seronegative at transplant at highest risk.[148,149 146] Additional antimicrobial prophylaxis after pancreas and islet transplant includes trimethoprim/sulfamethoxazole to protect against *Pneumocystis jarvii*.

Despite advances in immunosuppressive therapy, transplanted pancreas or islet grafts remain at risk for alloimmune rejection or recurrence of β cell autoimmunity. In pancreas transplant, serum amylase and lipase (or urinary amylase and lipase in the case of bladder drainage) are routinely monitored, with elevations raising concern for possible rejection.[150] In SPK transplantation, an elevated serum creatinine can be a surrogate marker for pancreas rejection, and thus rejection is often detected and treated more effectively in combined pancreas-kidney transplant than pancreas transplant alone.[1] When rejection is suspected, radiology-guided pancreas biopsy can be performed to definitively diagnose acute cellular and/or antibody-mediated rejection.[150] If diagnosed sufficiently early, before fibrosis of the pancreas graft and need for resumption of insulin therapy, acute rejection in pancreas transplant can often be treated successfully with a short course of added immunosuppression. High-dose corticosteroids and/or antithymocyte globulin are commonly used to treat acute cellular rejection, whereas antibody-mediated rejection is treated with intravenous immunoglobulin and plasmapheresis or anti-B cell therapies for refractory cases.[150] Unfortunately, in islet transplantation efficient biomarkers for rejection are largely lacking, and, because of patchy distribution of islets in the liver, biopsy diagnosis is not feasible. For islet transplants, acute rejection is often undiagnosed until the recipient presents with sudden-onset hyperglycemia, at which point islet damage is likely irreversible. Donor-specific antibodies may be present at diagnosis or preceding acute rejection.[151,152] Islet graft rescue for presumptive cases of cellular and antibody-mediated rejection have been reported.[153,154]

Recurrence of β cell autoimmunity remains a concern, and in pancreas transplantation may be distinguished from classic rejection based on biopsy features.[155] In islet transplant, recurrent autoimmunity may be suspected in the case of rapid loss of graft function in the absence of donor-specific antibodies detected in blood samples. T-cell autoreactivity has been shown to correlate with both pancreas and islet graft loss, but these cellular assays are cumbersome to perform and not clinically available.[127,156] β cell autoantibodies are less sensitive as markers of recurrent autoimmunity, but new positive antibodies or increasing antibody titers have shown some correlation with biopsy-proven type 1 diabetes recurrence in pancreas transplant recipients.[155,157] Circulating islet exosomes can delineate cell type–specific islet damage and in at least one case has provided proof of concept to distinguish specific β cell loss typical of recurrent autoimmunity from the more general islet destruction associated with classic alloimmune rejection.[158]

KEY POINTS

- Both pancreas and alloislet transplant recipients must take immunosuppressive drugs to prevent rejection and recurrence of autoimmunity (in type 1 diabetes). Immunosuppression must be carefully monitored to provide sufficient dosing to mitigate risk for pancreas or islet rejection, while minimizing drug toxicity. Minor side effects (e.g., gastrointestinal symptoms) are common, while less common but serious risks of treatment include renal insufficiency, opportunistic infections, or malignancy.

AUTOLOGOUS ISLET (AUTOISLET) TRANSPLANTATION

Patient Selection for Autoislet Transplantation

Autologous islet transplantation is most often performed for patients with recurrent acute and/or chronic pancreatitis who are undergoing total pancreatectomy, completion pancreatectomy after prior pancreas surgery, or (less commonly) distal pancreatectomy.[159] Candidates for a pancreatectomy must have clear evidence that abdominal pain is secondary to pancreatitis based on either recurrent episodes of acute pancreatitis documented by elevated lipase or imaging or a clear diagnosis of chronic pancreatitis by imaging.[11] Calcifications in the pancreas are clearly diagnostic for chronic pancreatitis in a patient with compatible clinical symptoms, but pancreatic calcifications are present only at a late stage of disease.[160] Noncalcific chronic pancreatitis is more difficult to diagnose. Endoscopic ultrasound, magnetic resonance imaging with secretin-stimulated magnetic resonance cholangiopancreatography, and endoscopic pancreatic function tests measuring bicarbonate secretion following secretin stimulation are all recommended for the diagnostic evaluation for noncalcific chronic pancreatitis.[161-163] In the absence of clear preceding history of recurrent acute pancreatitis, two positive diagnostic tests are preferred to establish a definitive diagnosis of noncalcific chronic pancreatitis before proceeding to surgery.[164]

Total pancreatectomy with autoislet transplant is only indicated in patients who continue to have abdominal pain from either chronic pancreatitis or recurrent episodes of acute pancreatitis when the pain is significant enough to interfere with quality of life despite optimal medical care and, when indicated, endoscopic therapy.[11] Although many patients are reliant on chronic opioids to manage pain at the time of total pancreatectomy with autoislet transplant, it is not necessary or recommended to wait for a patient to become opioid-dependent before considering total pancreatectomy with autoislet transplant. A shorter duration of disease, less severe pancreatic fibrosis, and a shorter history of opioid use are all likely to convey a better prognosis for both resolution of pain after surgery and higher islet mass for transplant.[165-168] Total pancreatectomy with autoislet transplant is often preferred over other conventional duct drainage surgeries like a Puestow or Frey for small-duct pancreatitis, children with pancreatitis, and genetic pancreatitis, because of high risks for disease recurrence and technical inability to perform a drainage with a small pancreatic duct.[33,169-172] Patients who have had a prior drainage and/or partial resection procedure are candidates for total pancreatectomy with autoislet transplant to treat pain, but they should be counseled that these prior surgeries impair islet yield, and the risk is high that they will need at least partial supplementation with exogenous insulin lifelong after surgery.[11,173] Distal pancreatic resections are rarely indicated for chronic pancreatitis but might be considered in specific settings—for example, distal pancreatectomy may be performed if pancreatitis is only present in the tail because of a disruption of the mid pancreatic duct. These patients can be considered for an autoislet transplant using the resected portion of the pancreas.

Because total pancreatectomy with autoislet transplant is a major surgical procedure with a physically, emotionally, and cognitively demanding recovery process, medical and psychological comorbidities need to be assessed and managed before surgery.[11] The patient needs to be psychologically and cognitively able to manage exocrine insufficiency and diabetes to be a candidate for surgery. The majority of patients undergoing total pancreatectomy with autologous islet transplant do not have diabetes mellitus before surgery, although most have at least some loss of islet mass because of the fibrotic destruction of exocrine and endocrine tissue that occurs with chronic pancreatitis.[174] Approximately 8% of autoislet transplant recipients will have mild preexisting pancreatogenic diabetes at time of surgery.[11] In these cases autoislet transplant is considered if sufficient C-peptide production can be demonstrated, with the surgical goal to preserve residual β cell mass.[175] Patients with chronic pancreatitis who have no or low C-peptide are not candidates for an islet infusion but can be considered for pancreatectomy alone to treat pain and associated disability, which is the primary goal of the surgery.

Rarely, autologous islet transplantation is also performed at time of extended distal or total pancreatectomy for other benign pancreatic disease, including trauma, benign neoplasm, or, more rarely, premalignant conditions.[176-179] Autologous islet transplantation is not recommended for malignant disease outside the context of a clinical trial because of the risk for establishing new metastatic disease in the liver when residual exocrine tissue is coinfused with the islet product.[180,181]

Surgical Approaches to Autoislet Transplantation

The surgical procedure of total pancreatectomy with autoislet transplantation (Fig. 41.13) carries significantly higher morbidity when compared with an alloislet transplant because of the pancreatectomy and associated anatomic reconstructions (Fig. 41.14).[182] However, the procedures of islet isolation and infusion are technically similar for autoislet and alloislet transplantation, with the exception that higher tissue volume and less pure islet tissue is used for autoislets to maximize islet availability.[183]

For the total pancreatectomy portion of the procedure, the pancreas is entirely resected, taking care to preserve the blood supply to the pancreas for as long as possible to minimize warm ischemia time.[11] Because of the vascular anatomy, the spleen and a portion of the duodenum are also resected. There are several approaches to restoring gastrointestinal and biliary tract continuity following pancreatectomy. Most often the gastrointestinal anatomy is reconstructed with a Roux-en-Y duodenojejunostomy or gastrojejunostomy.[170,172,184] The biliary tract is reanastomosed most commonly via choledochojejunostomy or hepaticojejunostomy.[172] One early surgical approach was end-to-end duodenoduodenostomy with choledochoduodenostomy, but this technique is rarely used in the current era because of a high risk for bile reflux gastritis.[185] In some centers a gastrojejunal or jejunal feeding tube may be placed during surgery to provide nutrition during surgical recovery.[172]

Once the pancreas is resected, it is transported immediately to the islet isolation facility. Islet isolation is most often performed at the same center where the pancreatectomy is occurring, but rarely the pancreas is transported to a remote islet lab for isolation.[186-188] Current data suggest similar or nearly equivalent outcomes with remote versus on-site islet isolation. Similar to the procedures for a cadaveric donor pancreas, the pancreatic duct is cannulated, and a collagenase solution is dispersed through the pancreas parenchyma for enzymatic digestion, followed by mechanical digestion of the pancreas using the semiautomated Ricordi method.[189] Autoislet grafts infrequently undergo a purification step by density centrifugation. Tissue volumes of up to 0.25 mL/kg recipient body weight are considered acceptable for autoislet transplantation, and because purification will reduce islet recovery and result in a lower islet dose transplanted to the patient, it is only used when necessary to reduce tissue volume to an acceptable threshold.[183] Islets can be infused intraoperatively by directly accessing a tributary of the portal vein or can be infused with a percutaneous transhepatic approach with interventional radiology following the surgical procedure.[172,190] As with allogeneic islet transplantation, heparin is administered with the islets and subsequently followed by continuous intravenous infusion or subcutaneous administration to minimize the risk for portal branch vein thrombosis.[11] Intensive insulin therapy is instituted to protect the transplanted islets from the toxic effects of hyperglycemia during the relative hypoxic period of engraftment until revascularization occurs.[10]

Patient and Graft Survival After Autoislet Transplant

The primary rationale for total pancreatectomy with autoislet transplant is to lessen pancreatic pain and improve quality of life.

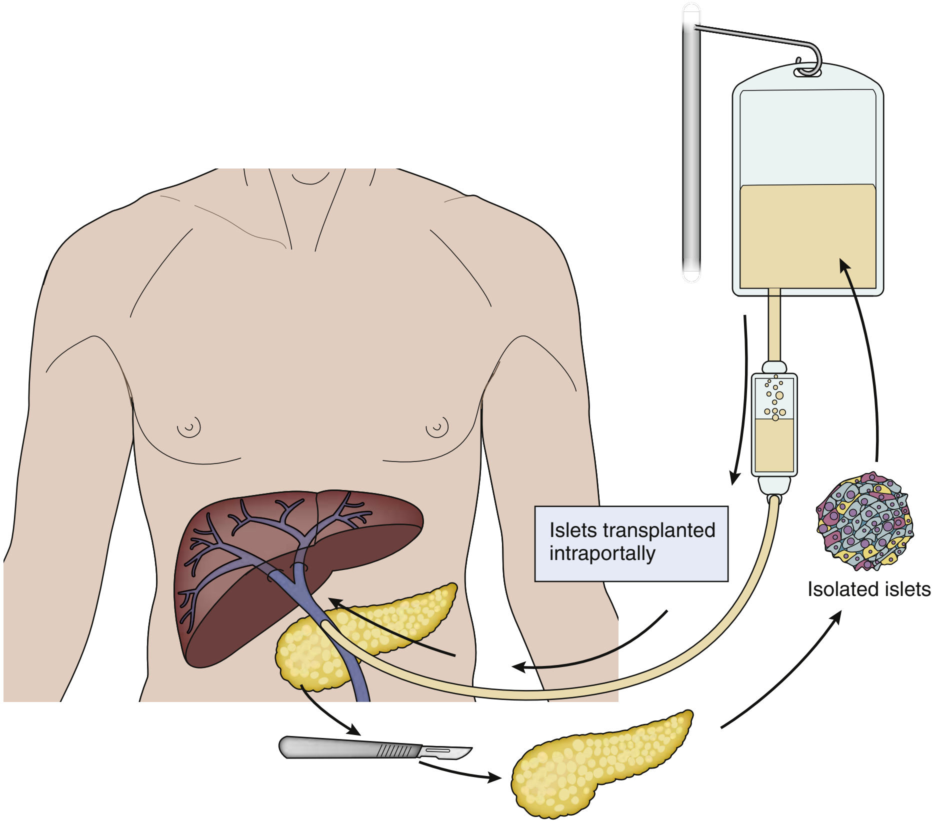

Islets transplanted intraportally
Isolated islets

Fig. 41.13 **Total Pancreatectomy With Autologous Lslet Transplantation.** The procedure involves removal of the diseased pancreas for collagenase digestion and centrifuge purification in a current good manufacturing practice or current good tissue practice islet isolation facility. The final islet product is transplanted intraportally via the splenic vein stump, a mesenteric artery, or recannulation of the umbilical vein. (Reproduced with permission from Rickels MR, Robertson RP. Pancreatic islet transplantation in humans: recent progress and future directions. *Endocr Rev.* 2019;40:631–668.)

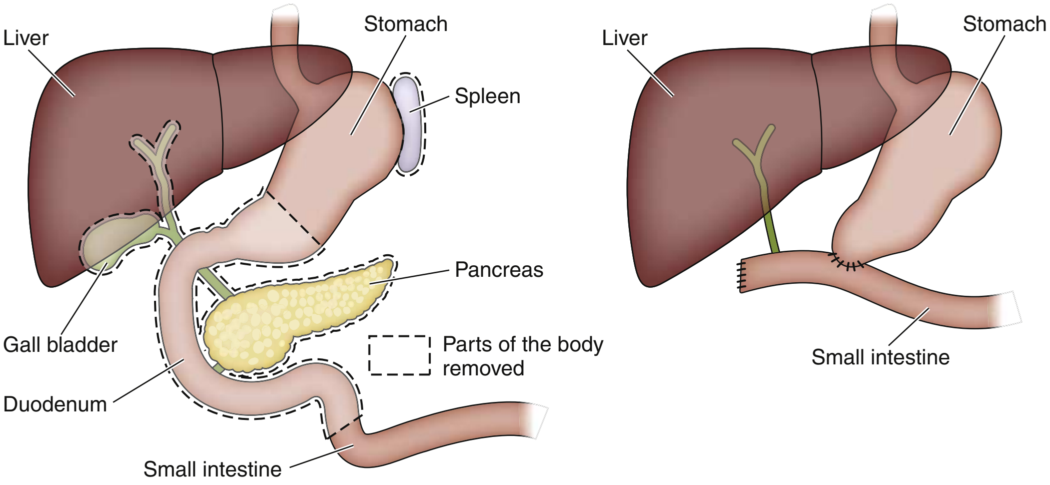

Liver
Stomach
Spleen
Pancreas
Gall bladder
Parts of the body removed
Duodenum
Small intestine
Liver
Stomach
Small intestine

Fig. 41.14 **Gastrointestinal Reconstruction Following Total Pancreatectomy.** The pancreas is removed together with the duodenum and spleen that share their arterial blood supply with the pancreas. While the pancreas is processed for islet isolation, Roux-en-Y reconstruction is performed as shown here with gastrojejunostomy, or with duodenojejunostomy, and choledochojejunostomy. (Reproduced with permission from Lund A, Bagger JI, Wewer Albrechtsen NJ, et al. Evidence of extrapancreatic glucagon secretion in man. *Diabetes.* 2016;65:585–597.)

Multiple centers have demonstrated improved quality of life and reduced opioid use after this procedure, with approximately 50% or more of adults and 80% of children opioid free at 1 year and significantly improved health-related quality of life scores overall.[11,172] Patient survival historically is 95% at 1 year and 72% at 10 years after surgery.[191]

Survival of autoislets has been defined by C-peptide positivity or by insulin independence. Depending on the patient population, insulin independence has been reported in 15% to 40% of patients at time intervals of 1 to 5 years posttransplant.[192] The largest 10-year follow-up cohort reported that 22% of patients were insulin independent at 5 years, and 13% were insulin independent at 10 years.[191] Factors potentially increasing likelihood for insulin independence and durability of insulin independence include younger patient age, higher islet mass transplanted, lack of prior pancreas surgeries, and shorter duration of pancreatitis.[192] At 1 year post–total pancreatectomy with islet autotransplantation, nearly 100% have clinically significant C-peptide (meal-stimulated value ≥0.6 ng/mL) when greater than 5,000 IEQ/kg are transplanted, 97% with 2500 to 5000 IEQ/kg, and 79% with less than 2500 IEQ/kg.[11]

Metabolic Outcomes

As with alloislet transplantation, successful autoislet transplantation with a sufficient number of islets (~350,000 islets or 5000 IEQ/kg body weight) achieves normal or nearly normal levels of fasting blood glucose and HbA1c. However, in the case of autoislet transplantation specifically, a frequent problem following the procedure is hypoglycemia associated with exercise and following high-carbohydrate meals.[193-195] This posttransplant complication is thought to be primarily associated with the use of Roux-en-Y surgery for pancreatectomy. This causes significant alterations in the normal physical relationships between the stomach outlet and the small intestine (Fig. 41.14). This results in a mismatch between glucose and insulin levels, such that the autoislets release excessive insulin given the ambient glucose levels in venous blood early after ingestion and insufficient glucagon during the late postprandial period. Interestingly, failure of an appropriate glucagon response to insulin-induced hypoglycemia seen with intrahepatic autoislet transplantation was not observed when autoislets were transplanted into nonhepatic sites in addition to the usual intrahepatic site (Fig. 41.15).[195] This led to the suggestion to provide autoislet recipients with intact glucagon responses to hypoglycemia by transplanting at least a portion of the autoislets in a nonhepatic site, such as the omentum.[195-197]

Metabolic tests have been used to estimate the number of intrahepatic islets that survive intrahepatic transplantation. The test used most frequently involves glucose potentiation of GPAIS.[196] Intravenous arginine alone can be used in both diabetic and nondiabetic subjects to induce an acute insulin response (see Chapter 32). This response is augmented when a prior glucose infusion is used to elevate blood glucose levels (Fig. 41.16). GPAIS has thus been used to estimate insulin secretory reserve and has been applied to recipients of intrahepatic autoislets. Because these recipients no longer have a native pancreas, insulin responses come exclusively from their transplanted islets. Correlations between the number of intrahepatic islets transplanted and the magnitude of GPAIS responses are significant (Fig. 41.17). The number of islets transplanted is a known value, and the slope of the line reflecting this relationship is close to the theoretical slope observed in normal subjects, therefore this justifies the use of the GPAIS value as an estimate of the number of islets surviving transplantation. Repeated use of GPAIS in the years following islet transplantation may provide a useful estimate of survival of transplanted allo- and autoislets.[117]

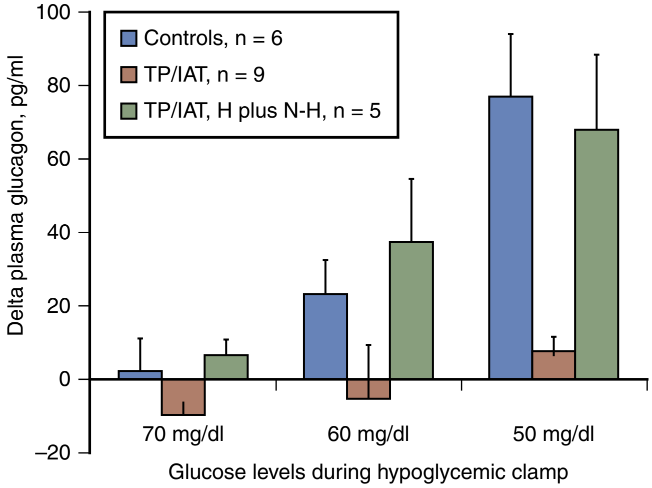


Fig. 41.15 Hypoglycemic clamp studies in patients who underwent total pancreatectomy with islet autotransplantation (*TP/IAT*) demonstrated minimal glucagon responsiveness to hypoglycemia with conventional intraportal islet transplant but restoration of normal glucagon counterregulation when a portion of the islets was transplanted into a nonhepatic site, usually the peritoneal cavity. *H*, Hepatic; *N-H*, nonhepatic site. (Reproduced with permission from Bellin MD, Parazzoli S, Oseid E, et al. Defective glucagon secretion during hypoglycemia after intrahepatic but not nonhepatic islet autotransplantation. *Am J Transplant.* 2014;14:1880–1886.)

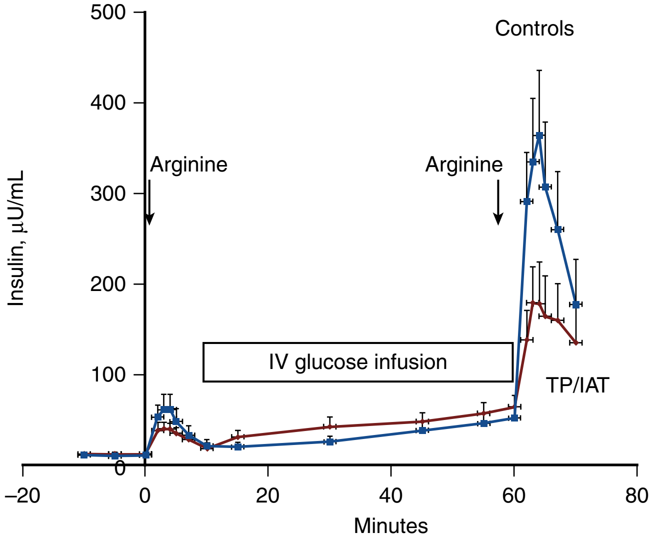


Fig. 41.16 Data demonstrating the acute insulin response to arginine and glucose-potentiated arginine stimulation in patients who underwent total pancreatectomy with islet autotransplantation (*TP/IAT*). *IV,* Intravenous. (Reproduced with permission from Robertson RP, Bogachus LD, Oseid E, et al. Assessment of beta-cell mass and alpha- and beta-cell survival and function by arginine stimulation in human autologous islet recipients. *Diabetes.* 2015;64:565–572.)

KEY POINTS

- Pancreatectomy with islet autotransplantation is indicated for severe, painful forms of chronic and recurrent acute pancreatitis. Most patients exhibit islet graft function (C-peptide–positive), with insulin independence in 15% to 40% of patients and most common with earlier surgical intervention, less pancreatic fibrosis, and high islet mass. Hypoglycemia posttransplant can occur with exercise or high carbohydrate meals because of alterations of the gastrointestinal track (Roux-en-Y reconstruction after pancreatectomy) and failure of an appropriate glucagon response to hypoglycemia.

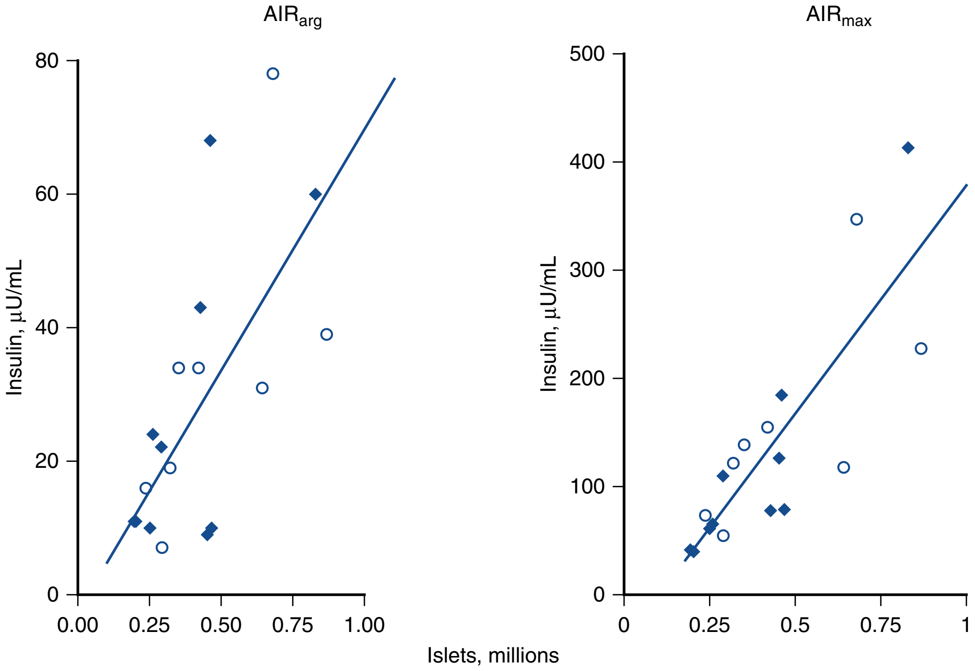


Fig. 41.17 Correlation between number of autoislets transplanted and the acute insulin response to arginine injected under fasting (*AIRarg*; *left panel*) and glucose-potentiated (*AIRmax*; *right panel*) conditions. (Reproduced with permission from Robertson RP, Bogachus LD, Oseid E, et al. Assessment of beta-cell mass and alpha- and beta-cell survival and function by arginine stimulation in human autologous islet recipients. *Diabetes.* 2015;64:565–572.)

EFFECT OF SUCCESSFUL PANCREAS OR ISLET TRANSPLANTATION ON COMPLICATIONS OF DIABETES

The risk for microvascular complications in T1DM is mediated by chronic hyperglycemia. Successful pancreas and islet transplantation restores normoglycemia and protects against development of diabetes-related changes in transplanted kidneys.[198,199] Reversal of diabetic lesions in the native kidney is apparent on renal biopsy at 10 years after successful pancreas transplantation.[5] Retinopathy stabilized or reversed in a cohort of 33 patients after pancreas transplant alone, whereas a medically treated control group had progression of retinopathy in 14%.[200] This was consistent with earlier findings in a study comparing fundoscopic changes and the need for further laser treatment in patients receiving kidney alone versus kidney and pancreas transplantation.[201] A similar benefit has been reported for neuropathy, with improved motor, sensory, and autonomic nerve function reported over 10 years after pancreas transplantation.[202] Importantly, this was also associated with a decreased death rate associated with autonomic insufficiency over the same time period.[203] Pancreas transplantation has potential benefits for cardiovascular health, with reduced cardiovascular mortality in pancreas-kidney transplant recipients compared with kidney transplant alone.[204] In a prospectively controlled study, regression of carotid artery intimal thickening was observed in a pancreas and kidney transplant group compared with no regression of arterial lesions in a kidney transplant alone group.[205]

Although the research on secondary diabetes complications after islet transplant is less extensive, available data suggest a similar protective effect against progression of microvascular complications. In a single cohort of 26 patients with type 1 diabetes, islet transplant alone with low-dose tacrolimus and mycophenolate resulted in better preserved renal function when compared with medical treatment (glomerular filtration decline of –1.4 mL/min/year/1.73m^2 vs. –3.6 mL/min/year/1.73m^2), and no patient had progression of retinopathy after islet transplant.[6,79] A separate cohort of patients followed for 5 years after allotransplantation showed improved sensory nerve conduction velocities after 3 years posttransplant.[206]

KEY POINTS

- Successful pancreas or islet allotransplantation stabilizes microvascular complications of diabetes, including retinopathy, nephropathy, and neuropathy. Biopsy-established diabetic nephropathy shows reversal after 10 years of successful pancreas transplantation, and reversal of autonomic neuropathy after pancreas transplant is associated with reduced mortality.

FUTURE DIRECTIONS AND CONCLUSIONS

Because the pool of deceased pancreas donors is limited, to offer transplant therapies to a larger population of patients with diabetes, a renewable cell source will be necessary. The two most promising avenues of research for a replenishable islet supply are porcine islets and stem cell–derived islets.

Adult and neonatal porcine islets have successfully restored insulin independence in nonhuman primate models when treated with conventional immunosuppression plus costimulatory blockade with an anti-CD154 monoclonal antibody (not available for clinical use at time of this writing).[207,208] Porcine xenografts are subject to a robust, instant blood-mediated inflammatory response that may be lessened by genetic manipulation of porcine islets, including development of genetically engineered galactose-alpha1, 3-galactose–deficient donor pigs.[209] Effective immunosuppression or immunoprotection by encapsulation will be necessary for porcine islets to become a therapeutic option for patients. Encapsulation of islets is challenging, as it requires a semipermeable barrier that can prevent immune cells from infiltrating the islet graft while still allowing adequate diffusion of oxygen, nutrients, and insulin.

One concern specific to porcine islets is the risk for transmission of xenotic infectious organisms. The porcine endogenous retroviruses (PERVs) are ubiquitous in pigs, and transmission from porcine to human cells has been demonstrated *in vitro*.[210] However, no transmission of PERV has been observed in very limited early clinical trials in which patients with T1DM received porcine islets microencapsulated in alginate or in subcutaneous collagen-covered macroencapsulation devices.[211,212] Clinical-grade source pigs for islet transplant would need to be housed in a pathogen-free secured environment with routine monitoring for PERV and other infectious organisms.

Stem cell–derived islets are the other promising future source for cell therapy, either by use of embryonic stem cell lines or induced pluripotent stem cells. Embryonic stem cells have the advantage of being a single renewable cell line that can be readily scaled up for large-batch manufacturing, but the disadvantage of posing a higher risk for teratoma transformation, immunogenicity, ethical constraints, and potential unacceptability of therapy for some patients. Embryonic cells reprogrammed *in vitro* to immature pancreatic endocrine progenitor cells (with further maturation *in vivo*) or differentiated *in vitro* into fully mature pancreatic islets with glucose-responsive insulin secretion have both shown success in reversing diabetes in murine models.[213,214] More recently, phase I/II clinical trials have been initiated with encapsulated embryonic stem cell–derived pancreas endoderm cells implanted subcutaneously with an immunoprotective membrane without immunosuppression, or a vascularized membrane with immunosuppression. Embryonic stem cell–derived islets are an allogeneic cell transplant and will require either immunosuppression or encapsulation to prevent rejection.[215]

Induced pluripotent stem cells are derived from a patient's own cells and thus are HLA-identical. Functional β cells have been developed using induced pluripotent stem cells both from nondiabetic controls and patients with T1DM.[216] Although autologous islets or β cells derived from one's own induced pluripotent stem cells offers promise for an immunosuppression-free approach to cell therapy for postpancreatectomy diabetes or T2DM, patients with T1DM would continue to require immunosuppressive treatment or immunoencapsulation to prevent recurrent β cell autoimmunity.

In conclusion, allogeneic whole pancreas and isolated islet transplantation are established procedures for patients with insulin-dependent diabetes complicated by excessive glycemic lability and/or hypoglycemia unawareness experiencing severe hypoglycemia events or by the requirement for immunosuppression in support of another organ transplant, most often a kidney transplant for end-stage diabetic nephropathy. Availability of both procedures is dependent on national and regional expertise and policy governing the regulation of cellular therapy. Most often performed in patients with T1DM, pancreas transplantation is increasingly offered to patients with insulin-dependent T2DM or pancreatogenic forms of diabetes at the time of or following another organ transplant. Islet transplantation offers a minimally invasive alternative to pancreas transplantation that has been used mainly for the treatment of T1DM owing to challenges in isolating and engrafting a sufficient islet mass to eliminate insulin requirements. Both approaches have demonstrated long-term improvement in metabolic control and quality of life and reduction in diabetes-related complications balanced against the risk for immunosuppressive drug therapy. Autologous islet transplantation may be offered to patients requiring total pancreatectomy for treatment of recurrent acute or chronic pancreatitis in order to mitigate or prevent postsurgical diabetes and offers improvement in metabolic control and quality of life without the associated risks of immunosuppression. Future work in genetically engineered porcine islets and human embryonic or inducible stem cell–derived islets offer hope for a one-day inexhaustible supply of islet tissue available for cell therapy to treat diabetes.

KEY POINTS

- β cell replacement therapy with pancreas or islet transplants can restore normoglycemia and prevent acute and chronic complications of disease, but its application is currently limited by the small number of suitable deceased pancreas donors; potential renewable sources of islets in development include human stem cell–derived islets and xenogeneic porcine islets.

REFERENCES

1. Sutherland DE, Gruessner RW, Dunn DL, et al. Lessons learned from more than 1,000 pancreas transplants at a single institution. *Ann Surg*. 2001;233:463–501.
2. Kandaswamy R, Stock PG, Gustafson SK, et al. OPTN/SRTR 2018 annual data report: pancreas. *Am J Transplant*. 2020;20:131–192.
3. Hering BJ, Clarke WR, Bridges ND, et al. Phase 3 trial of transplantation of human islets in type 1 diabetes complicated by severe hypoglycemia. *Diabetes Care*. 2016;39:1230–1240.
4. Bellin MD, Barton FB, Heitman A, et al. Potent induction immunotherapy promotes long-term insulin independence after islet transplantation in type 1 diabetes. *Am J Transplant*. 2012;12:1576–1583.
5. Fioretto P, Steffes MW, Sutherland DE, et al. Reversal of lesions of diabetic nephropathy after pancreas transplantation. *N Engl J Med*. 1998;339:69–75.
6. Thompson DM, Meloche M, Ao Z, et al. Reduced progression of diabetic microvascular complications with islet cell transplantation compared with intensive medical therapy. *Transplantation*. 2011;91:373–378.
7. Dunn TB, Kirchner V, Bellin MD. Beta-cell replacement therapy: current outcomes and future landscape. *Curr Opin Organ Transplant*. 2015;20:681–690.
8. Vantyghem MC, de Koning EJP, Pattou F, et al. Advances in beta-cell replacement therapy for the treatment of type 1 diabetes. *Lancet*. 2019;394:1274–1285.
9. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43:S98–S110.
10. Rickels MR, Robertson RP. Pancreatic islet transplantation in humans: recent progress and future directions. *Endocr Rev*. 2019;40:631–668.
11. Sutherland DE, Radosevich DM, Bellin MD, et al. Total pancreatectomy and islet autotransplantation for chronic pancreatitis. *J Am Coll Surg*. 2012;214:409–424.
12. Sutherland DE, Matas AJ, Goetz FC, et al. Transplantation of dispersed pancreatic islet tissue in humans: autografts and allografts. *Diabetes*. 1980;29:31–44.
13. Kelly WD, Lillehei RC, Merkel FK, et al. Allotransplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. *Surgery*. 1967;61:827–837.
14. Kelly WBL, Lillehei RC, Merkel FK, Idezuki Y, Goetz FC. Allotrannsplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. *Transplantation*. 1968;6:145.
15. Gruessner AC, Gruessner RW. Pancreas transplantation of US and non-US cases from 2005 to 2014 as reported to the United Network for Organ Sharing (UNOS) and the International Pancreas Transplant Registry (IPTR). *Rev Diabet Stud*. 2016;13:35–58.
16. Ballinger WF, Lacy PE. Transplantation of intact pancreatic islets in rats. *Surgery*. 1972;72:175–186.
17. Reckard CR, Barker CF. Transplantation of isolated pancreatic-islets across strong and weak histocompatibility barriers. *Transplant Proc*. 1973;5:761–763.

18. Kretschmer GJ, Sutherland DE, Matas AJ, et al. Islet autotransplantation in pancreatectomized dogs: liver vs spleen as a site of pancreatic tissue implantation. *Surg Forum*. 1977;28:292–294.
19. Najarian JS, Sutherland DE, Matas AJ, et al. Human islet transplantation: a preliminary report. *Transplant Proc*. 1977;9:233–236.
20. Sutherland DE, Matas AJ, Najarian JS. Pancreatic islet cell transplantation. *Surg Clin North Am*. 1978;58:365–382.
21. Najarian JS, Sutherland DE, Matas AJ, et al. Human islet autotransplantation following pancreatectomy. *Transplant Proc*. 1979;11:336–340.
22. Najarian JS, Sutherland DE, Baumgartner D, et al. Total or near total pancreatectomy and islet autotransplantation for treatment of chronic pancreatitis. *Ann Surg*. 1980;192:526–542.
23. Ricordi C, Lacy PE, Finke EH, et al. Automated method for isolation of human pancreatic islets. *Diabetes*. 1988;37:413–420.
24. Scharp DW, Lacy PE, Santiago JV, et al. Insulin independence after islet transplantation into type I diabetic patient. *Diabetes*. 1990;39:515–518.
25. Shapiro AM, Lakey JR, Ryan EA, et al. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med*. 2000;343:230–238.
26. Hering BJ, Kandaswamy R, Ansite JD, et al. Single-donor, marginal-dose islet transplantation in patients with type 1 diabetes. *J Am Med Assoc*. 2005;293:830–835.
27. Markmann JF, Rickels MR, Eggerman TL, et al. Phase 3 trial of human islet-after-kidney transplantation in type 1 diabetes. *Am J Transplant*. 2021;21:1477–1492.
28. Collaborative Islet Transplant Registry: 10th Annual Report. Available from: https://citregistry.org/system/files/10th_AR.pdf. Accessed 2/20/2020.
29. Blondet JJ, Carlson AM, Kobayashi T, et al. The role of total pancreatectomy and islet autotransplantation for chronic pancreatitis. *Surg Clin North Am*. 2007;87:1477–1501.
30. Bellin MD, Balamurugan AN, Pruett TL, et al. No islets left behind: islet autotransplantation for surgery-induced diabetes. *Curr Diab Rep*. 2012;12:580–586.
31. Pyzdrowski KL, Kendall DM, Halter JB, et al. Preserved insulin secretion and insulin independence in recipients of islet autografts. *N Engl J Med*. 1992;327:220–226.
32. Sutherland DE, Gruessner AC, Carlson AM, et al. Islet autotransplant outcomes after total pancreatectomy: a contrast to islet allograft outcomes. *Transplantation*. 2008;86:1799–1802.
33. Bellin MD, Freeman ML, Schwarzenberg SJ, et al. Quality of life improves for pediatric patients after total pancreatectomy and islet autotransplant for chronic pancreatitis. *Clin Gastroenterol Hepatol*. 2011;9:793–799.
34. Collaborative Islet Transplant Registry. *Inaugural Report on Autologous Islet Transplantation*; 2017. Accessed 2/20/2020. Available from: https://citregistry.org/system/files/1st_AR_Auto.pdf.
35. Concepcion BP, Feurer ID, Rega SA, et al. Geographic disparities in access to simultaneous pancreas and kidney transplant in the pre- and post-pancreas allocation system eras. *Transplantation*. 2020;104:623–631.
36. Gruessner RW, Sutherland DE, Drangstveit MB, et al. Pancreas allotransplants in patients with a previous total pancreatectomy for chronic pancreatitis. *J Am Coll Surg*. 2008;206:458–465.
37. Cerise A, Nagaraju S, Powelson JA, et al. Pancreas transplantation following total pancreatectomy for chronic pancreatitis. *Clin Transplant*. 2019;33:e13731.
38. Usatin DJ, Perito ER, Posselt AM, et al. Underutilization of pancreas transplants in cystic fibrosis recipients in the United Network Organ Sharing (UNOS) data 1987-2014. *Am J Transplant*. 2016;16:1620–1625.
39. Esmeijer K, Hoogeveen EK, van den Boog PJM, et al. Superior long-term survival for simultaneous pancreas-kidney transplantation as renal replacement therapy: 30-year follow-up of a nationwide cohort. *Diabetes Care*. 2020;43:321–328.
40. Choudhary P, Rickels MR, Senior PA, et al. Evidence-informed clinical practice recommendations for treatment of type 1 diabetes complicated by problematic hypoglycemia. *Diabetes Care*. 2015;38:1016–1029.
41. Reynoso JF, Gruessner CE, Sutherland DE, et al. Short- and long-term outcome for living pancreas donors. *J Hepatobiliary Pancreat Sci*. 2010;17:92–96.
42. Sibley RK, Sutherland DE, Goetz F, et al. Recurrent diabetes mellitus in the pancreas iso- and allograft. A light and electron microscopic and immunohistochemical analysis of four cases. *Lab Invest*. 1985;53:132–144.
43. Kirchner VA, Finger EB, Bellin MD, et al. Long-term outcomes for living pancreas donors in the modern era. *Transplantation*. 2016;100:1322–1328.
44. Kendall DM, Sutherland DE, Najarian JS, et al. Effects of hemipancreatectomy on insulin secretion and glucose tolerance in healthy humans. *N Engl J Med*. 1990;322:898–903.
45. Seaquist ER, Robertson RP. Effects of hemipancreatectomy on pancreatic alpha and beta cell function in healthy human donors. *J Clin Invest*. 1992;89:1761–1766.
46. Robertson RP, Sutherland DE, Seaquist ER, et al. Glucagon, catecholamine, and symptom responses to hypoglycemia in living donors of pancreas segments. *Diabetes*. 2003;52:1689–1694.
47. Robertson RP, Lanz KJ, Sutherland DE, et al. Relationship between diabetes and obesity 9 to 18 years after hemipancreatectomy and transplantation in donors and recipients. *Transplantation*. 2002;73:736–741.
48. Lam NN, Schnitzler MA, Segev DL, et al. Diabetes mellitus in living pancreas donors: use of integrated national registry and pharmacy claims data to characterize donation-related health outcomes. *Transplantation*. 2017;101:1276–1281.
49. Berney T, Boffa C, Augustine T, et al. Utilization of organs from donors after circulatory death for vascularized pancreas and islet of Langerhans transplantation: recommendations from an expert group. *Transpl Int*. 2016;29:798–806.
50. Muthusamy AS, Mumford L, Hudson A, et al. Pancreas transplantation from donors after circulatory death from the United Kingdom. *Am J Transplant*. 2012;12:2150–2156.
51. Serrano OK, Vock DM, Dunn TB, et al. Maximizing utilization in pancreas transplantation: phenotypic characteristics differentiating aggressive from nonaggressive transplant centers. *Transplantation*. 2018;102:2108–2119.
52. Larsen JL. Pancreas transplantation: indications and consequences. *Endocr Rev*. 2004;25:919–946.
53. Becker BN, Becker YT, Odorico JS, et al. Simultaneous pancreas-kidney and pancreas transplantation. *Minerva Urol Nefrol*. 2002;54:213–226.
54. Diem P, Abid M, Redmon JB, et al. Systemic venous drainage of pancreas allografts as independent cause of hyperinsulinemia in type I diabetic recipients. *Diabetes*. 1990;39:534–540.
55. Hampson FA, Freeman SJ, Ertner J, et al. Pancreatic transplantation: surgical technique, normal radiological appearances and complications. *Insights Imaging*. 2010;1:339–347.
56. Petruzzo P, Badet L, Lefrancois N, et al. Metabolic consequences of pancreatic systemic or portal venous drainage in simultaneous pancreas-kidney transplant recipients. *Diabet Med*. 2006;23:654–659.
57. Katz H, Homan M, Velosa J, et al. Effects of pancreas transplantation on postprandial glucose metabolism. *N Engl J Med*. 1991;325:1278–1283.
58. Gruessner AC, Gruessner RWG. Pancreas transplantation for patients with type 1 and type 2 diabetes mellitus in the United States: a registry report. *Gastroenterol Clin North Am*. 2018;47:417–441.
59. Gruessner AC, Gruessner RW. Long-term outcome after pancreas transplantation: a registry analysis. *Curr Opin Organ Transplant*. 2016;21:377–385.
60. Burke 3rd GW, Vendrame F, Virdi SK, et al. Lessons from pancreas transplantation in type 1 diabetes: recurrence of islet autoimmunity. *Curr Diab Rep*. 2015;15:121.
61. Kopp W, van Meel M, Putter H, et al. Center volume is associated with outcome after pancreas transplantation within the Eurotransplant region. *Transplantation*. 2017;101:1247–1253.
62. Alhamad T, Malone AF, Brennan DC, et al. Transplant center volume and the risk of pancreas allograft failure. *Transplantation*. 2017;101:2757–2764.
63. Kim Y, Dhar VK, Wima K, et al. The center volume-outcome effect in pancreas transplantation: a national analysis. *J Surg Res*. 2017;213:25–31.
64. Alhamad T, Kunjal R, Wellen J, et al. Three-month pancreas graft function significantly influences survival following simultaneous pancreas-kidney transplantation in type 2 diabetes patients. *Am J Transplant*. 2020;20:788–796.

65. Fridell JA, Niederhaus S, Curry M, et al. The survival advantage of pancreas after kidney transplant. *Am J Transplant*. 2019;19:823–830.
66. Barlow AD, Saeb-Parsy K, Watson CJE. An analysis of the survival outcomes of simultaneous pancreas and kidney transplantation compared to live donor kidney transplantation in patients with type 1 diabetes: a UK Transplant Registry study. *Transpl Int*. 2017;30:884–892.
67. Paty BW, Lanz K, Kendall DM, et al. Restored hypoglycemic counterregulation is stable in successful pancreas transplant recipients for up to 19 years after transplantation. *Transplantation*. 2001;72:1103–1107.
68. Diem P, Redmon JB, Abid M, et al. Glucagon, catecholamine and pancreatic polypeptide secretion in type I diabetic recipients of pancreas allografts. *J Clin Invest*. 1990;86:2008–2013.
69. Barrou Z, Seaquist ER, Robertson RP. Pancreas transplantation in diabetic humans normalizes hepatic glucose production during hypoglycemia. *Diabetes*. 1994;43:661–666.
70. Flatt AJS, Bennett D, Counter C, et al. beta-Cell and renal transplantation options for diabetes. *Diabet Med*. 2020;37:580–592.
71. Wojtusciszyn A, Branchereau J, Esposito L, et al. Indications for islet or pancreatic transplantation: statement of the TREPID working group on behalf of the Societe Francophone du Diabete (SFD), Societe Francaise d'Endocrinologie (SFE), Societe Francophone de Transplantation (SFT) and Societe Francaise de Nephrologie - Dialyse - Transplantation (SF-NDT). *Diabetes Metab*. 2019;45:224–237.
72. Lablanche S, Borot S, Wojtusciszyn A, et al. Five-year metabolic, functional, and safety results of patients with type 1 diabetes transplanted with allogenic islets within the Swiss-French GRAGIL network. *Diabetes Care*. 2015;38:1714–1722.
73. Badet L, Benhamou PY, Wojtusciszyn A, et al. Expectations and strategies regarding islet transplantation: metabolic data from the GRAGIL 2 trial. *Transplantation*. 2007;84:89–96.
74. Lablanche S, David-Tchouda S, Margier J, et al. Randomised, prospective, medico-economic nationwide French study of islet transplantation in patients with severely unstable type 1 diabetes: the STABILOT study protocol. *BMJ Open*. 2017;7:e013434.
75. Barton FB, Rickels MR, Alejandro R, et al. Improvement in outcomes of clinical islet transplantation: 1999-2010. *Diabetes Care*. 2012;35:1436–1445.
76. Shapiro AM, Ricordi C, Hering BJ, et al. International trial of the Edmonton protocol for islet transplantation. *N Engl J Med*. 2006;355:1318–1330.
77. Lablanche S, Vantyghem MC, Kessler L, et al. Islet transplantation versus insulin therapy in patients with type 1 diabetes with severe hypoglycaemia or poorly controlled glycaemia after kidney transplantation (TRIMECO): a multicentre, randomised controlled trial. *Lancet Diabetes Endocrinol*. 2018;6:527–537.
78. Fensom B, Harris C, Thompson SE, et al. Islet cell transplantation improves nerve conduction velocity in type 1 diabetes compared with intensive medical therapy over six years. *Diabetes Res Clin Pract*. 2016;122:101–105.
79. Warnock GL, Thompson DM, Meloche RM, et al. A multi-year analysis of islet transplantation compared with intensive medical therapy on progression of complications in type 1 diabetes. *Transplantation*. 2008;86:1762–1766.
80. Klee P, Dirlewanger M, Lavallard V, et al. Combined pancreatic islet-lung-liver transplantation in a pediatric patient with cystic fibrosis-related diabetes. *Horm Res Paediatr*. 2018;90:270–274.
81. Spijker HS, Wolffenbuttel BH, van der Bij W, et al. Islet-after-lung transplantation in a patient with cystic fibrosis-related diabetes. *Diabetes Care*. 2014;37:e159–160.
82. Kessler L, Bakopoulou S, Kessler R, et al. Combined pancreatic islet-lung transplantation: a novel approach to the treatment of end-stage cystic fibrosis. *Am J Transplant*. 2010;10:1707–1712.
83. Kessler L, Greget M, Metivier AC, et al. Combined pancreatic islets-lung transplantation in cystic fibrosis-related diabetes: case reports. *Transplant Proc*. 2010;42:4338–4340.
84. Koh A, Imes S, Shapiro AM, et al. Successful treatment of brittle diabetes following total pancreatectomy by islet allotransplantation: a case report. *JOP*. 2013;14:428–431.
85. Balamurugan AN, Naziruddin B, Lockridge A, et al. Islet product characteristics and factors related to successful human islet transplantation from the Collaborative Islet Transplant Registry (CITR) 1999-2010. *Am J Transplant*. 2014;14:2595–2606.
86. Markmann JF, Deng SP, Desai NM, et al. The use of non-heart-beating donors for isolated pancreatic islet transplantation. *Transplantation*. 2003;75:1423–1429.
87. Saito T, Gotoh M, Satomi S, et al. Islet transplantation using donors after cardiac death: report of the Japan Islet Transplantation Registry. *Transplantation*. 2010;90:740–747.
88. Lakey JR, Warnock GL, Rajotte RV, et al. Variables in organ donors that affect the recovery of human islets of Langerhans. *Transplantation*. 1996;61:1047–1053.
89. Ihm SH, Matsumoto I, Sawada T, et al. Effect of donor age on function of isolated human islets. *Diabetes*. 2006;55:1361–1368.
90. Niclauss N, Bosco D, Morel P, et al. Influence of donor age on islet isolation and transplantation outcome. *Transplantation*. 2011;91:360–366.
91. Brooks AM, Walker N, Aldibbiat A, et al. Attainment of metabolic goals in the integrated UK islet transplant program with locally isolated and transported preparations. *Am J Transplant*. 2013;13:3236–3243.
92. Goss JA, Goodpastor SE, Brunicardi FC, et al. Development of a human pancreatic islet-transplant program through a collaborative relationship with a remote islet-isolation center. *Transplantation*. 2004;77:462–466.
93. Goss JA, Schock AP, Brunicardi FC, et al. Achievement of insulin independence in three consecutive type-1 diabetic patients via pancreatic islet transplantation using islets isolated at a remote islet isolation center. *Transplantation*. 2002;74:1761–1766.
94. Kessler L, Bucher P, Milliat-Guittard L, et al. Influence of islet transportation on pancreatic islet allotransplantation in type 1 diabetic patients within the Swiss-French GRAGIL network. *Transplantation*. 2004;77:1301–1304.
95. O'Connell PJ, Holmes-Walker DJ, Goodman D, et al. Multicenter Australian trial of islet transplantation: improving accessibility and outcomes. *Am J Transplant*. 2013;13:1850–1858.
96. Ricordi C, Goldstein JS, Balamurugan AN, et al. National Institutes of Health-sponsored Clinical Islet Transplantation consortium phase 3 trial: manufacture of a complex cellular product at eight processing facilities. *Diabetes*. 2016;65:3418–3428.
97. Hering BJ, Kandaswamy R, Harmon JV, et al. Transplantation of cultured islets from two-layer preserved pancreases in type 1 diabetes with anti-CD3 antibody. *Am J Transplant*. 2004;4:390–401.
98. Bucher P, Mathe Z, Morel P, et al. Assessment of a novel two-component enzyme preparation for human islet isolation and transplantation. *Transplantation*. 2005;79:91–97.
99. Szot GL, Lee MR, Tavakol MM, et al. Successful clinical islet isolation using a GMP-manufactured collagenase and neutral protease. *Transplantation*. 2009;88:753–756.
100. Ricordi C, Lacy PE, Scharp DW. Automated islet isolation from human pancreas. *Diabetes*. 1989;38:140–142.
101. Markmann JF, Tomaszewski J, Posselt AM, et al. The effect of islet cell-culture in vitro at 24-degrees-C on graft-survival and MHC antigen expression. *Transplantation*. 1990;49:272–277.
102. Piemonti L, Leone BE, Nano R, et al. Human pancreatic islets produce and secrete MCP-1/CCL2: relevance in human islet transplantation. *Diabetes*. 2002;51:55–65.
103. Moberg L, Johansson H, Lukinius A, et al. Production of tissue factor by pancreatic islet cells as a trigger of detrimental thrombotic reactions in clinical islet transplantation. *Lancet*. 2002;360:2039–2045.
104. Bertuzzi F, Marzorati S, Maffi P, et al. Tissue factor and CCL2/monocyte chemoattractant protein-1 released by human islets affect islet engraftment in type 1 diabetic recipients. *J Clin Endocrinol Metab*. 2004;89:5724–5728.
105. Marzorati S, Antonioli B, Nano R, et al. Culture medium modulates proinflammatory conditions of human pancreatic islets before transplantation. *Am J Transplant*. 2006;6:2791–2795.
106. Bucher P, Mathe Z, Bosco D, et al. Morbidity associated with intraportal islet transplantation. *Transplant Proc*. 2004;36:1119–1120.

107. Owen RJ, Ryan EA, O'Kelly K, et al. Percutaneous transhepatic pancreatic islet cell transplantation in type 1 diabetes mellitus: radiologic aspects. *Radiology*. 2003;229:165–170.
108. Robertson RP. Islet transplantation as a treatment for diabetes - a work in progress. *N Engl J Med*. 2004;350:694–705.
109. Shapiro AM. Islet transplantation in type 1 diabetes: ongoing challenges, refined procedures, and long-term outcome. *Rev Diabet Stud*. 2012;9:385–406.
110. Cabric S, Sanchez J, Lundgren T, et al. Islet surface heparinization prevents the instant blood-mediated inflammatory reaction in islet transplantation. *Diabetes*. 2007;56:2008–2015.
111. Moberg L. The role of the innate immunity in islet transplantation. *Ups J Med Sci*. 2005;110:17–55.
112. Bennet W, Groth CG, Larsson R, et al. Isolated human islets trigger an instant blood mediated inflammatory reaction: implications for intraportal islet transplantation as a treatment for patients with type 1 diabetes. *Ups J Med Sci*. 2000;105:125–133.
113. Bellin MD, Clark P, Usmani-Brown S, et al. Unmethylated insulin DNA is elevated after total pancreatectomy with islet autotransplantation: assessment of a novel beta cell marker. *Am J Transplant*. 2017;17:1112–1118.
114. Kawahara T, Kin T, Kashkoush S, et al. Portal vein thrombosis is a potentially preventable complication in clinical islet transplantation. *Am J Transplant*. 2011;11:2700–2707.
115. Olsson R, Olerud J, Pettersson U, et al. Increased numbers of low-oxygenated pancreatic islets after intraportal islet transplantation. *Diabetes*. 2011;60:2350–2353.
116. Lee MH, Ward GM, MacIsaac RJ, et al. Mortality in people with type 1 diabetes, severe hypoglycemia, and impaired awareness of hypoglycemia referred for islet transplantation. *Transplant Direct*. 2018;4:e401.
117. Rickels MR, Liu C, Shlansky-Goldberg RD, et al. Improvement in β-cell secretory capacity after human islet transplantation according to the CIT07 protocol. *Diabetes*. 2013;62:2890–2897.
118. Foster ED, Bridges ND, Feurer ID, et al. Improved health-related quality of life in a phase 3 islet transplantation trial in type 1 diabetes complicated by severe hypoglycemia. *Diabetes Care*. 2018;41:1001–1008.
119. Meier JJ, Hong-McAtee I, Galasso R, et al. Intrahepatic transplanted islets in humans secrete insulin in a coordinate pulsatile manner directly into the liver. *Diabetes*. 2006;55:2324–2332.
120. Rickels MR, Kong SM, Fuller C, et al. Insulin sensitivity index in type 1 diabetes and following human islet transplantation: comparison of the minimal model to euglycemic clamp measures. *Am J Physiol Endocrinol Metab*. 2014;306:E1217–E1224.
121. Rickels MR, Markmann E, Naji A. Successful pregnancies after islet transplantation for type 1 diabetes. *Am J Transplant*. 2019;19:298–299.
122. Rickels MR, Fuller C, Dalton-Bakes C, et al. Restoration of glucose counterregulation by islet transplantation in long-standing type 1 diabetes. *Diabetes*. 2015;64:1713–1718.
123. Ang M, Meyer C, Brendel MD, et al. Magnitude and mechanisms of glucose counterregulation following islet transplantation in patients with type 1 diabetes suffering from severe hypoglycaemic episodes. *Diabetologia*. 2014;57:623–632.
124. Rickels MR, Peleckis AJ, Markmann E, et al. Long-term improvement in glucose control and counterregulation by islet transplantation for type 1 diabetes. *J Clin Endocrinol Metab*. 2016;101:4421–4430.
125. Naji A, Silvers WK, Bellgrau D, et al. Spontaneous diabetes in rats: destruction of islets is prevented by immunological tolerance. *Science*. 1981;213:1390–1392.
126. Sutherland DE, Goetz FC, Sibley RK. Recurrence of disease in pancreas transplants. *Diabetes*. 1989;38:85–87.
127. Huurman VA, Hilbrands R, Pinkse GG, et al. Cellular islet autoimmunity associates with clinical outcome of islet cell transplantation. *PloS One*. 2008;3:e2435.
128. Pascual J, Pirsch JD, Odorico JS, et al. Alemtuzumab induction and antibody-mediated kidney rejection after simultaneous pancreas-kidney transplantation. *Transplantation*. 2009;87:125–132.
129. Bazerbachi F, Selzner M, Boehnert MU, et al. Thymoglobulin versus basiliximab induction therapy for simultaneous kidney-pancreas transplantation: impact on rejection, graft function, and long-term outcome. *Transplantation*. 2011;92:1039–1043.
130. Bellin MD, Barton FB, Heitman A, et al. Potent induction immunotherapy promotes long-term insulin independence after islet transplantation in type 1 diabetes. *Am J Transplant*. 2012;12:1576–1583.
131. Lopez M, Clarkson MR, Albin M, et al. A novel mechanism of action for anti-thymocyte globulin: induction of CD4+CD25+Foxp3+ regulatory T cells. *J Am Soc Nephrol*. 2006;17:2844–2853.
132. Ricordi C, Zeng YJ, Alejandro R, et al. In vivo effect of FK506 on human pancreatic islets. *Transplantation*. 1991;52:519–522.
133. Drachenberg CB, Klassen DK, Weir MR, et al. Islet cell damage associated with tacrolimus and cyclosporine: morphological features in pancreas allograft biopsies and clinical correlation. *Transplantation*. 1999;68:396–402.
134. Naesens M, Kuypers DR, Sarwal M. Calcineurin inhibitor nephrotoxicity. *Clin J Am Soc Nephrol*. 2009;4:481–508.
135. Trinanes J, Ten Dijke P, Groen N, et al. Tacrolimus-induced BMP/SMAD signaling associates with metabolic stress-activated FOXO1 to trigger beta-cell failure. *Diabetes*. 2020;69:193–204.
136. Posselt AM, Szot GL, Frassetto LA, et al. Islet transplantation in type 1 diabetic patients using calcineurin inhibitor-free immunosuppressive protocols based on T-cell adhesion or costimulation blockade. *Transplantation*. 2010;90:1595–1601.
137. Posselt AM, Bellin MD, Tavakol M, et al. Islet transplantation in type 1 diabetics using an immunosuppressive protocol based on the anti-LFA-1 antibody efalizumab. *Am J Transplant*. 2010;10:1870–1880.
138. Stock PG, Mannon RB, Armstrong B, et al. Challenges of calcineurin inhibitor withdrawal following combined pancreas and kidney transplantation: results of a prospective, randomized clinical trial. *Am J Transplant*. 2020;20:1668–1978.
139. Berney T, Secchi A. Rapamycin in islet transplantation: friend or foe? *Transpl Int*. 2009;22:153–161.
140. Garrett GL, Blanc PD, Boscardin J, et al. Incidence of and risk factors for skin cancer in organ transplant recipients in the United States. *JAMA Dermatol*. 2017;153:296–303.
141. Senior PA, Zeman M, Paty BW, et al. Changes in renal function after clinical islet transplantation: four-year observational study. *Am J Transplant*. 2007;7:91–98.
142. Senior PA, Paty BW, Cockfield SM, et al. Proteinuria developing after clinical islet transplantation resolves with sirolimus withdrawal and increased tacrolimus dosing. *Am J Transplant*. 2005;5:2318–2323.
143. Leitao CB, Cure P, Messinger S, et al. Stable renal function after islet transplantation: importance of patient selection and aggressive clinical management. *Transplantation*. 2009;87:681–688.
144. Cure P, Pileggi A, Faradji RN, et al. Cytomegalovirus infection in a recipient of solitary allogeneic islets. *Am J Transplant*. 2006;6:1089–1090.
145. Shah AP, Chen JM, Fridell JA. Incidence and outcomes of cytomegalovirus in pancreas transplantation with steroid-free immunosuppression. *Clin Transplant*. 2015;29:1221–1229.
146. Dierickx D, Habermann TM. Post-transplantation lymphoproliferative disorders in adults. *N Engl J Med*. 2018;378:549–562.
147. Gala-Lopez BL, Senior PA, Koh A, et al. Late cytomegalovirus transmission and impact of T-depletion in clinical islet transplantation. *Am J Transplant*. 2011;11:2708–2714.
148. Issa N, Amer H, Dean PG, et al. Posttransplant lymphoproliferative disorder following pancreas transplantation. *Am J Transplant*. 2009;9:1894–1902.
149. Peters A, Olateju T, Deschenes J, et al. Posttransplant lymphoproliferative disorder after clinical islet transplantation: report of the first two cases. *Am J Transplant*. 2017;17:2474–2480.
150. Redfield RR, Kaufman DB, Odorico JS. Diagnosis and treatment of pancreas rejection. *Curr Transplant Rep*. 2015;2:169–175.
151. Brooks AM, Carter V, Liew A, et al. De novo donor-specific HLA antibodies are associated with rapid loss of graft function following islet transplantation in type 1 diabetes. *Am J Transplant*. 2015;15:3239–3246.
152. Rickels MR, Kamoun M, Kearns J, et al. Evidence for allograft rejection in an islet transplant recipient and effect on {beta}-cell secretory capacity. *J Clin Endocrinol Metab*. 2007;92:2410–2414.

153. Kessler L, Parissiadis A, Bayle F, et al. Evidence for humoral rejection of a pancreatic islet graft and rescue with rituximab and IV immunoglobulin therapy. *Am J Transplant*. 2009;9:1961–1966.
154. Moreau F, Toti F, Bayle F, et al. Rescue of a pancreatic islet graft after steroid therapy. *Transplantation*. 2012;93:e10–e11.
155. Burke 3rd GW, Vendrame F, Pileggi A, et al. Recurrence of autoimmunity following pancreas transplantation. *Curr Diab Rep*. 2011;11:413–419.
156. Vendrame F, Pileggi A, Laughlin E, et al. Recurrence of type 1 diabetes after simultaneous pancreas-kidney transplantation, despite immunosuppression, is associated with autoantibodies and pathogenic autoreactive CD4 T-cells. *Diabetes*. 2010;59:947–957.
157. Burke 3rd GW, Chen LJ, Ciancio G, et al. Biomarkers in pancreas transplant. *Curr Opin Organ Transplant*. 2016;21:412–418.
158. Korutla L, Rickels MR, Hu RW, et al. Noninvasive diagnosis of recurrent autoimmune type 1 diabetes after islet cell transplantation. *Am J Transplant*. 2019;19:1852–1858.
159. McEachron KR, Bellin MD. Total pancreatectomy and islet autotransplantion for chronic and recurrent acute pancreatitis. *Curr Opin Gastroenterol*. 2018;34:367–373.
160. Young MC, Theis JR, Hodges JS, et al. Preoperative computerized tomography and magnetic resonance imaging of the pancreas predicts pancreatic mass and functional outcomes after total pancreatectomy and islet autotransplant. *Pancreas*. 2016;45:961–966.
161. Trikudanathan G, Munigala S, Barlass U, et al. Evaluation of Rosemont criteria for non-calcific chronic pancreatitis (NCCP) based on histopathology—a retrospective study. *Pancreatology*. 2017;17:63–69.
162. Trikudanathan G, Vega-Peralta J, Malli A, et al. Diagnostic performance of endoscopic ultrasound (EUS) for non-calcific chronic pancreatitis (NCCP) based on histopathology. *Am J Gastroenterol*. 2016;111:568–574.
163. Trikudanathan G, Walker SP, Munigala S, et al. Diagnostic performance of contrast-enhanced MRI with secretin-stimulated MRCP for non-calcific chronic pancreatitis: a comparison with histopathology. *Am J Gastroenterol*. 2015;110:1598–1606.
164. Sutherland DER, Radosevich DM, Bellin MD, et al. Total pancreatectomy and islet autotransplantation for chronic pancreatitis. *J Am Coll Surg*. 2012;214:409–424.
165. Chinnakotla S, Beilman GJ, Dunn TB, et al. Factors predicting outcomes after a total pancreatectomy and islet autotransplantation lessons learned from over 500 cases. *Ann Surg*. 2015;262:610–622.
166. Kobayashi T, Manivel JC, Carlson AM, et al. Correlation of histopathology, islet yield, and islet graft function after islet autotransplantation in chronic pancreatitis. *Pancreas*. 2011;40:193–199.
167. Kobayashi T, Manivel JC, Bellin MD, et al. Correlation of pancreatic histopathologic findings and islet yield in children with chronic pancreatitis undergoing total pancreatectomy and islet autotransplantation. *Pancreas*. 2010;39:57–63.
168. Takita M, Lara LF, Naziruddin B, et al. Effect of the duration of chronic pancreatitis on pancreas islet yield and metabolic outcome following islet autotransplantation. *J Gastrointest Surg*. 2015;19:1236–1246.
169. Bellin MD, Prokhoda P, Hodges JS, et al. Age and disease duration impact outcomes of total pancreatectomy and islet autotransplant for PRSS1 hereditary pancreatitis. *Pancreas*. 2018;47:466–470.
170. Chinnakotla S, Radosevich DM, Dunn TB, et al. Long-term outcomes of total pancreatectomy and islet auto transplantation for hereditary/genetic pancreatitis. *J Am Coll Surg*. 2014;218:530–543.
171. Bellin MD, Forlenza GP, Majumder K, et al. Total pancreatectomy with islet autotransplantation resolves pain in young children with severe chronic pancreatitis. *J Pediatr Gastroenterol Nutr*. 2017;64:440–445.
172. Chinnakotla S, Bellin MD, Schwarzenberg SJ, et al. Total pancreatectomy and islet autotransplantation in children for chronic pancreatitis: indication, surgical techniques, postoperative management, and long-term outcomes. *Ann Surg*. 2014;260:56–64.
173. Morgan KA, Theruvath T, Owczarski S, et al. Total pancreatectomy with islet autotransplantation for chronic pancreatitis: do patients with prior pancreatic surgery have different outcomes? *Am Surg*. 2012;78:893–896.
174. Rickels MR, Norris AW, Hull RL. A tale of two pancreases: exocrine pathology and endocrine dysfunction. *Diabetologia*. 2020;63:2030–2039.
175. Bellin MD, Beilman GJ, Dunn TB, et al. Islet autotransplantation to preserve beta cell mass in selected patients with chronic pancreatitis and diabetes mellitus undergoing total pancreatectomy. *Pancreas*. 2013;42:317–321.
176. Balzano G, Maffi P, Nano R, et al. Diabetes-free survival after extended distal pancreatectomy and islet auto transplantation for benign or borderline/malignant lesions of the pancreas. *Am J Transplant*. 2019;19: 920–928.
177. Jin SM, Oh SH, Kim SK, et al. Diabetes-free survival in patients who underwent islet autotransplantation after 50% to 60% distal partial pancreatectomy for benign pancreatic tumors. *Transplantation*. 2013;95:1396–1403.
178. Ris F, Niclauss N, Morel P, et al. Islet autotransplantation after extended pancreatectomy for focal benign disease of the pancreas. *Transplantation*. 2011;91:895–901.
179. Thakor AS, Sangha BS, Ho SG, et al. Percutaneous autologous pancreatic islet cell transplantation for traumatic pancreatic injury. *J Clin Endocrinol Metab*. 2015;100:1230–1233.
180. Muratore S, Zeng X, Korc M, et al. Metastatic pancreatic adenocarcinoma after total pancreatectomy islet autotransplantation for chronic pancreatitis. *Am J Transplant*. 2016;16:2747–2752.
181. Balzano G, Maffi P, Nano R, et al. Autologous islet transplantation in patients requiring pancreatectomy: a broader spectrum of indications beyond chronic pancreatitis. *Am J Transplant*. 2016;16:1812–1826.
182. Lund A, Bagger JI, Wewer Albrechtsen NJ, et al. Evidence of extrapancreatic glucagon secretion in man. *Diabetes*. 2016;65:585–597.
183. Wilhelm JJ, Bellin MD, Dunn TB, et al. Proposed thresholds for pancreatic tissue volume for safe intraportal islet autotransplantation after total pancreatectomy. *Am J Transplant*. 2013;13:3183–3191.
184. Fan CJ, Hirose K, Walsh CM, et al. Laparoscopic total pancreatectomy with islet autotransplantation and intraoperative islet separation as a treatment for patients with chronic pancreatitis. *JAMA Surg*. 2017;152:550–556.
185. Barrera KSS, Schwartzman A, Gruessner RWG. Surgical techniques for total pancreatectomy and islet autotransplantation. In: *Transplantation, Bioengineering, and Regeneration of the Endocrine Pancreas*. Vol 2. Academic Press; 2020:101–116.
186. Kesseli SJ, Wagar M, Jung MK, et al. Long-term glycemic control in adult patients undergoing remote vs. local total pancreatectomy with islet autotransplantation. *Am J Gastroenterol*. 2017;112:643–649.
187. Tai DS, Shen N, Szot GL, et al. Autologous islet transplantation with remote islet isolation after pancreas resection for chronic pancreatitis. *JAMA Surg*. 2015;150:118–124.
188. Johnston PC, Lin YK, Walsh RM, et al. Factors associated with islet yield and insulin independence after total pancreatectomy and islet cell autotransplantation in patients with chronic pancreatitis utilizing off-site islet isolation: Cleveland Clinic experience. *J Clin Endocrinol Metab*. 2015;100:1765–1770.
189. Balamurugan AN, Loganathan G, Bellin MD, et al. A new enzyme mixture to increase the yield and transplant rate of autologous and allogeneic human islet products. *Transplantation*. 2012;93:693–702.
190. Morgan KA, Nishimura M, Uflacker R, et al. Percutaneous transhepatic islet cell autotransplantation after pancreatectomy for chronic pancreatitis: a novel approach. *HPB (Oxford)*. 2011;13:511–516.
191. Bellin MD, Beilman GJ, Sutherland DE, et al. How durable is total pancreatectomy and intraportal islet cell transplantation for treatment of chronic pancreatitis? *J Am Coll Surg*. 2019;228:329–339.
192. Abu-El-Haija M, Anazawa T, Beilman GJ, et al. The role of total pancreatectomy with islet autotransplantation in the treatment of chronic pancreatitis: a report from the International Consensus Guidelines in chronic pancreatitis. *Pancreatology*. 2020;20:762–771.
193. Bogachus LD, Oseid E, Bellin M, et al. Deficient endogenous glucose production during exercise after total pancreatectomy/islet autotransplantation. *J Clin Endocrinol Metab*. 2017;102:3288–3295.
194. Bogachus LD, Bellin MD, Vella A, et al. Deficient glucagon response to hypoglycemia during a mixed meal in total pancreatectomy/islet autotransplantation recipients. *J Clin Endocrinol Metab*. 2018;103:1522–1529.

195. Bellin MD, Parazzoli S, Oseid E, et al. Defective glucagon secretion during hypoglycemia after intrahepatic but not nonhepatic islet autotransplantation. *Am J Transplant*. 2014;14:1880–1886.
196. Robertson RP, Bogachus LD, Oseid E, et al. Assessment of beta-cell mass and alpha- and beta-cell survival and function by arginine stimulation in human autologous islet recipients. *Diabetes*. 2015;64:565–572.
197. Baidal DA, Ricordi C, Berman DM, et al. Bioengineering of an intraabdominal endocrine pancreas. *N Engl J Med*. 2017;376:1887–1889.
198. Lindahl JP, Reinholt FP, Eide IA, et al. In patients with type 1 diabetes simultaneous pancreas and kidney transplantation preserves long-term kidney graft ultrastructure and function better than transplantation of kidney alone. *Diabetologia*. 2014;57:2357–2365.
199. Jenssen T, Hartmann A, Birkeland KI. Long-term diabetes complications after pancreas transplantation. *Curr Opin Organ Transplant*. 2017;22:382–388.
200. Giannarelli R, Coppelli A, Sartini MS, et al. Pancreas transplant alone has beneficial effects on retinopathy in type 1 diabetic patients. *Diabetologia*. 2006;49:2977–2982.
201. Koznarová R, Saudek F, Sosna T, et al. Beneficial effect of pancreas and kidney transplantation on advanced diabetic retinopathy. *Cell Transplant*. 2000;9:903–908.
202. Navarro X, Sutherland DE, Kennedy WR. Long-term effects of pancreatic transplantation on diabetic neuropathy. *Ann Neurol*. 1997;42:727–736.
203. Navarro X, Kennedy WR, Loewenson RB, et al. Influence of pancreas transplantation on cardiorespiratory reflexes, nerve conduction, and mortality in diabetes mellitus. *Diabetes*. 1990;39:802–806.
204. Lindahl JP, Hartmann A, Aakhus S, et al. Long-term cardiovascular outcomes in type 1 diabetic patients after simultaneous pancreas and kidney transplantation compared with living donor kidney transplantation. *Diabetologia*. 2016;59:844–852.
205. Jukema JW, Smets YF, van der Pijl JW, et al. Impact of simultaneous pancreas and kidney transplantation on progression of coronary atherosclerosis in patients with end-stage renal failure due to type 1 diabetes. *Diabetes Care*. 2002;25:906–911.
206. Vantyghem MC, Quintin D, Caiazzo R, et al. Improvement of electrophysiological neuropathy after islet transplantation for type 1 diabetes: a 5-year prospective study. *Diabetes Care*. 2014;37:e141–142.
207. Hering BJ, Wijkstrom M, Graham ML, et al. Prolonged diabetes reversal after intraportal xenotransplantation of wild-type porcine islets in immunosuppressed nonhuman primates. *Nat Med*. 2006;12:301–303.
208. Bottino R, Knoll MF, Graeme-Wilson J, et al. Safe use of anti-CD154 monoclonal antibody in pig islet xenotransplantation in monkeys. *Xenotransplantation*. 2017;24. Available from: https://doi.org/10.1111/xen.12283.
209. Thompson P, Badell IR, Lowe M, et al. Islet xenotransplantation using gal-deficient neonatal donors improves engraftment and function. *Am J Transplant*. 2011;11:2593–2602.
210. Niu D, Wei HJ, Lin L, et al. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science*. 2017;357:1303–1307.
211. Valdes-Gonzalez R, Dorantes LM, Bracho-Blanchet E, et al. No evidence of porcine endogenous retrovirus in patients with type 1 diabetes after long-term porcine islet xenotransplantation. *J Med Virol*. 2010;82:331–334.
212. Matsumoto S, Abalovich A, Wechsler C, et al. Clinical benefit of islet xenotransplantation for the treatment of type 1 diabetes. *E Bio Med*. 2016;12:255–262.
213. Rezania A, Bruin JE, Arora P, et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotechnol*. 2014;32:1121–1133.
214. Rezania A, Bruin JE, Riedel MJ, et al. Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. *Diabetes*. 2012;61:2016–2029.
215. van der Torren CR, Zaldumbide A, Duinkerken G, et al. Immunogenicity of human embryonic stem cell-derived beta cells. *Diabetologia*. 2017;60:126–133.
216. Millman JR, Xie C, Van Dervort A, et al. Generation of stem cell-derived beta-cells from patients with type 1 diabetes. *Nat Commun*. 2016;7:11463.

42

Diabetes and Long-Term Complications

Enrico Cagliero

OUTLINE

History, 730
Classification, 730
  Rheumatic, Dermatologic, and Infectious Manifestations, 730
Epidemiology, 731
Pathophysiology, 732
Role of Glucose Control in the Prevention of Diabetic Complications, 732
Role of Glucose Control in the Prevention of Cardiovascular Disease, 735
Can We Prevent Macrovascular Disease?, 735
Effects of Specific Hypoglycemic Agents on Cardiovascular Disease and Renal Outcomes in Patients With Type 2 Diabetes, 737
Conclusion, 737

HISTORY

Although the term "diabetes" was initially used by Aretaeus of Cappadocia, the first Greek physician to accurately describe the symptoms of diabetes, in the second century of the Common Era, the first known description of the disease is in the Ebers Papyrus, dating 1552 BCE.[1] It was not until the nineteen century that the initial descriptions of diabetic complications started to appear in the literature,[2] but several characteristic diabetic complications started to gain recognition with the discovery of insulin and the long-term survival of diabetic patients.[3-5] Initially, patients with longstanding diabetes were observed to have a high incidence of heart disease and atherosclerosis,[2,6] but in the 1950s the concept of a specific diabetic angiopathy was proposed, consisting of retinopathy, nephropathy, coronary artery disease (CAD), and peripheral vascular disease,[6] and the frequent triad of retinopathy, nephropathy, and neuropathy was also described.[7] Multiple theories were proposed to account for these vascular lesions observed in diabetic patients, however the observation that patients with "poor control" (as defined by symptoms of polydipsia and polyuria, frequent glycosuria, and ketosis) had a higher incidence of complications strongly suggested that these complications could be prevented by increasing the control of the metabolic disturbances associated with diabetes.[8,9] With these data and clinical observations in mind, Dr. Elliott P. Joslin already wrote in 1950 that "these young patients must be told that if they do not keep their diabetes controlled they are liable to serious complications in their eyes, their kidneys, their hearts or their brains."[10] However, debate regarding the role of diabetic control, or, more precisely, glycemic control, on the pathogenesis of diabetic complication persisted for several decades. Until the late 1970s groups of investigators were openly debating whether good diabetic control had any role in the prevention of the vascular complications of diabetes.[11,12]

CLASSIFICATION

Long-term complications of diabetes can be divided into *microvascular*, or diabetes-specific, due to damage to capillaries and small vessels (retinopathy, nephropathy, and neuropathy, described in detail later in Chapters 43, 44, and 45) and *macrovascular*, reflecting the increased risk of accelerated atherosclerosis with its clinical manifestations such as CAD, cerebrovascular accident (CVA), and peripheral vascular disease. These complications are responsible for the increased morbidity and mortality associated with both type 1 and type 2 diabetes.[13-19] However, other long-term complications of diabetes, including certain rheumatic manifestations, increased susceptibility to some infections, and some skin disorders, have also been reported, some with higher incidence in patients with diabetes and others that occur almost exclusively in diabetic patients. Although these complications are less well known, they can have a significant impact on morbidity and mortality, with survival studies having shown an increased incidence of death from cancer [19] and infectious diseases[19,20] in diabetic patients.

Rheumatic, Dermatologic, and Infectious Manifestations

Several rheumatic manifestations of diabetes mellitus have been described, affecting up to 30% of diabetic patients.[21] Although the majority of these syndromes are not specific to diabetic patients, but occur more frequently in them (e.g., adhesive capsulitis of the shoulder, Dupuytren disease, flexor tenosynovitis), diabetic cheiroarthopathy or "syndrome of limited joint mobility" is a diabetes-specific complication characterized by painless limitations of multiple joints in the hand and skin changes similar to those seen in scleroderma, with thick and rigid skin.[21,22]

The increased susceptibility to infections in patients with diabetes was already described by Elliott Joslin in 1950,[10] but more contemporary studies have shown that diabetic patients are more prone to certain specific infections and organisms, among them respiratory tract infections (*Staphylococcus aureus, Streptococcus pneumoniae, Mycobacterium tuberculosis*) and urinary tract infections (bacteriuria and upper urinary tract infections, as well as fungal infections).[23] The reasons for this predisposition are not known, however several studies, mostly in animal models of diabetes, have focused on impaired neutrophil function; the data are inconclusive at best.[24] Fournier gangrene, a fulminant form of infective necrotizing fasciitis of the perineal, genital, or perianal regions, as well as other forms of necrotizing fasciitis have diabetes as the major risk factor,[25] and an increased prevalence and severity of periodontal disease has been described in diabetic patients.[26] Some infections seem to occur almost exclusively in patients with diabetes, including invasive (or malignant) otitis externa caused by *Pseudomonas aeruginosa* in most cases, rhinocerebral mucormycosis caused by *Rhizopus oryzae*, with ketoacidosis being the most important risk factor, and emphysematous infections such as emphysematous cholecystitis and pyelonephritis.[23]

Necrobiosis lipoidica is a characteristic skin lesion associated with diabetes and consists of bilateral, hyperpigmented plaques on the anterior tibial surface with telangiectasias and violaceous/brown periphery. Together with bullosis diabeticorum (asymptomatic bullae containing sterile fluid, usually present on the dorsa and sides of lower legs and feet), they are cutaneous manifestations of diabetes. Moreover, patients with diabetes also have a higher prevalence of other dermatologic conditions, including ichthyosiform skin changes, thick, scleroderma-like changes of the hands and fingers, and toenail onychomycosis.[27,28]

KEY POINTS

- Diabetic complications can be divided into microvascular (retinopathy, nephropathy, and neuropathy) and macrovascular (increased incidence of coronary artery disease, cerebrovascular disease, and peripheral arterial disease).
- Diabetes is also associated with rheumatic (adhesive capsulitis, Dupuytren disease, and flexor tenosynovitis), dermatologic (necrobiosis lipoidica and bullosis diabeticorum), and infectious manifestations (Fournier gangrene, malignant otitis externa, and rhinocerebral mucormycosis).

EPIDEMIOLOGY

The prevalence of the microvascular complications of diabetes is associated with duration of disease and degree of glycemic control. An early prospective study of more than 4000 diabetic patients followed for 25 years in Brussels, Belgium showed that the prevalence of retinopathy and neuropathy in patients who had had diabetes for 25 years was more than 50%, whereas the prevalence of diabetic nephropathy was significantly lower, at 14%.[29] More recent epidemiological studies have confirmed these initial data: after several decades of diabetes, the majority of patients will have at least some mild retinopathy and neuropathy, whereas only a subset of genetically predisposed subjects will develop diabetic nephropathy.

Large epidemiological and clinical trial studies have shown that the prevalence of any diabetic retinopathy approximates 80% after 30 years of diabetes, whereas the prevalence of vision-threatening proliferative retinopathy is around 30% to 50%.[30,31] These values are heavily influenced by the degree of metabolic control, as shown by the Diabetes Control and Complication Trial (DCCT), with 50% of type 1 diabetic patients having proliferative retinopathy after 30-year disease duration in the conventional treatment cohort and 21% in the intensive therapy cohort.[31] The data for peripheral neuropathy are similar. The DCCT experience showed that, after 25 years of type 1 diabetes, 61% of patients had abnormal nerve conduction, and 37% of patients had clinically significant neuropathy,[32] a percentage similar to other epidemiological studies in Europe (28% in the EURODIAB IDDM Complications study and 37% in a large multicenter study conducted in the United Kingdom).[33,34]

The epidemiology of diabetic nephropathy is significantly different from that of the other microvascular complications of diabetes, with fewer diabetic patients being affected by renal problems than other complications. Data from the DCCT/Epidemiology of Diabetes Interventions (EDIC) study in young type 1 diabetic patients show that, after 30 years of type 1 diabetes, only 25% of conventional treatment patients and 9% of intensive therapy patients showed microalbuminuria or worse renal outcomes.[35] Data from type 2 diabetic patients older than 40 years of age in the United States showed a frequency of microalbuminuria of 35% versus 12% of the nondiabetic population.[36] Although it is not clear why there is a difference in the epidemiology of these microvascular complications, evidence from familial aggregation studies, ethnic specific prevalence rates, and, more recently, genetic studies, has shown that the development of diabetic complications is at least partially under the control of genetic factors, with several candidate genes having been identified.[37,38]

It is well known that cardiovascular disease (CVD) is the leading cause of death in patients with both type 1 and type 2 diabetes, and that the prevalence of CVD is increased by the disease. CAD, peripheral artery disease, and CVD rates are increased 2- to 4-fold in diabetic patients, with a higher relative risk in women than in men. Moreover, in most studies, clinical outcomes of CVD are worse in patients with diabetes as compared with nondiabetic subjects.[39,40] Whereas type 2 diabetes is associated with metabolic changes that are known to have a proatherosclerotic profile (obesity, hypertension, dyslipidemia), recent studies have shown that the risk of CVD is similar in type 1 and type 2 diabetes, although the relative role of diabetic-specific risks could be different in type 1 versus type 2 diabetes.[41,42] In type 1 diabetes the essential role of diabetic nephropathy as a risk factor for CVD is well established and has been demonstrated in several studies that have shown mortality in type 1 diabetic patients without microalbuminuria or more severe stages of renal disease similar to that observed in the nondiabetic population.[43,44]

As a consequence of the high rate of CVD and overall diabetic complications, mortality rates are increased in diabetic patients, both in the younger and older populations. In the Framingham Heart Study, individual with diabetes and a mean age of 60 years were shown to have a 2-fold increase in all-cause mortality and a 3-fold increase in CVD mortality compared with the nondiabetic population.[45] A recent study in Sweden of young diabetic patients (15 to 34 years of age) showed a 2- to 3-fold increase in mortality rates.[46] There is a significant difference in type 1 versus type 2 diabetes, with type 2 diabetic patients having a double hazard ratio (HR) for total mortality as compared with age-matched type 1 diabetic patients,[46,47] and this is likely because of the marked excess in macrovascular diabetic complications found in the type 2 diabetic population.[47]

Significant improvements in diabetes management over the last several years regarding glycemic control and cardiovascular risk factors management have resulted in a changing pattern of diabetic complications. Observational and epidemiological studies have demonstrated a decreased incidence of severe retinopathy, nephropathy and end-stage renal disease, neuropathy, CVD, and overall mortality in patients with a recent diagnosis of diabetes as compared with patients with an earlier diagnosis, likely the result of decreases in glycated hemoglobin (HbA1c), blood pressure, low-density lipoprotein (LDL) cholesterol, and smoking status.[48-51] Unfortunately, while the incidence of advanced complications is decreasing, there is an epidemic of obesity and type 2 diabetes, which means that the overall numbers of diabetic patients reaching the stage of late microvascular and macrovascular complications is increasing, putting a strain on healthcare resources.

KEY POINTS

- Although diabetic retinopathy and neuropathy are present in most patients with more than 25 years of diabetes, diabetic nephropathy is only present is less than 25% of patients with long diabetes duration and is likely under the control of genetic factors.
- The prevalence of cardiovascular disease is increased by 2- to 4-fold in patients with diabetes, with a higher relative risk in women than men.

PATHOPHYSIOLOGY

It has long been known that exposure to the diabetic milieu is responsible for the pathogenesis of the microvascular complications of diabetes, and among the multiple metabolic abnormalities observed in diabetic patients chronic exposure to high glucose levels is the common feature of all forms of diabetes. Animal models of diabetes have also helped to support the hypothesis that high glucose levels are responsible for the vascular complications. Galactose-fed dogs develop retinopathy that is histologically identical to diabetic retinopathy, thus suggesting that high hexose levels are by themselves able to cause retinopathy[52]; and good glucose control achieved by insulin injections or islet cell transplantation reduces the incidence and severity of retinopathy and nephropathy in dogs and rats, respectively.[53-54]

To elucidate the mechanisms by which high glucose levels are toxic to the small vessels, multiple animal models have been studied, with the most commonly used ones being rodents. Rodents made diabetic by the use of the toxic drugs streptozotocin or alloxan or by pancreatectomy are useful models of insulin-deficient diabetes. Genetically altered models such as db/db mice (leptin receptor mutation) and ob/ob mice (leptin mutation) have been used to mimic the abnormalities seen in obesity and type 2 diabetes. Although these rodent model have been very useful in the study of diabetic complications, they do not completely resemble their human counterparts. Rodents do not usually exhibit the more advanced lesions of diabetic nephropathy (nodular glomerulosclerosis), retinopathy (preretinal neovascularization and retinal neuron degeneration), or neuropathy (segmental demyelination and fiber loss).[55] Larger animals have also been studied as models for the vascular complications of diabetes, including dogs (alloxan, pancreatectomy, or galactose-fed), cats, pigs, and nonhuman primates, animals that have a macula like humans and can be used for the study of diabetic maculopathy. However, these larger animal models of diabetic complications have disadvantages, such as long time lags before the development of diabetic lesions (studies in dogs have been performed after 5 years of diabetes and in monkeys up to 15 years later) and the absence of advanced retinopathy (no neovascularization or neuronal damage).[56]

Although there is universal agreement that high glucose levels cause the microvascular lesions of diabetes, the exact molecular mechanisms by which glucose levels induce vascular damage are still not completely understood. It is well known that several molecular pathways are activated by high glucose levels in cell cultures or animal models of diabetes, including increased flux in the polyol pathway, increased formation of advanced glycation end-products (AGEs), activation of protein kinase C (PKC), and increased flux in the hexosamine pathway.[57] More recently, a common element between these pathways has been found in the increased production of reactive oxygen species (ROS) in the mitochondria and overall increased oxidative stress induced by hyperglycemia[57-59]; these changes have been linked to the activation of transforming growth factor β signaling, decreased activity of endothelial nitric oxide synthase, decreased activity of prostacyclin synthase, and increased activity of several proinflammatory pathways.[58-59]

The role of increased mitochondrial ROS production in explaining the phenomenon called *metabolic memory* (another phenomenon in the pathogenesis of diabetic vascular complications) has also been recently postulated and supported by experimental data. It has long been known that high glucose levels in tissue culture lead to changes in gene expression that are then perpetuated *in vitro*.[60] In experimental animal models of diabetes, the same phenomenon has been described, with retinopathy progressing even with good glycemic control after an initial period of poor control.[61] Recent data from randomized clinical trials such as the DCCT have shown that the benefits from tight glucose control in terms of reduction in diabetic complications persist for several years, thus confirming that glycemic or metabolic memory also occurs in clinical diabetes.[62,63] Epigenetic modifications such as DNA methylation and histone posttranslational modifications can be induced by oxidative stress and have been detected in models of diabetic complications.[64,65] More recently, epigenetic changes in lymphocytes and monocytes from DCCT patients have been reported, thus confirming the potential role of high glucose–induced changes in chromatin structure in the pathogenesis of the clinical complications of diabetes.[66]

Specific inhibitors of these pathways have been used for more than 30 years in an attempt to find pharmacological agents that could prevent or delay the development of the vascular complications of diabetes. Although successful in preclinical animal studies, the clinical trials of such agents in diabetic patients have been disappointing, and none of them has made it to clinical use. Clinical trials of aldose reductase inhibitors have been performed for more than 30 years, but a recent metaanalysis showed no statistically significant difference in placebo-controlled trials of diabetic neuropathy.[67] AGE inhibitors have also been studied for many years, again with either significant side effects or no efficacy.[68,69] A PKC inhibitor was also shown not to have statistically significant efficacy in diabetic retinopathy in phase 3 studies[70] and was dropped from commercial development. More recently, benfotiamine, an agent shown to inhibit three of the proposed pathways in the pathogenesis of diabetic complications (hexosamine pathways, PKC activation, and AGE formation) was also found to have no effect in diabetic neuropathy.[71] While research in this field continues, these results have underlined the difficulties of translating results from animal models of diabetic complications to clinical use, as well as the long timeline involved in bringing new agents to therapeutic use.

KEY POINTS

- High glucose levels lead to multiple biochemical changes in vascular cells, including increased production of reactive oxygen species.
- Specific inhibitors of these biochemical pathways have been studied in animal models of diabetes but have not been successful in preventing vascular disease in patients with diabetes.

ROLE OF GLUCOSE CONTROL IN THE PREVENTION OF DIABETIC COMPLICATIONS

Poor glycemic control has been associated in early studies with increased severity of microvascular complications,[29] but it was not until the early 1980s that, with the introduction of home glucose monitoring, widespread use of HbA1c measurements, and the use of insulin pump therapies, some initial small clinical trials suggested the possibility that tighter glucose control could prevent or delay the vascular complications of diabetes.[72] However, it was the seminal DCCT, whose results were published in 1993, that definitively established the essential role of tight glucose control in preventing or delaying the microvascular complications of diabetes.[73] This study, initiated in 1983, was designed to answer the question of whether intensive diabetes management would affect the development (in a primary prevention cohort of 726 type 1 diabetic patients with no complications and only 1–5 years of diabetes duration) or delay the progression (in a secondary intervention cohort of 715 type 1 diabetic patients with at least one microaneurysm but no more than moderate nonproliferative retinopathy and duration of diabetes of 1–15 years) of the microvascular or neurological complications of diabetes. Patients were randomly assigned to conventional treatment (in 1983 defined as 1–2 daily injections of insulin, daily blood glucose monitoring, and the goal of avoiding hypoglycemia or hyperglycemia symptoms) or intensive therapy (goal of achieving blood glucose levels as close to nondiabetic range as possible, including fasting values of 70–120 mg/dL, postprandial levels <180 mg/dL, and HbA1c levels in the nondiabetic range). To achieve these goals, intensive therapy patients used three or more insulin injections per day or insulin pump therapy. Over the 6.5-year mean follow-up time, compliance was excellent, with subjects adhering to assigned treatment 97% of study time and more than 99% of subjects completing the trial. The separation between groups in terms of HbA1c was excellent, at approximately 2%, with the conventional treatment group maintaining their HbA1c at approximately 9%, while the intensive therapy group achieved a nadir HbA1c of 6.9% at 6 months and then maintained it at approximately 7% for the remainder of the time (Fig. 42.1).

The results published in the initial report showed a 76% and 54% risk reduction for retinopathy in the primary prevention and secondary intervention groups, respectively (Fig. 42.2) and a risk reduction of 39% for microalbuminuria, 54% for albuminuria, and 60% for clinical neuropathy, at the cost of a 3-fold increase incidence of severe hypoglycemia.[73,74] Although 65% of patients in the intensive therapy group had at least one episode of severe hypoglycemia by study end, as compared with 35% of patients in the conventional therapy group,[74] the data showed no deterioration in quality of life for patients undergoing intensive treatment.[75] More detailed analysis of diabetic complications showed favorable results in almost all subgroups of patients, as defined by gender, age, and other baseline characteristics, in terms of retinopathy, nephropathy, peripheral neuropathy, and autonomic neuropathy.[76-79] The small number of macrovascular events was not statistically decreased by intensive therapy, but some cardiovascular risk factors improved.[80] Moreover, the mean HbA1c levels during the trial were the most significant predictor for the development of complications, with no level of glycemic control below which the risk of complications was eliminated, and there was a 44% risk reduction in the development of retinopathy for each 10% decrease in HbA1c levels (Fig. 42.3).[81]

At the end of the DCCT trial, all participants in the conventional treatment arm were offered training in intensive therapy, and then diabetes care was returned to the subjects' own doctors. The entire group was followed in the EDIC observational study, which started in 1994 and published the 30-year follow-up data.[82] Although HbA1c was nicely separated in the two groups at the end of the DCCT (mean 9.1% in the conventional therapy group vs. mean 7.2% in the intensive therapy group), the difference in HbA1c between the two groups disappeared during the 20 years of EDIC follow-up, and the mean HbA1c overall for both groups at the end of the EDIC study was 8.0%. The significant differences in complications observed at the end of the DCCT and the subsequent disappearance of the difference in glycemic control between the two groups would have logically predicted a parallel development in complications during the

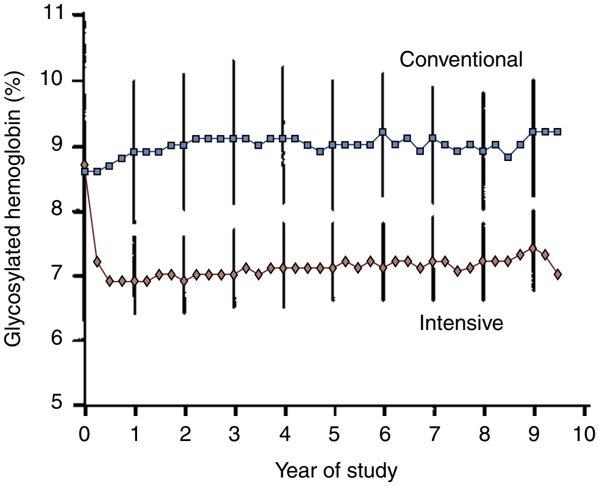


Fig. 42.1 Glycated hemoglobin results from the Diabetes Control and Complications Trial (DCCT). Medians of all quarterly measurements with the 25th and 75th percentiles of the yearly values (*vertical lines*) are shown. The differences between treatment groups became significant by 3 months and remained significantly different over the course of the study (*P* < .001). (From Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med.* 1993;329:977–986.)

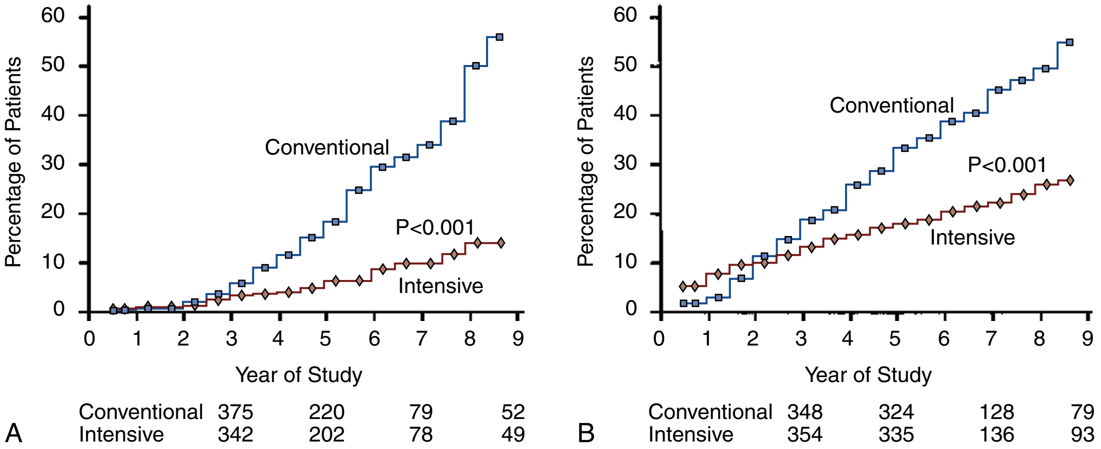


Fig. 42.2 Cumulative incidence of Diabetes Control and Complications Trial primary retinopathy outcomes (three-step change in retinopathy from baseline value, sustained for at least 6 months) measured by fundus photography, comparing intensive and conventional therapy groups. **A,** In the primary prevention cohort, intensive therapy reduced the cumulative incidence by 76% (*P* < .001). **B,** In the secondary intervention cohort, intensive therapy reduced the cumulative incidence by 54% (*P* < .001). (From Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med.* 1993;329:977–986.)

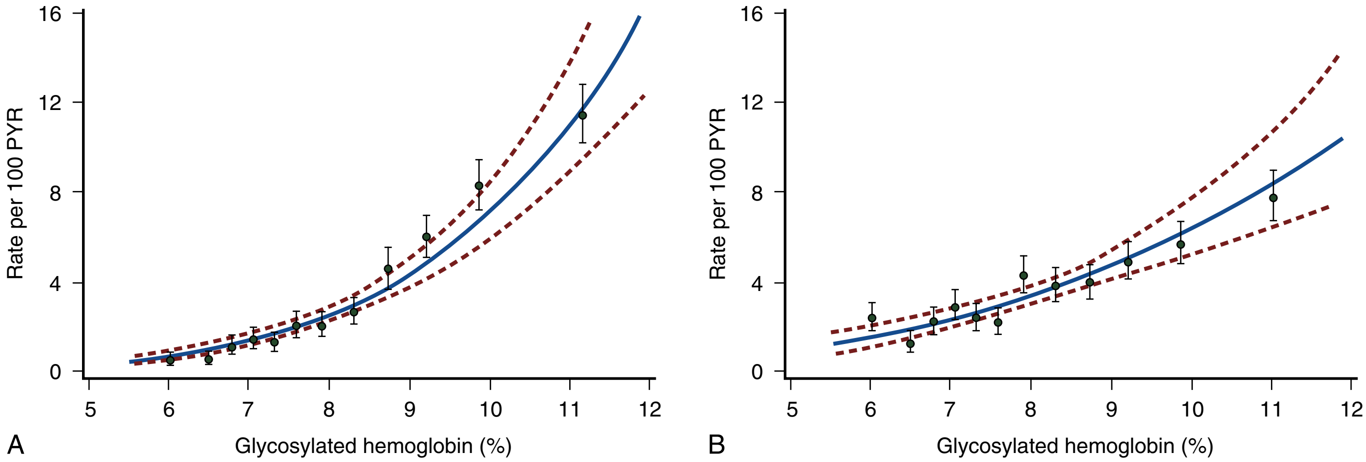


Fig. 42.3 Relationship between updated mean glycated hemoglobin levels and complications in the Diabetes Control and Complications Trial, with examination of combined treatment groups with Poisson regression models (95% confidence bands shown with *dashed lines*). The relationship between mean HbA1c and complications is continuous over the entire range of HbA1c levels, with no inflection or break point at levels of HbA1c above the nondiabetic range. **A,** Risk for sustained (for at least 6 months) retinopathy progression measured as rate per 100 patient-years (PYR). **B,** Risk for developing microalbuminuria (rate per 100). (From Diabetes Control and Complications Trial Research Group. The absence of a glycemic threshold for the development of long-term complications: the perspective of the Diabetes Control and Complications Trial. *Diabetes*. 1996;45:1289–1298.)

EDIC observational study. However, during the initial years of the EDIC there was a further widening in the differences in microvascular outcomes, referred to as “the metabolic memory phenomenon.”[83] The most recent data from the DCCT/EDIC study show that, after a 30-year duration of diabetes, good glycemic control during the interventional trial resulted in an approximately 50% reduction in microvascular complications[84,85] (Fig. 42.4) and a 57% reduction in clinically significant cardiovascular events (nonfatal myocardial infarction, stroke, or death from CVD).[86]

Whether the lessons from the DCCT/EDIC could be applied to type 2 diabetic patients, and whether different therapies for the management of type 2 diabetes would provide advantage in terms of reduced complications was the goal of the seminal United Kingdom Prospective Diabetes Study (UKPDS), which was planned in the 1970s and published in 1998.[87,88] Although a much smaller clinical trial in a Japanese population of type 2 diabetic patients showed results almost identical to the DCCT trial,[89] the patient population (lean type 2 diabetic patients using much less insulin than the DCCT cohorts used) was thought not to be representative of the majority of type 2 diabetic patients. The UKPDS study randomized approximately 4000 newly diagnosed type 2 diabetic patients to conventional treatment (dietary intervention) versus intensive treatment with one of three different sulphonylureas, insulin, or metformin. However, because of the progressive β cell dysfunction characteristic of type 2 diabetes[90] and the stepwise design of the study, a significant proportion of patients initially assigned to diet, sulphonylureas, or metformin ended the study with another added medication, making the assessment of whether one therapy was better than other impossible. The study achieved a 1% separation in HbA1c levels between the conventional and intensive therapy groups. However, after an initial decline from a baseline of 7%, HbA1c levels rose continuously over the 10 years following randomization. Although the median HbA1c over the 10-year study was 7% versus 7.9%, during the last period the HbA1c had increased to 8.1% and 8.7% in the intensive and conventional groups, respectively.[87] The study design and the data analysis were somewhat controversial,[91] however, the overall findings showed that intensive treatment improved microvascular outcomes (driven by a decrease in the need for retinal photocoagulation) but did not affect cardiovascular outcomes, and that there were no clear differences between the different intensive therapies. Of interest, a similar relationship between glycemic control and retinopathy was observed in the UKPDS and in the DCCT; for each 1% decrease in HbA1c, there was a 39% decrease in the risk for retinopathy.[92]

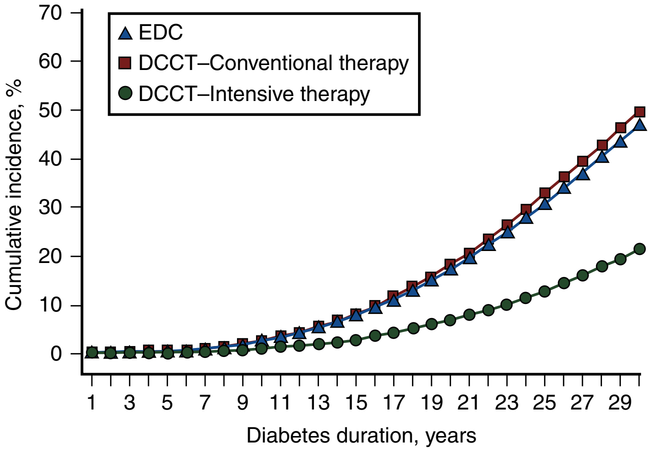


Fig. 42.4 Cumulative incidence of any major eye disease end point in relation to diabetes duration. Diabetes Control and Complications Trial (DCCT) conventional therapy groups (*squares*) and intensive therapy groups (*circles*) are presented. Also presented is the cumulative incidence of these major eye disease points observed in the observational Pittsburgh Epidemiology of Diabetes Complication (EDC) study (*triangles*). (From Nathan DM, Zinman B, Cleary PA, et al.; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Research Group. Modern-day clinical course of type 1 diabetes mellitus after 30 years’ duration: the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications and Pittsburgh Epidemiology of Diabetes Complications experience (1983– 2005). Arch Intern Med 2009;169:1307–1316.)

At the end of the UKPDS study, follow-up questionnaires were sent annually for 10 years, and, although differences in the HbA1c levels between the two groups were lost at the end of the study, the risk reduction in terms of microvascular events remained significant. A reduction in myocardial infarction and death from any cause became significant, likely secondary to the increased number of events.[93] Thus, a "memory" effect of good glycemic control was observed in type 2 diabetic patients as well, similar to the results of the DCCT/EDIC study in type 1 diabetic patients.

KEY POINTS

- Intensive glucose control has been demonstrated to decrease the risk of microvascular complications in patients with type 1 and type 2 diabetes.
- The effects of several years of tight glucose control persist for many years, the so-called "metabolic memory phenomenon."

ROLE OF GLUCOSE CONTROL IN THE PREVENTION OF CARDIOVASCULAR DISEASE

The role of tight glucose control in the prevention of CVD is still an unresolved issue. Although large amounts of epidemiological data show a tight link between glycemic levels and cardiovascular outcomes, the data from randomized controlled trials are more contradictory. In newly diagnosed type 1 diabetic patients studied in the DCCT/EDIC trial, it took 17 years of follow-up to detect a statistically significant decrease in cardiovascular events in the intensive therapy group,[86] whereas in newly diagnosed type 2 diabetic patients studied in the UKPDS trial a statistically significant decrease in myocardial infarction was observed only 20 years after randomization.[93] More recently, three large randomized clinical trials have been performed with the goal of determining whether achieving normal glycemic control in type 2 diabetic patients at high risk for CVD would result in a reduction in cardiac events [94-96]; unfortunately, the results have been somewhat unexpected. The Action to Control Cardiovascular Risk in Diabetes study randomized more than 10,000 type 2 diabetic patients with a median duration of diabetes of 10 years and either previous cardiovascular events (35% of the group) or two other cardiovascular risk factors to standard therapy (target HbA1c of 7.0%–7.9%) or intensive therapy (goal HbA1c of <6.0%). After 3.5 years of follow-up the study was terminated early because of increased mortality in the intensive therapy group, while the primary outcome was not statistically different between the two groups. The HbA1c separation between the two groups was good, with a median of 7.5% in the standard treatment and a median of 6.4% in the intensive therapy group.[94] Not surprisingly, the intensive therapy group was characterized by a 3-fold increase in the incidence of severe hypoglycemia and a doubling of patients gaining more than 10 kg.[94] At study end, after the intensive therapy group was terminated but patients were followed for up to 5 years, the use of intensive therapy reduced nonfatal myocardial infarctions but increased total mortality.[95] Of interest, the tighter control in the intensive group was achieved with multiple glucose-lowering drugs, including insulin and oral agents. Moreover, the effects of intensive therapy on microvascular disease were less evident than in prior studies of newly diagnosed type 2 diabetic patients, with no reductions in retinopathy, although there was a decrease in cataract extractions, some improvements in neuropathy, and delayed onset of albuminuria.[96] The cause of the excess mortality associated with intensive therapy has still not been identified, however, subsequent analysis found the effect to be robust and observed in most subgroups.[97] The Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial randomized 11,000 type 2 diabetic patients with duration of diabetes of approximately 8 years and history of CVD or at least one risk factor for vascular disease to standard glucose control (target HbA1c by local guidelines) or intensive glucose control (target HbA1c of <6.5%) achieved with the addition of glicazide 30 to 120 mg daily. After a median follow-up of 5 years, mean HbA1c levels were 7.3% and 6.5% in the two groups, respectively, but there was no effect on major macrovascular disease or death from any cause.[98] The two groups were then followed for an additional 6 years after trial end, and no differences were observed in the rate of macrovascular disease or death from any cause, consistent with the findings of the original trial.[99] The Veterans Affairs Diabetes Trial was a similarly designed trial, smaller in size (almost 1800 type 2 diabetic patients randomized), in type 2 diabetic patients with an 11-year history of diabetes, 40% with CVD. After a median follow-up of 5.6 years, and despite robust difference in HbA1c level (8.4% in standard therapy and 6.9% in the intensive therapy group), there was no significant difference in the primary or secondary cardiovascular outcomes.[100] After the conclusion of the trial, participants were followed for an additional 5 years, with both groups showing a difference in HbA1c levels of only 0.2% to 0.3% at the end of follow-up. At the end of the observation period the intensive therapy group showed a significant decrease in the composite primary cardiovascular outcome (HR 0.83, $P = .04$) but no reduction in cardiovascular mortality.[101]

The results of these studies suggest that aggressive management of hyperglycemia may not be appropriate in all diabetic patients; but in certain populations, especially older adults and patients with established microvascular disease, a less stringent HbA1c goal would be appropriate. Moreover, in patients with established CVD, tight glucose control does not seem able to decrease CVD outcomes, but still plays a role in the beneficial effects on microvascular disease.

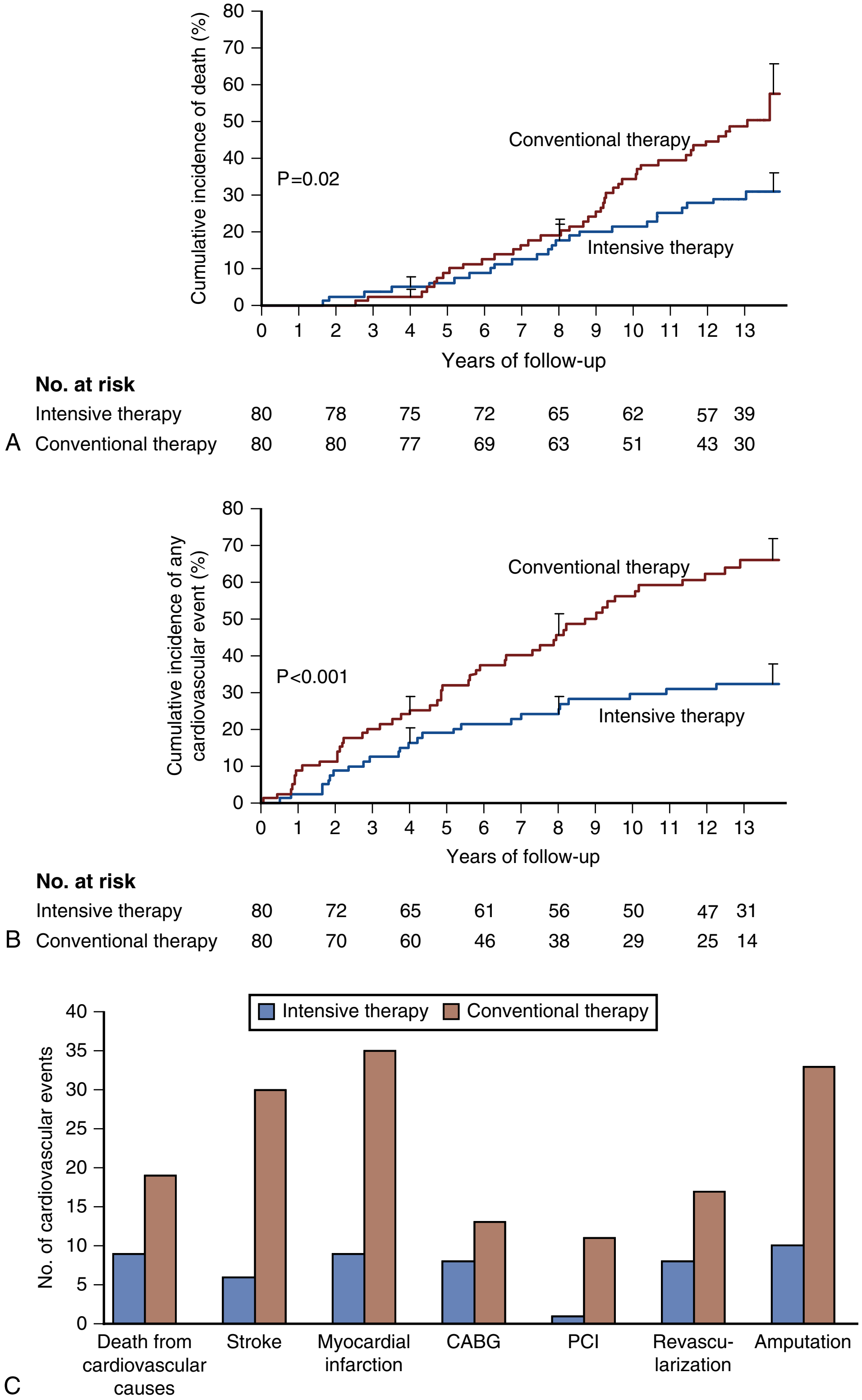
KEY POINTS

- Intensive diabetes treatment leads to a reduction in the incidence of cardiovascular disease in patients with newly diagnosed diabetes but requires a much longer time to take effect than it does for glucose control to have a positive effect on the incidence of microvascular disease.
- Tight glucose control is not effective in reducing cardiovascular disease in patients with long duration of diabetes and may be linked to higher mortality.

CAN WE PREVENT MACROVASCULAR DISEASE?

Although randomized clinical trials evaluating the effects of tight glucose control have shown that CVD can be prevented after a long period of normoglycemia in patients with newly diagnosed type 1 or type 2 diabetes,[80,86,93] multiple other risk factors predispose diabetic patients to CVD, and it is therefore not surprising that interventions addressing multiple risk factors could be more effective in reducing the burden of CVD in these patients. One study that has proven this point is the small Steno type 2 randomized study conducted in Denmark in the 1990s. A total of 160 type 2 diabetic patients with a mean age of 55 years and a duration of diabetes of 6 years were randomized in this open, parallel trial to standard treatment (according to the 1988 recommendations of the Danish Medical Association) or intensive multifactorial intervention (diet and exercise, counseling on smoke cessation, captopril 50 mg twice/day, vitamin C 250 mg/day, vitamin E 100 mg/day, aspirin 150 mg/day for patients with history of CVD) plus intensification of therapy with stricter goals for blood pressure, HbA1c, and fasting LDL cholesterol.[102] At the end of 3.8 years of follow-up, there were statistically significant differences in HbA1c, systolic blood pressure, and cholesterol levels, and these changes led to a significant decrease in microvascular complications.[102] The study continued for another 5 years of follow-up, during which intensive therapy patients maintained a 0.7% decrease in HbA1c, a 11-mm Hg decrease in systolic blood pressure, and a 30-mg/dL decrease in LDL cholesterol, and this led to a greater than 50% reduction in a primary composite cardiovascular endpoint (death from CVD, nonfatal myocardial infarction, coronary artery bypass grafting, percutaneous coronary interventions, nonfatal stoke, amputations, or vascular surgery for peripheral vascular disease).[103]

After the end of this interventional study, these patients were observed in a follow-up study for another 5.5 years, during which the differences in metabolic parameters narrowed and lost statistical significance. Intensive treatment was, however, associated with a significant decrease in total and cardiovascular mortality, with HRs of 0.43 and 0.41, respectively, and the decrease in cardiovascular events was seen in all categories, from stroke and myocardial infarction to amputation[104] (Fig. 42.5).



No. at risk (A)

Intensive therapy	80	78	75	72	65	62	57	39
Conventional therapy	80	80	77	69	63	51	43	30

No. at risk (B)

Intensive therapy	80	72	65	61	56	50	47	31
Conventional therapy	80	70	60	46	38	29	25	14

Fig. 42.5 A, Kaplan-Meier estimates of the risk for death from any cause and B, of the incidence of any cardiovascular events (death from cardiovascular causes, nonfatal stroke, nonfatal myocardial infarction, coronary artery bypass grafting [CABG], percutaneous coronary intervention [PCI], revascularization for peripheral atherosclerotic artery disease, and amputation). The bars represent standard error. C shows the number of events for each component of the composite end point. (From Gaede P et al. N Engl J Med. 2008;358:580-591.)

KEY POINTS

- A combined approach to the prevention of cardiovascular disease by lowering glycated hemoglobin, blood pressure, and low-density lipoprotein cholesterol has been shown to successfully reduce cardiovascular disease, cardiovascular mortality, and total mortality.

EFFECTS OF SPECIFIC HYPOGLYCEMIC AGENTS ON CARDIOVASCULAR DISEASE AND RENAL OUTCOMES IN PATIENTS WITH TYPE 2 DIABETES

Several large cardiovascular outcome trials, mandated by the US Food and Drug Administration to prove CVD safety for novel hypoglycemic agents, have recently shown that two classes of drugs, sodium-glucose cotransporter 2 (SGLT-2) inhibitors and glucagon-like peptide 1 receptor agonists (GLP-1 RAs), can significantly reduce the progression of CVD events and the decline in renal function in patients with type 2 diabetes.

These large-scale, randomized, placebo-controlled clinical trials have been performed in patients with type 2 diabetes and high risk of CVD (either established CVD and/or multiple risk factors for atherosclerotic disease), with the primary outcome being a composite major adverse cardiovascular event (MACE), including nonfatal myocardial infarction, nonfatal stroke, and cardiovascular death.

The first study to be published was the Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients showing that the addition of the SGLT-2 inhibitor empagliflozin to standard care induced a significant decrease in MACE (HR 0.86, $P < .001$), death from CVD (HR 0.62, $P < .001$), hospitalization for heart failure (HR 0.65, $P < .002$), and death from any causes (HR 0.68, $P < .001$).[105] These results were also replicated in the studies of canagliflozin (HR for composite primary outcome 0.86, $P < .001$), whereas dapagliflozin did not decrease MACE (HR 0.93, $P = .17$), but induced a statistically significant decreases in CVD death and hospitalization for heart failure.[106-107] The most recent SGLT-2 inhibitor, ertugliflozin, did not induce any changes in cardiovascular events.[108] These agents reduce hyperglycemia by increasing urinary glucose excretion and have also been shown to reduce renal intraglomerular pressure, and they have been evaluated for their effects on renal outcomes. These studies have been conducted in patients with type 2 diabetes and an estimated glomerular filtration rate greater than 30 mL/min/1.73 m^2 of body surface area, with a composite renal outcome including doubling of serum creatinine, initiation of renal-replacement therapy, or death from renal disease. Their efficacy in reducing renal outcomes has been demonstrated for empagliflozin (HR 0.61, $P < .001$), canagliflozin (HR 0.66, $P < .001$), and dapagliflozin (HR 0.61, $P < .001$).[109,110,111]

The first CVD outcome trial of a GLP-1 RA was published in 2015 and showed that lixisenatide did not significantly change the rate of major CVD events.[112] However, in 2016, the Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results trial demonstrated that liraglutide induced a reduction in MACE (HR 0.87, $P = .01$) in patients with type 2 diabetes and high risk for CVD.[113] Similar results were observed for the GLP-1 RA semaglutide, either the subcutaneous or the oral formulations (HR 0.74, $P < .001$ and 0.79, $P < .001$, respectively),[114-115] but not for once-weekly exenatide.[116] Finally, dulaglutide was also shown to induce a significant reduction in MACE (HR 0.88, $P = 0.026$).[117] The effects of GLP-1 RA on renal outcomes has also been studied, and, although a reduction in new-onset macroalbuminuria has been demonstrated,[118] the rates of other renal adverse events were not significantly decreased with these agents. Moreover, the overall effects of GLP-1 RA on composite renal outcomes were of a much lesser degree than those observed in SGLT-2 inhibitor studies.[119] These results have changed the treatment of patients with type 2 diabetes; current guidelines now recommend that patients with type 2 diabetes and CVD use SGLT-2 inhibitors or GLP-1 RA as second agents after metformin monotherapy.[120] Specifically, patients with chronic kidney disease or congestive heart failure should preferentially add an SGLT-2 inhibitor, whereas patients with CVD or high risk for CVD should add a GLP-1 RA or an SGLT-2 inhibitor with proven CVD benefit as a second agent.[120]

KEY POINTS

- Several sodium-glucose cotransporter 2 (SGLT-2) inhibitors and glucagon-like peptide 1 receptor agonists (GLP-1 RAs) have been demonstrated to reduce cardiovascular events in patients with type 2 diabetes who are at high risk for cardiovascular disease.
- SGLT-2 inhibitors are particularly effective at reducing hospitalizations in patients with congestive heart failure.
- SGLT-2 inhibitors have also been shown to reduce composite renal outcomes, while GLP-1 RAs have shown reduction of new-onset macroalbuminuria.
- Current guidelines recommend that patients with type 2 diabetes and cardiovascular disease use SGLT-2 inhibitors or GLP-1 RAs as second agents after metformin monotherapy.

CONCLUSION

It has been clearly demonstrated that hyperglycemia affects the microvasculature through several different biochemical mechanisms and is the major cause of microvascular disease in diabetes. While several drugs have been helpful in preventing or reducing microvascular disease in animal models of diabetes, the results of clinical trials in diabetic patients have been disappointing so far. Near-normoglycemia, especially from close to the onset of type 1 or type 2 diabetes, is very effective in preventing or delaying the development of diabetic retinopathy, nephropathy, and neuropathy, and is the mainstay of current diabetes management. The increased prevalence of CVD in diabetic patient is likely the result of a combined effect of the metabolic abnormalities associated with the metabolic syndrome (i.e., obesity, hypertension, hyperlipidemia, and hyperglycemia). Conversely, focusing on the management of hyperglycemia alone is not a successful strategy in the management of the cardiovascular complications of diabetes. New hypoglycemic agents such as SGLT-2 inhibitors and GLP-1 RAs have been shown to be effective in the secondary prevention of CVD events in patients with type 2 diabetes and have been incorporated in the standard of care for such patients. Current guidelines in the management of diabetic patients focus on risk-factor control in a similar way to the Steno type 2 trial approach, and recent data have shown the efficacy of this approach. Over the last 10 years, the proportion of diabetic subjects achieving HbA1c, blood pressure, and LDL cholesterol goals has increased, even though almost half of US adults with diabetes are still not reaching the recommended goals.[121] More importantly, and at least partially as a consequence of these changes, over the last 20 years there has been a significant decline in the incidence of diabetes-related complications in the United States, with the greatest changes in myocardial infarction and death from hyperglycemic crisis (relative decreases of 68% and 64%, respectively), strokes and lower extremity amputations (decreases of 53% and 51%, respectively), and end-stage renal disease (decrease of 28%)[122] (Fig. 42.6).

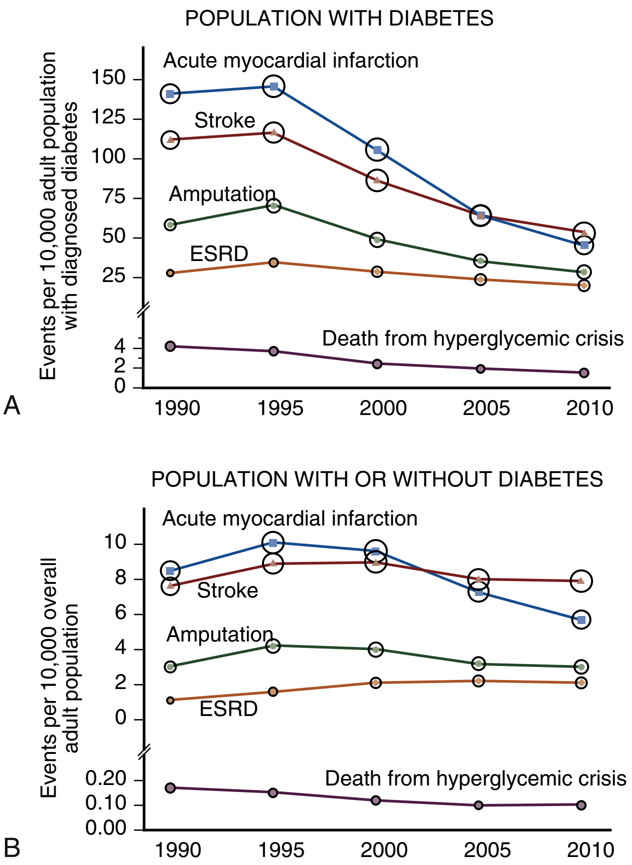


Fig. 42.6 Trends in age-standardized rates of diabetes-related complications among US adults with diagnosed diabetes, 1990 to 2010. Circle size is proportional to the absolute number of cases. *ESRD*, End-stage renal disease. (From Gregg EW, Li Y, Wang J, et al. Changes in diabetes-related complications in the United States, 1990-2010. *N Engl J Med.* 2014;370:1514–1523.)

REFERENCES

1. Sandres LJ. From Thebes to Toronto and the 21st century: an incredible journey. *Diabet Spectr*. 2002;15:56–60.
2. Mayers J. On diabetes and its connection with heart disease. *Br Med J*. 1888;1:949951.
3. Kimmelstiel P, Wilson C. Intercapillary lesions in the glomeruli of the kidney. *Am J Pathol*. 1936;12:83–97.
4. Ballantyne AJ, Loewenstein A. Retinal micro-aneurysms and punctate haemorrhages. *Br J Ophthalmol*. 1944;28:593–598.
5. Rundles RW. Diabetic neuropathy. General review with report of 125 cases. *Medicine*. 1945;24:111–160.
6. Lundbaek K. Diabetic angiopathy, a specific vascular disease. *Lancet*. 1954;1:377–379.
7. Root HF, Pote WH, Frehner H. Triopathy of diabetes. *Arch Intern Med*. 1954;94:931–941.
8. Dunlop DM. Are diabetic degenerative complications preventable? *Br Med J*. 1954;2:383–385.
9. Matthews JD. Vascular disease in diabetes mellitus. *Lancet*. 1954;2:573–575.
10. Joslin EP. A half-century's experience in diabetes mellitus. *Br Med J*. 1950;1:1095–1098.
11. Cahill GF, Etzwiler DD, Freinkel N. "Control" and diabetes. *N Engl J Med*. 1976;294:1004–1005.
12. Siperstein MD, Foster DW, Knowles Jr HC, et al. Control of blood glucose and diabetic vascular disease. *N Engl J Med*. 1977;296:1060–1063.
13. Ding J, Wong TY. Current epidemiology of diabetic retinopathy and diabetic macular edema. *Curr Diab Rep*. 2012;12:346–354.
14. Antonetti DA, Klein R, Gardner TW. Diabetic retinopathy. *N Engl J Med*. 2012;366:1227–1239.
15. Reutens AT. Epidemiology of diabetic kidney disease. *Med Clin N Am*. 2013;97:1–18.
16. Boulton AJM. Diabetic neuropathy: is pain God's greatest gift to mankind? *Semin Vasc Surg*. 2012;25:61–65.
17. Vinik AI, Maser RE, Mitchell BD, et al. Diabetic autonomic neuropathy. *Diabet Care*. 2003;26:1553–1579.
18. Stone PH, Muller JE, Hartwell T, et al. The effect of diabetes mellitus on prognosis and serial left ventricular function after acute myocardial infarction: contribution of both coronary disease and diastolic left ventricular dysfunction to adverse prognosis. *J Am Coll Cardiol*. 1989;14:49–57.
19. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med*. 2011;364:829–841.
20. Shah BR, Hux JE. Quantifying the risk of infectious diseases for people with diabetes. *Diabet Care*. 2003;26:510–513.
21. Cagliero E, Apruzzese W, Perlmutter GS, et al. Musculoskeletal disorders of the hand and shoulder in patients with diabetes mellitus. *Am J Med*. 2002;112:487–490.
22. Cagliero E. Rheumatic manifestations of diabetes mellitus. *Curr Rheumatol Rep*. 2003;5:189–194.
23. Joshi N, Caputo GM, Weitekamp MR, et al. Infections in patients with diabetes mellitus. *N Engl J Med*. 1999;341:1906–1912.
24. Knapp S. Diabetes and infections: is there a link? A mini-review. *Gerontology*. 2013;59:99–104.
25. Thwaini A, Khan A, Malik A, et al. Fournier's gangrene and its emergency management. *Postgrad Med J*. 2006;82:516–519.
26. Taylor GW, Borgnakke WS. Periodontal disease: associations with diabetes, glycemic control and complications. *Oral Dis*. 2008;14:191–203.
27. Yosipovitch G, Hodak E, Vardi P, et al. The prevalence of cutaneous manifestations in IDDM patients and their association with diabetes risk factors and microvascular complications. *Diabet Care*. 1998;21:506–509.
28. Nern K. Dermatologic conditions associated with diabetes. *Curr Diab Rep*. 2002;2:53–59.
29. Pirart J. Diabetes mellitus and its degenerative complications: a prospective study of 4,400 patients observed between 1947 and 1973. *Diabet Care*. 1978;1:168–188.
30. Roy MS, Klein R, O'Colmain BJ, et al. The prevalence of diabetic retinopathy among adult type 1 diabetic persons in the United States. *Arch Ophthalmol*. 2004;122:546–551.
31. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Research Group: Modern day clinical course of type 1 diabetes mellitus after 30 years' duration. *Arch Intern Med*. 2009;169:1307–1316.
32. Martin CL, Waberski BH, Pop-Busui R, et al. Vibration perception threshold as a measure of distal symmetrical peripheral neuropathy in type 1 diabetes. Results from the DCCT/EDIC study. *Diabetes Care*. 2010;33:2635–2641.
33. Tesfaye S, Stevens LK, Stephenson JM, et al. Prevalence of diabetic peripheral neuropathy and its relation to glycemic control and potential risk factors: the EURODIAB IDDM Complications study. *Diabetologia*. 1996;39:1377–1384.
34. Young MJ, Boulton AJM, Macleod AF, et al. A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population. *Diabetologia*. 1993;36:150–154.
35. de Boer IH, for the DCCT/EDIC Research Group. Kidney disease and related findings in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study. *Diabet Care*. 2014;37:24–30.
36. Kramer HJ, Nguyen QD, Curham G, et al. Renal insufficiency in the absence of albuminuria and retinopathy among adults with type 2 diabetes mellitus. *JAMA*. 2003;289:3273–3277.
37. Sandholm N, Groop PH. Genetic basis of diabetic kidney disease and other diabetic complications. *Curr Opin Genetics Dev*. 2018;50:17–24.
38. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16:377–390.
39. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis. Epidemiology, Pathophysiology and management. *JAMA*. 2002;287:2570–2581.

40. Pasterkamp G. Methods of accelerated atherosclerosis in diabetic patients. *Heart*. 2013;99:743–749.
41. Juutilainen A, Seppo L, Ronnemaa T, et al. Similarity of the impact of type 1 and type 2 diabetes on cardiovascular mortality in middle-aged subjects. *Diabet Care*. 2008;31:714–719.
42. Duca L, Sippl R, Snell-Bergeon JK. Is the risk and nature of CVD the same in tyep1 and type 2 diabetes? *Curr Diab Rep*. 2013;13:350–361.
43. Groop PH, Thomas MC, Moran JL, et al. The presence and severity of chronic kidney disease predicts all-cause mortality in type 1 diabetes. *Diabetes*. 2009;58:1651–1658.
44. Orchard TJ, Secrest AM, Miller RG, et al. In the absence of renal disease, 20 year mortality risk in type 1 diabetes is comparable to that of the general population: a report from the Pittsburgh Epidemiology of Diabetes Complications Study. *Diabetologia*. 2010;53:2312–2319.
45. Rosner Preis S, Hwang SJ, Coady S, et al. Trends in all-cause and cardiovascular disease mortality among women and men with and without diabetes mellitus in the Framingham Heart Study, 1950 to 2005. *Circulation*. 2009;119:1728–1735.
46. Waerbaum I, Blohme G, Ostman J, et al. Excess mortality in incident cases of diabetes mellitus aged 15 to 34 years at diagnosis: a population-based study (DISS) in Sweden. *Diabetologia*. 2006;49:653–659.
47. Constantino MI, Molyneaux L, Limacher-Gisler F, et al. Long-term complications and mortality in young-onset diabetes. *Diabet Care*. 2013;36:3863–3869.
48. Rossing P. The changing epidemiology of diabetic microangiopathy in type 1 diabetes. *Diabetologia*. 2005;48:1439–1444.
49. Harding JL, Pavkov ME, Magliano DJ, et al. Global trends in diabetes complications: a review of current evidence. *Diabetologia*. 2019;62:3–16.
50. Miller RG, Secrest AM, Ellis D, et al. Changing impact of modifiable risk factors on the incidence of major outcomes of type 1 diabetes. The Pittsburgh epidemiology of diabetes complications study. *Diabet Care*. 2013;36:3999–4006.
51. LeCaire T, Klein BEK, Howard KP, et al. Risk for end-stage renal disease over 25 years in the population-based WESDR cohort. *Diabet Care*. 2014;37:381–388.
52. Engerman RL, Kern TS. Experimental galactosemia produces diabetic-like retinopathy. *Diabetes*. 1984;33:97–100.
53. Engerman R, Bloodworth Jr IMB, Nelson S. Relationship of microvascular disease in diabetes to metabolic control. *Diabetes*. 1977;26:760–769.
54. Mauer SM, Steffes MW, Sutherland DER, et al. Studies of the rate of regression of the glomerular lesions in diabetic rats treated with pancreatic islet transplantation. *Diabetes*. 1975;24:280–285.
55. Forbes JM, Cooper ME. Mechanisms of diabetic complications. *Physiol Rev*. 2013;93:137–188.
56. Robinson R, Barathi VA, Chaurasia SS, et al. Update on animal models of diabetic retinopathy: from molecular approaches to mice and higher mammals. *Dis Model Mech*. 2012;5:444–456.
57. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature*. 2001;414:813–820.
58. Giacco F, Brownlee M. Oxidative stress and diabetic complications. *Circ Res*. 2010;107:1058–1070.
59. Badal SS, Danesh FR. New insights into molecular mechanisms of diabetic kidney disease. *Am J Kidney Dis*. 2014;63(suppl 2):S63–83.
60. Roy S, Sala R, Cagliero E, et al. Overexpression of fibronectin induced by diabetes or high glucose: phenomenon with a memory. *Proc Natl Acad Sci USA*. 1990;87:404–408.
61. Engerman RL, Kern TS. Progression of incipient diabetic retinopathy during good glycemic control. *Diabetes*. 1987;36:808–812.
62. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: The Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA*. 2003;290:2159–2167.
63. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus: 10 years after the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 2008;126:1707–1715.
64. Cooper ME, El-Osta A. Epigenetics. Mechanisms and implications for diabetic complications. *Circ Res*. 2010;107:1403–1413.
65. Keating ST, El-Osta A. Glycemic memories and the epigenetic component of diabetic nephropathy. *Curr Diab Rep*. 2013;13:574–581.
66. Miao F, Chen Z, Genuth S, DCCT/EDIC Research Group, et al. Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. *Diabetes*. 2014;63:1748–1762.
67. Chalk C, Benstead TJ, Moore F. Aldose reductase inhibitors for the treatment of diabetic polyneuropathy. *Cochrane Database Syst Rev*. 2007;4:CD004572.
68. Bolton WK, Cattran DC, Williams ME, et al. Randomized trial of an inhibitor of formation of advanced glycation end products in diabetic nephropathy. *Am J Nephrol*. 2004;24:32–40. s.
69. Turgut F, Bolton WK. Potential new therapeutic agents for diabetic kidney disease. *Am J Kidney Dis*. 2010;55:928–940.
70. Sheets MJ, Aiello LP, Davis MD, et al. The effect of the oral PKC β inhibitor ruboxistaurin on vision loss in two phase 3 studies. *Invest Ophthalmol Vis Sci*. 2013;54:1750–1757.
71. Fraser DA, Diep LM, Hovden IA, et al. The effect of long term oral benfotiamine supplementation on peripheral nerve function and inflammatory markers in patients with type 1 diabetes: a 24-month, double blind, randomized, placebo controlled trial. *Diabet Care*. 2012;35:1095–1097.
72. Hanssen KF, Dahl-Jorgensen K, Lauritzen T, et al. Diabetic control and microvascular complications: the near normoglycemic experience. *Diabetologia*. 1986;29:677–684.
73. DCCT Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *N Engl J Med*. 1993;329:977–986.
74. DCCT Research Group. Hypoglycemia in the Diabetes Control and Complications Trial. *Diabetes*. 1997;46:271–286.
75. DCCT Research Group. Influence of intensive diabetes treatment on quality of life outcomes in the diabetes control and complications trial. *Diabet Care*. 1996;19:195–203.
76. DCCT Research Group. Progression of retinopathy with intensive versus conventional treatment in the diabetes control and complications trial. *Ophthalmology*. 1995;102:647–661.
77. DCCT Research Group. Effect of intensive therapy on the development and progression of diabetic nephropathy in the diabetes control and complications trial. *Kidney Int*. 1995;47:1703–1720.
78. DCCT Research Group. The effect of intensive diabetes therapy on the development and progression of neuropathy. *Ann Intern Med*. 1995;122:561–568.
79. DCCT Research Group. The effect of intensive diabetes therapy on measures of autonomic nervous system functions in the diabetes control and complications trial. *Diabetologia*. 1998;41:416–423.
80. DCCT Research Group. Effect of intensive diabetes management on macrovascular events and risk factors in the diabetes control and complications trial. *Am J Cardiol*. 1995;75:894–903.
81. DCCT Research Group. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. *Diabetes*. 1995;44:968–983.
82. Nathan DM for the DCCT/EDIC Research Group. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabet Care*. 2014;37:9–16.
83. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med*. 2000;342:381–390.
84. Aiello LP for the DCCT/EDIC Research Group. Diabetic retinopathy and other ocular findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabet Care*. 2014;37:17–23.
85. De Boer IH, for the DCCT/EDIC Research Group. Kidney disease and related findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabet Care*. 2014;37:24–30.
86. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group.

Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353:2643–2653.

87. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352:837–853.
88. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet*. 1998;352:854–865.
89. Ohkubo Y, Kishikawa H, Araki E, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diabet Res Clin Pract*. 1995;28:103–117.
90. UK Prospective Diabetes Study (UKPDS) Group. UK Prospective Diabetes Study 16. Overview of 6 years' therapy of type II diabetes: a progressive disease. *Diabetes*. 1995;44:1249–1258.
91. Nathan DM. Some answers, more controversy, from UKPDS. *Lancet*. 1998;352:823–833.
92. Stratton IM, Adler AI, Neil AW, et al. Association of glycemic with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ*. 2000;321:405–412.
93. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577–1589.
94. The ACCORD Study group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358:2545–2559.
95. The ACCORD Study group. Long-term effects of intensive glucose lowering on cardiovascular outcomes. *N Engl J Med*. 2011;364:818–828.
96. Ismail-Beigi F, Craven T, Banerji MA, et al. Effects of intensive treatment of hyperglycemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomized trial. *Lancet*. 2010;376:419–430.
97. Calles-Escandon J, Lovato LC, Simon-Morton DG, et al. Effects of intensive compared with standard glycemia treatment strategies on mortality by baseline subgroup characteristics. The ACCORD trial. *Diabet Care*. 2010;33:721–727.
98. The ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008;358:2560–2572.
99. Zoungas S, Chalmers J, Neal B, et al. Follow-up of blood pressure lowering and glucose control in type 2 diabetes. *N Engl J Med*. 2014;371:1392–1406.
100. Duckworth W, Abraira C, Moritz T, et al. Glucose control and vascular complications in Veterans with type 2 diabetes. *N Engl J Med*. 2009;360:129–139.
101. Hayward RA, Reaven PD, Wiitala WL, et al. Follow-up of glycemic control and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2015;372:2197–2206.
102. Gaede P, Vedel P, Parving HH, et al. Intensified multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: the Steno type 2 randomised study. *Lancet*. 1999;353:617–622.
103. Gaede P, Vedel P, Larsen N, et al. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. *N Engl J Med*. 2003;348:383–393.
104. Gaede P, Lund-Andersen H, Parving HH, et al. Effect of a multifactorial intervention on mortality in type2 diabetes. *N Engl J Med*. 2008;358:580–591.
105. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373:2117–2128.
106. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*. 2017;377:644–657.
107. Wiviott SD, Raz I, Bonaca MP, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2019;380:347–357.
108. Cannon CP, Pratley R, Dagogo-Jack S, et al. Cardiovascular outcomes with ertugliflozin in type 2 diabetes. *N Engl J Med*. 2020;383:1425–1435.
109. Wanner C, Inzucchi SE, Lachin JM, et al. Empagliflozin and progression of kidney disease in type 2 diabetes. *N Engl J Med*. 2016;375:323–334.
110. Perkovic V, Neal JB, Bompoint S, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*. 2019;380:2295–2306.
111. Heerspink HJL, Stefansson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383:1436–1446.
112. Pfeffer MA, Clagett B, Diaz R, et al. Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. *N Engl J Med*. 2015;373:2247–2257.
113. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016;375:311–322.
114. Marso SP, Bain SC, Consoli A, et al. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2016;375:1834–1844.
115. Husain M, Birkenfeld AL, Donsmark M, et al. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2019;381:841–851.
116. Holman RR, Bethel MA, Mentz RJ, et al. Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2017;377:1228–1239.
117. Gerstein HC, Colhoun HM, Dagenais GR, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomized placebo-controlled trial. *Lancet*. 2019;394:121–130.
118. Mann JFE, Orsted DD, Brown-Frandsen K, et al. Liraglutide and renal outcomes in type 2 diabetes. *N Engl J Med*. 2017;377:839–848.
119. Zelniker TA, Wiviott SD, Raz I, et al. Comparison of the effects of glucagon-like peptide receptor agonists and sodium-glucose cotransporter 2 inhibitors for prevention of major adverse cardiovascular and renal outcomes in type 2 diabetes mellitus. *Circulation*. 2019;139:2022–2031.
120. American Diabetes Association. Pharmacologic approaches to glycemic treatment: Standards of medical care in diabetes-2021. *Diabet Care*. 2021;44(suppl 1):S111–S124.
121. Ali MK, McKeever Bullard K, Saaddine JB, et al. Achievement of goals in US diabetes care, 1999-2010. *N Engl J Med*. 2013;368:1613–1624.
122. Gregg EW, Li Y, Wang J, et al. Changes in diabetes related complications in the United States, 1990–2010. *N Engl J Med*. 2014;2370:1514–1523.

43

Diabetic Eye Disease

Lloyd Paul Aiello, Paolo S. Silva, Jerry D. Cavallerano, and Ronald Klein

OUTLINE

Pathogenesis, 741
Natural History of Diabetic Retinopathy, 742
  Nonproliferative Diabetic Retinopathy, 742
  Proliferative Diabetic Retinopathy, 743
  Diabetic Macular Edema, Ischemia, and Traction, 743
Clinical Classification of Diabetic Retinopathy Severity, 744
  Nonproliferative Diabetic Retinopathy Levels, 744
  Proliferative Diabetic Retinopathy Levels, 744
Epidemiology, 745
  Prevalence and Incidence of Retinopathy, 746
  Risk Factors for Diabetic Retinopathy, 746
  Diabetes-Related Risk Factors, 746
  Other Risk Factors, 748
Other Ocular Diseases Associated With Diabetes, 749
  Cataract, 749
  Glaucoma, 749
  Visual Impairment, 749
Detection, 749
  Telemedicine Programs for Diabetic Retinopathy, 750
Management, 750
  Medical, 750
  Ophthalmic, 751
  Surgical, 751
  Rehabilitation, 751
  Protective and Risk Factors for Diabetic Retinopathy, 751
Novel Therapies, 751

Diabetes mellitus affects virtually all structures of the eye and many aspects of visual function. Diabetic retinopathy (DR) remains an important cause of new-onset blindness in the United States and other industrialized nations.[1] Additionally, coexisting morbidities of diabetes, including hypertension, renal disease, and dyslipidemia, are associated risk factors for the progression of DR. Multicenter clinical trials have demonstrated the efficacy of glycemic and blood pressure control in reducing the risk for onset and progression of DR[2-4] and have demonstrated the effectiveness of laser photocoagulation in preserving vision and reducing the risk of vision loss.[5,6] Laser photocoagulation, however, is usually reserved for vision-threatening DR, and although laser photocoagulation significantly reduces the risk for severe visual loss (SVL) in most patients, it often does not restore vision, and it may be associated with side effects and complications. Intravitreal injections of vascular endothelial growth factor (VEGF) inhibitors have been shown to reduce the risk for moderate visual loss and improve vision in patients with diabetic macular edema (DME), but this treatment typically requires multiple repeated injections over an extended period of time, especially over the first year of treatment. Currently, laser photocoagulation and intravitreal injections are the only ophthalmic interventions proven to reduce the risk for visual loss once vision-threatening retinopathy is present; as a significant number of individuals with diabetes at risk for visual loss from vision-threatening retinopathy do not receive necessary eye care, public health efforts need to be directed at earlier detection and timely intervention.[7-11] The purpose of this chapter is to present current understanding of the pathogenesis, natural history, epidemiology, detection, and management of DR.

PATHOGENESIS

KEY POINTS

- Elevated blood glucose levels result in anatomic, functional, and biochemical changes that alter cellular metabolism, retinal blood flow, and retinal capillary permeability.

The pathogenesis of DR is not fully understood; however, many mechanisms have been suggested and are summarized in Fig. 43.1.[12-15] The development and progression of DR probably result from a complex interplay of these and other factors, which may vary from person to person. It is likely that the relative contributions of different mechanisms vary in importance at different stages of retinopathy. Glycosylation, protein kinase C, and polyol pathways and changes in retinal blood flow may be particularly important early in the course of the disease, even before the development of microaneurysms or other clinically evident findings. Angiogenesis factors such as VEGF and insulin-like growth factor-1 are more likely to be important later in the course of the disease, just before and during the development of proliferative diabetic retinopathy (PDR) or DME. In addition, interindividual and intraindividual variations in biochemical or physiologic responses to hyperglycemia (perhaps as a result of differences in genetic susceptibility) may exist among people at different stages of diabetic disease. This variability may explain why a few diabetic patients have minimal retinopathy despite years of severe hyperglycemia, whereas in others severe retinopathy develops in a short period despite relatively good glycemic control. Furthermore, these pathogenetic factors have not often been studied together in a prospective fashion, making it difficult to prove a causal relationship.

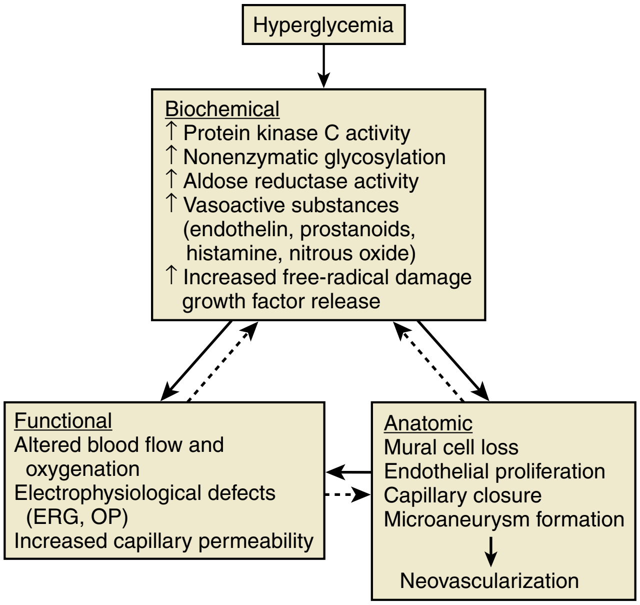

Hyperglycemia
Biochemical
↑ Protein kinase C activity
↑ Nonenzymatic glycosylation
↑ Aldose reductase activity
↑ Vasoactive substances (endothelin, prostanoids, histamine, nitrous oxide)
↑ Increased free-radical damage growth factor release
Functional
Altered blood flow and oxygenation
Electrophysiological defects (ERG, OP)
Increased capillary permeability
Anatomic
Mural cell loss
Endothelial proliferation
Capillary closure
Microaneurysm formation
Neovascularization

Fig. 43.1 **Hypothesized pathogenetic mechanisms for the development and progression of diabetic retinopathy.** *ERP,* Electroretinography; *OP,* oscillatory potentials.

NATURAL HISTORY OF DIABETIC RETINOPATHY

KEY POINTS

- Diabetic retinopathy can be broadly classified into nonproliferative diabetic retinopathy and proliferative diabetic retinopathy. Based on the presence and degree of retinal lesions, nonproliferative diabetic retinopathy is clinically classified as mild, moderate, severe, or very severe. Proliferative diabetic retinopathy is marked by proliferation of new vessels on the optic disc, new vessels elsewhere on the retina, preretinal hemorrhage, vitreous hemorrhage, or fibrous tissue proliferation. Diabetic macular edema, consisting of fluid and swelling in the central macula region, can be present with any level of diabetic retinopathy.
- Proper diagnosis of diabetic retinopathy severity establishes the risk of progression to sight-threatening retinopathy and appropriate clinical management both in terms of follow-up schedule and therapeutic options

Nonproliferative Diabetic Retinopathy

The earliest diabetes-induced changes in the retina are biochemical, hemodynamic, and cellular in nature and are imperceptible clinically. These include changes in biochemical pathways, enzyme activation, retinal blood flow, as well as cellular loss, particularly that of pericytes. The first clinical signs of DR are *microaneurysms,* which are saccular outpouchings of retinal capillaries.[16] These lesions usually appear as round red dots ranging in size from 20 to 200 μm and represent an outpouching of the retinal capillaries. They often appear first in the macular area in areas of capillary closure. It is unusual to detect retinal microaneurysms within 3 years of the diagnosis of type 1 diabetes mellitus (T1DM); however, they are often present at the time of diagnosis in people with type 2 diabetes mellitus (T2DM).[17] In the United Kingdom Prospective Diabetes Study (UKPDS), in which subjects were enrolled at the time of diagnosis of T2DM, nearly 40% of the enrollees had some level of DR at entrance into the study.[18] Moreover, after 10 years of diabetes, 69% of people with T1DM and 55% of people with T2DM have microaneurysms present.[19,20]

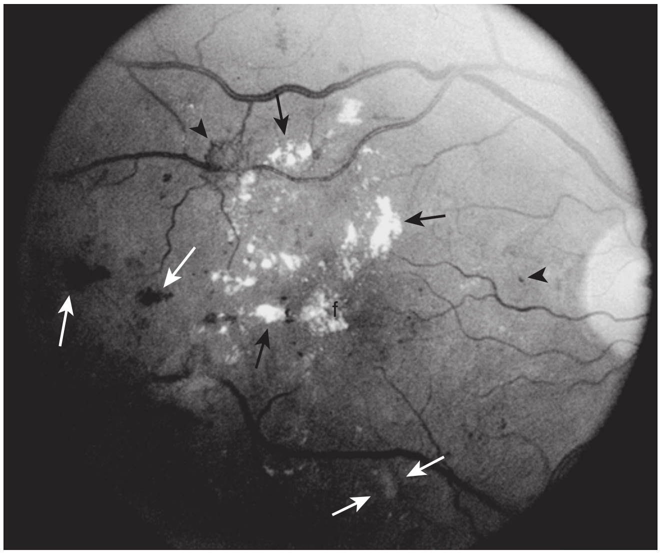


Fig. 43.2 **Fundus photograph of the right eye.** A number of retinal microaneurysms (*small black arrowheads*) appear as small dark spots with sharp margins, and retinal blot hemorrhages (*large white arrows*) appear as dark spots of varying size with irregular margins and uneven densities. Retinal hard exudates appear as white deposits with sharp margins either scattered, "ringlike," or aggregated in their distributions (*small black arrows*) in the superior, temporal, and foveal (*f*) areas. A cotton-wool spot or soft exudate (*small white arrows*) appears as a grayish white area with ill-defined edges. A retinal new vessel superior and temporal to the fovea (*larger black arrowhead*) originates from a small retinal venule.

Retinal microaneurysms are not pathognomonic of DR because they may also be associated with essential hypertension, retinal venous stasis caused by atherosclerotic carotid artery disease, acquired immunodeficiency syndrome, and other systemic and ocular conditions.[21] The appearance of a microaneurysm or two in only one eye of an individual with T2DM should not be regarded as specific for DR; however, when larger numbers of microaneurysms are present (four or more in an eye, or their presence in both eyes), they are more likely caused by diabetes, and the likelihood of progression to more severe nonproliferative DR (NPDR) is greater.[22]

Microaneurysms have abnormal permeability to fluorescein, red blood cells, and lipoproteins.[16] By themselves, microaneurysms are not a threat to vision; however, as the disease progresses, hard exudates (HEs) and retinal dot or blot hemorrhages appear. Dot hemorrhages are frequently indistinguishable from microaneurysms, and for the purposes of grading the severity of retinopathy they are frequently grouped with microaneurysm, being referred to as *hemorrhages and/or microaneurysms* (H/Ma). The blot hemorrhages are round with blurred edges and result from extravasation of blood from retinal capillaries or microaneurysms into the inner nuclear layer of the retina (Fig. 43.2). Retinal blot hemorrhages usually disappear within 3 to 4 months.[23] Ruptured microaneurysms, decompensated capillaries, and intraretinal microvascular abnormalities (IRMAs) can result in *intraretinal hemorrhages.* The clinical appearance of these hemorrhages reflects the retinal architecture at the level at which the hemorrhage occurs. Hemorrhages in the nerve fiber layer assume a more flame-shaped appearance, coinciding with the structure of the nerve fiber layer that runs parallel to the retinal surface. Hemorrhages deeper in the retina, where the arrangement of cells is more or less perpendicular to the surface of the retina, assume a pinpoint or dot shape.

Retinal HEs are sharply defined, yellow, and variable in size; they may be aggregated or scattered and partially or fully circinate in their distribution (see Fig. 43.2). A ring of HEs generally reflects the border of an area of retinal leakage. HEs result from leakage of lipoprotein material from retinal microaneurysms or capillaries into the outer retinal layer, and they may persist for months to years.[23] The exudate is usually found in the posterior layer of the retina, and if they extend into the foveal area, they may reduce visual acuity.

With closure of the retinal capillaries and arterioles, whitish or grayish swellings appear in the nerve fiber layer of the retina. These changes, termed *cotton-wool spots* (CWSs) or *soft exudates,* arise from microinfarcts in the nerve fiber layer (see Fig. 43.2). CWSs may remain only a few weeks to months. After they disappear, the retina may appear normal on ophthalmoscopy, but fluorescein angiography reveals a corresponding area of nonperfusion of the retinal arterioles.

Dilated capillaries called IRMAs are another manifestation of focal retinal ischemia. They are found in areas of capillary nonperfusion and may be abnormally permeable to plasma proteins. IRMAs represent either new vessel growth within the retina or preexisting vessels that have developed abnormal morphology, thought to include endothelial cell proliferation. IRMAs may be seen adjacent to CWSs. Multiple IRMAs identify a severe stage of NPDR.

Venous caliber abnormalities (VCABs) are indicators of severe retinal hypoxia. These abnormalities can take the form of venous dilatation, beading, reduplication, or loop formation. There are often large areas of nonperfusion adjacent to these abnormalities. Treatment with scatter (panretinal) photocoagulation (PRP) may cause these abnormal veins to become less dilated and more regular. IRMAs, intraretinal hemorrhages, and venous beading represent significant retinal ischemia. These changes were referred to as *preproliferative retinopathy* under previous classification systems and are clearly associated with more severe stages of NPDR, being a warning sign of impending retinal neovascularization. Late in the course of the disease, thinly sheathed sclerotic "white, threadlike" arterioles may be present.

Proliferative Diabetic Retinopathy

Proliferative DR (PDR) is characterized by proliferating retinal vessels, the growth of which is variable. They are commonly identified according to their retinal location: at or near the optic disc (*neovascularization of the disc* [NVD]; Fig. 43.3) or elsewhere in the retina (*neovascularization elsewhere* [NVE]; see Fig. 43.2). Retinal neovascularization may be difficult to detect when the vessels first appear as fine tufts of "naked" vessels on the surface of the retina or optic nerve head.[24] They are prone to proliferating on the posterior surface of the vitreous and hemorrhaging into the vitreous. With time, new vessels often fibrose and may contract, resulting in traction detachment of the retina. Regression of new vessels occurs as a result of photocoagulation, treatment with VEGF inhibitors, or the natural course of the disease, although fibrous tissue proliferations (FPs) typically remain.

PDR poses a significant risk for vision loss. Patients with *high-risk PDR* generally require prompt PRP. High-risk PDR is characterized by one or more of the following lesions: (1) NVD that is approximately one-quarter to one-third the disc area or more in size (i.e., greater than or equal to NVD in Early Treatment Diabetic Retinopathy Study (ETDRS) standard photograph 10A); (2) any amount of NVD if fresh vitreous or preretinal hemorrhage (PRH) is present; or (3) NVE greater than or equal to one-half the disc area in size if fresh vitreous or PRH is present. Therefore, attention must be paid to the presence, location, and severity of new vessels, as well as the presence or absence of preretinal or vitreous hemorrhages (VHs).[25]

Diabetic Macular Edema, Ischemia, and Traction

Diabetes affects the macula in a number of ways. First, increased permeability of retinal capillaries and microaneurysms may result in the accumulation of extracellular fluid and thickening of the normally compact macular tissue, with or without the development of cystoid spaces. Macular evaluation requires stereoscopic examination with slit-lamp biomicroscope, stereoscopic fundus photography, and/or optical coherence tomography. The edema may be easily missed, especially if not evaluated with appropriate equipment and by an experienced examiner. The edema is often associated with HE distributed in rings, clumps, or large deposits. Accumulation of exudate is often gradual, and spontaneous resolution may occur. Involvement of the foveal area can be associated with a profound drop in visual acuity.

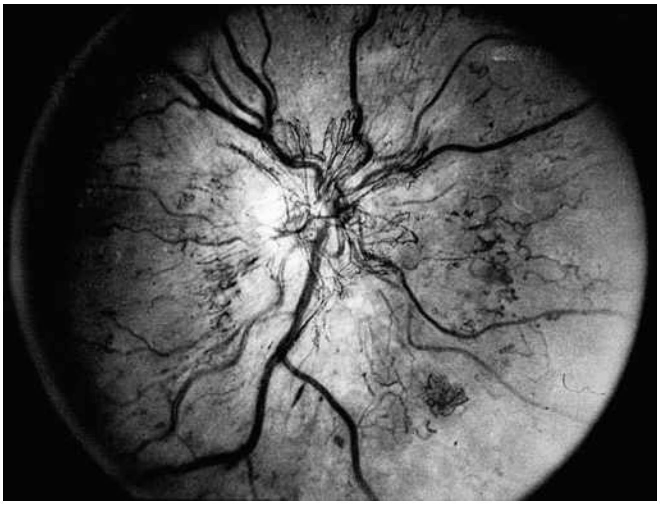


Fig. 43.3 Fundus photograph demonstrating new retinal vessels on the optic nerve head. The retinal veins are also dilated.

When macular edema threatens or involves the center of the macula, the edema is labelled clinically significant macular edema (CSME). CSME can be present with any level of NPDR or PDR, but is increasingly more common with more severe DR. The ETDRS found that CSME is associated with a 30% risk of visual loss over a 3- to 5-year period if left untreated with focal photocoagulation. This risk is reduced by 50% or more with appropriate focal laser photocoagulation.[6] CSME is defined as the presence of any of the following: (1) thickening of the retina at or within 500 µm of the center of the macula, (2) HE at or within 500 µm of the center of the macula with thickening of the adjacent retina, or (3) a zone or zones of retinal thickening one disc area or larger in size, any part of which is within one disc diameter of the center of the macula (see Fig. 43.2). The vast majority of vision loss occurs only once macular edema involves the center of the macula. The presence or absence of center-involved macular edema was the primary factor in determining short- and long-term visual acuity outcomes in the ETDRS. The current standard for DME evaluation is more focused on determining the presence or absence of edema involving the center of the macula. If center involvement of the macula is identified along with reduced vision, a regimen of intravitreal injections of VEGF inhibitors are recommended, which can reduce the risk for moderate visual loss to less than 5% and can significantly improve vision in nearly 50% of eyes when managed appropriately.

The underlying cause of macular edema is not known. It may be a result of both increased leakage and impaired removal of fluid. Breakdown of the blood-retinal barrier has been postulated as an important cause of fluid accumulation in the macula.[16] Reduced osmotic pressure resulting from decreased serum albumin levels, increased intravascular fluid load, increased arterial perfusion pressure, and tissue hypoxia have been postulated to lead to breakdown of the blood-retinal barrier. The retinal pigment epithelium normally serves to "pump" fluid out of the sensory retina, and this function is postulated to be impaired in persons with hyperglycemia. Key growth factors involved in the progression of DR, such as VEGF, have been shown to be potent permeability factors, and contribute significantly to this problem.[26,27] Clinical trials have shown that intravitreal injections of agents that inhibit VEGF rapidly induce the regression of retinal neovascularization[28] and ameliorate macular edema.[29-31]

In addition to macular edema, diabetes may cause macular capillary nonperfusion, retinal hemorrhage or PRH, lamellar or full-thickness hole formation, or dragging or detachment of the macula caused by contraction of fibrovascular tissue. These changes may occur in isolation or in combination and generally result in the loss of central vision if the fovea is involved.

CLINICAL CLASSIFICATION OF DIABETIC RETINOPATHY SEVERITY

DRs can be broadly classified as NPDR or PDR. NPDR lesions include H/Ma, dot and blot hemorrhages, CWS, HE, VCAB, and IRMA. Based on the presence and degree of retinal lesions, NPDR is clinically classified as mild, moderate, severe, or very severe NPDR. PDR is marked by ocular neovascularization as manifested by NVD, NVE, or new vessels on the iris and anterior segment of the eye, PRH, VH, or FP. DME can be present with any level of DR. DME that involves or threatens the center of the macula is classified as CSME. The involvement of the center of the macula has been associated with a significantly increased risk for visual loss. Accurate diagnosis of the severity of DR is essential because the risk for progression to PDR and high-risk PDR is closely correlated with each specific NPDR level. Proper diagnosis of DR severity establishes the risk for progression to sight-threatening retinopathy and appropriate clinical management both in terms of follow-up schedule and therapeutic options. For example, it is important to consider treatment as DR reaches severe NPDR, early PDR, or high-risk PDR.

Nonproliferative Diabetic Retinopathy Levels

Mild NPDR is marked by at least one retinal microaneurysm, but H/Ma are less than those in ETDRS standard photograph No. 2A, and no other retinal lesion or abnormality associated with diabetes mellitus is present. Those with mild NPDR have a 5% risk for progression to PDR within 1 year and a 15% risk for progression to high-risk PDR within 5 years.[32,33] *Moderate NPDR* is characterized by H/Ma greater than those pictured in ETDRS standard photograph No. 2A. Soft exudates, venous beading, and IRMA are definitely present to a mild degree. The risk for progression to PDR within 1 year is 12% to 27%, and the risk for progression to high-risk PDR within 5 years is 33%. These risks are based on estimates from a clinical trial done in the 1980s, the ETDRS, and may be lower now based on changes in management of diabetes resulting in generally improved glycemic and blood pressure control.[34] Patients with mild or moderate NPDR generally are not candidates for PRP and can be followed safely at 6- to 12-month intervals. The presence of macular edema, even with mild or moderate degrees of NPDR, requires follow-up in a shorter period. If CSME that does not involve the center of the macula is present, focal laser treatment is to be considered, and if the DME involves the center of the macula, a regimen of intravitreal injections of VEGF inhibitors is recommended, particularly when vision is also adversely affected. Coincident medical problems or pregnancy will reduce the period until reevaluation. *Severe NPDR,* based on the severity of H/Ma, IRMA, and venous beading, is characterized by any one of the following lesions: (1) H/Ma greater than standard photograph No. 2A in **four** quadrants *or* (2) venous beading in **two** or more quadrants *or* (3) IRMA greater than standard photograph No. 8A in at least **one** quadrant. Clinically, severe NPDR is diagnosed by applying the "4-2-1 rule," reflected in the definition provided earlier. Eyes with severe NPDR have a 52% risk for developing PDR within 1 year and a 60% risk for developing high-risk PDR within 5 years. These patients require follow-up evaluation in 2 to 4 months. Treatment of CSME is strongly indicated in these patients because of the high risk for the development of PDR requiring PRP.

Eyes with *very severe NPDR* have two or more characteristics of severe NPDR but no frank neovascularization. There is a 75% risk for developing PDR within 1 year. Patients with very severe NPDR may be candidates for PRP, and macular edema, if present, generally should be treated. Follow-up evaluation at 2- to 3-month intervals is important. For patients with T2DM, early PRP may be considered for patients with severe or very severe NPDR.[35]

Proliferative Diabetic Retinopathy Levels

DR with new vessel growth on the optic disc or elsewhere on the retina or with FP is designated PDR. *Early PDR* does not meet the definition of high-risk PDR. Eyes with early PDR have a 75% risk for developing high-risk PDR within 5 years. Patients with *severe* or *very severe NPDR,* or *early PDR,* may be considered for early PRP. In the presence of macular edema, patients with severe NPDR or worse DR should be considered for early focal treatment of macular edema or with intravitreal injections of anti-VEGF agents if the edema involves the center of the macula (especially when vision is reduced) in preparation for the probable need of scatter laser photocoagulation.

The ETDRS severity scale was based on the modified Airlie House classification of DR and is a recognized standard for grading severity of DR. Its use in everyday clinical practice, however, poses difficulty, both in its complexity and difficulty because definitions of the levels are detailed, require comparison with standard photographs, and are complex to remember and apply in a clinical setting. A DR severity scale was developed by the Global Diabetic Retinopathy Group at the International Congress of Ophthalmology in Sydney in April 2002.[36]

This International Classification scale defines five levels of DR. The first level is "no apparent retinopathy," and the second level is "mild NPDR," corresponding to ETDRS stage 20 (microaneurysms only). The risk for significant progression over several years is very low in both groups. The third level, "moderate NPDR," includes eyes with ETDRS levels 35 to 47, and the risk for progression increases

TABLE 43.1 International Clinical Diabetic Retinopathy Disease Severity Scale

Diabetic Retinopathy Severity	Findings Present on Ophthalmoscopy
No apparent DR	No abnormalities
Mild NPDR	Microaneurysm only
Moderate NPDR	More than microaneurysms only, but less than severe NPDR
Severe NPDR	Any of the following with no PDR: • >20 intraretinal hemorrhages in each four quadrants • definite VB in two or more quadrants • prominent IRMA in one or more quadrants
PDR	One or more of: • NV • VH • PRH

DR, Diabetic retinopathy; *IRMA,* intraretinal microvascular abnormalities; *NPDR,* nonproliferative diabetic retinopathy; *NV,* neovascularization; *PDR,* proliferative diabetic retinopathy; *PRH,* preretinal hemorrhage; *VB,* venous beading; *VH,* vitreous hemorrhage.

(From Adamis AP, Miller JW, Bernal MT, et al. Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am J Ophthalmol* 1994;118:445–450.)

significantly by level 47. Still, the fourth level, "severe NPDR" (ETDRS stage 53), carries with it the most ominous prognosis for progression to PDR. The fifth level, "PDR," includes all eyes with definite neovascularization or vitreous hemorrhage/PRH (Table 43.1). There was no attempt to subdivide this level as a function of ETDRS "high-risk characteristics," because significant rates of progression are expected to occur in all cases.

The International Classification DME severity scale separates eyes with apparent DME from those with no apparent thickening or lipid in the macula. For eyes with apparent DME, three categories classify DME as not threatening the center of the macula (mild), threatening the center of the macula (moderate), or involving the center of the macula (severe). The clinical disease severity scale is intended to be a practical and valid method of grading severity of DR and DME (Table 43.2).

TABLE 43.2 International Clinical Diabetic Macular Edema Disease Severity Scale

Diabetic Macular Edema Disease Severity	Findings on Ophthalmoscopy
DME apparently absent	No apparent retinal thickening or HE in the posterior pole
DME apparently present	Some apparent retinal thickening or HE in the posterior pole
Mild DME	Some retinal thickening or HE in the posterior pole but distant from the center of the macula
Moderate DME	Retinal thickening or HE approaching the center of the macula but not involving the center
Severe DME	Retinal thickening or HE involving the center of the macula

DME, Diabetic macular edema; *HE,* hard exudates.
(From Adamis AP, Miller JW, Bernal MT, et al. Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am J Ophthalmol* 1994;118:445–450.)

EPIDEMIOLOGY

KEY POINTS

- Diabetic retinopathy and diabetic macular edema affect individuals with both type 1 and type 2 diabetes.
- Type 1 diabetic patients, because of the earlier onset and longer duration of diabetes, experience more frequent and severe ocular complications. After 5 years of diabetes, 23% of type 1 diabetic patients have diabetic retinopathy. After 10 and 15 years, this prevalence increases to almost 60 and 80%, respectively. Proliferative diabetic retinopathy, the most sight-threatening stage of the disease, is present in 25 and 56% of patients with type 1 diabetes after 15 and 20 years, respectively, and often remains asymptomatic until it has progressed substantially beyond the optimal stage for initiating treatment.
- In type 2 diabetes, diabetic retinopathy was present in 20% of cases at diagnosis, increasing to 60–85% after 15 years, as assessed by the Wisconsin Epidemiologic Study of Diabetic Retinopathy. Proliferative diabetic retinopathy was present in 3–4% of patients within 4 years, 5–20% of patients after 15 years, and >50% after 20 years. Thus, patients with type 2 diabetes are more likely to have diabetic retinopathy at diagnosis and are more likely to develop diabetic retinopathy sooner after diagnosis than patients with type 1 diabetes.

Epidemiologic studies of DR are useful in developing public health strategies to prevent or reduce the occurrence or progression of this complication. In addition, epidemiologic data concerning DR, visual loss, and associated risk factors may be used in projecting costs, developing etiologic insight, designing future studies such as controlled clinical trials of treatment or prevention, and estimating the need for rehabilitative services.

One epidemiologic study that has provided data on DR, visual loss, and associated risk factors is the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR). This population-based study has been described in detail.[14,16,22,23] Data from other epidemiologic studies are also cited. Standardized examination protocols and questionnaires, photographic documentation, photographic standards for grading the severity of retinal lesions, and standardized retinopathy severity scales have permitted, in some cases, comparisons among studies.[37,38]

TABLE 43.3 Prevalence and Severity of Retinopathy by Sex at the Baseline Examination in the Wisconsin Epidemiologic Study of Diabetic Retinopathy (1980–1982)

	YOUNGER ONSET, TAKING INSULIN			OLDER ONSET, NOT TAKING INSULIN			OLDER ONSET, TAKING INSULIN		
Retinopathy Status	Male (%)	Female (%)	Total (%)	Male (%)	Female (%)	Total (%)	Male (%)	Female (%)	Total (%)
	(n = 512)	*(n = 484)*	*(n = 996)*	*(n = 321)*	*(n = 352)*	*(n = 673)*	*(n = 313)*	*(n = 379)*	*(n = 692)*
None	31.1	27.5	29.3	26.8	32.7	29.9	64.5	58.6	61.3
Mild NPDR	26.4	34.7	30.4	34.0	27.6	30.6	25.9	28.5	27.3
Moderate to severe NPDR	18.2	16.9	17.6	27.7	23.9	25.7	6.4	10.3	8.5
PDR without high-risk characteristics	12.3	14.0	13.2	8.1	9.9	9.1	1.9	1.1	1.4
PDR with high-risk characteristics		12.1	6.8	9.5	3.4	6.0	4.8	1.3	1.6

NPDR, Nonproliferative diabetic retinopathy; *PDR,* proliferative diabetic retinopathy.
(From Klein R, Moss SE, Klein BE. New management concepts for persons with diabetes. *Diabetes Care*. 1987;10:633–638.)

Prevalence and Incidence of Retinopathy

The prevalence of DR and CSME by age, gender, and diabetes group in the WESDR is presented in Table 43.3. The highest frequencies of DR and PDR were found in the younger-onset group using insulin; the lowest frequencies were in the older-onset group not using insulin. CSME was most frequent in the younger-onset group using insulin. The prevalence of DR has been reported in other selected population-based studies.[37-58] Pooled data from eight studies, including the WESDR, estimate that, among individuals 40 years of age and older, the prevalence of DR was 40%, and the prevalence of severe retinopathy (severe–very severe NPDR and PDR or macular edema) was 8%. Projection of these rates to the diabetic population 40 years of age or older in the United States resulted in an estimate of 4 million persons with DR, of whom 900,000 have signs of vision-threatening retinopathy.[1]

In the WESDR, the highest incidence and rate of progression of DR were found in the younger-onset group using insulin; the lowest incidence and rate of progression were found in the older-onset group not using insulin.[59,60] Based on the WESDR data, it is estimated that approximately 63,000 new cases of PDR occur annually nationwide, with 29,000 having PDR with DRS high-risk characteristics. In addition, approximately 50,000 new cases of DME occur each year in the United States. It should be noted that the burden of these severe complications cited at that time may currently be different owing to the increasing frequency of T2DM and decreasing incidence of PDR and CSME.[61,62] In the WESDR, this decreasing incidence of PDR and CSME is reflected in the markedly lower estimated annual incidence of any visual impairment in the most recent period of the study over a 25-year time frame.[63]

Risk Factors for Diabetic Retinopathy

Gender, Race, Genetics, and Age. Few differences are found in the risk for development and progression of DR in men and women with diabetes. However, differences among race/ethnic groups have been reported. Results from the study of Pima Indians with T2DM suggest that they are at increased risk for PDR in comparison to White Americans with T2DM.[57] After adjusting for all confounding risk factors, diabetic Mexican-Americans in San Antonio had a 2.4 times higher frequency of DR than diabetic non-Hispanic White Americans studied in the WESDR.[37] Similarly, Mexican Americans with T2DM participating in the National Health and Nutrition Examination Survey III (NHANES had an 84% higher frequency of DR than non-Hispanic White Americans.[58] The higher frequency of DR in Mexican Americans than White Americans remained after adjusting for the duration of diabetes, glycated hemoglobin (HbA1c) level, insulin and oral agent use, and hypertension in that study. However, Hamman and colleagues failed to find a difference in the frequency of DR between Hispanics and non-Hispanic White Americans examined in the San Luis Valley study.[38] West and associates also reported a similar prevalence of retinopathy in Mexican Americans with T2DM living in Arizona, of whom 48% had any retinopathy, 6% had PDR and 5% had CSME.[43] However, a higher prevalence of PDR and macular edema was found in Mexican Americans living in Los Angeles than in White Americans living in Beaver Dam.[64] In the 4-year follow-up, Mexican Americans with T2DM participating in the Los Angeles study had significantly higher annual incidence of PDR, CSME, and visual loss than reported in population-based studies of non-Hispanic White Americans.[65] It has been suggested that Black Americans with T2DM may have more severe DR and loss of vision than White Americans with T2DM.[66] In the NHANES the prevalence of DR in individuals with T2DM was 46% higher in non-Hispanic Black Americans than in non-Hispanic White Americans.[58] However, after adjustment for HbA1c, the duration of diabetes, insulin and oral agent use, and hypertension, the rates for DR were similar between White and Black Americans in that study.

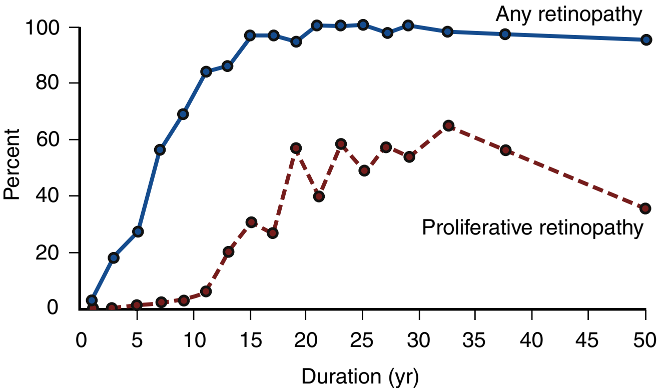


Fig. 43.4 Frequency of retinopathy or proliferative retinopathy by duration of diabetes (years) in 996 insulin-taking individuals in whom diabetes was diagnosed when younger than 30 years of age and who participated in the Wisconsin Epidemiologic Study of Diabetic Retinopathy, 1980–1982. (From Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Arch Ophthalmol.* 1984;102:520–526.)

Reports of a relationship between genetic factors and the prevalence of DR have been inconsistent.[35,67-70] Supporting such a relationship has been the observation that the severity and onset of DR are similar among concordant identical twins, which suggests that the tendency for the development of DR and possibly its progression are influenced by genetic factors.[71] In addition, Hanis demonstrated an 8.3-fold increased risk for DR in 46 Mexican American siblings of probands who had DR compared with the siblings of those who did not.[72] A high degree of familial aggregation and concordance in the development of PDR and end-stage renal disease in diabetic patients implies that common genetic factors may be important in susceptibility or resistance to these complications.[73] Case-control studies have examined other various candidate gene single nucleotide polymorphisms (SNPs) and suggested that chromosome locus 7q21 is a modifier for the risk for nephropathy in diabetes.[74-77] A specific SNP has been identified in the promoter of the erythropoietin gene, which is located at 7q21, that is associated with higher rates of development of severe diabetic eye and kidney complications.[78]

It is uncommon to find clinical evidence of DR in children younger than 10 years of age, regardless of the duration of T1DM; the frequency of any DR or more severe DR increases after 13 years of age.[79-82] This age effect has been postulated to result from a protective effect lost after the start of puberty. In the WESDR, menarchal status at the time of the baseline examination was associated with the prevalence of DR.[83] After adjusting for other factors such as diastolic blood pressure and duration of T1DM, those who were postmenarchal in the WESDR were 3.2 times more likely to have DR than those who were premenarchal.

A number of changes occurring at puberty have been thought to explain the higher risk for DR. These changes include increases in insulin-like growth factor-1, growth hormone, sex hormones, and blood pressure and poorer glycemic control. Increased insulin resistance, inadequate insulin dosage, and poorer compliance in attempts to control blood sugar may result in poorer glycemic control in postpubertal teenagers.[84-89]

Diabetes-Related Risk Factors

Multiple factors have been suggested as risk factors for onset or progression of DR. These factors include duration of diabetes,

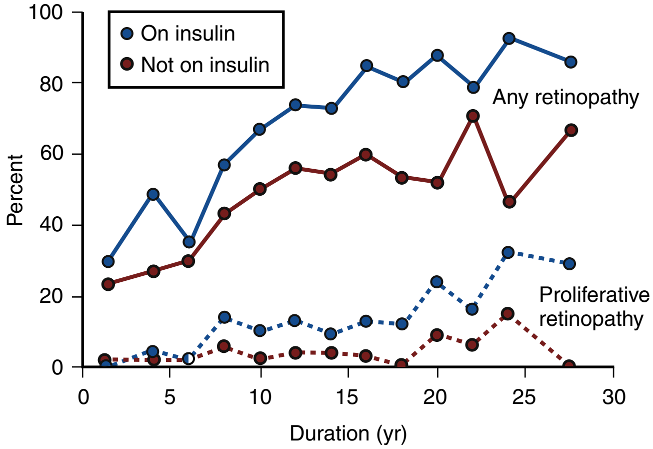


Fig. 43.5 Frequency of retinopathy or proliferative retinopathy by duration of diabetes (years) in 673 individuals taking insulin and 697 people not taking insulin in whom diabetes was diagnosed when older than 29 years of age and who participated in the Wisconsin Epidemiologic Study of Diabetic Retinopathy, 1980–1982. (From Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Arch Ophthalmol.* 1984;102:520–526.)

control of diabetes, hypertension, level of cholesterol and other lipids, renal disease, and anemia. The data associated with these risk factors have been extensively reviewed elsewhere[90] and are only presented briefly here.

Duration of Diabetes. The prevalence of DR (Fig. 43.4), macular edema, and PDR (Fig. 43.5) is significantly related to the duration of diabetes in all three diabetic groups studied in the WESDR. This observation is consistent with all previous epidemiologic studies.[39-45,47-49,55,91] The relationship between the duration of diabetes at the baseline examination and the incidence, progression of NPDR, or progression to PDR in the WESDR has been presented elsewhere.[59,60] At the time of diagnosis, DR was less common, and PDR or macular edema was not present in younger-onset diabetes, whereas the reverse was true in older-onset diabetes mellitus not taking insulin. These findings are consistent with studies that reported relatively high rates of DR at the time of diagnosis of T2DM.[92,93]

These findings have important public health implications. First, they suggest that patients with younger-onset diabetes do not need ophthalmologic evaluation for DR before puberty or before 5 years of diabetes because of the lack of vision-threatening retinopathy. For older-onset individuals, because the onset of diabetes may have been years before its diagnosis, it is important to perform a comprehensive ophthalmic examination, which includes a dilated retinal examination at diagnosis to detect possible PDR or macular edema. These findings have been used to develop guidelines recommending ophthalmologic care for patients with diabetes[94,95] (Table 43.3).

Glycemic Status. A growing body of epidemiologic data has shown that hyperglycemia is related to the incidence and progression of DR.[37,40,41,43,44,46,47,49,50,52-55,58] In the WESDR, the HbA1c level at baseline was found to be a significant predictor of the incidence of any DR, progression, incidence of PDR, or incidence of macular edema in all three diabetic groups studied.[96,97] WESDR findings and those of other studies suggest that, even in people with longer duration of diabetes, good glycemic control is more likely to be associated with lower rates

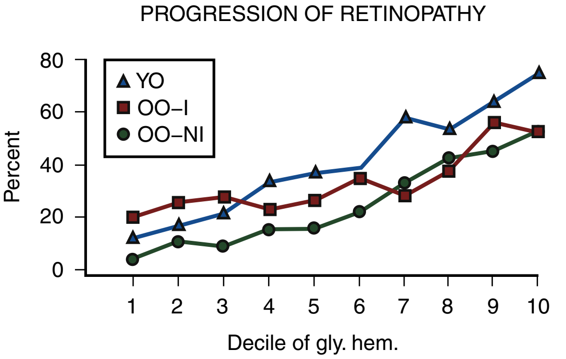


Fig. 43.6 Relationship of the 4-year progression of diabetic retinopathy to deciles of glycated hemoglobin at baseline examination in three diabetic groups participating in the Wisconsin Epidemiologic Study of Diabetic Retinopathy. *YO,* Younger onset; *OO-I,* older onset taking insulin; *OO-NI,* older onset not taking insulin.

of progression of DR than poor glycemic control.[52,97-99] In addition, in the WESDR, no threshold level of glycemia, as measured by deciles of HbA1c at baseline, was seen below which DR did not progress (Fig. 43.6).

The Diabetes Control and Complications Trial (DCCT) was a large randomized, controlled clinical trial of more than 1400 patients with T1DM[2] (see Chapter 51). This multicenter trial showed that intensive insulin treatment reduced the risk for the development and progression of DR. In that study, those assigned to intensive glycemic control had a 60% reduction in three-step or greater progression of DR than did the group assigned to conventional treatment. In the secondary intervention arm of the DCCT, those assigned to intensive glycemic control had a 34% reduction in the progression of DR, a 47% reduction in the incidence of severe levels of NPDR or PDR, a 22% reduction in the incidence of macular edema, and a 54% reduction in laser photocoagulation treatment compared with the group assigned to conventional insulin treatment. The Epidemiology of Diabetes Interventions and Complications (EDIC) study was an extension of the DCCT and followed 1395 (95%) of the 1441 original patients enrolled in the DCCT study. Participants who had been assigned to intensive treatment were encouraged to continue, and participants originally assigned to conventional treatment were advised to change to intensive treatment. The beneficial effects observed in the DCCT were sustained despite a continuously narrowing difference in HbA1c between groups that was not statistically significant by 5 years of follow-up.[99] The risk reductions observed in the DCCT between the rates of microvascular complications in the intensive treatment group compared with the conventional treatment group were sustained throughout the now 30 years of follow-up in the DCCT/EDIC cohort.[100,101]

The UKPDS was a randomized controlled clinical trial involving 3867 patients with newly diagnosed T2DM.[3] After 3 months of diet treatment, patients with a mean of two fasting plasma glucose concentrations of 6.1 to 15.0 mmol/L were randomly assigned to an intensive glycemic control group with either a sulfonylurea (chlorpropamide, glibenclamide, or glipizide) or insulin, or to a conventional glycemic control group with diet. After 10 years of follow-up, HbA1c was 7.0% in the intensive group and 7.9% in the conventional group, and the data conclusively showed that intensive treatment with either sulfonylureas or insulin significantly reduced the risk for progression of DR in individuals with T2DM.

Both the findings of the DCCT/EDIC and the UKPDS underscore the clear benefit of and need for intensive diabetes management as soon as safely possible, which should be sustained with a target HbA1c level of 7.0% or less. Although intensive therapy does not completely prevent DR, when begun early before retinal complications are present it is effective in significantly reducing the risk for development of PDR or macular edema, or the need for photocoagulation.[2,3] However, reducing glycemic control is not without adverse effects. The DCCT found a 60% risk for weight gain and a 330% increased risk for severe hypoglycemic episodes in the intensive glycemic control group versus the conventional group.[2] The Action to Control Cardiovascular Risk in Diabetes (ACCORD) study was a randomized, controlled clinical trial that evaluated the effects of management approaches for "blood glucose levels, serum lipid levels, and blood pressure on cardiovascular events in participants with T2DM who had either established cardiovascular disease (CVD) or known cardiovascular risk factors."[102] The trial showed the efficacy of glycemic control in reducing the incidence and progression of DR. It also showed that in these patients who were at higher risk for a CVD event, intensive glycemic control treatment aimed at lowering the HbA1c level to below 6.0% was associated with a statistically significant 22% increase in mortality compared with those in the standard management group.[103] This resulted in relaxing the guidelines for treatment goals for older adults with T2DM and CVD risk factors or disease or other comorbid conditions to an HbA1c of less than 8.0% or less than 8.5%.[104]

Other Risk Factors

Blood Pressure. Clinical studies suggest a possible relationship between elevated blood pressure and the presence of DR.[85] High blood pressure, through an effect on blood flow, has been postulated to damage capillary endothelial cells, possibly contributing to the development or progression of retinopathy.[105] However, recent epidemiologic studies on the relationship of blood pressure to the incidence and progression of DR have been conflicting.[37,38,40,41,43,44,46,49,50,52,54,55,107-109] These data have been reviewed previously.[90] The UKPDS randomized 1148 patients with hypertension (mean blood pressure, 160/94 mm Hg) to a regimen of tight control with either captopril or atenolol and another 390 patients to less tight control of their blood pressure to evaluate whether tight control of blood pressure with either a β-adrenergic blocker or an angiotensin-converting enzyme inhibitor was beneficial in reducing the macrovascular and microvascular complications associated with T2DM.[4] Tight blood pressure control resulted in a 35% reduction in retinal photocoagulation compared with conventional control. After 7.5 years of follow-up, there was a 34% reduction in the rate of progression of DR by two or more steps on the modified ETDRS severity scale and a 47% reduction in moderate visual loss. The effect was largely caused by a reduction in the incidence of DME. Atenolol and captopril were equally effective in reducing the risk for development of these microvascular complications. The effects of blood pressure control were independent of those of glycemic control. These findings strongly support tight blood pressure control in people with T2DM as a means of preventing visual loss resulting from the progression of DR. However, normalization of blood pressure in the ACCORD trial (systolic blood-pressure control target to <120 vs. <140 mm Hg) showed no benefit in reducing progression of DR in individuals with T2DM and CVD or at risk for CVD.[102]

Regardless of whether blood pressure is a causal agent for the development or progression of DR, it is imperative to measure and treat high blood pressure in diabetic patients because of the known association of high blood pressure with a higher risk for heart attack, stroke, and the development of diabetic nephropathy.[110-115]

Serum Lipids. Epidemiologic data suggest a relationship between higher levels of lipids and the presence of DR or the development of HE.[38,43,49,115,116] The WESDR showed a significant trend toward increasing severity of DR and retinal HE with increasing cholesterol in the younger- and older-onset groups using insulin.[116] The ETDRS found a positive relationship between serum lipids (triglycerides, low-density lipoprotein cholesterol, and very low-density lipoprotein cholesterol) and the development of HE.[103] Although data from earlier clinical trials suggested a beneficial effect of clofibrate (Atromid-S, a lipid-lowering agent with significant hepatic toxicity) in reducing the presence of HE, it did not lead to an improvement in vision.[117,118] Results of the ACCORD-EYE study have shown that management of dyslipidemia with fenofibrate in individuals with T2DM will reduce the incidence and progression of DR.[102]

Proteinuria and Renal Disease. Diabetic nephropathy may lead to lipid, platelet, and rheologic abnormalities, all of which have been hypothesized to be pathogenetic factors for the development of DR.[119,120] Therefore, it is not surprising that most epidemiologic studies have found a strong association between the presence and severity of DR and microalbuminuria and gross proteinuria.[43,49,53,115] These relationships are independent of blood pressure.[121]

Cigarette Smoking and Alcohol Consumption. Most data from epidemiologic studies have failed to confirm earlier reports of a positive relationship between cigarette smoking and DR.[37,38,43,46,122,123] Regardless of the relationship with DR, diabetic patients should be advised to not smoke, because smoking is an important risk factor for respiratory and CVD, as well as for cancer.[124,125] Few epidemiologic studies have investigated the relationship of alcohol consumption to DR.[127-129] One might anticipate a possible protective effect of alcohol as a result of decreased platelet aggregation and adhesiveness.[129] Data from one study suggested a beneficial effect of alcohol, whereas data from another study suggested an increased risk for PDR in people with diabetes.[126,127] In the WESDR, alcohol consumption was associated with a lower frequency of PDR in the younger-onset group.[128] However, no relationship was found between alcohol consumption at the 4-year examination and the incidence and progression of DR in either the younger- or older-onset groups at the 10-year follow-up.[128]

Pregnancy. Epidemiologic studies suggest that pregnancy is a significant risk factor for progression of DR.[130] In a review of the literature, Rodman and colleagues reported that 8% of women with T1DM who had no or early NPDR at the onset of pregnancy had progression of DR during pregnancy.[131] In a case-control study of women with T1DM, the frequency of progression to PDR retinopathy was higher in those who were pregnant than in those who were not (7.3% vs. 3.7%).[130] Women in this study were similar in age, duration of diabetes, and DR status at the baseline examination. Pregnancy remained a significant risk factor for progression of DR after controlling for HbA1c. Severe DR is also an indicator of a higher risk for congenital abnormalities in children born of mothers with T1DM.[132]

Comorbidity. In the WESDR, diabetic individuals with PDR were at higher risk for the development of diabetic nephropathy, heart attack, stroke, and amputation than were those with minimal or no DR present.[121] This observation is consistent with the association of severe DR with CVD risk factors such as increased fibrinogen, increased platelet aggregation, hyperglycemia, and hypertension.[116,119,120,133,134] Retinal vessel diameters, independent of retinopathy severity levels, have also been shown to be associated with stroke, diabetic nephropathy, and lower extremity amputation.[135,136]

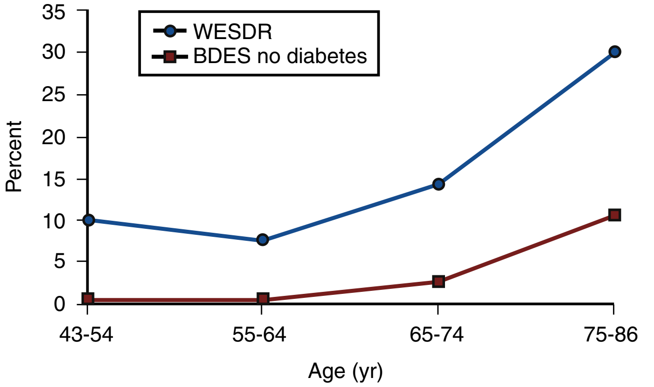


Fig. 43.7 Visual impairment (best-corrected visual acuity in the better eye of 20/40 or worse) in those with diabetes participating in the Wisconsin Epidemiologic Study of Diabetic Retinopathy, 1980–1982, and in those without diabetes participating in the Beaver Dam Eye Study, 1988–1990. *WESDR,* Wisconsin Epidemiologic Study of Diabetic Retinopathy; *BDES,* Beaver Dam Eye Study.

OTHER OCULAR DISEASES ASSOCIATED WITH DIABETES

KEY POINTS

- Diabetic eye disease affects structures of the eye that are susceptible to the deleterious effects of diabetes. The most common nonretinal ocular disease due to diabetes include the development of earlier and more rapidly progressing cataracts and a higher incidence of glaucoma.

Cataract

Cataract is an opacification of the normally clear crystalline lens of the eye. Individuals with diabetes tend to develop cataracts at a younger age, and the cataracts develop more rapidly, compared with individuals without diabetes.[137] In both the NHANES and the Framingham Eye Study, diabetic participants younger than 65 years of age had a respective 3.0 and 4.0 increased risk for having a cataract compared with similarly aged nondiabetic participants. Individuals with diabetes appear to have differences in the frequency and severity of specific lens opacities. In the Beaver Dam Eye Study, after adjusting for age and sex, cortical opacities and posterior subcapsular cataract were significantly more common among individuals with older-onset diabetes than in the rest of the Beaver Dam population.[138] In the younger-onset group in the WESDR, the prevalence of cataract surgery and cataract was significantly related to older age at examination, longer duration of diabetes, diuretic usage, higher HbA1c concentration, and more severe DR.[139] In the older-onset groups, the prevalence of cataract was associated with older age, diuretic usage, smoking, and more severe retinopathy. In the UKPDS, intensive glycemic control was associated with a significant reduction in cataract extraction compared with those with conventional control.[3]

Glaucoma

Glaucoma is usually defined by the presence of characteristic visual field loss associated with damage to the retinal nerve fiber layer as a result of intraocular pressure higher than the eye can tolerate. Diabetes has been suggested as increasing the risk for glaucoma.[140] However, neither the Framingham Eye Study nor a study in Dalby, England, found a relationship between diabetes and glaucoma.[141,142] In the WESDR, a history of glaucoma was more frequent in younger and older diabetic individuals than in the general population studied in the Health Interview Survey or in nondiabetic participants in the WESDR.[142-145] In the Beaver Dam Eye Study, after controlling for age and sex, diabetes was associated with an 84% increase in the risk for glaucoma.[141] The Nurse's Health Study, which enrolled 76,318 women who were followed from 1980 to 2000, showed that T2DM was associated with a 1.23 to 2.7 increased risk for developing primary open-angle glaucoma.[145]

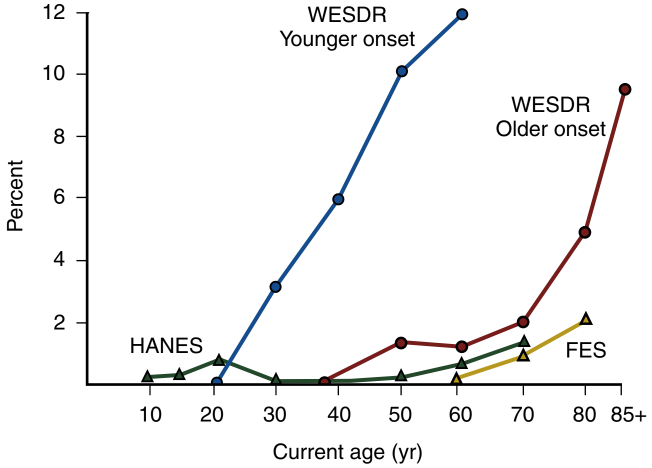


Fig. 43.8 Percentage of individuals with visual acuity of 20/200 or worse in the better eye in the Wisconsin Epidemiologic Study of Diabetic Retinopathy (*WESDR*), the National Health and Nutrition Examination Survey (*NHANES*), and the Framingham Eye Study (*FES*) by current age. (From Klein R, Klein BE, Moss SE. Visual impairment in diabetes. *Ophthalmology*. 1984;91:1–9.)

Visual Impairment

In the WESDR, individuals with diabetes had a higher age-specific prevalence of visual impairment (best-corrected visual acuity in the better eye of 20/40 or worse) than did the nondiabetic population in the Beaver Dam Eye Study (Fig. 43.7), as well as a higher age-specific prevalence of legal blindness (best-corrected visual acuity in the better eye of 20/200 or worse) than found in the general population in either the Framingham Eye Study or the NHANES[146] (Fig. 43.8). For the younger-onset group using insulin, DR was responsible for 86% of the visual loss; for the older-onset group, it was responsible for 33% of the visual loss. In the WESDR, the annual incidence of blindness caused by diabetes mellitus was estimated to be 3.3 per 100,000 general population.[147] The 4-year incidence of legal blindness was higher in the older-onset groups (3.2% in those using insulin and 2.7% in those not using insulin) than in the younger-onset group (1.5%). The incidence of visual impairment because of DR appears to be decreasing in individuals with T1DM in the WESDR.[148] However, the prevalence of visual impairment in T2DM appears to have increased from the 1999 to 2002 to the 2005 to 2008 NHANES and is thought to be attributable in part to an increase in the prevalence during this time period.[149,150]

DETECTION

KEY POINTS

- Detection of diabetic retinal complications is best achieved by ophthalmoscopic examination through a dilated pupil by a retinal specialist or an eye doctor with expertise in diabetes eye care. Alternative means for retinal evaluations include ocular telehealth programs that use validated means of retinal imaging to identify eyes at risk for visual loss. Ocular telehealth programs have the potential to preserve vision and prevent vision loss by increasing access to evaluation, educating patients, and promoting appropriate follow-up.

TABLE 43.4 Recommendations for Eye Care for Diabetic Patients

Primary care physician informs the patient at the time of diagnosis of diabetes that:
- Ocular complications are associated with diabetes and may threaten sight
- Timely detection and treatment may reduce the risk for decreased vision-Referral to an eye doctor competent in diabetes eye care:
- All patients 10–30 years of age who have had diabetes for 5 or more years
- In patients in whom diabetes was diagnosed when older than 30 years of age, examination at the time of diagnosis or shortly thereafterReferral to an ophthalmologist:
- All women with insulin-dependent diabetes mellitus planning pregnancy within 12 months, in the first trimester, and thereafter at the discretion of the ophthalmologist
- Patients found to have reduced corrected visual acuity, elevated intraocular pressure, and any other vision-threatening ocular abnormalities

Detection of DR is best achieved by ophthalmoscopic examination through a dilated pupil by a retinal specialist or an eye doctor with expertise in diabetes eye care.[95] Stereoscopic color fundus photography increases the sensitivity of detection of vision-threatening retinopathy relative to ophthalmoscopy and is especially indicated when monitoring patients with signs of NPDR, PDR, or macular edema.[152-154] Fluorescein angiography is generally used by retinal specialists before laser photocoagulation treatment of eyes with CSME to detect areas of retinal ischemia and leaking microaneurysms. Angiograms may occasionally be used to provide further information regarding the amount of ischemia and to demonstrate the cause of unexplained visual loss. However, fluorescein angiography is not routinely necessary at the time of diagnosis or in diabetic patients with no or minimal NPDR.

In the WESDR in 1980 to 1982, 33% of individuals with either PDR with a high risk for SVL or CSME had not seen an ophthalmologist within 2 years of the examination.[154] In addition, 51% of eyes with PDR had not been treated with PRP.[155] Based on these data, it is estimated that 35,000 Americans with vision-threatening DR who might benefit from photocoagulation treatment had not received such treatment in 1980 to 1982. Little change was found when the group was reexamined,[156] and similar findings have also been reported by others.[158-160] Many possible reasons have been proposed to explain the high rate of patients with serious DR who do not receive appropriate eye care.[11,24,157,161-165] Guidelines for the detection and management of diabetic eye disease have been developed and distributed[95] (Table 43.4). However, studies have suggested that these referral guidelines for DR are not being sufficiently followed.[7,158,159,164] Cost-effectiveness studies using the WESDR and other data suggest that earlier detection of PDR by ophthalmologic or photographic screening is a cost-effective approach for diabetic individuals requiring insulin.[166-168] The American Diabetes Association has recognized the value of retinal imaging in assessing DR and states that examinations can be done with retinal photographs with or without pupil dilation read by experienced experts, but that in-person exams are still necessary when photos are unacceptable and for follow-up of abnormalities detected.[168]

Telemedicine Programs for Diabetic Retinopathy

Despite multiple national and global initiatives, a significant number of individuals with diabetes mellitus who are at risk for visual loss from vision-threatening retinopathy do not receive necessary eye care.[13-15] Public health efforts need to be directed at earlier detection and providing access to appropriate levels of eye care. Ocular telehealth programs for DR have the potential to expand access to highly effective evidence-based diabetes eye care and provide cost-effective alternative methods of care.[169,170] Studies across multiple populations demonstrate that the prevalence of blindness and visual impairment among patients with diabetes is lowest among populations with a national program that provides retinal evaluations for all patients with diabetes.[172-175] The implementation of national coverage of universal retinal evaluation for all patients with diabetes has been shown to reduce the incidence of blindness by as much as 95%.[176-178] Ocular telehealth programs for DR have the potential to preserve vision and prevent vision loss by increasing access to evaluation, educating patients, and promoting appropriate follow-up and treatment.

MANAGEMENT

KEY POINTS
- Medical management relies on achieving control of blood glucose, blood pressure, and blood lipids which have been shown to greatly reduce the risk for development and progression of diabetic retinopathy and to limit visual loss in patients with diabetes.
- Nationwide, multicenter, randomized clinical trials have largely determined the appropriate strategies for ophthalmic management of patients with diabetic retinopathy. Panretinal laser photocoagulation should be considered as an eye approaches the high-risk stage of proliferative retinopathy and intravitreal injections of vascular endothelial growth factor (VEGF) inhibitors have been shown to be as effective as laser photocoagulation for treating proliferative retinopathy. When therapy is initiated at the first sign of appropriate severity of diabetic retinopathy, benefits may include a reduction of more than 90% in severe visual loss.
- Results for the Diabetic Retinopathy Clinical Research Network and pharmaceutical studies have shifted the initial therapeutic approach for patients with center-involved diabetic macular edema and visual loss to a regimen of repeated intravitreal injections of vascular endothelial growth factor (VEGF) inhibitors. These inhibitors can reduce moderate visual loss from DME to less than 5% and lead to visual improvement of 10 letters or more in more than half of these patients.

Medical

As discussed earlier in the chapter, the DCCT and UKPDS studies have demonstrated that it is important to manage glycemic and hypertension control carefully for optimum care of patients with diabetes mellitus. Achieving control of blood glucose and blood pressure has been shown to greatly reduce the risk for development and progression of DR and to limit visual loss in patients with diabetes. However, intensive glycemic control is often difficult or impossible to achieve in certain patients. Strict glycemic control is also associated with an increased frequency of severe complications.[2] An unexpected worsening of DR in the first 3 to 12 months following the initiation of intensive insulin therapy (termed "early worsening") was observed in the DCCT and several other smaller clinical trials.[179-182] In the DCCT, clinically significant early worsening (defined as a three-step progression of DR within the first 12 months) was observed in 13.1% of 711 patients assigned to intensive treatment and in 7.6% of 728 patients assigned to conventional treatment (odds ratio, 2.06; $P < .001$). In both groups, recovery occurred at the 18-month visit in 51% and 55% of patient, respectively ($P = .39$). The risk factors for early worsening were higher baseline HbA1c and greater reduction of HbA1c after enrollment. In the DCCT, early worsening did not lead to serious visual loss, but data support the conclusion that early worsening may be more common and more sight-threatening in patients with more severe retinopathy and/or very poor glycemic control. It is recommended that patients with moderate or worse NPDR should be monitored closely before and for several months after initiation of intensive insulin treatment.[178] Initiation of laser treatment should be considered when factors suggest a particular need to protect against advancing severe retinopathy, such as in patients with actively progressing DR, longstanding very poor glycemic control, and high likelihood of suboptimal follow-up.[178] Given the long-term benefits of improved glycemic control in reducing the risk for DR progression and the availability of highly effective treatment for vision-threatening DR, intensive glycemic control should generally not be discouraged for fear of DR progression.[182,183]

Ophthalmic

Several nationwide, multicenter, randomized clinical trials have largely determined the appropriate strategies for clinical management of patients with DR. The Diabetic Retinopathy Study conclusively demonstrated that PRP significantly reduces the risk for SVL from PDR, particularly when high-risk PDR was present.[5,184,185] The ETDRS provided valuable information concerning the timing of PRP for advancing DR and conclusively demonstrated that focal/grid photocoagulation for CSME reduces the risk for moderate visual loss by 50% or more.[5,32,186] Furthermore, the ETDRS demonstrated that both early PRP and deferral of PRP "until and as soon as high-risk PDR developed are effective in reducing the risk for SVL." PRP, therefore, should be considered as an eye approaches the high-risk stage and "usually should not be delayed if the eye has reached the high-risk proliferative stage."[32] Detailed analysis of the ETDRS data shows that early PRP is beneficial for patients with T2DM.[187,188] When therapy is initiated at the first sign of appropriate severity of DR, benefits may include a reduction of more than 90% in SVL.[32,188] Results for the Diabetic Retinopathy Clinical Research Network[29-31] and pharmaceutical studies[189,190] have shifted the initial therapeutic approach for patients with center-involved DME and visual loss to a regimen of repeated intravitreal injections of VEGF inhibitors. Focal/grid photocoagulation is then added secondarily as needed; however, approximately 50% of the patients will never require focal grid photocoagulation after anti-VEGF therapy. Intravitreal injections, when used according to the protocol described in these studies, can result in limiting moderate visual loss from DME to less than 5% and lead to visual improvement of 10 letters or more in more than half of these patients.

Surgical

Vitrectomy surgery has been effective in reducing vision loss when traction retinal detachment threatens the center of the macula or VH fails to clear. The use of laser delivered from within the eye at the time of surgery and other advanced surgical techniques make vitrectomy a valuable procedure to prevent vision loss and restore vision.

Neovascular glaucoma, in which retinal ischemia leads to neovascularization of the anterior chamber angle of the eye, is a severe and difficult-to-treat complication of advanced DR. Such eyes are at high risk for SVL. Aggressive PRP is usually the preferred initial treatment, but other procedures such as goniophotocoagulation of neovascularization in the filtration angle, filtration surgery, vitrectomy surgery, or combinations of these procedures may be necessary. The use of intravitreal anti-VEGF agents initially in eyes with severe neovascularization of the iris angle have dramatically improved the treatment of those cases where the severity of the condition precluded laser treatment.[191] In these cases, the use of anti-VEGF agents induce a rapid, temporary resolution of the vessels, allowing laser treatment to be subsequently performed.[192]

Rehabilitation

Diabetic patients with severe visual impairment are confronted with difficulties in glucose monitoring, insulin administration, and monitoring of systemic changes in the skin such as foot ulcers. In addition, these patients are often ill with renal disease, CVD, or amputations. Many have lost their jobs and health insurance and are without adequate financial support. These problems may lead to anger, anxiety, loss of self-esteem, and difficulties in social adjustment.[193]

Patients with reduced vision frequently benefit from referral to low-vision specialists or eye doctors specializing in vision rehabilitation. Spectacle, handheld, or stand magnifiers may be useful in near tasks such as insulin dosage, reading, computer work, job-related tasks, and hobbies. Telescopes may improve distance vision, and in some cases permit individuals with reduced vision to resume driving. A variety of electronic devices is available to enhance reading activities, and computer software is available to assist in the use of work- and leisure-related tasks.

Primary care physicians should be actively involved in a team including psychologists, orientation and mobility instructors, rehabilitation teachers, and social workers to deal with the problems faced by visually impaired diabetic patients. Such support facilitates acceptance of visual loss, development of coping strategies, and planning of living arrangements.

Protective and Risk Factors for Diabetic Retinopathy

Recent interest has focused on possible genetic risk factors that may delay or hasten the progression to severe PDR. There are published reports on two unique cohorts of T1DM patients with more than 50 years of diabetes. The Golden Years cohort has shown a possible genetically determined protective effect against the development of diabetic nephropathy and large vessel disease.[194] The 50-Year Medalist Study at Joslin Diabetes Center has demonstrated that substantial proportions of individuals may survive diabetes with a duration of 50 years or more and remain free of advanced diabetic vasculopathy, including PDR (49.4%). Longitudinal data from a subgroup of 97 Medalists followed for an average of 20.6 years and 39.4 visits suggest that retinopathy worsening occurs almost entirely within the first 2 decades of follow-up, and eyes that do not develop PDR have a slower rate of retinopathy progression. The data from these two cohorts point to a possible genetic susceptibility or protective factors against the development of PDR.[195,196] The identification of novel genes may potentially help understand the pathogenic pathways underlying DR. SNPs such as the one observed in the promoter region of the erythropoietin gene has been shown to result in the dramatic increase in expression of erythropoietin, which is a known angiogenic factor in the eye.[78] This polymorphism is associated with higher rates of development of severe diabetic eye and kidney complications and represents one of the first gene mutations specifically associated with DR.

NOVEL THERAPIES

With the emerging global epidemic of diabetes and rapid advances in scientific discovery, there has been a dramatic shift and remarkable improvements in the prevention of visual loss from diabetic retinal disease. Despite these advances, it remains critical that we continue to identify the most promising new therapies and evaluate them in as rigorous and timely a manner as possible. Collaborative efforts between eye care specialists in the highly successful National Eye Institute–sponsored Diabetic Retinopathy Clinical Research Network (www.DRCR.net) have allowed basic science discoveries to be rapidly translated into the clinical setting for the benefit of the patient. It is through collaborative efforts such as these and the close involvement of a multidisciplinary team of healthcare providers that we ultimately may be able to prevent eye and other complications of this systemic disease.

REFERENCES

1. Kempen JH, O'Colmain BJ, Leske MC, et al. The prevalence of diabetic retinopathy among adults in the United States. *Arch Ophthalmol*. 2004;122:552–563.
2. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med*. 1993;329:977–986.
3. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*. 1998;352:837–853.
4. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. *BMJ*. 1998;317:703–713.
5. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical application of Diabetic Retinopathy Study (DRS) findings, DRS Report Number 8. The Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1981;88:583–600.
6. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch Ophthalmol*. 1985;103:1796–1806.
7. Witkin SR, Klein R. Ophthalmologic care for persons with diabetes. *JAMA*. 1984;251:2534–2537.
8. Klein R, Moss SE, Klein BE. New management concepts for persons with diabetes. *Diabetes Care*. 1987;10:633–638.
9. National Eye Institute. National Eye Health Education Program: From Vision Research to Eye Health: Planning the Partnership. NIH: Bethesda, MD 1999.
10. Smith RE, Patz A. Diabetes 2000—Closing the Gap. *Ophthalmology*. 1990;97:153–154.
11. Herman WH, Teutsch SM, Sepe SJ, et al. An approach to the prevention of blindness in diabetes. *Diabetes Care*. 1983;6:608–613.
12. Klein R, Klein BE. Diabetic eye disease. *Lancet*. 1997;350:197–204.
13. Schoenfeld ER, Greene JM, Wu SY, et al. Patterns of adherence to diabetes vision care guidelines: baseline findings from the Diabetic Retinopathy Awareness Program. *Ophthalmology*. 2001;108:563–571.
14. Bressler NM, Varma R, Doan QV, et al. Underuse of the health care system by persons with diabetes mellitus and diabetic macular edema in the United States. *JAMA Ophthalmol*. 2014;132:168–173.
15. Maclennan PA, McGwin Jr G, Heckemeyer C, et al. Eye care use among a high-risk diabetic population seen in a public hospital's clinics. *JAMA Ophthalmol*. 2014;132:162–167.
16. Bresnick GH. Diabetic retinopathy. In: Peyman GA, Sanders DR, Goldber MF, eds. *Principles and practice of ophthalmology*. Philadelphia: WB Saunders; 1977.
17. Klein R, Klein BE, Moss SE. Prevalence of diabetes in southern Wisconsin. *Am J Epidemiol*. 1984;119:54–61.
18. Kohner EM, Aldington SJ, Stratton IM, et al. United Kingdom Prospective Diabetes Study, 30: diabetic retinopathy at diagnosis of non-insulin-dependent diabetes mellitus and associated risk factors. *Arch Ophthalmol*. 1998;116:297–303.
19. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Arch Ophthalmol*. 1984;102:520–526.
20. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Arch Ophthalmol*. 1984;102:527–532.
21. Gass JDM. *Stereoscopic Atlas of Macular Disease.* 3rd ed. Mosby: St Louis 1997.
22. Klein R, Meuer SM, Moss SE, et al. The relationship of retinal microaneurysm counts to the 4-year progression of diabetic retinopathy. *Arch Ophthalmol*. 1989;107:1780–1785.
23. Dobree JH. Simple diabetic retinopathy. Evolution of the lesions and therapeutic considerations. *Br J Ophthalmol*. 1970;54:1–10.
24. Sussman EJ, Tsiaras WG, Soper KA. Diagnosis of diabetic eye disease. *JAMA*. 1982;247:3231–3234.
25. Four risk factors for severe visual loss in diabetic retinopathy. The third report from the Diabetic Retinopathy Study. The Diabetic Retinopathy Study Research Group. *Arch Ophthalmol*. 1979;97:654–655.
26. Aiello LP, Avery RL, Arrigg PG, et al. Vascular endothelial growth factor in ocular fluid of patients with diabetic retinopathy and other retinal disorders. *N Engl J Med*. 1994;331:1480–1487.
27. Adamis AP, Miller JW, Bernal MT, et al. Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am J Ophthalmol*. 1994;118:445–450.
28. Ip MS, Domalpally A, Hopkins JJ, et al. Long-term effects of ranibizumab on diabetic retinopathy severity and progression. *Arch Ophthalmol*. 2012;130:1145–1152.
29. Elman MJ, Aiello LP, Beck RW, et al. Randomized trial evaluating ranibizumab plus prompt or deferred laser or triamcinolone plus prompt laser for diabetic macular edema. *Ophthalmology*. 2010;117:1064–1077.
30. Elman MJ, Qin H, Aiello LP, et al. Intravitreal ranibizumab for diabetic macular edema with prompt versus deferred laser treatment: three-year randomized trial results. *Ophthalmology*. 2012;119:2312–2318.
31. Elman MJ, Bressler NM, Qin H, et al. Expanded 2-year follow-up of ranibizumab plus prompt or deferred laser or triamcinolone plus prompt laser for diabetic macular edema. *Ophthalmology*. 2011;118:609–614.
32. Early photocoagulation for diabetic retinopathy. ETDRS report number 9. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:766–785.
33. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:786–806.
34. Hovind P, Tarnow L, Rossing K, et al. Decreasing incidence of severe diabetic microangiopathy in type 1 diabetes. *Diabetes Care*. 2003;26:1258–1264.
35. Cruickshanks KJ, Vadheim CM, Moss SE, et al. Genetic marker associations with proliferative retinopathy in persons diagnosed with diabetes before 30 yr of age. *Diabetes*. 1992;41:879–885.
36. Wilkinson CP, Ferris III FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110:1677–1682.
37. Haffner SM, Fong D, Stern MP, et al. Diabetic retinopathy in Mexican Americans and non-Hispanic whites. *Diabetes*. 1988;37:878–884.
38. Hamman RF, Mayer EJ, Moo-Young GA, et al. Prevalence and risk factors of diabetic retinopathy in non-Hispanic whites and Hispanics with NIDDM. San Luis Valley Diabetes Study. *Diabetes*. 1989;38:1231–1237.
39. Nielsen NV. Diabetic retinopathy II. The course of retinopathy in diabetics treated with oral hypoglycaemic agents and diet regime alone. A one year epidemiological cohort study of diabetes mellitus. The Island of Falster. Denmark. *Acta Ophthalmol (Copenh)*. 1984;62:266–273.
40. Dorf A, Ballintine EJ, Bennett PH, et al. Retinopathy in Pima Indians. Relationships to glucose level, duration of diabetes, age at diagnosis of diabetes, and age at examination in a population with a high prevalence of diabetes mellitus. *Diabetes*. 1976;25:554–560.
41. Bennett PH, Rushforth NB, Miller M, et al. Epidemiologic studies of diabetes in the Pima Indians. *Recent Prog Horm Res*. 1976;32:333–376.
42. Kahn HA, Leibowitz HM, Ganley JP, et al. The Framingham Eye Study. I. Outline and major prevalence findings. *Am J Epidemiol*. 1977;106:17–32.
43. West KM, Erdreich LJ, Stober JA. A detailed study of risk factors for retinopathy and nephropathy in diabetes. *Diabetes*. 1980;29:501–508.
44. King H, Balkau B, Zimmet P, et al. Diabetic retinopathy in Nauruans. *Am J Epidemiol*. 1983;117:659–667.
45. Dwyer MS, Melton III LJ, Ballard DJ, et al. Incidence of diabetic retinopathy and blindness: a population-based study in Rochester, Minnesota. *Diabetes Care*. 1985;8:316–322.
46. Ballard DJ, Melton III LJ, Dwyer MS, et al. Risk factors for diabetic retinopathy: a population-based study in Rochester, Minnesota. *Diabetes Care*. 1986;9:334–342.
47. Danielsen R, Jonasson F, Helgason T. Prevalence of retinopathy and proteinuria in type 1 diabetics in Iceland. *Acta Med Scand*. 1982;212:277–280.

48. Constable IJ, Knuiman MW, Welborn TA, et al. Assessing the risk of diabetic retinopathy. *Am J Ophthalmol*. 1984;97:53–61.
49. Knuiman MW, Welborn TA, McCann VJ, et al. Prevalence of diabetic complications in relation to risk factors. *Diabetes*. 1986;35:1332–1339.
50. Sjolie AK. Ocular complications in insulin treated diabetes mellitus. An epidemiological study. *Acta Ophthalmol Suppl*. 1985;172:1–77.
51. Nielsen NV. Diabetic retinopathy I. The course of retinopathy in insulin-treated diabetics. A one year epidemiological cohort study of diabetes mellitus. The Island of Falster. Denmark. *Acta Ophthalmol (Copenh)*. 1984;62:256–265.
52. Teuscher A, Schnell H, Wilson PW. Incidence of diabetic retinopathy and relationship to baseline plasma glucose and blood pressure. *Diabetes Care*. 1988;11:246–251.
53. Jerneld B. Prevalence of diabetic retinopathy. A population study from the Swedish island of Gotland. *Acta Ophthalmol Suppl*. 1988;188:3–32.
54. McLeod BK, Thompson JR, Rosenthal AR. The prevalence of retinopathy in the insulin-requiring diabetic patients of an English country town. *Eye*. 1988;2:424–430.
55. Kostraba JN, Klein R, Dorman JS, et al. The epidemiology of diabetes complications study. IV. Correlates of diabetic background and proliferative retinopathy. *Am J Epidemiol*. 1991;133:381–391.
56. Fujimoto W, Fukuda M. Natural history of diabetic retinopathy and its treatment in Japan. In: Fukui I, ed. *Diabetes mellitus in Asia*. Excerpta Medica: Amsterdam;225–231 1967.
57. Nelson RG, Wolfe JA, Horton MB, et al. Proliferative retinopathy in NIDDM. Incidence and risk factors in Pima Indians. *Diabetes*. 1989;38:435–440.
58. Harris MI, Klein R, Cowie CC, et al. Is the risk of diabetic retinopathy greater in non-Hispanic blacks and Mexican Americans than in non-Hispanic whites with type 2 diabetes? A U.S. population study. *Diabetes Care*. 1998;21:1230–1235.
59. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. IX. Four-year incidence and progression of diabetic retinopathy when age at diagnosis is less than 30 years. *Arch Ophthalmol*. 1989;107:237–243.
60. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. X. Four-year incidence and progression of diabetic retinopathy when age at diagnosis is 30 years or more. *Arch Ophthalmol*. 1989;107:244–249.
61. Klein R, Knudtson MD, Lee KE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XXII. The twenty-five-year progression of retinopathy in persons with type 1 diabetes. *Ophthalmology*. 2008;115:1859–1868.
62. Klein R, Knudtson MD, Lee KE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy XXIII: The twenty-five-year incidence of macular edema in persons with type 1 diabetes. *Ophthalmology*. 2009;116:497–503.
63. Klein R, Lee KE, Gangnon RE, et al. The 25-year incidence of visual impairment in type 1 diabetes mellitus the Wisconsin Epidemiologic Study of Diabetic Retinopathy. *Ophthalmology*. 2010;117:63–70.
64. Varma R, Torres M, Pena F, et al. Prevalence of diabetic retinopathy in adult Latinos: the Los Angeles Latino eye study. *Ophthalmology*. 2004;111:1298–1306.
65. Varma R, Choudhury F, Klein R, et al. Four-year incidence and progression of diabetic retinopathy and macular edema: the Los Angeles Latino Eye Study. *Am J Ophthalmol*. 2010;149:752–761.
66. Rabb MF, Gagliano DA, Sweeney HE. Diabetic retinopathy in blacks. *Diabetes Care*. 1990;13:1202–1206.
67. Barbosa J, Ramsay RC, Knobloch WH, et al. Histocompatibility antigen frequencies in diabetic retinopathy. *Am J Ophthalmol*. 1980;90:148–153.
68. Dornan TL, Ting A, McPherson CK, et al. Genetic susceptibility to the development of retinopathy in insulin-dependent diabetics. *Diabetes*. 1982;31:226–231.
69. Rand LI, Krolewski AS, Aiello LM, et al. Multiple factors in the prediction of risk of proliferative diabetic retinopathy. *N Engl J Med*. 1985;313:1433–1438.
70. Jervell J, Solheim B. HLA-antigens in long standing insulin dependent diabetics with terminal nephropathy and retinopathy with and without loss of vision. *Diabetologia*. 1979;17:391.
71. Leslie RD, Pyke DA. Diabetic retinopathy in identical twins. *Diabetes*. 1982;31:19–21.
72. Hanis CL. Genetics of non-insulin dependent diabetes mellitus among Mexican Americans. Approaches and Perspectives. In: Berg K, Boulyjenkov V, Christen Y, eds. *Genetic Approaches To Noncommunicable Diseases*. Springer-Verlag: Berlin:65–77 1996.
73. Clustering of long-term complications in families with diabetes in the diabetes control and complications trial. The Diabetes Control and Complications Trial Research Group. *Diabetes*. 1997;46:1829–1839.
74. Iyengar SK, Abboud HE, Goddard KA, et al. Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: the family investigation of nephropathy and diabetes (FIND). *Diabetes*. 2007;56:1577–1585.
75. Imperatore G, Knowler WC, Nelson RG, et al. Genetics of diabetic nephropathy in the Pima Indians. *Curr Diab Rep*. 2001;1:275–281.
76. Kankova K, Stejskalova A, Pacal L, et al. Genetic risk factors for diabetic nephropathy on chromosomes 6p and 7q identified by the set-association approach. *Diabetologia*. 2007;50:990–999.
77. Placha G, Canani LH, Warram JH, et al. Evidence for different susceptibility genes for proteinuria and ESRD in type 2 diabetes. *Adv Chronic Kidney Dis*. 2005;12:155–169.
78. Tong Z, Yang Z, Patel S, et al. Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc Natl Acad Sci U S A*. 2008;105:6998–7003.
79. Frank RN, Hoffman WH, Podgor MJ, et al. Retinopathy in juvenile-onset type I diabetes of short duration. *Diabetes*. 1982;31:874–882.
80. Knowles Jr HC, Guest GM, Lampe J, et al. The course of juvenile diabetes treated with unmeasured diet. *Diabetes*. 1965;14:239–273.
81. Palmberg P, Smith M, Waltman S, et al. The natural history of retinopathy in insulin-dependent juvenile-onset diabetes. *Ophthalmology*. 1981;88:613–618.
82. Klein R, Klein BE, Moss SE, et al. Retinopathy in young-onset diabetic patients. *Diabetes Care*. 1985;8:311–315.
83. Klein BE, Moss SE, Klein R. Is menarche associated with diabetic retinopathy? *Diabetes Care*. 1990;13:1034–1038.
84. Pieters GF, Smals AG, Kloppenborg PW. Defective suppression of growth hormone after oral glucose loading in adolescence. *J Clin Endocrinol Metab*. 1980;51:265–270.
85. Klein R, Klein BE, Moss SE, et al. Blood pressure and hypertension in diabetes. *Am J Epidemiol*. 1985;122:75–89.
86. Blethen SL, Sargeant DT, Whitlow MG, et al. Effect of pubertal stage and recent blood glucose control on plasma somatomedin C in children with insulin-dependent diabetes mellitus. *Diabetes*. 1981;30:868–872.
87. Allen C, Zaccaro DJ, Palta M, et al. Glycemic control in early IDDM. The Wisconsin Diabetes Registry. *Diabetes Care*. 1992;15:980–987.
88. Sizonenko PC. Endocrinology in preadolescents and adolescents. I. Hormonal changes during normal puberty. *Am J Dis Child*. 1978;132:704–712.
89. Haffner SM, Klein R, Dunn JF, et al. Increased testosterone in type I diabetic subjects with severe retinopathy. *Ophthalmology*. 1990;97:1270–1274.
90. Aiello LP, Cahill MT, Wong JS. Systemic considerations in the management of diabetic retinopathy. *Am J Ophthalmol*. 2001;132:760–766.
91. Deleted in review.
92. Klein R, Klein BE, Moss SE, et al. The Beaver Dam Eye Study. Retinopathy in adults with newly discovered and previously diagnosed diabetes mellitus. *Ophthalmology*. 1992;99:58–62.
93. UK Prospective Diabetes Study 6. Complications in newly diagnosed type 2 diabetic patients and their association with different clinical and biochemical risk factors. *Diabetes Res*. 1990;13:1–11.
94. Guidelines for eye care in patients with diabetes mellitus. Results of a symposium. The Kentucky Diabetic Retinopathy Group. *Arch Intern Med*. 1989;149:769–770.
95. Diabetic retinopathy. American Diabetes Association. *Diabetes Care*. 1998;21:157–159.
96. Klein R, Klein BE, Moss SE, et al. Glycosylated hemoglobin predicts the incidence and progression of diabetic retinopathy. *JAMA*. 1988;260:2864–2871.
97. Klein R, Klein BE, Moss SE, et al. Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. *Arch Intern Med*. 1994;154:2169–2178.

98. Chase HP, Jackson WE, Hoops SL, et al. Glucose control and the renal and retinal complications of insulin-dependent diabetes. *JAMA*. 1989;261:1155–1160.
99. Effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus. *JAMA*. 2002;287:2563–2569.
100. Nathan DM. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*. 2014;37:9–16.
101. Aiello LP. Diabetic retinopathy and other ocular findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabetes Care*. 2014;37:17–23.
102. Chew EY, Ambrosius WT, Davis MD, et al. Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*. 2010;363:233–244.
103. Chew EY, Klein ML, Ferris III FL, et al. Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy. Early Treatment Diabetic Retinopathy Study (ETDRS) Report 22. *Arch Ophthalmol*. 1996;114:1079–1084.
104. American Diabetes Association. Standards of medical care in diabetes—2013. *Diabetes Care*. 2013;36:S11–S66.
105. Kohner EM. Diabetic retinopathy. *Br Med Bull*. 1989;45:148–173.
106. Knowler WC, Bennett PH, Ballintine EJ. Increased incidence of retinopathy in diabetics with elevated blood pressure. A six-year follow-up study in Pima Indians. *N Engl J Med*. 1980;302:645–650.
107. Klein R, Klein BE, Moss SE, et al. Is blood pressure a predictor of the incidence or progression of diabetic retinopathy? *Arch Intern Med*. 1989;149:2427–2432.
108. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XVII. The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes. *Ophthalmology*. 1998;10:1801–1815.
109. Fuller JH, Shipley MJ, Rose G, et al. Mortality from coronary heart disease and stroke in relation to degree of glycaemia: the Whitehall study. *Br Med J (Clin Res Ed)*. 1983;287:867–870.
110. Jarrett RJ, McCartney P, Keen H. The Bedford survey: ten year mortality rates in newly diagnosed diabetics, borderline diabetics and normoglycaemic controls and risk indices for coronary heart disease in borderline diabetics. *Diabetologia*. 1982;22:79–84.
111. Palumbo PJ, Elveback LR, Whisnant JP. Neurologic complications of diabetes mellitus: transient ischemic attack, stroke, and peripheral neuropathy. *Adv Neurol*. 1978;19:593–601.
112. Asplund K, Hagg E, Helmers C, et al. The natural history of stroke in diabetic patients. *Acta Med Scand*. 1980;207:417–424.
113. Mogensen CE. Long-term antihypertensive treatment (over six years) inhibiting the progression of diabetic nephropathy. *Acta Endocrinol Suppl (Copenh)*. 1981;242:31–32.
114. Parving HH, Andersen AR, Hommel E, et al. Effects of long-term antihypertensive treatment on kidney function in diabetic nephropathy. *Hypertension*. 1985;7:II114–II117.
115. Nielsen NV. Diabetic retinopathy I. The course of retinopathy in insulin-treated diabetics. A one year epidemiological cohort study of diabetes mellitus. The Island of Falster, Denmark. *Acta Ophthalmol (Copenh)*. 1984;62:256–265.
116. Klein BE, Moss SE, Klein R, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. XIII. Relationship of serum cholesterol to retinopathy and hard exudate. *Ophthalmology*. 1991;98:1261–1265.
117. Cullen JF, Ireland JT, Oliver MF. A controlled trial of Atromid therapy in exudative diabetic retinopathy. *Trans Ophthalmol Soc U K*. 1964;84:281–295.
118. Duncan LJ, Cullen JF, Ireland JT, et al. A three-year trial of atromid therapy in exudative diabetic retinopathy. *Diabetes*. 1968;17:458–467.
119. Borch-Johnsen K, Kreiner S. Proteinuria: value as predictor of cardiovascular mortality in insulin dependent diabetes mellitus. *Br Med J (Clin Res Ed)*. 1987;294:1651–1654.
120. Winocour PH, Durrington PN, Ishola M, et al. Influence of proteinuria on vascular disease, blood pressure, and lipoproteins in insulin dependent diabetes mellitus. *Br Med J (Clin Res Ed)*. 1987;294:1648–1651.
121. Klein R, Klein BE, Moss SE. Epidemiology of proliferative diabetic retinopathy. *Diabetes Care*. 1992;15:1875–1891.
122. Klein R, Klein BE, Davis MD. Is cigarette smoking associated with diabetic retinopathy? *Am J Epidemiol*. 1983;118:228–238.
123. Moss SE, Klein R, Klein BE. Association of cigarette smoking with diabetic retinopathy. *Diabetes Care*. 1991;14:119–126.
124. The Health Consequences of Smoking. A Report to the Surgeon General. U.S. Department of Health, Education and Welfare: Washington, DC.
125. Doyle JT. Risk factors in arteriosclerosis and cardiovascular disease with special emphasis on cigarette smoking. *Prev Med*. 1979;8:264–270.
126. Kingsley LA, Dorman JS, Doft BH, et al. An epidemiologic approach to the study of retinopathy: the Pittsburgh diabetic morbidity and retinopathy studies. *Diabetes Res Clin Pract*. 1988;4:99–109.
127. Young RJ, McCulloch DK, Prescott RJ, et al. Alcohol: another risk factor for diabetic retinopathy? *Br Med J (Clin Res Ed)*. 1984;288:1035–1037.
128. Moss SE, Klein R, Klein BE. Alcohol consumption and the prevalence of diabetic retinopathy. *Ophthalmology*. 1992;99:926–932.
129. Jakubowski JA, Vaillancourt R, Deykin D. Interaction of ethanol, prostacyclin, and aspirin in determining human platelet reactivity in vitro. *Arteriosclerosis*. 1988;8:436–441.
130. Klein BE, Moss SE, Klein R. Effect of pregnancy on progression of diabetic retinopathy. *Diabetes Care*. 1990;13:34–40.
131. Rodman HM, Singerman LJ, Aiello LM, et al. Diabetic retinopathy and its relationship to pregnancy.In: Merkatz ER, Adams PAJ, eds. *The Diabetic Pregnancy: A Perinatal Perspective*. Grune and Stratton: New York. 1979
132. Klein BE, Klein R, Meuer SM, et al. Does the severity of diabetic retinopathy predict pregnancy outcome? *J Diabet Complications*. 1988;2:179–184.
133. Dornan TL, Carter RD, Bron AJ, et al. Low density lipoprotein cholesterol: an association with the severity of diabetic retinopathy. *Diabetologia*. 1982;22:167–170.
134. Miccoli R, Odello G, Giampietro O, et al. Circulating lipid levels and severity of diabetic retinopathy in type I diabetes mellitus. *Ophthalmic Res*. 1987;19:52–56.
135. Wong TY, Klein R, Klein BE, et al. Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality. *Surv Ophthalmol*. 2001;46:59–80.
136. Wong TY, Shankar A, Klein R, et al. Retinal vessel diameters and the incidence of gross proteinuria and renal insufficiency in people with type 1 diabetes. *Diabetes*. 2004;53:179–184.
137. Ederer F, Hiller R, Taylor HR. Senile lens changes and diabetes in two population studies. *Am J Ophthalmol*. 1981;91:381–395.
138. Klein BE, Klein R, Wang Q, et al. Older-onset diabetes and lens opacities. The Beaver Dam Eye Study. *Ophthalmic Epidemiol*. 1995;2:49–55.
139. Klein B, Klein R, Moss SE. Prevalence of cataracts in a population-based study of persons with diabetes mellitus. *Ophthalmology*. 1985;92:1191–1196.
140. Klein R, Klein BE. Vision Disorders. In: Harrris, MWH, ed. *Diabetes in America*. 2nd Edition. NIH-NIDDK. 1995. 293-338.
141. Kahn HA, Leibowitz HM, Ganley JP, et al. The Framingham Eye Study. II. Association of ophthalmic pathology with single variables previously measured in the Framingham Heart Study. *Am J Epidemiol*. 1977;106:33–41.
142. Bengtsson B. Aspects of the epidemiology of chronic glaucoma. *Acta Ophthalmol Suppl*. 1981;146:1–48.
143. Klein BE, Klein R, Jensen SC. Open-angle glaucoma and older-onset diabetes. The Beaver Dam Eye Study. *Ophthalmology*. 1994;101:1173–1177.
144. Howie LJ, Drury TF. Current estimates from the Health Interview Survey: United States—1977. *Vital Health Stat*. 1978;10:1–98.
145. Pasquale LR, Kang JH, Manson JE, et al. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology*. 2006;113:1081–1086.
146. Klein R, Klein BE, Moss SE. Visual impairment in diabetes. *Ophthalmology*. 1984;91:1–9.
147. Moss SE, Klein R, Klein BE. The incidence of vision loss in a diabetic population. *Ophthalmology*. 1988;95:1340–1348.
148. Klein R, Klein BE. Are individuals with diabetes seeing better?: a long-term epidemiological perspective. *Diabetes*. 2010;59:1853–1860.
149. Ko F, Vitale S, Chou CF, et al. Prevalence of nonrefractive visual impairment in US adults and associated risk factors, 1999-2002 and 2005-2008. *JAMA*. 2012;308:2361–2368.

150. Klein R, Klein BE. Is the prevalence of visual impairment rising or falling in the people with diabetes mellitus? It depends on who you study. *JAMA Ophthalmol*. 2013;131:948–950.
151. Pugh JA, Jacobson JM, van Heuven WA, et al. Screening for diabetic retinopathy. The wide-angle retinal camera. *Diabetes Care*. 1993;16:889–895.
152. Nathan DM, Fogel HA, Godine JE, et al. Role of diabetologist in evaluating diabetic retinopathy. *Diabetes Care*. 1991;14:26–33.
153. Klein R, Klein BE, Neider MW, et al. Diabetic retinopathy as detected using ophthalmoscopy, a nonmydriatic camera and a standard fundus camera. *Ophthalmology*. 1985;92:485–491.
154. Klein R, Klein BE, Moss SE. The Wisconsin epidemiological study of diabetic retinopathy: a review. *Diabetes Metab Rev*. 1989;5:559–570.
155. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. *VI. Retinal photocoagulation. Ophthalmology*. 1987;94:747–753.
156. Klein R, Moss SE, Klein BE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: VIII. The incidence of retinal photocoagulation. *J Diabet Complications*. 1988;2:79–87.
157. Stross JK, Harlan WR. The dissemination of new medical information. *JAMA*. 1979;241:2622–2624.
158. Sprafka JM, Fritsche TL, Baker R, et al. Prevalence of undiagnosed eye disease in high-risk diabetic individuals. *Arch Intern Med*. 1990;150:857–861.
159. Brechner RJ, Cowie CC, Howie LJ, et al. Ophthalmic examination among adults with diagnosed diabetes mellitus. *JAMA*. 1993;270:1714–1718.
160. Payne TH, Gabella BA, Michael SL, et al. Preventive care in diabetes mellitus. Current practice in urban health-care system. *Diabetes Care*. 1989;12:745–747.
161. Newcomb PA, Klein R. Factors associated with compliance following diabetic eye screening. *J Diabet Complications*. 1990;4:8–14.
162. Moss SE, Klein R, Klein BE. Factors associated with having eye examinations in persons with diabetes. *Arch Fam Med*. 1995;4:529–534.
163. Klein R, Klein BE, Moss SE, et al. The validity of a survey question to study diabetic retinopathy. *Am J Epidemiol*. 1986;124:104–110.
164. Hess RG, Lengyel MC, Hess GE, et al. Diabetes in communities. University of Michigan: Ann Arbor. 1986
165. Javitt JC, Aiello LP. Cost-effectiveness of detecting and treating diabetic retinopathy. *Ann Intern Med*. 1996;124:164–169.
166. Javitt JC, Canner JK, Frank RG, et al. Detecting and treating retinopathy in patients with type I diabetes mellitus. A health policy model. *Ophthalmology*. 1990;97:483–494.
167. Dasbach EJ, Fryback DG, Newcomb PA, et al. Cost-effectiveness of strategies for detecting diabetic retinopathy. *Med Care*. 1991;29:20–39.
168. Fong DS, Aiello L, Gardner TW, et al. Diabetic retinopathy. *Diabetes Care*. 2003;26:S99–S102.
169. Kirkizlar E, Serban N, Sisson JA, et al. Evaluation of telemedicine for screening of diabetic retinopathy in the Veterans Health Administration. *Ophthalmology*. 2013;120:2604–2610.
170. Silva PS, Cavallerano JD, Aiello LM, et al. Telemedicine and diabetic retinopathy: moving beyond retinal screening. *Arch Ophthalmol*. 2011;129:236–242.
171. Olafsdottir E, Andersson DK, Stefansson E. Visual acuity in a population with regular screening for type 2 diabetes mellitus and eye disease. *Acta Ophthalmol Scand*. 2007;85:40–45.
172. Fong DS, Sharza M, Chen W, et al. Vision loss among diabetics in a group model Health Maintenance Organization (HMO). *Am J Ophthalmol*. 2002;133:236–241.
173. Prasad S, Kamath GG, Jones K, et al. Prevalence of blindness and visual impairment in a population of people with diabetes. *Eye (Lond)*. 2001;15:640–643.
174. Kirkizlar E, Serban N, Sisson JA, et al. Evaluation of telemedicine for screening of diabetic retinopathy in the Veterans Health Administration. *Ophthalmology*. 2013;120:2604–2610.
175. Zoega GM, Gunnarsdottir T, Bjornsdottir S, et al. Screening compliance and visual outcome in diabetes. *Acta Ophthalmol Scand*. 2005;83:687–690.
176. Forster AS, Forbes A, Dodhia H, et al. Changes in detection of retinopathy in type 2 diabetes in the first 4 years of a population-based diabetic eye screening program: retrospective cohort study. *Diabetes Care*. 2013;36:2663–2669.
177. Trautner C, Haastert B, Giani G, et al. Incidence of blindness in southern Germany between 1990 and 1998. *Diabetologia*. 2001;44:147–150.
178. Early worsening of diabetic retinopathy in the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 1998;116:874–886.
179. Dahl-Jorgensen K, Brinchmann-Hansen O, Hanssen KF, et al. Rapid tightening of blood glucose control leads to transient deterioration of retinopathy in insulin dependent diabetes mellitus: the Oslo study. *Br Med J (Clin Res Ed)*. 1985;290:811–815.
180. Dandona P, Bolger JP, Boag F, et al. Rapid development and progression of proliferative retinopathy after strict diabetic control. *Br Med J (Clin Res Ed)*. 1985;290:895–896.
181. Funatsu H, Yamashita H, Ohashi Y, et al. Effect of rapid glycemic control on progression of diabetic retinopathy. *Jpn J Ophthalmol*. 1992;36:356–367.
182. White NH, Sun W, Cleary PA, et al. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus: 10 years after the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 2008;126:1707–1715.
183. Holman RR, Paul SK, Bethel MA, et al. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med*. 2008;359:1565–1576.
184. Preliminary report on effects of photocoagulation therapy. The Diabetic Retinopathy Study Research Group. *Am J Ophthalmol*. 1976;81:383–396.
185. Photocoagulation treatment of proliferative diabetic retinopathy. The second report of diabetic retinopathy study findings. *Ophthalmology*. 1978;85:82–106.
186. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study Report no. 4. The Early Treatment Diabetic Retinopathy Study Research Group. *Int Ophthalmol Clin*. 1987;27:265–272.
187. Ferris F. Early photocoagulation in patients with either type I or type II diabetes. *Trans Am Ophthalmol Soc*. 1996;94:505–537.
188. Ferris III FL, Davis MD, Aiello LM. Treatment of diabetic retinopathy. *N Engl J Med*. 1999;341:667–678.
189. Nguyen QD, Brown DM, Marcus DM, et al. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*. 2012;119:789–801.
190. Brown DM, Nguyen QD, Marcus DM, et al. Long-term outcomes of ranibizumab therapy for diabetic macular edema: the 36-month results from two phase III trials: RISE and RIDE. *Ophthalmology*. 2013;120:2013–2022.
191. Avery RL. Regression of retinal and iris neovascularization after intravitreal bevacizumab (Avastin) treatment. *Retina*. 2006;26:352–354.
192. Wakabayashi T, Oshima Y, Sakaguchi H, et al. Intravitreal bevacizumab to treat iris neovascularization and neovascular glaucoma secondary to ischemic retinal diseases in 41 consecutive cases. *Ophthalmology*. 2008;115:1571–1580.
193. Sinzato R, Fukino O, Tamai H, et al. Coping behaviors of severe diabetics. *Psychother Psychosom*. 1985;43:219–226.
194. Bain SC, Gill GV, Dyer PH, et al. Characteristics of Type 1 diabetes of over 50 years duration (the Golden Years Cohort). *Diabet Med*. 2003;20:808–811.
195. Keenan HA, Costacou T, Sun JK, et al. Clinical factors associated with resistance to microvascular complications in diabetic patients of extreme disease duration: the 50-year medalist study. *Diabetes Care*. 2007;30:1995–1997.
196. Sun JK, Keenan HA, Cavallerano JD, et al. Protection from retinopathy and other complications in patients with type 1 diabetes of extreme duration: the Joslin 50-year medalist study. *Diabetes Care*. 2011;34:968–974.

44

Diabetic Neuropathy

Lynn Ang, Kara Mizokami-Stout, and Rodica Pop-Busui

OUTLINE

Definition of Diabetic Neuropathies, 756
Classification of Diabetic Neuropathies, 756
  Pathogenesis, 756
Distal Symmetrical Polyneuropathy, 757
  Epidemiology, 757
  Screening and Diagnosis in Clinical Care, 758
Distal Symmetric Polyneuropathy Evaluation in Clinical Research, 759
  Secondary Distal Symmetric Polyneuropathy Complications: Diabetic Foot Ulcers/Charcot Neuroarthropathy, 760
  Distal Symmetric Polyneuropathy Treatment, 761
Diabetic Autonomic Neuropathies, 762
  Cardiovascular Autonomic Neuropathy, 762
  Cardiovascular Autonomic Neuropathy Evaluation in Clinical Research, 763
    Treatment, *765*
  Gastrointestinal Autonomic Neuropathy, 765
  Urogenital Autonomic Neuropathy, 766
  Sudomotor Dysfunction, 767
Atypical Neuropathies, 767
  Mononeuropathies, 767
  Diabetic Radiculoplexus Neuropathy, 767
Conclusions, 767

KEY POINTS

- Early recognition and appropriate management of neuropathy in all patients with diabetes is essential to prevent serious complications.
- Diabetic neuropathy is a diagnosis of exclusion.
- Nondiabetic neuropathies mimicking diabetic neuropathy may be present in patients with diabetes and may be treatable by specific measures.
- Up to 50% of diabetic peripheral neuropathies may be asymptomatic.
- Various forms of peripheral and autonomic neuropathies have been described in patients with prediabetes/metabolic syndrome and obesity, as well as in youth.
- If not recognized, and if preventive foot care is not implemented, patients are at risk of injuries to their insensate feet.
- Several treatment options exist for painful diabetic peripheral neuropathy and some forms of autonomic neuropathy.
- Recognition and treatment of autonomic neuropathy may improve symptoms, reduce sequelae, and improve quality of life.

DEFINITION OF DIABETIC NEUROPATHIES

The consensus in the field is that diabetic neuropathies are defined by the presence of symptoms and/or signs of peripheral and autonomic nerve dysfunction in people with diabetes after the exclusion of other causes confirmed through a careful clinical examination and appropriate differential tests.[1,2] However, emerging evidence reports that various forms of peripheral and autonomic neuropathies similar to diabetic neuropathies have been described in patients with prediabetes/metabolic syndrome and obesity.[1,3]

CLASSIFICATION OF DIABETIC NEUROPATHIES

Among the various forms of diabetic neuropathy, distal symmetric polyneuropathy (DSPN) and diabetic autonomic neuropathies, particularly cardiovascular autonomic neuropathy (CAN), are by far the most studied.[1,4,5] There are several atypical forms of diabetic neuropathy as well.[1,6] A comprehensive classification of diabetic neuropathies has been recently updated by the American Diabetes Association (ADA) in the 2017 Statement on Diabetic Neuropathy and is shown in Fig. 44.1.[1]

Pathogenesis

Diabetic neuropathy is a unique neurodegenerative disorder of the peripheral nervous system that preferentially targets sensory axons, autonomic axons, and later, to a lesser extent, motor axons.[3] Experimental studies suggest a multifactorial pathogenesis of diabetic neuropathy, but how diabetes targets sensory neurons remains debatable. Although diabetic neuropathy is not considered primarily a demyelinating neuropathy, Schwann cells are targeted by chronic hyperglycemia, and more severe forms of human diabetic neuropathy do include features of demyelination.[3] There is significant experimental evidence to support the idea that the entire neuronal tree, from perikarya to terminal, including alterations in sensory neurons within dorsal root ganglia, may be targeted by diabetes.[3] A prevailing view of the pathogenesis is that hyperglycemia and hyperlipidemia lead to increased flux of glucose metabolized through the polyol and hexosamine pathways, oxidative and inflammatory stress, excessive free fatty acids as well as alterations in substrate energy metabolism, mitochondrial dysfunction, and a maladaptive integrated stress response, as well as lipid oxidation and nonenzymatic glycation in the context of metabolic dysfunction, all leading to damage to nerve

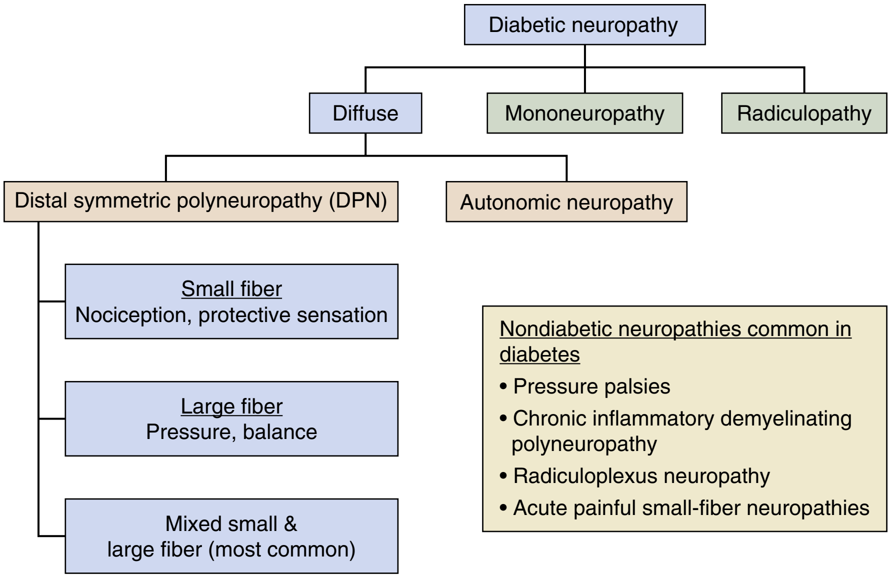

Diabetic neuropathy
Diffuse
Mononeuropathy
Radiculopathy
Distal symmetric polyneuropathy (DPN)
Autonomic neuropathy
Small fiber
Nociception, protective sensation
Large fiber
Pressure, balance
Mixed small & large fiber (most common)
Nondiabetic neuropathies common in diabetes
• Pressure palsies
• Chronic inflammatory demyelinating polyneuropathy
• Radiculoplexus neuropathy
• Acute painful small-fiber neuropathies

Figure 44.1 Classification of diabetic neuropathies.

cells.[1,3] A detailed review of the intricate pathogenic mechanisms is however beyond the scope of this chapter, and the reader is referred to other comprehensive reviews.[3]

KEY POINTS

- The mechanisms leading to the development of diabetic neuropathies are intricate, involving complex interactions between hyperglycemia, hyperlipidemia, and other risk factors leading to inflammatory and oxidative stress, mitochondrial dysfunction, impaired substrate energy metabolism with disruption in whole-nerve bioenergetics, and ultimately loss of function and degeneration of small and large nerve fibers.

DISTAL SYMMETRICAL POLYNEUROPATHY

DSPN is the most common form of diabetic neuropathy, and most of the available evidence regarding the epidemiology, pathogenesis, and treatment is relevant to this form.[1,2]

Epidemiology

DSPN is the most common complication of both type 1 diabetes (T1D) and type 2 diabetes (T2D), with an estimated lifetime prevalence exceeding 50%.[1,2] Precise prevalence and incidence estimates for DSPN have been also hampered by limitations in the outcome measures selected to define DSPN in various large cohorts, as well as by a rather high heterogeneity in the diagnostic criteria. There are differences in the epidemiology and natural history of diabetes neuropathy between T1D and T2D, despite no major structural differences in nerve pathology.

As demonstrated by the Diabetes Control and Complications Trial (DCCT), arguably the largest and best phenotyped cohort of patients with T1D with comprehensive and repeated measures of DSPN, the prevalence of DSPN is quite low in patients with early T1D, at approximately 5% at baseline.[1,7] However, among DCCT participants who were considered to be nonneuropathic at baseline, the prevalence of an abnormal neurologic exam was almost 20% in those on conventional treatment and almost 10% in those on intensive treatment after approximately 5 years of follow-up.[4,7] Moreover, the prevalence of DSPN increased during observational follow-up of the DCCT cohort in the Epidemiology of Diabetes Interventions and Complications (EDIC) study, and the most recent analysis reported that, over 23 years of mean follow-up, 33% of DCCT/EDIC participants developed confirmed DSPN.[4] Beyond glycemia as defined by time-weighted hemoglobin A1c and diabetes duration, additional risk factors for development of DSPN in the DCCT/EDIC cohort were diabetic kidney disease, elevated blood pressure (BP), height, and elevated triglycerides.[4] The European Insulin-Dependent Diabetes Mellitus Prospective Complications Study (EURODIAB IDDM), a large cross-sectional cohort of randomly selected patients with T1D from 16 European countries, reported mean prevalence rates of DSPN of 28%, but with significant DSPN trends associated with age, diabetes duration, and hemoglobin A1c, as well as with hypertension and hyperlipidemia.[8] However, these earlier cohorts may not be entirely representative of contemporary populations. Thus, the most recent data from the T1D Exchange, using the validated Michigan Neuropathy Screening Instrument (MNSI) questionnaire in a large cohort of approximately 6000 T1D participants, reported a DSPN prevalence of 11% after a median of 32 years of diabetes duration.[5] Risk factors for DSPN in this cohort include traditional risk factors such as older age, longer duration of diabetes, poor glycemic control, other microvascular complications, established cardiovascular disease (CVD), and CVD risk factors including body mass index, smoking history. and high triglycerides. However, the T1D Exchange analyses also unveiled a strong association of DSPN with female gender, as well as with several socioeconomic factors that had not been described before.[5] Interestingly, very similar findings were unveiled in another contemporary T1D cohort, the Scottish T1D cohort, that used the same MNSI questionnaire to assess DSPN in approximately 6000 participants, reporting prevalence rates of approximately 13% and also confirming similar risk factors, as well as the role of socioeconomic factors.[9]

DSPN may be present in at least 10% to 15% of newly diagnosed patients with T2D, with rates increasing to 50% after 10 years of disease duration.[1] For instance, in more than 5000 T2D participants in the Glycemia Reduction Approaches in Diabetes - A Comparative Effectiveness (GRADE) trial, approximately 21% of this contemporary cohort with a mean known T2D duration of only 4 years presented with DSPN at baseline as diagnosed with the MNSI (both examination and questionnaire).[10] In a cohort of greater than 1500 individuals with screen-detected T2D participating in the ADDITION-Denmark Study, the prevalence of DSPN assessed with the MNSI questionnaire was approximately 13% at baseline, with cumulative incidence rates of 10% over 13 years of follow-up.[11] In contrast, the Bypass Angioplasty Revascularization Intervention 2 Diabetes (BARI 2D), which enrolled greater than 2300 participants with T2D with a mean diabetes duration of approximately 10 years and confirmed coronary artery disease, found that the prevalence of DSPN at baseline as assessed by MNSI was approximately 50%.[12] The rates of DSPN are higher in various ethnic groups. Recent data show, for instance, that American Indians, besides having disproportionally higher rates of T2D, also develop more chronic complications, including DSPN.[13]

In addition, emerging evidence demonstrates that DSPN rates in youth with T1D and T2D approach those observed in adult populations. The most recent report from the SEARCH study including approximately 2000 youth with T1D and T2D reported DPSN prevalence rates as high as 22% in youth with T2D.[14] There is also emerging evidence that DSPN, especially the painful small fiber neuropathy subtype, may be present in 10% to 30% of subjects with impaired glucose tolerance, prediabetes, or metabolic syndrome.[1]

KEY POINTS

- Distal symmetric polyneuropathy (DSPN) is the most common form of diabetic neuropathy.
- DSPN prevalence increases with disease duration, with an estimated lifetime prevalence of ~35% in type 1 diabetes (T1D) and exceeding 50% in type 2 diabetes (T2D).
- DSPN remains prevalent in contemporary cohorts of T1D and T2D across the world, despite changes in standards of care.
- High rates of DSPN have been reported in prediabetes and in youth.

Screening and Diagnosis in Clinical Care

DSPN involves a loss of sensory function beginning distally that is also characterized by pain and substantial morbidity.[3] It is usually of gradual onset and may be present at the diagnosis of T2D in greater than 10% of subjects.[1,2]

DSPN Symptoms: Symptoms of DSPN vary depending on the class of sensory nerve fibers involved. The earliest affected nerve fibers are usually the small myelinated and unmyelinated fibers, thus the most common early symptoms are pain and dysesthesias (unpleasant sensations of burning).[1,2] Neuropathic pain may be the first symptom that prompts patients to seek medical care and is present in up to 25% to 30% of individuals with DSPN.[1-3] Characteristically, the pain is burning, lancinating, tingling, or shooting (electric shock–like), with paresthesias, occurring in varying combinations, and is typically worse at night.[1,2] Neuropathic pain may be accompanied by an exaggerated response to painful stimuli (hyperalgesia) and pain evoked by contact, e.g., with socks, shoes, and bedclothes (allodynia).[1,2] Neuropathic pain can lead to interference with daily activities, disability, psychosocial impairment, and reduced health-related quality of life.[15-17] The direct and indirect economic burden associated with neuropathic pain is substantial.[17,18]

The involvement of large fibers may cause numbness, tingling without pain, and loss of protective sensation. Patients can also initially present with a completely insensate, numb foot, stating their feet feel like they are wrapped in wool, or they are walking on thick socks.[1,19]

DSPN Signs: There are several established clinical tests that may be easily used to assess small- and large-fiber function, as summarized in Table 44.1. In clinical care, assessment of pinprick and temperature (mostly cold) sensation provides reliable information on small-fiber function, while assessment of vibration perception (with a 128-Hz tuning fork), proprioception, light touch to a 10-g monofilament, and ankle reflexes allow assessments of large-fiber function,[1] all in a "stocking and glove" distribution. Pinprick is tested using a sharp object, such as a safety pin, that is discarded after each patient, temperature is tested by application of a cool material such as a metallic object, and proprioception is examined by small movements of the distal interphalangeal joint of the great toe.[1,3] Assessment of light touch perception using a 10-g monofilament should be performed on the dorsal aspect of the great toe bilaterally.[1] However, it is important to consider that the 10-g monofilament is a useful clinical tool mainly for detecting advanced neuropathy and identifying patients at increased risk of ulceration and amputation.[1,2] Loss of ankle reflexes occurs early, and later weakness of small foot muscles and dorsiflexors is also observed.[1,3]

TABLE 44.1 Clinical Symptoms and Signs of DSPN

	Large Myelinated Nerve Fibers	Small Nerve Fibers
Function	Pressure, balance	Nociception; protective sensation
Symptoms	Numbness Tingling Poor balance	Pain: Burning Electric shocks Stabbing pain
Examination (clinically diagnostic)**	Ankle reflexes Reduced Abolished Vibration perception* Reduced Absent 10-g monofilament* Reduced Absent Proprioception impaired	Thermal (cold/hot) discrimination* Reduced Absent Pinprick sensation* Reduced Absent

*Documented impairment/loss of symmetry, as well as a distal to proximal pattern.

Additional Clinical Pearls: All these assessments should follow the typical DSPN pattern, starting distally (the dorsal aspect of the hallux) on both sides and move proximally until a sensory threshold is identified[1,2] (Table 44.1). These sensory modalities are tested initially by application of the sensory stimulus to a body site where normal responses are expected, such as the forehead. Then the stimulus is applied to the great toe, and then moved proximally up the limb to the level where the sensation is felt to be normal. Combining at least two examinations will increase the sensitivity and specificity of detecting DSPN, as demonstrated in several cohorts of patients with T1D and T2D including children and adolescents.[1,2,7,14,20-22]

The diagnosis of DSPN is principally a clinical one (Table 44.1). A combination of typical symptomatology and symmetrical distal sensory loss, or typical signs in the absence of symptoms, in a patient with diabetes is highly suggestive of DSPN, and may not require additional evaluation or referral.[1,2] Electrophysiological testing or referral to a neurologist is rarely needed for diagnosis, except in situations where the clinical features are atypical, the diagnosis is unclear, and a different etiology is suspected.[1,23] However, the presence of atypical features that include motor deficits greater than sensory, asymmetry of symptoms and signs, and rapid progression warrants a timely referral.[1]

It is also very important to remember that several peripheral neuropathies due to causes other than diabetes may coexist in patients with diabetes, may mimic diabetic neuropathy, and may be treatable (Fig. 44.2). Thus, undertaking a thorough family and medication history and performing relevant investigations are helpful to assess other potential etiologies, such as alcohol abuse, genetic neuropathies, neoplasia, toxic exposure, and amyloidosis. Laboratory screening includes vitamin B12 levels (methylmalonic acid with or without homocysteine may provide additional information) to evaluate for B12 deficiency (particularly in patients treated with metformin), thyroid function

Metabolic disease Thyroid Renal	**Systemic disease** Vasculitis Paraproteinemia Amyloidosis	**Infectious** HIV Hepatitis B Lyme disease
Inflammatory Chronic inflammatory demyelinating polyradiculoneuropathy	**Nutritional** Vitamin B12 deficiency* Pyridoxine Thiamine Tocopherol	**Industrial agents** Acrylamide Organophosphorous agents
Drugs Alcohol Amiodarone Chemotherapy	**Heavy metals** Arsenic Mercury	**Hereditary**

*B12 deficiency is more commonly associated with malabsorption rather than nutritional deficiency

Figure 44.2 Differential diagnosis of DSPN.

tests, complete blood count, metabolic panel, and serum immunoelectrophoresis with immunofixation to evaluate for a monoclonal gammopathy and may ensure timely treatment.[1,3] Although an increasing number of genetic forms of polyneuropathy are being discovered, the role of routine genetic testing in clinical care is unclear unless patients present with distinct phenotypes suggesting genetic polyneuropathy.

Recommendations for Screening: Patients with T1D for 5 or more years and all patients with T2D should be assessed annually for DSPN using medical history and simple clinical tests.[1] As up to half of patients may be asymptomatic, a diagnosis may only be made on examination or, in some cases, when the patient presents with a painless foot ulcer.[1] Clinicians should be aware that some patients may not volunteer symptoms, but on targeted inquiry may reveal that they are experiencing numbness or other symptoms of DSPN. In addition, healthcare providers should consider screening as above patients with prediabetes/metabolic symptoms and with obesity, as well as children and adolescents.[1,3,14,24,25]

KEY POINTS

- Symptoms and signs of distal symmetric polyneuropathy (DSPN) vary depending on the class of sensory nerve fibers involved and follow a typical symmetrical, distal-to-proximal (stocking and glove) pattern.
- Neuropathic pain (burning, lancinating, tingling, or shooting electric shock–like) may be the first symptom that prompts patients to seek medical care and is due to small-fiber damage.
- Involvement of large fibers may cause numbness and tingling without pain.
- Signs of DSPN include loss of protective sensation to pain, temperature discrimination, touch, and reflexes, eventually leading to loss of balance, falls, fractures, and loss of physical function.
- Up to 50% of diabetic peripheral neuropathies may be asymptomatic.
- Early recognition and appropriate management of neuropathy in all patients with diabetes is essential to prevent serious complications.
- The diagnosis of DSPN is, in the vast majority of cases, a clinical one.
- A combination of at least two evaluations, with at least one targeting small fibers and one targeting large fibers, is recommended to screen and diagnose DSPN in routine clinical care.
- Electrophysiological testing or referral to a neurologist is not needed for diagnosis, except in situations where the clinical features are atypical, the diagnosis is unclear, and a different etiology is suspected.
- Nondiabetic neuropathies mimicking diabetic neuropathy may be present in patients with diabetes and may be treatable by specific measures; thus, a comprehensive differential is recommended.

DISTAL SYMMETRIC POLYNEUROPATHY EVALUATION IN CLINICAL RESEARCH

A valid and careful diagnosis for DSPN in clinical research is critical for correctly identifying the appropriate patient population, either for targeting for a specific intervention or for determining prognostic implications.[1,3] This is particularly relevant given that there are currently no approved disease-modifying therapies for any forms of diabetic neuropathy, and a very large number clinical trials evaluating promising experimental targets for these conditions have failed.[1] Important contributing factors include a lack of agreement and uniformity in the use of the most sensitive DSPN measures that correctly capture the natural history of the disease and may detect timely repair in the specific nerve fiber populations, as well as the inclusion of appropriate patient populations.[1]

As recommended by the Toronto Consensus on Diabetic Neuropathy,[26] for research purposes, a diagnosis of confirmed diabetic neuropathy requires a *combination of symptoms, signs, and abnormality of objective tests* such as changes in nerve conduction studies (NCSs) or, if all NCS parameters are normal, a validated measure of small nerve fibers (Table 44.2). The symptoms and signs of DSPN were broadly covered earlier in the Screening and Diagnosis in Clinical Care section. However, for research, a range of assessments including more objective measures and, importantly, patient-reported outcomes, have been developed over time and validated. The use of *validated clinical instruments* such as the MNSI (most widely used in large cohorts of patients with T1D and T2D),[3,7,10,12,14,27] the modified Toronto Clinical Neuropathy Scale,[28] the Utah Early Neuropathy Scale,[20] or the Neuropathy Disability Score[29] may be most feasible in large cohorts of patients. Table 44.3 includes the main components of and comparisons among the most sensitive and specific instruments. Among these validated clinical instruments, the MNSI has been used most consistently in largest cohorts of patients with T1D, such as DCCT/EDIC, T1D Exchange, and Scottish Registry,[5,7,9] and T2D, including the Action to Control Cardiovascular Risk in Diabetes (ACCORD), BARI 2D, GRADE, and German Diabetes Cohorts,[10,12,22,27] or in youth such as SEARCH for Diabetes in Youth Cohort (SEARCH).[14] However, for reliable results, careful training and standardizations are required.

There are multiple instruments for assessing neuropathic pain in general, with the numerical rating scale and Short-Form McGill Pain Questionnaire being widely used in most neuropathic pain trials. More recently, the DN4 (Douleur Neuropathique en 4 Questions) consisting of interview questions (DN4-interview) and a few physical tests has been validated specifically for painful DSPN, supporting its usefulness as both a screening tool for neuropathic pain and a reliable component of the diagnostic workup for painful DSPN.[30] In addition, the Quality of Life in Neurological Disorders and the Norfolk neuropathy instrument are validated instruments that capture measures of quality of life specific to peripheral neuropathy in several domains, including pain, lost/reduced feeling, diffuse sensory-motor symptoms, restrictions in activities of daily living, disruptions in social relationships, and emotional distress.[31,32]

The NCSs are performed with surface stimulating and recording techniques testing motor and sensory nerve fibers in the upper and lower limbs.[3] In diabetic neuropathy, the first changes in NCS are decrease in sural nerve amplitude, followed by mild reduction in sensory nerve conduction velocity.[3] Motor changes typically follow sensory nerve changes, and lower limb changes precede upper limb changes, as diabetic neuropathy is a length-dependent process. NCSs may be completely normal in those with primarily small fiber neuropathy, and loss of small fibers typically precedes loss of large nerve fibers.[3] The gold standard for small fiber neuropathy is assessment of intraepidermal nerve fiber density (IENFD) measurements by skin punch biopsy,[3] while other measures include quantitative sensory thermal thresholds for reduced cooling detection thresholds or elevated heat thresholds,[6] laser Doppler flare imaging studies,[33] and corneal confocal microscopy to measure nerve fiber length in Bowman's layer of the cornea, which was found to be reduced in diabetic neuropathy.[34,35] The validity of the majority of the small fiber tests, perhaps with the exception of IENFD, is not as well defined as it is for NCS, and for the moment there is no clear role for these tests in the routine clinical diagnosis of small fiber neuropathy, unless patients present with atypical phenotypes.[1,3]

KEY POINTS

- For clinical research purposes, a diagnosis of confirmed distal symmetric polyneuropathy (DSPN) requires a combination of symptoms, signs, and abnormality of objective tests, such as changes in nerve conduction studies and/or a validated measure of small nerve fibers, such as intraepidermal nerve fiber density.
- Several clinical instruments have been validated to assess DSPN in large observational or interventional cohorts.
- The Michigan Neuropathy Screening Instrument has been most widely used in large cohorts of patients with type 1 diabetes or type 2 diabetes and yields the most consistent data.
- Several instruments have also been validated to assess neuropathic pain associated with DSPN and for DSPN-associated quality of life.

Secondary Distal Symmetric Polyneuropathy Complications: Diabetic Foot Ulcers/Charcot Neuroarthropathy

DSPN is the most important cause of foot ulceration, and it is also a prerequisite in the development of Charcot neuroarthropathy, both of which are recognized as late complications of DSPN.[1] These late complications drive amputation risk and economic costs of diabetic neuropathy and are also predictors of mortality. The reader is referred to several other reviews that cover this topic.[36]

The simple yet comprehensive clinical exam is principally designed to identify those at risk of late complications who need education on preventative foot self-care and regular podiatric foot care. Recently an even simpler foot exam, the "3-minute diabetic foot exam" has been proposed.[1,36] This is intended not only for physicians, but also for other healthcare professionals who may only have 15 minutes for the entire diabetes annual review. It requires no equipment and provides simple advice on education on preventative foot self-care.

DSPN is also a major contributor to falls and fractures, through more advanced small and large fiber dysfunction, with loss of sensory, proprioception, temperature discrimination, and pain, all ultimately leading to unsteadiness, recurrent minor injuries, and an increased risk of falls.[1]

TABLE 44.2 **Distal Symmetric Polyneuropathy (DSPN) Diagnostic Criteria for Research**

Probability of DSPN	Signs, Symptoms, and Test Results
Confirmed	Sign(s) or symptom(s) of neuropathy **AND** Abnormal electrophysiology/small fiber (intraepidermal nerve fiber density)
Probable	**≥2 of the following:** Neuropathic symptoms Decreased distal sensation Decreased/absent ankle reflexes
Possible	**Any of the following:** Symptoms: decreased distal sensation, distal neuropathic sensory symptoms **OR** Signs: symmetric decreased distal sensation or decreased/absent ankle reflexes

Adapted from Tesfaye, Boulton, Dyck et al: Toronto Diabetic Neuropathy Expert Group. Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care*. 2010; 33:2285–2293.

TABLE 44.3 **Screening Instruments for Distal Symmetric Polyneuropathy**

	Michigan Neuropathy Screening Instrument (MNSI) Index	Toronto Clinical Neuropathy Score (TCNS)	Neuropathy Disability Score (NDS)
Symptoms	15 items Self-administered	6-item symptom score	None
Exam	Inspection Vibration testing Ankle reflexes	Reflex score (knee, ankle) Sensory test score (pinprick, temperature, light touch, vibration and position sense)	Sensation (vibration, temperature, pinprick) Ankle reflexes
Scoring	≥4 symptoms >2 clinical	6–8: Mild neuropathy 9–11: Moderate neuropathy ≥12: Severe neuropathy	0–2: Mild neuropathy 6–8: Moderate neuropathy 9–10: Severe neuropathy

TABLE 44.4 Number Needed to Treat and Dosing Algorithm for Most Effective Agents for Distal Symmetric Polyneuropathy Pain

Drug	Drug Class	DOSE Initial	DOSE Effective	Number Needed to Treat*
Pregabalin**	Anticonvulsant	25–75 mg one to three times daily	300–600 mg/day	3.3–7.7
Duloxetine**	Antidepressant	20–30 mg/day	60–120 mg/day	3.8–11
Gabapentin	Anticonvulsant	100–300 mg one to three times daily	1800–3600 mg/day	3.3–7.2
Amitriptyline/tricyclics	Antidepressant	10–25 mg daily	25–100 mg/day	2.1–4.2
Capsaicin 8% patch	Topical	Apply on mapped painful areas of the feet, up to a total combined surface area of 1120 cm^2 for both feet for 30 minutes		Not available

*Based on 30%–50% improvement in symptoms.

**US Food and Drug Administration–approved for distal symmetric polyneuropathy treatment.

Adapted from Pop-Busui, Boulton, Feldman et al. Diabetic neuropathy: a position statement by the American Diabetes Association. *Diabetes Care.* 2017;40:136–154.

A clinical diagnosis of small fiber neuropathy is made unless patients present with atypical phenotypes.[1,3]

KEY POINTS

- Distal symmetric polyneuropathy is the most important cause of foot ulcerations, and it is also a prerequisite in the development of Charcot neuroarthropathy.
- If not recognized, and if preventive foot care is not implemented, patients are at risk of injuries to their insensate feet.

Distal Symmetric Polyneuropathy Treatment

Disease-Modifying Therapies. Despite the recent major advances in elucidating the pathogenesis of diabetic neuropathy, there remains a lack of treatment options that effectively target the natural history of DSPN or reverse DSPN once established.[1,3] Several pathogenetic pharmacotherapies have been investigated, but evidence from randomized clinical trials is very limited.[1,3] Advances in DSPN disease modification need to be confirmed with further robust evidence from clinical trials, together with a better understanding of the mechanisms of action of promising treatment.

Prevention. Prevention of diabetic neuropathies focuses on glucose control and lifestyle modifications. Enhanced glucose control in people with T1D dramatically reduces the incidence of DSPN (78% relative risk reduction).[1,4,7] Therefore, tight glucose control targeting near-normal glycemia in patients with T1D is recommended for DSPN prevention in T1D.[1]

In contrast, enhanced glucose control in people with T2D modestly reduces the risk of developing DSPN (5%9% relative risk reduction).[1] In a small trial of Japanese patients with early T2D, intensive insulin treatment was associated with improvement in selected DSPN measures,[1] and in the ACCORD trial individuals with T2D reported a modest but significant DSPN risk reduction with the glycemia intervention after 5 years of follow-up,[27] but other trials reported no effects.[1] This discrepancy highlights the differences between T1D and T2D and emphasizes the point that many people with T2D develop DSPN despite adequate glucose control.[1] Specific glucose-lowering strategies may contribute to the discrepancy. For example, participants, particularly men, in the BARI 2D trial treated with insulin sensitizers had a lower incidence of DSPN over 4 years than those treated with insulin/sulfonylurea.[12]

There is also emerging evidence that lifestyle modifications—either exercise alone (supervised aerobics ± resistance training) or combined dietary modification and exercise—may have beneficial effects on preventing DSPN in some patients with T2D or prediabetes.[1,3]

Management of Pain in Distal Symmetric Polyneuropathy. Currently, there is no convincing evidence supporting glucose control or lifestyle management as therapies for neuropathic pain in diabetes or prediabetes.[1]

At present, pregabalin and duloxetine have received regulatory approval for the treatment of neuropathic pain associated with DSPN by the US Food and Drug Administration (FDA), Health Canada, and the European Medicines Agency.[1] Recent evidence from two large, 12-week, randomized, multicenter trials led to the FDA approval of a cutaneous concentrated capsaicin 8% patch that works by desensitizing and interfering with the function of the transient receptor potential vanilloid 1 receptor, a protein involved in pain signaling. Pain reduction was approximately 30% in these trials, but the patch needs to be applied for approximately 30 minutes. The opioid tapentadol has regulatory approval in the United States and Canada, but the evidence of its use is weak, and the ADA and other societies strongly recommend against using any opioids for management of DSPN pain.[1] The ADA Statement on Diabetic Neuropathy evaluated large datasets from published and unpublished studies to decrease the chance of publication bias and concluded that several other agents of different classes, including gabapentin and some of the monoamine reuptake inhibitors, may also be effective in reducing DSPN pain, alone or in combination, and thus recommended a stepwise approach taking into account patients' socioeconomic status, comorbidities, and potential drug interactions.[1] Table 44.4 summarizes the evidence on number needed to treat as well as dosing considerations and adverse effects for most effective agents for DSPN pain, while Fig. 44.3 shows a proposed stepwise approach. Given the high risk of addiction, abuse, sedation, and other complications and psychosocial issues even with short-term opioid use, opioids are not recommended in the treatment of painful DSPN.[1]

KEY POINTS

- There remains a lack of treatment options that effectively reverse distal symmetric polyneuropathy (DSPN), once established.
- Tight glucose control targeting near-normal glycemia in patients with type 1 diabetes (T1D) is recommended for DSPN prevention in T1D.
- Emerging evidence supports that lifestyle modifications—either exercise alone (supervised aerobics ± resistance training) or combined dietary modification and exercise—may have beneficial effects on preventing DSPN in some patients with type 2 diabetes or prediabetes.
- Several treatment options exist for pain associated with DSPN.
- Pregabalin, duloxetine, and the capsaicin 8% patch have received regulatory approval for the treatment of neuropathic pain associated with DSPN by the US Food and Drug Administration and the European Medicines Agency.
- Given the high risk of addiction, abuse, sedation, and other complications and psychosocial issues even with short-term opioid use, opioids are not recommended in the treatment of painful DSPN.

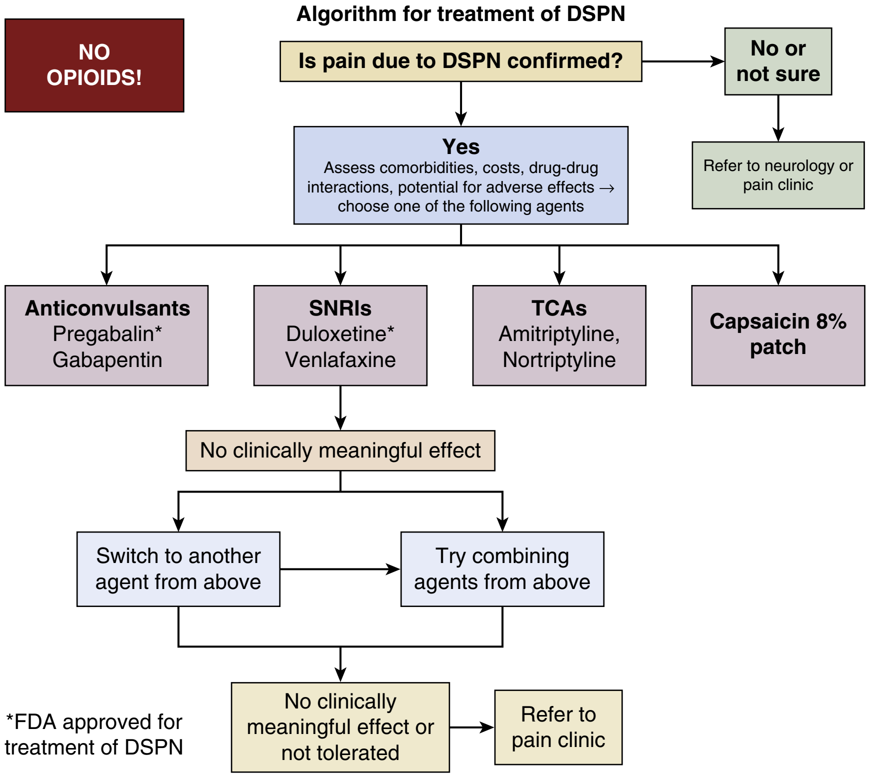

Algorithm for treatment of DSPN
NO OPIOIDS!
Is pain due to DSPN confirmed?
No or not sure
Refer to neurology or pain clinic
Yes
Assess comorbidities, costs, drug-drug interactions, potential for adverse effects → choose one of the following agents
Anticonvulsants
Pregabalin*
Gabapentin
SNRIs
Duloxetine*
Venlafaxine
TCAs
Amitriptyline,
Nortriptyline
Capsaicin 8% patch
No clinically meaningful effect
Switch to another agent from above
Try combining agents from above
*FDA approved for treatment of DSPN
No clinically meaningful effect or not tolerated
Refer to pain clinic

Figure 44.3 Algorithm for painful DSPN. (Adapted from Pop-Busui R, Boulton AJ, Feldman EL, et al. Diabetic neuropathy: a position statement by the American Diabetes Association. Diabetes Care. 2017;40:136–154.)

DIABETIC AUTONOMIC NEUROPATHIES

Autonomic neuropathies affect the autonomic neurons (parasympathetic, sympathetic, or both) and are associated with a variety of condition-specific symptoms and signs that should be evaluated during the medical history and physical examination of all individuals with diabetes[1]. While CAN is the most studied and clinically relevant of the diabetic autonomic neuropathies, gastrointestinal, genitourinary, and sudomotor dysfunction may also develop during diabetes course, with their direct impact in patient's quality of life and optimal management.

Cardiovascular Autonomic Neuropathy

CAN is a common and deadly complication of diabetes mellitus, which is frequently overlooked in clinical practice due to its characteristic subtle presentation earlier in disease. Yet, timely detection of CAN may help implementation of tailored interventions to prevent its progression and mitigate the risk of associated complications, including CVD, cardiac arrhythmias, myocardial dysfunction leading to congestive heart failure, and all-cause mortality.[37]

Epidemiology. There are variations in the reported prevalence of CAN depending on the population(s) studied (e.g., T1D vs. T2D, intervention trials vs. epidemiological observations, older vs. contemporary cohorts), the available outcome measures reported, and/or the diagnostic criteria used. In earlier cohorts of patients, particularly those with longstanding T1D, potential candidates for pancreas transplantation, abnormalities in cardiovascular reflex tests (CARTs) suggestive of CAN were detected in up to 90% of patients.[37] In contrast, prevalence rates are very low in patients with newly diagnosed T1D, as documented by the primary prevention arm of the DCCT cohort, arguably the best phenotyped cohort for CAN with CARTs obtained repeatedly over time.[1,4,37,38] During the observational epidemiological follow-up of the same cohort, known as the EDIC study, the prevalence of CAN increased steadily over time to as high as 44% of the cohort over 23 years of mean follow-up.[4] These data are comparable with data reported in other observational cohorts with T1D such as the EURODIAB IDDM study, which reported baseline prevalence rates of autonomic dysfunction, defined as an abnormality in one or both of the CARTs used, of 36%,[39] or the Pittsburgh Epidemiology of Diabetes Complications Study (EDC), which reported a cumulative incidence of symptomatic autonomic neuropathy over 25 years of approximately 25%.[40] Similar to T1D, prevalence rates as high as 60% have been reported in earlier cohorts of patients with longstanding T2D.[1,37] In more recent T2D cohorts, however, lower prevalence rates of 9% to 25% have been observed. For instance, in Europe, the Diabetic Cardiovascular Autonomic Neuropathy Multicenter Study Group reported CAN prevalence rates of approximately 22% in T2D using strict criteria of abnormal findings in CARTs and heart rate variability (HRV), while the observational prospective follow-up of the Danish arm of the Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care (ADDITION-Denmark) reported prevalence rates of 9% and 15% for manifest CAN, defined based on CARTs, at the 6-year and 13-year follow-up examinations, respectively.[41] In the United States, in more than 8000 T2D patients with a mean age of approximately 62 years and mean duration of 10 years of participants with T2D in the ACCORD trial, the prevalence of CAN at study entry was approximately 20%,[42] while the GRADE trial reported that approximately 10% of this contemporary cohort with a mean known T2D duration of 4 years presented with CAN at baseline as diagnosed with the same HRV studies.[10] In addition, CAN is present in patients with impaired glucose tolerance, insulin resistance, and/or the metabolic syndrome,[1] as well as in youth, as documented by the SEARCH for Diabetes in Youth Cohort study that included approximately 2000 youth with T1D and T2D and reported

CAN prevalence rates of approximately 12% in youth with T1D (mean age 18 years, duration of diabetes 8 years) and 17% in youth with T2D (mean age 22 years, duration of diabetes 8 years).[14]

Similarly with DSPN, notable risk factors for CAN include older age and markers of higher glycemic exposure (including longer diabetes duration and poor glucose control), diabetic kidney disease, hypertension, elevated triglycerides, and use of some medications.[1,4,37] More recently, socioeconomic factors including lower income and education, as well as depression, are emerging as risk factors for CAN.

KEY POINTS

- Cardiovascular autonomic neuropathy (CAN) is a prevalent complication in type 1 diabetes and type 2 diabetes (T2D), including in newly diagnosed T2D.
- High prevalence rates were described in individuals with prediabetes/metabolic syndrome, as well as in youth.
- Notable risk factors for CAN include older age, higher glycemic exposure, diabetic kidney disease, hypertension, elevated triglycerides, and smoking.
- Socioeconomic factors, female gender, and depression are emerging as risk factors for CAN.

Clinical Impact. A timely diagnosis of CAN has important clinical implications, as CAN is an independent risk factor for cardiovascular mortality, arrhythmia, silent ischemia, or any major cardiovascular event. Additional evidence demonstrates that CAN is an important risk factor for the development of myocardial dysfunction and heart failure in both T1D and T2D, which are emerging as underdiagnosed and important health problems in diabetes, and a predictor of the development of diabetic kidney disease.[1,37]

Screening and Diagnosis in Clinical Care. It could be quite challenging to detect CAN in its early stages in routine clinical care, as patients may be completely asymptomatic (subclinical CAN) and may only present with decreased HRV.[1,37]

Symptoms of CAN: The most common symptoms of CAN occur upon standing and include lightheadedness, weakness, palpitations, faintness, and syncope[1,37] (Table 44.5). Thus, targeted questions to unveil above symptoms should be included while a medical history is taken in the office, although the specificity of these symptoms for CAN is quite low, requiring an appropriate differential[1,37] (Table 44.5). In addition, these symptoms occur late in the disease course.[1,37] Providers may also consider screening patients at risk for developing CAN for hypoglycemia unawareness and vice versa, as this may be associated with CAN.[1,43]

Clinical Signs of Cardiovascular Autonomic Neuropathy. Physiologic HRVs re a product of both sympathetic and parasympathetic activity.[1,37] Patients with subclinical CAN present with a decrease in HRV, usually with deep breathing or change in posture, considered the earliest clinical indicator of CAN.[1,37] Although HRV testing is largely confined to the research setting, it may be done in the office by either 1) taking an electrocardiogram (ECG) recording as a patient begins to rise from a seated position or 2) taking an ECG recording during 1 to 2 minutes of deep breathing with calculation of HRV.[1] However, even these relatively simple methods could be challenging in some clinical settings, highlighting the need for easily accessible diagnostic tools and/or biomarkers for screening and diagnosis of CAN for general clinical care.

As CAN progresses, patients may present with resting tachycardia with fixed heart rate (>100 bpm; although one should be aware that resting heart rate is not specific), exercise intolerance, nondipping and reverse dipping, and in most advanced cases orthostatic hypotension (a fall in systolic or diastolic BP by >20 mm Hg or >10 mm Hg, respectively, upon standing without an appropriate increase in heart rate).[1,44] Orthostatic hypotension is usually easy to document in the office by measuring the BP supine and after standing, keeping also in mind that, in most cases of CAN, there is no compensatory increase in the heart rate, despite hypotension.[1] To document nondipping or reverse dipping, ambulatory BP monitoring, a standard tool in the current hypertension management, allows the assessment of the diurnal BP patterns that is mainly regulated by sleep-awake changes in the autonomic cardiovascular function.[1,44]

CAN diagnosis includes documentation of symptoms and signs of CAN (Table 44.5). In a symptomatic patient with a history of poor glucose control presenting with resting tachycardia or postural hypotension, clinicians may not need to perform additional CAN tests, given costs and burden, after excluding other potential causes.

As with DSPN, diagnosis of CAN requires an extensive differential, as many other comorbidities, such as CVD, adrenal insufficiency, thyroid dysfunction, acute blood loss and anemia, dehydration, and drug effects/interactions may present with symptoms and/or signs mimicking CAN, as illustrated in Table 44.5.[1,37]

Recommendations for screening in clinical care:

- All patients should be assessed for CAN starting 5 years after diagnosis or in the presence of other forms of diabetic neuropathy and/or other diabetic complications with symptoms and signs.
- Tests excluding other comorbidities or drug effects/interactions that could mimic CAN should be performed.
- Screening for CAN should also be considered in patients with hypoglycemia unawareness and high glucose variability, prior to insulin dose adjustments and/or perioperatively.[1]

KEY POINTS

- In the earliest stages, cardiovascular autonomic neuropathy (CAN) may be completely asymptomatic (subclinical CAN) and may only present with decreased heart rate variability.
- Symptoms of CAN, including palpitations, lightheadedness, weakness, palpitations, faintness, and syncope with standing, may occur only in late stages of disease.
- Signs of CAN include: resting tachycardia with fixed heart rate (>100 bpm), exercise intolerance, nondipping and reverse dipping, and orthostatic hypotension (a fall in systolic or diastolic blood pressure by >20 m mHg or >10 mm Hg, respectively).
- Cardiovascular reflex tests remain the gold standard diagnostic test.
- Diagnosis of CAN requires an extensive differential, as symptoms and signs are nonspecific (Table 44.5).
- All individuals with diabetes should be assessed for CAN starting 5 years after diagnosis or in the presence of other forms of diabetic neuropathy and/or other diabetic complications with symptoms and signs.

Cardiovascular Autonomic Neuropathy Evaluation in Clinical Research

Several tests are available to diagnose CAN, including: CARTs, ideally under paced breathing; HRV studies that derive indices in the time and frequency domains using various algorithms; 24-hour BP profiles; baroreflex sensitivity testing; and cardiac sympathetic imaging.[1,37]

CARTs remain the gold standard for autonomic testing in both clinical and research settings[1,37,44] and evaluate changes in heart rate and BP during several physiological maneuvers such as: 1) deep

TABLE 44.5 Stepwise Approach for Cardiovascular Autonomic Neuropathy Diagnosis

	Symptoms	Signs/ Diagnostic Tests	Differential Workup

Resting tachycardia	Palpitations
	Could be asymptomatic
		Clinical exam:
		Resting HR >100 bpm
			• Anemia
			• Hypothyroidism
			• Fever
			• CVD (atrial fibrillation, flutter, other)
			• Dehydration
			• Adrenal insufficiency
			• Medications
			  • Sympathomimetic agents (asthma)
			  • Over-the-counter cold agents containing ephedrine or pseudoephedrine
			  • Dietary supplements (e.g., ephedra alkaloids)
			• Smoking, alcohol, caffeine
			• Recreational drugs (cocaine, amphetamines,methamphetamine, mephedrone)

Orthostatic hypotension	Light headedness
	Weakness
	Faintness
	Visual impairment
	Syncope
		Clinical Exam:
		A reduction of >20 mm Hg in systolic BP or >10 mm Hg in diastolic BP
			• Adrenal insufficiency
			• Intravascular volume depletion
			  • Blood loss/acute anemia
			  • Dehydration
			• Pregnancy/postpartum
			• CVD
			  • Arrhythmias
			  • Heart failure
			  • Myocarditis
			  • Pericarditis
			  • Valvular heart disease
			• Alcohol
			• Medication
			  • Antiadrenergics
			  • Antianginals
			  • Antiarrhythmics
			  • Anticholinergics
			  • Diuretics
			• ACEi/ARBs
			• Narcotics
			• Neuroleptics
			• Sedatives

HR, Heart rate; *CVD*, cardiovascular disease; *BP*, blood pressure; *ACEi*, angiotensin-converting enzyme inhibitor; *ARB*, angiotensin receptor blocker.

breathing; 2) standing; and 3) a Valsalva maneuver.[1,37] Several factors may influence the results of CARTs, such as intake of caffeine, tobacco products, food, and some medications like beta-blockers. Therefore, these tests should be performed by uniformly trained personalized personnel and using standardized protocols to minimize variability and bias. In addition, given the very large number of recording devices and particularly software operating in "black-box" mode that have emerged in recent years, a careful selection of the most transparent and verifiable technologies to be used is recommended for reliable and reproducible data. In general, it is currently accepted that one abnormal test may indicate early or subclinical CAN, while two or more abnormal tests are recommended for a definite diagnosis of CAN.[44] Given that age is one of the most important factors affecting heart rate, age-related normative are needed for correct classification settings.[1,37,44]

Indices of HRV are additional CAN outcome measures that have been used extensively. The HRV can be evaluated either as time domain measures such as: the difference between the longest and shortest R–R interval, the standard deviation of 5-min average of normal R–R intervals, and the root-mean square of the difference of successive R–R intervals, or as frequency domain measures obtained by spectral analysis or other mathematical algorithms, such as: low, very low, and high frequency power.[1,37,44] Although traditionally these indices were derived from longer ECG recordings, more recently HRV-derived indices from standard 10-second 12-lead ECG recordings have emerged as more feasible alternative measures of CAN in population studies and have been shown to independently predict CVD morbidity and mortality in several large cohorts of people with and without diabetes[10,45-47] and to define prevalence rates.

Cardiac sympathetic imaging using iodine-123-labeled metaiodobenzylguanidine scintigraphy or ^{11}C-meta-hydroxyephedrine positron emission tomography, 24-hour ambulatory blood pressure monitoring profiles, baroreflex sensitivity testing to assess cardiac vagal and sympathetic baroreflex function, and microneurography, while valid and useful in some specific research scenarios, require sophisticated infrastructure and highly trained personnel and are quite expensive and time-consuming, but are widely used in research protocols.[1,37]

Several clinical instruments such as the Autonomic Symptoms Profile, the Composite Autonomic Symptom Score, and more recently the Survey of Autonomic Symptoms, which have various degrees of complexity, have also been validated to assess autonomic neuropathy in general and indirectly CAN in large cohorts.

KEY POINTS

- Cardiovascular reflex tests evaluating changes in heart rate and blood pressure during several physiological maneuvers remain the gold standard for clinical research.
- Heart rate variability–derived indices derived from standard 10-second 12-lead electrocardiogram recordings have emerged as more feasible alternative measures of cardiovascular autonomic neuropathy (CAN) in large cohorts and population studies.
- Several validated clinical instruments that offer indirect information on CAN may also be used in large cohorts.

Treatment

As it is the case with DSPN, to date no disease-modifying agents are available for CAN, likely due to the very complex pathogenic pathways contributing to CAN development.[1,37] Thus, the management of CAN includes consideration of glucose control and modification of other risk factors.

In patients with T1D, the conclusive data obtained by the DCCT, and later the EDIC study, as well as evidence from several smaller trials, strongly support that tight control of blood glucose implemented as early as possible in the disease course is very effective for preventing CAN and slowing its progression in patients with T1D.[1,4,7,37,38,48] For instance, during over an average of 6.5 years' DCCT follow-up, intensive glucose control, with lower hemoglobin A1c levels, reduced the risk of CAN by 45%, a benefit that persisted during EDIC, with a 30% reduction in incident CAN over an additional 14 years of follow-up.[4,7,38] The evidence for patients with T2D is less conclusive, as in several large cohorts intensive control of blood glucose had no effect on the risk of CAN,[49-51] although one could argue that these trials may have not included the most reliable CAN outcomes, and most enrolled patients with T2D had multiple associated comorbidities and other risk factors, such as hypertension, hyperlipidemia, and obesity, and required polypharmacy. In contrast, glucose control as part of a multifactorial intervention that also targeted hyperglycemia, hypertension, dyslipidemia, and lifestyle demonstrated a 63% reduction in the rate of progression to CAN in a small T2D cohort participating in the STENO-2 trial.[52] Most recently, analyses from the ACCORD trial reported that, after adjusting for multiple other risk factors, intensive glucose treatment significantly reduced CAN risk by 16% compared with standard intervention in a large cohort of more than 8000 T2D participants in the ACCORD trial over a mean of 5 years' follow-up.[42] These data confirmed a beneficial effect of intensive glycemic therapy, particularly in individuals without CVD at baseline. A similar benefit on CAN was found for intensive BP control,[42] while intensive treatment of hyprelipidemia with statins and fenofibrate did not. The population-based Coronary Artery Risk Development in Young Adults study reported benefit of lifestyle interventions in prevention of CAN progression.[53] In the Diabetes Prevention Program, lifestyle interventions improved measures of HRV over time in participants with prediabetes.[54]

Management of orthostatic hypotension involves both behavioral and pharmacological interventions. Behavioral supportive measures include: a) avoiding abrupt changes in body position; b) avoiding actions that elevate intraabdominal and intrathoracic pressures; c) avoiding medications that would exacerbate hypotension, such as tricyclic antidepressants, phenothiazines, and diuretics; d) raising the head of the bed during sleep; e) following a schedule of small and frequent meals to minimize postprandial hypotension; f) considering physical counter-pressure maneuvers such as leg crossing and squatting; g) hydrating with fluids and salt, if not contraindicated.[1,37] Pharmacological therapy includes midodrine and the more recent droxidopa, both of which are FDA-approved for the management of orthostatic hypotension and may be considered in patients who fail nonpharmacological interventions. However, it is recommended to proceed with a very slow titration and use the minimally effective dose to avoid undesirable side effects. In selected severe cases low-dose fludrocortisone may also be an option.[1,37]

KEY POINTS

- Similar to distal symmetric polyneuropathy, there are no available disease-modifying therapies for cardiovascular autonomic neuropathy (CAN).
- Tight glucose control is recommended to prevent CAN in type 1 diabetes.
- Glucose and other risk-factor control such as hypertension and lifestyle modifications have been shown to reduce CAN risk in type 2 diabetes.
- Management of orthostatic hypotension involves both behavioral and pharmacological interventions.
- Midodrine and droxidopa are US Food and Drug Administration–approved for the management of orthostatic hypotension, but very slow titration and using the minimally effective dose to avoid undesirable side effects are recommended.

Gastrointestinal Autonomic Neuropathy

Gastrointestinal neuropathies may involve any portion of the gastrointestinal tract, with manifestations including esophageal dysmotility, gastroparesis (delayed gastric emptying), constipation, diarrhea, and fecal incontinence.[1] Among these, gastroparesis may the most common condition providers may consider in clinical practice.

Epidemiology. Earlier prevalence data on gastroparesis are limited, given that most initial reports were from selected case series rather than larger populations, and there was inconsistency in the outcome measures use. In a large community-based study in Olmsted County, Minnesota, the cumulative incidence of gastroparesis over 10 years was higher in T1D patients (5%) compared with T2D patients (1%) and controls (1%).[55] The age–adjusted incidence of gastroparesis per 100,000 person–years was higher for women compared with men (9.8 for women and 2.4 for men), and the overall survival was found to be significantly lower than the age– and sex–specific survival of general population.[1] The prevalence of gastrointestinal symptoms that could mimic gastroparesis increased substantially in recent years in the United States and other countries, although these are nonspecific.[55,56] Additionally, the number of gastroparesis-related hospitalizations has been increasing in the United States, suggesting an increasing prevalence of gastroparesis.

KEY POINTS

- The contemporary prevalence rates of confirmed gastroparesis due to type 1 or type 2 diabetes are low.
- The prevalence of gastrointestinal symptoms that could mimic gastroparesis has increased substantially in recent years, although these are nonspecific and mainly due to iatrogenic causes.

Screening and Diagnosis in Clinical Care. Gastroparesis, one of the most common diabetic complications that providers may consider in clinical practice, can manifest with a broad spectrum of symptoms and signs.[1,57] As part of a medical history, providers are encouraged to document symptoms of gastroparesis, such as early satiety, fullness, bloating, nausea, vomiting, dyspepsia, and abdominal pain. However, all these symptoms are nonspecific and resemble many other conditions, do not correspond with severity of gastroparesis, and are poorly associated with abnormal gastric emptying.[1] Thus, gastroparesis may be clinically silent in the majority of cases and symptoms,[1] and many patients may only present with unexplained early postprandial hypoglycemia followed by later hyperglycemia.

Importantly, besides organic causes such as gastric outlet obstruction or peptic ulcer disease, hyperglycemia, hypoglycemia, and acute changes in blood glucose[58,59] are well documented to alter gastric emptying.[1] A critical consideration is that several medications widely used in patients with diabetes, especially opioids, which have been unfortunately prescribed extensively for pain management, and glucagon-like peptide-1 (GLP-1) receptor agonists, directly affect gastric emptying and could mimic gastroparesis or cause iatrogenic gastroparesis.[1]

Therefore, a thorough differential to exclude all these factors known to affect gastric emptying (including esophagogastroduodenoscopy and/or a barium study of the stomach, careful documentation of medication intake, and performing continuous glucose monitoring if available) should always be considered before considering more specialized testing is performed for gastroparesis and a firm diagnosis is established.

The diagnostic gold standard is the measurement of gastric emptying, with scintigraphy of digestible solids at 15-minute intervals for 4 hours after food intake. Optimization of glucose levels prior to scanning is needed[58,59] to avoid false-positive results. However, this test is burdensome, time-consuming, and costly, and is not readily available. Recently, the use of 13C octanoic acid breath test has been FDA-approved and emerged as a much easier to use alternative.[1]

Evaluation in Clinical Research: gastric emptying with scintigraphy of liquids and digestible solids, ideally under stable, clamped glucose levels, remains the gold standard, but is not feasible in large cohorts of participants, for which the breath tests may be more feasible. In addition, collection of full-thickness gastric tissue for detecting specific cellular changes associated with diabetic gastroparesis was developed by the National Institute of Diabetes and Digestive and Kidney Diseases Gastroparesis Clinical Research Consortium, although these studies are invasive and require specialized infrastructure, and thus may only be applicable in selected situations.

KEY POINTS

- Symptoms such as early satiety, fullness, bloating, nausea, vomiting, dyspepsia, and abdominal pain are not specific for gastroparesis.
- Gastroparesis may be clinically silent in the majority of cases, and most patients may only present with unexplained early postprandial hypoglycemia followed by later hyperglycemia.
- Several medications widely used in patients with diabetes, especially opioids, which have been unfortunately prescribed extensively for pain management, and glucagon-like peptide-1 receptor agonists, directly affect gastric emptying and could mimic gastroparesis or cause iatrogenic gastroparesis.
- Gastric emptying with scintigraphy of digestible solids remains the gold standard diagnostic test, but is burdensome, time-consuming, and costly, and is not readily available..
- Recently, the use of 13C octanoic acid breath test has been US Food and Drug Administration–approved and emerged as a much easier to use alternative.

Treatment. Treatment for diabetic gastroparesis may be very challenging and includes nonpharmacological and pharmacological options. Dietary changes, including eating multiple small meals and decreasing dietary fat and fiber intake, may be useful. In addition, improving glucose control and blunting large glucose fluctuations by optimizing the relationship between food intake and type of and pharmacokinetic profiles of glucose-lowering medications, including the use of sensor-augmented insulin pumps, will improve gastric emptying. Importantly, withdrawing drugs with effects on gastrointestinal motility, particularly opioids, but also some anticholinergics and tricyclic antidepressants, and providing appropriate education on the expected symptoms and dietary changes needed when initiating GLP-1 receptor agonists, pramlintide, and possibly dipeptidyl peptidase-4 inhibitors, will prevent unnecessary evaluations.[1] In cases of severe confirmed diabetic gastroparesis, pharmacological interventions are needed. To date, only metoclopramide, a prokinetic agent, is approved by the FDA for the treatment of gastroparesis, although with a weak level of evidence for benefits in the management of gastroparesis. Given the risk for serious adverse effects (extrapyramidal symptoms, such as acute dystonic reactions; drug-induced parkinsonism; akathisia; and tardive dyskinesia), its use in the treatment of gastroparesis beyond 5 days is no longer recommended by the FDA and the European Medicines Agency.[1] It should be reserved for severe cases that are unresponsive to other therapies.

KEY POINTS

- Dietary changes, including eating multiple small meals and decreasing dietary fat and fiber intake, may be useful.
- Improving glucose control and blunting large glucose fluctuations will improve gastric emptying.
- It is recommended to withdraw drugs with effects on gastrointestinal motility, particularly opioids, and to provide appropriate education on the expected symptoms and dietary changes needed when initiating glucagon-like peptide-1 receptor agonists, pramlintide, and possibly dipeptidyl peptidase-4 inhibitors to prevent unnecessary evaluations.
- Metoclopramide is approved by the US Food and Drug Administration (FDA) for the treatment of gastroparesis, but given the risk for serious adverse effects its use in the treatment of gastroparesis beyond 5 days, it is no longer recommended by the FDA.

Urogenital Autonomic Neuropathy

The genitourinary disturbances associated with diabetic autonomic neuropathy include sexual dysfunction in both men and women, as well as bladder dysfunction.

Epidemiology. Erectile dysfunction (ED) is three times more common in men with diabetes than those without the disease.[1,60] Sexual dysfunction is also more common in women with diabetes.[1,61] Emerging evidence also demonstrate associations between ED and other diabetes complications, including CAN and DSPN.[1,62,63] Glucose control was associated with a lower incidence of ED in men with T1D.[1] Evidence is less strong for T2D. Control of other risk factors such as hypertension and hyperlipidemia may also improve the condition.[1]

Screening in Clinical Care. Men may present ED and/or retrograde ejaculation, given that autonomic neurotransmission controls the cavernosal and detrusor smooth muscle tone and function.[1] The etiology, however, is multifactorial, and clinicians should also evaluate other vascular risk factors such as hypertension, hyperlipidemia, obesity, endothelial dysfunction, smoking, CVD, concomitant medication, and psychogenic factors.[1] A diagnosis should be made after establishing the signs and symptoms of ED and after excluding alternate causes. Clinicians should consider performing hormonal evaluation (luteinizing hormone, testosterone, free testosterone, prolactin) to rule out hypogonadism. In addition, a variety of medications (such as beta-blockers) and organic causes should be excluded.[1]

It is recommended that clinicians screen men with other forms of diabetic neuropathy annually for ED with simple questions about a patient's libido and ability to reach and maintain an erection.[1]

Female sexual dysfunction (FSD) presents as decreased sexual desire, increased pain during intercourse, decreased sexual arousal, and inadequate lubrication.

Bladder dysfunction may manifest as nocturia, frequent urination, urination urgency, and weak urinary stream, and may also coexist with lower urinary tract symptoms (LUTS) such as urinary incontinence, and recurrent urinary tract infections.[1] Thus, it is also recommended that clinicians screen patients with other forms of diabetic neuropathy for FSD and LUTS. Evaluation of bladder function should be performed for individuals with diabetes who have recurrent urinary tract infections, pyelonephritis, urinary incontinence, nocturia, or a palpable bladder. In addition, FSD could be documented using targeted questioning regarding symptoms, such as nocturia or pain during intercourse.[1]

Treatment. Pharmacological treatment for ED includes phosphodiesterase type 5 inhibitors as first-line therapy, and in more advanced cases, transurethral prostaglandins, intracorneal injections, vacuum devices, and penile prosthesis.[1] FSD treatment is much more complex and challenging.

KEY POINTS

- Erectile dysfunction (ED), female sexual dysfunction (FSD), and bladder dysfunction are common in men and women with diabetes and have direct links with autonomic neuropathy.
- Clinicians should consider screening men who present with other forms of diabetic neuropathy annually for ED with simple questions about a patient's libido and ability to reach and maintain an erection.
- Bladder dysfunction may manifest as nocturia, frequent urination, urination urgency, and weak urinary stream.
- FSD could be documented using targeted questioning regarding symptoms, such as nocturia or pain during intercourse.
- Phosphodiesterase type 5 inhibitors remain the first-line therapy for ED.
- FSD treatment is much more complex and challenging.

Sudomotor Dysfunction

Sudomotor dysfunction may manifest as dry skin, anhidrosis, heat intolerance, or in rare cases gustatory sweating.[1] The latter is a very rare form of sudomotor dysfunction comprising excessive sweating limited exclusively to the head and neck region usually triggered by food consumption, and is described in patients with more advanced autonomic neuropathy and/or on dialysis.[64]

Based on the available evidence, routine screening for sudomotor dysfunction in clinical practice is not recommended at this time.[1]

ATYPICAL NEUROPATHIES

Mononeuropathies

Mononeuropathies occur as a result of involvement of the median, ulnar, radial, and common peroneal nerves and are more common in patients with diabetes.[1] Cranial neuropathies are very rare, primarily involve cranial nerves III, IV, VI, and VII, and usually resolve spontaneously over several months.[1] Electrophysiological studies are most helpful in identifying nerve conduction slowing or conduction block at the site of nerve entrapment. Nerve entrapments may require surgical decompression. The improvement in symptom severity and functional status score is no different between patients with and without diabetes.[1]

Diabetic Radiculoplexus Neuropathy

Diabetic radiculoplexus neuropathy (diabetic amyotrophy or diabetic polyradiculoneuropathy) typically involves the lumbosacral plexus.[1] People with the condition routinely present with extreme unilateral thigh pain and weight loss, followed by motor weakness. Electrophysiological assessment is required to document the extent of disease and alternative etiologies, including degenerative disc disease, neoplastic, infectious, and inflammatory spinal disease.[1] The disorder is usually self-limiting, patients improve over time with medical management and physical therapy, and there is presently no evidence from randomized trials to support any recommendation on the use of any immunotherapy treatment in this condition.[1]

CONCLUSIONS

Diabetic neuropathies are prevalent and have a diverse clinical spectrum of symptoms and signs. By far the most common forms are DSPN and CAN, and pain may be an initial symptom that triggers medical attention. Given its high morbidity and mortality risk, an earlier diagnosis is recommended. In most cases, accurate recording of symptoms, as described by the patient, a careful clinical examination, use of simple instruments (e.g., the MNSI), and foot inspection are sufficient for daily clinical practice. More sophisticated diagnostic tests or procedures, including referral to neurologists or electrophysiology, are rarely needed and are reserved for atypical and rapidly progressing cases. Treatment of diabetic neuropathies is challenging, often requiring a team approach. Glucose control (especially for patients with T1D), effective treatment of pain, and measures to prevent foot ulcerations remain the mainstay of treatment.

REFERENCES

1. Pop-Busui R, Boulton AJ, Feldman EL, et al. Diabetic neuropathy: a position statement by the American Diabetes Association. *Diabetes Care*. 2017;40:136–154.
2. Pop-Busui R, Boulton A, Sosenko JM. Peripheral and autonomic neuropathy in diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, et al., eds. *Diabetes in America*. 3rd ed. Vol. 23. Bethesda: National Institutes of Health; 2018:21–20.
3. Feldman EL, Callaghan BC, Pop-Busui R, et al. Diabetic neuropathy. *Nat Rev Dis Prim*. 2019;5:41.
4. Braffett BH, Gubitosi-Klug RA, Albers JW, et al. Risk factors for diabetic peripheral neuropathy and cardiovascular autonomic neuropathy in the diabetes control and complications trial/epidemiology of diabetes interventions and complications (DCCT/EDIC) study. *Diabetes*. 2020;69:1000–1010.
5. Mizokami-Stout KR, Li Z, Foster NC, et al. The contemporary prevalence of diabetic neuropathy in type 1 diabetes: findings from the T1D Exchange. *Diabetes Care*. 2020;43:806–812.
6. Dyck PJ, Albers JW, Andersen H, et al. Diabetic polyneuropathies: update on research definition, diagnostic criteria and estimation of severity. *Diabetes Metab Res Rev*. 2011;27:620–628.
7. Martin CL, Albers JW, Pop-Busui R. Neuropathy and related findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabetes Care*. 2014;37:31–38.
8. Tesfaye S, Stevens LK, Stephenson JM, et al. Prevalence of diabetic peripheral neuropathy and its relation to glycaemic control and potential risk factors: the EURODIAB IDDM Complications Study. *Diabetologia*. 1996;39:1377–1384.

9. Jeyam A, McGurnaghan SJ, Blackbourn LAK, et al. Diabetic neuropathy is a substantial burden in people with type 1 diabetes and is strongly associated with socioeconomic disadvantage: a population-representative study from Scotland. *Diabetes Care*. 2020;43:734–742.
10. Mather KJ, Bebu I, Baker C, et al. Prevalence of microvascular and macrovascular disease in the Glycemia Reduction Approaches in Diabetes - a Comparative Effectiveness (GRADE) study cohort. *Diabetes Res Clin Pract*. 2020;165:108235.
11. Andersen ST, Witte DR, Andersen H, et al. Risk-factor trajectories preceding diabetic polyneuropathy: ADDITION-Denmark. *Diabetes Care*. 2018;41:1955–1962.
12. Pop-Busui R, Lu J, Brooks MM, et al. Impact of glycemic control strategies on the progression of diabetic peripheral neuropathy in the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) Cohort. *Diabetes Care*. 2013;36:3208–3215.
13. Jaiswal M, Fufaa GD, Martin CL, et al. Burden of diabetic peripheral neuropathy in Pima Indians with type 2 diabetes. *Diabetes Care*. 2016;39:e63–e64.
14. Dabelea D, Stafford JM, Mayer-Davis EJ, et al. Association of type 1 diabetes vs type 2 diabetes diagnosed during childhood and adolescence with complications during teenage years and young adulthood. *JAMA*. 2017;317:825–835.
15. Vileikyte L, Rubin RR, Leventhal H. Psychological aspects of diabetic neuropathic foot complications: an overview. *Diabetes Metab Res Rev*. 2004;20:S13–S18.
16. Vinik E, Silva MP, Vinik AI. Measuring the relationship of quality of life and health status, including tumor burden, symptoms, and biochemical measures in patients with neuroendocrine tumors. *Endocrinol Metab Clin North Am*. 2011;40:97–109, viii.
17. O'Connor AB. Neuropathic pain: quality-of-life impact, costs and cost effectiveness of therapy. *Pharmacoeconomics*. 2009;27:95–112.
18. O'Brien JA, Patrick AR, Caro J. Estimates of direct medical costs for microvascular and macrovascular complications resulting from type 2 diabetes mellitus in the United States in 2000. *Clin Therapeut*. 2003;25:1017–1038.
19. Boulton AJ, Malik RA, Arezzo JC, Sosenko JM. Diabetic somatic neuropathies. *Diabetes Care*. 2004;27:1458–1486.
20. Singleton JR, Bixby B, Russell JW, et al. The Utah Early Neuropathy Scale: a sensitive clinical scale for early sensory predominant neuropathy. *J Peripher Nerv Syst*. 2008;13:218–227.
21. Zilliox LA, Ruby SK, Singh S, et al. Clinical neuropathy scales in neuropathy associated with impaired glucose tolerance. *J Diabetes Complications*. 2015;29:372–377.
22. Ziegler D, Strom A, Lobmann R, et al. High prevalence of diagnosed and undiagnosed polyneuropathy in subjects with and without diabetes participating in a nationwide educational initiative (PROTECT study). *J Diabetes Complications*. 2015;29:998–1002.
23. Callaghan BC, Kerber KA, Lisabeth LL, et al. Role of neurologists and diagnostic tests on the management of distal symmetric polyneuropathy. *JAMA Neurol*. 2014;71:1143–1149.
24. Ziegler D, Rathmann W, Dickhaus T, et al. Prevalence of polyneuropathy in pre-diabetes and diabetes is associated with abdominal obesity and macroangiopathy: the MONICA/KORA Augsburg Surveys S2 and S3. *Diabetes Care*. 2008;31:464–469.
25. Ziegler D, Rathmann W, Dickhaus T, et al. Neuropathic pain in diabetes, prediabetes and normal glucose tolerance: the MONICA/KORA Augsburg Surveys S2 and S3. *Pain Medicine*. 2009;10:393–400.
26. Tesfaye S, Boulton AJ, Dyck PJ, et al. Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care*. 2010;33:2285–2293.
27. Ismail-Beigi F, Craven T, Banerji MA, et al. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial. *Lancet*. 2010;376:419–430.
28. Bril V, Perkins BA. Validation of the Toronto clinical scoring system for diabetic polyneuropathy. *Diabetes Care*. 2002;25:2048–2052.
29. Young MJ, Boulton AJ, MacLeod AF, et al. A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population. *Diabetologia*. 1993;36:150–154.
30. Spallone V, Morganti R, D'Amato C, et al. Validation of DN4 as a screening tool for neuropathic pain in painful diabetic polyneuropathy. *Diabet Med*. 2012;29:578–585.
31. Vileikyte L, Peyrot M, Bundy C, et al. The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. *Diabetes Care*. 2003;26:2549–2555.
32. Vinik EJ, Hayes RP, Oglesby A, et al. The development and validation of the Norfolk QOL-DN, a new measure of patients' perception of the effects of diabetes and diabetic neuropathy. *Diabetes Technol Ther*. 2005;7:497–508.
33. Krishnan ST, Rayman G. The LDIflare: a novel test of C-fiber function demonstrates early neuropathy in type 2 diabetes. *Diabetes Care*. 2004;27:2930–2935.
34. Ziegler D, Papanas N, Zhivov A, et al. Early detection of nerve fiber loss by corneal confocal microscopy and skin biopsy in recently diagnosed type 2 diabetes. *Diabetes*. 2014;63:2454–2463.
35. Perkins BA, Lovblom LE, Bril V, et al. Corneal confocal microscopy for identification of diabetic sensorimotor polyneuropathy: a pooled multinational consortium study. *Diabetologia*. 2018;61:1856–1861.
36. Boulton AJ. Diabetic neuropathy and foot complications. *Handb Clin Neurol*. 2014;126:97–107.
37. Ang L, Dillon B, Mizokami-Stout K, et al. Cardiovascular autonomic neuropathy: a silent killer with long reach. *Auton Neurosci: Basic & Clinical*. 2020;225:102646.
38. Pop-Busui R, Low PA, Waberski BH, et al. Effects of prior intensive insulin therapy on cardiac autonomic nervous system function in type 1 diabetes mellitus: the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study (DCCT/EDIC). *Circulation*. 2009;119:2886–2893.
39. Kempler P, Tesfaye S, Chaturvedi N, et al. Autonomic neuropathy is associated with increased cardiovascular risk factors: the EURODIAB IDDM Complications Study. *Diabet Med*. 2002;19:900–909.
40. Pambianco G, Costacou T, Ellis D, et al. The 30-year natural history of type 1 diabetes complications: the Pittsburgh Epidemiology of Diabetes Complications Study experience. *Diabetes*. 2006;55:1463–1469.
41. Andersen ST, Witte DR, Fleischer J, et al. Risk Factors for the presence and progression of cardiovascular autonomic neuropathy in type 2 diabetes: ADDITION-Denmark. *Diabetes Care*. 2018;41:2586–2594.
42. Tang Y, Shah H, Bueno Junior CR, et al. Intensive risk factor management and cardiovascular autonomic neuropathy in type 2 diabetes: the ACCORD Trial. *Diabetes Care*. 2021;44:164–173.
43. Lin YK, Fisher SJ, Pop-Busui R. Hypoglycemia unawareness and autonomic dysfunction in diabetes: lessons learned and roles of diabetes technologies. *J Diabetes Investig*. 2020;11:1388–1402.
44. Spallone V, Bellavere F, Scionti L, et al. Recommendations for the use of cardiovascular tests in diagnosing diabetic autonomic neuropathy. *Nutr Metabol Cardiovasc Dis: Nutr Metabol Cardiovasc Dis*. 2011;21:69–78.
45. O'Neal WT, Chen LY, Nazarian S, et al. Reference ranges for short-term heart rate variability measures in individuals free of cardiovascular disease: the Multi-Ethnic Study of Atherosclerosis (MESA). *J Electrocardiol*. 2016;49:686–690.
46. Pop-Busui R, Evans GW, Gerstein HC, et al. Effects of cardiac autonomic dysfunction on mortality risk in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. *Diabetes Care*. 2010;33:1578–1584.
47. Ziegler D, Zentai CP, Perz S, et al. Prediction of mortality using measures of cardiac autonomic dysfunction in the diabetic and nondiabetic population: the MONICA/KORA Augsburg Cohort Study. *Diabetes Care*. 2008;31:556–561.
48. Margolis KL, O'Connor PJ, Morgan TM, et al. Outcomes of combined cardiovascular risk factor management strategies in type 2 diabetes: the ACCORD randomized trial. *Diabetes Care*. 2014;37(6):1721–1728.
49. Azad N, Emanuele NV, Abraira C, et al. The effects of intensive glycemic control on neuropathy in the VA cooperative study on type II diabetes mellitus (VA CSDM). *J Diabetes Complications*. 1999;13:307–313.
50. Charles M, Fleischer J, Witte DR, et al. Impact of early detection and treatment of diabetes on the 6-year prevalence of cardiac autonomic neuropathy in people with screen-detected diabetes: ADDITION-Denmark, a cluster-randomised study. *Diabetologia*. 2013;56:101–108.

51. Duckworth W, Abraira C, Moritz T, et al. Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med*. 2009;360:129–139.
52. Gaede P, Vedel P, Larsen N, et al. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. *N Engl J Med*. 2003;348:383–393.
53. Carnethon MR, Jacobs Jr DR, Sidney S, et al. A longitudinal study of physical activity and heart rate recovery: cardia, 1987-1993. *Med Sci Sports Exerc*. 2005;37:606–612.
54. Carnethon MR, Prineas RJ, Temprosa M, et al. The association among autonomic nervous system function, incident diabetes, and intervention arm in the diabetes prevention program. *Diabetes Care*. 2006;29:914–919.
55. Jung HK, Choung RS, Locke 3rd GR, et al. The incidence, prevalence, and outcomes of patients with gastroparesis in Olmsted County, Minnesota, from 1996 to 2006. *Gastroenterology*. 2009;136:1225–1233.
56. Wang YR, Fisher RS, Parkman HP. Gastroparesis-related hospitalizations in the United States: trends, characteristics, and outcomes, 1995-2004. *Am J Gastroenterol*. 2008;103:313–322.
57. Kempler P, Amarenco G, Freeman R, et al. Gastrointestinal autonomic neuropathy, erectile-, bladder- and sudomotor dysfunction in patients with diabetes mellitus: clinical impact, assessment, diagnosis, and management. *Diabetes Metab Res Rev*. 2011;27:665–677.
58. Schvarcz E, Palmer M, Aman J, et al. Physiological hyperglycemia slows gastric emptying in normal subjects and patients with insulin-dependent diabetes mellitus. *Gastroenterology*. 1997;113:60–66.
59. Parkman HP, Camilleri M, Farrugia G, et al. Gastroparesis and functional dyspepsia: excerpts from the AGA/ANMS meeting. *Neuro Gastroenterol Motil*. 2010;22:113–133.
60. Giugliano F, Maiorino M, Bellastella G, et al. Determinants of erectile dysfunction in type 2 diabetes. *Int J Impot Res*. 2010;22:204–209.
61. Enzlin P, Rosen R, Wiegel M, et al. Sexual dysfunction in women with type 1 diabetes: long-term findings from the DCCT/ EDIC study cohort. *Diabetes Care*. 2009;32:780–785.
62. Pop-Busui R, Braffett BH, Hotaling J, et al. Diabetic neuropathy and urologic complications in men with type 1 diabetes in the diabetes control and complications trial/epidemiology of diabetes intervention and complications study (DCCT/EDIC). *Diabetes*. 2014;63:A148–A149.
63. Pop-Busui R, Hotaling J, Braffett BH, et al. Cardiovascular autonomic neuropathy, erectile dysfunction and lower urinary tract symptoms in men with type 1 diabetes: findings from the DCCT/EDIC. *J Urol*. 2015;193:2045–2051.
64. Shaw JE, Parker R, Hollis S, et al. Gustatory sweating in diabetes mellitus. *Diabet Med*. 1996;13:1033–1037.

45

Diabetic Kidney Disease

Radica Z. Alicic, Joshua J. Neumiller, and Katherine R. Tuttle

OUTLINE

Epidemiology, 770
Risk Factors, 770
Diagnosis and Classification, 772
- Biomarkers, 773

Natural History, 773
Structural Changes of the Kidney, 775
Pathophysiology, 775
- Increased Glucose Uptake in the Diabetic Kidney, 775
- Hemodynamic Abnormalities, 776
- Mechanotransduction, 776
- Inflammation and Fibrosis, 777
- Epigenetic Changes, 778

Treatment, 779
- Glycemic Control, 780
- Glucose-Lowering Agents, 780
- Blood Pressure Control, 782
- Cardiovascular Risk Reduction, 784
- Nutrition Management, 784
- Additional Interventions and Management Approaches, 784

Summary and Future Directions, 785

Diabetic kidney disease (DKD) is one of the most consequential complications of diabetes. DKD is defined as a sustained elevation of urinary albumin excretion (urine albumin-to-creatinine ratio [UACR] >30 mg/g) or a reduction in estimated glomerular filtration rate (eGFR) to less than 60 ml/min/1.73 m^2.[1] DKD develops in approximately one-third of patients with type 1 diabetes mellitus (T1DM) and half of patients with type 2 diabetes mellitus (T2DM).[1-3] The epidemiology of this microvascular complication has substantially changed over time. Whereas in the early 1970s DKD was classified as a relatively rare cause of kidney disease, constituting only 3% of patients transitioning to kidney replacement therapy (KRT), several decades later DKD has become both the leading cause of kidney failure classified as end-stage kidney disease (ESKD) and the most frequent indication for KRT.[4-6] Closely mirroring the diabetes pandemic, DKD has become a major public health problem worldwide.[7] In contrast to the morbidity and mortality associated with other diabetic complications (myocardial infarction, stroke, limb amputation) that are trending downward, the number of cases of DKD progressing to kidney failure or ESKD and kidney-related deaths continues to climb.[8-10] Besides the impending risk for progression to kidney failure, patients with DKD are at notably higher risk of experiencing cardiovascular events, stroke, infections, hospitalizations, and death.[11-14] A disproportionate burden of the increase in DKD (63%) has occurred in lower-middle income countries.[15] In these regions, KRT is only accessible to 9% to 16% of patients with kidney failure.[16-17] Soberingly, it is estimated that in the year 2010 alone more than 2.2 million individuals with kidney failure worldwide died owing to lack of KRT access.[18] Rates of kidney failure, largely driven by diabetes, continue to increase and are projected to double by the year 2030.[7,16]

Given the enormous humanistic and societal impact of DKD, improved approaches to prevention and treatment are critical to improving outcomes. This chapter outlines the complex pathophysiology and natural history of DKD. It offers an overview of best practices for DKD diagnosis, classification, and comprehensive care. Treatment of DKD is discussed with emphasis on the established standard of care, along with the newer classes of glucose-lowering agents within the sodium-glucose cotransporter 2 (SGLT-2) inhibitor and glucagon-like peptide-1 (GLP-1) receptor agonist classes.

EPIDEMIOLOGY

The reported prevalence of DKD varies depending on the methodology and the geographic setting. In the United States, studies have reported the prevalence of DKD to range between 26% and 44%.[19] These data align with occurrence of DKD in approximately one-third of patients with T1DM and up to half of patients with T2DM. The increase in prevalence of DKD parallels the increased prevalence of diabetes in the same time periods.[19] Progressively increasing trends continued between 2002 and 2016, with the total burden of chronic kidney disease (CKD) increasing by 53%, which was primarily driven by DKD. Moreover, 90% of all new cases of diabetes-attributable kidney failure occur in T2DM.[6,20] From 1990 to 2016, the global incidence and prevalence of CKD increased by 89% and 87%, respectively, and the death rate because of CKD increased by 98%.[15] Similar to trends reported in the United States, this upsurge is attributed to growth in the global age-standardized prevalence of DKD.[15] The United Kingdom Prospective Diabetes Study (UKPDS), which enrolled newly diagnosed T2DM participants, offered an opportunity to determine incident DKD. After median follow-up of 15 years, 38% developed albuminuria, 29% developed creatinine clearance less than 60 mL/min or doubling of plasma creatinine, and 14% developed both conditions.[3] A metaanalysis of 71 studies reported the annual incidence of albuminuria at 8% in T2DM and 2% to 3% in T1DM, while the incidence of developing eGFR less than 60 mL/min/1.73 m^2 was 2% to 4% per year, regardless of diabetes type[21] (Box 45.1).

BOX 45.1 EPIDEMIOLOGY OF DIABETIC KIDNEY DISEASE

- The number of people with diabetic kidney disease (DKD) is increasing globally in parallel to the diabetes pandemic.
- The worldwide incidence and prevalence of DKD increased by 90% from 1990 to 2016.
- DKD is the leading cause of kidney failure and need for kidney replacement therapy.

KEY POINTS

- In parallel to the diabetes pandemic, the incidence and prevalence of diabetic kidney disease in the 1990–2016 time span increased by 90%. This establishes diabetic kidney disease as a worldwide leading cause of kidney failure and kidney replacement therapy.

RISK FACTORS

Risk factors for DKD can be classified as those promoting susceptibility, initiation, and progression. These risk factors may influence more than one process (Table 45.1).[22] Demographic risk factors such as age, sex, and race/ethnicity are nonmodifiable. Both the incidence and the prevalence of DKD increase with age. Indeed, older age is an important predictor of progressive albuminuria and eGFR decline in diabetes.[23,24] Male sex is a risk factor for eGFR decline in T2DM, although women have a higher prevalence of diabetes.[25-27] In women, DKD typically develops more than 10 years later than in men, such that postmenopausal women have higher risk for developing DKD and progressing to kidney failure.[28-30] Global reports indicate that some racial or ethnic groups have greater susceptibility and worse prognosis for DKD.[31] People of South Asian, African American, Native American, Pacific Islander ancestry, or Hispanic ethnicity have a higher risk of eGFR decline, compared with people of European descent.[32-40] However, determining an independent association of race or ethnicity with DKD is complicated by potential differences in socioeconomic status or access to care that may influence these relationships.[35]

In terms of hereditary risk factors, early studies found an increased risk of DKD in family groupings, suggesting inherited susceptibility. The diabetic offspring of parents with DKD have a cumulative DKD risk of greater than 70%, with a 3- to 4-fold higher incidence compared with children of diabetic parents without DKD.[41-43] Although genome-wide association studies for genetic predisposition to DKD have thus far not produced conclusive results, epigenetic changes as a predisposition to DKD are being increasingly recognized.[44,45] Epigenetic predisposition has been demonstrated in multiple studies. For instance, in patients with T1DM, levels of DNA methylation correlate with time to DKD development.[46] A study that compared DNA methylation levels in Black American and Latino patients with kidney failure owing to DKD to those without DKD found that approximately 20% of the genes exhibiting different methylation between the two groups had been previously implicated in CKD or DKD in other genome-wide association or transcription profiling studies.[47] A majority (24 out of 30) of the top DNA methylation markers associated with GFR changes were related to aging, inflammation, or lipid pathways.[48]

Exposure to maternal diabetes during gestation is associated with higher risk of DKD and high birth weight (≥4500 g).[31] When compared with individuals not exposed to gestational hyperglycemia, the offspring of mothers with T1DM demonstrate alterations in the DNA methylation profile and kidney hemodynamics (GFR and blood flow) indicating reduced kidney reserve that implies susceptibility to DKD.[49,50] Notably, both low (<2500 g) and high (≥4500 g) birth weight have been associated with higher risk of albuminuria in the offspring of diabetic mothers.[51-54] For low birth weight infants, the proposed mechanism underlying higher risk of DKD is growth retardation with low nephron number that predisposes to development of kidney disease and hypertension.[51-54] In contrast, the Finnish Diabetic Nephropathy study did not find a relationship between low birth weight and DKD.[55]

The presence of obesity often precedes the diagnosis of T2DM by years or decades and may exert an additive effect on risk of DKD.[56] One study of obese participants with T2DM and preserved kidney function demonstrated a significantly faster age-adjusted annual eGFR decline (–1.2 ± 0.1 mL/min/1.73 m^2 per year) compared with normal weight individuals.[57] The mechanism(s) by which obesity increases DKD risk is not clearly elucidated, but may include inflammation, abnormal fatty acid metabolism, hemodynamic abnormalities, and exacerbation of hypertension and insulin resistance.[58-61]

Key modifiable risk factors for development and progression of DKD are hyperglycemia and hypertension. The association of glycemic control and microvascular complications has been well described in both T1DM and T2DM. Evidence supporting intensive glycemic control early in the course of diabetes comes from two landmark trials: the Diabetes Control and Complications Trial (DCCT), with follow-up within the Epidemiology of Diabetes Interventions and Complications (EDIC) study in T1DM and the UKPDS in T2DM.[62-64] During the 20-year follow-up period of the ongoing EDIC study, the glycated hemoglobin (HbA1c) difference (7.3% vs. 9.1% during active study in the intervention and control groups, respectively) dissipated, such that the subsequent mean HbA1c was approximately 8% in both groups.[62,63] Despite equalization of overall glycemic control, an analysis including 9 years of EDIC data showed that the intensive glycemic control group exhibited 34% and 56% lower risks of developing microalbuminuria and macroalbuminuria, respectively.[65] After 22 years of follow-up, intensive glycemic control was associated with 50% lower risk of developing eGFR less than 60 mL/min/1.73 m^2 and kidney failure, as well as a slower rate of eGFR loss (–1.56 mL/min/1.73 m^2 per year in the standard therapy group versus –1.27 mL/min/1.73 m^2 per year in the intensive therapy group).[66] Between-group differences in glycemic control of newly diagnosed T2DM patients were lost 1 year into the posttrial follow-up stage of the UKPDS study.[67] Nevertheless, 10 to 12 years of posttrial follow-up showed a 24% reduction in risk of microvascular complications, including macroalbuminuria, as well as a reduction in the rate of serum creatinine doubling in the intensive compared with the conventional therapy group (0.9% vs. 3.5%, $P = .0028$).[68,69] This “metabolic memory” in response to early intensive glycemic control highlights its importance from diabetes onset to prevent DKD. Epigenetic alterations may be an important contributor to metabolic memory in DKD.[45] In a study of DCCT/EDIC participants who developed DKD, a set of abnormal DNA methylation patterns persisted for 17 years during follow-up into the EDIC study, irrespective of glycemic control.[70]

Blood pressure control is another central component of DKD prevention and treatment. An analysis of UKPDS participants followed for a median of 15 years showed that a target blood pressure of less than 150/85 mm Hg when compared with a target blood pressure of less than 180/105 mm Hg resulted in a 39% reduced risk for microvascular complications.[71] Overall, each 10 mm Hg increase in mean systolic blood pressure (SBP) was associated with a 13% increase in the hazard ratio for development of both microalbuminuria and macroalbuminuria and eGFR less than 60 mL/min/1.73 m^2 or a doubling of the serum creatinine.[71] A *post hoc* analysis of the Reduction of Endpoints in Non–Insulin-Dependent Diabetes Mellitus with the Angiotensin II

TABLE 45.1 Risk Factors for Diabetic Kidney Disease

Risk Factor	Susceptibility	Initiation	Progression
Demographic			
Older age	+		
Sex (male, postmenopausal female)	+		+
Race/ethnicity (Black American, Native American, Hispanic, Asian/Pacific Islanders)	+		+
Hereditary			
Family history of diabetic kidney disease	+	+	+
Genetic kidney disease	+	+	+
Alterations of developmental programming/low nephron number			
Low birth weight/premature birth	+		+
Small for gestational age	+		+
Prematurity	+		+
Exposure to maternal diabetes	+		+
Systemic conditions			
Hyperglycemia	+	+	+
Obesity	+	+	+
Hypertension	+		+
Kidney injuries			
Acute kidney injury		+	+
Toxins		+	+
Smoking	+		+
Dietary factors			
High protein intake	+		+

(Modified from Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol*. 2017; 12:2032–2045.)

Antagonist Losartan study demonstrated that SBP was an independent risk factor for kidney failure. A baseline SBP in the range of 140 to 159 mm Hg increased risk for ESKD or death by 38% compared with a SBP less than 130 mm Hg.[72] An analysis of the Irbesartan Diabetic Nephropathy Trial data showed that a baseline SBP greater than 149 mm Hg associated with 2.2-fold higher risk for doubling of serum creatinine or kidney failure compared with SBP less than 134 mm Hg.[73] Progressive lowering of SBP to 120 mm Hg was associated with reduced risk for doubling of serum creatinine or kidney failure, as well as for survival, independent of baseline kidney function.[73]

Diabetes is a risk factor for acute kidney injury (AKI). Patients with diabetes undergoing surgical procedures or suffering from acute illness have a high risk of AKI. Notably, children with T1DM hospitalized with diabetic ketoacidosis (DKA) also have increased risk of AKI.[74] However, AKI may also occur in the absence of an obvious precipitating event.[75-77] Single or multiple repetitive episodes of AKI significantly increase the risk of developing DKD and progression to kidney failure.[78,79]

High blood amino acids levels induce glomerular hyperfiltration, a major mechanism of hemodynamic injury to the diabetic kidney. As such, excess dietary protein intake can promote initiation and progression of DKD. High dietary protein intake is associated with development of albuminuria, decline in eGFR, and cardiovascular mortality.[80-82] In contrast, lower dietary protein intake reduces glomerular hyperfiltration and may slow progression of DKD.[83,84] In patients with T1DM, a diet with lower (0.89 g/kg/day) compared with higher (>1g/kg/day) protein intake resulted in a reduced rate of kidney failure (10% vs. 27%, $P = .042$).[85] A metaanalysis of eight studies showed that the benefit of a low-protein diet was mainly on albuminuria reduction in diabetic patients.[86] Notably, these studies were underpowered to assess effects on clinical outcomes. As albuminuria is highly associated with future risk of eGFR decline, a lower protein diet or avoidance of excessive protein intake are considered favorable for DKD overall (Box 45.2).

BOX 45.2 RISK FACTORS FOR DIABETIC KIDNEY DISEASE

- Nonmodifiable risk factors associated with diabetic kidney disease (DKD) include age, sex, and ethnicity.
- A family history of DKD increases DKD risk.
- Hyperglycemia and hypertension control are key modifiable risk factors for development and progression of DKD.
- Hyperglycemia early in the course of disease leads to "metabolic memory" that impacts future risk of DKD.

KEY POINTS

- Uncontrolled hyperglycemia early in the course of diabetes leads to "metabolic memory" and together with hypertension control are the most important modifiable risk factors for development and progression of diabetic kidney disease.

DIAGNOSIS AND CLASSIFICATION

DKD refers to a diagnosis of CKD in a patient with diabetes without another known or likely cause. CKD is established by two abnormal measurements of laboratory markers that indicate kidney damage or

TABLE 45.2 Categories of Chronic Kidney Disease by Estimated Glomerular Filtration Rate and Albuminuria

Stage	Description	Estimated Glomerular Filtration Rate (mL/min/1.73 m^2)
G1	Normal or high	≥90
G2	Mildly decreased	60–89
G3a	Mildly to moderately decreased	45–59
G3b	Moderately to severely decreased	30–44
G4	Severely decreased	15–29
G5	Kidney failure	<15

Stage	Description	Albuminuria	
A1	Normal to mildly increased	<30 mg/g	<3 mg/mmol
A2	Moderately increased	30–300 mg/g	3–30 mg/mmol
A3	Severely increased	>300 mg/g	>30 mg/mmol

(Modified from Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl.* 2013;3:1–150.)

loss of function, at least 3 months apart. These CKD markers are albuminuria, low eGFR, or both. Moreover, the presence of retinopathy, diabetes duration longer than 5 years in T1DM, or diabetes duration longer than 10 years in T2DM are highly predictive of classical diabetic glomerulopathy.[1,87,88] The preferred test for albuminuria is UACR performed on a random urine spot sample, if possible in the morning. Increased urinary albumin excretion is defined as a UACR of 30 mg/g or more. The eGFR is calculated from the serum creatinine concentration using either the Modification of Diet in Renal Disease or Chronic Kidney Disease-Epidemiologic Prognosis Initiative (CKD-EPI) equations. The CKD-EPI equation is now preferred because it has greater precision at higher levels of eGFR (>60 mL/min/1.73 m^2).[89] The serum cystatin C level can also be used to calculate eGFR by the CKD-EPI equation, either alone or combined with the serum creatinine level, which provides even greater precision for eGFR.[90] However, cystatin C is not standardized or available in many laboratories.

Kidney biopsy is not routinely performed to diagnose DKD. However, a diagnostic kidney biopsy may be considered if there is a clinical suspicion for an alternative diagnosis, or if there are atypical features of DKD such as rapid eGFR decline, abrupt increase in albuminuria, a nephritic syndrome, or signs or symptoms of another systemic disease.[1,87] Annual measurements of albuminuria and serum creatinine are recommended for people with T1DM beginning 5 years after the diagnosis, and for all T2DM patients beginning the time of diagnosis of diabetes (Table 45.2).[1,88,91,92] If there is a change in clinical status (progression to a more advanced CKD stage) or medication (renin angiotensin system inhibitor or an SGLT-2 inhibitor), then UACR and eGFR should be measured more frequently (every 3–6 months). Referral to a nephrologist is recommended if there is need for further diagnostic evaluation or management guidance, or if the eGFR is less than 30 mL/min/1.73 m^2 (Box 45.3).[1,87]

Biomarkers

Use of serum creatinine to measure eGFR has limitations, such as variability depending upon age, sex, and muscle mass. Addition of serum cystatin C to the CKD-EPI equation is recommend to increase precision for eGFR over 60 mL/min/1.73 m^2.[90] Limitations related to measurements of albuminuria include lack of specificity and sensitivity, with up to 40% day-to-day variability.[88,93-96]

BOX 45.3 DIAGNOSIS OF DIABETIC KIDNEY DISEASE

- Diabetic kidney disease (DKD) is diagnosed by microalbuminuria (urine albumin-to-creatinine ratio [UACR] 30–300 mg/g), macroalbuminuria (UACR >300 mg/g), estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2, or a combination of high albuminuria and low eGFR that persists for at least 3 months in a person with diabetes.
- Annual screening for DKD is recommended within 5 years of a diagnosis of type 1 diabetes mellitus and beginning at the time of diagnosis of type 2 diabetes mellitus.
- If there is a change in clinical status (progression to a more advanced chronic kidney disease category) or medications that can impact UACR and/or eGFR (renin angiotensin system inhibitor or sodium-glucose cotransporter 2 inhibitor), then UACR and eGFR should be measured more frequently (every 3–6 months).
- Referral to a nephrologist is recommended if there is need for further diagnostic evaluation or management guidance, or if the eGFR is <30 mL/min/1.73 m^2.

New biomarkers that will advance DKD diagnosis, risk stratification, monitoring, and prognosis are an active area of investigation. Several studies assessed individual or panels of biomarkers representing different pathways or processes involved in the pathogenesis of DKD using urinary proteomics or metabolomics.[97-99] Promising candidates include soluble tumor necrosis factor receptors (TNFR1 and TNFR2) and kidney injury molecule-1 (KIM-1).[100] These biomarkers were found to be 2-fold higher among those with advanced DKD compared with earlier stages of disease and predicted eGFR decline and kidney failure.[101] Similarly, urinary monocyte chemoattractant protein-1 (MCP1) levels were elevated in diabetic patients with macroalbuminuria and were prognostic for deterioration in eGFR.[102] Analysis of Chronic Renal Insufficiency Cohort Study participants with DKD and eGFR less than 60 mL/min/1.73 m^2 at baseline demonstrated that, in addition to TNFR1, TNFR2, and MCP1, higher plasma levels of soluble urokinase-type plasminogen activator receptor and YKL-40 were associated with increased risk of progression of DKD.[103] Epigenetic studies have identified signatures of specific DNA methylation sites, histone posttranslational modifications, and urinary microRNAs (miRNAs) that are associated with low eGFR or albuminuria.[45,104] However, no new candidates have meaningfully outperformed traditional DKD biomarkers for risk prediction. Albuminuria remains the best predictor of DKD progression and kidney failure.[105]

KEY POINTS

- Annual screening for diabetic kidney disease should start within 5 years of a diagnosis of type 1 diabetes mellitus or at the time of diagnosis of type 2 diabetes mellitus, and is established by presence of microalbuminuria (urine albumin-to-creatinine ratio [UACR] 30–300 mg/g), macroalbuminuria (UACR >300 mg/g), estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2, or a combination of high albuminuria and low eGFR that persists for at least 3 months.

NATURAL HISTORY

The traditional model for the natural history of DKD centers around development of glomerular hyperfiltration that progresses to increasing albuminuria followed by declining eGFR (Fig. 45.1).[106] However, it is clear that DKD can present with low eGFR and without albuminuria.[107]

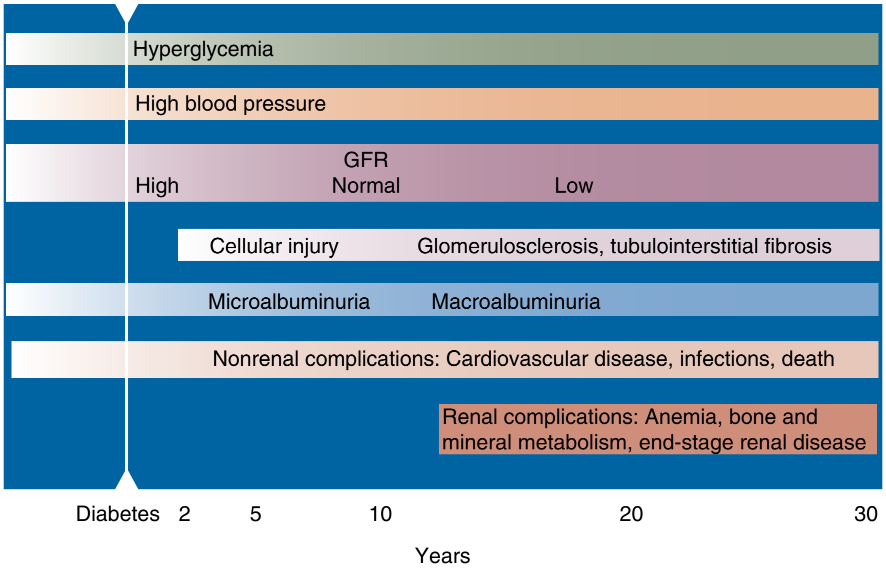

Hyperglycemia
High blood pressure
GFR
High
Normal
Low
Cellular injury
Glomerulosclerosis, tubulointerstitial fibrosis
Microalbuminuria
Macroalbuminuria
Nonrenal complications: Cardiovascular disease, infections, death
Renal complications: Anemia, bone and mineral metabolism, end-stage renal disease
Diabetes
2
5
10
20
30
Years

Fig. 45.1 Conceptual Model of The Natural History of Diabetic Kidney Disease. Duration of diabetes, in years, is presented on the horizontal axis. Timeline is well characterized for type 1 diabetes mellitus; for type 2 diabetes mellitus, timeline may depart from the illustration because of the variable timing of the onset of hyperglycemia. *Kidney complications: anemia, bone and mineral metabolism, retinopathy, and neuropathy. *GFR,* Glomerular filtration rate. (From Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol.* 2017; 12:2032–2045.)

For instance, over a median follow-up of 15 years of patients with newly diagnosed T2DM in the UKDPS, 40% developed albuminuria, and 30% progressed to an eGFR under 60 mL/min/1.73 m^2 or experienced a doubling in serum creatinine. Of participants experiencing a decline in kidney function, 60% did not have preceding albuminuria.[3] Further support for this observation came from the Developing Education on Microalbuminuria for Awareness of Renal and Cardiovascular Risk in Diabetes study, with approximately 21%, 31%, and 35% of individuals with normoalbuminuria, microalbuminuria, and macroalbuminuria, respectively, developing a creatinine clearance of less than 60 mL/min/1.73 m^2.[93] The clinical presentation of DKD has shifted such that the prevalence of low eGFR in the absence of albuminuria occurs more commonly. Indeed, a study comparing DKD presentation during the time periods of 1988 to 1994 and 2009 to 2014 reported that the prevalence of albuminuria decreased from 21% to 16%, while the prevalence of low eGFR (<60 mL/min/1.73 m^2) and severely reduced eGFR (<30 mL/min/1.73 m^2) increased from 9% to 14%, and 1% to 3%, respectively.[92] The classical paradigm of DKD is further challenged by findings that albuminuria may be dynamic and fluctuating rather than linearly progressive. For example, during nearly 8 years of follow-up in the Multifactorial Intervention for Patients with Type 2 Diabetes Study, 31% of participants with microalbuminuria progressed to macroalbuminuria, 38% remained microalbuminuric, and 31% regressed to normoalbuminuria.[108] Another study demonstrated remission of microalbuminuria to normoalbuminuria in 24% of cases and regression of macroalbuminuria to microalbuminuria in 30% of cases.[109]

Only limited data on the natural history of DKD by diabetes type are available. In general, albuminuria rates appear to be similar in patients with T1DM and T2DM. However, low eGFR is more common in T2DM.[110] In terms of progression to kidney failure, rates of eGFR decline can be highly variable between patients. Data are also mixed regarding whether the incidence of kidney failure differs by type of diabetes.[111-112] Data from patients with T1DM and T2DM registered in Australia's National Diabetes Services Scheme showed that the age-standardized annual incidence of kidney failure increased for people with T2DM and remained stable for patients with T1DM from 2002 to 2013 (Box 45.4).[112]

BOX 45.4 NATURAL HISTORY OF DIABETIC KIDNEY DISEASE

- Estimated glomerular filtration rate (eGFR) decline is faster in association with albuminuria, especially macroalbuminuria, but may also occur in non-albuminuric patients.
- The natural history of diabetic kidney disease (DKD) appears to be similar in both types of diabetes.
- Patients with DKD are more likely to have anemia compared with patients with nondiabetic chronic kidney disease at comparably reduced levels of eGFR.
- Insulin resistance in patients with diabetes contributes to lower parathyroid hormone levels with an increased predisposition to adynamic bone disease.

DKD produces hormonal changes that often contribute to complications of kidney failure earlier than in nondiabetic CKD. Because of a prominent tubulointerstitial component of DKD, damage to erythropoietin-producing peritubular interstitial cells occurs with consequent erythropoietin deficiency. Therefore, patients with DKD are nearly twice as likely to have anemia compared with patients with nondiabetic CKD at comparably reduced levels of eGFR.[22] Insulin resistance in patients with diabetes contributes to lower parathyroid hormone levels, with an increased predisposition to adynamic bone disease.[113]

KEY POINTS

- The natural history of diabetic kidney disease is similar in both types of diabetes. Estimated glomerular filtration rate declines faster in association with albuminuria, especially macroalbuminuria, but may also occur in nonalbuminuric patients. Patients with diabetic kidney disease are more likely to have anemia compared with patients with nondiabetic chronic kidney disease at comparably reduced levels of kidney function, and lower parathyroid hormone levels with an increased predisposition to adynamic bone disease.

STRUCTURAL CHANGES OF THE KIDNEY

Structural changes of DKD occur within multiple kidney compartments.[114] The earliest changes are enlargement of the kidney because of glomerular and tubular hypertrophy.[115,116] Additionally, glomerular basement membrane (GBM) thickening is evident as early as 1.5 to 2 years following a diagnosis of T1DM.[117-119] Glomerular and tubular hypertrophy may be driven by activation of growth factors within the diabetic milieu,[120] whereas increased formation of glycated proteins and mechanical stress of glomerular hyperfiltration are implicated in the thickening of GBM (Fig. 45.2). Other features of early DKD include expansion of the glomerular mesangium.[121-122] In part because of an adaptive response to the thickening GBM, podocytes hypertrophy, flatten, and detach. Eventually severe loss of podocytes ensues, leaving large subepithelial surfaces of the GBM denuded.[122-124] Progression of the structural glomerular changes include thickening of Bowman's capsule and greater glomerular mesangial expansion owing to increased mesangial matrix and, to a lesser extent, increases in mesangial cellularity.[125] The mesangial matrix expansion often develops into a form of nodular glomerulosclerosis known as Kimmelstiel-Wilson nodules. Plasma proteins, or hyaline material, often accumulate between the glomerular endothelium and the GBM and between the glomerular tuft and Bowman's capsule (Fig. 45.3).[126-128] The tubulointerstitium progressively develops fibrosis and inflammation, thickening of the tubular basement membranes, and tubular atrophy.[129] The blood vessels display arteriosclerosis and hyalinosis, particularly at the vascular pole. The tubulointerstitial damage progresses in concert with glomerulosclerosis. In advanced stages, global glomerular sclerosis with severe interstitial fibrosis and tubular atrophy are evident (Fig. 45.4).[128] The international Renal Pathology Society work group developed a histopathological classification system to describe and grade the severity of DKD.[130]

Structural changes in DKD are similar in the setting of T1DM and T2DM, but the pathological findings and correlations with the clinical presentation are less predictable in T2DM.[131] For instance, an autopsy study of predominantly T2DM (90%) patients found that twenty out of 106 histologically proven cases of DKD did not have a clinical diagnosis of DKD while alive, and specifically did not have albuminuria.[96] Structural changes in this cohort encompassed the spectrum of histopathological classes of DKD.[96] These findings were confirmed in diabetic kidney transplant donors (98% T2DM) in whom structural changes preceded clinical evidence of DKD (Box 45.5).[132]

KEY POINTS

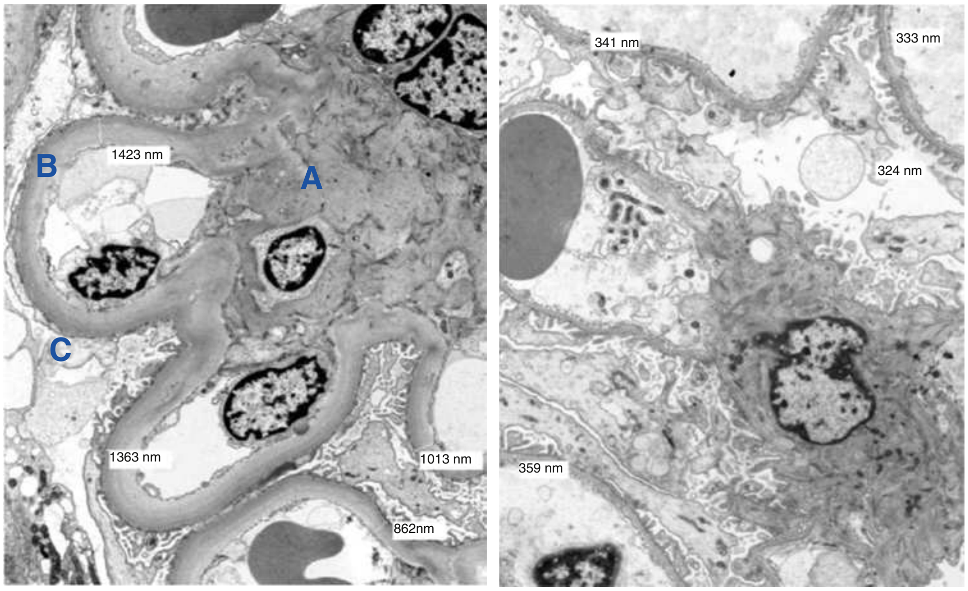
- Structural changes involve all kidney compartments and are similar in patients with type 1 diabetes mellitus and type 2 diabetes mellitus. The earliest changes are glomerular and tubular hypertrophy, followed by glomerular basement thickening, podocyte foot process effacement and detachment, mesangial expansion and formation of Kimmelstiel–Wilson nodules, glomerulosclerosis, tubulointerstitial fibrosis and inflammation, and arteriolar hyalinosis and arteriosclerosis.

PATHOPHYSIOLOGY

The metabolic, hemodynamic, inflammatory, fibrotic, and epigenetic changes induced by the diabetic milieu trigger a complex, multidirectional, self-perpetuated cascade leading to structural and functional abnormalities associated with DKD (Fig. 45.5, Box 45.6).

Increased Glucose Uptake in the Diabetic Kidney

Important contributions of the kidney to glucose homeostasis are reabsorption of glucose from the glomerular filtrate and gluconeogenesis.[133] Glucose reabsorption takes place in the cells of the proximal tubule via apically located SGLT-1 and SGLT-2, and through facilitative glucose transporters (GLUTs) located on the basolateral side of



Diabetic Glomerulopathy

Normal Glomerulus

Fig. 45.2 Electron microscope images of structural changes in diabetic kidney disease. Structural changes in diabetic glomerulopathy found with electron microscopy. **A,** Marked expansion of the mesangium. **B,** Marked diffuse thickening of capillary basement membranes (to three times the normal thickness in this case). **C,** Segmental effacement of the visceral epithelial foot processes. Original magnification, x3500. (From Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol.* 2017; 12:2032–2045.)

the tubular cells.[134-136] The majority of the reabsorption takes place in the early segment of the proximal tubule via coupling of SGLT-2 and GLUT-2 transporters. Only approximately 10% of glucose is reabsorbed via SGLT-1 and GLUT-1 in the S3 segment of the proximal tubule.[137] A landmark trial published in 1971 demonstrated increased uptake of glucose in the proximal tubule in diabetic patients, thus indicating enhanced reabsorption transport capacity in diabetes.[138] Indeed, the threshold for glucosuria in diabetes is in the range of 200 to 240 mg/dL, which is increased from the threshold of 100 to 240 mg/dL reported under physiological conditions.[135] The increase in resorptive capacity in diabetes is a result of increased SGLT and GLUT transporters driven by increased mRNA expression and production. Support for this hypothesis comes from experimental models of diabetes. Mouse and rat models, as well as studies using tubular epithelial cells isolated from the urine of diabetic patients, report increased expression of SGLT-1, SGLT-2, GLUT-1, and GLUT-2 proteins.[139-145] Increased glucose and sodium chloride uptake in the proximal tubule via SGLT-2 reduces solute delivery to the macula densa and significantly contributes to a vasoactive imbalance between the preglomerular (afferent) and postglomerular (efferent) arterioles, resulting in glomerular hyperfiltration (Fig. 45.6).[22,146]

Hemodynamic Abnormalities

Glomerular hyperfiltration is driven by an increase in the glomerular capillary pressure.[116,147] Glomerular hyperfiltration is present in about 75% of patients with T1DM and in approximately 40% of patients with early T2DM.[148-150] Even though increased glomerular plasma flow, hyperfiltration, and hypertension are typically associated with diabetes, these hemodynamic aberrations are also found in patients with prediabetes and obesity.[151-152] Although the exact mechanism is unclear, hyperglycemia, abnormal insulin levels, and dietary proteins are all important mediators.[115,153-155] Major factors contributing to increased efferent arteriolar resistance (vasoconstriction) include activation of the renin-angiotensin system, with increased production of angiotensin II and endothelin-1, whereas reduction of afferent arteriolar resistance (vasodilation) is driven by inhibition of tubuloglomerular feedback. Tubuloglomerular feedback is an autoregulatory mechanism that adjusts GFR in response to the concentration of sodium chloride delivered to the macula densa segment of the distal convoluted tubule. A byproduct of sodium reabsorption is adenosine produced by utilization of adenosine triphosphate for active sodium–potassium exchange on the basolateral side of macula densa cells. Adenosine release produces a paracrine effect to increase afferent arteriolar resistance (vasoconstriction). In diabetes, increased sodium and glucose reabsorption in the proximal tubule reduces the delivery of solutes to the macula densa, effectively inhibiting tubuloglomerular feedback and adenosine-mediated vasoconstriction of the preglomerular (afferent) arteriole (Fig. 45.6).[156] The imbalance between pre- and postglomerular arteriolar resistance leads to development of glomerular hypertension and hyperfiltration.[115]

Mechanotransduction

Increased mechanical stress caused by glomerular hyperfiltration influences both glomerular and tubular structures. Stress-induced expansion of the GBM prompts adaptive podocyte hypertrophy and foot process effacement, ultimately resulting in denudement of areas of the GBM and podocyte detachment. Podocytes respond to mechanical stress by modifying the actin cytoskeleton and activating various signaling pathways, including protein kinases, prostaglandins, angiotensin II, mTOR, and transforming growth factor-β (TGF-β.) This property,

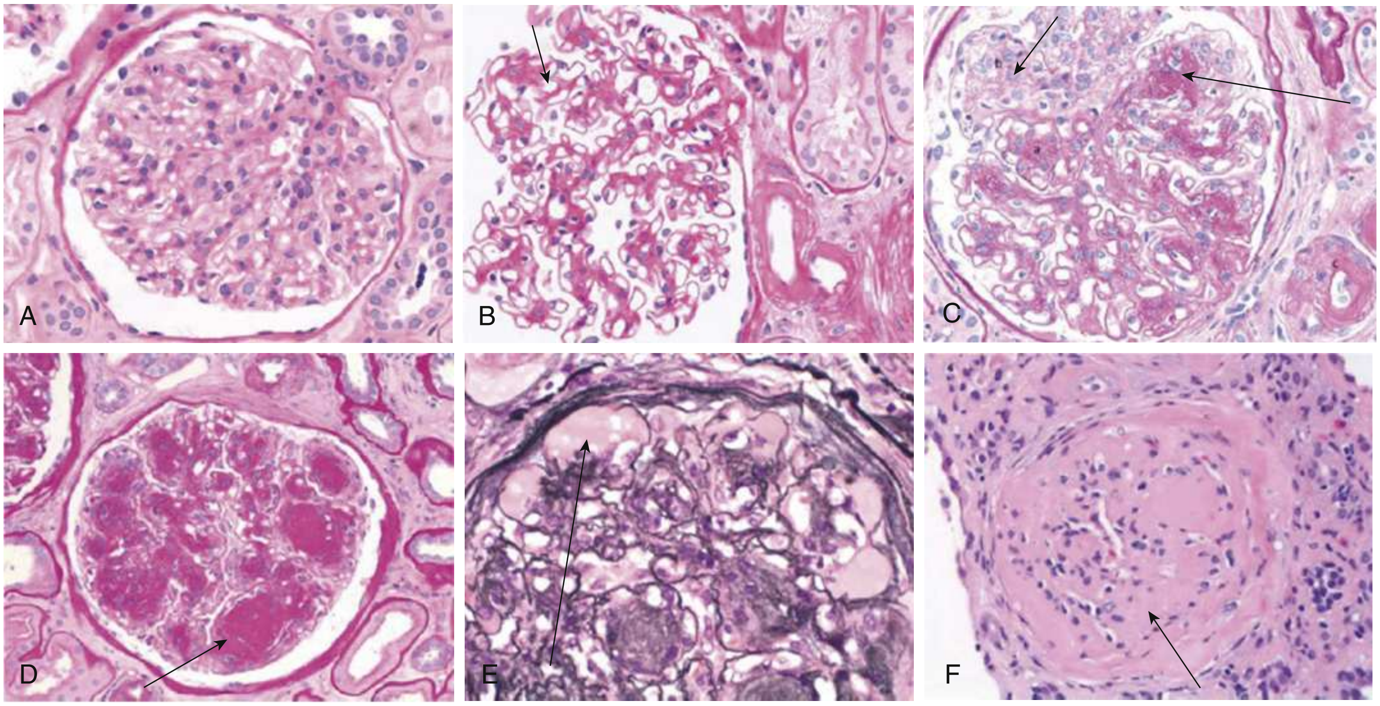


Fig. 45.3 **Diabetic glomerulopathy.** Changes in glomerular histology in diabetic glomerulopathy. **A,** Normal glomerulus. **B,** Diffuse mesangial expansion with mesangial cell proliferation. **C,** Prominent mesangial expansion with early nodularity and mesangiolysis. **D,** Accumulation of mesangial matrix forming Kimmelstiel-Wilson nodules. **E,** Dilation of capillaries forming microaneurysms, with subintimal hyaline (plasmatic insudation). **F,** Obsolescent glomerulus. **A**–**D** and **F** were stained with period acid-Schiff stain, and **E** was stained with Jones stain. Original magnification, 400×. (From Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol*. 2017; 12:2032–2045.)

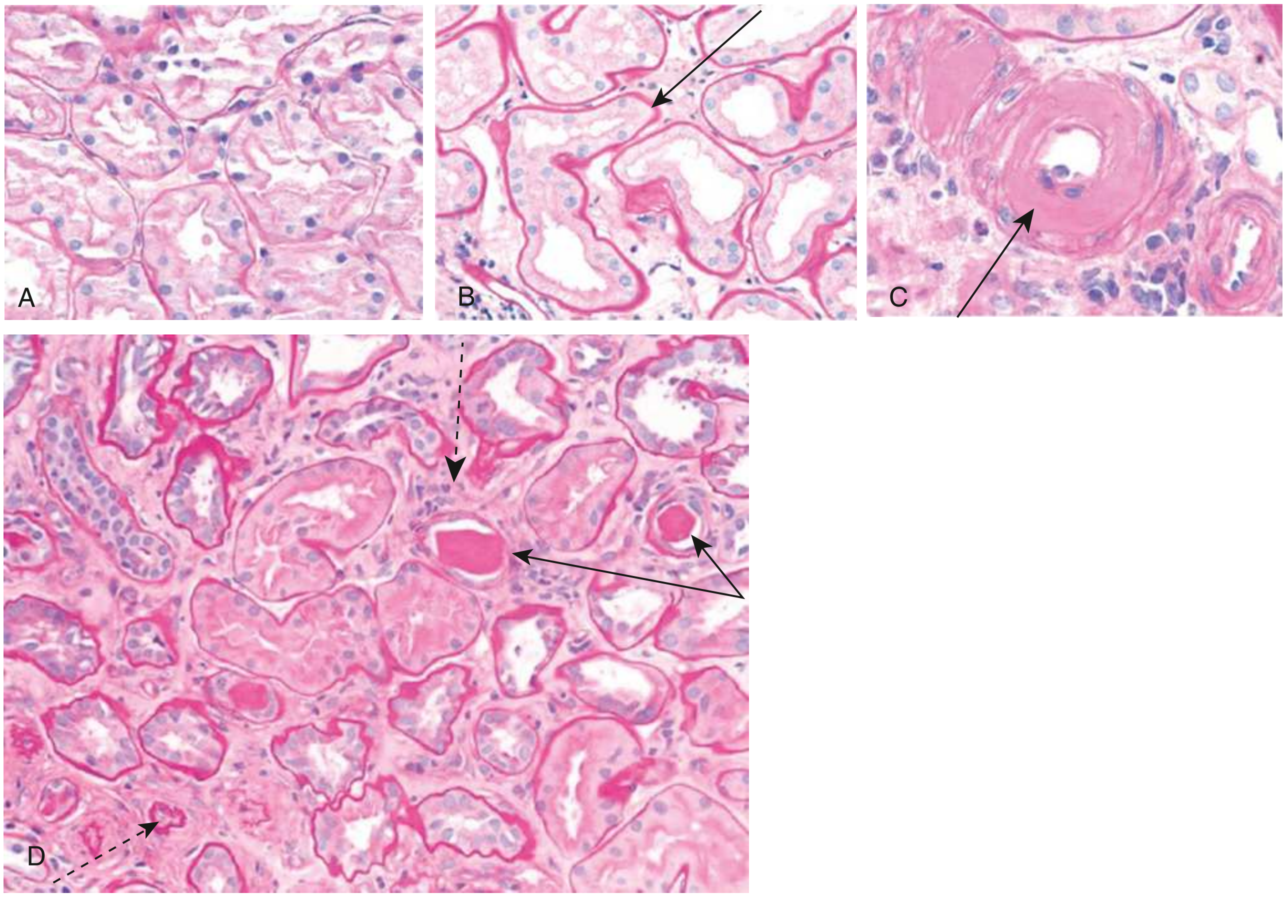


Fig. 45.4 **Tubulointerstitial changes and arteriolar hyalinosis in diabetic kidney disease.** Tubulointerstitial changes in diabetic kidney disease. **A,** Normal renal cortex. **B,** Thickened tubular basement membranes and interstitial widening. **C,** Arteriole with an intimal accumulation of hyaline material with significant luminal compromise. **D,** Renal tubules and interstitium in advancing diabetic kidney disease, with thickening and wrinkled tubular basement membranes (*solid arrows*), atrophic tubules (*dashed arrows*), some containing casts, and interstitial widening with fibrosis and inflammatory cells (*dotted arrow*). All sections were stained with period acid-Schiff stain. Original magnification, 200×. (From Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol*. 2017; 12:2032–2045.)

BOX 45.5 STRUCTURAL CHANGES OF THE KIDNEY IN DIABETIC KIDNEY DISEASE

- Structural changes in diabetic kidney disease (DKD) are similar in patients with type 1 diabetes mellitus and type 2 diabetes mellitus, but the pathological and clinical presentation is less predictable in type 2 diabetes mellitus.
- All kidney compartments display structural changes.
- The earliest changes are glomerular and tubular hypertrophy at the stage of glomerular hyperfiltration.
- Glomerular changes include glomerular basement membrane thickening, podocyte foot process effacement and detachment, mesangial expansion and formation of Kimmelstiel–Wilson nodules, and glomerulosclerosis.
- The tubulointerstitium develops fibrosis and inflammation and tubular atrophy.
- The arterioles display hyalinosis and arteriosclerosis.
- Structural changes may be present without clinically evident DKD.

known as "mechanotransduction," is now recognized as an important contributor to DKD pathogenesis.[157-160] Mechanotransduction also plays a role in tubulointerstitial injury, where increased shear stress on proximal tubule microvilli transmits to the brush border cytoskeleton. This process also signals activation of the luminal NHE3 transporter and the basolateral sodium-bicarbonate cotransporter to increase sodium reabsorption in the proximal tubule and thereby exacerbates glomerular hyperfiltration by decreasing distal sodium delivery and reducing tubuloglomerular feedback.[161-162]

Inflammation and Fibrosis

Activation of a network of inflammatory cells coexpressing markers specific for macrophages and dendritic interstitial cells, referred to as the renal mononuclear phagocyte system, precipitates the release of proinflammatory cytokines and paracrine signals in the diabetic kidney.[163-164] Proinflammatory cytokines, in turn, recruit additional monocytes and macrophages, perpetuating a vicious cycle of cytokine release and inflammatory cell recruitment.[165-168] The number of proinflammatory cells and cytokines correlates with DKD progression. The extent of glomerulosclerosis, tubulointerstitial inflammation, and rate of eGFR decline are correlated with macrophage infiltration.[169-170] Involvement of proinflammatory cytokines (e.g., interleukin [IL] 1, IL-6, IL-18, and tumor necrosis factor α) has been implicated in contributing to numerous DKD mechanisms, including oxidative stress, glomerular hemodynamics, recruitment and activation of inflammatory cells, extracellular matrix formation, GBM thickening, and podocyte and endothelial cell injury.[171-176] Similarly, chemokines and adhesion molecules, including MCP1/chemokine C-C motif-ligand 2, C-X3-C motif chemokine, and

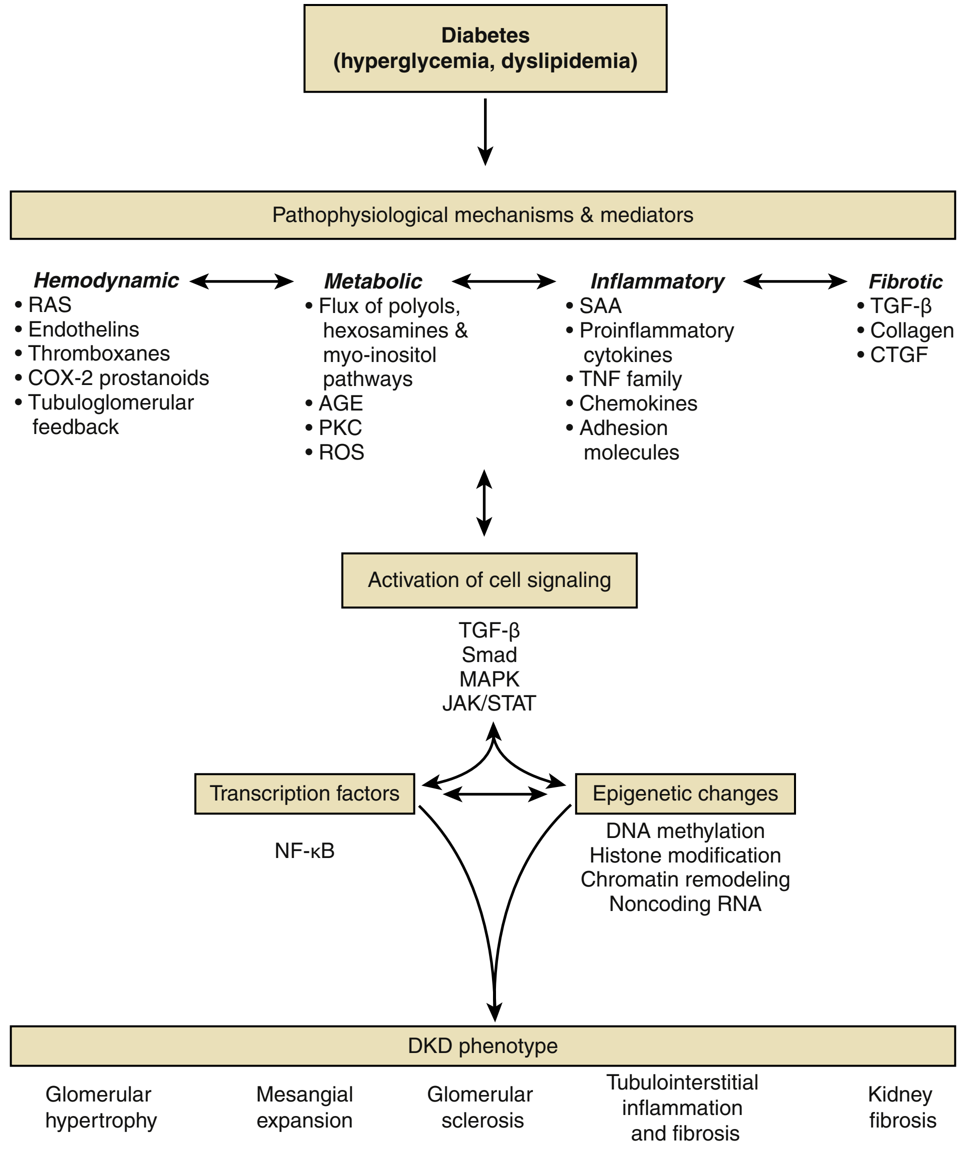

Diabetes
(hyperglycemia, dyslipidemia)
Pathophysiological mechanisms & mediators
Hemodynamic
• RAS
• Endothelins
• Thromboxanes
• COX-2 prostanoids
• Tubuloglomerular feedback
Metabolic
• Flux of polyols, hexosamines & myo-inositol pathways
• AGE
• PKC
• ROS
Inflammatory
• SAA
• Proinflammatory cytokines
• TNF family
• Chemokines
• Adhesion molecules
Fibrotic
• TGF-β
• Collagen
• CTGF
Activation of cell signaling
TGF-β
Smad
MAPK
JAK/STAT
Transcription factors
NF-κB
Epigenetic changes
DNA methylation
Histone modification
Chromatin remodeling
Noncoding RNA
DKD phenotype
Glomerular hypertrophy
Mesangial expansion
Glomerular sclerosis
Tubulointerstitial inflammation and fibrosis
Kidney fibrosis

Fig. 45.5 **Different pathways and networks involved in the initiation and progression of diabetic kidney disease.** *AGE*, Advanced glycation end product; *CTGF*, connective tissue growth factor; *JAK-STAT*, Janus kinase/signal transducer and activator of transcription; *MAPK*, mitogen-activated protein kinase; *NF-κB*, nuclear factor kappa-light-chain-enhancer of activated B cells; *PKC*, protein kinase C; *RAS*, renin-angiotensin system; *ROS*, reactive oxygen species; *SAA*, serum amyloid A; *TGF-β*, transforming growth factor β; *TNF*, tumor necrosis factor. (From Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol.* 2017; 12:2032–2045.)

C-C motif chemokine 5, are upregulated and contribute to inflammatory cell infiltration and consequent cellular injury in DKD.[164,177-179]

BOX 45.6 PATHOPHYSIOLOGY OF DIABETIC KIDNEY DISEASE

- Metabolic, hemodynamic, inflammatory, fibrotic, and epigenetic changes in the diabetic milieu lead to structural and functional abnormalities of diabetic kidney disease (DKD).
- Increased glucose and sodium chloride uptake in the proximal tubule via sodium-glucose cotransporter 2 reduces solute delivery to the macula densa, resulting in decreased tubuloglomerular feedback and glomerular hyperfiltration.
- Glomerular hyperfiltration is the earliest hemodynamic change in DKD.
- Mechanotransduction, induced by mechanical stress from glomerular hyperfiltration, is a key contributor to DKD pathogenesis.
- Inflammatory and profibrotic mediators contribute to a variety of structural and functional changes involved in DKD development and progression.

Epigenetic Changes

Epigenetic changes constitute a plausible mechanism underlying damage to the diabetic kidney stemming from metabolic memory following exposure to persistent hyperglycemia. The epigenetic modifications most studied in the context of DKD include DNA methylation, histone modification, noncoding miRNAs, and long noncoding RNAs. DNA methylation may be involved in the pathogenesis and progression of DKD by regulating the expression of many genes impacting glomerular and tubular cell function.[46,180,181] Both in experimental mouse models and in biopsy studies of diabetic patients, histone modifications are associated with inflammation in podocytes and upregulation of profibrotic genes that promote kidney fibrosis. Their silencing decreases

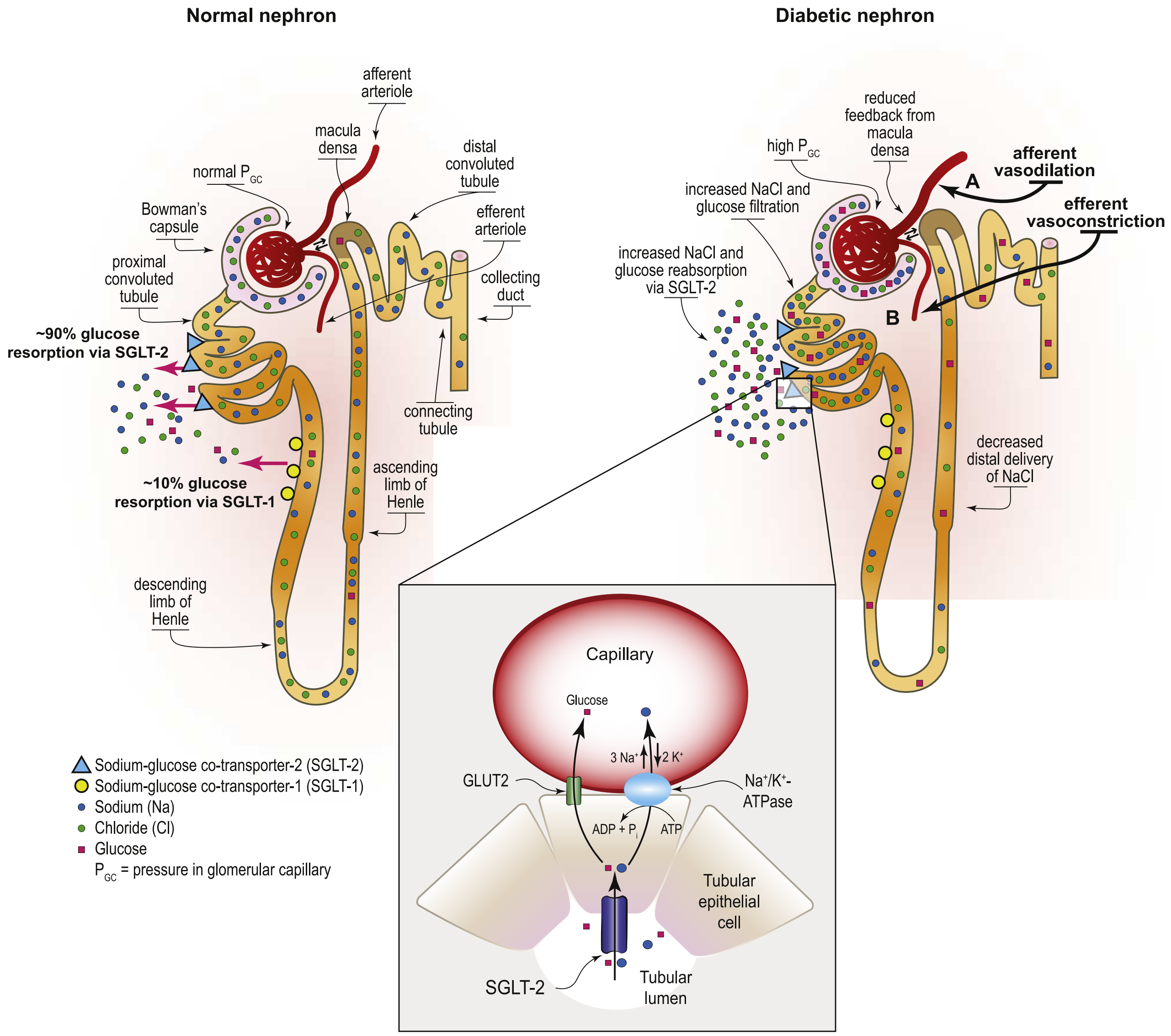

Normal nephron
Diabetic nephron
afferent arteriole
macula densa
distal convoluted tubule
normal P_{GC}
Bowman's capsule
efferent arteriole
proximal convoluted tubule
collecting duct
~90% glucose resorption via SGLT-2
connecting tubule
ascending limb of Henle
~10% glucose resorption via SGLT-1
descending limb of Henle
reduced feedback from macula densa
high P_{GC}
afferent vasodilation
A
efferent vasoconstriction
increased NaCl and glucose filtration
increased NaCl and glucose reabsorption via SGLT-2
B
decreased distal delivery of NaCl
Capillary
Glucose
3 Na^+
2 K^+
GLUT2
Na^+/K^+-ATPase
ADP + P_i
ATP
Tubular epithelial cell
SGLT-2
Tubular lumen
Sodium-glucose co-transporter-2 (SGLT-2)
Sodium-glucose co-transporter-1 (SGLT-1)
Sodium (Na)
Chloride (Cl)
Glucose
P_{GC} = pressure in glomerular capillary

Fig. 45.6 Normal and diabetic nephron with altered renal hemodynamics. A, Afferent vasodilation is promoted by hyperglycemia, hyperinsulinemia, elevated level of circulating amino acids, COX-2 prostanoids, and reduction of tubule glomerular feedback. Tubuloglomerular feedback is a kidney intrinsic autoregulatory mechanism that helps regulate the rate of glomerular filtration. Because of increased reabsorption of glucose and sodium via sodium-glucose cotransporters in diabetes, sodium chloride delivery to macula densa cells of juxtaglomerular apparatus is decreased, resulting in lower production of adenosin and consequent relative vasodilation of afferent arteriolar. **B,** Efferent vasoconstriction is promoted by high local angiotensin II level, endothelin I, reactive oxygen species, and thromboxane A2. (From Alicic RZ, Neumiller JJ, Johnson EJ, et al. Sodium-glucose cotransporter 2 inhibition and diabetic kidney disease. *Diabetes.* 2019; 68:248–257.)

glomerulosclerosis, inflammation, and podocyte injury.[182,183] TGF-β–mediated upregulation of miR-192 in proximal tubule cells has been linked to fibrosis and lower eGFR in patients with DKD. In mouse models of DKD, miR-192 inhibition reduced fibrosis and proteinuria.[184-187]

KEY POINTS

- Functional and structural abnormalities of diabetic kidney disease are result of a complex, self-perpetuated process driven by metabolic, hemodynamic, inflammatory, fibrotic, and epigenetic changes associated with the diabetic milieu.

TREATMENT

Treatment of DKD is centered around a multifactorial approach to addressing modifiable DKD risk factors. Interventions include glycemic control, use of glucose-lowering therapies with proven kidney protective benefits, blood pressure control with use of angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs), management of cardiovascular risk factors, and nutritional management. Current recommendations related to these treatment approaches will be discussed, as well as a brief review of novel therapies in the developmental pipeline and population health considerations

TABLE 45.3 Key American Diabetes Association Recommendations for the Management of Diabetic Kidney Disease

Treatment

- Optimize glucose control to reduce the risk or slow the progression of chronic kidney disease (CKD). (Level of evidence: **A**)
- For patients with type 2 diabetes mellitus and diabetic kidney disease, consider use of a sodium-glucose cotransporter 2 inhibitor in patients with an estimated glomerular filtration rate (eGFR) ≥30 mL/min/1.73 m^2 and urinary albumin >300 mg/g creatinine. (Level of evidence: **A**)
- In patients with type 2 diabetes and CKD, consider use of sodium-glucose cotransporter 2 inhibitors additionally for cardiovascular risk reduction when eGFR and UACR are ≥30 mL/min/1.73m^2 or >300 mg/g, respectively. (Level of evidence: **A**)
- In patients with CKD who are at increased risk for cardiovascular events, use of a glucagon-like peptide 1 receptor agonist reduces renal end point, primarily albuminuria, progression of albuminuria, and cardiovascular events. (Level of evidence: **A**)
- Optimize blood pressure control to reduce the risk or slow the progression of CKD. (Level of evidence: **A**)
- In nonpregnant patients with diabetes and hypertension, either an angiotensin-converting enzyme inhibitor or an angiotensin receptor blocker is recommended for those with modestly elevated urine albumin-to-creatinine ratio (UACR; 30–299 mg/g creatinine) (Level of evidence: **B**) and is strongly recommended for those with UACR ≥300 mg/g creatinine and/or eGFR <60 mL/min/1.73 m^2. (Level of evidence: **A**)

Referral

- Patients should be referred for evaluation by a nephrologist if they have an eGFR <30 mL/min/1.73 m^2. (Level of evidence: **A**)
- Promptly refer to a physician experienced in the care of kidney disease for uncertainty about the etiology of kidney disease, difficult management issues, and rapidly progressing kidney disease. (Level of evidence: **A**)

Recommendations are assigned evidence ratings per the ADA's evidence grading system.
(Modified from American Diabetes Association. Standards of Medical Care in Diabetes—2021. *Diabetes Care*. 2021; 44:S1–S232.)

for improving DKD outcomes. Table 45.3 provides a summary of key recommendations from the American Diabetes Association (ADA) for management of DKD (Box 45.7).[1]

Glycemic Control

Optimization of glycemic control to reduce risk or slow progression of DKD is considered the cornerstone of management.[1,91] Targeting lower HbA1c values through intensification of glucose-lowering therapy has shown benefit in reducing risk for development of diabetes complications, including DKD.[66,188-190] However, achieving therapeutic targets in the DKD population is challenged by reduced capacity for gluconeogenesis by the kidney and impaired clearance of glucose-lowering therapies, thus placing DKD patients at high risk for hypoglycemia or other adverse drug events.[88,133] The Kidney Disease: Improving Global Outcomes (KDIGO) 2020 Clinical Practice Guideline for Diabetes Management in CKD recommends a combination of metformin plus an SGLT-2 inhibitor as first-line pharmacotherapy in most patients with T2DM and CKD with an eGFR over 30 mL/min/1.73 m^2.[91] The addition of additional glucose-lowering agents, as needed, to achieve patient-specific glycemic goals is recommended, with a preference given to a long-acting agent from the GLP1 receptor agonist class.[91] Recommendations for use of agents from the SGLT-2 inhibitor and GLP1 receptor agonist classes are based on recent cardiovascular and kidney disease outcome trials demonstrating benefit with these agents.

Glycemic Monitoring. HbA1c is the recommended biomarker for long-term monitoring of glycemia in DKD.[1,91,133] The HbA1c value can be falsely decreased by factors that reduce the erythrocyte lifespan, such as anemia, red blood cell transfusion, and use of erythrocyte-stimulating agents or iron replacement, particularly in patients undergoing hemodialysis.[87,191,192] Conversely, HbA1c may be falsely increased by metabolic acidosis, carbamylation, and advanced glycation end product formation, which are common in advanced DKD stages. Nonetheless, HbA1c remains the biomarker of choice, as alternative methods such as fructosamine or glycated albumin do not have more accurate relationships to glycemia compared with HbA1c testing. Self-monitoring of blood glucose (SMBG) or continuous glucose monitoring (CGM) is additionally recommended to inform daily treatment decisions, especially in advanced DKD stages. SMBG and CGM are not biased by CKD and can be effective tools to validate HbA1c findings.[91] In patients for whom the reliability of the HbA1c is uncertain (advanced DKD stages or hemodialysis), CGM can be used to generate a glucose management indicator, a calculated measure of average glucose control, generated from CGM data to inform treatment decisions.[193]

Glycemic Targets. The 2020 KDIGO guideline for diabetes management in CKD recommends an individualized HbA1c target based on patient- and medication-specific factors target ranging from less than 6.5% to less than 8.0% in patients not treated with dialysis.[1,91] Targeting an HbA1c less than 6.5% may be appropriate in younger people with earlier DKD for whom prevention of additional complications is a major consideration, while a more liberal HbA1c target (e.g., <8.0%) may be more appropriate in those with longstanding diabetes and a high comorbidity burden or a high risk of hypoglycemia or hypoglycemia unawareness.[91] Table 45.4 provides a summary of factors to guide HbA1c target selection in patients with DKD.[1,91] When HbA1c may not be a reliable indicator of glycemic control in advanced DKD and use of CGM is an option, CGM metrics such as time in range and time spent in hypoglycemia (defined as time spent <70 mg/dL) may be helpful in guiding daily care and treatment modification.[91]

Glucose-Lowering Agents

The SGLT-2 inhibitor and GLP1 receptor agonist classes have demonstrated kidney- and heart-protective benefits extending beyond their impact on glycemia in large-scale clinical trials.

Sodium-Glucose Cotransporter 2 Inhibitors. The ADA and KDIGO clinical practice guidelines recommend addition of an SGLT-2 inhibitor in patients with CKD independent of the current HbA1c or individualized HbA1c target for patients with an eGFR greater than or equal to 30 mL/min/1.73 m^2.[1,91] The initial signal for a kidney-specific benefit of this medication class came from secondary outcomes reported in the cardiovascular outcome trials (CVOTs) with canagliflozin, dapagliflozin, and empagliflozin. All showed reduced risk of DKD onset and progression with the SGLT-2 inhibitors.[194-197] These findings prompted study of SGLT-2 inhibitors in patients with T2DM and CKD to determine the effects on kidney disease endpoints as primary outcomes (Table 45.5).[198,199] The Canagliflozin and Renal

BOX 45.7 TREATMENT OF DIABETIC KIDNEY DISEASE

Glycemic Control

- First-line therapy for patients with type 2 diabetes mellitus and chronic kidney disease with an estimated glomerular filtration rate >30 mL/min/1.73 m^2 is metformin plus a sodium-glucose cotransporter 2 inhibitor. For patients requiring additional glucose-lowering therapy, a long-acting glucagon-like peptide-1 receptor agonist is preferred.
- Glycated hemoglobin (HbA1c) is the recommended metric in diabetic kidney disease (DKD) for assessment of long-term glycemic control; however, limitations of HbA1c in this population reduces the reliability of this measurement.
- Self-monitoring of blood glucose and continuous glucose monitoring are recommended for glucose monitoring to inform daily self-management and prevent hypoglycemia.
- An individualized HbA1c target in the range of <6.5%–7.0% to <8.0% is recommended.

Blood Pressure Control

- A blood pressure target of <140/90 mm Hg is recommended for patients without albuminuria, while a more intensive target of <130/80 mm Hg is recommended for those with albuminuria.
- Angiotensin-converting enzyme (ACE) inhibitor/angiotensin receptor blocker (ARB) therapy at the highest approved and tolerated dose is recommended for patients with diabetes and albuminuria.
- Combination ACE inhibitor plus ARB therapy is not recommended, because of risks of hyperkalemia and acute kidney injury.

Cardiovascular Risk Reduction

- In addition to optimized blood pressure control, lifestyle interventions (nutrition management and physical activity), lipid management with a statin, smoking cessation, and guideline-directed antiplatelet therapy are recommended.

Nutritional Management

- Patients with DKD should be encouraged to adopt a generally healthy diet that stresses consumption of whole foods such as vegetables, fruits, whole grains, fiber, legumes, plant-based proteins, unsaturated fats, and nuts.
- Patients with DKD are recommended to maintain daily protein intake of 0.8 g/kg per day, with a higher daily protein intake of 1.0–1.2 g/kg per day recommended for those treated by dialysis.
- To assist with blood pressure management, patients with DKD are recommended to limit daily sodium consumption to <2 g/day.

Events in Diabetes with Established Nephropathy Clinical Evaluation enrolled participants with eGFR of 30 or higher and less than 90 mL/min/1.73 m^2 and UACR of 300 or higher and less than 5000 mg/g.[198] After a mean follow-up period of 2.6 years, the composite outcome inclusive of ESKD, doubling of serum creatinine, or death from kidney disease causes was lower in the intervention group (hazard ratio [HR]: 0.68; 95% confidence interval [CI]: 0.5–-0.81; $P < .001$). When analyzed alone, the relative risk of ESKD was lowered by 32% (HR: 0.68; 95% CI: 0.54–0.86; $P = .002$). Canagliflozin treatment was also shown to improve the cardiovascular composite outcomes of cardiovascular death or hospitalization for heart failure (HR: 0.69, 95% CI: 0.57–0.83; $P < .01$), and cardiovascular death, myocardial infarction, or stroke (HR: 0.80; 95% CI: 0.67–0.95; $P = .01$). The risk of hospitalization for heart failure was also reduced with intervention when compared with placebo (HR: 0.61; 95% CI: 0.47–0.80; $P < .001$).[198] Subsequently, the Dapagliflozin and Prevention of Adverse Outcomes in Chronic Kidney Disease enrolled patients with or without T2DM and with eGFR of 25 or more and less than 75 mL/min/1.73 m2 and UACR of 200 or more and less than 5000 mg/g.[199] The primary composite outcome of a sustained decline in eGFR of greater than 50%, ESKD, or death from kidney disease or cardiovascular causes was significantly reduced with dapagliflozin treatment (HR: 0.61; 95% CI: 0.51–0.72; $P < .001$).[199] Importantly, in both studies the benefit of SGLT-2 inhibitor treatment was on top of the standard of care with an ACE inhibitor or an ARB. The Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure (EMPEROR-Reduced) study enrolled diabetic and nondiabetic patients with heart failure and reduced systolic function. The risk of a primary outcome event was significantly reduced (HR for cardiovascular death or hospitalization for heart failure: 0.75; 95% CI: 0.65–0.86; $P < .001$) over a median of 16 months. The EMPEROR-Reduced study confirmed the beneficial effects of SGLT-2 inhibition on kidney function as the major secondary outcome. Independent of diabetes status, eGFR decline was significantly slower in participants treated with empagliflozin (–0.55 vs. –2.28 mL/min/1.73 m^2 per year in the intervention and placebo groups, respectively; $P < .001$). Empagliflozin-treated participants also had a 50% reduction in risk of kidney failure or requiring KRT (HR: 0.50; 95% CI: 0.32–0.77).[200] Although current guidelines recommend use of SGLT-2 inhibitors in appropriate patients to mitigate kidney risk, potential side effects and safety concerns with this class of medication must also be considered. Potential notable harms that have been associated with SGLT-2 inhibitor therapy include DKA and genital mycotic infections.[91] The

TABLE 45.4 Factors That May Guide Decisions on Individualized Glycated Hemoglobin Targets in Diabetic Kidney Disease

Factor/Consideration	Supports More Stringent HbA1c Target (e.g., <6.5% or 7.0%)	Supports Less Stringent HbA1c Target (e.g., <8.0%)
Chronic kidney disease (CKD) severity	Early-stage CKD (e.g., CKD G1)	Advanced-stage CKD (e.g., CKD G5)
Macrovascular complications	Absent/minor	Present/severe
Comorbidity burden	Low	High
Life expectancy	Long	Short
Hypoglycemia risk	Low	High
Hypoglycemia awareness	Present	Impaired
Resources/support	Available	Scarce
Patient motivation/preference for treatment intensity	Highly motivated	Low motivation

(Modified from Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2020 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int.* 2020; 98:S1–.S115; American Diabetes Association. Standards of Medical Care in Diabetes—2020. *Diabetes Care*. 2020; 43:S1–S212; Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2020 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int.* 2020; 98:S1–S115.)

risk of DKA with SGLT-2 inhibitors is rare in patients with T2DM but educating patients and caregivers about reporting symptoms suggestive of DKA is warranted. Likewise, counseling patients about appropriate self-care and hygiene, such as daily bathing, can reduce the risk of genital mycotic infections.[91] Because an increased risk of lower-limb amputations was observed in one clinical trial with canagliflozin, encouraging patients to engage in routine preventative foot self-care and maintain adequate hydration may reduce the risk for lower extremity complications.[91]

A principal mechanism by which SGLT-2 inhibitors convey kidney protection is believed to be reduced glomerular hyperfiltration via restoration of solute delivery to the macula densa and reversal of vasodilation of the afferent arteriole by increased tubuloglomerular feedback (Fig. 45.7).[156] SGLT-2 inhibitors may also improve DKD risk factors such as hyperglycemia, obesity, and hypertension.[156] These agents have other pleiotropic protective effects, including: increased erythropoietin, red blood cell mass, and oxygen delivery; mild ketosis providing β-hydroxybutyrate as a myocardial fuel; reduction of inflammation and oxidative stress; and inhibition of sodium-hydrogen exchanger activity in the kidney and heart.[201]

Glucagon-Like Peptide-1 Receptor Agonists. GLP1 receptor agonists are effective glucose-lowering agents even in advanced stages of DKD. KDIGO preferentially recommends the addition of a long-acting GLP1 receptor agonist in patients with T2DM and CKD who need additional glucose lowering if first-line metformin plus an SGLT-2 inhibitor is insufficient to meet individualized glycemic goals.[91] To date, kidney disease outcomes data with agents from the GLP1 receptor agonist class are limited to secondary outcomes from the CVOTs and a glycemic control trial showing benefits on albuminuria onset and progression[202-204] and eGFR decline.[205] A metaanalysis of GLP1 receptor agonists evaluated a composite kidney disease outcome inclusive of development of macroalbuminuria, decline in eGFR or increase in creatinine, progression to kidney failure, or death owing to kidney disease.[206] GLP1 receptor agonist treatment was associated with a 17% relative risk reduction in the composite kidney outcome (HR: 0.83; 95% CI: 0.78–0.89; $P < .0001$), which was primarily attributed to a reduction in albuminuria.[206] Although current kidney disease outcome data is limited to secondary outcomes, the Effect of Semaglutide Versus Placebo on the Progression of Renal Impairment in Subjects With Type 2 Diabetes and Chronic Kidney Disease (FLOW) study is prospectively testing the impact of injectable semaglutide on kidney disease as primary outcomes in patients with DKD.[207] The FLOW trial will enroll approximately 3200 participants with T2DM with eGFR of 25 or more to less than75 mL/min/1.73 m^2 and UACR of 300 or more to less than 5000 mg/g.[207] The primary outcome is a composite of kidney disease endpoints (≥50% eGFR decline, kidney failure, and death from kidney disease or cardiovascular disease). This is the first trial to include kidney disease endpoints as the primary outcome with an agent from the GLP-1 receptor agonist class. In terms of adverse events, GLP1 receptor agonists are commonly associated with gastrointestinal (GI) side effects (e.g., nausea, vomiting, diarrhea) and injection site reactions, which can both vary by agent used.[91] GI side effects tend to abate with continued use; however, patients should be encouraged to contact their healthcare provider if severe GI side effects occur, and to maintain hydration. GLP1 receptor agonists are contraindicated and should not be initiated in patients with a history of medullary thyroid cancer, multiple endocrine neoplasia 2, or a history of acute pancreatitis.[91]

Experimental data suggest several mechanisms by which GLP1 receptor agonists may provide benefit in DKD. GLP1 receptor agonists may reduce both systemic and local inflammation as plausible mechanisms to confer kidney protective effects.[208] In preclinical models of diabetic rats and mice, GLP1 receptor agonists reduce pathological changes of inflammation and fibrosis, along with decreased proinflammatory and profibrotic mediators in the glomerular and tubulointerstitial compartments.[208]

Blood Pressure Control

Antihypertensive therapy reduces risks for of DKD and related cardiovascular events.[209] For those at lower risk for cardiovascular disease (10-year atherosclerotic cardiovascular disease [ASCVD] risk of <15%) a target blood pressure of less than 140/90 mm Hg is recommended by the ADA for patients with diabetes and hypertension.[1] For patients at higher risk (existing ASCVD or 10-year ASCVD risk ≥15%), a target of less than 130/80 mm Hg is appropriate if it can be achieved safely. These targets are in line with those recommended by KDIGO.[210] The preferred antihypertensive classes include ACE inhibitors, ARBs, dihydropyridine calcium channel blockers, and thiazide-like diuretics. In the presence of albuminuria, an ACE inhibitor or an ARB is specifically recommended.

TABLE 45.5 Summary of Completed Sodium-Glucose Cotransporter 2 Inhibitor Trials With Primary Outcomes for Kidney Disease

Agent	Canagliflozin	Dapagliflozin
Study	CREDENCE (n = 4,401)	DAPA-CKD (n = 4,304)
Median follow-up (years)	2.6	2.4
Kidney-related enrollment criteria†	• eGFR ≥30 to <90 mL/min/1.73 m^2 • UACR: ≥300 to <5000 mg/g	• eGFR ≥25 to <75 mL/min/1.73 m^2 • UACR: ≥200 to 5000 mg/g
Mean baseline eGFR	56 mL/min/1.73 m^2	43 mL/min/1.73 m^2
Median baseline UACR	927 mg/g	949 mg/g
Kidney outcome(s)	Primary Outcome • ESKD (dialysis, transplantation, or sustained eGFR <15 mL/min/1.73 m^2), doubling of SCr, or death from renal causes: HR: 0.70 (0.59–0.82)	Primary Outcome • ≥50% decrease in eGFR, ESKD, or death from renal or cardiovascular causes: HR: 0.61 (0.51–0.72)

eGFR, Estimated glomerular filtration rate; *ESKD*, end-stage kidney disease; *HR*: hazard ratio; *SCr*, serum creatinine; *UACR*, urinary albumin-to-creatinine ratio.

(Modified from Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med.* 2019;380:2295–2306; Heerspink HJL, Stefansson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med* 2020;383:1436–1446.)

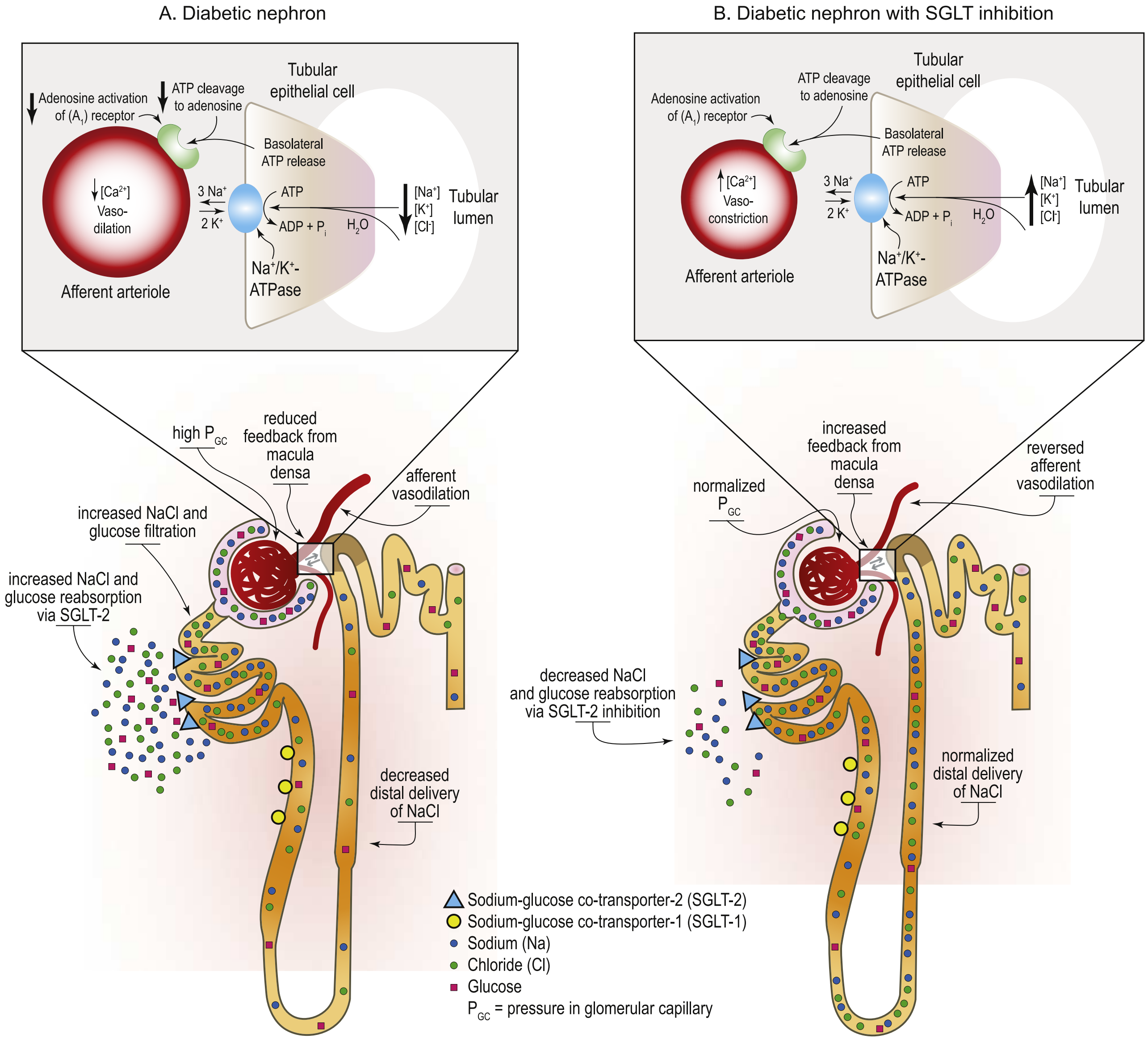

A. Diabetic nephron
Adenosine activation of (A1) receptor
ATP cleavage to adenosine
Tubular epithelial cell
Basolateral ATP release
[Ca2+]
Vaso-dilation
3 Na+
2 K+
ATP
ADP + Pi
H2O
[Na+]
[K+]
[Cl-]
Tubular lumen
Na+/K+-ATPase
Afferent arteriole
high PGC
reduced feedback from macula densa
afferent vasodilation
increased NaCl and glucose filtration
increased NaCl and glucose reabsorption via SGLT-2
decreased distal delivery of NaCl
B. Diabetic nephron with SGLT inhibition
Adenosine activation of (A1) receptor
ATP cleavage to adenosine
Tubular epithelial cell
Basolateral ATP release
[Ca2+]
Vaso-constriction
3 Na+
2 K+
ATP
ADP + Pi
H2O
[Na+]
[K+]
[Cl-]
Tubular lumen
Na+/K+-ATPase
Afferent arteriole
normalized PGC
increased feedback from macula densa
reversed afferent vasodilation
decreased NaCl and glucose reabsorption via SGLT-2 inhibition
normalized distal delivery of NaCl
Sodium-glucose co-transporter-2 (SGLT-2)
Sodium-glucose co-transporter-1 (SGLT-1)
Sodium (Na)
Chloride (Cl)
Glucose
PGC = pressure in glomerular capillary

Fig. 45.7 **Effects of diabetes and SGLT2 inhibition on nephron hemodynamics. A:** Increased reabsorption of glucose by SGLT2 in the proximal convoluted tubule decreases delivery of solutes to the macula densa. The resulting decrease in ATP release from the basolateral membrane of tubular epithelial cells reduces production of adenosine and produces a vasodilatation of the afferent arteriole. **B:** SGLT2 inhibitors restore solute delivery to the macula densa with resulting adenosine activation and reversal of vasodilation of the afferent arteriole. (From Alicic RZ, Neumiller JJ, Johnson EJ, et al. Sodium-glucose cotransporter 2 inhibition and diabetic kidney disease. *Diabetes*. 2019; 68:248–257.)

Renin-Angiotensin-Aldosterone System Inhibitors. In patients with DKD and T1DM or T2DM, use of an ACE inhibitor or ARB reduces risk of progression to kidney failure.[211-213] Clinical practice guidelines consistently recommend treatment with an ACE inhibitor or an ARB in patients with diabetes, hypertension, and albuminuria at the highest tolerated dose approved for use by regulatory agencies.[1,91] Combination ACE inhibitor plus ARB therapy is not recommended, as there are no additive benefits on cardiovascular or kidney disease outcomes, and the combination was associated with higher risks for hyperkalemia and AKI.[214,215]

The effect of Finerenone, the nonsteroidal mineralcorticoid inhibitor, on kidney and cardiovascular outcomes in patients with moderate to severe DKD was studied in two large trails. The Finerenone in Reducing Cardiovascular Mortality and Morbidity in Diabetic Kidney Disease (FIGARO-DKD) and companion study the Effect of Finerenone on Chronic Kidney Disease Outcomes in Type 2 Diabetes (FIDELIO-DKD) trial enrolled participants with eGFR >25 to <75 ml/min/1.73 m^2 and a UACR >300 to <5000 mg/g or eGFR >25 to <60 ml/min/1.73 m^2 and a UACR >30 to <300 mg/g with retinopathy.[216,217] Compared with placebo, finerenone reduced the risk of the composite primary kidney disease outcome (kidney failure, a sustained decrease in eGFR of >40%, or death from kidney disease) by 18% (HR: 0.82; 95% CI: 0.73–0.93; *P* = .001). The key secondary composite outcome (death from cardiovascular causes, nonfatal myocardial infarction, nonfatal stroke, or hospitalization for heart failure) was also reduced (HR: 0.86; 95% CI: 0.75–0.99; *P* = .03). Participants in the finerenone and placebo groups experienced similar rates of kidney injury–related events. The rate of hyperkalemia-related events was twice as high in those receiving finerenone versus placebo (18.3% vs. 9.0%, respectively), indicating that potassium monitoring will be important in patients treated with this agent. In FIGARO-DKD, finerenone treatment was positive for the primary cardiovascular composite outcome inclusive of: death from cardiovascular causes, nonfatal myocardial infarction, nonfatal stroke, and hospitalization for heart failure (HR: 0.87; 95% CI: 0.76-0.98; P=0.03). Event rates for hyperkalemia were overall low in the trial (2.3% versus 0.9% in the finerenone and placebo groups, respectively).[216] FIGARO-DKD and FIDELIO-DKD outcomes supported FDA approval of finerenone for risk reduction of sustained eGFR decline, reaching end stage kidney disease, cardiovascular death, non-fatal myocardial infarction, and hospitalization for heart failure in adults with chronic kidney disease associated with T2DM.[218] Clinical practice guidelines from major professional organizations are expected to be updated to recommend finerenone for kidney and heart protection in T2DM.[216]

Cardiovascular Risk Reduction

In patients with diabetes, the presence of CKD accounts for most of the increased cardiovascular risk.[13,219] Higher albuminuria and lower eGFR are independently and additively associated with an increased risk for cardiovascular events, mortality, and all-cause death.[88,220] In the UKPDS study, the overall death rate was nearly 20% per year in participants who were either receiving KRT or had a blood creatinine greater than 2 mg/dL.[221]

The American College of Cardiology recommends that adults with T2DM with ASCVD, heart failure, or DKD receive treatment with an SGLT-2 inhibitor or GLP1 receptor agonist with proven cardiovascular benefit.[222] In addition, they and KDIGO recommend optimization of guideline-directed medical therapy inclusive of lifestyle intensification (nutrition management and physical activity), blood pressure control, lipid management with a statin, glycemic control, smoking cessation, and antiplatelet therapy.[91,222]

Nutrition Management

Key considerations include adopting a generally healthy and balanced diet, maintaining appropriate protein intake, and avoiding excessive sodium intake. Like other patients with diabetes, those with DKD benefit from individualized diets that stress consumption of whole foods such as vegetables, fruits, whole grains, fiber, legumes, plant-based proteins, unsaturated fats, and nuts.[1,91] Minimal consumption of processed meats, refined carbohydrates, and sweetened beverages is also encouraged. While dietary protein restriction has been hypothesized to slow DKD progression,[223] protein restriction combined with carbohydrate limitation can contribute to overly restrictive diets and weight loss.[91] KDIGO recommends following the World Health Organization recommendation of a protein intake of 0.8 g/kg per day, with an emphasis on plant-based sources for patients with DKD.[91] For those treated by dialysis, a higher daily protein intake of 1.0 to 1.2 g/kg per day is recommended to compensate for amino acid losses via the dialysate and a catabolic state. To address dietary approaches to blood pressure control, patients with DKD are recommended to limit sodium intake to less than 2 g per day.[91] Evidence suggests that adherence to nutritional recommendations is low overall in patients with CKD.[224] Therefore, engaging patients in shared decision-making for an individualized nutrition plan is critical, as is providing nutrition education and promoting healthy behavior change.

Additional Interventions and Management Approaches

Multiple novel therapeutics for DKD are under study. Given the gaps in care, population health initiatives aiming to improve awareness, identification, and management of DKD are essential.

Novel Therapeutics. Despite recent therapeutic progress, new DKD treatments are needed to reduce residual risk by targeting disease pathways that may not be addressed by available approaches. Promising targets and drug candidates under study include further study of finerenone in earlier stage DKD, endothelin A receptor antagonism (atrasentan), and antiinflammatory signaling pathways (baracitinib, selonsertib) (Table 45.6).[225]

Health Delivery. Given the serious health consequences of DKD, early identification and initiation of interventions to prevent disease onset and progression are imperative. Even though an established standard of care has been in place for nearly two decades, patients with DKD often receive suboptimal care, reflected underuse of ACE inhibitors/ARBs and overuse of potentially nephrotoxic agents (nonsteroidal antiinflammatory drugs and proton pump inhibitors).[226] Healthcare delivery approaches for increased DKD awareness and initiation of evidence-based therapies are critical to improving clinical outcomes. Interventions at both the patient and health-systems level are recommended for patient-centered care through a multidisciplinary approach (Table 45.7).[91]

TABLE 45.6 Novel Therapeutic Targets for the Treatment of Diabetic Kidney Disease

Medication Class	Mechanism/Target	Candidate Drug(s)
Nonsteroidal mineralocorticoid receptor (MR) antagonist	Antagonism potentially leads to: • Reversal of ischemia-related kidney damage • Reversal of oxidative stress • Correction of impaired nitric oxide synthesis • Attenuation of macrophage-induced kidney injury	Finerenone
Endothelin A receptor antagonist	Antagonism reduces proteinuria, kidney fibrosis, and inflammation	Astrasentan
Apoptosis signal-regulating kinase 1 inhibitor	Antagonism reduces epithelial cell injury and kidney fibrosis	Selonsertib
Janus kinase 1/2 inhibitor	Inhibition prevents activation of STATs and reduces IgG, IgM, IgA, and C-reactive protein	Baricitinib

Ig, Immunoglobulin; *STAT*, signal transducer and activator of transcription.

(Modified from Hartman RE, Rao PSS, Churchwell MD, et al. Novel therapeutic agents for the treatment of diabetic kidney disease. *Expert Opin Investig Drugs*. 2020;29:1277–1293.)

TABLE 45.7 Management Approach Recommendations for Patients With Diabetes and Chronic Kidney Disease

- Structured self-management educational programs are recommended in people with diabetes and chronic kidney disease.
- Health systems should consider implementing structured self-management programs that consider local context, cultures, and resource availability.
- Implementation of team-based, integrated care that is focused on risk evaluation and patient empowerment.
- Team-based care should incorporate care delivered by physicians and nonphysician providers (e.g., nurses, dietitians, pharmacists, healthcare assistants, community workers, and peer supporters).

(Modified from Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2020 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int.* 2020; 98:S1–S115.)

KEY POINTS

- For patients with type 2 diabetes mellitus and chronic kidney disease, the recommended glycated hemoglobin target is individualized in the range of <6.5%–7.0% to <8.0%. In patients with an estimated glomerular filtration rate >30 mL/min/1.73 m^2, the first-line therapy is metformin plus a sodium-glucose cotransporter 2 inhibitor. In patients with albuminuria, a more intensive blood pressure target of <130/80 mm Hg is recommended, and use of angiotensin-converting enzyme inhibitor/angiotensin receptor blocker therapy is recommended at the highest approved and tolerated dose.

SUMMARY AND FUTURE DIRECTIONS

In contrast to other diabetic complications, the burden of DKD has steadily increased worldwide. Despite available therapies, the residual risk of DKD progression to death and kidney failure is unacceptably high. Notably, DKD awareness among patients and providers is low, and adherence to recommended standard-of-care interventions is poor. Recent therapeutic modalities shown to improve DKD outcomes, such as SGLT-2 inhibitors, GLP1 receptor agonists, and finerenone, show great potential to reduce the enormous human and societal burden of DKD. Together with efforts to implement harmonized multidisciplinary care, quality of life and survival in DKD can be improved, while also striving to develop and deliver even more effective approaches for this high-risk population.

REFERENCES

1. American Diabetes Association. Standards of Medical Care in Diabetes—2021. *Diabetes Care*. 2021;43:S1–S232.
2. Parving HH, Lewis JB, Ravid M, et al. Prevalence and risk factors for microalbuminuria in a referred cohort of type II diabetic patients: a global perspective. *Kidney Int*. 2006;69:2057–2063.
3. Retnakaran R, Cull CA, Thorne KI, et al. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. *Diabetes*. 2006;55:1832–1839.
4. Cameron JS. The discovery of diabetic nephropathy: from small print to centre stage. *J Nephrol*. 2006;19:S75–S87.
5. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: global dimension and perspectives. *Lancet*. 2013;382:260–272.
6. United States Renal Data System. *2019 USRDS Annual Data Report: Epidemiology of kidney disease in the United States*. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2019.
7. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019;157:107843.
8. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095–2128.
9. Gregg EW, Li Y, Wang J, et al. Changes in diabetes-related complications in the United States, 1990-2010. *N Engl J Med*. 2014;370:1514–1523.
10. Harding JL, Pavkov ME, Magliano DJ, et al. Global trends in diabetes complications: a review of current evidence. *Diabetologia*. 2019;62:3–16.
11. Cea Soriano L, Johansson S, Stefansson B, et al. Cardiovascular events and all-cause mortality in a cohort of 57,946 patients with type 2 diabetes: associations with renal function and cardiovascular risk factors. *Cardiovasc Diabetol*. 2015;14:38.
12. Scirica BM, Mosenzon O, Bhatt DL, et al. Cardiovascular outcomes according to urinary albumin and kidney disease in patients with type 2 diabetes at high cardiovascular risk: observations from the SAVOR-TIMI 53 trial. *JAMA Cardiol*. 2018;3:155–163.
13. Afkarian M, Sachs MC, Kestenbaum B, et al. Kidney disease and increased mortality risk in type 2 diabetes. *J Am Soc Nephrol*. 2013;24:302–308.
14. Daratha KB, Short RA, Corbett CF, et al. Risks of subsequent hospitalization and death in patients with kidney disease. *Clin J Am Soc Nephrol*. 2012;7:409–416.
15. Xie Y, Bowe B, Mokdad AH, et al. Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. *Kidney Int*. 2018;94:567–581.
16. Crews DC, Bello AK, Saadi G, et al. World Kidney Day Steering Committee. Burden, access, and disparities in kidney disease. *Kidney Int*. 2019;95:242–248.
17. Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease—A systematic review and meta-analysis. *PLoS One*. 2016;11:e0158765.
18. Liyanage T, Ninomiya T, Jha V, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. *Lancet*. 2015;385:1975–1982.
19. de Boer IH, Rue TC, Hall YN, et al. Temporal trends in the prevalence of diabetic kidney disease in the United States. *JAMA*. 2011;305:2532–2539.
20. Bowe B, Xie Y, Li T, et al. Changes in the US burden of chronic kidney disease from 2002 to 2016: an analysis of the Global Burden of Disease study. *JAMA Netw Open*. 2018;1:e184412.
21. Koye DN, Shaw JE, Reid CM, et al. Incidence of chronic kidney disease among people with diabetes: a systematic review of observational studies. *Diabet Med*. 2017;34:887–901.
22. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol*. 2017;12:2032–2045.
23. Boyle JP, Thompson TJ, Gregg EW, et al. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr*. 2010;8:29.
24. Scheven L, Halbesma N, de Jong PE, et al. Predictors of progression in albuminuria in the general population: results from the PREVEND cohort. *PLoS One*. 2013;8:e61119.
25. Sibley SD, Thomas W, de Boer I, et al. Gender and elevated albumin excretion in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) cohort: role of central obesity. *Am J Kidney Dis*. 2006;47:223–232.
26. Hanai K, Babazono T, Yoshida N, et al. Gender differences in the association between HDL cholesterol and the progression of diabetic kidney disease in type 2 diabetic patients. *Nephrol Dial Transplant*. 2012;27:1070–1075.
27. Narayan KM, Boyle JP, Thompson TJ, et al. Lifetime risk for diabetes mellitus in the United States. *JAMA*. 2003;290:1884–1890.
28. Cobo G, Hecking M, Port FK, et al. Sex and gender differences in chronic kidney disease: progression to end-stage renal disease and haemodialysis. *Clin Sci (Lond)*. 2016;130:1147–1163.
29. Earle KA, Ng L, White S, et al. Sex differences in vascular stiffness and relationship to the risk of renal functional decline in patients with type 2 diabetes. *Diab Vasc Dis Res*. 2017;14:304–309.

30. Maric C. Sex, diabetes and the kidney. *Am J Physiol Renal Physiol*. 2009;296:F680–F688.
31. Dreyer G, Hull S, Mathur R, et al. Progression of chronic kidney disease in a multi-ethnic community cohort of patients with diabetes mellitus. *Diabet Med*. 2013;30:956–963.
32. Chowdhury TA, Lasker SS. Complications and cardiovascular risk factors in South Asians and Europeans with early-onset type 2 diabetes. *QJM*. 2002;95:241–246.
33. Derose SF, Rutkowski MP, Crooks PW, et al. Racial differences in estimated GFR decline, ESRD, and mortality in an integrated health system. *Am J Kidney Dis*. 2013;62:236–244.
34. Tarver-Carr ME, Powe NR, Eberhardt MS, et al. Excess risk of chronic kidney disease among African-American versus white subjects in the United States: a population-based study of potential explanatory factors. *J Am Soc Nephrol*. 2002;13:2363–2370.
35. Lanting LC, Joung IMA, Mackenbach JP, et al. Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. *Diabetes Care*. 2005;28:2280–2288.
36. Raymond NT, O'Hare JP, Bellary S, et al. Comparative risk of microalbuminuria and proteinuria in UK residents of south Asian and white European ethnic background with type 2 diabetes: a report from UKADS. *Curr Med Res Opin*. 2011;27:47–55.
37. Kenealy T, Elley CR, Collins JF, et al. Increased prevalence of albuminuria among non-European peoples with type 2 diabetes. *Nephrol Dial Transplant*. 2012;27:1840–1846.
38. Bruce R, Williams L, Cundy T. Rates of progression to end stage renal failure in nephropathy secondary to type 1 and type 2 diabetes mellitus. *Aust N Z J Med*. 1994;24:390–395.
39. International Diabetes Federation. *IDF Diabetes Atlas update poster*. 6th ed. Brussels: International Diabetes Federation; 2014.
40. Bryson CL, Ross HJ, Boyko EJ, et al. Racial and ethnic variations in albuminuria in the US Third National Health and Nutrition Examination Survey (NHANES III) population: associations with diabetes and level of CKD. *Am J Kidney Dis*. 2006;48:720–726.
41. Seaquist ER, Goetz FC, Rich S, et al. Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med*. 1989;320:1161–1165.
42. Harjutsalo V, Katoh S, Sarti C, et al. Population-based assessment of familial clustering of diabetic nephropathy in type 1 diabetes. *Diabetes*. 2004;53:2449–2454.
43. Quinn M, Angelico MC, Warram JH, et al. Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia*. 1996;39:940–945.
44. Ahlqvist E, van Zuydam NR, Groop LC, et al. The genetics of diabetic complications. *Nat Rev Nephrol*. 2015;11:277–287.
45. Kato M, Natarajan R. Epigenetics and epigenomics in diabetic kidney disease and metabolic memory. *Nat Rev Nephrol*. 2019;15:327–345.
46. Bell CG, Teschendorff AE, Rakyan VK, et al. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics*. 2010;3:33.
47. Sapienza C, Lee J, Powell J, et al. DNA methylation profiling identifies epigenetic differences between diabetes patients with ESRD and diabetes patients without nephropathy. *Epigenetics*. 2011;6:20–28.
48. Bomotti SM, Smith JA, Zagel AL, et al. Epigenetic markers of renal function in African Americans. *Nurs Res Pract*. 2013;2013:687519.
49. Khalil AC, Travert F, Fetita S, et al. Fetal exposure to maternal type 1 diabetes is associated with renal dysfunction at adult age. *Diabetes*. 2010;59:2631–2636.
50. Gautier JF, Porcher R, Khalil CA, et al. Kidney dysfunction in adult offspring exposed in utero to type 1 diabetes is associated with alterations in genome-wide DNA methylation. *PLoS One*. 2015;10:e0134654.
51. Nelson RG, Morgenstern H, Bennett PH. Birth weight and renal disease in Pima Indians with type 2 diabetes mellitus. *Am J Epidemiol*. 1998;148:650–656.
52. Rossing P, Tarnow L, Nielsen FS, et al. Low birth weight. A risk factor for development of diabetic nephropathy? *Diabetes*. 1995;44:1405–1407.
53. Hoy WE, Rees M, Kile E, et al. A new dimension to the Barker hypothesis: low birthweight and susceptibility to renal disease. *Kidney Int*. 1999;56:1072–1077.
54. Lackland DT, Bendall HE, Osmond C, et al. Low birth weights contribute to high rates of early-onset chronic renal failure in the Southeastern United States. *Arch Intern Med*. 2000;160:1472–1476.
55. Fagerudd J, Carol Forsblom C, Kim Pettersson-Fernholm K, et al. Low birth weight does not increase the risk of nephropathy in Finnish type 1 diabetic patients. *Nephrol Dial Transpl*. 2006;21:2159–2165.
56. Macisaac RJ, Ekinci EI, Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. *Am J Kidney Dis*. 2014;63:S39–S62.
57. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. *Clin J Am Soc Nephrol*. 2012;7:401–408.
58. D'Agati VD, Chagnac A, de Vries APJ, et al. Obesity-related glomerulopathy: clinical and pathological characteristics and pathogenesis. *Nat Rev Nephrol*. 2016;12:453–471.
59. Hall JE, do Carmo JM, da Silva AA, et al. Obesity, kidney dysfunction and hypertension: mechanistic links. *Nat Rev Nephrol*. 2019;15:367–385.
60. Helal I, Fick-Brosnahan GM, Reed-Gitomer B, et al. Glomerular hyperfiltration: definitions, mechanisms and clinical implications. *Nat Rev Nephrol*. 2012;8:293–300.
61. Thomas MC, Brownlee M, Susztak K, et al. Diabetic kidney disease. *Nat Rev Dis Primers*. 2015;1:15018.
62. The Diabetes Control and Complications (DCCT) Research Group. Effect of intensive therapy on the development and progression of diabetic nephropathy in the diabetes control and complications trial. *Kidney Int*. 1995;47:1703–1720.
63. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA*. 2003;290:2159–2167.
64. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352:837–853.
65. Nathan DM, DCCT/EDIC Research Group. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*. 2014;37:9–16.
66. de Boer IH, DCCT/EDIC Research Group. Kidney disease and related findings in the Diabetes Control and Complications trial/Epidemiology of Diabetes Interventions and Complications study. *Diabetes Care*. 2014;37:24–30.
67. UK Prospective Diabetes Study (UKPDS). XI: Biochemical risk factors in type 2 diabetic patients at diagnosis compared with age-matched normal subjects. *Diabet Med*. 1994;11:534–544.
68. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577–1589.
69. Bilous R. Microvascular disease: What does the UKPDS tell us about diabetic nephropathy? *Diabet Med*. 2008;25:25–29.
70. Chen Z, Miao F, Paterson AD, et al. Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc Natl Acad Sci U S A*. 2016;113:E3002–E3011.
71. Adler AI, Stratton IM, Neil HA, et al. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. *BMJ*. 2000;321:412–419.
72. Bakris GL, Weir MR, Shanifar S, et al. Effects of blood pressure level on progression of diabetic nephropathy: results from the RENAAL study. *Arch Intern Med*. 2003;163:1555–1565.
73. Pohl MA, Blumenthal S, Cordonnier DJ, et al. Independent and additive impact of blood pressure control and angiotensin II receptor blockade on renal outcomes in the irbesartan diabetic nephropathy trial: clinical implications and limitations. *J Am Soc Nephrol*. 2005;16:3027–3037.
74. Hursh BE, Ronsley R, Islam N, et al. Acute kidney injury in children with type 1 diabetes hospitalized for diabetic ketoacidosis. *JAMA Pediatr*. 2017;171:e170020.

75. Venot M, Weis L, Clec'h C, et al. Acute kidney injury in severe sepsis and septic shock in patients with and without diabetes mellitus: a multicenter study. *PLoS One*. 2015;10:e170020.
76. Girman CJ, Kou TD, Brodovicz K, et al. Risk of acute renal failure in patients with type 2 diabetes mellitus. *Diabet Med*. 2012;29:614–621.
77. Hertzberg D, Sartipy U, Holzmann MJ. Type 1 and type 2 diabetes mellitus and risk of acute kidney injury after coronary artery bypass grafting. *Am Heart J*. 2015;170:895–902.
78. Thakar CV, Christianson A, Himmelfarb J, et al. kidney injury episodes and chronic kidney disease risk in diabetes mellitus. *Clin J Am Soc Nephrol*. 2011;6:2567–2572.
79. Monseu M, Gand E, Saulnier PJ, et al. Acute kidney injury predicts major adverse outcomes in diabetes: synergic impact with low glomerular filtration rate and albuminuria. *Diabetes Care*. 2015;38:2333–2340.
80. Almeida JC, Zelmanovitz T, Vaz JS, et al. Sources of protein and polyunsaturated fatty acids of the diet and microalbuminuria in type 2 diabetes mellitus. *J Am Coll Nutr*. 2008;27:528–537.
81. Wrone EM, Carnethon MR, Palaniappan L, et al. Association of dietary protein intake and microalbuminuria in healthy adults: Third National Health and Nutrition Examination Survey. *Am J Kidney Dis*. 2003;4:580–587.
82. Cirillo M, Lombardi C, Chiricone D, et al. Protein intake and kidney function in the middle-age population: contrast between cross-sectional and longitudinal data. *Nephrol Dial Transplant*. 2014;21:1733–1740.
83. Tuttle KR, Bruton JL, Perusek MC, et al. Effect of strict glycemic control on renal hemodynamic response to amino acids and renal enlargement in insulin-dependent diabetes melliuts. *N Engl J Med*. 1991;324:1626–1632.
84. Jheng HF, Hirotsuka M, Goto T, et al. Dietary low-fat soy milk powder retards diabetic nephropathy progression via inhibition of renal fibrosis and renal inflammation. *Mol Nutr Food Res*. 2017;61.
85. Hansen HP, Tauber-Lassen E, Jensen BR, et al. Effect of dietary protein restriction on prognosis in patients with diabetic nephropathy. *Kidney Int*. 2002;62:220–228.
86. Pan Y, Guo LL, Jin HM. Low-protein diet for diabetic nephropathy: a meta-analysis of randomized controlled trials. *Am J Clin Nutr*. 2008;88:660–666.
87. National Kidney Foundation. KDOQI Clinical Practice Guideline for Diabetes and CKD: 2012 update. *Am J Kidney Dis*. 2012;60:850–886.
88. Tuttle KR, Bakris GL, Bilous RW, et al. Diabetic kidney disease: a report from an ADA Consensus Conference. *Am J Kidney Dis*. 2014;64:510–533.
89. Michels WM, Grootendorst DC, Verduijn M, et al. Performance of the Cockcroft-Gault, MDRD, and new CKD-EPI formulas in relation to GFR, age, and body size. *Clin J Am Soc Nephrol*. 2010;5:1003–1009.
90. Levey AS, Coresh J, Tighiouart H, et al. Measured and estimated glomerular filtration rate: current status and future directions. *Nat Rev Nephrol*. 2020;16:51–64.
91. Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2020 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int*. 2020;98:S1–S115.
92. Afkarian M, Zelnick LR, Hall YN, et al. Clinical manifestations of kidney disease among US adults with diabetes, 1988-2014. *JAMA*. 2016;316:602–610.
93. Dwyer JP, Parving HH, Hunsicker LG, et al. Renal dysfunction in the presence of normoalbuminuria in type 2 diabetes: results from the DEMAND study. *Cardiorenal Med*. 2012;2:1–10.
94. Garg AX, Kiberd BA, Clark WF, et al. Albuminuria and renal insufficiency prevalence guides population screening: results from the NHANES III. *Kidney Int*. 2002;61:2165–2175.
95. Kramer HJ, Nguyen QD, Curhan G, et al. Renal insufficiency in the absence of albuminuria and retinopathy among adults with type 2 diabetes mellitus. *JAMA*. 2003;289:3273–3277.
96. Klessens CQF, Woutman TD, Veraar KAM, et al. An autopsy study suggests that diabetic nephropathy is underdiagnosed. *Kidney Int*. 2016;90:149–156.
97. Colhoun HM, Marcovecchio ML. Biomarkers of diabetic kidney disease. *Diabetologia*. 2018;61:996–1011.
98. Merchant ML, Niewczas MA, Ficociello LH, et al. Plasma kininogen and kininogen fragments are biomarkers of progressive renal decline in type 1 diabetes. *Kidney Int*. 2013;83:1177–1184.
99. Zhang Y, Zhang S, Wang G. Metabolomic biomarkers in diabetic kidney diseases – a systematic review. *J Diabetes Complications*. 2015;29:1345–1351.
100. Coca SG, Nadkarni GN, Huang Y, et al. Plasma biomarkers and kidney function decline in early and established diabetic kidney disease. *J Am Soc Nephrol*. 2017;28:2786–2793.
101. Pavkov ME, Nelson RG, Knowler WC, et al. Elevation of circulating TNF receptors 1 and 2 increases the risk of end-stage renal disease in American Indians with type 2 diabetes. *Kidney Int*. 2015;87:812–819.
102. Tam FWK, Riser BL, Meeran K, et al. Urinary monocyte chemoattractant protein-1 (MCP-1) and connective tissue growth factor (CCN2) as prognostic markers for progression of diabetic nephropathy. *Cytokine*. 2009;47:37–42.
103. Schrauben SJ, Shou H, Zhang X, et al. Association of multiple plasma biomarker concentrations with progression of prevalent diabetic kidney disease: findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. *J Am Soc Nephrol*. 2021;32:115–126.
104. Argyropoulos C, Wang K, Bernardo J, et al. Urinary microRNA profiling predicts the development of microalbuminuria in patients with type 1 diabetes. *J Clin Med*. 2015;4:1498–1517.
105. Koye DN, Magliano DJ, Reid CM, et al. Risk of progression of normoalbuminuric CKD to end-stage kidney disease in people with diabetes: the CRIC (Chronic Renal Insufficiency Cohort) study. *Am J Kidney Dis*. 2018;72:653–661.
106. Mogensen CE, Christensen CK, Vittinghus E. The stages in diabetic renal disease. With emphasis on the stage of incipient diabetic nephropathy. *Diabetes*. 1983;32:64–78.
107. Krolewski AS, Niewczas MA, Skupien J, et al. Early progressive renal decline precedes the onset of microalbuminuria and its progression to macroalbuminuria. *Diabetes Care*. 2014;37:226–234.
108. Gaede P, Tarnow L, Vedel P, et al. Remission to normoalbuminuria during multifactorial treatment preserves kidney function in patients with type 2 diabetes and microalbuminuria. *Nephrol Dial Transplant*. 2004;19:2784–2788.
109. Bentata Y, Karimi I, Benabdellah N, et al. Albuminuria in type 2 diabetes mellitus: from remission to progression. *Ren Fail*. 2016;38:481–483.
110. Hill CJ, Cardwell CR, Patterson CC, et al. Chronic kidney disease and diabetes in the national health service: a cross-sectional survey of the U.K. national diabetes audit. *Diabet Med*. 2014;31:448–454.
111. Hadjadj S, Cariou B, Fumeron F, et al. Death, end-stage renal disease and renal function decline in patients with diabetic nephropathy in French cohorts of type 1 and type 2 diabetes. *Diabetologia*. 2016;59:208–216.
112. Koye DN, Magliano DJ, Reid CM, et al. Trends in incidence of ESKD in people with type 1 and type 2 diabetes in Australia, 2002-2013. *Am J Kidney Dis*. 2019;73:300–308.
113. Moseley KF. Type 2 diabetes and bone fractures. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:1432–1435.
114. Kanwar YS, Sun L, Xie P, et al. A glimpse of various pathogenetic mechanisms of diabetic nephropathy. *Annu Rev Pathol*. 2011;6:395–423.
115. Vallon V, Komers R. Pathophysiology of the diabetic kidney. *Compr Physiol*. 2011;1:1175–1232.
116. Hostetter TH. Hypertrophy and hyperfunction of the diabetic kidney. *J Clin Invest*. 2001;107:161–162.
117. Kriz W, Lemley KV. A potential role for mechanical forces in the detachment of podocytes and the progression of CKD. *J Am Soc Nephrol*. 2015;26:258–269.
118. Caramori ML, Parks A, Mauer M. Renal lesions predict progression of diabetic nephropathy in type 1 diabetes. *J Am Soc Nephrol*. 2013;24:1175–1181.
119. Fioretto P, Mauer M. Histopathology of diabetic nephropathy. *Semin Nephrol*. 2007;27:195–207.
120. Bak M, Thomsen K, Christiansen T, et al. Renal enlargement precedes renal hyperfiltration in early experimental diabetes in rats. *J Am Soc Nephrol*. 2000;11:1287–1292.
121. Najafan B. Pathology of human diabetic nephropathy. In: Lai KNTS, ed. *Diabetes and the Kidney*. Basel: Karger; 2011:36–47.
122. Conti S, Perico N, Novelli R, et al. Early and late scanning electron microscopy findings in diabetic kidney disease. *Sci Rep*. 2018;8:4909.
123. Andeen NK, Nguyen TQ, Steegh F, et al. The phenotypes of podocytes and parietal epithelial cells may overlap in diabetic nephropathy. *Kidney Int*. 2015;88:1099–1107.
124. Jones DB. Correlative scanning and transmission electron microscopy of glomeruli. *Lab Invest*. 1977;37:569–578.

125. Wolf G, Reinking R, Zahner G, et al. Erk 1,2 phosphorylates p27(Kip1): functional evidence for a role in high-glucose-induced hypertrophy of mesangial cells. *Diabetologia*. 2003;46:1090–1099.
126. Saito Y, Kida H, Takeda S, et al. Mesangiolysis in diabetic glomeruli: its role in the formation of nodular lesions. *Kidney Int*. 1988;34:389–396.
127. Fioretto P, Caramori ML, Mauer M. The kidney in diabetes: dynamic pathways of injury and repair. The Camillo Golgi Lecture 2007. *Diabetologia*. 2008;51:1347–1355.
128. Alsaad KO, Herzenberg AM. Distinguishing diabetic nephropathy for other causes of glomerulosclerosis: an update. *J Clin Pathol*. 2007;60:18–26.
129. Tyagi I, Agrawal U, Amitabh V, et al. Thickness of glomerular and tubular basement membranes in preclinical and clinical stages of diabetic nephropathy. *Indian J Nephrol*. 2008;18:64–69.
130. Tervaert TWC, Mooyaart AL, Amann K, et al. Renal Pathology Society. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol*. 2010;21:556–563.
131. Parving HH. Diabetic nephropathy: prevention and treatment. *Kidney Int*. 2001;60:2041–2055.
132. Comai G, Malvi D, Angeletti A, et al. Histological evidence of diabetic kidney disease precede clinical diagnosis. *Am J Nephrol*. 2019;50:29–36.
133. Neumiller JJ, Alicic RZ, Tuttle KR. Therapeutic considerations for antihyperglycemic agents in diabetic kidney disease. *J Am Soc Nephrol*. 2017;28:2263–2274.
134. Rave K, Nosek L, Posner J, et al. Renal glucose excretion as a function of blood glucose concentration in subjects with type 2 diabetes–results of a hyperglycaemic glucose clamp study. *Nephrol Dial Transplant*. 2006;21:2166–2171.
135. Alsahli M, Gerich JE. Renal glucose metabolism in normal physiological conditions and in diabetes. *Diabetes Res Clin Pract*. 2017;133:1–9.
136. Mather A, Pollock C. Renal glucose transporters: novel targets for hyperglycemia management. *Nat Rev Nephrol*. 2010;6:307–311.
137. Wright EM. Renal Na(+)-glucose cotransporters. *Am J Physiol Renal Physiol*. 2001;280:F10–F18.
138. Mogensen CE. Maximum tubular reabsorption capacity for glucose and renal hemodynamics during rapid hypertonic glucose infusion in normal and diabetic subjects. *Scand J Clin Lab Invest*. 1971;28:101–109.
139. Marks J, Carvou NJC, Debnam ES, et al. Diabetes increases facilitative glucose uptake and GLUT2 expression at the rat proximal tubule brush border membrane. *J Physiol*. 2003;553:137–145.
140. DeFronzo RA, Hompesch M, Kasichayanula S, et al. Characterization of renal glucose reabsorption in response to dapagliflozin in healthy subjects and subjects with type 2 diabetes. *Diabetes Care*. 2013;36:3169–3176.
141. Gallo LA, Ward MS, Fotheringham AK, et al. Once daily administration of the SGLT2 inhibitor, empagliflozin, attenuates markers of renal fibrosis without improving albuminuria in diabetic db/db mice. *Sci Rep*. 2016;6:26428.
142. Norton L, Shannon CE, Fourcaudot M, et al. Sodium-glucose co-transporter (SGLT) and glucose transporter (GLUT) expression in the kidney of type 2 diabetic subjects. *Diabetes Obes Metab*. 2017;19:1322–1326.
143. Rahmoune H, Thompson PW, Ward JM, et al. Glucose transporters in human renal proximal tubular cells isolated from the urine of patients with non-insulin-dependent diabetes. *Diabetes*. 2005;54:3427–3434.
144. Ackermann TF, Boini KM, Völkl H, et al. SGK1-sensitive renal tubular glucose reabsorption in diabetes. *Am J Physiol Renal Physiol*. 2009;296:F859–F866.
145. Palmada M, Boehmer C, Akel A, et al. SGK1 kinase upregulates GLUT1 activity and plasma membrane expression. *Diabetes*. 2006;55:421–427.
146. Tuttle KR. Back to the future: glomerular hyperfiltration and the diabetic kidney. *Diabetes*. 2017;66:14–16.
147. Guinti S, Barit D, Cooper ME. Mechanisms of diabetic nephropathy: role of hypertension. *Hypertension*. 2006;48:519–526.
148. Anderson S, Brenner BM. Pathogenesis of diabetic glomerulopathy: hemodynamic considerations. *Diabetes Metab Rev*. 1988;4:163–177.
149. Premaratne E, Verma S, Ekinci EI, et al. The impact of hyperfiltration on the diabetic kidney. *Diabetes Metab*. 2015;41:5–17.
150. Vallon V, Richter K, Blantz RC, et al. Glomerular hyperfiltration in experimental diabetes mellitus: potential role of tubular reabsorption. *J Am Soc Nephrol*. 1999;10:2569–2576.
151. Praga M. Synergy of low nephron number and obesity: a new focus on hyperfiltration nephropathy. *Nephrol Dial Transplant*. 2005;20:2594–2597.
152. Chagnac A, Weinstein T, Korzets A, et al. Glomerular hemodynamics in severe obesity. *Am J Physiol Renal Physiol*. 2000;278:F817–F822.
153. Bankir L, Roussel R, Bouby N. Protein- and diabetes-induced glomerular hyperfiltration: role of glucagon, vasopressin, and urea. *Am J Physiol Renal Physiol*. 2015;309:F2–F23.
154. Tuttle KR, Puhlman ME, Cooney SK, et al. Effects of amino acids and glucagon on renal hemodynamics in type 1 diabetes. *Am J Physiol Renal Physiol*. 2002;282:F102–F112.
155. Tuttle KR, Bruton JL. Effect of insulin therapy on renal hemodynamic response to amino acids and renal hypertrophy in non-insulin-dependent diabetes. *Kidney Int*. 1992;42:167–173.
156. Alicic RZ, Neumiller JJ, Johnson EJ, et al. Sodium-glucose cotransporter 2 inhibition and diabetic kidney disease. *Diabetes*. 2019;68:248–257.
157. Endlich N, Kress KR, Reiser J, et al. Podocytes respond to mechanical stress in vitro. *J Am Soc Nephrol*. 2001;12:413–422.
158. Srivastava T, Thiagarajan G, Alon US, et al. Role of biomechanical forces in hyperfiltration-mediated glomerular injury in congenital anomalies of the kidney and urinary tract. *Nephrol Dial Transplant*. 2017;32:759–765.
159. Srivastava T, Celsi GE, Sharma M, et al. Fluid flow shear stress over podocytes is increased in the solitary kidney. *Nephrol Dial Transplant*. 2014;29:65–72.
160. Grabias BM, Konstantopoulos K. The physical basis of renal fibrosis: effects of altered hydrodynamic forces on kidney homeostasis. *Am J Physiol Renal Physiol*. 2014;306:F473–F485.
161. Tobar A, Ori Y, Benchetrit S, et al. Proximal tubular hypertrophy and enlarged glomerular and proximal tubular urinary space in obese subjects with proteinuria. *PLoS One*. 2013;8:e75547.
162. Weinbaum S, Duan Y, Satlin LM, et al. Mechanotransduction in the renal tubule. *Am J Physiol Renal Physiol*. 2010;299:F1220–F1236.
163. Nelson PJ, Rees AJ, Griffin MD, et al. The renal mononuclear phagocytic system. *J Am Soc Nephrol*. 2012;23:194–203.
164. Navarro-Gonzalez JF, Mora-Fernandez C, Muros de Fuentes M, et al. Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. *Nat Rev Nephrol*. 2011;7:327–340.
165. Kikuchi Y, Ikee R, Hemmi N, et al. Fractalkine and its receptor, CX3CR1, upregulation in streptozotocin-induced diabetic kidneys. *Nephron Exp Nephrol*. 2004;97:e17–e25.
166. Chow FY, Nikolic-Paterson DJ, Ozols E, et al. Intercellular adhesion molecule-1 deficiency is protective against nephropathy in type 2 diabetic db/db mice. *J Am Soc Nephrol*. 2005;16:1711–1722.
167. Pichler R, Afkarian M, Dieter BP, et al. Immunity and inflammation in diabetic kidney disease: translating mechanisms to biomarkers and treatment targets. *Am J Physiol Renal Physiol*. 2017;312:F716–F731.
168. Alicic RZ, Johnson EJ, Tuttle KR. Inflammatory mechanisms as new biomarkers and therapeutic targets for diabetic kidney disease. *Adv Chronic Kidney Dis*. 2018;25:181–191.
169. Furuta T, Saito T, Ootaka T, et al. The role of macrophages in diabetic glomerulosclerosis. *Am J Kidney Dis*. 1993;21:480–485.
170. Ninichuk V, Khandoga AG, Segerer S, et al. The role of interstitial macrophages in nephropathy of type 2 diabetic db/db mice. *Am J Pathol*. 2007;170:1267–1276.
171. Chung AC, Lan HY. Chemokines in renal injury. *J Am Soc Nephrol*. 2011;22:802–809.
172. Zhang JM, An J. Cytokines, inflammation, and pain. *Int Anesthesiol Clin*. 2007;45:27–37.
173. Pfeilschifter J, Muhl H. Interleukin 1 and tumor necrosis factor potentiate angiotensin II- and calcium ionophore-stimulated prostaglandin E2 synthesis in rat renal mesangial cells. *Biochem Biophys Res Commun*. 1990;169:585–595.
174. Jones SA. Directing transition from innate to acquired immunity: defining a role for IL-6. *J Immunol*. 2005;175:3463–3468.
175. Miyauchi K, Takiyama Y, Honjyo J, et al. Upregulated IL-18 expression in type 2 diabetic subjects with nephropathy: TGF-beta1 enhanced IL-18 expression in human renal proximal tubular epithelial cells. *Diabetes Res Clin Pract*. 2009;83:190–199.
176. Al-Lamki RS, Mayadas TN. TNF receptors: signaling pathways and contribution to renal dysfunction. *Kidney Int*. 2015;87:281–296.

177. Haller H, Bertram A, Nadrowitz F, et al. Monocyte chemoattractant protein-1 and the kidney. *Curr Opin Nephrol Hypertens*. 2016;25:42–49.
178. Lee EY, Chung CH, Khoury CC, et al. The monocyte chemoattractant protein-1/CCR2 loop, inducible by TGF-beta, increases podocyte motility and albumin permeability. *Am J Physiol Renal Physiol*. 2009;297:F85–F94.
179. Donadelli R, Zanchi C, Morigi M, et al. Protein overload induces fractalkine upregulation in proximal tubular cells through nuclear factor kappaB- and p38 mitogen-activated protein kinase-dependent pathways. *J Am Soc Nephrol*. 2003;14:2436–2446.
180. Wanner N, Bechtel-Walz W. Epigenetics of kidney disease. *Cell Tissue Res*. 2017;369:75–92.
181. Yang XH, Cao RF, Yu Y, et al. A study on the correlation between MTHFR promoter methylation and diabetic nephropathy. *Am J Transl Res*. 2016;8:4960–4967.
182. Wang X, Lui J, Zhen J, et al. Histone deacetylase 4 selectively contributes to podocyte injury in diabetic nephropathy. *Kidney Int*. 2014;86:712–725.
183. Noh H, Oh EY, Seo JY, et al. Histone deacetylase-2 is a key regulator of diabetes- and transforming growth factor-beta1-induced renal injury. *Am J Physiol Renal Physiol*. 2009;297:F729–F739.
184. Krupa A, Jenkins R, Luo DD, et al. Loss of MicroRNA-192 promotes fibrogenesis in diabetic nephropathy. *J Am Soc Nephrol*. 2010;21:438–447.
185. Deshpande SD, Putta S, Wang M, et al. Transforming growth factor-ß-induced cross talk between p53 and a microRNA in the pathogenesis of diabetic nephropathy. *Diabetes*. 2013;62:3151–3162.
186. Putta S, Lanting L, Sun G, et al. Inhibiting microRNA-192 ameliorates renal fibrosis in diabetic nephropathy. *J Am Soc Nephrol*. 2012;23:458–469.
187. Leung A, Amaram V, Natarajan R. Linking diabetic vascular complications with LncRNAs. *Vasc Pharmacol*. 2019;114:139–144.
188. DCCT/EDIC Research Group. Effect of intensive diabetes treatment on albuminuria in type 1 diabetes: long-term follow-up of the diabetes control and complications trial and epidemiology of diabetes interventions and complications study. *Lancet Diabetes Endocrinol*. 2014;2:793–800.
189. Zoungas S, Arima H, Gerstein HC, et al. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. *Lancet Diabetes Endocrinol*. 2017;5:431–437.
190. Zoungas S, Chalmers J, Ninomiya T, et al. Association of HbA1c levels with vascular complications and death in patients with type 2 diabetes: evidence of glycaemic thresholds. *Diabetologia*. 2012;55:636–643.
191. Little RR, Rohlfing CL, Tennill AL, et al. Measurement of Hba(1C) in patients with chronic renal failure. *Clin Chem Acta*. 2013;418:73–76.
192. Tarim O, Kucukerdogan A, Gunay U, et al. Effects of iron deficiency anemia on hemoglobin A1c in type 1 diabetes mellitus. *Pediatr Int*. 1999;41:357–362.
193. Bergenstal RM, Beck RW, Close KL, et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. *Diabetes Care*. 2018;41:2275–2280.
194. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*. 2017;377:644–657.
195. Wiviott SD, Raz I, Bonaca MP, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2019;380:347–357.
196. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373:2117–2128.
197. Wanner C, Inzucchi SE, Lachin JM, et al. Empagliflozin and progression of kidney disease in type 2 diabetes. *N Engl J Med*. 2016;375:323–334.
198. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*. 2019;380:2295–2306.
199. Heerspink HJL, Stefansson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383:1436–1446.
200. Packer M, Anker SD, Butler J, et al. EMPEROR-Reduced Trial Investigators. Cardiovascular and renal outcomes with empagliflozin in heart failure. *N Engl J Med*. 2020;383:1413–1424.
201. Tuttle KR, Brosius 3rd FC, Cavender MA, et al. SGLT2 inhibition for ckd and cardiovascular disease in type 2 diabetes: report of a scientific workshop sponsored by the national kidney foundation. *Diabetes. Am J Kidney Dis.* 2021;70:1–16.
202. Gerstein HC, Colhoun HM, Dagenais GR, et al. REWIND Investigators. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomized placebo-controlled trial. *Lancet*. 2019;394:121–130.
203. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016;375:311–322.
204. Marso SP, Bain SC, Consoli A, et al. SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2016;375:1834–1844.
205. Tuttle KR, Lakshmanan MC, Rayner B, et al. Dulaglutide versus insulin glargine in patients with type 2 diabetes and moderate-to-severe chronic kidney disease (AWARD-7): a multicentre, open-label, randomised trial. *Lancet Diabetes Endocrinol*. 2018;6:605–617.
206. Kristensen SL, Rorth R, Jhund PS, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Lancet Diabetes Endocrinol*. 2019;7:776–785.
207. US National Library of Medicine. ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03819153?term=semaglutide+-FLOW&draw=2&rank=1. Accessed October 15, 2020.
208. Alicic RZ, Cox EJ, Neumiller JJ, et al. Incretin drugs in diabetic kidney disease: biological mechanisms and clinical evidence. *Nature Neph Reviews*. 2021;17:227–244.
209. Emdin CA, Rahimi K, Neal B, et al. Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. *JAMA*. 2015;313:603–615.
210. Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO clinical practice guideline for the management of blood pressure in chronic kidney disease. *Kidney Int*. 2012;2:337–414.
211. Brenner BM, Cooper ME, de Zeeuw D, RENAAL Study Investigators, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med*. 2001;345:861–869.
212. Lewis EJ, Hunsicker LG, Bain RP, The Collaborative Study Group, et al. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *N Engl J Med*. 1993;329:1456–1462.
213. Lewis EJ, Hunsicker LG, Clark WR, Collaborative Study Group, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med*. 2001;345:851–860.
214. Yusuf S, Teo KK, Pogue J, ONTARGET Investigators, et al. Telmisartan, ramipril, or both in patients at high risk for vascular events. *N Engl J Med*. 2008;358:1547–1559.
215. Fried LF, Emanuele N, Zhang JH, VA NEPHRON-D Investigators, et al. Combined angiotensin inhibition for the treatment of diabetic nephropathy. *N Engl J Med*. 2013;369:1892–1903.
216. Bakris GL, Agarwal R, Anker S, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *N Eng J Med*. 2020;383:2219–2229.
217. Pitt B, Filippatos G, Agarwal R, et al. Cardiovascular events with finerenone in kidney disease and type 2 diabetes. *N Engl J Med*. 2021.
218. Kerendia (finerenone) tablets [product information]. Whippany, NJ: Bayer HealthCare Pharmaceuticals Inc.; 2021.
219. Jannson FJ, Forsblom C, Harjutsalo V, FinnDiane Study Group, et al. Regression of albuminuria and its association with incident cardiovascular outcomes and mortality in type 1 diabetes: the FinnDiane Study. *Diabetologia*. 2018;61:1203–1211.
220. Fox CS, Matsushita K, Woodward M, et al. Chronic Kidney Disease Prognosis Consortium. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet*. 2012;380:1662–1673.
221. Adler AI, Stevens RJ, Manley SE, UKPDS GROUP, et al. Development and progression of nephropathy in type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS 64). *Kidney Int*. 2003;63:225–232.
222. Das SR, Everett BM, Birtcher KK, et al. 2020 expert consensus decision pathway on novel therapies for cardiovascular risk reduction in patients with type 2 diabetes: a report of the American College of Cardiology Solution Set Oversight Committee. *J Am Coll Cardiol*. 2020;76:1117–1145.

223. Klahr S, Buerkert J, Purkerson ML. Role of dietary factors in the progression of chronic renal disease. *Kidney Int*. 1983;24:579–587.
224. Betz M, Steenes A, Peterson L, et al. Knowledge does not correspond to adherence of renal diet restrictions in patients with chronic kidney disease stage 3-5. *J Ren Nutr*. 2021;31:351–360.
225. Hartman RE, Rao PSS, Churchwell MD, et al. Novel therapeutic agents for the treatment of diabetic kidney disease. *Expert Opin Investig Drugs*. 2020;29:1277–1293.
226. Tuttle KR, Alicic RZ, Duru OK, et al. Clinical characteristics of and risk factors for chronic kidney disease among adults and children: an analysis of the CURE-CKD Registry. *JAMA Netw Open*. 2019;2:e1918169.

46

Diabetic Foot and Vascular Complications

Adam L. Isaac, Leonardo C. Clavijo, Gregory A. Magee, and David G. Armstrong

OUTLINE

Introduction, 791
Pathophysiology, 791
  Neuropathy, 792
  Infection, 792
  Ischemia, 792
  Bone and Joint Abnormalities, 792
Presentation and Diagnosis, 793
  Diabetic Foot Ulcers, 793
  Infection/Osteomyelitis, 793
  Peripheral Arterial Disease, 794
Medical Management, 795
  Prevention and Screening, 795
  Antibiotics, 795
  Wound Care, 795
  Offloading, 795
Surgical Management, 796
  Drainage Procedures, 796
  Lower Extremity Arterial Reconstruction, 796
  Endovascular Procedures, 796
  Amputations, 797
Multidisciplinary Approach, 797
Future Trends, 797

KEY POINTS

- Prevention of diabetic foot ulcers involves aggressive glycemic control, management of associated medical risk factors, periodic physical examination (including foot and vascular examination), and proper foot care and hygiene strategies.
- The goal of all foot care strategies (both medical and surgical) is to achieve complete closure of the "skin envelope" as quickly as possible while maintaining a functional, ambulatory foot.
- Vascular reconstruction has proven to be a safe and durable revascularization alternative in patients with diabetes and should be used to avoid ischemic diabetic foot complications and resulting limb amputations.

INTRODUCTION

Diabetic foot and vascular complications, including ulcers, infections, and amputations, continue to rise at an alarming rate, presenting an enormous challenge to providers, patients, and caretakers across the entire healthcare spectrum. In the United States, care of the lower extremity accounts for one-third of the annual direct costs for diabetes,[1-3] and the 5-year mortality rate following a major (proximal to ankle) lower extremity amputation (56.6%) is second only to that of lung cancer (80%).[4,5] Remarkably, lower extremity complications of diabetes now constitute a top ten global burden of disability based on criteria set by the World Health Organization.[6,7]

The lifetime risk of developing a foot ulcer in patients with diabetes is 34%,[1] and more than 50% of wounds become infected.[8] In fact, the most common diabetes-related complication leading to hospitalization and lower limb amputation is diabetic foot infection.[9-11] Indeed, seconds count, as every 1.2 seconds someone develops a diabetic foot ulcer (DFU), every 7 seconds around the world someone dies from diabetes, and every 20 seconds someone is amputated.[12]

As the economic, medical, and social impacts of diabetic foot and vascular complications grow, focused education and evidence-based treatment strategies are needed to increase awareness and standardize care in clinical practice. Many providers and health systems throughout the world have successfully adopted the recommended multidisciplinary approach to diabetic foot care. In addition, advances in research, innovation, and technology have shown great potential in preventing amputations and preserving limbs.

Management of diabetic foot complications requires a thorough understanding of the pathophysiology of the underlying disease, as well as a familiarity with the wide range of therapeutic options currently available. The goal of this chapter is to outline the current concepts of diabetic foot disease and the most effective treatment strategies for both medical and surgical management.

KEY POINTS

- Diabetic foot and vascular complications present an enormous challenge to providers, patients and caretakers across the entire healthcare spectrum. In the United States, care of the lower extremity accounts for one-third of the annual direct costs for diabetes. The 5-year mortality rate following a major (proximal to ankle) lower extremity amputation (56.6%) is second only to that of lung cancer (80%). Every 1.2 seconds someone develops a diabetic foot ulcer, every 7 seconds around the world someone dies from diabetes, and every 20 seconds someone is amputated.

PATHOPHYSIOLOGY

The epidemiology and pathophysiology of diabetes have been discussed in previous chapters, and the pathways that lead to major complications, including retinopathy, neuropathy, and nephropathy, exhibit many of the same underlying mechanisms as lower extremity complications. Although glycemic control is critical in diabetes management, it is simply not enough to prevent diabetic complications of the diabetic foot. Furthermore, it is important for clinicians to be aware of the so-called etiologic triad leading to diabetic foot complications (i.e., neuropathy, ischemia, and infection) and to recognize that these may occur in isolation, but more commonly occur in combination.

Neuropathy

Neuropathy that arises from diabetes is often multifactorial in that sensory, motor, and autonomic fibers may be affected. Sensory neuropathy affects the small-diameter pain and temperature fibers first, which can result in decreased sensation and greater susceptibility to injury caused by pressure-related microtrauma and other minor skin injuries. Motor neuropathy affects the longer fibers that innervate the foot, including intrinsic foot muscles and leg muscles. Atrophy, or muscle wasting, of the intrinsic foot muscles allows the strong flexor muscles to draw the toes into a "clawed" position, whereby new pressure points form along the toes and metatarsals. In addition, diminished joint mobility combined with increased rigidity from glycation of scleral proteins exacerbates the situation by altering the normal weight distribution on the foot. Lastly, autonomic neuropathy causes the skin to become dry through loss of sweat and apocrine gland function. The resulting dry, scaly skin can lead to cracks and fissures, thereby creating a portal of entry for bacteria. Additionally, patients with diabetes can have a blunted neuroinflammatory response and thus are missing a crucial component of the body's natural first-line defense against pathogens.[13]

Infection

Patients with diabetes typically have an altered response to infectious processes owing to defects in their host immune defense system.[14] Moreover, wound healing can be delayed in these patients as a result of abnormal cellular and inflammatory pathways involving fibroblasts, neutrophils, and advanced glycation end-products (AGEs). Glycation is a nonenzymatic chemical reaction whereby sulfhydryl protein linkages are replaced by glucose, causing impairment in normal cellular and tissue functions.[15] AGEs increase the stiffness of precapillary vessel walls and contribute to the development of diabetic microangiopathy.[16] Additionally, the presence of neuropathy and ischemia can potentiate the infectious process whereby the loss of protective sensation from neuropathy leads to unrecognized injury and secondary infection, and ischemia results in poor healing, blunts the inflammatory response, and may lead to inadequate tissue levels of antimicrobial agents.

Ischemia

Up to 50% of patients with diabetes and foot ulceration have concurrent peripheral artery disease (PAD).[8,17] Significant progress has been made in identifying the underlying cause of lower extremity ischemia in patients with diabetes, and results have challenged longstanding misconceptions. Thus, it is imperative that practitioners continue to reject the so-called "small-vessel disease" theory related to occlusions of the microcirculation, as espoused by a single histologic study in 1959,[18] and instead embrace the notion that ischemia results from both atherosclerotic macrovascular disease and microcirculatory dysfunction.[19] Patients with diabetes typically suffer from tibial and peroneal arterial disease with sparing of the foot arteries, especially the dorsalis pedis and its branches.

Research has demonstrated that diabetes causes structural and functional changes within the arteriolar and capillary systems, notably thickening of the basement membrane. However, despite thickening of the capillary basement membrane, there is no decrease in capillary luminal diameter.[20] A thickened membrane impairs leukocyte migration and hampers the normal hyperemic or vasodilatory response to injury, thus simultaneously increasing the susceptibility to injury while blunting the typical manifestations of such an injury.[21] Overall, this dysfunction of the microcirculation in diabetic patients creates a functionally ischemic foot despite conditions that would otherwise allow normal blood flow in healthy patients.

Besides causing specific structural changes in the microcirculation, diabetes causes an undesirable shift in the natural balance that exists between stress/ulceration and resistance to stress/ulceration. Thus, the diabetic foot is more prone to tissue injury from the repetitive low-intensity stress of normal weight bearing. The presence of neuropathy may necessitate revascularization under conditions of perfusion that would not require revascularization in the absence of neuropathy and is often described as neuroischemia. Because of the compromise caused by this neuroischemia, the diabetic foot requires even more perfusion than usual to resist ulceration and/or respond appropriately to injury.

Bone and Joint Abnormalities

As discussed previously, neuropathy plays an important role in the development of bone and joint abnormalities in patients with diabetes. These deformities lead to biomechanical changes and the formation of high-pressure areas, which increases the risk of ulceration (Fig. 46.1). Furthermore, changes in foot biomechanics, combined with limited joint mobility, may result in structural deformities, including hallux abducto valgus, claw- and hammertoes, and abnormalities of the metatarsophalangeal joints. Patients with hammertoes are vulnerable to even minor shoe trauma, and contracted toes cause increased retrograde pressure along the plantar metatarsals. In addition, irregular osseous changes and midfoot collapse owing to Charcot neuroarthropathy, as well as severe flatfoot, equinus, and other knee or back problems, may contribute to increased biomechanical stress encountered by the foot, worsen the degree of ulceration, and/or inhibit proper healing.[22]

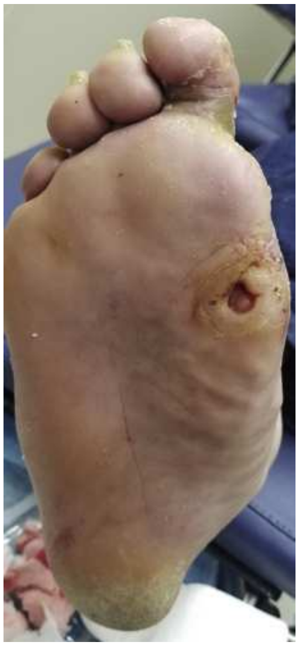


Fig. 46.1 Bone and joint abnormalities in the diabetic foot can result in the formation of high-pressure areas, increasing the risk of ulceration.

KEY POINTS

- It is important for clinicians to be aware of the so-called etiologic triad leading to diabetic foot complications (i.e., neuropathy, ischemia, and infection) and to recognize that these may occur in isolation, but more commonly occur in combination. Neuropathy that arises from diabetes is often multifactorial in that sensory, motor, and autonomic fibers may be affected. Patients with diabetes typically have an altered response to infectious processes owing to defects in their host immune defense system. Up to 50% of patients with diabetes and foot ulceration have concurrent peripheral artery disease. Bone and joint deformities in the foot lead to biomechanical changes and the formation of high-pressure areas, which increases the risk of ulceration.

PRESENTATION AND DIAGNOSIS

The mechanisms described previously place the diabetic patient at risk for lower extremity complications, ranging from simple ulcers to severe limb- or life-threatening infections and/or ischemic gangrene. As with most diseases, patient education and access to healthcare are integral to the timely diagnosis and treatment of these complications. Patients who are aware of their foot care needs are more likely to seek help during the early stages, and healthcare providers who regularly treat patients with diabetes are more likely to identify potential complications early. To decrease the morbidity and mortality associated with diabetic foot complications, healthcare providers must maintain heightened vigilance for these problems and execute appropriate judgment when making diagnoses in these patients.

Diabetic Foot Ulcers

Up to 25% of patients with diabetes will develop a foot ulcer during their lifetime,[23] of which one in five requires an amputation.[24] Many ulcers develop as a result of altered pressures created in the diabetic foot. These foot pressures are influenced by muscle atrophy, obesity, callus formation, other forms of local microtrauma (including improper footwear), and limited joint mobility. The presence of neuropathy, as discussed previously, increases the risk for ulcer formation, because patients with diabetes may not be aware of the damage they are inflicting on their feet through their normal activities of daily living. These patients are said to lack the “gift of pain,” and may develop severe complications rapidly.[25] Furthermore, impaired wound healing and blunted neuroinflammatory responses contribute to progression of the initial ulcer to a potentially more serious lesion.

Patients typically present with a wound that fails to heal or with pain at the site of a callus, pressure point, or other bony prominence. Important considerations for foot ulcers include the depth and extent of tissue loss, anatomic location, origin, presence of ischemia or infection, and clinical signs of systemic infection. Proper diagnosis of DFUs depends in large part on the education and reliability of the patient and physician. A patient must be instructed on the benefit of meticulous foot care and seek treatment regularly from a healthcare provider who is familiar with the diabetic foot. Similarly, the healthcare provider must be learned in the field of diabetes management so as not to miss the symptoms and signs of diabetic foot complications.

Infection/Osteomyelitis

As discussed earlier, patients with diabetes often have a blunted neuroinflammatory response and do not exhibit the typical physiologic reactions to infection. In fact, the usual manifestations of infection (e.g., fever, tachycardia, elevated white blood cell count) are frequently absent in diabetic patients. Therefore, patients with diabetes require extra vigilance so that providers do not underestimate potential limb- or even life-threatening conditions that may otherwise present as less severe. Unexplained hyperglycemia should prompt an aggressive search for an infectious source in patients with diabetes, as elevated glucose levels may be the only sign of impending problems.[26]

In many cases, patients do not recognize the signs of an infection until malodor or wound drainage becomes apparent on a shoe or sock (Fig. 46.2). By this point, the portal of entry for bacteria has already been well established, and a polymicrobial infection may overcome the diabetic patient’s defense system. It is therefore imperative that all ulcers or calluses be carefully probed and inspected, followed by unroofing of superficial eschar to search for potential deep space abscesses. It is not uncommon for clinicians to find unexpected purulent material or necrotic tissue underneath an area of dry crust or dry gangrene.

Deep abscess, osteomyelitis, and/or sepsis can develop rapidly in patients with diabetes. Diabetic foot infections are typically bacterial, and can be monomicrobial or polymicrobial, with *Staphylococcus* and *Streptococcus* representing the most common infecting organisms.[27,28] In patients with diabetes, osteomyelitis has been identified in up to 20% of outpatient cases and 67% of hospitalized patients requiring an amputation.[24,29] The majority of cases of osteomyelitis in the diabetic foot are the result of contiguous spread, with the ulcer forming the initial entry site of infection. Therefore, treatment of an acute infection may include antibiotics and surgery, depending on the severity and depth. Broad-spectrum empiric antibiotic therapy is started initially, followed by targeted therapy based on deep intraoperative tissue and bone cultures and sensitivities. Many times, multiple trips to the operating room are required before a wound is considered clean and uninfected. Once the wound is determined to be free of infection, planning of definitive procedures for closure can begin.

Although advanced imaging modalities can be helpful in diagnosing osteomyelitis, there may be significant costs associated with these tests, whereas the use of a simple metal probe may be sufficient in many cases. Grayson and coworkers revealed that, if a sterile probe hits bone, then osteomyelitis can be diagnosed with a sensitivity of 66%, a specificity of 85%, and a positive predictive value of 89%.[30]

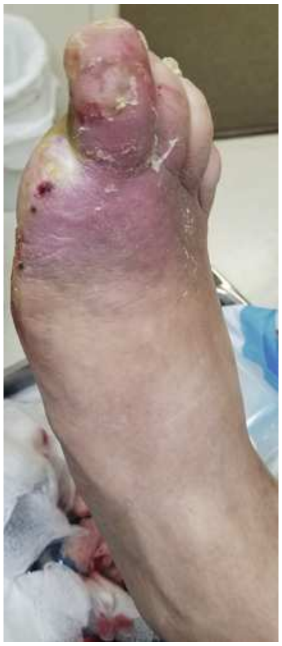


Fig. 46.2 Diabetic foot infection is often unrecognized until malodor or wound drainage becomes apparent on a patient’s shoe or sock.

Plain radiographs should be initially obtained to determine the extent of osseous erosion, as well as to assess anatomy for surgical planning.

Further advanced imaging with magnetic resonance imaging, computerized tomography scan, bone scintigraphy, or tagged white blood cell scan should be reserved for cases in which the metal probe test is equivocal, when an abscess or multifocal disease is suspected, or in patients with neuropathic osteoarthropathy, or Charcot's foot, in whom associated bony changes and inflammatory response can be misinterpreted as osteomyelitis. Definitive diagnosis of osteomyelitis is typically obtained by bone biopsy and deep intraoperative bone culture, but controversy exists regarding the consistency and validity of this approach.[31]

Peripheral Arterial Disease

The clinical presentation of peripheral arterial disease encompasses intermittent claudication, rest pain, and tissue loss (ulceration) with or without gangrene. The extent of ischemia and symptoms depends on the location of the vascular lesion, as well as the effectiveness of collateral circulation. Intermittent claudication presents as cramping, pain, or fatigue in the leg muscles, is seen with walking, and is relieved by rest. Of note, the anatomic location of the occlusive arterial lesion is usually one level above the clinically affected muscle group. Typically, aortoiliac disease causes buttock and thigh pain, and femoral disease causes calf discomfort. Tibial/peroneal disease usually does not cause claudication, although some patients will complain of foot pain or numbness while walking. Nocturnal muscle cramping, on the other hand, is a common complaint in diabetic patients and should not be mistaken for intermittent claudication, which results from exertion and is relieved by rest. Although claudication does cause disability from pain with walking, it rarely progresses to limb-threatening ischemia and often responds to conservative measures such as supervised exercise training[32] and smoking cessation.

Patients with more severe vascular disease may present with ischemic rest pain, which typically occurs in the distal foot and particularly the toes. Rest pain is exacerbated by recumbency and is relieved by dependency. Patients usually report pain when lying in bed or resting that is relieved by standing up and walking around. Diabetic neuropathy may sometimes mask symptoms of claudication or rest pain because these patients have blunted sensation in their lower extremities. More severe disease will result in tissue loss, including a foot ulceration or gangrene.

Once infection is ruled out, it is important to assess foot perfusion. Clinical examination alone is often insufficient in ruling out ischemia in patients with diabetes, and further investigation, such as ankle-brachial index (ABI), toe pressures, or skin perfusion pressure, may be warranted.[33] However, caution must be exercised in interpreting ABI results because of medial calcification, which is common in diabetes and can produce a falsely elevated or normal result.[34,35] Therefore, toe waveforms and systolic pressures are the preferred tests for patients with diabetes.

More recently, the Wound, Ischemia, and foot Infection (WIfI) classification has emerged as a valuable tool in evaluating patients at risk for limb loss. The WIfI system, based on wound depth, infection, and perfusion status, incorporates several other classification systems. Furthermore, it provides a comprehensive stratification model for determining amputation risk and aids in clinical decision-making.[36] It is critical that impaired perfusion be diagnosed promptly and staged based on the Society for Vascular Surgery WIfI classification. For example, WIfI stage 3 and 4 patients most often benefit from revascularization, and these patients should be identified early and treated aggressively. Any delay can result in loss of tissue and reduce the opportunity for limb preservation.[36,37] In addition to measuring wound characteristics, ischemia, and foot infection, functional deficit is now being added as a secondary survey after initial assessment of limb threat.[38] The WIfI plus Functional Impairment classification of limb threat is shown in Fig. 46.3.[36,38,39]

KEY POINTS

- Up to 25% of patients with diabetes will develop a foot ulcer during their lifetime, of which one in five requires an amputation. Impaired wound healing and blunted neuroinflammatory responses contribute to progression of the initial ulcer to a potentially more serious lesion. The usual manifestations of infection (e.g., fever, tachycardia, elevated white blood cell count) are frequently absent in diabetic patients. More recently, the Wound, Ischemia, and foot Infection (WIfI) classification has emerged as a valuable tool for evaluating patients at risk for limb loss. The WIfI system is based on wound depth, infection, and perfusion status and incorporates several other classification systems.

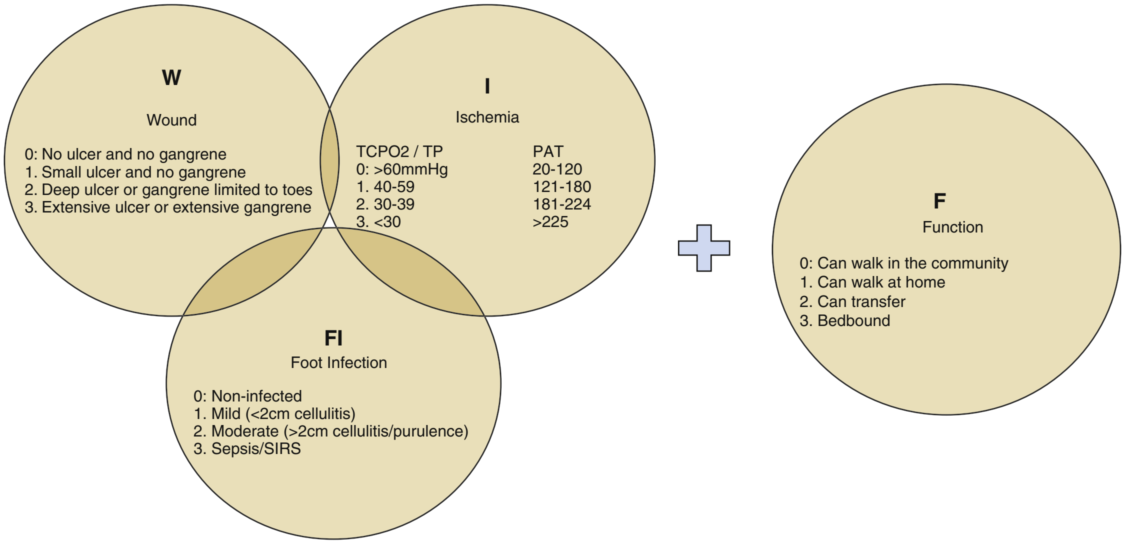

W
Wound
0: No ulcer and no gangrene
1. Small ulcer and no gangrene
2. Deep ulcer or gangrene limited to toes
3. Extensive ulcer or extensive gangrene
I
Ischemia
TCPO2 / TP
0: >60mmHg
1. 40-59
2. 30-39
3. <30
PAT
20-120
121-180
181-224
>225
FI
Foot Infection
0: Non-infected
1. Mild (<2cm cellulitis)
2. Moderate (>2cm cellulitis/purulence)
3. Sepsis/SIRS
F
Function
0: Can walk in the community
1. Can walk at home
2. Can transfer
3. Bedbound

Fig. 46.3 Wound, Ischemia, and foot Infection (WIfI) plus Functional Impairment classification of limb threat. *PAT,* pedal acceleration time; *TCPO2,* transcutaneous oxygen pressure; *TP,* toe pressure; *SIRS,* systemic inflammatory response syndrome.

MEDICAL MANAGEMENT

In addition to wound healing, the goals of diabetic foot management include infection control, amputation prevention, maintenance of health status, improved function and quality of life, and reduced costs.[40] Numerous modalities are available for achieving these goals, and healthcare providers who treat diabetic patients should be familiar with the most current therapeutic options.

Prevention and Screening

International guidelines suggest screening people at low risk for ulceration annually for loss of protective sensation and PAD, and persons at higher risk at higher frequencies for additional risk factors. For foot ulcer prevention, providers should educate patients about appropriate foot self-care and treat any preulcerative sign on the foot. Furthermore, moderate- to high-risk patients should be instructed to wear accommodative, properly fitting therapeutic footwear and consider monitoring foot skin temperature. Therapeutic footwear that has a demonstrated plantar pressure-relieving effect during walking to prevent plantar foot ulcer recurrence should be prescribed in higher risk patients.[41]

More specifically, prevention involves aggressive glycemic control; management of associated risk factors (such as smoking, hypertension, hyperlipidemia, and obesity); periodic physical examination, including a vascular examination; and perhaps most important, proper foot care and hygiene strategies. The importance of daily foot inspections should be emphasized to patients and their caretakers, with careful attention paid to calluses, fissures, and red or bruised areas, as well as open sores or blisters. Moisturizing creams should be used on dry areas of skin, and antifungal medication should be used as needed. In general, astringents and heat soaks should be avoided, as should patients' self-removal of eschar or other foot lesions. Toenails should be trimmed periodically to prevent penetration into adjacent toes and ulceration, in the hope that vigilance will prevent future complications.

Determining which patients are at risk for foot infection because of both neuropathy and ischemia is an important aspect of prevention.[23] Use of Semmes–Weinstein monofilament testing for sensory neuropathy and periodic vascular assessments (with or without the noninvasive vascular laboratory) are useful adjuncts in preventing foot complications.[42] The Ipswich touch test is a simplified sensory test that involves lightly touching the tips of the first, third, and fifth toes.[43]

Antibiotics

When DFUs become infected, deep (preferably, intraoperative) cultures should be obtained, and antibiotic therapy should be initiated. Common infective organisms include *Staphylococcus aureus, Streptococcus,* and occasionally gram-negative bacilli, and deeper infections may be polymicrobial, involving anaerobes and certain gas-forming species.[44] Initially, selection of broad-spectrum antibiotics should cover the most likely offending organisms and can be narrowed once gram stain or culture/sensitivity results have been reported. Other factors to bear in mind include local bacterial resistance patterns in a specific hospital or community, individual patient characteristics (such as comorbidities and allergies), and, most important, the severity of infection. Traditional therapy for osteomyelitis has been typically accepted as 4 to 6 weeks of intravenous antibiotics,[45] but recent studies have documented a greater than 30% recurrence rate using this modality alone.[46]

Wound Care

Local wound care is important to ensure adequate healing of DFUs, as well as to combat infection and prevent further complications. Debridement is an essential component of local wound care and many times can be performed at the bedside without anesthesia because of the neuropathy associated with diabetes. Debridement may take numerous forms, including mechanical, autolytic, enzymatic, and sharp. Mechanical debridement can involve various dressing types that remove adherent necrotic tissue, but is nonselective and can damage fragile tissue, whereas autolytic debridement depends on the patient's immune response and circulation, involves moist dressings, and is slow. Enzymatic debridement is selective for necrotic tissue and is achieved using commercial agents with various enzymes (such as collagenases). Lastly, sharp (surgical) debridement involves excision of tissue at the bedside or in the operating room.

The best wound care strategy combines sharp surgical debridement with a dressing type that effectively absorbs drainage and exudate while maintaining a moist wound environment. Numerous advanced wound-care modalities exist, such as negative pressure wound therapy (NPWT), cellular and tissue-based products, and hyperbaric oxygen therapy. Each has its own merits, but providers should be mindful of the economic implications and reimbursement challenges involved with using these dressings. Most importantly, any new wound-care modality should be regarded as adjunctive to frequent clinical examinations, high-quality local wound care, and offloading.

Prospective, randomized studies have demonstrated the effectiveness of NPWT in the treatment of DFUs.[47,48] In addition to maintaining a moist wound-healing environment, NPWT has been shown to decrease wound margins and promote granulation tissue formation and perfusion in DFUs. Furthermore, the mechanical stimulation of a wound by NPWT contributes to improved wound healing through micro- and macrostrain properties, and NPWT facilitates the removal of exudate and infectious materials.[49]

NPWT involves the application of local, subatmospheric pressure to a wound. Typically, a foam or gauze interface is applied directly to the wound site and connected to a pump via external tubing. The wound interface is secured in place with an adhesive dressing. A pressure of 75 to 125 mm Hg is commonly used, in a continuous or intermittent fashion, and can be adjusted based on the needs of the patient and healthcare team. Other uses include NPWT with fluid installation, closed-incision NPWT for complex, at-risk surgical sites, and ultraportable NPWT systems for skin graft bolstering.[50]

Offloading

Patients with diabetes and peripheral neuropathy are said to lack the "gift of pain," owing to decreased sensation and an inability to perceive emerging foot complications.[25] In addition, many patients do not adhere to recommendations for offloading, including protected weight-bearing (i.e., cast, boot), decreased activity level, and other preventative measures.

The "gold-standard" for offloading foot ulcers in patients with diabetes is total contact casting (TCC), and there is a large body of evidence in the literature to support its use. TCC can reduce pressure by up to 84% to 92% at the site of an ulcer and has been reported to heal most DFUs within 6 to 8 weeks.[51] Furthermore, Bus et al.[52] found that nonremovable offloading, that is, TCC, is more effective than removable devices for healing neuropathic foot ulcers. Yet, for all the evidence supporting the use of TCC for the treatment of DFUs, it is seldom used in practice. In a large multicenter study conducted by Wu and colleagues,[53] it was determined that, among 895 centers involved in the active treatment of DFUs, only 1.7% used TCC for the majority of cases. Several factors were noted to contribute to the lack of use of TCC, including patient tolerance, time needed to apply and remove the cast, materials cost, and reimbursement challenges.

One of the most important attributes of TCC is to ensure proper patient adherence because of its irremovability,[54] whereas removable cast walkers (RCWs), which allow for more frequent wound and skin inspection and dressing application, and, in some cases, provide nearly as much pressure reduction as TCC, are not as effective at healing DFUs.[55,56] A proposed alternative, the instant total contact cast (iTCC), has demonstrated healing results equivalent to TCC and may be more tolerable to patients. The iTCC consists of an RCW wrapped with cohesive bandage or fiberglass, rendering the device irremovable.[57]

KEY POINTS

- In addition to wound healing, the goals of diabetic foot management include infection control, amputation prevention, maintenance of health status, improvement of function and quality of life, and reduction in costs. For foot ulcer prevention, providers should educate patients about appropriate foot self-care and treat any preulcerative sign on the foot. Moderate- to high-risk patients should be instructed to wear accommodative, properly fitting therapeutic footwear and consider monitoring foot skin temperature. When diabetic foot ulcers become infected, deep (preferably intraoperative) cultures should be obtained, and antibiotic therapy should be initiated. Debridement is an essential component of local wound care and many times can be performed at the bedside without anesthesia because of the neuropathy associated with diabetes. The "gold-standard" for offloading foot ulcers in patients with diabetes is total contact casting.

SURGICAL MANAGEMENT

The goal of all foot care strategies is to achieve complete healing of the "skin envelope" as quickly as possible while maintaining a functional walking foot. When medical management is not enough, different surgical management strategies are available to treat the diabetic foot. These strategies range from wide debridement and drainage procedures to arterial reconstruction or amputation. The decision to operate necessitates careful preoperative planning by the surgeon, as well as a cooperative patient who is willing to accept the inherent risks of the procedure and the demanding rehabilitation period.

Drainage Procedures

When minor debridement strategies do not effectively relieve severe infection, it becomes imperative to perform more extensive drainage procedures (including partial open toe, ray, or forefoot amputation) to drain abscesses or remove necrotic tissue.[58] In fact, hidden infection should be suspected in any situation in which a well-perfused foot continues to experience necrosis. Adequate drainage should be obtained, even if foot function must be compromised initially, because function potentially can be restored later with numerous advanced wound-closure techniques; however, persistent infection can lead to severe limb- and life-threatening complications. Additionally, if ischemia appears to be contributing to the infection, then arteriography should be performed to determine whether perfusion is adequate, or if revascularization is necessary. Controlling active, spreading infection may delay revascularization initially, but longer waiting periods taken on in the hope of completely sterilizing wounds are not appropriate and may result in continued necrosis or tissue loss.[59] In fact, in some cases, inadequate blood flow may prevent the delivery of antibiotics, nutrients, or oxygen to the foot wound, thus requiring revascularization to correct the problem. Furthermore, bypass procedures can be performed safely in the presence of foot infection as long as sepsis has been controlled before surgery.[60]

Lower Extremity Arterial Reconstruction

Despite many preconceived notions that practitioners have about diabetic patients, research has shown that these patients usually tolerate revascularization extremely well and do not suffer from increased mortality or diminished graft patency.[61] The ultimate goal of any revascularization procedure should be to restore adequate perfusion distal to an occlusion to relieve symptoms and promote healing. The outflow target artery, or the site of the distal anastomosis, should be free of occlusive disease and should be in direct continuity with the arteries of the foot. Appropriate inflow and outflow vessels can be identified for revascularization via arteriography, including images of the pedal arteries. Selection of the appropriate conduit for the arterial reconstruction should be autologous vein because of superior patency rates compared with prosthetic grafts.[62] The decision to use saphenous vein in the reversed (*in situ*) or nonreversed position, as well as contralateral saphenous vein or arm vein when necessary, depends on the specific clinical circumstances and patient anatomy.

Occasionally, patients with diabetes may require reconstruction of the aortoiliac segment or will have disease limited to the superficial femoral artery above the knee, and in these cases prosthetic grafts are well suited to achieve this goal. However, because diabetic patients typically suffer from tibial and peroneal disease, the most effective bypass often involves restoring blood flow to the dorsalis pedis artery or the posterior tibial artery at the ankle using an autologous vein as the conduit.[61] Numerous reports have demonstrated the durability and success of distal origin grafts that obtain inflow from the popliteal artery and bypass over a shorter length to the arteries in the foot.[63,64] Although distal bypasses are technically challenging, success is attainable, especially when the surgeon maintains a flexible approach to preparation and placement of the vein graft.[65]

In one of the most comprehensive studies regarding dorsalis pedis revascularization in diabetic patients, Pomposelli and coworkers[66] reported results from more than 1000 bypasses spanning a decade. Primary patency, secondary patency, and limb salvage rates were 56.8%, 62.7%, and 78.2%, respectively, at 5 years, and 37.7%, 41.7%, and 57.7%, respectively, at 10 years. Patient survival was 48.6% and 23.8% at 5 and 10 years, respectively, and perioperative mortality was only 0.9%. The popliteal artery was the source of inflow in 53.2% of patients. Overall, pedal arterial reconstruction has proved to be a safe and durable revascularization alternative in diabetic patients and should be used to avoid ischemic diabetic foot complications.

Endovascular Procedures

Although surgical reconstruction remains the current gold standard for diabetic foot revascularization, recent advances in technology have allowed endovascular interventions to become integrated into the management strategies employed by vascular surgeons and interventionists. Research has definitively shown that balloon angioplasty and stenting are well suited to focal, short-segment iliac stenoses or occlusions, which exist in 10% to 20% of patients with diabetes.[67] Percutaneous intervention can be performed easily in the same setting as the preoperative diagnostic arteriogram.

Regarding outflow bypass procedures, the morbidity of open surgery can be quite significant and is not limited simply to local wound complications or myocardial infarctions. Readmission rates, reoperation, slow time to healing,[68] and time spent in rehabilitation must be factored into the risk-benefit analysis.[69] Nicoloff and colleagues[70] determined that the so-called "ideal outcome" (patent graft, healed wound, no additional operations in a fully ambulatory patient who can sustain independent living) was attainable only 14% to 22% of the time. When the potential pitfalls accompanying traditional surgical approaches to limb salvage are factored in, as well as the overall poor health and life expectancy of patients with peripheral vascular disease, less invasive endovascular therapy can represent a very attractive alternative.

Unfortunately, the literature supporting endovascular interventions consists mainly of individual case series and retrospective reviews and does not compare with the research describing traditional open bypasses. The Transatlantic Inter-Society Consensus Working Group restratified femoropopliteal and tibial lesions in 2007[67] and made recommendations for therapy based on lesion type (stenosis vs. occlusion), location, and length. Although the literature relating to endovascular interventions for critical limb ischemia is mixed in terms of positive and negative results,[71,72] the best and most recent scientific attempt to compare open and endovascular interventions was the Bypass Versus Angioplasty in Severe Ischemia of the Leg trial.[73] Although only 42% of the patients in this trial were known to have diabetes, the level of ischemia was comparable with the typical disease patterns seen in this patient population. Overall, the trialists concluded that, although the strategies are roughly equivalent at medium-term follow-up with regard to mortality and amputation-free survival, angioplasty should be used first for patients with significant comorbidities and with a life expectancy of less than 1 to 2 years, and bypass first for patients expected to live longer than 2 years. Longer follow-up from this trial will continue to provide statistically guided information in the years to come.

Amputations

When efforts at limb preservation prove unsuccessful, or if a situation becomes life-threatening, amputation must be considered. Minor amputations of toes, or transmetatarsal amputation, is practical after infection control and revascularization have been achieved. For cases in which the extent of tissue loss precludes a functional foot, when there are nonhealing wounds in the setting of patent grafts, and for control of sepsis, amputation below the knee is sometimes necessary. Surgeons should always strive to preserve the knee joint because of its functional significance in a patient's potential for rehabilitation. Above-the-knee amputations are reserved for debilitated patients with severe tissue loss or with no capacity to ambulate. Overall, because of modern advances in prostheses coupled with aggressive approaches to rehabilitation, amputation should be viewed as an acceptable modality to treat diabetic foot complications, and not as a treatment failure.

KEY POINTS

- The goal of all foot care strategies is to achieve complete healing of the "skin envelope" as quickly as possible while maintaining a functional walking foot. When medical management is not enough, different surgical management strategies are available to treat the diabetic foot. These strategies range from wide debridement and drainage procedures to arterial reconstruction or amputation. When efforts at limb preservation prove unsuccessful, or if a situation becomes life-threatening, amputation must be considered.

MULTIDISCIPLINARY APPROACH

As geographic and demographic differences persist in the delivery of care for patients at high risk for amputation,[74] a standardized, evidence-based approach is needed. A multidisciplinary approach to limb preservation has been repeatedly shown to reduce amputations,[75-79] with a growing number of national and international bodies supporting the establishment of these teams.[80,81] Because no one specialist possesses all the skills necessary for managing complex diabetic foot and vascular patients, a team of specialists with the required skills is necessary. For a multidisciplinary program to succeed, there must be a coordinated effort of physicians, including podiatric and vascular surgeons, nurses, allied health professionals, and administrators dedicated to the common goal of preserving functional limbs.

FUTURE TRENDS

Future trends in diabetic foot and vascular care must focus on the organization and integration of care, with an emphasis placed on identifying key points in healing and potential warning signs of impending recurrence. Technology that is portable, durable, automated, intelligent, ubiquitous, integrable, and affordable will have a significant impact on wound diagnostics and prevention programs.[49,82]

Recent advances in technology, including remote temperature monitoring (RTM), pressure monitoring and wearable technologies, have the potential to fundamentally change the management of patients with diabetes. One example is the Podimetrics RTM system (Podimetrics, Inc., Somerville, MA), which uses a wireless floor mat to record thermograms and detect temperature asymmetry that can be predictive of skin ulceration (Fig. 46.4). In a multicenter trial evaluating 129 participants with diabetes and history of DFU, the RTM system detected 97% of observed DFUs nearly 6 weeks before the clinical appearance of an ulcer. Of note, 86% of the study cohort used the system at least 3 days per week on average during the study.[83]

Another multicenter study found a significant decrease in resource utilization within a large integrated healthcare system using the RTM system. The authors reported a significant reduction in inpatient admissions, emergency room visits, incidence of ulcer recurrence, and amputations for patients who participated in a prevention program with RTM.[84] Further studies investigating the usefulness and practicality of remote and wearable technologies for the treatment of diabetic foot and vascular complications are needed.

KEY POINTS

- Future trends in diabetic foot and vascular care must focus on the organization and integration of care, with an emphasis placed on identifying key points in healing and potential warning signs of impending recurrence. Technology that is portable, durable, automated, intelligent, ubiquitous, integrable, and affordable will have a significant impact on wound diagnostics and prevention programs.

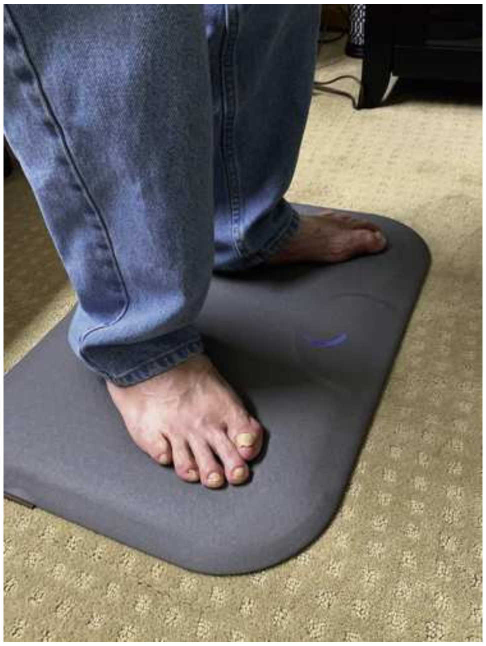


Fig. 46.4 Recent advances in technology, including a wireless remote temperature monitoring smart mat, have the potential to transform the way in which patients at high risk for diabetic foot complications are managed in the future.

REFERENCES

1. Armstrong DG, Boulton AJ, Bus SA. Diabetic foot ulcers and their recurrence. *N Engl J Med*. 2017;376:2367–2375.
2. Driver VR, Fabbi M, Lavery LA, et al. The costs of diabetic foot: the economic case for the limb salvage team. *J vasc surg*. 2010;52:17S–22S.
3. American Diabetes Association. Economic costs of diabetes in the US in 2012. *Diabetes Care*. 2013;36:1033–1046.
4. Armstrong DG, Swerdlow MA, Armstrong AA, et al. Five year mortality and direct costs of care for people with diabetic foot complications are comparable to cancer. *J Foot Ankle Res*. 2020;13:1–4.
5. Barshes NR, Sigireddi M, Wrobel JS, et al. The system of care for the diabetic foot: objectives, outcomes, and opportunities. *Diabetic Foot Ankle*. 2013;4:21847.
6. Zhang Y, Lazzarini PA, McPhail SM, et al. Global disability burdens of diabetes-related lower-extremity complications in 1990 and 2016. *Diabetes Care*. 2020;43:964–974.
7. Lazzarini PA, Pacella RE, Armstrong DG, et al. Diabetes-related lower-extremity complications are a leading cause of the global burden of disability. *Diabetic Med*. 2018;35:1297–1299.
8. Prompers L, Huijberts M, Apelqvist J, et al. High prevalence of ischaemia, infection and serious comorbidity in patients with diabetic foot disease in Europe. Baseline results from the Eurodiale study. *Diabetologia*. 2007;50:18–25.
9. Lavery LA, Armstrong DG, Murdoch DP, et al. Validation of the Infectious Diseases Society of America's diabetic foot infection classification system. *Clin Infect Dis*. 2007;44:562–565.
10. Ndosi M, Wright–Hughes A, Brown S, et al. Prognosis of the infected diabetic foot ulcer: a 12–month prospective observational study. *Diabetic Med*. 2018;35:78–88.
11. Lavery LA. Discussion: Off-loading the diabetic foot for ulcer prevention and healing. *Plast Reconstr Surg*. 2011;127:257S. –258S.
12. Armstrong DG, Kanda VA, Lavery LA, et al. Mind the gap: Disparity between research funding and costs of care for diabetic foot ulcers. *Diabetes Care*. 2013;36:1815–1817.
13. Parkhouse N, Le Quesne PM. Impaired neurogenic vascular response in patients with diabetes and neuropathic foot lesions. *N Engl J Med*. 1988;318:1306–1309.
14. Peters EJG, Lipsky BA. Diagnosis and management of infection in the diabetic foot. *Foot Diabetes*. 2020:265–286.
15. Yan SF, Ramasamy R, Naka Y, et al. Glycation, inflammation, and RAGE: A scaffold for the macrovascular complications of diabetes and beyond. *Circ Res*. 2003;93:1159–1169.
16. Lyons TJ, Jenkins AJ. Glycation, oxidation, and lipoxidation in the development of the complications of diabetes: A carbonyl stress hypothesis. *Diabetes Rev*. 1997;5:365.
17. Morbach S, Furchert H, Gröblinghoff U, et al. Long-term prognosis of diabetic foot patients and their limbs: Amputation and death over the course of a decade. *Diabetes Care*. 2012;35:2021–2027.
18. Goldenberg S, Alex M, Joshi RA, et al. Nonatheromatous peripheral vascular disease of the lower extremity in diabetes mellitus. *Diabetes*. 1959;8:261–273.
19. LoGerfo FW, Coffman JD. Vascular and microvascular disease of the foot in diabetes: Implications for foot care. *N Engl J Med*. 1984;311:1615–1619.
20. Leinonen H, Matikainen E, Juntunen J. Permeability and morphology of skeletal muscle capillaries in type 1 (insulin-dependent) diabetes mellitus. *Diabetologia*. 1982;22:158–162.
21. Rayman G, Williams SA, Spencer PD, et al. Impaired microvascular hyperaemic response to minor skin trauma in type I diabetes. *Br Med J (Clin Res Ed)*. 1986;292:1295–1298.
22. Veves A, Murray HJ, Young MJ, et al. The risk of foot ulceration in diabetic patients with high foot pressure: A prospective study. *Diabetologia*. 1992;35:660–663.
23. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. *JAMA*. 2005;293:217–228.
24. Lavery LA, Armstrong DG, Wunderlich RP, et al. Risk factors for foot infections in individuals with diabetes. *Diabetes Care*. 2006;29:1288–1293.
25. Yancey P, Brand PW. The gift of pain: why we hurt & what we can do about it. Michigan, USA: Zondervan; 1997.
26. Frykberg RG. An evidence-based approach to diabetic foot infections. *Am J Surg*. 2003;186:44–54.
27. Lipsky BA, Weigelt JA, Gupta V, et al. Skin, soft tissue, bone, and joint infections in hospitalized patients: Epidemiology and microbiological, clinical, and economic outcomes. *Infect Control Hospital Epidemiol*. 2007;28:1290–1298.
28. Katz DE, Friedman ND, Ostrovski E, et al. Diabetic foot infection in hospitalized adults. *J Infect Chemother*. 2016;22:167–173.
29. Balsells M, Viade J, Millan M, et al. Prevalence of osteomyelitis in non-healing diabetic foot ulcers: Usefulness of radiologic and scintigraphic findings. *Diabetes Res Clin Pract*. 1997;38:123–127.
30. Grayson ML, Gibbons GW, Balogh K, et al. Probing to bone in infected pedal ulcers: A clinical sign of underlying osteomyelitis in diabetic patients. *JAMA*. 1995;273:721–723.
31. Meyr AJ, Singh S, Zhang X, et al. Statistical reliability of bone biopsy for the diagnosis of diabetic foot osteomyelitis. *J Foot Ankle Surg*. 2011;50:663–667.
32. Gardner AW, Parker DE, Montgomery PS, et al. Efficacy of quantified home-based exercise and supervised exercise in patients with intermittent claudication: A randomized controlled trial. *Circulation*. 2011;123: 491–498.
33. Hingorani A, LaMuraglia GM, Henke P, et al. The management of diabetic foot: a clinical practice guideline by the society for vascular surgery in collaboration with the American Podiatric Medical Association and the Society for Vascular Medicine. *J Vasc Surg*. 2016;63:3S–21S.
34. Chung NS, Han SH, Lim SH, et al. Factors affecting the validity of ankle-brachial index in the diagnosis of peripheral arterial obstructive disease. *Angiology*. 2010;61:392–396.
35. Wukich DK, Shen W, Raspovic KM, et al. Noninvasive arterial testing in patients with diabetes: a guide for foot and ankle surgeons. *Foot Ankle Int*. 2015;36:1391–1399.
36. Mills Sr JL, Conte MS, Armstrong DG, et al. The society for vascular surgery lower extremity threatened limb classification system: risk stratification based on wound, ischemia, and foot infection (WIfI). *J Vasc Surg*. 2014;59:220–234.
37. Mills Sr JL. The application of the society for vascular surgery wound, ischemia, and foot infection (WIfI) classification to stratify amputation risk. *J Vasc Surg*. 2017;65:591–593.
38. Khan T, Plotkin A, Magee GA, et al. Functional ambulatory status as a potential adjunctive decision-making tool following wound, level of ischemia, and severity of foot infection assessment. *J Vasc Surg*. 2020;72:738–746.
39. Serino F, Fernando M, Montero-Baker M, et al. Toward an interdisciplinary approach to diabetic limbs in the era of functional limb preservation: "Can we preserve this limb?" meets "Should we preserve this limb?". *J Am Podiatr Med Assoc*. 2021:20–165.
40. American Diabetes Association. Consensus development conference on diabetic foot wound care: 7–8 April 1999, Boston, Massachusetts. American Diabetes Association. *Diabetes Care*. 1999;22:1354–1360.
41. Bus SA, Lavery LA, Monteiro–Soares M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update). *Diabetes Metab Res*. 2020;36:e3269.
42. Feng Y, Schlösser FJ, Sumpio BE. The Semmes Weinstein monofilament examination is a significant predictor of the risk of foot ulceration and amputation in patients with diabetes mellitus. *J Vasc Surg*. 2011;53:220–226.
43. Rayman G, Vas PR, Baker N, et al. The Ipswich touch test. *Diabetes Care*. 2011;34:1517–1518.
44. Lipsky BA, Berendt AR, Cornia PB, et al. 2012 Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. *Clin Infect Dis*. 2012;54:e132–e173.
45. Bamberger DM, Daus GP, Gerding DN. Osteomyelitis in the feet of diabetic patients: Long-term results, prognostic factors, and the role of antimicrobial and surgical therapy. *Am J Med*. 1987;83:653–660.
46. Tice AD, Hoaglund PA, Shoultz DA. Outcomes of osteomyelitis among patients treated with outpatient parenteral antimicrobial therapy. *Am J Med*. 2003;114:723–728.

47. Armstrong DG, Lavery LA. Diabetic Foot Study Consortium. Negative pressure wound therapy after partial diabetic foot amputation: a multicentre, randomised controlled trial. *Lancet*. 2005;366:1704–1710.
48. Blume PA, Walters J, Payne W, et al. Comparison of negative pressure wound therapy using vacuum-assisted closure with advanced moist wound therapy in the treatment of diabetic foot ulcers: a multicenter randomized controlled trial. *Diabetes Care*. 2008;31:631–636.
49. Isaac AL, Armstrong DG. Negative pressure wound therapy and other new therapies for diabetic foot ulceration: the current state of play. *Med Clin*. 2013;97:899–909.
50. Isaac AL, Rose J, Armstrong DG. Mechanically powered negative pressure wound therapy as a bolster for skin grafting. *Plast Reconstr Surg Global Open*. 2014;2:e103.
51. Cavanagh PR, Bus SA. Off-loading the diabetic foot for ulcer prevention and healing. *J Am Podiatr Med Assoc*. 2010;100:360–368.
52. Bus SA, Van Deursen RW, Armstrong DG, et al. Footwear and offloading interventions to prevent and heal foot ulcers and reduce plantar pressure in patients with diabetes: A systematic review. *Diabetes Metab Res*. 2016;32:99–118.
53. Wu SC, Jensen JL, Weber AK, et al. Use of pressure offloading devices in diabetic foot ulcers: Do we practice what we preach? *Diabetes Care*. 2008;31:2118–2119.
54. Armstrong DG, Isaac AL, Bevilacqua NJ, et al. Offloading foot wounds in people with diabetes. *Wounds*. 2014;26:13–20.
55. Lavery LA, Vela SA, Lavery DC, et al. Reducing dynamic foot pressures in high-risk diabetic subjects with foot ulcerations: a comparison of treatments. *Diabetes Care*. 1996;19:818–821.
56. Armstrong DG, Lavery LA, Kimbriel HR, et al. Activity patterns of patients with diabetic foot ulceration: patients with active ulceration may not adhere to a standard pressure off-loading regimen. *Diabetes Care*. 2003;26:2595–2597.
57. Katz IA, Harlan A, Miranda-Palma B, et al. A randomized trial of two irremovable off-loading devices in the management of plantar neuropathic diabetic foot ulcers. *Diabetes Care*. 2005;28:555–559.
58. Gibbons GW. The diabetic foot: Amputations and drainage of infection. *J Vasc Surg*. 1987;5:791–793.
59. Aulivola B, Pomposelli FB. Lower extremity arterial reconstruction in patients with diabetes mellitus: Principles of treatment. In: Veves A, Giurini J, LoGerfo F. (eds). *The Diabetic Foot. Contemporary Diabetes*. Totowa, NJ: Humana Press; 2012. Available from: https://doi.org/10.1007/978-1-61779-791-0_19.
60. Tannenbaum GA, Pomposelli Jr FB, Marcaccio EJ, et al. Safety of vein bypass grafting to the dorsal pedal artery in diabetic patients with foot infections. *J Vasc Surg*. 1992;15:982–990.
61. Pomposelli Jr FB, Marcaccio EJ, Gibbons GW, et al. Dorsalis pedis arterial bypass: Durable limb salvage for foot ischemia in patients with diabetes mellitus. *J Vasc Surg*. 1995;21:375–384.
62. Veith FJ, Gupta SK, Ascer E, et al. Six-year prospective multicenter randomized comparison of autologous saphenous vein and expanded polytetrafluoroethylene grafts in infrainguinal arterial reconstructions. *J Vasc Surg*. 1986;3:104–114.
63. Stonebridge PA, Tsoukas AI, Pomposelli Jr FB, et al. Popliteal-to-distal bypass grafts for limb salvage in diabetics. *Eur J Vasc Surg*. 1991;5:265–269.
64. Reed AB, Conte MS, Belkin M, et al. Usefulness of autogenous bypass grafts originating distal to the groin. *J Vasc Surg*. 2002;35:48–55.
65. Pomposelli FB, Jepsen SJ, Gibbons GW, et al. A flexible approach to infrapopliteal vein grafts in patients with diabetes mellitus. *Arch Surg*. 1991;126:724–729.
66. Pomposelli FB, Kansal N, Hamdan AD, et al. A decade of experience with dorsalis pedis artery bypass: Analysis of outcome in more than 1000 cases. *J Vasc Surg*. 2003;37:307–315.
67. Norgren L, Hiatt WR, Dormandy JA, et al. Inter-society consensus for the management of peripheral arterial disease (TASC II). *J Vasc Surg*. 2007;45:S5–S67.
68. Chung J, Bartelson BB, Hiatt WR, et al. Wound healing and functional outcomes after infrainguinal bypass with reversed saphenous vein for critical limb ischemia. *J Vasc Surg*. 2006;43:1183–1190.
69. Goshima KR, Mills Sr JL, Hughes JD. A new look at outcomes after infrainguinal bypass surgery: traditional reporting standards systematically underestimate the expenditure of effort required to attain limb salvage. *J Vasc Surg*. 2004;39:330–335.
70. Nicoloff AD, Taylor LM, McLafferty RB, et al. Patient recovery after infrainguinal bypass grafting for limb salvage. *J Vasc surg*. 1998;27:256–266.
71. Meltzer AJ, Evangelisti G, Graham AR, et al. Determinants of outcome after endovascular therapy for critical limb ischemia with tissue loss. *Ann Vasc Surg*. 2014;28:144–151.
72. Iida O, Soga Y, Hirano K, et al. Long-term results of direct and indirect endovascular revascularization based on the angiosome concept in patients with critical limb ischemia presenting with isolated below-the-knee lesions. *J Vasc Surg*. 2012;55:363–370.
73. Bradbury AW, Ruckley CV, Fowkes F, et al. Bypass Versus Angioplasty in Severe Ischaemia of the Leg (BASIL): Multicentre, randomised controlled trial. *Lancet*. 2005;366:1925–1934.
74. Henry AJ, Hevelone ND, Belkin M, et al. Socioeconomic and hospital-related predictors of amputation for critical limb ischemia. *J Vasc Surg*. 2011;53:330–339.
75. Van Gils CC, Wheeler LA, Mellstrom M, et al. Amputation prevention by vascular surgery and podiatry collaboration in high-risk diabetic and nondiabetic patients. the operation desert foot experience. *Diabetes Care*. 1999;22:678–683.
76. Wrobel JS, Robbins JM, Charns MP, et al. Diabetes-related foot care at 10 veterans affairs medical centers: must do's associated with successful microsystems. *Jt Comm J Qual Patient Saf*. 2006;32:206–213.
77. Driver VR, Madsen J, Goodman RA. Reducing amputation rates in patients with diabetes at a military medical center: the limb preservation service model. *Diabetes Care*. 2005;28:248–253.
78. Armstrong DG, Bharara M, White M, et al. The impact and outcomes of establishing an integrated interdisciplinary surgical team to care for the diabetic foot. *Diabetes Metab Res*. 2012;28:514–518.
79. Krishnan S, Nash F, Baker N, et al. Reduction in diabetic amputations over 11 years in a defined UK population: benefits of multidisciplinary team work and continuous prospective audit. *Diabetes Care*. 2008;31:99–101.
80. Sumpio BE, Armstrong DG, Lavery LA, et al. The role of interdisciplinary team approach in the management of the diabetic foot: a joint statement from the Society for Vascular Surgery and the American Podiatric Medical Association. *J Am Podiatr Med Assoc*. 2010;100:309–311.
81. Conte MS, Bradbury AW, Kolh P, et al. Global vascular guidelines on the management of chronic limb-threatening ischemia. *Eur J Vasc Endovasc Surg*. 2019;58:S1–S109.
82. Armstrong DG, Giovinco NA. Diagnostics, theragnostics, and the personal health server: Fundamental milestones in technology with revolutionary changes in diabetic foot and wound care to come. *Foot Ankle spec*. 2011;4:54–60.
83. Frykberg RG, Gordon IL, Reyzelman AM, et al. Feasibility and efficacy of a smart mat technology to predict development of diabetic plantar ulcers. *Diabetes Care*. 2017;40:973–980.
84. Isaac A, Swartz T, Miller M, et al. Lower resource utilization for patients with healed diabetic foot ulcers during participation in a prevention program with foot temperature monitoring. *BMJ Open Diabet Res Care*. 2020;8:e001440.

47

Parathyroid Hormone and the Parathyroid Hormone Receptor Type 1 in the Regulation of Calcium and Phosphate Homeostasis and Bone Metabolism

Thomas J. Gardella, Harald Jüppner, Michael Mannstadt, Edward M. Brown, Henry M. Kronenberg, and John T. Potts, Jr.

OUTLINE

Regulation of Mineral Ion Homeostasis—General Considerations, 802
Parathyroid Hormone, 803
Parathyroid Hormone Chemistry, 804
Evolution of Parathyroid Hormone Ligands, 805
The Parathyroid Hormone Gene and its mRNA, 805
Parathyroid Hormone Biosynthesis and Intraglandular Processing, 805
Regulation of Parathyroid Hormone Gene Expression, 806
Regulation of Parathyroid Hormone Secretion, 806
Physiologic Control of Parathyroid Hormone Secretion by Extracellular Calcium Ion, 806
Mechanism of Extracellular Calcium Ion Sensing by Parathyroid Cells and Other Cells Involved in Mineral Ion Homeostasis, 807
1, 25(OH)$_2$D, Phosphate, and Other Factors Regulating Parathyroid Hormone Secretion, 809
Metabolism of Parathyroid Hormone, 809
Parathyroid Hormone–Dependent Regulation of Mineral Ion Homeostasis Mediated Through the Parathyroid Hormone Receptor Type-1, 810
Actions of Parathyroid Hormone on the Kidney, 810
Renal Calcium Reabsorption, 810
Regulation of 1α- and 24-Hydroxylase Activity, 810
Renal Phosphate Transport, 810
Actions of Parathyroid Hormone on Bone, 811
Molecular Properties of The Parathyroid Hormone Receptor Type-1, 812
Gene Structure, Protein Topology, and Evolution, 812
Three-Dimensional Structures of the Parathyroid Hormone Receptor Type-1, 813
Conformationally Selective Parathyroid Hormone Ligands and Prolonged Signaling, 814
Disease Mutations in the PTHR-1 and its Ligands, 814
Small-Molecule Ligands for the Parathyroid Hormone Receptor Type-1, 815
The Parathyroid Hormone Receptor Type-2 and Tip39, 815
Summary, 815

Parathyroid hormone (PTH) and PTH-related peptide (PTHrP), along with other calciotropic hormones, play critical roles in calcium and bone biology. First discovered as a calcium-regulating hormone in the 1920s, PTH is secreted by the parathyroid glands and is one of the most important regulators of blood calcium concentration in all terrestrial vertebrate species from amphibians to mammals. PTH is produced by discrete endocrine glands. In contrast, PTHrP is produced as a paracrine/autocrine factor in many different adult and fetal tissues and, unlike PTH, has multiple functions (see Chapter 48).

PTH and PTHrP most likely evolved from a common ancestral precursor. Despite this common evolutionary origin, both peptides share only limited overall amino acid sequence identity, yet their N-terminal regions are sufficiently homologous to enable them to bind to and activate a common G protein–coupled receptor (GPCR), the PTH/PTHrP receptor (also referred to as PTHR-1). This chapter reviews (1) the comparative chemistry of PTH, the structure of its gene, and the key elements of function within the hormone; (2) the evolutionary history of the PTH family of ligands, as well as the PTH receptor; (3) the role of PTH in calcium and phosphate homeostasis and in bone turnover; (4) new molecular structures obtained for the PTHR-1 and conformational changes involved in ligand-induced activation; (5) disease-causing mutations in the PTH receptor and ligands; and (6) the PTH receptor type-2 and its ligand, TIP39.

REGULATION OF MINERAL ION HOMEOSTASIS—GENERAL CONSIDERATIONS

To ensure a multitude of essential cellular functions, the concentration of extracellular calcium ion ($Ca^{2+}{}_o$) is maintained within a narrow range. In terrestrial vertebrates, calcium is necessary for adequate mineralization of the skeleton, which provides mechanical support and protection for internal organs and acts as levers for the various muscle groups involved in locomotion. Because of its high calcium content, representing 99% of the body's supply, the skeleton also serves as the most important reservoir from which calcium can be rapidly mobilized. Because food intake, and thus the nutritional supply of calcium, is usually discontinuous, intestinal calcium absorption occurs only intermittently. Maintenance of a constant blood calcium concentration thus constitutes a major homeostatic challenge, which led during evolution to the development of highly efficient mechanisms to increase intestinal calcium absorption, reduce urinary calcium losses, and facilitate, if necessary, rapid mobilization of calcium from the skeletal reservoir (see Chapter 52).

Unlike most terrestrial vertebrates, marine animals are usually exposed to the high environmental calcium concentration of seawater (10 mM) and thus had to adopt mechanisms by which $Ca^{2+}{}_o$ could be reduced. In contrast, both terrestrial and marine animals had to develop efficient mechanisms to enhance intestinal phosphate absorption and to prevent urinary losses of this essential mineral, leading in the mammalian kidney to an impressive capacity to retain phosphate. To reduce blood calcium concentrations, fish use stanniocalcin, which is produced by the corpuscles of Stannius, as well as several other hormonal factors. A putative mammalian homolog of stanniocalcin has shown similar properties when tested in rodents, but it remains uncertain whether this peptide hormone has a significant physiologic role in mammalian mineral ion homeostasis. The corpuscles of Stannius are located adjacent to the pronephros, and are furthermore the only tissue in fish expressing fibroblast growth factor 23 (FGF23), which is a major regulator of phosphate homeostasis in mammals.[1-3] Current data (see later) indicate that FGF23, the FGF23 coreceptor Klotho, dentin matrix protein 1, ecto-nucleotide pyrophosphatase/phosphodiesterase 1, and matrix extracellular phosphoglycoprotein are more likely to be involved physiologically in the regulation of mammalian phosphate homeostasis (see Chapter 53).

PARATHYROID HORMONE

PTH and the active form of vitamin D, 1,25-dihydroxyvitamin D ($1,25(OH)_2D$) are the principal physiologic regulators of calcium homeostasis in humans and all terrestrial vertebrates. Synthesis and secretion of PTH are stimulated by any decrease in blood calcium, and, conversely, secretion of the hormone is inhibited by an increase in blood calcium. This rapid negative feedback regulation of PTH production, along with the biologic actions of the hormone on different target tissues, represents the most important homeostatic mechanism for the minute-to-minute control of calcium concentration in the extracellular fluid (ECF). The vital importance of PTH in maintaining calcium homeostasis and the rapid effects of PTH on blood levels of calcium, as well as phosphorous, are illustrated by the experiments performed in rats shown in Fig. 47.1.

In contrast to the rapid actions of PTH, $1,25(OH)_2D$ is of critical importance for long-term, day-to-day, and week-to-week calcium balance (see Chapter 50). The actions of both hormones are coordinated, and each influences the synthesis and secretion of the other. Calcitonin, the third of the calciotropic hormones known to be important in the regulation of vertebrate mineral ion homeostasis (see Chapter 49), may be vestigial in humans with respect to calcium homeostasis and will not be discussed in this brief review of physiology.

At least three distinct, but coordinated, actions of PTH increase the flow of calcium into the ECF and thus increase the concentration of blood calcium. Through its rapid actions on kidney and bone, which are all mediated through the PTHR-1 and subsequent secondary message molecules, such as cyclic 3′,5′-adenosine monophosphate (cAMP), that are activated in specific and highly specialized cells, PTH increases the release of calcium from bone, reduces the renal clearance of calcium, and stimulates the production of $1,25(OH)_2D$ by activating the gene encoding 25-hydroxyvitamin D-1α-hydroxylase (1α-hydroxylase) in the kidney. The relative importance of the first two actions of PTH on the rapid, minute-to-minute regulation of calcium is not definitively resolved, but most physiologists have stressed the importance of the effects of PTH on bone in maintaining hour-to-hour calcium homeostasis in the ECF. Several lines of evidence, such as that provided by calcium kinetic analyses, indicate a transfer between ECF and bone of as much as 500 mg of calcium daily, which is equivalent to one fourth to one half the total ECF calcium content. Besides regulating this transfer of calcium from bone through direct breakdown of bone tissue (mineral and matrix), PTH influences the rates of exchange of calcium adsorbed to the surface of bone; this exchangeable calcium pool can be stimulated to provide a rapid and substantial rate of entry of calcium into blood. In addition to these actions on bone, actions of PTH on the kidney may also be extremely important in the precise hourly regulation of ECF calcium. The third action of PTH on calcium homeostasis—namely, enhancement of intestinal calcium absorption—is indirect and involves the synthesis of $1,25(OH)_2D$ from the biologically inactive precursor 25(OH)D. However, it is difficult to quantitatively analyze or proportionately contrast the relative physiologic importance of the direct and indirect actions of PTH on the three principal target tissues: kidney, bone, and intestine.

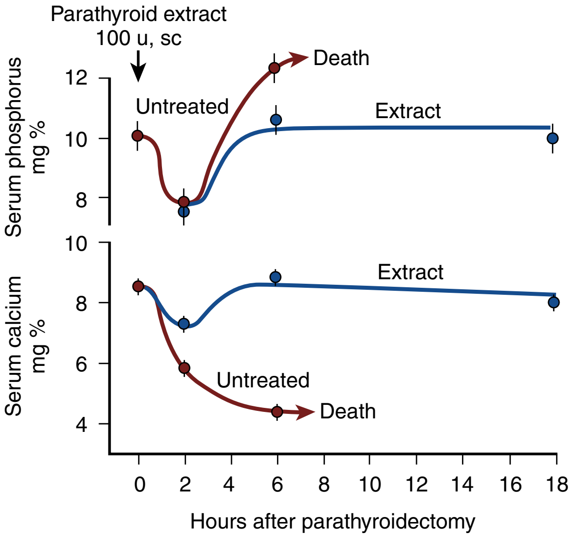


Fig. 47.1 **Biological actions of parathyroid hormone and calcium homeostasis.** Rate of change in blood phosphorus and calcium levels in rats with stressed calcium homeostasis (low-calcium diet) after parathyroidectomy without treatment or with parathyroid hormone extract given in addition at the time of parathyroidectomy. Rapid and usually fatal hypocalcemia and hyperphosphatemia result within hours unless hormone is given. *sc,* Subcutaneous. (Data from Munson PL. Studies on the role of the parathyroids in calcium and phosphorus metabolism. *Ann N Y Acad Sci.* 1955;60:776–796.)

The complexity of bone as a tissue and the many detectable rates of exchange of calcium between the skeleton and the ECF have made the action of PTH on the skeleton difficult to analyze. The physical state of calcium in blood is dynamic; much of the calcium is present as chelates or is bound to plasma proteins, but this bound or chelated form exchanges with the free ionized form of calcium that is also present in the blood at rates dependent on a variety of factors, such as the concentrations of counterions, such as phosphate (for detailed review see Chapter 52). Because actual filtered loads depend on the ratio of free and bound forms of calcium, it is difficult to calculate renal calcium clearance accurately. The different PTH-dependent actions to promote calcium entry into the ECF are most clearly defined in conditions of

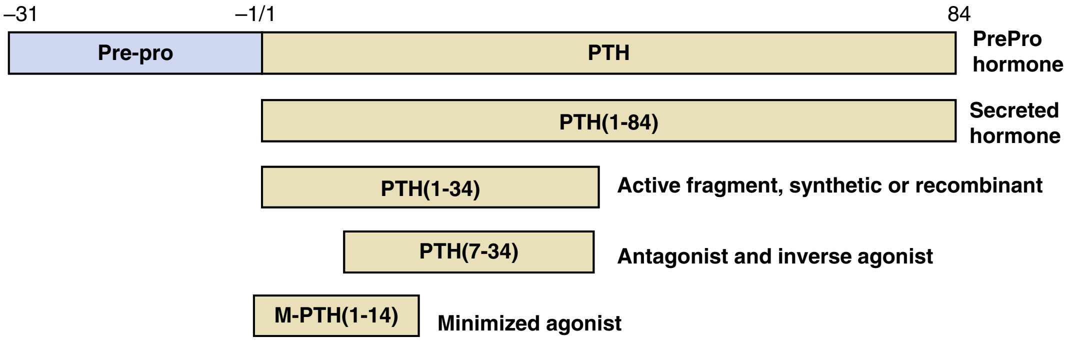


Fig. 47.2 **Parathyroid hormone (PTH) polypeptide and peptide derivatives.** Schematized are the biologically and pharmacologically relevant forms of the PTH polypeptide, including the nascent prepro PTH produced by parathyroid gland cells; mature PTH(1–84) that is secreted after removal of the prepro sequence; the PTH(1–34) fragment produced by chemical synthesis or recombinantly and which retains most if not all of the biological activity of intact PTH(1–84); the PTH(7–34) peptide that binds to but does not activate the PTH receptor type-1 and thus can function as a competitive antagonist, and, when modified with D-tryptophan at position 12, as an inverse agonist on constitutively active mutant PTH receptors, such as those found in patients with Jansen's metaphyseal chondrodysplasia,[88] and the N-terminal PTH(1–14) fragment that, when modified (M) with affinity-enhancing substitutions, functions *in vitro* as a potent agonist.

deficiency or excess of PTH, such as during experiments in animals or during controlled observations in patients with disorders of parathyroid gland function. The experimental data in these extremes abundantly affirm the crucial calcium homeostatic role of PTH. However, because of continuous and rapid adjustments in mineral ion concentration, it can be difficult to observe the consequences of hormone action under normal physiologic conditions. For example, the rate of PTH secretion changes continually and rapidly, so that the controlled variable, calcium, remains constant, and it may, therefore, be difficult to experimentally detect small corrective changes.

Teleologically, the action of PTH on the regulation of blood phosphate concentration in terrestrial species is best understood as a secondary, rather than a homeostatic, action. Phosphate is abundant in the food chain in terrestrial existence. Phosphate deficiency, unlike calcium deficiency, in the absence of specific organ dysfunction is, therefore, an unlikely environmental challenge (see Chapter 53 for detailed review of the regulation of phosphate homeostasis). To correct a deficiency in calcium, mineral stores in bone can be rapidly dissolved; such activity results, however, in the simultaneous liberation of ionic calcium and phosphate. Because a high blood phosphate level tends to lower the calcium concentration through multiple mechanisms, the rise in blood calcium that occurs after bone dissolution (desirable homeostatically) is, therefore, beneficial only if the concomitant increase in blood phosphate concentration (undesirable) can be rapidly corrected. To maximize the control of calcium homeostasis, PTH thus has divergent actions on renal tubular handling of the two mineral ions: it increases the retention of calcium and, at the same time, diminishes the reabsorption of phosphate. Through these mechanisms, namely, increased renal phosphate clearance to prevent hyperphosphatemia and increased tubular calcium reabsorption, PTH guarantees that an elevation in blood calcium results from the increased release of calcium from bone. The renal action of PTH on phosphate homeostasis is biologically predominant over the increased phosphate flux from bone. Consequently, parathyroidectomy or renal resistance to PTH, as in patients with pseudohypoparathyroidism or renal failure, can lead not only to severe or even fatal hypocalcemia, but also to an increase in blood phosphate and a marked reduction in urinary phosphate excretion (Fig. 47.1). This finding demonstrates the importance of the PTH-dependent action on phosphate homeostasis in the kidney, which becomes particularly important in disease states when high bone turnover is the result of dietary calcium deficiency or lack of biologically active vitamin D.

Parathyroid Hormone Chemistry

The first extracts from bovine parathyroid glands were described in the 1920s, and the content of biologically active PTH was assessed by their hypercalcemic and phosphaturic properties.[4,5] In the 1950s, extraction procedures were improved such that it became possible to purify sufficient quantities for biochemical analysis.[6] The amino acid sequences of the active purified PTH peptide from bovine, porcine, and human glands were obtained in the 1970s.[7,8] The sequences of the intact bovine, porcine, and human hormones each revealed a polypeptide chain of 84 amino acids (schematized in Fig. 47.2). Based on the derived amino acid sequences of the purified PTH(1–84) polypeptide, and clues derived from peptide fragmentation studies, the 1–34 fragments of PTH from several different species were chemically synthesized, and their biological activities *in vitro* and *in vivo* were found to be nearly equivalent to those of the corresponding intact hormone purified from the same species.[9] Molecular cloning techniques then led to the deduction of the amino acid sequences of PTH in several mammals[10] and subsequently in many other diverse species, including birds and fish, and more recently the identification of plausible PTH-like sequences in lower vertebrates and even invertebrates, as discussed in a later section of this chapter.

The potency of the PTH(1–34) fragment, observed both *in vitro* and *in vivo*, included its capacity to increase bone mass in animals. The latter findings led to the development of PTH(1–34) as a treatment for osteoporosis, and since 2001 the recombinant form of this peptide, teriparatide, has been in use as an effective therapy for that disease.[11] Extensive analysis of synthetic PTH peptides and a wide variety of derivative analogs (Fig. 47.2) in various types of pharmacological assays has led to the identification of the key structural elements in the molecule that determine its capacity to bind to the target PTH receptor and induce receptor activation responses that lead to downstream biologic effects. Truncated PTH fragment peptides, such as PTH(3–34), PTH(7–34), and PTH(15–34), established that the N- and C-terminal regions of PTH(1–34) contain the principal determinants of PTH receptor binding affinity and receptor activation potency, respectively. Analogs of the PTH(7–34) fragment thus function as competitive antagonists at the PTHR-1. Certain modified N-terminal peptides, such as the M-PTH(1–14) analog, have also been developed that contain affinity-enhancing modification and thus can adequately stimulate signaling responses in cells, further confirming the importance of the N-terminal region of PTH for receptor activation. Fig. 47.2 schematizes the various PTH analogs that help demonstrate the bipartite domain structure of the PTH(1–34) bioactive peptide. A more precise view of the mechanism by which PTH peptides interact with the PTH receptor has been made possible by the recent determination of the x-ray crystallography and cryogenic electron microscopy (cryo-EM) structures of the PTHR-1 obtained in complex with PTH(1–34) analog peptides, as discussed in a later section of this chapter.

KEY POINTS

- Parathyroid hormone (PTH) is an 84–amino acid polypeptide in which the first 34 amino acids comprise the biologically active region. Synthetic PTH(1–34) peptides are potent agonists at the PTH receptor type-1, a class B G protein–coupled receptor that mediates the key biological actions of PTH. PTH(7–34) peptides act as competitive antagonists.

Evolution of Parathyroid Hormone Ligands

Parathyroid glands, as discrete organs, first appeared in evolution with the amphibians, at the time of the migration of vertebrates from an aquatic to a terrestrial existence and the adaptation to an environment that is, by comparison to seawater, low in calcium. While fish lack parathyroid glands, immunologic and RNA hybridization studies provided evidence that PTH proteins are expressed in a variety of fish tissues, including the pituitary, brain, kidney, spinal cord, and developing jaw. Genomic studies have confirmed the presence of PTH coding sequences in a number of teleost fish, including the zebrafish (*Danio rerio*) and puffer fish (*Takifugu rubripes*).[12] The zebrafish genome contains six copies of *PTH*-like gene sequences, a number that presumably reflects a series of gene duplication events that occurred over the course of evolution of the vertebrate species.[13,14] Such gene duplication and gene loss models also explain the presence in humans and other higher vertebrates of genes encoding PTH, PTHrP, and a third homologous ligand called TIP39, which binds to the PTH-2 receptor subtype, as discussed in a later section of this chapter. Tracing the evolutionary path of the *PTH* and *PTHrP* gene-like sequences through diverse species can be confusing, given the occurrence of such genetic rearrangement events over time. Nevertheless, genome-based studies seem to provide ample support for the notion that PTH is an ancient hormone that was present in animals living at least as early as the time of the first appearance of the vertebrates, i.e., greater than approximately 500 million years ago. Indeed, gene sequences encoding putative peptides with at least weak or apparent similarity to vertebrate PTH ligands have even been identified in the genomes of several protochordates, including amphioxus (*Branchiostoma floridae*) and the sea squirt (*Ciona intestinalis*), as well as in the genomes of some insects.[13,15] These findings suggest that the ancestral *PTH*-like gene sequence emerged even prior to the divergence of the chordate lineages from the nonchordate metazoans.

Comparison of the amino acid sequences of PTH peptides from across species reveals a considerable amount of structural homology, particularly in the N-terminal portion of the hormone. Synthetic peptides corresponding to PTH-like sequences detected in the genomes of several lower vertebrates (e.g., zebrafish) as well as insects have been produced and shown to be active for inducing second-messenger (cAMP or intracellular calcium) signaling responses in experimental cells that heterologously express a target PTH receptor.[16] Such activities observed *in vitro* are at least consistent with a potential physiological role for the coded peptides in the parent organism, which also have been shown to contain genes for PTH-like receptors (discussed in a section that follows on the PTHR-1).[13,17] Analyses of messenger RNA or reporter gene expression levels in zebrafish indicate that at least some of the fish *PTH*-like genes are actively transcribed in certain tissues. The *PTH4* gene of zebrafish was thus found to be actively and exclusively expressed in the hypothalamus, both in early embryonic stages and in adults. This site of expression suggests an early neuroendocrine role for at least some PTH family ligands. Further functional analyses suggest that this role could involve a brain-to-bone signaling network that potentially ties in with mineral ion homeostasis, but such systems in lower vertebrates and invertebrates are not well understood.[12,14]

KEY POINTS

- Sequences encoding parathyroid hormone (PTH) or PTH-like peptides have been found in the genomes of several invertebrates, such as amphioxus (*Branchiostoma floridae*), indicating an early evolutionary origin.

The Parathyroid Hormone Gene and its mRNA

The human PTH gene consists of three exons located on chromosome 11p15.[10] The first exon is 85 bp in length and is noncoding. Exon 2 (90 bp) encodes most amino acids of the prepropeptide sequence. The third exon (612 bp) encodes the remainder of the propeptide sequence and all amino acids of the mature peptide. Exon 3 also contains the 3′ noncoding region. Two mRNAs that are 822 and 793 bp in length are derived from the human gene from the two transcriptional start sites that follow two different functional TATA boxes that are separated by 29 base pairs.[10] Genes encoding PTH ligands in the zebrafish exhibit at least some organizational similarity to mammalian PTH genes.[12,14]

KEY POINTS

- The parathyroid hormone gene in humans is located on chromosome 11p15 and is comprised of three exons, the first of which is noncoding, while exons 2 and 3 encode most of the prepro region and the remaining region to residue 84, respectively.

Parathyroid Hormone Biosynthesis and Intraglandular Processing

During the synthesis of the preproPTH molecule, the signal sequence, which comprises the 25–amino acid containing “pre” sequence, is cleaved off after entry of the nascent peptide chain into the intracisternal space bounded by the endoplasmic reticulum (ER). Heterozygous and homozygous missense mutations in this PTH leader sequence have been identified in patients with familial hypoparathyroidism.[18] Studies in cells indicate that such mutant polypeptide chains become trapped intracellularly, predominantly in the ER, leading to a marked upregulation of ER stress–responsive proteins (BiP and PERK) and the proapoptotic transcription factor CHOP. Such effects *in vivo* could likely give rise to apoptosis-mediated parathyroid cell death, thus explaining the observed hypoparathyroidism of the patients and the dominant nature of at least some of the mutations.

Subsequent to the removal of the presequence, the propeptide is transported to trans-Golgi network, where the prosequence (amino acid residues –6 through –1) is removed. This latter process may involve furin (paired basic amino acid cleaving enzyme) and/or proprotein convertase-7, which are both expressed in parathyroid tissue; their expression levels do not appear to be regulated by either calcium or $1,25(OH)_2D$. After removal of the basic prosequence, the mature polypeptide, PTH(1–84), is packaged into secretory granules. Two proteases, cathepsins B and H, are subsequently involved in the intraglandular generation of C-terminal PTH fragments from the intact hormone; no N-terminal PTH fragments appear to be released from the gland. Because small or intermediate size C-terminal fragments of PTH are unlikely to be involved in the regulation of calcium homeostasis, the intraglandular degradation of intact PTH is thought to represent an inactivating pathway, at least with regard to the regulation of mineral ion homeostasis. Consistent with this conclusion, hypercalcemia results in a substantial decrease in PTH secretion and, furthermore, favors the secretion of C-terminal PTH fragments (see later section on PTH metabolism). Several studies provide evidence that certain N-terminally truncated PTH fragments, such as PTH(7–84), can have hypocalcemic properties *in vivo*, and can furthermore reduce the formation of osteoclasts *in vitro*. While the underlying mechanisms for these observed effects remain to be established, it is at least conceivable that secreted C-terminal PTH fragments do contribute to the regulation of calcium homeostasis, at least under certain physiologic or pathophysiologic conditions.[19]

The pool of stored, intracellular PTH in parathyroid cells is small, and the cell must therefore have mechanisms to increase hormone synthesis and release in response to sustained hypocalcemia. One such adaptive mechanism is to reduce the intracellular degradation of the hormone, thereby increasing the net amount of intact, biologically active PTH that is available for secretion. During hypocalcemia the bulk of the hormone that is released from the parathyroid cell is intact PTH(1–84). As the level of $Ca^{2+}{}_o$ increases, a greater fraction of intracellular PTH is degraded, and with overt hypercalcemia most of the secreted immunoreactive PTH consists of biologically inactive C-terminal fragments.

KEY POINTS

- Parathyroid hormone is synthesized in cells of the parathyroid gland as a preprohormone, which is processed to the mature 1–84 peptide and stored in secretory vesicles until secretion is activated by low blood calcium levels.

Regulation of Parathyroid Hormone Gene Expression

Another adaptive mechanism of the parathyroid cell to sustained reductions in $Ca^{2+}{}_o$ is to increase cellular levels of PTH mRNA, a response that takes several hours. A reduction in $Ca^{2+}{}_o$ increases, whereas an elevation in $Ca^{2+}{}_o$ reduces, the cellular levels of PTH mRNA by affecting both its stability and the transcriptional rate of its gene. Phosphate ions also regulate, directly or indirectly, PTH gene expression. Hypophosphatemia and hyperphosphatemia in the rat lower and raise, respectively, the levels PTH mRNA through a mechanism that is independent of changes in $Ca^{2+}{}_o$ or $1,25(OH)_2D$. An elevated extracellular phosphate concentration could, thus, contribute importantly to the secondary hyperparathyroidism frequently encountered in patients with end-stage renal failure, who often have chronically elevated serum phosphate concentrations. FGF23, a major regulator of phosphate metabolism, has direct effects on the parathyroid gland to suppress the synthesis of PTH mRNA and protein.[1,3] In renal failure, in which high levels of both FGF23 and PTH are found, the parathyroid glands are resistant to the effects of FGF23 on PTH mRNA. This resistance is associated with decreased expression of FGF receptor 1 and α-Klotho, the receptor and coreceptor needed to transmit signals from FGF23, respectively, thereby explaining the paradox of increased levels of FGF23 and PTH in that condition.

Metabolites of vitamin D, principally $1,25(OH)_2D$, also play an important role in the long-term regulation of parathyroid function and may act at several levels: by suppressing the secretion of PTH and the transcription of its gene, by upregulating transcriptional activity of the genes encoding the calcium-sensing receptor (CaSR) (see later) and the vitamin D receptor (VDR), as well as by decreasing parathyroid cellular proliferation. $1,25(OH)_2D$ is by far the most important vitamin D metabolite that modulates parathyroid function. It acts through a nuclear receptor, the VDR, often in concert with other such receptors (i.e., those for retinoic acid or glucocorticoids), on DNA sequences upstream from the PTH gene (see Chapter 50). $1,25(OH)_2D$-induced upregulation of VDR and CaSR expression in the parathyroid could potentiate its inhibitory action(s) on PTH synthesis and secretion. Noncalcemic or less calcemic analogs of $1,25(OH)_2D$ inhibit PTH secretion while producing relatively little stimulation of intestinal calcium absorption and bone resorption, and may thus be attractive candidates for treating the hyperparathyroidism of chronic renal insufficiency.

Adjustment of the rate of parathyroid cellular proliferation is the third adaptive mechanism contributing to changes in the overall secretory activity of the parathyroid gland. Under normal conditions, parathyroid cells have little or no proliferative activity. The parathyroid glands, however, can enlarge greatly during states of chronic hypocalcemia, particularly in the setting of renal failure, probably because of a combination of hypocalcemia, hyperphosphatemia, and low levels of $1,25(OH)_2D$ in the latter condition.

Regulation of Parathyroid Hormone Secretion

The most important physiological factors that regulate PTH secretion are the concentrations of $Ca^{2+}{}_o$, $1,25(OH)_2D$, and extracellular phosphate ions. Of these three, $Ca^{2+}{}_o$ is most important in the minute-to-minute control of PTH secretion. Indeed, the actions of $1,25(OH)_2D$ and phosphate ions on the secretion concentrations of PTH probably result, at least in part, from their effects on hormonal biosynthesis rather than secretion *per se*. $Ca^{2+}{}_o$ also modulates several other aspects of parathyroid function that indirectly affect PTH secretion, including PTH gene expression, the hormone's intracellular degradation, and parathyroid cellular proliferation, as described previously. FGF23 and α-Klotho (a coreceptor for FGF receptors), also modulate parathyroid function, by inhibiting both PTH secretion and *PTH* gene expression. FGF23 also has important actions on $Ca^{2+}{}_o$ homeostasis, through inhibition of the formation of $1,25(OH)_2D$.

Physiologic Control of Parathyroid Hormone Secretion by Extracellular Calcium Ion

As illustrated in Fig. 47.3A, the relationship between PTH and $Ca^{2+}{}_o$ is represented by a steep inverse sigmoidal curve that can be quantitatively described by four parameters.[20] These are the maximal rate of PTH secretion at low $Ca^{2+}{}_o$ (parameter A); the slope of the curve at its midpoint (parameter B); the value of $Ca^{2+}{}_o$ at the midpoint (e.g., the "set point" or the level of $Ca^{2+}{}_o$ half-maximally suppressing PTH release; parameter C); and the minimal secretory rate at high $Ca^{2+}{}_o$ (parameter D). Parameter A *in vivo* is the sum of the maximal rates of PTH release from all individual parathyroid chief cells, as reflected by the resultant, maximally stimulated level of circulating PTH. Because of the steepness of the $Ca^{2+}{}_o$–PTH relationship, small alterations in $Ca^{2+}{}_o$ evoke large changes in PTH release, thereby contributing importantly to the near constancy of $Ca^{2+}{}_o$ *in vivo*. Indeed, parathyroid cells can readily detect and correct reductions in $Ca^{2+}{}_o$ of a few percentage points, leading to a percent coefficient of variation in $Ca^{2+}{}_o$ in humans that is less than 2%. The set point of the parathyroid gland is the key determinant of the level at which Ca^{2+}o is "set" *in vivo*, although the parathyroid set point is usually slightly lower than the ambient blood calcium concentration. Thus, the parathyroid cell is normally more than half-maximally suppressed at normal levels of Ca^{2+}o and has a large secretory reserve for responding to hypocalcemic stress. Nevertheless, PTH levels in vivo also fall dramatically (e.g., by 80%) when Ca^{2+}o rises to frankly hypercalcemic levels, which is thought to contribute importantly to the mineral ion homeostatic system's defense against hypercalcemia.[21] Evidence, however, from studies using mice with knockout of PTH or both PTH and the extracellular CaSR have shown that, even in the total absence of PTH, the high calcium-evoked increases in renal calcium excretion and calcitonin secretion provide a robust defense against hypercalcemia.[22]

Elevating Ca$^{2+}_{o}$ also decreases the proportion of secreted, intact PTH because of increased intraglandular degradation to inactive fragments (see the earlier section entitled Parathyroid Hormone Biosynthesis and Intraglandular Processing and the later section entitled Metabolism of Parathyroid Hormone). Even with severe hypercalcemia, however, some residual release of intact PTH(1–84) still occurs *in vivo* and persists at a level approximately 5% of that observed with a maximal hypocalcemic stimulus (see Fig. 47.3). This nonsuppressible basal component of PTH release may contribute to the hypercalcemia caused by hyperparathyroidism when the mass of abnormal parathyroid tissue is very great (e.g., in patients with renal failure).[20]

The parathyroid cell has a temporal hierarchy of responses to low Ca$^{2+}_{o}$ that permits it to secrete progressively larger amounts of hormone during prolonged hypocalcemia. To meet acute hypocalcemic challenges, PTH is released within seconds from preformed secretory vesicles by exocytosis as dictated by the sigmoidal curve (see Fig. 47.3). Sufficient PTH is stored in the parathyroid chief cell to sustain maximal, low Ca$^{2+}_{o}$-stimulated PTH release for approximately 60 to 90 minutes. Another rapid response of the parathyroid cell to hypocalcemia that enhances its net synthetic rate of PTH is reduced intracellular hormonal degradation—the opposite of what occurs at high levels of Ca$^{2+}_{o}$—which occurs within minutes to an hour. Hypocalcemia persisting for hours to days elicits increased PTH gene expression, whereas that lasting for days to weeks or longer stimulates parathyroid cellular proliferation. A greater secretory capacity for PTH on a per-cell basis (e.g., as a result of enhanced PTH gene expression) increases maximal hormonal secretion *in vivo*, as does an increase in cell number as a result of parathyroid cellular proliferation. In severe secondary hyperparathyroidism, very large increases in parathyroid cellular mass can elevate circulating PTH levels by 100-fold or more.

In addition to responding to changes in Ca$^{2+}_{o}$ *per se*, the parathyroid cell also appears to sense the rate of change in Ca$^{2+}_{o}$ such that rapid decrements in calcium promote more vigorous secretory responses than do changes of a similar magnitude occurring more slowly. Furthermore, during dynamic testing of parathyroid function *in vivo* by induced increases or decreases in Ca$^{2+}_{o}$, PTH in blood is higher at a given serum calcium concentration when Ca$^{2+}_{o}$ is falling than when it is rising (e.g., there is hysteresis in this relationship). The latter results in an apparent direction dependence of the secretory response, which, when combined with the rate dependence just described, may allow for a physiologically appropriate, more vigorous secretory response to large rapid decrements in Ca$^{2+}_{o}$. Also present are circadian and more rapid (i.e., occurring at rates of one to six pulses per hour) phasic changes in circulating PTH levels, but the physiologic significance of these changes is not known.

KEY POINTS

- The processes of parathyroid hormone (PTH) synthesis and secretion in parathyroid gland cells are tightly regulated by the calcium-sensing receptor (CaSR), a class C G protein–coupled receptor. The CaSR binds calcium cooperatively such that small changes in blood calcium levels lead to significant changes in blood PTH levels.

Mechanism of Extracellular Calcium Ion Sensing by Parathyroid Cells and Other Cells Involved in Mineral Ion Homeostasis

The molecular mechanism underlying Ca$^{2+}_{o}$-regulated PTH secretion involves the CaSR expressed on the surface of cells in the parathyroid glands.[22,23] The CaSR is a member of the class C subgroup of the GPCR protein superfamily.[24] Depending on cell type and experimental conditions, the CaSR can couple to several G protein signaling mediators, including $G_{q/11}$, G_i, $G_{12/13}$, and G_s, which, in turn, lead to the stimulation of phospholipase C (PLC), the inhibition of adenylate cyclase, the activation of Rho kinase, and the activation of adenylate cyclase, respectively. Patient-based studies described below indicate that G_{11} plays a particularly dominant role in mediating the control of PTH secretion via the CaSR.

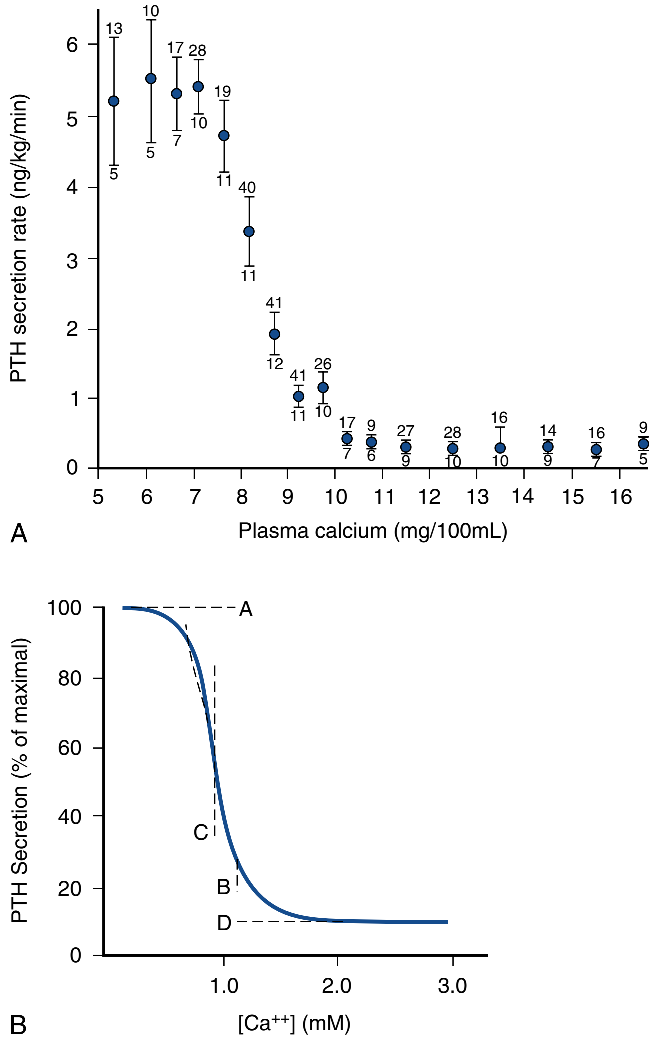


Fig. 47.3 Inverse sigmoidal relationship between calcium (Ca++) and parathyroid hormone (PTH) release and the four-parameter model describing these curves. A, Secretory response of bovine parathyroid glands to induced alterations in plasma calcium concentration. Calves were infused with calcium or ethylenediaminetetraacetic acid, and PTH secretion was assessed by measuring PTH levels in the parathyroid venous effluent. The symbols and vertical bars indicate the secretory rate (mean ± standard error) in calcium concentration ranges of 1.0 or 0.5 mg/100 mL. The number of calves and samples are indicated by numbers below and above, respectively. **B,** Sigmoidal curve generated by the equation Y = [(A − D)/(1 + (X/CB)] + D; the significance of A, B, C, and D are described in the text. (**A** from Mayer GP and Hurst JG. Sigmoidal relationship between parathyroid hormone secretion rate and plasma calcium concentration in calves. *Endocrinology*. 1978;102(4):1036–1042.)

The protein topology of the CaSR, as for other class C GPCRs, is characterized by a large N-terminal extracellular domain (ECD) that contains a bilobed, Venus flytrap (VFT)-like structure (Fig. 47.4). The

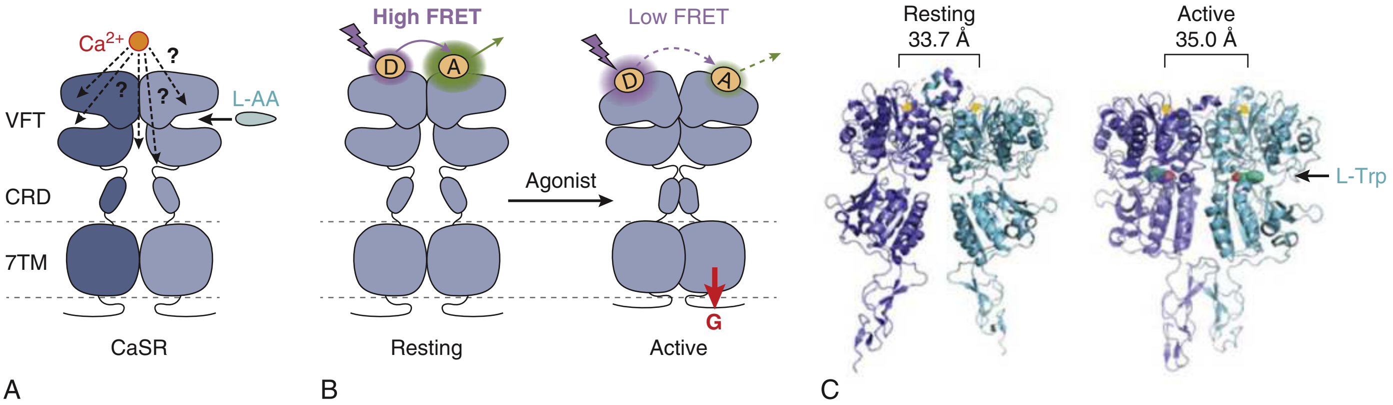


Fig. 47.4 **Calcium-sensing receptor (CaSR) structure and models of activation. A,** Schematic showing the dimeric organization of the CaSR; each receptor monomer contains a large extracellular domain (ECD) that is seen in x-ray crystallography structures.[25,26] The ECD is comprised of a two-lobed venus fly trap (VFT) domain and a cysteine-rich domain (CRD), which joins to the seven-helix bundle of the transmembrane domain (7TM). A binding site for an L-amino acid (L-AA) lies between the two lobes of each VFT. The location of binding sites for calcium (Ca^{2+}) remain uncertain, but mutational and Förster resonance energy transfer (FRET)-based studies are consistent with two functionally important Ca^{2+} binding sites in each VFT monomer. **B,** FRET-based studies using CaSR monomers tagged with fluorescent donor (D) and acceptor (A) probes support a closing of the VFTs upon binding agonist (e.g., Ca^{2+}), which leads to receptor activation and G protein coupling, a process potentiated by binding of an LAA.[27] **C,** Crystal structure–based models of the CaSR ECD in the inactive, open state obtained in the presence of low calcium (*left*, protein database [PDB] file 5K5T) and in the active, closed state obtained in the presence of high calcium (*right*, PDB 5K5S). Residue Asp23, tagged in each monomer with fluorescent FRET probes used to assess interlobe distances, measured as ångstrom (A˚), is shown as yellow spheres. (Reproduced with permission from Liu H, Yi P, Zhao W, et al. Illuminating the allosteric modulation of the calcium-sensing receptor. *Proc Natl Acad Sci U S A*. 2020;117(35):21711–21722.)

structure of the CaSR ECD has been determined by x-ray crystallography, and is thus seen to form a homodimer that is linked by a disulfide bond between cysteine residues located at position 129 at the top of each VFT domain.[25,26] A cysteine-rich domain links the VFT of each monomer to the corresponding heptahelical transmembrane domain (TMD) portion of the receptor, which mediates G protein coupling. The VFT of the CaSR plays a critical role in the receptor's capacity to bind extracellular calcium and likely other stimuli.[27] The x-ray crystallographic structures were obtained both in the absence and presence of high levels of calcium, and differences observed in the two structures provide clues about the mechanism of calcium-induced activation of the CaSR. In the calcium-free state, the two lobes of the VFT, which appear to be connected by a hinge, are open to form a cleft between the two lobes. In the calcium-bound state, the two lobes are closed together. It can thus be inferred that calcium binding leads to a shift of the VFT from an open, resting state to a closed, active-state conformation, and this closing, in turn, leads to a change in the TMD portion of the receptor and G protein activation (see Fig. 47.4).[27,28] Several bound calcium[26] or magnesium[25] atoms were observed in the ECD structures, although the bound divalent metal ions were not located at two sites that are predicted by modeling and mutational data to be particularly critical for calcium binding and function.[27] The structures also revealed an aromatic amino acid, L-tryptophan or a derivative, bound in the cleft between the two lobes of the VFT, and functional studies support a critical role in activation for binding of such an aromatic acid (see later). Intriguingly, the CaSR ECD structures also revealed several bound phosphate groups,[26] which has opened the possibility that this ion can act to negatively regulate CaSR function, as also discussed following.

The CaSR manifests marked positive cooperativity in its activation by $Ca^{2+}{}_o$, probably as a result of having at least two binding sites for $Ca^{2+}{}_o$ in each lobe of the VFT, as well as a binding site for an aromatic amino acid.[25-27] At least one key $Ca^{2+}{}_o$-binding site is thought to be located in the crevice between the two lobes of the VFT and adjacent to the amino acid binding site. Dynamic Förster resonance energy transfer–based studies of fluorescently labeled CaSR constructs show that such aromatic acids act as positive allosteric modulators (PAMs) for the CaSR and promote activation induced by calcium ions.[27] In contrast to L-amino acids, which bind to the VFT, small-molecule ligands, such as NPS R568, that also act as PAMs for the CaS, appear to bind to the TMD region of the receptor, suggesting that these compounds promote CaSR activation via a distinct mechanism.[29] Compounds that act as negative allosteric modulators for the CaSR have also been identified, but their binding sites have not been well characterized.[30] The full sequence of events by which the binding of calcium and other ligands to the CaSR alters the configuration of the receptor and its coupling to downstream signaling effectors remains to be established, as structures of the intact receptor have not yet been reported. Moreover, the potential role of phosphate in acting as a negative feedback modulator of CaSR signaling, as suggested by *ex vivo* functional studies,[31] also remains to be established.

The identification of hypercalcemic (e.g., familial hypocalciuric hypercalcemia [FHH] and neonatal severe hyperparathyroidism) and hypocalcemic (i.e., autosomal dominant hypocalcemia) disorders caused by inactivating and activating mutations of the CaSR, respectively, has provided incontrovertible proof of the receptor's central, nonredundant role in setting the serum calcium concentration.[22,28] Patients with these disorders have characteristic abnormalities in parathyroid and renal calcium sensing/handling. Heterozygous inactivating mutations of $G\alpha_{11}$ have been identified in patients with a form of FHH called FHH type 2, while activating mutations of $G\alpha_{11}$ have been shown to be the cause of autosomal dominant hypoparathyroidism type 2.[32,33] The $G\alpha_{11}$ protein, therefore, seems to be a key downstream effector of the CaSR in regulating PTH secretion, presumably by stimulating

PLC and modulating levels of intracellular calcium and other signaling molecules. Other second messenger pathways that might be involved as mediators of Ca^{2+}_o-induced suppression of PTH secretion remains uncertain, but these could potentially include arachidonic acid production by PLA_2, stimulation of ERK1/2, and elevation of cAMP.

The CaSR is also expressed in several additional tissues involved in systemic mineral ion homeostasis, including the calcitonin-secreting C-cells of the thyroid, diverse cells within the kidney, bone cells and/or their precursors, and intestinal epithelial cells. In the C-cell, the CaSR mediates the stimulatory action of high Ca^{2+}_o on calcitonin secretion, a Ca^{2+}_o-lowering hormone with unclear physiological significance in humans. In the kidney, the CaSR in the cortical thick ascending limb of the nephron mediates direct high Ca^{2+}_o–induced inhibition of the tubular reabsorption of calcium and magnesium. Therefore, raising Ca^{2+}_o directly inhibits renal tubular reabsorption of calcium via actions on the CaSR expressed in nephron segments involved in hormonal regulation of calcium reabsorption (e.g., by PTH) and indirectly inhibits it by reducing PTH secretion (see the later section Renal Calcium Reabsorption). The CaSR is expressed by chondrocytes as well as by osteoblasts and osteoclasts and/or their precursors, and some data suggest that it plays key roles in skeletal development, as well as bone anabolism.[34,35] It is not currently known whether the CaSR expressed in intestinal epithelial cells plays any role in regulating $1,25(OH)_2$ D-mediated absorption of calcium.

KEY POINTS

- The calcium-sensing receptor (CaSR) extracellular domain is characterized by a venus fly trap domain that contains binding sites for calcium as well as other ligands, including phosphate and an aromatic L-amino acid. The CaSR functions as a homodimer and couples to the Gα11 signaling cascade.

$1,25(OH)_2D$, Phosphate, and Other Factors Regulating Parathyroid Hormone Secretion

In addition to directly inhibiting *PTH* gene expression, $1,25(OH)_2D$ also reduces PTH secretion. It is not known whether this latter action is solely secondary to the effect of $1,25(OH)_2D$ on biosynthesis of the hormone and/or represents a direct action on the secretory process *per se*. Increasing the ambient level of phosphate *in vitro*, independent of concomitant changes in Ca^{2+}_o, enhances parathyroid cellular proliferation, *PTH* gene expression, and hormonal secretion. Phosphate-induced changes in PTH secretion, however, take several hours and may result secondarily from changes in hormonal biosynthesis rather than secretion *per se*. Finally, extracellular magnesium ion (Mg^{2+}_o) clearly functions as a CaSR agonist *in vitro*, although it is 2-fold to 3-fold less potent than Ca^{2+}_o.[27] Because levels of serum ionized Mg^{2+}_o are lower than those of Ca^{2+}_o, it is presently unclear whether Mg^{2+}_o acts as a physiologically relevant CaSR agonist at the parathyroid gland *in vivo* under normal circumstances. Patients with inactivating or activating CaSR mutations, however, can exhibit mild hypermagnesemia or hypomagnesemia, respectively, suggesting that the CaSR does contribute to setting Mg^{2+}_o *in vivo*.

KEY POINTS

- Parathyroid hormone acts in concert with $1,25(OH)_2D$ and FGF23 as part of a tightly integrated signaling network that operates to maintain blood levels of calcium as well as phosphorus.

Metabolism of Parathyroid Hormone

In addition to the full-length polypeptide PTH(1–84), which is the biologically active hormone, circulating PTH fragments that lack an intact N-terminus can generally be readily detected, and most of these fragments are devoid of biologic activity, at least with regard to the PTHR-1.[36] C-terminal PTH fragments are produced in and released from the parathyroid gland, but they are also derived from the circulating intact hormone by degradative systems active in the liver, kidney, and other peripheral sites. However, some PTH fragments, such as PTH(7–84), appear to be generated within the gland and may have some biological activity. The relative concentration of such C-terminal fragments released from the gland increases under conditions of systemic hypercalcemia, when overall secretion rates of intact hormone are lower. The secreted C-terminal PTH fragments are similar to those generated by peripheral metabolism (see later) but have not been chemically characterized. Studies in animals show that the clearance of intact PTH from plasma is very rapid (half-life: 2–4 min), and the liver and kidney are the major organ sites of PTH clearance. Hepatic mechanisms are estimated to account for 40% to 75% of intact hormone clearance, while renal mechanisms account for approximately 20% to 30% of total peptide clearance.

The current evidence therefore indicates that intact PTH and multiple C-terminal fragments (which are derived from glandular and peripheral cleavage and may not be identical) are the principal circulating hormonal forms. Biologically active, N-terminal fragments of PTH, if found in the circulation at all, are likely to circulate only at extremely low concentrations ($<10^{-13}$ to 10^{-14} mol/L).[36] Several different immunometric assays have been developed for detecting PTH in the circulation. These assays each use two different antibodies: an immobilized capturing antibody directed against the C-terminal portion of PTH(1–84) and an enzyme-labeled detection antibody directed against the N-terminal portion of the intact molecule. The precise epitopes may differ for the different antibodies used in the various assays used clinically, which can lead to ambiguity in the results obtained for patient samples analyzed with two or more of the assays.[37] Nevertheless, the different assays have been helped in demonstrating the presence of different forms of PTH fragments.

The careful immunometric analysis of circulating PTH levels has considerable significance for the management of patients with parathyroid dysfunction, especially in the presence of renal failure. Particularly during hypercalcemia, N-terminally truncated forms of PTH become a significant, if not the dominant, PTH species. Measurements of "intact" PTH by earlier assays had, therefore, overestimated the concentrations of biologically active PTH, and so could result in the overtreatment of uremic patients with vitamin D analogs and/or calcium. The resulting suppression of the parathyroid gland, even without evoking an inhibitory effect of a fragment similar to PTH(7–84), could be acting together to excessively reduce PTH-dependent bone turnover. When secretion of PTH is suppressed, as in hypercalcemia, and the fractional concentration of the large, N-truncated PTH increases, however, inhibition of the actions of native PTH on calcium or bone could occur. Experiments with synthetic PTH(7–84) support the conclusion that such actions occur *in vivo* and *in vitro*. For example, PTH(7–84) was shown to have hypocalcemic properties *in vivo*, and it reduces, potentially via an as yet unidentified receptor specific for the C-terminal portion of PTH, bone resorption that may be partly due to impaired osteoclast differentiation. Newer immunoassays that do not detect such amino terminally truncated PTH molecules likely offer a decisive advantage in diagnosis and management of secondary hyperparathyroidism (see Chapters 54 and 56).

KEY POINTS

- Parathyroid hormone (PTH) 1–84 exhibits a short half-life in the circulation of approximately 2–4 minutes, due in large part to cleavage and clearance mechanisms operating in the liver and kidney. C-terminal fragments such as PTH(7–84) can be detected as circulating metabolites, particularly in patients with renal insufficiency.

PARATHYROID HORMONE–DEPENDENT REGULATION OF MINERAL ION HOMEOSTASIS MEDIATED THROUGH THE PARATHYROID HORMONE RECEPTOR TYPE-1

PTH is, as outlined earlier, the most important peptide regulator of mineral ion homeostasis in mammals. Through its actions on the kidney and bone, PTH maintains blood calcium concentration within narrow limits. In bone it stimulates the release of calcium, while in the kidney, it enhances renal tubular reabsorption of calcium (and diminishes tubular reabsorption of phosphate). Furthermore, in the kidney it stimulates the synthesis of renal 1α-hydroxylase, which leads to an increase in the renal production of 1,25(OH)$_2$D, that, in turn, promotes the intestinal absorption of calcium (and phosphate). These direct and indirect endocrine actions of PTH are mediated through PTHR-1, a class B GPCR that is abundantly expressed in both major target tissues of PTH action, bone and kidney.

Actions of Parathyroid Hormone on the Kidney

In addition to regulating renal calcium and phosphate transport, PTH modifies the tubular handling of magnesium, sodium, potassium, bicarbonate, and water, and moreover stimulates renal gluconeogenesis. Specific regions of the nephron that are involved in each of the PTH-dependent actions have been defined through *in vivo* micropuncture analyses and through *in vitro* studies with nephron segments that have been dissected from the remainder of the kidney. Furthermore, cell lines from specific regions of the kidney have also been used. However, interpretation of these results has to take into account the complexity of the organization of the kidney as an organ (for example, the complex anatomic relationship involving different renal tubular segments spanning both the cortex and medulla and the effects of countercurrent distribution that modify solute and water transport), particularly because certain *in vivo* features of renal tubules are not readily imitated *in vitro* through the use of isolated tubules or cells.

Renal Calcium Reabsorption

Most of the calcium in the glomerular filtrate is reabsorbed. The bulk of this reabsorption (65%) occurs via passive, paracellular mechanisms, both in the proximal tubules and, to a lesser extent, in the thick ascending limb of Henle's loop and the distal convoluted tubule. The CaSR also plays an important, PTH-independent role in the adjustment of renal calcium reabsorption in the cortical thick ascending limb. The physiologically important stimulation of renal calcium reabsorption by PTH occurs almost entirely in the distal nephron. In the cortical thick ascending limb (TAL), PTH increases the magnitude of the positive charge potential of the lumen that drives the passive, paracellular reabsorption of calcium and magnesium. In the distal convoluted tubule, in contrast, PTH promotes increased transcellular calcium reabsorption by coordinate upregulation of molecular components of this transcellular pathway. PTH enhances uptake of calcium into the distal tubular cells via the (luminal) plasma membrane channel, TRPV5, as well as its active extrusion into the blood against a steep electrochemical gradient at the basolateral membrane. This latter process of extrusion involves two types of transporters. One is the plasma membrane calcium pump (Ca^{2+}/Mg^{2+}-ATPase, PMCA), and the second is a Na^+/Ca^{2+} exchanger (NCX1), which, in turn, is indirectly regulated by Na^+/K^+-ATPase(s), which maintains the transcellular sodium gradient. Studies performed with membrane vesicles from the distal region of the kidney show increased activity of the Na^+/Ca^{2+} exchanger in response to PTH. PTH also stimulates the translocation of preformed calcium channels sequestered within the interior of certain distal tubular cells, presumably TRPV5, to the apical (i.e., luminal) surface of the renal tubular epithelial cell and mediate increased cellular calcium uptake. Because PTH simultaneously enhances the activity of Na^+/Ca^{2+} exchangers in the basolateral (antiluminal) membrane, the overall process promotes an increase in transcellular calcium uptake from the lumen to blood, that is, from the apical to the basolateral membrane.

More recent studies in mice reveal that PTHR-1 signaling in the TAL downregulates claudin-14, a tight junction factor that inhibits paracellular calcium transport in the TAL.[38] Additionally, the membrane abundance of TRPV5 in cells of the TAL has been shown to be negatively regulated by phosphorylated claudin-16, mutations in which cause hypomagnesemia with hypercalciuria and nephrocalcinosis, and PTHR-1 signaling promotes claudin-16 phosphorylation, which would thus augment calcium reabsorption via TRPV5.[39]

Regulation of 1α- and 24-Hydroxylase Activity

PTH is a major inducer of the activity of proximal tubular 1α-hydroxylase, a microsomal cytochrome P-450 enzyme that synthesizes biologically active 1,25(OH)$_2$D from the substrate 25(OH)D. This effect of PTH on synthesis of the renal enzyme shows longer lag times than its effect on renal calcium transport and is mediated, at least in part, by the protein kinase A (PKA) signaling pathway of the PTHR-1. Although 1α-hydroxylase activity has been detected in several nonrenal tissues, its mRNA is found in abundant concentrations only in the kidney, confirming that this organ is the most important site for the generation of 1,25(OH)$_2$D. PTH-dependent synthesis of new 1α-hydroxylase protein, which only occurs in the renal proximal tubule, requires several hours and is blocked by 1,25(OH)$_2$D. Studies in mice have revealed a renal-specific transcriptional control element located distal to the *Cyp27b1* gene, which encodes the 1α-hydroxylase and is required for responsiveness to regulation by PTH, as well as FGF23.[40] The element appears to be involved in a longer-range control of *Cyp27b1* promoter activity, likely via effects on chromatin structure in that genetic locus. Hypophosphatemia is, similar to PTH, a major inducer of 1α-hydroxylase, though it is still unclear whether the effects of hypophosphatemia are independent of effects on FGF23 levels in the blood. In contrast, hypercalcemia, as would be generated by sustained increases in circulating levels of PTH or PTHrP, suppresses synthesis of the enzyme, thus limiting overall 1,25(OH)$_2$D synthesis in a homeostatic manner. The molecules 25(OH)D and 1,25(OH)$_2$D can also be hydroxylated by the 24-hydroxylase, resulting in the inactive metabolites, 24,25(OH)$_2$D and 1,24,25(OH)$_3$D (see Chapter 50). PTH has an inhibitory effect on the 24-hydroxylase, thus reducing the inactivation of 1,25(OH)$_2$D; in contrast, 1,25(OH)$_2$D stimulates the synthesis of 24-hydroxylase, thereby inducing its own metabolism.

Renal Phosphate Transport

PTH is of major importance for maintaining normal blood calcium levels, but it is not the principal regulator of the serum phosphate concentration, which appears to depend on specific phosphaturic factors, particularly FGF23 (see Chapter 53). However, when PTH causes an increase in bone resorption (as might occur with prolonged dietary calcium deprivation), calcium and phosphate increase simultaneously in the blood. Although calcium is needed, phosphate is best excreted, which is mainly accomplished by a PTH-stimulated increase in renal phosphate clearance.

PTH acts directly on proximal tubular cells, where it regulates the expression of the type II sodium-phosphate cotransporters, NPT2a and NPT2c, and the abundance of these protein in the apical brush border membrane (BBM). PTH thus induces phosphaturia by increasing the internalization of NPT2a and NPT2c, and hence reducing their functional expression at the apical, BBM surface of the proximal tubule cells. Both the cAMP/PKA- and intracellular calcium/IP_3/protein kinase C (PKC)-dependent signaling pathways appear to contribute to the PTH-induced reduction in proximal tubular phosphate reabsorption.[41,42] Evidence supporting a predominant role for the cAMP/PKA pathway, however, in the PTH-dependent regulation of renal phosphate transporters comes from patients with pseudohypoparathyroidism type Ia. This disease is caused by maternally inherited, inactivating mutations in *GNAS*, i.e., the gene encoding the stimulatory G protein Gsα.[43] Patients develop hyperphosphatemia due to a lack of functional Gsα in the proximal tubules, and consequently show impaired urinary cAMP and phosphate excretion in response to PTH (see Chapters 53 and 57).

NPT2a and NPT2c appear to have distinct roles in regulating phosphate transport in the kidney. NPT2a is expressed in segments S1 through S3 of the proximal tubule, whereas NPT2c is expressed only in the S1 segment. NPT2a protein undergoes lysosomal degradation upon internalization in response to PTH, whereas NPT2c appears to be recycled and reinserted into the BBM after initial internalization. The mechanisms underlying these diverse response profiles are uncertain at present. In mice, the mRNA encoding NTP2c is approximately 10-fold less abundant than that encoding NPT2a, suggesting that NTP2c plays only a minor biological role in mice, limited to some period during postnatal development. However, in humans, homozygous and compound heterozygous mutations in NPT2c are known to cause hereditary hypophosphatemic rickets with hypercalciuria, an autosomal recessive disorder. This disease linkage makes it certain that NPT2c serves important functions in biology and is not redundant with NPT2a. Homozygous, inactivating mutations in NPT2a are the cause of a more severe disease leading not only to renal phosphate wasting, but also to the impairment of other proximal tubular functions and severe hypercalcemia.

Intracellular scaffold proteins, in particular the NHERF family of proteins, play important roles in the PTH-mediated regulation of NPT2a and NPT2c expression at the cell surface.[42,44,45] NHERF-1-null mice display phosphate wasting due to a reduction in NPT2a expression at the BBM of proximal tubules. NHERF-1 interacts with the PTHR-1, as well as NPT2a, via PDZ domain–based scaffolding interactions that involve PDZ domain recognition motifs located on the C-terminal tail of the PTHR-1 and on cytoplasmic segments of the transporter proteins. PTHR-1 signaling induces phosphorylation of Ser290 in NHERF-1 as a key step in the mechanism by which PTH regulates NPT2a-mediated phosphate uptake, presumably in part by promoting dissociation of NPT2a–NHERF-1 complexes at the BBM membrane.[42,44] Other studies have shown that PTHR-1 coupling to NHERF proteins can lead to changes in the specific signaling pathways activated by the receptor, at least in certain cells. NPT2a expression in the renal BBM and its interaction with NHERF proteins can further be regulated by FGF23, as well as by $1,25(OH)_2D$, as shown by studies in the Hyp mouse model of human X-linked hypophosphatemia.[45]

KEY POINTS

- Parathyroid hormone (PTH) stimulates calcium reabsorption in the distal nephron through effects on both passive paracellular transport and active transcellular processes, the latter involving TRPV5 calcium channels. In proximal tubule cells, PTH stimulates the synthesis of the 1α-hydroxylase, and hence the formation of $1,25(OH)_2D$, as well as the downregulation of NPT2a and NPT2c, and hence the urinary excretion of phosphorus.

Actions of Parathyroid Hormone on Bone

PTH actions on bone cells are critical in the normal process of bone development and maintenance of skeletal integrity in the adult, as well as playing a crucial role in calcium and phosphate homeostasis through its actions on target cells. In bone, PTH affects several highly specialized cells, i.e., osteoblasts, osteocytes, and osteoclasts, that are critical in bone health and integrity and, as discussed later, are used in generating some of the homeostatic responses in mineral ion metabolism. The actions of PTH on cells of the osteoblast lineage reflect direct actions of PTH, while the effects on osteoclasts are indirect, mediated in an autocrine/paracrine manner through factors released by osteoblasts and osteocytes in response to PTH receptor activation. PTH actions on bone cells are critically influenced by the actions of vitamin D (see Chapters 50 and 51) and closely intertwined with the actions of PTHrP (see Chapter 48). PTH actions on bone are but one factor in a complex pattern of interactions involving multiple systemic and locally produced hormones and cytokines, for example, sex steroids.

The tools of mouse genetics have permitted remarkable advances in unraveling the complex pattern of interactions due to PTH by selective gene knockout of individual components of the interacting system, followed by observation of the effects *in vivo*. Mouse models in which PTH action has been selectively blocked exhibit a marked reduction in trabecular bone, indicating a role for PTH in bone development and a close interplay between PTH and PTHrP in the skeleton. PTH action can also be analyzed in a pharmacological mode. Administration of PTH leads within minutes to a lowering of calcium caused by uptake of the mineral into bone cells followed by increased mobilization of calcium from the mineral phase into the bloodstream.[46,47] The precise mechanisms responsible for these rapid bidirectional changes have not yet been elucidated, although clearly a major pathway for calcium release from bone involves osteoclasts; these cells undergo multiple cellular changes involving the activation of cellular transporters and pumps, as well as the secretion of enzymes such as cathepsin K and collagenases. Activation of the PTHR-1 in osteocytes also leads to release of calcium from the matrix surrounding osteocytes; the relative roles, quantitatively, of this osteocytic osteolysis versus osteoclastic osteolysis in responses to PTH is a focus of current research.

An additional insight regarding the effects of PTH actions on bone cells can be appreciated through results seen when PTH is used as a therapy for osteoporosis. Daily administration of PTH, i.e., given intermittently, increases bone mass by stimulating bone formation, an action that involves a complex set of cellular responses affecting principally stromal cell/osteoblast proliferation, differentiation, and cellular actions, as well as actions on osteocytes. On the other hand, continuous PTH administration *in vivo* leads to decreased bone mass. In this situation both bone formation and bone resorption are stimulated, but the latter effect is predominant. The pathologic condition of hyperparathyroidism is an example of this effect (see Chapters 54 and 51).

Effects on Osteoblasts. PTH actions on mature osteoblasts, which express the PTHR-1, involve a number of signaling pathways with complex interactions. PTH activates the cAMP/PKA and multiple additional signaling pathways, including the calcium/PKC pathway, the Wnt/beta-catenin pathway, and the mitogen-activated protein kinase pathway. Such potential for signaling diversity likely contributes to the complex patterns of direct and indirect responses that occur in the target cells of bone tissue. The upstream signaling events eventually lead to modulation of the actions of genes crucial for osteoblast differentiation, function, cell division, and apoptosis. A remarkable number of cellular activities of osteoblasts are influenced by PTH, including cellular metabolic activity, ion transport, cell shape, gene transcriptional activity, and secretion of multiple proteases. Among these actions are critical cellular responses including direction of precursor cells into the osteoblast lineage under the influence of Runx2, beta-catenin, and osterix. PTH action promotes osteoblast differentiation in part by promoting exit from the cell cycle. In addition to effects on cell proliferation, PTH controls cell survival by promoting inactivation of the proapoptosis factor BAD and upregulation of the cell survival factor Bcl2.

The salt-inducible kinase (SIK) family of proteins have recently been discovered to play critical roles in mediating both the anabolic and catabolic, as well as the developmental effects of PTHR-1 signaling in the skeleton, particularly in osteocytes, as described further in the following section.[48-50] Recent lineage tracing studies performed in mice that permit tracking, using genetically encoding fluorescent labels, of all descendants of a specifically targeted type of stem cell, reveal that PTH treatment stimulates the awakening of quiescent bone lining cells and hence their transformation into active osteoblasts. PTH also induces the progression of preosteoblastic mesenchymal stem cells along the osteoblast lineage pathway, while preventing their progression along an alternate pathway of adipogenesis.[51-54] The beta-catenin–Wnt ligand signaling pathway plays a key role in mediating responses in osteoblastic cells downstream of PTHR-1 signaling, although some of these effects likely occur indirectly through the direct downregulation by PTHR-1 signaling of expression in osteoblasts and or osteocytes (see later) of the *SOST* gene, which encodes the Wnt pathway inhibitor protein sclerostin.[55] Mature osteoblasts on the bone surface secrete collagen and other elements of the bone matrix (osteoid) and also direct the mineralization of osteoid. These critical steps, as well as the multiple factors in addition to PTH that influence the process, are thoroughly reviewed in Chapter 51.

Effects on Osteoclasts. Multiple factors, local and systemic, are involved in osteoclast generation and activity in addition to PTH (see Chapters 51). Regarding PTH actions, it is generally agreed that osteoclasts lack PTH receptors, and hence actions of PTH are indirect through stimulation of cytokines from osteoblasts and osteocytes. Osteoblasts and osteocytes (see later) produce RANKL, a membrane-associated or soluble, secreted protein with homology to the family of tumor necrosis factors (TNFs), and PTH can stimulate production of this cytokine in these cells.[50,56,57]. RANKL is a critical regulator of osteoclast function, directing differentiation of osteoclast precursors into mature bone-resorbing osteoclasts. These effects of RANKL are mediated through RANK, a member of the TNF receptor family, that induces the differentiation of osteoclast precursors into mature bone-resorbing osteoclasts upon cell-to-cell contact or by binding soluble RANKL (in the presence of macrophage colony-stimulating factor). An additional feedback regulator involved in the balance of bone formation and bone resorption is osteoprotegerin (OPG), a soluble protein with homology also to the TNF receptor family that acts as a decoy receptor and thus binds and inactivates RANKL.[58] PTH inhibits OPG production by osteoblasts. Transgenic expression of OPG leads to impaired osteoclastogenesis and thus to osteopetrosis, whereas genetic ablation of OPG leads to bone loss. Ablation of the gene for RANKL or the receptor, RANK, also leads to osteopetrosis.

Effects on Osteocytes. Osteocytes are the most numerous cells in bone, and there is a growing appreciation of the central biological role they play, not only in bone turnover but also as a source of circulating factors critical in calcium and phosphate homeostasis. Osteocytes are present throughout bone, encased within the mineral matrix.[59,60] However, they develop numerous dendritic processes, which permit communication with one another and with the vasculature and cells on the bone surface. Osteocytes are the key cells in bone that sense and respond to mechanical forces, which lead to changes in bone mass. They are also responsive to PTH, as well as other circulating hormones and locally produced cytokines that contribute to the control of bone mass.[61,62] The role of PTH action in osteocytes in controlling bone remodeling is demonstrated by the marked reduction in bone mass, particularly in the trabecular compartment, seen in mice that have osteocyte-selective deletion of the PTHR-1. Conversely, expression of a constitutively active PTHR-1 specifically in osteocytes leads to a high bone mass phenotype. Osteocytes produce a number of important proteins, many of which are secreted, that are involved in skeletal homeostasis and mineral ion homeostasis. One PTH-sensitive factor that plays a key role in controlling bone metabolism is sclerostin, encoded by the *SOST* gene. Sclerostin acts as a competitive inhibitor of Wnt receptor signaling, and reductions in sclerostin production lead to expansions in osteoblast populations and marked increases in bone mass.[55] The suppression of *SOST* gene expression is thus a major mechanistic component of the anabolic actions of PTH.

A key mechanistic link in the signaling pathway by which PTH reduces *SOST* expression in osteocytes was recently shown to involve the SIK family of proteins.[49,50] In response to PTHR-1–mediated activation of the cAMP/PKA pathway, SIK-2 is specifically phosphorylated by PKA, and this inhibits SIK-2 kinase activity. This, in turn, leads to a decrease in the phosphorylation states of SIK-2 target proteins, which include members of the class 2a family of histone deacetylases (HDACs). The consequent decreased phosphorylation of HDAC proteins causes them to translocate from the cytoplasm to the nucleus, where they act to control the expression of a variety of genes, including *SOST*. PTH-induced SIK-2 phosphorylation thus leads to reduced levels of phosphorylated HDAC-4 and HDAC-5, and hence an increase in the translocation of HDAC-4/5 to the nucleus and their subsequent complexing with certain transcriptional regulators, such as MEF2C, at a distal *SOST* gene enhancer to thus inhibit *SOST* expression. The critical role that SIK proteins play in mediating the anabolic effects of PTHR-1 signaling is demonstrated by two findings. First, treatment of rodents with a small-molecule SIK inhibitor induces increases in bone mass similar to those seen with PTH treatment.[48] Secondly, mice engineered to have genetic ablation of SIK-2 and SIK-3 in mature osteoblasts exhibit a striking increase in bone mass that resembles closely the increase seen in mice expressing a constitutively active PTHR-1 mutant.[49]

In parallel to *SOST* gene regulation, the SIK proteins also mediate the PTH-induced induction of *RANKL* gene expression in osteocytes, and thus act as critical regulators of an important driver of osteoclastogenesis.[50] In this case, instead of HDACs, the SIK-2 target protein that becomes less phosphorylated upon activation of PTHR-1 signaling is the cAMP response transcriptional coactivator 2, CRTC2, which then translocates to the nucleus and activates the *RANKL* gene promoter.

KEY POINTS

- Parathyroid hormone (PTH) acts on cells of the osteoblast lineage to induce changes in the expression of multiple genes involved in calcium and bone metabolism, such as *RANKL* and the gene encoding sclerostin. The salt-inducible kinase family of proteins are key downstream mediators of PTH signaling in bone cells.

MOLECULAR PROPERTIES OF THE PARATHYROID HORMONE RECEPTOR TYPE-1

Gene Structure, Protein Topology, and Evolution

The PTHR-1 belongs to a distinct subgroup of GPCRs called the class B GPCRs, which is comprised of 15 receptors, including those for glucagon, GLP-1, calcitonin, and several other peptide hormones that are each similar in size to PTH(1–34).[63] The gene encoding the human PTHR-1 is located on chromosome 3p and spans at least 20 kb of genomic DNA. The gene consists of 14 coding exons and at least three noncoding exons.

The size of the coding exons in the human gene ranges from 42 bp (exon M7) to more than 400 bp (exon T), and the introns vary in length from 81 bp (intron between exons M6/7 and M7) to more than 10,000 bp (intron between exons S and E1). Two promoters for the PTHR-1, P1 and P2, have been described in rodents. The activity of P1 is mainly restricted to the adult kidney, while that of P2 is detected in several fetal and adult tissues, including cartilage and bone.

The human PTHR-1 receptor protein is 593 amino acids in length, and consists of a 22–amino acid N-terminal signal peptide that is removed during receptor synthesis. The general protein topology is defined by a relatively large N-terminal extracellular domain (ECD) of approximately 160 amino acids, a TMD containing seven membrane-spanning helices with interconnecting loops, and a C-terminal tail of approximately 110 amino acids. In mammals, the PTHR-1 ECD is encoded by five exons: S, E1, E2, E3, and G. Exon S and G encode the signal peptide and a glycosylated segment, respectively. Within segments encoded by exons E1, E3, and G are six conserved cysteines that form a disulfide bond network that establishes the ECD structural fold. Exon E2 encodes a nonconserved 41–amino acid segment that is not essential for function, as it can be deleted or replaced by insertions, such as with sequences encoding green fluorescent protein, without affecting ligand binding or signaling. An equivalent of E2 is absent in the PTHR-1 homologs found in all nonmammalian species, as well as in the PTHR-2 and all other class B GPCRs. The E2 segment of the mammalian *PTHR-1* gene thus likely reflects an evolutionarily recent genetic insertion event. In apparent parallel to the evolutionary history described above for PTH ligands, DNA sequences corresponding to plausible homologs of the PTH receptor have been detected in the genomes of certain protochordates, such as amphioxus and the tunicate sea squirt *Ciona intestinalis,* as well as in some insects.[12,15,16]

Three-Dimensional Structures of the Parathyroid Hormone Receptor Type-1

The first two high-resolution three-dimensional structures of the PTHR-1 were recently reported; one was obtained by x-ray crystallography, and the other by cryo-EM.[64,65] The cryo-EM structure is shown in Fig. 47.5.[65] The structure was obtained as a complex of the receptor bound to a long-acting PTH ligand (LA-PTH; discussed further in a later section of this chapter) and coupled to a heterotrimeric G protein, Gsα/β/γ. The protein components were produced in SF-9 insect cells, and the complex was assembled in a discoid lipid micelle. The PTHR-1 structure determined by x-ray crystallography used an engineered PTHR-1 construct that contained several thermostabilizing mutations, which resulted in a loss of signaling capacity, and the PTHR-1 was bound to an engineered (e)PTH(1–34) analog, but not to a partner G protein.[64] In both studies, the overall structure of the PTHR-1 resembles closely the structures determined for several other class B GPCRs, which includes the glucagon receptor, GLP-1 receptor, secretin receptor, and calcitonin receptor.[63,66-68] The cryo-EM and x-ray crystallography structures of the PTHR-1 represent the receptor in an active, G protein–coupled state, and in a ligand-bound but inactive and G protein–uncoupled state, respectively. Comparing the two structures can thus yield insights into the molecular mechanisms by which PTHR-1 binds a peptide ligand, changes conformation, and couples to G proteins.

KEY POINTS

- X-ray crystallographic and cryogenic electron microscopy structures reveal that the parathyroid hormone 1–34–based ligand binds to parathyroid hormone receptor type-1 as a linear alpha-helix, with its C-terminal 15–34 region docked to the receptor's N-terminal extracellular domain and its N-terminal ~1–14 region projecting down into the core of the receptor's heptahelical transmembrane bundle. Upon activation, the cytoplasmic surface of the receptor opens to allow G protein coupling.

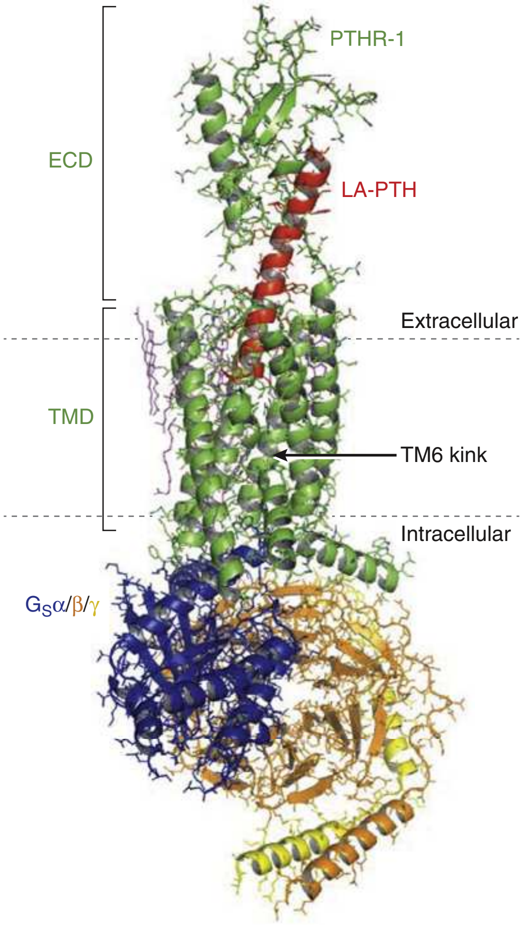


Fig. 47.5 Structure of the parathyroid hormone (PTH) receptor type-1 (PTHR-1) in complex with long-acting (LA)-PTH and Gsα/β/γ. Shown is the structure of the PTHR-1 bound to the LA-PTH, which is a hybrid peptide consisting of the M-PTH(1–14) segment joined to a modified PTH-related peptide(15–36) segment and coupled to a heterotrimeric G protein. The structure was determined by cryogenic electron microscopy at a global resolution of 3.0 Å. The protein backbones are shown in ribbon format, and amino acid sidechains are shown as sticks. The receptor *(green)* is comprised of an N-terminal extracellular domain (ECD) and a heptahelical TMD bundle region. The ligand *(red)* binds as a linear alpha helix, with its C-terminal (15–36) portion contacting the ECD, and its N-terminal (1–14) portion entering into the core of the TMD bundle. Interactions between the N-terminal residues of PTH and residues within the TMD core induce a kink in transmembrane helix 6 (TM6) specifically at the Pro415–Gly418 segment, which leads to an outward movement of the cytoplasmic terminus of TM6 and an opening in the receptor that enables coupling to the G protein trimer. The G protein subunits are Gsα *(blue)*, β *(orange)*, and γ *(yellow)*. Lipid components of the micelle used for assembling the complex are shown in purple. The model is based on the studies reported in[65] and was generated using the protein database file 6NBF and Pymol modeling software.

Mechanisms of Ligand Binding and Activation at the Parathyroid Hormone Receptor Type-1. The basic structure of the PTHR-1 consists of the N-terminal ECD region, which adopts a beta sheet–helix–beta sheet protein fold with a central groove, and a TMD region, which utilizes the classical heptahelical bundle motif present in all GPCRs. The TMD contains an extracellularly exposed pocket that accommodates the key portion of the ligand that activates the receptor. In both the x-ray crystallography and cryo-EM structures, the PTH peptide binds as a continuous alpha-helix with its C-terminal portion docked to the central groove of the ECD and its N-terminal portion projecting downward into TMD pocket, such that residue 1 of PTH is positioned at the approximate center of the TMD bundle and midway between the inner and outer surfaces of the plasma membrane. The total ligand–receptor contact surface area is extensive and ranges the full length of the ligand. Most, if not all, of the numerous intermolecular sidechain and backbone interactions that occur likely contribute to the overall affinity and stability of the complex.

A key difference between the inactive and active state structures of the PTHR-1, also observed in all other class B GPCR structures obtained in active and inactive conformations, is a pronounced kink and partial unfolding in the midsection of transmembrane helix (TM) 6 in the active-state receptor, whereas TM6 is fully helical in the inactive-state receptor (Fig. 47.5). The TM6 kink is centered at the segment Pro415–Gly418, which represents the PXXG motif that is highly conserved in the class B GPCRs. As proline and glycine are helix destabilizing residues, the PXXG motif likely mediates a key step in the PTHR-1 activation process, i.e., the partial unraveling and bending of TM6 upon binding of an agonist ligand. The Pro415–Gly418 segment thus serves as a molecular switch that acts to shift the receptor between inactive- and active-state conformations.

The precise mechanisms of ligand-induced activation remain to be elucidated, but one ligand residue that seems to be involved is Glu4, which is highly conserved in PTH and PTHrP ligands. Glu4 contacts several receptor residues near or just above the PXXG motif, specifically, Tyr195 in TM1, Arg233 in TM2, and Asn448 in TM7.[69] These residues form a conserved central polar network that likely rearranges upon interaction with the negatively charged carboxylate of Glu4 to thereby relay conformational changes to the PXXG motif. As a result of bending at the PXXG segment, the lower, cytoplasmic portion of TM6, as well as those of TM7 and TM5, move outward to open a cavity on the cytoplasmic face of the receptor that is then used for docking a G protein, thus establishing the ligand–receptor–G protein ternary complex.

A second PTHR-1 structural motif that plays a key role in controlling receptor activation is a distinct polar network formed at the cytoplasmic base of the TMD bundle. This motif involves four residues: His223 in TM2, Glu302 in TM3, Thr410 in TM6, and Tyr459 in TM7, which are again highly conserved in the class B GPCRs. This so-called HETY motif acts as yet a second relay switch that directly controls the outward movements of the lower portions of the TMD helices that is required for G protein coupling, as described above. Mutations impacting this HETY motif in the PTHR-1 have been identified in patients with Jansen's metaphyseal chondrodysplasia (JMC), which is discussed in a later section of this chapter.

Conformationally Selective Parathyroid Hormone Ligands and Prolonged Signaling

A series of kinetic binding and signaling assays performed using various PTH and PTHrP ligand analogs has shown that the PTHR-1 can adopt different protein conformations and thus be selectively bound by structurally distinct PTH peptide ligand analogs to induce different types of signaling responses in target cells.[70-72] A noteworthy ligand in this regard is LA-PTH. LA-PTH is a hybrid peptide comprised of a modified PTH(1–14) sequence joined to a modified PTHrP(15–36) sequence. LA-PTH and similar analogs mediate prolonged cAMP signaling responses in cells due to a capacity to bind with high affinity and stability to the G protein–uncoupled PTHR-1 conformation, called R^0. The functional data suggest that such R^0 complexes can, over time, convert to the G protein–coupled, active-state conformation, called RG, to thereby produce extended signaling responses.[71,72] Of further note, fluorescent ligand-tracking studies in cells show that these analogs move with the PTHR-1 into internalized vesicles during the times of prolonged signaling.[69,70] These observations suggested that the internalized PTHR-1–ligand–G protein complexes remained active after internalization, rather than dissociating and terminating signaling, as generally would be expected from classical GPCR theory. Evidence to support such prolonged endosomal signaling has since been demonstrated for at least several other class B GPCRs. In contrast to LA-PTH, other PTH peptides, including PTHrP(1–36), M-PTH(1–14), and abaloparatide, a PTHrP(1–34) analog used to treat osteoporosis (see later), bind poorly to the R^0 conformation of the PTHR-1, and thus mediate relatively transient signaling responses in cells, due to their relative selectivity for the RG conformation.[73]

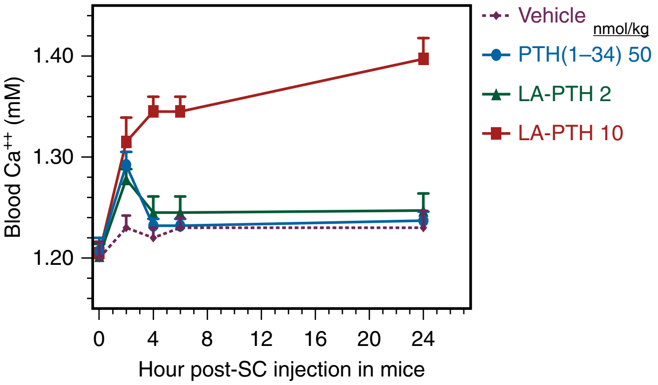


Fig. 47.6 Properties of a long-acting parathyroid hormone (PTH) analog (LA-PTH) *in vivo*. Shown are the capacities of PTH(1–34) and LA-PTH, which is comprised of a modified M-PTH(1–14) segment joined to a modified PTH-related peptide(15–36) segment, to stimulate increases of blood ionized calcium in mice following a single subcutaneous (SC) injection. PTH(1–34) was injected at a peptide dose of 50 nmol/kg, and LA-PTH was injected at doses of 2 and 10 nmol/kg; vehicle was injected as a control. The long duration of the calcemic responses induced by LA-PTH, relative to that of PTH(1–34), correlates with the prolonged cyclic 3′,5′-adenosine monophosphate response induced by LA-PTH in cells, and with the greater affinity with which it binds to the R^0 conformation of the PTH receptor type-1. (Adapted from Maeda A, Okazaki M, Baron DM, et al. Critical role of parathyroid hormone (PTH) receptor-1 phosphorylation in regulating acute responses to PTH. *Proc Natl Acad Sci U S A*. 2013;110:5864–5869.)

The biological relevance of such ligand-directed conformational selectivity at the PTHR-1 remains to be established. A potential clinical relevance, however, has been indicated, most notably by the capacity of LA-PTH to induce sustained elevations in serum calcium when injected into animals (Fig. 47.6).[71,72,74] Accordingly, LA-PTH has recently entered into a phase I clinical trial to evaluate the analog as a potential alternative therapy for hypoparathyroidism, a disease that requires careful therapeutic control of blood calcium levels for extended periods.[75] Abaloparatide, a PTHrP(1–34) analog in use as an osteoporosis therapy,[76] binds selectively to RG and mediates transient cAMP signaling responses in cells.[73] The distinct profile of receptor conformation selectivity exhibited by this peptide may facilitate its efficacy in osteoporosis patients, as PTH ligand–induced increases in bone mass are well known to be dependent on a pulsatile, rather than continuous, mode of exposure, although the underlying mechanisms for these effects are not fully understood.[77]

Disease Mutations in the PTHR-1 and its Ligands

Both gain-of-function and loss-of-function mutations in the PTHR-1 have been linked to disease pathologies that involve disturbances in skeletal development, and in some cases calcium homeostasis. Gain-of-function mutations have been identified in patients with JMC.[78] Although ultrarare, all JMC patients to date (n ≈ 25) have a heterozygous missense mutation that impacts the HETY motif in the PTHR-1 TMD region described earlier. This observation strikingly highlights the significance of the HETY motif in maintaining proper PTHR-1 function. The JMC mutations are: His223→Arg in TM2, which is the most common JMC allele, Thr410→Pro/Arg in TM6, and Ile458→Arg/Lys. The JMC mutations likely perturb the H-bond and electrostatic interaction network in the HETY motif to cause an increase in the rate of conversion of the receptor from the inactive state to the active state in the absence of a bound agonist ligand, and hence an excess in basal levels of cAMP in cells of bone, kidney, and growth-plate cartilage, and ultimately the severe abnormalities in skeletal development and mineral ion homeostasis, including dwarfism and hypercalcemia, that characterizes the JMC phenotype.[78]

Several pathologic conditions, again rare, have been linked to loss-of-function mutations in the PTHR-1. Blomstrand osteochondrodysplasia (BOCD), is a recessive neonatal lethal disorder characterized by premature ossification of the skeleton, as well as defects in development of the lung, aorta, and breasts. The physical defects and neonatal death associated with BOCD parallel those seen in mice in which the *PTHR-1* gene is ablated. The mutations of BOCD are homozygous or compound heterozygous changes that impact the structure or function of the PTHR-1 protein, e.g., of conserved Pro132 to Leu, or the splicing or stability of the mRNA. Heterozygous loss-of-function mutations in the PTHR-1, including Pro132Leu, have been identified in patients with familial primary failure of tooth eruption,[79] which correlates with mouse studies showing a clear role for PTHR-1 signaling and the paracrine actions of PTHrP in tooth development.[80] Ollier disease, a developmental disorder characterized by the presence of multiple enchondromas, has been linked to several heterozygous PTHR-1 mutations, including Gly121→Glu, Ala122→Thr, and Arg150→Cys, which are each located in the ECD. Functional assessment in cells of these three mutations suggest that they reduce ligand-binding affinity and/or expression of the receptor, but it remains unclear how the mutations lead to the enchondromas seen in patients.

Disease mutations have also been identified in the genes encoding the ligands PTH and PTHrP. Heterozygous mutations impacting expression or function of PTHrP have been identified in patients with brachydactyly type E, at least some of whom also exhibit tooth eruption failure, short stature, shortened limbs, and/or dysmorphic craniofacial features.[81] These phenotypes are again fully consistent with the critical role that paracrine-acting PTHrP and PTHR-1 signaling play in skeletal development. Mutations impacting PTH function have been found in patients with hypoparathyroidism. The mutations identified include dominant-acting missense mutations in the PTH prepro region, as well as two very rare homozygous missense mutations in the mature PTH polypeptide, Arg25→Cys and Ser1→Pro.[37,82] Here, the hypocalcemia characterizing the disease reflects the critical role that endocrine-acting PTH plays in maintaining calcium homeostasis.

KEY POINTS

- Heterozygous loss-of-function mutations in parathyroid hormone receptor type-1 are associated with a primary failure of tooth eruption, while gain-of-function mutations cause Jansen's metaphyseal chondrodysplasia (JMC), a disease of bone development and mineral ion homeostasis. JMC mutations have been found at only three receptor sites, consistent with the ultrarare frequency of the disease.

Small-Molecule Ligands for the Parathyroid Hormone Receptor Type-1

Several small-molecule ligands have been identified for the PTHR-1. This includes the agonist PCO371 and several antagonist compounds.[83,84] The precise mode of action used by such compounds remains to be elucidated; however, functional studies indicate that PCO371 binds to the TMD portion of the receptor. Although PCO371 binds to the PTHR-1 with a low-micromolar affinity, and is thus much weaker than PTH(1–34), it can be administered orally, whereas PTH peptides must be injected for therapeutic use. Based on this bioavailability and a prolonged pharmacokinetic profile, PCO371 has entered into early-stage clinical testing as a candidate therapy for hypoparathyroidism. These studies should provide further clues as to how new compounds can be developed that function as more potent and orally active agonists or antagonists on the PTHR-1 .

THE PARATHYROID HORMONE RECEPTOR TYPE-2 AND TIP39

The PTH receptor type-2 (PTHR-2) was originally identified by screening a human brain cDNA library for sequences homologous to the PTHR-1, and thus exhibits 51% identity to the PTHR-1 at the amino acid sequence level. The human PTHR-2 exhibits at least some reactivity towards PTH, but it does not bind PTHrP. Further studies on the rat PTHR-2 revealed weak, if any, binding of either PTH or PTHrP. This finding, and the abundant expression of the PTHR-2 mRNA in brain, led to a search for the endogenous cognate PTHR-2 ligand in brain tissue. Fractionation of hypothalamic extracts then revealed such a peptide ligand, which was called tubular infundibular peptide of 39 amino acids, or TIP39. This peptide efficiently binds to and activates both the rat and human PTH2 receptors and is thus presumed to be the natural agonist for the PTHR-2.

TIP39 exhibits weak amino acid sequence homology to the corresponding N-terminal regions of PTH and PTHrP and is extended N-terminally by two amino acids. TIP39 and the PTHR-2 are expressed in brain, testes, and skin. Studies in rodents provide evidence for a neuroendocrine role for the PTHR-2 and TIP39 in mediating pain and stress responses, as well as the lactation response to nursing.[85,86] A paracrine role in the development of the testis is suggested by the finding that male mice lacking the gene for either TIP39 or PTHR-2 are sterile due to a failure in spermatid formation. PTHR-2 and TIP39 are also expressed in keratinocytes, and female mice deleted for the *PTHR-2* gene exhibit a thickened epidermis and increased expression of keratin-14, suggesting a role in keratinocyte differentiation and in the maintenance and/or development of the dermal layer.[87] Genes corresponding to *TIP39* and *PTHR-2* are detected in the genomes of primitive fish, such as the sea lamprey, suggesting an emergence during early vertebrate evolution; however, neither gene is detected in birds.[12,13] No disease phenotype in humans has been linked to the PTHR-2 or TIP39.

SUMMARY

The actions of the N-terminal portion of PTH and PTHrP (or fragments of these regions) are mediated through a single receptor, PTHR-1, a class B GPCR that signals through the cAMP/PKA as well as the inositol triphosphate/calcium/diacylglycerol/PKC pathways. The PTHR-1 is expressed in a variety of fetal and adult tissues. The PTHR-1 in kidney and bone mediates the endocrine actions of PTH in mineral ion homeostasis, while in the growth plates and other developing tissues it mediates the autocrine/paracrine actions of PTHrP. Mutations in the PTHR-1 or its ligands are associated with several diseases of calcium homeostasis and skeletal development. The PTHR-1 adopts multiple conformational states, which can be selectively bound by structurally distinct peptide ligands to achieve disparate biological outcomes. The PTH-2 receptor and TIP39 appear to mediate neuroendocrine and potentially developmental functions but are still not well characterized biologically. A potential biological role for the C-terminal regions of either PTH or PTHrP has not been fully defined.

REFERENCES

1. Edmonston D, Wolf M. FGF23 at the crossroads of phosphate, iron economy and erythropoiesis. *Nat Rev Nephrol*. 2020;16:7–19.
2. Blau JE, Collins MT. The PTH-Vitamin D-FGF23 axis. *Rev Endocr Metab Disord*. 2015;16:165–174.
3. Christov M, Juppner H. Phosphate homeostasis disorders. *Best Pract Res Clin Endocrinol Metab*. 2018;32:685–706.
4. Collip JB. Extraction of a parathyroid hormone which will prevent or control parathyroid tetany and which regulates the level of blood calcium. *J Biol Chem*. 1925;63:395–438.
5. Albright F, Bauer W, Ropes M, et al. Studies of calcium and phosphorus metabolism. IV. The effect of the parathyroid hormone. *J Clin Invest*. 1929;7:139–181.
6. Aurbach GD. Isolation of parathyroid hormone after extraction with phenol. *J Biol Chem*. 1959;234:3179.
7. Brewer Jr HB, Ronan R. Bovine parathyroid hormone: amino acid sequence. *Proc Natl Acad Sci*. 1970;67:1862.
8. Niall HD, Keutmann HT, Sauer RT, et al. The amino-acid sequence of bovine parathyroid hormone I. *Hoppe-Seyler's Z Physiol Chem*. 1970;351:1586–1588.
9. Potts Jr JT, Tregear GW, Keutmann HT, et al. Synthesis of a biologically active N-terminal tetratriacontapeptide of parathyroid hormone. *Proc Natl Acad Sci USA*. 1971;68:63–67.
10. Kronenberg HK, Igarashi T, Freeman MW, et al. Structure and expression of the human parathyroid hormone gene. *Recent Prog Horm Res*. 1986;42:641–663.
11. Leder BZ. Parathyroid hormone and parathyroid hormone-related protein analogs in osteoporosis therapy. *Curr Osteoporos Rep*. 2017;15:110–119.
12. Suarez-Bregua P, Torres-Nuñez E, Saxena A, et al. Pth4, an ancient parathyroid hormone lost in eutherian mammals, reveals a new brain-to-bone signaling pathway. *Faseb J*. 2017;31:569–583.
13. On JS, Chow BK, Lee LT. Evolution of parathyroid hormone receptor family and their ligands in vertebrate. *Front Endocrinol*. 2015;6:28.
14. Suarez-Bregua P, Cal L, Canestro C, et al. PTH reloaded: a new evolutionary perspective. *Front Physiol*. 2017;8:776.
15. On JS, Duan C, Chow BK, et al. Functional pairing of class B1 ligand-GPCR in Cephalochordate provides evidence of the origin of PTH and PACAP/glucagon receptor family. *Mol Biol Evol*. 2015;32:2048–2059.
16. Xie J, Sang M, Song X, et al. A new neuropeptide insect parathyroid hormone iPTH in the red flour beetle Tribolium castaneum. *PLoS Genet*. 2020;16:e1008772.
17. Cardoso JC, Felix RC, Power DM. Nematode and arthropod genomes provide new insights into the evolution of class 2 B1 GPCRs. *PLoS One*. 2014;9:e92220.
18. Cinque L, Sparaneo A, Penta L, et al. Autosomal dominant PTH gene signal sequence mutation in a family with familial isolated hypoparathyroidism. *J Clin Endocrinol Metab*. 2017;102:3961–3969.
19. D'Amour P. Acute and chronic regulation of circulating PTH: significance in health and in disease. *Clin Biochem*. 2012;45:964–969.
20. Brown EM. Four-parameter model of the sigmoidal relationship between parathyroid hormone release and extracellular calcium concentration in normal and abnormal parathyroid tissue. *J Clin Endocrinol Metab*. 1983;56:572–581.
21. Brown EM. Extracellular Ca^{2+} sensing, regulation of parathyroid cell function, and role of Ca^{2+} and other ions as extracellular (first) messengers. *Physiol Rev*. 1991;71:371–411.
22. Hannan FM, Kallay E, Chang W, et al. The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. *Nat Rev Endocrinol*. 2018;15:33–51.
23. Brown EM, Gamba G, Riccardi D, et al. Cloning and characterization of an extracellular Ca^{2+}-sensing receptor from bovine parathyroid. *Nature*. 1993;366:575–580.
24. Pándy-Szekeres G, Munk C, Tsonkov TM, et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res*. 2017;46:D440–D446.
25. Zhang C, Zhang T, Zou J, et al. Structural basis for regulation of human calcium-sensing receptor by magnesium ions and an unexpected tryptophan derivative co-agonist. *Sci Adv*. 2016;2:e1600241.
26. Geng Y, Mosyak L, Kurinov I, et al. Structural mechanism of ligand activation in human calcium-sensing receptor. *Elife*. 2016;5.
27. Liu H, Yi P, Zhao W, et al. Illuminating the allosteric modulation of the calcium-sensing receptor. *Proc Natl Acad Sci U S A*. 2020;117:21711–21722.
28. Gorvin CM, Frost M, Malinauskas T, et al. Calcium-sensing receptor residues with loss- and gain-of-function mutations are located in regions of conformational change and cause signalling bias. *Hum Mol Genet*. 2018;27:3720–3733.
29. Keller AN, Kufareva I, Josephs TM, et al. Identification of global and ligand-specific calcium sensing receptor activation mechanisms. *Mol Pharmacol*. 2018;93:619–630.
30. Gregory KJ, Kufareva I, Keller AN, et al. Dual action calcium-sensing receptor modulator unmasks novel mode-switching mechanism. *ACS Pharmacol Transl Sci*. 2018;1:96–109.
31. Centeno PP, Herberger A, Mun HC, et al. Phosphate acts directly on the calcium-sensing receptor to stimulate parathyroid hormone secretion. *Nat Commun*. 2019;10:4693.
32. Roszko KL, Bi R, Gorvin CM, et al. Knockin mouse with mutant Galpha11 mimics human inherited hypocalcemia and is rescued by pharmacologic inhibitors. *JCI Insight*. 2017;2:e91079.
33. Gorvin CM, Hannan FM, Howles SA, et al. Galpha11 mutation in mice causes hypocalcemia rectifiable by calcilytic therapy. *JCI Insight*. 2017;2:e91103.
34. Goltzman D, Hendy GN. The calcium-sensing receptor in bone--mechanistic and therapeutic insights. *Nat Rev Endocrinol*. 2015;11:298–307.
35. Al-Dujaili SA, Koh AJ, Dang M, et al. Calcium sensing receptor function supports osteoblast survival and acts as a co-factor in PTH anabolic actions in bone. *J Cell Biochem*. 2016;117:1556–1567.
36. Bringhurst FR, Stern AM, Yotts M. Peripheral metabolism of PTH: Fate of the biologically active amino-terminus in vivo. *Am J Physiol*. 1988;255:E886–E893.
37. Lee S, Mannstadt M, Guo J, et al. A homozygous [Cys25]PTH(1-84) mutation that impairs PTH/PTHrP receptor activation defines a novel form of hypoparathyroidism. *J Bone Miner Res*. 2015.
38. Sato T, Courbebaisse M, Ide N, et al. Parathyroid hormone controls paracellular Ca^{2+} transport in the thick ascending limb by regulating the tight-junction protein Claudin14. *Proc Natl Acad Sci Unit States Am*. 2017;114:E3344.
39. Hou J, Renigunta V, Nie M, et al. Phosphorylated claudin-16 interacts with Trpv5 and regulates transcellular calcium transport in the kidney. *Proc Natl Acad Sci Unit States Am*. 2019;116:19176.
40. Meyer MB, Benkusky NA, Kaufmann M, et al. A kidney-specific genetic control module in mice governs endocrine regulation of the cytochrome P450 gene Cyp27b1 essential for vitamin D(3) activation. *J Biol Chem*. 2017;292:17541–17558.
41. Guo J, Khatri A, Maeda A, et al. Prolonged pharmacokinetic and pharmacodynamic actions of a pegylated parathyroid hormone (1-34) peptide fragment. *J Bone Miner Res*. 2017;32:86–98.
42. Sneddon WB, Ruiz GW, Gallo LI, et al. Convergent signaling pathways regulate parathyroid hormone and fibroblast growth factor-23 action on NPT2A-mediated phosphate transport. *J Biol Chem*. 2016;291:18632–18642.
43. Juppner H. Genetic and epigenetic defects at the GNAS locus cause different forms of pseudohypoparathyroidism. *Ann Endocrinol*. 2015;76:92–97.
44. Zhang Q, Xiao K, Paredes JM, et al. Parathyroid hormone initiates dynamic NHERF1 phosphorylation cycling and conformational changes that regulate NPT2A-dependent phosphate transport. *J Biol Chem*. 2019;294:4546–4571.
45. Martins JS, Liu ES, Sneddon WB, et al. 1,25-Dihydroxyvitamin D maintains brush border membrane NaPi2a and Attenuates phosphaturia in Hyp mice. *Endocrinology*. 2019;160:2204–2214.
46. Parsons JA, Robinson CJ. Calcium shift into bone causing transient hypocalcaemia after injection of parathyroid hormone. *Nature*. 1971;230:581–582.
47. Talmage RV, Doppelt SH, Fondren FB. An interpretation of acute changes in plasma 45Ca following parathyroid hormone administration to thyroparathyroidectomized rats. *Calcif Tissue Res*. 1976;22:117–128.

48. Wein MN, Liang Y, Goransson O, et al. SIKs control osteocyte responses to parathyroid hormone. *Nat Commun*. 2016;7:13176.
49. Nishimori S, O'Meara MJ, Castro CD, et al. Salt-inducible kinases dictate parathyroid hormone 1 receptor action in bone development and remodeling. *J Clin Invest*. 2019;129:5187–5203.
50. Wein MN, Foretz M, Fisher DE, et al. Salt-inducible kinases: physiology, regulation by cAMP, and therapeutic potential. *Trends Endocrinol Metab*. 2018;29:723–735.
51. Ono N, Kronenberg HM. Bone repair and stem cells. *Curr Opin Genet Dev*. 2016;40:103–107.
52. Balani DH, Ono N, Kronenberg HM. Parathyroid hormone regulates fates of murine osteoblast precursors in vivo. *J Clin Invest*. 2017;127:3327–3338.
53. Balani DH, Kronenberg HM. Withdrawal of parathyroid hormone after prolonged administration leads to adipogenic differentiation of mesenchymal precursors in vivo. *Bone*. 2018.
54. Maridas DE, Rendina-Ruedy E, Helderman RC, et al. Progenitor recruitment and adipogenic lipolysis contribute to the anabolic actions of parathyroid hormone on the skeleton. *Faseb J*. 2019;33:2885–2898.
55. Kim SW, Lu Y, Williams EA, et al. Sclerostin antibody administration converts bone lining cells into active osteoblasts. *J Bone Miner Res : The Official Journal of the American Society for Bone and Mineral Research*. 2017;32:892–901.
56. Xiong J, Piemontese M, Onal M, et al. Osteocytes, not osteoblasts or lining cells, are the main source of the RANKL required for osteoclast formation in remodeling bone. *PLoS One*. 2015;10:e0138189.
57. Xiong J, Cawley K, Piemontese M, et al. Soluble RANKL contributes to osteoclast formation in adult mice but not ovariectomy-induced bone loss. *Nat Commun*. 2018;9:2909.
58. Cawley KM, Bustamante-Gomez NC, Guha AG, et al. Local production of osteoprotegerin by osteoblasts suppresses bone resorption. *Cell Rep*. 2020;32:108052.
59. Robling AG, Bonewald LF. The osteocyte: new insights. *Annu Rev Physiol*. 2020;82:485–506.
60. Jilka RL, O'Brien CA. The role of osteocytes in age-related bone loss. *Curr Osteoporos Rep*. 2016;14:16–25.
61. Sato T, Verma S, Andrade CDC, et al. A FAK/HDAC5 signaling axis controls osteocyte mechanotransduction. *Nat Commun*. 2020;11:3282.
62. Wein MN. Parathyroid hormone signaling in osteocytes. *JBMR Plus*. 2018;2:22–30.
63. de Graaf C, Song G, Cao C, et al. Extending the structural view of class B GPCRs. *Trends Biochem Sci*. 2017;42:946–960.
64. Ehrenmann J, Schoppe J, Klenk C, et al. High-resolution crystal structure of parathyroid hormone 1 receptor in complex with a peptide agonist. *Nat Struct Mol Biol*. 2018;25:1086–1092.
65. Zhao L-H, Ma S, Sutkeviciute I, et al. Structure and dynamics of the active human parathyroid hormone receptor-1. *Science*. 2019;364:148–153.
66. Dong M, Deganutti G, Piper SJ, et al. Structure and dynamics of the active Gs-coupled human secretin receptor. *Nat Commun*. 2020;11:4137.
67. Hilger D, Kumar KK, Hu H, et al. Structural insights into differences in G protein activation by family A and family B GPCRs. *Science*. 2020;369:eaba3373.
68. Liang YL, Khoshouei M, Radjainia M, et al. Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature*. 2017;546:118–123.
69. Sutkeviciute I, Clark LJ, White AD, et al. PTH/PTHrP receptor signaling, allostery, and structures. *Trends Endocrinol Metab*. 2019;30:860–874.
70. Cheloha RW, Gellman SH, Vilardaga JP, et al. PTH receptor-1 signalling-mechanistic insights and therapeutic prospects. *Nat Rev Endocrinol*. 2015;11:712–724.
71. Shimizu M, Joyashiki E, Noda H, et al. Pharmacodynamic actions of a long-acting PTH analog (LA-PTH) in thyroparathyroidectomized (TPTX) rats and normal monkeys. *J Bone Miner Res*. 2016;7:1405–1412.
72. Noda H, Okazaki M, Joyashiki E, et al. Optimization of PTH/PTHrP hybrid peptides to derive a long-acting PTH analog (LA-PTH). *JBMR Plus*. 2020;4:e10367.
73. Hattersley G, Dean T, Corbin BA, et al. Binding selectivity of abaloparatide for PTH-type-1-receptor conformations and effects on downstream signaling. *Endocrinology*. 2016;157:141–149.
74. Bi R, Fan Y, Lauter K, et al. Diphtheria toxin- and GFP-based mouse models of acquired hypoparathyroidism and treatment with a long-acting parathyroid hormone analog. *J Bone Miner Res*. 2016;31:975–984.
75. Winer KK. Advances in the treatment of hypoparathyroidism with PTH 1-34. *Bone*. 2019;120:535–541.
76. Boyce EG, Mai Y, Pham C. Abaloparatide: review of a next-generation parathyroid hormone agonist. *Ann Pharmacother*. 2018;52:462–472.
77. Bilezikian JP, Hattersley G, Fitzpatrick LA, et al. Abaloparatide-SC improves trabecular microarchitecture as assessed by trabecular bone score (TBS): a 24-week randomized clinical trial. *Osteoporos Int*. 2018;29:323–328.
78. Saito H, Noda H, Gatault P, et al. Progression of mineral ion abnormalities in patients with jansen metaphyseal chondrodysplasia. *J Clin Endocrinol Metab*. 2018;103:2660–2669.
79. Roth H, Fritsche LG, Meier C, et al. Expanding the spectrum of PTH1R mutations in patients with primary failure of tooth eruption. *Clin Oral Investig*. 2014;18:377–384.
80. Takahashi A, Nagata M, Gupta A, et al. Autocrine regulation of mesenchymal progenitor cell fates orchestrates tooth eruption. *Proc Natl Acad Sci Unit States Am*. 2019;116:575–580.
81. Reyes M, Bravenboer B, Juppner H. A heterozygous splice-site mutation in PTHLH causes autosomal dominant shortening of metacarpals and metatarsals. *J Bone Miner Res*. 2019;34:482–489.
82. Gild ML, Bullock M, Luxford C, et al. Congenital hypoparathyroidism associated with elevated circulating nonfunctional parathyroid hormone due to novel PTH mutation. *J Clin Endocrinol Metab*. 2020;105.
83. Tamura T, Noda H, Joyashiki E, et al. Identification of an orally active small-molecule PTHR1 agonist for the treatment of hypoparathyroidism. *Nat Commun*. 2016;7:13384.
84. Arai Y, Kiyotsuka Y, Kagechika K, et al. Discovery of novel, potent, and orally bioavailable pyrido[2,3-d][1]benzazepin-6-one antagonists for parathyroid hormone receptor 1. *Bioorg Med Chem*. 2020:115524.
85. Cservenak M, Keller D, Kis V, et al. A thalamo-hypothalamic pathway that activates oxytocin neurons in social contexts in female rats. *Endocrinology*. 2017;158:335–348.
86. Gellen B, Zelena D, Usdin TB, et al. The parathyroid hormone 2 receptor participates in physiological and behavioral alterations of mother mice. *Physiol Behav*. 2017;181:51–58.
87. Sato E, Muto J, Zhang LJ, et al. The parathyroid hormone second receptor PTH2R and its ligand tuberoinfundibular peptide of 39 residues TIP39 regulate intracellular calcium and influence keratinocyte differentiation. *J Invest Dermatol*. 2016;136:1449–1459.
88. Noda H, Guo J, Khatri A, et al. An inverse agonist ligand of the PTH receptor partially rescues skeletal defects in a mouse model of Jansen's metaphyseal chondrodysplasia. *J Bone Miner Res*. 2020;35:540–549.
89. Maeda A, Okazaki M, Baron DM, et al. Critical role of parathyroid hormone (PTH) receptor-1 phosphorylation in regulating acute responses to PTH. *Proc Natl Acad Sci U S A*. 2013;110:5864–5869.

48

Parathyroid Hormone-Related Protein

Diego Y. Grinman and John J. Wysolmerski

OUTLINE

The Pthrp Gene and the Pth/Pthrp Gene Family, 818
Posttranslational Processing Of Parathyroid Hormone-Related Protein, 819
Parathyroid Hormone-Related Protein Receptors, 821
Nuclear Parathyroid Hormone-Related Protein, 821
Physiologic Functions of Parathyroid Hormone-Related Protein, 823
  The Skeleton, 823
  The Mammary Gland, 824
  The Skin, 826
  Placenta, 827
  Endocrine Pancreas, 827
  Smooth Muscle and the Cardiovascular System, 827
  Teeth, 828
  The Nervous System, 828
  The Immune System, 829
Parathyroid Hormone-Related Protein In Disease, 829
  Cancer, 829
  Skeletal Disease, 830
  Skin Disease, 830
  Diabetes, 830
Conclusion, 830

KEY POINTS

- Parathyroid hormone-related protein (PTHrP) was originally discovered as a cause of elevated calcium levels in patients with cancer.
- PTHrP is part of a small gene family related to parathyroid hormone (PTH).
- Both PTH and PTHrP bind to and activate the same G protein–coupled receptor known as the type 1 PTH/PTHrP receptor.
- PTHrP is expressed in many tissues during development and has been shown to be critical for the normal development of bones, cartilage, teeth, and the mammary gland.
- Mammary gland–derived PTHrP has also been shown to contribute to the regulation of calcium and bone metabolism during lactation.
- PTHrP contributes to the development and/or progression of breast cancer.
- PTHrP can regulate smooth muscle cell tone and contraction in blood vessels and in the uterus.
- PTHrP can regulate pancreatic β cell mass and may be useful in the treatment of diabetes.
- A synthetic PTHrP analog was approved for the treatment of postmenopausal women with osteoporosis.

The discovery of parathyroid hormone-related protein (PTHrP) was anticipated in 1941, when Fuller Albright postulated that tumors may produce a parathyroid hormone (PTH)-like humor.[1] We now understand that PTHrP and PTH are related molecules that can both stimulate the same type 1 PTH/PTHrP receptor (PTH1R).[1,2] As a result, when PTHrP is secreted by tumors, it mimics its cousin, PTH, and leads to excessive bone resorption and hypercalcemia. After its discovery, it became obvious that PTHrP was widely expressed, prompting efforts to understand its normal functions. Over the past 34 years, we have defined critical functions in some specific tissues, but it is remarkable how little we really understand of its function(s) in all the various tissues that produce PTHrP. The goal of this chapter is to outline the basic facts of PTHrP biology, especially its contributions to physiology and pathophysiology.

THE *PTHRP* GENE AND THE *PTH/PTHRP* GENE FAMILY

In humans, PTHrP is encoded by a single-copy gene (the official gene symbol is *PTHLH*) containing eight exons and three promoters that is located on the short arm of chromosome 12 (Fig. 48.1).[2] Alternative splicing at the 3′ end of the gene gives rise to three different classes of mRNAs coding for specific translation products of 139, 141, or 173 amino acids, each of which contains the same 36–amino acid prepro, or signal, sequence and the same initial 139 amino acids. Each class of mRNA also contains AU-rich regulatory elements within the 3′ untranslated region that are associated with rapid but regulated mRNA turnover.[2] The physiologic significance of the different classes of PTHLH transcripts remains unclear, and in rodents and lower vertebrates such as birds and fish PTHrP 1–173 does not exist.[2,3] Two studies examining breast cancer cells have suggested that certain patterns of promoter usage and/or splicing patterns may be associated with different behaviors of the cancer cells, such as the likelihood metastasizing to bone or their invasive behavior.[4,5]

PTHLH mRNA is also regulated by microRNAs. miR-182-5p and miR-195 have been shown to downregulate *PTHLH* mRNA and PTHrP protein levels, altering the differentiation and function of cultured human chondrocytes.[6,7] Moreover, compared with healthy tissue, miR-195 levels were shown to be increased in tibial plateaus samples from patients with osteoarthritis, which demonstrated reduced PTHrP mRNA and protein levels. Given that PTHrP has been suggested to prevent osteoarthritis, targeting miR-195 could be a novel therapeutic approach for the treatment of this condition.[6]

The *PTHLH* and *PTH* genes share structural elements and sequence homology, establishing that they are both members of the

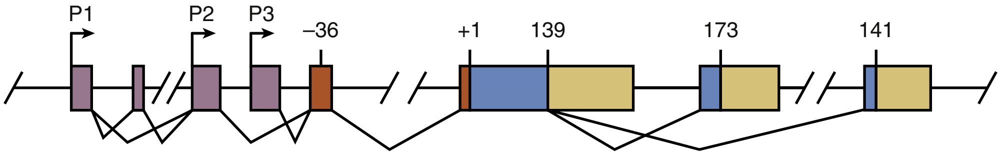


Fig. 48.1 **Diagram of the *PTHLH* gene.** The gene consists of eight exons *(colored boxes).* The 5′ untranslated region is represented in *purple*, the prepro sequences in *red*, the coding region in *blue*, and the 3′ untranslated region in *yellow*. The three promoter regions are denoted by *arrows,* and the possible splicing patterns are noted by the *lines* connecting the different exons.

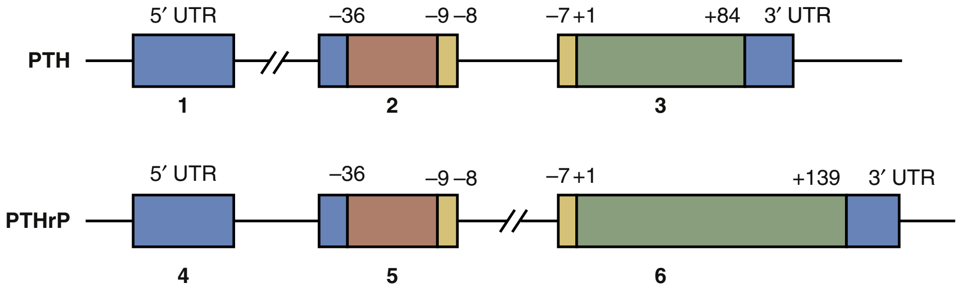


Fig. 48.2 **Side-by-side comparison of the prepro and coding regions of the *PTH* and *PTHrP* genes.** Note that the exon/intron configuration of both genes is identical for those portions encoding the signal peptides and N-terminal portions of each protein. *UTR,* Untranslated region.

same gene family.[1,2] The exon/intron organization of that portion of both genes encoding the prepro sequences and the initial portion of the mature peptides is identical (Fig. 48.2). Furthermore, there is high sequence homology at the N-terminal portion of both genes, such that the peptides share eight of the first 13 amino acids and a high degree of predicted secondary structure over the next 21 amino acids. These common sequences allow both peptides to bind and activate the same receptor, known as the *type 1 PTH/PTHrP receptor* (*PTH1R*) explaining the ability of PTHrP to cause hypercalcemia in humoral hypercalcemia of malignancy (HHM).[8] These structural similarities, together with the location of the two genes on related chromosomes in the human genome (short arm of chromosome 11 for *PTH*; short arm of chromosome 12 for *PTHLH*), indicate that the two genes likely arose from a common ancestor via gene duplication. The PTH family also contains the *PTH-L* gene and the more distantly related tuberoinfundibular peptide 39 (*TIP39*) gene. All of these genes were derived from a common ancestor and emerged concurrent with the evolution of vertebrates.[3] Furthermore, fish have two *Pth* genes, two *Pthlh* genes, one *Tip39* gene, and one *Pth-l* gene; amphibians have one *Pth* gene, one *Pthlh* gene, and one *Pth-l* gene; and mammals only retain one *PTH* gene and one *PTHLH* gene.[3] Thus, PTHrP is a member of an ancient family of PTH-like peptides that appears to be larger and more diverse in lower vertebrates than in mammals. The evolutionary relationships between members of the PTH gene family are also discussed in Chapter 47. There is a high degree of homology within PTHrPs of different species for the first 111 amino acids. There is less homology between amino acids 111 and 139, and PTHrP sequences from 141 to 173 are specific to the human gene.[2]

KEY POINTS

- Parathyroid hormone-related protein is part of a small gene family related to parathyroid hormone.

PTHLH mRNA has been found in almost every organ. A multitude of different hormones and growth factors have been described to regulate the transcription and/or stability of *PTHLH* mRNA. This chapter will not list all of these factors but will discuss the regulation of PTHrP in different cell types as part of the discussion of the functions of PTHrP in specific organs. The reader is referred to other reviews for a more comprehensive cataloging of the factors described to regulate PTHrP expression.[2,9] There are several broad themes to emphasize at this juncture. First, PTHrP is generally expressed at much higher levels and in many more sites during fetal development than it is during adult life.[2] Second, during development, PTHrP is expressed in close association with the PTH1R receptor (discussed later) in a “hand-in-glove” pattern in which PTHrP is produced by forming epithelial organs and the PTH1R is produced by the enveloping mesenchyme (Fig. 48.3).[10] Third, analogous to PTH in the parathyroid glands, the calcium-sensing receptor (CaSR) has been found to regulate *PTHrP* gene expression in many cells.[11] Finally, *PTHLH* mRNA levels are induced by mechanical deformation in various types of smooth muscle cells[2] and by cells in the osteoblast lineage.[11]

POSTTRANSLATIONAL PROCESSING OF PARATHYROID HORMONE-RELATED PROTEIN

The primary translation product of PTHrP undergoes posttranslational processing events to produce an overlapping series of biological peptides.[2] The basic outline of PTHrP processing was elucidated in the first decade after its discovery, but little work has been performed in this area for many years. As a result, the full details of cell-specific PTHrP processing and the biological significance of the different PTHrP peptides remain only partly understood. The primary translation product predicted by the *PTHLH* cDNAs contains multiple potential processing sites, consisting of clusters of basic amino acids in pairs, triplets, or quadruplets. It has been shown that an arginine at position 37 serves as a cleavage site to generate PTHrP 1–36, which is secreted from many cell types.[2] The N-terminus is necessary for interaction with the PTH1R and PTHrP 1–36 appears fully able to bind and activate this receptor. Moreover, PTHrP 1–36 can be further processed by matrix

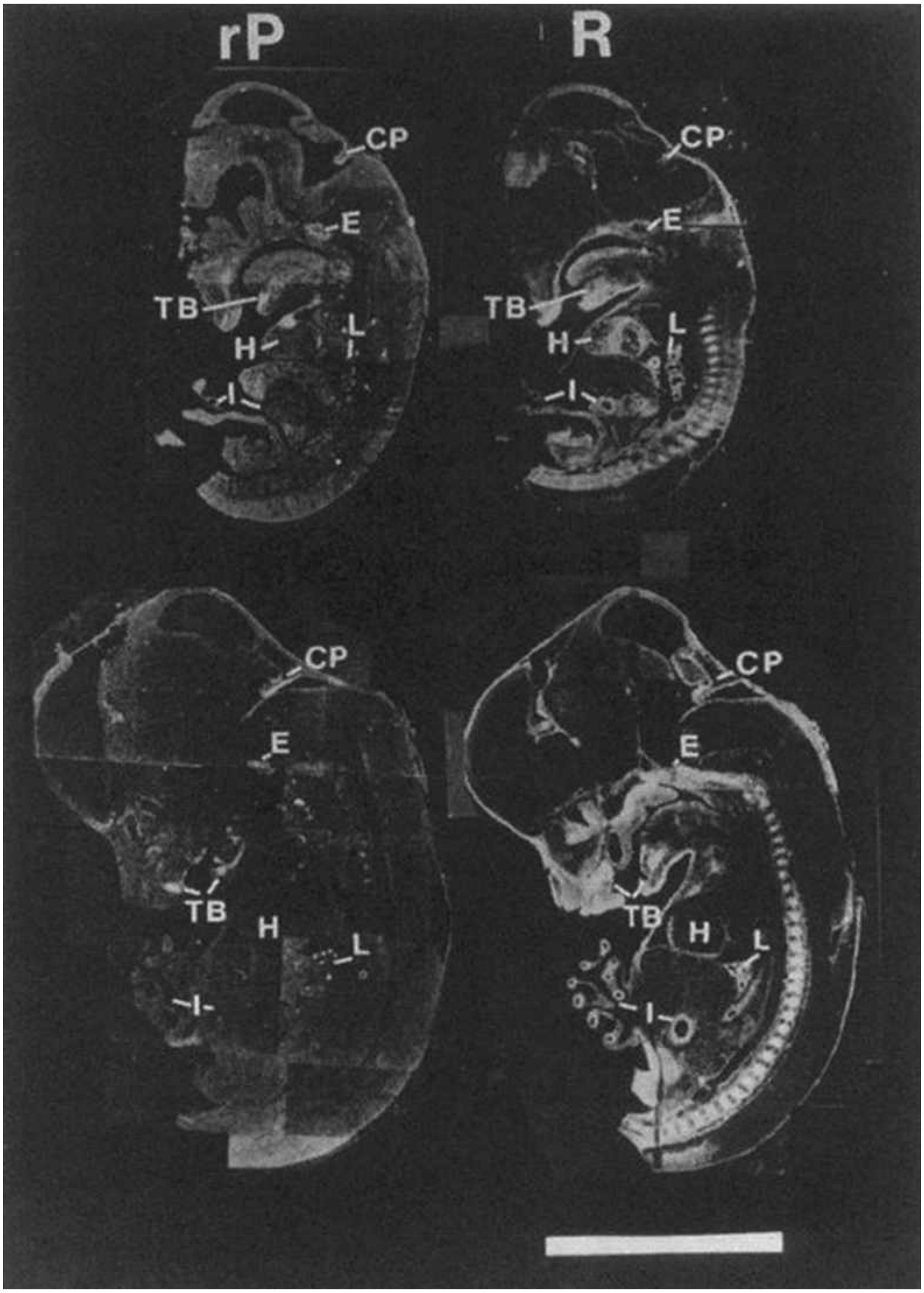


Fig. 48.3 Patterns of expression of parathyroid hormone-related protein (PTHrP) and type 1 parathyroid hormone/ parathyroid hormone-related protein receptor (PTH1R) mRNA in rat fetuses at embryonic day (E) 15 *(top)* and E16 *(bottom)*. Shown are dark-field photographs of ^{35}S-labeled complementary RNAs. Note that PTHrP is expressed in forming epithelial organs, and the PTH1R is expressed in the surrounding mesenchyme. *CP,* Choroid plexus; *E,* inner ear; *TB,* tooth bud; *L,* lung; *H,* heart; *I,* intestine; *rP*, parathyroid hormone-related protein; *R,* type 1 parathyroid hormone/ parathyroid hormone-related protein receptor. The scale bar represents 5000 microns. (Reproduced from Lee K, Deeds JD, Segre GV. Expression of parathyroid hormone-related peptide and its receptor messenger ribonucleic acids during fetal development in rats. *Endocrinology*. 1995;136:453–463.)

metalloproteinases to yield a stable product, PTHrP 1–17, secreted by rat and human cancer cells and still capable of activating the PTH1R to direct osteoblast differentiation and osteogenesis.[12] Longer forms of N-terminal–containing PTHrP are secreted from keratinocytes and mammary epithelial cells and circulate in patients with cancer and during lactation.[11] It is not clear what defines the C-terminus of these peptides, nor is it clear whether they differ in biological activity from PTHrP 1–36. Adding additional complexity, N-terminal PTHrP secreted by keratinocytes has been shown to be heavily glycosylated.[13] Furthermore, a recent study suggested that osteocytes secrete only full-length PTHrP without any cleavage products.[14] Clearly, there is great heterogeneity in the forms of PTHrP that contain the N-terminus of the molecule and that have the ability to interact with the PTH1R. Presumably, a similar variety of PTHrP peptides exists *in vivo*, but this question remains understudied and, therefore, the answer remains unknown.

KEY POINTS

- Parathyroid hormone-related protein undergoes posttranslational processing events to produce an overlapping series of biological peptides.

Cleavage of PTHrP at downstream sites leads to the secretion of midregion peptides including amino acids 38 to 94, 38 to 95, and 38 to 101.[2] The biology of these peptides is unclear, but the midregion of PTHrP stimulates placental calcium transport and modulates renal bicarbonate handling, and this portion of the molecule contains nuclear localization signals (see later).[15] Finally, C-terminal fragments consisting of amino acids 107 to 138 and 109 to 138 have been described. These peptides have been suggested to inhibit osteoclast function and stimulate osteoblast proliferation.[2,15] It is likely that the posttranslational processing of PTHrP is controlled for specific physiologic purposes, but, until the biology of specific PTHrP peptides (apart from those that contain the N-terminus) is better appreciated, little additional effort is likely to be expended on elucidating the specifics of PTHrP processing.

PARATHYROID HORMONE-RELATED PROTEIN RECEPTORS

N-terminal PTHrP binds to and activates the PTH1R a member of class B of the large family of seven transmembrane-spanning domain, G protein–coupled receptors.[9] This receptor is discussed at greater length in Chapter 47. Although binding of PTHrP 1–36 or PTH 1–34 to the receptor occurs via similar mechanisms, biochemical and biophysical data have demonstrated subtle differences in the binding of PTH versus PTHrP to the PTH1R. Binding of PTH favors a conformational state known as G^0, which is relatively stable and allows the receptor to undergo multiple rounds of G-protein activation generating cyclic adenosine monophosphate (cAMP) over a longer period of time. By contrast, the binding of PTHrP 1–36 to the receptor favors a more labile RG conformational state, which does not support multiple rounds of G-protein activation and, thus, generates less cAMP than the binding of PTH.[16] The three-dimensional crystal structures of PTH 1–34 and PTHrP 1–36 bound to the extracellular domain (ECD) of the PTH1R offer a potential explanation for these pharmacologic observations.[17,18] Portions of both peptides from amino acids 14 through 30 form an amphipathic helix, which binds to a specific cleft in the receptor's ECD. However, the α-helix formed by PTH is longer than the one formed by PTHrP, allowing PTH to fit more tightly into the receptor's binding cleft than PTHrP. Interestingly, a recent study showed that full-length PTHrP elicits a more sustained cAMP response than PTHrP 1–36. This action seems to be endocytosis-dependent, because blocking receptor internalization inhibited the cAMP activation response.[19] Thus, while the PTH1R can be activated by either PTH or PTHrP, the receptor appears to be able to tell the difference between the two peptides and, perhaps, even between different forms of PTHrP. These different ligand–receptor interactions may stimulate different biological processes in response to one protein versus the other. This seems to be the case in humans subjected to continuous infusion of PTH versus PTHrP for 3 to 7 days. Subjects became hypercalcemic with lower doses of PTH 1–34 than with PTHrP 1–36.[16] In addition, PTHrP was also less effective than PTH at stimulating the renal 1-α-hydroxylase enzyme and producing 1,25-dihydroxyvitamin D.

KEY POINTS

- Both parathyroid hormone (PTH) and parathyroid hormone-related protein (PTHrP) bind to and activate the same G protein–coupled receptor, known as the type 1 PTH/PTHrP receptor.

In addition to these ligand-specific differences in signaling, studies *in vitro* have suggested that PTH1R signaling can be modulated by other molecules. High extracellular calcium was recently shown to act as an allosteric modulator of the PTH1R prolonging the residence time of ligands on the receptor and, consequently, increasing both the duration of receptor activation and cAMP signaling.[20] In addition, adenosine triphosphate, whose levels were shown to increase in the bone extracellular fluid after mechanical loading, may potentiate PTH and PTHrP-induced adenylyl cyclase activation.[21] Therefore, alterations in the target cell microenvironment that may modulate the pharmacologic effects of PTHrP peptides, are important considerations for understanding PTHrP biology *in vivo*, as well as for strategies to potentiate current and future anabolic therapy for osteoporosis.

NUCLEAR PARATHYROID HORMONE-RELATED PROTEIN

In addition to being secreted and acting on the PTH1R PTHrP can also act within the nucleus. Two pathways through which PTHrP accesses the nucleus have been described: 1) PTHrP directly enters the cytoplasm rather than the endoplasmic reticulum (ER) for secretion; and 2) PTHrP is secreted from cells and then reenters the cytoplasm by endocytosis (Fig. 48.4). Two aspects of the *PTHLH* gene make the first pathway possible. First, *PTHLH* encodes an alternative downstream transcriptional start site that bypasses the signal peptide, allowing PTHrP to avoid secretion and remain in the cell.[15] Second, PTHrP's signal peptide is less hydrophobic than most, weakening its interaction with the signal recognition particle responsible for initiating translocation of the nascent polypeptide chain into the ER lumen.[22] Moreover, the proregion has positively charged residues that may affect the insertion of the growing nascent polypeptide into the ER translocation pore.[23] Either way, once in the cytoplasm, PTHrP shuttles into and out of the nucleus in a regulated fashion. There are classic nuclear localization sequences (NLSs) located between amino acids 84 to 93, and specific proteins have been shown to mediate entry into and export from the nucleus.[15] Importin β1 mediates the entry of PTHrP into the nucleus, in part by displacing PTHrP from microtubules and allowing it to transit the nuclear pore.[15] Nuclear export involves the actions of a related shuttle protein known as CRM1 and appears to involve specific sequences in the C-terminal region of the peptide.[15] Phosphorylation of PTHrP at Thr^{85} by the cell-cycle–associated cyclin-dependent kinase $p34^{cdc2}$ regulates nuclear import in a cell-cycle–dependent fashion.[15]

KEY POINTS

- In addition to being secreted and acting on the type 1 parathyroid hormone/parathyroid hormone-related protein receptor, parathyroid hormone-related protein can also act within the nucleus.

The second pathway through which PTHrP accesses the nucleus is by binding to the PTH1R at the cell surface. In ways that are poorly understood, PTHrP bound to the PTH1R can be internalized and transported into the nucleus.[24] The PTH1R has a potential NLS,

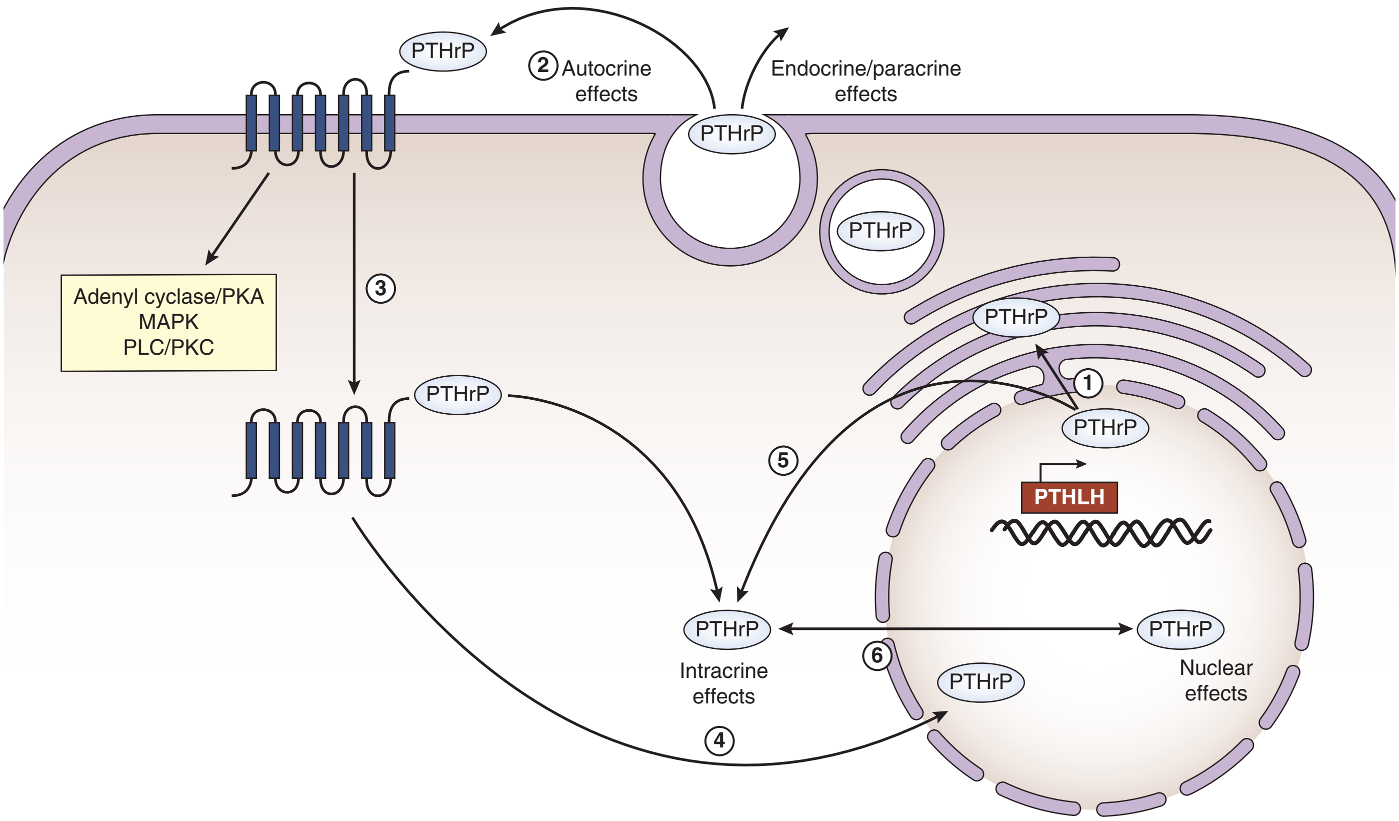


Fig. 48.4 Multiple pathways of parathyroid hormone-related protein (PTHrP) action. After being synthetized, PTHrP can be 1) targeted to the endoplasmic reticulum (ER) and secreted from the cell to act in a paracrine/endocrine or 2) autocrine fashion by binding to type 1 parathyroid hormone/parathyroid hormone-related protein receptor and activating signaling pathways. 3) PTHrP can also be internalized by endocytosis and, once in the cytoplasm, 4) imported to the nucleus. 5) PTHrP can also elude targeting into the ER and directly enter the cytoplasm, remain intracellular to act in an intracrine manner, and then 6) be transported to the nucleus by importin β. *PTHLH,* parathyroid hormone like hormone.

although it is not clear if this mediates active transport of the receptor into the nucleus together with PTHrP.[15] However, this second pathway may allow nuclear PTHrP signaling to operate in a paracrine fashion as well as through a cell autonomous pathway.

The function(s) of nuclear PTHrP are a focus of ongoing investigation, and very little is understood about its actions in the nuclear microenvironment. PTHrP has been shown to bind to RNA and, in some cells, it localizes to the nucleolus, suggesting that it may be involved in regulating RNA trafficking, ribosomal dynamics, and/or protein translation.[15] In cultured cells, nuclear PTHrP influences cell proliferation and/or apoptosis and often appears to oppose the effects of secreted PTHrP. In breast, colon, and prostate cancer cells, the nuclear pathway stimulates cell proliferation, protects cells from apoptosis or anoikis, and stimulates cell migration, while secreted PTHrP inhibits cell proliferation and promotes cell death.[15,25] This has been most carefully studied in vascular smooth muscle cells, where nuclear PTHrP stimulates proliferation by inducing the expression of c-myc and skp2, which, in turn, reduce levels of the cell-cycle inhibitor p27.[26] In contrast, secreted PTHrP has the opposite effects.

Two groups have examined the biology of nuclear PTHrP *in vivo* by replacing the endogenous mouse *Pthlh* gene with truncated versions, encoding PTHrP 1–84 or PTHrP 1–66, that exclude the nuclear localization signals and C-terminus.[27,28] This approach has complicated the interpretation of these experiments somewhat, because it is difficult to know whether the resulting phenotypes are due only to loss of nuclear signaling or also to loss of the C-terminal portion of PTHrP. Nevertheless, in both cases, while loss of nuclear and C-terminal PTHrP did not cause the developmental defects typical of PTHrP or PTH1R knockout mice, it did cause growth failure, premature osteoporosis, reduced hematopoiesis, altered energy metabolism, premature ageing of skin, abnormal brain development and, ultimately, premature death at approximately 2 weeks of age. There was a decrease in the proliferation of chondrocytes, osteoblasts, neurons, skin cells, and bone marrow cells, and an increase in apoptosis or senescence at these sites, as well as in the thymus and spleen. Loss of midregion and C-terminal PTHrP was associated with increased expression of senescence markers such as p21, p27, and $p16^{INK4a}$ and decreased expression of Bmi-1, which is involved in stem/progenitor cell maintenance.[11,29] This phenotype is consistent with the previous observations in cell culture and suggests that nuclear PTHrP may participate broadly in the regulation of cell proliferation and survival, as well as in stem/progenitor cell maintenance or self-renewal. However, experiments in the mammary gland have shown that *Pthlh* gene expression is severely reduced in mammary epithelial cells in PTHrP 1–84 knockin embryos, and that the resulting mammary phenotypes were the result of very low levels of PTHrP.[30] Although both groups documented normal production of PTHrP 1–84 or PTHrP 1–66 in some cell types, this is clearly not the case in all cells. Therefore, some of the reported findings in these mouse models may be due to reduced overall PTHrP production rather than loss of nuclear signaling.

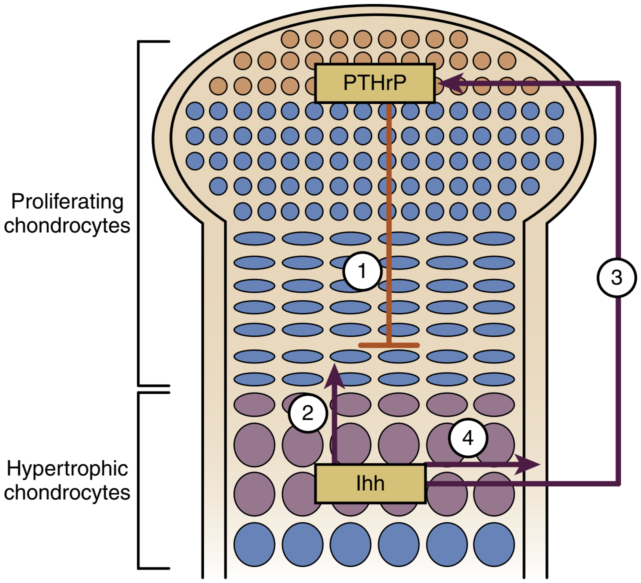


Fig. 48.5 Parathyroid hormone-related protein (PTHrP) and Indian hedgehog (Ihh) act as part of a negative feedback loop regulating chondrocyte proliferation and differentiation. The chondrocyte differentiation program proceeds from undifferentiated chondrocytes at the end of the bone, to proliferative chondrocytes within the columns and then to prehypertrophic and terminally differentiated, hypertrophic chondrocytes nearest the primary spongiosum. PTHrP is made by undifferentiated and proliferating chondrocytes at the ends of long bones. It acts through the type 1 parathyroid hormone/parathyroid hormone-related protein receptor (PTH1R) on proliferating and prehypertrophic chondrocytes to delay their differentiation, maintain their proliferation, and delay the production of Ihh, which is made by hypertrophic cells *(1)*. Ihh, in turn, increases the rate of chondrocyte proliferation *(2)* and stimulates the production of PTHrP at the ends of the bone *(3)*. Ihh also acts on perichondrial cells in order to generate osteoblasts of the bone collar *(4)*. Ihh, Indian Hedgehog (Reproduced from Ansari N, Ho PW, Crimeen-Irwin B, et al. Autocrine and paracrine regulation of the murine skeleton by osteocyte-derived parathyroid hormone-related protein. *J Bone Miner Res*. 2018;33:137–153.)

PHYSIOLOGIC FUNCTIONS OF PARATHYROID HORMONE-RELATED PROTEIN

PTHrP is expressed by many cells, and it has been shown to affect the proliferation, differentiation, and survival of many cell types *in vitro*. This chapter will not attempt to review all of these data. Rather, the discussion will focus on organs in which there are physiologic or genetic data *in vivo* implicating PTHrP in critical aspects of that organ's functions.

The Skeleton

One of the first-described and perhaps best-understood functions of PTHrP is to control the proliferation and differentiation of chondrocytes in the developing growth plate. Animal models of PTHrP overexpression and underexpression have documented that N-terminal PTHrP acts through the PTH1R to coordinate the rate of chondrocyte differentiation in order to maintain the orderly growth of long bones during development.[31] As illustrated in Fig. 48.5, the growth plate consists of columns of proliferating and differentiating chondrocytes that progressively enlarge to prehypertrophic and then hypertrophic chondrocytes. Both the production of new chondrocytes and their enlargement into hypertrophic cells together drive linear growth at the ends of bones. PTHrP is initially produced by cells at the periphery of the condensing mesenchyme that gives rise to the endochondral bones, and PTHrP expression becomes restricted to the ends of the forming bones as development proceeds.[32] The creation of a secondary ossification center after birth splits the zone of PTHrP production into that associated with the growth plate and that associated with the articular cartilage and joint (discussed later).[33] Within the growth plate, PTHrP is produced by immature, round chondrocytes at the top of the columns in response to another molecule known as *Indian Hedgehog* (*IHH*) that is produced by differentiating hypertrophic chondrocytes.[31] PTHrP, in turn, activates the PTH1R located on proliferating and prehypertrophic cells to maintain their proliferation and to slow their rate of differentiation into hypertrophic cells.[31] PTH1R expression is low in the round and proliferating chondrocytes but is upregulated within prehypertrophic chondrocytes. This system describes a local negative feedback loop that allows IHH and PTHrP to act together to regulate the orderly differentiation and hypertrophy of the chondrocyte columns and the elongation of the developing bones.[31] Genetic experiments in mice have demonstrated that loss of either PTHrP or the PTH1R leads to a foreshortening of the bones due to premature differentiation of the growth plate chondrocytes and loss of their proliferative capacity.[31,33] Interestingly, overexpression of PTHrP in chondrocytes also results in short bones, but this is caused by a severe impairment in chondrocyte differentiation and a disruption of the orderly architecture of the growth plate.[31] Loss of PTH1R function in human fetuses with Blomstrand's chondrodysplasia also impairs bone growth and mimics the phenotype of PTHrP- and PTH1R null mice.[34]

KEY POINTS

- Local parathyroid hormone-related protein controls the proliferation and differentiation of chondrocytes in the developing growth plate.

In recent years, many details about the network of signals though which PTHrP exerts it effects on chondrocyte differentiation have emerged. IHH increases PTHrP expression at the ends of the growth plate primarily through direct effects, but it may also act indirectly through TGF-β signaling.[31] IHH upregulates PTHrP expression in the growth plate by antagonizing the activity of the transcription factor Gli3.[31] PTHrP acts on chondrocytes through the PTH1R primarily by stimulating cAMP production and PKA activity, which, in turn, mediate a series of downstream events including phosphorylation of SIK3 and SOX9; inhibition of p57 expression; induction of Gli3, Bcl-2, and cyclin D1 expression; and, eventually, the phosphorylation and degradation of Runx2 and Runx3, two transcription factors necessary for chondrocyte differentiation.[31,35] It has also been demonstrated that PTHrP modulates chondrocyte differentiation by inducing the phosphorylation and inactivation of SIK3 kinase, which promotes the translocation of HDAC4 and HDAC5 into the nucleus. Once in the nucleus, HDAC4 and HDAC5 negatively regulate the activity of a network of transcription factors, including Zfp521, MEF2, and Runx2.[11,35,36] Multiple mouse genetic models have confirmed that this pathway is required for PTHrP's actions. As previously mentioned, PTHrP knockout mice die perinatally of accelerated chondrocyte hypertrophy, but simultaneous deletion of SIK3 in proliferating chondrocytes completely abrogated this phenotype.[35] Mice engineered to disrupt both HDAC4 and HDAC5 die at birth, with a phenotype similar to that of PTH1R knockout mice.[36] Finally, the lack of Zfp521 rescues the abnormal growth plate phenotype in mice that overexpress a constitutively active PTH1R.[35,37] In humans, mutations in the *PTHRP* and *HDAC4* genes cause a form of chondrodysplasia known as *brachydactyly* type E, which mimics skeletal abnormalities found in patients with mutations in the *GNAS* gene, suggesting that these genes are in the same genetic pathway in both mice and humans.[38] In summary, PTHrP signaling regulates a complex web of downstream events that affect chondrocyte proliferation and differentiation to ensure orderly bone growth during development.

There is a population of chondrocytes within the resting zone, located at the top of the growth plate, that persistently express PTHrP after fetal life.[33] Lineage tracing experiments showed that these PTHrP-positive resting-zone chondrocytes contributed to the pool of columnar chondrocytes but also became osteoblasts and marrow stromal cells beneath the growth plate, suggesting that they served as stem/progenitor cells for bone growth.[33] Importantly, whole-body ablation of the PTHrP-expressing cells or disruption of the *Pthlh* gene in chondrocytes after birth caused abnormal chondrocyte differentiation and premature loss of the growth plate.[33,39] These studies documented that PTHrP also regulates chondrocyte differentiation and maintenance in the mature growth plate, and raised the intriguing possibility that loss of PTHrP signaling may contribute to growth plate closure during the adolescent growth spurt, especially because IGF1 regulates PTHrP production by growth-plate chondrocytes.[40]

PTHrP is produced in other cartilaginous sites, such as the perichondrium that surrounds the costal cartilage and nasal cartilage, as well as by the subarticular chondrocyte population immediately subjacent to the hyaline cartilage lining the joint space.[11] These are sites of permanent cartilage, and the cells in these structures do not undergo differentiation, hypertrophy, and mineralization as in the growth plate. In these sites, it appears that PTHrP inhibits hypertrophic differentiation of chondrocytes and prevents the inappropriate encroachment of bone into these structures.[11] PTHrP is often regulated by mechanical forces and, interestingly, its expression within the articular cartilage of joints is load-dependent.[41] Deletion of PTHrP in these cells and/or unloading of the joints leads to abnormal mineralization of cells just under the articular surface, suggesting that PTHrP may normally maintain the undifferentiated state of articular chondrocytes and help prevent osteoarthritis.[11]

PTHrP is prominently expressed at the insertion sites of ligaments and tendons into bone, called *entheses*.[42] Specifically, it is expressed in the fibrous layer of the periosteum and in fibrous entheses that attach tendons and ligaments to the cortical bone surface. At these sites, PTHrP is induced by mechanical loading and upregulates RANKL production to induce osteoclast formation.[43] In the periosteum, these osteoclasts are uncoupled from osteoblast formation and serve to model the periosteal surface during growth and development. In the fibrous entheses, these osteoclasts erode the cortical surface to create the root system by which the tendons and ligaments are anchored to cortical bone. In general, this osteoclastic resorption is subsequently coupled to osteoblastic bone formation that anchors the insertion in place, but in certain sites, such as the tibial insertion of the medial collateral ligament, the resorption is uncoupled, which permits the insertion site to migrate to accommodate linear growth and thereby maintain its biomechanical function as a collateral ligament.[43]

A final function of PTHrP in the skeleton is the regulation of osteoblast and osteocyte anabolic function. There is some disagreement over the PTHrP-expressing population(s) of osteoblasts in the skeleton. Many studies in cultured osteoblasts demonstrate that these cells produce PTHrP and express the PTH1R providing the possibility for paracrine and/or autocrine regulation of osteoblast function.[44,45] Furthermore, there is extensive literature demonstrating that activation of the PTH1R in osteoblasts induces proliferation of osteoblasts and production of osteoclast-inducing cytokines, such as CSF-1 and RANKL, and inhibits full differentiation of osteoblasts (see Chapters 47 and 51).[44,45] However, studies using a PTHrP-lacZ knockin reporter mouse suggested that typical osteoblasts on trabecular and endosteal surfaces did not express PTHrP.[11] Nevertheless, genetic manipulation of the *PTHRP* gene in osteoblasts *in vivo* causes profound changes in bone homeostasis. First, it was shown that, while heterozygous PTHrP-null mice are normal at birth, they develop progressive osteopenia with age.[46] In addition, selective deletion of the *Pthlh* gene from osteoblasts caused decreased bone mass, reduced bone formation, and a reduction in the formation and survival of osteoblasts.[47] While osteoblasts clearly express the PTH1R studies in cultured cells have suggested that PTHrP can also affect osteoblast biology via a nuclear pathway. Loss of the nuclear PTHrP pathway in PTHrP 1–84 knockin or PTHrP 1–66 knockin mice causes defects in osteoblast proliferation and function and is associated with low bone mass.[28] Thus, manipulation of the *Pthlh* gene *in vivo* affects the behavior of osteoblasts and alters bone mass. However, it is not entirely clear whether PTHrP affects osteoblasts in an autocrine or intracrine fashion, or whether it is produced by other cell types, such as osteocytes, and acts on osteoblasts in a paracrine fashion.

Recent studies have disrupted either the *Pthlh* or the *Pth1r* genes in osteocytes in mice. Loss of osteocyte PTHrP production reduced bone mass, which was associated with low osteoblast numbers and mineralizing surfaces. Moreover, *in vitro* experiments in cultured osteocytes showed that they secrete full-length PTHrP, which, in turn, regulated the expression of genes involved in matrix mineralization through autocrine and/or paracrine activation of the osteocyte PTH1R (Fig. 48.6).[14] Consistent with these findings, ablation of the PTH1R in osteocytes has profound effects on the activity of surface osteoclasts and osteoblasts *in vivo,* although these results are likely a composite of loss of PTHrP signaling from osteocytes and osteoblasts, as well PTH signaling from the circulation.[48,49] Collectively, these studies suggest that osteocytes can stimulate bone formation by secreting PTHrP, which may act in an autocrine/paracrine manner to promote bone formation. What triggers PTHrP expression in osteocytes in the first place and whether PTHrP can be transported through the lacunocanalicular network remain unknown. Further research is needed to answer these questions. Nevertheless, the ability of local PTHrP to stimulate bone formation is exploited in the clinic for the treatment of osteoporosis. Abaloparatide, a US Food and Drug Administration (FDA)-approved synthetic PTHrP analog, as well as teriparatide, an FDA-approved fragment of PTH, both exert potent anabolic activity that increases bone formation and bone mass and are used to treat patients with osteoporosis. Both of these drugs mimic pathways that likely reflect the local functions of PTHrP on bone formation (see “Skeletal Diseases”).

The Mammary Gland

Soon after its discovery, PTHrP was reported to be expressed in mammary tissue and to be secreted in milk.[50] Subsequently, PTHrP was shown to be critically important for the development and functioning of the breast throughout life. PTHrP serves distinct functions during different stages of mammary development; we will first discuss PTHrP’s functions during embryonic breast development and then its functions during lactation.

In mice, embryonic mammary development begins with the formation of 10 mammary buds, each of which consists of a bulb-shaped collection of epithelial cells surrounded by several layers of specialized fibroblasts known as the *mammary mesenchyme.*[51] Once formed, the mammary buds remain quiescent until embryonic day 16, when they give rise to a mammary sprout that develops into the rudimentary ductal tree. At the time of birth, the gland consists of a simple epithelial ductal tree consisting of 15 to 20 branched tubes within a fatty stroma.[51] This initial pattern persists until puberty, at which time the mature virgin gland is formed through a second round of branching morphogenesis that is regulated by circulating hormones.

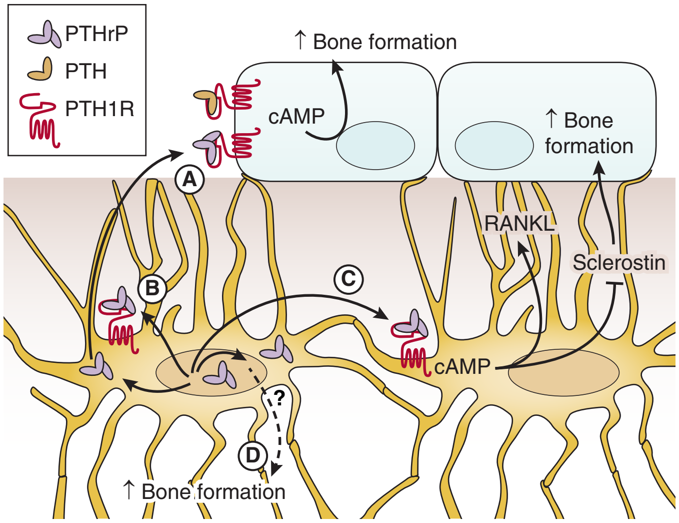


Fig. 48.6 Osteocytes can stimulate bone formation by secreting parathyroid hormone-related protein (PTHrP). (**A**), Full-length PTHrP produced and secreted by osteocytes promotes bone formation by activating the type 1 parathyroid hormone/parathyroid hormone-related protein receptor (PTHR1) from osteoblasts and their precursors, but also by acting in an autocrine (**B**) and paracrine (**C**) manner in osteocytes to regulate gene expression, including suppressing sclerostin production, which promotes bone formation by osteoblasts and stimulating RANKL production. (**D**), In addition, PTHrP promotes bone formation through its receptor-independent intracrine actions by mechanisms that are yet to be elucidated. PTH, Parathyroid hormone; cAMP, cyclic adenosine monophosphate. (Reproduced from Kronenberg HM. PTHrP and skeletal development. *Ann N Y Acad Sci*. 2006;1068:1–13.)

PTHrP is expressed in the nascent mammary gland, within epithelial cells coalescing into the developing mammary buds (Fig. 48.7).[52] The *Pthlh* gene continues to be expressed in all epithelial cells of the embryonic duct system,[51,52] but, after birth, *Pthlh* gene expression becomes restricted principally to myoepithelial cells.[52] In contrast to the *Pthlh* gene, the *Pth1r* gene is expressed within the mammary mesenchyme, but its expression is more widespread throughout the developing dermis.[51] It is not clear when the PTH1R is first expressed within the subepidermal mesenchyme. However, PTH1Rs are already present when the mammary bud begins to form, and the receptor continues to be expressed throughout embryonic development.[51] Similar patterns of epithelial PTHrP expression and mesenchymal PTH1R expression have been documented in the embryonic breast buds of human fetuses.[53,54]

Studies in genetically altered mouse models have demonstrated that PTHrP signaling is required for the differentiation of the specialized mammary mesenchyme and for the development of the embryonic duct system. As the mammary bud invaginates into the mesenchyme, PTHrP, produced by mammary epithelial cells, interacts with the PTH1R on the immature mesenchymal cells closest to the epithelial bud and triggers these cells to differentiate into mammary mesenchyme. PTHrP accomplishes this, in part, by facilitating local activation of Wnt and BMP signaling specifically within mesenchymal cells.[55] Although mammary buds form normally in PTHrP$^{-/-}$ and PTH1R$^{-/-}$ embryos, the lack of mammary mesenchyme leads to the subsequent failure of further mammary ductal development.[51] In the absence of PTHrP or the PTH1R the mammary buds never give rise to the neonatal mammary ducts, and these mice lack mammary epithelial structures. In addition, PTHrP$^{-/-}$ and PTH1R$^{-/-}$ mice lack nipples, because the mammary mesenchyme is also required to direct the epidermis to form these structures coincident with the outgrowth of the mammary ducts. Similarly, PTHrP is important for human fetal breast development. Blomstrand's chondrodysplasia is a fatal form of dwarfism caused by null mutations of the PTH1R gene,[34] and affected fetuses lack breast ducts and nipples.[54]

PTHrP also contributes to mammary physiology during lactation. *Pthlh* gene expression is upregulated in alveolar epithelial cells at the start of lactation under the control of both local and systemic factors.[56] Thiede and Rodan originally reported that PTHrP expression in rats depends on suckling and on serum prolactin concentrations.[57] However, prolactin most likely serves as a permissive factor, as additional studies have shown that the suckling response is a local one, and that PTHrP expression only rises in the milked gland.[56] Much of the PTHrP made during lactation ends up in milk, in which levels of PTHrP are up to 10,000-fold higher than in the circulation of normal individuals and 1000-fold higher than in patients suffering from humoral hypercalcemia of malignancy.[11] PTHrP concentrations in milk have generally been found to mirror RNA levels in the gland, increasing over the duration of lactation and rising acutely with suckling.[11] In addition, evidence shows that milk PTHrP levels may vary with the calcium content of milk.[58] In mice, the CaSR on the basolateral surface of mammary epithelial cells regulates PTHrP production, such that increased delivery of calcium to the mammary gland decreases PTHrP production and secretion into milk.[59]

Milk production requires a great deal of calcium, and the maternal skeleton serves as an important source of the calcium transported into milk.[58] Overall rates of bone turnover are elevated during lactation, but bone resorption outstrips bone formation so that bone mass declines rapidly. Nursing women lose between 5% and 10% of their bone mass over 6 months, while rodents lose up to one third of their skeletal mass over 21 days of lactation.[56,58] In mice, bone resorption is increased in great part due to the combination of decreased systemic estradiol concentrations and elevated circulating PTHrP levels.[58] Many studies have now documented elevated circulating PTHrP levels in lactating humans and rodents, and PTHrP levels correlate directly with biochemical markers of bone resorption and inversely with bone mass in lactating mice.[58,60] In addition, circulating PTHrP levels have been shown to correlate with bone density changes in lactating humans.[61] When the *Pthlh* gene was disrupted specifically in the lactating mammary gland in mice, circulating PTHrP levels declined, milk PTHrP became unmeasurable, rates of bone resorption decreased, and bone loss was reduced by 50%.[62] These data demonstrate that the lactating mammary gland secretes PTHrP both into milk and to the circulation, where it acts together with estrogen deficiency to stimulate bone resorption and cause bone loss. As noted previously, the production of PTHrP by the lactating breast is decreased by stimulation of the CaSR on mammary epithelial cells.[59] Therefore, as shown in Fig. 48.8, a classic endocrine negative feedback loop is created between bone and breast during lactation. If calcium delivery to the gland falls, mammary epithelial cells produce more PTHrP, which stimulates increased bone resorption. The delivery of more calcium to the mammary gland from the skeleton then stimulates the CaSR and decreases PTHrP production.

KEY POINTS

- Mammary gland–derived parathyroid hormone-related protein contributes to the regulation of calcium and bone metabolism during lactation.

The exact function of PTHrP in milk remains unclear. Genetic removal of PTHrP from the mammary gland does not alter milk calcium levels, and PTHrP-null mammary epithelial cells are able to transport calcium *in vitro* at the same rate as wild-type mammary

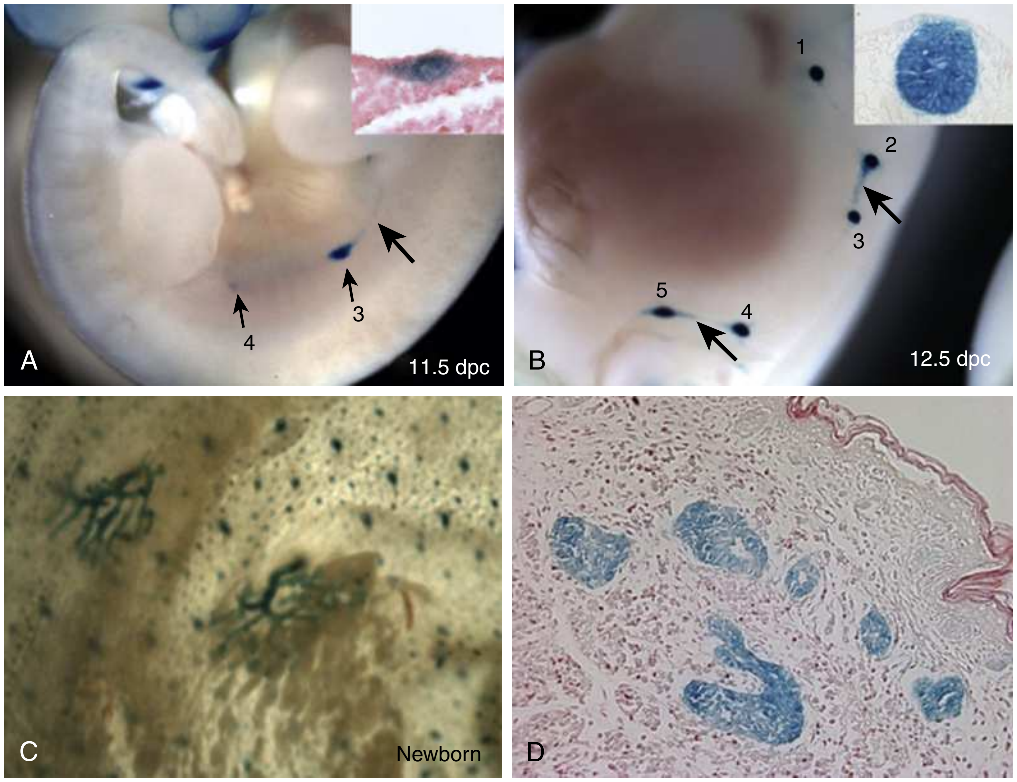


Fig. 48.7 Parathyroid hormone-related protein (PTHrP) is expressed in embryonic mammary glands and skin as indicated in PTHrP-lacZ knockin mice. A, At embryonic day (E) 11.5, PTHrP is expressed *(blue staining)* within the mammary line *(large arrow)* and in the forming mammary placodes *(arrows).* Inset shows section through mammary placode. **B,** By E12.5, PTHrP is expressed within the mammary buds *(inset)* and within cells migrating toward the forming buds *(double arrows).* **C,** At birth, PTHrP is expressed within the branching ductal tree of the neonatal mammary gland but also within the forming hair follicles in the skin *(punctate pattern).* **D,** Sections through the newborn skin show PTHrP expression within the epithelial cells of the mammary ducts. *dpc,* Days postconception. (Reproduced from Boras-Granic K, VanHouten J, Hiremath M, et al. Parathyroid hormone-related protein is not required for normal ductal or alveolar development in the post-natal mammary gland. *PLoS One.* 2011;6:e27278.)

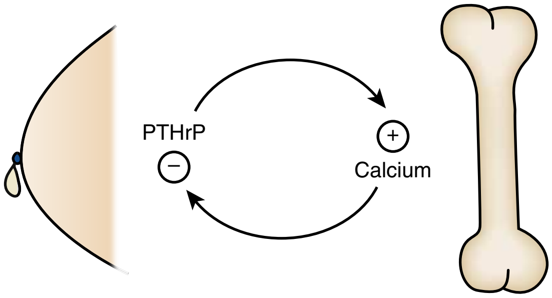


Fig. 48.8 The breast and the skeleton communicate during lactation in order to provide a steady supply of calcium for milk production. The lactating breast secretes parathyroid hormone-related protein (PTHrP) into the systemic circulation during lactation. PTHrP interacts with the type 1 parathyroid hormone/parathyroid hormone-related protein receptor in bone cells in order to increase the rate of bone resorption and liberate skeletal calcium stores. Mammary epithelial cells in the lactating breast express the calcium-sensing receptor and suppress PTHrP production in response to increased delivery of calcium, defining a classical endocrine negative feedback loop between breast and bone.

epithelial cells.[62] However, there is a dose-responsive, inverse relationship between milk PTHrP levels and neonatal ash calcium content in mice.[63] The mechanism(s) by which milk PTHrP modulates neonatal calcium and/or bone metabolism remains unexplored. Nevertheless, given that PTHrP levels vary inversely with maternal calcium availability and milk calcium content, these observations suggest that PTHrP may act as a "metabolic messenger" in milk to adjust neonatal bone and mineral metabolism to the mother's ability to supply calcium in milk.

The Skin

Normal human keratinocytes were the first nonmalignant cells demonstrated to produce PTHrP.[64] Subsequently, multiple studies have confirmed that cultured rodent and human keratinocytes secrete bioactive PTHrP (reviewed in[2]). The *Pthlh* gene is expressed most prominently within hair follicles, both during development and in the adult, where mRNA levels have been shown to vary with the hair cycle.[65] Low levels of *Pthlh* expression may also be found throughout the interfollicular epidermis. The PTH1R is expressed on dermal fibroblasts, and PTHrP has been shown to bind to skin fibroblasts and to elicit a variety of biological responses.[11] In addition to fibroblasts, studies using sensitive PCR-based detection methods have also reported expression of *Pth1r* mRNA in keratinocytes, although these studies used cultured cells.[66]

Several studies suggest that PTHrP participates in the regulation of the hair cycle. During development, disruption of the *Pthlh* or *Pth1r* genes does not seem to affect initial hair follicle morphogenesis or patterning in mice, except for the vicinity of the nipple.[11] However, in mature mice, systemic or topical administration of PTH1R antagonists perturbs the hair cycle by prematurely terminating telogen, prolonging anagen growth, and inhibiting catagen.[67] In a reciprocal fashion, mice overexpressing PTHrP in basal keratinocytes demonstrate a delayed emergence of dorsal hair and have shorter hairs, and their hair follicles enter catagen approximately 2 days early.[68] These findings were associated with a lower proliferation rate in the hair matrix and a less well-developed perifollicular vasculature during anagen. Together, these studies imply that PTHrP may regulate the anagen to catagen transition, acting to inhibit hair follicle growth by pushing growing hair follicles into the growth-arrested or catagen/telogen phase of the hair cycle. If this hypothesis were correct, one would expect PTHrP knockout mice to exhibit findings similar to PTH1R antagonist–treated mice. However, this does not appear to be the case. In mice that lack PTHrP in their skin, the hair cycle appears to be normal.[69] In fact, rather than a promotion of hair growth, these mice demonstrate a thinning of their coat over time. These conflicting results are difficult to reconcile, but suggest that, while PTHrP may contribute to the regulation of the hair cycle, it is unlikely to be necessary for this process to unfold normally.

PTHrP has also been implicated in the regulation of keratinocyte proliferation and/or differentiation. A careful comparison of the histology of PTHrP-null and PTHrP-overexpressing skin demonstrated reciprocal changes. In the absence of PTHrP, keratinocyte differentiation was accelerated, whereas in skin exposed to PTHrP overexpression, keratinocyte differentiation was delayed.[69] Therefore, in a physiologic context, PTHrP appears to slow the rate of keratinocyte differentiation and to preserve the proliferative basal compartment, effects that are analogous to those noted for chondrocyte differentiation in the growth plates of mice overexpressing PTHrP as compared with PTHrP- and PTH1R-null mice.[11]

Recent experiments have suggested that nuclear and/or C-terminal PTHrP may function to protect skin from aging. Knockin PTHrP 1–84 mice have a thinner epidermis and decreased subcutaneous fat and collagen bundles that suggest premature ageing. This is associated with decreased skin cell proliferation, increased apoptosis, and inappropriate senescence due to the upregulation of p27 and other CDK inhibitors. These findings suggest that the NLS and C-terminus of PTHrP are important to skin homoeostasis and raise the interesting possibility that decreased PTHrP expression could contribute to aging *in vivo*.[29]

Finally, it has been suggested that PTHrP contributes to innate immunity in the epidermis that protects against bacterial invasion. Muehleisen and colleagues reported that PTHrP signaling through the PTH1R in keratinocytes interacted synergistically with vitamin D to induce cathelicidin production and protect against experimental skin infection with group A *Streptococcus.*[70] The authors showed that PTH1R receptor expression was induced in keratinocytes by 1,25-vitamin D signaling. Furthermore, they showed that PTHrP expression was induced in keratinocytes by exposure to the bacterial product and TLR2 agonist lipoteichoic acid. These results suggest that local production of PTHrP in response to TLR activation serves to trigger the production of antimicrobial peptides, and that vitamin D may enhance the sensitivity of this regulatory loop by upregulating PTH1R expression. These are intriguing findings, and we will discuss other potential roles of PTHrP-PTH1R signaling in immunity later (see “The Immune System”).

Placenta

PTHLH mRNA and PTHrP protein have been detected in rat and human placenta in both cytotrophoblasts and syncytiotrophoblasts, as well as in mouse trophoblast giant cells.[11] The cuboidal cells of the amniotic membranes also express PTHrP, and these cells are likely the source of the PTHrP found in amniotic fluid.[71] PTHrP has effects on early trophoblast differentiation and is an important regulator of murine placental development, but the best-studied aspect of PTHrP in the placenta is its involvement in calcium transport.[72] Almost all of the calcium required for mineralization of the fetal skeleton is transferred from the mother to the fetus across the placenta during a relatively short interval in late pregnancy.[58] Both total and ionized calcium concentrations are higher in the fetal circulation than they are in the maternal circulation, demonstrating that calcium is actively transported across the placenta against an uphill concentration gradient. In mice homozygous for deletion of the *Pthrp* gene, fetal plasma calcium and the maternal-fetal calcium gradient are significantly reduced. When fetuses were injected *in utero* with fragments of PTHrP or PTH, calcium transport from the mother into the fetus was significantly restored only by treatment with a midmolecule PTHrP peptide that cannot stimulate the PTH1R.[58] The circulating PTHrP in fetal mice is thought to be derived from liver and placenta.[58] Other studies have demonstrated that PTH also contributes to the regulation of placental mineral transport and fetal mineral metabolism.[58] PTH appears to operate mainly on the fetal skeleton and kidneys to prevent fetal hypocalcemia, while PTHrP acts to stimulate transplacental calcium transport and raises the fetal calcium level above that in the maternal circulation, resulting in the typical suppression of PTH in the fetal circulation.

Endocrine Pancreas

Soon after its discovery, PTHrP expression was detected in all four cell types of pancreatic islets (α, β, δ, and pancreatic polypeptide cells).[73] More recent studies have verified that PTHrP is produced by rodent islet cells and human islet cells in primary tissue culture.[11] Islet cells also express the PTH1R and studies suggest that most of the observed biological actions of PTHrP in the endocrine pancreas are mediated by signaling from the PTH1R.[74]

PTHrP stimulates the proliferation of mouse and human pancreatic β cells during development and in the adult pancreas.[74] Vasavada and coworkers generated transgenic mice that overexpress PTHrP under the control of the rat insulin-II promoter (RIP).[75,76] RIP-PTHrP mice display striking degrees of islet hyperplasia and an increase in islet numbers compared with their control littermates (Fig. 48.9). This increased islet cell mass is the result of increased proliferation and decreased apoptosis, and it is associated with hyperinsulinemia and hypoglycemia, as well as resistance to the diabetogenic effects of the β cell toxin streptozotocin.[75,76] The effects of PTHrP on β cell proliferation have been further studied in rodent and human islet cells *in vitro.*[74] Proliferation is induced by N-terminal portions of PTHrP and depends on activation of the PTH1R, as well as downstream signaling via activation of PKC-ζ, cyclin E, and cdk2.[77,78] Furthermore, acute treatment with N-terminal PTHrP improved glucose-stimulated insulin secretion both in cultured β cells and in intact mice.[74,77,78] Increases in insulin expression and secretion in cultured rodent islet cells appear to be mediated by alterations in MAPK-associated phosphatase activity and inhibition of c-Jun N-terminal kinase.[79] Remarkably, daily subcutaneous injections of PTHrP 1–36 in normal adult mice or in a model of β cell deficiency increased the proliferation of β cells in islets *in vivo* and improved glucose tolerance, suggesting that PTHrP administration may prove useful in the maintenance and regeneration of islet cell mass for the treatment of diabetes.[78,80]

KEY POINTS

- Parathyroid hormone-related protein can regulate pancreatic β cell mass and may be useful in the treatment of diabetes.

Smooth Muscle and the Cardiovascular System

PTHrP is produced by smooth muscle cells in response to mechanical deformation and functions in a short feedback manner to relax the

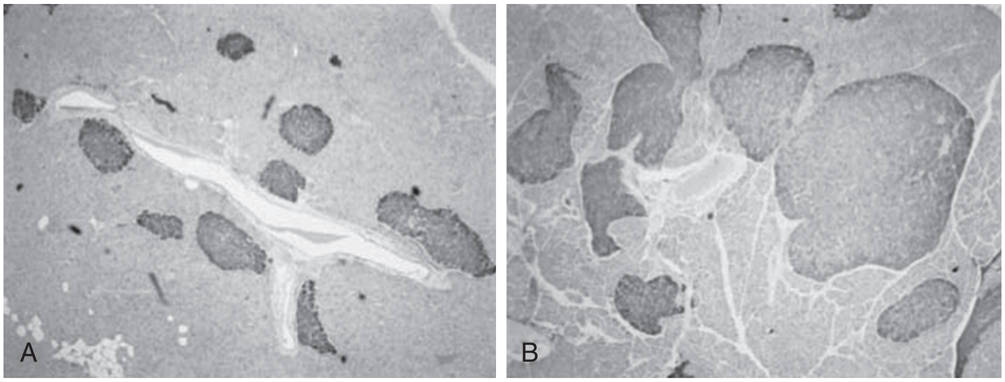


Fig. 48.9 **Overexpression of parathyroid hormone-related protein (PTHrP) in pancreatic islets causes islet hyperplasia.** Immunohistochemistry for insulin *(dark staining)* in normal mouse pancreas (**A**) and in RIP-PTHrP transgenic pancreas (**B**). Note that overexpression of PTHrP in β cells leads to an obvious increase in the number and size of pancreatic islets. (Reproduced from Porter SE, Sorenson RL, Dann P, et al. Progressive pancreatic islet hyperplasia in the islet-targeted, parathyroid hormone-related protein-overexpressing mouse. *Endocrinology*. 1998;139:3743–3751.)

stretched muscle.[1,2,11] This was initially studied in the uterus, where it was shown that PTHrP expression is induced by distension of the uterine horn by the developing fetus or by experimentally stretching the uterus.[81] In turn, treatment of rat, baboon, and human myometrium with PTHrP *ex vivo* results in reductions in isometric tension and inhibits spontaneous and electrically-induced contractions.[11] In baboon myometrium, PTHrP can counter the effects of oxytocin to induce uterine contraction.[82] These observations suggested that PTHrP may cause gradual relaxation of the uterus to accommodate the growth of the fetal-placental unit and/or to inhibit premature labor. Interestingly, the effects of PTHrP on myometrial contractivity wane with progression of pregnancy in rats, and the uterus becomes resistant to PTHrP after labor has been initiated.[11]

PTHrP was also shown to be expressed in both fetal and adult heart tissue from rats and humans, where it was suggested to have cardioprotective effects.[83] Infusions of PTHrP 1–34 improved the contractility of the rat and pig myocardium after stunning.[84] In primary cultures of endothelial cells and cardiomyocytes, PTHrP inhibited apoptosis.[85] Finally, the administration of PTHrP 87–139 enhanced cardiac angiogenesis and inhibited myocardial apoptosis and fibrosis after experimental myocardial infarction.[86] Although more research is needed to understand the role of PTHrP in normal and injured heart tissue, these results suggest that PTHrP may be cardioprotective and could be used therapeutically to preserve cardiac function in patients with myocardial ischemia or infarction.

PTHrP is induced by vasoconstrictive agents as well as stretch itself, and acts as a vasodilator in resistance blood vessels.[87] As a result, PTHrP may act to modulate local blood flow in a variety of tissues.[88] PTHrP has also been shown to regulate the proliferation of vascular smooth muscle cells. Secreted N-terminal PTHrP inhibits the proliferation of vascular smooth muscle cells by acting via the PTH1R on the cell surface. However, the midregion and C-terminal portions of PTHrP act in the nucleus to stimulate the proliferation of these cells *in vitro* and *in vivo* by inducing the expression of c-myc and skp2, which, in turn, reduce levels of the cell-cycle inhibitor p27, promoting progression through the G_1/S checkpoint.[26] This pathway appears to be active during development, because the rate of proliferation of smooth muscle cells in the aorta of PTHrP$^{-/-}$ embryos is reduced.[89] Furthermore, PTHrP expression is upregulated by vascular damage following balloon angioplasty and in atherosclerotic lesions in rodents and in humans. Several studies have suggested that PTHrP plays an important role in the response of these cells to injury and may contribute to the pathophysiologic development of a neointima following angioplasty.[90]

Teeth

Developing teeth become surrounded by alveolar bone but must maintain a bone-free cavity or crypt to allow for their morphogenesis. After teeth are formed, they must then erupt through the roof of the dental crypt in order to emerge into the oral cavity. The process of tooth eruption relies on geographically uncoupled bone turnover, in which osteoclasts form over the crown of the tooth in order to resorb the overlying bone, and osteoblasts at the base of the tooth propel it upward out of the crypt. PTHrP/PTH1R signaling is a key regulator of this process, because loss-of-function mutations of the PTH1R lead to a failure of tooth eruption in humans.[91,92] In mice, disruption of the *Pthlh* gene causes the same phenotype: teeth develop, but they do not erupt. PTHrP is normally produced by stellate reticulum cells and signals to dental follicle cells in order to drive the formation of osteoclasts above the crypt. In the absence of PTHrP, these osteoclasts do not appear, eruption fails to occur, and teeth become impacted.[93] In addition to being critical for tooth eruption, PTHrP/PTH1R signaling in dental mesenchymal progenitors has been shown to be essential for tooth root formation in mice. However, it is unclear whether this is also true in humans.[94]

The Nervous System

PTHrP and PTH1R are expressed in both the central nervous system (CNS) and peripheral nervous system. In the rat brain, PTHrP was found to be expressed mainly in the cerebral cortex, hippocampus, and cerebellum, whereas PTH1R is more widely distributed.[95] Interestingly, PTHrP is also found in human cerebrospinal fluid in levels that are at least 50 times those found in serum, and its abundance positively correlates with increased age.[96] At present, the exact role of PTHrP in the CNS remains unclear. Primary cultures of various neuronal and glial cell populations respond to exogenous PTHrP.[97-99] Furthermore, experiments in PTHrP knockout mice rescued from neonatal death by transgenic expression of PTHrP targeted to chondrocytes showed that lack of PTHrP was not associated with anatomical defects in the brain. However, primary cultures of cerebral cortical cells lacking the *Pthlh* gene showed increased susceptibility to excitotoxic agents, suggesting that PTHrP may have a neuroprotective function in the brain.[97] In contrast, PTHrP 1–84 knockin mice display abnormal brain development associated with decreased proliferation and increased apoptosis in neuronal cells that affects hippocampal synaptic transmission and plasticity, suggesting that the NLS and/or C-terminus of PTHrP may be important for normal brain development and function.[100]

The Immune System

The thymus is one of the primary lymphoid organs of the immune system in which T-cell development takes place. PTHrP and PTH1R have been shown to be expressed in thymic epithelial cells (TECs) from both human and mouse thymus. Treatment of primary cultures of human TECs with PTHrP stimulates the production of IL-6, which is important for the maturation and development of thymocytes. Moreover, PTHrP 1–84 knockin mice display abnormal development of T-cells, indicating that the NLS and C-terminus of PTHrP may be necessary for the normal formation of this cell type.[101-103]

Several *in vitro* and *in vivo* experiments have shown that PTHrP can trigger and mediate the inflammatory response in different tissues in disorders like pancreatitis, nephritis, atherosclerosis, and brain injury, among others.[104-108] Interestingly, PTHrP 107–111 was recently shown to have the opposite effect, improving arthritis by inhibiting immune activation, proinflammatory cytokines, osteoblast senescence, and osteoclastogenesis, indicating that it could modulate aspects of the pathophysiology of osteoarthritis.[109,110]

PARATHYROID HORMONE-RELATED PROTEIN IN DISEASE

Given the diverse biological effects of PTHrP, one might expect that it is involved in the pathogenesis of diseases. The most established role for PTHrP in disease states is to cause hypercalcemia in patients with cancer. This occurs because PTHrP and PTH share the same receptor, but normally PTHrP does not circulate. However, in some tumors, PTHrP is made in sufficient quantities to reach the circulation and to stimulate PTH1R receptors in the bone and kidney normally meant to mediate the functions of PTH, thus mimicking many aspects of primary hyperparathyroidism. The topic of hypercalcemia in malignancy is covered in detail in Chapter 55. What follows here is a short discussion of other conditions in which PTHrP may play a pathophysiologic role or conditions in which it may have therapeutic applications.

Cancer

Many studies have suggested that PTHrP modulates the proliferation, differentiation, and/or survival of cancer cell lines from a variety of different tumor types *in vitro*.[2,11,111] However, there are many fewer studies that address the possibility that PTHrP contributes to tumor cell growth in animal models *in vivo* or in human cancers. These data do exist for breast cancer, although studies in cultured breast cancer cell lines, animal models, and clinical studies have yielded conflicting results.[112] Some case series suggest that PTHrP expression predicts a more aggressive clinical course.[112] However, other studies have suggested just the opposite. First, Henderson and colleagues examined 526 consecutive cases of breast cancer and found that PTHrP expression was an independent predictor of a more benign clinical course.[113] More recently, in two additional patient cohorts totaling nearly 800 patients, the authors found that PTHrP levels were downregulated in malignant epithelium, and that lower levels of nuclear-localized PTHrP in cancer cells correlated with unfavorable clinical outcomes.[114] In addition to these case studies, several large genome-wide association studies (GWAS) have implicated the *PTHLH* gene as a breast cancer susceptibility locus.[115] While GWAS do not provide mechanistic insight, they do underscore the fact that the effects of PTHrP in breast cancer are likely to be clinically important.

KEY POINTS

- Parathyroid hormone-related protein contributes to the development and/or progression of breast cancer.

Conflicting results have also been reported in studies employing transgenic models of breast cancer. Fleming and colleagues demonstrated that mammary-specific disruption of the *Pthlh* gene increased the incidence of tumors in MMTV-Neu mice.[116] In contrast, Li and colleagues used an identical genetic approach and found that disruption of the *Pthlh* gene dramatically prolonged tumor latency, slowed tumor growth, and reduced metastases caused by transgenic overexpression of the polyoma middle T antigen (MMTV-PyMT mice).[117] Moreover, using the same MMTV-PyMT mice breast cancer model, Kim et al. demonstrated that CaSR signaling stimulates the growth of mammary tumors, at least in part, by upregulating PTHrP production *in vitro* and *in vivo*. These findings were further supported by a positive correlation between CaSR and PTHrP mRNA expression in human samples.[118] The opposing results from these different transgenic mouse studies suggest that the molecular context of transformation may be critical for determining PTHrP's actions, and this may help explain the variability in the clinical observations.

Another function of PTHrP in breast cancer is its contribution to the osteolysis that is required for the formation of bone metastases.[119] In addition, although prostate cancer is often associated with osteoblastic bone metastases, studies in xenograft models of human prostate cancer cell lines have suggested that PTHrP also promotes the survival and growth of prostate cancer in the skeleton.[120] In breast cancers, experiments have suggested that, in response to the bone microenvironment, cells metastatic to the skeleton produce more PTHrP than cells in the primary tumor.[121] TGF-β, released at sites of bone resorption, has been shown to upregulate PTHrP production through a signaling pathway involving the transcription factor Gli2.[121] PTHrP, in turn, increases the production of RANKL and decreases the production of OPG, increasing osteoclast numbers and activity.[121] This sets up a feedforward loop between PTHrP production and bone resorption, contributing to a vicious cycle of osteolysis. However, there is conflicting clinical evidence as to whether PTHrP production by a primary breast tumor is predictive of bone metastases in patients.[121] One large and carefully controlled study suggested that PTHrP production by the primary tumor is actually a negative predictor, not a positive predictor, of skeletal metastases.[113] In contrast, a study of 125 patients with breast cancer and long-term clinical follow-up suggested that PTHrP staining in the primary tumor was predictive of bone metastases.[122] In agreement with this result, PTHrP expression examined in 497 cases of early breast cancer was found to be associated with bone metastasis and, in the estrogen receptor-negative subgroup, those who stained positive for PTHrP had a poor prognosis.[123] Furthermore, Washam and colleagues showed that the presence of PTHrP 12–48 in the circulation of human breast cancer patients was predictive of bone metastases.[124] This same group showed that PTHrP 12–48 could affect the differentiation of hematopoietic cells as well as the activity of osteoclasts, providing evidence that this peptide might somehow modulate tumor cell–bone cell interactions.[125]

Tumor-derived PTHrP can also alter energy metabolism and lead to a debilitating metabolic disorder termed cancer anorexia cachexia syndrome (CACS). CACS is caused by the tumor release of soluble factors leading to pronounced loss of weight, appetite, skeletal muscle, and fat mass. Several studies have implicated PTHrP in the pathogenesis of CACS. In a longitudinal observational cohort of 624 cancer patients, circulating PTHrP was found to be an independent predictor of weight loss.[126] Studies in mice carrying xenograft tumors producing

PTHrP suggested that it reduced food intake and promoted weight loss, which could be reversed by the administration of anti-PTHrP antiserum, resulting in prolonged survival.[121] More recently, Kir et al. suggested that PTHrP, produced by lung cancer cells, caused weight loss, fat wasting, and muscle atrophy, and suggested that this was due to adipocyte "browning" and a hypermetabolic state but not changes in appetite.[127] Again, treatment with anti-PTHrP antibodies prevented fat and muscle wasting in this model, and mice lacking PTH1R in adipocytes were resistant to cachexia driven by the same lung cancer xenograft tumor.[128] These studies implicate PTHrP as a contributor to CACS, and the fact that neutralizing antibodies can prevent the development of CACS in these mouse models suggests that targeting PTHrP-PTH1R signaling may be useful to treat cachexia.

Skeletal Disease

Given the extensive involvement of PTHrP in normal chondrocyte and bone biology, it is not surprising that alterations in PTHrP function are associated with skeletal diseases. The most obvious case is HHM, where circulating PTHrP interacts with PTH1R receptors in the skeleton to induce widespread bone resorption and produce hypercalcemia.[1] Loss-of-function mutations in the *PTHLH* gene cause Blomstrand's chondrodysplasia, which is associated with bone abnormalities that mimic the PTHrP knockout mouse and cause fetal demise.[34] Gain-of-function mutations in the *PTHLH* gene cause a rare disease termed Jansen's metaphyseal chondrodysplasia (JMC). This is a form of short-limbed dwarfism that results from inhibition of chondrocyte differentiation due to overactive PTH1R signaling.[34] Inverse agonist ligands that are able to suppress the basal cAMP signaling activity of the PTH1R are a promising therapy for JMC, a disorder that is currently without effective treatment.[129] As described previously, loss-of-function mutations in the *PTHLH* gene have been shown to cause brachydactyly type E, a syndrome including short stature, shortened metacarpals and metatarsals, and learning disabilities.[38] The functions of PTHrP in the maintenance of articular cartilage and modeling at ligamentous and tendinous insertion sites suggest that it may contribute to the pathophysiology of osteoarthritis or enthesopathies. Finally, the anabolic actions of PTHrP on bone have been the focus of multiple studies that evaluated its potential as a treatment for osteoporosis.[130] In 2017, the FDA approved an injectable synthetic analog of PTHrP for the treatment of postmenopausal women with osteoporosis at high risk of fracture. Abaloparatide is an N-terminal analog of PTHrP 1–34 that shares 76% homology with PTHrP 1–34 and 41% homology with PTH 1–34. The first 21 amino acids of abaloparatide are identical to PTHrP, but multiple substitutions have been made between amino acids 22 to 34 that conferred more stability. In pivotal trials, abaloparatide has been shown to be effective in reducing both vertebral and nonvertebral fractures and appears to provide a more rapid increase in bone mineral density than teriparatide (PTH 1–34).[131]

KEY POINTS

- A synthetic parathyroid hormone-related protein analog was approved for the treatment of postmenopausal women with osteoporosis.

Skin Disease

It has been suggested that defects in the skin and skin appendages in the rescued PTHrP-null mice are reminiscent of a series of disorders collectively known as the *ectodermal dysplasias*,[132] but the *PTHLH* gene has not been formally linked to any of these diseases. It has also been noted that psoriatic skin may downregulate PTHrP expression, and a small trial of topically-administered PTH suggested that stimulation of the PTH1R may improve the histologic abnormalities in psoriatic plaques.[133] However, these potentially exciting results have not yet been confirmed in larger trials. Preliminary studies also suggest that topical application of PTH1R antagonists may be useful for stimulating hair growth.[67] These original observations date back many years, but no further evidence has emerged to demonstrate whether manipulation of PTHrP-PTH1R signaling is an effective treatment for alopecia. Furthermore, a more recent study suggested that, in the setting of chemotherapy-induced alopecia, PTH1R agonists, but not an antagonist, stimulated hair regrowth.[134] Therefore, despite some intriguing preliminary findings, it is not entirely clear yet whether PTHrP is involved in skin pathology.

Diabetes

There is no evidence that PTHrP is involved in the pathogenesis of diabetes. However, because PTHrP induces proliferation and improves glucose-stimulated insulin secretion in human β cells, it is possible that it may be useful for the treatment of diabetes, especially in the setting of islet cell transplantation.[74] In support of these ideas, daily subcutaneous injections of PTHrP 1–36 have been shown to increase the proliferation of β cells in mouse islets *in vivo* and improve glucose tolerance in intact mice, as well as in a mouse model of β cell deficiency.[78,80] Interestingly, the same therapeutic approach was shown to be effective in restoring bone mass and strength in diabetic mice.[135] Patients with diabetes mellitus have alterations in bone structure and strength that compromise bone quality, increasing the risk for osteoporotic fractures.[136] Therefore, these results suggest that PTHrP treatment could be a useful therapeutic strategy to induce β cell proliferation and prevent bone fractures in diabetic patients.

CONCLUSION

The identification of PTHrP as the primary cause of HHM was the culmination of efforts from several investigators and the combination of basic and clinical research studies that elucidated the pathophysiology of this syndrome. Since then, we have learned much about the normal functions of PTHrP in well-studied sites such as bone, cartilage, pancreas, and breast. However, PTHrP is expressed in virtually all tissues and in many developmental stages. Yet, we have very little knowledge of its functions in these other sites. While it is well established that PTHrP can be processed into a series of different peptides, the biological significance of these different forms of PTHrP is poorly understood. Likewise, it is established that, in addition to its actions through the PTHR1, PTHrP can translocate to the nucleus. While preliminary data suggest that nuclear PTHrP is necessary for life in mice, we have very limited knowledge of how nuclear PTHrP may regulate either physiology or disease. Therefore, although much progress has been made in the 34 or so years since the identification of PTHrP, we actually have only a rudimentary understanding of the biology of this remarkable and ancient protein.

There is cumulative evidence that PTHrP contributes to disease processes in a variety of organs. It has been shown to affect the development and/or behavior of breast and other cancers. However, the reported functions of PTHrP in cancer cells have often been contradictory. Nevertheless, it does appear to have an important role in the progression of osteolytic bone metastases from multiple types of cancer. Mutations in the *PTHLH* gene have been shown to cause rare bone diseases, and alterations in PTHrP expression have been noted in diabetes and other disorders. Given its widespread expression and functions in multiple organs, it is likely that dysregulation of PTHrP expression or signaling will be found to contribute to the pathophysiology of other disease states.

Finally, targeting PTHrP signaling is likely to have therapeutic possibilities. In fact, recent advances in our understanding of the effects of PTHrP in bone biology have now culminated in the first synthetic PTHrP analog (abaloparatide) approved for the treatment of postmenopausal osteoporosis, completing the translational research cycle from bedside to bench and back again. Clearly, there is yet much to learn about PTHrP and its functions in biology and disease, with the future promise of new therapeutic uses for PTHrP or its analogs.

Acknowledgment

This work was supported by the National Institutes of Health (AR073146, HD100751, HD100468) and the Department of Defense Breast Cancer Research Program (BC151665).

REFERENCES

1. Strewler GJ. The physiology of parathyroid hormone-related protein. *N Engl J Med*. 2000;342:177–185.
2. Philbrick WM, et al. Defining the roles of parathyroid hormone-related protein in normal physiology. *Physiol Rev*. 1996;76:127–173.
3. Pinheiro PLC, et al. Gene structure, transcripts and calciotropic effects of the PTH family of peptides in Xenopus and chicken. *BMC Evol Biol*. 2012;10(1):373.
4. Bouizar Z, Spyratos F, De Vernejoul MC. The parathyroid hormone-related protein (PTHrP) gene: use of downstream TATA promotor and PTHrP 1-139 coding pathways in primary breast cancers vary with the occurrence of bone metastasis. *J Bone Miner Res*. 1999;14:406–414.
5. Tost J, et al. Methylation of specific CpG sites in the P2 promoter of parathyroid hormone-related protein determines the invasive potential of breast cancer cell lines. *Epigenetics*. 2011;6:1035–1046.
6. Cao X, et al. miR-195 contributes to human osteoarthritis via targeting PTHrP. *J Bone Miner Metab*. 2019;37:711–721.
7. Bai M, Yin H, Zhao J, et al. miR-182-5p overexpression inhibits chondrogenesis by down-regulating PTHLH. *Cell Biol Int*. 2019;43:222–232.
8. Stewart AF, et al. Biochemical evaluation of patients with cancer-associated hypercalcemia: evidence for humoral and nonhumoral groups. *N Engl J Med*. 1980;303:1377–1383.
9. Gensure RC, Gardella TJ, Jüppner H. Parathyroid hormone and parathyroid hormone-related peptide, and their receptors. *Biochem Biophys Res Commun*. 2005;328:666–678.
10. Lee K, Deeds JD, Segre GV. Expression of parathyroid hormone-related peptide and its receptor messenger ribonucleic acids during fetal development of rats. *Endocrinology*. 1995;136:453–463.
11. Wysolmerski JJ. Parathyroid hormone-related protein: an update. *J Clin Endocrinol Metab*. 2012;97:2947–2956.
12. Frieling JS, et al. Matrix metalloproteinase processing of PTHrP yields a selective regulator of osteogenesis, PTHrP1-17. *Oncogene*. 2017;36:4498–4507.
13. Wu TL, Soifer NE, Burtis WJ, et al. Glycosylation of parathyroid hormone-related peptide secreted by human epidermal keratinocytes. *J Clin Endocrinol Metab*. 1991;73:1002–1007.
14. Ansari N, et al. Autocrine and paracrine regulation of the murine skeleton by osteocyte-derived parathyroid hormone-related protein. *J Bone Miner Res*. 2018;33:137–153.
15. Martin TJ. Parathyroid hormone-related protein, its regulation of cartilage and bone development, and role in treating bone diseases. *Physiol Rev*. 2016;96:831–871.
16. Sutkeviciute I, Clark LJ, White AD, et al. PTH/PTHrP receptor signaling, allostery, and structures. *Trends Endocrinol Metab*. 2019;30:860–874.
17. Ehrenmann J, et al. High-resolution crystal structure of parathyroid hormone 1 receptor in complex with a peptide agonist. *Nat Struct Mol Biol*. 2018;25:1086–1092.
18. Pioszak AA, Parker NR, Gardella TJ, et al. Structural basis for parathyroid hormone-related protein binding to the parathyroid hormone receptor and design of conformation-selective peptides. *J Biol Chem*. 2009;284:28382–28391.
19. Ho PWM, Chan AS, Pavlos NJ, et al. Brief exposure to full length parathyroid hormone-related protein (PTHrP) causes persistent generation of cyclic AMP through an endocytosis-dependent mechanism. *Biochem Pharmacol*. 2019;169:113627.
20. White AD, et al. Ca 2+ allostery in PTH-receptor signaling. *Proc Natl Acad Sci U S A*. 2019;116:3294–3299.
21. Kim BH, et al. Extracellular nucleotides enhance agonist potency at the parathyroid hormone 1 receptor. *Cell Signal*. 2018;46:103–112.
22. Nilsson I, et al. The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence. *J Mol Biol*. 2015;427:1191–1201.
23. Amaya Y, Nakai T, Miura S. Evolutionary well-conserved region in the signal peptide of parathyroid hormone-related protein is critical for its dual localization through the regulation of ER translocation. *J Biochem*. 2016;159:393–406.
24. García-Martín A, et al. Functional roles of the nuclear localization signal of parathyroid hormone-related protein (PTHrP) in osteoblastic cells. *Mol Endocrinol*. 2014;28:925–934.
25. Shen X, Qian L, Falzon M. PTH-related protein enhances MCF-7 breast cancer cell adhesion, migration, and invasion via an intracrine pathway. *Exp Cell Res*. 2004;294:420–433.
26. Sicari BM, et al. c-myc and skp2 coordinate p27 degradation, vascular smooth muscle proliferation, and neointima formation induced by the parathyroid hormone-related protein. *Endocrinology*. 2012;153:861–872.
27. Miao D, et al. Severe growth retardation and early lethality in mice lacking the nuclear localization sequence and C-terminus of PTH-related protein. *Proc Natl Acad Sci U S A*. 2008;105:20309–20314.
28. Toribio RE, et al. The midregion, nuclear localization sequence, and C terminus of PTHrP regulate skeletal development, hematopoiesis, and survival in mice. *FASEB J*. 2010;24:1947–1957.
29. Jiang M, et al. Deficiency of the parathyroid hormone-related peptide nuclear localization and carboxyl terminal sequences leads to premature skin ageing partially mediated by the upregulation of p27. *Exp Dermatol*. 2015;24:847–852.
30. Boras-Granic K, Dann P, VanHouten J, et al. Deletion of the nuclear localization sequences and C-terminus of PTHrP impairs embryonic mammary development but also inhibits PTHrP production. *PLoS One*. 2014;9(5):e90418.
31. Ohba S. Hedgehog signaling in skeletal development: roles of indian hedgehog and the mode of its action. *Int J Mol Sci*. 2020;21:1–17.
32. Kronenberg HM. PTHrP and skeletal development. *Ann N Y Acad Sci*. 2006;1068:1–13.
33. Serowoky MA, Arata CE, Crump JG, et al. Skeletal stem cells: Insights into maintaining and regenerating the skeleton. *Development*. 2020;147(5):dev179325.
34. Nissenson RA. Editorial: parathyroid hormone (PTH)/PTHrP receptor mutations in human chondrodysplasia. *Endocrinology*. 1998;139:4753–4755.
35. Nishimori S, et al. Salt-inducible kinases dictate parathyroid hormone 1 receptor action in bone development and remodeling. *J Clin Invest*. 2019;129:5187–5203.
36. Nishimori S, et al. PTHrP targets HDAC4 and HDAC5 to repress chondrocyte hypertrophy. *JCI Insight*. 2019;4(5):e97903.
37. Seriwatanachai D, et al. Deletion of Zfp521 rescues the growth plate phenotype in a mouse model of Jansen metaphyseal chondrodysplasia. *FASEB J*. 2011;25:3057–3067.
38. Reyes M, Bravenboer B, Jüppner H. A heterozygous splice–site mutation in PTHLH causes autosomal dominant shortening of metacarpals and metatarsals. *J Bone Miner Res*. 2019;34:482–489.
39. Hirai T, Chagin AS, Kobayashi T, et al. Parathyroid hormone/parathyroid hormone-related protein receptor signaling is required for maintenance of the growth plate in postnatal life. *Proc Natl Acad Sci U S A*. 2011;108:191–196.
40. Wang Y, et al. IGF-1R signaling in chondrocytes modulates growth plate development by interacting with the PTHrP/Ihh pathway. *J Bone Miner Res*. 2011;26:1437–1446.
41. Chen X, Macica CM, Nasiri A, et al. Regulation of articular chondrocyte proliferation and differentiation by Indian hedgehog and parathyroid hormone-related protein in mice. *Arthritis Rheum*. 2008;58:3788–3797.

42. Chen X, Macica C, Nasiri A, et al. Mechanical regulation of PTHrP expression in entheses. *Bone*. 2007;41:752–759.
43. Wang M, Vanhouten JN, Nasiri AR, et al. Periosteal PTHrP regulates cortical bone modeling during linear growth in mice. *J Anat*. 2014;225:71–82.
44. Datta NS, Abou-Samra AB. PTH and PTHrP signaling in osteoblasts. *Cell Signal*. 2009;21:1245–1254.
45. Goltzman D. Studies on the mechanisms of the skeletal anabolic action of endogenous and exogenous parathyroid hormone. *Arch Biochem Biophys*. 2008;473:218–224.
46. Amizuka N, et al. Haploinsufficiency of parathyroid hormone-related peptide (PTHrP) results in abnormal postnatal bone development. *Dev Biol*. 1996;175:166–176.
47. Miao D, et al. Osteoblast-derived PTHrP is a potent endogenous bone anabolic agent that modifies the therapeutic efficacy of administered PTH 1-34. *J Clin Invest*. 2005;115:2402–2411.
48. Saini V, et al. Parathyroid hormone (PTH)/PTH-related peptide type 1 receptor (PPR) signaling in osteocytes regulates anabolic and catabolic skeletal responses to PTH. *J Biol Chem*. 2013;288:20122–20134.
49. Rhee Y, et al. Resorption controls bone anabolism driven by parathyroid hormone (PTH) receptor signaling in osteocytes. *J Biol Chem*. 2013;288:29809–29820.
50. Budayr AA, et al. High levels of a parathyroid hormone-like protein in milk. *Proc Natl Acad Sci U S A*. 1989;86:7183–7185.
51. Hiremath M, Wysolmerski J. Parathyroid hormone-related protein specifies the mammary mesenchyme and regulates embryonic mammary development. *J Mammary Gland Biol Neoplasia*. 2013;18:171–177.
52. Boras-Granic K, VanHouten J, Hiremath M, et al. Parathyroid hormone-related protein is not required for normal ductal or alveolar development in the post-natal mammary gland. *PLoS One*. 2011;6:e27278.
53. Cormier S, Delezoide AL, Silve C. Expression patterns of parathyroid hormone-related peptide (PTHrP) and parathyroid hormone receptor type 1 (PTHR1) during human development are suggestive of roles specific for each gene that are not mediated through the PTHrP/PTHR1 paracrine signaling pathway. *Gene Expr Patterns*. 2003;3:59–63.
54. Wysolmerski JJ, et al. Absence of functional type 1 parathyroid hormone (PTH)/PTH-related protein receptors in humans is associated with abnormal breast development and tooth impaction. *J Clin Endocrinol Metab*. 2001;86:1788–1794.
55. Hiremath M, et al. Parathyroid hormone-related protein activates Wnt signaling to specify the embryonic mammary mesenchyme. *Dev*. 2012;139:4239–4249.
56. Wysolmerski JJ. Interactions between breast, bone, and brain regulate mineral and skeletal metabolism during lactation. *Ann N Y Acad Sci*. 2010;1192:161–169.
57. Thiede MA. The mRNA encoding a parathyroid hormone-like peptide is producedin mammary tissue in response to elevations in serum prolactin. *Mol Endocrinol*. 1989;3:1443–1447.
58. Kovacs CS. Maternal mineral and bone metabolism during pregnancy, lactation, and post-weaning recovery. *Physiol Rev*. 2016;96:449–547.
59. Kim W, Wysolmerski JJ. Calcium-sensing receptor in breast physiology and cancer. *Front Physiol*. 2016;7:440.
60. VanHouten JN, Wysolmerski JJ. Low estrogen and high parathyroid hormone-related peptide levels contribute to accelerated bone resorption and bone loss in lactating mice. *Endocrinology*. 2003;144:5521–5529.
61. Sowers MF, et al. Elevated parathyroid hormone-related peptide associated with lactation and bone density loss. *JAMA, J Am Med Assoc*. 1996;276:549.
62. VanHouten JN, et al. Mammary-specific deletion of parathyroid hormone-related protein preserves bone mass during lactation. *J Clin Invest*. 2003;112:1429–1436.
63. Mamillapalli R, et al. Mammary-specific ablation of the calcium-sensing receptor during lactation alters maternal calcium metabolism, milk calcium transport, and neonatal calcium accrual. *Endocrinology*. 2013;154:3031–3042.
64. Merendino JJ, Insogna KL, Milstone LM, et al. A parathyroid hormone-like protein from cultured human keratinocytes. *Science*. 1986;231:388–390.
65. Chen X, et al. Initial characterization of PTH-related protein gene-driven lacZ expression in the mouse. *J Bone Miner Res*. 2006;21:113–123.
66. Errazahi A, Bouizar Z, Lieberherr M, et al. Functional type I PTH/PTHrP receptor in freshly isolated newborn rat keratinocytes: identification by RT-PCR and immunohistochemistry. *J Bone Miner Res*. 2003;18:737–750.
67. Safer JD, Ray S, Holick MF. A topical parathyroid hormone/parathyroid hormone-related peptide receptor antagonist stimulates hair growth in mice. *Endocrinology*. 2007;148:1167–1170.
68. Diamond AG, et al. Parathyroid hormone hormone-related protein and the PTH receptor regulate angiogenesis of the skin. *J Invest Dermatol*. 2006;126:2127–2134.
69. Foley J, et al. PTHrP regulates epidermal differentiation in adult mice. *J Invest Dermatol*. 1998;111:1122–1128.
70. Muehleisen B, et al. PTH/PTHrP and vitamin D control antimicrobial peptide expression and susceptibility to bacterial skin infection. *Sci Transl Med*. 2012;4(135):135ra66.
71. Dvir R, et al. Amniotic fluid and plasma levels of parathyroid hormone-related protein and hormonal modulation of its secretion by amniotic fluid cells. *Eur J Endocrinol*. 1995;133:277–282.
72. Duval C, Dilworth MR, Tunster SJ, et al. PTHrP is essential for normal morphogenetic and functional development of the murine placenta. *Dev Biol*. 2017;430:325–336.
73. Asa SL, Henderson J, Goltzman D, et al. Parathyroid hormone-like peptide in normal and neoplastic human endocrine tissues. *J Clin Endocrinol Metab*. 1990;71:1112–1118.
74. Kondegowda NG, et al. Parathyroid hormone-related protein enhances human β-cell proliferation and function with associated induction of cyclin-dependent kinase 2 and cyclin E expression. *Diabetes*. 2010;59:3131–3138.
75. Vasavada RC, et al. Overexpression of parathyroid hormone-related protein in the pancreatic islets of transgenic mice causes islet hyperplasia, hyperinsulinemia, and hypoglycemia. *J Biol Chem*. 1996;271:1200–1208.
76. Porter SE, et al. Progressive pancreatic islet hyperplasia in the islet-targeted, parathyroid hormone-related protein-overexpressing mouse. *Endocrinology*. 1998;139:3743–3751.
77. Vasavada RC, et al. Protein kinase C-ζ activation markedly enhances β-cell proliferation: an essential role in growth factor-mediated β-cell mitogenesis. *Diabetes*. 2007;56:2732–2743.
78. Williams K, et al. Systemic and acute administration of parathyroid hormone-related peptide(1-36) stimulates endogenous beta cell proliferation while preserving function in adult mice. *Diabetologia*. 2011;54:2867–2877.
79. Zhang B, et al. Parathyroid hormone-related protein induces insulin expression through activation of MAP kinase-specific phosphatase-1 that dephosphorylates c-Jun NH2-terminal kinase in pancreatic β-cells. *Diabetes*. 2003;52:2720–2730.
80. Mozar A, et al. Parathyroid hormone-related peptide (1-36) enhances beta cell regeneration and increases beta cell mass in a mouse model of partial pancreatectomy. *PLoS One*. 2016;11:e0158414.
81. Williams ED, Leaver DD, Danks JA, et al. Effect of parathyroid hormone-related protein (PTHrP) on the contractility of the myometrium and localization of PTHrP in the uterus of pregnant rats. *J Reprod Fertil*. 1994;102:209–214.
82. Pitera AE, Smith GCS, Wentworth RA, et al. Parathyroid hormone-related peptide (1 to 34) inhibits in vitro oxytocin-stimulated activity of pregnant baboon myometrium. *Am J Obstet Gynecol*. 1998;179:492–496.
83. Grohé C, et al. Sex-specific differences in ventricular expression and function of parathyroid hormone-related peptide. *Cardiovasc Res*. 2004;61:307–316.
84. Jansen J, et al. Parathyroid hormone-related peptide improves contractile function of stunned myocardium in rats and pigs. *Am J Physiol Circ Physiol*. 2003;284:H49–H55.
85. Datta NS, et al. Parathyroid hormone-related peptide protects cardiomyocytes from oxidative stress-induced cell death: first evidence of a novel endocrine-cardiovascular interaction. *Biochem Biophys Res Commun*. 2015;468:202–207.

86. Zhang H, et al. The effects of parathyroid hormone-related peptide on cardiac angiogenesis, apoptosis, and function in mice with myocardial infarction. *J Cell Biochem*. 2019;120:14745–14755.
87. Benson T, et al. Mechanisms of vasodilation to PTH 1–84, PTH 1–34, and PTHrP 1–34 in rat bone resistance arteries. *Osteoporos Int*. 2016;27:1817–1826.
88. Massfelder T, Helwig JJ. Editorial: parathyroid hormone-related protein in cardiovascular development and blood pressure regulation. *Endocrinology*. 1999;140:1507–1510.
89. Massfelder T, et al. Opposing mitogenic and anti-mitogenic actions of parathyroid hormone-related protein in vascular smooth muscle cells: a critical role for nuclear targeting. *Proc Natl Acad Sci U S A*. 1997;94:13630–13635.
90. Ishikawa M, et al. Expression of parathyroid hormone-related protein in human and experimental atherosclerotic lesions: functional role in arterial intimal thickening. *Atherosclerosis*. 2000;152:97–105.
91. Decker E, et al. PTHR1 loss-of-function mutations in Familial, Nonsyndromic primary failure of tooth eruption. *Am J Hum Genet*. 2008;83:781–786.
92. Yamaguchi T, et al. Exome resequencing combined with linkage analysis identifies novel PTH1R variants in primary failure of tooth eruption in Japanese. *J Bone Miner Res*. 2011;26:1655–1661.
93. Zhang J, et al. Parathyroid hormone–related peptide (1–34) promotes tooth eruption and inhibits osteogenesis of dental follicle cells during tooth development. *J Cell Physiol*. 2019;234:11900–11911.
94. Ono W, Sakagami N, Nishimori S, et al. Parathyroid hormone receptor signalling in osterix-expressing mesenchymal progenitors is essential for tooth root formation. *Nat Commun*. 2016;7:1–16.
95. Weaver DR, Deeds JD, Lee K, et al. Localization of parathyroid hormone-related peptide (PTHrP) and PTH/PTHrP receptor mRNAs in rat brain. *Mol Brain Res*. 1995;28:296–310.
96. Kushnir MM, Peterson LK, Strathmann FG. Parathyroid hormone related protein concentration in human serum and CSF correlates with age. *Clin Biochem*. 2018;52:56–60.
97. Chatterjee O, et al. Endogenous parathyroid hormone-related protein functions as a neuroprotective agent. *Brain Res*. 2002;930:58–66.
98. Hashimoto H, Aino H, Ogawa N, et al. Identification and characterization of parathyroid hormone/parathyroid hormone-related peptide receptor in cultured astrocytes. *Biochem Biophys Res Commun*. 1994;200:1042–1048.
99. Fukayama S, Tashjian AH, Davis JN, et al. Signaling by N- and C-terminal sequences of parathyroid hormone-related protein in hippocampal neurons. *Proc Natl Acad Sci U S A*. 1995;92:10182–10186.
100. Gu Z, et al. Absence of PTHrP nuclear localization and carboxyl terminus sequences leads to abnormal brain development and function. *PLoS One*. 2012;7:e41542.
101. Gessi M, Monego G, Lauriola L, et al. Parathyroid hormone-related peptide (hPTHrP) and parathyroid hormone-related peptide receptor type 1 (PTHR1) expression in human thymus. *J Histochem Cytochem*. 2005;53:955–962.
102. Kramer S, et al. Immunological identification and distribution of parathyroid hormone-like protein polypeptides in normal and malignant tissues. *Endocrinology*. 1991;128:1927–1937.
103. Jia KZ, et al. Absence of PTHrP nuclear localization and C-terminus sequences leads to abnormal development of T cells. *Biochimie*. 2017;138:13–19.
104. Bhatia V, et al. Role of parathyroid hormone-related protein in the pro-inflammatory and pro-fibrogenic response associated with acute pancreatitis. *Regul Pept*. 2012;175:49–60.
105. Rámila D, et al. Parathyroid hormone-related protein promotes inflammation in the kidney with an obstructed ureter. *Kidney Int*. 2008;73:835–847.
106. Martín-Ventura JL, et al. Possible role of parathyroid hormone-related protein as a proinflammatory cytokine in atherosclerosis. *Stroke*. 2003;34:1783–1789.
107. Funk JL, Trout CR, Wei H, et al. Parathyroid hormone-related protein (PTHrP) induction in reactive astrocytes following brain injury: a possible mediator of CNS inflammation. *Brain Res*. 2001;915:195–209.
108. Hochane M, et al. Parathyroid hormone-related protein modulates inflammation in mouse mesangial cells and blunts apoptosis by enhancing COX-2 expression. *Am J Physiol Physiol*. 2018;314:C242–C253.
109. Nácher-Juan J, Terencio MC, Alcaraz MJ, et al. Osteostatin inhibits collagen-induced arthritis by regulation of immune activation, pro-inflammatory cytokines, and osteoclastogenesis. *Int J Mol Sci*. 2019;20:3845.
110. Platas J, et al. Anti-senescence and anti-inflammatory effects of the C-terminal moiety of PTHrP peptides in OA osteoblasts. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2017;72:624–631.
111. Luparello C. Parathyroid hormone-related protein (PTHrP): a key regulator of life/death decisions by tumor cells with potential clinical applications. *Cancers*. 2011;3:396–407.
112. Zhang R, et al. Parathyroid hormone-related protein (PTHrP): an emerging target in cancer progression and metastasis. *Adv Exp Med Biol*. 2019;1164:161–178.
113. Henderson MA, et al. Parathyroid hormone-related protein localization in breast cancers predict improved prognosis. *Cancer Res*. 2006;66:2250–2256.
114. Tran TH, et al. Loss of nuclear localized parathyroid hormone-related protein in primary breast cancer predicts poor clinical outcome and correlates with suppressed Stat5 signaling. *Clin Cancer Res*. 2018;24:6355–6366.
115. Lyra-Junior PCM, et al. GWAS in breast cancer. *Breast Cancer - From Biology to Medicine*. 2017;100–109.
116. Fleming NI, et al. Parathyroid hormone-related protein protects against mammary tumor emergence and is associated with monocyte infiltration in ductal carcinoma in situ. *Cancer Res*. 2009;69:7473–7479.
117. Li J, et al. PTHrP drives breast tumor initiation, progression, and metastasis in mice and is a potential therapy target. *J Clin Invest*. 2011;121:4655–4669.
118. Kim W, et al. Calcium-sensing receptor promotes breast cancer by stimulating intracrine actions of parathyroid hormone-related protein. *Cancer Res*. 2016;76:5348–5360.
119. Guise TA, et al. Evidence for a causal role of parathyroid hormone-related protein in the pathogenesis of human breast cancer-mediated osteolysis. *J Clin Invest*. 1996;98:1544–1549.
120. Park SI, McCauley LK. Nuclear localization of parathyroid hormone-related peptide confers resistance to anoikis in prostate cancer cells. *Endocr Relat Cancer*. 2012;19:243–254.
121. Martin TJ, Johnson RW. Multiple actions of parathyroid hormone-related protein in breast cancer bone metastasis. *Br J Pharmacol*. 2019;178(9):1923–1935.
122. Takagaki K, et al. Parathyroid hormone-related protein expression, in combination with nodal status, predicts bone metastasis and prognosis of breast cancer patients. *Exp Ther Med*. 2012;3:963–968.
123. Xu C, et al. Co-expression of parathyroid hormone related protein and TGF-beta in breast cancer predicts poor survival outcome. *BMC Cancer*. 2015;15:925.
124. Washam CL, et al. Identification of PTHrP(12-48) as a plasma biomarker associated with breast cancer bone metastasis. *Cancer Epidemiol Biomarkers Prev*. 2013;22:972–983.
125. Kamalakar A, et al. PTHrP(12-48) modulates the bone marrow microenvironment and suppresses human osteoclast differentiation and lifespan. *J Bone Miner Res*. 2017;32:1421–1431.
126. Lee J-Y, et al. Effects of serum albumin, calcium levels, cancer stage and performance status on weight loss in parathyroid hormone-related peptide positive or negative patients with cancer. *Endocrinol Metab*. 2018;33:97.
127. Kir S, et al. Tumour-derived PTH-related protein triggers adipose tissue browning and cancer cachexia. *Nature*. 2014;513:100–104.
128. Kir S, et al. PTH/PTHrP receptor mediates cachexia in models of kidney failure and cancer. *Cell Metab*. 2016;23:315–323.
129. Noda H, et al. An inverse agonist ligand of the PTH receptor partially rescues skeletal defects in a mouse model of Jansen's metaphyseal chondrodysplasia. *J Bone Miner Res*. 2020;35:540–549.
130. Horwitz MJ, et al. Parathyroid hormone-related protein for the treatment of postmenopausal osteoporosis: defining the maximal tolerable dose. *J Clin Endocrinol Metab*. 2010;95:1279–1287.

131. Barrionuevo P, et al. Efficacy of pharmacological therapies for the prevention of fractures in postmenopausal women: a network meta-analysis. *J Clin Endocrinol Metab*. 2019;104:1623–1630.
132. Lanske B, et al. PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science*. 1996;273:663–666.
133. Holick MF, Chimeh FN, Ray S. Topical PTH (1-34) is a novel, safe and effective treatment for psoriasis: a randomized self-controlled trial and an open trial. *Br J Dermatol*. 2003;149:370–376.
134. Katikaneni R, et al. Treatment for chemotherapy-induced alopecia in mice using parathyroid hormone agonists and antagonists linked to a collagen binding domain. *Int J Cancer*. 2012;131(5):E813–821.
135. Maycas M, et al. PTHrP-derived peptides restore bone mass and strength in diabetic mice: additive effect of mechanical loading. *J Bone Miner Res*. 2017;32:486–497.
136. Hofbauer LC, Brueck CC, Singh SK, et al. Osteoporosis in patients with diabetes mellitus. *J Bone Miner Res*. 2007;22:1317–1328.

49

Calcitonin

David M. Findlay and Rachel A. Davey

OUTLINE

Calcitonin Synthesis, Secretion, and Cells of Origin, 835
Calcitonin Chemistry, 836
Calcitonin Gene, 836
  Calcitonin Gene Alternative Product, 836
Calcitonin Receptors, 837
  Receptor Gene, 837
  Receptor Isoforms, 837
  Receptor Polymorphisms, 837
  Receptor Distribution, 838
  Calcitonin Receptors and Receptor Activity-Modifying Proteins, 838
  Calcitonin–Calcitonin Receptor Signaling, 839
  Receptor Regulation, 840
Calcitonin Physiology, 840
  Calcitonin in Growth and Development, 840
  Calcitonin in Calcium Stress: Pregnancy and Lactation, 841
  Renal Actions of Calcitonin, 842
  Calcitonin in the Central Nervous System, 843
Peptides Related to Calcitonin, 844
Calcitonin and its Receptors in Cancer, 844
  Medullary Carcinoma of the Thyroid, 844
  Calcitonin Receptors in Other Cancers, 844
Calcitonin as a Therapeutic, 845
  Hypercalcemia, 845
  Osteoporosis and Paget Disease of Bone, 845
  Osteoarthritis, 845
Conclusion, 846

CALCITONIN SYNTHESIS, SECRETION, AND CELLS OF ORIGIN

The detailed history of the discovery of calcitonin (CT) has been expounded by Paul Munson and is a fascinating story of endocrinology.[1] The narrative describes how Copp and colleagues[2] discovered CT as a factor that opposes the action of parathyroid hormone (PTH) to contribute to the tight control of serum calcium in mammals. By perfusing the thyroparathyroid glands of dogs and sheep with high calcium concentrations, they obtained evidence for the secretion of a factor that rapidly lowers blood calcium levels, and called this factor CT. Although initially thought to be produced by the parathyroid gland, CT was subsequently found to be secreted by the thyroid gland in mammals.[3,4]

CT is produced by the C cells of the mammalian thyroid, from which its secretion is stimulated by an increase in serum calcium levels.[5] Interestingly, the distribution of C cells varies considerably across animal species. In some animals, CT-producing cells are found in other parts of the neck, including the thymus, and in fish and most birds CT is produced by the ultimobranchial glands. Whereas in mammalian development the ultimobranchial bodies fuse with the posterior lobes of the developing thyroid to become the C cells, in submammalian vertebrates these bodies remain separate. Interestingly, the CTs of ultimobranchial origin are highly potent in their actions on mammalian targets. The physiological significance of CT in fish and birds remains uncertain, although CT has been reported to suppress osteoclastic activity in the scales of freshwater and seawater teleosts.[6] Increased serum calcium is an important secretagogue for CT in normal and malignant C cells, and as the same extracellular calcium-sensing receptor (CaSR) that mediates decreased PTH secretion from parathyroid cells[7] is also found in C cells, it is likely to represent the primary molecular entity through which C cells detect changes in extracellular calcium and control CT release. Further suggesting that the CaSR is a physiological regulator of CT, the dose-response curve of CT to increases in serum ionized calcium was shifted to the right in CaSR heterozygous-ablated ($Casr^{+/-}$) mice.[7] Also consistent with this notion is the recent observation that CT release is sensitized to extracellular calcium by L-amino acids, as the CaSR is positively modulated by L-amino acids in the context of PTH secretion.[8]

Agents that elevate cyclic adenosine monophosphate (cAMP) in C cells may stimulate CT secretion, as cAMP analogs have this effect both *in vitro* and *in vivo*. Aside from calcium, secretagogues include the gastrointestinal hormones, such as gastrin, suggesting that CT may have a physiologic role postprandially as a hormone that assists uptake of calcium after a calcium-rich meal by preventing the efflux of calcium from bone into blood. The gastrin analog pentagastrin has been used clinically as a provocative test for CT secretion in patients with medullary carcinoma of the thyroid. Other gastrointestinal hormones, including glucagon, cholecystokinin, and secretin, are also capable of promoting CT secretion. Long-acting glucagon-like peptide-1 receptor agonists also stimulate CT secretion in rats and mice but reportedly have minimal or no ability to stimulate CT secretion in the normal C cells of primates and humans.[9] Other hormones that influence calcium homeostasis may also directly or indirectly influence CT secretion. 1,25-dihydroxyvitamin D_3 ($1,25(OH)_2D_3$) administration has been reported to increase plasma CT levels. Both CT and $1,25(OH)_2D_3$ levels are raised in pregnancy and lactation, leading to the suggestion that CT may act to protect the skeleton in the face of increased calcium demand during lactation (see later).[9]

Basal serum levels of CT in normal human subjects are very low (<10 pg/mL [3 picomolar]), and their measurement requires sensitive assays that are specific for the monomeric form of the molecule.[10]

Circulating levels of CT are increased in several pathologic states, particularly CT-secreting tumors, so that patients with medullary carcinoma of the thyroid can have highly elevated CT levels.[11] In addition to the CT monomer (~3500 Daltons), high molecular weight forms circulate, and elevated levels of these molecules can be useful diagnostically in certain situations, such as acute pancreatitis and infection/inflammatory conditions.[12] It has been reported that procalcitonin (proCT) is toxic, and that immunoneutralization with immunoglobulin G that is reactive to this molecule significantly improves survival in animal models of sepsis.[13] A prospective, population-based study of 3322 individuals demonstrated an association between plasma proCT levels and all-cause and cancer mortality in apparently healthy men, but not women.[14]

CALCITONIN CHEMISTRY

The amino acid sequence of CT has been determined for many species, with the common features being that it is a 32–amino acid peptide with a C-terminal proline amide, as well as a disulfide bridge between cysteine residues at positions 1 and 7 creating an N-terminal ring structure.[15] Based on their amino acid sequence homologies, the different CTs (Fig. 49.1) are classified into three groups: (1) *artiodactyl*, which includes porcine, bovine, and ovine; (2) *primate/rodent*, which includes human and rat; and (3) *teleost/avian*, which includes salmon, eel, sardine, goldfish, and chicken. The common structural features of the CT molecule contribute importantly to biological activity. The standard assay for CT bioactivity that has been used since the discovery of CT is the measurement of the hypocalcemic response in young rats. Subsequently, receptor-based assays have been used also, and structure/function relationships are largely consistent in these various assays. The relative order of biological potency of the CTs is, in general, teleost≥artiodactyl≥human, although the absolute biological activities vary considerably between the CT receptors (CTRs) of different species and between receptor isoforms within species. Studies of substituted, deleted, and otherwise modified CTs have provided considerable information regarding structure/activity relationships of the CT molecule, showing, for example, that the ring structure serves to stabilize the molecule. The disulfide bridge of the ring can be chemically substituted by an N-N bond, as in aminosuberic eel CT, and this modification yields an extremely stable and fully potent CT variant.[16] The sequence differences among species are concentrated in the middle portion of the molecule, and these differences contribute to the wide variations in biological potency. However, the outcomes of studies of structural requirements for biological activity vary with the different biological assays used, and the type of receptor (see later sections on CTR isoforms and the influence of the coreceptor molecules, receptor activity–modifying proteins [RAMPS]). For example, residues in the C-terminal half of salmon CT are more important for binding competition with the two rat CTR isoforms and the human CTR, whereas residues in the N-terminus are more important for interaction with the porcine CTR.[17]

CALCITONIN GENE

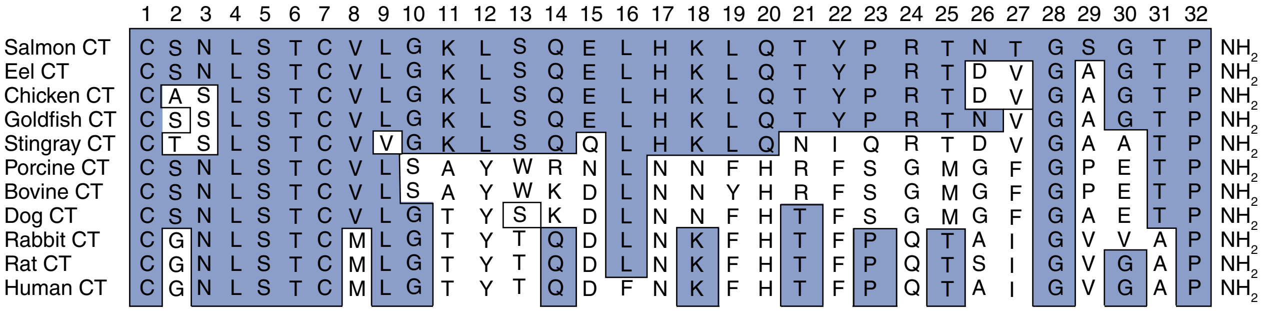
The human gene encoding CT is located in the p14-qter region of chromosome 11.[18] The complete sequences of the gene encoding human CT, and those of a number of other species, including rat, mouse, chicken, sheep, dog CT, as well as various species of fish CTs, have been determined. As with other peptide hormones, CT is synthesized as a larger precursor molecule, which is processed by cleavage and amidation before secretion. A 136–amino acid precursor has a leader sequence at the N-terminus that is cleaved during transport of the molecule into the endoplasmic reticulum. It is uncertain whether the N- and C-terminal flanking peptides that are released by the cleavage have any biological significance. A potentially important posttranslational modification of CT is that of glycosylation. The tripeptide sequence Asn-Leu-Ser found within the N-terminal ring structure of CT is invariate among the CTs of different species, and this sequence is an acceptor site for N-linked glycosylation. The biological significance of CT glycosylation has yet to be determined.

Calcitonin Gene Alternative Product

The CT gene encodes a second distinct peptide, known as CT gene–related peptide (CGRP), which is produced by tissue-specific alternative splicing (see ref 18 and Fig 49.2). The CGRP and CT mRNAs predict proteins that share sequence identity in the N-terminal regions but are almost entirely different in the C-terminal regions. The mature, secreted 32– and 37–amino acid CT and CGRP peptides, respectively, result from cleavage of both N-terminal and C-terminal flanking sequences at specific cleavage sites, as depicted in Fig 49.2. CT mRNA is found largely in the thyroid, and CGRP mRNA is found primarily in the nervous system,[19] although both mRNAs are expressed to a lesser extent in other tissues. Aberrant expression of CGRP may be seen in medullary thyroid carcinoma. Two different CT/CGRP genes, α and β, have been identified in humans and rats.

KEY POINTS

- Calcitonin (CT) is a peptide hormone that in mammals is produced in the thyroid gland, with a primary recognized action of regulating blood calcium, and its secretion is stimulated by an increase in serum calcium levels. The potency of CT action in different species varies considerably. CT is also produced in diverse tissues and acts in a paracrine fashion. The gene encoding calcitonin also encodes the peptide CGRP, which has a very different bioactivity profile.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
Salmon CT	C	S	N	L	S	T	C	V	L	G	K	L	S	Q	E	L	H	K	L	Q	T	Y	P	R	T	N	T	G	S	G	T	P	NH$_2$
Eel CT	C	S	N	L	S	T	C	V	L	G	K	L	S	Q	E	L	H	K	L	Q	T	Y	P	R	T	D	V	G	A	G	T	P	NH$_2$
Chicken CT	C	A	S	L	S	T	C	V	L	G	K	L	S	Q	E	L	H	K	L	Q	T	Y	P	R	T	D	V	G	A	G	T	P	NH$_2$
Goldfish CT	C	S	S	L	S	T	C	V	L	G	K	L	S	Q	E	L	H	K	L	Q	T	Y	P	R	T	N	V	G	A	G	T	P	NH$_2$
Stingray CT	C	T	S	L	S	T	C	V	V	G	K	L	S	Q	Q	L	H	K	L	Q	N	I	Q	R	T	D	V	G	A	A	T	P	NH$_2$
Porcine CT	C	S	N	L	S	T	C	V	L	S	A	Y	W	R	N	L	N	N	F	H	R	F	S	G	M	G	F	G	P	E	T	P	NH$_2$
Bovine CT	C	S	N	L	S	T	C	V	L	S	A	Y	W	K	D	L	N	N	Y	H	R	F	S	G	M	G	F	G	P	E	T	P	NH$_2$
Dog CT	C	S	N	L	S	T	C	V	L	G	T	Y	S	K	D	L	N	N	F	H	T	F	S	G	M	G	F	G	A	E	T	P	NH$_2$
Rabbit CT	C	G	N	L	S	T	C	M	L	G	T	Y	T	Q	D	L	N	K	F	H	T	F	P	Q	T	A	I	G	V	V	A	P	NH$_2$
Rat CT	C	G	N	L	S	T	C	M	L	G	T	Y	T	Q	D	L	N	K	F	H	T	F	P	Q	T	S	I	G	V	G	A	P	NH$_2$
Human CT	C	G	N	L	S	T	C	M	L	G	T	Y	T	Q	D	F	N	K	F	H	T	F	P	Q	T	A	I	G	V	G	A	P	NH$_2$

Fig. 49.1 **Alignment of calcitonin (*CT*) sequences from different species.** The shaded amino acids indicate identity with salmon CT. All CTs have a disulfide-bridged loop between cysteines (C) at position 1 and 7, a glycine at position 28, and a proline amide at position 32. Amino acids 4, 5, and 6 are also conserved across all species.

CALCITONIN RECEPTORS

Receptor Gene

The *CTR* gene sequence for human and other species, including rat, mouse, rabbit, guinea pig, and several species of fish, has been reported, and the phylogenetic relationships between CTRs of mammalian and nonmammalian species have been well described.[20-22] Analyses of the corresponding predicted proteins revealed that these receptors comprise approximately 500 amino acids and belong to the class B subclass of G protein–coupled receptors, which includes the receptors for other peptide hormones, such as secretin, PTH, glucagon, glucagon-like peptide 1, vasoactive intestinal polypeptide, pituitary adenylate cyclase–activating peptide, and gastric inhibitory peptide.

The CTR, encoded by the *CALCR* gene, is greater than 90 kb and comprises 14 exons.[23] The complexity of *CALCR* is shared by the genes encoding other class II G protein-coupled receptors, and the overall structural organization and size of the human gene are similar in other species.[23,24] Isolation of the mouse *Calcr* revealed that the locations of introns within the coding region of the *mCTR* gene (exons E3–E14) are identical to those of the porcine and human *CALCR* genes. The predicted structure of the mefugu fish gene is more complex than that of the human gene, with an additional nine exons.[22] It is interesting to note that both fish and mammalian genes code for the microRNA-489, in intron 3, the function of which remains to be determined. As detailed below, different CTR phenotypes are generated by alternate splicing of the primary mRNA transcript.

The regulatory portion of the *CALCR* gene contains three putative promoters (P1, P2, and P3), giving rise to multiple CTR mRNA transcripts, which differ in the 5′ region of the gene and generate 5′-untranslated regions (UTRs) of very different lengths, in a tissue-specific manner.[24] Promoters P1 and P2 are used in osteoclasts, brain, and kidney, while the proximal promoter of the human *CTR* (*hCTRP1*) was transcriptionally active in all cell lines tested, with high-level activity dependent on an 11–base pair Sp1/Sp3 binding site. In contrast, promoter P3 appears to be osteoclast-specific and is sensitive to the osteoclast-inducing cytokine receptor activator of nuclear factor kappa-β ligand (RANKL), as well as the RANKL-induced transcription factor NFATc1, consistent with a role for the latter in regulating the *CALCR* gene in osteoclasts.[24]

The human *CALCR* gene is located on chromosome 7, at 7q21.3. In mice it is in the proximal region of chromosome 6, while the pig gene is located in chromosomal band 9q11-12. These pig and mouse chromosomal regions are homologous to 7q in humans. It is interesting to note that the mouse *Calcr* gene is expressed preferentially from the maternal allele in brain, with no allelic bias detected in other tissues, indicating that the mouse *Calcr* gene is imprinted in a tissue-specific manner.[25]

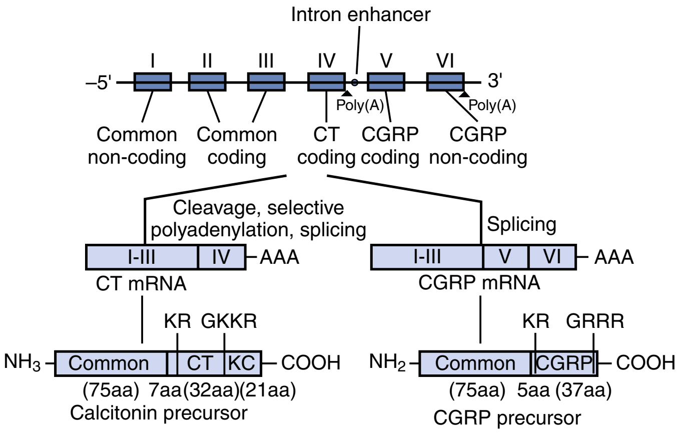


Fig. 49.2 Organization of the calcitonin (*CT*)/calcitonin gene–related peptide (*CGRP*) gene, illustrating alternative patterns of processing of the primary transcript and subsequent protein processing. The exons are denoted I through VI in Roman numerals, and introns are represented by a single line (not to scale).

Receptor Isoforms

CTRs show heterogeneity, with multiple receptor isoforms that result from alternative splicing of the CTR mRNA primary transcript. The human receptor, of which there are at least five splice variants, is the most extensively studied. The most common hCTR splice variant occurs in putative intracellular domain 1, generating a 16–amino acid insert in this domain (I_{1+} or CTb isoform); the insert-negative isoform is termed CTa.[26] Additional splice variants have been identified in other species. In the pig, an inserted sequence (the I1 insert) is generated by alternate splice sites located in exon 8, while in humans the I1 insert occurs on a separate exon. In rodents, alternate splicing leads to two receptor isoforms, which differ from the human isoforms: these have been termed C1a and C1b, and vary by the presence (C1b, or E_{2+}) or absence (C1a), respectively, of a 37–amino acid insert in the putative second extracellular domain.[27] There is no evidence for expression of the rodent C1b isoform in humans. In rabbits, an additional splice variant in which the exon encoding transmembrane domain 7 is spliced out (ΔTM7, or delta e13) has been isolated.[28]

The insertions and deletions in the CTR structure lead to altered receptor function, including ligand binding, signal transduction, and regulation of receptor expression. For the E_{2+} variants, there is a loss of affinity for peptides exhibiting weak α-helical secondary structure, such as human CT, whereas peptides with strong helical secondary structure, such as salmon CT, maintain high affinity at this receptor.[29] An interesting variant of the hCTR that lacks the first 47 residues of the N-terminus, the Δ1-47 variant, retains expression at the cell surface but has altered functionality in terms of ligand (amylin [AMY]) binding, depending on which RAMP coreceptor (see below) it associates with.[30] The CTb (I_{1+}) hCTR isoform displays similar affinity to the CTa (I_1) variant for CT-like peptides but has markedly altered G protein–coupling efficiency. However, the presence of the 16–amino acid insert leads to complete loss of intracellular calcium mobilization. CT signaling through the CTb hCTR via G_s is cell type–dependent but overall is decreased, leading to a 10- to 100-fold reduction in the potency of peptides for stimulation of cAMP production.[29] It is interesting to speculate about the functional significance of the different receptor isoforms, particularly those with apparently impaired signaling capacity. Accumulating evidence suggests that G protein–coupled receptors can form homodimers and heterodimers, and that these interactions can crossmodulate the activity of the partners. Indeed, in the case of the rabbit CTR (ΔTM7), Seck et al.[31] reported that this isoform can complex with the CTa isoform and inhibit its cell surface expression and activity, thus exerting an endogenous dominant negative effect on CTR signaling.

Receptor Polymorphisms

There are reports of polymorphisms in the *CTR* gene associating with bone disease, specifically fracture risk, [32-35] although these studies are often underpowered to draw conclusions. There is a reported association between the expression of the CTa receptor mRNA in circulating monocytes of postmenopausal, but not premenopausal, women and their serum levels of bone alkaline phosphatase and urinary deoxypyridinolone.[32] The authors suggested a link between CTR expression and increased bone resorption postmenopausally. In other work, a polymorphism within the *CTR* gene coding sequence arising from a single nucleotide substitution was identified that leads to either a proline (CC genotype) or a leucine (TT genotype) at amino acid 447 in the human CTa receptor (amino acid 463 of the hCTb receptor), with heterozygotes designated TC.[33,34] The prevalence of different polymorphisms varies markedly between different ethnic groups.

The influence of the *CTR* polymorphism on bone mineral density (BMD) has been studied by multiple investigators. A number of studies have identified significant correlations between *CTR* genotype and osteoporotic markers (BMD) and fracture risk; however, there is considerable divergence among studies regarding which genotype is more favorable, with the same polymorphism being correlated with both increased BMD and decreased BMD.[33,34] The *CTR* genotype has also been linked to body weight in premenopausal Japanese women, and to incidence of kidney stones. These studies have not identified significant functional differences in ligand binding or cAMP generation between the two polymorphic variants.[35] A large number of other, probably silent, polymorphisms in *hCTR*, whose frequency varies among ethnic groups, have been identified.

In assessing polymorphism association studies, it is important to appreciate issues of study power, ethnicity, gene interactions, and other confounders. Genome-wide association studies have so far not linked the CTR with metabolic bone changes. Surprisingly, however, linkage studies in rats have identified the *CALCR* gene as a candidate for regulating susceptibility to herpes simplex type 1 neuronal infection leading to encephalitis.[36]

Receptor Distribution

Direct evidence for binding of CT to osteoclasts was obtained from the work of Nicholson et al.,[37] who used *in vitro* autoradiography to demonstrate ^{125}I-sCT binding specifically to mono- and multinuclear osteoclasts and their precursor cells. Expression of the CTR became a required phenotypic feature to establish cells as osteoclasts. Radioligand binding studies in tissue sections, membranes, and cultured cells have revealed extensive extraskeletal distribution of CTRs. These extraskeletal sites include kidney, brain, pituitary, placenta, lung, lymphocytes, testis, and spermatozoa, as well as cancer-derived cells from lung, breast, pituitary, and embryonal carcinoma.[38]

The centrally mediated actions of CT are discussed later and correlate well with the locations of CT binding sites. Autoradiographic mapping, where ^{125}I-sCT was used as the radioligand, revealed high binding densities associated with parts of the ventral striatum and amygdala, the hypothalamic and preoptic areas, as well as most of the circumventricular organs. High-density binding also occurs in parts of the periaqueductal gray, the reticular formation, most of the midline raphe nuclei, parabrachial nuclei, locus coeruleus, and solitary tract nucleus.[39,40] As is discussed later, CT and AMY receptors derive from the same gene product, and consequently care needs to be taken in extrapolating the potential significance of CTR mRNA expression for CT biology. CTR mRNA expression has been confirmed in many of the sites with demonstrated ^{125}I-CT binding, including kidney, brain, lung, placenta, and osteoclasts, and a number of cancer cell lines. In addition, CTR mRNA studies have also provided evidence for previously uncharacterized sites of CTR expression, including prostate, normal and neoplastic breast, and thyroid.[38] There is also evidence for CTR expression in osteocytes in bone.[41]

Calcitonin Receptors and Receptor Activity-Modifying Proteins

Members of the CTR family, and the CTR itself, can assume different ligand binding phenotypes by two distinct mechanisms. Firstly, and as already discussed, receptor binding affinity and signaling depend on the specific receptor isoform. Secondly, different functional receptors arise from association of the CTR with coreceptor molecules. Specifically, molecules termed RAMPs can associate with the CTR to alter ligand affinity[42] and downstream signaling.[43] The RAMP family comprises three single transmembrane proteins, RAMP 1, 2, and 3, which act to assist trafficking of receptors to the cell surface and/or physically associate with the receptor to modulate ligand binding and specificity. CTRs do not require RAMPs to translocate to the cell surface or to exhibit a classic CTR phenotype, that is, with highest affinity for CT. In the case of the CTR, RAMP-CTR complexes give rise to receptors for AMY (Fig. 49.3), with distinct receptor phenotypes depending on which RAMP is complexed with the receptor.[44,45] The CTR/RAMP 1 complex, in addition to being a high-affinity AMY receptor, also interacts strongly with CGRP.[44] Both RAMPs and receptors are subject to dynamic regulation.[46]

KEY POINTS

- The calcitonin receptor (CTR) is a G-protein coupled protein that spans the cell membrane seven times. It is produced in a number of species-specific isoforms that have different activities. It also associates with the coreceptor molecules, receptor activity-modifying proteins, which confer different ligand recognition capacity. The CTR has a wide tissue distribution associated with the different actions of calcitonin (CT) and CT-like peptides.

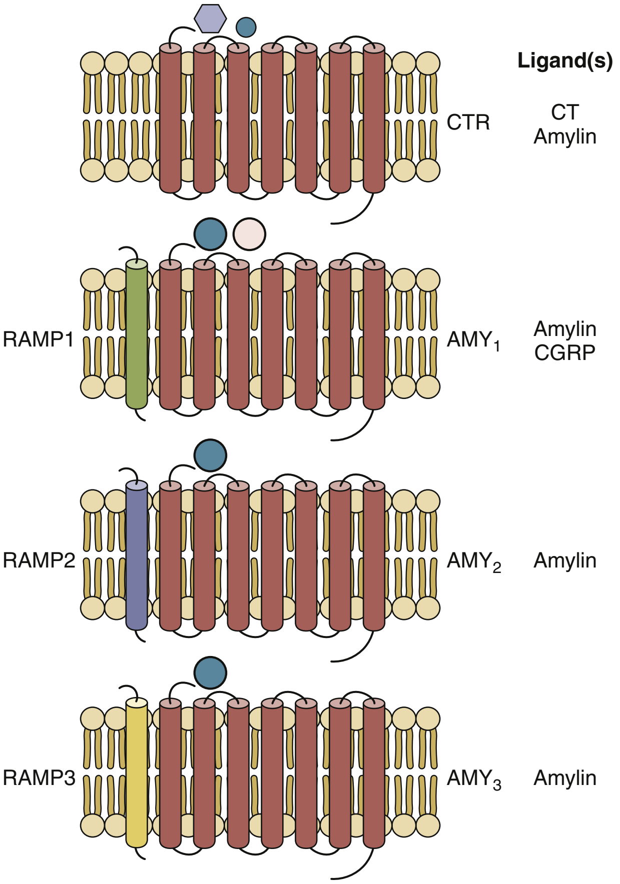


Fig. 49.3 **Calcitonin receptor (*CTR*) ligand specificity regulation by receptor associated–modifying proteins (*RAMPS*).** Shown are ligand specificities and relative affinities for calcitonin (*CT*), amylin, and calcitonin gene–related peptide (*CGRP*), respectively, of CTR alone or when associated with RAMPs 1, 2, or 3, respectively. (Modified from Hay DL, Garelja ML, Poyner DR, et al. Update on the pharmacology of calcitonin/CGRP family of peptides: IUPHAR Review 25. *Br J Pharmacol.* 2018;175:3–17.)

Calcitonin–Calcitonin Receptor Signaling

G protein–coupled receptors can interact with and signal through multiple G proteins (Fig. 49.4), with G_s activation stimulating adenylate cyclase activity, cAMP production, and protein kinase A (PKA)-induced signaling pathways. This occurs in osteoclasts and in the kidney, with the CT responsiveness spatially paralleling the distribution of CTRs in the latter tissue. CT induction of cAMP has now been documented in a large number of CTR-bearing cells, which include LLC-PK1 pig kidney cells and cancers of lung, breast, prostate, and bone. Receptor cloning and expression studies have confirmed that cAMP production is an important component of CTR-mediated signaling.[15,29,38]

In mice, the response of osteoclasts to CT appears to be mediated predominantly by the PKA pathway,[47,48] although inhibition of osteoclast-mediated bone resorption by CT can be mimicked by both dibutyryl cAMP and phorbol esters or blocked by protein kinase inhibitors. Coupling of the CTR to G_q can also activate phospholipase C, leading to increased cytosolic calcium and triphosphate levels and activation of protein kinase C.[49] Thus, in human osteoclasts, activation of protein kinase C pathways appears to be the predominant mechanism of CT-mediated osteoclast inhibition.

CT induction of interleukin (IL)-6 production in pituitary folliculostellate cells requires both PKA and protein kinase C (PKC) signaling pathways.[50] In hepatocytes, even at very low concentrations, CT is capable of increasing cytosolic calcium.[51] CT-induced differentiation of early rat embryos is dependent upon intracellular calcium mobilization.[52] It is interesting to note that, although CT inhibits proton extrusion from osteoclasts,[53] it can increase hydrogen ion efflux from nonosteoclastic cells in a PKC-dependent manner.[54] The biological significance of this activity is not clear. CT-induced changes mediated by cAMP or intracellular calcium in LLC-PK1 pig kidney cells are cell cycle–dependent.[55] Expression of cloned receptors in a variety of cell types has now conclusively shown that CTRs of human, rat, and porcine origin are capable of signaling through both cAMP- and calcium-activated second messenger systems.

CT treatment of CTR-expressing cells in the presence of extracellular calcium initiates a sustained rise in intracellular calcium, the extent of which is dependent on the concentration of the extracellular calcium and is proportional to the receptor density.[56] Because osteoclasts, which express high levels of CTR, are reportedly exposed to calcium concentrations as high as 26 mM during bone resorption, this phenomenon may have particular relevance for this cell type. CT and extracellular calcium can both cause intracellular calcium transients in isolated osteoclasts, and each agent greatly augments the signal produced by either agent alone.[57]

CTR-mediated activation of the mitogen-activated protein kinase (MAPK) pathway has been described,[58-60] with a role for both G_q and $Gi_{/o}$ proteins—the latter principally via βγ G protein subunits, as is commonly the case for other G protein–coupled receptors. There is also evidence that CT can influence cell attachment by modulating components of focal adhesions and the cytoskeleton.[59] CT treatment of osteoclasts disrupts the F-actin ring that is thought to correspond to the sealing zone[61,62] and increases the tyrosine phosphorylation of paxillin and focal adhesion kinase. In rabbit osteoclasts, these actions were mediated by PKC.[59] These actions of CT, together with decreased osteoclast motility, cytosol retraction, and disassembly of podosomes, have been linked to modulation of the activity and intracellular localization of Pyk2 and Src in osteoclasts.[63] It has also been shown, in nonosteoclastic cells, that CT can cause cell death by aniokis as the result of loss of cell attachment.[64]

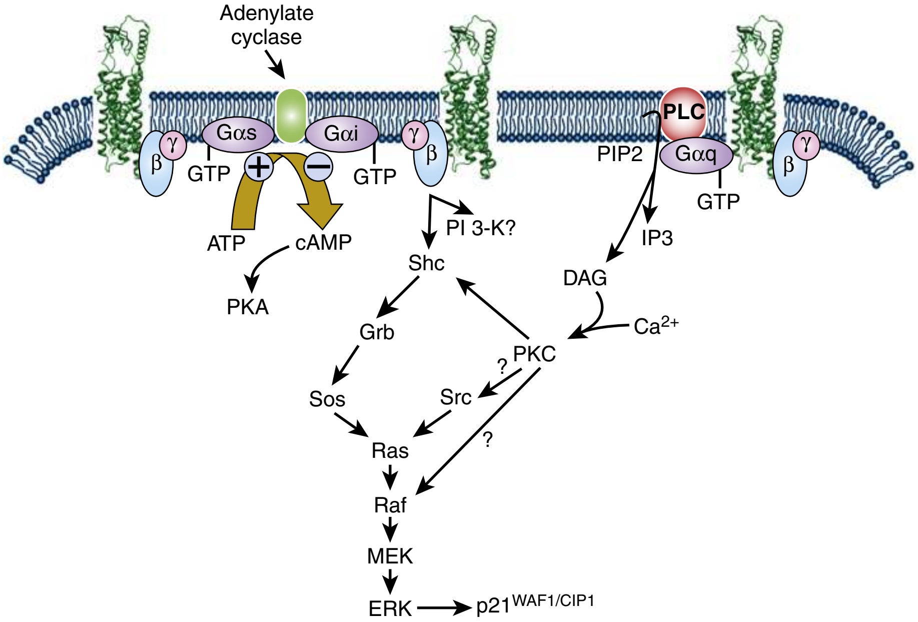


Fig. 49.4 Calcitonin (CT) activation of the calcitonin receptor activates multiple signaling pathways. Coupling to $G_{\alpha i}$ is seen less frequently and may be owing, at least in part, to a protein kinase C (*PKC*)-dependent inhibition of G_i. The balance of activation of different G proteins and effectors is critical for the cell-specific actions of CT. Also illustrated is the current understanding of CT-induced mitogen-activated protein kinase (MAPK) activation, as delineated in stably transfected HEK-293 cells. ERK1/2 phosphorylation requires both PTX-sensitive G proteins (presumably G_i via βγ subunits) and G_q. The latter leads to rises in intracellular calcium and activation of PKC, and both contribute to ERK phosphorylation. At least part of the activation occurs following phosphorylation of Shc and mobilization of its downstream effectors. Activation of ERK increases phosphorylation of $p21^{WAF/CIPI}$ and modulation of cell growth. Although strongly activated by CTa receptors, G_s is not involved in activation of MAPKs.

Receptor Regulation

Regulation of the level and/or affinity of cell surface receptors is a key component of the physiologic and pathophysiologic responses to both endogenous and pharmacologically administered agents. The CTR is subject to both homologous (CT-induced) and heterologous regulation. CT-induced downregulation of the CTR was initially demonstrated in various transformed cell lines and subsequently in primary kidney cell cultures. Given that the induction by CT of resistance to its own action in osteoclasts is such a specific process, and that it correlates so closely in dose dependence to CT efficacy as an agonist,[65] it appears to be an important accompaniment of CT action.

CT-induced receptor downregulation, resulting in a specific loss of cell surface receptors, occurs via energy-dependent internalization of the ligand-receptor complex, in which the principal internalization pathway involves processing of the receptor-ligand complex into lysosomes and subsequent degradation of the receptor. It is worth noting that the kinetics of internalization of human CTRs appears to be isoform-specific, with the I_{1+} variant being resistant to internalization.[26] Receptor regulatory responses to CT are also likely to be cell- or tissue-dependent. For example, in mouse or rat osteoclasts, a potent downregulation of CTR mRNA appears to be mediated by a cAMP-dependent mechanism, in addition to downregulation of the receptor by internalization.[66] Similar CT-mediated CTR regulation is also observed in human osteoclasts, but as stated above, activation of PKC pathways is thought to be the predominant mechanism of CT-mediated osteoclast inhibition in human osteoclasts. Different mechanisms may apply in nonosteoclastic cells, as downregulation of CTRs in UMR106-06 osteoblast-like cells and T47D breast cancer cells is not accompanied by changes in CTR mRNA levels.[66] The CT-induced reduction in CTR mRNA in osteoclasts appears to be due principally to destabilization of receptor mRNA.[67] The 3’-UTR of the mouse and rat CTR mRNA contains four AUUUA motifs, as well as other A/U--rich domains, which can function as signals for rapid mRNA inactivation. Thus, addition of A/U-rich CTR 3’-UTR sequences considerably shortens the mRNA half-life of a reporter gene, and evidence shows the involvement of AUF1 p40, HuR in the regulation of CTR mRNA.[67] Whether this regulation involves the participation of microRNAs, which have been shown to regulate the expression of other CGPRs, has not yet been reported.

Heterologous regulation of CTR includes regulation by glucocorticoids. In mouse osteoclast cultures, dexamethasone increased the level of cell surface CTR by upregulation of receptor mRNA levels, which was mediated at the level of transcription.[68] Moreover, the CT-mediated decrease in cell surface receptor and mRNA was attenuated by dexamethasone. Increased production of CTR in response to glucocorticoid stimulation also occurs in the human T47D breast cancer cell line, where cortisol is actually required for the expression of CTRs, suggesting that this may be a common regulatory mechanism for induction of CTR expression. It is worth noting the clinical evidence suggesting that glucocorticoids, given together with CT, might prevent to some extent the CT-induced resistance to its own action.[69]

KEY POINTS

- Calcitonin (CT) binding to its cognate receptor (CTR) results in intracellular signaling, predominantly via the protein kinase A– and protein kinase C–activated signaling pathways. Signaling stimulates different, cell-dependent actions. In addition, in at least some target cells, such as osteoclasts, CT signaling initiates downregulation of the CTR, via cell internalization of the receptor and reduced synthesis of the CTR.

CALCITONIN PHYSIOLOGY

The physiological roles of CT are not fully understood, although the repertoire of actions of the peptide has expanded considerably since its discovery. As a hormone, CT has been described as “enigmatic”[70] or “in the process of becoming vestigial,”[71] because there are no apparent deficiency syndromes and no obvious effects of high concentrations because of CT-producing tumors. However, there is clearly more to learn, and important roles for locally produced CT have been uncovered, as detailed here.

Calcitonin in Growth and Development

The presence of CT and its receptors in a large number of cell types and tissue sites suggests multiple physiological roles. CTRs have been identified in bone, kidney, brain, pituitary, testis, prostate, spermatozoa, breast, lung, and lymphocytes, as well as in cancer-derived cells from lung, breast, prostate, pituitary, bone (osteoclastoma, osteogenic sarcoma), and embryonal carcinoma.[40] One potential role of CT relates to its ability to potently modulate the growth and/or viability of some CTR-bearing cells. Treatment with CT has been shown to stimulate the growth of human prostate cancer (PC) cells, in which the peptide increases both intracellular calcium and cAMP levels.[72] Conversely, knockdown of CTR expression in these cells induced apoptosis and growth arrest.[73] The effects of CT on growth are clearly cell type–dependent, as CT treatment repressed the growth of T47D human breast cancer cells, an action proposed to be mediated by the specific activation of the type II isoenzyme of the cAMP-dependent protein kinase.[74] In HEK-293 kidney cells overexpressing the human CTR, CT inhibited cell growth and caused an accumulation of cells in the G2 phase of the cell cycle, associated with a prolonged increase in $p21^{WAF1/CIP1}$ expression[75] and sustained activation of the p42/44 MAPK proteins.[76]

There is evidence consistent with the involvement of CT also in the early events of embryo development. CT was shown to promote development of preimplantation embryos, an action that was antagonized by p38-MAPK inhibition or anti-CT antibody treatment.[77] CT may be involved in both blastocyst implantation and development of the early blastocyst,[78] and CT downregulates E-cadherin expression in rat uterine epithelium during implantation and upregulates integrin β3 expression in mouse endometrial epithelial cells.[79] CT also promotes the outgrowth of trophoblast cells on endometrial epithelial cells, perhaps thereby facilitating embryo implantation.[80] Reduced endometrial expression of CT and β3 integrin could contribute to unexplained infertility in humans.[81] Indeed, CT administration improves endometrial receptivity,[82] and intraperitoneal CT treatment of pregnant or pseudopregnant mice significantly increased blastocyst implantation.[83] These actions of CT may explain the reports of embryonic lethality of global deletion of the *Calcr* gene in mice.[61,84] The embryonic expression of mouse CTR suggests that CTRs may play important roles in morphogenesis.[85]

It has been reported that CT is expressed in the pregnant mouse mammary gland, exclusively in the luminal epithelial cells, from which it is secreted, and that its expression is progesterone-dependent.[86] It was found that CT induction spatiotemporally correlates with increases in progesterone-induced mammary gland proliferation and structural remodeling, suggesting that CT may be involved in one or both of these progesterone-dependent processes. It would be of interest to know whether the established central roles of RANKL in these processes involve CT. As mentioned, CT can inhibit cell proliferation or can have mitogenic actions. CT may also have a role in cell survival, and as discussed above, has been claimed to be protective of drug-induced apoptosis in osteoblast-like, osteocyte-like, and PC cell lines, and to promote the survival of osteoclasts.[87] On the other hand, CT has been found to potentiate neuronal death because of oxygen and glucose deprivation, and to be proapoptotic in serum-deprived cells.[88]

Calcitonin in Calcium Stress: Pregnancy and Lactation

CT was initially understood to be an inhibitor of bone resorption, functioning to regulate loss of calcium (and perhaps phosphate) from the skeleton, particularly in times of stress on skeletal calcium conservation, such as during pregnancy, lactation, and growth.[70] An early observation was that thyroidectomized patients had a significantly prolonged elevated serum calcium following a calcium infusion when compared with thyroid-intact control patients, which the authors were unable to account for mechanistically.[89] More recent studies suggest that the ability of CT to reduce serum calcium levels following an exogenous calcium load can be at least partly explained by reducing the efflux of calcium from bone. Thus, in genetically modified mice, in which the CTR is deficient either globally or specifically in osteoclasts, $1{,}25(OH)_2D_3$ treatment induced greater bone resorption and exacerbated hypercalcemia compared with littermate controls.[90] These data suggest that the action of CT to protect against hypercalcemia is largely mediated by the CTR on osteoclasts. In young and growing animals, in which rapid bone modeling and remodeling are required for morphogenesis of the skeleton, CT may act as a regulator of extracellular fluid calcium. Thus, exogenous CT causes a lowering of blood calcium levels in the young rat, a response which is used as the *in vivo* assay of CT, while its calcium-lowering effect becomes less marked with increasing age of the animal. Despite this, the ability of CT to counteract the effects of a calcium load is not impaired with age, at least in the rat.[91]

It is important that bone resorption is tightly regulated, and CT is the only hormone known to be capable of carrying out this function by acting directly on osteoclasts. However, in normal adult human subjects, even quite large doses of exogenous CT have little effect on basal serum calcium levels. This is not surprising, as the extent of bone resorption is much reduced in maturity. However, in those subjects, in whom bone remodeling is increased (e.g., in thyrotoxicosis and Paget disease), CT treatment acutely inhibits bone resorption and lowers serum calcium levels.[92] The relative physiological actions of CT and local inhibitors of osteoclastic resorption, such as osteoprotegerin, remain to be clarified.

An antiresorptive role for CT in maturity is likely to be more important in circumstances in which skeletal integrity is at particular risk, for example with the high calcium demand that occurs during lactation. In both humans and mice, circulating levels of CT are elevated during lactation,[93] supporting the notion that CT opposes the bone-resorbing actions of mammary-derived PTH-related peptide to protect the maternal skeleton from excessive resorption during lactation.[94] Evidence in support of such an important physiological role for endogenous CT in protecting against bone loss is provided by experiments showing that, in thyroparathyroidectomized rats (lacking CT) treated with PTH, cancellous bone loss was greater than that in similarly treated sham-operated (CT-replete) controls.[95] In addition, mice in which the *CT/CGRP* gene was ablated showed a severe drop in bone mineral content during lactation as a result of a doubling in osteoclast numbers, which was prevented by treatment with CT but not CGRP.[96] Somewhat surprisingly, in mice lacking the *Calcr* gene, bone resorption was not increased following lactation, suggesting that this protective effect of CT on bone in CT/CGRP knockout mice is not mediated via the activity of CTR on osteoclasts, but perhaps involves another member of the CTR family.[97] This study also uncovered a potentially important role for the CTR located on osteocytes to protect the maternal skeleton. Here, activation of the osteocyte CTR inhibited lactation-induced osteocytic osteolysis,[97] the process by which osteocytes mobilize calcium from the perilacunar bone matrix.[98] Recovery of the skeleton does not appear to be dependent on CT/CGRP[99] or CTR[99a] as the maternal skeleton of mice deficient in either ligand or receptor recovers to baseline postweaning. [99]

As discussed in more detail later, physiological roles in calcium homeostasis potentially also exist for other members of the CT family. These include AMY, CGRP, CTR-stimulating peptide (CRSP), and intermedin (IMD).[100]

KEY POINTS

- Calcitonin (CT) appears to have important roles in pregnancy and lactation. Evidence suggests its involvement in both blastocyst implantation and development of the early embryo. In addition, it appears that CT is involved in preparation of the mammary glands for lactation, as well as protection of the skeleton from excessive calcium loss during lactation.

Calcitonin in Bone: Osteoclasts. The first evidence of the mechanism of action of CT was obtained by showing in organ culture of bone that CT inhibited bone resorption.[101] Inhibition of bone resorption appeared to be explained by a direct action on osteoclasts because CT treatment of resorbing bone *in vitro* resulted in rapid loss of osteoclast ruffled borders[102] and decreased release of lysosomal enzymes. Evidence *in vivo* was also consistent with an inhibitory action upon bone resorption. Loss of ruffled borders in osteoclasts was seen in patients with Paget disease, from whom bone biopsies were taken before and 30 minutes after an injection of CT.[103] In the same clinical study, CT was noted to decrease the number of osteoclasts, in addition to altering their ultrastructure. CT infusion in rats led to an immediate reduction in the rate of excretion of hydroxyproline, consistent with inhibition of osteoclastic breakdown of bone collagen.[92] Other studies led to similar conclusions, with no evidence to suggest any increase by CT in the active uptake of calcium by bone.[104]

Mammalian osteoclasts possess abundant, specific, high-affinity receptors for CT, and CT stimulates cAMP formation in osteoclasts dose-dependently,[37] as well as increasing intracellular free calcium levels and PKC activity. Isolated osteoclast cells are exquisitely sensitive to CT, and the direct effect of CT is a rapid inhibition of cell activity, reflected in cessation of osteoclast motility and contraction.[102] Interestingly, although isolated osteoclasts remained quiescent as long as CT was present, they regained activity when osteoblasts were added to the culture. This "escape" of osteoclasts from inhibition by CT took place at a rate proportional to the number of osteoblasts, with which they were in contact. PTH had no effect on CT-reduced cytoplasmic spreading of isolated osteoclasts unless osteoblasts were cocultivated with the osteoclasts, in which case PTH caused a marked increase in cytoplasmic spreading of osteoclasts. These were the first direct observations of actions of hormones on isolated bone cell preparations containing osteoclasts.

The molecular mechanisms by which CT directly and rapidly decreases osteoclast function have been extensively investigated. The rapid effects of the hormone involve cytoskeletal molecules, following generation of several intracellular second messengers. Early events in CT signal transduction have been studied in a variety of cell types and are described in greater detail above. The other means by which CT might inhibit resorption is through inhibition of osteoclast formation, and here the evidence is conflicting. For example, a study in which developing mouse osteoclasts were treated with low concentrations of CT showed no reduction in osteoclast formation.[105] However, the multinucleated osteoclasts that formed in the continuous presence of exogenous CT did have fewer nuclei, suggesting that CT somewhat reduced the fusion of mononuclear cells. In contrast, several reports describe the inhibition of osteoclast-like cell formation by CT in bone marrow cultures of human, baboon, and mouse origin. More recent studies, in which osteoclasts were formed from mouse bone marrow macrophages under the influence of RANKL and macrophage colony stimulating factor, reported that 1 nM CT greatly reduced the formation of multinucleated osteoclasts, while having minimal effect on total tartrate-resistant acid phosphatase (TRAP) activity.[100] In elegant *in vivo* studies in mice, Ikegame et al.[65] showed that a CT-induced drop in serum calcium was linked temporally to the loss of osteoclast ruffled borders. Further, frequent dosing of the animals resulted in insensitivity to CT, with recovery of osteoclast ruffled borders and return of serum calcium to control levels in the presence of CT. It was significant that treatment of mice with CT initially rendered osteoclasts unable to bind ^{125}I-sCT, which recovered after a single treatment but not with repeated treatments.[65] These findings may be relevant to the mechanism of "escape" from CT that is observed clinically.

Calcitonin in Bone: Osteoblasts, Osteocytes, and Bone Formation. Although the best understood action of CT in bone is as an antiresorptive agent, numerous reports have described actions also on cells of the osteoblast lineage, and both direct and indirect effects on bone formation. CT increased cell proliferation, shown by [^{3}H]thymidine incorporation in embryonic chicken calvariae, in the transformed murine calvarial cell lines MMB and MC-3T3-E1, and in primary cultures of cells prepared from newborn mouse calvaria. CT was also shown to stimulate [^{3}H]thymidine incorporation in primary human osteoblasts[106] and to increase expression of insulin-like growth factors by human SaOS-2 cells. It should be noted that data failing to show such effects of CT have also been published. In addition to the fact that relatively high concentrations of CT were used in these early experiments, results with primary osteoblasts should be interpreted with caution, because calvarial osteoblast preparations are usually contaminated with other cell types, including osteoclast precursors and primary osteoblast preparations contain a substantial proportion of bone-specific macrophages, termed "osteomacs." Although at the level of resolution of CTR autoradiography, osteoclasts are the only discernible bone cell CT targets,[37] CTR mRNA and protein expression have since been demonstrated by immunohistochemistry in mouse osteocytes *in situ*, with rapid loss of detectable CTR following removal of these cells from bone.[41] The presence of the CTR in osteocytes can explain an increase in the osteocyte-secreted molecule, sclerostin, in the bones of CT-injected mice.[41] In addition, CT at high concentrations was associated with a small increase in cAMP and with protection of MLO-Y4 osteocyte-like cells from apoptosis induced by etoposide, tumor necrosis factor α, or dexamethasone. It has also been claimed that CT can hasten and improve the process of fracture healing in normal and osteoporotic rats. However, the mechanism for this "anabolic" response could be similar to that of bisphosphonates, which have also been shown to enhance the strength of the healed fracture, apparently by modulating resorption during the remodeling phase of bone repair. It could also relate to possible stimulation of angiogenesis by CT, which has been shown for human microvascular endothelial cells, albeit at supraphysiologic concentrations.

Experiments deleting the *CT* and *Calcr* genes in mice have provided important insights into the physiological significance of CT. Ablation of the *CT/CGRP* gene resulted in viable and fertile mice, with no production of CT or CGRPα.[99] These mice were much less able than wild-type mice to overcome the hypercalcemia induced by a calcium load,[99] and they lost excessive bone during lactation.[107] Surprisingly, however, the mice had increased bone mass, with histomorphometric parameters showing increased bone formation.[107] An increased bone formation phenotype was not found when CGRPα-deficient mice were examined, indicating that the increased bone formation was attributed to CT deficiency, and these results called for a reappraisal of the physiology of CT. The CT/CGRP-deficient mice were protected against ovariectomy-induced bone loss,[107,108] but it was striking that, after 6 months of age, these mice showed severe cortical bone porosity, even though indices of increased bone formation were maintained.[109,110] Taken together, the observations are indicative of an inhibitory effect of CT on both bone formation and bone resorption.

The concept of a negative regulatory role for CT on bone formation was strengthened by the outcome of studies by Dacquin et al.[84] in mice, in which the CTR was genetically manipulated. Homozygous deletion of the *Calcr* gene in mice was embryologically lethal, and this was thought to be because of a placental effect. The *Calcr*$^{+/-}$ mice, however, exhibited a bone phenotype virtually indistinguishable from the *CT/CGRP*$^{-/-}$ mice, with increased bone mass and increased bone formation on histomorphometry. Studies with mice that had substantial knockdown of the CTR also provided evidence of increased indices of bone formation.[65] Together, these results lead to the conclusion that, at least in mice, the removal of CT production or CT action via the CTR results in an increased amount of bone, again supporting a physiological role for CT as a tonic inhibitor of bone formation.

Several possible mechanisms for an effect of CT on bone formation have now been uncovered. Firstly, mouse and rat osteocytes express the CTR, and salmon CT treatment of rats increased sclerostin expression in the bones (Fig. 49.5).[41,111] Sclerostin is a negative regulator of bone formation and acts to inhibit anabolic responses in bone by blocking Wnt signaling. It was shown by microarray analysis of bone RNA and by immunohistochemistry that CT treatment both induced sclerostin mRNA in rat long bones and mouse calvariae and increased the number of sclerostin positive osteocytes in cortical bone from the tibia.[41] These actions are predicted to decrease local bone formation, but they may only be of physiological relevance in young animals. It was subsequently shown that, although the CTR was readily detected in osteocytes in 3-week-old mice, CTR levels declined with age and were undetectable by 49 weeks of age.[111] Secondly, CT has been shown to negatively regulate osteoclastic expression of the *Spns2* gene,[1112] which in turn encodes a transporter for sphingosine 1-phosphate (S1P), a reported osteoclast-osteoblast coupling factor.[113] Thus, CTR-deficient mice show increased S1P levels, which the authors demonstrated is linked to increased bone formation. There are multiple potential twists and caveats in this story, which have been expanded upon in separate publications,[114,115] in which the authors speculated that "a notable conclusion from this work is that the physiological function of CT in bone differs entirely from its pharmacological function."

KEY POINTS

- In bone, calcitonin has potent inhibitory actions on osteoclasts and additional actions on osteocytes, with evidence for both a protective role in the skeleton during lactation and for paradoxical inhibition of bone formation via induction of osteocyte production of the Wnt pathway inhibitor, sclerostin.

Renal Actions of Calcitonin

When infused into thyroparathyroidectomized rats, CT caused a dose-dependent phosphaturia, although this was minor in comparison with the phosphaturic effect of PTH. This was demonstrated in human subjects also, however in several species CT failed to have any effect on phosphate excretion. Thus it seems unlikely that the phosphaturic effect is of any major physiologic significance. A number of other renal effects of CT have

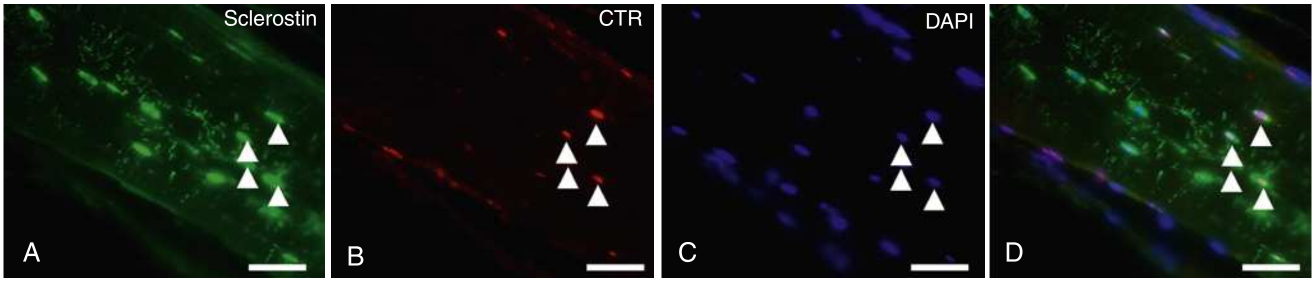


Fig. 49.5 Expression of calcitonin receptor (CTR) by mouse osteocytes. Immunohistochemical detection of CTR in osteocytes within young mouse calvariae. **A,** Green depicts sclerostin-positive osteocytes; **B,** red depicts CTR-positive osteocytes; **C,** blue depicts 4′,6-diamidino-2-phenylindole–stained nuclei in osteocytes; **D,** the merged image shows a subset of the sclerostin-positive osteocytes also positive for the CTR. (Used with permission from Gooi JH, Pompolo S, Karsdal MA, et al. Calcitonin impairs the anabolic effect of PTH in young rats and stimulates expression of sclerostin by osteocytes. *Bone.* 2010;46:1486–1497.)

been noted, including a transient increase in calcium excretion, probably caused by inhibition of renal tubular calcium reabsorption. Although this has not usually been regarded as an important effect of CT, it has been linked to the calcium-lowering effect of CT in hypercalcemic patients with metastatic bone disease. The use of CT in the treatment of hypercalcemia owing to cancer has been based exclusively on the inhibition of osteolysis by CT. However, there is evidence that failure of the kidneys to excrete the calcium load derived from bone breakdown is a major contributor to the hypercalcemia. This has prompted careful study of the relative contributions to the hypocalcemic effect of CT of its renal and skeletal components. It was concluded that inhibition of renal tubular reabsorption by CT can induce a rapid fall in serum calcium, and that the magnitude of this effect depends upon the correction of volume depletion that inevitably accompanies hypercalcemia.[110] In global CTR-deficient mice, hypercalcemia increased tubular reabsorption of calcium in the kidney,[61] whereas reabsorption was unchanged in mice in which *CTR* was deleted specifically in osteoclasts.[90] These data are consistent with the renal actions of CT to inhibit calcium reabsorption, to rapidly improve severe hypercalcemia, being mediated directly via the CTR. Thus, the calciuretic action of CT may assume greater importance than was hitherto suspected.

CTRs have been demonstrated in rat kidney,[116] and in humans the action of CT upon adenylate cyclase activity has been localized to the nephron, predominantly to the medullary and cortical portions of the thick ascending limb and to the early portion of the distal convoluted tubule. The colocalization of CTR mRNA expression and cell surface receptors with G protein–sensitive adenylate cyclase is consistent with cAMP being an important mediator of CT action in this organ. A possible role for CT in the kidney is to regulate $1,25(OH)_2D_3$ levels. The original observation was of enhanced 1-hydroxylation of 25(OH)D in the proximal straight tubule of the kidney by CT stimulation of the expression of the 25(OH)D 1α-hydroxylase.[117] Subsequently, Shinki et al.[118] showed that CT administration to rats induced renal CYP27B1 when serum calcium levels were normal or high, and this was supported by a report that CT treatment in rats increased renal production of *CYP27B1* mRNA.[119] CT stimulation of the expression of *CYP24* in CTR-transfected HEK-293 cells has also been reported.[120] The authors speculated that because $1,25(OH)_2D_3$ and CT synergistically stimulate *CYP24* mRNA production in kidney cells, this latter action of CT could be part of the process by which it regulates serum calcium, that is by controlling renal production of $1,25(OH)_2D_3$.

Calcitonin in the Central Nervous System

CTRs are widely distributed in the central nervous system.[39,40,121] Moreover, CT itself and immunoreactive CT-like peptides are produced in the central nervous system (CNS). Together, these findings suggest multiple autocrine or paracrine actions of CT and/or CT-related peptides in the brain, and effects on feeding,[122] pain sensation,[123] and regulation of pituitary prolactin secretion[124] have been documented. Intraventricular administration of CT generates potent effects that include analgesia and inhibition of appetite and gastric acid secretion, as well as modulation of hormone secretion and the extrapyramidal motor system.

Immunoreactive CT-like peptides have been demonstrated in the brain and nervous system of rats, humans, and other species.[27] Immunoreactive CT related antigenically to hCT (hCTI) has also been demonstrated in the nervous systems of protochordates, lizards, and pigeons, as well as at low levels in human and rat brain extracts. Low levels of hCTI also occur in extracts of postmortem human brain. However, in addition to hCTI, low levels of a molecular entity chromatographically and immunologically similar to sCT have been found, with concentrations in the brain approximately 10 times greater than those in serum and cerebrospinal fluid. Similarly, extracts of rat brain diencephalon contain an sCT-like peptide. Although hCTI has been detected in rat brain, only limited evidence has been found for the presence of rat CT mRNA.

Immunoreactive CT-like material also occurs in the pituitary of both mammals and lower vertebrates, and the CTR is present in the intermediate pituitary, at least in some species. CT has been described as participating in complex crosstalk between extracellular signaling molecules in the pituitary. For example, treatment of pituitary cells with CT in culture downmodulates the secretion of prolactin, whereas treatment of the cultures with an anti-CT antibody enhances it.[119] Accordingly, targeted overexpression of CT in gonadotrophs of mice leads to long-term hypoprolactinemia, decreased prolactin gene expression, female subfertility, and a selective underdevelopment of lactotrophs.

The complex effects of activating CTR in specific areas of the CNS linked to food intake are summarized elsewhere.[122] Overall, activation of the CTR by agonists in such areas leads to decreased food uptake. It has recently been demonstrated, for example, that the nucleus tractus solitarius (NTS) is crucial for integrating appetite, and that with ablation of CTR in the NTS, the long-term suppression of food intake was lost.[122] These results suggest a role for CTR in the regulation of energy balance.

The reported efficacy of CT in skeletal pain also appears to be mediated primarily in the CNS. Delivered either by injection or nasally, CT has been suggested as an adjunct treatment for acute, severe back pain secondary to fracture, as well as in Paget disease of bone and other painful conditions of the skeleton, such as osteoarthritis (see later), because of its analgesic properties. In a trial comparing elcatonin (a stabilized derivative of eel CT) with nonsteroidal anti-inflammatory (NSAID) medications for pain owing to vertebral fracture, elcatonin was reported to be more effective than NSAIDs.[123] The mechanism of CT-induced analgesia is not known, although several have been suggested, including increased secretion of β-endorphin from the pituitary gland, decreased synthesis of prostaglandins and other inflammatory factors, and modulation of serotonin transporter activity.[125]

KEY POINTS

- Calcitonin (CT) receptors are widely distributed in the brain, together with CT and CT-like peptides. This distribution is associated with multiple actions of CT and/or CT-related peptides in the brain, including effects on feeding, pain sensation, and regulation of pituitary prolactin secretion.

PEPTIDES RELATED TO CALCITONIN

CT is a member of a family of peptides, albeit with rather weak homology, that includes CGRP, AMY, CRSP, adrenomedullin (ADM), and IMD.[126] Greatest identity is between the teleost CTs and AMY (~33%). There is approximately 22% identity with the CGRPs and 16% identity with ADM. Even less homology is observed between the mammalian CTs and the other peptides. Consistent with the homology between peptides, there is a limited degree of overlap in specificity among the binding sites for the peptide receptors, with CT-like actions usually seen only at high concentrations of AMY and CGRP.

AMY is a 37–amino acid peptide that is cosecreted with insulin from pancreatic β cells following nutrient ingestion.[126] At physiological concentrations, AMY is important in the integrated control of nutrient influx, with potent actions that include inhibition of gastric emptying, gastric acid secretion, food intake, digestive enzyme secretion, and glucagon secretion. At higher concentrations, AMY also inhibits insulin secretion from the pancreas to promote glycogen breakdown and to decrease insulin-stimulated incorporation of glucose into glycogen in skeletal muscle. Thus, AMY is thought to act as a partner to insulin in metabolic regulation, although this effect may not occur at normal circulating levels of the peptide. In the kidney, AMY is proposed to have a diverse range of actions, including modulation of calcium excretion and thiazide receptor levels, proliferative effects on tubule epithelium, and increasing renin activity. *AMY*$^{-/-}$ mice have less bone, as the result of increased bone resorption, and *in vitro* tests indicated that this may be because of the release of an AMY-mediated attenuator of osteoclastogenesis. AMY receptors are widely expressed in brain, where administered peptide induces many potent effects. These include decreased appetite and gastric acid secretion, hyperthermia, adipsia, and reduction in growth hormone-releasing hormone. Central AMY injection may also modulate memory and the extrapyramidal motor system. The molecular nature of the AMY receptor is discussed later.

CGRP is a neuropeptide with a diverse range of actions, including potent dilation of vascular beds, as well as relaxation of other smooth muscle, inotropy and chronotropy in the heart, and paracrine regulation of pituitary hormone release, and many central effects, such as suppression of appetite and gastric acid secretion, modulation of body temperature, and modulation of sympathetic outflow. The actions of CGRP have been extensively reviewed elsewhere.[127] CGRP exists in two isoforms in humans: the main isoform present centrally is α-CGRP, whereas the β-CGRP isoform mainly functions in the periphery.[127] CGRP has been implicated in migraine, and a number of therapeutic agents have been trialed for efficacy in migraine treatment, including neutralizing CGRP antibodies and receptor antagonists. It was recently reported that salmon CT ameliorated migraine in a rat model by regulating central CGRP release from multiple cell types.[128]

ADM was originally isolated from human pheochromocytoma and is abundant in the normal adrenal medulla, hence its name. The full-length peptide of 52 amino acids shares approximately 25% homology with CGRP across its N-terminal 37 amino acids. ADM is a potent dilator of many vascular beds and is protective against conditions such as cardiac hypertrophy, perivascular fibrosis, renal damage, and pulmonary hypertension.[129]

As alluded to earlier, additional peptide members of the CT family have been identified. The CRSPs were originally identified from the porcine hypothalamus and being localized principally to brain and pituitary, have been proposed as potential endogenous ligands for central CTRs. However, not all CRSPs activate the CTR or related receptors, suggesting that they may also act at other receptors. A second ADM-like peptide (ADM 2 or IMD) has been identified in multiple species, including in humans. IMD can exert several actions similar to CGRP and ADM, including decreasing blood pressure, effects on cardiac function, and suppression of gastric emptying and food intake. IMD has also been found to cause renal effects similar to ADM. CRSP and IMD can inhibit bone resorption,[97] although their physiological actions remain to be identified.

CALCITONIN AND ITS RECEPTORS IN CANCER

Medullary Carcinoma of the Thyroid

The classic syndrome of CT excess is medullary carcinoma of the thyroid (MCT).[130] This tumor clearly differs in origin from all other thyroid cancers because it is a tumor of the CT-producing C cells. MCT may occur as a sporadic or a genetic (familial) tumor. The only effective treatment for MCT is surgical. The earlier this can be performed in the course of disease, the better the likelihood of a cure. Once the tumor becomes palpable, any treatment is unlikely to be curative. Because malignant C cells continue to produce CT, raised blood levels of CT provide an accurate diagnostic test for MCT, and continued raised levels after surgery suggest metastatic disease.[130] In rare cases, MCT is not accompanied by elevated serum CT. Genetic testing for the locus of mutation in the responsible gene (the *RET* protooncogene) often has considerable clinical relevance in management, as early surgery is key to survival, especially in the syndrome known as multiple endocrine neoplasia type 2b.

Calcitonin Receptors in Other Cancers

As stated, specific high-affinity receptors for CT have been demonstrated in human lymphoid cell lines, in a human lung cancer cell line, in several human breast cancer cell lines, including MCF-7, T47D, and ZR-75, in glioblastoma multiforme, and in PC cells. Indeed, CTR mRNA appears to be a frequent accompaniment of human tumors. The first clinical study of CTR in surgically obtained human breast cancers identified receptor mRNA production in all of 18 cancers, and receptor expression was confined to tumor cells in the sections.[131] In contrast, a subsequent study using laser capture microdissection and PCR found that the CTR in breast cancer tissue was actually reduced compared with adjacent normal breast tissue.[132]

The CT-CTR signaling pathway is implicated in human PC. Although CT is expressed endogenously in the prostate,[133] PCs display increased expression of CT and CTR compared with benign prostate tissue. Tumors with higher CT/CTR expression displayed metastatic disease and poor clinical outcome.[134] Forced expression of human CT in LNCaP human PC cells dramatically enhanced their oncogenic characteristics.[135] Moreover, CT promoted progression of PC cells from a localized cancer to an aggressive form by inducing a stem cell–like phenotype,[136] as well as increasing the concentration and activity of metalloproteinases that promote invasion, MMP-2 and MMP-9. A canonical

PDZ-binding site in the CTR cytoplasmic tail was shown to play a critical role in the CTR-mediated increase in invasiveness of PC cells and in the formation of distant metastases by orthotopically implanted PC cells in immunodeficient mice (Fig. 49.6).[137] Studies *in vitro* and *in vivo*, together with epidemiological data reporting an increased risk of cancer with CT use in osteoporosis patients, compared with those treated with bisphosphonates, prompted the European Medicines Agency to warn against the use of CT for this condition (see later).

KEY POINTS

- Calcitonin (CT) is important in a number of malignancies. The hallmark of medullary carcinoma of the thyroid is high circulating CT levels. The effects of CT on cell growth assume pathophysiological importance in some cancers, such as prostate cancer, with CT signaling inducing progression of primary tumors and metastatic invasiveness.

CALCITONIN AS A THERAPEUTIC

Hypercalcemia

CT has been used in the treatment of hypercalcemia in many different conditions, especially malignancy, but its effective use is limited by rapid tachyphylaxis, often within 24 hours of use. The development of resistance to this action of CT seems to involve downregulation of its receptor, which to some extent is blocked by corticosteroid use.[69]

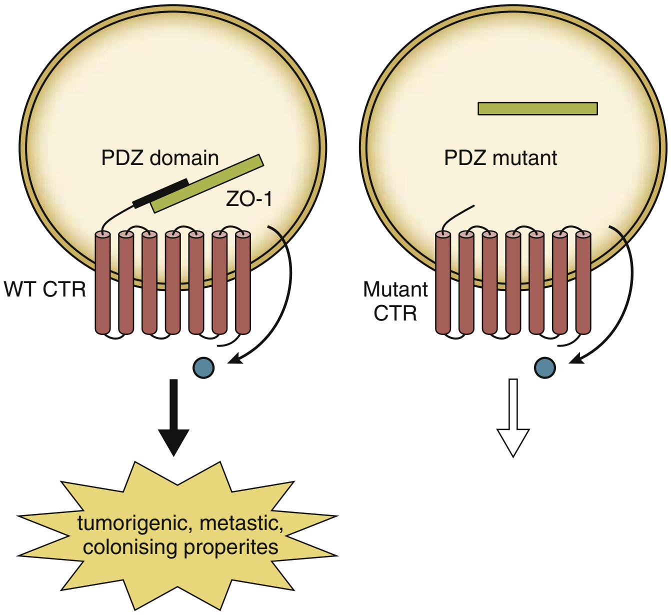


Fig. 49.6 **Calcitonin receptor (*CTR*)-mediated induction of stem-like cell phenotype of human prostate cancer (PC) cells.** Forced expression of wild type (*WT*) CTR in PC-3 and LNCaP PC-like cells induces tumorigenic, metastatic, and repopulation properties. CTR actions are dependent on the cytoplasmic tail PDZ domain, which interacts with the zonula occludens protein-1 (ZO-1). (From Chien J, Ren Y, Qing Wang Y, et al. Calcitonin is a prostate epithelium-derived growth stimulatory peptide. *Mol Cell Endocrinol.* 2001;181:69–79; Thakkar A, Bijnsdorp IV, Geldof AA, et al. Profiling of the calcitonin-calcitonin receptor axis in primary prostate cancer: clinical implications and molecular correlates. Oncol Rep. 2013;30:1265–1274; Thomas S, Chigurupati S, Anbalagan M, et al. Calcitonin increases tumorigenicity of prostate cancer cells: evidence for the role of protein kinase A and urokinase-type plasminogen receptor. *Mol Endocrinol.* 2006;20:1894–1911; Aldahish A, Kale A, Aljameeli A, et al. Calcitonin induces stem cell-like phenotype in prostate cancer cells. Endocr Relat Cancer. 2019;26:815–828; Aljameeli A, Thakkar A, Thomas S, et al. Calcitonin receptor-zonula occludens-1 interaction is critical for calcitonin-stimulated prostate cancer metastasis. *PLOS ONE.* 2016;e0150090.133–137.)

Osteoporosis and Paget Disease of Bone

The antiresorptive action of CT has been the basis for its use clinically in osteoporosis, with an additional reported advantage being its analgesic effect on pain because of vertebral fracture.[138] CT treatment has modest effects on bone quantity (BMD) or bone remodeling markers, although there is reported improvement in bone quality parameters after administration of salmon CT nasal spray (CT-NS) over 2 years. The results of the QUEST study, in which effects of CT on bone structure were assessed by high-resolution magnetic resonance imaging (MRI), suggest therapeutic benefit of CT-NS compared with placebo in maintaining trabecular microarchitecture at multiple skeletal sites.[139] Similar protection by CT of bone structure was obtained in ovariectomized sheep, again with MRI used to assess bone changes. High doses of eel CT were effective in protecting bone architecture and strength in ovariectomized rats, and additional effectiveness was reported in combination with alendronate. A problem with CT use clinically has been its bioavailability, and to address this, both orally and nasally administered formulations of CT have been developed.[140]

CT given by subcutaneous injection was recognized as effective in the treatment of Paget disease, with relief of pain, return of biochemical indices of disease activity toward normal, formation of normal lamellar bone, and filling in of osteolytic lesions.[141] In long-term treatment of Paget disease with CT, the number of osteoclasts declines progressively, either because of a direct effect on osteoclast precursors or as an indirect consequence of the acute inhibition of osteoclast function by CT. With respect to an effect of CT on osteoclast development, it was reported that eel CT can protect from the bone loss of skeletal unloading by inhibiting the fusion of TRAP$^+$ mononuclear preosteoclasts.[142] This experiment was performed in 7-week-old mice, and its implications for adult humans are not clear. Skeletal blood flow (SBF) in normal male volunteers is approximately 5% of blood volume/min or about 4 mL/min/100 g of bone. In untreated Paget disease of bone, there is a striking increase in SBF at affected sites. Intriguingly, treatment with CT was shown to bring a rapid return to normal perfusion rates,[143] even before a detectable biochemical response. This response was similarly observed with bisphosphonate treatment and was postulated to explain the rapid pain relief obtained after commencement of treatment. The use of CT in Paget disease has been largely supplanted by bisphosphonates.

Osteoarthritis

A large number of *in vitro* and *in vivo* animal studies[144] have suggested the potential benefit of CT therapy in osteoarthritis. CT was found to attenuate indices of progression of osteoarthritis in several commonly affected joints[144] and to reduce the pain of osteoarthritis.[145] Studies designed to elucidate the mechanism of joint protection by CT found that it reduced IL1β-induced catabolism in chondrocyte cultures, as well as preventing cartilage degradation and subchondral bone changes in an *in vivo* rat model of osteoarthritis.[144] Potential benefits of CT have also been suggested for inflammatory arthritis, as it was found to preserve bone morphology in a rat model of rheumatoid arthritis, particularly when used with prednisolone. Thus a study in which CT and hyaluronic acid were formulated together in nanocomplexes and injected intraarticularly into the knee in a mouse model of inflammatory arthritis showed an apparently synergistic effect of the two agents in reducing markers of the disease.[146] Although studies such as those cited here claim that CT has direct effects on articular chondrocytes via the CTR, other investigators have been unable to find CTR expression or responsiveness to CT, at least in human chondrocytes.[147] The latter findings align with the difficulty of demonstrating a consistent effect of CT treatment on osteoarthritis in humans. Any benefits of CT in osteoarthritis could result from the effects of CT on subchondral bone, such as have been described for bisphosphonate treatment of animal models of osteoarthritis. However, like CT, these effects have been difficult to observe in human studies.

In 2012, the European Medicines Agency (EMA) recommended against the long-term therapeutic use of CT.[148] A possible causal link between long-term CT administration and the progression of PC was the given rationale for the recommendation. The EMA concluded that the benefit-risk balance remains positive only for treatment of Paget disease, immobilization-induced bone loss, and cancer-induced hypercalcemia in which patients cannot be treated with alternative agents. A similar recommendation was brought the following year by the US Food and Drug Administration, on the basis of a lack of benefit in the face of concerns about a cancer risk.[149]

KEY POINTS

- Calcitonin (CT) has been used in diseases of bone, primarily osteoporosis and Paget disease of bone, where its target is the inhibition of osteoclast resorption. The suggested and controversial efficacy of CT in osteoarthritis is now moot, given the recommendation by several drug regulatory agencies against its long-term therapeutic use.

CONCLUSION

A large number of actions and potential actions of CT have been discussed in this chapter (Fig 49.7), reflecting the wide tissue location of the CTR. Clearly, CT has both local paracrine actions and hormonal actions, and the significance of these appears to vary across life and is context-dependent, for example in reproduction, and sometimes species-specific. Although the clinical use of CT has declined in the last several decades, recent research has brought interesting discoveries that show the value of continuing to pursue the physiology and pathophysiology of this fascinating molecule.

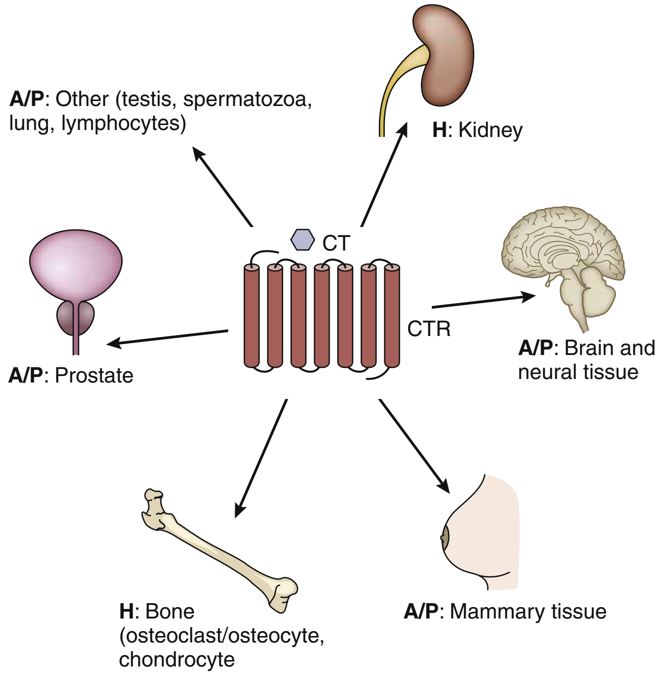

A/P: Other (testis, spermatozoa, lung, lymphocytes)
H: Kidney
CT
CTR
A/P: Brain and neural tissue
A/P: Prostate
H: Bone (osteoclast/osteocyte, chondrocyte
A/P: Mammary tissue

Fig. 49.7 **Physiology of calcitonin (CT).** Cartoon shows putative target tissues of either hormonal (H) or autocrine/paracrine (A/P) CT or CT-like peptides acting at the CT receptor (CTR).

Acknowledgments

Work from each of the authors' laboratories was supported by the National Health and Medical Research Council of Australia.

REFERENCES

1. Munson PL. Parathyroid hormone and calcitonin. In: McCann SM, ed. *Endocrinology: People and Ideas*. New York: Springer; 1988:239–284.
2. Copp DH, Cameron EC, Cheney BA, et al. Evidence for calcitonin- a new hormone from the parathyroid that lowers blood calcium. *Endocrinology*. 1962;70:638–649.
3. Hirsch PF, Voelkel EF, Munson PL. Thyrocalcitonin: hypocalcemic hypophosphatemic principle of the thyroid gland. *Science*. 1964;146:412–413.
4. Foster GV, Baghdiantz A, Kumar MA, et al. Thyroid origin of calcitonin. *Nature*. 1964;202:1303–1305.
5. Pearse AG, Carvalheira AF. Cytochemical evidence for an ultimobranchial origin of rodent thyroid C cells. *Nature*. 1967;214:929–930.
6. Suzuki N, Suzuki T, Kurokawa T. Suppression of osteoclastic activities by calcitonin in the scales of goldfish (freshwater teleost) and nibbler fish (seawater teleost). *Peptides*. 2000;21:115–124.
7. Fudge NJ, Kovacs CS. Physiological studies in heterozygous calcium sensing receptor (CaSR) gene-ablated mice confirm that the CaSR regulates calcitonin release in vivo. *BMC Physiol*. 2004;4:5.
8. Mun HC, Leach KM, Conigrave AD. L-amino acids promote calcitonin release via a calcium-sensing receptor: Gq/11-mediated pathway in human c-cells. *Endocrinology*. 2019;160:1590–1599.
9. Davey RA, Findlay DM. Calcitonin: physiology or fantasy? *J Bone Miner Res*. 2013;28:973–979.
10. Rong H, Sääf M, Tørring O, et al. Circulating monomer-like calcitonin in osteoporotic patients. *Osteoporos Int*. 1996;6:394–398.
11. Melvin KE, Tashjian Jr AH. The syndrome of excessive thyrocalcitonin produced by medullary carcinoma of the thyroid. *Proc Natl Acad Sci U S A*. 1968;59:1216–1222.
12. Canale DD, Donabedian RK. Hypercalcitoninemia in acute pancreatitis. *J Clin Endocrinol Metab*. 1975;40:738–741.
13. Becker KL, Nylen ES, Snider RH, et al. Immunoneutralization of procalcitonin as therapy of sepsis. *J Endotoxin Res*. 2003;9:367–374.
14. Cotoi OS, Manjer J, Hedblad B, et al. Plasma procalcitonin is associated with all-cause and cancer mortality in apparently healthy men: a prospective population-based study. *BMC Med*. 2013;11:180.
15. Martin TJ, Findlay DM, Moseley JM, et al. Calcitonin. In: Avioli LV, Krane SM, eds. *Metabolic Bone Disease and Clinically Related Disorders*. 3rd ed. St. Louis: Academic Press; 1998:95–121.
16. Yamauchi H, Shiraki M, Otani M, et al. Stability of [Asu1,7]-eel calcitonin and eel calcitonin in vitro and in vivo. *Endocrinol Jpn*. 1977;24:281–285.
17. Houssami S, Findlay DM, Brady CL, et al. Divergent structural requirements exist for calcitonin receptor binding specificity and adenylate cyclase activation. *Mol Pharmacol*. 1995;47:798–809.
18. Jacobs JW, Goodman RH, Chin WW, et al. Calcitonin messenger RNA encodes multiple polypeptides in a single precursor. *Science*. 1981;213:457–459.
19. Rosenfeld MG, Amara SG, Evans RM. Alternative RNA processing: determining neuronal phenotype. *Science*. 1984;225:1315–1320.
20. Gorn AH, Lin HY, Yamin M, et al. Cloning, characterization, and expression of a human calcitonin receptor from an ovarian carcinoma cell line. *J Clin Invest*. 1992;90:1726–1735.
21. Lin HY, Harris TL, Flannery MS, et al. Expression cloning of an adenylate cyclase-coupled calcitonin receptor. *Science*. 1991;254:1022–1024.
22. Nag K, Kato A, Sultana N, et al. Fish receptor has novel features. *Gen Comp Endocrinol*. 2007;154:48–58.
23. Anusaksathien O, Laplace C, Li X, et al. Tissue-specific and ubiquitous promoters direct the expression of alternatively spliced transcripts from the calcitonin receptor gene. *J Biol Chem*. 2001;276:22663–22674.

24. Shen Z, Crotti TN, Flannery MR, et al. A novel promoter regulates calcitonin receptor gene expression in human osteoclasts. *Biochim Biophys Acta*. 2007;1769:659–667.
25. Hoshiya H, Meguro M, Kashiwagi A, et al. Calcr, a brain-specific imprinted mouse calcitonin receptor gene in the imprinted cluster of the proximal region of chromosome 6. *J Hum Genet*. 2003;48:208–211.
26. Moore EE, Kuestner RE, Stroop SD, et al. Functionally different isoforms of the human calcitonin receptor result from alternative splicing of the gene transcript. *Mol Endocrinol*. 1995;9:959–968.
27. Sexton PM, Houssami S, Hilton JM, et al. Identification of brain isoforms of the rat calcitonin receptor. *Mol Endocrinol*. 1993;7:815–821.
28. Shyu JF, Inoue D, Baron R, et al. The deletion of 14 amino acids in the seventh transmembrane domain of a naturally occurring calcitonin receptor isoform alters ligand binding and selectively abolishes coupling to phospholipase C. *J Biol Chem*. 1996;271:31127–31134.
29. Houssami S, Findlay DM, Brady CL, et al. Divergent structural requirements exist for calcitonin receptor binding specificity and adenylate cyclase activation. *Mol Pharmacol*. 1995;47:798–809.
30. Qi T, Dong M, Watkins HA, et al. Receptor activity-modifying protein-dependent impairment of calcitonin receptor splice variant Δ(1-47) hCT((a)) function. *Br J Pharmacol*. 2013;168:644–657.
31. Seck T, Baron R, Horne WC. The alternatively spliced deltae13 transcript of the rabbit calcitonin receptor dimerizes with the C1a isoform and inhibits its surface expression. *J Biol Chem*. 2003;278:23085–23093.
32. Beaudreuil J, Taboulet J, Orcel P, et al. Calcitonin receptor mRNA in mononuclear leucocytes from post-menopausal women: decrease during osteoporosis and link to bone markers with specific isoform involvement. *Bone*. 2000;27:161–168.
33. Nakamura M, Zhang ZQ, Shan L, et al. Allelic variants of human calcitonin receptor in the Japanese population. *Hum Genet*. 1997;99:38–41.
34. Braga V, Mottes M, Mirandola S, et al. Association of CTR and COLIA1 alleles with BMD values in peri- and postmenopausal women. *Calcif Tissue Int*. 2000;67:361–366.
35. Wolfe 3rd LA, Fling ME, Xue Z, et al. In vitro characterization of a human calcitonin receptor gene polymorphism. *Mutat Res*. 2003;522:93–105.
36. Abdelmagid N, Bereczky-Veress B, Guerreiro-Cacais AO, et al. The calcitonin receptor gene is a candidate for regulation of susceptibility to herpes simplex type 1 neuronal infection leading to encephalitis in rat. *PLoS Pathog*. 2012;8:e1002753.
37. Nicholson GC, Moseley JM, Sexton PM, et al. Abundant calcitonin receptors in isolated rat osteoclasts. Biochemical and autoradiographic characterization. *J Clin Invest*. 1986;78:355–360.
38. Sexton PM, Findlay DM, Martin TJ. Calcitonin. *Curr Med Chem*. 1999;6:1067–1093.
39. Sexton PM. Central nervous system binding sites for calcitonin and calcitonin gene-related peptide. *Mol Neurobiol*. 1991;5:251–273.
40. Paxinos G, Chai SY, Christopoulos G, et al. In vitro autoradiographic localization of calcitonin and amylin binding sites in monkey brain. *J Chem Neuroanat*. 2004;27:217–236.
41. Gooi JH, Pompolo S, Karsdal MA, et al. Calcitonin impairs the anabolic effect of PTH in young rats and stimulates expression of sclerostin by osteocytes. *Bone*. 2010;46:1486–1497.
42. Hay DL, Garelja ML, Poyner DR, et al. Update on the pharmacology of calcitonin/CGRP family of peptides: IUPHAR Review 25. *Br J Pharmacol*. 2018;175:3–17.
43. Pham V, Zhu Y, Dal Maso E, et al. Deconvoluting the molecular control of binding and signaling at the amylin 3 receptor: RAMP3 alters signal propagation through extracellular loops of the calcitonin receptor. *ACS Pharmacol Transl Sci*. 2019;2:183–197.
44. Christopoulos G, Perry KJ, Morfis M, et al. Multiple amylin receptors arise from receptor activity-modifying protein interaction with the calcitonin receptor gene product. *Mol Pharmacol*. 1999;56:235–242.
45. Morfis M, Christopoulos A, Sexton PM. RAMPs: 5 years on, where to now? *Trends Pharmacol Sci*. 2003;24:596–601.
46. Hay DL, Christopoulos G, Christopoulos A, et al. Pharmacological discrimination of calcitonin receptor: receptor activity-modifying protein complexes. *Mol Pharmacol*. 2005;67:1655–1665.
47. Suzuki H, Nakamura I, Takahashi N, et al. Calcitonin-induced changes in the cytoskeleton are mediated by a signal pathway associated with protein kinase A in osteoclasts. *Endocrinology*. 1996;137:4685–4690.
48. Granholm S, Lundberg P, Lerner UH. Calcitonin inhibits osteoclast formation in mouse haematopoietic cells independently of transcriptional regulation by receptor activator of NF-κB and c-Fms. *J Endocrinol*. 2007;195:415–427.
49. Offermanns S, Iida-Klein A, Segre GV, et al. G alpha q family members couple parathyroid hormone (PTH)/PTH-related peptide and calcitonin receptors to phospholipase C in COS-7 cells. *Mol Endocrinol*. 1996;10:566–574.
50. Kiriyama Y, Tsuchiya H, Murakami T, et al. Calcitonin induces IL-6 production via both PKA and PKC pathways in the pituitary folliculo-stellate cell line. *Endocrinology*. 2001;142:3563–3569.
51. Yamaguchi M. Stimulatory effect of calcitonin on Ca2+ inflow in isolated rat hepatocytes. *Mol Cell Endocrinol*. 1991;75:65–70.
52. Wang J, Rout UK, Bagchi IC, et al. Expression of calcitonin receptors in mouse preimplantation embryos and their function in the regulation of blastocyst differentiation by calcitonin. *Development*. 1998;125:4293–4302.
53. Kajiya H, Okamoto F, Fukushima H, et al. Calcitonin inhibits proton extrusion in resorbing rat osteoclasts via protein kinase A. *Pflugers Arch*. 2003;445:651–658.
54. Santhanagopal A, Chidiac P, Horne WC, et al. Calcitonin (CT) rapidly increases Na(+)/H(+) exchange and metabolic acid production: effects mediated selectively by the C1A CT receptor isoform. *Endocrinology*. 2001;142:4401–4413.
55. Chakraborty M, Chatterjee D, Kellokumpu S, et al. Cell cycle-dependent coupling of the calcitonin receptor to different G proteins. *Science*. 1991;251:1078–1082.
56. Stroop SD, Thompson DL, Kuestner RE, et al. A recombinant human calcitonin receptor functions as an extracellular calcium sensor. *J Biol Chem*. 1993;268:19927–19930.
57. Malgaroli A, Meldolesi J, Zallone AZ, et al. Control of cytosolic free calcium in rat and chicken osteoclasts. The role of extracellular calcium and calcitonin. *J Biol Chem*. 1989;264:14342–14347.
58. Raggatt LJ, Evdokiou A, Findlay DM. Sustained activation of Erk1/2 MAPK and cell growth suppression by the insert-negative, but not the insert-positive isoform of the human calcitonin receptor. *J Endocrinol*. 2000;167:93–105.
59. Zhang Z, Baron R, Horne WC. Integrin engagement, the actin cytoskeleton, and c-Src are required for the calcitonin-induced tyrosine phosphorylation of paxillin and HEF1, but not for calcitonin-induced Erk1/2 phosphorylation. *J Biol Chem*. 2000;275:37219–37223.
60. Findlay DM, Sexton PM. *Calcitonin, growth factors*. 2004;22:217–224.
61. Davey RA, Turner AG, McManus JF, et al. The calcitonin receptor plays a physiological role to protect against hypercalcemia in mice. *J Bone Miner Res*. 2008;23:1182–1193.
62. Okumura S, Mizoguchi T, Sato N, et al. Coordination of microtubules and the actin cytoskeleton is important in osteoclast function, but calcitonin disrupts sealing zones without affecting microtubule networks. *Bone*. 2006;39:684–693.
63. Shyu JF, Shih C, Tseng CY, et al. Calcitonin induces podosome disassembly and detachment of osteoclasts by modulating Pyk2 and Src activities. *Bone*. 2007;40:1329–1342.
64. Findlay DM, Raggatt LJ, Bouralexis S, et al. Calcitonin decreases the adherence and survival of HEK-293 cells by a caspase-independent mechanism. *J Endocrinol*. 2002;175:715–725.
65. Ikegame M, Ejiri S, Ozawa H. Calcitonin-induced change in serum calcium levels and its relationship to osteoclast morphology and number of calcitonin receptors. *Bone*. 2004;35:27–33.
66. Wada S, Martin TJ, Findlay DM. Homologous regulation of the calcitonin receptor in mouse osteoclast-like cells and human breast cancer T47D cells. *Endocrinology*. 1995;136:2611–2621.
67. Yasuda S, Wada S, Arao Y, et al. Interaction between 3′ untranslated region of calcitonin receptor messenger ribonucleic acid (RNA) and adenylate/uridylate (AU)-rich element binding proteins (AU-rich RNA-binding factor 1 and Hu antigen R). *Endocrinology*. 2004;145:1730–1738.

68. Wada S, Udagawa N, Akatsu T, et al. Regulation by calcitonin and glucocorticoids of calcitonin receptor gene expression in mouse osteoclasts. *Endocrinology*. 1997;138:521–529.
69. Binstock ML, Mundy GR. Effect of calcitonin and glutocorticoids in combination on the hypercalcemia of malignancy. *Ann Intern Med*. 1980;93:269–272.
70. Hirsch PF, Lester GE, Talmage RV. Calcitonin, an enigmatic hormone: does it have a function? *J Musculoskel Neuron Interact*. 2001;1:299–305.
71. Hirsch PF, Baruch H. Is calcitonin an important physiological substance? *Endocrine*. 2003;21:201–208.
72. Shah GV, Rayford W, Noble MJ, et al. Calcitonin stimulates growth of human prostate cancer cells through receptor-mediated increase in cyclic adenosine 3′,5′-monophosphates and cytoplasmic Ca2+ transients. *Endocrinology*. 1994;134:596–602.
73. Thomas S, Muralidharan A, Shah GV. Knock-down of calcitonin receptor expression induces apoptosis and growth arrest of prostate cancer cells. *Int J Oncol*. 2007;31:1425–1437.
74. Ng KW, Livesey SA, Larkins RG, et al. Calcitonin effects on growth and on selective activation of type ii isoenzyme of cyclic adenosine 3′:5′-monophosphate-dependent protein kinase in T 47D human breast cancer cells. *Cancer Res*. 1983;43:794–800.
75. Evdokiou A, Raggatt LJ, Atkins GJ, et al. Calcitonin receptor-mediated growth suppression of HEK-293 cells is accompanied by induction of p21WAF1/CIP1 and G2/M arrest. *Mol Endocrinol*. 1999;13:1738–1750.
76. Raggatt LJ, Evdokiou A, Findlay DM. Sustained activation of Erk1/2 MAPK and cell growth suppression by the insert-negative, but not the insert-positive isoform of the human calcitonin receptor. *J Endocrinol*. 2000;167:93–105.
77. Wang FW, Zhang YM, Wang Z, et al. Calcitonin promotes mouse preimplantation development: involvement of calcium mobilization and P38 mitogen-activated protein kinase activation. *Reprod Domest Anim*. 2013;48:382–389.
78. Zhu LJ, Cullinan-Bove K, Polihronis M, et al. Calcitonin is a progesterone-regulated marker that forecasts the receptive state of endometrium during implantation. *Endocrinology*. 1998;139:3923–3934.
79. Xiong T, Zhao Y, Hu D, et al. Administration of calcitonin promotes blastocyst implantation in mice by up-regulating integrin β3 expression in endometrial epithelial cells. *Hum Reprod*. 2012;27:3540–3551.
80. Li HY, Shen JT, Chang SP, et al. Calcitonin promotes outgrowth of trophoblast cells on endometrial epithelial cells: involvement of calcium mobilization and protein kinase C activation. *Placenta*. 2008;29:20–29.
81. Dorostghoal M, Ghaffari HO, Shahbazian N, et al. Endometrial expression of β3 integrin, calcitonin and plexin-B1 in the window of implantation in women with unexplained infertility. *Int J Reprod Biomed (Yazd)*. 2017;15:3–40.
82. Shokrzadeh N, Alivand MR, Abedelahi A, et al. Calcitonin administration improves endometrial receptivity via regulation of LIF, Muc-1 and microRNA Let-7a in mice. *J Cell Physiol*. 2019;234:12989–13000.
83. Moghani-Ghoroghi F, Moshkdanian G, Sehat M, et al. Melatonin pretreated blastocysts along with calcitonin administration improved implantation by upregulation of heparin binding-epidermal growth factor expression in murine endometrium. *Cell J*. 2018;19:599–606.
84. Dacquin R, Davey RA, Laplace C, et al. Amylin inhibits bone resorption while the calcitonin receptor controls bone formation in vivo. *J Cell Biol*. 2004;164:509–514.
85. Jagger C, Chambers T, Pondel M. Transgenic mice reveal novel sites of calcitonin receptor gene expression during development. *Biochem Biophys Res Commun*. 2000;274:124–129.
86. Ismail PM, DeMayo FJ, Amato P, et al. Progesterone induction of calcitonin expression in the murine mammary gland. *J Endocrinol*. 2004;180:287–295.
87. Plotkin LI, Weinstein RS, Parfitt AM, et al. Prevention of osteocyte and osteoblast apoptosis by bisphosphonates and calcitonin. *J Clin Invest*. 1999;104:1363–1374.
88. Asrari M, Lobner D. Calcitonin potentiates oxygen-glucose deprivation-induced neuronal death. *Exp Neurol*. 2001;167:183–188.
89. Williams GA, Hargis GK, Galloway WB, et al. Evidence for thyrocalcitonin in man. *Proc Soc Exp Biol Med*. 1966;122:1273–1276.
90. Turner AG, Tjahyono F, Chiu WS, et al. The role of the calcitonin receptor in protecting against induced hypercalcemia is mediated via its actions in osteoclasts to inhibit bone resorption. *Bone*. 2011;48:354–361.
91. Harper C, Toverud SU. Ability of thyrocalcitonin to protect against hypercalcemia in adult rats. *Endocrinology*. 1973;93:1354–1359.
92. Martin TJ, Robinson CJ, MacIntyre I. The mode of action of thyrocalcitonin. *Lancet*. 1966;1:900–902.
93. Gonen E, Sahin I, Ozbek M, et al. Effects of pregnancy and lactation on bone mineral density, and their relation to the serum calcium, phosphorus, calcitonin and parathyroid hormone levels in rats. *J Endocrinol Invest*. 2005;28:322–326.
94. Kovacs CS. Calcium and bone metabolism during pregnancy and lactation. *J Mammary Gland Biol Neoplasia*. 2005;10:105–118.
95. Yamamoto M, Seedor JG, Rodan GA, et al. Endogenous calcitonin attenuates parathyroid hormone-induced cancellous bone loss in the rat. *Endocrinology*. 1995;136:788–795.
96. Woodrow JP, Sharpe CJ, Fudge NJ, et al. Calcitonin plays a critical role in regulating skeletal mineral metabolism during lactation. *Endocrinology*. 2006;147:4010–4021.
97. Clarke MV, Russell PK, Findlay DM, et al. A role for the calcitonin receptor to limit bone loss during lactation in female mice by inhibiting osteocytic osteolysis. *Endocrinology*. 2015;156:3203–3214.
98. Qing H, Ardeshirpour L, Divieti Pajevic P, et al. Demonstration of osteocytic perilacunar/canalicular remodeling in mice during lactation. *J Bone Miner Res*. 2012;27:1018–1029.
99. Hoff AO, Catala-Lehnen P, Thomas PM, et al. Increased bone mass is an unexpected phenotype associated with deletion of the calcitonin gene. *J Clin Invest*. 2002;110:1849–1857.
99a. Davey RA, Clarke MV, Golub SB, et al. The calcitonin receptor regulates osteocyte lacunae acidity during lactation in mice. *J Endocrinol*. 2021;249:31–41.
100. Granholm S, Henning P, Lerner UH. Comparisons between the effects of calcitonin receptor-stimulating peptide and intermedin and other peptides in the calcitonin family on bone resorption and osteoclastogenesis. *J Cell Biochem*. 2011;112:3300–3312.
101. Friedman J, Raisz LG. Thyrocalcitonin: inhibitor of bone resorption in tissue culture. *Science*. 1965;150:1465–1467.
102. Chambers TJ, Athanasou NA, Fuller K. Effect of parathyroid hormone and calcitonin on the cytoplasmic spreading of isolated osteoclasts. *J Endocrinol*. 1984;102:281–286.
103. Matthews JL, Martin JH. Immediate changes in the ultrastructure of bone cells following thyrocalcitonin administration. In: Talmage RV, Munson PL, eds. *Calcium, Parathyroid Hormone and the Calcitonins*. Amsterdam: Excerpta Medica; 1972:375–382.
104. Robinson CJ, Martin TJ, Matthews EW, et al. Mode of action of thyrocalcitonin. *J Endocrinol*. 1967;39:71–79.
105. Ikegame M, Rakopoulos M, Martin TJ, et al. Effects of continuous calcitonin treatment on osteoclast-like cell development and calcitonin receptor expression in mouse bone marrow cultures. *J Bone Miner Res*. 1996;11:456–465.
106. Farley J, Dimai HP, Stilt-Coffing B, et al. Calcitonin increases the concentration of insulin-like growth factors in serum-free cultures of human osteoblast-line cells. *Calcif Tissue Int*. 2000;67:247–254.
107. Woodrow JP, Sharpe CJ, Fudge NJ, et al. Calcitonin plays a critical role in regulating skeletal mineral metabolism during lactation. *Endocrinology*. 2006;147:4010–4021.
108. Schinke T, Liese S, Priemel M, et al. Decreased bone formation and osteopenia in mice lacking alpha-calcitonin gene-related peptide. *J Bone Miner Res*. 2004;19:2049–2056.
109. Huebner AK, Schinke T, Priemel M, et al. Calcitonin deficiency in mice progressively results in high bone turnover. *J Bone Miner Res*. 2006;21:1924–1934.
110. Huebner AK, Keller J, Catala-Lehnen P, et al. The role of calcitonin and alpha-calcitonin gene-related peptide in bone formation. *Arch Biochem Biophys*. 2008;473:210–217.
111. Gooi JH, Chia LY, Walsh NC, et al. Decline in calcitonin receptor expression in osteocytes with age. *J Endocrinol*. 2014;221:181–191.

112. Keller J, Catala-Lehnen P, Huebner AK, et al. Calcitonin controls bone formation by inhibiting the release of sphingosine 1-phosphate from osteoclasts. *Nat Commun*. 2014;5:5215.
113. Ryu J, Kim HJ, Chang EJ, et al. Sphingosine 1-phosphate as a regulator of osteoclast differentiation osteoclast-osteoblast coupling. *Embo J*. 2006;25:5840–5851.
114. Martin TJ, Sims NA. Calcitonin physiology, saved by a lysophospholipid. *J Bone Miner Res*. 2015;30:212–215.
115. Hosking DJ, Gilson D. Comparison of the renal and skeletal actions of calcitonin in the treatment of severe hypercalcaemia of malignancy. *Q J Med*. 1984;53:359–368.
116. Sexton PM, Adam WR, Moseley JM, et al. Localization and characterization of renal calcitonin receptors by in vitro autoradiography. *Kidney Int*. 1987;32:862–868.
117. Kawashima H, Torikai S, Kurokawa K. Calcitonin selectively stimulates 25-hydroxyvitamin D_3-1 alpha-hydroxylase in proximal straight tubule of rat kidney. *Nature*. 1981;291:327–329.
118. Shinki T, Ueno Y, DeLuca HF, et al. Calcitonin is a major regulator for the expression of renal 25-hydroxyvitamin D_3-1alpha-hydroxylase gene in normocalcemic rats. *Proc Natl Acad Sci U S A*. 1999;96: 8253–8258.
119. Murayama A, Takeyama K, Kitanaka S, et al. Positive and negative regulations of the renal 25-hydroxyvitamin D_3 1alpha-hydroxylase gene by parathyroid hormone, calcitonin, and 1alpha, 25(OH)2D3 in intact animals. *Endocrinology*. 1999;140:2224–2231.
120. Gao XH, Dwivedi PP, Omdahl JL, et al. Calcitonin stimulates expression of the rat 25-hydroxyvitamin D3–24-hydroxylase (CYP24) promoter in HEK-293 cells expressing calcitonin receptor: identification of signaling pathways. *J Mol Endocrinol*. 2004;32:87–98.
121. Becskei C, Riediger T, Zünd D, et al. Immunohistochemical mapping of calcitonin receptors in the adult rat brain. *Brain Res*. 2004;1030: 221–233.
122. Cheng W, Gonzalez I, Pan W, et al. calcitonin receptor neurons in the mouse nucleus tractus solitarius control energy balance via the non-aversive suppression of feeding. *Cell Metab*. 2020;31:301–312.e5.
123. Endo N, Fujino K, Doi T, et al. Effect of elcatonin versus nonsteroidal anti-inflammatory medications for acute back pain in patients with osteoporotic vertebral fracture: a multiclinic randomized controlled trial. *J Bone Miner Metab*. 2017;35:375–384.
124. Shah GV, Deftos LJ, Crowley WR. Synthesis and release of calcitonin-like immunoreactivity by anterior pituitary cells: evidence for a role in paracrine regulation of prolactin secretion. *Endocrinology*. 1993;132:1367–1372.
125. Yeh CB, Weng SJ, Chang KW, et al. Calcitonin alleviates hyperalgesia in osteoporotic rats by modulating serotonin transporter activity. *Osteoporos Int*. 2016;27:3355–3364.
126. Naot D, Musson DS, Cornish J. The activity of peptides of the calcitonin family in bone. *Physiol Rev*. 2019;99:781–805.
127. Henson B, Hollingsworth H, Nevois E, et al. Calcitonin gene-related peptide (CGRP) antagonists and their use in migraines. *J Pain Palliat Care Pharmacother*. 2019;25:1–10.
128. Kilinc E, Dagistan Y, Kukner A, et al. Salmon calcitonin ameliorates migraine pain through modulation of CGRP release and dural mast cell degranulation in rats. *Clin Exp Pharmacol Physiol*. 2018;45:536–546.
129. Niu P, Shindo T, Iwata H, et al. Protective effects of endogenous adrenomedullin on cardiac hypertrophy, fibrosis, and renal damage. *Circulation*. 2004;109:1789–1794.
130. Costante G, Meringolo D. Calcitonin as a biomarker of C cell disease: recent achievements and current challenges. *Endocrine*. 2020;67: 273–280.
131. Gillespie MT, Thomas RJ, Pu ZY, et al. Calcitonin receptors, bone sialoprotein and osteopontin are expressed in primary breast cancers. *Int J Cancer*. 1997;73:812–815.
132. Wang X, Nakamura M, Mori I, et al. Calcitonin receptor gene and breast cancer: quantitative analysis with laser capture microdissection. *Breast Cancer Res Treat*. 2004;83:109–117.
133. Chien J, Ren Y, Qing Wang Y, et al. Calcitonin is a prostate epithelium-derived growth stimulatory peptide. *Mol Cell Endocrinol*. 2001;181:69–79.
134. Thakkar A, Bijnsdorp IV, Geldof AA, et al. Profiling of the calcitonin-calcitonin receptor axis in primary prostate cancer: clinical implications and molecular correlates. *Oncol Rep*. 2013;30:1265–1274.
135. Thomas S, Chigurupati S, Anbalagan M, et al. Calcitonin increases tumorigenicity of prostate cancer cells: evidence for the role of protein kinase A and urokinase-type plasminogen receptor. *Mol Endocrinol*. 2006;20:1894–1911.
136. Aldahish A, Kale A, Aljameeli A, et al. Calcitonin induces stem cell-like phenotype in prostate cancer cells. *Endocr Relat Cancer*. 2019;26:815–828.
137. Aljameeli A, Thakkar A, Thomas S, et al. Calcitonin receptor-zonula occludens-1 interaction is critical for calcitonin-stimulated prostate cancer metastasis. *PLOS ONE*. 2016:e0150090.
138. Chen WC, Lin EY, Kang YN. Efficacy and safety of elcatonin in postmenopausal women with osteoporosis: a systematic review with network meta-analysis of randomized clinical trials. *Osteoporos Int*. 2019;30:1723–1732.
139. Chesnut 3rd CH, Majumdar S, Newitt DC, et al. Effects of salmon calcitonin on trabecular microarchitecture as determined by magnetic resonance imaging: results from the QUEST study. *J Bone Miner Res*. 2005;20:1548–1561.
140. Chen M, Li XR, Zhou YX, et al. Improved absorption of salmon calcitonin by ultraflexible liposomes through intranasal delivery. *Peptides*. 2009;30:1288–1295.
141. DeRose J, Singer FR, Avramides A, et al. Response of Paget's disease to porcine and salmon calcitonins: effects of long-term treatment. *Am J Med*. 1974;56:858–866.
142. Tsukamoto M, Menuki K, Murai T, et al. Elcatonin prevents bone loss caused by skeletal unloading by inhibiting preosteoclast fusion through the unloading-induced high expression of calcitonin receptors in bone marrow cells. *Bone*. 2016;85:70–80.
143. Wootton R, Reeve J, Spellacy E, et al. Skeletal blood flow in Paget's disease of bone and its response to calcitonin therapy. *Clin Sci Mol Med*. 1978;54:69–74.
144. Cheng T, Zhang L, Fu X, et al. The potential protective effects of calcitonin involved in coordinating chondrocyte response, extracellular matrix, and subchondral trabecular bone in experimental osteoarthritis. *Connect Tissue Res*. 2013;54:139–146.
145. Wen ZH, Tang CC, Chang YC, et al. Calcitonin attenuates cartilage degeneration and nociception in an experimental rat model of osteoarthritis: role of TGF-β in chondrocytes. *Sci Rep*. 2016;6:28862.
146. Ryan SM, McMorrow J, Umerska A, et al. An intra-articular salmon calcitonin-based nanocomplex reduces experimental inflammatory arthritis. *J Control Release*. 2013;167:120–129.
147. Lin Z, Pavlos NJ, Cake MA, et al. Evidence that human cartilage and chondrocytes do not express calcitonin receptor. *Osteoarthritis Cartilage*. 2008;16:450–457.
148. European Medicines Agency recommends limiting long-term use of calcitonin medicines. https://www.ema.europa.eu/en/news/european-medicines-agency-recommends-limiting-long-term-use-calcitonin-medicines.
149. Lowry F. FDA panel: long-Used osteoporosis drug too risky. https://www.webmd.com/osteoporosis/news/20130307/fda-panel-osteoporosis-drug-too-risky#1.

50

Vitamin D: From Photosynthesis, Metabolism, and Action to Clinical Applications

Roger Bouillon, Lieve Verlinden, and Geert Carmeliet

OUTLINE

Historic Overview, 850
Origin of Vitamin D: Nutrition and Photosynthesis, 850
Metabolism of Vitamin D, 851
  25-Hydroxylation, 851
  1α-Hydroxylation, 851
  24-Hydroxylation: Catabolism or Specific Function?, 852
  Other Metabolic Pathways for Vitamin D and its Metabolites, 852
Vitamin D Transport, 853
Action and Mode of Action, 853
  General Characteristics of the Vitamin D Receptor, 853
  Classic Target Tissues, 854
Clinical Consequences of Vitamin D Status, 855
  Skeletal Effects of the Vitamin D Endocrine System, 855
  Extraskeletal Actions of the Vitamin D Endocrine System, 856
Clinical Use of Vitamin D, 860
  Assays for Vitamin D and Metabolites: Methodology and Applications, 860
  Optimal Serum 25OHD Concentration, 861
  Worldwide Vitamin D Status, 862
  Vitamin D Requirements, 863
Summary, 865

HISTORIC OVERVIEW

Early in the twentieth century, it was demonstrated that rachitic children were cured after exposure to sunlight or by oral cod liver oil. The new essential nutrient was called *vitamin D*. The two discoveries of vitamin D were unified by the demonstration that irradiation of 7-dehydrocholesterol in the skin could produce the antirachitic vitamin D. Irradiation of yeast ergosterol provided a large supply of vitamin D_2 used for the prevention and virtual elimination of vitamin D–deficiency rickets in the Western world (for references see Bouillon and Antonio[1]). During the last century, extensive studies revealed (1) the complex metabolism of vitamin D, which involves approximately 50 metabolites, (2) the mode of action of the most active metabolite, 1,25-dihydroxyvitamin D ($1,25(OH)_2D$), as a ligand for a specific and quite ubiquitous nuclear transcription factor, vitamin D receptor (VDR), and (3) the regulation of the expression of a very large number of genes (between 1% and 5% of the human genome) either involved in calcium homeostasis or related to a wide variety of functions in all vertebrates, including humans.

ORIGIN OF VITAMIN D: NUTRITION AND PHOTOSYNTHESIS

Vitamin D can be obtained from dietary sources of vegetal (vitamin D_2, or ergocalciferol) or animal (vitamin D_3, or cholecalciferol) origin. Approximately 70% of dietary vitamin D is absorbed by the enterocytes and transported to the blood circulation via chylomicrons. This vitamin D is taken up by a variety of tissues (fat and muscle) before the chylomicron remnants and their vitamin D finally reach the hepatocytes. The best food sources are oily fish, but it is also found in small amounts in butter, cream, and egg yolks. Breast milk (even from mothers living in equatorial areas of the world) and cow's milk are poor sources of vitamin D, providing only 15 to 40 IU/L, and only contain minimal concentrations of 25OHD or $1,25(OH)_2D$. In most countries, the median vitamin D intake is less than 5 μg/d, which is insufficient for daily needs. Therefore, many countries have introduced a policy of vitamin D enrichment of selected food items, such as fortification of milk and dairy products in North America, so that the median vitamin D intake according to National Health and Nutrition Examination Survey (NHANES) data is between 6 and 9 μg/d. Vitamin D is remarkably stable and does not deteriorate when food is heated or stored for long periods. In view of the low vitamin D content of a vegetarian diet (natural vitamin D intake is indeed related to intake of animal fat), vitamin D deficiency and rickets are risks for strictly vegetarian children with insufficient sun exposure or vitamin D supplementation. In view of the low contribution from food, most vitamin D comes from ultraviolet (UV) B –driven synthesis in the skin. The very same UVB light needed for the photochemical conversion of 7DHC to previtamin D is also able to damage DNA and causes photodamage, and, with a long lag time, increases the risk of skin cancers, including melanoma. This was no real problem during human evolution, but with increasing life expectancy the benefits of UV light for the photosynthesis of vitamin D should be compared with the lifetime risk of skin damage, especially as vitamin D supplementation can safely replace synthesis in the skin. As described in this chapter, and because of a combination of several predisposing factors, variable degrees of vitamin D deficiency are highly prevalent around the world. By contrast, excess vitamin D is quite exceptional.

Hypervitaminosis can occur when excess pharmaceutical vitamin D is taken, with a wide variety of symptoms and signs related to hypercalciuria, hypercalcemia, and ectopic calcifications (Table 50.1). The toxic dosage has not been established, but toxicity should always be monitored when daily doses that markedly exceed the current upper limit of

TABLE 50.1 Symptoms of Vitamin D Toxicity

TABLE 50.1 Symptoms of Vitamin D Toxicity
Hypercalciuria and/or kidney stones/nephrocalcinosis
Hypercalcemia and/or hyperphosphatemia
Polyuria and polydipsia
Decalcification of bone
Ectopic calcification of soft tissues (kidney and lung)
Nausea, vomiting, anorexia, or constipation
Headache
Hypertension

more than 100 µg are given for a long period of time. Overproduction of renal 1,25(OH)$_2$D by abnormal hormonal stimuli (as seen in fibroblast growth factor-23 [FGF-23] or Klotho-null mice) or absence of CYP24A1 (see later), the main catabolizing enzyme, causes the same calcemic side effects, with severe multiple-organ calcification (especially kidney, vascular wall, and heart valves) leading to premature death. Apart from hypercalcemic consequences, some randomized controlled trials (RCTs) suggest that serum concentrations above 45 ng/mL may have negative effects on the risk of falls or fractures (see later).

Apart from substrate (7DHC) availability (which may be influenced by genetic polymorphisms of 7HDC-Δ7-reductase enzyme), the photochemical synthesis of vitamin D$_3$ in the skin largely depends on the amount of UVB photons that strike the basal epidermal layers. Glass, sunscreen, clothes, and skin pigment absorb UVB and blunt vitamin D$_3$ synthesis. Latitude, time of day, and season are factors that influence the intensity of solar radiation and the cutaneous production of vitamin D$_3$. Therefore, there is a risk for a shortage of vitamin D supply during winter and spring. In both the Northern and Southern hemispheres above 40 degrees' latitude, vitamin D$_3$ synthesis in the skin decreases or disappears during winter months, owing to the low inclination of the sun and the atmospheric filtration of the shortest UV waves of sunlight, which are necessary for vitamin D$_3$ synthesis. The importance of skin synthesis of vitamin D$_3$ to maintain normal vitamin D status is best reflected by the high prevalence of vitamin D deficiency in countries where exposure to sunlight is extremely low for cultural and religious reasons. Solar exposure of 2 hours per week of the face and hands is probably enough for maintaining normal 25OHD concentrations in children and adults but should be further fine-tuned according to the climate and latitude. There is no unanimity regarding the amount of vitamin D produced by UVB exposure, as older studies suggested that full-body exposure to UVB may produce more than 20,000 IU/d, whereas more recent studies of young adults exposed to intensive sun exposure during a winter holiday in the Canary Islands showed an increase in serum 25OHD equivalent to the daily supply of approximately 800 IU/d (for references see[2]).

Nature has built in several feedback mechanisms to minimize the risk of prolonged sun exposure causing vitamin D intoxication. Cutaneous vitamin D, and especially previtamin D, are photosensitive and are degraded to inactive sterols. Sunlight-induced melanin synthesis, acting as a natural sunscreen, provides additional negative feedback. This is best demonstrated by mean serum 25OHD levels of approximately 45 ng/mL found in Masai and other African tribes with extensive exposure to sunlight. These values are higher than mean values found in most other populations but lower than expected on the basis of intensive UVB exposure and skin capacity to produce vitamin D.

KEY POINTS

- Intake of vitamin D from food is low because most food items have a low vitamin D content and is insufficient for most people.
- Most vitamin D is synthesized in the skin after exposure to ultraviolet B sun rays. Unfortunately, these ultraviolet B rays are also photodamaging to the skin and can increase the risk of skin cancer in the long term.
- The vitamin D content of breast milk is low and largely insufficient to cover the needs of infants or children around the world.
- Vitamin D excess attributed to overexposure to sunlight is highly unlikely because evolution has provided many mechanisms to avoid vitamin D toxicity.

METABOLISM OF VITAMIN D

Vitamin D is biologically inert and requires two successive hydroxylations in the liver (on C$_{25}$) and kidney (on the α position of C$_1$), mediated by cytochrome P450 enzymes to form its hormonally active metabolite, 1,25(OH)$_2$D (Fig. 50.1).

25-Hydroxylation

Although the liver is probably the main tissue responsible for 25-hydroxylation of vitamin D, extrahepatic 25-hydroxylation has been observed in many tissues. The most important 25-hydroxylase is CYP2R1 because biallelic mutations of the gene encoding this protein were identified in patients with classical rickets with low 25OHD levels. Several other cytochrome P enzymes are, however, also able to perform the same reaction. In contrast to the widespread concept of constitutive *CYP2R1* expression, new data in mice have demonstrated that this gene is controlled by several metabolic conditions, as both gene and protein expression are strongly reduced in cases of fasting, diet-induced obesity, glucocorticoid exposure, or type 1 diabetes mellitus (T1DM) or type 2 diabetes mellitus (T2DM). PPAR gamma coactivator, estrogen-related receptor, and glucocorticoid receptor are all involved in the regulation of *Cyp2r1* in mice. If this also applies to humans, it may well help to explain the frequent association between low serum concentrations of 25OHD and many metabolic diseases (reviewed in Bouillon & Bikle[3]).

1α-Hydroxylation

25OHD is biologically inactive and requires further hydroxylation in the kidney to the active hormone, 1,25(OH)$_2$D, by 25-hydroxyvitamin D-1α-hydroxylase (CYP27B1). The production of 1,25(OH)$_2$D is regulated primarily at this final step by several factors. The proximal renal tubule is the principle site of 1α-hydroxylation, but high levels of CYP27B1 mRNA and enzyme activity have also been found in human keratinocytes, macrophages, and approximately ten other tissues. Renal CYP27B1 activity is tightly controlled by 1,25(OH)$_2$D (negative feedback), by parathyroid hormone (PTH), calcitonin, and insulin-like growth factor 1 (all positive feedback), as well as by phosphate, calcium, and especially FGF-23 (negative regulation). The promoters of the mouse and human *CYP27B1* genes are profoundly responsive to PTH and are negatively regulated by 1,25(OH)$_2$D.[4] Extrarenal CYP27B1 activity is regulated by different (yet to be identified) mechanisms in the skin and monocytes.

Pseudovitamin D–deficiency rickets, also known as vitamin D–dependency rickets type I, is an autosomal recessive disease characterized by failure to thrive, muscle weakness, skeletal deformities, hypocalcemia, secondary hyperparathyroidism, normal to high serum levels of 25OHD, and low serum 1,25(OH)$_2$D concentrations, all caused by impaired activity of renal CYP27B1. These patients recover well when treated with oral physiologic doses of 1,25(OH)$_2$D.[2,5,6]

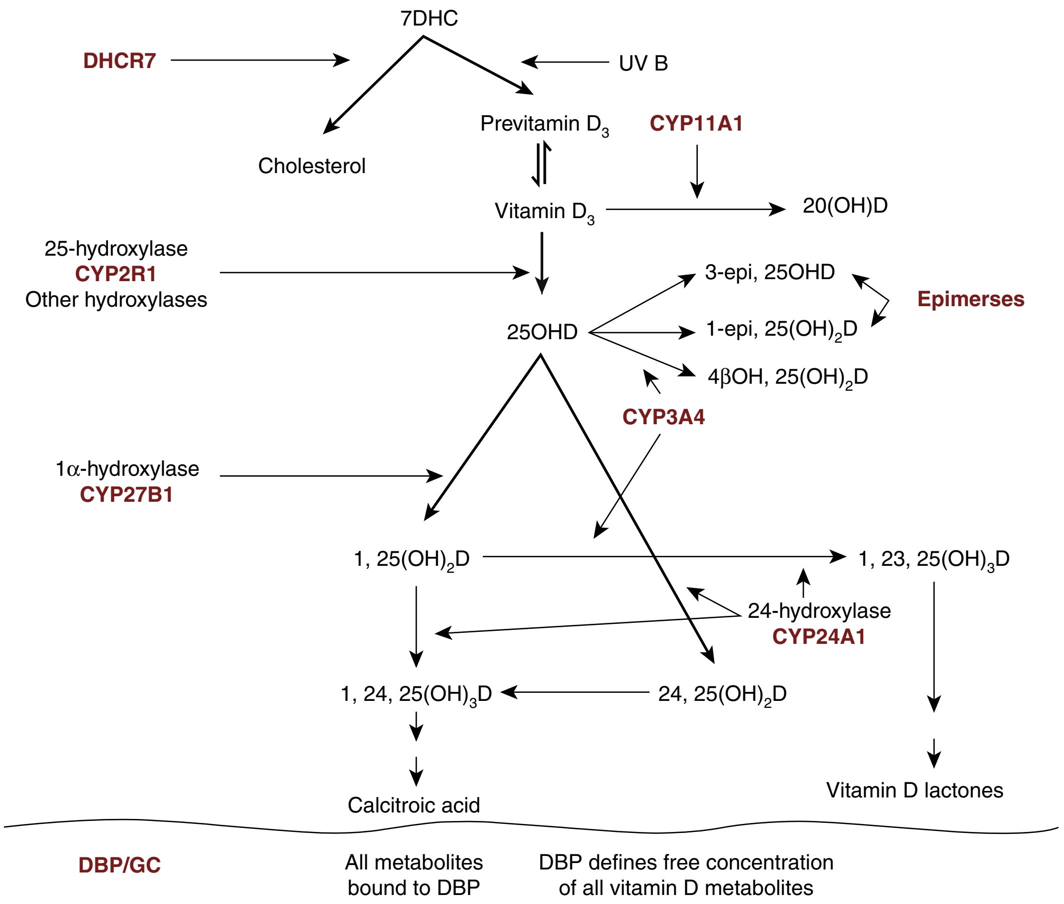

7DHC
DHCR7
UV B
Cholesterol
Previtamin D_3
CYP11A1
Vitamin D_3
20(OH)D
25-hydroxylase
CYP2R1
Other hydroxylases
3-epi, 25OHD
Epimerses
25OHD
1-epi, $25(OH)_2D$
4βOH, $25(OH)_2D$
CYP3A4
1α-hydroxylase
CYP27B1
1, $25(OH)_2D$
1, 23, $25(OH)_3D$
24-hydroxylase
CYP24A1
1, 24, $25(OH)_3D$
24, $25(OH)_2D$
Calcitroic acid
Vitamin D lactones
DBP/GC
All metabolites bound to DBP
DBP defines free concentration of all vitamin D metabolites

Fig. 50.1 Major enzymes and metabolic pathways of vitamin D. Vitamin D_3 originates from dietary sources or is synthesized in the skin from 7-dehydrocholesterol into previtamin D. Approximately one in three molecules of vitamin D is later found as 25OHD generated by CYP2R1 (mostly in the liver) and maybe to a lesser extent by some other cytochrome P enzymes. What happens to the other vitamin D molecules is not fully known (they may be esterified or metabolized into other D metabolites by CYP11A1). In normal subjects approximately one molecule of 25OHD is metabolized into $1,25(OH)_2D$ by CYP27B1 in the kidney under tight feedback control. The major $1,25(OH)_2D$ catabolic pathways are under the control of a multifunctional CYP24A1 that generates many intermediate metabolites, ultimately generating calcitriol or lactones. The role of several epimerases is unknown, and their endproducts are so far believed to be largely devoid of physiologic action. CYP3A4 probably has a minor role, except in case of drug-induced induction of this enzyme, in which case it accelerates 25OHD degradation. *DBP,* Vitamin D–binding protein; *GC*, group-specific component; *UV*, ultraviolet.

24-Hydroxylation: Catabolism or Specific Function?

An alternative hydroxylation of 25OHD occurs on carbon 24 mediated by the multifunctional enzyme 24-hydroxylase (CYP24A1). This enzyme not only initiates the catabolic cascade of 25OHD and $1,25(OH)_2D$ by 24-hydroxylation but catalyzes its further metabolism into calcitroic acid. Evidence for this catabolic role of 24-hydroxylase was provided by *Cyp24a1*-deficient mice, which exhibited pathology consistent with systemic excess of $1,25(OH)_2D$. Expression of the *CYP24A1* gene has been detected in virtually all nucleated cells. Induction of *CYP24A1* expression is a highly sensitive biomarker for responsiveness to vitamin D supplementation owing to the presence of several vitamin D–responsive elements in its promoter. Several mutations in the human *CYP24A1* gene have recently been found to be responsible for infantile hypercalcemia and increased risk of kidney stones in adults.[7]

Patients with loss-of-function *CYP24A1* mutations have been successfully treated by restriction of vitamin D intake or with inhibitors of vitamin D synthesis or inducers of *CYP3A4* expression.

KEY POINTS

- Vitamin D is an inactive precursor and requires hydroxylation by CYP2R1 to produce 25-hydroxyvitamin D (the main circulating metabolite) and by CYP27B1 to produce 1,25-dihydroxyvitamin D [1,25(OH)2D]. The catabolism of all vitamin D metabolites largely depends on CYP24A1. Genetic or acquired deficiency of each of these enzymes can generate important diseases related to disruption of calcium, phosphate, or bone homeostasis. There are, however, several other metabolic steps resulting in approximately 50 different (probably mostly inactive) metabolites.

Other Metabolic Pathways for Vitamin D and its Metabolites

Apart from the multifunctional 24-hydroxylation pathway, C_{23} and C_{26} hydroxylation of $1,25(OH)_2D$ are also possible in the absence of prior 24-hydroxylation. Both activities are necessary for the formation of 25OHD- or $1,25(OH)_2D$-23,26-lactones. A-ring metabolism involves C_{19} oxidation and 3-epimerisation. The latter reaction, which is irreversible, occurs only in a limited number of cells (e.g., keratinocytes, bone, and parathyroid cells) and is performed by hydroxysteroid dehydrogenases. 1-epi-25$(OH)_2D$ also circulates in the serum and exhibits a strong correlation with serum 25OHD concentrations. The enzyme or tissue responsible for its synthesis, and the implications, are still to be explored.[8]

The physiologic implications of 4-hydroxylation or 11-hydroxylation are presently unclear. CYP3A4 is another P450 enzyme with mainly a detoxicating function but can also strongly influence vitamin D metabolism. Excess activity (owing to either genetic mutation or drug-induced induction of this enzyme) increases vitamin D/25OHD catabolism, and can even lead to rickets (type 3)[9] (for a review see Bouillon and Bikle[3]) (Fig. 50.1). Vitamin D is mainly excreted in the bile after esterification in the liver, but some of its more polar metabolites (e.g., calcitroic acid) are excreted via the urine. The enterohepatic recirculation of vitamin D esters is probably devoid of biological relevance.

VITAMIN D TRANSPORT

Nutritional vitamin D is absorbed by the gut and then transported via the lymphatic system by chylomicrons and stored in several tissues (e.g., fat and muscle). Skin-produced vitamin D binds directly to an α-globulin known as vitamin D–binding protein (DBP). This protein is essential for transport to the liver, where vitamin D is hydroxylated and thereafter released as 25OHD, again bound to DBP. Human DBP, also known as group-specific component, or GC-globulin, is a highly polymorphic glycoprotein synthesized by the liver. Worldwide, over 120 *GC* alleles have been detected, making the *DBP* locus one of the most polymorphic known. *GC1F*, *GC1S*, and *GC2* are the three most common alleles.[10]

Dbp-null mice are viable and fertile and have a normal bone and calcium homeostasis, despite extremely low concentrations of 25OHD and 1,25$(OH)_2D$. This is a very strong argument in favor of the role of free rather than total concentrations because tissue samples from these mice show normal free 1,25$(OH)_2D$ concentrations. One only case of biallelic mutations with silencing of *DBP* has been described. This adult woman did not have vitamin D–related health problems, despite near-undetectable serum concentrations of all vitamin D metabolites.[11] DBP has a cleft on its surface that serves as a binding site for all vitamin D metabolites. Its affinity for D_2 metabolites is slightly lower than that for D_3 metabolites. As probably only non–DBP-bound vitamin D metabolites can readily cross the plasma membrane, and given that the VDR has a much higher affinity for 1,25$(OH)_2D$ than for 25OHD (100-fold difference), while the opposite is true for DBP, it is clear that 1,25$(OH)_2D$ has substantially higher cellular uptake than 25OHD. The half-lives of 25OHD and 1,25$(OH)_2D$ in the human circulation are approximately 2 weeks and 4 to 6 hours, respectively.[12] DBP's function in the vitamin D endocrine system is assumed to reflect the "free hormone" hypothesis, which states that the unbound (free), rather than the protein-bound, fraction of the active vitamin D hormone is responsible for its biological activity. The plasma concentration of DBP is increased by estrogens. In women, the DBP concentration therefore doubles at the end of pregnancy. Studies with *Megalin*-null mice indicate that megalin, a lipoprotein-like receptor present at the surface of the proximal tubular cells in the kidney, is responsible for the reabsorption of DBP and of DBP complexed with vitamin D sterols. This megalin reabsorption mechanism may control the availability of the 25OHD/DBP complex for the 25OHD-1α-hydroxylation enzyme and explain the severe bone disease of *Megalin*-deficient mice. *Dbp*-null mice are more sensitive to vitamin D deficiency and less sensitive to vitamin D excess, probably because of enhanced urinary loss of vitamin D metabolites.[10] The *Dbp* and *Megalin* null mice suggest that the main function of DBP is indeed to transport all vitamin D metabolites and preserve them from rapid clearance or urinary loss.

DBP binds globular actin with a high affinity ($K_a = 2 \times 10^9$ mmol/L). Actin is the most abundant intracellular protein. Cell motility, shape, and size depend on the ability of globular actin to polymerize into filaments (F-actin). Upon cell injury or cell necrosis, actin is released into the extracellular space. When actin is released from cells, it may rapidly form filaments, with detrimental effects for the microcirculation. Two plasma proteins, DBP and gelsolin, bind actin avidly, thereby acting as an "actin-scavenger" system.[10] Whether this has clinical implications for human health and disease is so far unknown.

KEY POINTS

- All vitamin D metabolites are bound in serum to a specific vitamin D–binding protein, DBP, the oldest member of the albumin family. Absence of DBP does not cause a clinical phenotype, but serum concentrations of all vitamin D metabolites are extremely low. This confirms that the free vitamin D hormone concentration is physiologically important.

ACTION AND MODE OF ACTION

General Characteristics of the Vitamin D Receptor

1,25$(OH)_2D$, the hormonally active form of vitamin D, exerts its effects mainly by activating the nuclear VDR, a member of the nuclear-receptor superfamily of ligand-activated transcription factors that has a structure similar to that of other steroid-binding nuclear receptors.[10,13] The human *VDR* gene on chromosome 12 encodes a 427–amino acid protein, with some minor variants containing an additional 23 or 50 amino acids at the N-terminus.

Genomic Actions. Binding of 1,25$(OH)_2D$ to VDR generates conformational changes in VDR followed by heterodimerization with unliganded RXR and binding to vitamin D response elements (VDREs) in the promoter region of vitamin D target genes, with subsequent release of corepressors and recruitment of coactivators and general transcription factors for the assembly of an active transcriptional complex. A crucial event in this respect is the mousetrap-like intramolecular folding of helix 12, which closes off the ligand-binding pocket and exposes the activation function domain for interaction with coactivators. Corepressors bind and silence unliganded steroid receptors via recruitment of histone deacetylases, maintaining chromatin in a transcriptionally repressed state. Coactivators are a group of proteins that allow gene transcription in several waves of activity. Gene expression can also be mediated by adenosine triphosphate (ATP)-dependent chromatin remodeling complexes such as SWI/SNF-type and ISWI-type complexes and the multiprotein complex WINAC. Furthermore, the expression or the recruitment of these regulatory proteins is regulated by several intracellular signaling pathways and by steroids themselves, with receptor agonists or antagonists inducing preferential recruitment of coactivators or corepressors, respectively. Finally, 1,25$(OH)_2D$ also controls gene regulation by regulating several miRNAs. A hexanucleotide direct repeat spaced by three nucleotides is the cognate VDRE of which RXR and VDR bind the 5′ and 3′ ends, respectively, although alternative options appear to be possible.

Nongenomic Actions. Several groups have described rapid effects of 1,25$(OH)_2D$ that are independent of transcription and are mediated by a membrane receptor for 1,25$(OH)_2D$. These so-called nongenomic effects are also well known for other nuclear receptor ligands such as estradiol or progesterone and include the opening of calcium or chloride channels and the activation of second-messenger signaling pathways. The *in vivo* effects of these nongenomic actions remain unexplained.

Classic Target Tissues

The action of 1,25$(OH)_2$D on bone, intestine, kidney, and the parathyroid glands and its role in mineral metabolism is the result of a complex interplay between calcium and phosphate, 1,25$(OH)_2$D, PTH, and phosphatonins.[14] PTH induces calcium mobilization from bone and stimulates 1,25$(OH)_2$D production, but its secretion is inhibited by the action of 1,25$(OH)_2$D on the parathyroid glands (negative feedback). In a second negative feedback loop, 1,25$(OH)_2$D limits its own availability by inhibition of CYP27B1 and stimulation of CYP24A1, inducing 1,25$(OH)_2$D catabolism. In the last few years, considerable progress has been made in understanding phosphate homeostasis. The phosphaturic hormone phosphatonin, or FGF-23, is produced by osteocytes and osteoblasts and inhibits the activity of the NPT2 protein. The *NPT2* gene encodes a renal sodium/phosphate cotransporter responsible for reabsorption of phosphate and represents a newly identified target gene for 1,25$(OH)_2$D. Phosphatonin can be indirectly inactivated by a protease encoded by the *PHEX* gene, which was identified as the gene that is defective in X-linked hypophosphatemic rickets. FGF-23 secretion is stimulated by 1,25$(OH)_2$D and impairs renal CYP27B1, creating an additional feedback system, such that the production of 1,25$(OH)_2$D is tightly feedback-regulated.

Effects on Intestine. The efficiency of the small intestine to absorb dietary calcium is increased by 1,25$(OH)_2$D, and VDR abundance is highest in the duodenum, followed by the jejunum and the ileum. Although the exact mechanism by which 1,25$(OH)_2$D alters the flux of calcium across the intestinal absorptive cell is not known, 1,25$(OH)_2$D increases the production and activity of several proteins in the small intestine, including transient receptor potential cation channel subfamily V (TRPV) member 6 (TRPV6) and TRPV5, calbindin-D9K, alkaline phosphatase, and low-affinity Ca-ATPase (PMCA). The entry of calcium from the intestinal lumen across the brush border membrane into the enterocyte is mainly regulated by the epithelial channels TRPV6 and TRPV5. The intracellular calcium transfer is considered to be dependent mainly on calbindin-D9K. The transfer of calcium from the cytoplasm to the extracellular space requires energy input because of an uphill concentration gradient and an unfavorable electrochemical gradient. Both the plasma membrane calcium pump and a sodium-calcium exchanger play important roles in this process. The stimulatory effect of 1,25$(OH)_2$D on the ATP-dependent uptake of calcium at the basolateral membrane involves an increase in *PMCA* expression. The essential role of the intestine for calcium and phosphate homeostasis was clearly demonstrated by the phenotype of *Vdr* knockout mice. Such *Vdr*-null mice are phenotypically normal at birth, but after weaning they develop hypocalcemia, secondary hyperparathyroidism, and hypophosphatemia, despite very high levels of 1,25$(OH)_2$D. They become growth-retarded and develop severe rickets. Mice deficient in *Cyp27b1* display a similar phenotype.[5,13] This bone and calcium phenotype can be largely corrected by a high dietary calcium intake (especially in combination with high lactose intake) in both knockout models or by 1,25$(OH)_2$D treatment in *Cyp27b1*-null mice. These findings confirm previous observations in humans. These data strongly suggest that the intestine is the primary target for 1,25$(OH)_2$D's action on calcium/bone homeostasis. This is largely confirmed by genetic mouse models of selective rescue or deletion of *Vdr* in the intestine of transgenic mice.[15] The primary molecular targets, however, merit further exploration; mice with ablation of *CaBP-9k* or *TRPV6*, or even both, can still partially increase their intestinal calcium absorption, and serum calcium levels are normal. Paracellular intestinal calcium transport may also be part of the picture of vitamin D's action, in that the expression of claudin 2 and claudin 12, both known to form paracellular calcium channels, is induced by 1,25$(OH)_2$D and decreased in the intestine of *Vdr*-null mice.[5]

Effects on Kidney. The kidney is important both for the metabolism of 1,25$(OH)_2$D and the reabsorption of calcium and phosphate, all processes that are regulated by 1,25$(OH)_2$D.[5] The kidney, and more specifically the proximal tubule, is the central tissue for 1α-hydroxylation of 25OHD. Chronic renal failure reduces CYP27B1 activity, which ultimately results in renal osteodystrophy or uremic bone disease. 1,25$(OH)_2$D also increased the distal tubular reabsorption of calcium; as in the intestine, TRP channels (especially TRPV5), calbindin-D9K and -28K, and the plasma membrane calcium ATPase are involved. The crucial role of 1,25$(OH)_2$D-regulated renal calcium reabsorption was demonstrated by persistent hypercalciuria and reduction in bone mass in TRPV5-deficient mice. The kidney is also the major component in phosphate homeostasis, as both PTH and FGF-23, in complex interplay with 1,25$(OH)_2$D, are able to reduce renal phosphate reabsorption

Effects on Bone. 1,25$(OH)_2$D has dual effects on bone: it can stimulate osteoclastogenesis and bone resorption, as well as modify osteoblast function and bone mineralization. The overall effects of vitamin D metabolites on bone are thus extremely complex. From observations in humans and animals, vitamin D deficiency or resistance impairs bone matrix mineralization, whereas osteoblast activity and matrix synthesis are even stimulated. Excess 1,25$(OH)_2$D can clearly enhance osteoclastogenesis and bone resorption. Because bone mineralization and bone structure can be largely normalized in vitamin D- or 1,25$(OH)_2$D-deficient or resistant mice by sufficient supply of minerals via active or passive intestinal calcium absorption, it seems that the direct effects of vitamin D metabolites on chondrocytes and bone cells are redundant if calcium and phosphate supply are guaranteed. However, most of the genes and proteins typically expressed in osteoblasts and osteoclasts are vitamin D–regulated, so it is likely that 1,25$(OH)_2$D can fine-tune bone mineral homeostasis.

The direct effects of the vitamin D endocrine system on bone cells are slowly being unraveled by studies employing cell–specific *Vdr* deletion: *Vdr* deficiency in osteoblasts modestly increases bone mineral mass, whereas *Vdr* deficiency in osteocytes does not affect basal bone mass or structure.[15] However, mice with osteocyte-selective *Vdr* deficiency exposed to a very low calcium diet or high 1,25$(OH)_2$D concentrations are protected against 1,25$(OH)_2$D-induced bone loss.[15] This is largely owing to 1,25$(OH)_2$D-mediated inhibition of mineral deposition (creating instead of solving osteomalacia) by regulating several genes (*ANK*, *ENPP1&3*, *TNAP*) as to increase PPi and other mineralization inhibitors such as osteopontin and other SIBLING proteins.[15] Therefore, it seems that 1,25$(OH)_2$D can have different effects on bone: in cases of sufficient calcium intake, it helps to stimulate calcium absorption and provide minerals and osteoblast function so as to improve mineral deposition; however, in cases of calcium deficiency, 1,25$(OH)_2$D will primarily defend serum calcium homeostasis at the expense of bone (increased bone resorption and inhibition of mineral deposition). This should not be a major surprise, as the vitamin D endocrine system, together with PTH, evolved during vertebrate evolution as to enable terrestrial animals to cope with increased gravity (necessitating a solid bone structure) while living in a calcium-poor environment.

Effects on Growth Plate. The absence of *Vdr* or *Cyp27b1* creates no detectable phenotype in overall growth or growth plate of prenatal animals, but the longitudinal growth of long bones is impaired after weaning. X-ray analysis reveals advanced rickets, including widening of the epiphyseal growth plate, with an increased width and marked disorganization of the growth plate on histology, including impaired mineralization of hypertrophic chondrocytes. This increased growth plate width in *Vdr*- or *Cyp27b1*-null mice is not caused by abnormal chondrocyte proliferation or differentiation, and the expression of major proteins such as collagen X and osteopontin is normal. The expansion of the growth plate can be largely explained by decreased apoptosis of hypertrophic chondrocytes. Based on analysis of several genetic models with abnormal phosphate homeostasis, serum phosphate levels are probably crucial for hypertrophic chondrocyte apoptosis *in vivo*. This is confirmed *in vitro*: apoptosis of hypertrophic chondrocytes is regulated by phosphate levels via the activation of the caspase-9–mediated mitochondrial pathway. In accordance with these findings, chondrocyte-specific inactivation of *Vdr* did not cause a growth plate phenotype, and certainly not rickets.

Parathyroid Glands. The parathyroid gland is an important target for vitamin D action, as it regulates the sensing of serum ionized calcium through a G-protein–coupled calcium-sensing receptor. Several parathyroid genes, including *CaSR* and *PTH*, are under the control of 1,25$(OH)_2$D. Tissue-selective knockout of *Vdr* in the parathyroid gland causes hyperparathyroidism and increased bone turnover. Parathyroid-selective deletion of *Cyp27b1* creates a complex mixture of hypocalcemia, low serum 1,25$(OH)_2$D levels and decreased bone mass, thus questioning the unique role of the kidney in the production of systemic 1,25$(OH)_2$D. The vitamin D endocrine system also plays a key role in the genesis of secondary hyperparathyroidism of chronic renal failure, including the suppression of local CYP27B1 activity by excess FGF-23.

KEY POINTS

- 1,25(OH)2D binds to a specific nuclear receptor, the vitamin D–binding receptor (VDR), and this complex recruits a set of other proteins that together regulates gene transcription, very similar to the mode of action of many other hormones that bind to nuclear receptors.
- The most important target tissue of 1,25(OH)2D is the intestine, as it stimulates the expression of several calcium-binding proteins that, combined, are responsible for intestinal calcium absorption. VDR is also expressed in all bone cells, but the direct effects of 1,25(OH)2D on these bone cells or chondrocytes are limited.
- VDR is expressed in many nonskeletal cells and regulates a very large number of genes (up to 5% of the human genome) that are unrelated to calcium or phosphate handling, and are mostly regulated in a coherent fashion. This suggest that the vitamin D endocrine system may have a wide spectrum of biological activities.

CLINICAL CONSEQUENCES OF VITAMIN D STATUS

Skeletal Effects of the Vitamin D Endocrine System

The absence of vitamin D, its active metabolite, 1,25$(OH)_2$D, or its receptor causes rickets in children and osteomalacia in adults. Nutritional rickets is mostly caused by poor vitamin D status, especially during rapid growth. In some areas of the world a poor calcium intake generates a very similar picture of rickets. Apart from nutritional rickets, genetic defects in vitamin D metabolism or action, genetic or acquired defects in phosphate homeostasis resulting in hypophosphatemia, or a series of genetic defects involved in mineral deposition (such as deficiency of alkaline phosphatase or excess pyrophosphate) are rare causes of rickets.[16] Nutritional rickets can be effectively prevented by appropriate vitamin D and calcium intake. However, nutritional rickets is still frequent in many countries or in high-risk groups.[1,17] There is fairly general consensus on how rickets could be eradicated, as daily supplementation with 400 IU of vitamin D (usually D_3) has been proven to eradicate rickets in countries with a systematic supplementation policy (reviewed in Bouillon & Antonio[1]). The minimal calcium requirements are age-dependent and well described in a consensus document.[18] The implementation of an adequate strategy is still poor or slow, and a dedicated program led by scientists, healthcare policy makers, and especially the World Health Organization (WHO), United Nations Children's Emergency Fund, and major charities is long overdue.[1,18,19] A less well-recognized consequence of vitamin D deficiency early in life deals with teeth development. Impaired mineralization of dentin in postnatal life results in tooth decay or tooth loss. As enamel, the hardest tissue of the body, is already deposited during late pregnancy, improvement of vitamin D status in the perinatal period may improve tooth (enamel and dentin) development, with permanent consequences for tooth health. There is much more debate on the role of vitamin D status on bone density and risks of fracture in postmenopausal or elderly subjects. A very large number of RCTs did not find a beneficial effect of calcium supplementation on fractures. Similarly, supplementation with vitamin D alone did not reduce fracture risk. In a highly publicized paper published in *Lancet Diabetes Endocrinology* in 2018,[20] the authors concluded that "there is little justification to use vitamin D supplements to maintain or improve musculoskeletal health." In their metaanalysis of 81 RCTs ($n = 53{,}537$ participants) that reported data on fracture ($n = 42$), falls ($n = 37$), or bone mineral density (BMD) ($n = 41$), vitamin D supplementation had no effect on (1) total fracture (36 trials; $n = 44{,}790$, relative risk [RR] 1.00, 95% confidence interval [CI]: 0.93–1.07), (2) hip fracture (20 trials; $n = 36{,}655$, RR 1.11, 95% CI: 0.97–1.26), or (3) falls (37 trials; $n = 34{,}144$, RR 0.97, 05% CI: 0.93–1.02). Results were similar in RCTs of high-dose versus low-dose vitamin D. However, the majority of the included studies lasted less than 12 months (and thus were unable to evaluate fracture outcome), the metaanalysis included two studies using very high bolus doses of vitamin D known to transiently increase fractures or falls, and a very large number of studies were performed in subjects with adequate vitamin D status at baseline. Above all, the authors deliberately did not include studies dealing with combined treatment with vitamin D and calcium supplements versus double placebo. However, in 2014 one of the authors reported[21] that vitamin D plus calcium versus placebo or no treatment resulted in a statistically significant reduction in (1) risk of hip fracture (9 trials; $n = 49{,}853$; RR 0.84; 95% CI: 0.74–0.96), (2) incidence of new nonvertebral fractures (8 trials; $n = 10{,}380$; RR 0.86; 95% CI: 0.78–0.96), and (3) incidence of any fracture (10 trials; $n = 49{,}976$; RR 0.95; 95% CI: 0.9–0.99). In addition, one of the other authors of the *Lancet Diabetes Endocrinology* paper concluded independently[22] that vitamin D plus calcium reduced hip fracture in institutionalized individuals (2 trials; $n = 3{,}853$) but did not reduce the risk of hip fracture by 15% or more in community-dwelling individuals (7 trials; $n = 46{,}237$). However, older studies, mostly performed in elderly women with rather poor vitamin D status and lower than optimal calcium intake, demonstrated that vitamin D supplementation alone did not reduce fracture risks, but that a combination of vitamin D and calcium supplementation reduced the risk of hip or nonvertebral fractures by 10% to 20% (reviewed in Weaver et al.[23] and Bouillon & Bikle[2]). The recent VITamin D and OmegA-3 TriaL (VITAL)[24] and the Vitamin

D Assessment Study (ViDA)[25,26] did not find beneficial effects on fractures, but these studies mostly included vitamin D–replete adults with generally good calcium intake. In a subanalysis, BMD increased in both studies relative to the placebo group in subjects with low total or free 25OHD. The Calgary study intended to explore the effects and safety of daily high-dose vitamin D (4000 and 10,000 IU vs. 400 IU) for 3 years on bone structure and quality (measured by high resolution peripheral quantitative computed tomography [HRpQCT]) in 311 Canadian adults.[27,28] Surprisingly, BMD at the radius and tibia decreased by 3.5% and 1.7%, respectively, in the 10,000 IU/d group, whereas the decrease at both sites was nonsignificantly lower in the 4000 IU/d group compared with the 400 IU/d group. This study does demonstrate that vitamin D supplementation of vitamin D–replete adults (baseline serum 25OHD of ~30 ng/mL) does not improve bone quality or quality. Moreover, very high doses may even have negative effects, as a small percentage of subjects developed hypercalciuria or hypercalcemia that quickly resolved after adjustment of dosing. Of course, this may imply that regular follow-up is desirable when using such dosages.[27,29]

The interpretation of older and newer data suggests that:

1. Vitamin D supplementation of subjects with a poor vitamin D status at baseline (<12–20 ng/mL) reduces the age-related loss of bone density.
2. Combined correction of poor vitamin D status and low calcium intake is recommended by most international guidelines, especially for the elderly population, and seems to be able to modestly reduce the risk of fracture. Most guidelines recommend a daily dose of 800 IU of vitamin D_3. The European Calcified Tissue society also recently supported this conclusion.[30]
3. Vitamin D supplementation (equivalent to 2000–3000 IU/d) given to vitamin D–replete adults does not generate additional bone health benefits.
4. High pulse doses of vitamin D (>300,000 IU) should be avoided because they may be associated with transient increased risk of fracture.[31]
5. Based on the results from the Calgary studies,[27,28] high daily doses (4000–10,000 IU) of vitamin D may result in increased urinary calcium excretion and sometimes even hypercalcemia, whereas BMD, as measured by the highly sensitive HRpQCT method may even decrease, rather than improve.
6. As most trials with active osteoporosis drugs (antiresorptive and others) added a calcium and vitamin D supplement in the active and control groups, the logic of evidence-based medicine requires that the active drug should be given to subjects with a good vitamin D and calcium status, and in the majority of cases this will require a daily supplement.

KEY POINTS

- Absence of vitamin D, its active metabolite 1,25(OH)2D, or vitamin D receptor causes rickets in infants or children, characterized by excess unmineralized bone (osteoid), too little solid mineralized bone, and an extremely delayed mineralization rate. Nutritional rickets is either as a result of too little vitamin D or too little calcium intake/absorption and can/must be prevented by appropriate access to vitamin D (~400 IU or 10 ug/d) and calcium (age-dependent). Unfortunately, nutritional rickets is still endemic in some parts of the world and in specific risk groups. Severe vitamin D deficiency in adults causes osteomalacia.
- A poor vitamin D status in elderly subjects may accelerate osteoporosis, whereas adequate vitamin D and calcium intake can modestly reduce the risk of osteoporotic fractures.

Extraskeletal Actions of the Vitamin D Endocrine System

The virtually ubiquitous expression of VDR in all nucleated cells, the presence of a functional CYP27B1 in at least 10 different tissues apart from the kidney, and the very large number of genes that are under direct or indirect control of $1,25(OH)_2D$ (even early in the evolution of vertebrates, as observed in zebrafish) all point toward a more universal role for the vitamin D endocrine system than just regulation of calcium/phosphate/bone metabolism. This is not totally unexpected, as most other nuclear receptor ligands, such as androgens, estrogens, glucocorticoids, and retinoids, also have a very wide spectrum of activities. Based on a massive number of observations in cells, tissues, and transgenic mice it seems likely that the total absence of vitamin D action may have many extraskeletal effects. This is in line with many observational studies in humans, but the final proof of causality requires data from Mendelian randomization (MR) or RCTs. Large-scale genetic testing has identified several polymorphisms in genes related to vitamin D metabolism or transports, or even genes apparently not related to vitamin D, that predispose to lower serum 25OHD concentrations. A combination of such polymorphisms can thus identify subjects who are genetically predisposed to lower 25OHD concentrations than the general population. Identification of such variants requires very large-scale population studies, and genetic variants can only be used in further analysis if they are highly significantly associated at a genome-wide level ($P < 5 \times 10^{-8}$) with 25OHD. MR then uses this information as an epidemiologic method to test whether genetically lower 25OHD concentrations are associated with increased risk of a specific disease. MR has several advantages in comparison with observational studies or even RCTs. First, MR detects genetic variants that are randomly assigned at conception, which avoids the introduction of confounding factors. This is highly relevant because serum 25OHD concentrations are prone to confounding factors such as differences in climate, lifestyle, or body mass index (BMI). Second, given that alleles are inherited at conception, preceding disease onset, MR avoids the detection of reverse causation (whereby the later disease causes low 25OHD concentrations). Finally, MR evaluates lifelong effects of lower 25OHD, thus extending the observation period well beyond what an observational study or RCT can achieve. The limitations of MR are also obvious because it assumes a linear effect between low 25OHD and health outcome (and thus cannot evaluate a threshold effect). It also assumes that polymorphisms mediate their effects through serum 25OHD levels and have no other direct effect on the outcome studied. Finally, most MR studies published to date have only predict been able to predict up to 5% of the variation in serum 25OHD levels; but this may improve soon.

Skin. Based on studies of *Cyp27b1*-deficient mice, repair of the barrier function of the skin is impaired in the absence of vitamin D action. The major skin phenotype of both *Vdr*-null mice and children with *VDR* mutations, however, is the development of total alopecia. Hair development at birth is normal, but hair loss starts after the first catagen and ultimately leads to alopecia totalis associated with large dermal cysts. This alopecia is caused by the VDR itself because it is not observed in vitamin D–deficient wild-type mice or in mice with *Cyp27b1* mutations. No clear skin disease is linked to simple vitamin D deficiency in humans.

The very same photons that can photoconvert 7-dehydrocholesterol into previtamin D are also able to cause DNA damage and, ultimately, photoaging, increasing the risk for skin cancer, so exposure to the UVB or sunlight needed to produce vitamin D always involves a small but cumulative risk of skin damage. This risk is especially relevant for individuals with a fair skin type (phototypes 1 and 2). There is thus a very diffcult trade-off between ensuring safe exposure to sunlight so as to produce suffcient vitamin while avoiding long-term risks of skin damage and skin cancer. *Vdr*-null mice are more prone to develop skin cancer when exposed to UVB light. Although 1,25(OH)$_2$D and some of its analogs are able to generate a strong photoprotective effect against UVB-mediated events in cultured keratinocytes,[32] the overall prooncogenic effect of UVB is dominant.

Cell Proliferation and Cancer. Exposure to 1,25(OH)$_2$D of virtually all normal cells and even most malignant cells results in accumulation in the G0/G1 phase of the cell cycle. A general downstream effect is the regulation of the E2F family of transcription factors, which act as master switch for a very large number of genes involved in cell cycle progression. These E2F factors are under the control of the retinoblastoma proteins, and their phosphorylation state is regulated by cyclins and cyclin-dependent kinases (p18, p19, p21, or p27), many of which are regulated by 1,25(OH)$_2$D. However, 1,25(OH)$_2$D may also inhibit cell growth by interfering with signaling pathways initiated by transforming growth factor–β, epidermal growth factor, prostaglandins, Wnt ligands, and *c-myc*).[2,33] When exposed to chemocarcinogens, oncogens, or UVB light, *Vdr*-null mice develop more cancers (mammary cancer–type lesions, skin tumors, and lymphomas).[5] In addition, *Cyp27b1*-deficient mice kept alive for more than one year develop diverse types of spontaneous tumors. This was associated with increased oxidative stress, cellular senescence, and senescence-associated secretory phenotype molecules. Furthermore, 1,25(OH)$_2$D$_3$ prevented spontaneous tumor development.[34]

In humans, absolute *VDR* or *CYP27B1* deficiencies are rare, but vitamin D deficiency is very frequent. This obviously raises the question of whether vitamin D deficiency is associated with increased risk of cancer in humans. In most cross-sectional or observational studies, a higher cancer risk was found in subjects with serum 25OHD levels below 20 ng/mL. However, a few studies have also linked better vitamin D nutritional status with a higher prevalence of cancer or more aggressive types of cancer.[35] The final question is, of course, whether serum 25OHD is a predictor or has a causative relation with the overall cancer risk. A few RCTs (three out of 10) suggested some beneficial effects on cancer incidence (reviewed in Bouillon & Bikle[2]). Two recent large RCTs (the New Zealand ViDA[36] and the US VITAL[37] trials) independently evaluated the effects of vitamin D supplementation on cancer incidence and cancer mortality. The VITAL study[37] recruited more than 25,000 adults from 44 centers in the United States. The trial had a two-by-two factorial design to evaluate the potential benefits of vitamin D$_3$ (2000 IU/d) or omega-3 marine fatty acids (1 g/d). Adult males (>50 years) or females (>55 years) were followed for a mean duration of 5.3 years. The primary endpoints were invasive cancer (any type) and major cardiovascular events (a composite of myocardial infarction, stroke, or death from cardiovascular events). Many secondary events were also examined. The mean baseline vitamin D status (serum 25OHD concentration of 30.8 ± 10 ng/mL) was higher than that of the average American adult (based on comparison with the latest NHANES data). Subjects taking vitamin D supplements at baseline had, as expected, higher baseline 25OHD (33 ng/mL) than the others (27 ng/mL). The ViDA study[36] evaluated the effects of monthly high-dose vitamin D supplementation on more than 5000 New Zealand adults followed for a mean duration of 3.3 years. Subjects were randomized to receive either placebo or a dose of 200,000 IU of vitamin D at the start of the study, followed by a monthly dose of 100,000 IU (~3300 IU/d). Most of the participants had a normal vitamin D status at baseline (mean serum 25OHD of 26 ng/mL). Neither study found an effect of vitamin D supplementation on cancer incidence at 3.3 or 5.3 years' follow-up. Combining all vitamin D and cancer studies involving a total of more than 70,000 subjects, therefore, suggests that vitamin D supplementation of mostly vitamin D-replete adults (mostly White people) for approximately 5 years does not change their cancer risk. However, five independent trials suggest, in *post hoc* analysis, potential benefits on cancer mortality if all data are combined or when the follow-up is longer than 4 years.[38] In view of the long latency period for carcinogenesis, the link between a good vitamin D status and cancer incidence or mortality probably requires a very long follow-up time, well beyond what most studies so far have been able to do. MR studies are more appropriate to answer the effect of lifelong vitamin D status on cancer incidence. A large number of such MR studies have looked at this potential link. Overall, the existing MR studies support the protective effect of vitamin D on ovarian cancer,[39] but not on other types of cancer such as breast, prostate, and colorectal cancer, among others.[2]

In conclusion, preclinical data suggest that severe vitamin D deficiency may predispose to carcinogenesis, whereas exposure to high 1,25(OH)$_2$D levels may reduce cell proliferation and induce cell differentiation. Numerous observational studies link poor vitamin D status to increased risk of cancer or poorer prognosis for existing cancers. Many MR studies did not demonstrate an effect of genetically lower serum 25OHD concentrations on the risk of major cancers, although the power of most of these MR studies to predict changes in serum 25OHD was usually below 5% of the variation of serum 25OHD. RCTs dealing with a total population of more than 70,000 mostly White adults, however, did not find a reduced risk of cancer associated with vitamin D supplementation. However, several independent trials suggest, in *post hoc* analysis, potential benefits on cancer mortality. This effect is significant if all data are combined or when the follow-up is longer than 4 years. In view of the long latency period for carcinogenesis, the link between a good vitamin D status and cancer incidence or mortality probably requires a very long follow-up time that is well beyond what most studies so far have been able to do.

There are thus confusing data regarding a possible link between vitamin D status and cancer, and several options are still open:

- Option 1. The vitamin D endocrine system is redundant; the cell cycle is under such complex control that the vitamin D status is largely redundant for the genesis or evolution of human cancers.
- Option 2. Only total or near-total absence of vitamin D action has an influence on cell proliferation and cancer, as shown in animal models of *Vdr* or *Cyp27b1* deficiency. Such very severe vitamin D deficiency is so infrequent in the human population that the benefits of correction of such severe deficiency cannot be picked up in RCTs.
- Option 3. Carcinogenesis is such a slow process that the duration of the RCTs (up to 5–7 years) is insufficient to reveal the true role of the vitamin D endocrine system. Long-term and or severe vitamin D deficiency may play a (minor) role in the pathogenesis or evolution of (some) human cancers. The available data are insufficient to provide a clear answer.

Immune Function and Vitamin D. All immune cells express a functional VDR at certain stages of their differentiation. Antigen-presenting cells (dendritic cells and equivalent resident cells, as well as monocytes/macrophages) can synthesize 1,25(OH)$_2$D using the same enzyme as the kidney but controlled by immune stimuli instead of calciotropic hormones.[40] Finally, 1,25(OH)$_2$D regulates a wide range of genes that play crucial roles in the immune system. Vitamin D metabolites stimulate the innate immune system (largely mediated by monocytes/macrophages) by enhanced expression of a number of critical genes such as those encoding the pattern recognition receptor nucleotide oligomerization domain protein, the Toll-like receptor cofactor CD14, the antimicrobial peptides cathelicidin, defensin beta 2, defensin beta 104B, as well as multiple cytokines and chemokines.[2,40,41] A recent analysis[42] evaluated 10,933 participants of 25 RCTs and found a significant overall reduction in acute respiratory infections (odds ratio [OR] 0.88, 95%CI: 0.81–0.96) with the greatest benefits in subjects with severe vitamin D deficiency (<10 ng/mL) at baseline (OR 0.58, 95% CI: 0.40–0.82), with a number needed to treat for benefit of eight subjects. Subgroup analysis revealed that intermittent (monthly or less frequent) doses of vitamin D did not generate protection, whereas daily or weekly vitamin D supplementation was more effective (OR 0.81, 95% CI: 0.072-0.91). In the ViDA trial, however, no effects of vitamin D supplementation were found in acute respiratory infections in older adults.[43] This is not a total surprise, as the lack of effects may be attributed to the intermittent dosing and/or adequate vitamin D status at baseline, and thereby not contradict the Martineau 2019 metaanalysis. Tuberculosis is frequently associated with poor vitamin D status, and vitamin D is also able to accelerate the killing of intracellular mycobacteria.[44,45] Some, but not all intervention studies suggested a more rapid clearance of tuberculosis with correction of vitamin D deficiency.[46] In a large study of severely vitamin D-deficient children, however, vitamin D supplementation did not reduce the risk of developing active tuberculosis.[47] A total of 8851 children underwent randomization, and all had a baseline serum 25OHD level of less than 20 ng/mL. The mean 25OHD level at the end of the trial was 31 ng/mL in the vitamin D group and 11 ng/mL in the placebo group. The two groups received a weekly oral dose of either 14,000 IU of vitamin D_3 or placebo for 3 years. Tuberculosis was diagnosed in 21 children in the vitamin D group and in 25 children in the placebo group (adjusted RR 0.87; 95% CI: 0.49–1.55). A total of 29 children in the vitamin D group and 34 children in the placebo group were hospitalized for treatment of acute respiratory infection (adjusted RR 0.86; 95% CI: 0.52–1.40). The incidence of adverse events did not differ significantly between the two groups. Vitamin D supplementation therefore did not result in a lower risk of tuberculosis infection, tuberculosis disease, or acute respiratory infection compared with placebo among vitamin D-deficient schoolchildren in Mongolia. Whether this was related to prior tuberculin vaccination is unknown.[47a]

The acquired immune system reacts in opposite way than does the native immune system. Indeed, $1{,}25(OH)_2D$ inhibits dendritic cell maturation and generates coordinated action on T cell gene expression of key cytokines (IL-1, IL-2, IL-12, IL-17, INF-γ) and genes needed for antigen presentation to T cells (MHC class II and cosignaling proteins). Secondly, in collaboration with dendritic cells, $1{,}25(OH)_2D$ decreases the activity of Th1 and Th17 cells, while upregulating regulatory T cells.[2,48] This should have beneficial effects on the occurrence or evolution of autoimmune diseases. Vitamin D-deficient mice more easily develop a more severe type of such autoimmune diseases. For example, genetically predisposed NOD mice exposed to transient vitamin D deficiency early in life have a much higher incidence of T1DM than vitamin D-replete mice,[49] and *Cyp27b1*-deficient mice are more prone to several types of inflammatory bowel disease. Moreover, $1{,}25(OH)_2D$ and more potent and selective analogs can significantly reduce spontaneous or experimental autoimmune diseases in rodents, such as T1DM in NOD mice and experimental allergic encephalitis, nephritis, or inflammatory bowel disease in rodents. Studies of human autoimmune diseases are, however, more complicated. Polymorphisms of the *CYP27B1* gene and other genes involved in the vitamin D endocrine system are associated with a risk of T1DM in cross-sectional, as well as in family, studies. Consistent with data from NOD mice, several epidemiologic studies in humans report that vitamin D supplementation early in life may reduce later risk of T1DM. Risk reduction varied between 26% with cod liver oil to 78% with 2000 IU/d, with an overall effect of 30% reduction in five published reports.

A low vitamin D status has also been repeatedly associated with a higher risk for multiple sclerosis. A large prospective, nested, case-control study among more than 7 million US military personnel revealed that low 25OHD level was a strong risk factor for later occurrence of multiple sclerosis (OR of ~2 for serum 25(OH)D <20 ng/mL, with possibly even greater "protection" by higher levels).[50] However, as for T1DM, no controlled, randomized intervention studies have proved yet a causal relation between vitamin D (deficiency or insufficiency) and later occurrence of autoimmune diseases, but randomized trials (especially in genetically at-risk groups) should receive high priority. MR studies, despite their inherent limitations, are the best proxy for RCTs, as they enable evaluation of the consequences of lifelong lower vitamin D status on specific outcomes. Four independent MR studies revealed that genetically predicted lower serum 25OHD concentrations are associated with a higher risk of juvenile or adult-onset multiple sclerosis (reviewed in Bouillon & Bikle[2] and Jacobs et al.[51]). All these observations, therefore, suggest that preventing vitamin D deficiency in the perinatal period, early childhood, or adolescence may have long-lasting effects on autoimmune diseases. Moreover, there are several observations in mouse models that could explain such effects, such as the increased apoptosis of (potentially autoreactive) thymocytes or lymphocytes and the generation of regulatory T cells and natural killer T cells after exposure to $1{,}25(OH)_2D$. Several studies have looked at the effects of high-dose vitamin D supplementation (up to 40,000 IU/d) as add-on therapy with the aim of slowing the progression of multiple sclerosis, but no clear beneficial effects on progression of disability or relapse–free rate have been observed so far.[52,53] Similarly, as for multiple sclerosis, there are many suggestions that inflammatory bowel disease is linked with vitamin D deficiency, but a high-dose supplementation study did not reveal any benefit regarding the cumulative postoperative recurrence of Crohn disease.[54]

Cardiovascular System. *Vd*- null mice, as well as *Cyp27b1*-null mice, develop high-renin hypertension and cardiac hypertrophy that can be prevented by treatment with an angiotensin blocker. $1{,}25(OH)_2D$ decreases renin production, probably by direct regulation of gene expression via a VDRE in its promoter. Many other genes such as plasminogen activator inhibitor and thrombomodulin are also under the control of $1{,}25(OH)_2D$. Several *in vitro* studies have also confirmed beneficial effects on vascular smooth muscle cells and endothelial cells and thereby suggest that vitamin D may have overall beneficial cardiovascular effects (reviewed in Bouillon[55]). However, exposure to excess $1{,}25(OH)_2D$ may result in transdifferentiation of vascular smooth muscle cells into osteoblast-like cells and cause calcification of the vascular wall. Observational studies are consistent with these preclinical data. In normotensive and hypertensive individuals, there is an inverse association between 25OHD concentration and cardiovascular risk factors (such as hypertension) and cardiovascular events (ischemic heart disease, cardiomyopathy, stroke, and cardiovascular mortality).[56,57] Indeed, a metaanalysis of nearly 850,000 individuals concluded that subjects with the lowest serum 25OHD concentrations were at increased risk (RR 1.43) of cardiovascular events.[56] Based on such observational data, a 10 ng/mL higher serum 25OHD was associated with a RR of 0.90 (95% CI: 0.86–0.94), but the RR was not linear, as the lowest risk of major cardiovascular diseases (including heart failure) was found in subjects with serum 25OHD between 20 and 25 ng/mL.[57] Many risk factors of cardiovascular diseases such as hypertension, obesity, diabetes, lipid profile, smoking, physical activity, and other lifestyle aspects are also linked to vitamin D status, either causally or by reverse causality (e.g., lower physical activity and sun exposure resulting in poorer vitamin D status). Therefore, observational links are only hypothesis-generating and need confirmation by controlled supplementation studies. The results of randomized trials have been inconsistent, but a 2015 metaanalysis of 46 trials including 4541 participants found no significant effects of vitamin D supplementation on systolic or diastolic blood pressure.[58] The New Zealand ViDA RCT did not find a beneficial effect on peripheral blood pressure with a monthly dose of 100,000 IU of vitamin D compared with placebo. In a more in-depth subanalysis, a significant decrease in central blood pressure was found in a small group of subjects with baseline vitamin D deficiency.[59] In humans, a low vitamin D status is associated with a number of cardiovascular risk factors, including all aspects of the metabolic syndrome. A metaanalysis of 19 prospective studies (65,994 patients) revealed an inverse relationship between serum 25OHD levels (8–24 ng/mL [20–60 nmol/L]) and risk of cardiovascular disease (RR 1.03, 95% CI: 1.00–1.60, per 10-ng/mL [25 nmol/L] decrement in serum 25OHD).[60] Randomized trials, however, could not demonstrate a persistent effect of vitamin D supplementation on cardiovascular outcomes, including myocardial infarction and stroke, lipids, glucose, or blood pressure. Two recent megatrials (VITAL[37] and ViDA[61]) evaluated the effect of vitamin D supplementation on a set of cardiovascular events. Combining all RCTs involving more than 80,000 subjects and up to 5 years of follow-up, vitamin D supplementation of mainly vitamin D-replete subjects generated no cardiovascular benefits or harm.[55,62,63] Patients with chronic renal failure have a high frequency of cardiovascular problems, whereby most patients suffer from combined deficiency of 25OHD and 1,25$(OH)_2$D. Whether correction of their vitamin D status would improve their outcome is still debatable, but such patients may well be an excellent target group to study the role of the vitamin D endocrine system in a high-risk group of combined vitamin D deficiency and cardiovascular mortality.

Muscle, Muscle Function, and Falls. VDR is expressed in myoblasts and stellate cells but is probably absent in mature striated muscle cells. *Vdr*-null mice show maturation problems of their muscle fibers, with smaller muscle fibers and expression of embryonic markers even after weaning.[64] Genes that are typically expressed early in life (e.g., *myf-5*) are under negative control by 1,25$(OH)_2$D. Muscle-specific deletion of *Vdr* results in sarcopenia and impaired muscle function.[65] Patients with chronic renal failure and vitamin D deficiency (and thus combined deficiency of 25OHD and 1,25$(OH)_2$D) can develop severe myopathy and an inability to walk that can be promptly restored by appropriate vitamin D and/or analog treatment, similar to effects seen in patients with congenital *CYP27B1* deficiency. Observational studies show a relationship between low (<20 ng/mL, and especially <10 ng/mL) 25OHD levels and muscle weakness in children and elderly subjects.[66] Sarcopenia (progressive loss of muscle mass and strength) is highly prevalent in the elderly and is frequently associated with vitamin D deficiency. Many intervention trials have yielded inconsistent results on muscle function and body sway.[2] Based on results from 30 RCTs, Beaudart et al.[67] concluded that muscle strength modestly improved with vitamin D supplementation, especially if it was given to elderly subjects with severe vitamin D deficiency (<12 ng/mL) at baseline. Two Cochrane reviews extensively examined the effects of vitamin D supplementation on falls. One review concluded that vitamin D supplementation reduced the risk of falls in institutionalized patients (RR 0.63, 95% CI: 0.46–0.85). In ambulatory subjects a reduced risk of falls and fallers was only found in vitamin D-deficient and not in vitamin D-replete subjects. A recent *post hoc* analysis of the large New Zealand ViDA study, however, did not find a beneficial effect on the risk of falls with monthly doses of 100,000 IU of vitamin D.[26] This conclusion does not totally contradict the previous results, as the ViDA participants were mainly vitamin D-replete adults. Similarly, the US VITAL trial reported no effects of daily supplementation with 2000 IU of vitamin D on fall risks.[24] High-dose vitamin D supplementation, however, may increase the risk of falling. This was first observed by Sanders et al.[68] in elderly subjects treated with a yearly dose of 500,000 IU of vitamin D. The vitamin D-treated group had significantly more falls and fractures during the first 3 months after each loading dose. In addition, high monthly doses of vitamin D (with or without 25OHD) also increased the risk of falls in the high-dose groups.[69] Other studies confirmed a fall risk owing to high-dose vitamin D.[70]

In conclusion, there is at present no consensus regarding the potential effects of vitamin supplementation on muscle function, balance, and risks of falls. Older studies performed in elderly, mostly vitamin D-deficient subjects, suggested a small reduction in risk of falls. Vitamin D supplementation of vitamin D-replete New Zealand or US adults did not convey benefits or risk. Finally, a few studies reported that high intermittent doses or even simply high serum 25OHD may enhance the risk of falls in elderly subjects. Therefore, it seems that correction of severe vitamin D deficiency with modest regular doses of vitamin D is desirable.

Diabetes and Obesity. Vitamin D deficiency in experimental animals (rodents and rabbits) impairs glucose tolerance, but this is not consistently confirmed in *Vdr*-null mice. Many observational cross-sectional or prospective studies suggest a close link between low vitamin D status and T2DM. As example, a metaanalysis of 21 prospective studies involving 76,220 participants and 4996 incident T2DM cases showed a RR of 0.62 (95% CI: 0.54–0.70) of T2DM in subjects with serum 25OHD concentrations above 25 ng/mL compared with subjects having serum 25OHD concentrations below 12 ng/mL.[71] Small-scale intervention studies provide inconsistent results, sometimes with small benefits for one aspect of insulin sensitivity or glucose homeostasis, sometimes with null results, as summarized in a systematic review.[72] The large US D2d study was designed to investigate this relationship and to evaluate the effects of a dose of vitamin D (4000 IU/d) for a mean duration of 2.5 years on the conversion of prediabetes to T2DM. The criteria used to define prediabetes (fasting plasma glucose between 100 and 125 mg/dL or a 2-hour post-oral glucose tolerance test plasma level between 140 and 199 mg/dL or a glycated hemoglobin level between 5.7% and 6.4%) and T2DM were in line with the American Diabetes Association guidelines. Most study participants were vitamin D-replete at baseline, as their mean baseline serum 25OHD concentration was approximately 28 ng/mL.[73] In the intention to treat analysis, the hazard ratio (HR) to develop T2DM in the vitamin D group was 0.88 (95% CI: 0.75–1.04; $P = 0.12$). The same conclusion was reached for subgroups of participants (according to race or ethnicity, BMI, glycemic baseline criteria, or gender). In a *post hoc* nonpreplanned analysis of a small ($n = 103$) subgroup of subjects with severe vitamin D deficiency at baseline (serum 25OHD <12 ng/mL), the HR to progress to T2DM was more than halved (HR 0.38, 95% CI: 0.18–0.80). In an exploratory per-protocol analysis censoring data from participants who did not follow the protocol exactly, the HR to progress to T2DM was 0.84 (95% CI: 0.71–1.00). In subjects with a baseline BMI below 30 (approximately one-third of all participants), the HR to progress to T2DM was just significant at 0.71 (95% CI: 0.53–0.95). These are all subgroup analyses that should not be considered confirmatory, as multiple statistical adjustments were required. Mice with a deletion of *Vdr* or *Cyp27b1* develop a lower fat mass owing to increased energy expenditure.[74] In contrast, most observational studies in humans link vitamin D insufficiency to nearly all aspects of the metabolic syndrome—including obesity, insulin resistance, fasting blood glucose or T2DM, hypertension, and hyperlipidemia. Intervention studies, however, had no effects on body weight evolution. Low serum 25OHD concentration in obese or diabetic subjects may well be attributed to the decreased expression of hepatic CYP2R1, as shown in mouse models (reviewed in Bouillon et al.[3]). Indeed, hepatic Cyp2r1 expression (at the mRNA and protein levels) and activity are markedly reduced in mice with diet-induced obesity or diabetes. If this also applies to human obesity/T2DM, it may well explain the discrepancy between strong observation data and the results from RCTs.

Reproduction. Serum 25OHD concentrations are frequently low in pregnant women. A Cochrane analysis revealed that vitamin D supplementation in pregnant women has some beneficial effects on maternal health (preeclampsia and gestational diabetes) and the health of their offspring (low birth weight).[75] An independent overview of 43 systemic reviews[76] dealing with 204 original studies concluded that observational data suggested a link between poor vitamin D status and maternal outcomes (preeclampsia or gestational diabetes) or fetal outcome (small for gestational age). However, this was not confirmed in an overview of intervention studies, except for a slightly increased risk of giving birth to a child who was small for gestational age. A detailed analysis focusing on preeclampsia concluded that vitamin D supplementation dose dependently decreased the risk of preeclampsia (OR 0.35, 95% CI: 0.25–0.5).[77] However, a large supplementation study in pregnant Bangladeshi women with severe vitamin D deficiency did not reveal beneficial effects.[78] Therefore, it is not yet possible to draw a definitive conclusion regarding vitamin D supplementation during pregnancy. In view of the high frequency of poor vitamin D status before or during pregnancy, further studies on the possible benefits of vitamin D supplementation during pregnancy or in the perinatal period should receive high priority.

Safety. Doses of vitamin D below the upper limit defined by the Institute of Medicine (IOM) or European Food Safety Agency (EFSA) seem to be safe. A small increase in the risk of kidney stones was observed in the Women's Health Initiative study, but not in the more recent large-scale studies. No changes in kidney function were reported. Very high single-bolus doses (>300,000 IU), however, have been associated with a transient increase in risk of falls or fractures and should be avoided.[79] Additional studies suggest that high continuous doses of vitamin D given to elderly subjects, increasing their 25OHD well above 45 ng/mL, may increase their risk of falls.[69] Individuals with *CYP24A1* mutations are of course very sensitive to toxicity, as mentioned earlier (in the section on Metabolism) and should avoid all exposure to vitamin D or even extensive sunlight because of the risk of hypercalcemia, kidney stones, or nephrocalcinosis. An extensive safety study (Calgary study) showed that daily doses between 4000 and 10,000 IU/d may occasionally result in hypercalciuria or hypercalcemia and even may modestly decrease BMD. No increase in vascular calcifications was seen in subjects receiving 4000 to 10,000 IU/d during the 3 years of follow-up in the Calgary study.[28] Higher doses of vitamin D (or vitamin D metabolites) may cause classical calcemic toxicity (Table 50.1), and such signs and symptoms are usually only seen when serum 25OHD is well above 100 ng/mL.[80]

Mortality. Because vitamin D status is associated with so many major diseases, it is worth exploring whether vitamin D deficiency is associated with increased mortality. NHANES studies, along with similar studies in Europe, demonstrate that low serum 25OHD concentrations (especially <20 ng/mL) are associated with increased mortality, even when deaths during the first 3 years of follow-up were removed from the dataset to avoid detecting reverse causality.[81] The recent megatrials did not demonstrate an effect on mortality, but the subjects were younger and more vitamin D–replete at baseline than in previous studies summarized in a Cochrane analysis showing a modest decrease in mortality in association with vitamin D_3 supplementation (RR 0.94, *P* = .002).[82]

KEY POINTS

- A large set of preclinical data suggests that the vitamin D endocrine system may influence cell proliferation and cancer, as well as the immune and cardiovascular systems and many other tissues. Observation studies are largely in line with these data and suggest that poor vitamin D status is linked to major human diseases such as cancer, infections, and autoimmune, cardiovascular ,and metabolic diseases. Large randomized trials, however, have failed to confirm the beneficial effects of vitamin D supplementation, although most subjects were not severely vitamin-deficient at baseline. Mendelian randomization studies have produced largely "null" results, except for a consistent link between genetically low 25OHD status and multiple sclerosis.

CLINICAL USE OF VITAMIN D

To define the use of vitamin D or its metabolites or analogs, we need to define what should be measured to evaluate the vitamin D status, how to define the optimal vitamin D status, and finally who should receive vitamin D supplements.

Assays for Vitamin D and Metabolites: Methodology and Applications

Vitamin D and approximately 30 of its metabolites are found in plasma. The measurement of serum concentrations of vitamin D_2 or D_3 is of little clinical value because it only reflects recent exposure to UV light or nutritional intake. Serum 25OHD concentration rather than serum vitamin D_3 is an excellent reflection of the vitamin D status because of the rapid conversion of vitamin D into 25OHD and its long plasma half-life.[12] Its plasma concentration varies widely in normal subjects because of large variations in endogenous and exogenous vitamin D supplies. Plasma 25OHD thus behaves as a true vitamin, in that its concentration depends on nutritional supply or synthesis in the skin after exposure to UVB light. Low exposure to UV light and low vitamin D intake are quite common in infants and elderly subjects if food sources are not supplemented with vitamin D. Plasma 25OHD concentrations are indeed low at birth (usually approximately half the maternal concentration because of the 2:1 ratio of maternal to fetal DBP concentration), and the natural vitamin D content of milk is low. During the evolution of vertebrates and humans, sun exposure was therefore the natural solution to prevent rickets. Members of African tribes that retain their traditional lifestyles have a mean serum 25OHD of approximately 45 ng/mL.

The metabolism of 25OHD into $1,25(OH)_2D$ and $24,25(OH)_2D$ is tightly controlled by hormones, ions, and humoral factors. The plasma concentration of $1,25(OH)_2D$ is therefore regulated as a true hormone, and measurements of its concentration can be useful for clinical exploration of unusual cases of rickets, osteopenia, and hypo- or hypercalcemia. The serum concentration of $24,25(OH)_2D$ usually reflects the concentration of 25OHD, and an abnormal $24,25(OH)_2D$:25OHD ratio is the best parameter for biochemically defining *CYP24A1* mutations. Measurements of the serum concentration of several vitamin D metabolites are essential for clinical or research purposes. Most laboratories prefer routine immunoassays for measurement of 25OHD and $1,25(OH)_2D$. The quality control of such assays has substantially improved thanks to interlaboratory quality-control studies. The major problem with routine 25OHD assays is accuracy, as vitamin D deficiency is defined by absolute concentration. Tandem mass spectrometry after liquid chromatography is now the gold standard for measuring 25OHD and other vitamin D metabolites. Above all, accuracy should be continuously monitored using validated control samples.[83] Whether free 25OHD or free 1,25$(OH)_2D$ is superior to total concentrations for clinical purposes is still an open question.[84] In countries where vitamin D_2 is (frequently) used, $25OHD_2$ should be measured as well as $25OHD_3$. In the case of immunoassays, this implies that the antibodies used should not discriminate between the two metabolites. When measured by tandem mass spectroscopy, then the results are preferably expressed as serum $25OHD_3$ and $25OHD_2$, both separately and combined. Because 3-epi-25OHD can make up a large percentage of total 25OHD, especially in younger children, care should be taken to minimize or eliminate the interference of this low-activity metabolite. The serum concentration of 25OHD varies widely between and even within the same population. Table 50.2 summarizes a number of factors influencing serum 25OHD. Serum 25OHD concentrations are higher in subjects with a higher vitamin D intake (e.g., higher serum 25OHD in Scandinavian countries because of higher consumption of oily fish) or higher exposure (e.g., as in Masai tribes or Brazil) or lower exposure (e.g., in many countries in North Africa, Middle East, or the Gulf states, or in countries with high latitude where no effective vitamin D synthesis is possible for many months) to UVB sunlight. Mismatch between skin color and sunlight hours (owing to migration) also markedly influences serum 25OHD. Sunscreen use may decrease endogenous vitamin D synthesis, but most subjects do not use the required amount of sunscreen or stay exposed to sunlight for longer than the recommended time. Therefore, in real-life situations, the use of sunscreen has very limited influence on serum 25OHD concentrations. True excess levels of 25OHD are mainly as a result of intake of pharmaceutical amounts of vitamin D or because of rare mutations of *CYP24A1*. Too-low serum 25OHD concentrations can be the result of abnormal metabolism (Table 50.2) or to a number of diseases that cause decreased intestinal absorption or enhanced urinary loss of vitamin D (metabolites). Table 50.2 also lists risk groups for poor vitamin D status. The thresholds defining vitamin D deficiency or excess are described in the following section.

The serum concentration of 1,25$(OH)_2D$ (and especially its free concentration) is feedback-regulated and is therefore more within a narrower range in normal subjects compared with serum 25OHD. Decreased serum 1,25$(OH)_2D$ concentrations are mainly attributed to substrate deficiency (as in nutritional rickets), genetic or acquired CYP27B1 deficiency, hormonal diseases (such as too little PTH or excess FGF-23 activity), or decreased serum DBP concentrations (Table 50.3). Excess 1,25$(OH)_2D$ is rarely attributed to substrate excess and mostly the consequence of hormonal imbalances (PTH excess/FGF-23 deficiency), DBP excess, or rarely vitamin D resistance (VDR mutations) (Table 50.3).

TABLE 50.2 **Serum Concentrations of 25OHD**

Normal Fluctuation According to:

Dietary intake (+)[a]
Sun (ultraviolet [UV] light) exposure (+) influenced by seasonal lifestyle and cultural habits
Age (–)
Skin pigmentation (–)
Latitude (–)
Sunscreen use (–)

Increased 25OHD Concentration

Exposure to pharmaceutical vitamin D[b]
Excess exposure to nutritional vitamin D
Excess exposure to UV light
CYP24A1 deficiency

Decreased 25OHD Concentration

- Genetic or acquired deficiency of CYP2R1
- Genetic or acquired excess of CYP4A3
- Combined deficiency of access/exposure to nutritional vitamin D and UV light
  - Major risk groups include:
    - Infants, especially when born in late winter
    - Women and children of immigrants with pigmented skin living in temperate climates
    - Elderly individuals with limited mobility
    - Subset of the population with low exposure to sunlight because of socioeconomic, religious, or cultural reasons
- Decreased intestinal absorption of vitamin D associated with fat malabsorption (e.g., associated with biliary cirrhosis)
- Short bowel syndrome
- Exocrine pancreas insufficiency
- Gluten enteropathy
- Increased loss or catabolism of vitamin D
- Nephrotic syndrome
- Chronic liver P450 activation by drugs (e.g., barbiturates or antiepileptic drugs)
- Low calcium intake or absorption

[a] Positive or negative effects are indicated by + or –, respectively.

[b] Vitamin D toxicity with hypercalciuria, hypercalcemia, nephrocalcinosis, kidney stones, metastatic calcification, etc., is only observed if 25OHD concentrations exceed 100 ng/mL. Without access to pharmaceutical vitamin D, it is therefore uncommon to develop clinical vitamin D toxicity.

Optimal Serum 25OHD Concentration

Vitamin D is a threshold nutrient, and thus a too-low supply of the substrate creates a deficiency of the active hormone and thus of VDR activation. Over the last decades, and even last century, the optimal vitamin D status has been a point of intense discussion. Many governmental and scientific societies have brought together a very large number of experts to define the minimal threshold for vitamin D status based on serum 25OHD concentrations. All guidelines are unanimous in defining serum 25OHD concentrations below 10 or 12 ng/mL as severe deficiency with increased risk for clinical consequences such as rickets or bone problems in general at any age (Fig. 50.2). This threshold is also in line with the intestinal absorption of calcium as defined by studies using the best technology for the estimation of intestinal calcium absorption. Indeed, calcium absorption and serum 1,25$(OH)_2D$ are markedly decreased when serum 25OHD drops below 5 ng/mL, and no further increase is observed in otherwise healthy subjects once serum 25OHD exceeds 10 to 12 ng/mL.[35,85] Apart from the UK guidelines (Scientific Advisory Committee on Nutrition [SACN]),[86] most other guidelines recommend a serum 25OHD concentration of at least 20 ng/mL for optimal bone health so as to avoid compensatory homeostatic mechanisms (mainly increased serum PTH). Most of these guidelines extensively describe the arguments for this choice, with specific focus on the results of randomized trials rather than on observational data. In addition, IOM adds another dimension to the discussion by arguing that serum 25OHD concentrations of 16 ng/mL would be sufficient for approximately half of the healthy North American population, and that 20 ng/mL would be sufficient

TABLE 50.3 **Serum Concentration of $1,25(OH)_2D$**

Decreased Concentrations	Increased Concentrations
Substrate Deficiency[a]	**Substrate Excess**[b]
(e.g., nutritional rickets, intestinal malabsorption)	
25OHD-1α-Hydroxylase Enzyme Deficiency	**25OHD-1α-Hydroxylase Enzyme Excess**
Inborn: Vitamin D–dependent rickets	Primary or tertiary hyperparathyroidism
Organic: Renal insufficiency or anephric patients	Hypothyroidism
Functional:	Glucocorticoid excess
Hypoparathyroidism	Acromegaly
Pseudohypoparathyroidism	Granulomatous diseases
Hypomagnesemia	Idiopathic hypercalciuria
Tumoral osteomalacia	Hypophosphatemic rickets type 2 (+hypercalciuria)
Hypercalcemia of malignancy	Pregnancy
Hyperthyroidism	Nutritional calcium deficiency
Morbus Addison (acute)	Williams syndrome
Severe insulin deficiency	
X-linked hypophosphatemia	
Rhabdomyolysis	
Tumoral calcinosis	
DBP Deficiency	**DBP Excess**
Fetus	Pregnancy
Nephrotic syndrome	Oral estrogen use
Liver cirrhosis	
Genetic mutation	
	End-Organ Resistance
	True vitamin D resistance (so-called vitamin D-dependent rickets type 2)

DBP, Vitamin D–binding protein.

[a] In many cases of rickets or osteomalacia, $1,25(OH)_2D$ concentrations are still measurable or even nearly normal. This may be as a result of recent (and insufficient) access to vitamin D after long-term vitamin D deficiency. Nevertheless, such concentrations are too low in comparison with the degree of secondary hyperparathyroidism. Regardless, 25(OH)D is a better marker for vitamin D deficiency than $1,25(OH)_2D$. A similar situation is observed in hypothyroidism, as the precursor hormone T_4 is a better marker for clinical hypothyroidism than the real hormone, T_3.

[b] Vitamin D excess only increases serum $1,25(OH)_2D$ when renal function remains normal and/or parathyroid hormone secretion is elevated. Frequently, $1,25(OH)_2D$ levels are low or normal in vitamin D toxicity.

for 97.5% of the population.[87] The Nordic (Scandinavian),[88] DACH (German-speaking countries),[89] and EFSA[90] guidelines also support this threshold of 20 ng/mL. Several scientific societies also supported 20 ng/mL as the definition of "normal vitamin D status or vitamin D sufficiency" (European Calcified Tissue Society, National Osteoporosis Society,[30,30a]). The Dutch Health council considered 12 ng/mL sufficient to prevent rickets and osteomalacia and could not find strong, hard data to recommend 20 ng/mL for adults, except for the elderly, and therefore recommended a mixture of SACN and IOM targets for all children and adults (SACN target), except for the elderly (IOM target). The Endocrine Society (US) and a few other scientific societies, however, recommend serum 25OHD concentrations above 30 ng/mL as being the minimal concentration to define vitamin D sufficiency. Thus, they define levels between 20 and 30 ng/mL as insufficient for optimal health in general.[91] A few regional scientific organizations, such as the Central European countries and the Italian Endocrinologists, support this interpretation. Some grass-root organizations recommend still higher 25OHD concentrations such as those found in subjects with extensive daily sun exposure (such as the Masai tribe in Africa), which are presumably similar to those of early humans.[92] There is no general agreement about the safe upper level of 25OHD. Values above 100 ng/mL (and especially >250 ng/mL) have been associated with hypercalcemia and its health implications (Table 50.1), and are thus to be avoided. Values above 50 ng/mL have been occasionally associated with increased risk of falls, fractures, or bone loss, and most guidelines therefore do not recommend values above this threshold (Fig. 50.2). In NHANES or European surveys, very few subjects not taking vitamin D supplements have spontaneous serum concentrations of 25OHD above 50 ng/mL and among the Masai tribe, who have extensive full-body exposure to sunlight, very few individuals have concentrations above 70 ng/mL. Nature has many inbuilt defense systems against vitamin D excess; nevertheless, in the absence of clear benefits of very high serum 25OHD concentrations (>50 ng/mL), it seems wise to avoid such levels.

Worldwide Vitamin D Status

There are an enormous amount of publications reporting the vitamin D status of individuals in different countries around the world. Unfortunately, most studies did not use assays validated for accuracy and did not use a sampling technique that enabled them to be representative of the country or region. Nevertheless, such data provide reasonably approximate data on vitamin D status around the world. NHANES data are an exception, as they are representative for the US population, and the assay results have recently been recalibrated according to the standards of The National Institute of Standards and Technology to provide accurate results.[81] In addition, NHANES has collected serum samples for many decades so that long-term changes can be calculated. The results reveal a fairly stable vitamin D status over many years (after correction for drift in assay results), despite an increasing use of vitamin D supplementation in the US population. Approximately 5% of the US population has a serum 25OHD concentration below 10 ng/mL, whereas approximately 18% has a level below 20 ng/mL. Although there are a very large number of reports on the vitamin D status of individuals in European countries, there has been no systematic survey using representative sampling and accurate methodology. On behalf of the European Calcified Tissue Society, Lips et al. provided a summary of approximately 100 studies conducted in different European or Middle-East countries.[30] The best situation, unexpectedly, is found in Northern Europe, where only between 1% and 8% of the population has a serum 25OHD concentration below 12 ng/mL, whereas this percentage is higher in most Western European countries (5%–31%). In Southern and Eastern Europe, there is such a wide variety of results that an easy summary is not possible, but overall the frequency of severe deficiency is high. In the Middle East, the Gulf states, and North African countries, between 30% and 90% of the population has severe vitamin D deficiency. When a conservative threshold of 20 ng/mL is used, then approximately 40% of the total European population can be considered "modestly" vitamin D-deficient. These data are, of course, compiled from studies that analyzed different populations and used different methodologies. Cashman et al., however, recalibrated the assay results from several European RCTs to provide a better comparison with NHANES data. According to this analysis, approximately 13% of subjects were severely vitamin D-deficient, and approximately 40% had a serum 25OHD concentration below 20 ng/mL.[93] One study recently generated data representative of the adult Portuguese population[94]: approximately 21% of the 4000 subjects had

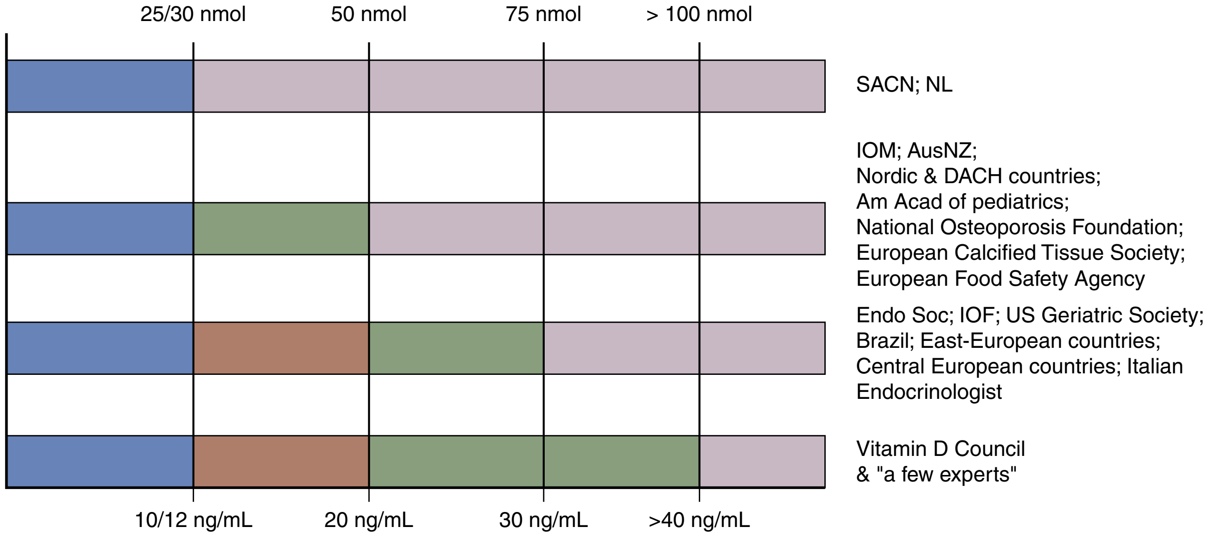


Fig. 50.2 Defining optimal vitamin D status as evaluated by serum 25OHD concentrations. Many guidelines (produced by governmental or scientific organizations) have tried to define thresholds for optimal vitamin D status in the normal population based on the most frequently used biomarker, serum 25OHD. All guidelines conclude that serum 25OHD concentrations should be above 10/12 ng/mL (25/30 nmol/L), and therefore levels below this threshold are indicated in red. Most experts now call this situation "severe vitamin D deficiency." The Institute of Medicine guidelines and many other guidelines consider a serum 25OHD level between 30 and 50 ng/mL as less than optimal. Indeed, these experts consider a serum 25OHD level of approximately 16 ng/mL to be sufficient for approximately 50% of the normal population and values above 20 ng/mL to be sufficient for approximately 97.5% of the population. Therefore, that range (12–20 ng/mL or 30–50 nmol/L) is indicated in orange. These guidelines consider serum 25OHD above 20 ng/mL as optimal, so we have indicated this range in green. The upper range of optimal 25OHD levels is disputed, but most guidelines consider 50 ng/mL to be the upper limit of normal values because of a combination of lack of proof of additional health benefits of concentrations higher than 50 ng/mL and some hints that such high values may have negative health consequences (especially with regards to falls and bone health). For details see the text. The Endocrine Society considers serum 25OHD between 20 and 30 ng/mL as "insufficient" and recommend all subjects (and especially subjects at risk of vitamin D deficiency) to aim for concentrations above 30 ng/mL. Therefore, we have indicated this range in yellow. A few organizations consider that serum 25OHD levels should be as high as they are in tribes with lifestyles that resemble those of early humans, and therefore suggest than we should all have serum 25OHD levels well above 40 or 50 ng/mL. Classical calcium-related toxicity is rarely found when serum 25OHD levels are lower than 100 ng/mL. We have therefore indicated higher 25OHD concentrations in red to denote their potential toxicity. *SACN,* Scientific Advisory Committee on Nutrition; *NL,* the Netherlands; *IOM,* Institute of Medicine; *NZ,* New Zealand; *IOF.* International Osteoporosis Foundation. (For references see Bouillon R. Comparative analysis of nutritional guidelines for vitamin D. *Nat Rev Endocrinol.* 2017;13:466–479.)

serum 25OHD concentrations below 10 ng/mL (the highest risk factors included living in the Azores, winter and spring season, obesity, and old age), whereas 67% had serum 25OHD concentrations below 20 ng/mL, with an overall concentration of 17 plus or minus 7 ng/mL (mean ± standard deviation). Asia is of course a large continent that comprises very different climates and lifestyles. Vitamin D deficiency is, however, frequent in some regions such as Mongolia (with one of the lowest mean 25OHD concentrations in children and a high prevalence of rickets), Northern China, or Northern India, and, of course, as mentioned earlier, in the Middle East and the Gulf states.

Africa, the cradle of mankind and relatively close the equator, was for a long time considered to be relatively free from vitamin D deficiency. However, a recent overview of available publications (with an unfortunate lack of standardization) revealed that approximately 18% of the population has severe vitamin D deficiency (especially in North and South Africa), whereas approximately one-third of the African population has levels below 20 ng/mL.[95] The Middle East and the Gulf states have the poorest vitamin D status, and severe vitamin D deficiency is also observed in some areas of Northern India and China.

KEY POINTS

- Serum 25OHD is the best marker of vitamin D status. Most international guidelines recommend serum 25OHD levels between 50 and 125 nmol/L (20L50 ng/ml). Severe vitamin D deficiency (serum 25OHD <30 nmol/L or 12 ng/mL) is, however, still frequent around the world, affecting more than 5% of the US population and 14% of the European population and an even higher percentage of people living in the Middle East, North Africa, Northern India, and China. Special risk groups include individuals with poor exposure to sunlight and individuals with dark skin living in temperate climate zones. A daily intake of 40–1000 IU of vitamin D can avoid such deficiency in most subjects.

Vitamin D Requirements

Most subjects synthesize sufficient vitamin D by regular exposure to sunlight combined with a (usually modest) dietary intake of vitamin D. Additional vitamin D supplementation is of course not necessary in this situation. Because UVB light is photodamaging to skin DNA,

TABLE 50.4 Recommended Daily Intake of Vitamin D According to Some Major Guidelines

Authority and/or Country (year)	RECOMMENDED INTAKE OF VITAMIN D (IU PER DAY)					
Guideline	**Age 0–1 year**	**Age 1–3 years**	**Adults**	**Age >75 years**	**Pregnancy**	**Lactation**
Institute of Medicine (2010)	400	600	600	800	600	600
Australia/New Zealand (2013)	400	600	600	800	600	600
DACH (2012)	400	800	800	800	800	800
Nordic countries (2012)	-/400	400	400	800	400	400
World Health Organization/Food and Agriculture Organization (2003/2012)	200	200	200	200	200	200
UK (Scientific Advisory Committee on Nutrition 2016)	340-400	400	400	400	400	400
The Netherlands (2012)	400	400	800	800	400	400
Belgium (2009)	400	400	400	800	800	800
France (Société Française de Nutrition; 2012)	800-1000	400	400-600	400-600	400	400
Endocrine society (2011)	400–1000	400–1000	600–2000	800–2000	600–2000	600–2000
European Food Safety Authority (2016)	400	600	600	600	600	600

(For details see text and Edouard T, Alos N, Chabot G, et al. Short- and long-term outcome of patients with pseudo-vitamin D deficiency rickets treated with calcitriol. *J Clin Endocrinol Metab.* 2011;96:82–89.)

with a host of associated long-term consequences, there is a difficult trade-off between avoiding excess exposure and spending just enough time in the sun to produce the necessary amount of vitamin D. It is not possible to formulate universal advice on reaching this optimal balance because it depends on many factors such as skin pigmentation, intensity of UVB light (which depends on geography, season, and climate), and extent of exposed skin. There are, however, a few general rules such as avoidance of sun burning or prolonged exposure. Indeed, UVB damage increases with increased duration of exposure, whereas vitamin D production rapidly reaches a plateau as the newly synthesized vitamin D is degraded by prolonged UVB exposure. Young children's skin is also more sensitive to UVB damage, and therefore it is generally advised to avoid direct exposure to sunlight for infants and small children.

The vitamin D requirements needed to meet these minimal serum 25OHD concentrations have been extensively evaluated in many guidelines.[96] For infants and small children there is wide unanimity that a daily supplement of 400 IU is sufficient to prevent rickets and increase serum 25OHD well above 12 ng/mL. According to the UK SACN experts, this dose should be sufficient at any age. Most other guidelines, however, recommend that older children, adults, and pregnant women who have minimal exposure to sunlight (or endogenous vitamin D production) take 600 IU daily. For elderly subjects a slightly higher dose of 800 IU/d is recommended by most guidelines. These conclusions, as summarized in the IOM report and subsequently in other guidelines, were based on mean serum 25OHD concentrations achieved by administering different doses of vitamin D in intervention studies. A reanalysis of the data using individual data rather than mean study results, however, suggests that higher daily doses may be needed, and are moreover age-dependent, whereby 1200 to 1600 IU/d may be needed to reach optimal values in 97.5% of subjects.[97] To achieve serum 25OHD concentrations above 30 ng/mL, higher doses with great interindividual variations may be needed. These recommendations are intended for the general population. Table 50.4 summarizes several recommendations for age-related daily vitamin D requirements for subjects with minimal exposure to sunlight. These guidelines of course specifically mention that, in some disease states different doses of vitamin D may be needed. Patients with intestinal malabsorption (including many subjects with bariatric surgery), patients with increased catabolism (such as patients taking antiepileptic drugs), or patients with renal loss of DBP/25OHD or chronic renal failure in general usually have higher daily replacement needs and require individual follow-up. Obese subjects frequently have a low vitamin D status, and thus a higher than usual intake can be recommended.

Choice of Vitamin D Supplementation. Vitamin D (either D_3 or D_2) for infants and children is usually given as a daily supplement or added to bottle milk. Vitamin D_3 and D_2 are nearly equipotent when given on a daily basis, but vitamin D_2 is somewhat less potent than D_3 when given intermittently in weekly or monthly doses. Vitamin D_3 given in weekly or monthly doses (calculated to be equivalent to daily doses), but is nearly equipotent to daily supplements. The absorption of vitamin D is slightly but inconsistently modified by simultaneous intake of food or fats. Much higher doses given once or a few times a year have been associated with transient increased risk for falls and fractures and are thus no longer recommended.

Calcifediol or 25OHD is used in a few countries as alternative for vitamin D. The potency of 25OHD is approximately 3-fold higher than that of vitamin D itself when given to otherwise healthy postmenopausal women.[98] Calcifediol more rapidly restores serum 25OHD concentrations because of its better intestinal absorption (nearly 100%) and because it bypasses the need for hepatic CYP2R1-mediated 25-hydroxylation of vitamin D. The higher solubility of 25OHD also assures very high intestinal absorption even in case of intestinal fat malabsorption, and thus may be the preferred option in such cases. The increase in serum 25OHD concentration after vitamin D intake does not increase linearly, as the "delta 25OHD" is higher when the baseline vitamin D status is low and decreases with increasing doses of vitamin D, reaching a plateau 25OHD value with high vitamin D intake. The increase in serum 25OHD is more linearly related to the dose of calcifediol.

The use of $1,25(OH)_2D$ or 1a-OHD is restricted to situations of (genetic or acquired) loss of CYP27B1 activity, such as in patients with CYP27B1 mutations, patient with advanced renal failure (to prevent severe secondary hyperparathyroidism), or patients with hypoparathyroidism.

Thousands of vitamin D analogs have been synthesized and evaluated for their ability to treat, among other conditions, osteoporosis, cancer, and immune disorders. Only a very few analogs have made it into clinical practice. Calcipotriol is the first-line topical treatment for psoriasis. A few analogs (especially paricalcitol) are used in patients with advanced chronic renal failure because of a modestly greater effect on serum PTH than on serum calcium. Eldecalcitol is used in Japan for the treatment of osteoporosis but is not available in most countries.

SUMMARY

Vitamin D has been used for almost a century to prevent or cure nutritional rickets. Although the success rate of supplementation in infants and children is excellent, implementation is still problematic in many areas of the world. A dedicated WHO-sponsored action program similar to that instituted to prevent iodine deficiency disorders is long overdue and should preferably be part of a more general childcare health project dealing with nutrition and vaccination programs.

Vitamin D is a precursor for a secosteroid hormone, 1,25(OH)$_2$D, which acts as a ligand for a specific receptor, VDR, and regulates a very large number of genes. It is a key regulator of calcium and phosphate homeostasis through regulation of calcium transporters in the intestine and other calciotropic hormones (PTH and FGF-23). Many other clusters of genes involved in a wide variety of actions such as cell proliferation or immunity are also under the control of 1,25(OH)$_2$D, making extraskeletal activity plausible.

Severe vitamin D deficiency, as defined by serum 25OHD concentrations below 10 ng/mL, is present in approximately 7% of the world population and is much more prevalent in North Africa, the Middle East, and parts of Asia, affecting both young and older individuals. There is probably not a simple strategy that would solve this issue in all these countries. Individualized regular intake of supplements may solve the problem, but vitamin D supplementation of frequently consumed food items may well be the best strategy.

Pregnancy does not increase the need for vitamin D; nevertheless, many young women have a poor vitamin D status. Because it is plausible that this deficiency has negative consequences for the health of both the mother and her offspring, a dedicated action program as part of prenatal care should be developed and implemented.

The elderly frequently have increased fracture risk, poor vitamin D status, and lower-than-optimal calcium intake. Vitamin D supplementation has not proven to reduce fracture risk, and isolated calcium supplementation was also not effective. However, combined calcium and vitamin D supplementation has been repeatedly shown in several metaanalyses to be able to reduce hip and major fractures by 10% to 20%. All guidelines and most science organizations therefore recommend a daily supplement of approximately 800 IU of vitamin D together with optimal calcium intake through diet and or calcium supplements.

Approximately one-third of the world population has a 25OHD serum concentrations between 10 and 20 ng/mL, which is considered to be suboptimal because this state stimulates compensatory mechanisms to maintain normal calcium and phosphate homeostasis. This situation may affect the general population but is more frequent in specific risks groups such as people with limited exposures to sunlight because of climatic reasons, those with religious or cultural practices that involve covering the body, individuals with dark skin living in moderate climates, or anyone who does not produce sufficient endogenous vitamin D owing to lack of exposure to sunlight.

Most current guidelines are based on a large set of studies dealing with bone health and did not consider the evidence of potential extraskeletal benefits of vitamin D supplementation of sufficient quality to adjust their conclusions. Over the last years the results of many RCTs, and especially of several major megatrials or metaanalyses, have been published and provide data on potential extraskeletal effects of vitamin D supplementation. These data clearly show that vitamin D supplements given to mostly vitamin D-replete subjects do not convey health benefits for the skeleton or for general health. Therefore, the widespread belief among many scientists and lay people that improving vitamin D status by taking regular supplements may decrease the risk of many major diseases and improve general health is not supported by large-scale clinical trials or by MR studies. Table 50.5 summarizes the potential extraskeletal health benefits of vitamin D supplementation, although most of the effects described were mainly observed in subjects with poor vitamin D status at baseline.

Whereas over the last two decades the most attention was paid to the potential benefits of higher than usual vitamin D supplements for mitigating many major extraskeletal disorders, it is time to reorient basic and clinical research in the vitamin D field.

First, severe vitamin D deficiency (defined as serum 25OHD <12 ng/mLs) is still widely present around the world. It can be rather easily eliminated by appropriate implementation policies using small daily doses of vitamin D (~400 IU). This would allow the eradication of nutritional rickets and other consequences of severe vitamin D deficiency.

Second, vitamin D deficiency (defined as serum 25OHD <20 ng/mL) affects more than one-third of the world population. Combined use of calcium and vitamin D by elderly subjects probably reduces their risk of bone loss and fractures, especially because many of them have poor calcium and vitamin D statuses. Other risk groups for (severe) vitamin D deficiency are well known, and specific approaches are needed to define the benefits of supplementation, as well as the best long-term strategies, to eliminate such deficiency.

Third, many pregnant women have a rather poor vitamin D status. Correction of such deficiency may have health benefits for mothers and their offspring, but better data are needed to define who should receive how much vitamin D and when.

Fourth, the extraskeletal health benefits of vitamin D are probably much lower than many experts and lay people have believed over the last decades. Nevertheless, there is enough evidence of such benefits (summarized in Table 50.5) to guide future research. All these extraskeletal effects should be interpreted with caution, as the observations were frequently based on *post hoc* or subgroup analysis and therefore will require validation in additional dedicated trials.

TABLE 50.5 **Potential Extraskeletal Benefits of Vitamin D Supplementation Based on Analysis of Secondary Endpoints or *Post Hoc* Analysis of Recent Major Supplementation Trial or Mendelian Randomization Studies**

1. Reduce the risk of upper respiratory infections
2. May improve lung function, especially in subjects with asthma, chronic obstructive pulmonary disease, or chronic smokers
3. May improve proximal muscle strength
4. May decrease the risk of falls of elderly subjects
5. May decrease the progression of prediabetes into type 2 diabetes
6. May improve maternal health (especially preeclampsia and gestational diabetes)
7. May decrease cancer mortality (without changing the incidence of cancer)
8. May modestly reduce mortality risk of elderly subjects
9. May decrease the risk of multiple sclerosis (as based on several Mendelian randomization studies)

The most important message generated by the recent megatrials is that vitamin D supplementation of mostly vitamin D-replete adults does not provide major health benefits. Vitamin D supplementation should thus be focused much more selectively on the millions of subjects with modest or severe vitamin D deficiency. The major risk factors for such poor vitamin D status are well-defined, and healthcare policy should address these risk groups with adequate strategies. There are still a large number of ongoing clinical trials that may convey additional insight in the coming years. Moreover, using the power of MR we can now predict larger lifetime changes in vitamin D status. Such studies may generate data that can further modify or fine-tune the present conclusions.

REFERENCES

1. Bouillon R, Antonio L. Nutritional rickets: historic overview and plan for worldwide eradication. *J Steroid Biochem Mol Biol*. 2020;198:105563.
2. Bouillon R, Marcocci C, Carmeliet G, et al. Skeletal and extraskeletal actions of vitamin D: current evidence and outstanding questions. *Endocr Rev*. 2019;40:1109–1151.
3. Bouillon R, Bikle D. Vitamin D metabolism revised: fall of dogmas. *J Bone Miner Res*. 2019;34:1985–1992.
4. Meyer MB, Benkusky NA, Kaufmann M, et al. Targeted genomic deletions identify diverse enhancer functions and generate a kidney-specific, endocrine-deficient. *J Biol Chem*. 2019;294:9518–9535.
5. Christakos S, Dhawan P, Verstuyf A, et al. Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects. *Physiol Rev*. 2016;96:365–408.
6. Edouard T, Alos N, Chabot G, et al. Short- and long-term outcome of patients with pseudo-vitamin D deficiency rickets treated with calcitriol. *J Clin Endocrinol Metab*. 2011;96:82–89.
7. Schlingmann KP, Cassar W, Konrad M. Juvenile onset IIH and CYP24A1 mutations. *Bone Rep*. 2018;9:42–46.
8. Pauwels S, Jans I, Billen J, et al. 1β,25-dihydroxyvitamin D. *J Steroid Biochem Mol Biol*. 2017;173:341–348.
9. Roizen JD, Li D, O'Lear L, et al. CYP3A4 mutation causes vitamin D-dependent rickets type 3. *J Clin Invest*. 2018;128:1913–1918.
10. Bouillon R, Schuit F, Antonio L, et al. Vitamin D binding protein: a historic overview. *Front Endocrinol (Lausanne)*. 2019;10:910.
11. Henderson CM, Fink SL, Bassyouni H, et al. Vitamin D-binding protein deficiency and homozygous deletion of the GC gene. *N Engl J Med*. 2019;380:1150–1157.
12. Jones KS, Assar S, Harnpanich D, et al. 25(OH)D2 half-life is shorter than 25(OH)D3 half-life and is influenced by DBP concentration and genotype. *J Clin Endocrinol Metab*. 2014;99:3373–3381.
13. Haussler MR, Whitfield GK, Kaneko I, et al. Molecular mechanisms of vitamin D action. *Calcif Tissue Int*. 2013;92:77–98.
14. Blau JE, Collins MT. The PTH-Vitamin D-FGF23 axis. *Rev Endocr Metab Disord*. 2015;16:165–174.
15. Lieben L, Masuyama R, Torrekens S, et al. Normocalcemia is maintained in mice under conditions of calcium malabsorption by vitamin D-induced inhibition of bone mineralization. *J Clin Invest*. 2012;122:1803–1815.
16. Carpenter TO, Shaw NJ, Portale AA, et al. Rickets. *Nat Rev Dis Primers*. 2017;3:17101.
17. Nutritional rickets: a review of disease burden, causes, diagnosis, prevention and treatment. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.
18. Munns CF, Shaw N, Kiely M, et al. Global consensus recommendations on prevention and management of nutritional rickets. *J Clin Endocrinol Metab*. 2016;101:394–415.
19. Uday S, Högler W. Nutritional rickets and osteomalacia in the twenty-first century: revised concepts, public health, and prevention strategies. *Curr Osteoporos Rep*. 2017;15:293–302.
20. Bolland MJ, Grey A, Avenell A. Effects of vitamin D supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis. *Lancet Diabetes Endocrinol*. 2018;6:847–858.
21. Avenell A, Mak JC, O'Connell D. Vitamin D and vitamin D analogues for preventing fractures in post-menopausal women and older men. *Cochrane Database Syst Rev*. 2014:CD000227.
22. Bolland MJ, Grey A, Gamble GD, et al. The effect of vitamin D supplementation on skeletal, vascular, or cancer outcomes: a trial sequential meta-analysis. *Lancet Diabetes Endocrinol*. 2014;2:307–320.
23. Weaver CM, Alexander DD, Boushey CJ, et al. Calcium plus vitamin D supplementation and risk of fractures: an updated meta-analysis from the National Osteoporosis Foundation. *Osteoporos Int*. 2016;27:367–376.
24. LeBoff MS, Chou SH, Murata EM, et al. Effects of supplemental vitamin d on bone health outcomes in women and men in the VITamin D and OmegA-3 TriaL (VITAL). *J Bone Miner Res*. 2020;35:883–893.
25. Reid IR, Horne AM, Mihov B, et al. Effect of monthly high-dose vitamin D on bone density in community-dwelling older adults substudy of a randomized controlled trial. *J Intern Med*. 2017;282:452–460.
26. Khaw KT, Stewart AW, Waayer D, et al. Effect of monthly high-dose vitamin D supplementation on falls and non-vertebral fractures: secondary and post-hoc outcomes from the randomised, double-blind, placebo-controlled ViDA trial. *Lancet Diabetes Endocrinol*. 2017;5:438–447.
27. Burt LA, Billington EO, Rose MS, et al. Effect of high-dose vitamin d supplementation on volumetric bone density and bone strength: a randomized clinical trial. *J Am Med Assoc*. 2019;322:736–745.
28. Billington EO, Burt LA, Rose MS, et al. Safety of high-dose vitamin d supplementation: secondary analysis of a randomized controlled trial. *J Clin Endocrinol Metab*. 2020;105:1261–1273.
29. Bouillon R. Safety of high-dose vitamin d supplementation. *J Clin Endocrinol Metab*. 2020;105:1290–1291.
30. Lips P, Cashman KD, Lamberg-Allardt C, et al. Current vitamin D status in European and Middle East countries and strategies to prevent vitamin D deficiency: a position statement of the European Calcified Tissue Society. *Eur J Endocrinol*. 2019;180:P23–P54.

30a. Cosman F, de Beur MS, LeBoff MS, et al. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int*. 2014;25:2359–2381.

31. Sanders KM, Seibel MJ. Therapy: new findings on vitamin D3 supplementation and falls - when more is perhaps not better. *Nat Rev Endocrinol*. 2016;12:190–191.
32. Gordon-Thomson C, Tongkao-on W, Song EJ, et al. Protection from ultraviolet damage and photocarcinogenesis by vitamin D compounds. *Adv Exp Med Biol*. 2014;810:303–328.
33. Feldman D, Krishnan AV, Swami S, et al. The role of vitamin D in reducing cancer risk and progression. *Nat Rev Cancer*. 2014;14:342–357.
34. Chen L, Yang R, Qiao W, et al. 1,25-Dihydroxy vitamin D prevents tumorigenesis by inhibiting oxidative stress and inducing tumor cellular senescence in mice. *Int J Cancer*. 2018;143:368–382.
35. Bouillon R, Van Schoor NM, Gielen E, et al. Optimal vitamin D status: a critical analysis on the basis of evidence-based medicine. *J Clin Endocrinol Metab*. 2013;98:E1283–E1304.
36. Scragg R, Khaw KT, Toop L, et al. Monthly high-dose vitamin d supplementation and cancer risk: a post hoc analysis of the vitamin d assessment randomized clinical trial. *JAMA Oncol*. 2018;4:e182178.
37. Manson JE, Cook NR, Lee IM, et al. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med*. 2019;380:33–44.
38. Manson JE, Bassuk SS, Buring JE, et al. Principal results of the VITamin D and OmegA-3 TriaL (VITAL) and updated meta-analyses of relevant vitamin D trials. *J Steroid Biochem Mol Biol*. 2020;198:105522.
39. Ong JS, Cuellar-Partida G, Lu Y, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. *Int J Epidemiol*. 2016;45:1619–1630.
40. Vanherwegen AS, Gysemans C, Mathieu C. Regulation of immune function by vitamin d and its use in diseases of immunity. *Endocrinol Metab Clin North Am*. 2017;46:1061–1094.
41. Dimitrov V, White JH. Vitamin D signaling in intestinal innate immunity and homeostasis. *Mol Cell Endocrinol*. 2017;453:68–78.
42. Martineau AR, Jolliffe DA, Greenberg L, et al. Vitamin D supplementation to prevent acute respiratory infections: individual participant data meta-analysis. *Health Technol Assess*. 2019;23:1–44.

43. Camargo CA, Sluyter J, Stewart AW, et al. Effect of monthly high-dose vitamin d supplementation on acute respiratory infections in older adults: a randomized controlled trial. *Clin Infect Dis*. 2020;71:311–317.
44. Liu PT, Stenger S, Li H, et al. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science*. 2006;311:1770–1773.
45. Wang TT, Nestel FP, Bourdeau V, et al. Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. *J Immunol*. 2004;173:2909–2912.
46. Wu HX, Xiong XF, Zhu M, et al. Effects of vitamin D supplementation on the outcomes of patients with pulmonary tuberculosis: a systematic review and meta-analysis. *BMC Pulm Med*. 2018;18:108.
47. Ganmaa D, Munkhzul B, Fawzi W, et al. High-dose vitamin D. *Am J Respir Crit Care Med*. 2017;196:628–637.
47a. Ganmaa D, Uyanga B, Zhou X, et al. Vitamin D supplements for prevention of tuberculosis infection and disease. *New Engl J Med*. 2020;383:359–368.
48. Bouillon R, Lieben L, Mathieu C, et al. Vitamin D action: lessons from VDR and Cyp27b1 null mice. *Pediatr Endocrinol Rev*. 2013;10(suppl 2):354–366.
49. Giulietti A, Gysemans C, Stoffels K, et al. Vitamin D deficiency in early life accelerates Type 1 diabetes in non-obese diabetic mice. *Diabetologia*. 2004;47:451–462.
50. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *J Am Med Assoc*. 2006;296:2832–2838.
51. Jacobs BM, Noyce AJ, Giovannoni G, et al. BMI and low vitamin D are causal factors for multiple sclerosis: a Mendelian Randomization study. *Neurol Neuroimmunol Neuroinflamm*. 2020;7:e662.
52. Feige J, Moser T, Bieler L, et al. Vitamin D supplementation in multiple sclerosis: a critical analysis of potentials and threats. *Nutrients*. 2020;12:783.
53. Dörr J, Bäcker-Koduah P, Wernecke KD, et al. High-dose vitamin D supplementation in multiple sclerosis—results from the randomized EVIDIMS (efficacy of vitamin D supplementation in multiple sclerosis) trial. *Mult Scler J Exp Transl Clin*. 2020;6:2055217320903474.
54. de Bruyn JR, Bossuyt P, Ferrante M, et al. High-dose vitamin d does not prevent postoperative recurrence of Crohn's disease in a randomized placebo-controlled trial. *Clin Gastroenterol Hepatol*. 2020;19:1573–1582.
55. Bouillon R. Vitamin D and cardiovascular disorders. *Osteoporos Int*. 2019;30:2167–2181.
56. Chowdhury R, Kunutsor S, Vitezova A, et al. Vitamin D and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies. *BMJ*. 2014;348:g1903.
57. Zhang R, Li B, Gao X, et al. Serum 25-hydroxyvitamin D and the risk of cardiovascular disease: dose-response meta-analysis of prospective studies. *Am J Clin Nutr*. 2017;105:810–819.
58. Beveridge LA, Struthers AD, Khan F, et al. Effect of vitamin d supplementation on blood pressure: a systematic review and meta-analysis incorporating individual patient data. *JAMA Intern Med*. 2015;175:745–754.
59. Sluyter JD, Camargo CA, Stewart AW, et al. Effect of monthly, high-dose, long-term vitamin d supplementation on central blood pressure parameters: a randomized controlled trial substudy. *J Am Heart Assoc*. 2017;6:e006802.
60. Wang L, Song Y, Manson JE, et al. Circulating 25-hydroxy-vitamin D and risk of cardiovascular disease: a meta-analysis of prospective studies. *Circ Cardiovasc Qual Outcomes*. 2012;5:819–829.
61. Scragg R, Stewart AW, Waayer D, et al. Effect of monthly high-dose vitamin d supplementation on cardiovascular disease in the vitamin d assessment study : a randomized clinical trial. *JAMA Cardiol*. 2017;2:608–616.
62. Hiemstra T, Lim K, Thadhani R, et al. Vitamin D and atherosclerotic cardiovascular disease. *J Clin Endocrinol Metab*. 2019;104:4033–4050.
63. Barbarawi M, Kheiri B, Zayed Y, et al. Vitamin D supplementation and cardiovascular disease risks in more than 83 000 individuals in 21 randomized clinical trials: a meta-analysis. *JAMA Cardiol*. 2019;4:765–776.
64. Girgis CM, Clifton-Bligh RJ, Hamrick MW, et al. The roles of vitamin D in skeletal muscle: form, function, and metabolism. *Endocr Rev*. 2013;34:33–83.
65. Girgis CM, Cha KM, So B, et al. Mice with myocyte deletion of vitamin D receptor have sarcopenia and impaired muscle function. *J Cachexia Sarcopenia Muscle*. 2019;10:1228–1240.
66. Rosen CJ, Adams JS, Bikle DD, et al. The nonskeletal effects of vitamin D: an Endocrine Society scientific statement. *Endocr Rev*. 2012;33:456–492.
67. Beaudart C, Buckinx F, Rabenda V, et al. The effects of vitamin D on skeletal muscle strength, muscle mass, and muscle power: a systematic review and meta-analysis of randomized controlled trials. *J Clin Endocrinol Metab*. 2014;99:4336–4345.
68. Sanders KM, Stuart AL, Williamson EJ, et al. Annual high-dose oral vitamin D and falls and fractures in older women: a randomized controlled trial. *J Am Med Assoc*. 2010;303:1815–1822.
69. Bischoff-Ferrari HA, Dawson-Hughes B, Orav EJ, et al. Monthly high-dose vitamin d treatment for the prevention of functional decline: a randomized clinical trial. *JAMA Intern Med*. 2016;176:175–183.
70. Smith LM, Gallagher JC, Suiter C. Medium doses of daily vitamin D decrease falls and higher doses of daily vitamin D3 increase falls: a randomized clinical trial. *J Steroid Biochem Mol Biol*. 2017;173:317–322.
71. Song Y, Wang L, Pittas AG, et al. Blood 25-hydroxy vitamin D levels and incident type 2 diabetes: a meta-analysis of prospective studies. *Diabetes Care*. 2013;36:1422–1428.
72. Rejnmark L, Bislev LS, Cashman KD, et al. Non-skeletal health effects of vitamin D supplementation: a systematic review on findings from meta-analyses summarizing trial data. *PloS One*. 2017;12:e0180512.
73. Pittas AG, Dawson-Hughes B, Sheehan P, et al. Vitamin D supplementation and prevention of type 2 diabetes. *N Engl J Med*. 2019;381:520–530.
74. Bouillon R, Carmeliet G, Lieben L, et al. Vitamin D and energy homeostasis: of mice and men. *Nat Rev Endocrinol*. 2014;10:79–87.
75. Palacios C, Kostiuk LK, Peña-Rosas JP. Vitamin D supplementation for women during pregnancy. *Cochrane Database Syst Rev*. 2019;7:CD008873.
76. Bialy L, Fenton T, Shulhan-Kilroy J, et al. Vitamin D supplementation to improve pregnancy and perinatal outcomes: an overview of 42 systematic reviews. *BMJ Open*. 2020;10:e032626.
77. Fogacci S, Fogacci F, Banach M, et al. Vitamin D supplementation and incident preeclampsia: a systematic review and meta-analysis of randomized clinical trials. *Clin Nutr*. 2020;39:1742–1752.
78. Roth DE, Morris SK, Zlotkin S, et al. Vitamin D supplementation in pregnancy and lactation and infant growth. *N Engl J Med*. 2018;379:535–546.
79. Gallagher JC. Vitamin D and falls—the dosage conundrum. *Nat Rev Endocrinol*. 2016;12:680–684.
80. Billington EO, Burt LA, Plett R, et al. Effect of high-dose vitamin D supplementation on peripheral arterial calcification: secondary analysis of a randomized controlled trial. *Osteoporos Int*. 2020;31:1995–2005.
81. Durazo-Arvizu RA, Dawson-Hughes B, Kramer H, et al. The reverse j-shaped association between serum total 25-hydroxyvitamin D concentration and all-cause mortality: the impact of assay standardization. *Am J Epidemiol*. 2017;185:720–726.
82. Bjelakovic G, Gluud LL, Nikolova D, et al. Vitamin D supplementation for prevention of mortality in adults. *Cochrane Database Syst Rev*. 2014:CD007470.
83. Sempos CT, Heijboer AC, Bikle DD, et al. Vitamin D assays and the definition of hypovitaminosis D: results from the first international conference on controversies in vitamin D. *Br J Clin Pharmacol*. 2018;84:2194–2207.
84. Bikle D, Bouillon R, Thadhani R, et al. Vitamin D metabolites in captivity? Should we measure free or total 25(OH)D to assess vitamin D status? *J Steroid Biochem Mol Biol*. 2017;173:105–116.
85. Need AG, O'Loughlin PD, Morris HA, et al. Vitamin D metabolites and calcium absorption in severe vitamin D deficiency. *J Bone Miner Res*. 2008;23:1859–1863.
86. *The Scientific Advisory Committee on Nutrition (SACN) Recommendations on Vitamin D. SACN Vitamin D and Health Report (Independent Report)*. 2016. Available at: https://www.gov.uk/government/publications/sacn-vitamin-d-and-health-report.
87. IOM (Institute of Medicine). 2011. *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: The National Academies Press.
88. *Nordic Nutrition Recommendations, 2012: Integrating nutrition and physical activity*. ISBN 978-92-893-2670-4; http://dx.doi.org/10.6027/Nord2014-002 Nord 2014:002 ISSN 0903-7004.
89. DACH, German Nutrition Society. New reference values for vitamin D. *Ann Nutr Metab*. 2012;60:241–246.

90. EFSEFSA Panel on Dietetic Products, Nutrition and Allergies, 2016. Scientific opinion on dietary reference values for vitamin D. *EFSA Journal*. 2016;14:4547.
91. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1911–1930.
92. Luxwolda MF, Kuipers RS, Kema IP, et al. Vitamin D status indicators in indigenous populations in East Africa. *Eur J Nutr*. 2013;52:1115–1125.
93. Cashman KD, Dowling KG, Škrabáková Z, et al. Vitamin D deficiency in Europe: pandemic? *Am J Clin Nutr*. 2016;103:1033–1044.
94. Duarte C, Carvalheiro H, Rodrigues AM, et al. Prevalence of vitamin D deficiency and its predictors in the Portuguese population: a nationwide population-based study. *Arch Osteoporos*. 2020;15:36.
95. Mogire RM, Mutua A, Kimita W, et al. Prevalence of vitamin D deficiency in Africa: a systematic review and meta-analysis. *Lancet Glob Health*. 2020;8:e134–e142.
96. Bouillon R. Comparative analysis of nutritional guidelines for vitamin D. *Nat Rev Endocrinol*. 2017;13:466–479.
97. Cashman KD, Kiely ME, Andersen R, et al. Individual participant data (IPD)-level meta-analysis of randomised controlled trials with vitamin D-fortified foods to estimate Dietary Reference Values for vitamin D. *Eur J Nutr*. 2020;59:741–753.
98. Quesada-Gomez JM, Bouillon R. Is calcifediol better than cholecalciferol for vitamin D supplementation? *Osteoporos Int*. 2018;29:1697–1711.

51

Bone Development and Remodeling

Christa Maes and Henry M. Kronenberg

OUTLINE

Mechanisms of Skeletal Development, 869
  Intramembranous Ossification, 869
  Endochondral Ossification, 870
Principles of Bone Remodeling and Skeletal Homeostasis, 872
  The Bone Remodeling Cycle, 872
  Unbalanced Bone Remodeling in Metabolic Bone Diseases, 872
Cellular and Molecular Control of Skeletal Development and Bone Remodeling, 873
  Chondrogenesis and Growth Plate Biology, 873
  Skeletal Vascularization, 876
  Skeletal Stem and Progenitor Cells, 876
  Osteoblastogenesis, 879
  Osteoclastogenesis and Skeletal Tissue Resorption, 880
Regulation of Bone Remodeling and Bone Mass, 883
  Coupling of Bone Resorption and Bone Formation During Remodeling, 883
  Local, Systemic, and Central Regulators of Bone Remodeling, 884
Conclusion, 886

Bone is a dynamic tissue that develops and maintains the skeleton to fulfill at least five primary functions in higher vertebrates: (1) the skeleton provides the crucial lever arms that allow muscular activity to generate movement; (2) the skeleton provides rigid protection for internal organs, including the brain and spinal cord; (3) the skeleton is the site for storage and controlled systemic release of calcium and other ions, which are essential for electrolyte homeostasis of the extracellular fluid; (4) bones house the bone marrow and participate in the genesis of hematopoietic cells by providing the essential stromal, vascular, and osteogenic niches that interact with hematopoietic precursors; and (5) the skeleton signals in ways that contribute to systemic glucose homeostasis and global energy metabolism.

A number of distinct cell types control the development of bone and regulate its functions. Skeletal development and growth are driven to a large extent by the differentiation of chondrocytes, which form and make up cartilage, and of osteoblasts, which build mineralized bone. These cells derive from mesenchyme-derived populations of skeletal stem and progenitor cells (SSPCs). Postnatally, the skeleton is continuously turning over in a process called bone remodeling. In this process, packets of bone are constantly being removed and replaced to meet the skeleton's mechanical demands by providing maximal strength from minimal material, while at the same time allowing the repair of microcracks and the accretion and release of calcium and phosphate. The principal cellular players in this process are the bone-forming osteoblasts and the osteoclasts, giant multinuclear cells derived from hematopoietic precursors that break down (resorb) bone. The quantitatively dominant cell type in adult bone and, in many ways, the master cell that coordinates the activity of the others is the osteocyte, which is derived from osteoblasts and buried within the bone matrix. The roles of osteocytes are not fully understood but include the sensing of mechanical forces, the regulation of bone formation and resorption, and the regulation of calcium and phosphate homeostasis. Bone development and maintenance are closely associated with and dependent on angiogenesis, putting vascular endothelial cells in the picture as a further crucial cell type in bone.

MECHANISMS OF SKELETAL DEVELOPMENT

During embryonic development, the skeleton is established by the coordinated formation of well over 200 separate bones at sites distributed all over the body. Two distinct mechanisms are responsible for the development of bones—intramembranous and endochondral ossification—both of which depend on the coordinate growth, differentiation, function, and interaction of various cell types. Intramembranous ossification is responsible primarily for the formation of most of the craniofacial elements and much of the scapulae and clavicles. In contrast, all the long bones of the limbs and the vertebrae and ribs of the trunk develop through endochondral ossification. This term connotes the formation of mineralized bone through the deposition of true bone matrix on top of scaffolding cartilaginous anlagen.

Intramembranous Ossification

The cranial and facial bones are derived largely from the neural crest through the process of intramembranous bone formation.[1] In areas where bones are formed, mesenchymal precursor cells aggregate and form regions of high cell density called condensations. In intramembranous bones, the cells in these condensations then differentiate directly into osteoblasts that deposit "osteoid" or bone matrix rich in type I collagen, a process that occurs in close spatial interaction with vascular tissue. As the osteoblasts mature, the bone matrix becomes progressively mineralized. Terminally differentiated osteoblasts eventually become entrapped in the bone as osteocytes. Membranous bones that derive from condensations that never form chondrocytes during development are restricted to the flat bones of the skull (calvarial bones and mandibles) and parts of the clavicles and scapulae. The subperiosteal bone collar or provisional cortex of long bones can be considered to form also by intramembranous bone formation, because the osteoblasts in this region differentiate directly from the perichondrial/periosteal mesenchyme and form bone adjacent to, instead of directly upon, a cartilage-derived matrix (see later). Of note, unlike the formation of intramembranous bone directly from condensations, the formation of this "intramembranous" bone in the long bones is regulated by signals from the chondrocytes, so can be thought of developmentally as part of the process of endochondral bone formation. Intramembranous bone formation also occurs during bone repair after fracture, particularly when the bones are stabilized (fixed).

KEY POINTS

- Intramembranous bone forms directly from mesenchymal condensations. Intramembranous ossification is the process used to form the bones of the calvarium and parts of the clavicle and scapula.

Endochondral Ossification

Endochondral ossification is the mechanism responsible for the formation of all long bones of the axial skeleton (vertebrae and ribs) and the appendicular skeleton (limbs). Most of the axial skeleton is derived from cells of the paraxial mesoderm that condense early in embryogenesis on both sides of the neural tube and the notochord. The appendicular skeleton arises from the lateral plate mesoderm. The mechanisms underlying the early condensation, segmentation, differentiation, and patterning events define the precise arrangement of the individual anatomic elements and their patterning along the body axes. These mechanisms involve the actions of and crosstalk among several morphogens, including fibroblast growth factors (FGFs), sonic hedgehog, bone morphogenetic proteins (BMPs), and Wnts, as well as control by Notch signaling and by transcription factors encoded by *HOX*, *PAX1*, and *TBX*.[2]

As in intramembranous ossification, the development of the long bones starts with mesenchymal progenitor cells forming condensations. In the mesenchymal condensations of endochondral bones, cells do not differentiate into osteoblasts but instead differentiate into chondrocytes that synthesize a characteristic extracellular matrix rich in type II collagen and specific proteoglycans. As such, a cartilaginous model or anlage is established that prefigures the future bone. These differentiated cartilage structures in the limb appear in sequence along the proximodistal axis (i.e., hip to toes, shoulder to fingers). The sequential steps of the endochondral ossification process starting from this stage are illustrated in Fig. 51.1. Initially, the cartilage further enlarges through chondrocyte proliferation and matrix production. Chondrocytes in the midportion of the bone model then stop proliferating, undergo further maturation, and ultimately become hypertrophic. These large hypertrophic chondrocytes secrete a distinct matrix, containing type X collagen, and then rapidly direct the calcification of the matrix. Concomitantly, the hypertrophic chondrocytes direct the cells surrounding the cartilage element called the perichondrium to differentiate into osteoblasts that deposit mineralized bone matrix—the "bone collar"—around the cartilage template. This bone collar forms the initiation site of the cortical bone,

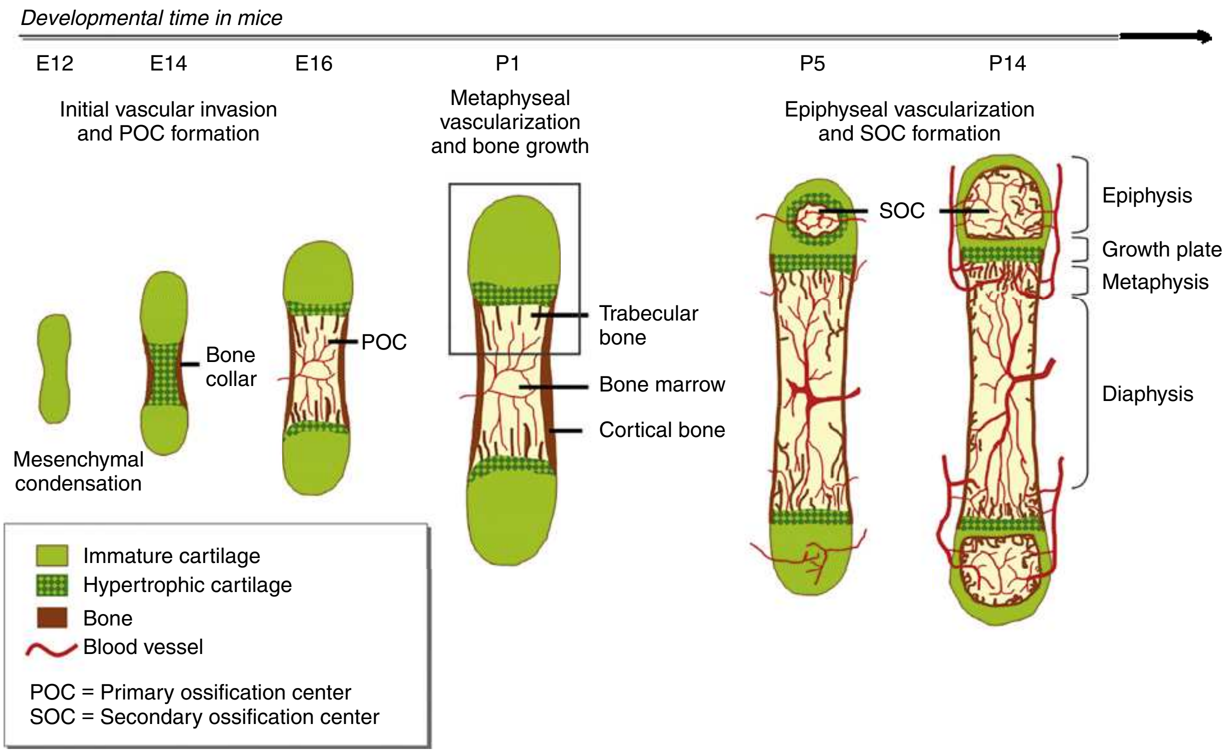

Developmental time in mice
E12
E14
E16
P1
P5
P14
Initial vascular invasion and POC formation
Metaphyseal vascularization and bone growth
Epiphyseal vascularization and SOC formation
Mesenchymal condensation
Bone collar
POC
Trabecular bone
Bone marrow
Cortical bone
SOC
Epiphysis
Growth plate
Metaphysis
Diaphysis
Immature cartilage
Hypertrophic cartilage
Bone
Blood vessel
POC = Primary ossification center
SOC = Secondary ossification center

Fig. 51.1 Stepwise Schematic Diagram of Long Bone Development Through Endochondral Ossification. Around embryonic day (E) 12 in mice, mesenchymal progenitor cells condense and differentiate into chondrocytes to form the cartilage anlagen that prefigure future long bones. Chondrocytes in the center become hypertrophic, while cells in the surrounding perichondrium differentiate into osteoblasts, forming a bone collar, the provisional cortical bone. The hypertrophic cartilage core subsequently is invaded by blood vessels along with osteoprogenitors and osteoclasts and becomes eroded and replaced by bone and marrow (primary ossification center). In the metaphysis, hypertrophic cartilage of the growth cartilage is continually replaced by trabecular bone, a process that relies on metaphyseal vascularization and mediates longitudinal bone growth (see postnatal day [P] 1). Around P5, epiphyseal vessels invade the avascular cartilage at the ends of the bone and initiate the secondary centers of ossification. Discrete layers of residual chondrocytes form growth plates between the epiphyseal and metaphyseal bone centers to support further postnatal longitudinal bone growth. Ultimately (in humans), the growth plates close, and growth stops. A detailed view of the perinatal bone structure (boxed area indicated at P1) is provided in Fig. 51.2.

the dense outer envelope of compact, lamellar bone that provides the long bone with most of its strength and rigidity (see Fig. 51.1).

At this time in development, the cartilage model starts to become replaced by bone, vascular, and marrow elements: the primary ossification center (Fig. 51.1). The transformation is initiated by the invasion of the hypertrophic cartilage core by blood vessels (around embryonic day 14 to 15 in mice) (see later). Lineage tracing genetic studies in mice have shown that, along with the invading vasculature, immature osteolineage cells coenter the developing primary ossification center, with at least a subset of them taking perivascular positions associating closely with the blood vessels.[3] This osteoangiogenic cartilage invasion is accompanied by apoptosis of terminally differentiated hypertrophic chondrocytes, resorption of the calcified cartilage matrix by invading osteoclasts or related "chondroclasts," and deposition of mineralized bone matrix on the remnants of calcified cartilage by perichondrium/periosteum-derived osteoblasts.

With the disappearance of the diaphyseal cartilage (see Fig. 51.1), the remaining chondrocytes, restricted to the opposing ends of the long bone, provide the engine for subsequent bone lengthening. This process is typified by precise temporal and spatial regulation of chondrocyte proliferation and differentiation, with the chondrocytes first flattening out and forming longitudinal columns of rapidly proliferating cells and next, as they reach the ends of the columns closest to the center of the bones, maturing further

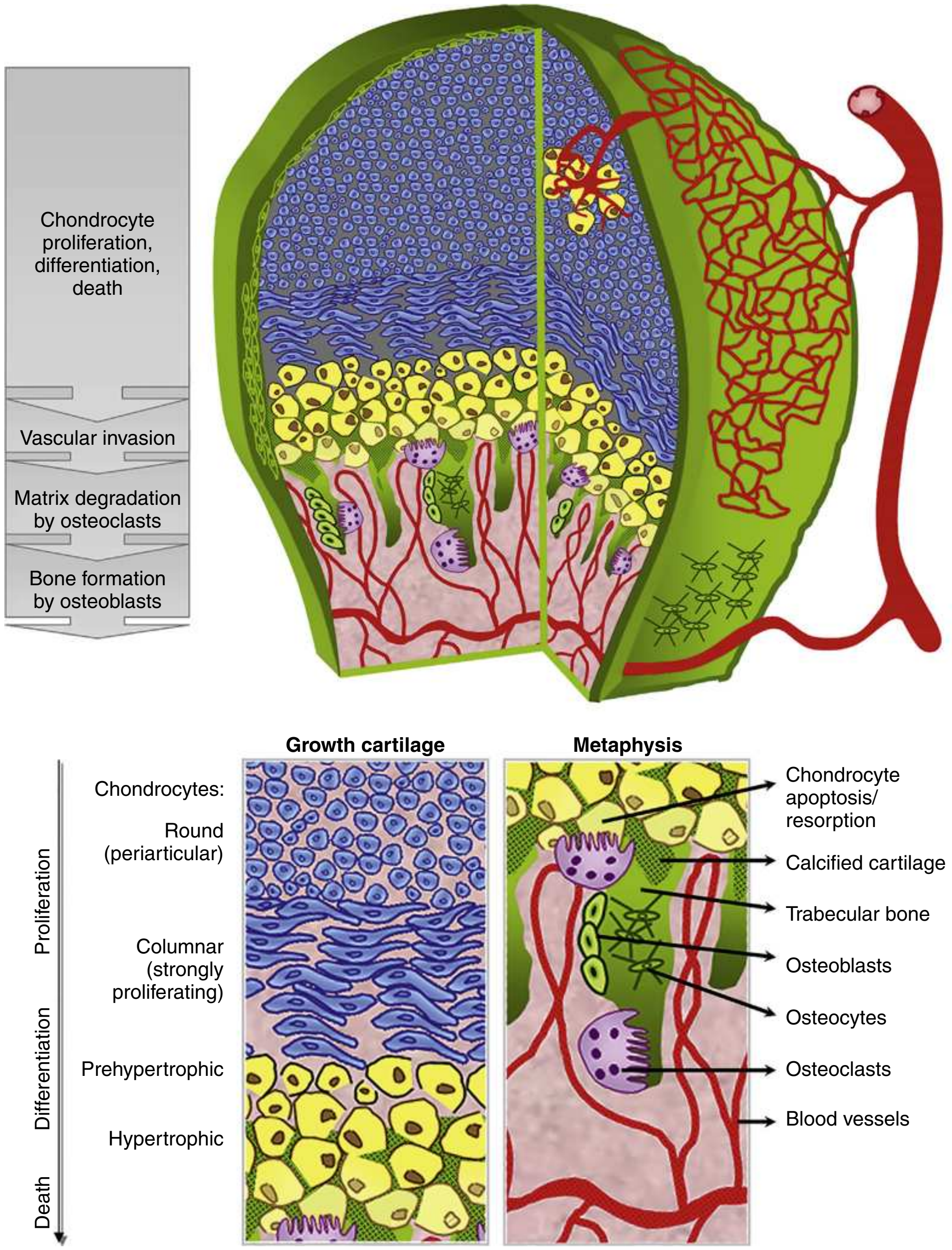

Chondrocyte proliferation, differentiation, death
Vascular invasion
Matrix degradation by osteoclasts
Bone formation by osteoblasts
Growth cartilage
Metaphysis
Chondrocytes:
Round (periarticular)
Columnar (strongly proliferating)
Prehypertrophic
Hypertrophic
Proliferation
Differentiation
Death
Chondrocyte apoptosis/ resorption
Calcified cartilage
Trabecular bone
Osteoblasts
Osteocytes
Osteoclasts
Blood vessels

Fig. 51.2 **Schematic View of the Cellular Structure of Developing Long Bones.** The epiphysis is composed of chondrocytes, organized in layers of proliferation (round periarticular and flat columnar proliferating cells), progressive differentiation toward the metaphysis (prehypertrophic and hypertrophic chondrocytes), and cell death (apoptosis). The avascular cartilage is supplied by the epiphyseal blood vessel network that overlays its surface. In the metaphysis, blood vessels invading the terminal hypertrophic chondrocytes, osteoclasts resorbing the cartilage, and osteoblasts building bone on the cartilage remnants all act coordinately to replace the cartilage anlagen with bone and marrow.

to hypertrophic chondrocytes (Fig. 51.2). Finally, at the border with the metaphysis, many of the terminally differentiated chondrocytes disappear through apoptosis, while a fraction of them likely evolve or transdifferentiate into bone marrow stromal cells (BMSCs) and osteoblast lineage cells (see below).[4,5] Meanwhile, the calcified hypertrophic cartilage matrix is digested by osteoclasts and progressively replaced with cancellous or trabecular bone (forming the primary spongiosa). This process of cartilage turnover and replacement by bone requires adequate neovascularization of the chondroosseous junction by metaphyseal capillaries (see Fig. 51.1 and cellular details in Fig. 51.2).[6] Thus, similar to the initial formation of the primary ossification center, endochondral bone formation at the growth cartilage involves rigorous coupling of vascular invasion with maturation and activity of chondrocytes, osteoclasts, and osteoblasts.

At a certain time (around postnatal day 5 in mice), epiphyseal vessels (Fig. 51.1) derived from the vascular network that overlays the cartilage tissue invade the growth cartilage and initiate the formation of the secondary center of ossification in larger long bones. As a result, discrete layers of residual chondrocytes form true growth plates between the epiphyseal and metaphyseal ossification centers, mediating further postnatal longitudinal bone growth. Ultimately, at least in humans, the growth plates completely disappear (close) at the end of adolescence in a process that actively requires the action of estrogen in both boys and girls, and lengthening stops. Remodeling of existing bone, replacing the primary spongiosa with lamellar bone in the secondary spongiosa and renewing the cortical bone, takes place throughout adult life, ensuring optimal mechanical properties of the skeleton and contributing to mineral ion homeostasis. This continual bone turnover is accomplished through the balanced action of osteoclasts and osteoblasts (see later and Fig. 51.3) and results in a dynamic organization of honeycomb platelike structures or trabeculae in the interior of the bone that are surrounded by blood vessels and bone marrow and housed within the cortical bone.

KEY POINTS

- Most bones form through the process of endochondral bone formation. All the cells on the inside of the mesenchymal condensations become chondrocytes that proliferate and lay down cartilage matrix. The central chondrocytes of the condensation stop dividing, become hypertrophic, and signal to allow vascular invasion, followed by osteoblast differentiation from cells associated with the blood vessels and from some chondrocytes; this establishes the primary ossification center. At the ends of long bones, chondrocytes proliferate, hypertrophy, and attract more blood vessels, leading to formation of the secondary ossification center. Chondrocytes between these centers of ossification form orderly, proliferating columns in the growth plate, leading both to bone lengthening and progressive formation of new bone from dying hypertrophic chondrocytes.

PRINCIPLES OF BONE REMODELING AND SKELETAL HOMEOSTASIS

The Bone Remodeling Cycle

Bone is constantly being renewed by undergoing a process of self-destruction followed by a regenerative process. Depending on mechanical and physiologic needs, osteoclasts resorb particular pockets of existing trabecular and cortical mineralized bone tissue, whereas osteoblasts lay down new bone at appropriate sites to replace the lost bone (Fig. 51.3). Packets of bone that are renewed during remodeling are called bone-remodeling units (BRUs) or bone multicellular units. It has been estimated that the adult human skeleton contains more than a million of these microscopic remodeling foci at any one time. The bone in an activated BRU is first removed by osteoclastic bone resorption through a process that takes a few weeks (see Fig. 51.3). Lost bone is replaced by osteoblastic bone formation, lasting 3 to 4 months for one packet. This discrepancy in the kinetics of bone resorption and formation explains in part how increased resorption, even when accompanied by coupled increased formation, can cause bone loss, for example, in estrogen deficiency or hyperparathyroidism.

How the bone resorption that starts the remodeling cycle itself is initiated remains enigmatic but may involve osteoblast lineage cells perceiving and responding to a remodeling signal by regulating osteoclast formation (see later). The mechanisms that regulate the recruitment of the hematopoietic precursors for the osteoclasts to the bone environment and the remodeling sites are also not very clear, although several findings indicate that the lipid mediator sphingosine-1-phosphate (S1P) may be involved in regulating their movement between the circulation and the bone marrow.[7] Osteoclast precursors, which highly express the S1P receptor 1 (S1PR1), appear to be attracted from the marrow to the bloodstream under physiologic conditions by S1P, whose concentration is higher in blood than in marrow. Conversely, the receptor S1PR2 appears to have a reciprocal function, mediating the attraction of the cells from the bloodstream to remodeling sites. Mice with deletion of *S1PR1* in osteoclast precursors developed excessive bone resorption and osteoporosis because the cells remain in the marrow and differentiate into osteoclasts; mice with deficiency in *S1PR2*, on the other hand, have moderate osteopetrosis owing to reduced bone resorption, because their osteoclast precursors remain in the bloodstream.[7,8]

The signal responsible for the completion of resorption in a remodeling cycle also has not been determined *in vivo*, but several factors that reduce osteoclast formation or activity could play a role (see later). In healthy bones, cessation of resorption is followed by bone formation. During the time lag that occurs between the end of resorption and the beginning of formation, called the reversal phase, small cells are seen on the resorbed surface, perhaps osteoblast precursors that are attracted to the BRU. At this stage of the remodeling cycle, plump cuboidal osteoblastic cells differentiate on the bone surface and deposit the organized matrix, which then becomes mineralized. The bone formation period is completed after approximately 4 months, when the respective packet of bone, either on the surface of a trabeculum or in the cortical haversian canal system, enters a quiescent period until the next remodeling cycle (see Fig. 51.3).

KEY POINTS

- Bone is constantly turning over in an orderly process undertaken by groups of cells called a bone multicellular unit. Probably initiating when osteoblast lineage cells sense microdamage or mechanical forces, signals including sphingosine-1-phosphate draw hematopoietic precursors from the bloodstream that form osteoclasts that then home to the site of injury. Bone resorption over several weeks is followed by a reversal phase with less active bone resorption, followed by several months of bone formation in the region where matrix had been removed by osteoclasts. In this complicated process, cells of the osteoblast and osteoclast lineages differentiate and signal to each other to coordinate the removal and new formation of bone.

Unbalanced Bone Remodeling in Metabolic Bone Diseases

Bone remodeling must be tightly controlled to maintain normal bone homeostasis. Failure of this normal skeletal functioning is one of the most common early manifestations of aging, for example, in osteoporosis and osteoarthritis. These and a variety of other pathologic conditions affecting the skeleton (e.g., rheumatoid arthritis, periodontitis, Paget disease, bone tumors) lead to perturbations in the bone remodeling process, predominantly shifting the balance toward increased degradation of bone owing to local and/or systemic alterations in the levels of hormones or

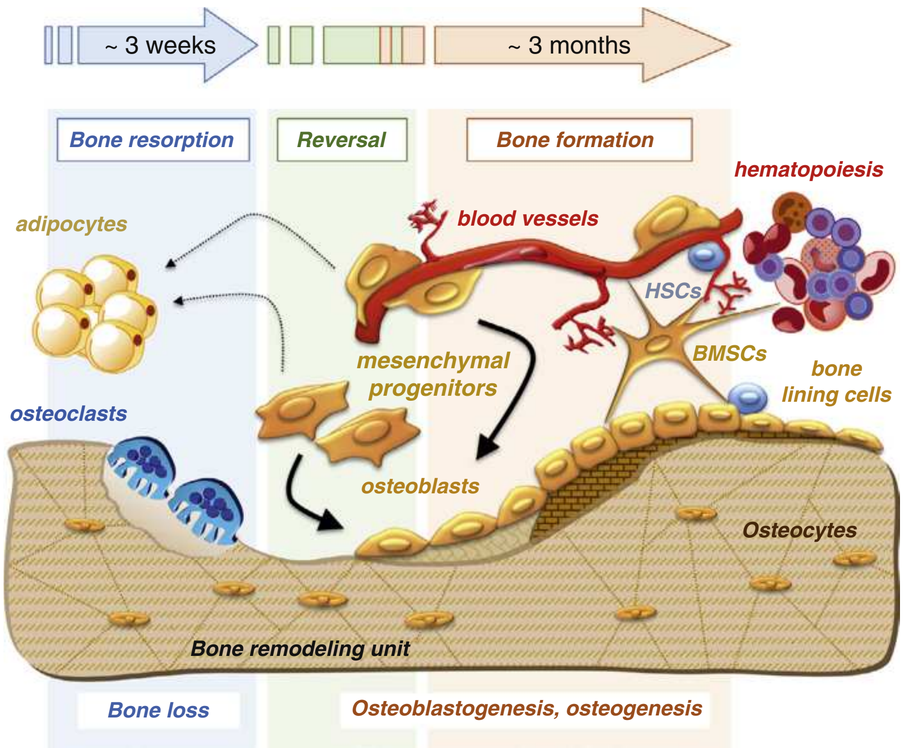


Fig. 51.3 Three-Phase Model of Bone Remodeling. The skeleton is a metabolically active organ that undergoes continuous remodeling throughout life in bone remodeling units. Bone remodeling involves the removal of mineralized bone by osteoclasts, followed by the formation and subsequent mineralization of bone matrix by osteoblasts. The remodeling cycle starts with recruitment of hematopoietic precursors and their differentiation to osteoclasts, induced by osteoblast lineage cells that express osteoclastogenic ligands such as receptor activator of nuclear factor-κB ligand. Osteoclasts become multinucleated and resorb bone. Reversal is marked by switching from bone resorption to bone formation; this is mediated by coupling factors, which can stimulate mesenchymal progenitor recruitment, osteoblast differentiation, and/or cellular activity. During the bone formation phase, the resorbed lacuna is refilled through osteogenesis by osteoblasts that later flatten to form a layer of lining cells on the bone surface or become osteocytes connected by canaliculi within the bone. The process of bone remodeling stands in close relationship with the skeletal vascularization and involves interplay with various cell types in the highly heterogeneous bone and bone marrow environment, including populations of perivascular and reticular bone marrow stromal cells, which can give rise to osteoblasts or marrow adipocytes, as well as form niches for hematopoietic stem cells and support hematopoiesis. *BMSC,* bone marrow stromal cell; *HSC* hematopoietic stem cell.

proinflammatory cytokines that stimulate bone resorption (see later). In cancellous bone, the loss results in thinner trabeculae and in trabecular discontinuity. In cortical bone, endosteal bone resorption causes thinning and sometimes trabecularization of the cortex, increasing the porosity of the cortical bone. The result of these changes is significant weakening of the bones. Osteoporosis, characterized by low bone mass and high risk for debilitating fractures of the vertebrae and long bones, afflicts millions of people, particularly postmenopausal women, and becomes increasingly prevalent with the aging of the general population. Estrogen deficiency increases both the number of sites at which remodeling is initiated and the extent of resorption at a given site (see later). Increased bone resorption is accompanied by increased bone formation as a result of signals that couple these processes (see later). However, the increase in bone formation is not sufficient to maintain bone balance and prevent bone loss.

KEY POINTS

- In metabolic bone disease, the balance between bone resorption and formation is disrupted. A variety of stimuli, for example, those unleashed by estrogen deficiency or aging, trigger an increase of osteoclastic bone resorption, while at the same time defects in bone formation can fail to keep up with the osteoclastic bone loss. The result in trabecular bone is thinning and decrease in number of trabeculae, and in cortical bone a decrease in the amount of cortical bone and increased cortical porosity.

CELLULAR AND MOLECULAR CONTROL OF SKELETAL DEVELOPMENT AND BONE REMODELING

Here, we will review some of the principal growth and transcription factors and signaling cascades known to affect key aspects of bone development and maintenance.

Chondrogenesis and Growth Plate Biology

A crucial step in endochondral bone development entails the differentiation of cells in the mesenchymal condensations into chondrocytes. Some of the mesenchymal cells are osteochondroprogenitor cells with the potential to differentiate into chondrocytes or osteoblasts. The decision to follow a given differentiation pathway is determined by the expression of key transcription factors, likely under the influence of canonical Wnt signaling.[9] As will be discussed in greater detail later, Wnt signaling results in the stabilization of β-catenin, which then can act as a transcription factor and regulate the expression of downstream target genes. Several studies using genetic modifications in mice have provided evidence that in the fetal perichondrium, where cells are destined to become osteoblasts, Wnt signaling is high (at least in part induced by Indian hedgehog [Ihh]), leading to high levels of β-catenin and inducing the expression of factors that mediate osteoblast differentiation (such as runt-related transcription factor 2 [Runx2], a master regulator of osteoblastogenesis, see later), while inhibiting transcription of genes required for chondrocyte differentiation (such as *Sox9*). In the absence of β-catenin during embryonic bone development, these cells become chondrocytes instead of osteoblasts; conversely, in the inner region of the condensations, Wnt signaling must be low in order for these cells to become chondrocytes.[9] The initial conversion of mesenchymal progenitor cells into chondrocytes is driven by the transcription factor Sox9, whereas the combined action of Sox9, 5, and 6 is required to direct the subsequent differentiation of chondrocytes throughout all phases of the chondrocyte lineage.[10]

In the growth cartilage, the most extensively studied cartilage, differentiation of committed cells along the chondrocyte lineage characteristically gives rise to a stratified organization of small, round periarticular chondrocytes (also previously termed resting, or reserve chondrocytes, because these cells proliferate only slowly after birth and can become columnar chondrocytes), flattened, columnar proliferating chondrocytes, and prehypertrophic and hypertrophic chondrocytes (see Figs. 51.2 and 51.4). Because progression of the chondrocytes through these stages is the driving force of actual bone development and growth, it is not surprising that the process is tightly controlled by a myriad of local signaling molecules, the best characterized being BMPs and transforming growth factor-β (TGFβ), FGFs, parathyroid hormone (PTH)-related protein (PTHrP), and Ihh. The importance of these molecules is reflected by the fact that mutations in their receptors have been found to cause severe human dwarfing conditions, such as constitutive activating mutations of FGF receptor (FGFR) 3 in human achondroplasia and thanatophoric dysplasia[11,12] and PTH/PTHrP receptor mutations in Jansen and Blomstrand chondrodysplasia.[13] Here, we will briefly review the basic mechanisms by which these signaling systems control the pace of proliferation and differentiation of growth chondrocytes.

BMPs, secreted proteins belonging to the TGFβ superfamily, transduce signals through serine/threonine kinase receptors, which are homomeric or heteromeric complexes composed of type I and type II receptor subtypes. The ligand binding activates intracellular proteins of the Smad family that relay the BMP signal to target genes in the nucleus. BMPs were first discovered as agents in bone matrix capable of inducing ectopic formation of cartilage and bone after injection into subcutaneous tissues. Subsequent studies have shown that BMPs, some of which are also called growth and differentiation factors, are, in fact, signaling proteins crucial for regulating development in almost all the principal organs and tissues. In the morphogenesis of the skeleton, BMPs have been implicated most strongly in early limb patterning, as well as in the subsequent processes of chondrogenesis and osteogenesis (also see further) (Fig. 51.4).[14]

The FGF pathway similarly involves multiple ligands and receptors in regulating skeletal development.[12] FGF18 and FGF9 are the most important FGF ligands in regulating chondrogenesis identified so far. These ligands activate FGFR3 expressed on proliferating chondrocytes (Fig. 51.4). Activating mutations in the *FGFR3* gene cause dominantly inherited human dwarfing chondrodystrophies because of impaired chondrocyte proliferation,[11] and FGFR3 inactivation in mice increases chondrocyte proliferation and prolongs growth.[12] This finding was

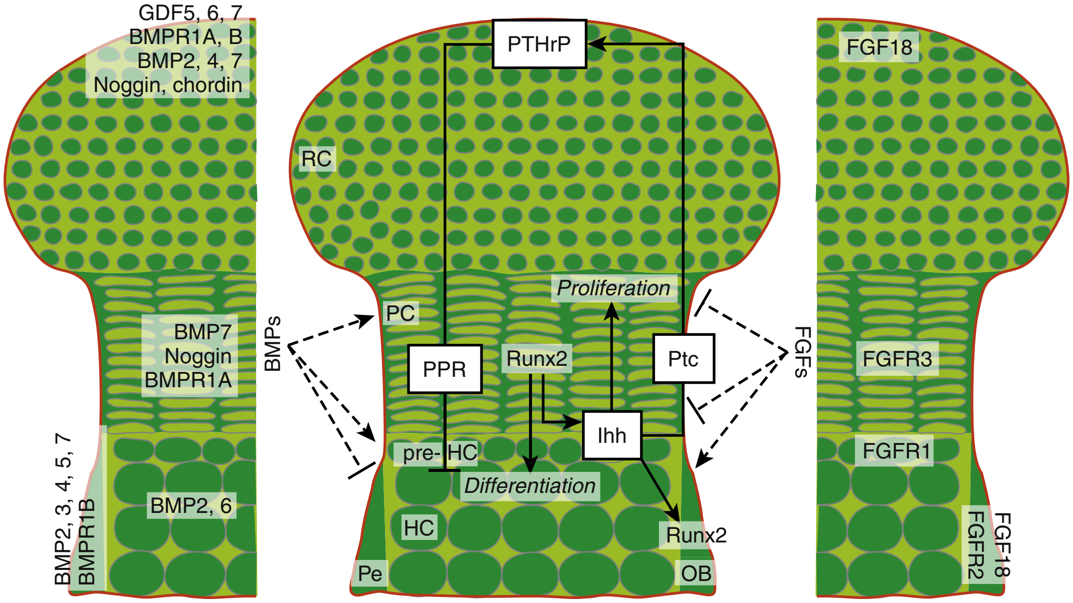


Fig. 51.4 Regulation of Chondrocyte Proliferation and Differentiation. Schematic representation of the epiphysis, with zones of periarticular round chondrocytes (*RC*), strongly proliferating columnar chondrocytes (*PC*), and (pre-)hypertrophic chondrocytes ([pre-]HC). Hypertrophic cartilage is surrounded by perichondrium (Pe), where osteoblasts (*OB*) develop. Centrally, the Indian hedgehog (Ihh)/parathyroid hormone-related protein (PTHrP) negative feedback loop is indicated. Cells at the articular end of the bone secrete PTHrP, acting through the PTH/PTHrP receptor (*PTHR1* or *PPR*) on columnar chondrocytes to keep them in the proliferating pool and delay their differentiation. When the distance from the bone end becomes sufficiently large, cells escape the differentiation block and become Ihh-expressing prehypertrophic chondrocytes. In an as yet unknown manner, Ihh induces PTHrP expression. In addition, Ihh directly stimulates chondrocyte proliferation and converts perichondrial cells into OBs by inducing runt-related transcription factor 2 (*RUNX2*), thereby determining the site of bone collar formation. RUNX2 also favors Ihh expression and hypertrophic differentiation. Several bone morphogenetic protein (*BMP*) and fibroblast growth factor (*FGF*) family members are expressed at specific sites in the growth cartilage (left and right panels). BMPs and FGFs act antagonistically to modulate chondrocyte proliferation, Ihh expression, and terminal hypertrophic differentiation. *GDF*, Growth and differentiation factor; *Ptc*, patched; *BMPR,* BMP receptor; *FGFR*, FGF receptor.

surprising, given that FGFs in most tissues act as potent mitogens; although this may pertain in early chondrogenesis, in which FGF18 stimulates proliferation, later in development signaling through FGFR3 constitutes a master block on chondrocyte proliferation through activation of STAT1, which activates the cell cycle inhibitor p21$^{Waf1/Cip1}$.[12] In addition, although previous *in vitro* experiments had indicated that FGF signaling accelerates the late steps of chondrocyte hypertrophy, mouse models of achondroplasia and thanatophoric dysplasia have shown delayed hypertrophic differentiation of chondrocytes, strongly suggesting that FGFR3 signaling *in vivo* inhibits chondrocyte differentiation.[12] This effect is relayed by activation of the mitogen-activated protein kinase (MAPK) pathway in chondrocytes, which affects longitudinal growth by regulating hypertrophic chondrocyte differentiation and matrix deposition.

A major regulatory system in growth chondrocytes is provided by Ihh, a member of the conserved family of hedgehog proteins, and by PTHrP.[13] The Ihh/PTHrP pathway forms a negative feedback loop that regulates the onset of hypertrophic differentiation (see Fig. 51.4). Hedgehog proteins, upon binding to the Patched (Ptc) receptor, signal through the seven-pass transmembrane protein Smoothened to regulate gene transcription through both derepression and activation of the Gli family of transcription factors. Ihh is produced by prehypertrophic and early hypertrophic chondrocytes and signals through Ptc to stimulate expression of PTHrP by chondrocytes located near the periarticular ends of bone. Increased production of Ihh leads to increased PTHrP production. PTHrP in turn signals back to its receptor PTHR1—the PTH/PTHrP receptor that also responds to PTH in osteoblasts and kidney—that is expressed at low levels by proliferating chondrocytes and at high levels by prehypertrophic cells. This signaling slows down chondrocyte differentiation and keeps chondrocytes in the proliferative state (see Fig. 51.4). This slowed differentiation consequently delays the generation of cells that can produce Ihh; therefore, the production of PTHrP is lowered. Thus, a negative feedback signaling pathway allows PTHrP and Ihh to regulate each other and, in turn, to control the pace of chondrocyte development in the growth plate.[13,15,16] It is important to note that the phenotype of *Ihh*$^{-/-}$ mice[17] revealed that Ihh has additional functions in endochondral bone formation, independent of regulation of PTHrP production. Indeed, these mutant mice also displayed a marked decrease in chondrocyte proliferation and absence of mature osteoblasts and bone collar formation.[17] The regulation of chondrocyte proliferation represents a direct action of Ihh signaling to chondrocytes. Ihh also accelerates the conversion of round periarticular chondrocytes to flat columnar chondrocytes. Moreover, Ihh (together with other, yet unidentified effectors) stimulates the expression of Runx2 in perichondrial cells. Under the influence of this essential transcription factor for osteoblast development (see later), the perichondrial cells are driven into the osteoblast lineage; hence, Ihh-induced direct actions also provide control of bone collar formation. Further, Runx2 in chondrocytes regulates Ihh expression and also regulates further chondrocyte differentiation into the hypertrophic stage, as revealed by chondrocyte maturation abnormalities in *Runx2*-null mice and conditional mutants deleting or overexpressing *Runx2* in cartilage.[18] Thus, it is evident that Ihh is a master regulator of endochondral bone development, coordinating chondrocyte proliferation, chondrocyte maturation, and osteoblast differentiation (see Fig. 51.4). A continued role for both PTHR1 and Ihh in the postnatal growth plate has been revealed by employing inducible mutagenesis in mice; inactivation of either gene in juvenile mice led to premature closure of the growth plate.[19,20]

Each of these pathways mentioned previously has multiple mechanisms for communicating with each other as they regulate chondrocyte and osteoblast differentiation.[21] Moreover, aside from these factors, chondrogenesis is affected by several other growth factors, cytokines, and hormones, including growth hormone (GH) and insulin-like growth factors (IGFs), C-natriuretic peptide, retinoids, thyroid hormone, estrogen, androgen, 1,25-dihydroxyvitamin D_3 [$1{,}25(OH)_2D_3$], glucocorticoids, and others.

In addition, importantly, the chondrocytes in the cartilage templates and growth plates of the endochondral bones typically develop over a prolonged period in the absence of blood vessels. As an avascular tissue, cartilage consequently faces the challenge of hypoxia.[22] The main mediator of cellular responses to hypoxia is the transcription factor hypoxia-inducible factor (HIF)-1.[23] HIF-1 consists of two subunits, HIF-1α and HIF-1β. HIF-1β is constitutively expressed in an oxygen-independent manner, while HIF-1α is activated only in hypoxic conditions (when oxygen levels drop to <5%). In conditions of higher oxygen tension, a class of 2-oxoglutarate–dependent and iron-dependent prolyl hydroxylases modify HIF-1α such that the protein becomes recognized by the von Hippel Lindau protein and targeted for destruction by this ubiquitination complex. The half-life of HIF-1α is consequently very short. Hence, specifically under hypoxic conditions, HIF-1α can accumulate, translocate to the nucleus, dimerize with HIF-1β, and bind to hypoxia response elements within the promoters of hypoxia-responsive genes.[22,23] To date, more than 100 putative HIF-1 target genes have been identified that are involved in a variety of biological processes; primary roles of HIF-1 transcriptional activity are induction of anaerobic metabolism (for instance, by regulating glycolytic genes), angiogenesis (through induction of vascular endothelial growth factor [VEGF]), and erythropoiesis (via increased erythropoietin expression).[23]

HIF-1α is absolutely essential for the survival and controlled differentiation of hypoxic chondrocytes: mice with conditional inactivation of HIF-1α displayed aberrant apoptosis of chondrocytes located in the center of the growth cartilage, farthest from blood vessels.[24] In addition, HIF-1α influences the cartilaginous matrix, such as by improving the efficiency of posttranslational modifications of type II collagen.[25] The mechanisms by which HIF-1 ensures the survival of hypoxic chondrocytes include the direct activation of genes that enable chondrocytes to switch to oxygen-sparing metabolic pathways, as well as the indirect consequences of induction of VEGF.[22,24] Deletion of *Vegf* from cartilage also results in massive apoptosis of nonhypertrophic, hypoxic chondrocytes located in the center of the growth plate, which could point at a cell-autonomous survival function effected by VEGF on chondrocytes. Alternatively, or additionally, VEGF expression in proliferating chondrocytes (in response to hypoxia and/or HIF-1) seems critical for stimulating vascularization in the immediate surroundings of the cartilage, thereby ensuring sufficient oxygen diffusion to the avascular cartilage to prevent cell death.[22] In addition to regulating blood vessel formation in the perichondrium, VEGF also induces blood vessel invasion and growth within the endochondral bones during skeletal development and growth, as discussed next.[6,26]

KEY POINTS

- Many paracrine factors and transcription factors regulate the survival, proliferation, and differentiation of growth cartilage. Bone morphogenetic proteins drive the generation of chondrocytes from mesenchymal cells and also their further proliferation and differentiation. Indian hedgehog (Ihh), made by prehypertrophic chondrocytes, directly stimulates the proliferation of adjacent chondrocytes and also stimulates the production of parathyroid hormone–related protein (PTHrP); PTHrP keeps chondrocytes proliferating. Ihh, along with subsequent Wnt signaling, also directs adjacent perichondrial cells to become osteoblasts. Perichondrial cells secrete fibroblast growth factors (FGFs) that activate FGF receptor 3 (FGFR3) on chondrocytes; this action slows the proliferation and differentiation of chondrocytes. Constitutively active mutations of FGFR3 lead to achondroplasia, a disease of short stature. Hypoxia inducible factor (HIF) is crucial for directing the metabolism of chondrocytes in the avascular growth plate and also induces the synthesis of vascular endothelial growth factor (VEGF).

Skeletal Vascularization

Both temporally and spatially, chondrocyte differentiation is followed by vascular invasion of the terminally differentiated, hypertrophic cartilage. This step is an absolute requirement for endochondral bone development and growth to proceed, as physical blockage of blood vessel invasion into the hypertrophic cartilage of fetal skeletal explants completely halts their development, and blocking the blood supply of bone *in vivo* results in reduced longitudinal growth.[6,26] During endochondral ossification, the initially avascular cartilage template is replaced by highly vascularized bone and marrow tissue through three consecutive vascularization events (see Figs. 51.1 and 51.5).[21,26] First, initial vascular invasion of the cartilage anlagen during embryonic development (sometimes called quiescent angiogenesis) involves endothelial cells invading from the perichondrial tissues and organizing into immature blood vessels in the primary center of ossification. Second, capillary invasion at the metaphyseal border of the growth cartilage is required for rapid bone lengthening (see Figs. 51.1, 51.2, and 51.5). Third, vascularization of the cartilage ends initiates the formation of secondary ossification centers (see Figs. 51.1 and 51.5). Hence, a process of cartilage neovascularization initiates each of the subsequent stages of endochondral replacement of cartilage by bone; at all times, this cartilage neovascularization is preceded by chondrocyte hypertrophy. Immature chondrocytes produce angiogenic inhibitors, such as chondromodulin-I and troponin-I, rendering the cartilage resistant to vascular invasion. In contrast, when chondrocytes become hypertrophic, they switch to production of angiogenic stimulators, including the potent angiogenic stimulator VEGF, and become a target for capillary invasion and angiogenesis. VEGF is expressed at very high levels by hypertrophic cartilage, as well as by osteoblasts and osteoclasts. Hypoxia represents a crucial trigger of VEGF expression in chondrocytes, osteoblasts, and possibly osteoclasts, through mechanisms that involve HIF both *in vitro* and *in vivo*. VEGF transcription may also be induced by Runx2 and by Osterix (Osx) (see later). In addition, several hormones (including PTH, GH, and 1,25$(OH)_2D_3$) and locally produced growth factors (e.g., FGFs, TGFβ, BMPs, IGFs, platelet-derived growth factor [PDGF]) have been demonstrated to be involved, at least *in vitro*, in the regulation of VEGF expression.[26]

Mouse genetic studies have demonstrated multiple essential roles of VEGF and its splice isoforms in endochondral ossification. Although ablation of one or both *Vegf* gene alleles in all cells causes early lethality in mice, several alternative mutagenesis approaches have been developed to investigate the role of VEGF in bone. These models include Cre/LoxP-mediated conditional inactivation of the *VEGF* gene (*VEGFa)* in type II collagen–expressing chondrocytes or in Osx-expressing osteoprogenitors, inactivation of two of the three major VEGF isoforms (leaving expression of only one isoform), and skeletal tissue-specific overexpression of VEGF in genetically engineered mice (reviewed in Maes[26]). These models demonstrated that VEGF is a key inducer of all three key vascularization stages of endochondral bone development (Fig. 51.5), but also acts as a direct modulator of bone development by affecting various cell types. Perichondrial cells, osteoblasts, and osteoclasts express VEGF receptors and respond to VEGF signaling through enhanced recruitment, differentiation, activity, and/or survival (for detailed reviews and references, see Maes [21,26]). In most of these functions, VEGF works in paracrine mode, but intracrine VEGF effects in (Osx-expressing) osteoprogenitors have also been reported, determining cell fate decisions in these cells; in the absence of intracellular VEGF signaling, the cells differentiated into adipocytes rather than osteoblasts.[27] Altogether, the pleiotrophic actions of VEGF on various bone cells contribute to the tight coordination of vascularization, ossification, and matrix resorption needed for endochondral ossification.

The prevailing model for VEGF action in endochondral bone development is as follows (Fig. 51.5): VEGF is secreted at particularly high levels by hypertrophic chondrocytes. VEGF is partially soluble (the short VEGF isoforms) and partially bound to constituents in the cartilage matrix (the long isoforms), from where it is further released as the terminal cartilage undergoes progressive decay in the endochondral turnover process. VEGF attracts blood vessels toward the chondroosseous junction and stimulates endothelial cells to form new blood vessels (angiogenesis), which will be indirectly associated with increased delivery of osteoblast and osteoclast progenitors.[26] At the same time, VEGF signaling directly stimulates the recruitment and differentiation of osteoblasts to form bone.[28] VEGF also works as a chemoattractant stimulating osteoclast invasion of cartilage, and enhances osteoclast differentiation, survival, and resorptive activity.[29] This likely activates a positive feedback system, as increased osteoclasts and osteoclast-derived matrix metalloproteinase (MMP) 9[30] can release more matrix-bound VEGF from the cartilage that is being resorbed (Fig. 51.5).[21,26]

The skeletal vascular system remains a vitally important player in the regulation of bone health beyond development and growth. Optimal tissue vascularization is essential for adequate homeostatic bone remodeling, and bone fracture repair and regeneration critically rely on revascularization of the injury site. In fact, any type of bone formation during growth, homeostasis, disease, or repair occurs in close spatial and temporal association with vascularization of the ossified tissue, a principle termed angiogenic-osteogenic coupling.[31] The vascular system is crucial, through its intrinsic function to supply oxygen, nutrients, and growth factors/hormones to bone cells as required for their specified activities. In addition, the blood vessels serve to bring in (precursors of) osteoclasts that will degrade the cartilage or bone extracellular matrix, to remove endproducts of the resorption processes, and to bring in progenitors of osteoblasts that will deposit bone.[3,26,31] Hypoxia-driven pathways, including HIF and VEGF signaling, and the Notch pathway represent major regulators of angiogenic-osteogenic coupling in postnatal bone remodeling and ageing.[22,32,33]

KEY POINTS

- Vascular endothelial growth factor (VEGF), secreted from hypertrophic chondrocytes, is essential for attracting the blood vessels and osteoblast precursors that generate the primary and secondary ossification centers. VEGF acts on receptors on the surface of endothelial cells and cells of the osteoblast lineage to direct their mobility, proliferation, and differentiation, and VEGF also attract osteoclasts. Hypoxia-inducible factor (HIF)-1α and various paracrine factors direct the synthesis of multiple spliced isoforms of VEGF.

Skeletal Stem and Progenitor Cells

Considerable progress has been made recently in the identification in both mice and humans of populations of skeletal stem cells (SSCs) and SSPCs, an overarching inclusive term encompassing bone fide SSCs as well as tri/bipotent and lineage-committed osteogenic progenitor cells (Fig. 51.6). SSCs have been identified in the growth plate and the fetal perichondrium and postnatal periosteum, as well as among the reticular and perivascular mesenchymal stromal cells dispersed throughout the bone marrow[34-41] (also see Fig. 51.3). They have the ability to give rise to cells of the prime skeletal lineages—osteoblasts, chondrocytes, and marrow adipocytes (fat cells)—with each of the distinct cell fate decisions and differentiation processes being driven by lineage-specific transcription factors (as indicated in Fig. 51.6).[42] BMSCs include a variety of osteogenesis-competent SSPC populations characterized by the expression of specific

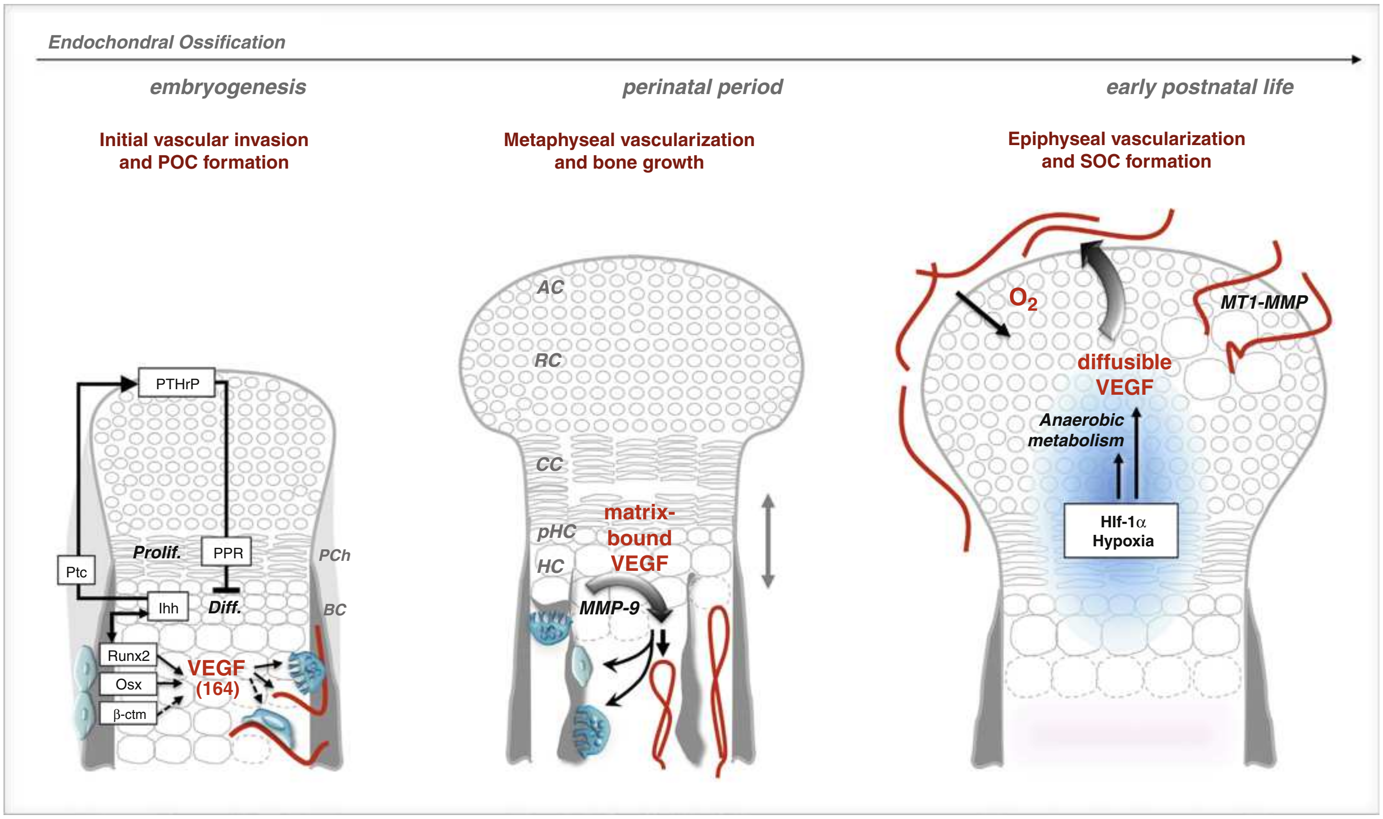


Fig. 51.5 Vascular Invasion During Endochondral Bone Development and the Role of Vascular Endothelial Growth Factor (*VEGF*). *Left panel,* During embryogenesis, the initiation of cartilage neovascularization is coordinated with proliferation and differentiation of chondrocytes in the avascular anlagen, osteoblastogenesis and bone collar (BC) formation in the surrounding perichondrium (PCh), and the expression of high levels of VEGF by hypertrophic chondrocytes in the middle portion of the bone model. Invasion by blood vessels is accompanied by chondrocyte apoptosis, osteoclastic cartilage resorption, and coinvasion by osteoprogenitors moving to the inside of the nascent primary ossification center (*POC*). For details on the molecular regulators, see text. *Middle panel,* Longitudinal bone growth is driven by the progressive replacement of cartilage with trabecular bone, a process that relies on metaphyseal vascularization mediated by VEGF. VEGF is produced at high levels by hypertrophic chondrocytes (*HCs*) and is sequestered in part in the cartilage matrix upon its secretion. Trapped VEGF can be released from the matrix by proteases such as matrix metalloproteinase (*MMP*)-9 secreted by osteoclasts/chondroclasts during cartilage resorption. VEGF then can bind to its receptors (*VEGFRs*) on endothelial cells and stimulate the guided attraction of blood vessels to invade the terminal cartilage. Osteoblasts and osteoclasts also express VEGF receptors, and VEGF can affect their differentiation and function both directly and indirectly by enhancing metaphyseal vascularization, thus playing a coordinative role in the key events that mediate bone development and growth. *AC,* Articular chondrocytes; *RC,* round chondrocytes; *CC,* columnar chondrocytes; *pHC,* prehypertrophic chondrocytes). *Right panel,* When the avascular cartilage exceeds a critical size during growth, the middle of the growth plate becomes hypoxic, stabilizing hypoxia-inducible factor (*HIF*)-1 and triggering VEGF expression in immature chondrocytes. Soluble VEGF isoforms are critical to diffuse to the surface of the cartilage, stimulating growth of the epiphyseal vascular network and tissue oxygenation. HIF-1 is additionally essential for chondrocyte survival by inducing oxygen-sparing anaerobic metabolism in the hypoxic cells. In early postnatal life, epiphyseal vessels invade the avascular cartilage and initiate a secondary center of ossification (*SOC*). (Reproduced from Maes C. Role and regulation of vascularization processes in endochondral bones. *Calcified Tissue Int.* 2013;92:307–323.)

markers, including CD146 in humans and leptin receptor, C-X-C motif ligand 12 (“CAR cells”), PDGFRβ, or PDGFRα and Sca-1 (“PαS” cells), among others, in mice.[34-36,43] Marker-based identification enables researchers to visualize the dedicated cell populations in the heterogeneous bone and bone marrow environment and isolate them for detailed studies, using advanced imaging, *in vivo* lineage tracing, flow cytometry, and high-resolution bulk and single-cell RNA sequencing technologies. These tools have also been providing new insights into the relationships among the cell lineages in bone, including the reactivation of quiescent bone lining cells upon bone anabolic stimuli such as intermittent PTH or antisclerostin antibody treatments,[44-46] adipogenic differentiation of presumed committed osteolineage cells, and transdifferentiation of hypertrophic chondrocytes into osteoblasts[4,5,47] (Fig. 51.6).

Functionally, besides providing pools of reserve cells for osteoblast differentiation and osteogenesis during homeostatic bone remodeling and fracture repair, SSPC populations in the bone marrow environment also critically support hematopoiesis by constituting functional niches regulating the maintenance and differentiation of hematopoietic stem cells (reviewed in Pinho & Frenette[48] and Crane et al.[49]). The existence of a perivascular niche for SSPCs, residing around bone marrow arteries, arterioles, and sinusoids may

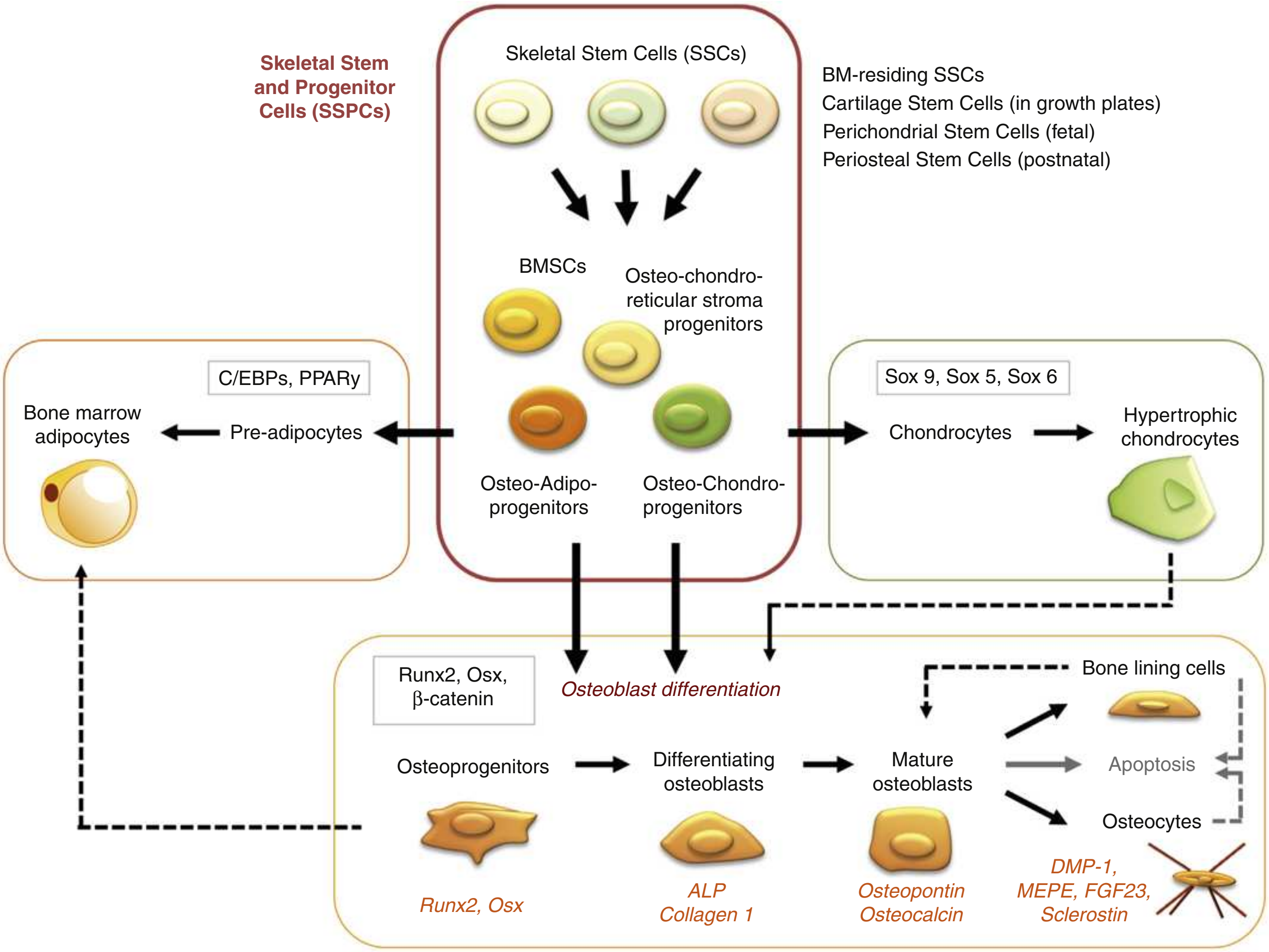

Skeletal Stem and Progenitor Cells (SSPCs)
Skeletal Stem Cells (SSCs)
BM-residing SSCs
Cartilage Stem Cells (in growth plates)
Perichondrial Stem Cells (fetal)
Periosteal Stem Cells (postnatal)
BMSCs
Osteo-chondro-reticular stroma progenitors
Osteo-Adipo-progenitors
Osteo-Chondro-progenitors
C/EBPs, PPARγ
Bone marrow adipocytes
Pre-adipocytes
Sox 9, Sox 5, Sox 6
Chondrocytes
Hypertrophic chondrocytes
Runx2, Osx, β-catenin
Osteoblast differentiation
Osteoprogenitors
Differentiating osteoblasts
Mature osteoblasts
Bone lining cells
Apoptosis
Osteocytes
Runx2, Osx
ALP
Collagen 1
Osteopontin
Osteocalcin
DMP-1,
MEPE, FGF23,
Sclerostin

Fig. 51.6 Differentiation of Skeletal Stem and Progenitor Cells (*SSPCs*) Into the Prime Skeletal Cell Lineages, With A Focus on Osteoblasts. Within the skeleton, various populations of skeletal stem cells (*SSCs*) and SSPCs exists in distinct locations. In endochondral bones SSCs have been identified in the fetal perichondrium, in the cartilaginous growth plate during the intense growth spurt of perinatal and juvenile mice, and in the bone marrow (*BM*) and periosteum during postnatal life, including adult bone homeostasis and fracture repair settings. The BM further harbors diverse SSPC populations that can give rise to osteoblasts, adipocytes, stroma, and/or chondrocytes, collectively designated BM stromal cells (*BMSCs*). The heterogeneous SSPC populations in the bone and BM environment appear to include reticular stromal cells, perivascular cells, osteochondroreticular cell progenitors, and more restricted bipotent progenitors with osteochondral differentiation potential (particularly important intermediate cells in fetal development) or osteoadipogenic capacities (found in postnatal bone). Lineage differentiation of the multipotent mesenchymal precursors into the three major skeletal lineages is controlled by respective sets of transcription factors as indicated. For osteolineage specification and differentiation, the actions of RUNX2, Osx and β-catenin are crucial. Progressive osteoblast differentiation is characterized by changes in gene expression typifying specific stages or subsets of osteolineage cells as indicated (*orange*). Mature osteoblasts ultimately undergo apoptosis, flatten to form a layer of bone lining cells on the bone surface, or become embedded within the bone as osteocytes. Recent studies indicate that, in addition to the classical, lineage-committed osteoprogenitors, hypertrophic chondrocytes may also constitute a source of osteoblasts during development and growth. In addition, quiescent bone lining cells can convert into mature osteoblasts in postnatal bone upon induction of active bone formation. Lastly, some populations of osteoprogenitors may also revert to an adipogenic differentiation state. (Reproduced from Mesnieres M, Maes C. Osteoblast lineage stem and progenitor cells. In: Zaidi M, ed. *Encyclopedia of Bone Biology. Section: Osteoblast Biology*. Oxford: Academic Press; 2020:383–396.)

also relate to the absolute necessity of blood vessels for bone formation and regeneration, a concept known as angiogenic-osteogenic coupling (see earlier).[22,33] Finally, SSPC cell fate decisions favoring adipogenesis can result in excessive bone marrow adipose tissue (BMAT) accumulation, as for instance seen with aging and pathophysiological conditions such as obesity, but also anorexia nervosa, bone trauma with marrow disruption, and cancer therapies such as irradiation and chemotherapy. Besides possibly shifting precursor cells away from osteoblastogenesis, leading to reduced bone formation, BMAT itself is thought to play a role in the skeletal contributions to systemic energy metabolism and glucose homeostasis and to affect hematopoiesis.[50,51]

KEY POINTS

- Investigators have identified cells in the bone marrow, perichondrium/periosteum, and growth plate that can form colonies *in vitro* and can differentiate into osteoblasts, chondrocytes, and adipocytes. These properties are those expected of stem/progenitor cells in bone. More recently, these studies have been combined with lineage tracing technology *in vivo*, along with single-cell sequencing and cell sorting technologies, to bring molecular rigor to the still-evolving characterization of skeletal stem and progenitor cells (SSPCs).

Osteoblastogenesis

A key step required to rebuild cartilage to bone during development consists of the differentiation of osteoblasts and the deposition of mineralized bone tissue by this specialized cell type (see Figs. 51.3 and 51.6). As outlined earlier, osteoblasts are mesenchymal cells that can descend from various SSPC populations that have the capacity to differentiate into osteoblast lineage cells through the expression of β-catenin, Runx2, and Osx; these transcription factors will be discussed in detail.[18,52,53]

Committed preosteoblasts further differentiate into mature osteoblasts, cuboidal cells on the bone surface that secrete large amounts of type I collagen and other matrix components, which together are called osteoid. Subsequently, osteoblasts direct the mineralization of the osteoid in a process that requires active alkaline phosphatase, expressed on the osteoblast membrane. These mature osteoblasts express characteristic genes, including the gene encoding osteocalcin. Ultimately, the cells die through apoptosis, are converted into thin, quiescent bone lining cells, or become embedded within the bone matrix as osteocytes. Osteocytes are found dispersed throughout the bone matrix and have the potential to live as long as the organism itself. In fact, osteocytes are the most abundant cell type in bone, constituting over 90% of adult bone cells.[54,55] As osteoblasts become encased by the mineralized bone matrix that they themselves synthesized, they dramatically change their shape, sending out numerous dendritic processes that run inside lacunar canaliculi. These processes allow osteocytes to communicate with each other, with the vasculature, and with cells on the bone surface.[55] Osteocytes are well positioned to sense and respond to mechanical forces, and a growing body of evidence documents that osteocytes mediate the actions of mechanical forces to increase bone mass (see later).[54-56] Osteocytes synthesize a series of proteins unique to this most differentiated member of the osteoblast lineage. These include dentin matrix protein 1, matrix extracellular phosphoglycoprotein, and FGF23, secreted proteins involved in systemic phosphate homeostasis, and sclerostin (SOST), an inhibitor of Wnt action (see later).[54] Through their actions to regulate bone resorption and formation, as well as phosphate and calcium homeostasis, osteocytes play a central role in the determination and maintenance of bone structure and function (see later).

A number of crucial transcription factors and signaling pathways determining osteoblastogenesis have been identified, and at least some of those play a role in inherited human bone disease.[18,52,53]

Wnt/β-Catenin Signaling. Wnts are a large family of secreted growth factors (19 different members in mouse and human genomes) that play essential roles in multiple developmental processes. Wnts are also required for adult tissue maintenance, and perturbations in Wnt signaling can lead to tumor formation and other diseases.[57] In the skeletal system, mutations in Wnt signaling components lead to skeletal malformations and diseases such as osteoporosis-pseudoglioma and osteoarthritis.

Wnts can transduce their signals through several different downstream signaling pathways. The best understood pathway is the canonical or Wnt/β-catenin pathway.[57] Central to this pathway is the regulation of the protein stability of β-catenin, which acts as a transcription factor in the canonical Wnt pathway and is also involved in cell adhesion by binding to cadherins. In the absence of Wnts, cytoplasmic β-catenin is constitutively degraded through its phosphorylation by glycogen synthase kinase 3-β in a large protein complex brought together by axin and adenomatous polyposis coli.[57] Phosphorylated β-catenin is recognized by a β-transducin repeat containing protein (β-TrCP) that targets it for proteasome-mediated degradation. In the presence of Wnt stimulation, the Wnt ligands bind to two synergistically acting families of Wnt (co)receptors: the Frizzled receptor family members and low-density lipoprotein receptor-related proteins (LRP5 or LRP6). This interaction results in the recruitment of axin to LRP5/6 in the plasma membrane; this sequestering of axin at the plasma membrane may lead to disassembly of the β-catenin destruction complex or, alternatively, induce the association of the complex with phosphorylated LRP5/6, blocking ubiquitination of β-catenin by β-TrCP. In either model, the degradation of β-catenin is inhibited, and the protein is stabilized. Stabilized β-catenin protein then accumulates in the cytoplasm and translocates to the nucleus, where it interacts with members of the T cell factor/lymphoid enhancer factor family of DNA-binding transcription factors to regulate the expression of downstream target genes.[57]

During the past two decades, canonical Wnt signaling has been shown to play a significant role in the control of osteoblastogenesis and bone formation. Genetic studies analyzing conditional β-catenin loss- and gain-of-function mouse models provided compelling evidence that β-catenin is a crucial transcription factor determining the osteoblast lineage commitment of mesenchymal progenitors.[58] During development, β-catenin is indispensible for suppressing chondrocytic differentiation of bipotent osteochondroprogenitor cells and for stimulating them to differentiate into mature osteoblasts.[9] Inactivation of β-catenin in mesenchymal progenitor cells blocked osteoblast differentiation, and mesenchymal cells in the perichondrium and calvarium differentiated into chondrocytes instead. In postnatal life, deletion of the gene encoding β-catenin in Osx-expressing osteoprogenitors inhibited their differentiation into mature osteoblasts, and instead induced the expression of adipocytic markers.[59,60]

The roles of the canonical Wnt pathway in osteoblasts have been clarified largely through genetic studies in mice.[58] Combined deletion of the key Wnt coreceptors LRP5 and LRP6 in the embryonic mesenchyme largely recapitulated the loss of β-catenin, the phenotypes being characterized by impaired osteoblast differentiation in the developing mutant mouse skeletons. Similarly, LRP5-deficient mice developed a low bone mass phenotype postnatally owing to reduced osteoblast number and function, and additional deletion of one allele of *LRP6* led to a further reduction in bone mass.[58]

Clinically, mutations in the Wnt receptor complexes have been amply associated with changes in bone mineral density and fractures.[61] Loss-of-function mutations in LRP5 receptors cause osteoporosis-pseudoglioma syndrome, a form of juvenile-onset osteoporosis, in humans.[62] Conversely, gain-of-function mutations that render LRP5 refractory to secreted antagonists of canonical Wnt signaling such as Dickopff-related 1 (DKK1) and sclerostin (encoded by the *SOST* gene), result in high bone mass phenotypes.[63] These protein inhibitors of Wnt signaling have additional clinical relevance. For example, increased production of DKK1 appears to contribute to the bone loss in patients with multiple myeloma.[64] Moreover, homozygous loss-of-function mutations in SOST have been identified in patients diagnosed with sclerosteosis, a disfiguring disease associated with high bone mass. Mutations affecting regulation of SOST transcription cause van Buchem disease, a disease that resembles sclerosteosis.[61] These findings highlight the important role of canonical Wnt signaling in regulating human bone mass.

Runx2 and Osx. Runx2 (previously known as Cbfa1, OSF2, or AML3), a transcription factor of the ancient runt family, has been demonstrated to be absolutely essential for the induction of osteoblast differentiation and endochondral and intramembranous bone formation. RUNX2-deficient mice completely lack osteoblasts and do not form bone at all; instead, a completely cartilaginous skeleton develops without any true bone matrix.[65,66] In heterozygous (haploinsufficient) *RUNX2* mutant mice, the defect in osteoblast differentiation is limited to intramembranous bones.[65] The resulting phenotype in these mice closely resembles the cleidocranial dysplasia syndrome in humans, a dominantly inherited developmental disorder of bone, in which *RUNX2* was found to be mutated in most patients.[67,68]

Runx2 can be sufficient to induce osteoblast differentiation *in vitro*. Moreover, *RUNX2*-null calvarial cells spontaneously differentiate into adipocytes and differentiate into chondrocytes in the presence of BMP-2 *in vitro*, but they do not differentiate into osteoblasts. Thus, Runx2 is both sufficient and essential for differentiation of mesenchymal cells into osteoblasts, and it inhibits their differentiation into adipocytes and chondrocytes. Runx2 mediates osteoblast differentiation by regulating the expression of a variety of bone matrix protein genes, and by stimulating mineralization in immature mesenchymal cells and osteoblastic cells. The DNA-binding sites of Runx2 have been identified in major osteoblast-specific genes, including the genes that encode type I collagen (Col1a1), osteopontin, osteonectin, bone sialoprotein, osteocalcin, and Runx2 itself, and Runx2 induced the expression of these genes or activated their promoters *in vitro* (for review, see Komori[18]).

Runx2 also regulates the expression of Osx (encoded by the *OSX* or *SP7* gene), a specificity protein family transcription factor with three zinc-finger motifs. Osx is expressed in osteoblast progenitors, in osteoblasts, and at a lower level in (pre-)hypertrophic chondrocytes.[69] Similar to Runx2-deficient mice, mice lacking OSX showed complete lack of osteoblasts and absence of both intramembranous and endochondral bone formation.[69] Thus, Osx is a third transcription factor that is essential for osteoblast differentiation. Because Runx2 is expressed in the mesenchymal cells of OSX-null mice, but OSX is not expressed in RUNX2-null mice, it can be concluded that Osx acts downstream of Runx2.[69] Furthermore, the *OSX* gene contains a consensus Runx2-binding site in its promoter region, suggesting that it might be a direct target of Runx2. The transcriptional activity of Osx involves its interaction with NFATc1, cooperatively forming a complex that binds to DNA and induces the expression of *COL1A1*. Overall, it is currently thought that Runx2 has a crucial role in the earliest determination stage of the osteoblast lineage, driving mesenchymal progenitors to become preosteoblasts, whereas Osx regulates at a later stage the differentiation of preosteoblasts to functional, bone-forming osteoblasts expressing high levels of osteoblast markers (see Fig. 51.6).[18]

Other Factors and Signaling Molecules Involved in Osteoblastogenesis. Signaling pathways involving BMPs, Ihh, and FGFs have been mentioned previously in this chapter for their roles in chondrogenesis, but these signals also exert important functions in osteoblastogenesis and bone formation. Although several BMPs are extremely powerful in inducing bone formation when administered exogenously, their biological functions in osteoblast differentiation and endochondral ossification are less clear. Among BMP family members, BMP-2, BMP-4, and BMP-7 have been studied most extensively for their roles in osteogenesis during development, postnatal bone formation and remodeling, and fracture repair (reviewed in Lowery & Rosen[14] and Salazar et al.[70]). Double knockout of BMP-2 and BMP-4 in the limb completely disrupted osteoblast differentiation, demonstrating the crucial roles of these two BMPs in osteoblast differentiation. Effects of BMP signaling in later stages of osteoblast differentiation are suggested by studies using BMP antagonists. Targeting of Noggin overexpression to differentiated osteoblasts by use of the osteocalcin promoter in an animal model results in osteopenia by 8 months of age. Likewise, overexpression of Gremlin, another BMP antagonist, in differentiated osteoblasts in mice results in reduced bone mineral density and fractures. Moreover, deletion of *SMAD4*, which encodes a downstream effector of BMP signaling, in mature osteoblasts led to reduced osteoblast function. BMP-2 in particular plays important roles in osteogenic differentiation, postnatal bone formation, and fracture healing; the activity of BMP-2 as a potent inducer of bone formation is being applied as a therapeutic strategy to repair bone defects in humans.[14,70] Regarding the receptors, it is known that activating mutations in the BMP type I receptor ALK2 cause fibrodysplasia ossificans progressive, a human disease characterized by ectopic bone formation that recapitulates the entire sequence of endochondral ossification.[71] Still, much remains to be learned about the physiologically relevant BMP receptors in osteoblasts.

The Hedgehog family member Ihh is required for endochondral, but not intramembranous, bone formation in the skull, by controlling osteoblast differentiation in the perichondrium of long bones. IHH-null mice completely lack endochondral ossification because of the lack of osteoblasts.[17,72] In these mice, RUNX2 expression and nuclear β-catenin are absent in the perichondrium. Thus, Ihh is required for inducing the initial activity of Runx2 and β-catenin in perichondrial cells and triggering them to become endochondral osteoblasts, thereby coupling chondrocyte maturation with osteoblast differentiation during endochondral ossification (reviewed in Maes[21]). The role for Hedgehog signaling in osteoblast-lineage cells proper during adult bone homeostasis is less clear.

FGF signaling has been implicated in the proliferation of immature osteoblasts and the anabolic function of mature osteoblasts *in vivo*, in both the fetal and postnatal stages.[12] In mice lacking FGF18, osteoblast progenitors could be identified in the bone collar, but their terminal differentiation was delayed resulting in delayed ossification. Mice lacking FGF2 showed a reduction in total bone mass at the adult stage, which was probably owing to reduced osteoprogenitor proliferation and decreased osteoblast function.[12] Several FGF receptors (FGFR1, FGFR2, and FGFR3) have been implicated in the osteogenic functions of FGF signaling, as revealed by genetic studies in mice, although FGF signaling through different receptors appears to have diverse roles in regulating osteoprogenitor proliferation, osteoblast differentiation, and the activity of mature osteoblasts.[12]

KEY POINTS

- The formation of osteoblasts and their descendants (bone lining cells and osteocytes) requires the actions of transcription factors, most crucially Runx2, Osterix (Osx), and β-catenin/T cell factors/lymphoid enhancer factor. The expression of these and other transcription factors are regulated by locally produced Indian hedgehog (Ihh), Wnts, bone morphogenetic proteins (BMPs), fibroblast growth factors (FGFs), and other paracrine factors, which in turn are also regulated by inhibitors of these pathways.

Osteoclastogenesis and Skeletal Tissue Resorption

At the time of vascular invasion of hypertrophic cartilage, the chondrocytes are generally thought to mostly die through apoptosis, and remaining cellular debris and matrix components become degraded and are digested by coinvading osteoclasts (or related cells termed chondroclasts that have been postulated in the context of resorption of cartilage[30]) and endothelial cells through their proteolytic enzyme products.[30,73] Beyond development, the osteoclast remains a key participant in the regulation of bone mass, as bone is constantly being remodeled throughout life (see before and Fig. 51.3). Several pathologic conditions are because of an imbalance between bone formation and resorption, most frequently involving excessive osteoclast activity (see later). Such skeletal diseases include osteoporosis, a common low bone mass disorder typically prevalent in postmenopausal women, as well as periodontal disease, rheumatoid arthritis, multiple myeloma, and metastatic cancer. On the other hand, osteopetrosis, a rare human disease characterized by increased bone mass and obliteration of the bone marrow cavity, is caused by impaired osteoclast differentiation and/or function. Many efforts have been made to dissect the regulatory pathways leading to functional osteoclasts (Fig. 51.7).[74,75]

Osteoclastogenesis. Osteoclasts are giant multinucleated cells that have the unique capacity to efficiently degrade mineralized tissues. Osteoclasts are derived from hematopoietic precursor cells of the myelomonocytic lineage and share a common precursor with macrophages. Recognition of the intertwining control and close interactions of the immune and skeletal systems has led to the emergence of the integrating field of osteoimmunology.[76,77] Early differentiation of the bipotential macrophage/osteoclast precursor cells is regulated by PU.1, a versatile hematopoietic cell-specific transcriptional regulator belonging to the ETS family of transcription factors. The commitment to macrophages/osteoclasts is dependent on a high level of activity of PU.1 (Fig. 51.7); PU.1 regulates the lineage fate decision of early progenitors by directly controlling expression of the *C-FMS* gene, which is a key determinant of differentiation into the macrophage/osteoclast lineage; PU.1-null mice lack both macrophages and osteoclasts and are osteopetrotic. PU.1 also regulates the transcription of another key osteoclastogenesis control gene, encoding receptor activator of nuclear factor-κB (*RANK*), in myeloid progenitors. Although mature macrophages and osteoclasts have some cell surface markers in common, the latter express high levels of tartrate-resistant acid phosphatase (TRAP), cathepsin K, vitronectin, calcitonin receptor, and $\alpha_v\beta_3$ integrin, which typify the osteoclast lineage (also see later).[74,75]

After the precursor cells have committed to the osteoclast lineage, they are subjected to a complex multistep process that culminates in the generation of mature, multinuclear, activated osteoclasts; these steps include proliferation, maturation, and fusion of differentiating precursor cells, and finally activation of resorption (see Fig. 51.7). Two critical cytokines are essential for osteoclastogenesis: macrophage-colony stimulating factor (M-CSF) and RANK ligand (RANKL).[74-77] Both these signaling molecules are expressed by bone marrow stromal cells (i.e., osteoblast progenitors) and osteoblasts; M-CSF is produced in both a soluble and a membrane-bound form, and RANKL is made as a membrane protein by osteoblasts. Both PTH and $1,25(OH)_2D_3$, as well as several other osteotropic factors stimulating resorption, increase the expression of RANKL on stromal/osteoblastic cells (see later). RANKL, together with M-CSF, is sufficient to induce *in vitro* osteoclast differentiation from spleen- or bone marrow–derived precursors in the complete absence of stromal/osteoblastic cells.

Osteocytes have also been recognized as a major source of RANKL driving osteoclast formation and activation during adult bone remodeling in mice,[78,79] and may thus be crucial regulators of osteoclast formation (Fig. 51.7). Furthermore, not only osteoblast lineage cells but also chondrocytes and T cells synthesize and secrete RANKL and are able to support osteoclastogenesis. Production of RANKL by

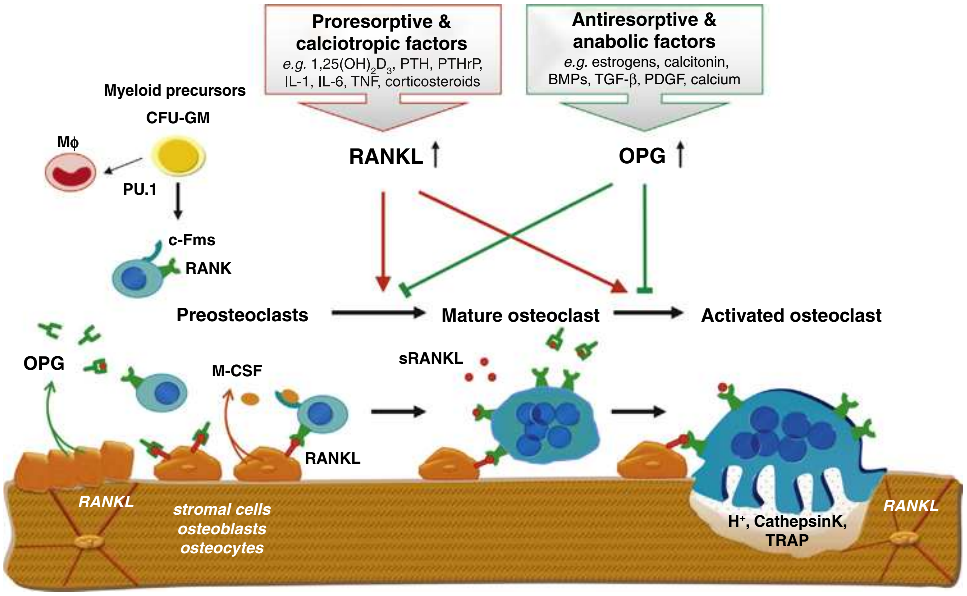

Proresorptive & calciotropic factors
e.g. 1,25(OH)₂D₃, PTH, PTHrP, IL-1, IL-6, TNF, corticosteroids
Antiresorptive & anabolic factors
e.g. estrogens, calcitonin, BMPs, TGF-β, PDGF, calcium
Myeloid precursors
CFU-GM
Mϕ
PU.1
c-Fms
RANK
RANKL ↑
OPG ↑
Preosteoclasts
Mature osteoclast
Activated osteoclast
OPG
M-CSF
sRANKL
RANKL
RANKL
stromal cells
osteoblasts
osteocytes
H⁺, CathepsinK, TRAP
RANKL

Fig. 51.7 Osteoclastogenesis. The transcription factor PU.1 is indispensable for development of the osteoclast and macrophage (*Mφ*) lineages from a common myeloid precursor. Macrophage-colony stimulating factor (*M-CSF*) acting upon the receptor c-Fms is required for the conversion of the progenitor to a preosteoclast. Signaling through the membrane-bound receptor activator of nuclear factor-κB (*RANK*) promotes further differentiation of preosteoclasts to mature polykaryotic osteoclasts and their subsequent activation to bone-resorbing cells that secrete protons and lytic enzymes. RANK signaling is induced by receptor activator of nuclear factor-κB ligand (*RANKL*), present on bone marrow stromal cells, osteoblasts, and osteocytes, as well as on lymphocytes and possibly in a soluble form in serum. The soluble antagonist osteoprotegerin (*OPG*) competes with RANK for RANKL binding, thereby functioning as a negative regulator of osteoclast differentiation, activation, and survival. RANKL expression is induced by proresorptive and calciotropic factors that stimulate osteoclastogenesis. Conversely, OPG is induced by factors that block bone catabolism and promote anabolic effects. *PTH,* parathyroid hormone; *PTHrP,* parathyroid hormone-related peptide; *IL,* interleukin; *TNF,* tumor necrosis factor; *BMPs,* bone morphogenetic proteins; *PDGF,* platelet-derived growth factor.

hypertrophic chondrocytes may be important physiologically for osteoclast differentiation and invasion of the hypertrophic cartilage during endochondral bone development and growth, and RANKL produced by T cells definitely plays a major role pathologically in numerous inflammatory conditions affecting the skeleton, including rheumatoid arthritis, periodontitis, and osteomyelitis, in which immune cells are activated in bones and joints. For instance, production of RANKL by T cells has been implicated as an activator of osteoclastic resorption in inflammation-mediated bone and cartilage destruction in rheumatoid arthritis, in which it likely works synergistically with tumor necrosis factor (TNF)-α and interleukin (IL)-17.[74-77]

M-CSF and RANKL affect several steps of the osteoclastogenesis cascade by binding to their respective receptors, C-FMS and RANK, which are expressed by osteoclast progenitors and osteoclasts at all stages of differentiation. M-CSF signaling is essential early on in the lineage for proliferation, differentiation, and survival of osteoclast/macrophage precursors. The essential roles of M-CSF were illustrated through analysis of mice carrying the *op/op* mutation, an inactivating point mutation in the *M-CSF* gene; these mice have osteopetrosis with low numbers of macrophages and a complete lack of mature osteoclasts.[74-77]

RANKL is an essential inducer of multiple aspects of osteoclastogenesis, including osteoclast differentiation, fusion, activation of mature osteoclasts to resorb mineralized bone, and survival.[74-77] Through these actions, RANKL potently stimulates bone resorption. RANKL acts by binding to its signal transducing receptor, RANK, a member of the TNF receptor family present on preosteoclasts and mature osteoclasts. Consistent with this, an activating mutation of the human *RANK* gene was found in patients with familial expansile osteolysis, a disease of excess bone resorption. Following the discovery of the RANK/RANKL system, numerous genetic and cell biological studies have been performed to elucidate the complex signaling cascade downstream of RANK/RANKL. Briefly stated and oversimplified, the binding of RANKL to RANK on the surface of osteoclast precursors recruits the adaptor protein/ubiquitin E3 ligase, TNF receptor-associated factor 6, to the cytoplasmic domain of RANK, leading to activation of NF-kB and its translocation to the nucleus. NF-kB increases C-FOS expression, and C-FOS, as a component of the activator protein complex, interacts with the master transcription factor for osteoclastogenesis, nuclear factor of activated T cells c1 (NFATc1), to induce osteoclast-specific genes (such as the genes encoding TRAP and calcitonin receptor). Costimulatory molecules act in concert with RANK to activate phospholipase C, thereby increasing intracellular calcium and activating NFATc1.[74]

The actions of RANKL are regulated negatively by a secreted soluble decoy receptor termed osteoprotegerin (OPG), also a member of the TNF receptor superfamily. OPG sequesters RANKL molecules, thereby blocking their binding to RANK. As such, OPG protects bone from excessive resorption. Correspondingly, certain homozygous deletions of *OPG* in humans can cause juvenile Paget disease, a disorder characterized by increased bone remodeling, osteopenia, and fractures. The relative concentrations of RANKL and OPG in bone thus are major determinants of bone mass and strength.[74-77]

Other regulatory molecules have been implicated in the late stages of osteoclast differentiation, the fusion process, and activation of resorption. The fusion of mononuclear osteoclast precursor cells into mature multinucleated osteoclasts is regulated by a membrane protein called dendritic cell-specific transmembrane protein (DC-STAMP). DC-STAMP–deficient cells failed to fuse, and these mononuclear osteoclasts had reduced resorptive efficiency in vitro. Consequently, DC-STAMP–deficient mice exhibited increased bone mass. These data suggest that multinucleation and osteoclast enlargement could be associated with a higher degree of resorption efficiency.[74-77] Pathologically huge osteoclasts containing substantially more nuclei than normal osteoclasts do are seen, for instance, in Paget disease of bone (up to 100 nuclei per cell; normal range is 3–20 nuclei per cell), in which localized excessive bone remodeling is initiated by increases in osteoclast-mediated bone resorption.[80]

In summary, RANKL and OPG act in an antagonistic fashion to regulate bone resorption, and their respective expression levels are under the control of proresorptive and antiresorptive factors, including several hormones, cytokines, and growth factors (see Fig. 51.7, top). The discovery of the RANK/RANKL/OPG signaling pathway has led to the development of drugs (e.g., denosumab, an anti-RANKL antibody) to suppress excessive osteoclast formation for clinical use in a variety of bone diseases, including osteoporosis, autoimmune arthritis, periodontitis, Paget disease, and bone tumors/metastases.[74,77]

Bone Resorption: Osteoclast Action and Proteolytic Enzymes.

Bone resorption involves both dissolution of bone mineral and degradation of organic bone matrix. Osteoclasts are highly specialized to perform both of these functions. Upon activation of mature multinucleated osteoclasts, the cells attach themselves firmly to the bone surface, using specialized actin-rich podosomes (actin ring), through cytoskeleton reorganization and cellular polarization.[81] Within these tightly sealed zones of adhesion to the mineralized matrix, the osteoclasts form convoluted, villus-like membranes called “ruffled borders,” which substantially increase the surface area of the cell membrane facing the resorption lacuna (Howship’s lacuna). Via these ruffled membranes, the osteoclasts secrete abundant hydrochloric acid (involving the vacuolar H^+-ATPase proton pump), mediating acidification of the compartment between the cell and the bone surface, as well as a myriad of enzymes such as lysosomal cathepsins, the phosphatase TRAP, and proteolytic MMPs (see later). The acidity of the environment leads to dissolution of the mineral phase (crystalline hydroxyapatite), activation of lytic enzymes, and digestion of organic matrix compounds (see Fig. 51.8). The sealing mechanism allows localized dissolving and degrading of the mineralized bone matrix, while simultaneously protecting neighboring cells from harm.[82] During the resorption process, dissolution of hydroxyapatite releases large amounts of soluble calcium, phosphate, and bicarbonate. Removal of these ions is needed (e.g., to maintain the acidic pH in the resorption lacuna) and involves vesicular pathways and direct ion transport via different ion exchangers, channels, and pumps. The degradation products of the organic matrix after enzymatic digestion are transcytosed through the cell for secretion at the basolateral membrane.[82]

These complex processes of osteoclast recruitment, polarization on the bone surface, and export of acid and enzymes are orchestrated by many factors, including RANKL, as well as by integrin-mediated signaling from the bone matrix itself.[81] The latter, which is particularly represented by the αvβ3 integrin in osteoclasts, was suggested to be important for osteoclast functioning based on the finding that inhibition of signaling through this αvβ3 integrin inhibited osteoclast-mediated bone resorption *in vitro* and in animal models of osteoporosis and malignant osteolysis. Integrins are heterodimeric cell-surface receptors, composed of an α and a β subunit, that mediate cell-matrix interactions and thus adhesion. The αvβ3 integrin, which is the most highly expressed integrin in osteoclasts, recognizes Arg-Gly-Asp-containing matrix proteins such as vitronectin, osteopontin, and bone sialoprotein. Several components of the αvβ3 integrin signaling pathway localize to the sealing zone of actively resorbing osteoclasts and play a role in linking the matrix adhesion of osteoclasts to cytoskeletal organization, cell polarization, and activation for bone resorption. Upon its activation, αvβ3 integrin stimulates an intracellular signaling complex involving the tyrosine kinases C-SRC and SYK. The importance of αvβ3, C-SRC and SYK in osteoclast activity is underscored by the development of osteopetrosis in mice deficient in each of these genes, because of lack of bone resorption.[81,82]

Many of the molecules that are important for osteoclast function *in vitro*, such as β3 integrin, C-SRC, cathepsin K, carbonic anhydrase II, TRAP, and several ion channel proteins, cause an osteopetrotic phenotype when deleted in mice or altered in humans. The absence of these genes does not affect the differentiation into morphologically normal osteoclasts; however, the osteoclasts are not functional, and they fail to resorb bone effectively.[83] For instance, cathepsin K, which is the key enzyme in the digestion of bone matrix owing to its activity in degrading type I collagen, is highly expressed by activated osteoclasts and secreted in the resorption lacuna. Its deletion in mice led to osteopetrosis, and mutations in the human cathepsin K gene cause pycnodysostosis. Highly selective and potent cathepsin K inhibitors have been developed and tested in clinical trials to assess their usefulness as antiresorptive agents for the treatment of osteoporosis and to reduce breast cancer-induced osteolysis and skeletal tumor burden, but adverse side effects and roles of cathepsin K beyond bone complicate the drug development.

Besides cathepsin K, several proteolytic enzyme groups are involved in the degradation of organic components (collagens and proteoglycans) of bone and cartilage matrices after the mineral is dissolved. One of these is the MMP family, which contains over 25 members, including secreted collagenases, stromelysins, gelatinases, and membrane-type (MT)-MMPs. Several MMPs, including MMP9 and MMP14 (also known as MT1-MMP), are highly expressed in osteoclasts/chondroclasts, but they are produced by many other cell types as well. Both of these molecules play a role in the cartilage resorption process associated with invasion by blood vessels and osteoclasts during endochondral ossification.[30,73,84]

KEY POINTS

- Osteoclasts are multinucleated cells with a hematopoietic origin shared by macrophages. They play crucial roles in resorbing both the mineral phase and matrix of bone by secreting acid and multiple proteolytic enzymes during development and bone remodeling. A complex series of receptors and transcription factors guide the differentiation of osteoclasts in response to macrophage-colony stimulating factor (M-CSF) and receptor activator of nuclear factor-κB ligand (RANKL), which are made by multiple cell subsets of the osteoblast lineage. In this way, the osteoblast lineage controls the activities of osteoclasts and the sites at which they act.

REGULATION OF BONE REMODELING AND BONE MASS

Coupling of Bone Resorption and Bone Formation During Remodeling

Skeletal homeostasis remains intact as long as the activities of both osteoclasts and osteoblasts are balanced (coupled), and the net bone mass is maintained (also see Fig. 51.3). This balance implies the existence of mechanisms that tightly coordinate the differentiation of osteoblasts and osteoclasts, as well as their migration to locations where they function. Increasing data support the model that osteoclasts and osteoblasts reciprocally regulate each other's recruitment, differentiation, and/or activity. In addition, osteocytes increasingly emerged as key players in the remodeling process by regulating both osteoclastic bone resorption and osteoblastic bone formation.

A first aspect of the coupling principle is provided by the direct control of osteoclastogenesis by cells of the osteoblast lineage. Indeed, as outlined above, osteoblasts and stromal cells control bone degradation by expressing M-CSF, required for the proliferation of osteoclast precursors, and RANKL, mediating the differentiation of hematopoietic osteoclast precursors toward mature multinucleated osteoclasts. Osteoblasts also express OPG, the soluble decoy receptor that binds to RANKL and prevents it from binding to RANK, thus providing a negative control mechanism to limit osteoclastogenesis (see Fig. 51.7).[74-76] Osteocytes, as sources of RANKL, are additional players in the regulation of osteoclast formation and activation during bone remodeling.[78,79]

Other mechanisms by which osteoblast lineage cells control osteoclastogenesis *in vivo* are being increasingly uncovered through the use of cell type and/or stage-specific mutagenesis in mice (for recent reviews, see Sims & Martin[85]and Kim et al.[86]). For instance, the noncanonical Wnt pathway (Wnt signaling relayed to intracellular responses by mechanisms that are independent of β-catenin actions) between osteoblast-lineage cells and osteoclast precursors appears to regulate bone resorption.[87] Another newly identified mechanism involves osteoblastic expression of semaphorin 3A (SEMA3A), a member of the family of secreted and membrane-associated axon guidance molecules that use plexins and neuropilins (NRPs) as their primary receptors; soluble SEMA3A produced by osteoblasts was found to exert an osteoprotective effect by inhibiting osteoclast precursor differentiation through the Plexin-A coreceptor NRP1.[88] In contrast, SEMA6D signaling through the receptor Plexin-A on osteoclasts was found to promote osteoclast formation and function.[89] Hence, this system's complexity, with multiple interacting ligands and receptors, calls for further studies to clarify their precise roles in osteoblast-osteoclast communication (also see following).

The second, converse aspect of the coupling is provided by the reciprocal activity of osteoclasts to modulate osteoblast recruitment, differentiation, and function to regulate the anabolic arm of the remodeling process that follows the resorption phase (see Fig. 51.3). In the coupling of osteoclasts to cells of the osteoblast lineage, three types of osteoblastogenic signals can be discriminated: (1) bone resorption–derived signals, most typically comprising growth factors that were embedded in bone matrix and become released upon osteoclastic bone resorption (e.g., TGFβ, BMPs, IGFs, PDGF); (2) osteoclast-derived and other soluble signals (e.g., hormones and diffusible growth factors, such as BMP-6, S1P, and Wnt10b); and (3) cell-surface interactions between osteoclasts and developing osteoblasts, mediated by membrane-bound molecules (e.g., ephrins, semaphorins).[85,86,90] These three classes of coupling mediators are discussed next.

Osteoclastic bone resorption is generally assumed to locally release a myriad of growth factors that are stored in the bone matrix, which can subsequently act as potent osteoblast stimuli by attracting osteoblastic cells or their progenitors to the remodeling site and/or stimulating their proliferation, differentiation and activity.[28] Growth factors like TGFβ, BMPs, IGFs, VEGF, and PDGF produced by osteoblasts are incorporated into the bone matrix in high concentrations; when released, they can elicit their anabolic effects on osteoblast lineage cells. One key aspect about which many questions remain is how osteoblast progenitors are initially recruited to the remodeling site. Constituents of the bone matrix, such as fragments of collagen or osteocalcin, complement fragments, inflammatory cytokines, and growth factors that are typically released, activated, or produced at sites of matrix resorption have all been reported to exert chemotactic potential toward osteoblasts and their precursors *in vitro* (reviewed Dirckx et al.[28]). For instance, release of TGFβ from the bone matrix during osteoclastic bone matrix degeneration was reported to exert a chemotactic effect on osteoblast precursors to these bone resorptive sites.[91] Similarly, IGF-1 released from the bone matrix upon resorption was found to promote osteoblastic differentiation of recruited SSCs by activation of mammalian target of rapamycin, thereby maintaining proper bone mass.[92]

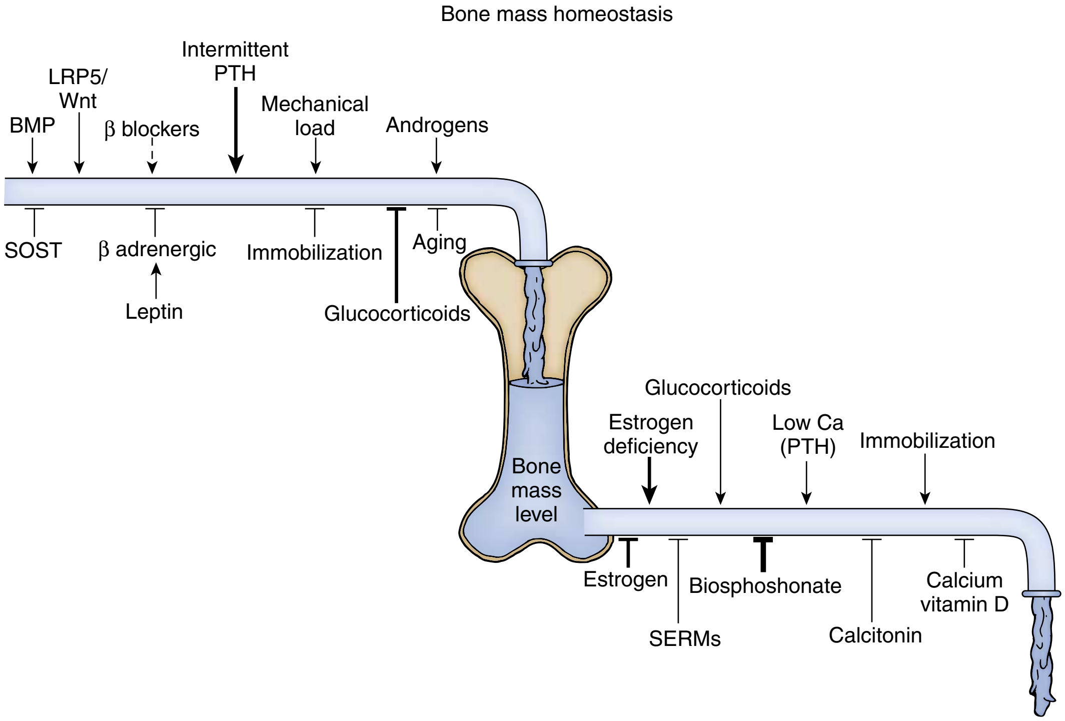


Fig. 51.8 **Control of Skeletal Homeostasis and Bone Mass.** Schematic representation of the "servo" system that maintains bone mass at steady-state levels. Physiologic and pharmacologic stimulators and inhibitors of bone formation and resorption are listed. The relative impact, where known, is represented by the thickness of the arrows. *SERM*, Selective estrogen receptor modulator; *BMP,* bone morphogenetic protein; *SOST,* sclerostin; *PTH,* parathyroid hormone; *LRP,* lipoprotein receptor–related protein; *Ca,* calcium, (Modified from Harada S, Rodan GA. Control of osteoblast function and regulation of bone mass. *Nature*. 2003;423:349–355.)

In addition to chemotactic signals released from the matrix upon resorption, osteoclasts can also play a direct role in regulating osteoblast recruitment. For instance, osteoclasts themselves secrete factors with chemotactic properties, such as PDGF, that can stimulate osteogenesis and angiogenesis in parallel.[28] Other molecules with anabolic potential that are expressed by osteoclasts include BMP-6, S1P, the Wnt agonist Wnt10b, and the adiponectin family member collagen triple helix repeat containing 1 (CTHRC).[85,86,90] When CTHRC was deleted only from osteoclasts in mice, osteopenia resulted; thus, this protein is likely to have a physiological role in the coupling mechanism.[93] Of note, osteocytes also play a role in determining osteoblastic differentiation and bone formation during remodeling by modulating Wnt signaling, for instance through their expression of the Wnt inhibitors DKK1 and sclerostin.

Lastly, osteoclasts can also signal directly to cells of the osteoblast lineage through cell-bound molecules.[85,86,90] For instance, the ephrin/Eph receptor system may allow bidirectional cell-surface contact based signaling between osteoclasts and osteoblasts. It was reported that osteoclasts express ephrin B2, whereas osteoblasts express its EphB4 receptor, both membrane-bound proteins. Signaling through EphB4 into osteoblasts (forward signaling) appeared to enhance osteogenic differentiation, whereas signaling through ephrin B2 into osteoclast precursors (reverse signaling) suppressed osteoclast differentiation, the overall outcome of such interaction thus being predicted to favor bone formation. Other studies have argued that ephrin B2 is produced by the osteoblast lineage, thus contesting the view of an osteoclast-derived coupling activity.[90] Another study showed that osteoclast-produced SEMA4D, a predominantly transmembrane semaphorin, contributes to balanced remodeling in adult bone by binding to its receptor Plexin-B1, expressed on osteoblasts. Altogether, although many details remain to be uncovered, it is clear that osteoclasts are capable of producing many molecules affecting the osteoblast lineage, both positively and negatively.[85,86,90]

KEY POINTS

- Osteoblasts and osteoclasts use multiple mechanisms to communicate with each other. Local factors produced by cells of the osteoblast lineage, such as semaphorins and noncanonical Wnts, complement macrophage-colony stimulating factor (M-CSF) and receptor activator of nuclear factor-κB ligand (RANKL) to regulate osteoclastogenesis and the activities of mature osteoclasts. Osteoclasts use several distinct types of mechanisms to influence osteoblastic cells. Osteoclastic bone resorption leads to release of cytokines and growth factors from the bone matrix that can attract osteoblast precursors and accelerate their differentiation and activities. Osteoclasts themselves also secrete factors that act on osteoblast lineage cells, and they express cell surface proteins that signal directly to osteoblast lineage cells through cell–cell contact.

Local, Systemic, and Central Regulators of Bone Remodeling

Bone remodeling and the resultant net bone mass are influenced by a myriad of signals (Fig. 51.8). In addition to the many local factors that were discussed previously, bone resorption and bone formation are also profoundly affected by mechanical stimuli, endocrine molecules, and central regulatory systems, that consequently impact strongly on the balance between the two processes, and hence on bone health.

Mechanical Loading. The skeletal system provides mechanical support and protection for the multiple organs of vertebrate organisms. To withstand loading in the most efficient way (maximal strength for minimal material), the skeleton constantly adjusts its bone mass and architecture in response to load through bone remodeling. In fact, adaptation and reshaping of the bone structure to mechanical forces require the replacement of existing bone packets or whole trabeculae with new ones appropriately oriented relative to prevailing mechanical loads. The close, dynamic relation between bone structure and the lines of force in bone may be maintained by responses to a mechanical strain threshold (force per unit area), below which osteoclastic removal of bone occurs, while a higher strain will stimulate the addition of bone. It has indeed long been recognized that mechanical stress induced by weight-bearing exercise increases osteoblast activity and induces bone formation. Conversely, the absence of mechanical stimulation resulting from prolonged immobilization or unloading causes severe bone loss. During immobilization, bone resorption increases and formation decreases, contributing to the bone loss seen to be associated with frailty and prolonged bed rest. Similarly, prolonged space flight, with the resulting lack of gravity forces on the skeleton, leads to marked loss of bone mass and increased bone fragility in astronauts.

Increasing evidence suggests that the main mechanosensing cell type of bone is the osteocyte (see earlier). The osteocyte is ideally situated within the lamellar bone to sense mechanical forces and communicates signals to other osteocytes, osteoblast lineage cells, and osteoclasts on the bone surface, likely through the connecting canalicular network.[54-56] Osteocytes also respond to increased load by altering their signaling, resulting in the translation of mechanical stimuli into biochemical signals that generally enhance osteoblastic bone formation and reduce osteoclastic bone resorption. For instance, osteocytes take part in the control of bone formation through the production of sclerostin, the protein derivative of the *SOST* gene that inhibits Wnt signaling through binding to LRP5/6 (see earlier); mechanical loading results in a rapid decrease in the *SOST* expression levels in osteocytes, suggesting that osteocytes transmit signals of mechanical loading to osteoblast lineage cells on the bone surface by modulating the Wnt/β-catenin pathway.[54-56] As also discussed earlier, osteocytes have the potential to stimulate bone resorption by producing RANKL, and the levels of RANKL versus OPG could be regulated by mechanical stimulation. Mechanical stimulation is also capable of maintaining osteocyte viability, which seems important in the response to mechanical signals. Experimental destruction of osteocytes in a genetic mouse model conferred resistance to unloading-induced bone loss.[94] The precise mechanisms by which a mechanical stimulus is transduced into biochemical signals in osteocytes and osteoblasts and the means whereby these cells then modulate bone remodeling activities have not been clearly identified yet.[54-56]

Parathyroid Hormone. Bone remodeling also represents a major mechanism used by land vertebrates to regulate mineral homeostasis. The removal of calcium from bone, to maintain calcium homeostasis acutely, is facilitated by osteoclastic resorption of calcium-containing bone. The principal mediators in this process are PTH and its downstream effector $1,25(OH)_2D_3$, the activities of which are discussed in detail in other chapters in this book. Both these hormones act on osteoblastic cells to increase the production of RANKL and thereby potently activate osteoclastogenesis.[95]

The actions of PTH on bone are complex, with multiple direct actions on cells of the osteoblast lineage that vary both with the state of differentiation of the cell and with its location in bone (periosteal vs. endosteal). The actions of PTH to increase osteoclast number and action dominate when levels of PTH are continuously elevated, but PTH also increases the rate of bone formation. In fact, when PTH is administered intermittently, this effect predominates and leads to a net increase in bone mass (see Fig. 51.8).[96] PTH action directly on osteocytes may be involved in the control of both bone resorption and bone formation. PTH and, during lactation, PTHrP can also stimulate osteocytes to directly dissolve bone matrix and release calcium into the bloodstream (osteocytic osteolysis).[54,55,97] How much osteoclastic osteolysis and osteocytic osteolysis each contribute to release of mineral from bone is an area of active investigation.

When administered intermittently, PTH enhances RUNX2 expression and activity; however, RUNX2 is also involved in the catabolic effect of PTH through the induction of RANKL.[95,98] Intermittent PTH promotes osteoblast differentiation also in part by its ability to promote exit from the cell cycle, to activate Wnt signaling in osteoblasts, and to inhibit production of the Wnt antagonist sclerostin in osteocytes. Furthermore, intermittent administration of PTH prolongs osteoblast survival, and increases osteoblast number by converting inactive lining cells to mature osteoblasts *in vivo*.[45] By controlling osteoblast function, PTH increases bone formation on cancellous, endocortical, and periosteal bone surfaces and has become a valuable clinical anabolic treatment strategy for osteoporotic conditions.[95,98]

Sex Steroids. Another significant endocrine input into bone remodeling and homeostasis is the pronounced influence of sex steroids, essential for maintenance of the adult skeleton (see Fig. 51.8; reviewed in Khosla et al.[99] and Manolagas et al.[100]) Skeletal preservation by estrogen in females may be related evolutionarily to the need for calcium stores for embryonic skeletal development. In mammalian adult males and females, including humans, estrogen has been identified as the major inhibitor of bone resorption that functions by reducing osteoclast number. This explains the profound bone loss associated with estrogen deficiency, with postmenopausal osteoporosis being the most prominent type of bone pathology in humans.

Estrogen actions on osteoclasts, which express estrogen receptors (ERα and ERβ), involve direct effects on their lifespan, through estrogen-mediated induction of the Fas/FasL system, causing apoptosis.[101] Furthermore, estrogen can have nongenomic effects, or rapid signaling effects, thus inducing the phosphorylation of components of various signaling pathways (e.g., the MAPK pathway) or calcium regulation. Such effects may contribute to the induction of osteoclast apoptosis by estrogen without the need for direct binding of ERα to DNA.[102] Besides affecting osteoclast lifespan, estrogen was reported to decrease the responsiveness of osteoclast progenitor cells to RANKL, thereby reducing osteoclastogenesis.[99,100]

Androgens are important not only as a source of estrogen, through the action of aromatase, but also for their direct effect in stimulating bone formation.[103] Testosterone, responsible for the male phenotype characterized by a larger skeleton, has complex effects on bone metabolism. As an anabolic steroid, it stimulates bone formation in both males and females. Androgens also increase periosteal bone formation, leading to larger and therefore stronger bones. Mice lacking the androgen receptor in osteoblasts and/or osteocytes displayed a phenotype characterized by reduced trabecular bone volume and accelerated age-related bone loss.[103] In addition, it is thought that testosterone can inhibit bone resorption directly, acting through the androgen receptor and through conversion by aromatase to estrogens.[103]

Other Systemic and Central Regulators of Bone Remodeling. In contrast to sex steroids, glucocorticoid excess is catabolic to bone, as is illustrated by glucocorticoid-induced osteoporosis, a devastating consequence of long-term use of glucocorticoids.[104] The mechanisms for glucocorticoid-induced bone loss are complex, including suppression of renal calcium reabsorption, reduction in intestinal calcium absorption, and hypogonadism, all of which lead to increased bone resorption and bone turnover. Perhaps more importantly, glucocorticoids have direct effects on the skeleton. Glucocorticoids impair the replication, differentiation, and function of osteoblasts and induce the apoptosis of mature osteoblasts and osteocytes, leading to dramatic suppression of bone formation. As well, glucocorticoids act directly on osteoclasts and prolong their lifespan; in addition, glucocorticoids may sensitize bone cells to regulators of bone remodeling and favor osteoclastogenesis, thus leading to overall increased bone resorption.[104] Influences on bone remodeling by the GH/IGF axis, thyroid hormones, and prostaglandins are reviewed in other chapters.

Signals from the central nervous system (CNS) and the sympathetic nervous system (SNS) also profoundly influence bone remodeling and resultant bone mass.[105,106] Leptin, an adipocyte-derived hormone mainly known for controlling appetite, energy metabolism, and reproduction, has been shown to act through the CNS (via the hypothalamus) and ultimately through the SNS to negatively regulate bone formation by osteoblasts. Sympathetic signaling via β_2-adrenergic receptors in osteoblasts inhibits proliferation through transcriptional factors such as Period, which also regulate the circadian clock, and additionally favors bone resorption by increasing osteoblastic expression of RANKL. These conclusions were drawn from a series of studies using mutant mice with deficiencies or lesions in portions of the leptin-hypothalamic-sympathetic pathways.[105] Furthermore, other neuroendocrine connections to bone remodeling have been made with the findings in mouse genetic and pharmacologic studies that both thyroid-stimulating hormone (TSH) and follicle-stimulating hormone (FSH) act directly on bone cells to influence remodeling.

Besides the mechanism of action of leptin through a central pathway, leptin appears to also act via peripheral pathways on osteoblasts to regulate bone mass and strength. The leptin receptor is present in bone marrow–resident stromal cells representing osteolineage progenitors (see earlier),[107] and the actions of leptin may be partly explained by direct effects on these cells.

Interestingly, whereas leptin represents an adipocyte-derived hormone regulating bone remodeling, through downstream effectors in bone that seem to be largely independent of its functions in energy metabolism and reproduction, a reciprocal regulatory function on energy metabolism and adiposity is exerted by osteoblasts. For instance, osteocalcin, one of the few osteoblast-specific gene products, is thought to act as a bone-derived signal modulating energy metabolism and adiposity. In its undercarboxylated form, which has a reduced affinity for the bone mineral, osteocalcin is released into the circulation and has been reported to function as an active hormone to increase pancreatic β cell proliferation, insulin expression and secretion, and insulin-sensitive glucose utilization and energy expenditure, and to reduce visceral fat.[108] In addition, recent studies using knockout mice indicated roles for osteocalcin in the regulation of brain development and adult brain function (including learning, memory, and emotional states), muscle mass, testosterone synthesis, and male fertility, as well as the acute response to stress, indicating that the endocrine actions of bone-derived osteocalcin are not limited to regulating energy metabolism.[109,110] On the other hand, recent studies of independently generated knockout rats and mice did not confirm the changes in insulin secretion, adiposity, and muscle mass noted in the initial osteocalcin knockout mice.[108,111-114] Hence, the consistency in the endocrine functions of osteocalcin among various species and genetic backgrounds and the full significance of these findings in humans have yet to be fully defined.

From many studies it is clear that the skeleton plays active roles in integrating organs systems and global functions within the organism, including in the control of energy metabolism. Besides osteocalcin, several other local regulators and pathways in bone, and in osteolineage cells in particular, have been linked to the regulation of systemic energy metabolism and glucose homeostasis, including the hypoxia-inducible pathway controlling osteolineage cell metabolism and skeletal glucose consumption, Wnt/β-catenin signaling, PPAR-γ activity in the osteolineage, and lipocalin and sclerostin production.[115,116]

KEY POINTS

- Mechanical forces, hormones, and the central nervous system complement the local factors discussed earlier to regulate bone formation and resorption. Receptors important for sensing mechanical forces are under active study. Osteocytes are the best studied cells that sense mechanical forces and signal to suppress sclerostin expression in response to these forces. Parthyroid hormone (PTH) increases both bone formation and resorption, whereas estrogens, glucocorticoids, and other steroid/thyroid hormones target multiple bone cell types to regulate bone turnover. The sympathetic nervous system, in response to multiple signals, can dramatically block osteoblast action. In turn, bone cells secrete a number of hormones that can affect mineral metabolism and other important physiological processes.

CONCLUSION

Bone formation during skeletal development involves complex spatiotemporal coordination among multiple molecules, cell types, and tissues. A large number of morphogens, signaling molecules, and transcriptional regulators have been implicated in the regulation of bone development. Many of the developmental mechanisms are recapitulated or remain important later in bone homeostasis and/or pathology. The skeleton is a vital organ system essential to survival and well-being. Its major functions, including mechanical support and mineral homeostasis, are achieved by continuous remodeling throughout life. Osteoclasts and osteoblasts normally act in a coordinated manner to resorb and rebuild bone. Most bone disorders are caused by bone loss resulting from an imbalance in the remodeling cycle, with bone resorption exceeding bone formation. A better understanding of the molecular mechanisms behind skeletal biology, largely derived from discoveries based on mouse and human genetics, has provided opportunities for developing therapies to treat diseases of bone. Additional advances in the molecular and genetic studies of bone, including integrative communication between pathways and cell types within and beyond the skeleton proper, may offer as yet unexplored targets for developing novel therapeutic options to treat various bone pathologies.

REFERENCES

1. Helms JA, Schneider RA. Cranial skeletal biology. *Nature*. 2003;423:326–331.
2. Tabin CJ, McMahon AP. Developmental biology. Grasping limb patterning. *Science*. 2008;321:350–352.
3. Maes C, Kobayashi T, Selig MK, et al. Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Dev Cell*. 2010;19:329–344.
4. Yang L, Tsang KY, Tang HC, et al. Hypertrophic chondrocytes can become osteoblasts and osteocytes in endochondral bone formation. *P Natl Acad Sci USA*. 2014;111:12097–12102.
5. Zhou X, von der Mark K, Henry S, et al. Chondrocytes transdifferentiate into osteoblasts in endochondral bone during development, postnatal growth and fracture healing in mice. *Plos Genetics*. 2014;10:e1004820.
6. Gerber HP, Vu TH, Ryan AM, et al. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *NatMed*. 1999;5:623–628.
7. Ishii M, Egen JG, Klauschen F, et al. Sphingosine-1-phosphate mobilizes osteoclast precursors and regulates bone homeostasis. *Nature*. 2009;458:524–528.

8. Ishii M, Kikuta J, Shimazu Y, et al. Chemorepulsion by blood S1P regulates osteoclast precursor mobilization and bone remodeling in vivo. *J Exp Med*. 2010;207:2793–2798.
9. Oichi T, Otsuru S, Usami Y, et al. Wnt signaling in chondroprogenitors during long bone development and growth. *Bone*. 2020;137:115368.
10. Lefebvre V. Roles and regulation of SOX transcription factors in skeletogenesis. *Curr Top Dev Biol*. 2019;133:171–193.
11. Ornitz DM, Legeai-Mallet L. Achondroplasia: development, pathogenesis, and therapy. *Dev Dyn*. 2017;246:291–309.
12. Ornitz DM, Marie PJ. Fibroblast growth factors in skeletal development. *Curr Top Dev Biol*. 2019;133:195–234.
13. Kronenberg HM. Developmental regulation of the growth plate. *Nature*. 2003;423:332–336.
14. Lowery JW, Rosen V. The BMP pathway and its inhibitors in the skeleton. *Physiol Rev*. 2018;98:2431–2452.
15. Lanske B, Karaplis AC, Lee K, et al. PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science*. 1996;273:663–666.
16. Vortkamp A, Lee K, Lanske B, et al. Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science*. 1996;273:613–622.
17. St-Jacques B, Hammerschmidt M, McMahon AP. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev*. 1999;13:2072–2086.
18. Komori T. Molecular mechanism of Runx2-dependent bone development. *Mol Cells*. 2020;43:168–175.
19. Maeda Y, Nakamura E, Nguyen MT, et al. Indian Hedgehog produced by postnatal chondrocytes is essential for maintaining a growth plate and trabecular bone. *Proc Natl Acad Sci U S A*. 2007;104:6382–6387.
20. Hirai T, Chagin AS, Kobayashi T, et al. Parathyroid hormone/parathyroid hormone-related protein receptor signaling is required for maintenance of the growth plate in postnatal life. *Proc Natl Acad Sci U S A*. 2011;108:191–196.
21. Maes C. Signaling pathways effecting crosstalk between cartilage and adjacent tissues: Seminars in cell and developmental biology: the biology and pathology of cartilage. *Semin Cell Dev Biol*. 2017;62:16–33.
22. Maes C, Carmeliet G, Schipani E. Hypoxia-driven pathways in bone development, regeneration and disease. *Nat Rev Rheumatol*. 2012;8:358–366.
23. Semenza GL. Hypoxia-inducible factors in physiology and medicine. *Cell*. 2012;148:399–408.
24. Schipani E, Ryan HE, Didrickson S, et al. Hypoxia in cartilage: HIF-1alpha is essential for chondrocyte growth arrest and survival. *Genes Dev*. 2001;15:2865–2876.
25. Stegen S, Laperre K, Eelen G, et al. HIF-1alpha metabolically controls collagen synthesis and modification in chondrocytes. *Nature*. 2019;565:511–515.
26. Maes C. Role and regulation of vascularization processes in endochondral bones. *Calcified Tissue Int*. 2013;92:307–323.
27. Liu Y, Berendsen AD, Jia S, et al. Intracellular VEGF regulates the balance between osteoblast and adipocyte differentiation. *J Clin Invest*. 2012;122:3101–3113.
28. Dirckx N, Van Hul M, Maes C. Osteoblast recruitment to sites of bone formation in skeletal development, homeostasis, and regeneration. *Birth Defects Res C Embryo Today*. 2013;99:170–191.
29. Clarkin CE, Gerstenfeld LC. VEGF and bone cell signalling: an essential vessel for communication? *Cell Biochem Fun*. 2013;31:1–11.
30. Vu TH, Shipley JM, Bergers G, et al. MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes. *Cell*. 1998;93:411–422.
31. Schipani E, Maes C, Carmeliet G, et al. Regulation of osteogenesis-angiogenesis coupling by HIFs and VEGF. *J Bone Miner Res*. 2009;24:1347–1353.
32. Kusumbe AP, Ramasamy SK, Adams RH. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature*. 2014;507:323–328.
33. Ramasamy SK, Kusumbe AP, Itkin T, et al. Regulation of hematopoiesis and osteogenesis by blood vessel-derived signals. *Annu Rev Cell Dev Biol*. 2016;32:649–675.
34. Sacchetti B, Funari A, Michienzi S, et al. Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell*. 2007;131:324–336.
35. Morikawa S, Mabuchi Y, Kubota Y, et al. Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. *J Exp Med*. 2009;206:2483–2496.
36. Greenbaum A, Hsu YM, Day RB, et al. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature*. 2013;495:227–230.
37. Chan CK, Seo EY, Chen JY, et al. Identification and specification of the mouse skeletal stem cell. *Cell*. 2015;160:285–298.
38. Worthley DL, Churchill M, Compton JT, et al. Gremlin 1 identifies a skeletal stem cell with bone, cartilage, and reticular stromal potential. *Cell*. 2015;160:269–284.
39. Chan CKF, Gulati GS, Sinha R, et al. Identification of the human skeletal stem cell. *Cell*. 2018;175:43–56 e21.
40. Mizuhashi K, Ono W, Matsushita Y, et al. Resting zone of the growth plate houses a unique class of skeletal stem cells. *Nature*. 2018;563:254–258.
41. Newton PT, Li L, Zhou B, et al. A radical switch in clonality reveals a stem cell niche in the epiphyseal growth plate. *Nature*. 2019;567:234–238.
42. Bianco P, Robey PG. Skeletal stem cells. *Development*. 2015;142:1023–1027.
43. Bohm AM, Dirckx N, Tower RJ, et al. Activation of skeletal stem and progenitor cells for bone regeneration is driven by PDGFRbeta signaling. *Dev Cell*. 2019;51:236–254.e12.
44. Kim SW, Lu Y, Williams EA, et al. Sclerostin antibody administration converts bone lining cells into active osteoblasts. *J Bone Miner Res*. 2017;32:892–901.
45. Kim SW, Pajevic PD, Selig M, et al. Intermittent parathyroid hormone administration converts quiescent lining cells to active osteoblasts. *J Bone Miner Res*. 2012;27:2075–2084.
46. Matic I, Matthews BG, Wang X, et al. Quiescent bone lining cells are a major source of osteoblasts during adulthood. *Stem Cells*. 2016;34:2930–2942.
47. Park J, Gebhardt M, Golovchenko S, et al. Dual pathways to endochondral osteoblasts: a novel chondrocyte-derived osteoprogenitor cell identified in hypertrophic cartilage. *Biol Open*. 2015;4:608–621.
48. Pinho S, Frenette PS. Haematopoietic stem cell activity and interactions with the niche. *Nat Rev*. 2019;20:303–320.
49. Crane GM, Jeffery E, Morrison SJ. Adult haematopoietic stem cell niches. *Nat Rev*. 2017;17:573–590.
50. Sebo ZL, Rendina-Ruedy E, Ables GP, et al. Bone marrow adiposity: basic and clinical implications. *Endocr Rev*. 2019;40:1187–1206.
51. Scheller EL, Cawthorn WP, Burr AA, et al. Marrow adipose tissue: trimming the fat. *Trends Endocrinol Metab*. 2016;27:392–403.
52. Long F. Building strong bones: molecular regulation of the osteoblast lineage. *Nat Rev Mol Cell Bio*. 2012;13:27–38.
53. Karsenty G, Kronenberg HM, Settembre C. Genetic control of bone formation. *Ann Rev Cell Dev Bio*. 2009;25:629–648.
54. Dallas SL, Prideaux M, Bonewald LF. The osteocyte: an endocrine cell ... and more. *Endocr Rev*. 2013;34:658–690.
55. Bonewald LF. The amazing osteocyte. *J Bone Miner Res*. 2011;26:229–238.
56. Klein-Nulend J, Bakker AD, Bacabac RG, et al. Mechanosensation and transduction in osteocytes. *Bone*. 2013;54:182–190.
57. Nusse R, Clevers H. Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. *Cell*. 2017;169:985–999.
58. Maupin KA, Droscha CJ, Williams BO. A comprehensive overview of skeletal phenotypes associated with alterations in Wnt/beta-catenin signaling in humans and mice. *Bone Res*. 2013;1:27–71.
59. Chen J, Long F. Beta-catenin promotes bone formation and suppresses bone resorption in postnatal growing mice. *J Bone Miner Res*. 2013;28:1160–1169.
60. Song L, Liu M, Ono N, et al. Loss of wnt/beta-catenin signaling causes cell fate shift of preosteoblasts from osteoblasts to adipocytes. *J Bone Miner Res*. 2012;27:2344–2358.
61. Huybrechts Y, Mortier G, Boudin E, et al. WNT signaling and bone: lessons from skeletal dysplasias and disorders. *Front Endocrinol (Lausanne)*. 2020;11:165.

62. Gong Y, Slee RB, Fukai N, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell*. 2001;107:513–523.
63. Boyden LM, Mao J, Belsky J, et al. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med*. 2002;346:1513–1521.
64. Tian E, Zhan F, Walker R, et al. The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma. *N Engl J Med*. 2003;349:2483–2494.
65. Otto F, Thornell AP, Crompton T, et al. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell*. 1997;89:765–771.
66. Komori T, Yagi H, Nomura S, et al. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell*. 1997;89:755–764.
67. Lee B, Thirunavukkarasu K, Zhou L, et al. Missense mutations abolishing DNA binding of the osteoblast-specific transcription factor OSF2/CBFA1 in cleidocranial dysplasia. *Nat Genet*. 1997;16:307–310.
68. Mundlos S, Otto F, Mundlos C, et al. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell*. 1997;89:773–779.
69. Nakashima K, Zhou X, Kunkel G, et al. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell*. 2002;108:17–29.
70. Salazar VS, Gamer LW, Rosen V. BMP signalling in skeletal development, disease and repair. *Nat Rev Endocrinol*. 2016;12:203–221.
71. Shore EM, Xu M, Feldman GJ, et al. A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet*. 2006;38:525–527.
72. Long F, Chung UI, Ohba S, et al. Ihh signaling is directly required for the osteoblast lineage in the endochondral skeleton. *Development*. 2004;131:1309–1318.
73. Romeo SG, Alawi KM, Rodrigues J, et al. Endothelial proteolytic activity and interaction with non-resorbing osteoclasts mediate bone elongation. *Nat Cell Biol*. 2019;21:430–441.
74. Tsukasaki M, Takayanagi H. Osteoimmunology: evolving concepts in bone-immune interactions in health and disease. *Nat Rev*. 2019;19:626–642.
75. Boyce BF. Advances in the regulation of osteoclasts and osteoclast functions. *J Dent Res*. 2013;92:860–867.
76. Lorenzo J, Horowitz M, Choi Y. Osteoimmunology: interactions of the bone and immune system. *Endocr Rev*. 2008;29:403–440.
77. Ono T, Hayashi M, Sasaki F, et al. RANKL biology: bone metabolism, the immune system, and beyond. *Inflamm Regen*. 2020;40:2.
78. Nakashima T, Hayashi M, Fukunaga T, et al. Evidence for osteocyte regulation of bone homeostasis through RANKL expression. *Nat Med*. 2011;17:1231–1234.
79. Xiong J, Onal M, Jilka RL, et al. Matrix-embedded cells control osteoclast formation. *Nat Med*. 2011;17:1235–1241.
80. Roodman GD, Windle JJ. Paget disease of bone. *J Clin Invest*. 2005;115:200–208.
81. Blangy A, Bompard G, Guerit D, et al. The osteoclast cytoskeleton - current understanding and therapeutic perspectives for osteoporosis. *J Cell Sci*. 2020;133:jcs244798.
82. Feng X, Teitelbaum SL. Osteoclasts: new insights. *Bone Res*. 2013;1:11–26.
83. Rachner TD, Khosla S, Hofbauer LC. Osteoporosis: now and the future. *Lancet*. 2011;377:1276–1287.
84. Holmbeck K, Bianco P, Caterina J, et al. MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. *Cell*. 1999;99:81–92.
85. Sims NA, Martin TJ. Osteoclasts provide coupling signals to osteoblast lineage cells through multiple mechanisms. *Annu Rev Physiol*. 2020;82:507–529.
86. Kim JM, Lin C, Stavre Z, et al. Osteoblast-osteoclast communication and bone homeostasis. *Cells*. 2020;9:2073.
87. Maeda K, Kobayashi Y, Udagawa N, et al. Wnt5a-Ror2 signaling between osteoblast-lineage cells and osteoclast precursors enhances osteoclastogenesis. *Nat Med*. 2012;18:405–412.
88. Hayashi M, Nakashima T, Taniguchi M, et al. Osteoprotection by semaphorin 3A. *Nature*. 2012;485:69–74.
89. Takegahara N, Takamatsu H, Toyofuku T, et al. Plexin-A1 and its interaction with DAP12 in immune responses and bone homeostasis. *Nat Cell Biol*. 2006;8:615–622.
90. Henriksen K, Karsdal MA, Martin JT. Osteoclast-derived coupling factors in bone remodeling. *Calcif Tissue Int*. 2014;94:88–97.
91. Tang Y, Wu X, Lei W, et al. TGF-beta1-induced migration of bone mesenchymal stem cells couples bone resorption with formation. *Nat Med*. 2009;15:757–765.
92. Xian L, Wu X, Pang L, et al. Matrix IGF-1 maintains bone mass by activation of mTOR in mesenchymal stem cells. *Nat Med*. 2012;18:1095–1101.
93. Takeshita S, Fumoto T, Matsuoka K, et al. Osteoclast-secreted CTHRC1 in the coupling of bone resorption to formation. *J Clin Invest*. 2013;123:3914–3924.
94. Tatsumi S, Ishii K, Amizuka N, et al. Targeted ablation of osteocytes induces osteoporosis with defective mechanotransduction. *Cell Metabolism*. 2007;5:464–475.
95. Goltzman D. Physiology of parathyroid hormone. *Endocrinol Metab Clin North Am*. 2018;47:743–758.
96. Potts JT, Gardella TJ. Progress, paradox, and potential: parathyroid hormone research over five decades. *Ann NY Acad Sci*. 2007;1117:196–208.
97. Tsourdi E, Jahn K, Rauner M, et al. Physiological and pathological osteocytic osteolysis. *J Musculoskelet Neuronal Interact*. 2018;18:292–303.
98. Jilka RL. Molecular and cellular mechanisms of the anabolic effect of intermittent PTH. *Bone*. 2007;40:1434–1446.
99. Khosla S, Monroe DG. Regulation of bone metabolism by sex steroids. *Cold Spring Harb Perspect Med*. 2018;8:a031211.
100. Manolagas SC, O'Brien CA, Almeida M. The role of estrogen and androgen receptors in bone health and disease. *Nat Rev Endocrinol*. 2013;9:699–712.
101. Nakamura T, Imai Y, Matsumoto T, et al. Estrogen prevents bone loss via estrogen receptor alpha and induction of fas ligand in osteoclasts 1. *Cell*. 2007;130:811–823.
102. Kousteni S, Chen JR, Bellido T, et al. Reversal of bone loss in mice by nongenotropic signaling of sex steroids. *Science*. 2002;298:843–846.
103. Vanderschueren D, Laurent MR, Claessens F, et al. Sex steroid actions in male bone. *Endocr Rev*. 2014;35:906–960.
104. Chotiyarnwong P, McCloskey EV. Pathogenesis of glucocorticoid-induced osteoporosis and options for treatment. *Nat Rev Endocrinol*. 2020;16:437–447.
105. Elefteriou F. Impact of the autonomic nervous system on the skeleton. *Physiol Rev*. 2018;98:1083–1112.
106. Harada S, Rodan GA. Control of osteoblast function and regulation of bone mass. *Nature*. 2003;423:349–355.
107. Zhou BO, Yue R, Murphy MM, et al. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell*. 2014;15:154–168.
108. Lee NK, Sowa H, Hinoi E, et al. Endocrine regulation of energy metabolism by the skeleton. *Cell*. 2007;130:456–469.
109. Karsenty G. Update on the biology of osteocalcin. *Endocr Pract*. 2017;23:1270–1274.
110. Obri A, Khrimian L, Karsenty G, et al. Osteocalcin in the brain: from embryonic development to age-related decline in cognition. *Nat Rev Endocrinol*. 2018;14:174–182.
111. Lambert LJ, Challa AK, Niu A, et al. Increased trabecular bone and improved biomechanics in an osteocalcin-null rat model created by CRISPR/Cas9 technology. *Dis Model Mech*. 2016;9:1169–1179.
112. Diegel CR, Hann S, Ayturk UM, et al. An osteocalcin-deficient mouse strain without endocrine abnormalities. *PLoS Genet*. 2020;16:e1008361.
113. Moriishi T, Ozasa R, Ishimoto T, et al. Osteocalcin is necessary for the alignment of apatite crystallites, but not glucose metabolism, testosterone synthesis, or muscle mass. *PLoS Genet*. 2020;16:e1008586.
114. Komori T. Functions of osteocalcin in bone, pancreas, testis, and muscle. *Int J Mol Sci*. 2020;21:7513.
115. Oldknow KJ, MacRae VE, Farquharson C. Endocrine role of bone: recent and emerging perspectives beyond osteocalcin. *J Endocrinol*. 2015;225:R1–R19.
116. Dirckx N, Moorer MC, Clemens TL, et al. The role of osteoblasts in energy homeostasis. *Nat Rev Endocrinol*. 2019;15:651–665.
117. Mesnieres M, Maes C. Osteoblast lineage stem and progenitor cells. In: Zaidi M, ed. *Encyclopedia of Bone Biology. Section: Osteoblast Biology*. 1st ed. Oxford: Academic Press; 2020:383–396.

52

Regulation of Calcium Homeostasis and Genetic Disorders That Affect Calcium Metabolism

Rajesh V. Thakker and Harald Jüppner

OUTLINE

Regulation of Calcium Homeostasis, 889
Distribution and Metabolic Actions of Calcium, 889
Calcium Absorption, 891
Mechanisms and Sites of Calcium Absorption, 891
Renal Calcium Excretion, 891
Sites and Mechanisms of Renal Calcium Reabsorption, 892
Regulation of Renal Calcium Reabsorption, 892
PTH Gene Structure and Function, 892
Hypercalcemic Diseases, 893
Parathyroid Tumors, 893
Syndromic Forms of Primary Hyperparathyroidism, 893
Nonsyndromic Forms of Primary Hyperparathyroidism, 896
Gene Testing for Primary Hyperparathyroidism in Clinical Practice, 897
Indications for Mutational Analysis in Primary Hyperparathyroidism, 897
Clinical Approach to Gene Testing in a Patient With Primary Hyperparathyroidism, 897
Familial Hypocalciuric Hypercalcemia and Disorders of the Calcium-Sensing Receptor, 898
Other Hereditary Hypercalcemic Disorders, 898
Hypocalcemic Disorders, 900
Hypoparathyroidism, 900
PTH Gene Abnormalities, 900
GCM2 Abnormalities, 901
X-Linked Recessive Hypoparathyroidism, 901
Pluriglandular Autoimmune Hypoparathyroidism, 901
DiGeorge Syndrome, 902
CHARGE Syndrome, 902
Hypoparathyroidism, Deafness, and Renal Anomalies Syndrome, 902
Mitochondrial Disorders Associated With Hypoparathyroidism, 903
Kenny–Caffey, Sanjad–Sakati, Kirk–Richardson, and Additional Familial Syndromes, 903
Calcium-Sensing Receptor Abnormalities, 903
Gα11 Abnormalities and Autosomal Dominant Hypocalcemia Type 2, 904
Blomstrand Disease, 904
Conclusion, 904

Calcium plays an important role in many physiologic pathways, including muscle contraction, the secretion of neurotransmitters and hormones, and the coagulation pathways, and is an important component of the skeleton.[1,2] Disturbances in extracellular calcium concentration therefore may cause a variety of symptoms, the most common of which reflect abnormal neuromuscular activity. Thus, hypercalcemia may lead to muscle weakness and areflexia, anorexia, constipation, vomiting, drowsiness, depression, confusion, other cognitive dysfunction, and coma.[2] Hypercalcemia leads to hypercalciuria, which can result in medullary calcification, nephrocalcinosis, and impaired kidney function. Hypocalcemia, conversely, may cause anxiety, seizures, muscle twitching, epilepsy, tetany, Chvostek's and Trousseau's signs, carpal or pedal spasm, stridor, bronchospasm, and intestinal cramps, as well as cataracts, basal ganglia calcifications, skeletal malformations, and abnormal dentition.[1] The control of body calcium involves a balance between the amounts that are absorbed from the gut, deposited into bone and into cells, and excreted from the kidney (Fig. 52.1). This fine balance, involving three organs, is chiefly under the control of parathyroid hormone (PTH), which is synthesized and secreted by the parathyroid glands.[1,2] Thus, hypocalcemia will lead to an increased secretion of PTH, whereas hypercalcemia will result in diminished PTH secretion. A number of clinical disorders characterized by derangements of calcium homeostasis are caused by abnormalities of the parathyroid glands themselves. Thus, PTH oversecretion attributed to parathyroid tumors, which affect 3:1000 in the general population, is a major cause of hypercalcemia, which may be associated with kidney stones, osteoporosis, and peptic ulcers. PTH deficiency resulting in hypocalcemia is thought to occur in 1:4000 live births, but precise data regarding the incidence are lacking. This chapter will review the physiologic and biochemical mechanisms underlying extracellular calcium homeostasis and the genetic basis for disorders of calcium metabolism.

KEY POINTS

- Parathyroid hormone is the most important peptide hormone for regulating calcium homeostasis. Its synthesis and secretion is controlled by the calcium-sensing receptor, a G protein–coupled receptor that is expressed in the parathyroid glands.

REGULATION OF CALCIUM HOMEOSTASIS

Distribution and Metabolic Actions of Calcium

The total-body content of calcium in a normal adult is approximately 1000 g, nearly all of which exists within the crystal structure of bone mineral, and less than 1% of which is dissolved in extracellular and intracellular fluid.[1,2] On average, bone mineral closely approximates the composition of hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$), which means

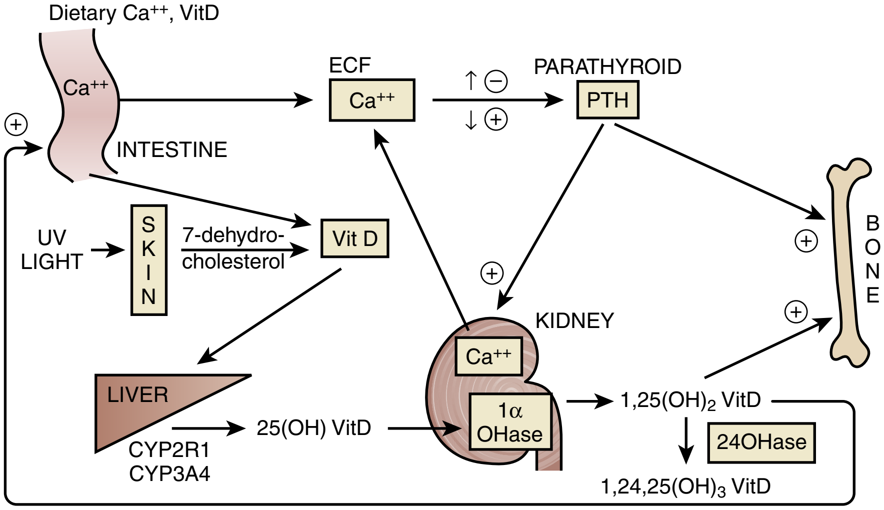


Fig. 52.1 Regulation of Extracellular Fluid (*ECF*) Calcium (Ca2^{++}) by Parathyroid Hormone (*PTH*) Action on Kidney, Bone, and Intestine. The calcium-sensing receptor (CaSR) senses a decrease in ECF Ca2^{++}, and this leads to an increase in PTH secretion, which acts directly on kidney and bone that express the PTH/PTHrP receptor (PTHR1; see Fig. 52.2). The skeletal effects of PTH are to increase (+) osteoclastic bone reabsorption, but as osteoclasts do not express the PTHR1, this action is mediated via the osteoblasts, which do have this receptor, and in response release cytokines and factors that activate osteoclasts. In the kidney, PTH stimulates (+) the 1α hydroxylase (1αOHase) to increase the conversion of 25-hydroxy vitamin D (25(OH)D) to the active metabolite 1,25-dihydroxy vitamin D (1,25(OH)$_2$D), which is subsequently inactivated by the 24-hydroxylase. In addition, PTH increases (+) the reabsorption of Ca2^{++} from the renal distal tubule and inhibits the reabsorption of phosphate from the proximal tubule, thereby leading to hypercalcemia and hypophosphatemia. PTH also inhibits Na$^+$–H$^+$ antiporter activity and bicarbonate reabsorption, thereby causing a mild hyperchloremic acidosis, and it reduces expression of two sodium-dependent phosphate cotransporters, NPT2a and NPT2c, thereby increasing urinary phosphate excretion. The elevated 1,25(OH)$_2$D acts on the intestine to increase (+) absorption of dietary calcium and phosphate, and it is important to note that PTH does not appear to have a direct action on the gut. Thus, in response to hypocalcemia and the increase in PTH secretion, all of these direct and indirect actions of PTH on the kidney, bone, and intestine will help to increase ECF Ca2^{++}, which in turn will act via the CaSR to decrease PTH secretion.

that 6 mmol of phosphate is released with every 10 mmol of calcium mobilized during bone resorption (or ~1:2 on a mg/mg basis).

In blood, 50% of calcium is bound to proteins. Albumin accounts for approximately 80% of the protein-bound fraction. Another portion (~10%) is associated with diffusable ion complexes. Thus, normally only half of total plasma calcium is freely ionized, but it is this fraction that is most important physiologically and thus subject to stringent endocrine regulation. The normal ranges of total and ionized serum calcium in the adult are 8.5 to 10.5 mg/dL and 4.25 to 5.25 mg/dL, respectively.[2] The usual 2:1 ratio of total to ionized calcium may be disturbed by disorders such as metabolic acidosis (calcium binding to proteins is reduced at acid pH) or by changes in serum protein concentrations, as in starvation, cirrhosis, dehydration, or multiple myeloma. Thus, ionized calcium concentrations can nowadays be routinely determined, if clinically required.[3]

Serum calcium is higher during infancy/childhood and adolescence than in the adult[7] but does not change at puberty or, in women, during the menstrual cycle.[2] During pregnancy, total serum calcium and albumin decline progressively, but ionized calcium is minimally affected. Fetal calcium content rises dramatically during the third trimester (to ~30 g at term), and fetal serum total and ionized calcium concentrations are higher than maternal levels, consistent with active placental transport of calcium.[4] Despite losses of 200 to 300 mg/day of calcium in breast milk, lactating women maintain normal levels of ionized calcium in blood by increasing intestinal calcium absorption in response to augmented production of 1,25-dihydroxy vitamin D (1,25(OH)$_2$D).

Calcium entry into cells is strongly favored by a steep electrochemical gradient. Thus, the cell interior is electronegative, and cytosolic free calcium is in the range of 10^{-7} M, which is 10,000-fold lower than extracellular calcium concentrations. Calcium traverses the plasma membrane through various channels, including voltage-, receptor-, and store-operated forms, the regulation of which is complex and tissue-specific.[5] Intracellularly, nearly all (99%) calcium is sequestered in pools within mitochondria, endoplasmic reticula, or sarcoplasmic reticula or is tightly bound to the inner surface of the plasma membrane. Sequestered calcium, especially that within the endoplasmic reticulum, may be released rapidly into the cytosol following activation of cell-surface receptors.[1,5] In this way, it plays a critical role in signal transduction and in controlling calcium entry via store-operated channels. The extremely low cytosolic free calcium concentration is maintained by active calcium transport into intracellular pools or by extrusion out of the cell via high-affinity, low-capacity Ca^{++}/H$^+$ ATPases and low-affinity, high-capacity Na$^+$/Ca^{++} exchangers driven by the transmembrane sodium gradient.

With the discovery of a calcium-sensing receptor (CaSR), a G protein–coupled receptor (GPCR) expressed in parathyroid, renal, epithelial, and numerous other cells, it has become clear that calcium can act as an extracellular ligand to directly control cellular function.[5] The principal actions of the CaSR with respect to calcium homeostasis include suppression of PTH secretion and, in the thick ascending loop of the renal tubule, inhibition of calcium, magnesium, and NaCl reabsorption.[5] Extracellular calcium also is used directly for normal matrix mineralization in bone and cartilage and is required for activation of important circulating or extracellular enzymes and proteases. Intracellular calcium exerts a broad range of effects via interaction with key enzymes and effector molecules, including kinases, phosphatases, calmodulins, transcription factors, ion channels (including calcium channels), and troponins and other proteins involved in contraction, microtubule and microfilament assembly, and motility.[5] The steep gradients between intracellular calcium and both extracellular and intracellular sequestered calcium are crucial for normal neuromuscular activity and provide the potential required for rapid transients and waves of cytosolic free calcium that serve key second-messenger functions in both excitable and nonexcitable cells.

Calcium Absorption

Intestinal calcium absorptive efficiency in humans ranges broadly between 20% and 70%, declines with age, and is strongly influenced by previous calcium intake, the presence of other nutrients, pregnancy, lactation, overall calcium balance, and the availability of activated vitamin D. Fecal calcium includes the residual fraction of dietary calcium that is not absorbed, as well as a contribution from secreted calcium present in bile and other digestive juices ("endogenous fecal calcium"). Endogenous fecal calcium in humans normally amounts to 100 to 200 mg/day and is relatively unaffected by changes in dietary or serum calcium.

Intestinal calcium absorption is adjusted physiologically in response to variations in calcium intake, as shown by radiocalcium kinetic and balance studies in normal subjects. Obligate renal and intestinal excretion of calcium is such that calcium balance cannot be maintained if dietary calcium consistently falls below 200 to 400 mg/day, even though the percentage of calcium absorbed may be very high (i.e., 70%). As calcium intake increases, overall calcium absorption (in mg/day) rises, but the fractional absorption of ingested calcium declines progressively such that, within the physiologic range of calcium intake, total net calcium absorption tends to plateau at approximately 400 mg/day. Consequently, urinary calcium excretion tends also to plateau at higher intakes. Additional buffering of changes in dietary calcium results from control of renal tubular calcium reabsorption and skeletal calcium release, but regulation of intestinal absorptive efficiency is critical for calcium balance. The mechanisms of this inverse regulation of intestinal absorptive efficiency by changes in calcium availability are considered further later in the chapter.

Mechanisms and Sites of Calcium Absorption

Calcium is absorbed throughout the intestine. In terms of rates of transport per unit length of mucosa, absorption is most efficient in the duodenum and proximal jejunum, which exhibit the highest levels of vitamin D–dependent calcium-binding proteins and in which lower luminal pH promotes dissociation of calcium from complexes with food constituents and other ions. On the other hand, longer residence times in the more distal small bowel segments may allow absorption of a larger proportion of total calcium intake in the distal jejunum and ileum. The ileum, for example, may become an important site of net calcium absorption during dietary restriction of calcium or when the residence time of luminal contents in more proximal bowel segments is reduced.

Calcium is absorbed by both a passive, paracellular route and by an active, transcellular mechanism. Paracellular absorption predominates at higher levels of dietary calcium, whereas transcellular absorption, which is strongly upregulated by activated vitamin D, is most important when dietary calcium is limiting. Transcellular calcium absorption necessarily involves three steps: entry into the cell, diffusion across the cell, and extrusion from the cell (see Chapter 50). The first step, entry through the apical brush border, occurs via a member of the vanilloid (TRPV) superfamily of channels, namely TRPV6.[6] Another closely related member of this family, TRPV5 (formerly ECaC1 or CaT2), appears to be dominantly involved in calcium reabsorption in the distal renal tubules (see later). The genes encoding TRPV6 and TRPV5 are located on chromosome 7q33-35, and the protein structure includes six transmembrane regions, a short hydrophobic region between segments 5 and 6 that likely forms the calcium pore, and large intracellular domains at the N- and C-terminals. The pore region is highly selective for calcium,[7] and mutations in this region have been shown to abolish calcium permeability.[8] Posttranslational modification and glycosylation of TRPV6 occurs, but the functional effects of these modifications are not well defined. Similarly, the mechanisms regulating the activity of these molecules are not well understood. *TRPV6* and *TRPV5* expression varies significantly among species and in humans, and both genes are expressed in a variety of tissues.[9] Specifically, TRPV5 seems to be the major isoform expressed in the kidney, while TRPV6 is most highly expressed in the small intestine (restricted to the apical surface of epithelial cells). Given the important role of $1,25(OH)_2D$ and other hormonal regulators in calcium absorption, the effects of these hormones on TRPV6 expression have been investigated. $1,25(OH)_2D$ increases TRPV6 expression[10] (see Chapter 50). Furthermore, mice with targeted inactivation of the vitamin D receptor demonstrate a 90% reduction in TRPV6 expression and a threefold reduction in intestinal calcium absorption.[11] Additionally, calcium exposure appears to independently affect TRPV6 expression because high calcium intake decreases TRPV6 expression, and low intake increases it. Finally, TRPV6 expression is decreased in mice with targeted inactivation of their estrogen receptor-α gene and is increased in mice receiving exogenous estradiol administration.[12]

After calcium has entered the cell, it must be transported through the cytoplasm and extruded from the cell against a steep gradient. The mechanism by which calcium moves through the cell involves the small cytosolic protein calbindin D9k. Calcium entering the cell via the apical TRPV6 channel becomes tightly associated with the calbindin, which buffers the relatively large mass of entering calcium and minimizes its impact upon cytosolic free calcium concentrations. The calbindin/calcium complex then diffuses across the cytosol to the basolateral membrane. Free calcium then dissociates into the low-cytosolic calcium environment maintained immediately subjacent to the basolateral membrane by high-affinity membrane Ca^{++}-ATPases located there. Finally, these calcium ATPases actively extrude calcium out of the cell. The known cellular concentrations and kinetic properties of the calbindin D9k molecule would support observed rates of duodenal calcium transport at submicromolar concentrations of cytosolic free calcium. Indeed, the importance of this buffered diffusional process predicts that enterocyte calbindin D9k content is likely to be a major determinant, along with TRPV6 activity, of the overall rate of enterocyte calcium transport.

Renal Calcium Excretion

Regulation of renal calcium excretion is an important mechanism for homeostatic control of blood ionized calcium in the face of fluctuations in filtered load, as derived from intestinal calcium absorption and net bone resorption.[5] When urinary calcium is viewed as a function of the amount of calcium actually absorbed by the gut (i.e., after regulation of intestinal absorptive efficiency has been factored out), the precision with which the kidney adjusts tubular calcium reabsorption to residual changes in filtered load becomes obvious. Ordinarily, the daily load of calcium filtered at the glomerulus (the product of glomerular filtration rate and ultrafilterable calcium) is approximately 10,000 mg/day in adult humans, which means that the extracellular calcium pool is completely filtered several times per day. As urinary calcium excretion (and net intestinal calcium absorption) is approximately 200 mg/day, only 2% of filtered calcium is excreted normally. This high ratio of filtered calcium to excreted calcium affords ample opportunity for finely tuned hormonal control of calcium excretion, even though it may be difficult to measure the small changes involved.

Sites and Mechanisms of Renal Calcium Reabsorption

Calcium is reabsorbed at multiple sites and by different mechanisms along the nephron.[5] As in the intestine, the challenge of transporting calcium across the renal epithelium requires that adequate rates of transport be achieved without substantially increasing the concentration of calcium within the cytosol of the epithelial cell. This is accomplished mainly by use of paracellular diffusional mechanisms that are supplemented, in some nephron segments, by active transcellular transport, especially in response to hormonal stimulation.

Approximately 60% to 70% of tubular calcium reabsorption, like that of sodium, occurs in the proximal tubule. Proximal tubular calcium reabsorption occurs mainly via passive diffusion along paracellular pathways, down the ambient (lumen-positive) electrochemical gradient. Another 20% to 25% of the filtered calcium load is reabsorbed in Henle's loop. This occurs mainly by paracellular diffusion in the cortical thick ascending limb (cTAL), although active transcellular transport may play a minor role as well. Calcium (and magnesium) reabsorption in this segment is severely impaired in patients with homozygous inactivating mutations in claudin 16 (CLDN16), previously referred to as *paracellin-1,* or claudin 19 (CLDN19), which are specifically expressed in tight junctions of the cTAL and are presumably required for normal divalent cation conductance.[5,13] Hypercalcemia suppresses renal tubular calcium reabsorption, even in the absence of PTH, an effect attributable to direct activation of basolateral CaSRs in the cTAL (and possibly other nephron segments). In the cTAL, activated CaSRs lower apical K^+ exit by inhibiting renal outer medullary K^+ channels (ROMKs), thereby blunting Na-K-2Cl co-transporter (NKCC2) activity and reducing the transepithelial voltage gradient that drives paracellular cation movement.[14] The critical importance of normal rates of Na-K-2Cl reabsorption for adequate calcium transport in the cTAL is highlighted clinically by the occurrence of dramatic hypercalciuria in patients with Bartter syndrome, which involves defects in NKCC2, ROMK, or other transporters or channels required for effective Na-K-2Cl reabsorption.[14] An analogous Bartter-like phenotype may occur in some patients with activating CaSR mutations.[15] CaSR activation in the cTAL also antagonizes PTH-stimulated increases in calcium reabsorption, possibly by impairing cyclic adenosine monophosphate (cAMP) generation (see later).

Only approximately 8% to10% of filtered calcium is reabsorbed in more distal segments of the tubule, but calcium transport in the distal nephron is a key control point for hormonal regulation and involves predominantly transcellular reabsorption.[5] Cells in the late distal tubule exhibit striking colocalization of several unique proteins that are critical for effective transcellular calcium reabsorption. These include an apical epithelial calcium channel, TRPV5, the cytosolic calcium-binding protein calbindin-D28K, basolateral plasma membrane calcium ATPase(s) (PMCAs), and basolateral Na^+/Ca^{++} exchanger-1 (NCX1).[16,17] As in the intestinal epithelium, these cells admit calcium across their apical membranes via opening of TRPV5. Expression of this channel is upregulated by PTH, thus enhancing calcium-reabsorption (see later). PTH also increases calbindin-D28K levels, which enhances transcellular calcium reabsorption via increased buffering of subapical Ca^{++} ions. Calcium bound to calbindin-D28K is ferried across the cell to the basolateral membrane, where it ultimately is extruded against a steep electrochemical gradient, predominantly via NCX1 (70% of Ca^{++} flux) but also by PMCA. As expected, mice with a TRPV5 mutation or deficiency[18] or lacking calbindin D28K[19] manifest severe hypercalciuria.

Additional calcium reabsorption may occur in more distal nephron segments, such as cortical and medullary collecting ducts. Proteins other than those described earlier may be involved in renal calcium reabsorption as well, including voltage-operated calcium channels and Cl– channels, as evidenced by the hypercalciuria seen in Dent's disease, an X-linked disorder caused by inactivating mutations in the Cl– channel ClC–5, which may be involved in endocytic vesicle function and protein trafficking.[20,21]

Regulation of Renal Calcium Reabsorption

Calcium excretion is affected by a variety of hormones, ions, nutrients, and drugs. Among these, PTH is the principal physiologic regulator of renal tubular calcium transport, as appreciated clinically by the markedly abnormal overall relation of serum to urinary calcium in hyperparathyroidism and hypoparathyroidism, and the reduction in calcium excretion observed after injection of PTH in healthy individuals or in patients affected by pseudohypoparathyroidism.[22] PTH acts to enhance tubular calcium reabsorption in the cTAL, distal convoluted tubules (DCT), and connecting tubules (CNT). It augments DCT calbindin-D28K expression[23] and basolateral Na^+/Ca^{++} exchange, increases the affinity for calcium of the basolateral Ca^{++}-ATPase, and hyperpolarizes the DCT cell, which then activates the TRPV5 channel.[24] In immortalized murine distal tubular cells, PTH causes insertion into the apical membrane of new dihydropyridine-sensitive membrane calcium channels, and it stimulates stretch-activated nonselective cation channels in rabbit CNT. These are distinct from TRPV5 channels and may represent other mechanisms of apical calcium entry. PTH increases the surface expression or activity of TRPV5 epithelial calcium channels. In the cTAL, PTH also may increase some form of active transcellular transport, and it may stimulate paracellular, diffusional calcium transport by augmenting the transepithelial voltage gradient, at least in some segments and species.

Studies in patients with inactivating vitamin D receptor mutations have demonstrated that $1,25(OH)_2D$ is required for the calcium-reabsorptive response to PTH. It furthermore accelerates the increase in DCT cell calcium entry initiated by PTH and directly increases DCT calcium reabsorption, apparently by driving higher expression of TRPV5, calbindin D28K, and PMCA[25] (see Chapter 50). Other hormones that affect calcium reabsorption include calcitonin, which, when given in large doses, acutely reduces proximal tubular calcium reabsorption by a mechanism independent of PTH; however, an important role for calcitonin in the physiologic regulation of calcium reabsorption is thought to be unlikely. Hypercalciuria observed in states of excess growth hormone or cortisol seems likely to be secondary to an increased filtered load of calcium rather than to direct tubular actions of these hormones. Estrogen treatment of normal postmenopausal women lowers urinary calcium excretion by increasing tubular calcium reabsorption. In DCT of ovariectomized rats, 17β-estradiol acutely increases mRNA expression of all major components of the calcium-transport pathway, including TRPV5, calbindin D28K, NCX1, and PMCA1b.[26] Reported effects on tubular calcium reabsorption of insulin, glucagon, antidiuretic hormone, and angiotensin II[27] are of uncertain physiologic or clinical significance.

PTH GENE STRUCTURE AND FUNCTION

The *PTH* gene is located on chromosome 11p15 and consists of three exons. Exon 1 is untranslated, whereas exons 2 and 3 encode the 115–amino acid pre-proPTH peptide (see Chapter 47) and the 3′ untranslated region. The 5′ regulatory sequence of the human *PTH* gene contains a vitamin D response element 125 bp upstream of the transcription start site, which downregulates PTH mRNA transcription in response to vitamin D receptor binding. *PTH* gene transcription (as well as PTH peptide secretion) is also dependent upon the extracellular calcium and phosphate concentration, although the presence of a specific upstream "calcium or phosphate response element(s)" has not yet been demonstrated. The secretion of mature PTH, which is an 84-amino acid peptide, from the parathyroid chief cell is regulated through the CaSR, which is also expressed in the renal tubules and in numerous other tissues, albeit at lower abundance.[1,2,5,28] PTH mRNA is first translated into a pre-proPTH peptide. After the mature PTH peptide is secreted from the parathyroid cell, it is cleared from the circulation with a short half-life of approximately 2 minutes, via nonsaturable hepatic and renal uptake (see also Chapter 47).

KEY POINTS

- The widely expressed parathyroid hormone (PTH)/parathyroid hormone–related peptide (PTHrP) receptor (PTHR1), a G protein–coupled receptor, mediates the actions of PTH and PTHrP. Because of its dual roles, *PTHR1* mutations affect mineral ion homeostasis as well as chondrocyte differentiation, and thus bone growth.

PTH mediates its actions through a receptor that it shares with PTH-related peptide (PTHrP, also known as PTHLH or PTH-like hormone).[1,5] This PTH/PTHrP receptor (Fig. 52.2), a GPCR whose gene is located on chromosome 3p21.3, is highly expressed in kidney and bone, where it mediates the endocrine actions of PTH. However, during embryonic and postnatal development, the PTH/PTHrP receptor is most abundantly expressed in chondrocytes of the metaphyseal growth plate, where it mediates predominantly the autocrine/paracrine actions of PTHrP.[29] Mutations involving the genes that encode PTH, the CaSR, the PTH/PTHrP receptor, AP2S1, Gα11, and G$_s\alpha$ all affect the regulation of calcium homeostasis and can thus be associated with genetic disorders characterized by hypercalcemia or hypocalcemia (Table 52.1).

KEY POINTS

- Studies of syndromic and nonsyndromic forms of primary hyperparathyroidism and hypoparathyroidism, which may occur as hereditary (i.e., familial) or nonfamilial (i.e., sporadic) diseases, have helped to identify some of the genetic abnormalities involved in the parathyroid tumorigenesis and in the embryonic development of the parathyroid glands.

HYPERCALCEMIC DISEASES

Genetic disorders resulting in hypercalcemia may be associated with parathyroid tumors, disorders of the CaSR or PTH/PTHrP receptors, impaired 1,25(OH)$_2$D inactivation, or chromosomal deletion syndromes (e.g., Williams syndrome).[2,30]

Parathyroid Tumors

Similar to the findings in other tumors, the abnormal expression of an oncogene or the loss of a tumor suppressor gene can result in an abnormal proliferative activity of parathyroid cells, and the molecular exploration of these genes has provided important novel insights into the pathogenesis of different forms of hyperparathyroidism.

Genetic abnormalities involving one of 12 genes may be found in approximately 10% of patients with parathyroid tumors, giving rise to primary hyperparathyroidism (PHPT), and resulting in syndromic and nonsyndromic forms of PHPT, which may occur as hereditary (i.e., familial) disorders or as nonfamilial (i.e., sporadic) diseases.[30,31] However, the distinction between sporadic and familial forms in PHPT patients is sometimes difficult because a family history may be absent in some sporadic cases, as the parent with the disease may not have been investigated or may have died before PHPT symptoms developed; or the PHPT may be attributed to a *de novo* germline mutation in the patient, which would account for an absent family history, but results in an increased risk for hereditary PHPT in the children of the patient. Syndromic forms of PHPT may occur as part of complex disorders (e.g., as part of multiple endocrine neoplasia [MEN] syndromes) and be inherited as autosomal dominant traits, (Table 52.1).[30,32] Such syndromic PHPT forms are MEN1, MEN2, MEN3 (also referred to as MEN2b), MEN4, and the hyperparathyroidism–jaw tumor (HPT-JT) syndrome.[30,32,33] PHPT as a nonsyndromic isolated endocrinopathy may occur as a familial disorder, which is referred to as *familial isolated hyperparathyroidism* (FIHP), or as a nonfamilial (sporadic) disorder.[30,32]

KEY POINTS

- So far, only approximately 10% of all patients with primary hyperparathyroidism have a defined mutation in one of 12 genes known to be associated with hyperparathyroidism. However, testing for mutations in these genes will help to confirm the diagnosis of a syndromic or nonsyndromic form of primary hyperparathyroidism, and thereby help in the clinical management and treatment of the disorder in patients and their relatives.

Syndromic Forms of Primary Hyperparathyroidism

Multiple Endocrine Neoplasia Type 1. MEN type 1 (MEN1) is characterized by the combined occurrence of tumors of the parathyroid glands, pancreatic islet cells, and anterior pituitary gland.[30,32] Parathyroid tumors occur in 95% of MEN1 patients, and the resulting hypercalcemia is the first manifestation of MEN1 in approximately 90% of patients. Pancreatic islet cell tumors occur in 40% of MEN1 patients and gastrinomas, leading to Zollinger–Ellison syndrome, are the most common type and also the important cause of morbidity and mortality in MEN1 patients. Anterior pituitary tumors occur in 30% of MEN1 patients, with prolactinomas representing the most common type. Associated tumors, which may also occur in MEN1, include adrenal cortical tumors, carcinoid tumors, lipomas, and cutaneous angiofibromas and collagenomas. The gene causing MEN1 is located on chromosome 11q13 and consists of 10 exons that encode a 610–amino acid protein referred to as *menin*. Over 1100 germline *MEN1* mutations have been identified, and the majority (>80%) are inactivating, consistent with its role as a tumor suppressor gene.[32,34] These mutations are diverse in their types: approximately 25% are nonsense, approximately 45% are frameshift deletions or insertions, approximately 9% are splice-site mutations, approximately 20% are missense mutations, and approximately 1% are whole or partial gene deletions.[30,34] In addition, *MEN1* mutations are scattered throughout the 1830-bp coding region of the genes with no evidence for clustering. Correlations between the MEN1 germline mutations and the clinical manifestations of the disorder appear to be absent.[34] Tumors from MEN1 patients and non-MEN1 patients have been observed to harbor the germline mutation together with a somatic loss of heterozygosity (LOH) involving chromosome 11q13, as expected from Knudson's model and the proposed role of the *MEN1* gene as a tumor suppressor.[34]

Menin is a nuclear protein in nondividing cells that interacts with a number of proteins that are involved in transcriptional regulation, genome stability, cell division, and proliferation[34,35] (see Chapter 133).

Multiple Endocrine Neoplasia Type 2 and Type 3. MEN type 2 (MEN2) describes the association of medullary thyroid carcinoma (MTC), pheochromocytomas, and parathyroid tumors.[30,32,33] Three clinical variants of MEN2 are recognized: MEN2a, MEN2b (also referred

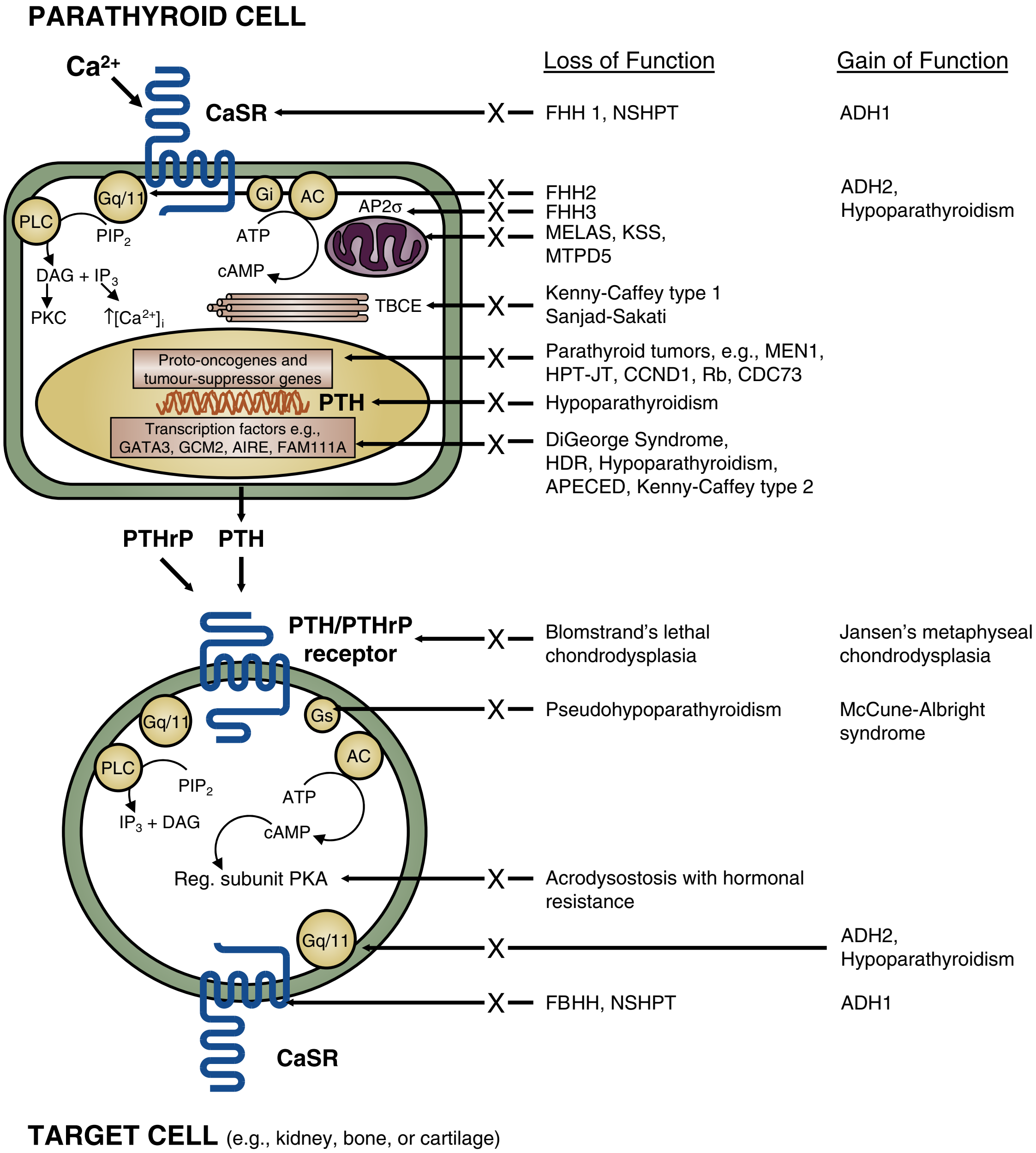

PARATHYROID CELL
Ca2+
CaSR
Loss of Function
Gain of Function
FHH 1, NSHPT
ADH1
Gq/11
Gi
AC
PLC
PIP2
ATP
AP2σ
cAMP
DAG + IP3
PKC
↑[Ca2+]i
TBCE
FHH2
FHH3
MELAS, KSS,
MTPD5
ADH2,
Hypoparathyroidism
Kenny-Caffey type 1
Sanjad-Sakati
Proto-oncogenes and tumour-suppressor genes
PTH
Transcription factors e.g., GATA3, GCM2, AIRE, FAM111A
Parathyroid tumors, e.g., MEN1, HPT-JT, CCND1, Rb, CDC73
Hypoparathyroidism
DiGeorge Syndrome, HDR, Hypoparathyroidism, APECED, Kenny-Caffey type 2
PTHrP
PTH
PTH/PTHrP receptor
Blomstrand's lethal chondrodysplasia
Jansen's metaphyseal chondrodysplasia
Gs
Pseudohypoparathyroidism
McCune-Albright syndrome
Gq/11
AC
PLC
PIP2
ATP
IP3 + DAG
cAMP
Reg. subunit PKA
Acrodysostosis with hormonal resistance
Gq/11
ADH2,
Hypoparathyroidism
FBHH, NSHPT
ADH1
CaSR
TARGET CELL (e.g., kidney, bone, or cartilage)

Fig. 52.2 Schematic Representation of Some of the Components Involved in Calcium Homeostasis. Alterations in extracellular calcium are detected by the calcium-sensing receptor (*CaSR*), which is a 1078–amino acid G protein–coupled receptor (*GPCR*). The PTH/PTHrP receptor, which mediates the actions of PTH and PTHrP, is also a GPCR. Thus, Ca^{++}, PTH, and PTHrP involve signaling pathways requiring an interaction between receptors that specifically recognize these ligands and couple to different G proteins, including $G\alpha s$, $G\alpha i$, and $G\alpha q/\alpha 11$, respectively. $G\alpha s$ couples to the adenylyl cyclase (*AC*), which catalyzes the formation of cAMP from ATP. $G\alpha i$ inhibits AC activity. cAMP stimulates PKA, which phosphorylates cell-specific substrates. Activation of the $G\alpha q/\alpha 11$ pathway stimulates PLC, which catalyzes the hydrolysis of the phosphoinositide (*PIP2*) to inositol triphosphate (*IP3*), which increases intracellular calcium, and diacylglycerol (*DAG*), which activates PKC. These proximal signals modulate downstream pathways, which result in specific physiologic effects. Abnormalities in several genes, which lead to mutations in proteins in these pathways, have been identified in specific disorders of calcium homeostasis (see Table 52.1).

to as MEN3), and MTC-only. MEN2a is the most common variant, and the development of MTC is associated with pheochromocytomas (50% of patients), which may be bilateral, and parathyroid tumors (20% of patients). MEN2b, which represents 5% of all MEN2 cases, is characterized by the occurrence of MTC and pheochromocytoma in association with a Marfanoid habitus, mucosal neuromas, medullated corneal fibers, and intestinal autonomic ganglion dysfunction leading to multiple diverticula and megacolon. Parathyroid tumors do not usually occur in MEN2b. MTC-only is a variant in which medullary thyroid carcinoma is the sole manifestation of the syndrome. The gene causing all three MEN2 variants is located on chromosome -10q11.2, and is the *RET* protooncogene, which encodes a tyrosine kinase receptor with cadherin-like and cysteine-rich extracellular domains, and a tyrosine kinase intracellular domain. Specific mutations of *C-RET* have

TABLE 52.1 Diseases of Calcium Homeostasis and Their Chromosomal Locations

Metabolic Abnormality	Disease	Inheritance	Gene Product	Chromosomal Location
Hypercalcemia	Multiple endocrine neoplasia type 1	Autosomal dominant	MENIN	11q13
	Multiple endocrine neoplasia type 2	Autosomal dominant	RET	10q11.2
	Hereditary hyperparathyroidism and jaw tumors	Autosomal dominant	PARAFIBROMIN	1q25
	Familial isolated primary hyperparathyroidism	Autosomal dominant	MEN1	11q13
			PARAFIBROMIN	1q25
			CaSR	3q21.1
			GCM2	6p24.2
	Hyperparathyroidism	Sporadic	PRAD1/CCND1	11q13
			Retinoblastoma	13q14
	Parathyroid carcinoma	Sporadic	PARAFIBROMIN	1q25
	Familial benign hypercalcemia (FBH)			
	FBH1	Autosomal dominant	CaSR	3q21.1
	FBH2	Autosomal dominant	GNA11	19p13
	FBH3 (FBH_{OK})	Autosomal dominant	AP2S1	19q13
	Neonatal hyperparathyroidism	Autosomal recessive	CaSR	3q21.1
		Autosomal dominant		
	Jansen disease	Autosomal dominant	PTHR1	3p21.3
	Williams syndrome	Autosomal dominant	Elastin, LIMK (and other proteins)	7q11.23
	Infantile hypercalcemia	Autosomal recessive	CYP24A1	20q13.2
		Autosomal recessive	SLC34A1	5q35
	McCune–Albright syndrome	Somatic mutations during early embryonic development?	GNAS	20q13.3
Hypocalcemia	Isolated hypoparathyroidism	Autosomal dominant	PTH, GCM2, GNA11	11p15 [a], 6p24.2, 19p13
		Autosomal recessive	PTH, GCM2	11p15*, 6p24.2
		X-linked recessive	SOX3[b]	Xq26-27
	Hypocalcemic hypercalciuria	Autosomal dominant	CaSR	3q21.1
	Hypoparathyroidism associated with polyglandular autoimmune syndrome	Autosomal recessive	AIRE-1	21q22.3
	Hypoparathyroidism associated with KSS, MELAS, and MTPDS	Maternal	Mitochondrial genome	
	Hypoparathyroidism associated with complex congenital syndromes			
	DiGeorge type 1	Autosomal dominant	TBX1	22q11.2/10p
	DiGeorge type 2	Autosomal dominant	NEBL	10p13-p12
	CHARGE	Autosomal dominant	CHD7,	8q12.1-12.2
			SEMA3E	7q21.11
	HDR syndrome	Autosomal dominant	GATA3	10p14
	Blomstrand lethal chondrodysplasia	Autosomal recessive	PTHR1	3p21.3
	Kenny–Caffey type 1, Sanjad–Sakati	Autosomal recessive	TBCE	1q42.3
	Kenny–Caffey type 2	Autosomal recessive	FAM111A	11q12.1
	Hypocalcemia caused by PTH resistance			
	Pseudohypoparathyroidism (type 1a)	Autosomal dominant parentally imprinted	GNAS exons 1-13	20q13.3
	Pseudohypoparathyroidism (type 1b)	Autosomal dominant parentally imprinted	Deletions upstream of or within GNAS complex locus	20q13.3
	Acrodysostosis with hormonal resistance	Autosomal dominant	PRKAR1A	17q23-q24

CHARGE, coloboma-heart anomaly-choanal atresia-retardation-genital-ear anomalies; *HDR,* hypoparathyroidism, deafness, renal dysplasia; *KSS,* Kearns–Sayre syndrome; *MELAS,* mitochondrial encephalopathy, stroke like episodes and lactic acidosis; *MTPDS,* mitochondrial trifunctional protein-deficiency syndrome; *PTH,* parathyroid hormone

[a] Most mutations of *PTH* gene affect the pre-pro sequence

[b] Deletional-insertional involving Xq27/2p25 located 67 kb downstream of *SOX3*

been identified for each of the three MEN2 variants. Thus, in 95% of patients, MEN2a is associated with mutations of the cysteine-rich extracellular domain, and mutations in codon 634 (Cys→Arg) account for 85% of MEN2a mutations.[33] However, a search for *C-RET* mutations in sporadic non-MEN2a parathyroid adenomas revealed no codon 634 mutations. MTC-only is also associated with missense mutations in the cysteine-rich extracellular domain, and most mutations are in codon 618. However, MEN2b is associated with mutations in codon 918 (Met→Thr) of the intracellular tyrosine kinase domain in 95% of patients. Interestingly, the *C-RET* protooncogene is also involved in the etiology of papillary thyroid carcinomas and in Hirschsprung disease. Mutational analysis of *C-RET* to detect mutations in codons 609, 611, 618, 634, 768, and 804 in MEN2a and MTC-only, and codon 918 in MEN2b, has been used in the diagnosis and management of patients and families with these disorders.[33] Indeed, the utilization of a genetic diagnosis has dramatically improved the outcomes for patients who are at risk for MTC and have undergone prophylactic thyroidectomy. Thus, the 10-year survival in patients with metastatic MTC is approximately 20%, but approximately 90% of young patients with a *C-RET* mutation who had a prophylactic thyroidectomy have no evidence of persistent or recurrent MTC at 7 years after surgery.[33,36]

Multiple Endocrine Neoplasia Type 4. Approximately 5% to 10% of patients with MEN1 do not have mutations of the *MEN1* gene,[32,34] and these patients may have mutations involving other genes. One of these genes is *CDNK1B*, which is located on chromosome 12p13, and encodes the 196–amino acid cyclin-dependent kinase inhibitor (CDKI) p27^{kip1} that regulates the cell cycle. Approximately 3% of these patients with MEN1–associated tumors, such as parathyroid adenomas, pituitary adenomas, and pancreatic NETs in association with gonadal, adrenal, renal, and thyroid tumors have *CDNK1B* mutations, and these patients are referred to as having MEN type 4 (MEN4).[37] To date, over 15 different heterozygous loss-of-function *CDNK1B* mutations, consistent with MEN4 being an autosomal dominant-disorder, have been identified in less than 50 patients with MEN4.[30,37] Germ line *CDNK1B* mutations may also rarely be found in patients with sporadic (i.e., nonfamilial) forms of primary hyperparathyroidism.[38]

The Hyperparathyroidism–Jaw Tumor Syndrome. The HPT-JT syndrome is an autosomal dominant disorder characterized by the development of parathyroid adenomas and carcinomas, and fibroosseous jaw tumors.[30,31] In addition, some patients may also develop uterine tumors and renal abnormalities, which include Wilms tumors, renal cysts, renal hamartomas, renal cortical adenomas, and papillary renal cell carcinomas. Other tumors, including pancreatic adenocarcinomas, testicular mixed germ cell tumors with a major seminoma component, and Hürthle cell thyroid adenomas, have also been reported in some patients.[31] It is important to note that the parathyroid tumors may occur in isolation and without any evidence of jaw tumors, and this may cause confusion with other hereditary hypercalcemic disorders such as MEN1, familial benign hypercalcemia (FBH), which is also referred to as familial hypocalciuric hypercalcemia (FHH), and FIHP.[30] HPT-JT syndrome can be distinguished from FBH, as FBH serum calcium levels are elevated during the early neonatal or infantile period, whereas in HPT-JT such elevations are uncommon in the first decade.[30] In addition, HPT-JT patients, unlike FBH patients, will have associated hypercalciuria. The distinction between HPT-JT patients and MEN1 patients, who have only developed the usual first manifestation of hypercalcemia (>90% of patients), is more difficult and is likely to be influenced by operative and histological findings, as well as by the occurrence of other characteristic lesions in each disorder. It is important to note that HPT-JT patients will usually have single adenomas or a carcinoma, whereas MEN1 patients will often have multiglandular parathyroid disease. The distinction between FIHP and HPT-JT in the absence of jaw tumors is difficult but important, as HPT-JT patients may be at a higher risk of developing parathyroid carcinomas.[30] These distinctions may be helped by the identification of additional features, and a search for jaw tumors and renal, pancreatic, thyroid, and testicular abnormalities may help to identify HPT-JT patients. The jaw tumors in HPT-JT are different from the brown tumors observed in some patients with primary hyperparathyroidism and do not resolve after parathyroidectomy.[31] Indeed, ossifying fibromas of the jaw are an important distinguishing feature of HPT-JT from FIHP, and the occurrence of these may occasionally precede the development of hypercalcemia in HPT-JT patients by several decades. The gene that causes HPT-JT is located on chromosome 1q31.2 and consists of 17 exons that encode a ubiquitously expressed and evolutionary conserved 531–amino acid protein, designated parafibromin.[39] This gene is also referred to as *CDC73* or *HRPT2* (i.e., hyperparathyroidism type 2). Over 60 heterozygous *HRPT2* germline mutations associated with HPT-JT have been reported, and these are scattered throughout the 1593-bp coding region, with the majority (>80%) predicting a functional loss through premature truncation.[30,31] A genotype–phenotype correlation was not apparent from these analyses.[31,39,40] The observation of LOH involving the chromosome 1q21.32 region in HPT-JT–associated tumors indicated that parafibromin may be acting as a tumor suppressor, consistent with Knudson's two-hit hypothesis.[31,39] Similar germline and somatic *HRPT2* mutations have also been found in sporadic parathyroid carcinomas, and the frequency of such mutations is high, ranging from 67% to 100%[31,40]; however, the frequency of *HRPT2* mutations in sporadic parathyroid adenomas is low, at 0% to 4%, indicating that *HRPT2* mutations likely confer an aggressive growth potential to the parathyroid cells.[31,34,39,40] *HRPT2* mutations and allelic imbalances have also been identified in sporadic renal tumors, and a loss or downregulation of *HRPT2* expression has been reported in both breast and gastric cancers.[31] These studies indicate that *HRPT2* and its encoded protein, parafibromin, play a critical role in inherited and sporadic parathyroid cancers, as well as other nonhereditary solid tumors. Parafibromin, which is predominantly a nuclear protein and a component of the polymerase-associated factor-1 (PAF-1) complex, likely has roles in transcriptional events that include histone modification, chromatin remodeling, initiation, and elongation.[31,34,41] Studies of a mouse deleted for *Hrpt2* have revealed that Hrpt2 expression and the parafibromin/PAF complex directly regulate genes (e.g., *H19, IgF1, Ifg3, Igfbp4, Hmga1, Hmga2*, and *Hmga3*) that are involved in cell growth and apoptosis.[42]

Nonsyndromic Forms of Primary Hyperparathyroidism

Familial Isolated Hyperparathyroidism. FIHP may represent an incomplete manifestation of a syndromic form such as MEN1, HPT-JT, or FHH because FIHP caused by germline mutations of the *MEN1*, *HRPT2 (CDC73),* or *CaSR* genes, respectively, has been reported in over 40 kindreds.[30,31,39] The sole occurrence of parathyroid tumors in these syndromic disorders is remarkable, and the mechanisms that determine the altered phenotypic expressions of these mutations remain to be authenticated. However, it is important to note that the genetic etiology of nonsyndromic FIHP in the majority of families remains to be elucidated.[31,34,43] Thus, search in 32 kindreds with nonsyndromic FIHP for mutations of the *MEN1*, *CaSR,* and *HRPT2* genes have revealed that only one family harbored a germline mutation, and this involved the *HRPT2* gene that encodes parafibromin.[43] Activating variants of the glial cells missing 2 (*GCM2*) gene, which encodes a parathyroid-specific transcription factor, have also been identified in some FIHP patients.[44,45] However, studies of other FIHP kindreds have indicated that another locus, referred to as *HRPT3,* on chromosome 2p13.3-p14, is likely to be involved in the etiology of nonsyndromic FIHP.[31] Thus, the genes and their underlying abnormalities that lead to nonsyndromic FIHP remain to be identified.

Nonfamilial (Sporadic) Primary Hyperparathyroidism Owing to Germline Mutations. Nonfamilial (sporadic) forms of PHPT may be associated with germline mutations involving the *MEN1*, *CDC73*, *CaSR*, *CDKIs*, or *PTH* genes. Thus, approximately 10% of patients presenting younger than 45 years of age with nonfamilial (sporadic) PHPT may have a *de novo* germline *MEN1, CDC73*, or *CaSR* mutation,[30-32] and this has implications for their future management, in requiring screening for the occurrence of tumors associated with the specific syndrome, as well as for screening their children who may inherit the germline mutation. Moreover, more than 5% of patients with typically presenting PHPT in the sixth to ninth decades of life with solitary parathyroid adenomas and without family histories of parathyroid, MEN, or other endocrine tumor syndromes may have germline mutations of genes encoding CDKIs; (CDKN1A (p21), CDKN2B (p15), or CDKN2C (p18)), thereby indicating that rare variants of these CDKIs may contribute to the etiology of nonfamilial (sporadic) PHPT, as reduced-penetrance predisposition alleles.[46] In addition, a nonsense mutation, Arg83Stop, has been reported that truncates the secreted PTH after the 52nd amino acid and occurred in association with LOH of the *PTH* locus in a parathyroid adenoma.[47] The patient, who had presented with hypercalcemia and an undetectable serum intact PTH concentration (measured by two-site sandwich assays), showed heterozygosity, with wild-type and mutant alleles present in the peripheral-blood leukocytes, which is consistent with a germline *PTH* mutation, but the parathyroid adenoma had a loss of the wild-type allele, while the mutant allele was retained. Following removal of the parathyroid adenoma, normocalcemia was restored.[47] These findings demonstrate that *PTH* nonsense mutations, which result in truncated forms of PTH not recognizable by standard hormone assays, may be associated with parathyroid adenoma and that endogenously produced N-terminal PTH fragments can be biologically active.

Somatic Abnormalities in Sporadic Parathyroid Tumors. Nonsyndromic sporadic parathyroid tumors may harbor somatic abnormalities of the *MEN1*, cyclin D1 (*CCND1*)/*PTH*, *GCM2*, retinoblastoma (*RB*), and retinoblastoma-interacting zinc finger protein (*RIZ1*) genes; and the LRP5 and Wnt/β-catenin pathway.[31,48] Whole-exome sequence analysis of sporadic parathyroid adenomas has identified that 35% to 50% of these tumors have a somatic mutation involving the *MEN1* gene, indicating that this is a major driver for tumorigenesis. Somatic rearrangements that juxtapose *CCND1*, an important cell-cycle regulator, and the 5′-PTH regulatory elements are found in at least 15% of sporadic parathyroid adenomas; this genetic rearrangement was previously referred to as parathyroid adenoma 1 and is associated with overexpression of cyclin D1. In addition, whole-exome sequencing of parathyroid carcinomas has revealed that greater than 45% harbor somatic *CDC73* mutations, approximately 30% have *CCND1* amplification, and greater than 20% have activation of the PI3K/AKT/mTOR pathway.[49] Allelic deletion of the *RB* gene has also been demonstrated in all parathyroid carcinomas and in 10% of parathyroid adenomas and was accompanied by abnormal staining patterns for the Rb protein in 50% of parathyroid carcinomas, but in no parathyroid adenomas. The Rb protein interacts with RIZ1, and 25% to 35% of parathyroid tumors are reported to have had LOH of the RIZ1 locus or hypermethylation of the RIZ1 promoter region, respectively.[31] Investigation of the Wnt-signaling pathway in parathyroid tumors has revealed that over 85% of adenomas and 100% of tumors from chronic renal failure patients have a shorter *LRP5* transcript, which encodes an in-frame deletion of 142 amino acids (residues 666 to 809) that encompasses the third YWTD beta propeller domain, and that this internally truncated LRP5 receptor is associated with activation of the β-catenin signaling in parathyroid tumors.

Gene Testing for Primary Hyperparathyroidism in Clinical Practice

Genetic testing for germline mutations, which are found in approximately 10% of all PHPT patients, is helpful in clinical practice in several ways, including: (1) confirmation of the clinical diagnosis so that appropriate screening for associated tumors can be undertaken; (2) implementation of appropriate treatment, such as early parathyroidectomy for patients with HPT-JT syndrome because of increased occurrence of parathyroid carcinomas, avoidance of minimally invasive parathyroid surgery in MEN1 patients who generally have multigland disease requiring open neck exploration, early prophylactic thyroidectomy in MEN2/MEN3 patients, and avoidance of surgery in FHH patients; (3) identification of family members who may be asymptomatic but harbor the mutation and therefore require screening for tumor detection and early/appropriate treatment; and (4) identification of the 50% of family members who do not harbor the familial germline mutation and can therefore be reassured and alleviated of the burden of anxiety about developing tumors.[30,32] This latter aspect cannot be overemphasized, as it helps to reduce the cost to individuals and their children, and also to health services in not having to undertake unnecessary biochemical and radiologic investigations.[30,32] Moreover, a notable example of the improvements in patient outcome resulting from genetic testing is provided by the results of prophylactic thyroidectomy in MEN2/MEN3 patients. Thus, the 10-year survival in patients with metastatic MTC is approximately 20%, and prophylactic thyroidectomy to prevent MTC and its metastases is recommended in patients with a *RET* mutation.[33] Prophylactic thyroidectomy has dramatically improved outcomes in MEN2/MEN3 patients, such that approximately 90% of young patients with a *RET* mutation who had prophylactic thyroidectomy have no evidence of persistent or recurrent MTC at 7 years after surgery.[33,36]

Indications for Mutational Analysis in Primary Hyperparathyroidism

Indications for testing for germline mutations in PHPT patients include: (1) PHPT occurring before 45 years of age, (2) multigland disease, (3) parathyroid carcinoma or atypical parathyroid adenomas (e.g., with fibrous bands or cysts), (4) being a first-degree relative of a known mutation carrier, and (5) being an index case with two or more MEN syndrome–associated endocrine tumors.[30,32] Genetic testing should use DNA obtained from leukocytes or salivary cells (i.e., nontumor cells), as DNA from parathyroid tumors is not clinically useful for establishing the diagnosis or staging, because such tumors may contain multiple mutations.[50,51]

Clinical Approach to Gene Testing in a Patient With Primary Hyperparathyroidism

A clinical approach to genetic testing in a patient who has PHPT and no manifestations of MEN-associated tumors or tumors associated with HPT-JT is as follows (Fig. 52.3).[30,32] PHPT patients in whom there is a high suspicion of a genetic etiology (e.g., young age of onset, multigland disease, parathyroid carcinoma, atypical parathyroid adenoma) should be offered genetic counseling and germline mutation testing of the *MEN1, CaSR, AP2S1, GNA11, CDC73, CDKN1A, CDKN1B, CDKN2B, CDKN2C, RET, GCM2*, and *PTH* genes using DNA obtained from nontumor cells (e.g., leukocytes).[30,32,33] Such patients may have *de novo* mutations, which occur in approximately 10% of patients, or they may have an undisclosed family history for the disease.[30,32] Furthermore, PHPT may be the first manifestation of MEN1 and HPT-JT in approximately 90% and greater than 95% of patients, respectively, with these disorders.[30,32] In addition, the clinical distinction between PHPT and FHH may be difficult, as approximately 20% of FHH patients may have elevated serum PTH concentrations and calcium:creatinine clearance of greater than 0.01,[5,30,32,52,53] and some PHPT patients may have hypocalciuria. Moreover, PHPT and familial PHPT (i.e., FIHP) may occasionally be caused by *CaSR* mutations.[5] The identification of a germline mutation should prompt entry into appropriate periodic clinical, biochemical, and radiologic screening programs, for example, for MEN- and HPT-JT–associated tumors.[30,32] The absence of clinical manifestations of hereditary or syndromic forms of PHPT and any genetic abnormalities within these genes would indicate that the likelihood of a MEN-syndrome, HPT-JT, or FHH was low (i.e., <5%).[30,32,33] First-degree relatives of a PHPT patient with a germline mutation should be identified and offered genetic counseling and appropriate gene testing, and individuals who have inherited the mutation should be offered periodic screening, even if asymptomatic.[30,32,33] First-degree relatives who have not inherited the causative mutation require no further follow-up and may be alleviated of the anxiety associated with the prospect of developing MEN- or HPT-JT–associated tumors.[30,32] For PHPT patients who present at a later age, a detailed family history for PHPT (i.e., FIHP), MEN syndrome, HPT-JT, or FHH should be undertaken, and appropriate clinical evaluations and gene testing should then be undertaken to determine the etiology of the PHPT.[30,32,33] Moreover, it is important to note that greater than 5% of patients with nonfamilial (sporadic) PHPT presenting in the sixth to ninth decade of life with solitary parathyroid adenoma may have a germline mutation involving *CDKN1A, CDKN2B*, or *CDKN2C*, thereby implying a higher risk for developing PHPT in first-degree relatives.[46] Thus, genetic assessment should be undertaken in their first-degree relatives for the presence of the rare variants/mutations of *CDKN1A, CDKN2B,* and *CDKN2C,* and those who have inherited the mutation should have periodic screening to detect the onset of hypercalcemia to facilitate appropriate earlier treatment aimed at preventing the skeletal and renal complications of PHPT.[30,32,46]

Familial Hypocalciuric Hypercalcemia and Disorders of the Calcium-Sensing Receptor

Three hypercalcemic disorders attributed to mutations and/or reduced activity of the CaSR have been reported[54]: FHH (FBH), neonatal severe hyperparathyroidism (NSHPT), and autoimmune hypocalciuric hypercalcemia.[5]

KEY POINTS

- Familial hypocalciuric hypercalcemia (FHH) is a genetically heterogeneous disorder, with FHH type 1 (FHH1), type 2 (FHH2), and type 3 (FHH3) owing to loss-of-function mutations of the calcium-sensing receptor, G protein α11 subunit, and adaptor protein 2 sigma subunit, respectively.

Familial Hypocalciuric Hypercalcemia. FHH is an autosomal dominant disorder characterized by the occurrence of lifelong elevations of serum calcium concentrations with low urinary calcium secretion (mean urinary calcium:creatinine clearance ratio <0.01) and inappropriately normal circulating PTH concentrations in 80% of patients.[5,30] FHH is genetically heterogeneous, with three reported variants, FHH1, FHH2, and FHH3, whose loci were mapped to chromosomes 3q21.1, 19p, and 19q13, respectively.

FHH1 is caused by heterozygous loss-of-function mutations of the CaSR, which signals via Gαq and Gα11.[5] Approximately two-thirds of FHH kindreds will have unique heterozygous mutations of the CaSR, and expression studies of these mutations have demonstrated a loss of CaSR function, whereby there is an increase in the calcium ion–dependent set point for PTH release from the parathyroid cell.

FHH2 is caused by loss-of-function mutations of Gα11, which decrease the sensitivity of cells expressing CaSR, probably by impairing the release of GDP.[5,52] Such Gα11 loss-of-function mutations may occur in less than 5% of FHH patients.

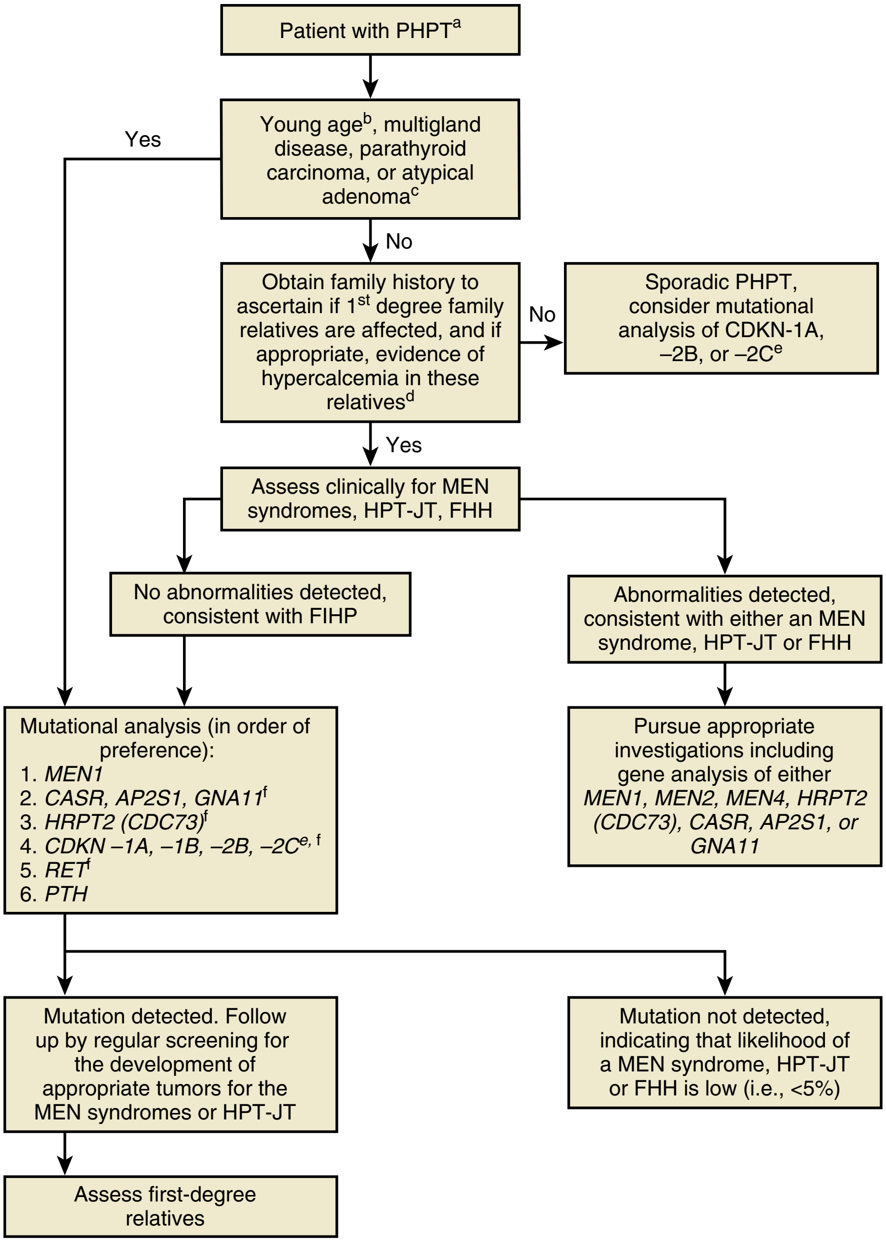
FHH3 is caused by loss-of-function mutations of the adaptor protein 2 (AP-2) sigma subunit (AP2σ).[5,53] AP-2, which is a heterotetramer of α, β, μ, and σ subunits, is a central component of clathrin-coated vesicles (CCVs) and is pivotal in clathrin-mediated endocytosis, that internalizes plasma membrane constituents such as GPCRs. The FHH3-associated AP2σ mutations, which all involve an Arg15 residue that forms key contacts with the dileucine-based motifs of CCV cargo proteins, result in a decreased sensitivity of CaSR-expressing cells to extracellular calcium and reduced CaSR endocytosis. Such AP2σ loss-of-function mutations occur in greater than 5% of FHH patients. FHH1, FHH2, and FHH3 have similar clinical features, and thus studies defining the molecular mechanisms underlying FHH1, FHH2, and FHH3 have helped to distinguish them by genetic analysis, as well as elucidate the CaSR signaling pathway.[5]

Neonatal Severe Primary Hyperparathyroidism. NSHPT occurring in the offspring of consanguineous FHH1 families has been shown to be as a result of homozygous CaSR mutations.[5] However, some patients with sporadic neonatal hyperparathyroidism have been reported to be associated with *de novo* heterozygous CaSR mutations, thereby suggesting that factors other than mutant gene dosage, such as the degree of set-point abnormality, the bony sensitivity to PTH, and the maternal extracellular calcium concentration, may also all play a role in the phenotypic expression of a CaSR mutation in the neonate.[5]

Autoimmune Hypocalciuric Hypercalcemia. Some patients, who have the clinical features of FHH but not *CaSR* mutations, may have AHH.[5,55] Such patients may have multiple clinical autoimmune manifestations, including antithyroid, antigliadin, or antiendomyseal antibodies. These patients were shown to have circulating antibodies to the extracellular domain of the CaSR, and these antibodies stimulated PTH release from dispersed human parathyroid cells *in vitro*, probably by inhibiting activation of the CaSR by extracellular calcium. The effects of treatment with glucocorticoids have been variable, with the hypercalcemia responding in one patient,[55] but not in another.[56] Thus, AHH is a condition of extracellular calcium sensing that should be considered in FHH patients who do not have CaSR mutations.

Other Hereditary Hypercalcemic Disorders

Jansen Disease. Jansen disease (Figs. 52.4 and 52.5) is an autosomal dominant disease that is characterized by abnormal regulation of chondrocyte proliferation and differentiation in the metaphyseal growth plate that leads to the radiographic appearance of severe rickets and adult short stature. During childhood and adolescence, affected patients furthermore present with (usually severe) hypercalcemia and hypophosphatemia, despite normal or undetectable serum levels of PTH or PTHrP, and with hypercalciuria that is associated with nephrocalcinosis; older patients have less severe hypercalcemia, but hypercalciuria persists and contributes to progressive loss of renal function.[57] These abnormalities are caused by mutations in the PTH/PTHrP receptor that lead to constitutive PTH-independent and PTHrP-independent receptor activation. Four different heterozygous mutations affecting three different amino acid residues of the PTH/PTHrP receptor have been identified in the severe form of Jansen disease, and these involve codon 223 (His→Arg), codon 410 (Thr→Pro), and codon 458 (Ile→Arg or Lys). A fifth PTH/PTHrP receptor mutation also affects codon 410 (Thr→Arg) but causes a less severe disease. Expression of the mutant receptors in COS-7 or HEK293 cells resulted in constitutive, ligand-independent accumulation of cAMP, whereas the basal accumulation of inositol phosphates was not measurably increased.[57] Because the PTH/PTHrP receptor is most abundantly expressed in kidney, bone, and the metaphyseal growth plate, these


Patient with PHPT[a]
Young age[b], multigland disease, parathyroid carcinoma, or atypical adenoma[c]
Yes
No
Obtain family history to ascertain if 1st degree family relatives are affected, and if appropriate, evidence of hypercalcemia in these relatives[d]
No
Sporadic PHPT, consider mutational analysis of CDKN-1A, –2B, or –2C[e]
Yes
Assess clinically for MEN syndromes, HPT-JT, FHH
No abnormalities detected, consistent with FIHP
Abnormalities detected, consistent with either an MEN syndrome, HPT-JT or FHH
Mutational analysis (in order of preference):
1. MEN1
2. CASR, AP2S1, GNA11[f]
3. HRPT2 (CDC73)[f]
4. CDKN –1A, –1B, –2B, –2C[e, f]
5. RET[f]
6. PTH
Pursue appropriate investigations including gene analysis of either MEN1, MEN2, MEN4, HRPT2 (CDC73), CASR, AP2S1, or GNA11
Mutation detected. Follow up by regular screening for the development of appropriate tumors for the MEN syndromes or HPT-JT
Mutation not detected, indicating that likelihood of a MEN syndrome, HPT-JT or FHH is low (i.e., <5%)
Assess first-degree relatives

[a] PHPT presenting without manifestations of MEN-associated tumors, or tumors associated with HPT-JT.

[b] Guidelines for MEN1 recommend MEN1 mutational analysis in patients with PHPT occurring before age of 30 years, and ~10% of PHPT patients below the age of 45 years have been reported to have a germline mutation involving the *MEN1, CASR* or *HRPT2* (cdc73) genes.

[c] Atypical parathyroid adenoma may have cysts or fibrous bands.

[d] PHPT may be the first manifestation of MEN1 and HPT-JT in ~90% and ~95% of patients, respectively, with these disorders.

[e] <5% of patients presenting with non-familial (sporadic) and non-syndromic PHPT, due to solitary parathyroid adenomas in the 6th–9th decade of life may have rare variants / mutations of *CDKN1A, CDKN2B,* or *CDKN2C.*

[f] *CASR, AP2S1, GNA11, HRPT2 (CDC73), CDKN1B*, and *RET* mutations are associated with FHH 1, FHH3, FHH2, HPT-JT,MEN4 and MEN2, respectively.

Fig. 52.3 Clinical Approach to Gene Testing in a Patient with Primary Hyperparathyroidism (*PHPT*). *AP2S1*, adaptor protein 2 sigma subunit; *CDKN*, cyclin-dependent kinase inhibitor; *CASR*, calcium sensing receptor;*CDC73*, cell division cycle 73; *FHH*, familial hypocalciuric hypercalcemia; *FIHP*, familial isolated primary hyperparathyroidism; *GCM2*, glial cells missing 2; *GNA11*, G protein alpha 11 subunit; *HPT-JT*, hyperparathyroidism–jaw tumor syndrome; *HRPT2*, hyperparathyroidism type 2; *MEN*, multiple endocrine neoplasia; *PTH*, parathyroid hormone *RET*, rearranged during transfection protooncogene. (Modified from Thakker RV. Familial and hereditary forms of primary hyperparathyroidism. In: Bilezikian JP, Marcus R, Levine M, et al, eds. *The parathyroids: basic and clinical concepts.* 3rd ed. Oxford: Elsevier; 2013:341–363.)

findings provide a likely explanation for the abnormalities observed in mineral homeostasis and growth plate development associated with this disorder. This conclusion is supported further by observations in mice that express the human PTH/PTHrP receptor with the His223Arg mutation under the control of the rat α_1(II) promoter, which resulted in a delayed differentiation of growth plate chondrocytes, abnormal bone elongation, excess trabecular bone mass, and increased bone turnover; bone histomorphometric and biochemical parameters were partially reversed by treatment with an inverse agonist at the mutant PTH/PTHrP receptor.[58]

Williams Syndrome. Williams syndrome is an autosomal dominant disorder characterized by supravalvular aortic stenosis, elfin-like facies, psychomotor retardation, and infantile hypercalcemia. The underlying abnormality causing hypercalcemia, which affects 5% to 50% of patients, remains unknown, but abnormal 1,25-$(OH)_2D$ metabolism or decreased calcitonin production have been implicated, although none have been consistently demonstrated.[59] Studies have demonstrated hemizygosity on chromosome 7q11.23, a region spanning 1.5 to 1.8 million base pairs and containing 26 to 28 genes that include the elastin gene, in over 90% of patients with the classical Williams phenotype. Interestingly, ablation of the elastin gene in mice results in vascular abnormalities similar to those observed in patients with Williams syndrome.[2] However, microdeletions have also been reported to involve the LIM-kinase gene, which is expressed in the central nervous system. Although deletion of the elastin and LIM-kinase genes can explain the respective cardiovascular and neurologic features of Williams syndrome, it seems likely that another, as yet uncharacterized, gene located within this contiguously deleted region is involved in this disorder and could explain the abnormalities of calcium metabolism.[2]

Infantile Hypercalcemia-CYP24A1 and NPT2a Deficiency. Initially referred to as *idiopathic infantile hypercalcemia* (*IIH*), this disorder was shown to be caused by homozygous or compound heterozygous mutations in the gene encoding 24-hydroxylase (CYP24A1), the enzyme that metabolizes and thus inactivates the biologically active 1,25$(OH)_2D$.[2,60] Affected infants develop often severe hypercalcemia and hypercalciuria, leading to nephrocalcinosis and failure to thrive. Besides elevated blood and urinary calcium levels, 1,25$(OH)_2D$ levels are increased, which was shown in one child to dramatically enhance intestinal calcium absorption, thereby suppressing circulating PTH concentrations through two mechanisms, that is, elevated blood calcium and 1,25$(OH)_2D$ concentration.

For several IIH patients, no CYP24A1 mutation could be identified, which led to the discovery of recessive mutations in SLC34A1, the most important sodium-dependent phosphate co-transporter (NPT2a) in the proximal renal tubules.[61] Urinary phosphate-wasting owing to these SLC34A1 mutations increases 1,25-$(OH)_2D$ levels, resulting in symptomatic hypercalcemia. Unlike IHH patients with CYP24A1 mutations, who are treated by reducing vitamin D intake, hypercalcemia in patients with SLC34A1 mutations responds to phosphate Supplementation.

HYPOCALCEMIC DISORDERS

Hypoparathyroidism

Hypoparathyroidism may occur in a syndromic form as part of a pluriglandular autoimmune disorder or as a complex congenital defect (e.g., DiGeorge syndrome). In addition, hypoparathyroidism may develop as a nonsyndromic solitary endocrinopathy, and this has been called *isolated* or *idiopathic hypoparathyroidism*.[1,28] Familial occurrences of isolated hypoparathyroidism with autosomal dominant, autosomal recessive, and X-linked recessive inheritances have been established (Fig. 52.6).

KEY POINTS

- Studies of hypoparathyroidism in humans and mice have revealed that transcriptional cascades involving TBX1, GCM2, GATA3, and SOX3 are likely involved in the development of parathyroid glands from the endoderm of the third pharyngeal pouch. In addition, mutations in the pre-pro portion of parathyroid hormone (PTH) or in the secreted form of PTH can lead to hypocalcemia.

PTH Gene Abnormalities

PTH gene abnormalities affect in most cases the pre-pro portion of PTH; only two mutations have thus far been identified in the secreted

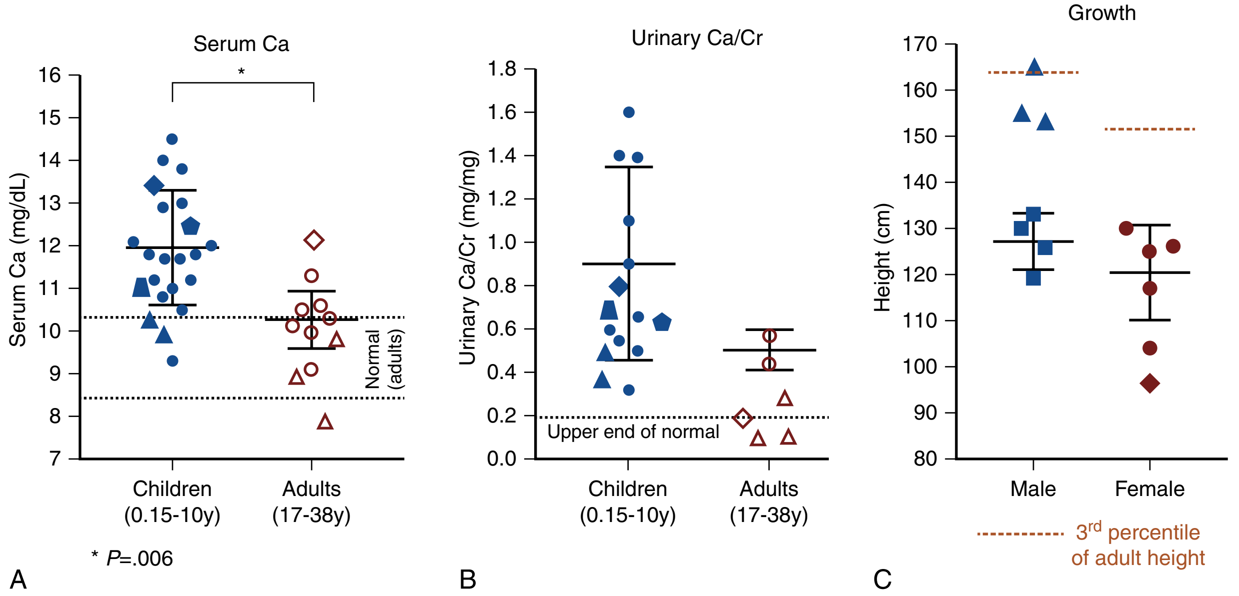


Fig. 52.4 A and B, Serum and urinary calcium measurements for multiple children and adults with Jansen disease attributed to different PTHR1 mutations. C, Adult heights for patients with Jansen disease; the red broken lines indicates the third percentile for normal adult heights.

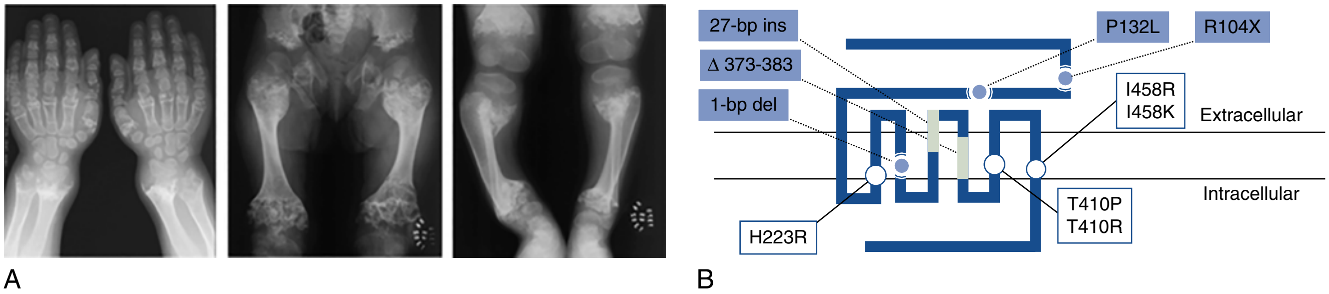


Fig. 52.5 A, Radiographic findings a patient affected by Jansen disease at 6 years of age. B, Schematic representation of the human PTH/PTHrP receptor. The approximate locations of heterozygous missense mutations that lead to constitutive receptor activation in patients with Jansen disease are indicated by *open circles.* Homozygous loss-of-function mutations identified in patients with Blomstrand disease are indicated by *striped boxes* or *gray circles*; the nucleotide exchange in exon M5 of the maternal PTH/ PTHrP receptor allele introduces a novel splice acceptor site that leads to the synthesis of an abnormal receptor protein that lacks portions of the fifth membrane-spanning domain (*stippled box*); for yet unknown reasons, the paternal allele is not expressed in this patient. (From Onuchi L, Ferraz-de-Souza B, Mendonca BB, et al., Potential effects of alendronate on fibroblast growth factor 23 levels and effective control of hypercalciuria in an adult with Jansen's metaphyseal chondrodysplasia. *JCEM.* 2012; 97:1098–1103.)

peptide. The pre-pro missense mutations Met1Cys (which resulted in use of Met7 to initiate translation and a deletion of six amino acids [Met1–Asp6]), Met14Lys, and Cys18Arg[1,28,62] and a donor splice mutation (IVS2 DS, G-C, +1) that resulted in skipping of exon 2 have been reported in patients with isolated hypoparathyroidism. These mutations generally disrupt cleavage of mutant pre-pro PTH to pro PTH, or of mutant pro PTH to the mature PTH, which results in abnormal processing of the mutant peptide in the endoplasmic reticulum (ER), thereby leading to inadequate PTH synthesis or secretion. However, the homozygous mutation Ser32Pro, that is, Ser1Pro of the mature PTH, appears to result in secretion of a PTH molecule, which was detected to be low-normal until mid-adulthood, when it progressively increased to over 100 times the upper limit of normal, that was presumed to be caused by parathyroid hyperplasia, secondary to chronic and appropriate mild undertreatment with calcium and calcitriol and/or chronic mild hyperphosphatemia.[63] In addition, a homozygous PTH mutation affecting a single amino acid residue of the secreted peptide, namely Arg23Cys (Arg56Cys of the pre-pro PTH) is another cause of hypoparathyroidism.[64] The mutation affects the epitope recognized by the detection antibody used in some PTH assays, but not in others, thus leading to descriptions of PTH levels as significantly elevated or undetectable. This further illustrates the complexities in diagnosis of hypoparathyroidism, as well as the value of genetic analysis and knowledge regarding the PTH assay characteristics.

GCM2 Abnormalities

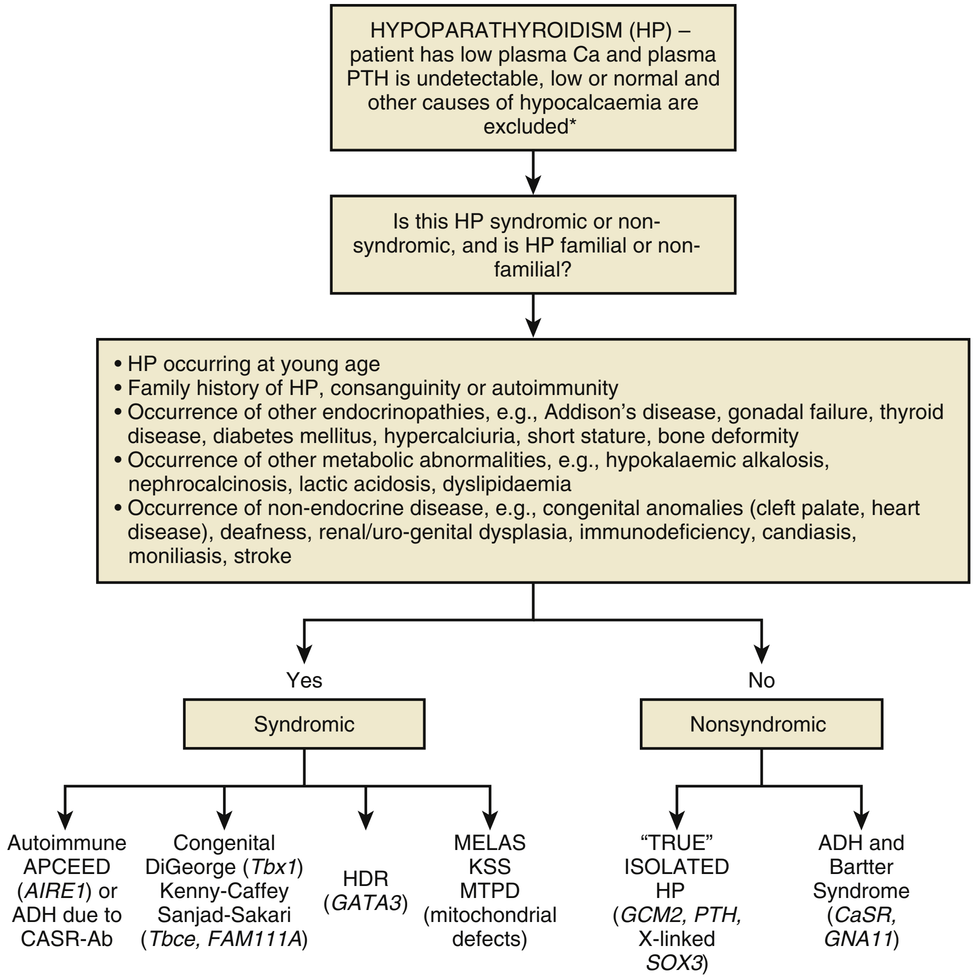
GCM2, also referred to as *GCMB*, is the human homolog of the Drosophila *Gcm* gene and of the mouse *Gcm2* gene. GCM2, which binds to the *PTH* gene 5′ promoter to regulate *PTH* transcription,[1,28] is expressed exclusively in the parathyroid glands, consistent with its role as a specific regulator of parathyroid gland development from the embryonic pharyngeal pouches. Studies of patients with isolated hypoparathyroidism have shown that *GCM2* mutation abnormalities comprising missense and nonsense mutations, and frameshifting and intragenic deletions are associated with autosomal recessive and dominant forms of the disease.[65-68] Studies have shown that these mutations result in a failure of the mutant GCM2 to localize to the nucleus, bind to DNA, or induce transactivation.[66] Moreover, mutant GCMB protein may also be unstable, hyperubiquinated, and degraded via the ubiquitin system.[69]

X-Linked Recessive Hypoparathyroidism

X-linked recessive hypoparathyroidism has been reported in two unrelated kindreds from the United States, and is associated with interstitial deletion-insertions that involve chromosome Xq27.1 downstream of *SOX3* and insertions of predominantly noncoding DNA from chromosome 2p25.3.[70] In this disorder, only males are affected, and they suffer from infantile onset of epilepsy and hypocalcemia, which is caused by an isolated defect in parathyroid gland development. These deletion-insertions are likely to exert a position effect on *SOX3*, expressed in the developing parathyroid glands of mouse embryos, and this indicates a likely role for *SOX3* in the embryonic development of the parathyroid glands.[71]

Pluriglandular Autoimmune Hypoparathyroidism

Hypoparathyroidism may occur in association with candidiasis and autoimmune Addison disease, and the disorder has been referred to as either autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) syndrome or autoimmune polyglandular syndrome (APS1) type 1.[1,28] Genetic studies identified *AIRE1* (autoimmune regulator 1), located on chromosome 21q22.3, to be the causative gene for APECED, *AIRE1* encodes a 545–amino acid protein that contains motifs associated with a transcriptional factor and includes two zinc-finger motifs, a proline-rich region, and three LXXLL motifs. Four *AIRE1* mutations are commonly found in APECED families: Arg257Stop in Finnish, German, Swiss, British, and Northern Italian families; Arg139Stop in Sardinian families; Tyr85Cys in Iranian Jewish families; and a 13-bp deletion in exon 8 in British, Dutch, German, and Finnish families. The AIRE1 protein is mainly localized to the nucleus and mediates E3 ubiquitin ligase activity, which is abolished by missense APECED-causing mutations.[72] AIRE1 has been shown to regulate the elimination of organ-specific T-cells in the thymus, and thus APECED is likely to be caused by a failure of this specialized mechanism for deleting forbidden T-cells and establishing immunologic tolerance.[73] Patients with APS1 may also develop autoantibodies that react with specific autoantigens that are not found in non-APS1 patients; examples of this are autoantibodies to type 1 interferon, which are present in all APS1 patients,[74] and to NACHT leucine-rich-repeat-protein 5 (NALP5), which is a parathyroid-specific autoantibody present in 49% of patients with APS1-associated hypoparathyroidism.[75] NALP proteins are essential components of the inflammasome and activate the innate immune system in different inflammatory and autoimmune disorders, such as vitiligo, which involves NALP1, and


HYPOPARATHYROIDISM (HP) – patient has low plasma Ca and plasma PTH is undetectable, low or normal and other causes of hypocalcaemia are excluded*
Is this HP syndromic or non-syndromic, and is HP familial or non-familial?
• HP occurring at young age
• Family history of HP, consanguinity or autoimmunity
• Occurrence of other endocrinopathies, e.g., Addison's disease, gonadal failure, thyroid disease, diabetes mellitus, hypercalciuria, short stature, bone deformity
• Occurrence of other metabolic abnormalities, e.g., hypokalaemic alkalosis, nephrocalcinosis, lactic acidosis, dyslipidaemia
• Occurrence of non-endocrine disease, e.g., congenital anomalies (cleft palate, heart disease), deafness, renal/uro-genital dysplasia, immunodeficiency, candiasis, moniliasis, stroke
Yes
Syndromic
No
Nonsyndromic
Autoimmune APCEED (*AIRE1*) or ADH due to CASR-Ab
Congenital DiGeorge (*Tbx1*) Kenny-Caffey Sanjad-Sakari (*Tbce, FAM111A*)
HDR (*GATA3*)
MELAS KSS MTPD (mitochondrial defects)
"TRUE" ISOLATED HP (*GCM2, PTH,* X-linked *SOX3*)
ADH and Bartter Syndrome (*CaSR, GNA11*)

*In pseudohypoparathyroidism the plasma PTH is high.

Fig. 52.6 Clinical Approach to Establishing the Genetic Etiology of Hypoparathyroidism. *HP,* hypoparathyroidism; *PTH,* Parathyroid hormone.

gout, which involves NALP3.[76] The precise role of NALP5 in APS1-associated hypoparathyroidism remains to be elucidated.

DiGeorge Syndrome

Patients with DiGeorge syndrome (DGS) typically suffer from hypoparathyroidism, immunodeficiency, congenital heart defects, and deformities of the ear, nose, and mouth.[1,28] The disorder arises from a congenital failure in the development of the derivatives of the third and fourth pharyngeal pouches, with resulting absence or hypoplasia of the parathyroid glands and thymus. Most cases are sporadic, but autosomal dominant inheritance of DGS has been observed, and an association between the syndrome and an unbalanced translocation and deletions involving 22q11.2 have also been reported; this is referred to as DGS type 1 (DGS1). Mapping studies of the DGS1-deleted region on chromosome 22q11.2 defined a 250-kb to 3000-kb critical region that contained approximately 30 genes. However, point mutations in DGS1 patients have only been detected in the *TBX1* gene,[77,78] and *TBX1* is now considered to be the causative gene for DGS1. TBX1 is a DNA-binding transcriptional factor of the T-Box family that is known to have an important role in vertebrate and invertebrate organogenesis and pattern formation.

Some patients may have a late-onset DGS1, and they develop symptomatic hypocalcemia in childhood or during adolescence, with only subtle phenotypic abnormalities. These late-onset DGS1 patients have similar microdeletions in the 22q11 region. It is of interest to note that the age of diagnosis in the families of the three DGS1 patients with inactivating TBX1 mutations ranged from 7 to 46 years, which is in keeping with late-onset DGS1.[77,78]

In some patients, deletions of another locus on chromosome 10p have been observed in association with DGS, and this is referred to as DGS type 2 (DGS2). The *nebulette* (*NEBL*) gene has been reported to be heterozygously deleted in cell lines from two female DGS2 patients, and thus may be the responsible gene.[79]

CHARGE Syndrome

Over 70% of patients with coloboma-heart-anomaly-choanal atresia-retardation-genital-ear anomalies (CHARGE) syndrome have hypocalcemia, likely caused by hypoparathyroidism.[80] CHARGE syndrome may be inherited as an autosomal dominant trait, and in most patients is attributed to *de novo* mutations of the chromodomain helicase DNA-binding protein 7 (CHD7), a transcriptional regulator that binds to enhancer elements in the nucleoplasm, and a minority of patients may have abnormalities involving semaphorin 3E (SEMA3E), which controls cell positioning during embryonic development.[1,28]

Hypoparathyroidism, Deafness, and Renal Anomalies Syndrome

The hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome is an autosomal dominant disorder.[1,28] Patients may have:

asymptomatic hypocalcemia with undetectable or inappropriately normal serum concentrations of PTH; bilateral, symmetrical, sensorineural deafness involving all frequencies; and renal abnormalities that consist mainly of bilateral renal cysts that compress the glomeruli and tubules, or dysplastic kidneys that may be associated with renal impairment. Genetic studies have established that HDR syndrome is caused by haploinsufficiency of GATA3,[81] which belongs to a family of zinc-finger transcription factors that are involved in vertebrate embryonic development. The majority (>75%) of HDR-associated mutations are predicted to result in truncated forms of the GATA3 protein. GATA3 has two zinc fingers, and the C-terminal finger (ZnF2) binds DNA, while the N-terminal finger (ZnF1) stabilizes this DNA binding and interacts with other zinc finger proteins, such as the Friends of GATA (FOG). HDR-associated mutations involving GATA3 ZnF2 or the adjacent basic amino acids result in a loss of DNA binding, while those involving ZnF1 either lead to a loss of interaction with FOG2 ZnFs or altered DNA binding affinity. Each proband and family will generally have its own unique mutation, and there does not appear to be any correlation with the underlying genetic defect and the phenotypic variation (e.g., the presence or absence of renal abnormalities). Over 90% of patients with two or three of the major clinical features of the HDR syndrome (i.e., hypoparathyroidism, deafness, or renal abnormalities) have a *GATA3* mutation[82]; some patients may also have vaginal and uterine abnormalities. The remaining 10% of HDR patients, who do not have a *GATA3* mutation of the coding region, may harbor mutations in the regulatory sequences flanking the *GATA3* gene, or else they may represent heterogeneity. The phenotypes of HDR patients with *GATA3* mutations appear to be similar to those without *GATA3* mutations.

Mitochondrial Disorders Associated With Hypoparathyroidism

Hypoparathyroidism has been reported to occur in three disorders associated with mitochondrial dysfunction: Kearns–Sayre syndrome (KSS), MELAS syndrome, and mitochondrial trifunctional protein deficiency syndrome (MTPDS).[1,28] KSS is characterized by progressive external ophthalmoplegia and pigmentary retinopathy before 20 years of age, and it is often associated with heart block or cardiomyopathy. MELAS syndrome consists of a childhood onset of mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes. In addition, varying degrees of proximal myopathy can be seen in both conditions. Both the KSS and MELAS syndromes have been reported to occur with insulin-dependent diabetes mellitus and hypoparathyroidism. A point mutation in the mitochondrial gene tRNA leucine (UUR) has been reported in one patient with the MELAS syndrome who also suffered from hypoparathyroidism and diabetes mellitus. Large deletions, consisting of 6741 and 6903 base pairs and involving more than 38% of the mitochondrial genome, have been reported in other patients who suffered from KSS, hypoparathyroidism, and sensorineural deafness. Rearrangements and duplication of mitochondrial DNA have also been reported in KSS. MTPDS is a disorder of fatty-acid oxidation that is associated with peripheral neuropathy, pigmentary retinopathy, and acute fatty liver degeneration in pregnant women who carry an affected fetus. Hypoparathyroidism has been observed in one patient with trifunctional protein deficiency. The role of these mitochondrial mutations in the etiology of hypoparathyroidism remains to be further elucidated.

Kenny–Caffey, Sanjad–Sakati, Kirk–Richardson, and Additional Familial Syndromes

Hypoparathyroidism has been reported to occur in over 50% of patients with Kenny–Caffey syndrome (KCS), which is associated with short stature, osteosclerosis and cortical thickening of the long bones, delayed closure of the anterior fontanel, basal ganglia calcification, nanophthalmos, and hyperopia.[1,28] There are two types of KCS, with KCS1 and KCS2 being autosomal recessive and dominant traits, respectively, and with KCS2 patients not being affected with mental retardation. In Kirk–Richardson and Sanjad–Sakati syndromes, which are similar and found in patients of consanguineous parents of Middle Eastern origin, hypoparathyroidism is associated with severe growth failure and dysmorphic features. Molecular genetic investigations have identified that mutations in the tubulin-specific chaperone (*TBCE*) gene are associated with KCS1 and Sanjad-Sakati syndromes.[83] *TBCE* encodes one of several chaperone proteins required for the proper folding of α-tubulin subunits and the formation of α-β tubulin heterodimers (see Fig. 52.2).[83] KCS2 is caused by heterozygous mutations of the family with sequence 111 member A (*FAM111A*) gene, which encodes a chromatin-associated protein involved in DNA replication.

Other familial syndromes in which hypoparathyroidism is a component have been reported, and the causative genetic abnormalities remain to be identified.

Calcium-Sensing Receptor Abnormalities

CaSR abnormalities are associated with three hypocalcemic disorders: autosomal dominant hypocalcemia type 1 (ADH-1), Bartter syndrome type 5 (i.e., ADH with a Bartter-like syndrome), and a form of autoimmune hypoparathyroidism attributed to CaSR autoantibodies.[5]

Autosomal Dominant Hypocalcemia Type 1. ADH1, which is caused by gain-of-function CaSR mutations, is characterized by hypocalcemia that is generally associated with normal serum intact PTH concentrations and hypomagnesemia, and sometimes hypercalciuria.[1,5,28] However, treatment with vitamin D or its active metabolites to correct the hypocalcemia results in marked hypercalciuria, nephrocalcinosis, nephrolithiasis, and renal impairment.

Bartter Syndrome Type 5. Bartter syndrome is a heterozygous group of electrolyte disorders characterized by hypokalemic alkalosis, renal salt wasting that may lead to hypotension, hyperreninemic hyperaldosteronism, increased urinary prostaglandin excretion, and hypercalciuria with nephrocalcinosis.[14] Mutations of ion transporters and channels are associated with Bartter syndrome, and five types are recognized. Thus, type 1 is caused by mutations involving the bumetanide-sensitive sodium-potassium-chloride co-transporter (NKCC2 or SLC12A2); type 2 is caused by mutations of the outwardly rectifying renal potassium channel (ROMK); type 3 is caused by mutations of the voltage-gated chloride channel (CLC-Kb); type 4 is caused by mutations of Barttin, which is a beta subunit that is required for trafficking of CLC-Kb and CLC-Ka (this form is also associated with deafness, as Barttin, CLC-Ka, and CLC-Kb are also expressed in the marginal cells of the scala media of the inner ear that secrete potassium ion-rich endolymph); and type 5 is caused by activating mutations of the CaSR. Patients with Bartter syndrome type 5 have the classical features of the syndrome (i.e., hypokalemic metabolic alkalosis, hyperreninemia, and hyperaldosteronism).[15,84] In addition, they develop hypocalcemia, which may be symptomatic and lead to carp-pedal spasm, and an elevated fractional excretion of calcium that may be associated with nephrocalcinosis.[15,84]

Autoimmune Hypoparathyroidism. Some 20% of patients who had hypoparathyroidism in association with autoimmune hypothyroidism were found to have autoantibodies to the extracellular domain of the CaSR.[5,85] The CaSR autoantibodies persisted for more than 5 years in only 14% of these patients, and a few of them also had features of APS1.

Gα11 Abnormalities and Autosomal Dominant Hypocalcemia Type 2

ADH2, which is caused by gain-of-function Gα11 mutations, has clinical features that are indistinguishable from those of ADH1.[1,5,28,52]

Blomstrand Disease

Blomstrand chondrodysplasia is an autosomal recessive disorder characterized by early lethality, dramatically advanced bone maturation, and accelerated chondrocyte differentiation.[1,28,86] Affected infants are typically born to consanguineous healthy parents (only in one instance did unrelated healthy parents have two affected offspring), show pronounced hyperdensity of the entire skeleton (Fig. 52.7) and markedly advanced ossification, and have long bones that are extremely short and poorly modeled. PTH/PTHrP receptor abnormalities comprising missense (Pro132Leu) and nonsense (Arg104Stop) mutations, a 1-bp deletion and a 27-bp insertion that resulted in a frameshift and a truncated protein, respectively, and a nucleotide exchange that introduced a novel splice acceptor site have been reported (Fig. 52.5). All of these mutations impaired the function of the PTH/PTHrP receptor, thereby indicating that they are the cause of Blomstrand disease. Fetuses with Blomstrand disease are likely to have secondary hyperplasia of the parathyroid glands, presumably attributed to hypocalcemia, and also have abnormal breast development and tooth impaction, highlighting the involvement of the PTH/PTHrP receptor in the normal development of breast and tooth.[87] Moreover, heterozygous PTH/PTHrP receptor mutations leading to haploinsufficiency or functionless receptors also cause the nonsyndromic isolated autosomal dominant disorder of primary failure of tooth eruption.

CONCLUSION

Considerable advances have been made in identifying key proteins that are involved, either directly or indirectly, in the regulation of PTH synthesis or secretion, and in mediating its hormonal actions in the different target tissues. The identification of mutations in several of these proteins furthermore provided a plausible molecular explanation for a variety of familial and sporadic disorders of mineral ion homeostasis and/or bone development. In addition to these advances in further defining the biological role(s) of known proteins, genetic loci and/or candidate genes have been identified for multiple inherited disorders, and these have provided further important insights into the regulation of body calcium.

Acknowledgments

Rajesh Thakker is supported by the Wellcome Trust and National Institute of Health (NIHR) Oxford Biomedical Research Centre Programme, and Harald Jüppner is supported by grants from the NIH and NIDDK (RO1-DK-46718 and PO1-DK-11794). The authors thank Tracey Walker and Diane Centofanti for their expert assistance in preparing this chapter.

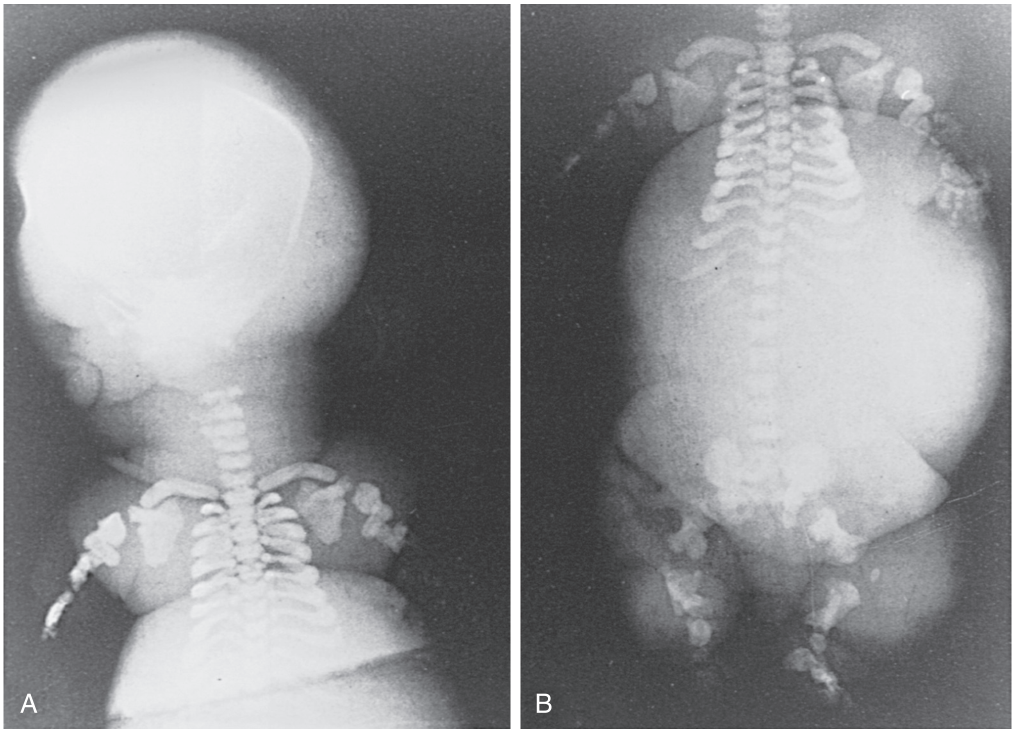


Fig. 52.7 Radiologic Findings in a Patient with Blomstrand Disease. Note the markedly advanced ossification of all skeletal elements, and the extremely short limbs, despite the comparatively normal size and shape of hands and feet. Furthermore, note that the clavicles are relatively long and abnormally shaped. (From Leroy JG, Keersmaeckers G, Coppens M, et al. Blomstrand lethal chondrodysplasia. *Am J Med Genet.* 1996;63:84–89.)

REFERENCES

1. Mannstadt M, Bilezikian JP, Thakker RV, et al. Hypoparathyroidism. *Nat Rev Dis Primers*. 2017;3:17080.
2. Stokes VJ, Nielsen MF, Hannan FM, Thakker RV. Hypercalcemic Disorders in Children. *J Bone Miner Res*. 2017;32:2157–2170.
3. Hannan FM, Thakker RV. Investigating hypocalcaemia. *BMJ*. 2013;346:f2213.
4. Kovacs CS. Fetal mineral homeostasis. In: Glorieux F, ed. *Pediatric Bone: Biology and Diseases*. San Diego: Academic Press; 2012:247–275.
5. Hannan FM, Kallay E, Chang W, et al. The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. *Nat Rev Endocrinol*. 2018;15:33–51.
6. Peng JB, Chen XZ, Berger UV, et al. Molecular cloning and characterization of a channel-like transporter mediating intestinal calcium absorption. *J Biol Chem*. 1999;274:22739–22746.
7. Gunthorpe MJ, Benham CD, Randall A, et al. The diversity in the vanilloid (TRPV) receptor family of ion channels. *Trends Pharmacol Sci*. 2002;23:183–191.
8. Nilius B, Vennekens R, Prenen J, et al. The single pore residue Asp542 determines Ca2+ permeation and Mg2+ block of the epithelial Ca2+ channel. *J Biol Chem*. 2001;276:1020–1025.
9. den Dekker E, Hoenderop JG, Nilius B, et al. The epithelial calcium channels, TRPV5 & TRPV6: from identification towards regulation. *Cell Calcium*. 2003;33:497–507.
10. Wood RJ, Tchack L, Taparia S. 1,25-Dihydroxyvitamin D3 increases the expression of the CaT1 epithelial calcium channel in the Caco-2 human intestinal cell line. *BMC Physiol*. 2001;1:11.
11. Van Cromphaut SJ, Dewerchin M, Hoenderop JG, et al. Duodenal calcium absorption in vitamin D receptor-knockout mice: functional and molecular aspects. *Proc Natl Acad Sci U S A*. 2001;98:13324–13329.
12. Van Cromphaut SJ, Rummens K, Stockmans I, et al. Intestinal calcium transporter genes are upregulated by estrogens and the reproductive cycle through vitamin D receptor-independent mechanisms. *J Bone Miner Res*. 2003;18:1725–1736.
13. Konrad M, Schaller A, Seelow D, et al. Mutations in the tight-junction gene claudin 19 (CLDN19) are associated with renal magnesium wasting, renal failure, and severe ocular involvement. *Am J Hum Genet*. 2006;79:949–957.
14. Howles SA, Thakker RV. Genetics of kidney stone disease. *Nat Rev Urol*. 2020;17:407–421.
15. Vargas-Poussou R, Huang C, Hulin P, et al. Functional characterization of a calcium-sensing receptor mutation in severe autosomal dominant hypocalcemia with a Bartter-like syndrome. *J Am Soc Nephrol*. 2002;13:2259–2266.
16. Reilly RF, Ellison DH. Mammalian distal tubule: physiology, pathophysiology, and molecular anatomy. *Physiol Rev*. 2000;80:277–313.
17. Loffing J, Kaissling B. Sodium and calcium transport pathways along the mammalian distal nephron: from rabbit to human. *Am J Physiol Renal Physiol*. 2003;284:F628–F643.
18. Hoenderop JG, van Leeuwen JP, van der Eerden BC, et al. Renal Ca2+ wasting, hyperabsorption, and reduced bone thickness in mice lacking TRPV5. *J Clin Invest*. 2003;112:1906–1914.
19. Lee CT, Huynh VM, Lai LW, et al. Cyclosporine A-induced hypercalciuria in calbindin-D28k knockout and wild-type mice. *Kidney Int*. 2002;62:2055–2061.
20. Silva IV, Cebotaru V, Wang H, et al. The ClC-5 knockout mouse model of Dent's disease has renal hypercalciuria and increased bone turnover. *J Bone & Min Res*. 2003;18:615–623.
21. Gorvin CM, Wilmer MJ, Piret SE, et al. Receptor-mediated endocytosis and endosomal acidification is impaired in proximal tubule epithelial cells of Dent disease patients. *Proc Natl Acad Sci U S A*. 2013;110:7014–7019.
22. Linglart A, Menguy C, Couvineau A, et al. Recurrent PRKAR1A mutation in acrodysostosis with hormone resistance. *N Engl J Med*. 2011;364:2218–2226.
23. Hemmingsen C. Regulation of renal calbindin-D28K. *Pharmacol Toxicol*. 2000;87:5–30.
24. Hoenderop JG, Nilius B, Bindels RJ. Molecular mechanism of active Ca2+ reabsorption in the distal nephron. *Annu Rev Physiol*. 2002;64:529–549.
25. Hoenderop JG, Muller D, Van Der Kemp AW, et al. Calcitriol controls the epithelial calcium channel in kidney. *J Am Soc Nephrol*. 2001;12:1342–1349.
26. Van Abel M, Hoenderop JG, Dardenne O, et al. 1,25-dihydroxyvitamin D(3)-independent stimulatory effect of estrogen on the expression of ECaC1 in the kidney. *J Am Soc Nephrol*. 2002;13:2102–2109.
27. Charbonneau A, Leclerc M, Brunette MG. Effect of angiotensin II on calcium reabsorption by the luminal membranes of the nephron. *Am J Physiol Endocrinol Metab*. 2001;280:E928–E936.
28. Clarke BL, Brown EM, Collins MT, et al. Epidemiology and diagnosis of hypoparathyroidism. *J Clin Endocrinol Metab*. 2016;101:2284–2299.
29. Kronenberg H. Developmental regulation of the growth plate. *Nature*. 2003;423:332–336.
30. Eastell R, Brandi ML, Costa AG, et al. Diagnosis of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3570–3579.
31. Cardoso L, Stevenson M, Thakker RV. Molecular genetics of syndromic and non-syndromic forms of parathyroid carcinoma. *Hum Mutat*. 2017;38:1621–1648.
32. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990–3011.
33. Kloos RT, Eng C, Evans DB, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid*. 2009;19:565–612.
34. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat*. 2008;29:22–32.
35. Frost M, Lines KE, Thakker RV. Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol*. 2018;14:216–227.
36. Skinner MA, Moley JA, Dilley WG, et al. Prophylactic thyroidectomy in multiple endocrine neoplasia type 2A. *N Engl J Med*. 2005;353:1105–1113.
37. Frederiksen A, Rossing M, Hermann P, et al. Clinical features of multiple endocrine neoplasia type 4: novel pathogenic variant and review of published cases. *J Clin Endocrinol Metab*. 2019;104:3637–3646.
38. Costa-Guda J, Marinoni I, Molatore S, et al. Somatic mutation and germline sequence abnormalities in CDKN1B, encoding p27Kip1, in sporadic parathyroid adenomas. *J Clin Endocrinol Metab*. 2011;96:E701–E706.
39. Carpten JD, Robbins CM, Villablanca A, et al. HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet*. 2002;32:676–680.
40. Shattuck TM, Valimaki S, Obara T, et al. Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma. *N Engl J Med*. 2003;349:1722–1729.
41. Rozenblatt-Rosen O, Nagaike T, Francis JM, et al. The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A*. 2009;106:755–760.
42. Wang P, Bowl MR, Bender S, et al. Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. *Mol Cell Biol*. 2008;28:2930–2940.
43. Simonds WF, Robbins CM, Agarwal SK, et al. Familial isolated hyperparathyroidism is rarely caused by germline mutation in HRPT2, the gene for the hyperparathyroidism-jaw tumor syndrome. *J Clin Endocrinol Metab*. 2004;89:96–102.
44. Guan B, Welch JM, Sapp JC, et al. GCM2-activating mutations in familial isolated hyperparathyroidism. *Am J Hum Genet*. 2016;99:1034–1044.
45. Song A, Yang Y, Wang Y, et al. Germline Gcm2 mutation screening in chinese primary hyperparathyroidism patients. *Endocr Pract*. 2020;26:1093–1104.
46. Costa-Guda J, Soong CP, Parekh VI, et al. Germline and somatic mutations in cyclin-dependent kinase inhibitor genes CDKN1A, CDKN2B, and CDKN2C in sporadic parathyroid adenomas. *Horm Cancer*. 2013;4:301–307.
47. Au AY, McDonald K, Gill A, et al. PTH mutation with primary hyperparathyroidism and undetectable intact PTH. *N Engl J Med*. 2008;359:1184–1186.
48. Mannstadt M, Holick E, Zhao W, et al. Mutational analysis of GCMB, a parathyroid-specific transcription factor, in parathyroid adenoma of primary hyperparathyroidism. *J Endocrinol*. 2011;210:165–171.

49. Pandya C, Uzilov AV, Bellizzi J, et al. Genomic profiling reveals mutational landscape in parathyroid carcinomas. *JCI Insight*. 2017;2:e92061.
50. Cromer MK, Starker LF, Choi M, et al. Identification of somatic mutations in parathyroid tumors using whole-exome sequencing. *J Clin Endocrinol Metab*. 2012;97:E1774–E1781.
51. Newey PJ, Nesbit MA, Rimmer AJ, et al. Whole-exome sequencing studies of nonhereditary (sporadic) parathyroid adenomas. *J Clin Endocrinol Metab*. 2012;97:E1995–E2005.
52. Nesbit MA, Hannan FM, Howles SA, et al. Mutations affecting G-protein subunit alpha11 in hypercalcemia and hypocalcemia. *N Engl J Med*. 2013;368:2476–2486.
53. Nesbit MA, Hannan FM, Howles SA, et al. Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3. *Nat Genet*. 2013;45:93–97.
54. Brown EM, MacLeod RJ. Extracellular calcium sensing and extracellular calcium signaling. *Physiol Rev*. 2001;81:239–297.
55. Pallais J, Kifor O, Chen Y, et al. Acquired hypocalciuric hypercalcemia due to autoantibodies against the calcium-sensing receptor. *N Engl J Med*. 2004;351:362–369.
56. Pallais JC, Kemp EH, Bergwitz C, et al. Autoimmune hypocalciuric hypercalcemia unresponsive to glucocorticoid therapy in a patient with blocking autoantibodies against the calcium-sensing receptor. *J Clin Endocrinol Metab*. 2011;96:672–680.
57. Saito H, Noda H, Gatault P, et al. Progression of mineral ion abnormalities in patients with jansen metaphyseal chondrodysplasia. *J Clin Endocrinol Metab*. 2018;103:2660–2669.
58. Noda H, Guo J, Khatri A, et al. An inverse agonist ligand of the PTH receptor partially rescues skeletal defects in a mouse model of Jansen's metaphyseal chondrodysplasia. *J Bone Miner Res*. 2020;35:540–549.
59. Pober BR. Williams-Beuren syndrome. N Engl J Med.362:239–252.
60. Schlingmann KP, Kaufmann M, Weber S, et al. Mutations in CYP24A1 and idiopathic infantile hypercalcemia. *N Engl J Med*. 2011;365:410–421.
61. Schlingmann KP, Ruminska J, Kaufmann M, et al. Autosomal-recessive mutations in SLC34A1 encoding sodium-phosphate cotransporter 2a cause idiopathic infantile hypercalcemia. *J Am Soc Nephrol*. 2016;27:604–614.
62. Lee JH, Davaatseren M, Lee S. Rare PTH gene mutations causing parathyroid disorders: a review. *Endocrinol Metab (Seoul)*. 2020;35:64–70.
63. Gild ML, Bullock M, Luxford C, et al. Congenital hypoparathyroidism associated with elevated circulating nonfunctional parathyroid hormone due to novel PTH mutation. *J Clin Endocrinol Metab*. 2020;105:dgaa279.
64. Lee S, Mannstadt M, Guo J, et al. A homozygous [Cys25]PTH(1-84) mutation that impairs PTH/PTHrP receptor activation defines a novel form of hypoparathyroidism. *J Bone Miner Res*. 2015;30:1803–1813.
65. Mannstadt M, Bertrand G, Muresan M, et al. Dominant-negative GCMB mutations cause an autosomal dominant form of hypoparathyroidism. *J Clin Endocrinol Metab*. 2008;93:3568–3576.
66. Bowl MR, Mirczuk SM, Grigorieva IV, et al. Identification and characterization of novel parathyroid-specific transcription factor Glial Cells Missing Homolog B (GCMB) mutations in eight families with autosomal recessive hypoparathyroidism. *Hum Mol Genet*. 2010;19:2028–2038.
67. Mirczuk SM, Bowl MR, Nesbit MA, et al. A missense glial cells missing homolog B (GCMB) mutation, Asn502His, causes autosomal dominant hypoparathyroidism. *J Clin Endocrinol Metab*. 2010;95:3512–3516.
68. Yi HS, Eom YS, Park IB, et al. Identification and characterization of C106R, a novel mutation in the DNA-binding domain of GCMB, in a family with autosomal-dominant hypoparathyroidism. *Clin Endocrinol*. 2012;76:625–633.
69. Xi X, Lu L, Zhuge CC, et al. The hypoparathyroidism-associated mutation in Drosophila Gcm compromises protein stability and glial cell formation. *Sci Rep*. 2017;7:39856.
70. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*. 2015;47:717–726.
71. Bowl M, Nesbit M, Harding B, et al. An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. *J Clin Invest*. 2005;115:2822–2831.
72. Uchida D, Hatakeyama S, Matsushima A, et al. AIRE functions as an E3 ubiquitin ligase. *J Exp Med*. 2004;199:167–172.
73. Liston A, Lesage S, Wilson J, et al. Aire regulates negative selection of organ-specific T cells. *Nat Immunol*. 2003;4:350–354.
74. Meager A, Visvalingam K, Peterson P, et al. Anti-interferon autoantibodies in autoimmune polyendocrinopathy syndrome type 1. *PLoS Med*. 2006;3:e289.
75. Alimohammadi M, Bjorklund P, Hallgren A, et al. Autoimmune polyendocrine syndrome type 1 and NALP5, a parathyroid autoantigen. *N Engl J Med*. 2008;358:1018–1028.
76. Eisenbarth SC, Colegio OR, O'Connor W, et al. Crucial role for the Nalp3 inflammasome in the immunostimulatory properties of aluminium adjuvants. *Nature*. 2008;453:1122–1126.
77. Yagi H, Furutani Y, Hamada H, et al. Role of TBX1 in human del22q11.2 syndrome. *Lancet*. 2003;362:1366–1373.
78. Li D, Gordon CT, Oufadem M, et al. Heterozygous mutations in TBX1 as a cause of isolated hypoparathyroidism. *J Clin Endocrinol Metab*. 2018;103:4023–4032.
79. Villanueva MP, Aiyer AR, Muller S, et al. Genetic and comparative mapping of genes dysregulated in mouse hearts lacking the Hand2 transcription factor gene. *Genomics*. 2002;80:593–600.
80. Jyonouchi S, McDonald-McGinn DM, Bale S, et al. CHARGE (coloboma, heart defect, atresia choanae, retarded growth and development, genital hypoplasia, ear anomalies/deafness) syndrome and chromosome 22q11.2 deletion syndrome: a comparison of immunologic and nonimmunologic phenotypic features. *Pediatrics*. 2009;123:e871–e877.
81. Van Esch H, Groenen P, Nesbit MA, et al. GATA3 haplo-insufficiency causes human HDR syndrome. *Nature*. 2000;406:419–422.
82. Ali A, Christie PT, Grigorieva IV, et al. Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet*. 2007;16:265–275.
83. Parvari R, Hershkovitz E, Grossman N, et al. Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. *Nat Genet*. 2002;32:448–452.
84. Watanabe S, Fukumoto S, Chang H, et al. Association between activating mutations of calcium-sensing receptor and Bartter's syndrome. *Lancet*. 2002;360:692–694.
85. Tomar N, Gupta N, Goswami R. Calcium-sensing receptor autoantibodies and idiopathic hypoparathyroidism. *J Clin Endocrinol Metab*. 2013;98:3884–3891.
86. Hoogendam J, Farih-Sips H, Wynaendts LC, et al. Novel mutations in the parathyroid hormone (PTH)/PTH-related peptide receptor type 1 causing Blomstrand osteochondrodysplasia types I and II. *J Clin Endocrinol Metab*. 2007;92:1088–1095.
87. Wysolmerski JJ, Cormier S, Philbrick W, et al. Absence of functional type 1 PTH/PTHrP receptors in humans is associated with abnormal breast development and tooth impactation. *J Clin Endocrinol Metab*. 2001;86:1788–1794.

53

Genetic Disorders of Phosphate Homeostasis

Kenneth E. White, Erik A. Imel, and Michael J. Econs

OUTLINE

Regulation of Phosphate Homeostasis, 907
Regulation of Phosphate Absorption, 907
  Phosphate Excretion, 907
  Regulation of Phosphate Reabsorption, 908
  Fibroblast Growth Factor-23, 908
  Fibroblast Growth Factor-23 Activity, 908
  Regulation of Fibroblast Growth Factor-23 Production, 908
Fibroblast Growth Factor-23–Associated Syndromes, 909
  Disorders Associated with Increased FGF23 Bioactivity, 909
  Disorders Associated With Reduced Fibroblast Growth Factor-23 Bioactivity, 916
  Genetic Hypophosphatemia Not Mediated by Elevated FGF23, 917
Possible Future Therapeutic Avenues, 919

The identification of genetic disorders associated with altered phosphate handling has been essential for understanding the mechanisms controlling phosphate homeostasis. Elevated fibroblast growth factor-23 (FGF23) is associated with syndromes manifested by hormonally mediated hypophosphatemia with paradoxically low or normal $1,25(OH)_2$ vitamin D ($1,25(OH)_2D$) and include autosomal dominant hypophosphatemic rickets (ADHR), X-linked hypophosphatemic rickets (XLH), tumor-induced osteomalacia (TIO), and several forms of autosomal recessive hypophosphatemic rickets (ARHR). Other phosphate-wasting diseases result directly in impaired phosphate reabsorption in the nephron. Heritable disorders of hyperphosphatemia and elevated $1,25(OH)_2D$, such as hyperphosphatemic familial tumoral calcinosis (hfTC)/hyperostosis-hyperphosphatemia syndrome (HHS), are associated with reduced FGF23 activity. These collective findings have provided unique insight into the activity of FGF23 on renal phosphate and vitamin D metabolism, as well as the pathophysiology of various disorders. This chapter reviews the clinical manifestations of these disorders, their pathophysiology, current treatment, and strategies for developing new therapies.

REGULATION OF PHOSPHATE HOMEOSTASIS

In contrast to calcium, phosphate is widely distributed in nonosseous tissues, both in inorganic form and as a component of numerous organic molecules ranging from nucleic acids and membrane phospholipids to small phosphoproteins and intermediates of carbohydrate metabolism. These soft-tissue phosphates nevertheless compose only approximately 15% of the total body content, with the remainder of inorganic phosphate being in the mineral phase of bone.

Phosphate in serum exists almost exclusively in association with cations. Unlike calcium, only 12% of phosphate is protein-bound.[1] Also in contrast to calcium, serum phosphate concentrations may vary substantially throughout the day. Carbohydrate ingestion may markedly reduce serum phosphate by moving phosphate from the extracellular to the intracellular space. Moreover, serum phosphate undergoes diurnal variation of as much as 1.5 mg/dL (0.5 mmol/L), with a nadir between 8:00 and 11:00 AM.[2] Hypertriglyceridemia,[3] hypergammaglobulinemia,[4] or mannitol therapy[5] may interfere with measurement of serum phosphate. Serum phosphate concentrations are relatively high in the newborn (5–7 mg/dL)[6] and decrease gradually through childhood to adult levels.

Phosphate is an important substrate for intracellular enzymes involved in glycolysis and respiration that synthesize high-energy phosphate bonds for storage of chemical energy in organophosphate compounds such as adenosine triphosphate (ATP), creatine phosphate, diphosphoglycerate, phosphoenolpyruvate, and others. Severe phosphate depletion leads to inhibition of glycolysis and decreased production of ATP. Adequate extracellular phosphate is required for normal mineralization of bone and cartilage,[7] and chronic hypophosphatemia leads to osteomalacia and, in children, rickets.

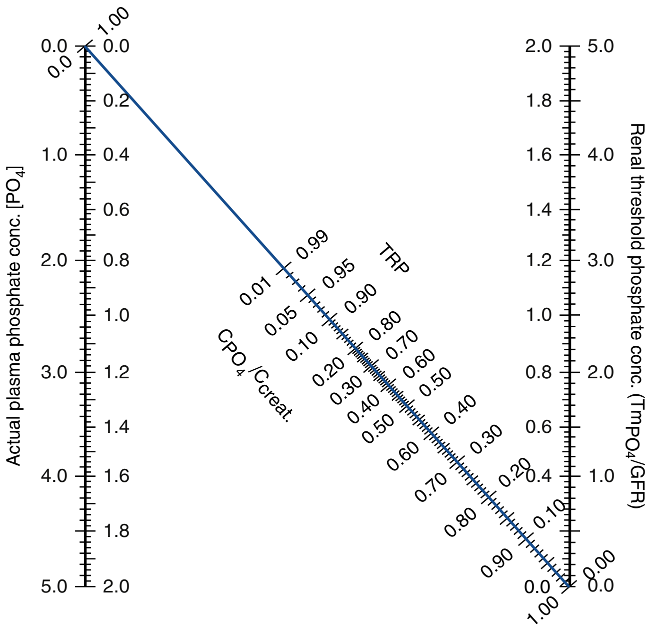
REGULATION OF PHOSPHATE ABSORPTION

The average dietary intake of phosphate—derived largely from dairy products, cereals, and meats—is roughly twice the estimated minimum requirement of 400 mg/day.[8] Absorptive efficiency is high, averaging approximately 70%, and may increase further if the intake of dietary phosphate decreases.[9] Phosphate is avidly absorbed throughout the small intestine, but especially in the jejunum.[10] Sodium-dependent active phosphate absorption is mediated by the NPT2b cotransporters present in the luminal brush-border membranes of enterocytes.[11] Absorption of phosphate is strikingly augmented by $1,25(OH)_2D$ through increased expression of NPT2b.[12]

Phosphate Excretion

Renal tubular reabsorption is the overriding determinant of serum phosphate concentration and is subject to elaborate regulation by a wide variety of hormonal and metabolic factors. As noted earlier, the efficiency of intestinal phosphate absorption is high. Consequently, urinary phosphate excretion is tightly regulated to correlate with enteral phosphate loads.

Eighty percent of phosphate filtered at the glomerulus is reabsorbed by the proximal tubule.[13] Additional phosphate may be reabsorbed in the distal tubule and cortical collecting tubule. Phosphate must



Nomogram for derivation of renal threshold phosphate concentration

Fig. 53.1 Nomogram for calculating the ratio of tubular maximum reabsorption (TMP) to glomerular filtration rate (GFR). Calculations are done from fasting samples obtained in the morning. The urine is collected for 1–2 hours (the time period is not critical). The C_{PO4}/Ccreat = (U_{PO4} × [Creat])/(U_{creat} × [Pi]). To use the nomogram to calculate TMP/GFR, a straight line is passed through the appropriate plasma phosphate concentration and the C C_{PO4}/Ccreat. The line will intersect with the corresponding TMP/GFR value. For SI units, use the inside scales. For metric (mg/dL), use the outside scales. *TRP,* Tubular reabsorption of phosphate. (From Walton RJ, Bijvoet OL. Nomogram for derivation of renal threshold phosphate concentration. *Lancet.* 1975;2:309–310.)

be actively transported across the luminal brush-border membrane against a steep electrochemical gradient. Consistent with this, renal phosphate transport requires luminal sodium ions and is blocked by inhibitors of Na^+/K^+-ATPase.[14]

The two primary transport proteins responsible for phosphate reabsorption in the kidney are the type II sodium-phosphate cotransporters NPT2a and NPT2c, expressed in the apical membrane of the proximal tubule. Ablation of the *Npt2a* gene in the mouse results in a decrease in proximal tubular sodium phosphate cotransport, hypophosphatemia, and loss of regulation of phosphate reabsorption by both parathyroid hormone (PTH) and dietary phosphate.[15,16,17] Ablation of NPT2c in mice does not result in a phosphate phenotype.[18] However, mutations in the gene coding NPT2c, *SLC34A3*, result in hereditary hypophosphatemic rickets with hypercalciuria (HHRH), indicating that this transporter plays an important role in maintaining phosphate homeostasis in humans.

Measurement of Renal Phosphate Transport. For clinical purposes, a useful way of assessing a patient's ability to reabsorb phosphate is to calculate the tubular maximum reabsorption (TMP) of phosphate divided by the glomerular filtration rate (GFR). A simple nomogram has been developed by Walton and Bijvoet[19] (Fig. 53.1), though it does not encompass the upper end of the normal range of serum phosphate or of TMP/GFR in young children.

Regulation of Phosphate Reabsorption

Parathyroid Hormone. PTH has a major effect on serum phosphate. The clinical relevance of this effect is demonstrated in patients with hyperparathyroidism, who develop phosphaturia and hypophosphatemia, as well as those with hypoparathyroidism, who have increased phosphate reabsorption and hyperphosphatemia. PTH rapidly (15–60 minutes) reduces the number of NPT2a cotransporters on the apical surface of the cells in the renal proximal tubule. Despite the fact that PTH clearly has a major effect on phosphate homeostasis, the primary role of PTH is to regulate serum calcium and not phosphate homeostasis, as PTH secretion is regulated more by serum calcium than serum phosphate.

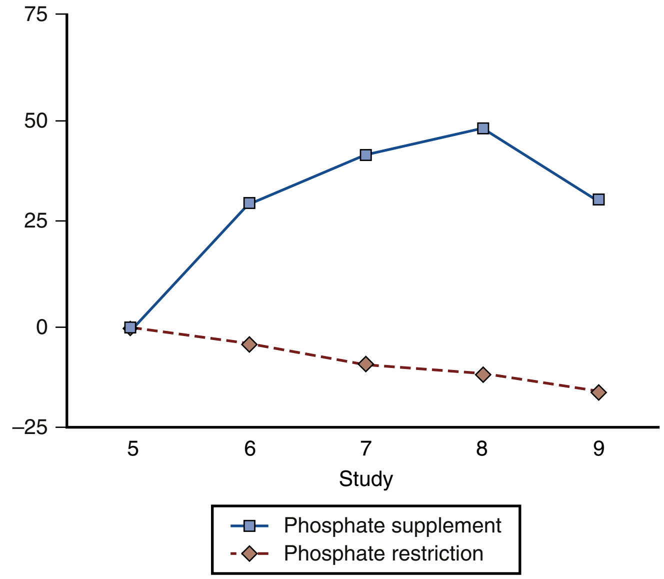


Fig. 53.2 Fibroblast growth factor-23 (FGF23) levels in response to diet. Percent change in FGF23 during the intervention period (♦, phosphate depletion intervention as measured with the intact FGF23 assay; ■, phosphate loading intervention as measured with the intact FGF23 assay). The intervention, either dietary phosphate depletion or loading, was started after the day 5 samples were collected. Values are presented as mean ± standard error. (Data from Burnett SM, Gunawardene SC, Bringhurst FR, et al. Regulation of C-terminal and intact FGF-23 by dietary phosphate in men and women. *J Bone Miner Res.* 2006;21:1187–1196.)

Fibroblast Growth Factor-23

FGF23, the gene identified as causative for autosomal dominant hypophosphatemic rickets and some forms of hfTC, does not appear to play a role in the rapid changes in phosphate reabsorption that occur following acute changes in intestinal phosphate content.[20] However, there is substantial evidence that FGF23 plays a role in maintenance of phosphate homeostasis over the course of days.

Fibroblast Growth Factor-23 Activity

FGF23 has similar functions as PTH to reduce renal phosphate reabsorption but has opposite effects on $1,25(OH)_2D$ production. FGF23 delivery leads to renal phosphate wasting through the downregulation of both NPT2a and NPT2c.[21] Under normal circumstances, hypophosphatemia is a strong positive stimulator for increasing serum $1,25(OH)_2D$ production. However, in mice, the expression of the $1\alpha(OH)$ase enzyme and the catabolic 24(OH)ase are reduced and elevated, respectively, when the animals are exposed to FGF23.[22] Thus, the effects of FGF23 on the renal vitamin D metabolic enzymes are responsible for the reductions in $1,25(OH)_2D$ seen in the setting of hypophosphatemia in ADHR, XLH, TIO, and ARHR patients.

Regulation of Fibroblast Growth Factor-23 Production

In humans, dietary phosphate supplementation increased FGF23, whereas phosphate restriction and the addition of phosphate binders suppressed serum FGF23 (Fig. 53.2),[23] indicating that FGF23 plays a role in maintaining phosphate homeostasis. In animal studies, the FGF23 response to serum phosphate appears more dramatic than in the human studies. Mice given high- and low-phosphate diets have increased and decreased *FGF23* gene expression, respectively, with corresponding changes in plasma FGF23 levels.[24]

Vitamin D has important regulatory effects on FGF23 *in vivo*. In mice, injections of 20 to 200 ng $1,25(OH)_2D$ led to dose-dependent increases in serum FGF23 concentrations.[22] These changes in FGF23 occurred before detectable changes in serum phosphate, indicating that FGF23 is probably directly regulated by vitamin D. Physiologically, this feedback mechanism is consistent with the role of FGF23 in vitamin D metabolism, in that FGF23 has been shown to downregulate the 1α(OH)ase mRNA.[21,22] Thus, as $1,25(OH)_2D$ is elevated in the blood as a product of 1α(OH)ase activity, vitamin D would then increase FGF23 production, which completes the negative feedback loop and downregulates 1α(OH)ase.

Enzyme-Linked Immunosorbent Assays Specific for Fibroblast Growth Factor-23. FGF23 can be measured in the bloodstream via several assays. One widely used assay is a C-terminal FGF23 enzyme-linked immunosorbent assay (ELISA), with both the capture and detection antibodies binding C-terminal to the $FGF23_{176}RXXR_{179}/S$ cleavage site.[25] This assay thus recognizes full-length FGF23 as well as C-terminal fragments that arise through proteolytic processing. The normal mean for this assay is 55 ± 50 reference units (RU)/mL, and the upper limit of normal is 150 RU/mL. Plasma FGF23 is detectable in normal individuals.[25] The mean C-terminal FGF23 value was greater than 10-fold elevated in TIO patients, falling rapidly after surgical resection. Importantly, most XLH patients (13 out of 21) had elevated FGF23 compared to controls,[25] and in those with "normal" FGF23, these levels may be considered "inappropriately normal" in the setting of hypophosphatemia.

There are multiple intact FGF23 assays using monoclonal monoclonal antibodies that bind to either side of the protein's $_{176}RXXR_{179}/S_{180}$ subtilisin-like proprotein convertase (SPC) cleavage site to recognize only the full-length protein.[26] In normal individuals, the Kainos intact assay has a mean circulating concentration of 29 pg/mL. The published upper limit of normal is 54 pg/mL.[26] The results of these two assays generally have good agreement regarding the relative elevations of FGF23 concentrations in XLH and in TIO patients. However, in diseases such as hfTC, and during iron deficiency, the intact and C-terminal ELISAs do not correspond, which is reflective of the underlying disease mechanisms. Based upon limited data from two TIO patients undergoing resection, the Kainos intact assay was used to determine that the half-life of FGF23 is between 20 and 50 minutes.[27] Additional intact FGF23 assays have been developed, but normal ranges with each intact assay may differ. The intact FGF23 level using the Kainos assay is generally similar when measuring plasma or serum samples. However, serum C-terminal FGF23 or intact FGF23 levels measured using assays from Immutopics are systematically much lower than those measured from plasma.[28] Thus, appropriate ranges need to account for the specific assay and manufacturer and the type of blood sample being measured.

KEY POINTS

- Phosphate metabolism is regulated at the level of the kidney, bone, and intestine and involves the hormones $1,25(OH)_2$ vitamin D ($1,25(OH)_2D$), parathyroid hormone, and fibroblast growth factor-23 (FGF23).
- FGF23 is produced in osteocytes and acts in the kidney though its coreceptor alpha-Klotho.
- FGF23 acts to down regulate $1,25(OH)_2D$ production and directly inhibit phosphate reabsorption in the kidney through internalization of proximal tuble sodium-phopshate cotransport proteins.
- Enzyme-linked immunosorbent assays have been developed for the intact, bioactive form of FGF23, as well as an assay that recognizes the C-terminal portions of FGF23.

FIBROBLAST GROWTH FACTOR-23–ASSOCIATED SYNDROMES

FGF23-associated syndromes, summarized in Table 53.1, which are discussed further as follows.

Disorders Associated with Increased FGF23 Bioactivity

Autosomal Dominant Hypophosphatemic Rickets. ADHR is a rare disorder characterized by low serum phosphate concentrations due to decreased TMP/GFR and inappropriately low or normal circulating vitamin D concentrations.[29] A large kindred demonstrated the wide phenotypic variability of ADHR. In contrast to the other genetic renal phosphate-wasting disorders, ADHR displays incomplete penetrance and variable age of onset. Some patients presented during childhood with phosphate wasting, rickets, and lower-extremity deformity similar to the classic presentation of XLH. Other individuals had grown normally without rickets but presented during adolescence or adulthood with new-onset hypophosphatemia. These individuals had bone pain, weakness, and insufficiency fractures.[29] The patients with adult-onset ADHR had clinical presentations essentially identical in clinical severity to patients with TIO. To date, most patients described with delayed onset of clinically evident disease are female. In addition, we found unaffected individuals with the ADHR mutation and patients who were treated for hypophosphatemia and rickets but later lost the phosphate-wasting defect.[29]

FGF23 was identified from genome-wide linkage studies in this and other kindreds and using exon prediction programs using genomic DNA sequence from the Human Genome Project.[30] The *FGF23* gene is composed of three coding exons and contains an open reading frame of 251 residues.[30] The tissue with the highest FGF23 expression is bone, where FGF23 mRNA is observed in osteoblasts, osteocytes, flattened bone-lining cells, and osteoprogenitor cells.[31] Quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) showed that FGF23 mRNA was most highly expressed in long bone, followed by thymus, brain, and heart.[32]

Western blot analysis demonstrated that wild-type FGF23 is secreted as a full-length 32-kD protein, as well as cleavage products of 20 kD (N-terminal) and 12 kD (C-terminal).[32-34] Cleavage of FGF23 occurs within a SPC proteolytic site ($_{176}RXXR_{179}/S_{180}$) that separates the conserved FGF-like N-terminal domain from the variable C-terminal tail (Fig. 53.3).

The ADHR mutations replace arginine (R) residues at FGF23 positions 176 or 179 with glutamine (Q) or tryptophan (W) within the SPC cleavage site, $_{176}RXXR_{179}/S_{180}$ (Table 53.2 and see Fig. 53.3).[30,33-36] The SPCs are a family of serine proteases that process a wide variety of proteins, including neuropeptides, peptide hormones, growth factors, membrane-bound receptors, blood coagulation factors, and plasma proteins.[37] SPC substrates are cleaved C-terminal to the basic motif K7R-X,,-K7R, where X_n = 2, 4, or 6 residues of any amino acid.[38,39] The SPCs, such as the furin protease, are largely expressed in the trans-Golgi network and possess similar, but not exact, substrate specificities.

Following insertion of these mutations into wild-type FGF23, FGF23 secreted from mammalian cells was primarily produced as the full-length (32 kD) active polypeptide, as opposed to the cleavage products observed for wild-type FGF23 expression.[33] Peptide sequencing demonstrated that the 32-kD FGF23 form corresponded to full-length FGF23 after cleavage of the signal peptide (residues 25 to 251) and that the 12-kD isoform was the C-terminal portion of FGF23 downstream from the SPC cleavage site after R179 (residues 180–251) (see Fig. 53.3).[34] These studies show that the RXXR motif in FGF23 is central to its intracellular processing.

TABLE 53.1 Heritable and Acquired Disorders Involving Fibroblast Growth Factor-23 (FGF23)

Disorder	Gene Mutated	Mutation Consequence	Relationship to FGF23	Effect on Serum Phosphate	Effect on Serum 1,25D	Intact ELISA (FGF23)	C-Terminal ELISA (FGF23)	Respective Animal Models
ADHR	*FGF23*	Gain of function	Stabilize full-length, active FGF23	↓	↔	↔ or ↑	↔ or ↑	ADHR FGF23 R176Q knockin mouse
XLH	*PHEX*	Loss of function	Increased FGF23 production in osteocytes	↓	↔	↔ or ↑	↔ or ↑	Hyp mouse (*Phex*-null)
ARHR	*DMP1*	Loss of function	Increased FGF23 production in osteocytes	↓	↔	↔ or ↑	↔ or ↑	*Dmp1*-null mouse
ARHR	*ENPP1*	Loss of function	Increased FGF23 production	↓	↔	↔ or ↑	↔ or ↑	*Enpp1*-null mouse
ARHR (variant of Raine syndrome)	*Fam20c*	Loss of Function	Stabilzation due to loss of FGF23 phosphorylation	↓	↔	↔ or ↑	↔ or ↑	*Fam20c* osteoblast- and osteocyte-specific conditional null mouse
αKlotho overexpression	*KLOTHO*	Gain of function	Increased FGF23 mRNA production	↓	↔	↑	↑	*KL* overexpression through AAV
TIO	—	—	FGF23 highly produced by tumor	↓	↔	↔ or ↑	↔ or ↑	FGF23 general and bone-specific transgenics
TC	*FGF23*	Loss of function	Destabilize full-length, active FGF23	↑	↔ or ↑	↓	↑	*Fgf23*-null mice
TC/HHS	*GALNT3*	Loss of function	Destabilize full-length, active FGF23	↑	↔ or ↑	↓	↑	*Galnt3*-null mouse
TC	*KL*	Loss of function	Decreased FGF23-dependent signaling	↑	↔ or ↑	↑	↑	*KL* knockdown and *KL*-null mice

ELISA, Enzyme-linked immunosorbent assay; *1,25D*, 1,25$(OH)_2$ vitamin D; *ADHR*, autosomal dominant hypophosphatemic rickets; *XLH*, X-linked hypophosphatemic rickets; *ARHR*, autosomal recessive hypophosphatemic rickets; *TIO*, tumor-induced osteomalacia; *TC*, tumoral calcinosis; *HHS*, hyperostosis-hyperphosphatemia syndrome.

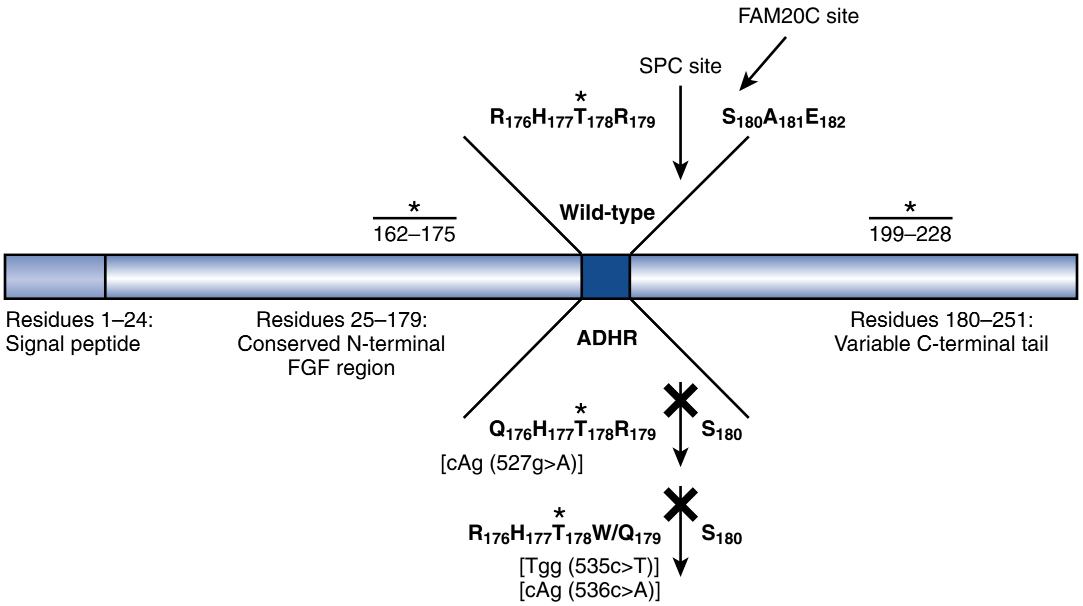

FAM20C site
SPC site
*
$R_{176}H_{177}T_{178}R_{179}$
$S_{180}A_{181}E_{182}$
Wild-type
*
162–175
*
199–228
Residues 1–24: Signal peptide
Residues 25–179: Conserved N-terminal FGF region
ADHR
Residues 180–251: Variable C-terminal tail
*
$Q_{176}H_{177}T_{178}R_{179}$
S_{180}
[cAg (527g>A)]
*
$R_{176}H_{177}T_{178}W/Q_{179}$
S_{180}
[Tgg (535c>T)]
[cAg (536c>A)]

Fig. 53.3 Model of fibroblast growth factor-23 (FGF23) protein domains and effect of the autosomal dominant hypophosphatemic rickets (ADHR) mutations. FGF23 has a 24-residue signal peptide followed by residues 25–179 that comprise the conserved N-terminal fibroblast growth factor (FGF)-like domain. The subtilisin-like proprotein convertase (SPC)-like site is interrupted by the ADHR mutations at R_{176} and R_{179} and divides the FGF-like domain from the variable C-terminal tail region. FGF23 undergoes glycosylation at three regions (denoted *), with T_{178} most likely being the glycosylated residue that protects the mature protein from SPC proteolysis between R_{179} and S_{180}, and is therefore critical for maintaining intact, active FGF23. The FAM20C phosphorylation site at S180 blocks GALNT3-mediated glycosylation of FGF23, leading to destabilization of intact bioactive FGF23.

TABLE 53.2 **Mutations in the Human Fibroblast Growth Factor-23 (FGF23) Gene**

FGF23 Mutation	Predicted Amino Acid Substitution	Demonstrated or Predicted Effect on FGF23	Disorder	Reference
527G>A	R176Q	Stabilized	ADHR	108
535C>T	R179W	Stabilized	ADHR	108
536G>A	R179Q	Stabilized	ADHR	108, 114, 115
526 C>T	R176W	Stabilized	ADHR	113
123C>T	H41Q	Destabilized	FTC	239
160C>A	Q54K	Destabilized	FTC	205
211A>G	S71G	Destabilized	FTC	202, 203
287C>T	M96T	Destabilized	FTC	206
386C>T	S129F	Destabilized	FTC	204

ADHR, Autosomal dominant hypophosphatemic rickets; *FTC,* familial tumoral calcinosis.

Cleavage of FGF23 at the RXXR motif appears to be inactivating. In this regard, when full-length FGF23 or N-terminal (residues 25–179) and C-terminal (residues 180 –251) fragments were injected into rodents, only the full-length protein lowered circulating phosphate concentrations.[34] Because the full-length form of FGF23 induces hypophosphatemia, it is likely that the ADHR mutations increase the biological activity of FGF23 by stabilizing the full-length form and increasing its concentrations in the serum. Indeed, severely affected ADHR patients have increased circulating levels of FGF23.[40] However, until recently it was unclear why ADHR patients were not able to simply reduce *FGF23* gene expression. The answer to this dilemma came from an unexpected connection between iron and FGF23.

Iron and Fibroblast Growth Factor-23 Interactions. Studies in ADHR kindreds demonstrated that disease status varied with FGF23 concentrations.[40] Subjects with ADHR mutations that did not manifest hypophosphatemia had normal FGF23 concentrations, while hypophosphatemic patients had elevated FGF23 concentrations. Cross-sectional and longitudinal studies demonstrated an inverse correlation between serum phosphate and FGF23 concentrations in ADHR patients, but not controls. In ADHR patients, low iron concentrations resulted in elevations in both total (intact and fragments as measured by the C-terminal assay) and intact FGF23 concentrations. However, in control individuals, low iron concentrations were also associated with elevated total FGF23 concentrations measured by the C-terminal assay, while intact FGF23 concentrations were normal.[41]

Studies in the ADHR mouse provide a detailed mechanism for the above observations. When placed on an iron-deficient diet, both control and R176Q-Fgf23 knockin mice (the animal model of ADHR) had marked elevations in bone Fgf23 mRNA. Wild-type mice on the low-iron diet maintained normal serum intact Fgf23 and phosphate concentrations but displayed elevations in serum C-terminal Fgf23 levels. In contrast, the ADHR mice receiving the low-iron diet had elevated intact, biologically active Fgf23 as well as C-terminal Fgf23, which resulted in hypophosphatemic osteomalacia.[42] Additionally, *in vitro* studies using the osteoblastic cell line UMR-106 demonstrated that iron chelation with 25 or 50 uM deferoxamine increased Fgf23 mRNA 12- and 20-fold, respectively.[42] The mechanism is mediated by increased expression of hypoxia inducible factor 1α.[42] Additional studies[43] demonstrate that hypoxia can also increase FGF23 production. In this regard, UMR-106 cells grown under reduced oxygen tension, but normal iron concentrations, increase FGF23 mRNA over 3-fold by 24 hours. Furthermore, normal Sprague-Dawley rats exposed to hypoxia for 2 weeks resulted in over 6-fold increase in C-terminal Fgf23 compared with control rats at normal oxygen tension, with no change in intact Fgf23 or serum phosphate, supporting the concept that hypoxia increases Fgf23 expression, but normal individuals are able to cleave FGF23 to prevent hypophosphatemia. Interestingly, studies done in neonatal mice whose mothers were fed an iron-deficient diet demonstrate increased intact FGF23 and hypophosphatemia in wild-type mice, although not to the same extent observed in ADHR mice.[43] These findings raise the possibility that iron deficiency during pregnancy could have negative effects on the FGF23/phosphate/calcitriol/bone axis in normal humans. Additionally, inflammation that causes a functional iron deficiency,[44] and other regulators of iron handling, such as erythropoietin,[45] also increase circulating FGF23. Although, without an ADHR mutation, acute intact FGF23 increases, but with longer exposure the major species detected in plasma from normal individuals and rodents are C-terminal fragments.[44]

Taken together, studies in human subjects and in animal models demonstrate that iron deficiency and hypoxia increase expression of FGF23 mRNA and protein in bone. In individuals without an FGF23 mutation, the excess FGF23 protein is cleaved into inactive fragments (inactive with respect to phosphate and calcitriol homeostasis). However, patients and mice with ADHR mutations are unable to cleave the excess mutant FGF23 and manifest elevated intact FGF23 concentrations resulting in hypophosphatemia and inappropriately normal calcitriol concentrations. Thus, disease status in ADHR is a product of gene–environment interactions in which the combined presence of an FGF23 mutation and iron deficiency results in the biochemical and phenotypic manifestations of ADHR.

To close the loop on this pathophysiology, we conducted an open-label clinical trial treating symptomatic ADHR patients and individuals at risk for becoming symptomatic based on serum iron levels, starting with a daily dose of 65 mg of oral iron (ferrous sulfate 325 mg tablets). Doses were adjusted according to a prespecified algorithm. Oral iron was chosen in light of studies indicating that intravenous iron could trigger transient acute increased intact FGF23 and hypophosphatemia in normal iron-deficient individuals.[46] By the end of the 12-month trial, all subjects were normophosphatemic, with normal C-terminal and intact FGF23 concentrations, and those that were treated with calcitriol and phosphate therapy were able to discontinue these while maintaining normophosphatemia.[47] Thus, treatment with oral iron to replete iron stores was able to normalize intact FGF23 and serum phosphate in patients with ADHR.

KEY POINTS

- Fibroblast growth factor-23 (FGF23) gene expression is increased in the setting of iron deficiency.
- Cleavage of FGF23 before secretion is a regulated step, and during iron deficiency FGF23 cleavage maintains normal intact FGF23 and phosphate despite the elevated gene expression.
- Some types of intravenous iron can paradoxically acutely increase the intact FGF23 transiently for weeks after dosing in the setting of iron deficiency. This is thought to be caused by a secondary effect of the carbohydrate moiety of iron infusions to impair FGF23 cleavage combined with the setting of iron deficiency–induced elevated *FGF23* gene expression.
- Autosomal dominant hypophosphatemic rickets (ADHR) is caused by mutations that impair the cleavage of FGF23, resulting in increased FGF23 concentrations that appear to manifest primarily during iron deficiency.
- Oral iron repletion to normal iron stores can normalize the FGF23 and phosphate levels in ADHR.

Tumor-Induced Osteomalacia. TIO is an acquired disorder of renal phosphate wasting that is associated with tumors. Patients with TIO present with hypophosphatemia with inappropriately suppressed $1,25(OH)_2D$ concentrations and elevated alkaline phosphatase levels.[48] Clinical symptoms include gradual onset of muscle weakness, fatigue, and bone pain, especially from ankles, legs, hips, and back.[48,49] Osteomalacia is observed in bone biopsies. Insufficiency fractures are common, and proximal muscle weakness can become severe enough for patients to require a wheelchair or become bedbound.[48] FGF23 mRNA and protein have been demonstrated to be highly expressed in these tumors, and plasma levels are usually correspondingly elevated.[25,50,51]

X-Linked Hypophosphatemic Rickets (OMIM No. 307800). XLH is an X-linked dominant disorder and the most common form of heritable rickets. XLH is fully penetrant, with variable severity. XLH patients present with laboratory findings that include hypophosphatemia with normocalcemia and inappropriately normal or low $1,25(OH)_2D$ concentrations.[52] Skeletal defects include lower-extremity deformities from rickets, short stature, dental abscesses, bone pain, osteomalacia, fracture, enthesopathy (calcification of the tendons and ligaments), and early-onset osteoarthritis. XLH is caused by inactivating mutations in *PHEX* (***p****hosphate-regulating gene with* ***h****omologies to* ***e****ndopeptidases on the* ***X*** *chromosome*).[53] Based upon sequence homology, PHEX encodes a protein that is a member of the M13 family of membrane-bound metalloproteases. Other members of this enzyme class include neutral endopeptidase and endothelin-converting enzymes 1 and 2.[53,54] This family is known to cleave small peptides. Over 300 inactivating *PHEX* mutations have been described in XLH patients, including genomic variations causing missense, nonsense, frameshift, and splicing changes (see PHEX Locus database: www.phexdb.mcgill.ca). PHEX shows the highest expression in bone cells such as osteoblasts, osteocytes, and odontoblasts in teeth, as well as lower expression in the parathyroid glands, lung, brain, and skeletal muscle, but no expression in kidney.[55]

A valuable tool for the study of the pathophysiology of XLH has been the *Hyp* mouse, which has 3′ deletion in the *Phex* gene from intron 15 through the 3′ untranslated region and does not make a stable Phex transcript.[55] This rodent model parallels the XLH biochemical phenotype, as well as having growth retardation and bone mineralization defects.[56]

X-Linked Hypophosphatemic Rickets and Fibroblast Growth Factor-23. Several reports have established that FGF23 is elevated in most XLH patients.[25,26] Quantitative real-time RT-PCR indicates that FGF23 mRNA in bone from *Hyp* mice is elevated compared with control mice,[32] and serum concentrations of FGF23 are up to 10-fold higher in *Hyp* mice compared with normal mice.[57] This finding indicates that there is a cellular connection between inactive PHEX mutants (or lack of Phex expression in *Hyp* mice) and the upregulation of FGF23 mRNA in bone cells resulting in overproduction and secretion of FGF23 by skeletal cells. Although the interactions between FGF23 and PHEX are most likely indirect, the encoded proteins are coexpressed in osteoblasts and osteocytes.[31,55] Although one report[58] provides evidence that PHEX can cleave osteopontin, its physiologically relevant substrate and the mechanisms for phosphate sensing are unknown. A further complication is that, not only are FGF23 mRNA and protein expression increased, but the ratio of intact FGF23 to total FGF23, as measured by the C-terminal assay, is increased in both Hyp mice[59] and XLH patients.[60] Although PHEX does not cleave FGF23, absence of PHEX may influence not only FGF23 expression, but reduce cleavage of intact, biologically active FGF23.

Interestingly, despite persistent hypophosphatemia during therapy with active vitamin D and phosphate salts, serum FGF23 concentrations generally rise in response to therapy.[61] To explore this phenomenon further, Ichikawa et al.[59] crossed mice with *Phex* mutations with *Galnt3*-null mice. GALNT3/Galnt3 O-glycosylates FGF23, protecting the FGF23 protein from degradation, and GALNT3 mutations lead to tumoral calcinosis (TC).[62] Humans and mice with these mutations have markedly decreased intact, biologically active FGF23 and manifest hyperphosphatemia and soft tissue calcificaitons (see later). Hyp mice displayed hypophosphatemia and elevated total (intact plus fragments) and intact FGF23 concentrations, whereas *Galnt3*-null mice were hyperphosphatemic with low intact FGF23 concentrations, but had high FGF23 mRNA expression and high total FGF23 concentrations (almost all inactive fragments). In the absence of altered phosphate responsiveness, FGF23 mRNA expression in the double mutant would be expected to be similar to the Hyp. However, the Hyp/Galnt3 double mutant remained hypophosphatemic with serum phosphate concentrations slightly, but significantly, higher than Hyp, and with markedly upregulated FGF23 gene expression even higher than the Hyp mouse (up to 24-fold higher than the 12-week-old Hyp mice, and 113-fold higher than control mice), with corresponding total FGF23 levels.[59] This study further supports the concept that PHEX deficiency lowers the regulated serum phosphate set point.

KEY POINTS

- X-linked hypophosphatemic rickets (XLH) is a lifelong condition resulting in chronic hypophosphatemia due to fibroblast growth factor-23 excess.
- Children with XLH generally present with features of rickets and poor growth.
- Adults with XLH manifest with bone pain, pseudofractures, and progressive osteoarthritis and enthesopathy, impairing mobility and quality of life.
- Dental complications are common in both children and adults.

Clinical Treatment of X-Linked Hypophosphatemic Rickets. Current treatment options of XLH include "conventional therapy" using a combination of phosphate salts and active vitamin D or monotherapy with burosumab.

Conventional Therapy. For several decades, the conventional therapy for XLH, ADHR, and TIO has involved administration of oral phosphate salts in combination with high-dose active forms of vitamin D (using calcitriol or alfacalcidol).[52] This regimen ameliorates much of the metabolic bone disease, but it does not directly "cure" the disorder, in that it does not reverse the underlying molecular defects in kidney and in bone. Therapy is labor-intensive, requiring multiple doses per day to be effective, as well as frequent laboratory monitoring for safety. Unfortunately, the conventional therapy has often been mistaken as taking nutritional "supplements," leading insurance companies to often deny coverage and patients and providers to overlook the real risks of important side effects of conventional therapy.

In fact, nutritional vitamin D supplementation has long been known to be ineffective at treating XLH, leading to one of its old names, as a "vitamin D–resistant rickets." In reality, XLH has nothing to do with resistance to vitamin D at its receptor. Thus, this name was not mechanistically accurate. In XLH, rickets and osteomalacia persist even despite very high doses of vitamin D, which are complicated by risk for vitamin D toxicity. Attempts to treat with phosphate salts alone, however, suffered from lack of effect to heal the rickets or osteomalacia, and are accompanied by very high risk for hyperparathyroidism. Eventually, multiple studies demonstrated that the combination of active forms of vitamin D (calcitriol, alfacacidol) with phosphate salts is necessary for improving the rickets and osteomalacia. However, medical therapy requires an ability to carefully monitor labs for safety. The goals of conventional therapy are not to normalize serum phosphate. In fact, overly aggressive attempts to normalize phosphate with conventional therapy may be associated with complication of therapy.[52]

The objectives of therapy in children are to correct or prevent deformity from rickets, promote growth, and lessen bone pain. In some cases, marked improvement in lower-extremity deformity is noted with medical therapy. Thus, surgical approaches to correcting lower-extremity deformities (using osteotomies or guided growth procedures) can often be delayed for a period of time in young children to see the magnitude of benefit from medical therapy alone. However, the response to conventional therapy is quite variable, and the need for corrective surgery remains quite common. Growth remains impaired, and the mean adult height is less than –2.5 standard deviation scores.[63] Most affected children probably should be treated if a system exists to safely administer and monitor therapy, including frequent laboratory assessments (usually every 3–4 months).

The indications for treatment of adult patients are more controversial.[52] Historically, most children have stopped treatment at the end of their growth, due to the complex balance of risks versus benefits. Some adults do well for a period of time without pain or fractures, while others develop these symptoms, which may be mild or severe and develop early or later in adulthood. Unfortunately, after stopping treatment at the end of growth, patients may wrongly assume that symptoms such as pain or fracture are due to something else, leading to great delays in reestablishing care and treatment. Additional delays may be caused by providers assuming that XLH is a disease of childhood, when in fact adults bear a large burden of disease. In a burden of disease study of 232 adults, 97% reported bone pain, joint pain, and stiffness, and 44 % had a history of fracture.[64] Of these adult respondents, 86% had abnormal gait (likely due to a combination of the residual effect of childhood deformities, as well as of their bone pain, osteoarthritis, and enthesopathy). The adult features of XLH are dominated by bone pain, pseudofracture, osteoarthritis, and enthesopathy. However, there are no data to suggest that conventional therapy for XLH prevents enthesopathy (calcifications of tendons and ligaments)[65] or osteoarthritis. XLH patients frequently complain of bone pain,[64] presumably due to osteomalacia. However, pseudofractures are common in moderately to severely affected adult XLH patients and are often unrecognized as a source of bone pain; thus, radiographs should be routine in evaluating patients with bone pain.[66] Because pseudofractures are often painful, may lead to complete fracture, and may be improved by treatment, medical therapy is generally recommended for patients with pseudofractures. Patients with bone pain, even without pseudofracture, may benefit from a trial of therapy. Additionally, it is advisable to treat patients who have nonunion after fractures or osteotomies, or after undergoing other orthopedic procedures, to help with bone healing.[52]

Once the decision is made to initiate conventional therapy, it is best to start at a low dose of calcitriol and phosphate (to avoid diarrhea from phosphate) and gradually increase therapy to target doses over several weeks for tolerability. It is important not to treat XLH with phosphate alone. Guidelines have recommended phosphate salts in doses ranging from 20 to 60 mg/kg daily be combined with either calcitriol 20 to 40 ng/kg daily or alfacalcidol 30 to 50 ng/kg daily. Phosphate doses usually need to be divided across four doses in a day, while calcitriol is usually given twice a day. Some clinicians have recommended administering phosphate over five doses per day; however, we do not ask patients to wake up to take medication at night, for the following reasons: (1) serum phosphate concentrations tend to rise at night in XLH patients, (2) there is an exaggerated nocturnal rise in PTH concentrations in XLH patients, and (3) it is very difficult for patients to wake up to take medication every night for a chronic condition.[67] Doses are individualized based on a combination of factors to optimize and balance tolerability, risks, and benefits.[52] Some patients may require higher doses for a period of time to improve bone healing. This has to be balanced with noting that higher doses may be associated with more side effects, including nephrocalcinosis and hyperparathyroidism. In addition, calcitriol and phosphate intake stimulate FGF23 expression, further increasing plasma FGF23 levels even in XLH patients.[68] There are currently no data as to whether it is clinically useful to monitor FGF23 concentrations during conventional therapy. However, increases in FGF23 concentrations are a possible unwanted side effect of standard therapy as well as any therapy that increases phosphate and/or calcitriol concentrations without correcting the hypothesized osteocyte phosphate set-point defect.

Careful laboratory monitoring is required when employing conventional therapy. Serum calcium, phosphorus, and creatinine levels, as well as urine calcium and creatinine concentrations, should be routinely monitored for safety, on a monthly basis during dose escalation and then every 3 to 4 months afterward. Alkaline phosphatase can be monitored as a sign of healing of rickets or osteomalacia. PTH should also be monitored every 6 months. The doses of calcitriol and phosphate are adjusted based on the laboratory results, but not necessarily targeting normal serum phosphate. Administration of phosphate leads to secondary hyperparathyroidism, while administration of active vitamin D can lead to hypercalciuria and hypercalcemia. Hence, a patient on conventional therapy can develop hypercalciuria as a consequence of either too much $1,25(OH)_2D$ or too little phosphate. Hypercalciuria may also be the first sign of healing of osteomalacia (theoretically, osteomalacic bone is capable of taking up more calcium and phosphate than bone that is histologically normal). Patients may require higher doses for initial healing, and then doses may be somewhat decreased during the long-term "maintenance" phase, especially to minimize calciuria and effects on PTH and renal function.

Diarrhea and gastrointestinal upset frequently develop from phosphate administration. To minimize diarrhea, the phosphate dose should be increased gradually. If diarrhea develops, it is best to decrease (or stop) the dose of phosphate and slowly increase therapy as tolerated. Because hypercalcemia may develop from calcitriol therapy, patients should be given instructions on the symptoms of hypercalcemia, such as depression, confusion, anorexia, nausea, vomiting, polyuria, and dehydration. Serum biochemistries should be measured if a patient experiences any symptoms of hypercalcemia.

The most common and potentially serious complication of therapy is the development of nephrocalcinosis.[52] With the widespread use of renal ultrasound, nephrocalcinosis has been noted with increasing frequency. Studies that are mostly cross-sectional indicate that nephrocalcinosis occurs in 50% to 80% of patients with XLH on conventional therapy, including adults and children. Hypercalciuria and hypercalcemia may contribute to (or at least aggravate) this problem. In most cases, the nephrocalcinosis is observed without evident changes in GFR; however, its long-term effect on kidney function remains uncertain. Data are sparse regarding chronic kidney disease in general in XLH, but some studies have indicated that 8% to 9% of adults with XLH have at least some degree of chronic kidney disease, which complicates management of their bone disease.[69] End-stage renal disease appears to be rare, but has been reported.

Hyperparathyroidism is a complex feature of XLH and its treatment. Prior to starting any treatment patients commonly have secondary hyperparathyroidism, likely the consequence of FGF23-mediated impairments in vitamin D activation. This often improves as calcitriol treatment begins. However, phosphate salts specifically trigger rises in PTH via their effect to decrease serum calcium after dosing, also contributing to the development of hyperparathyroidism. High phosphate doses, especially with insufficient calcitriol, can lead to prolonged secondary hyperparathyroidism with multigland hyperplasia, which converts to tertiary (or hypercalcemic) hyperparathyroidism in some patients. Secondary hyperparathyroidism occurs in up to 83% of patients with XLH, which can sometimes resolve with dose adjustment.[69] However, tertiary hyperparathyroidism occurs in 4% of children and 25% to 30% of adults.[69] In tertiary hyperparathyroidism, renal function may decline, necessitating a reduction in the dose of phosphate and calcitriol or cessation of therapy, with corresponding limitations in effectiveness at treating osteomalacic bone disease. Typically, all the parathyroid glands are abnormal in tertiary hyperparathyroidism, and removing a single gland is insufficient to resolve hypercalcemia. The recurrence rate after surgery was high (75%) in a small study at a median follow-up of 6 years, and hypercalcemia has recurred even in some patients with four-gland resection with reimplantation of a piece of a gland.[69] One risk of the procedure is complete hypoparathyroidism, which greatly complicates management, as patients have both hypocalcemia and hypophosphatemia, with both hypercalciuria and hyperphosphaturia. Cinacalcet has been used to manage tertiary hyperparathyroidism, with mixed results.[69]

Although treatment complications cannot be prevented entirely, careful monitoring may minimize treatment-related problems. Serum calcium, phosphorus, and creatinine levels and urine calcium and creatinine concentrations should be monitored monthly while patients are on high-dose therapy and every 3 to 4 months while on maintenance therapy. A 24-hour urine phosphorus measurement is often useful to gauge compliance with therapy. Serum PTH concentration should be measured at 6- to 12-month intervals and should influence dose adjustments as appropriate. Renal ultrasounds should be assessed every 1 to 2 years. In adults restarting therapy, it is important to assess PTH and renal function and renal ultrasound at baseline, as these may be abnormal even years after stopping prior conventional therapy.

Burosumab. In preclinical animal studies, using murine antibodies to FGF23 improved the tubular reabsorption of phosphate, serum phosphorus, and 1,25(OH)$_2$D in the Hyp mouse model, resulting in improvements in mineral apposition rate, bone volume/total volume, and skeletal growth, and improved organization and mineralization of the Hyp mouse's rachitic growth plate, femur shape, and skeletal growth.[70]

Burosumab is a fully human IgG$_1$ monoclonal antibody against the N-terminal domain of FGF23; thus, it binds to both intact FGF23 and N-terminal FGF23 fragments. Burosumab was developed for treating XLH and TIO, administered subcutaneously. Its half-life is 13 to 19 days when administered subcutaneously. In children with XLH- the FDA approved dosing of burosumab is 0.8 to 1.2 mg/kg subcutaneously every 2 weeks, and in adults with XLH the approved dose is 1 mg/kg subcutaneously every 4 weeks, with a maximum dose of 90 mg (https://www.accessdata.fda.gov2020). Burosumab has also been approved for managing the osteomalacia and hypophosphatemia of TIO at doses from 0.4 to 2 mg/kg every 2 weeks in children or every 4 weeks in adults.

Unlike conventional therapy for XLH, by binding FGF23 and blocking its activity burosumab alters the renal transport of phosphorus, allowing clinicians to effectively and safely target normalizing serum phosphorus levels. Both adult and pediatric studies demonstrated that burosumab increases 1,25(OH)$_2$D, TMP/GFR, and serum phosphate after injection.[71,72] Following burosumab dosing, 1,25(OH)$_2$D peaks between 3 and 7 days later; TMP/GFR peaks approximately 7 days after injection, and serum phoshorus increases with a peak between 3 and 7 days after injection.[71] However, hyperphosphatemia may occur, requiring careful monitoring and dose decreases if this occurs. Burosumab is given as monotherapy; all clinical trials had patients undergo washout of conventional therapy for 1 week or more before starting burosumab.

The first phase II pediatric XLH trial was designed to establish both dose level and dose interval. Dosing burosumab every 2 weeks was found to provide a better serum phosphorus profile than every 4 weeks among children with XLH aged 5 to 12 years at enrollment.[72] Children dosed every 2 weeks also had greater improvement in rickets than those dosed every 4 weeks during this 64-week trial. Subsequently, every 2-week dosing in younger children aged 1 to 4 years also demonstrated increases in serum phosphorus and improvement in rickets and lower limb deformities.[73] The only active comparator randomized controlled trial of burosumab was conducted in children aged 1 to 12 years. In this trial, children aged 1 to 12 years with XLH had to have Rickets Severity Scores of 2 or more despite treatment with conventional therapy for a mean of 3.3 to 4.3 years between randomization groups.[74] After a washout period of 7 days, these children were randomized to restart conventional therapy or switch to burosumab for 64 weeks. The fasting serum phosphate remained low in the conventional therapy group, while the fasting serum phosphate increased into the normal range during treatment with burosumab. While both groups had improvements in rickets, greater improvements in radiographic rickets was seen in those receiving burosumab at week 40 (the primary endpoint) and at week 64. Lower limb deformity improved in these trials, though to a lesser degree than the changes in the growth plate during 64 weeks of treatment. There were also modest improvements in length/height noted during treatment with burosumab compared with conventional therapy that were similar to the growth improvements seen in the prior phase II study.

A phase III randomized double-blind placebo-controlled trial of burosumab was conducted in adults with XLH who had Brief Pain Inventory worst pain scores of greater than or equal to 4.[66] These patients had evidence of severe disease: at baseline approximately half of subjects had active fractures or psuedofractures, 99% had enthesopathy, and half had nephrocalcinosis. Participants received 1 mg/kg burosumab (up to 90 mg) or placebo every 4 weeks for 24 weeks, followed by open-label burosumab during a

24-week extension period. The primary outcome of normal mean serum phosphorus at midcycle of the dosing range (2 weeks post-dose) throughout the trial was achieved by 94% in the burosumab group versus 7.6% in the placebo group. Treatment with burosumab completely healed 43% of active fractures or psuedofractures at 24 weeks, versus only 7.7% in the placebo group. After 48 weeks of burosumab, 63.1% of active fractures and psuedofractures had completely healed.[66] This improved healing is an important treatment target, but the study also demonstrates the prolonged healing time that is often necessary for patients with XLH. Some patients may still require rodding for stabilization of psuedofractures. Burosumab also improved Brief Pain Inventory worst pain score, stiffness, physical function, and 6-Minute Walking Distance. During a parallel trial, 11 subjects had paired bone biopsies before and after 48 weeks of burosumab 1 mg/kg, demonstrating improvements in osteoid volume/bone volume, osteoid thickness, and mineralization lag time.[75]

Two clinical trials of burosumab for treatment of TIO have been conducted: one in the United States and one in Japan and Korea. In the trial in Japan and Korea, burosumab increased serum phosphorus, decreased pain, and improved walking distance.[76] Burosumab has been FDA-approved for TIO.[77]

The most common adverse event in children was injection site reactions, occurring in 57.7%.[72] Injection site reactions were less common in adults.[66] Restless legs syndrome emerged during the first multidose trial in adults[71] and occurred in approximately 11.8% of those in the burosumab arm and 7.6% of those in the placebo arm of the placebo-controlled trial. Since publication of these trials, it has become our practice to routinely query restless legs syndrome in all adults with XLH, and we have noted frequent occurrence even among patients not receiving any medical therapy, raising the possibility of restless legs as a feature of XLH. However, it is also clear that, in some patients, restless legs emerges or worsens after starting burosumab.

Hyperparathyroidism is not likely to be a long-term risk for burosumab, but long-term studies are needed. Actually, during clinical trials, PTH levels improved in adults, though the mean values remained elevated.[66] The data available from clinical trials so far do not indicate a difference in the rate of nephrocalcinosis with burosumab versus conventional therapy. Because the renal phosphate reabsorption is improved with burosumab, the risk of nephrocalcinosis long-term may be less. However, there are potential reasons for burosumab to at least carry some risk for nephrocalcinosis, including the rise in $1,25(OH)_2D$ after each burosumab dose and the potential risk for hyperphosphatemia with burosumab (small numbers of patients in adult clinical trials required dose decreases due to excessive phosphate levels), both of which could increase risk for nephrocalcinosis. Consequently, we recommend continuing to monitor renal ultrasounds in patients treated with burosumab. Kidney function did not worsen during the trials. However, burosumab is contraindicated in patients with moderate to severe renal disease. It should be noted that, in rat studies of chronic kidney disease, using an anti-FG23 antibody worsened hyperphosphatemia and vascular calcifications, leading to death.[78]

Many things remain unknown about the effects of burosumab in the long term. To date there is no evidence for any effect of burosumab on enthesopathy. In addition, the full effect of burosumab on development of adult height is also not yet known, nor to what degree burosumab may reduce the need for orthopedic surgeries in children or adults. Burosumab has not yet been adequately studied in pregnancy or lactation.

KEY POINTS

- The first step to appropriate medical therapy is making the correct diagnosis, as treatment and risks differ by condition.
- Chronic fibroblast growth factor-23 (FGF23)-mediated hypophosphatemia may be treated using either conventional therapy with active vitamin D analogs combined with phosphate salts or monotherapy with an anti-FGF23 antibody, burosumab.
- In children with X-linked hypophosphatemia, both conventional therapy and burosumab improved rickets, with greater improvement from burosumab.
- Medical therapy of chronic hypophosphatemic conditions require frequent and careful laboratory monitoring for both efficacy and safety.
- Important complications of therapy for chronic hypophosphatemic conditions, especially those mediated by FGF23, include nephrocalcinosis and hyperparathyroidism.
- Additional risks of anti-FGF23 therapy include restless legs syndrome and injection site reactions.

Autosomal Recessive Hypophosphatemic Rickets (OMIM No. 241520). ARHR is genetically heterogeneous, as ARHR has been found to be the result of inactivating mutations in the dentin matrix protein-1 gene (*DMP1*), mutations in the ectonucleotide pyrophosphatase/phosphodiesterase 1 gene (*ENPP1*), and mutations in the family with sequence similarity 20, member C (*FAM20C*) gene (which causes a variant of Raine syndrome). DMP1 is a member of the **s**mall **i**ntegrin-**b**inding ***li***gand, N-linked **g**lycoprotein (SIBLING) family and is highly and specifically expressed in osteocytes. Both *Dmp1*-null mice and patients with ARHR manifest rickets and osteomalacia with isolated renal phosphate wasting associated with elevated FGF23. Mutational analyses revealed that one ARHR family carried a mutation that ablated the DMP1 start codon (Met1Val), and a second family exhibited a deletion in the DMP1 C-terminus (1484–1490del).[79]

Mechanistic studies using the *Dmp1*-null mouse have demonstrated that loss of Dmp1 impairs osteocyte maturation, leading to markedly elevated FGF23 mRNA and protein expression. The hypophosphatemia results in pathological changes in bone mineralization[79] that can be largely, but not completely, abrogated by a high-phosphate diet in *Dmp1*-null mice. Importantly, *Dmp1*-null mice are biochemical phenocopies of the Hyp mouse, and patients with ARHR and XLH (as well as *Dmp1*-null and Hyp mice) share a distinctive bone histology characterized by periosteocytic lesions of nonmineralized bone.[79] Thus, these findings suggest that PHEX (mutated in XLH) may also have a role in osteocyte maturation in a parallel pathway to DMP1 that leads to overexpression of FGF23.

ENPP1 encodes a type II transmembrane glycoprotein that cleaves a variety of substrates, including phosphodiester bonds of nucleotides and nucleotide sugars and pyrophosphate bonds of nucleotides and nucleotide sugars. ENPP1 likely functions to hydrolyze nucleoside 5' triphosphates to monophosphates, and also hydrolyzes diadenosine polyphosphates. ARHR patients who have *ENPP1* homozygous missense mutations have modestly elevated intact FGF23 and hypophosphatemia.[80] In bone, ENPP1 hydrolyzes ATP to pyrophosphate (PPi); therefore, the ENPP1 mutations could indicate that a balance of PP_i to inorganic phosphate phosphate may have a role in FGF23 production. Mutations in this gene are also known to be causative for generalized infantile arterial calcification and ossification of the posterior longitudinal ligament of the spine.

ARHR is also a variant of the typically lethal Raine syndrome, which due to loss-of-function mutations in the *FAM20C* gene (also known as Golgi casein kinase). Through exome sequencing, it was determined that compound heterozygous mutations in *FAM20C* were associated with increases in plasma levels of intact FGF23 but tapering hypophosphatemia with age, severe tooth decay from infancy, ectopic brain calcifications, and osteosclerosis.[81] Interestingly, FAM20C knockout mice are characterized by elevated FGF23, hypophosphatemia with inappropriately normal 1,25(OH)2 vitamin D concentrations, severe tooth defects, and a rickets phenotype.[82] Due to the noted similarities between the DMP1 knockout mice and FAM20C knockout mice, it was discovered that DMP1 levels were reduced in FAM20C knockout cells, perhaps producing the parallel phenotypes between these murine models. It was determined that FAM20C is an "atypical" kinase localized to the Golgi apparatus that phosphorylates a number of secreted SIBLING family members, including DMP1.[83] Thus, the lack of FAM20C may reduce DMP1 phosphorylation and inhibit its expression or function, thereby producing a "DMP1 knockout–like" phenotype in the FAM20C knockout mice. Why patients with ARHR type 3 and Raine syndrome have an osteosclerotic disease while FAM20C knockout mice have rickets is currently unclear.

Other Disorders Involving Fibroblast Growth Factor-23

Fibrous Dysplasia. Fibrous dysplasia (FD) is a disorder caused by activating somatic mutations in the *GNAS1* gene that encodes the α subunit of the stimulatory G protein, G_s.[84] The skeletal defects in FD are characterized by fibrous lesions and colocalized mineralization defects, which contribute to the morbidity in these patients. A significant proportion of FD patients also manifest various degrees of renal phosphate wasting and subsequent hypophosphatemia, which can lead to hypophosphatemic rickets and osteomalacia. Extraskeletal clinical manifestations of FD can occur, such as abnormal skin pigmentation, premature puberty, and hyperthyroidism; the disease is then referred to as *McCune–Albright syndrome* (MAS). In one study of FD patients, increased FGF23 serum levels correlated negatively to serum phosphate and $1,25(OH)_2D_3$ but positively to skeletal disease burden.[31] FGF23 mRNA and protein were localized to fibrous cells in the fibrous bone lesions of FD, as well as osteogenic and endothelial cells associated with microvascular walls in bone.[31] Therefore, it is evident that FGF23 plays an important role in the pathogenesis of the phosphate wasting that is often seen in FD/MAS.

The reason why the bone cells associated with lesions in FD overexpress FGF23 is not understood. Normally, FGF23 is expressed in osteogenic cells, but it is possible that improperly differentiated cells of FD lesions may have lost regulatory mechanisms required to suppress FGF23 production. Speculatively, this is analogous to FGF23 overexpression in TIO tumors, which are not subject to normal regulatory mechanisms.

Osteoglophonic Dysplasia (OMIM No. 166250). Activating mutations in fibroblast growth factor receptors 1 to 3 (FGFR1–3) are responsible for a diverse group of skeletal disorders. In general, mutations in FGFR1-2 cause syndromes involving craniosynostosis, whereas the dwarfing syndromes are associated with FGFR3 mutations. Osteoglophonic dysplasia (OGD) is a "crossover" disorder that has skeletal phenotypes associated with FGFR1 and FGFR2 mutations, as well as with FGFR3 mutations. In this regard, OGD patients present with craniosynostosis, prominent supraorbital ridge, and depressed nasal bridge, as well as with rhizomelic dwarfism and nonossifying bone lesions that show some similarity on x-ray to FD lesions. OGD is an autosomal dominant disorder that is caused by missense mutations in a highly conserved region comprising the ligand-binding and transmembrane domain of FGFR1, thus defining novel roles for this receptor as a negative regulator of long-bone growth. Of significance, three out of the four OGD patients studied had isolated renal phosphate wasting with inappropriately low $1,25(OH)_2D$ concentrations.[85] In one of these patients, a sample was available for analysis of plasma FGF23 concentrations, which were significantly elevated above control levels.[5] It was hypothesized that the associated metaphyseal lesions, which may be similar to FD lesions, produce FGF23, which causes renal phosphate wasting. Although only a few patients have been assessed, owing to the rarity of the disorder, OGD may have parallels with FD whereby the lesional burden of a patient is proportional to the FGF23 production and the extent of phosphate wasting. Interestingly, FGF23 production is stimulated by activation of FGFR1 in osteoblasts or osteocytes.[86]

Linear Nevus Sebaceous Syndrome. Linear nevus sebaceous syndrome (LNSS) is a rare congenital disorder involving cutaneous lesions characterized by papillomatous epidermal hyperplasia and excess sebaceous glands. Additional aberrations are often present in LNSS patients, including developmental defects of the brain that are associated with seizures, as well as eye complications. Hypophosphatemic rickets, may also present as part of LNSS, which usually presents within the first years of life, often with rickets, skeletal pain, and insufficiency fractures. LNSS has been associated with postzygotic activating mutations in three different genes: *HRAS*, *KRAS*, and *NRAS*. Patients with hypophosphatemia have elevated serum FGF23. Reports of improvements in hypophosphatemia following resection of skin lesions are mixed, and it is likely that the actual source of high FGF23 in these patients is the skeleton, as there are FD-like lesions in bone.[87]

αKlotho Overexpression. A case of severe hypophosphatemic rickets with hyperparathyroidism was identified in which the patient was negative for mutations in other genes responsible for heritable hypophosphatemias, including *FGF23*, *DMP1*, and *PHEX*. The patient was found to have a *de novo* translocation (t9:13) with a breakpoint adjacent to the gene encoding αKlotho (KL),[88] which is a single-pass membrane-bound renal coreceptor for FGF23. This chromosomal abnormality was associated with elevated circulating levels of cleaved or "cut" form of alpha-Klotho (cKL) that is known to result from proteolytic cleavage of KL at its extracellular juxtatransmembrane domain to produce a circulating isoform (see later for description of KL isoforms). The translocation patient had markedly elevated FGF23 in the face of prevailing hypophosphatemia, and hyperparathyroidism that necessitated gland debulking.[88] When cKL was delivered to mice via adenoassociated virus, a parallel phenotype was observed in the murine model characterized by markedly increased FGF23 despite severe hypophosphatemia. It was determined that elevated blood cKL resulted in increased FGF23 mRNA production in bone (150-fold), which would explain the high FGF23 concentrations despite low blood phosphate. *In vitro* studies indicated that cKL–FGFR interactions were involved in mediating cKL bioactivity.[89]

Disorders Associated With Reduced Fibroblast Growth Factor-23 Bioactivity

Hyperphosphatemic Familial Tumoral Calcinosis (OMIM No. 211900). Familial TC is a heritable autosomal recessive disorder characterized by ectopic calcified tumoral masses, dental abnormalities, and soft-tissue ectopic and vascular calcification. Biochemical abnormalities include hyperphosphatemia, increased TMP/GFR, and inappropriately normal or elevated $1,25(OH)_2D$ concentrations. Calcium and PTH are typically within the normal reference range, although suppressed PTH levels also occur. Biochemically, TC is the mirror image of the phosphate-wasting disorders XLH and ADHR. HHS is a rare disorder in which patients present with a similar biochemical profile to TC and have localized hyperostosis.[90]

Hyperphosphatemic Familial Tumoral Calcinosis/Hyperostosis-Hyperphosphatemia Syndrome Due to GALNT3 Mutations. The gene first identified for the heritable form of TC was UDP-*N*-acetyl-alpha-D-galactosamine: polypeptide *N*-acetylgalactos-aminyl transferase-3 (*GALNT3*), in which inactivating mutations were detected.[62] GALNT3 is expressed in the trans-Golgi network and initiates O-linked glycosylation of nascent proteins. In another family with a reported autosomal dominant TC, the clinical symptoms were in fact caused by two different biallelic *GALNT3* mutations, further supporting an autosomal recessive inheritance of this disorder.[89] The patients with TC due to *GALNT3* mutations were originally reported to have serum FGF23 levels approximately 30-fold above the normal mean when assessed with the C-terminal FGF23 ELISA.[62] Importantly, it was demonstrated that the TC patients did indeed have elevated C-terminal FGF23, but the same individuals had low circulating FGF23 when measured with the intact FGF23 ELISA.[1] These findings were then confirmed by demonstrating that loss of GALNT3 resulted in the production of nonfunctional FGF23 protein due to intracellular degradation.[91] FGF23 is O-glycosylated on residues within the $_{176}RHTR_{179}/S_{180}$ site (specifically at threonine 178); thus, the lack of glycosylation at this residue is thought to expose the SPC site and lead to degradation and destabilization of intact, active FGF23.[92]

HHS is due to inactivating mutations in GALNT3, and these patients also manifest inappropriate ELISA ratios of C-terminal to intact FGF23.[90] It has been shown that some of the HHS mutations are the same as those that result in TC,[90] indicating that TC and HHS represent ranges of severity within the same disease.

Hyperphosphatemic Familial Tumoral Calcinosis Due to FGF23 Mutations. TC is also caused by recessive mutations in the *FGF23* gene. Groups reported mutations giving rise to an amino acid change from a serine to a glycine at residue 71 (S71G),[93] or from a serine to a phenylalanine at residue 129 (S129F).[94] Other missense mutations have also been identified, including and M96T.[96] These FGF23 TC mutations (summarized in Table 53.2) occur within the FGF23 N-terminal FGF-like domain (residues 25–180). The TC alterations apparently destabilize intact FGF23, as supported by the findings that TC patients with FGF23 mutations have the same FGF23 ELISA pattern as GALNT3-TC patients, that is, markedly elevated C-terminal concentrations in concert with low intact values,[93] and the fact that these mutants are proteolytically cleaved prior to cellular secretion.[93,94] This cleavage is in part mediated by SPC proteases, inasmuch as addition of the ADHR mutations in tandem with the S71G TC mutations in the FGF23 cDNA resulted in rescue of full-length FGF23 production.[97] Thus, the common element in GALNT3-TC and FGF23-TC is the lack of secretion of intact FGF23. This lack of active FGF23 then results in elevation of serum phosphate through increased renal reabsorption, which in turn most likely results in elevated secretion of nonfunctional FGF23 fragments through a positive feedback cycle.

Hyperphosphatemic Familial Tumoral Calcinosis Due to Alpha-Klotho Mutations. FGF23 is a member of a class of FGFs including FGF19 and FGF21 that are endocrine, as opposed to paracrine/autocrine, factors. FGF23 requires the coreceptor KL for bioactivity, whereas FGF19 and FGF21 require beta-Klotho.[98,99] KL-null mice have severe calcifications, as well as markedly elevated serum phosphate,[100] which parallels Fgf23-null mice[101] and TC patients. Both the KL-null and Fgf23-null mice, however, have more severe phenotypes than that observed in TC patients. Importantly, the defects in the KL-null and Fgf23-null mice can largely be ameliorated with a low-phosphate diet to reduce serum phosphate. In parallel with Fgf23-null mice, KL-null mice have increased NPT2a in the proximal tubule,[102] indicating that the hyperphosphatemia in these animals is secondary to increased renal reabsorption of phosphate.

KL is produced as two isoforms due to alternative splicing of the same gene. Membrane-bound KL (mKL) is a 130-kD single-pass transmembrane protein characterized by a large extracellular domain and a very short (10-residue) intracellular domain that does not possess signaling capabilities. A secreted form of KL (sKL) is approximately 80 kD and arises from splicing within the *KLOTHO* gene exon 3 to result in a KL protein that does not contain the transmembrane domain and could be secreted into the circulation.[103] This form of KL, however, was not detected in human or mouse blood or cerebrospinal fluid.[104] The primary circulating form of KL is cKL, which results from proteolytic cleavage of mKL at its extracellular juxtatransmembrane domain by ADAM10/17 and BACE protease activity.[105]

The most likely mechanism for FGF23 signaling through KL is the recruitment of canonical FGFRs to form active signaling complexes. One group has identified a specific complex between FGFR1c and KL.[98] In contrast, others demonstrated that multiple FGFRs (FGFR1c, FGFR3c, and FGFR4) can interact with KL and FGF23 and signal through mitogen-activated protein kinase cascades.[106] Within the kidney, KL localizes to the proximal and distal tubules.[102,107] FGF23 is known to mediate its effects on NPT2a, NPT2c, and the vitamin D metabolizing enzymes within the proximal tubule.[21,51] FGF23 may also have effects on maintaining calcium and sodium reabsorption in the distal tubule. Because KL is a coreceptor for FGF23, mutations in the gene encoding this protein were sought as candidate causes for TC in a 13-year-old female with hypothesized end-organ resistance to renal FGF23 activity. This patient presented with hyperphosphatemia, hypercalcemia, elevated PTH, elevated intact and C-terminal FGF23 over 100- to 550-fold elevation of the normal means[108] (in contrast to the differential C-terminal to intact ELISA ratios in GALNT3- and FGF23-TC), and ectopic calcifications in the heel and brain. She had normal pubertal development, and her disease paralleled KL-null mice with regard to ectopic calcifications and dramatic elevation of circulating FGF23.[99] This patient was shown to have a recessive mutation in a highly conserved residue (Histidine193Arginine, or H193R) in the extracellular domain of KL (the "KL1" domain). Mutant KL expression was markedly reduced *in vitro* compared to that of wild-type KL, which resulted in a striking reduction in the ability of KL to mediate FGF23-dependent signaling.[108] Thus, an inactivating H193R KL mutation results in a TC phenotype and demonstrates that KL is required for FGF23 bioactivity.

In sum, it is clear that, in human disorders of phosphate handling and the corresponding mouse models, increased FGF23 bioactivity is associated with renal phosphate wasting and inappropriately normal $1,25(OH)_2D$ concentrations, and that the converse biochemical relationships exist, with loss of FGF23 bioactivity in the hormone itself or in its receptor complex (summarized in Fig. 53.4).

Genetic Hypophosphatemia Not Mediated by Elevated FGF23

Several genetic conditions exist that cause renal phosphate losses and hypophosphatemia but are not mediated by elevated FGF23. Some examples are described in this section. A common feature of these conditions is that they are often associated with hypercalciuria, either due to compensatory rises in $1,25(OH)_2D$ concentrations in response to hypophosphatemia or due to concurrent renal calcium losses due to a generalized tubulopathy. In addition, some of these conditions, such as Dent disease and nephropathic cystinosis, are associated with progressive renal failure.

Hereditary Hypophosphatemic Rickets With Hypercalciuria.

Similar to the disease entities described earlier, HHRH is characterized by hypophosphatemia; however, HHRH is distinguished by elevated 1,25(OH)D concentrations and suppressed PTH. The distinguishing biochemical characteristic of this disorder compared with the FGF23-related diseases such as ADHR and XLH is the notable increase in serum 1,25(OH)D as a compensatory response to the prevailing

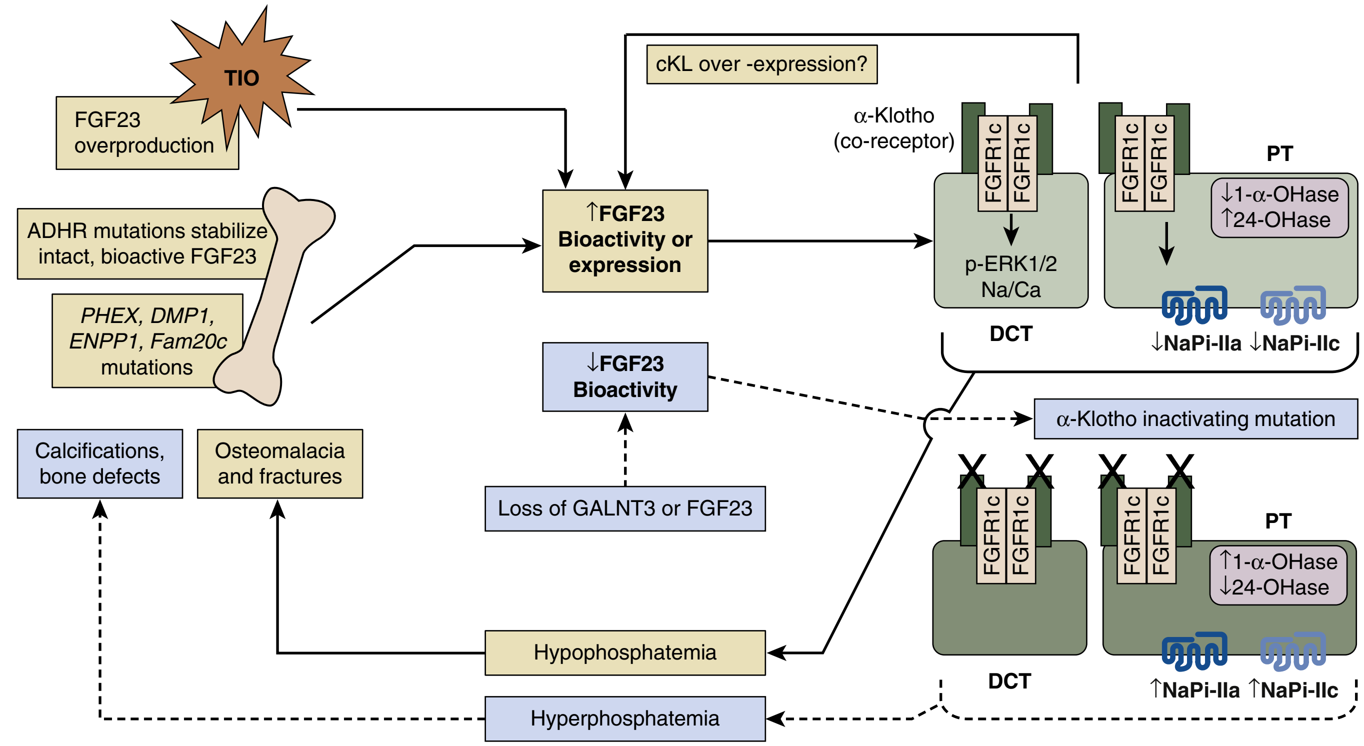

TIO
FGF23 overproduction
cKL over -expression?
ADHR mutations stabilize intact, bioactive FGF23
PHEX, DMP1, ENPP1, Fam20c mutations
↑FGF23 Bioactivity or expression
α-Klotho (co-receptor)
FGFR1c
FGFR1c
FGFR1c
FGFR1c
PT
↓1-α-OHase
↑24-OHase
p-ERK1/2
Na/Ca
DCT
↓NaPi-IIa ↓NaPi-IIc
↓FGF23 Bioactivity
α-Klotho inactivating mutation
Calcifications, bone defects
Osteomalacia and fractures
Loss of GALNT3 or FGF23
FGFR1c
FGFR1c
FGFR1c
FGFR1c
PT
↑1-α-OHase
↓24-OHase
Hypophosphatemia
DCT
↑NaPi-IIa ↑NaPi-IIc
Hyperphosphatemia

Fig. 53.4 Genes involved in fibroblast growth factor-23 (FGF23) activity. Summary of genes involved in gain of FGF23 bioactivity *(light gray boxes)* through intrinsic stabilization (*FGF23* gene [ADHR]) or overproduction (TIO; *PHEX* gene [XLH]; *DMP1* gene [ARHR]; *ENPP1* gene [ARHR]; FAM20c gene [Raine syndrome, ARHR]; cleaved alpha-Klotho overexpression) *(solid-line pathways),* and involved in loss of FGF23 bioactivity *(dark gray boxes)* through mutations that compromise intact FGF23 production (*FGF23* and *GALNT3* genes) or its signaling complex (*KL* gene) *(dashed-line pathways). DCT,* Distal convoluted tubule; *PT,* proximal tubule; *TIO,* tumor-induced osteomalacia; *ADHR,* autosomal dominant hypophosphatemic rickets.

hypophosphatemia. Affected individuals also have markedly elevated urine calcium excretion, with increased intestinal absorption of phosphorus and calcium that can lead to nephrolithiasis. Clinically, patients with HHRH have rickets and short stature. Additional analysis of this kindred revealed 21 members with "intermediate" phenotypes, which led to an initial hypothesis that HHRH was inherited through an autosomal codominant mode. This group of patients manifested hypercalciuria, mild hypophosphatemia, and elevated calcitriol levels but no apparent bony abnormalities.[109]

Using a genome-wide linkage scan in combination with homozygosity mapping, the disorder was mapped to a 1.6-Mb region on chromosome 9q34, which contains the *SLC34A3* gene that encodes NPT2c.[110] A recessive, single-nucleotide deletion was then identified in this gene (228delC). Subsequently, multiple additional loss of function mutations have been identified in NPT2c causing HHRH. One of the key physiologic implications of identifying disease-causing mutations in NPT2c was the fact that this transporter now appears to have a significant role in renal phosphate reabsorption throughout life and is not limited to perinatal phosphate homeostasis as previously hypothesized. The fact that NPT2c mutations lead to disease in humans but not in mice[18] may indicate that this transporter plays a more important role in humans than in mice.

Dent Disease. In contrast to XLH, which is X-linked dominant, Dent disease is an X-linked recessive form of hyphophosphatemia. Dent disease is also associated with hypophosphatemic rickets, but, because it is not FGF23-mediated, the $1,25(OH)_2D$ is elevated, with associated hypercalciuria. Dent disease has been demonstrated to be caused by inactivating mutations in the gene for chloride channel 5 or in *OCRL1*, and the resulting protein abnormalities impair lysosomal trafficking, causing renal tubulopathies.[111] Nephrocalcinosis and nephrolithiasis are important consequences of the hypercalciuria. Proteinuria is also a prominent feature, and consequently evaluation for proteinuria should be an important part of the clinical evaluation of children presenting with hypophosphatemia. Patients with Dent disease develop chronic kidney disease, which often progresses to end-stage kidney disease by adulthood.

Nephropathic Cystinosis. Nephropathic cystinosis is a systemic autosomal recessive disease caused by mutations in the *CTNS* gene that ecnodes the lysosomal cystine transporter cystinosin. Deficiency of cystinosin causes cystine to accumulate in lysosomes in many organs and tissues, resulting in tissue damage.[112] In the nephropathic form, Fanconi syndrome develops in infancy, resulting in urinary losses of electrolytes, glucose, amino acids, bicarbonate, calcium, and phosphate. The resulting hypophosphatemic rickets is complicated by hypercalciuria, and sometimes hypocalcemia may also occur. The resulting metabolic bone disease causes rickets and poor linear growth. Corneal cystine crystals can impair vision, while accumulation in the thyroid leads to hypothyroidism. Additional features can include a vacuolar myopathy and hypogonadism. The rickets is treated with phosphate, and sometimes active vitamin D may be added, but patients must be carefully monitored for hypercalciuria, which can limit this treatment.[111] Early recognition of cystinosis is important, because treatment with oral cysteamine can delay end-organ damage in the kidney and other tissues. Cysteamine enters lysosomes and converts cystine to cysteine and cysteine–cysteamine disulfide, which can then be transported out of the cell, thus removing the intracellular cystine.

However, cysteamine treatment does not generally alter the Fanconi syndrome.

Making these diagnoses is important, as they alter the management. In contrast to the FGF23-mediated hypophosphatemia associated with XLH or ADHR, phosphate supplementation is an effective treatment for HHRH, but oral calcitriol is contraindicated; calcitriol levels in this disease are already elevated, and hypercalciuria may worsen. However, treating the bone disease in HHRH may be limited by tolerability of the phosphate salts. Nephropathic cystinosis is treated with oral cysteamine, in addition to management of mineral metabolism abnormalities. Burosumab is not recommended in conditions that are not FGF23-mediated.

KEY POINTS

- Rare diseases of fibroblast growth factor-23 (FGF23) overproduction are due to a heterogeneous group of mutations.
- Hyperphosphatemic falmilial tumoral calcinosis is cause by mutations in FGF23 itself, alpha-Klotho, and genes responsible for intracellular processing of FGF23.
- Some diseases of phosphate metabolism are independent of FGF23, and these should be distinguished at the genetic level to ensure proper treatment courses.

POSSIBLE FUTURE THERAPEUTIC AVENUES

As noted above, the use of a neutralizing monoclonal antibody to FGF23 in XLH and TIO has become a reality. Whether a longer-term infusion derivative of burosumab would provide benefit remains to be determined. For hyperphosphatemic disorders, the most direct potential application for treatment is the use of recombinant FGF23 in hfTC. Several groups have demonstrated that inactivating mutations in *FGF23* lead to hfTC; thus, delivering recombinant FGF23 may completely resolve the disorder by directly treating the molecular defect through replacement of missing or mutant FGF23. Additional data are required to determine if FGF23 would be a potential treatment for GALNT3-mediated TC. Whether ADHR-mutant FGF23 (mutant at positions 176 and/or 179) would provide a "longer-acting" form of therapy as a result of stabilization of the full-length polypeptide compared to the labile wild-type form remains to be determined.

Acknowledgments

The authors would like to acknowledge the funding support of the National Institutes of Health through grants R21AR059278, R01DK112958, R01HL145528 (KEW), and AR068615 (MJE), as well as the Indiana CTSI. We thank Dr. F. Richard Bringhurst for contributions to earlier editions of this chapter.

REFERENCES

1. Walser M. Ion association. VI. Interactions between calcium, magnesium, inorganic phosphate, citrate and protein in normal human plasma. *J Clin Invest*. 1961;40:723–730.
2. Somell A, Alveryd A. Diurnal variations in the urinary excretion of calcium and phosphate in hyperparathyroidism. *Acta Chir Scand*. 1976;142:357–359.
3. Adam A, et al. Colorimetric vs enzymatic determination of serum phosphorus. *Clin Chem*. 1984;30:1724–1725.
4. Landowne RA. Immunoglobulin interference with phosphorus and chloride determinations with the Coulter Chemistry. *Clin Chem*. 1979;25:1189–1190.
5. Donhowe JM, et al. Factitious hypophosphatemia related to mannitol therapy. *Clin Chem*. 1981;27:1765–1769.
6. Reitz RE, et al. Calcium, magnesium, phosphorus, and parathyroid hormone interrelationships in pregnancy and newborn infants. *Obstet Gynecol*. 1977;50:701–705.
7. Lian JB, et al. Changes in phosphoproteins of chicken bone matrix in vitamin D-deficient rickets. *FEBS Lett*. 1982;149:123–125.
8. Marxhall DH, Nordin BE, Speed R. Calcium, phosphorus and magnesium requirement. *Proc Nutr Soc*. 1976;35:163–173.
9. Wilz DR, et al. Plasma 1,25-(OH)2-vitamin D concentrations and net intestinal calcium, phosphate, and magnesium absorption in humans. *Am J Clin Nutr*. 1979;32:2052–2060.
10. Wasserman RH, Taylor AN. Intestinal absorption of phosphate in the chick: effect of vitamin D and other parameters. *J Nutr*. 1973;103:586–599.
11. Feild JA, et al. Cloning and functional characterization of a sodium-dependent phosphate transporter expressed in human lung and small intestine. *Biochem Biophys Res Commun*. 1999;258:578–582.
12. Xu H, et al. Age-dependent regulation of rat intestinal type IIb sodium-phosphate cotransporter by 1,25-(OH)(2) vitamin D(3). *Am J Physiol Cell Physiol*. 2002;282:C487–C493.
13. Dennis VW, Brazy PC. Divalent anion transport in isolated renal tubules. *Kidney Int*. 1982;22:498–506.
14. Cheng L, Sacktor B. Sodium gradient-dependent phosphate transport in renal brush border membrane vesicles. *J Biol Chem*. 1981;256:1556–1564.
15. Beck L, et al. Targeted inactivation of Npt2 in mice leads to severe renal phosphate wasting, hypercalciuria, and skeletal abnormalities. *Proc Natl Acad Sci U S A*. 1998;95:5372–5377.
16. Hoag HM, et al. Effects of Npt2 gene ablation and low-phosphate diet on renal Na(+)/phosphate cotransport and cotransporter gene expression. *J Clin Invest*. 1999;104:679–686.
17. Zhao N, Tenenhouse HS. Npt2 gene disruption confers resistance to the inhibitory action of parathyroid hormone on renal sodium-phosphate cotransport. *Endocrinology*. 2000;141:2159–2165.
18. Segawa H, Onitsuka A, Kuwahata M, Hanabusa E, Furutani J, Kaneko I, Tomoe Y, Aranami F, Matsumoto N, Ito M, Matsumoto M, Li M, Amizuka N, Miyamoto K. Type IIc sodium-dependent phosphate transporter regulates calcium metabolism. *J Am Soc Nephrol*. 2009 Jan;20(1):104–13.
19. Walton RJ, Bijvoet OL. Nomogram for derivation of renal threshold phosphate concentration. *Lancet*. 1975;2:309–310.
20. Berndt T, et al. Evidence for a signaling axis by which intestinal phosphate rapidly modulates renal phosphate reabsorption. *Proc Natl Acad Sci U S A*. 2007;104:11085–11090.
21. Larsson T, et al. Transgenic mice expressing fibroblast growth factor 23 under the control of the alpha1(I) collagen promoter exhibit growth retardation, osteomalacia, and disturbed phosphate homeostasis. *Endocrinology*. 2004;145:3087–3094.
22. Shimada T, et al. FGF-23 is a potent regulator of vitamin D metabolism and phosphate homeostasis. *J Bone Miner Res*. 2004;19:429–435.
23. Burnett SM, et al. Regulation of C-terminal and intact FGF-23 by dietary phosphate in men and women. *J Bone Miner Res*. 2006;21:1187–1196.
24. Perwad F, et al. Dietary and serum phosphorus regulate fibroblast growth factor 23 expression and 1,25-dihydroxyvitamin D metabolism in mice. *Endocrinology*. 2005;146:5358–5364.
25. Jonsson KB, et al. Fibroblast growth factor 23 in oncogenic osteomalacia and X-linked hypophosphatemia. *N Engl J Med*. 2003;348:1656–1663.
26. Yamazaki Y, et al. Increased circulatory level of biologically active full-length FGF-23 in patients with hypophosphatemic rickets/osteomalacia. *J Clin Endocrinol Metab*. 2002;87:4957–4960.
27. Khosravi A, et al. Determination of the elimination half-life of fibroblast growth factor-23. *J Clin Endocrinol Metab*. 2007;92:2374–2377.
28. El-Maouche D, et al. Stability and degradation of fibroblast growth factor 23 (FGF23): the effect of time and temperature and assay type. *Osteoporos Int*. 2016;27:2345–2353.
29. Econs MJ, McEnery PT. Autosomal dominant hypophosphatemic rickets/osteomalacia: clinical characterization of a novel renal phosphate-wasting disorder. *J Clin Endocrinol Metab*. 1997;82:674–681.
30. ADHR Consortium. Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23. *Nat Genet*. 2000;26:345–348.

31. Riminucci M, et al. FGF-23 in fibrous dysplasia of bone and its relationship to renal phosphate wasting. *J Clin Invest*. 2003;112:683–692.
32. Liu S, et al. Regulation of fibroblastic growth factor 23 expression but not degradation by PHEX. *J Biol Chem*. 2003;278:37419–37426.
33. White KE, et al. Autosomal-dominant hypophosphatemic rickets (ADHR) mutations stabilize FGF-23. *Kidney Int*. 2001;60:2079–2086.
34. Shimada T, et al. Mutant FGF-23 responsible for autosomal dominant hypophosphatemic rickets is resistant to proteolytic cleavage and causes hypophosphatemia in vivo. *Endocrinology*. 2002;143:3179–3182.
35. Gribaa M, et al. An autosomal dominant hypophosphatemic rickets phenotype in a Tunisian family caused by a new FGF23 missense mutation. *J Bone Miner Metab*. 2010;28:111–115.
36. Negri AL, et al. Different forms of clinical presentation of an autosomal dominant hypophosphatemic rickets caused by a FGF23 mutation in one family. *Medicina (B Aires)*. 2004;64:103–106.
37. Seidah NG, Chretien M. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res*. 1999;848:45–62.
38. Molloy SS, et al. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J Biol Chem*. 1992;267:16396–16402.
39. Nakayama K. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem J*. 1997;327:625–635.
40. Imel EA, Hui SL, Econs MJ. FGF23 concentrations vary with disease status in autosomal dominant hypophosphatemic rickets. *J Bone Miner Res*. 2007;22:520–526.
41. Imel EA, et al. Iron modifies plasma FGF23 differently in autosomal dominant hypophosphatemic rickets and healthy humans. *J Clin Endocrinol Metab*. 2011;96:3541–3549.
42. Farrow EG, et al. Iron deficiency drives an autosomal dominant hypophosphatemic rickets (ADHR) phenotype in fibroblast growth factor-23 (Fgf23) knock-in mice. *Proc Natl Acad Sci U S A*. 2011;108:E1146–E1155.
43. Clinkenbeard EL, et al. Neonatal iron deficiency causes abnormal phosphate metabolism by elevating FGF23 in normal and ADHR mice. *J Bone Miner Res*. 2013;29:361–369.
44. David V, et al. Inflammation and functional iron deficiency regulate fibroblast growth factor 23 production. *Kidney Int*. 2015;22:1020–1032.
45. Clinkenbeard EL, et al. Erythropoietin stimulates murine and human fibroblast growth factor-23, revealing novel roles for bone and bone marrow. *Haematologica*. 2017;102:e427–e430.
46. Wolf M, Koch TA, Bregman DB. Effects of iron deficiency anemia and its treatment on fibroblast growth factor 23 and phosphate homeostasis in women. *J Bone Miner Res*. 2013;28:1793–1803.
47. Imel EA, et al. Oral iron replacement normalizes fibroblast growth factor 23 in iron-deficient patients with autosomal dominant hypophosphatemic rickets. *J Bone Miner Res*. 2020;35:231–238.
48. Ryan EA, Reiss E. Oncogenous osteomalacia. Review of the world literature of 42 cases and report of two new cases. *Am J Med*. 1984;77:501–512.
49. Schapira D, et al. Tumor-induced osteomalacia. *Semin Arthritis Rheum*. 1995;25:35–46.
50. White KE, et al. The autosomal dominant hypophosphatemic rickets (ADHR) gene is a secreted polypeptide overexpressed by tumors that cause phosphate wasting. *J Clin Endocrinol Metab*. 2001;86:497–500.
51. Shimada T, et al. Cloning and characterization of FGF23 as a causative factor of tumor-induced osteomalacia. *Proc Natl Acad Sci U S A*. 2001;98:6500–6505.
52. Carpenter TO, et al. A clinician's guide to X-linked hypophosphatemia. *J Bone Miner Res*. 2011;26:1381–1388.
53. Consortium PEX. A gene (PEX) with homologies to endopeptidases is mutated in patients with X-linked hypophosphatemic rickets. The HYP Consortium. *Nat Genet*. 1995;11:130–136.
54. Francis F, et al. Genomic organization of the human PEX gene mutated in X-linked dominant hypophosphatemic rickets. *Genome Res*. 1997;7:573–585.
55. Beck L, et al. Pex/PEX tissue distribution and evidence for a deletion in the 3' region of the Pex gene in X-linked hypophosphatemic mice. *J Clin Invest*. 1997;99:1200–1209.
56. Eicher EM, et al. Hypophosphatemia: mouse model for human familial hypophosphatemic (vitamin D-resistant) rickets. *Proc Natl Acad Sci U S A*. 1976;73:4667–4671.
57. Aono Y, et al. Therapeutic effects of anti-FGF23 antibodies in hypophosphatemic rickets/osteomalacia. *J Bone Miner Res*. 200924(11):1879–88.
58. Barros NM, et al. Proteolytic processing of osteopontin by PHEX and accumulation of osteopontin fragments in Hyp mouse bone, the murine model of X-linked hypophosphatemia. *J Bone Miner Res*. 2013;28:688–699.
59. Ichikawa S, et al. A Phex mutation in a murine model of X-linked hypophosphatemia alters phosphate responsiveness of bone cells. *J Bone Miner Res*. 2012;27:453–460.
60. Imel EA, et al. Iron and fibroblast growth factor 23 in X-linked hypophosphatemia. *Bone*. 2014;60:87–92.
61. Imel EA et al, J Clin Endocrinol Metab 2010 Vol. 95 Issue 4 Pages 1846-1850.
62. Topaz O, et al. Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nat Genet*. 2004;36:579–581.
63. Zivicnjak M, et al. Age-related stature and linear body segments in children with X-linked hypophosphatemic rickets. *Pediatr Nephrol*. 2011;26:223–231.
64. Skrinar A, et al. The lifelong impact of X-linked hypophosphatemia: results from a burden of disease survey. *J Endocr Soc*. 2019;3:1321–1334.
65. Connor J, et al. Conventional therapy in adults with X-linked hypophosphatemia: effects on enthesopathy and dental disease. *J Clin Endocrinol Metab*. 2015;100:3625–3632.
66. Portale AA, et al. Continued beneficial effects of burosumab in adults with X-linked hypophosphatemia: results from a 24-week treatment continuation period after a 24-week double-blind placebo-controlled period. *Calcif Tissue Int*. 2019;105:271–284.
67. Carpenter TO, et al. Nocturnal hyperparathyroidism: a frequent feature of X-linked hypophosphatemia. *J Clin Endocrinol Metab*. 1994;78:1378–1383.
68. Imel EA, et al. Treatment of X-linked hypophosphatemia with calcitriol and phosphate increases circulating fibroblast growth factor 23 concentrations. *J Clin Endocrinol Metab*. 2010;95:1846–1850.
69. DeLacey S, et al. Hyperparathyroidism and parathyroidectomy in X-linked hypophosphatemia patients. *Bone*. 2019;127:386–392.
70. Aono Y, et al. Therapeutic effects of anti-FGF23 antibodies in hypophosphatemic rickets/osteomalacia. *J Bone Miner Res*. 2009;24:1879–1888.
71. Imel EA, et al. Prolonged correction of serum phosphorus in adults with X-linked hypophosphatemia using monthly doses of KRN23. *J Clin Endocrinol Metab*. 2015;100:2565–2573.
72. Carpenter TO, et al. Burosumab therapy in children with X-linked hypophosphatemia. *N Engl J Med*. 2018;378:1987–1998.
73. Whyte MP, et al. Efficacy and safety of burosumab in children aged 1-4 years with X-linked hypophosphataemia: a multicentre, open-label, phase 2 trial. *Lancet Diabetes Endocrinol*. 2019;7:189–199.
74. Imel EA, et al. Burosumab versus conventional therapy in children with X-linked hypophosphataemia: a randomised, active-controlled, open-label, phase 3 trial. *Lancet*. 2019;393:2416–2427.
75. Insogna KL, et al. Burosumab improved histomorphometric measures of osteomalacia in adults with X-linked hypophosphatemia: a phase 3, single-arm, international trial. *J Bone Miner Res*. 2019;34:2183–2191.
76. Imanishi Y, et al. Interim analysis of a phase 2 open-label trial assessing burosumab efficacy and safety in patients with tumor-induced osteomalacia. *J Bone Miner Res*. 2021;36(2):262–270.
77. De Beur J et al. Burosumab for the Treatment of Tumor-Induced Osteomalacia. *J Bone Miner Res.* 2021;36(4):627–635.
78. Shalhoub V, et al. FGF23 neutralization improves chronic kidney disease-associated hyperparathyroidism yet increases mortality. *J Clin Invest*. 2012;122(7):2543–2553.
79. Feng JQ, et al. Loss of DMP1 causes rickets and osteomalacia and identifies a role for osteocytes in mineral metabolism. *Nat Genet*. 2006;38(11):1310–1315.

80. Lorenz-Depiereux B, et al. Loss-of-function ENPP1 mutations cause both generalized arterial calcification of infancy and autosomal-recessive hypophosphatemic rickets. *Am J Hum Genet*. 2010;86:267–272.
81. Rafaelsen SH, et al. Exome sequencing reveals FAM20c mutations associated with fibroblast growth factor 23-related hypophosphatemia, dental anomalies, and ectopic calcification. *J Bone Miner Res*. 2013;28:1378–1385.
82. Wang X, et al. Inactivation of a novel FGF23 regulator, FAM20C, leads to hypophosphatemic rickets in mice. *PLoS Genet*. 2012;8. e1002708.
83. Tagliabracci VS, et al. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science*. 2012;336:1150–1153.
84. Weinstein LS, et al. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*. 1991;325:1688–1695.
85. White KE, et al. Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *Am J Hum Genet*. 2005;76:361–367.
86. Han X, Xiao Z, Quarles LD. Membrane and integrative nuclear fibroblastic growth factor receptor (FGFR) regulation of FGF-23. *J Biol Chem*. 2015;290:10447–10459.
87. Ovejero D, et al. Cutaneous skeletal hypophosphatemia syndrome: clinical spectrum, natural history, and treatment. *Osteoporos Int*. 2016;27:3615–3626.
88. Brownstein CA, et al. A translocation causing increased alpha-klotho level results in hypophosphatemic rickets and hyperparathyroidism. *Proc Natl Acad Sci U S A*. 2008;105:3455–3460.
89. Smith RC, et al. Circulating alphaKlotho influences phosphate handling by controlling FGF23 production. *J Clin Invest*. 2012;122:4710–4715.
90. Ichikawa S, et al. Novel GALNT3 mutations causing hyperostosis-hyperphosphatemia syndrome result in low intact fibroblast growth factor 23 concentrations. *J Clin Endocrinol Metab*. 2007;92:1943–1947.
91. Garringer HJ, et al. The role of mutant UDP-N-acetyl-alpha-D-galactosamine-polypeptide N-acetylgalactosaminyltransferase 3 in regulating serum intact fibroblast growth factor 23 and matrix extracellular phosphoglycoprotein in heritable tumoral calcinosis. *J Clin Endocrinol Metab*. 2006;91:4037–4042.
92. Frishberg Y, et al. Hyperostosis-hyperphosphatemia syndrome: a congenital disorder of O-glycosylation associated with augmented processing of fibroblast growth factor 23. *J Bone Miner Res*. 2007;22:235–242.
93. Larsson T, et al. A novel recessive mutation in fibroblast growth factor-23 causes familial tumoral calcinosis. *J Clin Endocrinol Metab*. 2005;90:2424–2427.
94. Araya K, et al. A novel mutation in fibroblast growth factor 23 gene as a cause of tumoral calcinosis. *J Clin Endocrinol Metab*. 2005;90:5523–5527.
95. Garringer HJ, et al. Molecular genetic and biochemical analyses of FGF23 mutations in familial tumoral calcinosis. *Am J Physiol Endocrinol Metab*. 2008;295:E929–E937.
96. Chefetz I, et al. A novel homozygous missense mutation in FGF23 causes Familial Tumoral Calcinosis associated with disseminated visceral calcification. *Hum Genet*. 2005;118:261–266.
97. Larsson T, et al. Fibroblast growth factor-23 mutants causing familial tumoral calcinosis are differentially processed. *Endocrinology*. 2005;146:3883–3891.
98. Urakawa I, et al. Klotho converts canonical FGF receptor into a specific receptor for FGF23. *Nature*. 2006;444:770–774.
99. Kurosu H, et al. Tissue-specific expression of betaKlotho and fibroblast growth factor (FGF) receptor isoforms determines metabolic activity of FGF19 and FGF21. *J Biol Chem*. 2007;282:26687–26695.
100. Tsujikawa H, et al. Klotho, a gene related to a syndrome resembling human premature aging, functions in a negative regulatory circuit of vitamin D endocrine system. *Mol Endocrinol*. 2003;17:2393–2403.
101. Shimada T, et al. Targeted ablation of Fgf23 demonstrates an essential physiological role of FGF23 in phosphate and vitamin D metabolism. *J Clin Invest*. 2004;113:561–568.
102. Li SA, et al. Immunohistochemical localization of Klotho protein in brain, kidney, and reproductive organs of mice. *Cell Struct Funct*. 2004;29:91–99.
103. Matsumura Y, et al. Identification of the human klotho gene and its two transcripts encoding membrane and secreted klotho protein. *Biochem Biophys Res Commun*. 1998;242:626–630.
104. Imura A, et al. Secreted Klotho protein in sera and CSF: implication for post-translational cleavage in release of Klotho protein from cell membrane. *FEBS Lett*. 2004;565:143–147.
105. Bloch L, et al. Klotho is a substrate for alpha-, beta- and gamma-secretase. *FEBS Lett*. 2009;583:3221–3224.
106. Kurosu H, et al. Regulation of fibroblast growth factor-23 signaling by klotho. *J Biol Chem*. 2006;281:6120–6123.
107. Olauson H, et al. Tissue expression and source of circulating alphaKlotho. *Bone*. 2017;100:19–35.
108. Ichikawa S, et al. A homozygous missense mutation in human KLOTHO causes severe tumoral calcinosis. *J Clin Invest*. 2007;117:2684–2691.
109. Tieder M, et al. “Idiopathic” hypercalciuria and hereditary hypophosphatemic rickets. Two phenotypical expressions of a common genetic defect. *N Engl J Med*. 1987;316:125–129.
110. Bergwitz C, et al. SLC34A3 mutations in patients with hereditary hypophosphatemic rickets with hypercalciuria predict a key role for the sodium-phosphate cotransporter NaPi-IIc in maintaining phosphate homeostasis. *Am J Hum Genet*. 2006;78:179–192.
111. Devuyst O, Thakker RV. Dent’s disease. *Orphanet J Rare Dis*. 2010;5:28.
112. Hohenfellner K, et al. Management of bone disease in cystinosis: Statement from an international conference. *J Inherit Metab Dis*. 2019;42:1019–1029.

54

Primary Hyperparathyroidism

Marcella D. Walker and John P. Bilezikian

OUTLINE

Introduction, 922
Risk Factors, Pathology, and Anatomical Location, 923
Epidemiology, Diagnosis, and Overview of Clinical Features, 924
  Incidence, 924
  Diagnosis and Differential Diagnosis, 924
  Other Biochemical Features, 925
  Clinical Presentation, 925
The Skeleton, 926
  Bone Turnover Markers, 926
  Bone Densitometry, 926
  Bone Histomorphometry, 927
  New Imaging Technologies, 928
  Fractures, 928
Nephrolithiasis and Renal Function, 930
Other Organ Involvement, 930
  Cardiovascular System, 930
  Neurological, Psychologic and Cognitive Features, 931
  Gastrointestinal Manifestations, 931
  Other Systemic Involvement, 932
Unusual Presentations, 932
  Neonatal Primary Hyperparathyroidism, 932
  Primary Hyperparathyroidism in Pregnancy, 932
  Acute Primary Hyperparathyroidism, 932
  Parathyroid Cancer, 932
Evaluation, 933
Natural History, 933
  Natural History With Surgery, 933
  Natural History Without Surgery, 933
  Randomized Studies of the Natural History of Asymptomatic Primary Hyperparathyroidism, 933
  Guidelines for Parathyroidectomy, 934
Surgery, 935
  Preoperative Localization of Hyperfunctioning Parathyroid Tissue, 935
  Surgical Approach, 936
  Immediate Postoperative Course, 936
Medical Management, 937
  General Measures, 937
  Diet and Supplements, 937
  Pharmaceuticals, 937
Hydrochlorothiazide, 938
  Treatment of Parathyroid Cancer, 938
Summary, 939

KEY POINTS

- The clinical presentation of primary hyperparathyroidism (PHPT) has evolved from a highly symptomatic disorder to one that is asymptomatic in the majority of patients. Presentation, however, can vary depending on screening practices.
- Despite the rarity of overt bony complications in PHPT today, sensitive tools demonstrate evidence of skeletal involvement in both cortical and trabecular compartments.
- Surgery remains the only cure for PHPT. Although medical alternatives can be used to treat hypercalcemia or osteoporosis, there is no "medical parathyroidectomy." All symptomatic patients should undergo surgery. Evidence-based guidelines provide clinicians with criteria for recommending parathyroidectomy in asymptomatic PHPT.

INTRODUCTION

Primary hyperparathyroidism (PHPT) is characterized by hypercalcemia and elevated or inappropriately normal parathyroid hormone (PTH) levels. The disorder today bears few similarities to the severe condition described by Fuller Albright and others as a "disease of stones, bones, and groans" in the 1930s.[1] The skeletal hallmark of PHPT was *osteitis fibrosa cystica*, radiographically characterized by brown tumors of the long bones, subperiosteal bone resorption, distal tapering of the clavicles and phalanges, and "salt-and-pepper" erosions of the skull (Fig. 54.1). Nephrocalcinosis and nephrolithiasis were present in the majority of patients, and neuromuscular dysfunction with muscle weakness was also common. With the advent of the automated serum chemistry autoanalyzer in the 1970s, the diagnosis of PHPT was increasingly recognized, leading to a 4- to 5-fold increase in incidence.[2] Classic symptomatology, concomitantly, became much less frequent. In the United States and elsewhere in the developed world, symptomatic PHPT is now the exception, and more than three fourths of patients have no symptoms attributable to their disease, making PHPT a disease that has "evolved" from its classic presentation (Table 54.1). Symptomatic nephrolithiasis is still observed, although much less frequently than in the past. Now, radiologically evident bone disease is rare, but subclinical skeletal involvement can be readily detected by bone densitometry. This chapter describes the modern presentation, diagnosis, and management of PHPT.

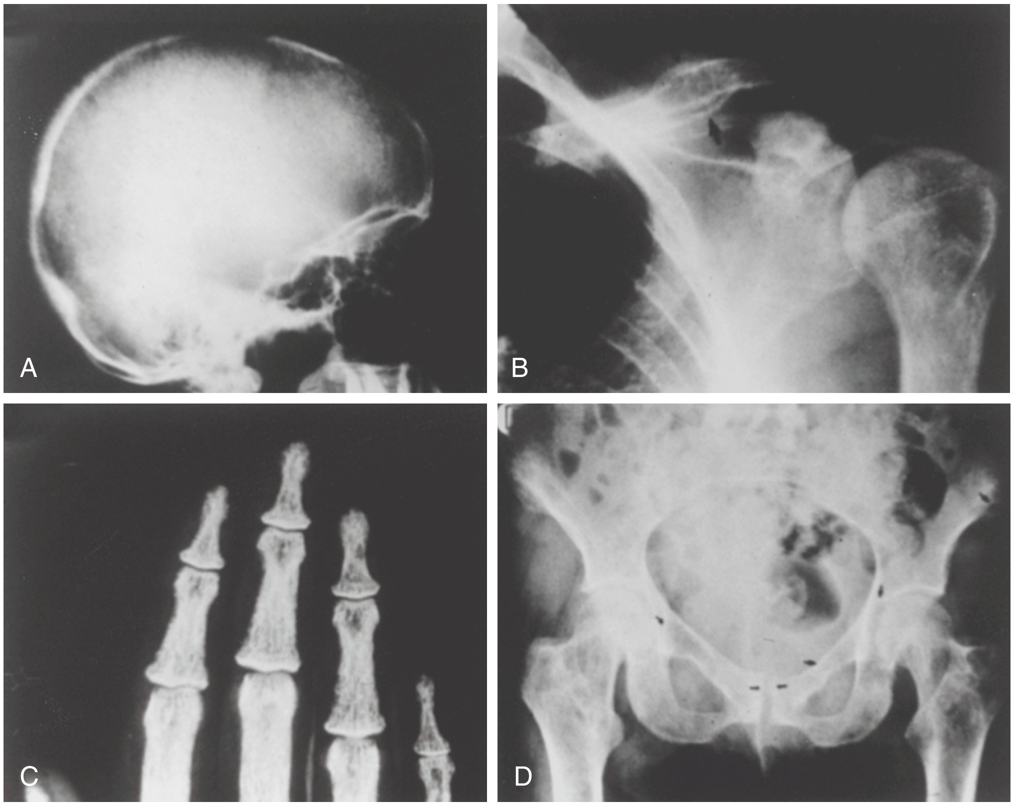


Fig. 54.1 Radiologic representation of osteitis fibrosa cystica in classic primary hyperparathyroidism. **A,** Salt-and-pepper skull. **B,** Cystic bone disease of the clavicle. **C,** Subperiosteal bone resorption of the digits. **D,** Cortical erosions.

TABLE 54.1 Changing Clinical Profile of Primary Hyperparathyroidism

	Cope (1930–1965)	Heath et al. (1965–1974)	Mallette et al. (1965–1972)	Silverberg et al. (1984–2009)
Nephrolithiasis (%)	57	51	37	17
Skeletal disease (%)	23	10	14	1.4
Hypercalciuria (%)	NR	36	40	39
Asymptomatic (%)	0.6	18	22	80

NR, Not reported.

RISK FACTORS, PATHOLOGY, AND ANATOMICAL LOCATION

PHPT results from excessive secretion of PTH from one or more of the parathyroid glands. The underlying cause of sporadic PHPT is unknown in most cases. Although external neck radiation and lithium therapy are risk factors for the development of sporadic PHPT, most patients do not report these exposures.[3,4] Chronically low calcium intake and higher body weight have also been recently described to be risk factors.[5,6] The genetic pathogenesis of sporadic PHPT is unclear in most patients, but genes regulating the cell cycle are thought to be important, given the clonal nature of sporadic parathyroid adenomas.

By far the most common pathological finding in patients with PHPT is a solitary parathyroid adenoma, occurring in 80% of patients.[7] In 2% to 4% of patients, PHPT is due to multiple adenomas. In approximately 15% of patients, all four parathyroid glands are involved.[8] Parathyroid carcinoma accounts for less than 1% of all cases of PHPT.[7] The etiology of four-gland parathyroid hyperplasia is multifactorial. There are no clinical features that definitively differentiate single versus multiglandular disease, but risk factors include inherited genetic syndromes such as multiple endocrine neoplasia (MEN) type 1 or type 2a and lithium exposure.

Parathyroid adenomas can be found in many unexpected anatomic locations. Parathyroid tissue embryonal migration patterns account for a plethora of possible sites of ectopic parathyroid adenomas. The most common locations are within the thyroid gland, the superior mediastinum, and the thymus.[8] Occasionally, adenomas are identified in the retroesophageal space, the pharynx, the lateral neck, and even in the alimentary submucosa of the esophagus. On histologic examination, most parathyroid adenomas are encapsulated and are composed of parathyroid chief cells. Adenomas containing mainly oxyphilic or oncocytic cells are rare, but can give rise to clinical PHPT.

KEY POINTS

- Primary hyperparathyroidism (PHPT) results from excessive secretion of parathyroid hormone from one or more of the parathyroid glands. The underlying cause of sporadic PHPT, the most common form, is unknown in most cases. A solitary parathyroid adenoma is the most common pathologic finding and is present in 80% of patients.

EPIDEMIOLOGY, DIAGNOSIS, AND OVERVIEW OF CLINICAL FEATURES

Incidence

The incidence of PHPT has changed dramatically over the last half century.[2,9] Before the advent of the multichannel autoanalyzer in the early 1970s, Heath et al. reported an incidence of 7.8 cases per 100,000 persons in Rochester, Minnesota. With the introduction of routine calcium measurements in the mid-1970s, this rate rose dramatically to 51.1 cases per 100,000 in the same community. After prevalent cases were diagnosed, the incidence declined to approximately 27 per 100,000 persons per year in the United States, until 1998, at which time another sharp increase was noted.[9] This second peak has been attributed to the introduction of osteoporosis screening guidelines and targeted testing in those with osteoporosis.[9] Recent works shows that the incidence of PHPT increases with age and is higher in women and Black Americans than in men and other racial groups, respectively.[10]

Greater appreciation of the catabolic potential of PTH in postmenopausal women with osteoporosis has led to measurement of PTH even in subjects who do not have hypercalcemia. This trend has led to the emergence of a new entity, normocalcemic PHPT (NPHPT).[11] This condition is characterized by normal serum calcium, elevated PTH, and exclusion of known causes of secondary hyperparathyroidism. The incidence of NPHPT is unknown, but recent studies suggest a prevalence ranging from 0.2% to 3.1%.[12,13]

Diagnosis and Differential Diagnosis

The diagnosis of PHPT is made when hypercalcemia and elevated PTH levels are present.

PTH levels that are inappropriately normal are also consistent with the diagnosis. The other major cause of hypercalcemia, malignancy, is readily discriminated from PHPT by a suppressed PTH level. Further, both the clinical presentation and biochemical profile of PHPT and hypercalcemia of malignancy help distinguish them. Patients with hypercalcemia of malignancy typically have severe and symptomatic hypercalcemia and advanced cancers that are clinically obvious. On the other hand, in PHPT most patients are asymptomatic, and the serum calcium level is typically mildly elevated (within 1 mg/dL of the upper limit of normal). Extremely rarely, a patient with malignancy will be shown to have elevated PTH levels resulting from ectopic secretion of native PTH from the tumor itself. Much more commonly, the malignancy is associated with the secretion of PTH-related protein (PTHrP), a molecule that does not crossreact in an intact PTH assay (discussed later). Finally, it is possible that a malignancy is present in association with PHPT. When the PTH level is elevated in someone with a malignancy, this is more likely to be the case than a true ectopic PTH syndrome.

Although 90% of patients with hypercalcemia have either PHPT or malignancy, the differential diagnosis of hypercalcemia includes a number of other etiologies, such as vitamin D intoxication, granulomatous disease, and others. With the exception of lithium and thiazide use and familial hypocalciuric hypercalcemia (FHH), virtually all other causes of hypercalcemia are associated with suppressed levels of PTH. If lithium and/or thiazides can be safely withdrawn, serum calcium and PTH levels that continue to be elevated 3 to 6 months later confirm the diagnosis of PHPT. FHH, on the other hand, is differentiated from PHPT by family history, typically (but not always) low urinary calcium excretion, and mutations in the calcium-sensing receptor gene or the more recently discovered associated genes *GNA11 and AP2S1*[14]. In addition, virtual complete genetic penetrance leads to its clinical appearance typically before the age of 30 years. It is extremely unusual for FHH to present without an antecedent history after the age of 50 years.

To distinguish PTH-mediated from non–PTH-mediated causes of hypercalcemia, PTH should be measured with an intact immunoradiometric (IRMA) or immunochemiluminometric assay, which readily discriminates between PHPT and hypercalcemia of malignancy. In PHPT, PTH concentrations are usually frankly elevated, but most often within two times the upper limit of normal. A minority may have PTH levels in the normal range, typically the upper range of normal. In PHPT, such values, although within the normal range, are clearly abnormal in a hypercalcemic setting. Several factors affect the PTH level in those with and without PHPT, including age, vitamin D levels, and renal function. Because PTH levels normally rise with age, the broad normal range (typically 10–65 pg/mL) reflects values for the entire population. In the younger individual (<45 years), one expects a narrower and lower normal range (10–45 pg/mL). Occasionally, the PTH level as measured will be as low as 20 to 30 pg/mL. Such unusual examples require a more careful consideration of other causes of hypercalcemia, but such individuals will usually be shown to have PHPT, because hypercalcemia that is not PTH-mediated suppresses the PTH concentration to levels that are either undetectable or at the lower limits of the reference range. Souberbielle et al.[15] have illustrated that the normal range is dependent on whether or not the reference population is vitamin D-deficient. When vitamin D-deficient individuals were excluded, the upper limit of the PTH reference interval decreased. Patients with PHPT and vitamin D deficiency have "heightened" PTH levels compared with those who are vitamin D-sufficient.[16]

On the other hand, renal dysfunction tends to elevate PTH levels via a number of mechanisms, including reduced clearance and degradation of PTH. Indeed, patients with PHPT and severe renal dysfunction (glomerular filtration rate <30mL/min) may also have higher PTH levels compared with those with better renal function.[17] In addition, the "intact" IRMA for PTH overestimates the concentration of biologically active PTH, particularly in renal failure. In 1998, Lepage et al.[18] demonstrated a large non-(1–84) PTH fragment that comigrated with a large aminoterminally truncated fragment (PTH[7–84]) and showed substantial crossreactivity in commercially available IRMAs. This large, inactive moiety constituted as much as 50% of immunoreactivity by IRMA for PTH in individuals with chronic renal failure. Recognition of this molecule led to the development of a new IRMA using affinity-purified polyclonal antibodies to PTH(39–84) and to the extreme N-terminal amino acid regions, PTH(1–4).[19] This "whole-PTH" or third-generation assay detects only the full-length PTH molecule, PTH(1–84). This assay has clear utility in uremic patients, but in PHPT both assays are equally useful.[18,20] Using the third-generation assay for PTH(1-84), a second molecular form of PTH(1-84) that is immunologically intact at both extremes has been identified. This molecule reacts only poorly in second-generation PTH assays. It represents less than 10% of immunoreactivity in normal individuals and up to 15% in renal failure patients. In a limited number of patients with a severe form of PHPT, or with parathyroid cancer, it may be overexpressed.[21]

PHPT can be discriminated from secondary and tertiary hyperparathyroidism by its different biochemical profile. Secondary hyperparathyroidism is associated with an appropriate elevation in PTH in response to a hypocalcemic provocation and either a frankly low or normal serum calcium level. Secondary hyperparathyroidism is often caused by vitamin D deficiency. Other causes include malabsorption, kidney disease, or hypercalciuria. Infrequently, patients with secondary hyperparathyroidism may become hypercalcemic, and will ultimately be found to have PHPT, when the underlying condition (for example, vitamin D deficiency) is corrected.[8] In these cases, the hypercalcemia of PHPT was "masked" by the coexisting condition. On the other hand, tertiary hyperparathyroidism describes a condition in which prolonged, severe

TABLE 54.2 Biochemical Profiles for Various Causes of Hypercalcemia and Hyperparathyroidism

Cause	Serum Calcium	Parathyroid Hormone (PTH)	Urine Calcium
Primary hyperparathyroidism	Elevated	Elevated or inappropriately normal	High or high normal
Hypercalcemia of malignancy and non–PTH-mediated hypercalcemia	Elevated	Suppressed	Typically high[a]
Secondary hyperparathyroidism	Normal or low	Elevated	Low in vitamin D deficiency, malabsorption, chronic renal failure High in idiopathic hypercalciuria
Tertiary hyperparathyroidism	Elevated	Elevated	Low before transplant
Familial hypocalciuric hypercalcemia	Elevated	Typically high-normal, or elevated	Typically low
Normocalcemic primary hyperparathyroidism	Normal	High	<350 mg/24 h[11]

[a] May vary by cause.

secondary hyperparathyroidism (as in end-stage renal disease) evolves into a hypercalcemic state owing to the development of autonomous functioning of one or more of the hyperplastic parathyroid glands. This can be observed in patients on dialysis or after renal transplant. Tertiary hyperparathyroidism is usually obvious from the history.

NPHPT describes a condition characterized by normal serum albumin-corrected calcium levels and ionized calcium values with an elevated PTH level. This condition can only be diagnosed when all known causes of secondary hyperparathyroidism have been excluded. Patients with NPHPT typically are diagnosed when PTH is measured in the course of an evaluation for low bone mass. NPHPT may represent the earliest manifestations of PHPT, a “forme fruste” of the disease. Several reports have appeared describing these individuals, with some patients progressing to overt hypercalcemia while under observation.[11,13]

Although the term “NPHPT” has been in use for decades, there has been considerable controversy concerning the accuracy of this designation. In many cases the increases in PTH levels were attributable to the limitations of available assay technology. The older midmolecule radioimmunoassay for PTH, previously in common use, measured hormone fragments in addition to the intact molecule. Spuriously elevated PTH levels, particularly in those with renal insufficiency in whom clearance of hormone fragments is impaired, were seen. Alternative explanations for hyperparathyroidism with NPHPT have been discovered, including medications, hypercalciuria, renal insufficiency, and certain forms of liver and gastrointestinal disease. In recent years, it has become clear that many patients designated as having NPHPT were vitamin D–deficient. Vitamin D deficiency with coexisting PHPT can give the semblance of normal calcium levels, when in fact the patients would have been hypercalcemic if the vitamin D levels were normal. Given that a possible view of NPHPT is a condition fostered by an element of vitamin D resistance, it is important to ensure vitamin D sufficiency. Although the Institute of Medicine stated that normal levels of vitamin D, as measured by 25-hydroxyvitamin D, are 20 ng/mL, it did not address conditions of abnormal mineral metabolism, such as PHPT. In particular, in NPHPT, we and others recommend that levels of 25-hydroxyvitamin D be raised, if necessary, to more than 30 ng/mL for at least 3 months to rule out an element of vitamin D insufficiency in this population. Biochemical profiles for the various causes of hypercalcemia and hyperparathyroidism are shown in Table 54.2.

Other Biochemical Features

In PHPT, serum phosphorus tends to be in the lower range of normal, but frank hypophosphatemia is present in less than one-fourth of patients. Hypophosphatemia, when present, is as a result of the phosphaturic actions of PTH. Average total urinary calcium excretion is at the upper end of the normal range, with approximately 40% of all patients experiencing hypercalciuria. Serum 25-hydroxyvitamin D levels tend to be at the lower end of the normal range. Although mean values of 1,25-dihydroxyvitamin D are in the high-normal range, approximately one third of patients have frankly elevated levels of 1,25-dihydroxyvitamin D.[8] This pattern is a result of the actions of PTH to increase expression of the 1-alpha hydroxylase that converts 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D. A typical biochemical profile is shown in Table 54.3.

TABLE 54.3 Biochemical Profile in Primary Hyperparathyroidism ($n = 137$)

	Patients (Mean ± SEM)	Normal Range
Serum calcium	10.7 ± 0.1 mg/dL	8.2–10.2 mg/dL
Serum phosphorus	2.8 ± 0.1 mg/dL	2.5–1.5 mg/dL
Total alkaline phosphatase	114 ± 5 IU/L	<100 IU/L
Serum magnesium	2.0 ± 0.1 mg/dL	1.8–2.4 mg/dL
PTH (IRMA)	119 ± 7 pg/mL	10–65 pg/mL
25(OH)D	19 ± 1 ng/mL	30–80 ng/mL
1,25(OH)2D	54 ± 2 pg/mL	15–60 pg/mL
Urinary calcium	240 ± 11 mg/g creatinine	
Urine DPD	17.6 ± 1.3 nmol/mmol creatinine	<14.6 nmol/mmol creatinine
Urine PYD	46.8 ± 2.7 nmol/mmol creatinine	<51.8 nmol/mmol creatinine

DPD, Deoxypyridinoline; *PTH,* parathyroid hormone; *IRMA,* immunoradiometric assay; *PYD,* pyridinoline; *25(OH)D,* 25-hydroxyvitamin D; *1,25(OH)2D,* 1,25-dihydroxyvitamin D; *SEM,* standard error of the mean.

Clinical Presentation

PHPT typically occurs in individuals in their middle years, with a peak incidence between ages 50 and 60 years.[8] However, the condition can occur at any age. Women are affected more frequently than men, in a ratio of approximately 3 to 4:1. Several different presentations of PHPT are possible and were originally described successively in time.[1,11,22] The classical symptomatic presentation was described first; later, asymptomatic PHPT emerged owing to biochemical screening, and most recently the normocalcemic variant was discovered, as described. However, these three forms of PHPT contemporaneously exist today. Which presentation predominates depends upon population- and geographic-specific screening practices. It is also postulated that vitamin D deficiency may affect clinical presentation. Vitamin D deficiency heightens PTH elevations, and this can worsen the hyperparathyroid process.[23] In regions of the world and populations where biochemical screening is not routine, and incidentally where vitamin D deficiency is endemic, symptomatic PHPT is the most common form, and PHPT will appear to be uncommon because it is only discovered when symptomatic.[24] In areas and populations where screening is routine, asymptomatic PHPT will predominate, and the incidence of PHPT is higher.[22]

This chapter focuses on asymptomatic PHPT, as it is the predominant form in the United States and in most of the developed world. At the time of diagnosis, most patients with PHPT do not exhibit classic symptoms or signs associated with disease. Clinically overt kidney stones and fractures are rare.[8] Constitutional complaints such as weakness, easy fatigability, depression, and intellectual weariness are seen with some regularity (see later discussion).[8] The physical examination is generally unremarkable. Band keratopathy, a hallmark of classic PHPT, occurs because of deposition of calcium phosphate crystals in the cornea, but is virtually never seen grossly. Even by slit-lamp examination, this finding is rare. The neck shows no masses. The neuromuscular system is normal. The sections following provide a detailed description of the multisystemic manifestations of PHPT.

Diseases associated epidemiologically with PHPT have included hypertension,[25] peptic ulcer disease, gout, or pseudogout.[26] More recently, celiac disease has been associated with PHPT.[27] Some concomitant disorders such as hypertension are commonly seen, but it is not established that any of these associated disorders are etiologically linked to the disease.

KEY POINTS

- Primary hyperparathyroidism is a common condition and most often occurs in postmenopausal women. Most often, the diagnosis is made when hypercalcemia is present with an elevated or inappropriately normal serum parathyroid hormone level. The clinical presentation has evolved from a highly symptomatic disorder to one that is asymptomatic in the majority of patients. Presentations, however, can vary depending on screening practices.

THE SKELETON

The classic radiologic bone disease of PHPT, *osteitis fibrosa cystica*, is rarely seen today in the United States.[8] Most series place the incidence of *osteitis fibrosa cystica* at less than 2% of patients with PHPT. The absence of classic radiographic features (salt-and-pepper skull, tapering of the distal third of the clavicle, subperiosteal bone resorption of the phalanges, brown tumors) does not mean that the skeleton is not affected. With more sensitive techniques, it has become clear that skeletal involvement in the hyperparathyroid process is actually quite common. This section reviews the profile of the skeleton in PHPT as it is reflected in assays for bone markers, bone densitometry, bone histomorphometry, and new skeletal imaging techniques.

Bone Turnover Markers

PTH stimulates both bone resorption and bone formation. Markers of bone turnover, which reflect those dynamics, provide clues to the extent of skeletal involvement in PHPT.

Bone Formation Markers. Osteoblast products, including bone-specific alkaline phosphatase activity, osteocalcin, and serum N-terminal propeptide of type I collagen, reflect bone formation. In PHPT, alkaline phosphatase, the most widely clinically available marker, can be mildly elevated, but in many patients total alkaline phosphatase values are within normal limits.[28] Osteocalcin is also generally increased in patients with PHPT.[29] Sclerostin is an important regulator of bone formation. Patients with PHPT have low sclerostin levels, suggesting that PTH downregulates sclerostin.[30]

Bone Resorption Markers. Markers of bone resorption include the osteoclast product, tartrate-resistant acid phosphatase, and collagen breakdown products such as hydroxyproline, hydroxypyridinium crosslinks of collagen, and N- and C-telopeptides of type 1 collagen (NTX and CTX). Urinary hydroxyproline, once the only available marker of bone resorption, no longer offers sufficient sensitivity or specificity to make it useful. Although urinary hydroxyproline was frankly elevated in patients with osteitis fibrosa cystica, in mild asymptomatic PHPT it is generally normal. Hydroxypyridinium cross-links of collagen, pyridinoline (PYD), and deoxypyridinoline (DPD), on the other hand, are often elevated in PHPT. They return to normal after parathyroidectomy.[31] DPD and PYD both correlate positively with PTH concentrations. Studies of NTX, CTX, and TRAP are limited. Thus, sensitive assays of bone formation and bone resorption are both elevated in mild PHPT.

Longitudinal Bone Turnover Marker Studies. Studies of bone turnover markers in the longitudinal follow-up of patients with PHPT indicate a reduction in these markers following parathyroidectomy. Bone markers decline after parathyroidectomy.[32] The kinetics of change in bone resorption versus bone formation following parathyroidectomy provide insight into skeletal recovery. We have found that markers of bone resorption decline rapidly following successful parathyroidectomy, whereas indices of bone formation follow a more gradual decrease.[31] Urinary PYD and DPD decreased significantly as early as 2 weeks following parathyroid surgery, preceding reductions in alkaline phosphatase. Similar data were reported from others. Minisola et al.[33] reported a drop in bone resorptive markers and no significant change in alkaline phosphatase or osteocalcin. The persistence of elevated bone formation markers coupled with rapid declines in bone resorption markers indicate a shift in the coupling between bone formation and bone resorption toward an accrual of bone mineral postoperatively.

Bone Densitometry

The advent of bone mineral densitometry as a major diagnostic tool for osteoporosis occurred at a time when the clinical profile of PHPT was changing from a symptomatic to an asymptomatic disease. This fortuitous timing allowed questions about skeletal involvement in PHPT to be addressed when specific gross radiologic features of PHPT had all but disappeared. Observations of skeletal health in PHPT made by bone densitometry have established the importance of this technology in the evaluation of all patients with PHPT. The Consensus Development Conference on Asymptomatic Primary Hyperparathyroidism in 1990 implicitly acknowledged this point when bone mineral densitometry was included as a separate criterion for clinical decision-making. Since that time, bone densitometry has become an indispensable component of both evaluating the patient and establishing clinical guidelines for management and monitoring.

The known physiologic proclivity of PTH to be catabolic at sites of cortical bone make a cortical site essential to any complete densitometric study of PHPT. By convention, the distal third of the radius

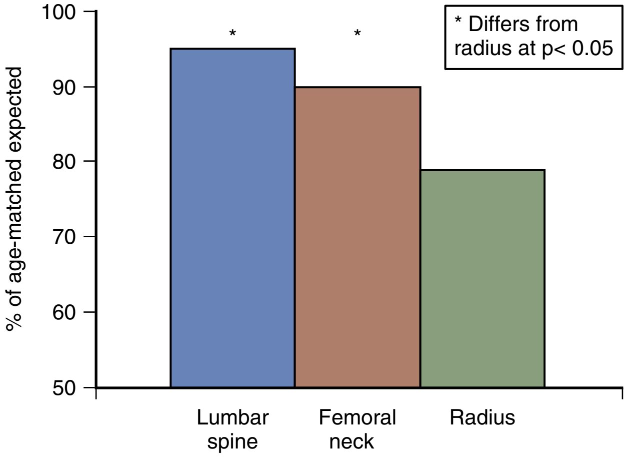


Fig. 54.2 Bone densitometry in primary hyperparathyroidism. Data are shown in comparison to age-matched and sex-matched normal subjects. Divergence from expected values is different at each site ($P < .0001$). (Data from Silverberg SJ, Shane E, de la Cruz L, et al. Skeletal disease in primary hyperparathyroidism, *J Bone Miner Res.* 1989;4:283–291.)

is the site used. The early densitometric studies in PHPT also showed another physiologic property of PTH, namely, to preserve bone at cancellous sites. The lumbar spine is an important site to measure, not only because it is predominantly cancellous bone, but also because postmenopausal women are at risk for cancellous bone loss. In PHPT, bone density at the distal third of the radius is diminished,[34] whereas at the lumbar spine it is only minimally reduced. The hip region, containing relatively equal amounts of cortical and cancellous elements, shows bone density intermediate between the cortical and cancellous sites (Fig. 54.2). The results support not only the notion that PTH is catabolic for cortical bone but also the view that PTH is generally protective against bone loss in cancellous bone.[35] In postmenopausal women, the same pattern was observed.[36] Postmenopausal women with PHPT, therefore, show a reversal of the pattern typically associated with postmenopausal bone loss. Rather than preferential loss of cancellous bone at the lumbar spine, the cortical site of the distal radius is more often affected in postmenopausal women with PHPT.

The bone density profile in which there is relative preservation of skeletal mass at the vertebrae and reduction at the more cortical distal radius is not always seen in PHPT. Although this pattern is evident in the vast majority of patients, some patients show evidence of vertebral osteopenia at the time of presentation. In our natural-history study, approximately 15% of patients had a lumbar spine Z score of less than –1.5 at the time of diagnosis.[37] Only half of these patients were postmenopausal women, so not all vertebral bone loss could be attributed entirely to estrogen deficiency. These patients are of interest with regard to changes in bone density following parathyroidectomy and are discussed in further detail later. The extent of vertebral bone involvement will vary as a function of disease severity. In the typical mild form of the disease, the pattern described earlier is seen. When PHPT is more advanced, there will be more generalized involvement, with involvement of the lumbar spine. When PHPT is severe or more symptomatic, all bones can be extensively involved. Vitamin D deficiency in mild asymptomatic PHPT seems to have minimal effect on bone mineral density (BMD), with only slightly reduced BMD at the one-third radius in those with low vitamin D.[38]

Bone Histomorphometry

Analyses of percutaneous bone biopsies from patients with PHPT have provided direct information that could only be indirectly surmised by bone densitometry and by bone markers. Both static and dynamic parameters present a picture of cortical thinning, maintenance of cancellous bone volume (Fig. 54.3), and a very dynamic process associated with high turnover and accelerated bone remodeling. Cortical thinning, inferred by bone mineral densitometry, is clearly documented in a quantitative manner by iliac crest bone biopsy.[39] Van Doorn et al.[40] demonstrated a positive correlation between PTH levels and cortical porosity. These findings are consistent with the known effect of PTH to be catabolic at endocortical surfaces of bone. Osteoclasts are thought to erode more deeply along the corticomedullary junction under the influence of PTH.

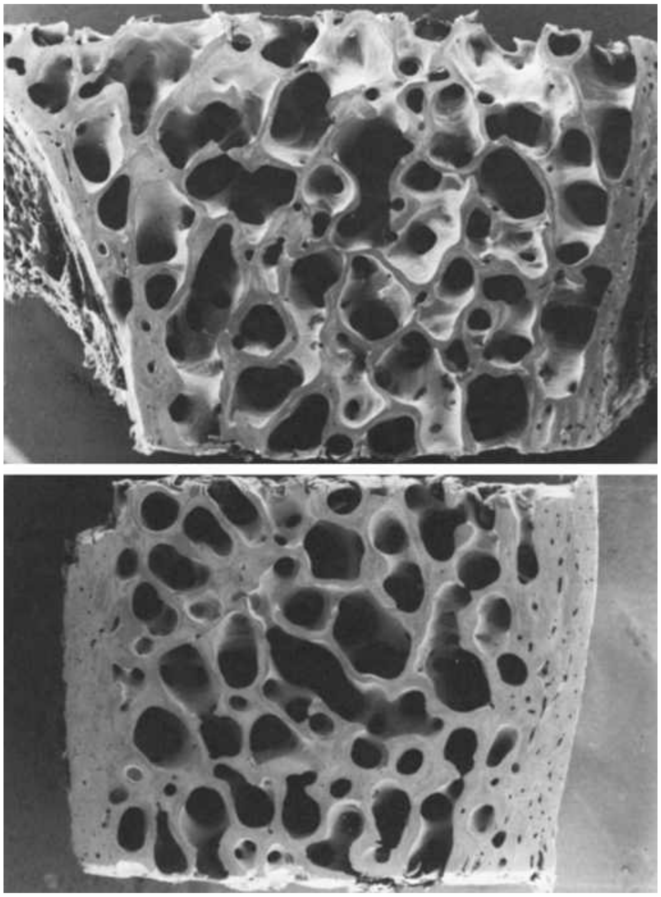


Fig. 54.3 Scanning electron micrographs of bone biopsy specimens from a normal subject (*bottom*) and an age- and sex-matched patient with primary hyperparathyroidism (*top*).The cortices of the hyperparathyroid sample are markedly thinned, but cancellous bone and trabecular connectivity appear to be well-preserved (×31.25). (From Parisien MV, Silverberg SJ, Shane E, et al. Bone disease in primary hyperparathyroidism, *Endocrinol Metab Clin North Am.* 1990;19:19–34.)

Histomorphometric studies have also contributed information about cancellous bone in PHPT. Again, as suggested by bone densitometry, cancellous bone volume is well preserved in PHPT. This is seen as well among postmenopausal women with PHPT. Several studies have shown that cancellous bone is actually increased in PHPT as compared with normal subjects.[41] When cancellous bone volume is compared among age- and sex-matched subjects with PHPT or postmenopausal osteoporosis, a dramatic difference is evident (Fig. 54.4). Whereas postmenopausal women with osteoporosis have reduced cancellous bone volume, women with PHPT have higher cancellous bone volume.[41] The region(s) of bone loss in PHPT is (are) directed toward the cortical bone compartment, with good maintenance of cancellous bone volume unless the PHPT is unusually active.

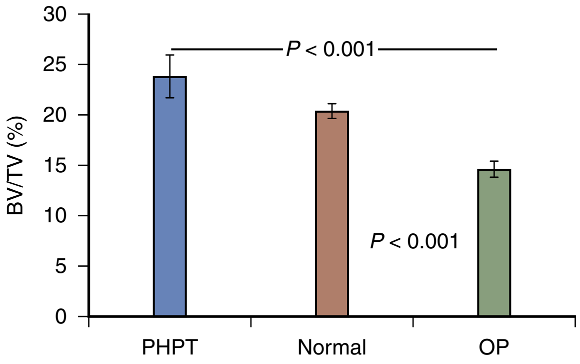


Fig. 54.4 Cancellous bone volume in primary hyperparathyroidism. Cancellous bone volume (*BV/TV*) was analyzed from bone biopsy specimens of the iliac crest. Comparisons are among 16 women with primary hyperparathyroidism (*PHPT*), 17 women with postmenopausal osteoporosis (*OP*), and 31 women with no known disorder of bone metabolism. Subjects were matched for age and other indices. (Data from Parisien M, Cosman F, Mellish RW, et al. Bone structure in postmenopausal hyperparathyroid, osteoporotic and normal women, *J Bone Miner Res.* 1995;10:1393–1399.)

In PHPT, age-related bone loss appears to be mitigated. In a study of 27 patients with PHPT (10 men and 17 women), static parameters of bone turnover (osteoid surface, osteoid volume, and eroded surface) were increased, as expected, in patients relative to control subjects.[42] However, in control subjects, trabecular number varied inversely with age, whereas trabecular separation increased with advancing age. These observations are expected concomitants of aging. In marked contrast, in the patients with PHPT, no such age dependency was seen. There was no relationship between trabecular number or separation and age in PHPT, suggesting that the actual plates and their connections were being maintained over time more effectively than one would have expected by aging *per se*. Thus, PHPT seems to retard the normal age-related processes associated with trabecular loss.

In PHPT, indices of trabecular connectivity are greater than expected, whereas indices of disconnectivity are decreased. When three matched groups of postmenopausal women were assessed (a normal group, a group with postmenopausal osteoporosis, and a group with PHPT), women with PHPT were shown to have trabeculae with less evidence of disconnectivity compared with normal women, despite increased levels of bone turnover.[42,43] Thus, cancellous bone is preserved in PHPT through the maintenance of well-connected trabecular plates. To determine the mechanism of cancellous bone preservation in PHPT, static and dynamic histomorphometric indices were compared between normal and hyperparathyroid postmenopausal women. In normal postmenopausal women, there is an imbalance in bone formation and resorption that favors excess bone resorption. In postmenopausal women with PHPT, on the other hand, the adjusted apposition rate is increased. Bone formation, thus favored, may explain the efficacy of PTH at cancellous sites in patients with osteoporosis. Assessment of bone remodeling variables in patients with PHPT shows increases in the active bone-formation period[41] (Table 54.4). The increased bone formation rate and total formation period may explain the preservation of cancellous bone seen in this disease.

More recently, further analysis of trabecular microarchitecture has taken advantage of newer technologies that have largely been confirmatory. In a three-dimensional analysis of transiliac bone biopsies using micro computed tomography technology, a highly significant correlation was observed with the conventional histomorphometry[44] described earlier. In comparison with age-matched control subjects without PHPT, postmenopausal women with PHPT had higher bone volume (BV/TV), higher bone surface area (BS/TV), higher connectivity density (Conn.D), and lower trabecular separation. There were also less marked age-related declines in BV/TV and Conn.D as compared with controls, with no decline in BS/TV. Using the technique of backscattered electron imaging to evaluate trabecular BMD distribution in iliac crest bone biopsies, Roschger et al.[45] showed reduced average mineralization density and an increase in the heterogeneity of the degree of mineralization, consistent with reduced mean age of bone tissue. Studies of collagen maturity using Fourier Transform Infrared Spectroscopy provide further support for these observations.[46] Bone strength, therefore, in PHPT has to take into account a number of factors related to skeletal properties of bone besides BMD.

New Imaging Technologies

Newer noninvasive skeletal imaging technologies offer new insight into the skeletal manifestations of PHPT beyond observations made by dual-energy x-ray absorptiometry (DXA) and radiography. The trabecular bone score (TBS) provides an indirect assessment of trabecular microstructure from the DXA image. In those without PHPT, it has been shown to predict fracture independently of BMD. In PHPT, TBS reveals trabecular microstructural deterioration at the spine, despite preserved BMD by DXA at this site.[47] High-resolution peripheral quantitative computed tomography (HRpQCT) is a technology that noninvasively and directly measures skeletal microstructure at the distal radius and tibia. Utilizing this technology, studies in patients with PHPT indicate not only cortical thinning, but additionally trabecular deficits at the radius and tibia (Fig. 54.5).[48] At the radius, trabeculae were fewer, thinner, and more widely and heterogeneously spaced. At the tibia, trabeculae were more heterogeneously spaced.[48] These deficits led to reduced stiffness when images were analyzed using microfinite element analysis, a technique that integrates structural and densitometric information from the HRpQCT image into an estimated of mechanical competence. These recent findings, pointing to abnormalities in the trabecular compartment of bone, help to account for the increased risk of vertebral fractures (see later) observed in PHPT that had remained unexplained before the advent of such technologies.[49,50] The difference in microskeletal abnormalities between the iliac crest bone biopsy data and HRpQCT may well reflect site-specific sampling differences.

Fractures

Fractures were a common clinical event in classic PHPT. In modern PHPT, one would anticipate, based on the BMD patterns observed with DXA, an increased risk of peripheral fractures but a reduction in vertebral fractures. Although not all studies are consistent, and many of the data are retrospective and/or cross-sectional, the majority of studies suggest an increased risk for vertebral fractures in patients with PHPT.[49,50] Moreover, recent data indicates that many vertebral fractures are in fact clinically silent.[50] The paradox of increased vertebral fracture risk despite preserved lumbar spine BMD in PHPT had remained unclear until the advent of TBS and HRpQCT, which clearly document trabecular deficits in addition to previously recognized cortical skeletal deterioration. Using the Danish National Patient registry and a nested case-control design, Ejlsmark-Svensson et al. recently showed that vertebral fractures in patients with PHPT occur at a higher BMD than in patients without PHPT, again pointing to the importance of other elements of bone quality in PHPT.[51] In one recent study, among asymptomatic PHPT patients, only those who met surgical guidelines showed a higher incidence of vertebral fractures compared with controls.[50]

TABLE 54.4 Wall Width and Remodeling Variables in Primary Hyperparathyroidism and Control Groups (Mean ± Standard Error of the Mean)

Variable	Primary Hyperparathyroidism (*n* = 19)	Control (*n* = 34)	*P*
Wall width (μm)	40.26 ± 0.36	34.58 ± 0.45	< .0001
Eroded perimeter (%)	9.00 ± 0.86	4.76 ± 0.39	< .0001
Osteoid perimeter (%)	26.84 ± 2.79	15.04 ± 1.09	< .0001
Osteoid width (μm)	13.39 ± 0.54	9.92 ± 0.36	< .0001
Single-labeled perimeter (%)	11.56 ± 1.63	4.47 ± 0.48	< .0001
Double-labeled perimeter (%)	10.41 ± 1.28	4.45 ± 0.65	< .0001
Mineralizing perimeter (%)	16.19 ± 1.75	6.68 ± 0.83	< .0001
Mineralizing perimeter/osteoid perimeter (%)	63.0 ± 5.0	44.04 ± 4.0	< .01
Mineral apposition rate (μm/day)	0.63 ± 0.03	0.63 ± 0.02	NS
Bone formation rate (μm 2/μm/day)	0.10 ± 0.01	0.042 ± 0.006	< .0001
Adjusted apposition rate (μm/day)	0.40 ± 0.04	0.29 ± 0.03	< .015
Activation frequency/year	0.95 ± 0.12	0.45 ± 0.06	< .0002
Mineralization lag time (days)	44.0 ± 6.5	57.0 ± 8.9	NS
Osteoid maturation time (days)	22.5 ± 1.8	16.6 ± 0.9	< .003
Total formation period (days)	129.2 ± 21.0	208.8 ± 32.5	NS
Active formation period (days)	67.8 ± 5.1	57.3 ± 2.3	< .05
Resorption period (days)	48.4 ± 7.3	84.8 ± 25.0	NS
Remodeling period (days)	172.5 ± 25.2	299.9 ± 55.1	NS

NS, Not significant.

Modified from Dempster DW, Parisien M, Silverberg SJ, et al. On the mechanism of cancellous bone preservation in postmenopausal women with mild primary hyperparathyroidism, *J Clin Endocrinol Metab.* 1999;84:1562–1566.

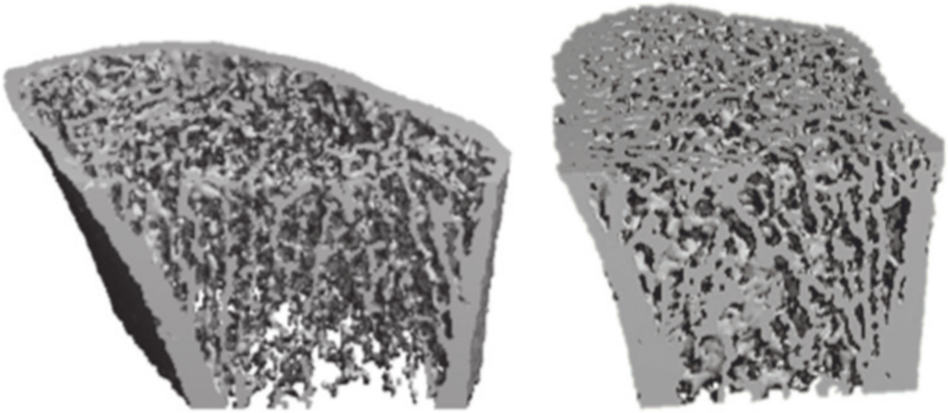


Fig. 54.5 High-resolution peripheral quantitative computed tomography images of the radius in a patient with primary hyperparathyroidism (PHPT; *left*) and a normal control (*right*). Trabecular deterioration is evident in PHPT. (From Stein EM, Silva BC, Boutroy S, et al. Primary hyperparathyroidism is associated with abnormal cortical and trabecular microstructure and reduced bone stiffness in postmenopausal women. *J Bone Miner Res.* 2012;28:1029–1040.)

The risk for hip fracture is not clearly increased in PHPT. In a study that focused on hip fracture, a population-based prospective analysis (mean of 17 years' duration; 23,341 person-years) showed women with PHPT in Sweden not to be at increased risk.[52] The Mayo Clinic experience with PHPT and risk of fracture reviewed 407 cases of PHPT recognized during the 28-year period between 1965 and 1992.[49] Fracture risk was assessed by comparing fractures at a number of sites with numbers of fractures expected based on gender and age from the general population. The clinical presentation of these patients with PHPT was typical of the mild form of the disease, with the serum calcium being only modestly elevated at 10.9 ± 0.6 mg/dL. The data from this retrospective epidemiologic study indicate that overall fracture risk was significantly increased at many sites such as the vertebral spine, the distal forearm, the ribs, and the pelvis. There was no increase in hip fractures. One might expect to see an increased incidence of distal forearm fractures as seen in the Mayo study, because the hyperparathyroid process tends to lead to a reduction of cortical bone (distal forearm) in preference to cancellous bone (vertebral spine). Unfortunately, there were no densitometric data provided in this study, so bone density could not be related to fracture incidence.

The impact of PHPT on fracture incidence appears complex and may be site-specific. This relationship is likely influenced by site-specific changes in areal bone density, bone size, and microstructure. Excess PTH would induce cortical thinning because of endosteal bone resorption but would also increase periosteal apposition, thus increasing bone diameter. Decreased areal bone density would increase fracture risk, whereas increased bone diameter and preserved microstructure at certain sites may protect against fractures. Prospective studies are needed to elucidate the site-specific risk of fracture in PHPT.

KEY POINTS

- The classic radiologic disease of primary hyperparathyroidism (PHPT), *osteitis fibrosa cystica*, is rarely seen today in the United States, but subclinical skeletal disease is often present. Dual-energy x-ray absorptiometry typically shows preferential loss of bone mineral density (BMD) at the distal one-third radius, a site that is rich in cortical bone. BMD at the lumbar spine, predominantly comprised of trabecular bone, tends to be preserved. The hip region, containing relatively equal amounts of cortical and trabecular elements, shows intermediate BMD values. Most data suggest fracture risk is increased in patients with PHPT.

NEPHROLITHIASIS AND RENAL FUNCTION

In the past, classic clinical descriptions of PHPT emphasized kidney stones as a principal complication of the disease. The cause of nephrolithiasis in PHPT is probably multifactorial. An increase in the amount of calcium filtered at the glomerulus resulting from hypercalcemia may lead to hypercalciuria despite the physiologic actions of PTH to facilitate calcium reabsorption. A component of absorptive hypercalciuria exists in this disorder. The enhanced intestinal calcium absorption is likely a result of increased production of 1,25-dihydroxyvitamin D, a consequence of PTH's action to increase the synthesis of this active metabolite.[53] Indeed, urinary calcium excretion is correlated with 1,25-dihydroxyvitamin D levels.[53] The skeleton provides yet another possible source for the increased levels of calcium in the glomerular filtrate. Hyperparathyroid bone resorption may contribute to hypercalciuria, and subsequently to nephrolithiasis, even though there is no convincing evidence to support this hypothesis. Finally, alteration in local urinary factors, such as a reduction in inhibitor activity or an increase in stone-promoting factors, may predispose some patients with PHPT to nephrolithiasis. It remains unclear whether the urine of patients with hyperparathyroid stone disease is different in this regard from that of other stone-formers.

Studies in the 1970s and 1980s documented a higher incidence of renal stone disease than do more recent reports.[8] Although the incidence of symptomatic nephrolithiasis today is much less common than it was in classical PHPT, kidney stones remain the most common manifestation of symptomatic PHPT, and symptomatic stones occur in 15% to20% of all patients (see Table 54.1).[8] Screening for asymptomatic nephrolithiasis indicates that the prevalence is actually much higher, and this is now recommended in the most recent set of guidelines on the management of asymptomatic PHPT.[54-56]

The etiology of why stones develop in some individuals with PHPT, but not others, has been debated since the 1930s, but is still not well understood. In the 1930s, it was generally accepted that bone and stone disease did not coexist in the same patient with classic PHPT.[1,57] Albright and Reifenstein[57] theorized that low dietary calcium intake led to bone disease, whereas adequate or high dietary calcium levels caused stone disease. Dent et al.,[58] who provided convincing evidence against this construct, proposed the existence of two forms of circulating PTH, one causing renal stones and the other causing bone disease. A host of mechanisms, including differences in dietary calcium, calcium absorption, forms of circulating PTH, and levels of 1,25-dihydroxyvitamin D, were proposed to account for the clinical distinction between bone and stone disease in PHPT.[58] Today, there is no clear evidence for two distinct subtypes of PHPT or that the process affecting the skeleton and kidneys occur in different subsets of patients. Cortical bone demineralization is as common and as extensive in those with and without nephrolithiasis.[59]

Although more recent work has suggested that 1,25-dihydroxyvitamin D plays an etiologic role in the development of nephrolithiasis in PHPT, not all studies are consistent with this premise.[59,60] Other investigations have shown risk factors for nephrolithiasis include younger age and male sex, whereas degree of hypercalcemia and hypercalciuria, PTH levels, and other urinary factors have less consistently been associated.[56,61,62] Hypomagnesuria has recently been associated with silent kidney stones in PHPT.[63]

Other renal manifestations of PHPT include hypercalciuria, which is seen in approximately 40% of patients, and nephrocalcinosis. The frequency of nephrocalcinosis is unknown, but it appears to be relatively uncommon today.[62] Although there were clear reports of renal impairment in early descriptions of PHPT, the prevalence of renal dysfunction (estimated glomerular filtration rate [eGFR] <60 mL/min) today in patients with mild PHPT is low, with recent studies suggesting rates of 15% to 17%.[64] Neither the severity of PHPT nor having a history of nephrolithiasis were risk factors for reduced eGFR in a 2014 study in those with mild PHPT; instead ,traditional risk factors, such as age, hypertension, use of antihypertensive medication, and fasting glucose levels were associated with poorer kidney function.[64] Longitudinal data are reassuring in this regard, as renal function remains stable in PHPT over long periods of follow-up.[65]

KEY POINTS

- Symptomatic nephrolithiasis is present in approximately 20% of patients today in the United States. Hypercalciuria is also common. With noninvasive imaging, nephrocalcinosis can also be seen, while overt renal dysfunction is rare.

OTHER ORGAN INVOLVEMENT

Cardiovascular System

Interest in the effect of PHPT on cardiovascular function is rooted in pathophysiologic observations of the hypercalcemic state. Hypercalcemia has been associated with increases in blood pressure, left ventricular hypertrophy (LVH), heart muscle hypercontractility, and arrhythmias.[8,66] Furthermore, evidence of calcium deposition has been documented in the form of calcifications in the myocardium, heart valves, and coronary arteries. The association of overt cardiovascular symptomatology with modern-day PHPT is unclear because of inconsistencies between studies. An explanation for the inconsistent results reported in the literature on the cardiovascular manifestations of PHPT relates to the fact that the clinical profile of the disease has changed. As a result, the cohorts that have been studied have varied greatly in the severity of their underlying disease. This is particularly true in terms of the serum calcium and PTH concentrations, with data from cohorts with marked hypercalcemia and hyperparathyroidism showing the most cardiovascular involvement. Because it is thought that both calcium and PTH can independently affect the cardiovascular system, such variability among cohorts can give rise to inconsistent results. Recent studies have focused on not only cardiovascular mortality, but also hypertension, coronary artery disease (CAD), valve calcification, LVH, carotid disease, and vascular stiffness.

Cardiovascular Mortality. There is little doubt that, in very active PHPT, cardiovascular mortality is increased.[8,66] Of some interest are postoperative observations in which the higher cardiovascular mortality rate persists for years after cure.[67] These observations differ markedly from those in which asymptomatic PHPT has been studied. Although limited, studies have not shown any increase in mortality.[68] The Mayo Clinic studies help to bring these observations together. In mildly hypercalcemic individuals, overall and cardiovascular mortality were reduced, but in those whose serum calcium was in the highest quartile, cardiovascular mortality was increased.[68] The idea that the more common asymptomatic form of PHPT is not associated with increased mortality is supported by data from studies in which more recently enrolled subjects had better survival than those who entered earlier and presumably had more active disease.[69]

Hypertension. Hypertension, a common feature of PHPT when it is part of a MEN syndrome with pheochromocytoma or hyperaldosteronism, has also been reported as being more prevalent in sporadic, asymptomatic PHPT than in appropriately matched control groups. The mechanism of this association is unknown, and the condition does not clearly remit following cure of the hyperparathyroid state.[66]

Coronary Artery Disease and Valve Calcification. Both calcium and PTH have independently been shown to be associated with coronary heart disease.[66] Aside from studies in which coronary atherosclerosis was seen in patients with marked hypercalcemia (16.8–27.4 mg/dL), the more recent literature has been controversial.[66] When CAD is present in PHPT, it is most likely owing to traditional risk factors rather than the disease itself.[66,70] Some studies have actually shown that in mild PHPT there is better exercise tolerance as determined by electrocardiogram.[71] Valve calcification, which is present in severe PHPT, has been shown to be more extensive (greater valve area) when present in those with mild PHPT versus controls[72,73] and is associated with increased PTH levels, but is not reversible with parathyroidectomy.[73]

Left Ventricular Hypertrophy. LVH is a strong predictor of cardiovascular disease and mortality.[66] LVH has been associated with PHPT in many, but not all, studies.[66] A 2015 metaanalysis indicated that parathyroidectomy is associated with a decline in left ventricular mass, and that higher levels of PTH preoperatively predict a greater cardiovascular benefit. However, dissociating disease severity from study design (randomized controlled trials [RCTs] included individuals with lower levels of calcium and PTH than those included in observational studies) was not possible.[74]

Vascular Function. Conflicting data exist regarding whether intima media thickness is increased in PHPT.[66] Multiple studies have reported increased vascular stiffness, sometimes associated with PTH levels, in mild PHPT, but its reversibility with parathyroidectomy is inconsistent.[66] Given conflicting data, most experts do not consider cardiovascular disease to be an indication for parathyroidectomy.[55]

Neurological, Psychologic and Cognitive Features

Descriptions of classical PHPT do indeed indicate neuropsychological features.[1] The extent to which these features remain a part of the modern picture of PHPT, as well as the exact mechanisms underlying them, is unclear. Perhaps the most common complaints have been those of weakness and easy fatigability. Classic PHPT was formerly associated with a distinct neuromuscular syndrome characterized by type II muscle cell atrophy.[75] Originally described by Vicale in 1949, the syndrome consisted of easy fatigability, symmetric proximal muscle weakness, and muscle atrophy. Both the clinical and electromyographic features of this disorder were reversible after parathyroid surgery.[76] In the milder, less symptomatic form of the disease that is common today, this disorder is rarely seen.

The "psychic groans" described by early observers of patients with classic PHPT remain a source of controversy today. Patients with PHPT often report some degree of behavioral and/or psychiatric symptomatology. A retrospective inquiry in patients with more severe disease showed a 23% incidence of psychiatric symptomatology ($n = 441$).[77] A number of studies suggest, however, that even "mild" PHPT (serum calcium <12 mg/dL) is associated with nonspecific symptoms such as depression, anxiety, fatigue, decreased quality of life (QoL), sleep disturbance, and cognitive dysfunction. Many, but not all, observational studies have indicated that these features improve after parathyroidectomy.[78] Three RCTs have investigated the reversibility of reduced QoL and psychiatric symptoms.[79-81] Despite being of similar design and using similar assessment tools, all three RCTs came to different conclusions; one RCT suggested that parathyroidectomy prevents worsening of QoL and improves psychiatric symptoms,[80] another RCT indicated no benefit, and a third RCT demonstrated improvement in QoL.[79,81]

Although less well-studied, some, but not all, studies have demonstrated reduced memory or impairment in other cognitive domains.[78,82] It is unclear if cognition improves after parathyroidectomy, because the results from studies in which longitudinal surgical control groups are compared with those undergoing PTX are inconsistent.[78] Recent work has turned to investigating the potential mechanisms that contribute to cognitive dysfunction in PHPT. Our latest studies have addressed this issue. We hypothesized that cerebrovascular dysfunction (i.e., vascular stiffness) may underlie cognitive changes in patients with PHPT. Although PTH correlated with cerebrovascular function as measured by transcranial Doppler, there was no consistent association between cerebrovascular function and cognitive performance.[83] In a separate study, we used functional magnetic resonance imaging (fMRI) to assess whether cerebral activation was altered by PHPT. fMRI is a noninvasive tool that maps brain function based on changes in blood flow.[84] We found that PHPT was associated with differences in task-related neural activation patterns, but that there no difference in cognitive performance. This may indicate compensation to maintain the same cognitive function, but there was no clear improvement in neural activation after PTX.[84] At present, most experts do not recognize cognitive or psychiatric symptoms as a sole indication for parathyroidectomy. Reasons for this include the failure to clearly demonstrate reversibility in RCTs, the inability to predict which patients may improve, and the lack of a clear mechanism.[55]

Gastrointestinal Manifestations

PHPT has long been considered to be associated with an increased incidence of peptic ulcer disease. Most recent studies suggest that the incidence of peptic ulcer disease in PHPT is approximately 10%, a figure similar to its percentage in the general population. An increased incidence of peptic ulcer disease is seen in patients with PHPT resulting from MEN1, in which approximately 40% of patients have clinically apparent gastrinomas (Zollinger–Ellison syndrome). In those patients, PHPT is associated with increased clinical severity of gastrinoma, and treatment of the associated PHPT has been reported to benefit patients with Zollinger–Ellison syndrome.[85] Despite this, current recommendations (Consensus Conference Guidelines for Therapy of MEN1) state that the coexistence of Zollinger–Ellison syndrome does not represent sufficient indication for parathyroidectomy, because medical therapy is so successful.[85]

Although hypercalcemia can underlie pancreatitis, most large series have not reported an increased incidence of pancreatitis in patients with PHPT with serum calcium levels less than 12 mg/dL. The Mayo Clinic experience from 1950 to 1975 showed that only 1.5% of those with PHPT exhibited coexisting pancreatitis, and alternative explanations for pancreatitis were found for several patients. Regarding pancreatitis in pregnancy in patients with PHPT, these conditions may coexist, but there is no evidence for a causal relationship between the disorders.

Other Systemic Involvement

Many organ systems were affected by the hyperparathyroid state in the past. Anemia, band keratopathy, and loose teeth are no longer part of the clinical syndrome of PHPT. Gout and pseudogout are seen infrequently, and their etiologic relationship to PHPT is not clear.

KEY POINTS

- Subclinical cardiovascular disease, neuropsychiatric and neurocognitive symptoms such as fatigue, reduced concentration and memory, and depression, and reduced quality of life have been reported in patients with primary hyperparathyroidism (PHPT). Specific attribution of these symptoms to PHPT and their reversibility with successful parathyroidectomy remain controversial. Gastrointestinal and rheumatological manifestations are rare, with an uncertain etiologic relationship to PHPT.

UNUSUAL PRESENTATIONS

Neonatal Primary Hyperparathyroidism

Neonatal PHPT is a rare form of the disorder caused by homozygous inactivation of the calcium-sensing receptor.[86] When present in a heterozygous form, it is a benign hypercalcemic state known as familial hypocalciuric hypercalcemia (FHH). However, in the homozygous neonatal form, hypercalcemia is severe, and the outcome is fatal unless recognized early. The treatment of choice is early subtotal parathyroidectomy to remove the majority of hyperplastic parathyroid tissue.

Primary Hyperparathyroidism in Pregnancy

PHPT in pregnancy is primarily of concern for its potential effect on the fetus and neonate. Potential complications of PHPT in pregnancy include spontaneous abortion, low birth weight, supravalvular aortic stenosis, and neonatal tetany. The last condition is a result of fetal parathyroid gland suppression by high levels of maternal calcium, which readily crosses the placenta during pregnancy. These infants, accustomed to hypercalcemia *in utero*, have functional hypoparathyroidism after birth, and can develop hypocalcemia and tetany in the first few days of life. Today, with most patients (pregnant or not) presenting with a mild form of PHPT, an individualized approach to the management of the pregnant patient with PHPT is advised. A recent retrospective study suggested that PHPT did not increase the risk of abortion, birth weight, length, or Apgar score.[87] Thus, many of those with mild disease can be followed safely, with successful neonatal outcomes without surgery. However, parathyroidectomy during the second trimester remains the traditional recommendation for this condition.

Acute Primary Hyperparathyroidism

Acute PHPT is known variously as *parathyroid crisis*, *parathyroid poisoning*, *parathyroid intoxication*, and *parathyroid storm*. Acute PHPT describes an episode of life-threatening hypercalcemia of sudden onset in a patient with PHPT. Clinical manifestations of acute PHPT are mainly those associated with severe hypercalcemia. Nephrocalcinosis or nephrolithiasis is frequently seen. Radiologic evidence of subperiosteal bone resorption is also commonly present. Laboratory evaluation is remarkable, not only for very high serum calcium levels, but also for extremely high PTH levels up to approximately 20 times normal. In this way, acute PHPT resembles, in biochemical terms, parathyroid carcinoma. A history of persistent mild hypercalcemia has been reported in 25% of patients. However, given the rarity of this condition, the risk of developing acute PHPT in a patient with mild asymptomatic PHPT is very low. Intercurrent medical illness with immobilization may precipitate acute PHPT. Early diagnosis with aggressive medical management followed by surgical cure is essential for a successful outcome. The initial impression of patients who present in this manner, without an antecedent history of hypercalcemia, is often mistaken for malignancy. However, the PTH level usually quickly clarifies the diagnosis as PHPT in most situations.

Parathyroid Cancer

An indolent yet potentially fatal disease, parathyroid carcinoma accounts for less than 0.5% of cases of PHPT. In contrast to patients with PHPT caused by benign parathyroid pathology, patients with parathyroid carcinoma typically have marked elevations in serum calcium and PTH. The cause of the disease is unknown, and no clear risk factors have been identified, except for hereditary syndromes. There is no evidence to support the malignant degeneration of previously benign parathyroid adenomas. Parathyroid carcinoma has been reported particularly in hyperparathyroidism-jaw tumor (HPT-JT) syndrome,[89] a rare autosomal disorder in which as many as 15% of patients will have malignant parathyroid disease. Because cystic changes are common, this disorder has also been referred to as *cystic parathyroid adenomatosis.* In HPT-JT, ossifying fibromas of the maxilla and mandible are seen in 30% of cases. Less commonly, kidney lesions, including cysts, polycystic disease, hamartomas, or Wilms tumors, can be present. Parathyroid carcinoma has also been reported in familial isolated hyperparathyroidism, in MEN1, and with somatic *MEN1* mutations.[90-92] However, recurrent parathyroid disease in MEN1 may mimic but might not actually be a result of malignancy. Only one case of parathyroid carcinoma has been reported in the MEN2A syndrome.[93]

Loss of the retinoblastoma tumor suppressor gene was formerly considered a marker for parathyroid cancer, but more recent studies do not unequivocally support this impression.[94] Work by Shattuck et al.[95] has provided new insights into the molecular pathogenesis of parathyroid cancer. Parathyroid carcinomas from 10 of 15 patients with sporadic parathyroid cancer carried a mutation in the *HRPT2* gene. The *HRPT2* gene that encodes for the parafibromin protein was shown to be mutated in a substantial number of patients with parathyroid cancer. Marcocci et al.[96] have reviewed this topic, pointing out a potential role for parafibromin in parathyroid cancer. In three of 15 patients with parathyroid cancer, Shattuck et al.[95] showed that the mutation was in the germline. The presence of the mutation in the germline suggests that this disease may be related in some way to the HPT-JT syndrome, in which this gene has been implicated.[95] In addition, there is an increased risk of parathyroid cancer in the HPT-JT syndrome. In fact, certain clinical features in patients with a germline mutation and in their relatives are indicative of the HPT-JT syndrome or phenotypic variants.[89]

Manifestations of hypercalcemia are the primary effects of parathyroid cancer. The disease tends not to have a bulk tumor effect, spreading slowly in the neck. Metastatic disease is a late finding, with lung (40%), liver (10%), and lymph node (30%) involvement seen most commonly. The clinical profile of parathyroid cancer differs from that of benign PHPT in several important ways.[96] First, no female predominance is seen among patients with carcinoma. Second, elevations in serum calcium and PTH are far greater. Consequently, the hyperparathyroid disease tends to be much more severe, with the classic targets of PTH excess involved in most cases. Nephrolithiasis or nephrocalcinosis is seen in up to 60% of patients; overt radiologic evidence of skeletal involvement is seen in 35% to 90% of patients. A palpable neck mass, distinctly unusual in benign PHPT, has been reported in 30% to 76% of patients with parathyroid cancer. Grossly, malignant glands are large, often exceeding 12 g. They tend to be adherent to adjacent structures. Microscopically, thick fibrous bands divide the trabecular arrangement of the tumor cells. Capsular and blood vessel invasion is common by these cells, which often contain mitotic figures.[96] Treatment is reviewed later.

KEY POINTS

- Parathyroid crisis, an episode of life-threatening hypercalcemia of sudden onset in a patient with primary hyperparathyroidism, is rarely seen today. Parathyroid carcinoma, a rare and indolent, but potentially fatal, condition is characterized by marked elevations in calcium and parathyroid hormone. Men and women are equally affected, with an average age at onset approximately 10 years younger than that of the much more common benign counterpart.

EVALUATION

The diagnosis of PHPT is confirmed by demonstrating an elevated or inappropriately normal PTH level in the face of hypercalcemia. Further biochemical assessment should include serum phosphorus, alkaline phosphatase activity, vitamin D metabolites, albumin, and creatinine. A morning 2-hour or 24-hour urine collection should be obtained for calcium and creatinine. A urinary bone resorption marker such as serum CTX or urinary N-telopeptide can be helpful. Bone densitometry is performed in all patients. It is important to obtain densitometry at three sites: the lumbar spine, the hip, and the distal third of the radius. Because of the differing amounts of cortical and cancellous bone at the three sites and the different effects of PTH on cortical and cancellous bone, measurement at all three sites gives the most accurate clinical assessment of skeletal involvement in PHPT. Bone biopsy is not routinely obtained in the evaluation of PHPT but is essential in research. In the most recent guidelines, spinal imaging is recommended to assess for clinically silent vertebral fractures.[55] This can be vertebral x-rays, vertebral fracture assessment, or TBS score, with the latter two obtained from the DXA image. Although symptomatic kidney stones are present in 15% to 20% of patients by history, the finding that many more have clinically silent nephrolithiasis has led to the recommendation to obtain renal ultrasound, computed tomography (CT), or abdominal x-ray to assess for either nephrolithiasis or nephrocalcinosis.

KEY POINTS

- Primary hyperparathyroidism is confirmed by demonstrating hypercalcemia and an elevated or inappropriately normal parathyroid hormone level. Further evaluation includes serum phosphorus, alkaline phosphatase activity, 25-hydroxyvitamin D, albumin, creatinine, estimated glomerular filtration rate, 24-hour urine calcium, and bone densitometry of the lumbar spine, hip and distal one-third radius. The newest guidelines also recommend obtaining spine and renal imaging to assess for subclinical vertebral fractures, kidney stones, and nephrocalcinosis.

NATURAL HISTORY

Since the early 1990s, new knowledge of the natural history of PHPT with or without surgery has been very helpful in guiding decisions regarding surgery in patients with asymptomatic PHPT. The authors and their colleagues have conducted the longest prospective observational trial.[22,97] This project began in 1984 in an effort to define the natural history of asymptomatic PHPT. The study included detailed analyses of pathophysiologic, densitometric, histomorphometric, and other skeletal features of PHPT.[22,97] Much of the information gleaned from that study has been summarized already in this chapter. The 15-year follow-up to this study constitutes the longest natural-history study of this disorder.[97]

Recommendations for surgery or observation were made based on the 1990 set of National Institutes of Health (NIH) guidelines, but both groups included patients who did or did not meet surgical guidelines. This is because some patients opted for surgery even if they did not meet the guidelines, whereas others opted for a conservative approach even if they did meet guidelines for surgery. As will be described in the following, this imperfectly designed study was followed by three studies that were truly randomized but were of shorter duration. The results with regard to natural history from all studies are remarkably concordant.

Natural History With Surgery

Successful parathyroidectomy results in permanent normalization of the serum calcium and PTH levels. Postoperatively, there is a marked improvement in BMD at all sites (lumbar spine, femoral neck, and distal one-third radius) amounting to gains greater than 10%.[22] The improvement is most rapid at the lumbar spine, followed by gains at the hip regions and the distal one-third radius during the 15-year follow-up (Fig. 54.6).[97] The improvements were seen in those who met and did not meet surgical criteria at study entry, confirming the salutary effect of parathyroidectomy in this regard on all patients.

Natural History Without Surgery

In subjects who did not undergo parathyroid surgery, serum calcium remained stable for approximately 12 years, with a tendency for the serum calcium level to rise in years 13 to 15.[97] Other biochemical indices such as PTH, vitamin D metabolites, and urinary calcium did not change for the entire 15 years of follow-up in the group as a whole. Bone density at all three sites remained stable for the first 8 to 10 years. However, after this period of stability, declining cortical BMD was seen at the hip and more dramatically at the distal one-third radius site. Although the numbers became limiting after 10 years of follow-up, it is noteworthy that a small majority of the subjects lost more than 10% of their BMD during the 15 years of observation. Even though this decline was observed in the majority of subjects, only 37% of subjects met one or more guidelines for surgery after the 15 years of follow-up.

Randomized Studies of the Natural History of Asymptomatic Primary Hyperparathyroidism

The long natural-history study of asymptomatic PHPT has added much to our knowledge about this disease throughout time. Subsequent randomized trials confirm the observational data but are limited by their shorter duration. In 2004, Rao et al.[98] reported on 53 subjects, assigned either to parathyroid surgery ($n = 25$) or to no surgery ($n = 28$). The follow-up lasted for at least 2 years. There was a significant effect of parathyroidectomy on BMD at the hip and femoral neck, but not the spine or forearm. In 2007 Bollerslev et al.[79] reported the interim results of their randomized trial of parathyroidectomy versus no surgery. This study from three Scandinavian countries was larger, with 191 patients who were randomized to medical observation or to surgery. After surgery, biochemical indices normalized, and BMD increased. In the group that did not undergo parathyroid surgery, BMD did not change. Also in 2007, Ambrogini et al.[81] reported the results of their randomized controlled trial of parathyroidectomy versus observation. Surgery

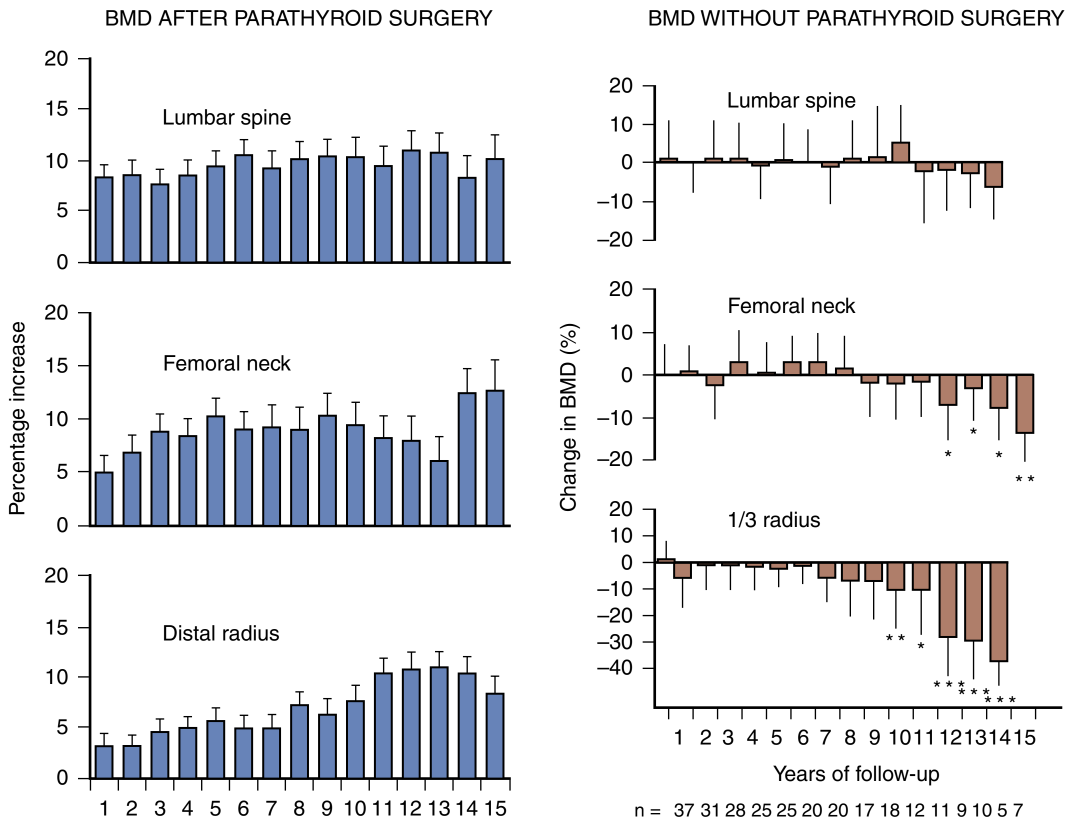


Fig. 54.6 Longitudinal course of bone density in primary hyperparathyroidism. Data are presented as percentage change from preoperative baseline bone mineral density (*BMD*) measurement by site following parathyroidectomy (*left*) or in patients followed with no intervention (*right*). (Data from Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years, *J Clin Endocrinol Metab.* 2008;93:3462–3470.)

was associated with a significant increase in BMD of the lumbar spine and hip after 1 year.

Whether fracture risk decreases after parathyroidectomy is not clear. The study by Bollerslev reported on vertebral fracture risk reduction at 5 years after initial treatment allocation. That study indicated that successful PTX versus observation was associated with a reduction in vertebral fracture risk that was of borderline statistical significance.[99]

KEY POINTS

- Successful parathyroidectomy results in cure. Serum calcium and parathyroid hormone levels become normal. Postoperatively, bone mineral density (BMD) improves at all sites. In those who are observed without parathyroidectomy, serum calcium and other biochemical indices remain stable for many years, but BMD may decline at cortical sites after 8 to 10 years. Approximately 40% of patients will develop an indication for surgery with long-term follow-up.

Guidelines for Parathyroidectomy

Parathyroidectomy remains the only currently available option to cure PHPT. As the disease profile has changed, questions have arisen concerning the advisability of surgery in asymptomatic patients. If asymptomatic patients have a benign natural history, the surgical alternative is not an attractive one. On the other hand, asymptomatic patients may display levels of hypercalcemia or hypercalciuria that cause concern for the future. Similarly, bone-mass measurements can be frankly low. In an effort to address such issues, there have been four consensus development conferences (in 1991, 2002, 2008, and 2013) on the management of asymptomatic PHPT. The most recent guidelines that emerged from the 2013 conference are helpful to the clinician faced with the asymptomatic hyperparathyroid patient: *all* symptomatic patients are advised to undergo parathyroidectomy.[55] Surgery is advised in asymptomatic patients with (1) serum calcium greater than 1 mg/dL higher than the upper limit of normal; (2) renal guidelines: reduction in creatinine clearance to less than 60 mL/min; urinary calcium excretion greater than 400 mg/24 h with increased stone risk; or presence of nephrolithiasis or nephrocalcinosis on renal imaging; (3) skeletal guidelines: reduced bone density T-score less than −2.5 at any site; or vertebral compression fracture on an imaging study; and (4) age younger than 50 years. The most recent guidelines are shown in Table 54.5. A noteworthy change in the guidelines reflects the fact that asymptomatic kidney stones and vertebral compression fractures are now considered indications for parathyroidectomy.

A number of points were discussed that did not lead to specific recommendations, including the issues of the neurocognitive and cardiovascular aspects of PHPT. The workshop panel also acknowledged a potential role of vitamin D deficiency in fueling processes associated with abnormal parathyroid glandular activity. Finally, the panel also reaffirmed the entity of NPHPT, but noted that there are insufficient data to provide evidence-based guidelines for management.

KEY POINTS

- Parathyroidectomy remains the only cure for primary hyperparathyroidism. *All* symptomatic patients should undergo parathyroidectomy. Surgery is advised in asymptomatic patients with (1) serum calcium greater than 1 mg/dL higher than the upper limit of normal; (2) reduction in creatinine clearance to less than 60 mL/min; urinary calcium excretion over 400 mg/24 hours with increased stone risk; or presence of nephrolithiasis or nephrocalcinosis on renal imaging; (3) reduced bone density T-score less than –2.5 at any site; or vertebral compression fracture on imaging; and (4) age younger than 50 years.

SURGERY

A large percentage of those patients who meet the surgical guidelines listed in Table 54.5 are asymptomatic. Some asymptomatic patients who meet surgical guidelines elect not to have surgery for varying reasons, including personal choice, intercurrent medical conditions, and previous unsuccessful parathyroid surgery. Conversely, there are patients who meet none of the NIH guidelines for parathyroidectomy but opt for surgery nevertheless. Physician and patient input remain important factors in the decision regarding parathyroid surgery.

Preoperative Localization of Hyperfunctioning Parathyroid Tissue

A number of imaging tests have been developed and have been applied singly or in combination to address the challenge of preoperative localization. The rationale for locating abnormal parathyroid tissue before surgery is that the glands can be notoriously unpredictable in their location. Although most parathyroid adenomas are identified in regions proximate to their embryologically intended position (the four poles of the thyroid gland), many are not. In such situations, previous surgical experience and skill are needed to locate the ectopic parathyroid gland. In such hands, 95% of abnormal parathyroid glands will be discovered and removed at the time of initial parathyroid surgery. However, in the patient with previous neck surgery, even expert parathyroid surgeons do not generally achieve such high success rates. Preoperative localization of the abnormal parathyroid tissue can be extremely helpful under these circumstances. Preoperative imaging is also necessary for any patient who will undergo parathyroidectomy using a minimally invasive approach. Imaging studies should not be used for the diagnosis of PHPT, because the sensitivity and specificity of various imaging modalities varies, with some having false-positive rates as high as 25%.[100]

Noninvasive Imaging. Noninvasive parathyroid imaging studies include technetium (Tc)-99m sestamibi scintigraphy, ultrasound, CT scanning, magnetic resonance imaging (MRI), and positron emission tomography (PET) scanning (see also Chapter 56). Tc-99m sestamibi is generally regarded to be the most sensitive and specific imaging modality, especially when it is combined with single-photon emission CT (SPECT). For the single parathyroid adenoma, sensitivity has ranged from 80% to 100%, with a 5% to 10% false-positive rate. On the other hand, sestamibi scintigraphy and the other localization tests have a relatively poor record in the context of multiglandular disease.[100] The success of ultrasonography is highly operator-dependent. In centers where there is great expertise, this noninvasive approach is most attractive. Abnormalities identified by ultrasound as possible parathyroid tissue may prove to be a thyroid nodule or lymph node, which underscores the importance of the skill and experience of the ultrasonographer. Rapid spiral thin-slice CT scanning of the neck and mediastinum with evaluation of axial, coronal, and sagittal views can add much to the search for elusive parathyroid tissue, albeit with attendant higher radiation exposure. Four-dimensional (4D) CT has emerged as a promising method and consists of multiphase CT acquired at noncontrast, contrast-enhanced, arterial, and delayed phases. In a recent study, 4D CT was superior to sestamibi SPECT/CT.[101] MRI can also identify abnormal parathyroid tissue, but it is time-consuming and expensive. It is also less sensitive than the other noninvasive modalities. It can nonetheless be useful when the search with these other noninvasive approaches has been unsuccessful. PET with or without simultaneous CT scan can be used, but, like MRI, it is expensive and does not have the kind of experiential basis that make it attractive.[100] There are also specificity issues, because (^{18}F)-fluorodeoxyglucose, the scanning agent, accumulates in the thyroid, making differentiation between parathyroid adenoma and thyroid nodules difficult.

Invasive Imaging

Parathyroid Fine-Needle Aspiration. Fine-needle aspiration of a parathyroid gland, identified by any of the aforementioned modalities, can be performed, and the aspirate analyzed for PTH. This technique is not recommended for routine *de novo* cases. A theoretical concern with this approach is the possibility that parathyroid cells could be

TABLE 54.5 Comparison of New and Old Guidelines for Surgery in Asymptomatic Primary Hyperparathyroidism

	1990 National Institutes of Health Consensus Conference	2002 National Institutes of Health Workshop	2008 International Workshop	2013 International Workshop
Serum calcium	1- to 1.6-mg/dL elevation	1.0-mg/dL elevation	1.0-mg/dL elevation	1.0-mg/dL elevation
Renal	24-h urine calcium >400 mg Creatinine clearance reduced by 30%	24-h urine calcium >400 mg Creatinine clearance reduced by 30%	No 24-h urine Creatinine clearance: <60 mL/min	24-h urine for FHH/stone risk Urine calcium >400 mg/day Creatinine clearance: <60 mL/min Calcification on renal imaging
Bone	Z-score <–2.0 in forearm	T-score <–2.5 at any site	T-score <–2.5 Fragility fracture	T-score <–2.5 Vertebral fracture on imaging
Age	<50	<50	<50	<50

FHH, Familial hypercalciuric hypercalcemia.

Columns 3 and 4 modified from Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab.* 2014;99:3561–3569.

deposited outside the parathyroid gland in the course of the aspiration. Autoseeding of parathyroid tissue would be an unwanted consequence of this procedure if it were to occur.

Arteriography and Selective Venous Sampling for Parathyroid Hormone. In situations where the gland has not been identified by any of the techniques described, the combination of arteriography and selective venous sampling can provide both anatomic and functional localization of abnormal parathyroid tissue. This approach, however, is costly and requires an experienced interventional radiologist. It is also performed in only a few centers in the United States. This approach is reserved for those individuals who have undergone previous unsuccessful parathyroid surgery in whom all other localization techniques have failed.

Surgical Approach

In the hands of an expert parathyroid surgeon, parathyroidectomy is successful, with infrequent complications. A standard surgical approach is the four-gland parathyroid exploration under general or local anesthesia, with or without preoperative localization. This approach has been reported to lead to surgical cure in more than 95% of cases (see Chapter 56).[102] Several alternative approaches have emerged that focus on the single gland and not the total four-gland neck exploration that was routinely used in the past. Unilateral approaches are appealing for a disease in which most often only a single gland is involved. These procedures include a unilateral operation in which the gland on the same side that harbors the adenoma is ascertained to be normal. Because multiple parathyroid adenomas are unusual, a normal parathyroid gland is considered by some to be sufficient evidence for single-gland disease. Another limited surgical approach that has emerged in many centers as the approach of choice is the minimally invasive parathyroidectomy (MIP). Preoperative parathyroid imaging is necessary, and the procedure is directed only to the site where the abnormal parathyroid gland has been visualized. Preoperative blood is obtained for comparison of the PTH concentration with an intraoperative sample(s) obtained after removal of the "abnormal" parathyroid gland. The availability of a rapid PTH assay in or near the operating room is necessary for this procedure. If the 10-minute postexcision PTH level falls by more than 50% compared with baseline, and into the normal range, the gland that has been removed is considered to be the sole source of overactive parathyroid tissue, and the operation is terminated. If the PTH level does not fall by more than 50%, into the normal range, the operation is extended to a more traditional one in a search for other overactive parathyroid tissue. There is a risk (albeit small) that the minimally invasive procedure may miss other overactive gland(s) that are suppressed in the presence of a dominant gland. The MIP procedure seems to be as successful, in the range of 95% to 98%, as more standard approaches. According to a recent metaanalysis that included more than 12,000 patients, MIP was associated with similar rates of success, disease recurrence, persistence, overall failure, and reoperation.[103] The operative time was significantly shorter, with a lower overall complication rate for MIP compared with bilateral neck exploration. In Europe, MIP is being performed with an endoscopic camera, but this does not offer any advantage other than a smaller incision.[104] Yet another variation on this theme is the use of preoperative sestamibi scanning with an intraoperative gamma probe to help locate enlarged parathyroid glands.

Immediate Postoperative Course

After surgery, biochemical indices return rapidly to normal.[22,97] Serum calcium and PTH levels normalize, and urinary calcium excretion falls by as much as 50%. Serum calcium levels no longer fall rapidly into the hypocalcemic range, a situation characteristic of an earlier time when PHPT was a symptomatic disease with overt skeletal involvement. The acute reversal of PHPT was associated with a robust deposition of calcium into the skeleton at a pace that could not be compensated for by supplemental calcium. Thus, postoperative hypocalcemia was routine and was sometimes a serious short-term complication ("hungry bone syndrome"). Occasionally, postoperative hypocalcemia still occurs, especially if preoperative bone turnover markers are markedly elevated or there is concomitant vitamin D deficiency. More typically, however, the early postoperative course is not complicated by symptomatic hypocalcemia.

After successful parathyroid surgery, biochemical indices of the disease return to normal, and BMD improves, as mentioned (Fig. 54.7). The capacity of the skeleton to restore itself is seen dramatically in

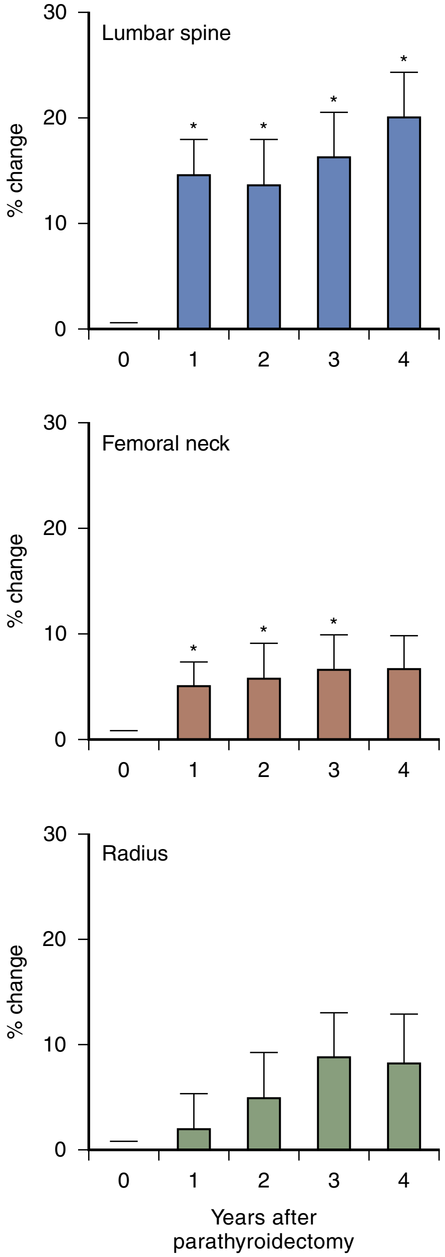


Fig. 54.7 Bone mineral density following parathyroidectomy in 14 primary hyperparathyroid patients with vertebral osteopenia. The cumulative percentage change from preoperative baseline (year 0) is shown for each site. *Asterisk* denotes change from baseline at $P < .05$. (Data from Silverberg SJ, Locke FG, Bilezikian JP. Vertebral osteopenia: a new indication for surgery in primary hyperparathyroidism, *J Clin Endocrinol Metab.* 1996;81:4007–4012.)

young patients with severe PHPT. Kulak et al.[105] reported two patients with *osteitis fibrosa cystica* who experienced increases in bone density that ranged from 260% to 430% in a period of 3 to 4 years following surgery.

KEY POINTS

- In the hands of an expert parathyroid surgeon, parathyroidectomy is successful in >95% of patients, with infrequent complications. The standard surgical approach is the four-gland parathyroid exploration. Most experts recommend preoperative parathyroid imaging for localization. Alternative approaches have emerged that focus on removal of the single abnormal gland, if identified preoperatively, as well as intraoperative parathyroid hormone measurements.

TABLE 54.6 Comparison of New and Old Management Guidelines for Patients With Asymptomatic Primary Hyperparathyroidism Who Do Not Undergo Parathyroid Surgery

Measurement	Older Guidelines	Newer Guidelines
Serum Calcium	*Semiannually*	*Annually*
24-h urinary calcium	Annually	Not recommended
Creatinine clearance	Annually	Not recommended
Serum creatinine	Annually	Annually
Bone density	Annually	Annually or biannually
Abdominal x-ray	Annually	Not recommended

From Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab.* 2014;99:3561–3569.

MEDICAL MANAGEMENT

Patients who do not meet any surgical guidelines are often followed without intervention. The most recent guidelines for management of asymptomatic PHPT restated the position that it is reasonable to pursue a nonsurgical course of management for those who do not meet criteria for surgery, at least for a period of years. In those patients who are not going to have parathyroid surgery, the workshop[55] suggested a set of monitoring steps that are summarized in Table 54.6. This includes annual measurements of the serum calcium concentration, a calculated creatinine clearance, and regular monitoring of BMD.

Ideal medical therapy of PHPT would provide the equivalent to a medical parathyroidectomy. Such an agent would normalize serum calcium and PTH levels, as well as urinary calcium excretion, increase BMD and lower fracture risk, and reduce the risk of kidney stones. Unfortunately, no currently available single drug meets all these criteria. The following medications can achieve some of these goals and may be considered in patients not having surgery in whom it is desirable to lower serum or urinary calcium levels or increase BMD.

General Measures

Patients should be instructed to remain well hydrated and to avoid, if possible, medications that can increase serum calcium (e.g., thiazide diuretics). Prolonged immobilization, which can raise the serum calcium concentration further and induce hypercalciuria, should also be avoided.

Diet and Supplements

Dietary management of PHPT has long been an area of controversy. Many patients are advised to limit their dietary calcium intake because of the hypercalcemia. However, it is well known that low dietary calcium can lead to increased PTH levels in normal individuals.[106] In patients with PHPT, even though the abnormal PTH tissue is not as sensitive to slight perturbations in the circulating calcium concentration, it is still possible that PTH levels will rise when dietary calcium is tightly restricted. Dietary calcium could also be variably influenced by ambient levels of 1,25-dihydroxyvitamin D. In patients with normal levels of 1,25-dihydroxyvitamin D, Locker et al.[107] noted no difference in urinary calcium excretion between those on high (1000 mg/day) and low (500 mg/day) calcium intake diets. On the other hand, in those with elevated levels of 1,25-dihydroxyvitamin D, high-calcium diets were associated with worsening hypercalciuria. This observation suggests that dietary calcium intake in patients can be liberalized to 1000 mg/day if 1,25-dihydroxyvitamin D levels are not increased but should be more tightly controlled if 1,25-dihydroxyvitamin D levels are elevated. Although calcium supplements are not specifically recommended in those with PHPT and osteoporosis, small doses do not seem to exacerbate hypercalcemia or hypercalciuria if the diet is deficient.[8] Most experts recommend that patients with PHPT who are going to be followed without surgery have an intake of calcium that is consistent with nutritional guidelines for a normal population.

Recent guidelines recommend maintaining 25-hydroxyvitamin D at levels of 21 to 30 ng/mL, with conservative doses of vitamin D (600–1000 IU daily) based on data showing that vitamin D repletion lowers PTH levels. Higher levels of vitamin D may be beneficial. A 2014 RCT of cholecalciferol (2800 IU daily versus placebo) indicated that treatment increased 25-hydroxyvitamin D levels from 20 ng/mL to 37.8 ng/mL, lowered levels of PTH, and increased lumbar spine BMD, without having a deleterious effect on serum or urinary calcium levels.[108]

Pharmaceuticals

Phosphate. Oral phosphate can lower the serum calcium by up to 1 mg/dL.[109] A complex interplay of mechanisms leads to this moderating effect of oral phosphate. First, calcium absorption falls in the presence of intestinal phosphorus. Second, concomitant increases in serum phosphorus will tend to reduce circulating 1,25-dihydroxyvitamin D levels. Third, phosphate can be an antiresorptive agent. Finally, increased serum phosphorus reciprocally lowers serum calcium. Problems with oral phosphate include limited gastrointestinal tolerance, possible further increase in PTH levels, and the possibility of soft-tissue calcifications after long-term use. It is essentially not used any longer in the management of PHPT.

Estrogens and Selective Estrogen-Receptor Modulators. The earliest studies on the use of estrogen replacement therapy in PHPT date back to the early 1970s and indicated a 0.5- to 1.0-mg/dL reduction in total serum calcium levels, along with a lowering of urinary calcium.[8] Most studies indicated no change in PTH.[8] An RCT of conjugated estrogen (0.625 mg daily plus medroxyprogesterone 5 mg daily) versus placebo indicated that hormone-replacement therapy effectively increases BMD at all skeletal sites in patients with PHPT, with the greatest increases at the lumbar spine.[110] This RCT, however, did not confirm the calcium-lowering effect of earlier uncontrolled studies. In view of concerns expressed about chronic estrogen use in the Women's Health Study, estrogen use is not often recommended for medical management of hyperparathyroidism.

Raloxifene, a selective estrogen-receptor modulator, has been studied in PHPT, but the data are sparse. In a short-term (8-week) trial of 18 postmenopausal women, raloxifene (60 mg/day) was associated with a statistically significant, although small (0.5 mg/dL), reduction in the serum calcium concentration and in markers of bone turnover.[111] No long-term data or data on bone density are available.

Bisphosphonates and Denosumab. Bisphosphonates are conceptually attractive in PHPT because they are antiresorptive agents with an overall effect of reducing bone turnover. Although they do not affect PTH secretion directly, bisphosphonates could reduce serum and urinary calcium levels. Early studies with the first-generation bisphosphonates were disappointing, showing either no effect or only transient reductions in serum and urinary calcium.[8,112]

Alendronate has been studied most extensively in PHPT.[8] Several randomized controlled studies demonstrate that BMD of the lumbar spine and hip regions increases along with reductions in bone turnover markers.[8] In most studies, serum calcium was unchanged.[8] These results suggest that bisphosphonates may be useful in patients with low bone density in whom parathyroid surgery is not to be performed. One small study suggests that denosumab increases BMD in women with PHPT to a greater extent than patients without PHPT but with osteoporosis.[113]

Inhibition of Parathyroid Hormone. The most specific pharmacologic approach to PHPT is to inhibit the synthesis and secretion of PTH from the parathyroid glands, such as those that act on the parathyroid cell calcium-sensing receptor. This G protein-coupled receptor recognizes calcium as its cognate ligand. When activated by increased extracellular calcium, the calcium-sensing receptor signals the cell via a G protein-transducing pathway to raise the intracellular calcium concentration, which inhibits PTH secretion. Molecules that mimic the effect of extracellular calcium by altering the affinity of calcium for the receptor could activate this receptor and inhibit parathyroid cell function. Cinacalcet is one such compound and is now approved for the treatment of severe hypercalcemia in PHPT when surgery cannot be pursued. Studies conducted by the authors and their colleagues[114] indicate that this drug can reduce the serum calcium concentration to normal in PHPT (Fig. 54.8), but despite normalization of the serum calcium concentration, PTH levels do not return to normal; they do fall by 35% to 50% after administration of the drug. Urinary calcium excretion does not change; serum phosphorus levels increase but are maintained in the lower range of normal; and 1,25-dihydroxyvitamin D levels do not change. The average BMD does not change, even after 5 years of administration of cinacalcet.[114] Silverberg et al.[115] have shown that cinacalcet reduces calcium levels effectively in inoperable parathyroid carcinoma.

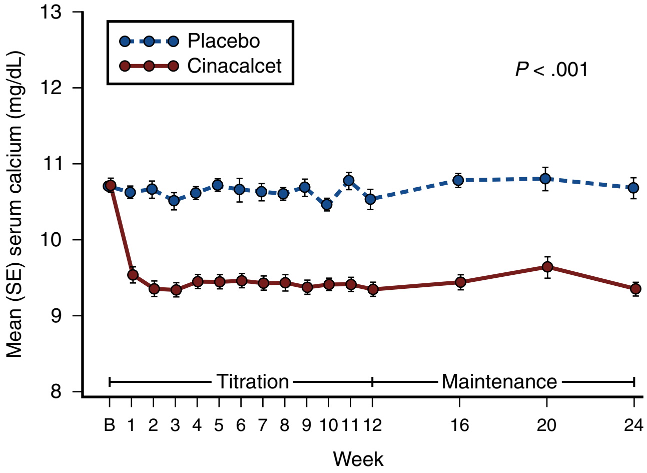


Fig. 54.8 Changes in serum calcium concentrations with administration of the calcimimetic cinacalcet (*solid line*) or placebo (*broken line*) in patients with primary hyperparathyroidism. *SE,* Standard error. (Data from Shoback DM, Bilezikian JP, Turner SA, et al. The calcimimetic cinacalcet normalizes serum calcium in subjects with primary hyperparathyroidism, *J Clin Endocrinol Metab.* 2003;88:5644–5649.)

HYDROCHLOROTHIAZIDE

Though avoiding agents that exacerbate hypercalcemia is generally recommended in patients with PHPT, a recent study has recently reexamined the role of thiazides in patients with PHPT. This retrospective analysis of 72 patients suggested that thiazides may not increase serum levels of calcium in PHPT as they can do in normal individuals. Hydrochlorothiazide (12.5–50 mg daily for 3.1 years on average) was associated with a decrease in urinary calcium excretion but no change in serum levels of calcium.[116] It is unclear if hydrochlorothiazide reduces the risk of nephrolithiasis. Given the heterogeneity of doses used, and the absence of data from larger, (preferably) randomized trials, recommending thiazide use routinely in PHPT is premature. However, thiazides could be considered in those who refuse surgery or who are poor surgical candidates but at high risk of nephrolithiasis in whom the benefit is thought to outweigh the risk, as long as serum levels of calcium are monitored regularly.

Treatment of Parathyroid Cancer

Surgery is the only effective therapy currently available for parathyroid cancer. The greatest chance for cure occurs with the first operation. After the disease recurs, cure is unlikely, although the disease may smolder for many years thereafter. The tumor is not radiosensitive, although there are isolated reports of tumor regression with localized radiation therapy. Traditional chemotherapeutic agents have not been useful. When metastasis occurs, isolated removal is an option, especially if only one or two nodules are found in the lung. Such isolated metastasectomies are never curative, but they can lead to prolonged remissions, sometimes lasting for several years. Similarly, local debulking of tumor tissue in the neck can be palliative, although malignant tissue is invariably left behind.

Chemotherapy has had a very limited role in this disease. Bradwell and Harvey[117] have attempted an immunotherapeutic approach by injecting a patient who had severe hypercalcemia resulting from parathyroid cancer with antibodies raised to her own circulating PTH. Coincident with a rise in antibody titer to PTH, previously refractory hypercalcemia fell impressively. A more recent report[118] provided evidence of the antitumor effect in a single case of PTH immunization in metastatic parathyroid cancer. Attention has been focused instead on control of hypercalcemia. Intravenous bisphosphonates have been used to treat severe hypercalcemia. Although efficacious in the short term, they do not provide an approach that allows long-term outpatient normalization of serum calcium levels. Denosumab has more recently been reported to treat resistant hypercalcemia in patients with parathyroid carcinoma.[119]

The calcimimetic agents offer a newer approach. Silverberg et al.[115] showed that cinacalcet is useful in the management of parathyroid cancer. The US Food and Drug Administration approved this calcimimetic for the treatment of hypercalcemia in patients with parathyroid cancer. Use of this agent in parathyroid cancer led to improvement in serum calcium levels and a decrease in symptoms of nausea, vomiting, and mental lethargy, which are common concomitants of marked hypercalcemia. There are no data on the effect of cinacalcet on tumor growth in parathyroid cancer. Similarly, there are no data on the use of a combination of cinacalcet and bisphosphonates in parathyroid cancer, the former used to decrease PTH secretion from the cancer, and the latter used to decrease release of calcium from the skeleton. Cinacalcet offers an option for control of intractable hypercalcemia when surgical removal of cancerous tissue is no longer an option.

KEY POINTS

- For those who will not or cannot undergo parathyroid surgery, medical alternatives include cinacalcet to reduce the serum calcium or bisphosphonates or denosumab to increase bone density. These approaches are not curative.

SUMMARY

This chapter has presented a comprehensive summary of the modern-day presentation of PHPT. Typically an asymptomatic disorder in countries that are economically more developed, the disorder's presentation has raised issues regarding the extent to which such patients may nevertheless show subclinical target organ involvement, who should be recommended for parathyroid surgery, and who can be safely followed without surgical intervention, as well as questions regarding the role of medical therapy. Questions about the natural history and pathophysiology of the disorder continue to be of great interest. Inasmuch as this disorder continues to evolve, it is clear that additional careful studies are required continually to gain new insights into this disease.

Acknowledgment

This work was supported in part by National Institutes of Health grants NIDDK 32333, DK084986, RR 00645, and R21DK104105.

REFERENCES

1. Albright F, Aub J, Bauer W. Hyperparathyroidism: common and polymorphic condition as illustrated by seventeen proven cases in one clinic. *J Am Med Assoc*. 1934;102:1276.
2. Heath 3rd H, Hodgson SF, Kennedy MA. Primary hyperparathyroidism. Incidence, morbidity, and potential economic impact in a community. *N Engl J Med*. 1980;302:189–193.
3. Bendz H, Sjodin I, Toss G, Berglund K. Hyperparathyroidism and long-term lithium therapy—a cross-sectional study and the effect of lithium withdrawal. *J Intern Med*. 1996;240:357–365.
4. Rao SD, Frame B, Miller MJ, et al. Hyperparathyroidism following head and neck irradiation. *Arch Intern Med*. 1980;140:205–207.
5. Vaidya A, Curhan GC, Paik JM, et al. Body size and the risk of primary hyperparathyroidism in women: a cohort study. *J Bone Miner Res*. 2017;32:1900–1906.
6. Paik JM, Curhan GC, Taylor EN. Calcium intake and risk of primary hyperparathyroidism in women: prospective cohort study. *BMJ*. 2012;345:e6390.
7. Bilezikian JP, Bandeira L, Khan A, et al. Hyperparathyroidism. *Lancet*. 2018;391:168–178.
8. Walker MD, Silverberg SJ. Primary hyperparathyroidism. *Nat Rev Endocrinol*. 2018;14:115–125.
9. Griebeler ML, Kearns AE, Ryu E, et al. Secular trends in the incidence of primary hyperparathyroidism over five decades (1965-2010). *Bone*. 2015;73:1–7.
10. Yeh MW, Ituarte PH, Zhou HC, et al. Incidence and prevalence of primary hyperparathyroidism in a racially mixed population. *J Clin Endocrinol Metab*. 2013;98:1122–1129.
11. Lowe H, McMahon DJ, Rubin MR, et al. Normocalcemic primary hyperparathyroidism: further characterization of a new clinical phenotype. *J Clin Endocrinol Metab*. 2007;92:3001–3005.
12. Cusano NE, Maalouf NM, Wang PY, et al. Normocalcemic hyperparathyroidism and hypoparathyroidism in two community-based nonreferral populations. *J Clin Endocrinol Metab*. 2013;98:2734–2741.
13. Schini M, Jacques RM, Oakes E, et al. Normocalcemic hyperparathyroidism: study of its prevalence and natural history. *J Clin Endocrinol Metab*. 2020;105:e1171–e1186.
14. Thakker RV. Genetics of parathyroid tumours. *J Intern Med*. 2016;280:574–583.
15. Souberbielle JC, Cormier C, Kindermans C, et al. Vitamin D status and redefining serum parathyroid hormone reference range in the elderly. *J Clin Endocrinol Metab*. 2001;86:3086–3090.
16. Walker MD, Cong E, Lee JA, et al. Vitamin D in primary hyperparathyroidism: effects on clinical, biochemical, and densitometric presentation. *J Clin Endocrinol Metab*. 2015;100:3443–3451.
17. Tassone F, Gianotti L, Emmolo I, et al. Glomerular filtration rate and parathyroid hormone secretion in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2009;94:4458–4461.
18. Lepage R, Roy L, Brossard JH, et al. A non-(1-84) circulating parathyroid hormone (PTH) fragment interferes significantly with intact PTH commercial assay measurements in uremic samples. *Clin Chem*. 1998;44:805–809.
19. John MR, Goodman WG, Gao P, et al. A novel immunoradiometric assay detects full-length human PTH but not amino-terminally truncated fragments: implications for PTH measurements in renal failure. *J Clin Endocrinol Metab*. 1999;84:4287–4290.
20. Silverberg SJ, Gao P, Brown I, et al. Clinical utility of an immunoradiometric assay for parathyroid hormone (1-84) in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2003;88:4725–4730.
21. Rubin MR, Silverberg SJ, D'Amour P, et al. An N-terminal molecular form of parathyroid hormone (PTH) distinct from hPTH(1 84) is overproduced in parathyroid carcinoma. *Clin Chem*. 2007;53:1470–1476.
22. Silverberg SJ, Shane E, Jacobs TP, et al. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. *N Engl J Med*. 1999;341:1249–1255.
23. Silverberg SJ, Shane E, Dempster DW, et al. The effects of vitamin D insufficiency in patients with primary hyperparathyroidism. *Am J Med*. 1999;107:561–567.
24. Zhao L, Liu JM, He XY, et al. The changing clinical patterns of primary hyperparathyroidism in Chinese patients: data from 2000 to 2010 in a single clinical center. *J Clin Endocrinol Metab*. 2013;98:721–728.
25. Broulik PD, Horky K, Pacovsky V. Blood pressure in patients with primary hyperparathyroidism before and after parathyroidectomy. *Exp Clin Endocrinol*. 1985;86:346–352.
26. Bilezikian JP, Connor TB, Aptekar R, et al. Pseudogout after parathyroidectomy. *Lancet*. 1973;1:445–446.
27. Ludvigsson JF, Kampe O, Lebwohl B, et al. Primary hyperparathyroidism and celiac disease: a population-based cohort study. *J Clin Endocrinol Metab*. 2012;97:897–904.
28. Silverberg SJ, Gartenberg F, Jacobs TP, et al. Longitudinal measurements of bone density and biochemical indices in untreated primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1995;80:723–728.
29. Deftos LJ, Parthemore JG, Price PA. Changes in plasma bone GLA protein during treatment of bone disease. *Calcif Tissue Int*. 1982;34:121–124.
30. van Lierop AH, Witteveen JE, Hamdy NA, et al. Patients with primary hyperparathyroidism have lower circulating sclerostin levels than euparathyroid controls. *Eur J Endocrinol*. 2010;163:833–837.
31. Seibel MJ, Gartenberg F, Silverberg SJ, et al. Urinary hydroxypyridinium cross-links of collagen in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1992;74:481–486.
32. Guo CY, Thomas WE, al-Dehaimi AW, et al. Longitudinal changes in bone mineral density and bone turnover in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1996;81:3487–3491.
33. Minisola S, Romagnoli E, Scarnecchia L, et al. Serum carboxy-terminal propeptide of human type I procollagen in patients with primary hyperparathyroidism: studies in basal conditions and after parathyroid surgery. *Eur J Endocrinol*. 1994;130:587–591.
34. Silverberg SJ, Shane E, de la Cruz L, et al. Skeletal disease in primary hyperparathyroidism. *J Bone Miner Res*. 1989;4:283–291.
35. Dempster DW, Cosman F, Parisien M, et al. Anabolic actions of parathyroid hormone on bone. *Endocr Rev*. 1993;14:690–709.

36. Bilezikian JP, Silverberg SJ, Shane E, et al. Characterization and evaluation of asymptomatic primary hyperparathyroidism. *J Bone Miner Res*. 1991;6(suppl 2):S85–S89; discussion S121–S124.
37. Silverberg SJ, Locker FG, Bilezikian JP. Vertebral osteopenia: a new indication for surgery in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1996;81:4007–4012.
38. Walker MD, Nishiyama KK, Zhou B, et al. Effect of low vitamin D on volumetric bone mineral density, bone microarchitecture, and stiffness in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2016;101:905–913.
39. Parisien M, Silverberg SJ, Shane E, et al. The histomorphometry of bone in primary hyperparathyroidism: preservation of cancellous bone structure. *J Clin Endocrinol Metab*. 1990;70:930–938.
40. van Doorn L, Lips P, Netelenbos JC, et al. Bone histomorphometry and serum concentrations of intact parathyroid hormone (PTH(1-84)) in patients with primary hyperparathyroidism. *Bone Miner*. 1993;23:233–242.
41. Dempster DW, Parisien M, Silverberg SJ, et al. On the mechanism of cancellous bone preservation in postmenopausal women with mild primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1999;84:1562–1566.
42. Parisien M, Mellish RW, Silverberg SJ, et al. Maintenance of cancellous bone connectivity in primary hyperparathyroidism: trabecular strut analysis. *J Bone Miner Res*. 1992;7:913–919.
43. Parisien M, Cosman F, Mellish RW, et al. Bone structure in postmenopausal hyperparathyroid, osteoporotic, and normal women. *J Bone Miner Res*. 1995;10:1393–1399.
44. Dempster DW, Muller R, Zhou H, et al. Preserved three-dimensional cancellous bone structure in mild primary hyperparathyroidism. *Bone*. 2007;41:19–24.
45. Roschger P, Dempster DW, Zhou H, et al. New observations on bone quality in mild primary hyperparathyroidism as determined by quantitative backscattered electron imaging. *J Bone Miner Res*. 2007;22:717–723.
46. Zoehrer R, Dempster DW, Bilezikian JP, et al. Bone quality determined by Fourier transform infrared imaging analysis in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2008;93:3484–3489.
47. Silva BC, Boutroy S, Zhang C, et al. Trabecular bone score (TBS)--a novel method to evaluate bone microarchitectural texture in patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2013;98:1963–1970.
48. Stein EM, Silva BC, Boutroy S, et al. Primary hyperparathyroidism is associated with abnormal cortical and trabecular microstructure and reduced bone stiffness in postmenopausal women. *J Bone Miner Res*. 2012;28:1029–1040.
49. Khosla S, Melton 3rd LJ, Wermers RA, et al. Primary hyperparathyroidism and the risk of fracture: a population-based study. *J Bone Miner Res*. 1999;14:1700–1707.
50. Vignali E, Viccica G, Diacinti D, et al. Morphometric vertebral fractures in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2009;94:2306–2312.
51. Ejlsmark-Svensson H, Bislev LS, Lajlev S, et al. Prevalence and risk of vertebral fractures in primary hyperparathyroidism: a nested case-control study. *J Bone Miner Res*. 2018;33:1657–1664.
52. Larsson K, Ljunghall S, Krusemo UB, et al. The risk of hip fractures in patients with primary hyperparathyroidism: a population-based cohort study with a follow-up of 19 years. *J Intern Med*. 1993;234:585–593.
53. Kaplan RA, Haussler MR, Deftos LJ, et al. The role of 1 alpha, 25-dihydroxyvitamin D in the mediation of intestinal hyperabsorption of calcium in primary hyperparathyroidism and absorptive hypercalciuria. *J Clin Invest*. 1977;59:756–760.
54. Cipriani C, Biamonte F, Costa AG, et al. Prevalence of kidney stones and vertebral fractures in primary hyperparathyroidism using imaging technology. *J Clin Endocrinol Metab*. 2015;100:1309–1315.
55. Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3561–3569.
56. Tay YD, Liu M, Bandeira L, et al. Occult urolithiasis in asymptomatic primary hyperparathyroidism. *Endocr Res*. 2018;43:106–115.
57. Albright F, Reifenstein EC. The Parathyroid Glands and Metabolic Bone Disease. Baltimore: Williams and Wilkins;1949.
58. Dent CE, Hartland BV, Hicks J, et al. Calcium intake in patients with primary hyperparathyroidism. *Lancet*. 1961;2:336–338.
59. Silverberg SJ, Shane E, Jacobs TP, et al. Nephrolithiasis and bone involvement in primary hyperparathyroidism. *Am J Med*. 1990;89:327–334.
60. Pak CY, Nicar MJ, Peterson R, et al. A lack of unique pathophysiologic background for nephrolithiasis of primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1981;53:536–542.
61. Cassibba S, Pellegrino M, Gianotti L, et al. Silent renal stones in primary hyperparathyroidism: prevalence and clinical features. *Endocr Pract*. 2014;20:1137–1142.
62. Rejnmark L, Vestergaard P, Mosekilde L. Nephrolithiasis and renal calcifications in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2011;96:2377–2385.
63. Saponaro F, Marcocci C, Apicella M, et al. Hypomagnesuria is associated with nephrolithiasis in patients with asymptomatic primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2020;105:e2789–e2795.
64. Walker MD, Nickolas T, Kepley A, et al. Predictors of renal function in primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2014;99:1885–1892.
65. Rao DS, Wilson RJ, Kleerekoper M, et al. Lack of biochemical progression or continuation of accelerated bone loss in mild asymptomatic primary hyperparathyroidism: evidence for biphasic disease course. *J Clin Endocrinol Metab*. 1988;67:1294–1298.
66. Walker MD, Silverberg SJ. Cardiovascular aspects of primary hyperparathyroidism. *J Endocrinol Invest*. 2008;31:925–931.
67. Hedback G, Oden A, Tisell LE. The influence of surgery on the risk of death in patients with primary hyperparathyroidism. *World J Surg*. 1991;15:399–405. discussion 6–7.
68. Wermers RA, Khosla S, Atkinson EJ, et al. Survival after the diagnosis of hyperparathyroidism: a population-based study. *Am J Med*. 1998;104:115–122.
69. Nilsson IL, Yin L, Lundgren E, et al. Clinical presentation of primary hyperparathyroidism in Europe--nationwide cohort analysis on mortality from nonmalignant causes. *J Bone Miner Res*. 2002;17(suppl 2):N68–N74.
70. Vestergaard P, Mollerup CL, Frokjaer VG, et al. Cardiovascular events before and after surgery for primary hyperparathyroidism. *World J Surg*. 2003;27:216–222.
71. Nilsson IL, Aberg J, Rastad J, et al. Maintained normalization of cardiovascular dysfunction 5 years after parathyroidectomy in primary hyperparathyroidism. *Surgery*. 2005;137:632–638.
72. Iwata S, Walker MD, Di Tullio MR, et al. Aortic valve calcification in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2012;97:132–137.
73. Walker MD, Rundek T, Homma S, et al. Effect of parathyroidectomy on subclinical cardiovascular disease in mild primary hyperparathyroidism. *Eur J Endocrinol*. 2012;167:277–285.
74. McMahon DJ, Carrelli A, Palmeri N, et al. Effect of parathyroidectomy upon left ventricular mass in primary hyperparathyroidism: a meta-analysis. *J Clin Endocrinol Metab*. 2015;100:4399–4407.
75. Vicale CT. Diagnostic features of muscular syndrome resulting from hyperparathyroidism, osteomalacia owing to renal tubular acidosis and perhaps related disorders of calcium metabolism. *Trans Am Neurol Assoc*. 1949;74:143–147.
76. Rollinson RD, Gilligan BS. Primary hyperparathyroidism presenting as a proximal myopathy. *Aust N Z J Med*. 1977;7:420–421.
77. Joborn C, Hetta J, Johansson H, et al. Psychiatric morbidity in primary hyperparathyroidism. *World J Surg*. 1988;12:476–481.
78. Walker MD, Silverberg SJ. Parathyroidectomy in asymptomatic primary hyperparathyroidism: improves “bones” but not “psychic moans”. *J Clin Endocrinol Metab*. 2007;92:1613–1615.
79. Bollerslev J, Jansson S, Mollerup CL, et al. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. *J Clin Endocrinol Metab*. 2007;92:1687–1692.

80. Talpos GB, Bone 3rd HG, Kleerekoper M, et al. Randomized trial of parathyroidectomy in mild asymptomatic primary hyperparathyroidism: patient description and effects on the SF-36 health survey. *Surgery*. 2000;128:1013–1020. discussion 20–21.
81. Ambrogini E, Cetani F, Cianferotti L, et al. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. *J Clin Endocrinol Metab*. 2007;92:3114–3121.
82. Walker MD, McMahon DJ, Inabnet WB, et al. Neuropsychological features in primary hyperparathyroidism: a prospective study. *J Clin Endocrinol Metab*. 2009;94:1951–1958.
83. Liu M, Sum M, Cong E, et al. Cognition and cerebrovascular function in primary hyperparathyroidism before and after parathyroidectomy. *J Endocrinol Invest*. 2020;43:369–379.
84. Gazes Y, Liu M, Sum M, et al. Functional magnetic resonance imaging in primary hyperparathyroidism. *Eur J Endocrinol*. 2020;183:21–30.
85. Brandi ML, Gagel RF, Angeli A, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab*. 2001;86:5658–5671.
86. Marx SJ, Fraser D, Rapoport A. Familial hypocalciuric hypercalcemia. Mild expression of the gene in heterozygotes and severe expression in homozygotes. *Am J Med*. 1985;78:15–22.
87. Abood A, Vestergaard P. Pregnancy outcomes in women with primary hyperparathyroidism. *Eur J Endocrinol*. 2014;171:69–76.
88. Bayat-Mokhtari F, Palmieri GM, Moinuddin M, et al. Parathyroid storm. *Arch Intern Med*. 1980;140:1092–1095.
89. Marx SJ, Simmons WF, Agarwal SK, et al. Hyperparathyroidism in hereditary syndromes: special expressions and special managements. *J Bone Miner Res*. 2002;17:N37–N43.
90. Wassif WS, Moniz CF, Friedman E, et al. Familial isolated hyperparathyroidism: a distinct genetic entity with an increased risk of parathyroid cancer. *J Clin Endocrinol Metab*. 1993;77:1485–1489.
91. Agha A, Carpenter R, Bhattacharya S, et al. Parathyroid carcinoma in multiple endocrine neoplasia type 1 (MEN1) syndrome: two case reports of an unrecognised entity. *J Endocrinol Invest*. 2007;30:145–149.
92. Haven CJ, van Puijenbroek M, Tan MH, et al. Identification of MEN1 and HRPT2 somatic mutations in paraffin-embedded (sporadic) parathyroid carcinomas. *Clin Endocrinol*. 2007;67:370–376.
93. Jenkins PJ, Satta MA, Simmgen M, et al. Metastatic parathyroid carcinoma in the MEN2A syndrome. *Clin Endocrinol*. 1997;47:747–751.
94. Dotzenrath C, Teh BT, Farnebo F, et al. Allelic loss of the retinoblastoma tumor suppressor gene: a marker for aggressive parathyroid tumors? *J Clin Endocrinol Metab*. 1996;81:3194–3196.
95. Shattuck TM, Valimaki S, Obara T, et al. Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma. *N Engl J Med*. 2003;349:1722–1729.
96. Marcocci C, Cetani F, Rubin MR, et al. Parathyroid carcinoma. *J Bone Miner Res*. 2008;23:1869–1880.
97. Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*. 2008;93:3462–3470.
98. Rao DS, Phillips ER, Divine GW, et al. Randomized controlled clinical trial of surgery versus no surgery in patients with mild asymptomatic primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2004;89:5415–5422.
99. Lundstam K, Heck A, Mollerup C, et al. Effects of parathyroidectomy versus observation on the development of vertebral fractures in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2015;100:1359–1367.
100. Cheung K, Wang TS, Farrokhyar F, et al. A meta-analysis of preoperative localization techniques for patients with primary hyperparathyroidism. *Ann Surg Oncol*. 2012;19:577–583.
101. Yeh R, Tay YD, Tabacco G, et al. Diagnostic performance of 4D CT and sestamibi SPECT/CT in localizing parathyroid adenomas in primary hyperparathyroidism. *Radiology*. 2019;291:469–476.
102. Clark OH. How should patients with primary hyperparathyroidism be treated? *J Clin Endocrinol Metab*. 2003;88:3011–3014.
103. Jinih M, O'Connell E, O'Leary DP, et al. Focused versus bilateral parathyroid exploration for primary hyperparathyroidism: a systematic review and meta-analysis. *Ann Surg Oncol*. 2017;24:1924–1934.
104. Miccoli P, Berti P, Materazzi G, et al. Endoscopic bilateral neck exploration versus quick intraoperative parathormone assay (qPTHa) during endoscopic parathyroidectomy: a prospective randomized trial. *Surg Endosc*. 2008;22:398–400.
105. Kulak CA, Bandeira C, Voss D, et al. Marked improvement in bone mass after parathyroidectomy in osteitis fibrosa cystica. *J Clin Endocrinol Metab*. 1998;83:732–735.
106. Calvo MS, Kumar R, Heath H. Persistently elevated parathyroid hormone secretion and action in young women after four weeks of ingesting high phosphorus, low calcium diets. *J Clin Endocrinol Metab*. 1990;70:1334–1440.
107. Locker FG, Silverberg SJ, Bilezikian JP. Optimal dietary calcium intake in primary hyperparathyroidism. *Am J Med*. 1997;102:543–550.
108. Rolighed L, Rejnmark L, Sikjaer T, et al. Vitamin D treatment in primary hyperparathyroidism: a randomized placebo controlled trial. *J Clin Endocrinol Metab*. 2014;99:1072–1080.
109. Broadus AE, Magee JS, Mallette LE, et al. A detailed evaluation of oral phosphate therapy in selected patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 1983;56:953. 561.
110. Grey AB, Stapleton JP, Evans MC, et al. Effect of hormone replacement therapy on bone mineral density in postmenopausal women with mild primary hyperparathyroidism. A randomized, controlled trial. *Ann Intern Med*. 1996;125:360–368.
111. Rubin MR, Lee KH, McMahon DJ, et al. Raloxifene lowers serum calcium and markers of bone turnover in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2003;88:1174–1178.
112. Shane E, Baquiran DC, Bilezikian JP. Effects of dichloromethylene diphosphonate on serum and urinary calcium in primary hyperparathyroidism. *Ann Intern Med*. 1981;95:23–27.
113. Eller-Vainicher C, Palmieri S, Cairoli E, et al. Protective effect of denosumab on bone in older women with primary hyperparathyroidism. *J Am Geriatr Soc*. 2018;66:518–524.
114. Peacock M, Bolognese MA, Borofsky M, et al. Cinacalcet treatment of primary hyperparathyroidism: biochemical and bone densitometric outcomes in a five-year study. *J Clin Endocrinol Metab*. 2009;94:4860–4867.
115. Silverberg SJ, Rubin MR, Faiman C, et al. Cinacalcet hydrochloride reduces the serum calcium concentration in inoperable parathyroid carcinoma. *J Clin Endocrinol Metab*. 2007;92:3803–3808.
116. Tsvetov G, Hirsch D, Shimon I, et al. Thiazide treatment in primary hyperparathyroidism—a new indication for an old medication? *J Clin Endocrinol Metab*. 2016;102:1270–1276.
117. Bradwell AR, Harvey TC. Control of hypercalcaemia of parathyroid carcinoma by immunisation. *Lancet*. 1999;353:370–373.
118. Betea D, Bradwell AR, Harvey TC, et al. Hormonal and biochemical normalization and tumor shrinkage induced by anti-parathyroid hormone immunotherapy in a patient with metastatic parathyroid carcinoma. *J Clin Endocrinol Metab*. 2004;89:3413–3420.
119. Vellanki P, Lange K, Elaraj D, et al. Denosumab for management of parathyroid carcinoma-mediated hypercalcemia. *J Clin Endocrinol Metab*. 2014;99:387–390.

55

Malignancy-Associated Hypercalcemia

Diana Athonvarangkul and John J. Wysolmerski

OUTLINE

Introduction and History, 942
Normal Physiology of Calcium Metabolism, 942
    Parathyroid Hormone and the Calcium-Sensing Receptor, 943
    Vitamin D, 943
    Parathyroid Hormone–Related Protein, 943
Malignancy-Associated Hypercalcemia, 944
    Humoral Hypercalcemia of Malignancy, 944
    Local Osteolytic Hypercalcemia, 946
    Overproduction of 1,25-Dihydroxyvitamin D, 947
    Ectopic Production of Parathyroid Hormone, 948
Treatment, 948
    Hydration, 948
    Inhibition of Bone Resorption, 948
    Glucocorticoids, 950
    Dialysis, 950
    Other treatments, 950

INTRODUCTION AND HISTORY

Hypercalcemia is a frequent complication of cancer and has been estimated to occur in up to 20% to 30% of patients during the course of their malignant disease.[1] The incidence of hypercalcemia depends on the tumor type, with higher rates in patients with nonsmall cell lung cancer, breast cancer, and multiple myeloma, and lower rates in squamous head and neck cancers, urothelial carcinomas, and ovarian cancers. However, hypercalcemia has been reported in almost all tumor types.[2] Rates of hypercalcemia may be decreasing over time owing to the overall better therapy for cancers of all types, as well as the prophylactic use of bisphosphonates or the receptor activator of nuclear factor kappa B ligand (RANKL) inhibitor denosumab to prevent skeletal-related events, including hypercalcemia, in patients with bone metastases. For instance, recent retrospective studies using national electronic medical record databases have reported an incidence of malignancy-associated hypercalcemia (MAHC) between 2% and 3% in US cancer patients.[3,4] MAHC generally predicts a poor prognosis, and this has not changed substantially over time. Older studies had estimated that patients with this disorder have a 30-day mortality rate of 50%.[2] More recent series have suggested that hypercalcemia is associated with a median survival of 25 to 52 days, although patients with hematologic malignancies or breast cancer appeared to survive longer than those with other solid tumor types.[5-7] These data demonstrate that hypercalcemia remains a late-stage complication of most tumors.

MAHC was initially recognized in the 1920s, soon after the development of methods to measure blood calcium levels.[8] In 1936, the first series of cases was published on patients with MAHC.[9] The reported patients suffered from myeloma and breast cancer with skeletal involvement, and their hypercalcemia was attributed to skeletal destruction caused by local tumor invasion. In 1941, Fuller Albright suggested an alternative mechanism: that a circulating humoral factor might cause some cases of MAHC.[10] Studies published in the 1950s and 1960s supported this hypothesis by describing patients with hypercalcemia who lacked bony metastases.[11,12] These patients had squamous, renal, and genitourinary cancers, as opposed to breast cancer or myeloma. Lafferty concluded that these tumors must secrete parathyroid hormone (PTH) and named this process *pseudohyperparathyroidism*.[13] At that point, it was thought that tumors could cause hypercalcemia through direct skeletal involvement or by acting on the skeleton in a humoral fashion. We now refer to these conditions as two subtypes of MAHC: the first is local osteolytic hypercalcemia (LOH), and the second is humoral hypercalcemia of malignancy (HHM).[14,15]

In the 1980s, investigators made great strides in understanding the pathophysiology of both LOC and HHM. First, it was appreciated that tumor cells could make cytokines that stimulated the local production and activity of bone-resorbing osteoclasts within the bone microenvironment surrounding tumor cells. Second, investigations into HHM led to the discovery of a new hormone, PTH-related peptide (PTHrP), which is evolutionarily related to PTH and uses the same receptor. As a result, when tumors secrete PTHrP into the circulation, it acts like PTH to increase osteoclast differentiation and activity throughout the skeleton. It has also become clear that two additional types of MAHC exist: 1,25-dihydroxyvitamin D–induced hypercalcemia in patients with lymphomas, and true ectopic production of PTH. These will be discussed further later in this chapter.

KEY POINTS

- Hypercalcemia is a common complication of malignancy that occurs with all known tumor types and usually portends a poor prognosis.

NORMAL PHYSIOLOGY OF CALCIUM METABOLISM

Calcium can be found in the body in two predominant forms. First, 99% of body calcium is found in bone hydroxyapatite, which provides structural integrity to the skeleton. Second, approximately 1% of body calcium is found as soluble calcium ions in the extracellular fluid (ECF) and cytosol, which contributes to a series of diverse cellular functions.[16] Skeletal and ECF calcium are exchangeable, allowing bone to provide calcium to the ECF when calcium concentrations are low and also to store excess calcium when needed. Ionized, or free, calcium represents approximately 50% of the circulating pool, and the remaining fraction is bound to serum proteins. The ionized fraction is the physiologically important one, but it is difficult to measure, as it requires anaerobic conditions for collection and handling. Therefore, total serum calcium is most commonly used as an indirect measurement of the ionized fraction, although some medical centers now routinely measure the ionized fraction. If measuring total calcium only, correction for the patient's albumin should be made as needed.

The parathyroid glands, kidneys, skeleton, and gut are the organs involved in calcium homeostasis and their interactions are coordinated through the actions of calcium, itself, PTH, vitamin D, and PTHrP. These interactions will be described in the following sections.

Parathyroid Hormone and the Calcium-Sensing Receptor

PTH is produced from the four parathyroid glands located near the upper and lower poles of the thyroid gland. It is an 84–amino acid protein with a very short half-life (minutes) in the circulation and is degraded by the liver and kidney. PTH secretion is regulated by extracellular ionized calcium concentrations, and small changes in ionized calcium concentrations are detected by parathyroid cells through the actions of a G-protein–coupled-receptor known as the calcium-sensing receptor (CaSR).[17-23] Calcium binds to the CaSR and activates a downstream signaling cascade that suppresses PTH synthesis and subsequent secretion.[19,20,24] Persistently elevated calcium levels can result in degradation of PTH within parathyroid cells.[25] The CaSR is also expressed in the kidney, where it regulates calcium handling by the renal tubules. Hypercalcemia activates the CaSR, resulting in suppression of renal calcium reabsorption and excretion of excess calcium in the urine.[26]

The actions of PTH are mediated through a different G-protein–coupled-receptor known as the type 1 PTH/PTHrP receptor (PTH1R), because it responds to both PTH and PTHrP. Activation of the receptor by PTH initiates several signaling cascades, but the best documented has been the adenylyl cyclase pathway. In the kidney, PTH acts on proximal tubular cells to inhibit reabsorption of phosphate through removal and degradation of the sodium-phosphate cotransporters 2a and 2c from the luminal membranes.[27] It also stimulates the activity of renal 25-hydroxyvitamin D 1-alpha-hydroxylase (CYP27B1) in proximal tubular cells and inhibits 24-hydroxylase (CYP24A1), which results in an increase in the biologically active 1,25-dihydroxyvitamin D.[28] PTH increases the reabsorption of calcium from the glomerular filtrate in the kidney by increasing the positive charges on the luminal side of the tubule in the cortical thick ascending loop of Henle and by regulating the trafficking of calcium channels into the apical membrane of cells in the distal tubule.[29] In the skeleton, PTH activates bone turnover and releases stored calcium. It does this by stimulating transport of calcium across bone lining cells from a pool of calcium at the surface of bone. It also stimulates both bone resorption and bone formation. PTH increases the number and activity of bone-forming osteoblasts and bone-resorbing osteoclasts, increases the size of the osteoblast precursor pool,[30] increases the bone-forming activity of mature osteoblasts, and promotes the release of colony-stimulating factor 1 and RANKL, which stimulate the formation of new osteoclasts and activate mature osteoclasts.[31] PTH inhibits osteoblast production of osteoprotegerin, which is a soluble decoy receptor for RANKL that inhibits osteoclast development. The ultimate effects of PTH on bone turnover depend on the pattern of PTH exposure. Continued long-term elevations in PTH stimulate net bone resorption, while intermittent PTH stimulation can result in net bone formation. See Chapter 47 for further discussion on PTH.

Vitamin D

In the skin, keratinocytes absorb ultraviolet light and convert 7-dehydrocholesterol to vitamin D3.[32] This biologically inert form of vitamin D is then hydroxylated in the liver by vitamin D 25-hydroxylase (CYP2R1) to 25-hydroxyvitamin D3.[33] The second hydroxylation occurs in the kidney by CYP27B1 to convert 25-hydroxyvitamin D3 to 1,25-dihydroxyvitamin D3.[33] This enzymatic process in the kidney is stimulated by PTH and phosphate and inhibited by hypercalcemia and FGF23. CYP27B1 is also expressed in immune cells, especially activated macrophages. Therefore, 1,25-dihydroxyvitamin D3 can be produced by macrophages, classically in granulomas.

Vitamin D is a steroid hormone that interacts with nuclear receptors to regulate gene expression. When 1,25-dihydroxyvitamin D is present, the vitamin D receptor (VDR) heterodimerizes with the retinoid acid X receptor and binds to specific DNA sequences within the target gene.[34,35] 1,25-dihydroxyvitamin D stimulates calcium absorption from the diet by upregulating expression of the components of the transport mechanism by which calcium is transferred across enterocytes from the lumen into the ECF.[35] At pathologic levels, 1,25 dihydroxyvitamin D can increase bone resorption by binding to the VDR on osteoblasts and stimulating osteoblastic production of RANKL, leading to osteoclast activation and bone resorption.[36,37] See Chapter 50 for a detailed discussion of vitamin D chemistry and biology.

Parathyroid Hormone–Related Protein

PTHrP is a "cousin" of PTH; both peptides are derived from related genes that, together with *TIP-39*, define a small gene family. PTHrP was initially identified as a cause of MAHC. Much research has now shown that PTHrP is a widely expressed cytokine that is involved in a variety of physiologic functions. PTHrP is discussed at greater length in Chapter 48. What follows here is a short discussion enabling the reader to appreciate its pathophysiologic role in HHM.

In humans, PTHrP is encoded by a single-copy gene containing eight exons and at least three promoters located on the short arm of chromosome 12.[38-40] Many different hormones and growth factors regulate the transcription and/or stability of PTHrP mRNA. As with PTH, the CaSR has been found to regulate PTHrP gene expression in many cells.[20,41]

The amino terminus of PTHrP is remarkably similar to that of PTH. The two peptides share eight of the first 13 amino acids and a high degree of predicted secondary structure over the next 21 amino acids. These common sequences allow both proteins to bind and activate PTH1R. The vast majority of studies *in vitro* suggest that this receptor binds PTHrP and PTH with equal affinity, and that both activate the receptor in an identical manner. However, the human PTH1R may respond slightly differently to PTH and PTHrP. There appear to be physical differences in the binding of the two peptides to different conformational states of the receptor, so that the duration of cyclic adenosine monophosphate (cAMP) production is shorter for PTHrP(1–36) than for PTH(1–34).[42] Nevertheless, when PTHrP enters the systemic circulation, it can mimic the effects of PTH and cause hypercalcemia in HHM.[15]

PTHrP has been found in at least some cells of almost all organs, and a variety of functions have been ascribed to PTHrP. These are discussed at greater length in Chapter 48. Pertinent to its role in causing hypercalcemia in cancer, it is worth highlighting PTHrP's function during lactation. Secreting milk increases maternal calcium requirements, and a significant portion of the calcium used for milk production comes from the mother's skeleton. As a result, lactating women lose 5% to 8% of their bone mass over the first 6 months of full-time nursing, and lactating rodents can lose 25% to 30% of their skeletal mass in as little as 12 to 14 days.[43] Bone loss during this time is mediated, in part, by the secretion of PTHrP from breast epithelial cells into the systemic circulation. This is the only time that PTHrP circulates in normal individuals in appreciable amounts, and its physiologic function is to increase bone resorption to liberate skeletal calcium stores for milk production. This PTH-like function during reproductive cycles is an important evolutionary pressure that helped to maintain PTHrP's ability to share the same receptor with PTH, even though, in every other aspect, the two peptides function much differently. Given that HHM exists because PTHrP and PTH share the same receptor, ultimately this evolutionary pressure also helps to explain why hypercalcemia occurs in many malignancies.[44]

KEY POINTS

- Calcium regulation involves intricate interactions between ionized calcium levels, parathyroid hormone (PTH), vitamin D, and bone cells that regulate calcium release from bones. PTH-related peptide (PTHrP) is a local growth factor that shares homology with PTH and that can activate the same PTH/PTHrP receptors that PTH uses to regulate calcium metabolism. As a result, excess production of PTHrP by cancer cells can activate PTH/PTHrP receptors in bone and kidney to cause hypercalcemia.

MALIGNANCY-ASSOCIATED HYPERCALCEMIA

Historically, MAHC has been subdivided into two main categories. The first occurs in patients with few or no bone metastases but with diffuse upregulation of bone resorption. These tumors produce a humoral factor(s) (usually PTHrP) that circulates and causes increased osteoclast numbers and activity. This type of MAHC has been referred to as *HHM*. The second main category includes tumors with osteolytic bone metastases referred to as *LOH*. In these tumors, the metastatic cells within the skeleton produce local paracrine factors that increase osteoclastic bone resorption around the tumor deposits, but not systemically. These tumors require an extensive skeletal tumor burden to cause hypercalcemia. Two additional, although less common, subtypes of MAHC have been described: 1,25-dihydroxyvitamin D induced hypercalcemia in patients with lymphoma and some solid tumors, and ectopic hyperparathyroidism produced by nonparathyroid tumors.

Humoral Hypercalcemia of Malignancy

HHM has been considered to be the more common type of MAHC, representing 75% to 80% of cases in consecutive series of patients with MAHC.[14,15] However, two more recent retrospective studies examining elevated PTHrP levels have suggested that HHM represents only 32% to 38% of MAHC.[45,46] In contrast, other recent studies of MAHC suggested that 57% to 74% of cases had HHM based on the absence of bone metastases.[7,47] It is not clear whether these different estimates reflect different patient populations or differences in patient identification and/or the sensitivity of PTHrP measurements. Classically, HHM occurs most commonly in squamous, renal, and genitourinary malignancies, some breast cancers, and human T-cell leukemia virus 1–associated adult T-cell leukemia/lymphomas.[48] However, almost all tumor types have been reported to occasionally cause HHM. It is difficult to determine the exact incidence of HHM in patients with different types of cancer, as the frequency generally increases as cancer progresses. For instance, approximately 15% of patients with renal cell carcinoma were shown to have hypercalcemia before treatment.[49] In contrast, up to 50% of patients with squamous cell tumors of the head and neck followed for several years eventually developed HHM.[50] In fact, hypercalcemia is typically a complication of advanced cancer, and the prognosis is poor in patients with HHM, with reports suggesting a median survival of 1 to 3 months.[51,52]

HHM is most commonly caused by a tumor, distant from the skeleton, releasing PTHrP into the systemic circulation. PTHrP is often a normal product of the cells of origin of the tumor, but in the setting of malignancy PTHrP production is upregulated by tumor cells, and the tumors evade the normal barriers that usually keep PTHrP within the local tissue environment. As a result, PTHrP enters the systemic circulation and gains access to the PTHR1 pool in the skeleton and kidney that is normally stimulated by PTH. This is associated with a characteristic pattern of biochemical abnormalities first defined by Stewart and colleagues in 1980.[15] Patients typically demonstrate elevated serum calcium levels associated with suppressed PTH levels and low 1,25-dihydroxyvitamin D levels. As a result, calcium clearance in the kidneys is increased, and gut calcium absorption is decreased, contributing to a profound negative calcium balance. Serum phosphorus levels are low-normal or low, with a decreased renal tubular reabsorption of phosphate. Nephrogenous cAMP (NcAMP) excretion rates are elevated, reflecting activation of the renal proximal tubular PTH receptor despite the reduction in circulating PTH levels. The elevated NcAMP levels in HHM formed the basis for biochemical assays that resulted in the purification of PTHrP from tumors and the cloning of the *PTHrP* gene in the 1980s.[15]

Hypercalcemia in HHM is caused by the release of calcium from the skeleton at a rate that exceeds the kidney's ability to excrete the excess load (Fig. 55.1). Patients initially develop hypercalciuria but eventually develop dehydration owing to an osmotic diuresis combined with the effects of calcium to activate the CaSR in the distal tubule and collecting duct of the kidney, which, in turn, inhibits aquaporin trafficking and induces a form of nephrogenic diabetes insipidus. Once patients become dehydrated, serum calcium levels typically rise rapidly, and patients develop symptoms of lethargy, anorexia, and confusion from the elevated ionized calcium levels that only worsen the dehydration and accelerate the increase in serum calcium levels.

Although PTH and PTHrP act on the same receptor, it was originally reported that several biochemical aspects of HHM differed from changes typical of primary hyperparathyroidism. First, as noted earlier, renal calcium clearance was higher in HHM than that observed in hyperparathyroidism. Second, although PTHrP levels are elevated in HHM, 1,25-dihydroxyvitamin D levels were described to be low. By contrast, the elevations in PTH in primary hyperparathyroidism were accompanied by increased 1,25-dihydroxyvitamin D levels. Finally, bone histomorphometry studies in patients with HHM demonstrated very elevated rates of bone resorption but suppressed bone formation rates.[53] Typically, in primary hyperparathyroidism, rates of both bone formation and bone resorption are elevated. Similar differences in bone formation rates in HHM as compared with hyperparathyroidism are noted when biochemical markers of bone turnover such as total deoxypyridinoline, pyridinoline cross-linked carboxyterminal telopeptide of type 1 collagen, serum osteocalcin, and serum bone type alkaline phosphatase were measured.[54] It is still not entirely clear why elevations in PTHrP produce different effects on the kidney and bone cells if both peptides act on the same receptor. In addition, more recent studies have suggested that some patients with HHM have elevated levels of both PTHrP and 1,25-dihydroxyvitamin D (discussed later). Studies by Horwitz and colleagues revealed that infusions of PTHrP in normal human subjects are less effective at stimulating 1,25-dihydroxyvitamin D production than are infusions of PTH.[55-57] As for the uncoupling of bone formation from bone resorption, tumor production of IL-6 has been implicated in human squamous carcinoma cells that also secrete PTHrP, and this cytokine may accelerate osteoclastic bone resorption and suppress osteoblastic function.[58] Finally, although PTHrP and PTH appear to act equivalently in many experimental systems, there are differences in the way they bind to the PTH1R, providing a potential mechanism through which the two proteins may have tissue-specific differences in signaling that would explain divergent effects on 1,25-dihydroxyvitamin D production and the stimulation of bone formation by osteoblasts.[15,42,59,60]

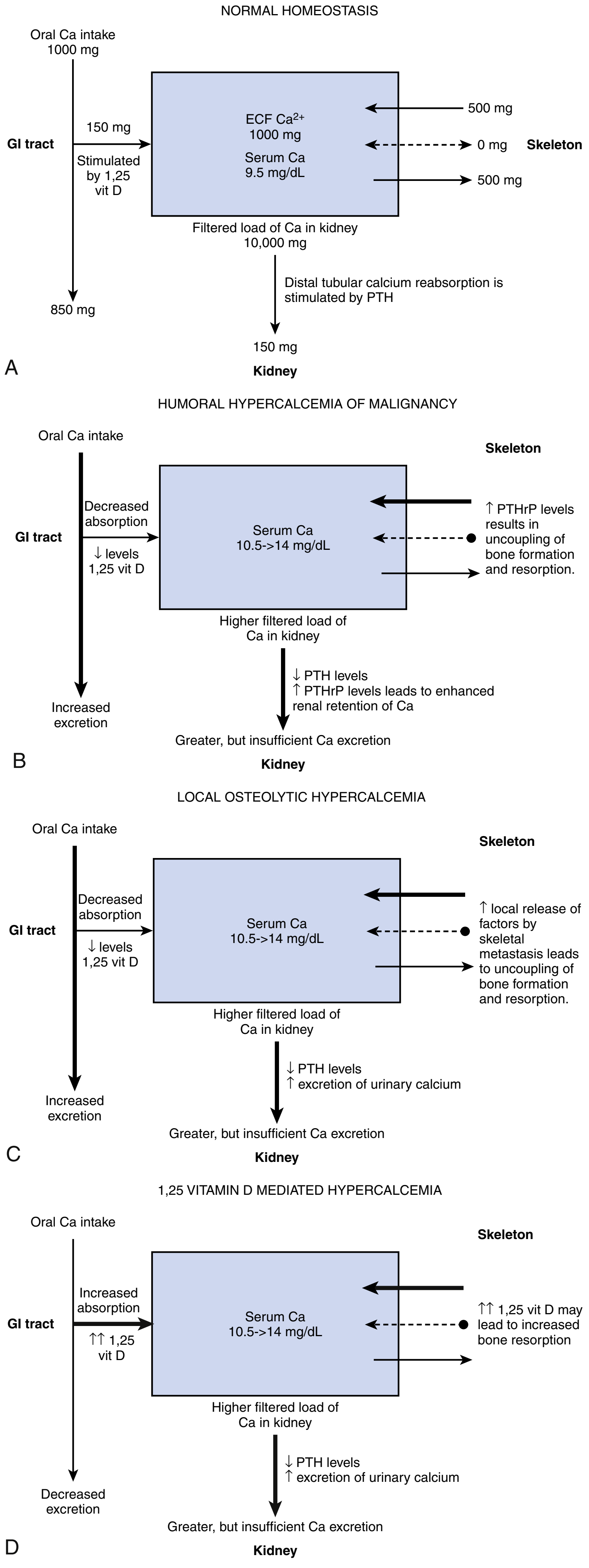

NORMAL HOMEOSTASIS
Oral Ca intake
1000 mg
GI tract
150 mg
Stimulated by 1,25 vit D
850 mg
ECF Ca2+
1000 mg
Serum Ca
9.5 mg/dL
500 mg
0 mg
Skeleton
500 mg
Filtered load of Ca in kidney
10,000 mg
Distal tubular calcium reabsorption is stimulated by PTH
150 mg
Kidney
A
HUMORAL HYPERCALCEMIA OF MALIGNANCY
Oral Ca intake
GI tract
Decreased absorption
↓ levels 1,25 vit D
Increased excretion
Serum Ca
10.5->14 mg/dL
Skeleton
↑ PTHrP levels results in uncoupling of bone formation and resorption.
Higher filtered load of Ca in kidney
↓ PTH levels
↑ PTHrP levels leads to enhanced renal retention of Ca
Greater, but insufficient Ca excretion
Kidney
B
LOCAL OSTEOLYTIC HYPERCALCEMIA
Oral Ca intake
GI tract
Decreased absorption
↓ levels 1,25 vit D
Increased excretion
Serum Ca
10.5->14 mg/dL
Skeleton
↑ local release of factors by skeletal metastasis leads to uncoupling of bone formation and resorption.
Higher filtered load of Ca in kidney
↓ PTH levels
↑ excretion of urinary calcium
Greater, but insufficient Ca excretion
Kidney
C
1,25 VITAMIN D MEDIATED HYPERCALCEMIA
Oral Ca intake
GI tract
Increased absorption
↑↑ 1,25 vit D
Decreased excretion
Serum Ca
10.5->14 mg/dL
Skeleton
↑↑ 1,25 vit D may lead to increased bone resorption
Higher filtered load of Ca in kidney
↓ PTH levels
↑ excretion of urinary calcium
Greater, but insufficient Ca excretion
Kidney
D

Fig. 55.1 Comparison of normal calcium fluxes with malignancy-associated hypercalcemia. **A,** The central box represents extracellular fluid (*ECF*), which contains approximately 1000 mg of calcium (*Ca*). This compartment interfaces with the gastrointestinal (*GI*) tract, the kidney, and the skeleton. The fluxes into and out of the ECF are measured in mg/d. There is a zero calcium balance under homeostatic conditions. **B,** In humoral hypercalcemia of malignancy, elevated circulating parathyroid hormone (*PTH*)–related protein (*PTHrP*) levels result in an uncoupling of bone resorption from bone formation and a net output of calcium from bone. There is decreased intestinal absorption of calcium secondary to reduced 1,25-dihydroxyvitamin D levels. The kidney filters an increased load of calcium, and PTHrP acts on the kidney to enhance renal retention of calcium. However, this is insufficient to reduce serum calcium concentrations to normal levels.**C,** In local osteolytic hypercalcemia, local release of factors from skeletal metastases leads to the local uncoupling of bone resorption from formation. If there are enough metastases, there is a net output of calcium from bone. There is decreased intestinal absorption of calcium secondary to reduced 1,25-dihydroxyvitamin D levels. The kidney filters an increased load of calcium, but this is insufficient to reduce serum calcium concentrations to normal levels. **D,** In 1,25-dihydroxyvitamin D–mediated hypercalcemia, there is a marked increase in intestinal calcium absorption, and there also may be an increase in bone resorption. The kidney filters an increased load of calcium, but this is insufficient to reduce serum calcium concentrations to normal levels. (**A,** Modified from Stewart AF. Normal physiology of bone and mineral homeostasis. In Andreoli TE, et al, eds. *Andreoli and Carpenter's Cecil essentials of medicine.* Philadelphia: Elsevier Saunders; 2010.)

Local Osteolytic Hypercalcemia

LOH is a second major type of MAHC, comprising approximately 20% to 30% of patients in consecutive series of patients with MAHC.[14,15,61] In patients with LOH, the hypercalcemia is caused by tumor growth within bone, which leads to the stimulation of osteolysis around the metastatic cancer cells with release of calcium into the systemic circulation. To develop hypercalcemia, patients must have extensive skeletal metastases. Classically, the most common tumors presenting with this syndrome include myeloma, breast cancer, and lymphoma.

The biochemical findings in LOH include hypercalcemia with reduced PTH and 1,25-dihydroxyvitamin D levels, similar to HHM. However, in this condition patients have normal serum phosphate concentrations and a normal renal phosphate threshold. In contrast to HHM, PTHrP levels are undetectable, so NcAMP levels are reduced secondary to the suppressed PTH levels. Urinary calcium excretion is increased as a result of an increased filtered load of calcium found in hypercalcemic conditions and a decreased distal tubular calcium reabsorption in the setting of lower PTH levels. Intestinal absorption of calcium is also reduced because of the decreased circulating 1,25-dihydroxyvitamin D levels. As with HHM, this results in a net negative calcium balance.

The pathophysiology of LOH is an extension of the pathophysiology of osteolytic bone metastases.[62,63] In order for tumor cells to survive and grow in bone, they influence the behavior of normal bone cells to remodel the bone microenvironment, often activating bone resorption and causing osteolysis. This leads to the release of factors from the bone matrix that, in turn, stimulate further growth of the tumor cells, causing more osteolysis. This vicious cycle of peritumor osteolysis and tumor growth has been the focus of many studies. In short, it appears that tumor cells produce soluble factors that lead to increased osteoclast recruitment, differentiation, and activity.[64-67] Some tumors also elaborate factors that inhibit osteoblast-mediated bone formation. As a result, the tumors cause progressive bone destruction, allowing them to expand within the skeleton. If the number and size of the osteolytic metastases is great enough, then the release of calcium from the skeleton will exceed the kidney's ability to excrete the filtered load, and patients will develop clinical hypercalcemia, usually associated with renal insufficiency (see Fig. 55.1). As opposed to HHM, bone resorption does not happen throughout the skeleton in LOH, just around the individual tumor deposits. The general pathophysiology of bone metastases has been reviewed in great detail previously. In this setting, we will only briefly review some of the main cytokines that have been implicated in driving the vicious cycle of osteolysis, but the reader is referred to other reviews for a more complete discussion of this topic.[68-70]

One of the tumors that frequently causes extensive bone destruction and hypercalcemia is myeloma. In 1974, Mundy described an osteoclast-activating factor that was produced by myeloma cells.[71] Many subsequent studies have identified a series of cytokines produced by myeloma cells that have the ability to cause osteolysis. These include interleukin (IL)-1α, IL-1β, tumor necrosis factor (TNF)-α, TNF-β (lymphotoxin), transforming growth factor (TGF)-α, macrophage inflammatory protein-1α, IL-3, IL-6, IL-17, and CD-47.[71] Many of these cytokines increase osteoclast differentiation and activity by increasing the production of RANKL by osteoblasts.[72-74] RANKL is the dominant factor produced by osteoblasts to regulate the production and activity of osteoclasts, and this factor acts through a receptor, receptor activator of nuclear factor kappa B (RANK), found on osteoclast progenitors and mature osteoclasts. Recently, direct interactions of myeloma cells with osteocytes and bone marrow stromal cells have been shown to be important for the upregulation of the above cytokines and increased RANKL production or decreased secretion of the RANKL inhibitor, osteoprotegerin (OPG).[75] An important mediator of these direct interactions appears to be activation of Notch signaling pathways. It has also become clear that suppression of bone formation contributes significantly to the pathogenesis of myeloma bone disease. This appears to be related to the secretion of Dickkopf-1, an inhibitor of the Wnt signaling pathway, by myeloma cells. In addition, myeloma bone disease is associated with increased production of sclerostin, a dual Wnt/bone morphogenic protein inhibitor that is made by osteocytes. Inhibition of Wnt signaling suppresses osteoblast activity and bone formation, augmenting the local osteolysis and release of calcium from the bone.[76-78]

Breast cancer is the second tumor type classically associated with LOH. Bone metastases are common in breast cancer, occurring in up to 70% of patients with advanced disease.[62,63] As in myeloma, bone metastases from breast cancers are typically osteolytic, and if the skeletal burden is great enough, a significant level of calcium can be released from the skeleton to overwhelm the body's adaptive mechanisms, resulting in hypercalcemia. It is difficult to ascertain the current incidence of hypercalcemia in patients with breast cancer and bone metastases. Older studies suggested that 10% to 15% of all patients with breast cancer developed hypercalcemia, the majority of whom had bone metastases.[79] More recent studies (usually drug trials) group together hypercalcemia, pathologic fracture, radiation treatment for impending fracture, and bone pain under the term *skeletal-related events* (*SREs*) and do not often report the incidence of hypercalcemia separately. In these trials, the control group of patients with bone

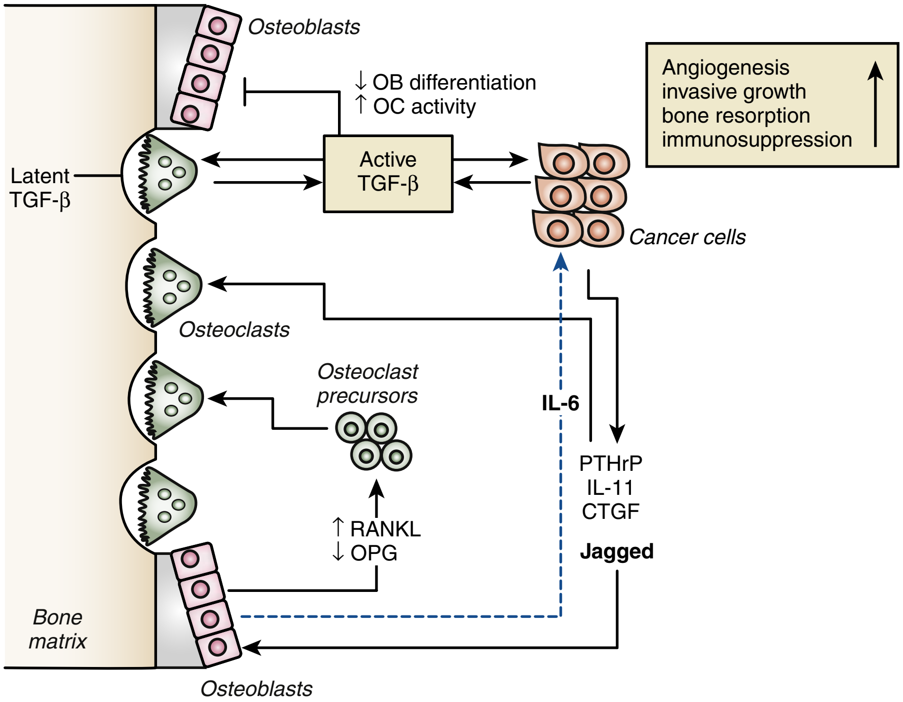


Fig. 55.2 Osteolytic bone metastasis in breast cancer. In skeletal metastases, the tumor cells make factors, including parathyroid hormone–related peptide (*PTHrP*), that upregulate osteoclastic bone resorption. This results in the release of transforming growth factor (*TGF*)-β from the bone matrix, which acts on cancer cells to further stimulate the production of osteolytic factors, such as PTHrP, connective tissue growth factor (*CTGF*), interleukin (IL)-6, and IL-11. These factors increase the receptor activator of nuclear factor kappa B ligand (RANKL)/ osteoprotegerin (*OPG*) expression ratio in bone stromal cells such as osteoblasts, resulting in osteoclastogenesis. Moreover, active TGF-β stimulates Jagged 1 expression in cancer cells, which in turn stimulates Notch signaling in osteoclasts (*OC*) and osteoblasts (*OB*), resulting in increased osteoclastogenesis and production of the cytokine IL-6 by osteoblasts (*blue dotted line*), which stimulates proliferation of tumor cells. TGF-β also inhibits osteoblast differentiation. Thus, once osteolysis is established, a vicious cycle of bone resorption and tumor cell proliferation is fed by the release of growth factors from bone matrix, tumor cells, and bone cells. (Reproduced with permission from Buijs JT, Stayrook KR, Guise TA. The role of TGF-beta in bone metastasis: novel therapeutic perspectives. *Bonekey Rep*. 2012;1:96.)

metastases had an SRE rate of approximately 60%. The use of prophylactic bisphosphonates is associated with an aggregate 15% reduction in SREs, so it is likely that the widespread use of these medications in patients with breast cancer and bone metastases has reduced the incidence of hypercalcemia by at least a similar amount.[80,81]

Like myeloma cells, breast cancer cells have been shown to produce a variety of factors that induce local bone resorption (Fig. 55.2). One of the best studied of these factors is PTHrP. It is well documented that PTHrP is produced by a number of primary breast carcinomas, and that this sometimes leads to classical HHM.[82] However, animal models have suggested that localized PTHrP production by breast tumor cells contributes to the osteolysis associated with skeletal metastases.[83,84] These studies have suggested that, in response to the bone microenvironment, breast cancer cells metastatic to the skeleton produce more PTHrP than cells in the primary tumor.[84,85] TGF-β, released at sites of bone resorption, has been shown to upregulate PTHrP production through a signaling pathway involving the transcription factor Gli2.[86] PTHrP, in turn, increases the production of RANKL and decreases the production of OPG, increasing osteoclast numbers and activity.[85] This sets up a feedforward loop between PTHrP production and bone resorption, contributing to a vicious cycle of osteolysis. In addition, unlike in normal breast cells, CaSR signaling in breast cancer cells stimulates PTHrP production and can synergize with the effects of TGF-β.[87,88] Finally, the mechanical stiffness of bone tissue may also increase PTHrP production.[89] Thus, several aspects of the bone microenvironment conspire to increase PTHrP production in skeletal metastases. However, PTHrP is not the only factor important in this process. Breast cancer cells have been shown to produce other osteolytic factors, such as IL-6, IL-8, IL-11, prostaglandin E2, Jagged 1, macrophage colony-stimulating factor (MCSF), and macrophage-stimulating protein (MSP).[66] MCSF and MSP may act directly to stimulate osteoclast production and activity, while the other factors act with PTHrP to stimulate RANKL production by osteoblasts.[66,90,91] As depicted in Fig. 55.2, the growth of osteolytic metastases depends on a vicious cycle of bone resorption and tumor cell proliferation that is fed by the release of growth factors from the bone matrix, the breast cancer cells, and normal bone cells.[62]

Overproduction of 1,25-Dihydroxyvitamin D

Another less common cause of MAHC is the overproduction of 1,25-dihydroxyvitamin D by lymphomas, both Hodgkin and non-Hodgkin types.[92-96] In this setting, local activation of CYP27B1 activity within or around the tumor results in the overproduction of the active form of vitamin D. In at least one carefully studied patient, the source of the CYB27B1 was the tumor-associated macrophages and not the lymphoma cells themselves.[97] This would suggest that the tumor cells make some paracrine factor that upregulates CYP27B1 expression in macrophages, although it is not clear whether lymphoma cells can also sometimes express CYB27B1. Unlike the kidney, there is no systemic feedback inhibition of the tumor-associated CYB27B1 activity, but 1,25-dihydroxyvitamin D production is responsive to 25-hydroxyvitamin D_3 levels. Hypercalcemia in these patients is largely secondary to the hyperabsorption of calcium in the intestine, although high 1,25-dihydroxyvitamin D levels can also directly cause bone resorption (see Fig. 55.1). Patients typically have normal serum phosphate levels and increased renal calcium excretion. PTH levels are appropriately suppressed in the face of hypercalcemia, and PTHrP concentrations are unmeasurable.[96]

Some solid tumors also cause hypercalcemia associated with elevated levels of 1,25-dihydroxyvitamin D, suppressed PTH levels, and low PTHrP levels. This has most commonly been reported in patients with seminomas or dysgerminomas, but also in a patient with small-cell cervical cancer.[98-101] While the majority of the reports have been single cases, Evans and colleagues examined a series of 12 dysgerminomas and documented increased expression of CYP27B1 in tumor cells and associated macrophages at both the mRNA and protein levels.[98] In addition, some cases of MAHC appear to be related to overproduction of both 1,25-dihydroxyvitamin D and PTHrP in the same patient. The cooccurrence of elevated levels of both hormones in MAHC has been reported in ovarian carcinoma, renal cell carcinoma, seminoma, adult T-cell leukemia/lymphoma, and pancreatic neuroendocrine tumors.[102-107] The authors have also recently seen similar patients with aggressive urothelial carcinoma and breast cancer at Yale New Haven Hospital. It is not clear whether the production of 1,25-dihydroxyvitamin D is occurring from the tumor itself as in lymphomas, or by tumor-associated macrophages or from the kidney. However, in one instructive case report of an ovarian clear cell adenocarcinoma, venous sampling at the time of resection demonstrated that the tumor produced PTHrP but not 1,25-dihydroxyvitamin D.[108] Nonetheless, both PTHrP and 1,25-dihydroxyvitamin D levels were normalized after surgery, suggesting that PTHrP or some other tumor product led to increased 1,25-dihydroxyvitamin D production by the kidney. In fact, despite the original descriptions of low 1,25-dihydroxyvitamin D levels in HHM, the cooccurrence of elevated PTHrP and 1,25-dihydroxyvitamin D levels may not be that unusual. In a recently published retrospective examination of 101 patients with MAHC from Memorial Sloan Kettering Cancer Center, 34 patients had simultaneous elevations of both PTHrP and 1,25-dihydroxyvitamin D. In addition, these patients were significantly more likely to have an incomplete response to antiresorptive therapy than patients with normal 1,25-dihydroxyvitamin D levels.[109] Thus, it is still important to measure 1,25-dihydroxyvitamin D levels in all patients with MAHC, as elevated levels may have therapeutic implications.

Ectopic Production of Parathyroid Hormone

Although HHM was originally thought to be secondary to ectopic PTH production, it is quite rare that MAHC is actually caused by ectopic production of PTH. Approximately 20 cases exist in the literature of ectopic hyperparathyroidism caused by a nonparathyroid tumor. These tumors include neuroendocrine tumors, hepatocellular carcinoma, medullary thyroid cancer, ovarian carcinoma, small-cell lung carcinoma, squamous carcinoma, rhabdoid tumor of the kidney, endometrial adenosquamous carcinoma, and thymoma.[59,110-126] Immunoradiometric two-site assays for intact PTH and PTHrP allow for the determination as to whether PTH or PTHrP is causing hypercalcemia in malignancy. There is no crossreactivity between PTH and PTHrP in these assays.[120] Nussbaum and coworkers described a patient with ovarian carcinoma, in which genetic evaluation of the tumor cells revealed DNA amplification and rearrangement in the upstream regulatory region of the *PTH* gene. This resulted in increased PTH expression.[120] VanHouten and associates described a patient with a high-grade neuroendocrine carcinoma of the pancreas found to have ectopic PTH production. Genetic evaluation of these tumor cells revealed transactivation of the PTH gene promoter associated with hypomethylation of the *PTH* gene locus, indicative of reactivation of *PTH* gene transcription in the tumor cells.[125]

KEY POINTS

- Bone resorption is the key pathophysiologic mechanism by which excess calcium is released into the circulation in malignancy-associated hypercalcemia (MAHC). The four main mechanisms underlying MAHC are: (1) elevation in circulating parathyroid hormone–related protein secreted by tumors outside the skeleton, known as humoral hypercalcemia of malignancy (HHM), (2) local cytokine production by tumor cells located within the bone microenvironment, known as local osteolytic hypercalcemia (LOH), (3) overproduction of 1,25-dihydroxyvitamin D, and (4) ectopic production of parathyroid hormone.

TREATMENT

Hydration

Hypercalcemia is associated with polyuria. In addition, hypercalcemic patients also suffer from reduced oral intake because of anorexia, nausea, and vomiting. These factors lead to dehydration and a reduction in glomerular filtration rate.[127] An impairment in renal function, in turn, limits the ability to excrete the increased calcium load coming from bone and accelerates the development of hypercalcemia. As a result, the first goal of therapy is to correct volume depletion (Fig. 55.3). The initial rate and duration of fluid administration should be determined by clinical signs of dehydration, the duration and severity of hypercalcemia, and underlying medical conditions, especially cardiovascular disease. It has typically been recommended to hydrate patients with 0.9% sodium chloride solution (also known as normal saline), initially with a 1- to 2-liter bolus followed by continuous infusion at 200 to 500 mL/hour intravenously, titrating to a urine output of 100 to 150 mL/hour.[128,129] After initial volume repletion and establishment of adequate urine output, fluids can be continued at greater than 100 mL/hour, but it is important to carefully monitor vital signs and watch for signs of fluid overload, especially in cases with existing comorbidities such as congestive heart failure.[127,130] Although intravenous fluid administration can transiently lower serum calcium concentrations on the order of 1 to 1.5 mg/dL, additional therapies are required to lower calcium further and to maintain the lower calcium levels.

It has been common practice to add loop diuretics, such as furosemide at a dose of 20 to 40 mg intravenously, after rehydration has been achieved to promote calcium excretion through a forced saline diuresis. The concept is to take advantage of the fact that sodium delivery to the distal tubule promotes urinary calcium excretion and helps to accelerate calcium clearance by the kidneys.[1,127] However, LeGrand and colleagues reviewed the literature on this topic and argued that routine use of loop diuretics in the treatment of MAHC was not helpful.[131] Clearly, if loop diuretics are given before adequate volume repletion, they run the risk of prolonging dehydration and may actually impair the excretion of a calcium load on this basis. Therefore, if diuretics are to be used, their administration must be delayed until volume status has been fully restored. In addition, they are valuable in managing hypercalcemic patients who develop fluid overload after aggressive fluid resuscitation.[131]

Inhibition of Bone Resorption

The majority of patients with MAHC are hypercalcemic because of excessive bone resorption. Therefore, the mainstay of therapy is the administration of powerful antiresorptive agents, primarily

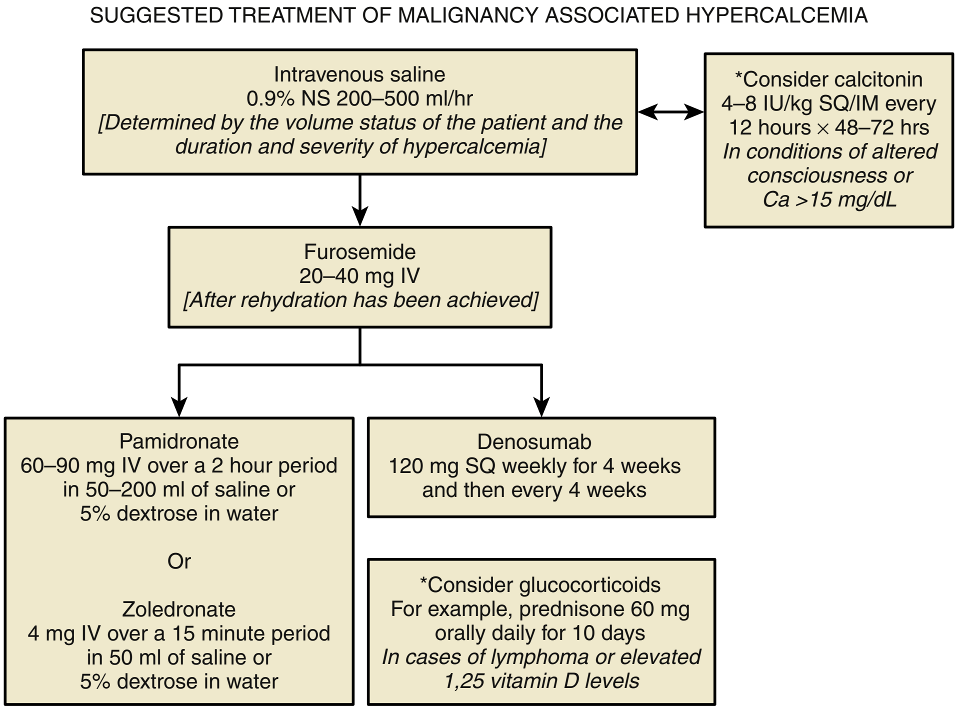

SUGGESTED TREATMENT OF MALIGNANCY ASSOCIATED HYPERCALCEMIA
Intravenous saline
0.9% NS 200–500 ml/hr
[Determined by the volume status of the patient and the duration and severity of hypercalcemia]
*Consider calcitonin
4–8 IU/kg SQ/IM every 12 hours × 48–72 hrs
In conditions of altered consciousness or Ca >15 mg/dL
Furosemide
20–40 mg IV
[After rehydration has been achieved]
Pamidronate
60–90 mg IV over a 2 hour period in 50–200 ml of saline or 5% dextrose in water
Or
Zoledronate
4 mg IV over a 15 minute period in 50 ml of saline or 5% dextrose in water
Denosumab
120 mg SQ weekly for 4 weeks and then every 4 weeks
*Consider glucocorticoids
For example, prednisone 60 mg orally daily for 10 days
In cases of lymphoma or elevated 1,25 vitamin D levels

Fig. 55.3 Suggested treatment for malignancy-associated hypercalcemia. *NS,* Normal saline; *IV,* intravenous; *SQ,* subcutaneous; *IM,* intramuscular.

bisphosphonates and denosumab. However, other agents, such as calcitonin, gallium nitrate, and mithramycin, have also been described to be helpful in lowering calcium levels. In the sections that follow, we will discuss the available agents for inhibiting bone resorption, concentrating on the most clinically useful and available agents.

Calcitonin. Calcitonin is a peptide hormone secreted by the parafollicular cells of the thyroid gland that acts to inhibit osteoclast activity. It has been used as an antiresorptive therapy for both hypercalcemia of malignancy and Paget disease.[127,130] It also has been shown to decrease the renal tubular resorption of calcium, thus promoting renal calcium excretion.[132,133] The benefit of calcitonin is its rapid onset of action, but its limitations are a short period of effectiveness and its relatively modest calcium-lowering actions. Salmon calcitonin has a longer half-life in the circulation than human calcitonin, and it is given as a subcutaneous or intramuscular injection at a dose of 4 to 8 IU per kilogram every 8 to 12 hours.[1] Administration of calcitonin results in a rapid reduction in serum calcium concentrations with a maximal response within 12 to 24 hours. However, the reductions are small (approximately 1 mg/dL), and the effects are usually lost within 48 to 96 hours, probably because of the downregulation of calcitonin receptors.[134] Nevertheless, it can be a valuable bridging agent that can rapidly lower calcium levels acutely while waiting for the more prolonged onset of other antiresorptive agents. This can be an important initial therapy for patients with extreme elevations in serum calcium levels. Side effects most commonly experienced with calcitonin include nausea, vomiting, and pain or irritation at the injection site.[127]

Bisphosphonates. Bisphosphonates are the most commonly used medications to treat MAHC in the United States. These drugs are analogs of pyrophosphates, which resist enzymatic hydrolysis and have a high affinity for hydroxyapatite.[135] Bisphosphonates concentrate at areas of high bone turnover and inhibit osteoclast function. Most patients are treated either with pamidronate (aminohydroxypropylidene bisphosphonate) or zoledronate (zoledronic acid), both potent nitrogen-containing bisphosphonates that interfere with protein prenylation through the inhibition of the mevalonate pathway enzyme, farnesyl disphosphate synthase, which is downstream of HMG-CoA reductase. These medications inhibit osteoclast function by inducing apoptosis.[135] The standard dose of pamidronate is 60 to 90 mg given intravenously over a 2-hour period in a solution of 50 to 200 mL of saline or 5% dextrose in water.[1,136] Zoledronate is given as a 4-mg intravenous infusion over 15 minutes in a solution of 50 mL of saline or 5% dextrose in water.[1] The serum calcium level normalizes in most patients, but the response can take 3 to 7 days and last from 1 to 3 weeks. In one study examining the dose-response to treatment with pamidronate in patients with MAHC, corrected serum calcium normalized in 61% of patients receiving 60 mg and 100% of patients receiving 90 mg. The mean duration of normalization in these two groups was 13.3 and 10.8 days in the 60-mg and 90-mg treatment groups, respectively.[136] Zoledronate has been shown to be superior to pamidronate in both clinical response and time to hypercalcemia relapse. One study, which was a pooled analysis from two randomized, double-blind, parallel-group trials, showed that zoledronate was more effective at normalizing serum calcium levels in patients with HCM, with a significantly longer time to relapse than pamidronate.[137,138] In this study, two treatment doses of zoledronic acid, 4 mg and 8 mg, were compared with 90 mg of pamidronate. The complete response rates were 88.4%, 86.7%, and 69.7%, respectively.[138] The corrected serum calcium normalized by day 4 in 50% of the zoledronic acid–treated patients and 33.3% of the pamidronate-treated patients.[138] The median duration of complete response was 32 and 43 days in those treated with 4 mg or 8 mg of zoledronic acid, respectively, and 18 days in those treated with pamidronate.[138]

The dosing interval of bisphosphonates to control MAHC is typically every 3 to 4 weeks, depending on the recurrence of hypercalcemia. There may be a reduction in efficacy with repeated dosing. The intravenous bisphosphonates are excreted unchanged by the kidneys and must be adjusted for impaired renal function. Zoledronate is generally contraindicated in patients with a creatinine clearance of less than 30 mL/min. The most common adverse effects are renal failure and 24 to 48 hours of transient flulike symptoms, including fever, chills, mild bone aches, and pain precipitated by the initial infusion. The fever is often mild and usually occurs only with the first infusion.

In addition to the acute treatment of MAHC, bisphosphonates are now routinely given to patients with myeloma and breast cancer to prevent the occurrence of SREs, including hypercalcemia. A comprehensive metaanalysis demonstrated that treatment of women with breast cancer and bone metastases with regular infusions of bisphosphonates reduced the overall rate of SREs by 15% in randomized trials and also delayed the onset of the first SRE.[81] However, this treatment did not prolong survival, and adjuvant treatment of breast cancer without bone metastases was not associated with enhanced survival.[139]

Denosumab. Denosumab is a humanized monoclonal antibody that binds to RANKL and prevents it from binding to the receptor RANK on osteoclast precursors and mature osteoclasts. This leads to the inhibition of the differentiation, activation, and function of osteoclasts which, in turn, leads to a reduction in bone resorption.[140] It has a profound suppressive effect on biochemical markers of bone resorption and is the most powerful antiresorptive agent currently available.[141] Denosumab is cleared through the reticuloendothelial system, independent from the kidneys.[140] The lack of renal toxicity makes it an attractive treatment option in patients with impaired renal function, who cannot tolerate bisphosphonates. Hypocalcemia is one of the main adverse effects of denosumab. Patients in osteoporosis trials also had a slightly increased incidence of skin infections as compared with patients treated with placebo, although this has not been specifically noted in cancer patients.[142]

Denosumab has been shown to prevent SREs or hypercalcemia of malignancy in phase III studies of advanced malignancies.[143-146] Administration of denosumab was associated with an 18% reduction in the risk for SREs[143] and a 70% success rate in normalization of serum calcium levels in a small study.[147] It also appears that denosumab may be useful in patients with MAHC who do not respond to bisphosphonates. In a study by Hu et al., patients with hypercalcemia of malignancy refractory to bisphosphonate treatment received subcutaneous denosumab on days 1, 8, 15, and 29, and then every 4 weeks at a dose of 120 mg given as a subcutaneous injection.[148] In this study, by day 10, 12 of 15 patients had responded with albumin-corrected serum calcium (CSC) levels less than 11.5 mg/dL, and the median duration of response was 26 days. In 10 of 15 patients, the CSC was less than 10.8 mg/dL.[148]

Glucocorticoids

Glucocorticoids can be an effective treatment for hypercalcemia associated with an overproduction of 1,25-dihydroxyvitamin D in patients with lymphoma.[127] This is likely because of the ability of glucocorticoids to suppress local cytokine production, resulting in the suppression of 1,25-dihydroxyvitamin D production through a reduction in vitamin D CYB27B1 activity.[130] As discussed previously, some patients with HHM and elevated PTHrP levels may also have elevations in 1,25-dihydroxyvitamin D, and this may predict an incomplete response to bisphosphonates.[109] Although it has not been systematically studied, it would be reasonable to consider adding steroids to the treatment of these patients as well. Treatment can consist of 60 mg/day of prednisone for 10 days.[1] This treatment can interfere with some chemotherapeutic agents, and treatment can be associated with hypokalemia, hyperglycemia, hypertension, Cushing syndrome, immunosuppression, peptic ulcer disease, steroid-induced psychosis, and increased risk of tumor lysis syndrome.[1,149] If glucocorticoids are not effective, or if they are not tolerated, another option for lowering 1,25-dihydroxyvitamin D production is to use the P450 inhibitor, ketoconazole, which has been shown to reduce 1,25-dihydroxyvitamin D levels and hypercalcemia.[150]

Dialysis

Patients with hypercalcemia and acute kidney injury with oliguria may require dialysis to lower their calcium levels. In these patients, a saline diuresis is not possible and could lead to significant fluid overload. Hemodialysis or peritoneal dialysis using a very low calcium or calcium-free dialysate can be considered in such situations[127] and can result in a rapid decrease in serum total calcium levels.[151]

Other treatments

Gallium nitrate was initially evaluated as a cancer chemotherapeutic and was noted to cause hypocalcemia by inhibiting bone resorption.[152] It was given as a continuous intravenous infusion of 100 to 200 mg/m^2 of body surface area over a 24-hour period for 5 consecutive days.[1,127] Gallium nitrate can cause nephrotoxicity, and it is important that adequate urine output be maintained on this therapy.[153] In one study comparing treatment with gallium nitrate to calcitonin, 18 of 24 patients receiving gallium nitrate achieved normocalcemia, and the median duration of normocalcemia was 6 days.[153] In another study comparing gallium nitrate therapy to treatment with etidronate, 28 of 34 patients (82%) who received gallium nitrate achieved normocalcemia, and the median duration of normocalcemia was 8 days.[152] Gallium nitrate was withdrawn from the market in 1995 and discontinued by the manufacturer in 2012 in the United States.

Mithramycin, otherwise known as plicamycin, is an antibiotic agent that was initially used to treat testicular neoplasms and was found to lower serum calcium levels.[154] It was used for the treatment of hypercalcemia of malignancy before the introduction of bisphosphonates. It interferes with osteoclastic function and can reduce serum calcium levels within 24 hours.[155] It is given as a single dose of 25 mcg/kg of body weight over a 4- to 6-hour period in saline. Its side effects include thrombocytopenia, platelet defects resulting in bleeding diathesis, azotemia, proteinuria, and hepatic toxicity.[156] It is rarely, if ever, used currently.

New treatment options for MAHC are being developed but have not yet been approved for clinical treatment. Cinacalcet is a calcimimetic that works by decreasing PTH production. It has been used to successfully to treat hypercalcemia secondary to parathyroid carcinoma and is approved for this indication by the US Food and Drug Administration.[157,158] Cinacalcet has also been reported in small case series to be successful for off-label use to treat MAHC.[159,160] Side effects of this treatment include nausea, vomiting, and headaches. Additional studies are needed to determine cinacalcet's therapeutic potential for MAHC. Lastly, anti-PTHrP and anti-PTH1R antibodies are being tested in animal models.[157,158,161,162] These novel therapies may offer additional treatment options in patients who are refractory or nonresponsive to current MAHC therapies.

Finally, it must be underscored that all the treatment options for MAHC described earlier can be used to stabilize the patient until the underlying malignancy can be appropriately treated, as successful antitumor therapy is the only true cure for MAHC.

KEY POINTS

- Hypercalcemia of malignancy can be effectively treated by lowering bone resorption and by treating the underlying malignancy if possible.

REFERENCES

1. Stewart AF. Clinical practice. Hypercalcemia associated with cancer. *N Engl J Med*. 2005;352(4):373–379.
2. Ralston SH, Gallacher SJ, Patel U, et al. Cancer-associated hypercalcemia: morbidity and mortality. Clinical experience in 126 treated patients. *Ann Intern Med*. 1990;112(7):499–504.
3. Bhandari S, Kumar R, Tripathi P, et al. Outcomes of hypercalcemia of malignancy in patients with solid cancer: a national inpatient analysis. *Med Oncol*. 2019;36(10):90.
4. Gastanaga VM, Schwartzberg LS, Jain RK, et al. Prevalence of hypercalcemia among cancer patients in the United States. *Cancer Med*. 2016;5(8):2091–2100.
5. Donovan PJ, Achong N, Griffin K, et al. PTHrP-mediated hypercalcemia: causes and survival in 138 patients. *J Clin Endocrinol Metab*. 2015;100(5):2024–2029.
6. Jin J, Chung JO, Chung MY, et al. Clinical Characteristics, causes and survival in 115 cancer patients with parathyroid hormone related protein-mediated hypercalcemia. *J Bone Metab*. 2017;24(4):249–255.
7. Ramos REO, Perez Mak M, Alves MFS, et al. Malignancy-Related hypercalcemia in advanced solid tumors: survival outcomes. *J Glob Oncol*. 2017;3(6):728–733.
8. Zondek SG. Die Bedeutung der Calcium- und Kaliumionen bei Giftwirkungen am Herzen. *Arch Exp Path Pharm*. 1920;88(3):158–171.
9. Gutman AB, Tyson TL, Gutman EB. Serum calcium, inorganic phosphorus and phosphatase activity: in hyperparathyroidism, Paget's disease, multiple myeloma and neoplastic disease of the bones. *Arch Intern Med*. 1936;57(2):379–413.
10. C, CR and MTB. Case 27461. *N Engl J Med*. 1941;225(20):789–791.
11. Gellhorn A, Plimpton CH. Hypercalcemia in malignant disease without evidence of bone destruction. *Am J Med*. 1956;21(5):750–759.
12. Connor TB, Thomas WC, Howard JE. Etiology of hypercalcemia associated with lung carcinoma. *J Clin Invest*. 1956;35:697–698.
13. Lafferty FW. Pseudohyperparathyroidism. *Medicine (Baltim)*. 1966;45(3):247–260.
14. Burtis WJ, Brady TG, Orloff JJ, et al. Immunochemical characterization of circulating parathyroid hormone-related protein in patients with humoral hypercalcemia of cancer. *N Engl J Med*. 1990;322(16):1106–1112.
15. Stewart AF, Horst R, Deftos LJ, et al. Biochemical evaluation of patients with cancer-associated hypercalcemia: evidence for humoral and nonhumoral groups. *N Engl J Med*. 1980;303(24):1377–1383.
16. Wysolmerski J, Insogna K. In: Goldman L, Schafer A, eds. *The parathyroid glands, hypercalcemia, and hypocalcemia. , in Goldman's Cecil Medicine*. Philadelphia: Elsevier Saunders; 2011:1591–1601.
17. Brenner B, Stein H. In: *Divalent Ion Homeostasis. (Contemporary Issues in Nephrology)*. 2nd ed. 1997.
18. Brown EM. Four-parameter model of the sigmoidal relationship between parathyroid hormone release and extracellular calcium concentration in normal and abnormal parathyroid tissue. *J Clin Endocrinol Metab*. 1983;56(3):572–581.
19. Brown EM, Gamba G, Riccardi D, et al. Cloning and characterization of an extracellular Ca(2+)-sensing receptor from bovine parathyroid. *Nature*. 1993;366(6455):575–580.
20. Chattopadhyay N, Mithal A, Brown EM. The calcium-sensing receptor: a window into the physiology and pathophysiology of mineral ion metabolism. *Endocr Rev*. 1996;17(4):289–307.
21. Hofer AM, Gerbino A, Caroppo R, et al. The extracellular calcium-sensing receptor and cell-cell signaling in epithelia. *Cell Calcium*. 2004;35(3):297–306.
22. Mayer GP, Hurst JG. Sigmoidal relationship between parathyroid hormone secretion rate and plasma calcium concentration in calves. *Endocrinology*. 1978;102(4):1036–1042.
23. Riccardi D, Kemp PJ. The calcium-sensing receptor beyond extracellular calcium homeostasis: conception, development, adult physiology, and disease. *Annu Rev Physiol*. 2012;74:271–297.
24. Chen RA, Goodman WG. Role of the calcium-sensing receptor in parathyroid gland physiology. *Am J Physiol Renal Physiol*. 2004;286(6):F1005–F1011.
25. Habener JF, Rosenblatt M, Potts Jr JT. Parathyroid hormone: biochemical aspects of biosynthesis, secretion, action, and metabolism. *Physiol Rev*. 1984;64(3):985–1053.
26. Kantham L, Quinn SJ, Egbuna OI, et al. The calcium-sensing receptor (CaSR) defends against hypercalcemia independently of its regulation of parathyroid hormone secretion. *Am J Physiol Endocrinol Metab*. 2009;297(4):E915–E923.
27. Murer H, Forster I, Biber J. The sodium phosphate cotransporter family SLC34. *Pflugers Arch*. 2004;447(5):763–767.
28. Bisello A, Friedman P. In: Bilezikian JP, Raisz LG, Martin TJ, eds. *PTH and PTHrP Actions on Kidney and Bone, in Principles of Bone Biology, Two-Volume Set*. USA: Elsevier Science; 2008:665–712.
29. Friedman PA, Gesek FA. Calcium transport in renal epithelial cells. *Am J Physiol*. 1993;264(2 Pt 2):F181–F198.
30. Nishida S, Yamaguchi A, Tanizawa T, et al. Increased bone formation by intermittent parathyroid hormone administration is due to the stimulation of proliferation and differentiation of osteoprogenitor cells in bone marrow. *Bone*. 1994;15(6):717–723.
31. Teitelbaum SL, Ross FP. Genetic regulation of osteoclast development and function. *Nat Rev Genet*. 2003;4(8):638–649.
32. MacLaughlin JA, Anderson RR, Holick MF. Spectral character of sunlight modulates photosynthesis of previtamin D3 and its photoisomers in human skin. *Science*. 1982;216(4549):1001–1003.
33. Bikle D. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Vitamin D: Production, Metabolism, and Mechanisms of Action, in Endotext*. South Dartmouth, MA: MDText.com, Inc; 2000.
34. Fleet JC, DeSmet M, Johnson R, et al. Vitamin D and cancer: a review of molecular mechanisms. *Biochem J*. 2012;441(1):61–76.
35. Fleet JC, Schoch RD. Molecular mechanisms for regulation of intestinal calcium absorption by vitamin D and other factors. *Crit Rev Clin Lab Sci*. 2010;47(4):181–195.
36. Kearns AE, Khosla S, Kostenuik PJ. Receptor activator of nuclear factor kappaB ligand and osteoprotegerin regulation of bone remodeling in health and disease. *Endocr Rev*. 2008;29(2):155–192.
37. Takeda S, Yoshizawa T, Nagai Y, et al. Stimulation of osteoclast formation by 1,25-dihydroxyvitamin D requires its binding to vitamin D receptor (VDR) in osteoblastic cells: studies using VDR knockout mice. *Endocrinology*. 1999;140(2):1005–1008.
38. Mangin M, Webb AC, Dreyer BE, et al. Identification of a cDNA encoding a parathyroid hormone-like peptide from a human tumor associated with humoral hypercalcemia of malignancy. *Proc Natl Acad Sci U S A*. 1988;85(2):597–601.
39. Philbrick WM, Wysolmerski JJ, Galbraith S, et al. Defining the roles of parathyroid hormone-related protein in normal physiology. *Physiol Rev*. 1996;76(1):127–173.
40. Suva LJ, Winslow GA, Wettenhall RE, et al. A parathyroid hormone-related protein implicated in malignant hypercalcemia: cloning and expression. *Science*. 1987;237(4817):893–896.
41. VanHouten J, Dann P, McGeoch G, et al. The calcium-sensing receptor regulates mammary gland parathyroid hormone-related protein production and calcium transport. *J Clin Invest*. 2004;113(4):598–608.
42. Dean T, Vilardaga JP, Potts Jr JT, et al. Altered selectivity of parathyroid hormone (PTH) and PTH-related protein (PTHrP) for distinct conformations of the PTH/PTHrP receptor. *Mol Endocrinol*. 2008;22(1):156–166.
43. Kovacs CS. Maternal mineral and bone metabolism during pregnancy, lactation, and post-weaning recovery. *Physiol Rev*. 2016;96(2):449–547.
44. VanHouten JN, Dann P, Stewart AF, et al. Mammary-specific deletion of parathyroid hormone-related protein preserves bone mass during lactation. *J Clin Invest*. 2003;112(9):1429–1436.
45. Fritchie K, Zedek D, Grenache DG. The clinical utility of parathyroid hormone-related peptide in the assessment of hypercalcemia. *Clin Chim Acta*. 2009;402(1–2):146–149.
46. Szymanski JJ, Otrock ZK, Patel KK, et al. Incidence of humoral hypercalcemia of malignancy among hypercalcemic patients with cancer. *Clin Chim Acta*. 2016;453:190–193.
47. Wright JD, Tergas AI, Ananth CV, et al. Quality and Outcomes of treatment of hypercalcemia of malignancy. *Cancer Invest*. 2015;33(8):331–339.

48. Watanabe T, Yamaguchi K, Takatsuki K, et al. Constitutive expression of parathyroid hormone-related protein gene in human T cell leukemia virus type 1 (HTLV-1) carriers and adult T cell leukemia patients that can be trans-activated by HTLV-1 tax gene. *J Exp Med*. 1990;172(3):759–765.
49. Palapattu GS, Kristo B, Rajfer J. Paraneoplastic syndromes in urologic malignancy: the many faces of renal cell carcinoma. *Rev Urol*. 2002;4(4):163–170.
50. Alsirafy SA, Sroor MY, Al-Shahri MZ. Hypercalcemia in advanced head and neck squamous cell carcinoma: prevalence and potential impact on palliative care. *J Support Oncol*. 2009;7(5):154–157.
51. Le Tinier F, Vanhuyse M, Penel N, et al. Cancer-associated hypercalcaemia in squamous-cell malignancies: a survival and prognostic factor analysis. *Int J Oral Maxillofac Surg*. 2011;40(9):938–942.
52. Penel N, Dewas S, Doutrelant P, et al. Cancer-associated hypercalcemia treated with intravenous diphosphonates: a survival and prognostic factor analysis. *Support Care Cancer*. 2008;16(4):387–392.
53. Stewart AF, Vignery A, Silverglate A, et al. Quantitative bone histomorphometry in humoral hypercalcemia of malignancy: uncoupling of bone cell activity. *J Clin Endocrinol Metab*. 1982;55(2):219–227.
54. Nakayama K, Fukumoto S, Takeda S, et al. Differences in bone and vitamin D metabolism between primary hyperparathyroidism and malignancy-associated hypercalcemia. *J Clin Endocrinol Metab*. 1996;81(2):607–611.
55. Horwitz MJ, Tedesco MB, Sereika SM, et al. Direct comparison of sustained infusion of human parathyroid hormone-related protein-(1-36) [hPTHrP-(1-36)] versus hPTH-(1-34) on serum calcium, plasma 1,25-dihydroxyvitamin D concentrations, and fractional calcium excretion in healthy human volunteers. *J Clin Endocrinol Metab*. 2003;88(4):1603–1609.
56. Horwitz MJ, Tedesco MB, Sereika SM, et al. Continuous PTH and PTHrP infusion causes suppression of bone formation and discordant effects on 1,25(OH)2 vitamin D. *J Bone Miner Res*. 2005;20(10):1792–1803.
57. Syed MA, Horwitz MJ, Tedesco MB, et al. Parathyroid hormone-related protein-(1--36) stimulates renal tubular calcium reabsorption in normal human volunteers: implications for the pathogenesis of humoral hypercalcemia of malignancy. *J Clin Endocrinol Metab*. 2001;86(4):1525–1531.
58. Nagai Y, Yamato H, Akaogi K, et al. Role of interleukin-6 in uncoupling of bone in vivo in a human squamous carcinoma coproducing parathyroid hormone-related peptide and interleukin-6. *J Bone Miner Res*. 1998;13(4):664–672.
59. Cupp ME, Nayak SK, Adem AS, et al. Parathyroid hormone (PTH) and PTH-related peptide domains contributing to activation of different PTH receptor-mediated signaling pathways. *J Pharmacol Exp Ther*. 2013;345(3):404–418.
60. Datta NS, Abou-Samra AB. PTH and PTHrP signaling in osteoblasts. *Cell Signal*. 2009;21(8):1245–1254.
61. Budayr AA, Nissenson RA, Klein RF, et al. Increased serum levels of a parathyroid hormone-like protein in malignancy-associated hypercalcemia. *Ann Intern Med*. 1989;111(10):807–812.
62. Brook N, Brook E, Dharmarajan A, et al. Breast cancer bone metastases: pathogenesis and therapeutic targets. *Int J Biochem Cell Biol*. 2018;96:63–78.
63. Guise TA, Mohammad KS, Clines G, et al. Basic mechanisms responsible for osteolytic and osteoblastic bone metastases. *Clin Cancer Res*. 2006;12(20 Pt 2):6213s–6216s.
64. Clohisy DR, Palkert D, Ramnaraine ML, et al. Human breast cancer induces osteoclast activation and increases the number of osteoclasts at sites of tumor osteolysis. *J Orthop Res*. 1996;14(3):396–402.
65. Pederson L, Winding B, Foged NT, et al. Identification of breast cancer cell line-derived paracrine factors that stimulate osteoclast activity. *Cancer Res*. 1999;59(22):5849–5855.
66. Winding B, Misander H, Sveigaard C, et al. Human breast cancer cells induced angiogenesis, recruitment, and activation of osteoclasts in osteolytic metastasis. *J Cancer Res Clin Oncol*. 2000;126(11):631–640.
67. Yoneda T, Sasaki A, Mundy GR. Osteolytic bone metastasis in breast cancer. *Breast Cancer Res Treat*. 1994;32(1):73–84.
68. Buijs JT, Stayrook KR, Guise TA. The role of TGF-β in bone metastasis: novel therapeutic perspectives. *BoneKEy Rep*. 2012;1:96.
69. Eyob H, Ekiz HA, Welm AL. RON promotes the metastatic spread of breast carcinomas by subverting antitumor immune responses. *OncoImmunology*. 2013;2(9):e25670.
70. Papachristou DJ, Basdra EK, Papavassiliou AG. Bone metastases: molecular mechanisms and novel therapeutic interventions. *Med Res Rev*. 2012;32(3):611–636.
71. Mundy GR, Raisz LG, Cooper RA, et al. Evidence for the secretion of an osteoclast stimulating factor in myeloma. *N Engl J Med*. 1974;291(20):1041–1046.
72. Abe M, Hiura K, Wilde J, et al. Role for macrophage inflammatory protein (MIP)-1alpha and MIP-1beta in the development of osteolytic lesions in multiple myeloma. *Blood*. 2002;100(6):2195–2202.
73. Gao JL, Kuhns DB, Tiffany HL, et al. Structure and functional expression of the human macrophage inflammatory protein 1 alpha/RANTES receptor. *J Exp Med*. 1993;177(5):1421–1427.
74. Shimizu H, Monden T, Tomotsune T, et al. A case of myeloma with hypercalcemia caused by high serum concentrations of both parathyroid hormone-related peptide (PTHrP) and macrophage inflammatory protein-1α (MIP-1α). *Intern Med*. 2011;50(24):2993–2996.
75. Delgado-Calle J. Osteocytes and their messengers as targets for the treatment of multiple myeloma. *Clin Rev Bone Miner Metab*. 2017;15(1):49–56.
76. Heider U, Kaiser M, Mieth M, et al. Serum concentrations of DKK-1 decrease in patients with multiple myeloma responding to anti-myeloma treatment. *Eur J Haematol*. 2009;82(1):31–38.
77. Tian E, Zhan F, Walker R, et al. The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma. *N Engl J Med*. 2003;349(26):2483–2494.
78. Zhou F, Meng S, Song H, et al. Dickkopf-1 is a key regulator of myeloma bone disease: opportunities and challenges for therapeutic intervention. *Blood Rev*. 2013;27(6):261–267.
79. Bower M, Cox S. In: Doyle D, Hanks G, Cherny N, et al., eds. *Endocrine and Metabolic Complications of Advanced Cancer, in Oxford Textbook of Palliative Medicine*. New York, NY: Oxford University Press; 2004:688–690.
80. O'Carrigan B, Wong MH, Willson ML, et al. Bisphosphonates and other bone agents for breast cancer. *Cochrane Database Syst Rev*. 2017;10(10):Cd003474.
81. Wong MH, Stockler MR, Pavlakis N. Bisphosphonates and other bone agents for breast cancer. *Cochrane Database Syst Rev*. 2012;(2):Cd003474.
82. Isales C, Carcangiu ML, Stewart AF. Hypercalcemia in breast cancer. Reassessment of the mechanism. *Am J Med*. 1987;82(6):1143–1147.
83. Guise TA, Yin JJ, Taylor SD, et al. Evidence for a causal role of parathyroid hormone-related protein in the pathogenesis of human breast cancer-mediated osteolysis. *J Clin Invest*. 1996;98(7):1544–1549.
84. Yin JJ, Selander K, Chirgwin JM, et al. TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. *J Clin Invest*. 1999;103(2):197–206.
85. Akhtari M, Mansuri J, Newman KA, et al. Biology of breast cancer bone metastasis. *Cancer Biol Ther*. 2008;7(1):3–9.
86. Johnson RW, Nguyen MP, Padalecki SS, et al. TGF-beta promotion of Gli2-induced expression of parathyroid hormone-related protein, an important osteolytic factor in bone metastasis, is independent of canonical Hedgehog signaling. *Cancer Res*. 2011;71(3):822–831.
87. Mamillapalli R, VanHouten J, Zawalich W, et al. Switching of G-protein usage by the calcium-sensing receptor reverses its effect on parathyroid hormone-related protein secretion in normal versus malignant breast cells. *J Biol Chem*. 2008;283(36):24435–24447.
88. Sanders JL, Chattopadhyay N, Kifor O, et al. Extracellular calcium-sensing receptor expression and its potential role in regulating parathyroid hormone-related peptide secretion in human breast cancer cell lines. *Endocrinology*. 2000;141(12):4357–4364.
89. Ruppender NS, Merkel AR, Martin TJ, et al. Matrix rigidity induces osteolytic gene expression of metastatic breast cancer cells. *PLoS One*. 2010;5(11):e15451.
90. Thomas RJ, Guise TA, Yin JJ, et al. Breast cancer cells interact with osteoblasts to support osteoclast formation. *Endocrinology*. 1999;140(10):4451–4458.

91. Welm AL, Sneddon JB, Taylor C, et al. The macrophage-stimulating protein pathway promotes metastasis in a mouse model for breast cancer and predicts poor prognosis in humans. *Proc Natl Acad Sci U S A*. 2007;104(18):7570–7575.
92. Adams JS. Vitamin D metabolite-mediated hypercalcemia. *Endocrinol Metab Clin North Am*. 1989;18(3):765–778.
93. Breslau NA, McGuire JL, Zerwekh JE, et al. Hypercalcemia associated with increased serum calcitriol levels in three patients with lymphoma. *Ann Intern Med*. 1984;100(1):1–6.
94. Mundy GR, Luben RA, Raisz LG, et al. Bone-resorbing activity in supernatants from lymphoid cell lines. *N Engl J Med*. 1974;290(16):867–871.
95. Rosenthal N, Insogna KL, Godsall JW, et al. Elevations in circulating 1,25-dihydroxyvitamin D in three patients with lymphoma-associated hypercalcemia. *J Clin Endocrinol Metab*. 1985;60(1):29–33.
96. Seymour JF, Gagel RF. Calcitriol: the major humoral mediator of hypercalcemia in Hodgkin's disease and non-Hodgkin's lymphomas. *Blood*. 1993;82(5):1383–1394.
97. Hewison M, Kantorovich V, Liker HR, et al. Vitamin D-mediated hypercalcemia in lymphoma: evidence for hormone production by tumor-adjacent macrophages. *J Bone Miner Res*. 2003;18(3):579–582.
98. Evans KN, Taylor H, Zehnder D, et al. Increased expression of 25-hydroxyvitamin D-1alpha-hydroxylase in dysgerminomas: a novel form of humoral hypercalcemia of malignancy. *Am J Pathol*. 2004;165(3):807–813.
99. Iqbal I, Khan MAA, Khan Y, et al. Hypercalcemia due to paraneoplastic production of 1,25- dihydroxyvitamin D in a young female with dysgerminoma. *Cureus*. 2019;11(11):e6097.
100. Osorio JC, Jones MG, Schatz-Siemers N, et al. Twist on a classic: vitamin D and hypercalcaemia of malignancy. *BMJ Case Rep*. 2017;2017.
101. Rodríguez-Gutiérrez R, Zapata-Rivera MA, Quintanilla-Flores DL, et al. 1,25-dihydroxyvitamin D and PTHrP mediated malignant hypercalcemia in a seminoma. *BMC Endocr Disord*. 2014;14(1):32.
102. Hoekman K, Tjandra YI, Papapoulos SE. The role of 1,25-dihydroxyvitamin D in the maintenance of hypercalcemia in a patient with an ovarian carcinoma producing parathyroid hormone-related protein. *Cancer*. 1991;68(3):642–647.
103. Nemr S, Alluri S, Sundaramurthy D, et al. Hypercalcemia in lung cancer due to simultaneously elevated PTHrP and ectopic calcitriol production: first case report. *Case Rep Oncol Med*. 2017;2017:2583217.
104. Ogawa T, Miyata J, Fukunaga K, et al. Successful treatment for hypercalcemia due to cosecretion of parathyroid hormone-related protein and 1,25-dihydroxyvitamin D(3) in non-small-cell lung cancer: a case report and literature review. *Case Rep Pulmonol*. 2020;2020:2475725. 2475725.
105. Rodríguez-Gutiérrez R, Zapata-Rivera MA, Quintanilla-Flores DL, et al. 1,25-dihydroxyvitamin D and PTHrP mediated malignant hypercalcemia in a seminoma. *BMC Endocr Disord*. 2014;14:32.
106. Shivnani SB, Shelton JM, Richardson JA, et al. Hypercalcemia of malignancy with simultaneous elevation in serum parathyroid hormone—related peptide and 1,25-dihydroxyvitamin D in a patient with metastatic renal cell carcinoma. *Endocr Pract*. 2009;15(3):234–239.
107. Van den Eynden GG, Neyret A, Fumey G, et al. PTHrP, calcitonin and calcitriol in a case of severe, protracted and refractory hypercalcemia due to a pancreatic neuroendocrine tumor. *Bone*. 2007;40(4):1166–1171.
108. Benit A, Allard J, Rimailho J, et al. Persistent and moderate hypercalcemia related to an ovarian clear cell adenocarcinoma: pre- and postoperative parathyroid hormone related-peptide and 1,25-dihydroxyvitamin D3 levels. *J Endocrinol Invest*. 2006;29(5):443–449.
109. Chukir T, Liu Y, Hoffman K, et al. Calcitriol elevation is associated with a higher risk of refractory hypercalcemia of malignancy in solid tumors. *J Clin Endocrinol Metab*. 2020;105(4):e1115–e1123.
110. Arps H, Dietel M, Schulz A, et al. Pancreatic endocrine carcinoma with ectopic PTH-production and paraneoplastic hypercalcaemia. *Virchows Arch A Pathol Anat Histopathol*. 1986;408(5):497–503.
111. Botea V, Edelson GW, Munasinghe RL. Hyperparathyroidism, hypercalcemia, and calcified brain metastatic lesions in a patient with small cell carcinoma demonstrating positive immunostain for parathyroid hormone. *Endocr Pract*. 2003;9(1):40–44.
112. Buller R, Taylor K, Burg AC, et al. Paraneoplastic hypercalcemia associated with adenosquamous carcinoma of the endometrium. *Gynecol Oncol*. 1991;40(1):95–98.
113. Demura M, Yoneda T, Wang F, et al. Ectopic production of parathyroid hormone in a patient with sporadic medullary thyroid cancer. *Endocr J*. 2010;57(2):161–170.
114. Doyle MA, Malcolm JC. An unusual case of malignancy-related hypercalcemia. *Int J Gen Med*. 2013;7:21–27.
115. Gabriel FGC, Picar RE. A rare case of hypercalcemic encephalopathy from ectopic secretion of parathyroid hormone. *Clin Case Rep*. 2020;8(3):423–425.
116. Kandil E, Noureldine S, Khalek MA, et al. Ectopic secretion of parathyroid hormone in a neuroendocrine tumor: a case report and review of the literature. *Int J Clin Exp Med*. 2011;4(3):234–240.
117. Koyama Y, Ishijima H, Ishibashi A, et al. Intact PTH-producing hepatocellular carcinoma treated by transcatheter arterial embolization. *Abdom Imaging*. 1999;24(2):144–146.
118. Mahoney EJ, Monchik JM, Donatini G, et al. Life-threatening hypercalcemia from a hepatocellular carcinoma secreting intact parathyroid hormone: localization by sestamibi single-photon emission computed tomographic imaging. *Endocr Pract*. 2006;12(3):302–306.
119. Mayes LC, Kasselberg AG, Roloff JS, et al. Hypercalcemia associated with immunoreactive parathyroid hormone in a malignant rhabdoid tumor of the kidney (rhabdoid Wilms' tumor). *Cancer*. 1984;54(5):882–884.
120. Nussbaum SR, Gaz RD, Arnold A. Hypercalcemia and ectopic secretion of parathyroid hormone by an ovarian carcinoma with rearrangement of the gene for parathyroid hormone. *N Engl J Med*. 1990;323(19):1324–1328.
121. Rizzoli R, Pache JC, Didierjean L, et al. A thymoma as a cause of true ectopic hyperparathyroidism. *J Clin Endocrinol Metab*. 1994;79(3):912–915.
122. Samaan NA, Ordonez NG, Ibanez ML, et al. Ectopic parathyroid hormone production by a squamous carcinoma of the tonsil. *Arch Otolaryngol*. 1983;109(2):91–94.
123. Strewler GJ, Budayr AA, Clark OH, et al. Production of parathyroid hormone by a malignant nonparathyroid tumor in a hypercalcemic patient. *J Clin Endocrinol Metab*. 1993;76(5):1373–1375.
124. Uchimura K, Mokuno T, Nagasaka A, et al. Lung cancer associated with hypercalcemia induced by concurrently elevated parathyroid hormone and parathyroid hormone-related protein levels. *Metabolism*. 2002;51(7):871–875.
125. VanHouten JN, Yu N, Rimm D, et al. Hypercalcemia of malignancy due to ectopic transactivation of the parathyroid hormone gene. *J Clin Endocrinol Metab*. 2006;91(2):580–583.
126. Yoshimoto K, Yamasaki R, Sakai H, et al. Ectopic production of parathyroid hormone by small cell lung cancer in a patient with hypercalcemia. *J Clin Endocrinol Metab*. 1989;68(5):976–981.
127. Rosner MH, Dalkin AC. Onco-nephrology: the pathophysiology and treatment of malignancy-associated hypercalcemia. *Clin J Am Soc Nephrol*. 2012;7(10):1722–1729.
128. Carroll R, Matfin G. Endocrine and metabolic emergencies: hypercalcaemia. *Ther Adv Endocrinol Metab*. 2010;1(5):225–234.
129. Dellay B, Groth M. Emergency management of malignancy-associated hypercalcemia. *Adv Emerg Nurs J*. 2016;38(1):15–25quiz E1.
130. Clines GA. Mechanisms and treatment of hypercalcemia of malignancy. *Curr Opin Endocrinol Diabetes Obes*. 2011;18(6):339–346.
131. LeGrand SB, Leskuski D, Zama I. Narrative review: furosemide for hypercalcemia: an unproven yet common practice. *Ann Intern Med*. 2008;149(4):259–263.
132. Cochran M, Peacock M, Sachs G, et al. Renal effects of calcitonin. *Br Med J*. 1970;1(5689):135–137.
133. Hosking DJ, Gilson D. Comparison of the renal and skeletal actions of calcitonin in the treatment of severe hypercalcaemia of malignancy. *Q J Med*. 1984;53(211):359–368.
134. Binstock ML, Mundy GR. Effect of calcitonin and glutocorticoids in combination on the hypercalcemia of malignancy. *Ann Intern Med*. 1980;93(2):269–272.

135. Zaidi M, Blair HC, Moonga BS, et al. Osteoclastogenesis, bone resorption, and osteoclast-based therapeutics. *J Bone Miner Res*. 2003;18(4):599–609.
136. Nussbaum SR, Younger J, Vandepol CJ, et al. Single-dose intravenous therapy with pamidronate for the treatment of hypercalcemia of malignancy: comparison of 30-, 60-, and 90-mg dosages. *Am J Med*. 1993;95(3):297–304.
137. Cheer SM, Noble S. Zoledronic acid. *Drugs*. 2001;61(6):799–805. discussion 806.
138. Major P, Lortholary A, Hon J, et al. Zoledronic acid is superior to pamidronate in the treatment of hypercalcemia of malignancy: a pooled analysis of two randomized, controlled clinical trials. *J Clin Oncol*. 2001;19(2):558–567.
139. Coleman RE, Marshall H, Cameron D, et al. Breast-cancer adjuvant therapy with zoledronic acid. *N Engl J Med*. 2011;365(15):1396–1405.
140. Baron R, Ferrari S, Russell RG. Denosumab and bisphosphonates: different mechanisms of action and effects. *Bone*. 2011;48(4):677–692.
141. Boikos SA, Hammers HJ. Denosumab for the treatment of bisphosphonate-refractory hypercalcemia. *J Clin Oncol*. 2012;30(29):e299.
142. Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*. 2009;361(8):756–765.
143. Fizazi K, Carducci M, Smith M, et al. Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study. *Lancet*. 2011;377(9768):813–822.
144. Henry DH, Costa L, Goldwasser F, et al. Randomized, double-blind study of denosumab versus zoledronic acid in the treatment of bone metastases in patients with advanced cancer (excluding breast and prostate cancer) or multiple myeloma. *J Clin Oncol*. 2011;29(9):1125–1132.
145. Lipton A, Fizazi K, Stopeck AT, et al. Superiority of denosumab to zoledronic acid for prevention of skeletal-related events: a combined analysis of 3 pivotal, randomised, phase 3 trials. *Eur J Cancer*. 2012;48(16):3082–3092.
146. Stopeck AT, Lipton A, Body JJ, et al. Denosumab compared with zoledronic acid for the treatment of bone metastases in patients with advanced breast cancer: a randomized, double-blind study. *J Clin Oncol*. 2010;28(35):5132–5139.
147. Dietzek A, Connelly K, Cotugno M, et al. Denosumab in hypercalcemia of malignancy: a case series. *J Oncol Pharm Pract*. 2015;21(2):143–147.
148. Hu MI, Glezerman I, Leboulleux S, et al. Denosumab for patients with persistent or relapsed hypercalcemia of malignancy despite recent bisphosphonate treatment. *J Natl Cancer Inst*. 2013;105(18):1417–1420.
149. Malangone S, Campen CJ. Hypercalcemia of malignancy. *J Adv Pract Oncol*. 2015;6(6):586–592.
150. Glass AR, Eil C. Ketoconazole-induced reduction in serum 1,25-dihydroxyvitamin D and total serum calcium in hypercalcemic patients. *J Clin Endocrinol Metab*. 1988;66(5):934–938.
151. Camus C, Charasse C, Jouannic-Montier I, et al. Calcium free hemodialysis: experience in the treatment of 33 patients with severe hypercalcemia. *Intensive Care Med*. 1996;22(2):116–121.
152. Warrell Jr RP, Murphy WK, Schulman P, et al. A randomized double-blind study of gallium nitrate compared with etidronate for acute control of cancer-related hypercalcemia. *J Clin Oncol*. 1991;9(8):1467–1475.
153. Warrell Jr RP, Israel R, Frisone M, et al. Gallium nitrate for acute treatment of cancer-related hypercalcemia. A randomized, double-blind comparison to calcitonin. *Ann Intern Med*. 1988;108(5):669–674.
154. Perlia CP, Gubisch NJ, Wolter J, et al. Mithramycin treatment of hypercalcemia. *Cancer*. 1970;25(2):389–394.
155. Ralston SH, Gardner MD, Dryburgh FJ, et al. Comparison of aminohydroxypropylidene diphosphonate, mithramycin, and corticosteroids/calcitonin in treatment of cancer-associated hypercalcaemia. *Lancet*. 1985;2(8461):907–910.
156. Stewart AF. Therapy of malignancy-associated hypercalcemia: 1983. *Am J Med*. 1983;74(3):475–480.
157. Mirrakhimov AE. Hypercalcemia of malignancy: an update on pathogenesis and management. *N Am J Med Sci*. 2015;7(11):483–493.
158. Sternlicht H, Glezerman IG. Hypercalcemia of malignancy and new treatment options. *Ther Clin Risk Manag*. 2015;11:1779–1788.
159. Asonitis N, Kassi E, Kokkinos M, et al. Hypercalcemia of malignancy treated with cinacalcet. *Endocrinol Diabetes Metab Case Rep*. 2017;2017:17–118.
160. Sheehan M, Tanimu S, Tanimu Y, et al. Cinacalcet for the treatment of humoral hypercalcemia of malignancy: an introductory case report with a pathophysiologic and therapeutic review. *Case Rep Oncol*. 2020;13(1):321–329.
161. Kukreja SC, Shevrin DH, Wimbiscus SA, et al. Antibodies to parathyroid hormone-related protein lower serum calcium in athymic mouse models of malignancy-associated hypercalcemia due to human tumors. *J Clin Invest*. 1988;82(5):1798–1802.
162. Mirza AM, Choppin A, Bedinger D, et al. Abstract LB-306: impacting Humoral Hypercalcemia of Malignancy (HHM) and associated PTH1R-mediated morbidities: characterization of an anti-PTH1R antagonist monoclonal antibody to reverse hypercalcemia. *Cancer Res*. 2017;77(13 Supplement):LB-306. LB-306.

56

Surgical Management of Hyperparathyroidism

Neeta J. Erinjeri, Pascual De Santis, and Robert Udelsman

OUTLINE

Primary Hyperparathyroidism, 955
- Clinical Presentation, 955
- Indications for Surgery, 955
- Preoperative Preparation, 957
- Operative Approaches, 960
- Intraoperative Adjuncts, 962
- Postoperative Management, 963
- Outcomes of Parathyroid Surgery, 963

Special Circumstances in the Surgical Management of Primary Hyperparathyroidism, 964
- Reoperative Parathyroid Surgery, 964
- Hypercalcemic Crisis, 966
- Inherited Syndromes, 966
- Parathyroid Carcinoma, 967
- Pediatric Parathyroid Surgery, 967
- Hyperparathyroidism in Pregnancy, 968

Secondary Hyperparathyroidism, 968

Tertiary Hyperparathyroidism, 969

COVID-19 and Parathyroid Surgery, 970

Summary, 970

PRIMARY HYPERPARATHYROIDISM

Primary hyperparathyroidism (PHPT) is demonstrated by a biochemical diagnosis resulting from excess synthesis and secretion of parathyroid hormone (PTH) owing to a loss of normal feedback regulation of the PTH receptor. Historically, it has been characterized by hypercalcemia in conjunction with elevated, or inappropriately normal, serum PTH levels in the setting of normal renal function. Recently, the normocalcemic PHPT phenotype has been recognized and categorized.[1] Causes of secondary hyperparathyroidism (SHPT) and tertiary hyperparathyroidism (THPT) must be excluded to unequivocally establish the diagnosis.

KEY POINTS: Primary Hyperparathyroidism

- Primary hyperparathyroidism requires biochemical diagnosis based on serum calcium and parathyroid hormone levels and absence of secondary hyperparathyroidism.
- Ideal management is a team approach (medical and surgical endocrinologists, imaging specialists).
- Surgery is the only curative treatment option.
- All symptomatic patients should be referred for surgical intervention.
- Asymptomatic patients need careful evaluation, as a subset will have indications for surgery.

Clinical Presentation

Before the development of automated laboratory equipment allowing for inexpensive and rapid measurement of serum calcium, the diagnosis of PHPT was rare. Generally, patients were evaluated after presenting with significant signs of advanced disease such as nephrolithiasis, peptic ulcers, or bone disease. With the development of multichannel serum chemical analyzers in the 1960s, measurement of serum calcium became routine, and by the 1970s PHPT was most often found because of incidental identification of hypercalcemia, particularly in the United States and Europe.[2,3] The overall prevalence in the United States is estimated at approximately 0.86%. It affects women significantly more than men, with an incidence ratio of 4:1 and peak incidence in the early postmenopausal period.[4]

Although the vast majority of patients in the United States do not present with the classic symptoms and signs, PHPT is associated with a multisystem clinical syndrome manifesting in the skeletal, renal, gastrointestinal, neurologic, and psychiatric systems. The previous hallmarks of the disease are now much less common, with osteitis fibrosis cystica occurring in less than 2% and nephrolithiasis in less than 20% of cases.[5] Once PHPT is diagnosed biochemically and causes of SHPT/THPT have been ruled out, providers must determine the appropriate next steps in management.

Indications for Surgery

Because surgery is the only curative treatment for PHPT, parathyroidectomy was previously identified as the standard of care for all patients presenting with PHPT. This is still the case for patients presenting with significant symptoms and overt signs of disease, such as fragility fractures, nephrolithiasis, peptic ulcer disease, or significant neurocognitive deficits. However, over the last four decades, with the change in severity of the disease at the time of diagnosis, questions have arisen regarding whether all patients need surgical intervention, especially in the growing cohort with asymptomatic PHPT. As such, in 1990, 2002, 2008, and 2013 international multidisciplinary expert workshops were assembled to discuss this issue, evaluate the relevant literature, and develop guidelines for the diagnosis, recognition, and management of this disease.

Following the fourth international workshop in Florence in 2013, updated guidelines were published. In these, Bilezikian et al.[6] indicate that all patients with confirmed PHPT and traditional signs and symptoms should undergo parathyroidectomy. In asymptomatic patients, defined as those with biochemically confirmed PHPT who lack overt signs and symptoms, guidelines for surgical recommendation are based on markers of more severe disease or disease more likely to progress. Observation is an acceptable option in those who do not meet these criteria, as long as they are able to comply.[6] Subsequently, in 2016, the American Association of Endocrine Surgeons convened a multidisciplinary panel and, using the American College of Physicians grading system, published consensus recommendations based on the literature.[7] Similarly, parathyroidectomy was recommended for all symptomatic patients. A summary of common symptoms and signs is detailed in Table 56.1.[7,8] Surgery is also recommended for those with evidence of severe disease, those at risk for progression, and those who are unreliable, unable, or unwilling to comply with surveillance. A summary of surgical recommendations for asymptomatic patients can be found in Table 56.2.[7]

Despite these consensus guidelines and straightforward recommendations, interpretation is more nuanced. Primarily, the issue lies in distinguishing between symptomatic and asymptomatic PHPT in those not presenting with obvious classic symptoms and signs. There are an array of nonspecific, subtle, common symptoms that are also associated with PHPT, including fatigue, exhaustion, weakness, polydipsia, polyuria, nocturia, constipation, bone pain, back pain, joint pain, depression, memory loss, loss of appetite, nausea, heartburn, and pruritus.[8] These symptoms, while difficult to quantify and unequivocally attribute to PHPT, affect the perception of quality of life. The true determination of causality and severity is only brought to light post parathyroidectomy.[9-15] Accordingly, those recommending surgical intervention must delve into the history to determine whether a patient is symptomatic. Additionally, although the guidelines suggest that asymptomatic patients that do not meet criteria for severe disease may be observed without surgery, they also state that surgery is always an acceptable option. Patients and providers should be given the opportunity to discuss the potential benefits of intervention, as well as the costs, success rates, and complication rates associated with surgery.

The natural history of PHPT has been well studied. Even in those with asymptomatic PHPT, although it occurs slowly, 25% to 33% of patients will have disease progression and eventually meet criteria for surgery, and some of the more serious consequences of unchecked disease, including bone loss, renal dysfunction, and cardiovascular risk, are preventable with early intervention. Additionally, prospective randomized controlled trials have shown that symptoms, specifically neurocognitive function and quality of life, improve following parathyroidectomy.[3,12-14,16,17] Accordingly, patients with a clear diagnosis of PHPT should be thoroughly evaluated and given the opportunity to discuss surgical and medical management with an experienced parathyroid surgeon, including the risks and benefits of intervention. Patients who have symptoms or who are asymptomatic and meet surgical criteria should receive

TABLE 56.1 Common Symptoms and Signs of Primary Hyperparathyroidism

System	Sign/Symptom
Renal	Symptomatic nephrolithiasis
	Recurrent nephrolithiasis
	Polyuria
	Nocturia
	Polydipsia
Skeletal	Osteitis fibrosa cystica (rare)
	Long bone pain
Gastrointestinal	Gastroesophageal reflux
	Acute pancreatitis with severe hypercalcemia
	Constipation
	Loss of appetite
	Nausea
Neuromuscular	Proximal muscle weakness (rare)
	Muscular atrophy (rare)
	Easy muscle fatigue
	Generalized weakness
Neuropsychiatric/cognitive	Mental status changes (rare)
	Obtundation (rare)
	Frank psychosis (rare)
	Progressive fatigue
	Exhaustion
	Depression
	Anxiety
	Emotional lability
	Sleep disturbances
	Worsening memory/memory loss
	Inability to concentrate

(From Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons Guidelines for Definitive Management of Primary Hyperparathyroidism. *JAMA Surg.* 2016;151:959–968; and Stephen AE, Mannstadt M, Hodin RA. Indications for surgical management of hyperparathyroidism: a review. *JAMA Surg.* 2017;152:878–882.)

TABLE 56.2 Guidelines for Parathyroidectomy in Asymptomatic Patients

Measure	Surgical Indication[a]
Age	Less than 50 years
Serum Calcium	>1 mg/dL (>0.25 mmol/L) above upper limit of normal
Skeletal	a. T-score of ≤–2.5 (osteoporosis)
	b. Fragility fracture
	c. Vertebral fracture by imaging
Renal	a. Creatinine clearance <60 mL/min
	b. 24-hr urine calcium >400 mg/dL (>10mmol/dL) AND increased stone risk by biochemical stone risk analysis
	c. Imaging evidence of nephrolithiasis – x-ray, ultrasound, or computed tomography
	d. Nephrocalcinosis on imaging

[a] Surgery also indicated in patients unwilling or unable to comply with medical surveillance.

(Data from Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons Guidelines for Definitive Management of Primary Hyperparathyroidism. *JAMA Surg.* 2016;151(10):959–968.)

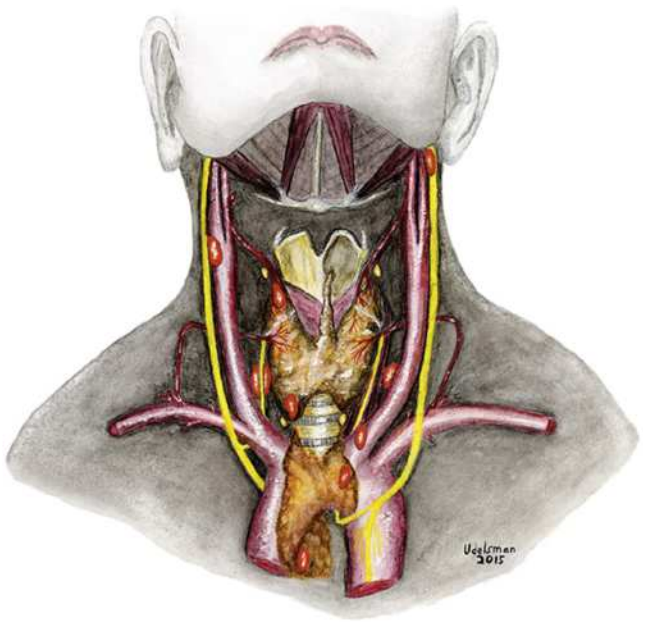


Fig. 56.1 Anatomical relationships of eutopic and ectopic parathyroid glands. (© Robert Udelsman.)

a strong recommendation for surgery. Patients who are asymptomatic and do not meet guideline criteria should be informed that surgery is an option that may offer benefit when compared with observation.

Preoperative Preparation

Diagnosis. One should consider referral of all patients with hyperparathyroidism (HPT) to an endocrinologist before surgical intervention. From the surgical perspective, the first step in preoperative preparation of a patient with PHPT is confirmation of the diagnosis. Although most endocrinologists perform exquisite evaluations confirming the diagnosis, less experienced practitioners can miss subtle confounding issues, including mild forms of SHPT or familial forms of HPT. Because this is a biochemical diagnosis, the surgeon needs to review the levels of serum calcium and intact PTH and assess renal function. They must also evaluate the patient's serum creatinine, 25-hydroxy-vitamin D, and 24-hour urine collection of calcium. Baseline bone mineral density examination is also beneficial, as the values can affect recommendations for surgery. Based on review of these values, combined with a detailed history, surgeons will make a recommendation regarding intervention.

Parathyroid Imaging. Parathyroid imaging studies, which can be broadly classified into noninvasive and invasive, are designed to provide surgeons with a map to assist exploration. They are not diagnostic studies. They provide an anatomic landscape to locate enlarged parathyroid glands in eutopic and ectopic positions. Some of these common locations are illustrated in Fig. 56.1.[7] Historically, the surgical approach for parathyroidectomy was bilateral cervical exploration with examination of all glands. However, PHPT is caused by a single parathyroid adenoma in approximately 85% of patients, and resection of this gland alone results in a sustainable cure. The remaining patients have multigland disease, generally four-gland hyperplasia (10%–15%) or double adenomas (3%–5%) and require resection of more than one gland to achieve operative cure.[3,18] The purpose of localization studies is to identify a target adenoma in those with single-gland disease, allowing the surgeon to perform a focused exploration, or to suggest the presence of multigland disease.

Noninvasive preoperative localization studies include cervical ultrasound, sestamibi-technetium-99m scintigraphy, usually with single-photon emission computed tomography (SPECT), four-dimensional parathyroid computed tomography (4D CT), and magnetic resonance imaging (MRI). In a patient who has not had prior anterior cervical exploration, localization studies are used to determine whether the patient is a candidate for focused parathyroidectomy or if bilateral neck exploration should be planned from the outset (Fig. 56.2). The aforementioned imaging techniques have advantages and disadvantages, and the selection of a study is dependent on the overall clinical picture, local availability of quality equipment, and experience of those interpreting the studies. The most commonly used protocols are either ultrasound, often with a corresponding sestamibi-SPECT, or ultrasound with 4D CT. Invasive preoperative localization studies include image-guided fine-needle aspiration of an identified lesion for cytologic evaluation and PTH measurement and parathyroid angiography with selective venous sampling and are reserved for patients in whom surgical intervention is required in a reoperative field. Furthermore, in the reoperative case, noninvasive studies are often adequate, and invasive studies can be obviated. As such, invasive localization studies will be reviewed in a later section when discussing reoperative cases.

KEY POINTS: Principles of Parathyroid Imaging

- Parathyroid imaging is not a diagnostic procedure.
- All imaging studies are not created equally: quality varies dramatically in different centers.
- The purpose of imaging is surgical planning, and imaging should only be done if surgery is indicated/planned.
- Negative imaging is not a contraindication for surgery.
- Non-Invasive imaging generally consists of ultrasound, sestamibi +/- single-photon emission computed tomography, four-dimensional parathyroid computed tomography, or magnetic resonance imaging.
- Invasive imaging, including image-guided aspiration and parathyroid angiography with selective venous sampling, is reserved for reoperative cases.

Ultrasound. Ultrasound is noninvasive and inexpensive, avoids ionizing radiation exposure, and has the benefit of screening for concomitant thyroid disease that could be addressed simultaneously during surgery. It is highly dependent on the skills of both the technician and the interpreter, with a true-positive rate ranging between 50% and 75%, which improves when combined with sestamibi or 4D CT to around 90%.[19-21] It is more accurate with larger glands and in the setting of adenoma rather than hyperplasia. Typically, a positive ultrasound demonstrates an enlarged parathyroid as a hypoechoic homogeneous extrathyroidal mass, often with a polar feeding blood vessel (Fig. 56.3). Normal parathyroid glands are unlikely to be visualized on ultrasound because of their small size. However, enlarged glands can also be missed, especially in the setting of mild disease, multigland disease, obese patients, mediastinal glands, or posterior glands. Although all imaging techniques are affected by the quality of the equipment, this technique relies heavily on the experience and proficiency of the ultrasonographer.

Sestamibi–Single-Photon Emission Computed Tomography. Sestamibi-SPECT is the most commonly used study for localization of parathyroid disease. Sestamibi-technitium-99m scans are dependent on mitochondrial uptake in parathyroid tissue (Fig. 56.4). Uptake is also noted in salivary glands, lymph nodes, and thyroid nodules, which

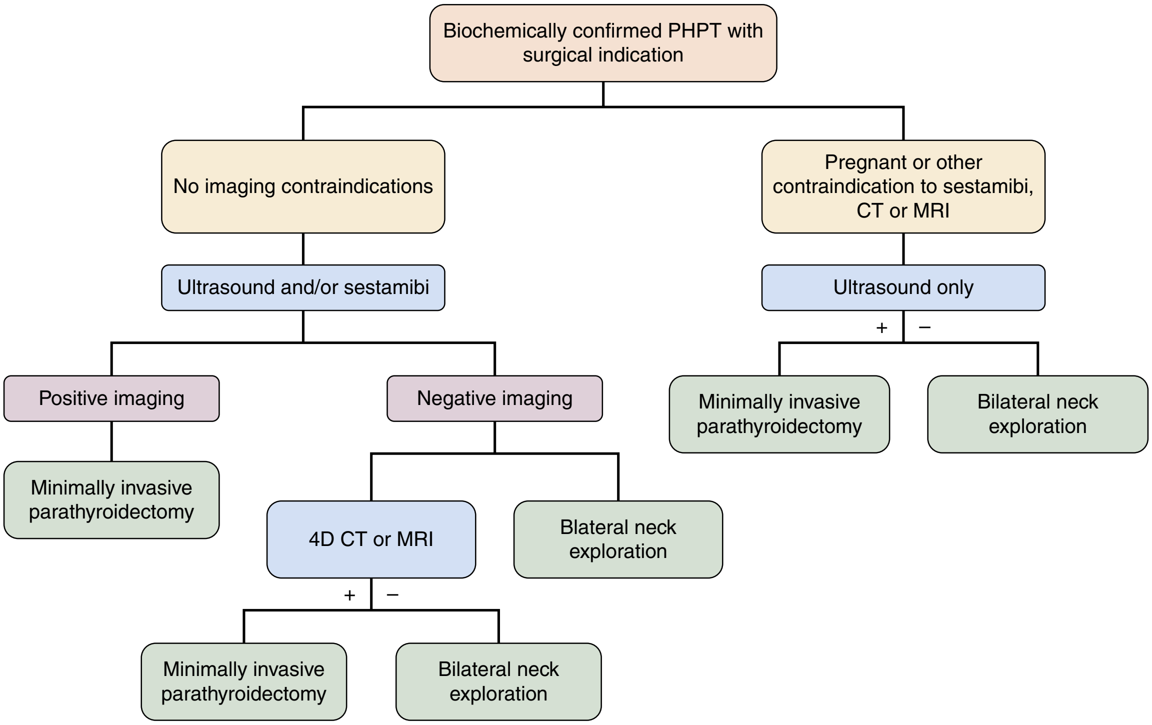

Biochemically confirmed PHPT with surgical indication
No imaging contraindications
Pregnant or other contraindication to sestamibi, CT or MRI
Ultrasound and/or sestamibi
Ultrasound only
+
−
Positive imaging
Negative imaging
Minimally invasive parathyroidectomy
Bilateral neck exploration
Minimally invasive parathyroidectomy
4D CT or MRI
Blateral neck exploration
+
−
Minimally invasive parathyroidectomy
Bilateral neck exploration

Fig. 56.2 Imaging algorithm for parathyroid localization in patients without prior parathyroid or thyroid surgery. *CT*, Computed tomography; *4D*, four-dimensional; *MRI*, magnetic resonance imaging; *PHPT*, primary hyperparathyroidism.

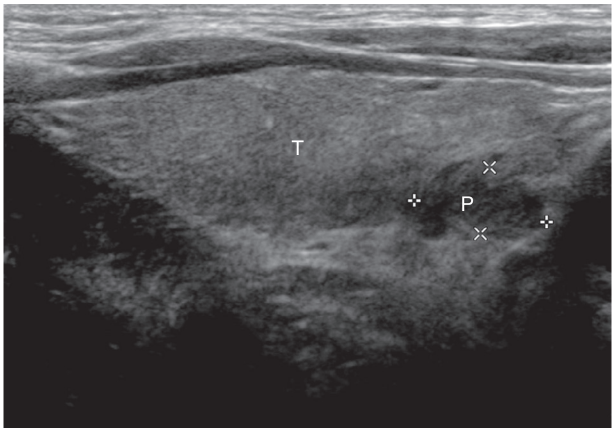


Fig. 56.3 An enlarged parathyroid gland appears as a hypoechoic homogeneous extrathyroidal mass by ultrasound. *P*, Parathyroid adenoma; *T*, thyroid lobe.

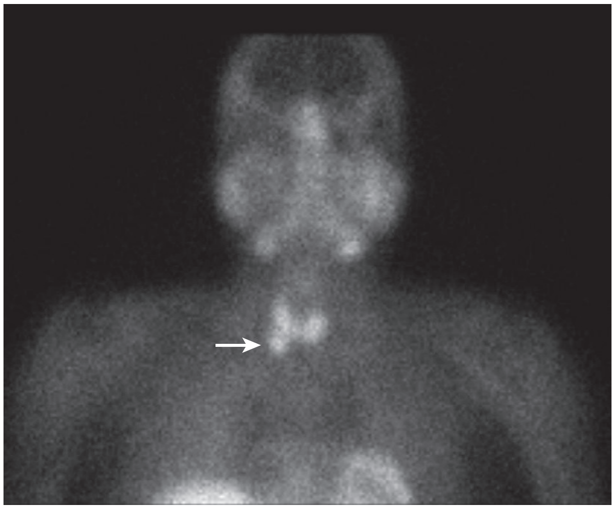


Fig. 56.4 Sestamibi–single-photon emission computed tomography demonstrates uptake consistent with a parathyroid adenoma (*arrow*).

can lead to false-positive studies. When performed in conjunction with a low-dose CT, the images can be fused for improved localization with fewer false-positive findings (Fig. 56.5). Reported sensitivities range widely, generally between 80% and 90%.[22] Sestamibi-SPECT allows for improved visualization of ectopic glands compared with ultrasound, specifically of the mediastinal and posterior glands. Similar to ultrasound, this technique has limited ability to identify multigland disease and small adenomas.[22-24]

Four-Dimensional Computed Tomography. 4D CT uses multiple runs and intravenous (IV) contrast to add the dimension of time to three-dimensional cross-sectional CT imaging. By scanning before and sequentially following contrast administration, structures in the neck and mediastinum are imaged and evaluated for the rapidity of uptake and washout of contrast, and thin cuts provide detailed images with anatomic and functional information (Fig. 56.6). Relying on their vascularity, hyperfunctioning parathyroid glands have low attenuation precontrast, with rapid uptake and washout of contrast in the arterial and delayed images, respectively, compared with surrounding structures and normal glands.[25] The sensitivity of 4D CT is better than that of sestamibi and ultrasound (88% vs. 65% and 57%, respectively) and

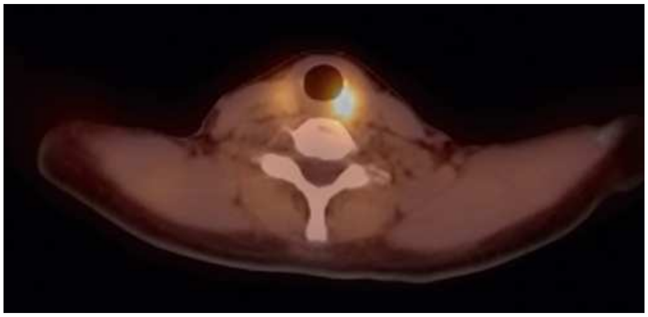


Fig. 56.5 Fused single-photon emission computed tomography/computed tomography demonstrates uptake consistent with a parathyroid adenoma.

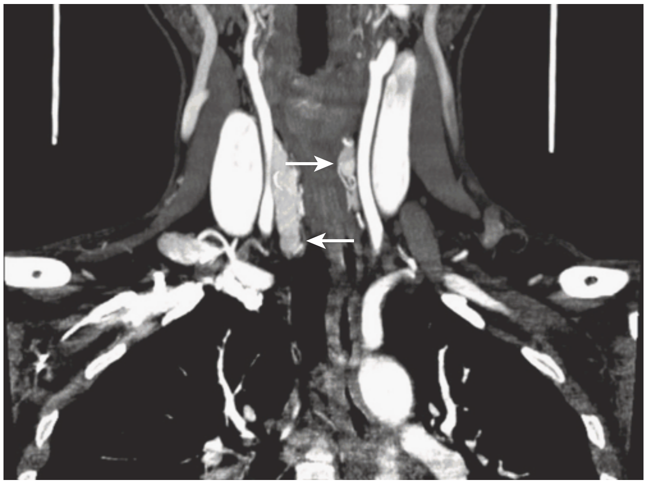


Fig. 56.6 Arterially-enhancing parathyroid adenomas on four-dimensional computed tomography (*arrows*).

predicts multigland disease better than sestamibi (85% vs. 25%–45%,).[25-27] Contrary to ultrasound and sestamibi, 4D CT exposes the patient to ionizing radiation and requires iodinated contrast administration. Accordingly, its use is minimized in certain populations, such as patients under the age of 30 years to reduce the exposure of the thyroid gland to ionizing radiation, those with renal insufficiency to avoid the contrast load to compromised kidneys, those with documented contrast allergy, or those with differentiated thyroid carcinoma with the anticipated need for postoperative radioactive iodine treatment in whom the iodinated contrast load would delay this treatment.[28]

Magnetic Resonance Imaging. MRI images demonstrate parathyroid adenomas as isointense structures relative to the surrounding muscle, with low intensity on T1-weighted images, intense contrast enhancement, and hyperintensity on T2-weighted images (Fig. 56.7). Although some prefer this method because of the lack of ionizing radiation, with the development of 4D CT it is not as commonly used, except in pregnant patients. Its sensitivity is reported to be 69% to 88%.[25]

Imaging is helpful for operative planning; however, negative imaging should not affect the decision for initial surgery because the biochemical diagnosis and surgical indications are not dependent upon imaging. Imaging is used for surgical preparation, strategy, and deciding between focused versus bilateral exploration. With or without positive imaging, the surgeon must be prepared for bilateral exploration and evaluation of all four glands in all cases. In the setting of negative preoperative imaging, the incidence of multigland disease occurs in as many as 37% to 50% of cases, although the glands are likely to be asymmetrically mildly enlarged.[29,30]

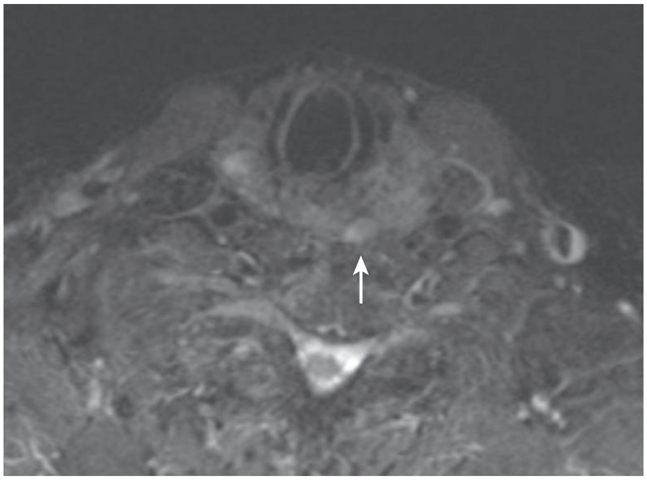


Fig. 56.7 Parathyroid adenomas appear bright on T2-weighted magnetic resonance imaging (*arrow*).

KEY POINTS: Imaging Shortcomings

- Imaging quality is highly variable because of different protocols, devices, and technicians.
- Ultrasound is particularly user dependent.
- All imaging studies are less sensitive in multigland disease.
- Reported imaging sensitivities and specificities rarely include operative and pathologic findings with biochemical confirmation of cure, thus likely are inaccurate estimates.

Parathyroid Development. The embryological development and migration of the parathyroid glands play a role in surgical planning. The inferior parathyroid glands are derived from the endoderm of the third pharyngeal pouch, developing with the thymus. During gestation, these glands descend inferomedially, following the path of the thymus, until they reach their final location, typically posterior to the inferior aspect of the thyroid gland, anterior to the recurrent laryngeal nerve (RLN). The superior parathyroid glands are derived in the fourth pharyngeal pouch with the c-cells of the thyroid and descend during gestation, attaching to the posterior surface of the thyroid gland, posterior to the RLN. Because of the gestational migration of these glands, their final positions can be quite varied, and may be found anywhere along the track from the angle of the mandible to the mediastinum, as shown in Fig. 56.1. As they travel a farther distance, the inferior parathyroid glands are more variable in location than the superior glands.

Knowledge of this embryologic migration and resultant anatomical variation is quite important in surgical planning and during exploration. Superior and inferior parathyroid glands identified during surgery are classified as such based on their relationship to the RLN. A parathyroid gland identified posterior and lateral to the nerve is a superior gland, and a gland identified anterior and medial to the nerve is an inferior parathyroid gland, regardless of which is more caudad or cephalad. Ectopic superior glands can be found in the paraesophageal or retroesophageal space, within the thyroid gland, in the carotid sheath, or in the posterior mediastinum. Ectopic inferior parathyroid glands can be found within the thymus, perithymic fat, thyrothymic ligament, thyroid gland, anterior mediastinum, or carotid sheath.

Preoperative Vocal Cord Evaluation. In the assessment of the patient, one additional factor for the surgeon to consider, is vocal cord mobility. RLN palsy, either temporary or permanent, is a rare but potential complication of parathyroidectomy, and can result in significant morbidity. There is no clear consensus regarding routine preoperative assessment to evaluate vocal cord mobility and identify potential asymptomatic palsies or other issues that impair motion in the previously unexplored patient with a normal voice.[31] In contrast, all reoperative patients, as well as those with complaints of hoarseness or voice changes, should undergo preoperative assessment of vocal cord function.[31,32] This practice does not change or affect the rate of injury; however, it does provide critical information to the surgeon.

Indirect mirror laryngoscopy or direct laryngoscopy, generally with a flexible fiberoptic nasopharyngoscope, allows for visualization of the vocal cords and evaluation of their movement with phonation, breathing, and coughing.[33] With this, the surgeon is able to assess for bilateral movement and symmetry of motion. Alternatively, transcutaneous ultrasound can be used to view the larynx and the vocal cords. This technique also allows for real-time cord evaluation, as well as assessment for other abnormalities; however, it is less reliable, especially in older patients, and other subtle abnormalities can be missed.[33] CT and MRI images can be examined for vocal cord symmetry and size and be used as an indication of function, but, without real-time input, are less reliable and more difficult to interpret. Although occasionally useful, it is unlikely that imaging-based evaluation of the vocal cords will ever replace the gold standard of direct or indirect laryngoscopy.

Operative Approaches

Intraoperative Parathyroid Hormone Assay. Understanding and interpreting PTH metabolism and levels are critical in the evaluation of this disease. Although it is not always used or available, one cannot discuss parathyroid surgery without reviewing intraoperative PTH measurement, as this is a major factor in determination of surgical approach. When considering use of an intraoperative PTH assay, the surgeon needs to be adept at interpretation of the results. Additionally, there are other factors that need to be considered for ideal utilization. One needs a rapid assay platform for evaluation of intact PTH. Ideally, a turnover time of less than 15 minutes from blood draw to result allows for optimal surgical decision-making. The assay should be located in proximity to the operating room, preferably close enough for swift transfer of the blood specimen from the patient to the machine for analysis. Proper interpretation requires serial values, so the assay should have the ability to run multiple samples simultaneously. The surgeon must also know the causes of falsely elevated and low values, as well as how to interpret and troubleshoot should errors occur. The instruments are delicate and require maintenance and standard curve calibration, and a properly trained technician is required to centrifuge the specimen and run the assay. Thus, in areas with resource scarcity, this may not be easily available or implementable.

Intraoperative PTH measurement requires user interpretation, which can be complex and requires experience. A baseline PTH level is obtained immediately before surgery, run on the same assay. Generally, two criteria are used to indicate biochemical surgical cure: (1) the PTH level should fall from preoperative baseline by more than 50%, 10 to 15 minutes following resection of the target gland, and (2) the post-resection fall should result in a PTH level in the normal range. Additionally, the rate of change

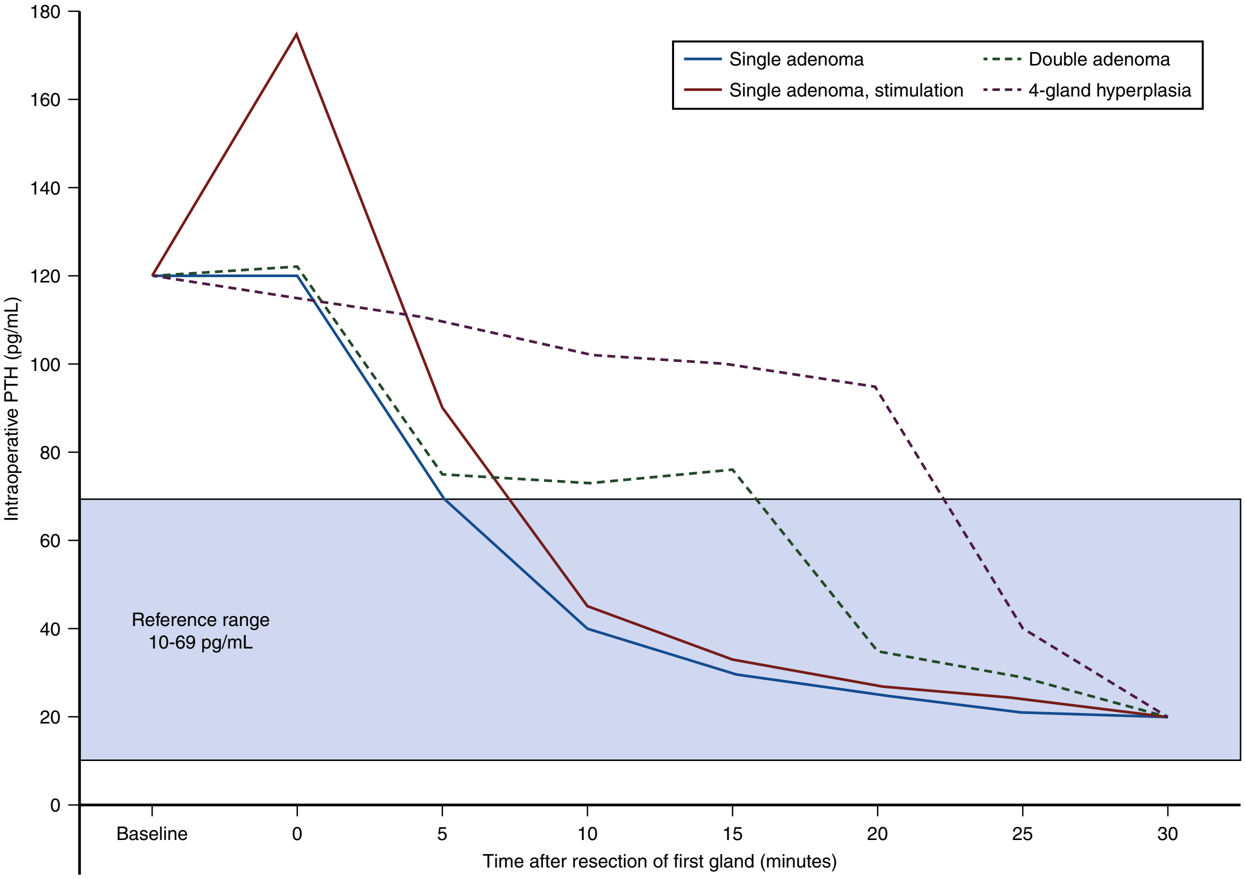


Fig. 56.8 Graphical representation of intraoperative parathyroid hormone (*PTH*) measurements.

and slope of the curve identified when plotting sequential intraoperative PTH levels gives significantly more information than a single value. In single-gland disease, resection of the culprit lesion generally results in a quick initial decline that slows or plateaus near the low end of normal. If the levels remain elevated or plateau at the upper end of normal, this generally indicates residual disease, and additional exploration and resection are necessary (Fig. 56.8). A computerized mathematical model was developed to use intraoperative PTH measures in individual cases to predict the chance of cure. This model could be used as a tool to aid the interpretation of these measurements, specifically in difficult cases.[34]

KEY POINTS: Intraoperative Parathyroid Hormone Assay

- Intraoperative parathyroid hormone (PTH) assay allows for focal exploration, dramatically altering surgical management.
- Surgeons must be adept at interpreting results, including false positives and false negatives .
- A rapid platform is optimal for intraoperative decision-making, ideally with results in <15 minutes.
- Assay should preferably be in proximity to the operating room to facilitate quick transfer of specimens.
- Interpretation is best with serial values, and the assay should be able to run simultaneous samples.
- Intraoperative PTH measurement showing a drop from immediate preoperative baseline by >50%, 10–15 minutes following target gland resection and into the normal range indicates intraoperative cure.

Operating Room Preparation. Before surgery, regardless of the planned approach, positioning is critical. The patient is placed in a comfortable semi-Fowler position with a thyroid pillow under the shoulder blades to allow for neck extension.[35] The use of an inflatable pillow allows for precise control over the pillow size to avoid stress on or overextension of the neck. Patients with cervical spine disease may require neutral positioning. A head ring is placed under the patient's occiput to stabilize the position during surgery. The arms are tucked to provide room for the surgical team to work.[36] Generally, a single large-bore peripheral IV can be used to administer fluids and medications intraoperatively, as well as to draw blood samples for intraoperative measurements of PTH. It is imperative to draw back blood diluted by IV fluids or medications before obtaining blood samples during intraoperative PTH measurements because this can otherwise artificially lower measured circulating PTH levels.[34]

Bilateral Cervical Exploration. A bilateral cervical exploration under general anesthesia is the traditional approach, with a rate of cure of greater than 95% and a low risk of complications when performed by experienced surgeons.[7] It is generally the planned approach in the setting of negative imaging or when family history or imaging studies suggest multigland disease. With this technique, the operator identifies and examines all parathyroid glands, then resects enlarged glands. Even when planned, bilateral exploration is not always necessary if single-gland disease can be confirmed. If, on exploration of the first side, the surgeon identifies one normal and one enlarged parathyroid gland, the enlarged gland can be resected and a rapid intraoperative PTH assay used to measure post-resection levels. If the PTH levels drop appropriately, indicating biochemical cure, no further exploration is required. Alternatively, if the post-resection levels do not show a sufficient decrease in circulating PTH, exploration should continue with further excision as indicated. In the case of four-gland hyperplasia, a subtotal parathyroidectomy is performed, where a sufficient volume (~30–50 mg) of one viable parathyroid gland is left at a single site *in situ*.[37] Although theoretically unnecessary in the setting of formal bilateral exploration and resection for four-gland hyperplasia, intraoperative PTH measurement is still useful to confirm sufficient debulking and adequate remnant function. A PTH drop into the range of 20 to 30 pg/mL is ideal in this circumstance.[37]

Minimally Invasive Parathyroidectomy. Given the prevalence of single-gland disease in PHPT, many prefer a focused approach with minimally invasive parathyroidectomy (MIP). In this technique, a unilateral cervical exploration is planned based on preoperative identification of a target lesion, presumed to be the incident parathyroid adenoma. This allows the surgeon to perform a limited exploration, identify the target gland, and resect the target gland only. With this approach, intraoperative PTH measurements are used to confirm single-gland disease and adequate resection and to demonstrate intraoperative biochemical cure. If these results do not reflect cure, additional exploration is required.[36,38] Some advocate for avoiding the use of intraoperative PTH monitoring in cases with good preoperative localization; however, studies show that, while cure is achieved at high rates in these cases, they may require additional surgery for multigland disease that was not identified with the initial exploration.[39-41]

The majority of MIPs throughout the world are performed under general anesthesia; however, in the proper setting, they can be performed on awake patients with regional anesthesia. Well-localized initial parathyroidectomies and selective reoperative parathyroidectomies can be performed by experienced teams without general anesthesia. Success with this approach requires an experienced anesthesia team to titrate sedation and a surgeon or anesthesiologist capable of administering a superficial cervical nerve block. Additionally, the surgeon will also often perform a field block. The positioning and preparation of the patient are the same as described above, with the addition of protective eyewear, and an ether screen is used to elevate the drapes off of the patient's face. Oxygen is administered, preferably by nasal cannula, and suction tubing is positioned to scavenge excess oxygen from the operative field and minimize fire hazard.[36,38] This set-up is depicted in Fig. 56.9.[35]

Endoscopic Approaches to Parathyroidectomy. As with many other surgical procedures, as technology has advanced, so has the drive to create further minimally invasive approaches. In 1996 an endoscopic approach to parathyroidectomy was described using video assistance and a laparoscope through a 1.5-cm incision in the neck.[42] This endoscopic technique for parathyroid and thyroid surgery evolved to permit remote access and avoid a neck incision altogether to improve cosmetic results, occasionally utilizing robotic assistance, including transaxillary, transbreast, combined axillary and breast, or retroauricular approaches. These techniques, while avoiding the visible cervical incision, imposed incremental risks compared with conventional surgery, and, because of the large flaps and extensive dissection required, are actually more invasive than conventional parathyroidectomy.[43] More recently, a transoral technique was developed that maintains the cosmetic benefits of remote access but is less invasive. This approach uses incisions in the oral vestibule of the lower lip and traverses the premandibular space to access the subplatysmal space opened in conventional parathyroidectomy.[43] These robotic and endoscopic approaches are more expensive, time-consuming, and difficult to learn and perform, as well as requiring sophisticated technology, solely for cosmetic benefit. Although they may play a role in select cases at select institutions, these are unlikely to become standard practice, especially as the majority of parathyroidectomies can be performed through an abbreviated Kocher incision measuring 2.5 to 3.5 cm in length. If the surgeon is able to place the incision parallel with a natural skin crease, the incision is often nearly invisible once fully healed.[44,45]

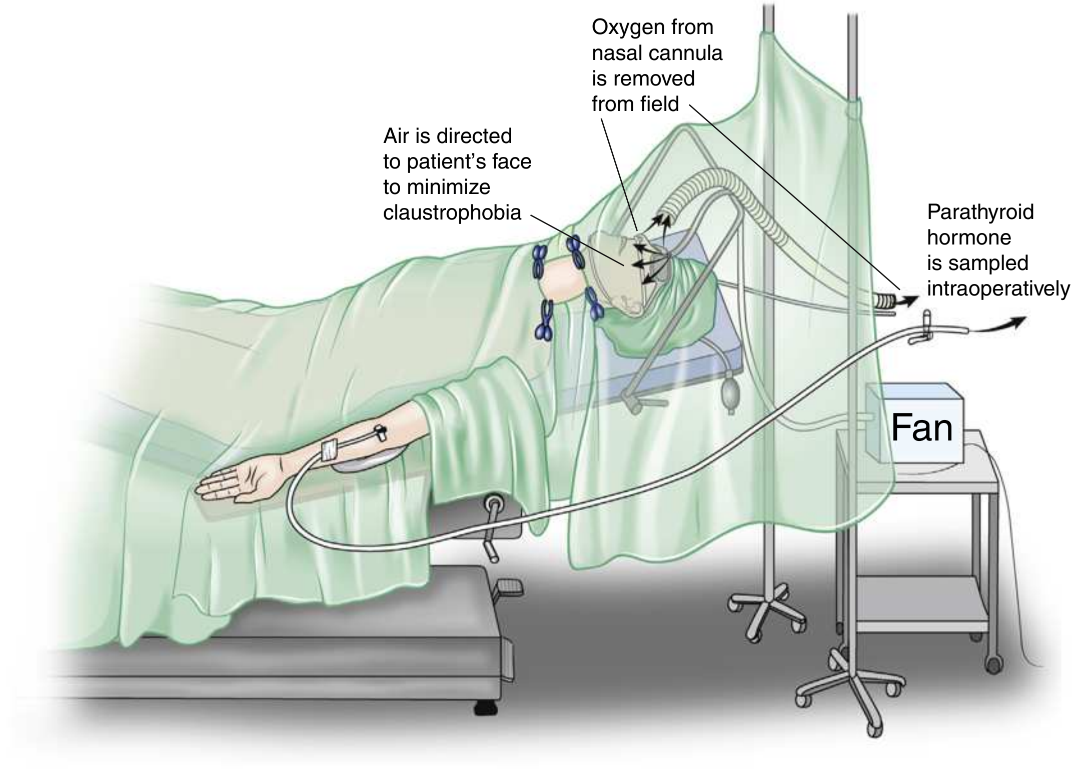

Oxygen from nasal cannula is removed from field
Air is directed to patient's face to minimize claustrophobia
Parathyroid hormone is sampled intraoperatively
Fan

Fig. 56.9 Patient positioning for minimally invasive parathyroidectomy under regional anesthesia. (Reprinted with permission from Carling T, Udelsman R. Minimally invasive parathyroidectomy. In: Oertli D, Udelsman R, eds. *Surgery of the Thyroid and Parathyroid Glands*. Berlin and Heidelberg: Springer; 2012:475–485.)

Intraoperative Adjuncts

Localization. There are several techniques that can be used to aid with identification or confirm intraoperative cure following resection of the target gland. The first is the use of intraoperative PTH measurements, as discussed. Additionally, this can also be used to confirm the tissue resected is actually a parathyroid gland, with *ex vivo* aspiration of the specimen sent for PTH measurement. Parathyroid tissue will have significantly elevated measurements of PTH, dramatically higher than serum levels and often higher than the scale of the assay. Lymph nodes, exophytic thyroid tissue, or other cervical soft tissues can appear similar to parathyroid tissue, and aspiration can confirm the presence or absence of parathyroid tissue in the specimen. Frozen section analysis can also be used to identify parathyroid tissue and can yield additional useful information such as gland weight and cellularity.

Some centers use preoperative sestamibi injection to aid with identification and resection. In these instances, a gamma probe is used intraoperatively to localize and identify the target lesion, then counts are used to measure the resected gland when compared with the background tissue to suggest single-gland disease and intraoperative cure. The excised lesion should contain more than 20% of the postexcision background radioactivity.[46-48] A large–field of view portable gamma camera has also been used in a similar way to show absent activity in the neck immediately following resection to indicate adenoma resection.[49]

Visualization of autofluorescence and indocyanine green is being used with increasing frequency in certain centers. This technique aids with localization and can also be used to assess gland perfusion following subtotal resection. In PHPT, indocyanine green angiography shows moderate to strong increased vascularization.[50] The use of IV methylene blue has been described for colorimetric localization to identify the parathyroid glands.[51]

Some prefer the use of high-resolution ultrasound guidance to aid in identifying target lesions. This is the least invasive imaging modality, does not use ionizing radiation, and can be reproduced intraoperatively, preincision. It is somewhat limited, however, following incision, as air in the tissues disrupts the appearance. Ultrasound has also been coupled with local methylene blue injection to aid with intraoperative identification.[51]

Some optical technologies described to evaluate the viability of parathyroid glands in the setting of thyroid surgery can also be used for localization of hyperfunctioning glands, including imaging with autofluorescence, 50-aminolevulinic acid, optical coherence tomography, laser speckle contrast imaging, dynamic optical contrast imaging, and Raman spectroscopy.[52]

Neuromonitoring. One of the major, although rare, risks of parathyroidectomy is RLN palsy. Intraoperative neurophysiological monitoring is being used increasingly. Neuromonitoring can only be done in procedures performed under general anesthesia with endotracheal intubation using a specialized endotracheal tube, and without the use of long-acting muscle relaxation. This adjunct to visual identification can assist in nerve localization and theoretically could reduce the incidence of trauma to the RLN. It does rely on correct implementation and interpretation of output, and training in analysis is required to prevent false security and increased harm.[53,54]

Postoperative Management

The immediate management and disposition of the postoperative patient are dependent on the extent of surgery performed. If a focused unilateral exploration was sufficient, there is no risk for permanent hypoparathyroidism, and if patients meet criteria, they may be discharged the day of surgery. Prophylactic supplementation with oral calcium is often initiated at this time. Furthermore, it is critical to instruct the patient on the symptoms of hypocalcemia and ensure access to reliable 24-hour communication with the surgical team to foster ease of remote management of potential postoperative hypocalcemia and minimize potentially dangerous hypocalcemic events. Having an experienced surgical team that can routinely follow up with the patient in the immediate postoperative period is imperative.

KEY POINTS: Postoperative Management

- Most patients have a straightforward postoperative course and are discharged home the same day.
- Postoperative events including severe hypocalcemia, bleeding, and airway obstruction can occur.
- A team approach is integral in inpatient and outpatient postoperative care.
- Hypocalcemia is treated with oral calcium and, if severe, intravenous calcium.
- Calcitriol can aid with stabilizing serum calcium levels in the immediate postoperative period.
- Serum calcium/intact PTH levels 6 months postoperatively indicate surgical cure if normocalcemic.

If surgery required bilateral exploration, or if the patient has had previous parathyroidectomy or contralateral cervical exploration, the risk for significant postoperative hypoparathyroidism is higher. In these instances, the patient should be considered for overnight hospital observation and measurement of postoperative serum calcium levels. With the lengthier half-life of calcium compared with PTH, it takes longer for hypocalcemia to manifest following an intraoperative drop in PTH levels. Hypocalcemia while in the hospital should be treated with oral calcium supplementation or, if severe, IV calcium. Calcitriol can also be used, with the goal of attaining a stable serum calcium level before discharge. Those with preoperative vitamin D deficiency, significantly elevated PTH with evidence of bone disease, normal to minimally elevated preoperative serum calcium levels, and an intraoperative drop in PTH levels of greater than 80% in 10 minutes are at increased risk for postoperative hypocalcemia.[55]

The patient is seen postoperatively, generally 1 to 2 weeks following surgery, for evaluation of the incision and measurement of serum calcium and intact PTH. The vast majority of patients demonstrate operative cure. However, those with evidence of persistent PHPT should be followed for consideration of reoperation and for specialized localization studies if surgery is planned. Those with initial results indicating cure should have medical endocrinology follow-up 6 months after surgery, with measurement of serum calcium and intact PTH. Although intraoperative PTH levels are used to demonstrate intraoperative or immediate postoperative cure, surgical cure is defined as the attainment of normocalcemia 6 months following parathyroidectomy. Those with an elevated serum calcium at this time may have persistent PHPT. Alternatively, if these values demonstrate normocalcemia, and patients develop hypercalcemia more than 6 months after surgery, it is considered recurrent disease. Patients with persistent or recurrent disease should be evaluated for reoperative parathyroidectomy, again with a detailed history and physical examination to determine the severity of disease, assessment of indications for surgery, and potential risks, which will be detailed later in this chapter.

A subset of patients following parathyroidectomy will demonstrate an elevated serum PTH level with a normal or low serum calcium level. This is usually attributed to SHPT in the setting of successful treatment of PHPT. Oral calcium supplementation resulting in suppression of PTH levels proves the diagnosis and serves as treatment. Some patients have elevated PTH levels postoperatively that eventually normalize without evidence of SHPT; however, the cause of this is yet unknown. Those in whom oral calcium supplementation does not suppress PTH levels into the normal range likely have early or mild multigland disease not identified at the initial operation.

Outcomes of Parathyroid Surgery

Operative Complications. In general, when performed by experienced high-volume surgeons, parathyroidectomy is a safe and effective surgery, with cure rates exceeding 95%. Major risks of surgery include RLN injury, hypocalcemia, and neck hematoma. Much less common complications included pneumothorax, stroke, and need for tracheostomy from bilateral RLN injury. RLN injury occurs in 0.5% to 1% of cases and usually results in hoarseness. This rate may be underestimated, as postoperative laryngoscopy is not routinely performed in the absence of symptoms. It can be temporary or permanent, depending on the degree of injury, and additional interventions such as vocal cord injection or medialization thyroplasty may be necessary.[44,45]

Some degree of hypocalcemia is common following parathyroidectomy owing to hypoparathyroidism but is usually mild, and temporary, especially if only one side of the neck is explored. It is managed with oral calcium and, occasionally, calcitriol. Permanent hypoparathyroidism with insufficient residual parathyroid function is rare. When it occurs, it results in long-term hypocalcemia in 0.1% of cases. It is managed with oral calcium and calcitriol supplementation.[44,45]

KEY POINTS: Operative Approach to Primary Hyperparathyroidism

- Sporadic primary hyperparathyroidism: ~85% single-gland disease, 10%–15% multigland hyperplasia, 3%–5% double adenomas.
- When localized disease is present, may attempt focal exploration with intraoperative parathyroid hormone (PTH) measurements.
- Four-gland exploration may be necessary for poorly localized cases, if PTH does not drop following focal exploration/resection, or if no access to intraoperative PTH assay/intraoperative adjuncts.
- In four-gland hyperplasia, treatment is subtotal parathyroidectomy, leaving a single-site remnant of ~30–50 mg of viable parathyroid gland .
- Intraoperative adjuncts including gamma probes are used by some to aid with gland identification.
- Endoscopic approaches to parathyroidectomy are done in select centers for cosmetic benefit.
- High-volume centers have operative success rates >95% and complication rates of 1%–2%; however, low-volume centers have poorer outcomes.

Postoperative bleeding resulting in a neck hematoma is a rare but potentially emergent situation. It is estimated to occur in 0.2% of cases and may require immediate return to the operating room for evacuation and hemostasis.[44,45] The airway can be very challenging in these cases, even in the absence of stridor or overt tracheal compression, because the hematoma can cause venous hypertension leading to laryngeal edema, making the airway difficult or impossible to visualize for standard intubation. Occasionally, in the setting of acute airway distress, it is necessary to open the incision at the bedside to protect the airway. In these cases, it is sometimes preferred or necessary to decompress the hematoma before attempting endotracheal intubation. Once the hematoma has been evacuated, the airway crisis is averted, and intubation of the trachea is usually unencumbered. Rarely, it is necessary to perform an awake, upright, fiberoptic intubation by the most experienced person available, given these potential issues.

Operative Cure. Success rates following parathyroidectomy with biochemical cure are usually excellent but are dependent on the experience of the surgeon and the center. Studies at high-volume centers report cure rates of around 99% for minimally invasive patients and 97% for bilateral explorations.[45,56] When discussing these success rates and contemplating surgery, it is important to consider that these studies are based on results of experienced surgeons in centers where the disease is well-understood, and high-quality imaging and intraoperative adjuncts are available. However, most parathyroid surgery is not performed by high-volume endocrine surgeons, and it is likely that outcomes from surgeons who do not perform this type of surgery often are not as favorable.[57]

The Health Care Utilization Project's National Inpatient Sample, maintained by the Agency for Healthcare Research and Quality, represents a stratified 20% sample of inpatient discharges from hospitals across the United States. Using this database, Adam et al. reported that, from 1988 to 2000, only 5% of parathyroid procedures were performed by a surgeon with 75% or greater endocrine case volume, and 78% of these cases were performed by surgeons with endocrine cases composing 25% or less of their practice.[58] Direct comparisons between surgeon volume and outcome following parathyroidectomy vary with regard to surgeon volume stratification, but demonstrate improved outcomes with increased surgeon volume.[57] Additionally, a higher percentage of operative failures occur at low-volume centers, and, based on the reoperative results, a proportion are considered potentially preventable.[57] For example, low-volume centers performed the initial surgery in 78% of cases considered to be avoidable following reoperative parathyroidectomy at a high-volume center.[59]

In a Scandinavian study, Malmaeus et al., using data from 1971 through 1980, directly compared high- and low-volume centers and reported that the cure rate in endocrine surgery centers was 90% versus 76% in general surgery clinics and 70% in centers that performed fewer than 10 parathyroid operations each year.[60] A subsequent Scandinavian quality register report did not categorize data based on surgeon or center volume; however, the overall cure rate was 91.9%, and postoperative hypocalcemia occurred in 11.4% of patients.[61] These results are not nearly as excellent as those reported by high-volume American centers and reflect that a substantial proportion of parathyroid surgeries in Scandinavia are not performed at high-volume centers.

SPECIAL CIRCUMSTANCES IN THE SURGICAL MANAGEMENT OF PRIMARY HYPERPARATHYROIDISM

Reoperative Parathyroid Surgery

Reoperative, remedial, or redo parathyroid surgery is performed in the setting of persistent disease or recurrent disease, or in patients who have undergone prior cervical surgery, such as thyroidectomy or anterior cervical fusion. These cases pose additional challenges and risks and are more complex. As such, it is beneficial to have these performed at a high-volume center by an experienced parathyroid surgeon.[57] These cases must be carefully reviewed to confirm the diagnosis and ensure an appropriate surgical indication. The recommendation for surgery in these cases is more complex, owing to increased operative risk, particularly in those with asymptomatic PHPT. Although these patients are likely to benefit from initial parathyroidectomy, remedial surgery has a higher rate of complications and a lower cure rate, altering the risk/benefit ratio. In general, reoperative parathyroidectomy is reserved for those with more symptomatic disease. Additionally, careful deliberation of the clinical history is warranted, with the consideration of the possibility of syndromic disease and a predisposition to recurrence.

KEY POINTS: Reoperative Parathyroid Surgery

- Patients with primary hyperparathyroidism and history of prior parathyroidectomy, thyroidectomy, or anterior cervical surgery are considered for remedial, reoperative, or redo parathyroid surgery.
- These cases have increased complexity, increased potential morbidity, and decreased rates of success.
- Reoperative cases are best managed at specialized centers.
- Imaging studies begin with noninvasive modalities, but invasive studies may be necessary.
- In high-volume centers, outcomes are still good despite increased difficulty.

The surgeon must carefully review all operative, pathologic, and interim data to have a clear picture of the previous operation(s) performed and subsequent course. Preoperative imaging searching for missed glands is also more challenging because the location of missed parathyroid glands is highly variable, as illustrated in Fig. 56.10.[62] Again, thorough knowledge of embryologic development and migration of the parathyroid glands is of utmost importance in these cases. Initial localization techniques are identical to unexplored cases, using ultrasound, sestamibi-SPECT, 4D CT, or MRI, ideally with two concordant studies before surgery. If these studies are not convincing, discordant, or nonlocalizing, invasive imaging may be necessary. If a lesion is seen on ultrasound, a fine-needle aspiration analyzed by cytologic evaluation and measurement of PTH in the aspirate can be performed to confirm that the identified lesion is a parathyroid gland. If the ultrasound cannot identify a target lesion, or if the lesion is not amenable to aspiration, parathyroid angiography with selective venous sampling can be performed, preferably with on-site rapid PTH measurement.[63] The PTH gradient between sites of venous sampling will regionalize the site of hormone drainage, suggesting the location of residual disease. These localization studies then provide the surgeon with a target for exploration. Reoperative exploration without identification of a target lesion is avoided whenever possible and should only be performed in severely symptomatic patients with failed localization studies, because of the added risk posed to these patients. This process is depicted in Fig. 56.11.

The success rates of reoperative parathyroid surgery are excellent when performed by high-volume surgeons, with cure rates reported to be 94% to 98%.[64-66] The rate of RLN injury in this patient population is higher, estimated at 2.3% based on a consecutive series of 130 reoperative parathyroidectomies.[62] Looking specifically at reoperations for persistent or sporadic adenoma, the rate of permanent RLN injury was 1.8%.[67] In reoperative cases, permanent hypoparathyroidism is of high concern because of the possibility that the target lesion may be the patient's only remaining functional parathyroid tissue. This risk is dependent on the extent of prior surgical exploration, as, even

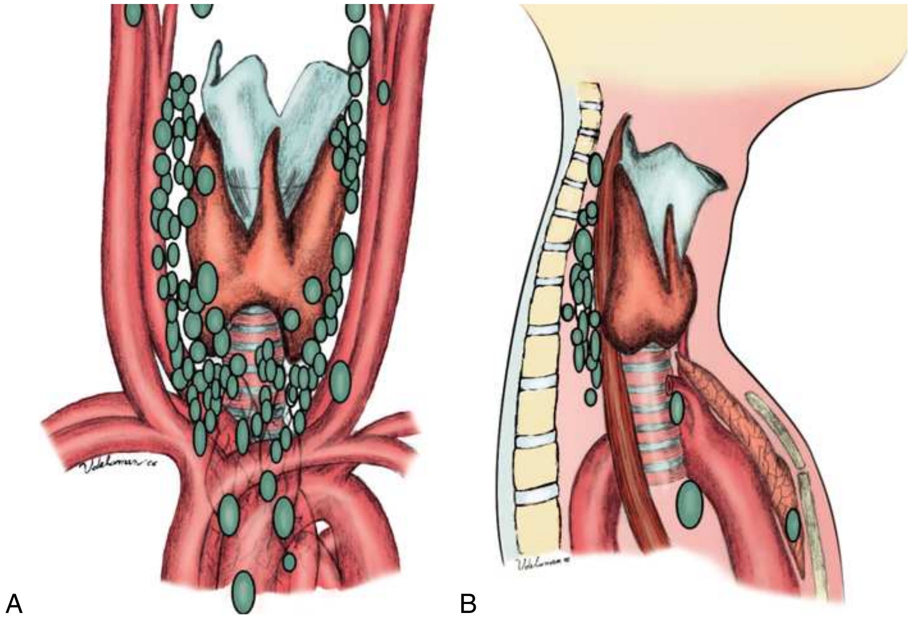
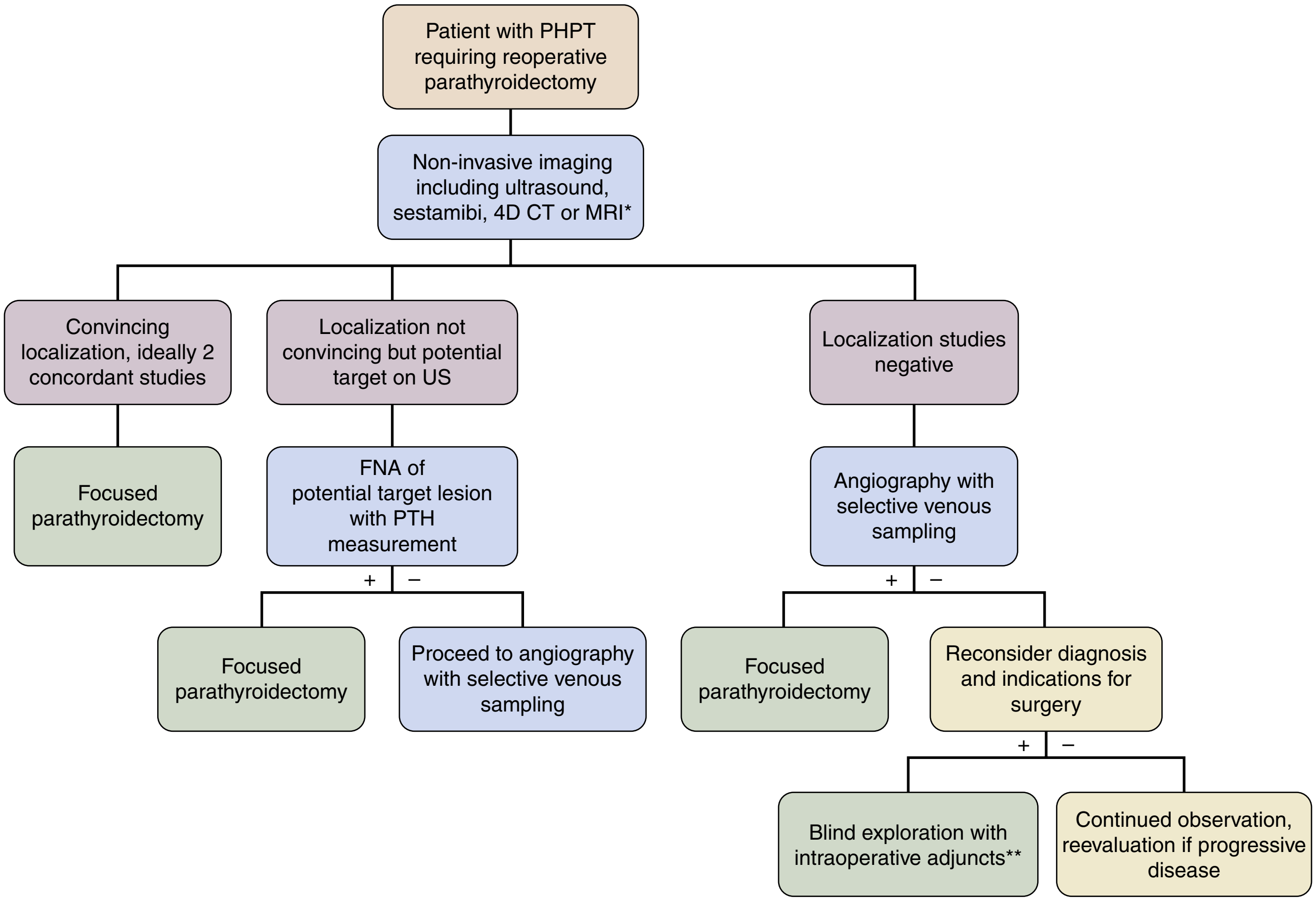


Fig. 56.10 **Locations of parathyroid glands found during reoperative parathyroid surgery. A**, Anterior-posterior view. **B**, Lateral view. (Reprinted with permission from Udelsman R, Donovan PI. Remedial parathyroid surgery: changing trends in 130 consecutive cases. *Ann Surg.* 2006;244:471–479.)


Patient with PHPT requiring reoperative parathyroidectomy
Non-invasive imaging including ultrasound, sestamibi, 4D CT or MRI*
Convincing localization, ideally 2 concordant studies
Localization not convincing but potential target on US
Localization studies negative
Focused parathyroidectomy
FNA of potential target lesion with PTH measurement
Angiography with selective venous sampling
+
−
Focused parathyroidectomy
Proceed to angiography with selective venous sampling
+
−
Focused parathyroidectomy
Reconsider diagnosis and indications for surgery
+
−
Blind exploration with intraoperative adjuncts**
Continued observation, reevaluation if progressive disease

* Local expertise and equipment dictate imaging selection.
** Blind exploration is generally avoided except in the most severe or symptomatic cases.

Fig. 56.11 Imaging algorithm for parathyroid localization in reoperative patients. *4D CT*, Four-dimensional computed tomography; *FNA*, fine-needle aspiration; *MRI*, magnetic resonance imaging; *PHPT*, Primary hyperparathyroidism; *PTH*, parathyroid hormone; *US*, ultrasound.

without identification or resection, it is possible to inadvertently injure or devascularize the end-arterial blood supply to the other parathyroid glands. Dependent on the initial and reoperative course, it may be necessary to perform immediate parathyroid autotransplantation or parathyroid cryopreservation.

Hypercalcemic Crisis

A severe manifestation of PHPT, occurring in 1% to 6% of patients, is hypercalcemic crisis. This life-threatening, progressive event is characterized by severe hypercalcemia associated with dehydration, vomiting, renal failure, mental status deterioration, coma, and death. This is a medical emergency requiring immediate management with hydration and reduction of hypercalcemia.[68-70] Hospitalization is required for close monitoring and hydration with large volumes of IV fluid, occasionally with the addition of furosemide. If the hypercalcemia cannot be adequately improved with initial therapy, it may be necessary to add bisphosphonates, cinacalcet, or calcitonin to the regimen. Surgery is performed once the calcium levels have been appropriately decreased and renal function improves. It is ideal if surgery can be performed early in this course, circumventing the need for long-acting pharmacologic treatments, such as bisphosphonates, as these can complicate postoperative management because of severe hypocalcemia.

Inherited Syndromes

Although accounting for a minority of cases of PHPT, inherited forms occur in several syndromes. This is more likely in young patients who present with apparent sporadic disease, as up to 10% will have an inherited condition. The identification of those with inherited syndromes is quite important, because surgical management of inherited disease differs significantly from that of sporadic cases.[71]

Multiple endocrine neoplasia type 1 (MEN1), an autosomal dominant inherited syndrome resulting from mutations within the *MEN1* gene, is associated with PHPT, anterior pituitary tumors, pancreatic neuroendocrine tumors, foregut carcinoids (thymic, bronchial, or gastric), adrenal adenomas, a variety of skin lesions, and other benign tumors. PHPT is the hallmark of this syndrome, with relentless multigland disease and high penetrance, as over 98% of MEN1 patients have or will develop PHPT.[72] Parathyroidectomy is the principal management in these patients, but it must be done aggressively, being mindful of the age of the patient and likelihood of recurrence, and with the goal of minimizing morbidity, including permanent hypoparathyroidism.

KEY POINTS: Inherited Syndromes

- Although the majority of primary hyperparathyroidism (PHPT) cases are sporadic, there are a number of inherited forms.
- Up to 10% of young patients presenting with apparent sporadic disease have an inherited condition.
- Inherited syndromes should be suspected in patients with known family history of PHPT.
- Those with inherited disease are at increased risk for initial failure and recurrent disease.
- Surgery varies based on inherited disease and is best managed at specialized centers.

The 2020 National Comprehensive Cancer Network guidelines describe two surgical options when treating PHPT associated with MEN1: (1) subtotal parathyroidectomy leaving approximately 30 to 50 mg of parathyroid tissue *in situ*, with possible cryopreservation and possible transcervical thymectomy, or (2) total parathyroidectomy with immediate autotransplantation of parathyroid tissue, generally into the nondominant forearm, possibly with cryopreservation, and possible transcervical thymectomy.[73] There is still controversy as to which of these approaches is best in this population. In MEN1, all parathyroid tissue is abnormal; thus, there is a high risk of recurrent disease. Some studies quote recurrence rates of 30% to 40% with either approach, whereas others found higher rates following subtotal resections. If subtotal parathyroidectomy is performed, leaving a small remnant *in situ*, cervical reoperation would be necessary in the case of recurrence. However, following total parathyroidectomy with autotransplantation as many as one-third of patients have failure of the autograft, resulting in permanent hypoparathyroidism, which carries significant morbidity and rarely occurs following subtotal resection.[74]

Experienced surgeons generally also advocate for transcervical thymectomy at the initial operation to reduce the risk of thymic carcinoid tumor development, and because the thyrothymic ligament and thymus are common sites for parathyroid rests or supernumerary parathyroid glands. Also, transcervical thymectomy is much easier at the initial operation than at subsequent explorations. Thus, a commonly adapted approach to minimize cervical reoperation and permanent hypoparathyroidism is to perform a careful but aggressive subtotal parathyroidectomy with transcervical thymectomy at the initial operation, and, in the case of recurrence, perform a completion total parathyroidectomy with parathyroid autotransplantation to the forearm. If the patient develops subsequent episodes of recurrence, the transplanted tissue can be debulked.[75]

Multiple endocrine neoplasia type 2A (MEN2A) is an autosomal dominant inherited syndrome resulting from mutations in the *RET* protooncogene, predisposing patients to medullary thyroid carcinoma (MTC), PHPT, and pheochromocytoma. In MEN2A, the hallmark disorder is MTC, affecting nearly every patient, with PHPT occurring in 20% to 30%. PHPT in these patients is overall similar to that seen in sporadic patients, in that most cases result from apparent single-gland enlargement at any one time; however, the management is more complicated, because most have already undergone a prophylactic or therapeutic thyroidectomy for MTC. In these patients, if PHPT has not developed at the time of thyroidectomy, surgeons typically leave normal-appearing parathyroid glands *in situ*. Given the lower penetrance of PHPT in these patients, only a few centers advocate for total parathyroidectomy and autotransplantation at the time of the initial thyroid surgery.[76]

HPT–jaw tumor (HPTJT) syndrome is a rare autosomal dominant disease resulting from mutations in the *CDC73/HRPT2* gene, with patients predisposed to PHPT, ossifying fibromas of the maxilla or mandible, cystic and neoplastic renal lesions, and hyperplastic and neoplastic uterine involvement.[77] PHPT occurs with the highest penetrance, affecting approximately 80% of patients. These cases generally present with apparent single-gland disease.[77] Importantly, the syndrome is associated with parathyroid carcinoma in approximately 15% of patients.[78] With this association, the surgeon must carefully consider an oncologic resection during operative exploration.[79]

Familial isolated HPT (FIHPT) has had an evolving description and associations over the last 85 years.[80] It is a condition of HPT not associated with other endocrine disorders. Although this has been actively studied, it is still unclear if a specific gene mutation is associated with this disease. Recent studies suggest that over 80% of families with FIHPT have no identifiable gene mutations; however, some families with FIHPT have shown germline mutations in *MEN1*, *CASR*, or *CDC73*. In small cohorts of those without any of the aforementioned mutations, whole-exome sequencing identified germline *GCM2* mutations.[80] FIHPT has an inheritance pattern consistent with autosomal dominant transmission. In the absence of an identifiable mutation, surgical management is the same as that for sporadic PHPT. If a genetic mutation is identified, surgical management is tailored to the underlying disease.

Mutations in the *CASR* gene result in a number of calcium disorders. Inactivation mutations can lead to familial hypocalciuric hypercalcemia (FHH) if heterozygous, or neonatal severe PHPT (NSPHPT) if homozygous.[81,82] FHH results in lifelong elevated serum calcium levels with normal to mildly elevated PTH concentration and low renal calcium excretion. The hypercalcemia is generally mild and does not require treatment, and surgery is rarely indicated.[81] Occasionally, patients developed symptomatic PHPT and require surgical resection. NSPHPT is rare, presenting at birth or in early infancy with marked hypercalcemia, skeletal demineralization, and failure to thrive. If left untreated, it can be fatal.[81] These patients require hydration and urgent progression to surgery. Total parathyroidectomy with forearm parathyroid autotransplantation is the recommended surgical approach. Bisphosphonates, and increasingly cinacalcet, have been used with good effect with specific gene mutations, although generally as a temporizing measure.[83,84]

Parathyroid Carcinoma

Parathyroid carcinoma is a rare malignancy affecting less than 1% of those presenting with PHPT and accounting for 0.005% of all cancers in the United States.[85,86] Nearly all parathyroid carcinomas present with functional tumors and symptoms and signs of significant hypercalcemia. This diagnosis is suspected when significant preoperative elevation of calcium and PTH levels are noted.[86] Of those with parathyroid carcinoma, 30% to 76% have a palpable cervical mass on physical exam.[87] Although the majority are sporadic, parathyroid carcinomas can be associated with familial syndromes, HPTJT, MEN1, MEN2A, and FIHPT.[85] In apparent sporadic cases, inactivation of the *CDC73* gene has been identified in up to 70% of cases of parathyroid carcinoma, one-third of which are germline.

During surgery, parathyroid carcinoma often manifests as an enlarged, firm mass, adherent to surrounding structures, and is grey or white in color.[85] The optimal surgical approach remains *en bloc* oncologic resection with the ipsilateral thyroid lobe and other adherent structures that can be sacrificed for the best chance of cure.[88] Regional lymph node involvement occurs in only 10% of cases, and a careful exploration of the ipsilateral central neck and resection of abnormal-appearing or pathologically involved lymph nodes is recommended. Although still under investigation, prophylactic ipsilateral central or lateral neck dissections have not shown a survival benefit.[89,90] Resection of the ipsilateral parathyroid gland is also considered to clear all parathyroid tissue from the affected side.

Local recurrence rates vary greatly, ranging from 33% to 78%; however, over 75% of these patients had standard parathyroidectomy at the initial operation, and thus may not have had adequate resection.[89-91] Although locoregional disease is more common, the lungs and liver are the most common sites for distant metastasis. Surgery in cases with distant disease is often debulking to reduce tumor burden and relieve symptoms; rarely is it curative (Fig. 56.12). The overall survival rates, however, are generally good, with 5-year survival rates ranging from 76% to 85% and 10-year rates ranging from 49% to 77%. In cases of complete tumor resection at the initial operation, these rates are 90% and 67.5%, respectively.[90,92,93]

KEY POINTS: Parathyroid Carcinoma

- Parathyroid carcinoma is best managed at the initial operation.
- Suspected when preoperative calcium and parathyroid hormone are severely elevated.
- Intraoperatively often appears as an enlarged, firm, adherent grey/white structure.
- Oncologic resection with *en bloc* resection of the ipsilateral thyroid lobe and any other adherent resectable structures offers the best chance of cure.
- When suspicious, avoiding capsule disruption can prevent seeding of malignant parathyroid cells.

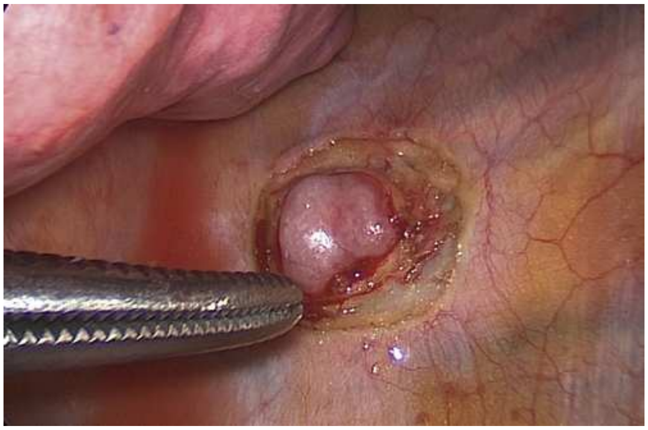


Fig. 56.12 Intrathoracic metastatic implant of parathyroid carcinoma, on diaphragm at tip of instrument.

Pediatric Parathyroid Surgery

In the pediatric population, PHPT is rare, affecting two to five per 100,000 children; thus, the literature is comparatively limited.[94] Although the incidence is less, it is more likely to be clinically significant.[95,96] Most high-volume institutional studies have only a few pediatric cases to evaluate. A 2013 literature review of English-language studies identified 230 reported cases of pediatric PHPT since 1987 and demonstrated etiologies proportionally similar to adults. Single adenomas accounted for 80% of cases, multigland hyperplasia for 16.5%, and double adenomas for 0.9%. This population also has a higher incidence of hereditary disease, and half of the multigland hyperplasia patients were attributed to MEN1, MEN 2A, or familial non-MEN hyperparathyroidism.[97]

Presentation in pediatric patients varies, as serum calcium is not routinely checked in children; therefore, asymptomatic PHPT is not often diagnosed. Children can present with severe hypercalcemic symptoms such as hypotonia, poor feeding, vomiting, constipation, abdominal pain, lethargy, failure to thrive, polyuria, dehydration, and seizures.[98] Furthermore, likely because of the severity of hypercalcemia at the time of diagnosis, children are more likely to present with end-organ damage, such as nephrocalcinosis, nephrolithiasis, pancreatitis, or bone disease.[99] Studies estimate that 77% to 94% of pediatric patients are symptomatic at the time of diagnosis, and that 61% have major signs and symptoms, including 42% with nephrolithiasis, 27% with bone disease, and 3% with pancreatitis.[100,101]

Despite the severity of symptoms, similar to adults, the majority of sporadic pediatric PHPT cases are because of single adenomas. As such, the operative approach is similar to that for adult cases, with preoperative imaging for localization and surgical planning and intraoperative adjuncts including a PTH assay to determine the extent of surgery necessary. In cases with known or suspected familial syndromes, SHPT, or THPT, presentation at a very young age, or other suspicion of multigland disease, four-gland exploration may be required, but otherwise, local exploration is usually sufficient. Preoperative imaging modalities in this population relies on ultrasonography and technetium-99m-sestamibi, with MRI and 4D CT reserved for cases in which other modalities have failed to localize.[97,101] 4D CT utilization should be minimized, as it is associated with a greater radiation dose to the thyroid than sestamibi, even with low-dose CT imaging.

Although the literature is limited, several studies suggest that pediatric image-guided MIP results in similar cure rates to those reported in adults, at 92% to 100%.[94,102-106] Ectopic gland incidence is higher in the pediatric population, with a recent multiinstitutional series estimating 25% ectopic adenomas, and prior reported rates ranging from

10% to 37%.[94,104,105] As in the adult patient, with intraoperative PTH measurement, a reduction of levels to less than 50% of baseline and into the normal range is used as an indication of intraoperative biochemical cure.[106]

Familial forms of hyperparathyroidism in the pediatric population are associated with a higher risk of recurrence, because any parathyroid tissue is left *in situ* is genetically abnormal, resulting in a recurrence rate of 10% to 30%.[103] The goals of treatment are to achieve and maintain eucalcemia for the longest period, avoid permanent hypoparathyroidism, and facilitate future surgery for recurrent disease.[79,107] In the pediatric population, genetic multigland disease has two main surgical options: subtotal resection leaving a small marked remnant *in situ* or total parathyroidectomy with autotransplantation of parathyroid tissue into the forearm. The reimplanted tissue is marked with metallic clips to facilitate future exploration. The surgical approach can be based upon pubertal status. Four-gland total parathyroidectomy with autotransplantation may be performed in the postpubertal population. This is associated with at least a 3- to 4-week period of hypoparathyroidism during which the reimplanted tissue is revascularized, and there is a 25% to 50% risk of permanent hypoparathyroidism.[105] In the prepubertal population, it is thought to be advantageous to leave approximately 30 to 50 mg of parathyroid tissue *in situ* to avoid the risk of permanent hypoparathyroidism.

Compared with the adult population, children have higher rates of general and endocrine-specific complications following parathyroidectomy.[108] Database studies report rates of temporary and permanent hypocalcemia of 56% and 4%, respectively.[94] One contribution to the significant difference in intraoperative and postoperative complications is the level of experience of the surgeon.[57,109] Procedures performed by high-volume surgeons are associated with shorter lengths of stay, lower costs, and fewer complications compared with those performed by low-volume surgeons. Because PHPT is rare in the pediatric population, referral to a high-volume endocrine center for collaboration between pediatric and endocrine surgeons is strongly recommended for this group.[106,109]

Hyperparathyroidism in Pregnancy

The majority of published data regarding PHPT in pregnancy consist of case reports or small case series with limited power and high potential for bias.[110] As such, consensus multidisciplinary guidelines for evaluation and treatment have not explicitly addressed this population. PHPT in pregnancy occurs with an incidence of 0.15% to 1.4%; however, the risks to the mother and fetus are significant.[111] The physiologic changes that occur with pregnancy, including hypoalbuminemia, increased glomerular filtration rate, calcium transport across the placenta, and decreased PTH-mediated bone resorption from estrogen inhibition, coupled with high fetal calcium requirements, can mask or delay the diagnosis.[110,111] Additionally, many early symptoms may be dismissed because of the pregnancy, including fatigue, lethargy, and weakness.[110,111] Maternal complications such as nephrolithiasis, hyperemesis, preeclampsia, miscarriage, polyhydramnios, hypercalcemic crisis, and pancreatitis occur in as many as 67% of untreated patients.[110-112] Fetal or neonatal complications occur in up to 80% of cases in which the mother goes untreated, including intrauterine growth retardation, low birth weight, preterm delivery, intrauterine fetal demise, or postpartum neonatal hypocalcemic tetany.[110-112]

Management of PHPT in pregnancy must take into consideration the effect of treatment on the fetus. Furthermore, the risk of continued hypercalcemia and hyperparathormonemia to the patient and fetus must be deliberated, in conjunction with patient symptoms, prior obstetric history, and gestational age of the fetus when determining management strategy. Medical treatment is an option, consisting of low calcium diet and oral or IV hydration, occasionally with added diuretics, oral phosphates, and vitamin D supplementation. Bisphosphonates are contraindicated in pregnancy because they can affect fetal skeletal development. Cinacalcet has been advocated for use in these patients; however, there are insufficient safety data with regards to effects on the fetus.[110] Calcitonin is not known to cross the placental barrier; however, it is short-acting and therefore unlikely to provide a sustained benefit.[110] Alternatively, parathyroidectomy is the definitive treatment for PHPT in pregnancy. Surgery is recommended when the serum calcium is greater than 11 mg/dL (2.75 mmol/L), especially in those with previous loss of pregnancy.[111] It should also be considered in those with less severe hypercalcemia, specifically because of the complications noted with untreated disease. With surgery, patients can avoid potential harmful medications and the long-term maternal and fetal effects of hypercalcemia and hyperparathormonemia.[110] Following treatment, these patients have lower rates of preeclampsia and preterm labor compared with those managed medically.[113]

Similarly to the general population, PHPT in the pregnant population is caused by a solitary parathyroid adenoma in approximately 85% of cases.[111] The surgical approach is still dependent on preoperative localization. Imaging in this patient population is limited to ultrasound and MRI. Sestamibi scans were considered contraindicated in pregnancy; however, some authors suggest that they can be used selectively by utilizing techniques to reduce fetal irradiation.[111] 4D CT should be avoided because of the risk of radiation exposure to the fetus. If imaging can localize a candidate lesion, a focal exploration with intraoperative PTH measurement without general anesthesia can be performed.[111] If imaging is unable to identify a candidate lesion, bilateral cervical exploration under general anesthesia is still a safe option.

The timing to proceed to surgery is still debated. For asymptomatic patients with minimal serum calcium elevation, surgery may be delayed until the postpartum period. Concern for fetal outcomes with surgical intervention is the reasoning behind opting for medical treatment during pregnancy.[111] Additionally, limitations on localizing imaging studies further discourage surgical intervention. Finally, there are additional risks during pregnancy related to general anesthesia. In cases where it is deemed necessary to proceed with surgery during pregnancy, because of incomplete organogenesis in the first trimester and the possibility of inducing preterm labor during the third trimester, parathyroidectomy is preferably performed in the second trimester.[113] In more severe cases, when medical management is unsuccessful or there is evidence of maternal or fetal complications, surgery can be performed safely regardless of gestational age. Reviews of medical compared with surgical treatment during pregnancy show increases in both maternal complications and neonatal tetany in those managed medically.[111] Following surgical intervention, the rate of neonatal complication was lower than in those who had medical treatment alone.[111,113]

SECONDARY HYPERPARATHYROIDISM

In SHPT, chronic hypocalcemia results in persistent stimulation of the parathyroid glands, leading to multigland hypertrophy. Most commonly, it is seen with end-stage renal failure (uremic hyperparathyroidism), but it is also present with certain bone and metabolic conditions such as malnutrition, malabsorption, vitamin D deficiency, chronic cholestasis, and with use of lithium.

In the renal failure population, patients generally have low serum calcium and 1,25-hydroxy-vitamin D levels, and significantly elevated serum phosphate and intact PTH levels. Although approximately 75% of the renal failure population have biochemically evident SHPT, the Healthcare Cost and Utilization Project's Nationwide Inpatient Sample estimates that, overall, 5.4/1000 renal dialysis patients require parathyroidectomy each year.[114] This risk increases with the duration of dialysis treatment, to 15% of patients after 10 years of therapy and 38% after 20 years.[115] Patients are initially managed medically, with calcitriol and calcimimetics, reserving surgery for those who fail medical therapy. Surgical indications for uremic hyperparathyroidism include uncontrollable hyperphosphatemia, worsening osteoporosis with bone pain and fractures, development of bone erosions or osteitis fibrosa cystica, and calciphylaxis.[116]

With vitamin D deficiency-associated SHPT, diminished calcium absorption in the gastrointestinal system results in hypocalcemia, causing chronically elevated PTH levels. Although occurring widely, it is common in the obese population. Unique to this population, bariatric surgery with subsequent malabsoprtive syndromes can exacerbate vitamin D deficiency and resultant SHPT in as many as 20% to 30% of patients. This exacerbation occasionally progresses to a level at which patients require surgical intervention.[117] As the prevalence of bariatric surgery increases, this phenomenon is seen with increasing frequency. Wei et al. described SHPT in 21% of the preoperative obese cohort, in 35% 1 year following bariatric surgery, and in 63% 5 years postoperatively.[117]

The pathophysiology of SHPT results in multigland disease. Accordingly, the surgical management generally consists of either subtotal parathyroidectomy or total parathyroidectomy with parathyroid autotransplantation. Total parathyroidectomy without autotransplantation is undesirable because of resulting severe hypoparathyroidism.[118] There do not appear to be significant differences in complication rates between these two operations; however, total parathyroidectomy is associated with increased length of stay and increased readmission rates compared with subtotal parathyroidectomy.[119] Trends suggest that subtotal parathyroidectomy is the preferred approach in the United States, whereas total parathyroidectomy with autotransplantation is preferred in Japan; however, there are additional factors that must be considered.[116] Patients unlikely to progress to renal transplantation will continue to have lifelong stimulation of any remaining parathyroid tissue; thus, total parathyroidectomy with autotransplantation into the forearm may be preferable to avoid the need for cervical reoperation. In those who may be candidates for renal transplant, a subtotal parathyroidectomy, with preservation of approximately 30 mg of parathyroid tissue *in situ*, is preferable, as it is sufficient to provide normal parathyroid function. Additionally, by leaving only a small amount of parathyroid tissue in a single location in the neck, should cervical reoperation be required, it can be expedited. Following autotransplantation of parathyroid tissue, it can take 12 to 18 weeks for circulating PTH levels to return to normal, and these patients may have profound hypoparathyroidism and hypocalcemia that is exacerbated by renal osteodystrophy and hungry bone syndrome.

Irrespective of the surgical approach, 15% to 30% of SHPT patients have supernumerary glands and need a thorough parathyroid exploration for optimal results.[120] Transcervical thymectomy should be a standard part of surgical management of SHPT because of the high incidence of supernumerary glands in the thyrothymic ligament and thymus. If not resected, these sites can cause persistent disease and, with chronic stimulation, grow and result in disease recurrence. One study found that over 28% of uremic hyperparathyroid patients undergoing reoperative parathyroidectomy had an intrathymic parathyroid gland removed at reoperation, suggesting that this could have been prevented had thymectomy been performed at the initial operation.[121]

Following surgery, these patients require meticulous management because they are prone to significant postoperative hypocalcemia, requiring inpatient admission for several days with oral and IV calcium supplementation and calcitriol. Because most have lived with chronic hypocalcemia, they often do not exhibit overt symptoms until their calcium levels are considerably abnormal. Accordingly, it is extremely important to have an experienced team for management to prevent serious, potentially life-threatening, hypocalcemia.

Preoperative imaging in these cases is controversial. Contrary to PHPT, in SHPT four-gland exploration should always be performed. Some advocate for no preoperative imaging studies, as gland localization to limit surgery is unnecessary, given known multigland disease. However, sestamibi or CT imaging may be able to identify ectopic or supernumerary glands that could be missed with standard exploration. Additionally, an ultrasound would screen for concomitant thyroid pathology that should be addressed simultaneously.

The use of intraoperative PTH measurement during these cases is also controversial, but not without benefit. Because PTH is renally cleared, and these patients often have a profoundly increased baseline PTH, intraoperative drop in PTH to the normal range generally does not occur quickly. However, analysis of the changes can still be used to predict cure. Although there is no consensus, several studies suggest that a decrease in intraoperative PTH levels from baseline by 80% or more at 20 minutes following completion of either subtotal parathyroidectomy or total parathyroidectomy with autotransplantation predicts cure in long-term follow-up.[122-124]

TERTIARY HYPERPARATHYROIDISM

THPT is a condition in which prolonged stimulation from SHPT results in autonomous parathyroid hyperfunction. It most commonly occurs in patients with uremic hyperparathyroidism and in end-stage renal patients following renal transplantation. Although mild hypercalcemia may be present following renal transplantation, it most commonly resolves within the first year. Persistence beyond 2 years, significant hypercalcemia, or the development of hypercalcemia following a normocalcemic posttransplant period is indicative of this disease.[118,125-127]

Surgery is indicated in the vast majority of THPT patients because of the risk of hypercalcemic toxicity to the renal transplant and severe bone disease. Because of the mechanism of development, four-gland exploration should be performed because all parathyroid glands will be abnormal. Subtotal parathyroidectomy is the procedure of choice, although some opt for a more limited image-guided parathyroidectomy.[118,125,126] Similar to SHPT, if bilateral exploration is performed, imaging for localization is not necessary; however, it can be useful in the setting of ectopic glands. Intraoperative PTH measurement is not necessary but can be used to confirm intraoperative cure and predict long-term cure.[122,128,129]

KEY POINTS: Secondary and Tertiary Hyperparathyroidism

- Secondary hyperparathyroidism (SHPT) is attributed to chronic hypocalcemia causing persistent parathyroid stimulation.
- Tertiary hyperparathyroidism (THPT) is a result of prolonged SHPT resulting in autonomous parathyroid hyperfunction.
- There are several causes of SHPT and THPT; end-stage renal disease is the most common.
- Multigland hyperplasia must be anticipated and appropriately addressed during surgery.

COVID-19 AND PARATHYROID SURGERY

First identified as a pneumonia epidemic spreading through Wuhan, the capital of the Hubei province China, the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a novel RNA coronavirus. It rapidly spread across China, to Italy and other European countries, then to the United States and throughout the world.[130] It was named coronavirus disease 2019 (COVID-19) by the World Health Organization and declared a pandemic in early 2020 owing to the rapidity and degree of its spread. There is significant variability in presentation, from asymptomatic carriers to severe interstitial pneumonia and acute respiratory distress syndrome with multiorgan failure leading to severe acute respiratory failure and death.[130] Although this disease is still being studied, it has had a significant effect on the medical system and will continue to do so for an indeterminate period of time.[131] As such, it is pertinent to examine how this disease affects operative intervention for HPT.

COVID-19 is generally spread through respiratory droplets, from direct, indirect, or close contact with infected persons, and medical procedures can lead to aerosol formation and viral dissemination, putting providers and patients at risk for infection.[132] Procedures involving instrumentation of the oropharyngeal, nasopharyngeal, and respiratory mucosa are high risk for aerosolization.[132,133] Additionally, those with active infections are at higher risk for postoperative complications. As such, it is important to consider infection status when planning surgery.[134]

Because of the fraction of patients with COVID-19 that become profoundly ill, regions fluctuate in local disease burden and need for hospitalization. When infection rates peak, cities have reported hospitals and intensive care units full to overflowing with COVID-19 patients, as well as cessation of many other inpatient services, including nonemergency surgeries, owing to lack of available space, staff, and equipment.[133,135] The vast majority of parathyroid surgery is elective. As such, depending on the local disease burden and hospital capacity, the majority of these procedures may be delayed until the resources are available.

When resources are available to proceed with surgery, for the safety of the patient, care providers, and others in the hospital, it is now recommended to test for COVID-19 preoperatively.[135] If the patient is asymptomatic and the test is negative, proceed with surgery as per hospital guidelines. However, if the patient is symptomatic or the test is positive, additional consideration is necessary. In emergent cases that cannot be delayed, providers should take additional precautions for personal protection, but in most cases, surgery can be deferred until it is safe to proceed, reducing the operative risk and potential for virus transmission.[131,134,135] Both direct and indirect laryngoscopy for vocal cord evaluation should also be avoided without confirmation of COVID-19 status for provider protection.

SUMMARY

Hyperparathyroidism encompasses a complex collection of diseases requiring careful diagnosis and, when surgery is indicated, meticulous, thoughtful, surgical management. When indicated, parathyroidectomy is curative in the vast majority of patients, increasingly so when managed by high-volume centers/surgeons. Preoperative imaging and intraoperative adjuncts allow for minimally-invasive approaches. Patients with persistent and recurrent disease, HPT with a genetic basis, or parathyroid carcinoma and the pediatric patient population represent the most complex surgical patients and require additional consideration and expertise in management.

Acknowledgments

The authors thank Dr. Glenda G. Callender, Dr. Tobias Carling, and Dr. Emily Christison-Lagay for their contributions in the authorship of the chapter in the previous edition.

REFERENCES

1. Bilezikian JP, Silverberg SJ. Normocalcemic primary hyperparathyroidism. *Arq Bras Endocrinol Metabol*. 2010;54:106–109.
2. Sivula A, Ronni-Sivula H. The changing picture of primary hyperparathyroidism in the years 1956-1979. *Ann Chir Gynaecol*. 1984;73:319–324.
3. Bilezikian JP, Cusano NE, Khan AA, et al. Primary hyperparathyroidism. *Nat Rev Dis Primers*. 2016;2:16033–16033.
4. Silva BC, Cusano NE, Bilezikian JP. Primary hyperparathyroidism. *Best Pract Res Clin Endocrinol Metab*. 2018;32:593–607.
5. Walker MD, Silverberg SJ. Primary hyperparathyroidism. *Nat Rev Endocrinol*. 2018;14:115–125.
6. Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3561–3569.
7. Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons guidelines for definitive management of primary hyperparathyroidism. *JAMA Surg*. 2016;151:959–968.
8. Stephen AE, Mannstadt M, Hodin RA. Indications for surgical management of hyperparathyroidism: a review. *JAMA Surg*. 2017;152:878–882.
9. Chan AK, Duh QY, Katz MH, et al. Clinical manifestations of primary hyperparathyroidism before and after parathyroidectomy. a case-control study. *Ann Surg*. 1995;222:402–412;discussion 412–414.
10. Hasse C, Sitter H, Bachmann S, et al. How asymptomatic is asymptomatic primary hyperparathyroidism? *Exp Clin Endocrinol Diabetes*. 2000;108:265–274.
11. Walker MD, McMahon DJ, Inabnet WB, et al. Neuropsychological features in primary hyperparathyroidism: a prospective study. *J Clin Endocrinol Metab*. 2009;94:1951–1958.
12. Rao DS, Phillips ER, Divine GW, et al. Randomized controlled clinical trial of surgery versus no surgery in patients with mild asymptomatic primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2004;89:5415–5422.
13. Ambrogini E, Cetani F, Cianferotti L, et al. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. *J Clin Endocrinol Metab*. 2007;92:3114–3121.
14. Bollerslev J, Jansson S, Mollerup CL, et al. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. *J Clin Endocrinol Metab*. 2007;92:1687–1692.
15. Pasieka JL, Parsons LL. Prospective surgical outcome study of relief of symptoms following surgery in patients with primary hyperparathyroidism. *World J Surg*. 1998;22:513–518;discussion 518–519.
16. Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*. 2008;93:3462–3470.
17. Silverberg SJ, Shane E, Jacobs TP, et al. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. *N Engl J Med*. 1999;341:1249–1255.
18. Bilezikian JP, Bandeira L, Khan A, et al. Hyperparathyroidism. *Lancet*. 2018;391:168–178.
19. Barczynski M, Golkowski F, Konturek A, et al. Technetium-99m-sestamibi subtraction scintigraphy vs. ultrasonography combined with a rapid parathyroid hormone assay in parathyroid aspirates in preoperative localization of parathyroid adenomas and in directing surgical approach. *Clin Endocrinol (Oxf)*. 2006;65:106–113.
20. Lo CY, Lang BH, Chan WF, et al. A prospective evaluation of preoperative localization by technetium-99m sestamibi scintigraphy and ultrasonography in primary hyperparathyroidism. *Am J Surg*. 2007;193:155–159.
21. Xue J, Liu Y, Ji T, et al. Comparison between technetium-99m methoxyisobutylisonitrile scintigraphy and ultrasound in the diagnosis of

parathyroid adenoma and parathyroid hyperplasia. *Nucl Med Commun*. 2018;39:1129–1137.

22. Mohebati A, Shaha AR. Imaging techniques in parathyroid surgery for primary hyperparathyroidism. *Am J Otolaryngol*. 2012;33:457–468.
23. Wong KK, Fig LM, Gross MD, et al. Parathyroid adenoma localization with 99mTc-sestamibi SPECT/CT: a meta-analysis. *Nucl Med Commun*. 2015;36:363–375.
24. Nafisi Moghadam R, Amlelshahbaz AP, Namiranian N, et al. Comparative diagnostic performance of ultrasonography and 99mtc-sestamibi scintigraphy for parathyroid adenoma in primary hyperparathyroidism; systematic review and meta-analysis. *Asian Pac J Cancer Prev*. 2017;18:3195–3200.
25. Bann DV, Zacharia T, Goldenberg D, et al. Parathyroid localization using 4D-computed tomography. *Ear Nose Throat J*. 2015;94:E55–E57.
26. Rodgers SE, Hunter GJ, Hamberg LM, et al. Improved preoperative planning for directed parathyroidectomy with 4-dimensional computed tomography. *Surgery*. 2006;140:932–940;discussion 940–941.
27. Starker LF, Mahajan A, Björklund P, et al. 4D parathyroid CT as the initial localization study for patients with de novo primary hyperparathyroidism. *Ann Surg Oncol*. 2011;18:1723–1728.
28. Mahajan A, Starker LF, Ghita M, et al. Parathyroid four-dimensional computed tomography: evaluation of radiation dose exposure during preoperative localization of parathyroid tumors in primary hyperparathyroidism. *World J Surg*. 2012;36:1335–1339.
29. Calo PG, Pisano G, Loi G, et al. Surgery for primary hyperparathyroidism in patients with preoperatively negative sestamibi scan and discordant imaging studies: the usefulness of intraoperative parathyroid hormone monitoring. *Clin Med Insights Endocrinol Diabetes*. 2013;6:63–67.
30. Elaraj DM, Sippel RS, Lindsay S, et al. Are additional localization studies and referral indicated for patients with primary hyperparathyroidism who have negative sestamibi scan results? *Arch Surg*. 2010;145:578–581.
31. Fowler GE, Chew PR, Lim CBB, et al. Is there a role for routine laryngoscopy before and after parathyroid surgery? *Surgeon*. 2019;17:102–106.
32. Dhillon VK, Rettig E, Noureldine SI, et al. The incidence of vocal fold motion impairment after primary thyroid and parathyroid surgery for a single high-volume academic surgeon determined by pre- and immediate post-operative fiberoptic laryngoscopy. *Int J Surg*. 2018;56:73–78.
33. Sinclair CF, Bumpous JM, Haugen BR, et al. Laryngeal examination in thyroid and parathyroid surgery: an American Head and Neck Society consensus statement. *Head Neck*. 2016;38:811–819.
34. Udelsman R, Donovan P, Shaw C. Cure predictability during parathyroidectomy. *World J Surg*. 2014;38:525–533.
35. Carling T, Udelsman R. Minimally invasive parathyroidectomy. In: Oertli D, Udelsman R, eds. *Surgery of the Thyroid and Parathyroid Glands*. Berlin and Heidelberg: Springer; 2012:475–485.
36. Carling T, Udelsman R. Focused approach to parathyroidectomy. *World J Surg*. 2008;32:1512–1517.
37. Callender GG, Udelsman R. Surgery for primary hyperparathyroidism. *Cancer*. 2014;120:3602–3616.
38. Starker LF, Fonseca AL, Carling T, et al. Minimally invasive parathyroidectomy. *Int J Endocrinol*. 2011;2011:206502.
39. Schneider R, Hinrichs J, Meier B, et al. Minimally invasive parathyroidectomy without intraoperative PTH performed after positive ultrasonography as the only diagnostic method in patients with primary hyperparathyroidism. *World J Surg*. 2019;43:1525–1531.
40. Mownah OA, Pafitanis G, Drake WM, et al. Contemporary surgical treatment of primary hyperparathyroidism without intraoperative parathyroid hormone measurement. *Ann R Coll Surg Engl*. 2015;97:603–607.
41. Soyder A, Ünübol M, Ömürlü İK, et al. Minimally invasive parathyroidectomy without using intraoperative parathyroid hormone monitoring or gamma probe. *Ulus Cerrahi Derg*. 2015;31:9–14.
42. Gagner M. Endoscopic subtotal parathyroidectomy in patients with primary hyperparathyroidism. *Br J Surg*. 1996;83:875.
43. Erinjeri NJ, Udelsman R. Transoral endoscopic thyroidectomy vestibular approach. *Adv Surg*. 2019;53:179–193.
44. Kunstman JW, Udelsman R. Superiority of minimally invasive parathyroidectomy. *Adv Surg*. 2012;46:171–189.
45. Udelsman R, Lin Z, Donovan P. The superiority of minimally invasive parathyroidectomy based on 1650 consecutive patients with primary hyperparathyroidism. *Ann Surg*. 2011;253:585–591.
46. Norman J, Politz D. 5,000 Parathyroid operations without frozen section or PTH assays: measuring individual parathyroid gland hormone production in real time. *Ann Surg Oncol*. 2009;16:656–666.
47. Norman J, Politz D. Measuring individual parathyroid gland hormone production in real-time during radioguided parathyroidectomy. Experience in over 8,000 operations. *Minerva Endocrinol*. 2008;33:147–157.
48. Bellantone R, Raffaelli M, De Crea C, Traini E, Lombardi CP. Minimally-invasive parathyroid surgery. *Acta Otorhinolaryngol*. 2011;31:207–215.
49. Hall NC, Plews RL, Agrawal A, et al. Intraoperative scintigraphy using a large field-of-view portable gamma camera for primary hyperparathyroidism: initial experience. *BioMed Res Int*. 2015;2015:930575.
50. Jin H, Cui M. New advances of ICG angiography in parathyroid identification. *Endocr Metab Immune Disord—Drug Targets*. 2019;19:936–940.
51. Mazeh H, Chen H. Intraoperative adjuncts for parathyroid surgery. *Expert Rev Endocrinol Metab*. 2011;6:245–253.
52. Aygün N, Uludağ M. Intraoperative adjunct methods for localization in primary hyperparathyroidism. *Sisli Etfal Hastan Tip Bul*. 2019;53:84–95.
53. Kartush JM, Naumann I. Laryngeal nerve monitoring. *Neurodiagn J*. 2014;54:227–259.
54. Randolph GW, Dralle H, Abdullah H, et al. Electrophysiologic recurrent laryngeal nerve monitoring during thyroid and parathyroid surgery: international standards guideline statement. *Laryngoscope*. 2011;121(suppl 1):S1–S16.
55. Steen S, Rabeler B, Fisher T, et al. Predictive factors for early postoperative hypocalcemia after surgery for primary hyperparathyroidism. *Proc (Bayl Univ Med Cent)*. 2009;22:124–127.
56. Udelsman R. Six hundred fifty-six consecutive explorations for primary hyperparathyroidism. *Ann Surg*. 2002;235:665–670;discussion 670–672.
57. Erinjeri NJ, Udelsman R. Volume–outcome relationship in parathyroid surgery. *Best Pract Res Clin Endocrinol Metab*. 2019:101287.
58. Adam MA, Thomas S, Youngwirth L, et al. Is there a minimum number of thyroidectomies a surgeon should perform to optimize patient outcomes? *Ann Surg*. 2017;265:402–407.
59. Mitchell J, Milas M, Barbosa G, et al. Avoidable reoperations for thyroid and parathyroid surgery: effect of hospital volume. *Surgery*. 2008;144:899–906;discussion 906–907.
60. Malmaeus J, Granberg PO, Halvorsen J, et al. Parathyroid surgery in Scandinavia. *Acta Chir Scand*. 1988;154:409–413.
61. Bergenfelz A, Jansson S, Mårtensson H, et al. Scandinavian quality register for thyroid and parathyroid surgery: audit of surgery for primary hyperparathyroidism. *Langenbeck's Arch Surg*. 2007;392:445–451.
62. Udelsman R, Donovan PI. Remedial parathyroid surgery: changing trends in 130 consecutive cases. *Ann Surg*. 2006;244:471–479.
63. Lebastchi AH, Aruny JE, Donovan PI, et al. Real-time super selective venous sampling in remedial parathyroid surgery. *J Am Coll Surg*. 2015;220:994–1000.
64. Prescott JD, Udelsman R. Remedial operation for primary hyperparathyroidism. *World J Surg*. 2009;33:2324–2334.
65. Wang TS, Udelsman R. Remedial surgery for primary hyperparathyroidism. *Adv Surg*. 2007;41:1–15.
66. Hessman O, Stålberg P, Sundin A, et al. High success rate of parathyroid reoperation may be achieved with improved localization diagnosis. *World J Surg*. 2008;32:774–781;discussion 782–783.
67. Powell AC, Alexander HR, Chang R, et al. Reoperation for parathyroid adenoma: a contemporary experience. *Surgery*. 2009;146:1144–1155.
68. Ahmad S, Kuraganti G, Steenkamp D. Hypercalcemic crisis: a clinical review. *Am J Med*. 2015;128:239–245.
69. Cannon J, Lew JI, Solórzano CC. Parathyroidectomy for hypercalcemic crisis: 40 years' experience and long-term outcomes. *Surgery*. 2010;148:807–812;discussion 812–813.

70. Singh DN, Gupta SK, Kumari N, et al. Primary hyperparathyroidism presenting as hypercalcemic crisis: twenty-year experience. *Indian J Endocrinol Metab*. 2015;19:100–105.
71. Starker LF, Akerström T, Long WD, et al. Frequent germ-line mutations of the MEN1, CASR, and HRPT2/CDC73 genes in young patients with clinically non-familial primary hyperparathyroidism. *Horm Cancer*. 2012;3:44–51.
72. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990–3011.
73. Shah MH, Goldner WS, Benson AB, et al. *Neuroendocrine and Adrenal Tumors*. NCCN Guidelines; 2020. Available at: www.nccn.org/professionals/physician_gls/pdf/neuroendocrine.pdf.
74. Elaraj DM, Skarulis MC, Libutti SK, et al. Results of initial operation for hyperparathyroidism in patients with multiple endocrine neoplasia type 1. *Surgery*. 2003;134:858–864;discussion 864–865.
75. Lambert L, Shapiro S, Lee J, et al. Surgical treatment of hyperparathyroidism in patients with multiple endocrine neoplasia type 1. *Arch Surg*. 2005;140:374–382.
76. Skinner MA, Norton JA, Moley JF, et al. Heterotopic autotransplantation of parathyroid tissue in children undergoing total thyroidectomy. *J Pediatr Surg*. 1997;32:510–513.
77. Torresan F, Iacobone M. Clinical features, treatment, and surveillance of hyperparathyroidism-jaw tumor syndrome: an up-to-date and review of the literature. *Int J Endocrinol*. 2019;2019:1761030.
78. Cetani F, Pardi E, Marcocci C. Parathyroid carcinoma. *Front Horm Res*. 2019;51:63–76.
79. Iacobone M, Citton M, Viel G, et al. Surgical approaches in hereditary endocrine tumors. *Updates Surg*. 2017;69:181–191.
80. Marx SJ. New concepts about familial isolated hyperparathyroidism. *J Clin Endocrinol Metab*. 2019;104:4058–4066.
81. Gorvin CM. Molecular and clinical insights from studies of calcium-sensing receptor mutations. *J Mol Endocrinol*. 2019;63:R1–R16.
82. Vahe C, Benomar K, Espiard S, et al. Diseases associated with calcium-sensing receptor. *Orphanet J Rare Dis*. 2017;12:19.
83. Mayr B, Schnabel D, Dörr HG, et al. Genetics in endocrinology: gain and loss of function mutations of the calcium-sensing receptor and associated proteins: current treatment concepts. *Eur J Endocrinol*. 2016;174:R189–R208.
84. Sun X, Huang L, Wu J, et al. Novel homozygous inactivating mutation of the calcium-sensing receptor gene in neonatal severe hyperparathyroidism responding to cinacalcet therapy: a case report and literature review. *Medicine (Baltim more)*. 2018;97:e13128.
85. Salcuni AS, Cetani F, Guarnieri V, et al. Parathyroid carcinoma. *Best Pract Res Clin Endocrinol Metab*. 2018;32:877–889.
86. Sadler C, Gow KW, Beierle EA, et al. Parathyroid carcinoma in more than 1,000 patients: a population-level analysis. *Surgery*. 2014;156:1622–1629;discussion 1629–1630.
87. Shane E. Parathyroid carcinoma. *J Clin Endocrinol Metab*. 2001;86:485–493.
88. Cetani F, Pardi E, Marcocci C. Update on parathyroid carcinoma. *J Endocrinol Invest*. 2016;39:595–606.
89. Hsu KT, Sippel RS, Chen H, et al. Is central lymph node dissection necessary for parathyroid carcinoma? *Surgery*. 2014;156:1336–1341;discussion 1341.
90. Ferraro V, Sgaramella LI, Di Meo G, et al. Current concepts in parathyroid carcinoma: a single Centre experience. *BMC Endocr Disord*. 2019;19:46.
91. Givi B, Shah JP. Parathyroid carcinoma. *Clin Oncol (R Coll Radiol)*. 2010;22:498–507.
92. Kleinpeter KP, Lovato JF, Clark PB, et al. Is parathyroid carcinoma indeed a lethal disease? *Ann Surg Oncol*. 2005;12:260–266.
93. Lee PK, Jarosek SL, Virnig BA, et al. Trends in the incidence and treatment of parathyroid cancer in the United States. *Cancer*. 2007;109:1736–1741.
94. Kollars J, Zarroug AE, van Heerden J, et al. Primary hyperparathyroidism in pediatric patients. *Pediatrics*. 2005;115:974–980.
95. Stokes VJ, Nielsen MF, Hannan FM, et al. Hypercalcemic disorders in children. *J Bone Miner Res*. 2017;32:2157–2170.
96. McNeilly JD, Boal R, Shaikh MG, et al. Frequency and aetiology of hypercalcaemia. *Arch Dis Child*. 2016;101:344–347.
97. Belcher R, Metrailer AM, Bodenner DL, et al. Characterization of hyperparathyroidism in youth and adolescents: a literature review. *Int J Pediatr Otorhinolaryngol*. 2013;77:318–322.
98. Lietman SA, Germain-Lee EL, Levine MA. Hypercalcemia in children and adolescents. *Curr Opin Pediatr*. 2010;22:508–515.
99. Makhdoomi KR, Chalmers J, Campbell IW, et al. Delayed diagnosis of juvenile primary hyperparathyroidism. *J R Coll Surg Edinb*. 1996;41:351–353.
100. Hsu SC, Levine MA. Primary hyperparathyroidism in children and adolescents: the Johns Hopkins Children's Center experience 1984-2001. *J Bone Miner Res*. 2002;17:N44–N50.
101. Mallet E. Primary hyperparathyroidism in neonates and childhood. The French experience (1984-2004). *Horm Res*. 2008;69:180–188.
102. Mancilla EE, Levine MA, Adzick NS. Outcomes of minimally invasive parathyroidectomy in pediatric patients with primary hyperparathyroidism owing to parathyroid adenoma: a single institution experience. *J Pediatr Surg*. 2017;52:188–191.
103. Nicholson KJ, McCoy KL, Witchel SF, et al. Comparative characteristics of primary hyperparathyroidism in pediatric and young adult patients. *Surgery*. 2016;160:1008–1016.
104. Rampp RD, Mancilla EE, Adzick NS, et al. Single gland, ectopic location: adenomas are common causes of primary hyperparathyroidism in children and adolescents. *World J Surg*. 2020;44:1518–1525.
105. Burke JF, Jacobson K, Gosain A, et al. Radioguided parathyroidectomy effective in pediatric patients. *J Surg Res*. 2013;184:312–317.
106. Durkin ET, Nichol PF, Lund DP, et al. What is the optimal treatment for children with primary hyperparathyroidism? *J Pediatr Surg*. 2010;45:1142–1146.
107. Iacobone M, Carnaille B, Palazzo FF, et al. Hereditary hyperparathyroidism—a consensus report of the European Society of Endocrine Surgeons (ESES). *Langenbeck's Arch Surg*. 2015;400:867–886.
108. Wang TS, Roman SA, Sosa JA. Predictors of outcomes following pediatric thyroid and parathyroid surgery. *Curr Opin Oncol*. 2009;21:23–28.
109. Tuggle CT, Roman SA, Wang TS, et al. Pediatric endocrine surgery: who is operating on our children? *Surgery*. 2008;144:869–877;discussion 877.
110. DiMarco AN, Meeran K, Christakis I, et al. Seventeen cases of primary hyperparathyroidism in pregnancy: a call for management guidelines. *J Endocr Soc*. 2019;3:1009–1021.
111. McMullen TPW, Learoyd DL, Williams DC, et al. Hyperparathyroidism in pregnancy: options for localization and surgical therapy. *World J Surg*. 2010;34:1811–1816.
112. McCarthy A, Howarth S, Khoo S, et al. Management of primary hyperparathyroidism in pregnancy: a case series. *Endocrinol Diabetes Metab Case Rep*. 2019;2019:19-0039. Epub ahead of print.
113. Rigg J, Gilbertson E, Barrett HL, et al. Primary hyperparathyroidism in pregnancy: maternofetal outcomes at a quaternary referral obstetric hospital, 2000 through 2015. *J Clin Endocrinol Metab*. 2019;104:721–729.
114. Kim SM, Long J, Montez-Rath ME, et al. Rates and outcomes of parathyroidectomy for secondary hyperparathyroidism in the United States. *Clin J Am Soc Nephrol*. 2016;11:1260–1267.
115. Schneider R, Slater EP, Karakas E, et al. Initial parathyroid surgery in 606 patients with renal hyperparathyroidism. *World J Surg*. 2012;36:318–326.
116. Lau WL, Obi Y, Kalantar-Zadeh K. Parathyroidectomy in the management of secondary hyperparathyroidism. *Clin J Am Soc Nephrol*. 2018;13:952–961.
117. Wei JH, Lee WJ, Chong K, et al. High incidence of secondary hyperparathyroidism in bariatric patients: comparing different procedures. *Obes Surg*. 2018;28:798–804.
118. van der Plas WY, Noltes ME, van Ginhoven TM, et al. Secondary and tertiary hyperparathyroidism: a narrative review. *Scand J Surg*. 2019;4:271–278.
119. Anderson Jr K, Ruel E, Adam MA, et al. Subtotal vs. total parathyroidectomy with autotransplantation for patients with renal hyperparathyroidism have similar outcomes. *Am J Surg*. 2017;214:914–919.
120. Pattou FN, Pellissier LC, Noël C, et al. Supernumerary parathyroid glands: frequency and surgical significance in treatment of renal hyperparathyroidism. *World J Surg*. 2000;24:1330–1334.
121. Schneider R, Bartsch DK, Schlosser K. Relevance of bilateral cervical thymectomy in patients with renal hyperparathyroidism: analysis of 161 patients undergoing reoperative parathyroidectomy. *World J Surg*. 2013;37:2155–2161.

122. Ohe MN, Santos RO, Kunii IS, et al. Intraoperative PTH cutoff definition to predict successful parathyroidectomy in secondary and tertiary hyperparathyroidism. *Braz J Otorhinolaryngol*. 2013;79:494–499.
123. Amza AB, Muntean V, Dindelegan G, et al. Surgery outcomes in patients with secondary hyperparathyroidism and impact of intra-operative PTH measurement. *Acta Endocrinol (Buchar)*. 2017;13:322–328.
124. El-Husseini A, Wang K, Edon A, et al. Value of intraoperative parathyroid hormone assay during parathyroidectomy in dialysis and renal transplant patients with secondary and tertiary hyperparathyroidism. *Nephron*. 2018;138:119–128.
125. Messa P, Alfieri CM. Secondary and tertiary hyperparathyroidism. *Front Horm Res*. 2019;51:91–108.
126. Dulfer RR, Franssen GJH, Hesselink DA, et al. Systematic review of surgical and medical treatment for tertiary hyperparathyroidism. *Br J Surg*. 2017;104:804–813.
127. D'Alessandro AM, Melzer JS, Pirsch JD, et al. Tertiary hyperparathyroidism after renal transplantation: operative indications. *Surgery*. 1989;106:1049–1055;discussion 1055–1056.
128. Triponez F, Dosseh D, Hazzan M, et al. Accuracy of intra-operative PTH measurement during subtotal parathyroidectomy for tertiary hyperparathyroidism after renal transplantation. *Langenbeck's Arch Surg*. 2006;391:561–565.
129. Park JH, Kang SW, Jeong JJ, et al. Surgical treatment of tertiary hyperparathyroidism after renal transplantation: a 31-year experience in a single institution. *Endocr J*. 2011;58:827–833.
130. Pascarella G, Strumia A, Piliego C, et al. COVID-19 diagnosis and management: a comprehensive review. *J Intern Med*. 2020;288:192–206.
131. Elizabeth Brindle M, Gawande A. Managing COVID-19 in surgical systems. *Ann Surg*. 2020;272:e1–e2.
132. Harding H, Broom A, Broom J. Aerosol-generating procedures and infective risk to healthcare workers from SARS-CoV-2: the limits of the evidence. *J Hosp Infect*. 2020;105:717–725.
133. Moletta L, Pierobon ES, Capovilla G, et al. International guidelines and recommendations for surgery during Covid-19 pandemic: a systematic review. *Int J Surg*. 2020;79:180–188.
134. Al-Jabir A, Kerwan A, Nicola M, et al. Impact of the coronavirus (COVID-19) pandemic on surgical practice—Part 1. *Int J Surg*. 2020;79:168–179.
135. Flemming S, Hankir M, Ernestus RI, et al. Surgery in times of COVID-19-recommendations for hospital and patient management. *Langenbecks Arch Surg*. 2020;405:359–364.

57

Pseudohypoparathyroidism, Albright's Hereditary Osteodystrophy, and Progressive Osseous Heteroplasia: Disorders Caused by *GNAS* Mutations that Reduce Gsα Activity

Murat Bastepe and Harald Jüppner

OUTLINE

Diagnosis and Progression of Parathyroid Hormone Resistance, 974
Pseudohypoparathyroidism Type-Ia, 975
  Albright's Hereditary Osteodystrophy and Multihormone Resistance, 975
  Coding Gsα Mutations as the Underlying Cause of Pseudohypoparathyroidism Type-Ia, 976
The Complex *Gnas* Locus, 977
  The Alpha Subunit of the Stimulatory G Protein (Gsα), 977
Clinically Distinct, Genetically Related Pseudohypoparathyroidism Variants, 979
  Pseudopseudohypoparathyroidism, 979
  Progressive Osseous Heteroplasia, 980
  Disorders of Concurrent Gsα Inactivity and Hyperactivity, 980
Pseudohypoparathyroidism Type-Ib, 982
  Autosomal Dominant PHP1B, 982
  Sporadic Pseudohypoparathyroidism Type-1B, 984
Pseudohypoparathyroidism Type-II, 984
Acrodysostosis With Hormonal Resistance, 985
Treatment, 986
Summary, 986

In 1942, Albright et al.[1] described a group of patients who displayed certain physical features, including obesity, short stature, brachydactyly, and cognitive impairment, combined with hypocalcemia and hyperphosphatemia. In these patients, exogenous, biologically active parathyroid hormone (PTH) extracts failed to result in a full phosphaturic response; hence, the term pseudohypoparathyroidism (PHP) was introduced, indicating that hypocalcemia and hyperparathyroidism in these patients resulted from target-organ resistance to, rather than deficiency of, PTH. Consistent with resistance to the actions of this hormone, it was later shown that patients affected by PHP have elevated concentrations of immunoreactive PTH.[2] Subsequently, it was also shown that some patients affected by PHP have resistance toward other hormones; however, PTH resistance is the most prominent feature of the disease.

The primary site of PTH resistance in PHP is the renal proximal tubule, as the actions of PTH in bone and the distal tubule appear normal.[3,4] Patients with PHP have reduced or inappropriately normal serum concentrations of 1,25-dihydroxyvitamin D (1,25(OH)$_2$D), which is the main cause of hypocalcemia.[5,6] Both low serum 1,25(OH)$_2$D and hyperphosphatemia are the direct results of PTH resistance at the proximal tubule. Hyperphosphatemia is typically worsened by the elevation of PTH in the circulation and the absence of resistance to the bone resorptive actions of this hormone, which leads to the release not only of calcium, but also of phosphate. Besides these unimpaired actions of PTH on bone, as well as unimpaired PTH-dependent calcium reabsorption in the renal distal tubule,[7,8] elevated serum PTH levels can prevent symptomatic hypocalcemia in PHP patients for extended periods of time. However, at some point in their lives, most patients manifest hypocalcemia and, therefore, present with associated clinical findings.

DIAGNOSIS AND PROGRESSION OF PARATHYROID HORMONE RESISTANCE

PTH exerts its actions by binding to a seven transmembrane domain, G protein-coupled receptor (the PTH/PTH-related protein receptor, PTHR1).[9] Although PTHR1 can couple to several different G proteins,[10] most PTH actions are mediated primarily through the stimulatory G protein, which acts on adenylyl cyclase, thereby increasing the formation of intracellular second messenger cyclic adenosine monophosphate (cAMP). PTH-induced cAMP formation is used as an important indicator of renal tubular PTH function, as most PHP patients display an inadequate or absent increase of urinary cAMP in response to exogenous, biologically active PTH (Fig. 57.1).[11] In fact, assessing the nephrogenous cAMP response to exogenously administered PTH is used as a test for establishing the diagnosis of PHP, and is more sensitive than measuring the increase in urinary phosphate excretion used in the original Ellsworth–Howard test. However, currently used high-sensitivity PTH assays often suffice to make the diagnosis when serum PTH is elevated in the presence of hypocalcemia and hyperphosphatemia. Nonetheless, depending on the nature of the nephrogenous response to exogenous PTH, PHP is subdivided into two main types. PHP type-I (PHP1) is defined by blunted urinary excretion of both cAMP and phosphate, and PHP type-II (PHP2) by blunted urinary excretion of phosphate only.[1,11]

Signs and symptoms of decreased serum calcium levels often reflect secondary prolongation of the QT interval on electrocardiogram and increased neuromuscular excitability leading to Trousseau's and Chvostek's sign, as well as laryngospasm. Although the most common manifestations of hypocalcemia include muscle tetany and spasms, findings vary markedly among patients. In more severe cases,

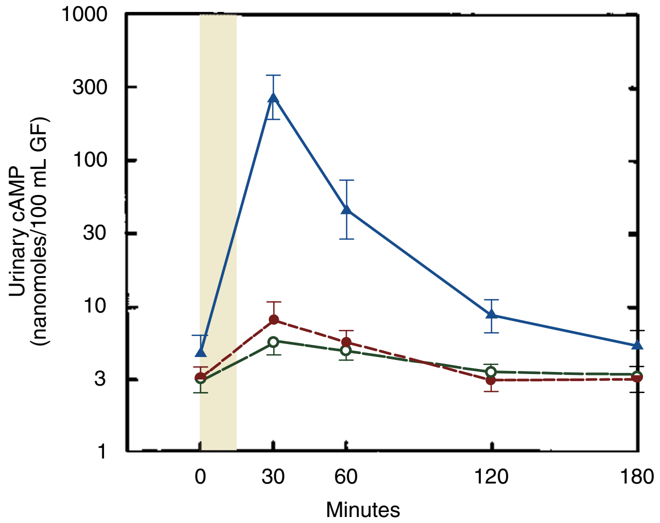


Fig. 57.1 Time-Course of Urinary Cyclic Adenosine Monophosphate (cAMP) Levels in Response to Parathyroid Hormone (PTH) Infusion. Upon injection of PTH, normal subjects (triangles) exhibit a 50- to 100-fold peak over the basal, while patients with pseudohypoparathyroidism type-Ia (open circles) or pseudohypoparathyroidism type-Ib (solid circles) demonstrate a blunted response. (Modified from Levine MA, Jap TS, Mauseth RS, et al. Activity of the stimulatory guanine nucleotide-binding protein is reduced in erythrocytes from patients with pseudohypoparathyroidism and pseudopseudohypoparathyroidism: Biochemical, endocrine, and genetic analysis of Albright's hereditary osteodystrophy in six kindreds. *J Clin Endocrinol Metab.* 1986;62:497–502.)

patients present with seizures. Other neurological symptoms can also arise from hypocalcemia, and some patients with PHP can initially be diagnosed with movement disorders.[12] In one report, two sisters with PHP1B (see later) presented with paroxysmal kinesigenic choreoathetosis, and the diagnosis of PTH resistance in one sister was made after approximately 4 years of oral antiepileptic treatment, when biochemical evaluation and genetic testing was performed.[13] Some patients presenting with seizures demonstrate epileptiform electroencephalogram changes, and, because this activity typically responds to antiepileptic treatment, the diagnosis of PHP can be delayed.[12] As another complication of the changes in serum calcium and phosphorous, brain imaging studies frequently show intracranial calcifications in PHP patients.[12]

PHP is a congenital disorder, but only few reports describe findings consistent with PTH resistance during the neonatal period.[14,15] Clinical manifestation of hypocalcemia typically occurs only later in childhood, suggesting that PTH resistance and the resultant changes in serum calcium and phosphorous develop gradually.[16-18] In a longitudinal study of a child with PHP1A (see later), cAMP response to PTH was found to be normal at age 3 months, whereas it was blunted at age 2.6 years.[8] In another PHP1A case, a gradual decline of serum calcium, preceded by increasing serum phosphorous and PTH levels, was demonstrated.[16] In addition, a PHP1B patient diagnosed by genetic analysis (see later) at birth was shown to have normal serum PTH levels until age 18 months, when an elevation of PTH was first detected despite normal serum calcium and phosphorous.[17] It thus appears that PTH responses are intact during the early postnatal period despite the existence of the molecular defect underlying PHP. The mechanisms that allow normal PTH signaling during this developmental stage remain unknown.

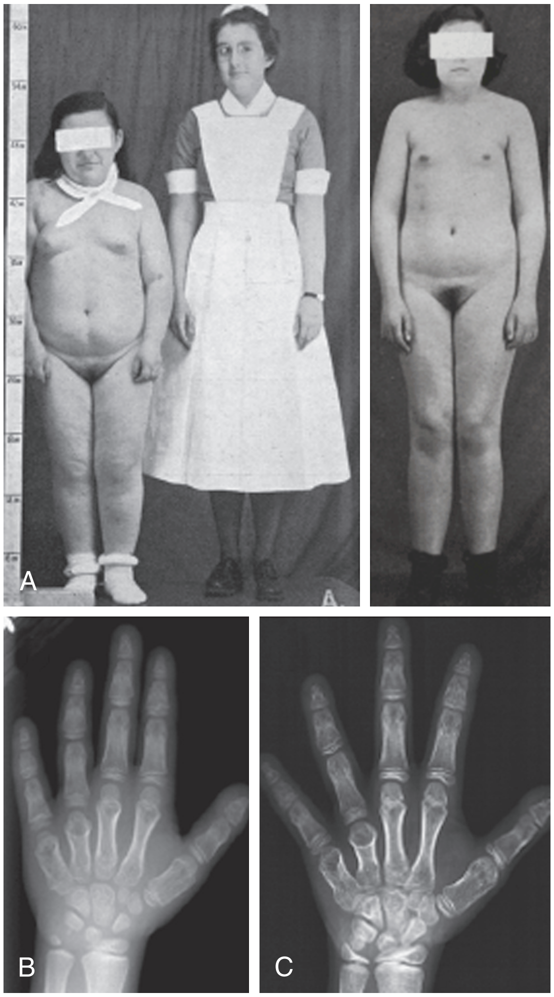


Fig. 57.2 Albright's hereditary osteodystrophy (AHO). **A,** Short stature and obesity are among the typical features of AHO. **B,** Hand radiograph of a child with AHO demonstrating short fourth and fifth metacarpals at age 5 11/12 years. **C,** The same patient's hand radiograph at age 8 10/12 years. (**A** Modified from Albright F, Burnett CH, Smith PH, et al. Pseudohypoparathyroidism — an example of "Seabright-Bantam syndrome". *Endocrinology.* 1942;30:922–932.)

PSEUDOHYPOPARATHYROIDISM TYPE-Ia

Of the two main PHP types, PHP1 is much more common. Various clinical variants of PHP1 have been described based on the presence or absence of clinical manifestations that coexist with PTH resistance, diminished stimulatory G protein activity in easily accessible cells, and silenced expression of the α-subunit of the stimulatory G protein (Gsα) from the *GNAS* complex gene (Table 57.1).

Albright's Hereditary Osteodystrophy and Multihormone Resistance

As described originally by Albright et al.,[1] PHP1 patients often exhibit characteristic physical features that may include obesity, round facies, short stature, brachydactyly, ectopic ossification, and intellectual disability (Fig. 57.2). These features are termed Albright's hereditary osteodystrophy (AHO) and occur primarily in those PHP patients, who are now classified as having PHP type-Ia (PHP1A). The brachydactyly observed in patients with AHO typically involves the metacarpal and/or metatarsal bones, and thus the pattern of shortening of the hand

TABLE 57.1 **Clinical and molecular features of the different PHP forms**

	PTH-resistance	Additional hormone resistance	Typical AHO features	Genetic defect
PHP1A/1C	Yes	Yes	Yes	*GNAS* mutations involving Gsα coding exons
PPHP	No	No	Yes	*GNAS* mutations involving Gsα coding exons
POH[a]	No	No	No	*GNAS* mutations involving Gsα coding exons
PHP1B	Yes	Some cases	No[b]	Microdeletions, duplications, inversions affecting *GNAS* imprinting, or patUPD20q[c]
PHP2	Yes	No	No	Unknown[d]

[a] Progressive osseous heteroplasia. Note that POH-like severe heterotopic ossifications have been reported in some patients with multiple hormone resistance.

[b] Some recent studies have demonstrated co-existence of AHO features and *GNAS* imprinting defects.

[c] Paternal uniparental disomy of chromosome 20q

[d] Mutations of *PRKAR1A*, *PDE4D*, or *PDE3A* are responsible for some variants of PHP2 associated with skeletal findings.

bones differs from other disorders with brachydactyly, such as familial brachydactyly and Turner's syndrome.[19] Because of shortened metacarpals, dimpling over the knuckles of a clenched fist (Archibald sign) is often observed.[20] The shortening of the distal phalanx of the thumb, however, is the most common skeletal abnormality (called "murderer's" or "potter's" thumb), and some patients have shortening of all digits.[21] Intellectual disability is mild, often presenting as cognitive defects. It is possible that the cause of cognitive impairment is the deficiency of Gsα signaling in the brain. Although hypocalcemia and/or hypothyroidism may also play a role in this phenotype, correction of these biochemical defects typically does not prevent it. There is remarkable patient-to-patient variability in AHO, even among patients that carry the same genetic mutation and belong to the same family (see later for discussion of the underlying genetic defect). Some patients may exhibit a single AHO feature only, such as obesity, while others may present with multiple different AHO features.[22] Furthermore, the severity of each feature differs vastly among the patients. In addition, individual AHO features are not unique to PHP, as they can be observed in other unrelated disorders, for example, acrodysostosis caused by *PDE4D*, *PDE3A*, *PTHLH*, or *PRKAR1A* mutations[23–26] or Brachydactyly-Mental Retardation Syndrome caused by *HDAC4* mutations.[27] The variable expressivity and the lack of specificity of individual features can make the AHO diagnosis challenging. Although the coexistence of hormone resistance in PHP1A patients is often helpful, it can also be misleading. This is particularly important in differential diagnosis of different PHP forms characterized by the presence of AHO features alone or hormone resistance alone (see later).

In addition to having PTH resistance and AHO, patients with PHP1A show clinical evidence that is consistent with target-organ resistance to other hormones. The most common additional hormone resistance involves the actions of thyroid-stimulating hormone (TSH), leading to hypothyroidism.[28,29] In fact, unlike PTH resistance, which typically develops later in life, TSH resistance can be present at birth.[30,31] Resistance towards gonadotropins and growth hormone–releasing factor has been reported,[32,33] whereas resistance toward other peptide hormones that also mediate their actions through Gsα-coupled receptors, such as vasopressin, calcitonin, or adrenocorticotropic hormone, does not appear to become clinically overt[12].

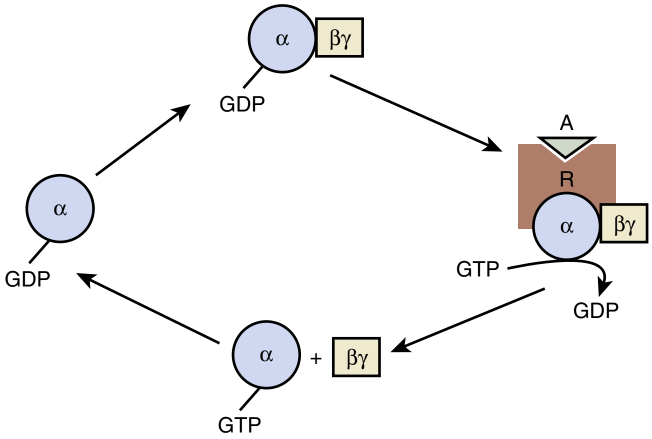


Fig. 57.3 Heterotrimeric G protein activation cycle. The heterotrimeric complex is assembled at the basal state, with the α-subunit being associated with a GDP molecule. Upon binding of an agonist (A) to its Gs-coupled receptor (R), the GDP molecule bound to the α-subunit is replaced with a GTP molecule, that is, the receptor acts as a guanine nucleotide exchange factor for the α-subunit. The GTP-bound form of the α-subunit dissociates from the βγ subunits and thereby stimulates its downstream effectors, which, in the case of Gsα, includes adenylyl cyclase. Note that the free Gβγ dimer can also stimulate different downstream effectors. The intrinsice GTP hydrolase activity of the α-subunit converts the GTP into GDP, resulting in the reassociation of the heterotrimer and, thereby, termination of effector stimulation.

Coding Gsα Mutations as the Underlying Cause of Pseudohypoparathyroidism Type-Ia

The genetic mutation causing PHP1A involves Gsα-coding *GNAS* exons.[34,35] A protein that is essential for the actions of many hormones, Gsα primarily mediates agonist-induced generation of intracellular cAMP. Activation of a stimulatory G protein coupled receptor by an agonist, such as PTH, leads to a GDP-GTP exchange on Gsα, causing dissociation of the latter from Gβγ subunits (Fig. 57.3). This allows both Gsα and Gβγ to stimulate their respective effectors. In its GTP-bound state, Gsα can directly activate several different effectors, such as Src tyrosine kinase,[36] and certain calcium channels.[37,38] Apart from these effectors, however, adenylyl cyclase is by far the most ubiquitous and the most extensively investigated effector molecule stimulated by Gsα. An integral membrane protein, adenylyl cyclase catalyzes the synthesis of the ubiquitous second messenger cAMP, which then triggers various cell-specific responses. The activation of adenylyl cyclase and other effectors by Gsα is regulated by the intrinsic GTP hydrolase (GTPase) activity of Gsα. Conversion of GTP into GDP results in the reassembly of the G protein heterotrimer, thereby preventing further effector stimulation (Fig. 57.3).

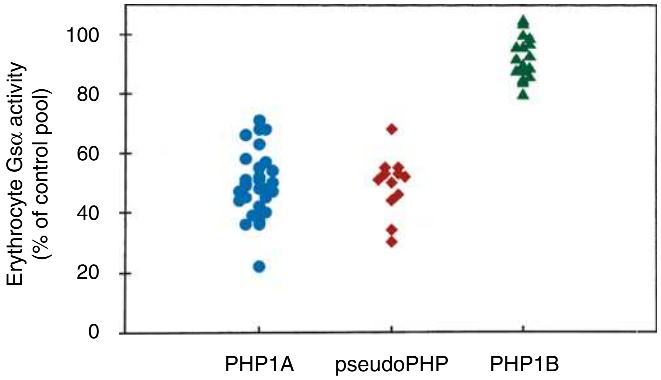


Fig. 57.4 Gsα erythrocyte activity in pseudohypoparathyroidism type-1a (PHP1A), pseudopseudohypoparathyroidism (*pseudo PHP*), and pseudohypoparathyroidism type-1b (PHP1B) patients. The measurement is conducted by complementation of patient-derived erythrocyte membranes with membranes from S49cyc- lymphoma cells, which genetically lack Gsα but retain all other components necessary for hormone-activated adenylyl cyclase response. (Modified from Levine M. Hypoparathyroidism and pseudohypoparathyroidism. In: DeGroot LJ, Jameson J, eds. *Endocrinology.* Philadelphia: W.B. Saunders; 2000:1133–1153.)

Mutations identified in PHP1A patients are heterozygous and scattered throughout all the 13 *GNAS* exons encoding Gsα and the intervening sequences, including missense and nonsense amino acid changes, insertions/deletions that cause frameshifts, and nucleotide alterations that disrupt pre-mRNA splicing. Constitutional deletions of the chromosomal arm containing *GNAS* have also been identified.[39] Among the many different inactivating *GNAS* mutations, a 4-bp deletion in exon 7 has been identified in numerous families, representing a genetic "hot spot." An extensive list of these mutations can be found at Online Mendelian Inheritance in Man entry #139320 (https://www.omim.org) and other gene-specific databases, such as Leiden Open Variation Database, https://databases.lovd.nl/shared/genes/GNAS). Consistent with this mutational spectrum, Gsα level/activity is reduced by approximately 50% in easily accessible tissues from PHP1A patients, such as erythrocytes, skin fibroblasts, and platelets.[40,41] Deficiency of Gsα has also been demonstrated in renal membranes from a patient with PHP.[42] A complementation assay is typically used for examining Gsα activity, involving patient-derived erythrocyte membranes and membranes from turkey erythrocyte that lack endogenous Gsα (Fig. 57.4). Gsα hypofunction can also be detected in patient-derived platelets by using a functional test that examines the ability of prostaglandin E1 or prostacyclin to inhibit platelet aggregation.[43] The detection of reduced Gsα activity is important for the establishment of PHP1A diagnosis, particularly in cases where genetic analysis fails to reveal a *GNAS* mutation. Reduction of Gsα activity in PHP1A is consistent with the fact that PTH and the other hormones with impaired actions in this disorder act via cAMP-mediated signaling pathways.

PHP1C has been described as a distinct variant of PHP1A,[44] but the clinical and laboratory features of patients with PHP1C are identical to those with PHP1A, as they have both AHO and multihormone resistance. In contrast to PHP1A, however, biochemical assays demonstrate no reduction in Gsα activity in erythrocytes obtained from PHP1C patients, suggesting the absence of mutations within the *GNAS* gene. Nevertheless, recent molecular characterizations have revealed Gsα mutations at least in some PHP1C patients. However, the Gsα mutants show only impaired coupling to different G protein–coupled receptors, yet show normal Gsα activity in complementation assays that use nonhydrolyzable GTP analogs for stimulation of Gsα activity rather than a receptor agonist.[21,45] Thus, the mutant Gsα protein causing PHP1C allows basal cAMP formation but is unable to trigger an increase in response to hormones. Hence, at least some PHP1C cases are allelic variants of PHP1A. Although it remains possible that some patients who match the description of PHP1C develop hormone resistance and AHO because of a novel gene mutation that affects cAMP production without functionally impairing Gsα itself—for example, inactivating mutations affecting one of the adenylyl cyclases or activating mutations in one of the phosphodiesterases—it appears unlikely that such putative mutations affect only few G protein-coupled receptors and thus cause limited hormonal resistance as observed in PHP1A and PHP1C, but instead dictate the phenotype by the tissue-specificity of their expression levels.

THE COMPLEX *GNAS* LOCUS

The Alpha Subunit of the Stimulatory G Protein (Gsα)

GNAS is a complex locus giving rise to multiple different coding and noncoding transcripts that show monoallelic, parent-of-origin–specific expression profiles (Fig. 57.5). *GNAS* maps to the telomeric end of the long arm of chromosome 20 (20q13.2-20q13.3).[46-48] Gsα is encoded by 13 exons,[49] but, because of alternative pre-mRNA splicing, the Gsα transcript has several variants. The long and short Gsα variants (Gsα-L and Gsα-S, respectively) differ from each other by the inclusion or exclusion of exon 3[49-51]; these Gsα variants are typically detected as 52- and 45-kDa protein bands on Western blots. Showing further complexity, each Gsα form either includes or excludes a CAG trinucleotide (encoding serine) at the start of exon 4. Small, but potentially important, differences have been reported between the activities of Gsα-L and Gsα-S. For example, Gsα-L has been shown to display greater ability to mediate receptor signaling than Gsα-S when partially purified proteins from rabbit liver were examined,[52] although the opposite finding was reported upon the use of cultured pancreatic islet cells.[53] Furthermore, the GDP release rate from Gsα-L appears to be approximately 2-fold higher than that of Gsα-S,[54] and, accordingly, fusion proteins involving the β2-adrenergic receptor and Gsα-L have shown higher constitutive activity than those involving the receptor and Gsα-S.[55] Moreover, differences in the subcellular targeting of these two Gsα variants have been reported.[56-58] Currently, it remains unclear whether these differences translate into biologically significant effects, such as divergence in the variety of effectors and/or the efficiency of effector activation. Nonetheless, a single nucleotide insertion in exon 3 leading to frameshift and early termination has been recently identified in two siblings affected by a mild form of PHP1A.[59] The mildness of the phenotype is consistent with generation of one of the two main Gsα variants, that is, Gsα-S. It remains unknown whether this exon 3 mutation impairs agonist responses in an effector- and tissue-specific manner, and this possibility depends on the putative effector selectivity and relative expression levels of Gsα-L and Gsα-S in different tissues.

Imprinted Products of the *GNAS* Locus. Recent studies have revealed that *GNAS* gives rise to multiple additional gene products that show parent-of-origin–specific, monoallelic expression. Besides Gsα, at least two translated *GNAS* transcripts exist, using distinct upstream promoters and alternative first exons that splice onto exons 2 to 13 encoding Gsα. The most upstream promoter relative to the Gsα promoter drives expression of a neuroendocrine secretory protein with an apparent molecular mass of 55 kDa (NESP55).[60] The paternal *NESP55* promoter is methylated, and transcription occurs from the maternal allele.[61,62] In humans, NESP55 is encoded by a single exon, so that Gsα exons 2 to 13 comprise the 3' untranslated region.[61,62] Expressed in neuroendocrine tissues, the peripheral and central nervous system,

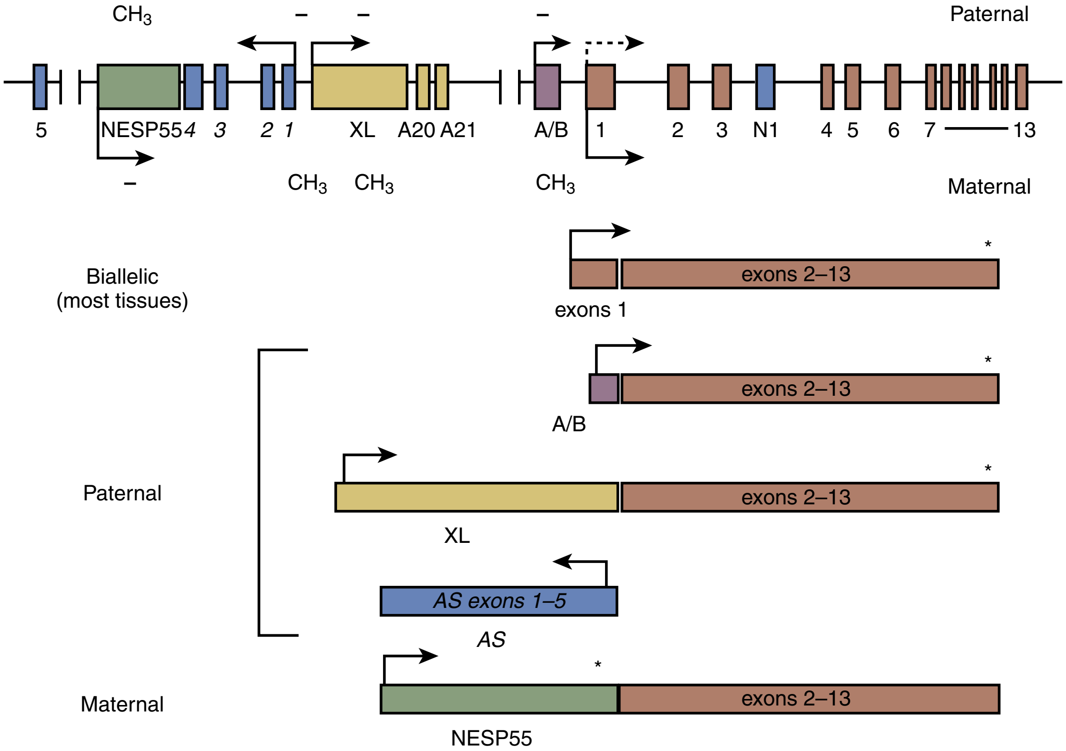


Fig. 57.5 The *GNAS* locus. The complex *GNAS* locus gives rise to multiple transcripts. Boxes and connecting lines indicate exons and introns, respectively. Arrows show the direction of transcription for each transcript. The main transcriptions derived from *GNAS* and the used exons are depicted as rectangles below the gene schematic. Gsα is biallelically expressed, except in a small number of tissues, including renal proximal tubules, thyroid, gonads, and pituitary, in which expression from the paternal allele is silenced (*dashed arrow*). XLαs, A/B, and antisense (*AS*) transcripts are derived from the paternal allele, and the NESP55 transcript from the maternal allele. Promoters of XLαs, A/B, antisense, and NESP55 transcripts are methylated on the silenced allele, as indicated by CH_3 (methylated CpGs) and - (unmethylated CpGs).

and some endocrine tissues,[60,63] NESP55 is a chromogranin-like protein associated with the constitutive secretory pathway.[64] Loss of the expression of NESP55 does not seem to have an overt clinical outcome in humans, as evidenced in patients with PHP1B (see later). However, its disruption in mice results in a subtle behavioral phenotype characterized by increased reactivity to novel environments.[65] Transcription from the *NESP55* promoter is likely to play a role in the regulation of *GNAS* imprinting (see later).

Another *GNAS* product is XLαs, which is expressed from the paternal allele.[62,66,67]. XLαs is expressed abundantly in various parts of the brain and neuroendocrine tissues, and its expression can be readily detected in a variety of other fetal and postnatal tissues.[68,69] XLαs also uses a distinct upstream promoter and a unique first exon that splices onto Gsα exons 2 to 13.[62,67] Unlike in *NESP55* transcript, however, the latter portion of the *XLαs* transcript is included in the translated product, resulting in a protein that is partially identical to Gsα.[66] Consistent with its structural similarity to Gsα, XLαs can mediate receptor-activated adenylyl cyclase stimulation in transfected cells and in transgenic mice.[70-72] In fact, it appears to be basally more potent than Gsα when expressed at levels comparable to the latter and is able to prolong PTH signaling in transfected cells.[73] However, the phenotype of mice with targeted ablation of paternal XLαs suggests that XLαs has unique, albeit as-yet-undefined, cellular functions. These mice show high early postnatal mortality because of poor adaptation to feeding and impairment in the glucose and energy metabolism,[74] in contrast to Gsα knockout mice, which to a large extent recapitulate findings in patients with PHP1A.[75,76] Findings from adult mice with disrupted XLαs expression indicate that this protein is a negative regulator of increased sympathetic nervous system activity in mice.[77] Recent studies have suggested novel cellular actions of XLαs, including the stimulation of PLC/PKC signaling and inhibition of clathrin-mediated endocytosis.[78,79]

The paternal *GNAS* allele also gives rise to two additional transcripts. From the sense strand originates the A/B transcript (also termed 1A or 1'), which, similar to NESP55 and XLαs, uses an upstream promoter and an alternative first exon (exon A/B) that splices onto Gsα exons 2 to 13.[80,81] Exon A/B does not comprise an in-frame translation initiation codon but, as demonstrated *in vitro*, the translation can be started through the use of an AUG located within exon 2, thereby giving rise to an N-terminally truncated protein that localizes to the plasma membrane when transiently expressed in cell lines.[81] Until recently, no evidence has supported the existence of an endogenous A/B protein, and therefore it was thought that the *A/B* transcript was a noncoding RNA. Endogenous A/B protein has recently been detected, however, in human fetal kidney by Western blot using an antibody against the C-terminal end of Gsα.[82] It remains possible that A/B has functions both as a protein and as a transcript regulating the expression within the *GNAS* locus (see later). Another paternal *GNAS* transcript is derived from the antisense strand.[83,84] The *GNAS* antisense transcript (*GNAS-AS1*), which is formed in humans by at least five primary exons that show multiple alternative splicing patterns,[83,85] is also considered to be noncoding. Data from mouse models indicate that *GNAS-AS1* plays a critical role in silencing the paternal *NESP55* promoter in *cis*.[86] The promoter of *GNAS-AS1* is immediately upstream of the promoter of *XLαs*. Although the promoters of *XLαs* and *GNAS-AS1* are located together within a large differentially methylated region (DMR), the female germline-specific imprint is established at the antisense promoter only.[87] The *A/B* promoter is likewise methylated in the female, but not in the male, germ cells.[88] Thus, the two gamete-specific imprint marks at the *GNAS* locus include the promoters of the *GNAS-AS1* and *A/B* transcripts. Accordingly, data from different genetically manipulated mouse models show that these noncoding transcripts are essential for the regulation of imprinted gene expression from *GNAS*.[86,89,90] Imprinted *A/B* transcription, that is, expression from only the nonmethylated maternal allele, is particularly important for the development of hormone resistance seen in patients with PHP1B (see later).

Unlike the promoters of the *NESP55*, *AS1*, *XLαs*, and *A/B* transcripts, the promoter of *Gsα* is not differentially methylated, and, accordingly, Gsα expression is biallelic in most tissues.[61,67,91] Biallelic Gsα expression has been specifically shown in human bone and white adipose tissue.[92] However, paternal Gsα expression is silenced in a small subset of tissues through as-yet-undefined mechanisms, so that the maternal allele is the predominant source of Gsα expression. These tissues include the renal proximal tubule, thyroid, pituitary, and gonads.[76,93-95] Although devoid of differential methylation, the Gsα promoter exhibits parent-of-origin specific histone modifications in those tissues where it is monoallelic. The active maternal Gsα promoter shows a greater ratio of tri- to dimethylated histone-3 Lys4 compared with the silenced paternal promoter in the proximal tubule, whereas the amount of methylated histones is similar in maternal and paternal Gsα promoters in liver, a tissue in which Gsα is biallelic.[96] As discussed later, the tissue-specific, paternal Gsα silencing has a key role in the development of PTH resistance in patients with PHP1A and PHP1B.

KEY POINTS

- The *GNAS* complex locus encodes the stimulatory G protein alpha-subunit (Gsα), which is critical for basal and receptor-activated cyclic adenosine monophosphate generation.
- Additional transcripts are also derived from *GNAS*, showing exclusively paternal or maternal expression (monoallelic) and differential methylation at their unique promoters. In contrast, Gsα expression is biallelic in most tissues, including bone and white adipose tissue; however, the paternal *Gsα* allele is silenced in a small set of tissues, including the renal proximal tubule, thyroid, brown adipocytes, gonads, and certain parts of the hypothalamus.
- Inactivating mutations affecting Gsα-coding *GNAS* exons cause multihormone resistance and Albright's hereditary osteodystrophy when located on the maternal allele, that is, pseudohypoparathyroidism type-Ia (PHP1A), whereas the same mutations cause some features of Albright's hereditary osteodystrophy (AHO) without hormone resistance when they occur on the paternal allele, that is, pseudopseudohypoparathyroidism (PPHP).
- AHO refers to a stellate of physical features including early-onset obesity, round face, short stature, brachydactyly, heterotopic ossification, and intellectual disability. Although obesity and intellectual disability appear to be present only when the Gsα mutation is maternal (i.e., part of PHP1A), the other AHO features are observed regardless of the parent of origin of the Gsα mutations (i.e., both PHP1A and PPHP).
- The underlying cause of this imprinted mode of inheritance in patients carrying Gsα mutations is the tissue-specific paternal silencing of the *Gsα* allele. In those tissues, the maternal inactivating mutation causes a dramatic loss of activity when combined with the already silenced paternal *Gsα* allele. The tissue-specific monoallelic expression of Gsα also explains the finding that, despite the ubiquitous expression of Gsα, hormone resistance is confined to a limited number of tissues.

CLINICALLY DISTINCT, GENETICALLY RELATED PSEUDOHYPOPARATHYROIDISM VARIANTS

Pseudopseudohypoparathyroidism

Physical abnormalities similar to those observed in patients with PHP1A but without evidence for an abnormal regulation of calcium and phosphate homeostasis were first reported in 1952.[97] Because of the lack of an abnormal regulation of mineral ion homeostasis, the name pseudopseudohypoparathyroidism (PPHP) was coined to describe this disorder.[97] Interestingly, patients with PPHP also carry *GNAS* mutations that lead to diminished Gsα function, and these mutations can be found in the same kindred as those with PHP1A. However, both disorders are never seen in the same sibling kinship, and a careful analysis of multiple families has revealed that the mode of inheritance of each disorder depends on the gender of the parent transmitting the Gsα mutations.[98] Thus, an inactivating Gsα mutation causes PHP1A (i.e., hormone resistance and AHO) when inherited maternally, whereas the same mutation results in PPHP (AHO only) if inherited paternally. Most AHO features, except obesity and intellectual disability, appear to develop regardless of the parent of origin, and it is therefore primarily hormone resistance that displays an imprinted mode of inheritance. Recent studies have furthermore revealed that most PPHP patients are considerably smaller at birth, particularly if their inactivating Gsα mutation is located in *GNAS* exons 2 to 13 of the paternal allele.[99,100]

The tissue-specific imprinting of Gsα expression can explain the parent-of-origin–specific inheritance of hormone resistance. In those tissues where Gsα expression is paternally silenced (i.e., Gsα is expressed exclusively or predominantly from the maternal allele), an inactivating mutation located on the paternal allele is not predicted to alter Gsα function, whereas the same mutation located on the maternal allele is predicted to abolish Gsα function completely (Fig. 57.6). The tissue-specific silencing of Gsα expression also explains why the target organ resistance involves only a small subset of tissues, despite the involvement of Gsα signaling in a multitude of physiological responses. Hormone resistance is observed in those tissues where Gsα

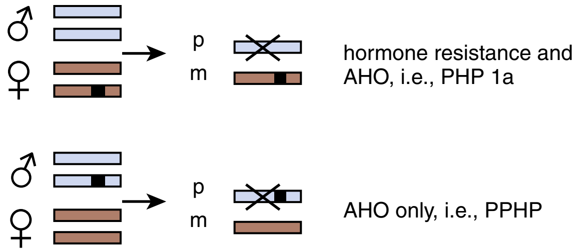


Fig. 57.6 The effect of paternal Gsα silencing in the development of hormone resistance. Gsα is biallelic in most tissues, and a heterozygous inactivating Gsα mutation therefore causes an approximately 50% reduction in Gsα activity/expression, regardless of the parent of origin of the mutation. However, in some tissues, such as the proximal tubule and thyroid, paternally inherited Gsα transcript is silenced (X). Thus, if a Gsα mutation (*black square*) is inherited from a female individual, this mutation nearly abolishes the expression or the activity of Gsα in those tissues, thus leading to hormone resistance (pseudohypoparathyroidism type-Ia; PHP1A). In contrast, upon paternal inheritance, the same mutation in Gsα does not lead to a significant change in the expression/activity, and thus the hormone responses are unimpaired. Paternal and maternal Gsα alleles are depicted by white or grey rectangles. *AHO*, Albright's hereditary osteodystrophy; *PPHP*, pseudopseudohypoparathyroidism.

is derived mainly from the maternal *GNAS* allele, such as the proximal renal tubule and the thyroid, while hormone responses are unimpaired in those tissues where Gsα is biallelic, such as the distal renal tubules and bone cells. The role of tissue-specific Gsα silencing in the development of PTH resistance has been demonstrated through the generation of mice heterozygous for maternal or paternal disruption of *GNAS*.[93] A recent study furthermore showed that the reduced paternal Gsα expression in the renal proximal tubule develops after the early postnatal period in mice,[101] thus providing a plausible explanation for the finding that the manifestation of PTH resistance occurs mostly after infancy in patients with PHP1A and PHP1B.[8,12,17]

The finding that most AHO features develop regardless of the parent transmitting a Gsα mutation has led to the hypothesis that the inheritance of AHO is caused by Gsα haploinsufficiency in various tissues, which appears to be true in certain settings. PTH-related protein–induced cAMP generation is critical for proper control of hypertrophic differentiation of growth plate chondrocytes,[102] and Gsα haploinsufficiency has been demonstrated in this tissue through the study of mice chimeric for wild-type cells and mutant cells heterozygous for disruption of *Gnas* exon 2.[103] Regardless of the parental origin of the *Gnas* exon 2 disruption, the mutant cells displayed premature hypertrophy compared with their wild-type neighbors, although the paternal disruption (i.e., loss of one *Gsα* allele combined with a complete loss of *XLαs*) resulted in significantly more premature hypertrophy than the maternal disruption (loss of one *Gsα* allele only). Thus, the brachydactyly and/or short stature observed in the context of AHO likely result from diminished Gsα signaling in growth-plate chondrocytes. Although these data correlate well with the notion that AHO develops after both maternal and paternal inheritance of a Gsα mutation, recent evidence suggests that individual AHO features can also be subject to imprinting. A careful analysis of multiple patients with PHP1A and PPHP revealed that obesity is primarily a feature of PHP1A patients, developing after maternal inheritance of the Gsα defect.[104] Considering that Gsα is biallelic in white adipose tissue,[92] it was proposed that Gsα may also be imprinted—predominantly maternal expression—in areas of the central nervous system that control satiety and body weight.[104] Recent reports have also demonstrated that cognitive impairment is encountered in PHP1A, but is not particularly prevalent in PPHP, thus indicating that tissue-specific Gsα imprinting may involve additional brain regions.[12,105] On the other hand, imprinted inheritance has not been reported regarding short stature, despite the finding that Gsα is maternally expressed in the pituitary[94,95] and the fact that PHP1A patients can display growth hormone-releasing hormone (GHRH) resistance that results in growth hormone (GH) deficiency.[33,106] Conversely, as outlined above, the small-for-gestational age phenotype is associated more strongly with *Gsα* mutations on the paternal allele than with *Gsα* mutations on the maternal allele.[100] A single study also showed that PPHP patients carrying mutations within exons 2 to 13 are born smaller than those carrying mutations within exon 1,[100] thus implicating both Gsα haploinsufficiency and the deficiency of a paternally expressed *GNAS* product, such as XLαs, in the pathogenesis. Future analyses of patients with PHP1A and PPHP will be helpful in determining the relative roles of genomic imprinting and haploinsufficiency in the development of individual AHO features.

Progressive Osseous Heteroplasia

A disorder termed progressive osseous heteroplasia (POH) has been described in patients with severe extraskeletal ossifications that involve deep connective tissue and skeletal muscle[107] (Fig. 57.7). In POH, the ectopic bone is primarily intramembranous, as opposed to a similar disease of heterotopic bone formation, termed fibrodysplasia ossificans progressiva, in which extraskeletal bone formation occurs via endochondral mechanisms and is accompanied by congenital skeletal malformations.[108] Few patients with POH demonstrate AHO features, and, consistent with the occasional coexistence of these two sets of clinical defects, heterozygous inactivating Gsα mutations have been identified as a cause of POH.[109,110] Several of the identified mutations are identical to those identified in PHP1A/PPHP kindreds.[109,110] Gsα activity and downstream signaling has been implicated in the control of osteogenic differentiation. Thus, patients who are mosaic for heterozygous *GNAS* mutations that result in constitutive Gsα activity develop fibrous dysplasia of bone characterized by irregular woven bone disrupted by fibrous tissue.[111] Moreover, in human mesenchymal stem cells, reduction of Gsα protein levels has been shown to cause osteogenic differentiation, while inhibiting the formation of adipocytes.[112,113] In addition, Runx2, a key regulator of osteoblast-specific gene expression, appears to suppress Gsα expression.[114] Thus, osteoprogenitor formation and the early stages of osteoblastic differentiation require reduced levels of cAMP signaling, consistent with the association of inactivating Gsα mutations with the severe ectopic bone formation observed in POH.

Because of the presence of *GNAS* mutations in both AHO and POH, it appears that additional factors, such as genetic background, epigenetic events, second hits, or environmental factors may determine the penetrance and severity of the ectopic ossifications in these patients who show approximately 50% loss of Gsα activity. Nevertheless, clinical and genetic data demonstrate several important differences between patients with AHO and those with POH. First, the ectopic bone in AHO is limited to subcutaneous tissue. In addition, in nearly all patients with POH, the severe ectopic bone formation is isolated (i.e., other typical AHO features are not manifest).[110,115] Moreover, mutations leading to isolated POH occurred either *de novo* on the paternal *GNAS* allele or are inherited from male obligate gene carriers only (i.e., the inheritance pattern is exclusively paternal).[110] In fact, in one large kindred, paternal inheritance of a *GNAS* mutation caused POH, while maternal inheritance of the same mutation caused typical AHO findings (without severe heterotopic ossification).[110] These findings indicate that the disease mechanism underlying POH is significantly different from that underlying AHO, and that deficiency of a *GNAS* product with exclusive paternal expression, such as XLαs, contributes to the molecular pathogenesis of POH. In addition, a study analyzing 12 patients with POH revealed that the lesions follow a predominantly dermomyotomal distribution that often shows a lateralization bias.[116] Based on these findings, it was hypothesized that the pathogenesis involves progenitor cells of somitic origin, which may undergo loss of heterozygosity at the *GNAS* locus and thereby cause severe Gsα deficiency.[116] Accordingly, complete ablation of Gsα in mouse mesenchymal tissues leads to POH-like skeletal findings.[117] In the ectopic bone lesions of POH patients, Hedgehog signaling is activated, which appears to be necessary and sufficient for the formation of Gsα-related ectopic bone in mouse models.[118]

Disorders of Concurrent Gsα Inactivity and Hyperactivity

Complex phenotypes have been described in several patients carrying *GNAS* mutations in whom the phenotype variably included PHP-like features combined with clinical findings that reflect hyperactivity of Gsα signaling and/or impaired Gsα activity in atypical tissues.

Testotoxicosis and Pseudohypoparathyroidism. A missense mutation in exon 13 (A366S) was identified in two unrelated boys who presented with PHP1A and precocious puberty.[119,120] This mutant Gsα

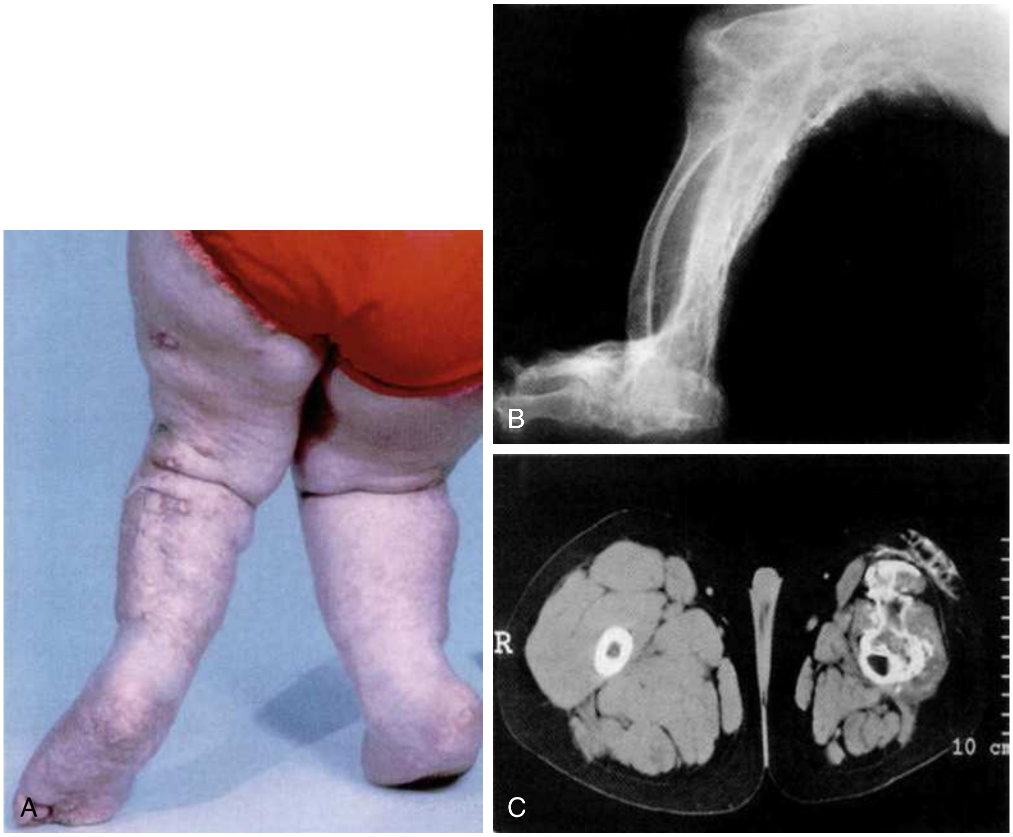


Fig. 57.7 Clinical and radiographic appearance of progressive osseous heteroplasia (POH). **A,** Posterior view of the legs and feet of a 5-year-old girl with POH, showing severe maculopapular lesions caused by extensive dermal and subcutaneous ossification. **B,** A lateral radiogram of the right leg of an 11-year-old girl with POH demonstrating severe heterotopic ossification of the soft tissues. **C,** Computed tomographic image of the thighs of a 10-year-old boy with POH, showing atrophied soft tissues in the left leg and extensive ossification of the skin, subcutaneous fat, and quadriceps muscles. (Modified from Shore EM, Ahn J, Jan de Beur S, et al. Paternally inherited inactivating mutations of the *GNAS1* gene in progressive osseous heteroplasia. *N Engl J Med*. 2002;346:99–106.)

protein is temperature-sensitive and thus rapidly degraded at normal body temperature. The amino acid substitution, however, renders the protein constitutively active, resulting in elevated cAMP signaling in the cooler temperature of the testis. More recently, two unrelated male patients have been reported to have another Gsα missense mutant (F376V) associated with severe early-onset precocious puberty combined with evidence of hormonal resistance.[121] Both patients presented with asymptomatic infantile-onset hyponatremia because of nephrogenic syndrome of inappropriate antidiuresis and also developed skeletal and growth plate abnormalities. As revealed by *in vitro* functional studies, the mutant protein exhibits enhanced receptor-dependent basal cAMP for several G protein-coupled receptors tested, including the lutropin/choriogonadotropin receptor, the PTHR1, and the vasopressin receptor type-2, while showing blunted ligand-induced activation. Thus, the observed clinical findings reflect both a loss and a gain of Gsα activity in a wide variety of tissues.

Other Complex Phenotypes of Altered Gsα Activity. Another mutant Gsα protein has been described in a unique case of familial PHP1A and transient neonatal diarrhea.[122] The mutation, which entails a repeated AVDT sequence at residues 366 to 369, generates an unstable but constitutively active Gsα mutant because of enhanced GDP-GTP exchange and reduced GTPase activity. Although hormone resistance results from the instability of the Gsα-AVDT mutant, the diarrhea is attributed to increased plasma membrane localization of the mutant protein in the small intestine epithelium. Another pediatric case has been described in which a *de novo* missense mutation (R231C) on the paternal allele was found together with a maternally inherited combination of three C-to-T substitutions resulting in aberrant *GNAS* splicing.[123] The patient with these compound heterozygous mutations had morbid obesity, TSH and PTH resistance, and a prothrombotic state owing to marked Gsα hypofunction in platelets.

KEY POINTS

- Heterozygous Gsα-coding mutations are rarely associated with severe ectopic bone formation, termed progressive osseous heteroplasia (POH). In those cases, the heterotopic bone invades the skeletal muscle and deep connective tissue, thus causing severe restriction of motion.
  - POH is mostly inherited paternally or found *de novo* in association with the mutant paternal *GNAS* allele.
    - The pathogenesis of POH involves progenitor cells of somitic origin, which may undergo loss of heterozygosity at the GNAS locus and thereby have severe Gsα deficiency.

PSEUDOHYPOPARATHYROIDISM TYPE-Ib

Another form of PHP was described by Peterman and Garvey[124], by Reynolds et al[124] and by Farfel et al.[43] Now known as PHP type-Ib (PHP1B), this PHP form is characterized by the presence of PTH resistant hypocalcemia and hyperphosphatemia, but without any evidence of AHO. In addition to increased serum PTH, patients with PHP1B can demonstrate elevated serum alkaline phosphatase activity, which suggests normal PTH-dependent bone turnover.[126] In fact, hyperparathyroid bone disease can be observed in association with PHP1B, especially in patients with sporadic PHP1B or in the index cases of the autosomal dominant form of PHP1B (AD-PHP1B). This occurs less frequently, however, in PHP1A, as this variant is associated with AHO features and is therefore diagnosed earlier in life. The intact PTH response in the bone is consistent with the lack of Gsα imprinting in bone[92] and led to the introduction of the term "pseudohypo-hyperpa rathyroidism."[127-129]

The hormone resistance observed in PHP1B patients develops only after maternal inheritance of the genetic defect[130] (i.e., the mode of inheritance is identical to the hormone resistance in PHP1A). PTH resistance and related changes in calcium and phosphate homeostasis are the major laboratory findings in PHP1B, but some of these patients also display mild hypothyroidism with slightly elevated TSH levels,[131-134] as well as some elevation in calcitonin level.[131] Hypothyroidism, as in patients with PHP1A, likely results from mild TSH resistance and is consistent with the predominantly maternal Gsα expression in the thyroid.[95,133] Nevertheless, evidence for resistance to other hormones, such as gonadotropins, whose actions also involve tissues in which paternal Gsα expression can be silenced, has not been reported for PHP1B patients. A study assessed GH response to GHRH plus arginine stimulation in PHP1B, revealing a normal response in 9 of 10 patients.[134] On the other hand, in addition to PTH and mild TSH resistance, hypouricemia because of increased fractional excretion of uric acid has been reported in the affected individuals of two unrelated PHP1B kindreds.[135,136] This finding implicates impaired PTH actions in the development of reduced uric acid levels in these patients, an interpretation that is consistent with two previous reports describing hyperuricemia in association with hyperparathyroidism.[137,138] However, hypouricemia resolved in one of the PHP1B kindreds following treatment with calcium and calcitriol.[136]

Patients with PHP1B display normal Gsα bioactivity/levels in easily accessible tissues.[43] Accordingly, coding Gsα mutations are ruled out in these patients. In one family, however, a mutation located in exon 13 (in-frame deletion of Ile^{382}, del382Ile) was reported, leading to the uncoupling of Gsα from the PTHR1 but not other receptors, including TSH receptor, luteinizing hormone receptor, and β-adrenergic receptor,[139] that were expressed in LLCPK cells, which are of renal origin and express endogenous Gsα. These findings suggested an isolated PTH resistance as seen in PHP1B, leading to the conclusion that the del382Ile mutation in Gsα represents a rare cause of PHP1B. However, this conclusion has been questioned, as the use of transfected mouse embryonic fibroblasts null for endogenous Gsα showed that del382Ile leads to uncoupling from not only PTHR1 but also the β-adrenergic receptor.[21] Because of a lack of Gsα mutations, and because Gsα activity/levels in easily accessible tissues are normal in PHP1B patients, inactivating mutations that affect the gene encoding PTHR1 was considered in the past.[140] However, several different studies have excluded such mutations as the cause of this disease.[141-144] Instead, homozygous or compound heterozygous inactivating mutations of *PTHR1* have been revealed as the cause of Blomstrand chondrodysplasia, an embryonic lethal disorder involving severe skeletal abnormalities.[145] Homozygous *PTHR1* mutations were furthermore identified in patients with Eiken syndrome, an autosomal recessive skeletal dysplasia,[146] and heterozyous inactivating *PTHR1* mutations were observed in several families with an autosomal dominant form of delayed tooth eruption,[146-149] that is, disorders without evidence for hypocalcemia and hyperphosphatemia because of impaired actions of PTH.

Based on genome-wide linkage analysis, the genetic defect underlying PHP1B was mapped to a region of chromosome 20q13.3 that comprises the *GNAS* locus,[130] but the critical interval excluded all the coding *GNAS* exons, including those that encode Gsα.[27,132] On the other hand, patients with PHP1B display epigenetic abnormalities within the *GNAS* locus.[132,150] The most consistent epigenetic defect is a loss of imprinting at exon A/B (also termed exon 1A), which is primarily found as an isolated defect in most familial PHP1B cases.[132,150] In addition, many sporadic, and some familial, PHP1B cases show additional loss of imprinting at the DMR comprising the XLαs and antisense promoters and gain of imprinting at the DMR comprising the NESP55 exon.[150,151] These abnormalities are associated with biallelic expression of A/B, XLαs, and antisense transcripts and the silencing of NESP55 transcript. Together with the genetic linkage data, these imprinting defects suggest that the mutation causing PHP1B disrupts a regulatory element that controls *GNAS* imprinting. However, evidence for incomplete penetrance regarding the *GNAS* imprinting defects has been reported in one kindred, in whom some individuals lacked loss of imprinting and were healthy despite maternally inheriting the disease-associated haplotype.[152] Thus, methylation abnormalities of *GNAS* are required for the development of PHP1B.

Autosomal Dominant PHP1B

In multiple familial PHP1B cases with isolated exon A/B loss of imprinting, a unique 3-kb microdeletion at the centromeric neighboring *STX16* locus has been identified[153] (Fig. 57.8). The deleted region comprises *STX16* exons 4 to 6 and is flanked by two direct repeats, which may underlie the mechanism whereby this deletion occurs. This is consistent with the presence of the same microdeletion in many unrelated families with different ethnic and racial origin.[12] In a single kindred, a different microdeletion within *STX16* has been reported, removing exons 2 through 4 and overlapping with the 3-kb microdeletion by approximately 1.3 kb.[17] Recently, larger maternally inherited deletions comprising either *STX16* exons 2 to 8 or all *STX16* and *NPEPL1* exons have been discovered in single families with this disorder.[154–156] Thus, the disruption of *STX16* appears to be a frequent genetic defect in cases with isolated loss of exon A/B imprinting. The parental origin of these *STX16* deletions correlates well with the mode of inheritance of PHP1B. It is maternally inherited in affected individuals and paternally inherited in unaffected carriers. This gene encodes syntaxin-16, a member of the SNARE family of proteins. However, *STX16* does not appear to be imprinted,[17,153] and it is therefore unlikely that the loss of one *STX16* copy leads to PHP1B. Instead, because the maternal inheritance is associated with loss of exon A/B imprinting on the same allele, these deletions are presumed to disrupt a cis-acting element controlling the establishment or maintenance of exon A/B imprinting. Other than genetic evidence, however, no currently available data corroborate this prediction. A mouse model carrying a deletion equivalent to the 3-kb *STX16* deletion in humans has been generated, but neither maternal nor paternal inheritance of this genetic alteration causes PTH resistance or any alterations in *GNAS* imprinting[157]; animals with the homozygous *Stx16* deletion are also healthy. It thus appears plausible that the imprinting control element of *GNAS* located within *STX16* in the human is not precisely at the same location in the mouse. Nonetheless, the absence of a phenotype in the *Stx16* deletion mice argues against a model in which syntaxin-16, the product of this gene, is required in the oocyte for proper exon A/B imprinting.

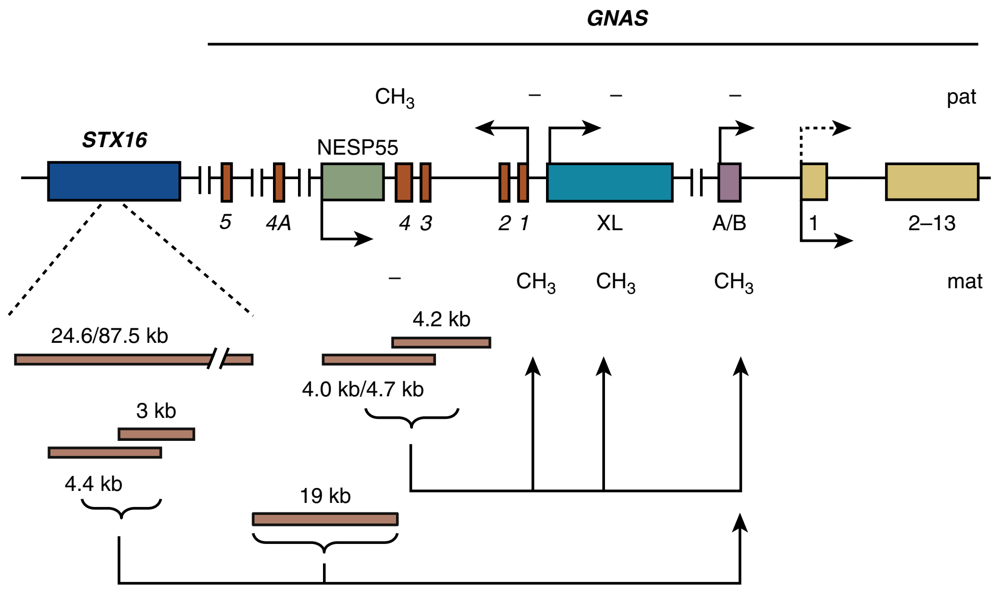


Fig. 57.8 Mutations identified in autosomal dominant pseudohypoparathyroidism type-Ib (AD-PHP1B) patients and their effects on *GNAS* imprinting. Most frequent mutation is a 3-kb deletion within *STX16*, a gene located more than 200 kb upstream of *GNAS*. This deletion and other overlapping deletions in the same gene are predicted to disrupt a cis-acting control element of *GNAS* that is required for the imprint mark located at exon A/B. The same prediction is also true for a 19-kb deletion of NESP55 and upstream genomic regions, which was identified in affected individuals of an AD-PHP1B kindred with isolated loss of A/B methylation. In addition, large duplications or genomic rearrangements involving the *GNAS* locus, such as a large inversion with a breakpoint in the intron between XLαs and A/B, have also been described as a cause of certain PHP1B cases that show loss of methylation at exon A/B alone. Deletions of the NESP55 DMR including exons 3 and 4 of the antisense transcript and a more recently identified deletion that only includes antisense transcript exons 3 and 4 have been identified in some AD-PHP1B kindreds. These reveal a cis-acting element controlling imprinting of the entire maternal *GNAS* allele. Boxes and connecting lines indicate exons and introns, respectively. *STX16* exons and *GNAS* exons 2–13 are shown as single rectangles for simplicity. Paternal (pat) and maternal (mat) methylation (CH_3) and parental origin of transcription (*arrows*) are indicated. Tissue-specific silencing of paternal Gsα transcription is depicted by a dotted arrow. The identified deletions are shown by brown horizontal bars.

In two unrelated familial cases of PHP1B in whom the affected individuals carried broad *GNAS* imprinting defects, maternally inherited deletions of the entire NESP55 DMR, including exons 3 and 4 of the antisense transcript, have been identified (Fig. 57.8).[85] The deletions are 4 kb and 4.7 kb in size and have breakpoints located in similar locations. The unaffected carriers in these families display an apparent loss of *NESP55* methylation because of the loss of this region from the normally methylated paternal allele, but do not show other imprinting *GNAS* defects. The affected individuals show a loss of imprinting in the entire maternal allele. The presence of similarly large deletions at the *NESP55* DMR has been excluded in a number of sporadic PHP1B cases.[85,151] However, a different 4.2-kb deletion has been identified in the affected individuals of a different PHP1B kindred who displayed broad *GNAS* imprinting defects (Fig. 57.8).[158] The latter deletion also includes antisense exons 3 and 4 but spares exon NESP55, overlapping with the previously identified two deletions by approximately 1.5 kb.[158] A small deletion adjacent to, but not affecting, NESP55 was also found in a kindred with two affected siblings who showed complete paternalization of the *GNAS* methylation pattern.[159] However, the clinically and epigenetically unaffected mother of the affected siblings had inherited the variant from her own mother, thus raising concerns that an as-yet identified mutation, rather than the identified deletion, is responsible for the disease in the siblings.[159]

A study revealed a 19-kb genomic deletion removing exon NESP55 and a large portion of the intron between NESP55 and antisense exon 4, but not including this exon, in a kindred with AD-PHP1B with isolated loss of A/B methylation.[160] Another 9-kb deletion removing NESP55 and the upstream genomic region has also been identified, which is also associated with isolated A/B methylation loss.[156] In addition, a 1.9-Mb genomic inversion has been discovered in a PHP1B kindred with isolated loss of *A/B* methylation, whose centromeric breakpoint disrupts the *GNAS* gene in its large intron separating the first exon of XLαs and A/B.[161] Furthermore, a family has been described in which a duplication within the *GNAS* locus spanning NESP55 and the first exons of XLαs led to loss-of-methylation at exon A/B alone, with methylation alterations in the other DMRs reflecting the duplication event.[162] Similarly, a duplication-triplication-duplication rearrangement involving approximately 95 kb of the *GNAS* locus, including all DMRs but not *STX16* or the Gsα coding region, has been identified in a PHP1B kindred, and this genetic alteration was associated with a complete loss of *A/B* methylation that could not be ascribed to the copy number changes.[163] Because all these patients lacked broad *GNAS* methylation abnormalities, the identified genetic alterations suggest the existence of another control element of specific to A/B methylation. The study of a mouse model in which Nesp55 transcription was prematurely truncated revealed loss of imprinting at exon 1A (the equivalent of human exon A/B) and, less consistently, the antisense promoter and exon XL.[164] Taken together with the genetic findings in PHP1B patients, it appears that the establishment of methylation imprints on the maternal *GNAS* allele is essential for allowing expression of Gsα in the proximal renal tubule and other tissues, including the thyroid, in which this *GNAS* product is derived only from the maternal allele. Furthermore, the evidence that exon *A/B* methylation requires transcription from the *NESP55* promoter[164] raises the possibility that even small deletions or point mutations can prevent the generation of *NESP55* pre-mRNA, thus leading to PHP1B.

Sporadic Pseudohypoparathyroidism Type-1B

Most cases of PHP1B are sporadic, (i.e., there are no affected family members). These cases all show broad *GNAS* methylation defects. Accordingly, approximately 10% of sporadic PHP1B cases have paternal uniparental isodisomy involving all of chromosome 20 or the parts that comprise the *GNAS* locus[131,165] Few patients were described to have *de novo* or maternally inherited genetic alterations affecting this gene. A large deletion has been identified in one case, comprising two adjacent deletions of approximately 38 kb and 1.4 kb and removing the region between the antisense exons 3 and 5, including exon NESP55; the deletion was found to be maternally inherited.[166] Additionally, a *de novo* large duplication (~320 kb) has been discovered, including the entire *GNAS* locus, as well as several neighboring genes, in a patient who displayed broad alterations in *GNAS* methylation.[162]

However, the maternal allele is frequently shared between affected and unaffected siblings, suggesting that these cases may carry small *de novo* mutations in this region. However, in some families the affected female passed either allele #1 or allele #2 to their children, who are all unaffected by PHP1B and show no methylation abnormalities at the *GNAS* locus.[167,168] It is therefore possible that some of the sporadic PHP1B cases carry homozygous or compound heterozygous mutations in an entirely different genomic location, resulting in a putative autosomal recessive mode of inheritance.[167] Consistent with this conclusion, few patients affected by imprinting disorders without obvious PTH resistance show *GNAS* methylation changes.[169,170] Moreover, methylation changes in some other imprinted genomic loci have been revealed in a study of sporadic PHP1B cases.[171] It has also been suggested that the broad *GNAS* imprinting defects observed in sporadic PHP1B patients result from stochastic defects in the regulation of imprinting.[172] Consistent with this hypothesis, the broad *GNAS* methylation defects are partial in some sporadic PHP1B cases, suggesting that those cases are mosaic for aberrant and normal cells. Investigating individual patients with incomplete methylation abnormalities, a recent study found no significant difference between organs and tissues in terms of the extent of the partial *GNAS* methylation defects, suggesting that those are acquired during early postzygotic development.[173]

Despite having distinct epigenetic abnormalities at the *GNAS* locus (i.e., isolated A/B loss of imprinting vs. broad imprinting defects that involve exon A/B and at least one other *GNAS* DMR), PHP1B patients seem to have similar clinical findings with respect to serum calcium, phosphate, and PTH levels.[151] Analysis of 20 families in which the affected individuals show an isolated loss of A/B imprinting revealed that a significant portion of such familial cases are asymptomatic at the time of diagnosis. In some of these cases, the diagnosis was made only based on elevated serum PTH. Comparison of male and female patients among sporadic PHP1B cases who exhibit *GNAS* imprinting defects at two or more *GNAS* DMRs also reveals that female patients have significantly higher serum PTH levels than male patients, suggesting that hormone resistance is more severe in females.[151]

Albright's Hereditary Osteodystrophy Features Associated With Cases With *GNAS* Methylation Defects. By definition, PHP1B patients do not show AHO features. However, some reports identified patients who carry genetic and epigenetic defects associated with PHP1B yet present with mild AHO features, particularly the shortness of metacarpal bones or Madelung-like deformities.[174,175]. This may suggest that Gsα imprinting occurs in more tissues than currently recognized, although there are alternative explanations. Considering that individual AHO features can be observed in other disorders, the presence of AHO features may be unrelated to the molecular genetic defects underlying PHP1B in these cases. In one case, the mother of two affected siblings with short metacarpals and loss of *A/B* methylation also exhibited short metacarpals, despite lacking any *GNAS* epigenetic abnormalities,[174] suggesting that the finding of short metacarpals is unrelated to the epigenetic defect in that family. In addition, the observed coexistence of *GNAS* imprinting defects and AHO can result from a large genomic deletion comprising at least the promoter of *Gsα* and one or more differentially methylated regions. In fact, such a large deletion leading to misdiagnosis of PHP1B has been discovered in a case with apparent loss of *A/B* methylation and AHO features.[176]

KEY POINTS

- Patients with pseudohypoparathyroidism type Ib (PHP1B) show renal parathyroid hormone resistance, often with mild thyroid-stimulating hormone resistance, typically in the absence of Albright's hereditary osteodystrophy. However, some PHP1B patients can have physical features that resemble Albright's hereditary osteodystrophy.
- PHP1B is often sporadic but can also be inherited in a maternal-specific autosomal dominant manner (AD-PHP1B).
- Alterations of the *GNAS* methylation pattern are associated with PHP1B. All PHP1B cases display a loss of methylation at *GNAS* exon A/B (also known as 1A). Most sporadic PHP1B (and a few AD-PHP1B) cases furthermore show methylation abnormalities at the other differentially methylated regions of *GNAS*.
- Various deletions, duplications, or rearrangements within or near the *GNAS* locus have been identified as the cause of the methylation alterations in PHP1B.
- The most frequent genetic cause of AD-PHP1B is a 3-kb microdeletion located within the neighboring *STX16* gene located ~200 kb centromeric of *GNAS*.
- The loss of *GNAS A/B* methylation, which is the most consistent epigenetic alteration, is predicted to silence, in cis, the expression of Gsα from the maternal allele. This leads to a dramatically reduced Gsα level in those tissues in which the paternal Gsα allele is already silenced via normal genomic mechanisms.

PSEUDOHYPOPARATHYROIDISM TYPE-II

Dissociation regarding the impairment of PTH-induced nephrogenous cAMP formation and phosphaturia (i.e., PHP2) appears to be the least common form of PHP. Although typically sporadic, a case with familial form of PHP2 type has been reported,[177] and several reports describe evidence for a self-limited form of this disease in newborns, which could indicate that it is transient in nature.[178-180] The molecular defect and pathophysiological mechanisms underlying this PHP variant remain to be discovered. Because the defect underlying PHP2 is associated with normal cAMP generation in response to exogenous PTH, it was postulated that it is caused by molecular defects that involve factors downstream of cAMP generation, such as protein kinase A (PKA).[181] In fact, mutations in the regulatory subunit of PKA have been identified in some patients who show, in association with characteristic skeletal abnormalities, biochemical defects similar to those seen in PHP2 (see later). Alternatively, the PTH signaling pathways that use other G proteins, such as Gq or G_{11}, may be defective in patients with PHP2. The signaling mediated by Gq/G_{11} pathway involves activation of phospholipase C, which in turn leads to the formation of second messengers inositol 1,4,5 tris-phosphate (IP3) and diacylglycerol. This signaling pathway, which results in the stimulation of protein kinase C (PKC) and an increase in intracellular calcium ions, was shown to be important in sustaining the phosphaturic actions on PTH, as recently shown for mice expressing a PTHR1 mutant that fails to activate IP3/PKC signaling.[182] Serum calcium levels, which may affect the efficient utilization of intracellular calcium signaling pathways, appear to be important for restoring PTH responsiveness in PHP2, as shown in some patients who normalized their phosphaturic response to PTH following normalization of serum calcium.[183] It is also possible that the sodium-phosphate transporters in the proximal renal tubule are nonresponsive to PTH, thereby resulting in a defective phosphaturic, but not cAMP, response to exogenous PTH. Such a defect, however, should preserve the action of PTH on 25(OH)D-1-α-hydroxylase and lead to normal serum 1,25$(OH)_2$D, unless it is combined with vitamin D deficiency. Hypocalcemia as a result of vitamin D deficiency has also been associated with PTH resistance that entailed the phosphaturic effect of this hormone without altering its potential to raise urinary cAMP,[184] suggesting that some PHP2 cases may in fact reflect vitamin D deficiency.[185,186]

ACRODYSOSTOSIS WITH HORMONAL RESISTANCE

A recent report described three patients who had resistance to PTH and some other hormones but showed unimpaired urinary cAMP excretion in response to exogenously administered recombinant PTH.[187] Consistent with resistance to the actions of PTH downstream of cAMP/PKA, PTH-induced urinary phosphate excretion was lost, and PTH-induced inhibition of urinary fractional calcium excretion was impaired. These unrelated patients, who could thus be classified as having a variant of PHP2 based on their renal PTH responsiveness, also displayed acrodysostosis, a skeletal dysplasia that resembles, but is more severe than, the skeletal defects observed in AHO. Acrodysostosis, also known as Maroteaux–Malamut syndrome, include severe brachydactyly, peripheral dysostosis, and nasal hypoplasia.[188,189] Analysis of DNA from these patients revealed a recurrent heterozygous nonsense mutation of the gene encoding cAMP-dependent protein kinase type I-alpha regulatory subunit (PRKAR1A; p.Arg368X). This mutation leads to a mutant PRKAR1A protein missing the last 14 C-terminal amino acids. Accordingly, cAMP cannot induce dissociation of this mutant regulatory subunit from the catalytic subunit, thereby keeping PKA in its inactive form. Several other patients with acrodysostosis have been described in more recent studies in which additional PRKAR1A mutations have been documented.[190,191] Note that the distal tubular actions of PTH appear to be impaired in patients with acrodysostosis, unlike in patients with PHP1A and PHP1B, who appear to show normal biallelic expression of Gsα in the distal renal tubule. Nevertheless, the similarity between this disorder and PHP1A with respect to the repertoire of hormone resistance is quite striking and highlights the critical nature of the cAMP signaling pathway in the actions of these hormones (Fig. 57.9).[192]

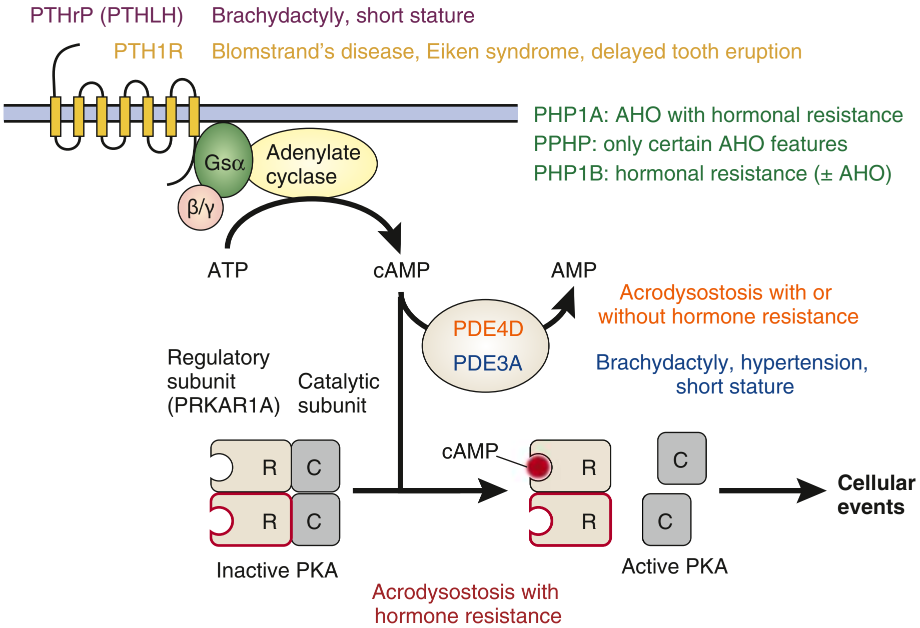

PTHrP (PTHLH)
Brachydactyly, short stature
PTH1R
Blomstrand's disease, Eiken syndrome, delayed tooth eruption
PHP1A: AHO with hormonal resistance
PPHP: only certain AHO features
PHP1B: hormonal resistance (± AHO)
Gsα
Adenylate cyclase
β/γ
ATP
cAMP
AMP
Acrodysostosis with or without hormone resistance
PDE4D
PDE3A
Brachydactyly, hypertension, short stature
Regulatory subunit (PRKAR1A)
Catalytic subunit
R
C
R
C
Inactive PKA
cAMP
R
R
C
C
Active PKA
Cellular events
Acrodysostosis with hormone resistance

Fig. 57.9 Mutated components of the cyclic adenosine monophosphate (cAMP) signaling pathway in pseudohypoparathyroidism and acrodysostosis. Gsα is essential for stimulation of cAMP generation, which binds to the regulatory subunits of protein kinase A (*PKA*) and thereby renders the catalytic subunits active. cAMP-specific phosphodiesterases 4D (*PDE4D*) and 3A (*PDE3A*) hydrolyze cAMP and thereby control its intracellular levels. Heterozygous mutations of *PTHLH*, encoding parathyroid hormone–related peptide (PTHrP), are responsible for brachydactyly and short stature, while homozygous or compound heterozygous mutations in the PTH/PTHrP receptor gene (*PTHR1*) are the cause of Blomstrand's lethal chondrodysplasia. Homozygous *PTHR1* deletions are associated with Eiken syndrome, and several heterozygous *PTHR1* mutations cause autosomal dominant delayed tooth eruption. Maternally inherited heterozygous inactivating mutations in Gsα-coding *GNAS* exons or deletions disrupting the imprinting of *GNAS* lead to pseudohypoparathyroidism type-Ia or pseudohypoparathyroidism type-Ib, respectively, resulting in resistance to parathyroid hormone, and frequently other hormones. Heterozygous mutations in *PRKAR1A* or *PDE4D* can also lead to resistance toward multiple hormones and are the cause of acrodysostosis. *PDE3A* mutations are responsible for severe hypertension and brachydactyly with short stature. AHO, Albright's hereditary osteodystrophy; ATP, adenosine triphosphate. (Modified from Silve C, Jüppner H. Genetic disorders caused by mutations in the PTH/PTHrP receptor and downstream effector molecules. In: Bilezikian JP, ed. *The Parathyroids*. 2014 Academic Press, San Diego.)

Defects in the gene encoding another protein within the cAMP signaling pathway, cAMP phosphodiesterase type 4D (PDE4D), have also been identified upon exome sequencing of DNA from some patients with acrodysostosis.[23,24,191,193,194] The heterozygous missense mutations in PDE4D are associated with acrodysostosis, but most patients lack endocrine abnormalities. The underlying biochemical mechanisms have yet to be elucidated. It is predicted, however, that the PDE4D mutations are gain-of-function, given that PDE4D limits the intracellular level of cAMP.[195] Interestingly, another subtype of cAMP phosphodiesterase is affected in another disorder: patients with heterozygous PDE3A mutations show severe hypertension and brachydactyly type-E with short stature.[25,196]

TREATMENT

The primary goal of treatment entails correction of abnormal serum biochemistries that result from PTH and, in some cases, other hormone resistance, such as TSH resistance leading to hypothyroidism, which can be treated by thyroid hormone replacement. GH deficiency can also be treated with recombinant human GH (rhGH) and is found to be efficacious in prepubertal patients with PHP1A.[12] PPHP patients who carry *de novo* mutations in the paternal *GNAS* allele are often diagnosed late in life (often when the children of female patients with PPHP are diagnosed with PHP1A), making it difficult to assess GH deficiency, and thus the benefits of rhGH treatment in PPHP patients. Clinical management of hypocalcemia in patients with PHP1A and PHP1B is less difficult than in patients with hypoparathyroidism because the distal tubular actions of PTH in PHP1 patients are not impaired, providing sufficient calcium reabsorption from the glomerular filtrate. The treatment involves oral calcium supplements and activated vitamin D analogs, such as $1,25(OH)_2D$ (calcitriol). Note that the active form of vitamin D is required, because of the lowered capacity of the proximal tubule to convert 25(OH)D into the biologically active $1,25(OH)_2D$. In addition, treatment of patients with PTH resistance should aim at keeping the serum PTH level within or close to the normal range, rather than simply avoiding symptomatic hypocalcemia, as persistent elevation of serum PTH will increase bone resorption and may eventually lead to hyperparathyroid bone disease.[197] Because of intact PTH actions in the distal tubule, urinary calcium levels are usually low, and affected individuals do not have a significant risk of developing kidney stones and nephrocalcinosis. In fact, during the course of the treatment, elevation of urinary calcium typically does not occur. Nevertheless, blood chemistries and urinary calcium excretion in patients undergoing treatment should be monitored annually, but more frequently during pubertal development and once skeletal growth is completed, as the requirements for treatment with calcium and $1,25(OH)_2D$ may need to be reduced.

SUMMARY

PHP refers to a group of disorders characterized by PTH resistance associated with hypocalcemia, hyperphosphatemia, and elevated serum PTH. Proximal tubular resistance to PTH is the most prominent hormonal defect, but, depending on the underlying genetic defect, resistance to other hormones is also observed. PTH and these other hormones all exert their actions via receptors that couple to Gsα. The primary genetic causes of PHP1 and related disorders are mutations that affect the complex *GNAS* locus, the gene encoding Gsα. These mutations result in decreased expression/activity of Gsα but also affect some of the other gene products derived from *GNAS*. The nature and the parental origin of the *GNAS* mutation is an important determinant of the clinical manifestations. Mutations that affect coding Gsα exons lead to broader clinical abnormalities than mutations that disrupt parent-specific *GNAS* methylation. Because of the tissue-specific silencing of paternal Gsα expression, hormone resistance develops only after maternal inheritance. AHO typically occurs after both maternal and paternal inheritance of coding Gsα mutations, although the repertoire of some AHO features also follows an imprinted mode of inheritance. PPHP is used to describe those patients with AHO who lack hormone resistance. AHO features vary markedly. An extreme manifestation of AHO is POH, which appears to develop when Gsα mutations are present on the paternal allele. PHP2 is rare, and the molecular determinants of at least few variants with concurrent skeletal dysplasia, that is, acrodysostosis, involve the regulatory subunit of PKA and cAMP phosphodiesterases.

REFERENCES

1. Albright F, Burnett CH, Smith PH, et al. Pseudohypoparathyroidism - an example of “Seabright-Bantam syndrome. *Endocrinology*. 1942;30:922–932.
2. Tashjian Jr AH, Frantz AG, Lee JB. Pseudohypoparathyroidism: assays of parathyroid hormone and thyrocalcitonin. *Proc Natl Acad Sci U S A*. 1966;56:1138–1142.
3. Ish-Shalom S, Rao LG, Levine MA, et al. Normal parathyroid hormone responsiveness of bone-derived cells from a patient with pseudohypoparathyroidism. *J Bone Miner Res*. 1996;11:8–14.
4. Murray T, Gomez Rao E, Wong MM, et al. Pseudohypoparathyroidism with osteitis fibrosa cystica: direct demonstration of skeletal responsiveness to parathyroid hormone in cells cultured from bone. *J Bone Miner Res*. 1993;8:83–91.
5. Breslau NA, Weinstock RS. Regulation of $1,25(OH)_2D$ synthesis in hypoparathyroidism and pseudohypoparathyroidism. *Am J Physiol*. 1988;255:E730–E736.
6. Drezner MK, Neelon FA, Haussler M, et al. 1,25-dihydroxycholecalciferol deficiency: the probable cause of hypocalcemia and metabolic bone disease in pseudohypoparathyroidism. *J Clin Endocrinol Metab*. 1976;42:621–628.
7. Balachandar V, Pahuja J, Maddaiah VT, et al. Pseudohypoparathyroidism with normal serum calcium level. *Am J Dis Child*. 1975;129:1092–1095.
8. Usardi A, Mamoune A, Nattes E, et al. Progressive development of PTH resistance in patients with inactivating mutations on the maternal allele of GNAS. *J Clin Endocrinol Metab*. 2017;102:1844–1850.
9. Jüppner H, Abou-Samra AB, Freeman MW, et al. A G protein-linked receptor for parathyroid hormone and parathyroid hormone-related peptide. *Science*. 1991;254:1024–1026.
10. Gardella TJ, Jüppner H, Brown EM, et al. Parathyroid hormone and parathyroid hormone receptor type 1 in the regulation of calcium and phosphate homeostasis and bone metabolism. In: DeGroot LJ, Jameson JL, eds. *Endocrinology*. Philadelphia, PA: W.B. Saunders Company; 2016:969–990.
11. Chase LR, Melson GL, Aurbach GD. Pseudohypoparathyroidism: defective excretion of 3’,5’-AMP in response to parathyroid hormone. *J Clin Invest*. 1969;48:1832–1844.
12. Mantovani G, Bastepe M, Monk D, et al. Diagnosis and management of pseudohypoparathyroidism and related disorders: first international Consensus Statement. *Nat Rev Endocrinol*. 2018;14:476–500.
13. Mahmud FH, Linglart A, Bastepe M, et al. Molecular diagnosis of pseudohypoparathyroidism type Ib in a family with presumed paroxysmal dyskinesia. *Pediatrics*. 2005;115:e242–e244.
14. Narang M, Salota R, Sachdev SS. Neonatal pseudohypoparathyroidism. *Indian J Pediatr*. 2006;73:97–98.
15. Sajitha S, Krishnamoorthy PN, Shenoy UV. Psedohypoparathyroidism in newborn - a rare presentation. *Indian Pediatr*. 2003;40:47–49.
16. Tsang R, Venkataraman P, Ho M, et al. The development of pseudohypoparathyroidism. Involvement of progressively increasing serum parathyroid hormone concentrations, increased 1,25-dihydroxyvitamin

D concentrations, and 'migratory' subcutaneous calcifications. *Am J Dis Child*. 1984;138:654–658.

17. Linglart A, Gensure RC, Olney RC, et al. A novel STX16 deletion in autosomal dominant pseudohypoparathyroidism type ib redefines the boundaries of a cis-acting imprinting control element of GNAS. *Am J Hum Genet*. 2005;76:804–814.
18. Gelfand IM, Eugster EA, Dimeglio LA. Presentation and clinical progression of pseudohypoparathyroidism with multi-hormone resistance and Albright hereditary osteodystrophy: a case series. *J Pediatr*. 2006;149:877–880.
19. Poznanski AK, Werder EA, Giedion A, et al. The pattern of shortening of the bones of the hand in PHP and PPHP—A comparison with brachydactyly E, Turner Syndrome, and acrodysostosis. *Radiology*. 1977;123:707–718.
20. Archibald RM, Finby N, De Vito F. Endocrine significance of short metacarpals. *J Clin Endocrinol Metab*. 1959;19:1312–1322.
21. Linglart A, Mahon MJ, Kerachian MA, et al. Coding GNAS mutations leading to hormone resistance impair in vitro agonist- and cholera toxin-induced adenosine cyclic 3',5'-monophosphate formation mediated by human XLas. *Endocrinology*. 2006;147:2253–2262.
22. Mendes de Oliveira E, Keogh JM, Talbot F, et al. Obesity-associated GNAS mutations and the melanocortin pathway. *N Engl J Med*. 2021;385:1581–1592.
23. Lee H, Graham Jr JM, Rimoin DL, et al. Exome sequencing identifies PDE4D mutations in acrodysostosis. *Am J Hum Genet*. 2012;90:746–751.
24. Michot C, Le Goff C, Goldenberg A, et al. Exome sequencing identifies PDE4D mutations as another cause of acrodysostosis. *Am J Hum Genet*. 2012;90:740–745.
25. Maass PG, Aydin A, Luft FC, et al. PDE3A mutations cause autosomal dominant hypertension with brachydactyly. *Nat Genet*. 2015;47:647–653.
26. Klopocki E, Hennig BP, Dathe K, et al. Deletion and point mutations of PTHLH cause brachydactyly type E. *Am J Hum Genet*. 2010;86:434–439.
27. Williams SR, Aldred MA, Der Kaloustian VM, et al. Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am J Hum Genet*. 2010;87:219–28.
28. Levine MA, Downs Jr RW, Moses AM, et al. Resistance to multiple hormones in patients with pseudohypoparathyroidism. Association with deficient activity of guanine nucleotide regulatory protein. *Am J Med*. 1983;74:545–556.
29. Mallet E, Carayon P, Amr S, et al. Coupling defect of thyrotropin receptor and adenylate cyclase in a pseudohypoparathyroid patient. *J Clin Endocrinol Metab*. 1982;54:1028–1032.
30. Levine MA, Jap TS, Hung W. Infantile hypothyroidism in two sibs: an unusual presentation of pseudohypoparathyroidism type Ia. *J Pediatr*. 1985;107:919–922.
31. Weisman Y, Golander A, Spirer Z, et al. Pseudohypoparathyroidism type 1a presenting as congenital hypothyroidism. *J Pediatr*. 1985;107:413–415.
32. Wolfsdorf JI, Rosenfield RL, Fang VS, et al. Partial gonadotrophin-resistance in pseudohypoparathyroidism. *Acta Endocrinol*. 1978;88:321–328.
33. Germain-Lee EL, Groman J, Crane JL, et al. Growth hormone deficiency in pseudohypoparathyroidism type 1a: another manifestation of multi-hormone resistance. *J Clin Endocrinol Metab*. 2003;88:4059–4069.
34. Weinstein LS, Gejman PV, Friedman E, et al. Mutations of the Gs alpha-subunit gene in Albright hereditary osteodystrophy detected by denaturing gradient gel electrophoresis. *Proc Natl Acad Sci U S A*. 1990;87:8287–8290.
35. Patten JL, Johns DR, Valle D, et al. Mutation in the gene encoding the stimulatory G protein of adenylate cyclase in Albright's hereditary osteodystrophy. *N Engl J Med*. 1990;322:1412–1419.
36. Ma YC, Huang J, Ali S, Lowry W, et al. Src tyrosine kinase is a novel direct effector of G proteins. *Cell*. 2000;102:635–646.
37. Yatani A, Imoto Y, Codina J, et al. The stimulatory G protein of adenylyl cyclase, Gs, also stimulates dihydropyridine-sensitive Ca2+ channels. Evidence for direct regulation independent of phosphorylation by cAMP-dependent protein kinase or stimulation by a dihydropyridine agonist. *J Biol Chem*. 1988;263:9887–9895.
38. Mattera R, Graziano MP, Yatani A, et al. Splice variants of the alpha subunit of the G protein Gs activate both adenylyl cyclase and calcium channels. *Science*. 1989;243:804–807.
39. Aldred MA, Aftimos S, Hall C, et al. Constitutional deletion of chromosome 20q in two patients affected with Albright hereditary osteodystrophy. *Am J Med Genet*. 2002;113:167–172.
40. Farfel Z, Brickman AS, Kaslow HR, et al. Defect of receptor-cyclase coupling protein in pseudohypoparathyroidism. *N Engl J Med*. 1980;303:237–242.
41. Levine MA, Downs Jr RW, Singer M, et al. Deficient activity of guanine nucleotide regulatory protein in erythrocytes from patients with pseudohypoparathyroidism. *Biochem Biophys Res Commun*. 1980;94:1319–1324.
42. Downs Jr RW, Levine MA, Drezner MK, et al. Deficient adenylate cyclase regulatory protein in renal membranes from a patient with pseudohypoparathyroidism. *J Clin Invest*. 1983;71:231–235.
43. Freson K, Izzi B, Labarque V, et al. GNAS defects identified by stimulatory G protein alpha-subunit signalling studies in platelets. *J Clin Endocrinol Metab*. 2008;93:4851–4859.
44. Farfel Z, Brothers VM, Brickman AS, et al. Pseudohypoparathyroidism: inheritance of deficient receptor-cyclase coupling activity. *Proc Natl Acad Sci U S A*. 1981;78:3098–3102.
45. Linglart A, Carel JC, Garabedian M, et al. GNAS1 lesions in pseudohypoparathyroidism ia and ic: genotype phenotype relationship and evidence of the maternal transmission of the hormonal resistance. *J Clin Endocrinol Metab*. 2002;87:189–197.
46. Gejman PV, Weinstein LS, Martinez M, et al. Genetic mapping of the Gs-a subunit gene (GNAS1) to the distal long arm of chromosome 20 using a polymorphism detected by denaturing gradient gel electrophoresis. *Genomics*. 1991;9:782–783.
47. Rao VV, Schnittger S, Hansmann I. G protein Gs alpha (GNAS 1), the probable candidate gene for Albright hereditary osteodystrophy, is assigned to human chromosome 20q12-q13.2. *Genomics*. 1991;10:257–261.
48. Levine MA, Modi WS, O'Brien SJ. Mapping of the gene encoding the alpha subunit of the stimulatory G protein of adenylyl cyclase (GNAS1) to 20q13.2—q13.3 in human by *in situ* hybridization. *Genomics*. 1991;11:478–479.
49. Kozasa T, Itoh H, Tsukamoto T, et al. Isolation and characterization of the human Gsa gene. *Proc Natl Acad Sci USA*. 1988;85:2081–2085.
50. Bray P, Carter A, Simons C, et al. Human cDNA clones for four species of G alpha s signal transduction protein. *Proc Natl Acad Sci U S A*. 1986;83:8893–8897.
51. Robishaw JD, Smigel MD, Gilman AG. Molecular basis for two forms of the G protein that stimulates adenylate cyclase. *J Biol Chem*. 1986;261:9587–9590.
52. Sternweis PC, Northup JK, Smigel MD, et al. The regulatory component of adenylate cyclase. Purification and properties. *J Biol Chem*. 1981;256:11517–11526.
53. Walseth TF, Zhang HJ, Olson LK, et al. Increase in Gs and cyclic AMP generation in HIT cells. Evidence that the 45-kDa alpha-subunit of Gs has greater functional activity than the 52-kDa alpha-subunit. *J Biol Chem*. 1989;264:21106–21111.
54. Graziano MP, Freissmuth M, Gilman AG. Expression of Gs alpha in Escherichia coli. Purification and properties of two forms of the protein. *J Biol Chem*. 1989;264:409–418.
55. Seifert R, Wenzel-Seifert K, Lee TW, et al. Different effects of Gsalpha splice variants on beta2-adrenoreceptor-mediated signaling. The beta2-adrenoreceptor coupled to the long splice variant of Gsalpha has properties of a constitutively active receptor. *J Biol Chem*. 1998;273:5109–5116.
56. Kvapil P, Novotny J, Svoboda P, et al. The short and long forms of the alpha subunit of the stimulatory guanine-nucleotide-binding protein are unequally redistributed during (-)-isoproterenol-mediated desensitization of intact S49 lymphoma cells. *Eur J Biochem*. 1994;226:193–199.
57. el Jamali A, Rachdaoui N, Jacquemin C, et al. Long-term effect of forskolin on the activation of adenylyl cyclase in astrocytes. *J Neurochem*. 1996;67:2532–2539.

58. Bourgeois C, Duc-Goiran P, Robert B, et al. G protein expression in human fetoplacental vascularization. Functional evidence for Gs alpha and Gi alpha subunits. *J Mol Cell Cardiol*. 1996;28:1009–1021.
59. Thiele S, Werner R, Ahrens W, et al. A disruptive mutation in exon 3 of the GNAS gene with Albright hereditary osteodystrophy, normocalcemic pseudohypoparathyroidism, and selective long transcript variant Gsalpha-L deficiency. *J Clin Endocrinol Metab*. 2007;92:1764–1768.
60. Ischia R, Lovisetti-Scamihorn P, Hogue-Angeletti R, et al. Molecular cloning and characterization of NESP55, a novel chromogranin-like precursor of a peptide with 5-HT1B receptor antagonist activity. *J Biol Chem*. 1997;272:11657–11662.
61. Hayward BE, Moran V, Strain L, et al. Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc Natl Acad Sci U S A*. 1998;95:15475–15480.
62. Peters J, Wroe SF, Wells CA, et al. A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proc Natl Acad Sci U S A*. 1999;96:3830–3835.
63. Li JY, Lovisetti-Scamihorn P, Fischer-Colbrie R, et al. Distribution and intraneuronal trafficking of a novel member of the chromogranin family, NESP55, in the rat peripheral nervous system. *Neuroscience*. 2002;110:731–745.
64. Fischer-Colbrie R, Eder S, Lovisetti-Scamihorn P, et al. Neuroendocrine secretory protein 55: a novel marker for the constitutive secretory pathway. *Ann N Y Acad Sci*. 2002;971:317–322.
65. Plagge A, Isles AR, Gordon E, et al. Imprinted Nesp55 influences behavioral reactivity to novel environments. *Mol Cell Biol*. 2005;25:3019–3026.
66. Kehlenbach RH, Matthey J, Huttner WB. XLas is a new type of G protein (Erratum in Nature 1995 375:253). *Nature*. 1994;372:804–809.
67. Hayward B, Kamiya M, Strain L, et al. The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc Natl Acad Sci U S A*. 1998;95:10038–10043.
68. Plagge A, Kelsey G, Germain-Lee EL. Physiological functions of the imprinted Gnas locus and its protein variants Galpha(s) and XLalpha(s) in human and mouse. *J Endocrinol*. 2008;196:193–214.
69. Bastepe M. Relative functions of Galphas and its extra-large variant XLalphas in the endocrine system. *Horm Metab Res*. 2012;44:732–740.
70. Klemke M, Pasolli H, Kehlenbach R, et al. Characterization of the extra-large G protein alpha-subunit XLalphas. II. Signal transduction properties. *J Biol Chem*. 2000;275:33633–33640.
71. Bastepe M, Gunes Y, Perez-Villamil B, et al. Receptor-mediated adenylyl cyclase activation through XLalphas, the extra-large variant of the stimulatory G protein alpha-subunit. *Mol Endocrinol*. 2002;16:1912–1919.
72. Liu Z, Segawa H, Aydin C, et al. Transgenic overexpression of the extra-large Gs{alpha} variant XL{alpha}s enhances Gs{alpha}-mediated responses in the mouse renal proximal tubule in vivo. *Endocrinology*. 2011;152:1222–1233.
73. Liu Z, Turan S, Wehbi VL, et al. Extra-long Galphas variant XLalphas protein escapes activation-induced subcellular redistribution and is able to provide sustained signaling. *J Biol Chem*. 2011;286:38558–38569.
74. Plagge A, Gordon E, Dean W, et al. The imprinted signaling protein XLalphas is required for postnatal adaptation to feeding. *Nat Genet*. 2004;36:818–826.
75. Chen M, Gavrilov O, Liu J, et al. Alternative Gnas gene products have opposite effects on glucose and lipid metabolism. *Proc Natl Acad Sci U S A*. 2005;102:7386–7391.
76. Germain-Lee EL, Schwindinger W, Crane JL, et al. A mouse model of Albright hereditary osteodystrophy generated by targeted disruption of exon 1 of the *Gnas* gene. *Endocrinology*. 2005;146:4697–4709.
77. Xie T, Plagge A, Gavrilova O, et al. The alternative stimulatory G protein alpha-subunit XLalphas is a critical regulator of energy and glucose metabolism and sympathetic nerve activity in adult mice. *J Biol Chem*. 2006;281:18989–18999.
78. He Q, Zhu Y, Corbin BA, et al. The G protein alpha subunit variant XLalphas promotes inositol 1,4,5-trisphosphate signaling and mediates the renal actions of parathyroid hormone in vivo. *Sci Signal*. 2015;8:ra84.
79. He Q, Bouley R, Liu Z, et al. The large G protein alpha-subunit XLαs limits clathrin-mediated endocytosis and regulates tissue iron levels in vivo. *Proc Natl Acad Sci U S A*. 2017;114:E9559–E9568.
80. Swaroop A, Agarwal N, Gruen JR, et al. Differential expression of novel Gs alpha signal transduction protein cDNA species. *Nucleic Acids Res*. 1991;19:4725–4729.
81. Ishikawa Y, Bianchi C, Nadal-Ginard B, et al. Alternative promoter and 5' exon generate a novel G_sa mRNA. *J Biol Chem*. 1990;265:8458–8462.
82. Puzhko S, Goodyer CG, Mohammad AK, et al. Parathyroid hormone signaling via Galphas is selectively inhibited by an NH(2) -terminally truncated Galphas: implications for pseudohypoparathyroidism. *J Bone Miner Res*. 2011;26:2473–2485.
83. Hayward BE, Bonthron DT. An imprinted antisense transcript at the human GNAS1 locus. *Hum Mol Genet*. 2000;9:835–841.
84. Wroe SF, Kelsey G, Skinner JA, et al. An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. *Proc Natl Acad Sci U S A*. 2000;97:3342–3346.
85. Bastepe M, Fröhlich LF, Linglart A, et al. Deletion of the NESP55 differentially methylated region causes loss of maternal GNAS imprints and pseudohypoparathyroidism type-Ib. *Nat Genet*. 2005;37:25–37.
86. Williamson CM, Turner MD, Ball ST, et al. Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster. *Nat Genet*. 2006;38:350–355.
87. Coombes C, Arnaud P, Gordon E, et al. Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Mol Cell Biol*. 2003;23:5475–5488.
88. Liu J, Yu S, Litman D, et al. Identification of a methylation imprint mark within the mouse gnas locus. *Mol Cell Biol*. 2000;20:5808–5817.
89. Williamson CM, Ball ST, Nottingham WT, et al. A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. *Nat Genet*. 2004;36:894–899.
90. Liu J, Chen M, Deng C, et al. Identification of the control region for tissue-specific imprinting of the stimulatory G protein alpha-subunit. *Proc Natl Acad Sci U S A*. 2005;102:5513–5518.
91. Zheng H, Radeva G, McCann JA, et al. Gas transcripts are biallelically expressed in the human kidney cortex: implications for pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*. 2001;86:4627–4629.
92. Mantovani G, Bondioni S, Locatelli M, et al. Biallelic expression of the Gsalpha gene in human bone and adipose tissue. *J Clin Endocrinol Metab*. 2004;89:6316–6319.
93. Yu S, Yu D, Lee E, et al. Variable and tissue-specific hormone resistance in heterotrimeric G_s protein a-subunit (G_sa) knockout mice is due to tissue-specific imprinting of the G_sa gene. *Proc Natl Acad Sci U S A*. 1998;95:8715–8720.
94. Hayward BE, Barlier A, Korbonits M, et al. Imprinting of the G(s) alpha gene GNAS1 in the pathogenesis of acromegaly. *J Clin Invest*. 2001;107:R31–R36.
95. Mantovani G, Ballare E, Giammona E, et al. The Gsalpha gene: predominant maternal origin of transcription in human thyroid gland and gonads. *J Clin Endocrinol Metab*. 2002;87:4736–4740.
96. Sakamoto A, Liu J, Greene A, et al. Tissue-specific imprinting of the G protein Gsalpha is associated with tissue-specific differences in histone methylation. *Hum Mol Genet*. 2004;13:819–828.
97. Albright F, Forbes AP, Henneman PH. Pseudo-pseudohypoparathyroidism. *Trans Assoc Am Physicians*. 1952;65:337–350.
98. Davies AJ, Hughes HE. Imprinting in Albright's hereditary osteodystrophy. *J Med Genet*. 1993;30:101–103.
99. Liu JJ, Russell E, Zhang D, et al. Paternally inherited gsalpha mutation impairs adipogenesis and potentiates a lean phenotype in vivo. *Stem Cell*. 2012;30:1477–1485.
100. Richard N, Molin A, Coudray N, et al. Paternal GNAS mutations lead to severe intrauterine growth retardation (IUGR) and provide evidence for a role of XLalphas in fetal development. *J Clin Endocrinol Metab*. 2013;98:E1549–E1556.
101. Turan S, Fernandez-Rebollo E, Aydin C, et al. Postnatal establishment of allelic Galphas silencing as a plausible explanation for delayed onset of parathyroid hormone resistance owing to heterozygous Galphas disruption. *J Bone Miner Res*. 2014;29:749–760.
102. Kronenberg HM. Developmental regulation of the growth plate. *Nature*. 2003;423:332–336.

103. Bastepe M, Weinstein LS, Ogata N, et al. Stimulatory G protein directly regulates hypertrophic differentiation of growth plate cartilage in vivo. *Proc Natl Acad Sci U S A*. 2004;101:14794–14799.
104. Long DN, McGuire S, Levine MA, et al. Body mass index differences in pseudohypoparathyroidism type 1a versus pseudopseudohypoparathyroidism may implicate paternal imprinting of Galpha(s) in the development of human obesity. *J Clin Endocrinol Metab*. 2007;92:1073–1079.
105. Mouallem M, Shaharabany M, Weintrob N, et al. Cognitive impairment is prevalent in pseudohypoparathyroidism type Ia, but not in pseudopseudohypoparathyroidism: possible cerebral imprinting of Gsalpha. *Clin Endocrinol*. 2008;68:233–239.
106. Mantovani G, Maghnie M, Weber G, et al. Growth hormone-releasing hormone resistance in pseudohypoparathyroidism type ia: new evidence for imprinting of the Gs alpha gene. *J Clin Endocrinol Metab*. 2003;88:4070–4074.
107. Kaplan FS, Shore EM. Progressive osseous heteroplasia. *J Bone Miner Res*. 2000;15:2084–2094.
108. Shore EM, Kaplan FS. Insights from a rare genetic disorder of extraskeletal bone formation, fibrodysplasia ossificans progressiva (FOP). *Bone*. 2008;43:427–433.
109. Yeh GL, Mathur S, Wivel A, et al. GNAS1 mutation and Cbfa1 misexpression in a child with severe congenital platelike osteoma cutis. *J Bone Miner Res*. 2000;15:2063–2073.
110. Shore EM, Ahn J, Jan de Beur S, et al. Paternally inherited inactivating mutations of the GNAS1 gene in progressive osseous heteroplasia. *N Engl J Med*. 2002;346:99–106.
111. Weinstein LS, Yu S, Warner DR, et al. Endocrine manifestations of stimulatory G protein alpha-subunit mutations and the role of genomic imprinting. *Endocr Rev*. 2001;22:675–705.
112. Lietman SA, Ding C, Cooke DW, et al. Reduction in Gsalpha induces osteogenic differentiation in human mesenchymal stem cells. *Clin Orthop Relat Res*. 2005:231–238.
113. Zhao Y, Ding S. A high-throughput siRNA library screen identifies osteogenic suppressors in human mesenchymal stem cells. *Proc Natl Acad Sci U S A*. 2007;104:9673–9678.
114. Bertaux K, Broux O, Chauveau C, et al. Runx2 regulates the expression of GNAS on SaOs-2 cells. *Bone*. 2006;38:943–950.
115. Adegbite NS, Xu M, Kaplan FS, et al. Diagnostic and mutational spectrum of progressive osseous heteroplasia (POH) and other forms of GNAS-based heterotopic ossification. *Am J Med Genet*. 2008;146A:1788–1796.
116. Cairns DM, Pignolo RJ, Uchimura T, et al. Somitic disruption of GNAS in chick embryos mimics progressive osseous heteroplasia. *J Clin Invest*. 2013;123:3624–3633.
117. Huso DL, Edie S, Levine MA, et al. Heterotopic ossifications in a mouse model of Albright hereditary osteodystrophy. *PloS One*. 2011;6:e21755.
118. Regard JB, Malhotra D, Gvozdenovic-Jeremic J, et al. Activation of Hedgehog signaling by loss of GNAS causes heterotopic ossification. *Nat Med*. 2013;19:1505–1512.
119. Iiri T, Herzmark P, Nakamoto JM, et al. Rapid GDP release from Gs in patients with gain and loss of function. *Nature*. 1994;371:164–168.
120. Nakamoto JM, Zimmerman D, Jones EA, et al. Concurrent hormone resistance (pseudohypoparathyroidism type Ia) and hormone independence (testotoxicosis) caused by a unique mutation in the G alpha s gene. *Biochem Mol Med*. 1996;58:18–24.
121. Biebermann H, Kleinau G, Schnabel D, et al. A new multisystem disorder caused by the Galphas mutation p.F376V. *J Clin Endocrinol Metab*. 2019;104:1079–1089.
122. Makita N, Sato J, Rondard P, et al. Human G(salpha) mutant causes pseudohypoparathyroidism type Ia/neonatal diarrhea, a potential cell-specific role of the palmitoylation cycle. *Proc Natl Acad Sci U S A*. 2007;104:17424–17429.
123. Freson K, Izzi B, Jaeken J, et al. Compound heterozygous mutations in the GNAS gene of a boy with morbid obesity, thyroid-stimulating hormone resistance, pseudohypoparathyroidism, and a prothrombotic state. *J Clin Endocrinol Metab*. 2008;93:4844–4849.
124. Peterman MG, Garvey JL. Pseudohypoparathyroidism; case report. *Pediatrics*. 1949;4:790.
125. Reynolds TB, Jacobson G, Edmondson HA, et al. Pseudohypoparathyroidism: report of a case showing bony demineralization. *J Clin Endocrinol Metab*. 1952;12:560.
126. Elrick H, Albright F, Bartter FC, et al. Further studies on pseudohypoparathyroidism: report of four new cases. *Acta Endocrinol*. 1950;5:199–225.
127. Costello JM, Dent CE. Hypo-hyperparathyroidism. *Arch Dis Child*. 1963;38:397–407.
128. Allen EH, Millard FJC, Nassim JR. Hypo-hyperparathyroidism. *Arch Dis Child*. 1968;43:295–301.
129. Frame B, Hanson CA, Frost HM, et al. Renal resistance to parathyroid hormone with osteitis fibrosa: "pseudohypohyperparathyroidism". *Am J Med*. 1972;52:311–321.
130. Jüppner H, Schipani E, Bastepe M, et al. The gene responsible for pseudohypoparathyroidism type Ib is paternally imprinted and maps in four unrelated kindreds to chromosome 20q13.3. *Proc Natl Acad Sci U S A*. 1998;95:11798–11803.
131. Bastepe M, Lane AH, Jüppner H. Paternal uniparental isodisomy of chromosome 20q (patUPD20q)—and the resulting changes in *GNAS1* methylation—as a plausible cause of pseudohypoparathyroidism. *Am J Hum Genet*. 2001;68:1283–1289.
132. Bastepe M, Pincus JE, Sugimoto T, et al. Positional dissociation between the genetic mutation responsible for pseudohypoparathyroidism type Ib and the associated methylation defect at exon A/B: evidence for a long-range regulatory element within the imprinted *GNAS1* locus. *Hum Mol Genet*. 2001;10:1231–1241.
133. Liu J, Erlichman B, Weinstein LS. The stimulatory G protein a-subunit Gsa is imprinted in human thyroid glands: implications for thyroid function in pseudohypoparathyroidism types 1A and 1B. *J Clin Endocrinol Metabol*. 2003;88:4336–4341.
134. Mantovani G, Bondioni S, Linglart A, et al. Genetic analysis and evaluation of resistance to thyrotropin and growth hormone-releasing hormone in pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*. 2007;92:3738–3742.
135. Laspa E, Bastepe M, Jüppner H, et al. Phenotypic and molecular genetic aspects of pseudohypoparathyroidism type Ib in a Greek kindred: evidence for enhanced uric acid excretion due to parathyroid hormone resistance. *J Clin Endocrinol Metab*. 2004;89:5942–5947.
136. Unluturk U, Harmanci A, Babaoglu M, et al. Molecular diagnosis and clinical characterization of pseudohypoparathyroidism type-Ib in a patient with mild Albright's hereditary osteodystrophy-like features, epileptic seizures, and defective renal handling of uric acid. *Am J Med Sci*. 2008;336:84–90.
137. Pepersack T, Jabbour N, Fuss M, et al. Hyperuricemia and renal handling of urate in primary hyperparathyroidism. *Nephron*. 1989;53:349–352.
138. Westerdahl J, Valdemarsson S, Lindblom P, et al. Urate and arteriosclerosis in primary hyperparathyroidism. *Clin Endocrinol*. 2001;54:805–811.
139. Wu WI, Schwindinger WF, Aparicio LF, et al. Selective resistance to parathyroid hormone caused by a novel uncoupling mutation in the carboxyl terminus of Gas: a cause of pseudohypoparathyroidism type Ib. *J Biol Chem*. 2001;276:165–171a.
140. Silve C, Santora A, Breslau N, et al. Selective resistance to parathyroid hormone in cultured skin fibroblasts from patients with pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*. 1986;62:640–644.
141. Schipani E, Weinstein LS, Bergwitz C, et al. Pseudohypoparathyroidism type Ib is not caused by mutations in the coding exons of the human parathyroid hormone (PTH)/PTH-related peptide receptor gene. *J Clin Endocrinol Metab*. 1995;80:1611–1621.
142. Suarez F, Lebrun JJ, Lecossier D, et al. Expression and modulation of the parathyroid hormone (PTH)/PTH-related peptide receptor messenger ribonucleic acid in skin fibroblasts from patients with type Ib pseudohypoparathyroidism. *J Clin Endocrinol Metab*. 1995;80:965–970.
143. Fukumoto S, Suzawa M, Takeuchi Y, et al. Absence of mutations in parathyroid hormone (PTH)/PTH-related protein receptor complementary deoxyribonucleic acid in patients with pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*. 1996;81:2554–2558.
144. Jan de Beur S, Ding C, LaBuda M, et al. Pseudohypoparathyroidism 1b: exclusion of parathyroid hormone and its receptors as candidate disease genes. *J Clin Endocrinol Metab*. 2000;85:2239–2246.

145. Blomstrand S, Claësson I, Säve-Söderbergh J. A case of lethal congenital dwarfism with accelerated skeletal maturation. *Pediatr Radiol*. 1985;15:141–143.
146. Duchatelet S, Ostergaard E, Cortes D, et al. Recessive mutations in PTHR1 cause contrasting skeletal dysplasias in Eiken and Blomstrand syndromes. *Hum Mol Genet*. 2005;14:1–5.
147. Decker E, Stellzig-Eisenhauer A, Fiebig BS, et al. PTHR1 loss-of-function mutations in familial, nonsyndromic primary failure of tooth eruption. *Am J Hum Genet*. 2008;83:781–786.
148. Frazier-Bowers SA, Simmons D, Wright JT, et al. Primary failure of eruption and PTH1R: the importance of a genetic diagnosis for orthodontic treatment planning. *Am J Orthod Dentofacial Orthop*. 2010;137:160 e161–e167; discussion 160–161.
149. Yamaguchi T, Hosomichi K, Narita A, et al. Exome resequencing combined with linkage analysis identifies novel PTH1R variants in primary failure of tooth eruption in Japanese. *J Bone Miner Res*. 2011;26:1655–1661.
150. Liu J, Litman D, Rosenberg M, et al. A GNAS1 imprinting defect in pseudohypoparathyroidism type IB. *J Clin Invest*. 2000;106:1167–1174.
151. Linglart A, Bastepe M, Jüppner H. Similar clinical and laboratory findings in patients with symptomatic autosomal dominant and sporadic pseudohypoparathyroidism type Ib despite different epigenetic changes at the GNAS locus. *Clin Endocrinol*. 2007;67:822–831.
152. Jan de Beur S, Ding C, Germain-Lee E, et al. Discordance between genetic and epigenetic defects in pseudohypoparathyroidism type 1b revealed by inconsistent loss of maternal imprinting at GNAS1. *Am J Hum Genet*. 2007;73:314–322.
153. Bastepe M, Fröhlich LF, Hendy GN, et al. Autosomal dominant pseudohypoparathyroidism type Ib is associated with a heterozygous microdeletion that likely disrupts a putative imprinting control element of GNAS. *J Clin Invest*. 2003;112:1255–1263.
154. Elli FM, de Sanctis L, Peverelli E, et al. Autosomal dominant pseudohypoparathyroidism type Ib: a novel inherited deletion ablating STX16 causes loss of imprinting at the A/B DMR. *J Clin Endocrinol Metab*. 2014;99:E724–E728.
155. Yang Y, Chu X, Nie M, et al. A novel long-range deletion spanning STX16 and NPEPL1 causing imprinting defects of the GNAS locus discovered in a patient with autosomal-dominant pseudohypoparathyroidism type 1B. *Endocrine*. 2020;69:212–219.
156. Danzig J, Li D, Jan de Beur S, et al. High-throughput molecular analysis of pseudohypoparathyroidism 1b patients reveals novel genetic and epigenetic defects. *J Clin Endocrinol Metab*. 2021;106:e4603–e4620.
157. Fröhlich LF, Bastepe M, Ozturk D, et al. Lack of Gnas epigenetic changes and pseudohypoparathyroidism type Ib in mice with targeted disruption of syntaxin-16. *Endocrinology*. 2007;148:2925–2935.
158. Chillambhi S, Turan S, Hwang DY, et al. Deletion of the GNAS antisense transcript results in parent-of-origin specific GNAS imprinting defects and phenotypes including PTH-resistance (Abstract No. 1052). Presented at the *30th Ann Meet Am Soc Bone Minerl Res*. Montreal. 2008.
159. Rezwan FI, Poole RL, Prescott T, et al. Very small deletions within the NESP55 gene in pseudohypoparathyroidism type 1b. *Eur J Hum Genet*. 2015;23:494–499.
160. Richard N, Abeguile G, Coudray N, et al. A new deletion ablating NESP55 causes loss of maternal imprint of A/B GNAS and autosomal dominant pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*. 2012;97:E863–E867.
161. Grigelioniene G, Nevalainen PI, Reyes M, et al. A large inversion involving GNAS exon A/B and all exons encoding Gsalpha is associated with autosomal dominant pseudohypoparathyroidism type Ib (PHP1B). *J Bone Miner Res*. 2017;32:776–783.
162. Perez-Nanclares G, Velayos T, Vela A, et al. Pseudohypoparathyroidism type Ib associated with novel duplications in the GNAS locus. *PloS One*. 2015;10:e0117691.
163. Nakamura A, Hamaguchi E, Horikawa R, et al. Complex genomic rearrangement within the GNAS region associated with familial pseudohypoparathyroidism type 1b. *J Clin Endocrinol Metab*. 2016;101:2623–2627.
164. Chotalia M, Smallwood SA, Ruf N, et al. Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev*. 2009;23:105–117.
165. Colson C, Decamp M, Gruchy N, et al. High frequency of paternal iso or heterodisomy at chromosome 20 associated with sporadic pseudohypoparathyroidism 1B. *Bone*. 2019;123:145–152.
166. Takatani R, Molinaro A, Grigelioniene G, et al. Analysis of multiple families with single individuals affected by pseudohypoparathyroidism type Ib (PHP1B) reveals only one novel maternally inherited GNAS deletion. *J Bone Miner Res*. 2016;31:796–805.
167. Fernandez-Rebollo E, Perez de Nanclares G, Lecumberri B, et al. Exclusion of the GNAS locus in PHP-Ib patients with broad GNAS methylation changes: evidence for an autosomal recessive form of PHP-Ib? *J Bone Miner Res*. 2011;26:1854–1863.
168. Cavaco BM, Tomaz RA, Fonseca F, et al. Clinical and genetic characterization of Portuguese patients with pseudohypoparathyroidism type Ib. *Endocrine*. 2010;37:408–414.
169. Mackay DJ, Callaway JL, Marks SM, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet*. 2008;40:949–951.
170. Bliek J, Verde G, Callaway J, et al. Hypomethylation at multiple maternally methylated imprinted regions including PLAGL1 and GNAS loci in Beckwith-Wiedemann syndrome. *Eur J Hum Genet*. 2009;17:611–619.
171. Perez-Nanclares G, Romanelli V, Mayo S, et al. Detection of hypomethylation syndrome among patients with epigenetic alterations at the GNAS locus. *J Clin Endocrinol Metab*. 2012;97:E1060–E1067.
172. Liu J, Nealon JG, Weinstein LS. Distinct patterns of abnormal GNAS imprinting in familial and sporadic pseudohypoparathyroidism type IB. *Hum Mol Genet*. 2005;14:95–102.
173. Elli FM, Bordogna P, Arosio M, et al. Mosaicism for GNAS methylation defects associated with pseudohypoparathyroidism type 1B arose in early post-zygotic phases. *Clin Epigenetics*. 2018;10:16.
174. de Nanclares GP, Fernandez-Rebollo E, Santin I, et al. Epigenetic defects of GNAS in patients with pseudohypoparathyroidism and mild features of Albright's hereditary osteodystrophy. *J Clin Endocrinol Metab*. 2007;92:2370–2373.
175. Sanchez J, Perera E, Jan de Beur S, et al. Madelung-like deformity in pseudohypoparathyroidism type 1b. *J Clin Endocrinol Metab*. 2011;96:E1507–E1511.
176. Fernandez-Rebollo E, Garcia-Cuartero B, Garin I, et al. Intragenic GNAS deletion involving exon A/B in pseudohypoparathyroidism type 1A resulting in an apparent loss of exon A/B methylation: potential for misdiagnosis of pseudohypoparathyroidism type 1B. *J Clin Endocrinol Metab*. 2010;95:765–771.
177. van Dop C. Pseudohypoparathyroidism: clinical and molecular aspects. *Semin Nephrol*. 1989;9:168–178.
178. Kruse K, Kustermann W. Evidence for transient peripheral resistance to parathyroid hormone in premature infants. *Acta Paediatr Scand*. 1987;76:115–118.
179. Lee CT, Tsai WY, Tung YC, et al. Transient pseudohypoparathyroidism as a cause of late-onset hypocalcemia in neonates and infants. *J Formos Med Assoc*. 2008;107:806–810.
180. Manzar S. Transient pseudohypoparathyroidism and neonatal seizure. *J Trop Pediatr*. 2001;47:113–114.
181. Drezner M, Neelon FA, Lebovitz HE. Pseudohypoparathyroidism type II: a possible defect in the reception of the cyclic AMP signal. *N Engl J Med*. 1973;289:1056–1060.
182. Guo J, Song L, Liu M, et al. Activation of a non-cAMP/PKA signaling pathway downstream of the PTH/PTHrP receptor is essential for a sustained hypophosphatemic response to PTH infusion in male mice. *Endocrinology*. 2013;154:1680–1689.
183. Kruse K, Kracht U, Wohlfart K, et al. Biochemical markers of bone turnover, intact serum parathyroid horn and renal calcium excretion in patients with pseudohypoparathyroidism and hypoparathyroidism before and during vitamin D treatment. *Eur J Pediatr*. 1989;148:535–539.
184. Rao DS, Parfitt AM, Kleerekoper M, et al. Dissociation between the effects of endogenous parathyroid hormone on adenosine 3',5'-monophosphate generation and phosphate reabsorption in hypocalcemia due to vitamin D depletion: an acquired disorder resembling pseudohypoparathyroidism type II. *J Clin Endocrinol Metab*. 1985;61:285–290.

185. Srivastava T, Alon US. Stage I vitamin D-deficiency rickets mimicking pseudohypoparathyroidism type II. *Clin Pediatr (Phila)*. 2002;41:263–268.
186. Shriraam M, Bhansali A, Velayutham P. Vitamin D deficiency masquerading as pseudohypoparathyroidism type 2. *J Assoc Physicians India*. 2003;51:619–620.
187. Linglart A, Menguy C, Couvineau A, et al. Recurrent PRKAR1A mutation in acrodysostosis with hormone resistance. *N Engl J Med*. 2011;364:2218–2226.
188. Robinow M, Pfeiffer RA, Gorlin RJ, et al. Acrodysostosis. A syndrome of peripheral dysostosis, nasal hypoplasia, and mental retardation. *Am J Dis Child*. 1971;121:195–203.
189. Maroteaux P, Malamut G. [Acrodysostosis]. *Presse Med*. 1968;76:2189–2192.
190. Silve C. Acrodysostosis: a new form of pseudohypoparathyroidism? *Ann Endocrinol*. 2015;76:110–112.
191. Michot C, Le Goff C, Blair E, et al. Expanding the phenotypic spectrum of variants in PDE4D/PRKAR1A: from acrodysostosis to acroscyphodysplasia. *Eur J Hum Genet*. 2018;26:1611–1622.
192. Silve C, Jüppner H, In: Bilezikian JP eds. *Genetic Disorders Caused by Mutations in the Parathyroid Hormone/parathyroid Hormone-Related Peptide Receptor, its Ligands, and Downstream Effector Molecules.* San Diego: Academic Press; 2020:1335–1360.
193. Linglart A, Fryssira H, Hiort O, et al. PRKAR1A and PDE4D mutations cause acrodysostosis but two distinct syndromes with or without GPCR-signaling hormone resistance. *J Clin Endocrinol Metab*. 2012;97:E2328–E2338.
194. Lynch DC, Dyment DA, Huang L, et al. Identification of novel mutations confirms PDE4D as a major gene causing acrodysostosis. *Hum Mutat*. 2013;34:97–102.
195. Conti M, Beavo J. Biochemistry and physiology of cyclic nucleotide phosphodiesterases: essential components in cyclic nucleotide signaling. *Annu Rev Biochem*. 2007;76:481–511.
196. Boda H, Uchida H, Takaiso N, et al. A PDE3A mutation in familial hypertension and brachydactyly syndrome. *J Hum Genet*. 2016;61:701–703.
197. Farfel Z. Pseudohypohyperparathyroidism-pseudohypoparathyroidism type ib. *J Bone Miner Res*. 1999;14:1016.
198. Levine M. Hypoparathyroidism and pseudohypoparathyroidism. In: DeGroot LJ, Jameson J, eds. *Endocrinology*. Philadelphia: W.B. Saunders; 2000:1133–1153.

58

Genetic Defects in Vitamin D Metabolism and Action

René St-Arnaud, Glenville Jones, and Francis H. Glorieux

OUTLINE

Overview of Vitamin D Metabolism, 992
Rickets and Osteomalacia, 993
Vitamin D–Deficiency Rickets, Vitamin D Hydroxylation–Deficient Rickets, and Vitamin D–Dependent Rickets, 993
  Clinical Features of Vitamin D Hydroxylation–Deficient Rickets Type 1B, 993
  Clinical Features of Vitamin D Hydroxylation-Deficient Rickets Type 1A, 994
  Clinical Features of Vitamin D–Dependent Rickets Type 2A, 995
Molecular Etiology of Vitamin D Hydroxylation–Deficient Rickets Type 1B, 995
  CYP2R1 Mutations, 996
Treatment of Vitamin D Hydroxylation–Deficient Rickets, Type 1B, 996
Molecular Etiology of Vitamin D Hydroxylation–Deficient Rickets Type 1A, 997
  CYP27B1 Mutations, 997
Treatment of Vitamin D Hydroxylation–Deficient Rickets, Type 1A, 998
  Molecular Etiology of Vitamin D–Dependent Rickets Type 2A, 998
  Vitamin D Receptor Mutations, 999
Treatment of Vitamin D-Dependent Rickets, Type 2A, 1001
Idiopathic Infantile Hypercalcemia, 1001
  Molecular Etiology of Idiopathic Infantile Hypercalcemia, 1002
  CYP24A1 Mutations, 1002
Treatment of Idiopathic Infantile Hypercalcemia, 1003
Future Directions, 1004

Vitamin D is a key regulator of mineral homeostasis[1,2] and bone development,[3] and perturbation of the vitamin D endocrine system leads to impaired calcium regulation that can lead to rickets and/or osteomalacia. In contrast, failure to destroy the active form of vitamin D leads to overexpression of the vitamin D–regulated genes and hypercalcemia. To gain a complete perspective of the clinical manifestations of genetic anomalies involving vitamin D endocrine function, this chapter will first present a short overview of the salient aspects of the vitamin D metabolic pathway. These notions will lay the groundwork for the discussion of the clinical, pathophysiological, and molecular aspects of hereditary diseases involving the vitamin D endocrine system.

OVERVIEW OF VITAMIN D METABOLISM

Following exposure to sunlight, both plants and animals are able to synthesize vitamin D. Vitamin D_2 is generated in yeast and plants; vitamin D_3 is produced in fish and mammals. The slight differences in the chemical structure of the two compounds do not affect function or metabolism. The generic term of vitamin D (without subscript) will be used hereafter.

In humans, a sizeable proportion of vitamin D requirements can be produced endogenously in the skin upon exposure to ultraviolet light (sunlight). It has been shown, however, that at latitudes where vitamin D synthesis is reduced or absent during winter months, there is a seasonal variation in the photosynthesis of vitamin D.[4] People who receive an ample supply of sunlight during the rest of the year are not at risk of developing vitamin D deficiency, because excess cutaneously produced vitamin D is stored in fat and muscle and released at times of need.[5] Dietary sources such as fish, plants, and grains can help meet vitamin D requirements. In the absence of any exposure to sunlight, however, such as in elderly people, a daily multivitamin that contains 800 IU of vitamin D is indicated.[6]

Ultraviolet B photons (270–290 nm) penetrate the epidermis and photolize 7-dehydrocholesterol into previtamin D, which rapidly becomes a more thermodynamically stable molecule, vitamin D.[7] Vitamin D then exits the keratinocyte cells and enters the dermal capillary bed, where it becomes bound to vitamin D–binding protein (DBP). Once associated with DBP in the circulation, vitamin D is transported to the liver, where the cytochrome P450 25-hydroxylase enzyme CYP2R1 adds a hydroxyl group on carbon 25 to produce 25-hydroxyvitamin D [25(OH)D]. Although several cytochrome P450 molecules can mediate the 25-hydroxylation of vitamin D to 25(OH)D *in vitro*,[8-11] findings from patients with vitamin D hydroxylation–deficient rickets, type 1B (VDDR1B), discussed later, confirm that CYP2R1 is the principal human vitamin D 25-hydroxylase.[12-15]

The 25(OH)D metabolite also circulates in the bloodstream bound to DBP. It is an abundant, but relatively inactive, vitamin D metabolite. Its circulating concentration provides the most readily available evaluation of the vitamin D status in a given individual. 25(OH)D must be further hydroxylated on a different carbon residue in the convoluted and straight portions of the proximal kidney tubule to gain hormonal bioactivity. Hydroxylation at position 1α by the mitochondrial cytochrome P450 enzyme 25-hydroxyvitamin D-1α-hydroxylase (CYP27B1) converts 25(OH)D to 1α,25-dihydroxyvitamin D [1,25(OH)$_2$D] (Fig. 58.1), the active, hormonal form of vitamin D that plays an essential role in mineral homeostasis, cellular differentiation, and bone mineralization.[16,17]

Upon reaching a target tissue, $1,25(OH)_2D$ binds a specific receptor (vitamin D receptor [VDR]) that is a member of the nuclear hormone receptor superfamily.[16-18] The VDR is considered a class II nuclear hormone receptor, because it needs to form a heterodimer with the retinoid X receptor (RXR) to bind specific DNA sequence elements with high affinity. These target sequences are termed vitamin D response elements, and the best characterized of these binding sites consist of two tandemly repeated hexanucleotide sequences separated by three base pairs.[18] Transcriptional coactivators and components of the basal transcriptional machinery interact with the liganded, DNA-bound VDR-RXR heterodimer to activate the transcription of vitamin D target genes responsible for carrying out the physiological actions of $1,25(OH)_2D$.[16,18] Amongst several target genes, the $1,25(OH)_2D$ hormone induces in target cells the expression of the gene encoding the key effector of its catabolic breakdown: 25-hydroxyvitamin D-24-hydroxylase (CYP24A1).[19,20] This ensures attenuation of the $1,25(OH)_2D$ biological signal inside target cells and helps regulate vitamin D homeostasis.

As detailed in this chapter, disease-causing mutations have been identified in genes encoding all components of this metabolic and signal transduction cascade: *CYP2R1*, *CYP27B1*, *VDR*, and *CYP24A1*.

KEY POINTS

- Disease-causing mutations have been identified in genes encoding all components of the vitamin D metabolic and signal transduction cascade: *CYP2R1*, *CYP27B1*, *VDR*, and *CYP24A1*.

RICKETS AND OSTEOMALACIA

The term "rickets" is often used to describe all of the skeletal abnormalities associated with defective mineralization in the growing skeleton, but it is more precise to restrict the term to changes in the growth plate and adjacent metaphysis. When mineralization is impaired, the accumulation of unmineralized osteoid at sites other than the growing metaphysis should be referred to as osteomalacia, not as rickets. Thus, defective mineralization can lead to both rickets and osteomalacia in the growing skeleton, but only to osteomalacia in the mature skeleton.

Rickets is characterized by the inadequate calcification of the growth plate and adjacent metaphysis. The impaired mineralization of the growth plate cartilage in the zone of provisional calcification prevents this zone from being resorbed. As the cartilage continues to be formed, but not resorbed, the growth plate begins to widen. Simultaneously, the trabecular bone directly underneath the cartilage fails to mineralize properly, and vascularization of this tissue becomes aberrant. These defects are accompanied by similar abnormalities in cortical bone, leading to the full spectrum of skeletal symptoms associated with the pathology.

Some nomenclature: "rickets resistant to vitamin D therapy" was described clinically by Albright et al. in 1937.[21] Subsequent reports[22,23] suggested a variant of resistant rickets that was fully clinically described by Prader et al. in 1961.[24] The absence of knowledge of the underlying molecular etiology of these diseases led to numerous, somewhat confusing designations: pseudo vitamin D–deficiency rickets,[24] vitamin D dependency,[25] vitamin D–dependent rickets type I, or 1α-hydroxylase deficiency.[26] In 1997, remarkable progress was made in the understanding of the molecular etiology of the disease through the identification of mutations in the gene encoding the 25-hydroxyvitamin D-1α hydroxylase (*CYP27B1*) in affected patients, leading to hydroxylation defects and the inability to synthesize $1,25(OH)_2D$.[27] For this reason, and to align with current Online Mendelian Inheritance in Man (OMIM) compendium nomenclature, we will use the OMIM designation of vitamin D hydroxylation–deficient rickets, type 1A (VDDR1A, OMIM 264700) to describe the disease.

On the other hand, VDDR1B (OMIM 600081) is caused by mutations in the *CYP2R1* gene[12-15] encoding the microsomal vitamin D 25-hydroxylase[28] and involves a defect in vitamin D 25-hydroxylation. Mutations in the *VDR* gene cause vitamin D–dependent rickets type 2A (VDDR2A, OMIM 277440), which is associated with true resistance to vitamin D action.[29] This condition was previously referred to as hereditary vitamin D–resistant rickets.

KEY POINTS

- Vitamin D hydroxylation–deficient rickets, type 1A (Online Mendelian Inheritance in Man compendium [OMIM] 264700) now designates mutations in the gene encoding the 25-hydroxyvitamin D-1α hydroxylase (*CYP27B1*) and replaces previous nomenclature for the disease (pseudo vitamin D–deficiency rickets, vitamin D dependency, vitamin D–dependent rickets type I, or 1α-hydroxylase deficiency).
- Vitamin D hydroxylation–deficient rickets, type 1B (OMIM 600081) is caused by mutations in the gene encoding the vitamin D 25-hydroxylase, *CYP2R1*.
- Vitamin D–dependent rickets, type 2A (OMIM 277440) is caused by mutations in the vitamin D receptor (nomenclature replaces hereditary vitamin D–resistant rickets).

VITAMIN D–DEFICIENCY RICKETS, VITAMIN D HYDROXYLATION–DEFICIENT RICKETS, AND VITAMIN D–DEPENDENT RICKETS

It is evident from the overview presented above that vitamin D metabolism can be affected at several levels: inadequate exposure to sunlight, inadequate dietary intake, and malabsorption of dietary vitamin D are classified as vitamin D–deficiency rickets. Reviews on these non-hereditary disorders of the vitamin D endocrine system have been published.[30-33] Genetic disorders affecting vitamin D metabolism and function leading to rickets or osteomalacia include impaired hepatic 25-hydroxylation, defects in renal 1α-hydroxylation, or defects in receptor function (Fig. 58.1).

The prime metabolic consequence of vitamin D deficiency is reduced net intestinal absorption of calcium.[1,2] Calcium malabsorption leads to a fall in plasma calcium, secondary hyperparathyroidism, reduced renal tubular reabsorption of phosphate, hypophosphatemia, and thus a reduction in the calcium X phosphate product. Eventually, deposition of mineral in osteoid is impaired because the supply of the relevant ions is reduced. The impaired mineralization triggers the development of the rachitic and/or osteomalacic phenotype.

Clinical Features of Vitamin D Hydroxylation–Deficient Rickets Type 1B

VDDR1B is a very rare condition that has been reported to affect six families to date.[12-15] It presents with later onset (~4 to 5 years of age) than VDDR1A and is characterized biochemically by hypocalcemia, hypophosphatemia, elevated parathyroid hormone (PTH), high alkaline phosphatase, and low-to-undetectable levels of 25(OH)D. Surprisingly, circulating concentrations of $1,25(OH)_2D$ are in the normal range or elevated. Radiographic findings include features of rickets, with classical cupping and fraying of the metaphyseal region and bowing of the legs. The serum biochemistry can induce misdiagnosis in the absence of genetic analysis: the low-to-undetectable 25(OH)D levels can be interpreted as classical nutritional vitamin D–deficiency

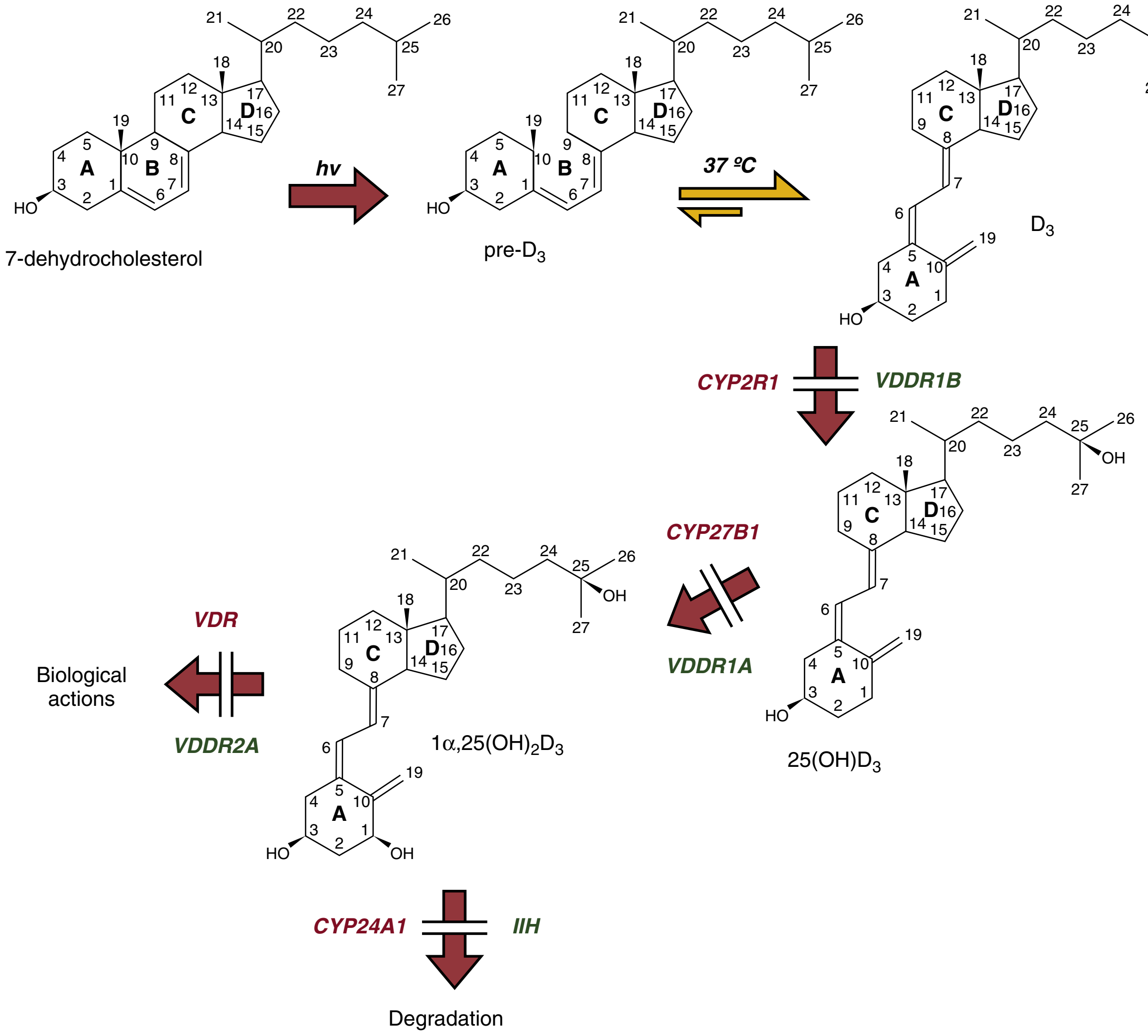


Fig. 58.1 Vitamin D metabolic pathway where genetic aberrations lead to rickets or hypercalcemia. Breaks in the arrows flanking CYP2R1, CYP27B1, VDR, and CYP24A1 represents the defects leading to VDDR1B, VDDR1A, VDDR2A, and IIH, respectively. *VDDR1A*, Vitamin D hydroxylation–deficient rickets type 1A; *VDDR1B,* vitamin D hydroxylation–deficient rickets type 1B; *VDDR2A*, vitamin D–dependent rickets type 2A; *IIH*, idiopathic infantile hypercalcemia; *VDR,* vitamin D receptor; *hv,* photon energy.

rickets, while the normal/high 1,25(OH)$_2$D levels can lead to diagnostic uncertainty.[14]

Clinical Features of Vitamin D Hydroxylation-Deficient Rickets Type 1A

The clinical symptoms of VDDR1A are similar to those of common vitamin D–deficiency rickets, including failure to thrive, hypotonia, and growth retardation. Affected babies lay supine because of severe muscle weakness and bone pain. At this age, gross skeletal deformities are rare; however, if diagnosis and treatment are delayed, severe deformities of the spine and long bones occur, together with generalized muscle weakness simulating myopathy. Motor problems translate into regression in head control and ability to stand. In some patients, the initial event is generalized convulsions, tremulations, and Bravais–Jacksonian fits or tetany. Pathological fractures may happen. The onset in most cases occurs early during the third trimester of life; the patients look healthy at birth.

Physical examination reveals a small, hypotonic child with wide anterior fontanel, frontal bossing, and frequent craniotabes (easy depression of the softened parietooccipital region). Tooth eruption is delayed, and erupted teeth show evidence of enamel hypoplasia. A rachitic rosary is either visible or palpable. In limbs, widening of the metaphyseal areas is evidenced by enlargement of wrists and ankles, and there is a variable degree of deformity (bowing) of long bones diaphyses. Deformities of the thorax may interfere with ventilation and predispose to pulmonary infection; infant death by pulmonary infections was not infrequent in the past when the diagnosis was either missed (confused with a neurological or respiratory condition) or made too late. The Chvostek sign (twitching of the upper lip upon light finger tapping of the facial nerve) reflects nerve irritability, a consequence of a rapid fall in serum calcium.

The x-ray features include diffuse osteopenia (mild to severe hypomineralization of the skeleton) and classic rachitic metaphyseal changes: fraying, cupping, widening, and fuzziness of the zone of provisional calcification immediately under the growth plate (Fig. 58.2). These changes are seen better and detected earlier in the most active growth plates, namely the distal ulna and femur and the proximal and distal tibia. Changes in the diaphyses may not be evident when metaphyseal changes are first detected. However, they will appear a few weeks later as rarefaction, coarse trabeculation, cortical thinning, and subperiosteal erosion. Looser–Milkman's pseudofractures and

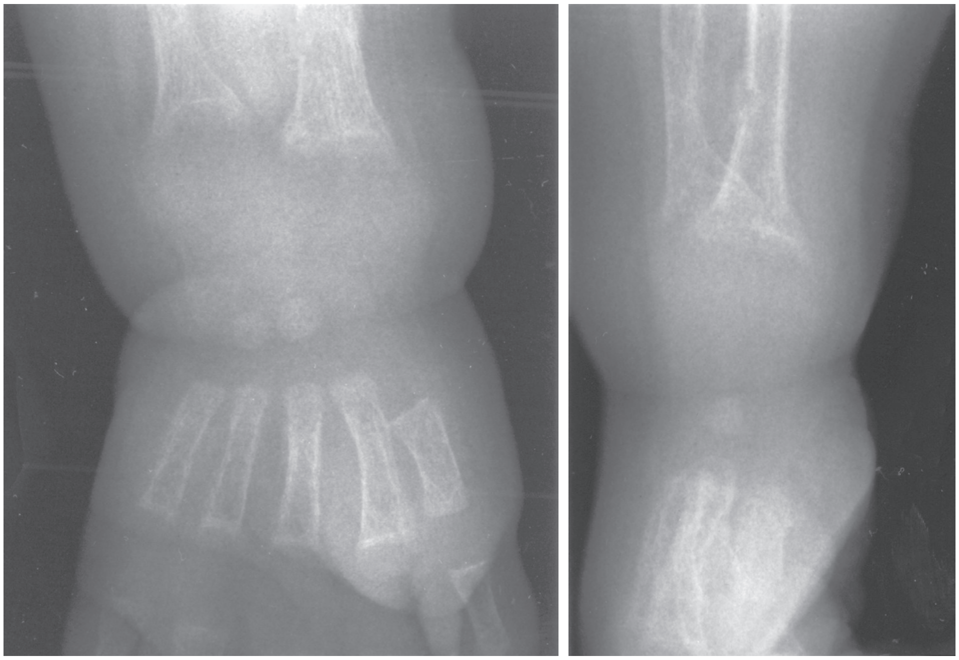


Fig. 58.2 Radiograph of the wrist of a 19-month-old boy with vitamin D hydroxylation–deficient rickets type 1A at diagnosis. Severe rickets and demineralization are evident. Pronounced hypocalcemia and secondary hyperparathyroidism were documented, with the latter causing a metaphyseal pseudofracture clearly seen on the lateral view.

curvature of the shafts of long bones may be observed, especially in children more than 1 to 2 years of age.

Hypocalcemia is the main biochemical feature in VDDR1A. A rapid decrease in serum calcium concentration may give rise to tetany and convulsions, which may occur before any radiological evidence of rickets. Prolonged hypocalcemia triggers secondary hyperparathyroidism and hyperaminoaciduria.[34] Urinary calcium excretion is very low, whereas fecal calcium is high as a consequence of impaired intestinal calcium absorption. Elevated urinary cyclic adenosine monophosphate is not a consistent finding, and normal values have been measured in VDDR1A patients despite high circulating PTH levels.[35]

Serum phosphate concentrations may be normal or low. When reduced, the hypophosphatemia is usually of a lesser degree than that measured in X-linked hypophosphatemic rickets.[36] It results from the combination of impaired intestinal absorption and increased urinary loss induced by the secondary hyperparathyroidism. Serum alkaline phosphatase activity is consistently elevated, and its increase precedes the appearance of clinical symptoms.

Patients with VDDR1A have normal serum levels of 25(OH)D after exposure to sunlight or oral intake of small doses of vitamin D; the concentrations increase if higher doses are given.[35] Circulating levels of 24,25-dihydroxyvitamin D [24,25$(OH)_2$D] are normal and correlate with those of 25(OH)D.[37] Serum levels of 1,25$(OH)_2$D are low to undetectable in untreated patients.[35,38] This is evident immediately after birth, months before any clinical evidence of rickets appears. Even when patients are treated with high doses of vitamin D, causing major increases in circulating levels of 25(OH)D, the blood concentration of 1,25$(OH)_2$D does not reach the normal range. These characteristic features of serum vitamin D metabolites have provided key insight into the pathogenesis of VDDR1A.

Clinical Features of Vitamin D–Dependent Rickets Type 2A

Many of the clinical findings in patients with VDDR 2A are identical to those described for VDDR1A, including bone pain, muscle weakness, hypotonia, and occasional convulsions.[39] Children are often growth-retarded, and tooth enamel hypoplasia is observed.[40] The radiological features of rickets are present. A major difference is that many children with VDDR2A have sparse body hair, and some have total scalp and body alopecia, sometimes even including eyebrows and eyelashes. Hair loss may be evident at birth or occur during the first months of life. Patients with alopecia generally have more severe resistance to vitamin D. In families with a prior history of the disease, the absence of scalp hair in newborns can provide initial diagnostic clues for VDDR2A. A defect in epithelial-mesenchymal communication that is required for normal hair cycling has been shown to be the cause of the alopecia in an animal model of VDDR2A.[41]

Serum biochemistry includes low concentrations of calcium and phosphate and elevated alkaline phosphatase activity. Secondary hyperparathyroidism with elevated circulating PTH is measurable. The key difference concerns circulating levels of vitamin D metabolites. The 25(OH)D values are normal, and, in the cases in which it has been measured, 24,25$(OH)_2$D levels have been low. Importantly, serum levels of 1,25$(OH)_2$D are elevated. This clinical feature clearly distinguishes VDDR2A from VDDR1A, where circulating concentrations of 1,25$(OH)_2$D are depressed. Additionally, patients with VDDR2A are resistant to supraphysiologic doses of all forms of vitamin D therapy. Table 58.1 outlines the similarities and differences between the forms of hereditary rickets involving the vitamin D endocrine system.

MOLECULAR ETIOLOGY OF VITAMIN D HYDROXYLATION–DEFICIENT RICKETS TYPE 1B

Several cytochrome P450 molecules have been shown to hydroxylate vitamin D on carbon 25,[8-11,28] but the identification of the most physiologically relevant enzyme remained uncertain for a considerable time. The initial main candidate was CYP27A1, which has some vitamin D 25-hydroxylase activity but mainly hydroxylates sterol carbons 26 and 27 to initiate bile acid synthesis.[42] Mutations in *CYP27A1* cause

TABLE 58.1 Genomic Defects and Clinical Manifestations of Vitamin D Hydroxylation–Deficient Rickets, Type 1B (VDDR1B), Vitamin D Hydroxylation–Deficient Rickets, Type 1A (VDDR1A), and Vitamin D-Dependent Rickets, Type 2A (VDDR2A)

Feature	VDDR1B	VDDR1A	VDDR2A
Mutations	*CYP2R1*	*CYP27B1*	*VDR*
Genetic inheritance	autosomal recessive	autosomal recessive	autosomal recessive
Age of onset	late (4–5 years of age)	early	early
Rickets	yes	yes	yes
Hypocalcemia	yes	yes	yes
Serum alkaline phosphatase	elevated	elevated	elevated
Secondary hyperparathyroidism	yes	yes	yes
Alopecia	no	no	yes
Serum 25(OH)D	low to undetectable	normal	normal
Serum $1,25(OH)_2D$	normal/high	low	elevated
Response to 25(OH)D therapy	yes	no	no
Response to $1,25(OH)_2D$ therapy	partial	yes	no

cerebrotendinous xanthomatosis, and patients with this disease or mice deficient for the *Cyp27a1* gene have normal calcium and vitamin D homeostasis.[43-45] These data suggested that enzyme(s) other than CYP27A1 must catalyze the 25-hydroxylation of vitamin D *in vivo*.

Cyp2r1 was identified as a gene encoding a cytochrome P450 vitamin D 25-hydroxylase using an elegant ligand activation assay strategy with a cDNA library from liver tissue of *Cyp27a1*-deficient mice.[28] Soon thereafter, mutations in both alleles of *CYP2R1* were identified in patients with low serum calcium concentrations, low circulating levels of 25(OH)D, and rickets.[12,46] These findings confirmed CYP2R1 as a physiologically relevant vitamin D 25-hydroxylase and identified *CYP2R1* mutations as the genetic defect in VDDR1B. The mode of transmission is autosomal recessive, although one family with matching clinical manifestations showed autosomal dominant transmission without *CYP2R1* coding sequence mutations, suggesting that different pathways may affect the function of the 25-hydroxylase enzyme.[47]

CYP2R1 Mutations

Since the identification and characterization of the first CYP2R1 mutation (p.L99P) in two patients of Nigerian origin,[13] four additional mutations in five other families of Nigerian, Saudi, French, and Moroccan descent have been reported in the literature (Table 58.2).[12,14,15] The type of mutations include frameshift (p.G42-L46delinsR and p.L257Sfs*6), missense (p.L99P and p.K242N), and splice site (c.367+1G>A). The splice donor mutation and the p.L257Sfs*6 mutation are predicted to yield truncated proteins that would be inactive, but have not been tested using expression strategies.[12] The p.K242N mutant showed decreased activity in a ligand activation assay with $1\alpha(OH)D$ as substrate.[15] The p.L99P and p.G42-L46delinsR mutations are devoid of enzymatic activity.[13,14]

An *in silico* approach was used to identify *CYP2R1* variants from publicly available databases and predict their pathogenicity. Mutations that were predicted as damaging by at least five separate computational tools yielded 22 potentially pathogenic *CYP2R1* mutations,[48] emphasizing that VDDR1B may be an underdiagnosed condition.[14]

TREATMENT OF VITAMIN D HYDROXYLATION–DEFICIENT RICKETS, TYPE 1B

Different therapeutic modalities have been attempted in patients with VDDR1B. Calcium supplementation alone showed poor efficiency.[15] High doses of vitamin D (cholecalciferol, D_3), with or without supplemental calcium, is able to cure rickets, but the blood biochemistry remains abnormal, mainly for elevated PTH and high alkaline phosphatase.[15] Treatment with $1\alpha(OH)D_3$ (alfacalcidol) and calcium is also beneficial, but the same abnormal biochemical indices are observed.[14] Success has been achieved by treatment of *CYP2R1*-deficient patients with $1,25(OH)_2D$, but this requires frequent monitoring to avoid hypercalciuria and nephrocalcinosis.[47]

TABLE 58.2 CYP2R1 Mutations

Mutation	Change	Exon	Reference
c.124-138delinsCGG	p.G42-L46delinsR		14
c.296T>C	p.L99P	exon 2	13
c.367+1G>A	-	Splice donor/ exon 2	12
c.726A>C	p.K242N	exon 3	15
c.768insT	p.L257Sfs*6	exon 3	12

Based on existing reports, and aligned with the concept of personalized medicine, the treatment of choice is replacement therapy with 25(OH)D to bypass the enzymatic defect. Such treatment leads to total amelioration in clinical symptoms and blood biochemistry data.[14] Replacement therapy administered to one genetically confirmed patient before symptoms appeared prevented bone deformation.[14]

An interesting finding associated with vitamin D or $1\alpha(OH)D_3$ therapy is the increase in circulating $1,25(OH)_2D$. Although it could be attributed to residual activity of the missense mutant CYP2R1 proteins, a preferred interpretation is redundancy of function from the many cytochrome P450 molecules that show 25-hydroxylase activity.[8-11] This is supported by studies in animal models where compound deletion of both *Cyp2r1* and *Cyp27a1* did not fully abrogate circulating levels of 25(OH)D.[49] The alternate enzymes appear not able to correct the metabolic defect in patients but may contribute to vitamin D metabolism in the context of high substrate availability.[14]

KEY POINTS

- The treatment of choice for vitamin D hydroxylation–deficient rickets, type 1B is replacement therapy with 25(OH)D to bypass the CYP2R1 enzymatic defect.

MOLECULAR ETIOLOGY OF VITAMIN D HYDROXYLATION–DEFICIENT RICKETS TYPE 1A

Serum levels of 25(OH)D are normal in untreated patients with VDDR1A and elevated in patients receiving large daily amounts of vitamin D.[35] These results indicate that intestinal absorption of vitamin D and its hydroxylation in the liver are not impaired in VDDR1A. Circulating levels of $24,25(OH)_2D$ are also normal and highly correlated with those of 25(OH)D, indicating a fully functional 25-hydroxyvitamin D-24-hydroxylase enzyme.[37] However, serum concentrations of $1,25(OH)_2D$ are low to undetectable in untreated patients and remain low even when they are treated with high doses of vitamin D.[35,38] This clearly identifies defective activity of the 25-hydroxyvitamin D-1α-hydroxylase enzyme (CYP27B1) as the basic abnormality in VDDR1A and differentiates it from VDDR1B and VDDR2A.

The CYP27B1 enzyme catalyzes the addition of an hydroxyl group at position 1α of the secosteroid backbone of 25(OH)D. The 1α-hydroxylase is a mitochondrial cytochrome P450 enzyme that requires electrons from NADPH to promote catalysis. These are delivered to the P450 moiety by the flavoprotein NADPH-ferredoxin reductase[50] and the nonheme iron protein ferredoxin.[51] The expression of these cofactors is ubiquitous, and their genes were mapped to chromosomes 17 and 11,[50,51] respectively, excluding them rapidly in the search for the VDDR1A mutations because the disease locus was mapped early on to chromosome 12.[52]

VDDR1A is inherited as a simple autosomal recessive trait.[25] No phenotypic abnormalities are observed in heterozygotes.[34] Although generally rare, VDDR1A may be more common in genetically isolated populations as a result of founder effects. For example, VDDR1A occurs at an unusually high frequency in the French Canadian population originating from the Charlevoix-Saguenay-Lac Saint Jean region of the province of Quebec.[53,54] Founder effects have also been reported for eastern Canadian (Acadian) patients[55] and for patients from Korea.[56]

Tremendous progress has been achieved in the study of the molecular etiology of VDDR1A through the cloning of the cDNA encoding for the CYP27B1 enzyme.[27,57-59] The human gene has also been cloned, sequenced, and mapped to 12q13.1-13.3 by fluorescence *in situ* hybridization,[58,60,61] consistent with the earlier mapping of the disease to this locus by linkage analysis.

The ultimate proof that mutations in the *CYP27B1* gene were responsible for the VDDR1A phenotype required the identification of such mutations in VDDR1A patients and carriers of the disease. The first identified mutation was reported by Fu et al.[27]; several additional mutations in various ethnic groups have since been published (Table 58.3). These findings unequivocally establish the molecular genetic basis of VDDR1A as inactivating mutation(s) in the *CYP27B1* gene. Further proof was provided by developing valid animal models of the disease using targeted inactivation of the *Cyp27b1* gene in mice.[62,63]

CYP27B1 Mutations

The human CYP27B1 enzyme is encoded by a single gene of nine exons and eight introns that is 5 kb in length.[64] The ferredoxin-binding domain is encoded by sequences contained in exons 6 and 7, while the heme-binding domain is contained in exon 8.[64,65]

At the time of updating of this chapter, 80 distinct *CYP27B1* mutations have been reported (Table 58.3). Identified disease-causing sequence changes include missense and nonsense mutations, deletions, duplications, and splice mutations.[27,56,60,61,66-87] These mutations are dispersed throughout the *CYP27B1* gene, affecting all exons and five introns (Table 58.3). There are a fair

TABLE 58.3 Mutations in *CYP27B1* Reported up to November 2021

Nucleotide Change[a]	Change	Exon	Reference
c.48_60del13	p.E20Pfs*2	1	68
c.85G>T	p.E29*	1	86
c.165delG	p.K55fs*22	1	80
c.170G>T	p.G57V	1	68
c.171delG	p.L58Cfs*20	1	79
c.171dupG	p.L58Afs*275	1	79a
c.195G>T	p.Q65H	1	80
c.195+2T>A	-	intron 1	87
c.195+2T>G	-	intron 1	70
c.201_204delinsCTTCG	p.Q67Hfs*89	2	81
c.217G>T	p.G73W	2	68
c.217G>C	p.G73R	2	71
c.242G>A	p.G81E	2	83
c.258delG	p.Y87Tfs*24	2	61,80
c.262delG	p.V88Wfs*24	2	71
c.286_300del15	p.E95delELLRQ	2	84
c.305G>A	p.G102E	2	66
c.311_321del11	p.R104Lfs*225	2	67
c.320G>A	p.R107H	2	60
c.335C>T	p.P112L	2	56
c.374G>A	p.G125E	2	60
c.386+1G>A	-	intron 2	81
c.398_400dupAAT	p.W134*	3	81a
c.403C>T	p.Q135*	3	76
c.403delC	p.Q135Kfs*8	3	71
c.413G>T	p.R138L	3	75
c.428C>T	p.P143L	3	74
c.473T>C	p.L158P	3	67
c.490G>A	p.D164N	3	74
c.565G>A	p.E189K	3	80
c.566A>G	p.E189G	3	74
c.574A>G	p.K192E	3	79
c.580G>A	p.G194R	3	86
c.580G>T	p.G194*	3	86
c.590+1G>A	-	intron3	74
c.590G>A	p.G197D	4	79
c.626delG	p.C209Sfs*9	4	27
c.692delC	p.T232Rfs*1	4	27
c.723G>A	p.W241*	4	80
c.934_935delAC	p.T312Rfs*6	5	69
c.962C>G	p.T321R	5	74
c.968C>A	p.S323Y	6	78
c.983G>A	p.W328*	6	56
c.997C>T	p.L333F	6	68
c.1004G>C	p.R335P	6	60
c.1022-1037del16	p.T341Rfs*2	6	69
c.1027C>T	p.L343F	6	81
c.1079C>A	p.S360*	6	70
c.1136+1G>T	-	intron 6	77
c.1136+2T>A	-	intron 6	69
c.1144C>T	p.P382S	7	80
c.1144C>G	p.P382A	7	83
c.1165C>T	p.R389C	7	73
c.1165C>G	p.R389G	7	74
c.1166G>A	p.R389H	7	80
c.1215+1G>A	-	intron 7	56
c.1521delC	p.D507Efs*34	9	85a

TABLE 58.3 Mutations in *CYP27B1* Reported up to August 2020—cont'd.

Nucleotide Change[a]	Change	Exon	Reference
c.1215+2T>A	-	intron 7	87
c.1226C>T	p.T409I	8	80
c.1232G>A	p.C411Y	8	75
c.1238A>G	p.Y413C	8	71
c.1286G>C	p.R429P	8	80
c.1294C>T	p.R432C	8	68
c.1294C>A	p.R432S	8	72
c.1299G>A	p.W433*	8	74
c.1310delG	p.G437Vfs*12	8	68
c.1319-1325dup7	p.F443Pfs*24	8	87
c.1321C>T	p.H441Y	8	82
c.1325-1332ins7	p.H441fs*4	8	80
c.1325-1334ins9	p.H441fs	8	80
c.1357C>T	p.R453C	8	56
c.1358G>A	p.R453H	8	71
c.1375C>T	p.R459C	8	68
c.1376G>T	p.R459L	8	82
c.1433T>G	p.V478G	9	78
c.1442delA	p.E481Gfs*179	9	86
c.1446delA	p.G483Vfs*19	9	68
c.1474C>T	p.R492W	9	68
c.1475G>C	p.R492P	9	71
c.1490C>G	p.P497R	9	80
c.1494delA	p.R499Gfs*14	9	56
c.1504delA	p.N502Tfs*141	9	86
c.1510C>T	p.Q504*	9	85

[a] Nucleotide numbers refer to coding sequence (CDS; National Center for Biotechnology Information consensus CDS database CCDS8954.1) and are numbered from the translation start site. Amino acid replacement mutations are indicated according to codon number. Nomenclature follows standard mutation nomenclature in molecular diagnostics.[172] Reference indicates first report.

number of compound heterozygous mutations (different mutation on each *CYP27B1* allele) in affected patients. A frequent mutated allele is the 7-bp insertion (CCCACCC) in exon 8 (c.1325-1332ins7), present in several unrelated ethnic groups.[61,68,78-80] This mutation alters the downstream reading frame, causing a premature stop codon 16 bp downstream. One patient was also reported with a 9-bp insertion at the same position (c.1325_1334ins9), again causing a frameshift and further suggesting that this area is a mutational hot spot.[80]

A number of assays have been used to study the enzymatic activity of the missense CYP27B1 mutants, including expression of the recombinant protein in heterologous cells that express the electron carriers adrenodoxin reductase and adrenodoxin,[80] activation of vitamin D receptor–dependent gene transcription,[60,74] or enzymatic activity in peripheral blood mononuclear cells.[75,78] In the few cases where missense CYP27B1 mutant proteins retained residual capacity to hydroxylate 25(OH)D on carbon 1 (assayed by expressing the mutant cDNA in heterozygous cells), the VDDR1A patients exhibited mild clinical biochemistry changes.[66,81] However, most of the mutant proteins tested did not retain enzymatic activity, a surprising finding considering that residual amounts of $1,25(OH)_2D$ were measured in the serum of patients harboring the studied mutations.[74] Whether this reflects spurious 1α-hydroxylation of 25(OH)D by other cytochrome P450 proteins when CYP27B1 is defective, or detection of food-derived $1,25(OH)_2D$ is unclear.[74] Because of this absence of complete genotype-phenotype correlation, it has been suggested that vitamin D hydroxylation deficiency may be relatively more common than is currently appreciated.[88]

TREATMENT OF VITAMIN D HYDROXYLATION–DEFICIENT RICKETS, TYPE 1A

The treatment of choice is long-term (lifelong) replacement therapy with $1,25(OH)_2D_3$.[35,89,90] This results in rapid and complete correction of the abnormal phenotype, eliminating hypocalcemia, secondary hyperparathyroidism, and radiographic evidence of rickets. Strikingly, the myopathy disappears within days after initiation of therapy. The restoration of bone mineral content is equally rapid, and histological evidence of healing of the bone structure has been reported.[35] Correction of tooth enamel hypoplasia is only partial. An important aspect of treatment is to ensure adequate calcium intake during the bone healing phase. Needs can be monitored by frequent assessment of urinary calcium excretion. It should be noted that hypercalciuria is common during treatment with $1,25(OH)_2D_3$, particularly during the first year of administration. Its close monitoring is used to adjust the needs in $1,25(OH)_2D_3$. Although the initial dose is recommended to be 1 to 2 µg per day, the maintenance dose will vary between 0.5 and 1 µg per day. High levels of calcium excretion may amplify the pattern of calcium deposition in the kidney; thus, frequent renal imaging and assessment of renal function are essential during the course of treatment.

Before $1,25(OH)_2D_3$ became available from commercial sources, several investigators used the monohydroxylated analog $1\alpha(OH)D_3$,[91] which only requires liver hydroxylation at position 25 to be activated to the hormonally active metabolite. Remember that the 25-hydroxylation step is not affected by the VDDR1A mutation. Response to treatment with $1\alpha(OH)D_3$ is rapid, with healing of rickets in 7 to 9 weeks, and this compound is still used in several countries. On a weight basis, $1\alpha(OH)D_3$ is about half as potent as $1,25(OH)_2D_3$, nullifying any possible economic advantage in favor of the monohydroxylated compound.

KEY POINTS

- Vitamin D hydroxylation–deficient rickets, type 1a (VDDR1A) is inherited as a simple autosomal recessive trait; heterozygotes are not affected.
- Generally rare, VDDR1A may be more common in genetically isolated populations as a result of founder effects.
- The treatment of choice is lifelong replacement therapy with $1,25(OH)_2D_3$.

Molecular Etiology of Vitamin D–Dependent Rickets Type 2A

Patients with VDDR2A have normal 25(OH)D and low, but measurable, $24,25(OH)_2D$ serum values. The circulating levels of $1,25(OH)_2D$ are elevated from three to five times the normal values. These biochemical findings demonstrate that all the vitamin D metabolic enzymes (25-hydroxylase, 24-hydroxylase, and 1α-hydroxylase) are active in patients with VDDR2A. Most patients with the disease are resistant to all forms of vitamin D therapy. This lack of response to vitamin D treatment led Albright et al.[21] to introduce the concept of hormonal resistance. The molecular basis of vitamin D end-organ resistance became clearer as the mechanism of action of the hormonal form of vitamin D was elucidated.

Ligand-binding studies first established that the $1,25(OH)_2D$ hormone, like other sex steroid hormones studied at the time, binds to

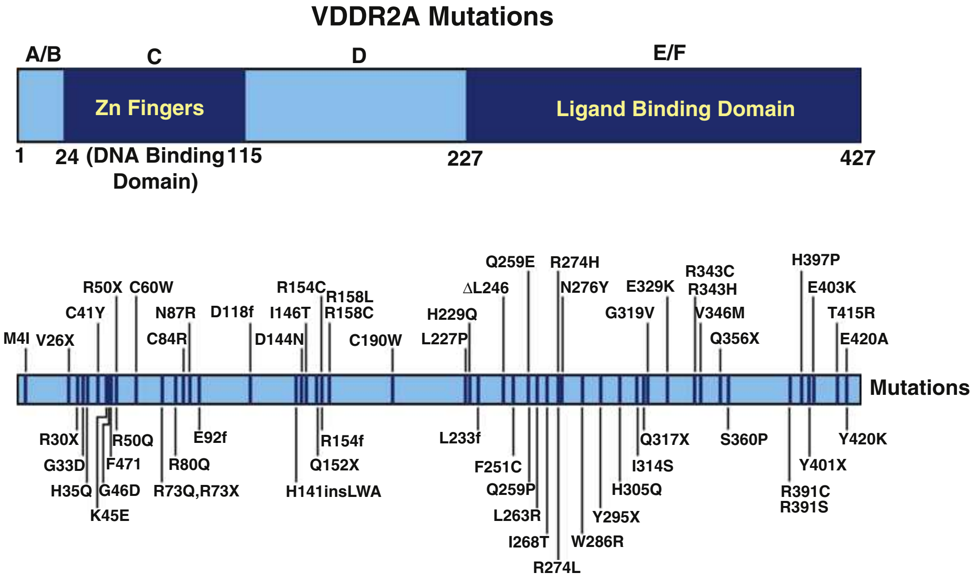


Fig. 58.3 Mutations associated with vitamin D–dependent rickets type 2A (*VDDR2A*). A schematic view of the vitamin D receptor is presented. The 427–amino acid protein can be separated into domains designated A/B, C, D, and E/F. Point mutations in human patients with VDDR2A are indicated below (one-letter amino acid code). *X*, Premature stop codon; *f*, frameshift mutation; *ins*, insertion.

a high-affinity receptor located in the nucleus.[92] It was later discovered that this receptor could bind to DNA.[93] This property was used to purify sufficient quantities of the VDR to raise antibodies.[94,95]

The observation that binding sites for 1,25$(OH)_2$D could be detected in many tissues in addition to the classical vitamin D target tissues helped to define the etiology of VDDR2A. Investigators began to study the VDR from cultured fibroblasts of patients and relatives, using ligand-binding assays, radio-immunoassays, and DNA-cellulose chromatography. Other fibroblast responses measured included induction of the 24-hydroxylase enzyme activity, and vitamin D-mediated growth arrest. These methodologies led to several milestone observations. In the first studies reported, $[^3H]$-1,25$(OH)_2D_3$ binding was undetectable in fibroblasts from VDDR2A patients, and high doses of 1,25$(OH)_2$D failed to induce the 24-hydroxylase bio-marker.[96] The diminished hormone binding provided a clear rationale for the end-organ resistance reported in these patients. Subsequent reports continued to describe a lack of response of the patients' cells to 1,25$(OH)_2$D, but some patients' fibroblasts exhibited normal $[^3H]$-1,25$(OH)_2D_3$ binding.[97] From these early reports, it was concluded that at least two classes of VDDR2A patients could be recognized: "receptor-negative" patients and "receptor-positive" patients. The development of a sensitive radio-immunoassay for the VDR protein[98] demonstrated that these semantic differences were incorrect. Using this assay, it was shown that fibroblasts from "receptor-negative" patients expressed normal levels of receptor protein. Pike et al.[99] hypothesized that the VDR defect in the patients' cells was because of a structural abnormality in the ligand-binding domain preventing $[^3H]$-1,25$(OH)_2D_3$ from binding to the receptor, and not from defective synthesis of the VDR protein. The two classes of VDDR2A were more adequately described by the terminology "ligand binding–positive" and "ligand binding–negative."

A second type of VDR structural abnormality was identified in cultured cells from patients displaying the "ligand binding–positive" VDDR2A phenotype. The VDR from these cells showed reduced affinity for heterologous DNA as measured by DNA-cellulose chromatography.[100] It was suspected that the defect in those patients would likely be a point mutation in the DNA binding domain (DBD) of the VDR.

Eventually, the purified receptor was used to obtain monoclonal antibodies that led to the cloning of the VDR cDNA from various species.[101-103] In turn, the structure of the *VDR* gene was analyzed. Eight exons comprise the entire coding region, spanning approximately 50 kb of genomic DNA.[29] The translation start site is contained within exon 2.

Analysis of the VDR cDNA sequence soon established that the VDR is a member of the nuclear hormone receptor superfamily, and its mechanism of action was subsequently unraveled.[104] The ligand-bound receptor forms a heterodimer with the RXR, and the dimer contacts specific sites within the regulatory domains of responsive genes. This results in positive or negative modulation of the transcription of target genes. The VDR is essential to transduce the biological effects of 1,25$(OH)_2$D, such as promoting calcium and phosphate transport across the small intestine and maintaining calcium homeostasis. The underlying molecular basis of the hormone resistance described in patients with VDDR2A is mutations in the VDR that render the receptor nonfunctional or less functional than the wild-type VDR. Several laboratories have independently engineered valid animal models of VDDR2A by inactivating the *Vdr* gene through gene targeting (reviewed in Bouillon et al.[105]).

VDDR2A follows an autosomal recessive pattern of inheritance. Parents of patients who are heterozygous for the mutation show no symptoms and have normal bone development. In many cases parental consanguinity is associated with the disease. Families often have several affected children, and males and females are affected equally.

Vitamin D Receptor Mutations

The VDR belongs to the nuclear hormone receptor superfamily. The different domains of nuclear receptors have been labeled A to F (Fig. 58.3). Some of these domains exhibit high sequence similarity between individual family members, while others vary considerably or are altogether absent. This diversity is manifested most strongly in domains A/B, D, and F. Domain A/B includes all residues that are N-terminal to the receptor's DBD. Domain C comprises the highly conserved zinc finger DBD, the hallmark feature of nuclear receptor family members. The DBD of the VDR consists of two zinc finger motifs located between residues 24 to 90. The zinc fingers are of the C2C2 type, with two zinc atoms tetrahedrally coordinated through four cysteine residues each that serve to stabilize the finger structure itself. The alpha-helical motifs residing on the C-terminal side of each zinc finger constitute the DNA-recognition and phosphate backbone–binding helices, respectively.[104] The region immediately C-terminal to the second zinc finger, covering residues 91 to 115, also forms an alpha-helical structure that could contribute to DNA contacts.

Region D serves as a flexible hinge domain between the DBD and domain E, and is the least conserved among nuclear receptors. The E region encodes the ligand-binding domain (LBD) of ligand-activated receptors and also contributes to transactivation and dimerization. The small F domain is not highly conserved between nuclear receptor family members.

The E/F region of the VDR (Fig. 58.3) represents a complex multifunctional domain involved in binding of the 1,25$(OH)_2$D ligand, heterodimerization with RXR, and transactivation. The structure of the LBD of the VDR has been modeled following x-ray crystallographic analysis.[106] It consists of 13 alpha-helices and several short β-strands organized as a "sandwich" around a lipophilic hormone-binding pocket. Once the ligand enters the pocket, a "lid" formed by Helix12 closes over the pocket. Several of the mutations characterized in VDDR2A coincide with hormone contact sites.

A number of different VDR mutations have been described in VDDR2A patients (Fig. 58.3 and Table 58.4). Interestingly, a mutation in the first translation initiation codon (p.M1T) is asymptomatic, as an in-frame methionine is encoded four codons downstream.[107] Some of the disease-causing genetic alterations are nonsense (X) or frameshift (fs) mutations that introduce premature stop codons in the receptor. The premature translation termination codons lead to truncated VDRs that lack the DNA-binding domain (R30X and R73X), the LBD (E92fs, Q152X, R154fs, L233fs, Y295X, and Q317X), or the AF-2 activation domain (Y401X); in most cases, the mRNAs for these truncated receptors are unstable.[108]

The first VDR mutation ever reported was described by Hughes et al.[29] and affected the DBD of the receptor. The mutation substitutes a polar uncharged glutamine for a positively charged arginine at position 73 (R73Q) within the second zinc finger of the DBD. The VDR mutation identified by Hughes et al.[29] was the first natural disease-causing mutation reported for the entire steroid-thyroid-retinoid receptor gene superfamily.

Several additional missense mutations affecting the DBD have been reported since (Table 58.4). Based on the crystal structure of the related GR, RXR, and TR molecules,[109-111] the H35Q, K45E, R50Q, R73Q, and R80Q mutations are thought to affect residues that contact DNA.[39,104,112] The conversion of residue glycine-46 to aspartic acid (G46D) introduces a bulky, charged amino acid that would create unfavorable electrostatic interactions with the negatively-charged phosphate backbone of the DNA helix.[113] The G33D mutation would be expected to have the same effect.[112] The F47I mutation is a relatively conserved substitution, but the loss of the phenylalanine ring structure may disrupt the hydrophobic core of the DBD and affect the proposed alpha-helical structure at the base of the first zinc finger.[112]

The first identified mutation affecting the LBD was the Y295X mutation reported by Ritchie et al.[114] This mutation truncates 132 amino acids from the C-terminus of the VDR, which results in the deletion of a major portion of the LBD. This mutation has been described in seven families forming a large kindred in which consanguineous marriages occurred.[108] The mutation was also identified in three additional families unrelated to the extended kindred.[115,116] As previously mentioned, the mutant mRNA is unstable, and the truncated VDR is undetectable using immunology-based assays.[108]

TABLE 58.4 Mutations in *VDR* Reported up to August 2020

Mutation	Domain	Reference
IVS3-2A>G	splicing	173
IVS3+4del2	splicing	174
p.M1T	A/B	107
p.M4I	A/B	175
p.V26M	C (DBD)	176
p.R30X	C (DBD)	177
p.G33D	C (DBD)	29
p.H35Q	C (DBD)	178
p.C41Y	C (DBD)	179
p.K45E	C (DBD)	112
p.G46D	C (DBD)	180
p.F47I	C (DBD)	112
IVS4-2A>T	splicing	181
p.R50Q	C (DBD)	107
p.R50X	C (DBD)	182
p.C60W	C (DBD)	174
p.R73Q	C (DBD)	29
p.R73X	C (DBD)	116
p.R80Q	C (DBD)	183
p.C84R	C (DBD)	184
p.N87R	C (DBD)	185
p.E92fs	C (DBD)	186
p.D118Sfs*7	D (hinge)	173
p.N118Sfs*7	D (hinge)	187
p.H141insLWA	D (hinge)	188
p.D144N	D (hinge)	174
(366delC)fs*54	D (hinge)	189
p.I146T	D (hinge)	190
p.Q152X	D (hinge)	191
IVS6+1G>C	splicing	192
p.R154C	D (hinge)	190
p.R154fs	D (hinge)	192
p.R158C	D (hinge)	174
p.R158L	D (hinge)	173
p.C190W	D (hinge)	39
p.L227P	E/F (LBD)	193
p.H229Q	E/F (LBD)	175
c.702C>G (cryptic splice site)	splicing	113
IVS8+1G>T	splicing	173
p.L233fs	E/F (LBD)	113
p.ΔL246	E/F (LBD)	194
p.F251C	E/F (LBD)	195
p.Q259E	E/F (LBD)	196
p.Q259P	E/F (LBD)	113
p.L263R	E/F (LBD)	197
p.I268T	E/F (LBD)	198
p.R274H	E/F (LBD)	199
p.R274L	E/F (LBD)	191
p.N276Y	E/F (LBD)	174
p.W286R	E/F (LBD)	200
p.Y295X	E/F (LBD)	114
p.H305Q	E/F (LBD)	123
p.I314S	E/F (LBD)	122
p.Q317X	E/F (LBD)	201
p.G319V	E/F (LBD)	196
p.E329K	E/F (LBD)	189

Continued

TABLE 58.4 Mutations in *VDR* Reported up to August 2020—cont'd.

Mutation	Domain	Reference
IVS10+1G>T	splicing	202
p.R343C	E/F (LBD)	203
p.R343H	E/F (LBD)	204
p.V346M	E/F (LBD)	205
p.Q356X	E/F (LBD)	206
p.S360P	E/F (LBD)	207
p.R391C	E/F (LBD)	122
p.R391H	E/F (LBD)	203
p.R391S	E/F (LBD)	197
p.H397P	E/F (LBD)	40
p.Y401X	E/F (LBD)	208
p.E403K	E/F (LBD)	185
p.T415R	E/F (LBD)	209
p.E420A	E/F (LBD)	210
p.E420K	E/F (LBD)	117

DBD, DNA-binding domain; *LBD*, ligand-binding domain.

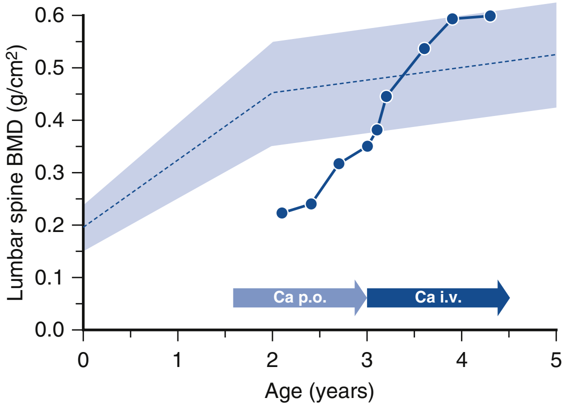


Fig. 58.4 Bone mineral density (*BMD*) response in a patient with vitamin D–dependent rickets type 2A treated with oral and intravenous calcium. BMD of the lumbar spine was measured by dual energy x-ray absorptiometry. The patient received oral calcium (Ca p.o., 2 grams/day) for 17 months before treatment with intravenous calcium (Ca i.v.). The shaded area represents the reference values for the corresponding ages. Note the rapid correction of the BMD upon initiation of intravenous calcium therapy.

The VDR crystallographic data[106] can be used to understand the structural consequences of the missense mutations in the VDR LBD described in VDDR2A patients. These mutations should perturb ligand binding and/or dimerization. Mutation E420K modifies a residue that provides coactivator dimerization interfaces, and thus results in loss of transactivation.[117]

TREATMENT OF VITAMIN D-DEPENDENT RICKETS, TYPE 2A

Therapies with pharmacological doses of vitamin D metabolites, including vitamin D itself, 25(OH)D, 1α(OH)D, and 1,25(OH)$_2$D have been used to attempt to overcome the target organ resistance to vitamin D associated with VDDR2A. Patients with VDDR2A without alopecia are generally more responsive to treatment with high doses of vitamin D preparations than patients with alopecia.[118] Doses reported effective range from 5,000 to 40,000 IU/day for vitamin D, 20 to 200 μg/day for 25(OH)D, and 17 to 20 μg/day for 1,25(OH)$_2$D.[119-121] The efficacy of treatment with high doses of vitamin D metabolites can be reconciled with the molecular cause of VDDR2A. When the resistance to vitamin D is caused by mutations in the VDR that moderately decrease the affinity of the receptor for its ligand, such as the H305Q or I314S mutations,[122,123] high doses of the hormone can apparently overcome the low-affinity binding defect and achieve adequate VDR occupancy to mediate normal 1,25(OH)$_2$D responses. A few patients with VDDR2A with alopecia have also been treated successfully using vitamin D metabolites.[29,97,124-128]

When vitamin D therapy is proven ineffective, intensive calcium therapy is the alternative treatment of choice. Some success has been achieved with high-dose oral calcium (Fig. 58.4).[129] To bypass the calcium absorption defect in the intestine caused by the mutant VDR, long-term intravenous (IV) calcium infusions should be considered (Fig. 58.4). High doses of calcium are infused IV during the night over a several-month period. The original protocol used a central catheter,[130] but success has been reported using intermittent peripheral IV administration.[131] Rapid disappearance of bone pains has been documented, as well as gradual improvement of calcemia and parathyroid function, followed by improvement of rickets (Fig. 58.5), and weight and height gains ensue.[130,132-134] The syndrome can recur when the IV infusions are discontinued.[130] After the serum calcium normalizes and radiological control of the rickets has been achieved with IV calcium infusions, high-dose oral calcium therapy is effective in maintaining normocalcemia.[133] For VDDR2A patients who do not respond to high-dose vitamin D therapy, the two-step calcium protocol (IV infusions followed by high oral doses) appears to be the preferred therapeutic approach.

It is interesting to note that there have been several reports of spontaneous improvement of disease in VDDR2A patients.[97,135,136] This spontaneous healing of rickets usually happens between 7 and 15 years of age, and has not been consistently associated with the onset of puberty. Spontaneous recovery does not appear to be related to treatment, as therapy was often ineffective, and improvement occurred after the treatment was discontinued. The patients appear to remain normocalcemic without therapy and show no evidence of rickets or osteomalacia. Spontaneous improvement has been reported for both ligand binding–positive and ligand binding–negative VDDR2A patients.[97,135,136] The alopecia persists despite healing of the rickets.[97,135,136] It is not uncommon for children to "outgrow" genetic diseases, and the organism appears able to compensate for the loss of VDR function after skeletal growth has been completed. This apparent functional redundancy does not apply to the hair follicle, however, a situation analogous to what has been described in the VDR-ablated animal model of the disease.[137]

KEY POINTS

- Vitamin D–dependent rickets, type 2A is autosomal recessive; individuals who are heterozygous for mutation in the vitamin D receptor show no symptoms and have normal bone development.
- Therapies with pharmacological doses of vitamin D metabolites are moderately effective when the mutation decreases the affinity of the receptor for its ligand.
- Long-term intravenous (IV) calcium infusion is the alternative treatment of choice. High doses of calcium are infused IV during the night over a several-month period.

IDIOPATHIC INFANTILE HYPERCALCEMIA

In a classic endocrine negative feedback loop, the vitamin D hormone, 1,25(OH)$_2$D, induces in target tissues the expression of the main effector of its catabolic breakdown, the CYP24A1 enzyme. This VDR-mediated transcriptional response ensures attenuation of the

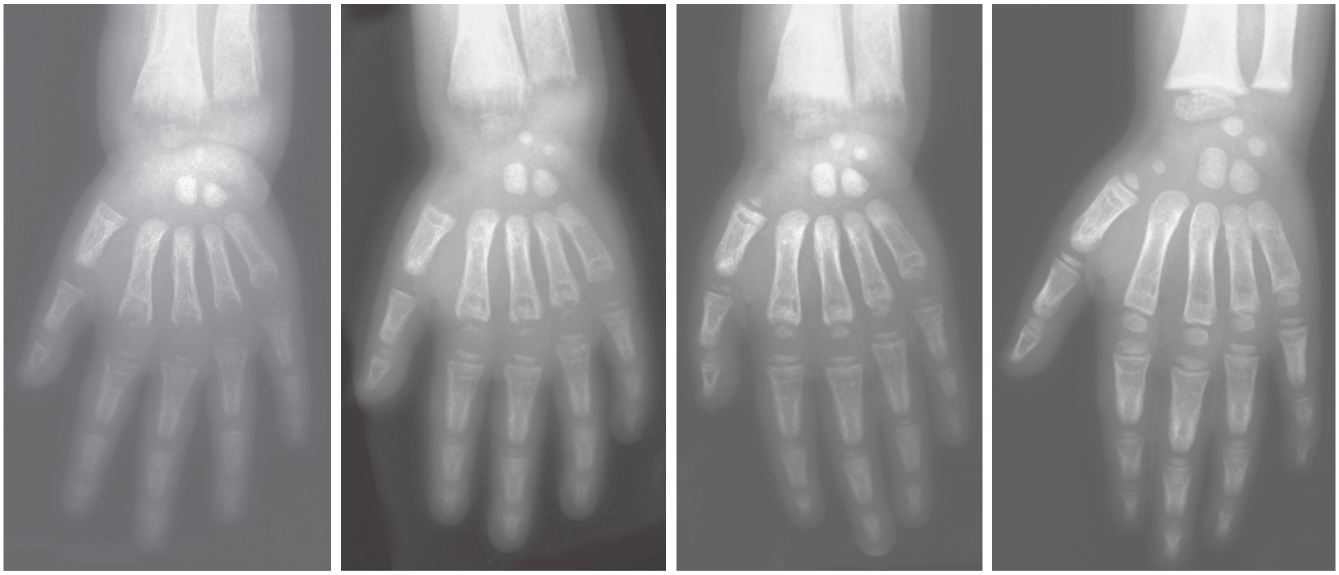


Fig. 58.5 Radiographic analysis of response to calcium therapy in a patient with vitamin D–dependent rickets type 2A. From left to right. First panel: at referral, the patient had been treated for 3 months with $1,25(OH)_2D_3$ (15 µg/day). Rickets is still active, and the bone is demineralized. Second panel: after 6 months on $1,25(OH)_2D_3$ (30 µg/day) and oral calcium (2 g/day), no significant improvement. Third panel: after 2 months of continuous intravenous calcium (without $1,25(OH)_2D_3$), mineralization is improved, but the growth plate remains irregular. Last panel: following 11 months of continuous infusion, mineralization is adequate. Clinical status is satisfactory and maintained with oral calcium (2 g/day).

$1,25(OH)_2D$ biological signal inside target cells and helps regulate vitamin D homeostasis. CYP24A1 is a mitochondrial inner-membrane cytochrome P450 enzyme that catabolizes $1,25(OH)_2D$ through modification of the secosteroid's aliphatic side-chain. The pathways catalyzed by the protein have been mapped out.[138]

Idiopathic infantile hypercalcemia[139-141] (IIH; OMIM 1438880, incidence of ~1 in 47, 000) was an orphan disease until Schlingmann et al.[142] identified loss-of-function mutations in the *CYP24A1* gene as the cause of the disease. IIH typically causes hypercalcemia and hypercalciuria in young infants. Those affected present with failure to gain weight, irritability, and nephrocalcinosis.[143-145] If undiagnosed, infants may die or remain stunted; they are also at risk for renal sequelae, including renal failure. Prompt intervention reduces these sequelae.[143-145] Hypercalciuria and risk of renal stones continue in most subjects throughout adult life.[143-145] Familial forms have been described; infant siblings of probands may be equally hypercalcemic without symptoms and are only detected by screening.[142,146] Moreover, the disorder may first manifest as hypercalciuria and painful kidney stones in adults, without childhood symptoms.[138,147-150]

The initial report of mutations in CYP24A1 causing IIH[142] was followed by various publications reporting the same consequences in adults,[149,151] including one in a 70-year-old male.[152] Thus the name of the disease is a misnomer because the cause is no longer unknown, and it is not confined to childhood.[153]

Molecular Etiology of Idiopathic Infantile Hypercalcemia

Schlingmann et al.[142] demonstrated biallelic mutations causing complete loss of function of CYP24A1 activity in European children with IIH. This leads to the symptoms of hypervitaminosis D: increased intestinal absorption of calcium and symptomatic hypercalcemia. Interestingly, not all cases of IIH are caused by *CYP24A1* mutations. In fact, there are at least three known diseases within the IIH constellation:

1. Loss-of-function mutations of *CYP24A1* at cytogenetic location 20q13.2 (sometimes known as Lightwood syndrome);[139]
2. Williams–Beuren Syndrome (WBS) caused by a large deletion on chromosome 7 (cytogenetic location 7q11.23) of 28 genes including the elastin gene (OMIM 194050); and
3. Loss-of function mutations of the sodium-phosphate cotransporter encoded at cytogenetic location 5q35.3 (also known as *SLC34A1*).[154]

It is likely that more genetic bases for hypercalcemia will be described in the future.

Because of this diagnostic uncertainty, the value of a reliable, rapid screening test for IIH has also been a principal research focus. Kaufmann et al.[155] described such a liquid chromatography tandem mass spectroscopy (LC-MS/MS) method for measuring $24,25(OH)_2D_3$ in 100 µL of serum or less, based upon derivatization with DMEQ-TAD, a dienophile that couples to all vitamin D metabolites in the blood. $24,25(OH)_2D_3$ is the main product of CYP24A1 found in the blood; it is indicative of the enzyme activity of the kidney enzyme and is roughly proportional to the level of the substrate $25(OH)D_3$. The absolute concentration of $24,25(OH)_2D_3$ is a fairly reliable measure of CYP24A1 enzyme activity *in vivo*, and the metabolite is low in most IIH patients with CYP24A1 mutations, even though most untreated IIH patients have elevated serum $25(OH)D_3$ values.[156] However, a more reliable indicator of IIH because of *CYP24A1* mutations is the ratio of the metabolite to its precursor $25(OH)D_3$, namely serum $25(OH)D_3$:$24,25(OH)_2D_3$.[155-158] We have found that this ratio is even more conclusive than serum $24,25(OH)_2D_3$ alone, because it eliminates the possibility that the patient might have a low serum $24,25(OH)_2D_3$ level because of vitamin D deficiency. Our experience of this analytical test using the serum $25(OH)D_3$:$24,25(OH)_2D_3$ ratio in over 100 IIH patients has convinced us that it does not produce false-negative results.[155,156] The other known causes of IIH, namely WBS mutations and sodium-phosphate cotransporter defects,[154] have normal serum $25(OH)D_3$:$24,25(OH)_2D_3$ ratios and are thus distinguished from CYP24A1 defects by this rapid test.

CYP24A1 Mutations

As of November 2021, there were 47 missense and deletion mutations of *CYP24A1* described in the literature (Table 58.5). These mutations must be

TABLE 58.5 Mutations in *CYP24A1* Reported up to November 2021

Nucleotide Change[a]	Protein Change	Exon	Reference
c.62delC	p.P21Rfs*8	1	211
c.107delC	p.P36Lfs*11	1	212
c.247C>G	p.H83D	1	213
c.400T>G	p.W134G	2	214
c.428_430delAAG	p.E143del	2	142
c.443T>C	p.L148P	2	150
c.445_449+1delATCCTg	p.I149Gfs*31	intron 2	142
c.449+1G>T	-	intron 2	142a
c.451G>T	p.E151*	3	142
c.464G>A	p.W155*	3	156
c.469C>T	p.R157W	3	211
c.476G>A	p.R159Q	3	142
c.628T>C	p.W210R	4	215
c.660T>A	p.Y220*	5	161
c.667A>T	p.R223*	5	216
c.732+1G>A	-	intron 5	151
c.758T>A	p.M253K	6	156
c.804G>A	p.W268*	6	149
c.823T>C	p.W275R	6	217
c.931delC	p.L311fs	7	151a
c.845-2A>G	-	intron 6	151
c.943G>T	p.E315*	7	214
c.964G>A	p.E322K	7	142
c.965A>C	p.E322A	7	156
c.980C>T	p.A327V	7	156
c.989C>T	p.T330M	7	156
not determined	p.K351Nfs*21	8	156a
c.999_1006delCAGTCTAA	p.S334Vfs*9	8	218
c.1003dupC	p.L335Pfs*11	8	156
c.1121T>C	p.M374T	8	211
c.1138T>C	p.C380R	8	156
c.1147G>C	p.E383Q	8	185
c.1157+1G>A	-	intron 8	212
c.1166C>T	p.P389L	9	156
c.1186C>T	p.R396W	9	142
c.1187G>A	p.R396Q	9	211
c.1206delA	p.V403F*15	9	156
c.1310C>A	not determined	10	142a
c.1226T>C	p.L409S	9	142
large deletion	-	9–11	219
10.9-kb large deletion	-	9–downstream	220
c.1315C>T	p.R439C	10	156
c.1366G>C	p.G456R	10	156
c.1396C>T	p.R466*	10	221
c.1406_1407delAG	p.E469Afs*22	10	211
c.1426_1427delCT	p.C477Lfs*14	10	142
c.1508C>T	p.P503L	11	161

[a] Nucleotide numbers refer to coding sequence (NG_008334.1) and are numbered from the translation start site. Amino acid replacement mutations are indicated according to codon number. Nomenclature follows standard mutation nomenclature in molecular diagnostics.[172] Reference indicates first report.

distinguished from numerous polymorphisms of the CYP24A1 gene. Loss-of-function mutations can be predicted by various computer programs,[159] but comprehensive confirmation of their effects on enzyme activity only comes from expression of the defective gene. Schlingmann et al.[142] used site-directed mutagenesis of the wild-type human *CYP24A1* gene followed by transient or stable expression of the mutant in V-79 Chinese hamster lung fibroblasts and incubation of the cells with $[1\beta\text{-}^3H]1,25\text{-}(OH)_2D_3$ to demonstrate loss of enzyme activity. Others[150] have used cultured fibroblasts to show loss of CYP24A1 enzyme function. Molin et al.[156] used LC-MS/MS analysis of the $1,24,25\text{-}(OH)_3D_3$ product, instead of a radioactive substrate, to demonstrate loss of CYP24A1 enzyme activity.

CYP24A1 mutations can impact CYP24A1 structure and activity in different ways. Mutations are predicted to alter the substrate (e.g., p.L409S; p.L148P) or heme (e.g., p.R159Q; p.R396W; p.S334Vfs*9) binding site, to alter binding of the ferredoxin cofactor (e.g., p.E383Q), to alter access of the substrate or exit of the product (e.g., p.E143del), or generally to affect protein folding (e.g., p.E322K).[142,150,159-161]

Interestingly, a missense mutation predicted to cause loss of function (p.M245I) was detected in an asymptomatic child.[161] This asymptomatic child never received vitamin D supplementation,[161] which can unmask a latent IIH condition.[142] It is also possible that this mutation could cause late-onset manifestation of the disease or a latent phenotype. Along these lines, there have been several reports of adult patients presenting with mild/moderate degrees of hypercalcemia but recurrent nephrolithiasis or nephrocalcinosis.[147-151] Based upon gene frequency studies, Nesterova and colleagues estimate that the frequency of kidney stones in the general population because of *CYP24A1* mutations might be as high as 4% to 20%.[150]

TREATMENT OF IDIOPATHIC INFANTILE HYPERCALCEMIA

All IIH patients with *CYP24A1* mutations should be placed on restricted vitamin D intake, and female patients should be counseled on their increased risk of hypercalcemia during pregnancy. Successful treatments for acute hypercalcemic episodes in IIH have included use of general cytochrome P450 inhibitors such as ketoconazole and fluconazole that block many endogenous drug and steroidal hydroxylases, as well as CYP27B1.[150,162] These agents have significant side-effect profiles and the posology is cumbersome,[163,164] stimulating the search for improved therapies.

It has been reported that gain-of-function mutations in CYP3A4 result in enhanced bioinactivation of $1,25(OH)_2D$.[165] Based on this observation, pharmacologic intervention to enhance CYP3A4 expression has been tested to treat IIH.[166] Patients were treated with 10 mg/kg/d of rifampin, a potent inducer of CYP3A4.[167] The treatment was well tolerated and normalized or improved all clinical and biochemical parameters. Although additional testing is required, rifampin treatment presents an interesting alternative for inactivation of vitamin D metabolites in IIH.[166]

KEY POINTS

- Idiopathic infantile hypercalcemia (IIH) causes hypercalcemia, hypercalciuria, failure to gain weight, irritability, and nephrocalcinosis in young infants. Adult patients can present with mild/moderate degrees of hypercalcemia and recurrent nephrolithiasis or nephrocalcinosis.
- Female IIH patients are at risk of hypercalcemic episodes during pregnancy because of placental production of $1,25(OH)_2D_3$.
- Molecular etiology is mutation of the vitamin D 24-hydroxylase gene, *CYP24A1*, although other mutations can generate hypercalcemia symptoms. A reliable indicator of IIH because of *CYP24A1* mutations is the serum $25(OH)D_3{:}24,25(OH)_2D_3$ ratio.
- Treatments include restricted vitamin D intake and use of general cytochrome P450 inhibitors such as ketoconazole and fluconazole. Pharmacologic intervention with rifampin to induce CYP3A4 has shown promising results for inactivation of vitamin D metabolites in IIH.

FUTURE DIRECTIONS

Vitamin D homeostasis relies on the equilibrium between the activation of the vitamin D precursor and the catabolic breakdown of the hormonal 1,25(OH)$_2$D form. As we have seen, mutations in the enzymes regulating this pathway lead to metabolic bone disease. A new genetic form of vitamin D–dependent rickets was uncovered in two patients with mutations in the cytochrome P450 enzyme CYP3A4.[165] Although CYP3A4 was identified as a vitamin D 25-hydroxylase,[9] it was not this activation step that was perturbed in the patients. The unrelated subjects had early-onset rickets with reduced serum levels of 25(OH)D and 1,25(OH)$_2$D and did not respond to vitamin D or its activated forms. Characterization of the activity of the mutant enzyme showed that it oxidized 1,25(OH)$_2$D, with 10-fold increased activity over wild-type CYP3A4 and 2-fold greater activity than CYP24A1. The authors concluded that accelerated inactivation of vitamin D metabolites represents a mechanism for vitamin D deficiency and named the disease VDDR3.[165]

VDR-dependent gene transcription involves accessory molecules such as RXR, coactivators, and components of the basic transcriptional machinery.[17] It is also regulated by heterogeneous nuclear ribonucleoproteins (hnRNPs) that bind to vitamin D response elements to reduce gene transcription.[168,169] A patient with typical 1,25(OH)$_2$D resistance was found to have normal VDR sequence and function but suffered from overexpression of hnRNPs, notably hnRNPA1.[170,171] This form of genetic vitamin D–dependent rickets is classified as VDDR2B (OMIM 600785).

Conceptually, vitamin D resistance could also be the result of a mutation affecting other accessory proteins involved in 1,25(OH)$_2$D-dependent transcriptional activation, such as transcriptional coactivators.[16,17] As has happened with the examples discussed above, the identification and characterization of additional clinical manifestations of rickets or hypercalcemia may continue to further enhance our understanding of vitamin D biology.

REFERENCES

1. Canadillas S, Rodriguez-Benot A, Rodriguez M. More on the bone-kidney axis--lessons from hypophosphataemia. *Nephrol Dial Transplant*. 2007;22:1521–1523.
2. DeLuca HF. Overview of general physiologic features and functions of vitamin D. *Am J Clin Nutr*. 2004;80:1689S–1696S.
3. St-Arnaud R. The direct role of vitamin D on bone homeostasis. *Arch Biochem Biophys*. 2008;473:225–230.
4. Webb AR, Kline L, Holick MF. Influence of season and latitude on the cutaneous synthesis of vitamin D3: exposure to winter sunlight in Boston and Edmonton will not promote vitamin D3 synthesis in human skin. *J Clin Endocrinol Metab*. 1988;67:373–378.
5. Mawer EB, Backhouse J, Holman CA, et al. The distribution and storage of vitamin D and its metabolites in human tissues. *Clin Sci*. 1972;43:413–431.
6. Institute of Medicine. *Dietary reference intakes for calcium and vitamin D*. Washington, D.C.: The National Academies Press; 2011.
7. Norman AW. Sunlight, season, skin pigmentation, vitamin D, and 25-hydroxyvitamin D: integral components of the vitamin D endocrine system. *Am J Clin Nutr*. 1998;67:1108–1110.
8. Guo YD, Strugnell S, Back DW, et al. Transfected human liver cytochrome P-450 hydroxylates vitamin D analogs at different side-chain positions. *Proc Natl Acad Sci U S A*. 1993;90:8668–8672.
9. Gupta RP, Hollis BW, Patel SB, et al. CYP3A4 is a human microsomal vitamin D-25-hydroxylase. *J Bone Miner Res*. 2004;19:680–688.
10. Rahmaniyan M, Patrick K, Bell NH. Characterization of recombinant CYP2C11: a vitamin D 25-hydroxylase and 24-hydroxylase. *Am J Physiol Endocrinol Metab*. 2005;288:E753–E760.
11. Yamasaki T, Izumi S, Ide H, et al. Identification of a novel rat microsomal vitamin D3 25-hydroxylase. *J Biol Chem*. 2004;279:22848–22856.
12. Al Mutair AN, Nasrat GH, Russell DW. Mutation of the CYP2R1 vitamin D 25-hydroxylase in a Saudi Arabian family with severe vitamin D deficiency. *J Clin Endocrinol Metab*. 2012;97:E2022–E2025.
13. Cheng JB, Levine MA, Bell NH, et al. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci U S A*. 2004;101:7711–7715.
14. Molin A, Wiedemann A, Demers N, et al. Vitamin D-dependent rickets type 1b (25-hydroxylase deficiency): a rare condition or a misdiagnosed condition? *J Bone Miner Res*. 32:1893–1899.
15. Thacher TD, Fischer PR, Singh RJ, et al. CYP2R1 mutations impair generation of 25-hydroxyvitamin D and cause an atypical form of vitamin D deficiency. *J Clin Endocrinol Metab*. 2015;100:E1005–E1013.
16. Christakos S, Dhawan P, Verstuyf A, et al. Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects. *Physiol Rev*. 2016;96:365–408.
17. Christakos S, Li S, De La Cruz J, et al. New developments in our understanding of vitamin metabolism, action and treatment. *Metabolism*. 2019;98:112–120.
18. Haussler MR, Haussler CA, Bartik L, et al. Vitamin D receptor: molecular signaling and actions of nutritional ligands in disease prevention. *Nutr Rev*. 2008;66:S98–S112.
19. Omdahl JL, Bobrovnikova EA, Choe S, et al. Overview of regulatory cytochrome P450 enzymes of the vitamin D pathway. *Steroids*. 2001;66:381–389.
20. Makin G, Lohnes D, Byford V, et al. Target cell metabolism of 1,25-dihydroxyvitamin D3 to calcitroic acid. Evidence for a pathway in kidney and bone involving 24-oxidation. *Biochem J*. 1989;262:173–180.
21. Albright F, Buttler A, Bloomberg E. Rickets resistant to vitamin D therapy. *Am J Dis Child*. 1937;54:529–547.
22. Fraser D, Salter RB. The diagnosis and management of the various types of rickets. *Pediatr Clin North Am May*. 1958:417–441.
23. Royer P. [Study on idiopathic hypophosphatemic vitamin-resistant rickets]. *Acta Clin Belg*. 1960;15:499–517.
24. Prader A, Illig R, Heierli E. [An unusual form of primary vitamin D-resistant rickets with hypocalcemia and autosomal-dominant hereditary transmission: hereditary pseudo-deficiency rickets]. *Helv Paediatr Acta*. 1961;16:452–468.
25. Scriver CR. Vitamin D dependency. *Pediatrics*. 1970;45:361–363.
26. Portale AA, Miller WL. Hereditary rickets revealed. *Kidney Int*. 1998;54:1762–1764.
27. Fu GK, Lin D, Zhang MY, et al. Cloning of human 25-hydroxyvitamin D-1 alpha-hydroxylase and mutations causing vitamin D-dependent rickets type 1. *Mol Endocrinol*. 1997;11:1961–1970.
28. Cheng JB, Motola DL, Mangelsdorf DJ, et al. De-orphanization of Cytochrome P450 2R1: a microsomal vitamin D 25-hydroxylase. *J Biol Chem*. 2003;278:38084–38093.
29. Hughes MR, Malloy PJ, Kieback DG, et al. Point mutations in the human vitamin D receptor gene associated with hypocalcemic rickets. *Science*. 1988;242:1702–1705.
30. Antonucci R, Locci C, Clemente MG, et al. Vitamin D deficiency in childhood: old lessons and current challenges. *J Pediatr Endocrinol Metab*. 2018;31:247–260.
31. Bouillon R, Antonio L. Nutritional rickets. Historic overview and plan for worldwide eradication. *J Steroid Biochem Mol Biol*. 2020;198:105563.
32. Uday S, Hogler W. Nutritional rickets and osteomalacia in the twenty-first century: revised concepts, public health, and prevention strategies. *Curr Osteoporos Rep*. 2017;15:293–302.
33. Walker A, El Demellawy D, Davila J. Rickets. Historical, epidemiological, pathophysiological, and pathological perspectives. *Acad Forensic Pathol*. 2017;7:240–262.
34. Arnaud C, Maijer R, Reade T, et al. Vitamin D dependency: an inherited postnatal syndrome with secondary hyperparathyroidism. *Pediatrics*. 1970;46:871–880.
35. Delvin EE, Glorieux FH, Marie PJ, et al. Vitamin D dependency: replacement therapy with calcitriol. *J Pediatr*. 1981;99:26–34.

36. Feldman D, Glorieux FH, Pike JW. *Vitamin D*. San Diego: Academic Press; 1997.
37. Mandla S, Jones G, Tenenhouse HS. Normal 24-hydroxylation of vitamin D metabolites in patients with vitamin D-dependency rickets type I. Structural implications for the vitamin D hydroxylases. *J Clin Endocrinol Metab*. 1992;74:814–820.
38. Scriver CR, Reade TM, DeLuca HF, et al. Serum 1,25-dihydroxyvitamin D levels in normal subjects and in patients with hereditary rickets or bone disease. *N Engl J Med*. 1978;299:976–979.
39. Malloy PJ, Pike JW, Feldman D. The vitamin D receptor and the syndrome of hereditary 1,25- dihydroxyvitamin D-resistant rickets. *Endocr Rev*. 1999;20:156–188.
40. Hanna AE, Sanjad S, Andary R, et al. Tooth development associated with mutations in hereditary vitamin D-resistant rickets. *JDR Clin Trans Res*. 2018;3:28–34.
41. Sakai Y, Kishimoto J, Demay MB. Metabolic and cellular analysis of alopecia in vitamin D receptor knockout mice. *J Clin Invest*. 2001;107:961–966.
42. Cali JJ, Russell DW. Characterization of human sterol 27-hydroxylase. A mitochondrial cytochrome P-450 that catalyzes multiple oxidation reactions in bile acid biosynthesis. *J Biol Chem*. 1991;266:7774–7778.
43. Cali JJ, Hsieh CL, Francke U, et al. Mutations in the bile acid biosynthetic enzyme sterol 27-hydroxylase underlie cerebrotendinous xanthomatosis. *J Biol Chem*. 1991;266:7779–7783.
44. Leitersdorf E, Reshef A, Meiner V, et al. Frameshift and splice-junction mutations in the sterol 27-hydroxylase gene cause cerebrotendinous xanthomatosis in Jews of Moroccan origin. *J Clin Invest*. 1993;91:2488–2496.
45. Rosen H, Reshef A, Maeda N, et al. Markedly reduced bile acid synthesis but maintained levels of cholesterol and vitamin D metabolites in mice with disrupted sterol 27-hydroxylase gene. *J Biol Chem*. 1998;273:14805–14812.
46. Casella SJ, Reiner BJ, Chen TC, et al. A possible genetic defect in 25-hydroxylation as a cause of rickets. *J Pediatr*. 1994;124:929–932.
47. Tosson H, Rose SR. Absence of mutation in coding regions of CYP2R1 gene in apparent autosomal dominant vitamin D 25-hydroxylase deficiency rickets. *J Clin Endocrinol Metab*. 2012;97:E796–E801.
48. Sunkar S, Neeharika D. CYP2R1 and CYP27A1 genes: An in silico approach to identify the deleterious mutations, impact on structure and their differential expression in disease conditions. *Genomics*. 2020;112:3677–3686.
49. Zhu JG, Ochalek JT, Kaufmann M, et al. CYP2R1 is a major, but not exclusive, contributor to 25-hydroxyvitamin D production in vivo. *Proc Natl Acad Sci U S A*. 2013;110:15650–15655.
50. Solish SB, Picado-Leonard J, Morel Y, et al. Human adrenodoxin reductase: two mRNAs encoded by a single gene on chromosome 17cen---q25 are expressed in steroidogenic tissues. *Proc Natl Acad Sci U S A*. 1988;85:7104–7108.
51. Chang CY, Wu DA, Mohandas TK, et al. Structure, sequence, chromosomal location, and evolution of the human ferredoxin gene family. *DNA Cell Biol*. 1990;9:205–212.
52. Labuda M, Morgan K, Glorieux FH. Mapping autosomal recessive vitamin D dependency type I to chromosome 12q14 by linkage analysis. *Am J Hum Genet*. 1990;47:28–36.
53. Bouchard G, Laberge C, Scriver CR. [Hereditary tyrosinemia and vitamin-dependent rickets in Saguenay. A genetic and demographic approach]. *Union Med Can*. 1985;114:633–636.
54. De Braekeleer M, Larochelle J. Population genetics of vitamin D-dependent rickets in northeastern Quebec. *Ann Hum Genet*. 1991;55:283–290.
55. Labuda M, Labuda D, Korab-Laskowska M, et al. Linkage disequilibrium analysis in young populations: pseudo-vitamin D- deficiency rickets and the founder effect in French Canadians. *Am J Hum Genet*. 1996;59:633–643.
56. Kim CJ, Kaplan LE, Perwad F, et al. Vitamin D 1alpha-hydroxylase gene mutations in patients with 1alpha-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2007;92:3177–3182.
57. Shinki T, Shimada H, Wakino S, et al. Cloning and expression of rat 25-hydroxyvitamin D3-1alpha-hydroxylase cDNA. *Proc Natl Acad Sci U S A*. 1997;94:12920–12925.
58. St-Arnaud R, Messerlian S, Moir JM, et al. The 25-hydroxyvitamin D 1-alpha-hydroxylase gene maps to the pseudovitamin D-deficiency rickets (PDDR) disease locus. *J Bone Miner Res*. 1997;12:1552–1559.
59. Takeyama K, Kitanaka S, Sato T, et al. 25-Hydroxyvitamin D3 1alpha-hydroxylase and vitamin D synthesis. *Science*. 1997;277:1827–1830.
60. Kitanaka S, Takeyama K, Murayama A, et al. Inactivating mutations in the 25-hydroxyvitamin D3 1alpha-hydroxylase gene in patients with pseudovitamin D-deficiency rickets. *N Engl J Med*. 1998;338:653–661.
61. Yoshida T, Monkawa T, Tenenhouse HS, et al. Two novel 1alpha-hydroxylase mutations in French-Canadians with vitamin D dependency rickets type I. *Kidney Int*. 1998;54:1437–1443.
62. Dardenne O, Prud'homme J, Arabian A, et al. Targeted inactivation of the 25-hydroxyvitamin D(3)-1(alpha)- hydroxylase gene (CYP27B1) creates an animal model of pseudovitamin D- deficiency rickets. *Endocrinology*. 2001;142:3135–3141.
63. Panda DK, Miao D, Tremblay ML, et al. Targeted ablation of the 25-hydroxyvitamin D 1alpha -hydroxylase enzyme: evidence for skeletal, reproductive, and immune dysfunction. *Proc Natl Acad Sci U S A*. 2001;98:7498–7503.
64. Fu GK, Portale AA, Miller WL. Complete structure of the human gene for the vitamin D 1alpha- hydroxylase, P450c1alpha. *DNA Cell Biol*. 1997;16:1499–1507.
65. Monkawa T, Yoshida T, Wakino S, et al. Molecular cloning of cDNA and genomic DNA for human 25-hydroxyvitamin D3 1 alpha-hydroxylase. *Biochem Biophys Res Commun*. 1997;239:527–533.
66. Alzahrani AS, Zou M, Baitei EY, et al. A novel G102E mutation of CYP27B1 in a large family with vitamin D-dependent rickets type 1. *J Clin Endocrinol Metab*. 2010;95:4176–4183.
67. Cao L, Liu F, Wang Y, et al. Novel vitamin D 1alpha-hydroxylase gene mutations in a Chinese vitamin-D-dependent rickets type I patient. *J Genet*. 2011;90:339–342.
68. Cui N, Xia W, Su H, et al. Novel mutations of CYP27B1 gene lead to reduced activity of 1alpha-hydroxylase in Chinese patients. *Bone*. 2012;51:563–569.
69. Demir K, Kattan WE, Zou M, et al. Novel CYP27B1 gene mutations in patients with vitamin d-dependent rickets type 1a. *PLoS One*. 2015;10:e0131376.
70. Durmaz E, Zou M, Al-Rijjal RA, et al. Clinical and genetic analysis of patients with vitamin D-dependent rickets type 1A. *Clin Endocrinol (Oxf)*. 2012;77:363–369.
71. Edouard T, Alos N, Chabot G, et al. Short- and long-term outcome of patients with pseudo-vitamin D deficiency rickets treated with calcitriol. *J Clin Endocrinol Metab*. 2011;96:82–89.
72. Fuchtbauer L, Brusgaard K, Ledaal P, et al. Case report: vitamin D-dependent rickets type 1 caused by a novel CYP27B1 mutation. *Clin Case Rep*. 2015;3:1012–1016.
73. Hu WW, Ke YH, He JW, et al. A novel compound mutation of CYP27B1 in a Chinese family with vitamin D-dependent rickets type 1A. *J Pediatr Endocrinol Metab*. 2014;27:335–341.
74. Kitanaka S, Murayama A, Sakaki T, et al. No enzyme activity of 25-hydroxyvitamin D3 1alpha-hydroxylase gene product in pseudovitamin D deficiency rickets, including that with mild clinical manifestation. *J Clin Endocrinol Metab*. 1999;84:4111–4117.
75. Koek WN, Zillikens MC, van der Eerden BC, et al. Novel compound heterozygous mutations in the CYP27B1 gene lead to pseudovitamin d-deficient rickets. *Calcif Tissue Int*. 2016;99:326–331.
76. Ozcabi B, Tahmiscioglu Bucak F, Jaferova S, et al. The first case of vitamin D dependent rickets type 1A from Uzbek population with a novel mutation. *J Clin Res Pediatr Endocrinol*. 2016;8:484–489.
77. Porcu L, Meloni A, Casula L, et al. A novel splicing defect (IVS6+1G>T) in a patient with pseudovitamin D deficiency rickets. *J Endocrinol Invest*. 2002;25:557–560.
78. Smith SJ, Rucka AK, Berry JL, et al. Novel mutations in the 1alpha-hydroxylase (P450c1) gene in three families with pseudovitamin D-deficiency rickets resulting in loss of functional enzyme activity in blood-derived macrophages. *J Bone Miner Res*. 1999;14:730–739.
79. Tahir S, Demirbilek H, Ozbek MN, et al. Genotype and phenotype characteristics in 22 patients with vitamin D-dependent rickets type I. *Horm Res Paediatr*. 2016;85:309–317.

79a. Kim YM, Jang YY, Jeong JE, et al. A case of vitamin D hydroxylation-deficient rickets type 1A caused by 2 novel pathogenic variants in CYP27B1 gene. *Ann Pediatr Endocrinol Metab*. 2019;24:137–141.

80. Wang JT, Lin CJ, Burridge SM, et al. Genetics of vitamin D 1alpha-hydroxylase deficiency in 17 families. *Am J Hum Genet*. 1998;63:1694–1702.

81. Wang X, Zhang MY, Miller WL, et al. Novel gene mutations in patients with 1alpha-hydroxylase deficiency that confer partial enzyme activity in vitro. *J Clin Endocrinol Metab*. 2002;87:2424–2430.

81a. Zou M, Guven A, BinEssa HA, et al. Molecular analysis of CYP27B1 mutations in vitamin D-dependent rickets Type 1A: c.590G > A (p.G197D) missense mutation causes a RNA splicing error. *Front Genet*. 2020;11:607517.

82. Zalewski A, Ma NS, Legeza B, et al. Vitamin D-dependent rickets type 1 caused by mutations in CYP27B1 affecting protein interactions with adrenodoxin. *J Clin Endocrinol Metab*. 2016;101(9):3409–3418.

83. Giannakopoulos A, Efthymiadou A, Chrysis D. A case of vitamin-D-dependent rickets type 1a with normal 1,25-dihydroxyvitamin D caused by two novel mutations of the CYP27B1 gene. *Horm Res Paediatr*. 2016;87:58–63.

84. Yan Y, Calikoglu AS, Jain N. Vitamin D-dependent rickets type 1: a rare, but treatable, cause of severe hypotonia in infancy. *J Child Neurol*. 2011;26:1571–1575.

85. Babiker AM, Al Gadi I, Al-Jurayyan NA, et al. A novel pathogenic mutation of the CYP27B1 gene in a patient with vitamin D-dependent rickets type 1: a case report. *BMC Res Notes*. 2014;7:783.

85a. Ozden A, Doneray H. The genetics and clinical manifestations of patients with vitamin D dependent rickets type 1A. *J Pediatr Endocrinol Metab*. 2021;34:781–789.

86. Chi Y, Sun J, Pang L, et al. Mutation update and long-term outcome after treatment with active vitamin D3 in Chinese patients with pseudovitamin D-deficiency rickets (PDDR). *Osteoporos Int*. 2019;30:481–489.

87. Dursun F, Ozgurhan G, Kirmizibekmez H, et al. Genetic and clinical characteristics of patients with vitamin D dependent rickets type 1a. *J Clin Res Pediatr Endocrinol*. 2019;11:34–40.

88. Miller WL. Genetic disorders of vitamin D biosynthesis and degradation. *J Steroid Biochem Mol Biol*. 2016;165:101–108.

89. Glorieux FH. Calcitriol treatment in vitamin D-dependent and vitamin D-resistant rickets. *Metabolism*. 1990;39(suppl):10–12.

90. Fraser D, Kooh SW, Kind HP, et al. Pathogenesis of hereditary vitamin-D-dependent rickets. An inborn error of vitamin D metabolism involving defective conversion of 25-hydroxyvitamin D to 1 alpha,25-dihydroxyvitamin D. *N Engl J Med*. 1973;289:817–822.

91. Reade TM, Scriver CR, Glorieux FH, et al. Response to crystalline 1alpha-hydroxyvitamin D3 in vitamin D dependency. *Pediatr Res*. 1975;9:593–599.

92. Haussler MR, Norman AW. Chromosomal receptor for a vitamin D metabolite. *Proc Natl Acad Sci U S A*. 1969;62:155–162.

93. Pike JW, Haussler MR. Purification of chicken intestinal receptor for 1,25-dihydroxyvitamin D. *Proc Natl Acad Sci U S A*. 1979;76:5485–5489.

94. Pike JW, Marion SL, Donaldson CA, et al. Serum and monoclonal antibodies against the chick intestinal receptor for 1,25-dihydroxyvitamin D3. Generation by a preparation enriched in a 64,000-dalton protein. *J Biol Chem*. 1983;258:1289–1296.

95. Dame MC, Pierce EA, Prahl JM, et al. Monoclonal antibodies to the porcine intestinal receptor for 1,25-dihydroxyvitamin D3: interaction with distinct receptor domains. *Biochemistry*. 1986;25:4523–4534.

96. Feldman D, Chen T, Cone C, et al. Vitamin D resistant rickets with alopecia: cultured skin fibroblasts exhibit defective cytoplasmic receptors and unresponsiveness to 1,25(OH)2D3. *J Clin Endocrinol Metab*. 1982;55:1020–1022.

97. Hirst MA, Hochman HI, Feldman D. Vitamin D resistance and alopecia: a kindred with normal 1,25-dihydroxyvitamin D binding, but decreased receptor affinity for deoxyribonucleic acid. *J Clin Endocrinol Metab*. 1985;60:490–495.

98. Dokoh S, Haussler MR, Pike JW. Development of a radioligand immunoassay for 1,25-dihydroxycholecalciferol receptors utilizing monoclonal antibody. *Biochem J*. 1984;221:129–136.

99. Pike JW, Dokoh S, Haussler MR, et al. Vitamin D3--resistant fibroblasts have immunoassayable 1,25-dihydroxyvitamin D3 receptors. *Science*. 1984;224:879–881.

100. Malloy PJ, Hochberg Z, Pike JW, et al. Abnormal binding of vitamin D receptors to deoxyribonucleic acid in a kindred with vitamin D-dependent rickets, type II. *J Clin Endocrinol Metab*. 1989;68:263–269.

101. Baker AR, McDonnell DP, Hughes M, et al. Cloning and expression of full-length cDNA encoding human vitamin D receptor. *Proc Natl Acad Sci U S A*. 1988;85:3294–3298.

102. Burmester JK, Maeda N, DeLuca HF. Isolation and expression of rat 1,25-dihydroxyvitamin D3 receptor cDNA. *Proc Natl Acad Sci U S A*. 1988;85:1005–1009.

103. McDonnell DP, Mangelsdorf DJ, Pike JW, et al. Molecular cloning of complementary DNA encoding the avian receptor for vitamin D. *Science*. 1987;235:1214–1217.

104. Haussler MR, Whitfield GK, Haussler CA, et al. The nuclear vitamin D receptor: biological and molecular regulatory properties revealed. *J Bone Miner Res*. 1998;13:325–349.

105. Bouillon R, Van Cromphaut S, Carmeliet G. Intestinal calcium absorption: Molecular vitamin D mediated mechanisms. *J Cell Biochem*. 2003;88:332–339.

106. Rochel N, Wurtz JM, Mitschler A, et al. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell*. 2000;5:173–179.

107. Saijo T, Ito M, Takeda E, et al. A unique mutation in the vitamin D receptor gene in three Japanese patients with vitamin D-dependent rickets type II: utility of single-strand conformation polymorphism analysis for heterozygous carrier detection. *Am J Hum Genet*. 1991;49:668–673.

108. Malloy PJ, Hochberg Z, Tiosano D, et al. The molecular basis of hereditary 1,25-dihydroxyvitamin D3 resistant rickets in seven related families. *J Clin Invest*. 1990;86:2071–2079.

109. Bourguet W, Ruff M, Chambon P, et al. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature*. 1995;375:377–382.

110. Renaud JP, Rochel N, Ruff M, et al. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature*. 1995;378:681–689.

111. Wagner RL, Apriletti JW, McGrath ME, et al. A structural role for hormone in the thyroid hormone receptor. *Nature*. 1995;378:690–697.

112. Rut AR, Hewison M, Kristjansson K, et al. Two mutations causing vitamin D resistant rickets: modelling on the basis of steroid hormone receptor DNA-binding domain crystal structures. *Clin Endocrinol (Oxf)*. 1994;41:581–590.

113. Cockerill FJ, Hawa NS, Yousaf N, et al. Mutations in the vitamin D receptor gene in three kindreds associated with hereditary vitamin D resistant rickets. *J Clin Endocrinol Metab*. 1997;82:3156–3160.

114. Ritchie HH, Hughes MR, Thompson ET, et al. An ochre mutation in the vitamin D receptor gene causes hereditary 1,25-dihydroxyvitamin D3-resistant rickets in three families. *Proc Natl Acad Sci U S A*. 1989;86:9783–9787.

115. Malloy PJ, Hughes MR, Pike JW, et al. Vitamin D receptor mutations and hereditary 1,25-dihydroxyvitamin D resistant rickets. In: Norman AW, Bouillon R, Thomasset M, eds. *Vitamin D: gene regulation, structure-function analysis, and clinical application. Eighth Workshop on vitamin D*. New York: Walter de Gruyter; 1991:116–124.

116. Wiese RJ, Goto H, Prahl JM, et al. Vitamin D-dependency rickets type II: truncated vitamin D receptor in three kindreds. *Mol Cell Endocrinol*. 1993;90:197–201.

117. Malloy PJ, Xu R, Peng L, et al. A novel mutation in helix 12 of the vitamin D receptor impairs coactivator interaction and causes hereditary 1,25-dihydroxyvitamin D-resistant rickets without alopecia. *Mol Endocrinol*. 2002;16:2538–2546.

118. Marx SJ, Bliziotes MM, Nanes M. Analysis of the relation between alopecia and resistance to 1,25-dihydroxyvitamin D. *Clin Endocrinol (Oxf)*. 1986;25:373–381.

119. Brooks MH, Bell NH, Love L, et al. Vitamin-D-dependent rickets type II. Resistance of target organs to 1,25-dihydroxyvitamin D. *N Engl J Med*. 1978;298:996–999.

120. Marx SJ, Spiegel AM, Brown EM, et al. A familial syndrome of decrease in sensitivity to 1,25-dihydroxyvitamin D. *J Clin Endocrinol Metab*. 1978;47:1303–1310.

121. Zerwekh JE, Glass K, Jowsey J, et al. An unique form of osteomalacia associated with end organ refractoriness to 1,25-dihydroxyvitamin D and apparent defective synthesis of 25-hydroxyvitamin D. *J Clin Endocrinol Metab*. 1979;49:171–175.
122. Whitfield GK, Selznick SH, Haussler CA, et al. Vitamin D receptors from patients with resistance to 1,25-dihydroxyvitamin D3: point mutations confer reduced transactivation in response to ligand and impaired interaction with the retinoid X receptor heterodimeric partner. *Mol Endocrinol*. 1996;10:1617–1631.
123. Malloy PJ, Eccleshall TR, Gross C, et al. Hereditary vitamin D resistant rickets caused by a novel mutation in the vitamin D receptor that results in decreased affinity for hormone and cellular hyporesponsiveness. *J Clin Invest*. 1997;99:297–304.
124. Kudoh T, Kumagai T, Uetsuji N, et al. Vitamin D dependent rickets: decreased sensitivity to 1,25-dihydroxyvitamin D. *Eur J Pediatr*. 1981;137:307–311.
125. Balsan S, Garabedian M, Liberman UA, et al. Rickets and alopecia with resistance to 1,25-dihydroxyvitamin D: two different clinical courses with two different cellular defects. *J Clin Endocrinol Metab*. 1983;57:803–811.
126. Tsuchiya Y, Matsuo N, Cho H, et al. An unusual form of vitamin D-dependent rickets in a child: alopecia and marked end-organ hyposensitivity to biologically active vitamin D. *J Clin Endocrinol Metab*. 1980;51:685–690.
127. Castells S, Greig F, Fusi MA, et al. Severely deficient binding of 1,25-dihydroxyvitamin D to its receptors in a patient responsive to high doses of this hormone. *J Clin Endocrinol Metab*. 1986;63:252–256.
128. Takeda E, Kuroda Y, Saijo T, et al. 1 alpha-hydroxyvitamin D3 treatment of three patients with 1,25-dihydroxyvitamin D-receptor-defect rickets and alopecia. *Pediatrics*. 1987;80:97–101.
129. Sakati N, Woodhouse NJ, Niles N, et al. Hereditary resistance to 1,25-dihydroxyvitamin D: clinical and radiological improvement during high-dose oral calcium therapy. *Horm Res*. 1986;24:280–287.
130. Balsan S, Garabedian M, Larchet M, et al. Long-term nocturnal calcium infusions can cure rickets and promote normal mineralization in hereditary resistance to 1,25-dihydroxyvitamin D. *J Clin Invest*. 1986;77:1661–1667.
131. Ersoy B, Kiremitci S, Isojima T, et al. Successful intermittent intravenous calcium treatment via the peripheral route in a patient with hereditary vitamin D-resistant rickets and alopecia. *Horm Res Paediatr*. 2015;83:67–72.
132. Weisman Y, Bab I, Gazit D, et al. Long-term intracaval calcium infusion therapy in end-organ resistance to 1,25-dihydroxyvitamin D. *Am J Med*. 1987;83:984–990.
133. Hochberg Z, Tiosano D, Even L. Calcium therapy for calcitriol-resistant rickets. *J Pediatr*. 1992;121:803–808.
134. Bliziotes M, Yergey AL, Nanes MS, et al. Absent intestinal response to calciferols in hereditary resistance to 1,25-dihydroxyvitamin D: documentation and effective therapy with high dose intravenous calcium infusions. *J Clin Endocrinol Metab*. 1988;66:294–300.
135. Hochberg Z, Benderli A, Levy J, et al. 1,25-dihydroxyvitamin D resistance, rickets, and alopecia. *Am J Med*. 1984;77:805–811.
136. Chen TL, Hirst MA, Cone CM, et al. 1,25-dihydroxyvitamin D resistance, rickets, and alopecia: analysis of receptors and bioresponse in cultured fibroblasts from patients and parents. *J Clin Endocrinol Metab*. 1984;59:383–388.
137. Li YC, Amling M, Pirro AE, et al. Normalization of mineral ion homeostasis by dietary means prevents hyperparathyroidism, rickets, and osteomalacia, but not alopecia in vitamin D receptor-ablated mice. *Endocrinology*. 1998;139:4391–4396.
138. Jones G, Prosser DE, Kaufmann M. 25-Hydroxyvitamin D-24-hydroxylase (CYP24A1): its important role in the degradation of vitamin D. *Arch Biochem Biophys*. 2012;523:9–18.
139. Lightwood R, Stapleton T. Idiopathic hypercalcaemia in infants. *Lancet*. 1953;265:255–256.
140. Martin ND, Snodgrass GJ, Cohen RD, et al. Vitamin D metabolites in idiopathic infantile hypercalcaemia. *Arch Dis Child*. 1985;60:1140–1143.
141. Stapleton T, Macdonald WB, Lightwood R. The pathogenesis of idiopathic hypercalcemia in infancy. *Am J Clin Nutr*. 1957;5:533–542.
142. Schlingmann KP, Kaufmann M, Weber S, et al. Mutations in CYP24A1 and idiopathic infantile hypercalcemia. *N Engl J Med*. 2011;365:410–421.
142a. Sun Y, Shen J, Hu X, et al. CYP24A1 Variants in two chinese patients with idiopathic infantile hypercalcemia. *Fetal Pediatr Pathol*. 2019;38:44–56.
143. Huang J, Coman D, McTaggart SJ, et al. Long-term follow-up of patients with idiopathic infantile hypercalcaemia. *Pediatr Nephrol*. 2006;21:1676–1680.
144. Nguyen M, Boutignon H, Mallet E, et al. Infantile hypercalcemia and hypercalciuria: new insights into a vitamin D-dependent mechanism and response to ketoconazole treatment. *J Pediatr*. 2010;157:296–302.
145. Pronicka E, Rowinska E, Kulczycka H, et al. Persistent hypercalciuria and elevated 25-hydroxyvitamin D3 in children with infantile hypercalcaemia. *Pediatr Nephrol*. 1997;11:2–6.
146. McTaggart SJ, Craig J, MacMillan J, et al. Familial occurrence of idiopathic infantile hypercalcemia. *Pediatr Nephrol*. 1999;13:668–671.
147. Dauber A, Nguyen TT, Sochett E, et al. Genetic defect in CYP24A1, the vitamin D 24-hydroxylase gene, in a patient with severe infantile hypercalcemia. *J Clin Endocrinol Metab*. 2012;97:E268–E274.
148. Streeten EA, Zarbalian K, Damcott CM. CYP24A1 mutations in idiopathic infantile hypercalcemia. *N Engl J Med*. 2011;365:1741–1742. author reply 1742–1743.
149. Dinour D, Beckerman P, Ganon L, et al. Loss-of-function mutations of CYP24A1, the vitamin D 24-hydroxylase gene, cause long-standing hypercalciuric nephrolithiasis and nephrocalcinosis. *J Urol*. 2013;190:552–557.
150. Nesterova G, Malicdan MC, Yasuda K, et al. 1,25-(OH)2D-24 hydroxylase (CYP24A1) deficiency as a cause of nephrolithiasis. *Clin J Am Soc Nephrol*. 2013;8:649–657.
151. Tebben PJ, Milliner DS, Horst RL, et al. Hypercalcemia, hypercalciuria, and elevated calcitriol concentrations with autosomal dominant transmission due to CYP24A1 mutations: effects of ketoconazole therapy. *J Clin Endocrinol Metab*. 2012;97:E423–E427.
151a. Rousseau-Nepton I, Jones G, Schlingmann K, et al. CYP24A1 and SLC34A1 pathogenic variants are uncommon in a canadian cohort of children with hypercalcemia or hypercalciuria. *Horm Res Paediatr*. 2021;94:124–132.
152. Jacobs TP, Kaufman M, Jones G, et al. A lifetime of hypercalcemia and hypercalciuria, finally explained. *J Clin Endocrinol Metab*. 2014;99:708–712.
153. Tray KA, Laut J, Saidi A. Idiopathic infantile hypercalcemia, presenting in adulthood—no longer idiopathic nor infantile: two case reports and review. *Conn Med*. 2015;79:593–597.
154. Schlingmann KP, Ruminska J, Kaufmann M, et al. Autosomal-recessive mutations in SLC34A1 encoding sodium-phosphate cotransporter 2a cause idiopathic infantile hypercalcemia. *J Am Soc Nephrol*. 2016;27:604–614.
155. Kaufmann M, Gallagher JC, Peacock M, et al. Clinical utility of simultaneous quantitation of 25-hydroxyvitamin D and 24,25-dihydroxyvitamin D by LC-MS/MS involving derivatization with DMEQ-TAD. *J Clin Endocrinol Metab*. 2014;99:2567–2574.
156. Molin A, Baudoin R, Kaufmann M, et al. CYP24A1 mutations in a cohort of hypercalcemic patients: evidence for a recessive trait. *J Clin Endocrinol Metab*. 2015;100:E1343–E1352.
156a. Macdonald C, Upton T, Hunt P, et al. Vitamin D supplementation in pregnancy: a word of caution. Familial hypercalcaemia due to disordered vitamin D metabolism. *Ann Clin Biochem*. 2020;57:186–191.
157. Cools M, Goemaere S, Baetens D, et al. Calcium and bone homeostasis in heterozygous carriers of CYP24A1 mutations: A cross-sectional study. *Bone*. 2015;81:89–96.
158. Ketha H, Kumar R, Singh RJ. LC-MS/MS for Identifying Patients with CYP24A1 Mutations. *Clin Chem*. 2016;62:236–242.
159. De Paolis E, Scaglione GL, De Bonis M, et al. CYP24A1 and SLC34A1 genetic defects associated with idiopathic infantile hypercalcemia: from genotype to phenotype. *Clin Chem Lab Med*. 2019;57:1650–1667.
160. Ji HF, Shen L. CYP24A1 mutations in idiopathic infantile hypercalcemia. *N Engl J Med*. 2011;365:1741. author reply 1742–1743.
161. Gigante M, Santangelo L, Diella S, et al. Mutational spectrum of CYP24A1 gene in a cohort of Italian patients with idiopathic infantile hypercalcemia. *Nephron*. 2016;133:193–204.
162. Sayers J, Hynes AM, Srivastava S, et al. Successful treatment of hypercalcaemia associated with a CYP24A1 mutation with fluconazole. *Clin Kidney J*. 2015;8:453–455.

163. Castinetti F, Guignat L, Giraud P, et al. Ketoconazole in Cushing's disease: is it worth a try? *J Clin Endocrinol Metab*. 2014;99:1623–1630.
164. Wingard JR, Leather H. Hepatotoxicity associated with antifungal therapy after bone marrow transplantation. *Clin Infect Dis*. 2005;41:308–310.
165. Roizen JD, Li D, O'Lear L, et al. CYP3A4 mutation causes vitamin D-dependent rickets type 3. *J Clin Invest*. 2018;128:1913–1918.
166. Hawkes CP, Li D, Hakonarson H, et al. CYP3A4 induction by rifampin: an alternative pathway for vitamin D inactivation in patients with CYP24A1 mutations. *J Clin Endocrinol Metab*. 2017;102:1440–1446.
167. Kolars JC, Schmiedlin-Ren P, Schuetz JD, et al. Identification of rifampin-inducible P450IIIA4 (CYP3A4) in human small bowel enterocytes. *J Clin Invest*. 1992;90:1871–1878.
168. Adams JS, Chen H, Chun R, et al. Response element binding proteins and intracellular vitamin D binding proteins: novel regulators of vitamin D trafficking, action and metabolism. *J Steroid Biochem Mol Biol*. 2004;89:461–465.
169. Chen H, Hu B, Allegretto EA, et al. The vitamin D response element-binding protein. A novel dominant-negative regulator of vitamin D-directed transactivation. *J Biol Chem*. 2000;275:35557–35564.
170. Chen H, Hewison M, Hu B, et al. Heterogeneous nuclear ribonucleoprotein (hnRNP) binding to hormone response elements: a cause of vitamin D resistance. *Proc Natl Acad Sci U S A*. 2003;100:6109–6114.
171. Hewison M, Rut AR, Kristjansson K, et al. Tissue resistance to 1,25-dihydroxyvitamin D without a mutation of the vitamin D receptor gene. *Clin Endocrinol (Oxf)*. 1993;39:663–670.
172. Ogino S, Gulley ML, den Dunnen JT, et al. Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. *J Mol Diagn*. 2007;9:1–6.
173. Demir K, Zou M, Al-Rijjal RA, et al. Novel Vdr mutations in patients with vitamin D-dependent rickets type 2a: a mild disease phenotype caused by a novel canonical splice-site mutation. *Endocr Pract*. 2020;26:72–81.
174. Malloy PJ, Tasic V, Taha D, et al. Vitamin D receptor mutations in patients with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Mol Genet Metab*. 2014;111:33–40.
175. Lim LM, Zhao X, Chao MC, et al. Novel vitamin D receptor mutations in hereditary vitamin D resistant rickets in Chinese. *PLoS One*. 2015;10. e0138152.
176. Malloy PJ, Wang J, Srivastava T, et al. Hereditary 1,25-dihydroxyvitamin D-resistant rickets with alopecia resulting from a novel missense mutation in the DNA-binding domain of the vitamin D receptor. *Mol Genet Metab*. 2010;99:72–79.
177. Mechica JB, Leite MO, Mendonca BB, et al. A novel nonsense mutation in the first zinc finger of the vitamin D receptor causing hereditary 1,25-dihydroxyvitamin D3-resistant rickets. *J Clin Endocrinol Metab*. 1997;82:3892–3894.
178. Yagi H, Ozono K, Miyake H, et al. A new point mutation in the deoxyribonucleic acid-binding domain of the vitamin D receptor in a kindred with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *J Clin Endocrinol Metab*. 1993;76:509–512.
179. Pang Q, Qi X, Jiang Y, et al. Clinical and genetic findings in a Chinese family with VDR-associated hereditary vitamin D-resistant rickets. *Bone Res*. 2016;4:16018.
180. Lin NU, Malloy PJ, Sakati N, et al. A novel mutation in the deoxyribonucleic acid-binding domain of the vitamin D receptor causes hereditary 1,25-dihydroxyvitamin D-resistant rickets. *J Clin Endocrinol Metab*. 1996;81:2564–2569.
181. Casey G, McPherson T, Kini U, et al. Hereditary vitamin D-resistant rickets presenting as alopecia. *Pediatr Dermatol*. 2014;31:519–520.
182. Supornsilchai V, Hiranras Y, Wacharasindhu S, et al. Two siblings with a novel nonsense mutation, p.R50X, in the vitamin D receptor gene. *Endocrine*. 2011;40:62–66.
183. Sone T, Marx SJ, Liberman UA, et al. A unique point mutation in the human vitamin D receptor chromosomal gene confers hereditary resistance to 1,25-dihydroxyvitamin D3. *Mol Endocrinol*. 1990;4:623–631.
184. Asunis I, Marini MG, Porcu L, et al. A novel missense mutation (C84R) in a patient with type II vitamin D-dependent rickets. *Exp Clin Endocrinol Diabetes*. 2010;118:177–179.
185. Braun DA, Lawson JA, Gee HY, et al. Prevalence of monogenic causes in pediatric patients with nephrolithiasis or nephrocalcinosis. *Clin J Am Soc Nephrol*. 2016;11:664–672.
186. Hawa NS, Cockerill FJ, Vadher S, et al. Identification of a novel mutation in hereditary vitamin D resistant rickets causing exon skipping. *Clin Endocrinol (Oxf)*. 1996;45:85–92.
187. Bayramoglu E, Savas Erdeve S, Shi Y, et al. Experience of intravenous calcium treatment and long-term responses to treatment in a patient with hereditary vitamin D-resistant rickets resulting from a novel mutation. *J Pediatr Endocrinol Metab*. 2019;32:647–651.
188. Malloy PJ, Xu R, Cattani A, et al. A unique insertion/substitution in helix H1 of the vitamin D receptor ligand binding domain in a patient with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *J Bone Miner Res*. 2004;19:1018–1024.
189. Miller J, Djabali K, Chen T, et al. Atrichia caused by mutations in the vitamin D receptor gene is a phenocopy of generalized atrichia caused by mutations in the hairless gene. *J Invest Dermatol*. 2001;117:612–617.
190. Song JK, Yoon KS, Shim KS, et al. Novel compound heterozygous mutations in the vitamin D receptor gene in a Korean girl with hereditary vitamin D resistant rickets. *J Korean Med Sci*. 2011;26:1111–1114.
191. Kristjansson K, Rut AR, Hewison M, et al. Two mutations in the hormone binding domain of the vitamin D receptor cause tissue resistance to 1,25 dihydroxyvitamin D3. *J Clin Invest*. 1993;92:12–16.
192. Katavetin P, Katavetin P, Wacharasindhu S, et al. A girl with a novel splice site mutation in VDR supports the role of a ligand-independent VDR function on hair cycling. *Horm Res*. 2006;66:273–276.
193. Huang K, Malloy P, Feldman D, et al. Enteral calcium infusion used successfully as treatment for a patient with hereditary vitamin D resistant rickets (HVDRR) without alopecia: a novel mutation. *Gene*. 2013;512:554–559.
194. Zhou Y, Wang J, Malloy PJ, et al. Compound heterozygous mutations in the vitamin D receptor in a patient with hereditary 1,25-dihydroxyvitamin D-resistant rickets with alopecia. *J Bone Miner Res*. 2009;24:643–651.
195. Malloy PJ, Zhu W, Zhao XY, et al. A novel inborn error in the ligand-binding domain of the vitamin D receptor causes hereditary vitamin D-resistant rickets. *Mol Genet Metab*. 2001;73:138–148.
196. Macedo LC, Soardi FC, Ananias N, et al. Mutations in the vitamin D receptor gene in four patients with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Arq Bras Endocrinol Metabol*. 2008;52:1244–1251.
197. Nguyen M, d'Alesio A, Pascussi JM, et al. Vitamin D-resistant rickets and type 1 diabetes in a child with compound heterozygous mutations of the vitamin D receptor (L263R and R391S): dissociated responses of the CYP-24 and rel-B promoters to 1,25-dihydroxyvitamin D3. *J Bone Miner Res*. 2006;21:886–894.
198. Malloy PJ, Xu R, Peng L, et al. Hereditary 1,25-dihydroxyvitamin D resistant rickets due to a mutation causing multiple defects in vitamin D receptor function. *Endocrinology*. 2004;145:5106–5114.
199. Aljubeh JM, Wang J, Al-Remeithi SS, et al. Report of two unrelated patients with hereditary vitamin D resistant rickets due to the same novel mutation in the vitamin D receptor. *J Pediatr Endocrinol Metab*. 2011;24:793–799.
200. Nguyen TM, Adiceam P, Kottler ML, et al. Tryptophan missense mutation in the ligand-binding domain of the vitamin D receptor causes severe resistance to 1,25-dihydroxyvitamin D. *J Bone Miner Res*. 2002;17:1728–1737.
201. Malloy PJ, Zhu W, Bouillon R, et al. A novel nonsense mutation in the ligand binding domain of the vitamin D receptor causes hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Mol Genet Metab*. 2002;77:314–318.
202. Ma NS, Malloy PJ, Pitukcheewanont P, et al. Hereditary vitamin D resistant rickets: identification of a novel splice site mutation in the vitamin D receptor gene and successful treatment with oral calcium therapy. *Bone*. 2009;45:743–746.
203. Mazen I, Ismail S, Amr K, et al. Hereditary 1,25-dihydroxyvitamin D-resistant rickets with alopecia in four Egyptian families: report of three novel mutations in the vitamin D receptor gene. *J Pediatr Endocrinol Metab*. 2014;27:873–878.

204. Tseng MH, Huang SM, Lo FS, et al. Functional analysis of VDR gene mutation R343H in a child with vitamin D-resistant rickets with alopecia. *Sci Rep*. 2017;7:15337.
205. Arita K, Nanda A, Wessagowit V, et al. A novel mutation in the VDR gene in hereditary vitamin D-resistant rickets. *Br J Dermatol*. 2008;158:168–171.
206. Papadopoulou A, Bountouvi E, Gole E, et al. Identification of a novel nonsense mutation in the ligand-binding domain of the vitamin d receptor gene and clinical description of two Greek patients with hereditary vitamin d-resistant rickets and alopecia. *Horm Res Paediatr*. 2014;82:206–212.
207. Tamura M, Ishizawa M, Isojima T, et al. Functional analyses of a novel missense and other mutations of the vitamin D receptor in association with alopecia. *Sci Rep*. 2017;7:5102.
208. Malloy PJ, Wang J, Peng L, et al. A unique insertion/duplication in the VDR gene that truncates the VDR causing hereditary 1,25-dihydroxyvitamin D-resistant rickets without alopecia. *Arch Biochem Biophys*. 2007;460:285–292.
209. Ben Ameur S, Silve C, Chabchoub I, et al. Clinical and genetic characterization of Tunisian children with hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Horm Res Paediatr*. 2017;87:23–29.
210. Malloy PJ, Zhou Y, Wang J, et al. Hereditary vitamin D-resistant rickets (HVDRR) owing to a heterozygous mutation in the vitamin D receptor. *J Bone Miner Res*. 2011;26:2710–2718.
211. Figueres ML, Linglart A, Bienaime F, et al. Kidney function and influence of sunlight exposure in patients with impaired 24-hydroxylation of vitamin D due to CYP24A1 mutations. *Am J Kidney Dis*. 2015;65:122–126.
212. Pronicka E, Ciara E, Halat P, et al. Biallelic mutations in CYP24A1 or SLC34A1 as a cause of infantile idiopathic hypercalcemia (IIH) with vitamin D hypersensitivity: molecular study of 11 historical IIH cases. *J Appl Genet*. 2017;58:349–353.
213. Madsen JOB, Sauer S, Beck B, et al. CYP24A1 mutation in a girl infant with idiopathic infantile hypercalcemia. *J Clin Res Pediatr Endocrinol*. 2018;10:83–86.
214. Dinour D, Davidovits M, Aviner S, et al. Maternal and infantile hypercalcemia caused by vitamin-D-hydroxylase mutations and vitamin D intake. *Pediatr Nephrol*. 2015;30:145–152.
215. Meusburger E, Mundlein A, Zitt E, et al. Medullary nephrocalcinosis in an adult patient with idiopathic infantile hypercalcaemia and a novel CYP24A1 mutation. *Clin Kidney J*. 2013;6:211–215.
216. Ferraro PM, Minucci A, Primiano A, et al. A novel CYP24A1 genotype associated to a clinical picture of hypercalcemia, nephrolithiasis and low bone mass. *Urolithiasis*. 2017;45:291–294.
217. Tang JCY, Nicholls H, Piec I, et al. Reference intervals for serum 24,25-dihydroxyvitamin D and the ratio with 25-hydroxyvitamin D established using a newly developed LC-MS/MS method. *J Nutr Biochem*. 2017;46:21–29.
218. Woods GN, Saitman A, Gao H, et al. A young woman with recurrent gestational hypercalcemia and acute pancreatitis due to CYP24A1 deficiency. *J Bone Miner Res*. 2016;31:1841–1844.
219. Castanet M, Mallet E, Kottler ML. Lightwood syndrome revisited with a novel mutation in CYP24 and vitamin D supplement recommendations. *J Pediatr*. 2013;163:1208–1210.
220. Molin A, Nowoczyn M, Coudray N, et al. Molecular characterization of a recurrent 10.9kb CYP24A1 deletion in Idiopathic Infantile Hypercalcemia. *Eur J Med Genet*. 2019;62:103577.
221. Murphy JD, Joseph M, Larsen CP. Infantile nephrocalcinosis resulting from a pathogenic CYP24A1 mutation. *Kidney Int Rep*. 2019;4:893–897.

59

Hereditary Disorders of the Skeleton

Michael P. Whyte

OUTLINE

Dense Bone Disorders, 1010
  Osteopetroses, 1010
  Pycnodysostosis, 1012
  Kenny–Caffey Syndrome, 1013
  Progressive Diaphyseal Dysplasia (Camurati–Engelmann Disease), 1013
  Infantile Cortical Hyperostosis (Caffey Disease), 1014
  Endosteal Hyperostoses, 1015
  Osteopoikilosis, 1015
  Osteopathia Striata, 1016
  Pachydermoperiostosis, 1016
Osteogenesis Imperfecta and Heritable Osteoporoses, 1016
Paget Disease of Bone, 1017
Disorders of Increased RANK/OPG/RANKL/NF-κB Signaling, 1018
  Familial Expansile Osteolysis, 1018
  Osteoprotegerin Deficiency, 1018

Endocrinologists not infrequently encounter patients who represent the great diversity of Mendelian disorders of the skeleton. Some of the entities are clinical curiosities, whereas some are lethal.[1,2] Some cause focal bony abnormalities, while some feature generalized disturbances of skeletal growth, modeling (shaping), and remodeling (turnover). Several derange mineral homeostasis. Cumulatively, the number of affected individuals is substantial. All of the disorders are also important because they reveal genes and gene products that influence skeletal biology.[2] As these conditions became understood molecularly, referrals to endocrinologists for diagnosis and treatment have increased. This chapter reviews several of the more common or instructive hereditary disorders of the skeleton, most manifesting in the pediatric population.

KEY POINTS

- Although each is rare, the many heritable diseases of the skeleton affect a significant number of individuals.
- Now, with revelation of the genetic bases of most of these disorders, their pathogeneses are better understood.
- Accordingly, such patients are increasingly referred to endocrinologists for diagnosis and treatment.

DENSE BONE DISORDERS

Increased bone density is a key feature of a number of heritable osteochondrodysplasias,[1-3] but sometimes occurs in various acquired endocrine, metabolic, dietary, hematologic, infectious, and neoplastic diseases (Table 59.1).[3] Osteosclerosis and hyperostosis refer to trabecular versus cortical bone thickening, respectively.[4] A radiographic skeletal survey will show where, and sometimes how, bone density increased, and often other characteristics necessary for diagnosis.[4] Now, the presumptive diagnosis is increasingly verified by mutation analysis.[1,2]

Osteopetroses

Osteopetrosis (OPT, or “marble bone disease”) was reported first by Albers-Schönberg in 1904.[5] Traditionally, two principal forms are encountered: the autosomal recessive infantile (malignant) type that, if untreated, is typically fatal during the first decade of life, and the autosomal dominant adult (benign) type, which has milder manifestations.[3] Especially rare, autosomal recessive childhood (intermediate) OPT has a less certain prognosis. Now, the genes underlying nearly all types of OPT are known, and molecular nosology has advanced the Mendelian classification.[6]

Although considerable clinical and genetic heterogeneity characterizes OPT,[1,6] all true forms reflect failure of osteoclasts to resorb skeletal tissue.[7] Persistence of primary spongiosa, calcified cartilage deposited during endochondral bone formation, is the histopathologic hallmark. The skeleton is fragile for several reasons.[3] Nevertheless, the term “*osteopetrosis*” is still sometimes used generically to describe radiodense skeletons. Now, OPT must be precisely applied based on this pathogenesis, partly because treating for other dense bone disorders could be unsuccessful or harmful.

Clinical Features. The manifestations and complications of OPT are largely predictable. Infantile OPT, which is caused by loss-of-function mutation in one of several genes,[1,6] manifests in the first years of life. There is a large head, frontal bossing, “adenoid” appearance, nystagmus, hepatosplenomegaly, short stature, and genu valgum. Nasal stuffiness caused by underdeveloped sinuses can be an early symptom. Cranial foramina do not widen, often compressing growing auditory, oculomotor, facial, and optic nerves. Blindness can also follow retinal degeneration or raised intracranial pressure. Hydrocephalus or sleep apnea may occur. Tooth eruption is delayed. Anemia, recurrent infections, and spontaneous bruising or bleeding are common and explained by myelophthisis from excessive bone tissue, osteoclasts, and fibrosis crowding what little marrow space has formed. Extramedullary hematopoiesis with hypersplenism and hemolysis may exacerbate

TABLE 59.1 **Dense Bone Diseases**

Dysplasias	Metabolic and Other
Central osteosclerosis with ectodermal dysplasia	Carbonic anhydrase II deficiency
Craniodiaphyseal dysplasia	Fluorosis
Craniometaphyseal dysplasia	Heavy metal poisoning
Dysosteosclerosis	Hepatitis C–associated osteosclerosis
LRP5/6 activation	Hypervitaminosis A, D
Van Buchem disease	Hyper-, hypo-, and pseudohypoparathyroidism
Sclerosteosis	Hypophosphatemic osteomalacia
Frontometaphyseal dysplasia	Milk-alkali syndrome
Juvenile Paget disease	Renal osteodystrophy
Infantile cortical hyperostosis (Caffey disease)	Axial osteomalacia
Lenz–Majewski syndrome	Fibrogenesis imperfecta ossium
Melorheostosis	Ionizing radiation
Metaphyseal dysplasia (Pyle disease)	Lymphoma
Mixed-sclerosing bone dystrophy	Mastocytosis
Oculodentoosseous dysplasia	Multiple myeloma
Osteodysplasia of Melnick and Needles	Myelofibrosis
Osteopathia striata	Osteomyelitis
Osteopetroses	Osteonecrosis
Osteopoikilosis	Paget disease of bone
Progressive diaphyseal dysplasia (Camurati–Engelmann disease)	Sarcoidosis
Pycnodysostosis	Skeletal metastases
	Tuberous sclerosis

(Modified from Whyte MP. Sclerosing bone disorders. In: Bilezikian JP, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism,* 9th ed. Hoboken, NJ: Wiley-Blackwell; 2019:826.)

already severe anemia. There is failure to thrive. Untreated patients succumb, usually during the first decade of life, to sepsis, pneumonia, severe anemia, or hemorrhage.

Intermediate OPT features macrocephaly and short stature, sometimes with cranial nerve palsies, ankylosed teeth leading to osteomyelitis of the jaw, fractures, and mild or occasionally moderately severe anemia. One type, carbonic anhydrase II (CA II) deficiency, the first OPT understood molecularly,[8] also features proximal or distal renal tubular acidosis (RTA) and cerebral calcification. Fractures, failure to thrive, developmental delay, and short stature can manifest in infancy or early childhood. Mental subnormality is common. Optic nerve compression and dental malocclusion are further complications. Metabolic acidosis presents as early as birth. The RTA, typically distal (type 1), may explain any hypotonia, apathy, and muscle weakness. Periodic hypokalemic paralysis can occur.[9]

Adult OPT (Albers–Schönberg disease) can fracture the axial as well as the appendicular skeleton, and sometimes cause blindness or deafness, facial nerve palsy, mandibular osteomyelitis, psychomotor delay, carpal tunnel syndrome, slipped capital femoral epiphysis, and osteoarthritis. Some asymptomatic "carriers" of one of the underlying chloride channel 7 gene (*CLCN7*) mutations[6] show no clinical or radiographic abnormalities.

Additional OPTs include enigmatic transient infantile OPT, which inexplicably resolves during the first months of life.[3] OPT, lymphedema, anhydrotic ectodermal dysplasia, and immunodeficiency (OL-EDA-ID) is an X-linked recessive disorder.[6] Additional severe OPTs feature epilepsy and neurodegeneration.[6] Dysosteosclerosis[1] is one of the "osteoclast-poor" types of OPT.[10]

Radiographic Findings. Typically, the three components of skeletal development are compromised: bone growth, modeling, and remodeling. Generalized symmetrical increase in bone density is the principal radiographic finding.[4] Cortical and trabecular bone are thickened. The calvarium is wide and dense, and the paranasal and mastoid sinuses are poorly pneumatized. Some bones, including vertebrae, may show a "bone-in-bone" (endobone) configuration. Rachitic changes can occur at physes if there is hypocalcemia.[11]

In CA II deficiency, findings can be subtle at birth. Remarkably, the subsequent osteosclerosis and defective skeletal modeling may diminish over years, perhaps from the accompanying RTA and metabolic acidosis. However, cerebral calcification appears on computed tomography between 2 and 5 years of age, affecting gray matter of the cortex and basal ganglia, and increases during childhood.[8]

In Albers–Schönberg disease, abnormalities appear during childhood, including an especially dense skull base, "rugger-jersey" spine, and sometimes enigmatic alternating sclerotic and lucent horizontal bands in the metaphyses of major long bones (Fig. 59.1). Defective modeling causes wide metaphyses with a club shape or "Erlenmeyer flask" appearance.[4] Pathologic "chalk-stick" fractures involve major long bones. Rarely, there is acroosteolysis, which is more common in pycnodysostosis.

Skeletal scintigraphy can reveal fractures and osteomyelitis. Magnetic resonance imaging helps assess hematopoietic stem cell transplantation treatment,[6] because engraftment restores medullary spaces and signals.[12]

Laboratory Findings. In infantile OPT, circulating calcium levels depend largely on gastrointestinal absorption because bone resorption is halted. Hypocalcemia can occur, especially if there is achlorhydria. The term osteopetrorickets denotes associated physeal widening.[11]

In adult OPT, biochemical indices of mineral homeostasis are usually within normal ranges, although circulating parathyroid hormone can be increased.

Numerous and dysfunctional osteoclasts explain why serum acid phosphatase and the "brain" isoenzyme of creatine kinase are often elevated.[13]

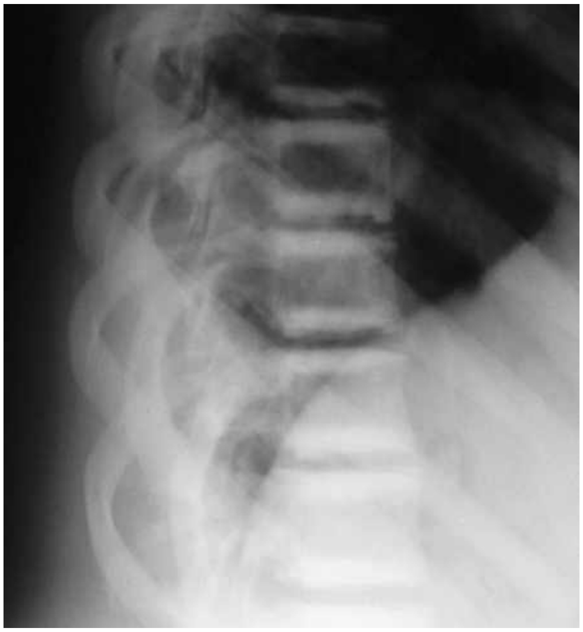


Fig. 59.1 **Osteopetrosis.** This teenage girl with Albers–Schönberg disease has a "rugger-jersey" spine typical of adult ("benign") osteopetrosis. The vertebral endplates are markedly thickened.

Histopathologic Findings. Osteoclast failure provides a pathognomonic finding in genuine OPTs,[7] that is, primary spongiosa synthesized during endochondral bone formation persisting as "islands" or "bars" of mineralized cartilage encased within trabecular bone. Osteoclasts are usually numerous, especially with many nuclei, but "ruffled borders" or "clear zones" may be absent. Fibrosis often crowds marrow spaces. Importantly, in several especially rare OPTs osteoclasts are few or absent.[6]

Etiology and Pathogenesis. The osteoclasts of most patients with OPT cannot produce hydrochloric acid for bone resorption because they harbor biallelic mutations of the *TCIRG1* (*ATP6i*) gene encoding the α3 subunit of the vacuolar proton (H+) pump, or mono- or biallelic defects of *CLCN7,* encoding chloride channel 7.[1,6] Whereas heterozygous mutations of *CLCN7* cause Albers–Schönberg disease, biallelic *CLCN7* mutations lead to severe or intermediate OPT.[14] Biallelic defects in *OSTM1* underlie especially severe OPT.[6] Mutations in *SNX10* have variable clinical severity.[15] OL-EDA-ID reflects disruption of an essential modulator of NF-κB.[1,6] Autosomal recessive transmission of mutations within the genes that encode receptor activator of nuclear factor-κB (RANK),[16] its ligand RANKL,[17] and SLC29A3, underlying dysosteosclerosis, cause the "osteoclast-poor" OPTs.

Ultimately, impaired skeletal resorption, and thus impaired bone modeling and remodeling, in OPT can cause both myelophthisis and bone fragility. The skeleton becomes brittle because calcified cartilage persists, collagen is not interwoven between osteons, hydroxyapatite "hardens," and microfractures may not heal.

Now, antiresorptive pharmaceuticals have become so potent that prolonged administration of excessive doses to a growing child can lead to the skeletal features of OPT.[18]

Treatment. Early prenatal diagnosis of OPT by sonography has generally been unsuccessful. Conventional radiographs occasionally reveal malignant OPT late in pregnancy. Now, mutation analysis is available, including in commercial laboratories, and can be used to detect OPT in utero.[6,19]

Because the etiology and pathogenesis for the failed osteoclast action and prognosis for the various OPTs differ, a precise diagnosis is crucial for planning therapy.[6,19] For example, infants or young children with CA II deficiency can have radiographic features of malignant OPT that then gradually resolve.[8] Now, diagnosis of OPT involves mutation analysis.[3,6,9]

Hyperbaric oxygenation can benefit osteomyelitis of the jaw. Surgical decompression of the optic and facial nerves may help some patients. Joint replacement is possible. Internal fixation may be necessary for femoral fractures.

Some success has been reported using a calcium-deficient diet. However, calcium supplementation would be necessary for accompanying hypocalcemic rickets.[11] Large oral doses of calcitriol to stimulate defective osteoclasts, together with dietary calcium restriction to prevent hypercalciuria/hypercalcemia, have not proven useful. Infusion of parathyroid hormone helped one infant.[3] Diminished cellular production of superoxide was the initial rationale for recombinant human interferon γ-1b (Actimmune) treatment,[20] now used supportively for malignant OPT before bone marrow transplantation.[6]

High-dose glucocorticoid treatment stabilizes pancytopenia and hepatosplenomegaly. One case report describes reversal of malignant OPT after prednisone therapy alone.[21] Prednisone and a low-calcium/high-phosphate diet may be helpful.[22]

CA II–deficiency RTA has been treated with bicarbonate, but the long-term impact is unknown.

Bone marrow transplantation from human leukocyte antigen (HLA)-identical donors has remarkably improved patients with infantile OPT, and cured some.[6,19] However, this approach is not always appropriate because it replaces dysfunctional osteoclasts through donor hematopoietic stem cell engraftment rather than correcting defects in osteoclastogenesis (e.g., RANKL deficiency).[17] Bone marrow transplantation corrects the OPT of CA II deficiency and slows the cerebral calcification, but does not address the RTA.[23] Because severely crowded medullary spaces appear less likely to engraft, rapid molecular diagnosis and early intervention seem best.[24] Marrow from HLA-nonidentical donors is having improved success.[6] Progenitor cells from cord blood have been useful.[25] Marked hypercalcemia can occur as osteoclast function begins.

KEY POINTS

- "Osteopetrosis" now refers selectively to the dense bone disorders that result from generalized failure of osteoclasts to resorb skeletal tissue.
- "Malignant (infantile)" and "benign (adult)" osteopetrosis imprecisely describe such entities, which are increasingly classified by mutation analysis.
- Distinguishing the especially rare "osteoclast-poor" from "osteoclast-rich" osteopetroses is crucial when considering treatment using hematopoietic stem cell transplantation.

Pycnodysostosis

Most reports of pycnodysostosis have been from the United States or Europe, but its prevalence seems greatest in Japan. Parental consanguinity explains approximately 30% of cases. The etiology is biallelic mutation of the *CTSK* gene that encodes cathepsin K.[26]

Clinical Features. In infancy or early childhood there is disproportionate short stature, a relatively large cranium with frontooccipital prominence, small facies and chin, proptosis, bluish sclera, beaked nose, high-arched palate, obtuse mandibular angle, and dental malocclusion with retained deciduous teeth.[27] The anterior fontanel and

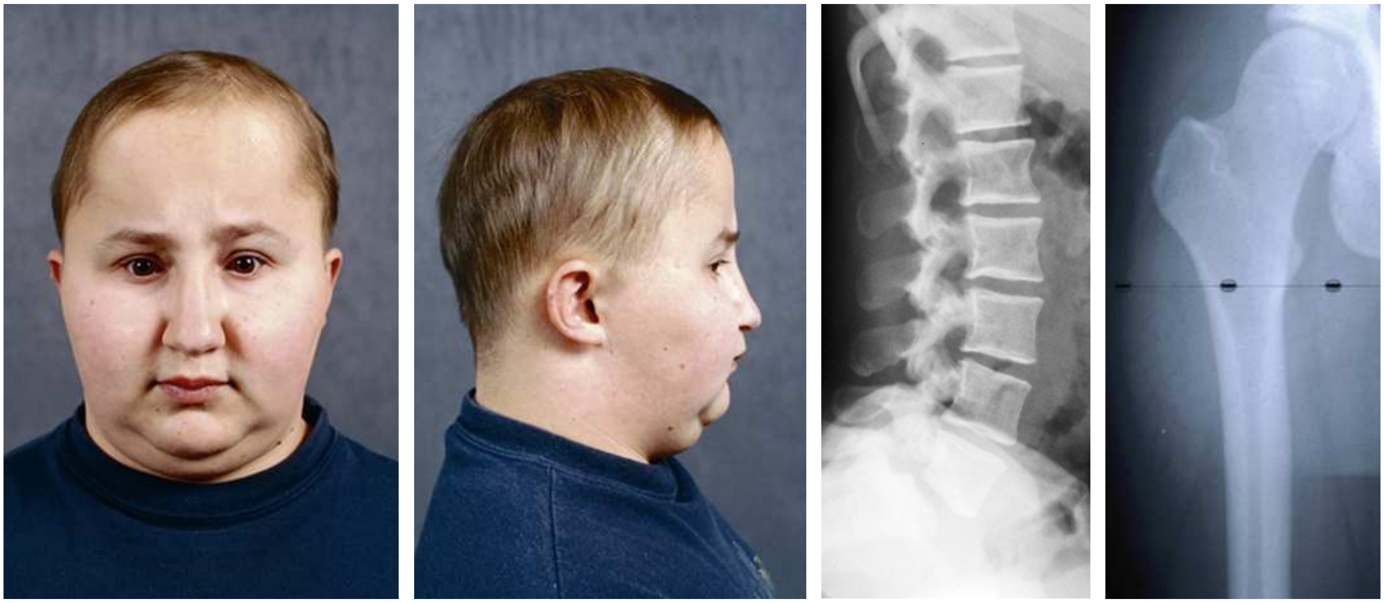


Fig. 59.2 **Kenny–Caffey Syndrome.** This 20-year-old man with Kenny–Caffey syndrome has mild brachycephaly, increased biparietal diameter, bifrontal narrowing, prominent zygomas, hypotelorism, and microretrognathia. His skeleton is diffusely osteosclerotic and hyperostotic.

cranial sutures remain open. Hands are small and square, with short and clubbed fingers from acroosteolysis or aplasia of terminal phalanges. Adult height ranges from 4 feet 3 inches to 4 feet 11 inches. Recurrent fractures typically involve the lower limbs and cause genu valgum deformity, although affected individuals usually walk independently. Pectus excavatum may occur. Respiratory infections and right heart failure can reflect upper airway obstruction from micrognathia. Approximately 10% of patients have compromised mentation.[27]

Radiographic Findings. Pycnodysostosis resembles most OPTs by manifesting recurrent fractures with generalized uniform osteosclerosis in childhood that increases with age. The calvarium and skull base are sclerotic, and the orbital ridges are dense. Although long bones have narrow medullary canals, striking modeling defects, endobones, and radiodense striations are absent. Distinguishing findings include delayed closure of cranial sutures and fontanels (prominently the anterior), wormian bones, obtuse mandibular angle, and hypoplastic facial bones, sinuses, distal phalanges, ribs, and lateral gracile clavicles.[4] Vertebral bodies are dense, with anterior and posterior concavities, but transverse processes are uninvolved. Lumbosacral spondylolisthesis as in the OPTs is not uncommon, and there can be lack of segmentation of the atlas and axis.

Laboratory Findings. Anemia is not expected. Serum calcium, phosphate, and alkaline phosphatase are usually normal. Cortical bone histopathology features diminished activity of osteoblasts and osteoclasts.[28] Abnormal inclusions are found in chondrocytes. Electron microscopy suggests impaired degradation of collagen.[28]

Cause and Pathogenesis. *CTSK* encodes cathepsin K, a lysosomal cysteine protease that is highly expressed in osteoclasts.[29] Thus, pycnodysostosis can be considered an OPT, because impaired collagen degradation by osteoclasts is a fundamental pathogenetic defect. Bone remodeling, and therefore quality, is compromised. Additionally, circulating monocytes have defective killing activity and interleukin-1 secretion. Virus-like inclusions in osteoclasts[30] and defective growth hormone secretion with low circulating insulin-like growth factor 1 levels[31] have been reported.

Treatment. There is no established medical therapy for pycnodysostosis.[32] Marrow cell transplantation has not been reported. The orthopedic challenges have been briefly reviewed.[33] Long bone fractures are typically transverse and heal at a satisfactory rate, but delayed union and massive callus formation can occur. Intramedullary fixation of long bones is formidable because of their hardness.

Extraction of teeth is difficult, and mandibular fracture has occurred.[27] Osteomyelitis of the jaw may require antibiotic and surgical treatment. Severe obstructive sleep apnea can be treated with upper airway surgery together with positive pressure ventilation.

Kenny–Caffey Syndrome

Kenny–Caffey syndrome features marked short stature, delayed closure of fontanels, frontal bossing, triangular face, small eyes (Fig. 59.2), defective dentition, dense bones, and often recurrent hypocalcemia with low serum parathyroid hormone levels.[1] Inheritance is usually autosomal dominant, involving heterozygous mutations of *FAM111A*, although an autosomal recessive form reflects biallelic mutation of *TBCE*.[1,34]

Progressive Diaphyseal Dysplasia (Camurati–Engelmann Disease)

Progressive diaphyseal dysplasia (PDD), an autosomal dominant disorder, is caused by activating mutations in *TGF-β1*.[35]

Clinical Presentation. PDD typically presents during childhood with limping or a broad-based and waddling gait, as well as lower limb pain, muscle wasting, and diminished subcutaneous fat mimicking a muscular dystrophy (Fig. 59.3). However, the clinical and radiographic expressivity is wide-ranging.[36] In severe occurrences, osteosclerosis also involves the skull and axial skeleton, and these patients have a characteristic body habitus featuring an enlarged head with prominent forehead, proptosis, and thin limbs with thickened bones. Cranial nerve palsies can develop. Raised intracranial

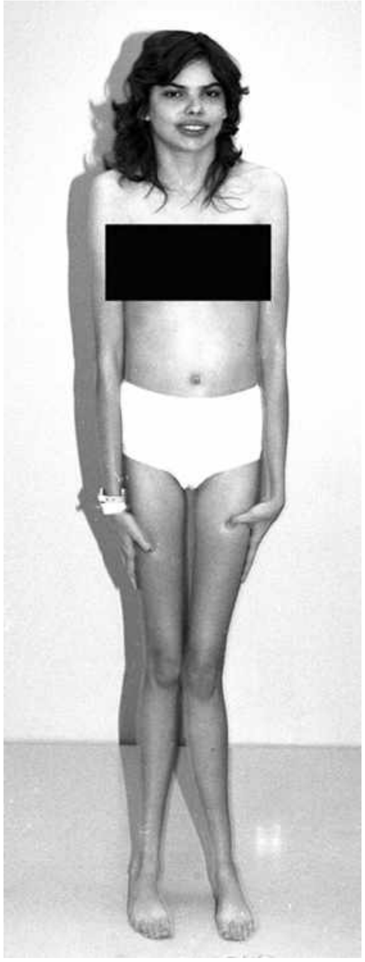


Fig. 59.3 **Progressive diaphyseal dysplasia.** This 16-year-old woman with Camurati–Engelmann disease has characteristic sarcopenia, most evident in her lower limbs, where there is knock-knee deformity.

pressure may occur. Puberty is sometimes delayed, and some patients are unusually tall. Bony thickening with skeletal tenderness is palpable. Hepatosplenomegaly, as well as Raynaud's phenomenon, can suggest vasculitis.[37] The course is variable, and symptom remission sometimes occurs in adult life.

Radiologic Features. Painful hyperostosis gradually involves the periosteal and endosteal surfaces of major long bones. Radiographic studies of PDD typically show progressive disease during growth.[4] Age of onset, rate of progression, and degree of bony involvement are highly variable. Tibias and femurs are most often involved; less frequently the humeri radii, ulnae, or short tubular bones. Diaphyses gradually widen and develop irregular surfaces. Bone sclerosis is fairly symmetrical and spreads to involve metaphyses, yet the epiphyses are characteristically spared (Fig. 59.4). The clavicles, scapulae, and pelvis may also thicken. Maturation of the new bone increases the hyperostosis. However, in severely affected children, some skeletal areas can appear osteopenic. In mild PDD, scintigraphic abnormalities may be confined to the lower limbs. Clinical, radiographic, and scintigraphic findings are generally concordant.[38] Occasionally, however, bone scans are unremarkable despite marked radiographic changes. This seems to reflect advanced but quiescent disease. Sometimes there are no radiographic changes, yet bone scintigraphy is abnormal. This can represent early skeletal involvement.

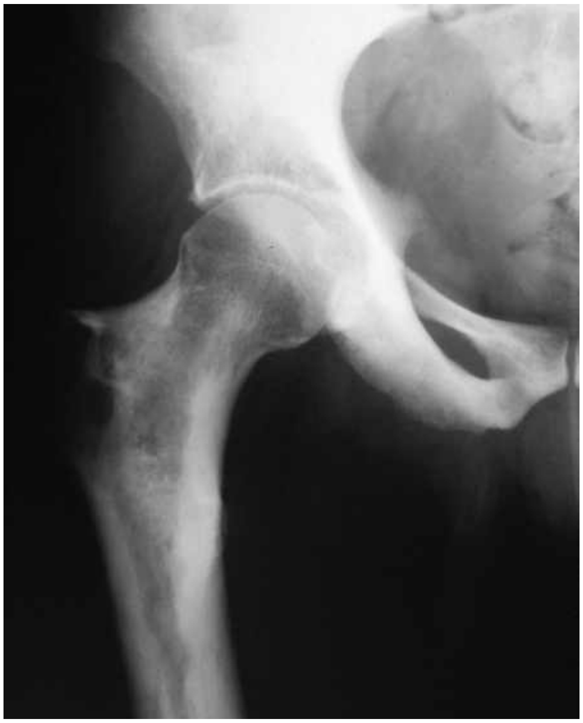


Fig. 59.4 **Progressive diaphyseal dysplasia.** This man with Camurati–Engelmann disease has in his proximal femur irregular hyperostosis that characteristically does not extend to the ends of long bones.

Laboratory Findings. Routine biochemical parameters of bone and mineral metabolism are typically normal in PDD, although serum alkaline phosphatase activity can be elevated. Mild hypocalcemia and significant hypocalciuria may occur in severe disease, likely reflecting positive calcium balance.[39] When present, mild anemia and leukopenia and an elevated erythrocyte sedimentation rate seem to reflect a poorly characterized systemic disturbance.[37]

Skeletal histopathology shows that new bone formation along the diaphysis is disorganized woven bone that matures with incorporation into the cortex. Electron microscopy has revealed myopathic changes and vascular abnormalities.

Etiology and Pathogenesis. The etiology of PDD is heterozygous mutation within a specific region of TGF-β1 that compromises dimerization and stability of the latent TGF-β1 complex, keeping this enhancer of bone formation active.[40] The clinical and laboratory features of severe PDD and its responsiveness to glucocorticoid treatment suggest an inflammatory connective tissue disease.[37] Excess active TGF-β1 in the marrow, released by osteoclasts, may induce migration,[41] yet aberrant differentiation,[42] of osteogenic stem cells.

Treatment. PDD is a chronic and somewhat unpredictable disorder. Symptoms may remit during adolescence or adult life. Management of cranial base hyperostosis has been reviewed.[43] Glucocorticoid therapy, typically small doses of prednisone on alternate days, can relieve bone pain and improve bone histology.[44] Bisphosphonate therapy can increase skeletal pain,[45] which may instead be controlled with salmon calcitonin nasal spray.[46]

Infantile Cortical Hyperostosis (Caffey Disease)

Infantile cortical hyperostosis features focal self-limiting, massive, subperiosteal new bone formation, typically of the long bones, mandible, or clavicles, together with fever, soft tissue swelling, and irritability.[47] A heterozygous missense mutation has been found in *COL1A1* that encodes type 1 collagen.[48]

Endosteal Hyperostoses

These are rare, highly informative, Mendelian disorders that feature excesses of good-quality cortical bone.[49] Van Buchem and associates in 1976 described *hyperostosis corticalis generalisata*.[50] Subsequently, additional hyperostoses involving primarily endosteal surfaces were characterized.[49]

Van Buchem disease is an autosomal recessive disorder[50] that is rarer than case reports might suggest. The principal clinical feature is marked progressive asymmetrical thickening of the jaw during puberty. The mandible develops a wide angle, yet there is no prognathism, and dental malocclusion is uncommon. There may be no symptoms, but cranial sclerosis develops as early as infancy, often causing recurrent facial nerve palsy, deafness, and optic atrophy from narrowing of cranial foramina. Long bones may hurt when palpated, yet are strong.[50] Endosteal cortical thickening leads to homogeneously dense diaphyses with narrowed medullary canals, yet long bones appear properly shaped. Osteosclerosis affects the skull base, facial bones, vertebrae, pelvis, and ribs.[4] There may be hyperphosphatasemia from bone, but serum calcium and phosphate levels are normal. Glucocorticoid therapy may be useful.[51]

Sclerosteosis, like Van Buchem disease, is an autosomal recessive disorder most prevalent in people of Dutch ancestry. Affected individuals additionally have congenital bony or soft tissue syndactyly and are excessively tall. At birth, only bony fusion of the middle and index fingers may be noted. During early childhood, skeletal overgrowth involves especially the skull, causing facial disfigurement. Progressive bone thickening widens the jaw resulting in prognathism.[52] Intelligence is normal. Deafness and facial palsy are significant complications. A small skull vault may increase intracranial pressure, cause headaches, and compress the brainstem,[53] shortening life expectancy.[54] Radiographs show long bones that widen as cortices thicken. Vertebral pedicles, ribs, and the pelvis may become dense. Middle ear ossicles can fuse, with narrowing of the internal auditory canal and cochlear aqueduct. No abnormality of mineral homeostasis or of pituitary function has been documented.

Deactivating mutations of the gene *SOST*, which encodes the inhibitor of bone formation sclerostin, cause sclerosteosis, type 1.[1,55] Van Buchem disease results from a 52-kb deletion of a downstream enhancer of *SOST*.[56] Thus, these disorders compromise the quality or amount, respectively, of sclerostin. Mutations in the portion of *SOST* that encodes the secretion signal cause autosomal dominant craniodiaphyseal dysplasia.[57] Sclerosteosis type 2 involves mutation of *LRP4*.[58]

Healthy sclerostin disengages low-density lipoprotein receptor–related protein (LRP) 5 and 6 from canonical Wnt signaling.[59] Deficient inhibitory action of sclerostin in these disorders enhances bone formation by osteoblasts, with failure of osteoclasts to compensate.[53]

There is no specific medical treatment for these four endosteal hyperostoses. Surgical correction of bony syndactyly is difficult. Management of the neurologic dysfunction, including craniectomy, has been reviewed.[53]

Worth type endosteal hyperostosis,[60] now called "high bone mass,"[61,62] is an autosomal dominant disorder that, like sclerosteosis types 1 and 2, Van Buchem disease, and craniodiaphyseal dysplasia, increases WNT/β-catenin signaling, leading to excessive high-quality bone formation. The etiology is gain-of-function mutation of *LRP5* or *LRP6* encoding the cognate LRP5 and LRP6 receptors, respectively.[62,63] LRP5 high bone mass is no longer incorrectly considered a type of autosomal dominant OPT.[64] In one view, LRP5 activation decreases systemic serotonin thereby enhancing bone formation.[65] However, it is generally accepted that compromised sclerostin binding to LRP 4, 5, or 6 activates WNT/β-catenin signaling and osteoblast activty.[66] LRP5 and 6 high bone mass can cause cranial nerve palsies, skeletal pain, and oropharyngeal exostoses.[62,67]

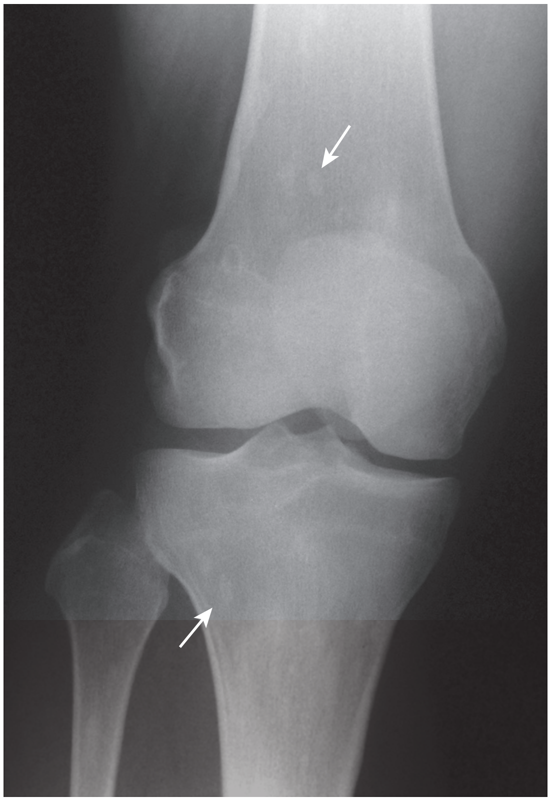


Fig. 59.5 **Osteopoikilosis.** Faint round or oval areas of osteosclerosis *(arrows)* are present in the metaphysis of the distal femur and proximal tibia.

Osteopoikilosis

Osteopoikilosis (spotted bones) is usually a radiographic curiosity transmitted as a highly penetrant autosomal dominant trait.[68] However, joint contractions and limb length inequality can occur, especially if there are radiographic changes of melorheostosis ("dripping candle wax bone").[69] When connective tissue nevi called *dermatofibrosis lenticularis disseminata* are present, this is Buschke–Ollendorff syndrome.[70] The spotted bone is benign and asymptomatic, but if misunderstood can lead to investigation for metastatic disease. Hence, family members at risk should be counseled. Radiographs show numerous small foci of usually round or oval osteosclerosis[4] involving the ends of short tubular bones, the metaepiphyseal regions of the long bones, and the tarsal, carpal, and pelvic bones (Fig. 59.5). Radionuclide accumulation is not increased on bone scanning.[71] These thickened trabeculae or islands of cortical bone do not change over decades.

The nevi usually appear before puberty on the lower trunk or extremities and are small, asymptomatic papules or yellow or white discs or plaques, deep nodules, or streaks. They represent excessive, unusually broad, and markedly branched elastin fibers in the dermis.[70] Heterozygous deactivating mutations in the *LEMD3* gene cause osteopoikilosis and the Buschke–Ollendorff syndrome,[72] but there may be genetic heterogeneity.[73] LEMD3 is an inner nuclear membrane protein that antagonizes TGF-β1 and bone morphogenetic protein signaling.

Osteopathia Striata

Osteopathia striata features gracile linear striations of trabecular bone, especially in the metaepiphyses of major long bones and in the periphery of the iliac bones, that are unchanged for years.[4] Radionuclide accumulation is not increased during bone scanning.[71] It is a curiosity when an isolated autosomal dominant finding, but also occurs in several clinically important syndromes, including osteopathia striata with cranial sclerosis,[74] where cranial nerve palsies are common because of *WTX* gene mutations.[75] Osteopathia striata with focal dermal hypoplasia (Goltz syndrome) is an X-linked dominant disorder complicated by bony defects in the limbs and widespread linear areas of dermal hypoplasia through which adipose tissue herniates because of mutation of the *PORCN* gene.[76]

Pachydermoperiostosis

Pachydermoperiostosis (hypertrophic osteoarthropathy: primary or idiopathic) manifests a classic clinical triad of hyperhidrosis, clubbing of the digits, and thickening of the skin, especially of the face and forehead (*cutis verticis gyrata*).[77] However, some affected individuals have one or two of these findings. Autosomal dominant inheritance with variable expression is established,[77] but autosomal recessive transmission also occurs.[78] Black people seem more severely affected than White people, and men more so than women. Presentation is typically during adolescence, but variable.[77] Symptoms emerge over a decade, but can then abate. Progressive enlargement of the hands and feet may cause a "pawlike" appearance. Palms may be dripping wet. Arthralgias of the elbows, wrists, knees, and ankles are common. Stiffness of the appendicular and axial skeleton can develop, and cranial or spinal nerves can become compressed. Fatigue is common. Cutaneous changes include coarsening, thickening, furrowing, pitting, and oiliness of especially the scalp and face, with some affected individuals described as "acromegalic." Myelofibrosis, myelophthisic anemia with extramedullary hematopoiesis, and renal amyloidosis may occur.

Radiologic Features. Radiographs reveal periosteal new bone formation, especially in the distal limbs, that roughens the surface of cortical bone.[4,79] The radius, ulna, tibia, and fibula are most often involved. Clubbing is obvious, and acroosteolysis can occur. Ankylosis of joints, especially in the hands and feet, may trouble older patients. The principal diagnostic challenge is secondary hypertrophic osteoarthropathy (pulmonary or otherwise), which instead features a smooth, undulating periostitis.[4] In pachydermoperiostosis, periosteal proliferation is more exuberant, irregular, and often involves epiphyses.[4] Bone scintigraphy in either condition reveals symmetrical, diffuse, regular uptake along the cortical margins of long bones, especially in the legs, causing a "double stripe" sign.

Laboratory Findings. Histopathologic examination shows that the new periosteal tissue undergoes cancellous compaction, rendering it difficult to distinguish from the original cortex.[79] Trabecular osteopenia reflects quiescent bone formation. Mild cellular hyperplasia and thickening of subsynovial blood vessels occur near synovial membranes. Increased circulating levels of prostaglandin E2 may be detected.[80]

Cause and Pathogenesis. Pachydermoperiostosis is caused by mutations of the *HPGD* gene that encodes 15-hydroxyprostaglandin dehydrogenase, which is involved in prostaglandin degradation,[81] or *SLC02A1* that encodes solute carrier organic anion transporter 2A1, which is involved in prostaglandin transport.[1]

Treatment. There is no specific medical treatment for pachydermoperiostosis. Painful synovial effusions may respond to nonsteroidal antiinflammatory drugs,[81] and myelofibrosis to prednisolone.[82] Contractures or neurovascular compression may require surgical intervention.

KEY POINTS

- Like the osteopetroses, most other heritable dense bone disorders can now be recognized by mutation analysis.
- The endosteal hyperostoses (sclerosteosis, van Buchem disease, LRP5 and 6 activation) are informative osteoblast-mediated (anabolic) bone disorders, including for pharmacologic treatment of osteoporosis.

OSTEOGENESIS IMPERFECTA AND HERITABLE OSTEOPOROSES

Osteogenesis imperfecta (OI) is increasingly a generic term for heritable "brittle bone disease."[83-85] The nosology for OI devised by Sillence and colleagues in 1979,[86] according to the principal features and the apparent inheritance pattern, was useful for prognostication and recurrence risk counseling. Then, collagen protein and later DNA-based findings focused on defective type 1 collagen. Germline mosaicism with autosomal dominant inheritance explained the especially severe type 2 form. Now, the clinical heterogeneity of classic OI is explained by many mutations throughout the two large genes that encode the pro-α_1 and pro-α_2 chains that combine to form the type 1 collagen heterotrimer.[1] Quantitative, and often qualitative, abnormalities compromise this most abundant protein in bone.[2] More recently, the genetic bases for especially rare forms of "OI," some no longer disrupting type 1 collagen, have been revealed.[83-85] Nevertheless, OI remains a heritable osteoporosis that causes fractures and skeletal deformity. Understandably, the severity of OI is extremely broad-ranging and spans neonatal death to asymptomatic mutation carriers.

Clinical Presentation. The differential diagnosis for OI in infants and children includes idiopathic juvenile osteoporosis, congenital indifference to pain, child abuse, homocystinuria, and a variety of other disorders. Classic OI has many typical features for diagnosis found in the medical history, physical examination, and radiographic findings. A positive family history is especially helpful, as many patients represent sporadic disease from a new mutation. The severity of this OI is often variable within families. Type 1 collagen is abundant not only in bone, but also in teeth, skin, ligaments, and sclerae, thereby explaining dentinogenesis imperfecta and abnormalities of other tissues that contain this fibrillar protein. Affected individuals can manifest ligamentous laxity, with joint hypermobility, diaphoresis, bruising, fragile and discolored teeth, and hearing loss (in ~50% of those patients <30 years and in nearly all who are older).[87] Deafness typically has conductive or mixed, but sometimes sensorineural, pathogenesis.[87] Scleral color can be blue or gray and subtle or striking. In severe classic OI, there is often a disproportionately large head, triangular face, high-pitched voice, short stature, scoliosis, herniae, and chest deformity. Mitral valve "clicks" are not uncommon, but cardiac disease is unusual. Intelligence is typically normal.

Type 1 OI features bluish sclerae (especially during childhood), relatively mild osteopenia with infrequent fractures (deformity is uncommon or slight), and deafness that manifests during early adult life.[86] Older women with type 1 OI can be mistaken as having postmenopausal osteoporosis. Subclassification into 1-A and 1-B is based on absence or rarely presence, respectively, of dentinogenesis imperfecta. Approximately one-third of type 1 patients represent new type 1 collagen mutations.

Type 2 OI causes prematurity, small size for gestational age, markedly soft skull, short and bowed limbs, numerous fractures, and a small thorax leading to death within the first weeks of life because of respiratory complications.[87]

Type 3 OI features recurrent fractures, including of long bone growth plates, leading to progressive skeletal deformity and short stature during childhood. Dentinogenesis imperfecta is common. Thoracic distortion predisposes to pneumonia.[86]

Type 4 OI features sclerae that appear normal, but with skeletal fractures and deformity, dental disease, and hearing loss.[86]

Now, especially rare autosomal recessive, autosomal dominant, and X-linked "OI"[83-85,88] includes severe forms because of defective enzymatic cross-linking of type 1 collagen.[89] The other disorders can be considered familial osteoporoses.

Radiographic Features. Radiographic findings in severe OI include generalized osteopenia, modeling defects featuring gracile long bones because of impaired periosteal bone formation, perhaps from weakness or non–weight-bearing, and deformities from recurrent fractures. Some severely affected infants have micromelia with wide major long bones. Recurrent fractures can deform vertebrae. Wormian bones in the skull, of significant number and size, are common but not pathognomonic.[90] Platybasia that can progress to basilar impression and excessive pneumatization of the frontal and mastoid sinuses are common with severe disease.[90] Appendicular abnormalities may worsen markedly during growth, a characteristic of "progressively deforming" type 3 OI. Here, "popcorn calcifications" within metaphyses, likely from traumatic fragmentation of growth-plate cartilage, appear during childhood in major long bones,[91] compromising their growth and contributing to short stature. They resolve when endochondral cartilage fully mineralizes and is replaced by bone at skeletal maturity. Fractures in OI often heal normally. Occasionally, hypertrophic callus, one of the hallmarks of type 5 OI because of a heterozygous mutation of the 5′-UTR of *IFITM5*, is mistaken for skeletal malignancy.

Osteoarthritis is frequent in ambulatory adults when there is bone deformity.

Laboratory Findings. Routine assessments of mineral metabolism are typically normal in OI, except hypercalciuria commonly occurs in severely affected children,[92] and nephrolithiasis seems prevalent during adult life. Elevations of markers of bone turnover characterize severe OI.

Bone histology usually reflects the abnormal skeletal matrix, especially in severely affected patients. Polarized-light microscopy often shows abundant disorganized woven bone, or thin collagen bundles in lamellar osseous tissue. Numerous osteocytes in cortical bone of some patients[93] could signify decreased bone produced by individual osteoblasts, yet many that are active. Hence, skeletal turnover can be rapid.[94]

Cause and Pathogenesis. The large size and complex nature of type 1 collagen is such that many and various heterozygous mutations are found within the genes that encode either pro-α_1 or pro-α_2 type 1 collagen.[1] The newer forms of OI reflect a considerable variety of causal genes.[83-85] A hallmark of classic OI is the paucity of often defective type 1 collagen synthesized by skin fibroblasts in culture.

Treatment. Support groups (e.g., the Osteogenesis Imperfecta Foundation in the United States) are important sources of comfort and lay-language information for OI patients and their families.

Genetic counseling for OI increasingly involves gene mutation analysis available in "OI panels" at commercial laboratories. Sporadic cases can reflect germline mosaicism for new dominant mutations in the genes that encode type 1 collagen. The recurrence risk for severe type 2 OI is 5% to 10%.[96] Some mildly affected individuals are mosaic, passing on severe disease to their child.

Prenatal detection of severe OI is possible by ultrasound examination at 14 to 18 weeks' gestation.[95] Now, mutation analysis is increasingly available for diagnosis.[83-85]

Treatment of OI often requires expert orthopedic, rehabilitative, and dental care.[96-97] Rodding of long bones and bracing of the lower limbs has enabled some severely affected children to walk. Stapes surgery has helped hearing loss. Bisphosphonate treatment has become commonplace for pediatric OI, but few trials have been placebo controlled.[98] Efficacy of bisphosphonates or teriparatide is less certain for affected adults. Increasing numbers of mouse models for OI are improving ways to test potential therapies. Clinical studies involve denosumab, parathyroid hormone analogs, and monoclonal antibodies to sclerostin and TGF-β1 (see ClinicalTrials.Gov).

KEY POINTS

- Mendelian osteoporoses have autosomal dominant osteogenesis imperfecta ("OI") as the paradigm "brittle bone disease."
- OI is typically caused by mutation of either of the two genes that encode the type I collagen heterotrimer. Some newer, autosomal recessive OIs feature gene defects that disrupt modification of this collagen.
- Heritable osteoporoses are increasingly identified by mutation analyses using panels of associated genes.
- Treatment may involve drugs or biologics used for acquired osteoporoses.

PAGET DISEASE OF BONE

Paget disease of bone (PDB) was once common in the United States (i.e., prevalence at least 1%, and perhaps 2%),[99] but now seems less so. The pathogenesis centers on focally accelerated skeletal remodeling in adults. Initially, a "wave" of osteoclast-mediated osteolysis moves slowly but relentlessly through a bone, followed by disorganized skeletal repair featuring bony expansion with hyperostosis and osteosclerosis. Pagetic bone is weak and can cause pain, fracture, and deformity. Skull involvement can lead to deafness of multifactorial pathogenesis. Rarely, pagetic lesions transform to osteosarcoma or chondrosarcoma.

Although the precise cause of all PDB remains uncertain, environmental factors acting upon a genetic predisposition (i.e., "diathesis") seem important. PDB prevalence in first-degree relatives is 12% to 40%, representing a 7-fold enhanced risk.[99] In fact, despite its classic focal appearance, PDB can manifest generalized acceleration in skeletal remodeling. Within PDB lesions, paramyxovirus infection could activate osteoclasts and their precursor cells. Here, the marrow microenvironment is especially osteoclastogenic and contains measles virus transcripts and proteins, perhaps explaining the inclusion bodies seen in PDB osteoclasts.[100]

Increasing evidence indicates that PDB is sometimes heritable. Characterization of familial expansile osteolysis (FEO), with its resemblance to PDB and obvious autosomal dominant inheritance,[101] rekindled earlier brief interest to identify a genetic basis for PDB (see later). Then, the rare syndrome called inclusion body myopathy with early-onset PDB and frontotemporal dementia was shown to be caused by autosomal dominant loss of function mutation of the gene that encodes valosin-containing protein.[102] PDB, particularly in French Canadians, was found to reflect heterozygous loss-of-function mutation in the *SQSTM1* gene that encodes sequestosome.[103] *SQSTM1* defects have been identified worldwide in large numbers of familial and some sporadic instances of PDB. Valosin-containing protein and SQSTM1 both seem to participate in ubiquination.[104] Now, predisposition to PDB is understood to be even more genetically heterogeneous.[105]

KEY POINTS

- Paget disease of bone (PDB), once "the second most common metabolic bone disease," seems to be waning in prevalence and severity worldwide.
- PDB diathesis ("predisposition") has an increasing number of associated genes.
- Paramyxovirus infection seems to underlie the characteristic focal bone lesions.
- Bone antiresorptive therapy, particularly the bisphosphonate zoledronic acid given intravenously, substantially controls active PDB.

DISORDERS OF INCREASED RANK/OPG/RANKL/NF-κB SIGNALING

Several other ultrarare but highly informative heritable disorders feature generalized acceleration of bone remodeling, specifically because of genetic defects within the receptor activator of nuclear factor-κB (RANK)/osteoprotegerin (OPG)/RANK ligand (RANKL)/NF-κB signaling pathway.[106]

Discovery of this regulatory system for osteoclast formation and action began with identification of new ligands and receptors in the tumor necrosis factor (TNF) superfamily, and then characterization of their knockout and transgenic mouse models.[107] Subsequently, clinical appreciation came from revelation of the genetic basis for FEO and several unique disease phenotypes from constitutive activation of RANK, and for juvenile Paget disease (JPD) owing to OPG deficiency.

In 2000, heterozygous gain-of-function 18-bp and 27-bp tandem duplications in the gene *TNFRSF11A* that encodes RANK were discovered to cause FEO and early-onset PDB,[101] respectively. In 2002, homozygous selective deletion of the *TNFRSF11B* gene that encodes OPG was found to cause JPD.[108] Then, in 2002 and 2014, similar 15-bp and 12-bp duplications in the gene encoding RANK explained expansile skeletal hyperphosphatasia[109] and panostotic expansile bone disease,[110] respectively.

Familial Expansile Osteolysis

FEO features deafness in childhood followed by generalized acceleration of skeletal remodeling and then focal lytic expansion of one or more major appendicular bones causing pain, fracture, deformity, and sometimes osteosarcoma in adult life.[106] Perhaps skeletal injury partly explains the focal bone lesions, which initially resemble the clinical, radiographic, and histopathologic appearance of PDB in its early osteolytic phase, starting near the end of a long bone and then slowly advancing. A tibia is most often involved. However, expanded bones eventually instead become shell-like and fat-filled, rather than exhibiting the hyperostosis and osteosclerosis of advanced PDB. These lesions are unresponsive to otherwise effective antiresorptive therapy given early on. Generalized osteopenia featuring a coarse trabecular pattern in adult life discloses the systemic nature of FEO bone disease. Conductive deafness presents in FEO as young as 4 years of age, but more commonly in the second decade, and ultimately becomes mixed-type. An especially remarkable finding in FEO is "idiopathic external resorption" that destroys teeth. The teeth have small pulp chambers and root canals, and fracture of adult dentition can also occur. Elevation in serum alkaline phosphatase activity and other biochemical markers of bone remodeling are partly because of the diffuse acceleration in skeletal turnover. Light microscopy of focal FEO bone lesions early on shows filigree-like trabeculae of woven bone, abundant osteoclasts and osteoblasts lining trabeculae, giant osteoclasts with bizarre shapes and numerous nuclei, and fibrous marrow. Electron microscopy reveals nuclear inclusions with ultrastructure like paramyxoviridae, as reported in PDB.[100] However, "mosaic bone," the hallmark of advanced PDB, is rare, perhaps explained by the extreme remodeling. Woven bone seems unable to mature or remodel into cortical bone. Perhaps focal loss of trabecular bone and instead fat-occupied medullary spaces in advanced FEO reflects mesenchymal stem cell differentiation into adipocytes. The molecular defect causing FEO was identified in 2000.[101] In three kindreds, an in-frame 18-bp tandem duplication was found within exon 1 of *TNFRSF11A* encoding RANK. Transfection studies suggested the RANK signal peptide extension trapped this receptor intracellularly, thereby increasing NF-κB activity.[101] Subsequent studies could not explain the enhanced osteoclastogenesis. Similar in-frame duplications in exon 1 encoding the signal peptide of RANK explain other disorders of constitutively enhanced RANK signaling.[109,110]

Osteoprotegerin Deficiency

JPD, also called idiopathic or hereditary hyperphosphatasia,[1] is usually diagnosed in infants or young children. A relatively mild form includes mental retardation.[111]

JPD affects the entire skeleton and causes bone pain, fracture, and deformity. Premature loss of teeth and deafness are typical. Radiographs show marked undertubulation of long bones with thin cortices (Fig. 59.6). Rapid skeletal remodeling, confirmed by

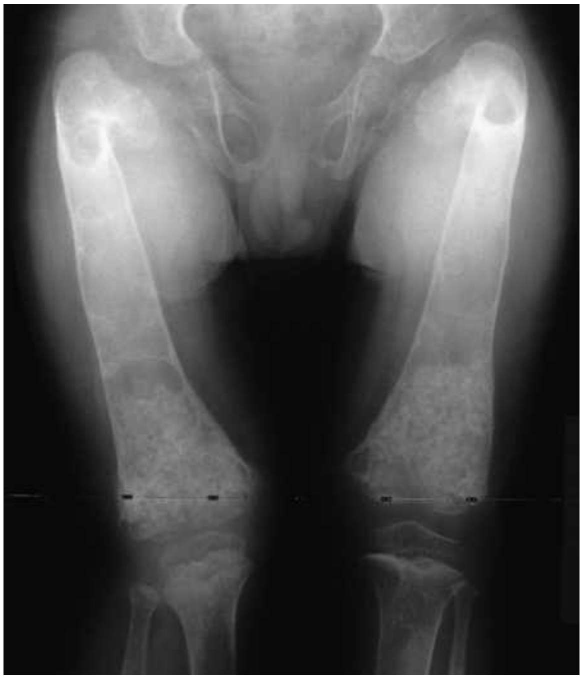


Fig. 59.6 Osteoprotegerin deficiency. The femurs of this boy with juvenile Paget disease are widened, with thin cortices and irregular ossifications. Marked angular deformity is present proximally.

histopathologic findings, leads to substantially elevated biochemical markers of bone turnover.[108,111] In fact, the mosaic pattern of bone characteristic of advanced PDB is not found in JPD or in the disorders of constitutive RANK activation.

Mild JPD features fewer fractures, less bony deformity, hyperostosis and osteosclerosis, as well as biochemical and histological evidence of generalized accelerated skeletal turnover,[108] which early on prompted objection to calling it a type of PDB. In 2002, OPG deficiency from homozygous absence of *TNFRSF11B* provided an obvious mechanism for the osteopathy.[108] OPG is normally on, or released into the marrow space by, preosteoblasts and osteoblasts, and acts as a decoy receptor for RANKL.[107] Thus, OPG deficiency enhances RANKL activity, which markedly accelerates osteoclastogenesis and bone turnover.[112]

TNFRSF11B knockout mice reveal a role for OPG also preventing vascular calcification.[112] Indeed, JPD patients can have striking histopathologic changes consistent with pseudoxanthoma elasticum,[113] including granular and coarse deposits of calcium in the membranes and intima of the muscular arteries and arterioles seen in all autopsy tissues from a young man with JPD.[113] Patients with JPD from OPG deficiency develop blindness from a retinopathy that seems to derive from microvascular calcification.[114]

Antiresorptive treatment using calcitonin or bisphosphonates has been beneficial for JPD,[103] but hypocalcemia may develop rapidly, especially with denosumab administration. Recombinant OPG was effective for two affected adult siblings.[115]

KEY POINTS

- Especially rare Mendelian disorders from gain- or loss-of-function mutations of the genes that encode receptor activator of NF-κB (RANK) or the decoy receptor osteoprotegerin (OPG) for RANK ligand (RANKL), respectively, illustrate the crucial role OPG/RANKL/RANK/NF-κB signaling has in humans for osteoclastogenesis and osteoclast action.
- All feature generalized acceleration of bone remodeling, some with focal lesions resembling those of Paget disease of bone, and respond to antiresorptive therapies.

REFERENCES

1. Online Mendelian Inheritance in Man. OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). Available at: http://www.ncbi.nlm.nih.gov/omim/.
2. Thakker RV, Whyte MP, Eisman J, et al., eds. *Genetics of Bone Biology and Skeletal Disease*. 2nd ed. San Diego: Academic Press; 2018.
3. Whyte MP. Sclerosing bone disorders. In: Bilezikian JP, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*. 9th ed. Ames: Wiley-Blackwell; 2019:825–838.
4. Frame B, Honasoge M, Kottamasu SR. *Osteosclerosis, Hyperostosis, and Related Disorders*. St Louis: Elsevier; 1987:1–375.
5. Albers-Schönberg H. Rontgenbilder einer seltenen, Knochenerkrankung. *Munch Med Wochenschr*. 1904;51:365.
6. Penna S, Capo V, Palagano E, et al. One disease, many genes: implications for the treatment of osteopetroses. *Front Endocrinol*. 2019;10:85.
7. Marks Jr SC. Osteopetrosis. Multiple pathways for the interception of osteoclast function. *Appl Pathol*. 1987;5:172–183.
8. Sly WS, Whyte MP, Sundaram V, et al. Carbonic anhydrase II deficiency in 12 families with the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification. *N Engl J Med*. 1985;313:139–145.
9. Awad M, Al-Ashwal AA, Sakati N, et al. Long-term follow up of carbonic anhydrase II deficiency syndrome. *Saudi Med J*. 2002;23:25–29.
10. Whyte MP, Wenkert D, McAlister WH, et al. Dysosteosclerosis presents as an "osteoclast-poor" form of osteopetrosis: comprehensive investigation of a 3-year-old girl and literature review. *J Bone Miner Res*. 2010;25:2527–2539.
11. Di Rocco M, Buoncompagni A, Loy A, et al. Osteopetrorickets: case report. *Eur J Paediatr Neurol*. 2000;159:579–581.
12. Rao VM, Dalinka MK, Mitchell DG, et al. Osteopetrosis: MR characteristics at 1.5 T. *Radiology*. 1986;161:217–220.
13. Whyte MP, Chines A, Silva Jr DP, et al. Creatine kinase brain isoenzyme (BB-CK) presence in serum distinguishes osteopetrosis among the sclerosing bone disorders. *J Bone Miner Res*. 1996;11: 1438–1443.
14. Campos-Xavier AB, Casanova JL, Doumaz Y, et al. Intrafamilial phenotypic variability of osteopetrosis due to chloride channel 7 (CLCN7) mutations. *Am J Med Genet*. 2005;133:216–218.
15. Pangrazio A, Fasth A, Sbardellati A, et al. SNX10 mutations define a subgroup of human autosomal recessive osteopetrosis with variable clinical severity. *J Bone Miner Res*. 2013;28:1041–1049.
16. Guerrini MM, Sobacchi C, Cassani B, et al. Human osteoclast-poor osteopetrosis with hypogamma-globulinemia due to TNFRSF11A (RANK) mutations. *Am J Hum Genet*. 2008;83:64–76.
17. Sobacchi C, Frattini A, Guerrini MM, et al. Osteoclast-poor human osteopetrosis due to mutations in the gene encoding RANKL. *Nat Genet*. 2007;39:960–962.
18. Whyte MP, McAlister WH, Novack DV, et al. Bisphosphonate-induced osteopetrosis: novel bone modeling defects, metaphyseal osteopenia, and osteosclerosis fractures after drug exposure ceases. *J Bone Miner Res*. 2008;23:1698–1707.
19. Teti A, Econs MJ. Osteopetroses, emphasizing potential approaches to treatment. *Bone*. 2017;102:50–59.
20. Key LL, Rodriguiz RN, Willi SM, et al. Recombinant human interferon gamma therapy for osteopetrosis. *N Engl J Med*. 1995;332:1594–1599.
21. Iacobini M, Migliaccio S, Roggini M, et al. Case report: apparent cure of a newborn with malignant osteopetrosis using prednisone therapy. *J Bone Miner Res*. 2001;16:2356–2360.
22. Dorantes LM, Mejia AM, Dorantes S. Juvenile osteopetrosis: effects of blood and bone of prednisone and low calcium, high phosphate diet. *Arch Dis Child*. 1986;61:666–670.
23. McMahon C, Will A, Hu P, et al. Bone marrow transplantation corrects osteopetrosis in the carbonic anhydrase II deficiency syndrome. *Blood*. 2001;97:1947–1950.
24. Driessen GJ, Gerritsen EJ, Fischer A, et al. Long-term outcome of haematopoietic stem cell transplantation in autosomal recessive osteopetrosis: an EBMT report. *Bone Marrow Transplant*. 2003;32:657–663.
25. Tsuji Y, Ito S, Isoda T, et al. Successful nonmyeloablative cord blood transplantation for an infant with malignant infantile osteopetrosis. *J Pediatr Hematol Oncol*. 2005;27:495–498.
26. Gelb BD, Brömme D, Desnick RJ, et al. Pycnodysostosis: cathepsin K deficiency. In: Scriver CR, Beaudet AL, Sly WS, eds. *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed. New York: McGraw-Hill; 2001:3453–3468.
27. Wolpowitz A, Matisson A. A comparative study of pycnodysostosis, cleidocranial dysostosis, osteopetrosis and acro-osteolysis. *S Afr Med J*. 1974;48:1011–1118.
28. Everts V, Aronson DC, Beertsen W. Phagocytosis of bone collagen by osteoclasts in two cases of pycnodysostosis. *Calcif Tissue Int*. 1985;37: 25–31.
29. Motyckova G, Fisher DE. Pycnodysostosis: role and regulation of cathepsin K in osteoclast function and human disease. *Curr Mol Med*. 2002;2:407–421.
30. Beneton MNC, Harris S, Kanis JA. Paramyxovirus-like inclusions in two cases of pycnodysostosis. *Bone*. 1987;8:211–217.
31. Soliman AT, Rajab A, AlSalmi I, et al. Defective growth hormone secretion in children with pycnodysostosis and improved linear growth after growth hormone treatment. *Arch Dis Child*. 1996;75:242–244.
32. Bizaqui V, Michat C, Baujat G, et al. Pycnodysostosis: natural history and management guidelines from 27 French cases and a literature review. *Clin Genet*. 2019;96:309–316.

33. Edelson JG, Obad S, Geiger R, et al. Pycnodysostosis: orthopedic aspects, with a description of 14 new cases. *Clin Orthop*. 1992;280:263–276.
34. Yerawar C, Kabde A, Deokar P. Kenny-Caffey syndrome type 2. *QJM*. 2021;114:267–269.
35. Saito T, Kinoshita A, Ki Y, et al. Domain-specific mutations of a transforming growth factor (TGF)-β1 latency-associated peptide cause Camurati-Engelmann disease because of the formation of a constitutively active form of TGF-β1. *J Biol Chem*. 2001;276:11469–11472.
36. Wallace SE, Lachman RS, Mekikian PB, et al. Marked phenotypic variability in progressive diaphyseal dysplasia (Camurati-Engelmann disease): report of a four-generation pedigree, identification of a mutation in TGFβ1, and review. *Am J Med Genet*. 2004;129:235–247.
37. Crisp AJ, Brenton DP. Engelmann's disease of bone: a systemic disorder? *Ann Rheum Dis*. 1982;41:183–188.
38. Kumar B, Murphy WA, Whyte MP. Progressive diaphyseal dysplasia (Englemann's disease): scintigraphic-radiologic-clinical correlations. *Radiology*. 1981;140:87–92.
39. Smith R, Walton RJ, Corner BD, et al. Clinical and biochemical studies in Engelmann's disease (progressive diaphyseal dysplasia). *Q J Med*. 1977;46:273–294.
40. Walton KL, Makanji Y, Chen J, et al. Two distinct regions of latency-associated peptide coordinate stability of the latent transforming growth factor-β1 complex. *J Biol Chem*. 2010;285:17029–17037.
41. Tang Y, Wu X, Lei W, et al. TGF-β1-induced migration of bone mesenchymal stems cells couples bone resorption and formation. *Nat Med*. 2009;15:757–765.
42. Labat ML, Bringuier AF, Seebold C, et al. Monocytic origin of fibroblasts: spontaneous transformation of blood monocytes into neo-fibroblastic structures in osteomyelosclerosis and Engelmann's disease. *Biomed Pharmacother*. 1991;45:289–299.
43. Carlson ML, Beatty CW, Neff BA, et al. Skull base manifestations of Camurati-Engelmann disease. *Arch Otolaryngol Head Neck Surg*. 2010;136:566–575.
44. Naveh Y, Alon U, Kaftori JK, et al. Progressive diaphyseal dysplasia: evaluation of corticosteroid therapy. *Pediatrics*. 1985;75:321–323.
45. Inaoka T, Shuke N, Sato J, et al. Scintigraphic evaluation of pamidronate and corticosteroid therapy in a patient with progressive diaphyseal dysplasia (Camurati-Engelmann disease). *Clin Nucl Med*. 2001;26:680–682.
46. Trombetti A, Cortes F, Kaelin A, et al. Intranasal calcitonin reducing bone pain in a patient with Camurati-Engelmann disease. *Scand J Rheumatol*. 2011;41:75–77.
47. Shandilya R, Gadre KS, Sharma J, et al. Infantile cortical hyperostosis (Caffey disease): a report and review of the literature—where are we after 70 years? *J Oral Maxillofac Surg*. 2013;71:1195–1201.
48. Gensure RC, Makitie O, Barclay C, et al. A novel COL1A1 mutation in infantile cortical hyperostosis (Caffey disease) expands the spectrum of collagen-related disorders. *J Clin Invest*. 2005;115:1250–1257.
49. Johnson ML, Recker RR. Exploiting the WNT signaling pathway for clinical purposes. *Curr Osteoporos Rep*. 2017;15:153–161.
50. Van Buchem FSP, Prick JJG, Jaspar HHJ. Hyperostosis Corticalis Generalisata Familiaris (Van Buchem's Disease). Excerpta: Amsterdam, The Netherlands; 1976.
51. Van Lierop AH, Hamdy NA, Papapoulos SE. Glucocorticoids are not always deleterious for bone. *J Bone Miner Res*. 2010;25:2796–2800.
52. Beighton P, Cremin BJ, Hamersma H. The radiology of sclerostenosis. *Br J Radiol*. 1976;49:934–939.
53. Stein SA, Witkop C, Hill S, et al. Sclerostenosis, neurogenetic and pathophysiologic analysis of an American kinship. *Neurology*. 1983;33:267–277.
54. Hamersma H, Gardner J, Beighton P. The natural history of sclerosteosis. *Clin Genet*. 2003;63:192–197.
55. Brunkow ME, Gardner JC, Van Ness J, et al. Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. *Am J Hum Genet*. 2001;68:577–589.
56. Balemans W, Patel N, Ebeling M, et al. Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J Med Genet*. 2002;39:91–97.
57. Kim SJ, Bieganski T, Sohn YB, et al. Identification of signal peptide domain SOST mutations in autosomal dominant craniodiaphyseal dysplasia. *Hum Genet*. 2011;129:497–502.
58. Whyte MP, Amalnath SD, McAlister WH, et al. Sclerosteosis: report of type 1 or 2 in three Indian Tamil families and literature review. *Bone*. 2018;116:321–332.
59. Li X, Zhang Y, Kang H, et al. Sclerostin binds to LRP5/6 and antagonizes canonical Wnt signaling. *J Biol Chem*. 2005;280:19883–19887.
60. Perez-Vicente JA, Rodriguez de Castro E, Lafuente J, et al. Autosomal dominant endosteal hyperostosis. Report of a Spanish family with neurological involvement. *Clin Genet*. 1987;31:161–169.
61. Whyte MP. Searching for gene defects that cause high bone mass (editorial). *Am J Hum Genet*. 1997;60:1309–1311.
62. Whyte MP, McAlister WH, Zhang F, et al. New explanation for autosomal dominant high bone mass: mutation of low-density lipoprotein receptor-related protein 6. *Bone*. 2019;127:228–243.
63. Van Wesenbeeck L, Cleiren E, Gram J, et al. Six novel missense mutations in the LDL receptor-related protein 5 (LRP5) gene in different conditions with an increased bone density. *Am J Hum Genet*. 2003;72:763–771.
64. Bollerslev J, Steiniche T, Melsen F, et al. Structural and histomorphometric studies of iliac crest trabecular and cortical bone in autosomal dominant osteopetrosis: a study of two radiological types. *Bone*. 1986;10:19–24.
65. Yadav VK, Ryu JH, Suda N, et al. Lrp5 controls bone formation by inhibiting serotonin synthesis in the duodenum. *Cell*. 2008;135:825–837.
66. Cui Y, Niziolek PJ, MacDonald BT, et al. Lrp5 functions in bone to regulate bone mass. *Nat Med*. 2011;17:684–691.
67. Rickels MR, Zhang X, Mumm S, et al. Oropharyngeal skeletal disease accompanying high bone mass and novel LRP5 mutation. *J Bone Miner Res*. 2005;20:878–885.
68. Berlin R, Hedensio B, Lilja B, et al. Osteopoikilosis: a clinical and genetic study. *Acta Med Scand*. 1967;18:305–314.
69. Whyte MP, Griffith M, Trani L, et al. Melorheostosis: exome sequencing of an associated dermatosis implicates postzygotic mosaicism of mutated KRAS. *Bone*. 2017;101:145–155.
70. Uitto J, Santa Cruz DJ, Starcher BC, et al. Biochemical and ultrastructural demonstration of elastin accumulation in the skin of the Buschke-Ollendorff syndrome. *J Invest Dermatol*. 1981;76:284–287.
71. Whyte MP, Murphy WA, Seigel BA. 99m Tc-pyrophosphate bone imaging in osteopoikilosis, osteopathia striata, and melorheostosis. *Radiology*. 1978;127:439–443.
72. Hellemans J, Preobrazhenska O, Willaert A, et al. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat Genet*. 2004;36:1213–1218.
73. Yadegari M, Whyte MP, Mumm S, et al. Buschke-Ollendorff syndrome: absence of LEMD3 mutation in an affected family. *Arch Dermatol*. 2010;146:63–68.
74. Rabinow M, Unger F. Syndrome of osteopathia striata, macrocephaly, and cranial sclerosis. *Am J Dis Child*. 1984;138:821–823.
75. Perdu B, Lakeman P, Mortier G, et al. Two novel WTX mutations underscore the unpredictability of male survival in osteopathia striata with cranial sclerosis. *Clin Genet*. 2011;80:383–388.
76. Clements SE, Wessagowit V, Lai-Cheong JE, et al. Focal dermal hypoplasia resulting from a new nonsense mutation, p.E300X, in the PORCN gene. *J Dermatol Sci*. 2008;49:39–42.
77. Zhang Z, Shang C, Zhang Z. Primary hypertrophic osteoarthropathy: an update. *Front Med*. 2013;7:60–64.
78. Seifert W, Kuhnisch J, Tuysuz B, et al. Mutations in the prostaglandin transporter encoding gene SLCO2A1 cause primary hypertrophic osteoarthropathy and isolated digital clubbing. *Hum Mutat*. 2012;33: 660–664.
79. Vogl A, Goldfischer S. Pachydermoperiostosis: primary or idiopathic hypertrophic osteoarthropathy. *Am J Med*. 1962;33:166–187.
80. Nakahigashi K, Otsuka A, Doi H, et al. Prostaglandin E2 increase in pachydermoperiostosis without 15-hydroprostaglandin dehydrogenase mutations. *Acta Derm Venereol*. 2013;93:118–119.
81. Uppal S, Diggle CP, Carr IM, et al. Mutations in 15-hydroxyprostaglandin dehydrogenase cause primary hypertrophic osteoarthropathy. *Nat Genet*. 2008;40:789–793.
82. Ninomiya S, Hara T, Tsurumi H, et al. Myelofibrosis successfully treated with prednisolone in a patient with pachydermoperiostosis. *Intern Med*. 2011;50:2207–2211.

83. Marini JC, Forlino A, Bächinger HP, et al. Osteogenesis imperfecta. *Nat Rev Dis Primers*. 2017;3:17052.
84. Lim J, Grafe I, Alexander S, et al. Genetic causes and mechanisms of osteogenesis imperfecta. *Bone*. 2017;102:40–49.
85. Mäkitie RE, Costantini A, Kämpe A, et al. New insights into monogenic causes of osteoporosis. *Front Endocrinol (Lausanne)*. 2019;10:70.
86. Sillence D. Osteogenesis imperfecta: an expanding panorama of variants. *Clin Orthop*. 1981;159:11–25.
87. Pedersen U. Hearing loss in patients with osteogenesis imperfecta. *Scand Audiol*. 1984;13:67–74.
88. Laine CM, Joeng KS, Campeau PM, et al. WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. *N Engl J Med*. 2013;368:1809–1816.
89. Baldridge D, Schwarze U, Morello R, et al. CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. *Hum Mutat*. 2008;29:1435–1442.
90. Cremin B, Goodman H, Prax M, et al. Wormian bones in osteogenesis imperfecta and other disorders. *Skeletal Radiol*. 1982;8:35–38.
91. Goldman AB, Davidson D, Pavlor H, et al. Popcorn calcifications: a prognostic sign in osteogenesis imperfecta. *Radiology*. 1980;136:351–358.
92. Chines A, Boniface A, McAlister W, et al. Hypercalciuria in osteogenesis imperfecta: a follow-up study to assess renal effects. *Bone*. 1995;16:333–339.
93. Falvo KA, Bullough PG. Osteogenesis imperfecta: a histometric analysis. *J Bone Joint Surg Am*. 1973;55:275–286.
94. Baron R, Gertner JM, Lang R, et al. Increased bone turnover with decreased bone formation by osteoblasts in children with osteogenesis imperfecta tarda. *Pediatr Res*. 1983;17:204–207.
95. Pepin M, Atkinson M, Starman BJ, et al. Strategies and outcomes of prenatal diagnosis for osteogenesis imperfecta: a review of biochemical and molecular studies completed in 129 pregnancies. *Prenat Diagn*. 1997;17:559–570.
96. Marr C, Seasman A, Bishop N. Managing the patient with osteogenesis imperfecta: a multidisciplinary approach. *J Multidiscip Healthc*. 2017;10:145–155.
97. Franzone JM, Shah SA, Wallace MJ, et al. Osteogenesis imperfecta: a pediatric orthopedic perspective. *Orthop Clin North Am*. 2019;50:193–209.
98. Marom R, Rabenhorst BM, Morello R. Osteogenesis imperfecta: an update on clinical features and therapies. *Eur J Endocrinol*. 2020;183:R95–R106.
99. Morales-Piga AA, Rey-Rey JS, Corres-Gonzalez J, et al. Frequency and characteristics of familial aggregation of Paget's disease of bone. *J Bone Miner Res*. 1995;10:663–670.
100. Roodman GD, Windle JJ. Paget disease of bone. *J Clin Invest*. 2005;115:200–208.
101. Hughes AE, Ralston SH, Marken J, et al. Mutations in TNFRSF11A, affecting the signal peptide of RANK, cause familial expansile osteolysis. *Nat Genet*. 2000;24:45–48.
102. Watts GD, Wymer J, Kovach MJ, et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet*. 2004;36:377–381.
103. Laurin N, Brown JP, Morissette J, et al. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. *Am J Hum Genet*. 2002;70:1582–1588.
104. Asai T, Tomita Y, Nakatsuka S, et al. VCP (p97) regulates NfκB signaling pathway, which is important for metastasis of osteosarcoma cell line. *Jpn J Cancer Res*. 2002;93:296–304.
105. Scotto di Carlo F, Pazzaglia L, Mumm S, et al. ZNF687 mutations in an extended cohort of neoplastic transformations in Paget's disease of bone. *J Bone Miner Res*. 2020;35:1974–1980.
106. Whyte MP. Mendelian disorders of RANKL/OPG/RANK/NF-κB signaling. In: Thakker RV, Whyte MP, Eisman J, et al., eds. *Genetics of Bone Biology and Skeletal Disease*, 2nd ed. San Diego: Elsevier; 2018:453–468.
107. Martin TJ. Paracrine regulation of osteoclast formation and activity: milestones in discovery. *J Musculoskelet Neuronal Interact*. 2004;4:243–253.
108. Whyte MP, Obrecht SE, Finnegan PM, et al. Osteoprotegerin deficiency and juvenile Paget's disease. *N Engl J Med*. 2002;347:174–184.
109. Whyte MP, Hughes AE. Expansile skeletal hyperphosphatasia is caused by a 15-base pair tandem duplication in TNFRSF11A encoding RANK and is allelic to familial expansile osteolysis. *J Bone Miner Res*. 2002;17:26–29.
110. Schafer AL, Mumm S, Ivan El-Sayed I, et al. Panostotic expansile bone disease with massive jaw tumor formation and a novel mutation in the signal peptide of RANK. *J Bone Miner Res*. 2014;29:911–921.
111. Golob DS, McAlister WH, Mills BG, et al. Juvenile Paget disease: life-long features of a mildly affected young woman. *J Bone Miner Res*. 1996;11:132–142.
112. Bucay N, Narosi I, Dunstan CR, et al. Osteoprotegerin-deficient mice develop early onset osteoporosis and arterial calcification. *Genes Dev*. 1998;12:260–268.
113. Mitsudo SM. Chronic idiopathic hyperphosphatasia associated with pseudoxanthoma elasticum. *J Bone Joint Surg Am*. 1971;53:303–314.
114. Whyte MP, Singhellakis P, Petersen MB, et al. Juvenile Paget's disease: the second reported, oldest patient is homozygous for the TNFRSF11B Balkan mutation (966_969delTGACinsCTT) which elevates circulating immunoreactive osteoprotegerin levels. *J Bone Miner Res*. 2007;22:938–946.
115. Cundy T, Davidson J, Rutland MD, et al. Recombinant osteoprotegerin for juvenile Paget's disease. *N Engl J Med*. 2005;353:918–923.

60

Osteoporosis

Ian R. Reid and Emma O. Billington

OUTLINE

What Is Osteoporosis?, 1022
Origins of Skeletal Fragility, 1022
  Bone Density, 1023
  Bone Quality, 1024
  Falls and Frailty, 1025
  Nutrition, 1025
  Genetics, 1025
Epidemiology, 1025
  Bone Mineral Density, 1025
  Fractures, 1026
Assessing Patients at Risk of Fracture, 1026
  Fracture Risk Assessment, 1026
  Bone Turnover Markers, 1028
  Skeletal Imaging, 1028
  Other Investigations, 1028
  Lifestyle, 1028
  Antiresorptive Medications, 1029
  Anabolic Medications, 1033
Treatment Decisions, 1035
  Whom to Treat, 1035
  Selection of Treatment, 1035
  How Long to Treat, 1036
  Follow-Up, 1036
Management of Subtypes of Osteoporosis, 1036

WHAT IS OSTEOPOROSIS?

In 1993, a consensus conference defined osteoporosis as:

> *"… a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture."*[1]

The appearance of osteoporotic bone is illustrated in Fig. 60.1. This definition has stood the test of time fairly well. However, how the definition is operationalized in clinical practice has changed considerably since then. Initially, osteoporosis was diagnosed in those with spontaneous fractures or fractures after minimal trauma. The advent of bone densitometry resulted in osteoporosis being redefined in terms of an areal bone mineral density (BMD) more than 2.5 standard deviations below the mean value in the young normal population (i.e., a T-score < -2.5).[2] This approach was widely endorsed, but had the shortcoming of taking account of only one risk factor for fractures, namely bone density. It has subsequently become apparent that only approximately 20% of postmenopausal women with fractures have bone density T-scores less than –2.5.[3] This realization stimulated the development of fracture risk calculators,[4,5] which are now widely used, but with a variety of different intervention thresholds endorsed by various academic and funding bodies. Some groups, such as the National Osteoporosis Foundation (NOF) in the United States, recommend treatment based on several criteria, such as fracture history, bone density, or calculated risk of either hip or major osteoporotic fractures. Sometimes intervention thresholds are equated with diagnostic thresholds, but this is inconsistent. Because intervention thresholds are usually based on pharmacoeconomic analyses, they will vary not only with fracture risk, but also with cost of medication and willingness to pay, factors that would not usually impact on diagnostic criteria.

These problems of operational definition stem from the decision to regard osteoporosis as a disease. If, instead, it is regarded as a condition or process defined by the previous description, then the question of diagnosis does not arise. Rather, the physician's task is to assess fracture risk and make treatment recommendations based on that. This philosophical approach will be adopted in the present chapter.

KEY POINTS

- Osteoporosis is characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in fracture risk. Operational definitions vary and may be unhelpful. We suggest that physicians assess fracture risk and make treatment recommendations accordingly.

ORIGINS OF SKELETAL FRAGILITY

Several processes contribute to the increased fracture risk in osteoporosis. Most emphasis has focused on the reduction in the amount of bone (measured as bone mass or bone density), but there is also a deterioration in bone quality. Also, frailty contributes to increased numbers of falls with age, which also contribute to fracture risk. The many influences at play can be summarized by saying that the key requirements for healthy bone are: (1) a balanced diet to maintain normal lean and fat masses, (2) normal sex hormone levels, (3) safe physical activity, (4) normal levels of hormones that influence bone turnover, and (5) absence of chronic inflammatory disease.

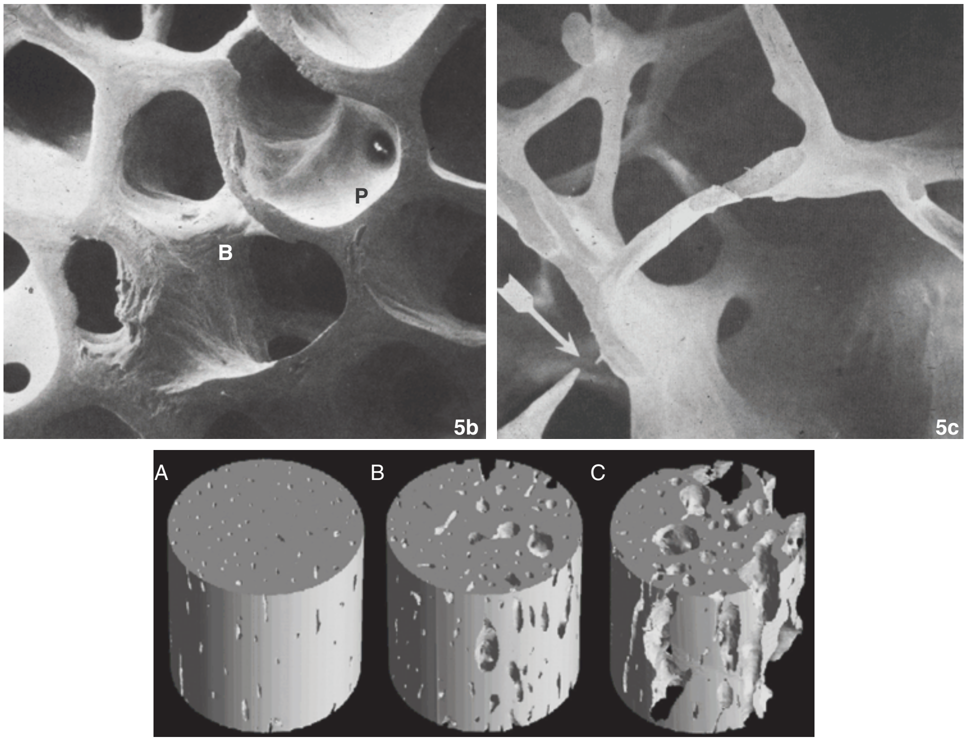


Fig. 60.1 **Normal and osteoporotic bone.** The upper panels show trabecular bone from the iliac crest in normal *(left)* and osteoporotic *(right)* individuals. The transition in architecture from exclusively plate-like structures to predominantly rod-like structures is noted, and results from trabecular perforation. Also note the osteoclast lacuna on the surface of the rod in the middle of the right-hand micrograph where perforation is imminent, and the wasting bone spicules in the lower left of the picture *(arrow)* where a rod has previously been perforated. The lower image demonstrates the normal loss of cortical bone in women aged 20, 61, and 87 years, respectively. P, trabecular plate; B, trabecular bar. (From Dempster DW, Shane E, Horbert W, et al. A simple method for correlative light and scanning electron microscopy of human iliac crest bone biopsies: qualitative observations in normal and osteoporotic subjects. *J Bone Mineral Res.* 1986;1:15–21; Cooper DML, Thomas CDL, Clement JG, et al. Age-dependent change in the 3D structure of cortical porosity at the human femoral midshaft. *Bone.* 2007;40:957–965; used with permission.)

Bone Density

At any age, bone density is determined by the amount of bone laid down during growth and the subsequent changes resulting from aging, drugs, or disease. Thus, factors that maximize bone mass gain during childhood and adolescence (e.g., global nutritional status, physical activity, timing of puberty, genes), those that maintain it during adulthood (normal gonadal function, nutrition, exercise), and those that delay or slow its loss during menopause and old age (timing of menopause, fat and lean masses, physical activity, vitamin D status) will reduce fracture risk. Throughout life, bone mass is also influenced by intercurrent illness, body weight, lifestyle factors such as exercise, and some medications, particularly glucocorticoids. Some of the common pathological and pharmacological contributors to osteoporosis are listed in Table 60.1.

Bone is a dynamic tissue and is constantly remodeled by the coordinated activities of osteoclasts (which remove old bone) and osteoblasts (which manufacture new collagen matrix, which is subsequently mineralized to produce bone). The activities of these cell types are coordinated by osteocytes, which sit in lacunae embedded in bone and have cell processes connecting adjacent osteocytes and communicating with the bone surface. This three-dimensional osteocyte network extends throughout bone and senses skeletal loading, as well as regulating the activities of both osteoblasts and osteoclasts. Osteocytes produce receptor activator of nuclear factor kappa-B ligand (RANKL; a protein that stimulates osteoclast development) and sclerostin (a protein that inhibits osteoblastic activity). Production of these factors is regulated by skeletal loading, circulating hormones, and local cytokines. At any time, approximately 20% of the trabecular bone surface is undergoing remodeling and, on average, the skeleton is replaced approximately every 10 years as a result of this process.

Bone is lost when bone resorption by osteoclasts exceeds the bone-forming activity of osteoblasts. In postmenopausal osteoporosis, the prime driver of bone loss is increased osteoclast activity resulting from the loss of the restraining effect of estrogen on the recruitment and activity of these cells. This results in losses of 20% to 30% of bone density at trabecular sites and 5% to 10% at cortical sites in the decade following menopause. The pivotal role of estrogen is demonstrated by the long-term maintenance of bone density with postmenopausal estrogen replacement.[6] Androgen status is also important in maintaining bone density in older men, and decline in free testosterone concentrations with age closely parallels bone loss in older men, which occurs at annual rates of 0.5% to 1%. Conversion of androgens to estrogens may be an important mechanism of their action on bone in men.[7] Osteoclast overactivity can also lead to perforation of trabeculae, resulting in deterioration in the microarchitecture of bone and leading to greater increases in fracture risk than the decline in bone density alone would predict.[8] It is likely that not all bone loss in older men and women is attributable to falling sex hormone levels. Multiple other aging processes result in cell senescence, including oxidative stress from increased mitochondrial production of reactive oxygen species,

TABLE 60.1 Conditions and Drugs Associated With Osteoporosis

Inflammatory Disorders
Rheumatoid arthritis
Inflammatory bowel disease
Cystic fibrosis
Bone Marrow Disorders
Multiple myeloma
Mastocytosis
Leukaemia
Hypogonadism
Athletic amenorrhoea
Haemochromatosis
Turner syndrome
Klinefelter syndrome
Postchemotherapy
Hypopituitarism
Malabsorption
Coeliac disease
Gastrectomy
Bariatric surgery
Liver disease
Total parenteral nutrition
Endocrine Disorders
Cushing syndrome
Diabetes
Thyrotoxicosis
?Hyperparathyroidism

Low Body Weight
Anorexia nervosa
Human immunodeficiency virus infection
Immobilization
Parkinson disease
Poliomyelitis
Cerebral palsy
Paraplegia
Defective Synthesis of Connective Tissue
Osteogenesis imperfecta
Marfan syndrome
Homocystinuria
Miscellaneous
Chronic obstructive lung disease
Congestive heart failure
Pregnancy/lactation
Ankylosing spondylitis
Hypercalciuric nephrolithiasis
Depression
Drugs
Glucocorticoids, Alcohol,
Medroxyprogesterone acetate
Anticonvulsants, Methotrexate
Heparin, Cyclosporin, Omeprazole
Aromatase inhibitors, Glitazones
Androgen deprivation therapy

telomere shortening, and impairment of DNA repair mechanisms. Bone cell senescence is associated with reduced osteoblast and osteocyte differentiation, as well as alterations in the stiffness and quality of the bone matrix. A more detailed account of cellular pathogenesis of bone loss is provided by Eastell et al.[9]

In contrast to postmenopausal bone loss, glucocorticoid treatment causes apoptosis of osteoblasts and osteocytes, resulting in inhibition of bone formation, although transient increases in bone resorption have also been reported early during treatment with these drugs.

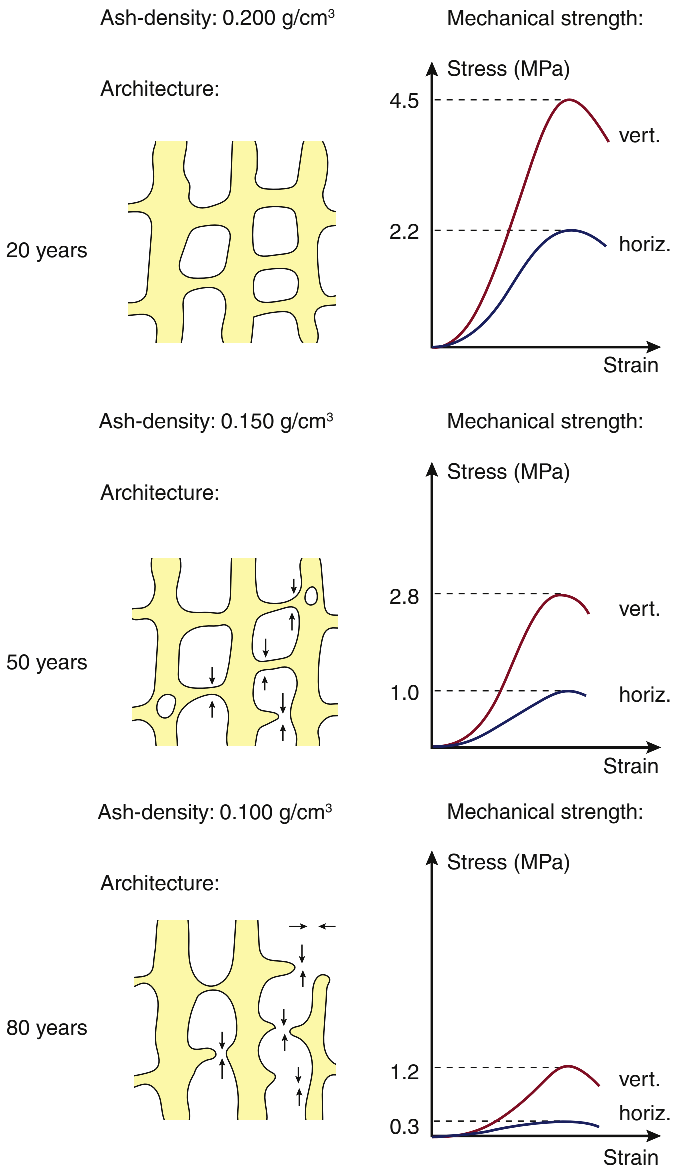


Fig. 60.2 Typical relationships among ash-density, architecture, and mechanical strength (vertical and horizontal) for vertebral trabecular bone obtained from three different age groups: 20 years, 50 years, and 80 years. (From Mosekilde L. Vertebral structure and strength in vivo and in vitro. *Calcif Tissue Int.* 1993;53:S121–S126; used with permission.)

Bone Quality

Epidemiological studies have made clear that bone density is not the only bone parameter that determines fracture risk. For instance, an 80-year-old woman with a femoral neck T-score of –3 is six times more likely to have a hip fracture than a 50-year-old woman with the same bone density, and a similar increase in fracture risk independent of bone density has been demonstrated in glucocorticoid users.[10] Both these observations indicate that there are elements of bone structure that are not captured with bone density measurement, and these are collectively referred to as bone quality. The components of bone quality are continuing to be defined, but include the material properties of both the protein matrix and the mineral phase of bone, as well as bone architecture. The integrity of collagen cross-linking, the state of cross-link isomerization, and content of advanced glycation end-products (resulting from nonenzymatic glycation of collagen) impact on skeletal strength. These aspects of bone matrix are influenced by the rate of bone turnover, which is regulated by sex hormones, cytokines, and genetic factors. Advanced glycation end-products are also influenced by ambient glucose concentrations and are increased in diabetes mellitus. This may contribute to the higher fracture risk than predicted by BMD in type 2 diabetes,[11] although in type 1 diabetes BMD is also reduced. Animal studies have suggested that the presence of microcracks in bone related to low turnover also reduce skeletal strength,[12] but a human study found that the compressive strength of vertebral trabecular bone was primarily determined by the amount of bone, not by cross-link profile, microdamage, or mineralization.[13] In pathological situations, such as severe vitamin D deficiency, bone mineralization can be substantially impaired, with a resultant decrease in bone strength.

The microarchitecture of bone impacts on its strength. As bone mass is lost, perforations occur in individual trabeculae (Fig. 60.1), resulting in a much greater decrease in bone strength than the decrease in mass would predict (Fig. 60.2). Consistent with this, loss of trabecular continuity is more frequently observed in patients with osteoporotic fractures than in controls. High bone turnover contributes to trabecular perforation by increasing the likelihood of resorption lacunae coinciding on opposite sides of a trabecula. Even if this proximity of lacunae does not lead to perforation, high turnover can increase trabecular fragility by causing critical trabecular narrowing where resorption lacunae are located close to one another on opposite sides of trabeculae. This may explain the independent association of bone turnover with fracture risk, which is probably also contributed to by the greater bone loss observed in high turnover states. In the cortex, increased osteoclast activity causes increased porosity (Fig. 60.1) and loss of strength, a process sometimes referred to as *trabecularization*.[14] In the first 15 years after menopause the greater surface area of trabecular bone results in its rate of loss being greater than in cortical bone. Subsequently, the increasing porosity of the cortex increases surface area in this compartment, whereas loss of trabecular elements reduces that surface area, so cortical loss exceeds that of trabecular bone.[14] Measures based on trabecular microarchitecture and cortical porosity appear to provide superior fracture prediction in older women than bone density measurement or the FRAX algorithm.[15]

Macroarchitecture is also important in determining skeletal fragility. One example of this is the positive association of hip axis length with hip fracture risk.[16] Increases in hip axis length and height (also a risk factor for fracture) probably have contributed to the secular increases in hip fracture risk as global nutrition has improved over the last 60 years. Macroarchitectural changes may partially offset the effects of age-related bone loss because long bones expand with age, which increases resistance to both bending and torsion. Other macroarchitectural parameters, such as outer diameter of the proximal femur and neck-shaft angle, have been associated with fracture risk in some populations.

Falls and Frailty

Epidemiology studies clearly demonstrate that the frequency of falls is closely related to the risk of fracture, and falls are the principal cause of skeletal trauma in the elderly. Propensity to fall is influenced by frailty, neuromuscular abnormalities, eyesight, sedative drugs, postural hypotension, and the safety of the home environment (presence of hazards such as loose carpets and electric cords, absence of safety devices such as handrails). Thus, understanding the etiology of frailty and of the resultant falls is a key part of understanding the pathogenesis of fractures in older people; but a detailed dissection of this is beyond the scope of the current chapter.

Nutrition

Nutrition has historically been regarded as a key contributor to the development of osteoporosis, but formal investigation provides only partial support for this hypothesis. Body weight is related to bone density and fracture risk throughout life[17] and is a key element in most fracture risk calculators. This is a dynamic relationship, with changes in weight directly impacting on BMD.

Interestingly, calcium intake has not being found to be useful in calculating fracture risk, and most recent studies do not demonstrate an effect of dietary calcium intake on adult bone density,[18] rates of bone loss in older people,[19] or fracture risk.[20] Earlier reports suggesting that calcium intake played an important role in bone health across the lifespan were substantially based on short-term *calcium* balance studies, which are contradicted by more recent studies of *bone* balance (using dual-energy x-ray absorptiometry [DXA]) over periods of up to 6 years.[19] Because bone mass and BMD are predictive of fracture, these recent results are more compelling. Consistent with these findings, calcium supplements do not impact on fracture risk, although they do have small effects on bone turnover and bone density.[20]

The vitamin D endocrine system is the principal regulator of intestinal calcium absorption. Its role in calcium homeostasis, together with the parathyroid glands, is to maintain normal concentrations of calcium in the extracellular fluid, thus permitting normal neuromuscular and cardiac function, blood coagulation, and other enzymatic activities. It is also critical that calcium concentrations do not exceed the normal range because this can result in deposition of calcium in soft tissues, with their resultant dysfunction. Vitamin D itself is without metabolic activity, but it is a substrate for production of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D. When supply of vitamin D is low enough to compromise production of 1,25-dihydroxyvitamin D, intestinal calcium absorption is impaired, and hypocalcemia and hypophosphatemia result. Hypocalcemia results in secondary hyperparathyroidism, which accelerates bone loss. Ultimately, bone mineralization is compromised, with resultant reduction in bone density and bone strength. Thus, severe vitamin D deficiency (25-hydroxyvitamin D <25 nmol/L) does have a profound effect on bone health, and such deficiency is observed in frail elderly individuals who seldom venture outside, in those who are habitually veiled, and in those with dark skin living at high latitude. In healthy community-dwelling adults with 25-hydroxyvitamin D greater than 25 nmol/L, vitamin D levels do not affect bone health, and supplements do not impact on bone density or fracture risk.[21-23]

Genetics

Genetic factors are important in bone health, with the heritability of bone density being 50% to 70%, and that for indices of bone architecture being 30% to 70%. Heritability of fracture risk is lower (20%–50%) and, at the hip, is highly age-dependent, dropping to near zero for first hip fracture after the age of 79 years.[24] These effects appear to be polygenic, with up to 200 different gene loci being implicated and accounting for up to 12% of the variation in BMD. Some loci are involved in pathways with a recognized role in bone (e.g., RANK and Wnt pathways), but many others are not, pointing to as yet undiscovered regulatory influences on bone. A much smaller number of loci have been shown to be associated with fracture.[25]

KEY POINTS

- Bone loss is a universal process in postmenopausal women and older men resulting from imbalance between bone formation and resorption. Bone architecture and the material properties of the bone matrix also change with age, as does frailty and risk of falling.

EPIDEMIOLOGY

Bone Mineral Density

Fig. 60.3 shows the distribution of BMD across the population, and the impact of age on it. BMD has a normal distribution throughout life, with young women by definition having a mean T-score of 0. By the age of 50 years, the mean T-score has diminished by only 0.38, but bone loss progresses more rapidly from that time through to the end of life. However, BMD remains normally distributed, with a standard deviation very similar to that in the young adult population. This demonstrates the universality of bone loss and strongly suggests that those who start at the upper end of the normal distribution in young adulthood will maintain that relative position subsequently, unless they suffer some specific disease or intervention that promotes bone loss (e.g., substantial weight loss, illness, long-term use of glucocorticoid drugs). BMD can also be presented as a Z-score, which is the number of standard deviations above or below the age-specific mean BMD.

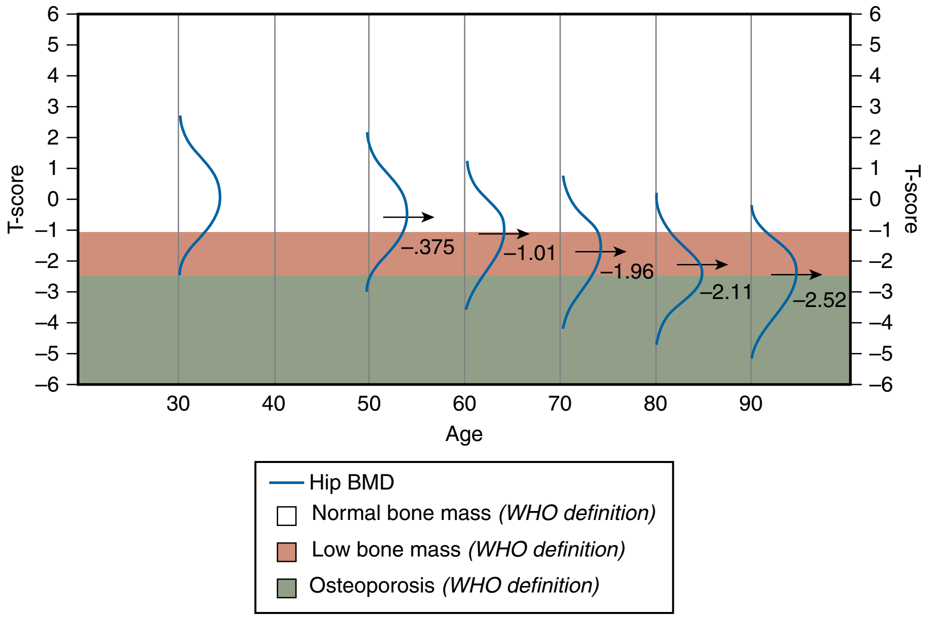


Fig. 60.3 The distributions of bone mineral density (*BMD*) and the decline with age for hip bone mineral density measured at the hip in White women in North America. *WHO*, World Health Organization. (From Cummings SR. Osteoporosis: Review of the evidence for prevention, diagnosis, and treatment and cost-effectiveness analysis — Status report. *Osteoporos Int.* 1998;8:S1–S85; used with permission.)

These data reinforce the suggestion that it is more appropriate to regard bone loss as part of the aging process rather than characterizing osteoporosis as being a "*disease*" because diseases usually affect individuals within a population rather than being experienced across the entire population. The universality of bone loss and the continuous distribution of bone density suggest that "*diagnosing*" osteoporosis is not an appropriate concept because any dichotomization of the population is necessarily arbitrary. Rather, bone density is one of a number of continuously distributed risk factors for fracture.

Fractures

The fall in bone density after the menopause is mirrored in a progressive rise in fracture numbers (Fig. 60.4). As a consequence, approximately one in five White women suffer a hip fracture between the age of 50 years and death, with similar incidences for clinical vertebral fractures and wrist fractures. Although these fractures are regarded as the "classical" osteoporotic fractures, they only represent approximately half of the total number of fractures that occur in this age group.[26] The remaining lifetime risk of any fracture in 60-year-olds is 56% for women and 29% for men.[27] Fracture prevalence is highest in Europe and North America and lowest in Africa, but hip fracture rates are increasing in much of the rest of the world, particularly in Asia.[28] This is contributed to by progressive increases in age-adjusted incidence rates and by a steady expansion in the numbers of individuals in the population over the age of 70 years.

Hip fractures are more common in winter, and usually result from falls, particularly falls to the side. There is an excess mortality at 1 year after hip fracture of approximately 20%, slightly higher in men than in women. About 50% of patients do not recover their former levels of mobility or ability to live independently after hip fracture, so this fracture has the highest costs in terms of loss of quality of life and need for ongoing care.

Vertebral fractures often occur after lifting rather than after falls. Only approximately one-third of vertebral deformities come to medical attention, hence the difference in rates of clinical versus radiographic vertebral fractures. Mortality is increased after vertebral fracture, to a lesser extent than after hip fracture, but the effect persists for greater than 5 years. Vertebral fractures result in height loss and kyphosis, with a resultant decrease in quality of life.

Risk Factors. Many observational studies have documented risk factors for fracture. As Fig. 60.4 indicates, age is a very important risk factor. BMD is also critical, with risk increasing by 1.5- to 2.0-fold for each standard deviation decrease in BMD.[29] Previous fracture increases risk almost 2-fold, with more recent fractures and clinical vertebral fractures carrying greater risk.[30] For this reason, fracture liaison services operate in many hospitals to ensure that those with fractures are offered interventions to prevent subsequent fractures. Almost one-third of women aged 55 to 90 years suffering a new fracture have had a previous fragility fracture,[30] highlighting the need for intervention following first fractures. Low body weight, family history, smoking, fall history, and markers of frailty are also important risk factors. Many risk factors are interdependent, so adjustment for other risk factors usually diminishes the predictive value of a given factor. This has led to the development of fracture risk calculators, which attempt to select the most parsimonious combination of predictive factors.

KEY POINTS

- Fracture risk climbs steeply with age and is also dependent on body weight, fracture history, falls, and bone density. The remaining lifetime risk of any fracture in White 60-year-olds is 56% for women and 29% for men but is lower in non-White persons. Increasing longevity is resulting in increasing total numbers of fractures.

ASSESSING PATIENTS AT RISK OF FRACTURE

Fracture Risk Assessment

The physician's goal in the evaluation and management of osteoporosis is to identify and offer treatment to those at the highest risk of fragility fracture. The cornerstone of this process is fracture risk assessment.

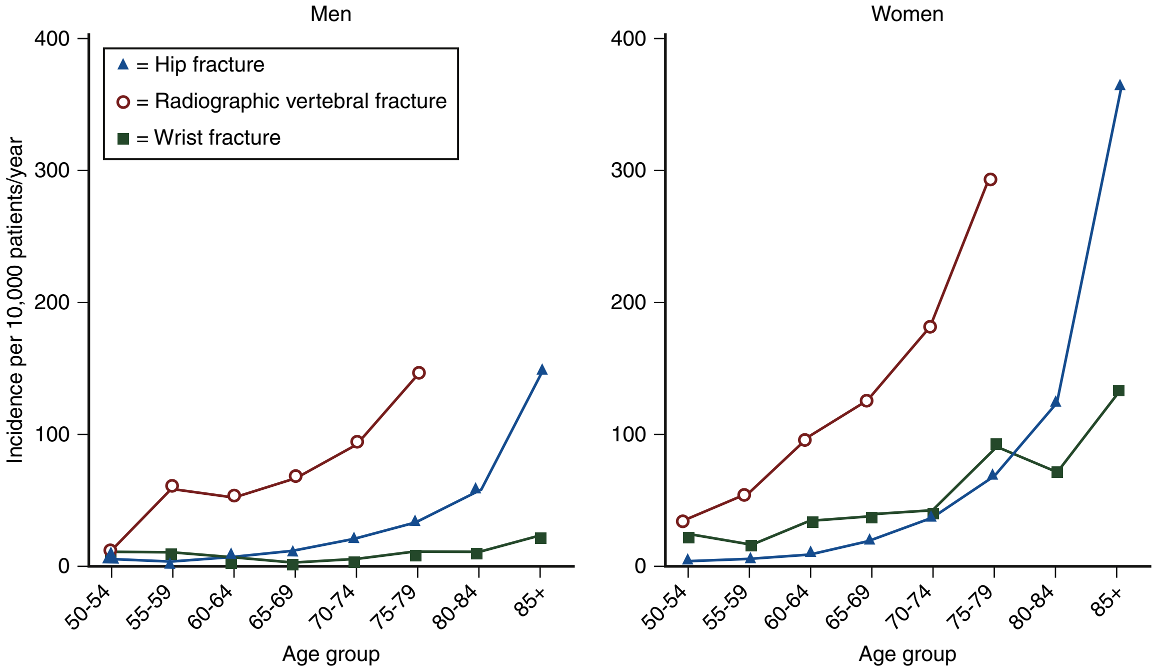


Fig. 60.4 **Fracture incidence by age and gender in England and Wales from 1988 to 1998.** (From van Staa TP, Dennison EM, Leufkens HG, et al. Epidemiology of fractures in England and Wales. *Bone.* 2001;9:517–522; used with permission.)

TABLE 60.2 Characteristics of Two Widely-Used Fracture Risk Calculators

	FRAX	Garvan
Clinical risk factors	Age (40–90 years) Sex Weight Height Previous fracture in adulthood (yes or no) Parent fractured hip Current smoking Glucocorticoid use Rheumatoid arthritis Secondary osteoporosis Alcohol Bone mineral density (optional) Country ± ethnicity	Age (50–96 years) Sex Fractures since age of 50 years (up to three) Falls over last 12 months (up to three) Bone mineral density or weight
Competing risk of death	Allowed for	Not allowed for
Fracture risk outputs	10-year risk of hip fracture or of a major osteoporotic fracture (i.e., hip, clinical spine, humerus, or forearm)	5- or 10-year risk of hip or any fragility fracture

Assessment of fracture risk involves the evaluation of a person's clinical risk factors for fracture (such as age, body mass index, prior low-trauma fracture) and the use of this information, often in conjunction with BMD results, to estimate the individual's absolute risk of sustaining a fragility fracture over a defined time period. The relationships between clinical risk factors, BMD, and fracture incidence are complex. As such, several fracture risk calculators have been developed to facilitate the assessment of fracture risk.[4,5] There are major differences in the variables that have been incorporated into the commonly used calculators, partly because different variables were available in the cohorts on which these calculators were based (Table 60.2). Both the FRAX and Garvan calculators can be used with or without entering BMD. For FRAX, inclusion of BMD makes surprisingly little difference to its predictive efficacy.[31]

In general, calculators produce similar estimates of risk in population studies but may generate disparate results for a given patient as a result of each focusing on different risk factors. Thus, FRAX underestimates risk (in comparison with Garvan) in those with falls or multiple fractures, but ethnicity differences in fracture risk are ignored by other calculators. In glucocorticoid-treated patients, FRAX does not take dosage into account, but a method has been described for revising FRAX estimates according to dose.[32] Similarly, FRAX risk estimates can be adjusted in patients with a considerably lower spine T-score than that at the femoral neck.[33] In general, the magnitude by which adjustment for high glucocorticoid doses and lumbar spine discordance change the initial FRAX estimate is small (on the order of 15%–20%), so the effects of adjustment are most clinically relevant when the initial estimate lies close to the treatment threshold. Fracture risk calculators were developed and validated for use in treatment-naïve individuals, but also appear to be reasonably accurate in those currently or previously receiving osteoporosis treatment.[34] FRAX has been validated in a number of populations, and country- and ethnicity-specific versions have been developed.

Risk assessment with FRAX can be done with or without a BMD result. In the United States, the NOF recommends BMD measurement as the entry point to fracture risk assessment, and, as such, BMD results are incorporated in fracture risk estimates.[35] Conversely, the UK National Osteoporosis Guidelines Group recommends initial fracture risk assessment without BMD, obtaining a BMD only if it has the potential to change management.[36]

When to Assess Fracture Risk. The age and other circumstances that are an indication for fracture risk assessment vary between countries. In Western countries, fracture risk assessment is typically recommended for all women aged 65 years or older[35,36] and in men older than 65 to 75 years. In women or men with clinical risk factors for fracture, risk is usually assessed in those over 50 years, although in high-risk contexts such as sustained glucocorticoid use, risk should be assessed in all age groups.

Bone Turnover Markers

Bone turnover markers are circulating factors that reflect either bone formation or resorption rates. Procollagen type I N-propeptide (PINP), a by-product of collagen synthesis, and the osteoblastic enzyme bone-specific alkaline phosphatase both reflect bone formation. Collagen degradation products can be measured in blood or urine as indices of bone resorption—serum C-terminal telopeptide of type 1 collagen (CTX) or urine N-terminal telopeptide of type 1 collagen. Higher bone turnover tends to be associated with higher fracture risk, but study results are inconsistent, and markers are not used in fracture risk assessment. They are increasingly used to monitor therapy, particularly oral bisphosphonates, and a CTX or PINP level in the lower half of the premenopausal reference range or a reduction of 40% to 50% from baseline indicates adequate delivery of drug to bone. CTX levels are markedly affected by time of day and meals, so are assessed during fasting in the morning.

Skeletal Imaging

DXA is the most widely used imaging modality for fracture risk assessment, but other radiographic techniques have been evaluated. Ultrasound of the heel is cheap and noninvasive but does not predict fracture as well as DXA does.[37] Assessment of BMD using quantitative computed tomography at the spine and femur has similar predictive ability to DXA for fragility fracture, but involves higher radiation exposure.[37] Therefore, performing quantitative computed tomography for the principal purpose of fracture risk assessment is not recommended, although opportunistic review of scans performed for other medical indications can identify individuals with low bone density who should be targeted for fracture risk assessment.[37] High-resolution peripheral quantitative computed tomography (HR-pQCT) has emerged as another promising tool for assessing both BMD and bone quality. This technique involves scanning of the radius and/or tibia and can provide a three-dimensional (volumetric) assessment of BMD. Scans obtained with HR-pQCT are of high enough resolution to permit differentiation between cortical and trabecular bone compartments and provide information about bone microarchitecture. Data obtained with HR-pQCT can also be used to estimate bone strength by finite element analysis. Although HR-pQCT can predict risk of fracture independent of DXA, its use is primarily limited to research at present, owing to the cost and scarcity of the required hardware and expertise.

Currently, the most clinically useful radiographic adjunct to DXA for estimation of fracture risk is trabecular bone score (TBS). TBS is a quantitative measure of the texture of a DXA image. It is determined using lumbar spine images and provides a reflection of trabecular bone quality. Incorporation of TBS has been shown to improve the ability of FRAX to predict fracture.

Other Investigations

In patients with markedly reduced BMD for their age (i.e., Z-score < –2), a search for underlying causes of bone loss should be undertaken. A clinical and medication history, together with physical examination, are important, as is consideration of the conditions listed in Table 60.1. Transiliac bone biopsy with tetracycline labeling is the gold standard for assessment of bone turnover and also provides detailed information about bone quality, but is seldom used, except in patients with severe unexplained osteoporosis or suspicion of osteomalacia.

KEY POINTS

- Fracture risk calculators are used to quantify risk, which guides the need for treatment. In patients with markedly reduced bone mineral density for their age (i.e., Z-score < –2), a search for underlying causes of bone loss should also be undertaken.

Interventions for Reduction of Fracture Risk. The most obvious approach to fracture prevention is to correct any risk factors for fracture that are present. These measures are grouped under the heading *Lifestyle*. The other broad group of interventions is the use of pharmaceuticals, including calcium and/or vitamin D supplements. Pharmaceutical agents for fracture prevention are classified as those inhibiting bone resorption and those that promote bone formation. Some, such as strontium, do not fit into either category, but this agent is now little used because of its marginal effect on nonvertebral fractures and its 60% increase in myocardial infarction, so will not be considered further here.

Lifestyle

Smoking cessation, moderation of alcohol intake, and treatment of the contributing diseases set out in Table 60.1 are obvious measures for reducing fracture risk. Discontinuation of the bone-damaging drugs listed in Table 60.1 is desirable from the point of view of skeletal health, but not always practical. However, dose minimization of glucocorticoids is a goal that should be actively pursued because the osteopenic effects of these drugs are dose related.

The skeleton is immediately responsive to the loads placed upon it, via the osteocyte stress-sensing network and the resultant secretion of regulatory factors by those cells, particularly sclerostin, which directly regulates bone formation. Thus, many studies have demonstrated that regular weight-bearing exercise increases bone density by a few percent, and this is maintained as long as the exercise is continued. Exercise is also of value in strengthening muscles, improving balance, and preventing falls. Metaanalyses have now demonstrated that exercise for falls prevention also results in reduced numbers of fractures.[38] In frail individuals who are falling frequently, pads (hip protectors) worn over the greater trochanters have been used. Some trials have shown major protective benefits to hip fracture from this intervention, although data are inconsistent, and compliance with the hip protectors is sometimes a problem.

Vitamin D. Vitamin D is often regarded as a nutrient, but unfortified diets contain very little vitamin D, and its circulating levels are mainly determined by sunshine exposure, habitual dress, latitude, and exercise. Adequate levels of vitamin D are required for the normal regulation of intestinal calcium absorption. Reduced vitamin D status results in secondary hyperparathyroidism leading to increased bone resorption; more severe vitamin D deficiency leads to low levels of calcium and phosphate in the extracellular fluid that are inadequate for bone mineralization. This is referred to as osteomalacia and results in reduced bone density, bone pain, and muscle weakness, plus, in children, the skeletal deformities of rickets. Vitamin D treatment in osteomalacia increases BMD by as much as 50%. This has led some to regard vitamin D as a potential bone anabolic agent, but the bulk of human data suggest that its role is restricted to the facilitation of intestinal calcium absorption and the prevention of secondary hyperparathyroidism.

Many studies have assessed the effects of vitamin D supplementation on BMD and fractures, but metaanalyses show no beneficial effect.[22] However, when vitamin D supplements are given to those with baseline 25-hydroxyvitamin D levels less than 30 nmol/L, clinically and statistically significant increases in BMD are demonstrable.[21,23] The study by Chapuy et al. carried out in elderly nursing home residents with marked vitamin D deficiency indicated that a combined supplement of calcium plus vitamin D did reduce fracture rates in these women, many of whom are likely to have had osteomalacia, based on their reported 25-hydroxyvitamin D levels.[39]

Routine measurement of vitamin D is not widely supported, because of its expense and the unreliability of some assays. In nursing homes, where patients are seldom outdoors, routine provision of a supplement is appropriate. Among the ambulant population, latitude, skin color, frequency of sunlight exposure, and habitual dress will allow most individuals at risk of severe vitamin D deficiency to be identified. Vitamin D supplements are usually given in a dose of 400 to 1000 IU/day (10–25 μg/day), and up to 2000 IU/day is regarded as safe. The long half-life of vitamin D has led to weekly, monthly, and annual dosing regimens delivering a similar average dose. Annual dosing is now discouraged, as the large bolus required increases the risk of fractures and falls. Daily doses of 4000 IU or more also increase falls and accelerate bone loss.[40,41] Vitamin D_2 and vitamin D_3 are often used interchangeably, but D_3 appears to have marginally greater efficacy, and D_2 is not detected by all 25-hydroxyvitamin D assays. Calcitriol or other 1-hydroxylated vitamin D metabolites have been assessed as treatments for osteoporosis but have not been consistently demonstrated to reduce fracture numbers, so are not used in most countries. Their use can result in hypercalcemia or hypercalciuria.

Diet. Many supplements have been advocated to increase bone health, but substantive data relating to most is lacking. Body weight and changes in weight directly impact on BMD. Fracture risk, particularly at the hip, climbs steeply when body mass index (BMI) is less than 20, although some limb fractures appear to be more common in obesity. Therefore, the optimal diet for bone health is one that has a balance of nutrients and maintains a healthy BMI.

Calcium. Consistent with the evidence cited above that calcium intake has little impact on bone health, introduction of calcium-rich foods or use of calcium supplements as a sole intervention has not been shown to influence fracture rates. Although the introduction of a calcium supplement does reduce bone turnover and lead to a BMD benefit of approximately 1% over the first year of use,[42] this BMD gain does not increase with longer-term use, accounting for the lack of effect of calcium supplements on fracture.[20] The United States Preventive Services Taskforce recommends against calcium or vitamin D supplementation for fracture prevention in noninstitutionalized postmenopausal women.[43]

Falls Prevention. The majority of fractures in older individuals occurs following falls—calculated fracture risk is doubled in those with falls in the previous 12 months. Therefore, fall prevention is an obvious way to reduce fracture numbers, and management of fallers involves both assessment to detect underlying disease causing falls (e.g., parkinsonism) and generic- and disease-specific interventions to reduce fall risk. Balance and strengthening exercises reduce falls and also appear to prevent fractures.[44] Optimization of vision (such as cataract removal and counseling regarding use of multifocal lenses), reduction or elimination of sedative drugs, prevention/treatment of postural hypotension, and provision of a safe home environment (including the provision of handrails and removal of falls hazards) are effective as part of multicomponent intervention programs. Vitamin D has been advocated for fall prevention, but evidence for its efficacy is inconsistent,[22] possibly because many studies have been carried out in populations not meeting current criteria for vitamin D deficiency.

Antiresorptive Medications

Because increased bone resorption is the principal mechanism of bone loss following menopause and in inflammatory conditions, inhibition of osteoclast activity is an obvious strategy for bone preservation. These were the earliest interventions used, in the form of calcium supplements (which reduce bone turnover by ~20%), calcitonin, and estrogen therapy in postmenopausal women. From the 1980s, bisphosphonates have been used in osteoporosis and remain the most widely used agents, being joined more recently by denosumab (Table 60.3). Calcitonin is now little used because of lack of consistent evidence for fracture prevention. A new class of antiresorptive drugs, the cathepsin K inhibitors, have excellent antifracture efficacy, but concerns regarding cardiovascular safety have caused odanacatib not to progress to the marketplace.

Bisphosphonates. The bisphosphonate nucleus consists of two phosphate groups linked through a central carbon atom. This P-C-P bond is stable in biological fluids, and there is effectively no metabolism of bisphosphonates in humans. The cluster of oxygen atoms in the bisphosphonate nucleus gives it a strong negative charge, which results in these molecules binding very tightly to the positively charged surface of bone. Thus, bisphosphonate in the circulation is rapidly taken up by bone, where it remains bound for years. When osteoclasts resorb bone coated by an aminobisphosphonate, they ingest the accompanying bisphosphonate, which then inhibits an enzyme, farnesyl pyrophosphate synthase, in the pathway leading to cholesterol synthesis (the mevalonate pathway). Blockade of this pathway results in cytoskeletal dysfunction and, ultimately, cell death. Non–nitrogen-containing bisphosphonates have other cellular targets, but these first-generation drugs are now little used in the treatment of osteoporosis. The specificity of the bisphosphonates for the skeleton arises from their avid uptake onto bone surfaces, the rapid excretion of the balance of the dose unaltered in the urine, and the fact that cells other than osteoclasts do not resorb bone.

Efficacy. Bisphosphonate treatment is accompanied by increases in BMD, as illustrated in Fig. 60.5. These tend to plateau after 3 to 4 years, and the extent of the increase varies between agents and between sites of BMD measurement. Because there is greater bone turnover in trabecular bone, sites rich in trabecular bone, such as the lumbar

TABLE 60.3 Principal Medications for Management of Osteoporosis

Medication	Dose	Indications	Dosing Instructions	Comments
Alendronate	70 mg/week PO 10 mg/day PO	Treatment of postmenopausal, male and glucocorticoid osteoporosis	Take fasting with a glass of water while sitting or standing. Remain upright and fasting for 30 minutes	
	35 mg/week PO 5 mg/day PO	Prevention of postmenopausal osteoporosis		
Risedronate	5 mg/day, 35 mg/week, 75 mg ×2 per month, 150 mg/month, all PO	Treatment or prevention of postmenopausal osteoporosis; treatment of male and glucocorticoid osteoporosis		
Ibandronate	150 mg/month PO	Treatment or prevention of postmenopausal osteoporosis	As above, but remain upright and fasting for 60 minutes	
	3 mg/3 months IV	Treatment of postmenopausal osteoporosis	Inject over 15–30 seconds	
Zoledronate or zoledronic acid	5 mg/year IV	Treatment of postmenopausal, male, and glucocorticoid osteoporosis, fracture prevention after hip fracture	IV infusion in 100 mL over at least 15 minutes	
	5 mg every 2 years	Prevention of postmenopausal osteoporosis		
Raloxifene	60 mg/day PO	Prevention and treatment of postmenopausal osteoporosis		
Denosumab	60 mg every 6 months SC	Treatment of postmenopausal, male, and glucocorticoid osteoporosis patients at high fracture risk		
Teriparatide	20 µg/day SC for 18–24 months	Treatment of postmenopausal, male, and glucocorticoid osteoporosis patients at high fracture risk	Follow with antiresorptive agent	FDA recommends considering >2 years of cumulative use during a patient's lifetime only if fracture risk remains high
Abaloparatide	80 µg/day SC for 18–24 months	Treatment of postmenopausal osteoporosis patients at high fracture risk	Inject into the periumbilical region. Administer initial doses where the patient can lie down if orthostatic hypotension occurs. Follow with antiresorptive agent	See above. Declined registration in Europe
Romosozumab	210 mg/month SC for 12 months	Treatment of postmenopausal osteoporosis at high fracture risk or patients who have failed or are intolerant to other available osteoporosis therapy	Follow with antiresorptive agent	

IV, Intravenously; *PO*, by mouth; *SC*, subcutaneously; *FDA*, US Food and Drug Administration.

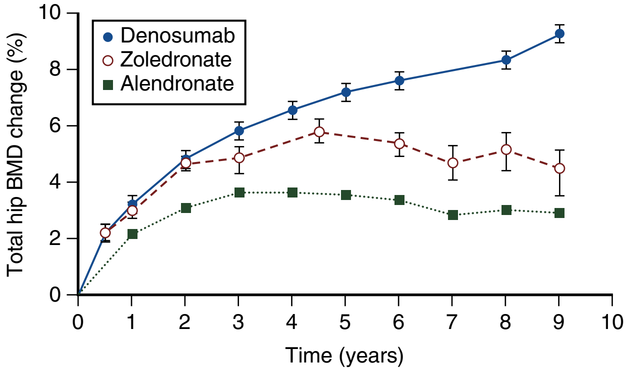


Fig. 60.5 **Long-term effects of osteoporosis treatments on total hip bone mineral density (*BMD*).** Data derived from the FLEX trial (alendronate, 5–10 mg per day), the HORIZON trial (zoledronate, 5 mg per year), and the FREEDOM trial (denosumab, 60 mg subcutaneous injection every 6 months). As data are derived from separate studies, formal comparisons between changes in BMD have not been made. (Modified from Reid IR. Short-term and long-term effects of osteoporosis therapies. *Nature Rev Endocrinol.* 2015;11:418–428; used with permission.)

spine, show greater BMD responses than more cortical sites, such as the mid-forearm and femoral neck. Despite differences between agents in antiresorptive potency, most of the aminobisphosphonates produce similar reductions in fractures, as demonstrated in Fig. 60.6. Although these drugs are very effective in reducing vertebral fractures (by as much as 70% with zoledronate), the reduction in total fracture numbers is a more modest 20% to 30%.

The prolonged residence time of bisphosphonates in bone raises the question of for how long these drugs need to be administered. Potential safety concerns with long-term continuous administration make this an important issue. Two studies have attempted to address the question of optimal duration of therapy for fracture prevention. In the FLEX study, women already on alendronate for 5 years were randomized to continue at their existing dose (10 mg/day), to continue at half that dose, or to cross over to placebo. Both groups continuing on alendronate maintained hip bone density, whereas those randomized to placebo tended to return to baseline BMD levels over the 5-year study. Rates of nonvertebral fractures were similar in all groups, but *post hoc* analyses indicated that, in those with a femoral neck BMD T-score less than −2.5 at the time they entered FLEX, continuation of

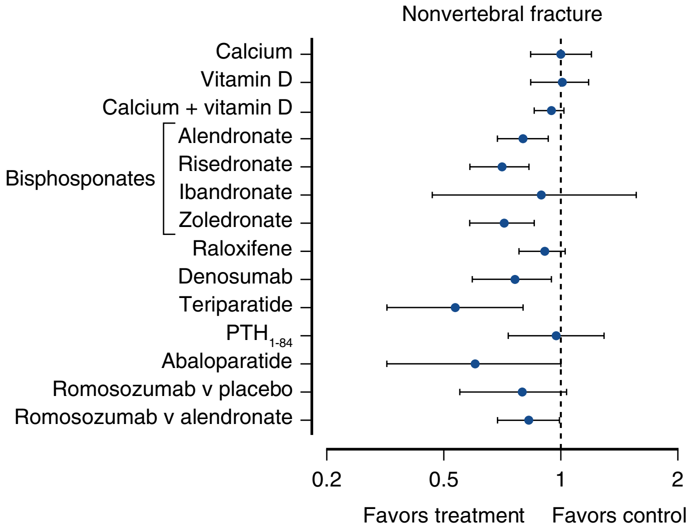


Fig. 60.6 Efficacy of treatments for the prevention of nonvertebral fractures. Abaloparatide data (hazard ratios) are at 18 months from Miller et al.[71] Romosozumab versus placebo data (hazard ratios) are at 12 months from Cosman et al.,[77] and romosozumab versus alendronate (hazard ratios) are at 24 months from Saag et al.[78] Parathyroid hormone (1–84) (*PTH1–84*) results are relative risks. Other results are odds ratios from the network metaanalysis of Murad et al.[82] Results are shown ± 95% confidence interval. (Modified from Reid IR. Short-term and long-term effects of osteoporosis therapies. *Nature Rev Endocrinol.* 2015;11:418–428; used with permission.)

alendronate reduced nonvertebral fractures by 50%.[45] Clinical vertebral fractures were also reduced by approximately one half in the alendronate groups compared with placebo. A similar study (to 9 years) was carried out in the phase III trial program for zoledronate, although it had only limited power to assess effects on fractures.[46] There was a partial loss of vertebral fracture efficacy if zoledronate was discontinued at year 3 (versus year 6), particularly in women with femoral neck T-score less than –2.5 at year 3. Bone loss in those discontinuing treatment at 6 years was less than 1% in the following 3 years, and there was no evidence of additional benefit with 9 years of therapy.

Common Side Effects. The most common side effect of oral bisphosphonates is upper *gastrointestinal irritation,* seen in 20% to 30% of users. Accordingly, patients are advised to remain upright for 30 minutes after ingesting bisphosphonates to minimize the risk of esophageal reflux. Oral bisphosphonates are contraindicated in those with active upper gastrointestinal disease. With intravenous aminobisphosphonates, the most common side effect is the *acute phase response*, comprised of fever, musculoskeletal pain, and malaise, frequently referred to as a “flu-like illness.” It can result in inflammation in almost any tissue, including the gastrointestinal system, the upper respiratory tract, and the eye.[47] With zoledronate, symptoms occur in 30% to 40% of patients in the week after their first dose and fade over a few days. Symptoms are only severe in 2% to 3% of those treated and decrease in frequency and severity after subsequent infusions. They are less frequent in those who have had oral aminobisphosphonates previously. Symptom frequency and severity are reduced by approximately half with either paracetamol/acetaminophen or a nonsteroidal antiinflammatory drug. Eye inflammation, typically uveitis, occurs in 1% of bisphosphonate-naïve subjects and is rapidly responsive to topical corticosteroids. It does not preclude further treatment with intravenous bisphosphonates, because, like other aspevects of the acute phase response, it does not necessarily recur with subsequent dosing.

Potent antiresorptive drugs sometimes cause *hypocalcemia* if administered to those with marked vitamin D deficiency (e.g., 25-hydroxyvitamin D <25 nmol/L). Vitamin D supplements should be provided to those judged clinically to be at risk of deficiency. Zoledronate is *nephrotoxic* in those with preexisting renal dysfunction and should not be administered when estimated glomerular filtration rate is less than 30 to 35 mL/minute. The use of a 1-mg or 2.5-mg dose of zoledronate, together with longer infusion times, could theoretically decrease renal toxicity in patients with marginal renal function.

In addition to these well-recognized side-effects, other safety concerns have arisen since bisphosphonates came into use. An increase in cases of atrial fibrillation was found in one phase III zoledronate study, but not in other major studies with this drug, nor with the other bisphosphonates. A small case series of carcinoma of the esophagus in oral bisphosphonate users was published, but other studies did not confirm this as a problem. The principal concerns that remain are osteonecrosis of the jaw (ONJ) and atypical femoral fractures.

Osteonecrosis of the Jaw. ONJ is defined as the presence of exposed bone in the mouth that fails to heal after appropriate therapy over a period of 6 to 8 weeks.[48] It usually develops following an invasive dental procedure, such as a tooth extraction, and the pathology is characterized by dead bone with overlying microbial biofilms that are very resistant to antimicrobial therapy. It was first described in the early 2000s in cancer patients receiving monthly infusions of pamidronate or zoledronate but is also clearly an issue in cancer patients receiving monthly denosumab injections, indicating that the development of ONJ is the result of potent suppression of bone resorption and is not specific to bisphosphonates. Whether ONJ is increased to a clinically significant extent in patients using antiresorptive drugs for osteoporosis treatment remains contentious. No cases were reported from the risedronate and alendronate development programs, but there was no awareness of ONJ as a clinical entity at that time. Postmarketing reports have suggested that the incidence of ONJ in patients receiving these drugs for osteoporosis is less than 2/100,000 patient-years. During the zoledronate phase III program, there was prospective adjudication of all potential cases, finding one case in each of the zoledronate and placebo groups (12,000 patient-years per group) and one during 1914 patients-years of observation in the study extensions. Studies of large databases have followed.[49] Although they clearly show that ONJ occurs in osteoporosis patients treated with antiresorptive drugs, they also demonstrate similar lesions at a similar frequency in osteoporosis patients not receiving these therapies. Case series in osteoporotic patients indicate that these ONJ-like lesions are lower grade than those seen in cancer patients, and that they are more likely to heal. Thus, the reality of this problem remains open to debate, although concern regarding it in the minds of the public and physicians remains a significant barrier to the more widespread use of these medications. It is also a major concern for dentists undertaking invasive procedures, so it is best not to start patients on bisphosphonates if such intervention is planned. Similarly, some dentists may wish to withhold bisphosphonates before tooth extractions. The long half-life of the bisphosphonates means that this is unlikely to change bone turnover at the extraction site, but, equally, the patient is unlikely to come to significant harm from the bone point of view if the medication is withdrawn for several months. Therefore, ascertaining a patient's current dental health status is advisable when considering initiating bisphosphonate therapy.[48]

Atypical Femoral Fractures. In 2008, reports appeared from Singapore and then the United States of the occurrence of transverse fractures of the upper femoral shaft in patients treated with bisphosphonates. Although femoral shaft fractures have always represented a few percent of fractures in osteoporotic subjects, and their frequency has not changed since the introduction of bisphosphonates to clinical practice, fractures with a specific radiologic appearance do seem to be more common in patients receiving antiresorptive drugs. These atypical femoral fractures start as stress fractures of the lateral femoral cortex that may be associated with local cortical thickening and thigh pain, then can progress to a complete transverse fracture with minimal comminution. Fractures are bilateral in approximately one-quarter of cases and are not usually the result of major trauma. Such appearances are sometimes seen in those who have never received antiresorptive drugs, and they have also been reported in other bone diseases. Most cases have been associated with alendronate use, although this might just reflect the wide use of this drug. None were reported in a recent 6-year study of 1000 women greater than 65 years receiving zoledronate,[50] and none in the extensions to the zoledronate phase III trial. A recent large Californian study suggests that incidence in the first year of bisphosphonate use is 0.2 per 10,000 patient-years, rising to 2.5 in years 3 to 5 and to 13 beyond 8 years of treatment. Importantly, there is consistent evidence that the risk drops rapidly after drug discontinuation, by more than half in the first year and by greater than 80% at 3 years.[51] It is important to remain mindful of the comparative rarity of these fractures. For instance, in a recent study of osteopenic postmenopausal women over the age of 65 years,[50] 20% of those on placebo had fragility fractures during 6 years' follow-up, a fracture rate of 333 per 10,000 patient-years, two to three orders of magnitude greater than the incidence of atypical fractures in California. Because zoledronate reduced fracture rates by one-third, the potential benefits from bisphosphonate use in appropriate patients clearly dwarf the problems associated with atypical femoral fractures. However, patients who develop thigh pain while taking bisphosphonates (or denosumab) should have appropriate radiology to determine whether there is evidence of a stress fracture. If there is, then bisphosphonate withdrawal and management in consultation with an orthopedic surgeon are appropriate. Prevention of these fractures is now the principal reason for drug holidays when using bisphosphonates, as discussed below.

Denosumab.

Denosumab is a monoclonal antibody directed against the protein RANKL, a key regulator of osteoclast development that is produced mainly in osteocytes. Denosumab is given by subcutaneous injection and produces near total arrest of bone resorption within days of administration. Bone formation indices decline by more than 60% over the 3 months following drug initiation, and bone histomorphometry confirms profound reductions in bone turnover (both formation and resorption), with a median zero bone formation rate after 2 to 3 years of use. The duration of effect is dose-dependent,[52] and in the conventionally used osteoporosis dose (60 mg) there is a marked offset of suppression of bone resorption beyond 6 months postinjection, with some variation in timing between individuals. Thus, if a denosumab injection is missed, bone resorption rises rapidly over 1 to 3 months after the injection was due, reaching levels more than 50% above baseline at 6 months. This is accompanied by rapid bone loss, with BMD returning to baseline within a year of drug discontinuation. Thus, the pharmacology of denosumab is profoundly different from that of the bisphosphonates, and this must be borne in mind when using these drugs clinically.

Efficacy. As expected for such a potent antiresorptive agent, denosumab produces substantial increases in BMD, which continue over 10 years of use (Fig. 60.5). Its effects on fracture are shown in Fig. 60.6, where the profile at the end of the three-year phase III FREEDOM study was very similar to that found in the phase III trial of zoledronate.[53] Ongoing follow-up of the active group from FREEDOM demonstrated that reductions in both vertebral and nonvertebral fractures were maintained to 10 years.[54]

Safety. The FREEDOM study found an excess of *skin infections* in those randomized to denosumab, although this has not been confirmed in the FREEDOM extension study, nor with extensive clinical use of the drug. FREEDOM also suggested that mortality was reduced by approximately 20% ($P = .08$), although the mechanisms of this are unclear. There were no *atypical femoral fractures* nor cases of ONJ during the core FREEDOM study. During the 7-year extension, there were two atypical femoral fractures (0.8 per 10,000 participant-years) and 13 cases of ONJ (5.2 per 10,000 participant-years), with all 11 patients with follow-up showing resolution. These rates are similar to those with other antiresorptive drugs, such as zoledronate. Potent antiresorptive drugs cause *hypocalcemia* in patients with marked vitamin D deficiency. Hypocalcemia has also been reported with denosumab use in those with renal impairment. Data from the Amgen postmarketing database covering 1,252,566 patient-years of use found four atypical femoral fractures, 32 cases of ONJ (with resolution in one-third), eight cases of severe symptomatic hypocalcemia (seven of whom had chronic kidney disease), and five cases of anaphylaxis.[55] These findings indicate that denosumab is a well-tolerated pharmaceutical and do not point to detectable adverse consequences from the sustained profound inhibition of bone turnover that it produces.

At present, the occurrence of *rebound fractures* following denosumab discontinuation is in the principal safety concern with this drug, particularly because long-term compliance with all osteoporosis drugs is poor. Most evidence for this problem comes from case reports and series. However, the FREEDOM extension study demonstrates that 15% of women discontinuing denosumab experienced a vertebral fracture subsequently, and in two-thirds of these (i.e., 10% of discontinuing patients) fractures were multiple. As many as nine fractures in a single patient have been reported, suggesting the possibility of substantial long-term disability from this problem.

The approach to rebound fractures is multifaceted. Patients need to be made aware of this problem before starting denosumab. It should not be initiated in those who are not committed to its long-term use or who are unlikely to be able to have follow-up injections in a timely manner. At the time of initiation, the prescribing physician should have an exit strategy in mind and have identified a medication to which the patient could transition following drug discontinuation.

In most instances, this will be a bisphosphonate, although which bisphosphonate is best and whether standard doses are adequate remain uncertain. Patients using denosumab because bisphosphonates are contraindicated represent a particular problem because it is not clear to what they should transition. Transitioning to teriparatide results in significant bone loss so is not usually appropriate.

Estrogen and Selective Estrogen Receptor Modulators. The key role of reduced ovarian estrogen production in postmenopausal bone loss is discussed above, so it is not surprising that administration of estrogen maintains BMD and reduces fracture risk.[6] In the Women's Health Initiative, hip and spine fractures were each reduced by 34% and total fractures by 24%.[56] Because estrogen receptors are present in many tissues, estrogen effects are widespread. This was demonstrated in the Women's Health Initiative, where beneficial effects on fractures and colon cancer were counterbalanced by changes in cardiovascular, cerebrovascular, and venous thromboembolic events.[56] The effects of estrogen alone appear to be different from those of combined treatment with a progestin, and may vary with specific agents within each class. This complexity has reduced enthusiasm for estrogen in management of osteoporosis, although its efficacy is firmly established. However, in the small proportion of women who have an elevated fracture risk in the decade following the menopause, the balance of benefit versus risk with estrogen appears to be positive.[57]

The 1990s saw the emergence of a new class of drugs called selective estrogen receptor modulators (SERMs), which bind to the estrogen receptor, producing tissue-specific agonist or antagonist effects. Thus, it is theoretically possible to have an agent that will mimic the beneficial effects of estrogen (such as maintenance of BMD) yet avoid its adverse effects, such as the promotion of breast cancer. Raloxifene is a SERM that has been widely used in osteoporosis, although its effects on BMD are significantly less than those of estrogen,[58] and it reduces only vertebral, not nonvertebral, fractures.[59] It reduces breast cancer risk, but, like estrogen, it increases the risk of thromboembolic disease. Another SERM, bazedoxifene, has been marketed in combination with conjugated estrogens for management of menopausal symptoms. Although it increases BMD, its antifracture efficacy has not been established. Because the available SERMs lack evidence of efficacy against nonvertebral fractures, they have only a minor role in osteoporosis management.

Tibolone is a synthetic steroid with weak estrogenic, progestogenic, and androgenic activity, and it prevents vertebral and nonvertebral fractures.[60] There is inconsistent evidence of its effects on breast cancer, and it increases stroke risk. It is used in Europe for osteoporosis treatment and management of menopausal symptoms, but not in North America.

Anabolic Medications

Anabolic medications stimulate bone formation. This action offers the possibility of increasing bone mass and, possibly, restoring microarchitecture, whereas antiresorptive agents essentially preserve bone structures that are already present. Anabolic treatments may be particularly appropriate in those with a high fracture risk. Modified androgens (anabolic steroids) have been used in the past in this role, but their androgenic side effects have limited their use, and antifracture efficacy has not been established. Until recently, analogs of parathyroid hormone (PTH) were the only such agents available, acting predominantly via PTH type 1 receptors on osteoblasts and osteocytes. The field has broadened recently with the advent of abaloparatide, an analog of PTH-related peptide, and with the development of antisclerostin antibodies (Table 60.3).

Teriparatide. Until the 1980s, PTH was regarded as having a negative effect on bone, being the principal stimulator of bone resorption. Thus, in patients with primary hyperparathyroidism, management of osteoporosis was a reason for parathyroidectomy. However, it is now clear that PTH stimulates both bone formation and bone resorption, the balance between these being determined by the dosing regimen. Sustained high levels of PTH, as in primary hyperparathyroidism, result in greater stimulation of resorption, particularly in cortical bone, leading to bone loss. In contrast, daily pulses appear to be anabolic, particularly in axial trabecular bone. Intact PTH has 84 amino acids, although most of its biological activity appears to reside in its N-terminus. A variety of PTH analogs have been assessed in osteoporosis, but PTH(1–34), known as teriparatide, is the principal one in current use.

Efficacy. Teriparatide transiently increases the activation frequency of bone remodeling, followed by progressive increases in bone density over the 18 to 36 months, particularly in the spine.[61] In the first year of therapy, PTH has little beneficial effect on femoral neck BMD, although beyond 12 months more positive changes are observed.[61] In the radius, the conventional dose of teriparatide (20 μg) has no beneficial effect, and 40 μg/day accelerates bone loss. When compared directly with alendronate, teriparatide produces greater increases in spine BMD but is less effective than alendronate in the proximal femur during the first year of use.[61] However, teriparatide reduced numbers of vertebral fractures by two-thirds and nonvertebral fractures by one-third over a median treatment period of 21 months in the phase III trial.[62] Only nine hip fractures occurred in this trial, so this aspect of efficacy was not addressed, although a subsequent metaanalysis suggested that teriparatide does prevent hip fractures.[63] The VERO trial directly compared fracture prevention with teriparatide (20 μg/day) and risedronate (35 mg/week) in postmenopausal women with existing vertebral fractures.[64] Over 2 years, new vertebral fracture incidence was reduced by more than half and new nonvertebral fractures by approximately one-third in the teriparatide group compared with risedronate, although the latter difference was not statistically significant.

A Japanese study has assessed the efficacy of teriparatide given weekly (dose 56.5 μg) over 72 weeks. Vertebral fractures were reduced by 80%, but the relative risk of nonvertebral fracture was 0.98 (0.47, 2.07).[65] The efficacy of this regimen has not been assessed elsewhere. Teriparatide has BMD effects in men similar to those in women, although fracture prevention has not been rigorously studied. In glucocorticoid-induced osteoporosis, teriparatide produces larger BMD increases than alendronate and substantially reduces vertebral fracture risk.[66]

Because the actions of teriparatide on bone formation potentially complement those of antiresorptive agents, there has been interest in studying combination therapy. An early study combining PTH(1–84) with alendronate did not suggest much advantage from this approach, but subsequent trials of teriparatide plus zoledronate[67] and of teriparatide plus denosumab[68] have both suggested additivity of BMD effects, although fractures were not endpoints of these studies. However, clinical fractures (reported as adverse events) were less frequent in the combination therapy group compared with zoledronate alone ($P = .04$).

Treatment sequences have also been evaluated. Cosman et al. randomized osteoporotic women on alendronate to addition of or switching to teriparatide and found that continuation of the alendronate after starting teriparatide produced more positive changes in bone density than did a switch of therapy.[69] Transitioning from teriparatide to denosumab produces positive effects on bone density, but the reverse sequence is associated with initial bone loss, so is not recommended.[70] Because duration of therapy with teriparatide is restricted to 18 or 24 months in most jurisdictions, it is important to have a transition plan in mind when teriparatide is initiated, as its BMD effects wear off over 1 to 2 years after drug cessation. Transition to either a bisphosphonate or denosumab is the usual practice.

Another PTH analog, PTH(1–84), was assessed in a phase III trial. There were high dropout rates, and an equivocal reduction in vertebral fractures was found, but nonvertebral fracture rates were not changed. Thus, PTH(1–84) is now little used for osteoporosis treatment.

Safety. In the phase III study, the 20 μg/day dose was well tolerated, but the 40 μg/day dose was associated with nausea and headache.[62] Serum calcium levels 4 to 6 hours postinjection were above the normal range in 2% of the placebo group and in 11% of the 20 μg group, and dose reduction for this reason was required in 3% of this group. The frequency of hypercalciuria was not increased by teriparatide. In the study of weekly teriparatide injections, nausea, headache, and abdominal pain led to 19% of the teriparatide group stopping treatment, compared with 7% in the placebo group.[65] These studies did not find an increase in the frequency of arthralgia or other musculoskeletal symptoms in the active treatment groups, although this has been anecdotally reported as a problem in clinical practice.

The phase III teriparatide trial was stopped because rat studies found a high incidence of osteosarcoma after high doses of teriparatide over several years. This led to restrictions in the permissible duration of administration of PTH peptides to 18 to 24 months. However, a decade of clinical surveillance has not produced evidence that this is a problem in humans. This suggests that second courses of treatment with teriparatide may be safe, and there is evidence of BMD increases with retreatment. Teriparatide is contraindicated in those at increased risk of osteosarcoma (e.g., previous skeletal radiation and, possibly, Paget disease), in hypercalcemia, in metabolic bone disease other than osteoporosis, and in patients with skeletal malignancies or bone metastases.

Abaloparatide. Abaloparatide is an analog of human PTH-related peptide(1-34). Like teriparatide, it acts through the PTH1 receptor, stimulating bone turnover. Some studies suggest that it has a preferential effect on bone formation over resorption, but findings are inconsistent. Abaloparatide is available in the United States but has been declined approval in Europe because the decrease in nonvertebral fractures is of marginal significance, because data from two study centers were judged to be unreliable, and because of an increase in palpitations in the abaloparatide group.

Efficacy. In the phase 3 trial, 2463 osteoporotic postmenopausal women were randomized to 18 months' treatment with placebo, abaloparatide 80 μg/day, or to open-label teriparatide 20 μg/day.[71] Teriparatide tended to increase bone turnover markers (both formation and resorption) more than abaloparatide, and increases in BMD were greater with abaloparatide. Vertebral fracture risk compared with placebo was reduced by more than 80% by both interventions, but nonvertebral fracture was only significantly reduced in the abaloparatide group. Hypercalcemia was less common with abaloparatide (3.4%) than with teriparatide (6.4%). At 19 months, participants from both the abaloparatide and the placebo groups were switched to active treatment with alendronate and continued to 43 months.[72] Vertebral fracture rates remained suppressed by more than 80% during this extension. For nonvertebral fractures there was a 39% reduction in the abaloparatide/alendronate group compared with the placebo/alendronate group over the entire 43 months ($P = .04$).

Safety. Abaloparatide was generally well tolerated in the phase III trial, with complaints of nausea and headache, as with teriparatide. Palpitations were reported by 5% of abaloparatide participants, compared with 0.4% in the placebo group and 1.6% in those receiving teriparatide. Hypercalcemia occurred less frequently than in the teriparatide group (6.4% vs. 3.4%). As with teriparatide, rat toxicology studies demonstrated an increased risk of osteosarcoma. Therefore, the contraindications to use and the limitations on treatment duration that apply to teriparatide also apply to abaloparatide.

Antisclerostin Antibodies. Sclerostin is produced in osteocytes and is regulated by bone loading and by hormonal influences such as PTH. It is released at the bone surface via the osteocytic cell processes, where it acts as a potent inhibitor of Wnt signaling in osteoblasts, dramatically reducing bone formation. The Wnt pathway also stimulates secretion of osteoprotegerin, an osteoblast protein that reduces osteoclast formation and, thus, bone resorption. Thus, sclerostin is an attractive target for monoclonal antibodies. This potential has been fulfilled with the development of romosozumab, which acts as a potent stimulator of bone formation but also reduces bone resorption, resulting in accelerated fracture healing in rats.[73]

Efficacy. Single doses of romosozumab in humans have anabolic and antiresorptive effects on biochemical markers that persist for 1 to 2 months. The phase II study in postmenopausal women with low bone density showed dose-related effects on BMD, with increases in spine and hip BMD of up to 11.3% and 4.1%, respectively, at 1 year, substantially greater than was seen with teriparatide or alendronate in the same study.[74] BMD increases continued in the second study year to 15% and 7%, respectively. Crossover to placebo at 24 months resulted in loss over the subsequent year of more than half of the BMD gained, a deficit that was reversed with a further year of romosozumab in study year 4. Other participants were crossed over to denosumab in year 3 and showed continuing gradual BMD gain, which continued in year 4 with crossing back to romosozumab. Romosozumab was without effect on BMD at the distal third of the radius. The BMD changes were preceded by a 100% increase in PINP (a marker of bone formation) and a 30% decline in the bone resorption marker, C-telopeptide, in the first study month. Subsequently, these markers were close to baseline levels during the period when BMD was increasing most. A subsequent BMD study comparing romosozumab with teriparatide also demonstrated superiority of romosozumab.[75]

Bone histomorphometry in older women treated with romosozumab showed that trabecular and endosteal bone forming surfaces were increased 3- to 4-fold at 2 months, but that this was no longer evident at 12 months. Osteoclast numbers on these surfaces were reduced to near zero at both timepoints. As a result, trabecular bone volume was increased by approximately 50% at 12 months, mainly as a result of increased trabecular thickness. Cortical bone was little affected by romosozumab.[76]

There have been two phase III trials of romosozumab in postmenopausal women. The FRAME study enrolled 7180 postmenopausal women who had BMD T-scores of –2.5 to –3.5 at the total hip or femoral neck.[77] Patients were randomly assigned to receive romosozumab (210 mg) or placebo monthly for 12 months. Thereafter, patients in both groups received only denosumab (60 mg every 6 months) to month 36. At 12 months, romosozumab reduced the risk of new vertebral fractures by 73% ($P < .001$) and nonvertebral fractures by 25% ($P = .1$), with similar results at 24 months. Clinical fractures were reduced by 36% at 12 months and 27% at 3 years compared with the placebo/denosumab group.

The ARCH study enrolled 4093 postmenopausal women who had osteoporosis plus a fragility fracture and randomly assigned them to receive monthly romosozumab or weekly oral alendronate (70 mg) for

12 months, followed by open-label alendronate alone in both groups.[78] Over 24 months, new vertebral fractures were reduced by 48% ($P < .001$), nonvertebral fractures by 19% ($P = .04$), and hip fractures by 38% ($P = .02$) in the romosozumab-to-alendronate group compared with the placebo-to-alendronate group. Clinical fractures were significantly reduced by 28% at 1 year and by 27% at 3 years.

Safety. In the FRAME study, romosozumab was well tolerated, but one atypical femoral fracture and two cases of ONJ were observed in the romosozumab group. In ARCH, total adverse events and serious adverse events (including atypical femoral fractures and ONJ) were balanced between the two groups. However, during year 1 serious cardiovascular adverse events were observed more often with romosozumab than with alendronate (2.5% vs. 1.9%), an imbalance not seen in FRAME. The reason for the different cardiovascular profiles in these two studies is unclear, but it is possible that alendronate had a cardioprotective effect in ARCH, although this has not been clearly demonstrated in the many other trials using alendronate. As a result, in some jurisdictions romosozumab carries a warning regarding cardiac safety, and its use is discouraged in those with recent cardiovascular events.

KEY POINTS

- Correction of reversible risk factors is the first line of intervention. Antiresorptive drugs are the most commonly used drug treatments, with 1- to 2-year courses of anabolic agents used as adjunctive therapy in those at highest fracture risk.

TREATMENT DECISIONS

Whom to Treat

In the introduction to this chapter, reference was made to the progressive evolution of the definition of osteoporosis. A similar path has been followed in determining whom should receive treatment, and there is not yet consensus on this question. When osteoporosis was diagnosed from the presence of a low-trauma fracture, then this was also the indication for treatment. A similar course was followed when the diagnosis moved to a BMD T-score of less than –2.5. The advent of fracture risk calculators permitted treatment decisions to be based on pharmacoeconomic analyses, resulting in a 10-year hip fracture risk of 3% or a major osteoporotic fracture risk of 15% to 20% being proposed as thresholds. These determinations were made at a time when conventional therapies cost several hundred dollars annually. We now have a wide range of costs, with generic bisphosphonates available for only a few dollars annually in some countries, but injectable anabolic therapies costing tens of thousands of dollars in others. This means that the former are cost-effective in women with a hip fracture risk of only 1% over 10 years, but the latter are only cost-effective in those at very high fracture risk.

A recent analysis of 55,000 women aged over 40 years who were followed for more than 5 years in the Manitoba BMD registry compared the sensitivities and specificities of commonly used thresholds. Among women in their 60s at baseline, 3% to 30% met the criteria for intervention, but these criteria only identified the 6% to 44% of those who fractured during follow-up. Among women in their 70s, intervention thresholds were met in 15% to 56%, identifying 23% to 71% of those who subsequently fractured.[31] The more inclusive National Osteoporosis Foundation criteria identified a greater proportion of women as being at risk and showed greater sensitivity (although lower specificity) in identifying those who subsequently had fractures. These data highlight the shortcomings of the calculators currently available and indicate that a broad use of antifracture medications is necessary if a substantial reduction in total fracture numbers is to be achieved. The numbers needed to treat in the Manitoba study suggest that broader treatment strategies are likely to be cost-effective for inexpensive medications such as bisphosphonates. Similar conclusions come from the recent trials of screening strategies for osteoporosis.[79] Screening resulted in intervention in only 6% to 18% of participants, with no change in total fracture numbers resulting, although some fracture subtypes were reduced.

In conclusion, whom to treat remains uncertain. Of the various strategies proposed, the multifaceted proposals of the National Osteoporosis Foundation appear to be best in terms of their sensitivity for future fracture. However, a UK cost-effectiveness analysis (National Institute for Health and Care Excellence) suggests that an even more inclusive strategy of treating women with a 10-year hip fracture risk of more than 1% is justified for the use of generic bisphosphonates. In such low-risk groups, the possibility of side effects becomes a key factor in determining whether to recommend intervention. Patients' enthusiasm for intervention will vary widely and is not necessarily related to absolute fracture risk or to arbitrary treatment thresholds.

Selection of Treatment

An initial step in providing osteoporosis treatment is a consideration of the likelihood of vitamin D deficiency in that patient, and correction of that as required. If pharmaceutical intervention is justified, the first line will usually be a *bisphosphonate*, on the basis of their efficacy, safety, and low cost. The choice between oral and intravenous bisphosphonates is a pragmatic one based substantially on patient preference. Intravenous bisphosphonates require infusion facilities and carry the risk of an acute phase response. They are more expensive in some countries. Intravenous bisphosphonates should not be administered to patients with glomerular filtration rates less than 30 to 35 mL/min. However, their major advantage is the certainty that the active drug has been delivered to the bone surface, plus the infrequency of dosing, particularly with zoledronate, where dosing as little as every 5 years can maintain bone density in postmenopausal women.[80] Oral bisphosphonates carry the risk of upper gastrointestinal symptoms and require adherence to the correct dosing requirements (taken fasting with a glass of water, at least 30 minutes before other fluids or food, with the patient remaining upright during this period).

Denosumab is also used as a first-line antiresorptive therapy, although it is more expensive than most bisphosphonates. It has the convenience of subcutaneous administration and is without any postdose side effects. The principal caution in its use is the rebound increase in bone resorption that occurs if the 6-monthly injection is delayed or the drug is discontinued without transition to another agent. The occurrence of clusters of vertebral fractures in as many as 10% of discontinuing patients requires that an exit strategy is carefully considered, and discussed with the patient, *before* denosumab is initiated. Denosumab can be used in those with marked renal impairment, although hypocalcemia may be a problem in this context, requiring use of calcium supplements.

The few head-to-head studies suggest that anabolic agents have superior antifracture efficacy to bisphosphonates. Therefore, if available and affordable, a *PTH analog* or *romosozumab* could be first-line treatment options for a patient with a high fracture risk. Cost and the requirement for daily or monthly subcutaneous injections are the main downsides of this approach. Teriparatide has nil or negative effects on proximal femoral BMD in the first year of treatment, raising questions over its use as a first-line monotherapy in those at high risk of femoral neck or other cortical fractures.

Combination Therapy. Some of the largest increases in BMD have resulted from combination therapy with a PTH analog plus an antiresorptive. Whether the large BMD benefit translates into greater antifracture efficacy is unknown, but it is a strategy for consideration in the individual with markedly reduced BMD. Sequential therapy is a further way of deploying multiple agents to manage patients with severe osteoporosis. A variety of scenarios are possible. A common one is the patient who has further fractures while on treatment with an antiresorptive. Starting an anabolic agent in this situation is appealing, and bone density changes are more positive if the anabolic is added to the bisphosphonate rather than substituted for it. Switching from denosumab to teriparatide produces marked adverse effects on BMD, so is contraindicated.

A second scenario is that of a patient coming to the end of a course of an anabolic agent. Both PTH analogs and romosozumab show substantial offset of effects following discontinuation. Therefore, transition to another antiosteoporosis medication is appropriate, and this usually is an antiresorptive. Oral or intravenous bisphosphonates or denosumab have been shown to perform satisfactorily in this context. Raloxifene has been used following teriparatide and is superior to placebo in its BMD effects.

In summary, therapy usually starts with an oral or intravenous bisphosphonate, with denosumab being substituted if there are specific indications to do so. In severe disease, use of an anabolic as a first- or second-line therapy may be appropriate but will require transition to an antiresorptive at the end of the 1- or 2-year course.

How Long to Treat

Low BMD or high fracture risk are not self-limiting conditions, so lifelong management is likely to be needed. Management, however, may include periods off drug, made possible by the long duration of action of bisphosphonates, which continue to exert antiresorptive effects for months to years after their discontinuation.

Concern regarding atypical femoral fractures (and to a lesser extent ONJ), together with the long duration of effect of bisphosphonates, has led to the common practice of punctuating bisphosphonate treatment with *drug holidays*. The risk of atypical femoral fractures is 0.2 to 2 per 10,000 patient-years in the early years of bisphosphonate use but climbs steeply after 5 years of continuous treatment. Observational data suggest that risk is reduced by two-thirds in the months following bisphosphonate cessation and by almost 90% with a drug holiday of more than 15 months. Thus, it is common to withdraw alendronate for a period of 1 to 2 years after approximately 5 years of continuous therapy. Risedronate has a more rapid offset, and drug holidays of 6 to 12 months are recommended with that agent. Drug holidays are contraindicated with denosumab because of its rapid offset—denosumab requires transition to another antiresorptive drug, usually a bisphosphonate.

Practice with the long-term use of zoledronate varies. In the phase III clinical trial program, six annual infusions prevented more vertebral fractures than three annual infusions, the main benefit being in individuals whose femoral neck T-score was still less than –2.5 at the end of year 3. Extending annual infusions from 6 years to 9 years did not appear to confer benefit in terms of turnover markers, bone density, or fracture numbers, although the latter were small. Accordingly, many clinicians give three or six annual infusions of zoledronate, followed by a pause for several years. The present authors' practice is to initiate therapy at 18-month intervals for several years because this has been shown to produce comparable antifracture efficacy to annual dosing, and then to increase the interdose interval to 2 or 3 years subsequently if BMD response is satisfactory. In patients receiving this medication long-term, suppression of PINP (to 35–40 μg/L) and maintenance of BMD increases can be sustained over periods of more than a decade with injections every 3 to 5 years.[80]

Safety concerns (i.e., osteosarcoma) have limited the duration of use of PTH analogs to 18 to 24 months in most clinical trials. Transition to an antiresorptive agent at the end of this time is the conventional practice. The phase III trials of romosozumab only administered the drug for 12 months, and again transition to an antiresorptive is the recommended course.

Estrogen and SERMs have no specific time limits on their use, although the risk of breast cancer appears to increase linearly with duration of estrogen (but not SERM) exposure. These agents are not usually used as a primary therapy for osteoporosis.

Follow-Up

Practice varies with respect to the use of specific investigations during follow-up. In patients treated with oral bisphosphonates, it is wise to measure a bone turnover marker 3 to 6 months after initiation to ensure that the desired antiresorptive effect has been achieved. Serum PINP levels less than 35 to 40 μg/L are indicative of a satisfactory antiresorptive effect. Failure to do this could result in unsatisfactory dosing practices or low absorption of the drug not being detected until bone density is measured some years later. Many physicians remeasure BMD 3 to 5 years after treatment initiation. Measurement earlier than this is likely to be counterproductive in patients treated with antiresorptives because the expected changes are close to the precision limits of bone densitometers. Failure of BMD to increase during treatment raises the possibility of an underlying condition contributing to ongoing bone loss and suggests the need for appropriate investigations. In patients receiving oral agents, noncompliance and inadequate absorption of medication are the most common reasons for suboptimal BMD effects, so these problems should be detectable early in the course of treatment through measurement of a bone turnover marker.

KEY POINTS

- No single indication for treatment is universally accepted, although 3% 10-year hip fracture risk or 15%–20% major osteoporotic fracture risk are commonly used thresholds. With inexpensive generic bisphosphonates, treatment may be cost-effective at a 10-year hip fracture risk as low as 1%.

MANAGEMENT OF SUBTYPES OF OSTEOPOROSIS

Although the pathogenesis of *glucocorticoid-induced osteoporosis* is distinct from that of postmenopausal osteoporosis (see above), the principles of its management are similar. Fracture risk is evaluated based on demographic parameters, bone density, and current steroid dose. Further adjustment can be made for higher doses of steroids.[32] The range of medications is similar, as noted in Table 60.3. Studies in glucocorticoid osteoporosis have assessed BMD as the primary endpoint, and where fracture prevention has been observed, this is only for vertebral fracture.[81]

In premenopausal women and in men under 70 years with osteoporosis, it is important to make a thorough search for underlying causes of bone loss. As many as 50% of men will have some identifiable cause for their bone loss. Treatment efficacy data in men have the limitations noted above for glucocorticoid osteoporosis. There are very little trial data to address efficacy of treatments in premenopausal women, so management of these patients is not strictly evidence-based. However, the same sequence of treatment outlined for postmenopausal osteoporosis can be followed in these younger women, as long as underlying conditions have been sought and addressed, to the extent that this is possible.

REFERENCES

1. Peck WA, Burckhardt P, Christiansen C, et al. Consensus Development Conference - diagnosis, prophylaxis, and treatment of osteoporosis — rapid publication. *Am J Med*. 1993;94:646–650.
2. Kanis JA, Melton LJ, Christiansen C, et al. Perspective: the diagnosis of osteoporosis. *J Bone Miner Res*. 1994;9:1137–1141.
3. Reid IR. A broader strategy for osteoporosis interventions. *Nat Rev Endocrinol*. 2020;16:333–339.
4. Kanis JA, Johnell O, Oden A, et al. FRAX (TM) and the assessment of fracture probability in men and women from the UK. *Osteoporos Int*. 2008;19:385–397.
5. Nguyen ND, Frost SA, Center JR, et al. Development of a nomogram for individualizing hip fracture risk in men and women. *Osteoporos Int*. 2007;18:1109–1117.
6. Lindsay R. Prevention and treatment of osteoporosis with ovarian hormones. *Ann Chir Gynaecol*. 1988;77:219–223.
7. Khosla S. New insights into androgen and estrogen receptor regulation of the male skeleton. *J Bone Miner Res*. 2015;30:1134–1137.
8. Samelson EJ, Broe KE, Xu H, et al. Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the Bone Microarchitecture International Consortium (BoMIC): a prospective study. *Lancet Diabetes Endocrinol*. 2019;7:34–43.
9. Eastell R, O'Neill TW, Hofbauer LC, et al. Postmenopausal osteoporosis. *Nat Rev Dis Primers*. 2016;2:16069.
10. Van Staa TP, Laan RF, Barton IP, et al. Bone density threshold and other predictors of vertebral fracture in patients receiving oral glucocorticoid therapy. *Arthritis Rheum*. 2003;48:3224–3229.
11. Schwartz AV, Vittinghoff E, Bauer DC, et al. Association of BMD and FRAX score with risk of fracture in older adults with type 2 diabetes. *J Am Med Assoc*. 2011;305:2184–2192.
12. Allen MR, Iwata K, Phipps R, et al. Alterations in canine vertebral bone turnover, microdamage accumulation, and biomechanical properties following 1-year treatment with clinical treatment doses of risedronate or alendronate. *Bone*. 2006;39:872–879.
13. Follet H, Viguet-Carrin S, Burt-Pichat B, et al. Effects of preexisting microdamage, collagen cross-links, degree of mineralization, age, and architecture on compressive mechanical properties of elderly human vertebral trabecular bone. *J Orthop Res*. 2011;29:481–488.
14. Zebaze RMD, Ghasem-Zadeh A, Bohte A, et al. Intracortical remodelling and porosity in the distal radius and post-mortem femurs of women: a cross-sectional study. *Lancet*. 2010;375:1729–1736.
15. Chapurlat R, Bui M, Sornay-Rendu E, et al. Deterioration of cortical and trabecular microstructure identifies women with osteopenia or normal bone mineral density at imminent and long-term risk for fragility fracture: a prospective study. *J Bone Miner Res*. 2019;35:833–844.
16. Faulkner KG, Cummings SR, Black D, et al. Simple measurement of femoral geometry predicts hip fracture — the Study of Osteoporotic Fractures. *J Bone Miner Res*. 1993;8:1211–1217.
17. Johansson H, Kanis JA, Odén A, et al. A meta-analysis of the association of fracture risk and body mass index in women. *J Bone Miner Res*. 2014;29:223–233.
18. Anderson JJB, Roggenkamp KJ, Suchindran CM. Calcium intakes and femoral and lumbar bone density of elderly U.S. Men and Women: National Health and Nutrition Examination Survey 2005-2006 analysis. *J Clin Endocrinol Metab*. 2012;97:4531–4539.
19. Bristow SM, Horne AM, Gamble GD, et al. Dietary calcium intake and bone loss over 6 years in osteopenic postmenopausal women. *J Clin Endocrinol Metab*. 2019;104:3576–3584.
20. Bolland MJ, Leung W, Tai V, et al. Calcium intake and risk of fracture: systematic review. *BMJ*. 2015;351:h4580.
21. Reid IR, Horne AM, Mihov B, et al. Effect of monthly high-dose vitamin D on bone density in community-dwelling older adults substudy of a randomized controlled trial. *J Intern Med*. 2017;282:452–460.
22. Bolland MJ, Grey A, Avenell A. Effects of vitamin D supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis. *Lancet Diabetes Endocrinol*. 2018;6:847–858.
23. Macdonald HM, Reid IR, Gamble GD, et al. 25-Hydroxyvitamin D threshold for the effects of vitamin d supplements on bone density: secondary analysis of a randomized controlled trial. *J Bone Miner Res*. 2018;33:1464–1469.
24. Michaelsson K, Melhus H, Ferm H, et al. Genetic liability to fractures in the elderly. *Arch Intern Med*. 2005;165:1825–1830.
25. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet*. 2012;44:491–501.
26. Stone KL, Seeley DG, Lui LY, et al. BMD at multiple sites and risk of fracture of multiple types: long-term results from the Study of Osteoporotic Fractures. *J Bone Miner Res*. 2003;18:1947–1954.
27. Jones G, Nguyen T, Sambrook PN, et al. Symptomatic fracture incidence in elderly men and women: the Dubbo Osteoporosis Epidemiology Study (DOES). *Osteoporos Int*. 1994;4:277–282.
28. Cooper C, Cole ZA, Holroyd CR, et al. Secular trends in the incidence of hip and other osteoporotic fractures. *Osteoporos Int*. 2011;22:1277–1288.
29. Cummings SR. Osteoporosis: review of the evidence for prevention, diagnosis, and treatment and cost-effectiveness analysis - status report. *Osteoporos Int*. 1998;8:S1–S85.
30. Toth E, Banefelt J, Åkesson K, et al. History of previous fracture and imminent fracture risk in Swedish women aged 55 to 90 years presenting with a fragility fracture. *J Bone Miner Res*. 2020;35:861–868.
31. Crandall CJ, Schousboe JT, Morin SN, et al. Performance of FRAX and FRAX-based treatment thresholds in women aged 40 years and older: the Manitoba BMD registry. *J Bone Miner Res*. 2019;34:1419–1427.
32. Kanis JA, Johansson H, Oden A, et al. Guidance for the adjustment of FRAX according to the dose of glucocorticoids. *Osteoporos Int*. 2011;22:809–816.
33. Leslie WD, Lix LM, Johansson H, et al. Spine-hip discordance and fracture risk assessment: a physician-friendly FRAX enhancement. *Osteoporos Int*. 2011;22:839–847.
34. Leslie WD, Lix LM, Johansson H, et al. Does osteoporosis therapy invalidate FRAX for fracture prediction? *J Bone Miner Res*. 2012;27:1243–1251.
35. Cosman F, de Beur SJ, LeBoff MS, et al. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int*. 2014;25:2359–2381.
36. Compston J, Cooper A, Cooper C, et al. UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos*. 2017;12:43.
37. Shepherd JA, Schousboe JT, Broy SB, et al. Executive summary of the 2015 ISCD position development conference on advanced measures from DXA and QCT: fracture prediction beyond BMD. *J Clin Densitom*. 2015;18:274–286.
38. de Souto Barreto P, Rolland Y, Vellas B, et al. Association of long-term exercise training with risk of falls, fractures, hospitalizations, and mortality in older adults: a systematic review and meta-analysis. *JAMA Intern Med*. 2018;179:394–405.
39. Chapuy MC, Arlot ME, Duboeuf F, et al. Vitamin-D3 and calcium to prevent hip fractures in elderly women. *N Engl J Med*. 1992;327:1637–1642.
40. Smith LM, Gallagher JC, Suiter C. Medium doses of daily vitamin D decrease falls and higher doses of daily vitamin D3 increase falls: a randomized clinical trial. *J Steroid Biochem Mol Biol*. 2017;173:317–322.
41. Burt LA, Billington EO, Rose MS, et al. Effect of high-dose vitamin D supplementation on volumetric bone density and bone strength: a randomized clinical trial. *J Am Med Assoc*. 2019;322:736–745.
42. Tai V, Leung W, Grey A, et al. Calcium intake and bone mineral density: systematic review and meta-analysis. *BMJ*. 2015;351:h4183.
43. Grossman DC, Curry SJ, Owens DK, et al. Vitamin D, calcium, OR combined supplementation for the primary prevention of fractures in community-dwelling adults. US Preventive Services Task Force recommendation statement. *J Am Med Assoc*. 2018;319:1592–1599.
44. Cauley JA, Giangregorio L. Physical activity and skeletal health in adults. *Lancet Diabetes Endocrinol*. 2020;8:150–162.
45. Schwartz AV, Bauer DC, Cummings SR, et al. Efficacy of continued alendronate for fractures in women with and without prevalent vertebral fracture: the FLEX trial. *J Bone Miner Res*. 2010;25:976–982.
46. Black DM, Reid IR, Cauley JA, et al. The effect of 6 versus 9 years of zoledronic acid treatment in osteoporosis: a randomized second extension to the HORIZON-Pivotal Fracture Trial (PFT). *J Bone Miner Res*. 2015;30:934–944.

47. Reid IR, Gamble GD, Mesenbrink P, et al. Characterization of and risk factors for the acute-phase response after zoledronic acid. *J Clin Endocrinol Metab*. 2010;95:4380–4387.
48. Khan AA, Morrison A, Hanley DA, et al. Diagnosis and management of osteonecrosis of the jaw: a systematic review and international consensus. *J Bone Miner Res*. 2015;30:3–23.
49. Lin TC, Yang CY, Kao Yang YH, et al. Incidence and risk of osteonecrosis of the jaw among the Taiwan osteoporosis population. *Osteoporos Int*. 2014;25:1503–1511.
50. Reid IR, Horne AM, Mihov B, et al. Fracture prevention with zoledronate in older women with osteopenia. *N Engl J Med*. 2018;379:2407–2416.
51. Black DM, Geiger EJ, Eastell R, et al. Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. *N Engl J Med*. 2020;383:743–753.
52. McClung MR, Lewiecki EM, Cohen SB, et al. Denosumab in postmenopausal women with low bone mineral density. *N Engl J Med*. 2006;354:821–831.
53. Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*. 2009;361:756–765.
54. Bone HG, Wagman RB, Brandi ML, et al. 10 years of denosumab treatment in postmenopausal women with osteoporosis: results from the phase 3 randomised FREEDOM trial and open-label extension. *Lancet Diabetes Endocrinol*. 2017;5:513–523.
55. Geller M, Wagman RB, Ho PR, et al. Early findings from Prolia® post-marketing safety surveillance for atypical femoral fracture, osteonecrosis of the jaw, severe symptomatic hypocalcemia, and anaphylaxis. *Osteoporos Int*. 2014;25:56–57.
56. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women - principal results from the Women's Health Initiative randomized controlled trial. *J Am Med Assoc*. 2002;288:321–333.
57. Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *J Am Med Assoc*. 2007;297:1465–1477.
58. Reid IR, Eastell R, Fogelman I, et al. A comparison of the effects of raloxifene and conjugated equine estrogen on bone and lipids in healthy postmenopausal women. *Arch Intern Med*. 2004;164:871–879.
59. Ettinger B, Black DM, Mitlak BH, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene – results from a 3-year randomized clinical trial. *J Am Med Assoc*. 1999;282:637–645.
60. Cummings SR, Ettinger B, Delmas PD, et al. The effects of tibolone in older postmenopausal women. *N Engl J Med*. 2008;359:697–708.
61. McClung MR, San Martin J, Miller PD, et al. Opposite bone remodeling effects of teriparatide and alendronate in increasing bone mass. *Arch Intern Med*. 2005;165:1762–1768.
62. Neer RM, Arnaud CD, Zanchetta JR, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med*. 2001;344:1434–1441.
63. Díez-Pérez A, Marin F, Eriksen EF, et al. Effects of teriparatide on hip and upper limb fractures in patients with osteoporosis: a systematic review and meta-analysis. *Bone*. 2019;120:1–8.
64. Kendler DL, Marin F, Zerbini CAF, et al. Effects of teriparatide and risedronate on new fractures in post-menopausal women with severe osteoporosis (VERO): a multicentre, double-blind, double-dummy, randomised controlled trial. *Lancet*. 2018;391:230–240.
65. Nakamura T, Sugimoto T, Nakano T, et al. Randomized teriparatide human parathyroid hormone (PTH) 1-34 once-weekly efficacy research (TOWER) trial for examining the reduction in new vertebral fractures in subjects with primary osteoporosis and high fracture risk. *J Clin Endocrinol Metab*. 2012;97:3097–3106.
66. Saag KG, Shane E, Boonen S, et al. Teriparatide or alendronate in glucocorticoid-induced osteoporosis. *N Engl J Med*. 2007;357:2028–2039.
67. Cosman F, Eriksen EF, Recknor C, et al. Effects of intravenous zoledronic acid plus subcutaneous teriparatide [rhPTH(1–34)] in postmenopausal osteoporosis. *J Bone Miner Res*. 2011;26:503–511.
68. Leder BZ, Tsai JN, Uihlein AV, et al. Two years of denosumab and teriparatide administration in postmenopausal women with osteoporosis (The DATA Extension Study): a randomized controlled trial. *J Clin Endocrinol Metab*. 2014;99:1694–1700.
69. Cosman F, Wermers RA, Recknor C, et al. Effects of teriparatide in postmenopausal women with osteoporosis on prior alendronate or raloxifene: differences between stopping and continuing the antiresorptive agent. *J Clin Endocrinol Metab*. 2009;94:3772–3780.
70. Leder BZ, Tsai JN, Uihlein AV, et al. Denosumab and teriparatide transitions in postmenopausal osteoporosis (the DATA-Switch study): extension of a randomised controlled trial. *Lancet*. 2015;386:1147–1155.
71. Miller PD, Hattersley G, Riis BJ, et al. Effect of abaloparatide vs placebo on new vertebral fractures in postmenopausal women with osteoporosis: a randomized clinical trial. *J Am Med Assoc*. 2016;316:722–733.
72. Bone HG, Cosman F, Miller PD, et al. ACTIVExtend: 24 months of alendronate after 18 months of abaloparatide or placebo for postmenopausal osteoporosis. *J Clin Endocrinol Metab*. 2018;103:2949–2957.
73. Ominsky MS, Li C, Li X, et al. Inhibition of sclerostin by monoclonal antibody enhances bone healing and improves bone density and strength of nonfractured bones. *J Bone Miner Res*. 2011;26:1012–1021.
74. McClung MR, Grauer A, Boonen S, et al. Romosozumab in postmenopausal women with low bone mineral density. *N Engl J Med*. 2014;370:412–420.
75. Langdahl BL, Libanati C, Crittenden DB, et al. Romosozumab (sclerostin monoclonal antibody) versus teriparatide in postmenopausal women with osteoporosis transitioning from oral bisphosphonate therapy: a randomised, open-label, phase 3 trial. *Lancet*. 2017;390:1585–1594.
76. Chavassieux P, Chapurlat R, Portero-Muzy N, et al. Bone-forming and antiresorptive effects of romosozumab in postmenopausal women with osteoporosis: bone histomorphometry and microcomputed tomography analysis after 2 and 12 months of treatment. *J Bone Miner Res*. 2019;34:1597–1608.
77. Cosman F, Crittenden DB, Adachi JD, et al. Romosozumab treatment in postmenopausal women with osteoporosis. *N Engl J Med*. 2016;375:1532–1543.
78. Saag KG, Petersen J, Brandi ML, et al. Romosozumab or alendronate for fracture prevention in women with osteoporosis. *N Engl J Med*. 2017;377:1417–1427.
79. Merlijn T, Swart KMA, van der Horst HE, et al. Fracture prevention by screening for high fracture risk: a systematic review and meta-analysis. *Osteoporos Int*. 2020;31:251–257.
80. Grey A, Horne A, Gamble G, et al. Ten years of very infrequent zoledronate therapy in older women: an open-label extension of a randomized trial. *J Clin Endocrinol Metab*. 2020;105.
81. Saag K, Adler RA. Glucocorticoid–induced osteoporosis. In: Bilezikian JP, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, 9th ed. Hoboken, NJ: Wiley; 2019:467–473.
82. Murad MH, Drake MT, Mullan RJ, et al. Comparative effectiveness of drug treatments to prevent fragility fractures: a systematic review and network meta-analysis. *J Clin Endocrinol Metab*. 2012;97:1871–1880.

61

CKD-MBD Syndrome: Chronic Kidney Disease Produced Disorders in Skeletal, Vascular, and Cardiac Tissues and Mineral Metabolism

Keith A. Hruska and Matthew James Williams

OUTLINE

Pathobiology, 1039
  Vascular, 1039
  Cardiac, 1040
  Skeletal, 1040
  Plasma, 1042
Pathogenesis, 1042
Pathogenic Factors in the Chronic Kidney Disease–Mineral and Bone Disorder, 1043
  FGF23, 1043
  Klotho, 1043
  Sclerostin, 1044
  Hyperphosphatemia, 1044
  Calcitriol Deficiency, 1044
  Hypocalcemia, 1045
  Hyperparathyroidism, 1045
  Hypogonadism, 1045
  Other Factors, 1045
Pathology of Renal Osteodystrophy, 1045
  Predominant Hyperparathyroid Bone Disease, High-Turnover Renal Osteodystrophy, Osteitis Fibrosa, 1045
  Low-Turnover Bone Disease, Adynamic Bone Disorder, 1045
  Mixed Uremic Osteodystrophy, High-Turnover Renal Osteodystrophy Plus a Mineralization Defect, 1046
Associated Features, 1046
  Osteoporosis and Osteosclerosis, 1046
  Bone Aluminum, Iron, Lanthanum, Bisphosphonate Accumulation, 1046
  Management of Osteoporosis in Chronic Kidney Disease, 1047
  Clinical Manifestations, 1047
  Heterotopic Mineralization, Calciphylaxis, and Tumoral Calcinosis, 1047
  Bone Pain, Fractures, and Skeletal Deformities, 1047
Conclusion, 1047

In 2006, the disorders of mineral and bone metabolism in chronic kidney disease (CKD) were enshrined in a new syndrome called chronic kidney disease–mineral and bone disorder (CKD-MBD). The CKD-MBD name was coined by the Kidney Disease Improving Global Outcomes (KDIGO) foundation[1,2] for the syndrome associated with chronic kidney disease in which disorders of mineral metabolism and renal osteodystrophy (ROD) are key contributors to excess mortality.[3-6] The pandemic of CKD and end-stage kidney disease (ESKD) and the role of the CKD-MBD in the associated mortality make the syndrome a major health issue for Americans and all developed societies.[5,7-10] Recent advances define the CKD-MBD as a uniform complication of renal injury with wide variation in severity. The syndrome begins with renal injury early in CKD and contributes to the high mortality rates associated with CKD.[11] This chapter reviews recent pathophysiologic advances in the CKD-MBD and the classic pathophysiology of secondary hyperparathyroidism and ROD of CKD that have been incorporated into the syndrome. The result is the new understanding of a disordered systems biology involving the kidney, vasculature, heart, skeleton, and parathyroid glands in CKD, contributing mortality risk beyond that due to kidney disease itself (Fig. 61.1).[12]

KEY POINTS

- The chronic kidney disease–mineral and bone disorder is a syndrome incorporating the skeletal, vascular, and cardiac effects of chronic kidney disease as they are contributed by disorders in the metabolism of fibroblast growth factor 23, αKlotho, phosphate, calcium, vitamin D, parathyroid hormone, and activin receptor ligands.

PATHOBIOLOGY

Vascular

Vascular calcification (VC) is a strong risk factor for cardiovascular outcome in CKD.[13,14] The past few decades have seen extensive research into VC, revealing that the mechanism of VC is more than a consequence of a high-phosphorous milieu. In CKD, VC occurs via disordered biologic processes, including an imbalance between osteochondrogenic signaling and anticalcific events. In addition to traditionally established osteogenic signaling, dysfunctional calcium homeostasis is a prerequisite in the development of VC. Moreover,

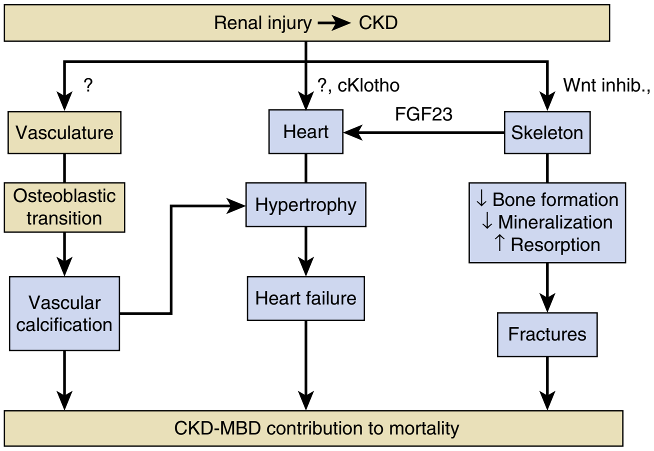


Figure 61.1 The disordered systems biology of the chronic kidney disease–mineral and bone disorder (CKD-MBD). A multiorgan systems biology "wired" together in part by circulating communications is disordered in chronic kidney disease (CKD). Renal injury increases circulating levels of factors affecting multiple organs peripheral to the kidneys. Key among these are the vasculature affecting very early in kidney disease by osteoblastic transition of cells in vessel walls; the heart's cardiomyocytes, which undergo hypertrophy due to a multitude of factors, including: 1) unknown in nature, 2) due to vascular stiffness produced in part by vascular calcification, and 3) by high levels of fibroblast growth factor 23 acting through fibroblast growth factor receptor 4.[87] The skeleton is immediately impacted during attempted renal repair by inhibitors of bone formation such as Dickkopf-1 and other Wnt inhibitors. The effects of peripheral organ damage lead to secondary pathologies of vascular calcification, cardiac hypertrophy, and renal osteodystrophy, culminating in the contributions of the CKD-MBD to the exaggerated mortality of CKD.

loss of defense mechanisms, by microorganelle dysfunction, including hyperfragmented mitochondria, mitochondrial oxidative stress, defective autophagy or mitophagy, and endoplasmic reticulum stress, may all contribute to VC.

Abnormalities of the vasculature found in early CKD produce vascular stiffness and contribute to left ventricular hypertrophy.[15,16] In addition, osteoblastic transition in arterial walls during early CKD progresses to VC.[17,18] VC further intensifies stiffness and development of left ventricular hypertrophy, all processes that contribute to cardiovascular risk and excess cardiac mortality. However, CKD directly stimulates cardiac hypertrophy independent of arterial stiffness, further contributing to mortality.[19]

The pathogenesis of VC in CKD is complex,[20] and pathologically is of two types: neointimal and arterial medial. The pathogenesis of atherosclerotic neointimal calcification is multifactorial, but it involves activation of an osteoblastic differentiation program in cells of the neointima of atherosclerotic plaques[2122] and in cells of arterial media. Diffuse calcification of the arterial tunica media is referred to as Mönckeberg sclerosis.[23] CKD is the most common cause of Mönckeberg sclerosis, especially when it complicates diabetes mellitus. In translational models of atherosclerosis and diabetes (low-density lipoprotein receptor deficiency [*ldlr-/-*] with high-fat feeding), mild renal insufficiency (equivalent to human stage 2 CKD) reduces the levels of aortic proteins involved in the contractile apparatus of smooth muscle.[17] This, coupled with evidence that vascular smooth muscle cells move into a dedifferentiated synthetic state from their normal contractile differentiated state,[24] represents significant evidence of large artery smooth muscle dedifferentiation in very early diabetic CKD.[25] Within the developmental program of mesenchymal stem cells and early progenitor cells, dedifferentiated vascular smooth muscle cells are susceptible to transition to the differentiation program of osteoblasts, which causes VC in CKD.[23] Osteoblastic transition of cells in the arterial walls produces CKD stimulated calcification of atherosclerotic plaques and the vascular media.[18,21,23,26-28] The source of these cells may be mesenchymal progenitors from the adventia moving into the media and expressing both smooth muscle and osteoblastic transcriptional programs,[29] although other lineage tracing studies support vascular smooth muscle cells.[30,31]

KEY POINTS

- Vascular calcification (VC) is highly prevalent in chronic kidney disease. The calcifications are in the media of large and medium arteries and in the neointima of atherosclerotic plaques. In the heart, coronary arteries and valves, especially aortic, are affected. VC is a major cause of arterial stiffness contributing to cardiac hypertrophy. VC is produced by osteoblastic differentiation of cells in the vascular wall and by disordered mineral metabolism producing the skeletal paradox of decreased orthotopic mineralization (deposition of hydroxylapatite in bone) and increased heterotopic mineralization (VC).

Cardiac

Cardiac hypertrophy is highly prevalent in CKD. We found in a cohort of patients with CKD stage 3 from a single-center clinic and a mean estimated glomerular filtration rate (eGFR) of 50 ml/min/mm^2, an 80% incidence of left ventricular hypertrophy by computed tomography scanning.[32] This is in concert with larger population studies.[7,19] Cardiac hypertrophy is an early stage of cardiovascular disease leading to high rates of heart failure, sudden death, and even ischemic myocardial infarction.[7] The causes of cardiac hypertrophy in CKD are multifactorial including hypertension, anemia, vascular stiffness, fibroblast growth factor 23 (FGF23),[33] and decreased αKlotho—the latter three are components of the CKD-MBD (Fig. 61.2). Vascular stiffness is a result of VC, vascular remodeling, and vascular smooth muscle dedifferentiation. The high levels of FGF23 causing cardiomyocyte hypertrophy are also late in CKD compared with the early elevations in FGF23 referred to above and elsewhere in this chapter.[34] We have found a new and novel cause of cardiac hypertrophy in our studies of activin receptor stimulation in CKD.[18,35] In three mouse models of CKD, cardiac hypertrophy develops during CKD with reductions in eGFR equivalent to human CKD stage 3. Inhibition of activin receptor type 2A prevents cardiac hypertrophy, despite not affecting FGF23 levels or fully restoring circulating Klotho.[18,35] The mechanism of this effect will introduce a new therapeutic target for cardiovascular disease in CKD and the CKD-MBD.

KEY POINTS

- The chronic kidney disease–mineral and bone disorder (CKD-MBD) contributes nontraditional risk factors to the elevated cardiovascular morbidity/mortality associated with chronic kidney disease (CKD). Cardiac hypertrophy is highly prevalent in CKD, and it leads to heart failure with both preserved and decreased ejection fraction. It contributes to arrhythmias and sudden death. Several components of the CKD-MBD (FGF23, circulating Klotho, vascular stiffness, and activin receptor activation) contribute to cardiac hypertrophy and cardiovascular risk. These direct cardiac effects of the CKD-MBD brings cardiovascular disease into the CKD-MBD syndrome.

Skeletal

Abnormalities of bone in the CKD-MBD begin with the renal response to injury due to activation of Wnt pathways in the kidney.[36,37]

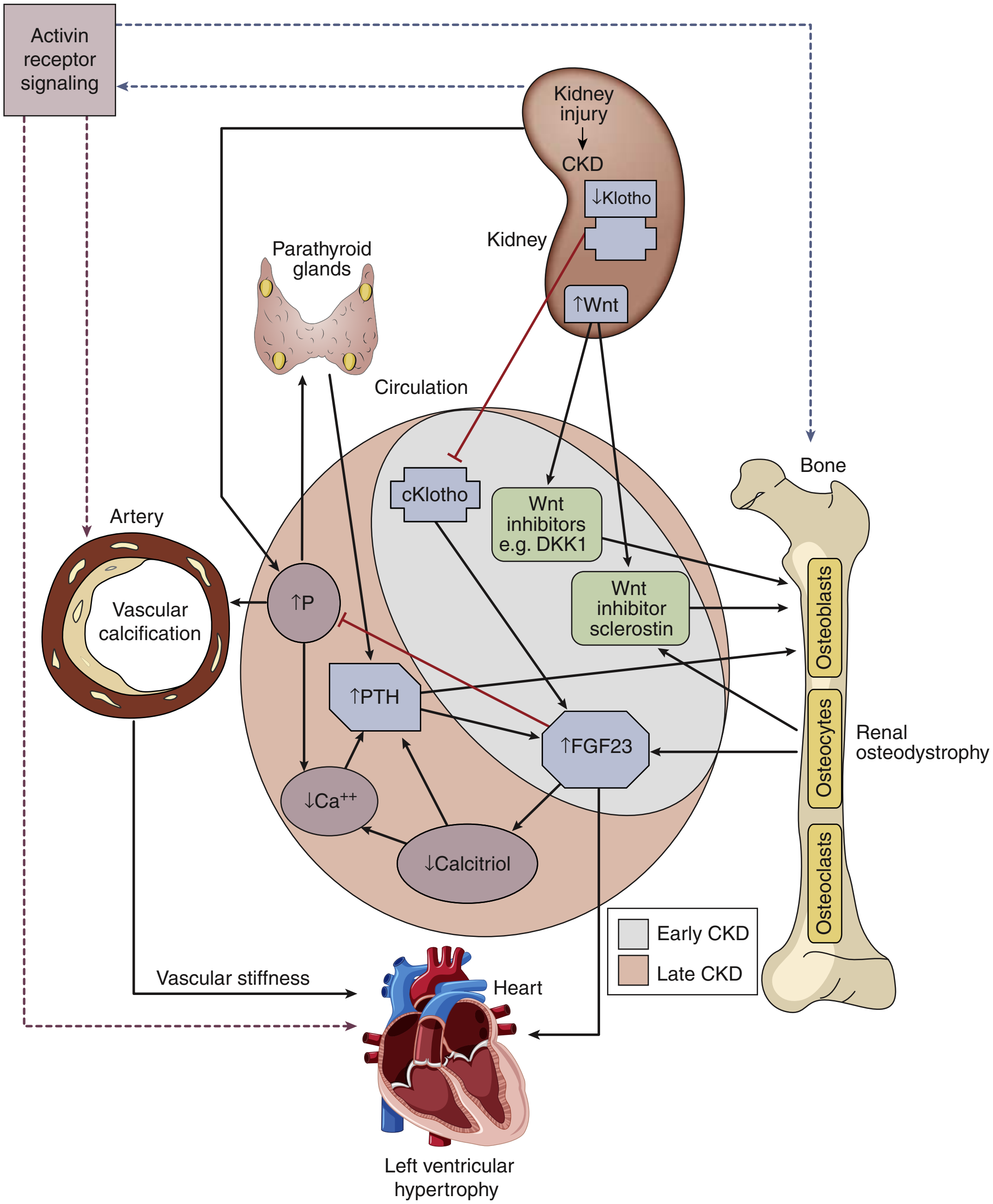

Activin receptor signaling
Kidney injury
CKD
↓Klotho
Kidney
↑Wnt
Parathyroid glands
Circulation
Bone
cKlotho
Wnt inhibitors e.g. DKK1
Wnt inhibitor sclerostin
Artery
Vascular calcification
↑P
↑PTH
↑FGF23
↓Ca++
↓Calcitriol
Osteoblasts
Osteocytes
Osteoclasts
Renal osteodystrophy
Early CKD
Late CKD
Vascular stiffness
Heart
Left ventricular hypertrophy

Figure 61.2 The disordered systems biology of the chronic kidney disease–mineral and bone disorder (CKD-MBD) (part 2). A multiorgan systems biology "wired" together in part by circulating communications is disordered in chronic kidney disease (CKD). Kidney injury leading to CKD increases levels of some factors reaching the circulation such as the Wnt inhibitor family, and decreases others such a αKlotho. Renal αKlotho is a transmembrane protein whose extracellular domain is cleaved by ADAMS 10 and 17,[80] producing circulating Klotho, whose hormonal functions include regulation of fibroblast growth factor 23 (FGF23) secretion by skeletal osteocytes.[79,80] The decrease in αKlotho produced by renal injury immediately disorders regulation of FGF23 secretion. The increases in FGF23 affect proximal tubular phosphate excretion and calcitriol production through activation of fibroblast growth factor receptor 1C through the coreceptor function of αKlotho.[202,203] The loss of Klotho disorders phosphate excretion,[100,103] leading to hyperphosphatemia, calcitriol deficiency, hypocalcemia, and hypeparathyroidism as kidney disease progresses. Hyperphosphatemia stimulates vascular calcification already in progress from osteoblastic transition.[122,204] CKD activation of activin receptor signaling systemically has been shown to contribute to vascular calcification, decreased bone formation, increased bone resorption, and cardiac hypertrophy without affecting FGF23 but increasing αKlotho.[18,35,47]

Canonical Wnt signaling transcriptionally stimulates production of multiple Wnt inhibitors, which are secreted systemic factors.[38,39] As a result, kidney injury directly inhibits bone remodeling, which is homeostatically Wnt-dependent (see Fig. 61.2).[40-42] Skeletal inhibition of β-catenin signaling and increased expression of osteocytic FGF23 are seen after a relatively mild reduction in the eGFR, stage 2 CKD.[43-45] These abnormalities affect cortical bone predominately in early CKD and progress to decreased cortical bone volume and porosity.[17] Abnormalities of remodeling, mineralization, and the material properties of bone develop in CKD, leading to major decreases in structural strength, fractures, and deformity associated with long-term disease (see Fig. 61.1).[46] Recent studies demonstrate the progressive development of a remodeling imbalance during CKD due to stimulation of osteoclast number and function, but a relative decrease in osteoblast function in view of increased osteoblast number that fails to increase bone formation as expected.[18,47] This is fueled by hyperparathyroidism in collaboration with activin/activin receptor type II signaling (see Fig. 61.2).[18] The progression of the skeletal remodeling disorder in CKD to the familiar high-turnover state related to hyperparathyroidism begins from a low-turnover state produced by CKD induced Wnt inhibition in early CKD.[39,44,45]

KEY POINTS

- The skeleton was the first organ shown to be affected by release of renal repair or injury factors from the kidney—the Wnt inhibitors. This established the paradigm that renal injury/repair produces systemic disease through release of factors to the circulation and loss of other factors. How early chronic kidney disease causes increased secretion of FGF23 and sclerostin from skeletal osteocytes remains unclear. Loss of hormonal circulating Klotho contributes to protean homeostatic disturbances.

Plasma

In addition, elevated FGF23 and sclerostin levels largely from osteocytes (although the diseased kidney produces sclerostin[48]) and elevated parathyroid hormone (PTH) levels are observed in early CKD before detectable changes in the serum inorganic phosphate (Pi), calcium, or calcitriol levels.[49-51] The early increase in FGF23 in CKD[43,49] makes it a powerful biomarker, indicating that renal injury has affected osteocyte function and secretion.[43] If the hyperparathyroidism of early CKD is prevented, the prevalence of a low turnover osteodystrophy (the adynamic bone disorder) increases, further demonstrating the effects of kidney injury on the skeleton.[52] By the time ESKD develops, skeletal pathology is nearly universally present.[53]

PATHOGENESIS

The response to renal injury early in CKD causes the CKD-MBD. Three effects of early CKD stand out from studies over the past two decades. First, renal injuries produce circulating signals that affect the vasculature, the skeleton, and the myocardium. Second, skeletal osteocytes increase FGF23 secretion. Third, reductions in renal αKlotho affect the hormonal soluble Klotho in the circulation (cKlotho) (see Fig. 61.1). Renal injury and disease produce reactivation of developmental programs of nephrogenesis in an attempt at kidney repair. These programs are generally silent in the normal adult human kidney. The best-studied example is the reactivation of the Wingless/Integration 1 (Wnt, a portmanteau of Wg and int) pathway[54,55] that controls tubular epithelial differentiation, proliferation, and polarity during nephrogenesis.[56-60] Wnt reactivation is a driving force in renal fibrosis in disease.[61-63] In the canonical Wnt signaling pathway, the transcriptional targets include a complex family of inhibitory proteins,[64] the Wnt inhibitors, including Dickkopf-1 (Dkk1),[65] soluble frizzled-related proteins, and sclerostin.[48] While Wnts are strictly autocrine/paracrine factors,[66] the Wnt inhibitors are secreted/circulating proteins.[67] The role of Wnt in renal development largely precedes the invasion of the microcirculation forming the glomerulus and the peritubular capillaries. Thus, the Wnt inhibitors are not circulating humoral substances produced by the normal kidney. Their release into the circulation during kidney injury/repair represents a systemic pathologic insult inhibiting the physiologic roles of Wnt in the rest of the body. This has been shown to have major consequences in the vasculature and the skeleton.[39,45,68]

The role of the Wnt inhibitor families in vascular function during early kidney disease was recently demonstrated by finding significantly increased renal expression of Wnt inhibitors in the injured kidney,[39,54] their increased presence in the circulation,[39] and the inhibition of one, Dkk1 by a monoclonal antibody.[39] Dkk1 neutralization was sufficient to decrease the CKD-stimulated dedifferentiation of the vasculature. Dedifferentiation was shown by decreased expression levels of proteins involved in smooth muscle contraction. Dkk1 neutralization increased smooth muscle contractile protein levels and inhibited osteoblastic transition. The levels of Runx2, the critical osteoblast transcription factor, and its transcriptional targets were decreased by Dkk1 neutralization, and the expression of Klotho was increased. The result was that Dkk1 neutralization inhibited stimulation of vascular calcium levels in an atherosclerotic calcification model stimulated by early CKD.[17,39]

The skeleton is especially sensitive to circulating Wnt inhibition, because Wnts have been shown to be a major stimuli of normal skeletal remodeling.[69] Genetic diseases produced by activating mutations in the Wnt coreceptor proteins low-density lipoprotein receptor–related proteins 5 and 6 (LRP5/6)[42,70] and by inactivating mutations in Wnt inhibitor proteins, such as sclerostin, involved in the pathogenesis of sclerosteosis[71,72] and van Buchem disease demonstrate the critical role of Wnt in skeletal remodeling and bone formation.[73] Wnts are also critical factors of the hematopoietic stem cell niche,[66,74,75] and their inhibition through circulating Wnt inhibitors is not tolerable, leading to adaptive production of another niche factor, PTH.[76,77] The adaptive increase in PTH prevents the expected inhibition of skeletal remodeling produced by circulating Wnt inhibitors.[45] Therefore, low-turnover osteodystrophy is not a generalized observation in early CKD, but it is observed when the increase in PTH is prevented.[52] In a preclinical model of atherosclerosis and type 2 diabetes, early CKD additively suppresses bone formation rates, especially in cortical bone, that is reversed by Dkk1 neutralization.[39]

Early CKD stimulates osteocyte production of FGF23 and sclerostin.[43,49] Recent studies show that the combination of dietary phosphate restriction and phosphate binders decreased FGF23 levels in early CKD with normophosphatemia.[78] The concept that changes in urinary (tubular fluid) phosphate regulates osteocytic FGF23 secretion is in agreement with the demonstrated hormonal function of cKlotho.[79] The physiologic paradigm is that regulation of disintergin metalloproteases ADAMS 10 and 17 by tubular fluid phosphate affects cleavage of membrane Klotho, producing the hormone cKlotho. In turn, cKlotho regulates osteocyte secretion of FGF23.[80] The problem with the physiologic system is that CKD decreases tubular epithelial membrane Klotho levels,[81] eliminating the adaptive potential of increasing cKlotho levels in CKD and relegating regulation of FGF23 secretion to increases in S[P] and other mechanisms. The stimulation of FGF23 and sclerostin in early CKD produce an inhibition to skeletal remodeling that becomes a background as adaptive hyperparathyroidism maintains and even increases bone remodeling rates.[45,82] In addition to FGF23, the increase in sclerostin in early CKD represents biomarker evidence of skeletal inhibition produced by renal injury. The loss of skeletal anabolism in CKD occurs in the presence of normal PTH, vitamin D, S[Ca], and S[P] levels, but it is not usually observed as the adynamic bone disorder, because disturbed homeostasis in these factors stimulates PTH secretion. The sustained increase in PTH levels produced through adaptation to CKD increases remodeling rates and eventually produces an unwanted high-turnover disorder of skeletal remodeling, osteitis firbosa.

PATHOGENIC FACTORS IN THE CHRONIC KIDNEY DISEASE–MINERAL AND BONE DISORDER

FGF23

FGF23 is the original phosphatonin (phosphate excretion–regulating hormone) discovered in studies of autosomal dominant hypophosphatemic rickets and oncogenic osteomalacia.[83,84] FGF23 is a bone-derived (osteocyte) hormone that regulates phosphorus balance by enhancing urinary phosphorus excretion and inhibiting the synthesis of 1’,25-dihydroxyvitamin D (1’,25(OH)2D). FGF23 plays an important role in maintaining phosphorus balance in states of phosphorus excess such as CKD.[85] FGF23 concentrations increase early in kidney failure and steadily rise with kidney function decline. Although this represents an appropriate physiologic response to maintain normal phosphorus balance, an important consequence is progressive 1,25(OH)$_2$D decline. Thus, while increased FGF23 secretion may be critical for avoiding phosphorus overload in early CKD, the tradeoff of lower 1,25(OH)$_2$D supports the notion that increased FGF23 may be the key factor underlying the genesis of secondary hyperparathyroidism in CKD.

Epidemiologic data have established that higher FGF23 concentrations are independently associated with higher risk of cardiovascular disease events and death in individuals across the spectrum of kidney function.[86] High FGF23 concentrations are strongly associated with higher left ventricular mass, independent of blood pressure and kidney function.[87] In addition, FGF23 induces hypertrophy of cardiomyocytes *in vitro* and *in vivo*,[87] both through fibroblast growth factor receptor (FGFR)–mediated stimulation of hypertrophy[87] and through disturbances in calcium handling in cardiac cells.[88] The prohypertrophic effects of FGF23 on cardiomyocytes are mediated by binding of FGF23 to FGFR4, even in the absence of Klotho, which activates phospholipase Cγ/calcineurin/NF of activated T-cell signaling.[89] The adverse effects of elevated FGF23 on cardiac hypertrophy in animals with experimentally induced CKD were partially reversible through targeted blockade of FGF23 binding to FGFR4 without precipitating hyperphosphatemia.[88] More recent data show that FGF23 can adversely affect cardiovascular health by stimulating sodium retention, upregulating the renin-angiotensin-aldosterone system, and enhancing the arrhythmogenic potential of heart cells via disruption of calcium trafficking.[88] These data provide clinical and mechanistic evidence for a direct link between higher FGF23 and cardiovascular disease. However, they are not without controversy,[90] and randomized trials that target pathways shown to be involved in the pathophysiology of FGF23-induced cardiovascular disease are necessary.[91]

KEY POINTS

- Fibroblast growth factor 23 (FGF23) is a relatively new hormone that was discovered in 2000. It is a major regulator of renal phosphate excretion, establishing the paradigm that signals emanating from the skeletal reservoir regulate phosphate excretion. FGF23 is dysregulated early in renal injury—both acute kidney injury and chronic kidney disease (CKD). Its elevations are a strong biomarker of cardiovascular outcome including mortality. FGF23 signals through fibroblast growth factor receptor (FGFR) requiring a coreceptor αKlotho, except for FGFR4, especially in the heart. Cardiac FGF23 effects include cardiac hypertrophy in CKD.

Klotho

αKlotho (Klotho hereafter) is highly expressed in the kidney as a transmembrane protein. In membrane-bound form, Klotho serves as a requisite coreceptor for FGF23 in conjunction with FGFRs.[92] Secondly, Klotho is a single–transmembrane pass protein with a large extracellular domain that possesses homology with glycosidases and functions as a glucuronidase and a sialidase.[93,94] FGF23 signaling through FGFRs requires the coreceptor function of Klotho. High Klotho expression has very limited tissue distribution, defining the tissue targets of FGF23 as a hormone.[95,96] These targets are the proximal and distal renal tubules, the parathyroid glands, and the brain.[95,97] Cleavage of the extracellular domain of Klotho[80] and its shedding into the urine and blood suggested hormonal function of the protein.[98,99] This was defined by its function in FGF23 secretion by osteocytes.[79] Soluble cKlotho acts as an endocrine or paracrine factor affecting multiple organs, such as the kidney, bone, brain, heart, lungs, and endothelium.[100-102]

Due to the major decrease in Klotho expression produced by renal injury,[103,104] the role of cKlotho in osteocyte regulation is limited.[102] This leaves hyperphosphatemia as the main regulator of FGF23 secretion in CKD. Furthermore, the loss of Klotho expression in CKD limits FGF23-stimulated signal transduction through FGFR/Klotho complexes. One result is the loss of negative feedback to FGF23 secretion and the continual production of FGF23 and secretion by the osteocyte. In late CKD, the very high levels of FGF23 permit anomalous FGFR activation independent of Klotho. This results in FGF23-stimulated pathologic stimuli such as cardiac myocyte hypertrophy (see Fig. 61.2).[87]

FGF23 suppresses PTH secretion and *PTH* gene expression through FGFR/Klotho signaling,[97] and PTH affects FGF23 secretion. This physiologic system of osteocyte-parathyroid and parathyroid-osteocyte regulation does not operate appropriately in CKD,[105] and the kidney-osteocyte, osteocyte-kidney signaling for phosphate homeostasis is not adaptive in CKD.

Experimental data establish a clear relationship between Klotho and the cardiovascular system. *Klotho*-null mice develop CKD-like cardiovascular phenotypes, such as VC, cardiac hypertrophy, and fibrosis. However, Klotho mRNA is not detected in the myocardium, aside from the sinoatrial node. Thus, loss of Klotho in the myocardium may not directly contribute to cardiac hypertrophy. Leifheit-Nestler et al.[106] revealed the presence of Klotho protein in cardiac muscle, which suggests that the heart can take up soluble Klotho from the circulation. A series of studies found that soluble Klotho inhibits the prohypertrophic signaling pathway and exerts unique biologic functions that are independent of FGF23.[107] For example, soluble Klotho directly binds to the type II transforming growth factor (TGF)-β receptor,[108] blocking TGF-β1 signaling. TGF-β1 is highly involved in cardiac hypertrophy.

Intriguingly, cKlotho suppresses cardiac hypertrophy that is induced by Ang-II, TGF-β, or isoprenaline.[109] Thus, cKlotho demonstrates a broad spectrum of antihypertrophic effects. Soluble Klotho inhibits the production of reactive oxygen species (ROS), and cKlotho deficiency may lead to an increase in ROS in myocardial tissue and induce cardiac hypertrophy.[110] In addition, several signaling pathways that may be involved in CKD-induced cardiac injury, including the PI3K, Wnt, and NF-κB pathways, can be regulated by cKlotho. Recently, a specific receptor for soluble Klotho, monosialoganglioside, was identified; cKlotho binds to it and inhibits raft-dependent PI3K signaling.[111] These studies show that soluble Klotho can inhibit the translocation of transient receptor potential cation channel, subfamily C, member 6 (TRPC6), an ion channel that is involved in the formation of cardiac hypertrophy.

KEY POINTS

- αKlotho is factor with widespread homeostatic functions. Most circulating Klotho (cKlotho) comes from the kidney, where the extracellular domain of renal tubular membrane Klotho is cleaved, producing the hormonal cKlotho. Membrane Klotho is the coreceptor with the fibroblast growth factor receptors for fibroblast growth factor 23 (FGF23). cKlotho regulates FGF23 secretion. cKlotho is a key cardiac homeostatic factor, and its loss in chronic kidney disease contributes to cardiac hypertrophy, vascular calcification, and osteoporosis.

Sclerostin

Sclerostin is a 22-kDa glycoprotein encoded by the *SOST* gene. Except in CKD, sclerostin is almost exclusively synthesized by osteocytes that are not present near the bone surface, but lie in the mineralized cortical and cancellous bone. Sclerostin can exert its inhibitory effects on bone formation not only by inhibiting the proliferation, differentiation, and function of osteoblast cells, but also by facilitating their apoptosis. Wnt/β-catenin signalling pathway participates in the pathogenesis of atherosclerosis and VC. Serum sclerostin levels in CKD patients are remarkably higher than those in the normal population, with their values increasing across the CKD stages. However, studies on the association of serum sclerostin with VC and mortality in renal disease patients have yielded conflicting results. Some investigations showed a positive correlation, whereas others suggested no or even negative correlation. VC occurs frequently in CKD patients, and its incidence increases across the CKD stages. Notably, VC is associated with cardiovascular events and poor prognosis in CKD. It is well-recognized that VC occurs in the early years of kidney disease, even prior to the occurrence of disordered phosphate homeostasis. VC is recognized as a pathological process of osteogenesis initiated by inflammatory factors in vessels. The upregulation of sclerostin in calcified tissues led to the hypothesis that sclerostin may be related to the pathogenic mechanism of VC in renal disease patients, leading to a large number of studies examining the association between sclerostin and VC in CKD patients. However, these studies reported inconsistent results, with some studies showing a positive association between sclerostin and VC in CKD patients, whereas others showing a negative association or no association at all.

As described above, Wnt signaling activity is a major player in physiological and adaptive bone metabolism. Wnt signaling is regulated by soluble inhibitors, with sclerostin being the most widely studied. Sclerostin's main origin is the osteocyte, and its major function is blockade of osteoblast differentiation and function. Therefore, sclerostin is a potent inhibitor of bone formation and mineralization. Consequently, blocking sclerostin via human monoclonal antibodies represents a new therapy for the treatment of postmenopausal osteoporosis,[112,113] and will be used in the treatment of ROD. In preclinical studies, sclerostin antibodies were effective in enhancing bone mass and ameliorating mechanical properties only if hyperparathyroidism was treated sufficiently by high calcium intake.[114] Only in the low-PTH group did sclerostin antibody treatment reveal remarkable changes in ROD, as evidenced by an increased cortical thickness and bone volume, as well as an increase in bone quality/strength (as measured by ultimate load and energy to failure).[114] However, sclerostin's physiology and the effects of sclerostin monoclonal antibody treatment are not limited to the skeleton. Specifically, the potential roles of sclerostin in CKD and associated pathologies of the CKD-MBD, including accelerated cardiovascular calcification, warrant specific attention. In the CKD-MBD, sclerostin antibodies may interfere at different levels and influence the multiform interplay of hyperparathyroidism, ROD, and VC, but the clinical sequelae remain obscure.[115] Research targets have been identified that need to be addressed before a valid risk/benefit ratio for sclerostin antibodies can be established in the clinical setting of CKD.

Hyperphosphatemia

As renal injury decreases nephron number, phosphate excretion is maintained by reductions in the fraction of filtered phosphate reabsorbed by the remaining nephrons under the influence of FGF23 and PTH.[116] The effects of FGF23 stimulated by cKlotho before changes in S[P] are limited by the loss of proximal tubular Klotho.[103] The combination of FGF23's and PTH's effects to decrease proximal tubular phosphate reabsorption are sufficient to maintain phosphate homeostasis through stage 3 to stage 4 CKD when hyperphosphatemia ensues.[105] This is at the cost of higher PTH and FGF23 levels. In stage 4 and stage 5 CKD, when renal injury is severe enough that the eGFR reaches levels of less than 30% of normal, hyperphosphatemia develops due to decreased renal [P] excretion despite high PTH and FGF23 levels.[105,117]

Studies demonstrate that failure of calcium and phosphorus deposition into the skeleton or excess resorption of the skeleton also contribute to abnormal S[P] and S[Ca] levels in CKD and end-stage renal disease (ESRD).[118-120] The skeleton contributes to the pathogenesis of VC through contributing to hyperphosphatemia.[16,118] Signals deriving from the skeleton are direct causes of vascular mineralization. The best-documented signal deriving from the skeleton is hyperphosphatemia.[121,122] Hyperphosphatemia stimulates osteoblastic transition in the vasculature[122,123] and directly contributes to mineralization through the calcium-phosphorus product (see Fig. 61.2).[121] Phosphorus is a signaling mechanism working through the Pit 1 and 2 sodium phosphate cotransport proteins as receptors for stimulation of MAP kinase signal transduction and stimulation of heterotopic mineralization of the vasculature in CKD.[121,124]

Hyperphosphatemia also decreases S[Ca] through physicochemical complexation and suppresses 1α-hydroxylase activity, which results in further lowering of circulating calcitriol levels (see Fig. 61.2). Moreover, a direct stimulatory effect of phosphorus on parathyroid gland cells, independent of S[Ca] and calcitriol levels, produces increased secretion and nodular hyperplasia of parathyroid gland cells (see Fig. 61.2).[125,126] A direct stimulatory effect of Pi, perhaps in the exchangeable Pi pool, regulates osteocyte FGF23 secretion.

Treatment of hyperphosphatemia employs a series of chemical salts that bind Pi with varying affinity and capacity. The clinical outcomes of this approach have been disappointing, and the reductions of serum Pi are usually approximately 1 mg/dL. In CKD predialysis, phosphate binders have been especially disappointing. The results of the IMPROVE-CKD trial are likely to lead to reduction in their use prior to dialysis.[127]

KEY POINTS

- Hyperphosphatemia develops in stage 4 chronic kidney disease (CKD) due to loss of fibroblast growth factor 23 and parathyroid hormone (PTH) effectiveness. Recent studies prove that stimulation of phosphate excretion in late CKD could control serum phosphate normal.[205] Hyperphosphatemia signals through the Pit class of phosphate transporters to stimulate vascular osteoblastic cells to mineralize the extracellular matrix (vascular calcification). Hyperphosphatemia also stimulates PTH secretion.

Calcitriol Deficiency

The physiologic actions of FGF23 from the osteocyte include inhibition of proximal tubular CYP27B1, 25-OH vitamin D 1α hydroxylase, and stimulation of CYP24A1, vitamin D 24-hydroxylase, decreasing calcitriol production and producing vitamin D deficiency in early CKD (see Fig. 61.2). As CKD advances, the functioning nephron mass is decreased, and this, combined with an increased phosphate load in the remaining nephrons and increased FGF23 levels, results in calcitriol deficiency.[128] Calcitriol deficiency in turn decreases intestinal calcium absorption, leads to the development of hypocalcemia, and produces increased *PTH* gene transcription. Calcitriol deficiency in cases of advanced CKD diminishes tissue levels of vitamin D receptors (VDR), in particular, the VDR of parathyroid gland cells.[129] Because the chief cell VDR suppresses the expression of prepro-PTH mRNA, lower circulating calcitriol levels, together with a low number of VDRs, in patients with ESRD result in stimulation of both synthesis and secretion of PTH[130] and resistance to the effects of calcitriol. The hypocalcemia in stage 5 CKD resulting from these pathophysiologic processes unchecked by medical intervention is severe and resistant to therapy with calcium and calcitriol supplementation, due to VDR deficiency.

KEY POINTS

- Due to the actions of fibroblast growth factor 23 to inhibit the 1α hydroxylase and stimulate the 24-hydroxylase, vitamin D and calcitriol are prevalent in chronic kidney disease. This stimulates parathyroid hormone secretion and eventually leads to hypocalcemia.

Hypocalcemia

As CKD progresses, hypocalcemia develops due to decreased intestinal calcium absorption. Low blood levels of ionized calcium stimulate, while high S[Ca] suppresses, PTH secretion. The action of calcium on parathyroid gland chief cells is mediated through a calcium sensor (calcium-sensing receptor [CASR]), a G protein–coupled plasma membrane receptor expressed in chief cells and kidney tubular epithelia and widely throughout the body at lower levels.[131,132] The short-term stimulation of PTH secretion induced by low S[Ca] is due to exocytosis of PTH packaged in granules, and longer-term stimulation results from an increase in the number of cells that secrete PTH. More prolonged hypocalcemia induces changes in intracellular PTH degradation and mobilization of a secondary storage pool. Within days or weeks of the onset of hypocalcemia, prepro-PTH mRNA expression is stimulated. This effect is exerted through a negative calcium response element located in the upstream flanking region of the gene encoding PTH. Expression of the calcium receptor has been shown to be suppressed by calcitriol deficiency and stimulated by calcitriol administration, suggesting an additional regulatory mechanism of the active vitamin D metabolite on PTH production. The decreased number of CASRs with low circulating calcitriol may, at least in part, explain the relative insensitivity of parathyroid gland cells to calcium in patients undergoing dialysis.

Hyperparathyroidism

All of the mechanisms discussed above result in increased production of PTH and nodular hyperplasia of the parathyroid glands in CKD. The size of the parathyroid glands progressively increases during CKD and in dialyzed patients, paralleling serum PTH levels. This increase in gland size is mainly due to diffuse cellular hyperplasia. Monoclonal chief cell growth also develops, resulting in the formation of nodules. Nodular hyperplastic glands have fewer VDRs and CASRs compared with diffusely hyperplastic glands, promoting parathyroid gland resistance to calcitriol and calcium. Sustained elevation in PTH levels, while adaptive to maintain osteoblast surfaces, produces an abnormal phenotype of osteoblast function and osteocyte stimulation, with relatively less type 1 collagen and more RANKL production than anabolic osteoblasts. A key component of the RANKL stimulation is osteocytic in origin. This leads to a high-turnover osteodystrophy, PTH receptor desensitization, and excess bone resorption.

Hypogonadism

Patients with ESRD have various states of gonadal dysfunction.[133-135] Estrogen and testosterone deficiency significantly contribute to the pathogenesis of ROD.[136,137].

Other Factors

Inflammatory mediators, acidosis[138,139], aluminum, leptin and retained catabolites are all potentially critical factors in the CKD-MBD that have not been well studied, or recently have become less important clinically. Some patients with CKD are treated with glucocorticoids, which affect bone metabolism. Patients maintained on chronic dialysis have retention of β2- microglobulin. Additionally, alterations in growth factors and other hormones involved in the regulation of bone remodeling may be disordered in CKD and ESRD, thus affecting bone remodeling and contributing to the development of ROD.

PATHOLOGY OF RENAL OSTEODYSTROPHY

With the advent of the CKD-MBD syndrome, the term ROD is restricted to the skeletal pathology of CKD and the modeling/remodeling disorders it causes. ROD is not a uniform disease. Depending on the relative contribution of the different pathogenic factors discussed above and their treatment, various pathologic patterns of bone remodeling are expressed in CKD and ESRD patients.[140]

Predominant Hyperparathyroid Bone Disease, High-Turnover Renal Osteodystrophy, Osteitis Fibrosa

Sustained excess PTH action results in increased bone turnover.[140-144] Osteoclasts, osteoblasts, and osteocytes are found in abundance. Disturbed osteoblastic activity results in a disorderly production of collagen, which results in formation of woven bone.[145] Accumulation of fibroblastic osteoprogenitors not in the osteoblastic differentiation program results in collagen deposition and in the development of fibrosis in the peritrabecular and marrow space.[27,140,146] The nonmineralized component of bone, osteoid, is increased, and the normal three-dimensional architecture of osteoid is frequently lost.[147] Osteoid seams no longer exhibit their usual birefringence under polarized light. Instead, a disorderly arrangement of woven osteoid and woven bone with a typical crisscross pattern under polarized light is seen. The mineral apposition rate and number of actively mineralizing sites are increased, as documented under fluorescent light after the administration of time-spaced fluorescent (tetracycline) markers.[148]

Low-Turnover Bone Disease, Adynamic Bone Disorder

Low-turnover uremic osteodystrophy is the other end of the spectrum of ROD.[149] The histologic hallmark of these disorders is a profound decrease in bone turnover, due to a low number of active remodeling sites and suppression of bone formation and resorption.[150,151] Bone resorption is not as decreased as formation. The result is a low-turnover osteopenic condition. The majority of trabecular bone is covered by lining cells, with few osteoclasts and osteoblasts. Bone structure is predominantly lamellar. The extent of mineralizing surfaces is markedly reduced. Usually only a few thin, single tetracycline labels are observed. Two subgroups can be identified in this type of ROD, depending on the cause of events leading to a decline in osteoblast activity: first, the adynamic bone disorder, and second, low-turnover osteomalacia from aluminum intoxication, bisphosphonate administration, or other factors.[152-154]

We postulated that the adynamic bone disorder was caused by loss of skeletal anabolic activity,[140] which has subsequently been shown to be true. First, we demonstrated that, following renal injury, if S[C], S[P], PTH, and calcitriol were maintained at normal levels, the adynamic bone disorder resulted,[52] and that it was corrected by a skeletal anabolic factor made in the normal kidney. Subsequently, we demonstrated that renal diseases increase inhibitors of Wnt in the systemic circulation,[39] and Sabbagh et al. showed that skeletal Wnt activity was suppressed in early CKD.[45] More recently, we showed that neutralization of a key Wnt inhibitor, Dkk1, was sufficient to correct the low-turnover osteodystrophy seen in a model of type 2 diabetes and early CKD.[39] In this model with further aging, resistance to PTH action is overcome by very high PTH levels, and the osteodystrophy begins to convert to higher turnover rates. Thus, the complex pathophysiology of ROD has been further clarified.

Low-turnover osteomalacia is characterized by an accumulation of unmineralized matrix in which a diminution in mineralization precedes or is more pronounced than the inhibition of collagen deposition.[155-158] Unmineralized bone represents a sizable fraction of trabecular bone volume. The increased lamellar osteoid volume is due to the presence of wide osteoid seams that cover a large portion of the trabecular surface. The occasional presence of woven bone buried within the trabeculae indicates past high bone turnover. When osteoclasts are present, they are usually seen within trabecular bone or at the small fraction of trabecular surface left without osteoid coating.

Mixed Uremic Osteodystrophy, High-Turnover Renal Osteodystrophy Plus a Mineralization Defect

Mixed uremic osteodystrophy is caused primarily by hyperparathyroidism and defective mineralization with or without increased bone formation.[148,151,155] These features may coexist in varying degrees in different patients. Increased numbers of heterogeneous remodeling sites can be seen. The number of osteoclasts is usually increased. Because active foci with numerous cells, woven osteoid seams, and peritrabecular fibrosis coexist next to lamellar sites with more reduced activity, greater production of lamellar or woven osteoid causes an accumulation of osteoid with normal or increased thickness of osteoid seams. While active mineralizing surfaces increase in woven bone with a higher mineralization rate and diffuse labeling, mineralization surfaces may be reduced in lamellar bone with a decreased mineral apposition rate.

ASSOCIATED FEATURES

Osteoporosis and Osteosclerosis

With progressive loss of renal function, cancellous bone volume may be increased along with a loss of cortical bone, but this is in part due to deposition of woven immature collagen fibrils instead of lamellar fibrils. Thus, bone strength suffers despite the increase in mass detected by dual-energy x-ray absorptiometry. Patients undergoing chronic dialysis may have a loss or gain in bone volume depending on bone balance. When the bone balance is positive, osteosclerosis may be observed when osteoblasts are active in depositing new bone, especially woven bone, thus superseding bone resorption. This is rare in the twenty-first century due to improved therapy of secondary hyperparathyroidism.[145]

In the case of negative bone balance, bone loss occurs in cortical and cancellous bone and is more rapid when bone turnover is high. In those cases, bone densitometry will detect osteopenia or osteoporosis.[159,160] The prevalence of osteoporosis in the population with CKD exceeds the prevalence in the general population.[161-163] Osteoporosis is observed in CKD before dialysis is required for ESRD.[164] When bone turnover is high, as in secondary hyperparathyroidism with osteitis fibrosa, bone resorption rates are in excess of bone formation, and osteopenia progressing to osteoporosis may result. When bone turnover is low, although both bone formation rates and bone resorption may be reduced, resorption exceeds bone formation, and loss of bone mass occurs. Thus, osteoporosis may be observed with either high-turnover[164-167] or low-turnover[168] forms of osteodystrophy. When bone resorption exceeds bone formation rates in CKD, positive phosphorus and calcium balance results in the development of hyperphosphatemia and hypercalcemia without an increase in skeletal mineral deposition, but with a stimulation of heterotopic mineralization, especially of the vasculature. The failure of the skeleton to absorb positive phosphate balance in CKD is an important stimulus to heterotopic mineralization and links the skeleton and osteoporosis in CKD to cardiovascular events and mortality.[122]

It is apparent that four forms of osteoporosis associate with the CKD-MBD: high-turnover ROD causing osteoporosis, low-turnover ROD causing osteoporosis, osteoporosis preceding CKD, and osteoporosis due to gonadal hormone deficiency. In patients with osteoporosis and decreased estimated eGFR, it is important to determine the presence of CKD and the state of the CKD-MBD.

Bone Aluminum, Iron, Lanthanum, Bisphosphonate Accumulation

Aluminum, iron, lanthanum,[169] and bisphosphonates accumulate in bone at the mineralization front, at the cement lines, or diffusely. The extent of stainable aluminum at the mineralization front correlates with histologic abnormalities in mineralization. Aluminum deposition is most severe in cases of low-turnover osteomalacia. However, aluminum deposition can be observed in all histologic forms of ROD. In patients in whom an increased aluminum burden develops, bone mineralization and bone turnover progressively decrease. These abnormalities are reversed with removal of the aluminum. Iron also accumulates at the mineralization front and can cause low-turnover forms of ROD and fractures,[170,171] similar to aluminum, although much less is known regarding skeletal iron intoxication than skeletal aluminum intoxication. Lanthanum has been added as a rare earth ion administered to CKD and ESRD patients as a phosphate binder. It is poorly absorbed, and its levels in bone are much lower than aluminum and not clinically relevant.[172] Therapy with lanthanum has not been shown to have long-term toxic effects, and studies suggest skeletal efficacy of lanthanum therapy.[173-175] Eight-year data suggest that the levels of skeletal accumulation remain below those associated with any biologic or toxic effects.[176] Lanthanum disappearance from bone deposits is slow, but not as slow as bisphosphonate disappearance.[172]

Bisphosphonates are drugs used in the treatment of osteoporosis and hypercalcemia.[177,178] There are increasing instances of bisphosphonate use in patients with CKD and ESRD, especially for treatment of VC.[179] Many patients in the multiple bisphosphonate registration trials were retrospectively found to have reduced eGFR.[180,181] Generally, these were elderly women who most likely did not have kidney disease or the CKD-MBD, as they did not have elevated PTH levels. Therefore, there is concern regarding extrapolating data from these trials to patients with osteoporosis and CKD. The nature of the bone remodeling abnormalities in CKD, especially with woven bone formation and mineralization defects, conveys a high level of risk for skeletal deposition of a substance that, once deposited, may not be removed. Such a risk of long-term retention of an active drug inhibiting bone turnover is now being recognized with use of bisphosphonates in osteogenesis imperfecta, and the rare side effect of the drugs in osteonecrosis of the jaw and atypical femoral fractures.[179,182] The US Food and Drug

Administration has added to its bisphosphonate warnings, due to nephrotoxicity of potent bisphosphonates.[177]

Management of Osteoporosis in Chronic Kidney Disease

Recently, prospective studies have shown that measurement of bone mineral density by dual-energy x-ray absorptiometry predicts incident fracture, providing nephrologists with the ability to risk-classify patients for skeletal fragility and targeted antifracture strategies for the first time. Furthermore, an expanding body of literature and anecdotal evidence suggest that pharmacologic agents used to treat osteoporosis in the general population can be safely used in patients with CKD. The KDIGO has updated guidelines for the management of CKD-associated osteoporosis and reviews recent investigations on the effects of antiosteoporotic agents in patients with CKD, including an overview of novel antiosteoporosis agents and the potential challenges related to their use in CKD.[183] This topic is beyond the scope of this chapter, and the reader is referred to recent reviews for an in-depth discussion.[184]

KEY POINTS

- Renal osteodystrophy is highly prevalent in chronic kidney disease (CKD), especially in end-stage renal disease (ESRD). It is characterized by assessing mineralization, turnover, and volume. Osteoporosis can be present in both high- and low-turnover disease but is most prevalent with low-turnover disease. New research suggests that treatment of osteoporosis in CKD until ESRD develops can be similar to its therapy with normal renal function.

Clinical Manifestations

Patients with early-stage CKD are rarely symptomatic due to ROD and its skeletal pathology, but they are usually hypertensive. If the CKD-MBD contributes to hypertension, the pulse pressure may be high due to vascular stiffness and increased pulse wave velocity.[185] However, fracture risk for CKD patients is increased more than 2-fold above age-matched cohorts without CKD.[186,187] Stimulation of VC in the setting of disordered skeletal remodeling is a life-threatening complication of the CKD-MBD.[188,189] VC produces vascular stiffness.[15] Vascular stiffness in CKD causes an increase in systolic blood pressure, a widening of the pulse pressure, and an increase in pulse wave velocity. All of these contribute to cardiac hypertrophy, heart failure, and cardiovascular mortality.[190] However, CKD causes multifactorial direct effects on the heart, contributing to progression through cardiac hypertrophy to heart failure and culminating in cardiovascular mortality. These factors include FGF23, loss of Klotho, and the newly discovered effects of Activin A in CKD.

Heterotopic Mineralization, Calciphylaxis, and Tumoral Calcinosis

Heterotopic tissue calcification may occur in the eyes and manifest as band keratopathy in the sclera or induce an inflammatory response known as the red eye syndrome in the conjunctiva. Calcium deposits are also found in the lungs, leading to restrictive lung disease. Deposits in the myocardium may cause arrhythmias, annular calcifications, valvular calcification, or myocardial dysfunction. Calcification of the kidney may contribute to progression of CKD. Most soft tissue calcifications are attributed to the increased calcium phosphate product contributed to by ROD and excess bone resorption.

The syndrome of calciphylaxis is characterized by VC in the tunica media of peripheral arteries.[191,192] These calcifications induce painful violaceous skin lesions that progress to ischemic necrosis.[193] This syndrome is associated with serious complications and often death.

Tumoral calcinosis is a form of soft tissue calcification that involves the periarticular tissues. Calcium deposits may grow to enormous size and interfere with the function of adjacent joints and organs. Although this type of calcification is usually associated with high calcium phosphate products, its exact pathogenesis is poorly understood. The recent discoveries of three single gene mutations in *FGF23*, *Klotho*, and *GALNT3* causing inherited tumoral calcinosis shed light on the role of hyperphosphatemia in the pathogenesis of tumoral calcinosis.[194-196] The role of GALNT3 in hyperphosphatemia is associated with enzymatic degradation of FGF23.

Bone Pain, Fractures, and Skeletal Deformities

Symptoms of ROD related to the skeleton appear in patients with advanced CKD.[197] Clinical manifestations are preceded, however, by an abnormal biochemical profile that should alert the physician and prompt steps to prevent more severe complications. When symptoms related to the skeleton occur, they are usually insidious, subtle, nonspecific, and slowly progressive.

Bone pain is usually vague, ill-defined, and deep-seated. Bone pain may be diffuse or localized in the lower part of the back, hips, knees, or legs. Weight-bearing and changes in position commonly aggravate bone pain. Bone pain may progress slowly to the degree that patients are completely incapacitated. Bone pain in patients with ESRD usually does not cause physical signs. Local tenderness, however, may be apparent with pressure. Occasionally, pain can occur suddenly at one joint of the lower extremities and mimic acute arthritis or periarthritis not relieved by heat or massage. A sharp chest pain may indicate rib fracture. Spontaneous fractures or fractures after minimal trauma may also occur in vertebrae (crush fractures) and in tubular bones.

Bone pain and bone fractures can be observed in all patients with ESRD independent of the underlying histologic bone disease, especially when osteoporosis is present.[161] However, low-turnover osteomalacia and aluminum-related bone disease are associated with the most severe bone pain and the highest incidence of fractures and incapacity.

Skeletal deformities can be observed in children and adults. Most children with ESRD have growth retardation, and bone deformities may develop from vitamin D deficiency (rickets) or secondary hyperparathyroidism.[138] In rickets, bowing of the long bones is seen, especially the tibiae and femora, with typical genu valgum that becomes more severe with adolescence. Longstanding secondary hyperparathyroidism in children may be responsible for slipped epiphyses secondary to impaired transformation of growth cartilage into regular metaphyseal spongiosa. This complication most commonly affects the hips, becomes obvious in preadolescence, and causes limping but is usually painless. When the radius and ulna are involved, ulnar deviation of the hands and local swelling may occur. In adults, skeletal deformities can be observed in cases of severe osteomalacia or osteoporosis and include lumbar scoliosis, thoracic kyphosis, and recurrent rib fractures.[198]

CONCLUSION

CKD produces a complex syndrome of disrupted systems biology that affects the vasculature, heart, skeleton, muscle, and mineral metabolism and that contributes a major component of its mortality. The advances in our understanding of the CKD-MBD syndrome and its pathogenesis bring into focus the great need for new therapeutics. The 2017 update of clinical practice guidelines by KDIGO[6] demonstrates that our therapies are focused on the late components of the syndrome, and have not improved its mortality in clinical trials (EVOLVE, OPERA, PRIMO, IMPROVE-CKD) despite control of secondary hyperparathyroidism.[127,199-201] Thus, there is great need for discoveries related to therapeutic targets for the CKD-MBD cardiovascular components and thedevelopment of therapeutic agents.

REFERENCES

1. Moe S, Drueke T, Cunningham J, et al. Definition, evaluation, and classification of renal osteodystrophy: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int*. 2006;69:1945–1953.
2. Olgaard K. In: Olgaard K, ed. *KDIGO:Clinical Guide to Bone and Mineral Metabolism in CKD*. New York: National Kidney Foundation; 2006.
3. Stevens LA, Djurdjev O, Cardew S, et al. Calcium, phosphate, and parathyroid hormone levels in combination and as a function of dialysis duration predict mortality: evidence for the complexity of the association between mineral metabolism and outcomes. *J Am Soc Nephrol*. 2004;15:770–779.
4. Block GA, Hulbert-Shearon TE, Levin NW, et al. Association of serum phosphorus and calcium X phosphate product with mortality risk in chronic hemodialysis patients: a national study. *Am J Kidney Dis*. 1998;31:607–617.
5. Slinin Y, Foley RN, Collins AJ. Calcium, phosphorus, parathyroid hormone, and cardiovascular disease in hemodialysis patients: the USRDS waves waves 1, 3, and 4 study. *J Am Soc Nephrol*. 2005;16:1788–1793.
6. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update Work Group. KDIGO 2017 clinical practice guideline update for the Diagnosis, evaluation, prevention, and treatment of chronic kidney disease-mineral and bone disorder (CKD-MBD). *Kidney Int Suppl*. 2017;7:1–59.
7. Foley RN, Parfrey PS, Sarnak MJ. Clinical epidemiology of cardiovascular disease in chronic renal disease. *Am J Kidney Dis*. 1998;32:S112–S119.
8. Shlipak MG, Sarnak MJ, Katz R, et al. Cystatin C and the risk of death and cardiovascular events among elderly persons. *New Engl J Med*. 2005;352:2049–2060.
9. Sarnak MJ, Levey AS, Schoolwerth AC, et al. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on kidney in cardiovascular disease, high blood pressure research, clinical Cardiology, and epidemiology and prevention. *Hypertension*. 2003;42:1050–1065.
10. Peralta CA, Shlipak MG, Judd S, et al. Detection of chronic kidney disease with creatinine, cystatin c, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. *J Am Med Assoc*. 2011;305:1545–1552.
11. Tyson CC, Smith PJ, Sherwood A, et al. Association between normal or mildly reduced kidney function, cardiovascular risk and biomarkers for atherosclerosis: results from the ENCORE trial. *Clin Kidney J*. 2017;10:666–671.
12. Graciolli FG, Neves KR, Barreto F, et al. The complexity of chronic kidney disease-mineral and bone disorder across stages of chronic kidney disease. *Kidney Int*. 2017;91:1436–1446.
13. Okuno S, Ishimura E, Kitatani K, et al. Presence of abdominal aortic calcification is significantly associated with all-cause and cardiovascular mortality in maintenance hemodialysis patients. *Am J Kidney Dis*. 2007;49:417–425.
14. Verbeke F, Van Biesen W, Honkanen E, et al. Prognostic value of aortic stiffness and calcification for cardiovascular events and mortality in dialysis patients: outcome of the Calcification Outcome in Renal Disease (CORD) study. *Clin J Am Soc Nephrol*. 2011;6:153.
15. Ix JH, de Boer IH, Peralta CA, et al. Serum phosphorus concentrations and arterial stiffness among individuals with normal kidney function to moderate kidney disease in MESA. *Clin J Am Soc Nephrol*. 2009;4:609–615.
16. Mitsnefes MM, Kimball TR, Kartal J, et al. Cardiac and vascular adaptation in pediatric patients with chronic kidney disease: role of calcium-phosphorus metabolism. *J Am Soc Nephrol*. 2005;16:2796–2803.
17. Fang Y, Ginsberg C, Sugatani T, et al. Early chronic kidney disease-mineral bone disorder stimulates vascular calcification. *Kidney Int*. 2014;85:142–150.
18. Williams MJ, Sugatani T, Agapova OA, et al. The activin receptor is stimulated in the skeleton, vasculature, heart, and kidney during chronic kidney disease. *Kidney Int*. 2018;93:147–158.
19. Di Lullo L, Gorini A, Russo D, et al. Left ventricular hypertrophy in chronic kidney disease patients: from pathophysiology to treatment. *Cardiorenal Med*. 2015;5:254–266.
20. Shanahan CM, Crouthamel MH, Kapustin A, et al. Arterial calcification in chronic kidney disease: key roles for calcium and phosphate. *Circ Res*. 2011;109:697–711.
21. Shanahan CM, Cary NRB, Metcalfe JC, et al. High expression of genes for calcification-regulating proteins in human atherosclerotic placques. *J Clin Invest*. 1994;93:2393–2402.
22. Demer LL. A skeleton in the atherosclerosis closet. *Circulation*. 1995;92:2029–2032.
23. Shanahan CM, Cary NRB, Salisbury JR, et al. Medial localization of mineralization-regulating proteins in association with Monckeberg's sclerosis: evidence for smooth muscle cell-mediated vascular calcification. *Circulation*. 1999;100:2168–2176.
24. Kokubo T, Ishikawa N, Uchida H, et al. CKD accelerates development of neointimal hyperplasia in arteriovenous fistulas. *J Am Soc Nephrol*. 2009;20:1236–1245.
25. Towler DA, Bidder M, Latifi T, et al. Diet-induced diabetes activates an osteogenic gene regulatory program in the aortas of low density lipoprotein receptor-deficient mice. *J Biol Chem*. 1998;273:30427–30434.
26. Giachelli CM, Bae N, Almeida M, et al. Osteopontin is elevated during neointima formation in rat arteries and is a novel component of human atherosclerotic plaques. *J Clin Invest*. 1993;92:1686–1696.
27. Davies MR, Lund RJ, Hruska KA. BMP-7 is an efficacious treatment of vascular calcification in a murine model of atherosclerosis and chronic renal failure. *J Am Soc Nephrol*. 2003;14:1559–1567.
28. Moe SM, O'Neill KD, Duan D, et al. Medial artery calcification in ESRD patients is associated with deposition of bone matrix proteins. *Kidney Int*. 2002;61:638–647.
29. Kramann R, Goettsch C, Wongboonsin J, et al. Adventitial MSC-like cells are progenitors of vascular smooth muscle cells and drive vascular calcification in chronic kidney disease. *Cell Stem Cell*. 2016;19:628–642.
30. Speer MY, Yang HY, Brabb T, et al. Smooth muscle cells give rise to osteochondrogenic precursors and chondrocytes in calcifying arteries. *Circ Res*. 2009;104:733–741.
31. Durham AL, Speer MY, Scatena M, et al. Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness. *Cardiovasc Res*. 2018;114:590–600.
32. Seifert ME dlFL, Ginsberg C, Rothstein M, et al. Progressive increase in left ventricular mass in stage 3 CKD despite stable kidney function. *Am J Nephrology*. 2014;39:392–399.
33. Gutierrez OM, Januzzi JL, Isakova T, et al. Fibroblast growth factor 23 and left ventricular hypertrophy in chronic kidney disease. *Circulation*. 2009;119:2545–2552.
34. Chmielewski M, Carrero JJ, Stenvinkel P, et al. Metabolic abnormalities in chronic kidney disease that contribute to cardiovascular disease, and nutritional initiatives that may diminish the risk. *Curr Opin Lipidol*. 2009;20:3–9.
35. Agapova OA, Fang Y, Sugatani T, et al. Ligand trap for the activin type IIA receptor protects against vascular disease and renal fibrosis in mice with chronic kidney disease. *Kidney Int*. 2016;89:1231–1243.
36. Zhou D, Tan RJ, Fu H, et al. Wnt/β-catenin signaling in kidney injury and repair: a double-edged sword. *Lab Invest*. 2016;96:156–167.
37. Surendran K, Schiavi SC, Hruska K. A putative Wnt antagonist secreted frizzled-related protein 4 (sFRP4) suppresses the progression of renal fibrosis follwoing unilateral ureteral obstruction. *J Am Soc Nephrol*. 2004;15:245A.
38. Aaron M. Wnt signalling: antagonistic Dickkopfs. *Curr Biol*. 2001;11:R592–R595.
39. Fang Y, Ginsberg C, Seifert M, et al. CKD-induced wingless/integration1 inhibitors and phosphorus cause the CKD-mineral and bone disorder. *J Am Soc Nephrol*. 2014;25:1760–1763.

40. Bafico A, Liu G, Yaniv A, et al. Novel mechanism of Wnt signalling inhibition mediated byDickkopf-1 interaction with LRP6/Arrow. *Nat Cell Biol*. 2001;3:683–686.
41. Gong Y, Slee RB, Fukai N, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell*. 2001;107:513–523.
42. Little RD, Carulli JP, DelMastro RG, et al. A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am J Hum Genet*. 2002;70:11–19.
43. Pereira RC, Juppner H, Azucena-Serrano CE, et al. Patterns of FGF-23, DMP1 and MEPE expression in patients with chronic kidney disease. *Bone*. 2009;45:1161–1168.
44. Graciolli FG, de Oliveira RB, dos Reis LM, et al. Wnt pathway inhibition: another actor in CKD-MBD pathophysiology? *J Am Soc Nephrol*. 2010;21:774A.
45. Sabbagh Y, Graciolli FG, O'Brien S, et al. Repression of osteocyte Wnt/β-catenin signaling is an early event in the progression of renal osteodystrophy. *J Bone Miner Res*. 2012;27:1757–1772.
46. Moe SM, Nickolas TL. Fractures in patients with CKD: time for action. *Clin J Am Soc Nephrol: CJASN*. 2016;11:1929–1931.
47. Sugatani T, Agapova OA, Fang Y, et al. Ligand trap of the activin receptor type IIA inhibits osteoclast stimulation of bone remodeling in diabetic mice with chronic kidney disease. *Kidney Int*. 2017;91:86–95.
48. Hruska KA, Sugatani T, Agapova O, et al. The chronic kidney disease — mineral bone disorder (CKD-MBD): advances in pathophysiology. *Bone*. 2017;100:80–86.
49. Isakova T, Wahl P, Vargas GS, et al. Fibroblast growth factor 23 is elevated before parathyroid hormone and phosphate in chronic kidney disease. *Kidney Int*. 2011;79:1370–1378.
50. Craver L, Marco MP, Martinez I, et al. Mineral metabolism parameters throughout chronic kidney disease stages 1-5: achievement of K/DOQI target ranges. *Nephrol Dial Transplant*. 2007;22:1171–1176.
51. Levin A, Bakris GL, Molitch M, et al. Prevalence of abnormal serum vitamin D, PTH, calcium, and phosphorus in patients with chronic kidney disease: results of the study to evaluate early kidney disease. *Kidney Int*. 2006;71:31–38.
52. Lund RJ, Davies MR, Brown AJ, et al. Successful treatment of an adynamic bone disorder with bone morphogenetic protein-7 in a renal ablation model. *J Am Soc Nephrol*. 2004;15:359–369.
53. Malluche HH, Faugere MC. Renal bone disease 1990: challenge for nephrologists. *Kidney Int*. 1990;38:193–211.
54. Surendran K, Schiavi S, Hruska KA. Wnt-dependent- β-catenin signaling is activated after unilateral ureteral obstruction, and recombinant secreted frizzled-related protein 4 alters the progression of renal fibrosis. *J Am Soc Nephrol*. 2005;16:2373–2384.
55. Terada Y, Tanaka H, Okado T, et al. Expression and function of the developmental gene Wnt-4 during experimental acute renal failure in rats. *J Am Soc Nephrol*. 2003;14:1223–1233.
56. Kispert A, Vainio S, McMahon AP. Wnt-4 is a mesenchymal signal for epithelial transformation of metanephric mesenchyme in the developing kidney. *Development*. 1998;125:4225–4234.
57. Lin Y, Liu A, Zhang S, et al. Induction of ureter branching as a response to Wnt-2b signaling during early kidney organogenesis. *Dev Dyn*. 2001;2001:26–39.
58. Stark K, Vainio S, Vassileva G, et al. Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature*. 1994;372:679–683.
59. Herzlinger D, Qiao J, Cohen D, et al. Induction of kidney epithelial morphogenesis by cells expressing Wnt-1. *Dev Biol*. 1994;166:815–818.
60. Iglesias DM, Hueber P-A, Chu L, et al. Canonical WNT signaling during kidney development. *Am J Physiol Ren Physiol*. 2007;293:F494–F500.
61. Surendran K, McCaul SP, Simon TC. A role for Wnt-4 in renal fibrosis. *Am J Physiol Renal Physiol*. 2002;282:F431–F441.
62. He W, Dai C, Li Y, et al. Wnt/{beta}-Catenin signaling Promotes renal Interstitial fibrosis. *J Am Soc Nephrol*. 2009;20:765–776.
63. Kawakami T, Ren S, Duffield JS. Wnt signalling in kidney diseases: dual roles in renal injury and repair. *J Pathol*. 2013;229:221–231.
64. Eh J, Zhang T, Domon C, et al. Wnt/{beta}-Catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol Cell Biol*. 2002;22:1172–1183.
65. Niida A, Hiroko T, Kasai M, et al. DKK1, a negative regulator of Wnt signaling, is a target of the β-catenin/TCF pathway. *Oncogene*. 2004;23:8520–8526.
66. Reya T, Duncan AW, Ailles L, et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature*. 2003;423:409–414.
67. Pietilä I, Ellwanger K, Railo A, et al. Secreted Wnt antagonist Dickkopf-1 controls kidney papilla development coordinated by Wnt-7b signalling. *Dev Biol*. 2011;353:50–60.
68. de Oliveira RB, Graciolli FG, dos Reis LM, et al. Disturbances of Wnt/β-catenin pathway and energy metabolism in early CKD: effect of phosphate binders. *Nephrol Dial Transplant*. 2013;28:2510–2517.
69. Kramer I, Halleux C, Keller H, et al. Osteocyte Wnt/β-Catenin signaling is required for normal bone homeostasis. *Mol Cell Biol*. 2010;30:3071–3085.
70. Boyden LM, Mao J, Belsky J, et al. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med*. 2002;346:1513–1521.
71. Balemans W, Ebeling M, Patel N, et al. Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). *Hum Mol Genet*. 2001;10:537–543.
72. Brunkow ME, Gardner JC, Van Ness J, et al. Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. *Am J Hum Genet*. 2001:577–589.
73. Staehling-Hampton K, Proll S, Paeper BW, et al. A 52-kb deletion in the SOST-MEOX1 intergenic region on 17q12-q21 is associated with van Buchem disease in the Dutch population. *Am J Med Genet*. 2002;110:144–152.
74. Reya T, Clevers H. Wnt signalling in stem cells and cancer. *Nature*. 2005;434:843–850.
75. ten Berge D, Kurek D, Blauwkamp T, et al. Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol*. 2011;13:1070–1075.
76. Calvi LM, Adams GB, Weibrecht KW, et al. Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature*. 2003;425:841–846.
77. Kuznetsov SA, Riminucci M, Ziran N, et al. The interplay of osteogenesis and hematopoiesis: expression of a constitutively active PTH/PTHrP receptor in osteogenic cells perturbs the establishment of hematopoiesis in bone and of skeletal stem cells in the bone marrow. *J Cell Biol*. 2004;167:1113–1122.
78. Isakova T, Barchi-Chung A, Enfield G, et al. Effects of dietary phosphate restriction and phosphate binders on FGF23 levels in CKD. *Clin J Am Soc Nephrol*. 2013;8:1009–1018.
79. Smith RC, O'Bryan LM, Farrow EG, et al. Circulating αKlotho influences phosphate handling by controlling FGF23 production. *J Clin Invest*. 2012;122:4710–4715.
80. Chen C-D, Podvin S, Gillespie E, et al. Insulin stimulates the cleavage and release of the extracellular domain of Klotho by ADAM10 and ADAM17. *Proc Natl Acad Sci Unit States Am*. 2007;104:19796–19801.
81. Hu MC, Shi M, Zhang J, et al. Klotho deficiency causes vascular calcification in chronic kidney disease. *J Am Soc Nephrol*. 2011;22:124–136.
82. Moe SM, Radcliffe JS, White KE, et al. The pathophysiology of early-stage chronic kidney disease–mineral bone disorder (CKD-MBD) and response to phosphate binders in the rat. *J Bone Miner Res*. 2011;26:2672–2681.
83. White KE, Evans WE, O'Riordan JLH, et al. Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23. *Nat Genet*. 2000;26:345–348.
84. White KE, Jonsson KB, Carn G, et al. The autosomal dominant hypophophatemic rickets (ADHR) gene is a secreted polypeptide overexpressed by tumors that cause phosphate wasting. *J Clin Endocrinol Metab*. 2001;86:497–500.
85. Gutiérrez OM. Fibroblast growth factor 23 and disordered vitamin d metabolism in chronic kidney disease: updating the "trade-off" hypothesis. *Clin J Am Soc Nephrol*. 2010;5:1710.

86. Marthi A, Donovan K, Haynes R, et al. Fibroblast growth factor-23 and risks of cardiovascular and noncardiovascular diseases: a meta-analysis. *J Am Soc Nephrol*. 2018;29:2015–2027.
87. Faul C, Amaral AP, Oskouei B, et al. FGF23 induces left ventricular hypertrophy. *J Clin Invest*. 2011;121:4393–4408.
88. Faul C. Cardiac actions of fibroblast growth factor 23. *Bone*. 2017.
89. Grabner A, Amaral Ansel P, Schramm K, et al. Activation of cardiac fibroblast growth factor receptor 4 causes left ventricular hypertrophy. *Cell Metabol*. 2015;22:1020–1032.
90. Bao J-F, Hu P-P, She Q-Y, et al. A land of controversy: fibroblast growth factor-23 and uremic cardiac hypertrophy. *J Am Soc Nephrol*. 2020;31:1423.
91. Gutiérrez OM. Fibroblast growth factor 23 and the last mile. *Clin J Am Soc Nephrol*. 2020;15:1355.
92. Kurosu H, Ogawa Y, Miyoshi M, et al. Regulation of fibroblast growth factor-23 signaling by Klotho. *J Biol Chem*. 2006;281:6120–6123.
93. Chang Q, Hoefs S, van der Kemp AW, et al. The β-glucuronidase klotho hydrolyzes and activates the TRPV5 channel. *Science*. 2005;310:490–493.
94. Cha SK, Ortega B, Kurosu H, et al. Removal of sialic acid involving Klotho causes cell-surface retention of TRPV5 channel via binding to galectin-1. *Proc Natl Acad Sci Unit States Am*. 2008;105:9805–9810.
95. Kuro-o M, Matsumura Y, Aizawa H, et al. Mutation of the mouse klotho gene leads to a syndrome resembling ageing. *Nature*. 1997;390:45–51.
96. Imura A, Tsuji Y, Murata M, et al. α-Klotho as a regulator of calcium homeostasis. *Science*. 2007;316:1615–1618.
97. Ben-Dov IZ, Galitzer H, Lavi-Moshayoff V, et al. The parathyroid is a target organ for FGF23 in rats. *J Clin Invest*. 2007;117:4003–4008.
98. Kurosu H, Yamamoto M, Clark JD, et al. Suppression of aging in mice by the hormone Klotho. *Science*. 2005;309:1829–1833.
99. Imura A, Iwano A, Tohyama O, et al. Secreted Klotho protein in sera and CSF: implication for post-translational cleavage in release of Klotho protein from cell membrane. *FEBS (Fed Eur Biochem Soc) Lett*. 2004;565:143–147.
100. Hu MC, Shi M, Gillings N, et al. Recombinant a-Klotho may be prophylactic and therapeutic for acute to chronic kidney disease progression and uremic cardiomyopathy. *Kidney Int*. 2017;91:1104–1114.
101. Ravikumar P, Li L, Ye J, et al. αKlotho deficiency in acute kidney injury contributes to lung damage. *J Appl Physiol*. 2016;120:723–732.
102. Neyra JA, Hu MC, Moe OW. Klotho in clinical nephrology. *Clin J Am Soc Nephrol*. 2021;16:162–176.
103. Hu MC, Shi M, Zhang J, et al. Klotho: a novel phosphaturic substance acting as an autocrine enzyme in the renal proximal tubule. *Faseb J*. 2010;24:3438–3450.
104. Koh N, Fujimori T, Nishiguchi S, et al. Severely reduced production of klotho in human chronic renal failure kidney. *Biochem Biophys Res Commun*. 2001;280:1015–1020.
105. Silver J, Rodriguez M, Slatopolsky E. FGF23 and PTH—double agents at the heart of CKD. *Nephrol Dial Transplant*. 2012;27:1715–1720.
106. Leifheit-Nestler M, Große Siemer R, Flasbart K, et al. Induction of cardiac FGF23/FGFR4 expression is associated with left ventricular hypertrophy in patients with chronic kidney disease. *Nephrol Dial Transplant*. 2016;31:1088–1099.
107. Xie J, Yoon J, An S-W, et al. Soluble Klotho Protects against uremic cardiomyopathy independently of fibroblast growth factor 23 and phosphate. *J Am Soc Nephrol*. 2015;26:1150–1160.
108. Doi S, Zou Y, Togao O, et al. Klotho inhibits transforming growth factor-beta1 (TGF-beta1) signaling and suppresses renal fibrosis and cancer metastasis in mice. *J Biol Chem*. 2011;286:8655–8665.
109. Hu MC, Shi M, Cho HJ, et al. Klotho and phosphate are modulators of pathologic uremic cardiac remodeling. *J Am Soc Nephrol*. 2015;26:1290–1302.
110. Yamamoto M, Clark JD, Pastor JV, et al. Regulation of oxidative stress by the anti-aging hormone Klotho. *J Biol Chem*. 2005;280:38029–38034.
111. Dalton G, An S-W, Al-Juboori SI, et al. Soluble klotho binds monosialoganglioside to regulate membrane microdomains and growth factor signaling. *Proc Natl Acad Sci Unit States Am*. 2017;114:752.
112. Cosman F, Crittenden DB, Adachi JD, et al. Romosozumab treatment in postmenopausal women with osteoporosis. *N Engl J Med*. 2016;375:1532–1543.
113. Fabre S, Funck-Brentano T, Cohen-Solal M. Anti-sclerostin antibodies in osteoporosis and other bone diseases. *J Clin Med*. 2020;9:3439.
114. Newman CL, Chen NX, Smith E, et al. Compromised vertebral structural and mechanical properties associated with progressive kidney disease and the effects of traditional pharmacological interventions. *Bone*. 2015;77:50–56.
115. Brandenburg VM, Verhulst A, Babler A, et al. Sclerostin in chronic kidney disease–mineral bone disorder think first before you block it. *Nephrol Dial Transplant*. 2019;34:408–414.
116. Slatopolsky E, Gradowska L, Kashemsant C. The control of phosphate excretion in uremia. *J Clin Invest*. 1966;45:672–677.
117. Slatopolsky E, Robson AM, Elkan I, et al. Control of phosphate excretion in uremic man. *J Clin Invest*. 1968;47:1865–1874.
118. Davies MR, Lund RJ, Mathew S, et al. Low turnover osteodystrophy and vascular calcification are amenable to skeletal anabolism in an animal model of chronic kidney disease and the metabolic syndrome. *J Am Soc Nephrol*. 2005;16:917–928.
119. Kurz P, Monier-Faugere MC, Bognar B, et al. Evidence for abnormal calcium homeostasis in patients with adynamic bone disease. *Kidney Int*. 1994;46:855–861.
120. Block GA, Bone HG, Fang L, et al. A single-dose study of denosumab in patients with various degrees of renal impairment. *J Bone Miner Res*. 2012;27:1471–1479.
121. Jono S, McKee MD, Murry CE, et al. Phosphate regulation of vascular smooth muscle cell calcification. *Circ Res*. 2000;87:e10–e17.
122. Mathew S, Tustison KS, Sugatani T, et al. The mechanism of phosphorus as a cardiovascular risk factor in chronic kidney disease. *J Am Soc Nephrol*. 2008;19:1092–1105.
123. El-Abbadi MM, Pai AS, Leaf EM, et al. Phosphate feeding induces arterial medial calcification in uremic mice: role of serum phosphorus, fibroblast growth factor-23, and osteopontin. *Kidney Int*. 2009;75:1297–1307.
124. Li X, Yang HY, Giachelli CM. Role of the sodium-dependent phosphate cotransporter, Pit-1, in vascular smooth muscle cell calcification. *Circ Res*. 2006;98:905–912.
125. Moallem E, Kilav R, Silver J, et al. RNA-protein binding and post-transcriptional regulation of parathyroid hormone gene expression by calcium and phosphate. *J Biol Chem*. 1998;273:5253–5259.
126. Naveh-Many T, Rahamimov R, Livni N, et al. Parathyroid cell proliferation in normal and chronic renal failure rats. *J Clin Invest*. 1995;4:1786–1793.
127. Toussaint ND, Pedagogos E, Lioufas NM, et al. A randomized trial on the effect of phosphate reduction on vascular end points in CKD (IMPROVE-CKD). *J Am Soc Nephrol*. 2020;31:2653.
128. Goodman WG, Quarles LD. Development and progression of secondary hyperparathyroidism in chronic kidney disease: lessons from molecular genetics. *Kidney Int*. 2007:276–288.
129. Naveh-Many T, Marx R, Keshet E, et al. Regulation of 1,25-dihydroxyvitamin D3 receptor gene expression by 1,25-dihydroxyvitamin D3 in the parathyroid in vivo. *J Clin Invest*. 1990;86:1968–1975.
130. Silver J, Russell J, Sherwood LM. Regulation by vitamin D metabolites of messenger ribonucleic acid for preproparathyroid hormone in isolated bovine parathyroid cells. *Proc Natl Acad Sci USA*. 1985;82:4270–4273.
131. Brown EM, Gamba G, Riccardi D, et al. Cloning and characterization of an extracellular Ca2+-sensing receptor from bovine parathyroid. *Nature*. 1993;366:575–580.
132. Brown EM, Hebert SC. A cloned Ca 2+ -sensing receptor; a mediator of direct effects of extracellular Ca 2+ on renal function? *J Am Soc Nephrol*. 1995;6:1530–1540.
133. Handelman DJ. Hypothalamic-pituitary gonadal dysfunction in renal failure, dialysis and renal transplantation. *Endocr Rev*. 1985;6:151–182.
134. Lim VS, Fang VS. Gonadal dysfunction in uremic men: a study of the hypothalamo-pituitary-testicular axis before and after renal transplantation. *Am J Med*. 1975;58:655–662.
135. Lindberg J. The effects of hypogonadism and kidney failure on renal osteodystrophy. *Contemp Dial Nephrol*. 1994;15:22–24.
136. Saito O, Saito T, Asakura S, et al. Involvement of estrogen deficiency in bone disease in female hemodialysis patients. In: Bezerra FCL and Oliveira CA, eds. *Menopause: Management, Complications and Social Perceptions*. Hauppauge, NY: Nova Science Publishers Inc; 2012:121–132.

137. Gluhovschi GH, Gluhovschi A, Anastasiu D, et al. Chronic kidney disease and the involvement of estrogen hormones in its pathogenesis and progression. *Rom J Intern Med*. 2012;50:135–144.
138. Bacchetta J, Harambat J, Cochat P, et al. The consequences of chronic kidney disease on bone metabolism and growth in children. *Nephrol Dial Transplant*. 2012;27:3063–3071.
139. Bushinsky DA. The contribution of acidosis to renal osteodystrophy. *Kidney Int*. 1995;47:1816–1832.
140. Hruska KA, Teitelbaum SL. Renal osteodystrophy. *New Engl J Med*. 1995;333:166–175.
141. Hruska KA, Teitelbaum SL, Kopelman R, et al. The predictability of the histological features of uremic bone disease by non-invasive techniques. *Metab Bone Dis Relat Res*. 1978;1:39–44.
142. Bordier PHJ, Marie PJ, Arnaud CD. Evolution of renal osteodystrophy: correlation of bone histomorphometry and serum mineral and immunoreactive parathyroid hormone values before and after treatment with calcium carbonate or 25-hydroxycholecalciferol. *Kidney Int*. 1975;7:S102–S112.
143. Wang M, Hercz G, Sherrard DJ, et al. Relationship between intact 1-84 parathyroid hormone and levels for bone turnover in patients on chronic maintenance dialysis. *Am J Kidney Dis*. 1995;26:836–844.
144. Malluche HH, Monier-Faugere MC. Risk of adynamic bone disease in dialyzed patients. *Kidney Int*. 1992;38:S62–S67.
145. Malluche HH, Ritz E, Lange HP. Bone histology in incipient and advanced renal failure. *Kidney Int*. 1976;9:355–362.
146. Gonzalez EA, Lund RJ, Martin KJ, et al. Treatment of a murine model of high-turnover renal osteodystrophy by exogenous BMP-7. *Kidney Int*. 2002;61:1322–1331.
147. Malluche HH, Mawad HW, Monier-Faugere MC. Renal osteodystrophy in the first decade of the new millennium: analysis of 630 bone biopsies in black and white patients. *J Bone Miner Res*. 2011;26:1368–1376.
148. Malluche H, Faugere M. Renal osteodystrophy. *N Engl J Med*. 1989;321:317–319.
149. Moriniere P, Cohen-Solal M, Belbrik S. Disappearance of aluminic bone disease in a long-term asymptomatic dialysis population restricting Al(OH)3 intake: emergence of an idiopathic adynamic bone disease not related to aluminum. *Nephron*. 1989;53:93–101.
150. Sherrard DJ. Renal osteodystrophy. *Semin Nephrol*. 1986;6:56–67.
151. Sherrard DJ, Hercz G, Pei Y, et al. The spectrum of bone disease in end-stage renal failure: an evolving disorder. *Kidney Int*. 1993;43:436–442.
152. Hodsman AB, Sherrard DJ, Wong EGC. Vitamin D resistant osteomalacia in hemodialysis patients lacking secondary hyperparathyroidism. *Ann Intern Med*. 1981;94:629–637.
153. Hercz G, Pei Y, Greenwood C, et al. Aplastic osteodystrophy without aluminum: the role of "suppressed" parathyroid function. *Kidney Int*. 1993;44:860–866.
154. Coburn J, Slatopolsky E. Vitamin D, PTH and renal osteodystrophy. In: Brenner B, Rector F, eds. *The Kidney*. Philadelphia: Saunders; 1981:2213.
155. Ott SM, Maloney NA, Coburn JW. The prevalence of bone aluminum deposition in renal osteodystrophy and its relation to the response to calcitriol therapy. *New Engl J Med*. 1982;307:709–713.
156. Andress DL, Sherrard DJ. The osteodystrophy of chronic renal failure. In: Schrier RW, ed. *Diseases of the Kidney and Urinary Tract*. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:2735–2767.
157. Andreoli SP, Bergstein JM, Sherrard DJ. Aluminum intoxication from aluminum-containing phosphate binders in children with axotemia not undergoing dialysis. *N Engl J Med*. 1984;310:1079–1084.
158. Coburn JW, Sherrard DJ, Brickman AS. A skeletal mineralizing defect in dialysis patients: a syndrome resembling osteomalacia but unrelated to vitamin D. *Contrib Nephrol*. 1980;18:172–183.
159. West S, Lok C, Jamal S. Fracture Risk Assessment in Chronic Kidney Disease, Prospective Testing Under Real World Environments (FRACTURE): a prospective study. *BMC Nephrol*. 2010;11:17.
160. Jamal S, Gilbert J, Gordon C, et al. Cortical pQCT measures are associated with fractures in dialysis patients. *J Bone Miner Res*. 2006;21:543–548.
161. Alem AM, Sherrard DJ, Gillen DL, et al. Increased risk of hip fracture among patients with end-stage renal disease. *Kidney Int*. 2000;58:396–399.
162. Cunningham J, Sprague SM, Cannata-Andia J, et al. Osteoporosis in chronic kidney disease. *Am J Kidney Dis*. 2004;43:566–571.
163. Stehman-Breen C. Osteoporosis and chronic kidney disease. *Semin Nephrol*. 2004;24:78–81.
164. Rix M, Andreassen H, Eskildsen P, et al. Bone mineral density and biochemical markers of bone turnover in patients with predialysis chronic renal failure. *Kidney Int*. 1999;56:1084–1093.
165. Bonyadi M, Waldman SD, Liu D, et al. Mesenchymal progenitor self-renewal deficiency leads to age-dependent osteoporosis in Sca-1/Ly-6A null mice. *Proc Natl Acad Sci USA*. 2003;100:5840–5845.
166. Stehman-Breen C. Bone mineral density measurements in dialysis patients. *Semin Dial*. 2001;14:228–229.
167. Stehman-Breen C, Sherrard D, Walker A, et al. Racial differences in bone mineral density and bone loss among end-stage renal disease patients. *Am J Kidney Dis*. 1999;33:941–946.
168. Coco M, Rush H. Increased incidence of hip fractures in dialysis patients with low serum parathyroid hormone. *Am J Kidney Dis*. 2000;36:1115–1121.
169. Behets GJ, Verberckmoes SC, Oste L, et al. Localization of lanthanum in bone of chronic renal failure rats after oral dosing with lanthanum carbonate1. *Kidney Int*. 2005;67:1830–1836.
170. Norrdin RW, Hoopes KJ, O'Toole D. Skeletal changes in hemochromatosis of salers cattle. *Vet Pathol*. 2004;41:612–623.
171. Van de Vyver F, Visses WJ, D'Haese P, et al. Iron overload and bone disease in chronic dialysis patients. *Nephrol Dial Transplant*. 1990;5:781–787.
172. Spasovski GB, Sikole A, Gelev S, et al. Evolution of bone and plasma concentration of lanthanum in dialysis patients before, during 1 year of treatment with lanthanum carbonate and after 2 years of follow-up. *Nephrol Dial Transplant*. 2006;21:2217–2224.
173. Freemont T, Malluche HH. Utilization of bone histomorphometry in renal osteodystrophy: demonstration of a new approach using data from a prospective study of lanthanum carbonate. *Clin Nephrol*. 2005;63:138–145.
174. D'Haese PC, Spasovski GB, Sikole A, et al. A multicenter study on the effects of lanthanum carbonate (Fosrenol ƒ) and calcium carbonate on renal bone disease in dialysis patients. *Kidney Int*. 2003;63: S73–S78.
175. Malluche HH, Siami GA, Swanepoel C, et al. Improvements in renal osteodystrophy in patients treated with lanthanum carbonate for two years. *Clin Nephrol*. 2008;70:284–295.
176. Hutchison AJ, Maes B, Vanwalleghem J, et al. Long-term efficacy and tolerability of lanthanum carbonate: results from a 3-year study. *Nephron Clin Pract*. 2006;102:c61–c71.
177. Whitaker M, Guo J, Kehoe T, et al. Bisphosphonates for osteoporosis — where do we go from here? *N Engl J Med*. 2012;366:2048–2051.
178. Berenson JR. Treatment of hypercalcemia of malignancy with bisphosphonates. *Semin Oncol*. 2002;29:12–18.
179. Toussaint ND, Elder GJ, Kerr PG. Bisphosphonates in chronic kidney disease; balancing potential benefits and adverse effects on bone and soft tissue. *Clin J Am Soc Nephrol*. 2009;4:221–233.
180. Miller PD. Is there a role for bisphosphonates in chronic kidney disease? *Semin Dial*. 2007;20:186–190.
181. Miller PD, Roux C, Boonen S, et al. Safety and efficacy of risedronate in patients with age-related reduced renal function as estimated by the Cockcroft and Gault method: a pooled analysis of nine clinical trials. *J Bone Miner Res*. 2005;20:2105–2115.
182. Khosla S, Burr D, Cauley J, et al. Bisphosphonate-associated osteonecrosis of the jaw: report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res*. 2007;22:1479–1491.
183. Ketteler M, Block GA, Evenepoel P, et al. Executive summary of the 2017 KDIGO chronic kidney disease - mineral and bone disorder (CKD-MBD) guideline update: what's changed and why it matters. *Kidney Int*. 2017;92:26–36.
184. Khairallah P, Nickolas TL. Management of osteoporosis in CKD. *Clin J Am Soc Nephrol*. 2018;13:962.
185. James MA, Watt PAC, Potter JF, et al. Pulse pressure and resistance artery structure in the elderly. *Hypertension*. 1995;26:301–306.
186. Nickolas TL, Leonard MB, Shane E. Chronic kidney disease and bone fracture: a growing concern. *Kidney Int*. 2008;74:721–731.

187. Yenchek RH, Ix JH, Shlipak MG, et al. Bone mineral density and fracture risk in older individuals with CKD. *Clin J Am Soc Nephrol*. 2012;7:1130–1136.
188. Goodman WG, London G, Amann K, et al. Vascular calcification in chronic kidney disease. *Am J Kid Dis*. 2004;43:572–579.
189. London GM, Marty C, Marchais SJ, et al. Arterial calcifications and bone histomorphometry in end-stage renal disease. *J Am Soc Nephrol*. 2004;15:1943–1951.
190. London GM, Pannier B, Guerin AP, et al. Alterations of left ventricular hypertrophy in and survival of patients receiving hemodialysis: follow-up of an interventional study. *J Am Soc Nephrol*. 2001;12:2759–2767.
191. Hafner J, Keusch G, Wahl C, et al. Uremic small-artery disease with medical calcification and intimal hyperplasia (so-called calciphylaxis): a complication of chronic renal failure and benefit from parathyroidectomy. *J Am Acad Dermatol*. 1995;33:954–962.
192. Selye H. *Calciphylaxis*. Chicago: University of Chicago Press; 1962.
193. Edwards RB, Jaffe W, Arrowsmith J, et al. Calciphylaxis: a rare limb and life threatening cause of ischaemic skin necrosis and ulceration. *Br J Plast Surg*. 2000;53:253–254.
194. Ichikawa S, Imel EA, Kreiter ML, et al. A homozygous missense mutation in human KLOTHO causes severe tumoral calcinosis. *J Clin Invest*. 2007;117:2684–2691.
195. Ichikawa S, Lyles KW, Econs MJ. A novel GALNT3 mutation in a pseudoautosomal dominant form of tumoral calcinosis: evidence that the disorder is autosomal recessive. *J Clin Endocrinol Metab*. 2005;90:2420–2423.
196. Benet-Pages A, Orlik P, Strom TM, et al. An FGF23 missense mutation causes familial tumoral calcinosis with hyperphosphatemia. *Hum Mol Genet*. 2005;14:385–390.
197. Martin KJ, González EA. Metabolic bone disease in chronic kidney disease. *J Am Soc Nephrol*. 2007;18:875–885.
198. Applbaum Y. Imaging the skeleton and joints in CKD. In: Olgaard K, Salusky IB, Silver J, eds. *The Spectrum of Mineral and Bone Disorders in Chronic Kidney Disease*. 2nd ed. Oxford University Press; 2010.
199. Thadhani R, Appelbaum E, Pritchett Y, et al. Vitamin D therapy and cardiac structure and function in patients with chronic kidney disease. *J Am Med Assoc: JAMA*. 2012;307(7):674–684.
200. Chertow GM, Parfrey P, Block G, et al. Effect of Cinacalcet on cardiovascular disease in patients undergoing dialysis. *N Engl J Med*. 2012;367:2482–2494.
201. Wang AY-M, Fang F, Chan J, et al. Effect of Paricalcitol on left ventricular mass and function in CKD—the OPERA trial. *J Am Soc Nephrol*. 2014;25:175–186.
202. Shimada T, Urakawa I, Yamazaki Y, et al. FGF-23 transgenic mice demonstrate hypophosphatemic rickets with reduced expression of sodium phosphate cotransporter type IIa. *Biochem Biophys Res Commun*. 2004;314:409–414.
203. Larsson T, Marsell R, Schipani E, et al. Transgenic mice expressing fibroblast growth factor 23 under the control of the à1 (I) collagen promoter exhibit growth retardation, osteomalacia, and disturbed phosphate homeostasis. *Endocrinology*. 2004;145:3087–3094.
204. Giachelli CM, Speer MY, Li X, et al. Regulation of vascular calcification: roles of phosphate and osteopontin. *Circ Res*. 2005;96:717–722.
205. Clerin V, Saito H, Filipski K, et al. Selective pharmacological inhibition of the sodium-dependent phosphate cotransporter NPT2a promotes phosphate excretion. *J Clin Investig*. 2020;130:6510–6522.

62

Disorders of Mineralization

Eva S. Liu and Marie B. Demay

OUTLINE

Mineralization, 1053
Clinical Features of Rickets and Osteomalacia, 1054
Radiologic Features, 1055
Nutritional Osteomalacia and Rickets, 1057
Gastrointestinal Diseases, 1058
Disorders of Vitamin D Metabolism, 1058
Hypophosphatemia, 1059
  FGF23-Mediated Hypophosphatemia, 1059
  Non–FGF23-mediated Hypophosphatemia, 1060
Miscellaneous Causes of Impaired Mineralization, 1060
  Acidosis, 1060
  Fanconi Syndrome, 1060
  Chronic Kidney Disease, 1061
  Aluminum, Fluoride, and Heavy Metal Intoxication, 1061
  Hypophosphatasia, 1061
  Osteopetrosis, 1062
  Fibrogenesis Imperfecta Ossium, 1062
Treatment, 1062
Conclusion, 1063

MINERALIZATION

Mineralization of bone is a complex process in which the calcium-phosphate inorganic mineral phase is deposited in relation to the organic protein matrix in a highly ordered fashion. Optimal mineralization requires synthesis of normal matrix at a normal rate by osteoblasts; an adequate supply of calcium and inorganic phosphate (Pi) from the extracellular fluid; an appropriate pH (~7.6) at sites of mineralization; and physiologic amounts of inhibitors of mineralization, notably, elevation in the pyrophosphate/phosphate (PPi/Pi) ratio inhibits mineralized matrix formation. Clinical disorders of mineralization can be attributed to defects at several of these steps, examples of which are shown in Table 62.1.

Mineralization of the skeleton is dependent on a normal calcium x phosphate product. Serum calcium is regulated by $1,25(OH)_2D$ and parathyroid hormone (PTH). $1,25(OH)_2D$ binds the vitamin D receptor (VDR) and promotes calcium absorption predominantly in the duodenum.[2] PTH increases calcium absorption in the distal renal tubule, and both $1,25(OH)_2D$ and PTH stimulate osteoclastic bone resorption to release calcium into the circulation. In the intestine, phosphate from the diet is absorbed passively between cells and transcellularly by the sodium phosphate transporter (NPT) 2b (NPT2b). Phosphate is filtered by the kidney and predominantly resorbed in the proximal renal tubules via the sodium phosphate transporters NPT2a and NPT2c. Fibroblast growth factor 23 (FGF23), produced by osteocytes and osteoblasts, is a major regulator phosphate homeostasis. Although both FGF23 and PTH decrease circulating phosphate by increasing renal phosphate wasting, FGF23 also impairs the production of $1,25(OH)_2D$ by inhibiting the vitamin D 1-alpha-hydroxylase and increasing inactivation of vitamin D metabolites by inducing the 24-hydroxylase.[3]

The precise mechanisms that govern mineralization have yet to be elucidated. It is thought that early mineralization occurs in matrix vesicles that accumulate calcium and Pi, initiating the formation of hydroxyapatite crystals.[4] Phosphatase orphan 1 (PHOSPHO1), a potent phosphohydrolase whose substrates include phosphoethanolamine and phosphocholine, is thought to release phosphate from the membrane phospholipids of matrix vesicles, thus providing phosphate for formation of crystals within matrix vesicles. Ablation of PHOSPHO1 in mice markedly impairs mineralization, supporting an important *in vivo* role for this protein.[5] The earliest calcium-phosphate (Ca-P) mineral phase is comprised of small apatite crystals that are thought to form complexes with phosphoproteins of the organic matrix.[6] The ultimate size and shape of the Ca-P crystals are also influenced by anionic electrolytes in the environment.

The major driving force for mineralization is the concentration of Pi, which, at normal sites of mineralization, is derived predominantly from the plasma. Therefore, control of renal Pi reabsorption is a critical process regulating mineralization. Tissue-nonspecific alkaline phosphatase (TNAP) generates additional Pi to promote mineralization.[7] Loss of function of TNAP results in accumulation of phosphorylated osteopontin, a known inhibitor of matrix mineralization,[8] as well as accumulation of inorganic pyrophosphate (PPi), a potent inhibitor of mineralization, both of which contribute to osteomalacia. Ectonucleoside triphosphate/diphosphohydrolase has also been shown to hydrolyze extracellular diphosphates into monophosphates and Pi. Mutations of this gene impair skeletal mineralization in zebrafish.[9] PPi is generated by osteoblasts and chondrocytes through the action of two enzymes: the ectonucleoside pyrophosphatase phosphodiesterase 1 (ENPP1), which releases PPi from nucleoside triphosphates, and the transmembrane protein ANK, which shuttles PPi between intracellular and extracellular compartments.[10] The specificity of normal bone mineralization is in part explained by the exclusive coexpression of the predominant organic matrix protein, type I collagen, and the enzyme that

TABLE 62.1 Examples of Disorders That Lead to Impaired Mineralization

Disorder	Mechanism
Postoperative hyperparathyroidism	Rate of matrix synthesis exceeds rate of mineralization
Fibrogenesis imperfecta ossium	Defective collagenous matrix
Renal tubular phosphate leak	Phosphate concentration deficient at mineralization sites
Vitamin D deficiency	Insufficient calcium and phosphate
Systemic acidosis	pH inadequate for mineralization
Hypophosphatasia	Excess inhibitors including pyrophosphate

hydrolyzes the inhibitor, PPi. Furthermore, the protein mineralization inhibitor matrix GLA protein is expressed in several tissues, including the growth plate, but not in bone.[11]

The mechanism of defective mineralization varies in disorders associated with osteomalacia and rickets, and biochemical indices such as serum levels of calcium and phosphorus also differ. The relative imbalance in matrix synthesis and its mineralization depend upon the underlying disease mechanism. Estimates from tetracycline labeling indicate that the appositional growth rate in normal bone is about 1 μm/day,[12] and that complete mineralization of osteoid requires approximately 10 to 21 days.[13] Thus, the thickness of the osteoid seam is dependent upon the rate of bone formation and the rate of mineralization. The major histologic criteria for the diagnosis of osteomalacia are an increased osteoid surface and increased osteoid thickness (Fig. 62.1).

The architecture of the cells and matrix in osteomalacic bone is usually normal. Osteocytes, chondrocytes, and osteoblasts all play roles in skeletal mineralization. Osteocytes have been shown to resorb their surrounding mineralized matrix in a PTH/PTH-related protein–dependent fashion, in states of high calcium demand like lactation.[14] Hypomineralized periosteocytic lesions have been also been observed in X-linked hypophosphatemia (XLH).[15] Improvement in serum phosphate and 1,25(OH)$_2$D levels in the mouse model of XLH attenuates these periosteocytic lesions.[16] In addition, there are clear abnormalities in the cellular composition of the rachitic growth plate. Whereas the resting and proliferative zones show normal histologic features (Fig. 62.2), the number of cells per column in the hypertrophic chondrocyte layer is increased, and the cells are irregularly aligned. This is accompanied by widening of the transverse diameter, resulting in characteristic cupping or flaring of the metaphysis. Osteoblast function is impaired in many forms of human rickets and osteomalacia, leading to abnormal matrix formation; notably, the hydroxylation of certain collagen lysyl residues is increased in vitamin D–deficient bone, as well as in other experimental hypocalcemic states.[17]

KEY POINTS

- Calcium and phosphate balance are critical for normal skeletal mineralization. Hormones including vitamin D, parathyroid hormone, and fibroblast growth factor 23 regulate calcium and phosphate homeostasis.

CLINICAL FEATURES OF RICKETS AND OSTEOMALACIA

The clinical manifestations of rickets are mainly related to skeletal pain and deformity, slippage of epiphyses, disturbances in growth,

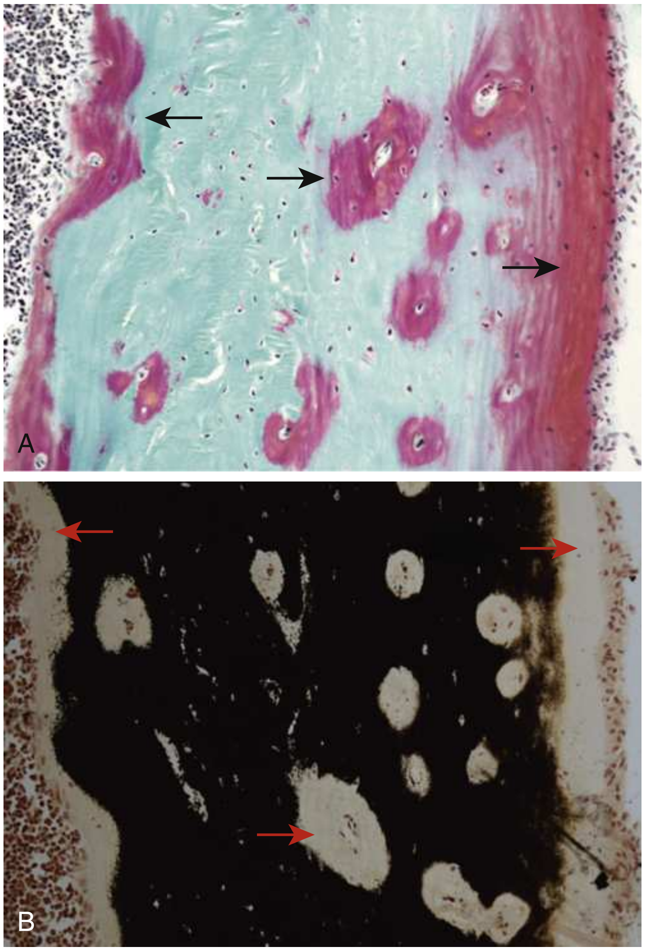


Fig. 62.1 Osteomalacia in the cortical bone of a Hyp mouse, the mouse model of X-linked hypophosphatemia. **A,** Trichrome stain, osteoid in red, mineralized bone in green. **B,** Von Kossa stain, osteoid in beige, mineralized bone in black. Osteoid is noted by arrows, pointing to the periosteal and endosteal cortical surfaces and periosteocytic halos.

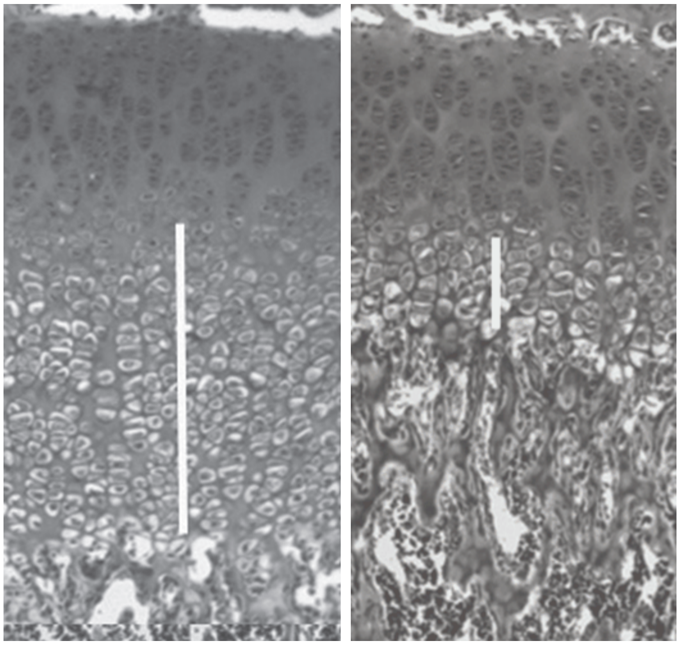


Fig. 62.2 Histology of the growth plate in a mouse model of renal phosphate wasting *(left)* and control normal mouse *(right)*. The extent of the hypertrophic chondrocyte layer is indicated by the white bar.

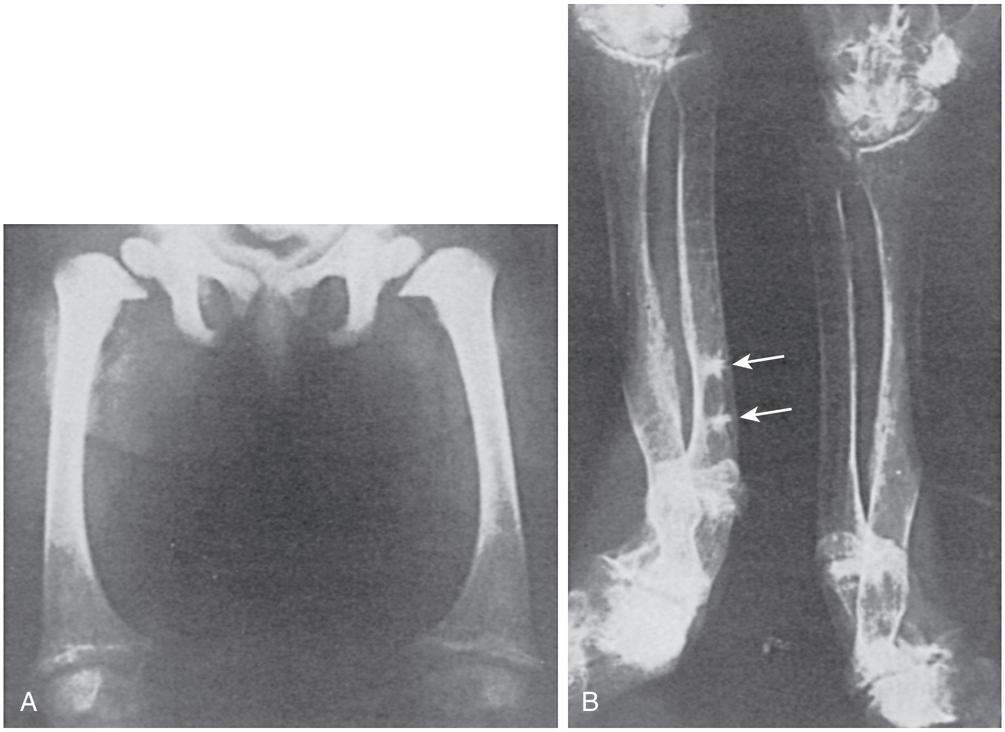


Fig. 62.3 A, Rickets in a child with Fanconi syndrome showing typical cupping of distal femoral epiphyses. B, Osteomalacia in an 80-year-old woman who had a history compatible with hypophosphatemic rickets dating to early childhood. Note multiple pseudofractures *(arrows)*.

and fracture of the osteomalacic bones. Hypocalcemia, when it occurs, may be symptomatic. Depending on the degree of hypophosphatemia, muscular weakness and hypotonia may be prominent features. Dent and Stamp have indicated several factors that underlie the clinical manifestations of rickets and osteomalacia.[18] Rickets affects predominantly rapidly growing parts of the skeleton: endochondral bone is more affected than intramembranous bone, presumably because of the more rapid growth of the former. Because the rate of growth of specific bones differs at different stages of development, the time when rickets is active determines the clinical expression. For example, the skull is growing rapidly at birth; therefore, craniotabes is a manifestation of congenital rickets. During the first year, the upper limbs and rib cage grow rapidly; thus, abnormalities at these sites are prominent, for example, rachitic rosary. Because the growth rate of the distal ulnar epiphysis is greater than that of the distal radial epiphysis, rickets is most prominent at the ulna. Similarly, proximal and distal ends of bones do not grow at the same rate; thus, rickets affects the most rapidly growing region. Rickets that occurs at the time of the pubertal growth spurt produces dramatic disturbances and results in knock-knee. Deformities that occur before the age of 4 years correct themselves if the rickets is successfully treated; if rickets persists to a later age, the deformities are permanent. Adult manifestations of osteomalacia, such as Looser's zones, are seen in young children only when the rickets is very severe. In some forms of rickets, the radiologic changes include those of secondary hyperparathyroidism (subperiosteal resorption, most commonly at the metaphyses).

In infants and young children with severe rickets, listlessness and irritability are common. In infants, hypophosphatemic myopathy is observed and may be accompanied by hypotonia. In older children and adults, the weakness may present as a proximal myopathy. Other findings in infants include parietal flattening or frontal bossing, softening of the calvarium (craniotabes), and widening of the sutures. The expanded growth plates may be evident clinically as a rachitic rosary at the sternocostal junction. Indentation or "bowing" of the lower ribs at the site of attachment of the diaphragm is known as Harrison's groove or sulcus. Pelvic deformities also occur, and the skeleton is more prone to fractures. Dental eruption may be delayed, and enamel defects are common.

Osteomalacia in adults may be difficult to diagnose on clinical grounds alone. Diffuse skeletal pain and muscular weakness may be present. Fractures occur in the ribs, vertebral bodies, and long bones, and progressive deformities may result from impaired fracture healing. Symmetric polyarthralgias resembling those of rheumatoid arthritis or polymyalgia rheumatica are occasionally observed. Proximal muscular weakness, often associated with a waddling gait and hypotonia, is thought to be a consequence of hypophosphatemia. The molecular basis remains elusive, with no difference observed in the relative concentrations of skeletal muscle phosphocreatine, adenosine triphosphate, or Pi estimated by phosphorus nuclear magnetic resonance (MR) spectroscopy.[19] Thus, the etiology of the neuromuscular features of osteomalacia is not clearly defined; however, therapy of the underlying disorder, such as vitamin D in nutritional osteomalacia, alkalinization in acidosis, and phosphate repletion in hypophosphatemic osteomalacia, results in resolution of these features.

RADIOLOGIC FEATURES

Radiologic features of rickets and osteomalacia reflect the histopathologic changes. In rickets, the epiphyseal growth plate is widened, with a cupped metaphysis and an ill-defined diaphyseal border because of decreased mineralization (Fig. 62.3). The trabecular pattern of the metaphyses is poorly defined, and the cortices of the diaphyses may be thinned and bowed.

Osteomalacia is caused by decreased mineralization; therefore, is associated with a decrease in bone density, loss of trabecular patterning, and variable degrees of thinning of the cortices. In some patients, radiologic changes are indistinguishable from those seen in osteoporosis. The characteristic finding that specifically suggests osteomalacia is the presence of radiolucent bands, ranging from a few millimeters to several centimeters in length, usually oriented perpendicularly to the surface of the bone, known as pseudofractures or Looser's zones (Fig. 62.4). They occur symmetrically and are common at the inner aspects of the femur near the femoral neck, in the pelvis, in the outer edge of the scapula, in the upper fibula, and in the metatarsals. Pseudofractures are most often seen at sites where major arteries cross the bones, and arteriography in some cases suggests that the origins of the pseudofractures correspond to the location of major arteries, like the pseudofractures in the iliac crest observed in Fig. 62.4. Local trauma, whether related to arterial pulsation or other factors, such as weight-bearing stress, is likely responsible for the symmetry of the lesions and their predilection for the described sites. The histology of Looser's zones is that of premalacic lamellar bone, some of which is surrounded by lamellar osteoid at the edge of the defect.[20] In addition, there are foci of woven bone with variable mineralization, accounting for the lower radiologic density of the pseudofracture relative to the surrounding bone. Subperiosteal erosions along the cortices may be seen when secondary hyperparathyroidism is present. Widening (or pseudowidening) of the sacroiliac joints with hazy margins has also been observed.

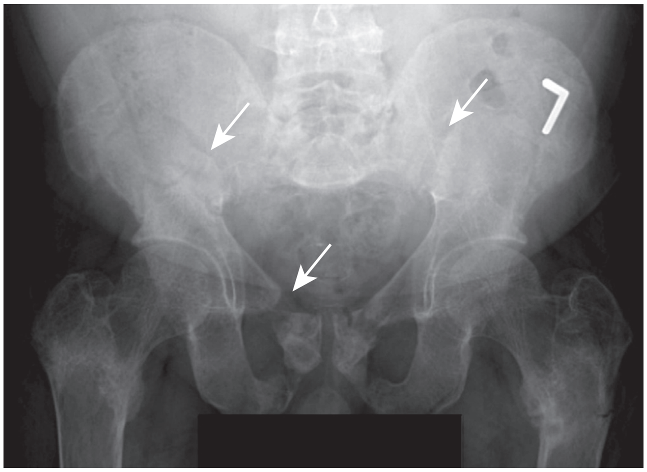


Fig. 62.4 Radiograph of the pelvis and proximal femora in an adult with tumor-induced osteomalacia. Numerous pseudofractures (Looser's zones) are observed. Arrows point to those in the superior pubic ramus and iliac crests. A left femoral fracture and right femoral stress fracture are also seen.

In some patients with osteomalacia, increased radiopacity of bones may be observed.[21] This is seen particularly in patients with renal tubular phosphate leaks, but not in vitamin D deficiency or generalized defects in proximal renal tubular reabsorption (Fig. 62.5). Despite the increase in mass of bone per unit volume, microscopically the trabeculae are covered with abnormally thickened osteoid seams typical of osteomalacia. Similar findings may be noted in patients with chronic kidney disease. The reason for the hyperostosis is unknown; however, the bone is architecturally abnormal and predisposed to low-trauma fractures. In patients with XLH, an additional finding has been the presence of a generalized involvement of the entheses, with mineralization of tendon and ligament insertions.[22] The absence of inflammatory cells, as well as other clinical features, differentiates this disorder from degenerative joint disease and the seronegative spondyloarthropathies.

A comprehensive classification of rickets and osteomalacia is shown in Table 62.2.

KEY POINTS

- Clinical and radiologic findings in children with rickets are often associated with impaired growth and growth plate maturation.
- Osteomalacia in adults can be associated with nonspecific clinical symptoms like skeletal pain, muscle weakness, and fractures.

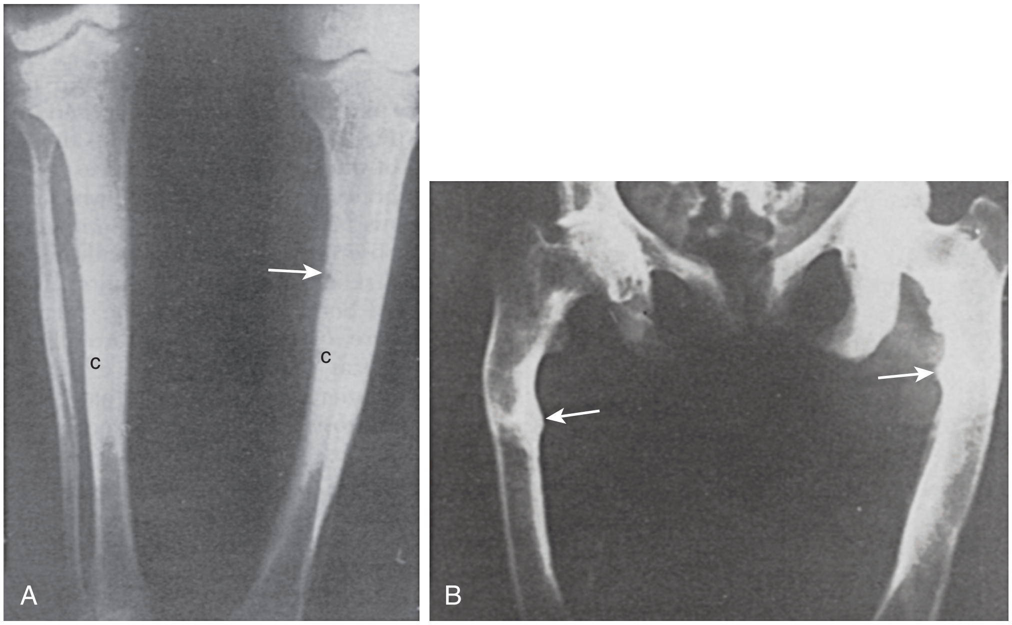


Fig. 62.5 Increased bone mass in patients with osteomalacia. **A,** Radiographs of tibiae of a 15-year-old boy with X-linked hypophosphatemia. Note the thick tibial cortex *(c)* and Looser's zone *(arrow).* **B,** Radiograph of pelvis and femora of a 38-year-old woman with hypophosphatemia present since childhood. Note Looser's zones *(arrows).*

TABLE 62.2 Classification of Rickets and Osteomalacia

I. Nutritional
- A. Vitamin D deficiency
  - 1. Dietary deficiency
  - 2. Deficient endogenous synthesis
- B. Calcium deficiency
  - 1. Dietary deficiency

II. Gastrointestinal diseases
- A. Intestinal
  - 1. Small intestine diseases with malabsorption
  - 2. Partial or total gastrectomy
  - 3. Obesity and Intestinal bypass
- B. Hepatobiliary
  - 1. Cirrhosis
  - 2. Biliary fistula
  - 3. Biliary atresia
- C. Pancreatic
  - 1. Chronic pancreatic insufficiency

III. Disorders of vitamin D metabolism
- A. Hereditary
  - 1. Isolated 25-hydroxyvitamin D deficiency
  - 2. Pseudovitamin D–deficiency rickets
  - 3. Hereditary vitamin D–resistant rickets
  - 4. CYP3A4 gain-of-function mutations
  - 5. FGF23 excess
- B. Acquired
  - 1. Hypoparathyroidism
  - 2. Anticonvulsants
  - 3. Renal insufficiency
  - 4. FGF23 excess

IV. Hypophosphatemia
- A. FGF23-dependent
  - 1. Genetic
    - a. XLH, ADHR, ARHR1, ARHR2
    - b. FAM20C mutation (Raine syndrome)
    - c. Fibrous dysplasia
  - 2. Acquired
    - b. Tumor-induced osteomalacia
- B. Non–FGF23-dependent
  - 1. Genetic
    - a. Hereditary hypophosphatemic rickets with hypercalciuria
  - 2. Dietary
    - a. Low phosphate intake with gastrointestinal disease
    - b. Alcohol use disorder
  - 3. Hyperparathyroidism

V. Miscellaneous causes of impaired mineralization
- A. Acidosis
  - 1. Distal renal tubular acidosis
- B. General renal tubular disorders (Fanconi syndrome)
  - 1. Primary renal
  - 2. Systemic disorder with associated tubular dysfunction
- C. Chronic kidney disease
- D. Aluminum, fluoride, and heavy metal intoxication
- E. Hypophosphatasia
- F.Osteopetrosis
- G. Fibrogenesis imperfecta ossium

NUTRITIONAL OSTEOMALACIA AND RICKETS

In seventeenth century Scotland and England, the association of poverty and malnutrition with the occurrence of infantile rickets was vividly documented. The link between rickets, dietary vitamin D deficiency, and correction of vitamin D deficiency by solar radiation was established in 1923 by the work of the Vienna Council. Following this discovery, fortification of certain foods with vitamin D reduced the incidence of nutritional rickets in Europe and the United States to negligible levels, so that by the 1940s vitamin D deficiency was no longer regarded as an important cause of osteomalacia and rickets.

Studies in Glasgow and London documented the reappearance of nutritional osteomalacia and rickets as a public health problem in Britain.[23,24] Since the 1950s, the population at risk primarily involved immigrants from East Asia, most commonly neonates, infants, and adolescents during pubertal growth.[25] Multiple factors have been implicated in this, including insufficient intake of calcium and vitamin D, skin pigmentation (which decreases cutaneous synthesis of vitamin D), genetic factors, and social customs (such as avoidance of sun exposure and consumption of a diet high in phytate that interferes with calcium absorption).[26,27] Based on studies demonstrating that hyperparathyroidism accelerates the rate of inactivation of vitamin D in the liver,[28] calcium deficiency leading to secondary hyperparathyroidism in this population can also exacerbate vitamin D deficiency.

Analogous increases in nutritional rickets have been reported in North America. Studies from the Mayo clinic demonstrate a dramatic increase in nutritional rickets in Olmsted county from 2.2 to 24.1 per 100,000 children between 1980 and 2000.[29] Associated risk factors included breastfeeding, low birth weight, and African ancestry. Low–birth weight infants are particularly at risk for rickets, with the reported incidence ranging from 13% to 32%. The condition should be detected early because prompt nutritional supplementation is required. Infants of vitamin D–deficient mothers are at greatest risk for neonatal rickets, which can be averted by maternal vitamin D supplementation during pregnancy.[30] Special consideration should be given to screening for vitamin D deficiency during pregnancy in women whose social and dietary history reveal inadequate sun exposure and food sources of vitamin D. Infants who are entirely breastfed and have inadequate exposure to sunlight also are at risk for rickets because the amount of vitamin D and 25(OH)D in human milk is inadequate.[31]

Vitamin D sources include dietary supplementation with ergocalciferol (vitamin D2), a plant-based provitamin, and cholecalciferol (vitamin D3), which is produced in skin by the action of ultraviolet light on 7-dehydrocholesterol. Because most foods (with the exception of wild fatty fish) contain only small amounts of vitamin D3, individuals must rely on either adequate sunlight exposure or dietary supplements to avoid vitamin D deficiency. There is marked seasonal variation in plasma 25(OH)D, independent of age and sex.[32] These variations parallel changes in duration and intensity of sun exposure, with higher values in late summer months in the northern hemisphere. In the United States, plasma 25(OH)D levels are almost twice as high in American women of European ancestry than in those of African ancestry during winter months, and the increment during summer months is also greater.[33] Plasma 25(OH)D and PTH levels are inversely correlated, implying that the changes in circulating 25(OH)D are physiologically significant.[34] Thus, adequate vitamin D supplementation is essential when exposure to sunlight is marginal.

In addition to housebound and elderly subjects, those on vegetarian or fat-free diets[35,36] are susceptible to vitamin D deficiency. Vitamin D requirements are greater in the elderly than in young adults. This difference is attributed to an age-related decline in dermal production of 7-dehydrocholesterol, the precursor for vitamin D3,[37] diminished renal production of 1,25(OH)$_2$D, and diminished intestinal absorption of calcium caused by lower levels of the VDR in the intestine.[38] Other contributing factors include previous gastric surgery and occult malabsorption, in addition to altered vitamin D metabolism. How much vitamin D supplementation is necessary varies widely based on dietary factors, solar exposure, and comorbid conditions including medications. Guidelines from the US Institute of Medicine[39] and the Endocrine Society[40] differ. The former suggests ingestion of approximately 800 IU of vitamin D per day in healthy adults and a level of 25-hydroxyvitamin D greater than 20 ng/mL (50 nmol), whereas the latter favors a level great than 30 mg/mL (75 nmol). The major limitation of the recommendations of both these guidelines with respect to skeletal health is the absence of consideration of comorbid conditions and calcium intake, as calcium deficiency *per se* can result in rickets and osteomalacia.[41] Serum 1,25(OH)$_2$D is not a reliable indicator of vitamin D status because values may be normal in individuals with vitamin D deficiency.[42]

Nutritional rickets caused by calcium deficiency was first reported in 1978 by Pettifor and coworkers,[43] who described findings in children in rural South Africa. These children had adequate sunlight exposure, normal calcium absorption, and normal serum 25(OH)D and serum 1,25(OH)$_2$D values. The biochemical and radiographic changes of rickets were entirely corrected by treatment with calcium alone. In a prospective, randomized, double-blind study of 123 Nigerian children with nutritional rickets, treatment with calcium alone or calcium and vitamin D together was more effective than treatment with vitamin D alone.[44] Calcium deficiency is also prevalent in children in North America and can contribute to nutritional rickets. Review of the records of 43 infants and toddlers who presented with rickets in New Haven, Connecticut (86% of whom were of African, Hispanic, or Middle Eastern ancestry, and more than 93% of whom had been breastfed) revealed serum 25(OH)D levels less than 15 ng/mL in only 22%. However, 86% of those with a food history had been weaned to diets with minimal calcium content.[45]

GASTROINTESTINAL DISEASES

Malabsorption associated with diseases of the small intestine, hepatobiliary tree, and pancreas is the most common cause of severe vitamin D deficiency in the United States. Disorders of the small intestine that cause malabsorption of vitamin D and resultant osteomalacia include celiac disease, Crohn's disease, scleroderma, multiple jejunal diverticula, and blind-loop syndrome. Impaired absorption of calcium *per se* also contributes to the development of rickets and osteomalacia.[41] Rickets occurs in infants and children with cholestatic liver disease[46] and biliary atresia because of markedly impaired vitamin D absorption.[47] In adults, osteomalacia is most common in patients with cholestasis (particularly primary biliary cirrhosis), but is also observed in patients with alcoholic liver disease and other forms of cirrhosis. In most patients, the low serum levels of 25(OH)D are attributable to impaired intestinal absorption of vitamin D; however, reduced exposure to ultraviolet light and reduced dietary intake also contribute.[48] Treatment with vitamin D metabolites can heal hepatic osteomalacia.

Other populations at risk for development of nutritional osteomalacia are the morbidly obese and those who have undergone bariatric surgery for treatment of severe obesity.[49-52] The etiology for the osteomalacia is thought to be as a result of the presence of vitamin D deficiency before surgery combined with postoperative impairment in intestinal calcium and vitamin D absorption. The low vitamin D levels lead to hypocalcemia, hypophosphatemia, and osteomalacia. Because adequate mixing of food with bile acids and pancreatic enzymes in the small intestine is needed for vitamin D absorption, patients who undergo Roux-en-Y gastric bypass or biliopancreatic diversion/duodenal switch are particularly at risk for postoperative vitamin D deficiency.[53] In one series, iliac bone biopsies from patients who had undergone intestinal bypass revealed the presence of osteomalacia in six of 21 patients.[52] Treatment with vitamin D2 (36,000 IU/day), as well as supplemental calcium (27 mmol/day), was required to promote a more positive calcium balance in some individuals.[50] Patients who undergo sleeve gastrectomy are at risk for nutritional osteomalacia because of inadequate dietary intake of vitamin D and calcium.[54,55] An increased frequency of osteomalacia is also observed postgastrectomy as well, the severity of the mineralization defect being positively correlated with serum 25(OH)D, but not serum 1,25(OH)$_2$D, levels.[56]

DISORDERS OF VITAMIN D METABOLISM

Rickets or osteomalacia can be caused by gene mutations in enzymes responsible for vitamin D metabolism (see Chapter 58). Isolated deficiency of 25(OH)D is a rare autosomal recessive disorder caused by impaired hepatic production of 25(OH)D. The disease is characterized by low serum 25(OH)D, normal serum 1,25(OH)$_2$D, hypocalcemia, increases in serum alkaline phosphatase, secondary hyperparathyroidism, and early onset of rickets. Homozygous L99P and K242N inactivating mutations of CYP2R1 were found to be the cause of 25(OH)D deficiency and rickets in Nigerian children with this disorder.[57] These findings provide evidence that CYP2R1 is the key enzyme catalyzing the conversion of vitamin D to 25-hydroxyvitamin D.

Pseudovitamin D–deficiency rickets is an autosomal recessive disorder caused by impaired production of 1,25(OH)$_2$D. The disease is caused by mutations of the *CYP27B1* gene that encodes the 25(OH) D 1-alpha-hydroxylase, the mitochondrial enzyme that is responsible for conversion of 25(OH)D to 1,25(OH)$_2$D.[58-60] Affected individuals appear normal at birth and develop the clinical and biochemical changes of rickets during the first year of life. Hypocalcemia, hypophosphatemia, elevated serum PTH, and elevated serum alkaline phosphatase are consistent findings. The diagnosis is made by genetic analyses and/or demonstration of a normal serum 25(OH)D and low, or undetectable, serum 1,25(OH)$_2$D in a patient with rickets who responds to physiologic doses of 1,25(OH)$_2$D3.

Hereditary vitamin D–resistant rickets is a rare inherited disorder characterized by resistance of target organs to 1,25(OH)$_2$D. The disease is caused by mutations of the *VDR* gene,[61] is transmitted as an autosomal recessive trait, and is often associated with consanguineous parentage. Onset of symptoms can occur at any age from infancy to adolescence. Patients exhibit hypocalcemia, hypophosphatemia, elevation of serum alkaline phosphatase and PTH, normal or increased serum 25(OH)D, and marked elevation of serum 1,25(OH)$_2$D. Patients may also develop alopecia. Because the pathophysiologic abnormality underlying this disorder is a receptor defect, no treatment has been uniformly successful. Therapeutic responses to pharmacologic doses of vitamin D metabolites and oral calcium supplements vary. Because the main physiologic effect of 1,25(OH)$_2$D is to promote intestinal calcium absorption, treatment with parenteral calcium infusions, circumventing the intestinal resistance to 1,25-dihydroxyvitamin D, can normalize the biochemical abnormalities and lead to clinical and radiologic remission of the osteomalacic lesions.[62]

The molecular basis for a novel form of vitamin D–resistant rickets was reported to be gain-of-function mutations in CYP3A4, a liver cytochrome P450 enzyme. Although this enzyme is thought to play a minor role in normal vitamin D physiology, induction of CYP3A4 underlies the accelerated metabolism of vitamin D observed with several pharmacological agents. The mutation in the substrate recognition site of CYP3A4 seen in affected individuals increases the rate of 1,25-dihydroxyvitamin D inactivation 10-fold versus wild-type CYP3A4. Affected individuals with this disorder have low serum calcium, phosphate, 25(OH)D, and 1,25$(OH)_2$D levels, with secondary hyperparathyroidism. Because this condition leads to accelerated catabolism of 25(OH)D and 1,25$(OH)_2$D, affected patients are treated with high doses of vitamin D or calcitriol.[63]

Hypoparathyroidism is a rare cause of osteomalacia.[64,65] The skeletal disease is attributed to a decrease in 1,25$(OH)_2$D synthesis owing to of lack of PTH and impaired PTH-mediated bone resorption and renal calcium reabsorption, all of which lead to hypocalcemia. Thus, appropriate therapy for hypoparathyroidism should prevent the mineralization defect. Osteomalacia occurs infrequently in pseudohypoparathyroidism, attributed to decreased 1,25$(OH)_2$D, a consequence of impaired renal response to PTH-induced formation of cyclic adenosine monophosphate, as well as hypocalcemia.[66,67] These parathyroid-related disorders are discussed in greater detail in Chapters 52 and 57.

KEY POINTS

- Nutritional rickets can result from inadequate dietary calcium or vitamin D intake.

HYPOPHOSPHATEMIA

FGF23-Mediated Hypophosphatemia

Impaired tubular reabsorption of phosphate is the underlying pathogenesis of hypophosphatemia in many disorders. In 1937, Albright and coworkers reported their studies of a 16-year-old boy with longstanding rickets in whom standard doses of vitamin D failed to produce clinical improvement.[68] Healing of the skeletal disease observed after administration of extremely high doses of vitamin D introduced the concept of "vitamin D resistance" in certain types of rickets. Subsequently, vitamin D–resistant rickets has been classified into several clinical and biochemical subtypes, the most common of which is XLH, characterized by elevated levels of FGF23 that result in renal phosphate wasting and inhibition of the vitamin D 1-alpha-hydroxylase, thus impairing production of 1,25$(OH)_2$D.[69] Affected individuals have hypophosphatemia, low renal resorption of phosphate, inappropriately normal 1,25$(OH)_2$D levels (because hypophosphatemia normally increases 1,25$(OH)_2$D), and modest elevations in serum PTH. Children usually present with clinical and radiographic evidence of rickets, growth retardation, and dental caries. In adults, the clinical course is often complicated by osteoarthritis and mineralization of the tendon-bone insertion sites (enthesopathy).[22] XLH is traditionally treated with a combination oral phosphate and 1,25$(OH)_2$D (calcitriol). More recently, a novel anti-FGF23 blocking antibody (burosumab) was approved for the treatment of children and adults with XLH. While antibody treatment is able to increase serum phosphate levels into the low normal range without increasing serum PTH levels,[70] no therapy has been shown to normalize growth and skeletal mineralization.[71,72]

Positional cloning used to identify the molecular basis for XLH[73] revealed a mutation in the *PHEX* gene, which is structurally similar to genes encoding a group of membrane-bound metalloendopeptidases.[74] It is still unclear how the mutations in this gene lead to increased levels of FGF23.

Autosomal dominant hypophosphatemic rickets (ADHR) is also characterized by isolated renal tubular phosphate wasting and inappropriately normal plasma 1,25$(OH)_2$D levels. Positional cloning identified *FGF23* as the gene mutated in ADHR.[75] These mutations, localized to a subtilisin-like proprotein convertase cleavage site, decrease inactivation of FGF23, thereby leading to elevated circulating levels, which result in renal phosphate wasting. In contrast to XLH, ADHR displays variable and incomplete penetrance and may not present until adolescence/adulthood. Iron deficiency increases FGF23 mRNA expression; thus, iron deficiency can lead to hypophosphatemia in normal individuals and unmask ADHR.[76] Treatment of iron deficiency in patients with ADHR has been shown to normalize serum FGF23 and phosphate levels.[77]

Two forms of autosomal recessive hypophosphatemic rickets (ARHR) have been described. The genetic basis for ARHR1 is a mutation in dentin matrix protein 1 (DMP1), which, like FGF23, is expressed in osteocytes. The clinical features are similar to those seen in XLH, including hypophosphatemia, inappropriately normal or low 1,25-dihydroxyvitamin D, and high circulating FGF23.[78] ARHR2 is secondary to mutations in ectonucleotide pyrophosphatase/phosphodiesterase1 (ENPP1), which cleaves adenosine triphosphate to release the mineralization inhibitor PPi. Clinical characteristics are similar to ADHR1; however, FGF23 levels are not always elevated. Loss-of-function mutations of ENPP1 are the molecular basis for generalized arterial calcification of infancy, an often fatal disorder.[79] Mutations in FAM20c, a kinase that can phosphorylate FGF23, osteopontin and DMP1, have been associated with an FGF23-dependent hypophosphatemic syndrome characterized by dental abnormalities and ectopic mineralization.[80]

Dent and Gertner also described three patients with fibrous dysplasia with hypophosphatemic osteomalacia or rickets.[81] Subsequent investigations revealed that dysplastic tissue produces FGF23, implicating this hormone in the pathogenesis of the renal phosphate leak.[82] Treatment of these and other rare inherited disorders of phosphate homeostasis are discussed in Chapter 53.

Tumor-induced osteomalacia is an acquired hypophosphatemic syndrome secondary to tumor secretion of phosphaturic factors, including FGF23. In 1959, Prader and coworkers described the case of an 11-year-old girl who, over the course of a year, developed severe symptomatic rickets accompanied by hypophosphatemia, renal phosphate wasting, and mild hypocalcemia.[83] The child was found to have a large tumor that was pathologically diagnosed as a reparative giant cell granuloma of a rib. Following excision of the tumor, the rickets healed without any specific therapy. It was postulated that the tumor may have produced a "rachitogenic substance." Since that time, numerous patients have been described with tumor-induced osteomalacia. One review reported that, of 72 tumors, 40 were localized to bone and 31 to soft tissues; two-thirds of the tumors occurred in the extremities.[84] More than one-third of the tumors were classified as vascular tumors, and half of these were hemangiopericytomas. All of the tumors exhibited prominent vascularity, multinucleated giant cells, and primitive stromal cells. Ten of the tumors were classified as malignant. Although most neoplasms associated with this syndrome are of mesenchymal origin, cases associated with prostate carcinoma and pheochromocytoma have been reported. In most of the reported cases, the removal of the tumor results in clinical cure of the osteomalacia or rickets; however, the tumors are usually small and often are not identified until years after development of clinical osteomalacia. Recurrence of the tumor or inadequate removal of the malignancy may prevent a complete clinical response.

Patients with tumor-induced osteomalacia exhibit hypophosphatemia, renal phosphate wasting, and normal serum calcium levels. Serum PTH levels are variable. Tumors that have been isolated express high levels of FGF23 mRNA,[85] and many have a fibronectin:FGFR1 or fibronectin:FGF1 fusion thought to be pathogenic in the increased FGF23 production.[86] Whereas other phosphaturic factors have been found in these tumors, normalization of circulating FGF23 is seen after removal of the tumor and correlates with resolution of the hypophosphatemia. As in XLH, inappropriately low levels of serum 1,25$(OH)_2$D are observed owing to expression of FGF23,[85] and tumor resection may result in a dramatic increase in 1,25$(OH)_2D$, resulting in transient hypercalcemia. Tumors secreting phosphaturic factors may be small and in obscure locations, making tumor localization challenging. The ^{68}Ga-DOTATATE positron emission tomography (PET)/computed tomography (CT) scan is the most sensitive and specific for localization of these tumors.[87] Other imaging modalities used include octreoscan–single-photon emission CT/CT, (^{18}F)-fluorodeoxyglucose–PET/CT, and 3T magnetic resonance imaging (MRI).[87,88] If the tumor secrets a high amount FGF23, venous sampling may be used for tumor localization.[88] If the tumor cannot be localized or resected, affected individuals, like those with hypophosphatemic rickets, are traditionally treated with combination oral phosphate and calcitriol therapy. The phosphate in oral supplements binds serum calcium, resulting in a transient decrease in serum ionized calcium levels and rise in serum PTH levels.[89] The calcium-sensing receptor agonist cinacalcet has been used to treat patients with tumor-induced osteomalacia or hypophosphatemic rickets as adjunctive therapy to improve hypophosphatemia and prevent hyperparathyroidism.[89,90] The anti-FGF23 antibody has recently been approved as a therapy for tumor-induced osteomalacia but has not been shown to cure the osteomalacia.

KEY POINTS

- Genetic and acquired disorders characterized by high serum levels of fibroblast growth factor 23 lead to hypophosphatemia, and thus rickets and osteomalacia.

Non–FGF23-mediated Hypophosphatemia

A mutation in the phosphate transporter SLC34A3 (NptIIc) is the genetic basis for hereditary hypophosphatemic rickets with hypercalciuria. Hypercalciuria is secondary to increased 1,25$(OH)_2D$, which leads to increased intestinal calcium absorption. This hypercalciuria and low serum FGF23 levels distinguish it from the FGF23-related hypophosphatemic rickets.[91] Similarly, homozygous mutations in SLC34A1 (NptIIa) are associated with hypophosphatemic rickets and Fanconi syndrome.[92]

It is difficult to produce selective deficiency of phosphorus by dietary means alone because most foods contain this element in concentrations that are sufficient to prevent hypophosphatemia and skeletal disease. Hypophosphatemia is seen in both chronic and acute alcohol use disorder.[93,94] Bone densitometry and tetracycline-labeled bone biopsies obtained from patients with chronic alcohol use disorder have revealed changes consistent with mixtures of osteoporosis, osteomalacia, and osteitis fibrosa.[95] Multiple factors, including hypomagnesemia, metabolic alkalosis or acidosis, and renal tubular phosphate wasting, contribute to the hypophosphatemia and skeletal disease.

In individuals with primary hyperparathyroidism who have excessive bone resorption and high rates of bone formation, thick osteoid seams may be seen because the rate of bone formation exceeds that of mineralization. This may present clinically as hungry bone syndrome following surgical cure of hyperparathyroidism. However, frank vitamin D deficiency and osteomalacia may coexist with primary hyperparathyroidism because hyperparathyroidism accelerates the rate of vitamin D catabolism.[28]

MISCELLANEOUS CAUSES OF IMPAIRED MINERALIZATION

Acidosis

The mechanism of bone loss and the mineralization defects seen with acidosis are complex and not completely understood. Based on normal subjects in whom metabolic acidosis is induced, it has been proposed that excess retained hydrogen ion is balanced by bone buffering and loss of skeletally derived calcium in the urine.[96] Because the hypercalciuria and increased bone resorption that accompany most acidotic states do not directly result in osteomalacia, other mechanisms must be invoked to explain the occurrence of clinically significant skeletal mineralization defects. Osteoclasts function optimally at a pH of approximately 6.9 and are inactive at a pH greater than 7.3[97]; thus, chronic acidosis may activate the vacuolar hydrogen ion pumps in osteoclasts. In addition, voltage-gated proton channels in osteoclasts can sense changes in pH.[98] As a consequence, calcium release from bone induced by acidosis may be caused by increased osteoclast activity. In parallel, metabolic acidosis inhibits the function of osteoblasts: in osteoblast culture systems, lowering pH to 6.9 decreases formation of mineralized nodules. Although acidosis has been shown to increase renal phosphate wasting, the role of phosphaturic hormones in this process has not been established.[99]

Rickets and osteomalacia secondary to acidosis are most frequently a complication of inherited distal renal tubular acidosis (RTA).[100] Distal or type I RTA is characterized by impaired secretion of hydrogen ions from the distal nephron, resulting in metabolic acidosis, nephrocalcinosis, hypokalemia, rickets, and osteomalacia. Distal RTA can be attributed to genetic defects; systemic disease including hyperparathyroidism, sickle cell disease, and autoimmune disorders; and drugs including lithium, foscarnet, and amphotericin B. Autosomal dominant distal RTA is caused by mutations in the *SLC4A1* gene that encodes a chloride-bicarbonate exchanger.[101] Rarely, mutations in the *SLC4A1* gene occur in recessive RTA as well.

Autosomal recessive distal RTA can be caused by homozygous mutations in the *ATP6VB1* or *ATP6V0A4* genes, which encode the b1 and a4 subunits of the collecting duct apical proton pump vacuolar H^+/ATPase, respectively, and may be associated with progressive bilateral sensorineural hearing loss.[102] Homozygous mutations in the FOXI1 transcription factor have been shown to result in a similar phenotype.[103] Mutations in WDR72 have also been implicated in the pathogenesis of autosomal recessive distal RTA.[104]

Autosomal dominant RTA is usually diagnosed in adults who present with nephrolithiasis and acidosis. Osteomalacia occurs infrequently. In contrast, autosomal recessive RTA presents in infancy or early childhood and is often accompanied by rickets. Healing of the skeletal disease is achieved by correcting acidosis with sodium bicarbonate (5–10 g/day). The response may be enhanced by the addition of vitamin D or 1,25$(OH)_2D3$. Occasionally, vitamin D toxicity can develop; therefore, patients must be monitored carefully.

Another form of inherited mixed distal/proximal RTA, caused by mutations carbonic anhydrase II (CAII), is associated with osteopetrosis and cerebral calcifications.[105,106] The osteopetrosis is as a result of impaired osteoclast function because CAII deficiency impairs the acidification required for bone resorption. It is of interest that bone marrow transplantation does not affect the acidosis, but reverses the osteopetrosis because osteoclasts are of hematopoietic origin.[107]

Osteomalacia may be a complication of the acidosis produced by ureterosigmoidostomy, a procedure that was formerly used in the treatment of patients with carcinoma of the bladder. Reabsorption of chloride and hydrogen ions from urine in the colon is responsible for the acidosis.

Fanconi Syndrome

Renal Fanconi syndrome is characterized by generalized dysfunction of the proximal renal tubules with loss of amino acids, glucose, phosphate, uric acid, and bicarbonate in the absence of abnormal glomerular function. This results in systemic acidosis, hypophosphatemia, and dehydration, which lead to growth disturbance, rickets, and osteomalacia. Fanconi syndrome may be further classified into that caused by primary renal abnormalities, in which the underlying defect is located within the renal tubular cells, and that because of prerenal disorders, in which metabolic substances from outside the kidney lead to derangements in tubular function. Autosomal dominant Fanconi syndrome, associated with rickets and osteomalacia, is seen with mutations in *EHHADH*, a gene involved in peroxisomal oxidation of fatty acids,[108] and mutations in *HNF4A*, which also cause maturity-onset diabetes of the young 1.[109] Genetic and physical mapping studies have also identified a locus for autosomal dominant Fanconi syndrome at chromosome 15q15.3.[110]

Dent disease is an X-linked renal tubular disorder characterized by low–molecular weight proteinuria, hypercalciuria, nephrocalcinosis, nephrolithiasis, rickets, and renal failure and may include other features of the Fanconi syndrome. The disease results from mutations of the X-linked *CLCN5* gene that encodes a 74–amino acid protein chloride channel with 12 to 13 transmembrane domains.[111] Patients usually present with signs and symptoms of hypokalemia and osteomalacia. Healing of the skeletal disease has occurred with vitamin D, alkali, and potassium treatment.

Fanconi syndrome associated with systemic metabolic processes is seen in Lowe syndrome (also known as Dent-2), a rare X-linked disorder that involves the eyes, kidney, and nervous system and is manifested by congenital cataracts, cognitive impairment, and renal tubular dysfunction. The disorder results from mutations of the *OCRL* gene,[112] which leads to abnormalities in the actin cytoskeleton, impairing tight and adherens junction formation, resulting in tubular dysfunction.[113]

Nephropathic cystinosis is an autosomal recessive lysosomal storage disease characterized by chronic kidney disease and other complications, including the Fanconi syndrome. The disease is caused by mutation of the *CTNS* gene that encodes a transmembrane protein thought to transport cystine out of lysosomes.[114] Cystine accumulates in many tissues, including the kidneys, resulting in tubular dysfunction and subsequently glomerular failure.[115] Fanconi-Bickel syndrome, characterized by renal tubular phosphate wasting and hepatorenal glycogen accumulation, is a result of homozygous mutations in the SLC2A2 (GLUT2) glucose transporter.[116]

Inborn errors of metabolism can also cause Fanconi syndrome, most notably Wilson disease and hereditary tyrosinemia. Wilson disease is an autosomal recessive disease caused by mutations in the *ATP7B* gene that encodes a copper-transporting P-type ATPase that is required for incorporation of this metal into ceruloplasmin. Clinical and pathologic manifestations in Wilson disease are attributed to the toxic effects of copper accumulation in liver, kidney, brain, and cornea. Fanconi syndrome and associated rickets and osteomalacia may predominate or be the presenting feature.[117] Treatment of Wilson disease with D-penicillamine and zinc salts significantly attenuates the phenotype.[118] Hereditary tyrosinemia type I is an autosomal recessive disease caused by mutations in the fumarylacetoacetate hydrolase gene that encodes an enzyme involved in the last step of tyrosine degradation.[119] Tyrosine accumulates in the kidney, peripheral nerves, and liver, which can lead to the development of cirrhosis and hepatoma.

In addition to other rare inherited causes of Fanconi syndrome, the development of Fanconi syndrome in multiple myeloma has been attributed to the toxic effects of monoclonal light chains on the proximal renal tubule.[120] Although extremely rare, osteomalacia may develop in patients with multiple myeloma as a consequence of the renal phosphate wasting and hypophosphatemia. Osteomalacia associated with RTA has also been reported in patients with Sjögren syndrome[121] and in patients with heavy metal poisoning, particularly cadmium[122] and lead.[123] Pharmacologic agents including valproate, tenofovir,[124,125] adefovir,[126] ifosfamide,[127] and outdated tetracycline[128] have been reported to cause Fanconi syndrome as well. The skeletal disease in these settings is attributed to hypophosphatemia and systemic acidosis produced by impaired tubular function.

KEY POINTS

- The kidney is a critical regulator of mineral metabolism. Inadequate renal resorption of phosphate and systemic acidosis can lead to impaired skeletal mineralization in Fanconi syndrome.

Chronic Kidney Disease

Renal osteodystrophy comprises a heterogeneous group of skeletal disorders associated with chronic kidney disease. Before the introduction of hemodialysis, the major accompanying skeletal disorder was a high turnover state because of secondary or tertiary hyperparathyroidism. Following the introduction of hemodialysis and the widespread use of aluminum-containing gels as phosphate binders,[129] two patterns of osteomalacia emerged. In the first type, associated with high-turnover bone disease, increased osteoid accompanied by low circulating calcium and phosphorus levels was seen. The second type, "adynamic" bone disease, was associated with sparse osteoid, a reduction in bone formation rate and normal or slightly elevated calcium and phosphorus levels accompanied by skeletal pain and fractures. Chronic kidney disease mineral and bone disorder is discussed in greater detail in Chapter 61.

Aluminum, Fluoride, and Heavy Metal Intoxication

Aluminum is deposited on the mineralization surface of bone and impairs osteoblastic bone formation.[130] At high concentrations, it can inhibit the vitamin D 1-alpha-hydroxylase.[131]

In high concentrations, fluoride incorporates into the skeleton as calcium fluorhydroxyapatite crystals, increasing apparent bone density on radiographs and bone density exams. Ingestion of water and food high in fluoride, in combination with a low calcium diet, can result in severe skeletal deformities and rickets in children.[132]

Exposure to heavy metals like cadmium, lead, and mercury can also cause phosphaturia, impairing skeletal mineralization.[133]

Hypophosphatasia

Hypophosphatasia is a heritable disorder characterized by mutations in the tissue-nonspecific (bone/liver/kidney) isoenzyme of alkaline phosphatase (TNSALP).[134] Biochemical diagnosis is made with low serum nonspecific alkaline phosphatase levels. Autosomal recessive infantile hypophosphatasia presents with hypercalcemia, renal failure, and increased intracranial pressure. Skeletal features include enlarged cranial sutures, craniostenosis, dental abnormalities, enlarged epiphyses, and prominent costochondral junctions. Serum calcium and phosphorus levels are normal in the less severe forms of hypophosphatasia that present in older children. Genotype-phenotype correlations demonstrate that the extreme heterogeneity in phenotype among patients with hypophosphatasia is because of residual enzyme activity with some mutations.[135] Bone marrow transplants have been performed for life-threatening disease.[136] Currently, enzyme replacement therapy with asfotase alfa is used to treat infantile and juvenile-onset hypophosphatasia.[137] This therapy improves skeletal mineralization, mobility, muscle strength, and pain.[138]

Autosomal dominant hypophosphatasia presents during adult life and is characterized by variable penetrance. Although impaired bone mineralization may be seen, the disorder tends to be mild. The presence of low bone mineral density or fractures in middle-aged patients can often result in a misdiagnosis of osteoporosis. Inappropriate treatment with bisphosphonates can worsen the hypomineralization by impairing residual alkaline phosphatase activity.[139]

The decrease in TNSALP enzyme activity in blood and tissues results in an accumulation of PPi, impairing skeletal mineralization. TNSALP generates Pi locally by cleaving substrates including PPi into Pi to promote mineralization. Because PPi is a potent inhibitor of mineralization, accumulation of PPi in hypophosphatasia inhibits mineralization at sites of bone formation. TNSALP has broad substrate activity; thus, in patients with hypophosphatasia, phosphorylated molecules such as phosphorylethanolamine and phosphorylcholine are excreted in excessive amounts in the urine, and circulating levels of pyridoxal-5'-phosphate, another TNSALP substrate, are markedly elevated.[140,141] Hypophosphatasia is discussed in more detail in Chapter 59.

KEY POINTS

- Presentation of hypophosphatasia in adults can be mistaken for osteoporosis.

Osteopetrosis

Rickets also may occur in infantile osteopetrosis,[142] which is most commonly caused by mutation of the *TCIRG1* gene that encodes the a3 subunit of V0 complex of the V-ATPase proton pump, expressed primarily in osteoclasts and gastric epithelium. Although a low local calcium × phosphate product owing to impaired osteoclastic bone resorption has been implicated in the development of rickets, affected children may have impaired intestinal calcium absorption, leading to secondary hyperparathyroidism and hypophosphatemia, which leads to rickets. Mutations in SNX10, which account for a small percentage of affected individuals,[143] also impair both osteoclast activity and intestinal calcium absorption.[144] Osteopetrosis is discussed in greater detail in Chapter 59.

Fibrogenesis Imperfecta Ossium

Fibrogenesis imperfecta ossium is an extremely rare disorder characterized by progressively disabling skeletal pain, bone fragility, and skeletal deformities, seen primarily in middle-aged individuals. Serum calcium and phosphorus are normal, but alkaline phosphatase activity is increased. The radiographic abnormalities include thickened and amorphous-appearing trabeculae with spotty increases in density, reduced cortical thicknesses, and occasional pseudofractures. The abnormal “fishnet” trabecular pattern seen on conventional radiographs, combined with MRI of low–signal intensity bone marrow on T1- and T2-weighted images, is helpful in diagnosis.[145] The primary defect is thought to be abnormal collagen synthesis, leading to defective fibrillar organization and impaired mineralization of bone. The increased osteoid lacks the lamellar structure and typical birefringence that is characteristic of other osteomalacic states. Electron microscopy reveals that immature small-diameter collagen fibers are arranged in loops, as well as dense “whorls.” The abnormal bone matrix leads to increased bone fragility and fractures. Of interest, paraproteinemia is seen in approximately one-third of individuals.[146] No uniformly effective treatment has been found.

TREATMENT

Vitamin D deficiency impairs intestinal calcium absorption, resulting in secondary hyperparathyroidism, which leads to hypophosphatemia, rickets in the growing skeleton, and osteomalacia. Guidelines from the US Institute of Medicine suggest that, for over 97.5% of the population, the recommended vitamin D intake from age 1 to 70 years should be 600 IU, with an increase to 800 IU for those over 70 years of age.[39] Although this recommended intake is an increase from previous guidelines, there remains controversy as to whether or not these recommendations are sufficient.[40] Higher doses are required in those with obesity, intestinal disease, or short gut syndrome, as well as to treat rickets and osteomalacia. The goals for treatment of patients with osteomalacia or rickets are to correct hypocalcemia if it exists, to correct secondary hyperparathyroidism, and to correct hypophosphatemia. Careful monitoring may be required to prevent vitamin D intoxication, which leads to hypercalcemia, hypercalciuria, nephrocalcinosis, and nephrolithiasis. The dose of vitamin D must be tailored to a given individual, and calcium intake must be optimized to prevent secondary hyperparathyroidism and the resultant renal phosphate losses. Vitamin D is available in 400-IU (10-mcg) to 50,000-IU (1.25-mg) capsules for oral administration and liquid form (250–2000 IU or 6.25–50 mcg per drop) for oral administration. In patients with malabsorption, vitamin D can also be administered sublingually.[147] The advantages of vitamin D are its low cost and the endogenous regulation of activation to 1,25$(OH)_2$D. The disadvantages are that an optimal therapeutic response may take several weeks, and, in cases of toxicity, its biologic activity persists for several weeks after it is discontinued, because it has a long half-life.

1,25$(OH)_2$D3 (calcitriol) is available in 0.25-mcg and 0.50-mcg capsules, as well as in oral and intravenous preparations containing 1 to 2 mcg/mL. The advantages of calcitriol are its rapid onset of action and short half-life (>6 hours). However, hypercalcemia may occur even long after an apparently stable maintenance dose has been established. Because hypercalcemia during treatment with 1,25$(OH)_2$D3 is unpredictable, patients must be monitored closely. 1-alpha-(OH)D2 is approved for use in the treatment of secondary hyperparathyroidism in patients with chronic kidney disease.[148] It is marketed in 0.5- to 2.5-mcg capsules and solutions containing 2 mcg/mL. After administration, it undergoes conversion to 1,25(OH)2D2.

1-alpha hydroxylated metabolites are indicated for the treatment of diseases in which renal 1-alpha hydroxylation is diminished or impaired (hypoparathyroidism, pseudohypoparathyroidism, chronic kidney disease, pseudovitamin D–deficiency rickets, tumor-induced osteomalacia, and other FGF23-dependent hypophosphatemic disorders), or when there is resistance to 1,25$(OH)_2$D at the cellular level. Osteomalacia or rickets associated with hereditary vitamin D–resistant rickets, characterized by mutation of the VDR, occasionally responds to high-dose 1,25$(OH)_2$D3. However, long-term parenteral administration of calcium may be required to normalize PTH and phosphorus levels and heal skeletal disease[62] (also see Chapter 58).

In children with nutritional rickets caused by calcium deficiency, increasing the calcium intake to between 800 and 1500 mg per day was shown to heal the skeletal disease.[44,149] Several protocols are recommended for treatment of vitamin D–deficient skeletal disease. Daily oral doses of vitamin D, 50 mcg (2000 IU) for 3 to 12 weeks, when accompanied by an adequate calcium intake, may result in a biochemical response in classic rickets and osteomalacia resulting from simple vitamin D deficiency. However, this regimen is insufficient to treat rickets associated with FGF23 excess, malabsorption, and/or drugs that increase vitamin D metabolism. Alternative recommendations include treatment with larger doses of vitamin D (50,000–500,000 IU given in divided doses over 1–3 months), followed by maintenance doses of 600 IU per day, to achieve a normal circulating 25(OH)D level. Patients on chronic anticonvulsant therapy often require higher maintenance doses.[150] Administration of vitamin D to patients with vitamin D–deficiency rickets or osteomalacia invariably results in healing of the skeletal disease when calcium intake is adequate. Biochemical responses include normalization of serum calcium and phosphorus levels. There is often an initial increase in serum alkaline phosphatase, indicative of healing of the osteomalacic lesions, followed by a gradual fall in both alkaline phosphatase and PTH.

Hypercalcemia and hypercalciuria are potential complications of treatment with vitamin D. Hypercalcemia resulting from vitamin D intoxication is caused by a combination of increased intestinal absorption of calcium and increased osteoclastic bone resorption. Impaired kidney function may develop in the setting of hypercalcemia, hypercalciuria, nephrocalcinosis, and nephrolithiasis. Patients may exhibit classic hypercalcemic symptoms, including anorexia, nausea, vomiting, weight loss, headache, constipation, polyuria, polydipsia, and altered mental status. Aggressive hydration with intravenous fluids is indicated, but because elevated serum 25(OH)D levels persist, glucocorticoids may be required to lower serum calcium levels, as may intravenous bisphosphonates to inhibit osteoclastic bone resorption. Patients on long-term treatment with pharmacologic vitamin D metabolites should be carefully monitored with measurement of serum and urinary calcium and serum creatinine. This is critical because there is no way of predicting when or in whom vitamin D intoxication will develop and/or recur.

Treatment of hypophosphatemia not associated with vitamin D deficiency and secondary hyperparathyroidism involves oral supplementation with neutral phosphate salts, including sodium or potassium salts. The rise in serum phosphorus after a single oral dose of phosphate is transient; therefore, phosphate supplements must be administered at frequent intervals. The precise dose must be individualized and is dependent on the degree of renal loss. In individuals with severe renal losses, as much as 1000 mg of elemental phosphorus is required every 4 to 6 hours to promote sustained elevations of serum phosphorus values. Treatment response cannot be assessed with a single fasting determination of serum phosphorus; multiple measurements of serum values at various times after each dose are required. Emptying the capsules and dissolving the contents into water or other liquid may improve intestinal absorption. When therapy is initiated, most patients experience some degree of gastrointestinal distress, including cramps and diarrhea. Therefore, initial doses should be low, and increments should be gradually introduced as tolerated.

Oral phosphate can bind calcium, leading to impaired intestinal calcium absorption and a rise in PTH levels, which further increases renal phosphate losses. Chronic phosphate therapy can lead to secondary hyperparathyroidism and parathyroid autonomy; therefore, chronic oral phosphate therapy should be combined with $1,25(OH)_2D$ to prevent the rise in PTH and promote the intestinal absorption of calcium and phosphate. The use of a calcimimetic to decrease PTH secretion has been reported as adjuvant treatment for hereditary and acquired hypophosphatemia associated with impaired renal tubular reabsorption.[89,90] In the various types of prerenal Fanconi syndrome, the specific treatment of the underlying disorder often results in disappearance of or improvement in the renal tubular dysfunction, thus allowing healing of the associated skeletal disease. In several of these disorders, improvement in the skeletal disease can also be achieved by correcting the acidosis and hypophosphatemia (i.e., supplementation with alkali, oral phosphate, and vitamin D). Targeted therapies for other rare diseases are discussed in the relevant sections.

KEY POINTS

- Adequate vitamin D levels are needed to maintain calcium and phosphate absorption and prevent secondary hyperparathyroidism.
- Treatment of hypophosphatemia with phosphate supplementation should always be accompanied with concomitant administration of calcitriol to promote phosphate absorption and prevent secondary hypoparathyroidism.

CONCLUSION

Optimal skeletal mineralization and strength are dependent on normal circulating levels of calcium and phosphate, as well as normal pH. Although expansion of the growth plate in rickets is because of impaired phosphate-mediated hypertrophic chondrocyte apoptosis in the setting of hypophosphatemia, the regulation of mineral deposition into bone is more complex and less well understood. Many of the factors regulating mineralization have been elucidated by identifying the molecular basis for rare genetic disorders, complemented by detailed analyses in naturally occurring or genetically engineered murine homologs. These studies have also identified a novel role for osteocytes in the regulation of bone resorption and are expected to further elucidate the molecular mechanisms regulating mineralization.

REFERENCES

1. Sabbagh Y, Carpenter TO, Demay MB. Hypophosphatemia leads to rickets by impairing caspase-mediated apoptosis of hypertrophic chondrocytes. *Proc Natl Acad Sci U S A*. 2005;102:9637–9642.
2. Christakos S, Dhawan P, Verstuyf A, et al. Vitamin D. Metabolism, molecular mechanism of action, and pleiotropic effects. *Physiol Rev*. 2016;96:365–408.
3. Farrow EG, White KE. Recent advances in renal phosphate handling. *Nat Rev Nephrol*. 2010;6:207–217.
4. Millan JL. The role of phosphatases in the initiation of skeletal mineralization. *Calcif Tissue Int*. 2012;93:299–306.
5. Yadav MC, Simao AM, Narisawa S, et al. Loss of skeletal mineralization by the simultaneous ablation of PHOSPHO1 and alkaline phosphatase function: a unified model of the mechanisms of initiation of skeletal calcification. *J Bone Miner Res*. 2011;26:286–297.
6. Wu Y, Ackerman JL, Strawich ES, et al. Phosphate ions in bone: identification of a calcium-organic phosphate complex by 31P solid-state NMR spectroscopy at early stages of mineralization. *Calcif Tissue Int*. 2003;72:610–626.
7. Hessle L, Johnson KA, Anderson HC, et al. Tissue-nonspecific alkaline phosphatase and plasma cell membrane glycoprotein-1 are central antagonistic regulators of bone mineralization. *Proc Natl Acad Sci U S A*. 2002;99:9445–9449.
8. Narisawa S, Yadav MC, Millan JL. In vivo overexpression of tissue-nonspecific alkaline phosphatase increases skeletal mineralization and affects the phosphorylation status of osteopontin. *J Bone Miner Res*. 2013;28:1587–1598.
9. Huitema LF, Apschner A, Logister I, et al. Entpd5 is essential for skeletal mineralization and regulates phosphate homeostasis in zebrafish. *Proc Natl Acad Sci U S A*. 2012;109:21372–21377.
10. Ho AM, Johnson MD, Kingsley DM. Role of the mouse ank gene in control of tissue calcification and arthritis. *Science*. 2000;289:265–270.
11. Murshed M, Harmey D, Millan JL, et al. Unique coexpression in osteoblasts of broadly expressed genes accounts for the spatial restriction of ECM mineralization to bone. *Genes Dev*. 2005;19:1093–1104.
12. Lee WR. Bone formation in Paget's disease. A quantitative microscopic study using tetracycline markers. *J Bone Joint Surg Br*. 1967;49:146–153.
13. Frost HM. Tetracycline-based histological analysis of bone remodeling. *Calcif Tissue Res*. 1969;3:211–237.
14. Qing H, Ardeshirpour L, Pajevic PD, et al. Demonstration of osteocytic perilacunar/canalicular remodeling in mice during lactation. *J Bone Miner Res*. 2012;27:1018–1029.
15. Marie PJ, Glorieux FH. Relation between hypomineralized periosteocytic lesions and bone mineralization in vitamin D-resistant rickets. *Calcif Tissue Int*. 1983;35:443–448.
16. Tokarz D, Martins JS, Petit ET, et al. Hormonal regulation of osteocyte perilacunar and canalicular remodeling in the hyp mouse model of X-linked hypophosphatemia. *J Bone Miner Res*. 2018;33:499–509.
17. Barnes MJ, Constable BJ, Morton LF, et al. Bone collagen metabolism in vitamin D deficiency. *Biochem J*. 1973;132:113–115.

18. Dent CE. Rickets (and osteomalacia), nutritional and metabolic (1919-69). *Proc R Soc Med*. 1970;63:401–408.
19. Smith R, Newman RJ, Radda GK, et al. Hypophosphataemic osteomalacia and myopathy: studies with nuclear magnetic resonance spectroscopy. *Clin Sci (Lond)*. 1984;67:505–509.
20. Ball J, Garner A. Mineralisation of woven bone in osteomalacia. *J Pathol Bacteriol*. 1966;91:563–567.
21. Steinbach HL, Noetzli M. Roentgen appearance of the skeleton in osteomalacia and rickets. *Am J Roentgenol Radium Ther Nucl Med*. 1964;91:955–972.
22. Carpenter TO, Imel EA, Holm IA, et al. A clinician's guide to X-linked hypophosphatemia. *J Bone Miner Res*. 2011;26:1381–1388.
23. Arneil GC, Crosbie JC. Infantile rickets returns to Glasgow. *Lancet*. 1963;2:423–425.
24. Benson PF, Stroud CE, Mitchell NJ, et al. Rickets in immigrant children in London. *Br Med J*. 1963;1:1054–1056.
25. Lawson M, Thomas M. Vitamin D concentrations in Asian children aged 2 years living in England: population survey. *BMJ*. 1999;318:28.
26. Dent CE, Round JM, Rowe DJ, et al. Effect of chapattis and ultraviolet irradiation on nutritional rickets in an Indian immigrant. *Lancet*. 1973;1:1282–1284.
27. Loomis WF. Skin-pigment regulation of vitamin-D biosynthesis in man. *Science*. 1967;157:501–506.
28. Clements MR, Davies M, Fraser DR, et al. Metabolic inactivation of vitamin D is enhanced in primary hyperparathyroidism. *Clin Sci (Lond)*. 1987;73:659–664.
29. Thacher TD, Fischer PR, Tebben PJ, et al. Increasing incidence of nutritional rickets: a population-based study in Olmsted County, Minnesota. *Mayo Clin Proc*. 2013;88:176–183.
30. Heckmatt JZ, Peacock M, Davies AE, et al. Plasma 25-hydroxyvitamin D in pregnant Asian women and their babies. *Lancet*. 1979;2:546–548.
31. Specker BL, Tsang RC, Hollis BW. Effect of race and diet on human-milk vitamin D and 25-hydroxyvitamin D. *Am J Dis Child*. 1985;139:1134–1137.
32. Stamp TC, Round JM. Seasonal changes in human plasma levels of 25-hydroxyvitamin D. *Nature*. 1974;247:563–565.
33. Dawson-Hughes B, Harris SS, Dallal GE. Plasma calcidiol, season, and serum parathyroid hormone concentrations in healthy elderly men and women. *Am J Clin Nutr*. 1997;65:67–71.
34. Harris SS, Dawson-Hughes B. Seasonal changes in plasma 25-hydroxyvitamin D concentrations of young American black and white women. *Am J Clin Nutr*. 1998;67:1232–1236.
35. Finch PJ, Ang L, Colston KW, et al. Blunted seasonal variation in serum 25-hydroxy vitamin D and increased risk of osteomalacia in vegetarian London Asians. *Eur J Clin Nutr*. 1992;46:509–515.
36. Henderson JB, Dunnigan MG, McIntosh WB, et al. Asian osteomalacia is determined by dietary factors when exposure to ultraviolet radiation is restricted: a risk factor model. *Q J Med*. 1990;76:923–933.
37. MacLaughlin J, Holick MF. Aging decreases the capacity of human skin to produce vitamin D3. *J Clin Invest*. 1985;76:1536–1538.
38. Ebeling PR, Sandgren ME, DiMagno EP, et al. Evidence of an age-related decrease in intestinal responsiveness to vitamin D: relationship between serum 1,25-dihydroxyvitamin D3 and intestinal vitamin D receptor concentrations in normal women. *J Clin Endocrinol Metab*. 1992;75:176–182.
39. Ross AC, Manson JE, Abrams SA, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab*. 2011;96:53–58.
40. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1911–1930.
41. Thacher TD, Fischer PR, Isichei CO, et al. Prevention of nutritional rickets in Nigerian children with dietary calcium supplementation. *Bone*. 2012;50:1074–1080.
42. Adams JS, Clemens TL, Parrish JA, et al. Vitamin-D synthesis and metabolism after ultraviolet irradiation of normal and vitamin-D-deficient subjects. *N Engl J Med*. 1982;306:722–725.
43. Pettifor JM, Ross P, Wang J, et al. Rickets in children of rural origin in South Africa: is low dietary calcium a factor? *J Pediatr*. 1978;92:320–324.
44. Thacher TD, Fischer PR, Pettifor JM, et al. A comparison of calcium, vitamin D, or both for nutritional rickets in Nigerian children. *N Engl J Med*. 1999;341:563–568.
45. DeLucia MC, Mitnick ME, Carpenter TO. Nutritional rickets with normal circulating 25-hydroxyvitamin D: a call for reexamining the role of dietary calcium intake in North American infants. *J Clin Endocrinol Metab*. 2003;88:3539–3545.
46. Daum F, Rosen JF, Roginsky M, et al. 25-Hydroxycholecalciferol inthe management of rickets associated with extrahepatic biliary atresia. *J Pediatr*. 1976;88:1041–1043.
47. Heubi JE, Tsang RC, Steichen JJ, et al. 1,25-Dihydroxyvitamin D3 in childhood hepatic osteodystrophy. *J Pediatr*. 1979;94:977–982.
48. Klein GL, Soriano H, Shulman RJ, et al. Hepatic osteodystrophy in chronic cholestasis: evidence for a multifactorial etiology. *Pediatr Transplant*. 2002;6:136–140.
49. Botella-Carretero JI, Alvarez-Blasco F, Villafruela JJ, et al. Vitamin D deficiency is associated with the metabolic syndrome in morbid obesity. *Clin Nutr*. 2007;26:573–580.
50. Charles P, Mosekilde L, Sondergard K, et al. Treatment with high-dose oral vitamin D2 in patients with jejunoileal bypass for morbid obesity. Effects on calcium and magnesium metabolism, vitamin D metabolites, and faecal lag time. *Scand J Gastroenterol*. 1984;19:1031–1038.
51. Crowley LV, Seay J, Mullin G. Late effects of gastric bypass for obesity. *Am J Gastroenterol*. 1984;79:850–860.
52. Parfitt AM, Podenphant J, Villanueva AR, et al. Metabolic bone disease with and without osteomalacia after intestinal bypass surgery: a bone histomorphometric study. *Bone*. 1985;6:211–220.
53. Lespessailles E, Toumi H. Vitamin D alteration associated with obesity and bariatric surgery. *Exp Biol Med (Maywood)*. 2017;242:1086–1094.
54. Emile Sameh Hany EH. Nutritional deficiency after sleeve gastrectomy: a comprehensive literature review. *Eur Med J*. 2017;6:99–105.
55. Mihmanli M, Isil RG, Isil CT, et al. Effects of laparoscopic sleeve gastrectomy on parathyroid hormone, vitamin D, calcium, phosphorus, and albumin levels. *Obes Surg*. 2017;27:3149–3155.
56. Bisballe S, Eriksen EF, Melsen F, et al. Osteopenia and osteomalacia after gastrectomy: interrelations between biochemical markers of bone remodelling, vitamin D metabolites, and bone histomorphometry. *Gut*. 1991;32:1303–1307.
57. Cheng JB, Levine MA, Bell NH, et al. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci U S A*. 2004;101:7711–7715.
58. Fu G, Lin D, Zang M, et al. Cloning of human 25-hydroxyvitamin D-1-a-hydroxylase and mutations causing vitamin D-dependent rickets type 1. *Mol Endocrinol*. 1997:1961–1970.
59. Kitnaka S, Takeyama K-I, Muryama A, et al. Inactivating mutations in the 25-hydroxyvitamin D, 1a-hydroxylase gene in patients with pseudo-vitamin d-deficiency rickets. *N Engl J Med*. 1998;338:653–661.
60. St-Arnaud R, Messerlian S, Moir J, et al. The 25-hydroxyvitamin D 1-alphahydroxylase gene maps to the pseudovitamin D-deficiency rickets (PDDR) disease locus. *J Bone Miner Res*. 1997;12:1552–1559.
61. Hughes M, Malloy P, Kieback D, et al. Human vitamin D receptor mutations: identification of molecular defects in hypocalcemic vitamin D resistant rickets. *Adv Exp Med Biol*. 1989;255:491–503.
62. Balsan S, Garabedian M, Larchet M, et al. Long-term nocturnal calcium infusions can cure rickets and promote normal mineralization in hereditary resistance to 1,25-dihydroxyvitamin D. *J Clin Invest*. 1986;77:1661–1667.
63. Roizen JD, Li D, O'Lear L, et al. CYP3A4 mutation causes vitamin D-dependent rickets type 3. *J Clin Invest*. 2018;128:1913–1918.
64. Albright F. Hypoparathyroidism as a cause of osteomalacia. *J Clin Endocrinol Metab*. 1956;16:419–425.
65. Drezner MK, Neelon FA, Jowsey J, et al. Hypoparathyroidism: a possible cause of osteomalacia. *J Clin Endocrinol Metab*. 1977;45:114–122.
66. Epstein S, Meunier PJ, Lambert PW, et al. 1 alpha,25-dihydroxyvitamin D3 corrects osteomalacia in hypoparathyroidism and pseudohypoparathyroidism. *Acta Endocrinol (Copenh)*. 1983;103:241–247.

67. Wilson JD, Hadden DR. Pseudohypoparathyroidism presenting with rickets. *J Clin Endocrinol Metab*. 1980;51:1184–1189.
68. Albright F, Butler AM, Bloomberg E. Rickets resistant to vitamin D therapy. *J Clin Dis Child*. 1937;54:529–547.
69. Jonsson KB, Zahradnik R, Larsson T, et al. Fibroblast growth factor 23 in oncogenic osteomalacia and X-linked hypophosphatemia. *N Engl J Med*. 2003;348:1656–1663.
70. Carpenter TO, Imel EA, Ruppe MD, et al. Randomized trial of the anti-FGF23 antibody KRN23 in X-linked hypophosphatemia. *J Clin Invest*. 2014;124:1587–1597.
71. Whyte MP, Carpenter TO, Gottesman GS, et al. Efficacy and safety of burosumab in children aged 1-4 years with X-linked hypophosphataemia: a multicentre, open-label, phase 2 trial. *Lancet Diabetes Endocrinol*. 2019;7:189–199.
72. Imel EA, Glorieux FH, Whyte MP, et al. Burosumab versus conventional therapy in children with X-linked hypophosphataemia: a randomised, active-controlled, open-label, phase 3 trial. *Lancet*. 2019;393:2416–2427.
73. A gene (PEX) with homologies to endopeptidases is mutated in patients with X-linked hypophosphatemic rickets. The HYP Consortium. *Nat Genet*. 1995;11:130–136.
74. Sabbagh Y, Boileau G, Campos M, et al. Structure and function of disease-causing missense mutations in the PHEX gene. *J Clin Endocrinol Metab*. 2003;88:2213–2222.
75. White KE, Carn G, Lorenz-Depiereux B, et al. Autosomal-dominant hypophosphatemic rickets (ADHR) mutations stabilize FGF-23. *Kidney Int*. 2001;60:2079–2086.
76. Farrow EG, Yu X, Summers LJ, et al. Iron deficiency drives an autosomal dominant hypophosphatemic rickets (ADHR) phenotype in fibroblast growth factor-23 (Fgf23) knock-in mice. *Proc Natl Acad Sci U S A*. 2011;108:E1146–E1155.
77. Imel EA, Liu Z, Coffman M, et al. Oral iron replacement normalizes fibroblast growth factor 23 in iron-deficient patients with autosomal dominant hypophosphatemic rickets. *J Bone Miner Res*. 2020;35:231–238.
78. Lorenz-Depiereux B, Bastepe M, Benet-Pages A, et al. DMP1 mutations in autosomal recessive hypophosphatemia implicate a bone matrix protein in the regulation of phosphate homeostasis. *Nat Genet*. 2006;38:1248–1250.
79. Lorenz-Depiereux B, Schnabel D, Tiosano D, et al. Loss-of-function ENPP1 mutations cause both generalized arterial calcification of infancy and autosomal-recessive hypophosphatemic rickets. *Am J Hum Genet*. 2010;86:267–272.
80. Rafaelsen SH, Raeder H, Fagerheim AK, et al. Exome sequencing reveals FAM20c mutations associated with fibroblast growth factor 23-related hypophosphatemia, dental anomalies, and ectopic calcification. *J Bone Miner Res*. 2013;28:1378–1385.
81. Dent CE, Gertner JM. Hypophosphataemic osteomalacia in fibrous dysplasia. *Q J Med*. 1976;45:411–420.
82. Imel EA, Econs MJ. Fibrous dysplasia, phosphate wasting and fibroblast growth factor 23. *Pediatr Endocrinol Rev*. 2007;4:434–439.
83. Prader A, Illig R, Uehlinger E, et al. Rickets following bone tumor. *Helv Paediatr Acta*. 1959;14:554–565.
84. Nuovo MA, Dorfman HD, Sun CC, et al. Tumor-induced osteomalacia and rickets. *Am J Surg Pathol*. 1989;13:588–599.
85. Shimada T, Mizutani S, Muto T, et al. Cloning and characterization of FGF23 as a causative factor of tumor-induced osteomalacia. *Proc Natl Acad Sci U S A*. 2001;98:6500–6505.
86. Lee JC, Su SY, Changou CA, et al. Characterization of FN1-FGFR1 and novel FN1-FGF1 fusion genes in a large series of phosphaturic mesenchymal tumors. *Mod Pathol*. 2016;29:1335–1346.
87. El-Maouche D, Sadowski SM, Papadakis GZ, et al. (68)Ga-DOTATATE for tumor localization in tumor-induced osteomalacia. *J Clin Endocrinol Metab*. 2016;101:3575–3581.
88. Kawai S, Ariyasu H, Furukawa Y, et al. Effective localization in tumor-induced osteomalacia using (68)Ga-DOTATOC-PET/CT, venous sampling and 3T-MRI. *Endocrinol Diabetes Metab Case Rep*. 2017;2017:16. 0005.
89. Alon US, Levy-Olomucki R, Moore WV, et al. Calcimimetics as an adjuvant treatment for familial hypophosphatemic rickets. *Clin J Am Soc Nephrol*. 2008;3:658–664.
90. Geller JL, Khosravi A, Kelly MH, et al. Cinacalcet in the management of tumor-induced osteomalacia. *J Bone Miner Res*. 2007;22:931–937.
91. Bergwitz C, Roslin NM, Tieder M, et al. SLC34A3 mutations in patients with hereditary hypophosphatemic rickets with hypercalciuria predict a key role for the sodium-phosphate cotransporter NaPi-IIc in maintaining phosphate homeostasis. *Am J Hum Genet*. 2006;78:179–192.
92. Magen D, Berger L, Coady MJ, et al. A loss-of-function mutation in NaPi-IIa and renal Fanconi's syndrome. *N Engl J Med*. 2010;362:1102–1109.
93. Stein JH, Smith WO, Ginn HE. Hypophosphatemia in acute alcoholism. *Am J Med Sci*. 1966;252:78–83.
94. Territo MC, Tanaka KR. Hypophosphatemia in chronic alcoholism. *Arch Intern Med*. 1974;134:445–447.
95. de Vernejoul MC, Bielakoff J, Herve M, et al. Evidence for defective osteoblastic function. A role for alcohol and tobacco consumption in osteoporosis in middle-aged men. *Clin Orthop Relat Res*. 1983:107–115.
96. Lemann Jr J, Litzow JR, Lennon EJ. The effects of chronic acid loads in normal man: further evidence for the participation of bone mineral in the defense against chronic metabolic acidosis. *J Clin Invest*. 1966;45:1608–1614.
97. Arnett T. Regulation of bone cell function by acid-base balance. *Proc Nutr Soc*. 2003;62:511–520.
98. Mori H, Sakai H, Morihata H, et al. Regulatory mechanisms and physiological relevance of a voltage-gated H+ channel in murine osteoclasts: phorbol myristate acetate induces cell acidosis and the channel activation. *J Bone Miner Res*. 2003;18:2069–2076.
99. Khairallah P, Isakova T, Asplin J, et al. Acid load and phosphorus homeostasis in CKD. *Am J Kidney Dis*. 2017;70:541–550.
100. Batlle D, Haque SK. Genetic causes and mechanisms of distal renal tubular acidosis. *Nephrol Dial Transplant*. 2012;27:3691–3704.
101. Devonald MA, Smith AN, Poon JP, et al. Non-polarized targeting of AE1 causes autosomal dominant distal renal tubular acidosis. *Nat Genet*. 2003;33:125–127.
102. Smith AN, Borthwick KJ, Karet FE. Molecular cloning and characterization of novel tissue-specific isoforms of the human vacuolar H(+)-ATPase C, G and d subunits, and their evaluation in autosomal recessive distal renal tubular acidosis. *Gene*. 2002;297:169–177.
103. Enerback S, Nilsson D, Edwards N, et al. Acidosis and deafness in patients with recessive mutations in FOXI1. *J Am Soc Nephrol*. 2018;29:1041–1048.
104. Rungroj N, Nettuwakul C, Sawasdee N, et al. Distal renal tubular acidosis caused by tryptophan-aspartate repeat domain 72 (WDR72) mutations. *Clin Genet*. 2018;94:409–418.
105. Sly WS, Hewett-Emmett D, Whyte MP, et al. Carbonic anhydrase II deficiency identified as the primary defect in the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification. *Proc Natl Acad Sci U S A*. 1983;80:2752–2756.
106. Whyte MP, Murphy WA, Fallon MD, et al. Osteopetrosis, renal tubular acidosis and basal ganglia calcification in three sisters. *Am J Med*. 1980;69:64–74.
107. McMahon C, Will A, Hu P, et al. Bone marrow transplantation corrects osteopetrosis in the carbonic anhydrase II deficiency syndrome. *Blood*. 2001;97:1947–1950.
108. Klootwijk ED, Reichold M, Helip-Wooley A, et al. Mistargeting of peroxisomal EHHADH and inherited renal Fanconi's syndrome. *N Engl J Med*. 2014;370:129–138.
109. Hamilton AJ, Bingham C, McDonald TJ, et al. The HNF4A R76W mutation causes atypical dominant Fanconi syndrome in addition to a beta cell phenotype. *J Med Genet*. 2014;51:165–169.
110. Lichter-Konecki U, Broman KW, Blau EB, et al. Genetic and physical mapping of the locus for autosomal dominant renal Fanconi syndrome, on chromosome 15q15.3. *Am J Hum Genet*. 2001;68:264–268.
111. Cox JP, Yamamoto K, Christie PT, et al. Renal chloride channel, CLCN5, mutations in Dent's disease. *J Bone Miner Res*. 1999;14:1536–1542.
112. Monnier N, Satre V, Lerouge E, et al. OCRL1 mutation analysis in French Lowe syndrome patients: implications for molecular diagnosis strategy and genetic counseling. *Hum Mutat*. 2000;16:157–165.
113. Zhang X, Hartz PA, Philip E, et al. Cell lines from kidney proximal tubules of a patient with Lowe syndrome lack OCRL inositol polyphosphate 5-phosphatase and accumulate phosphatidylinositol 4,5-bisphosphate. *J Biol Chem*. 1998;273:1574–1582.

114. Anikster Y, Shotelersuk V, Gahl WA. CTNS mutations in patients with cystinosis. *Hum Mutat*. 1999;14:454–458.
115. Nesterova G, Gahl WA. Cystinosis: the evolution of a treatable disease. *Pediatr Nephrol*. 2013;28:51–59.
116. Santer R, Groth S, Kinner M, et al. The mutation spectrum of the facilitative glucose transporter gene SLC2A2 (GLUT2) in patients with Fanconi-Bickel syndrome. *Hum Genet*. 2002;110:21–29.
117. Morgan HG, Stewart WK, Lowe KG, et al. Wilson's disease and the Fanconi syndrome. *Q J Med*. 1962;31:361–384.
118. Chang H, Xu A, Chen Z, et al. Long-term effects of a combination of D-penicillamine and zinc salts in the treatment of Wilson's disease in children. *Exp Ther Med*. 2013;5:1129–1132.
119. Bergeron A, D'Astous M, Timm DE, et al. Structural and functional analysis of missense mutations in fumarylacetoacetate hydrolase, the gene deficient in hereditary tyrosinemia type 1. *J Biol Chem*. 2001;276:15225–152231.
120. El Hamel C, Aldigier JC, Oblet C, et al. Specific impairment of proximal tubular cell proliferation by a monoclonal kappa light chain responsible for Fanconi syndrome. *Nephrol Dial Transplant*. 2012;27:4368–4377.
121. Monte Neto JT, Sesso R, Kirsztajn GM, et al. Osteomalacia secondary to renal tubular acidosis in a patient with primary Sjogren's syndrome. *Clin Exp Rheumatol*. 1991;9:625–627.
122. Adams RG, Harrison JF, Scott P. The development of cadmium-induced proteinuria, impaired renal function, and osteomalacia in alkaline battery workers. *Q J Med*. 1969;38:425–443.
123. Chisolm Jr JJ. Aminoaciduria as a manifestation of renal tubular injury in lead intoxication and a comparison with patterns of aminoaciduria seen in other diseases. *J Pediatr*. 1962;60:1–17.
124. Haverkort ME, van der Spek BW, Lips P, et al. Tenofovir-induced Fanconi syndrome and osteomalacia in two HIV-infected patients: role of intracellular tenofovir diphosphate levels and review of the literature. *Scand J Infect Dis*. 2011;43:821–826.
125. Wanner DP, Tyndall A, Walker UA. Tenofovir-induced osteomalacia. *Clin Exp Rheumatol*. 2009;27:1001–1003.
126. Kim DH, Lee SW, Kweon YO, et al. Bone scintigraphic findings of hypophosphatemic osteomalacia associated with adefovir therapy: involvement of lower extremity skeletons. *Clin Nucl Med*. 2013;38:378–380.
127. Ingemi AI, Bota VM, Peguero A, et al. Fanconi's syndrome and nephrogenic diabetes insipidus in an adult treated with ifosfamide. *Pharmacotherapy*. 2012;32:e12–e16.
128. Gross JM. Fanconi syndrome (adult type) developing secondary to the ingestion of outdated tetracycline. *Ann Intern Med*. 1963;58:523–528.
129. Coladonato JA. Control of hyperphosphatemia among patients with ESRD. *J Am Soc Nephrol*. 2005;16:S107–S114.
130. Hercz G, Andress DL, Nebeker HG, et al. Reversal of aluminum-related bone disease after substituting calcium carbonate for aluminum hydroxide. *Am J Kidney Dis*. 1988;11:70–75.
131. Henry HL, Norman AW. Interactions between aluminum and the actions and metabolism of vitamin D3 in the chick. *Calcif Tissue Int*. 1985;37:484–490.
132. Teotia M, Teotia SP, Singh KP. Endemic chronic fluoride toxicity and dietary calcium deficiency interaction syndromes of metabolic bone disease and deformities in India: year 2000. *Indian J Pediatr*. 1998;65:371–381.
133. Rodriguez J, Mandalunis PM. A review of metal exposure and its effects on bone health. *J Toxicol*. 2018;2018:4854152.
134. Whyte MP. Hypophosphatasia. In: Scriver CR, Beaudet AL, Sly WS, et al., eds. *The Metabolic and Molecular Bases of Inherited Disease*. 8th ed. New York: McGraw-Hill; 2001:5313–5329.
135. Henthorn PS, Raducha M, Fedde KN, et al. Different missense mutations at the tissue-nonspecific alkaline phosphatase gene locus in autosomal recessively inherited forms of mild and severe hypophosphatasia. *Proc Natl Acad Sci U S A*. 1992;89:9924–9928.
136. Whyte MP, Kurtzberg J, McAlister WH, et al. Marrow cell transplantation for infantile hypophosphatasia. *J Bone Miner Res*. 2003;18:624–636.
137. Whyte MP, Greenberg CR, Salman NJ, et al. Enzyme-replacement therapy in life-threatening hypophosphatasia. *N Engl J Med*. 2012;366:904–913.
138. Whyte MP, Madson KL, Phillips D, et al. Asfotase alfa therapy for children with hypophosphatasia. *JCI Insight*. 2016;1:e85971.
139. Choida V, Bubbear JS. Update on the management of hypophosphatasia. *Ther Adv Musculoskelet Dis*. 2019;11:1759720X19863997.
140. Birtwell Jr WM, Riggs L, Peterson LF, et al. Hypophosphatasia in an adult. *Arch Intern Med*. 1967;120:90–93.
141. Whyte MP, Mahuren JD, Vrabel LA, et al. Markedly increased circulating pyridoxal-5'-phosphate levels in hypophosphatasia. Alkaline phosphatase acts in vitamin B6 metabolism. *J Clin Invest*. 1985;76:752–756.
142. Donnelly LF, Johnson 3rd JF, Benzing G. Infantile osteopetrosis complicated by rickets. *AJR Am J Roentgenol*. 1995;164:968–970.
143. Aker M, Rouvinski A, Hashavia S, et al. An SNX10 mutation causes malignant osteopetrosis of infancy. *J Med Genet*. 2012;49:221–226.
144. Ye L, Morse LR, Zhang L, et al. Osteopetrorickets due to Snx10 deficiency in mice results from both failed osteoclast activity and loss of gastric acid-dependent calcium absorption. *PLoS Genet*. 2015;11:e1005057.
145. Coursey C, Weber T, Dodd L, et al. Fibrogenesis imperfecta ossium: MR imaging of the axial and appendicular skeleton and correlation with a unique radiographic appearance. *Skeletal Radiol*. 2007;36:1077–1084.
146. Bhadada SK, Dhaliwal R, Dhiman V, et al. Fibrogenesis imperfecta ossium. *Calcif Tissue Int*. 2019;104:561–569.
147. McCullough P, Heaney R. Correction of vitamin D deficiency using sublingually administered vitamin D2 in a Crohn's disease patient with mal-absorption and a new ileostomy. *J Steroid Biochem Mol Biol*. 2017;173:211–214.
148. Maung HM, Elangovan L, Frazao JM, et al. Efficacy and side effects of intermittent intravenous and oral doxercalciferol (1alpha-hydroxyvitamin D(2)) in dialysis patients with secondary hyperparathyroidism: a sequential comparison. *Am J Kidney Dis*. 2001;37:532–543.
149. Okonofua F, Gill DS, Alabi ZO, et al. Rickets in Nigerian children: a consequence of calcium malnutrition. *Metabolism*. 1991;40:209–213.
150. Collins N, Maher J, Cole M, et al. A prospective study to evaluate the dose of vitamin D required to correct low 25-hydroxyvitamin D levels, calcium, and alkaline phosphatase in patients at risk of developing antiepileptic drug-induced osteomalacia. *Q J Med*. 1991;78:113–122.

63

Paget Disease of Bone

Frederick R. Singer

OUTLINE

Incidence and Epidemiology, 1067
Pathophysiology, 1068
Clinical Features, 1070
  Regional Manifestations, 1070
  Sarcoma, Giant-Cell Tumors, and Nonskeletal Malignancies, 1071
Biochemical Features, 1072
  Indices of Bone Resorption, 1072
  Indices of Bone Formation, 1072
  Calciotropic Hormones, 1072
Systemic Complications and Associated Diseases, 1072
  Primary Hyperparathyroidism, Nephrolithiasis, Hypercalciuria, Hypercalcemia, 1072
  Hyperuricemia and Gout, 1072
  Cardiovascular Abnormalities, 1072
Drug Treatment, 1073
  Indications for Treatment, 1073
  Pretreatment Laboratory Evaluation, 1073
  Calcitonin, 1073
  Bisphosphonates, 1073
Treatment and Posttreatment Laboratory Evaluation, 1074
  Surgery, 1074
  Causes, 1075
  Animal Models of Paget Disease, 1075

Paget disease of bone (osteitis deformans) is a skeletal disorder of a focal nature that is an example of physiology gone awry. The primary disturbance appears to be an exaggeration of osteoclastic bone resorption that initially produces localized loss of bone. Usually, the disorder is not appreciated until secondary bone formation becomes so pronounced that it results in one or more enlarged and deformed bones. The clinical findings depend on which bones are affected and vary from discovery of an incidental radiologic or biochemical abnormality to devastating musculoskeletal disabilities with a variety of neurologic and systemic complications. Great progress has been made in our ability to treat this disorder, and knowledge of its pathogenesis is increasing.

INCIDENCE AND EPIDEMIOLOGY

Studies of the incidence of Paget disease are inherently imprecise, because many affected individuals are asymptomatic and have normal serum total alkaline phosphatase levels.[1] On the basis of autopsy studies[2,3] and review of radiographs, the prevalence of the disease has been 3% or greater in individuals older than 40 years of age in countries where the disease is common.[4]

A striking feature of the epidemiology of Paget disease is the great variability in prevalence estimates in different regions of the world, and even within a single country. A survey of hospital radiographs in patients older than 55 years of age in 31 British towns revealed a prevalence of Paget disease ranging from 2.3% in Aberdeen, Scotland to 8.3% in Lancaster, England.[5] A similar survey done throughout Europe found that only in France did the prevalence equal the lowest prevalence rates in Britain.[6] Australia, New Zealand, and the United States[7] have a relatively high prevalence, perhaps because of British migration. The disease appears to be rare in Asia. In Japan, the prevalence has been estimated to be 2.8 per million of the population.[8] In most studies, the prevalence of Paget disease in men slightly exceeds that in women.[2-7] A meta-analysis of the temporal changes in prevalence of Paget disease in a variety of countries indicates that there has been a considerable decrease in the populations studied, but that this is heterogeneous, even in the same country.[9]

Much emphasis has been placed on the increase in prevalence of Paget disease with aging. It has been estimated that the prevalence is nearly 10% by the ninth decade.[2-4] Conversely, the disease has rarely been reported in individuals younger than 20 years of age. Although Paget disease is most often recognized after 50 years of age, it is probably misleading to conclude that the disease is rare in younger individuals. As is discussed later, the obvious manifestations, such as skeletal deformity, probably evolve over decades. Failure to diagnose the earlier phases of the disease is no doubt because of lack of symptoms and minimal use of the radiologic and biochemical tests that would lead to an early diagnosis in younger individuals.

Since 1883, it has been appreciated that Paget disease may affect more than one member of a family.[10] In large studies, a family history of the disease has been obtained in 12.3% of 788 cases[11] (United States), 13.8% of 407 cases[12] (Great Britain), and 22.8% of 658 cases[13] (Australia). In the former two studies, a 7- to 10-fold increase in Paget disease was noted in relatives of patients in comparison with control groups. In a small study in Spain in which relatives of patients were screened with bone scans, 40% of patients had at least one first-degree relative with Paget disease.[14] Examination of the overall pattern of apparent transmission suggests an autosomal dominant mode of inheritance.[10]

The search for potential environmental factors in the pathogenesis of Paget disease has led to consideration of whether past ownership of dogs might be a risk factor,[15] but this has not been confirmed by subsequent studies. Occupational exposure to lead has been proposed as a possible factor in Paget disease.[16] In one study, levels of cortical bone lead were higher than those in control subjects, but trabecular bone lead was lower.[17] A more recent Canadian study reported that exposure to wood burning during childhood could be a pathogenic factor.[18] The possible role of viral infections will be reviewed later.

PATHOPHYSIOLOGY

KEY POINTS

- The primary underlying pathology of Paget disease is a localized increase in osteoclastic bone resorption.

Consideration of the radiologic and pathologic evolution of the lesions of Paget disease strongly suggests that the primary disturbance is localized acceleration of osteoclastic bone resorption. At the interface of normal bone and an advancing lesion, numerous osteoclasts are found in Howship's lacunae in cortical or trabecular bone.[2] Many of the osteoclasts are larger than normal (Fig. 63.1) and may have up to 100 nuclei in cross-section rather than the several found in normal osteoclasts.[19] Examination of the ultrastructure of osteoclasts in specimens of Paget disease reveals a striking and characteristic feature: the presence of microfilaments in the nucleus and, less frequently, in the cytoplasm.[20,21] These structures have not been observed in osteoclasts from normal subjects or in the bones of patients with primary or secondary hyperparathyroidism, osteoporosis, or osteomalacia. They also have not been observed in osteoblasts, osteocytes, or bone marrow cells in the lesions of Paget disease. The inclusions have been found in a small percentage of the multinucleated giant cells (osteoclasts) in giant-cell tumors of bone and in the osteoclasts of some patients with osteopetrosis and pyknodysostosis.[22] The structure of the microfilaments most closely resembles the nucleocapsids of viruses of the *Paramyxoviridae* family, a group of RNA viruses known to cause common childhood infections. Evidence of *Paramyxoviridae* nucleocapsid proteins[23,24] and messenger RNA[25,26] has been found in pagetic lesions. The full-length sequence of the measles virus nucleocapsid (*MVNP*) gene has been sequenced from a pagetic lesion of one patient.[27] The relevance of these findings to the cause of Paget disease is discussed later.

As osteoclastic resorption progresses in the cortex, individual osteons widen and become confluent with adjacent osteons. The resorptive area thereby may extend to the endosteal and periosteal surfaces. In trabecular bone of the medullary cavities, the osteolytic process results in a marked reduction in bone volume. Associated with both early cortical and trabecular lesions is a remarkable proliferation of fibrous tissue that replaces normal fatty or hematopoietic bone marrow. The fibrous stroma is highly vascular, and although arteriovenous shunts were previously thought to be present, this feature has not been confirmed with radiolabeled albumin microspheres.[28]

The earliest recognizable radiologic feature of Paget disease is a focal osteolytic lesion. The skull was first appreciated to be affected by circumscribed osteolytic lesions, and Schuller[29] applied the term *osteoporosis circumscripta* to this finding. One or more osteolytic foci may be present, most often in the frontal and occipital regions, and may be observed to coalesce slowly over a period of years (Fig. 63.2). Pure osteolytic lesions also may be detected in other regions of the skeleton, but less frequently. They are seldom observed in the vertebral column or pelvis. In the long bones, osteolytic lesions usually develop at either end of the bone, less often in the diaphysis.[29] Occasionally, osteolytic foci can be observed simultaneously at both ends of a bone. The junction of normal bone and the osteolytic lesion shows a characteristic radiologic appearance that is nearly diagnostic of Paget disease. The edge of the lesion usually assumes the shape of a flame or an inverted V (Fig. 63.3). Serial radiologic follow-up of untreated lesions has documented an average rate of extension of approximately 1 cm annually.[30] Bone biopsies taken during this earliest phase of Paget disease not unexpectedly reveal a marked increase in osteoclastic activity and thinning of the trabeculae.[31] Other striking features include fibrovascular marrow, numerous osteoblasts lining the trabeculae, and prominent woven bone. Thus, a discrepancy is found in the results of radiologic and pathologic examination of an osteolytic lesion. Although focal density is decreased on radiographs, histology demonstrates bone formation that is very active but is not sufficient to overcome the remarkable degree of osteoclastic bone resorption.

A more commonly observed stage of Paget disease is the mixed phase, in which osteoblastic (or osteosclerotic) features are intermixed with osteolytic features in an individual bone. This phase is best appreciated in long bones, in which one may observe the advancing osteolytic front adjacent to normal bone and, trailing this front,

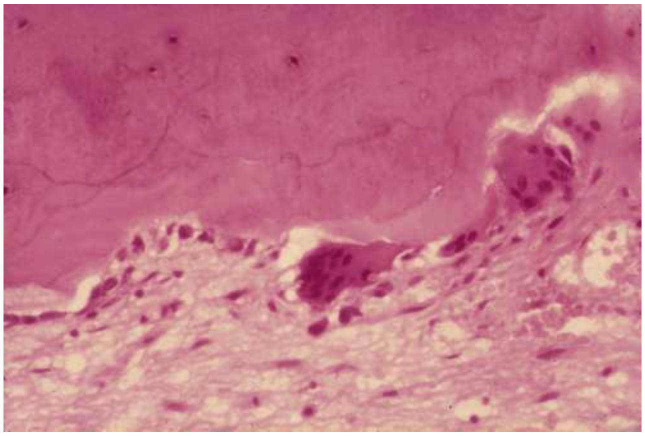


Fig. 63.1 **Bone Biopsy of Paget Disease with Large Multinucleated Osteoclasts in Howship's Lacunae with Adjacent Plump Osteoblasts in Previously Resorbed Sites.** Hematoxylin and Eosin Stain.

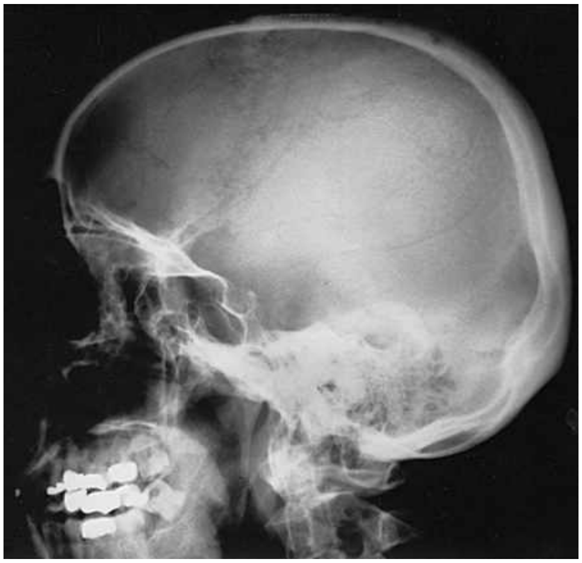


Fig. 63.2 **Radiograph Demonstrating Osteoporosis Circumscripta of the Skull Affecting the Frontal and Temporal Regions.**

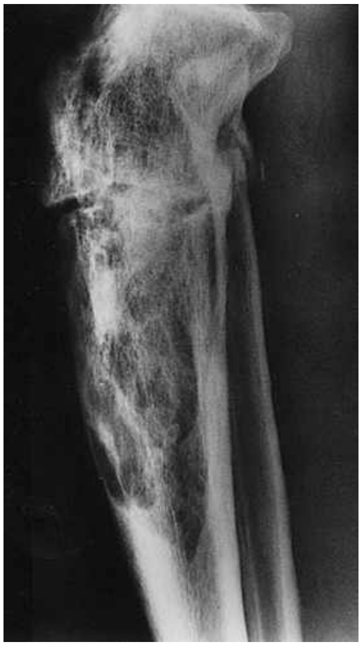


Fig. 63.3 **Radiograph of a Tibia Exhibiting a Distal Advancing Osteolytic Front with Proximal Sclerotic Bone.** A Partially Healed Pathologic Fracture is Present Proximally.

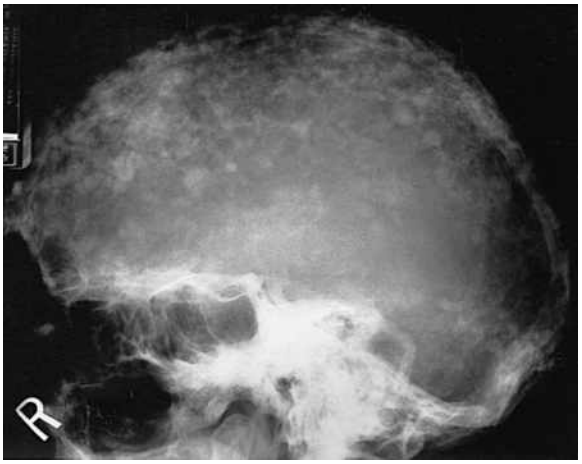


Fig. 63.4 Radiograph of the Sclerotic Phase of Paget Disease in the Skull Exhibiting a "Cotton-Wool" Appearance.

a heterogeneous region of osteosclerosis superimposed on the region that had previously been dominated by the osteolytic process (see Fig. 63.3). Biopsies of the mixed phase reveal a characteristic abnormality of lamellar bone in both cortical and trabecular bone. The matrix is transformed into a bizarre "mosaic" pattern of irregularly juxtaposed pieces of lamellar bone separated by cement lines that have a scalloped outline. The irregularity probably reflects areas of previous osteoclastic resorption. The structure of the involved cortex is so disordered that complete osteons are rare, and the outer and inner circumferential lamellae and the interstitial lamella may be totally disrupted.

The same disordered matrix structure is seen in trabecular bone. Interspersed among the chaotic lamellae are patches of woven bone characterized by a random pattern of deposition of collagen fibers and a larger number of osteocytes per unit area of matrix. It has been suggested that the lacunae surrounding the osteocytes are larger than normal, and that this finding represents osteocytic osteolysis.[32] However, it is more likely that the increased size of the periosteocytic lacunae is simply a characteristic of woven bone and not a second type of bone resorption in Paget disease. At the surfaces of bone formation, plump osteoblasts are found in great number adjacent to abundant osteoid. This type of bone is seldom found in adults, except when associated with rapid remodeling of bone, such as occurs after a fracture or in response to tumor invasion of bone. Studies using quantitative histomorphometry of bone documented the marked degree of cellular activity underlying the dramatic changes in bone structure in Paget disease.[33] The total amount of osteoid and the percentage of the bone surface covered by osteoid may be increased 4-fold to 5-fold. The increase in osteoid is not associated with an increase in osteoid seam width, because the rate of calcification also is increased, as established by double labeling of bone-forming surfaces with tetracycline. No dynamic means of defining the rate of bone resorption is available, but the extent of the total bone surface exhibiting evidence of bone resorption averages approximately 6-fold that of normal individuals, and the number of osteoclasts may be increased as much as 10-fold. In the medullary cavities, the intense resorptive process may produce hemorrhagic cysts with encircling fibrous marrow containing macrophages filled with hemosiderin. These cysts are believed to result from rupture of multiple dilated vessels and ensuing microinfarctions.[34] In the mixed phase of Paget disease, not only does patchy sclerosis of bone become apparent on radiography, but a bone also may be enlarged. If the osteolytic process extends to the subperiosteal layer, bone formation may be stimulated to such an extent that the thickness and circumference of the bone are increased as a result of periosteal new bone formation. When the skull is affected, this process can produce as much as a 4-fold thickening of the calvarium, as was reported by Paget.[35] A patchy form of sclerosis is often of a "cotton-wool" character (Fig. 63.4). The skull may be so severely affected that platybasia, or basilar impression, may be a complication. Long bones may be shortened because of lateral bowing of the femur or anterior bowing of the tibia, or both. Later in the course of the disease, the tibia may exhibit severe lateral bowing. The pathogenesis of the slowly progressive deformity is not known but must be related to the state of abnormal remodeling of the bone. Frequently, fissure fractures are associated with the bowing deformity. These fractures are linear transverse radiolucencies that usually are present in the cortex of the convex aspect of the deformed bone. They often are multiple and may remain stable in appearance for years. They can be found in either osteopenic or osteosclerotic cortices and may be present even in the absence of deformity. Histologic examination of these lesions suggests that they are incomplete fractures.[2] Only a small percentage of these lesions progress to a complete transverse fracture, which has been seen more often in patients with a sclerotic cortex.

Even after osteosclerotic bone has invaded the previously osteolytic regions of affected bone, evidence of ongoing abnormal bone resorption can be seen in the form of secondary resorption fronts. These secondary fronts are commonly seen as clefts in the cortex of long bones and trail in the path of the primary front.

In the final stage of Paget disease, termed *osteoblastic* or *sclerotic*, the affected bone remains dense and retains the "mosaic" matrix pattern characteristic of Paget disease. Tubular bones show a loss of differentiation between cortical and trabecular bone, even with enlargement of the bone as a result of periosteal new bone formation, because the new bone is no longer compact bone. Much less cellular activity is present in sclerotic lesions. Osteoclasts are few or absent, but osteoblasts may still be seen to line bone surfaces. The marrow may remain fibrous, but the number of blood vessels is greatly reduced. Scattered chronic inflammatory cells may be present. Occasionally, parts of lesions are totally devoid of bone cells, and for this reason the concept of "burned-out" Paget disease has arisen. However, it is very unlikely that extensive

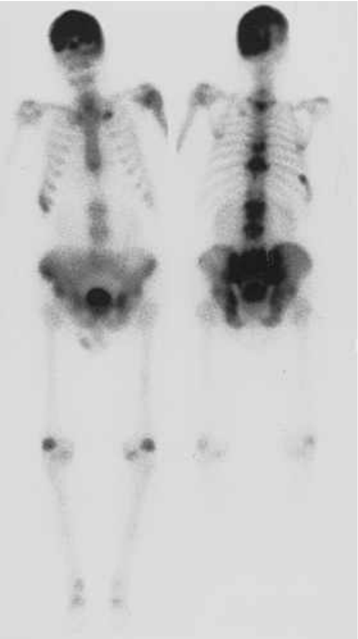


Fig. 63.5 **Anterior and Posterior Views of a Bone Scan in a Patient with Polyostotic Paget Disease.** Increased Uptake of Tracer Can be Noted in the Skull, Multiple Vertebrae, and the Pelvis, Areas in Which Paget Disease was Observed on Radiographs. The Other Abnormal Areas Probably Represent Degenerative Arthritis and a Healed Rib Fracture.

lesions of Paget disease ever achieve an entirely burned-out state. On the contrary, the presence of all stages of the disease in a single bone is much more likely.

The evolution of Paget disease can also be observed by administration of radioactive tracers and scanning of the entire skeleton or selected regions. Bone scans use technetium-labeled bisphosphonates, which, after intravenous injection, localize to skeletal sites in proportion to the relative blood flow and the rate of bone formation. The scans usually, but not always, demonstrate high uptake of radioactivity in the areas of the skeleton noted to be radiographically abnormal[36] (Fig. 63.5). In a small proportion of patients, increased uptake may be seen when the radiograph is normal, thus illustrating the great sensitivity of bone scans. On the contrary, a small percentage of sclerotic lesions may not be picked up by a bone scan. These lesions appear to represent areas of inactive disease. Bone scans can be analyzed semiquantitatively or by computer and thus may be used to monitor response to treatment.

Gallium scans, most often used to detect occult infection or tumors, also have been shown to delineate the lesions of Paget disease.[37] Evidence indicates that tracer gallium is localized to the nuclei of osteoclasts.[38] Therefore, the gallium scan may serve as a direct index of cellular activity in Paget disease.

CLINICAL FEATURES

KEY POINTS

- A significant proportion of patients have neither symptoms nor signs of the disease.

A considerable proportion of individuals with Paget disease have neither symptoms nor signs of the disease.[1] The disease is accidentally discovered in these cases because of radiologic or biochemical abnormalities uncovered during investigation of another disorder.

The most common clinical problems are pain and deformity. The bone pain of Paget disease, when present, is seldom severe. Usually it is a dull pain, is located deep below the soft tissues, and often persists during the night. In weight-bearing bone, the pain may be slightly worsened by ambulation, but to a lesser extent than pain originating from the joints or from nerve impingement. Deformity of the skeleton is most often noted in the skull and the lower extremity. Over many years, the hat size of an affected individual may be noted to increase. Bowing and enlargement of the femur and tibia also may evolve over a period of years.

Regional Manifestations

The Skull. In the absence of an enlarged cranium, symptoms in the skull are uncommon. Even with an enlarged cranium, symptoms often are absent. Certainly, the most common symptom (30%–50%) is hearing loss,[39] which is slowly progressive in untreated patients. Vertigo or tinnitus or both are much less common. The main mechanism of hearing loss in Paget disease has been attributed to a reduction in bone mineral density of the cochlear capsule.[40]

Much more serious complications of Paget disease may occur in the advanced stage of the disease in a small number of patients. The weight of the skull may be so great that the ability of the patient to keep the head erect is impaired. Muscle spasm may then produce pain in the neck and tension headaches. Neurologic abnormalities may be found in such patients as a consequence of basilar invagination or platybasia. Although they are unusual complications, even in the presence of radiologic evidence of basilar invagination compression of structures in the posterior fossa or cerebellar tonsillar herniation may produce ataxia, muscle weakness, and impaired respiration. Hydrocephalus also has been noted to be a rare complication and may be manifested by impaired gait, urinary incontinence, and some degree of dementia.

Finally, severe skull disease may be associated with the vague findings of a withdrawn individual who is somnolent and weak. It has been suggested that this manifestation might be a consequence of shunting of blood from the brain vessels to the external carotid artery system, a possible pagetic steal syndrome.[41] These symptoms also could represent a psychological response to disability, inasmuch as nearly 50% of patients in one study have been reported to have depression.[42]

The Jaws. Paget disease may affect the facial bones and jaw, but such involvement is uncommon. *Leontiasis ossea* is the descriptive term applied to a patient with enlargement of all the facial bones, but such deformity is more likely to be found in fibrous dysplasia.

Involvement of the mandible or maxilla may produce progressive root resorption leading to the loss of teeth.[43] In the more advanced stages of Paget disease, excessive formation of the cementum is associated with absence of the lamina dura and periodontal membranes. Facial disfigurement may occur from enlargement of the maxilla or mandible or both and is associated with spreading of the teeth and malocclusion. Edentulous patients have difficulty acquiring properly fitting dentures. Oral surgery may be complicated by excessive intraoperative bleeding and postoperative osteomyelitis. Tooth extractions may prove difficult because of ankylosis resulting from hyperplasia of the cementum.

The Spine. Neck pain and back pain are common complaints in an aging population of patients with Paget disease. Visualization of Paget disease in one or more vertebrae on radiographic examination often leads clinicians to conclude that they are dealing with bone pain, and treatment is instituted. However, most patients with moderate to severe pain have a complication associated with Paget disease as the cause of the pain rather than bone pain alone.

Paget disease affects the lumbar and sacral regions most frequently. One vertebra or many vertebrae may be involved. In the early osteolytic phase, which often is not recognized, the vertebral body appears osteoporotic and, in rare cases, may undergo so much resorption that it takes the shape of a thin transverse rod. Much more often, the vertebral bodies become enlarged overall, with thickened margins and coarse vertical striations centrally. Compression of a sclerotic vertebral body may develop because of the abnormal mechanical properties produced by the chaotic microarchitecture.

Severe pain or impaired neurologic function or both may result from compression of the spinal cord or nerve roots.[44] This complication can arise from enlargement of the vertebral bodies, pedicles, or laminae, as well as from compression fractures. It also has been suggested that shunting of blood may occur from the spinal arteries to the highly vascular bone.[45] Neurologic syndromes are more likely to develop with thoracic involvement. Symptoms include back pain, difficult ambulation, numbness, paresthesias of the feet, and progressive paresis of the legs. Later problems can include impaired bladder and bowel function, as well as spastic paraparesis and loss of sensation. Computed tomography (CT) and magnetic resonance imaging (MRI) are particularly helpful in resolving the anatomic abnormalities producing the disturbed function.

A rare complication in the spine is the development of a discrete paraspinal mass consisting of a central marrow cavity surrounded by pagetic bone that extends from the vertebrae.[46] It may appear that the lesion represents a neoplasm, but careful analysis of prior radiologic studies may reveal the chronic nature of the lesion and may make it unnecessary to perform a biopsy.

The other major causes of back pain in Paget disease are intervertebral disc disease and degenerative arthritis. No evidence indicates that disc degeneration is more common in patients with Paget disease, but it has been reported that the pagetic process can invade the disc and produce bony bridging across the disc space.[47] Back pain in the lumbar region is frequently associated with degenerative arthritis,[48] particularly when distortion of the facet joints is associated with Paget disease. Large osteophytes also may be found in association with enlarged vertebral bodies or where a compression fracture has occurred. A syndrome mimicking ankylosing spondylitis may occur in the presence of extensive osteophyte formation or with ossification of spinal ligaments,[49] but the human leukocyte antigen–B27 antigen is absent. However, classic ankylosing spondylitis has been found in association with Paget disease.[48]

The Pelvis and Extremities. The main symptoms associated with pelvic and lower extremity involvement are pain and impaired ambulation. Pain is seldom a significant symptom in the osteolytic phase of the disease. Hip pain is most common when both the acetabulum and the proximal end of the femur are affected by the sclerotic phase of the disease.[49] Bowing of the femur and protrusio acetabuli are often associated with pain aggravated by weight bearing. Many patients are relatively comfortable when not weight bearing, unlike patients with bone pain, who usually have nocturnal discomfort.

Knee pain and occasionally joint effusions may occur with sclerotic disease affecting the femur or tibia or both. Distortion of the knee joint produced by enlargement of the distal end of the femur or the proximal part of the tibia and severe bowing of either bone can induce mechanical strains on the articular cartilage and thereby accelerate the degenerative process. The pagetic process in subchondral bone also may contribute to joint disease. A similar set of circumstances may account for ankle pain.

Fractures of the lower extremity are more likely to affect the femur than the tibia. In the largest series of reported femoral fractures, the subtrochanteric region was the most common site of fracture (49 of 182), and the rate of nonunion was noted to be 40%, a figure considerably greater than was previously appreciated.[50] Nonunion appears to be less common after tibial fractures.

Involvement of the upper-extremity long bones is much less likely to produce symptoms, although deformity may be apparent. At the shoulder, impaired rotator cuff function may be noted when overgrowth of bone leads to anatomic distortion of the glenohumeral joint.

Occasionally, patients who have Paget disease affecting the foot have pain on weight bearing. Symptoms are seldom encountered in individuals with radiologic evidence of Paget disease in the hands.

Sarcoma, Giant-Cell Tumors, and Nonskeletal Malignancies

The most feared complication in Paget disease is sarcoma. It has been estimated that 10% of patients with extensive disease may experience this problem, but if all affected individuals are considered, the incidence is probably less than 1%.[52] Sarcomas have rarely been reported to develop in multiple members of families with Paget disease. This will be discussed in the Genetics section.

Patients in whom sarcomas develop usually have pain and swelling, always in an area previously affected by Paget disease. Occasionally, fracture at the tumor site may lead to discovery of the neoplasm. Tumors most often arise in the pelvis, femur, humerus, skull, and facial bones.[52] Multifocal sarcomas are found only in patients with advanced and widespread polyostotic disease and are thought to represent tumors of independent origin rather than metastases.[53]

The histology of sarcomas is quite variable. Fibrosarcomas, chondrosarcomas, osteogenic sarcomas, and anaplastic sarcomas may be found. Variable numbers of multinucleated giant cells (probably osteoclasts) may be scattered throughout the tumor stroma and most likely are not neoplastic. The nuclear inclusions typical of Paget disease have been observed in giant cells but not in the tumor cells.[54] It is not unusual for several histologic patterns to be present in a single tumor, which suggests that a common stem cell may give rise to a variety of better differentiated mesenchymal cells.

Lymphomas and multiple myelomas have been found in association with Paget disease[51] but probably represent chance occurrences rather than a complication of the pathologic process.

It is difficult to detect early sarcoma formation by radiologic examination, because of the underlying distortion of pagetic bone. Because they appear to arise in medullary bone, an early finding may be a subcortical osteolytic lesion. Only when a radiolucent focus with speckled areas of calcification has broken through the confines of the cortex is it apparent that a malignant neoplasm is present. CT or MRI is the best means of determining the extent of the tumor mass.

The rate of change in serum alkaline phosphatase activity has not proved to be a useful marker for the development of sarcomas, despite early reports that this might be the case.

The life expectancy for the average patient in whom a sarcoma develops is sadly brief, perhaps because of the difficulty associated with early detection. In one study, only 7.5% of patients survived 5 years, whereas a 37% 5-year survival rate was noted in older adult patients who were free of underlying Paget disease.[55]

Giant-cell tumors of bone, which usually follow a benign course and most often are found at the ends of long bones in otherwise normal individuals, may arise in the lesions of Paget disease.[56] They appear to be much less common than sarcoma in Paget disease and have frequently been noted to originate in the skull and facial bones. Rarely, these tumors have been reported in multiple family members who have Paget disease.[57]

These tumors are characterized by spindle-shaped cells with fusiform nuclei and clumped chromatin or nuclei and by scattered multinucleated giant cells. Mitoses are rarely found in either the mononuclear or giant cells. The giant cells contain the nuclear inclusions of Paget disease, but the stromal cells do not.[56] The opinion has been expressed that many of the reported cases of giant-cell tumor in Paget disease actually represent giant-cell reparative granulomas, which are common lesions arising in the jaw.[58]

Surgery and radiation have been used to treat symptomatic giant-cell tumors in Paget disease. In one patient, high doses of dexamethasone were effective in shrinking an extraskeletal tumor.[59] Another patient responded to the antiresorptive agent, denosumab.[60]

BIOCHEMICAL FEATURES

The intense cellular activity in active lesions of Paget disease may be reflected in various biochemical markers of bone resorption and bone formation. In most patients who come to clinical attention, biochemical markers do reflect the extent and activity of the disease, although patients with only a small percentage of affected skeleton have no biochemical abnormalities.

Indices of Bone Resorption

The increased bone resorption typical of active Paget disease might be expected to produce an increase in serum and urinary calcium levels, but in the absence of fractures or immobilization, hypercalcemia or hypercalciuria is not a prominent feature of Paget disease.[61] It is generally believed that this finding is explained by a concomitant increase in bone formation that is demonstrable histologically and by kinetic analysis of plasma disappearance rates and skeletal uptake of radiocalcium[61] or other skeletal tracers. A variety of bone collagen matrix breakdown products have been used as indices of bone resorption. These products include urinary hydroxyproline, total and free pyridinoline and deoxypyridinoline, type 1 collagen N-telopeptide, and type 1 collagen C-telopeptide. The telopeptide assays appear to be most specific for bone collagen resorption. Measurement of non-isomerized fragments of collagen type 1 C-telopeptide may be the most sensitive means of evaluating bone resorption in patients with Paget disease.[62]

Indices of Bone Formation

Since 1929, it has been appreciated that serum total alkaline phosphatase activity may be increased in patients with Paget disease.[63] The enzyme is localized at the plasma membrane in osteoblasts and may participate in the mineralization of bone matrix. In Paget disease, enzyme activity in the circulation correlates with the extent of disease on radiographic skeletal surveys, as well as with total urinary hydroxyproline.[49] Effective treatment of Paget disease reduces total alkaline phosphatase activity by 50% or more.

During long-term follow-up of untreated patients with Paget disease, alkaline phosphatase activity usually exhibits a gradual increase or no significant change.[64] In some patients, major fluctuations may represent technical errors rather than changes in disease activity. In the presence of liver disease, hepatic alkaline phosphatase activity may interfere with an accurate assessment of Paget disease activity. In such patients, measurement of procollagen type 1 N-terminal peptide levels by immunoassay is preferable.[65]

Serum osteocalcin or bone Gla protein is a nearly specific product of osteoblasts that may be elevated in Paget disease, but not to the same degree as serum alkaline phosphatase activity.[66] Paradoxically, treatment of patients with drugs that reduce turnover may produce a transient increase in osteocalcin levels.[65] A potential explanation for these observations is that osteocalcin gene expression is reduced in active Paget disease.[67]

Serum levels of procollagen type 1 N-terminal peptide correlate significantly with quantitative bone scintigraphy and decline with treatment of Paget disease to a greater extent than bone-specific alkaline phosphatase.[65]

Because of the cost-effectiveness of the serum total alkaline phosphatase assay, this remains a reasonable choice for the biochemical assessment of the average patient with Paget disease.

Calciotropic Hormones

Calcitonin secretion is normal in patients with Paget disease.[68] Parathyroid hormone concentrations generally are within the normal range, but elevated concentrations have been found in subsets of patients.[69,70] This observation could be related to renal failure, vitamin D deficiency, or subtle hypocalcemia related to a higher rate of bone formation than bone resorption. Primary hyperparathyroidism may be a coincidental disorder.

Patients with adequate vitamin D intake or sufficient ultraviolet light exposure or both have normal 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels. In two studies, serum 24,25-dihydroxyvitamin D was reported to be low and to correlate inversely with disease activity.[71,72] The pathogenesis and clinical relevance of this finding remain unknown.

Interleukin-6, an important local modulator of osteoclast function, has been reported to be elevated in the bone marrow of patients with Paget disease[73] and in serum by some,[73-75] but not all, investigators.[76] Receptor activator of nuclear factor-κ beta ligand (RANKL) and osteoprotegerin (OPG), a decoy receptor for RANKL, are key modulators of osteoclast function. Serum measurements in patients with Paget disease have also been variable, but one study indicates that both RANKL and OPG levels may be highly elevated, and that bisphosphonate therapy can increase OPG, and to a lesser extent reduce RANKL levels, thereby reducing the RANKL/OPG ratio.[77]

SYSTEMIC COMPLICATIONS AND ASSOCIATED DISEASES

Primary Hyperparathyroidism, Nephrolithiasis, Hypercalciuria, Hypercalcemia

In a recent large study in Italy, it was reported that the prevalence of nephrolithiasis and primary hyperparathyroidism was greater than in a control group.[78] Hyperoxaluria was a prominent feature in the nephrolithiasis patients. The explanation for these findings remains to be established.

Hypercalciuria is readily provoked by immobilization after fracture[79] or neurologic injury. In this setting bone resorption increases, whereas bone formation decreases.

In addition to primary hyperparathyroidism,[78,80] hypercalcemia also may occur as a consequence of immobilization[79] or malignancy.[81]

Hyperuricemia and Gout

Serum uric acid concentrations have been reported to be elevated primarily in men with relatively severe Paget disease.[49] Nearly half of the hyperuricemic men had clinical episodes of gouty arthritis. Paget disease also was found to be present in 23% of a group of patients with gout.[82] It is possible that a high turnover of nucleic acids in the lesions of Paget disease could increase the urate pool sufficiently to account for these observations.[83]

Cardiovascular Abnormalities

Increased cardiac output has been found to be present in patients who have at least 15% of their skeleton affected by Paget disease.[84] This

TABLE 63.1 Indications for Drug Treatment in Paget Disease

Bone pain
Treatment of high-output congestive heart failure
Preparation for orthopedic surgery
Hypercalciuria
Hypercalcemia
Neurologic deficit from vertebral disease
Prevention of hearing loss
Prevention of complications in young patients

abnormality is often associated with left ventricular hypertrophy. The excessive vascularity of the soft tissue and adjacent pagetic bone must be factors contributing to these phenomena. In addition, a reduction in peripheral vascular resistance may lead to increased cardiac output.[85] High-output cardiac failure may occur but seems to be unusual.

Calcific aortic stenosis appears to be four to six times more common in patients with Paget disease than in a control population.[86,87] It is more likely to be found in patients with severe disease, which suggests that increased cardiac output producing turbulence across the valve may induce calcification. Intracardiac calcification also may occur in the interventricular septum and may produce a complete heart block.[87,88] Arterial calcification of peripheral vessels in patients with Paget disease is also greater than in control subjects.[89]

DRUG TREATMENT

Indications for Treatment

Indications for drug treatment of Paget disease are listed in Table 63.1. Perhaps the most common reason for treatment is bone pain. When Paget disease occurs adjacent to a joint, it may be difficult to distinguish bone pain from joint pain. In such cases, drug therapy may be particularly useful in clarifying the origin of the pain. Pretreatment of patients who require elective orthopedic surgery may prevent complications such as intraoperative or postoperative hemorrhage and immobilization hypercalcemia. Hypercalciuria and hypercalcemia can be reversed or prevented by drug treatment, but these indications are uncommon.

KEY POINTS

- If bone pain is present, or if there is skeletal involvement likely to produce significant complications, a single 5-mg infusion of zoledronic acid can sup-press disease activity for up to 6.5 years.

Some patients with neurologic deficits associated with vertebral disease may experience dramatic remission of their signs and symptoms. Although hearing loss is seldom reversed, preservation of auditory acuity is expected. Reduction of disease activity produces a decrease in cardiac output. It is possible that early treatment could prevent future complications such as skeletal deformity, but no long-term randomized clinical trials have been completed.

Certainly, many patients do not need to be treated. The decision to treat must take into consideration the present symptoms, the likelihood of future complications, the cost of therapy, and the mode of administration.

Pretreatment Laboratory Evaluation

Measurement of serum total alkaline phosphatase activity in a reliable laboratory is probably the only biochemical test needed for most patients. Radiographs of known lesions of Paget disease should be performed to ensure awareness of osteolytic lesions. In patients whose extent of disease is unknown, a bone scan is the best means of defining the regions of the skeleton requiring radiographic evaluation.

Calcitonin

Calcitonin is a peptide hormone that binds to receptors on osteoclasts and can rapidly inhibit bone resorption. Salmon calcitonin was approved for clinical use in the United States in 1975 and was the first therapy useful in treating Paget disease. Subcutaneous injections three to seven times weekly were effective in reducing bone pain, cardiac output, and vascularity of affected bones, reversing some neurologic deficits, and stabilizing hearing deficits.[90] Bone resorption and formation parameters were reduced by 50% in 3 to 6 months. Unfortunately, high antibody titers induced by the hormone caused loss of responsiveness in approximately 26% of patients.[90] Human calcitonin was effective in these patients until its manufacture was terminated. At this time salmon calcitonin is seldom used because of the availability of more effective therapy.

Bisphosphonates

Bisphosphonates are analogs of inorganic pyrophosphate, a compound thought to participate in the mineralization of bone. By substituting a P-C-P bond for the naturally occurring P-O-P bond, a family of metabolically stable compounds has been produced that bind to hydroxyapatite and inhibit bone resorption and, secondarily, bone formation in experimental animals and humans. The earliest bisphosphonates, such as etidronate, induce osteoclast apoptosis by producing nonhydrolyzable analogs of adenosine triphosphate, whereas the more potent aminobisphosphonates inhibit farnesyl-diphosphate synthase in the mevalonate pathway, which produces osteoclast apoptosis by inhibition of protein prenylation.[91]

Four oral bisphosphonates taken daily are etidronate[92] (5 mg/kg body weight for 6 months), alendronate[93] (40 mg for 6 months), tiludronate[94] (400 mg for 3 months), and risedronate[95] (30 mg for 2 months). Pamidronate is available in intravenous form and is commonly infused once over a period of several hours at a dose of 60 mg for patients with less than 5-fold elevations of serum total alkaline phosphatase activity[96] and at an intermittent dose of 60 or 90 mg on 2 or more days, depending on the level of alkaline phosphatase and the response to each infusion. Zoledronic acid, the most recently approved agent, is administered as a 5-mg dose intravenously over 15 minutes.[97]

The bisphosphonates taken orally must be ingested with water only on an empty stomach, because they are poorly absorbed. Generally, side effects are not a major problem and, when present, include abdominal distress, diarrhea, and a temporary increase in bone pain. Patients who receive pamidronate or zoledronic acid intravenously may experience fever and myalgias for approximately 24 hours or more. This acute-phase response is a consequence of cytokine release from gamma delta T cells and appears to be accentuated in patients with serum 25OHD levels below 30 ng/mL.[98] This side effect seldom occurs with subsequent infusions. Allergic reactions are rare and most often are inflammatory eye reactions associated with pamidronate use. Etidronate is the only bisphosphonate reported to produce significant osteomalacia, usually at a dose greater than 5 mg/kg body weight daily. This drug is now rarely used, given the existence of better agents. Renal toxicity is not observed with intravenous zoledronic acid if the drug is infused as approved, and the creatinine clearance is 35 cc/min or higher.[99] Osteonecrosis of the jaw, an unusual event in cancer patients treated monthly with intravenous bisphosphonates, is extremely uncommon in patients with Paget disease or osteoporosis.[100]

The potent aminobisphosphonates can induce biochemical remissions in the great majority of patients with mild to moderate disease activity. Only zoledronic acid can consistently normalize biochemical parameters in the most severely affected patients,[97] and the response

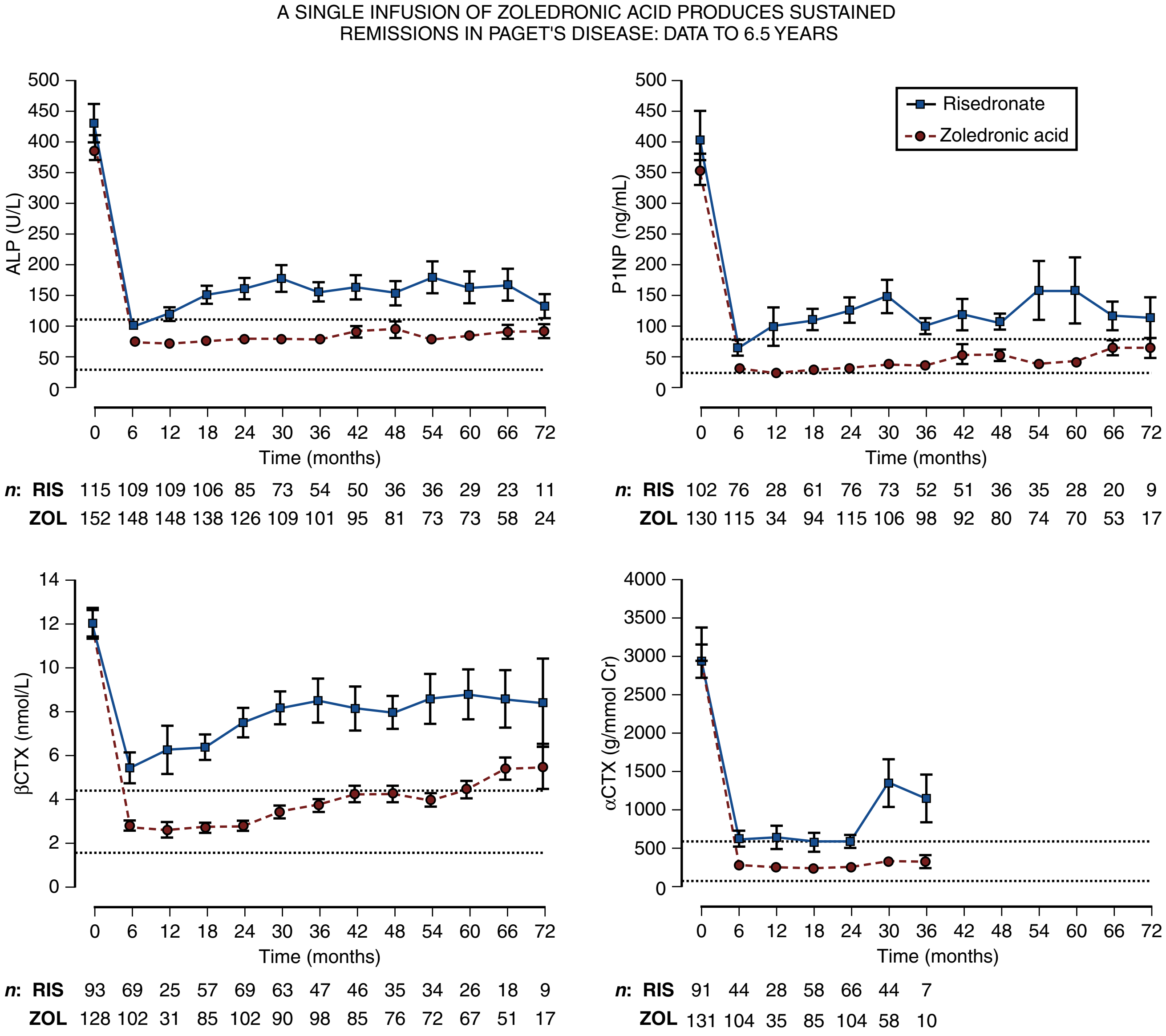


Fig. 63.6 **Comparison of the Long-Term Suppression of Bone Turnover Markers by a 5-mg Single Infusion of Zoledronic Acid (*ZOL*) or a 2-Month Course of Risedronate (*RIS*) tablets, 30 mg/day.** Bone Formation Markers, Alkaline Phosphatase (*ALP*) and N-Terminal Propeptide of Type 1 Collagen (*P1NP*), and Bone Resorption Markers, Serum Beta-C-Telopeptide (*beta CTX*) and Urine Alpha–C-Telopeptide (*alpha-CTX*) Remained Suppressed Up to 6.5 Years After Zoledronic Acid Infusion. Risedronate Had a Lesser Effect, and Patients Dropped Out of the Study After Relapse. The *Dotted Lines* Indicate the Reference Ranges. The Data are Presented as Mean ± Standard Error of the Mean. (From Reid IR, Lyles K, Su G, et al. A single infusion of zoledronic acid produces sustained remissions in Paget disease: data to 6.5 years. *J Bone Miner Res.* 2011;26:2261–2270.)

can persist for up to 6.5 years[101] (Fig. 63.6). It remains to be seen whether long-term biochemical suppression, an achievable goal in most patients, can reduce the incidence of complications in patients who are most at risk. These individuals include those with skull, vertebral, pelvic, and lower-extremity involvement.

If there is a contraindication to use of a bisphosphonate such as renal failure, denosumab may be effective and safe, as reported in one such patient.[102]

TREATMENT AND POSTTREATMENT LABORATORY EVALUATION

For most patients, measurement of total serum alkaline phosphatase activity is sufficient to determine the success of treatment. In patients with known osteolytic lesions on radiologic examination, an annual evaluation should be adequate.

Surgery

Certainly, the benefits of surgery for the appropriate indications in patients with Paget disease outweigh the potential complications of excessive hemorrhage and impaired healing. Hemorrhage is less likely if patients are treated with an effective drug preoperatively. Probably the most common reason for orthopedic surgery is total hip replacement.[103] The rate of success in relieving intractable hip pain and improving mobility is excellent. Heterotopic ossification may be somewhat more common postoperatively but is seldom a major problem. Total knee replacement is also a successful procedure.[104] Tibial and fibular osteotomies to correct varus deformity of the tibia are impressive in relieving knee and ankle pain associated with marked deformity.[105] Because the rate of nonunion is relatively high in femoral fractures, open reduction and fixation of these fractures may prove necessary.

Much less commonly required are suboccipital craniectomy and upper cervical vertebral laminectomy in patients with symptomatic basilar impression. Equally uncommon is the need for ventricular shunting in patients with hydrocephalus. Attempted relief of hearing loss in patients with skull loss by stapes mobilization or stapedectomy has been of questionable benefit. Surgery to correct spinal stenosis or nerve root compression has generally been successful.

Causes

Genetics. In 2002, a mutation in the sequestosome 1 (*SQSTM1*) gene on chromosome 5 was reported in 11 of 24 French Canadian families with Paget disease and in 18 of 112 patients with apparently sporadic disease (see Chapter 59).[106] By 2010, at least 20 separate mutations of this gene were documented in a variety of countries.[107] The mutations are mainly clustered around the ubiquitin-binding domain of the SQSTM1 protein. The SQSTM1 protein modulates activity of the nuclear factor kappa B pathway, an important mediator of osteoclast function, and is involved in the process of autophagy. It is unlikely that mutations of this gene entirely account for a pagetic phenotype, because many of the family members of these pedigrees exhibit no evidence of Paget disease yet harbor one of the mutations.[108-111] For example, in New Zealand only four in 23 children with the same mutation inherited from a parent with Paget disease had bone scan evidence of the disorder at an average age of 44 years.[110] Similarly, in the Netherlands only 7.1% of individuals with *SQSTM1* mutations developed Paget disease when followed over a 15-year period.[111] In the past 10 years, 14 new susceptibility loci for Paget disease have been discovered.[112] These are likely to influence bone metabolism (*CSF1,* TNFRSF11A, *PML, TM7SF4, UCMA/GRP, DKK1, CTHRC1, OPTN, RIN3, hnRNPA2B7, FKBP5, ZNF687, BER,* and *C9ORF72* hexanucleotide repeat expansion frequency). Also, it has been indicated that patients who harbor *SQTM1* mutations plus mutation at one of these loci may have a more severe phenotype.[113] In Italy two groups have reported loss-of-function mutations in the profilin 1 gene that were associated with the early onset of a polyostotic Paget-like disease in family members.[114,115] In one kindred, sarcomas were a complication in three out of 11 family members.[114] In the other family, one out of eight patients developed a sarcoma, and there was considerable difficulty in successfully treating the patients with standard drug therapy, including multiple doses of zoledronate[115]

A rare syndrome of Paget disease, inclusion body myositis and frontotemporal dementia, also has a genetic association. Mutations in valosine-containing protein (VCP) are found in this syndrome.[116] This protein also contains a ubiquitin-binding domain and has effects on autophagy. As in the case of patients with classical familial Paget disease who have *SQTM1* mutations, there is incomplete penetrance of the skeletal disorder in individuals who have a VCP mutation; approximately 50% are affected.[116] Several rare heritable osteolytic disorders superficially resemble Paget disease.[117] These are caused by mutations that affect RANK or OPG (see Chapter 59).

Slow Virus Infection. The presence of osteoclast nuclear and cytoplasmic microfilaments essentially identical in structure to nucleocapsids of the *Paramyxoviridae* virus family initially suggested the possibility that Paget disease was caused by a slow virus infection.[20,21] Subsequent immunochemical studies[23,24] and sequence analysis of *MVNP* transcripts[25-27] have added support to the hypothesis, although not all studies have positively demonstrated a viral presence.[118,119] Further evidence for a role of measles virus in the pathogenesis of the disorder is indicated by *in vitro* studies carried out on pelvic bone marrow aspirations from patients with Paget disease. Osteoclasts generated in culture from osteoclast precursors exhibit a "pagetic phenotype" that distinguishes them from normal osteoclasts.[120] There are increased numbers generated, increased nuclei/osteoclast, increased bone resorption capacity/osteoclast, increased 1,25$(OH)_2D_3$ sensitivity, increased TAF-12 expression, and increased RANKL and tumor necrosis factor (TNF) sensitivity. They also secrete high levels of IL-6. When normal osteoclasts were transduced with MVNP, they were transformed into cells with the "pagetic phenotype."[121] Finally, when antisense-*MVNP* was tested on osteoclasts generated from the marrow of patients expressing MVNP, the "pagetic phenotype" was abolished.[122]

Animal Models of Paget Disease

Transgenic mice have been used to further address the pathogenesis of Paget disease.[122-127] In two studies, the most common *SQSMT1* mutation found in familial Paget disease (*SQSTM1/p62P394L*) was inserted into transgenic mice. In the first study, increased osteoclastic activity without increased bone formation was reported in lumbar vertebrae,[124] while in the second study generalized osteoclastic activity was found in the femur and tibia, with small areas of osteoblastic activity at the ends of these long bones.[127] In both studies, osteoclast precursors were hyperresponsive to RANKL, but in the first study the precursors were not hyperresponsive to 1,25$(OH)_2D_3$, a feature of pagetic osteoclasts. This was not tested in the second study. Transgenic mice were also generated with VCP mutations in two studies.[125,126] In both studies, the dominant lesions in the long bones were severe osteopenia with occasional hypersensitivity to TNF or RANKL but were not assessed for 1,25$(OH)_2D_3$ sensitivity. In a recent study, global deletion of optineurin in mice altered type 1 interferon signaling and resulted in increasing bone resorption and bone lesions resembling Paget disease.[128] Targeting the *MVNP* gene to cells of the osteoclast lineage in transgenic mice produced bone lesions in the vertebrae similar to those of Paget disease.[123] Transgenic mice coexpressing both MVNP and the *SQSTM1/p62P394L* mutation developed a particularly severe form of Paget disease.[122] The first animal model to examine the mechanism of increased bone formation in pagetic lesions used transgenic mice that expressed MVNP.[129] It was shown that IL-6 was induced in osteoclasts, which produced an increase in IGF1 expression and enhanced EphB2/EphB4 expression, a likely stimulus of osteoblasts.

There is general agreement that genes play a role in determining susceptibility to developing Paget disease, but family studies clearly demonstrate that a genetic abnormality alone is not sufficient to produce the disorder. Although the role of viruses remains controversial, the observation that Paget disease is less common than in the past, coupled with the fact that measles vaccine first became available 57 years ago, lends further support for a viral role in this disorder.[130]

- For your free Expert Consult eBook with bibliographic citations as well as the ability to take notes, highlight important content, search the full text, and more, visit http://www.ExpertConsult.Inkling.com.

REFERENCES

1. Eekhoff EMW, van der Klift M, Kroon HM, et al. Paget's disease of bone in the Netherlands: a population-based radiological and biochemical survey; the Rotterdam study. *J Bone Miner Res.* 2004;19:566–570.
2. Schmorl G. Über Osteitis deformans Paget. *Virchows Arch.* 1932;283:694–751.
3. Collins DH. Paget's disease of bone: incidence and subclinical forms. *Lancet.* 1956;2:51–57.
4. Pygott F. Paget's disease of bone: the radiological incidence. *Lancet.* 1956;1:1170–1171.
5. Barker DJP, Chamberlain AT, Guyer PB, et al. Paget's disease of bone: the Lancashire focus. *Br Med J.* 1980;1:1105–1107.

6. Detheridge FM, Guyer PB, Barker DJP. European distribution of Paget's disease of bone. *Br Med J*. 1982;285:1005–1008.
7. Altman RD, Bloch DA, Hochberg MC, et al. Prevalence of pelvic Paget's disease of bone in the United States. *J Bone Miner Res*. 2000;15:461–465.
8. Hashimoto J, Ohno I, Nkatsuka K, et al. Prevalence and clinical features of Paget's disease of bone in Japan. *J Bone Miner Metab*. 2006;24:186–190.
9. Corral-Gudino L, Borao-Cengotita-Bengoa M, Del Pino-Montes J, et al. Epidemiology of Paget's disease of bone: a systematic review and meta-analysis of secular changes. *Bone*. 2013;55:347–352.
10. McKusick VA. *Heritable Disorders of Connective Tissue*. St. Louis: Mosby; 1972.
11. Siris ES, Ottman R, Flaster E, et al. Familial aggregation of Paget's disease of bone. *J Bone Miner Res*. 1991;6:495–500.
12. Sofoer JA, Holloway SM, Emery AEH. A family study of Paget's disease of bone. *J Epidemiol Community Health*. 1983;37:226–231.
13. Posen S. Paget's disease: current concepts. *Aust N Z J Surg*. 1992;62:17–23.
14. Morales-Piga AA, Rey-Rey JS, Corres-Gonzales J, et al. Frequency and characteristics of familial aggregation of Paget's disease of bone. *J Bone Miner Res*. 1995;10:663–670.
15. O'Driscoll JB, Anderson DC. Past pets and Paget's disease. *Lancet*. 1985;2:919–921.
16. Spencer H, O'Sullivan V, Sontag SJ. Does lead play a role in Paget's disease of bone? A hypothesis. *J Lab Clin Med*. 1992;120:798–800.
17. Adachi JD, Arlen D, Webber CE, et al. Is there any association between the presence of bone disease and cumulative exposure to lead? *Calcif Tissue Int*. 1998;63:429–432.
18. Audet MC, Jean S, Beaudoin C, et al. Environmental factors associated with familial or non-familial forms of Paget's disease of bone. *Joint Bone Spine*. 2017;84:719–723.
19. Rubinstein MA, Smelin A, Freedman AL. Osteoblasts and osteoclasts in bone marrow aspiration. *Arch Intern Med*. 1953;92:684–696.
20. Rebel A, Malkani K, Basle M. Anomalies nucleaires de la maladie osseuse de Paget. *Nouv Presse Med*. 1974;3:1299–1301.
21. Mills BG, Singer FR. Nuclear inclusions in Paget's disease of bone. *Science*. 1976;194:201–202.
22. Singer FR. Paget's disease of bone: possible viral basis. *Trends Endocrinol Metab*. 1996;7:258–261.
23. Rebel A, Basle M, Pouplard A, et al. Viral antigens in osteoclasts from Paget's disease of bone. *Lancet*. 1980;2:344–346.
24. Mills BG, Singer FR, Weiner LP, et al. Evidence for both respiratory syncytial virus and measles virus antigens in the osteoclasts of patients with Paget's disease of bone. *Clin Orthop*. 1984;183:303–311.
25. Basle MF, Fournier JG, Rozenblatt S, et al. Measles virus RNA detected in Paget's disease bone tissue by *in situ* hybridization. *J Gen Virol*. 1986;67:907–913.
26. Gordon MT, Sharpe PT, Anderson DC. Canine distemper virus localised in bone cells of patients with Paget's disease. *Bone*. 1991;12:195–201.
27. Friedrichs WE, Reddy SV, Bruder JM, et al. Sequence analysis of measles virus nucleocapsid transcripts in patients with Paget's disease. *J Bone Miner Res*. 2002;17:145–151.
28. Rhodes BA, Greyson ND, Hamilton Jr CR, et al. Absence of anatomic arteriovenous shunts in Paget's disease of bone. *N Engl J Med*. 1972;287:686–689.
29. Schuller A. Ueber circumscripte Osteoporose des Schädels. *Med Klin*. 1929;25:631–632.
30. Maldague B, Malghem J. Dynamic radiologic patterns of Paget's disease of bone. *Clin Orthop*. 1987;217:126–151.
31. Jacobs P. Osteolytic Paget's disease. *Clin Radiol*. 1974;25:137–144.
32. Belanger LF, Jarry L, Uhthoff HK. Osteocytic osteolysis in Paget's disease. *Rev Can Biol*. 1968;27:37–44.
33. Meunier PJ, Coindre JM, Edouard CM, et al. Bone histomorphometry in Paget's disease: quantitative and dynamic analysis of pagetic and nonpagetic bone tissue. *Arthritis Rheum*. 1980;23:1095–1103.
34. Jaffe HL. *Metabolic, Degenerative and Inflammatory Diseases of Bones and Joints*. Philadelphia: Lea & Febiger; 1972.
35. Paget J. On a form of chronic inflammation of bones (osteitis deformans). *Med Chir Trans*. 1877;60:37–64.
36. Vellenga CJLR, Bijvoet OLM, Pauwels EKJ. Bone scintigraphy and radiology in Paget's disease of bone: a review. *Am J Physiol*. 1988;3:154–168.
37. Waxman AD, McKee D, Siemsen JK, et al. Gallium scanning in Paget's disease of bone: effect of calcitonin. *Am J Roentgenol*. 1980;134:303–306.
38. Mills BG, Masuoka LS, Graham Jr CC, et al. Gallium-67 citrate localization in osteoclast nuclei of Paget's disease of bone. *J Nucl Med*. 1988;29:1083–1087.
39. Nager GT. Paget's disease of the temporal bone. *Ann Otol Rhinol Laryngol*. 1975;84:1–32.
40. Monsell EM, Cody DD, Bone HG, et al. Hearing loss in Paget's disease of bone: the relationship between pure tone thresholds and mineral density of the cochlear capsule. *Hearing Res*. 1995;83:114–120.
41. Blotman F, Blard JM, Labauge R, et al. Exploration ultrasonique de la circulation encephalique chez le Pagetique. *Rev Rhum*. 1975;42:647–651.
42. Gold DT, Boisture J, Shipp KM, et al. Paget's disease of bone and quality of life. *J Bone Miner Res*. 1996;11:1897–1904.
43. Smith NHH. Monostotic Paget's disease of the mandible presenting with progressive resorption of the teeth. *Oral Surg Oral Med Oral Pathol*. 1978;46:246–253.
44. Hadjipavlou A, Lander P. Paget's disease of the spine. *J Bone Joint Surg Am*. 1991;73:1376–1381.
45. Douglas DL, Duckworth T, Kanis JA, et al. Spinal cord dysfunction in Paget's disease of bone. *J Bone Joint Surg Br*. 1981;63:495–503.
46. Samuels MA, Schiller AL. Case records of the Massachusetts General Hospital. *N Engl J Med*. 1981;304:1411–1421.
47. Lander P, Hadjipavlou A. Intradiscal invasion of Paget's disease of the spine. *Spine*. 1991;16:46–51.
48. Altman RD, Collins B. Musculoskeletal manifestations of Paget's disease of bone. *Arthritis Rheum*. 1980;23:1121–1127.
49. Franck WA, Bress NM, Singer FR, et al. Rheumatic manifestations of Paget's disease of bone. *Am J Med*. 1974;56:592–603.
50. Dove J. Complete fractures of the femur in Paget's disease of bone. *J Bone Joint Surg Br*. 1980;62:12–17.
51. Hadjipavlou A, Lander P, Srolovitz H, et al. Malignant transformation in Paget disease of bone. *Cancer*. 1992;70:2802–2808.
52. Haibach H, Farrell C, Dittrich BS. Neoplasms arising in Paget's disease of bone: a study of 82 cases. *Am J Clin Pathol*. 1985;83:594–601.
53. Choquette D, Haraoui B, Altman RD, et al. Simultaneous multifocal sarcomatous degeneration in Paget's disease of bone. *Clin Orthop*. 1983;179:308–311.
54. Seret P, Basle MF, Rebel A, et al. Sarcomatous degeneration in Paget's bone disease. *J Cancer Res Clin Oncol*. 1987;113:392–399.
55. Huvos AG. Osteogenic sarcoma of bones and soft tissues in older persons. *Cancer*. 1986;57:1442–1449.
56. Singer FR, Mills BG. Giant cell tumor in Paget's disease of bone: recurrence after 36 years. *Clin Orthop*. 1993;293:293–301.
57. Gianfrancesco F, Rendina D, Merlotti D, et al. Giant cell tumor occurring in familial Paget's disease of bone: report of clinical characteristics and linkage analysis of a large pedigree. *J Bone Miner Res*. 2013;28:341–350.
58. Upchurch KS, Simon LS, Schiller AL, et al. Giant cell reparative granuloma of Paget's disease of bone: a unique clinical entity. *Ann Intern Med*. 1983;98:35–40.
59. Ziambaras K, Totty WA, Teitelbaum SL, et al. Extraskeletal osteoclastomas responsive to dexamethasone treatment in Paget bone disease. *J Clin Endocrinol Metab*. 1997;82:3826–3834.
60. Cosso R, Nuzzo V, Zuccoli A, et al. Giant cell tumor in a case of Paget's disease of bone: an aggressive benign tumor exhibiting a quick response to an innovative therapeutic agent. *Clin Cases Miner Bone Metab*. 2010;7:145–152.
61. Nagant de Deuxchaisnes CN, Krane SM. Paget's disease of bone: clinical and metabolic observations. *Medicine (Baltimore)*. 1964;43:233–266.
62. Alexandersen P, Peris P, Guanabens N, et al. Non-isomerized C-telopeptide fragments are highly sensitive markers for monitoring disease activity and treatment efficacy in Paget's disease of bone. *J Bone Miner Res*. 2005;20:588–595.
63. Kay HD. Plasma phosphatase in osteitis deformans and in other disease of bone. *Br J Exp Pathol*. 1929;10:253–256.

64. Woodard HQ. Long term studies of the blood chemistry in Paget's disease of bone. *Cancer*. 1959;12:1226–1237.
65. Alvarez L, Guanabens N, Peris P, et al. Usefulness of biochemical markers of bone turnover in assessing response to the treatment of Paget's disease. *Bone*. 2001;29:447–452.
66. Papapoulos SE, Frolich M, Mudde AH, et al. Serum osteocalcin in Paget's disease of bone: basal concentrations and response to bisphosphonate treatment. *J Clin Endocrinol Metab*. 1987;65:189–194.
67. Naot D, Bava U, Matthews B, et al. Differential gene expression in cultured osteoblasts and bone marrow stromal cells from patients with Paget's disease of bone. *J Bone Miner Res*. 2007;22:298–309.
68. Kanis JA, Heynen G, Walton RJ. Plasma calcitonin in Paget's disease of bone. *Clin Sci*. 1977;52:329–332.
69. Chapuy MC, Zucchelli P, Meunier PJ. Parathyroid function in Paget's disease of bone. *Bone Miner Electrolyte Metab*. 1981;6:112–118.
70. Siris ES, Clemens TP, McMahon D, et al. Parathyroid function in Paget's disease of bone. *J Bone Miner Res*. 1989;4:75–79.
71. Guillard-Cumming DF, Beard DJ, Douglas DL, et al. Abnormal vitamin D metabolism in Paget's disease of bone. *Clin Endocrinol*. 1985;22:559–566.
72. Castro-Errecaborde N, de la Piedra C, Rapado A, et al. Correlation between serum osteocalcin and 24,25-dihydroxyvitamin D levels in Paget's disease of bone. *J Clin Endocrinol Metab*. 1991;72:462–466.
73. Roodman GD, Kurihara N, Ohsaki Y, et al. Interleukin 6: a potential autocrine/paracrine factor in Paget's disease of bone. *J Clin Invest*. 1992;89:46–52.
74. Schweitzer DH, Oostendorp-Van de Ruit M, van der Plujim G, et al. Interleukin-6 and the acute phase response during treatment of Paget's disease with the nitrogen-containing bisphosphonate dimethylamino-hydroxypropylidene bisphosphonate. *J Bone Miner Res*. 1995;10:956–962.
75. Rendina D, Postiglione L, Vuotto P, et al. Clodronate treatment reduces serum levels of interleukin-6 soluble receptor in Paget's disease of bone. *Clin Exp Rheumatol*. 2002;20:359–364.
76. Neale SD, Schulze E, Smith R, et al. The influence of serum cytokines and growth factors on osteoclast formation in Paget's disease. *Q J Med*. 2002;95:233–240.
77. Martini G, Gennari L, Merlotti D, et al. Serum OPG and RANKL levels before and after intravenous bisphosphonate treatment in Paget's disease of bone. *Bone*. 2007;40:457–463.
78. Rendina D, DeFillipo GD, Merlottu D, et al. Increased prevalence of nephrolithiasis and hyperoxaluria in Paget's disease of bone. *J Clin Endocrinol Metab*. 2020;105:dgaa576.
79. Reifenstein Jr EC, Albright F. Paget's disease: its pathologic physiology and the importance of this in the complications arising from fracture and immobilization. *N Engl J Med*. 1944;231:343–355.
80. Gutteridge DH, Gruber HE, Kermode DG, et al. Thirty cases of concurrent Paget's disease and primary hyperparathyroidism: sex distribution, histomorphometry, and prediction of the skeletal response to parathyroidectomy. *Calcif Tissue Int*. 1999;65:427–435.
81. Rosenkrantz JA, Gluckman EC. Coexistence of Paget's disease of bone and multiple myeloma. *Am J Roentgenol*. 1957;78:30–38.
82. Lluberas-Acosta G, Hansell JR, Schumacher Jr HR. Paget's disease of bone in patients with gout. *Arch Intern Med*. 1986;146:2389–2392.
83. Fennelly JJ, Hogan A. Pseudouridine excretion: a reflection of high RNA turnover in Paget's disease. *Ir J Med Sci*. 1972;141:103–107.
84. Arnalich F, Plaza I, Sobrino JA, et al. Cardiac size and function in Paget's disease of bone. *Int J Cardiol*. 1984;5:491–505.
85. Morales-Piga AA, Moya JL, Bachiller FJ, et al. Assessment of cardiac function by echocardiography in Paget's disease of bone. *Clin Exp Rheumatol*. 2000;18:31–37.
86. Strickberger SA, Schulman SP, Hutchins GM. Association of Paget's disease of bone with calcific aortic valve disease. *Am J Med*. 1987;82:953–956.
87. Hultgren HN. Osteitis deformans (Paget's disease) and calcific disease of the heart valves. *Am J Cardiol*. 1998;81:1461–1464.
88. Harrison CV, Lennox B. Heart block in osteitis deformans. *Br Heart J*. 1948;10:167–176.
89. Laroche M, Delmotte A. Increased arterial calcification in Paget's disease of bone. *Calcif Tissue Int*. 2005;77:129–133.
90. Singer FR, Krane SM. Paget's disease of bone. In: Avioli LV, Krane SM, eds. *Metabolic Bone Disease and Clinically Related Disorders*. San Diego: Academic Press; 1990:545–605.
91. Rogers MJ. New insights into the molecular mechanisms of action of bisphosphonates. *Curr Pharm Des*. 2003;9:2643–2658.
92. Khairi MRA, Altman RD, DeRosa GP, et al. Sodium etidronate in the treatment of Paget's disease of bone: a study of long-term results. *Ann Intern Med*. 1977;87:656–663.
93. Siris E, Weinstein RS, Altman R, et al. Comparative study of alendronate versus etidronate for the treatment of Paget's disease of bone. *J Clin Endocrinol Metab*. 1996;81:961–967.
94. Reginster JY, Calson F, Morlock G, et al. Efficacy and safety of oral tiludronate in Paget's disease of bone: a double-blind, multiple-dosage, placebo-controlled study. *Arthritis Rheum*. 1992;35:967–974.
95. Siris ES, Chines AA, Altman RD, et al. Risedronate in the treatment of Paget's disease of bone: an open label multicenter study. *J Bone Miner Res*. 1998;13:1032–1038.
96. Thiebaud D, Jaeger P, Gobelet C, et al. A single infusion of the bisphosphonate AHPrBP (APD) as treatment of Paget's disease of bone. *Am J Med*. 1988;85:207–212.
97. Reid IR, Miller P, Lyles K, et al. Comparison of a single infusion of zoledronic acid with risedronate for Paget's disease. *N Engl J Med*. 2005;353:898–908.
98. Bertoldo BF, Pancheri S, Zenari S, et al. Serum 25-hydroxyvitamin D levels modulate the acute-phase response associated with the first nitrogen-containing bisphosphonate infusion. *J Bone Miner Res*. 2010;25:447–454.
99. Boonen S, Sellmayer DE, Lippuner K, et al. Renal safety of annual zoledronic acid infusions in osteoporotic postmenopausal women. *Kidney Int*. 2008;74:641–648.
100. Rizzoli R, Burlet N, Cahall D, et al. Osteonecrosis of the jaw and bisphosphonate treatment for osteoporosis. *Bone*. 2008;42:841–847.
101. Reid IR, Lyles K, Su G, et al. A single infusion of zoledronic acid produces sustained remissions in Paget disease: data to 6.5 years. *J Bone Miner Res*. 2011;26:2261–2270.
102. Schwarz P, Rasmussen AQ, Kvist TM, et al. Paget's disease of bone after treatment with denosumab: a case report. *Bone*. 2012;50:1023–1025.
103. Sochart DH, Porter ML. Charnley low-friction arthroplasty for Paget's disease of the hip. *J Arthroplasty*. 2000;15:210–219.
104. Lee GC, Sanchez-Sotelo J, Berry DJ. Total knee arthroplasty in patients with Paget's disease at the knee. *J Arthroplasty*. 2005;20:689–693.
105. Meyers M, Singer F. Osteotomy for tibia vara in Paget's disease under cover of calcitonin. *J Bone Joint Surg Am*. 1978;60:810–814.
106. Laurin N, Brown JP, Morissette J, et al. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget's disease of bone. *Am J Hum Genet*. 2002;70:1582–1588.
107. Michou L, Morissette J, Gagnon ER, et al. Novel SQSTM1 mutations in patients with Paget's disease of bone in an unrelated multiethnic American population. *Bone*. 2011;48:456–460.
108. Morissette J, Laurin N, Brown JP. Sequestosome 1: mutation frequencies, haplotypes, and phenotypes in familial Paget's disease of bone. *J Bone Miner Res*. 2006;21:P38–P44.
109. Johnson-Pais TL, Wisdom JH, Weldon KS, et al. Three novel mutations in SQSTM1 identified in familial Paget's disease of bone. *J Bone Miner Res*. 2003;18:1748–1753.
110. Bolland MJ, Tong PC, Naot D, et al. Delayed development of Paget's disease in offspring inheriting *SQSTM 1* mutations. *J Bone Miner Res*. 2007;22:411–415.
111. Peeters JJM, DeRidder R, Hamoen EC, et al. Long-term followup of unaffected relatives with and without sequestosome 1 mutations. *Bone*. 2019;128:115044.
112. Singer FR. Paget's disease of bone. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext (Internet)*. Bethesda, MD: NIH/NLM/MNCI books; 2020.
113. Albagha OME, Visconti MR, Alonso N, et al. Common susceptibility alleles and SQSTM1 mutations predict disease extent and severity in a multinational study of patients with Paget's disease. *J Bone Miner Res*. 2013;28:2338–2346.

114. Scotto di Carlo F, Pazzaglia L, Esposito T, et al. The loss of profilin1 causes early onset Paget's disease of bone. *J Bone Miner Res*. 2020;35:1387–1398.
115. Merlotti D, Materozzi M, Bianciardi S, et al. Mutation of PFN1 gene in early onset, polyostotic Paget-like disease. *J Clin Endocrinol Metab*. 2020;105:2553–2565.
116. Mehta SG, Khare M, Simon M, et al. Genotype-phenotype studies of VCP-associated inclusion body myopathy with Paget's disease of bone and/or frontotemporal dementia. *Clin Genet*. 2013;83:422–431.
117. Ralston SH. Juvenile Paget's disease, familial expansile osteolysis and other genetic osteolytic disorders. *Best Pract Res Clin Rheumatol*. 2008;22:101–111.
118. Helfrich MH, Hobson RP, Grabowski PS, et al. A negative search for a paramyxoviral etiology of Paget's disease of bone: molecular, immunological, and ultrastructural studies in UK patients. *J Bone Miner Res*. 2000;15:2315–2329.
119. Matthews BG, Afzal MA, Minor PD, et al. Failure to detect measles virus ribonucleic acid in bone cells from patients with Paget's disease. *J Clin Endocrinol Metab*. 2008;93:1398–1401.
120. Singer FR, Roodman GD. Paget's disease of bone: pathogenesis and treatment. In: Bronner F, Farach-Carson M, Roach HI, eds. *Bone-Metabolic Functions and Modulators*. London: Springer; 2012:197–215.
121. Kurihara N, Reddy SV, Menaa C, et al. Osteoclasts expressing the measles virus nucleocapsid gene display a pagetic phenotype. *J Clin Invest*. 2000;105:607–614.
122. Kurihari N, Hiruma Y, Yamana K, et al. Contributions of the measles virus nucleocapsid gene and the SQSTM1/p62^{P392L} mutation to Paget's disease. *Cell Metab*. 2011;13:23–34.
123. Kurihara N, Zhou H, Reddy SV, et al. Expression of measles virus nucleocapsid protein in osteoclasts induces Paget's disease-like bone lesions in mice. *J Bone Miner Res*. 2006;21:446–455.
124. Kurihara N, Hiruma Y, Zhou H, et al. Mutation of the sequestosome 1 (p62) gene increases osteoclastogenesis but does not induce Paget disease. *J Clin Invest*. 2007;117:133–142.
125. Daroszewska A, van't Hof RJ, Rojas JA, et al. A point mutation in the ubiquitin-associated domain of SQSMT1 is sufficient to cause a Paget's disease-like disorder in mice. *Hum Mol Genet*. 2011;20:2734–2744.
126. Custer SK, Neumann M, Lu H, et al. Transgenic mice expressing mutant forms VCP/p97 recapitulate the full spectrum of IBMPFD including degeneration in muscle, brain and bone. *Hum Mol Genet*. 2010;19:1741–1755.
127. Badadani M, Nalbandian A, Watts GD, et al. VCP associated inclusion body myopathy and Paget disease of bone knock-in mouse model exhibits tissue pathology typical of human disease. *PLoS One*. 2010;5:1–15.
128. Wong SW, Huang BW, Hu X, et al. Global deletion of optineurin results in altered type 1 IFN signaling and abnormal bone remodeling in a model of Paget's disease. *Cell Death Differ*. 2020;27:71–84.
129. Miyagawa K, Ohata Y, Delgado-Calle J, et al. Osteoclast-derived IGF1 is required for pagetic lesion formation in vivo. *JCI Insight*. 2020;5:e133113.
130. Singer FR. Paget's disease of bone-genetic and environmental factors. *Nat Rev Endocrinol*. 2015;11:662–671.

64

Anatomy and Development of the Thyroid

Mario De Felice and Roberto Di Lauro

OUTLINE

The Anatomy Of The Thyroid Gland, 1080
- Gross Anatomy, 1080
- Blood Supply, 1080
- Lymphatics, 1080
- Innervation, 1081
- Anatomic Variants, 1081

Thyroid Follicular Cells And Thyroid Follicles, 1081

Other Thyroid Cells, 1083
- C Cells, 1083
- Ultimobranchial Body-Derived Epithelial Cells, 1084
- Thyroid Stem Cells, 1084

The Development Of The Thyroid Gland, 1084
- Specification of the Thyroid Follicular Cells: The Thyroid Anlage and the Thyroid Placode, 1085
- Early Stages of Thyroid Morphogenesis: Budding, Migration, and Lobulation of Thyroid Primordium, 1085
- Late Stages of Thyroid Morphogenesis: Functional Differentiation and Histogenesis, 1087
- Assembling the Gland: Development and Merging of the Ultimobranchial Bodies, 1088
- Differentiation of C Cells, 1089

Molecular Genetics Of Thyroid Gland Development, 1089
- Functional Genomics of Thyroid-Enriched Transcription Factors, 1089
- From Animal Models to Human Diseases, 1094
- Generation of Functional Thyroid Follicles *In Vitro*, 1095

Phylogenesis Of The Thyroid Gland, 1096

THE ANATOMY OF THE THYROID GLAND

Gross Anatomy

The thyroid gland was first described by Galen (130–210 A.D.) in his work “De Voce.” The gland was named “thyroid” by Thomas Whorton (1614–1673), due its proximity to the thyroid cartilage. Despite its name (thyreòs in Greek means “shield”; also, the German name “Schilddrüse” means “shield gland”), the characteristic shape of the thyroid, consisting of two lateral lobes connected by a narrow isthmus, is more reminiscent of a butterfly or a capital H than a shield (Fig. 64.1). The lateral lobes are 3 to 4 cm long and 15 to 20 mm wide and are located between the larynx and the trachea medially and the carotid sheath and the sternomastoid muscles laterally. The upper pole of the lobes reaches the level of the thyroid cartilage, while the lower pole reaches the V to VI tracheal ring. The isthmus is 12 to 20 mm long and 20 mm wide and crosses the trachea between the I and II rings. In a normal adult the entire gland is approximately 6 to 7 cm wide and 3 to 4 cm long, and its weight ranges between 15 and 25 g. The thyroid is generally asymmetrical: the right lobe may be twice as large as the left, extending higher and lower in the neck than the left pole. Interestingly, in patients with dextrocardia, the position of the lobes is reversed, suggesting that the asymmetry in the thyroid lobes could be due to the position of the heart. A thin connective capsule encloses the thyroid. Fibrous septa are occasionally detached from this capsule and penetrate into the parenchyma to produce an incomplete lobulation. This inner capsule is connected to an outer capsule (also called the false capsule of the thyroid) that is continuous with the pretracheal fascia. Vessels, the parathyroids, and the recurrent laryngeal nerves are located in the space between the two capsules and are in close contact with the thyroid: the parathyroids on the posterior surface of the gland, and the recurrent laryngeal nerves just medial to the lateral lobes.

Blood Supply

The thyroid is a highly vascularized organ: four arteries provide the gland with an abundant blood supply. Frequent anastomoses among these vessels and an arteriolar network are present on the surface of the gland; from this network, small arteries arborize and enter deeply into the tissue. The capillaries are localized in the interfollicular connective tissue, forming a basket-like network that surrounds each follicle. The capillary endothelial cells are fenestrated, like those of other endocrine glands. Each fenestration is approximately 50 nm in diameter. Stimulation with thyroid-stimulating hormone (TSH) increases the number and density of the fenestrations. The veins emerge from the thyroid parenchyma and form a plexus of three groups of veins: superior, middle, and inferior thyroid veins.

Lymphatics

A rich plexus of lymphatic capillaries surrounds the thyroid follicles and communicates with small lymphatic vessels found in the interlobular connective tissue. These deep vessels give rise to a surface network of lymphatics that drain into several groups of nodes. The uppermost group of nodes is situated just above the thyroid isthmus and is a constant group of one to five nodes called the Delphian nodes. These are readily felt when involved in cancer or Hashimoto thyroiditis. Other variable nodes in the thyroid area include the pretracheal nodes below the isthmus, those on the thyroid surface, and a last group found along the lateral vein, the recurrent laryngeal nerve, or the carotid sheath.

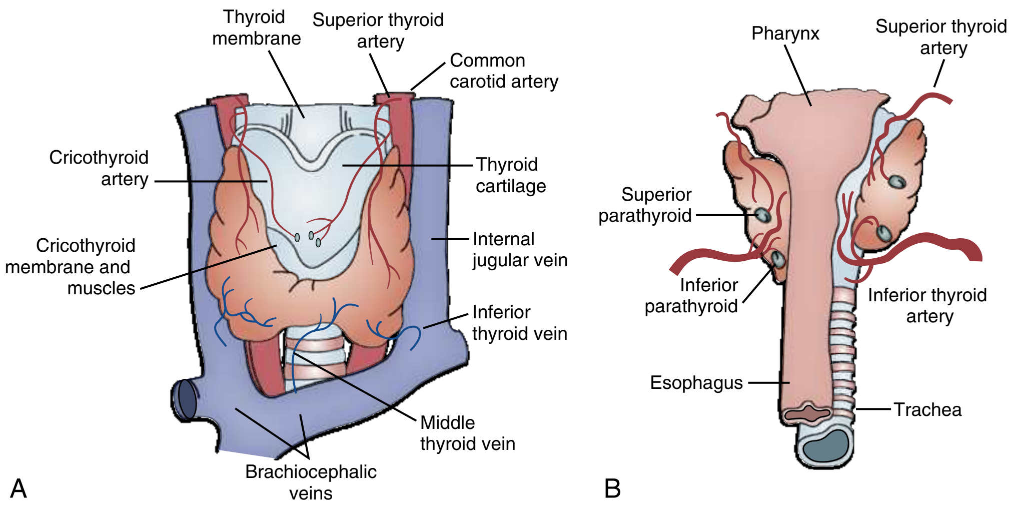


Fig. 64.1 Gross anatomy of the thyroid. A, Anterior view. B, Posterior view.

Innervation

The innervation of the thyroid is provided by sympathetic, parasympathetic, and peptidergic fibers, although few fibers enter the gland. Both sympathetic and parasympathetic fibers extend throughout the tissue among the follicles in close relation to the follicular cells or around a blood vessel.

Some neuropeptides, such as vasoactive intestinal peptide, neuropeptide Y, substance P, or galanin, are exclusively produced in the nerve fibers distributed throughout the thyroid. Presumably, these neuropeptides regulate follicular cell functions via a paracrine pathway. Thyroid innervation is very poor compared to that of other endocrine glands.

Anatomic Variants

The anatomical variants most frequently described in healthy individuals are caused by defects in the regression of the thyroglossal duct. This group of anomalies is generally characterized by the presence of an accessory lobe (pyramidal lobe) attached to the upper part of the isthmus of the thyroid (15% of the population). The pyramidal lobe is a rostral-directed stalk that results from the retention and growth of the caudal end of the thyroglossal duct. Other anomalies are associated with a defective atrophy of the duct. In some instances, the entire thyroglossal duct persists as an epithelial cord connecting the foramen cecum of the tongue to the larynx; in other cases, remainders of the duct form isolated or multiple cysts that are found along the line of descent of the duct. Persistent portions of the thyroglossal duct may differentiate into thyroid tissue and form structures called accessory thyroids. A different developmental defect results in the formation of an ectopic thyroid gland (see later, Late Stages of Thyroid Morphogenesis: Functional Differentiation and Histogenesis). The other frequent anatomical variant (5% of the population) is a thyroid gland that has not developed as a unique mass. These variants include the absence of the isthmus (the thyroid consists of two independent lateral lobes, a physiological condition in nonmammalian vertebrates) or the absence of a significant portion of a lateral lobe, frequently the lower half of the left lobe.

KEY POINTS

- The functional unit of the thyroid gland is the thyroid follicle, surrounded by a network of capillaries. Several of these units make up the thyroid gland. The assembly of the units determines the amount, not the type, of activity.

THYROID FOLLICULAR CELLS AND THYROID FOLLICLES

The thyroid gland displays a peculiar highly-organized architecture characterized by the presence of spheroidal structures known as follicles. The thyroid follicles, supported by an interfollicular extracellular matrix and a capillary network, are composed of a single layer of epithelial cells, the thyroid follicular cells (TFCs), that surround a closed cavity, the follicular lumen, which is filled with colloid, a concentrated solution of thyroglobulin. TFCs represent the largest cell population of the thyroid gland and are responsible for thyroid hormone synthesis. When observed under the light microscope TFCs show a neutrofilic cytoplasm, a basal nucleus, and periodic acid-Schiff–positive vacuoles (phagosomes). Follicular cells appear as cuboidal epithelial cells, whose height is approximately 15 μm. Cells become flatter (squamous) or higher (columnar) depending on whether or not they undergo TSH stimulation. Electron microscopy (Fig. 64.2) reveals the characteristic features of cells actively engaged in protein synthesis, with the rough endoplasmic reticulum and the Golgi apparatus as the dominant organelles in the cell. Furthermore, several vesicles are localized in the apical or subapical cytoplasm. Smaller (150–200 nm) vesicles are exocytotic vesicles containing newly synthesized thyroglobulin. The fusion of these vesicles with the apical plasmamembrane leads to the delivery of thyroglobulin into the follicle lumen. The larger vesicles (500–4000 nm) are filled with dense material, called colloid droplets, and are the result of the uptake of the iodinated thyroglobulin stored in the follicular lumen. The reabsorption of the colloid involves a macropinocytosis mechanism whose first step is the formation of pseudopods at the apical pole. The pseudopods close, and a portion of the colloid is internalized into the cell.

At the molecular level (Fig. 64.3) a follicular cell can be identified by the presence of a specific set of proteins—and the corresponding mRNAs—that are indispensable for its specialized functions. Among such proteins, thyroglobulin and thyroperoxidase are remarkably specific, being detectable exclusively in TFCs. Other proteins, such as TSH receptor (Tshr), sodium/iodide symporter (NIS), pendrin, and Duox/DuoxA complex, though expressed at high level in TFCs, are also present in few other tissues. The exclusive or prevalent expression of genes necessary for thyroid hormone biosynthesis in TFCs appears to be due to a combination of transcription factors unique to this cell type, comprising Nkx2.1 (also called Ttf-1, Titf1, and T/EBP) Pax8, and Foxe1. The molecular features and the functions of these factors will be described in the section on Functional Genomics of Thyroid-Enriched Transcription Factors.

A distinctive feature of follicular cells is their specific polarity (see[1] and references therein). The surface of the cells is divided by the junctional complexes between cells into two functionally distinct but physically contiguous regions: an apical domain that faces the follicle lumen and a basolateral domain orientated toward the extracellular matrix. The apical domain is characterized by the presence of apical microvilli, pseudopods and by the localization of thyroperoxidase, sodium or chloride channels, pendrin, and Duox. The basal domain is characterized by the expression of NIS, Na^+/K^+ ATPase, EGF, and TSH receptor. Thyroid hormone synthesis requires basal to apical transport of iodide and thyroglobulin; conversely, hormone secretion is based on apical to basal transport of thyroglobulin. In addition, the follicular organization allows the biochemical steps required for hormone production (secretion of proteins in the follicular lumen as exocrine cells do; reabsorption and hydrolysis of proteins; release of hormones into blood by endocrine secretion) to occur as a functional chain of events.

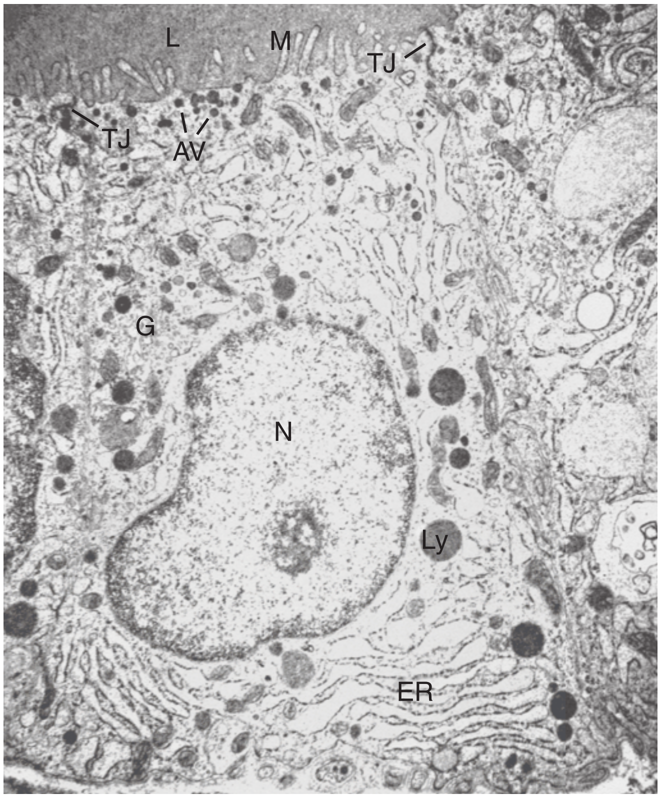


Fig. 64.2 **Electron micrograph of a rat follicular cell.** *AV,* Apical vesicles; *ER,* endoplasmic reticulum; *G,* Golgi apparatus; *L,* lumen; *Ly,* lysosome; *M,* microvilli; *N,* nucleus; *Tg,* thyroglobulin; *TJ,* tight junction. (Courtesy Professor L. Nitsch.)

In the follicles, thyroid cells form a barrier between the extrafollicular space and the lumen. A tight barrier is critical, because it promotes cell polarity, guarantees efficient transport, and prevents passive back-diffusion. The strong intercellular adhesion is mediated by a junctional complex that consists of tight junctions (zonula occludens), adherens junctions, and desmosomes. All these adhesion structures present the same overall organization: adhesive transmembrane proteins linked to the cytoskeleton by cytosolic adapter proteins. Among the different types of cell junctions, the tight junctions, located close to the apical border of the cells, are the structures that seal cells together. Tight junctions control the permeability of the paracellular space and define the boundary between the apical and basolateral domain of the follicular plasma membrane. They appear as a complex network of anastomosing fibrils consisting of junctional proteins responsible for cell–cell contact. This macromolecular complex interacts with different cellular structures. The major transmembrane proteins in a tight junction are occludin and claudins. The cytoplasmic portion of these proteins interact with intracellular peripheral membrane proteins such as ZO-1 and ZO-2, which link the junction to the microfilaments. The microtubules can also be functionally linked to the tight junctions via cytoskeletal-associated proteins, such as cingulin, 7H6 antigen, and actin.

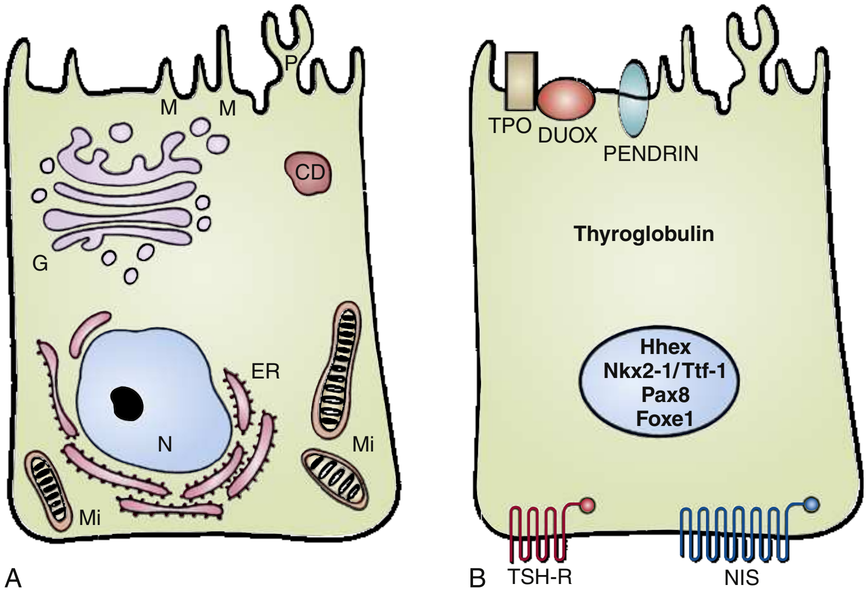


Fig. 64.3 **Thyroid follicular cell. A,** Schematic structural representation. **B,** Schematic molecular representation. *CD,* Colloid droplets; *ER,* endoplasmic reticulum; *G,* Golgi apparatus; *M,* microvilli; *MI,* mitochondrion; *N,* nucleus; P, pseudopod.

The anchoring junctions connect the cytoskeleton of a cell to the cytoskeleton of its neighbor or to the extracellular matrix. The epithelial adherens junctions are a type of anchoring junction located below the tight junctions. These junctions are formed by transmembrane proteins, adhesion proteins that belong to the cadherins family (cadherins and transmembrane calcium-dependent adhesion molecules). In thyroid tissue, as well as in other epithelial tissues, E-cadherin accumulates in adherens junctions and plays a central role in the induction of these stable adhesions forming homophilic contacts between neighboring cells. E-cadherin is linked to an intracellular anchor protein, catenin, which in turns binds to the actin cytoskeleton. The catenins participate in intracellular signal transduction pathways, creating a functional bridge that couples physical adhesion to intracellular signaling events. Desmosomal junctions, another type of anchoring junction, are the third type of junction present in epithelia. These structures occur on the plasma membrane below the adherens junction. The transmembrane proteins found in desmosomal junctions are the desmogleins and desmocollins, members of the cadherins superfamily. The cytoplasmic tails of these adhesion proteins bind intracellular proteins (plakoglobin, desmosplakin), which in turn bind to the intermediate cytoskeleton filaments. These protein–protein interactions lead to the formation of a network throughout the tissue that connects different adjacent cells. Follicular cells in the epithelia communicate via the gap junctions. Gap junctions are formed by channel forming proteins (connexins). The channels they form allow inorganic ions and other small water-soluble molecules to pass directly from the cytoplasm of one cell to another, coupling the cells both electrically and metabolically.

Primary cultures of porcine TFCs provided a useful *in vitro* model to understand the mechanisms leading to folliculogenesis. Follicular cells freshly isolated from pig thyroid glands cultured in absence of TSH aggregate only transiently. In presence of TSH, these cells assemble epithelial junctions, polarize, and organize themselves into follicle-like structures in which apical poles of cells delineate a lumen cavity. The first and essential step for *in vitro* folliculogenesis is cell aggregation mediated by E-cadherin. During the first few hours in culture, E-cadherin expression increases in the lateral cell surface and accumulates in the subapical regions, where the adherens junctions will be assembled. The earliest stage of cell surface differentiation is marked by the redistribution of two proteins, ZO-1 and Na^+/K^+ ATPase: ZO-1 is recruited around the future pole of the cell, and the ATPase is confined to the basolateral cell surface. At this stage, apical domain–associated proteins are detected in intracellular vacuoles that later fuse with the cell surface at the nascent apical pole. The follicular lumen is generated in two different steps, both consequences of the polarized phenotype of thyroid cells. The first step is triggered by the lack of adhesive properties of the apical cell surface. A second step, which is required for the control of the follicular size, is driven by a bidirectional ion transport system that secrets chloride in a basal to apical direction and absorbs sodium in an apical to basal direction as well.

The formation and maintenance of follicle-like structures in cultured porcine thyroid cells is dependent on TSH. In absence of TSH, or cyclic adenosine monophosphate (cAMP) stimulators, follicles undergo a morphological conversion: thyroid cells spread and form epithelial monolayers. The activation of the extracellular signal-regulated kinase (ERK) signaling pathway is involved in this conversion. TSH, acting through its second messenger cAMP, inhibits ERK activation and represses the epithelial conversion. In addition, TSH modulates other steps required to maintain follicle organization. Stimulation of cAMP/PKA stabilizes E-cadherin–dependent cell–cell adhesions and inhibits the production of thrombospondin 1, a matricellular protein that is a negative modulator of cell–cell adhesion, thereby inhibiting the dissociation of tight and adherens junctions. Furthermore, TSH downregulates the expression of TGFb1, which induces the loss of epithelial polarization. TSH may also control follicular lumen generation, because chloride channels, located at the apical pole, are regulated by cAMP. One limitation of the data on follicle formation from *in vitro* studies is that they can not be directly extrapolated to comprehend folliculogenesis *in vivo*. Cultured rat follicles invert their polarity and manifest a dramatic change in functional properties unless they are cultured with a gel consisting of extracellular matrix proteins. In addition, TSH is not required for the formation of follicles *in vivo*, because folliculogenesis is not affected in mutant mice with impaired TSH/Tshr signaling (see later, Late Stages of Thyroid Morphogenesis: Functional Differentiation and Histogenesis). Furthermore, follicles in the thyroid gland are surrounded by a basal lamina and present adjacent capillaries, strongly remodeled when iodine uptake is altered, both of which are absent in cultured follicles. As the microvasculature plays a key role in thyroid homeostasis, the angiofollicular unit (AFU), formed by TFCs and capillaries, has been proposed as the functional and morphological unit of the thyroid. Morphological changes in the endothelial compartment in response to iodine deficiency have been described that anticipate by a few days the changes visible in the epithelial cells and are related to vascular endothelial growth factor (VEGF) release from thyroid epithelial cells. Thus, thyroid hormones synthesis will be controlled by both follicular cell homeostatic processes and the vascular plasticity that, in turn, is controlled by TSH plasma levels in a delayed phase.[2] Even if AFUs are able to respond synchronously to goitrogen stimuli, they do not work uniformly, but as heterogeneous clusters of units each presenting its own functional properties, allowing the gland to adapt to fast-changing physiological conditions.

KEY POINTS

- The thyroid follicular cell (TFC) is a highly specialized cell type only found in the thyroid gland. TFCs are characterized by a unique combination of proteins devoted to the synthesis of thyroid hormones and by regulatory factors that control their ontogeny and function.

OTHER THYROID CELLS

In addition to the follicular cells, the thyroid gland is composed of other epithelial cell populations of different origin: i) the parafollicular C cells, devoted to calcitonin production; ii) the epithelial cells, vestiges of the ultimobranchial body; iii) the thyroid stem cells.

C Cells

The other endocrine cells present in the thyroid gland of mammals are C cells, which synthesize and secrete calcitonin, a polypeptide hormone involved in calcium metabolism. In lower vertebrates C cells form an organ called the ultimobranchial gland that is separate from the thyroid gland.

C cells are known as parafollicular cells because of their distribution among follicular cells. But, in spite of their name, C cells are not solely located between follicular cells and the basement membrane (a real parafollicular position), but are also found among the follicles (interfollicular) or in an intrafollicular position. C cells are found dispersed as individual cells closely associated with follicular cells, or in more complex structures consisting of both follicular and C cells. The number of parafollicular cells in the thyroid differs among species. In humans, the number of C cells decreases with age; indeed, the neonatal thyroid has ten times more C cells than the adult thyroid. In the adult gland the C cells represent less than 1% of the follicular cells and are usually distributed in the upper two thirds of the lateral lobe, in an intrafollicular and parafollicular position.

C cells are characterized by clear cytoplasm and small, compact nuclei. Electron microscopy reveals that these contain cytoplasmatic secretion granules that are 100 to 200 nm in diameter. At the molecular level, C cells are identified chiefly by the presence of calcitonin. The calcitonin protein is encoded by *calcitonin/calcitonin-related polypeptide, alpha* (*Calca*) gene, which is capable of expressing, by tissue-specific alternative splicing, three proteins: calcitonin gene–related peptide (CGRP), katacalcin I, and katacalcin II. The splicing of the first three exons to the fourth exon gives rise to an mRNA that encodes for a protein precursor, which is subsequently processed to yield calcitonin and a peptide known as katacalcin I. CGRP and katacalcin II are the products of alternative splicing and are by far less abundant than calcitonin in C cells. CGRP is a 37–amino acid vasoactive peptide with no known effect on calcium metabolism. The embryological origin of the C cells has been investigated in detail (see later, Differentiation of C Cells).

Ultimobranchial Body-Derived Epithelial Cells

In addition to the TFCs and C cells, another cell population has been described in the mammalian thyroid. These cells do not display a clear follicular organization, but form epithelial structures that were called a "second kind of thyroid follicles" or "ultimobranchial follicle"[3] in rodents. The finding that the second kind of thyroid follicles is absent in the avian thyroid (where ultimobranchial bodies never merge with the thyroid) has suggested that they represent remnants of the ultimobranchial body endodermal component. These structures show a notable interspecies differences.[4] In humans, ultimobranchial body remnants are known as solid cell nests (SCNs). SCNs are frequently present in the thyroid gland, preferentially located in the middle and upper third of the lobes. SCNs appear as para- or intrafollicular clusters and cords of epithelial cells clearly separated from the follicles by a basal lamina. SNCs are composed of two cell types: both C cells and "main" cells, the most important cell population of these structures. In most cases, the SNCs are found admixed with another structure, known as "mixed follicle," in which follicular cells and main cells underline a lumen filled with colloid-like material. Main cells are polygonal cells showing squamoid features, with oval nuclei and eosinophilic cytoplasm, lacking intercellular bridges. The molecular phenotype of main cells is peculiar. Indeed, these cells do not express markers of differentiation common to either follicular cells or C cells. It is worth noting that recent studies in mouse embryos have revealed that p63-positive, Nkx2.1-negative cells are present both in the epithelium of the fourth pharyngeal pouch and in the ultimobranchial body, confirming that the SNCs are ultimobranchial body remnants.

Thyroid Stem Cells

Human TFCs have a slow turnover and divide approximately five times over the course of a lifetime. However, in animal models, it has been demonstrated that the thyroid preserves the ability to both regrow and restore follicular organization after severe perturbations, such as severe experimental thyroiditis or partial thyroidectomy.[5] Recently, in a genetically modified mouse model of thyroid injury,[6] significant transient expression of stemness markers such as Nanog and Oct4 was reported after the massive cell death of follicular cells. Then, clusters of thyroid precursor cells appeared, and finally thyroid function was restored.

These phenomena could be explained by the presence of a population of "adult" stem cells (often referred to as "progenitor" cells) capable both of self-renewal and of differentiating into a more mature cell type.

Adult stem cells have been identified in several organs, where they have a role in tissue repair after disease or injury. Although thyroid adult stem cells have been postulated since early 1990s, their presence in the thyroid gland has been receiving firm support only in the last few years.

Side population (SP) cells are a small subset of cells, also identified in the mouse thyroid gland, that are mostly located in the interfollicular space, where they represent less than 1% of the total cells of the gland. SP cells share many features with *bona fide* stem cells, including the expression of stem cell markers Oct4 and nucleostemin, while SCA1 antigen is expressed only in a half of SP cells. Because of their stemness, they do not express detectable levels of thyroid-enriched transcription factors (Nkx2.1, Pax8, and Foxe1) or functional markers (Tg and TPO). In humans, SP cells have been isolated from cultures of thyrocytes obtained from nodular goiters and display a molecular signature that is similar to that of the mouse SP cells. Usually, SP cells undergo tissue-specific differentiation when maintained in culture. Notably, SP cells isolated from normal mouse thyroid stay in an undifferentiated state in culture and are not able to arrange follicle-like structures. In contrast, human SP cells are able to proliferate *in vivo* and acquire the thyroid phenotype in response to TSH. It is unclear whether disparities between mouse and human SP cells depend on different culture conditions or on the different origin of the cells, from normal or goitrous thyroid, respectively.

In addition to SP cells, the main cells of SCNs (see earlier) also display many characteristics of stemness. Indeed, in these cells, genes specific for either thyroglobulin- or calcitonin-producing cells are not expressed, while markers usually detected in basal/stem cells of several multilayered epithelia are present. Furthermore, these cells show an increased expression of Bcl-2, an antiapoptotic protein that is constitutively expressed in TFCs. This molecular signature corresponds to the definition of stem cell and suggests that the main cells of SCNs could be a source of multipotent cells able to differentiate towards hormone-producing cells.

The data summarized earlier seem to suggest that the regeneration of the thyroid could be due to proliferation and differentiation of thyroid stem/progenitor cells suitable of replacing the pool of the disrupted follicular cells. However, even if the evidence for the existence of thyroid stem cells is convincing, we still do not know what the role of these cells is in the physiology and pathology of the gland.

KEY POINTS

- The thyroid contains a minority of other cell types, of which the most relevant are C cells, devoted to calcitonin synthesis.

THE DEVELOPMENT OF THE THYROID GLAND

The adult thyroid gland in mammals is assembled from two different embryological structures. This composite origin reflects the dual endocrine function of the gland. The thyroglobulin-producing follicular cells derive from a small group of endodermal cells of the primitive pharynx (the thyroid anlage), whereas the calcitonin-producing parafollicular cells derive from the endoderm lining the ultimobranchial bodies, transient embryonic structures that are derived from the fourth pharyngeal pouches.

The thyroid anlage and the ultimobranchial bodies migrate from their original sites, reach their final position anterior to the trachea, and fuse to form the definitive thyroid gland. The thyroid follicles derive from the thyroid anlage cells, while the C cells are scattered within the interfollicular space.

Data on thyroid organogenesis in humans are scarce. In contrast, the morphogenesis and differentiation of the thyroid have been extensively studied in animal models, mostly in mice and rats. Studies on patients affected by congenital hypothyroidism with thyroid dysgenesis (CHTD) have confirmed that identical genetic mechanisms are involved in thyroid organogenesis in humans and in mice. Furthermore, as the molecular pathways involved in the formation of the thyroid and in the differentiation of follicular cells are essentially conserved among all vertebrates, the use of zebrafish as a tool for the analysis of thyroid development has also been introduced in the last years.

Here we will describe in detail the morphological and molecular aspects of thyroid development in mice (summarized in Fig. 64.4). These data can reasonably be extended to humans; the known relevant differences will be highlighted.

Specification of the Thyroid Follicular Cells: The Thyroid Anlage and the Thyroid Placode

Gastrulation involves dramatic changes in the overall structure of the embryo, converting it into a complex three-dimensional structure. The anterior and posterior endoderm invaginates, forming two pockets that extend and fuse, generating a primitive gut tube that runs along the anterior-posterior axis of the embryo. The foregut is the most anterior (cranial) region of this tube. The endoderm of the foregut gives rise to a number of cell lineages that eventually form the epithelial components of the thyroid, thymus, lungs, stomach, liver, and pancreas. All organs derived from the gut tube undergo a rather similar process. First, a restricted group of cells differentiate from their neighbors, as can be visualized both by the appearance of specific molecular markers (usually transcription factors) and by the almost simultaneous appearance of a multilayered structure. Subsequently, all organs undergo a morphogenetic process that entails, depending on the organ, proliferation, branching morphogenesis, migration, or a combination of these processes. Finally, each organ forms a specific cellular organization that is directly correlated to the organ function (i.e., follicular organization in the case of the thyroid). The first differentiation step in this process will be called *specification* throughout this chapter, even though the term usually refers to a reversible process; in the case of all organs derived from the gut, it is at present unknown whether the first differentiation event is reversible or irreversible. While the inductive mechanisms by which endodermal cells produce liver, pancreas, and lung cell lineage are beginning to be understood, the events that commit a group of multipotent endodermal cells to a thyroid fate are not known in similar level of detail.

In mice, which undergo a 19.5 day–long gestation period, the first evidence of thyroid specification is the detection of a monolayer of cells (*thyroid anlage*) that differentiate from the surrounding epithelium and simultaneously express the transcription factors Nkx2.1, Pax8, and Hhex at embryonic day (E) 8 to 8.5. It is worth noting that the thyroid anlage can be detected only by the expression of molecular markers. The morphological consequence of thyroid specification is the appearance of a midline endodermal thickening in the ventral wall of the primitive pharynx, the *thyroid placode*. In the mouse, this is visible at E9 to E9.5, and in humans at E22. The thyroid placode is caudal to the region of the first branchial arch that forms the tuberculum impar, the region of the first branchial arch that gives rise to the median portion of the tongue (Fig. 64.4). At this stage the cells of the thyroid placode simultaneously express four transcription factors: Hhex, Nkx2.1, Pax8, and Foxe1.[7] This unique molecular signature hallmarks these cells, which can be called precursors of TFCs (pTFCs), including all developing thyroid cells that do not yet synthesize thyroid hormones. It is widely accepted that TFCs are derived from the endodermal cells of the thyroid anlage.

Any genes relevant in the patterning process of the foregut, such as *Nodal*, members of the *Gata* family, or *Sox* genes could play a role in thyroid specification. As yet, there have been no reports of either cell fate mapping studies or genetically modified mice clarifying the dissection of the early steps of the thyroid specification. Actually, mouse embryos carrying targeted inactivation of genes involved in foregut patterning usually show developmental arrest at stages that preclude the assessment of the thyroid anlage. In zebrafish, Nodal signaling, required for endoderm specification, seems to be specifically involved in the commitment of thyroid fate. Indeed, as a consequence of the impairment of this signaling, endodermal cells form a reduced gut tissue but do not contribute to the thyroid.[8]

It has been reported that signals from cardiac tissue adjacent to the endodermal layer have an inductive role in liver, pancreas, and lung specification. Similarly, endodermal cells could acquire thyroid fate as a consequence of inductive signals from the mesenchyme or from the adjacent mesoderm. The study of the developing thyroid in mice has shown that, at E8.5, the earliest identifiable thyroid anlage appears in close apposition to the aortic sac, the cardiac region that gives rise to the embryonic heart outflow and pharyngeal arteries. This observation suggests that short-range inductive signals from cardiac mesoderm or from the endothelial lining of the aortic sac could have an inductive role to specify undifferentiated endodermal cells towards a thyroid fate. Indeed, in mice, alterations in the foregut have been observed as a consequence of impaired heart development. Consistent with the hypothesis of a role of cardiac tissue in thyroid development are findings showing that cardiac malformations represent the most frequent birth defects associated with thyroid dysgenesis (TD) in humans. Data obtained from zebrafish and other species are proving invaluable for the study of the extrinsic factors (acting outside the pTFCs) that control thyroid specification. In zebrafish, pharmacological upregulation of Wnt/beta catenin signaling during gastrula stage results both in a block of early cardiac mesoderm differentiation and an extreme reduction of Nkx2.4b-positive thyroid precursor cells. The cardiac and thyroid defects are rescued by enhanced BMP signaling. It is worth noting that, both in zebrafish and in mice, BMP4, a BMP ligand, is expressed in cardiac mesoderm adjacent to developing thyroid. The role of cardiac mesoderm in thyroid specification is reinforced by the findings that genetically mutated fishes in which heart formation is blocked display the same thyroid phenotype.[9] Other reports indicate that thyroid specification requires the combined action of BMP and FGF signaling. Both in zebrafish[10] and in chick embryos, pharmacological manipulations of FGF signaling disrupt endoderm patterning along the anterior-posterior axis. Also, in mice it has been demonstrated that FGF signals are required for correct thyroid development; however, these factors appear to play a role after the specification step (see later, Assembling the Gland: Development and Merging of the Ultimobranchial Bodies). Finally, the relevance of both BMP and FGF signaling has received strong support from the finding that these two pathways are required for thyroid lineage specification of embryonic stem cells (see later, Generation of Functional Thyroid Follicles *In Vitro*).

Early Stages of Thyroid Morphogenesis: Budding, Migration, and Lobulation of Thyroid Primordium

As described above, in mice, by E9.5, the thyroid placode first appears as a multilayered epithelium. Shortly after, it forms a bud that, evaginating from the floor of the pharynx, invades the surrounding mesenchyme as an endodermal extroflession close to the aortic sac. Other endodermal organs follow the same developmental pattern as the

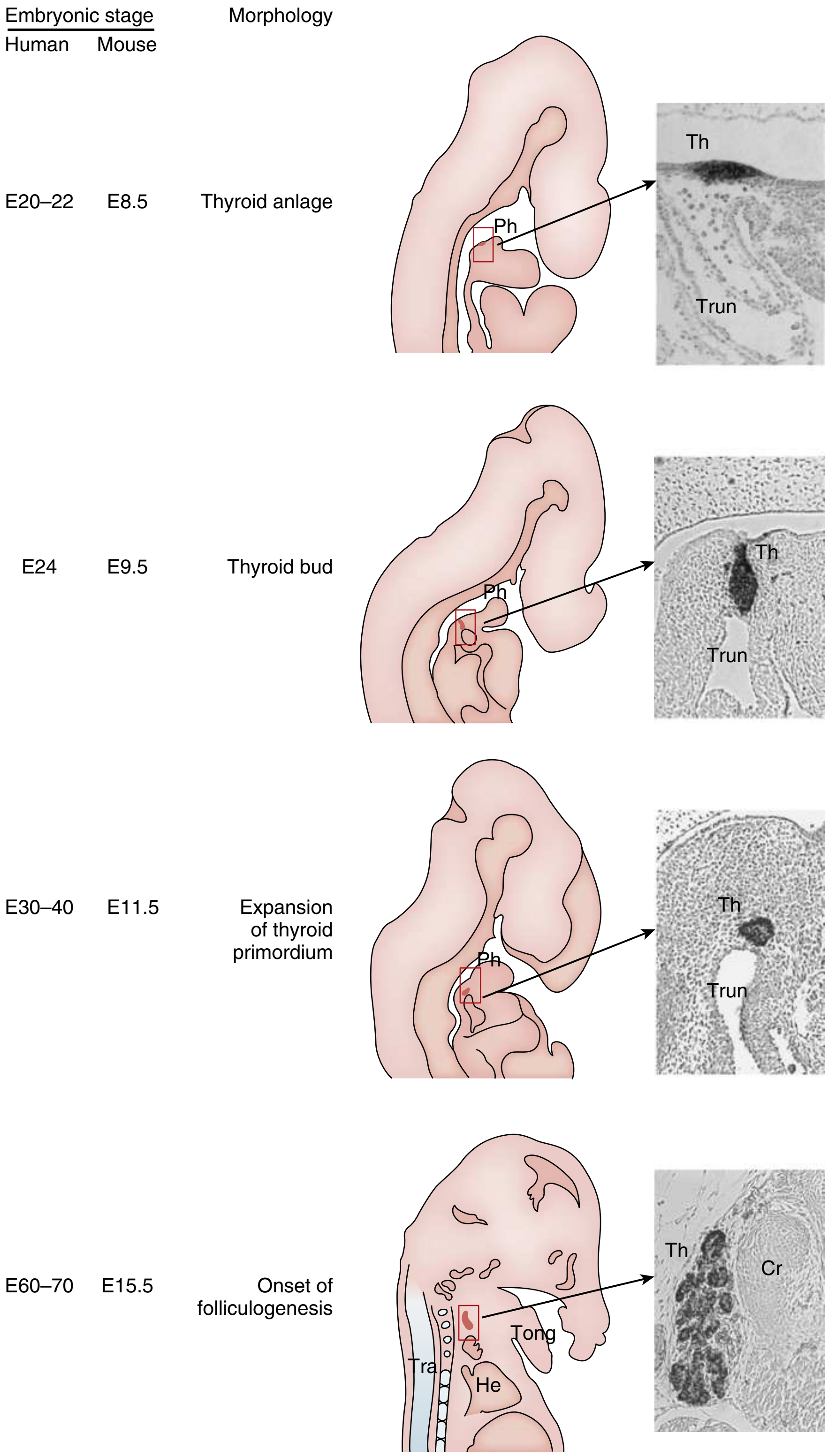

Embryonic stage
Human
Mouse
Morphology
E20–22
E8.5
Thyroid anlage
Ph
Th
Trun
E24
E9.5
Thyroid bud
Ph
Th
Trun
E30–40
E11.5
Expansion of thyroid primordium
Ph
Th
Trun
E60–70
E15.5
Onset of folliculogenesis
Tra
Tong
He
Th
Cr

Fig. 64.4 Thyroid development. In the *middle* is a schematic representation of the different steps of thyroid organogenesis in mouse embryos; on the *left,* the correspondent embryonic stages *(E,* embryonic day) in mice and humans are indicated; on the *right,* relevant sagittal sections of mouse embryos stained with an anti-Nkx2.1 are shown. *Cr,* Cricoid cartilage; *He,* heart; *Ph,* pharynx; *Th,* thyroid; *Tong,* tongue; *Tra,* trachea; *Trun,* truncus arteriosus.

thyroid: an endodermal domain of specified cells, then a pseudostratified epithelium that eventually invaginates to form a bud.

In these early stages of morphogenesis, the expansion of the thyroid placode does not seem to be due to an enhanced proliferation of pTFCs; indeed, these cells show a very low cell proliferation rate as compared with that of the cells of the surrounding endoderm and mesoderm.[2] Other cells from the pharyngeal endoderm could be recruited into the developing thyroid, contributing to the increase of the number of thyroid progenitor cells. At least for some organs, such as the pancreas, it has been demonstrated that the number of the precursor cells determines the final size of the organ. Thus, perturbations of the genetic mechanisms, still unknown, underlying the recruitment of endodermal cells in the thyroid placode could be a cause of a hypoplastic thyroid in the adult.

At E10.5, the thyroid bud has a flask-shaped structure and is still connected to the floor of the pharynx by a thin cord, the thyroglossal duct, a transient narrow channel. One day later, the thyroid bud is visible as a cap-like shape that has migrated caudally into the mesenchyme, losing all connections with the floor of the pharynx (Fig. 64.4). At this stage the developing thyroid begins to expand laterally, the first step of the process that eventually lead to the formation of the two lobes. At E12.5 the thyroid primordium appears as an elongated structure, extended laterally, in contact to the third pharyngeal arch arteries that will participate the formation of the definitive carotid vessels. By E13, the thyroid bud continues its downward relocation into the mesenchyme and approaches the ultimobranchial bodies, which are accomplishing their ventrocaudal migration from the fourth pharyngeal pouch. By E13.5, the developing thyroid, already formed by rudimental lobes connected by a thin central portion, reaches its definitive pretracheal position, where it merges with the ultimobranchial bodies containing the precursors of C cells.

In humans, as described in mice, the thyroid anlage appears as a bud invading the mesenchyme at E26; in a few days it appears as a migrating primordium, connected to the pharynx by the thyroglossal duct that disappears around E37. At this stage the thyroid bud acquires a bilobed shape. The developing thyroid merges with the ultimobranchial bodies by the sixth week and reaches its final position in front of the trachea around the seventh week.[11]

The translocation of TFC precursors to reach the sublaryngeal position is a process that lasts for almost 4 days in the mouse and almost 4 weeks in the human embryo. The mechanisms involved in the movement of the thyroid primordium are still a matter of debate. Because the definitive location of the thyroid gland is rather distant from the site of primitive specification, translocation mostly involves active migration of the precursors. However, other morphogenetic events occurring in the neck region and the remodeling of the surrounding mesenchyme could contribute to the definitive location of the thyroid. The expression of transcription factors such as either Hhex or Pax8, or Nkx2.1, is not sufficient for thyroid migration, while Foxe1 plays a crucial role, because the presence of this factor in the thyroid bud is required to allow the cells to move.[7,12]

In many processes of embryogenesis, such as gastrulation, neural crest migration, and heart formation, cell migration is involved. However, in these processes, migrating cells lose the epithelial phenotype and acquire mesenchymal features. This phenomenon, called epithelial-mesenchymal transition, is hallmarked by an increased expression of N-cadherin and the downregulation of E-cadherin. In contrast, TFC precursors maintain their epithelial phenotype (E-cadherin stays expressed) throughout the entire translocation process and never acquire a mesenchymal identity.[13] It has been hypothesized[14] that TFC precursors could translocate following a model of collective cell migration featured by the presence of migrating cells that remain connected during the movement, even though no conclusive data have been reported.

Once the gland reaches its final location, the two rudimentary paratracheal lobes expand, and, by E15 to E16, the thyroid gland assumes its definitive shape (Fig. 64.4). In the last stages of thyroid organogenesis, the gland increases further in size, probably due to the high proliferation of TFCs.

The genetic basis of the formation of two lateral symmetric lobes (lobulation) is beginning to be understood. When the lobulation process begins, at E11.5, the developing thyroid is in contact with the third pharyngeal arch arteries that will participate the formation of the definitive carotid vessels located most closely to the mature thyroid lobes. Signals originating from adjacent vessels or factors regulating vasculogenesis could instruct the process of lobulation by non–cell-autonomous mechanisms. The study of animal models seems to confirm this hypothesis. Mice deficient in either *Shh* (a key regulator of embryogenesis)[15] or *Tbx1* (a factor regulated by Shh itself)[16] display a disturbed morphogenetic patterning of vessels adjacent to the developing thyroid. In these mutated embryos, as a consequence of the absence of caudal pharyngeal arch arteries, the thyroid bud is never in close contact with vessels, and the lobulation process is impaired: the thyroid fails to separate into two distinct lobes, maintaining the shape of a single tissue mass throughout development. Along similar lines, *ntn1a*-deficient zebrafish embryos present severe cardiac, vascular, and thyroid defects. In these embryos the morphology of developing thyroid is perturbed and shows lateral or anteroposterior expansion. *Ntn1a* (a gene involved in Sonic hedgehog signaling) is not expressed in thyroid precursor cells, but rather in the mesenchyme surrounding the aortic arch arteries. Interestingly, loss-of-function mutation in *netrin-1* (*NTN1*), the human homolog of *ntn1a,* was reported in a patient affected by congenital hypothyroidism (CH) and cardiac defects.[17] Finally, TD is not uncommon in patients affected by DiGeorge syndrome, which is characterized by congenital anomalies of the heart and great vessels.

However, the inductive signals from adjacent tissues must interact with events restricted to the thyroid cells to accomplish the lobulation process. Indeed, thyroid hemiagenesis is very frequent in mice double heterozygous for the null allele of *Nkx2.1* and *Pax8*, genes that are expressed in thyroid precursor cells and are absent in other structures close to the developing thyroid.[18]

Late Stages of Thyroid Morphogenesis: Functional Differentiation and Histogenesis

When the thyroid primordium reaches the sublaryngeal position, TFC precursors accomplish their functional differentiation. Notably, the normal final location of TFCs in front of the trachea is not an essential requirement for functional differentiation, because an ectopic, sublingual thyroid expresses thyroglobulin in both human patients and mutated mice.[12] The functional differentiation of TCFs is hallmarked by the expression of a series of proteins, essential for thyroid hormone biosynthesis, such as thyroglobulin (Tg), thyroid peroxidase (TPO), Tshr, NIS, thyroid NADPH oxidases (Duox proteins), and pendrin. This program requires almost three days, between E14 and E16.5, and results in the differentiation of the thyroid primordium in a functional thyroid gland that is able to produce and release hormones. *Tg* and *Tshr* are expressed by E14[19]; expression of *Tpo* and *Nis,* which encode the two key enzymes involved in the process of Tg iodination, appear one day later, probably because their expression is absolutely dependent on the pathway activated by the binding of TSH to its receptor, Tshr.[20] *Duox* expression appears at E15.5,[21] and finally thyroxine expression by E16.5.[22]

The molecular mechanisms involved in the differentiation of the human thyroid are not much different from those found in mice.

Functional differentiation of TFCs requires almost 3 weeks. It begins after the developing thyroid is located in front of the trachea, at E48, when TFCs express *TG* and *TSHR*, and T4 synthesis is detected at the tenth week.[11]

The binding of TSH to Tshr triggers a signaling pathway that regulates many functions of the thyroid gland in postnatal life. In the mouse embryo, *Tshr* is detected in TFC precursors by E14 to E14.5,[19] when the developing thyroid has reached its final position and *Tg* is expressed. At later stages of development *Tshr* expression increases, and it remains expressed in adult life. The availability of mice that carry mutations in the *Tshr* gene enabled elucidation of the role of the TSH/Tshr signaling during embryonic life.[20] Both $Tshr^{hyt/hyt}$ (carrying a loss-of-function mutation in the *Tshr* gene) and $Tshr^{-/-}$ adult mice[23] display severe hypothyroidism with a hypoplastic thyroid. During embryonic life, in the absence of a functional Tshr, the developing thyroid does not show any alterations in either size or histological structure. However, the expression of both *Tpo* mRNA and *Nis* is strongly downregulated.[20] Thus, TSH/Tshr signaling is required to complete the TFC differentiation program. However, the TSH-induced cAMP pathway is not involved in controlling the growth of the embryonic gland, while this pathway is the main regulator of the growth of the adult thyroid. It is worth noting that the requirements for growth of the embryonic thyroid seem to be different between mouse and human; in the latter, TSH/Tshr signaling during fetal life is necessary for development of the thyroid.[24]

During the process of functional differentiation, the thyroid gland also accomplishes its characteristic histological organization. By E15.5, TFCs start to form small rudimentary follicles, as revealed by the expression of ZO-1, a tight junction marker. At E16.5 the gland displays an evident follicular organization. The histogenesis is complete in late fetal life, E17 to E18: the thyroid parenchyma is organized into small follicles, surrounded by a capillary network, that contain thyroglobulin in their lumens.[2] At birth, the thyroid gland is able to produce and release thyroid hormones, though the regulation of its growth and function by the hypothalamic-pituitary axis is fully active only after birth.[24]

In humans, the establishment of the characteristic histological organization lasts several weeks and can be divided into three phases: the precolloid phase, the beginning colloid phase, and the follicular growth phase, which occur from 7 to 10 weeks, from 10 to 11 weeks, and after 11 weeks of gestation, respectively. In the precolloid phase, small intracellular canaliculi develop as an accumulation of colloid material. These small canaliculi enlarge, and the colloid organizes itself into extracellular spaces. In the last phase primary follicles are clearly visible, and the fetal thyroid is able to concentrate iodide and synthesize thyroid hormones. The human thyroid continues to expand until term, and, contrary to mice, the hypothalamic-pituitary-thyroid axis starts functioning at midgestation.

Genetically modified mouse models have given insights on the genetic mechanisms controlling histogenesis. A fully active Nkx2.1 is required for the formation of follicles. The absence of Nkx2.1 in differentiated TFCs impairs the normal architecture of the thyroid. Along the same lines, a hypomorph Nkx2.1 is not able to support proper organization of TFCs.[25] In addition, normal production of miRNAs is required to maintain a correct thyroid structure.[26]

Coordinated interactions between epithelial thyroid cells and endothelium are involved in establishing thyroid follicles.[2] Thus, both cell-autonomous and non–cell-autonomous mechanisms cooperate to achieve the peculiar morphofunctional structures of the thyroid gland.[27]

Assembling the Gland: Development and Merging of the Ultimobranchial Bodies

In vertebrates, calcitonin-producing cells differentiate from the ultimobranchial body, a structure that derives from the fourth pharyngeal pouch. The ultimobranchial body is a definitive organ in all vertebrates except in placental mammals, where it is an embryonic, transient structure fated to join the medial thyroid bud.

In mice, the fourth pharyngeal pouches are evident for the first time by E10. They appear as a lateral extroflexion of the primitive foregut. Shortly after, the caudal portion of the pouches grows, and at E11.5 the fourth pharynx-branchial duct is pinched off, forming an ultimobranchial body primordium visible as an ovoid vesicle with a lumen lined by columnar epithelium.

By E11.5 ultimobranchial body primordia migrate, and at E13 they appear as solid clusters of cells in contact with the midline primordium of the thyroid. By E14.5 ultimobranchial body cells begin to disperse within the thyroid parenchyma, and 1 day later only remnants of ultimobranchial bodies can be distinguished in the thyroid gland. C cells complete their differentiation program through the expression of a series of proteins according to a precise temporal pattern: the basic helix-loop-helix transcription factor Mash1 is expressed by E12.5; the neuronal markers TuJ1, CGRP, and somatostatin are expressed by E14.5; and 1 day later, Mash1 expression disappears, and calcitonin-producing cells can be detected between follicular cells.

In humans, the development of ultimobranchial bodies is similar to that seen in mice. At E24 the ultimobranchial body primordia appear as an outpouching of the ventral component of the fourth pharyngeal pouch. At this stage, the primordium of parathyroid IV is visible as a dorsal evagination of the same pouch. Some authors describe a transient fifth pouch as the endodermal origin of the ultimobranchial body. Probably the shape of the ultimobranchial body anlage itself, which appears as an incomplete pouch, has generated these different interpretations. By E35 the ventral extroflexion is a long-necked flask still attached to the pharynx; a few days later the ultimobranchial body primordium loses its connection with the pharyngeal cavity, starts its migration, and, at E40, reaches the posterior surface of the median thyroid. A connective layer separates these two buds, which display a different histological organization: the lateral bud is composed of a compact mass of cells, while the median bud is composed of interconnecting sheets of epithelium. Finally, at E55, the ultimobranchial bodies are incorporated with the lateral lobes of the thyroid, and the cells from both structures mix with each other.

A number of intrinsic and extrinsic factors play a role in controlling the development and the merging of the ultimobranchial body. Among the cell-autonomous mechanisms, it is worth noting the essential role of Nkx2.1 for the survival of ultimobranchial body cells, but not for their specification. Nkx2.1 functions are in part dosage-sensitive. Indeed $Nkx2.1^{+/-}$ mice display an abnormal fusion of the ultimobranchial bodies with the thyroid diverticulum. Ultimobranchial body cells are incompletely incorporated into the thyroid parenchyma and remain at the dorsal part of the thyroid lobe. Another factor required for the proper merging of ultimobranchial bodies is Hes-1, a transcriptional repressor expressed in ultimobranchial body cells by E11.5. In the absence of Hes-1, ultimobranchial bodies form but show hypoplasia, as well a 3-day delay in fusion with the medial developing thyroid.[28]

Ultimobranchial defects have been reported in mice carrying mutations in *Hox3* paralog genes. These factors regulate the regionalization of the embryo along its major axes and are also involved in the morphogenesis of several structures. *Hoxa3* is expressed in the pharynx, in the developing thyroid, and in the mesenchymal and endodermal cells of the fourth pharyngeal pouch. In *Hoxa3*-null mice, C cells differentiate correctly, but their number is significantly reduced. In many cases, ultimobranchial bodies fail to fuse with the thyroid bud and remain as bilateral vesicles composed exclusively of calcitonin-producing cells (persistent ultimobranchial bodies). Variable thyroid anomalies are described, including hypoplasia, hemiagenesis, reduction of the number of follicular cells, and absence of isthmus. The phenotype appears more severe

TABLE 64.1 Expression of Relevant Genes and Capacity to Produce Thyroid Hormones at Different Stages of Thyroid Development in Mice

	CONTROLLER GENES		FUNCTIONAL DIFFERENTIATION			
Embryonic Day	*Nkx2.1* *Foxe1* *Pax8* *Hhex*	*Ffgr2*	*Tg,* *Tshr*	*Tpo* *Nis*	*Duox*	Thyroid Hormones
E8.5	+	-	-	-	-	-
E11.5	+	+	-	-	-	-
E14 -14.5	+	+	+	-	-	-
E15-15.5	+	+	+	+	-	-
E15.5-16	+	+	+	+	+	-
E16.5	+	+	+	+	+	+

in mice carrying various mutant combinations in *Hoxa3* and its paralogs *Hoxb3* and *Hoxd3*.[29] It is thought that the impaired organogenesis of the thyroid is secondary to defects in the ultimobranchial bodies.

Finally, in absence of Tbx1-dependent signals from the mesoderm, the developing thyroid appears as a hypoplastic single tissue mass.[16] Tbx1 signals are executed by FGF8, whose expression in the mesenchyme surrounding the thyroid is regulated by Tbx1.[30]

Differentiation of C Cells

The ontogenesis and differentiation of C cells was initially studied in birds, using as model chick–quail chimeras.[31] The analysis of this model was taken by the entire community as demonstrating that avian C cell precursors derived from the neural crest colonize ultimobranchial bodies. However no formal proof of the neural crest origin of these cells was reported in mice and humans. Recently, very elegant studies both in the developing mouse embryo and in human cancers have clearly shown that, in these organisms, C cells derive from the endoderm lining the fourth pharyngeal pouches,[32] in agreement with several other lines of evidence pointing in the same direction, such as a common essential role of Nkx2.1 in the development of the thyroid proper and of the ultimobranchial body. It would be now essential to revisit the chick–quail chimera studies in order to evaluate whether some details of those studies could be reinterpreted in order to definitively assess whether the same anatomical structure could have different embryonic origins in diverse organisms.

KEY POINTS

- The thyroid gland derives from progenitor cells located mostly in the middle of the pharyngeal cavity, in the area that will be covered by the tongue. Subsequently these cells migrate to reach the final location of the thyroid in front to the trachea. Hence, the sublingual thyroid is a block in normal thyroid development.
- The calcitonin-producing C cells reach the developing thyroid via two lateral pouches emanating from the primitive pharynx. At variance from what believed up to now, it has been demonstrated, at least in mice and men, that they are of endodermal origin, as are the follicular cells.

MOLECULAR GENETICS OF THYROID GLAND DEVELOPMENT

The discovery of transcription factors, which regulate the expression of follicular cell specific genes in the mature thyroid and are also expressed in the thyroid primordium, offered a useful tool to explore the genetic basis of the developmental process of the thyroid gland. At E8.5 epithelial cells fated to become TFCs are unequivocally individualized in the endodermal layer of the primitive pharynx by the expression of *Nkx2.1,*[19] *Foxe1,*[33] *Pax8,*[34] and *Hhex.*[35] Although these factors are also expressed in other embryonic tissues, the coexpression of all four is only seen in both precursor and differentiated TFCs. Experiments in animal models have conclusively demonstrated that the expression of *Nkx2.1*, *Foxe1*, *Pax8,* and *Hhex* is required for the development of the thyroid.[36,37] It is worth noting that, in addition to *Nkx2.1*, *Foxe1*, *Pax8,* and *Hhex,* other genes, both thyroid-enriched and ubiquitous, are required to guarantee a correct organogenesis and functions of the gland (Table 64.1). A list of genes that are known (or are candidates) to be responsible for a thyroid defect, mainly as deduced by the phenotype of knockout animals, is presented in Table 64.2. The role that Tshr plays in thyroid differentiation was summarized earlier in the section on Late Stages of Thyroid Morphogenesis: Functional Differentiation and Histogenesis. The functional genomics of this molecule and the consequences of its mutations in the pathogenesis of thyroid diseases will be exhaustively treated in later chapters. In the sections that follow, we will focus on *Nkx2.1*, *Foxe1*, *Pax8,* and *Hhex,* which can be considered hallmarks of differentiated TFCs.

Functional Genomics of Thyroid-Enriched Transcription Factors

Nkx2.1 Nkx2.1 (formerly known as TTF-1 for thyroid transcription factor–1, or T/EBP) is a transcription factor that recognizes and binds to specific DNA sequences via a 61–amino acid–long DNA binding domain called the homeodomain. The homeodomain sequence is conserved from fruitflies to humans. Nkx2.1 was initially identified in a rat thyroid cell line as a nuclear protein able to bind to specific sequences in the *Tg* promoter.

Nkx2.1 is a member of the Nkx2 class of transcription factors and is encoded by a single gene, whose official name is *Nkx2.1* in mice and *NKX2.1* in humans, which are located on chromosome 12 and chromosome 14q13, respectively.[36] The gene is formed by at least three exons that express multiple transcripts. The most abundant is the shorter isoform, a 2.3-Kb mRNA that encodes a 42-kDa protein that is 371 amino acids long in humans; the longer protein has also been detected in the lung. However, the biological relevance of the various proteins is unclear. Functional studies have determined that the homeodomain is responsible only for binding to DNA, while the transactivating activity resides in two domains, N and C, which are localized at the N-terminus and the C-terminus of the protein, respectively. These

TABLE 64.2 Currently Available Mouse Models of Thyroid Dysgenesis and Summary of Known and Potential Genes Involved in the Pathogenesis of the Disease

Stage of Morphogenesis Impaired	Thyroid Phenotype Expected	Mouse Model	Other Candidate Genes	Genetic Defect in Human Diseases
Formation of thyroid anlage	Agenesis	Not available	Unknown genes responsible for specification	Unknown
Translocation of thyroid bud	Ectopic thyroid	*Foxe1* knockout	Genes controlled by Fox1	Unknown
Survival and expansion of immature thyroid cells	Athyreosis	*Foxe1* knockout *Nkx2.1* knockout *Pax8* knockout *Fgf10* knockout *Fgfr2* knockout *Hhex* knockout	Genes controlled by Nkx2.1, Foxe1, Pax8, and Hhex	*FOXE1* mutations
Expansion of differentiated thyroid cells	Hypoplasia	*Tshr* knockout $Tshr^{hyt/hyt}$ mouse $Tshr^{dw/dw}$ mouse	TSH-induced genes	*PAX8* mutations[a] *NKX2.1* mutations[a] *TSHR* mutations TSH-induced genes
Interactions with ultimobranchial body	Hypoplasia	*ET-1* knockout *Hoxa3* knockout *Eya1* knockout *Pax3* knockout (*splotch*) *Tbx* knockout	Other *Hox* genes	Unknown

[a]In humans, unlike mice, the abnormal phenotype is detected in heterozygotes.

two domains appear redundant in *in vitro* assays, but could have different functions *in vivo*. In fact, a number of Nkx2.1 interactors have been identified that bind specifically to either the N domain or the C domain. Nkx2.1 is phosphorylated at several serine residues. This posttranslational modification could regulate Nkx2.1 activity. Indeed, a mouse model harboring a mutated *Nkx2.1* allele encoding a protein that could not be phosphorylated showed impaired differentiation of both thyroid and lung.[25]

The expression pattern of Nkx2.1 has been exhaustively studied in rodents. Nkx2.1 is expressed in the thyroid, lung, and brain. In mouse embryos, Nkx2.1 is detected in the developing thyroid as soon as the thyroid anlage is visible (E8.5).[36] Interestingly, Nkx2.1 is present also in the epithelial cells of the fourth pharyngeal pouches forming the ultimobranchial body.[38] In the adult thyroid, Nkx2.1 maintains its expression in both follicular and parafollicular C cells.[39] During embryonic life, Nkx2.1 is detected in the epithelial cells of the developing trachea and lungs, while in adults it is present in bronchiolar epithelial Clara cells and in type II alveolar cells. As for the brain, Nkx2.1 is expressed in some areas of the developing diencephalon, such as the hypothalamic areas and the infundibulum, from which the neurohypophysis develops. Nkx2.1 expression in the adult hypothalamus is faint, but is upregulated before the onset of puberty.[40] In the dorsomedial and lateral hypothalamic nuclei, Nkx2.1 interacts with Sir2, a member of the sirtuin family, to control expression of orexin type 2 receptor (Ox2r), which plays a crucial role in the regulation of aging and longevity in mammals.[41]

In human embryos, the expression pattern of NKX2.1 is similar to that seen in mice, except that NKX2.1 is not detected in the fourth pharyngeal pouches.[42]

The mechanisms that induce the expression of Nkx2.1 in a restricted pool of foregut endodermal cells are still under investigation; data suggest that Nkx2.1 expression is regulated by many transcription factors, such as Sox2, FoxA1, FoxA2, and GATA6. Then, primordial Nkx2.1-positive foregut endodermal cells are induced to adopt a thyroid fate by the combination of FGF and BMP, whereas Wnt and BMP signaling promote lung specification.[43]

Gene-targeting experiments have enabled assessment of the *in vivo* role of this transcription factor. In the absence of Nkx2.1, newborn mice immediately die at birth and are characterized by impaired morphogenesis of both lung and brain, as well as lack of thyroid and of the entire pituitary[44] (Fig. 64.5). In *Nkx2.1*-null newborns, the lungs appear as dilated sac-like structures without normal pulmonary parenchyma.[44] Furthermore, the trachea has a reduced number of cartilage rings and it is not separated from the esophagus.[45] In the case of the brain, alterations in the ventral region of the forebrain are evident.[46] In addition, both neurohypophysis and anterior hypophysis are absent. As a consequence of the lack of signals from the pituitary, the morphogenesis of adrenal glands is impaired too.[44] In *Nkx2.1*-null embryos, the thyroid anlage forms in its correct position, but at an early stage the morphogenesis of the gland is impaired. The thyroid primordium by E10 is hypoplastic; subsequently, it undergoes degeneration, probably due to apoptosis. The ultimobranchial bodies undergo the same process; they form, but degenerate by E12.[38] Hence, Nkx2.1 is dispensable for the initial commitment of both follicular and parafollicular thyroid cells. However, all cell types that will form the thyroid gland require Nkx2.1 for their survival; in the absence of this transcription factor, no thyroid rudiment is detectable. Another mouse strain, in which *Nkx2.1* is disrupted only in the thyroid gland late in organogenesis, presents different phenotypes: either the thyroid follicles appear atrophic, or the thyroid shows a reduced number of dilated follicles.[47] This indicates the requirement of Nkx2.1 in adult life to maintain the follicular structure of the gland.

The transactivating capability of Nkx2.1 has been clearly demonstrated through studies on cultured cell lines. Sequences recognized by Nkx2.1 have been identified in the promoters of thyroid-specific/-enriched genes, such as *Tg*, *TPO*, *Tshr*, and *Nis*, as well as of the lung-specific genes *SP-A*, *SP-B*, *SP-C*, and *CC10*. However, the genetic pathways controlled by Nkx2.1, especially those that are active during organogenesis, are poorly understood.

Because cell proliferation and differentiation are hallmarks of both embryogenesis and cancer, the study of the role of Nkx2.1 in cancerous

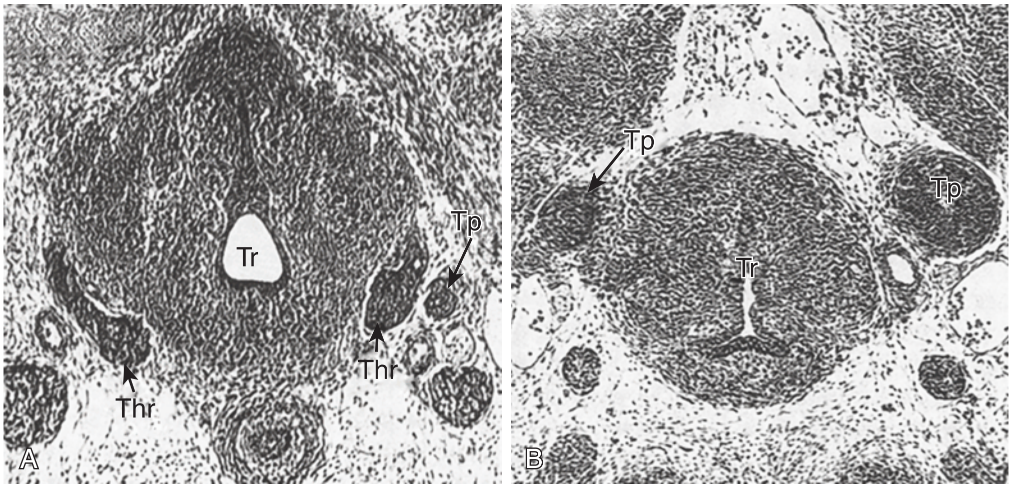


Fig. 64.5 **Thyroid gland phenotype in *Nkx2.1*-null mouse embryos.** Transversal sections of embryonic day E13 wild-type (**A**) and *Nkx2.1*$^{-/-}$ (**B**) mouse embryos stained with hematoxylin/eosin. In *Nkx2.1*-null embryos, the thyroid primordium is absent. *Thr,* Thyroid; *Tr,* trachea; *Tp,* thymic primordium. (From Kimura S, Hara Y, Pineau T, et al. The *T/ebp* null mouse: Thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* 1996;10:60–69.)

cells could provide additional information on the biological functions of this transcription factor. A germline mutation leading to a mutated Nkx2.1 protein (A339V) has been found in patients affected by papillary thyroid cancer (PTC) with a history of mononodular goiter and in their first-degree relatives affected by goiter without PTC.[48] Thyroid cells overexpressing the mutated Nkx2.1 protein display increased cell proliferation, activation of cell cycle regulator genes, and impaired expression of thyroid-specific genes.[48]

Pax8. *Pax8* (paired box gene 8) is a member of a family of transcription factors characterized by the presence of a 128–amino acid–long domain that recognizes and binds to specific DNA sequences. This DNA-binding domain is called a paired domain because it was identified for the first time in the *Drosophila* segmentation gene *paired*. Pax8 was identified in the mouse as a protein expressed in the developing thyroid gland.[34] Further studies demonstrated that the Pax8 paired domain recognizes and binds to a single site present in the *Tg* and in *Tpo* promoters.[49]

The gene encoding Pax8 is located on chromosome 2 in mice,[34] and the human ortholog, *PAX8*, is located on chromosome 2q12-q14. It consists of 12 exons that encode for different alternative spliced transcripts, at least six different transcripts in the mouse and five in humans. All the isoforms generated share the paired domain located near the N-terminus and differ in their C-terminal regions. Pax8a, the most abundant isoform, is a 457–amino acid–long protein in mice[34] and is 450 amino acids long in humans.

Pax8 is expressed in kidneys, in the nervous system, and in the thyroid.[34] In the developing kidney, at an early stage, *Pax8* is expressed in the nephrogenic cord and in mesonephric tubules; then it is present in the cortex of the metanephros, whereas in the adult kidney Pax8 is clearly detected in the medullar zone. In embryos, *Pax8* is transiently expressed in the myelecephalon,[34] in the uterine epithelium, and in the luminal epithelium of the oviduct and vagina. In males, strong *Pax8* mRNA expression was observed in the epithelium of the epididymis.[50]

As for the thyroid, Pax8 is expressed in the TFCs and in their precursors from the early stages of gland morphogenesis.

In humans, PAX8 is detected in the developing thyroid, kidneys, otic vesicle, central nervous system at E32,[42] and in both the luminal and glandular epithelium of the endometrium.[50]

Analysis of *Pax8* $^{-/-}$ mice[51](Fig. 64.6) offers the unique possibility of studying the *in vivo* role of this transcription factor. *Pax8*-null pups are born without any apparent brain or kidney defects. In contrast, the thyroid appears as a rudimental structure lacking TFCs, composed almost completely of calcitonin-producing C cells. The animals are affected by a severe hypothyroidism, show growth retardation, and die within 2 to 3 weeks after birth. The administration of thyroxine to these mice leads to their survival. However, in T4-treated female *Pax8*-null animals, the development of reproductive system is severely affected: the uterus appears as remnants of myometrium lacking endometrial structures, and there is no vaginal opening.[50] In *Pax8*-null embryos, the thyroid anlage is correctly formed, evaginates from the endoderm, and begins to migrate into the mesenchyme. However, by E11 the thyroid bud is much smaller compared to that of wild-type mice. In addition, in the absence of Pax8, other transcription factors, such as Foxe1 and Hhex, are downregulated in thyroid cell precursors.[7] At E12.5, TFCs are not detectable.[51] In *Pax8*-null mice, mature TFCs are absent, making it difficult to understand the role of this factor in the control of adult thyroid function. Another mouse strain, in which *Pax8* is disrupted late in gland development, exhibits severe hypothyroidism, hypoplasia of the gland, and reduced expression of thyroid-specific genes.[52] Thus, animal models point to a double essential role of Pax8 for both the survival and the differentiation of TFCs. It is worth noting that, in the developing thyroid, *Bcl-2*, an antiapoptotic gene, is downregulated in the absence of Pax8.[53]

As to the role of Pax8 in TFC differentiation, it has been well established by many reports. Functional assays have demonstrated that Pax8 is required to activate the *Tpo* promoter, and to a lesser extent the *Tg* promoter and the Nis enhancer.[36] In addition, Pax8 is able to activate expression of *Tg*, *Tpo* and *Nis* at their chromosomal loci.[54] A recent paper[55] indicates that, in primary culture of adult human thyrocytes, TSH-dependent upregulation of thyroid-specific genes (*TG*, *TPO*, *NIS*, *DIO2*, *TSHR*) required the simultaneous presence of *NKX2.1*, *FOXE1*, and *PAX8*. Furthermore, expression of these transcription factors does not required TSH, confirming the finding that mutated mice null for TSH/Tshr signaling do not show a reduced expression of Pax8.[20]

Through exhaustive genomic analyses, many other genes modulated by this factor have been identified, suggesting that *Pax8* can be

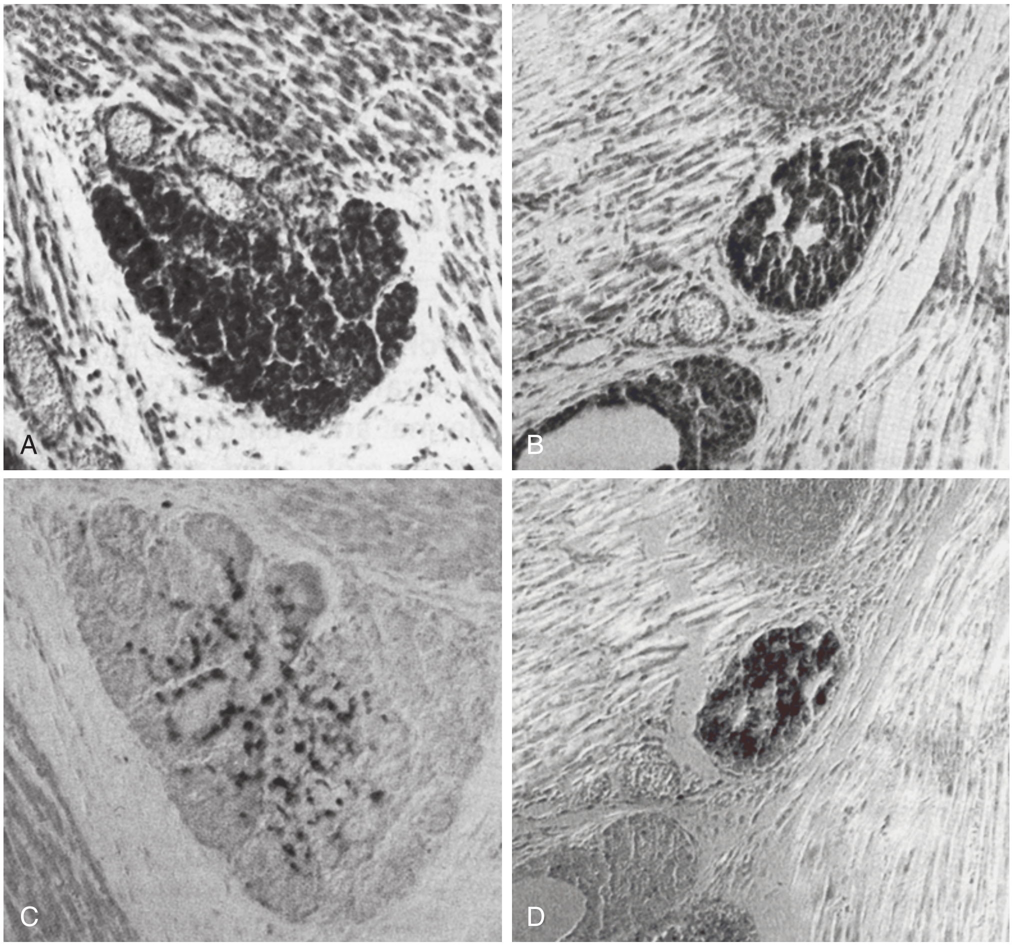


Fig. 64.6 Thyroid gland phenotype in *Pax8*$^{-/-}$ mouse embryos. Sagittal sections of embryonic day 18 wild-type (**A, C**) and *Pax8*$^{-/-}$ (**B, D**) mouse embryos stained with hematoxylin/eosin (**A, B**) and anticalcitonin-specific antibody (**C, D**). The thyroid gland of *Pax8*$^{-/-}$ embryos is smaller than that of wild-type mice and composed almost exclusively of C cells. (From Mansouri A, Chowdhury K, Gruss P. Follicular cells of the thyroid gland require *Pax8* gene function. *Nat Genet.* 1998;19:87–90.)

considered a master gene in both development and differentiation of the thyroid gland.

Foxe1. Foxe1 (formerly called TTF-2 for thyroid transcription factor–2) was originally identified as a thyroid-specific nuclear protein that can bind to a sequence present in both the *Tg* and *Tpo* promoters under insulin, IGF-1, or TSH stimulation. Subsequently, rat *Foxe1* cDNA was cloned, allowing characterization of the salient features of this protein.[33] Foxe1 belongs to the winged helix/forkhead family of transcription factors, which are characterized by a 100–amino acid–long DNA binding domain homologous to that of the *Drosophila forkhead* gene. The official name for the genetic locus encoding this transcription factor is *Foxe1* in mice (located on chromosome 4[33]) and *FOXE1* in humans (located on chromosome 9q22[12]). *Foxe1* is an intronless gene coding for a 42-kDa protein that is phosphorylated.[56] The protein contains an alanine stretch of variable length, ranging from 11 to 22 alanines. In humans, the most frequent *FOXE1* allele contains 14 alanine residues.[57] Interestingly, the presence of 16 alanines seems to be inversely associated with the risk of TD.[58]

Like *Nkx2.1* and *Pax8*, *Foxe1* is detected in the thyroid primordium, and its expression is maintained in TFCs during all stages of development and in adulthood. However, during embryonic life, *Foxe1* has a wide domain of expression. Indeed, at early stages of development, *Foxe1* is detected in the endodermal epithelium lining the primitive pharynx, the arches, and the foregut, as well as transiently in Rathke's pouch. Subsequently, *Foxe1* is expressed in tissues that develop from the pharynx and pharyngeal arches: thyroid, tongue, epiglottis, palate, choanae, and esophagus. In addition, Foxe1 is also detected in the whiskers and hair follicles, which derive from the ectoderm.[56] In humans, *FOXE1* is expressed in the thyroid, foregut embryonic thymus,[42] outer follicular hair sheath, and the seminiferous tubules of prepubertal testes.[59]

Studies in mutant mice have shown that *Foxe1* is tightly regulated in the thyroid bud by Pax8 and in the pharyngeal cells by Shh.[7] In differentiated rat thyroid cell lines, transcription of the *Foxe1* gene is under the tight control of TSH and cAMP, as well as that of insulin and IGF-1.[33] These data suggest that Foxe1 plays an important role in controlling interactions between hormone signaling pathways and thyroid-specific gene expression. In the developing thyroid, these controls do not seem to be effective, because in mutated mice that lack TSH, GH, and IGF-1, the expression of *Foxe1* is not affected.[20]

Analysis of *Foxe1*-null mice (Fig. 64.7) revealed the role of this transcription factor during thyroid development. Targeted inactivation of *Foxe1* showed that homozygous *Foxe1*$^{-/-}$ mice are born at the expected Mendelian ratio but die within 48 hours of birth. These mice display a severe cleft palate, probably responsible for the perinatal death, an absent thyroid or ectopic thyroid, lack of thyroid hormones, and elevated TSH levels in the bloodstream.[12] The early stages of thyroid morphogenesis,

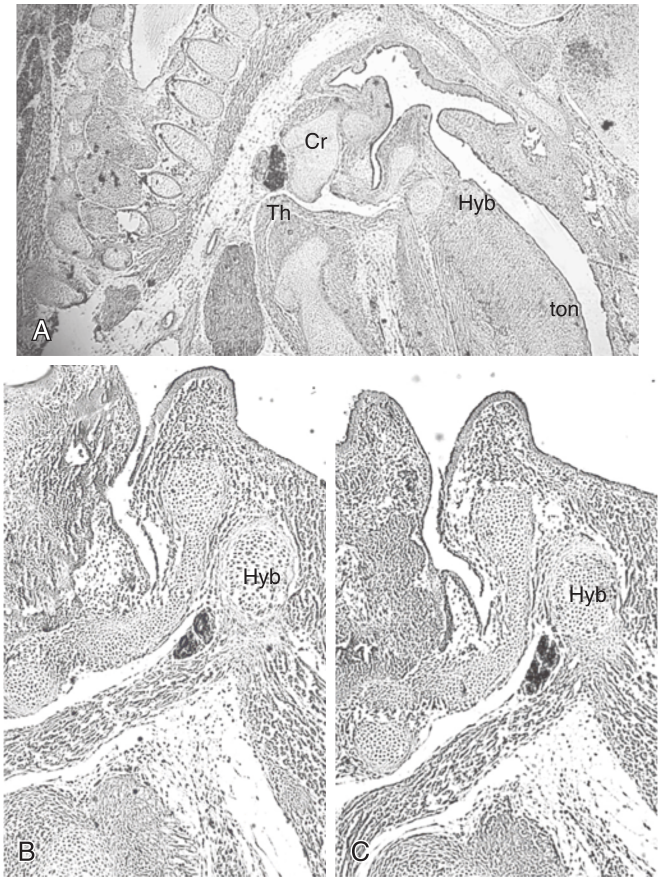


Fig. 64.7 Thyroid gland phenotype in *Foxe1*$^{-/-}$ mouse embryos. Sagittal sections of embryonic day 15.5 *Foxe1*$^{+/-}$ (A) and *Foxe1*$^{-/-}$ (B, C) embryos stained with an anti-*Nkx2.1* (A, B) or antithyroglobulin (C) antibody. The thyroid gland of the *Foxe1*$^{-/-}$ embryo is in a sublingual position. *Cr,* Cricoid cartilage; *Hyb,* hyoid bone; *Th,* thyroid; *ton,* tongue. (From De Felice M, Ovitt C, Biffali E et al. A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat Genet.* 1998;19:395–398.)

when the thyroid anlage is the formed, are not affected in *Foxe1*$^{-/-}$ embryos. However, at E10 thyroid precursor cells are still on the floor of the pharynx in *Foxe1*-null embryos, whereas in wild-type embryos the thyroid primordium begins to descend towards its final location. At later stages of development, in the absence of Foxe1, TFCs either disappear or form a small thyroid remnant still attached to the pharyngeal floor. In this case, the cells are able to complete their differentiation program, as shown by their ability to synthesize thyroglobulin. These data indicate that during embryonic life Foxe1 has a specific role in controlling the migration of TFC precursors, most likely by controlling the expression of target genes required for the migration process, but is not relevant for the specification and differentiation of thyroid anlage. In addition, Foxe1 could be involved in the survival of TFCs, because in many *Foxe1*-null embryos the thyroid primordium disappears.[7,12]

Recently, a genomic-wide screen performed in *Foxe1*-silenced thyroid cells indicated that both Duox2 and NIS are transcriptional targets of Foxe1.[60]s However, it is still hard to define the functions of this gene in the physiology of the gland, because *Foxe1*-null newborns die at birth.

Hhex. Hhex (formerly known as Hex for hematopoietically expressed homeobox, or Prh for proline-rich homeobox) is a homeodomain-containing transcription factor that was first identified in multipotent hematopoietic cells. It was later demonstrated that Hhex is expressed in other tissues, including the thyroid.[35] Hhex is encoded by a gene called *Hhex* in mice and *HHEX* in humans, which are located on chromosome 19 and chromosome 10q23.32, respectively. The gene has four exons and codes for a protein that is 270 amino acids long. *Hhex* is an orphan homeobox–containing gene, because the sequence of its homeodomain, which is responsible for binding to DNA, shows some differences with respect to other homeodomains. Outside the homeodomain, Hhex contains an N-terminal proline-rich region and a C-terminal acid region. These two regions are probably involved in repressing the transcription of the target genes.

In mouse embryos, *Hhex* is expressed very early in the primitive and then in the definitive endoderm. At E7.0 *Hhex* is detected in the developing blood islands, in the ventral foregut endoderm at E8.5 to E9.0, and in the endothelium of the developing vasculature and heart at E9.0. Hhex is an early marker of thyroid cells, because it is already expressed in the thyroid anlage at E8.5,[35] at the same stage in which *Nkx2.1*, *Foxe1,* and *Pax8* are detected. In the adult, in addition to the thyroid, *Hhex* expression is maintained in liver and lungs.

Analysis of mutated mice has revealed that this factor is absolutely necessary in many developmental processes. *Hhex*$^{-/-}$ embryos die at midgestation (E13.5–E15.5) and show severe defects in liver, forebrain, heart, and thyroid morphogenesis.[61] In *Hhex*-null embryos, thyroid precursor cells are present and express Nkx2.1, Pax8, and Foxe1 until E9. At E10 the thyroid primordium is composed of a few cells that do not express Nkx2.1, Foxe1, or Pax8.[61] Hence, Hhex guarantees the survival of TFC precursors and maintains the expression of Foxe1, Nkx2.1, and Pax8. Studies in differentiated thyroid cells suggest that the network controlling Hhex and the other thyroid-specific transcription factors is complex. *Hhex* seems to be regulated by Nkx2.1,[36] and overexpression of *Hhex* partly inhibits *Tg* promoter activity. These data are consistent with the hypothesis that Hhex is a transcriptional repressor in thyroid cells, as reported in other systems.

Thyroid-Enriched Transcription Factor Interactions. The analysis of the phenotype of *Nkx2.1*, *Hhex*, *Pax8*, and *Foxe1* knockout mice highlights a complex network of reciprocal regulatory interactions among these transcription factors.[7] At an early stage of thyroid development, the initial *Nkx2.1*, *Hhex*, and *Pax8* expression seem to be independent from each other, because the absence of any of these genes does not affect the expression of the others. The expression of *Foxe1*, on the contrary, is tightly controlled by *Pax8*. By E10, in the thyroid placode, each affects expression of the others. *Foxe1* seems to be located downstream in this regulatory network, because it is expressed only if Titf1, Hhex, and Pax8 are simultaneously present, while in *Foxe1*-null mice *Nkx2.1*, *Hhex,* and *Pax8* are correctly expressed in the thyroid bud. Interestingly, in the human thyroid the expression of both *NKX2.1* and *PAX8* precedes the onset of *FOXE1* expression.[42] This genetic cascade has been conserved during evolution. In zebrafish, at an early stage of thyroid development, *hhex*, *nk2.1,* and *pax2.1* (which has a role comparable to murine *Pax8* in the zebrafish thyroid) expression are not dependent on each other, while at a later stage the expression of each gene requires the presence of the other genes.[62]

The network summarized above is due, at least in part, to regulatory interplay at the transcriptional level. *Pax8* is a target of Nkx2.1, and reciprocally, Nkx2.1 is a target of Pax8.[63] The *Hhex* promoter is activated by both Nkx2.1 and Pax8.[64] Finally, autoregulatory loops contribute to maintenance of *Nkx2.1*, *Pax8,* and *Hhex* expression, because each factor stimulates transcription from its own promoter[65] (Fig. 64.8). In addition to the transcriptional network, physical and functional interactions between Nkx2.1 and Pax8 have been demonstrated. Nkx2.1 and Pax8 form a protein complex and cooperate in activating the *Tg* promoter. The cooperation between these two factors

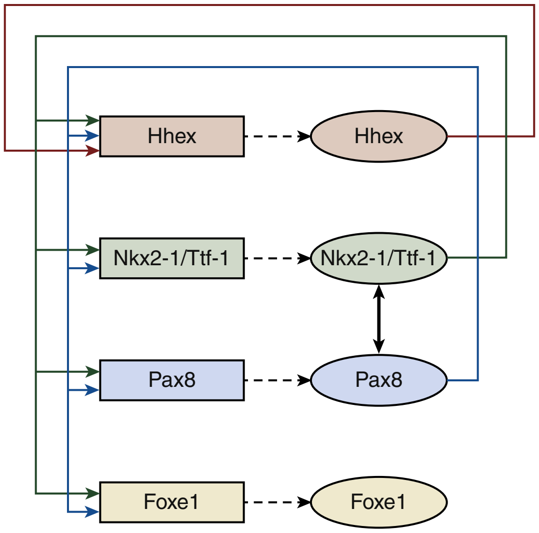


Fig. 64.8 **Physical and regulatory interactions among Hhex, Nkx2.1/Ttf-1, Pax8, and Foxe1.** Factors are represented by *ovals*; gene promoters are represented by *rectangles*. Binding of a regulator to a promoter is indicated by a *solid arrow*. Genes encoding regulators are linked to their respective regulators by *dashed arrows*. Physical interaction between factors is indicated by a *thick line arrow*. (From De Felice M, Di Lauro R. Intrinsic and extrinsic factors in thyroid gland development: an update. *Endocrinology*. 2011;152:2948–2956.)

is confirmed *in vivo*. Actually, double heterozygous mice for the null allele of *Nkx2.1* and *Pax8* show severe hypothyroidism characterized by thyroid hypoplasia,[18] although *Pax8* or *Nkx2.1* single heterozygous null mice do not show defects in thyroid morphogenesis.

The extreme relevance of the functional interaction between these two proteins has received strong support by the finding that aggregated mouse embryonic stem cells overexpressing both *Pax8* and *Nkx2.1* form thyroid follicular organoids that express thyroid-specific genes[66] (see later, Generation of Functional Thyroid Follicles *In Vitro*).

From Animal Models to Human Diseases

Eighty-five percent of cases of CH, the most frequent endocrine disorder in newborns, are due to TD. The term "thyroid dysgenesis" indicates an ectopic or hypoplastic thyroid, as well as the absence of thyroid (thyroid agenesis or athyreosis). Thyroid dysgenesis is due to anomalies during gland organogenesis: defects in growth and/or differentiation of the thyroid primordium can result in an absent or hypoplastic thyroid, and impaired migration of thyroid precursor cells causes an ectopic gland.

The discovery that the disruption of genes involved in thyroid morphogenesis causes impaired development of the gland in mice has led researchers to look for mutations in the homologous human genes in patients affected by CHTD. These studies have been successful, because a minority of cases of CHTD are associated with mutations in genes relevant for thyroid development, confirming data obtained in animal models.[11,36] Although these cases account only for a small minority (3%–5%) of cases of CHTD, this finding proves unequivocally that CHTD can be a genetic and inheritable affection due to defects in genes involved in thyroid development. New technologies such as next-generation sequencing have made it possible to identify a number of mutated genes in patients affected by CHTD.[67] However, in many cases the causative role of these mutations has to be confirmed. The genetic landscape of CHTD has been recently and exhaustively reported in a number of excellent reviews.[68]

In this section we will focus only on cases of CHTD associated with mutations in *NKX2.1*, *FOXE1*, and *PAX8* in which a clear pattern of Mendelian inheritance can be observed. The integration of these studies with data obtained from animal models played a valuable role in providing insights into the molecular mechanisms that underlie normal and pathological thyroid development.

Thyroid Dysgenesis. Both athyreosis and agenesis indicate the absence of thyroid tissue. However, agenesis should be used to define the absence of the gland due to a defective initiation of thyroid morphogenesis.[36] The term athyreosis indicates a dysgenesis characterized by the disappearance of the thyroid due to alterations in any step following the specification of the thyroid anlage. In several patients with athyreosis, the presence of cystic structures, probably resulting from degeneration of the TFC clusters, suggests that thyroid specification was not impaired in these subjects. At the moment, there are no animal models that display a *bona fide* agenesis. Indeed, gene-targeting experiments have demonstrated that Nkx2.1, Foxe1, Pax8, and Hhex do not play a role in specification of the thyroid anlage, which forms correctly in embryos deprived of any of these proteins. Agenesis could be due to defects in genes involved in the early regionalization of the endoderm, including the genes controlling the onset of *NKX2.1*, *FOXE1*, *PAX8*, and *HHEX* expression. However, the relevance of these genes in CH is difficult to demonstrate, because mutations in genes that are widely expressed in the endoderm at early stages of development could cause embryonic-lethal phenotypes and/or many additional defects. On the contrary, the knockout mice described in the previous section are good models of athyreosis. In these mutants, the morphogenesis of the gland begins, but the thyroid bud disappears, probably due to a defective survival program or due to proliferation of the precursors of the follicular cells.

Thyroid hypoplasia (a gland containing a reduced number of follicular cells) is probably a genetically heterogeneous disorder and could be a consequence of alterations in any of the steps that control the expansion and/or survival of thyroid cells during organogenesis. Genes involved in early stages of development (*NKX2.1*, *FOXE1*, *PAX8*, and *HHEX*), as well as in the late steps, could be good candidates for this dysgenesis. Indeed, mutations in *NKX2.1*, *PAX8*, or *FOXE1* have been found in patients with CH associated with hypoplasia.

The Ectopic Thyroid is a Consequence of an Impaired Thyroid Bud Migration. Familial cases of TD have been described, in which affected members show either ectopic thyroid or athyreosis. This observation raises the possibility that athyreosis and thyroid ectopy have a common underlying mechanism. This hypothesis is consistent with the data from mice lacking the Foxe1 protein, which show either ectopy with a very small thyroid or no thyroid at all.[12]

Thyroid-Enriched Transcription Factors and Thyroid Dysgenesis. Consistent with its expression pattern, mutations have been found in *NKX2.1* in patients affected by brain-lung-thyroid syndrome (OMIM 610978), a complex syndrome characterized by thyroid alterations, choreoathetosis, and lung abnormalities.

All affected *NKX2.1* individuals are heterozygous for the mutation; patients carrying loss-of-function mutations in both alleles have never been reported. It is probable that such a condition would be not compatible with the life, like the corresponding *Nkx2.1*-null mice. After initial reports of two patients with a large chromosomal deletion that includes the *NKX2.1* gene,[69,70] the first mutations were identified several years ago.[71] Until now many different mutations have been identified, including chromosomal deletions and point mutations, with the latter mainly resulting in either truncated proteins or proteins with impaired DNA binding and transactivation properties.[56] In mice, *Nkx2.1* functions are partially dosage-sensitive. Indeed, functional studies have revealed that $Nkx2.1^{+/-}$ mice present decreased coordination and mild hyperthyrotropinemia.[72] According to this finding, the dominant effect of *NKX2.1* mutations in humans could be due to a haploinsufficiency.

NKX2.1$^{+/-}$ subjects display a highly variable phenotype. The classic triad, brain-lung-thyroid syndrome, is not always present. Furthermore, the severity both of chorea and of respiratory symptoms varies largely.

As to thyroid alterations, these have not been reported in many patients; in other cases, patients show elevated TSH levels with mild or severe hypoplasia of the gland or athyreosis, albeit in a minority of cases.[73] There is no clear correlation between the phenotype severity and the type of mutations. A variable phenotype has been described also in the same familial cluster; in addition, it has been reported that a pair of monozygotic twins carrying a loss-of-function mutation at the splice consensus of intron 2 of *NKX2.1* gene are strongly discordant for thyroid phenotype.[73] Actually, one twin showed hemiagenesis of the gland and mild hypothyroidism, while the other twin presented a severe hypothyroidism and athyreosis. Taken together, these data suggest that modifier genes and epigenetic mechanisms could be responsible for the incomplete penetrance and variability of the phenotype.

Mutations in *PAX8* associated with sporadic and familial cases of CHTD were reported for the first time in the late 1990s.[74] By now, more than 50 variants have been reported,[75] especially in the exons coding the DNA-binding paired domain. In only a few cases, missense or truncating mutations in the sequences encoding for C-terminus of the proteins have been described. *In vitro* studies have demonstrated that some of these mutations impair PAX8–NKX2.1 interactions, which are required for synergistic activation of thyroid-specific genes such as *TPO* and *Tg*.[76]

While *Pax8*$^{+/-}$ mice display a normal phenotype,[51] all but one patient reported until now are heterozygous for the mutations, and in the familial cases the mode of inheritance is dominant. Thus, in humans, both fully functional *PAX8* alleles are necessary for correct thyroid morphogenesis, and haploinsufficiency could cause CHTD. In the majority of cases, the mutated protein shows diminished transactivation ability due to loss of DNA-binding capability. In the case of proteins transcriptionally inactive but still able to bind the target sequences, the diminished transactivation capability has been attributed to an impaired synergy with the coactivator p300 or NKX2.1 in activating thyroid-specific promoters.

In most cases *PAX8* mutations are associated to nonsyndromic CHTD. However, some patients display other defects, such as unilateral renal hemiagensis or abnormalities in the urogenital tract.[77] In two cases, deletions encompassing *PAX8* were associated with hypothyroidism and Mullerian agenesis.[78]

The thyroid phenotype is variable, ranging from a normal-sized gland to a severe hypoplasia and even athyreoisis. Because the phenotype shows discrepancy even in familial cases in which related individuals carry the same mutation, effects of other modifier genes could be hypothesized. In some patients, morphological and functional defects of the thyroid become evident only after birth, and overt hypothyroidism develops some years later.[79,80] This finding demonstrates that, in the postnatal thyroid, the survival and proliferation of the follicular cells require a full dosage of PAX8.

In accordance with the phenotype of *Foxe1*-null mice, homozygous loss-of-function mutation in the *FOXE1* gene was first reported[81] in two siblings affected by syndromic CH, characterized by athyreosis, cleft palate, bilateral choanal atresia, and spiky hair (Bamforth–Lazarus syndrome[82]). Until now many different mutations have been identified.[68] All affected individuals are homozygous for the mutations, and in the familial cases the mode of inheritance is recessive. The mutations are inherited from heterozygous carrier parents, apart from a case in which the mechanism of transmission was uniparental isodisomy.

All identified mutations cause substitutions within the FOXE1 forkhead domain, and the mutated proteins show reduced binding and transcriptional activity. However, a novel gain-of-function mutation of the *FOXE1* gene has recently been reported. Indeed, even if the patient presents features typical of Bamforth–Lazarus syndrome, functional studies show that mutated FOXE1 enhances transcriptional activity at both the *TG* and *TPO* promoters.[83] This finding seems to be consistent with the observation that both *Foxe1*-null[12] and *Foxe1* overexpressing mice[84] present cleft palate, one of the hallmark signs of Bamforth–Lazarus syndrome. The thyroid phenotype associated with *FOEX1* mutations ranges from athyreosis to severe hypoplasia. Notably, in mice the absence of this factor causes either athyreosis or defects in thyroid migration, while in humans *FOXE1* mutations associated with an ectopic thyroid have not been described.

Generation of Functional Thyroid Follicles *In Vitro*

The large amount of information collected on the molecules involved in the formation of a functional thyroid gland, recapitulated in the previous sections, together with the increasing availability of several types of cells (for brevity here called stem cells, of which there are however different types) capable of differentiation into diverse cell types, prompted several groups to use these cells to try to form thyroid tissue *in vitro*. The first attempts[85,86] showed that it was possible to promote partial thyroid differentiation starting from mouse embryonic stem cells. Even though no thyroglobulin expression was present, and hence no thyroid hormone was formed, the induction of other thyroid differentiation markers in these experiments established the important principle that it was possible to obtain some thyroid features starting from multipotent, nondifferentiated stem cells. Using a different strategy, a more successful attempt resulted in functional thyroid tissue starting from mouse embryonic stem cells in culture.[66] The approach taken involved forced but transient expression of *Nkx2.1* and *Pax8*, followed by TSH treatment and culturing in Matrigel to allow formation of follicular structures. The follicles that formed therein, in addition to expressing the entire set of thyroid differentiation markers (*Tg, TPO, NIS, Tshr,* and *Foxe1*, in addition to endogenous *Nkx2.1* and *Pax8*), were able, when transplanted under the kidney capsule, to rescue the severe hypothyroidism consequent to thyroid ablation by ^{131}I treatment. These stunning results showed both the relevance of basic research in promoting potentially new, revolutionary applications of regenerative medicine, as well as the courage and the perseverance of the researcher responsible. Of course it will take still some time before a patient with CH will be able to do without the daily doses of thyroid hormones, but the path is clear. It should be emphasized that, in the study presenting such results, no additional information was provided on the mechanisms that induce the beginning of thyroid development *in vivo*, because the authors skipped the phase of thyroid commitment that occurs *in vivo*.

A different experimental approach attempted to recapitulate early events in foregut development and from there try to establish functional thyroid tissue. A first set of observations demonstrated that it was possible to obtain foregut endoderm *in vitro* starting from mouse embryonic stem cells, using a combination of growth factors.[87] Based on those observations, a differentiation protocol was set up that resulted in functional thyroid follicles capable of rescuing severe hypothyroidism in mice.[88] These results also provide important data on the essential role of combined FGF and BMP signaling in promoting very early phases of thyroid specification.

KEY POINTS

- Normal development of the thyroid gland requires a set of regulatory molecules (transcription factors) whose role has been firmly assessed in animal models. A few patients with mutations in these transcription factors show associated thyroid dysgenesis, thus demonstrating that congenital hypothyroidism can be a heritable condition, even though this is only true for a very small number of cases.
- Functional thyroid cells organized into follicles have been recently obtained *in vitro*. The cells are able to rescue severe hypothyroidism in mice, opening the way to a replacement therapy based on cells for the lack or severe reduction of thyroid tissue.

PHYLOGENESIS OF THE THYROID GLAND

The morphological bases of thyroid function are the same in all vertebrates, as reflected by the conserved follicular structure throughout vertebrate evolution. On the contrary, the gross anatomy of the thyroid gland is different in the vertebrate classes. In placental mammals, and in some reptiles, the thyroid is composed of two lobes connected by an isthmus that crosses the trachea. In nonplacental mammals, birds, and amphibians, the thyroid consists of two isolated lobes. In cartilaginous fish and in some teleosts (mostly marine ones), the thyroid is massed into a compact organ. In the remaining marine and almost all freshwater teleosts, the thyroid consists of non-encapsulated follicles scattered in the subpharyngeal connective tissue. Often, heterotopic follicles are found in nonpharyngeal areas such as the kidneys, heart, esophagus, or spleen in these animals. Interestingly, the ability of thyroid cells to proliferate in these ectopic areas has been shown to be genetically controlled in platyfish, and low iodine intake represents a strong inductive stimulus towards this condition. Even though iodine-concentrating activity and iodamino acid production are ancient phenomena that have also been described in nonvertebrate animals, only cyclostomes (such as the lamprey) have a *bona fide* thyroid characterized by follicles scattered in the subpharyngeal connective tissue.

The ontogeny of the thyroid follows the same pattern in all vertebrates, including teleosts[89]: the thyroid anlage arises from an outpouching of the primitive pharynx, migrates caudally, and reaches its definitive position. The development of the thyroid in cyclostomes follows a different pattern from that used by the higher vertebrates. Thyroid follicles in the adult lamprey derive from the endostyle, an organ present only during larval life. In the larval lamprey, at the early stages of development, the endostyle forms as a ciliated groove in the ventral part of the pharynx. During development the groove becomes a cylinder and then a complex structure connected to the pharyngeal cavity by the ductus hypobranchialis.[90] The epithelium of the endostyle differentiates into different types of cells, a group of which displays the ability to trap iodine, exhibits peroxidase activity, and expresses thyroxine.[91] These cells are fated to become “typical” follicular cells only after metamorphosis. A structure similar to the lamprey endostyle is present in the ventral pharynx of protochordates such as cephalocordates (amphioxus) and urochordates (such as ascidians). This organ, which is also called endostyle, is composed of different cell types, the majority of which are involved in the food capturing process.[90] More than a century ago, the homology between the cyclostome endostyle and the thyroid suggested the hypothesis that the endostyle of protochordates was a primitive antecedent of the vertebrate thyroid gland.[92] This hypothesis, based on morphological observation, has gained support from the presence of iodine trapping, thyreoperoxidase activity, and *Ci-Duox*[93] (the *Ciona* homolog of the vertebrate *Duox*) in a group of cells of the endostyle of protochordates. Thus, these cells can be considered “protothyroid” cells.[90]

The molecular mechanisms involved in the differentiation of thyroid cells of mammals are conserved throughout the evolution of protochordates to vertebrate chordates. In urochordates, (*Ciona intestinalis*) the group of cells that express the *TPO* homolog[94] also express *CiFoxE*, a *Foxe1* ortholog,[95] but not *CiNkx2.1*.[96] On the contrary, in amphioxus, both *Nkx2.1*- and *TPO*-homologous genes are coexpressed in the same cells of the endostyle,[97] while the *Foxe1* ortholog *AmphiFoxE4* is not detected in the endostyle, but rather in another pharyngeal-derived structure.[98] Finally, in urochordates, cephalochordates, and lampreys, *Pax2/5/8* (*Pax8* ortholog) is always expressed in the group of cells of the endostyle that exhibit iodide uptake and TPO activities.[99]

These data suggest that, during the evolution of protochordates, a group of genes whose function was to contribute to the specification of the anterior part of the gut were selected as part of the genetic program that leads to thyroid specification. In particular, *Pax8* and its ortholog gene could have had a specific role in the genetic mechanism controlling the differentiation of both protothyroid and true follicular cells, revealing a specific master gene of thyroid cells.

REFERENCES

1. Rousset B, Dupuy C, Miot F, et al. Chapter 2: thyroid hormone synthesis and secretion. In: Feingold KR, Anawalt B, Boyce A, et al., ed. *Endotext*. South Dartmouth: MDText.com, Inc.; 2000.
2. Fagman H, Andersson L, Nilsson M. The developing mouse thyroid: embryonic vessel contacts and parenchymal growth pattern during specification, budding, migration, and lobulation. *Dev Dyn*. 2006;235:444–455.
3. Wollman SH, Neve P. Postnatal development and properties of ultimobranchial follicles in the rat thyroid. *Anat Rec*. 1971;171:247–258.
4. Vázquez-Román V, Utrilla JC, Fernández-Santos JM, et al. Postnatal fate of the ultimobranchial remnants in the rat thyroid gland. *J Morphol*. 2013;274:725–732.
5. Iwadate M, Takizawa Y, Shirai YT, et al. An in vivo model for thyroid regeneration and folliculogenesis. *Lab Invest*. 2018;98:1126–1132.
6. Ma R, Morshed SA, Latif R, et al. A stem cell surge during thyroid regeneration. *Front Endocrinol (Lausanne)*. 2020;11:606269.
7. Parlato R, Rosica A, Rodriguez-Mallon A, et al. An integrated regulatory network controlling survival and migration in thyroid organogenesis. *Dev Biol*. 2004;276:464–475.
8. Elsalini OA, von Gartzen J, Cramer M, et al. Zebrafish hhex, nk2.1a, and pax2.1 regulate thyroid growth and differentiation downstream of Nodal-dependent transcription factors. *Dev Biol*. 2003;263:67–80.
9. Vandernoot I, Haerlingen B, Gillotay P, et al. Enhanced canonical Wnt signaling during early zebrafish development perturbs the interaction of cardiac mesoderm and pharyngeal endoderm and causes thyroid specification defects. *Thyroid*. 2021;31:420–438.
10. Haerlingen B, Opitz R, Vandernoot I, et al. Small-molecule screening in zebrafish embryos identifies signaling pathways regulating early thyroid development. *Thyroid*. 2019;29:1683–1703.
11. Polak M, Sura-Trueba S, Chauty A, et al. Molecular mechanisms of thyroid dysgenesis. *Horm Res*. 2004;62:14–21.
12. De Felice M, Ovitt C, Biffali E, et al. A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat Genet*. 1998;19:395–398.
13. Fagman H, Grande M, Edsbagge J, et al. Expression of classical cadherins in thyroid development: maintenance of an epithelial phenotype throughout organogenesis. *Endocrinology*. 2003;144:3618–3624.
14. Fagman H, Nilsson M. Morphogenesis of the thyroid gland. *Mol Cell Endocrinol*. 2010;323:35–54.
15. Fagman H, Grande M, Gritli-Linde A, et al. Genetic deletion of sonic hedgehog causes hemiagenesis and ectopic development of the thyroid in mouse. *Am J Pathol*. 2004;164:1865–1872.
16. Fagman H, Liao J, Westerlund J, et al. The 22q11 deletion syndrome candidate gene Tbx1 determines thyroid size and positioning. *Hum Mol Genet*. 2007;16:276–285.

17. Opitz R, Hitz MP, Vandernoot I, et al. Functional zebrafish studies based on human genotyping point to netrin-1 as a link between aberrant cardiovascular development and thyroid dysgenesis. *Endocrinology*. 2015;156:377–388.
18. Amendola E, Sanges R, Galvan A, et al. A locus on mouse chromosome 2 is involved in susceptibility to congenital hypothyroidism and contains an essential gene expressed in thyroid. *Endocrinology*. 2010;151:1948–1958.
19. Lazzaro D, Price M, De Felice M, et al. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development*. 1991;113:1093–1104.
20. Postiglione MP, Parlato R, Rodriguez-Mallon A, et al. Role of the thyroid-stimulating hormone receptor signaling in development and differentiation of the thyroid gland. *Proc Natl Acad Sci U S A*. 2002;99:15462–15467.
21. Milenkovic M, De Deken X, Jin L, et al. Duox expression and related H2O2 measurement in mouse thyroid: onset in embryonic development and regulation by TSH in adult. *J Endocrinol*. 2007;192:615–626.
22. Meunier D, Aubin J, Jeannotte L. Perturbed thyroid morphology and transient hypothyroidism symptoms in Hoxa5 mutant mice. *Dev Dyn*. 2003;227:367–378.
23. Marians RC, Ng L, Blair HC, et al. Defining thyrotropin-dependent and -independent steps of thyroid hormone synthesis by using thyrotropin receptor-null mice. *Proc Natl Acad Sci U S A*. 2002;99:15776–15781.
24. De Felice M, Postiglione MP, Di Lauro R. Minireview: thyrotropin receptor signaling in development and differentiation of the thyroid gland: insights from mouse models and human diseases. *Endocrinology*. 2004;145:4062–4067.
25. Silberschmidt D, Rodriguez-Mallon A, Mithboakar P, et al. In vivo role of different domains and of phosphorylation in the transcription factor Nkx2.1. *BMC Dev Biol*. 2011;11:9.
26. Frezzetti D, Reale C, Cali G, et al. The microRNA-processing enzyme Dicer is essential for thyroid function. *PLoS One*. 2011;6:e27648.
27. De Felice M, Di Lauro R. Intrinsic and extrinsic factors in thyroid gland development: an update. *Endocrinology*. 2011;152:163–175.
28. Carre A, Rachdi L, Tron E, et al. Hes1 is required for appropriate morphogenesis and differentiation during mouse thyroid gland development. *PLoS One*. 2011;6:e16752.
29. Manley NR, Capecchi M. Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands. *Dev Biol*. 1998;195:1–15.
30. Lania G, Zhang Z, Huynh T, et al. Early thyroid development requires a Tbx1–Fgf8 pathway. *Dev Biol*. 2009;328:109–117.
31. Le Douarin N, Fontaine J, LeLievre C. New studies on the neural crest origin of the avian ultimobranchial glandular cells. Interspecific combinations and cytochemical characterization of C cells based on the uptake of biogenic amine precursors. *Histochemie*. 1974;38:297–305.
32. Johansson E, Andersson L, Ornros J, et al. Revising the embryonic origin of thyroid C cells in mice and humans. *Development*. 2015;142:3519–3528.
33. Zannini M, Avantaggiato V, Biffali E, et al. TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation. *EMBO J*. 1997;16:3185–3197.
34. Plachov D, Chowdhury K, Walther C, et al. Pax8, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development*. 1990;110:643–651.
35. Thomas PQ, Brown A, Beddington R. Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development*. 1998;125:85–95.
36. De Felice M, Di Lauro R. Thyroid development and its disorders: genetics and molecular mechanisms. *Endocr Rev*. 2004;25:722–746.
37. De Felice M, Di Lauro R. Murine models for the study of thyroid gland development. *Endocr Dev*. 2007;10:1–14.
38. Kusakabe T, Hoshi N, Kimura S. Origin of the ultimobranchial body cyst: T/ebp/Nkx2.1 expression is required for development and fusion of the ultimobranchial body to the thyroid. *Dev Dyn*. 2006;235:1300–1309.
39. Suzuki K, Kobayashi Y, Katoh R, et al. Identification of thyroid transcription factor-1in C cells and parathyroid cells. *Endocrinology*. 1998;139:3014–3017.
40. Lee BJ, Cho GJ, Norgren Jr RB, et al. TTF-1, a homeodomain gene required for diencephalic morphogenesis, is postnatally expressed in the neuroendocrine brain in a developmentally regulated and cell-specific fashion. *Mol Cell Neurosci*. 2001;17:107–126.
41. Satoh A, Brace CS, Rensing N, et al. Sirt1 extends life span and delays aging in mice through the regulation of Nk2 homeobox 1 in the DMH and LH. *Cell Metab*. 2013;18:416–430.
42. Trueba SS, Auge J, Mattei G, et al. PAX8, TITF1 and FOXE1 gene expression patterns during human development: new insights into human thyroid development and thyroid dysgenesis associated malformations. *J Clin Endocrinol Metab*. 2004;90:455–462.
43. Serra M, Alysandratos KD, Hawkins F, et al. Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung- versus thyroid-lineage specification. *Development*. 2017;144:3879–3893.
44. Kimura S, Hara Y, Pineau T, et al. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev*. 1996;10:60–69.
45. Minoo P, Su G, Drum H, et al. Defects in tracheoesophageal and lung morphogenesisn in Nkx2.1(-/-) mouse embryos. *Devel Biol*. 1999;209:60–71.
46. Sussel L, Marin O, Kimura S, et al. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular rerspecification within the basal telencephalon: evidence for a transformation of the pallidum into the. *Development*. 1999;126:3359–3370.
47. Kusakabe T, Kawaguchi A, Hoshi N, et al. Thyroid-specific enhancer-binding protein/NKX2.1 is required for the maintenance of ordered architecture and function of the differentiated thyroid. *Mol Endocrinol*. 2006;20:1796–1809.
48. Ngan ES, Lang BH, Liu T, et al. A germline mutation (A339V) in thyroid transcription factor-1 (TITF-1/NKX2.1) in patients with multinodular goiter and papillary thyroid carcinoma. *J Natl Cancer Inst*. 2009;101:162–175.
49. Francis-Lang H, Zannini MS, De Felice M, et al. Multiple mechanisms of interference between transformation and differentiation in thyroid cells. *Mol Cell Biol*. 1992;12:5793–5800.
50. Mittag J, Winterhager E, Bauer K, et al. Congenital hypothyroid female pax8-deficient mice are infertile despite thyroid hormone replacement therapy. *Endocrinology*. 2007;148:719–725.
51. Mansouri A, Chowdhury K, Gruss P. Follicular cells of the thyroid gland require *Pax8* gene function. *Nat Genet*. 1998;19:87–90.
52. Marotta P, Amendola E, De Luca P, et al. Pax8 expression is required for the differentiation of thyroid follicular cells. *Eur Thyroid J*. 2012;1:83.
53. Fagman H, Amendola E, Parrillo L, et al. Gene expression profiling at early organogenesis reveals both common and diverse mechanisms in foregut patterning. *Dev Biol*. 2011;359:163–175.
54. Pasca di Magliano M, Di Lauro R, Zannini MS. Pax8 has a key role in thyroid cell differentiation. *Proc Natl Acad Sci U S A*. 2000;97:13144–13149.
55. Jang D, Marcus-Samuels B, Morgan SJ, et al. Thyrotropin regulation of differentiated gene transcription in adult human thyrocytes in primary culture. *Mol Cell Endocrinol*. 2020;518:111032.
56. Dathan N, Parlato R, Rosica A, et al. Distribution of the titf2/foxe1 gene product is consistent with an important role in the development of foregut endoderm, palate, and hair. *Dev Dyn*. 2002;224:450–456.
57. Macchia PE, Mattei MG, Lapi P, et al. Cloning, chromosomal localization and identification of polymorphisms in the human thyroid transcription factor 2 gene (TITF2). *Biochimie*. 1999;81:433–440.
58. Castanet M, Polak M. Spectrum of human Foxe1/TTF2 mutations. *Horm Res Paediatr*. 2010;73:423–429.
59. Sequeira M, Al-Khafaji F, Park S, et al. Production and application of polyclonal antibody to human thyroid transcription factor 2 reveals thyroid transcription factor 2 protein expression in adult thyroid and hair follicles and prepubertal testis. *Thyroid*. 2003;13:927–932.
60. Fernández LP, López-Márquez A, Martínez AM, et al. New insights into FoxE1 functions: identification of direct FoxE1 targets in thyroid cells. *PLoS One*. 2013;8:e62849.
61. Martinez Barbera JP, Clements M, Thomas P, et al. The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development*. 2000;127:2433–2445.

62. Wendl T, Adzic D, Schoenebeck JJ, et al. Early developmental specification of the thyroid gland depends on han-expressing surrounding tissue and on FGF signals. *Development*. 2007;134:2871–2879.
63. Nitsch R, Di Dato V, di Gennaro A, et al. Comparative genomics reveals a functional thyroid-specific element in the far upstream region of the PAX8 gene. *BMC Genom*. 2010;11:306.
64. Puppin C, Presta I, D'Elia AV, et al. Functional interaction among thyroid-specific transcription factors: pax8 regulates the activity of Hex promoter. *Mol Cell Endocrinol*. 2010;214:117–125.
65. di Gennaro A, Spadaro O, Baratta MG, et al. Functional analysis of the murine Pax8 promoter reveals autoregulation and the presence of a novel thyroid-specific DNA-binding activity. *Thyroid*. 2013;23:488–496.
66. Antonica F, Kasprzyk DF, Opitz R, et al. Generation of functional thyroid from embryonic stem cells. *Nature*. 2012;491:66–71.
67. Stoupa A, Kariyawasam D, Muzza M, et al. New genetics in congenital hypothyroidism. *Endocrine*. 2021;71:696–705.
68. Mio C, Grani G, Durante C, et al. Molecular defects in thyroid dysgenesis. *Clin Genet*. 2020;97:222–231.
69. Devriendt K, Vanhole C, Matthijs G, et al. Deletion of thyroid transcription factor-1 gene in an infant with neonatal thyroid dysfunction and respiratory failure. *N Engl J Med*. 1998;338:1317–1318.
70. Iwatani N, Mabe H, Devriendt K, et al. Deletion of NKX2.1 gene encoding thyroid transcription factor-1 in two siblings with hypothyroidism and respiratory failure. *J Pediatr*. 2000;137:272–276.
71. Krude H, Schutz B, Biebermann H, et al. Choreoathetosis, hypothyroidism, and pulmonary alterations due to human NKX2.1 haploinsufficiency. *J Clin Invest*. 2002;109:475–480.
72. Moeller LC, Kimura S, Kusakabe T, et al. Hypothyroidism in thyroid transcription factor 1 haploinsufficiency is caused by reduced expression of the thyroid-stimulating hormone receptor. *Mol Endocrinol*. 2003;17:2295–2302.
73. Carré A, Szinnai G, Castanet M, et al. Five new TTF1/NKX2.1 mutations in brain-lung-thyroid syndrome: rescue by PAX8 synergism in one case. *Hum Mol Genet*. 2009;18:2266–2276.
74. Macchia PE, Lapi P, Krude H, et al. PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat Genet*. 1998;19:83–86.
75. Nettore IC, Cacace V, De Fusco C, et al. The molecular causes of thyroid dysgenesis: a systematic review. *J Endocrinol Invest*. 2013;36:654–664.
76. Iwahashi-Odano M, Nagasaki K, Fukami M, et al. Congenital hypothyroidism due to truncating PAX8 mutations: a case series and molecular function studies. *J Clin Endocrinol Metab*. 2020;105:e4055–e4065.
77. Carvalho A, Hermanns P, Rodrigues AL, et al. A new PAX8 mutation causing congenital hypothyroidism in three generations of a family is associated with abnormalities in the urogenital tract. *Thyroid*. 2013;23:1074–1078.
78. Smol T, Ribero-Karrouz W, Edery P, et al. Mayer-Rokitansky-Kunster-Hauser syndrome due to 2q12.1q14.1 deletion: PAX8 the causing gene? *Eur J Med Genet*. 2020;63:103812.
79. Meeus L, Gilbert B, Rydlewski C, et al. Characterization of a novel loss of function mutation of PAX8 in a familial case of congenital hypothyroidism with in-place, normal-sized thyroid. *J Clin Endocrinol Metab*. 2004;89:4285–4291.
80. Narumi S, Yoshida A, Muroya K, et al. PAX8 mutation disturbing thyroid follicular growth: a case report. *J Clin Endocrinol Metab*. 2011;96:E2039–E2044.
81. Clifton-Bligh RJ, Wentworth JM, Heinz P, et al. Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate, and choanal atresia. *Nat Genet*. 1998;19:399–401.
82. Bamforth JS, Hughes IA, Lazarus JH, et al. Congenital hypothyroidism, spiky hair, and cleft palate. *J Med Genet*. 1989;26:49–60.
83. Carré A, Hamza RT, Kariyawasam D, et al. First FOXE1 mutation in Bamforth syndrome causing thyroid-gene overactivation. *Thyroid*. 2014;24:649–654.
84. Meng T, Shi JY, Wu M, et al. Overexpression of mouse TTF-2 gene causes cleft palate. *J Cell Mol Med*. 2012;16:2362–2368.
85. Lin RY, Kubo A, Keller GM, et al. Committing embryonic stem cells to differentiate into thyrocyte-like cells *in vitro*. *Endocrinology*. 2003;144:2644–2649.
86. Arufe MC, Lu M, Kubo A, et al. Directed differentiation of mouse embryonic stem cells into thyroid follicular cells. *Endocrinology*. 2006;147:3007–3015.
87. Longmire TA, Ikonomou L, Hawkins F, et al. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell*. 2012;10:398–411.
88. Kurmann AA, Serra M, Hawkins F, et al. Regeneration of thyroid function by transplantation of differentiated pluripotent stem cells. *Cell Stem Cell*. 2015;17:527–542.
89. Wendl T, Lun K, Mione M, et al. Pax2.1 is required for the development of thyroid follicles in zebrafish. *Development*. 2002;129:3751–3760.
90. Ericson LE, Fredriksson G. Phylogeny and ontogeny of the thyroid gland. In: Freer M, ed. *The Thyroid Gland*. New York: Raven Press; 1990:1–35.
91. Kluge B, Renault N, Rohr KB. Anatomical and molecular reinvestigation of lamprey endostyle development provides new insight into thyroid gland evolution. *Dev Genes Evol*. 2005;215:32–40.
92. Dohrn A. Thyroidea bei Petromyzon, Amphioxus und den Tunicaten. *Mitt zool Stat Neapel*. 1885;6:49–92.
93. Hiruta J, Mazet F, Ogasawara M. Restricted expression of NADPH oxidase/peroxidase gene (Duox) in zone VII of the ascidian endostyle. *Cell Tissue Res*. 2006;326:835–841.
94. Ogasawara M, Di Lauro R, Satoh N. Ascidian homologs of mammalian thyroid peroxidase genes are expressed in the thyroid-equivalent region of the endostyle. *J Exp Zool*. 1999;285:58–69.
95. Ogasawara M, Satou Y. Expression of FoxE and FoxQ genes in the endostyle of Ciona intestinalis. *Dev Genes Evol*. 2003;213:416–419.
96. Ristoratore F, Spagnuolo A, Aniello F, et al. Expression and functional analysis of Cittf1, an ascidian NK-2 class gene, suggest its role in endoderm development. *Development*. 1999;i126:149–159.
97. Ogasawara M. Overlapping expression of amphioxus homologs of the thyroid transcription factor-1 gene and thyroid peroxidase gene in the endostyle: insight into evolution of the thyroid gland. *Dev Genes Evol*. 2000;210:231–242.
98. Yu JK, Holland LZ, Jamrich M, et al. AmphiFoxE4, an amphioxus winged helix/forkhead gene encoding a protein closely related to vertebrate thyroid transcription factor-2: expression during pharyngeal development. *Evol Dev*. 2002;4:9–15.
99. Ogasawara M, Shigetani Y, Suzuki S, et al. Expression of thyroid transcription factor-1 (TTF-1) gene in the ventral forebrain and endostyle of the agnathan vertebrate, Lampetra japonica. *Genesis*. 2001;30:51–58.

65

Thyroid-Stimulating Hormone and TSH Receptor

Susanne Neumann, Gunnar Kleinau, and Marvin C. Gershengorn

OUTLINE

Introduction, 1099
Regulation of Thyroid-Stimulating Hormone Biosynthesis, 1099
  Hypothalamic Regulation of Thyroid-Stimulating Hormone Subunit Biosynthesis, 1099
  Peripheral Regulation of Thyroid-Stimulating Hormone Subunit Biosynthesis, 1101
  Regulation of Thyroid-Stimulating Hormone Glycosylation, 1101
Regulation of Thyroid-Stimulating Hormone Secretion, 1101
  Hypothalamic Regulation of Thyroid-Stimulating Hormone Secretion, 1101
  Peripheral Regulation of Thyroid-Stimulating Hormone Secretion, 1102
Thyrostimulin, 1103
Thyroid-Stimulating Hormone Receptor, 1103
  Protein Structure, 1103
  Intracellular Primary and Secondary Signal Transduction Mechanisms, 1103
  Thyroid-Stimulating Hormone Receptor Homodimerization, 1104
Thyroidal Actions of Thyroid-Stimulating Hormone, 1104
  Regulation of Thyroid-Specific Gene Expression, 1104
Thyroid-Stimulating Hormone Receptor/Insulin-Like Growth Factor 1 Receptor Crosstalk, 1106
  Regulation of Thyroid Hormone Synthesis, 1106
  Regulation of Thyroid Hormone Secretion, 1107
Extrathyroidal Actions Mediated By Thyroid-Stimulating Hormone Receptor, 1107

INTRODUCTION

Thyroid-stimulating hormone (TSH, also known as thyrotropin) is the pituitary hormone that is the major regulator of the thyroid gland, and because of this, is a key player in the hypothalamic-pituitary-thyroid axis that regulates thyroid metabolism overall (Fig. 65.1). TSH acts to control cell function by binding to specific cell-surface receptors. Its major physiologic function is mediated by TSH receptors (TSHRs) on the thyroid gland, where it stimulates the production of the thyroid hormones L-thyroxine (T_4) and L-triiodothyronine (T_3). They act on multiple organs and tissues to modulate many metabolic processes, as well as to cause negative feedback inhibition of TSH synthesis and secretion from the pituitary.

TSH is a glycoprotein produced by the thyrotrope cells of the anterior pituitary gland. TSH, luteinizing hormone (LH), and follicle-stimulating hormone (FSH), as well as chorionic gonadotropin (CG) in the placenta, are heterodimers consisting of two noncovalently linked subunits, α (glycoprotein hormone α, GPH α) and β (TSH β) (Fig. 65.2). The α-subunit is common to all four glycoproteins hormones (GPHs), whereas the β-subunit is unique to each and confers specificity of action. Glycoprotein hormones are cystine knot proteins.

Each TSH subunit is encoded by a separate gene and is coordinately transcribed and regulated mainly by hypothalamic thyrotropin-releasing hormone (TRH) stimulation and thyroid hormone feedback inhibition. Production of bioactive TSH involves a process of cotranslational glycosylation and folding that enables dimerization between the nascent α- and β-subunits. TSH is stored in secretory granules and is released into the circulation.

In recent years, it has become evident that TSH receptors (TSHRs) are expressed on a number of extra-thyroidal tissues also. Most prominently, TSHRs that are expressed on cells in the retroorbital space, with a so far unknown physiological relevance, but are involved in the pathogenesis of thyroid eye disease (TED, also termed Graves ophthalmopathy or Graves orbitopathy; see Chapter 72).

REGULATION OF THYROID-STIMULATING HORMONE BIOSYNTHESIS

TSH biosynthesis is regulated by coordinated signals from the central nervous system and feedback from the peripheral circulation. The most important positive input for TSH biosynthesis is hypothalamic TRH, and the most powerful negative regulator is circulating thyroid hormones. However, additional hypothalamic factors and circulating hormones provide important modifying effects. Most of these factors have independent effects on the biosynthesis of the two subunits of TSH (Fig. 65.1).

Hypothalamic Regulation of Thyroid-Stimulating Hormone Subunit Biosynthesis

TRH is a tripeptide secreted from the paraventricular nucleus (PVN) of the hypothalamus and transported to the pituitary via the hypothalamic-hypophyseal portal system, and is a major activator of TSH production, with a marked increase on the transcription of both TSH β- and α-subunit mRNAs[1] (Fig. 65.1). Although TRH may not be required for normal thyrotrope development during ontogeny, TRH appears required for the postnatal maintenance of TSH activation.[2]

TRH binding to a specific membrane receptor, TRHR, on thyrotropes initiates a cascade of intracellular events. TRHR is a member of the class A family of the G protein-coupled receptor (GPCR)

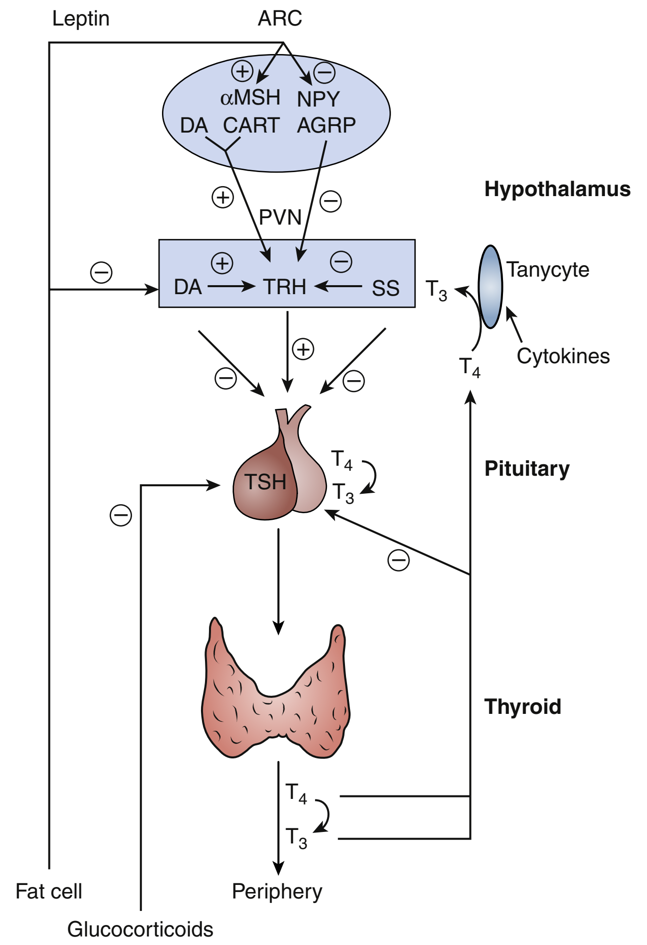


Fig. 65.1 **Neuroendocrine and peripheral control of thyroid-stimulating hormone (TSH) secretion.** Positive and negative stimuli are indicated at various levels within the hypothalamic-pituitary-thyroid axis. *AGRP*, Agouti-related protein; *MSH*, melanocyte-stimulating hormone; *ARC*, arcuate nucleus; *CART*, cocaine and amphetamine regulated transcript; *D2*, deiodinase type II; *DA*, dopamine; *NPY*, neuropeptide Y; *PVN*, paraventricular nucleus; *SS*, somatostatin; *T_3*, triiodothyronine; *T_4*, thyroxine; *TRH*, thyrotropin-releasing hormone. (Adapted from Roelfsema F, Veldhuis JD. Thyrotropin secretion patterns in health and disease. *Endocr Rev.* 2013;34:619–657.)

superfamily. Intracellular signal transduction is mediated primarily by coupling to $G_{q/11}$ proteins. TRH binding results in the activation of phosphoinositide-specific phospholipase C, which stimulates phosphatidylinositol 4,5-P2 hydrolysis to form inositol 1,4,5- trisphosphate (InsP3) and 1,2-diacylglycerol (1,2-DG). These second messengers stimulate increases in intracellular calcium and activation of protein kinase C.[3,4,5] TRHR activation also stimulates calcium/calmodulin-dependent protein kinase[6,7] and mitogen-activated protein kinase.[8] The TRHR has also been shown to couple to G_{i2}, G_{i3}, and to a G_s-like protein that does not activate adenylyl cyclase,[9] and to increase cGMP.[10,11] Thus, TRHR activation signals via several intracellular transduction pathways.

TRH stimulates GPH α-subunit biosynthesis through a distinct further mechanism. A cyclic adenosine monophosphate (cAMP) response element–binding protein that binds to the human α-subunit promoter is important for TRH regulation.[12] This mechanism involves two transcription factors that when stimulated by TRH cooperate to activate α-subunit promoter activity resulting from protein–protein interactions.[13] One factor binds to the α-subunit promoter directly, but the

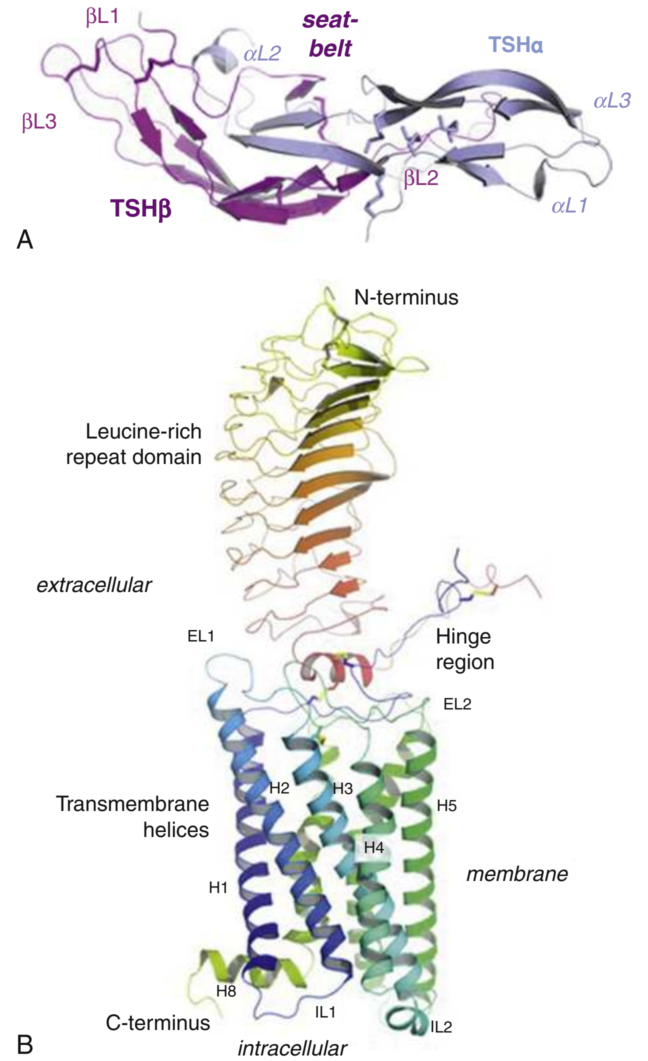


Fig. 65.2 **Structural homology models of human thyroid-stimulating hormone (TSH) and TSH receptor (TSHR). A,** The heterodimeric TSH is constituted by two noncovalently bound subunits, termed alpha-TSH (TSH α) and beta-TSH (TSH β). The human TSH α and TSH β-subunits comprise 116 and 138 amino acids, respectively, and together they form one of the largest endogenous ligands for G protein–coupled receptors known so far. They are intrinsically subdivided into several loop and β-sheet regions according to predicted and observed secondary structures of determined homologous glycoprotein hormones such as follicle-stimulating hormone or chorionic gonadotrophin. The "seat-belt" region is directly involved in TSH binding to TSHR, glycoprotein hormone receptor selectivity, and activity regulation. **B,** The structure of the TSHR comprises 764 amino acids and is subdivided into distinct functional and self-folding units, called domains. In the extracellular domain, the leucine-rich repeat domain (LRRD), with approximately 288 amino acids, is mainly involved in TSH binding. The hinge region with approximately 130 amino acids connects the LRRD with the transmembrane-spanning domain, constituted by seven helices (H1–7) and their interrelated loops, extracellular loops (ELs) and intracellular loops (ILs) 1–3. The hinge region is not a domain by itself because of missing self-folding properties but plays an essential role in the arrangement and function of the entire receptor, in both the ligand-free and ligand-occupied receptor conformations. Each domain or interconnecting structural part plays crucial roles for receptor function, like folding, transport to the cell surface, intermolecular activity (e.g., ligand or G protein binding), or intramolecular functions like signal transduction by structural rearrangements initiated by ligand contacts.

other does not bind DNA, so it must be recruited to the promoter via protein–protein interactions.

Dopamine, acting via DA2 dopamine receptors, inhibits GPH α- and TSH β-subunit gene transcription by, at least in part, decreasing intracellular cAMP.[1] GPH α-subunit gene expression in thyrotropes is inhibited by dopamine, in coordination with the expression of the TSH β-subunit gene.

Peripheral Regulation of Thyroid-Stimulating Hormone Subunit Biosynthesis

Thyroid hormone acts predominantly, if not exclusively, through a classical thyroid hormone receptor (THR) that binds to the upstream region of the β-subunit gene. The thyroid hormone enters the cell via at least four families of membrane transporter proteins and diffuses from the plasma membrane to the nucleus.[14] T_4 serves as a minimally active prohormone that is a substrate for a family of tissue iodothyronine deiodinases, termed DIO1, DIO2, and DIO3. These homodimeric, membrane-bound, selenoproteins can activate (DIO1, DIO2) or inactivate (DIO3) T_4.[15,16] DIO2 is the major T_4-activating deiodinase with a high substrate affinity. It is present on the endoplasmic reticulum close to the nucleus and produces 3,5,3'-triiodothyronine (T_3) by the removal of an iodine residue from the outer ring of thyroxine. Serum TSH levels in normal mice are suppressed by administration of either T_4 or T_3, although only T_3 was effective in the mouse with a targeted disruption of the *DIO2* gene,[17] or in animals treated with the DIO2 inhibitor amiodarone. The observed phenotype of pituitary resistance to T_4 demonstrated the critical importance of DIO2 in controlling negative thyroid hormone regulation of TSH.

TSH β- and α-subunit gene transcription rates are markedly inhibited by treatment with T_3. Studies using mouse TtT-97 thyrotropic tumors have demonstrated that T_3 suppression of TSH β- and α-subunit gene transcription rates, as measured by nuclear run-on assays, is evident by 30 minutes after treatment and is maximal by 4 hours.[18] Other studies using mouse and rat pituitaries along with mouse thyrotropic tumors have demonstrated that steady-state mRNA levels of TSH β- and α-subunit are dramatically decreased by T_3.[19] The mechanism of action of T_3 involves interaction with nuclear THRs that act mainly at the transcriptional level.[20] The T_3 inhibitory effect on the TSH β-subunit gene requires ligand occupied THR, mainly the THRβ1 or THRβ2 isoform, because patients with thyroid hormone resistance and inappropriate secretion of TSH generally have abnormalities in the THRβ gene,[21] although few mutations in the THRα gene have been described (see later).

In general, the mechanism(s) involved in gene upregulation by T_3 are known. THRs bind to cis-acting DNA response elements in the absence of ligand, interacting with a family of nuclear receptor corepressor molecules that recruit histone deacetylases and locally modify chromatin structure to result in repression of the target gene.[22] In the presence of T_3, the corepressor complexes rapidly dissociate and are replaced by coactivator complexes that bind to THRs, increasing histone methylation and acetylation locally on the chromosomal DNA, which unwinds the chromatin.[23]

In contrast, the molecular mechanisms involved in negative T_3 regulation, including that of TSH subunit genes, have not been as well characterized. Liganded THRβ has been reported to recruit histone deacetylase 3 and to reduce histone H4 acetylation, which modifies histones and results in a fully repressed chromosomal state of the TSH β-subunit gene.[24] For the human TSH α-subunit gene, a corepressor complex containing TSH β-subunit was shown to bind the promoter and to determine the basal level of histone acetylation and transcription in the absence of T_3. In the presence of T_3, the complex was released from the THR.[25] Although thyrotrope cells contain all THRs (THRα1, THRβ1, and THRβ2), as well as a non–T_3-binding variant α2,[26] it is THRβ2 that is expressed predominantly in the pituitary[27] and is most critical for the regulation of TSH.[28] Binding studies using chromatin immunoprecipitation assays in TαT1.1 thyrotrope cells showed that THRβ, and not THRα, could interact with the TSH β-subunit promoter. Furthermore, depletion of only THRβ and not THRα abolished T_3-mediated downregulation of TSH β-subunit transcripts.[29] THR accessory proteins, such as retinoid X receptor (RXR), may also bind to DNA[30] and constitute heterodimeric complexes with THR that have different affinities for specific DNA sequences and different functional activities. A particular RXR isoform, RXRγ1, is uniquely expressed in thyrotropes and appears to mediate the inhibition by 9-*cis*-retinoic acid.[31]

Because thyroid hormones can act at the level of both the pituitary and the hypothalamus to regulate TSH, studies were performed to establish their relative contributions. Using genetic knockout mouse models where either TRH or THRβ genes alone, or both together, were eliminated showed a dominant role for TRH *in vivo* in regulating the hypothalamic-pituitary-thyroid axis. The hypothalamic TRH-producing cells are absolutely required for both TSH and thyroid hormone synthesis and control the set points in the hypothalamic-pituitary-thyroid axis.

Thyroid hormone inhibition of α-subunit gene transcription is observed in thyrotropes in coordination with that of the TSH β-subunit. T_3 inhibition may be mediated by different isoforms of the THR[32] in combination with corepressors.[33] It has been suggested that mutations of the T_3 response element of the human α-subunit promoter that eliminate THR binding do not abrogate the inhibitory effect of T_3, suggesting that protein–protein interactions may be more important than protein–DNA binding.[34]

Regulation of Thyroid-Stimulating Hormone Glycosylation

Glycosylation of both TSH subunits is a regulated process that is primarily modulated by TRH and thyroid hormone.[35] Primary hypothyroidism[36] and TRH administration[37] have been found to increase oligosaccharide addition, which results in an increased bioactivity of TSH.[38] The same was noted in patients with resistance to thyroid hormone.

TSH glycosylation patterns were also found to differ in several pathological states, such as central hypothyroidism,[39] TSH-producing pituitary adenomas, and euthyroid sick syndrome.[40] Also observed were changes in the sulfation and sialylation of the oligosaccharide residues, which modulate bioactivity.[41] Thyroid hormone was shown to increase TSH bioactivity via decreased sialylation.[42]

KEY POINTS

- Thyrotropin-releasing hormone (TRH) is the major activator of thyroid-stimulating hormone (TSH) production, and circulating thyroid hormones are negative regulators of TSH synthesis. Thyroid hormones can act at the level of both the pituitary and the hypothalamus to regulate TSH, but the hypothalamic TRH-producing cells are required for both TSH and thyroid hormone synthesis, and control the set points in the hypothalamic-pituitary-thyroid axis.

REGULATION OF THYROID-STIMULATING HORMONE SECRETION

Hypothalamic Regulation of Thyroid-Stimulating Hormone Secretion

TRH from the PVN of the hypothalamus directly affects pituitary TSH secretion *in vivo* (Fig. 65.1). Concentrations that exist in the pituitary portal blood also cause TSH secretion *in vitro*.[43] Immunoneutralization of TRH in animals leads to a decline in thyroid function,[44] and TRH knockout mice have a reduced postnatal TSH surge, followed by impaired baseline thyroid function with a poor TSH response to hypothyroidism. Lesions of the PVN decrease circulating TRH and TSH levels in normal or hypothyroid animals and cause hypothyroidism.[45] Similar results were found in humans with central hypothyroidism.[39] Although baseline levels of TSH are reduced in animals with lesions of the PVN, TSH levels still show appropriate responses to changes in circulating thyroid hormone levels. Thus, TRH may determine the set point of feedback control by thyroid hormones.

Acute intravenous administration of TRH to human subjects causes a dose-related release of TSH from the pituitary. This occurs within 5 minutes and is maximal at 20 to 30 minutes. Serum TSH levels return to basal levels by 2 hours. More prolonged (2- to 4-hour) infusions of TRH lead to biphasic increases in serum TSH levels in humans and animals.[46] The early phase may reflect release of stored TSH, whereas the later phase may reflect release of newly synthesized TSH. An interpretation of TSH responses to even more prolonged TRH infusions is complicated by the increase in serum T_3, which feeds back to further suppress TSH release.[47] Continuous TRH administration *in vitro* also causes desensitization of TSH responses, which may further explain decreased TSH levels with long-term TRH exposure.[48]

Somatostatin (SS) in humans and animals inhibits basal and TRH-stimulated TSH secretion *in vivo* and *in vitro* at concentrations that exist in the pituitary portal blood.[49] In the hypothalamus, the highest concentrations of SS occur in the anterior paraventricular region.[50] From this region, axonal processes of SS-containing neurons project to the median eminence. Animals that have undergone sectioning of these fibers experience depleted SS content of the median eminence and increased serum TSH levels.[51] Similarly, immunoneutralization of SS in animals increases basal TSH levels and TSH responses to TRH.[52] In humans, SS infusions suppress TSH pulse amplitude and abolish the nocturnal TSH surge.[53] Thus, TSH secretion is regulated through a simultaneous dual control system of TRH stimulation and SS inhibition from the hypothalamus.

SS binds to specific, high-affinity receptors in the anterior pituitary gland. Five subtypes of the SS receptor (SST1–5) have been identified,[54,55] and SST2 and 5 have been localized to thyrotropes.[56] Binding of SS to its receptor inhibits adenylate cyclase via the inhibitory subunit of the guanine nucleotide regulatory protein (G_i), which inhibits cAMP production, leading to lowering of protein kinase A activity,[54] and decreases TSH secretion. SS may also exert some effects by cAMP-independent actions on intracellular calcium levels. Hypothyroidism reduces the efficacy of SS in decreasing TSH secretion *in vitro*, which is reversed by thyroid hormone administration.[57] Further studies in mouse thyrotropic tumors indicate that the synthesis of both SST2 and SST5 are markedly downregulated in hypothyroidism and are induced by thyroid hormone.[56] Although short-term infusions of SS lead to pronounced suppression of TSH secretion in humans, long-term treatment with SS or its analogs does not cause hypothyroidism.[58] Growth hormone (GH) deficiency is associated with increased TSH responses to TRH, whereas GH administration decreases basal and TRH-stimulated TSH secretion,[59] possibly as a result of GH stimulation of hypothalamic SS release.

Dopamine also inhibits basal and TRH-stimulated TSH secretion *in vivo* and *in vitro* at concentrations that exist in pituitary portal blood.[60] Exogenous dopamine antagonists, including those that do not penetrate the blood–brain barrier, increase TSH levels.[61] In humans, dopamine infusions rapidly suppress TSH pulse amplitude, do not affect TSH pulse frequency, and abolish the nocturnal TSH surge,[53] whereas administration of a dopamine antagonist results in the opposite effect.[62] Dopamine also has direct effects on hypothalamic hormone secretion that may indirectly impact TSH secretion. For example, dopamine and dopamine-agonist drugs stimulate both TRH and SS release from the rat hypothalamus.[63]

In the hypothalamus, dopamine is secreted by neurons in the arcuate nucleus.[64] From the arcuate nucleus, neuronal processes project to the median eminence. Dopamine acts by binding to type 2 dopamine receptors (DA_2) on thyrotrope cells.[65] This leads to inhibition of adenylate cyclase, which decreases the synthesis and secretion of TSH. TSH may downregulate its own release through the activation of DA_2 receptors on thyrotrope cells.[66] The inhibitory effects of dopamine on TSH secretion vary according to sex steroids, body mass, and thyroid status. Dopamine-antagonist drugs cause greater increases in serum TSH levels in women than in men. Studies show that obesity is associated with enhanced TSH secretion, which may be mediated via blunted central dopaminergic tone.[67] Dopamine inhibition of TSH release is greater in patients with mild hypothyroidism than in normal subjects, although subjects with severe hypothyroidism may be less responsive.[68] Although short-term infusions of dopamine lead to pronounced suppression of TSH secretion, long-term treatment with dopamine agonists does not cause hypothyroidism.

Acute opioid administration in humans may slightly stimulate TSH levels, whereas acute administration of naloxone, an opioid receptor antagonist, has little effect.[69] In contrast to these acute studies, when naloxone is administered throughout the day, 24-hour TSH secretion decreases, primarily because of a decrease in nocturnal TSH pulse amplitude.[70] TSH responses to TRH are also decreased. Serum T_3 levels are decreased as well, suggesting that the magnitude of TSH suppression is sufficient to affect thyroid gland function. These findings suggest that endogenous opioids may play a role in tonic stimulation of TSH secretion.

Peripheral Regulation of Thyroid-Stimulating Hormone Secretion

Thyroid hormones directly block pituitary secretion of TSH.[71] Acute administration of T_3 suppresses TSH levels within hours, whereas chronic administration leads to further suppression. Slight changes in serum thyroid hormone levels within the normal range alter basal and TRH-stimulated TSH levels, confirming the sensitivity of the pituitary gland to thyroid hormone feedback. Thyroid hormones alter tonic TSH secretion and TSH pulse amplitude without affecting pulse frequency, based on findings that subjects with primary hypothyroidism have a near-normal number of TSH pulses, and T_4 replacement leads to a decrease in TSH pulse amplitude without changing pulse frequency.[72]

In addition to direct effects on TSH secretion, thyroid hormones include other actions that impact TSH secretion. At the pituitary level, thyroid hormones decrease the number of TRH receptors[73] and stimulate the activity of the pituitary TRH-degrading enzyme,[74] which act in concert to decrease TRH stimulation of TSH secretion. At the hypothalamic level, PVN TRH mRNA is increased in hypothyroidism and reduced by T_3 and T_4.[75] In contrast, hypothalamic SS content is decreased in hypothyroid rats and is restored by T_3 treatment. Finally, T_3 directly stimulates SS release from hypothalamic tissue[76] and increases the number of SS receptors on thyrotrope cells. These combined effects of thyroid hormones on TRH and SS decrease TRH release from the hypothalamus and indirectly decrease TSH secretion.

Glucocorticoids at pharmacologic doses or high endogenous cortisol levels (Cushing syndrome) suppress basal TSH levels, blunt TSH responses to TRH, and diminish the nocturnal TSH surge in humans and animals.[77] In healthy subjects, infusions of hydrocortisone that increase serum cortisol to levels seen in mild to moderate stress suppress 24-hour TSH secretion.[78,79] Glucocorticoid-induced changes in TSH levels result from decreased TSH pulse amplitude without altering pulse frequency, with more profound suppression of nocturnal TSH secretion and abolition of the TSH surge. Physiologic glucocorticoid levels also affect TSH secretion.[80] Untreated patients with adrenocortical insufficiency can have elevated serum TSH levels that resolve with steroid replacement. Complementary studies of metyrapone (an inhibitor of cortisol synthesis) administration to healthy subjects confirm that endogenous cortisol levels suppress TSH secretion, and physiologic hydrocortisone replacement in patients with adrenal insufficiency decrease daytime TSH levels back to those seen in healthy subjects.

Glucocorticoid suppression of TSH levels may occur directly at the pituitary gland. Animal studies suggest that glucocorticoids exert direct effects on thyrotropes to impair TSH secretion, although these appear to be highly dependent on dose and time-course of administration.[81,82] It does not appear that glucocorticoids directly affect TSH gene transcription. In humans, TSH pulse frequency is maintained during glucocorticoid administration, whereas TSH pulse amplitude is reduced, and TSH responses to exogenous TRH are attenuated, suggesting a direct effect on TSH secretion.

Leptin is primarily a product of adipocytes, although it is also present in thyrotropes. It regulates food intake and energy expenditure, decreasing acutely with fasting in animals and humans.[83,84] Exogenous leptin administration to fed rats raises serum TSH levels, probably by increasing TRH gene expression and TRH release. Similarly, leptin administration to fasting rats or humans reverses fasting-induced decrements in TSH levels, also by increasing *TRH* gene expression and TRH release. This suggests that fasting-related reductions in leptin levels play a role in suppressing TSH secretion. In healthy subjects, leptin and TSH have coordinate pulsatility and circadian rhythms in plasma, with a nadir in the late morning and a peak in the early morning. L-T_4 treatment of patients with subclinical hypothyroidism reduced both serum TSH and leptin levels.[85] In thyroid cancer patients on suppressive doses of levothyroxine, recombinant TSH injections increase serum leptin levels.[86]

Cytokines are circulating mediators of the inflammatory response that are produced by many cells and have systemic effects on the hypothalamic-pituitary-thyroid axis. Administration of tumor necrosis factor (TNF) or interleukin-6 decreases serum TSH levels in healthy human subjects, and TNF and interleukin-1 (IL-1) decrease TSH levels in animals.[87,88,89] Administration of these cytokines recapitulates the alterations in thyroid hormone and TSH levels seen in acute nonthyroidal illness. In rats, TNF reduces hypothalamic TRH content and pituitary *TSH* gene transcription. IL-1 stimulates type II 5'-deiodinase activity in rat brain, which may decrease TSH secretion by increasing intrapituitary T_3 levels.

KEY POINTS

- Thyroid-stimulating hormone (TSH) secretion is regulated through a simultaneous dual control system of thyrotropin-releasing hormone (TRH) stimulation and somatostatin (SS) inhibition from the hypothalamus. Dopamine also has direct effects on hypothalamic hormone secretion that may indirectly impact TSH secretion. Thyroid hormones directly affect pituitary secretion of TSH by combined effects on TRH release from the hypothalamus and SS levels and SS receptor expression on thyrotrope cells.

THYROSTIMULIN

Thyrostimulin is a second glycoprotein that binds to TSHR and can activate signaling by TSHR.[90] Indeed, thyrostimulin binds to TSHR with higher affinity than TSH. Like TSH, thyrostimulin is a heterodimer comprised of an α- and β-subunit, both of which are distinct from those of TSH. Thyrostimulin has been shown to affect several tissues in animal models.[91] However, no effect of thyrostimulin has been shown in humans, and there has not been any report of measurable levels of thyrostimulin in human blood. Thus, the role of thyrostimulin in humans is not known.

THYROID-STIMULATING HORMONE RECEPTOR

TSHR[9,92,93] is a member of the class A subfamily of the ubiquitous family of cell-surface receptors known as GPCRs, or seven transmembrane-spanning domain receptors.[94,95] TSHR is highly conserved across species. In this chapter, TSHR will usually refer to the human receptor.

Protein Structure

Like all GPCRs, TSHR is comprised of an extracellular portion (or ectodomain) and a transmembrane domain comprised of seven alpha-helices, connected by extra- and intracellular loops, respectively (Fig. 65.2). TSHR is most homologous to the receptors for the two other GPHs (receptors for LH/CG and FSH), where the highest homology is within the transmembrane domains.[92]

Another feature that is shared by the GPH receptors is the presence of a large ectodomain to which their cognate ligands bind. The binding sites of TSH (and TSHR-stimulating antibodies found in patient with Graves disease), like binding of GPHs to their cognate receptors, are primarily, if not exclusively, to the extracellular part of TSHR (Fig. 65.2). But in contrast to FSHR and LH/CGR, as well as other GPCRs, which are present on the surface of cells as single polypeptide chains, TSHR may be cleaved within its ectodomain when translocated from the endoplasmic reticulum to the cell surface to generate a receptor composed of two subunits, A/α and B/β, that are held together by disulfide bridges.[96] Thereby, the cleaved sequence of approximately 50 amino acid residues (C-peptide) is excised and lost (Fig. 65.3). It is likely that TSHR is present on the cell surface in both cleaved and uncleaved forms. The biological role of receptor cleavage is still controversial.

Also, like many GPCRs,[97,98] TSHRs may be present on the cell surface as monomers and homodimers.[99,100] The presence of homodimers appears to be important in allowing TSHR to initiate signaling via multiple transduction pathways (see later).

Intracellular Primary and Secondary Signal Transduction Mechanisms

TSH and thyrostimulin (and TSHR-stimulating antibodies found in Graves disease patients; see Chapter 71) activate TSHR by binding primarily, if not exclusively, to its large extracellular N-terminal part. Initiation of TSHR signaling involves coupling to and activation of members of all four G-protein families (G_s, $G_{q/11}$, $G_{i/o}$, $G_{12/13}$)[101] and β-arrestin 1.[102,103] The best-characterized signaling cascades initiated by TSHR are activation of G_s and G_q. G_s activates adenylyl cyclase to increase production of the second messenger cAMP, which in turn activates proteinase kinase A. G_q activates phospholipase C to generate inositol 1,4,5-trisphosphate, which increases intracellular calcium that in turn activates calcium/calmodulin-dependent protein, and 1,2-diacylglycerol, which activates protein kinase C. Activation of $G_{i/o}$ leads to inhibition of cAMP production and activation of $G_{12/13}$ leads to activation of a series of signaling protein kinases. β-arrestin 1–mediated signaling leads to activation of multiple protein kinases like ERK1/2, AKT1, and p38α.[102] Downstream of these initial activations/inhibitions, multiple protein kinase cascades mediate the responses at every level of cellular function. For example, animal cell studies have

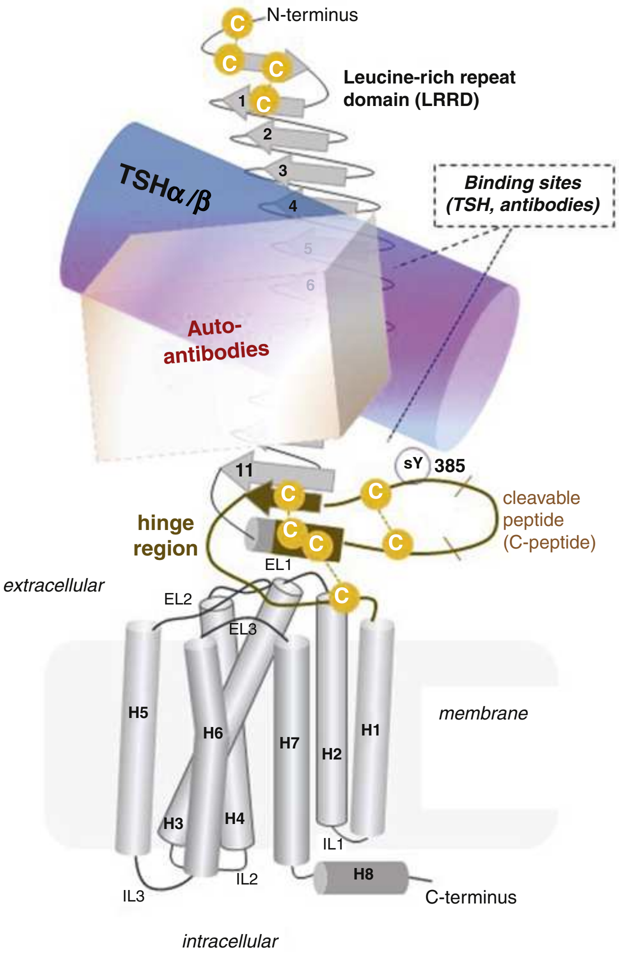


Fig. 65.3 Structural scheme of thyroid-stimulating hormone (TSH) and autoantibodies bound to the TSH receptor (TSHR). The TSHR exposes an extracellular N-terminus of approximately 400 amino acids. TSH interacts with two binding sites, first in the concave site of the leucine-rich repeat domain (LRRD) constituted by 11 leucine-rich repeats, and secondly in the hinge region, e.g., at a sulfated tyrosine (sY385). The TSH-induced signal is conveyed via structural rearrangements between specific interacting components of the hinge region, the LRRD and extracellular loops (ELs), which are termed in their entirety "intramolecular agonist." The extracellular part is characterized by several disulfide bridges participating in the stabilization of the receptor fold. Pathogenic autoantibodies also bind to the extracellular domain of the TSHR and cause Graves hyperthyroidism and thyroid eye disease by persistent, unregulated stimulation of the TSHR. The transmembrane spanning region is constituted by seven helices (H1–7), connected by ELs and intracelluar loops (ILs). The eighth helix is located parallel to the membrane and anchors the receptor to the membrane. The α-subunit of the heterotrimeric G protein binds between the intracellular loops and adjacent transitions to the helices after agonist interaction with the receptor or under basally/constitutively active conditions.

led to the conclusion that increased intracellular cAMP is involved in iodide uptake, thyroid hormone secretion, and cell proliferation. Phospholipase C–dependent pathways are required for iodine organification and thyroid hormone secretion. In addition, the $G_{q/11}$ pathway seems to have an essential role in the adaptive growth of the thyroid gland. It deserves reemphasis that most of the data and conclusions drawn regarding the molecular mechanisms of TSHR signal transduction come from studies of animal cells and engineered model cell systems in culture. Nevertheless, the many steps in the signal transduction pathways that mediate the entirety of TSHR signaling leading to the multiple biologic responses have not been completely elucidated. Moreover, although these studies have provided important information, it has become clear from studies of human thyrocytes in primary culture[104] that the details of the mechanisms elucidated in animal cells do not precisely model signal transduction in human cells.

Thyroid-Stimulating Hormone Receptor Homodimerization

A number of GPCRs have been found to exist on the surface of cells as monomers, homodimers, and heterodimers.[97,98] TSHR is capable of forming homodimers and heterodimers.[99,100] However, to date no physiological function has been found for TSHR heterodimers, which have been shown only in artificial systems. TSHR homodimers, by contrast, have been found in normal human thyrocytes and have effects on signaling.[103] Binding of TSH to TSHR homodimers exhibits the phenomenon of negative cooperativity; that is, after binding of one molecule of TSH to one of the two promoters that comprise the homodimer, binding of a second molecule of TSH to the second protomer exhibits decreased binding affinity.[99] This is a mechanism that allows TSHR to be activated over a large range of TSH concentrations and to signal via different pathways at different TSH doses. For example, low of doses of TSH allow only a single TSH molecule to bind to a monomer or to a homodimer to activate G_s, whereas high doses of TSH allow two molecules of TSH to bind to a homodimer to activate $G_{q/11}$ or Gi/o.[105-108] Laurent et al. were the first to demonstrate that increasing doses of TSH generated a biphasic dose response for cAMP production in normal human thyroid tissue slices *in vitro*. In the cAMP signaling system, coupling to G_s at low TSH doses and to $G_{i/o}$ at high TSH doses leads to a biphasic response of cAMP production to TSH. This biphasic stimulation of cAMP production at low TSH doses and partial inhibition of stimulated production at high TSH doses[107] has been termed an inverted U-shaped dose response curve (IUDRC) (Fig. 65.4). This contrasts the monophasic dose response curve usually reported in cell systems stimulated by TSH, in which increasing doses of TSH lead to progressively increased cAMP production that plateaus at high TSH doses. It is now appreciated that a wide variety of receptor activators may generate IUDRCs.[109,110] IUDRCs have been shown to be exhibited by several other GPCRs[111,112] and in a wide range of pharmaceutical products and approved drug responses.[113] For TSH, the alternative cell-autonomous mechanism of an IUDRC in the cAMP signaling pathway may represent a mechanism to prevent overstimulation of cAMP-mediated responses in thyrocytes.

KEY POINTS

- The thyroid-stimulating hormone receptor (TSHR) is a G protein–coupled receptor with a large extracellular domain to which thyroid-stimulating hormone (TSH), thyrostimulin, and pathogenic TSHR antibodies bind. TSHR can activate members of all four G-protein families and β-arrestin 1, depending on the cell context. Homodimerization of TSHR plays a pivotal role in signaling, including the cell-autonomous biphasic response to TSH for cyclic adenosine monophsphate production.

THYROIDAL ACTIONS OF THYROID-STIMULATING HORMONE

Regulation of Thyroid-Specific Gene Expression

Much of the information regarding regulation of transcription of thyroid-specific genes was initially generated in studies of thyroid development in mice.[114] Although there have been studies of

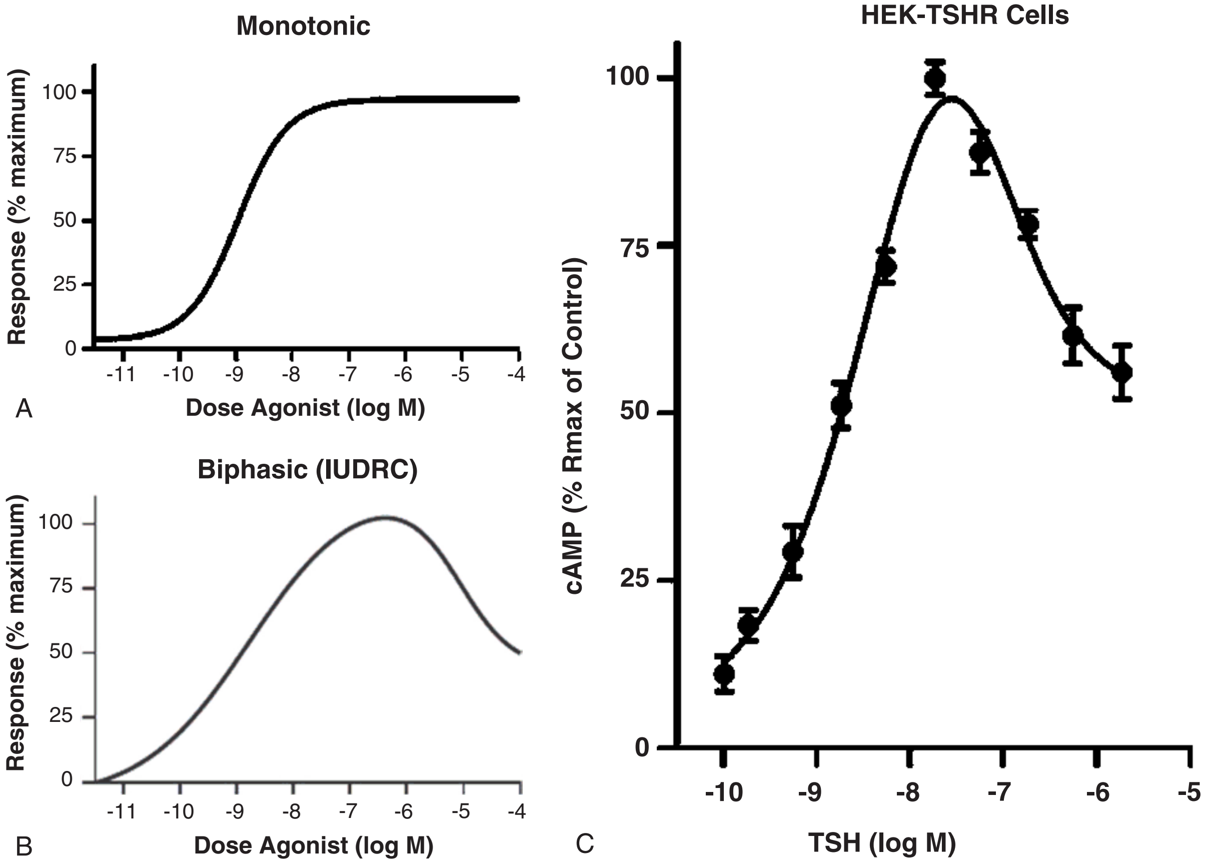


Fig. 65.4 **Biphasic regulation of thyroid-stimulating hormone (TSH)-mediated cyclic adenosine monophosphate (cAMP) production. A,** Typical monotonic dose response curve. **B,** Biphasic (inverted U-shaped dose response curve, IUDRC). **C,** Biphasic regulation of cAMP production by TSH in HEK cells expressing TSH receptors (HEK-TSHR cells). (Adapted from Neumann S, Malik SS, Marcus-Samuels B, et al. Thyrotropin causes dose-dependent biphasic regulation of cAMP production mediated by G_s and $G_{i/o}$ proteins. *Mol Pharmacol.* 2020;97:2–8.)

regulation by TSH in rodent thyroid cell lines and other animal cells in culture, only those findings confirmed in (*in vivo*) rodent studies and in human thyrocytes in short-term primary cell cultures will be emphasized here. This approach was selected, as there have been contradictory and variable findings in those other systems. For this chapter, the thyroid-specific genes whose regulation will be described include thyroglobulin (*TG*), thyroperoxidase (*TPO*), sodium-iodide symporter (*NIS*), deiodinase type 2 (*DIO2*) and *TSHR*. These were chosen because they are regulated by TSH in thyrocytes and are important in thyroid hormone biosynthesis in the thyroid gland, even though several of these genes are expressed in other tissues as well. Although the individual transcription factors that regulate thyrocyte function may be found in other tissues, their expression together appears to occur only in thyroid cells. It is known that four thyroid transcription factors (TTFs) must be expressed simultaneously to allow for the differentiation of stem/precursor cells to thyrocytes during mouse development.[114] These are homeobox protein NKX2-1 (also known as TTF-1), forkhead box protein FOXE1 (also known as TTF-2), paired box protein PAX8, and hematopoietically expressed homeobox protein HHEX. In contrast to the well-described roles of these TTFs in mouse development, their roles and targets in adult differentiated thyroid cells are less well understood. There have been several reports of the effects of deleting individual TTFs on expression of genes in rodent thyroid cell lines, for example, rat FRTL-5 and PCCL3 cells but few in human cell lines.[114-117] For example,[115] showed that *NIS* transcription is upregulated by TSH in normal thyroid cells in culture, and that NKX2-1 overexpression increased *TG* but not *NIS* or *TSHR* in human cells.[116] However, the effects of all four TTFs in TSH-induced regulation of thyroid-specific genes was not delineated. Recently, it was found[117] that NKX2-1, FOXE1, and PAX8, but not HHEX, are required for maximal TSH-induced upregulation of the transcription of the thyroid-specific genes in adult differentiated human thyrocytes. In human thyrocytes in primary culture, individual knockdown of NKX2-1, FOXE1, or PAX8 expression decreased the expression of the mRNAs for *TG*, *TPO*, *DIO2*, *NIS*, and *TSHR*. Knockdown of HHEX had no effect on upregulation of these genes by TSH, but it was found that knockdown of HHEX increased *TG* mRNA in the absence of TSH, and therefore HHEX appears to be a repressor of basal *TG* transcription.[117,118] Thus, it is clear that NKX2-1, FOXE1, and PAX8 are required for optimal expression of thyroid-specific genes induced by TSH, but the role of HHEX on gene transcription in the adult thyroid remains undetermined.[114]

It has recently been revealed that TSH regulation of transcription of genes involved in thyroid hormone biosynthesis in human thyrocytes elicits IUDRCs.[119] The three TTFs—NKX2-1, FOXE1, and PAX8—are required for optimal upregulation of the transcription of *TG*, *TPO*, and *NIS* in rat cell lines.[120] The role of HHEX is least well understood, but it has been reported that it is a repressor of the *TG* promoter and that it is able to abolish the activating effects of PAX8 and NKX2-1 in engineered cells.[119] In primary cultures of human thyrocytes,[120] TSH elicited a biphasic response in terms of stimulation of transcription of *TG, TPO, DIO2, NIS,* and *TSHR*. Stimulation by TSH of transcription of these five genes was dependent on NKX2-1, FOXE1, and PAX8, as knocking down each of these TTFs inhibited upregulation of *TG, TPO, DIO2, NIS*, and *TSHR*.[119] In human thyrocytes, HHEX does not appear to be involved in upregulation of thyroid specific genes but does appear to repress basal levels of TG. However, as noted above, the full signaling pathways participating in these regulations have not been elucidated, but it is clear that multiple pathways are involved in the transcriptional regulation of each of these genes. Thus, it has been proposed that cell-autonomous, biphasic regulation of thyroidal responses to TSH complements the well-known feedback inhibition by thyroid hormones on TSH secretion from the pituitary gland to prevent hyperstimulation of thyroid gland function.

KEY POINTS

- A cell-autonomous, biphasic regulation of thyroid-stimulating hormone (TSH) receptor–mediated thyroid-specific gene expression by different doses of TSH has been shown in normal human thyroid cells in primary cultures. This biphasic response to TSH may be another fine-tuning regulatory mechanism, in addition to the hypothalamic-pituitary-thyroid axis negative feedback and inhibition by thyroglobulin, to limit hyperstimulation of thyroid gland function.

THYROID-STIMULATING HORMONE RECEPTOR/INSULIN-LIKE GROWTH FACTOR 1 RECEPTOR CROSSTALK

Crosstalk between cell-surface GPCRs and receptor tyrosine kinases (RTKs) is a ubiquitous phenomenon that leads to broadening of the signaling responses initiated by the specific activation of one receptor.[94] In the case of GPCRs and RTKs, crosstalk may occur after activation of either receptor; this has been termed bidirectional transactivation.[121,122] Although these interactions may occur at several steps in the signal transduction pathways, a large number of these involve GPCR–RTK interactions at the cell surface membrane. Crosstalk of this type may be dependent on the generation of a ligand for the second receptor by activation of the first receptor, or may be independent of ligand generation.[123] Crosstalk independent of ligand generation may result from receptor–receptor interactions: when the binding of a ligand to one receptor causes, via direct interactions, a change in the ligand recognition, decoding, and trafficking processes of another receptor.[124] However, an alternative to direct receptor–receptor interaction may be the formation of a multiprotein complex ("signalosome") in which two receptors are brought in close proximity and may influence each other's signaling. This type of receptor–receptor interaction appears to be the mechanism of crosstalk between the TSHR, a GPCR, and the insulin-like growth factor 1 receptor (IGF1R), an RTK (Fig. 65.5). Functional interactions between TSHR and IGF1R in thyrocytes *in vitro*[125-128] have been reported for many years; however, the majority of these studies were performed in rodent cell lines. Functional interactions between TSHR and IGF1R have also been shown *in vitro* in orbital fibroblasts from patients with Graves disease.[129-132] Studies in primary cultures have recently confirmed TSHR/IGF1R crosstalk in normal, differentiated, human thyrocytes[133] (Fig. 65.5). Crosstalk was shown in two ways. First, simultaneous stimulation by TSH and IGF1 led to a synergistic increase in the upregulation of *NIS* gene transcription and NIS protein expression. And second, an IGF1R kinase inhibitor inhibited stimulation by TSH.[133] These findings were confirmed and expanded by showing that β-arrestin 1 anchors TSHR and IGF1R in a signalosome, and that disruption of the signalosome by knocking down β-arrestin 1 inhibits TSHR/IGF1R crosstalk and the synergistic increase in NIS expression.[134] Thus, TSHR/IGF1R crosstalk has an important role in regulation of thyrocyte function under physiological stimulation by TSH, but also in pathophysiological stimulation by TSHR-stimulating antibodies, especially in TED (see Chapter 72).

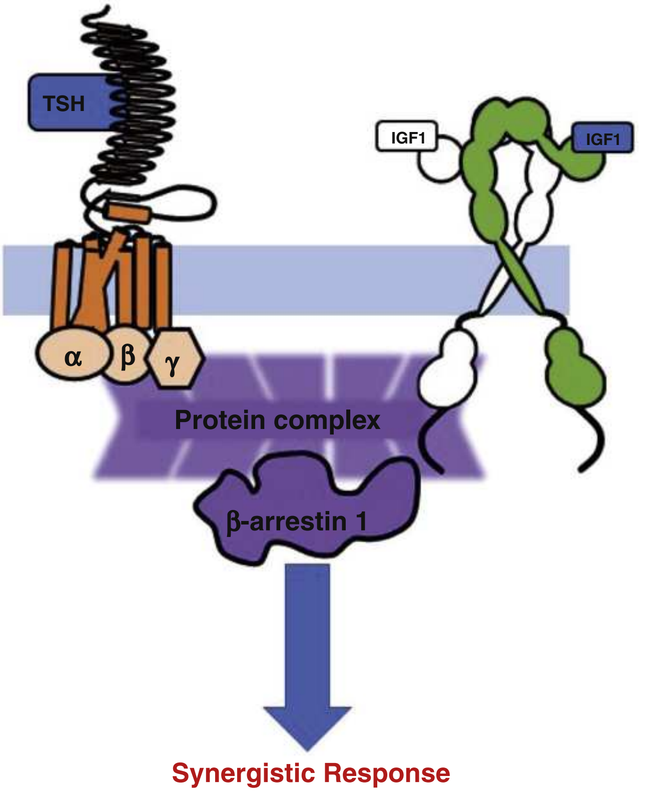


Fig. 65.5 Model of thyroid-stimulating hormone receptor (TSHR)/insulin-like growth factor receptor 1 (IGF1R) crosstalk. Activation of TSHR by thyroid-stimulating hormone (TSH) activates two signal transduction pathways—one that is independent of the IGF1R and another that is dependent on IGF1R. TSHR and IGF1R are in close proximity within a signalosome for which β-arrestin 1 acts as a scaffold. TSHR- and IGF1R-mediated pathways combine, leading to a more synergistic stimulation of *NIS* gene and protein expression in primary cultures of human thyrocytes. TSHR/IGF1R crosstalk leads also to a synergistic increase in hyaluronan secretion in Graves orbital fibroblasts, which is a major component of thyroid eye disease (see Chapter 72). (Modified from Krieger CC, Place RF, Bevilacqua C, et al. TSH/IGF-1 receptor cross talk in Graves' ophthalmopathy pathogenesis. *J Clin Endocrinol Metab*. 2016;101:2340–2347 and Krieger CC, Neumann S, Gershengorn MC. TSH/IGF1 receptor crosstalk: Mechanism and clinical implications. *Pharmacol Ther*. 2020;209:107502.)

KEY POINTS

- TSHR/IGF1R crosstalk initiated by TSHR activation has been observed in cell culture in thyrocytes and Graves orbital fibroblasts, and it has been demonstrated that it plays a pivotal role in the pathophysiology of Graves hyperthyroidism and TED. The activated TSHR engages the IGF1R in a signalosome that is scaffolded by β-arrestin 1, and this crosstalk induces a synergistic increase in NIS expression in thyrocytes and hyaluronan secretion in Graves orbital fibroblasts.

Regulation of Thyroid Hormone Synthesis

Thyroid hormone synthesis requires iodide uptake, TG biosynthesis, oxidation of iodide, and coupling of two iodotyrosines into iodothyronines within TG.[135] Iodide is actively transported into the thyrocyte by NIS within the thyrocyte basal membrane and diffuses down an electrical gradient to the apical membrane. As mentioned above, TSH stimulates expression of TG and NIS by increasing the rates of their transcription. In some species, TSH also has been reported to increase iodide uptake by increasing thyroid blood flow. Iodide oxidation and coupling to tyrosine residues within TG, and iodotyrosine coupling to iodothyronines are catalyzed by the same enzyme, TPO, which is within the apical membrane of the thyrocyte. Hydrogen peroxide is generated by two apical membrane dual oxidases. TSH stimulates the rate of transcription of TPO and of dual oxidase 2 (DUOX2), which is the major dual oxidase in the thyrocyte. In human thyrocytes, hydrogen peroxide, which oxidizes iodide for coupling, but not TPO, is also stimulated by direct activation of preformed DUOX2 TSH.[136]

Regulation of Thyroid Hormone Secretion

Secretion of thyroid hormones requires endocytosis of follicle TG via the apical membrane, its hydrolysis within the cell, and release of thyroid hormones from the basolateral membrane of the cell. The major endocytotic pathway involves pseudopod extrusion from the apical surface of the thyrocyte and engulfment of particles of TG. This phenomenon is mediated by direct activation of intracellular signaling pathways stimulated by TSH (for review, see[137]). The endocytosed vesicles then fuse with lysosomes in which TG is hydrolyzed. Thyroxine is the major iodothyronine in TG. In this process, a small amount of T_4 is deiodinated by 5-deiodinases to T_3, thereby increasing the amount of T_3 that is secreted. Free thyroid hormones are released by the thyroid hormone transporter SLC16A2. It has also been suggested that thyroid hormones may be released by transcytosis around the lateral sides of the thyrocyte. TG is also released by the thyroid gland. Like thyroid hormones, TG secretion appears to be via endocytosis, but transcytosis has been proposed also.

EXTRATHYROIDAL ACTIONS MEDIATED BY THYROID-STIMULATING HORMONE RECEPTOR

TSHR has been shown to be expressed in a number of extrathyroidal tissues, including fat cells,[138] fibroblasts, preadipocytes, and adipocytes in the retroorbital space,[131,139] osteoblasts and osteoclasts,[140] lymphocytes,[141] erythrocytes,[142] brain,[143] and pituitary.[144] Expression of TSHR has often been demonstrated at the level of mRNA and in fewer tissues at the level of functional TSHR. Nevertheless, the physiologic role(s) of TSHR in most of these tissues is not well-defined. However, evidence has been reported for a pathophysiological role of TSHR in the orbital space (see Chapter 72).

Functional TSHRs are expressed in adipocytes,[139] and TSH was shown to stimulate proliferation and inhibit differentiation of rat preadipocytes *in vitro*.[145] Recently, a study using adipose tissue–specific partial knockdown in mice[146] reported that loss of TSHRs in white and brown adipose tissue led to a predisposition to weight gain and effects on regulation of glucose tolerance, lipid metabolism, adipokine profile, and thermogenesis. Nevertheless, at this time, the physiologic role of the TSHR in adipose tissue has not been well-defined.

There is abundant evidence of an important role of thyroid hormones on bone metabolism.[147] By contrast, although TSHRs have been found in bone, the question of whether there is a direct, thyroid hormone–independent action of TSH on bone remains unresolved.[140,148] When TSHR is overexpressed in a human preosteoblast cell line,[102,105,149] several responses to TSH have been described. These include: 1) TSHR couples to several different G proteins to induce upregulation of genes associated with osteoblast activity—interleukin 1, osteopontin, and alkaline phosphatase; 2) TSH can stimulate β-arrestin 1–mediated signal transduction, and β-arrestin 1 in part mediates TSH-induced preosteoblast differentiation; and 3) TSHR/IGF1R crosstalk synergistically increased osteopontin secretion by TSH and IGF1.

TSHR expression has been demonstrated in rodent osteoblasts and osteoclasts. Direct effects of TSH were described in bone,[150] where it modulates TNF alpha and receptor activator of nuclear factor-κB ligand production to decrease osteoclast differentiation.[151] TSHR knockout causes osteoporosis and focal osteosclerosis in mice.[152] TSH administration inhibits bone loss in adult, ovariectomized rodents, consistent with the idea that TSH is an activator of bone formation.[153] Overall, *in vivo* studies in mice and rats propose TSH as a fine-tuning regulator of bone homeostasis, and these findings have been comprehensibly reviewed.[140,147,154,155]

Several clinical studies have also pointed to direct action of TSH on bone. Administration of recombinant human TSH (rhTSH) in postmenopausal women increased serum N-terminal propeptide of type-I procollagen, a marker of bone formation, demonstrating an anabolic effect of TSH in humans.[156] Mazziotti et al. showed that short–term rhTSH stimulation leads to a reversible inhibition of bone resorption in postmenopausal women, suggesting a role for TSH in patients with bone loss and a high bone turnover rate.[157] Furthermore, epidemiological studies demonstrated a tight relationship between low TSH levels and parameters of bone loss and fracture risk. However, the role of TSH on the adult skeleton and its mechanisms of action in human bone have yet to be definitively demonstrated and are still controversial (reviewed in[140,155,158]).

Also, although TSHR is expressed in human bone, TSH is not expressed in primary human osteoblasts or osteoclasts.[91] Pituitary TSH works systemically, and therefore it is likely that it can activate TSHRs in bone. Thyrostimulin has been considered as a regulator of bone formation, perhaps acting in an autocrine/paracrine fashion. In contrast to TSH, thyrostimulin is expressed in osteoblasts and osteoclasts.[150] The combination of *in vivo* and *in vitro* studies demonstrated a role for thyrostimulin during skeletal development, but an essential role in the adult skeleton has not been shown.[150]

The presence of TSHRs in orbital fibroblasts, preadipocytes, and adipocytes in the retroorbital tissue led to speculation about their possible pathophysiological role in TED (see Chapter 72). The hypothesis generated was that the thyroid-stimulating antibodies (or TSHR-stimulating antibodies [TSHRSAbs]), the well-recognized mediators of Graves hyperthyroidism, were a major component of TED by activating TSHRs on resident retroorbital cells.[131,16] It has been suggested that a second source of these target cells may be circulating, bone marrow–derived cells termed fibrocytes.[159] Furthermore, it has been proposed that the pathogenesis of TED involves IGF1R also[160,161] (see the earlier section on Thyroid-Stimulating Hormone Receptor /Insulin-like Growth Factor 1 Receptor Crosstalk). It is important to note that these hypotheses are supported primarily by studies in primary cultures of cells obtained at surgical decompression from retroorbital tissue termed Graves orbital fibroblasts or fibrocytes, and not from studies in intact humans. An interesting controversy regarding the pathogenesis of TED has arisen as to whether activation of the IGF1R is because it is a direct target of stimulating antibodies or whether IGF1R is activated secondarily after activation of TSHRs by TSHRSAbs via TSHR/IGF1R crosstalk. We favor the hypothesis that TED pathogenesis involves TSHR/IGF1R crosstalk. (A more complete discussion of the pathogenesis of TED can be found in Chapter 72.)

KEY POINTS

- Extrathyroidal thyroid-stimulating hormone receptor (TSHR) expression has been reported in orbital fibroblasts, bone, brain, kidney, testis, heart, and adipose tissue. Thyroid-stimulating hormone (TSH) is a fine-tuning regulator of bone homeostasis. However, despite evidence of an important role of thyroid hormones on bone metabolism, the question of whether there is a direct, thyroid hormone-independent action of TSH on bone remains unresolved. In contrast, TSHR in orbital fibroblasts has been shown to play a pivotal role in the pathogenesis of thyroid eye disease. TSHR-induced signaling is mediated through an insulin-like growth factor 1 receptor (IGF1R)-independent pathway and an IGF1R-dependent pathway.

REFERENCES

1. Shupnik MA, Greenspan SL, Ridgway EC. Transcriptional regulation of thyrotropin subunit genes by thyrotropin-releasing hormone and dopamine in pituitary cell cultures. *J Biol Chem*. 1986;261:12675–12679.
2. Shibusawa N, Yamada M, Hirato J, et al. Requirement of thyrotropin-releasing hormone for the postna tal functions of pituitary thyrotrophs: ontogeny study of congenital tertiary hypothyroidism in mice. *Mol Endocrinol*. 2000;14:137–146.
3. Gershengorn MC. Mechanism of signal transduction by TRH. *Ann N Y Acad Sci*. 1989;553:191–196.
4. Hsieh KP, Martin TF. Thyrotropin-releasing hormone and gonadotropin-releasing hormone receptors activate phospholipase C by coupling to the guanosine triphosphate-binding proteins Gq and G11. *Mol Endocrinol*. 1992;6:1673–1681.
5. Kiley SC, Parker PJ, Fabbro D, et al. Differential regulation of protein kinase C isozymes by thyrotropin-releasing hormone in GH4C1 cells. *J Biol Chem*. 1991;266:23761–23768.
6. Jefferson AB, Travis SM, Schulman H. Activation of multifunctional Ca2+/calmodulin-dependent protein kinase in GH3 cells. *J Biol Chem*. 1991;266:1484–1490.
7. Cui ZJ, Gorelick FS, Dannies PS. Calcium/calmodulin-dependent protein kinase-II activation in rat pituitary cells in the presence of thyrotropin-releasing hormone and dopamine. *Endocrinology*. 1994;134:2245–2250.
8. Ohmichi M, Sawada T, Kanda Y, et al. Thyrotropin-releasing hormone stimulates MAP kinase activity in GH3 cells by divergent pathways. Evidence of a role for early tyrosine phosphorylation. *J Biol Chem*. 1994;4:2693783–2693788.
9. Gershengorn MC, Osman R. Molecular and cellular biology of thyrotropin-releasing hormone (TRH) receptors. *Physiol Rev*. 1996;76:175–191.
10. Mailman RB, Frye GD, Mueller RA, et al. Thyrotropin-releasing hormone reversal of ethanol-induced decreases in cerebellar cGMP. *Nature*. 1978;272:832–833.
11. Nakayama T, Hashimoto T, Nagai Y. Involvement of glutamate and gamma-aminobutyric acid (GABA)-ergic systems in thyrotropin-releasing hormone-induced rat cerebellar cGMP formation. *Eur J Pharmacol*. 1996;316:157–164.
12. Kim DS, Yoon JH, Ahn SK, et al. A 33kDa Pit-1-like protein binds to the distal region of the human thyrotrophin alpha-subunit gene. *J Mol Endocrinol*. 1995;14:313–322.
13. Hashimoto K, Zanger K, Hollenberg AN, et al. cAMP response element-binding protein-binding protein mediates thyrotropin-releasing hormone signaling on thyrotropin subunit genes. *J Biol Chem*. 2000;275:33365–33372.
14. Visser WE, Friesema EC, Visser TJ. Minireview: thyroid hormone transporters: the knowns and the unknowns. *Mol Endocrinol*. 2011;25:1–14.
15. St Germain DL, Galton VA. The deiodinase family of selenoproteins. *Thyroid*. 1997;7:655–668.
16. Arrojo e, Drigo R, Bianco AC. Type 2 deiodinase at the crossroads of thyroid hormone action. *Int J Biochem Cell Biol*. 2011;43:1432–1441.
17. Rosene ML, Wittmann G, Arrojo e, Drigo R, et al. Inhibition of the type 2 iodothyronine 5'-deiodinase underlies the elevated plasma TSH associated with amiodarone treatment. *Endocrinol*. 2010;151:5961–5970.
18. Shupnik MA, Chin WW, Habener JF, et al. Transcriptional regulation of the thyrotropin subunit genes by thyroid hormone. *J Biol Chem*. 1985;260:2900–2903.
19. Shupnik MA, Ridgway EC. Thyroid hormone control of thyrotropin gene expression in rat anterior pituitary cells. *Endocrinology*. 1987;121:619–624.
20. Shupnik MA, Ardisson LJ, Meskell MJ, et al. Triiodothyronine (T3) regulation of thyrotropin subunit gene transcription is proportional to T3 nuclear receptor occupancy. *Endocrinology*. 1986;118:367–371.
21. Refetoff S. Resistance to thyroid hormone. *Curr Ther Endocrinol Metab*. 1997;6:132–134.
22. Ordentlich P, Downes M, Evans RM. Corepressors and nuclear hormone receptor function. *Curr Top Microbiol Immunol*. 2001;254:101–116.
23. Liu Y, Xia X, Fondell JD, et al. Thyroid hormone-regulated target genes have distinct patterns of coactivator recruitment and histone acetylation. *Mol Endocrinol*. 2006;20:483–490.
24. Matsushita A, Sasaki S, Kashiwabara Y, et al. Essential role of Gata2 in the negative regulation of thyrotropin beta gene by thyroid hormone and its receptors. *Mol Endocrinol*. 2007;21:865–884.
25. Wang D, Xia X, Weiss RE, et al. Distinct and histone-specific modifications mediate positive versus negative transcriptional regulation of TSHα promoter. *PLoS One*. 2010; 5:e9853.
26. Wood WM, Ocran KO, Gordon DF, et al. Isolation and characterization of mouse complementary DNAs encoding alpha and beta thyroid hormone receptors from thyrotrope cells: the mouse pituitary specific beta-2 isoform differs at the amino terminus from the corresponding species from rat pituitary tumor cells. *Mol Endocrinol*. 1991;5:1049–1061.
27. Hodin RA, Lazar MA, Wintman BI, et al. Identification of a thyroid hormone receptor that is pituitary-specific. *Science*. 1989;244:76–79.
28. Langlois MF, Zanger K, Monden T, et al. A unique role of the beta-2 thyroid hormone receptor isoform in negative regulation by thyroid hormone: mapping of a novel amino-terminal domain important for ligand-independent activation. *J Biol Chem*. 1997;272:24927–24933.
29. Chiamolera MI, Sidhaye AR, Matsumoto S, et al. Fundamentally distinct roles of thyroid hormone receptor isoforms in a thyrotroph cell line are due to differential DNA binding. *Mol Endocrinol*. 2012;26:926–939.
30. Hallenbeck PL, Phyillaier M, Nikodem V. Divergent effects of 9-cis-retinoic acid receptor on positive and negative thyroid hormone receptor-dependent gene expression. *J Biol Chem*. 1993;268:3825–3828.
31. Haugen BR, Brown NS, Wood WM, et al. The thyrotrope-restricted isoform of the retinoid X receptor (gamma 1) mediates 9-cis retinoic acid suppression of thyrotropin beta promoter activity. *Mol Endocrinol*. 1997;11:481–489.
32. Sarapura VD, Wood WM, Bright TM, et al. Reconstitution of triiodothyronine inhibition in non-triiodothyronine responsive thyrotropic tumor cells using transfected thyroid hormone receptor isoforms. *Thyroid*. 1997;7:453–461.
33. Tagami T, Madison LD, Nagaya T, et al. Nuclear receptor corepressors activate rather than suppress basal transcription of genes that are negatively regulated by thyroid hormone. *Mol Cell Biol*. 1997;17:2642–2648.
34. Madison LD, Ahlquist JA, Rogers SD, et al. Negative regulation of the glycoprotein hormone alpha gene promoter by thyroid hormone: mutagenesis of a proximal receptor binding site preserves transcriptional repression. *Mol Cell Endocrinol*. 1993;94:129–136.
35. Persani L. Hypothalamic thyrotropin-releasing hormone and thyrotropin biological activity. *Thyroid*. 1998;8:941–946.
36. Persani L, Borgato S, Romoli R, et al. Changes in the degree of sialylation of carbohydrate chains modify the biological properties of circulating thyrotropin isoforms in various physiological and pathological states. *J Clin Endo and Metab*. 1998;83:2486–2492.
37. Taylor T, Weintraub BD. Altered thyrotropin (TSH) carbohydrate structures in hypothalamic hypothyroidism created by paraventricular nuclear lesions are corrected by in vivo TSH-releasing hormone administration. *Endocrinology*. 1989;125:2198–2203.

38. Menezes-Ferreira MM, Petrick PA, Weintraub BD. Regulation of thyrotropin bioactivity by thyrotropin-releasing hormone and thyroid hormone. *J Endocrinol*. 1986;118:2125–2130.
39. Persani L, Cangiano B, Bonomi M. The diagnosis and management of central hypothyroidism in 2018. *Endocr Connect*. 2019;8:R44–R54.
40. Papandreou MJ, Persani L, Asteria C, et al. Variable carbohydrate structures of circulating thyrotropin as studied by lectin affinity chromatography in different clinical conditions. *J Clin Endo and Metab*. 1993;77:393–398.
41. Persani L, Ferretti E, Borgato S, et al. Circulating thyrotropin bioactivity in sporadic central hypothyroidism. *J Clin Endocrinol Metab*. 2000;85:3631–3635.
42. Oliveira JHA, Barbosa ER, Kasamatsu T, et al. Evidence for thyroid hormone as a positive regulator of serum thyrotropin bioactivity. *J Clin Endocrinol Metab*. 2007;92:3108–3113.
43. Sheward WJ, Harmar AJ, Fraser HM, et al. TRH in rat pituitary stalk blood and hypothalamus: studies with high performance liquid chromatography. *Endocrinology*. 1983;113:1865–1869.
44. Fraser HM, McNeilly AS. Effect of chronic immunoneutralisation of thyrotropin-releasing hormone on the hypothalamic-pituitary thyroid axis, prolactin and reproductive function in the Ewe. *Endocrinology*. 1982;111:1964–1971.
45. Aizawa T, Green M. Delineation of the hypothalamic area controlling thyrotropin secretion in the rat. *Endocrinology*. 1981;109:1731–1738.
46. Chan V, Wang C, Yeung TT. Thyrotropin, α and β-subunits of thyrotropin and prolactin response to four hour constant infusions of thyrotropin releasing hormone in normal subjects and patients with pituitary-thyroid disorders. *J Clin Endo and Metab*. 1979;49:127–133.
47. Samuels MH, Henry P, Luther M, et al. Pulsatile TSH secretion during 48-hour continuous TRH infusions. *Thyroid*. 1993;3:201–206.
48. Sheppard MC, Shennan KI. Desensitisation of rat anterior pituitary gland to thyrotrophin releasing hormone. *Endocrinology*. 1984;101:101–105.
49. Vale W, Brazeau P, Rivier C, et al. Somatostatin. *Recent Prog Horm Res* 1975;31:365–397.
50. Reichlin S. Somatostatin. *N Engl J Med*. 1983;309:1495–1501.
51. Urman S, Critchlow V. Long-term elevations in plasma thyrotropin, but not growth hormone, concentrations associated with lesion-induced depletion of median eminence somatostatin. *Endocrinology*. 1983;112:659–664.
52. Arima A, Schally AV. Increase in basal and thyrotropin-releasing hormone stimulated secretion of thyrotropin by passive immunization with antiserum to somatostatin. *Endocrinology*. 1976;98:1069–1075.
53. Samuels MH, Henry P, Ridgway EC. Effects of dopamine and somatostatin on pulsatile pituitary glycoprotein secretion. *J Clin Endo and Metab*. 1992;74:217–222.
54. Reisine T, Bell GI. Molecular biology of the somatostatin receptors. *Endocrine Rev*. 1995;16:427–442.
55. Theodoropoulou M, Stalla GK. Somatostatin receptors: from signaling to clinical practice. *Front Neuroendocrinol*. 2013;34:228–252.
56. James RA, Sarapura VD, Bruns C, et al. Thyroid hormone-induced expression of specific somatostatin receptor subtypes correlates with involution of the TtT-97 murine thyrotrope tumor. *Endocrinology*. 1997;138:719–724.
57. Ridgway EC, Klibanski A, Martorana MA, et al. The effect of somatostatin on the release of thyrotropin and its subunits from bovine anterior pituitary cells in vitro. *Endocrinology*. 1983;112:1937–1942.
58. Page MD, Millward ME, Taylor A, et al. Long-term treatment of acromegaly with a long-acting analogue of somatostatin, octreotide. *Q J Med*. 1990;74:189–201.
59. Lippe BM, Van Herle AJ, La Franchi SH, et al. Reversible hypothyroidism in growth hormone-deficient children treated with human growth hormone. *J Clin Endo and Metab*. 1975;40:612–618.
60. Cooper DS, Klibanski A, Ridgway EC. Dopaminergic modulation of TSH and its subunits: in vivo and in vitro studies. *Clin Endocrinol (Oxf)*. 1983;18:265–272.
61. Pourmand M, Rodriguez-Arnao MD, Weightman DR, et al. Domperidone: a novel agent for the investigation of anterior pituitary function and control in man. *Clin Endocrinol (Oxf)*. 1980;12:211–215.
62. Samuels MH, Kramer P. Effects of metoclopramide on fasting-induced TSH suppression. *Thyroid*. 1996;6:85–89.
63. Lewis BM, Dieguez C, Lewis MD, et al. Dopamine stimulates release of thyrotrophin-releasing hormone from perfused intact rat hypothalamus via hypothalamic D2 receptors. *J Endocrinol*. 1987;115:419–424.
64. Morley JE. Neuroendocrine control of thyrotropin secretion. *Endocrine Rev*. 1981;2:396–436.
65. Foord SM, Peters JR, Dieguez C, et al. Dopamine receptors on intact anterior pituitary cells in culture: functional association with the inhibition of prolactin and thyrotropin. *Endocrinology*. 1983;112:1567–1571.
66. Foord SM, Peters JR, Dieguez C, et al. TSH regulates thyrotroph responsiveness to dopamine in vitro. *Endocrinology*. 1985;118:1319–1324.
67. Kok P, Roelfsema F, Frolich M, et al. Spontaneous diurnal thyrotropin secretion is enhanced in proportion to circulating leptin in obese premenopausal women. *J Clin Endocrinol Metab*. 2005;90:6185–6191.
68. Scanlon MF, Chan V, Heath M, et al. Dopaminergic control of thyrotropin, α-subunit and prolactin in euthyroidism and hypothyroidism: dissociated responses to dopamine receptor blockade with metoclopramide in euthyroid and hypothyroid subjects. *J Clin Endo and Metab*. 1981;53:360–365.
69. Morley JE, Baranetsky NG, Wingert TD, et al. Endocrine effects of naloxone-induced opiate receptor blockade. *J Clin Endo and Metab*. 1980;50:251–257.
70. Samuels MH, Kramer P, Wilson D, et al. Effects of naloxone infusions on pulsatile thyrotropin secretion. *J Clin Endo and Metab*. 1994;78:1249–1252.
71. Larsen PR. Thyroid-pituitary interaction. *N Engl J Med*. 1982;306:23–32.
72. Samuels MH, Veldhuis JD, Henry P, et al. Pathophysiology of pulsatile and copulsatile release of thyroid-stimulating hormone, luteinizing hormone, follicle-stimulating hormone, and α-subunit. *J Clin Endo and Metab*. 1990;71:425–432.
73. Hinkle PM, Goh KBC. Regulation of thyrotropin-releasing hormone receptors and responses by L-triiodothyronine in dispersed rat pituitary cell cultures. *Endocrinology*. 1982;110:1725–1731.
74. Ponce G, Charli JL, Pasten JA, et al. Tissue-specific regulation of pyroglutamate aminopeptidase II activity by thyroid hormones. *Neuroendocrinol*. 1988;48:211–214.
75. Kakucska I, Rand W, Lechan RM. Thyrotropin-releasing hormone gene expression in the hypothalamic paraventricular nucleus is dependent upon feedback regulation by both triiodothyronine and thyroxine. *Endocrinology*. 1992;130:2845–2850.
76. Berelowitz M, Maefa K, Harris S, et al. The effect of alterations in the pituitary-thyroid axis on hypothalamic content and in vitro release of somatostatin-like immunoreactivity. *Endocrinology*. 1980;107:24–29.
77. Bartalena L, Martino E, Petrini L, et al. The nocturnal serum thyrotropin surge is abolished in patients with adrenocorticotropin (ACTH)-dependent or ACTH-independent Cushing's Syndrome. *J Clin Endo and Metab*. 1991;72:1195–1199.
78. Samuels MH, Luther M, Henry P, et al. Effects of hydrocortisone on pulsatile pituitary glycoprotein secretion. *J Clin Endo and Metab*. 1994;78:211–215.
79. Samuels MH, McDaniel PA. Thyrotropin levels during hydrocortisone infusions that mimic fasting-induced cortisol elevations—a clinical research center study. *J Clin Endo and Metab*. 1997;82:3700–3704.
80. Samuels MH. Effects of variations in physiological cortisol levels on thyrotropin secretion in subjects with adrenal insufficiency: a clinical research center study. *J Clin Endocrinol Metab*. 2000;85:1388–1393.
81. Pamenter RW, Hedge GA. Inhibition of thyrotropin secretion by physiological levels of corticosterone. *Endocrinology*. 1980;106:162–166.
82. Mitsuma T, Nogimori T. Effects of dexamethasone on the hypothalamic-pituitary-thyroid axis in rats. *Acta Endocrinol (Copenh)*. 1982;100:51–56.
83. Ortiga-Carvalho TM, Oliveira KJ, Soares BA, et al. The role of leptin in the regulation of TSH secretion in the fed state: in vivo and in vitro studies. *J Endocrinol*. 2002;174:121–125.
84. Mantzoros CS, Ozata M, Negrao AB, et al. Synchronicity of frequently sampled thyrotropin (TSH) and leptin concentrations in healthy adults and leptin-deficient subjects: evidence for possible partial TSH regulation by leptin in humans. *J Clin Endocrinol Metab*. 2001;86:3284–3291.

85. Yildiz BO, Aksoy DY, Harmanci A, et al. Effects of L-thyroxine therapy on circulating leptin and adiponectin levels in subclinical hypothyroidism: a prospective study. *Arch Med Res*. 2013;44:317–320.
86. Santini F, Galli G, Maffei M, et al. Acute exogenous TSH administration stimulates leptin secretion in vivo. *Eur J Endocrinol*. 2010;163:63–67.
87. Hermus RM, Sweep CG, van der Meer MJ, et al. Continuous infusion of interleukin-1 induces a nonthyroidal illness syndrome in the rat. *Endocrinology*. 1992;131:2139–2146.
88. Van der Poll T, Romijn JA, Wiersinga WM, et al. Tumor necrosis factor: a putative mediator of the sick euthyroid syndrome in man. *J Clin Endo and Metab*. 1990;71:1567–1572.
89. Torpy DJ, Tsigos C, Lotsikas AJ, et al. Acute and delayed effects of a single-dose injection of interleukin-6 on thyroid function in healthy humans. *Metabolism*. 1998;47:1289–1293.
90. Karponis D, Ananth S. The role of thyrostimulin and its potential clinical significance. *Endocr Regul*. 2017;51:117–128.
91. Bassett JHD, van der Spek A, Logan JG, et al. Thyrostimulin regulates osteoblastic bone formation during early skeletal development. *Endocrinology*. 2015;156:3098–3113.
92. Vassart G, Pardo L, Costagliola S. A molecular dissection of the glycoprotein hormone receptors. *Trends Biochem Sci*. 2004;29:119–126.
93. Briet C, Suteau-Courant V, Munier M, et al. Thyrotropin receptor, still much to be learned from the patients. *Best Pract Res Clin Endocrinol Metabol*. 2018;32:155e164.
94. Hilger D, Masureel M, Kobilka BK. Structure and dynamics of GPCR signaling Complexes. *Nat Struct Mol Biol*. 2018;4:4–12.
95. Wingler LM, Lefkowitz RJ. Conformational basis of G protein-coupled receptor signaling versatility. *Trends Cell Biol*. 2020;30:736–747.
96. Rapoport B, McLachlan SM. TSH receptor cleavage into subunits and shedding of the a-subunit; a molecular and clinical perspective. *Endocr Rev*. 2016;2016:23–42.
97. Maurice P, Kamal M, Jockers R. Asymmetry of GPCR oligomers supports their functional relevance. *Trends Pharmacol Sci*. 2011;32:514–520.
98. Ferré S, Casadó V, Devi LA, et al. G protein–coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacol Rev*. 2014;66:413–434.
99. Urizar E, Montanelli L, Loy T, et al. Glycoprotein hormone receptors: link between receptor homodimerization and negative cooperativity. *EMBO J*. 2005;24:1954–1964.
100. Persani L, Calebiro D, Bonomi M. Technology Insight: modern methods to monitor protein–protein interactions reveal functional TSH receptor oligomerization. *Nat Clin Pract Endocrinol Metabol*. 2007;3:180–190.
101. Laugwitz KL, Allgeier A, Offermanns S, et al. The human thyrotropin receptor: a heptahelical receptor capable of stimulating members of all four G protein families. *Proc Natl Acad Sci U S A*. 1996;93:116–120.
102. Boutin A, Eliseeva E, Gershengorn MC, et al. beta-Arrestin-1 mediates thyrotropin-enhanced osteoblast differentiation. *FASEB J*. 2014;28:3446–3455.
103. Boutin A, Gershengorn MC, Neumann S. β-arrestin 1 in thyrotropin receptor signaling in bone: studies in osteoblast-like cells. *Front Endocrinol*. 2020;11:312.
104. Morgan SJ, Neumann S, Gershengorn MC. Normal human thyrocytes in culture. *Methods Mol Biol*. 2018;1817:1–7.
105. Boutin A, Krieger CC, Marcus-Samuels B, et al. TSH receptor homodimerization in regulation of cAMP production in human thyrocytes in vitro. *Front Endocrinol*. 2020;11:276.
106. Allen MD, Neumann S, Gershengorn MC. Occupancy of both sites on the thyrotropin (TSH) receptor dimer is necessary for phosphoinositide signaling. *FASEB J*. 2011;25:3687–3694.
107. Neumann S, Malik SS, Marcus-Samuels B, et al. Thyrotropin causes dose-dependent biphasic regulation of cAMP production mediated by Gs and Gi/o proteins. *Mol Pharmacol*. 2020;97:2–8.
108. Laurent E, Van Sande J, Ludgate M, et al. Unlike thyrotropin, thyroid-stimulating antibodies do not activate phospholipase C in human thyroid slices. *J Clin Invest*. 1991;87:1634–1642.
109. Calabrese EJ. Biphasic dose responses in biology, toxicology and medicine: accounting for their generalizability and quantitative features. *Environ Pollut*. 2013;182:452e460.
110. Vandenberg LN, Colborn T, Hayes TB, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr Rev*. 2012;33:378–455.
111. Jamshidi RJ, Jacobs BA, Sullivan LC, et al. Functional selectivity of kappa opioid receptor agonists in peripheral sensory neurons. *J Pharmacol Exp Ther*. 2015;355:174–182.
112. Riters LV, Pawlisch BA, Kelm-Nelson CA, et al. Inverted-U shaped effects of D1 dopamine receptor stimulation in the medial preoptic nucleus on sexually motivated song in male European starlings. *Eur J Neurosci*. 2014;39:650–662.
113. Kozumbo WJ, Calabrese EJ. Two decades (1998-2018) of research progress on hormesis: advancing biological understanding and enabling novel applications. *J Cell Commun Signal*. 2019;13:273–275.
114. Fernandez LP, Lopez-Marquez A, Santisteban P. Thyroid transcription factors in development, differentiation and disease. *Nat Rev Endocrinol*. 2015;11:29–42.
115. Taki K, Kogai T, Kanamoto Y, et al. A thyroid-specific far-upstream enhancer in the human sodium/iodide symporter gene requires Pax-8 binding and cyclic adenosine 3',5'-monophosphate response element-like sequence binding proteins for full activity and is differentially regulated in normal and thyroid cancer cells. *Mol Endocrinol*. 2002;16:2266–2282.
116. Espadinha C, Cavaco BM, Leite V. PAX8PPARgamma stimulates cell viability and modulates expression of thyroid-specific genes in a human thyroid cell line. *Thyroid*. 2007;17:497–509.
117. Jang D, Marcus-Samuels B, Morgan SJ, et al. Thyrotropin regulation of differentiated gene transcription in adult human thyrocytes in primary culture. *Mol Cell Endocrinol*. 2020;518:111032.
118. Pellizzari L, D'Elia A, Rustighi A, et al. Expression and function of the homeodomain-containing protein Hex in thyroid cells. *Nucleic Acids Res*. 2000;28:2503–2511.
119. Jang D, Morgan SJ, Klubo-Gwiezdzinska J, et al. TSH, but not thyroid-stimulating antibodies, induces biphasic regulation of gene expression in human thyrocytes. *Thyroid*. 2020;30:270–276.
120. Ortiz L, Zannini M, Di Lauro R, et al. Transcriptional control of the forkhead thyroid transcription factor TTF-2 by thyrotropin, insulin, and insulin-like growth factor I. *J Biol Chem*. 1997;272:23334–23339.
121. Delcourt N, Bockaert J, Marin P. GPCR-jacking: from a new route in RTK signalling to a new concept in GPCR activation. *Trends Pharmacol Sci*. 2007;28:602–607.
122. Pyne NJ, Pyne S. Receptor tyrosine kinase-G-protein-coupled receptor signalling platforms: out of the shadow? *Trends Pharmacol Sci*. 2011;32:443–450.
123. Di Liberto V, Mudò V, Belluardo N. Crosstalk between receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCR) in the brain: focus on heteroreceptor complexes and related functional neurotrophic effects. *Neuropharmacology*. 2019;152:67–77.
124. Kenakin T, Agnati LF, Caron M, et al. International Workshop at the Nobel Forum, Karolinska Institutet on G protein-coupled receptors: finding the words to describe monomers, oligomers, and their molecular mechanisms and defining their meaning. Can a consensus be reached? *J Recept Signal Transduct Res*. 2010;30:284–286.
125. Eggo MC, Bachrach LK, Burrow GN. Interaction of TSH, insulin and insulin-like growth factors in regulating thyroid growth and function. *Growth Factors*. 1990;2:99–109.
126. Kimura T, Van Keymeulen A, Golstein J, et al. Regulation of thyroid cell proliferation by TSH and other factors: a critical evaluation of in vitro models. *Endocr Rev*. 2001;22:631–656.
127. Santisteban P, Kohn LD, Di Lauro R. Thyroglobulin gene expression is regulated by insulin and insulin-like growth factor I, as well as thyrotropin, in FRTL-5 thyroid cells. *J Biol Chem*. 1987;262:4048–4052.
128. Tramontano D, Cushing GW, Moses AC, et al. Insulin-like growth factor-I stimulates the growth of rat thyroid cells in culture and synergizes the stimulation of DNA synthesis induced by TSH and Graves'-IgG. *Endocrinology*. 1986;119:940–942.
129. Chen H, Mester T, Raychaudhuri N, et al. Teprotumumab, an IGF-1R blocking monoclonal antibody inhibits TSH and IGF-1 action in fibrocytes. *J Clin Endocrinol Metab*. 2014;99:E1635–E1640.

130. Hoa N, Tsui S, Afifiyan NF, et al. Nuclear targeting of IGF-1 receptor in orbital fibroblasts from Graves' disease: apparent role of ADAM17. *PLoS One*. 2012;7:e34173.
131. Kumar S, Iyer S, Bauer H, et al. A stimulatory thyrotropin receptor antibody enhances hyaluronic acid synthesis in graves' orbital fibroblasts: inhibition by an IGF-I receptor blocking antibody. *J Clin Endocrinol Metab*. 2012;97:1681–1687.
132. Tsui S, NaikV HN, et al. Evidence for an association between thyroid-stimulating hormone and insulin-like growth factor 1 receptors: a tale of two antigens implicated in Graves' disease. *J Immunol*. 2008;181:4397–4405.
133. Morgan SJ, Neumann S, Marcus-Samuels, et al. Thyrotropin and insulin-like growth factor 1 receptor crosstalk upregulates sodium–iodide symporter. expression in primary cultures of human thyrocytes. *Thyroid*. 2016;26:1794–1803.
134. Krieger CC, Boutin A, Jang D, et al. Arrestin-beta-1 physically scaffolds TSH and IGF1 receptors to enable crosstalk. *Endocrinology*. 2019;160:1468–1479.
135. Brix C, Führer D, Biebermann H. Molecules important for thyroid hormone synthesis and action - known facts and future perspectives. *Thyroid Res*. 2011;4:S9.
136. Rigutto C, Sandice H, Grasberger H, et al. Activation of dual oxidases duox1 and duox2 differential regulation mediated by cAMP-dependent protein kinase and protein kinase c-dependent phosphorylation. *J Biol Chem*. 2009;284:6725–6734.
137. Ericson LE. Exocytosis and endocytosis in the thyroid follicle cell. *Mol Cell Endocrinol*. 1981;22:1–24.
138. Endo T, Ohta K, Haraguchi K, et al. Cloning and functional expression of thyrotropin receptor cDNA from fat cells. *J Biol Chem*. 1995;270:10833–10837.
139. Krieger CC, Neumann S, Gershengorn MC. Is there evidence for IGF1RAbs in Graves' orbitopathy pathogenesis? *Intl J Mol Sci*. 2020;21:6561.
140. Baliram R, Latif R, Zaidi M, et al. Expanding the role of thyroid-stimulating hormone in skeletal physiology. *Front in Endocrinology*. 2017;8:252.
141. Peele ME, Carr FE, Baker JR, et al. TSHβ subunit gene expression in human lymphocytes. *Am J Med Sci*. 1993;305:1–7.
142. Balzan S, Nicolini G, Forini F, et al. Presence of a functional TSH receptor on human erythrocytes. *Biomed Pharmacother*. 2007;61:463–467.
143. Crisanti P, Omri B, Hughes EJ, et al. The expression of thyrotropin receptor in the brain. *Endocrinol*. 2001;142:812–822.
144. Prummel MF, Brokken LJS, Meduri G, et al. Expression of the thyroid-stimulating hormone receptor in the folliculo-stellate cells of human anterior pituitary. *J Clin Endocrinol Metab*. 2000;85:4347–4353.
145. Haraguchi K, Shimura H, Kawaguchi A, et al. Effects of thyrotropin on the proliferation and differentiation of cultured rat preadipocytes. *Thyroid*. 1999;9:613–619.
146. Lundbäck V, Kulyté A, Dahlman I, et al. Adipose-specific inactivation of thyroid stimulating hormone receptors in mice modifies body weight, temperature and gene expression in adipocytes. *Physiol Rep*. 2020;8:e14538.
147. Bassett JHD, Williams GR. Role of thyroid hormones in skeletal development and bone maintenance. *Endocr Rev*. 2016;37:135–187.
148. Bassett JHD, Williams AJ, Murphy E, et al. A lack of thyroid hormones rather than excess thyrotropin causes abnormal skeletal development in hypothyroidism. *Mol Endocrinol*. 2008;22:501–512.
149. Boutin A, Neumann S, Gershengorn MC. Multiple transduction pathways mediate thyrotropin receptor signaling in preosteoblast-like cells. *Endocrinology*. 2016;157:2173–2181.
150. Bassett JHD, Williams GR. Critical role of the hypothalamic-pituitary-thyroid axis in bone. *Bone*. 2008;43:418–426.
151. Hase H, Ando T, Eldeiry L, et al. TNFα mediates the skeletal effects of thyroid-stimulating hormone. *Proc Natl Acad Sci U S A*. 2006;103:12849–12854.
152. Abe E, Marians RC, Yu W, et al. TSH is a negative regulator of skeletal remodeling. *Cell*. 2003;115:151–162.
153. Sun L, Vukicevic S, Baliram R, et al. Intermittent recombinant TSH injections prevent ovariectomy-induced bone loss. *Proc Natl Acad Sci*. 2008;105:4289–4294.
154. de Lloyd A, Bursell J, Gregory JW, et al. TSH receptor activation and body composition. *J Endocrinol*. 2010;204:13–20.
155. Williams GR, Bassett JHD. Thyroid diseases and bone health. *J Endocrinol Invest*. 2018;41:99–109.
156. Martini G, Gennari L, De Paola V, et al. The effects of recombinant TSH on bone turnover markers and serum osteoprotegerin and RANKL levels. *Thyroid*. 2008;18:455–460.
157. Mazziotti G, Sorvillo F, Piscopo M, et al. Recombinant human TSH modulates in vivo C-telopeptides of type-1 collagen and bone alkaline phosphatase, but not osteoprotegerin production in postmenopausal women monitored for differentiated thyroid carcinoma. *J Bone Mineral Res*. 2005;20:480–486.
158. Williams GR, Bassett JHD. Thyroid diseases and bone health. *J Endocrinol Invest*. 2018;41:99–109.
159. Smith TJ. TSH-receptor-expressing fibrocytes and thyroid-associated ophthalmopathy. *Nat Rev Endocrinol*. 2015;11:171–181.
160. Smith TJ, Janssen JAMJL. Insulin-like growth factor-i receptor and thyroid-associated ophthalmopathy. *Endocr Rev*. 2019;40:236–267.
161. Krieger CC, Neumann S, Marcus-Samuels B, et al. TSHR/IGF-1R cross-talk, not IGF-1R stimulating antibodies, mediates Graves' ophthalmopathy pathogenesis. *Thyroid*. 2017;27:746–747.

66

Thyroid Hormone Transporters and Metabolism

Erik Schoenmakers and W. Edward Visser

OUTLINE

Thyroid Hormone Transport and Metabolism, 1112
Disorders of Thyroid Hormone Transport, 1116
MCT8 Deficiency, 1116
Disorders of Thyroid Hormone Metabolism, 1119
Iodothyronine Deiodinase I Deficiency, 1119
SBP2 Deficiency, 1120
Sec-tRNA[Ser]Sec, 1124

THYROID HORMONE TRANSPORT AND METABOLISM

The key determinants of intracellular thyroid hormone bioavailability are (1) thyroid hormone transporter proteins at the plasma membrane and (2) thyroid hormone deiodinating enzymes (deiodinases [DIOs]) with their catalytic sites located intracellularly.[1,2] Together, thyroid hormone transporters and DIOs govern intracellular concentrations of the bioactive hormone 3,5,3'-tri-iodothyronine (T3). As such, the genomic actions of T3 are exerted through binding to the nuclear T3 receptor, which functions as a ligand-dependent transcription factor.[3]

The majority of the thyroid hormone transporters that have been identified accept a broad range of substrates, with only a few transporters exhibiting high specificity towards iodothyronines.[1] Table 66.1 displays the main characteristics of human thyroid hormone transporters. The monocarboxylate transporter 8 (MCT8; solute carrier family 16A2 [SLC16A2]) transports 3,5,3',5'-tetra-iodothyronine (thyroxine [T4]), T3 as well as 3,5',3'-tri-iodothyronine T3 (reverse T3 [rT3]) and 3,3'-di-iodothyronine (3,3'-T2).[4] MCT8 is expressed in different tissues, including the developing and adult brain (endothelial cells of the blood-brain barrier [BBB], neuronal cells, choroid plexus), thyroid, liver, kidney, and pituitary. MCT8 is physiologically relevant, as evidenced by the severe clinical sequelae in patients with defective MCT8 (see later section on MCT8 Deficiency). The highly homologous MCT10 (SLC16A10) preferentially transports T3 over T4, but also facilitates transport of aromatic amino acids.[5] MCT10 has a broad tissue expression, including skeletal muscle, kidney, heart, and placenta. The organic anion transporting polypeptide 1C1 (OATP1C1) is also regarded as a T4 transporter.[6] Interestingly, there is a notable species-specific difference in expression: murine *Oatp1c1* is expressed at the BBB, and as such is redundant with Mct8 function; in contrast, human *OATP1C1* is highly expressed in astrocytes, but not in endothelial cells of the BBB. The transmembrane transporter SLC17A4 transfers T3 and T4 as effectively as MCT8[7] and, given its expression along the gastrointestinal tract, one may anticipate that it is involved in intestinal transport of thyroid hormone. While MCT8 and OATP1C1 are essential for thyroid hormone regulation in the brain, the physiological relevance of MCT10 and SLC17A4 is unknown.

There are different pathways of thyroid hormone metabolism (Fig. 66.1), of which deiodination is the most prominent.[2] Therefore, the bioactivity of thyroid hormone is importantly dictated by the three DIOs that facilitate outer-ring deiodination (activation) or inner-ring deiodination (inactivation) of iodothyronines. These DIOs are small (30-kDa) membrane-bound enzymes that function as homodimers, with two distinct domains: an N-terminal transmembrane domain (TMD) and a C-terminal globular cytosolic catalytic domain. The molecular structure of cytosolic DIO3 domain has been solved[26] and shows a thioredoxin-like fold with a selenocysteine-containing active center. The three DIOs are highly homologous, but differ in biochemical characteristics, expression patterns, and regulation (Table 66.2).

DIO1 possesses both outer-ring and inner-ring deiodination activity, although it is mostly effective in degrading rT3. It is present in the plasma membrane and highly expressed in liver, kidneys, and thyroid,[2] with both hepatic and renal DIO1 activity positively regulated by systemic serum thyroid hormone concentration. DIO1 is involved in the clearance of rT3 from the circulation and contributes to the production of T3 from T4 present in the circulation. Sulfated T4 and T3S are rapidly degraded by DIO1, possibly serving as a mechanism to retrieve iodide before the T3 or T4 molecules are removed from the body.

DIO2 catalyzes conversion of T4 to T3 and of rT3 to 3,3'-T2 and is expressed in the brain, pituitary, brown adipose tissue, and skeletal muscle. Together with DIO1, DIO2 is responsible for the production of serum T3 from T4, with DIO2 having the larger contribution.[2] In addition, DIO2 facilitates tissue-specific conversion of T4 to T3 independent of systemic thyroid hormone concentrations. This role of DIO2 allows dynamic control of intracellular hormone availability at different developmental stages and under different conditions.[2]

DIO3 degrades T3 to 3,3'-T2 and T4 to rT3, thereby preventing T4 from being activated. DIO3 is highly expressed in fetal tissues, in uterus and placenta; in adult life, expression is mainly limited to the brain, skin, and β cells of the pancreas. DIO3 can be reactivated under specific pathological conditions, such as severe illness and fasting, with impact on systemic thyroid hormone homeostasis. Cell-specific reactivation in certain conditions, such as myocardial infarction, tissue regeneration, or hypoxia, can alter cell-specific thyroid hormone signaling independent of systemic thyroid hormone concentrations.[2]

Despite considerable differences in thyroid hormone economy between animals and humans, mice defective in one or more DIOs

TABLE 66.1 **Characteristics of Human Thyroid Hormone Transporters**

Transporter	Gene Name	Aliases[1]	Tissue/Cells and Subcellular Localization	Substrate (Km in µM) Uptake	Substrate Efflux	Other Substrates (Km)	Inhibitor
OATP1A2	*SLCO1A2*	OATP-A, OATP, OATP1	Brain, endothelial cells of BBB, kidney (AM of distal nephron), liver (cholangiocytes), lung, ciliary body, placenta (syncytiotrophoblasts)	T3 (~6.5) > T4 (~8.0), rT3; T3S > T4S; rT3S	n.d.	Tauroursodeoxycholate (~19 µM), DHEAS (~7 µM), cholate (~93 µM), taurocholate (~60 µM), cholate (~93 µM), BSP (~20 µM), E3S (~16-59 µM), ouabain, N-methyl quinine, PGE2, glycocholate	BSP, rifampicin, flavanones, polymethoxyflavanones
OATP1B1	*SLCO1B1*	OATP-C, LST-1, OAPT2, OATP6	Liver (BLM of hepatocytes)	T4S, T3S, rT3S > rT3 >T4 (~3.0), T3 (~2.7)	n.d.	BSP (~0.3 µM), cholate (~11 µM), taurocholate (~10-34 µM), glycocholate, E3S (various), DHEAS (~22 µM), PGE2, E217βG (~8.2 µM)	BSP, flavanones, polymethoxyflavanones
OATP1B3	*SLCO1B3*	OATP8, LST-2	Liver (BLM of hepatocytes, central vein>portal vein)	T3 (~6.4) > T4	n.d.	Bilirubin (~39 nM), BSP (~0.4 µM), cholate (~42 µM), taurocholate (~6-42 µM), glycocholate (~43 µM), E3S (~73 µM), methotrexate (~25 µM), DHEAS, ouabain, digoxin, E217βG, PGE2	BSP
OATP1C1	*SLCO1C1*	OATP-F, OATP-RP5 (human), BSAT1 (rat), Oatp2 (mouse), Oatp14	Brain (glial and neuronal (precursor) cells), BBB, choroid plexus, testis (Leydig cells), ciliary body, adipose tissue	T4 (~0.09-0.12) >T4S (~3) > rT3 (~0.13) > T3	n.d.	E3S, BSP, E217βG, taurocholate	BSP
OATP2B1	*SLCO2B1*	OATP-B, OATP-RP2 (human), moat1 (rat), Oatp9	Ubiquitously (including hepatocytes (BLM) syncytiotrophoblasts (BLM), enterocytes (AM, BLM), BBB (AM)	[T4 (~0.31-0.77)]	n.d.	BSP (~0.7 µM), E3S (~6.3 µM), DHEAS	BSP, flavanones, polymethoxyflavanones
OATP3A1 (two protein isoforms: V1 and V2)	*SLCO3A1*	OATP-D, OATP-RP3 (human), Pgt2 (rat), MJAM (mouse), Oatp11	Brain (white/grey matter, choroid plexus), testes, ciliary body, heart, ovary	T4	n.d.	E3S, PGE2 (~0.22-0.37 µM), vasopressin	
OATP4A1	*SLCO4A1*	OATP-E, OATP-RP1 (human), oatpE (rat), Oatp12	Ubiquitous (except gastrointestinal tract and brain)	T3 (~0.9)> rT3>T4	n.d.	Taurocholate (~14.9 µM), PGE2	BSP
OATP4C1	*SLCO4C1*	OATP-H	Kidney	T4, T3 (~0.9)	T3: no T4: n.d.	Digoxin, ouabain, methotrexate	

Continued

TABLE 66.1 Characteristics of Human Thyroid Hormone Transporters—cont'd.

Transporter	Gene Name	Aliases[1]	Tissue/Cells and Subcellular Localization	Substrate (Km in μM) Uptake	Substrate Efflux	Other Substrates (Km)	Inhibitor
NTCP	*SLC10A1*		Liver (basolateral membrane of hepatocytes)	T4S, T3S > T4, T3	No	Eprotirome, cholate, taurocholate (~6-34 μM), glycocholate (~27 μM), chenodeoxycholate-3-sulfate, E3S (~27-60 μM), DHEAS, BSP (~3.7 μM)	Myrcludex B, BSP, furosemide
SLC17A4	*SLC17A4*	NPT homolog	Gastrointestinal tract, liver, pancreas, kidney	T3, T4	n.d.	p-[glycyl-2-3H]p-aminohippuric acid (PAH), (uric acid)	n.d.
LAT1	*SLC7A5*		Brain, placenta, testis, leukocytes, fetal liver, bone marrow	3,3′-T2 > rT3 > T3 > T4; MIT (~13)	3,3′-T2	Broad spectrum of (neutral) amino acids (excluding Gly, Ala, Ser) (~12-120 μM), L-dopa (~34 μM), pregabalin (~288 μM), IMT (22.6 μM)	BCH, JPH203
LAT2	*SLC7A8*		Kidney, placenta, brain, spleen, prostate, testis, ovary	3,3′-T2 > T3; MIT	None	Broad spectrum of (neutral) amino acids (including Gly, Ala, Ser) (~35–200 μM)	BCH, NEM
LAT3	*SLC43A1*		Placenta, liver, kidney, pancreas, skeletal muscle, heart	None	3,3′-T2, MIT, (rT3)	Neutral amino acids (Leu, Ile, Val, Phe > Met, Tyr > others) (~8–30 μM)	BCH, NEM
LAT4	*SLC43A2*		Placenta, leukocytes, skeletal muscle, spleen, kidney and heart	None	3,3′-T2, MIT, (rT3)	Neutral amino acids (Ile, Leu, Met, Phe >>> others) (~100–200 μM)	BCH, NEM
MCT8	*SLC16A2*	XPCT, MCT7	Developing and adult brain: vascular endothelial cells of BBB, choroid plexus, liver, kidney, thyroid, pituitary	T3 (~7.5 μM short isoform, ~1 μM long isoform), T4 (~3 μM) > rT3 > 3,3′-T2 > 3,5-T2 >3′,5′-T2	T3, T4	-	Silychristin
MCT10	*SLC16A10*	TAT1	Kidney, skeletal muscle, placenta, heart, developing brain, hypothalamus, choroid plexus	T3 >3,5-T2 >>T4, rT3	T3	Phe, Tyr, Trp (~400–700 μM), L-dopa (~1.2 mM)	Trp

[1] Ambiguous protein names (similar names used for distinct proteins, tissue-specific names) are indicated in bold.

LST, Liver-specific transporter, *OATP*, organ anion transporting polypeptide; *OATP-RP*, organ anion transporting polypeptide–related protein; *BSAT*, brain-specific anion transporter; *E3S*, estrone-3-sulfate; *PGE2*, prostaglandin E2; *E217βG*, 17β-glucuronosyl estradiol; *IMT*, 3-iodo-α-methyl-L-tyrosine; *NEM*, N-ethylmaleimide; *XPTC*, X-Linked PEST-containing transporter; *TAT*, T-type amino acid transporter; *BLM*, basolateral membrane; *AM*, apical membrane; BSP, sulfobromophthalein; BBB, blood-brain barrier; BCH, 2-aminobicyclo-(2,2,1)-heptane-2-carboxylic acid; DHEAS, dehydroepiandrosterone sulfate; MIT, monoiodotyrosine.

If the transport of different iodothyronines has been studied in parallel, the substrates are listed in order of apparent substrate preference with a , (comma sign) denoting equal efficacy, and > denoting greater efficacy. If the transport of iodothyronines has not been studied in parallel, a ; (semicolon sign) is used, and substrates are not ranked in a specific order.

have been helpful to dissect their role.[8] Table 66.2 provides the main features of Dio knockout (KO) mice models. *Dio1* null mice exhibit raised T4 and rT3, with normal T3 and thyroid-stimulating hormone (TSH) concentrations.[9] The *Dio2* null mice has high T4 and TSH but normal T3 concentrations and shows growth retardation and defective auditory function.[10,11] Murine *Dio3* deficiency is associated with partial embryonic and neonatal lethality, and surviving mice show severe growth retardation, impaired fertility, and central hypothyroidism.[12] The phenotypes of the murine *Dio1* and *Dio2* double KO mice are surprisingly mild, with serum T3 level, general health, and reproductive capacity apparently not affected, and brain function only somewhat impaired.[13] However, the serum rT3 is elevated 6-fold in the *Dio1/Dio2* null mouse, while only 2-fold in the *Dio1* knock-out mouse and not at all in the *Dio2* KO mouse. A mouse model with complete DIO deficiency, the triple *Dio1/Dio2/Dio3* KO, has survival and growth issues that occur mainly in the embryonic and neonatal period, and many of the mice are infertile. However, the healthy surviving *Dio1/Dio2/Dio3* mice have normal growth, can be bred, and have little gross neurological phenotype, and serum thyroid hormone concentrations are unremarkable.[8,13]

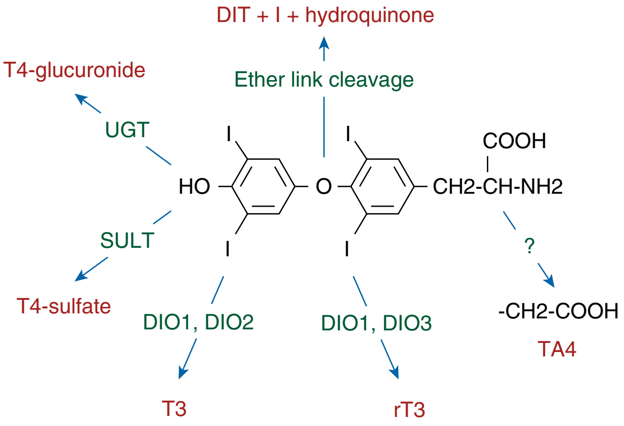


Fig. 66.1 **Pathways of thyroid hormone metabolism.** Deiodination by the three types of deiodinating enzymes (DIO1, DIO2, DIO3) catalyzes removal of iodine moieties. Sulfation and glucuronidation are catalyzed by sulfotransferases (SULTs) and UDP-glucuronosyltransferases (UGTs), respectively. Ether link cleavage and side-chain modification result in alternative products.

The systemic and localized control of thyroid hormone signaling is essential for development, growth, and normal adult life, requiring dynamic regulation to allow for constant adjustment to thyroid hormone signaling according to endogenous and environmental demands. However, the versatility of the system also can lead to the dysregulation of DIO expression, resulting in abnormal systemic or local thyroid hormone concentrations. A striking clinical condition that affects DIO activity is observed in critically ill patients with a substantial drop in circulating T3, at least in part attributed to changed DIO expression and activity in one or more affected tissues.[14,15] Other

TABLE 66.2 Characteristics of Iodothyronine Deiodinases

Characteristics		DIO1	DIO2	DIO3
Function		Plasma T3 production rT3 clearance	Local T3 production	T3 and T4 clearance rT3 production
Reaction catalyzed		T4 to T3, rT3 to T2 T4 to rT3, T3 to T2	T4 to T3 rT3 to T2	T4 to rT3 T3 to T2
Substrate preference		rT3 >>T4=T3	T4 >rT3	T3 >T4
Inhibitor	PTU	++++	+	+/-++
	Aurochioglucose	++++	++	+++
	Iopanoic acid	+++	++++	
Human chromosomal location		1p32-p33	14q24.3	14q32
Tissues		Liver, kidney, thyroid, pituitary, white adipose tissue, brain	Pituitary, brain, brown adipose tissue, skeletal muscle, osteoblast, thyroid, heart	Fetus, brain, skin, uterus, decidual tissue, placenta, pancreatic β cell
Activity in hypothyroidism		↓ Liver, kidney ↑ thyroid	↑ all tissues	↓brain
Activity in hyperthyroidism		↑ Liver, kidney, thyroid	↓ most tissues ↑ thyroid	↑ brain
Phenotypic features of knockout mice		Viable Normal growth Normal fertility ↑ Serum T4 and rT3 in adults Normal serum T3 and TSH in adults Enhanced fecal excretion endogenous iodothyroines	Viable Mild growth delay in males Normal fertility ↑ Serum T4 and TSH Normal serum T3 Pituitary resistance to T4 Impaired thermogenesis Impaired hearing Mild impairment neurocognition Insulin resistance Enhanced susceptibility to diet induced obesity	Increased perinatal mortality Marked growth retardation Impaired fertility Hyperthyroidism in the perinatal period Moderate hypothyroidism in adulthood Impaired responsiveness of the thyroid gland, pituitary gland and hypothalamus Impaired retinal development Impaired hearing

++++: highly sensitive; +++: moderate sensitive; ++: low sensitivity; + very low sensitivity; ↑-↓ increased - decreased levels.
PTU, Propylthiouracil; *TSH*, thyroid-stimulating hormone.

examples include patients with Graves disease[16] and McCune–Albright syndrome,[17] in which higher serum T3 concentrations correlated with higher thyroidal DIO1 and DIO2 activities. Changes in DIO activity are also reported in several neoplasias, with depletion or excess of thyroid hormone promoting modifications in tumoral growth and development.[2] One of the most striking examples is a severe form of hypothyroidism due to high levels of DIO3 activity in vascular tumors such as hepatic hemangiomas.[18] Depending on tumor size and the levels of *DIO3* expression, the DIO3 activity may inactivate circulating thyroid hormones faster than the thyroid gland can secrete, resulting in a condition known as consumptive hypothyroidism.

The knowledge that DIO activity can affect thyroid hormone concentrations dramatically, influencing patient treatment and potentially outcome, raises the possibility that patients can benefit from genome-wide association studies that identify single nucleotide polymorphisms (SNPs) in *DIO* genes associated with serum thyroid hormone concentrations. These studies show that common variations in *DIO1*, but not *DIO2* or *DIO3*, have a moderate-to-strong relationships with systemic thyroid hormone concentrations.[19,20] With regard to identifying SNPs in *DIO3*, their impact may be affected by genetic imprinting, with *DIO3* being preferentially expressed from the paternal allele.[21] There are several studies describing an association between *DIO2* SNPs and clinical syndromes, including hypothyroidism, hypertension, type 2 diabetes, and mental disorders; however, these associations have not been universally reproduced in all population studies.[2,19] This observation confirms the notion that serum thyroid hormone concentrations and related phenotypes are controlled by a multitude of factors, including genetic background, diet, the interplay between gene networks, and environmental factors. While there is increasing evidence that *DIO2* polymorphisms may play a role in dysregulation of thyroid hormone metabolism, further research should be performed to clarify these associations. Several independent studies suggest that *DIO1* polymorphisms are linked with abnormal thyroid hormone metabolism, and further investigation will determine if this should be taken in consideration for treatment of affected patients.[19]

The human *DIO* genes have been subject investigation for a long time; however, the identification of pathogenic genetic defects has remained elusive. Only recently, two unrelated families were identified to be heterozygous for pathogenic variants in *DIO1*[22] (see later section on Iodothyronine Deiodinase I Deficiency).

All three DIOs have the amino acid selenocysteine (Sec) in their catalytic site, providing enhanced substrate affinity and a fast turnover rate for the deiodination reaction. In humans, more than 25 different Sec-containing proteins, referred to a selenoproteins, have been identified. The Sec residue is encoded by a UGA triplet that in functions as a stop codon in nonselenoproteins. Accordingly, recoding of UGA, enabling incorporation of Sec during selenoprotein synthesis, depends on the presence of a SElenoCysteine Insertion Sequence (SECIS) element, a stem-loop structure in the 3′ untranslated region of the mRNA. The SECIS element attracts a number of cofactors, including SECIS-binding protein 2 (SECISBP2 or SBP2) and Sec-tRNA, the selenocysteine tRNA encoded by *TRU-TCA1-1*. Genetic defects in this pathway, e.g., *SECISBP2, SEPSECS,* and *TRU-TCA1-1*, affects generation of most if not all selenoproteins and, consequently, results in a multisystem disorder. Homozygous and compound heterozygous mutations have been described in *SEPSECS* and are linked with a disorder classified as autosomal recessive pontocerebellar hypoplasia type2D (OMIM 613811), also known as progressive cerebellocerebral atrophy.[27-29] The effect of *SEPSECS* mutations on selenoprotein expression and thyroid hormone metabolism has not been studied in detail and suggests that, in contrast to *SECISBP2* and *TRU-TCA1-1* mutations, selenoprotein deficiency is not generalized; it is unclear if thyroid hormone metabolism is affected, and therefore this will not be discussed any further. However, in patients with *SECISBP2* and *TRU-TCA1-1* mutations, decreased levels of active DIOs mediate abnormal thyroid hormone metabolism, with elevated free T4 (FT4), low or low-normal free T3, raised rT3, and normal or slightly high TSH concentrations.[29]

KEY POINTS

- Cellular thyroid hormone regulation is governed at multiple levels: (1) plasma membrane transporters facilitate influx and efflux of thyroid hormones, (2) intracellular metabolism of thyroid hormone is catalyzed by three deiodinating, selenocysteine-containing enzymes, and (3) genomic actions of thyroid hormone are mediated through binding of 3,5,3′-tri-iodothyronine to its nuclear receptor.

DISORDERS OF THYROID HORMONE TRANSPORT

MCT8 Deficiency

Clinical Phenotype. Disrupted thyroid hormone transport due to mutations in *MCT8* (*SLC16A2*) underlies the severe clinical features of MCT8 deficiency. In 2004, mutations in *MCT8* were linked to patients with an X-linked form of intellectual and motor disability and abnormal thyroid function tests.[30,31] Soon afterwards, it was realized that this disorder had been recognized decades before by Allan, Herndon, and Dudley, and ever since was documented as the Allan–Herndon–Dudley syndrome.[32] Over 200 families have been reported, and the estimated prevalence is approximately 1 in 70,000 males.[33,34] Most children are born after an uneventful pregnancy and delivery. The age of presenting symptoms is approximately 4 months, with developmental delay, hypotonia, poor weight gain, and feeding problems most commonly reported. At examination, most patients have a global hypotonia, with the inability to hold the head upright, as well as upper trunk slipping through. Dystonic posturing of limbs starts in the first year of life. Both dystonia and spasticity contribute to hypertonia and exaggerated deep tendon reflexes. As a consequence, scoliosis is present in the large majority of cases. Primitive reflexes (e.g., glabellar reflex) do not disappear over time. Electroencephalogram-proven seizures are present in approximately a quarter of patients. Patients exhibit moderate-to-severe intellectual disability with a pronounced delay in motor and language development (e.g., speech, as defined by the ability to speak one word, is present in less than 10% of cases).

Body weight (corrected for age) shows progressive deterioration over time, with the majority of patients being severely underweight (Fig. 66.2A). Gastroesophageal reflux disease is commonly encountered. Blood pressure is frequently elevated. In the context of being immobilized, tachycardia is present in a large proportion of patients, with the frequent occurrence of premature atrial and ventricular contractions (Fig. 66.2B). Also, atrial fibrillation and nonsustained ventricular tachycardia have been reported in affected children. In addition, second-degree and incomplete right bundle branch blocks have been reported in up to 12% of patients.

The median survival is 35 years, with 30% of patients dying in childhood. The main causes of death are pulmonary infections, aspiration pneumonia, and sudden death. There is a strongly increased risk for death in patients not attaining full head control compared with patients who do attain head control (Fig. 66.2C) and for young children (1–3 years of age) who are underweight compared with patients who are of normal body weight.

Biochemical evaluation reveals abnormal thyroid function tests. Although TSH concentrations are within the normal range, free and total T4 as well as rT3 concentrations are lower than normal in 90%

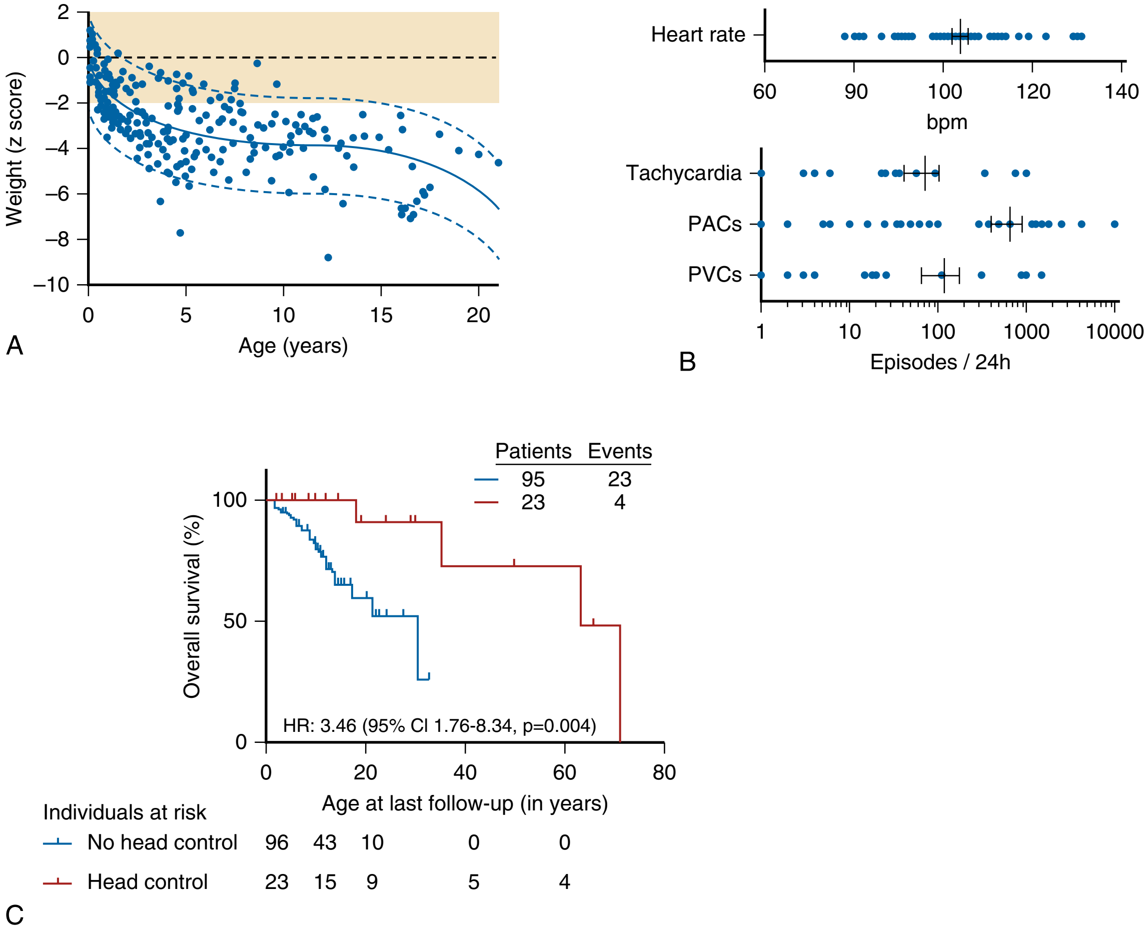


Fig. 66.2 **A,** Natural course of bodyweight change in patients with MCT8 deficiency. Blue dots represent available historical body weight measurements (n = 300) in 86 untreated patients. **B,** Mean (standard error of the mean; *black lines*) occurrence of indicated features during 24-hour cardiac monitoring. *Blue dots* represent the measurements of mean heart rate during 24 hours in 47 patients, or the frequency (per 24 hours) of indicated parameter in 45 patients (premature atrial contractions [PACs] and premature ventricular contractions [PVCs]) or 41 patients (tachycardia) with available data. **C,** Kaplan-Meier estimates of MCT8-specific survival in patients who attained head control *(red line)* by the age of 1.5 years versus those who did not *(blue line)*.

of patients, while total T3 concentrations are above normal in 95% of patients. This results in highly abnormal T3:T4 and T3:rT3 ratios. T4 concentrations obtained from neonatal screening cards were low in the majority of patients, while TSH concentrations were under 15 mU/L in all tested patients.

Serum concentrations of sex hormone–binding globulin were elevated in most patients. Serum alanine aminotransferase, aspartate aminotransferase, and glutamyl transferase concentrations were mildly elevated in a large proportion of patients.

Magnetic resonance imaging (MRI) scans of the brain consistently show a global delay in myelination,[33,35] which improves over age. In addition, diffuse atrophy is present with concomitant dilatation of the ventricles. Magnetic resonance spectroscopy shows an increased choline peak and a decreased N-acetyl aspartate peak, which is compatible with aberrant myelination and general atrophy.

Molecular Genetics. The *SLC16A2* gene is located on the X chromosome and encodes two possible translation start sites, resulting in a long MCT8 isoform of 613 amino acids or a short isoform of 539 amino acids. The relevance of the extended intracellular N-terminal part of the long isoform is unclear. MCT8 consists of 12 TMDs that are organized in two symmetric bundles of six TMDs linked through a large intracellular loop.[1] Approximately half of the mutations are missense mutations, all of which are located in the TMDs. The remainder of the mutations include deletions and nonsense mutations resulting in truncated proteins. An overview of mutations is shown in Fig. 66.3.

For certain mutations (e.g., large deletions, early premature stop codons) it can be inferred that no functional MCT8 is produced. *In vitro* thyroid hormone transport capacity assessment of missense mutations indicates that the majority render the transporter fully inactive, with some mutations showing residual transport activity.[1,36] Although the definitive genotype–phenotype relationship remains to be elucidated, individuals that have MCT8 mutations with residual thyroid hormone transport appear to have a less severe phenotype, including less affected motor and cognitive abilities.

Female carriers have FT4 concentrations between those seen in affected patients and healthy individuals,[37] and the majority appear asymptomatic. However, the occurrence of minor symptoms in female carriers has never been systematically studied and thus remains speculative. MCT8 deficiency mainly occurs in males due to the X-linked inheritance. In rare cases, traits of MCT8 deficiency can be present in females as a result of chromosomal translocations and unfavorable nonrandom X inactivation in female carriers.[38]

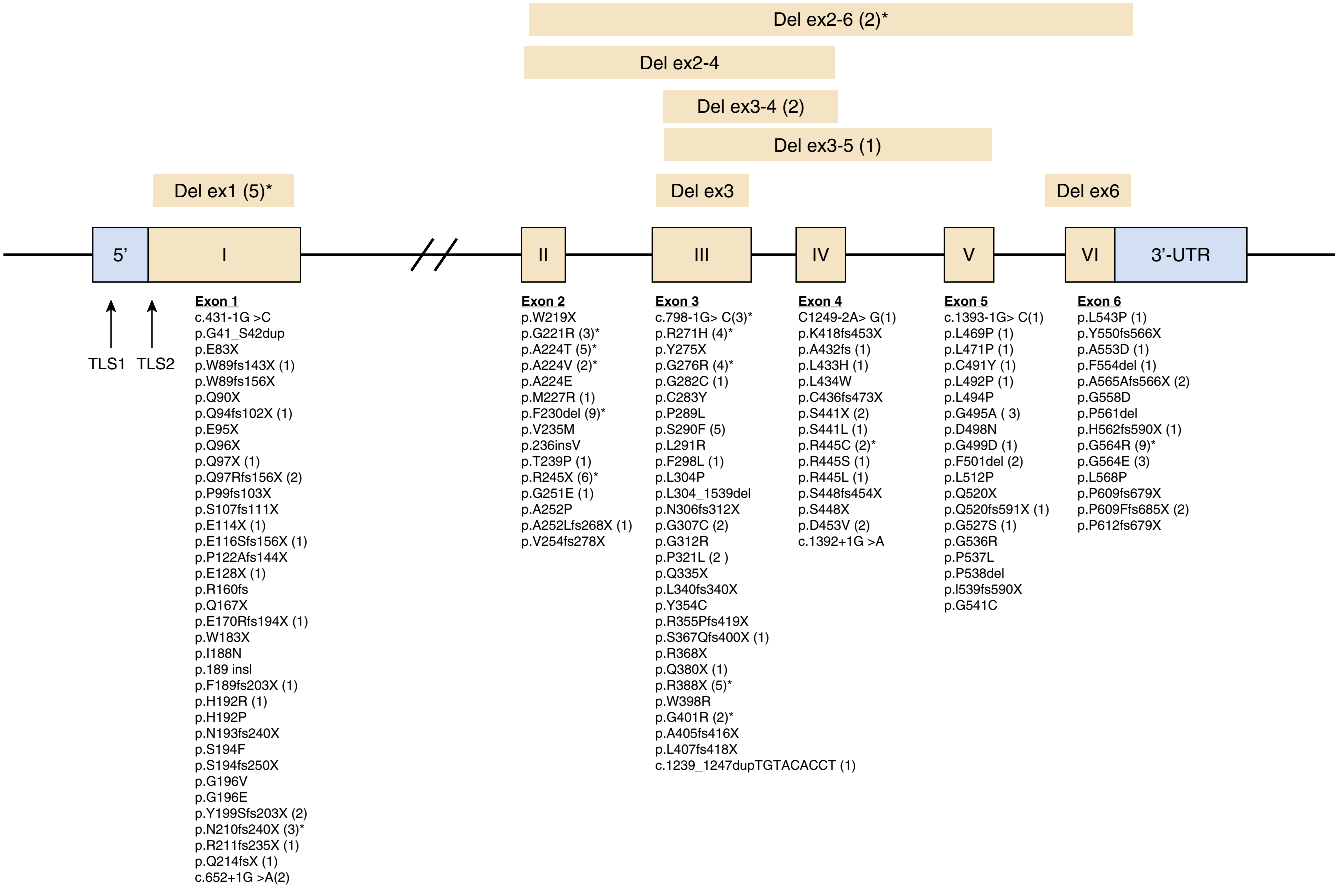

Del ex2-6 (2)*
Del ex2-4
Del ex3-4 (2)
Del ex3-5 (1)
Del ex1 (5)*
Del ex3
Del ex6
5'
I
II
III
IV
V
VI
3'-UTR
TLS1
TLS2
Exon 1
c.431-1G >C
p.G41_S42dup
p.E83X
p.W89fs143X (1)
p.W89fs156X
p.Q90X
p.Q94fs102X (1)
p.E95X
p.Q96X
p.Q97X (1)
p.Q97Rfs156X (2)
p.P99fs103X
p.S107fs111X
p.E114X (1)
p.E116Sfs156X (1)
p.P122Afs144X
p.E128X (1)
p.R160fs
p.Q167X
p.E170Rfs194X (1)
p.W183X
p.I188N
p.189 insI
p.F189fs203X (1)
p.H192R (1)
p.H192P
p.N193fs240X
p.S194F
p.S194fs250X
p.G196V
p.G196E
p.Y199Sfs203X (2)
p.N210fs240X (3)*
p.R211fs235X (1)
p.Q214fsX (1)
c.652+1G >A(2)
Exon 2
p.W219X
p.G221R (3)*
p.A224T (5)*
p.A224V (2)*
p.A224E
p.M227R (1)
p.F230del (9)*
p.V235M
p.236insV
p.T239P (1)
p.R245X (6)*
p.G251E (1)
p.A252P
p.A252Lfs268X (1)
p.V254fs278X
Exon 3
c.798-1G> C(3)*
p.R271H (4)*
p.Y275X
p.G276R (4)*
p.G282C (1)
p.C283Y
p.P289L
p.S290F (5)
p.L291R
p.F298L (1)
p.L304P
p.L304_1539del
p.N306fs312X
p.G307C (2)
p.G312R
p.P321L (2)
p.Q335X
p.L340fs340X
p.Y354C
p.R355Pfs419X
p.S367Qfs400X (1)
p.R368X
p.Q380X (1)
p.R388X (5)*
p.W398R
p.G401R (2)*
p.A405fs416X
p.L407fs418X
c.1239_1247dupTGTACACCT (1)
Exon 4
C1249-2A> G(1)
p.K418fs453X
p.A432fs (1)
p.L433H (1)
p.L434W
p.C436fs473X
p.S441X (2)
p.S441L (1)
p.R445C (2)*
p.R445S (1)
p.R445L (1)
p.S448fs454X
p.S448X
p.D453V (2)
c.1392+1G >A
Exon 5
c.1393-1G> C(1)
p.L469P (1)
p.L471P (1)
p.C491Y (1)
p.L492P (1)
p.L494P
p.G495A (3)
p.D498N
p.G499D (1)
p.F501del (2)
p.L512P
p.Q520X
p.Q520fs591X (1)
p.G527S (1)
p.G536R
p.P537L
p.P538del
p.I539fs590X
p.G541C
Exon 6
p.L543P (1)
p.Y550fs566X
p.A553D (1)
p.F554del (1)
p.A565Afs566X (2)
p.G558D
p.P561del
p.H562fs590X (1)
p.G564R (9)*
p.G564E (3)
p.L568P
p.P609fs679X
p.P609Ffs685X (2)
p.P612fs679X

Fig. 66.3 Genetic pathogenic mutations in *SLC16A2*, the gene encoding MCT8.

Mechanisms of Disease. The mechanisms underlying the phenotypic features have been investigated in different animal models.

Mct8 KO mice have elevated T3, decreased T4, and normal or slightly increased TSH concentrations compared with wild-type mice, replicating human thyroid function tests. Different mechanisms may contribute to the abnormal endocrine profile. First, the increased T3:T4 ratio suggests that thyroid hormone metabolism is changed. Dio1 activity is increased in the liver and kidney of *Mct8* KO mice, and additional inactivation of *Dio1* in *Mct8* KO (*Mct8/Dio1* double KO mice) normalizes the T3:T4 ratio, indicating the prominent role of Dio1 to the disturbed T3:T4 ratio.[39] However, selective inactivation of hepatic Dio1 activity did not normalize thyroid hormone concentrations in *Mct8* KO mice.[40] Second, the thyroid gland may contribute to the abnormal serum thyroid hormones. Intrathyroidal T4 and T3 concentrations are increased in *Mct8* KO mice, and the ability to secrete T4 is reduced.[41,42] Those observations indicate a role of MCT8 in releasing thyroid hormones into the bloodstream, which is compatible with the basolateral expression of MCT8 in thyrocytes. The partial retention of T4 in the thyroid may lead to an enhanced conversion to T3 by thyroidal Dio1 activity. However, athyroid *Pax8/Mct8* KO mice substituted with T4 had similar low T4 concentrations to *Mct8* KO mice, indicating that extrathyroidal mechanisms are involved. Third, T4 accumulates in kidneys of *Mct8* KO mice; however, it is unclear how the absence of MCT8 in the kidney contributes to the endocrine phenotype.

In *Mct8* KO mice, thyrotropin-releasing hormone (TRH) is upregulated in the hypothalamus,[41] possibly reflecting the inability of locally produced T3 to exert its effects on the hypophysiotropic TRH-producing cells. The pituitary is relative insensitive towards thyroid hormone, although the precise mechanisms are unknown.[41,42]

The elevated T3 concentrations in *Mct8* KO mice have adverse effects on energy metabolism, with increased energy expenditure, associated with a lower fat mass and enhanced glucose metabolism in muscle as result.[43] In addition, *Mct8/Oatp1c1* KO mice have episodes of tachycardia and bradycardia, possibly linked to defective input from the central nervous system.[44]

Although *Mct8* KO mice mimic the endocrine phenotype seen in patients, no neurological phenotype was observed. The double *Mct8/Oatp1c1* KO mouse model did, however, result in cerebral hypothyroidism associated with abnormal brain development due to defective transport of T3 and T4 across the BBB as described in patients.[45] Patient-derived induced pluripotent stem cells differentiated towards endothelial cells of the BBB also indicate the importance of MCT8 for the translocation of both T4 and T3 across the BBB.[46] In addition to the expression at the BBB, MCT8 is expressed in other cells of the brain (e.g., neurons, astrocytes, and tanycytes lining the third ventricle).[47] Although the functional role in those cells remains to be elucidated, data indicate that MCT8 has a cell-autonomous role in neurons.[48]

The clinical phenotype is a composite of tissues with different thyroid states. Depending on the expression of MCT8 and other thyroid hormone transporters, tissues either have insufficient thyroid hormone concentrations (e.g., brain) or are exposed to toxic T3 concentrations (e.g., liver and muscle). With MCT8 being expressed at the BBB and in neural cells, defective MCT8 precludes entry of thyroid hormone into the brain, resulting in abnormal cerebral thyroid hormone concentrations. Postmortem histological analysis of brain regions in patients with MCT8 deficiency are consistent with a hypothyroid state.[49] Therefore, given the critical role of thyroid hormone for many processes involved in brain development, absence of MCT8 will result in abnormal neurodevelopment. The strongly elevated serum T3 levels cause thyrotoxic signs and symptoms in tissues that are not dependent on MCT8; the clinical sequelae predispose to significant morbidity and mortality.

Clinical Management. Supportive care is recommended to address all common clinical features. The occurrence of contractures and scoliosis warrants referral to rehabilitation physicians early in life. Physiotherapists should acknowledge the dominance of the movement disorder in childhood compared with a stronger spastic component later in life. Empirical symptomatic treatment can be initiated on demand. Seizures may warrant antiepileptic drugs, and treatment with drugs to alleviate dystonia and drooling (e.g., anticholinergic drugs) can provide relief, but usually to a limited extent. The progressive decline in body weight due to the thyrotoxicosis and difficulties in swallowing, as well as the strong association with mortality, calls for close monitoring of nutritional status by dieticians, with percutaneous endoscopic gastrostomy feeding often required to meet daily calorie requirements. Given the high prevalence of cardiovascular risk factors, referral for cardiac evaluation is indicated.

Levothyroxine replacement therapy has been offered empirically in many patients because of suspected central hypothyroidism (i.e., low serum FT4 concentrations with normal TSH concentrations). Typically, this approach is ineffective in normalizing FT4 concentrations, and rather aggravates the elevated serum T3 concentrations due to immediate conversion of T4 to T3.[1] One case report has been described in which prenatal intraamniotic treatment with a high dosage of levothyroxine improved neurodevelopment.[50]

The combinatory therapy of propylthiouracil (PTU) (but not methimazol) with levothyroxine replacement can improve features of peripheral thyrotoxicosis.[51,52] This strategy can have clinical benefits, although effects on neurodevelopment are not expected. In addition, PTU is associated with a risk of severe hepatotoxicity. Accordingly, PTU is not recommended as therapy for hyperthyroidism, and its use, particularly in children, is discouraged by current guidelines. The unfavorable safety profile of PTU is particularly relevant in the context of the frequent need to use other drugs (e.g., antiepileptic drugs) with hepatotoxic side effects in MCT8-deficient patients. Therefore, risks and benefits counseling is appropriate if this therapy is considered, in particular the lifelong need for PTU.

Effective therapy should aim to alleviate toxic effects of thyroid hormone in peripheral tissues while improving decreased thyroid hormone signaling in brain. Different possibilities have been explored as therapeutic options for MCT8 deficiency. Preclinical studies have investigated restoring MCT8 function through gene therapy and through application of chaperone molecules that have the capacity to rescue misfolded MCT8 mutants.[53,54] The clinical utilization of those potential options needs to be further elucidated.

Thyromimetic molecules that act as thyroid hormone receptor agonist, but whose cellular entry is not dependent on MCT8, have the potential to prevent or even reverse the neurological phenotype in patients with MCT8 deficiency. Also, such T3 analogs can negatively regulate TSH concentrations, thereby reducing endogenous thyroid hormone production and secretion. The T3 analogs Triac (triiodothyroacetic acid), DITPA (diiodothyropropionic acid), and sobetirome and the sobetirome prodrug Sob-AM2 have been investigated in preclinical studies, with varying positive effects on different outcomes.[55-58] Clinical experience with DITPA application in four children with MCT8 deficiency has been reported.[59] Serum T3 and sex hormone–binding globulin concentrations improved in all patients, although no consistent effect on body weight was observed. Triac treatment was evaluated in 46 (pediatric and adult) patients in an international trial.[60] Triac treatment induced a strong and sustained reduction of the elevated T3 concentrations and, consequently, improved biochemical markers of thyroid hormone action in different tissues. Also, sustained improvements in clinically relevant outcomes including body weight for age, heart rate, and blood pressure were observed. Severe underweight and cardiovascular dysfunction are important clinical sequelae of chronic peripheral thyrotoxicosis, causing substantial morbidity and mortality in patients with MCT8 deficiency. Amelioration of the thyrotoxicosis with Triac treatment could benefit patients with MCT8 deficiency, irrespective of their age. Ongoing studies may reveal if Triac administration early in life can improve the neurocognitive phenotype.

KEY POINTS

- Mutations in the X-linked thyroid hormone transporter MCT8 cause profound neurodevelopmental delay due to cerebral hypothyroidism, as well as a wide range of severe clinical sequelae secondary to chronic peripheral thyrotoxicosis. 3,5,3′-tri-iodothyronine analog (e.g., Triac) therapy has the potential to ameliorate key clinical features of this rare disorder.

Organic Anion Transporting Polypeptide 1C1 Deficiency. One individual with a homozygous mutation in *OATP1C1* has been reported.[61] Cognitive and motor function were delayed in childhood, followed by the progressive loss of attained skills, resulting in the absence of verbal communication and spasticity, as well as difficulties in swallowing. Profound cold intolerance was noted, with normal serum thyroid function parameters. (^{18}F)-fluorodeoxyglucose–positron emission tomography scan showed a largely decreased glucose metabolism, and MRI examination of the brain showed progressive atrophy in the cerebral cortex, subcortical white matter, and cerebellum. The homozygous D252N mutation had decreased cellular T4 transport capacity due to impaired plasma membrane localization.

Brain T4 concentrations were reduced in *Oatp1c1* KO mice but did not result in neurological abnormalities, probably due to redundancy of other thyroid hormone transporters, including Mct8.[62] Zebrafish *oatp1c1* KOs showed abnormal locomotor changes, alterations in radial glial cell development, and shorter neuronal axons.[63] Of note, *oatp1c1* KO zebrafish demonstrated a large goiter, which was fully reversed in the presence of Triac, but not by T3 or T4. As OATP1C1 expression is different in human versus in rodents and fish, it is unclear to what extent findings in animal models can be extrapolated to humans.[1] Moreover, it is unclear if the clinical phenotype can be fully attributed to dysregulation of thyroid hormone in the brain. Possibly, based on the OATP1C1 expression in astrocytes, impaired T4 accumulation in astrocytes results in less conversion to T3 by the DIO2 present in those cells. However, with OATP1C1 present in other cell types as well, defective OATP1C1 in such cells may also contribute to the clinical phenotype.

The combination of levothyroxine and Triac resulted in improvement of alertness and also of swallowing.[61] Future identification of more patients with *OATP1C1* mutations will help to define the clinical features and to understand the mechanisms of disease underlying OATP1C1 deficiency.

DISORDERS OF THYROID HORMONE METABOLISM

Iodothyronine Deiodinase I Deficiency

Two unrelated families heterozygous for pathogenic variants in *DIO1* mutations have been identified.[22] In the first family, a 3-year-old girl with Down syndrome, from nonconsanguineous parents of Hispanic origin, had slightly elevated serum TSH and was positive for thyroperoxidase antibodies, often associated with Down syndrome.[23] The second proband, from nonconsanguineous Jewish parents, was also referred to the clinic with resistance to TSH. Genetic analysis identified in each family one rare heterozygous and one predicted pathogenic variant in *DIO1*: N94K and M201I, respectively. Multiple family members were found to be heterozygous for the *DIO1* mutations, and all affected individuals showed no obvious abnormalities other than characteristic thyroid tests with elevated serum rT3 concentrations and rT3/T3 ratios compared with unaffected family members. One family was investigated in more detail and exhibited normal liver enzymes (alanine aminotransferase and aspartate aminotransferase), blood urea nitrogen, and creatinine concentrations. However, total cholesterol concentrations were high/normal to very high in the affected individuals, and need to be considered as a potential parameter for identifying patients.

In vitro activity assays suggest that the mutations affect substrate affinity and enzyme velocity compared with the wild-type enzyme. The underlying mechanism for the abnormal serum thyroid hormone levels in heterozygous *DIO1*-patients is thought to be haploinsufficiency, similar to what is described in two mouse models. Dio1 deficiency, naturally occurring in the C3H/He mice strain, resulted in elevated serum rT3 level and rT3/T3 ratio.[24,25] These results were confirmed in the genetically engineered heterozygous *Dio1*-deficient mice,[22] with the homozygous *Dio1* KO exhibiting raised T4 and rT3 with normal T3 and TSH levels.[9]

Both Asn94 and Met201 are highly conserved amino acids within the DIO family over different species, suggesting an important function. In the absence of a protein structure for DIO1, the effect of these mutations can be assessed using a model generated from the highly homologous DIO3 structure (Fig. 66.4). The N94K and M201I mutations are fairly conservative, both in size and polarity, and their positions are not expected to result in major changes in the overall structure, but may affect specific local function of DIO1. Met201 is located near the substrate-binding and catalytic activity core of the enzyme, with His174, part of the proposed proton transfer chain for reductive 5'-deiodination, in close proximity. Asn94 is situated close to the hinge segment, between the TMD and the globular substrate-binding domain.

KEY POINTS

- The human deiodinase genes have been subject to investigation for a long time; however, only pathogenic genetic defects in DIO1 have recently been described. Thyroid tests show elevated serum 3,5′,3′-tri-iodothyronine concentrations and 3,5′,3′-tri-iodothyronine/3,5,3′-tri-iodothyronine ratio, with a suggestion that total cholesterol concentrations are elevated.

SBP2 Deficiency

Clinical Phenotype. Compound heterozygous or homozygous *SECISBP2* defects have been described in patients from diverse ethnic backgrounds, exhibiting similar clinical phenotypes[29] (Table 66.3). Most cases were diagnosed in childhood with growth retardation (e.g., failure to thrive, short stature, delayed bone age) and developmental delay (e.g., delayed speech, intellectual and motor coordination deficits) as common features that brought probands to clinical attention.

The biochemical phenotype in all affected cases suggests a defect in T4 to T3 conversion, with raised circulating FT4, normal to low free T3 (FT3), raised rT3 concentrations, and normal or slightly high TSH concentrations with no noticeable goiter.[64] This pattern reflects deficiency of all three selenocysteine-containing DIO enzymes, and, consistent with this hypothesis, higher than normal amounts of exogenous T4 are required to suppress TSH levels, whereas the response to T3 is normal.[65] Such abnormal thyroid function, together with low plasma selenium levels, reflecting decreased levels of SELENOP and GPX3, which constitute the major circulating selenoproteins, provides the biochemical signature that facilitates identification of putative cases of SBP2 deficiency due to *SECISBP2* mutation.

Fatigue and muscle weakness, at least partially due to progressive rigid spine muscular dystrophy affecting axial and proximal limb muscles, are recognized features in several patients and are very similar to the phenotype of selenoprotein N–deficient myopathies.[66] Circulating skeletal muscle–specific creatine kinase levels can be elevated, and selective fatty infiltration of the muscle groups (adductor, biceps femoralis) is visible on MRI scan even in childhood, prior to onset of clinical symptoms. Histological analysis shows type 1 oxidative fiber predominance and areas of sarcomere disorganization, termed minicores, in muscle from the *SECISBP2*-mutated adult proband.[67] More marked abnormalities and symptoms in adult patients suggest that the myopathy is progressive and lead to reduced lung function (expiratory and inspiratory flow; total vital capacity), possibly requiring positive pressure ventilatory support.

Mild bilateral high-frequency sensorineural hearing loss is seen in some patients, with adults being most severely affected and sometimes requiring hearing aids, perhaps reflecting a progressive course.

Increased whole-body fat mass, with the excess adipose tissue being predominantly subcutaneous with relatively normal visceral fat, and high circulating adiponectin, possibly of peripubertal onset, have been noted in some cases. Paradoxically, these features are associated with low intrahepatic lipid, a favorable blood lipid profile, and preserved or enhanced tissue insulin sensitivity with possible propensity to spontaneous hypoglycemia.[67]

One adult *SECISBP2*-mutated patient was investigated in great detail, and at presentation (age 35 years) exhibited additional features beside the phenotypes described already, including azoospermia (spermatogenic maturation arrest with preservation of early cell types, e.g., spermatogonia and spermatocytes, but lack of mature spermatids and spermatozoa), skin photosensitivity, severe Raynaud disease (digital vasospasm), rotatory vertigo, and reduced red blood cell and total lymphocyte counts, with impaired mononuclear cell cytokine secretion and T-cell proliferation.[67] Decreased circulating white blood cells and neutrophils were also described in a Japanese patient.[68] An additional characteristic in patients is that many phenotypes (e.g., metabolic phenotype, hearing, muscle weakness, photosensitivity) worsen with advancing age.

Molecular Genetics. Selenium is an essential micronutrient that is incorporated as the amino acid Sec into at least 25 human selenoproteins. Most selenoproteins function as oxidoreductases, with the Sec residue involved in catalytic activity, and recognized functions include maintenance of redox potential, regulating redox sensitive biochemical pathways, protection of genetic material, proteins, and membranes from oxidative damage, metabolism of thyroid hormones, regulation of gene expression, and control of protein folding.[69,70] Biosynthesis of selenoproteins requires an UGA codon within its mRNA to be recoded as the amino acid Sec, preventing its recognition as a premature stop signal. This process is achieved via unique Sec-insertion machinery comprising cis-acting SECIS elements located in the 3′ untranslated region of all selenoprotein mRNAs and the UGA codon, interacting with transacting factors (SECISBP2, Sec tRNA–specific eukaryotic elongation factor and Sec-tRNA$^{[Ser]Sec}$) (Fig. 66.5).[29,70]

From the early stages of selenoprotein research it was clear that dietary selenium intake affects systemic selenium status and selenoprotein expression. However, not all selenoproteins are affected in the same way, with synthesis of housekeeping selenoproteins (e.g., TrxR1, TrxR3, GPx4) being less affected by reduced circulating selenium levels than stress-related selenoproteins (e.g., GPx1, GPx3, SEPW1). Such differential preservation of selenoprotein expression is attributed to the existence of a "hierarchy of selenoprotein synthesis," the underlying molecular basis of which is unclear.[71]

SECISBP2 is an obligate and limiting factor for biosynthesis of selenoproteins and is a large (854 amino acids, 120 kDa) protein. The first 400 N-terminal residues are dispensable for its function *in vitro*, whereas the C-terminal region (amino acids 399–784) is both necessary and sufficient for binding to the SECIS element in the selenoprotein RNA and Sec incorporation. Alternative splicing events and multiple ATG start codons for initiation of translation have been described for SECISBP2, generating different protein isoforms.[72] These events alter content of the dispensable N-terminal region, but not the essential C-terminal domain and are thought to play a role in regulation of SECISBP2-dependent Sec incorporation and the hierarchy of selenoprotein expression *in vivo*.

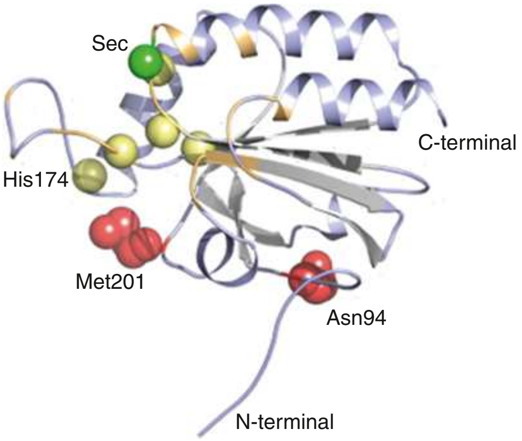


Fig. 66.4 Structural model of the deiodinase 1 with positioning of the mutated amino acids. The model represents the globular, active domain of deiodinase 1 (DIO1). The amino acid selenocysteine is shown in *green*, amino acids described to be involved in proton transfer during reductive 5-deiodination are in *yellow*, those involved in substrate binding and catalytic activity in *orange,* and the two affected amino acids (Met201 and Asn94) are in *red*. The C-terminal and N-terminal ends of the protein are indicated, with the latter linking to the transmembrane domain. The model was generated using the phyre2-server[101] and is based on the crystal structure of DIO3 (protein data bank [PDB]: 4TR3 and 4TR4[26]).

Homozygous or compound heterozygous mutations in *SECISBP2* have been described in 13 individuals from 11 families (Table 66.3). Consistent with a recessive mode of inheritance, heterozygous individuals do not present with any of the described phenotypes. Most *SECISBP2* mutations identified to date are premature stop codons, resulting in absence of full-length SECISBP2 protein. However, for premature stops located in the N-terminal part of the protein, initiation of translation from alternative, downstream ATG codons permits low-level synthesis of shorter SECISBP2 isoforms.[67,73,74] Some premature stop mutations are situated downstream of Met300 and may completely eliminate synthesis of functional protein. Conversely, stop mutations (e.g., R770X, Q782X), distal to the minimal functional domain might generate C-terminally truncated proteins whose functions remain partially intact (Fig. 66.6). In one patient with defective mRNA splicing due to an intronic mutation IVS8ds+29G>A,[64] expression of the correctly spliced transcript was only reduced by 50%, and it is possible that a similar mechanism operates with other splice site mutations, resulting in some preservation of normally spliced SECISBP2 mRNA.

Three missense SECISBP2 mutations (R540Q, E679D, and C691R) are situated in the RNA-binding domain of the minimal functional protein. The R540Q mutation fails to bind only a subset of SECIS-elements, and a mouse model revealed a tissue-specific pattern of SECISBP2 protein expression, correlating with varying loss or preservation of synthesis of different selenoproteins.[64,75,76] The C691R mutation is subject to enhanced proteasomal degradation,[67] confirmed in a mouse model, which also suggests that the C691R mutant is unable to bind RNA and is nonfunctional.[76] The E679D mutation is predicted to be deleterious (PolyPhen-2 algorithm score of 0.998), possibly affecting its binding to RNA, but this has not been investigated.[73]

With only a small number of patients described, *SECISBP2* mutations identified hitherto are uniformly distributed, with no particular mutation "hot spots" in the protein, although most result in a premature stop in the dispensable N-terminal region (Fig. 66.6). All patients described to date have at least one allele that either directs synthesis of SECISBP2 at reduced levels or directs synthesis of a partially functional SECISBP2, due to synthesis of shorter N- or C-terminally truncated

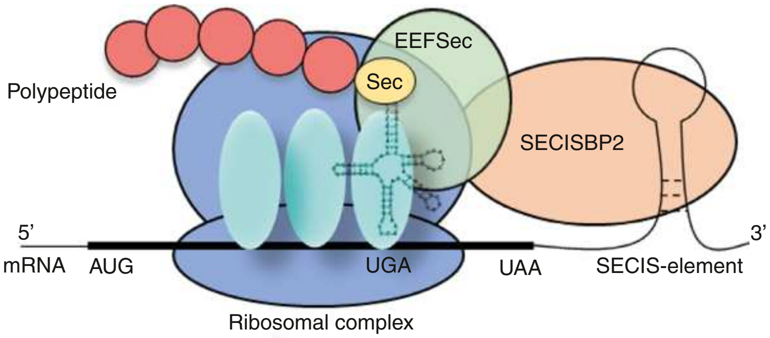


Fig. 66.5 Mechanism of selenoprotein biosynthesis. The 3′ untranslated region of selenoprotein mRNAs contains a stem-loop RNA structure (SElenoCysteine Insertion Sequence element) that interacts with a multiprotein complex which includes SECISBP2 and selenocysteine (Sec)-specific elongation factor. This enables ribosomal recruitment of a specific transfer RNA (tRNA$^{[Ser]Sec}$) to the recoded UGA codon, enabling Sec incorporation into the nascent polypeptide. Failure of this mechanism results in miscoding of the UGA as a stop codon, terminating protein synthesis.[70]

TABLE 66.3 **Human *SECISBP2* Mutations**

Family	Mutation	Protein Change	Alleles Affected	Suggested Mechanism	Ethnicity
A	c.1619 G>A	R540Q	Homozygous	Predicted to affect SECIS and ribosome binding	Saudi Arabian
B	c.1312A>T c.IVS8ds+29 G>A	K438X fs431X	Compound heterozygous	Premature stop, no/decreased full-length protein	Irish/Kenyan
C	c.382 C>T	R128X	Homozygous	Premature stop, no full-length protein	Ghanaian
D	c.358 C>T	R120X	Compound heterozygous	Premature stop, no full-length protein	Brazilian
	c.2308 C>T	R770X		Premature stop, no full-length protein	
E	c.668delT	F223fs255X	Compound heterozygous	Premature stop, no full-length protein	British
	c.IVS7 -155, T>A	fs295X+fs302X		Premature stop, no or decreased full-length protein	
F	c. 2017T>C	C691R	Compound	Predicted to affect SECIS and ribosome binding, increased degradation	British
	1-5 intronic SNPs	fs65X + fs76X	Heterozygous	Premature stop, no full-length protein/splice variants affected	
G	c.1529_1541dup CCAGCGCCCCACT	M515fs563X	Compound heterozygous	Premature stop, no full-length protein	Japanese
	c.235 C>T	Q79X			
H	c.2344 C>T c.2045-2048 delAACA	Q782X K682fs683X	Compound heterozygous	Premature stop, no full-length protein	Turkish
I	c.589 C>T	R197X	Compound heterozygous	Premature stop, no full-length protein/splice variants affected	Argentinian
	c.2108 G> T or C	E679D		Predicted to affect SECIS and ribosome binding	
J	c.800_801insA	K267Kfs*2	Homozygous	Premature stop, no full-length protein	Turkish
K	c.283delT c.589 C>T	T95Ifs31* R197X	Compound heterozygous	Premature stop, no full-length protein	N/A

N/A, Not available; *SNP*, single nucleotide polymorphism; SECIS, SElenoCysteine Insertion Sequence.

forms of the protein (Table 66.3), suggesting that they are hypomorphic rather than completely null for loss of gene function. Nevertheless, because SECISBP2 is rate-limiting for Sec incorporation, reduction in functional SECISBP2 protein compromises selenoprotein synthesis. The small number of patients described from different ethnic and geographical backgrounds, often with compound heterozygous mutations whose functional consequences have not been defined, together with limited knowledge of their phenotypes, makes it impossible to assess the effect of a specific *SECISBP2* mutation or its correlation with the severity of phenotype.

Mechanisms of Disease. Disruption of SECISBP2 function prevents appropriate Sec incorporation into most if not all selenoproteins, resulting in a multisystem disorder. Some clinical phenotypes are attributable to deficiencies of particular selenoproteins, while others may have a complex, multifactorial basis, possibly due to unbalanced oxidoredox and/or protein folding pathways, reflected in increased cellular oxidative stress measured in patients' cells, or unknown function of selenoproteins. SECISBP2 is essential for survival, as murine *Secisbp2* deletion is embryonic lethal.[77] Accordingly, conditional and tissue-specific murine *Secisbp2* or specific selenoprotein KO models have provided informative insights into pathogenesis of the human disorder.

Selenocysteine is present in the active center and required for enzymatic activity of all three iodothyronine DIOs, and, consonant with *SECISBP2* deficiency, DIO2 activity was reduced in fibroblasts from one family.[64] DIO2 converts T4 to T3, and its activity can change very rapidly, as its half-life (approximately 20 minutes) is much shorter (15-fold) than that of DIO1 and DIO3. The combination of reduced DIO2 synthesis, together with its short half-life, likely explains its low activity in patient cells. The thyroid phenotype of *SECISBP2*-mutated patients is most closely recapitulated by a murine *Dio1* and *Dio2* double KO,[13] suggesting that it is mediated by combined, partial deficiency of both DIOs. However, unlike human *SECISBP2* patients, both double *Dio1/Dio2* and triple *Dio1/Dio2/Dio3* KO mice have relative normal serum T3 levels, suggesting that a combination of partial deficiency of all three DIOs and improper fine-tuning of specific DIO expression mediates the abnormal thyroid function test pattern seen in this disorder.

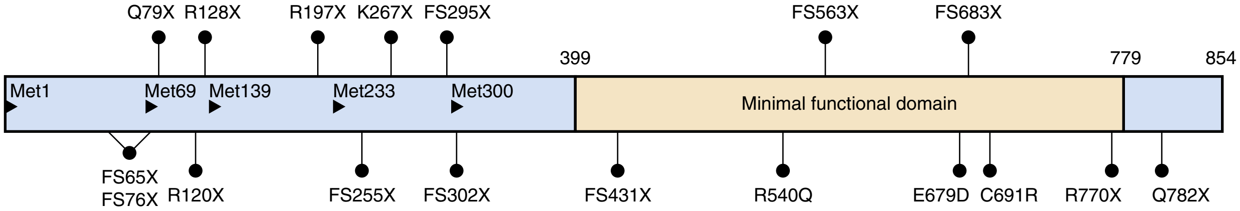


Fig. 66.6 Schematic representation of the SECISBP2 protein with the location of mutations. The position of all mutations is superimposed, and details of mutations in patients are shown in Table 66.3. Neither homozygous or different, compound heterozygous, mutation combinations are predicted to completely eliminate synthesis of the minimal functional domain (amino acids 399–779, *shaded*) of SECISBP2. Internal methionine residues from which translation of shorter SECISBP2 protein isoforms can be initiated are shown.

The thyroid hormone biochemical phenotype seen in patients is recapitulated in an inducible mouse model of *Secisbp2* deficiency, albeit associated with more robust increase in serum TSH levels.[78]

Growth retardation and developmental delay are the commonest symptoms in *SECISBP2*-mutated patients, and the pathogenesis is probably multifactorial, with no clear link to deficiency of a single selenoprotein or pathway. Growth retardation is a recognized feature in *Dio2* and *Dio3* null mice,[11,12] suggesting that the human phenotype is, at least in part, mediated by abnormal thyroid hormone metabolism. Consistent with this notion, treatment of SECISBP2-deficient children with T3 alone (family C, F) or in conjunction with growth hormone (family G) resulted in catch-up growth. Neuron-specific *Secisbp2* KO mice present with growth retardation, with unclear mechanistic basis.[79] Selective depletion of multiple selenoproteins in murine osteochondroprogenitor cells results in epiphyseal abnormalities and chondronecrosis,[80] suggesting an important, cell-autonomous role for selenoproteins in skeletal development. However, in a single *SECISBP2*-mutated case (family E), radiological skeletal survey and bone mineral density were found to be normal.

Axial muscular dystrophy, leading to fatigue and muscle weakness, is a musculoskeletal phenotype highly analogous to that seen in patients with a myopathic disorder (rigid spine muscular dystrophy) due to mutations in selenoprotein N.[81] Motor incoordination in some *SECISBP2*-mutated patients could be due to selenoprotein N myopathy and axial muscle weakness. However, neuron-specific *Secisbp2* null mice[79] also exhibit an obvious locomotor phenotype, together with degenerative loss of PV+/Gad67+ interneurons in basal ganglia, providing a possible alternative explanation for the movement disorder and raising the possibility that this phenotype is progressive.

Azoospermia with spermatogenic maturation arrest described in the adult *SECISBP2*-mutated patient can be linked with deficiency of three selenoproteins (mGPX4, TXNRD3, SELV), with recognized roles in spermatogenesis. Mitochondrial GPX4 (mGPX4) is a structural component of the mitochondrial capsule in the midpiece of spermatozoa,[82] with murine mGpx4 KO resulting in male infertility,[83] and reduced human seminal mGPX4 activity correlates with oligospermia.[84] TXNRD3 is highly enriched in spermatids and may catalyze protein disulfide bond isomerization in sperm development.[85] Human SELV expression is known to be testis-restricted, but its function is unknown.[86] Consistent with the role of mGPX4 and TXNRD3 in latter spermatogenic stages, testicular histology in the adult *SECISBP2*-mutated patient showed maturation arrest, with preservation of early cell types (e.g., spermatogonia and spermatocytes) but lack of mature spermatids and spermatozoa. Male and female infertility are described in a mouse model of *Secisbp2* deficiency.[78]

The bilateral high-frequency sensorineural hearing loss in *SECISBP2*-mutated cases could be due to DIO2 deficiency, with hearing loss and retarded cochlear development being recognized features in *Dio2* null mice,[11] whereas cochlear anatomy is normal and DIO2 activity only partially reduced in human *SECISBP2*-mutated cases. An alternative hypothesis links hearing loss to damage mediated by elevated cellular reactive oxygen species (ROS), similar to the noise-induced hearing loss observed in *Gpx1* null mice.[87] ROS-mediated cochlear damage can be cumulative,[88] which could explain the progressive nature of hearing loss, resulting in a more severe deficit in the older, adult, *SECISBP2*-mutated case.

Several other recorded phenotypes (enhanced insulin sensitivity and increased adipose mass, reduced red blood cell and total lymphocyte counts, cutaneous photosensitivity, developmental delay, growth and skeletal phenotype) probably have a multifactorial basis, comprising a combination of abnormal thyroid hormone metabolism, loss of antioxidant and endoplasmic reticulum stress defense pathways, and deficiency of specific selenoprotein(s), resulting in irreparable cell and organ damage due to cumulative cellular stress. Many studies in mouse models and in humans provide a substantial body of evidence to suggest a link between selenoproteins and most of these phenotypes.[67,89-92]

It is conceivable that cumulative cellular stress (oxidative, endoplasmic reticulum) and irreparable cell damage could also predispose to other phenotypes (e.g., premature ageing, cancer) that have not yet manifested in the relatively young cohort of patients identified hitherto.

Clinical Management. Oral selenium supplementation in some *SECISBP2*-mutated patients raised total serum selenium levels, but without clinical[74,93,94] or biochemical (circulating GPXs, SELENOP, thyroid hormone metabolism) effect.[95] This is probably because the disorder involves defective incorporation of selenium into the selenoproteins, rather than deficiency of this trace element *per se*, although the effect could be mutation-dependent.

Treatment of some probands with T3 alone (T4 was not effective in one case) or in combination with growth hormone resulted in some improvement in growth, development, and bone maturation, suggesting that T4 or T3 treatment at younger age might be beneficial.

Treatment of proband G with a combination of alpha tocopherol (vitamin E) and T3 resulted in the most promising response. Serum levels of products of lipid peroxidation decreased, FT4 and FT3 concentrations ameliorated, and circulating white blood cells and neutrophils increased, all of which were reversed after treatment withdrawal.[68] These observations suggest that treatment with antioxidants, to compensate for loss of antioxidant selenoproteins, is a possible therapeutic option for this disorder. However, due to limited number of *SECISBP2*-mutated cases and a paucity of evidence, clinical trials of antioxidants, perhaps in combination with T3 and growth hormone, are warranted, to develop a rational therapeutic approach in this disorder.

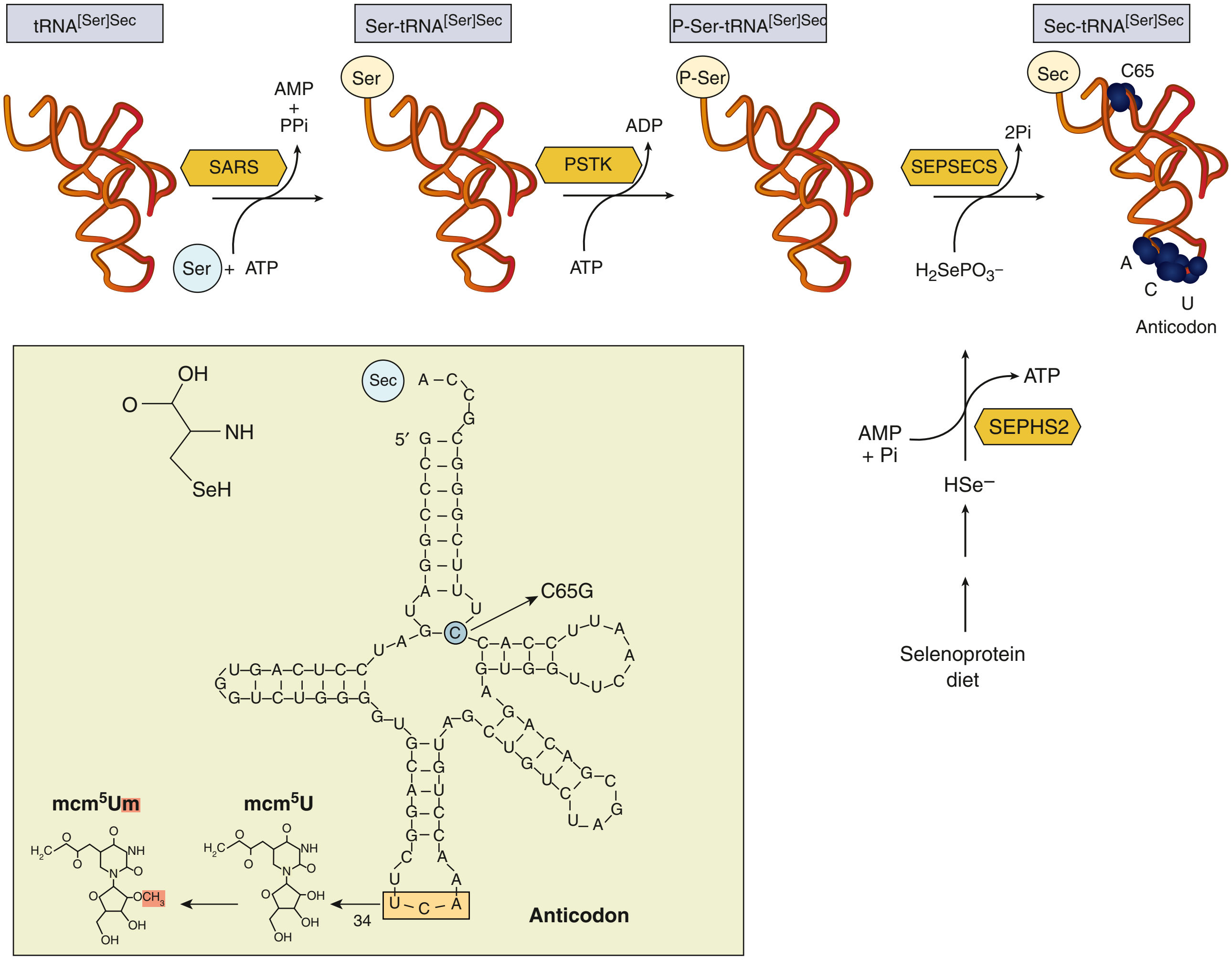


Fig. 66.7 Pathway of Sec-tRNA$^{[Ser]Sec}$ synthesis and clover leaf model of Sec-tRNA$^{[Ser]Sec}$. **A,** Synthesis of selenocysteine (Sec) occurs using its own tRNA, starting with attachment of serine to tRNA$^{[Ser]Sec}$ by seryl-tRNA synthetase (SARS). This residue is phosphorylated by phosphoseryl-tRNA kinase (PSTK), and the phosphoserine is subsequently converted to Sec-tRNA$^{[Ser]Sec}_{mcmU}$ by Sep (O-phosphoserine) tRNA:Sec tRNA synthase (SEPSECS) using selenophosphate. The selenophosphate is synthesized from selenide, originating from diet and degradation of Sec-containing proteins, by selenophosphate synthetase 2 (SEPHS2).[70] The position of the mutated base (C65) and the anticodon bases (U34, C35, and A36) are highlighted as *black spheres* in Sec-tRNA$^{[Ser]Sec}$. **B,** The amino acid Sec is depicted, as well as a clover leaf model of Sec-tRNA$^{[Ser]Sec}$ with the position of the mutated base (C65G, *ringed*) and the modification on uridine 34 in the anticodon *(boxed)* to Sec-tRNA$^{[Ser]Sec}$ $_{mcm5U}$ or Sec-tRNA$^{[Ser]Sec}$ $_{mcm5Um}$ by an unidentified Um34 methylase.[99,100]

Sec-tRNA$^{[Ser]Sec}$

The amino acid Sec is different from other amino acids in that it is synthesized using its own tRNA, tRNA$^{[Ser]Sec}$, encoded by *TRU-TCA1-1*, via a well-described pathway involving SEPSECS (Fig. 66.7).[70] Two major isoforms of the mature Sec-tRNA$^{[Ser]Sec}$ have been identified, containing either 5-methoxycarbonyl-methyluridine (mcm^{5}U) or its methylated form, 5-methoxycarbonylmethyl-2’-O-methyluridine (mcm^{5}Um), at position 34 (Fig. 66.7). Uridine 34 is located in the anticodon loop, and its methylation may contribute to stabilization of the codon–anticodon interaction. These Sec-tRNA$^{[Ser]Sec}$ subtypes have a role in controlling the hierarchy of selenoprotein synthesis, with production of essential, cellular housekeeping selenoproteins (e.g., TXNRDs, GPX4) being dependent on the mcm^{5}U isoform and cellular, stress-related selenoprotein (e.g., GPX1, GPX3) synthesis requiring the mcm^{5}Um isoform.[96,97]

A single patient homozygous for a single nucleotide change (C65G) in *TRU-TCA1-1* has been identified (Fig. 66.7)[98] and exhibits a similar clinical and biochemical phenotype (abdominal pain, fatigue and muscle weakness, raised FT4, normal T3, raised rT3 and TSH, low plasma selenium concentrations) to that seen in patients with *SECISBP2* deficiency. However, the pattern of cellular expression of selenoproteins in the two disorders differs, with synthesis of housekeeping selenoproteins (e.g., TXNRDs, GPX4) being preserved in the *TRU-TCA1-1* case, while levels of cellular stress–related selenoproteins (e.g., GPX1, GPX3) were similarly reduced in both disorders.

In the proband with the *TRU-TCA1-1* C65G mutation, lower total tRNA$^{[Ser]Sec}$ expression with disproportionately greater diminution in Sec-tRNA$^{[Ser]Sec}$ mcm^{5}Um levels were found. The mechanism of this is unresolved, with possibilities including effects of the mutation on stability or posttranscriptional maturation of Sec-tRNA$^{[Ser]Sec}$ or the stability of the Sec-tRNA$^{[Ser]Sec}$–SEPSECS complex. Similar to differential preservation of selenoprotein synthesis seen in murine tRNAsec mutant models,[70,96] it is thought that the low levels of tRNA$^{[Ser]Sec}$ seen in the proband are sufficient for normal synthesis of housekeeping selenoproteins, whereas diminution of Sec-tRNA$^{[Ser]Sec}$ mcm^{5}Um levels accounts for reduced synthesis of stress-related selenoproteins.

Given that systemic selenium status is known to influence relative proportions of the tRNA$^{[Ser]Sec}$ isoforms, with Sec-tRNA$^{[Ser]Sec}$ mcm^{5}Um being low in selenium deficiency but enriched in the selenium-replete state,[96,99,100] it is possible that altering the selenium status of the patient with selenium supplementation to restore particular selenoprotein deficiencies, with careful monitoring for toxicity, could represent a rational therapeutic approach.

KEY POINTS

- Selenium is an essential micronutrient incorporated into at least 25 human selenoproteins as the amino acid selenocysteine (Sec), and SECISBP2 is an obligate and limiting factor for their biosynthesis. Sec incorporation is also dependent on its own transfer RNA (tRNA[Ser]Sec), encoded by TRU-TCA1-1.
- Individuals with SECISBP2 defects exhibit a multisystem phenotype including growth retardation, fatigue and muscle weakness, sensorineural hearing loss, increased whole-body fat mass with enhanced insulin sensitivity, azoospermia, and cutaneous photosensitivity. Characteristic biochemical abnormalities in patients are raised FT4, normal to low FT3, raised rT3, and normal/slightly high thyroid-stimulating hormone, reflecting abnormal thyroid hormone metabolism due to deficiency of all three Sec-containing deiodinase enzymes and low plasma selenium levels due to deficiency of circulating SELENOP and GPX3.
- A single individual with a TRU-TCA1-1 mutation exhibits the same biochemical phenotype and some clinical features as patients with SECISBP2 defects. However, selenoprotein expression in this patient's cells differ, with relative preservation of essential housekeeping versus stress-related selenoproteins.

REFERENCES

1. Groeneweg S, van Geest FS, Peeters RP, et al. Thyroid hormone transporters. *Endocr Rev*. 2020;41:1-55.
2. Bianco AC, Dumitrescu A, Gereben B, et al. Paradigms of dynamic control of thyroid hormone signaling. *Endocr Rev*. 2019;40:1000–1047.
3. Cheng SY, Leonard JL, Davis PJ. Molecular aspects of thyroid hormone actions. *Endocr Rev*. 2010;31:139–170.
4. Friesema EC, Ganguly S, Abdalla A, et al. Identification of monocarboxylate transporter 8 as a specific thyroid hormone transporter. *J Biol Chem*. 2003;278:40128–40135.
5. Friesema EC, Jansen J, Jachtenberg JW, et al. Effective cellular uptake and efflux of thyroid hormone by human monocarboxylate transporter 10. *Mol Endocrinol*. 2008;22:1357–1369.
6. Pizzagalli F, Hagenbuch B, Stieger B, et al. Identification of a novel human organic anion transporting polypeptide as a high affinity thyroxine transporter. *Mol Endocrinol*. 2002;16:2283–2296.
7. Teumer A, Chaker L, Groeneweg S, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat Commun*. 2018;9:4455.
8. Galton VA, de Waard E, Parlow AF, et al. Life without the iodothyronine deiodinases. *Endocrinology*. 2014;155:4081–4087.
9. Schneider MJ, Fiering SN, Thai B, et al. Targeted disruption of the type 1 selenodeiodinase gene (Dio1) results in marked changes in thyroid hormone economy in mice. *Endocrinology*. 2006;147:580–589.
10. Schneider MJ, Fiering SN, Pallud SE, et al. Targeted disruption of the type 2 selenodeiodinase gene (DIO2) results in a phenotype of pituitary resistance to T4. *Mol Endocrinol*. 2001;15:2137–2148.
11. Ng L, Goodyear RJ, Woods CA, et al. Hearing loss and retarded cochlear development in mice lacking type 2 iodothyronine deiodinase. *Proc Natl Acad Sci U S A*. 2004;101:3474–3479.
12. Hernandez A, Martinez ME, Fiering S, et al. Type 3 deiodinase is critical for the maturation and function of the thyroid axis. *J Clin Invest*. 2006;116:476–484.
13. Galton VA, Schneider MJ, Clark AS, et al. Life without thyroxine to 3,5,3'-triiodothyronine conversion: studies in mice devoid of the 5'-deiodinases. *Endocrinology*. 2009;150:2957–2963.
14. Fliers E, Bianco AC, Langouche L, et al. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol*. 2015;3:816–825.
15. Peeters RP, Wouters PJ, Kaptein E, et al. Reduced activation and increased inactivation of thyroid hormone in tissues of critically ill patients. *J Clin Endocrinol Metab*. 2003;88:3202–3211.
16. Ito M, Toyoda N, Nomura E, et al. Type 1 and type 2 iodothyronine deiodinases in the thyroid gland of patients with 3,5,3'-triiodothyronine-predominant Graves' disease. *Eur J Endocrinol*. 2011;164:95–100.
17. Celi FS, Coppotelli G, Chidakel A, et al. The role of type 1 and type 2 5'-deiodinase in the pathophysiology of the 3,5,3'-triiodothyronine toxicosis of McCune-Albright syndrome. *J Clin Endocrinol Metab*. 2008;93:2383–2389.
18. Huang SA, Tu HM, Harney JW, et al. Severe hypothyroidism caused by type 3 iodothyronine deiodinase in infantile hemangiomas. *N Engl J Med*. 2000;343:185–189.
19. Kus A, Chaker L, Teumer A, et al. The genetic basis of thyroid function: novel findings and new approaches. *J Clin Endocrinol Metab*. 2020;105:1707-1721.
20. Medici M, Visser WE, Visser TJ, et al. Genetic determination of the hypothalamic-pituitary-thyroid axis: where do we stand? *Endocr Rev*. 2015;36:214–244.
21. Hernandez A, Fiering S, Martinez E, et al. The gene locus encoding iodothyronine deiodinase type 3 (Dio3) is imprinted in the fetus and expresses antisense transcripts. *Endocrinology*. 2002;143:4483–4486.
22. Franca MM, German A, Fernandes GW, et al. Human type 1 iodothyronine deiodinase (DIO1) mutations cause abnormal thyroid hormone metabolism. *Thyroid*. 2020:202-207.
23. Amr NH. Thyroid disorders in subjects with down syndrome: an update. *Acta Biomed*. 2018;89:132–139.
24. Berry MJ, Grieco D, Taylor BA, et al. Physiological and genetic analyses of inbred mouse strains with a type I iodothyronine 5' deiodinase deficiency. *J Clin Invest*. 1993;92:1517–1528.
25. Schoenmakers CH, Pigmans IG, Poland A, et al. Impairment of the selenoenzyme type I iodothyronine deiodinase in C3H/He mice. *Endocrinology*. 1993;132:357–361.
26. Schweizer U, Schlicker C, Braun D, et al. Crystal structure of mammalian selenocysteine-dependent iodothyronine deiodinase suggests a peroxiredoxin-like catalytic mechanism. *Proc Natl Acad Sci U S A*. 2014;111:10526–10531.
27. Agamy O, Ben Zeev B, Lev D, et al. Mutations disrupting selenocysteine formation cause progressive cerebello-cerebral atrophy. *Am J Hum Genet*. 2010;87:538–544.
28. Ben-Zeev B, Hoffman C, Lev D, et al. Progressive cerebellocerebral atrophy: a new syndrome with microcephaly, mental retardation, and spastic quadriplegia. *J Med Genet*. 2003;40:e96.
29. Schoenmakers E, Chatterjee K. Human disorders affecting the selenocysteine incorporation pathway cause systemic selenoprotein deficiency. *Antioxid Redox Signal*. 2020;33:481–497.
30. Friesema EC, Grueters A, Biebermann H, et al. Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet*. 2004;364:1435–1437.
31. Dumitrescu AM, Liao XH, Best TB, et al. A novel syndrome combining thyroid and neurological abnormalities is associated with mutations in a monocarboxylate transporter gene. *Am J Hum Genet*. 2004;74:168–175.
32. Allan WHC, Dudley FC. Some examples of the inheritance of mental deficiency: apparently sex-linked idiocy and microcephaly. *Am J Mental Defic*. 1944;48:325–334.
33. Groeneweg S, van Geest FS, Abaci A, et al. Disease characteristics of MCT8 deficiency: an international, retrospective, multicentre cohort study. *Lancet Diabetes Endocrinol*. 2020;8:594–605.

34. Visser WE, de Rijke YB, van Toor H, et al. Thyroid status in a large cohort of patients with mental retardation: the TOP-R (Thyroid Origin of Psychomotor Retardation) study. *Clin Endocrinol*. 2011;75:395–401.
35. Vancamp P, Demeneix BA, Remaud S. Monocarboxylate transporter 8 deficiency: delayed or permanent hypomyelination? *Front Endocrinol*. 2020;11:283.
36. Braun D, Schweizer U. Thyroid hormone transport and transporters. *Vitam Horm*. 2018;106:19–44.
37. Friesema EC, Jansen J, Heuer H, et al. Mechanisms of disease: psychomotor retardation and high T3 levels caused by mutations in monocarboxylate transporter 8. *Nat Clin Pract Endocrinol Metab*. 2006;2:512–523.
38. Frints SG, Lenzner S, Bauters M, et al. MCT8 mutation analysis and identification of the first female with Allan-Herndon-Dudley syndrome due to loss of MCT8 expression. *Eur J Hum Genet*. 2008;16:1029–1037.
39. Liao XH, Di Cosmo C, Dumitrescu AM, et al. Distinct roles of deiodinases on the phenotype of Mct8 defect: a comparison of eight different mouse genotypes. *Endocrinology*. 2011;152:1180–1191.
40. Wirth EK, Rijntjes E, Meyer F, et al. High T3, low T4 serum levels in Mct8 deficiency are not caused by increased hepatic conversion through type I deiodinase. *Eur Thyroid J*. 2015;4:87–91.
41. Trajkovic-Arsic M, Muller J, Darras VM, et al. Impact of monocarboxylate transporter-8 deficiency on the hypothalamus-pituitary-thyroid axis in mice. *Endocrinology*. 2010;151:5053–5062.
42. Di Cosmo C, Liao XH, Dumitrescu AM, et al. Mice deficient in MCT8 reveal a mechanism regulating thyroid hormone secretion. *J Clin Invest*. 2010;120:3377–3388.
43. Di Cosmo C, Liao XH, Ye H, et al. Mct8-deficient mice have increased energy expenditure and reduced fat mass that is abrogated by normalization of serum T3 levels. *Endocrinology*. 2013;154:4885–4895.
44. Herrmann B, Harder L, Oelkrug R, et al. Central hypothyroidism impairs heart rate stability and prevents thyroid hormone-induced cardiac hypertrophy and pyrexia. *Thyroid*. 2020;30:1205–1216.
45. Mayerl S, Muller J, Bauer R, et al. Transporters MCT8 and OATP1C1 maintain murine brain thyroid hormone homeostasis. *J Clin Invest*. 2014;124:1987–1999.
46. Vatine GD, Al-Ahmad A, Barriga BK, et al. Modeling psychomotor retardation using iPSCs from MCT8-deficient patients indicates a prominent role for the blood-brain barrier. *Cell Stem Cell*. 2017;20:831–843.e5.
47. Lopez-Espindola D, Garcia-Aldea A, Gomez de la Riva I, et al. Thyroid hormone availability in the human fetal brain: novel entry pathways and role of radial glia. *Brain Struct Funct*. 2019;224:2103–2119.
48. Mayerl S, Heuer H, Ffrench-Constant C. Hippocampal neurogenesis requires cell-autonomous thyroid hormone signaling. *Stem Cell Reports*. 2020;14:845–860.
49. Lopez-Espindola D, Morales-Bastos C, Grijota-Martinez C, et al. Mutations of the thyroid hormone transporter MCT8 cause prenatal brain damage and persistent hypomyelination. *J Clin Endocrinol Metab*. 2014;99:E2799–E2804.
50. Refetoff S, Pappa T, Williams MK, et al. Prenatal treatment of thyroid hormone cell membrane transport defect caused by MCT8 gene mutation. *Thyroid*. 2020:713-720.
51. Wemeau JL, Pigeyre M, Proust-Lemoine E, et al. Beneficial effects of propylthiouracil plus L-thyroxine treatment in a patient with a mutation in MCT8. *J Clin Endocrinol Metab*. 2008;93:2084–2088.
52. Visser WE, Vrijmoeth P, Visser FE, et al. Identification, functional analysis, prevalence and treatment of monocarboxylate transporter 8 (MCT8) mutations in a cohort of adult patients with mental retardation. *Clin Endocrinol (Oxf)*. 2013;78:310–315.
53. Braun D, Schweizer U. Efficient activation of pathogenic DeltaPhe501 mutation in monocarboxylate transporter 8 by chemical and pharmacological chaperones. *Endocrinology*. 2015;156:4720–4730.
54. Iwayama H, Liao XH, Braun L, et al. Adeno associated virus 9-based gene therapy delivers a functional monocarboxylate transporter 8, improving thyroid hormone availability to the brain of Mct8-deficient mice. *Thyroid*. 2016;26:1311–1319.
55. Barez-Lopez S, Hartley MD, Grijota-Martinez C, et al. Sobetirome and its amide prodrug Sob-AM2 exert thyromimetic actions in Mct8-deficient brain. *Thyroid*. 2018;28:1211–1220.
56. Kersseboom S, Horn S, Visser WE, et al. In vitro and mouse studies supporting therapeutic utility of triiodothyroacetic acid in MCT8 deficiency. *Mol Endocrinol*. 2014;28:1961–1970.
57. Di Cosmo C, Liao XH, Dumitrescu AM, et al. A thyroid hormone analog with reduced dependence on the monocarboxylate transporter 8 for tissue transport. *Endocrinology*. 2009;150:4450–4458.
58. Zada D, Tovin A, Lerer-Goldshtein T, et al. Pharmacological treatment and BBB-targeted genetic therapy for MCT8-dependent hypomyelination in zebrafish. *Dis Model Mech*. 2016;9:1339–1348.
59. Verge CF, Konrad D, Cohen M, et al. Diiodothyropropionic acid (DITPA) in the treatment of MCT8 deficiency. *J Clin Endocrinol Metab*. 2012;97:4515–4523.
60. Groeneweg S, Peeters RP, Moran C, et al. Effectiveness and safety of the tri-iodothyronine analogue Triac in children and adults with MCT8 deficiency: an international, single-arm, open-label, phase 2 trial. *Lancet Diabetes Endocrinol*. 2019;7:695–706.
61. Stromme P, Groeneweg S, Lima de Souza EC, et al. Mutated thyroid hormone transporter OATP1C1 associates with severe brain hypometabolism and juvenile neurodegeneration. *Thyroid*. 2018;28:1406–1415.
62. Mayerl S, Visser TJ, Darras VM, et al. Impact of Oatp1c1 deficiency on thyroid hormone metabolism and action in the mouse brain. *Endocrinology*. 2012;153:1528–1537.
63. Admati I, Wasserman-Bartov T, Tovin A, et al. Neural alterations and hyperactivity of the hypothalamic-pituitary-thyroid axis in Oatp1c1 deficiency. *Thyroid*. 2020;30:161–174.
64. Dumitrescu AM, Liao XH, Abdullah MS, et al. Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat Genet*. 2005;37:1247–1252.
65. Dumitrescu AM, Refetoff S. The syndromes of reduced sensitivity to thyroid hormone. *Biochim Biophys Acta*. 2013;1830:3987–4003.
66. Silwal A, Sarkozy A, Scoto M, et al. Selenoprotein N-related myopathy: a retrospective natural history study to guide clinical trials. *Ann Clin Transl Neurol*. 2020;7:2288–2296.
67. Schoenmakers E, Agostini M, Mitchell C, et al. Mutations in the selenocysteine insertion sequence-binding protein 2 gene lead to a multisystem selenoprotein deficiency disorder in humans. *J Clin Invest*. 2010;120:4220–4235.
68. Saito Y, Shichiri M, Hamajima T, et al. Enhancement of lipid peroxidation and its amelioration by vitamin E in a subject with mutations in the SBP2 gene. *J Lipid Res*. 2015;56:2172–2182.
69. Zoidis E, Seremelis I, Kontopoulos N, et al. Selenium-dependent Antioxidant Enzymes: Actions and Properties of Selenoproteins. *Antioxidants (Basel)*. 2018;7 doi:10.3390/antiox7050066.
70. Labunskyy VM, Hatfield DL, Gladyshev VN. Selenoproteins: molecular pathways and physiological roles. *Physiol Rev*. 2014;94:739–777.
71. Sunde RA, Raines AM. Selenium regulation of the selenoprotein and nonselenoprotein transcriptomes in rodents. *Adv Nutr*. 2011;2:138–150.
72. Papp LV, Lu J, Holmgren A, et al. From selenium to selenoproteins: synthesis, identity, and their role in human health. *Antioxid Redox Signal*. 2007;9:775–806.
73. Fu J, Korwutthikulrangsri M, Gonc EN, et al. Clinical and molecular analysis in 2 families with novel compound heterozygous SBP2 (SECISBP2) mutations. *J Clin Endocrinol Metab*. 2020;105:e6-e11.
74. Di Cosmo C, McLellan N, Liao XH, et al. Clinical and molecular characterization of a novel selenocysteine insertion sequence-binding protein 2 (SBP2) gene mutation (R128X). *J Clin Endocrinol Metab*. 2009;94:4003–4009.
75. Bubenik JL, Driscoll DM. Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2. *J Biol Chem*. 2007;282:34653–34662.
76. Zhao W, Bohleber S, Schmidt H, et al. Ribosome profiling of selenoproteins in vivo reveals consequences of pathogenic Secisbp2 missense mutations. *J Biol Chem*. 2019;294:14185–14200.
77. Seeher S, Atassi T, Mahdi Y, et al. Secisbp2 is essential for embryonic development and enhances selenoprotein expression. *Antioxid Redox Signal*. 2014;21:835–849.

78. Fu J, Fujisawa H, Follman B, et al. Thyroid hormone metabolism defects in a mouse model of SBP2 deficiency. *Endocrinology*. 2017;158:4317–4330.
79. Seeher S, Carlson BA, Miniard AC, et al. Impaired selenoprotein expression in brain triggers striatal neuronal loss leading to co-ordination defects in mice. *Biochem J*. 2014;462:67–75.
80. Downey CM, Horton CR, Carlson BA, et al. Osteo-chondroprogenitor-specific deletion of the selenocysteine tRNA gene, Trsp, leads to chondronecrosis and abnormal skeletal development: a putative model for Kashin-Beck disease. *PLoS Genet*. 2009;5:e1000616.
81. Moghadaszadeh B, Petit N, Jaillard C, et al. Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nat Genet*. 2001;29:17–18.
82. Ursini F, Heim S, Kiess M, et al. Dual function of the selenoprotein PHGPx during sperm maturation. *Science*. 1999;285:1393–1396.
83. Schneider M, Forster H, Boersma A, et al. Mitochondrial glutathione peroxidase 4 disruption causes male infertility. *FASEB J*. 2009;23:3233–3242.
84. Foresta C, Flohe L, Garolla A, et al. Male fertility is linked to the selenoprotein phospholipid hydroperoxide glutathione peroxidase. *Biol Reprod*. 2002;67:967–971.
85. Su D, Novoselov SV, Sun QA, et al. Mammalian selenoprotein thioredoxin-glutathione reductase. Roles in disulfide bond formation and sperm maturation. *J Biol Chem*. 2005;280:26491–26498.
86. Kryukov GV, Castellano S, Novoselov SV, et al. Characterization of mammalian selenoproteomes. *Science*. 2003;300:1439–1443.
87. McFadden SL, Ohlemiller KK, Ding D, et al. The influence of superoxide dismutase and glutathione peroxidase deficiencies on noise-induced hearing loss in mice. *Noise Health*. 2001;3:49–64.
88. Riva C, Donadieu E, Magnan J, et al. Age-related hearing loss in CD/1 mice is associated to ROS formation and HIF target proteins up-regulation in the cochlea. *Exp Gerontol*. 2007;42:327–336.
89. Liao C, Carlson BA, Paulson RF, et al. The intricate role of selenium and selenoproteins in erythropoiesis. *Free Radic Biol Med*. 2018;127:165–171.
90. Verma S, Hoffmann FW, Kumar M, et al. Selenoprotein K knockout mice exhibit deficient calcium flux in immune cells and impaired immune responses. *J Immunol*. 2011;186:2127–2137.
91. Sengupta A, Lichti UF, Carlson BA, et al. Selenoproteins are essential for proper keratinocyte function and skin development. *PLoS One*. 2010;5:e12249.
92. Schweikert K, Gafner F, Dell'Acqua G. A bioactive complex to protect proteins from UV-induced oxidation in human epidermis. *Int J Cosmet Sci*. 2010;32:29–34.
93. Azevedo MF, Barra GB, Naves LA, et al. Selenoprotein-related disease in a young girl caused by nonsense mutations in the SBP2 gene. *J Clin Endocrinol Metab*. 2010;95:4066–4071.
94. Catli G, Fujisawa H, Kirbiyik O, et al. A novel homozygous selenocysteine insertion sequence binding protein 2 (SECISBP2, SBP2) gene mutation in a Turkish boy. *Thyroid*. 2018;28:1221–1223.
95. Schomburg L, Dumitrescu AM, Liao XH, et al. Selenium supplementation fails to correct the selenoprotein synthesis defect in subjects with SBP2 gene mutations. *Thyroid*. 2009;19:277–281.
96. Carlson BA, Yoo MH, Tsuji PA, et al. Mouse models targeting selenocysteine tRNA expression for elucidating the role of selenoproteins in health and development. *Molecules*. 2009;14:3509–3527.
97. Shetty SP, Copeland PR. Selenocysteine incorporation: a trump card in the game of mRNA decay. *Biochimie*. 2015;114:97–101.
98. Schoenmakers E, Carlson B, Agostini M, et al. Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *J Clin Invest*. 2016;126:992–996.
99. Hatfield D, Lee BJ, Hampton L, et al. Selenium induces changes in the selenocysteine tRNA[Ser]Sec population in mammalian cells. *Nucleic Acids Res*. 1991;19:939–943.
100. Diamond AM, Choi IS, Crain PF, et al. Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *J Biol Chem*. 1993;268:14215–14223.
101. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009;4:363–371.

67

Thyroid Hormone Action

Gregory A. Brent and Anthony N. Hollenberg

OUTLINE

Components of Thyroid Hormone Action, 1128
Regulation of Ligand Availability in Development and the Adult, 1128
Thyroid Hormone Binding to Serum Proteins, 1129
Local Regulation of Thyroid Hormone Availability, 1129
Thyroid Hormone Receptors and Their Response Elements, 1130
  Thyroid Hormone Receptor Coregulatory Proteins, 1132
Positive And Negative Gene Regulation By Thyroid Hormone, 1134
  Thyroid Hormone Receptor Posttranslational Modifications, 1135
Nongenomic Thyroid Hormone Action, 1135
Thyroid Hormone Tissue-Specific Actions, 1135
  Thyroid Hormone Receptor Isoform–Specific Expression, 1135
  Growth, 1136
  Stem Cell Proliferation, 1136
  Thyroid Hormone and Somatic Growth, 1136
  Brain, 1136
  Heart and Blood Vessels, 1137
  Skeletal Muscle, 1137
  Skeletal System, 1137
  Small Intestine, 1137
  Metabolic Regulation and Actions in Liver, 1137
  White and Brown Adipose Tissue, 1138
  Skin, 1138
Lessons from Resistance to Thyroid Hormone Receptor Beta and Alpha, 1138
  Resistance to Thyroid Hormone Beta, 1139
  Resistance to Thyroid Hormone Alpha, 1139
Summary and Future Directions, 1139

COMPONENTS OF THYROID HORMONE ACTION

Thyroid hormone (TH) action is mediated by the two principal THs, thyroxine (T4) and triidothyronine (T3), that are synthesized within thyroid follicular cells such that the human adult thyroid synthesizes and releases THs in a ratio of approximately 85% T4 and 15% T3. TH signaling is regulated at multiple levels (Fig. 67.1). Circulating serum binding proteins deliver T4 and T3 to tissues, and binding proteins help to ensure consistent hormone delivery throughout the body. Higher hormone levels are generally correlated with increased TH response. However, there are differences in availability of hormone in peripheral tissues that are the consequence of variations in cellular transport and local conversions of T4 to T3 by the deiodinase enzymes.[4] In some tissues, as best characterized in the brain, this tissue uptake is regulated by specific TH transporters.[2,3] TH receptors (TRs) mediate the majority of TH signaling by nuclear actions and regulate gene expression. Tissue sensitivity is further adjusted by expression of coregulatory proteins that interact with TRs. The relative activity of these pathways, and the variation by tissue type, means that the correlation between plasma hormone levels and response is not absolute, consistent with the variety of clinical manifestations of thyroid disease. All components of TH signaling will be described, as well as the associated defects that have been reported with the various TH signaling pathways.

KEY POINTS

- Thyroid hormone signaling is influenced by ligand availability, tissue uptake, nuclear receptor isoform type, and interaction with corepressors and coactivators.

REGULATION OF LIGAND AVAILABILITY IN DEVELOPMENT AND THE ADULT

TH is required for normal brain and somatic development, and regulation of ligand availability occurs at multiple levels. While TH synthesis is covered in another chapter, it is important to note that the fetal thyroid (in particular the follicular cells) develops from anterior endoderm and is functional in context of TH synthesis in the late stages of the first trimester.[5] Until that time, the developing fetus is dependent upon maternal THs, which cross the placenta in small amounts. The importance of the maternal transfer of TH for development is best delineated by syndromes of congenital hypothyroidism, where TH synthesis is not possible.[6,7] As long as TH therapy is begun at birth, fetal development, including neurologic development, is normal. In contrast to the ligand deficiency, which is compensated by maternal TH, fetuses with rare defects in the TH transporter (monocarboxylate transporter 8 [MCT8]) are born with significant irreversible neurologic deficits due to an inability of maternal or fetal T4 to impact the brain at critical stages of development.[8]

---

*Portions of this chapter were taken from an earlier edition, authored by Dr. Fredric E. Wondisford, and these contributions are acknowledged.

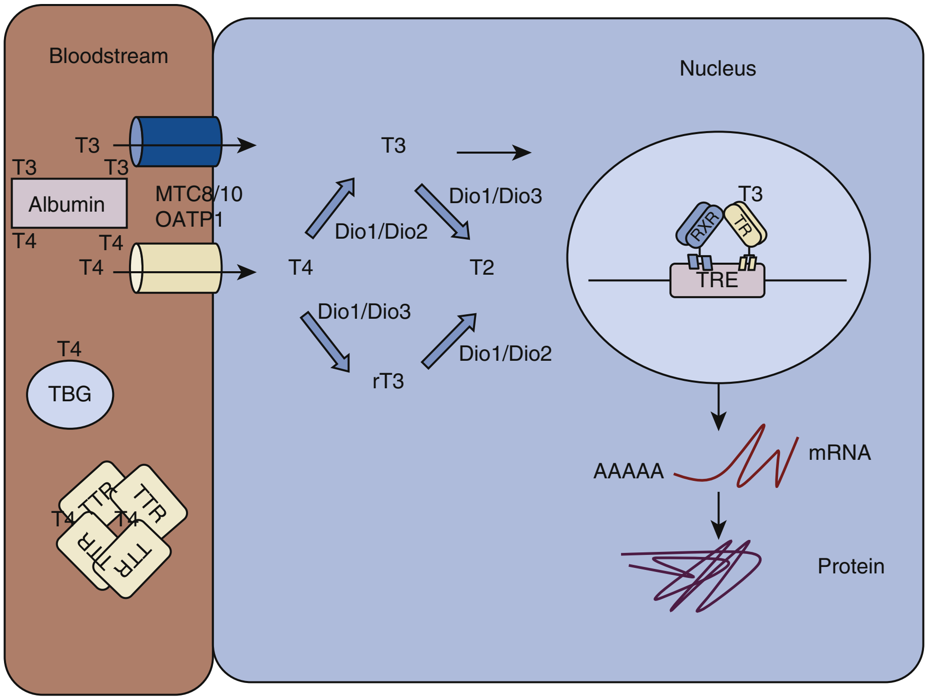


Fig. 67.1 **Triiodothyronine (T3) modulates gene expression in virtually every vertebrate.** Thyroxine (T4)/T3 circulates attached to serum proteins, including thyroxine-binding globulin (TBG), transthyretin (TTR), and albumin. A small fraction of circulating T4 is free to be transported into the cytoplasm, where it is activated to T3 by outer-ring deiodination catalyzed by type 1 deiodinase (Dio1) or type 2 deiodinase (Dio2). The resulting T3 is thought to be diffused into the nucleus to bind the thyroid hormone receptors (TRs). Upon binding to the TRs, T3 modulates the rate of mRNA synthesis, and ultimately the protein levels of thousands of genes in virtually every cell. *RXR*, Retinoid X receptor; *TRE*, thyroid hormone response element. (From Mendoza A, Hollenberg AN. New insights into thyroid hormone action. *Pharmacol Ther*. 2017;173:135–145.)

THYROID HORMONE BINDING TO SERUM PROTEINS

TH circulates predominantly bound to serum protein, which provides a significant reservoir for distribution to target tissues (Fig. 67.1). Both major forms of TH (T4 and T3) are transported in the circulation in complex with plasma proteins.[1] The ratio of total T4 to T3 in plasma is approximately 60:1. This is higher than the 15-fold ratio of T4 to T3 that is initially secreted by the thyroid gland, because of greater plasma binding of T4 versus T3, resulting in greater clearance rates of T3. Approximately 99.98% of total circulating T4 and 99.7% of T3 form noncovalent interactions with serum proteins: mostly thyroxine-binding globulin, but to a lesser extent transthyretin (also referred to as thyroxine-binding prealbumin) and albumin. The fact that less T3 circulates in complex with plasma proteins means that the ratio of free T4 to T3 is around 4-fold, with typical circulating free hormone levels around 20 picomolar (pM) and 6 pM, respectively. The free fraction is considered biologically active and can enter target tissues.

LOCAL REGULATION OF THYROID HORMONE AVAILABILITY

As outlined, thyroid follicular cells synthesize primarily T4 but also some T3. Critical, however, to the development of certain tissues is the amount of locally available T3, which is the primary driver of the physiologic response to the THs due to its genomic signaling capabilities. The regulation of local cellular ligand availability is mediated by the availability of circulating of THs in conjunction with: 1) cellular transporters that allow for the entry of T4 and T3 into cells, and 2) deiodinases that determine the amount of intracellular T3 made available for genomic and nongenomic signaling (Fig. 67.1).

The importance of specific cell surface TH transporters came to light in 2003 when MCT8 was identified, in cloning experiments performed with *Xenopus* oocytes, as being able to transport T4 and T3 across cell membranes through a channel specific to the iodothyronines.[9] Soon after the identification of the MCT8, its relevance in humans was quickly identified based upon the fact that relevant mutations that impaired its function were discovered in patients with the Allan–Herndon–Dudley syndrome (AHDS), a rare X-linked syndrome leading to a severe intellectual and motor delay syndrome in affected males.[10,11] Indeed, besides the severe clinical syndrome often resulting in early mortality in this syndrome, AHDS patients had long been characterized by the presence of relatively normal thyroid-stimulating hormone (TSH) with a low T4 and high T3. Extensive work on the function of the MCT8 suggests that the syndrome is the result of severe neurologic hypothyroidism mediated by the inability of THs to cross the blood–brain barrier and neurons *in utero*, leading to the clinical syndrome after birth. The peripheral TH abnormalities are at least partially explained by a central resistance to feedback at the level of the hypothalamus and pituitary in the regulation of TSH production, coupled with the requirement of the MCT8 for the movement of T4 out of the thyroid follicular cell via the basolateral membrane into the circulation, leading to enhanced thyroidal T3 production and release. In addition, the consumption of available T4 by the type 1 deiodinase (Dio1; see later) in the liver may also play a role in its low levels and the elevated T3 levels, and there may also be a role for enhanced renal metabolism in AHDS patients.[12] Importantly, if TH analog therapy is used in patients with AHDS, the circulating levels of T3 fall with proper regulation of central TSH production.[13]

While MCT8 is the best-characterized TH transporter, other family members exist, and their cell-specific expression can certainly help to determine cellular TH levels. Currently, many other TH transporters have been identified, including monocarboxlate transporter 10 (MCT10), organic anion transporting polypeptide 1C1 (OATP1C1), and SLC17A4. Interestingly, a mutation in OATP1C1 has been identified in a teenager with neurocognitive and motor deficits. Its further importance as a transporter became apparent in mouse studies, where MCT8 knockout mice were found to have a normal neurologic phenotype likely because of the presence of OATP1C1 that allowed for TH transport into the brain.[13a] Importantly, OATP1C1 does not appear to play a developmental role in TH transport in humans. Significantly, MCT8/OATP1C1 double knockout mice develop severe neurologic hypothyroidism akin to that seen in AHDS.[14] In addition to their role as transporters of T4 and T3 into cells, it is likely that TH transporters also play a role in the efflux of THs from cells, and this in particular may be the role of MCT10. Much work remains to be done in order to elucidate the cell-specific mechanisms of TH transport.

Once TH reaches a particular cell type, the resulting T3 availability for signaling is mediated by the deiodinase family. Also, as reviewed in another chapter, the deiodinases can be classified into activating and deactivating subtypes in context of their ability to produce T3. While Dio1, which is preferentially expressed in liver, can metabolize T4 to either T3 or the inactive rT3, its action is primarily to produce T3. Similarly, the type 2 deiodinase (Dio2) plays a significant role in T3 production in many tissues, including neurons, the cochlea, and skeletal muscle (Fig. 67.1). In contrast, the type 3 deiodinase (Dio3) inactivates T4 to rT3 and T3 to T2 in a variety of tissues, especially developmentally, but is also expressed ectopically in the syndrome of infantile hepatic hemangiomatosis, leading to “consumptive hypothyroidism” and demonstrating its role. Certainly the fine tuning of expression of intracellular deiodinases play a significant role in T3 availability for signaling.[15]

THYROID HORMONE RECEPTORS AND THEIR RESPONSE ELEMENTS

The majority of the actions of the THs are mediated by the interactions of T3 with its cognate nuclear receptors, the thyroid hormone receptor (TR) isoforms (Fig. 67.2). While the bulk of T3 action is mediated by the genomic actions of the TRs, evidence is certainly accumulating that there are fast-acting nongenomic actions of T3 that are mediated by the TRs via their actions in the cytoplasm. Evidence of the nuclear actions of T3 first came into view in the 1960s, when Tata demonstrated the significant action of T3 on RNA synthesis in the liver and the requirement for RNA synthesis for the metabolic effects of T3.[16] Subsequent work by Oppenheimer and Samuels in the 1970s demonstrated specific nuclear binding sites for T3 in rat liver, paving the way for the isolation and cloning of the TRs. Interestingly, the pioneering biochemical work in the 1960s and through the 1980s played little role in the first identification of a functional TH receptor. In 1985, the glucocorticoid receptor was the first nuclear receptor identified, and its sequence seemed analogous to the v-erb-A gene of the avian erythroblastosis virus.[17] Subsequently, the Vennstrom laboratory identified human versions of v-erb-A, termed c-erb-A, which is now known as thyroid receptor α, or *Thra,* in humans.[18] At the same time, the Evans laboratory identified a homologous receptor that is now known as the thyroid receptor β, or *Thrb,* in humans.[19] Both receptors were shown at the time to be nuclear and to bind T3 with high affinity. Importantly, the functional concept of similar receptors engaging with a variety of steroid and THs brought about a new era in physiology and therapeutics that continues through this day based on the ability to develop compounds that acts as antagonists or agonists to these nuclear receptors.

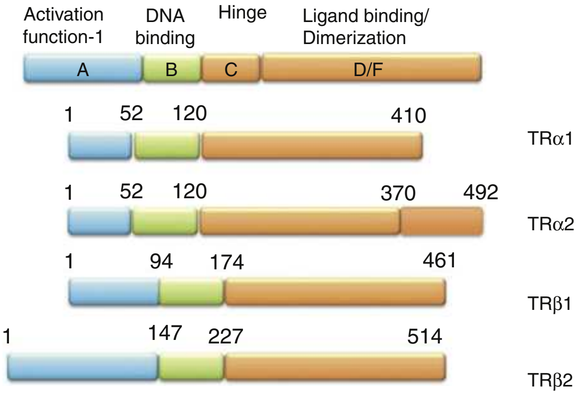


Fig. 67.2 Schematic representation of the thyroid hormone receptor isoforms. Thyroid hormone receptors (TRs) are encoded by the *THRA* and *THRB* genes, which produce multiple isoforms. Depicted are the major isoforms that modulated the actions of triiodothyronine (T3), with exception of TRα2, which does not bind T3 but acts as negative dominant by competing for the thyroid hormone response element with other TRs. TR isoforms share high sequence homology within their functional domains. (From Mendoza A, Hollenberg AN. New insights into thyroid hormone action. *Pharmacol Ther.* 2017;173:135–145.)

Soon after the identification of the original TR isoforms it subsequently became clear that both human *Thra* and *Thrb* were subject to splicing and to the creation of different isoforms (Fig. 67.2). The TRβ gene, located on chromosome 3, produces two major isoforms termed TRβ1 and TRβ2 that differ structurally in their N-termini. In contrast, the TRα gene undergoes splicing in its C-terminus to generate a nonbinding TH receptor isoform termed TRα2, while the principal TH receptor produced from this gene is TRα1. Other isoforms of the TRs have been detected in rodents but are not found or are not functional in humans.[20,21]

Importantly, the tissue of expression more than likely determines the role of the TH receptor isoform, rather than the isoform itself. However, the cell-specific expression of each of the TR isoforms has been difficult to delineate because of their relatively low levels of expression, and the ability of isoform-specific antibodies to identify *in vivo* expression is not strong. Recent genetic approaches to tagging the endogenous alleles in mice has been successful and will yield important information in the future.[22,23] Most of our knowledge of the actions of individual isoforms comes from pioneering mouse genetic studies coupled with the human syndrome of resistance to TH (RTH). TRβ1 is specifically expressed in the liver, where it is a key regulator of cholesterol metabolism and hepatic fat content. Indeed, TRβ1 has long been sought as and continues to be a drug target for the treatment of metabolic diseases, including hypercholesterolemia and nonalcoholic fatty liver disease and its complications. TRβ2 functions specifically in the hypothalamus and the pituitary to regulate the central components of the thyroid axis, including thyrotropin-releasing hormone (TRH) expression in the paraventricular nucleus of the hypothalamus and TSH α and β subunit expression in the pituitary (Fig. 67.3). Interestingly, the exact mechanism by which TRβ2 regulates TRH remains unclear. It is presumed that TRβ2 is expressed in the same neurons as TRH, but because of the lack of reagents available for detecting cell-specific TRβ2, this remains unclear and is an area of important investigation. In addition to its role centrally, TRβ2 plays a key role in the retina, where it controls the proper differentiation of rods and cones to allow for color vision. In contrast to their separate roles, TRβ1 and TRβ2

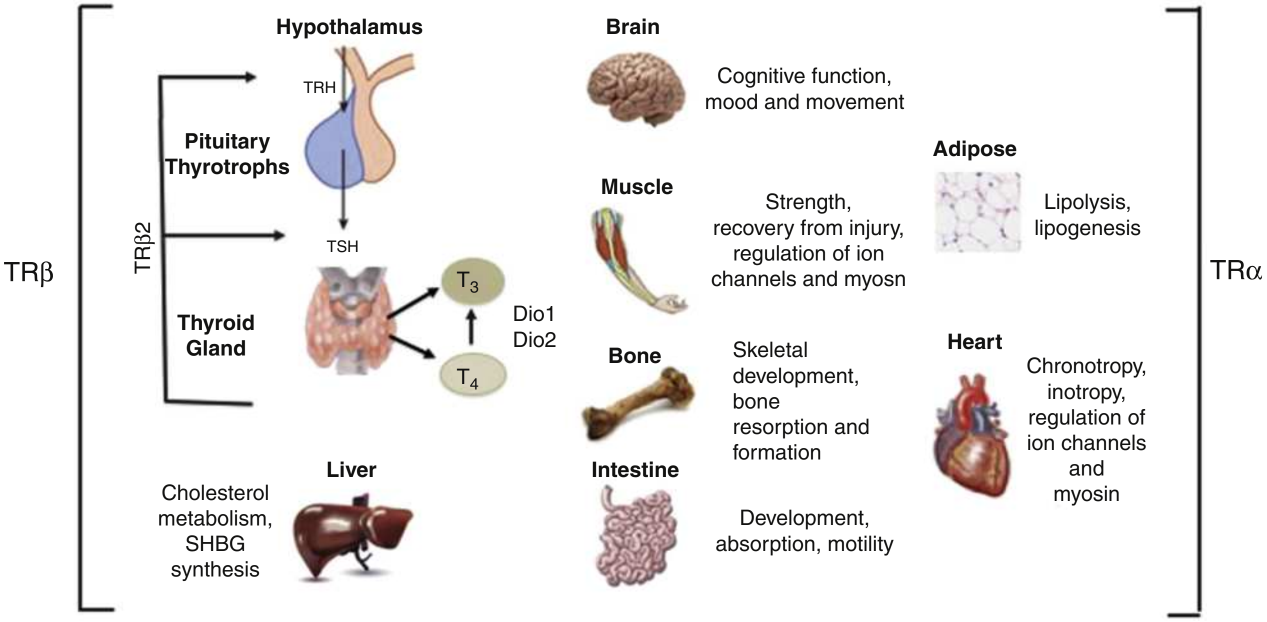

TRβ
TRβ2
Hypothalamus
TRH
Pituitary Thyrotrophs
TSH
Thyroid Gland
T_3
T_4
Dio1
Dio2
Liver
Cholesterol metabolism, SHBG synthesis
Brain
Cognitive function, mood and movement
Muscle
Strength, recovery from injury, regulation of ion channels and myosn
Bone
Skeletal development, bone resorption and formation
Intestine
Development, absorption, motility
Adipose
Lipolysis, lipogenesis
Heart
Chronotropy, inotropy, regulation of ion channels and myosin
TRα

Fig. 67.3 **Thyroid hormone receptor (TR) isoforms and tissue-specific actions.** Tissues with predominant TRα or TRβ expression and effects of thyroid hormone signaling. *SHBG*, Sex hormone–binding globulin; *Dio1*, type 1 5′-deiodinase; *Dio2*, type 2 5′-deiodinase.

appear to collaborate in the regulation of hearing, based on their function in the cochlea.[24] TRα1, while present in the liver and pituitary, plays little role in isolation in these tissues. Its major roles include actions in the atrium and ventricle of the heart, as well as in bone, where its expression in a variety of cell types in that tissue is likely to be more than 10-fold more than that of the TRβ isoform.[25] In addition, TRα1 has principal actions in the gastrointestinal tract, where it plays a strong role in function and cell renewal. Additionally, TRα1 is important in the regulation of many actions of THs in the brain, but again, cell-specific actions are limited by available tools. While TRα2 has always been viewed as a potential dominant inhibitor of TRα1, its role *in vivo* does not appear to be paramount in determining cell-specific TH actions. Taken together, it is clear that each of the isoforms have cell-specific functions, likely based on their ratio of expression. However, it remains to be determined in future work whether the isoforms' structural differences delineate a unique genomic response. This would ideally require experiments where TRβ1 would replace TRα1 in a particular target tissue. In experiments performed in cell culture where TRβ1 and TRα1 were expressed at a similar level, there was evidence of isoform specificity in terms of target activation.[26]

As outlined, the TRβ and TRα isoforms are distinct at the nucleic acid and amino acid levels but share a high degree of structural homology that is also shared with other members of the nuclear receptor superfamily. The TRs all possess both a similar DNA-binding domain in the central region of the protein, followed by a ligand-binding domain that is specific for T3 at nM concentrations (with TRα1 having a slight degree of enhanced sensitivity for T3) and overlaps with the activating function 2 domain. What differentiates the TR isoforms are unique N-terminal domains (also referred to as activating function 1 [AF-1] domains) that are relatively unstructured but may allow for differential DNA binding and also possess independent transcriptional properties. The DNA-binding domains of the TR isoforms consist of two zinc finger protein domains that mediate contact with DNA and allow for receptor binding. The specificity of TR binding to target DNA elements is encoded for by the P-box in the first zinc finger. The importance of the P-box is highlighted by experimental mutations in mouse models that have allowed investigators to interrogate the role of the TR in the absence of its ability to bind DNA.[27] The second zinc finger, in conjunction with the C-terminal extension (or hinge region), also plays a critical role in DNA response element recognition of the TR isoforms and in their ability to heterodimerize with the retinoid X receptor (RXR) isoforms, their preferred partners when bound to DNA response elements. Importantly, this interaction with RXR in all cases has RXR bound 5', or upstream, to the TR. This interaction via the DNA-binding domain and C-terminal extension and hinge region with the TR and RXR, confirmed in structural studies, likely explains the preference of the RXR/TR heterodimer for response elements that contain two classic nuclear receptor half-sites (AGGTCA) spaced by four nucleotides (Fig. 67.4). This is in contrast to other nuclear receptor heterodimer pairs such as the vitamin D receptor with RXR and the retinoic acid receptor isoforms with RXR, which prefer elements that have spacings of three and five base pairs, respectively.[28]

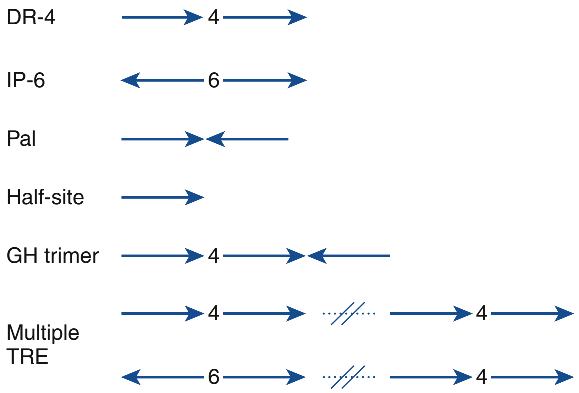

DR-4
4
IP-6
6
Pal
Half-site
GH trimer
4
Multiple TRE
4
4
6
4

Fig. 67.4 **The diverse organization of thyroid hormone response element (TREs).** A schematic of orientations *(arrows)* and spacing *(numbers between arrows)* of AGGTCA half-sites observed in natural TREs. With multiple TREs, the orientations can differ between elements.

The majority of TR isoform binding is likely as a heterodimer with RXR, but it remains possible that the TR can also function to bind DNA as a homodimer, or even as a monomer. To exerts its actions, the TR must engage T3 via its ligand-binding domain (LBD). Structural studies of the TR LBD demonstrate that it, like other nuclear receptors, contains a ligand-binding pocket that is surrounded by 12 alpha helices[29] (Fig. 67.5). That it exerts functions both in the presence and absence of T3, helix 12 (H12) demonstrates an ability to be in an open position in the absence of ligand. When T3 moves into the ligand-biding pocket, H12 shifts and closes over the pocket.[30] This movement of H12 creates two distinct grooves to allow for separate groups of TR-interacting proteins to engage the receptor, either in the presence of or absence of T3. Most likely, in the presence of low levels of T3, the unliganded structure is preferred, while in the presence of excess ligand the liganded structure is preferred (Fig. 67.5). The importance of a variety of these helices in forming the ligand binding pocket for T3 is shown by the many mutations found RTH syndrome, where mutations in the LBD usually lead to a reduced or absent ability to bind T3.[31]

Importantly, structural and biochemical studies have determined the nature of the RXR/TR heterodimer and its ability to engage with target response elements *in vitro* and in cell culture models. Indeed, these techniques identified a variety of T3 response elements such as the DR4 element described previously, as well as palindromic and inverted palindromic arrangements of the AGGTCA half-site. However, identification of the true response elements for the TR *in vivo* was not possible until the development of *in vivo* chromatin immunoprecipitation, which allows for the identification of DNA sequences that engage the TR directly. Experiments performed in mouse liver using an overexpressed TRβ1 isoform that can be biotinylated *in vivo,* and thus precipitated by streptavidin, demonstrated thousands of binding sites for the TR, with many of the sites located at great distances from regulated genes or within introns. Classically regulated T3 targets, such as the Dio1 gene, was shown to have TR binding sites in its first intron, while the spot 14 and fatty acid synthase genes had more traditional sites within 10 kB of their transcription start sites. Importantly, using motif enrichment tools, the most common binding site seen was the DR+4 element, and there was good overlap with previously known RXR binding sites, demonstrating the action of the heterodimer *in vivo*. However, not all TR binding sites were also occupied by RXR, raising the possibility that the TR homodimer and monomer could also play a role in TH action. While DR+4 sites were associated with genes induced or positively regulated by T3, there were fewer binding sites associated with genes negatively regulated by T3.[32] Similar results were seen in experiments in mouse liver utilizing an antibody against the TR itself, again with the DR+4 motif predominating.[33] Interestingly, in both of these studies a percentage of TR-binding sites were enhanced in the presence of T3, suggesting that the traditional view of the TR being engaged on DNA response elements irrespective of the concentration of ligand is not correct for every response element. The significance of the DR+4 motif in TR binding has been further supported by cell culture studies using tagged TRs that could either be biotinylated or immunoprecipitated. Finally, a recent mouse model utilized a strategy whereby the endogenous TRβ1 allele was tagged at its N-terminus by hemagglutinin (HA) sequences and would thus be expressed with a tag that could be immunoprecipitated by an antibody directed against HA. In liver chromatin, this model confirmed the importance of the DR+4 element as the primary TRβ1 binding site but also demonstrated a role for the DR+4 binding site in mediating negative regulation or the downregulation of genes in the presence of T3. Interestingly, this novel model also suggested that T3-mediated recruitment of the receptor did play a significant role across the genome.[23] Thus, these *in vivo* unbiased approaches to examining TR binding to DNA are beginning to articulate a picture whereby the DR+4 element is paramount for TR binding, especially in liver. Further work will be required to demonstrate that the binding sites identified (especially in the context of negative regulation) actually mediate regulation of the gene in closest proximity. Indeed, it remains possible that distinctly different binding sites that are less frequently seen and are at great distances from target genes may play critical roles in regulation.

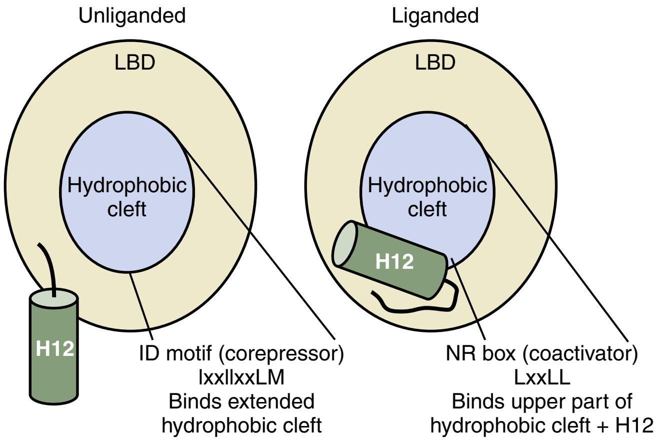


Fig. 67.5 Helix 12 position and coregulator binding. Thyroid hormone receptor TR C-terminal helix 12 is displaced away from the hydrophobic cleft without bound hormone, exposing a large hydrophobic cleft, which serves as an interaction site for α-helical motifs from corepressors (IDs). With hormone, H12 packs over the lower part of the corepressor binding surface, blocking corepressor binding and creating a new binding site for coactivator NR boxes (LxxLL). *LBD,* Ligand-binding domain.

Thyroid Hormone Receptor Coregulatory Proteins

Beginning in the 1990s it became clear that the TR recruited activity that was dependent upon the presence or absence of ligand. In these

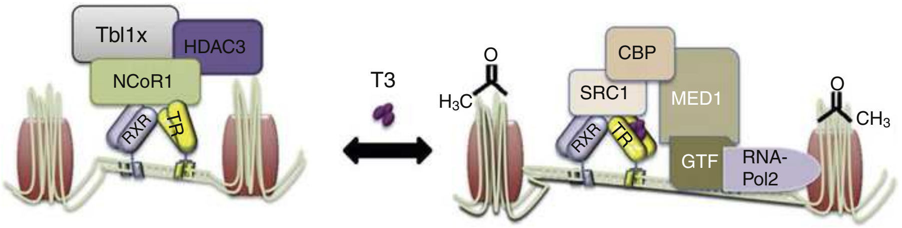


Fig. 67.6 CoRepressors and CoActivators are recruited to the TR based on the presence of ligand. In the absence or in the presence of limiting amounts of T3 the TR recruits the nuclear corepressors (ie NCoR1) as part of a multiprotein complex to repress transcription. In the presence of T3 the corepressor complex is dismissed and coactivators including SRC1 are recruited to activate trasncription. Coregulators, corepressors, and coactivators harbor nuclear receptor–interacting domains. NCoR1 and SMRT harbor three nuclear receptor interacting domains; N3 and N2 interact preferentially with unliganded thyroid hormone receptor (TR), whereas steroid receptor coactivator (SRC) harbors an RID region that interacts with the TR upon adoption of the liganded conformation. (From Mendoza A, Hollenberg AN. New insights into thyroid hormone action. *Pharmacol Ther*. 2017;173:135–145.)

studies the TR-LBD was fused to a heterologous DNA-binding domain (in this case yeast Gal4), and its function could be altered by the presence of a full-length TR that would compete for the activity allowed for by the Gal4-TR hybrid.[34] Central to these experiments was the long-held observation that, in cellular transcriptional assays on artificial DNA-binding elements, the unliganded TR was a potent repressor of gene expression for positive target genes, while it activated the expression of the regulatory elements of negatively regulated target genes. With the addition of T3, the TR then activated positively regulated genes and repressed negatively regulated genes, demonstrating that the TR possessed repressing and activating functions, depending upon the presence of ligand. The importance of the function of the unliganded TR was further highlighted by genetic studies demonstrating that mice could survive the deletion of all TRs, while mice that lacked TH at birth that was not replaced could not survive.[35]

To identify proteins that could mediate differential functions of nuclear receptors and the TR based on the presence or absence of ligand, pivotal experiments were performed initially using an immobilized estrogen receptor LBD to screen radiolabeled cellular lysates for target-interacting proteins or yeast-two hybrid assays using the LBD of a nuclear receptor as bait to identify interacting proteins either in the presence or absence of ligand. The results of these experiments and many others that followed was the identification of groups of proteins that could interact with the TR based on ligand availability, and as a field include hundreds of proteins. Strikingly, as will be outlined later, a central quality of the majority of these interacting proteins was an ability to enzymatically alter histone acetylation or methylation, implying that the TR functions as an epigenetic switch to modify chromatin in the presence or absence of its ligand.

Two general classes of proteins that interact with the TR are fundamental to its actions and are termed corepressors and coactivators (Fig. 67.6).[36,37] As their names suggest, they offer distinct regulatory functions based on their recruitment. Importantly, the structure of the TR-LBD, with the location of H12, predicts how they may be differentially recruited. Indeed, coactivators, which include members of the steroid receptor coactivator (SRC-1, 2, and 3) family, CREB-binding protein (CBP)/p300 and coactivator-associated arginine methyltransferase 1 (CARM1), and protein arginine N-methyl transferase 1 (PRMT1) all contain motifs termed nuclear receptor–interacting domains, which are specified by leucine-rich motifs (LxxLL) that dock in a groove created by a closed H12 as T3 binds.[38] In contrast, the corepressors heralded by nuclear receptor corepressor 1 (NCoR1) and the silencing mediator of the retinoic acid and TH receptor (SMRT or NCoR2) possess in their C-terminus isoleucine-rich domains (LxxH/IxxII), also termed CoRNR boxes, which bind to the groove in the unliganded TR created by the open position of H12.[39,40] It is likely that, in the presence of limiting amounts of T3 or in the hypothyroid state, the majority of TRs exist in the unliganded state, and as T3 concentration increases, the liganded-conformation is favored, corepressors are dismissed, and coactivators are recruited. However, mouse genetic experiments and human genome-wide association study (GWAS) data suggest that the amount of specific coregulator present dictates the ability of the TR to be activated or repressed. For example, in humans with excess NCoR1, T4 levels must be higher to achieve normal TSH levels. Similarly, mice that lack functional NCoR1 have lower TH levels with a normal TSH. Taken together, these data suggest that the corepressor/coactiviator ratio *in vivo* regulates sensitivity to TH.[41,42]

As discussed, the principal coactivators of the TR include members of the SRC family, CBP/P300, the methyltransferases, and also members of the Mediator complex. There are likely others also that play a role, depending on the target gene and cell type where T3-dependent regulation is occurring. The SRC family is recruited to the T3-bound RXR/TR heterodimer via their nuclear receptor–interacting domains, which demonstrate specificity to particular nuclear receptors. It is likely that CBP/p300 and other coactivators are scaffolded on the principal SRC–TR interaction, and thus play an essential role. The Mediator complex appears to link the activated TR complex to RNA polymerase. It has also been suggested in biochemical assays that SRC-1 makes contact with TR isoforms via the N-terminal or AF-1 domain, which could play a role in T3-mediated negative regulation. Genetic analysis of coactivator function in mice has demonstrated a critical role for SRC-1 in T3 action both in the liver and centrally, where these mice display a phenotype of resistance to TH, implying that SRC-1 is required for determining the set point of the hypothalamic-pituitary-thyroid (HPT) axis.[43] When deleted in conjunction with individual TR isoforms, SRC-1 has also been shown to play a role in growth and body weight that could be T3-dependent. However, SRC-1 may not play a role in the regulation of heart rate by T3. In contrast to SRC-1, murine knockout studies of SRC-2 and SRC-3 demonstrate no specific defect in T3 action, though they have numerous other metabolic phenotypes. Much work remains to be done to ascertain the specific coactivators that allow for the cell-specific actions of TR isoforms.[36]

The principal TR corepressors NCoR1 and SMRT also form a multiprotein complex that mediates histone deacetylation and transcriptional repression. The large, 270-kD proteins are paralogs and share homology in terms of their domain structure but have specificity in terms of their ability to interact with nuclear receptors, including the TR. The N-terminal regions of NCoR1 and SMRT contain repression domains that recruit proteins such as G protein pathway suppressor (GPS2) and transducin β-like 1 (TBL1 or TBL1X) and its homolog, TBL-related 1 (TBLR1or TBL1XR1), which all together form the core repression complex, together with HDAC3, which is recruited and activated by a specific domain (deacetylase activation domain [DAD]) present in both NCoR1 and SMRT. Mice that lack a functional DAD domain can no longer mediate deacetylation but still have repressive properties, indicating that NCoR1 and SMRT have activity independent of their ability to recruit and activate HDAC3. The C-termini of NCoR1 and SMRT contain three nuclear receptor–interacting domains that mediate interactions with nuclear receptors. Early biochemical studies suggested a preference of NCoR1 for the TR, but testing of this was hampered *in vivo* by the fact that global deletion of NCoR1 or SMRT is embryonic-lethal in the mouse. To get around this lethality, a mouse model was developed that created a hypomorphic NCoR1 allele, termed NCoRΔID, that encoded the full NCoR1 protein except for the two interacting domains that are required for interactions with the TR.[44] Using this targeted allele, tissue-specific or global models could be built that would live and express an NCoR1 protein that was nonfunctional in terms of TR action. In each model developed where NCoRΔID was expressed in lieu of NCoR1, mice displayed evidence of increased sensitivity to T3. For example, in mice where NCoRΔID was expressed in all tissues, both T4 and T3 levels were significantly low in the sitting of normal TSH, yet energy expenditure was increased, growth was normal, and T3-mediated gene expression was normal, all of which is consistent with increased sensitivity to T3. Further proof of this concept was provided by mice with the PV mutation causing RTH present in their Tβ or TRα alleles. When crossed with NCoRΔID mice, the syndrome of RTH was corrected in both models, again consistent with NCoR1 mediating sensitivity to T3.[45] Similarly, in NCoRΔID mice that lacked SRC-1, increased sensitivity to T3 was reversed, confirming the notion that the balance between available corepressors and coactivators determines sensitivity to T3.[46] Further support for this hypothesis comes from the identification of a variety of NCoR1 splice variants *in vivo* that, in turn, alter hormone sensitivity.[47] The importance of this observation likely carries

TABLE 67.1 Thyroid Hormone Action Classification

Action Type	Category	Description
Type 1	TR-dependent signaling of TH with direct binding to DNA	Canonical TH signaling, TR binds directly to consensus TRE
Type 2	TR-dependent signaling of TH with indirect binding to DNA	TR indirectly modifies gene expression by chromatin remodeling, possibly tethered to DNA by other transcription factors
Type 3	TR-dependent signaling of TH without DNA binding	Cytoplasmic TR can activate kinases and promote membrane signal transduction pathways
Type 4	TR-independent TH signaling	TH interaction with membrane receptors, including integrins and actin polymerization

TR, Thyroid hormone receptor; *TH*, thyroid hormone; *TRE*, thyroid hormone receptor response element.
From Flamant F, Cheng SY, Hollenberg AN, et al. Thyroid hormone signaling pathways: time for a more precise nomenclature. *Endocrinology*. 2017;158:2052–2057.

through to humans, where GWAS and quantitative trait locus studies demonstrate that increased free T4 levels are associated with increased NCoR1 in order to overcome the decreased sensitivity due to higher levels of NCoR1.[42] Interestingly, human mutations in TBLX1 have also been described, and in these patients TH levels are low and the TSH is normal. Currently, they are classified as having central hypothyroidism, but it remains possible that, if the corepressor complex is dysfunctional, they would in fact have increased sensitivity to TH.[48] The specificity of NCoR1 for the TR has been further confirmed in mouse models that either delete SMRT action in the liver or globally in adulthood to get around the embryonic lethality of early SMRT deletion. In both of these models, SMRT had no effect on the regulation of T3 target genes or in the systemic regulation of the HPT axis.[49,50]

POSITIVE AND NEGATIVE GENE REGULATION BY THYROID HORMONE

The majority of TH's actions are mediated by its genomic regulation via the TR isoforms and their ability to recruit coregulators to modify the histone environment to stimulate or repress target genes, referred to as type 1 action. While each of the players are now known, it is likely that multiple mechanisms regulate target genes. Indeed, the regulation of TH action by the TR has been classified into three distinct subtypes: 1) direct regulation of target genes by bound TR; 2) regulation of gene expression by tethered TR, i.e., via another transcription factor; and 3) regulation that is independent of TR recruitment or due to nongenomic mechanisms (Table 67.1).[51]. A number of recent studies have addressed positive regulation using genetic mouse models and genome-wide chromatin assays to clarify the mechanism by which T3 induces gene expression, particularly in the liver, where hundreds of genes are both upregulated and downregulated by T3.[23,52,53] To date, most positive regulation by the TR is explained via its direct binding to genomic response elements (Fig. 67.1). TR action occurs at many regulated genes, whereby, in the absence of T3, there is decreased histone acetylation (H3K27 and H3K9) in regions in proximity to the regulated gene that correspond to a TR binding site. The acetylation status of these sites is controlled by the recruitment of NCoR1 and HDAC3, and the presence of T3 leads to hyperacetylation via the accrual of coactivators such as SRC-1 and CBP, leading to enhanced histone acetylation. Examples of such classically regulated targets include spot14, fatty acid synthase, and malic enzyme. Interestingly, many direct targets of the TR are repressed independently of NCoR1, such as the Dio1 or cyp17a1 genes, suggesting that either repression is mediated by a separate corepressor, or repression of these targets is mediated by a lack of TR binding that only occurs when T3 is present, which then leads to the recruitment of coactivators that mediate histone acetylation and gene expression. Importantly, chromatin immunoprecipitation analysis of histone acetylation discloses many areas of the genome that are acetylated in response to T3 but that are not near a TR binding site. It remains possible that remote TR binding sites act from a great distance to regulate these areas of histone acetylation and potentially target genes.

KEY POINTS

- The structure of the thyroid hormone receptor isoforms ligand-binding domain changes secondary to the amount of ligand present, which determines the type of coregulator complex recruited and confers tissue-level sensitivity.

While the mechanisms underlying positive regulation have become clearer with genome-wide techniques, the mechanisms underlying negative regulation by TH remain in question. Understandably, many studies have focused on how the TRβ isoform regulates TRH and TSH subunit gene expression, given the importance of this in determining the set point of the thyroid axis. The regulation of the central axis is known to be dependent on TRβ2 via its ability to bind to DNA, as mouse models that either lack the isoform or express a TRβ isoform that cannot bind DNA have evidence of central resistance to TH, with elevated TRH mRNA expression and increased TSH subunit gene expression. Furthermore, the coactivator SRC-1 also appears to play a role in the regulation of the central axis, as its deletion leads to central resistance to TH. In contrast, expression of NCoRΔID in the absence of SRC-1 reestablishes the proper set point of the axis, implying that TRH and TSH subunit gene expression is mediated by the TRβ2 isoform via its interactions with NCoR1 and SRC-1. However, studies to date have not defined this at the cellular level in terms of demonstrating intact DNA-binding sites for the TR in the hypothalamus or pituitary.

Early experiments using cotransfection assays in heterologous cell lines identified TR binding sites in regions close to the transcriptional start site of the TRH and TSHβ genes. Indeed, site 4 in the TRH promoter recruits the RXR/TR heterodimer, while a strong monomeric TR binding site exists just downstream of the transcriptional start site of the TSHβ gene. However, the importance of these sites *in vivo* has been difficult to test because of the difficulty of isolating hypothalamic and pituitary cell types for genome-wide analyses utilizing chromatin immunoprecipitation and other such techniques. Recent work, therefore, has focused on the liver, which is more accessible to *in vivo* approaches, to define the mechanism.

As outlined earlier, hundreds of genes are regulated negatively by T3 in the liver and follow the classic example of gene activation in hypothyroidism and then repression in hyperthyroidism. Deletion of TRβ isoforms in hepatocytes alone demonstrates its requirement for negative regulation in the vast majority of cases. Analysis of the requirement of TRβ binding in the vicinity of negatively regulated targets suggests that its binding is important for a significant number of genes, but not all. Furthermore, motif enrichment analysis has suggested the presence of either a DR+0 motif, or more recently a DR+4 motif, that would recruit the TR.[23,32] While direct TR binding is likely important, there is also clearly a role for negative regulation via independent actions of the TR in context of inducing another pathway (i.e., microRNA regulation) or interacting with another transcription factor.[54] The roles of the corepressors and coactivators are less clear, as other SRC isoforms appear to compensate for the loss of SRC-1 when it is ablated, as negative regulation is preserved in the liver of SRC-1 knockout mice. When NCoR1 is disrupted in the liver, negative regulation is also preserved, and SMRT appears to play little compensatory role. Still, recent work that demonstrates DR+4 TRβ binding sites in enhancers near negatively regulated genes suggests that it is the ratio of coactivators to corepressors recruited that determines target gene regulation, such that, in the hypothyroid state, there is enhanced coactivator recruitment leading to histone acetylation and enhanced gene expression, while the presence of T3 leads to enhanced corepressor recruitment and gene repression. Further *in vivo* work will be required to clarify the exact mechanisms of negative regulation.

Thyroid Hormone Receptor Posttranslational Modifications

KEY POINTS

- Most positive regulation by the thyroid hormone receptor (TR) is explained via its direct binding to genomic response elements. TR action occurs at many regulated genes, whereby, in the absence of triidothyronine (T3), there is decreased histone acetylation (H3K27 and H3K9) in proximity to the regulated gene. The acetylation status of these sites is controlled by the recruitment of NCoR1 and HDAC3, and the presence of T3 leads to hyperacetylation via the accrual of coactivators such as SRC-1 and CBP, leading to enhanced histone acetylation.

While the TR isoforms are strongly influenced by ligand both in terms of coregulator recruitment and DNA binding, it is also clear that other signaling pathways impact their function. The TRα isoform is known to possess a protein kinase A site within its DNA-binding domain that may interact with its ability to bind to DNA, though this has not been shown *in vivo*. Additionally, TRβ can be phosphorylated, but the significance is not known.[55] Most interestingly, both TR isoforms can be sumoylated, with TRα having two sites and TRβ three.[56] The mutation of sumoylation sites on either isoform impairs their ability to mediate ligand-dependent gene activation or repression, potentially through aberrant or diminished coregulator recruitment. Furthermore, active introduction of TR isoform sumoylation mutants interfered with the normal differentiation of preadipocytes into adipocytes. Importantly, the sumoylation sites present on TRβ appear to govern distinct pathways involved in adipogenesis, including Wnt signaling and nuclear receptor signaling via coregulator recruitment.[57]

NONGENOMIC THYROID HORMONE ACTION

While the majority of TH action is mediated by T3 engaging its cognate receptor in the nucleus, there have been significant advances in understanding rapid or nongenomic actions of T3 via a cytoplasmic TR isoform (Table 67.1). However, some actions of T3 are felt to occur too quickly to be explained by a genomic process. Indeed, *in vitro* and *in vivo*, when present in the cytoplasm, the unliganded TRβ can interact with p85, the regulatory subunit of PI3 kinase. This interaction, mediated by one of the zinc fingers in the DNA-binding domain, is blocked by T3, and PI3 kinase signaling is induced. If the site necessary for PI3 kinase interaction is mutated *in vivo* in a mouse knockin model, normal nuclear TRβ signaling is preserved in terms of the HPT axis and induction liver target genes. However, synaptic strength and plasticity in hippocampal neurons studied in slices was significantly less in the mutant mice, implying that PI3 kinase activation via the TRβ isoform is critical.[58] Further support for rapid nongenomic effects of the TR isoforms comes from more recent experiments, where *in vivo* mouse models were developed that possessed mutations in the first zinc finger of both the TRα and TRβ isoforms that prevented DNA binding, and T3 signaling was compared with wild-type controls. Strikingly, TRβ mutant mice could still selectively regulate glucose and triglyceride lowering, and additionally temperature elevation, while TRα mutant mice could still regulate heart rate in response to *in vivo* circulating T3 levels. Importantly, the effects in the TRβ DNA-binding mutant mice on glucose, triglycerides, and temperature were blocked in the mouse model where T3-mediated PI3 kinase activation was inhibited. Taken together, these data suggest that it is likely that the TR isoforms can have noncanonical effects (so called type 3 signaling) that impact physiology and are likely regulated through the activation of PI3 kinase.[58]

THYROID HORMONE TISSUE-SPECIFIC ACTIONS

Thyroid Hormone Receptor Isoform–Specific Expression

As described previously, TRs are expressed in essentially all tissues, although well-defined tissues, such as liver, heart, and brain, express higher levels of TRs than other organs[21,59] (Fig. 67.3). The estimates of TR isoform content across tissues have varied depending on the tool that is used to detect TR, which have included T3 binding, TR isoform mRNA levels, western blot with antibody to native TR protein, or introduction of specific immunogenic “tags” into the TR protein for detection by a standard antibody that has a high affinity for that tag. There are limitations to all of the approaches used, which result in wide variations in estimates of TR isoform content, especially those relying on mRNA measurement rather than protein. The affinity and specificity of antibodies for TR isoforms also varies, and there are additional differences in tissue TR content based on the species that is studied. A “knockin” strategy involves introducing a standard HA tag sequence into the *Thra* and *Thrb* genes in mice and then measuring protein content in various tissues with a high-affinity antibody that recognizes the HA tag.[22] TRα was the predominant form in all tissue, except liver, where TRβ was predominant. In metabolically-responsive tissues, including fat and muscle, TRα protein levels were 10 times higher than those of TRβ protein. In the brain, there were low levels of TRα and TRβ, but high levels of TRα2, which does not bind T3. TRβ2 was very high in the pituitary, as has been previously reported, but an interesting sex difference was seen, with TRβ2 content greater in females than males, and females had lower serum TSH and *Tshb* mRNA. TR isoform expression can also vary within a tissue. The heart has predominantly TRα in the atria, mediating the chronotropic actions of TH, and the ventricle contains more TRβ. The phenotypes of individuals with RTH beta and alpha have largely been validated as being associated with TR isoform–specific functions, which have also been identified in mouse, *Xenopus,* and zebrafish models.[60] A further variation is TR intracellular distribution. Although TR resides predominantly in the nucleus, there is rapid shuttling between the cytoplasm and nucleus that varies by specific TR isoform.[61]

Growth

TH is necessary for proper growth and development of the fetus, as well as for normal child development, and exerts widespread influences on multiple aspects of metabolism and organ function in adults. TH regulation of gene expression in development is complex, and several important underlying principles have been identified.[21,62] First, developmental defects that arise in hypothyroid mothers are not reversed by later hormone replacement, whereas adults with TH imbalances exhibit metabolic disturbances that in most cases can be reversed by restoring TH to correct levels. TH triggers key developmental events in defined temporal windows, with abnormalities seen with both deficient and excess TH, especially as has been shown in sensory development of the inner ear and retina.[63] These developmental windows are regulated by specific expression of TR isoforms and the Dio2 enzyme, which activates T4 to T3, and the Dio3 enzyme, which converts T4 to the inactive rT3. TH-regulated genes important for organ function in the adult, such as those involved in metabolic regulation and brain function, remain sensitive to alterations in hormone levels, and gene expression is linked to thyroid status.

Gene expression patterns reveals that some genes are induced in multiple tissues, whereas other target genes are regulated in a manner that is highly tissue- and gene-specific. Another additional level of regulation is TH action on expression of microRNAs, an emerging area of TH regulation.[64] Thus, TH is a primary regulator of some genes, but the hormone must cooperate with other factors to induce expression of other genes. An example is TH metabolic regulation, which involves "crosstalk" with other metabolic regulators, including peroxisome proliferator-activated receptor (PPAR)α and PPARγ, as well as PPARγ coactivator 1-α (PGC-1α).[65,66]

Specific effects of TH will be discussed, arranged by the tissue or site of action.

Stem Cell Proliferation

In most tissues, TH stimulates stem cell proliferation as well as differentiation, in contrast to morphogens, like retinoic acid, that predominantly promote differentiation.[67] Direct TH stimulation of stem cell proliferation, predominantly mediated by TRα, includes stem cells in the intestine, skeletal muscle, pancreatic islets, heart, liver, adipose tissue, bone, and neurons[67,68] (Table 67.2). In many tissues, such as skeletal muscle and pancreatic islet cells, the stimulation of stem cells requires TR, but is also regulated by local production of T3, by actions of Dio2 and Dio3.[69] Stimulation of stem cell proliferation is important in development, but also to mediate regeneration after injury, as has been shown in skeletal muscle, cardiac muscle, and brain.[70] These findings have led to an interest in TH treatment in a variety of conditions to promote regeneration after injury, such as traumatic brain injury, stroke, myocardial infarction, and muscle injury.

KEY POINTS

- Stimulation of stem cell proliferation is important in development, but also to mediate regeneration after injury, as has been shown in skeletal muscle, cardiac muscle, and brain.

Thyroid Hormone and Somatic Growth

TH deficiency in both the mother and fetus results in reduced somatic growth. Normal maternal thyroid status is sufficient to compensate for fetal thyroid deficiency in congenital hypothyroidism, with maternal TH crossing the placenta from high to low levels in the fetus. Combined maternal and fetal hypothyroidism, as seen in profound iodine deficiency, results in irreversible somatic growth deficit. Hypothyroidism in children is associated with reduced height. TH directly stimulates production of growth hormone in some species, including mice and rats. In children with deficiency of both growth hormone and TH, treatment with both hormones works to restore normal growth.

Brain

TH is required for normal brain development, initially from transport of maternal TH across the placenta and then from TH produced by the human fetus, at around 13 weeks, when the thyroid gland develops.[71] Maternal TH can compensate for defective fetal thyroid gland development in the condition referred to as congenital hypothyroidism. The infant, however, must be identified by screening at birth and started promptly on TH replacement to support normal brain development. The combination of maternal and fetal TH deficiency, as is seen with profound iodine deficiency during development, is associated with profound and irreversible abnormalities of brain and nervous system development.

TRα1 is the predominant TR isoform in most of the brain, although TRβ2 is the TR isoform expressed in the pituitary and hypothalamus[21,22] (Fig. 67.3). TH is important for neural stem cell proliferation and expression of both neurons and glial cells.[72,73] TH promotes neurogenesis and oligodendrocyte differentiation and maturation.[74-76] A number of TH-stimulated gene targets have been identified in the cortex, hippocampus, and cerebellum. TH induction of a number of genes in the brain is antagonized by the expression of chicken ovalbumin upstream transcription factor 1 (COUPTF1), which has a developmental pattern of expression and blunts the expression of neural differentiation in key developmental windows.[77] TH induction of neural proliferation is not only important developmentally, but has also been shown to play a role in recovery after brain injury. TH treatment reduces hypoxia-induced damage and promotes recovery.[78] TH treatment within a few hours of traumatic brain injury in rodent models significantly reduces the extent of injury and promotes recovery.

Genes have been identified in rodent models that are TH-responsive and important in neurogenesis, including T-box brain transcription factor 1 (*Tbr1*), *Sox2*, and neurogenin1.[79,80] A TH gene target in the hippocampus, and area important for memory and emotion, is calcium calmodulin ATPase IV (CamKIV). CamKIV is upregulated by T3, but stimulation is inhibited by COUPTF1.[81]

Oligodendrocyte differentiation is promoted by TH, and a number of gene targets have been identified. Myelination is promoted by the TH-stimulated, and TH analog-induced expression of myelin basic protein has been reported.[82] Another TH gene target is glial fibrillary acidic protein (GFAP), which is expressed in astrocytes. GFAPs form the intermediate filament and play an important role in neural support. TH and TH analog promotion of myelination have been used to promote myelination in models of demyelinating illness, such as multiple sclerosis. In addition to stimulating myelination, TH and thyromimetics have been shown to act to inhibit myelin and axonal degeneration and oligodendrocyte loss in models of autoimmune encephalitis.[83]

Mouse models with THRA1 mutations have various defects, depending on the mutation, but most have shown some defects in brain development, neurological impairment, and behavioral abnormalities.[84] Defects include those of neuronal and glial cell lines.

TH regulates brain development, and hypothyroidism results in intellectual disability and multiple neurologic defects.[59] Slow mentation and other central nervous system disturbances characterize

TABLE 67.2 **Thyroid Hormone Receptors and Proliferation/Differentiation**

Cell Type	Proliferation	Primary Receptor	Genes/Pathways	Response Modulated
Hepatocyte	Increased	TRβ	Cyclin D1	Liver hyperplasia, liver regeneration, regression of neoplastic nodules
Pancreatic β cells	Increased	TRα, TRβ	Cyclin D1/CDK/Rb/E2F PI3K/Akt	Cell proliferation and survival during pancreatic development
Intestinal epithelial cells	Increased	TRα	Wnt/β-catenin Notch, BMP	Intestinal maturation and renewal
Cardiomyocytes	Increased	TRα	MHCα, ANF, SERCA2, β-1 adrenergic receptors, K+ channels PI3K/AKT	Contractility, heart rate, cardiac hypertrophy, cardiomyocyte maturation
Skeletal muscle cells	Increased	TRα	MyoD and contractile genes	Muscle contraction, repair in response to injury

TR, Thyroid hormone receptor.
From Pascual A, Aranda A. Thyroid hormone receptors, cell growth and differentiation. *Biochim Biophys Acta*. 2013;1830:3908–3916.

hypothyroidism. In hyperthyroidism, there can be episodes of anxiety and even psychosis. CamKIV, a TH-responsive gene expressed in the hippocampus, is important in learning and memory, and reduced expression may be responsible for some of the cognitive deficit seen in patients with hypothyroidism.

KEY POINTS

- Thyroid hormone (TH) is important for neural stem cell proliferation and expression of both neurons and glial cells. TH promotes neurogenesis and oligodendrocyte differentiation and maturation. TH and TH analog promotion of myelination have been used to promote myelination in models of demyelinating illness and recovery from brain injury.

Heart and Blood Vessels

THs have multiple effects on the cardiovascular system.[85] One action is to increase cardiac output via an increase in cardiac contractility and heart rate. TH regulates genes involved in cardiac contractility; there is induction of the sarcoplasmic reticulum Ca^{2+} ATPase 2, involved in calcium reuptake during the diastolic phase, and α myosin heavy chain (MHC), a fast ATPase required for heart contractility that is expressed in adult heart. Conversely, TH inhibits expression of βMHC, a slow ATPase expressed in embryonic heart and upregulated in stress conditions. Excess TH, in contrast, is associated with atrial arrhythmias, as well as development of high-output heart failure. TH and its analogs have been used to increase cardiac output in heart failure patients and after coronary artery revascularization, with modest improvement.

Skeletal Muscle

THs promote muscle catabolism and increase skeletal muscle energy expenditure in adults.[15] TH also induces the insulin-sensitive glucose transporter in muscle and promotes fat burning. These effects may help to sensitize muscle to insulin response. Accordingly, human Dio2 gene polymorphisms are correlated with insulin resistance in human populations, although this finding remains controversial.[86] TH has a specific role in promoting proliferation of satellite cells and skeletal muscle stem cells, during development and during regeneration following injury.[70] TH activation, T4 to T3 conversion by Dio2,[87] and the presence of the TRα are required for normal development and response to injury.

Skeletal System

TH is required for normal bone growth and maturation.[88] Juvenile hypothyroidism causes delayed bone formation and short stature, whereas thyrotoxicosis leads to increased growth and advanced skeletal development. In adults, however, TH excess promotes bone resorption. TH excess can lead to osteoporosis, especially in postmenopausal women.

Small Intestine

TH is required for normal maturation of the small intestine. A prominent phenotype of mice with TRα deletions and mutations is abnormal development of the small intestine and malabsorption. TRα1 is required for proliferation of intestinal epithelial progenitor cells via induction of the β-catenin protooncogene and underlies TH action.[89]

Metabolic Regulation and Actions in Liver

One of the most important effects of TH involves metabolic regulation, including changes in basal metabolic rate (BMR) in multiple tissues (Table 67.3).[66] THs stimulate oxygen consumption (indicative of enhanced metabolism) in multiple locations, including skeletal muscle, liver, kidney, and intestine. TH potentiates adrenergic signaling.[90] Increases in BMR are probably partly related to increased mitochondrial activity and number, but details of how TH mediates these effects remain unknown. Importantly, THs also induce expression of uncoupling proteins (UCPs), which are mitochondrial membrane proteins that dissipate the proton gradient as heat, rather than allow the gradient to be used to generate adenosine triphosphate. TH does not always enhance BMR. For example, TH does not enhance BMR in most regions of the brain, and the hormone actually suppresses metabolic activity in the pituitary gland. The TH receptors in these tissues are functional, suggesting that key mediators of TH regulation of BMR are blocked or absent.

THs exert multiple effects on the liver and are potent mitogens in this tissue, especially in growing animals. TH also influences multiple metabolic processes such as fatty acid β-oxidation and gluconeogenesis, both key aspects of the fasting response. However, TH can also stimulate expression of enzymes involved in lipogenesis and generation of NADPH-reducing equivalents that are required for fat synthesis and protection against reactive oxygen species. Because it is thought that fat oxidation and synthesis do not occur simultaneously, these processes are probably separated spatially or temporally, and TH must cooperate with other signaling mechanisms to regulate these effects. TR interacts directly with PPARα and PPARγ in regulation of some genes and may be a mediator for an indirect influence of thyroid status on metabolism.[66]

TH plays a major role in regulation of cholesterol metabolism in liver in rodents, and studies of patients with thyroid excess and deficiency states suggest that some of these pathways are also TH-regulated

TABLE 67.3 Thyroid Hormone Regulation of Metabolism

Tissue	Process	TR Isoform	Impact
BAT	Potentiation of adrenergic stimulation of BAT activity	TRα	Increased thermogenesis
BAT	Stimulation of UCP1	TRβ	Increased thermogenesis
WAT	Direct and adrenergic potentiation, induction of lipolysis	TRα/TRβ	Reduce lipid stores, increased energy expenditure
Liver/WAT	Lipogenesis	TRβ	Restore lipid stores
Liver	Coupling of autophagy to mitochondrial fatty acid oxidation	TRβ	Reduce hepatic lipid stores
Liver	Reverse cholesterol transport	TRβ	Reduce serum cholesterol
Liver	Increase expression of LDL Receptor	TRβ	Reduce serum cholesterol
Liver	Reduced hepatic fat deposition	TRβ	Reduce hepatic lipid stores

BAT, Brown adipose tissue; *WAT*, white adipose tissue; *TR*, thyroid hormone receptor.
From Sinha RA, Singh BK, Yen PM. Direct effects of thyroid hormones on hepatic lipid metabolism. *Nat Rev Endocrinol*. 2018;14:259–269 and Mullur R, Liu YY, Brent GA. Thyroid hormone regulation of metabolism. *Physiol Rev*. 2014;94:355–382.

in humans.[91] TH induces expression of enzymes involved in cholesterol synthesis but also increases levels of the low-density lipoprotein (LDL) cholesterol receptor, which promotes cholesterol uptake from the circulation. Likewise, TH increases expression of apolipoprotein A1, a key protein component of high-density lipoprotein (HDL), and increases expression of an HDL receptor (SR-B1). Consequently, TH excess may promote increased cholesterol flux from the plasma to the liver through both the LDL and HDL pathways. Additionally, TH stimulates cholesterol efflux and cholesterol-to–bile acid conversion. More recently, TH has been shown to directly stimulate autophagy. Collectively, these mechanisms account for the net reductions of serum cholesterol levels that are observed with TH and the net reverse cholesterol transport reflected by an increased flow of bile acids into the gut.

TH activation of lipolysis, reduction in serum cholesterol, as well as autophagy in the liver and reduction in liver fat storage, are attractive targets to treat a range of metabolic disorders.[91] TH analogs with more selective activation of TRβ and hepatic action were very effective in lowering cholesterol in statin failure patients, but use was limited by action on cartilage in animal models. TH and analogs have been effective for regression of hepatic steatosis in animal models and some clinical studies and remain an active area of clinical development.

KEY POINTS

- Thyroid hormone induces expression of the low-density-lipoprotein cholesterol receptor, apolipoprotein A1, a key protein component of high-density lipoprotein (HDL), and increases expression of an HDL receptor (SR-B1), with a net reduction of serum cholesterol levels and net reverse cholesterol transport, reflected by an increased flow of bile acids into the gut.

White and Brown Adipose Tissue

THs promote differentiation of precursors into white fat and induces lipogenic enzymes in preadipocytes from young rats and cell lines, but they also increase lipolysis in animals and humans.[66] However, the lipolytic effects predominates, because there is a net loss of fat in thyrotoxic states and a gain in hypothyroidism. It is thought that the lipogenesis is in response to the lipolysis in order to restore fat deposits. Hyperthyroidism is associated with loss of both fat and lean body mass, in most patients proportional to the elevation in TH levels. In contrast, hypothyroidism is generally associated with only a modest increase in fat mass.

TH stimulates adaptive thermogenesis in brown adipose tissues. In cold or in response to overeating, there is increased local production of T3 via induction of Dio2, shown to be due to a ubiquitination/deubiquitination, where adrenergic signaling and low T4 promote deubiquitination of Dio2 and increase local Dio2 activity. Thus, T3 cooperates with norepinephrine outputs from the sympathetic nervous system to induce UCPs that promote dissipation of the mitochondrial protein gradient as heat. There is a TR isoform specificity, with TRβ mediating TH stimulation of UCP1 gene expression,[92] and TRα potentiating TH-mediated adrenergic signaling.[93] Whereas previous reports suggested that adaptive thermogenesis was only important for temperature regulation in human newborns and in small mammals, studies have suggested that brown fat may also be important in adult humans.

Skin

TH is important for skin function, and TH imbalances are often first manifested in changes in appearance.[94] Hypothyroidism leads to cold, dry, and thickened skin. There is also increased hair loss. Conversely, hyperthyroidism leads to warm, moist, and smooth skin with fine, soft hair. Hormone effects are a combination of inhibitory changes in keratin expression, sterol biosynthesis, diminished sebaceous gland secretion, and increased collagen breakdown.[95] TH, when applied topically, has been shown to promote hair growth and wound repair.[94,96] TH has specific action on hair follicle stem cells and mediates the effects seen on hair changes as a result of changes in thyroid status.[97] The role of TH metabolism by Dio2 and Dio3 in the regulation of skin cancer growth and differentiation has been reported, including in melanoma and in basal and squamous cell carcinoma. T3 has been shown to be a significant inducer of the epithelial to mesenchymal transition in squamous cell carcinoma of the skin.[98] An increase in local T3 from Dio2 action, and reduction of inactivation by Dio3, promotes tumor growth and invasion.

LESSONS FROM RESISTANCE TO THYROID HORMONE RECEPTOR BETA AND ALPHA

The majority of studies that have identified TR isoform–specific actions were based on models with introduced mutations or TR gene deletions in mice, zebrafish, and frogs. Although these approaches have provided important insights, ultimately, the phenotypes of humans with mutations in genes important in TH signaling are most informative for understanding TH action. The genetic defects of TH signaling involving TR genes are now classified as RTH beta, associated with mutations in the *Thrb* gene, and RTH alpha, due to mutations in the *Thra* gene. These syndromes are discussed in detail elsewhere (see Chapter 81), but the connection between genotype and phenotype in individuals with RTH will be briefly discussed for the insights provided into tissue-specific actions of TH and the role of TR isoforms.

Resistance to Thyroid Hormone Beta

RTH beta is due to mutations in the *THRB* locus and is inherited in an autosomal dominant fashion. In most cases, patients are heterozygous for *Thrb* mutations, and the clinical signs and symptoms suggest that there is inhibition of endogenous functional TRs by the mutant receptor through a “dominant negative effect.” This explains the lack of negative feedback in the HPT axis, mediated by TRβ, which results in elevation of serum TSH and circulating TH levels. Since the central feedback regulation is blunted, circulating serum T4 and T3 are elevated, producing a “compensation” for reduced signaling at tissues with TRβ predominance. In tissues with TRα predominance, such as the heart, then there are signs of excess TH, with tachycardia. Metabolic regulation is also predominantly mediated by TRα, and RTH beta individuals have increased metabolism, higher resting energy expenditure, and a higher calorie intake compared with unaffected matched controls, without an increase in weight.[99] In a range of genetic models, there was a recapitulation of the primary biochemical phenotype of RTH, elevated blood levels of T4 and T3, and a high, nonsuppressible TSH. The high TSH levels in mice induces goiter, which is also a common clinical sign found in RTH beta patients.

Resistance to Thyroid Hormone Alpha

Thra mutations have no significant alterations in circulating TSH or T4 levels, suggesting that central TH feedback at the level of the hypothalamus and pituitary is not impaired. RTH alpha patients have an increase in serum T3 concentrations, increased T3/T4 ratios, and reduced serum reverse T3 concentrations, consistent with a reduction in Dio3 activity. This finding supports an earlier study that showed that upregulation of Dio3 by T3 is mediated by TRα.[100] Patients with *Thra* mutations present with signs suggesting hypothyroidism: short stature, delayed bone age, and clinical findings of dry skin and slow reflexes, and evidence for reduced TH action in key tissues, including delay in linear growth and tooth eruption, low serum insulin-like growth factor-1 levels, reduced muscle tone, impaired fine motor coordination, and severe constipation.[84] *In vitro* analysis of these TRα mutations reveals very low T3 binding affinity, defective activation from canonical TH response elements (TREs), repression of basal promoter activity, irreversible binding of corepressors, and a clear dominant negative effect. The response of these clinical manifestations of hypothyroidism to treatment with levothyroxine have been variable, with some improvement in growth if initiated in children and improvement of the metabolic and some cognitive manifestations in adults.[84]

KEY POINTS

- Findings from animal models regarding thyroid hormone (TH) receptor (TR) isoform specificity is confirmed by individuals with resistance to TH (RTH). In RTH beta, central feedback regulation is blunted and circulating serum T4 and T3 are elevated, producing a “compensation” for reduced signaling at tissues with TRβ predominance. In tissues with TRα predominance, there are signs of excess thyroid hormone, such as tachycardia and hypermetabolism. Patients with *Thra* mutations do not have this “compensation” and have manifestations of reduced TH action, short stature, delayed bone age, low serum insulin-like growth factor-1 levels, impaired fine motor coordination, and severe constipation.

SUMMARY AND FUTURE DIRECTIONS

TH is essential for normal development and regulates the function of a wide range of tissues in the adult. The use of animal models, including mice, zebrafish, and *Xenopus*, as well as analysis of human genetic disorders of thyroid signaling, has led to the recognition of multiple sites of regulation of TH action. The HPT axis closely regulates TH production and secretion from the thyroid gland. Additional sites of regulation include local tissue activation of the prohormone, T4, to the active form, T3, TH transporters in specific tissues, the amount and relative distribution of TR isoforms, and an important role of TR coactivators and corepressors. Newer areas of the mechanism of regulation that have been identified include posttranslational modification of TR by phosphorylation and sumoylation, subcellular localization of TR isoforms, and TH regulation of microRNAs. In addition to nuclear action, nongenomic actions of TH through membrane integrin receptors as well as modification of intracellular signaling pathways have been reported. A core action of TH is metabolic regulation, and many of these actions involve crosstalk with other metabolic signaling pathways, including PPARα and PPARγ. Genome-wide approaches have confirmed previous findings, such as the classic DNA TRE sequence, but have also shown indirect effects of TR, especially on chromatin remodeling and actions in regions without consensus TREs. TH signaling has a central role in metabolism, stem cell proliferation in development and regeneration after injury, brain development and function, retinal and ear development, cardiovascular function, and cancer. Greater understanding of these complex factors that regulate TH signaling will permit more detailed description of these interacting pathways, with the potential for selective therapeutic targets that impact these important processes.

REFERENCES

1. Mimoto MS, Refetoff S. Clinical recognition and evaluation of patients with inherited serum thyroid hormone-binding protein mutations. *J Endocrinol Invest*. 2020;43:31–41.
2. Groeneweg S, Visser WE, Visser TJ. Disorder of thyroid hormone transport into the tissues. *Best Pract Res Clin Endocrinol Metab*. 2017;31:241–253.
3. Braun D, Schweizer U. Thyroid hormone transport and transporters. *Vitam Horm*. 2018;106:19–44.
4. Dentice M, Marsili A, Zavacki A, et al. The deiodinases and the control of intracellular thyroid hormone signaling during cellular differentiation. *Biochim Biophys Acta*. 2013;1830:3937–3945.
5. Hollenberg AN, Choi J, Serra M, et al. Regenerative therapy for hypothyroidism: mechanisms and possibilities. *Mol Cell Endocrinol*. 2017;445:35–41.
6. Balhara B, Misra M, Levitsky LL. Clinical monitoring guidelines for congenital hypothyroidism: laboratory outcome data in the first year of life. *J Pediatr*. 2011;158:532–537.
7. Rose SR, Brown RS, Foley T, et al. Update of newborn screening and therapy for congenital hypothyroidism. *Pediatrics*. 2006;117:2290–2303.
8. Visser WE, Vrijmoeth P, Visser FE, et al. Identification, functional analysis, prevalence and treatment of monocarboxylate transporter 8 (MCT8) mutations in a cohort of adult patients with mental retardation. *Clin Endocrinol (Oxf)*. 2013;78:310–315.
9. Friesema EC, Ganguly S, Abdalla A, et al. Identification of monocarboxylate transporter 8 as a specific thyroid hormone transporter. *J Biol Chem*. 2003;278:40128–40135.
10. Dumitrescu AM, Liao XH, Best TB, et al. A novel syndrome combining thyroid and neurological abnormalities is associated with mutations in a monocarboxylate transporter gene. *Am J Hum Genet*. 2004;74:168–175.
11. Friesema EC, Grueters A, Biebermann H, et al. Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet*. 2004;364:1435–14371.
12. Groeneweg S, van Geest FS, Peeters RP, et al. Thyroid hormone transporters. *Endocr Rev*. 2020;41:1–55.
13. Groeneweg S, Peeters RP, Moran C, et al. Effectiveness and safety of the tri-iodothyronine analogue Triac in children and adults with MCT8

deficiency: an international, single-arm, open-label, phase 2 trial. *Lancet Diabetes Endocrinol*. 2019;7:695–706.

13a. Strømme P, Groeneweg S, Lima de Souza EC. Mutated Thyroid Hormone Transporter OATP1C1 Associates with Severe Brain Hypometabolism and Juvenile Neurodegeneration. *Thyroid*. 2018;28(11):1406–1415. https://doi.org/10.1089/thy.2018.0595.

14. Mayerl S, Muller J, Bauer R, et al. Transporters MCT8 and OATP1C1 maintain murine brain thyroid hormone homeostasis. *J Clin Invest*. 2014;124:1987–1999.
15. Bianco AC, Kim BW. Deiodinases: implications of the local control of thyroid hormone action. *J Clin Invest*. 2006;116:2571–2579.
16. Tata JR. Inhibition of the biological actions of thyroid hormones by actinomycin D and puromycin. *Nature*. 1963;197:1167–1168.
17. Weinberger C, Hollenberg SM, Ong ES, et al. Identification of human glucocorticoid receptor complementary DNA clones by epitope selection. *Science*. 1985;228:740–742.
18. Sap J, Munoz A, Damm K, et al. The c-erb-A protein is a high-affinity receptor for thyroid hormone. *Nature*. 1986;324:635–640.
19. Weinberger C, Thompson CC, Ong ES, et al. The c-erb-A gene encodes a thyroid hormone receptor. *Nature*. 1986;324:64–646.
20. Lazar MA. Thyroid hormone receptors: multiple forms, multiple possibilities. *Endocr Rev*. 1993;14:184–193.
21. Mendoza A, Hollenberg AN. New insights into thyroid hormone action. *Pharmacol Ther*. 2017;173:135–145.
22. Minakhina S, Bansal S, Zhang A, et al. A direct comparison of thyroid hormone receptor protein levels in mice provides unexpected insights into thyroid hormone action. *Thyroid*. 2020;30:1193–1204.
23. Shabtai Y, Nagaraj NK, Batmanov K, et al. A coregulator shift, rather than the canonical switch, underlies thyroid hormone action in the liver. *Genes Dev*. 2021;35:367–378.
24. Forrest D, Vennstrom B. Functions of thyroid hormone receptors in mice. *Thyroid*. 2000;10:41–52.
25. Williams GR, Bassett JHD. Thyroid diseases and bone health. *J Endocrinol Invest*. 2018;41:99–109.
26. Chatonnet F, Guyot R, Benoit G, Flamant F. Genome-wide analysis of thyroid hormone receptors shared and specific functions in neural cells. *Proc Natl Acad Sci U S A*. 2013;110:E766–E775.
27. Shibusawa N, Hashimoto K, Nikrodhanond AA, et al. Thyroid hormone action in the absence of thyroid hormone receptor DNA-binding in vivo. *J Clin Invest*. 2003;112:588–597.
28. Naar AM, Boutin JM, Lipkin SM, et al. The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell*. 1991;65:1267–1279.
29. Wagner RL, Apriletti JW, McGrath ME, et al. A structural role for hormone in the thyroid hormone receptor. *Nature*. 1995;378:690–697.
30. Glass CK, Rosenfeld MG. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev*. 2000;14:121–141.
31. Moran C, Chatterjee K. Resistance to thyroid hormone alpha-emerging definition of a disorder of thyroid hormone action. *J Clin Endocrinol Metab*. 2016;101:2636–2639.
32. Ramadoss P, Abraham BJ, Tsai L, et al. Novel mechanism of positive versus negative regulation by thyroid hormone receptor beta1 (TRbeta1) identified by genome-wide profiling of binding sites in mouse liver. *J Biol Chem*. 2014;289:1313–1328.
33. Grontved L, Waterfall JJ, Kim DW, et al. Transcriptional activation by the thyroid hormone receptor through ligand-dependent receptor recruitment and chromatin remodelling. *Nat Commun*. 2015;6:7048.
34. Qi JS, Desai-Yajnik V, Greene ME, et al. The ligand-binding domains of the thyroid hormone/retinoid receptor gene subfamily function *in vivo* to mediate heterodimerization, gene silencing, and transactivation. *Mol Cell Biol*. 1995;15:1817–1825.
35. Gothe S, Wang Z, Ng L, et al. Mice devoid of all known thyroid hormone receptors are viable but exhibit disorders of the pituitary-thyroid axis, growth, and bone maturation. *Genes Dev*. 1999;13:1329–1341.
36. Astapova I. Role of co-regulators in metabolic and transcriptional actions of thyroid hormone. *J Mol Endocrinol*. 2016;56:73–97.
37. Glass CK, Rose DW, Rosenfeld MG. Nuclear receptor co-activators. *Curr Opin Cell Bio*. 1997;9:222–232.
38. Heery DM, Kalkhoven E, Hoare S, et al. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*. 1997;387:733–736.
39. Cohen RN, Brzostek S, Kim B, et al. The specificity of interactions between nuclear hormone receptors and corepressors is mediated by distinct amino acid sequences within the interacting domains. *Mol Endocrinol*. 2001;15:1049–1061.
40. Hu X, Lazar MA. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature*. 1999;402:93–96.
41. Astapova I, Vella KR, Ramadoss P, et al. The nuclear receptor corepressor (NCoR) controls thyroid hormone sensitivity and the set point of the hypothalamic-pituitary-thyroid axis. *Mol Endocrinol*. 2011;25:212–224.
42. Teumer A, Chaker L, Groeneweg S, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat Commun*. 2018;9:4455.
43. Weiss RE, Xu J, Ning G, et al. Mice deficient in the steroid receptor co-activator 1 (SRC-1) are resistant to thyroid hormone. *EMBO J*. 1999;18:1900–1904.
44. Astapova I, Lee LJ, Morales C, et al. The nuclear corepressor, NCoR, regulates thyroid hormone action *in vivo*. *Proc Natl Acad Sci U S A*. 2008;105:19544–19549.
45. Fozzatti L, Lu C, Kim DW, et al. Resistance to thyroid hormone is modulated in vivo by the nuclear receptor corepressor (NCOR1). *Proc Natl Acad Sci U S A*. 2011;108:17462–17467.
46. Vella KR, Ramadoss P, Costa ESRH, et al. Thyroid hormone signaling in vivo requires a balance between coactivators and corepressors. *Mol Cell Biol*. 2014;34:1564–1575.
47. Goodson ML, Young BM, Snyder CA, et al. Alteration of NCoR corepressor splicing in mice causes increased body weight and hepatosteatosis without glucose intolerance. *Mol Cell Biol*. 2014;34:4104–4114.
48. Heinen CA, Losekoot M, Sun Y, et al. Mutations in TBL1X are associated with central hypothyroidism. *J Clin Endocrinol Metab*. 2016;101:4564–4573.
49. Shimizu H, Astapova I, Ye F, et al. NCoR1 and SMRT play unique roles in thyroid hormone action in vivo. *Mol Cell Biol*. 2015;35:555–565.
50. Shimizu H, Lu Y, Vella KR, et al. Nuclear corepressor SMRT is a strong regulator of body weight independently of its ability to regulate thyroid hormone action. *PLoS One*. 2019;14:e0220717.
51. Flamant F, Cheng SY, Hollenberg AN, et al. Thyroid hormone signaling pathways: time for a more precise nomenclature. *Endocrinology*. 2017;158:2052–2057.
52. Mendoza A, Astapova I, Shimizu H, et al. NCoR1-independent mechanism plays a role in the action of the unliganded thyroid hormone receptor. *Proc Natl Acad Sci U S A*. 2017;114:E8458–E8467.
53. Praestholm SM, Siersbaek MS, Nielsen R, et al. Multiple mechanisms regulate H3 acetylation of enhancers in response to thyroid hormone. *PLoS Genet*. 2020;16:e1008770.
54. van Rooij E, Sutherland LB, Qi X, et al. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science*. 2007;316:575–579.
55. Liu YY, Brent GA. Posttranslational modification of thyroid hormone nuclear receptor by phosphorylation. *Methods Mol Biol*. 2018;1801:39–46.
56. Liu YY, Brent GA. Posttranslational modification of thyroid hormone nuclear receptor by sumoylation. *Methods Mol Biol*. 2018;1801:47–59.
57. Liu YY, Ayers S, Milanesi A, et al. Thyroid hormone receptor sumoylation is required for preadipocyte differentiation and proliferation. *J Biol Chem*. 2015;290:7402–7415.
58. Martin NP, Marron Fernandez de Velasco E, Mizuno F, et al. A rapid cytoplasmic mechanism for PI3 kinase regulation by the nuclear thyroid hormone receptor, TRbeta, and genetic evidence for its role in the maturation of mouse hippocampal synapses in vivo. *Endocrinology*. 2014;155:3713–3724.
59. Bernal J. Thyroid hormone receptors in brain development and function. *Nat Clin Pract Endocrinol Metab*. 2007;3:249–259.
60. Pappa T, Refetoff S. Resistance to thyroid hormone beta: a focused review. *Front Endocrinol (Lausanne)*. 2021;12:656551.

61. Anyetei-Anum CS, Roggero VR, Allison LA. Thyroid hormone receptor localization in target tissues. *J Endocrinol*. 2018;237:R19–R34.
62. Brent GA. Mechanisms of thyroid hormone action. *J Clin Invest*. 2012;122:3035–3043.
63. Ng L, Kelley MW, Forrest D. Making sense with thyroid hormone--the role of T(3) in auditory development. *Nat Rev Endocrinol*. 2013;9:296–307.
64. Aranda A. MicroRNAs and thyroid hormone action. *Mol Cell Endocrinol*. 2021;525:111175.
65. Liu YY, Brent GA. Thyroid hormone crosstalk with nuclear receptor signaling in metabolic regulation. *Trends Endocrinol Metab*. 2010;21:166–173.
66. Mullur R, Liu YY, Brent GA. Thyroid hormone regulation of metabolism. *Physiol Rev*. 2014;94:355–382.
67. Pascual A, Aranda A. Thyroid hormone receptors, cell growth and differentiation. *Biochim Biophys Acta*. 2013;1830:3908–3916.
68. Verga Falzacappa C, Mangialardo C, Raffa S, et al. The thyroid hormone T3 improves function and survival of rat pancreatic islets during in vitro culture. *Islets*. 2010;2:96–103.
69. Aguayo-Mazzucato C, Zavacki AM, Marinelarena A, et al. Thyroid hormone promotes postnatal rat pancreatic beta-cell development and glucose-responsive insulin secretion through MAFA. *Diabetes*. 2013;62:1569–1580.
70. Milanesi A, Lee JW, Kim NH, et al. Thyroid hormone receptor alpha plays an essential role in male skeletal muscle myoblast proliferation, differentiation, and response to injury. *Endocrinology*. 2016;157:4–15.
71. Liu YY, Brent GA. Thyroid hormone and the brain: mechanisms of action in development and role in protection and promotion of recovery after brain injury. *Pharmacol Ther*. 2018;186:176–185.
72. Bernal J, Guadano-Ferraz A, Morte B. Thyroid hormone transporters-functions and clinical implications. *Nat Rev Endocrinol*. 2015;11:690.
73. Wallis K, Dudazy S, van Hogerlinden M, et al. The thyroid hormone receptor alpha1 protein is expressed in embryonic postmitotic neurons and persists in most adult neurons. *Mol Endocrinol*. 2010;24:1904–1916.
74. Lemkine GF, Raj A, Alfama G, et al. Adult neural stem cell cycling *in vivo* requires thyroid hormone and its alpha receptor. *FASEB J*. 2005;19:863–865.
75. Billon N, Tokumoto Y, Forrest D, et al. Role of thyroid hormone receptors in timing oligodendrocyte differentiation. *Dev Biol*. 2001;235:110–120.
76. Morte B, Manzano J, Scanlan TS, et al. Aberrant maturation of astrocytes in thyroid hormone receptor alpha 1 knockout mice reveals an interplay between thyroid hormone receptor isoforms. *Endocrinology*. 2004;145:1386–1391.
77. Teng X, Liu YY, Teng W, et al. COUP-TF1 modulates thyroid hormone action in an embryonic stem-cell model of cortical pyramidal neuronal differentiation. *Thyroid*. 2018;28:667–678.
78. Li J, Abe K, Milanesi A, et al. Thyroid hormone protects primary cortical neurons exposed to hypoxia by reducing DNA methylation and apoptosis. *Endocrinology*. 2019;160:2243–2256.
79. Kapoor R, Desouza LA, Nanavaty IN, et al. Thyroid hormone accelerates the differentiation of adult hippocampal progenitors. *J Neuroendocrinol*. 2012;24:1259–1271.
80. Attardo A, Fabel K, Krebs J, et al. Tis21 expression marks not only populations of neurogenic precursor cells but also new postmitotic neurons in adult hippocampal neurogenesis. *Cereb Cortex*. 2010;20:304–314.
81. Liu YY, Brent GA. A complex deoxyribonucleic acid response element in the rat Ca(2+)/calmodulin-dependent protein kinase IV gene 5’-flanking region mediates thyroid hormone induction and chicken ovalbumin upstream promoter transcription factor 1 repression. *Mol Endocrinol*. 2002;16:2439–2451.
82. Hartley MD, Banerji T, Tagge IJ, et al. Myelin repair stimulated by CNS-selective thyroid hormone action. *JCI Insight*. 2019;4:e126329.
83. Chaudhary P, Marracci GH, Calkins E, et al. Thyroid hormone and thyromimetics inhibit myelin and axonal degeneration and oligodendrocyte loss in EAE. *J Neuroimmunol*. 2021;352:577468.
84. Moran C, Agostini M, McGowan A, et al. Contrasting phenotypes in resistance to thyroid hormone alpha correlate with divergent properties of thyroid hormone receptor alpha1 mutant proteins. *Thyroid*. 2017;27:973–982.
85. Ahmadi N, Ahmadi F, Sadiqi M, et al. Thyroid gland dysfunction and its effect on the cardiovascular system: a comprehensive review of the literature. *Endokrynol Pol*. 2020;71:466–478.
86. Mentuccia D, Proietti-Pannunzi L, Tanner K, et al. Association between a novel variant of the human type 2 deiodinase gene Thr92Ala and insulin resistance: evidence of interaction with the Trp64Arg variant of the beta-3-adrenergic receptor. *Diabetes*. 2002;51:880–883.
87. Marsili A, Tang D, Harney JW, et al. Type II iodothyronine deiodinase provides intracellular 3,5,3’-triiodothyronine to normal and regenerating mouse skeletal muscle. *Am J Physiol Endocrinol Metab*. 2011;301:E818–E824.
88. Wojcicka A, Bassett JH, Williams GR. Mechanisms of action of thyroid hormones in the skeleton. *Biochim Biophys Acta*. 2013;1830:3979–3986.
89. Godart M, Frau C, Farhat D, et al. The murine intestinal stem cells are highly sensitive to the modulation of the T3/TRalpha1-dependent pathway. *Development*. 2021 148:dev194357.
90. Silva JE, Bianco SD. Thyroid-adrenergic interactions: physiological and clinical implications. *Thyroid*. 2008;18:157–165.
91. Sinha RA, Singh BK, Yen PM. Direct effects of thyroid hormones on hepatic lipid metabolism. *Nat Rev Endocrinol*. 2018;14:259–269.
92. Ribeiro MO, Bianco SD, Kaneshige M, et al. Expression of uncoupling protein 1 in mouse brown adipose tissue is thyroid hormone receptor-beta isoform specific and required for adaptive thermogenesis. *Endocrinology*. 2010;151:432–440.
93. Ribeiro MO, Carvalho SD, Schultz JJ, et al. Thyroid hormone--sympathetic interaction and adaptive thermogenesis are thyroid hormone receptor isoform--specific. *J Clin Invest*. 2001;108:97–105.
94. Safer JD. Thyroid hormone action on skin. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:388–393.
95. Mancino G, Miro C, Di Cicco E, et al. Thyroid hormone action in epidermal development and homeostasis and its implications in the pathophysiology of the skin. *J Endocrinol Invest*. 2021 44:1571–1579.
96. Safer JD. Thyroid hormone and wound healing. *J Thyroid Res*. 2013;2013:124538.
97. Contreras-Jurado C, Lorz C, Garcia-Serrano L, et al. Thyroid hormone signaling controls hair follicle stem cell function. *Mol Biol Cell*. 2015;26:1263–1272.
98. Miro C, Di Cicco E, Ambrosio R, et al. Thyroid hormone induces progression and invasiveness of squamous cell carcinomas by promoting a ZEB-1/E-cadherin switch. *Nat Commun*. 2019;10:5410.
99. Mitchell CS, Savage DB, Dufour S, et al. Resistance to thyroid hormone is associated with raised energy expenditure, muscle mitochondrial uncoupling, and hyperphagia. *J Clin Invest*. 2010;120:1345–1354.
100. Barca-Mayo O, Liao XH, Alonso M, et al. Thyroid hormone receptor alpha and regulation of type 3 deiodinase. *Mol Endocrinol*. 2011;25:575–583.

68

Thyroid Function Testing (Thyrotropin, Triiodothyronine, and Thyroxine)

James D. Faix

OUTLINE

Introduction, 1142
Thyrotropin, 1142
  Immunoassay for Thyrotropin, 1142
  Thyrotropin Reference Intervals, 1143
  Standardization of Thyrotropin Immunoassays, 1144
Free Thyroid Hormones, 1145
  Indirect Estimation of Free Thyroid Hormones, 1145
  Free Thyroid Hormones by Equilibrium Dialysis, 1145
  Immunoassay of Free Thyroid Hormones, 1146
  Free Thyroid Hormone Reference Intervals, 1147
  Standardization of Free Thyroid Hormones, 1147
Total Thyroxine and Triiodothyronine, 1148
Interference in the Measurement of Thyroid Hormones, 1149
Future Directions, 1149

INTRODUCTION

Thyroid disease is common, and the signs and symptoms of thyroid disease may resemble those of other disorders. Therefore, physicians frequently request so-called thyroid function tests even when they have a low suspicion of thyroid disease. Although these may include tests that directly assess thyroid gland activity, the most commonly requested thyroid function tests are those which determine the levels of thyrotropin (thyroid stimulating hormone [TSH]) and/or the individual thyroid hormones thyroxine (T4) and triiodothyronine (T3).

TSH is the key thyroid function test, because it is at the center of the hypothalamic-pituitary-thyroid axis. TSH is elevated if the thyroid gland fails to produce adequate thyroid hormone and suppressed if the thyroid gland produces too much. But there are other causes of an elevated or suppressed TSH, and measurement of thyroid hormone levels is often needed as well to verify that the cause of the abnormal TSH result is thyroid disease. Originally, this meant determining the total levels of T4 and T3, but the establishment of the so-called free hormone hypothesis—that the free or unbound hormone concentration, not the total, reflects thyroid hormone activity—resulted in assays designed to estimate the free T4 and free T3 fractions. There still is a role for measuring total T4 and T3, but free T4 is the most commonly requested test when the need to confirm the TSH level result arises.

This chapter describes measurement of the major thyroid function tests: TSH and free T4, as well as total T4 and T3 and free T3. The other laboratory tests that help in the diagnosis of thyroid disease (thyroglobulin and autoantibodies against the TSH receptor, thyroid peroxidase [TPO], and thyroglobulin) are described in the relevant chapters discussing the disorders they help diagnose.

THYROTROPIN

TSH is a glycoprotein hormone with a structure similar to other anterior pituitary glycoprotein hormones. They are noncovalently linked heterodimers of a common (alpha) chain and a unique (beta) chain, which contains the specific hormone's biological activity. TSH is considered the primary thyroid function test.[1] It is a useful screen for evidence of thyroid disorders and helps to monitor thyroid hormone replacement in patients with hypothyroidism, as well as thyroid hormone suppressive therapy in patients with thyroid carcinoma.

KEY POINTS

- Measurement of thyrotropin is the key thyroid function test, because it is at the center of the hypothalamic-pituitary-thyroid axis.

Immunoassay for Thyrotropin

The major advances in TSH testing involved moving from competitive to noncompetitive immunoassay approaches (Fig. 68.1). Routine measurement of TSH in clinical practice initially used competitive immunoassay techniques, in which TSH in the patient's serum competed with labeled TSH for anti-TSH antibody. Once separation of the bound from the unbound label was accomplished, the amount of label bound to antibody was inversely proportional to the concentration of TSH in the sample. These first-generation assays were useful for detecting elevated TSH levels, but they could not differentiate euthyroid from hyperthyroid patients with very low TSH levels. Noncompetitive immunoassays introduced in the 1980s achieved a level of analytical sensitivity that accomplished this differentiation (Fig. 68.2). These assays use one antibody to capture TSH and another, which is labeled, to bind to this complex. Once the unbound labeled antibody is removed, the amount of label bound to the solid phase is directly proportional to the concentration of TSH in the sample. This approach not only enhanced the ability to measure lower levels of TSH, but it expanded the measuring range on the high end as well. These second-generation TSH immunoassays quickly replaced the older ones. And by increasing the signal of the label on the labeled antibody and washing the solid-phase more efficiently, even greater analytical sensitivity was possible. Soon, so-called third-generation (and even fourth-generation) TSH assays were introduced. Today, almost all commercially available TSH immunoassays are capable of third-generation performance. Fourth-generation assays, including so-called single molecule immunoassays, are not widely used clinically.

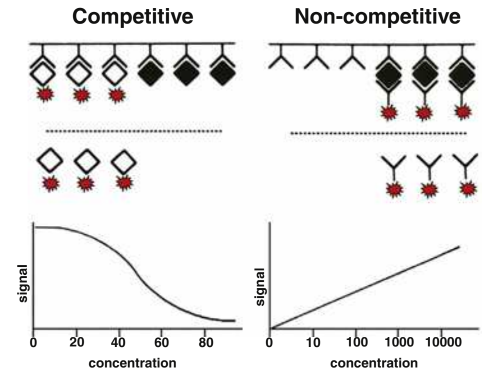


Fig. 68.1 Competitive versus noncompetitive immunoassay. The first immunoassays for thyrotropin (TSH) used a competitive approach (left) in which labeled TSH competes with the patient's hormone for a limited number of binding sites. The bound label is separated from the unbound label and the bound label (top) is inversely proportional to the patient's hormone concentration. Using a non-competitive approach (right) in which patient hormone is captured by bound and labeled antibodies, the bound label is proportional to the TSH concentration. This approach allows for measurement over a wider range and, more importantly, the ability to reliably detect lower levels of TSH is improved.

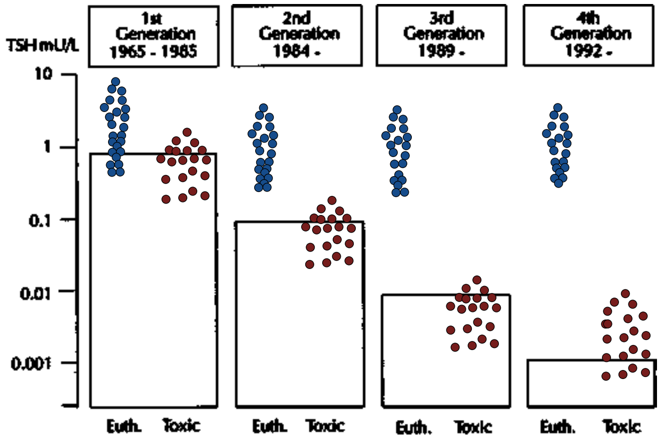


Fig. 68.2 Thyrotropin immunoassay generations. The first-generation assays were competitive immunoassays with limited analytical sensitivity. Differentiation of euthyroid (Euth) from hyperthyroid (Toxic) individuals was possible only when noncompetitive approaches were used. Enhancements improved the sensitivity of these second-generation assays, producing so-called third and even fourth generations, with even better separation.

KEY POINTS

- Today, almost all commercially available thyrotropin (TSH) immunoassays are capable of measuring TSH, with a reliable analytical sensitivity of 0.01 mU/L.

Performance of noncompetitive immunoassays for TSH is defined by so-called functional, rather than analytical, sensitivity. More modern terms for these distinctions are "limit of detection" (LOD, analogous to analytical sensitivity) and "limit of quantitation" (LOQ, analogous to functional sensitivity). The LOD is the lowest level of the TSH that can be distinguished from the background noise of the assay. The LOQ is the lowest level of TSH that can be measured with some degree of reliability, usually defined as reproducibility or precision over an extended period of time. The original LOQ target for TSH was the level below which the precision of the assay exceeded a coefficient of variability of 20%. By convention, second-, third-, and fourth-generation assays detect TSH with this degree of reproducibility down to levels of approximately 0.1, 0.01, and 0.001 mU/L, respectively. However, precision in the very low range for commercially available TSH immunoassays may not be as robust as it appears when functional sensitivity is initially evaluated. Also, many laboratories probably do not regularly monitor their performance in this range, and few proficiency programs regularly challenge laboratories at the limit of third-generation performance. Nonetheless, although the ability to better differentiate euthyroid from hyperthyroid patients is a useful feature, third-generation TSH assays are useful because the precision at higher levels (0.1–0.4 mU/L), where the results often impact clinical decisions, is markedly improved.

Thyrotropin Reference Intervals

Determining reference intervals for TSH is not straightforward. TSH does show diurnal variation, and significantly higher values have been recorded during the late evening and early night. However, because outpatient measurements are usually made during the day, this should have no significant impact on testing. Age does have an effect on TSH levels. Although only minimal changes occur during early adult life, observations that TSH levels rise with advanced age (without evidence of thyroid dysfunction) have raised the question of whether age-specific reference intervals should be implemented. In the absence of pregnancy, no significant gender differences have been observed. However, within these defined groups of euthyroid individuals, TSH levels do show significant interindividual variability.

There appears to be a consensus that the lower limit of the euthyroid reference interval for TSH should be approximately 0.4 mU/L. Therefore, the improved reproducibility in TSH measurement just below this cutoff is helpful. This range, representing potentially mild thyroid hormone excess but not overt hyperthyroidism, is an important one for clinical purposes. It may help to identify patients with subclinical hyperthyroidism and, when monitoring thyroid hormone replacement after thyroid ablation, to detect overtreatment with increased risk of atrial fibrillation. Significant hyperthyroidism due to Graves disease, functioning adenoma, or other causes is likely when the TSH level is truly "suppressed." This usually means less than 0.01 mU/L, because most commercially available TSH assays cannot reliably measure lower than this.

TSH assays are also especially important when hypothyroidism is suspected, because the inverse logarithmic relationship between TSH and T4 means that TSH levels will rise long before free T4 concentrations fall. Experts have disagreed with regard to the correct upper limit. The cutoff for an elevated TSH with first-generation TSH assays was approximately 10 mU/L. This fell to approximately 5 mU/L with the introduction of second- and third-generation assays. The major reason for this change was probably the reduced crossreactivity afforded by the use of monoclonal antibodies. A major problem with early TSH competitive immunoassays was crossreactivity with gonadotropins (luteinizing hormone, follicle-stimulating hormone, and human chorionic gonadotropin [hCG]), which share a common alpha subunit with TSH. For many years, most physicians have considered TSH levels greater than 10 mU/L as evidence of thyroid failure and levels between 5 and 10 mU/L as evidence of mild (or subclinical) hypothyroidism. During the past decade, however, there has been debate about the correct upper limit of the reference interval for TSH, as well as the approach to patients with subclinical hypothyroidism.

An analysis of thyroid function test results from a large survey of individuals chosen to represent the population of the United States (National Health and Nutrition Examination Survey) revealed that the mean TSH level in the general population was approximately 1.5 mU/L, with only a small degree of variability.[2] This finding prompted

TABLE 68.1 **Discrepancies Between Thyrotropin (TSH) and Free Thyroid Hormone Levels**

Elevated TSH Level Without Low FT4 or FT3 Levels	Low TSH Levels Without Elevated FT4 or FT3 Levels
Drugs that increase TSH concentration (cimetidine, dopamine antagonists, lithium, spironolactone, etc.)	Drugs that decrease TSH concentration (dopamine glucocorticoids, opiates, somatostatin, etc.)
Laboratory artifact (i.e., human antimouse antibody)	Laboratory artifact (i.e., biotin ingestion)
Nonthyroidal illness	Nonthyroidal illness
Recent increase in thyroid hormone replacement therapy	Recent decrease in suppressive thyroid hormone replacement
Subclinical hypothyroidism	Recent treatment for hyperthyroidism
Thyroid hormone resistance syndrome	Subclinical hyperthyroidism

FT3, Free triiodothyronine; *FT4,* free thyroxine.

organizations to call for lowering the upper limit of TSH to 4.0 mU/L, 3.0 mU/L, or even lower. These calls were resisted by many who worried that a significant number of patients would be unnecessarily labeled as having thyroid dysfunction, especially given the fact that there was no evidence that treatment of these individuals would provide any benefit. Because of concern that the population used to determine the range of TSH may have included individuals with occult thyroid disease, the data from the survey were reanalyzed to better clarify the relationship of TSH and antibodies to thyroid peroxidase (TPO), a recognized marker of autoimmune thyroid disease. TSH levels correlated with anti-TPO positivity, and the investigators found that TSH reference intervals determined with a population from which patients with occult autoimmune thyroid disorders had been excluded would indeed support the lower upper limit.[3] The issue of the upper limit for TSH remains controversial, especially because several studies since have demonstrated that most individuals with results slightly above these lower cutoffs have no evidence of thyroid disease.[4]

KEY POINTS

- There appears to be a consensus that the lower limit of the euthyroid reference interval for thyrotropin should be approximately 0.4 mIU/L, but experts have disagreed with regard to the correct upper limit.

Therefore, many other factors, including diet and nutrition, weight and body composition, genetic polymorphisms, and stressful stimuli apparently influence set points for the hypothalamic-pituitary-thyroid axis. A recently recognized factor, discussed later, that may account for some of the variability is the performance of the commercially available immunoassays for TSH. Several special situations have focused attention on other factors that influence TSH levels. One area that has always been problematic is pregnancy, which has a significant effect on thyroid function. TSH levels fall during the first trimester due to the thyroid-stimulating effect of hCG and then rise to baseline as hCG levels fall in the third trimester. Guidelines for the diagnosis and management of thyroid disease during pregnancy issued by the American Thyroid Association recommend trimester-specific reference intervals for TSH, as well as TSH targets for the diagnosis and treatment of thyroid disease during pregnancy.[5] As neonatal screening programs shift from T4 to TSH screening for congenital hypothyroidism, the appropriate cutoff for this indication is the subject of much debate. If screening is done during the first 24 hours after birth, programs must account for the neonatal surge in TSH production during this period. Also, even programs that screen at the traditional 2 to 4 days after birth are lowering the cutoff used in order to detect cases that would otherwise be missed.[6] Finally, a variety of stressful stimuli may suppress TSH production, including depression and other psychiatric diseases, sleep disorders, and the effects of severe systemic illness and/or major surgery (so-called nonthyroidal illness). Recovery from the latter may also result in a temporary elevation of TSH levels. Consequently, measurement of TSH in severely ill hospitalized patients should be done with caution.[7]

The inverse relationship between TSH and free T4 is log-linear in most individuals,[8] and this allows TSH to be considered as a single-test screen of thyroid function. In patients with primary hypothyroidism of whatever cause, levels should be elevated. The magnitude of serum TSH elevation correlates with the severity and, in part, the duration of thyroid hormone deficiency. Similarly, in patients with hyperthyroidism of whatever cause, levels should be lowered. If levels are below 0.01 mU/L (or the lowest limit reported by the laboratory), significant hyperthyroidism is probably present. TSH abnormalities may need to be further investigated using free T4 because of the variety of conditions that may cause TSH to be abnormal in the absence of either primary hypothyroidism or significant hyperthyroidism (Table 68.1).

KEY POINTS

- Thyrotropin (TSH) abnormalities may need to be further investigated using free T4, because a variety of conditions may cause TSH to be abnormal in the absence of either primary hypothyroidism or significant hyperthyroidism.

The hypothalamic tripeptide thyrotropin-releasing hormone (TRH) plays a central role in the regulation of pituitary TSH secretion. The TRH test measures the increase of pituitary TSH in serum in response to the administration of synthetic TRH. A direct correlation between basal serum TSH values and the maximal response to TRH has been observed. It could be useful in some situations, such as when there is suspicion of central hypothyroidism or a TSH-secreting pituitary adenoma. However, the use of modern sensitive immunoassays for TSH, plus the fact that TRH testing is no longer available in many areas, has decreased the need to perform this test.

Standardization of Thyrotropin Immunoassays

All strategies for defining abnormal TSH levels in different patient populations will suffer if all TSH immunoassays do not produce comparable results. Variability in the performance of TSH immunoassays has been apparent for several years. Data from proficiency testing programs have revealed poor correlation between observed method bias and the reference intervals recommended by the manufacturers (which are largely equivalent). Possible sources of variability include the specificity of the antibodies used, interferences from patient antibodies against the mouse monoclonal antibodies used, interferences from other endogenous substances such as biotin (for immunoassays that use biotin-streptavidin binding), and the TSH material used for calibration. TSH has three sites for the addition of oligosaccharides, and carbohydrate may constitute up to 25% of its mass. The pituitary releases a heterogeneous mixture of TSH glycoforms with regard to these side chains, and its nature changes when thyroid disease is present. Because the same assay is used to measure TSH in all patients, it is important that immunoassays measure TSH in an equimolar fashion.

All of these considerations were the impetus for the establishment of a Committee for Standardization of Thyroid Function Tests to study comparability of TSH assays (as well as total and free T4 and T3 immunoassays) under the auspices of the International Federation for Clinical Chemistry and Laboratory Medicine. Over the course of a decade, this committee worked with many *in vitro* diagnostic companies who provide TSH immunoassays to examine potential sources of variability, and potential solutions for standardization. Standardization of laboratory testing aims to ensure that all assays are traceable to a reference measurement procedure (RMP) and, consequently, all assays give equivalent results for any patient's specimen. The standard material for TSH is a World Health Organization International Reference Preparation obtained from purified cadaver pituitary. Because no RMP existed for TSH, the committee approached the problem using an alternative scheme called harmonization, in which a system consisting of methods and materials act as the reference. A panel of single-donation human blood serum specimens that spanned the TSH measurement range was tested by all manufacturers and TSH values were assigned to each sample using the "all-procedure trimmed mean." This target was used by assay manufacturers to recalibrate their individual methods. This approach has been well documented as a reliable way to ensure uniform results among different immunoassays.[9] The initial panel consisted of specimens from euthyroid individuals. The investigators showed that currently available TSH assays are precise, and that most agreed within plus or minus 10%. Recalibration was successful in reducing this variability. Recalibration with a panel of specimens from patients with thyroid disease was similarly successful and also demonstrated that there does not appear to be any effect of different TSH glycoforms on different assay performance. Finally, a study was conducted using a new panel of 120 specimens from presumably healthy individuals with negative testing for anti-TPO antibodies and TSH less than 10 mU/mL.[10] Recalibration was able to narrow the ranges of the reference intervals for each commercially available assay tested so that they showed much better agreement (Fig. 68.3). This indicates that TSH may be successfully harmonized in the near future, which may help the resolve some of the controversy regarding the appropriate upper limit to use for screening.

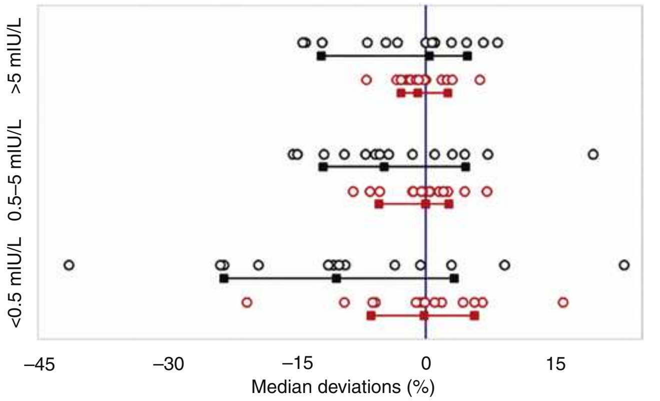


Fig. 68.3 **Harmonization of thyrotropin (TSH) immunoassays.** Median deviations of commercially available TSH immunoassays before *(black)* and after *(red)* recalibration using targets derived from a factored analysis of all of the results. The improvement is shown for specimens with low, normal, and high TSH levels. (From Thienpont LM, Van Uytfanghe K, De Grande LAC, et al. Harmonization of serum thyroid-stimulating hormone measurement paves the way for the adoption of a more uniform reference interval. *Clin Chem.* 2017;63:1248–1260. Reproduced with permission.)

FREE THYROID HORMONES

Thyroid hormones in the blood are bound to serum protein carriers, thus leaving only a minute fraction of free hormone in the circulation that is capable of mediating biological activities. A reversible equilibrium exists between bound and unbound hormone, and it is the latter that represents the fraction of the hormone capable of traversing cellular membranes to exert its effects on body tissues. This free hormone hypothesis forms the basis of the clinical approach to a number of endocrine disorders but this is especially true for thyroid disease. Most (approximately 99%) of the total T4 present in blood is bound, the majority to thyroxine-binding globulin (TBG). Even though it is present in a much lower concentration than the other major thyroid hormone–binding proteins (transthyretin and albumin), TBG has a significantly higher affinity for T4. Most experts agree that, in euthyroid individuals, the free T4 concentration is held constant, even when the concentration of the binding proteins change, while in thyroid disease it is abnormal. But estimating or measuring the very low concentration of free T4 has proved difficult over the years, and approaches have evolved.

Indirect Estimation of Free Thyroid Hormones

The first attempts to measure the free T4 concentration relied on an estimate of the number of binding sites available on the serum binding proteins. This was accomplished by adding labeled thyroid hormone to serum and, after incubation, using an absorbent material of some kind to collect all of the unbound labeled hormone. Labeled T3, for which TBG has a lower affinity, was used rather than labeled T4 in order to prevent the labeled hormone from displacing T4 bound to the protein. The so-called "T3 uptake" (T3U) test produced a result that correlated with the saturation of TBG with T4. If TBG saturation was high (as would be the case in hyperthyroidism), the uptake of labeled hormone by the absorbent would be high; if TBG saturation was low (as would be the case in hypothyroidism), the uptake of labeled hormone by the absorbent would be low. However, when the total T4 level was increased or decreased because of changes in TBG concentration, a baseline level of TBG saturation would have been achieved, and the uptake of the labeled hormone would be in the opposite direction. Calculation of a unitless figure called the free thyroxine index (FTI), the mathematical product of the total T4 (originally the protein-bound iodine) and the T3U, could thus be used to normalize changes in total T4 due to increases of decreases in TBG. Currently, labeled T4 analogs are used in place of T3 (so-called "T-uptake" assays), and another approach relies on measurement of TBG, with calculation of the T4/TBG ratio. Although FTI determinations were once commonly requested laboratory tests, they have been supplanted by more direct approaches, including equilibrium dialysis (or ultrafiltration) and immunoassay.

Free Thyroid Hormones by Equilibrium Dialysis

Early attempts to directly measure free T4 used equilibrium dialysis to physically separate free T4 from bound T4. But the initial approach was not direct but indirect. In direct equilibrium dialysis, free hormone in the specimen moves into a second compartment on the other side of a semipermeable membrane that excludes the protein-bound hormone. After an equilibrium between the two compartments is reached, the concentration of hormone in the second compartment is measured. However, because the free T4 is such a small percentage of the total, T4 concentrations in the second compartment were far

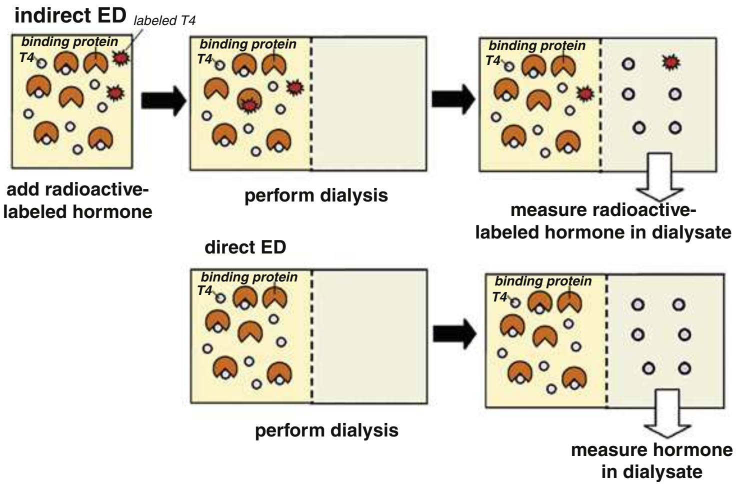

indirect ED
labeled T4
binding protein
T4
add radioactive-labeled hormone
perform dialysis
measure radioactive-labeled hormone in dialysate
direct ED
perform dialysis
measure hormone in dialysate

Fig. 68.4 **Equilibrium dialysis (ED) measurement of free thyroxine (T4).** The first approach to equilibrium dialysis for free T4 used an indirect approach *(top)* in which radioactively labeled T4 was added to the sample, and the amount of label in the dialysate was used along with the total T4 measured separately to estimate the free T4. In the modern direct approach *(bottom)*, the T4 in the dialysate is measured using mass spectrometry. Note that the amount of protein-bound T4 in actual samples would be much greater than the percentage shown, so disruption of the equilibrium between bound and free T4 would not be significantly disturbed.

below the ability of early immunoassays to measure it. Consequently, most investigators took an indirect approach by adding a trace amount of radioactive-labeled T4 to the specimen (Fig. 68.4). This labeled T4 equilibrated with the respective bound and free endogenous T4 and the free fraction could be calculated from the proportion of labeled hormone in the dialysate.

An equilibrium method that used immunoassay to directly measure T4 in the dialysate was introduced in the 1990s. Unfortunately, this product is no longer widely commercially available. More recently, a direct equilibrium method was developed and proposed as an RMP. The details of this procedure include strict requirements for the buffer used, the pH of the specimen, the dialysate membrane cutoff size, the size of the two compartments, and the temperature. The T4 in the dialysate is measured by isotope dilution–liquid chromatography/tandem mass spectrometry (LC-MS/MS), not immunoassay. Other alternative direct approaches, such as ultrafiltration followed by LC-MS/MS, have also been described. There is debate regarding which is the best approach, but either is probably too laborious to be implemented for routine clinical use. Therefore, most clinical laboratories rely on commercially available immunoassays.[11]

KEY POINTS

- A direct equilibrium method for measurement of free thyroxine has been developed and proposed as a reference measurement procedure.

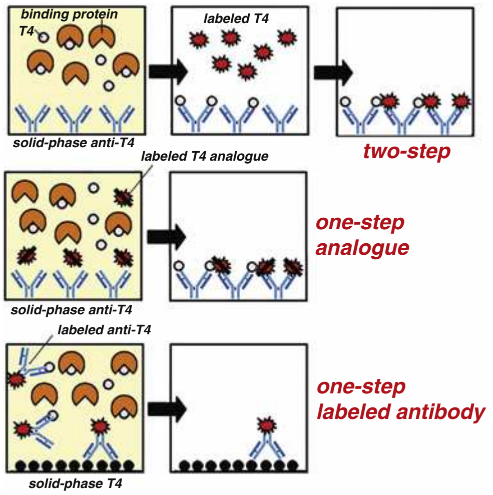


Fig. 68.5 **Free thyroxine (T4) immunoassay.** All of the commercially available immunoassays for free T4 use one of these three approaches (described in the text).

Immunoassay of Free Thyroid Hormones

There have been three formats used for free T4 immunoassay. All utilize competitive immunoassay with a solid support that can be washed, removing unbound labeled reagent. But each is very different (Fig. 68.5).

In the so-called two-step method, free hormone in the specimen interacts with antihormone antibody on a solid support, which becomes occupied by endogenous free hormone to a smaller or larger extent based on the free hormone concentration. The specimen is then removed, and excess labeled hormone is added, which fills any empty antibody binding sites. This is also often referred to as "back-titration." In the labeled analogue approach (also commonly referred to as a "one-step" method), free hormone in the specimen interacts with antihormone antibody on a solid support *in the presence of* labeled hormone. In order to prevent interaction of labeled hormone with binding proteins, the T4 used as the labeled hormone is altered in a way that is meant to interfere with binding to endogenous proteins but not with antibody binding. The nature of these so-called hormone "analogs" is proprietary. Finally, the "labeled antibody" method (which is also a "one-step" assay) avoids the use of labeled hormone entirely. Free hormone in the specimen interacts with labeled *antibody* in the presence of solid-phase hormone. Some critics have pointed out that this solid-phase hormone is also an "analog," but the fact that it is not able to enter the free hormone pool in the same way as the soluble analog probably limits its ability to compete with the endogenous free hormone for endogenous binding sites. For all three formats, the amount of labeled reagent left on the solid-phase after final washing is inversely proportional to the original concentration of free hormone in the specimen.[12]

There are significant discrepancies between different commercially available immunoassays for the measurement of free T4. Although different calibration may explain some of the overall differences (see later), the ways in which different assays influence the equilibrium between protein-bound and free hormone in individual patients may also explain many of the discrepancies. Although the labeled T4 analog used in one-step analog assays should not bind to any of the endogenous binding proteins, it initially appeared that some analogs did exhibit significant binding to albumin. Free T4 concentrations in patient specimens with albumin levels different from those of the calibrators used in the assay would either be overestimated or underestimated. Because patients who are critically ill often have low albumin levels, free T4 assays frequently underestimated the true free T4 concentration. However, most manufacturers have addressed this problem, either by adding albumin to the reagent (so that any variation in the patient's albumin level will be compensated) or by improving the behavior of the T4 analog.

Another problem is the influence of the total concentration of protein-bound T4, especially the fraction bound to TBG, because of its high affinity for T4, as well as the fact that it may be present in a wide range of concentrations. Generally speaking, immunoassays have been shown to underrecover free T4 when TBG levels are elevated, although the validity of the design of these investigations (using artificial T4 solutions with added T4-binding proteins) has been criticized. This limitation has often been cited as a reason to question the accuracy of free T4 concentrations in pregnant women, especially in the third trimester (see later).

Finally, for both one-step analog and two-step back-titration immunoassays, consideration must also be given to the amount of anti-T4 antibody used on the solid phase. It is very important that binding of free T4 to the anti-T4 antibody occur without significantly disturbing the equilibrium between free T4 and protein-bound T4. If this happens, protein-bound T4 may dissociate. This phenomenon is called sequestration. The antibody reagent should not remove more than 1% of the total T4 in the specimen. Even though this represents more than the free moiety, it represents only a small percentage of the total T4 present and is unlikely to significantly disturb the equilibrium. Dilution has been traditionally used to gauge the degree to which a free T4 immunoassay is susceptible to this effect, either because of the amount of reagent antibody present or the affinity of the antibody used. Assays are considered robust as long as this effect is kept to a minimum.[14]

Commercially available immunoassays for free T3 have been developed, using approaches similar to those used for free T4, and free T3 is often requested instead of total T3. The same issues that may influence assays for free T4 can impact assays for free T3. The Clinical Laboratory Standards Institute has issued an approved guideline for the measurement of free thyroid hormones.[13] Although it is unlikely that many clinical laboratories will be able to utilize the proposed RMPs for free T4 and free T3, they should be able to implement several of this guideline's recommendations for ensuring that free thyroid hormone results using commercially available immunoassays are as accurate and precise as possible. These include adding increasing amounts of albumin and TBG to specimens from euthyroid individuals, as well as diluting specimens from euthyroid individuals using inert buffer in order to assess possible analog binding, the effect of binding proteins, and the degree of antibody sequestering. The guideline also recommends verifying the performance of the free T4 immunoassay in specimens from the following groups of patients: euthyroid; hyperthyroid; hypothyroid; patients with elevated TBG concentrations; and patients who are taking drugs known to compete for T4 binding sites on TBG and albumin.

KEY POINTS

- Guidelines exist for ensuring that free thyroid hormone results using commercially available immunoassays are as accurate and precise as possible, but there are significant discrepancies between different commercially available immunoassays for the measurement of free thyroxine.

Free Thyroid Hormone Reference Intervals

Reference intervals for both free T4 and free T3 vary with the manufacturers of commercially available assays, but are approximately 0.8 to 2.0 ng/dL (10–26 pmoL/L) for free T4 and 0.2 to 0.4 ng/dL (3-6 pmol/L) for free T3. With few exceptions, the free T4 and free T3 concentrations are high in hyperthyroidism, low in hypothyroidism, and normal in euthyroidism, even in the presence of changes in TBG concentration. The condition that has raised the most concern about free thyroid hormone reference intervals is pregnancy. The influence of elevated TBG on immunoassays for free T4 is especially pronounced in pregnancy. Although so-called hypothyroxinemia of pregnancy does exist and needs to be recognized, euthyroid reference intervals in the second and third trimesters should probably be lower than previously thought. Guidelines of the American Thyroid Association for management of thyroid disease during pregnancy recommend that the laboratories develop trimester-specific reference intervals for the free T4 assay used. However, because of concerns about the performance of free T4 immunoassays, they also include a recommendation to use either equilibrium dialysis or ultrafiltration. Because neither direct approach is likely to be widely available, the guidelines suggest alternatives such as the use of total T4 (with trimester-specific reference intervals) or the free thyroxine index using T-uptake assays.[5] However, the use of either alternative suggested by the guidelines may not properly classify pregnant patients.[15] Immunoassays have been shown to correlate with equilibrium dialysis; although levels using immunoassay may be lower, the trending is similar,[16] but variability between different commercially available assays may classify patients differently.[17]

KEY POINTS

- There is probably still a lack of consensus regarding the use of free thyroxine assays in pregnancy.

Standardization of Free Thyroid Hormones

The same group that looked at harmonization of TSH measurements also examined commercially available immunoassays for free T4 and free T3, working in partnership with the same manufacturers. Except for one, the results were all considerably lower than those using the equilibrium dialysis RMP. However, the good news was that most of the bias appeared not to be related to any of the assay issues discussed previously, but rather to the calibration employed by the individual manufacturers. Mathematical recalibration using the regression of the results by each assay to the target values determined by the equilibrium dialysis RMP was able to remove most of the aforementioned considerable systematic bias (Fig. 68.6). This approach to standardization was pursued extensively for free T4 immunoassays in a similar fashion to that used for TSH and culminated with a similar study showing that removal of assay-specific bias may be able to align the different reference intervals to that of the RMP. The ability to share the RMP reference interval was not initially fully successful using a stringent requirement for the location of the upper and lower cutoffs to be within the confidence limits of the RMP. This was attributed to the susceptibility of free hormone immunoassays to a variety of individual

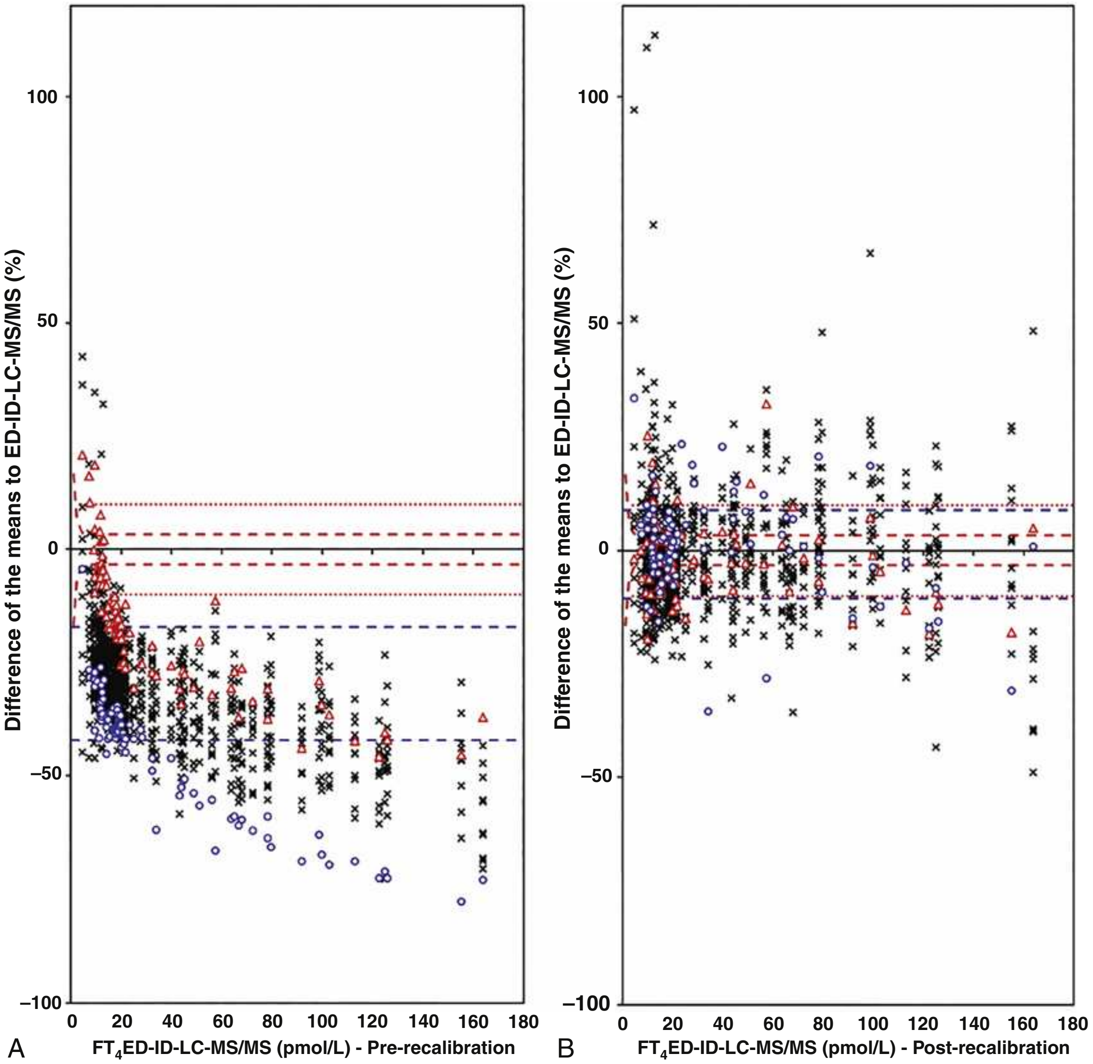


Fig. 68.6 **Standardization of free thyroxine (T4) immunoassays.** Comparison of commercially available immunoassays for free T4 to the equilibrium dialysis reference measurement procedure before *(left)* and after *(right)* recalibration. The initially most discrepant assays (high or low) are shown in red and blue, respectively. Although there was still significant variability for some of the assays, recalibration shifted the overall percentiles (15th and 18th) to within the bias expected based on biological variation *(red dotted lines). ED-ID-LC-MS/MS*, isotope dilution–liquid chromatography/tandem mass spectrometry; *FT_4*, free thyroxine. (From De Grande LAC, Van Uytfanghe K, Reynders D, et al. Standardization of free thyroxine measurements allows the adoption of a more uniform reference interval. *Clin Chem.* 2017;63:1642–1652. Reproduced with permission.)

sample related effects and/or variability in the lots of the reagents used by the manufacturers. However, using a slightly less stringent requirement, the hope for an eventual uniform reference interval for free T4 appeared within reach.[18]

TOTAL THYROXINE AND TRIIODOTHYRONINE

Although the primary focus of the use of thyroid function tests is the combination of TSH and free T4, there are still situations where the traditional measurement of the total level of T4 and T3 is helpful. The major secretory product of the thyroid gland is T4 in a ratio favoring T4 over T3 by 10- to 20-fold. Because they are present in higher concentrations and are not subject to the same specimen issues as free hormones, total T4 and T3 are easier to measure. Assays for total T4 and T3 use competitive immunoassay and include a reagent such as 8-anilinonaphthalene sulfonic acid that blocks T4 and T3 binding to serum proteins, so that total hormone is available for competition with the assay antibody.

The usual concentration of total T4 in adults ranges from 5 to 12 microg/dL (64–154 nmol/L). The usual concentration of total T3 in adults ranges from 80 to 190 ng/dL (1.2–2.9 nmol/L). Although levels of total T4 and T3 above and below these ranges are usually associated with hyperthyroidism and hypothyroidism, respectively, the total levels do not always correspond to the free hormone concentrations. When concentrations are below or above these ranges in the absence of thyroid dysfunction, they are usually the result of an abnormal level of serum TBG. Such abnormalities are commonly seen during pregnancy or the administration of estrogen-containing compounds, which result in a significant elevation of serum total T4 levels in euthyroid individuals. Similarly, total T4 levels are low in conditions that are associated with decreased TBG concentrations, but the free T4 is usually normal. The total T4 concentration in serum may be also be altered by alterations in T4 binding to TBG or compensatory changes due to defects in the conversion of T4 to T3. Although changes in transthyretin concentration rarely give rise to significant alterations in total T4 concentration, the presence of a variant serum albumin in patients with familial dysalbuminemic hyperthyroxinemia with high affinity for T4 produce apparent elevations in the measured total T4 concentration, whereas the free T4 levels remains normal.

The ratio of total T3 to total T4 may sometimes be helpful in the evaluation of patients with hyperthyroidism. A low ratio may favor thyroiditis or iodine-induced hyperthyroidism over Graves disease. A high ratio may be a helpful prognostic marker after initial of antithyroid treatment of Graves disease.

KEY POINTS

- Although the primary focus of the use of thyroid function tests is the combination of thyrotropin and free thyroxine (T4), there are still situations where the traditional measurement of the total level of T4 and triiodothyronine is helpful.

INTERFERENCE IN THE MEASUREMENT OF THYROID HORMONES

When there is discrepancy between the results of a laboratory test and the patient's clinical presentation, interferences with the assay itself must be considered. Traditional problems such as crossreactivity of the antibody used have largely been eliminated. Some interferences specific for TSH and free T4 immunoassays have been discussed, but there are several categories that may influence the results of all assays. Endogenous antibodies that react with any of the assay reagents may cause a falsely low or high result, depending on the format of the assay. This category includes human antimouse antibody, antibodies against the hormone being measured, or rheumatoid factor. Other endogenous substances may interfere in other ways. A recent focus has been the effect of biotin, used in the treatment of rare metabolic disorders but also taken in high amounts as a nutritional supplement. Immunoassays that use the binding of biotin to streptavidin in order to measure the bound label may produce erroneous results in patients with high levels, again in a way that is dependent on the format of the assay.[19] Finally, a large number of drugs may produce discrepant results. These effects are less likely to be assay interferences, but rather alterations in thyroid hormone physiology. These include agents that inhibit thyroid hormone synthesis and secretion, affect thyroid hormone binding proteins, or alter thyroid hormone metabolism.[20]

If some assay interference is suspected, a good first step is to repeat the testing using another manufacturer's method that has a different assay format. Diluting the specimen in the hopes of diluting out the effect of the interfering substance without affecting the measuring range of the assay or attempting removal of immunoglobulin in case the interference is due to an endogenous antibody may also help. When the presence of a drug interfering with endogenous thyroid function or metabolism is suspected, it may be necessary to temporarily withhold the drug and repeat the testing.[21]

FUTURE DIRECTIONS

TSH will continue to be the key thyroid function test when screening for thyroid dysfunction, and the efforts directed at harmonization of the commercially available assays will be successful. This will help investigators continue to evaluate the relationship between TSH levels and thyroid hormone activity in the general population and may allow the controversy regarding the appropriate upper limit of the reference interval to be resolved. Free T4, rather than total T4 or either total T3 or free T3, will continue to be used to confirm abnormal TSH results. Some have advocated more routine use of equilibrium dialysis for free T4 measurement, but this will not be feasible in the near future, and laboratories will continue to use commercially available immunoassays. The efforts directed at standardization of free T4 immunoassays will be hampered by several factors, including the various formats used, the differences in the influence of specimen-specific effects, and the fact that most of the current methods produce results that are significantly lower than the equilibrium dialysis RMP. Although standardization will allow for better patient care, a major education initiative will be needed if it is accomplished. Finally, trimester-specific reference intervals for thyroid function tests in pregnancy will be widely adopted.

REFERENCES

1. Schneider C, Feller M, Bauer DC, et al. Initial evaluation of thyroid dysfunction – are simultaneous TSH and fT4 tests necessary? *PLoS One*. 2018;13:e0196631.
2. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES II). *J Clin Endocrinol Metab*. 2002;87:489–499.
3. Spencer CA, Hollowell JG, Kazarosyan M, et al. National Health and Nutrition Examinations Survey III. Thyroid-stimulating hormone (TSH)-thyroperoxidase antibody relationships demonstrate that TSH upper limit reference limits may be skewed by occult thyroid dysfunction. *J Clin Endocrinol Metab*. 2007;92:4236–4240.
4. Feldt-Rasmussen U. Laboratory measurement of thyroid-related hormones, proteins and autoantibodies in serum. In: Braverman LE, Cooper DS, Kopp PA, eds. *Werner & Ingbar's the Thyroid: A Fundamental and Clinical Text*. Philadelphia: Walters Kluwer; 2021:267–299.
5. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the post-partum. *Thyroid*. 2017;27:315–389.
6. Kaplowitz PB. Neonatal thyroid disease. *Pediatr Clin N Am*. 2019;66:343–352.
7. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. *Thyroid*. 2014;24:1456–1465.
8. Rothacker KM, Brown SJ, Hadlow NC, et al. Reconciling the log-linear and non-log-linear nature of the TSH-free T4 relationship: Intra-Individual analysis of a large population. *J Clin Endocrinol Metabol*. 2016;101:1151–1158.
9. Stockl D, Van Uytfanghe K, Van Aelst S, et al. A statistical basis for harmonization of thyroid stimulating hormone immunoassays using a robust factor analysis model. *Clin Chem Lab Med*. 2014;52:965–972.
10. Thienpont LM, Van Uytfanghe K, De Grande LAC, et al. Harmonization of serum thyroid-stimulating hormone measurement paves the way for the adoption of a more uniform reference interval. *Clin Chem*. 2017;63:1248–1260.
11. Thienpont LM, Van Uytfanghe K, Poppe K, et al. Determination of free thyroid hormones. *Best Pract Res Clin Endo Metabol*. 2013;27:689–700.
12. Faix JD. Principles and pitfalls of free hormone measurements. *Best Pract Res Clin Endo Metabol*. 2013;27:631–645.
13. Thienpont L, Bunk DM, Christofides ND, et al. *Measurement of Free Thyroid Hormones*: Approved Guideline. Clinical and Laboratory Standards Institute (CLSI) document C45-A; Wayne, PA 2004.
14. Christofides ND. Free analyte immunoassay. In: Wild D, ed. *The Immunoassay Handbook*. 4th ed. Oxford: Elsevier; 2013:123–137.
15. Geno KA, Reed MS, Cervinski MA, et al. Evaluation of thyroid function in pregnant women using automated immunoassays. *Clin Chem*. 2021;67:772–780.

16. Anckaert E, Poppe K, Van Uytfanghe K, et al. FT4 immunoassays may display a pattern during pregnancy similar to the equilibrium dialysis ID-LC/tandem MS candidate reference measurement procedure in spite of susceptibility towards binding protein alterations. *Clin Chim Acta*. 2010;411:1348–1353.
17. Andersen SL, Christensen PA, Kronsgaard L, et al. Classification of thyroid dysfunction in pregnant women differs by analytical method and type of thyroid function tests. *J Clin Endo Metabol*. 2020;105:e4012–e4022.
18. De Grande LAC, Van Uytfanghe K, Reynders D, et al. Standardization of free thyroxine measurements allows the adoption of a more uniform reference interval. *Clin Chem*. 2017;63:1642–1652.
19. Bowen R, Benavides R, Colon-Franco JM, et al. Best practices in mitigating the risk of biotin interference with laboratory testing. *Clin Biochem*. 2019;74:1–11.
20. Burch HB. Drug effects on the thyroid. *New Eng J Med*. 2019;381:749–761.
21. Favresse J, Burlacu MC, Maiter D, et al. Interferences with thyroid function immunoassays: clinical implications and detection algorithm. *Endocrine Rev*. 2018;39:830–850.

69

Thyroid Imaging

Jennifer A. Sipos

OUTLINE

Ultrasonography, 1151
  Indications for Sonographic Assessment, 1151
  Physics of Ultrasonography, 1151
  Diffuse Thyroid Disorders, 1152
Primary Thyroid Lymphoma, 1153
  Subacute Thyroiditis, 1153
  Sonographic Risk Stratification Systems, 1154
  American Thyroid Association Sonographic Risk Stratification System, 1154
  Korean Thyroid Imaging Reporting and Data System, 1156
  American College of Radiology Thyroid Imaging Reporting and Data System, 1156
  Diagnostic Performance of Sonographic Risk Stratification Systems, 1157
  Specialized Sonographic Techniques, 1157
Scintigraphy, 1159
  Thyroid Nodules, 1159
  Diffuse Thyroid Disorders, 1160
  Congenital Hypothyroidism, 1161
  Perchlorate Discharge Test, 1161
Computed Tomography And Magnetic Resonance Imaging, 1161
Incidentalomas, 1162
  Computed Tomography and Magnetic Resonance Imaging, 1162
  (^{18}F)-Fluorodeoxyglucose–Positron Emission Tomography, 1162
Future Directions, 1163

ULTRASONOGRAPHY

Indications for Sonographic Assessment

Thyroid nodules are a common clinical entity, with up to 70% of patients over the age of 70 years harboring this neoplasm.[1] While the majority of nodules represent a benign neoplasm, and most malignant lesions are associated with an excellent prognosis, up to a quarter of patients with carcinoma have larger, higher-stage tumors at initial diagnosis.[2] Thus, it is incumbent upon the clinician to triage nodules efficiently and effectively. Current guidelines, therefore, uniformly recommend sonography as the initial step in the evaluation of a patient suspected of having a thyroid nodule.[3-5]

The physical examination of the neck is an insensitive means of identification of thyroid nodules. Palpation alone only identifies nodules in approximately 5% of patients,[1] whereas screening ultrasound (US) may reveal a nodule in over 37% of patients, with incidence escalating with age.[6] When a nodule is suspected on exam, US is essential to confirm the exam finding. One retrospective study of patients with suspected thyroid nodules identified on palpation found no corresponding lesion in 16% of patients. Further, another nodule that was not palpated was identified elsewhere in the gland in 23% of patients.[7] In this particular study, the US findings changed the management plan in 63% of patients.[7]

Physics of Ultrasonography

It is important to have at least a basic understanding of the verbiage used in thyroid sonography and the physics behind the findings. Echogenicity is essentially describing the color of a nodule or the thyroid itself with specific modifiers to compare to surrounding normal anatomy. Isoechoic describes echogenicity that is of the same color as normal thyroid tissue. The term hypoechoic is used to describe a lesion or gland that is darker than normal thyroid tissue, whereas hyperechoic is brighter. Echotexture is a term used to describe the pattern of tissue being examined; a normal thyroid is defined as having a homogeneous echotexture (Fig. 69.1A). In contrast, the echotexture of a gland involved with lymphocytic thyroiditis may have a heterogeneous echotexture (see Fig. 69.1B). This concept will be discussed in greater detail in the section later on diffuse thyroid disorders.

The sonographic appearance of a normal thyroid is one of a uniformly medium gray echotexture that is midway in hue between the lighter fascia and the darker musculature surrounding the gland (Fig. 69.e1). A healthy thyroid appears relatively bright on US due to its follicular architecture, which is composed of lakes of colloid surrounded by a thin layer of follicular cells. The degree of reflected sound waves directly impacts the brightness of the resulting image; greater reflection results in a brighter image. As the US wave travels through soft tissues of any type, a portion of the sound wave is reflected back to the probe when it encounters the interface of two tissues with unequal acoustic impedance (or resistance to the US wave traveling through a tissue). There is a dramatic mismatch in acoustic impedance between follicular cells and colloid; consequently, a large fraction of the US wave is reflected back to the probe when a sound wave encounters a follicle. Because there are innumerable follicles in a healthy thyroid, there is a significant amount of reflection of the US wave back to the probe. The result is a relatively bright-appearing thyroid. In contrast, tissue that is relatively homogeneous in histologic composition will have minimal reflection of the US wave back to the probe. The resultant image is uniformly dark. Understanding the physics behind this phenomenon of acoustic impedance is important when examining thyroid tissue and nodules, as it provides insight into the underlying pathophysiology of the disease process (Fig. 69.e2).

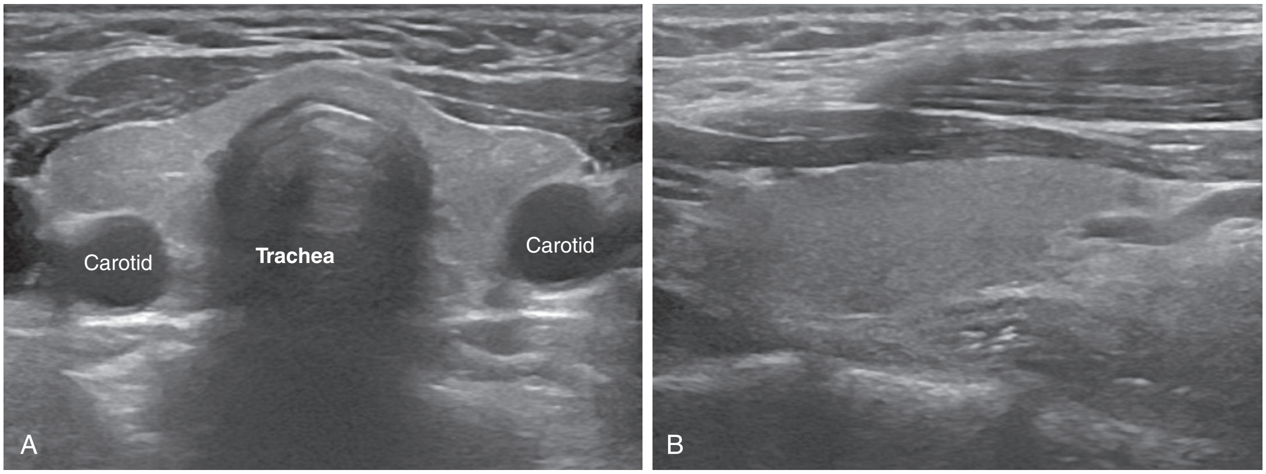


Fig. 69.e1 A, Diagnostic image of an ultrasound of a normal thyroid gland on transverse imaging with labels identifying the surrounding anatomic landmarks. B, Diagnostic image of an ultrasound of a normal thyroid gland in the sagittal view.

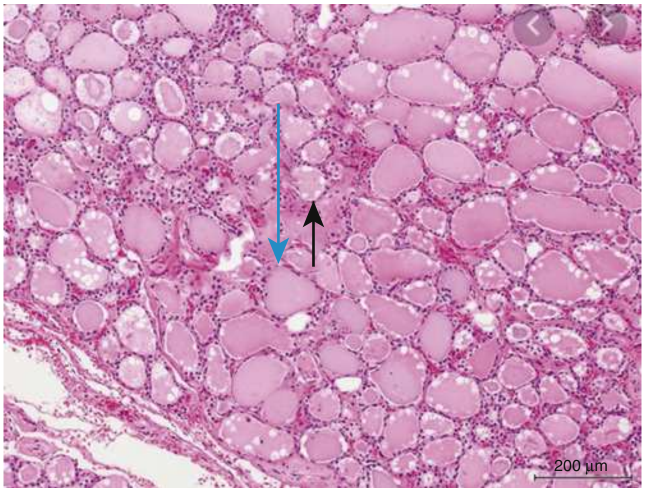


Fig. 69.e2 Photomicrograph of normal thyroid follicular architecture. The larger, blue arrow represents ultrasound waves traveling from the probe. When it encounters the follicle, some of the US wave is reflected back to the probe due to the change in acoustic impedance represented by the smaller, black arrow.

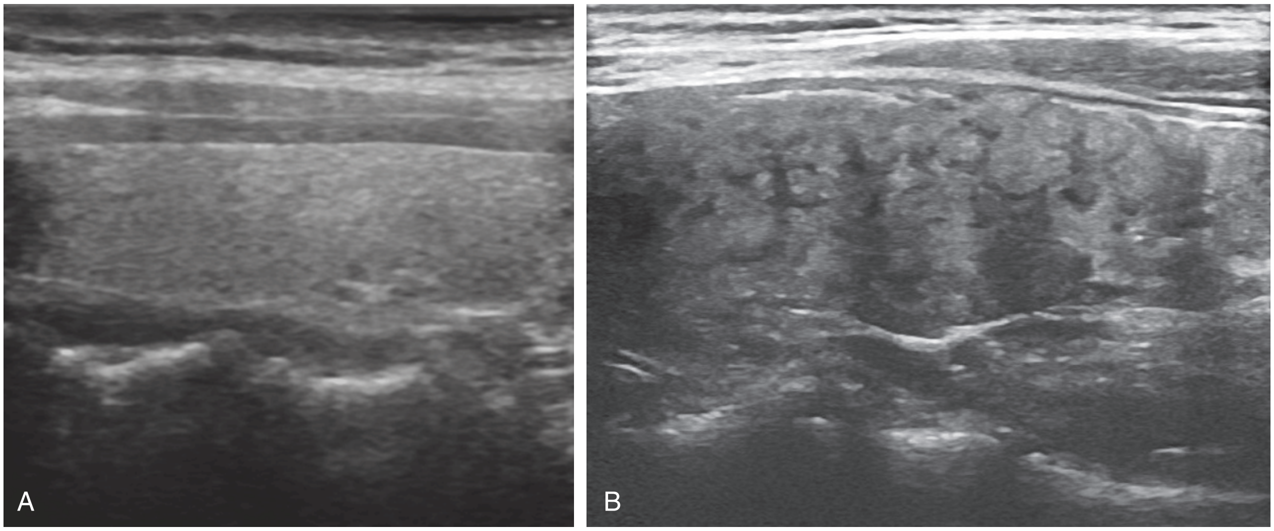


Fig. 69.1 A, Diagnostic image of a normal thyroid ultrasound in the sagittal view. Notice the medium gray color of the gland compared to the surrounding musculature and fascia. B, Diagnostic image of a thyroid ultrasound showing Hashimoto's thyroiditis. The thyroid is diffusely heterogeneous in echotexture. This diffuse micronodular appearance has been called a "giraffe pattern."

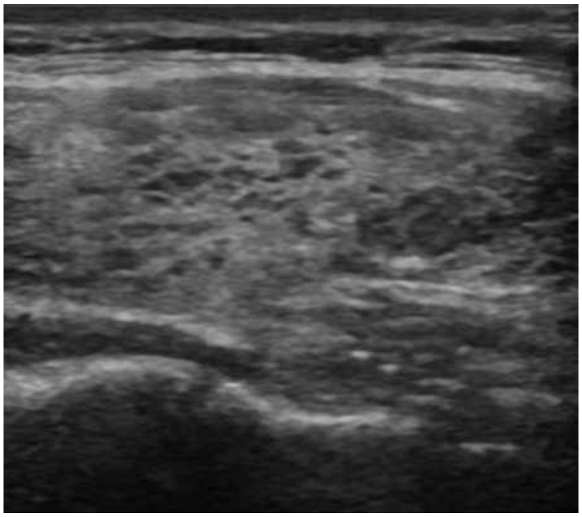


Fig. 69.2 Diagnostic image of a thyroid ultrasound showing Hashimoto's thyroiditis. The gland is heterogeneous with a microcystic echotexture due to small areas of germinal centers.

Diffuse Thyroid Disorders

Lymphocytic Thyroiditis. Lymphocytic thyroiditis is readily visible on sonography; the identification of diffuse heterogeneity in thyroid echotexture is a cardinal feature. Additional descriptors that may be reported in lymphocytic thyroiditis include a micronodular or microcystic echotexture (Fig. 69.2). The lymphocytic infiltrate produces a change in the histologic architecture of the normal thyroid tissue described earlier. This destruction of the normal follicular unit and replacement with sheets of lymphocytes alters the dynamics of US waves as they travel through the gland. The lymphocytic infiltrates appear as dark areas on US because the acoustic interface of follicular cells and colloid is replaced with a more uniform inflammatory process. Consequently, the US wave is not reflected back to the probe with the same intensity, and the lymphocytic infiltrate appears dark. In the early stages of the disease, the spared areas of the gland appear isoechoic, while the germinal centers appear dark, resulting in an overall heterogeneous echotexture of the gland (see Fig. 69.1B). As the disease progresses, the entirety of the gland is encompassed by the inflammatory process and may become diffusely hypoechoic (Fig. 69.e3A–C).

Hashimoto's Thyroiditis. US of the neck can have an important role in the diagnosis of Hashimoto's thyroiditis. Reliance on clinical and biochemical findings alone to diagnose Hashimoto's thyroiditis would miss half of patients with disease; including the sonographic analysis significantly improves the diagnostic sensitivity and specificity.[8,9] The characteristic sonographic appearance of Hashimoto's thyroiditis includes a diffusely heterogeneous parenchyma with hypoechoic areas interspersed with normal thyroid tissue; identification of this US feature has a positive predictive value of 95%[10] (see Fig. 69.1B). The vascularity of the gland may be variable, but is often increased early in the disease course. In the later stages of the disease, there may be a reduction in the vascularity with eventual avascularity as the degree of fibrosis increases (Fig. 69.e4). Echogenic septations, caused by fibrosis of the parenchyma, may be seen throughout the gland.

Another important feature associated with Hashimoto's thyroiditis includes surrounding reactive or hypertrophic lymphadenopathy. These nodes appear plump and may have a characteristic hilar stripe. They are typically seen anterior to the trachea and inferior to the thyroid gland, though they may also be seen in the lateral neck at levels II and IV (Fig. 69.e5).

KEY POINTS

- Hashimoto's thyroiditis is easily identified on ultrasound. The gland is marked by varying degrees of hypoechogenicity, with pseudonodules and septations. The presence of surrounding reactive adenopathy is also a clue to the underlying diagnosis of Hashimoto's.

Graves Disease. The sonographic appearance of Graves disease is similar to that seen in Hashimoto's thyroiditis but tends to be more diffusely involved; the micronodular pattern is less commonly seen (Fig. 69.3A). Generally, the gland is enlarged with a diffuse hypoechoic echotexture. Increased vascularity is a prominent component of the examination, particularly in untreated glands. This sonographic

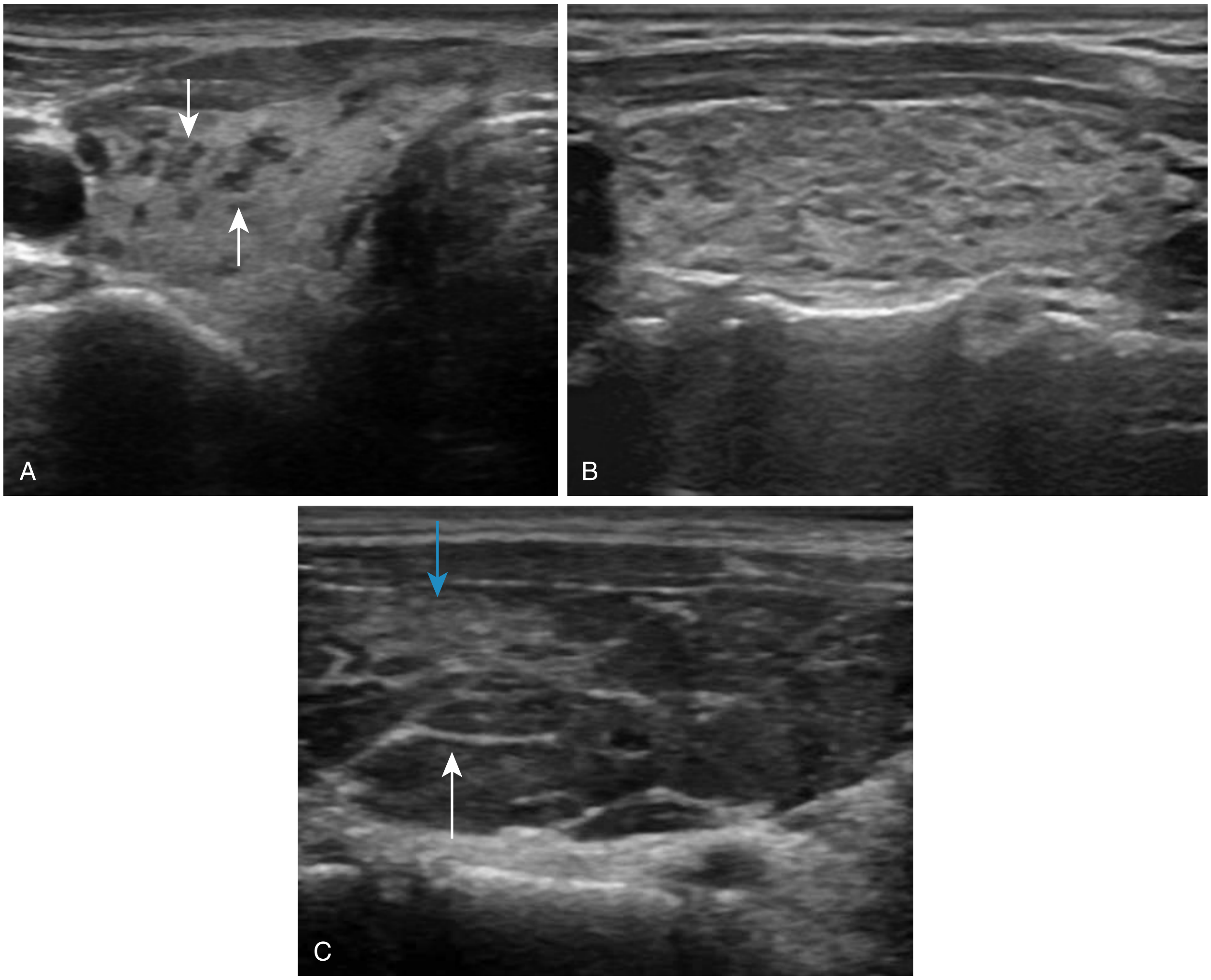


Fig. 69.e3 A, Diagnostic image of an ultrasound demonstrating early stage Hashimoto's thyroiditis with a microcystic or "moth-eaten" pattern marked by the small white arrows. B, Diagnostic image of an ultrasound of the thyroid illustrating progression of lymphocytic thyroiditis. Notice the more diffuse involvement of the lobe and the microcystic pattern. C, Diagnostic image of an ultrasound of a thyroid gland with chronic lymphocytic thyroiditis (Hashimoto's thyroiditis). The gland is diffusely hypoechoic with bright linear bands (white arrow), which represent fibrotic septations. Anteriorly an area of hyperechogenicity represents a pseudonodule with an area of more 'normal appearing' thyroid parenchyma (blue arrow).

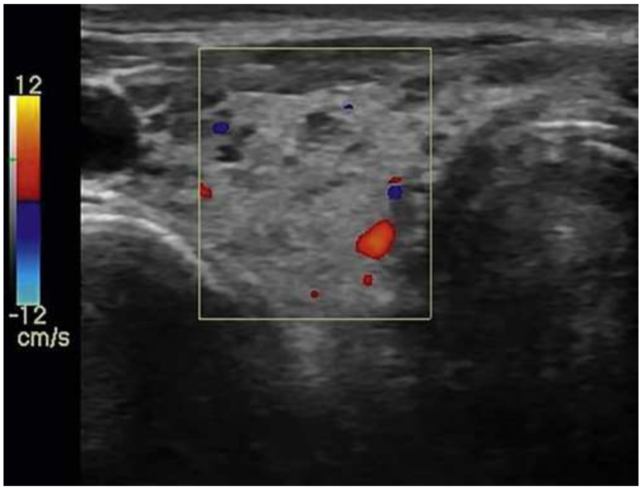


Fig. 69.e4 Image of Hashimoto's thyroiditis demonstrating reduced vascular flow.

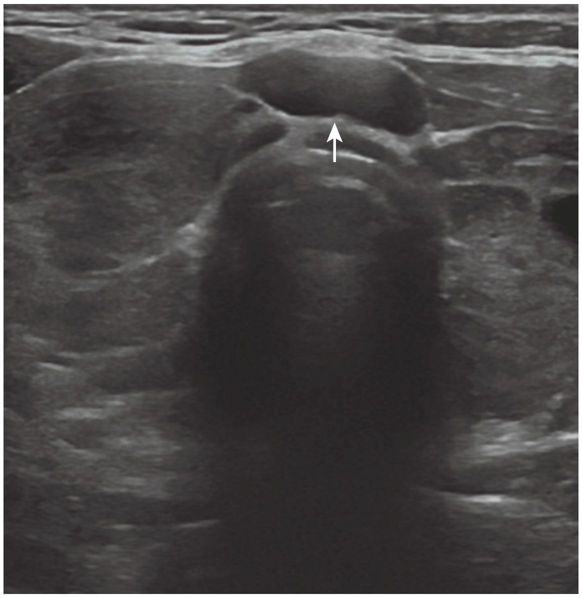


Fig. 69.e5 Diagnostic image of a thyroid ultrasound showing a hypertrophic delphian node in Hashimoto's thyroiditis (white arrow). Notice the hypoechogenicity of the thyroid gland and its similarity in color to the surrounding strap musculature.

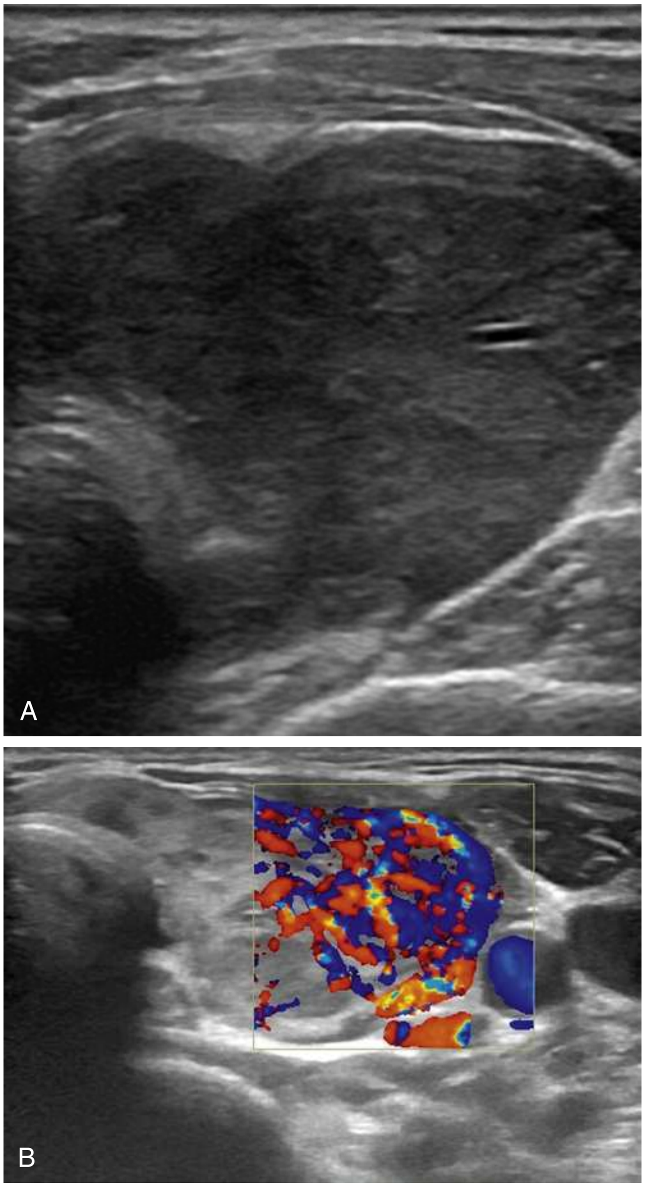


Fig. 69.3 A, Diagnostic image of a thyroid ultrasound showing lymphocytic thyroiditis involving the right lobe of the thyroid. Examination of the left lobe reveals marked enlargement and diffuse hypoechogenicity of the gland. Findings are consistent with primary thyroid lymphoma. B, Diagnostic image of a thyroid ultrasound demonstrating Graves disease. The gland is diffusely heterogeneous with markedly increased Doppler flow, also called thyroid inferno.

finding explains the bruit that often may be heard on physical examination and has been termed "thyroid inferno" when seen on US (see Fig. 69.3B).[11]

Thyroid nodules may be identified in up to 35% of patients with Graves disease when screening US is performed.[12] There is significant debate regarding whether nodules harbor an increased risk of malignancy in the setting of coexistent autoimmune thyroiditis.[13,14] Similar controversy exists regarding the prognostic significance of those nodules that are malignant; some studies suggest that the cancer is more aggressive with simultaneous Graves disease,[15] while others found no difference in clinical outcomes.[16] Although US identifies more nodules and cancers than palpation or scintigraphy,[17] screening US examination for all patients with Graves disease currently is not recommended.[18] The majority of cancers identified with screening sonography are subcentimeter tumors that have questionable clinical significance,[16] and the diagnostic evaluation can lead to increased morbidity if surgery is pursued.[19,20]

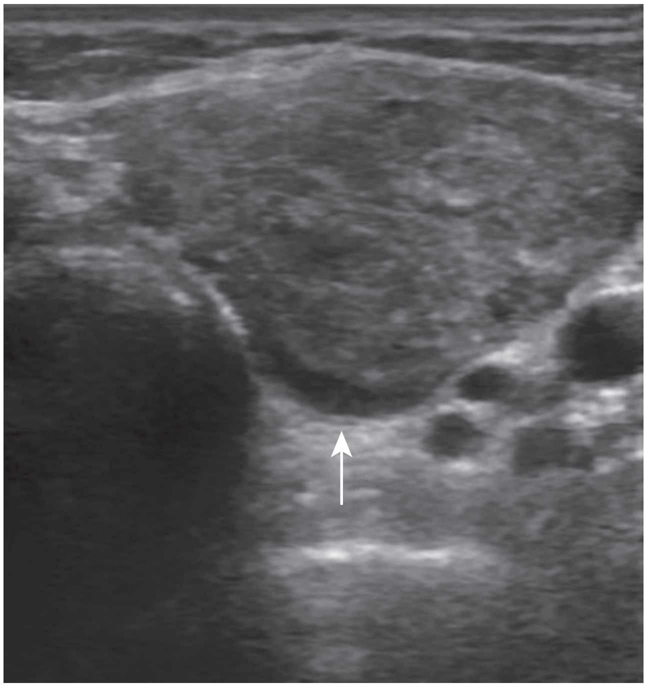


Fig. 69.4 **Diagnostic image of a thyroid ultrasound showing Graves disease.** The gland is diffusely hypoechoic and mildly enlarged with a prominent inferior thyroidal vein *(white arrow).*

PRIMARY THYROID LYMPHOMA

Primary thyroid lymphoma is a rare malignancy that represents less than 2% of all thyroid carcinomas.[21] It is more commonly seen in White women over the age of 60 years with a prior history of Hashimoto's thyroiditis.[22] Fortunately, most cases are diagnosed at an early stage and are associated with an excellent prognosis, but it is important to recognize the cardinal sonographic features of primary thyroid lymphoma, as a delay in diagnosis can negatively impact survival.[22] The clinical features are generally the first clue to the diagnosis; patients describe a several-week history of rapidly enlarging neck mass that may cause worsening dysphagia and a pressure sensation. Some patients also report "B symptoms" such as night sweats, fever, and weight loss.[23] Largely occurring in the setting of preexisting Hashimoto's thyroiditis, the US appearance can appear as an extreme manifestation of lymphocytic thyroiditis. A uniform, markedly hypoechoic gland with significant posterior acoustic enhancement that is significantly enlarged should increase the suspicion for primary thyroid lymphoma (Fig. 69.4). Depending on the stage and extent of disease, all or part of the gland may be involved. In higher-stage disease, the presence of enlarged locoregional adenopathy may also be seen.[24]

Subacute Thyroiditis

Subacute thyroiditis is a self-limited inflammatory disease that occurs in response to an antecedent viral illness. Patients present with acute neck pain and sudden-onset (often bilateral) thyromegaly, as well as fatigue and fever. In the acute phase, patients also may experience symptoms of thyrotoxicosis. Though the clinical presentation and significant anterior neck pain often is sufficiently diagnostic, US can aid in the distinction from other causes of thyrotoxicosis. The typical sonographic findings include focal, nonovoid areas of markedly

TABLE 69.1 American Thyroid Association Risk Stratification System

Risk Category	Description	Malignancy Risk	Threshold for FNA
High	Solid hypoechoic nodule or solid hypoechoic component in partially cystic nodule with one or more suspicious features*	>70%–90%	≥1 cm
Intermediate	Hypoechoic solid nodule with smooth margins lacking suspicious features**	10%–20%	≥1 cm
Low	Isoechoic or hyperechoic solid or partially cystic with eccentric solid areas lacking suspicious features**	5%–10%	≥1.5 cm
Very low	Spongiform or partially cystic without suspicious features	<3%	Consider ≥2 cm or observe
Benign	Purely cystic (no solid component)	<1%	No FNA

*Irregular margins (infiltrative, microlobulated), microcalcifications, taller-than-wide shape, rim calcifications with small extrusive soft tissue component, evidence of extrathyroidal extension.

**Microcalcifications, extrathyroidal extension, or taller-than-wide shape.

FNA, Fine-needle aspiration.

decreased echogenicity with significantly reduced vascularity (Fig. 69.e6A–E).[25] These poorly defined, hypoechoic areas may be seen in bilateral lobes and may be migratory if followed with serial examinations.[26] The inflammatory component typically produces enlargement of the gland.[27] Upon recovery from the disorder, there is characteristically complete resolution of sonographic abnormalities and return to normal thyroid volume.[28]

Sonographic Risk Stratification Systems

A number of sonographic risk stratification systems (SRSSs) have been developed to aid the clinician in assessing the need for fine-needle aspiration (FNA) and/or surveillance imaging upon identification of a thyroid nodule. The three main SRSSs presently in use are the American Thyroid Association (ATA), the American College of Radiology Thyroid Imaging Reporting and Data System (ACR-TIRADS), and the Korean Thyroid Imaging Reporting and Data System (K-TIRADS).[3,29]

American Thyroid Association Sonographic Risk Stratification System

Individual sonographic features have varying degrees of specificity for malignancy. The three individual features with the highest specificity for malignancy include a taller-than-wide shape (nonparallel orientation), presence of microcalcifications, and irregular (infiltrative, speculated, or microlobulated) margins; each of these features carries a median risk of malignancy of nearly 90%.[30] Independently, these high-risk features lack sufficient sensitivity, however, to effectively rule out malignancy in their absence. Furthermore, reliance on these individual features to assign malignancy risk is fraught with interobserver variability.[31] Consequently, the ATA system (Table 69.1) employs a pattern-based approach to the stratification of nodules.[3] This reliance on the assembly of sonographic features is more robust in terms of the reproducibility of the risk assessment.[32]

The ATA SRSS stratifies nodules into one of five risk categories. The benign category of nodules is comprised of simple cysts (Fig. 69.5). Because these lesions are benign, they do not require FNA for diagnostic purposes, regardless of nodule size. Therapeutic aspiration may be pursued if there are compressive symptoms or cosmetic concerns regarding their appearance. It is important to note, however, these lesions often reaccumulate fluid after such aspiration. The utility and interval for follow-up imaging of cystic nodules is unclear; if surveillance US is repeated, the interval should be greater than or equal to 24 months.[3]

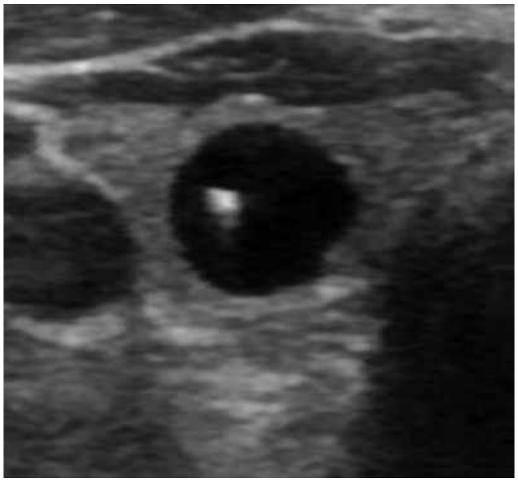


Fig. 69.5 **Diagnostic image of a thyroid ultrasound showing an American Thyroid Association benign nodule.** This nodule is a simple cyst.

The very low-risk suspicion nodules (Fig. 69.6) are comprised of spongiform nodules and partially cystic nodules without any suspicious US features (irregular margins, microcalcifications, taller-than-wide shape, or evidence of extrathyroidal extension). Spongiform nodules are defined as an isoechoic nodule with a microcystic aggregate that comprises greater than 50% of the lesion. The risk of malignancy for this category is less than 3%. Because of the low risk of malignancy for these nodules, a threshold of greater than or equal to 2 cm for FNA is recommended. Observation without FNA is also considered a reasonable option, depending on the clinical scenario. For those nodules that are aspirated and found to harbor benign cytology, the need for surveillance imaging is unclear, given the very low risk of missed malignancy. If US is repeated, the interval may be greater than or equal to 24 months.[3]

The ATA low-suspicion category includes isoechoic or hyperechoic solid nodules with no suspicious US features, or partially cystic nodules with an eccentric solid portion that lacks the suspicious features described above (Fig. 69.7).[3] Only 5% to 10% of these nodules are malignant, most of which are a follicular-patterned tumor (follicular carcinoma, follicular variant of papillary carcinoma, and noninvasive follicular tumor with papillary-like nuclear features [NIFTP]). Recognizing that metastatic spread is rare among follicular carcinomas under 2 cm, these lesions are recommended for aspiration at or above a size threshold of 1.5 cm. If aspirated and cytologically benign, these

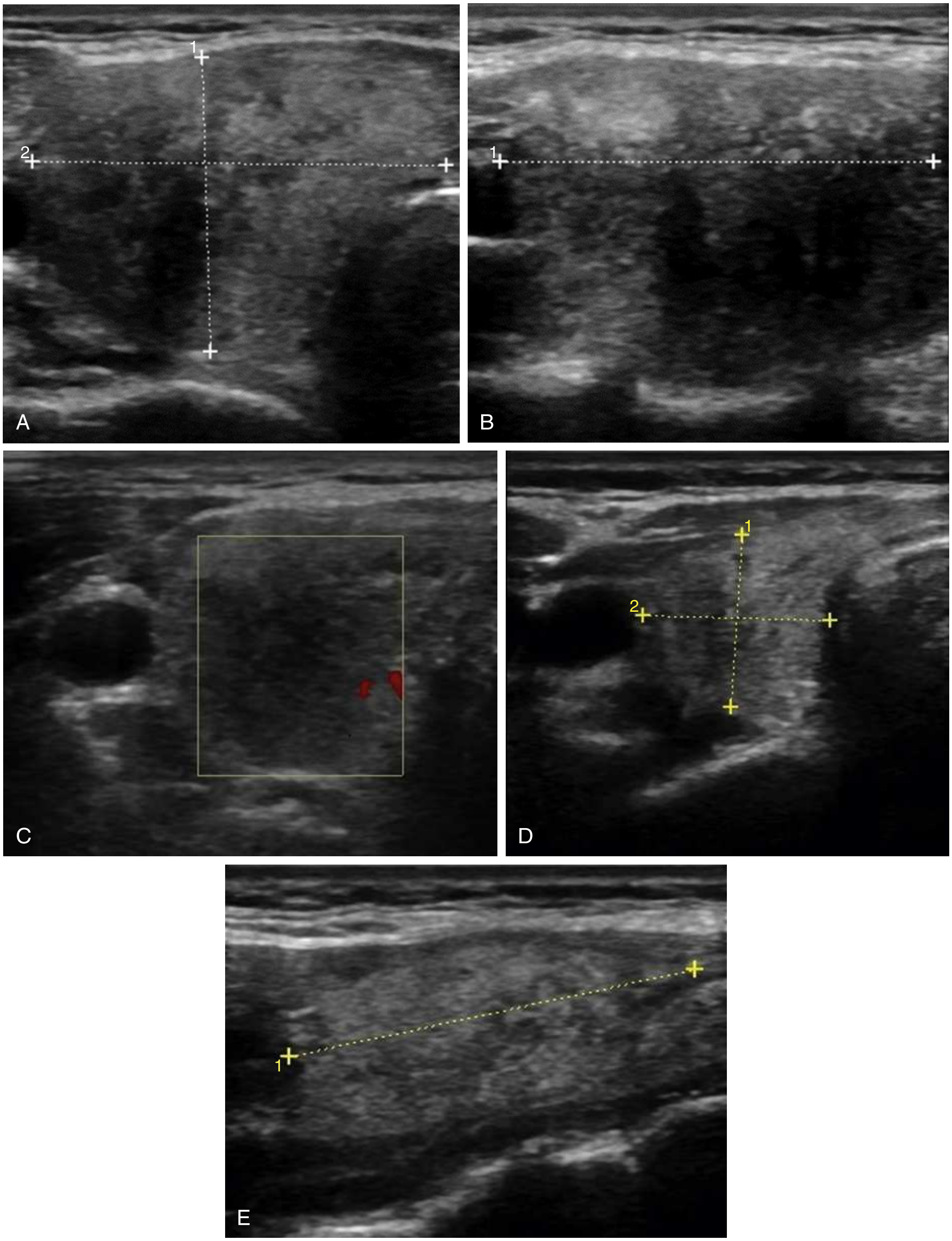


Fig. 69.e6 A, Diagnostic image of thyroid ultrasound showing the acute phase of subacute thyroiditis in the transverse view. The poorly defined markedly hypoechoic areas in the lower left corner of the image represent diffuse inflammation. Also note that the gland is significantly enlarged. B, Diagnostic image of a sagittal view of the thyroid ultrasound demonstrating the acute phase of subacute thyroiditis. The markedly hypoechoic area in the lower portion of the image represents the significant inflammatory component. C, Diagnostic image of color Doppler flow in DeQuervain's (subacute) thyroiditis. The significantly reduced Doppler flow is typical of subacute thyroiditis. D, Diagnostic image of thyroid ultrasound demonstrating the resolving phase of subacute thyroiditis. Repeat imaging of the same patient two months later reveals a significantly brighter echogenicity of the thyroid compared to image e4a. The gland is also returning to a more normal size as the inflammation wanes. E, Diagnostic image of thyroid ultrasound demonstrates the resolving phase of subacute thyroiditis.

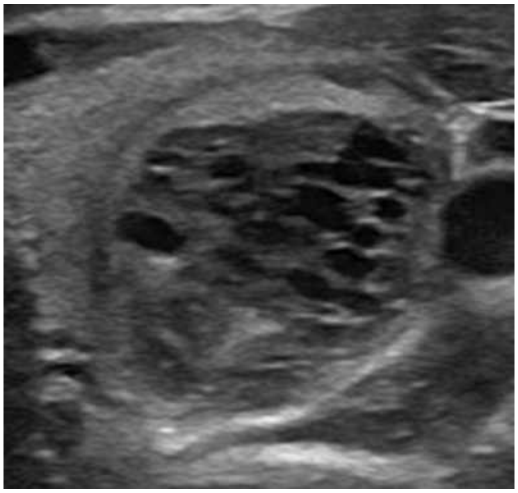


Fig. 69.6 **Diagnostic image of a thyroid ultrasound showing an American Thyroid Association very low–suspicion nodule.** This is an example of a spongiform nodule.

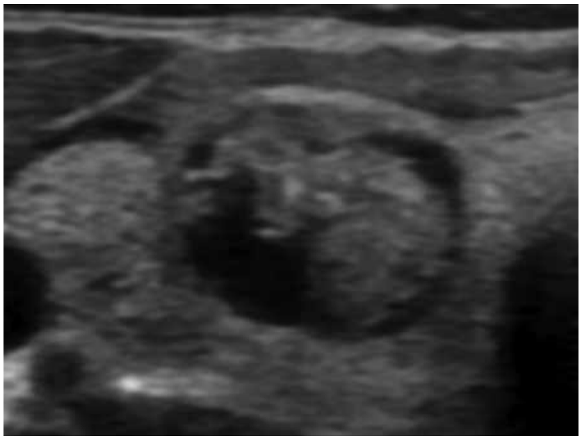


Fig. 69.7 **Diagnostic image of a thyroid ultrasound demonstrating an American Thyroid Association low-suspicion nodule.** This nodule is partially cystic with no suspicious ultrasound features.

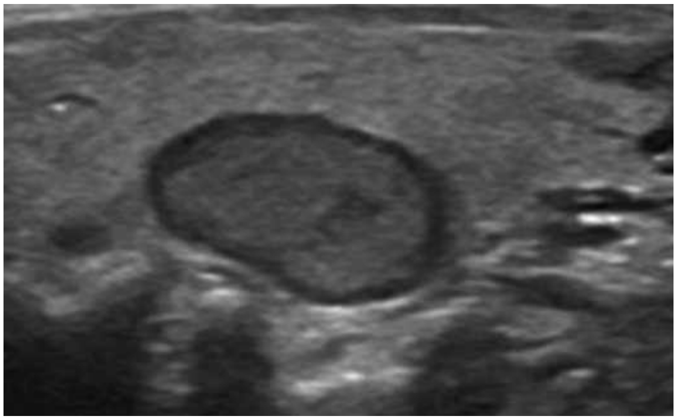


Fig. 69.8 **Diagnostic image of a thyroid ultrasound showing an American Thyroid Association intermediate-suspicion nodule.** This is a hypoechoic solid nodule with regular margins and lacking suspicious features.

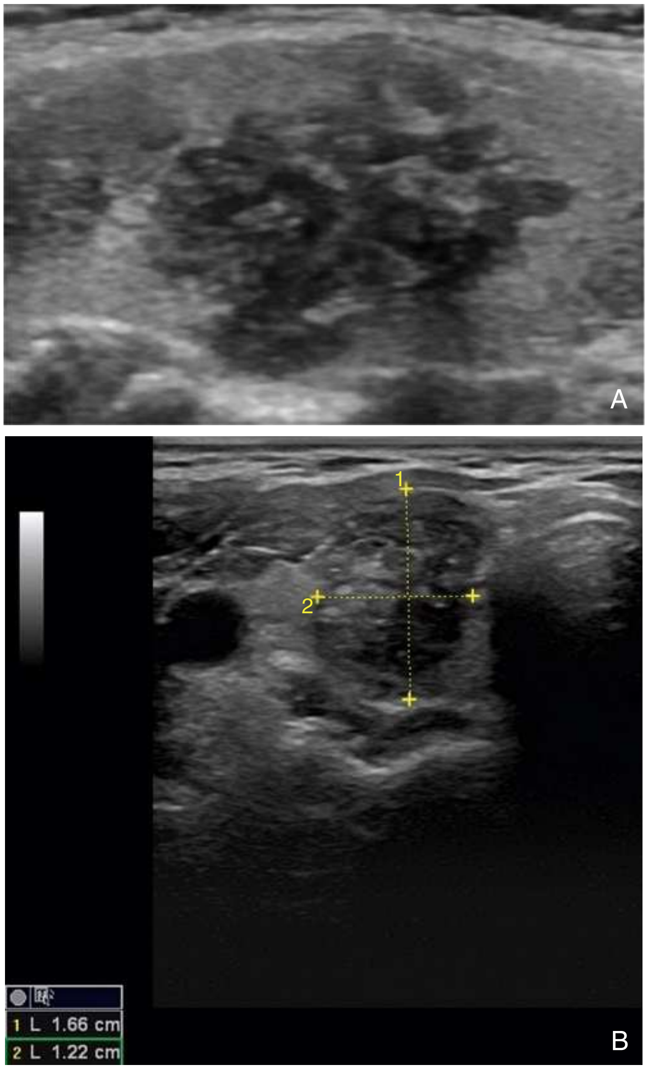


Fig. 69.9 **A,** Diagnostic image of a thyroid ultrasound demonstrating an American Thyroid Association high-suspicion nodule. This image shows irregular or infiltrative margins in a markedly hypoechoic nodule. **B,** Diagnostic image of a thyroid ultrasound showing a hypoechoic solid nodule with taller-than-wide shape.

nodules are recommended for surveillance imaging at 12 to 24 months after FNA.[3]

ATA intermediate-suspicion nodules are hypoechoic and solid with a regular margin, but lacking suspicious US features (Fig. 69.8). While this pattern has the highest sensitivity for detecting malignancy, it lacks specificity; approximately 10% to 20% of these nodules are malignant.[3] The threshold for aspiration is 1 cm. Nodules with a benign FNA are recommended for repeat US at 12 to 24 months.[3]

The ATA high-suspicion category includes solid, hypoechoic nodules (or cystic nodules with a hypoechoic solid component) and one or more suspicious US features, including microcalcifications, taller-than-wide shape, evidence of extrathyroidal extension (ETE), irregular margins, or a disrupted rim calcification with extrusive hypoechoic soft tissue component (Figs. 69.9A and B and 69.e7A–C). This is the category with the highest specificity, and identification of one of these types of nodules has a risk of malignancy of 70% to 90%. Nodules with this pattern measuring 1 cm or more are recommended for consideration of FNA.[3] For those highly suspicious nodules measuring less than 1 cm, lacking evidence of ETE, suspicious locoregional nodes, or concern for distant metastatic disease, FNA or observation may be considered based on patient age and clinical scenario. If a highly suspicious nodule is aspirated and found to have benign cytology, consideration should be given to repeat the FNA within 12 months, as the risk of a false-negative result is higher in this population.[33] Nodules with a high-suspicion pattern that do not meet the threshold for FNA should be considered for repeat surveillance US in 6 to 12 months.[3]

One important limitation to this pattern-based approach is that not all types of nodules are captured with this system. Approximately 4% to 18% of nodules are nonclassifiable by the ATA system.[34,35] In particular, isoechoic nodules with a suspicious US feature are not assessable with the current system (Figs. 69.10 and 69.e8A–C). The malignancy risk associated with these types of nodules is unclear, with widely disparate rates reported between 9% and 68%,[36-39] though most report the malignancy rate on the lower end of that range.

Korean Thyroid Imaging Reporting and Data System

The K-TIRADS (Table 69.2), initially created in 2011[40] and revised in 2016,[41] is similar to the ATA system in its size thresholds for FNA but differs in that it employs a four-tiered stratification. Additionally, K-TIRADS recommends FNA for some high-risk nodules measuring 5 mm or more (in contrast to the 1-cm threshold for all other systems) for certain high-risk patients.

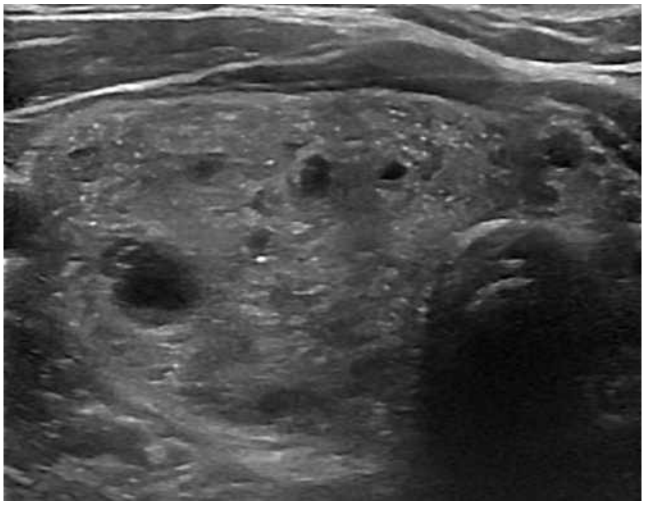


Fig. 69.10 Diagnostic image of a thyroid ultrasound showing an isoechoic mixed composition nodule with microcalcifications.

K-TIRADS emphasizes nodule composition followed by echogenicity for assignment of malignancy risk. The algorithm stratifies solid hypoechoic nodules as the two highest-risk tiers. The presence of sonographically suspicious features (microcalcification, nonparallel orientation, spiculated/lobulated margins) elevates a nodule from K-TIRADS 4 to the highest risk category, K-TIRADS 5. Partially cystic or iso/hyperechoic solid nodules with a suspicious feature are considered intermediate-suspicion or K-TIRADS 4, whereas the absence of a suspicious feature is considered low-suspicion, or K-TIRADS 3. Those nodules that are pure cysts, spongiform, or partially cystic with comet-tail artifact (Fig. 69.e9) are considered benign, or K-TIRADS 2.[41]

The K-TIRADS 3 and 4 nodules that are biopsied and found to have benign cytology are recommended for repeat imaging at 12 to 24 months. K-TIRADS 5 nodules with benign cytology are recommended for repeat FNA at 6 to 12 months of follow-up, in recognition of the high rate of false-negative cytology with a suspicious US appearance and benign cytology.[41]

American College of Radiology Thyroid Imaging Reporting and Data System

A more recent system endorsed by the American College of Radiology utilizes a summation of scores associated with individual features to derive a malignancy risk[29] (Table 69.3). This ACR-TIRADS system specifically examines five nodule features: composition, echogenicity, shape, margin, and echogenic foci. Within each category, the types of findings are assigned a score; higher values are associated with increasing malignancy risk.[29]

The ACR-TIRADS is composed of five levels of thyroid nodules. Those nodules that fall into the TR1 (0 points) and TR2 (2 points) categories are considered benign and not suspicious, respectively, and as such require neither FNA nor follow-up. Those nodules with 3 points are deemed “mildly suspicious,” or TR3; FNA is recommended if the nodule measures 2.5 cm or more. Nodules of this class that measure 1.5 cm or more are advocated for follow-up. Nodules with an aggregate of 4 to 6 points are considered “moderately suspicious,” or TR4, and FNA is advised if greater than or equal to 1.5 cm, with follow-up for those greater than or equal to 1 cm. Nodules that have accrued 7 or more points are regarded as “highly suspicious,” or TR5. FNA is recommended for those measuring greater than or equal to 1 cm, while follow-up is encouraged for those measuring 5 mm or more. Similar to the other two SRSSs, routine FNA of highly suspicious nodules measuring 5 mm or more is not recommended. However, in certain circumstances, FNA may be indicated.[29]

The ACR-TIRADS system also provides a framework for surveillance of nodules, both those meeting criteria for FNA and those that do not. For TR 5 nodules, the committee recommends annual US exam

TABLE 69.2 Korean Thyroid Imaging Reporting and Data System Risk Stratification System

Risk Category	Description	Malignancy Risk	Threshold for FNA
High	Solid, hypoechoic *with* any suspicious feature*	>60%	≥1 cm; (>0.5 cm selective^)
Intermediate	Solid hypoechoic *without* suspicious features OR Partially cystic or iso/hyperechoic *with* any suspicious feature*	15%–50%	≥1 cm
Low	Partially cystic or iso/hyperechoic *without* any suspicious feature*	3%–15%	≥1.5 cm
Benign	Spongiform OR Partially cystic with comet-tail artifact OR Pure cyst	<3% <1%	≥2 cm N/A

*Suspicious features include: microcalcification, nonparallel orientation (taller-than-wide), or spiculated/microlobulated margin.

^FNA of nodules >5 mm may be performed for those patients with risk factors requiring immediate surgery, such as those with evidence of extrathyroidal extension, nodal metastases, or distant metastases. Additionally, those nodules with evidence of invasion into the trachea or recurrent laryngeal nerve may be biopsied.

FNA, Fine-needle aspiration.

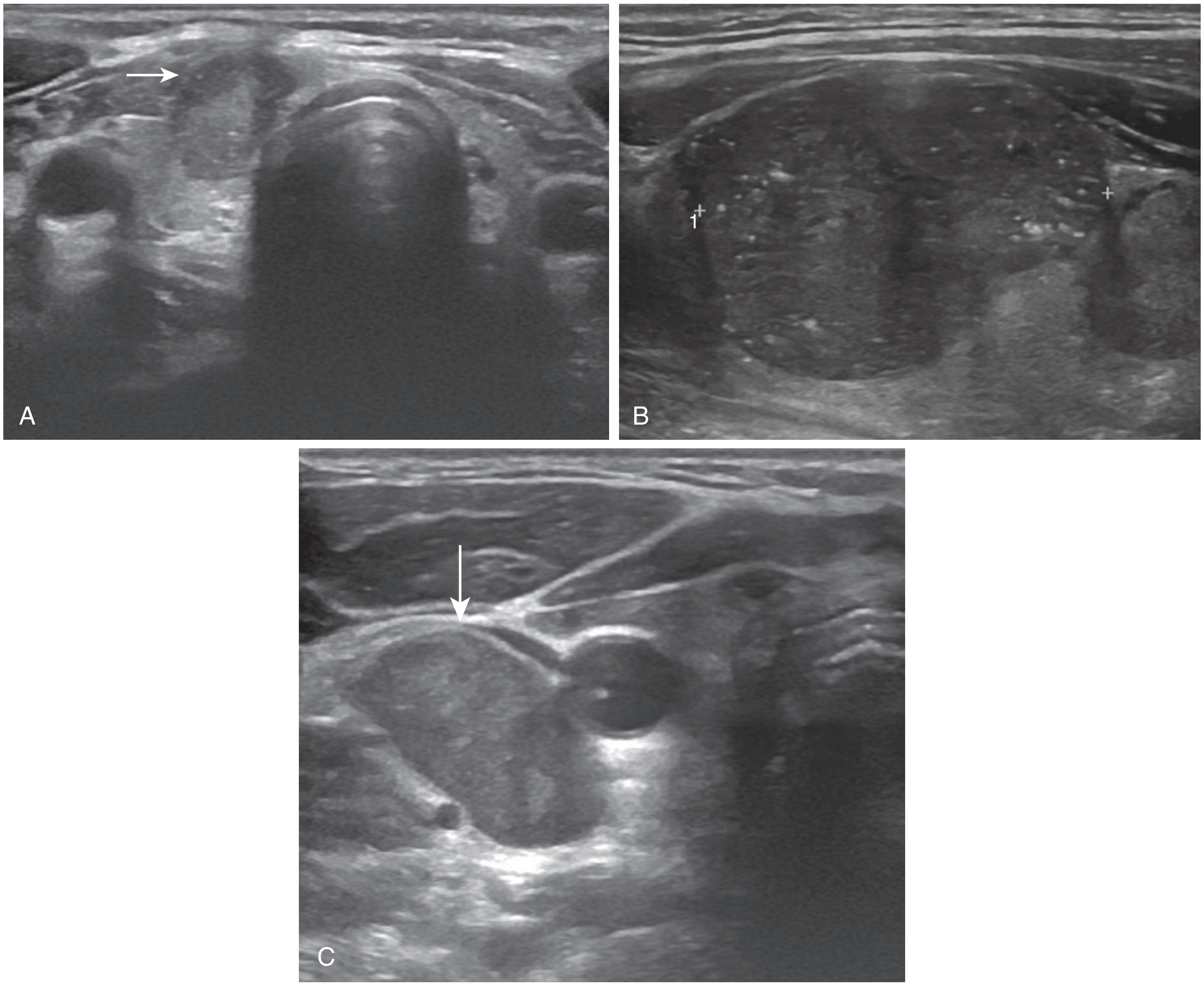


Fig. 69.e7 A, Diagnostic image of a thyroid ultrasound showing a hypoechoic nodule with microcalcifications and possible extrathyroidal extension (white arrow). B, Diagnostic image of a thyroid ultrasound showing a hypoechoic solid nodule with microcalcifications which are also called punctate echogenic foci. C, Diagnostic image of a thyroid ultrasound demonstrating a malignant lateral lymph node demarcated with the white arrow.

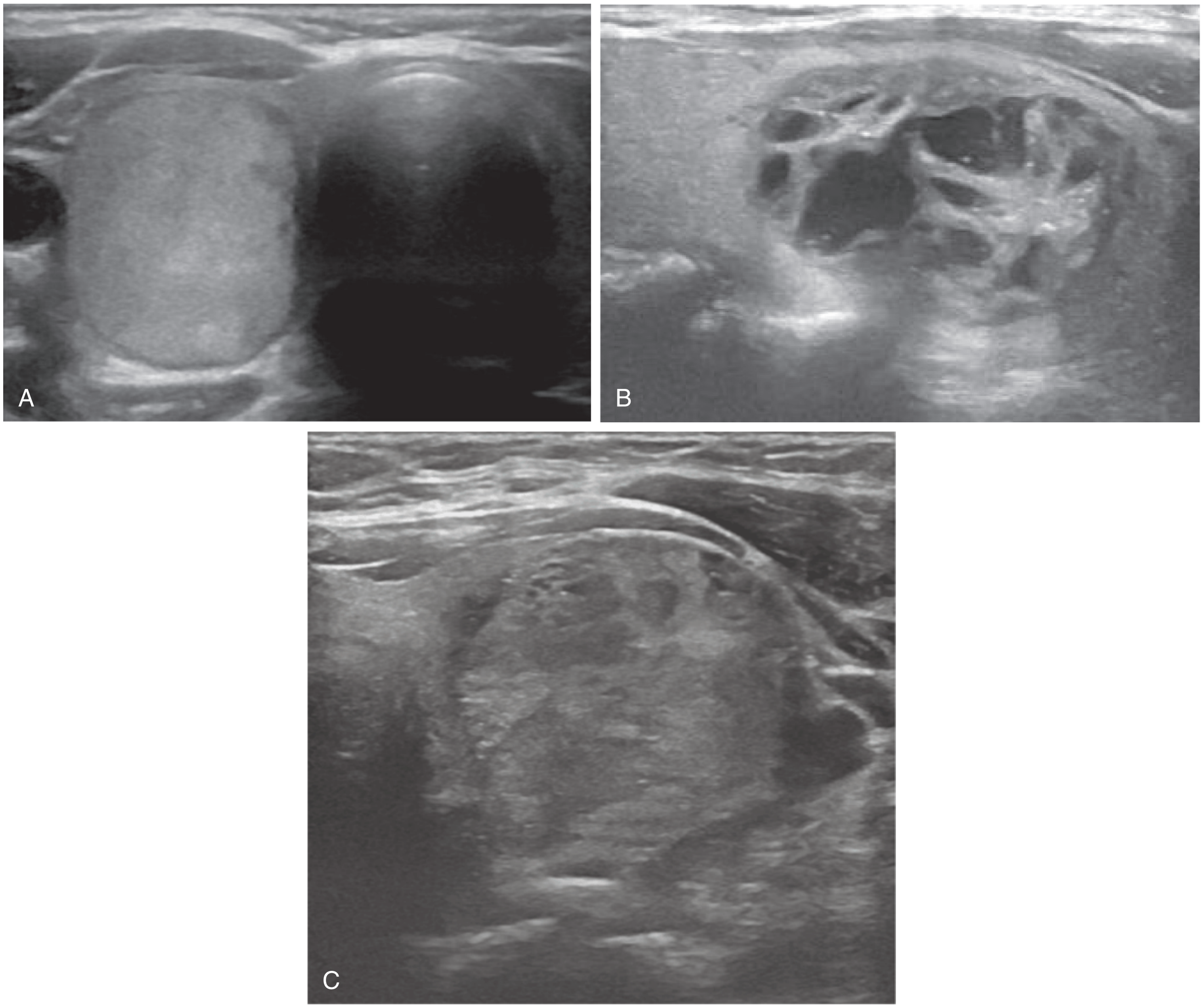


Fig. 69.e8 A, Diagnostic image of a thyroid ultrasound showing an isoechoic nodule with taller-than wide shape. B, Diagnostic image of a thyroid ultrasound showing an isoechoic mixed composition nodgule with microcalcifications. C, Diagnostic image of a thyroid ultrasound showing a solid nodule with heterogeneous echogenicity.

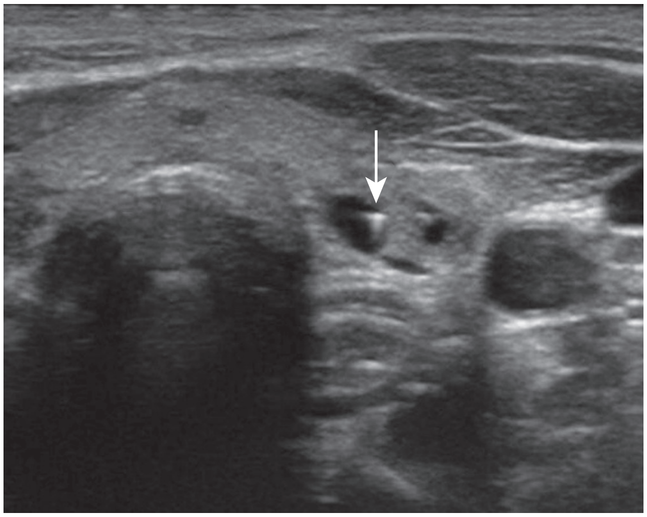


Fig. 69.e9 Diagnostic image of thyroid ultrasound showing a comet-tail sign (white arrow) or punctate echogenic focus with a vee-shaped posterior tail. These findings are reassuring in a purely cystic nodule.

TABLE 69.3 American College of Radiology Thyroid Imaging Reporting and Data System

Nodule features and associated points for each characteristic.

Composition	*Echogenicity*	*Shape*	*Margin*	*Echogenic Foci*
Cystic 0	Anechoic 0	Wider than tall 0	Smooth 0	None 0
Spongiform 0	Hyper-/isoechoic 1	Taller than wide 3	Ill-defined 0	Comet tails 0
Mixed 1	Hypoechoic 2		Irregular 2	Macrocalcifications 1
Solid 2	Markedly hypoechoic 3		ETE 3	Peripheral/rim 2
				Punctate 3

ACR-TIRADS scores and FNA thresholds

Score	*0*	*2*	*3*	*4–6*	*≥7*
TIRADS Class	TR1	TR2	TR3	TR4	TR5
Description	Benign	Not suspicious	Mildly suspicious	Moderately suspicious	Highly suspicious
FNA Threshold	Not recommended	Not recommended	≥2.5 cm	≥1.5 cm	≥1 cm

FNA, Fine-needle aspiration; *ACR,* American College of Radiology; *TIRADS,* Thyroid Imaging Reporting and Data System.

every year for up to 5 years. Those nodules categorized as TR4 should be scanned at years 1, 2, 3, and 5, whereas TR3 lesions may be imaged at 1, 3, and 5 years of follow-up. This committee proposes that imaging can stop at 5 years if there is no change in size, as a lack of growth over that length of time is highly consistent with a benign-behaving neoplasm. Furthermore, they recommend that if a nodule's TIRADS level increases during follow-up, repeat imaging in 12 months is warranted, regardless of the initial grade.[29]

The major limitation of ACR-TIRADS is the need to memorize a number of features and their associated scores. The attendant interobserver variability associated with each feature further constrains the reliability of this system for accurately characterizing malignancy risk. Another criticism of ACR-TIRADS is that using it can become time-consuming, particularly in the setting of having to categorize multiple nodules; assessing nodules with a pattern-based system is significantly quicker.

Diagnostic Performance of Sonographic Risk Stratification Systems

While each of these systems have their individual limitations and strengths, their performance for distinguishing benign from malignant nodules in prospective and retrospective analyses is very similar. Generally, the ATA and K-TIRADS have a high sensitivity for detecting malignancy but do so at the cost of a lower specificity. Consequently, these two systems are associated with a higher "unnecessary biopsy rate" or recommend more biopsies of what are ultimately benign nodules.[42] In contrast, the ACR-TIRADS typically has a higher specificity and lower sensitivity than the ATA and K-TIRADS; the unnecessary biopsy rate for ACR-TIRADS is accordingly lower.[42]

An important caveat to all SRSSs is that they are designed to identify papillary thyroid carcinoma. The sonographically suspicious features of taller-than-wide shape, microcalcifications, and infiltrative margins are more specific for papillary carcinomas.[43] In contrast, follicular-patterned tumors generally appear as solid, isoechoic, or mildly hypoechoic nodules with a regular oval shape and lacking microcalcifications.[43,44] Consequently, the papillary thyroid carcinomas tend to be represented by the "high-risk" lesions on all SRSSs, whereas the follicular tumors are represented in the "low-risk" and "intermediate-risk" categories. It is also noteworthy that nearly all studies examining the performance of various SRSSs include very limited numbers of follicular-patterned tumors in their cohort of nodules. While the decreased prevalence of follicular carcinoma partly contributes to this lack of representation, many of the studies also exclude nodules with indeterminate cytology, which disproportionately eliminates follicular tumors.[45] As a result, it is unclear how well these SRSSs perform with identification of malignancy in follicular-patterned tumors.[45]

KEY POINTS

- Though each of the three main sonographic risk stratification systems have individual strengths and weaknesses, they perform similarly to categorize the malignancy risk of thyroid nodules.

Specialized Sonographic Techniques

Contrast-Enhanced Ultrasonography. Contrast-enhanced ultrasonography (CEUS) is a technique that is helpful to uncover tissue microvascularization, a critical component of neoplastic growth. Microbubbles, composed of a shell with a gas core, are employed as contrast agents and act as echo enhancers to improve visualization of tissue microperfusion. These 1- to 10-µm microbubbles are capable of passing through capillary vessels, allowing for improved characterization of tumor vascularity.

These microbubbles resonate in response to a specific transmitted acoustic wave frequency, which is dictated by the radius of the microbubble. Those agents used in thyroid applications respond to the US frequencies used in evaluation of thyroid tissues, 7 to 15 MHz.[46] The contrast agent is then detected by the US device because it has acoustical characteristics that are different from the surrounding soft tissues. The difference in acoustic impedance compared with the surrounding tissues results in a mismatch with reflection of the US wave back to the probe, which enhances the imaging characteristics and leads to improved detection of subtle features.[46]

The contrast agent is typically powdered and requires reconstitution in normal saline. It is administered intravenously in the antecubital fossa for each CEUS examination. The microbubble perfusions can be seen within a few seconds in the targeted nodule and typically can be observed for 2 minutes or less. After injection of the contrast agent, the US machine is switched to contrast mode and directed to the region of interest, with the focal zone placed at the posterior edge of the nodule (Fig. e69.10). Because of the short interval that the microbubbles remain within the thyroid nodule, it is important to record cine-loops after the administration of contrast for later, more careful inspection of the nodule.

Being rich in microvessels, the normal thyroid parenchyma demonstrates a rapid, uniform enhancement after injection of the contrast. Nodules, on the other hand, exhibit a different vascular structure and consequently stand out from the surrounding normal tissue. The

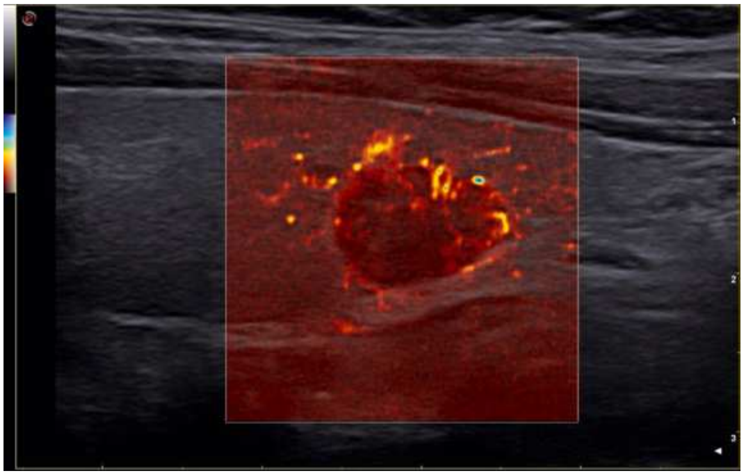


Fig. 69.e10 Contrasted ultrasonography

resultant images of the nodules are then qualitatively analyzed by their vascularization features.

Quantitative analysis also may be performed upon completion of the image acquisition. Using the quantification software, the operator manually distinguishes a region of interest and automatically calculates various parameters including: peak intensity, time to peak, mean transit time, and total volume of blood in the region of interest.[47] The role that these features play in distinguishing benign and malignant neoplasms is still under investigation, however.

Elastography. On palpation examination, the firmness of a thyroid nodule is assessed; increasing hardness of the tissue is associated with greater concern for the possibility of malignancy. US elastography (USE) is a bedside tool that can be used to objectively measure this tissue stiffness. This technology may aid in the assessment of malignancy risk of thyroid nodules, with stiffer nodules having a higher likelihood of harboring a cancer. While utilizing USE requires specific software and a special probe, it allows for assessment of nonpalpable nodules. The measure of tissue elasticity is compared with that of adjacent musculature for a quantitative assessment. There are two main forms of applied force: external compression of the probe applied by the user or shear wave generated by the US machine.

The external compression form of elastography (also known as strain USE [SUSE]) is manually performed by pressing the probe into the skin of the patient, generating tissue strain perpendicular to the skin surface. The nodule and surrounding parenchyma are highlighted with a box on the screen. US images are taken before and after the tissue compression. Software then measures the degree of tissue displacement with the pressure; the degree of stiffness is inversely proportional to the amount of tissue deformation. The data are then processed to create a corresponding color scale that is superimposed on the US image. The colors range from blue (which has the lowest degree of elasticity and the highest risk of malignancy) to green (intermediate elasticity) to red (corresponding to the highest degree of elasticity and lowest risk of malignancy) (Fig. 69.e11). SUSE is older technology that is prone to bias; its utility is limited by operator experience and tissue depth.

The newer technology, shear wave elastography (SWE), creates particle motion tangential to the skin surface. Utilizing an acoustic radiation force pulse generated from the probe, shear waves that are perpendicular to the US beam cause transient displacement of the tissues. The distribution of shear-wave velocities is directly related to the tissue's elastic properties, as shear waves propagate faster through stiffer tissue. The US probe detects these changes and translates the data into an image that conveys the degree of tissue elasticity. The benefit of this technology is that it is independent of user experience.

There are important limitations of elastography that likely have contributed to its failure to achieve widespread adoption into clinical practice. USE cannot be used in nodules with calcifications, as their hardness will artificially reduce the tissue strain. Similarly, the presence of a cystic component in a nodule precludes the use of this technology, as it will falsely elevate the compressibility of the nodule tissue. Many studies also exclude those patients with Hashimoto's thyroiditis due to the potential for the lymphocytic infiltrate and tissue fibrosis to influence the tissue compressibility.

It is unclear how location of the nodules impacts the performance of SUSE, particularly those that are posteriorly located or in the isthmus. The nodules in the isthmus or near the trachea may not be assessed with this technology due to the interference produced by the tracheal cartilage. One recent, small, retrospective analysis[48] examined the question of nodule depth on performance of SUSE. The authors found that, while the depth of the nodule did not impact the ability to measure tissue stiffness, proximity to the carotid did influence the accuracy. The elasticity contrast index was underestimated for those nodules located within 15 mm of the carotid.[48] Suspicious nodules are associated with a higher elasticity contrast index, and underestimation of this value may result in a false reduction of the malignancy risk. This "geographic" limitation is of particular significance in women, as their thyroid lobes tend to be smaller, measuring between 15 and 20 mm in the transverse direction. A smaller lobe positions the majority of nodules within those regions where the tissue strain may be misregistered.

Initial analyses of the performance of elastography revealed that the technology had remarkably high sensitivity and specificity for distinguishing benign from malignant nodules.[49,50] While useful for a proof of concept, subsequent studies inclusive of a more diverse nodule population were unable to replicate such a highly discerning performance.[51]

There are noteworthy methodologic flaws in many of the studies examining the utility of USE. Due to the retrospective nature and small sample size of most of the studies, there is the significant potential for selection bias. Many of the analyses examined patients with a solitary nodule, further limiting the diagnostic relevance of this technology in the general population, particularly those with multinodular goiter.[51] Additionally, many of the early studies excluded nodules with indeterminate cytology from the analysis of the performance of USE. However, the ability to discern the malignant potential of nodules with indeterminate cytology would be an ideal utility for this technology. Nodules with a suspicious appearance on traditional gray-scale sonography do not require further risk stratification, as the specificity of these findings is already satisfactorily elevated.[3] Instead, it is nodules with an intermediate- or low-risk appearance on gray-scale US that would benefit from the augmented diagnostic accuracy of identifying the few malignancies within this class of nodules. Those nodules that are malignant in these low- and intermediate-risk sonographic categories tend to represent follicular-patterned tumors such as follicular carcinoma, follicular variant of papillary thyroid carcinoma, or NIFTP. Consequently, exclusion of nodules with indeterminate cytology not only limits the utility of this testing, but also provides an incomplete picture of its performance. Unfortunately, those few analyses that include indeterminate cytology provide conflicting results regarding how well this technology performs in this population,[52-55] likely owing to small sample sizes and potential selection bias. Nevertheless, a recent metaanalysis that included 1734 nodules did show a modest diagnostic performance with USE in indeterminate nodules; the pooled sensitivity and specificity were 76.6% and 86.7%, respectively.[56] An important limitation of this particular study is that SUSE and SWE were combined into one analysis. Thus, it remains unclear how these individual technologies perform for the malignancy risk prediction with indeterminate nodules. Nonetheless, in spite of the abovementioned concerns, the potential of this technology to play a role in the assessment of nodules in conjunction with the gray-scale sonographic risk stratification to help stratify malignancy risk remains intriguing.[57-60]

Computer-Aided Diagnosis Systems. One of the major concerns regarding traditional gray-scale US is that its diagnostic performance relies heavily on the expertise of the interpreting sonographer.[61] The sonographic features used to describe nodules are fraught with interobserver variability in reporting.[62,63] In areas where proficiency in US interpretation is lacking, there is a need for technology that will deliver diagnostic accuracy using sonographic analysis in a timely fashion. Artificial intelligence is being increasingly investigated as a possible adjunctive tool to assist in the decision-making or to provide a second opinion for the diagnosis of malignant nodules with operators who have less experience using US.[64]

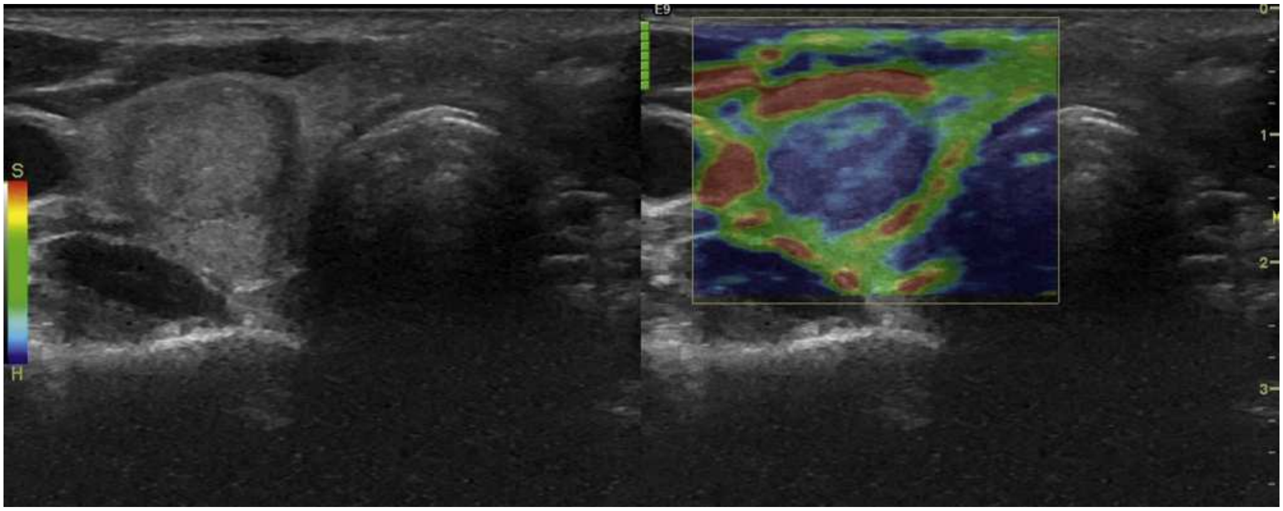


Fig. 69.e11 B-mode US and elastography image of a high-risk thyroid nodule. The blue regions demonstrate areas of decreased compressibility.

There are two main types of machine learning methods: the classic machine learning (CML) method and the deep learning (DL) method. The CML model utilizes features defined by human experts; it merely simulates the diagnostic strategy of radiologists. Those sonographic images of nodules with known histologic diagnoses of malignancy or benignity are fed into the system by the radiologist to "train" the computer with these associations. Consequently, the CML model is limited in its ability to transcend the imperfections of its human teachers. The DL model overcomes this important obstacle by decreasing subjectivity during the training process: it allows computational models, composed of multiple processing layers, to learn representations of data with compound levels of abstraction.[65] Specifically, DL uses numerous levels of training with representative images. Each layer of learning adds a level of complexity to aid in the discriminatory power of the machine. The key aspect of this learning method is that it is independent of human input; the learning is solely derived from the data or raw images entered into the system.[65]

A recent metaanalysis examined the diagnostic performance of the two computer-aided diagnosis (CAD) systems and compared their performance with that of experienced radiologists. In an analysis of six separate studies, the pooled sensitivity and specificity were 86% and 85%, respectively, for CML-based CAD.[66] The DL-based CAD systems performed similarly, with a pooled sensitivity and specificity of 89% and 84%, respectively.[66] The pooled sensitivity, specificity, area under the curve, and diagnostic odds ratio were comparable between the two CAD systems and the radiologists' interpretations.[66]

Given the rising incidence of thyroid cancer, much of which is due to increased detection of subclinical disease,[67] a tool that can rapidly and accurately sort through nodules with limited intervention would be of great value. Machine-based learning offers an attractive option for the rapid analysis of incidentally-detected nodules to help radiologists identify high-risk lesions and potentially prevent unnecessary FNA of low-risk nodules.

CAD technology has only recently become available outside of a research setting; there is one company that has created software that can be integrated into a US platform.[68] However, as yet, these tools are not ready for clinical practice. The existing studies are designed to prove technical feasibility, but real-world clinical performance has not yet been demonstrated.[68] Multicenter, prospective trials and observational outcome studies are needed to demonstrate efficacy.[68,69] Until such studies are performed, widespread adoption of this technology into clinical practice seems less likely.

SCINTIGRAPHY

Radionuclide scintigraphy was once the mainstay of evaluation of thyroid disorders. With the advent of bedside US to stratify malignancy risk in thyroid nodules, routine scintigraphic analysis of the thyroid has declined. However, this imaging technique remains the only methodology to evaluate the functional status of a thyroid nodule, and it remains a useful tool for the distinction of Graves disease from subacute and silent thyroiditis. Furthermore, it is an important component of the evaluation of select cases of congenital hypothyroidism. The use of nuclear scintigraphy for the evaluation and treatment of thyroid cancer is discussed in Chapter 78.

Radioiodine tests are composed of two distinct parts: the planar imaging and the measurement of the percent uptake of iodine. One or both components of radionuclide testing may be performed for determination of the underlying thyroid disorder. The radiotracer is given to the patient either intravenously or orally. The timing of the scan is dependent on the route of administration and the type of tracer utilized. The patient is placed on an examination table, and a gamma camera takes images of the thyroid region from three different angles: anterior, posterior, and oblique.

TABLE 69.4 Drugs Impacting Uptake of Tracer in the Thyroid

Impact	Agent
Decreased Uptake	Amiodarone
	Methimazole
	Perchlorate
	Propylthiouracil
	High dose salicylates
	Sulfonamides
	Thiocyanate
	Triiodothyronine (T3)
	Thyroxine (T4)
	Iodine-containing ointments
	Kelp
	Cough medicine and vitamins containing iodine
	Iodinated contrast agents
Increased Uptake	Lithium

The patient should be carefully prepared for the test, including a thorough review of medications and recent imaging studies. In patients who are already being treated for hyperthyroidism it is important to hold thionamide therapy for at least 5 days prior to the scan to ensure accurate results. A number of other drugs can falsely alter the uptake of tracer within the thyroid (Table 69.4). In the setting of iodine deficiency there can be falsely increased uptake of radiotracer. The administration of iodinated contrast with computed tomography (CT) can falsely lower the uptake of radioiodine tracer. Chronic renal failure can impair iodine clearance and lower radioiodine uptake. Also, increased ingestion of goitrogenic foods such as cabbage and turnips may artificially lower uptake of the tracer.

Radionuclide imaging is absolutely contraindicated in pregnancy; radiation exposure from the scan is harmful to the developing fetus. A negative pregnancy test should be obtained prior to performing this testing for all females with reproductive potential. In addition, because radioactive iodine is secreted in breast milk, lactating women should not undergo nuclear scintigraphy. If it is critical to perform scintigraphy, lactating women should be instructed to pump and discard the breastmilk for 5 days after the ^{123}I administration.

Thyroid Nodules

Thyroid scintigraphy is most commonly performed for evaluation of the functional status of a nodule. Nodules are described based on the pattern of uptake of the radiotracer compared with the surrounding thyroid parenchyma. Nodules that are hyperfunctioning are classified as "hot," while hypofunctioning lesions are termed "cold" (Fig. 69.11). Those nodules that demonstrate similar activity to the surrounding thyroid parenchyma are defined as "warm" or indeterminate. Hot nodules are almost exclusively benign; they do not routinely require FNA interrogation.[3] In contrast, cold or warm nodules may harbor a malignancy, and consequently FNA may be warranted based on their sonographic appearance.[3]

The two main radionuclides employed for the evaluation of thyroid nodules are ^{123}I (radioiodine) and ^{99m}TcO4^{-} (pertechnetate). The ^{123}I is both concentrated and organified within the thyroid, whereas ^{99m}TcO4^{-} is only concentrated in the gland, without organification. This

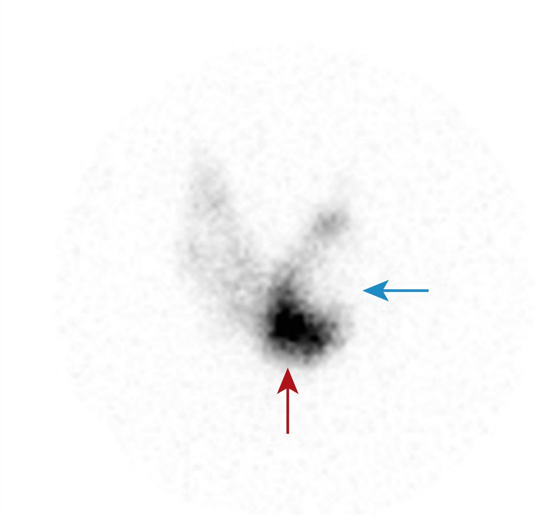


Fig. 69.11 Diagnostic image of ^{123}I scintigraphy demonstrating a hot nodule *(red arrow)* with an adjacent cold nodule *(blue arrow)*.

distinction is important for the evaluation of nodules. Approximately 5% of thyroid cancers concentrate pertechnetate but not radioiodine.[70] These nodules may falsely appear as functional on pertechnetate scans, while they actually are nonfunctioning on radioiodine scans. Consequently, the ATA guidelines preferentially recommend ^{123}I over $^{99m}TcO4^{-}$ for scintigraphic scanning.

There are benefits and disadvantages to the use of each of the tracers. The use of technetium exposes the patient to less radiation, but because it is not organified, it does not allow for the measurement of uptake. Because technetium is not retained within thyroid tissue, this type of scintigraphy must be performed within 20 to 30 minutes after injection of the tracer, providing rapid, convenient results for the patient. An additional benefit of technetium is that it is less expensive and more readily available than iodine isotopes. Radioiodine, on the other hand, allows for better image quality and resolution, particularly when thyroid uptake is low. Radioiodine is metabolized identically to naturally occurring iodine, and it is incorporated into thyroglobulin for storage in thyroid follicles. The emission of gamma rays from the ^{123}I allows for the imaging of the thyroid several hours after ingestion or infusion of this tracer. Measurements of uptake are performed 1 hour after intravenous injection or at 4 and 24 hours after oral administration. The timing of varying degrees of intrathyroidal uptake is then utilized to further differentiate the underlying pathophysiologic process. The radioiodine scans are more inconvenient to the patients; the delayed images over 24 hours require patients to return to the imaging center for multiple scans. Additionally, ^{123}I is a more expensive tracer, and at times it can be difficult to obtain the tracer due to interruptions in the supply chain.

A thyroid radionuclide scan is not essential for the investigation of all thyroid nodules. The initial step in the evaluation of sonographically-confirmed thyroid nodules involves measurement of a serum thyroid-stimulating hormone (TSH). If the TSH is low, a scintigraphic scan is beneficial to ascertain the functional status of the nodule. Because the malignancy risk is generally accepted to be extremely low in hot nodules, further analysis with FNA is largely unnecessary. On the other hand, nodules that are deemed warm or cold may require further investigation with FNA to rule out malignancy. Conversely, for patients with a normal or elevated TSH level, the likelihood of identifying a functional nodule on scintiscanning is very low. As such, a radionuclide study is generally not recommended in this setting.[3] Nevertheless, in iodine-deficient regions, TSH may remain unsuppressed even in the presence of nodule autonomy due the low production rates of thyroid hormone in the iodine-depleted gland. Indeed, a metaanalysis of eight studies evaluating scintigraphy revealed that up to 50% of autonomous nodules had a normal TSH.[71] Thus, it may be beneficial to perform scintigraphy in the setting of a nodule or multinodular goiter with a low-normal TSH in an area of iodine deficiency, as the risk of autonomy may be high in this population.[4]

It is important for clinicians to be familiar with the limitations of nuclear scintigraphy. The resolution of radionuclide scanning is very poor in comparison to ultrasonography. Cold nodules measuring less than 1 cm are very likely to be missed on scintigraphy. Furthermore, posteriorly located nodules may be difficult to characterize. Overlying normal thyroid tissue could distort the images and impart a warm appearance to a posterior, nonfunctioning nodule. Additionally, in cases of a multinodular goiter with nodules of varying functionality, it can be difficult to distinguish which lesion has corresponding activity, particularly in the event of overlapping nodules. In these settings, it is important to carefully perform scintigraphic imaging at multiple angles and closely correlate the findings with the US exam. Finally, concentration of the tracer by the esophagus can mimic ectopic thyroid tissue. If this situation is suspected, it is important to acquire additional images after swallowing water to determine if the tracer has cleared.

Diffuse Thyroid Disorders

Uptake scintigraphy is helpful for distinguishing the causes of thyrotoxicosis from diffuse thyroid disorders. Abnormal uptake values are described as either high or low. Those conditions with high or normal uptake are indicative of synthesis of thyroid hormone. In contrast, those conditions with low or absent uptake may be seen due to either destruction/inflammation of the thyroid gland with release of preformed hormone into the circulation or due to an external source of thyroid hormone (exogenous thyroid hormone ingestion or struma ovarii). Low uptake values may also be seen when there is increased exposure to iodine, such as with contrast dye administration or amiodarone use. If the uptake values are lower than would be expected by the clinical history, a 24-hour urine iodine may be obtained to demonstrate whether the patient has had exposure to excess iodine.[18]

The uptake of radiotracer is diffusely increased in the setting of hyperthyroidism due to Graves disease and toxic multinodular goiter. These two conditions also may be distinguished by antibody testing and US examination demonstrating the presence or absence of nodularity. Patients with clinically evident end-organ manifestations of Graves disease such as orbitopathy do not require scintigraphy to confirm the diagnosis. However, in mild cases of Graves disease lacking the characteristic clinical findings, the radionuclide study is a very useful test to rule out silent thyroiditis. In the thyrotoxic phase of silent or subacute thyroiditis, the radioiodine uptake is very low or undetectable. An experienced sonographer can distinguish these two entities, however, utilizing color Doppler interrogation (see earlier section). Furthermore, the clinical features of silent thyroiditis, including a painless thyrotoxicosis occurring within a year of childbirth in the setting of a positive family history for autoimmune thyroid disease, are sufficiently diagnostic in many cases. Similarly, nuclear scintigraphy is typically not needed to identify subacute thyroiditis, which is marked by an acutely painful gland with an elevated erythrocyte sedimentation rate.

Although most patients with Graves disease do not require scintigraphic studies, in the setting of a concomitant nodule a scintigraphic scan is beneficial to assess whether the nodule is nonfunctional and in need of possible FNA.[18] It should be noted, though, that a small, nonfunctioning nodule may be falsely reported as either warm or hot due to overlying, hyperfunctioning Graves parenchyma.

Congenital Hypothyroidism

The most common causes of congenital primary hypothyroidism are due to either a defect in thyroid gland development (dysgenesis) or a defect in thyroid hormone synthesis (dyshormonogenesis). While biochemical analysis is the mainstay of diagnostic confirmation, imaging may be necessary to identify the underlying cause or may be helpful to assess whether to treat versus observe.

Due to its ease of use, universal availability, and lack of radiation exposure, US is generally the first-line imaging technique in the evaluation of congenital hypothyroidism. In the case of thyroid dysgenesis, US most frequently identifies the location of the ectopic thyroid tissue. However, the small size of fetal thyroids means that they can be challenging to identify for an inexperienced sonographer, which may necessitate a nuclear study. Scintiscanning may also be crucial in cases of athyreosis or a lingual thyroid. For an accurate diagnosis of dyshormonogenesis, on the other hand, the combination of both US and scintigraphy may be needed. Either ^{123}I or $^{99m}TcO4^{-}$ may be used for fetal thyroid imaging, but ^{131}I should be avoided, because the radiation dose is significantly higher. Generally, ^{123}I is the preferred tracer, as it provides superior contrast resolution and allows for the concomitant administration of the perchlorate discharge test (see later section).[72]

Scintigraphic scanning for the evaluation of congenital hypothyroidism should be performed either before or shortly after the initiation of levothyroxine, because the uptake of radioiodine is dependent on TSH secretion. Normalization of serum TSH can result in misrepresentation of the underlying pathophysiology, with little or no uptake of radioactive iodine. If it cannot be performed within this narrow time window, the testing should be delayed until after the age of 3 years, when the interruption of hormone therapy no longer poses a threat to neurocognitive development.[73] To circumvent this limitation, some advocate for avoiding thyroid hormone withdrawal entirely by utilizing recombinant human TSH.[74]

Perchlorate Discharge Test

A perchlorate discharge test is used in conjunction with an ^{123}I scan to diagnose organification defects. In the setting of normal thyroid physiology, iodide is captured by thyrocytes and rapidly fixed, or organified, to thyroglobulin tyrosine residues, resulting in no free iodide within the cell. When there is a defect in organification, free iodide accumulates within the thyrocytes. Perchlorate is similarly taken into the thyrocytes but is not organified. In patients with an organification defect, administered perchlorate competes with the free iodide that is located inside of the thyrocytes and pushes it out of the cells,[72] whereas, in patients with normal organification, the perchlorate does not impact the iodide dynamics, because intrathyroidal iodine is fixed to the thyroglobulin.

A perchlorate discharge test involves administration of radioactive iodine followed by a baseline measurement of the amount of iodine captured by the thyrocytes. Oral perchlorate is then administered, and the ^{123}I activity within the thyroid is again measured after 1 hour. The ratio of the change in activity before and after perchlorate is compared with the initial activity. In an infant with normal organification, perchlorate does not change the amount of activity within the thyroid; this is termed a "negative" perchlorate discharge test. On the other hand, in children with organification defects, there is a rapid reduction in ^{123}I activity within the thyroid after the administration of perchlorate. A change in activity of greater than 10% is considered to be significant. A "positive" perchlorate discharge test therefore serves as confirmation of an organification defect.[75]

COMPUTED TOMOGRAPHY AND MAGNETIC RESONANCE IMAGING

Although CT and magnetic resonance imaging (MRI) are insensitive means of characterizing thyroid nodules, these imaging modalities do have an important role in the evaluation of patients with obstructive and substernal goiters. Cross-sectional imaging allows for determination of the presence and degree of intrathoracic extension, which is difficult, if not impossible, to assess with US. Furthermore, these modalities also aid in the assessment of the tracheal narrowing. This visualization of the effect of the goiter upon surrounding structures informs the evaluation of whether the patient's compressive symptoms are due to the thyroid or some other underlying pathophysiologic condition. Additionally, cross-sectional imaging aids in the decision to pursue surgical intervention versus observation.

Large goiters may be asymptomatic or cause varying degrees of compressive symptoms. Exertional dyspnea is present in 30% to 60%[76,77] of those with obstructive retrosternal goiters and typically occurs when the tracheal diameter is less than 8 mm.[78] Other potential compressive symptoms include cough, hoarseness, stridor, choking sensation, Horner syndrome, and dysphagia. Superior vena cava syndrome also may be seen due to compression of venous structures.

US should be performed in all patients with a goiter to assess the cervical component for findings concerning for malignancy. Once it is established that the inferior extent of the gland is not visible with US, CT or MRI is indicated. Depending on the length of the thyroid, imaging of both the neck and chest may be necessary to capture the entirety of the goiter. The decision to utilize CT versus MRI is based on provider preference. However, if CT is utilized, a noncontrasted study should be performed. If the goiter is autonomous or the patient is overtly hyperthyroid, the administration of iodinated contrast may precipitate or exacerbate preexisting thyrotoxicosis.

Cross-sectional imaging allows for the assessment of the degree of tracheal deviation and quantification of tracheal narrowing (Fig. 69.12). Pulmonary function testing with flow-volume loop study also is helpful to assess the degree of respiratory compromise, even if the patient is asymptomatic. It is also helpful for symptomatic patients to determine if other causes of dyspnea are responsible.

Cross-sectional imaging is also beneficial for preoperative evaluation of thyroid cancer for those with an invasive primary tumor or clinically apparent multiple or bulky lymph nodes.[5] In this scenario, contrasted (rather than noncontrasted) CT is recommended to improve the sensitivity of identification of nodal involvement.[5] Although there is the theoretical concern that iodinated contrast will delay the use of radioactive iodine imaging and therapy, the majority of patients can achieve sufficiently low iodine concentrations to proceed with such therapy within

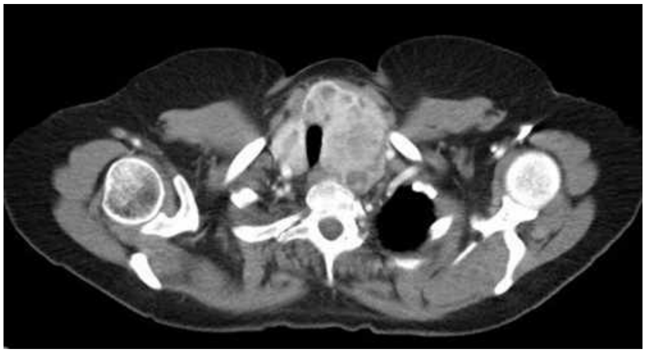


Fig. 69.12 Diagnostic image of computed tomography demonstrating tracheal deviation and narrowing due to a compressive goiter.

a month of receiving the contrast load.[79] Such a brief delay is unlikely to confer a survival disadvantage for the vast majority of thyroid cancer patients receiving radioactive iodine therapy. Alternatively, gadolinium-enhanced MRI of the neck may be used to avoid iodinated contrast entirely. The use of MRI, however, is limited by the potential for motion artifact due to swallowing and breathing.[5]

KEY POINTS

- Computed tomography or magnetic resonance imaging are important components of the evaluation of a substernal or compressive goiter to examine the degree of retrosternal extension and/or the presence of tracheal narrowing. Cross-sectional imaging is also an important component of the preoperative planning in thyroid cancer patients with bulky nodal disease.

INCIDENTALOMAS

Thyroid nodules identified on imaging studies not specifically intended for the assessment of thyroid pathology are termed thyroid incidentalomas, or incidental thyroid nodules (ITNs). It is important to recognize that the majority of these lesions represent a benign neoplasm. When these lesions do harbor a malignancy, papillary thyroid carcinoma constitutes the bulk of the cancers identified, which is generally associated with an excellent prognosis. This potentially favorable outcome is particularly important for those patients with another primary malignancy elsewhere that may have instigated the imaging study. Within this framework, diagnostic evaluation of ITNs is not always actively pursued, depending on the clinical scenario in which they are identified.

Computed Tomography and Magnetic Resonance Imaging

ITNs are one of the most common incidental findings on imaging of the neck; the prevalence is dependent on the type of imaging modality employed. Patient demographics also may have an impact on the incidence of incidentally noted thyroid nodules; women are twice as likely to have a thyroid nodule on exam compared with men,[80] and nodules are more commonly seen with increasing age.[81] Further, geographic location impacts the frequency of identification of thyroid pathology. Areas of low iodine intake are associated with higher rates of identification of thyroid nodules.[82]

Thyroid nodules are readily visualized on contrasted and noncontrasted CT; dedicated imaging of the chest is the most common source of ITNs.[83] However, the sensitivity and specificity for identifying thyroid nodules is low with CT. One study correlating the CT findings with US exams and surgical pathology found that, although CT identified the dominant nodule in half of patients, it missed mulitnodularity in 30% of patients.[84] Furthermore, CT overestimated the number of nodules in 2.2% and yielded a false-positive finding in 4.3% of patients.

While thyroid nodules are identified on screening US in up to 69% of asymptomatic individuals,[85] CT exam only uncovers thyroid abnormalities in 2% to 30% of patients.[80,86] MRI of the cervical spine identifies incidental thyroid nodules in up to 20% of scans.[87] The actual incidence of thyroid nodules on CT and MRI is unclear, however, as not all radiologists convey these findings on the final report. In retrospective studies, thyroid nodules are identified on CT in 10% to 15% of scans, but those findings are only reported in a small fraction of those exams.[87,88]

Recommendations regarding the evaluation of ITNs are sparse and nonspecific. Weighing the economic and emotional costs of pursuing a thorough diagnostic evaluation of what is most likely a benign neoplasm against the low but certainly highly concerning risk of missing a clinically significant malignancy is truly a conundrum for the radiologist. Furthermore, the exponentially rising rates of thyroid cancer diagnoses, many of which are largely attributed to increased diagnostic scrutiny,[67] have intensified the need for a well-defined management strategy regarding these common clinical entities. A white paper sponsored by the American College of Radiology recommended an algorithm for evaluation of ITNs detected on CT and MRI. In this flowchart, consideration is given to concomitant suspicious CT or MRI findings, patient life expectancy and comorbidities, patient age (<35 years or ≥35 years), and nodule size (<1–1.5 cm or ≥1–1.5 cm) when determining whether to recommend further diagnostic evaluation of ITNs with thyroid US.[89] Prospective analysis of the clinical and economic impact of such a strategy is currently lacking, however.

(^{18}F)-Fluorodeoxyglucose–Positron Emission Tomography

Incidental thyroid nodules are identified on (^{18}F)-fluorodeoxyglucose–positron emission tomography (PET) in approximately 2% to 3% of examinations (Fig. 69.13A,B).[90,91] Though studies report widely varying rates of malignancy, the finding of a focal area of increased activity in the thyroid generally is associated with a risk of malignancy of 35%.[91,92] However, the evaluation of an ITN on PET should be tempered based on the underlying malignancy that prompted the examination. In one retrospective study of ITNs assessed by PET, the median overall survival from the primary malignancy that initially prompted the imaging study was 20 months. On multivariate analysis, advanced stage of the known malignancy at the time of PET and a decision not to investigate the ITN were significant predictors of mortality.[91] Of the fraction of patients who had further investigation of the ITNs, the majority represented low-risk papillary thyroid carcinomas; 98% of the confirmed thyroid cancer cases were either stable or free of disease. The high mortality in this cohort was almost exclusively due to the poor prognosis associated with the underlying initial malignancy

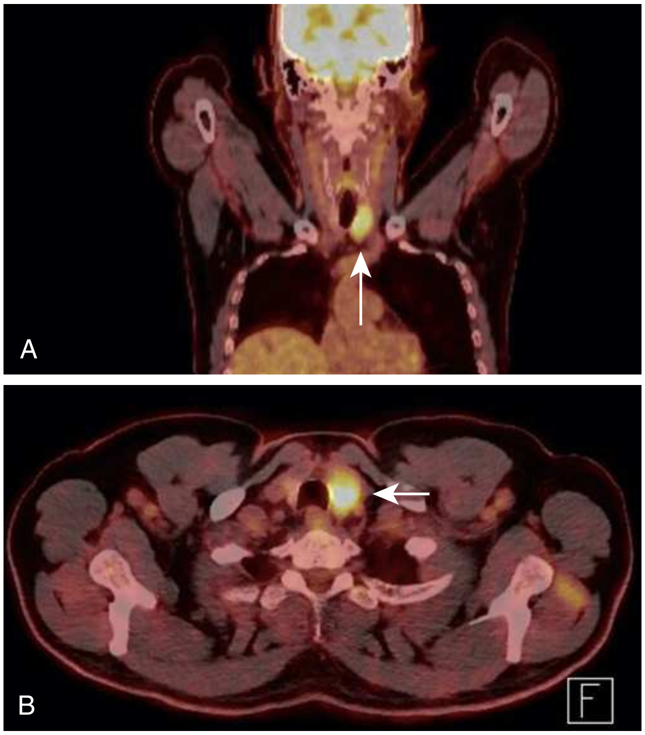


Fig. 69.13 A, Diagnostic image of an incidentally-detected (^{18}F)-fluorodeoxyglucose (FDG)-positive thyroid nodule identified on colon cancer staging scan. B, Diagnostic image of an incidentally detected FDG-avid thyroid nodule.

rather than to adverse outcomes related to the incidentally-detected thyroid cancer.[91] It is imperative, therefore, that careful consideration be given to the overall clinical context and prognosis of the underlying malignancy before undertaking further steps in the evaluation of an ITN by PET. Deliberation of whether to investigate an ITN by PET should include determination of a favorable prognosis with the underlying malignancy, consideration of patient age, and regard to comorbid conditions.

KEY POINTS

- Thyroid nodules are frequently detected incidentally on imaging of the head and neck for other reasons. These neoplasms are seldom associated with a clinically meaningful primary thyroid malignancy. Further evaluation of these incidental nodules should be performed based on the clinical context, including other primary malignancies, comorbid conditions, and patient life expectancy.

FUTURE DIRECTIONS

Although sonography and scintigraphy are well established in their role in imaging of thyroid disorders, the field continues to evolve. Continued investigation is needed to determine the appropriate applications of newer technologies such as elastography, contrasted sonography, and machine learning algorithms of sonographic risk stratification.

REFERENCES

1. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med*. 1993;328:553–559.
2. SEER Database. 2020. Accessed on 10/7/20. Available from: https://seer.cancer.gov/statfacts/html/thyro.html.
3. Haugen BR, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines Task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
4. Gharib H, et al. American association of clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules-2016 Update. *Endocr Pract*. 2016;22:622–639.
5. Patel KN, et al. The American association of Endocrine Surgeons guidelines for the Definitive surgical management of thyroid disease in Adults. *Ann Surg*. 2020;271:e21–e93.
6. Moon JH, et al. Prevalence of thyroid nodules and their associated clinical parameters: a large-scale, multicenter-based health checkup study. *Korean J Intern Med*. 2018;33(4):753–762.
7. Marqusee E, et al. Usefulness of ultrasonography in the management of nodular thyroid disease. *Ann Intern Med*. 2000;133:696–700.
8. Gutekunst R, et al. Ultrasonography related to clinical and laboratory findings in lymphocytic thyroiditis. *Acta Endocrinol*. 1989;121:129–135.
9. Nordmeyer JP, Shafeh TA, Heckmann C. Thyroid sonography in autoimmune thyroiditis. A prospective study on 123 patients. *Acta Endocrinol*. 1990;122:391–395.
10. Yeh HC, Futterweit W, Gilbert P. Micronodulation: ultrasonographic sign of Hashimoto's thyroiditis. *J Ultrasound Med*. 1996;15:813–819.
11. Ralls PW, et al. Color-flow Doppler sonography in Graves disease: "thyroid inferno." *AJR Am J Roentgenol*. 1988;150:781–784.
12. Kim WB, et al. Ultrasonographic screening for detection of thyroid cancer in patients with Graves' disease. *Clin Endocrinol*. 2004;60:719–725.
13. Staniforth JUL, Erdirimanne S, Eslick GD. Thyroid carcinoma in Graves' disease: a meta-analysis. *Int J Surg*. 2016;27:118–125.
14. Lee JH, et al. The association between papillary thyroid carcinoma and histologically proven Hashimoto's thyroiditis: a meta-analysis. *Eur J Endocrinol*. 2013;168:343–349.
15. Menon R, et al. The outcome of papillary thyroid cancer associated with Graves' disease: a case control study. *J Thyroid Res*. 2018;2018:8253094.
16. Kwon H, Moon BI. Prognosis of papillary thyroid cancer in patients with Graves' disease: a propensity score-matched analysis. *World J Surg Oncol*. 2020;18:266.
17. Cappelli C, et al. The role of imaging in Graves' disease: a cost-effectiveness analysis. *Eur J Radiol*. 2008;65:99–103.
18. Ross DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.
19. Fan C, et al. Risk factors for neck hematoma requiring surgical reintervention after thyroidectomy: a systematic review and meta-analysis. *BMC Surg*. 2019;19:98.
20. Rubio GA, et al. Postoperative outcomes in Graves' disease patients: results from the Nationwide Inpatient sample Database. *Thyroid*. 2017;27:825–831.
21. Hirokawa M, et al. Prevalence and diagnostic challenges of thyroid lymphoma: a multi-institutional study in non-Western countries. *Endocr J*. 2020;67:1085–1091.
22. Graff-Baker A, et al. Prognosis of primary thyroid lymphoma: demographic, clinical, and pathologic predictors of survival in 1,408 cases. *Surgery*. 2009;146:1105–1115.
23. Yi J, et al. A multicenter retrospective study of 58 patients with primary thyroid diffuse large B cell lymphoma. *Front Endocrinol*. 2020;11:542.
24. Luo J, et al. Is ultrasound combined with computed tomography useful for distinguishing between primary thyroid lymphoma and Hashimoto's thyroiditis? *Endokrynol Pol*. 2019;70:463–468.
25. Frates MC, et al. Subacute granulomatous (de Quervain) thyroiditis: grayscale and color Doppler sonographic characteristics. *J Ultrasound Med*. 2013;32:505–511.
26. Lee YJ, Kim DW. Sonographic characteristics and interval changes of subacute thyroiditis. *J Ultrasound Med*. 2016;35:1653–1659.
27. Cappelli C, et al. Ultrasound findings of subacute thyroiditis: a single institution retrospective review. *Acta Radiol*. 2014;55:429–433.
28. Sencar ME, et al. The contribution of ultrasonographic findings to the prognosis of subacute thyroiditis. *Arch Endocrinol Metab*. 2020;64:306–311.
29. Tessler FN, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol*. 2017;14:587–595.
30. Remonti LR, et al. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid*. 2015;25:538–550.
31. Liu H, et al. Variability in the interpretation of grey-scale ultrasound features in assessing thyroid nodules: a systematic review and meta-analysis. *Eur J Radiol*. 2020;129:109050.
32. Russ G, et al. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol*. 2013;168:649–655.
33. Rosario PW, Purisch S. Ultrasonographic characteristics as a criterion for repeat cytology in benign thyroid nodules. *Arq Bras Endocrinol Metabol*. 2010;54:52–55.
34. Ahmadi S, et al. Using the ATA and ACR TI-RADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. *Endocr Pract*. 2019;25:908–917.
35. Grani G, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "Right" TIRADS. *J Clin Endocrinol Metab*. 2019;104:95–102.
36. Ha EJ, et al. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: comparison of the 2015 American thyroid association guidelines with the 2016 Korean thyroid association/Korean Society of thyroid Radiology and 2017 American College of Radiology guidelines. *Thyroid*. 2018;28:1532–1537.
37. Yoon JH, et al. Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American thyroid association management guidelines. *Radiology*. 2016;278:917–924.
38. Peng JY, et al. Malignancy risk stratification and FNA recommendations for thyroid nodules: a comparison of ACR TI-RADS, AACE/ACE/AME and ATA guidelines. *Am J Otolaryngol*. 2020;41:102625.

39. Middleton WD, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of thyroid Radiology TIRADS, and American thyroid association guidelines. *AJR Am J Roentgenol*. 2018;210:1148–1154.
40. Moon WJ, et al. Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. *Korean J Radiol*. 2011;12:1–14.
41. Shin JH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of thyroid Radiology consensus statement and recommendations. *Korean J Radiol*. 2016;17:370–395.
42. Kim PH, et al. Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. *Thyroid*. 2020;30:1159–1168.
43. Jeh SK, et al. Evaluating the degree of conformity of papillary carcinoma and follicular carcinoma to the reported ultrasonographic findings of malignant thyroid tumor. *Korean J Radiol*. 2007;8:192–197.
44. Kim DS, et al. Sonographic features of follicular variant papillary thyroid carcinomas in comparison with conventional papillary thyroid carcinomas. *J Ultrasound Med*. 2009;28:1685–1692.
45. Trimboli P, et al. The ultrasound risk stratification systems for thyroid nodule have been evaluated against papillary carcinoma. A meta-analysis. *Rev Endocr Metab Disord*; 2021;22(2):453–460.
46. Lee H, et al. Microbubbles used for contrast enhanced ultrasound and theragnosis: a review of principles to applications. *Biomed Eng Lett*. 2017;7:59–69.
47. Zhan J, Ding H. Application of contrast-enhanced ultrasound for evaluation of thyroid nodules. *Ultrasonography*. 2018;37:288–297.
48. Xie X, Yu Y. Effect of the location and size of thyroid nodules on the diagnostic performance of ultrasound elastography: a retrospective analysis. *Clinics*. 2020;75:e1720.
49. Asteria C, et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. *Thyroid*. 2008;18:523–531.
50. Rago T, et al. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. *J Clin Endocrinol Metab*. 2007;92:2917–2922.
51. Hegedus L. Can elastography stretch our understanding of thyroid histomorphology? *J Clin Endocrinol Metab*. 2010;95:5213–5215.
52. Lippolis PV, et al. Is elastography actually useful in the presurgical selection of thyroid nodules with indeterminate cytology? *J Clin Endocrinol Metab*. 2011;96:E1826–E1830.
53. Rago T, et al. Real-time elastosonography: useful tool for refining the presurgical diagnosis in thyroid nodules with indeterminate or nondiagnostic cytology. *J Clin Endocrinol Metab*. 2010;95:5274–5280.
54. Moon HJ, et al. Diagnostic performance of gray-scale US and elastography in solid thyroid nodules. *Radiology*. 2012;262:1002–1013.
55. Yang BR, et al. Qualitative and semiquantitative elastography for the diagnosis of intermediate suspicious thyroid nodules based on the 2015 American Thyroid Association Guidelines. *J Ultrasound Med*. 2018;37:1007–1014.
56. Qiu Y, et al. Diagnostic reliability of elastography in thyroid nodules reported as indeterminate at prior fine-needle aspiration cytology (FNAC): a systematic review and Bayesian meta-analysis. *Eur Radiol*. 2020;30:6624–6634.
57. Huang S, et al. Diagnostic performances of the KWAK-TIRADS classification, elasticity score, and Bethesda System for Reporting Thyroid Cytopathology of TI-RADS category 4 thyroid nodules. *Int J Clin Exp Pathol*. 2020;13:1159–1168.
58. Luo J, et al. A retrospective study of reducing unnecessary thyroid biopsy for American College of Radiology Thyroid Imaging Reporting and Data Systems 4 assessment through applying shear wave elastography. *Arch Endocrinol Metab*. 2020;64:349–355.
59. Okasha HH, et al. Role of elastography strain ratio and TIRADS score in predicting malignant thyroid nodule. *Arch Endocrinol Metab*. 2021;64(6):735–742.
60. Pei S, et al. Ultrasound real-time tissue elastography improves the diagnostic performance of the ACR Thyroid Imaging Reporting and Data System in differentiating malignant from benign thyroid nodules: a summary of 1525 thyroid nodules. *Int J Endocrinol*. 2020;2020:1749351.
61. Brito JP, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2014;99:1253–1263.
62. Choi SH, et al. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid*. 2010;20:167–172.
63. Park CS, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound*. 2010;38:287–293.
64. Chung SR, et al. Computer-aided diagnosis system for the evaluation of thyroid nodules on ultrasonography: prospective non-inferiority study according to the experience level of radiologists. *Korean J Radiol*. 2020;21:369–376.
65. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
66. Xu L, et al. Computer-aided diagnosis systems in diagnosing malignant thyroid nodules on ultrasonography: a systematic review and meta-analysis. *Eur Thyroid J*. 2020;9:186–193.
67. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg*. 2014;140s:317–322.
68. Ha EJ, Baek JH. Applications of machine learning and deep learning to thyroid imaging: where do we stand? *Ultrasonography*. 2021;40:23–29.
69. Li X, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol*. 2019;20:193–201.
70. Reschini E, et al. The trapping-only nodules of the thyroid gland: prevalence study. *Thyroid*. 2006;16:757–762.
71. Treglia G, et al. Prevalence of normal TSH value among patients with autonomously functioning thyroid nodule. *Eur J Clin Invest*. 2015;45:739–744.
72. Keller-Petrot I, et al. Congenital hypothyroidism: role of nuclear medicine. *Semin Nucl Med*. 2017;47:135–142.
73. Leger J, et al. European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. *Horm Res Paediatr*. 2014;81:80–103.
74. Fugazzola L, et al. Thyroid scintigraphy and perchlorate test after recombinant human TSH: a new tool for the differential diagnosis of congenital hypothyroidism during infancy. *Eur J Nucl Med Mol Imaging*. 2007;34:1498–1503.
75. Dias VM, et al. Congenital hypothyroidism: etiology. *J Pediatr Endocrinol Metab*. 2010;23:815–826.
76. Banks CA, et al. Thyroid disease and compressive symptoms. *Laryngoscope*. 2012;122:13–16.
77. Katlic MR, Grillo HC, Wang CA. Substernal goiter. Analysis of 80 patients from Massachusetts general Hospital. *Am J Surg*. 1985;149:283–287.
78. Al-Bazzaz F, Grillo H, Kazemi H. Response to exercise in upper airway obstruction. *Am Rev Respir Dis*. 1975;111:631–640.
79. Padovani RP, et al. One month is sufficient for urinary iodine to return to its baseline value after the use of water-soluble iodinated contrast agents in post-thyroidectomy patients requiring radioiodine therapy. *Thyroid*. 2012;22:926–930.
80. Khasawneh A, et al. Incidental findings in the thyroid gland on computed tomography images of the oral and maxillofacial region. *Oncol Lett*. 2020;19:2005–2010.
81. Drake T, et al. Prevalence of thyroid incidentalomas from 1995 to 2016: a single-center, retrospective cohort study. *J Endocr Soc*. 2020;4:bvz027.
82. Guth S, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest*. 2009;39:699–706.
83. Grady AT, et al. Radiology reports for incidental thyroid nodules on CT and MRI: high variability across subspecialties. *AJNR Am J Neuroradiol*. 2015;36:397–402.
84. Shetty SK, et al. Significance of incidental thyroid lesions detected on CT: correlation among CT, sonography, and pathology. *AJR Am J Roentgenol*. 2006;187:1349–1356.
85. Ezzat S, et al. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. *Arch Intern Med*. 1994;154:1838–1840.
86. Lee JH, Jeong SY, Kim YH. Clinical significance of incidental thyroid nodules identified on low-dose CT for lung cancer screening. *Multidiscip Respir Med*. 2013;8:56.

87. Youserm DM, et al. Clinical and economic impact of incidental thyroid lesions found with CT and MR. *AJNR Am J Neuroradiol*. 1997;18:1423–1428.
88. Bahl M. Incidental thyroid nodules in the National Lung Screening Trial: estimation of prevalence, malignancy rate, and strategy for workup. *Acad Radiol*. 2018;25:1152–1155.
89. Hoang JK, et al. Managing incidental thyroid nodules detected on imaging: white paper of the ACR Incidental Thyroid Findings Committee. *J Am Coll Radiol*. 2015;12:143–150.
90. Chen W, et al. Evaluation of thyroid FDG uptake incidentally identified on FDG-PET/CT imaging. *Nucl Med Commun*. 2009;30:240–244.
91. Pattison DA, et al. 18)F-FDG-Avid thyroid incidentalomas: the importance of Contextual interpretation. *J Nucl Med*. 2018;59:749–755.
92. Bertagna F, et al. Diagnostic and clinical significance of F-18-FDG-PET/CT thyroid incidentalomas. *J Clin Endocrinol Metab*. 2012;97:3866–3875.

70

Pathology of the Thyroid Gland

Zubair W. Baloch

OUTLINE

Thyroid Gland, 1166
Pathologic Assessment of Thyroid Lesions, 1166
- Pathology of Thyroid Lesions, 1167
- Diffuse Enlargements of Thyroid Gland, 1167
- Nodular Enlargements and Nodules, 1168
- Benign Neoplasms, 1168
- Malignant Neoplasms, 1169
- Papillary Thyroid Carcinoma, 1169
- Variants of Papillary Cancer, 1170
- Follicular Carcinoma, 1172
- Hürthle Cell (Oncocytic Follicular) Tumors, 1172
- High-Grade Thyroid Carcinoma, 1173
- Poorly Differentiated Carcinoma, 1173
- Anaplastic Carcinoma, 1173
- Familial Follicular-Derived Thyroid Tumors, 1173
- Medullary Carcinoma, 1173

Tumors with Thymic or Related Branchial Pouch Differentiation, 1174
Rare Tumors of Thyroid Gland, 1174
- Malignant Tumors Arising in Ectopic Thyroid Tissue, 1174
- Secondary Tumors, 1174

Summary and Future Directions, 1174

THYROID GLAND

The *normal thyroid gland* consists of right and left lobes that are connected by an isthmus (the so called *"butterfly gland"*). The thyroid capsule is incomplete (especially posteriorly and in the region of the isthmus), thin, and cannot be easily separated from the thyroid gland. The normal thyroid is soft and light brown to red in color with a glistening cut surface owing to colloid. The functional unit of the thyroid gland is the follicle. These are oval to spherical structures of varying sizes formed by a single layer of low-cuboidal epithelium encircling a lumen containing colloid, partly composed of thyroglobulin. The individual follicular cells show nuclei with even nuclear chromatin and inconspicuous nucleoli.[1]

The calcitonin secreting C-cells (also known as [AKA] parafollicular cells) are housed next to follicular cells within the confines of the basement membranes. The C-cells are more numerous in infants and children than in adults and are mainly seen in the central portions of the middle and upper thirds of the thyroid lobes. Interestingly, the well-publicized neural crest origin of the C-cells is now challenged by studies conducted in mouse embryos showing that C-cells arise from pharyngeal endoderm.[1,2]

The Hürthle cells, also termed "oncocytic cells," are metaplastic follicular cells with cytoplasmic eosinophilia because of enlarged mitochondria and enlarged round nuclei with prominent nucleoli. These can be seen in normal thyroid glands, but are commonly encountered in diffuse toxic goiter, longstanding goiterous nodules, chronic lymphocytic thyroiditis, and neoplasms.[1,3]

PATHOLOGIC ASSESSMENT OF THYROID LESIONS

As mentioned, the pathologic assessment of thyroid lesions includes cytopathologic examination of preoperative fine needle aspiration (FNA) specimens, intraoperative assessment/frozen section, and gross pathologic and histopathologic evaluation of surgically excised specimens.

FNA of thyroid nodules is usually performed under ultrasound (US) guidance rather than palpation. The rapid on-site evaluation (ROSE) of thyroid FNA specimens is performed by a cytopathology professional to assess the adequacy of a specimen, render preliminary diagnosis in some cases, and help with specimen triage for ancillary studies (flow cytometry, immunohistochemistry, reflex molecular studies). It has been shown that ROSE can significantly decrease the nondiagnostic rate. FNA specimens can be prepared for cytopathologic evaluation by any of the following methods: air-dried smears and alcohol-fixed smears and monolayer preparation (usually Cytolyt® solution for Thin-Prep®), cytospin, and/or cellblock prepared from the needle rinse. Thyroid FNA specimens are usually classified using a tiered reporting scheme known as the Bethesda System for Reporting Thyroid Cytology (BSRTC). Each diagnostic category of the BSRTC is associated with an assigned risk of malignancy, along with recommendations for management (Table 70.1).[4,5]

It is well known that *thyroid frozen section* diagnosis is specific but not sensitive. Most experts agree that the initial approach to diagnosing a thyroid nodule should be FNA.[6,7] With an unequivocal diagnosis of malignancy on FNA, the surgeon should proceed with the appropriate surgery for that malignant diagnosis. For thyroid nodules diagnosed as "neoplasm/suspicious for neoplasm," the limited sampling at frozen section is unlikely to be fruitful in finding a random microscopic focus of capsular or vascular invasion required for the diagnosis of follicular carcinoma.[8] However, in the current era of adjunct molecular testing of thyroid FNA specimens and conservative surgical approach to thyroid lesions, frozen section is rarely performed.

TABLE 70.1 The Bethesda System for Reporting Thyroid Cytopathology

Diagnostic Category	Risk of Malignancy (%)	Possible Management Paradigm
Nondiagnostic or unsatisfactory	5–10	Repeat FNA with ultrasound guidance
Benign	0–3	Clinical and sonographic follow-up
Atypia of undetermined significance or follicular lesion of undetermined significance	~10–30	Repeat FNA, molecular testing, or lobectomy
Follicular neoplasm or suspicious for a follicular neoplasm	25–40	Molecular testing, lobectomy
Suspicious for malignancy	50–75	Near-total thyroidectomy or lobectomy[a]
Malignant	97–99	Near-total thyroidectomy

FNA, Fine-needle aspiration.

[a] Molecular testing can be used to inform the extent of surgery.

From The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid.* 2017;27:1341–1346

KEY POINTS

- The limited sampling associated with taking frozen sections is unlikely to be fruitful in finding a random microscopic focus of capsular or vascular invasion required for the diagnosis of follicular carcinoma.

Gross pathologic examination of a thyroid is usually performed on a fresh specimen, and tumor size and appearance are documented before sections are taken. In the case of a dominant encapsulated nodule, it is recommended that the entire circumference of the nodule be sectioned, whereas for a nonencapsulated nodule one section per 0.5 cm can be submitted.[9]

KEY POINTS

- It is recommended that entire circumference of a dominant encapsulated nodule be sectioned and submitted for histopathologic examination to exclude invasive characteristics.

The *final histopathologic report* should be detailed and include all of the known prognostic parameters according to the latest synoptic reporting system. Based on the pathologic assessment and provided clinical information, the tumor stage should be documented based on the current version of the tumor-node-metastasis staging system.[9]

Pathology of Thyroid Lesions

Developmental Variations. By light microscopy the thyroglossal duct is usually lined by ciliated pseudostratified epithelium; however, the cell type may be transitional or squamous if the duct is traumatized or infected, or it may be partially or completely lost and replaced by fibrous tissue. If fluid accumulates in part of the thyroglossal duct, a thyroglossal cyst may develop. Any type of diffuse thyroid disease can involve lingual thyroid and the thyroid tissue along the thyroglossal duct tract.[10,11] Parathyroid glands, thymic tissue, small collections of cartilage, and tiny glands lined by ciliated cells may be seen in normal thyroid glands, presumably related to the development of the branchial pouches.[11]

Because of developmental variations, strips of striated muscle and adipose tissue can be encountered embedded within the thyroid parenchyma. Thyroid tissue can also be seen in perithyroidal skeletal muscle, and they should not be confused with infiltrating or metastatic carcinoma.[1]

Diffuse Enlargements of Thyroid Gland

Graves Disease/Diffuse Toxic Goiter. Grossly, the thyroid gland is diffusely enlarged up to several times the normal size.[12] The cut surfaces are fleshy and lack normal translucence because of the loss of colloid.

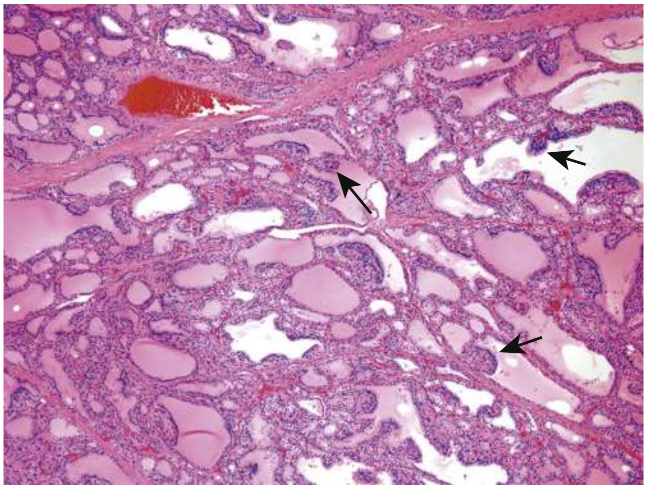


Fig. 70.1 Graves disease. Photomicrograph showing diffuse papillary hyperplasia. The individual papillae are highlighted by arrows (hematoxylin and eosin stain, 20x).

- Microscopically, there is follicular cell hypertrophy and hyperplasia. The follicular cells are tall and columnar and form papillary folds that extend into the lumina of the follicles. Infiltrates of lymphocytes are seen within the interfollicular stroma (Fig. 70.1).
- FNA biopsies are usually not performed in patients with Graves disease; only patients with one or more nodules suspicious on US are selected for FNA.[13] The FNA specimens are cellular and show features of hyperplastic goiter with lymphocytes and oncocytic cells.[14,15]

Chronic Lymphocytic Thyroiditis.

- Grossly, the thyroid gland may appear normal in size, as large goiter or atrophic, depending upon the severity and stage of involvement by chronic lymphocytic thyroiditis.[16]
- Microscopically, focal chronic lymphocytic thyroiditis is characterized by patches of lymphocytic infiltrates within thyroid parenchyma with or without associated Hürthle cell metaplasia. In more severe cases, there is diffuse lymphocytic infiltration of thyroid parenchyma with lymphoid germinal center formation, follicular cell atrophy, Hürthle cell metaplasia, and fibrosis (Fig. 70.2). In more advanced cases, little or no thyroid parenchyma is seen and is replaced by fibrosis and nodules of remaining atrophic thyroid tissue with lymphoid germinal centers.[16,17]
- FNA is usually performed in patients with chronic lymphocytic thyroiditis who present with distinct nodules or show a suspicious area on US examination (hypoechogenicity, increased vascularity,

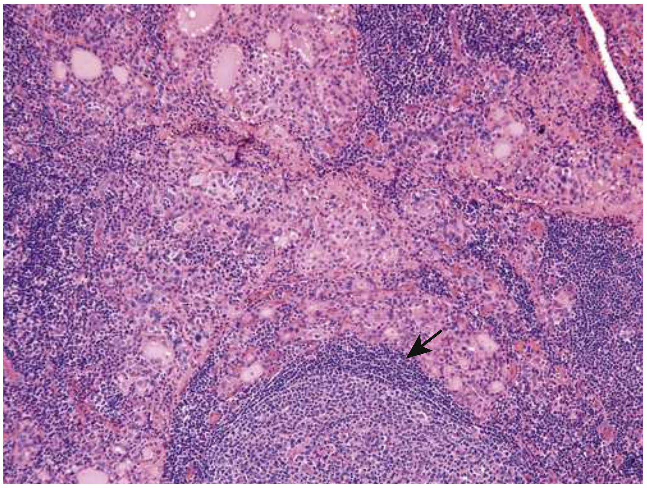


Fig. 70.2 **Chronic lymphocytic thyroiditis.** Photomicrograph showing islands of follicular cells with Hürthle cell metaplasia and brisk lymphocytic infiltrate with lymphoid germinal centers *(arrow)* (hematoxylin and eosin stain, 20×).

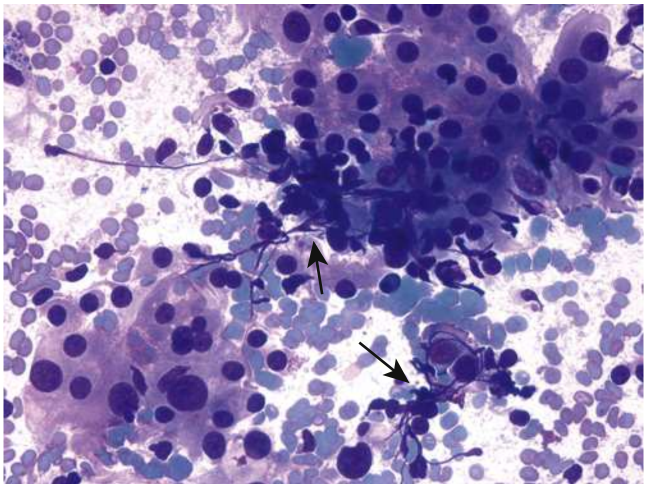


Fig. 70.3 **Chronic lymphocytic thyroiditis.** Photomicrograph of fine-needle aspiration specimen showing Hürthle cells infiltrated by lymphocytes *(arrows)* (air-dried smear preparation, Diff-Quik® stain, 40×).

microcalcifications). The cytologic preparations show follicular cells and Hürthle cells in a background of lymphocytes with little or no colloid. The lymphocytes are usually seen in the background and infiltrating the follicular and/or Hürthle cell groups (Fig. 70.3).[18]

- The other forms of thyroiditis encountered in thyroid surgical specimens include Riedel thyroiditis (AKA Riedel struma), combined Riedel's struma[19] and fibrosing chronic lymphocytic thyroiditis,[20] nonspecific-focal chronic lymphocytic thyroiditis, granulomatous/subacute thyroiditis,[21] palpation thyroiditis,[22] and, rarely, specific infectious thyroiditis.[23]

Nodular Enlargements and Nodules

Nontoxic Nodular Goiter.

- The pathologic stages of the nontoxic nodular goiter include one or more of the following microscopic features:[24,25]
    - Hyperplastic stage: This represents response of the thyroid follicular cells to thyroid-stimulating hormone or other growth and/or stimulants. The follicular cells are tall and columnar in shape, with enlarged and hyperchromatic nuclei. The epithelial hyperplasia may also lead to papillary architecture.
    - Colloid accumulation stage (AKA colloid goiter): The thyroid follicles may undergo involution, and the epithelium become low-cuboidal and flattened. Some follicles grow larger than normal and contain excessive colloid. On gross examination the gland has a characteristic glistening cut surface because of excess colloid accumulation.
    - Nodular stage: The persistence of hyperplasia along with involution of some thyroid follicles gives rise to a heterogeneous appearance of the thyroid gland. The vascular network is altered because of overgrowth, leading to degenerative changes with ensuing fibrosis. This combination leads to poorly demarcated yet well-circumscribed nodules known as hyperplastic/adenomatoid nodules.
- The FNA specimen from a hyperplastic/adenomatoid nodule in multinodular goiter shows the following cytomorphologic features:[26]
    - Abundant thin and watery colloid, small round- to oval-shaped follicular cells with dark nuclei arranged in monolayer sheets, follicular groups, or single cells (Fig. 70.4).
    - The cytoplasm is usually scant in follicular cells and can show numerous, small, blue-black granules.

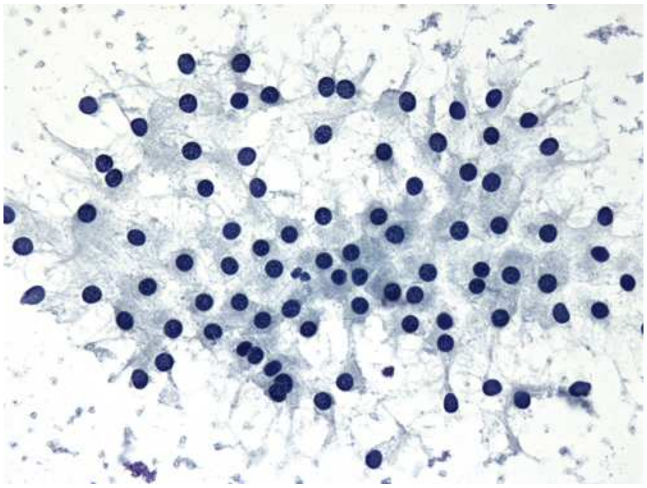


Fig. 70.4 **Nodular goiter.** Photomicrograph of fine-needle aspiration specimen showing loosely cohesive group of follicular cells with small round nuclei (alcohol-fixed smear preparation, Papanicolaou stain, 40×).

    - Macrophages are usually present, and their number correlates with presence or absence of degenerative changes or a cystic component
    - The FNA specimens from a hyperplastic/adenomatoid nodule usually are cellular and contain an admixture of follicular cells and oncocytic (Hürthle) cells arranged in monolayer sheets or three-dimensional groups in a background of thin watery colloid and macrophages.

Thyroid Neoplasms. All neoplasms that arise from thyroid epithelium can have some functional capacities. They may respond to TSH and even produce excessive amounts of thyroid hormones or, if medullary carcinoma, may release abnormal quantities of calcitonin or other peptide hormones.[1,27] Immunohistochemical localization of thyroid transcription factor-1 (TTF-1), thyroglobulin, paired box 8 (PAX8), or calcitonin aids in the classification of unusual thyroidal tumors and in providing definite identification of metastatic thyroid carcinomas.[1,27]

Benign Neoplasms

Adenomas and Adenomatous Nodules.

- Classically, the follicular adenoma is described as a solitary, encapsulated lesion having a uniform internal architecture that is different

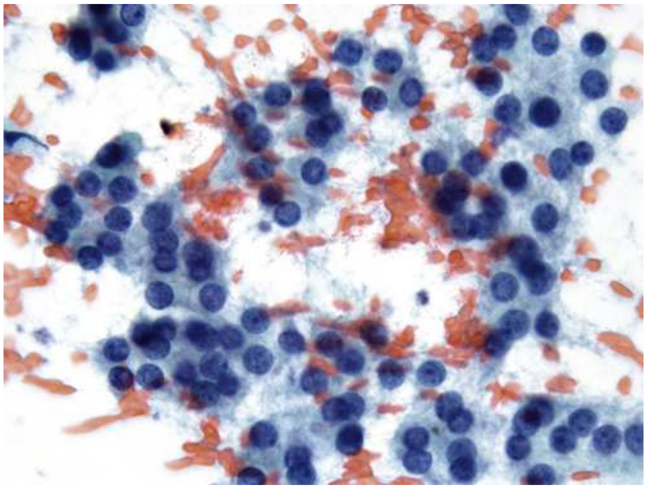


Fig. 70.5 Follicular neoplasm/suspicious for follicular neoplasm. Photomicrograph of fine-needle aspiration specimen showing a monotonous population of follicular cells arranged in microfollicles. Notice the round nuclei with even chromatin pattern (alcohol-fixed smear preparation, Papanicolaou stain, 40×).

from the surrounding thyroid based on growth pattern and cellular features.[1] However, it is often difficult based on morphologic features to clearly separate hyperplastic/adenomatous nodules arising in the background of goiter or thyroiditis from adenomas.

- By molecular analysis, up to 70% of the hyperplastic/adenomatoid nodules in goiter are clonal, thus representing neoplastic proliferations.[28] Therefore, it would not be farfetched to expand the pathologic definition of follicular adenoma to a follicular-derived partially or completely encapsulated or circumscribed nodule with a distinct growth pattern that is different from the surrounding thyroid parenchyma and lacks invasive features. These can be multiple and arise in a background of normal thyroid or in the setting of nodular goiter, toxic goiter, and thyroiditis.[1]
- By gross pathologic examination, follicular adenomas and adenomatous nodules are well-circumscribed and are often demarcated from the adjacent tissue. Their size varies from about 1 mm to several centimeters in diameter. A classic adenoma is fleshy and pale; hemorrhage, fibrosis, and cystic change may be evident. Calcifications may be seen in longstanding lesions.[1,27]
- By light microscopy, an adenoma or an adenomatous nodule is encapsulated, shows follicular, solid or rarely papillary growth patterns, and lacks capsular and/or lymphovascular invasion. The lesional cells can be cuboidal or columnar in shape with small, round nuclei with an even chromatin pattern.[27]
- In some cases, the adenoma can be hyperfunctioning (AKA toxic adenoma). In adolescents, especially females, many hot, or toxic, nodules contain numerous papillae, often sufficient in number to cause a pathologist to suggest a diagnosis of papillary carcinoma.[29]
- By immunohistochemistry, all adenomatous nodules and adenomas express TTF-1 and thyroglobulin and show cytokeratin expression similar to normal thyroid parenchyma.[30]
- FNA specimens obtained from follicular adenoma are cellular and show a monotonous population of follicular cells arranged in cohesive groups with nuclear overlapping and crowding and/or microfollicles. The individual cells are enlarged (compared with background red blood cells) and contain round nuclei with an even chromatin pattern (Fig. 70.5).
- Post-FNA, changes can be seen on gross examination as focal hemorrhage and on microscopic examination in the form of hemorrhage, fibrosis, endothelial proliferation, pseudovascular and capsular invasion, and even partial or total infarction.[31-33] These changes are often prominent in cases of Hürthle cell adenoma or adenomatous nodule.
- Several morphologic variants of follicular adenoma have been described. These include:
  - *Atypical follicular adenoma*: This term is reserved for cases that exhibit some atypical features, including foci of necrosis, excess cellularity, mitoses, and lack of capsular and/or vascular invasion. The recent World Health Organization (WHO) classification includes the term "follicular tumor of uncertain malignant potential" for the encapsulated follicular pattern lesions with irregular tumor capsule interface, incomplete capsular involvement, and mitotic foci. The great majority of such cases behave as benign adenomas.[27]
  - *Hyalinizing trabecular neoplasm of the thyroid* (HTN), AKA *hyalinizing trabecular adenoma*: This is a morphologically distinct follicular0derived lesion. Microscopically, the lesional cells grow in nests that are surrounded by dense hyaline stroma. The nuclear features of the follicular cells are similar to those seen in papillary carcinoma. However, a benign behavior has so far been described in all cases of hyalinizing trabecular adenoma that show classical histology. By immunohistochemistry, the cells of hyalinizing trabecular adenoma stain positive for thyroglobulin and cytokeratin-19 and negative for calcitonin. *RET/PTC* rearrangements have been detected in HTN; however, recent studies show that *PAX8-GLIS3* and *PAX8-GLIS1* fusions are specific alterations of HTN.[34,35]
  - The other morphologic variants of follicular adenoma include, signet ring cell follicular adenoma, spindle cell follicular adenoma, and lipoadenoma adenomlipoma.[35]

Malignant Neoplasms

The most common malignant neoplasms that originate in the thyroid are well-differentiated carcinomas of follicular cell origin; most are papillary carcinomas (Table 70.2).[1,27]

Papillary Thyroid Carcinoma

Papillary thyroid carcinoma (PTC) comprises approximately 80% of thyroid carcinomas affecting adults and children in countries with sufficient dietary iodine intake. PTC occurs frequently in women and is rarely familial.[1,13,36]

- Grossly, papillary cancers can present in various sizes, and are predominantly solid, although small cystic foci may be present. Cystic cancers can occur, with one or more cystic spaces occupying most of the neoplasm.[1]
- Small papillary carcinomas (*microcarcinomas*) are defined as being 1 cm or less in diameter. The prevalence of these lesions ranges from 6% to 36% (depending upon the studied population and how thoroughly the thyroid gland was examined). When visible on gross examination, they present as small, irregular, firm scars, as soft foci of discoloration, or as tiny calcified lesions.[1,27]
- Readily identifiable calcifications or specks of calcified material and crystals may be present in the cyst fluid. The cut surface of tumor may appear granular because of exuberant papillary formations.[27]
- Fibrosis is common in and around papillary carcinomas, and it may be distributed in an extremely irregular fashion, grossly and microscopically.
- Most PTCs are infiltrative within the thyroid gland; however, about 10% to 20% are grossly encapsulated. Encapsulation of the primary carcinoma is associated with a lower frequency of lymph node involvement.[1,37,38]

TABLE 70.2 World Health Organization Pathologic Classification of Thyroid Neoplasm of Follicular Cell Origin

Neoplasms

Other Encapsulated Follicular-Patterned Thyroid Tumors

1. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features
2. Follicular tumor of uncertain malignant potential
3. Well-differentiated tumor of uncertain malignant potential

Papillary Thyroid Carcinoma

1. Conventional type
2. Follicular
3. Encapsulated
4. Microcarcinoma
5. Tall cell
6. Columnar cell
7. Oncocytic
8. Hobnail
9. Diffuse sclerosing
10. Cribriform morular
11. Solid/trabecular
12. Warthin-like
13. Others

Follicular Thyroid Carcinoma

1. Minimally invasive
2. Encapsulated angioinvasive
3. Widely invasive

Hürthle Cell Carcinoma

1. Minimally invasive
2. Encapsulated angioinvasive
3. Widely invasive

Poorly Differentiated Carcinoma

Anaplastic/Undifferentiated Thyroid Carcinoma

From *WHO Classification of Tumours of Endocrine Organs.* 4th ed. Vol 10. Lyon: WHO/IARC Press; 2017.

- By light microscopy, PTC contain papillae. The papillae contain well-developed fibrovascular cores and are covered with a single layer of the epithelial cells; the amount of true papillae can be variable within a PTC (Fig. 70.6).[1,27]
- PTC may show a solid growth pattern, may be composed of follicles (follicular variant), or may be almost entirely papillary.[1,27]
- The microscopic diagnosis of PTC is based on characteristic nuclear cytology of lesional cells. These are usually cuboidal to low columnar and contains nuclei that are relatively large and irregular in shape, with pale, empty, or "ground glass" chromatin, folds/grooves, indentations, and cytoplasmic inclusions. The nucleolus is often inconspicuous and eccentric in location (lies near the nuclear membrane) (Fig. 70.6).[39,40 27,41]
- About one-third to one-half of papillary carcinomas form laminated calcific spherules known as psammoma bodies. Anytime a psammoma body is found in normal thyroid tissue, cervical lymph nodes, or perithyroidal soft tissue, PTC should be excluded by careful screening of histologic sections.[42,43] Structures resembling psammoma bodies are occasionally found within the colloid of follicles in adenomas or adenomatous nodules, especially those composed of oxyphilic cells, where they seem to arise from calcification of inspissated colloid; these structures should not be equated with psammoma bodies.[44,45]

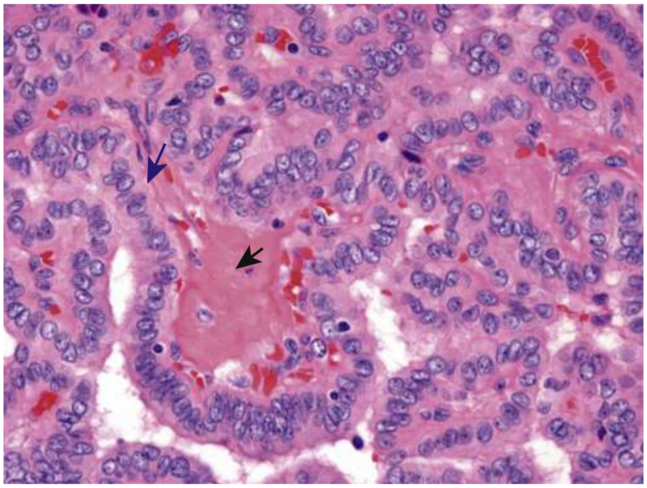


Fig. 70.6 **Papillary thyroid carcinoma, classic variant.** Photomicrograph showing complex papillae with fibrovascular core *(black arrow)* and tumor cells lining the papillae with nuclear features of papillary thyroid carcinoma *(blue arrow)* (hematoxylin and eosin stain, 40×).

- Lymphocytic infiltration is often present within and around papillary carcinomas.
- Multiple intrathyroidal foci of tumor and metastasis to cervical lymph nodes are common and likely attributed to lymphatic invasion by tumor.[46]
- If a nodal metastasis is cystic, it must be differentiated from other cystic benign lesions such as branchial cleft cyst.[47]
- Blood vessel invasion (angioinvasion) by papillary carcinoma is rare, and metastatic foci in distant sites are unusual (~5%–7% of cases), with the lungs most frequently involved.[48]
- The diagnosis of PTC in FNA specimens is based on characteristic nuclear morphology, regardless of cytoplasmic and architectural features.[26,49]
  - The specimens are usually cellular, and the tumor cells are arranged as papillary groups (Fig. 70.7A and B), three-dimensional tissue fragments, or single cells in the background of thin or thick colloid (AKA "ropy" or "chewing gum" colloid). Some cases can show calcifications, psammoma bodies, or variable numbers of macrophages as a result of cystic changes.
  - The individual tumor cells are enlarged and oval in shape, with nuclei showing chromatin clearing, membrane thickening, intranuclear grooves, and pseudoinclusions (Fig. 70.8A and B).
  - Some cases may also demonstrate multinucleated giant cells.

KEY POINTS

- The diagnosis of papillary thyroid carcinoma in fine-needle aspiration specimens is based on characteristic nuclear morphology, regardless of cytoplasmic and architectural features.

Variants of Papillary Cancer

Follicular variant of papillary thyroid carcinoma (FVPTC) is the most common histologic variant of papillary carcinoma after the usual or classic variant; it is characterized by formation of follicles lined by cells with nuclear features of papillary carcinoma. FVPTC clinically behaves similar to the conventional papillary carcinoma.[50] Although FVPTC is a variant of papillary carcinoma (based on cytology and clinical

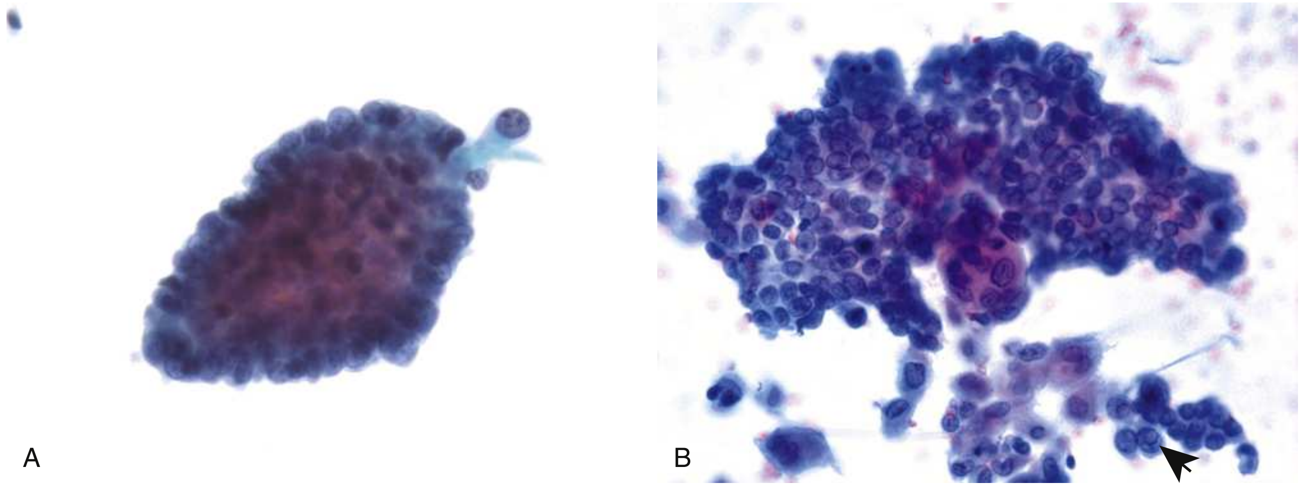


Fig. 70.7 Papillary carcinoma. Photomicrographs of fine-needle aspiration specimen showing papillary formation containing tumor cells. Note the intranuclear inclusion in B *(arrow)* (alcohol-fixed smear preparations, Papanicolaou stain, 40×).

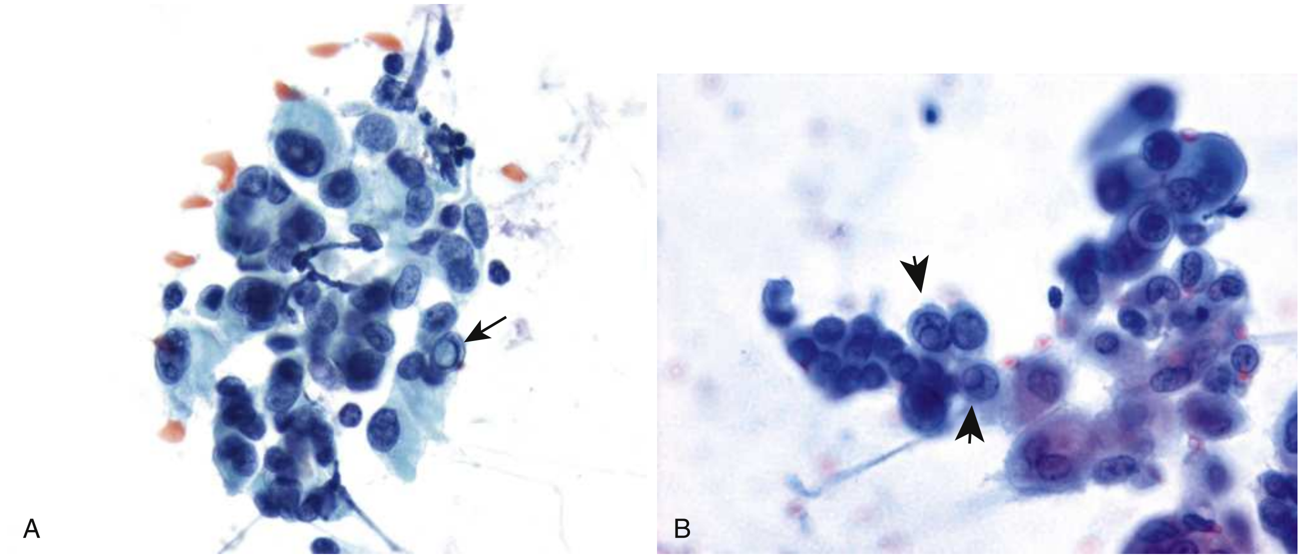


Fig. 70.8 Papillary thyroid carcinoma. Photomicrographs of fine needle aspiration specimen showing prominent intranuclear inclusions *(arrow)* (alcohol fixed smear preparations, Papanicolaou stain, 60×).

behavior), it does share some morphologic and clinical features with follicular carcinoma. These include a follicular growth pattern, encapsulation, capsular and vascular invasion, and distant metastases to lung and bone. Gene expression profiling studies have shown that some cases of FVPTC will have a similar profile to follicular carcinoma.[51,52]

FVPTC can present as the following:

- The *encapsulated variant* refers to the follicular variant, which is characterized by the presence of a capsule around the lesion. These lesions are associated with an excellent prognosis.[1] Some of these cases are readily identifiable as carcinoma because of presence of invasion into the tumor capsule and/or vascular invasion. However, cases without invasive characteristics and with the subtle presence of nuclear features of PTC can be mistaken for adenomatoid nodule or follicular adenoma.[40] In 2015, a multidisciplinary panel evaluated a large cohort of patients with FVPTC with clinical follow-up ranging from 10 to 26 years confirmed that noninvasive encapsulated FVPTC is free of recurrences and any metastatic involvement.[53] The panel suggested removing the word "carcinoma" for noninvasive encapsulated FVPTC and renaming it "noninvasive follicular thyroid neoplasm with papillary-like nuclear features" (NIFTP). This entity was defined by a set of histopathologic features including a noninvasive follicular patterned neoplasm with nuclear cytology of PTC, lack of any necrosis, and papillary architecture (Fig. 70.9).[53,54]
- Unencapsulated and infiltrative tumor.
- The diffuse follicular variant, which diffusely involves the thyroid, leads to total replacement of thyroid by tumor. Both lymph node and distant metastases (pulmonary) are common in this variant.[50,55]
- The macrofollicular variant of FVPTC is composed almost entirely of distended, colloid-filled follicles of moderately uniform size and shape, and the characteristic nuclei of PTC confirm the diagnosis.[56,57]

KEY POINTS

- Noninvasive follicular thyroid neoplasm with papillary-like nuclear features, formerly classified as noninvasive follicular variant of papillary thyroid carcinoma, is defined by a set of histopathologic features including a noninvasive follicular patterned neoplasm with nuclear cytology of papillary thyroid carcinoma, lack of any necrosis, and papillary architecture.

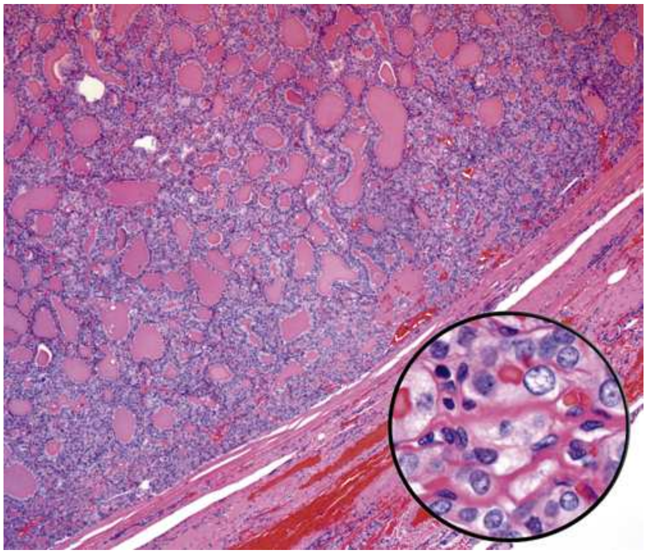


Fig. 70.9 **Noninvasive follicular thyroid neoplasm with papillary like nuclear features.** Photomicrograph showing a encapsulated follicular patterned neoplasm with nuclear features of papillary thyroid carcinoma *(inset)* (hematoxylin and eosin stain, 10×; inset 40×).

FNA specimens from FVPTC show enlarged follicular cells arranged in monolayer sheets, as well as follicles of varying sizes and shapes in a background of thin watery colloid. Usually, the diagnostic nuclear features are scarce, because of which some specimen will diagnosed as either follicular neoplasm/suspicious for follicular neoplasm (Bethesda IV) or suspicious for PTC (Bethesda V).[26]

The *tall cell variant* is considered as an aggressive variant of PTC. It is commonly occurs at older age, and a mortality rate of 25% has been reported.[58]

- The tumor cells are elongated (three times as long as they are wide) and contain eosinophilic cytoplasm.
- The tall cell variant often shows extrathyroidal soft tissue extension and vascular invasion (20%–25%).[59,60]
- The FNA specimen tall cell variant of PTC shows oncocytic tumor cells with readily identifiable nuclear features of PTC. The intranuclear inclusions can small and multiple, giving rise to so-called “soap bubble” appearance.[61,62]

The other less common variants of PTC include:

- *Diffuse sclerosis variant* (DSV), which diffusely permeates lymphatic channels within one lobe or of the entire thyroid and is accompanied by severe lymphocytic thyroiditis with interstitial fibrosis. Psammoma bodies are numerous in DSV. It occurs more often in young people, is virtually always accompanied by lymph node metastases, and often has pulmonary metastases.[63]
- *Solid variant*, which shows mainly or exclusively a solid growth pattern. It is usually seen in young people.[64]
- *Columnar cell variant* is a rare tumor composed of clear cells with marked nuclear stratification.[65]
- *Warthin-like papillary carcinoma* morphologically resembles a benign parotid lesion. This variant appears to have the same spectrum of behavior as classic PTC.[66]
- *Hobnail variant,* a rare and clinically aggressive variant of PTC that is characterized by a micropapillary growth pattern, similar to that encountered in tumors found at other biopsy sites (ovary, peritoneum, breast, bladder, kidney, and lung).[67-69]
- *Cribriform-morular variant.*[35,70]
- PTC with *Ewing family tumor elements.*[35,71]
- PTC with *ETV6-NTRK3 translocation.*[35,72]

Rarely PTC can show morphologic features indicative of aggressive clinical behavior. These include easily found mitotic figures, enlarged or hyperchromatic nuclei, abnormal DNA content, necrosis, and regions of nondescript neoplastic cells often growing in sheets (loss of differentiation).[50,73]

Follicular Carcinoma

Follicular carcinoma accounts for approximately 5% or less of all thyroid carcinomas in the United States. It is nearly always encapsulated and grossly resembles follicular adenoma.[1,27]

- It presents as a solid, encapsulated nodule with a well-developed capsule of variable thickness. In some cases tumor invasion beyond the capsule, and even venous invasion, can be seen grossly.
- On microscopic examination, follicular carcinomas most often have a microfollicular pattern and resemble a cellular follicular adenoma. Mitotic figures can be seen.
- As per WHO. follicular carcinomas can be classified as (1) localized, *minimally invasive* and (2) *widely invasive* cancers.[27]
- *Minimally invasive follicular carcinoma* is recognized by its penetration into and through the capsule that surrounds it and by its limited invasion into vessels at its periphery (less than four vessels). A minimally invasive carcinoma can present as a metastatic lesion. Multiple sections of the periphery of the neoplasm may be necessary to find the invasion.[74]
- *Widely invasive follicular carcinoma* extensively infiltrates the thyroid parenchyma and vascular channels (four or more vessels) within the gland or even in the perithyroidal tissues.[27]
- Follicular carcinoma has little tendency to invade lymphatic vessels and to spread to lymph nodes. Hematogenous spread to the skeleton, lungs, brain, liver, and other tissues occurs.
- Capsular “invasion” and “penetration” may be mimicked by post-FNA effect.
- FNA cannot distinguish between a follicular adenoma and carcinoma based on cytomorphologic features. The specimens are usually cellular and show a monotonous population of enlarged follicular cells arranged in three-dimensional groups with nuclear overlapping and crowding and microfollicles. Similar features can also be encountered in aspirates representing cellular adenomatoid nodules.[75]

KEY POINTS

- The minimally invasive follicular carcinoma is recognized by its penetration into and through the capsule that surrounds it and by its limited invasion into vessels at its periphery (less than four vessels).

Hürthle Cell (Oncocytic Follicular) Tumors

Hürthle cell tumors can be divided into benign and malignant categories by careful adherence to strict pathologic criteria defined for follicular carcinoma. More importantly, these pathologic distinctions predict clinical behavior.[3,27,76]

- Most oncocytic (Hürthle cell) neoplasms behave as follicular carcinomas; that is, pathologically the capsule or vessels should be assessed for invasion.[3]
- Oncocytic follicular (Hürthle cell) carcinomas should be considered a separate category of thyroid neoplasms from true follicular cancers. They can metastasize to regional lymph nodes as well as spread hematogenously; in addition, histologic evidence of invasive characteristics is found more commonly in oncocytic cancers than in nononcocytic follicular tumors.[50,77]

- FNA specimens of Hürthle cell neoplasms contain numerous Hürthle cells arranged in sheets and as single cells. The cells may demonstrate random nuclear pleomorphism; however, this feature is not helpful in differentiating between adenoma and carcinoma. The background shows mostly thick colloid, which on occasion maintains the round shape of the follicular lumen.[3]

KEY POINTS

- Hürthle cell (oncocytic follicular) carcinomas should be considered a separate category of thyroid neoplasms from true follicular cancers. They can metastasize to regional lymph nodes and spread hematogenously.

High-Grade Thyroid Carcinoma

The pathologic term "high-grade thyroid carcinoma" has been proposed (not endorsed by the current WHO classification of thyroid tumors) by expert thyroid pathologists to include malignant tumors of thyroid that show recognizable architectural and cytologic patterns of differentiated thyroid cancer, usually papillary, with high-grade pathologic features. These tumors often are large and show nuclear pleomorphism, foci of tumor necrosis and numerous mitotic figures, multifocal vascular invasion, and grossly identifiable extrathyroidal extension and are associated with regional and distant metastases.[50,78] It is important not to confuse these with poorly differentiated thyroid carcinomas (PDTCAs).

Poorly Differentiated Carcinoma

PDTCAs are a heterogeneous group of tumors.[27] They generally lack the usual histologic features and exceptional aggressiveness of anaplastic carcinomas.[79,80]

- The pathological features of PDTCA are solid/trabecular/insular growth, large size, frequent extrathyroidal extension, extensive vascular invasion, presence of necrosis, and increased mitotic activity (Fig. 70.10).[81]
- They may show elements of well-differentiated carcinoma, and less frequently anaplastic carcinoma.[82,83]
- Cytologic preparations representing PDTCA can be cellular, containing monotonous population of cells with nuclear overlapping and crowding with mitoses and single-cell necrosis (apoptosis). Some of these cases may resemble a secondary tumor of thyroid with neuroendocrine differentiation. Therefore, a thyroid origin should be confirmed by immunohistochemistry.[84,85]

Anaplastic Carcinoma

Fewer than 5% of thyroid carcinomas may be classified as anaplastic or undifferentiated. These usually occur in elderly people, more often women.[81,86]

- Most anaplastic carcinomas (50%–70%) will show remnants of well-differentiated follicular or papillary carcinoma or sometimes an adenoma or adenomatoid nodules, confirming the clinical impression that anaplastic carcinomas arise out of low-grade tumors.[81,87]
- Light microscopic examination shows varied histologic patterns, easily found mitotic figures, and regions of tumor necrosis.[88]
- The tumor cells show marked nuclear pleomorphism and can show squamous cell differentiation, sarcomatous appearance, bizarre giant cells, and, less commonly, giant cells resembling osteoclasts (Fig. 70.11).[88]

FNA specimens from anaplastic carcinoma can be readily classified as malignant because of readily identifiable features of poorly differentiated malignant neoplasm.[88]

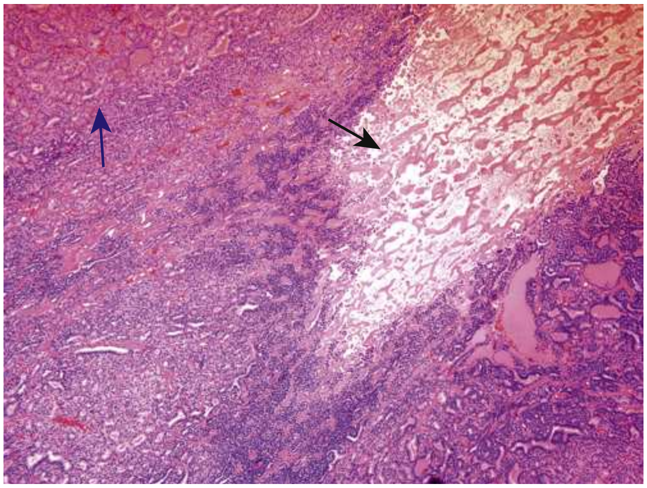


Fig. 70.10 Poorly differentiated carcinoma. Photomicrograph showing poorly differentiated carcinoma with solid growth pattern and foci of confluent tumor necrosis *(black arrow)*. Note the presence of elements of well-differentiated thyroid carcinoma *(blue arrow)* (hematoxylin and eosin stain, 20×).

Familial Follicular-Derived Thyroid Tumors

PTC and other follicular-derived thyroid tumors can be seen in other familial syndromes; these include familial adenomatous polyposis syndrome, PTEN hamartoma syndrome, McCune–Albright syndrome, Carney complex, and Peutz–Jeghers syndrome. Strong associations between *DICER1* mutations and differentiated thyroid carcinoma have also been reported.[72,81,89]

Medullary Carcinoma

Medullary carcinoma constitutes approximately 5% of thyroid carcinomas, originates from C-cells, may be sporadic or familial, and may be associated with disorders of other endocrine glands.[27,90]

- Medullary carcinomas on gross examination present as firm, white or yellow circumscribed nodules, some with infiltrative margins.[27]
- On light microscopic examination, the cells are rounded or polygonal or may be spindled. They are arranged as a diffuse solid mass, as islands separated by fibrous tissue (usually dense or hyalinized), as trabeculae or ribbons of cells, and (uncommonly) as glandular structures.[27]
- The tumor cell nuclei are rounded or elongated; occasional nuclei are large and irregular with "stippled" chromatin, also known as a salt and pepper pattern.[27] Amyloid is frequently present, both in the primary neoplasms and in the metastatic foci (~75% of cases).
- Both tumor stroma and the amyloid may undergo calcification (Fig. 70.12).[27]
- Medullary carcinomas nearly always produce calcitonin, although a few may lack this peptide.
- The tumor can invade both lymphatic and blood vessels; cervical nodes metastases are common.[91,92]
- The familial medullary carcinomas are usually bilateral and multicentric. Other family members may have C-cell hyperplasia and medullary carcinomas of microscopic size, some of which may have already spread to lymph nodes. In this situation, the so-called C-cell hyperplasia is premalignant.[1,27,93]
- A few neoplasms have been reported as mixed medullary and follicular carcinoma, which appears to represent joint C-cell and follicular cell proliferations, but these are rare.[1,27]
- FNA specimens of medullary carcinoma can show different morphologic patterns. A majority show round to oval tumor cells

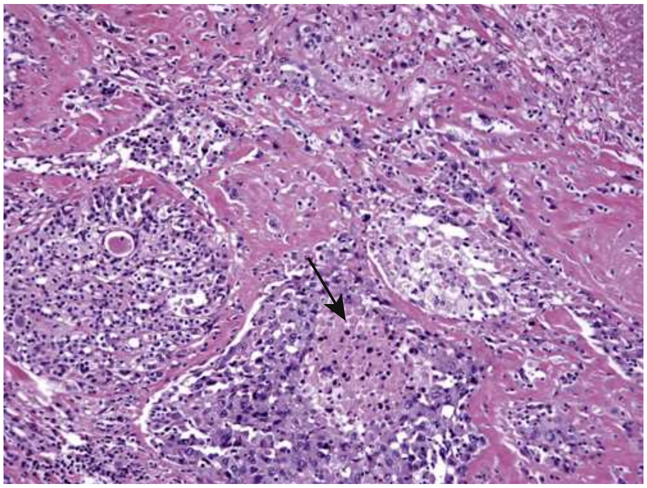


Fig. 70.11 Anaplastic carcinoma. Photomicrograph showing an undifferentiated malignant neoplasm with foci of tumor necrosis *(arrow)* (hematoxylin and eosin stain, 20×).

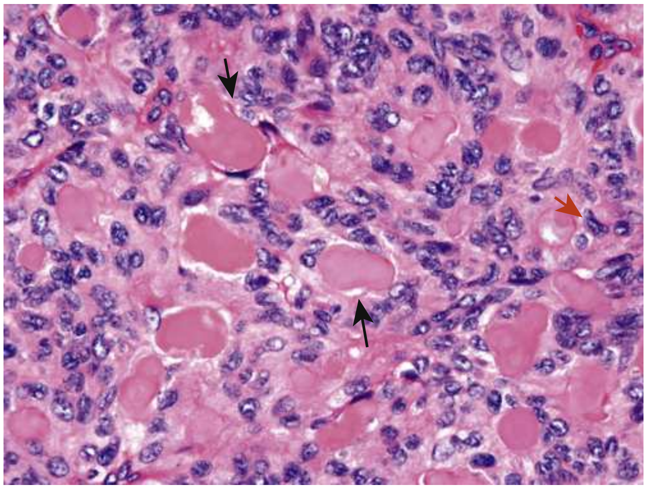


Fig. 70.12 Medullary carcinoma. Photomicrograph showing epithelioid and spindle-shaped tumor cells with stippled chromatin pattern *(red arrow)* and amyloid appearing an acellular eosinophilic deposits *(black arrows)* (hematoxylin and eosin stain, 20×).

arranged as single cells and loosely cohesive tissue fragments. The nuclear chromatin of the individual tumor cells is similar to that seen in neuroendocrine tumors, that is, stippled chromatin. Tumor cells of medullary carcinoma can either spindle-shaped or plasmacytoid in appearance (nucleus located at the edge of cell as seen in plasma cells) or contain oncocytic cytoplasm. Amyloid can be seen in the background or closely associated with tumor cells.[26,94] The diagnosis of medullary thyroid carcinoma is confirmed by performing immunostaining for either calcitonin and thyroglobulin.[26] An aliquot of the FNA specimen can also be submitted for calcitonin level determination.[95]

TUMORS WITH THYMIC OR RELATED BRANCHIAL POUCH DIFFERENTIATION

This category of primary thyroid tumors includes lesions that by morphology, immunohistochemistry, and electron microscopy show thymic or related branchial pouch differentiation. The list includes *intrathyroidal thymoma, spindle epithelial tumor with thymus-like differentiation,* and *carcinoma showing thymus-like differentiation.*[35]

RARE TUMORS OF THYROID GLAND

This list includes the following.

- *Mucoepidermoid carcinoma*, which microscopically resembles mucoepidermoid carcinoma of the salivary glands.[96]
- *Sclerosing mucoepidermoid carcinoma with eosinophilia.* This tumor usually originates in a thyroid gland with chronic lymphocytic thyroiditis and shows both squamous and glandular differentiation in a background of fibrotic stroma and inflammatory infiltrate rich in eosinophils.[97]
- *Mesenchymal tumors.* These tumors arise from the nonepithelial components of the thyroid gland and include leiomyoma, solitary fibrous tumor, angiosarcoma, and leiomyosarcoma.[35]
- *Lymphoma.* Primary lymphoma of the thyroid gland is rare, and most arise in the background of chronic lymphocytic thyroiditis. Most primary thyroid lymphomas are diffuse and are of B-cell type. Secondary involvement of thyroid gland by lymphoma can occur in 20% of patients with generalized lymphoma.[98,99]

Malignant Tumors Arising in Ectopic Thyroid Tissue

These include tumors arising in the thyroglossal duct tract, sublingual and lingual ectopic thyroid tissue, and struma ovarii. Papillary carcinoma of the classic type is the most common tumor arising in the thyroglossal duct tract. Because of this, carcinoma arising in this location can be easily diagnosed on FNA.[11,100,101]

Secondary Tumors

The most common tumors that metastasize to the thyroid are carcinomas of the breast, lung, colon, and kidney. These can also clinically mimic a primary thyroid neoplasm. The pathology diagnosis (FNA and histopathology) is based upon morphologic similarities between the primary and metastatic tumor and confirmation by appropriate immunohistochemical stain, such as thyroglobulin. In rare instances the secondary tumors can involve a preexisting primary thyroid neoplasm.[102,103]

SUMMARY AND FUTURE DIRECTIONS

This chapter has reviewed and illustrated the key histopathologic and cytological features of nonneoplastic and neoplastic lesions of the thyroid gland. Advances in cytologic diagnosis and classification of specific morphologic features has allowed FNA cytology of thyroid nodules to become part of the clinical approach to the treatment and follow-up of such nodules. Discussion has also included newly described entities such as NIFTP and its separation from the invasive follicular variant of papillary carcinoma, which continues to evolve with long term follow-up studies showing its extremely low risk prognosis.

REFERENCES

1. Rosai J, DeLellis RA, Carcangui ML, et al. *Tumors of The Thyroid Gland.* Vol 4. Washington, DC: Armed Forces Institute of Pathology; 2014.
2. Kameda Y. Cellular and molecular events on the development of mammalian thyroid C cells. *Dev Dyn.* 2016;245:323–341.
3. Jalaly JB, Baloch ZW. Hürthle-cell neoplasms of the thyroid: an algorithmic approach to pathologic diagnosis in light of molecular advances. *Semin Diagn Pathol.* 2020;37:234–242.
4. Pusztaszeri M, Rossi ED, Auger M, et al. The Bethesda system for reporting thyroid cytopathology: proposed modifications and updates for the second edition from an international panel. *Acta Cytol.* 2016;60:399–405.

5. Cibas ES, Ali SZ. The 2017 BeKthesda system for reporting thyroid cytopathology. *Thyroid*. 2017;27:1341–1346.
6. Menon SS, Tandon P, Ramaswamy B, et al. A retrospective study to assess the role of using fine needle aspiration cytology and frozen section in the diagnosis of thyroid swelling. *Indian J Otolaryngol Head Neck Surg*. 2018;70:471–476.
7. Siadati S, Rabiee SM, Alijanpour E, et al. The diagnostic value of fine needle aspiration in comparison with frozen section in thyroid nodules: a 20-year study. *Caspian J Intern Med*. 2017;8:301–304.
8. Bollig CA, Lesko D, Gilley D, et al. The futility of intraoperative frozen section in the evaluation of follicular thyroid lesions. *Laryngoscope*. 2018;128:1501–1505.
9. Seethala R, Asa S, Carty SE, et al. Protocol for the examination of specimens from patients with carcinomas of the thyroid gland. *CAP Cancer Protocols*. 2014;123:45–49.
10. Ross J, Manteghi A, Rethy K, et al. Thyroglossal duct cyst surgery: a ten-year single institution experience. *Int J Pediatr Otorhinolaryngol*. 2017;101:132–136.
11. Wei S, LiVolsi VA, Baloch ZW. Pathology of thyroglossal duct: an institutional experience. *Endocr Pathol*. 2015;26:75–79.
12. LiVolsi VA, Baloch ZW. The pathology of hyperthyroidism. *Front Endocrinol*. 2018;9.
13. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
14. Centeno BA, Szyfelbein WM, Daniels GH, et al. Fine needle aspiration biopsy of the thyroid gland in patients with prior Graves' disease treated with radioactive iodine. Morphologic findings and potential pitfalls. *Acta Cytol*. 1996;40:1189–1197.
15. Hang JF, Lilo MT, Bishop JA, et al. Diagnostic accuracy of fine needle aspiration in thyroid nodules arising in patients with graves disease. *Acta Cytol*. 2017;61:117–124.
16. Mizukami Y, Michigishi T, Nonomura A, et al. Pathology of chronic thyroiditis: a new clinically relevant classification. *Pathol Annu*. 1994;29:135–158.
17. Baloch Z, LiVolsi VA. Diagnostic dilemmas in thyroid pathology: follicular variant of papillary thyroid carcinoma and classic papillary thyroid carcinoma arising in lymphocytic thyroiditis. *Pathol Case Rev*. 2003;8: 47–56.
18. Canberk S, Griffin AC, Goyal A, et al. Oncocytic follicular nodules of the thyroid with or without chronic lymphocytic thyroiditis: an institutional experience. *Cytojournal*. 2013;10:2.
19. Lu L, Gu F, Dai WX, et al. Clinical and pathological features of Riedel's thyroiditis. *Chin Med Sci J*. 2010;25:129–134.
20. Baloch ZW, Feldman MD, LiVolsi VA. Combined Riedel's disease and Fibrosing Hashimoto's Thyroiditis: a report of three cases with two showing coexisting papillary carcinoma. *Endocr Pathol*. 2000;11:157–163.
21. Bastenie P, Neve P. Subacute and chronic granulomatous thyroiditis. In: Alexander P, Bacq ZM, eds. *Thyroiditis and Thyroid Function, Clinical, Morphological and Physicological Studies*. Oxford: Pergamon Press; 1972:69–97.
22. Carney J, Moore S, Northcutt R, et al. Palpation thyroiditis (multifocal granulomatous thyroiditis). *Am J Clin Pathol*. 1975;64:639–647.
23. Shah SS, Baum SG. Diagnosis and management of infectious thyroiditis. *Curr Infect Dis Rep*. 2000;2:147–153.
24. Welsh RA, Correa P. The comparative pathology of goiter in a nonendemic and an endemic area. *Arch Pathol*. 1960;69:694–700.
25. Correa P, Castro S. Survey of the pathology of thyroid glands from Cali, Colombia—a goiter area. *Lab Invest*. 1961;10:39–50.
26. The Bethesda System for Reporting Thyroid Cytopathology. *Definitions, Criteria, and Explanatory Notes*. 2nd ed. Switzerland; 2018.
27. Lloyd, RV, Osamura RY, Kloppel G, et al (eds). *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Vol 10. Lyon: WHO/IARC Press; 2017.
28. Harrer P, Brocker M, Zint A, et al. The clonality of nodules in recurrent goiters at second surgery. *Langenbecks Arch Surg*. 1998;383:453–455.
29. Khayyata S, Barroeta JE, LiVolsi VA, et al. Papillary hyperplastic nodule: pitfall in the cytopathologic diagnosis of papillary thyroid carcinoma. *Endocr Pract*. 2008;14:863–868.
30. Baloch Z, Mete O, Asa SL. Immunohistochemical biomarkers in thyroid pathology. *Endocr Pathol*. 2018;29:91–112.
31. Baloch ZW, LiVolsi VA. Post fine-needle aspiration histologic alterations of thyroid revisited. *Am J Clin Pathol*. 1999;112:311–316.
32. Sharma C, Krishnanand G. Histologic analysis and comparison of techniques in fine needle aspiration-induced alterations in thyroid. *Acta Cytol*. 2008;52:56–64.
33. Pandit AA, Phulpagar MD. Worrisome histologic alterations following fine needle aspiration of the thyroid. *Acta Cytol*. 2001;45:173–179.
34. Nikiforova MN, Nikitski AV, Panebianco F, et al. GLIS rearrangement is a genomic hallmark of hyalinizing trabecular tumor of the thyroid gland. *Thyroid*. 2019;29:161–173.
35. Cameselle-Teijeiro JM, Eloy C, Sobrinho-Simoes M. Pitfalls in challenging thyroid tumors: emphasis on differential diagnosis and ancillary biomarkers. *Endocr Pathol*. 2020;31:197–217.
36. Parisi MT, Eslamy H, Mankoff D. Management of differentiated thyroid cancer in children: focus on the American Thyroid Association Pediatric guidelines. *Semin Nucl Med*. 2016;46:147–164.
37. Baloch Z, LiVolsi VA. Fifty years of thyroid pathology: concepts and developments. *Hum Pathol*. 2020;95:46–54.
38. Rivera M, Tuttle RM, Patel S, et al. Encapsulated papillary thyroid carcinoma: a clinico-pathologic study of 106 cases with emphasis on its morphologic subtypes (histologic growth pattern). *Thyroid*. 2009;19:119–127.
39. Oertel YC, Oertel JE. Diagnosis of malignant epithelial thyroid lesions: fine needle aspiration and histopathologic correlation. *Ann Diagn Pathol*. 1998;2:377–400.
40. Baloch ZW, Livolsi VA. Follicular-patterned lesions of the thyroid: the bane of the pathologist. *Am J Clin Pathol*. 2002;117:143–150.
41. Hapke M, Dehner L. The optically clear nucleus: a reliable sign of papillary carcinoma of the thyroid? *Am J Surg Pathol*. 1979;3:31–38.
42. Vickery Jr AL. Thyroid papillary carcinoma. Pathological and philosophical controversies. *Am J Surg Pathol*. 1983;7:797–807.
43. Dugan J, Atkinson B, Avitable A, et al. Psammoma bodies in fine needle aspirate of the thyroid in lymphocytic thyroiditis. *Acta Cytol*. 1987;31:330–334.
44. Fiorella RM, Isley W, Miller LK, et al. Multinodular goiter of the thyroid mimicking malignancy: diagnostic pitfalls in fine-needle aspiration biopsy. *Diagn Cytopathol*. 1993;9:351–355;discussion 355–357.
45. Hunt JL, Barnes EL. Non-tumor-associated psammoma bodies in the thyroid. *Am J Clin Pathol*. 2003;119:90–94.
46. Kim H, Kim TH, Choe JH, et al. Patterns of initial recurrence in completely resected papillary thyroid carcinoma. *Thyroid*. 2017;27:908–914.
47. Landry CS, Grubbs EG, Busaidy NL, et al. Cystic lymph nodes in the lateral neck as indicators of metastatic papillary thyroid cancer. *Endocr Pract*. 2011;17:240–244.
48. Vuong HG, Kondo T, Duong UNP, et al. Prognostic impact of vascular invasion in differentiated thyroid carcinoma: a systematic review and meta-analysis. *Eur J Endocrinol*. 2017;177:207–216.
49. Kini SR. *Thyroid Cytopathology: An Atlas and Text*. 1st ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
50. Xu B, Ghossein R. Evolution of the histologic classification of thyroid neoplasms and its impact on clinical management. *Eur J Surg Oncol*. 2017;34:338–347.
51. Giordano TJ, Kuick R, Thomas DG, et al. Molecular classification of papillary thyroid carcinoma: distinct BRAF, RAS, and RET/PTC mutation-specific gene expression profiles discovered by DNA microarray analysis. *Oncogene*. 2005;24:6646–6656.
52. Asa SL, Giordano TJ, LiVolsi VA. Implications of the TCGA genomic characterization of papillary thyroid carcinoma for thyroid pathology: does follicular variant papillary thyroid carcinoma exist? *THYROID*. 2015;25:1–2.
53. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol*. 2016;2:1023–1029.

54. Xu B, Tallini G, Ghossein RA. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: historical context, diagnosis, and future challenges. *Endocr Pathol*. 2017;28:128–138.
55. Vinciguerra GL, Noccioli N, Bartolazzi A. Diffuse follicular variant of papillary thyroid carcinoma: a case report with a revision of literature. *Rare Tumors*. 2016;8:6536.
56. Albores-Saavedra J, Gould E, Vardaman C, et al. The macrofollicular variant of papillary thyroid carcinoma: a study of 17 cases. *Hum Pathol*. 1991;22:1195.
57. Fukushima M, Ito Y, Hirokawa M, et al. Macrofollicular variant of papillary thyroid carcinoma: its clinicopathological features and long-term prognosis. *Endocr J*. 2009;56:503–508.
58. LiVolsi VA. Papillary carcinoma tall cell variant (TCV): a review. *Endocr Pathol*. 2010;21:12–15.
59. Villar-Taibo R, Peteiro-Gonzalez D, Cabezas-Agricola JM, et al. Aggressiveness of the tall cell variant of papillary thyroid carcinoma is independent of the tumor size and patient age. *Oncol Lett*. 2017;13:3501–3507.
60. Liu Z, Zeng W, Chen T, et al. A comparison of the clinicopathological features and prognoses of the classical and the tall cell variant of papillary thyroid cancer: a meta-analysis. *Oncotarget*. 2017;8:6222–6232.
61. Bocklage T, DiTomasso J, Ramzy I, et al. Tall cell variant of papillary thyroid carcinoma: cytologic features and differential diagnostic considerations. *Diag Cytopathol*. 1997;17:25–29.
62. Solomon A, Gupta PK, LiVolsi VA, et al. Distinguishing tall cell variant of papillary thyroid carcinoma from usual variant of papillary thyroid carcinoma in cytologic specimens. *Diagn Cytopathol*. 2002;27:143–148.
63. Chereau N, Giudicelli X, Pattou F, et al. Diffuse sclerosing variant of papillary thyroid carcinoma is associated with aggressive histopathological features and a poor outcome: results of a large multicentric study. *J Clin Endocrinol Metab*. 2016;101:4603–4610.
64. Nikiforov YE, Erickson LA, Nikiforova MN, et al. Solid variant of papillary thyroid carcinoma: incidence, clinical- pathologic characteristics, molecular analysis, and biologic behavior. *Am J Surg Pathol*. 2001;25:1478–1484.
65. Bongiovanni M, Mermod M, Canberk S, et al. Columnar cell variant of papillary thyroid carcinoma: Cytomorphological characteristics of 11 cases with histological correlation and literature review. *Cancer Cytopathol*. 2017;125:389–397.
66. Hirokawa M, Nishihara E, Takada N, et al. Warthin-like papillary thyroid carcinoma with immunoglobulin G4-positive plasma cells possibly related to Hashimoto's thyroiditis. *Endocr J*. 2018;65:175–180.
67. Asioli S, Erickson LA, Sebo TJ, et al. Papillary thyroid carcinoma with prominent hobnail features: a new aggressive variant of moderately differentiated papillary carcinoma. A clinicopathologic, immunohistochemical, and molecular study of eight cases. *Am J Surg Pathol*. 2010;34:44–52.
68. Cameselle-Teijeiro JM, Rodriguez-Perez I, Celestino R, et al. Hobnail variant of papillary thyroid carcinoma: clinicopathologic and molecular evidence of progression to undifferentiated carcinoma in 2 cases. *Am J Surg Pathol*. 2017;41:854–860.
69. Teng L, Deng W, Lu J, et al. Hobnail variant of papillary thyroid carcinoma: molecular profiling and comparison to classical papillary thyroid carcinoma, poorly differentiated thyroid carcinoma and anaplastic thyroid carcinoma. *Oncotarget*. 2017;8:22023–22033.
70. Lam AK, Fridman M. Characteristics of cribriform morular variant of papillary thyroid carcinoma in post-Chernobyl affected region. *Hum Pathol*. 2018;74:170–177.
71. Eloy C, Oliveira M, Vieira J, et al. Carcinoma of the thyroid with ewing family tumor elements and favorable prognosis: report of a second case. *Int J Surg Pathol*. 2014;22:260–265.
72. Seethala RR, Chiosea SI, Liu CZ, et al. Clinical and morphologic features of ETV6-NTRK3 translocated papillary thyroid carcinoma in an adult population without radiation exposure. *Am J Surg Pathol*. 2017;41:446–457.
73. Shafique K, Baloch Z. Risk stratification of papillary thyroid carcinoma and its variants; from clinicopathologic features to molecular profiling. *Diagnostic Histopathology*. 2019;25:143–153.
74. Mete O, Asa SL. Pathological definition and clinical significance of vascular invasion in thyroid carcinomas of follicular epithelial derivation. *Mod Pathol*. 2011;24:1545–1552.
75. Lubitz CC, Faquin WC, Yang J, et al. Clinical and cytological features predictive of malignancy in thyroid follicular neoplasms. *Thyroid*. 2010;20:25–31.
76. Wenter V, Albert NL, Unterrainer M, et al. Clinical impact of follicular oncocytic (Hürthle cell) carcinoma in comparison with corresponding classical follicular thyroid carcinoma. *Eur J Nucl Med Mol Imaging*. 2021;48:449–460.
77. Zhou X, Zheng Z, Chen C, et al. Clinical characteristics and prognostic factors of Hürthle cell carcinoma: a population based study. *BMC Cancer*. 2020;20:407.
78. Xu B, Ibrahimpasic T, Wang L, et al. Clinicopathologic features of fatal non-anaplastic follicular cell-derived thyroid carcinomas. *Thyroid*. 2016;26:1588–1597.
79. Walczyk A, Kopczynski J, Gasior-Perczak D, et al. Poorly differentiated thyroid cancer in the context of the revised 2015 American Thyroid Association Guidelines and the updated American Joint Committee on Cancer/Tumor-Node-Metastasis Staging System (Eighth Edition). *Clin Endocrinol (Oxf)*. 2019;91:331–339.
80. Yu MG, Rivera J, Jimeno C. Poorly differentiated thyroid carcinoma: 10-year experience in a Southeast Asian population. *Endocrinol Metab (Seoul)*. 2017;32:288–295.
81. Romei C, Tacito A, Molinaro E, et al. Clinical, pathological and genetic features of anaplastic and poorly differentiated thyroid cancer: a single institute experience. *Oncol Lett*. 2018;15:9174–9182.
82. Volante M, Collini P, Nikiforov YE, et al. Poorly differentiated thyroid carcinoma: the Turin proposal for the use of uniform diagnostic criteria and an algorithmic diagnostic approach. *Am J Surg Pathol*. 2007;31:1256–1264.
83. Xu B, Ghossein R. Genomic landscape of poorly differentiated and anaplastic thyroid carcinoma. *Endocr Pathol*. 2016;27:205–212.
84. Tallini G, Gallo C. Fine-needle aspiration and intraoperative consultation in thyroid pathology: when and how? *Int J Surg Pathol*. 2011;19:141–144.
85. Sadow PM, Faquin WC. Poorly differentiated thyroid carcinoma: an incubating entity. *Front Endocrinol*. 2012;3:77.
86. Deeken-Draisey A, Yang GY, Gao J, et al. Anaplastic thyroid carcinoma: an epidemiologic, histologic, immunohistochemical, and molecular single-institution study. *Hum Pathol*. 2018;82:140–148.
87. Ranganath R, Shah MA, Shah AR. Anaplastic thyroid cancer. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:387–391.
88. Talbott I, Wakely Jr PE. Undifferentiated (anaplastic) thyroid carcinoma: practical immunohistochemistry and cytologic look-alikes. *Semin Diagn Pathol*. 2015;32:305–310.
89. Yoo SK, Lee S, Kim SJ, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet*. 2016;12:e1006239.
90. Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
91. Moo-Young TA, Traugott AL, Moley JF. Sporadic and familial medullary thyroid carcinoma: state of the art. *Surg Clin North Am*. 2009;89:1193–1204.
92. Brandao LG, Cavalheiro BG, Junqueira CR. Prognostic influence of clinical and pathological factors in medullary thyroid carcinoma: a study of 53 cases. *Clinics (Sao Paulo)*. 2009;64:849–856.
93. Essig Jr GF, Porter K, Schneider D, et al. Multifocality in sporadic medullary thyroid carcinoma: an international multicenter study. *Thyroid*. 2016;26:1563–1572.
94. Essig Jr GF, Porter K, Schneider D, et al. Fine needle aspiration and medullary thyroid carcinoma: the risk of inadequate preoperative evaluation and initial surgery when relying upon FNAB cytology alone. *Endocr Pract*. 2013;19:920–927.
95. Trimboli P, Guidobaldi L, Bongiovanni M, et al. Use of fine-needle aspirate calcitonin to detect medullary thyroid carcinoma: a systematic review. *Diagn Cytopathol*. 2016;44:45–51.
96. Baloch ZW, LiVolsi VA. Special types of thyroid carcinoma. *Histopathology*. 2018;72:40–52.
97. Shah AA, La Fortune K, Miller C, et al. Thyroid sclerosing mucoepidermoid carcinoma with eosinophilia: a clinicopathologic and molecular analysis of a distinct entity. *Mod Pathol*. 2017;30:329–339.

98. Zhou Y, Xu X, Xu H, et al. [Clinicopathological features and treatment outcome of 37 cases of primary thyroid lymphoma]. *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*. 2018;32:1551–1556.
99. Watanabe N, Narimatsu H, Noh JY, et al. Long-term outcomes of 107 cases of primary thyroid mucosa-associated lymphoid tissue lymphoma at a single medical institution in Japan. *J Clin Endocrinol Metab*. 2018;103:732–739.
100. Benameur Y, Ghfir I, Guerrouj H, et al. [Papillary carcinoma on an ectopic lingual thyroid]. *La Revue du Praticien*. 2019;69:65.
101. Wei S, Baloch ZW, LiVolsi VA. Pathology of struma ovarii: a report of 96 cases. *Endocr Pathol*. 2015;26:342–348.
102. HooKim K, Gaitor J, Lin O, et al. Secondary tumors involving the thyroid gland: a multi-institutional analysis of 28 cases diagnosed on fine-needle aspiration. *Diagn Cytopathol*. 2015;43:904–911.
103. Duskova J, Rosa P, Preucil P, et al. Secondary or second primary malignancy in the thyroid? metastatic tumors suggested clinically: a differential diagnostic task. *Acta Cytol*. 2014;58:262–268.

71

Graves Disease

Anthony P. Weetman and George J. Kahaly

OUTLINE

Historical Background, 1178
Etiology, 1179
  Genetic Factors, 1179
  Environmental Factors, 1180
  Existential Factors, 1181
Pathology, 1181
Pathogenesis, 1181
  The Role of Thyroid-Stimulating Hormone-Receptor Autoantibodies, 1181
  The Role of Cellular Immunity, 1182
Clinical Spectrum, 1184
Diagnosis, 1185
  Serology, 1185
  Imaging, 1185
Management, 1185
Antithyroid Drugs, 1185
  Adverse Events, 1187
  Pregnancy, 1187
  Long-Term Antithyroid Drug Therapy, 1188
  Children, Adolescents, and the Elderly, 1188
  Symptomatic Management, 1188
  Subclinical Graves Hyperthyroidism, 1188
Radioactive Iodine, 1188
Thyroidectomy, 1189
Perspective and Conclusion, 1189
Acknowledgment, 1190

Thyrotoxicosis is a biochemically defined state caused by an excess of circulating thyroid hormones: the term hyperthyroidism should only be used when the cause of thyrotoxicosis is excessive production of thyroid hormone by the thyroid gland. Graves disease (GD), accounts for 60% to 80% of cases of hyperthyroidism. Excessive levothyroxine dosage and destructive thyroiditis are the most common examples of thyrotoxicosis without hyperthyroidism.

GD is characterized by hyperthyroidism, a diffusely enlarged thyroid gland, and, in around one-quarter of patients, orbitopathy (GO). Pretibial myxedema (thyroid dermopathy) and thyroid acropachy are much less common extrathyroidal manifestations. Where present, these extrathyroidal features make the diagnosis possible clinically, but in many patients presenting only with biochemical evidence of thyrotoxicosis the diagnosis requires confirmation by further laboratory testing before a suitable therapeutic strategy can be chosen.

GD is caused by autoantibodies that mimic the action of thyroid-stimulating hormone (TSH) by binding to and activating the TSH receptor (TSH-R). As with most autoimmune disorders, a combination of many different genetic and environmental factors may contribute to the etiology. Family clustering and associated nonthyroidal autoimmune diseases are often found in GD patients. The cause of the extrathyroidal features is not yet entirely clear, which has hampered the development of optimal treatment for these complications, but recent developments have suggested new approaches.

HISTORICAL BACKGROUND

The earliest possible description of GD is in the *Thesaurus of the Shah of Khwarazm,* written around 1110 AD.[1] The first credited description of GD is that of Caleb Parry, who in 1825 described five patients with goiter and palpitations, including exophthalmos in one case.[2] Ten years later, Robert Graves identified three patients with tachycardia and goiter and a fourth who also had exophthalmos.[3] The other major name in the early history of GD is Carl Adolph von Basedow, who in 1840 described a syndrome comprising the "Merseburg triad" of exophthalmos, goiter, and palpitations; two of his patients also had pretibial myxedema.[4] Given this more comprehensive description, the condition was known initially as Basedow's disease, but the eponym honors finally went to Graves, whose textbook of medicine was popular.

Möbius was the first to identify that a thyroid disorder was central to GD, in 1886. Later work focused on the putative role of a thyrotropic pituitary factor with exophthalmic effects, but the appearance of GO in a patient with a total hypophysectomy finally demolished the pituitary theory of GO, although it remained reasonable to postulate that the thyroid overactivity in GD was secondary to pituitary overproduction of such a factor. However, by the 1950s, TSH had been purified, and it became obvious that excess TSH was not the cause of GD. Purves and Adams noted that TSH did not produce the typical lymphocytic infiltration of the thyroid, and therefore set about testing whether TSH really was elevated in GD by means of a bioassay that measured the *in vivo* release of radiolabelled thyroid hormone from the thyroid of guinea pigs. The serum from one patient with GD produced a release of thyroid hormone that was maximal at 24 hours, totally unlike the 3-hour peak effect of TSH; this new activity was therefore called long-acting thyroid stimulator (LATS).[5] By 1964, Kriss and colleagues had demonstrated that LATS was an IgG antibody, and the autoimmune basis of GD was established.

Initially, thyroidectomy (TX) was the only definitive treatment for GD, an operation rendered safer after the introduction of preoperative treatment with stable iodine in the 1920s. The discovery that thiocyanates had a goitrogenic effect led Astwood to develop antithyroid drugs

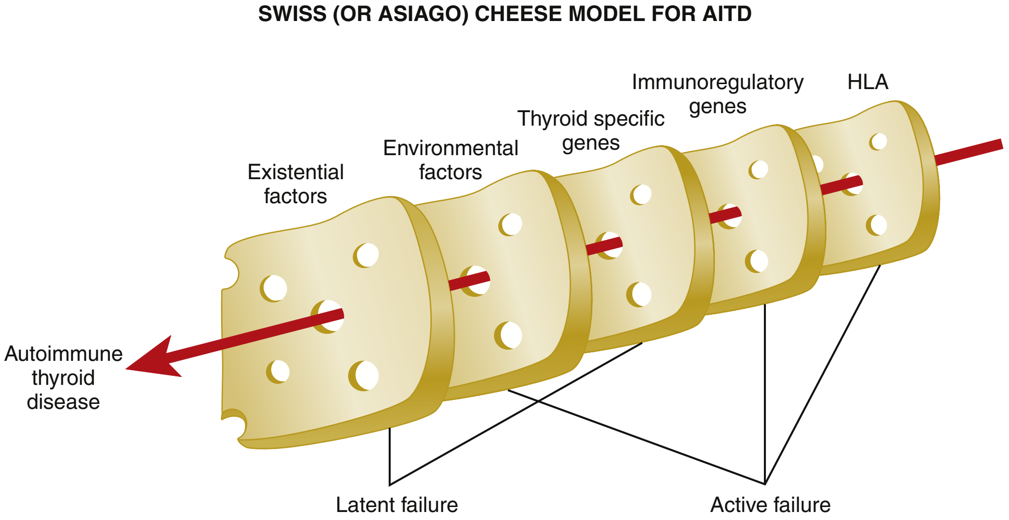


Fig. 71.1 **A model for the emergence of autoimmune thyroid disease.** This shows how the effect of multiple susceptibility factors may align to allow disease to occur in an individual, like the holes in the slices of Swiss cheese: each slice portrayed comprises many individual components, which must line up too. This is based on the well-known Swiss-cheese model for human accidents, which comprises active failures (e.g., pilot error) and latent failures (e.g., poor aircraft maintenance). *AITD*, autoimmune thyroid disease; *HLA,* human leukocyte antigen. (From Weetman AP. The immunopathogenesis of chronic autoimmune thyroiditis one century after Hashimoto. *Eur Thyroid J.* 2012;1:243–250; copyright the author.)

(ATDs), which were first prescribed in the 1940s. The introduction of radioactive iodine (RAI) treatment was based on the pioneering studies with iodine radioisotopes conducted at the Massachusetts Institute of Technology and the Harvard Medical School. Although the first GD patient so treated was in 1941, definitive publication was delayed until after the end of World War II.

ETIOLOGY

For GD to occur, a combination of genetic and environmental factors, working with a background of existential factors such as age, parity, and sex, have to interact in a favorable way to overcome the normal immunoregulatory mechanisms that prevent autoimmunity. This interaction can be thought of as analogous to the Swiss-cheese model for the multilayered causation of accidents (Fig. 71.1). Such a multifactorial model is typical for most autoimmune disorders. The sharing of similar genetic factors predisposing to autoimmunity helps to explain why around 20% of patients with GD have another organ-specific or non-organ-specific autoimmune disease.[6] Patients may also have a family history of various autoimmune problems besides those affecting the thyroid gland. The complexity involved, whereby the same disease may be the result of multiple combinations of susceptibility factors, helps explain why the phenotypic features of GD are diverse. As a simple example, smoking is associated with GD, but the association with GO is even stronger.

Genetic Factors

The role of heredity in GD was initially demonstrated by studies showing a higher frequency of autoimmune thyroid disease or thyroid antibodies in family members of patients with GD. A number of patterns of inheritance and candidate genes have been suggested, but ascertainment artifacts and the drawbacks of underpowered genetic association studies have produced many inconsistencies in results. Meticulous monozygotic twin studies have shown a concordance rate of only around 35% for GD, much lower than was previously thought.[7]

Attempts to understand which genes contribute to GD have typically used either a candidate gene approach, in which the frequency of polymorphisms in a possible candidate gene is determined in a cohort of patients and controls, or genome-wide association studies. The latter have the advantage of being hypothesis-free, thus potentially providing novel insights in etiology, but they demand huge numbers of patients to achieve sufficient statistical power. The susceptibility genes solidly identified so far in thyroid autoimmunity can be broadly classed as those that regulate the immune response and those that are thyroid-specific and thus increase target organ risk.

Immunoregulatory Genes. The first candidate genes to be studied, and the most important susceptibility genes that have been recognized so far, lie in the major histocompatibility complex (MHC), also known as the HLA (human leucocyte antigen), region in humans. Broadly speaking, these genes control the immune response at the level of antigen presentation. Class I HLA molecules, which present endogenous antigens (including those generated by intracellular viruses), are found on the surface of almost all cells and are involved in several immune responses, including cytotoxicity. Class II HLA molecules are expressed by antigen-presenting cells (APCs) and initiate antigen-specific stimulation of helper T-cells.

A consistent association has long been known to exist between the serologically-defined class II specificity HLA-DR3 and GD in Caucasians, with a relative risk of 2 to 3, but more detailed analysis has shown that the strongest association in this population is with the HLA class I allele HLA-C*07.[8] Molecular fine mapping has confirmed that polymorphisms in multiple HLA class I and class II genes independently contribute to disease risk, and this risk appears to be higher in Japanese patients than in European or other East Asian populations.[9] In Han Chinese, different HLA class I and II polymorphisms are associated with GD, but the majority of the risk is from three alleles encoding four amino acid changes, and these alleles may help determine the responsiveness of GD to treatment.[10] Together with many other studies, these results demonstrate a role for both the endogenous and exogenous antigen presentation pathways in determining susceptibility to

GD and show that their etiological role is heavily dependent on the ethnic background of the population, through interaction with other genetic or environmental factors.

The low relative risk conferred by HLA alleles obviously implies the existence of other susceptibility loci. Many candidates have been tested, including genes that encode immunoglobulins, T-cell receptors, and various cytokines, but the first robust association to be identified outside the HLA region was with two linked polymorphisms of the cytotoxic T-lymphocyte antigen-4 (*CTLA-4*) gene, which contribute to both GD and autoimmune hypothyroidism and confer a relative risk of around 2.[11] The CTLA-4 (or CD152) protein is expressed by T-cells and has a critical role in immunoregulation; engagement of CTLA-4 results in T-cell anergy, thus terminating immune and autoimmune responses. Additional polymorphisms in protein tyrosine phosphatase, nonreceptor type 22, *CD40*, and *CD25*, all of which are involved in T-cell regulation, contribute modestly to susceptibility to GD, as well as to many other autoimmune disorders.[11] Using a custom-made single-nucleotide polymorphism array of 186 loci known to be involved in susceptibility to other immune-mediated diseases, seven additional loci for both GD and Hashimoto thyroiditis (HT) have been identified.[12] The same loci have been associated with type 1 diabetes and rheumatoid arthritis, which have an increased prevalence in GD patients. Given the scale of the entire genome and the establishment of ever greater databases, the number of susceptibility loci involved in the etiology of GD looks set to expand much further.

The Thyroid-Stimulating Hormone-Receptor Gene. The genes considered previously are involved in the normal immune process and could exert their effects at the initiation of autoimmunity, through effects on central or peripheral tolerance, or they could vary the intensity of an established autoimmune process. None is specific for GD. In contrast, polymorphisms in the *TSH-R* gene, confirmed as a susceptibility locus for GD but not HT, represent the influence of a tissue-specific gene.[13] In families with an immunogenetic predisposition to developing thyroid autoimmunity, *TSH-R* polymorphisms may be important in determining which members will go on to develop GD rather than HT. It is unclear how these *TSH-R* polymorphisms predispose to GD, but this could be through (1) a less tolerogenic TSH-R being expressed in the thymus, leading to failure of central tolerance, or (2) coding for the increased release of soluble TSH-R fragments, which stimulate the autoimmune response (see later).

Environmental Factors

Iodine and Selenium. Several dietary factors have been implicated as risk factors for the development of both GD and HT. An increase in thyroid autoantibodies and HT has been observed in several populations after instituting iodination programs, and increased dietary iodine may exacerbate experimental disease in animal models of HT, suggesting that iodine has an immunological effect that enhances the thyroid autoimmune response in general.[14] Such an effect of iodine may be owing to altered immunogenicity of the thyroid autoantigen thyroglobulin, or excessive iodine may have a role in the intrathyroidal generation of toxic metabolites, leading to tissue damage. The specific situation in GD is complicated by the Jod–Basedow phenomenon, which is the stimulatory biochemical effect of iodine on thyroid hormone synthesis, especially in the presence of iodine deficiency. This phenomenon is demonstrated by the transient rise in both GD and toxic multinodular goiter after iodine supplementation.[15]

Selenium is found in all three deiodinase enzymes and in other proteins that protect thyroid cells from oxidative injury. Much attention has been paid recently to a possible role of selenium deficiency in autoimmune thyroid disease. So far, variable results have come from measuring selenium intake levels and conducting disease association studies, which in any case cannot directly show causation. There is also a lack of convincing data from selenium supplementation trials, which are heterogeneous and generally do not account for individual variation in dietary selenium intake. So far as GD specifically is concerned, a detailed analysis of existing trials has shown a possible beneficial effect of additional selenium in the rate at which ATD treatment restores euthyroidism, although any long-term effect remains to be demonstrated, and a single randomized control trial showed that selenium may have clinical benefit in mild GO.[16]

Infection. No convincing evidence has indicated a role for infection in GD. An association has been proposed between *Yersinia* infection and GD, and *Yersinia* contains proteins that mimic TSH-R immunologically at the level of TSH-R antibody (TSH-R–Ab) binding. The importance of any such effect seems minimal at best, given the relative frequencies of GD and *Yersinia* infection and the fact that molecular mimicry as a cause of autoimmunity is far more likely to operate at the level of T-cell rather than B-cell epitopes.[14] Hepatitis C is the only infectious agent that has been reliably implicated in the development of HT, but does not appear to be associated with GD. Indeed, epidemiologic studies suggest that inferior levels of prosperity and hygiene, with a consequently greater risk of infections in childhood, may actually protect against the development of thyroid autoimmunity.[17] Such an effect may be mediated by the response to infections skewing the immune system away from proautoimmune T-cell responses during development, the so-called hygiene hypothesis.

Stress. A number of retrospective studies have identified stress during the year before the initial presentation as an important risk factor for GD, but all such surveys suffer from potential flaws in their dependence on recall, as well as other sources of bias, and prospective studies are missing, owing to their difficulty.[18] Any effect of stress presumably results from the neuroendocrine responses to it that alter the regulation of autoreactive lymphocytes. Attacks of allergic rhinitis and elevated serum IgE concentrations (denoting an atopic diathesis) are associated with increased risk of relapse in those with treated GD, perhaps because of the pattern of cytokines produced by the immune system during an allergic episode also exacerbates autoimmunity.[19]

Smoking and Alcohol. Smoking is associated with increased susceptibility to GD, especially in women, with an overall relative risk of around 2: the risk is dose-dependent and disappears if smoking is stopped for several years.[20] The development of GO is even more strongly associated with smoking. On the other hand, alcohol consumption is a dose-dependent protective factor, an effect that is independent of the type of alcohol consumed, sex, iodine intake, and smoking behavior.[21] The reason for these effects is unknown, although smoking can alter normal thyroid function, and alcohol is protective in other autoimmune disease like rheumatoid arthritis.

Drugs. Exogenous cytokines given in massive doses therapeutically, particularly interferon (IFN)-α, exacerbate preexisting thyroid autoimmunity and lead in predisposed individuals to the development of HT, destructive thyroiditis causing transient thyrotoxicosis, and, much less commonly, GD.[22] Other immunologically active drugs can also cause GD. Alemtuzumab is a lymphocyte-depleting monoclonal antibody used to treat multiple sclerosis, and autoantibody-mediated diseases including GD may develop during the phase of immune reconstitution that follows treatment (see later).[23] Immune checkpoint inhibitors are increasingly being used as anticancer treatments. These drugs include CTLA-4 blockers (such as ipilimumab) and programmed death protein 1 blockers (such as pembrolizumab and nivolumab), which enhance T-cell–mediated immunity against tumors and, as a side effect, can induce HT or, rarely, GD.[22] Distinguishing the precise cause of thyrotoxicosis (destructive thyroiditis or GD) in all of these drug-related settings requires care to ensure optimal treatment.

Existential Factors

By far the biggest risk factor for developing GD is being female (as with most autoimmune diseases), with a 5- to 10-fold higher frequency in women. In contradistinction to HT, GD is more common in Black and Asian/Pacific Island people than in White people.[24] Age is also important, as the disease is rare in children, peaking in incidence between the ages of 40 and 50 years. There is a well-established pattern of postpartum occurrence of GD or exacerbation of preexisting GD in the year after delivery, akin to the pattern seen in postpartum thyroiditis, and women below 40 years old who have had live pregnancies or induced abortions are at increased risk of GD in the years after the postpartum period.[25] These effects may be related to reconstitution of the immune system after the alterations in immunoregulation during pregnancy, which ensure maternal tolerance to the fetal semiallograft. Microchimerism, caused by the transfer of fetal cells bearing paternal antigens at delivery and leading to the exacerbation of any ongoing autoimmune response, has also been implicated.

KEY POINTS

- The biggest risk factor for developing Graves disease is being female, with a 5- to 10-fold higher frequency in women.

PATHOLOGY

The thyroid is typically diffusely enlarged, with an increase in vascularity. Untreated GD is practically unseen histologically nowadays, as patients are rendered euthyroid before surgery to ensure their safety, and the typical features are attenuated by ATDs. The thyroid follicles are both hypertrophied and hyperplastic, with columnar epithelium that is often folded into the follicular lumen. New small follicles are formed containing scant colloid. There are varying degrees of lymphocytic and plasma cell infiltration, but, unlike HT, the generation of germinal centers is uncommon. Rarely lymphoid hyperplasia may also occur in the lymph nodes, thymus, and spleen; as with the thyroid lymphocytic infiltrate, these manifestations reverse with ATD treatment. Cases of unnecessary thymectomy have resulted from ignorance of these features.

PATHOGENESIS

GD results from the action of thyroid-stimulating antibodies (TSAbs) that bind to and activate the TSH-R. The primary role of TSAbs is most clearly shown by the occurrence of neonatal GD in the newborn infants of mothers who have a high circulating concentration of these antibodies; the infants' disease remits as the level of placentally transmitted maternal antibody declines. TSH-mediated activation of the TSH-R is characterized by a biphasic regulatory mechanism, which prevents overstimulation of the thyroid, whereas TSAb binding appear to show a monophasic effect, thus allowing chronic activation.[26] However, the degree of hyperthyroidism does not exactly correlate with the measured serum level of TSAb; some patients fluctuate between hyperthyroidism and hypothyroidism because of the simultaneous presence of TSH-R–Ab that block the receptor, and in around 15% of patients treated with ATDs, permanent hypothyroidism supervenes years later, indicating the concurrence of the same kind of tissue-destructive mechanisms that cause autoimmune hypothyroidism. Besides causing hyperthyroidism, TSAb also mediate goitrogenesis and increase the vascularity of the Graves thyroid by enhancing local expression of vascular endothelial growth factor and its receptor.

KEY POINTS

- Graves disease is caused by stimulatory autoantibodies that mimic the action of thyroid-stimulating hormone (TSH) by binding to and activating the autoantigen TSH receptor.

The Role of Thyroid-Stimulating Hormone-Receptor Autoantibodies

Thyroid-Stimulating Hormone-Receptor as an Autoantigen. The TSH-R is a member of the G protein-coupled receptor family; activation of the receptor by TSH or that subset of TSH-R–Ab with stimulatory activity leads to intracellular signaling via the cyclic adenosine monophosphate pathway, although other signaling pathways are activated at high ligand concentrations and by nonstimulating TSH-R–Ab.[27] The receptor is a 744–amino acid glycoprotein with an apparent molecular weight of 95 to 120 kD, depending on glycosylation. It has an extracellular domain, the A subunit, with leucine-rich repeat domains and a hinge region to which TSH and TSH-R–Ab bind, linked via a C peptide region to the B subunit, which comprises a looped transmembrane domain and an intracellular signaling domain[28] Uniquely for a G protein-coupled receptor, the TSH-R is cleaved during posttranslational processing at sites in the hinge region; further reduction of disulfide bonds leads to shedding of the A subunit from the thyroid cell surface. The soluble A subunit that is formed binds TSAb more strongly than the holoreceptor, and this is an important factor in the induction and affinity maturation of TSAb.[28] The enzyme responsible for TSH-R cleavage has not been identified but could have a critical pathological role in GD through regulation of autoantigen production. The extrathyroidal expression of TSH-R, particularly in the orbit, may serve to provide a cross-reactive autoantigen crucial to the complications of GD.

TSAb are high affinity and usually show restriction to the κ light-chain and the IgG1 subclass, implying that they originate from only a small number of B-cell clones. In turn, this indicates that clinical disease manifests itself quickly once the autoimmune process begins in GD, in contrast to the much slower progression of autoimmune hypothyroidism. B-cell epitopes on the TSH-R are generally conformational and discontinuous. The exception is a set of linear hinge region epitopes recognized by "neutral" TSH-R–Ab (Table 71.1), which occurs in some patients with GD. These antibodies may not actually be pathologically neutral, as they can induce thyroid cell apoptosis and alter posttranslational modification of the TSH-R *in vitro*.[27] Patients with GD may also have TSH-R–Abs that can block rather than stimulate the receptor, albeit at a level where the stimulating effect of TSAb supervenes; blocking TSH-R–Abs are more commonly found in HT. The characterization of B-cell epitopes made possible with the development of human monoclonal TSH-R–Abs has shown a greater overall similarity of binding sites for blocking and stimulatory antibodies than had previously been suspected, but there is still considerable heterogeneity in the detailed pattern of epitope recognition by TSAb, which is likely to determine the ultimate severity of clinical presentation. The antibody response to the TSH-R in GD is therefore complex and differs both between patients and in the same patient over time. In some patients, fluctuation in the relative proportion of the two functional types of TSH-R–Abs may produce a confusing clinical picture

TABLE 71.1 Classification of Main Thyrotropin Receptor Antibodies

Antibody	Assay
Long-acting thyroid stimulator	Original description of TSH-R antibodies; used bioassay in whole mouse to assess effects of stimulator antibodies on radioiodine release
TSAb	Bioassay; usually measurement of cAMP production by primary cultures of thyroid cells, thyroid cell lines, (e.g., FRTL-5), or TSH-R-transfected eukaryotic cells
Thyroid-blocking antibodies	Bioassay; measurement of inhibition of cAMP production after TSH-mediated stimulation of the TSH-R; may operate at the level of TSH binding or receptor signaling
TSH binding-inhibiting immunoglobulins (TBII)	Measurement of inhibition by patient antibodies of labeled TSH or labeled monoclonal TSH-R antibody binding to solubilized or recombinant TSH-R; unable to define functional activity
Neutral TSH-R antibodies	Detected by assays with TSH-R hinge region linear peptides; may not be TBII-positive, and no demonstrable effect on cAMP production

cAMP, Cyclic adenosine monophosphate; *FRTL*-5, Fischer rat thyroid line; *TSAb*, thyroid-stimulating antibodies; *TSH*, thyroid-stimulating hormone; *TSH-R*, thyroid-stimulating hormone receptor.

of alternating hyperthyroidism and hypothyroidism, and such fluctuation also may occur after pregnancy.[29]

Assays for Thyroid-Stimulating Hormone-Receptor–Autoantibodies. The nomenclature of TSH-R–Abs can be confusing because of the different names given to the various assays, which have been used in measurement; a summary of these assays is shown in Table 71.1. In essence, demonstrating the true functionality of any TSH-R–Ab requires the use of highly sensitive and specific bioassays that measure activation of the receptor *in vitro*. Such assays have been vital research tools but may be difficult to standardize and maintain in nonexperienced laboratories. Measurement of simple TSH-R binding by an antibody is available in all routine labs and is easier to perform using TSH-binding inhibitor immunoglobulin (TBII) assays; these are commercially available and serve well for clinical purposes, as the result can be interpreted in relation to the patient's presentation. For instance, biochemical thyrotoxicosis and a positive TBII result indicate that the antibodies are most likely TSAbs, and the patient has GD, whereas the same positive TBII result in a hypothyroid patient presenting with HT indicates that the TBII assay is most likely detecting blocking TSH-R–Abs.[29] Current TBII assays have sensitivities and specificities in excess of 95%.

Other Autoantibodies. Thyroglobulin and thyroid peroxidase antibodies occur up to 80% of GD patients. Up to one-third of GD sera contain antibodies that inhibit Na^+/I^- symporter mediated iodide uptake *in vitro*. Antibodies to pendrin, an apical protein responsible for mediating iodide efflux into the follicular lumen, occur in a smaller percentage of patients. All these autoantibodies are also found in autoimmune hypothyroidism and mark the presence of an underlying thyroiditis accompanying the development of TSAbs in GD.

The Role of Cellular Immunity

Self-Tolerance and Autoimmunity. Although GD is an autoantibody-mediated disorder, B-cells only produce antibodies after differentiation into plasma cells, a process that is regulated by cytokines and CD40/CD40 ligand interaction with an activated helper $CD4^+$ T-cell. Activated $CD4^+$ T-cells can follow two broad pathways of function, depending on their pattern of cytokine secretion. Type 1 helper T (Th1) cells mediate delayed-type hypersensitivity responses, essential to the kind of destructive process seen in most organ-specific endocrinopathies, whereas type 2 helper T (Th2) cells promote antibody production. Although this Th1/Th2 paradigm has been useful in murine immunology, many helper T-cells cannot be so neatly classified in humans.

TABLE 71.2 Mechanisms for Maintaining Nonresponsiveness of Autoreactive T-Cells to Target Tissues

Mechanism	Reversibility
Thymic Presentation of Self Antigens (Central Tolerance)	
Deletion	No
Anergy	Yes
Peripheral Tolerance	
Deletion	No
Anergy	Yes
Clonal ignorance	Yes
Immunologic privilege at the target site (e.g., cornea)	Yes
Active Suppression	
Cytokine network	Yes
T regulatory (Treg) cells	Yes

The immune system exists to eliminate foreign antigens but must remain tolerant of autoantigens. Self-tolerance depends critically on the absence of effective autoreactive T-cells. During development, the thymus is mainly responsible for eliminating autoreactive T-cells (clonal deletion) and for positively selecting appropriate T-cells to constitute the immune repertoire.[30] These processes depend on the presentation of endogenous peptides by thymic APCs to the naïve T-cell repertoire. Rare mutations in the autoimmune regulator gene that prevent such presentation of self peptides within the thymus give rise to autoimmune polyglandular syndrome type 1, which may include HT amongst its more usual manifestations. GD is unusual in this syndrome, suggesting additional pathways for intrathymic TSH-R presentation.

We now know, particularly from experiments with transgenic mice, that individual mechanisms for ensuring self-tolerance are never completely successful. In particular, the thymus only manages to prune, rather than eliminate, the autoreactive T-cell repertoire, and so such T-cells are present in all healthy individuals, although held in various states of nonreactivity to self by a number of checkpoints (Table 71.2). Tolerance to a self antigen can be broken at one or more of these controlling points, resulting in autoimmune disease if the response is sufficiently vigorous, and each point may be influenced by both genetic and environmental factors. The following have all been implicated in autoimmunity:

1. An excessive number of autoreactive T-cells may not be deleted or rendered anergic in the thymus (e.g., through inheritance of particular genes).

2. T-cells that escape thymic tolerance may fail to be deleted or rendered anergic in the periphery (e.g., because of abnormal provision of costimulatory signals; see later).
3. Failure of immunologic privilege may occur (e.g., in the cornea).
4. Molecular mimicry may occur with cross-reactive exogenous antigens, inducing a response against a normally "silent" autoantigen (e.g., myocarditis after streptococcal infection).
5. T regulatory (Treg) cells may not function adequately (see later).

There is no clear evidence for failure of immune privilege or the operation of molecular mimicry in GD, but the other mechanisms play a role, although the degree to which each is involved is unclear. Some of the genes associated with GD described earlier may alter intrathymic deletion, especially through thymic TSH-R expression, and the role of Treg cells is considered later. Regarding peripheral tolerance (point 2), only APCs normally express class II HLA molecules, but thyroid cells express class II molecules when stimulated by IFN-γ secreted by any thyroidal T-cells, for instance during an episode of viral thyroiditis. Unlike APCs, however, thyroid cells do not express the necessary costimulatory signals to activate naïve T-cells, and, in this setting, such T-cells are instead rendered anergic.[31] The situation differs when a T-cell has been previously activated: a costimulatory signal is then no longer required, and T-cell stimulation will occur. Thus, HLA class II expression by thyroid cells may be a means of ensuring peripheral T-cell tolerance under normal circumstances, but such expression may exacerbate already established thyroid autoimmunity (Fig. 71.2).

Insights From Animal Models. It is striking that, in contrast to HT, no spontaneous animal model of GD has been found. Such restriction may be because (1) the necessary predisposing genetic and environmental factors are exclusively human, or (2) the TSH-R is uniquely autoantigenic in humans. Attempts to create an induced animal model of GD with TSH-R immunization have also been frustrating, with initial experiments resulting in thyroiditis and blocking TSH-R–Abs. Hyperthyroidism and stimulating TSH-R–Ab can be induced in rodents by immunization with fibroblasts expressing MHC class II molecules and the human TSH-R, or with plasmid or adenovirus vectors encoding the holoreceptor or the A subunit.[32] NOD.H2h4 mice have spontaneous background HT, and using this strain to create transgenic animals with intrathyroidal expression of the human TSH-R A subunit results in the spontaneous development of human-specific TSAbs; similar expression of the mouse TSH-R A subunit does not result in TSH-R–Abs.[33] These results indicate that the human TSH-R A subunit may indeed be uniquely autoantigenic, but also show the importance of the necessary genetic background to initiate disease. This animal model has also provided evidence that the level of intrathymic expression of the TSH-R A subunit is critical in determining whether autoimmunity or central tolerance is induced.[34]

T-Cell Responses. Examining TSH-R–specific T-cell function in GD is currently prevented by the low number of such cells, their predominantly intrathyroidal location, and the distorting effects of ATDs on the autoimmune response. Many studies have used the circulating T-cell population, which may have little or no relevance to what is happening within the thyroid lymphocytic infiltrate. Various abnormalities in the phenotype of circulating T-cells have been described, such as a reduction in CD8$^+$ T-cells and an increase in activated HLA-DR$^+$ T-cells, but these are not specific to GD. Circulating T-cells respond to multiple T-cell epitopes on the TSH-R in GD. The intrathyroidal T-cells are a mix of CD4$^+$ and CD8$^+$ populations, the latter presumably contributing to the destructive thyroiditis that may eventually supervene. Attempts have also been made to understand the pattern of cytokines secreted in GD, with the overly simplified view that this is likely to be a Th2-mediated disorder. However, the predominance of the IgG1 subclass in TSAb indicates a robust Th1 type response, as does the pattern of chemokine expression, which is important in the active phase of disease.[35] No distinct Th1 or Th2 pattern of response of cytokines has been observed within the thyroid in GD, and some cytokines are derived from the thyroid cells themselves or from other immune system cells, including the proinflammatory Th17 T-cell subset.[36,37]

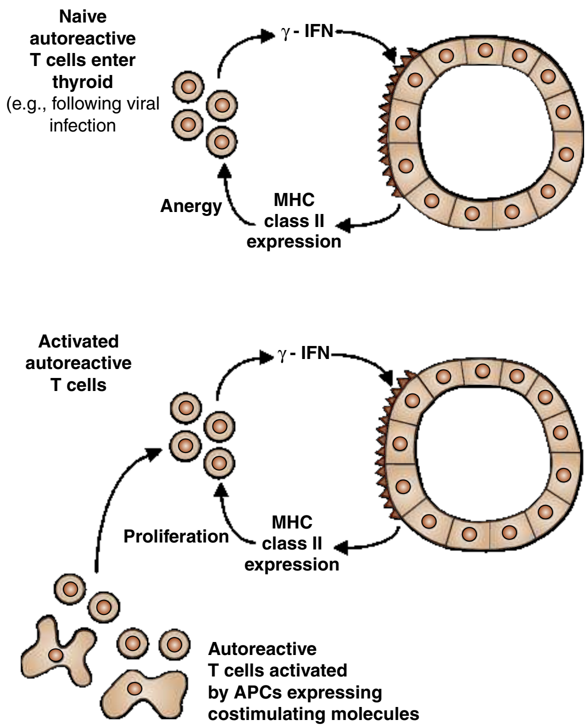

Naive autoreactive T cells enter thyroid (e.g., following viral infection
γ - IFN
Anergy
MHC class II expression
Activated autoreactive T cells
γ - IFN
Proliferation
MHC class II expression
Autoreactive T cells activated by APCs expressing costimulating molecules

Fig. 71.2 Alternative outcomes after major histocompatibility complex (*MHC*) class II expression by thyroid cells. Naïve T cells require costimulation for activation, and anergy can be induced by interaction with the MHC class II molecule/antigenic epitope alone (upper panel, peripheral tolerance). If T cells receive prior costimulation from classic antigen-presenting cells (*APCs*), class II expression by thyroid cells can enhance the T-cell response *(lower panel)*. *IFN*, Interferon. (From Jameson LJ, De Groot LJ, de Kretser DM, et al. *Endocinology: Adult and Pediatric,* 7th ed. Philadelphia: Elsevier; 2016.)

The levels of CD4$^+$FOXP3$^+$ "natural" Treg cells are either normal or lowered in the circulation of untreated patients with GD; abundant natural Treg cells can be detected in the thyroid infiltrate but seem to have lowered function.[37] Once activated, Treg cells exert their effects without antigen restriction, and so it is the intrathyroidal population that is most likely to be critical to pathogenesis. There is good evidence from animal models that systemic Treg cell depletion leads to the development of experimental HT and hypothyroidism, but the effects of such depletion on experimental GD are not known.[14] With the continuing description of new types of regulatory cells, which now encompass subpopulations of B-cells, CD8+ cells, and Th1, Th2, and Th17 cells with differing immunoregulatory capacities,[38] it is evident that the control of autoimmune T-cell responses in GD is at present poorly understood. MicroRNAs (noncoding small RNA molecules) regulate normal gene expression, including genes involved in the innate and adaptive immune response. Recently, dysregulation of specific microRNAs has been implicated in autoimmunity, and a signature profile of microRNAs within the thyroid has been identified that shows associations with the severity of GD, the response to treatment, and the occurrence of GO.[39] This area offers the potential of developing new insights into pathogenesis, as well as novel biomarkers and even therapies.

Immune Reconstitution and Graves Disease. GD (as well as HT and a spectrum of autoimmune disorders) has been reported in three particular settings of immune reconstitution after a period of induced lymphopenia.[23] First, bone marrow transplantation using a donor who has had GD may transfer the disease to the recipient. This is primarily the result of adoptive immunity transferring autoreactive lymphocytes, although there may be an additional role for disordered immunoregulation secondary to graft-versus-host disease. Second, alemtuzumab is a humanized monoclonal antibody against CD52 expressed by lymphocytes and macrophages that rapidly depletes T-cells and is mainly used for active relapsing-remitting multiple sclerosis. GD occurs in up to one-third of such patients during the phase of naïve T-cell expansion, which occurs around 18 months after therapeutic lymphocyte depletion. Fluctuating levels of TSAbs and blocking TSH-R–Abs occur in 20% of the Graves cases, and the disease may be quite resistant to treatment. Finally, GD occurs in up to 3% of patients with human immunodeficiency virus infection given highly active antiretroviral therapy, again during the phase of T-cell expansion. The mechanisms responsible for reconstitution GD are at present unclear but may include dysregulation of Treg function and the effects of interleukin-21, a pleiotropic cytokine that can promote autoimmune responses through its effects on Th17 and B-cell activation.[14]

CLINICAL SPECTRUM

Signs and symptoms include features that are common to any cause of thyrotoxicosis (Table 71.3), as well as those specific for GD.[40] The clinical presentation depends on the severity of thyrotoxicosis, the duration of the disease, individual susceptibility to excess thyroid hormone, and the age of the patient. In the elderly, features of thyrotoxicosis may be subtle or masked, and patients may present mainly with fatigue and weight loss, leading to apathetic hyperthyroidism. Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, and is because of the increased metabolic rate. Weight gain occurs in 5% to 10% of patients because of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatigability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly.[41]

In GD, the thyroid is diffusely enlarged to two to three times its normal size (Table 71.4). The consistency is firm, but less so than in multinodular goiter. There may be a thrill or bruit because of the increased vascularity of the gland and the hyperdynamic circulation. The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations and sometimes because of supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur, and can lead to worsening of angina or heart failure in the elderly or those with preexisting heart disease.[42] Atrial fibrillation is more common in patients older than 50 years. Treatment of the thyrotoxic state alone reverts atrial fibrillation to normal sinus rhythm in fewer than half of patients, suggesting the existence of an underlying cardiac problem in the remainder.

TABLE 71.3 Causes and Differential Diagnosis of Hyperthyroidism

Primary hyperthyroidism

- Graves disease
- Toxic multinodular goiter
- Toxic adenoma
- Amiodarone, iodine excess
- Ingestion of excess thyroid hormone (thyrotoxicosis factitia) or thyroid tissue
- Subacute thyroiditis
- Silent thyroiditis
- Activating mutation of the thyroid-stimulating hormone receptor (autosomal dominant)
- Struma ovarii
- Functioning thyroid carcinoma metastases

Secondary hyperthyroidism

- Thyroid-stimulating hormone–secreting pituitary adenoma
- Thyroid hormone resistance syndrome
- Chorionic gonadotropin–secreting tumors
- Gestational thyrotoxicosis

TABLE 71.4 Signs and Symptoms of Graves Hyperthyroidism

Symptoms

- Hyperactivity, irritability
- Heat intolerance and sweating
- Palpitations
- Dysphoria
- Fatigue and weakness
- Weight loss with increased appetite
- Diarrhea
- Polyuria
- Oligomenorrhea, loss of libido

Signs

- Tachycardia
- Atrial fibrillation in the elderly
- Tremor
- Goiter
- Warm, moist skin
- Muscle weakness, proximal myopathy
- Lid retraction or lag
- Exophthalmos
- Gynecomastia

The skin is usually warm and moist, and the patient complains of sweating and heat intolerance, particularly during warm weather. Palmar erythema, onycholysis, and ,less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and a diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Fine tremor is a very frequent finding, best elicited by asking patients to stretch out the fingers and feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is a rare feature. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men, there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in longstanding thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalcuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

KEY POINTS

- Graves disease is characterized by autoimmune-induced hyperthyroidism, a diffusely enlarged thyroid gland, and, in around 30%–40% of patients, orbitopathy.

DIAGNOSIS

Serology

Serum TSH measurement has the highest sensitivity and specificity of any single blood test used in the evaluation of suspected hyperthyroidism and is used as initial screening.[43-45] However, when hyperthyroidism is strongly suspected, diagnostic accuracy improves when both a serum TSH and free T4 are assessed at the time of the initial evaluation. In overt hyperthyroidism, both serum free T4 and T3 concentrations are elevated, and serum TSH is suppressed; however, in subclinical hyperthyroidism, serum total T4 and free T4 levels can be normal, with an undetectable or low serum TSH. Finally, measurement of serum T3 is helpful, as isolated T3 elevation ("T3-toxicosis") is observed in several patients with GD of recent onset. TSH-R–Abs are specific biomarkers for GD.[46-48] The likelihood that patients have GD is 1367- to 3,420-fold greater if they are TSH-R–Ab–positive than if they are TSH-R–Ab–negative.[46] Most immunoassays today use a competitive binding assay and measure TBIIs. Binding assays only report the presence or absence of TSH-R–Abs and their concentrations, rather than their functional activity.[49] A metaanalysis of 21 studies showed that the overall pooled sensitivity and specificity of the serum TSH-R–Ab concentration measured with binding assays were 98 and 99%, respectively.[50] In contrast, the highly sensitive cell-based bioassays[1-58] exclusively differentiate between the stimulating (TSAb) and TSH-R–blocking antibodies.[59-61] In addition, TSAb is a highly sensitive and predictive biomarker for the extrathyroidal manifestations of GD,[62-68] as well as a useful predictive measure of fetal or neonatal hyperthyroidism.[69,70] Further, the incorporation and early utilization of TSAb into diagnostic algorithms conferred a 46% shortened time to diagnosis of GD and a cost saving of 47%.[71] In summary, the measurement of TSH-R–Ab is a sensitive and specific tool for rapid and accurate diagnosis and differential diagnosis of Graves hyperthyroidism. When technically available, differentiation of TSH-R–Ab functionality is helpful and predictive in GD patients during pregnancy/postpartum, as well as for evaluating extrathyroidal manifestations.

KEY POINTS

- The thyroid-stimulating hormone receptor antibody (TSH-R–Ab), especially the stimulatory TSH-R–Ab, is the serum biomarker of Graves disease; it offers rapid and accurate diagnosis of Graves hyperthyroidism and is very useful for the differential diagnosis of thyrotoxicosis.

Imaging

In addition to thyroid function and TSH-R–Ab determination, most clinicians would request thyroid ultrasound (US) and, less often, isotope scanning.[47] Thyroid US is a convenient, noninvasive, rapid, and accurate tool in the initial work-up of GD patients. It aids in the diagnosis, without exposing the patient to ionizing irradiation, and assists in determining the underlying etiology of thyrotoxicosis and detecting concomitant thyroid nodules.[72,73] A high-frequency linear probe should be used. GD is often, but not invariably, characterized by diffuse thyroid enlargement and by hypoechogenicity, both assessed by US and conventional gray-scale analysis.[74] A color-flow or power Doppler examination characterizes vascular patterns and quantifies thyroid vascularity.[75] The latter is significantly increased in untreated GD and typically shows a pulsatile pattern called "thyroid inferno," (i.e., multiple small areas of increased intrathyroidal blood flow seen diffusely throughout the gland).[76] Typical US patterns combined with positive TSH-R–Ab obviate the need for scintigraphy in the vast majority of cases. However, RAI uptake and/or isotope scintigraphy may be useful in the assessment of patients before RAI treatment, especially when facing coexistent multinodular goiter.[74] In summary, thyroid US, comprising conventional gray-scale analysis and color-flow Doppler examination, is recommended as the imaging procedure to support the diagnosis of Graves hyperthyroidism, while thyroid scan is suggested when thyroid nodularity coexists with hyperthyroidism, and before RAI therapy.

MANAGEMENT

The hyperthyroidism of GD is treated by reducing thyroid hormone synthesis using thionamide ATDs or by reducing the amount of thyroid tissue with RAI treatment or total TX. Overall, the American guidelines[77] have been noncommittal in their approach to selecting an individual therapy and leave this to the patient and their physician after an assessment of individual preferences. In comparison, the European guidelines,[78] although offering "personalized medicine," favor ATDs as the treatment of choice.

KEY POINTS

- The thionamide antithyroid drugs (ATDs) are the most popular treatment for Graves hyperthyroidism, especially for the first onset of the disease.
- Methimazole (MMI) is the preferred ATD worldwide.
- Persistent hyperthyroidism after an 18-month first course of ATD or a relapse of Grave disease are indications for a definitive treatment, (e.g., radioactive iodine (RAI) administration or total thyroidectomy). However, long-term, low-dose MMI administration is a safe, reliable, and low-cost alternative
- *De novo* and/or exacerbation of extrathyroidal orbital disease occurs in 15%–25% of cases post-RAI.

ANTITHYROID DRUGS

ATDs have always been the first-line treatment for GD in Europe and Asia, but RAI was the preferred treatment modality in the United States.[79,80] This has changed dramatically in the last decade, and now ATDs are the most common treatment in the United States; they are used in 58% of patients, followed by RAI in 35%.[81] The reason for this shift away from RAI towards ATDs is the realization that RAI is associated with a definite risk of developing or exacerbating GO.[82,83] Numerous randomized clinical trials and metaanalyses comparing the effects of either RAI therapy or ATDs on the development or progression of GO have shown better outcomes using ATD therapy.[82,84] These differences are largely attributed to adverse effects of RAI on TSH-R autoimmunity, as evidenced by sustained increases in TSH-R–Ab titers after RAI compared with gradual decreases during a prolonged course of ATDs.[85]

The thionamide ATDs are ring-structured sulfur-containing compounds that are thiourea derivatives. ATDs inhibit the organification of inorganic iodine, an early step in thyroid hormone synthesis (Table 71.5). This process is catalyzed by thyroid peroxidase and requires

TABLE 71.5 Mechanism of Action of Thionamide Antithyroid Drugs

- Intrathyroidal inhibition of:
    - Iodine oxidation/organification
    - Iodotyrosine coupling
    - Thyroglobulin biosynthesis
    - Follicular cell growth
- Extrathyroidal inhibition of T4/T3 conversion (propylthiouracil)
- Immunological effects:
    - Direct effects on both intrathyroidal T-cells and human leukocyte antigen class II expression by thyrocytes
    - Increase numbers of suppressor T-cells and decrease intrathyroidal activated T-cells
    - Inhibit proinflammatory mediator release, including oxygen radicals, by thyroid cells

TABLE 71.6 Pharmacology and Pharmacokinetics of Thionamide Antithyroid Drugs

	Methimazole	Propylthiouracil
Absorption	Rapid	Rapid
Bioavailability	~100%	~100%
Peak serum level	60–120 minutes	60 minutes
Serum half-life	6–8 hours	90 minutes
Thyroidal concentration	5×10^{-5} mol/L	unknown
Thyroidal turnover	Slow	moderate
Duration of action	>24 hours	8–12 hours
Serum protein binding	nil	>75%
Crosses placenta	++	+
Levels in breast milk	++	+
Volume of distribution	40 L	20 L
Excretion	renal	renal
Metabolism during illness		
- Renal	nil	nil
- Liver	prolonged	Nil
Potency	10–50	1
Normalization T3/T4	6 weeks	12 weeks
Adverse events (AE, %)	15	20
Agranulocytosis (%)	0.6	1.8
Cross-reaction of AE (%)	13.8	15.2
Compliance	high	fair
Costs	low	moderate

endogenously generated hydrogen peroxide.[86,87] Three ATDs are available: methimazole (MMI), propylthiouracil (PTU), and carbimazole. Carbimazole is not an active drug and is rapidly metabolized in the blood to MMI, with 10 mg carbimazole yielding approximately 6 mg of MMI.[88] PTU, but not MMI, inhibits selenoprotein type I iodothyronine deiodinase, thereby reducing conversion of T4 to T3 in peripheral tissues. The mechanism of inhibition of type I deiodinase appears to be formation of a stable selenyl-sulfide moiety, which prevents the formation of a key selenyl-iodide intermediate.[89] ATDs have direct effects on both intrathyroidal T-cells[90] and HLA class II expression by thyrocytes,[91] as well as on increased numbers of suppressor T-cells and decreased intrathyroidal activated T-cells.[92] These effects are probably secondary to the release of drug action–inhibiting proinflammatory mediators, including oxygen radicals, by thyroid cells. MMI and PTU inhibit leukocyte production of oxygen radicals in a dose-dependent manner[93,94] and accelerate hydrogen peroxide scavenging in cultured thyroid cells.[93]

MMI can be given as a single daily dose, whereas PTU, the least potent ATD with the shortest duration of action, is initiated three times daily (Table 71.6). A PTU dose of 100 mg is equivalent to 5 to 6 mg of MMI (15–20:1). Initial dosing of MMI depends on the severity of hyperthyroidism and is based on the elevation in free T4, using 5 to 10 mg if free T4 is 1 to 1.5 times the upper limit of normal or 20 and 30 mg if free T4 is 2- and 3-fold increased.[77,78] MMI is the preferred ATD because of superior efficacy, the ability to use once daily, and a lower adverse effect profile.[95] Surveys show that less than 5% of endocrinologists choose PTU as the initial ATD.[47,79] However, PTU is still preferred during the first trimester of pregnancy because of a lower prevalence and severity of drug-associated embryopathy compared with MMI,[96] and in the treatment of thyroid storm, where the additional inhibition of T4-to-T3 conversion is critical for early reversal of this disorder.[7] Both MMI and PTU are nearly completely absorbed following oral administration, with peak serum levels obtained within 1 to 2 hours. Each drug is further concentrated within the thyroid. Intrathyroidal MMI levels are approximately 2- to 5-fold higher than peak plasma levels, and the intrathyroidal levels remain high 20 hours after ingestion.[98] Restoration of euthyroidism occurs gradually, as existing thyroid hormone stores are released and T4 is converted to T3 both within the thyroid and in the peripheral tissues.

The optimal duration of ATD therapy is approximately 18 months, with a high recurrence rate after only 6 months and no apparent additional benefit by extending therapy beyond 18 to 24 months.[99] The "block-and-replace" regimen, in which LT4 is added to ongoing ATD therapy, neither enhances the remission rate nor decreases the recurrence of hyperthyroidism.[100] [101] Hence, the evidence-based titration regimen is favored.[77,78] Approximately 50% of patients receiving ATD for 12 to 18 months will enter a period of drug-free remission, with remission rates ranging from 30% to 60%.[102] Although hyperthyroidism subsequently recurs in one-third of patients initially achieving a remission, approximately one in three patients treated achieves permanent remission (i.e., biochemical euthyroidism and negative TSH-R–Ab). Consistent features predictive of remission include TSH-R–Ab titers that normalize on therapy, a small goiter, and lesser degrees of thyrotoxicosis.[103] Features with a negative influence on remission rates include tobacco smoking, age under 30 years, and the postpartum state.[104,105] However, the predictive value of each of these risk factors is too low for accurate assessment of the remission chance before starting ATD in an individual patient. A recent prospective study was able to construct a predictive score by combining a number of independent risk factors. This so-called Graves' Recurrent Events after Therapy (GREAT) score provides a reasonable prediction of recurrent Graves hyperthyroidism after an 18-month course of ATDs.[106] The GREAT score takes into account four baseline characteristics that are already assessed routinely in the work-up of every patient with Graves hyperthyroidism: TSH-R–Ab, free T4, goiter size, and age. Patients for whom the GREAT score falls in class I have a rather high chance of remission (84%), and ATD would be a reasonable option. In contrast, the chance of remission is rather low in GREAT score class III (32%), and TX or RAI might be preferred.

In summary, patients with newly diagnosed GD are treated with ATDs. RAI therapy or TX may be considered in patients who prefer this approach. MMI should be used in every nonpregnant patient who choses ATDs. MMI is administered for 12 to 18 months, then discontinued if the TSH and TSH-R–Ab levels are normal. Measurement of TSH-R–Abs before stopping ATD therapy is recommended, as it aids in predicting which patients can be weaned from the medication, with normal levels indicating a greater chance of remission. Patients with persistently high TSH-R–Ab at 18 months can opt for definitive treatment or continue low-dose MMI therapy.

Adverse Events

Most common adverse events (AEs) ascribed to ATD include minor cutaneous allergic reactions, including rash (3%–6%), pruritis (2%–3%), and urticaria (1%–2%), as well as dyspepsia (3%–4%), nausea/gastric distress (2.4%), and arthralgia (1.6%). Cutaneous reactions to ATDs occur within the first few weeks of therapy.[107] Antihistamine therapy without stopping ATD in patients with mild skin reactions, an alternate ATD in the case of moderate AEs, and avoidance of ATDs in patients with severe cutaneous AE are recommended.[77] Major AEs (Table 71.7) include agranulocytosis, hepatotoxicity, pancreatitis, and vasculitis.[95,108] Agranulocytosis occurs in 0.2% to 0.5% of patients taking ATDs.[109] The onset is abrupt, with high fever and severe pharyngitis. Most cases occur within the first 90 days of treatment.[109] Agranulocytosis is dose-related with respect to MMI,[110], but not PTU. Agranulocytosis is caused by immune targeting of the granulocytic lineage in the bone marrow rather than a direct toxic effect.[111] ATDs are concentrated within granulocytes and metabolized by myeloperoxidase,[112] thus leading to alterations of the cell membrane. Specific HLA alleles in Asian and White persons[113,114] increase susceptibility to ATD-induced agranulocytosis. Management involves broad-spectrum antibiotics and granulocyte colony-stimulating factor.[115] Recovery of white blood cell counts occurs over 7 to 14 days. The mortality rate remains approximately 5%, with older age, comorbidities, and septicemia at the time of diagnosis being risk factors. Because of insufficient evidence and lack of cost-effectiveness, routine white blood cell count monitoring is not recommended.[77,95] PTU-induced vasculitis often occurs after years of treatment and is associated with perinuclear anticytoplasmic neutrophil antibody (ANCA).[116] Vasculitis usually resolves with drug discontinuation, but a few patients have required immunosuppressive therapy and hemodialysis. Routine screening for ANCA positivity is not recommended.[77] Hepatocellular toxicity has been described predominantly with PTU, with a cholestasis picture associated with MMI. This view has been challenged recently.[117,118] Among 71,379 patients newly prescribed ATDs, hepatotoxicity occurred in 0.3% of patients exposed to MMI and 0.15% of those given PTU. MMI and PTU were associated with noninfectious hepatitis in 0.25% and 0.08% of cases, respectively.[117] The rates of cholestasis were similar with PTU and MMI, but the rates of hepatic failure were higher with PTU than with MMI. Guidelines do not recommend routine monitoring of liver function in patients on ATDs. However, patients should be informed about possible hepatotoxicity, discontinue the offending drug, and contact their physician immediately with the onset of jaundice, dark urine, or clay-colored stools. Because up to one-third of patients with thyrotoxicosis have baseline transaminase elevation, liver function testing before starting ATDs is warranted, and ATDs should be avoided if baseline transaminase values are more than 5-fold increased.[77] In summary, patients should be informed of the potential side effects of ATDs and the necessity of informing the physician promptly if they should develop jaundice, light-colored stools, dark urine, fever, pharyngitis, or cystitis. In those cases, a differential white blood cell count and liver function should be assessed.

TABLE 71.7 Adverse Events of Thionamide Antithyroid Drugs

Common (1%–5%)

- Skin rash
- Urticaria
- Arthralgia
- Fever
- Transient mild leukopenia

Rare (0.2%–0.5%)

- Gastrointestinal
- Abnormalities of taste and smell
- Agranulocytosis (within the first 90 days)

Very rare

- Hepatitis (propylthiouracil, PTU)–avoid in children!
- Vasculitis, lupus-like (PTU)
- Aplastic anemia (PTU)
- Thrombocytopenia (PTU)
- Hypoglycemia (antiinsulin antibody) (PTU)
- Cholestatic jaundice (methimazole)
- Pancreatitis (?)

Pregnancy

GD affects approximately 0.1% of pregnancies[119] and is associated with a significantly increased risk of adverse outcomes in mother and child, as well as maternal–fetal complications.[120] Preterm labor, fetal loss, low birth weight, hydrops, pregnancy-induced hypertension, maternal congestive heart failure, and congenital malformations in offspring have been reported. Seizures and neurobehavioral disorders occurring later in life may also be related to fetal exposure to excess maternal thyroid hormones.[121] Thus, ideally, thyrotoxic women are rendered stably euthyroid before attempting pregnancy.

Treatment of GD during pregnancy and in the preconception period poses significant problems regarding indications for treatment, therapeutic options, and use of ATDs. ATDs cross the placenta and may produce fetal–neonatal hypothyroidism and teratogenic effects in the embryo.[122] MMI and PTU have similar kinetics of placental crossing and similar risks of inducing fetal and neonatal hypothyroidism.[123] Both MMI and PTU are associated with teratogenesis.[124,125] MMI-associated birth defects occur at a rate of 3% to 4% (e.g., choanal atresia, esophageal atresia, aplasia cutis), especially if the fetus is exposed during gestational weeks 6 to 10.[96,126] The risk of birth defects with PTU is 2% to 3%, although they are less severe (e.g., renal cysts, preauricular and branchial sinus or fistula).[127] Therefore, ATDs are ideally avoided in the first trimester of pregnancy, especially in the critical gestational weeks 6 to 10. This is feasible if hyperthyroidism is mild; however, ATDs are the mainstay of treatment for clinically significant GD during pregnancy. When necessary, PTU is generally favored, because of lower risk of severe birth defects. In the second and third trimester, given the risk of severe hepatotoxicity with PTU, the recommendations focus on reverting to the use of MMI.[128]

All patients with a history of autoimmune thyroid disease should have their TSH-R–Ab serum levels measured at the first presentation of pregnancy[77,78,119] using either a sensitive-binding or preferably a functional cell-based bioassay, and, if they are elevated, again at 18 to 22 weeks and, if still elevated, again at 30 to 34 weeks of gestation. If the maternal TSH-R–Ab concentration remains high (more than three times the cutoff), monitoring of the fetus for thyroid dysfunction throughout pregnancy is recommended. TSH-R–Abs are monitored in the same manner if the mother has been treated with RAI or TX before pregnancy. When GD is diagnosed before pregnancy and the patient is taking MMI, it should be switched to PTU or withdrawn as soon as pregnancy is confirmed.[119] The risk of relapse after ATD withdrawal in early pregnancy is high in those with high TSH-R–Abs, those who require greater than 10 mg of MMI per day, those with severe GO, and those who have been treated for less than 6 months. When this risk is considered high, then, instead of ATD withdrawal, therapy should be switched to PTU during the first trimester. During pregnancy, the lowest possible dose of ATDs should be given, and the block-and-replace ATD regimen is discouraged. Maternal free T4 and TSH should be measured every 2 weeks after the initiation of therapy and every 4 weeks after achieving the target value. A change from PTU to MMI should be considered if ATDs are required after 16 weeks of gestation. In women on a low dose of MMI (<5–10 mg/day) or PTU (<50–100 mg/day), ATDs may be stopped during gestation before weeks 6 to 10. Lactating women with GD should be offered the same treatments as nonlactating women. MMI is recommended during lactation, given the concerns about PTU-mediated hepatotoxicity.[77,78,119]

Long-Term Antithyroid Drug Therapy

If TSH-R–Abs are negative at the end of 12 to 18 months of ATD therapy, it is reasonable to discontinue ATDs, as the likelihood of remission is relatively high. In contrast, remission is relatively unlikely in the presence of high TSH-R–Ab levels at the end of ATD therapy. High TSH-R–Ab levels after treatment in the presence of normal serum thyroid hormone levels is often referred to as persistent serological hyperthyroidism.[78] This reflects the notion that Graves hyperthyroidism is not really cured as long as TSH-R–Abs are present. Both European and American guidelines acknowledge the feasibility of continuing ATD for a further 12 months if TSH-R–Abs are still high[78] or to proceed with long-term use of a low dose of MMI.[77] Indeed, patients increasingly report a preference for nonablative therapy.[81] The best argument, however, for long-term ATD therapy is a recent prospective randomized clinical trial showing a 4- to 5-fold increased cumulative incidence of relapse of hyperthyroidism in GD patients treated for an average of 19 months (±3 months) with MMI in comparison with a GD group treated for 95 months (±22 months, $P < .001$).[129] Further, there is a lower incidence of AEs on a low (maintenance) dose of MMI,[130] and most cases of agranulocytosis occur within the first 3 months of ATD therapy.[131] Conversely, ATD therapy requires serial reassessment of dosing adequacy, surveillance for possible overtreatment, and monitoring for rare late-occurring AEs related to ATDs.[132] A study comparing 10-year outcomes in patients randomized to either RAI or further ATD therapy after initially failing to achieve a remission on ATDs found that management costs were lower with ATDs, and episodes of hypothyroidism occurred more frequently after RAI than with ATDs. Studies of long-term ATD therapy with a block-and-replace regimen found that 90% of patients stayed continuously euthyroid during a mean follow-up of 6.7 years, with a paucity of ATD-induced AEs.[132] ATD therapy in patients treated for 2 to 11 years and then followed for an average of 4.5 years after stopping ATDs led to a remission rate of 63%.[133] A metaanalysis of patients treated with ATD for a minimum of 2 years found an AE rate of 19.1%, consisting of rash, gastric intolerance, or arthralgia, with only 1.5% of patients experiencing severe AEs.[105] Finally, patients treated with either continued low-dose ATDs or RAI following an initial relapse after ATD therapy showed that low-dose ATDs were associated with better maintenance of euthyroidism, less weight gain, less GO deterioration, and similar quality of life compared with RAI.[134] In summary, if a patient with GD becomes hyperthyroid after completing a first course of ATDs, definitive treatment with RAI or TX is recommended. However, continued long-term low-dose MMI can be considered in patients not in remission who prefer this approach.

Children, Adolescents, and the Elderly

PTU is contraindicated in children and adolescents because of its enhanced hepatotoxicity.[77,78,135] Long-term MMI should be the mainstay of treatment in children with GD. TX is the primary definitive therapy in childhood, but in postpubertal children RAI can be considered. Older patients who have had atrial fibrillation, cardiac failure, or cardiac ischemic symptoms precipitated by hyperthyroidism should undergo definitive therapy, usually RAI.[77,78] Long-term, low-dose MMI is considered a satisfactory treatment for older individuals with mild GD.

Symptomatic Management

Beta-adrenergic blockade is recommended in all patients with Graves hyperthyroidism.[42,136] Propranolol or longer-acting beta-blockers (bisoprolol/atenolol) are useful to control adrenergic symptoms (palpitations/tremor), especially in the early stages before ATDs take effect. High doses of propranolol (40 mg four times daily) inhibit peripheral conversion of T4 to T3.[77,78] Cardio-selective beta-blockers with higher cardioprotective effects and superior prevention of atrial fibrillation are an alternative, especially for patients with asthma. Anticoagulation with warfarin or direct oral anticoagulants should be considered in all patients with atrial fibrillation.[42,46,136]

Subclinical Graves Hyperthyroidism

Subclinical hyperthyroidism is defined as a low TSH and normal thyroid hormone levels in the absence of overt symptoms of hyperthyroidism. Endogenous subclinical hyperthyroidism is associated with increased risk of coronary heart disease mortality, atrial fibrillation, heart failure, fractures, and excess mortality in patients with serum TSH levels less than 0.1 mIU/L.[137-139] In addition, in the presence of TSH-R–Abs indicating "subclinical" GD, the rate of progression to overt hyperthyroidism is up to 30% in the subsequent 3 years.[140] Therefore, treatment of subclinical hyperthyroidism is recommended in GD patients in general and in those older than 65 years, in particular with completely suppressed serum TSH levels.[78] ATDs are the first choice.

RADIOACTIVE IODINE

American Thyroid Association guidelines stipulate that the goal of RAI therapy in GD is to control hyperthyroidism by rendering the patient hypothyroid.[77] Advantages of RAI include avoidance of surgical risk and potential avoidance of hormone replacement therapy. RAI treatment should be considered in the elderly and in those with significant comorbidities, contraindications for surgery, previous neck radiation, small goiters, and limited access to high-volume surgeons.[77] It is contraindicated in pregnancy or in those planning a pregnancy, as well as in those who cannot comply with radiation safety guidelines. RAI should be used with caution in GD, as it can exacerbate GO.[74,78,141] RAI treatment can be considered in the absence of GO and/or in clinically mild, inactive GO in nonsmokers, but it should be avoided in active and moderate to severe GO.[74,77,78] Clinical exacerbation of hyperthyroidism, and even thyroid storm, can occur with RAI, especially in the elderly and in those who have comorbidities that may increase the risk of thyrotoxicosis,[142] including cardiovascular disease, atrial fibrillation, heart failure, pulmonary hypertension, diabetes, renal failure, infection, trauma, or cerebrovascular disease. These patients should be given beta-blockers[42,77,136] and be considered for MMI treatment before and after RAI treatment. In individuals who will benefit from pretreatment with MMI, the recommendation is to give it before RAI but discontinue it 3 days before RAI treatment and then resume 3 days after RAI treatment with a taper, while thyroid function tests should simultaneously normalize. Young, healthy individuals who are clinically well compensated tolerate RAI without pretreatment.[77,78] In patients who are not pretreated, thyroid levels stabilize or decrease in the first month after RAI.

RAI is very effective, provided a sufficient radiation dose is delivered in the thyroid. This can be done equally well by a fixed activity of RAI (^{131}iodine) or by a calculated activity based on thyroid size and thyroid RAI uptake.[77] Balancing rapid relief of hyperthyroidism and postponement of hypothyroidism appears to be an elusive goal. No dose calculation can secure long-term euthyroidism.[78] Many centers therefore have given up meticulous dose calculation and use fixed doses of RAI (e.g., 185, 370, or 555 MBq). When evaluating the use of an ablative RAI dose in 576 GD patients 1 year after the first fixed RAI dose of 400 MBq, 17% were still hyperthyroid, 77% hypothyroid, and 6% euthyroid.[143] At 80-month follow-up, 81 patients had received a second dose, and eight patients a third dose; 3.2% were still hyperthyroid, 86.4% hypothyroid, 3.3% euthyroid, and 6.4% had died. Persistent thyrotoxicosis after the first RAI dose was associated with higher free T4 at diagnosis, higher pretreatment TSH-R–Ab, and post-RAI treatment with ATDs. New GO after RAI developed in 7.3%. RAI-associated GO is related to a steep increase in TSH-R–Abs; 3 years after RAI, TSH-R–Abs were still positive in 60% of patients.[85] Worsening or development of GO occurs more often after RAI than after treatment with ATDs (38% vs. 19%; relative risk, 1.94; 95% confidence interval: 1.4–2.7), as evident from two randomized clinical trials.[82] Hypothyroidism developed in 95% after RAI, but euthyroidism was not achieved by any participant.[82] One can conclude that elimination of Graves hyperthyroidism can be reached by RAI (although often a second, or sometimes a third, dose of RAI is required), but at the expense of a substantial risk of worsening or developing GO and creating a new disease—hypothyroidism.

Postoperative RAI for thyroid cancer increases the risk of second primary malignancies. as these doses are usually greater than 100 mCi.[144,145] Studies have been inconclusive for the risk of developing site-specific cancers after treatment with RAI for hyperthyroidism.[145,146] Doses of RAI are much less (10–15 mCi) for GD compared with postoperative cancer treatment, and absorption of radiation varies in different organs.[77] However, given the significant retained uptake within the thyroid gland in GD, the total body radiation exposure may actually be higher in treatment for GD versus cancer. A higher incidence of stomach, kidney, and breast cancers was reported in those who underwent RAI ablation for GD.[147] The multicenter Thyrotoxicosis Therapy Follow-up Study examined the radiation dose-response relationships for site-specific cancer death within the group of hyperthyroid patients (n = 18,805) who were treated with RAI.[145] This study did confirm an increased risk of secondary cancers (including breast) because of RAI treatment for hyperthyroidism. The average administered RAI activity was 10.1 mCi for GD. The relative risk of death from solid organ cancers (including breast) increases with greater doses of RAI. If patients are treated for hyperthyroidism at age 40 or 50 years, one in 31 to 52 patients or one in 32 to 55 would develop a secondary cancer. This long-term follow-up study may shift the decision-making process going forward about definitive treatment of GD.

In summary, there are no absolute indications for RAI therapy, but it is often recommended for patients with side effects from ATDs or recurrence after a course of ATDs. Verbal as well as written information regarding all aspects of efficacy and potential side effects of RAI therapy should be provided. If ATDs are used before RAI therapy, they should be paused around 1 week before and after therapy to avoid decreasing the efficacy of RAI therapy. No dose calculation can secure long-term euthyroidism, and it is fully acceptable to offer a fixed dose of RAI. Pregnancy and breastfeeding constitute to be absolute contraindications to RAI therapy. Conception should be postponed until at least 6 months after RAI in both males and females. If used in children, ablative doses aimed at rapid hypothyroidism should be administered.

THYROIDECTOMY

Advantages of total TX for hyperthyroidism allow for nearly 100% cure rates, identification of incidental cancers, and rapid achievement of euthyroid or hypothyroid states.[77,78] Total TX, when performed by a high-volume surgeon, is a cost-effective definitive treatment for GD,[148] with a failure rate of less than 1%.[77] Disadvantages of TX are perioperative and postoperative risks including bleeding, hypocalcemia, recurrent laryngeal nerve injuries, and the potential need for lifelong LT4 replacement. TX should be considered if there are compressive symptoms, suspicion for or known thyroid cancer, concurrent hyperparathyroidism, or large or substernal goiters, in pregnancy planning, pregnancy in the second trimester, or GD with nodular goiter and GO, if there is low RAI uptake, or if there is a need for fast, definitive therapy in thyrotoxicosis.[77,78,149] Preoperatively, patients should be on ATDs to establish a euthyroid state. Beta blockade is initiated for higher risk individuals, including elderly patients and those with severe thyrotoxicosis and cardiovascular disease.[42,136] In GD, inorganic iodide administration preoperatively decreases both blood flow to the thyroid and intraoperative blood loss, as well as improving the safety of TX, with decreased rates of transient hypoparathyroidism and hoarseness from transient recurrent laryngeal nerve injury.[150] The main risks of TX include anesthesia, bleeding, hypocalcemia with hypoparathyroidism, laryngeal edema, and recurrent laryngeal nerve injury. These risks, however, remain very low with experienced, high-volume surgeons (>100 thyroid surgeries per year).[151,152] Risks of recurrent laryngeal nerve injury and transient hypocalcemia are less than 1% and less than 10%, respectively.[153] In GD, preoperative vitamin D deficiency, female gender, and perioperative parathyroid autotransplantation are predictors for post-TX hypocalcemia.[154] Therefore, preoperative and postoperative supplementation with calcium carbonate and activated vitamin D, if needed, decrease rates of postoperative hypocalcemia.[155] Postoperatively, LT4 replacement is started on a weight-based calculation (1.4–1.6 μg/kg/day). TSH is measured by 6 to 8 weeks postoperatively to assess the dosing of LT4 replacement. Approximately 75% of patients will require subsequent dose adjustments before they are euthyroid.[156] In summary, if surgery is selected, total TX is the procedure of choice and should be performed by a skilled surgeon with high annual volumes of thyroidectomies. Euthyroidism should be restored by ATD before TX to avoid peri- or postoperative exacerbation of thyrotoxicosis. Vitamin D deficiency should be corrected to reduce the postoperative risk of hypocalcemia.

PERSPECTIVE AND CONCLUSION

A variety of new treatments are currently in clinical trials, which suggests that in the coming years we will have additional therapies in our armamentarium that hopefully avoid the disadvantages of our current approaches. Indeed, the novel and more disease-specific treatments for thyroidal GD aim to target primarily the TSH-R as the main autoantigen in GD and/or relevant for the pathogenesis of GD, as well as molecules playing an important role within the immunological synapse and expressed on the surface of B-cells (i.e., CD40 and CD20). In line with this, CD40 is abundantly expressed on the surface of thyrocytes and orbital target cells, and the CD40–CD154 interaction within the immunological synapse modulates humoral immunity.[157,158] Iscalimab (CFZ533), an anti-CD40 monoclonal antibody and a potent inhibitor of the CD40 pathway, led to an approximately 50% response rate when administered to patients with untreated Graves hyperthyroidism.[159]

However, in the future, causal treatment of GD will rely on monoclonal antibodies or small molecules inhibiting activation/hypertrophy of thyrocytes and orbital target cells. In detail, the disease-specific and most effective way to inhibit the increased release of free thyroid hormones induced by GD is to administer a TSH-R antagonist able to block the TSH-R and the stimulatory effect of the TSH-R–Abs. In this regard, a human anti-TSH-R monoclonal antibody (K1-70)[160,161] is being tested in a phase I trial in patients with GD. A significant concern is that this antibody has the potential for nontraditional signaling unrelated to TSH action. Further, a new, highly TSH-R–selective small molecule antagonist (S37a) was reported to inhibit TSH-R activation by both TSH itself and the human monoclonal stimulatory TSH-R–Ab M22.[162] Finally, the clinical effect of a combination of different TSH-R peptides called ATX-GD-59[163] able to stimulate the generation of regulatory T-cells and suppress the immune response against TSH-R was assessed in patients with untreated GD. A complete or partial short-term response was noted in 70%.[164]

ACKNOWLEDGMENT

We acknowledge the editorial assistance of Miss Lara Frommer, MSc, Thyroid Lab, JGU Medical Center, Mainz, Germany.

REFERENCES

1. Nabipour I. Clinical endocrinology in the Islamic civilization in Iran. *Int J Endocrinol Metab*. 2003;1:43–45.
2. Parry CH. Collections from the unpublished medical writings of the late Caleb Hillier Parry. London: Underwood; 1825.
3. Graves RJ. Clinical lectures. *Lond Med Surg J*. 1835;7:516–517.
4. Basedow von CA. [Exophthalmos durch Hypertrophie des Zellgewebes in der Augenhöhle.] *Wochenschr ges Heilk*. 1840;6:197–204.
5. Adams DD. Long-acting thyroid stimulator: how receptor autoimmunity was discovered. *Autoimmunity*. 1988;1:3–9.
6. Ferrari SM, Fallahi P, Ruffilli I, et al. The association of other autoimmune diseases in patients with Graves' disease (with or without ophthalmopathy): review of the literature and report of a large series. *Autoimmun Rev*. 2019;18:287–292.
7. Brix TH, Kyvik KO, Christensen K, et al. Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. *J Clin Endocrinol Metab*. 2001;86:930–934.
8. Simmonds MJ, Howson JM, Heward JM, et al. A novel and major association of HLA-C in Graves' disease that eclipses the classical HLA-DRB1 effect. *Hum Mol Genet*. 2007;16:2149–2153.
9. Okada Y, Momozawa Y, Ashikawa K, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat Genet*. 2015;47:798–802.
10. Chu X, Yang M, Song ZJ, et al. Fine mapping MHC associations in Graves' disease and its clinical subtypes in Han Chinese. *J Med Genet*. 2018;55:685–692.
11. Lee HJ, Li CW, Hammerstad SS, et al. Immunogenetics of autoimmune thyroid diseases: a comprehensive review. *J Autoimmun*. 2015;64:82–90.
12. Cooper JD, Simmonds MJ, Walker NM, et al. Seven newly identified loci for autoimmune thyroid disease. *Hum Mol Genet*. 2012;21:5202–5208.
13. Pujol-Borrell R, Gimenez-Barcons M, Marin-Sanchez A, et al. Genetics of Graves' disease: special focus on the role of TSHR gene. *Horm Metab Res*. 2015;47:753–766.
14. McLachlan SM, Rapoport B. Breaking tolerance to thyroid antigens: changing concepts in thyroid autoimmunity. *Endocr Rev*. 2014;35:59–105.
15. Leung AM, Braverman LE. Consequences of excess iodine. *Nat Rev Endocrinol*. 2014;10:136–142.
16. Bednarczuk T, Schomburg L. Challenges and perspectives of selenium supplementation in Graves' disease and orbitopathy. *Hormones (Athens)*. 2020;19:31–39.
17. Kondrashova A, Viskari H, Haapala AM, et al. Serological evidence of thyroid autoimmunity among schoolchildren in two different socioeconomic environments. *J Clin Endocrinol Metab*. 2008;93:729–734.
18. Bagnasco M, Bossert I, Pesce G. Stress and autoimmune thyroid diseases. *Neuroimmunomodulation*. 2006;13:309–317.
19. Komiya I, Yamada T, Sato A, et al. Remission and recurrence of hyperthyroid Graves' disease during and after methimazole treatment when assessed by IgE and interleukin 13. *J Clin Endocrinol Metab (Oxf)*. 2001;86:3540–3544.
20. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol (Oxf)*. 2013;79:145–151.
21. Carle A, Bulow Pedersen I, Knudsen N, et al. Graves' hyperthyroidism and moderate alcohol consumption: evidence for disease prevention. *Clin Endocrinol*. 2013;79:111–119.
22. Chalan P, Di Dalmazi G, Pani F, et al. Thyroid dysfunctions secondary to cancer immunotherapy. *J Endocrinol Invest*. 2018;41:625–638.
23. Muller I, Moran C, Lecumberri B, et al. 2019 European Thyroid Association Guidelines on the management of thyroid dysfunction following immune reconstitution therapy. *Eur Thyroid J*. 2019;8:173–185.
24. McLeod DS, Caturegli P, Cooper DS, et al. Variation in rates of autoimmune thyroid disease by race/ethnicity in US military personnel. *J Am Med Assoc*. 2014;311:1563–1565.
25. Carle A, Knudsen N, Jorgensen T, et al. Previous live births and induced abortions may precede later development of Graves' hyperthyroidism. *Eur Thyroid J*. 2019;8:70–78.
26. Jang D, Morgan SJ, Klubo-Gwiezdzinska J, et al. Thyrotropin, but not thyroid-stimulating antibodies, induces biphasic regulation of gene expression in human thyrocytes. *Thyroid*. 2020;30s:270–276.
27. Morshed SA, Davies TF. Graves' disease mechanisms: the role of stimulating, blocking, and cleavage region TSH receptor antibodies. *Horm Metab Res*. 2015;47:727–734.
28. Rapoport B, McLachlan SM. TSH receptor cleavage into subunits and shedding of the A-subunit; a molecular and clinical perspective. *Endocr Rev*. 2016;37:114–134.
29. Fallahi P, Ferrari SM, Ragusa F, et al. Th1 chemokines in autoimmune endocrine disorders. *J Clin Endocrinol Metab*. 2020;105.
30. McLachlan SM, Rapoport B. Thyrotropin-blocking autoantibodies and thyroid-stimulating autoantibodies: potential mechanisms involved in the pendulum swinging from hypothyroidism to hyperthyroidism or vice versa. *Thyroid*. 2013;23:14–24.
31. Bluestone JA, Bour-Jordan H, Cheng M, et al. T cells in the control of organ-specific autoimmunity. *J Clin Invest*. 2015;125:2250–2260.
32. Weetman A, DeGroot LJ. Autoimmunity to the thyroid gland. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *South Dartmouth*. Endotext; 2000.
33. Nagayama Y. Graves' animal models of Graves' hyperthyroidism. *Thyroid*. 2007;17:981–988.
34. McLachlan SM, Aliesky HA, Rapoport B. A mouse thyrotropin receptor a-subunit transgene expressed in thyroiditis-prone mice may provide insight into why Graves' disease only occurs in humans. *Thyroid*. 2019;29:1138–1146.
35. McLachlan SM, Aliesky HA, Banuelos B, et al. High-level intrathymic thyrotrophin receptor expression in thyroiditis-prone mice protects against the spontaneous generation of pathogenic thyrotrophin receptor autoantibodies. *Clin Exp Immunol*. 2017;188:243–253.
36. Lumachi F, Basso SM, Orlando R. Cytokines, thyroid diseases and thyroid cancer. *Cytokine*. 2010;50:229–233.
37. Ramos-Levi AM, Marazuela M. Pathogenesis of thyroid autoimmune disease: the role of cellular mechanisms. *Endocrinol Nutr*. 2016;63:421–429.
38. Singh B, Summers KL, Kerfoot SM. Novel regulatory Th17 cells and regulatory B cells in modulating autoimmune diseases. *Cell Immunol*. 2019;339:29–32.
39. Martinez-Hernandez R, Sampedro-Nunez M, Serrano-Somavilla A, et al. A microRNA signature for evaluation of risk and severity of autoimmune thyroid diseases. *J Clin Endocrinol Metab*. 2018;103:1139–1150.
40. Brent GA. Clinical practice. Graves' disease. *N Engl J Med*. 2008;358:2594–2605.
41. Boelaert K, Torlinska B, Holder RL, et al. Older subjects with hyperthyroidism present with a paucity of symptoms and signs: a large cross-sectional study. *J Clin Endocrinol Metab*. 2010;95:2715–2726.
42. Biondi B, Kahaly GJ. Cardiovascular involvement in patients with different causes of hyperthyroidism. *Nat Rev Endocrinol*. 2010;6:431–443.
43. de los Santos ET, Starich GH, Mazzaferri EL. Sensitivity, specificity, and cost-effectiveness of the sensitive thyrotropin assay in the diagnosis of thyroid disease in ambulatory patients. *Arch Intern Med*. 1989;149:526–532.

44. Spencer CA, LoPresti JS, Patel A, et al. Applications of a new chemiluminometric thyrotropin assay to subnormal measurement. *J Clin Endocrinol Metab*. 1990;70:453–460.
45. Grebe SK, Kahaly GJ. Laboratory testing in hyperthyroidism. *Am J Med*. 2012;125:S2.
46. Bartalena L. Diagnosis and management of Graves disease: a global overview. *Nat Rev Endocrinol*. 2013;9:724–734.
47. Bartalena L, Burch HB, Burman KD, et al. 2013 European survey of clinical practice patterns in the management of Graves' disease. *Clin Endocrinol (Oxf)*. 2016;84:115–120.
48. Kahaly GJ, Olivo PD. Graves' disease. *N Engl J Med*. 2017;376:184.
49. Kahaly GJ, Diana T. TSH receptor antibody functionality and nomenclature. *Front Endocrinol (Lausanne)*. 2017;8:28.
50. Tozzoli R, Bagnasco M, Giavarina D, et al. TSH receptor autoantibody immunoassay in patients with Graves' disease: improvement of diagnostic accuracy over different generations of methods. Systematic review and meta-analysis. *Autoimmun Rev*. 2012;12:107–113.
51. Lytton SD, Li Y, Olivo PD, et al. Novel chimeric thyroid-stimulating hormone-receptor bioassay for thyroid-stimulating immunoglobulins. *Clin Exp Immunol*. 2010;162:438–446.
52. Leschik JJ, Diana T, Olivo PD, et al. Analytical performance and clinical utility of a bioassay for thyroid-stimulating immunoglobulins. *Am J Clin Pathol*. 2013;139:192–200.
53. Li Y, Kim J, Diana T, et al. A novel bioassay for anti-thyrotrophin receptor autoantibodies detects both thyroid-blocking and stimulating activity. *Clin Exp Immunol*. 2013;173:390–397.
54. Araki N, Iida M, Amino N, et al. Rapid bioassay for detection of thyroid-stimulating antibodies using cyclic adenosine monophosphate-gated calcium channel and aequorin. *Eur Thyroid J*. 2015;4:14–19.
55. Kahaly GJ. Bioassays for TSH receptor antibodies: quo vadis? *Eur Thyroid J*. 2015;4:3–5.
56. Diana T, Kanitz M, Lehmann M, et al. Standardization of a bioassay for thyrotropin receptor stimulating autoantibodies. *Thyroid*. 2015;25:169–175.
57. Diana T, Li Y, Olivo PD, et al. Analytical performance and validation of a bioassay for thyroid-blocking antibodies. *Thyroid*. 2016;26:734–740.
58. Diana T, Krause J, Olivo PD, et al. Prevalence and clinical relevance of thyroid stimulating hormone receptor-blocking antibodies in autoimmune thyroid disease. *Clin Exp Immunol*. 2017;189:304–309.
59. Diana T, Wuster C, Kanitz M, et al. Highly variable sensitivity of five binding and two bio-assays for TSH-receptor antibodies. *J Endocrinol Invest*. 2016;39:1159–1165.
60. Diana T, Wuster C, Olivo PD, et al. Performance and specificity of 6 immunoassays for TSH receptor antibodies: a multicenter study. *Eur Thyroid J*. 2017;6:243–249.
61. Allelein S, Diana T, Ehlers M, et al. Comparison of a bridge immunoassay with two bioassays for thyrotropin receptor antibody detection and differentiation. *Horm Metab Res*. 2019;51:341–346.
62. Lytton SD, Ponto KA, Kanitz M, et al. A novel thyroid stimulating immunoglobulin bioassay is a functional indicator of activity and severity of Graves' orbitopathy. *J Clin Endocrinol Metab*. 2010;95:2123–2131.
63. Ponto KA, Kanitz M, Olivo PD, et al. Clinical relevance of thyroid-stimulating immunoglobulins in Graves' ophthalmopathy. *Ophthalmology*. 2011;118:2279–2285.
64. Diana T, Brown RS, Bossowski A, et al. Clinical relevance of thyroid-stimulating autoantibodies in pediatric graves' disease-a multicenter study. *J Clin Endocrinol Metab*. 2014;99:1648–1655.
65. Ponto KA, Diana T, Binder H, et al. Thyroid-stimulating immunoglobulins indicate the onset of dysthyroid optic neuropathy. *J Endocrinol Invest*. 2015;38:769–777.
66. Kampmann E, Diana T, Kanitz M, et al. Thyroid stimulating but not blocking autoantibodies are highly prevalent in severe and active thyroid-associated orbitopathy: a prospective study. *Int J Endocrinol*. 2015;2015:678194.
67. Kahaly GJ, Diana T, Glang J, et al. Thyroid stimulating antibodies are highly prevalent in Hashimoto's thyroiditis and associated orbitopathy. *J Clin Endocrinol Metab*. 2016;101:1998–2004.
68. Stozek K, Bossowski A, Ziora K, et al. Functional TSH receptor antibodies in children with autoimmune thyroid diseases. *Autoimmunity*. 2018;51:62–68.
69. Kiefer FW, Klebermass-Schrehof K, Steiner M, et al. Fetal/neonatal thyrotoxicosis in a newborn from a hypothyroid woman with Hashimoto thyroiditis. *J Clin Endocrinol Metab*. 2017;102:6–9.
70. Decallonne B, Martens PJ, Van den Bruel A, et al. Graves disease with thyroid-stimulating hormone receptor-blocking autoantibodies during pregnancy. *Ann Intern Med*. 2020;172:767–769.
71. McKee A, Peyerl F. TSI assay utilization: impact on costs of Graves' hyperthyroidism diagnosis. *Am J Manag Care*. 2012;18:e1–e14.
72. Hegedus L. Thyroid ultrasound. *Endocrinol Metab Clin North Am*. 2001;30:339–360, viii–ix.
73. Kahaly GJ, Bartalena L, Hegedus L. The American Thyroid Association/American Association of Clinical Endocrinologists guidelines for hyperthyroidism and other causes of thyrotoxicosis: a European perspective. *Thyroid*. 2011;21:585–591.
74. Bartalena L, Baldeschi L, Boboridis K, et al. The 2016 European Thyroid Association/European Group on Graves' orbitopathy guidelines for the management of graves' orbitopathy. *Eur Thyroid J*. 2016;5:9–26.
75. Erdogan MF, Anil C, Cesur M, et al. Color flow Doppler sonography for the etiologic diagnosis of hyperthyroidism. *Thyroid*. 2007;17:223–228.
76. Ralls PW, Mayekawa DS, Lee KP, et al. Color-flow Doppler sonography in Graves disease: "thyroid inferno". *AJR Am J Roentgenol*. 1988;150:781–784.
77. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.
78. Kahaly GJ, Bartalena L, Hegedus L, et al. 2018 European Thyroid Association guideline for the management of Graves' hyperthyroidism. *Eur Thyroid J*. 2018;7:167–186.
79. Burch HB, Burman KD, Cooper DS. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab*. 2012;97:4549–4558.
80. Bartalena L, Burch HB, Burman KD, et al. A 2013 European survey of clinical practice patterns in the management of Graves' disease. *Clin Endocrinol (Oxf)*. 2016;84:115–120.
81. Brito JP, Schilz S, Singh Ospina N, et al. Antithyroid drugs-the most common treatment for Graves' disease in the United States: a nationwide population-based study. *Thyroid*. 2016;26:1144–1145.
82. Ma C, Xie J, Wang H, et al. Radioiodine therapy versus antithyroid medications for Graves' disease. *Cochrane Database Syst Rev*. 2016;2:CD010094.
83. Fanning E, Inder WJ, Mackenzie E. Radioiodine treatment for Graves' disease: a 10-year Australian cohort study. *BMC Endocr Disord*. 2018;18:94.
84. Li HX, Xiang N, Hu WK, et al. Relation between therapy options for Graves' disease and the course of Graves' ophthalmopathy: a systematic review and meta-analysis. *J Endocrinol Invest*. 2016;39:1225–1233.
85. Laurberg P, Wallin G, Tallstedt L, et al. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*. 2008;158:69–75.
86. Taurog A, Riesco G, Larsen PR. Formation of 3,3'-diiodothyronine and 3',5',3-triiodothyronine (reverse T3) in thyroid glands of rats and in enzymatically iodinated thyroglobulin. *Endocrinology*. 1976;99:281–290.
87. Davidson B, Soodak M, Neary JT, et al. The irreversible inactivation of thyroid peroxidase by methylmercaptoimidazole, thiouracil, and propylthiouracil in vitro and its relationship to in vivo findings. *Endocrinology*. 1978;103:871–882.
88. Jansson R, Dahlberg PA, Lindstrom B. Comparative bioavailability of carbimazole and methimazole. *Int J Clin Pharmacol Ther Toxicol*. 1983;21:505–510.
89. Kuiper GG, Kester MH, Peeters RP, et al. Biochemical mechanisms of thyroid hormone deiodination. *Thyroid*. 2005;15:787–798.
90. Humar M, Dohrmann H, Stein P, et al. Thionamides inhibit the transcription factor nuclear factor-kappaB by suppression of Rac1 and inhibitor of kappaB kinase alpha. *J Pharmacol Exp Ther*. 2008;324:1037–1044.
91. Zantut-Wittmann DE, Tambascia MA, da Silva Trevisan MA, et al. Antithyroid drugs inhibit in vivo HLA-DR expression in thyroid follicular cells in Graves' disease. *Thyroid*. 2001;11:575–580.

92. Totterman TH, Karlsson FA, Bengtsson M, et al. Induction of circulating activated suppressor-like T cells by methimazole therapy for Graves' disease. *N Engl J Med*. 1987;316:15–22.
93. Imamura M, Aoki N, Saito T, et al. Inhibitory effects of antithyroid drugs on oxygen radical formation in human neutrophils. *Acta Endocrinol-Cop*. 1986;112:210–216.
94. Weetman AP, Holt ME, Campbell AK, et al. Methimazole and generation of oxygen radicals by monocytes: potential role in immunosuppression. *Br Med J*. 1984;288:518.
95. Cooper DS. Antithyroid drugs. *N Engl J Med*. 2005;352:905–917.
96. Andersen SL, Olsen J, Wu CS, et al. Birth defects after early pregnancy use of antithyroid drugs: a Danish nationwide study. *J Clin Endocrinol Metab*. 2013;98:4373–4381.
97. Burch HB, Wartofsky L. Life-threatening thyrotoxicosis. Thyroid storm. *Endocrinol Metab Clin North Am*. 1993;22:263–277.
98. Jansson R, Dahlberg PA, Johansson H, et al. Intrathyroidal concentrations of methimazole in patients with Graves' disease. *J Clin Endocrinol Metab*. 1983;57:129–132.
99. Abraham P, Avenell A, McGeoch SC, et al. Antithyroid drug regimen for treating Graves' hyperthyroidism. *Cochrane Database Syst Rev*. 2010;1:CD003420.
100. McIver B, Rae P, Beckett G, et al. Lack of effect of thyroxine in patients with Graves' hyperthyroidism who are treated with an antithyroid drug. *N Engl J Med*. 1996;334:220–224.
101. Glinoer D, de Nayer P, Bex M, et al. Effects of l-thyroxine administration, TSH-receptor antibodies and smoking on the risk of recurrence in Graves' hyperthyroidism treated with antithyroid drugs: a double-blind prospective randomized study. *Eur J Endocrinol*. 2001;144:475–483.
102. Okamoto Y, Tanigawa S, Ishikawa K, et al. TSH receptor antibody measurements and prediction of remission in Graves' disease patients treated with minimum maintenance doses of antithyroid drugs. *Endocr J*. 2006;53:467–472.
103. Cooper DS. Antithyroid drugs in the management of patients with Graves' disease: an evidence-based approach to therapeutic controversies. *J Clin Endocrinol Metab*. 2003;88:3474–3481.
104. Rotondi M, Cappelli C, Pirali B, et al. The effect of pregnancy on subsequent relapse from Graves' disease after a successful course of antithyroid drug therapy. *J Clin Endocrinol Metab*. 2008;93:3985–3988.
105. Azizi F, Malboosbaf R. Long-term antithyroid drug treatment: a systematic review and meta-analysis. *Thyroid*. 2017;27:1223–1231.
106. Vos XG, Endert E, Zwinderman AH, et al. Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with Graves' hyperthyroidism. *J Clin Endocrinol Metab*. 2016;101:1381–1389.
107. Nakamura H, Noh JY, Itoh K, et al. Comparison of methimazole and propylthiouracil in patients with hyperthyroidism caused by Graves' disease. *J Clin Endocrinol Metab*. 2007;92:2157–2162.
108. Brix TH, Lund LC, Henriksen DP, et al. Methimazole and risk of acute pancreatitis. *Lancet Diabetes Endocrinol*. 2020;8:187–189.
109. Nakamura A, Bak B, Silander TL, et al. Three novel IGSF1 mutations in four Japanese patients with X-linked congenital central hypothyroidism. *J Clin Endocrinol Metab*. 2013;98:E1682–E1691.
110. Takata K, Kubota S, Fukata S, et al. Methimazole-induced agranulocytosis in patients with Graves' disease is more frequent with an initial dose of 30 mg daily than with 15 mg daily. *Thyroid*. 2009;19:559–563.
111. Johnston A, Uetrecht J. Current understanding of the mechanisms of idiosyncratic drug-induced agranulocytosis. *Expert Opin Drug Metab Toxicol*. 2015;11:243–257.
112. Lam KS, Kasper CB. Selective phosphorylation of a nuclear envelope polypeptide by an endogenous protein kinase. *Biochemistry*. 1979;18:307–311.
113. Chen PL, Shih SR, Wang PW, et al. Genetic determinants of antithyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. *Nat Commun*. 2015;6:7633.
114. Hallberg P, Eriksson N, Ibanez L, et al. Genetic variants associated with antithyroid drug-induced agranulocytosis: a genome-wide association study in a European population. *Lancet Diabetes Endocrinol*. 2016;4:507–516.
115. Andres E, Zimmer J, Mecili M, et al. Clinical presentation and management of drug-induced agranulocytosis. *Expert Rev Hematol*. 2011;4:143–151.
116. Balavoine AS, Glinoer D, Dubucquoi S, et al. Antineutrophil cytoplasmic antibody-positive small-vessel vasculitis associated with antithyroid drug therapy: how significant is the clinical problem? *Thyroid*. 2015;25:1273–1281.
117. Wang MT, Lee WJ, Huang TY, et al. Antithyroid drug-related hepatotoxicity in hyperthyroidism patients: a population-based cohort study. *Br J Clin Pharmacol*. 2014;78:619–629.
118. Yang J, Li LF, Xu Q, et al. Analysis of 90 cases of antithyroid drug-induced severe hepatotoxicity over 13 years in China. *Thyroid*. 2015;25:278–283.
119. Alexander EK, Pearce EN, Brent GA, et al. 2017 guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27:315–389.
120. Kriplani A, Buckshee K, Bhargava VL, et al. Maternal and perinatal outcome in thyrotoxicosis complicating pregnancy. *Eur J Obstet Gynecol Reprod Biol*. 1994;54:159–163.
121. Andersen SL, Andersen S, Vestergaard P, et al. Maternal thyroid function in early pregnancy and child neurodevelopmental disorders: a Danish nationwide case-cohort study. *Thyroid*. 2018;28:537–546.
122. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:238–249.
123. Momotani N, Noh JY, Ishikawa N, et al. Effects of propylthiouracil and methimazole on fetal thyroid status in mothers with Graves' hyperthyroidism. *J Clin Endocrinol Metab*. 1997;82:3633–3636.
124. Seo GH, Kim TH, Chung JH. Antithyroid drugs and congenital malformations: a nationwide Korean cohort study. *Ann Intern Med*. 2018;168:405–413.
125. Andersen SL, Knosgaard L, Olsen J, et al. Maternal thyroid function, use of antithyroid drugs in early pregnancy, and birth defects. *J Clin Endocrinol Metab*. 2019;104:6040–6048.
126. Laurberg P, Andersen SL. Therapy of endocrine disease: antithyroid drug use in early pregnancy and birth defects: time windows of relative safety and high risk? *Eur J Endocrinol*. 2014;171:R13–R20.
127. Andersen SL, Olsen J, Wu CS, et al. Severity of birth defects after propylthiouracil exposure in early pregnancy. *Thyroid*. 2014;24:1533–1540.
128. Andersen SL, Andersen S. Timing of shift in antithyroid drug therapy and birth defects. *Thyroid*. 2019;29:155–156.
129. Azizi F, Amouzegar A, Tohidi M, et al. Increased remission rates after long-term methimazole therapy in patients with Graves' disease: results of a randomized clinical trial. *Thyroid*. 2019;29:1192–1200.
130. Sato S, Noh JY, Sato S, et al. Comparison of efficacy and adverse effects between methimazole 15 mg+inorganic iodine 38 mg/day and methimazole 30 mg/day as initial therapy for Graves' disease patients with moderate to severe hyperthyroidism. *Thyroid*. 2015;25:43–50.
131. Nakamura H, Miyauchi A, Miyawaki N, et al. Analysis of 754 cases of antithyroid drug-induced agranulocytosis over 30 years in Japan. *J Clin Endocrinol Metab*. 2013;98:4776–4783.
132. Laurberg P, Berman DC, Andersen S, et al. Sustained control of Graves' hyperthyroidism during long-term low-dose antithyroid drug therapy of patients with severe Graves' orbitopathy. *Thyroid*. 2011;21:951–956.
133. Elbers L, Mourits M, Wiersinga W. Outcome of very long-term treatment with antithyroid drugs in Graves' hyperthyroidism associated with Graves' orbitopathy. *Thyroid*. 2011;21:279–283.
134. Villagelin D, Romaldini JH, Santos RB, et al. Outcomes in relapsed Graves' disease patients following radioiodine or prolonged low dose of methimazole treatment. *Thyroid*. 2015;25:1282–1290.
135. Rivkees SA, Mattison DR. Ending propylthiouracil-induced liver failure in children. *N Engl J Med*. 2009;360:1574–1575.
136. Kahaly GJ, Dillmann WH. Thyroid hormone action in the heart. *Endocr Rev*. 2005;26:704–728.
137. Biondi B, Bartalena L, Cooper DS, et al. The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. *Eur Thyroid J*. 2015;4:149–163.
138. Wirth CD, Blum MR, da Costa BR, et al. Subclinical thyroid dysfunction and the risk for fractures: a systematic review and meta-analysis. *Ann Intern Med*. 2014;161:189–199.
139. Blum MR, Bauer DC, Collet TH, et al. Subclinical thyroid dysfunction and fracture risk: a meta-analysis. *J Am Med Assoc*. 2015;313:2055–2065.

140. Zhyzhneuskaya S, Addison C, Tsatlidis V, et al. The natural history of subclinical hyperthyroidism in Graves' disease: the rule of thirds. *Thyroid*. 2016;26:765–769.
141. Bartalena L, Baldeschi L, Dickinson A, et al. Consensus statement of the European Group on Graves' orbitopathy (EUGOGO) on management of GO. *Eur J Endocrinol*. 2008;158:273–285.
142. Burch HB, Solomon BL, Cooper DS, et al. The effect of antithyroid drug pretreatment on acute changes in thyroid hormone levels after (131)I ablation for Graves' disease. *J Clin Endocrinol Metab*. 2001;86:3016–3021.
143. Aung ET, Zammitt NN, Dover AR, et al. Predicting outcomes and complications following radioiodine therapy in Graves' thyrotoxicosis. *Clin Endocrinol (Oxf)*. 2019;90:192–199.
144. Teng CJ, Hu YW, Chen SC, et al. Use of radioactive iodine for thyroid cancer and risk of second primary malignancy: a nationwide population-based study. *J Natl Cancer Inst*. 2016;108.
145. Kitahara CM, Berrington de Gonzalez A, Bouville A, et al. Association of radioactive iodine treatment with cancer mortality in patients with hyperthyroidism. *JAMA Intern Med*. 2019;179:1034–1042.
146. Ryodi E, Metso S, Jaatinen P, et al. Cancer incidence and mortality in patients treated either with RAI or thyroidectomy for hyperthyroidism. *J Clin Endocrinol Metab*. 2015;100:3710–3717.
147. Metso S, Auvinen A, Huhtala H, et al. Increased cancer incidence after radioiodine treatment for hyperthyroidism. *Cancer*. 2007;109:1972–1979.
148. Genovese BM, Noureldine SI, Gleeson EM, et al. What is the best definitive treatment for Graves' disease? A systematic review of the existing literature. *Ann Surg Oncol*. 2013;20:660–667.
149. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet*. 2016;388:906–918.
150. Randle RW, Bates MF, Long KL, et al. Impact of potassium iodide on thyroidectomy for Graves' disease: implications for safety and operative difficulty. *Surgery*. 2018;163:68–72.
151. Dralle H. [Adrenocortical tumors > 8 cm should be treated by open surgery]. *Chirurg*. 2015;86:897.
152. Hauch A, Al-Qurayshi Z, Randolph G, et al. Total thyroidectomy is associated with increased risk of complications for low- and high-volume surgeons. *Ann Surg Oncol*. 2014;21:3844–3852.
153. Porterfield JR, Thompson GB, Young Jr WF, et al. Surgery for Cushing's syndrome: an historical review and recent ten-year experience. *World J Surg*. 2008;32:659–677.
154. Al Qubaisi M, Haigh PI. Hypocalcemia after total thyroidectomy in Graves disease. *Perm J*. 2019;23:18–188.
155. Oltmann SC, Brekke AV, Schneider DF, et al. Preventing postoperative hypocalcemia in patients with Graves disease: a prospective study. *Ann Surg Oncol*. 2015;22:952–958.
156. Elfenbein DM, Schaefer S, Shumway C, et al. Prospective intervention of a novel levothyroxine dosing protocol based on body mass index after thyroidectomy. *J Am Coll Surg*. 2016;222:83–88.
157. Cordoba F, Wieczorek G, Audet M, et al. A novel, blocking, Fc-silent anti-CD40 monoclonal antibody prolongs nonhuman primate renal allograft survival in the absence of B cell depletion. *Am J Transplant*. 2015;15:2825–2836.
158. Ristov J, Espie P, Ulrich P, et al. Characterization of the in vitro and in vivo properties of CFZ533, a blocking and non-depleting anti-CD40 monoclonal antibody. *Am J Transplant*. 2018;18:2895–2904.
159. Kahaly GJ, Stan MN, Frommer L, et al. A novel anti-CD40 monoclonal antibody, iscalimab, for control of Graves hyperthyroidism-a proof-of-concept trial. *J Clin Endocrinol Metab*. 2020:105.
160. Sanders P, Young S, Sanders J, et al. Crystal structure of the TSH receptor (TSHR) bound to a blocking-type TSHR autoantibody. *J Mol Endocrinol*. 2011;46:81–99.
161. Furmaniak J, Sanders J, Rees Smith B. Blocking type TSH receptor antibodies. *Auto Immun Highlights*. 2013;4:11–26.
162. Marcinkowski P, Hoyer I, Specker E, et al. A new highly thyrotropin receptor-selective small-molecule antagonist with potential for the treatment of Graves' orbitopathy. *Thyroid*. 2019;29:111–123.
163. Jansson L, Vrolix K, Jahraus A, et al. Immunotherapy with apitopes blocks the immune response to TSH receptor in HLA-DR transgenic mice. *Endocrinology*. 2018;159:3446–3457.
164. Pearce SHS, Dayan C, Wraith DC, et al. Antigen-specific immunotherapy with thyrotropin receptor peptides in Graves' hyperthyroidism: a phase i study. *Thyroid*. 2019;29:1003–1011.

72

Graves' Orbitopathy

Alan Chun Hong Lee and George J. Kahaly

OUTLINE

Epidemiology, 1194
Risk Factors, 1194
  Age, 1194
  Gender, 1194
  Ethnicity, 1195
  Smoking, 1195
  High Thyroid Stimulating Hormone Receptor Antibody Titer, 1195
  Radioactive Iodine Therapy, 1195
  Cholesterol Level and Statin Use, 1195
Pathogenesis, 1196
  Key Pathological Changes in Graves' Orbitopathy, 1196
  Thyroid Stimulating Hormone Receptor as the Putative Autoantigen, 1196
  Immune Dysregulation, 1196
  Orbital Fibroblasts as Key Effector Cells in Graves' Orbitopathy, 1197
  Signaling Pathways Mediating Hyaluronan Production and Adipogenesis, 1197
  Orbital Infiltration of CD34+ Fibrocytes, 1197
  Thyroid Stimulating Hormone Receptor and Insulin-Like Growth Factor-1 Receptor Crosstalk, 1197
  Genetic Susceptibility, 1198
Clinical Manifestations And Diagnosis, 1199
  Clinical Evaluation, 1199
  Thyroid Function Test, 1200
  Thyroid Stimulating Hormone Receptor Antibody, 1201
  Orbital Imaging, 1202
  Quality of Life Assessment, 1203
Management, 1203
  General Measures, 1203
  Mild Graves' Orbitopathy, 1205
  Active Moderate-to-Severe Graves' Orbitopathy, 1206
  Sight-Threatening Graves' Orbitopathy, 1207
  Corneal Breakdown and Other Forms of Sight-Threatening Graves' Orbitopathy, 1207
  Inactive Graves' Orbitopathy and the Role of Rehabilitative Surgery, 1207
  Novel Treatment Approaches for Immunosuppression in Graves' Orbitopathy, 1207
Brief Note On Graves Dermopathy And Thyroid Acropachy, 1210
Summary And Perspectives, 1210

EPIDEMIOLOGY

Graves' orbitopathy (GO), also known as thyroid eye disease, Graves ophthalmopathy, or thyroid-associated ophthalmopathy, is the most common extrathyroidal manifestation of Graves disease (GD). European Group on Graves' Orbitopathy (EUGOGO) estimated the prevalence of GO in Europe to be around 10 per 10,000 persons based on reported annual incidences of GO at approximately 5 per 100,000.[1] A recent meta-analysis involving 33 studies and over 20,000 patients suggested that the overall prevalence of GO in GD patients was 40%, and it did not significantly change with time when studies conducted before and after 2010 were compared.[2] Mild and moderate-to-severe GO accounted for around two-thirds and one-third of the cases, respectively, while 2% of GO patients suffered from sight-threatening GO.[1] Following the Amsterdam Declaration in 2009, which aimed at optimizing GO care, changes in referral patterns and epidemiology have been observed in European multidisciplinary GO centers. The increased proportion of patients with mild and inactive GO may be explained by several factors: (1) earlier referrals and assessments of GO in tertiary centres; (2) decline in the prevalence of smoking; (3) earlier diagnosis and treatment of Graves hyperthyroidism to maintain stable euthyroidism; and (4) more judicious (and probably less frequent) use of radioactive iodine (RAI) therapy in Graves hyperthyroidism.[3,4] Prompt multidisciplinary care is believed to improve both immediate and long-term outcomes of GO patients.

RISK FACTORS

Age

GO tends to affect middle-aged people, with an average age at diagnosis between 42 and 55 years.[3-5] Older age at onset of GO is associated with more severe disease presentation, including soft tissue involvement and ocular dysmotility, and predicts the development of dysthyroid optic neuropathy (DON), a form of sight-threatening GO.[6] A Danish study further showed that GD patients aged 40 to 60 years were at higher risk of developing moderate-to-severe GO compared with younger patients.[7]

Gender

Like Graves hyperthyroidism, GO is more common in women than in men. The overall female-to-male ratio was 3.7:1, according to large series from European and Australian tertiary referral centers.[3,4,6] It was shown previously that the ratio was inversely proportional to the severity of GO (mild 9.3:1; moderate 3.2:1; severe 1.4:1).[8] Recent DON cohorts revealed a female-to-male ratio of 2.8:1.[9-11] The association of male gender with more severe disease is well-recognized. Although the underlying pathophysiological mechanisms are not fully understood, the higher prevalence of smoking among men may at least partially explain the association.

Ethnicity

It has been suggested that certain ethnicities carry a higher risk of developing GO, but individual comparative studies showed conflicting results. A meta-analysis revealed a trend towards lower prevalence of GO among Caucasians (37%) compared with Asians (45%), but the difference was not statistically significant. The prevalence of GO stratified by continents/regions and ethnicities is shown in Fig. 72.1.[2]

Smoking

Cigarette smoking is associated with multiple negative effects on GO[12]: (1) the prevalence of smokers is higher in GD patients with GO compared with those without GO; (2) smokers suffer from more severe GO; (3) smoking renders immunomodulatory treatment of GO less effective; and (4) smoking increases the risk of progression or *de novo* development of GO after RAI. Smoking cessation may reduce the occurrence of proptosis and diplopia.

High Thyroid Stimulating Hormone Receptor Antibody Titer

In a multicenter prospective observational study, EUGOGO identified four independent predictors of GO development in patients with Graves hyperthyroidism, namely: current smoking, longer duration of hyperthyroid symptoms, clinical activity score (CAS) of at least 1, and high thyroidstimulating hormone receptor antibody (TSH-R–Ab) titer.[13] The relevance and utility of TSH-R–Ab in the management of GO will be discussed in the section "Thyroid Stimulating Hormone Receptor Antibody."

Radioactive Iodine Therapy

RAI is associated with a 4-fold increase in the risk of progression or *de novo* development of GO, affecting approximately 20% of those who receive RAI for Graves hyperthyroidism. The risk of post-RAI severe GO is around 7%.[14] Several factors further amplify the risk, including pre-existing GO, active smoking, high TSH-R–Ab titer, recent-onset hyperthyroidism, active or severe hyperthyroidism, and delayed correction of post-RAI hypothyroidism.[12,15] A recent Italian study suggested that patients with GD with a duration of more than 5 years were not at increased risk of exacerbation or *de novo* occurrence of GO. RAI results in a 4- to 5-fold increase in TSH-R–Ab titer, which peaks by 3 months post-treatment and then gradually returns to pre-treatment level by 1 year. This transient post-RAI deterioration in thyroid autoimmunity is believed to drive the progression or *de novo* development of GO.[16,17]

Cholesterol Level and Statin Use

In a cross-sectional study,[18] total cholesterol (TC) and LDL cholesterol (LDL-C) significantly correlated with the occurrence of GO in GD patients of relatively recent onset. TC above 191 mg/dL and LDL-C above 118.4 mg/dL were associated with a 47% and 28% increased risk of developing GO, respectively. A correlation was also found between CAS and TC/LDL-C levels. Another retrospective study confirmed the observation that TC/LDL-C levels were higher in GD patients with GO than those without GO.[19] Moreover, statin use for more than 60 days was associated with 40% decreased risk of GO in GD patients, whereas nonstatin cholesterol-lowering medications were not protective.[20] Future prospective trials will confirm whether statin use will impact on the occurrence and natural history of GO, as well as elucidate the mechanisms (e.g., cholesterol lowering, direct immunomodulatory actions) through which statin exerts its beneficial effects.[21]

KEY POINTS

- High thyroid stimulating hormone receptor antibody titer predicts the occurrence of Graves' orbitopathy. Smoking and radioactive iodine therapy are important modifiable risk factors. Older age and male gender are associated with more severe disease. Lower cholesterol levels and statin use are associated with lower risk of developing Graves' orbitopathy.

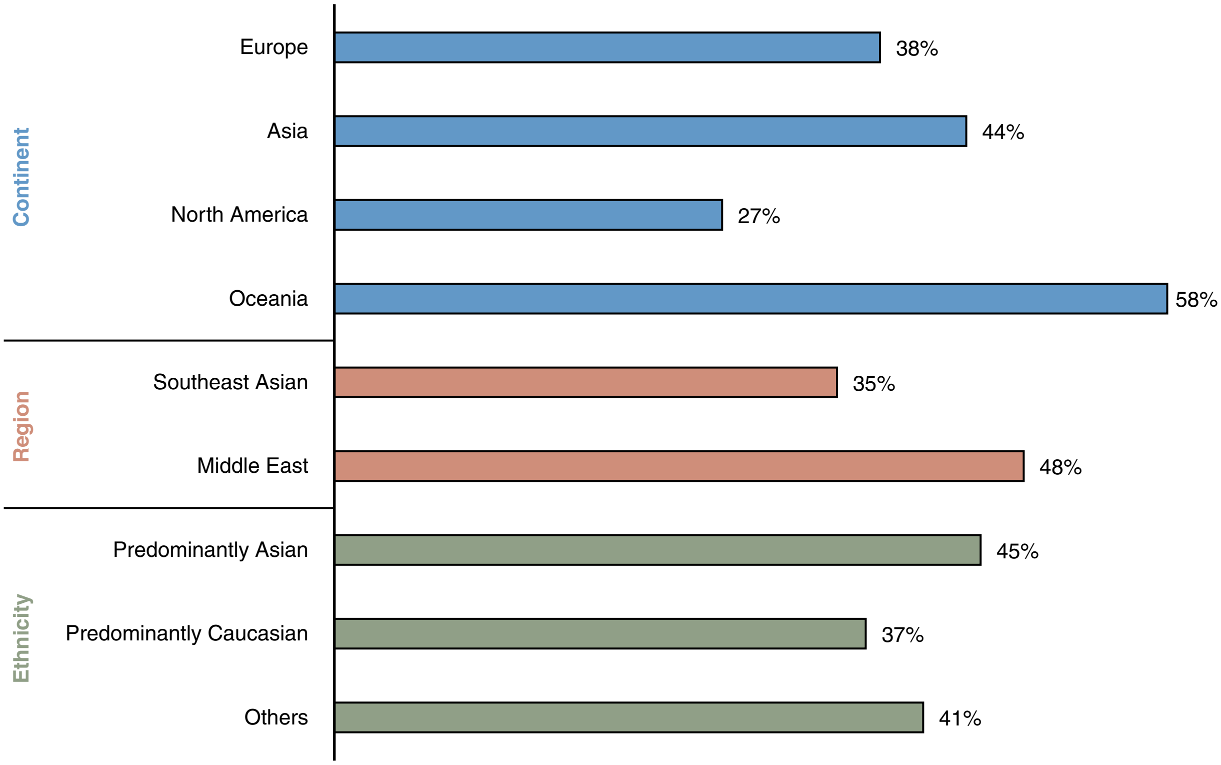


Fig. 72.1 Prevalence of Graves' orbitopathy in Patients with Graves Disease.

PATHOGENESIS

Key Pathological Changes in Graves' Orbitopathy

GO is characterized by three key pathological changes: (1) orbital inflammatory infiltration; (2) expansion of the adipose tissue within the connective tissue in extraocular muscles (sparing muscle fibers) and in other orbital adipose/connective tissues (i.e., *de novo* adipogenesis); and (3) increased production of hydrophilic glycosaminoglycans (GAG; mainly hyaluronan [HA]) by orbital fibroblasts (OFs), resulting in edema and further increase in the volume of various orbital tissues. These processes carry adverse mechanical consequences that explain most of the clinical manifestations of GO.

Thyroid Stimulating Hormone Receptor as the Putative Autoantigen

The close relationship between Graves hyperthyroidism and its associated orbitopathy suggests that thyroid stimulating hormone receptor (TSH-R) is the autoantigen shared by these two autoimmune disorders. Multiple immunohistochemical studies have confirmed the presence of TSH-R expression in orbital tissues (especially OFs and preadipocytes) of GO patients but not healthy individuals.[22] Similar TSH-R immunoreactivity was also demonstrated in pretibial fibroblasts from patients with Graves dermopathy, which is a less common extrathyroidal manifestation of GD.[23] The degree of TSH-R expression in GO orbital tissues correlated positively with serum TSH-R–Ab levels and negatively with disease duration.[22] In turn, serum TSH-R–Ab is a useful biomarker of disease activity and severity in GO (see "Thyroid Stimulating Hormone Antibody").

Immune Dysregulation

Pattern of Orbital Inflammatory Infiltration. The mononuclear cell infiltrates comprise predominantly T-cells (especially CD4+, also CD8+/CD45RO+ and CD45RB+) and monocytes/macrophages, consistent with the notion that GO is primarily a T-cell–mediated disease. Infiltration of mast cells and dendritic cells has been occasionally reported. In contrast, B-cells were much less prominent or even absent in certain immunohistochemical studies.[22] The magnitude of inflammatory infiltration correlated with the disease status.[22] The early or active phase of GO (when compared with the late or inactive phase) was associated with more prominent cellular infiltration by T-cells (especially CD4+), monocytes/macrophages, and major histocompatibility complex (MHC) class II+ cells (human leukocyte antigen [HLA]-DR, -DP, -DQ, or -DX). Severe GO (when compared with mild GO) was associated with more prominent T-cell infiltration, whereas CAS was positively correlated with the degree of lymphocyte or macrophage infiltration.

Interaction of T- and B-Cells in Graves' Orbitopathy. The exact mechanisms causing disruption of immune tolerance to self antigens (especially TSH-R) in GO are yet to be elucidated. Antigen-presenting cells (e.g., B-cells, monocytes/macrophages, dendritic cells) recognize and present TSH-R peptides through MHC class II molecules to T-cell receptors of autoreactive T helper (Th) cells. Upregulation of HLA-DR (a MHC class II molecule) expression was evident in GO orbital tissues compared with healthy controls.[22] Antigen recognition constitutes the first signal for T- or B-cell activation, and the second signal by costimulation is required for their full activation. The binding of CD40 ligand (CD40L or CD154) on the T-cell surface to CD40 on the B-cell surface activates B-cells. The CD40–CD40L interaction is followed by CD28 expression on the T-cell surface and B7 expression on the B-cell surface, and the B7–CD28 interaction leads to full activation of T-cells. The CD40–CD40L costimulatory pathway is essential for T-cell activation and the T-cell–dependent humoral immune response, playing an important role in activation of autoreactive B-cells and subsequent production of pathogenic TSH-R–Ab by plasma cells. The observations that rituximab resulted in both B- and T-cell depletion on orbital tissue immunohistochemistry and that the efficacy of rituximab in GO was not consistently associated with a decline in serum TSH-R–Ab level suggest that B-cell–T-cell interaction and antigen presentation are important components during the early phase of the pathogenesis.[22,24] This is in keeping with the impression from clinical trials that patients with GO of recent onset (instead of those with longer disease duration) benefited from rituximab therapy.[25]

Roles of B- and T-Cells in Graves' Orbitopathy. The principal role of B-cells in GO is believed to be antigen presentation to Th cells, which leads to activation of both types of lymphocytes through costimulation, resulting in production of autoantibodies and various proinflammatory cytokines as well as adhesion molecules. These stimuli not only amplify and maintain orbital inflammation, but also promote activation, proliferation, and differentiation of OFs. The consequences are GAG deposition and tissue edema, *de novo* adipogenesis, orbital tissue remodeling, and ultimately fibrosis.

CD4+ T-cells, which dominate orbital inflammatory infiltration, are classified into the following subtypes: Th1, Th2, Th17, and regulatory T (Treg) cells. Th1 cells mediate the cell-mediated immune response and produce interleukin (IL)-1β, IL-2, interferon (IFN)-γ, and tumor necrosis factor (TNF)-α. These proinflammatory cytokines result in OF activation and proliferation, as well as enhanced GAG production. Th2 cells, with a cytokine profile comprising IL-4, IL-5, IL-10, and IL-13, mediate humoral immunity and autoantibody production, as well as tissue remodeling and fibrosis. It is postulated that Th1 immunity is important in initiating GO development during the early stage of the disease, whereas Th2 immunity is responsible for the chronic phase of GO.[23 26]

Th17 is a recently identified CD4+ T-cell subset that is implicated in the pathogenesis of GO owing to its proinflammatory and profibrotic properties. Recent evidence suggests that there is a close interplay between Th17 cells and OFs.[26] IL-17A, the signature cytokine of Th17 cells, promoted the secretion of proinflammatory cytokines (IL-6, IL-8, monocyte chemotactic factor-1, regulated on activation normal T-cell expressed and secreted [RANTES], TNF-α, granulocyte macrophage colony stimulating factor), costimulatory molecules (CD40, MHC class II), and adhesion molecules (vascular cell adhesion molecule-1 [i.e., CD106], intracellular adhesion molecule-1 [ICAM-1, i.e., CD54]) by OF. IL-17A also enhanced profibrotic functions of OFs, including extracellular matrix deposition, transforming growth factor-β (TGF-β)-induced differentiation into myofibroblasts, and subsequent fibrosis. OFs in turn triggered Th17 differentiation and supported Th17 functions via prostaglandin E2 (PGE2). In contrast, IL-17A suppressed adipogenesis. Th17 cells also promoted proinflammatory and antigen-presenting functions of GO fibrocytes, which enhanced Th17 cell function and recruitment.[26]

Treg cells are important in controlling autoimmunity, and higher levels in peripheral circulation are associated with favorable treatment response and clinical improvement in GO. The importance of T-cell immunity has been reproduced in a recent mouse model of GO.[27] When administered an adenovirus expressing the human TSH-R A subunit, seven of 10 mice developed orbital changes associated with GO. Increased proportions of Th1 cells and decreased proportions of Th2/Treg cells in the splenocytes of GO mice, as detected by flow cytometry, were accompanied by the same change in CD4+ T-cell subgroups as detected by orbital immunohistochemical staining. Furthermore, there was an upregulation of genes involved in T-cell receptor signaling, proliferation, adhesion, inflammation, and cytotoxicity in GO mice.

Orbital Fibroblasts as Key Effector Cells in Graves' Orbitopathy

OFs mediate most of the pathological changes in GO under the influence of various stimuli, including cytokines, growth factors, CD40–CD40L interaction, and TSH-R–insulin-like growth factor-1 receptor (IGF-1R) crosstalk.[23] IFN-γ and IL-1β stimulate OFs to produce proinflammatory cytokines and chemoattractants (e.g., C-X-C motif ligand-9 (CXCL-9), CXCL-10, CXCL-11, C-C motif ligand-2 (CCL2), CCL5, IL-6, IL-8, IL-16, RANTES), leading to migration and infiltration of lymphocytes, monocytes, and neutrophils. The interaction between CD40L on T-cells and CD40 on OFs results in activation and proliferation of OFs, as well as the release of inflammatory molecules (e.g., IL-1α, IL-6, IL-8, CCL2, PGE2). PGE2 potentiates B-cell maturation, stimulates IL-6 synthesis by OFs and activates mast cells. Other cytokines (e.g., leukoregulin, IL-1β from macrophages or OFs, IFN-γ from activated T-cells) can also stimulate OFs to produce PGE2. B-cell activating factor (BAFF) plays an important role in promoting B-cell survival and maturation, as well as inducing immunoglobulin class-switching recombination. TNFα and IFN-γ enhance BAFF expression by OFs. IL-1α, IL-1β, TNF-α, IFN-γ, and CD40–CD40L interaction enhance expression of ICAM-1 on OFs, which augments T-cell recruitment during the active phase of the disease.

Platelet-derived growth factors (PDGFs) were found to be overexpressed by monocytes, macrophages, and mast cells (but not T- or B-cells) in orbital tissues of GO patients.[22] Different PDGF isoforms were capable of stimulating proliferation, HA and cytokine/chemokine (e.g., IL-6, IL-8, IL-16, RANTES, CCL2, CCL5, CCL7) production, and TSH-R expression by OFs. Of these PDGF isoforms, PDGF-BB exhibits the strongest OF-activating effects, which is likely related to its ability to bind both α and β chains of the PDGF receptor.[28]

The overproduction of hydrophilic HA by OFs was stimulated by TSH-R/IGF-1R activation (discussed in more detail below) and various inflammatory mediators, such as IL-1, TNF-α, IFN-γ, TGF-β, IGF-1, PDGF, and prostaglandins, through upregulation of HA synthases (the major isoform being HA synthase-2 [HAS2]). *In vitro* studies reported undetectable hyaluronidase activity in GO OFs. The imbalance between overproduction and impaired degradation probably explains the HA accumulation in OG orbital tissues.[8,23]

OFs can be classified into Thy1- (CD90-) and Thy1+ subpopulations based on the presence or absence of CD90 glycoprotein expression.[23] Thy1- OFs (preadipocytes) possess a strong capability to differentiate into mature adipocytes. Peroxisome proliferator-activated receptor-γ (PPARγ) is highly expressed in Thy1- OFs. PPARγ agonists (e.g., rosiglitazone) enhanced adipogenesis and expression of TSH-R in Thy1- OFs. In contrast, they inhibited orbital inflammation and GAG production. IL-1β, IL-6, and proadipogenic prostaglandins (PPARγ ligands) also stimulated Thy1- OFs to undergo adipogenesis, while certain Th1 cytokines (e.g., TNFα, IFNγ) inhibited the process. On the other hand, Thy1+ OFs are able to differentiate into myofibroblasts under stimulation by TGF-β (a type of Th2 cytokine). Myofibroblasts play an essential role in mediating deposition of extracellular matrix (e.g., collagen) and eventually tissue fibrosis. It was shown that Thy1+ OFs suppressed adipocytic differentiation of Thy1- OFs by producing antiadipocytic factors. Therefore, the balance and relative proportion of Thy1- and Thy1+ OFs modulate the presentation of tissue remodeling in GO.

Signaling Pathways Mediating Hyaluronan Production and Adipogenesis (Fig. 72.2)

TSH-R activation of GO OFs switched on G protein Gαs, which triggered adenylate cyclase activation and generation of cyclic adenosine monophosphate (cAMP), leading to activation of cAMP dependent protein kinase A (PKA). Direct phosphorylation of cAMP-response element binding protein (CREB), a transcription factor, was catalyzed by activated PKA. CREB then bound to the HAS1 and HAS2 promoters, leading to increased HA production. On the other hand, stimulation of IGF-1R activated phosphatidylinositol 3-kinase 1A (PI3K1A)-Akt/PKB-mammalian target of rapamycin complex 1 (mTORC1) pathway.[29] PI3K signaling specifically regulated HA production via HAS2, while mTORC1 was essential for adipogenesis.[30] Forkhead transcription factors, forkhead box O (FOXO) subtypes, were tightly regulated via Akt within the PI3K signaling cascade. Enhancing FOXO3a expression significantly decreased HAS2/HA production in GO OFs, whereas FOXO1 repressed adipogenesis. FOXOs were also able to counteract TSH-R–activated HAS2/HA production and block adipogenesis in GO OFs.[31] Enhanced TSH-R expression, which accompanied adipocytic differentiation, may thus be indirectly attenuated through inhibition of the PI3K/mTORC1 pathway. Therefore, FOXOs probably serve as convergence points for TSH-R and IGF-1R signaling pathways in GO.[29]

Orbital Infiltration of CD34+ Fibrocytes

Fibrocytes are CD34+ bone marrow–derived progenitor cells of the monocyte lineage. They normally circulate at low concentrations but become more abundant in GD patients. It is postulated that they infiltrate the orbits and initiate the development of GO. Fibrocytes present thyroid antigens (e.g., TSH-R, thyroglobulin, thyroperoxidase, sodium-iodide symporter) to antigen-specific autoreactive T-cells, which in turn activate B-cells and subsequent autoantibody generation. Fibrocytes differentiate into CD34+ OFs, which can further differentiate into myofibroblasts and adipocytes. Although only CD34- OFs are present in normal orbits, a mixture of CD34+ OFs (~50% of the whole population of OF) and the residential CD34- OFs coexist in GO orbits. Both fibrocytes and GO OFs express high levels of TSH-R/IGF1-R and produce various inflammatory molecules (e.g., proinflammatory cytokines, adhesion molecules, growth factors, etc.), leading to mononuclear cell infiltration. The importance of TSH-R/IGF-1R will be discussed below. Bidirectional stimulation between lymphocytes and GO OFs results in perpetuation of orbital inflammation and tissue remodeling. Interestingly, recent studies suggested that Slit2, a glycoprotein produced by residential CD34- OFs, appeared to downregulate the inflammatory phenotype of CD34+ OFs, and thus potentially attenuate the manifestations of GO.[32-34]

Thyroid Stimulating Hormone Receptor and Insulin-Like Growth Factor-1 Receptor Crosstalk (Fig. 72.3)

Multiple immunohistochemical studies have confirmed the presence of TSH-R expression in orbital tissues of GO patients but not in healthy individuals,[22] suggesting that TSH-R is one of the primary autoantigens in GO. Activation of TSH-R stimulated production of proinflammatory cytokines (e.g., IL-6, TNF-α) and GAG by GO OFs/fibrocytes. There was a positive correlation between TSH-R mRNA levels in GO orbital tissues and disease activity.[33,35]

IGF-1R was overexpressed in GO OFs and lymphocytes. Accumulating evidence suggests that IGF-1R forms a physical and functional complex with TSH-R and that its activity is important for mediating TSH-R downstream signaling[36,37]:

1. TSH-R and IGF-1R were in close proximity in a signalosome, in which β-arrestin 1 acted as a scaffold.
2. Simultaneous activation of TSH-R and IGF-1R (by TSH and IGF-1, respectively) synergistically increased HA secretion by GO OFs.
3. IGF-1–induced HA secretion was effectively inhibited by linsitinib (IGF-1R kinase inhibitor) or IGF-1R blocking antibodies (AF305 and 1H7). However, only linsitinib and 1H7 (but not AF305) suppressed

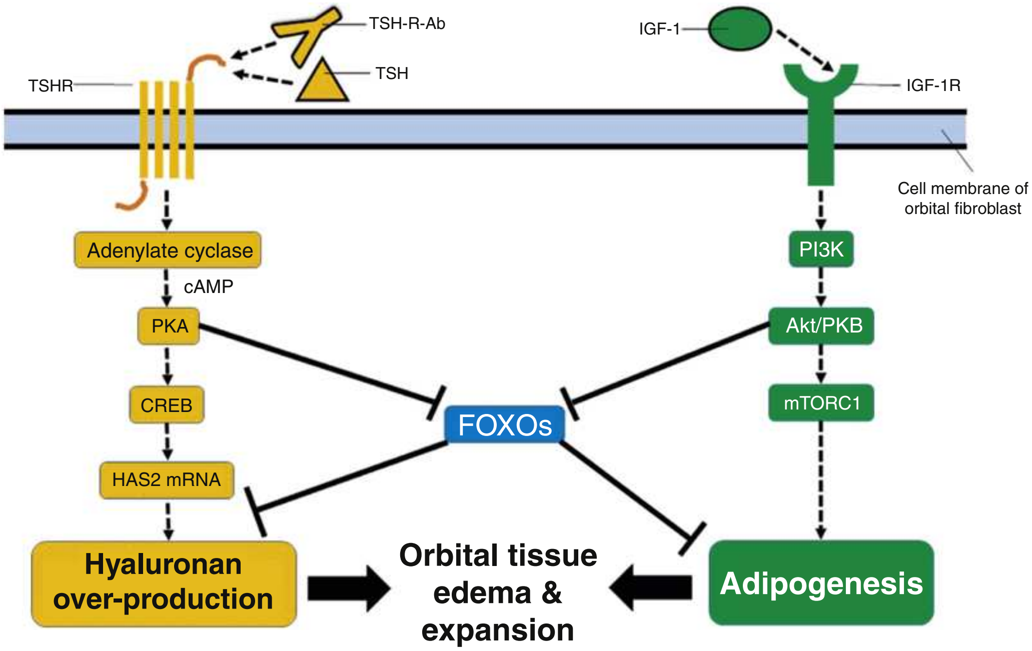


Fig. 72.2 Signaling Pathways of Thyroid-Stimulating Hormone Receptor (TSH-R) and Insulin-Like Growth Factor-1 Receptor (IGF-1R) in Graves' Orbitopathy. TSH-R and IGF-1R activation directly stimulate hyaluronan production and adipogenesis in orbital fibroblasts, respectively. Both pathways can also activate orbital fibroblasts through suppressing expression of FOXOs, which potentially serve as the convergence point of TSH-R/IGF-1R signaling pathways. *Akt/PKB*, Protein kinase B; *cAMP*, cyclic adenosine monophosphate; *CREB*, cAMP response element–binding protein; *FOXOs*, forkhead box O family of transcription factors; *HAS2*, hyaluronan synthase-2; *IGF-1*, insulin-like growth factor-1; *IGF-1R*, insulin-like growth factor-1 receptor; *mTORC1*, mammalian target of rapamycin complex 1; *PI3K*, phosphatidylinositol 3-kinase; *PKA*, protein kinase A; *TSH*, thyroidstimulating hormone; *TSH-R–Ab*, thyroidstimulating hormone receptor antibody.

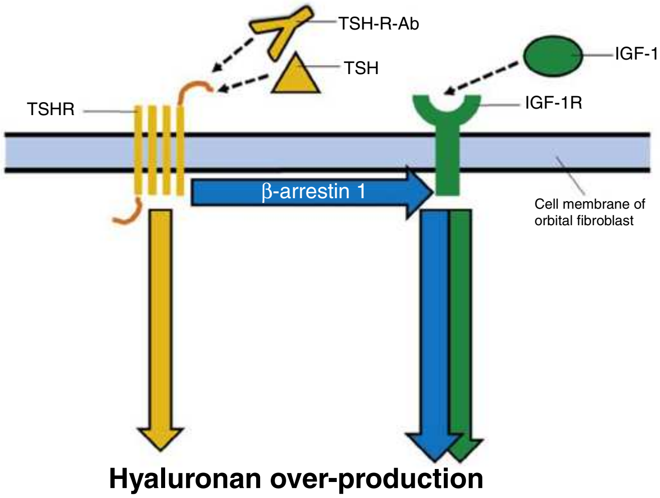

TSH-R-Ab
TSH
TSHR
IGF-1
IGF-1R
β-arrestin 1
Cell membrane of orbital fibroblast
Hyaluronan over-production

Fig. 72.3 Thyroidstimulating Hormone Receptor (TSH-R)/Insulin-Like Growth Factor-1 Receptor (IGF-1R) Crosstalk. TSH-R and IGF-1R are in close proximity within a signalosome, with β-arrestin 1 acting as a scaffold. TSH-R stimulation activates two signal transduction pathways, one being IGF-1R-independent *(orange arrow)* and the other IGF-1R-dependent *(blue arrow)*. Therefore, both TSH-R and IGF-1R pathways combine through crosstalk, resulting in more potent stimulation of hyaluronan production. *IGF-1*, Insulin-like growth factor-1; *IGF-1R*, insulin-like growth factor-1 receptor; *TSH*, thyroidstimulating hormone; *TSH-R–Ab*, thyroidstimulating hormone receptor antibody.

HA secretion by GO OFs induced by M22 (TSH-R stimulatory monoclonal antibody) or immunoglobulins from GO patients (GO-Ig).

4. TSH-R antagonist (ANTAG3) completely suppressed M22-induced HA secretion by GO OFs, regardless of M22 concentration. In contrast, linsitinib and 1H7 fully antagonized HA secretion induced by M22 at low concentration, but their efficacy diminished at high concentration of M22. Individually, ANTAG3, linsitinib, and 1H7 (all at half maximally effective inhibitory concentrations) incompletely inhibited GO-Ig–induced HA secretion, while the combination of ANTAG3 and linsitinib/1H7 additively suppressed HA secretion.

In summary, the stimulation of TSH-R activates two signal transduction pathways, one being IGF-1R–independent and the other IGF-1R–dependent (i.e., TSH-R/IGF-1R crosstalk pathway). TSH-R– and IGF-1R–mediated pathways cooperate through crosstalk, resulting in more intense activation of OFs. IGF-1R autoantibodies were present in both normal individuals and GD patients with or without GO. Although GO-Ig was able to displace IGF-1 from the binding sites of IGF-1R on the surface of GO OFs, the current consensus is that stimulatory IGF-1R autoantibodies are not directly involved in the pathogenesis of GO[38]

Genetic Susceptibility

The etiology of GD has a strong genetic component, as evidenced by greater concordance rate among identical twins compared with nonidentical twins and high sibling risk ratios among family members of GD patients. Several genetic loci have been shown to be associated with GD susceptibility, including *HLA* genes (e.g., HLA*DR3 and DQA*10501 haplotypes in Caucasians), immune-related genes (e.g., *CTLA-4, CD25, CD40, PTPN22, FOXP3*), and thyroid-specific genes encoding thyroglobulin and TSH-R.[39] In recent years, multiple genetic polymorphisms have also been reported to be associated with GO, involving *CTLA-4* (SNP49 polymorphism),[40] adipocyte-related immediate early genes (IEGs), and immunomodulatory genes (cysteine-rich, angiogenic inducer 61, zinc finger protein 36, C3H type, homolog mouse, and stearoyl-coenzyme A desaturase),[41] interleukin-1A (rs1800587, c.-889C>T),[42] *TSH-R* (rs179247 and rs12885526),[43] FC receptor-like-3 (FCRL3_3C, FCRL3_5C, FCRL3_6A),[44],methylenetetrahydrofolate reductase gene (TT genotype),[45] tumor necrosis factor super family-15 ((rs3810936 and rs4979462 polymorphisms),[46] and *HLA* genes.[47]

Fig. 72.4 summarizes the key components involved in the pathogenesis of GO.

KEY POINTS

- Thyroidstimulating hormone receptor (TSH-R) is believed to be the primary autoantigen in Graves' orbitopathy (GO). The loss of immune tolerance leads to orbital inflammatory infiltration and activation of orbital fibroblasts, resulting in perpetuation of inflammation, excessive glycosaminoglycan production and subsequent tissue edema, *de novo* adipogenesis, and ultimately tissue remodeling and fibrosis. TSH-R–insulin-like growth factor-1 receptor crosstalk and their signaling pathways are central to the pathophysiology of GO.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

GO can present with a wide spectrum of clinical features, from non-specific to pathognomonic symptoms and signs, and from trivial complaints to disabling and sight-threatening complications (Table 72.1).

The diagnosis of GO is usually established clinically based on compatible clinical features, while thyroid function test, TSH-R–Ab assessment, and orbital imaging will provide further supportive evidence. As GO has significant impact on patients' quality of life (QoL), QoL assessment forms an integral part in the evaluation of GO.

Clinical Evaluation

EUGOGO recommends the assessment of both activity and severity of GO according to standardized criteria, and subsequent classification as inactive or active, and mild, moderate-to-severe, or sight-threatening.[12] Accurate assessment is essential in determining the most appropriate treatment strategies for GO patients. CAS is the most well-established tool used to assess disease activity in both daily practice and clinical trial settings (Table 72.2). A CAS value of 3/7 or more or 4/10 or more indicates active GO, which necessitates immunomodulatory treatment. Disease severity is determined by five components: soft tissue involvement as reflected by 7-point CAS value (excluding pain parameters), upper lid retraction, diplopia, proptosis, and evidence of sight-threatening complications (Table 72.3).

Asymmetrical and Unilateral Graves' Orbitopathy. GO is considered a generalized orbital disease. Asymmetrical or unilateral involvement can occur, although there is no established definition of this clinical entity. In a recent prospective cross-sectional multicenter study involving new referrals to EUGOGO centers,[48] unilateral GO was defined as one or more clinical features of GO in one eye only. Asymmetrical GO was defined as bilateral disease with one or more of the following features: (1) difference in exophthalmometer readings of 2 mm or more; (2) difference in palpebral aperture of 2 mm or more; (3) difference in eyelid swelling by one grade or more; (4) difference in eyelid erythema by one grade or more; (5) difference in conjunctival redness by

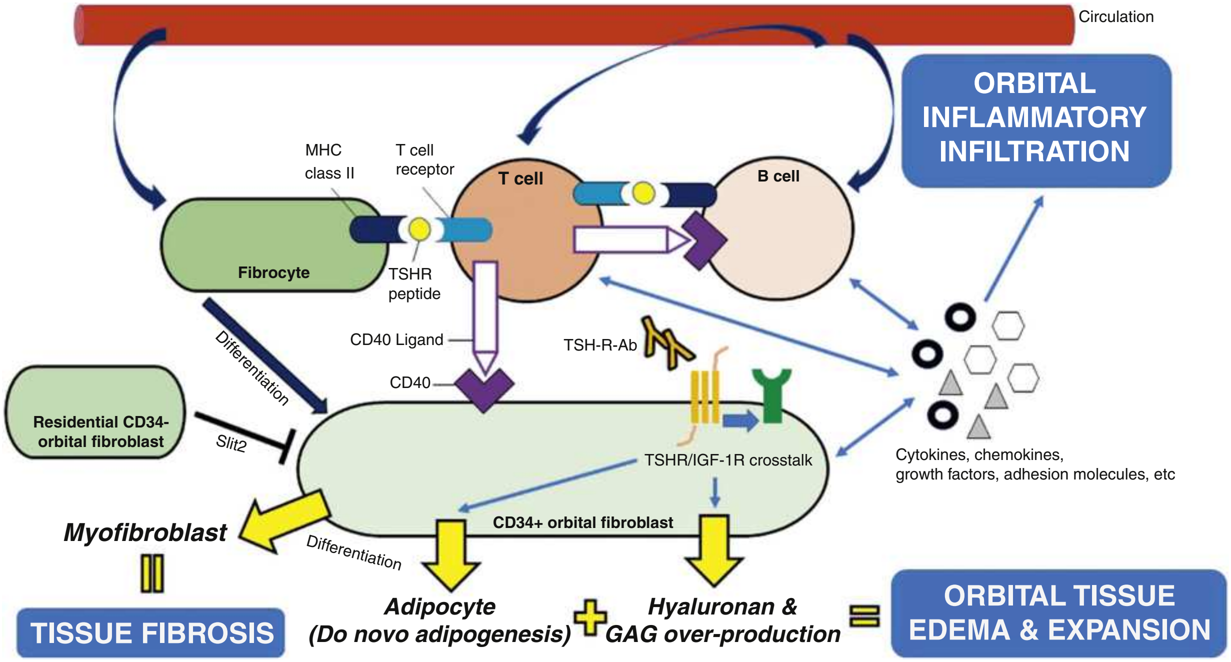

Circulation
ORBITAL INFLAMMATORY INFILTRATION
MHC class II
T cell receptor
T cell
B cell
Fibrocyte
TSHR peptide
CD40 Ligand
TSH-R-Ab
CD40
Differentiation
Residential CD34- orbital fibroblast
Slit2
TSHR/IGF-1R crosstalk
Cytokines, chemokines, growth factors, adhesion molecules, etc
Myofibroblast
Differentiation
CD34+ orbital fibroblast
TISSUE FIBROSIS
Adipocyte (Do novo adipogenesis)
Hyaluronan & GAG over-production
ORBITAL TISSUE EDEMA & EXPANSION

Fig. 72.4 **Pathogenesis of Graves' Orbitopathy (GO)—an Overview. TSH-R is Believed to be the Primary Autoantigen in GO.** The loss of immune tolerance generates autoreactive T- and B-cells, which are activated by costimulation and cytokines. Orbital fibroblasts becomes activated by various stimuli (cytokines/growth factors, TSH-R/IGF-1R signaling and crosstalk after stimulation by TSH-R–Ab, and CD40–CD40 ligand interaction with T-cells). Orbital fibroblasts play a central role in mediating the key pathological changes in GO, including orbital inflammatory infiltration, *de novo* adipogenesis, hyaluronan overproduction, with resultant tissue edema, expansion, remodeling, and fibrosis. *IGF-1R*, Insulin-like growth factor-1 receptor; *MHC*, major histocompatibility complex; *TSHR*, thyroid-stimulating hormone receptor; *TSH-R–Ab*, thyroidstimulating hormone receptor antibody; *GAG*, glycosaminoglycan.

TABLE 72.1 Clinical Features of Graves' Orbitopathy

Clinical Features	Pathophysiology and Clinical Significance
• Gritty or foreign body sensation, light sensitivity (photophobia), excess tearing	• Ocular surface irritation resulting from increased evaporative loss and/or impaired tear film formation because of proptosis or eyelid abnormalities • May mimic other common ocular conditions, e.g., allergic conjunctivitis, dry eye disease
• Eyelid swelling and redness • Conjunctival swelling (chemosis) and redness • Caruncle or plica swelling • Gaze-evoked pain or spontaneous retrobulbar pain • Elevated intraocular pressure	• Reflect orbital inflammation and congestion
• Upper eyelid retraction	• Sympathetic activation of Muller's muscle • Synkinetic activity of levator palpebrae superioris with superior rectus that tries to overcome the tight inferior rectus • Inflammation, degeneration, and scarring of eyelid structures
• Lower eyelid retraction/displacement	• Correlates with the degree of proptosis • Lower lid retractors apparently unaffected in Graves' orbitopathy
• Lid lag	• Defined as a static phenomenon in which the upper eyelid assumes and maintains a higher position (relative to its position in primary gaze) with the eye in downgaze
• Von Graefe's sign	• Defined as a dynamic phenomenon in which the upper eyelid fails to descend smoothly and lags behind the eyeball during the course of downgaze
• Lagophthalmos	• Incomplete eyelid closure because of eyelid abnormalities and significant proptosis • Predisposes to exposure keratopathy, especially when Bell's phenomenon is defective owing to tight inferior rectus
• Diplopia and squint	• The proliferation of orbital fibroblasts (*de novo* adipogenesis, secretion and accumulation of hydrophilic glycosaminoglycans) and inflammatory infiltration lead to enlargement of extraocular muscles, resulting in restrictive (instead of paralytic) strabismus
• Proptosis	• Expanded volume of retrobulbar structures pushes the eyeball forward • Globe subluxation may occur in extreme proptosis
• Visual loss	• Watch out for sight-threatening Graves' orbitopathy

TABLE 72.2 Clinical Activity Score

Parameters	7-Point CAS	10-Point CAS
1. Spontaneous retrobulbar pain 2. Pain on attempted upward/downward/side gaze 3. Redness of eyelids 4. Redness of conjunctiva 5. Swelling of eyelids 6. Swelling of conjunctiva (chemosis) 7. Swelling of caruncle or plica 8. Increase of ≥2 mm in proptosis 9. Decrease in uniocular excursion in any one direction of ≥8 degrees 10. Decrease in visual acuity equivalent to 1 Snellen line	• Applicable to first clinical assessment or during subsequent follow-ups (parameters 1-7)	• Applicable during subsequent follow-ups (parameters 1-10)

CAS, Clinical activity score.

- One point is given to each of the seven signs and symptoms in the 7-point CAS during initial assessment.
- On subsequent follow-up, either the 7-point or the 10-point CAS (which includes three additional parameters, namely deterioration in ocular dysmotility, proptosis, or visual acuity) can be used.

one grade or more; (6) presence of unilateral DON. Some 58% of GO patients had symmetrical disease, whereas 31% and 11% had asymmetrical and unilateral disease, respectively. Compared with the symmetrical and unilateral groups, the asymmetrical group had an older mean age, lower female-to-male ratio, and more active and severe disease. Whereas asymmetry is a potential indicator of more active and severe GO than other clinical variants, unilaterality is associated with less severe disease.

Dysthyroid Optic Neuropathy. DON is the most common form of sight-threatening GO. Direct compression of swollen extraocular muscles on the optic nerve at orbital apex is believed to be the pathogenic mechanism in over 90% of cases. However, some propose that optic nerve stretch in the presence of marked proptosis may explain a minority of cases, although this theory remains controversial. The most common symptom is visual blurring. Approximately 50% of affected patients have bilateral, but not necessarily symmetrical, involvement (simultaneous or sequential). Although there are no established diagnostic criteria validated for DON, the salient diagnostic features are summarized in Table 72.4.

Thyroid Function Test

Over 95% of GO cases are associated with Graves hyperthyroidism.[49,50] Although GO most commonly occurs simultaneously with or after the onset of hyperthyroidism, it precedes the onset of hyperthyroidism in 20% of patients (and most of them are diagnosed within

TABLE 72.3 Assessing Severity of Graves' Orbitopathy

	Mild Graves' Orbitopathy	Moderate-to-Severe Graves' Orbitopathy	Sight-Threatening Graves' Orbitopathy
General description	• Minimal impact on daily life • Usually present ≥1 of the following signs:	• No sight-threatening GO • Significant impact of daily life to justify the risks of immunosuppressants (if active) or surgical intervention (if inactive) • Usually present ≥2 of the following signs:	• Presence of any one of the following features: • Dysthyroid optic neuropathy • Corneal breakdown because of severe exposure • Globe subluxation • Choroidal folds • Postural visual darkening
Soft tissue involvement	Mild	Moderate or severe	
Lid retraction	<2 mm	≥2 mm	
Proptosis	<3 mm above normal for race and gender	≥3 mm above normal for race and gender	
Diplopia	Absent or intermittent	Inconstant or constant	
Corneal exposure	Responsive to lubricants		

TABLE 72.4 Diagnosis of Dysthyroid Optic Neuropathy

Clinical Features	Examples of Diagnostic Test	Frequency in Dysthyroid Optic Neuropathy	Remarks
1. Reduced BCVA	Snellen chart	• 82%	• No single BCVA threshold can exclude DON, and 20% of cases have BCVA >0.67
2. Reduced color vision	Ishihara test	• 83%	• One of the most specific features • Can be present even if BCVA appears normal
3. RAPD	Swinging flashlight test	• 44%	• Possible only in unilateral DON or bilateral DON with significant asymmetry
4. Papilloedema	Fundoscopy	• 40%	• One of the most specific features • Optic disc pallor (present only in ~4%) may be observed in longstanding severe optic nerve dysfunction
5. Visual field defect	Automated perimetry	• 64%	• Most common patterns include central scotoma, an arcuate or altitudinal defect
6. Abnormal electrophysiology	Visual-evoked potentials	• 73% (↑ latency) • 53% (↓ amplitude)	• Helpful to exclude DON if unequivocally normal • Interference by thyroid dysfunction and lack of reference range in elderly may make interpretation difficult • Time-consuming and not widely available
7. Apical crowding on imaging	CT or MRI orbits	• 95%	• Apical crowding is quantified by assessing the degree of perineural fat effacement on coronal images at the orbital apex • Less commonly imaging shows no apical crowding but reveals optic nerve stretching ± tenting at posterior eyeball because of severe proptosis

BCVA, Best corrected visual acuity; *CT*, computed tomography; *DON*, dysthyroid optic neuropathy; *MRI*, magnetic resonance imaging; *RAPD*, relative afferent pupillary defect.

- Expert opinion suggests that DON can be diagnosed if papilloedema (after excluding alternative causes) is present or if two of the following features are present: impaired visual acuity or color vision, RAPD, and visual field defect.
- A US group developed a mathematical formula, the Columbia thyroid eye disease-compressive optic neuropathy diagnostic formula (CTD formula), which helps guide clinicians in accurately diagnosing DON. The parameters include visual acuity, color vision, RAPD, visual field defect, ocular dysmotility, and proptosis. However, this has not been prospectively validated.

6 months before the diagnosis of hyperthyroidism).[51] Approximately 5% of GO cases are associated with primary hypothyroidism because of Hashimoto thyroiditis ("hypothyroid GO," i.e., hypothyroidism diagnosed before or within 6 months after GO onset) or euthyroidism ("euthyroid GO," i.e., euthyroid before and within 6–24 months after GO onset).[49,50] Two German cohorts reported that hypothyroid or euthyroid GO was different from hyperthyroid GO in several ways:[49,52] (1) less active and milder disease; (2) more likely to be unilateral or asymmetrical; (3) lower levels of TSH-R–Ab and higher prevalence of negative thyroid autoantibodies; (4) less likely to require intravenous glucocorticoids (GCs) or orbital radiotherapy; and (5) less sick leave or work disability observed and better QoL.

Thyroid Stimulating Hormone Receptor Antibody

TSH-R–Ab refers to any type of autoantibody that interacts specifically with TSH-R. There are three functional subtypes of TSH-R–Ab: (1) TSH-R–stimulating antibody (TSAb), which stimulates unregulated thyroid growth and thyroid hormone production; (2) TSH-R–blocking antibody (TBAb), which blocks the activity of the natural ligand TSH to TSH-R; and (3) neutral TSH-R–Ab, which carries uncertain clinical significance. Serum TSH-R–Ab level is commonly quantified by competitive binding immunoassays (TSH-R–binding inhibitory immunoglobulin, TBII), which do not differentiate the functionality of different subtypes. On the other hand, cell-based bioassays measure specifically the activity of TSAb and TBAb.[35,53]

TSH-R–Ab (especially TSAb) measurement complements clinical and radiological assessment of GO, adding another dimension to diagnosis, disease monitoring, and prediction of clinical course, as well as the risk of developing GO. TSAb is more sensitive and specific than TBII in GO,[54] and overall TSH-R–Ab is a useful marker to differentiate GO from other orbital diseases.[55] TSH-R–Ab levels correlated positively with the activity and severity of GO and negatively with the duration of GO. The association of TSAb with clinical features of GO was much stronger than TBII.[54,56-58] TSAb was also associated with DON of recent onset, and it may help identify patients who require urgent treatment.[59] A recent study measured TSH-R–Ab levels (by immunoassays and cell-based bioassay) in patients with thyroidal GD only and GD plus GO.[60] The baseline undiluted samples from the two groups revealed an overlap in the ranges of TSH-R–Ab levels. However, serial dilution analyses showed that, in both hyperthyroid and euthyroid states (after methimazole), GD plus GO samples remained positive for TSH-R–Ab at high dilutions, while GD samples turned negative at low dilutions. Therefore, high TSH-R–Ab titer revealed by dilution analyses may serve as a useful parameter to predict the risk of GD patients developing GO.

Orbital Imaging

The diagnosis of GO can be established in most cases by clinical evaluation and blood tests (thyroid function test, TSH-R–Ab assessment) alone. However, orbital imaging should be considered in the following scenarios: (1) atypical presentation (e.g., euthyroid, absence of thyroid autoantibodies, unilateral or grossly asymmetrical disease)—alternative orbital diseases have to be excluded; (2) suspected DON—imaging may show apical crowding, frank optic nerve compression, or, less commonly, optic nerve stretch; (3) operative planning before orbital decompression or eye muscle surgery; and (4) inconclusive clinical evaluation (e.g., CAS, severity parameters)—imaging can provide more objective evidence of disease status and aid future monitoring.

Expansion of orbital tissues is characteristic of GO. Expansion of fat component because of *de novo* adipogenesis is the main contributor to proptosis (Fig. 72.5). Extraocular muscle enlargement most often involves the inferior rectus, followed by the medial, superior, and lateral rectus. Multiple muscles are involved in most cases (Fig. 72.6A and B). Enlarged muscle belly with sparing of the tendinous insertion is typical but not pathognomonic, as up to 6% of GO patients may show thickening of extraocular muscle tendon, and non-GO ocular myositis can present without tendon involvement. The degree of apical crowding secondary to extraocular muscle enlargement, which can be semiquantified on imaging, correlates with the risk of DON (Fig. 72.6A and B). Expanded orbital volume and venous congestion can result in dilated superior ophthalmic vein, anterior displacement of the lacrimal gland, intracranial fat prolapse through the superior orbital fissure, and bone remodeling (e.g., bowing of the lamina papyracea because of enlarged medial rectus muscle). The lacrimal gland itself can become enlarged because of venous congestion and/or inflammatory infiltration.

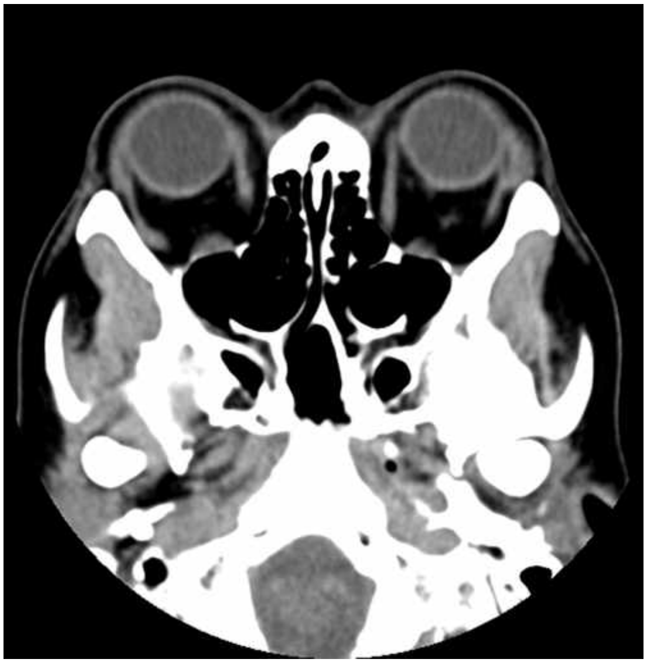


Fig. 72.5 Axial Orbital Computed Tomograpy Image of a 52-Year-Old with Inactive Graves' Orbitopathy, Showing Bilateral Proptosis and Expanded Retrobulbar Fat.

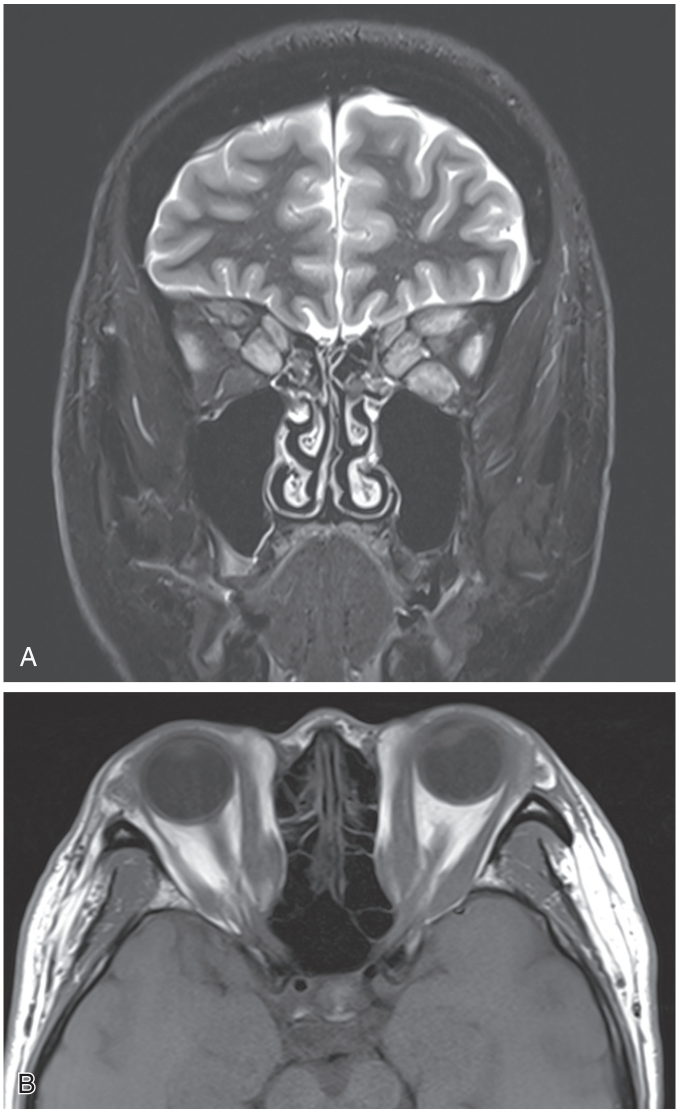


Fig. 72.6 Orbital Magnetic Resonance Imaging Scans of a 49-Year-Old With Left Dysthyroid Optic Neuropathy. **A,** Coronal T2-weighted short tau inversion recovery sequence image showing enlargement of all extraocular muscles, marked signal hyperintensity over bilateral rectus (except left inferior rectus) and superior oblique muscles, as well as bilateral apical crowding. Left orbit was more severely affected. **B,** Axial T1-weighted image, showing bilateral proptosis and apical crowding, which were more prominent over the left orbit.

Orbital computed tomography (CT) and magnetic resonance imaging (MRI) are the most commonly used imaging modalities in the evaluation of GO.[61,62] Compared with MRI, CT is quicker, more readily available, and much cheaper. CT provides simultaneous assessment of both bony and soft tissue structures, although it is less preferred in serial monitoring because of the radiation burden. In contrast, MRI offers superior resolution of orbital soft tissue, which is invaluable in the evaluation of disease activity and in the differential diagnoses of non-GO orbital diseases. Multiple MRI parameters, including short tau inversion recovery (STIR) sequence signal intensity ratio (SIR), T2-SIR, T1-SIR postgadolinium contrast, and apparent diffusion coefficient obtained from diffusion-weighted imaging, measure water content and vascularity of extraocular muscles (STIR-SIR of orbital fat as well), thus reflecting the degree of inflammatory edema[63-65] (Figs. 72.6A and B and Figs. 72.7). These measurements have been shown to correlate with CAS and response to treatment. They will be particularly helpful when CAS does not reliably reflect the actual disease activity (e.g., in patients with long disease duration,[66] darker complexion, or thicker skin).

Optical coherence tomography evaluates orbital structures affected by GO (e.g., retinal nerve fiber layer or choroidal thickness, optic nerve blood vessel density, size of rectus muscles), and the measurements may provide additional information regarding the disease status. Technetium-based scintigraphy is a novel modality that can potentially diagnose GO, assess disease activity and, reflect treatment response.[67]

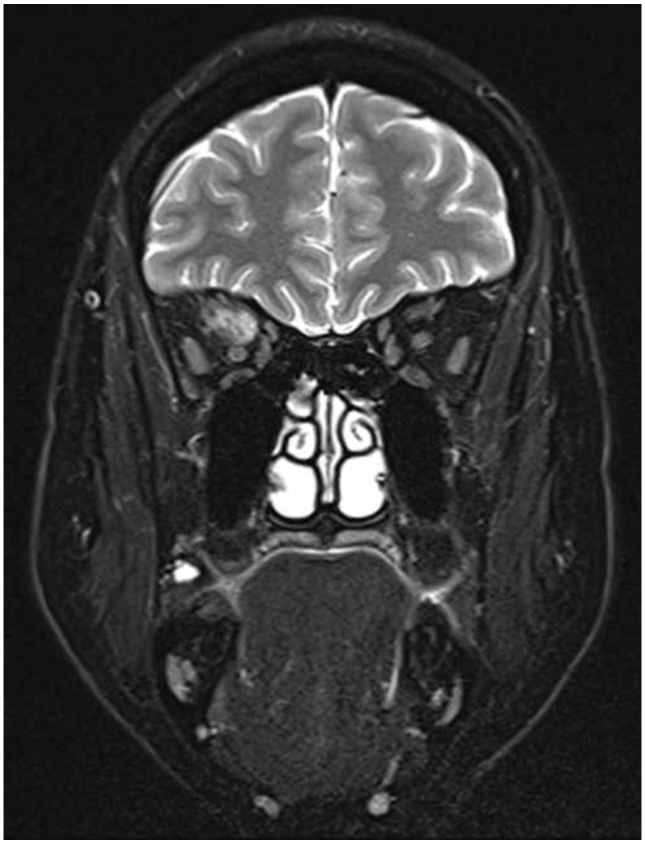


Fig. 72.7 Coronal T2-Weighted Orbital Magnetic Resonance Imaging of a 50-Year-Old With Active Moderate-To-Severe Graves' Orbitopathy, Revealing Markedly Asymmetrical Involvement With Isolated Enlargement and Signal Hyperintensity of Right Superior Rectus, Causing Downward Displacement of Right Optic Nerve.

Quality of Life Assessment

Significant impairment of QoL and psychological well-being has been well-recognized among GO patients, and the impact is evident across the whole disease spectrum. Patients with GO are more likely to require psychotherapy, more prone to develop mental features suggestive of underlying neuropsychiatric disorders, and at higher risk of death by suicide.[68] Clinical measures do not adequately predict physical and psychosocial functioning in daily life. Hence, EUGOGO recommends the use of a disease-specific questionnaire, Graves' Orbitopathy Quality of Life (GO-QoL), in routine clinical practice to assess and monitor the impact of the disease and its treatment on QoL and psychosocial well-being.[12] A recent systematic review concluded that GO-QoL was the best disease-specific QoL assessment tool, given its high validity, reliability, and cultural applicability.[69] A standard GO-QoL questionnaire and its scoring system are shown in Table 72.5. Originally developed in Dutch, GO-QoL is now available in 18 languages. All major treatment trials regarding GO over the past decade have included GO-QoL as one of the primary or secondary outcomes so that the benefits of certain interventions could be comprehensively evaluated.

Previous studies reported that certain demographic or clinical characteristics were associated with the visual functioning subscale (VFS) and the appearance subscale (AS) of the GO-QoL. Their results are inconsistent, probably because of the following confounders: (1) non-consecutive patient recruitment resulting in potential selection bias; (2) wide disease spectrum in the study population; (3) QoL assessed when the patients had not been rendered euthyroid; and (4) previous GO treatment may have altered QoL. Recently, a EUGOGO center performed a prospective follow-up trial, recruiting consecutive patients with active moderate-to-severe GO who were euthyroid and treatment-naïve before study inclusion.[68] The predictive factors for QoL at baseline and 6 months after standard intravenous GC treatment were studied using multivariate analysis. While increasing age, overall disease severity, and diplopia predicted worse baseline VFS score, CAS, upper lid retraction, and diplopia were associated with a worse VFS score post-treatment. Younger age, female gender, smoking history, and overall disease severity predicted worse baseline AS score, while only overall disease severity and upper lid retraction were associated with a worse AS score post-treatment. Both GO-QoL subscales significantly improved after treatment. It became evident that, while some demographic factors correlated with baseline QoL, the improved GO-QoL after treatment was exclusively associated with ophthalmic parameters. These findings suggest that treatment strategies that effectively ameliorate both disease activity and severity help optimize QoL in patients suffering from active moderate-to-severe GO.

KEY POINTS

- All patients with Graves' orbitopathy should undergo comprehensive clinical evaluation (disease activity and severity and quality-of-life assessment) and laboratory tests (thyroid function test and thyroid-stimulating hormone receptor antibody). Orbital imaging will be indicated in patients with atypical presentation or inconclusive clinical evaluation, as well as in preoperative planning.

MANAGEMENT

In this section, we will discuss the management of GO in terms of general measures which apply to all GO patients, specific pharmacological treatment, and the role of surgery in GO.

General Measures

Multidisciplinary Care. EUGOGO recommends referral of all GO patients (except for the mildest cases) to combined thyroid–eye clinics

TABLE 72.5 Graves' Orbitopathy Quality of Life Questionnaire

The following questions deal specifically with your thyroid eye disease. Please focus on the past week while answering these questions.

During the past week, to what extent were you limited in carrying out the following activities, because of your thyroid eye disease? (Questions 1-8)

Visual Functioning Subscale (VFS)	*1 point*	*2 points*	*3 points*
1. Bicycle (Skip if never learned to ride a bike)	*Yes, Seriously*	*Yes, a Little*	*No, Not at all*
2. Driving (Skip if no driver's license)			
3. Moving around the house			
4. Walking outdoors			
5. Reading			
6. Watching television			
7. Hobby or pastime			
8. During the past week, did you feel hindered from something that you wanted to do because of your thyroid eye disease?			

The following questions deal with your thyroid eye disease in general. (Questions 9-16)

Appearance subscale (AS)			
9. Do you feel that your appearance has changed because of your thyroid eye disease?	*Yes, Very Much So*	*Yes, a Little*	*No, Not at all*
10. Do you feel that you are stared at in the streets because of your thyroid eye disease?			
11. Do you feel that people react unpleasantly because of your thyroid eye disease?			
12. Do you feel that your thyroid eye disease has an influence on your self-confidence?			
13. Do you feel socially isolated because of your thyroid eye disease?			
14. Do you feel that your thyroid eye disease has an influence on making friends?			
15. Do you feel that you appear less often on photos than before you had thyroid eye disease?			
16. Do you try mask changes in appearance caused by your thyroid eye disease?			

- Calculation:
  - Total number of points scored in questions 1–8 = raw score for VFS (range 8-24); total number of points scored in questions 9–16 = raw score for AS (range 8–24)
  - Total score for VFS or AS (each range 0–100) = $100 \times (\text{raw score} - y)/2\,y$, where y represents the number of valid answers in each subscale.
  - For both VFS and AS, higher scores indicated better quality of life.

or specialized centers that provide endocrinological and ophthalmological expertise.[12] Early diagnosis and treatment are believed to have contributed to an increase in prevalence of mild and inactive disease since the implementation of the Amsterdam Declaration.[3,4] A UK survey demonstrated that significantly more patients who attended specialist GO clinics were satisfied with treatment than those who did not (67% vs. 52%, $P < .05$).

Smoking Cessation. The well-established harmful effects of smoking on GO are discussed in "Radioactive Iodine Therapy." According to a UK multicenter GO audit, only 63% of current smokers were offered smoking cessation advice, and only one-third of GO patients were aware of the deleterious effect of cigarette smoking on GO.[70] Physicians should urge all patients with Graves hyperthyroidism, irrespective of the presence/absence of GO, to refrain from smoking, if necessary with the help of specialized smoking cessation programs or clinics.

Thyroid Dysfunction. Prompt restoration and stable maintenance of euthyroidism are priorities in GO patients. As recommended by the European Thyroid Association, the first-line treatment of Graves hyperthyroidism is antithyroid drugs (ATDs).[71] Major guidelines favor the administration of ATDs as a titration regimen (T) instead of as a "block and replace" regimen (BR), which is associated with a higher rate of side effects.[71,72] Some argue that BR may be associated with more stable thyroid function, although there is little supportive evidence. In a recent EUGOGO prospective multicenter observational cohort study,[73] 344 patients with untreated Graves hyperthyroidism without overt GO were treated with either T ($n = 278$; 81%) or BR ($n = 66$; 19%) for 18 months. During the course of ATDs, the prevalence of euthyroidism was always higher in the T group than in the BR group (6 months: 47.8% vs. 24.5%, $P = .002$; 12 months: 81.4% vs. 58.2%, $P = .001$; 18 months: 85.6% vs. 62.7%, $P = .001$). There was no difference between both groups in TSH-R–Ab levels, which decreased significantly in all patients after treatment with ATDs. There was a trend of higher prevalence of *de novo* GO in the T group (17.8% vs. 9.1%, $P = .096$), although it was not statistically significant. It remains uncertain whether the choice of ATD regimen affects the risk of subsequent development of GO, or whether it alters the natural history of preexisting GO. Until larger randomized controlled trials (RCTs) resolve these issues, careful titration of either ATD regimen to maintain euthyroidism is essential for all GO patients, as both under- and overtreatment with ATDs will worsen GO. Likewise, clinicians should ensure adequate thyroid hormone replacement in patients with hypothyroid GO and watch out for emergence of thyroid dysfunction in those with euthyroid GO.

Regarding definitive treatments of Graves hyperthyroidism, thyroidectomy does not seem to affect the natural course of GO in the presence of adequate postoperative thyroid hormone replacement. However, RAI is a definite risk factor for progression or *de novo* occurrence of GO (discussed in "Radioactive Iodine Therapy"). The use of GC prophylaxis effectively reduces the risk. It is postulated that GCs work by blunting the surge in TSH-R–Ab levels, a well-documented phenomenon after RAI.[17] Apparently, patients with Graves hyperthyroidism for more than 5 years are not at increased risk of the

TABLE 72.6 Post–Radioactive Iodine Therapy Glucocorticoid Prophylaxis for Patients With Graves Hyperthyroidism

Indications According to Different Scenarios

Graves' orbitopathy (GO) status	*Other risk factors[a] present?*	*Radioactive iodone therapy (RAI) as one of the possible treatment options?*	*Glucocorticoid prophylaxis indicated?*	
			ATA 2016[b]	*ETA 2018[c]*
No pre-existing GO	No	Yes	No	No
No pre-existing GO	Yes	Yes	Insufficient data for recommendation	Yes
Inactive mild-to-severe GO	No	Yes	No	No
Inactive mild-to-severe GO	Yes	Yes	No	Yes
Active mild GO	No	Yes	Yes, consider risk-benefit ratio	Yes
Active mild GO	Yes	Yes	Yes	Yes
Active moderate-to-severe GO or sight-threatening GO		No, RAI should be avoided		

Regimens of Glucocorticoid Prophylaxis

Regimen	*Dosing*	*When to start?*	*Remarks*
Standard-dose oral prednisone	0.4–0.5 mg/kg body weight daily for 4 weeks, then taper over 8 weeks	1–3 days after RAI	• Most well-validated regimen
Low-dose oral prednisone	0.2–0.3 mg/kg body weight daily, reduce by 2.5–5 mg every 7–14 weeks, total 6 weeks	1 day after RAI	• Not consistently effective • Better reserved for low-risk patients (e.g., mild or absent GO)
Intravenous methylprednisolone	500 mg weekly for 2 weeks, then 250 mg weekly for 2 weeks	2 days after RAI	• Equally effective but associated with less insomnia and gastric symptoms when compared with standard-dose oral prednisone

[a] High serum thyroidstimulating hormone receptor antibody titer (>5-fold cutoff), acute or severe hyperthyroidism (high serum free T3 >3- to 5-fold), active smoking, recent-onset hyperthyroidism (duration <5 years), or delayed correction of post-RAI hypothyroidism.

[b] 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis[53]

[c] 2018 European Thyroid Association Guideline for the Management of Graves' Hyperthyroidism[52]

adverse effects on GO, and GC prophylaxis may not be necessary in these patients.[17] The indications and regimens for GC prophylaxis are shown in Table 72.6.[15,71,72]

Local Treatment Modalities. GO patients frequently complain of dry eyes owing to tear film instability, which is caused by lacrimal gland dysfunction, proptosis, eyelid abnormalities, and impaired Bell's phenomenon. Frequent use of artificial tears (while awake) and lubricant ointments (before sleep) is essential for ocular surface protection. Lifestyle modifications (e.g., limiting screen time, avoiding prolonged reading or driving) should be advised to reduce excessive strain on the eyes. Wearing sunglasses may relieve photophobia and minimize ocular surface irritation (e.g., by wind). Prisms as temporary measures can reduce troublesome diplopia while awaiting more definitive correction by immunomodulatory treatment or rehabilitative surgery. Periocular injections of GCs or botulinum toxin may offer temporary control of eyelid retraction, extraocular muscle inflammation, and ocular dysmotility.

KEY POINTS

- All Graves' orbitopathy patients should be evaluated and managed in dedicated multidisciplinary thyroid–eye clinics, which improve patient outcomes and satisfaction. Smoking cessation, maintenance of stable euthyroidism, and local treatment are essential general measures.

Mild Graves' Orbitopathy

Selenium. The importance of selenium to thyroid physiology is reflected by the high concentration of selenium in the thyroid gland. Thyrocytes express several selenoproteins, including deiodinase isoenzymes, members of the glutathione peroxidase, and the thioredoxin reductases, which play important roles in thyroid hormone metabolism, antioxidative functions, and antiinflammatory functions.[74] In a cell culture study, preincubation with selenium (in the form of selenium-methylselenocysteine, SeMCys) abolished some of the adverse effects (cellular proliferation and secretion of proinflammatory cytokines) of hydrogen peroxide–induced oxidative stress on OFs from GO patients through an increase in glutathione peroxidase activity. SeMCys also reduced HA release, which was not augmented by hydrogen peroxide–induced oxidative stress, from OFs.[75] In a EUGOGO-led multicenter randomized placebo-controlled trial,[76] selenium administration (sodium selenite 100 μg twice daily for 6 months) resulted in improved QoL ($P < .001$), greater improvement in CAS ($P < .001$), higher likelihood of overall improvement ($P = .01$), and lower risk of disease progression ($P = .01$) in patients with mild GO. The better overall ophthalmic outcomes in the selenium group were mainly driven by improvement in eyelid aperture and soft tissue involvement, but not by changes in proptosis or ocular motility. The beneficial effects of selenium persisted for 6 months after treatment withdrawal. Hence, EUGOGO recommends 6-month selenium supplementation for patients with mild GO of relatively short duration because it improves eye manifestations and QoL, as well as prevents GO progression to more severe forms.[12] Although this European study was conducted in selenium-deficient geographical regions, the selenium status of individual patients was not evaluated. This has raised the question of whether the beneficial effects of selenium are explained by replacement of underlying deficiency or if they are independent of patients' baseline selenium status. Clinicians should also be aware that excessive selenium intake is associated with certain health hazards, for example, malignancy (skin cancer, prostate cancer), type 2 diabetes mellitus, and increased overall mortality.[74] A 2018 European Thyroid Association survey[77] showed that most respondents (94.1%) recommended selenium use in GO. Although nearly half of them (44.9%) also recommended selenium in addition to standard treatment in patients with moderate-to-severe GO, this approach has yet to be proven.

KEY POINTS

- Six-month selenium supplementation ameliorates and prevents progression of mild Graves' orbitopathy.

Active Moderate-to-Severe Graves' Orbitopathy

Intravenous Glucocorticoids. Systemic GCs have been the most commonly used immunosuppressants in the treatment of active moderate-to-severe GO, with good efficacy, since the 1950s. They achieve anti-inflammatory and immunomodulatory functions by: (1) interfering with the functions of B and T lymphocytes; (2) suppressing the recruitment and functions of immunocompetent cells; (3) inhibiting the release of proinflammatory mediators; and (4) decreasing the synthesis and secretion of glycosaminoglycans from activated OFs.[78] The molecular mechanisms underlying GC actions can be classified as genomic and non-genomic pathways. Through the genomic pathway, the lipophilic GCs bind to their intracellular receptor. The activated GC–receptor complex is then transported into the cell nucleus, where it dissociates and interferes with the transcription processes. The production of anti-inflammatory mediators is increased by transactivation of anti-inflammatory genes, whereas that of proinflammatory mediators is reduced through transrepression of proinflammatory genes. Although the genomic mechanisms take time to consolidate, the rapid effects of GCs, especially when given via an intravenous route, have been attributed to the non-genomic pathway. The non-genomic pathway modulates intracellular signaling through: (1) interacting with cytosolic receptors; (2) interacting with membrane-bound steroid receptors and their second messenger systems; and (3) stabilizing the cell membrane by intercalation, which ultimately disturbs the physiological course of affected cells (exclusively occurs during high-dose GC treatment).[78,79]

EUGOGO recommends high-dose intravenous GCs (IVGCs) as the first-line treatment for active moderate-to-severe GO.[12] IVGCs demonstrated superior efficacy (overall response rate 70%–80% vs. 50%) and safety profile (rate of adverse events 39% vs. 81%) than oral GCs (OGCs).[12] IVGCs substantially improved CAS, ocular dysmotility, visual acuity, and QoL,[68] but only had a limited benefit on proptosis. They are most effective and disease-modifying when given early during the active phase of the disease. Successful treatment was associated with reduced need for rehabilitative surgery in the long run.[80] The typical dosing regimens, monitoring schedule, and absolute contraindications of IVGCs are shown in Table 72.7. The overall morbidity and mortality associated with IVGCs were 6.5% and 0.6%, respectively.[78] Severe cardiovascular (e.g., acute coronary syndrome, stroke) and hepatic toxicities (e.g., acute hepatitis, acute liver failure) were uncommon but potentially fatal,[78] and current consensus suggests several precautions to minimize the risks of such events: (1) each single dose of IVGCs (as methylprednisolone) should not exceed 750 mg; (2) the cumulative dose of each IVGC course should not exceed 8 g; and (3) administration of IVGCs on consecutive days should be avoided.[12]

TABLE 72.7 Intravenous Glucocorticoid Use in Graves' Orbitopathy—Regimen, Monitoring and Absolute Contraindications

Regimen according to Graves' orbitopathy (GO) status	• Active moderate-to-severe GO
	  ◦ Intravenous methylprednisolone 500 mg weekly for six doses, then 250 mg weekly for six doses
	  ◦ Cumulative dose 4.5 g, 12-week course
	• Active moderate-to-severe GO with diplopia or severe proptosis
	  ◦ Intravenous methylprednisolone 750 mg weekly for six doses, then 500 mg weekly for six doses
	  ◦ Cumulative dose 7.5 g, 12-week course
	• Dysthyroid optic neuropathy ("hit hard and short")
	  ◦ Intravenous methylprednisolone 750 mg alternate day for seven doses (2 weeks), (if dysthyroid optic neuropathy resolves), then 500 mg weekly for five doses
	  ◦ Cumulative dose 7.75 g, 7-week course

Monitoring and Precautions	• Before treatment:
	  ◦ Check fasting blood glucose, liver function test, and markers for viral hepatitis
	  ◦ Ensure stable control of hypertension and diabetes mellitus if any
	• During treatment:
	  ◦ Monitor fasting blood glucose, liver function test, and blood pressure at least every other week
	  ◦ Observe for mood changes or mental disturbance
	• Adjunctive treatments:
	  ◦ Proton pump inhibitors (to prevent peptic ulcer) if needed

Absolute contraindications	• Recent viral hepatitis or significant hepatic dysfunction • Severe cardiovascular disease(s) • Uncontrolled severe hypertension or unstable diabetes mellitus • Active psychiatric disorders

Orbital Radiotherapy. Orbital radiotherapy (ORT) has been established as a treatment modality of GO since the 1970s, following a series of reports describing its clinical efficacy. The clinical benefit of ORT in GO is believed to be mediated through depleting radiosensitive lymphocytes that infiltrate the orbits and drive orbital inflammation.[79] ORT may also induce terminal differentiation of OFs, thus limiting their activation (e.g., secretion of GAG and proinflammatory mediators, *de novo* adipogenesis) and resulting in durable therapeutic effect.[81] The typical dosing regimen is a cumulative dose of 20 Gy per orbit fractionated in 10 daily doses given over a 2-week period,[12] but an alternative regimen of 1 Gy per week over a 20-week period was shown to be equally effective and better tolerated.[82]

A recent comprehensive review has summarized the clinical efficacy of ORT.[81] ORT and OGC demonstrated similar efficacy as initial treatment of GO, but ORT was better tolerated. There has been ample evidence since the 1980s that the combination of ORT and OGC is more efficacious than either therapy alone. However, a recent multicenter, double-blind, RCT showed that concurrent ORT did not have added benefit in patients receiving a 24-week course of OGC.[83] In general, patients with a shorter duration of GO before treatment and with active moderate-to-severe disease were more likely to benefit from ORT, compared with those with long disease duration and mild inactive disease. ORT mainly improved ocular dysmotility and soft tissue involvement, while the effect on proptosis was limited. In addition, retrospective case series supported the disease-modifying role of ORT in the management of GO. ORT-GC combination lowered the risk of disease reactivation and developing DON compared with GC monotherapy in patients with active moderate-to-severe GO[84,85] and helped eliminate the need for urgent orbital decompression in patients with GC-responsive DON.[86] ORT given during the active phase of GO was also associated with better outcomes after rehabilitative orbital decompression, including greater improvement in proptosis and lower risk of postoperative increase in extraocular muscle volume, as well as diplopia.[87,88] Most clinical evidence of ORT-GC combination in active moderate-to-severe GO was based on patients who received OGC or an IVGC/OGC combination. There has not been any prospective randomized trial evaluating whether ORT offers additional benefit on top of standard IVGC therapy. In a retrospective study comparing ORT-IVGC and IVGC monotherapy (median cumulative dose in both groups 1.6–1.9 g only) in patients with active moderate-to-severe GO, ocular dysmotility improved only in the combination group, which also demonstrated more frequent overall ophthalmic improvement and lower risk of disease reactivation, compared with IVGC monotherapy.[89]

Overall, ORT is safe, without major safety concerns. Transient conjunctivitis and periorbital skin irritation may occur. Large retrospective studies showed no increased risk of cataract formation compared with nonirradiated GC-treated GO patients or normal controls.[90,91] Long-term follow-up analyses did not reveal any cancer within the radiation field nor increased risk of radiation-induced cancer death compared with the general population.[92,93] Nonetheless, ORT should be avoided in patients below the age of 35 years, in view of the theoretical risk of carcinogenesis.[79] Coexisting diabetes mellitus and hypertension were associated with the development of retinopathy after ORT, which could be solely radiation-induced or represent progression of underlying diabetic or hypertensive retinopathy.[90,91] Hence, ORT is contraindicated in patients with diabetic or hypertensive retinopathy.[79]

Glucocorticoid-Resistant, Active, Moderate-to-Severe Graves' Orbitopathy. If patients with active GO show no or only partial response to first-line IVGC therapy, EUGOGO recommends several second-line treatment options, namely a second course of IVGCs (as long as patients can tolerate it, with a normal liver function test), OGC-ORT combination, OGC-cyclosporin combination, and rituximab.[12] Although these strategies are efficacious in treatment-naïve cases, no prospective randomized trial has examined their role in patients resistant to IVGCs. Persistently high CAS may sometimes be caused by orbital congestion instead of active inflammation. Orbital decompression will provide rapid relief of congestion, which is usually resistant to immunomodulation. MRI may help distinguish between the two clinically similar conditions. Management for these difficult cases should be individualized, with detailed discussion between the multidisciplinary team and patients. Novel treatment approaches (see below) are potential solutions and will be elaborated in a later section.

Sight-Threatening Graves' Orbitopathy

Dysthyroid Optic Neuropathy. DON is the most common form of sight-threatening GO. Prompt recognition and timely treatment are essential in restoring optic nerve function. IVGC therapy was superior to orbital decompression as upfront treatment in terms of resolution of DON (response rate 56% vs. 17%) based on a small randomized clinical trial.[94] EUGOGO recommends high-dose IVGCs as the first-line treatment of DON if not contraindicated[12] (Table 72.7). If features of DON persist after 2 weeks of IVGCs, prompt orbital decompression should be performed to prevent irreversible visual loss. Urgent surgery can be avoided in up to 40% to 50% of DON cases after IVGCs.[10,95] The presence of optic disc swelling at diagnosis,[95] persistently active disease at 2 weeks after IVGCs,[95] and DON attributed to optic nerve stretch[96] predicted GC unresponsiveness, and patients with the above features will benefit from early orbital decompression. In a retrospective study, 104 patients with GC-responsive DON received ORT, and 94% of them did not require urgent orbital decompression. Therefore, ORT concurrent with IVGC therapy may prevent recurrent DON after GC withdrawal.[86]

KEY POINTS

- Intravenous glucocorticoid (IVGC) is the standard treatment for active moderate-to-severe Grave orbitopathy and dysthyroid optic neuropathy (DON). Orbital radiotherapy may further improve overall response (especially in ocular dysmotility) and prevent disease reactivation or progression. Urgent orbital decompressive surgery is indicated in IVGC-resistant DON.

Corneal Breakdown and Other Forms of Sight-Threatening Graves' Orbitopathy

Approximately 16% of GO patients develop corneal involvement.[3] They are at risk of corneal damage, especially when lagophthalmos and poor Bell's phenomenon coexist, meaning that the normal corneal protective mechanisms are defective. Corneal breakdown is a medical emergency that warrants immediate intensive treatment, including frequent topical lubricants and antibiotics, moisture chambers, and other temporary measures to improve eyelid closure (e.g., blepharroraphy, tarsorraphy, amnion membranes, botulinum toxin, emergency gluing). Corneal grafting may be necessary in case of irreversible damage. After the above acute management, permanent measures to improve eyelid closure are essential to prevent recurrent exposure keratopathy and corneal breakdown.

Urgent orbital decompression is indicated if other rare forms of sight-threatening GO are present, including globe subluxation (sudden anterior displacement of the globe with eyelid retraction behind the equator of the globe), choroidal folds (indentation on the globe by enlarged extraocular muscles), and postural visual darkening (borderline optic nerve perfusion because of high intraorbital pressure aggravated by transient fall in systolic blood pressure on upright posture[97]).

Inactive Graves' Orbitopathy and the Role of Rehabilitative Surgery

Despite successful immunomodulatory therapies leading to inactivation of the disease, GO patients may continue to suffer from disfigurement and disability owing to irreversible orbital tissue remodeling. Rehabilitative surgery aims at correcting disfigurement and restoring visual function during the inactive post-inflammatory phase, when pharmacological treatments have a limited disease-modifying role. GO should remain inactive and stable for at least 6 to 8 months before rehabilitative surgery is considered; otherwise, ongoing inflammation and tissue remodeling may render surgery ineffective. There are three main types of rehabilitative surgery: (1) orbital decompression, which reduces proptosis by removing orbital bone or adipose tissue; (2) eye muscle surgery, which ameliorates squint and diplopia; and (3) eyelid surgery, which corrects the abnormal eyelid position. If patients require more than one type of procedure, orbital decompression should be the first step, followed by eye muscle surgery, and then finally eyelid surgery. The rationale behind this sequence is that orbital decompression may cause deterioration in ocular dysmotility because of displacement of orbital soft tissue, and vertical squints may influence the optimal eyelid position. Rehabilitative surgery effectively improves patients' visual functioning, appearance, and overall QoL.[69]

Novel Treatment Approaches for Immunosuppression in Graves' Orbitopathy

Limitations of Current Therapy. IVGC is currently the standard of care in active GO, but approximately 20% to 30% of cases will have persistently active disease despite standard treatment. On the other hand, some patients develop disease reactivation after steroid withdrawal. A recent randomized trial that evaluated IVGC efficacy by a more stringent and comprehensive composite outcome found an overall response rate of 50% only (compared with 70%–80% in older studies) and a relapse rate of around 10%.[98] Disease severity (e.g., proptosis, diplopia, ocular dysmotility) tends to be less responsive to IVGCs compared with disease activity. Residual functional impairment, impaired QoL, and the requirement for rehabilitative surgery may therefore persist despite standard treatment. Severe toxicities and treatment-related mortality associated with IVGCs are uncommon but non-negligible. Novel treatment strategies for active moderate-to-severe GO that address some of the above limitations have emerged over the past decade.[29,36,39] The key findings from representative RCTs are discussed later and summarized in Table 72.8.

TABLE 72.8 **Summary of Key Randomized Controlled Trials on Novel Treatment Approaches of Graves' Orbitopathy**

Therapeutics	Mechanism of Action	Study Design/Patient population	Key Outcomes	Results (intervention vs. control group)
Mycophenolate mofetil[80]	Dual antiproliferative effect on both B- and T-cells	• Single-center RCT • 174 euthyroid patients with active moderate-to-severe GO • MMF 1 g daily vs. glucocorticoid therapy (both for 24 weeks)	**Primary** • Response rates at weeks 12 and 24 (improvement in ≥3 parameters of the composite index without deterioration in any objective parameters)	• Week 12: 79% vs. 51% • Week 24: 91% vs. 68%
Mycophenolate sodium[78]	Dual antiproliferative effect on both B- and T-cells	• Multicenter RCT • 164 euthyroid patients with active moderate-to-severe GO • MPS 720 mg daily (24 weeks) plus weekly IVGC (12 weeks) vs. weekly IVGC (12 weeks) alone	**Primary** • Response rate at week 12 (improvement in ≥2 parameters of the composite index without deterioration in any parameters) • Relapse rates at weeks 24 and 36 **Post hoc** • Response rate at week 24 • Sustained response at week 36 • GO-QoL scales	• No significant difference between groups • 71% vs. 53% • 67% vs. 43% • No significant difference between groups
Rituximab[83]	B-cell depletion	• Single-center RCT • 32 euthyroid patients with active moderate-to-severe GO • RTX (2 g/500 mg IV) vs. weekly IVGC (12 weeks with cumulative dose 7.5 g)	**Primary** • Decrease of CAS of ≥2 points or disease inactivation (CAS <3) at week 24 **Secondary** Improvement in severity signs (NOSPECS W class, proptosis and lid width, eye motility) • GO reactivation • Percentage of patients requiring surgical procedures after 12 months • GO-QoL scales	• 100% vs. 69% • Better ocular motility in RTX group • 0% vs. 33% • 33% vs. 75% • Visual functioning scores improved significantly in RTX group only
Rituximab[84]	B-cell depletion	• Single-center RCT • 25 euthyroid patients with active moderate-to-severe GO • RTX (2 g IV) vs. placebo	**Primary** • Decrease of CAS (assessed as a continuum and separately as an improvement by ≥2 points) at week 24	• No significant difference between groups in both primary and all secondary outcomes
Tocilizumab[100]	Anti-IL-6 receptor MAb	• Multicenter RCT • 32 euthyroid patients with glucocorticoid-resistant active moderate-to-severe GO • TCZ (8 mg/kg IV once every 4 weeks for 4 doses) vs. placebo	**Primary** • Decrease of CAS of ≥2 points at week 16 **Secondary/post hoc** • Decrease of CAS of ≥2 points at week 40 • CAS <3 at weeks 16 and 40 • Disease severity • GO-QoL scales	• 93% vs. 59% • 87% vs. 59% • Week 16: 87% vs. 35% • Week 40: 80% vs. 47% • No significant improvement of proptosis and diplopia in TCZ group • More patients in TCZ group reported improvement in GO-QoL (47% vs. 35%)

TABLE 72.8 Summary of Key Randomized Controlled Trials on Novel Treatment Approaches of Graves' Orbitopathy—cont'd

Therapeutics	Mechanism of Action	Study Design/Patient population	Key Outcomes	Results (intervention vs. control group)
Teprotumumab[103,104]	Anti-IGF1-R blocking MAb	• Two multicenter RCTs • 170 euthyroid patients with active moderate-to-severe GO in total • TPT (once every 3 weeks for 8 IV infusions; first dose 10 mg/kg, remaining seven doses 20 mg/kg)	• Overall responders (reduction of ≥2 points in CAS and reduction of ≥2 mm in proptosis) at week 24 • CAS of 0 or 1 • Reduction of ≥2 mm in proptosis • Diplopia response • GO-QoL scales	• 73% vs. 14% • 62% vs. 22% • 77% vs. 15% • 68% vs. 27% • TPT group reported higher scores and greater improvement in overall, visual functioning, and appearance scores

RCT, Randomized controlled trial; *GO*, Graves' orbitopathy; *MMF*, mycophenolate mofetil; *MPS*, mycophenolate sodium; *IVGC*, intravenous glucocorticoid; *GO-QoL*, Graves' orbitopathy quality of life questionnaire; *RTX*, rituximab; *IV*, intravenous; *CAS*, clinical activity score; *MAb*, monoclonal antibody; *TCZ*, tocilizumab; *IGF1-R*, insulin-like growth factor-1 receptor; *IV*, intravenous; *TPT*, teprotumumab

Mycophenolate. Pathophysiological basis and mechanisms of actions: Mycophenolic acid (MPA) inhibits inosine monophosphate dehydrogenase (IMPDH), which catalyzes the rate-limiting step in *de novo* synthesis of guanosine nucleotides in activated lymphocytes, and therefore exerts dual antiproliferative action on both B- and T-cells. In addition, *in vitro* studies suggest that MPA is capable of inhibiting fibroblast proliferation and function via both IMPDH-dependent and -independent (e.g., PI3K/mTOR) pathways. MPA is clinically available as mycophenolate mofetil (MMF) and enteric-coated mycophenolate sodium (MPS).

Clinical efficacy, safety, and therapeutic application: A single-center RCT demonstrated that MMF was superior to GC monotherapy (both for 24 weeks) in terms of overall response and individual ophthalmic parameters (e.g., soft tissue involvement, diplopia, proptosis)[99] (Table 72.8). A subsequent EUGOGO multicenter RCT compared standard 4.5 g weekly IVGC alone and IVGC with MPS for 24 weeks[98] (Table 72.8). The combination group achieved a higher overall response rate at week 24 and a higher sustained response rate at week 36, as well as more significant improvement in individual ophthalmic components (including CAS and its components, vertical ocular dysmotility, and visual functioning subscale of GO-QoL). Several important differences in baseline characteristics between the two study populations (higher mean age, more prevalent smoking, longer disease duration, and greater proportion of TSH-R–Ab–positive patients) possibly explain the lower response rates in the EUGOGO trial.[100] Low-dose mycophenolate use in GO is generally safe.[100] Mild gastrointestinal intolerance was slightly more prevalent among mycophenolate-treated patients. Compared with IVGC monotherapy, combination treatment did not result in increased risk of major toxicities, including serious infection, hepatotoxicity, cytopenia, and treatment-related mortality. Therefore, IVGC-mycophenolate combination will potentially supplant IVGC monotherapy as the first-line treatment of active moderate-to-severe GO, in view of its superior efficacy and reassuring safety profile.

Rituximab. Pathophysiological basis and mechanisms of actions: Rituximab (RTX), a chimeric murine/human monoclonal antibody targeting CD20, results in rapid depletion of both peripheral B-cells and intraorbital B-/T-cells.[22] As the clinical improvement of GO after RTX does not always correlate with the reduction in TSH-R–Ab level, the efficacy of RTX is probably explained by elimination of other B-cell functions apart from autoantibody production (e.g., cytokine production, antigen presentation, B-/T-cell costimulation, etc.).[24]

Clinical efficacy, safety, and therapeutic application: Two RCTs examined the role of RTX in active moderate-to-severe GO, but they produced grossly conflicting results (Table 72.8). In the Italian trial,[101] RTX demonstrated superior efficacy compared with weekly IVGC therapy in terms of disease inactivation, ocular dysmotility, and requirement of subsequent rehabilitative surgery. In contrast, the US trial[102] failed to show any difference between RTX and placebo. A subsequent joint *post hoc* analysis suggested that several differences in baseline characteristics (younger age, lower TSH-R–Ab levels, shorter duration of GO) may have accounted for the positive treatment outcomes in the Italian trial.[25] A few RTX-treated patients in these two RCTs developed transient deterioration in proptosis or visual loss, and even frank DON. It is postulated that massive lysis of intraorbital lymphocytes because of RTX therapy may cause cytokine release syndrome, with subsequent edema and expansion of orbital tissues, leading to progression of subclinical DON.[100] EUGOGO recommends RTX as one of the second-line treatment options after failure of IVGCs, but it is best avoided in patients with impending DON or long duration of disease.[12] This recommendation is in keeping with the conclusions of several recent retrospective studies.[103-105]

Tocilizumab. Pathophysiological basis and mechanisms of actions: IL-6 is a potent proinflammatory cytokine implicated in the pathogenesis of multiple autoimmune disorders. Activated GO OFs produced multiple chemoattractants, including IL-6, that support B-cell differentiation and synthesis of autoantibodies.[106-108] Although TSH-R activation enhanced the expression of IL-6 in GO OF/adipocytes, IL-6 in turn stimulated TSH-R expression in these orbital target cells.[109,110] IL-6 suppressed regulatory T-cell induction while promoting the development and functions of Th17 cells, a recently identified CD4+ T-cell subset that plays an important pathogenic role in GO.[111-113] GD patients with GO had higher serum levels of IL-6 than those without GO.[114] Greater lacrimal levels of IL-6 were also evident in GO patients compared with healthy controls.[115,116] Therefore, antagonizing the IL-6 pathway with tocilizumab (TCZ), a fully human IgG1 anti-IL6 receptor monoclonal antibody, represents a sound strategy in GO.

Clinical efficacy, safety, and therapeutic application: A small RCT evaluated the efficacy of TCZ versus placebo in patients with GC-resistant active moderate-to-severe GO[117] (Table 72.8). Significantly more patients in the TCZ group achieved improvement in CAS and disease inactivation, but not in proptosis or diplopia. Multiple preliminary reports and recent retrospective series[118,119] have also described the efficacy of TCZ in GC-resistant or -dependent cases. TCZ is overall safe and well-tolerated. The remarkable anti-inflammatory action of TCZ in active GO holds great promise, although larger clinical trials are needed to confirm its benefits and precisely define its role in GO management.

Teprotumumab. Pathophysiological basis and mechanisms of actions: The role of IGF-1R in the pathogenesis of GO has been discussed in the section "Thyroid Stimulating Hormone Receptor and Insulin-Like Growth Factor-1 Receptor Crosstalk." *In vitro* studies showed that blocking IGF-1R inhibited activation of OFs and fibrocytes. Teprotumumab (TPT) is a fully human IgG1 monoclonal blocking antibody that binds to the extracellular portion of IGF-1R and blocks its activation and signaling by endogenous ligands (IGF-1 and IGF-2). TPT also leads to near-complete internalization and degradation of IGF-1R on the cell surface. Once developed as an anticancer therapy, TPT has been subsequently repurposed for potential use in GO based on preclinical data.

Clinical efficacy, safety, and therapeutic application: The efficacy and safety of TPT versus placebo in treatment-naïve, active, moderate-to-severe GO were evaluated in two multicenter RCTs with similar trial design.[120,121] The primary endpoint of the initial phase II study[120] was a composite of reduction of at least 2/7 in CAS and a reduction of at least 2 mm in proptosis at week 24, whereas the subsequent phase III study[121] only focused on reduction in proptosis as the primary endpoint. The key findings of both studies are jointly summarized below. Approximately 70% of TPT-treated patients (~20% in the placebo group) achieved significant improvement in CAS, proptosis, or both. In particular, the proptosis response was both rapid (evident by week 6 in most responders) and marked (mean reduction of 3 mm, comparable to that achieved by single-wall orbital decompression). Furthermore, a 4-mm reduction in proptosis was achieved in 40% of the TPT group (phase II), and 72% of overall responders (phase II) were graded as having high response (reduction of at least 3 mm in proptosis and reduction of at least 3/7 in CAS). Significant improvement in diplopia and GO-QoL was also achieved in the TPT group. Patients in the phase III study who did not respond or relapse could enter an open-label extension study and receive second course of TPT. The follow-up and extension studies are currently ongoing. TPT was generally well tolerated, and the only significant side effect was hyperglycemia, which occurred in 7% of patients who received TPT. It mainly affected those with pre-existing diabetes mellitus and was easily managed with adjustment in antidiabetic medications.

In view of its impressive all-round efficacy (in terms of disease activity, severity, and QoL) and favorable safety profile, TPT has become the first pharmacotherapy approved by the US Food and Drug Administration for the treatment of adult GO since January 2020. Although there has been heated debate on whether TPT should replace IVGCs as the first-line treatment in active GO,[122] several issues of concern have to be adequately addressed before we incorporate TPT in our routine clinical practice[123]: (1) lack of long-term efficacy and safety data for TPT, including head-to-head comparison with IVGCs; (2) restricted availability (currently only commercially available in the United States); and (3) the unaffordability of TPT for most patients.

KEY POINTS

- Mycophenolate enhanced treatment efficacy of intravenous glucocorticoids. Tocilizumab is potentially useful in glucocorticoid-resistant active Graves' orbitopathy (GO). Teprotumumab resulted in marked improvement of active GO, in terms of both disease activity and severity.

BRIEF NOTE ON GRAVES DERMOPATHY AND THYROID ACROPACHY

Graves dermopathy (most commonly pretibial myxedema [PTM]) and thyroid acropachy are rare extrathyroidal manifestations of GD. According to case series that recruited patients before the twenty-first century, PTM occurs in 4% of patients with GO and in 13% of those with severe GO. Overall, PTM and thyroid acropachy affect approximately 1.5% and 0.3% of GD patients, respectively. Thyroid acropachy almost always occur in association with GO and Graves dermopathy, and PTM is associated with thyroid acropachy in 20% of cases.[124,125] However, prospective observational studies by EUGOGO reveal a significant reduction in the prevalence of Graves dermopathy in patients with GO, from 6% in 2010 to 0.7% in 2012.[3] The contemporary estimated prevalences of Graves dermopathy associated with GO and thyroid acropachy associated with GO in Europe are 0.15 and 0.03 per 10,000 population, respectively (compared with 10 for GO).[1] In general, PTM and acropachy share similar risk factors as GO. Local mechanical factors, for instance trauma or surgery and dependency (which may be aggravated by obesity) of lower extremities, also predispose to the development of PTM in susceptible patients.

Graves dermopathy most commonly affects the pretibial area in a symmetrical manner, with a raised waxy appearance similar to orange skin. Lesions in unusual sites are usually secondary to local trauma. Diffuse nonpitting edema or plaque forms are typical, but nodular or elephantiasic forms can occur. The lesions are usually asymptomatic but can be painful or pruritic occasionally. Thyroid acropachy is characterized by digital clubbing alone in most cases. Less commonly the presence of digital swelling and periosteal reaction of underlying bones constitute the full form, resembling hypertrophic pulmonary osteoarthropathy. The diagnosis of Graves dermopathy and thyroid acropachy is usually clinical, and biopsy is rarely needed. The histology of dermopathy reveals accumulation of GAG and mucin-like substances in the reticular dermis, together with marked edema similar to that of GO.

Smoking cessation should be strongly advised. Weight reduction in obese patients may improve PTM owing to reduction in venous stasis and dependency effect. Specific treatment of Graves dermopathy is indicated if symptomatic or raising cosmetic concerns. Local GC therapy in the form of topical GCs under occlusive dressing or intralesional injections can induce remission of cutaneous lesions and possibly prevent fibrosis or lymphatic obstruction. Systemic immunosuppressive therapies employed in active GO can also improve dermopathy. A recent case series reported that the administration of RTX or intravenous immunoglobulin resulted in improvement or stabilization of PTM.[126] Thyroid acropachy usually does not require any specific treatment. Long-term remission appears to be determined by the disease severity at presentation, instead of the effect of specific treatment.

SUMMARY AND PERSPECTIVES

Over the past two decades, we have witnessed remarkable advances in the understanding and clinical management of GO. It is reasonable to recommend the combination of IVGCs and mycophenolate as the new standard of care in active moderate-to-severe GO, in view of its practicability and superior efficacy to conventional treatment. Other combinations of widely available and affordable therapies may also be beneficial in GO, (e.g., IVGC-ORT [ongoing trial: ClinicalTrials.gov: NCT02339142] and IVGC with RTX or methotrexate).[36] Biologics (e.g,. TCZ, TPT) appear highly promising both in treatment-naïve and -resistant cases, although long-term efficacy and safety data are still pending. The roles of other existing targeted therapies in GO are also worth exploring (e.g., anti-IL23 or anti-IL17 agents for the IL-23/IL-17/Th17 axis, sirolimus for the mTOR pathway). Furthermore, several novel therapeutic agents can potentially treat Graves hyperthyroidism and associated orbitopathy simultaneously by targeting their shared immunological mechanisms (e.g., antigen-specific immunotherapy ATX-GD-59; anti-CD40 monoclonal antibody iscalimab; TSH-R antagonists).[36] Nonetheless, all new treatments should be carefully examined in RCTs (vs. placebo and/or standard of care), preferably with standardized primary and secondary outcomes, to draw sound conclusions on the efficacy of certain intervention and facilitate comparisons among different trials.[127] Therapeutic advances will continue to optimize clinical care of GO patients in a holistic manner.

REFERENCES

1. Perros P, Hegedus L, Bartalena L, et al. Graves' orbitopathy as a rare disease in Europe: a European Group on Graves' Orbitopathy (EUGOGO) position statement. *Orphanet J Rare Dis*. 2017;12:72.
2. Chin YH, Ng CH, Lee MH, et al. Prevalence of thyroid eye disease in Graves' disease: a meta-analysis and systematic review. *Clin Endocrinol*. 2020;93:363–374.
3. Perros P, Zarkovic M, Azzolini C, et al. PREGO (presentation of Graves' orbitopathy) study: changes in referral patterns to European Group on Graves' Orbitopathy (EUGOGO) centres over the period from 2000 to 2012. *Br J Ophthalmol*. 2015;99:1531–1535.
4. Quinn AS, Dujardin LR, Knight B, et al. Changing trend in referral to secondary care specialist thyroid eye disease clinic following the Amsterdam declaration. *Int Ophthalmol*. 2018;38:301–306.
5. Woo KI, Kim YD, Lee SY. Prevalence and risk factors for thyroid eye disease among Korean dysthyroid patients. *Korean J Ophthalmol*. 2013;27:397–404.
6. Khong JJ, Finch S, De Silva C, et al. Risk factors for graves' orbitopathy; the Australian thyroid-associated orbitopathy Research (ATOR) study. *J Clin Endocrinol Metab*. 2016;101:2711–2720.
7. Laurberg P, Berman DC, Bulow Pedersen I, et al. Incidence and clinical presentation of moderate to severe graves' orbitopathy in a Danish population before and after iodine fortification of salt. *J Clin Endocrinol Metab*. 2012;97:2325–2332.
8. Wiersinga WM, Kahaly G. *Graves' Orbitopathy: A Multidisciplinary Approach: Questions and Answers*. 3rd ed. Basel, New York: Karger; 2017:60–274.
9. McKeag D, Lane C, Lazarus JH, et al. Clinical features of dysthyroid optic neuropathy: a European Group on Graves' Orbitopathy (EUGOGO) survey. *Br J Ophthalmol*. 2007;91:455–458.
10. Wong Y, Dickinson J, Perros P, et al. A British Ophthalmological Surveillance Unit (BOSU) study into dysthyroid optic neuropathy in the United Kingdom. *Eye*. 2018;32:1555–1562.
11. Tramunt B, Imbert P, Grunenwald S, et al. Sight-threatening Graves' orbitopathy: twenty years' experience of a multidisciplinary thyroid-eye outpatient clinic. *Clin Endocrinol*. 2019;90:208–213.
12. Bartalena L, Baldeschi L, Boboridis K, et al. The 2016 European thyroid association/European group on Graves' orbitopathy guidelines for the management of graves' orbitopathy. *Eur Thyroid J*. 2016;5:9–26.
13. Wiersinga W, Zarkovic M, Bartalena L, et al. Predictive score for the development or progression of Graves' orbitopathy in patients with newly diagnosed Graves' hyperthyroidism. *Eur J Endocrinol*. 2018;178:635–643.
14. Acharya SH, Avenell A, Philip S, et al. Radioiodine therapy (RAI) for Graves' disease (GD) and the effect on ophthalmopathy: a systematic review. *Clin Endocrinol*. 2008;69:943–950.
15. Ponto KA, Zang S, Kahaly GJ. The tale of radioiodine and Graves' orbitopathy. *Thyroid*. 2010;20:785–793.
16. Laurberg P, Wallin G, Tallstedt L, et al. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*. 2008;158:69–75.
17. Vannucchi G, Covelli D, Campi I, et al. Prevention of orbitopathy by oral or intravenous steroid prophylaxis in short duration Graves' disease patients undergoing radioiodine ablation: a prospective randomized control trial study. *Thyroid*. 2019;29:1828–1833.
18. Sabini E, Mazzi B, Profilo MA, et al. High serum cholesterol is a novel risk factor for Graves' orbitopathy: results of a cross-sectional study. *Thyroid*. 2018;28:386–394.
19. Lanzolla G, Sabini E, Profilo MA, et al. Relationship between serum cholesterol and Graves' orbitopathy (GO): a confirmatory study. *J Endocrinol Invest*. 2018;41:1417–1423.
20. Stein JD, Childers D, Gupta S, et al. Risk factors for developing thyroid-associated ophthalmopathy among individuals with Graves disease. *JAMA Ophthalmol*. 2015;133:290–296.
21. Lanzolla G, Vannucchi G, Ionni I, et al. Cholesterol serum levels and use of statins in Graves' orbitopathy: a new starting point for the therapy. *Front Endocrinol*. 2019;10:933.
22. Hai YP, Lee ACH, Frommer L, et al. Immunohistochemical analysis of human orbital tissue in Graves' orbitopathy. *J Endocrinol Invest*. 2020;43:123–137.
23. Lacheta D, Miskiewicz P, Gluszko A, et al. Immunological aspects of Graves' ophthalmopathy. *BioMed Res Int*. 2019;2019:7453260.
24. Salvi M, Covelli D. B cells in Graves' Orbitopathy: more than just a source of antibodies? *Eye*. 2019;33:230–234.
25. Stan MN, Salvi M. Management of endocrine disease: rituximab therapy for Graves' orbitopathy—lessons from randomized control trials. *Eur J Endocrinol*. 2017;176:R101–R109.
26. Huang Y, Fang S, Li D, et al. The involvement of T cell pathogenesis in thyroid-associated ophthalmopathy. *Eye*. 2019;33:176–182.
27. Zhang M, Ding X, Wu LP, et al. A promising mouse model of Graves' orbitopathy induced by adenovirus expressing thyrotropin receptor A subunit. *Thyroid*. 2020;31:638–648.
28. Virakul S, van Steensel L, Dalm VA, et al. Platelet-derived growth factor: a key factor in the pathogenesis of Graves' ophthalmopathy and potential target for treatment. *Eur Thyroid J*. 2014;3:217–226.
29. Taylor PN, Zhang L, Lee RWJ, et al. New insights into the pathogenesis and nonsurgical management of Graves' orbitopathy. *Nat Rev Endocrinol*. 2020;16:104–116.
30. Zhang L, Grennan-Jones F, Draman MS, et al. Possible targets for non-immunosuppressive therapy of Graves' orbitopathy. *J Clin Endocrinol Metab*. 2014;99:E1183–E1190.
31. Zhang L, Ji QH, Ruge F, et al. Reversal of pathological features of Graves' orbitopathy by activation of forkhead transcription factors, FOXOs. *J Clin Endocrinol Metab*. 2016;101:114–122.
32. Smith TJ, Hegedus L. Graves' disease. *N Engl J Med*. 2016;375:1552–1565.
33. Smith TJ, Janssen J. Insulin-like growth factor-i receptor and thyroid-associated ophthalmopathy. *Endocr Rev*. 2019;40:236–267.
34. Smith TJ. Potential roles of CD34+ fibrocytes masquerading as orbital fibroblasts in thyroid-associated ophthalmopathy. *J Clin Endocrinol Metab*. 2019;104:581–594.
35. Diana T, Ponto KA, Kahaly GJ. Thyrotropin receptor antibodies and Graves' orbitopathy. *J Endocrinol Invest*. 2020;44:703–712.
36. Lee ACH, Kahaly GJ. Novel approaches for immunosuppression in Graves' hyperthyroidism and associated orbitopathy. *European Thyroid Journal*. 2020;9:17–30.
37. Neumann S, Krieger CC, Gershengorn MC. Targeting TSH and IGF-1 receptors to treat thyroid eye disease. *European Thyroid Journal*. 2020;9:59–65.
38. Krieger CC, Neumann S, Gershengorn MC. Is there evidence for IGF1R-stimulating Abs in Graves' orbitopathy pathogenesis? *Int J Mol Sci*. 2020;21:6561.
39. Davies TF, Andersen S, Latif R, et al. Graves' disease. *Nat Rev Dis Primers*. 2020;6:52.
40. Du P, Ma X, Wang C. Associations of CTLA4 gene polymorphisms with Graves' ophthalmopathy: a meta-analysis. *Int J Genomics*. 2014;2014:537969.
41. Planck T, Shahida B, Sjogren M, et al. Association of BTG2, CYR61, ZFP36, and SCD gene polymorphisms with Graves' disease and ophthalmopathy. *Thyroid*. 2014;24:1156–1161.
42. Wong KH, Rong SS, Chong KK, et al. Genetic associations of interleukin-related genes with Graves' ophthalmopathy: a systematic review and meta-analysis. *Sci Rep*. 2015;5:16672.

43. Bufalo NE, Dos Santos RB, Marcello MA, et al. TSHR intronic polymorphisms (rs179247 and rs12885526) and their role in the susceptibility of the Brazilian population to Graves' disease and Graves' ophthalmopathy. *J Endocrinol Invest*. 2015;38:555–561.
44. Wu S, Cai T, Chen F, et al. Genetic associations of FCRL3 polymorphisms with the susceptibility of Graves ophthalmopathy in a Chinese population. *Int J Clin Exp Med*. 2015;8:16948–16954.
45. Lee JY, Kim NK, Cho YW, et al. Association between methylenetetrahydrofolate reductase (MTHFR) polymorphisms and susceptibility to Graves' ophthalmopathy. *Mol Med Rep*. 2016;14:2276–2282.
46. Zhang M, Liu S, Xu J, et al. TNFSF15 polymorphisms are associated with Graves' disease and Graves' ophthalmopathy in a Han Chinese population. *Curr Eye Res*. 2020;45:888–895.
47. Huang X, Liu G, Mei S, et al. Human leucocyte antigen alleles confer susceptibility and progression to Graves' ophthalmopathy in a Southern Chinese population. *Br J Ophthalmol*. 2021; 105:1462–1468.
48. Perros P, Zarkovic MP, Panagiotou GC, et al. Asymmetry indicates more severe and active disease in Graves' orbitopathy: results from a prospective cross-sectional multicentre study. *J Endocrinol Invest*. 2020;43: 1717–1722.
49. Ponto KA, Binder H, Diana T, et al. Prevalence, phenotype, and psychosocial well-being in euthyroid/hypothyroid thyroid-associated orbitopathy. *Thyroid*. 2015;25:942–948.
50. Leo M, Menconi F, Rocchi R, et al. Role of the underlying thyroid disease on the phenotype of Graves' orbitopathy in a tertiary referral center. *Thyroid*. 2015;25:347–351.
51. Bartley GB, Fatourechi V, Kadrmas EF, et al. Chronology of Graves' ophthalmopathy in an incidence cohort. *Am J Ophthalmol*. 1996;121: 426–434.
52. Eckstein AK, Losch C, Glowacka D, et al. Euthyroid and primarily hypothyroid patients develop milder and significantly more asymmetrical Graves ophthalmopathy. *Br J Ophthalmol*. 2009;93:1052–1056.
53. Kahaly GJ, Diana T, Olivo PD. Tsh receptor antibodies: relevance & utility. *Endocr Pract*. 2020;26:97–106.
54. Lytton SD, Ponto KA, Kanitz M, et al. A novel thyroid stimulating immunoglobulin bioassay is a functional indicator of activity and severity of Graves' orbitopathy. *J Clin Endocrinol Metab*. 2010;95:2123–2131.
55. Marino M, Ionni I, Lanzolla G, et al. Orbital diseases mimicking Graves' orbitopathy: a long-standing challenge in differential diagnosis. *J Endocrinol Invest*. 2020;43:401–411.
56. Ponto KA, Kanitz M, Olivo PD, et al. Clinical relevance of thyroid-stimulating immunoglobulins in Graves' ophthalmopathy. *Ophthalmology*. 2011;118:2279–2285.
57. Kampmann E, Diana T, Kanitz M, et al. Thyroid stimulating but not blocking autoantibodies are highly prevalent in severe and active thyroid-associated orbitopathy: a prospective study. *Int J Endocrinol*. 2015;2015:678194.
58. Nicoli F, Lanzolla G, Mantuano M, et al. Correlation between serum anti-TSH receptor autoantibodies (TRAbs) and the clinical feature of Graves' orbitopathy. *J Endocrinol Invest*. 2020;44:581–585.
59. Ponto KA, Diana T, Binder H, et al. Thyroid-stimulating immunoglobulins indicate the onset of dysthyroid optic neuropathy. *J Endocrinol Invest*. 2015;38:769–777.
60. Kahaly GJ, Wuster C, Olivo PD, et al. High titers of thyrotropin receptor antibodies are associated with orbitopathy in patients with Graves disease. *J Clin Endocrinol Metab*. 2019;104:2561–2568.
61. Kahaly GJ. Imaging in thyroid-associated orbitopathy. *Eur J Endocrinol*. 2001;145:107–118.
62. Muller-Forell W, Kahaly GJ. Neuroimaging of Graves' orbitopathy. *Best Pract Res Clin Endocrinol Metab*. 2012;26:259–271.
63. Siakallis LC, Uddin JM, Miszkiel KA. Imaging investigation of thyroid eye disease. *Ophthalmic Plast Reconstr Surg*. 2018;34:S41–S51.
64. Lo C, Ugradar S, Rootman D. Management of Graves myopathy: orbital imaging in thyroid-related orbitopathy. *JAAPOS*. 2018;22:e1–e9. 256.
65. North VS, Freitag SK. A review of imaging modalities in thyroid-associated orbitopathy. *Int Ophthalmol Clin*. 2019;59:81–93.
66. Zhou M, Shen L, Jiao Q, et al. Role of magnetic resonance imaging in the assessment of active thyroid-associated ophthalmopathy patients with long disease duration. *Endocr Pract*. 2019;25:1268–1278.
67. Roos JCP, Murthy R. Update on the clinical assessment and management of thyroid eye disease. *Curr Opin Ophthalmol*. 2019;30:401–406.
68. Hoppe E, Lee ACH, Hoppe D, et al. Predictive factors for changes in quality of life after steroid treatment for active moderate-to-severe Graves' orbitopathy: a prospective trial. *European Thyroid Journal*. 2020;9:313–320.
69. Lee THB, Sundar G. Quality of life in thyroid eye disease: a systematic review. *Ophthalmic Plast Reconstr Surg*. 2019;36:118–126.
70. Mellington FE, Dayan CM, Dickinson AJ, et al. Management of thyroid eye disease in the United Kingdom: a multi-centre thyroid eye disease audit. *Orbit*. 2017;36:159–169.
71. Kahaly GJ, Bartalena L, Hegedus L, et al. European Thyroid Association guideline for the management of Graves' hyperthyroidism. *Eur Thyroid J*. 2018;7:167–186. 2018.
72. Ross DS, Burch HB, Cooper DS, et al. American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421. 2016.
73. Zarkovic M, Wiersinga W, Perros P, et al. Antithyroid drugs in Graves' hyperthyroidism: differences between "block and replace" and "titration" regimes in frequency of euthyroidism and Graves' orbitopathy during treatment. *J Endocrinol Invest*. 2021;44:371–378.
74. Winther KH, Rayman MP, Bonnema SJ, et al. Selenium in thyroid disorders—essential knowledge for clinicians. *Nat Rev Endocrinol*. 2020;16:165–176.
75. Rotondo Dottore G, Leo M, Casini G, et al. Antioxidant actions of selenium in orbital fibroblasts: a basis for the effects of selenium in Graves' orbitopathy. *Thyroid*. 2017;27:271–278.
76. Marcocci C, Kahaly GJ, Krassas GE, et al. Selenium and the course of mild Graves' orbitopathy. *N Engl J Med*. 2011;364:1920–1931.
77. Negro R, Hegedus L, Attanasio R, et al. European Thyroid Association survey on the use of selenium supplementation in Graves' hyperthyroidism and Graves' orbitopathy. *Eur Thyroid J*. 2018;8:7–15. 2019.
78. Zang S, Ponto KA, Kahaly GJ. Clinical review: intravenous glucocorticoids for Graves' orbitopathy: efficacy and morbidity. *J Clin Endocrinol Metab*. 2011;96:320–332.
79. Wiersinga WM, Kahaly G. *Graves' Orbitopathy: A Multidisciplinary Approach: Questions and Answers*. 3rd ed. Basel, New York: Karger; 2017:140–176.
80. Yang M, Wiersinga WM, Soeters MR, et al. What is the aim of immunosuppressive treatment in patients with Graves' orbitopathy? *Ophthalmic Plast Reconstr Surg*. 2014;30:157–161.
81. Godfrey KJ, Kazim M. Radiotherapy for active thyroid eye disease. *Ophthalmic Plast Reconstr Surg*. 2018;34:S98–S104.
82. Kahaly GJ, Rosler HP, Pitz S, et al. Low- versus high-dose radiotherapy for Graves' ophthalmopathy: a randomized, single blind trial. *J Clin Endocrinol Metab*. 2000;85:102–108.
83. Rajendram R, Taylor PN, Wilson VJ, et al. Combined immunosuppression and radiotherapy in thyroid eye disease (CIRTED): a multicentre, 2 x 2 factorial, double-blind, randomised controlled trial. *Lancet Diabetes Endocrinol*. 2018;6:299–309.
84. Shams PN, Ma R, Pickles T, et al. Reduced risk of compressive optic neuropathy using orbital radiotherapy in patients with active thyroid eye disease. *Am J Ophthalmol*. 2014;157:1299–1305.
85. Kim JW, Han SH, Son BJ, et al. Efficacy of combined orbital radiation and systemic steroids in the management of Graves' orbitopathy. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:991–998.
86. Gold KG, Scofield S, Isaacson SR, et al. Orbital radiotherapy combined with corticosteroid treatment for thyroid eye disease-compressive optic neuropathy. *Ophthalmic Plast Reconstr Surg*. 2018;34:172–177.
87. Kim JW, Lee KH, Woo YJ, et al. The effect of systemic steroids and orbital radiation for active Graves' orbitopathy on postdecompression extraocular muscle volume. *Am J Ophthalmol*. 2016;171:11–17.
88. Boulanouar L, Grunenwald S, Imbert P, et al. Effect of orbital radiotherapy on the outcome of surgical orbital decompression for thyroid-associated orbitopathy (TAO): a retrospective study in 136 patients. *Endocrine*. 2020;67:605–612.
89. Oeverhaus M, Witteler T, Lax H, et al. Combination therapy of intravenous steroids and orbital irradiation is more effective than intravenous steroids alone in patients with Graves' orbitopathy. *Horm Metab Res*. 2017;49:739–747.

90. Marcocci C, Bartalena L, Rocchi R, et al. Long-term safety of orbital radiotherapy for Graves' ophthalmopathy. *J Clin Endocrinol Metab*. 2003;88:3561–3566.
91. Wakelkamp IM, Tan H, Saeed P, et al. Orbital irradiation for Graves' ophthalmopathy: is it safe? A long-term follow-up study. *Ophthalmology*. 2004;111:1557–1562.
92. Marquez SD, Lum BL, McDougall IR, et al. Long-term results of irradiation for patients with progressive Graves' ophthalmopathy. *Int J Radiat Oncol Biol Phys*. 2001;51:766–774.
93. Schaefer U, Hesselmann S, Micke O, et al. A long-term follow-up study after retro-orbital irradiation for Graves' ophthalmopathy. *Int J Radiat Oncol Biol Phys*. 2002;52:192–197.
94. Kapamajian MA. Medical or surgical decompression of optic neuropathy in Graves' ophthalmopathy: the choice remains unclear. *Clin Endocrinol*. 2006;65:132.
95. Curro N, Covelli D, Vannucchi G, et al. Therapeutic outcomes of high-dose intravenous steroids in the treatment of dysthyroid optic neuropathy. *Thyroid*. 2014;24:897–905.
96. Dolman PJ. Dysthyroid optic neuropathy: evaluation and management. *J Endocrinol Invest*. 2020;44:421–429.
97. Rose GE. Postural visual obscurations in patients with inactive thyroid eye disease; a variant of 'hydraulic' disease. *Eye*. 2006;20:1178–1185.
98. Kahaly GJ, Riedl M, Konig J, et al. Mycophenolate plus methylprednisolone versus methylprednisolone alone in active, moderate-to-severe Graves' orbitopathy (MINGO): a randomised, observer-masked, multicentre trial. *Lancet Diabetes Endocrinol*. 2018;6:287–298.
99. Ye X, Bo X, Hu X, et al. Efficacy and safety of mycophenolate mofetil in patients with active moderate-to-severe Graves' orbitopathy. *Clin Endocrinol*. 2017;86:247–255.
100. Lee ACH, Riedl M, Frommer L, et al. Systemic safety analysis of mycophenolate in Graves' orbitopathy. *J Endocrinol Invest*. 2020;43:767–777.
101. Salvi M, Vannucchi G, Curro N, et al. Efficacy of B-cell targeted therapy with rituximab in patients with active moderate to severe Graves' orbitopathy: a randomized controlled study. *J Clin Endocrinol Metab*. 2015;100:422–431.
102. Stan MN, Garrity JA, Carranza Leon BG, et al. Randomized controlled trial of rituximab in patients with Graves' orbitopathy. *J Clin Endocrinol Metab*. 2015;100:432–441.
103. Du Pasquier-Fediaevsky L, Andrei S, Berche M, et al. Low-dose rituximab for active moderate to severe Graves' orbitopathy resistant to conventional Treatment. *Ocul Immunol Inflamm*. 2019;27:844–850.
104. Eid L, Coste-Verdier V, Longueville E, et al. The effects of Rituximab on Graves' orbitopathy: a retrospective study of 14 patients. *Eur J Ophthalmol*. 2020;30:1008–1013.
105. Deltour JB, d'Assigny Flamen M, Ladsous M, et al. Efficacy of rituximab in patients with Graves' orbitopathy: a retrospective multicenter nationwide study. *Graefes Arch Clin Exp Ophthalmol*. 2020;258:2013–2021.
106. Chen B, Tsui S, Smith TJ. IL-1 beta induces IL-6 expression in human orbital fibroblasts: identification of an anatomic-site specific phenotypic attribute relevant to thyroid-associated ophthalmopathy. *J Immunol*. 2005;175:1310–1319.
107. Hwang CJ, Afifiyan N, Sand D, et al. Orbital fibroblasts from patients with thyroid-associated ophthalmopathy overexpress CD40: CD154 hyperinduces IL-6, IL-8, and MCP-1. *Invest Ophthalmol Vis Sci*. 2009;50:2262–2268.
108. Khong JJ, McNab AA, Ebeling PR, et al. Pathogenesis of thyroid eye disease: review and update on molecular mechanisms. *Br J Ophthalmol*. 2016;100:142–150.
109. Kumar S, Schiefer R, Coenen MJ, et al. A stimulatory thyrotropin receptor antibody (M22) and thyrotropin increase interleukin-6 expression and secretion in Graves' orbital preadipocyte fibroblasts. *Thyroid*. 2010;20:59–65.
110. Jyonouchi SC, Valyasevi RW, Harteneck DA, et al. Interleukin-6 stimulates thyrotropin receptor expression in human orbital preadipocyte fibroblasts from patients with Graves' ophthalmopathy. *Thyroid*. 2001;11:929–934.
111. Fang S, Huang Y, Zhong S, et al. Regulation of orbital fibrosis and adipogenesis by pathogenic Th17 cells in Graves' orbitopathy. *J Clin Endocrinol Metab*. 2017;102:4273–4283.
112. Fang S, Huang Y, Wang N, et al. Insights into local orbital immunity: evidence for the involvement of the Th17 cell pathway in thyroid-associated ophthalmopathy. *J Clin Endocrinol Metab*. 2019;104:1697–1711.
113. Jones BE, Maerz MD, Buckner JH. IL-6: a cytokine at the crossroads of autoimmunity. *Curr Opin Immunol*. 2018;55:9–14.
114. Molnar I, Balazs C. High circulating IL-6 level in Graves' ophthalmopathy. *Autoimmunity*. 1997;25:91–96.
115. Kishazi E, Dor M, Eperon S, et al. Differential profiling of lacrimal cytokines in patients suffering from thyroid-associated orbitopathy. *Sci Rep*. 2018;8:10792.
116. Xu N, Cui Y, Fu D, et al. Tear inflammatory cytokines and ocular surface changes in patients with active thyroid eye disease treated with high-dose intravenous glucocorticoids. *J Endocrinol Invest*. 2020;43: 901–910.
117. Perez-Moreiras JV, Gomez-Reino JJ, Maneiro JR, et al. Efficacy of tocilizumab in patients with moderate-to-severe corticosteroid-resistant Graves' orbitopathy: a randomized clinical trial. *Am J Ophthalmol*. 2018;195:181–190.
118. Sanchez-Bilbao L, Martinez-Lopez D, Revenga M, et al. Anti-IL-6 receptor tocilizumab in refractory Graves' orbitopathy: national multicenter observational study of 48 patients. *J Clin Med*. 2020;9:2816.
119. Bennedjai A, Bouheraoua N, Gatfosse M, et al. Tocilizumab versus rituximab in patients with moderate to severe steroid-resistant Graves' orbitopathy. *Ocul Immunol Inflamm*. 2020:1—6.
120. Smith TJ, Kahaly GJ, Ezra DG, et al. Teprotumumab for thyroid-associated ophthalmopathy. *N Engl J Med*. 2017;376:1748–1761.
121. Douglas RS, Kahaly GJ. Teprotumumab for active thyroid eye disease. *Reply. N Engl J Med*. 2020;382:1959–1960.
122. Smith TJ, Bartalena L. Will biological agents supplant systemic glucocorticoids as the first-line treatment for thyroid-associated ophthalmopathy? *Eur J Endocrinol*. 2019;181:D27–D43.
123. Kahaly GJ. Management of Graves thyroidal and extrathyroidal disease: an update. *J Clin Endocrinol Metab*. 2020;105:3704–3720.
124. Fatourechi V. Thyroid dermopathy and acropachy. *Best Pract Res Clin Endocrinol Metab*. 2012;26:553–565.
125. Bartalena L, Fatourechi V. Extrathyroidal manifestations of Graves' disease: a 2014 update. *J Endocrinol Invest*. 2014;37:691–700.
126. Kotwal A, Turcu AF, Sonawane V, et al. Clinical experience with rituximab and intravenous immunoglobulin for pretibial myxedema: a case series. *Thyroid*. 2019;29:692–699.
127. Bartalena L, Wiersinga WM. Proposal for standardization of primary and secondary outcomes in patients with active, moderate-to-severe Graves' orbitopathy. *European Thyroid Journal*. 2020;9:3–16.

73

Thyroiditis

Ana M. Ramos-Leví and Mónica Marazuela

OUTLINE

Introduction: Concept and Definition of Thyroiditis, 1214
Subtypes of Thyroiditis, 1215
Hashimoto's Thyroiditis, 1215
  Definition and Incidence, 1215
  Etiopathogenesis and Physiopathologic Mechanisms in Autoimmune Thyroid Disease, 1216
  Stages in the Development of Autoimmune Thyroid Disease, 1220
  Natural History and Clinical Features, 1221
  Relationship With Other Types of Thyroiditis and Graves' Disease, 1221
  Relationship With Other Autoimmune Diseases, 1222
Hashimoto's Thyroiditis and Thyroid Malignancy, 1223
  Diagnosis, 1223
  Treatment and Follow-Up, 1224
  Hashimoto's Encephalopathy, 1225
Acute/Infectious Thyroiditis, 1225
  Definition and Incidence, 1225
  Clinical and Diagnostic Features, 1225
  Treatment and Follow-Up, 1226
Subacute Thyroiditis, 1226
  Definition and Incidence, 1226
  Clinical Features and Phases: Inflammatory, Recovery, Stabilization, 1226
  Diagnosis: Clinical, Laboratory, and Imaging Findings and Differential Diagnosis, 1226
  Pathology, 1227
  Treatment and Follow-Up, 1227
Silent Thyroiditis, 1227
  Definition and Incidence, 1227
  Clinical Features, 1227
  Diagnosis: Clinical, Laboratory, and Imaging Findings, 1228
  Treatment, 1228
Riedel's Thyroiditis, 1228
  Definition and Incidence, 1228
  Etiology, 1228
  Clinical Features, 1229
  Diagnosis: Laboratory and Imaging Findings, 1229
  Pathology, 1229
  Treatment, 1229
IgG4 Thyroiditis, 1229
  Definition, Incidence, and Relationships in the Setting of Other Subtypes of Thyroiditis, 1229
  Clinical Aspects, 1230
  Diagnosis, 1230
  Treatment and Follow-Up, 1230
Summary and Future Directions, 1230

INTRODUCTION: CONCEPT AND DEFINITION OF THYROIDITIS

The term "thyroiditis" refers to several generally benign thyroid disorders with the common feature of some form of inflammation of the thyroid gland, which may sometimes be accompanied by thyroid function perturbations. Its clinical spectrum may vary from an absence of clinically evident inflammation, apart from a biochemical thyroid dysfunction or goiter, to significant symptomatic subtypes, with acute pain and overall illness. The terminology used in thyroid autoimmune disorders may sometimes insufficiently reflect our knowledge on their physiopathology, and even regard as opposites diseases that are truly part of the same spectrum, for instance Hashimoto's thyroiditis (HT) and Graves' disease (GD). In any case, information regarding its pathogenesis is still being revealed.

Thyroidites have been classified according to different criteria, including known or suspected etiology, pathology, the existence of pain and tenderness, or the effect on thyroid function.

KEY POINTS

- The term "thyroiditis" refers to some form of thyroid inflammation with variable repercussions for thyroid function.
- Thyroidites are classified according to their etiology, pathology, the existence of pain, or the effect on thyroid function.
- The clinical course of subacute thyroiditis frequently fluctuates. The characteristic sequence involves an initial thyrotoxic phase, followed by hypothyroidism and eventual normalization.
- The presence of pain, the clinical course, evaluation of thyroid peroxidase antibody levels, erythrocyte sedimentation rate, white blood cell count, and thyroid hormones, as well as thyroid ultrasound and radioactive iodine uptake, are the mainstay for the differential diagnosis between the different subtypes of thyroiditis.

This chapter reviews the different subtypes of thyroiditis, with special reference to autoimmunity as the main cause of their development.

TABLE 73.1 Terminology for Thyroiditis

Type	Synonyms
Hashimoto's thyroiditis	Chronic lymphocytic thyroiditis Chronic autoimmune thyroiditis Lymphadenoid goiter
Postpartum thyroiditis	Painless postpartum thyroiditis Subacute lymphocytic thyroiditis
Painless sporadic thyroiditis	Silent sporadic thyroiditis Subacute lymphocytic thyroiditis Lymphocytic thyroiditis with spontaneously resolving hyperthyroidism
Painful subacute thyroiditis	Subacute de Quervain's thyroiditis Subacute nonsuppurative thyroiditis Giant-cell thyroiditis Subacute granulomatous thyroiditis Pseudogranulomatous thyroiditis
Suppurative thyroiditis	Infectious thyroiditis Acute suppurative thyroiditis Pyrogenic thyroiditis Bacterial thyroiditis
Drug-induced thyroiditis	Amiodarone, lithium, IFN-α, IL-2, tyrosine kinase inhibitors, checkpoint inhibitor immunotherapy
Radiation-induced thyroiditis	
Palpation- or trauma-induced thyroiditis	
Riedel's thyroiditis	Fibrous thyroiditis IgG4-induced thyroiditis

SUBTYPES OF THYROIDITIS

One of the main concerns when referring to thyroiditis is the controversy regarding the nomenclature used to categorize the different forms. However, in general, there is agreement on the use of several synonyms for equivalent entities (Table 73.1).

Most classifications categorize the different forms of thyroiditis according to the presence of thyroid pain and tenderness. In this regard, de Quervain's subacute thyroiditis (SAT), suppurative thyroiditis, radiation-induced thyroiditis, and palpation- or trauma-induced thyroiditis may involve pain to different extents, while the rest are generally clinically asymptomatic.[1] Table 73.2 summarizes the main differential characteristics between the various subtypes of thyroiditis, which will be discussed in detail in the corresponding sections. A frequent feature regarding the clinical course in thyroiditis is fluctuation; in this regard, painless postpartum thyroiditis, painless thyroiditis, and painful de Quervain's SAT may exhibit a characteristic sequence, which may involve thyrotoxicosis, hypothyroidism, or both (Fig. 73.1). On the contrary, although an initial euthyroid, or even slightly hyperthyroid, phase may occur in HT, the most frequent laboratory finding in this case is chronic hypothyroidism.

KEY POINTS

- Hashimoto's thyroiditis is a chronic autoimmune inflamatory infiltrate of the thyroid gland, with destruction of thyroid follicular cells, in the setting of thyroid auto-antibodies.

HASHIMOTO'S THYROIDITIS

Definition and Incidence

The term HT, or chronic autoimmune thyroiditis, implies an inflammatory infiltrate of the thyroid gland, destruction of thyroid follicular cells, and the presence of thyroid antibodies, mainly against thyroid peroxidase (TPOAb). It is a long-term and lifelong autoimmune disease in which the thyroid gland enlarges and often atrophies, leading to hypothyroidism.

Hakaru Hashimoto's, in 1912, was the first to report four patients with diffuse goiter and outline four histologic characteristics: diffuse lymphocytic infiltration, formation of lymphoid follicles, destruction of epithelial cells, and proliferation of fibrous tissue.

The term "Hashimoto's disease" or HT is sometimes used to refer only to goitrous thyroiditis, but it is usually considered to be a broad synonym for chronic thyroiditis or autoimmune thyroiditis, including atrophic and nongoitrous thyroiditis.

Hashimoto's disease is the world's most frequent cause of hypothyroidism in iodine-sufficient conditions. Overall, the disorder may affect up to 2% of the general population, is more common in older women, and is ten times more frequent in women than in men; in fact, 3% to 10% of women have been described as exhibiting chronic autoimmune thyroiditis.[2,3]

KEY POINTS

- Hashimoto's thyroiditis (HT) is a chronic autoimmune thyroiditis with thyroid inflammatory infiltration, destruction of thyroid follicular cells, and positivity for thyroid antibodies.
- HT is the most frequent cause of hypothyroidism on iodine-replete areas.
- Positivity for thyroid peroxidase antibody levels is the hallmark for diagnosis of HT.
- Several environmental and existential factors may trigger the onset of aberrant immune mechanisms in genetically predisposed individuals that eventually lead to autoimmune thyroid disease.

KEY POINTS

- Autoimmune thyroid disease is triggered and perpetuated due to a complex reciprocal relationship between humoral and cellular mechanisms.

TABLE 73.2 Main Differentiating Characteristics Between the Different Subtypes of Thyroiditis

	Etiology	Epidemiology	Clinical picture	Diagnosis	Treatment
Acute, suppurative	Infection (bacterial most frequent)	Prior upper respiratory infection No difference between sexes Children and 20–40 y. Association to pyriform sinus fistula	Sudden, local symptoms (pain, swelling, dysphagia, etc.) Generalized symptoms (fever) Mild thyrotoxicosis very rare	↑ESR. ↑WBC. TSH/FT4 normal TPOAb and TgAb negative FNA (gram-positive staining) and culture	Antibiotics FNA and/or drainage
Painful, subacute (de Quervain's)	Viral (almost)	Female predominance (5:1) 20–60 y. Possible seasonal incidence related to prior viral infection. Associated with HLA-B35	Prodromal generalized symptoms (flulike) Local pain, extending to jaw and ears Initial hyperthyroid phase followed by hypothyroidism	↓TSH/↑FT4. ↑↑ESR, ↑CRP Negative TPOAb and TgAb. ↓RAIU FNA: inflammatory infiltrate (+ giant cells) Biopsy: inflammatory infiltrate (+ giant cells), granulomas > fibrosis	NSAID, especially aspirin 1–3 g/day. Glucocorticoids Beta-blockers
Painless, silent, subacute	Autoimmune Postpartum	Female predominance (2:1) All ages, peak 30–40 y. Thyroid "aggressors" Potential progression to chronic	Initial hyperthyroid phase followed by hypothyroidism. No local symptoms	Abnormal TSH/FT4 ↑Thyroglobulin. (+) TPOAb and TgAb (persistent) (-) TSH-R–Ab. Normal inflammation reactants ↓RAIU	Expectant (observation) Beta-blockers for hyperthyroid symptoms L-T4 replacement therapy if needed
Chronic (Hashimoto's)	Autoimmune	Middle-aged women Thyroid dysfunction precipitating factors	Goiter in the early phase Later hypothyroidism	Initial goiter and mild hyperthyroidism Then, ↑TSH/↓FT4. (++) TPOAb and TgAb Biopsy: inflammatory infiltrate + fibrosis + necrosis	L-T4 replacement therapy
Riedel	Fibrous, IgG4		Insidious compressive local symptoms Fibrosis in other extrathyroidal locations No signs of systemic inflammation Hypothyroidism	Biopsy: fibrosis	Glucocorticoids in the early phase Surgery for compressive symptoms and signs

↑ elevated; ↓: decreased; (+): positive; (-): negative.

ESR, Erythrocyte sedimentation rate; *WBC*, white blood cell; *CRP*, C-reactive protein; *TSH*, thyroid-stimulating hormone; *FT4*, free T4; *TPOAb*, antithyroperoxidase antibodies; *TgAb*, antithyroglobulin antibodies; *FNA*, fine-needle aspiration; *NSAID*, nonsteroidal antiinflammatory drugs; *RAIU*, radioactive iodine uptake; *L-T4*, levothyroxine.

Etiopathogenesis and Physiopathologic Mechanisms in Autoimmune Thyroid Disease

Autoimmune Abnormalities. The human immune system is comprised of billions of B and T lymphocytes in relatively equal proportions. B lymphocytes (immunoglobulin-producing cells) develop from precursor cells in fetal liver and bone marrow and are found in all lymphoid organs and in the bloodstream. They exhibit antigen-specific receptors on their surface membranes, and they generate and secrete specific antibodies following antigenic stimulation. B cells are continuously recirculating between the blood and lymphoid tissue, thus maximizing the chances of recognizing antigen.

T cells also develop from precursor cells in fetal liver and bone marrow and migrate to the thymus. T cell differentiation then initiates in the thymus in response to exposure to specific antigens and subsequent production of cytokines. In normal conditions, selection of nonautoreactive cells occurs by central immune tolerance mechanisms in the thymus itself. These naïve differentiated T cells are released from the thymus into peripheral lymphoid tissue, where they will await specific signals to further mature in response to antigens, costimulators, and cytokines. This differentiation implies that each lymphocyte subtype will eventually secrete a battery of specific cytokines, and, thus, generate subsequent specific molecular reactions, apart from serving as a potential identifier of each individual subpopulation.

Selection processes throughout these maturations phases, for which multiple cellular and humoral mechanisms are necessary, lead to elimination of autoreactive cells. The main participants in these reactions, apart from B and T cells and their specific subtype populations, include multiple chemokines, growth factors, antigen-presenting cells (APCs), natural killer (NK) cells, membrane receptors, surface molecules, and major histocompatibility complex (MHC) molecules.[4,5]

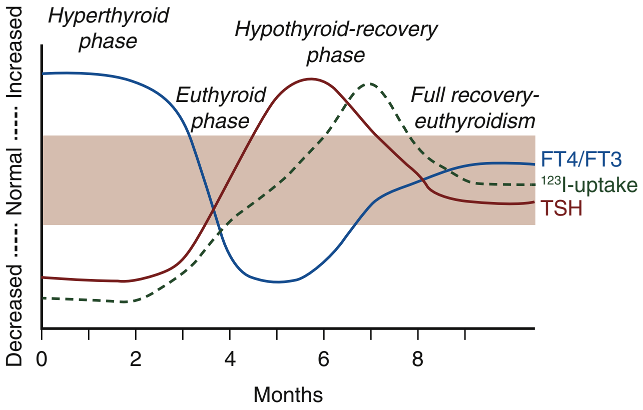


Fig. 73.1 **Graphical representation of the typical clinical course of various forms of thyroiditis, mainly painful subacute thyroiditis, painless postpartum thyroiditis, and painless sporadic thyroiditis.** Fluctuation is a frequent finding; patients with thyroiditis archetypally undergo an initial thyrotoxic phase, followed by hypothyroidism and eventual recovery. The horizontal axis represents time, and the vertical axis represents the quantity of thyroid-stimulating hormone *(green curve)*, FT4/FT3 *(orange curve),* and ^{123}I uptake *(dotted blue curve).* The green area represents the normal range of values. *TSH*, Thyroid-stimulating hormone.

However, some immune cells may escape these tolerance mechanisms, react to self antigens, and lead to proliferation of autoreactive clonal populations. Hypothesized explanations for this drift include genetic, environmental, and existential factors, which will be discussed in more detail later.[4,6,7] As a result, the concomitant action of B and T cells will generate autoantibodies and infiltrate the thyroid, reciprocally potentiating the autoimmune response with a mix of humoral and cellular mechanisms. Exosomes, microparticles, and miRNAs have also been associated with the onset and perpetuation of autoimmunity.[8-11] Some particularities occur in the specific setting of autoimmune thyroid disease (AITD).

KEY POINTS

- The typical clinical course of many forms of thyroiditis concerns an initial hyperthyroid thyrotoxic phase, followed by hypothyroidism and eventual recovery.

KEY POINTS

- The main auto-antibodies involved in thyroid autoimmune disease are anti-thyroid peroxidase (TPOAb) and anti-thyroglobulin (TgAb). Other antibodies such as anti-sodium/iodide symporter and anti-pendrin have been recently described.

Humoral Mechanisms. Detection of circulating thyroid autoantibodies has become the cornerstone for diagnosing AITD, in tandem with detection of autoreactive T cells. For clinical purposes, three main types of autoantibodies are considered, including those against TPOAb, those against thyroglobulin (TgAb), and those against the thyroid-stimulating hormone receptor (TSH-R–Ab); the latter is described in detail in the chapter on GD.[12]

Modern laboratory techniques allow a relatively precise quantification of these thyroid autoantibodies in serum, enabling a relatively specific diagnosis. Indeed, the more sensitive the assay, the more precise and antigen-specific it is. However, a considerable proportion of euthyroid individuals exhibit low titers of thyroid autoantibodies, meaning that the specificity of sensitive assays is reduced and making assessment of the total amount of thyroid autoantibodies more important. In this way, the higher the concentration of antibodies, the higher the specificity of the test. To optimally compare the results intra- and interindividually, there have been attempts to standardize assays, but each specific assay should be taken into account when considering them in the clinical follow-up.[13]

Both TPOAb and TgAb have been considered a result of thyroid damage, rather than the initial primary cause of thyroid destruction itself. In this regard, they are able to fix complement if they are of an appropriate IgG subclass, and thus facilitate complement-dependent antibody-mediated cytotoxicity, leading to thyroid tissue damage. Indeed, this cytotoxicity seems to lead to more damage to thyroid tissue than T cells and cytokine-mediated apoptosis.[13] In addition, antibody-mediated cytotoxicity also occurs through NK cells and macrophages. However, the resulting thyroid destruction may leave thyroglobulin (Tg) and thyroperoxidase (TPO) debris, perpetuating their significance as antigens in the subsequent development of thyroid autoantibodies. In fact, recent data show that thyroid antibodies may truly contribute to AITD outcome and chronicity.[14] TPOAb and TgAb are polyclonal IgG antibodies and are present in almost 100% of patients with AITD. In fact, almost all patients with HT have TPOAb. However, serum TPOAb and TgAb may be present in the general normal population, with a prevalence around 10% to 20% in relation to increasing age, with a clear difference between sexes (women are more prone to HT than men, with a 6:1 ratio), but with no clinically relevant thyroid disease. The significance of this finding may be related to an increased risk of future hypothyroidism, especially in the setting of a mildly elevated baseline thyroid-stimulating hormone (TSH).[15] For instance, subclinical autoimmune thyroiditis may occur after delivery in up to 50% of TPOAb-positive women. On the other hand, although few patients with AITD exhibit positivity for TgAb in the absence of TPOAb positivity, the opposite does not always occur. In this regard, because TPOAb levels are usually higher and bind with a higher affinity, they are usually considered the first choice if only one antibody is to be analyzed in screening for AITD.[13] All types of thyroid autoantibodies increase during AITD development if no treatment is established. Moreover, the process of thyroid hormone production itself may also contribute to the generation of immune-reactive C-terminal fragments, oxidative stress, presentation of epitopes by thyroid follicular cells, and further production of Tg and TPO autoimmune responsiveness.

Thyroid Peroxidase. TPO is expressed on the surface and cytoplasm of thyroid follicular cells and allows cell-surface antigen recognition, which may trigger complement-mediated cytotoxicity, as well as antibody-dependent cell-mediated cytotoxicity.[13] Binding to intracytoplasmic TPO suggests further access to this compartment, with potential immune consequences.

The role of TPO as a microsomal antigen was first described in 1985 by DeGroot's laboratory and Czarnocka et al. The interaction of human anti-TPO antisera and monoclonal antibodies denotes the presence of several B cell epitopes, with critical regions involving individual amino acids. In contrast, the search for T cell epitopes leading to AITD has yielded conflicting results, and only some sequences have been reported to exist in several patients.[12,13,15]

An interesting line of investigation is whether patients who develop hypothyroidism in the setting of autoimmunity differ in their epitope pattern from individuals with positivity to TPOAb but normal thyroid function. In fact, this could help predict which thyroid autoantibody-positive individuals will eventually become hypothyroid.[12,13,16]

A further possible role for TPOAb is its involvement in the controversial entity of Hashimoto's encephalopathy, as several reports have found that TPOAb binds to cultured astrocytes and may trigger some sort of autoimmune cross-reaction between the thyroid gland and the central nervous system.[17]

Thyroglobulin. Tg is a thyroid-synthesized protein that is the precursor of thyroid hormones. It behaves as one of the essential autoantigens in the pathogenesis of AITD through presenting several epitopes to B and T cells. For instance, a particular Tg T cell epitope, Tg2098, has been shown to bind strongly and specifically to the MHC class II disease-susceptibility HLA-DRβ-Arg74 molecule, which stimulates T cells and triggers the development of AITD.[18] In addition, the gene encoding Tg may confer susceptibility to autoimmune thyroiditis by coding for Tg variants of different immunogenicity.[19]

Other Antibodies. Several additional antibodies have been described in recent years. For instance, antibodies to the sodium/iodide symporter can potentially block its activity, although reports have described mixed results regarding its significance. Another reported autoantigen is pendrin, which is an apical protein involved in mediating iodine efflux from thyroid follicular cells into the lumen. In this case, pendrin antibodies have been found in around 80% of patients with AITD (mainly HT) when using immunoblotting techniques, but only in 10% when using radioligand binding assays. These antibodies have also been reported in up to 9% of healthy controls, thus further adding uncertainty to this finding. Thyroxine, triiodothyronine, and the insulin-like growth factor receptor have also been identified as potential autoantigens.[13] Furthermore, antibodies to other organs that are also susceptible to autoimmune disease, such as those directed against pancreatic islet cells, the adrenal cortex, the gastric mucosa, or the parathyroid, may be also detected with a higher frequency than in healthy controls, as will be discussed later in this chapter.

Cellular Mechanisms. Cellular immune mechanisms are closely related and cross-linked to humoral mechanisms in the development of AITD. In fact, once they are triggered, subsequent feedback circuits occur, with an eventual and reciprocal amplification and perpetuation of complex immune responses. Although cellular mechanisms have been traditionally attributed to HT, increasing evidence has proven that they are also related to the whole spectrum of AITD, including GD and Graves' orbitopathy.[5,20-23]

APCs, especially dendritic cells (DCs), are bone marrow-derived cells of both lymphoid and myeloid stem cell origin that populate all lymphoid organs, as well as almost all nonlymphoid tissues and organs. Human DCs are classified as immunogenic or tolerogenic, according to their functional role, depending on whether they activate autoreactive T cells or regulatory T cells (T regs), respectively. DCs are increased in thyroid-infiltrating cells in both GD and HT, where they can present immunogenic thyroid-related epitopes (e.g., Tg) to T cells. Moreover, an altered proportion and function of tolerogenic DCs has been suggested as an additional potential contributor to the pathogenesis of AITD.[5,24] Thyroid follicular cells may also act as APCs, favored by their ability to secrete specific recruitment chemokines, and in the setting of a highly vascularized endocrine gland.[25]

B cells are mainly involved in humoral immune mechanisms, so their role has traditionally been attributed to the production of TRAb and subsequent onset of GD. But in this same manner, B cells are also responsible for the production of TPOAb and TgAb by means of the usual mechanisms: they recognize soluble antigens through their specific cell membrane immunoglobulin receptors, synthesize antibodies, and act as APCs to CD4+ cells. B cell activation is further promoted by T helper (Th) cells, contributing to the spread and diversification of polyclonality. In addition, B cells take part in the thyroid lymphocytic infiltrate, where they also exert their role as antibody producers, enhancing the relevance of the thyroid itself in perpetuating AITD.

T cells also form part of the thyroid inflammatory infiltrate in AITD, and they are subclassified in two main subtypes according to their predominant role: cytotoxic T cells (Tc), mainly CD8+, and Th cells, mainly CD4+, which interact with other surrounding cells and secrete specific cytokines. Further differentiation of CD4+ naïve T cells leads to the emergence of specific subsets, namely Th1, Th2, Th17, and T reg.[5]

The Th1 response has been the one traditionally ascribed to HT. In this case, Th1 cells emerge following stimulation by IL-12, IFN-γ, and T-bet, and synthesize IL-1, IL-2, IFN-γ, and TGF-β, through activation of the CCR5 and CXCR3 receptors. This inflammatory infiltrate, together with macrophages and other subsets of T cells, will eventually lead to thyroiditis and destruction of the thyroid gland. On the other hand, Th2 cells, which emerge because of IL-4 stimulation, synthesize IL-4, IL-5, IL-6, IL-10, and IL-13 and mediate a predominantly humoral response. In fact, they interact with B and plasma cells and allow production of antibodies. In this regard, recent evidence proved that this occurred not only in GD, as it was classically described, but also in HT, where TPOAb and TgAb contribute to thyroid destruction by means of antibody-dependent cell-mediated cytotoxicity and complement activation.[5,21,23]

T regs are regulatory CD4+ CD25+ Foxp3+ cells. They recognize self antigens, proliferate only to a limited extent, and inhibit activation of, proliferation of, and cytokine synthesis by other lymphocytes. They also complete their regulatory role by interacting in cell–cell mechanisms and producing antiinflammatory cytokines, mainly TGF-β, IL-10, and IL-35. If T reg cells escape their own immunoregulatory processes, and autoreactive T reg cells prevail, with defects in their number and/or function, loss of self-tolerance and potentiation of AITD may occur. In this regard, for instance, Foxp3 mutations and defects in the immunosuppressive effect of CD69 have been recently reported as reasons for the inability to downmodulate the autoimmune response and tissue damage.[26,27]

Th17 cells derive from Th stimulation by IL-β, RORC2, and STAT3. They predominantly synthesize IL-17A, IL-17F, IL-21, IL-22, TNF-α, and IL-β, and contribute to the proinflammatory environment. If Th17 cells further undergo exposure to specific antigens, two potential pathways can originate. In this regard, in the setting of IL-23, pathogenic proinflammatory Th17 cells will arise that secrete IFN-γ and granulocyte-macrophage colony-stimulating factor; on the other hand, when TGF-β prevails, differentiation to nonpathogenic Th17 cells, which are involved in mechanisms against different extracellular bacteria and fungi and tissue repair, will occur.[5,28,29]

Another issue that heightens the complex interrelationship between cellular mechanisms is cross-regulation of Th1, Th17, and T reg cells, which subsequently promotes tissue damage in autoimmune conditions.[5,30] For instance, Th17 cells may differentiate into Th1-like cells or to Fop3+ Treg cells, but Treg cells may also evolve to Th17, and a mixed Th17/T reg phenotype has even been observed.[5,31-33]

This complexity in T cell regulation of has not been fully clarified yet in the particular setting of AITD. Recent findings point to a predominance of Th17 cells and their derived cytokines in peripheral blood and in the thyroid gland of patients with AITD, mainly those with HT,[25] which presumably would be of the pathogenic lineage and prevail over T reg cells. Moreover, circulating microvesicles, several Th17-related cytokines, leptin, and galectins, amongst others, could also contribute to a higher promotion of these proinflammatory Th7 cells, whilst the role of thyroid hormones still needs further investigation.[5,8,24,34]

Chemokines and Cytokines. Patients with HT seem to exhibit higher circulating levels of IL-6, IL-15, and IL-22, whereas IL-17 and IL-22 can be both found in thyroid tissue.

IL-17 is a proinflammatory cytokine, and numerous cells express IL-17R, suggesting that IL-17 serves as a bridge between the innate and adaptive immune responses. In fact, IL-17 stimulates NF-κB and mitogen-activated protein kinase pathways, and this induces proinflammatory cytokines, recruitment of neutrophils and macrophages at inflammation sites, and T cell proliferation. In addition, it will also trigger the upregulation of further proinflammatory cytokines, perpetuating the induction and exacerbation of the process. In this regard, IL-17 has been ascribed a significant role in AITD, mainly in HT and severe GD.[25]

IL-23 is produced by activated macrophages and DCs and has been described as a key Th17-polarizing mediator. As such, it can promote the proliferation of memory T cells and further production of IFN-γ. In addition, it helps maintain an adequate microenvironment for Th17 by mediating upregulation of IL-17.[34,35]

Apoptosis. Ultimately, autoimmune thyroiditis, and especially HT, is characterized by apoptosis of thyroid cells, eventually leading to follicular destruction, and not so much by hyperplasia. Depending on the degree of thyroid destruction, there may be certain residual thyroid function, and thus the clinical impact on thyroid hormones may be variable.

Both TPOAb and TgAb can bind to the complement system and trigger cytotoxic mechanisms, while B and T cells also infiltrate the thyroid gland. In fact, Tc cells, which are mainly CD8+, mediate direct apoptosis and destruction of thyroid follicular cells through activation of the CD95 membrane receptor (Fas–FasL). The reason for which thyroid cells express Fas may be related to T cell blastic transformation after exposure to thyroid antigens and subsequent release of cytokines. In this regard, cytokines have demonstrated to play a very relevant role on the onset of thyroid destruction, in the setting of an inflammatory infiltrate. Additionally, humoral mechanisms will come together in this complex autoimmune loop to perpetuate the autoimmune process.[5,34,35]

Clinical Relevance: Viewing Autoimmunity From a Perspective of Trying to Understand Autoimmune Thyroid Disease. Immune mechanisms involved in the onset and development of AITD are truly complex. Although there is currently a large amount of evidence regarding this condition, there is probably still much to be discovered and integrated, including other potential cell subsets, mediators, and functions. Fig. 73.2 represents a simplified summary of the currently known major components in an attempt to graphically show how

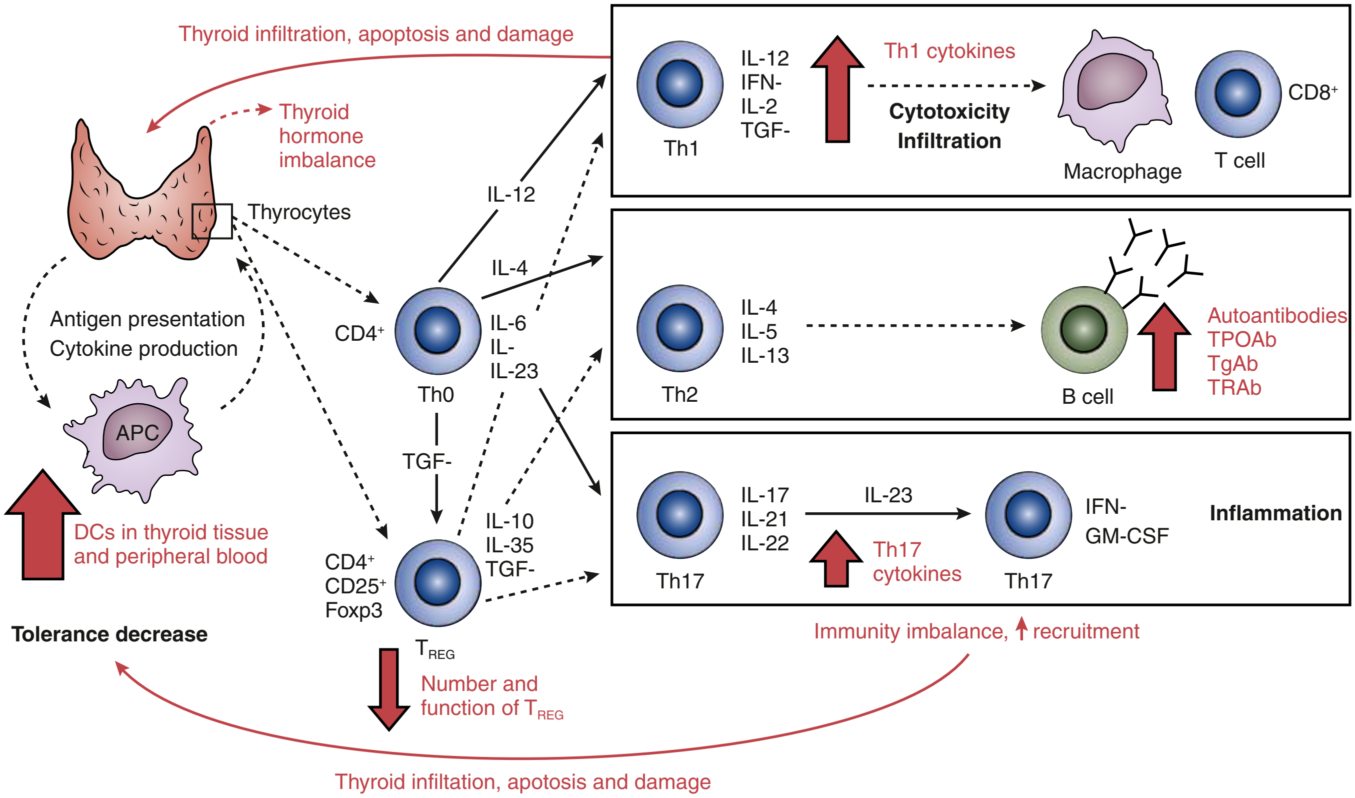

Thyroid infiltration, apoptosis and damage
Thyroid hormone imbalance
Thyrocytes
Antigen presentation
Cytokine production
APC
DCs in thyroid tissue and peripheral blood
Tolerance decrease
CD4⁺
Th0
IL-12
IL-4
IL-6
IL-
IL-23
TGF-
CD4⁺
CD25⁺
Foxp3
T_REG
IL-10
IL-35
TGF-
Number and function of T_REG
Th1
IL-12
IFN-
IL-2
TGF-
Th1 cytokines
Cytotoxicity
Infiltration
Macrophage
CD8⁺
T cell
Th2
IL-4
IL-5
IL-13
B cell
Autoantibodies
TPOAb
TgAb
TRAb
Th17
IL-17
IL-21
IL-22
IL-23
Th17 cytokines
Th17
IFN-
GM-CSF
Inflammation
Immunity imbalance, ↑ recruitment
Thyroid infiltation, apotosis and damage

Fig. 73.2 Schematic representation of cellular and humoral mechanisms involved in thyroid autoimmunity. Antigen presenting cells (*APCs*) in thyroid tissue and peripheral blood may recognize self antigens and present them to T cells (decreased self-tolerance), leading to specific T helper (Th) cell differentiation according to targeted cytokine production. Increased Th1 activity will mediate cytotoxicity and infiltration (mainly through IL-12 and IFN-γ). Increased Th2 activity (mainly mediated through IL-4) will lead to increased autoantibody production. Increased Th17 activity (mediated through IL-17) will lead to increased differentiation of pathogenic Th17 cells and subsequent inflammation. Decreased T regulatory (T reg) activity jeopardizes the immunosuppressive effect for other T effector cells. ↑, Increased; ↓, decreased. (Reproduced from Ramos-Levi AM, Marazuela M. Pathogenesis of thyroid autoimmune disease: the role of cellular mechanisms. Endocrinol Nutr. 2016;63:421 429. Copyright © 2016 Elsevier Masson SAS. All rights reserved; Li Q, Wang B, Mu K, et al. The pathogenesis of thyroid autoimmune diseases: new T lymphocytes - cytokines circuits beyond the Th1-Th2 paradigm. *J Cell Physiol.* 2019;234:2204–2216.)

cellular and immune mechanisms come together in AITD, as well as its clinical implications.

Environmental Factors. Exogenous environmental factors may contribute to the development of AITD. It is interesting to note that some factors may have opposite effects on the development of the different types of AITD (GD or HT) and, furthermore, that some factors may protect against, rather than promote, the onset of AITD.[14,36] The most frequently associated factors include smoking, alcohol, selenium,[37-39] microparticles,[8] vitamin D,[39-44] iodine and any form of iodine supplementation, including drugs like amiodarone or lithium,[39,45] stress,[6,13,14,36] infections (although without identification of a specific, unequivocal pathogen),[6,13,14,36] drugs and other toxic substances, and radiation.[6,13,14,34,36] The influence and relationship of these factors in AITD are described in the chapter on GD.

Genetic Background and Susceptibility Genes for Thyroid Autoimmunity. Epidemiological evidence of familial clustering of AITD foresees a genetic susceptibility. In fact, up to 30% of siblings of patients with AITD eventually develop the condition, there is an increased sibling risk ratio for AITD (up to 17), and there is an increased prevalence of thyroid autoantibodies (up to 50%) in family members of affected individuals.[46,47] Moreover, studies in twins showed an increased concordance rate in monozygotic twins versus dizygotic twins (0.3–0.6 vs. 0.0–0.1).[48,49] Additionally, a recent National Health and Nutritional Examination Surveys study highlighted racial variations in the prevalence of AITD, suggesting potential genetic differences and the role of genetic susceptibility in the etiology of GD and HT.[50,51]

Several genes have been identified as being significantly associated with AITD and the presence of thyroid antibodies.[34,46,52]

Case-control studies, tag single nucleotide polymorphism screening, genome-wide association studies, and immunochip studies have identified some genes potentially involved in the development of AITD (Table 73.3). It is worth noting that the majority of these genes are known are immunoregulatory genes involved in T cell function, strongly suggesting the relevance of these cells in the onset and perpetuation of AITD. In this regard, certain polymorphisms may uniquely predispose to GD, HT, or both, whereas others are not thyroid-specific and increase the likelihood of autoimmunity in general.[34,46]

Existential Factors. Age, sex, parity, breastfeeding, and X-chromosome inactivation have been linked to AITD,[13,14,34,36] as is explained in the chapter on GD. Additionally, some studies have observed an association between the month of birth and the future potential development of AITD, or even that a specific month of the year could be associated to a higher incidence of these diseases, in line with observations in type 1 diabetes mellitus (T1DM) and multiple sclerosis.[53] Implicated seasonal factors include viral epidemics, ultraviolet radiation, vitamin D, and country latitude.

KEY POINTS

- Exogenous environmental factors and a genetic susceptibility background contribute to the development of thyroid autoimmune disease.

Stages in the Development of Autoimmune Thyroid Disease

The information provided earlier demonstrates the vast complexity of AITD pathogenesis. Autoimmune abnormalities and environmental, genetic, and existential factors play a significant role in thyroid disturbances,[7] as represented in Fig. 73.3. In addition, Fig. 73.4 attempts to summarize the series of events occurring during the different stages of the development of AITD. Perpetuation of these mechanisms will eventually lead to clinically evident chronic thyroiditis.

KEY POINTS

- The autoimmune process in Hashimoto's thyroiditis will eventually lead to thyroid dysfunction

TABLE 73.3 **Susceptibility Genes for Autoimmune Disease**

Method of Identification	Gene	Function
Case-control studies and tag single nucleotide polymorphism screening	*PTPN22*	T cell signal transduction through interaction with molecules critical for T cell receptor signaling
	CTLA4	Inhibition of T cell signaling
	Histocompatibility antigen (*HLA*) class II molecules	Presentation of exogenous antigens for recognition by CD4+ T helper cells. T cell activation
	IL2RA	Encodes CD25, which is expressed on T regulatory cells and downregulates T cell activity (peripheral tolerance)
Case-control studies and confirmation by genome-wide association studies	*FCRL3*	Expressed during B cell maturation to regulate their function in both ways
	HLA class I molecules	Presentation of endogenous antigens (e.g., viral) for recognition by CD8+T cells. T cell activation
	TSHR	Receptor for thyroid-stimulating hormone, the primary autoantigenic target in Graves' disease
Genome-wide association studies and immunochip	*GDCG4p14*	Expressed in CD4+ T helper and CD8+ T cells
	BACH2	Expressed during B cell maturation, controls B cell development and antibody production
	RNASET2	Expressed in CD4+ T helper and CD8+ T cells
	FOXE1	Involved in thyroid gland morphogenesis. Binds response elements in the thyroglobulin and thyroid peroxidase gene promoters
	FOXP3	Both transcriptional repressor and activator for primarily immunological genes (e.g., inhibits IL-2 and IFN, or activates CD25)
		Crucial regulator of regulatory T cells differentiation and function
	Others yet to be described	Unknown functions

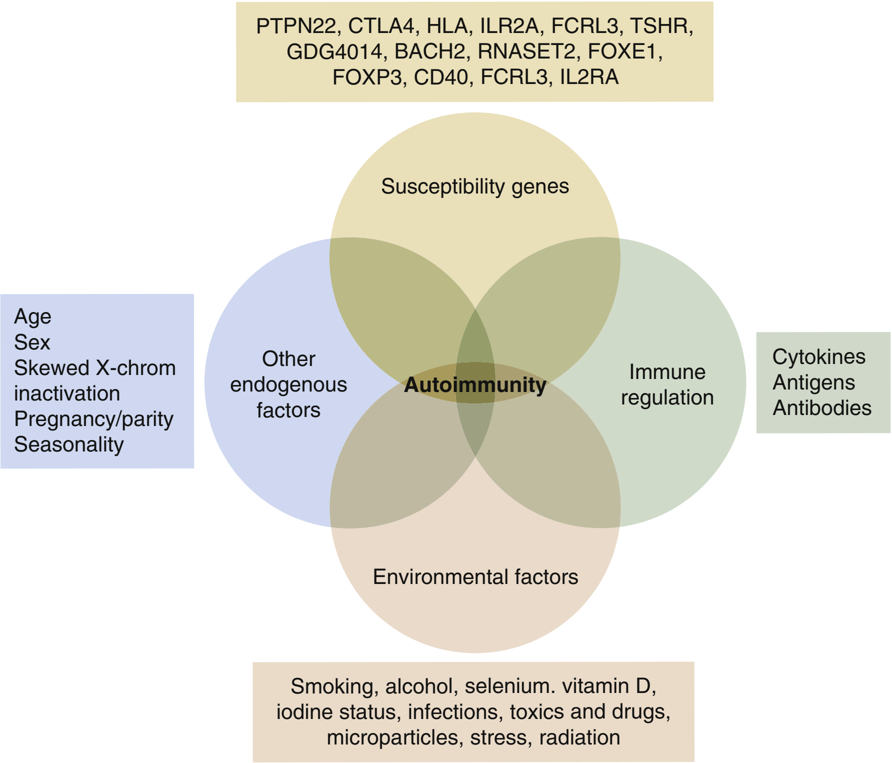

PTPN22, CTLA4, HLA, ILR2A, FCRL3, TSHR, GDG4014, BACH2, RNASET2, FOXE1, FOXP3, CD40, FCRL3, IL2RA
Susceptibility genes
Age
Sex
Skewed X-chrom inactivation
Pregnancy/parity
Seasonality
Other endogenous factors
Autoimmunity
Immune regulation
Cytokines
Antigens
Antibodies
Environmental factors
Smoking, alcohol, selenium. vitamin D, iodine status, infections, toxics and drugs, microparticles, stress, radiation

Fig. 73.3 Schematic representation of the mechanisms involved in the onset of autoimmunity, and their complex interrelationship. Certain genes may increase the susceptibility to autoimmune disease, for instance, *PTPN22, CTLA4, HLA, ILR2A, FCRL3, TSHR, GDG4014, BACH2, RNASET2, FOXE1, FOXP3, CD40, FCRL3,* and *IL2RA*. Additional endogenous factors, such as age, sex, skewed X-chromosome inactivation, pregnancy/parity and seasonality, and exogenous factors, including smoking, alcohol, selenium, vitamin D, iodine status, infections, toxic substances and drugs, microparticles, stress, and radiation, may contribute to dysregulation of the immune system and onset and perpetuation of autoimmunity.

Natural History and Clinical Features

Given the progressive nature of autoimmune disorders, patients with early-stage HT are usually euthyroid and with no goiter, or a very small one. In this regard, the presence of thyroid autoantibodies would be the only detectable evidence of autoimmune thyroiditis, with pathological thyroiditis being only subclinical. In fact, postmortem studies in individuals with no overt thyroid disease have revealed that positivity for thyroid antibodies, especially TPOAb, significantly correlates with the presence of thyroid lymphocytic infiltration. Progressive development of the disease usually leads to a more firm and diffuse goiter of variable size, a stage in which patients are usually referred as "having chronic autoimmune thyroiditis."[2]

Goiter may be an incidental finding at physical examination, but some patients may experience local discomfort. Neck palpation usually reveals a firm consistency, but sometimes it may also give the impression of being hard, requiring differential diagnosis with malignancy. In any case, compressive symptoms and pain are rare in HT, allowing its differentiation from Riedel's thyroiditis and SAT, respectively, but if a thyroid nodule is identified, further investigation must be performed to rule out the potential coexistence of thyroid cancer.[1,3]

At this time, findings regarding thyroid function may be variable, ranging from normal to a slight initial hyperthyroidism because of thyrotoxicosis and hormone liberation. When the disease progresses even more, goiter increases, histological fibrosis spreads, to the detriment of thyroid follicles, the gland atrophies, and hypothyroidism occurs, evolving into what is prototypically known as Hashimoto's disease in its clinical, final, lifelong phase (Fig. 73.5). Progression to overt hypothyroidism is reported to occur at an approximate rate of 5% per year; thus, aging in this setting may be associated with a higher rate of thyroid hypofunction. Some patients may seem to spontaneously return to euthyroidism after transient hypothyroidism; this may occur if thyroid destruction occurs at a slow pace, with a clinically evident goiter and high Tg levels, and in patients with subclinical autoimmune thyroiditis with an excessive iodine intake. In these cases, thyroid response following TSH administration may be used to evaluate the potential expected recovery of thyroid hormones, and reduction of iodine intake may reverse the problem. In some cases, patients may develop hyperthyroidism over time and after long-term replacement therapy with levothyroxine (L-T4), mainly owing to the existence of stimulating antibodies, or because of an acute aggravation of thyroid autoimmunity and subsequent destructive thyrotoxicosis. These latter episodes are then usually followed by transient hypothyroidism.[54]

HT is not as common in children, in whom thyroid antibodies are usually lower. In fact, the frequency of AITD in children and adolescents is low, and the disease is usually self-limited. Thus, when symptoms occur, they are generally mild, with small goiters, normal thyroid function or only subclinical dysfunction, and reversion to normality in up to half of cases.[55,56]

Relationship With Other Types of Thyroiditis and Graves' Disease

Clinical presentation of autoimmune thyroiditis as a whole may be variable; for instance, hypothyroidism may be a frequent finding that leads to diagnosis, but maintenance of euthyroidism for prolonged periods of time is not rare, as will be described later. Thus, considering the chronicity in the definition of HT, patients who suffer any of the

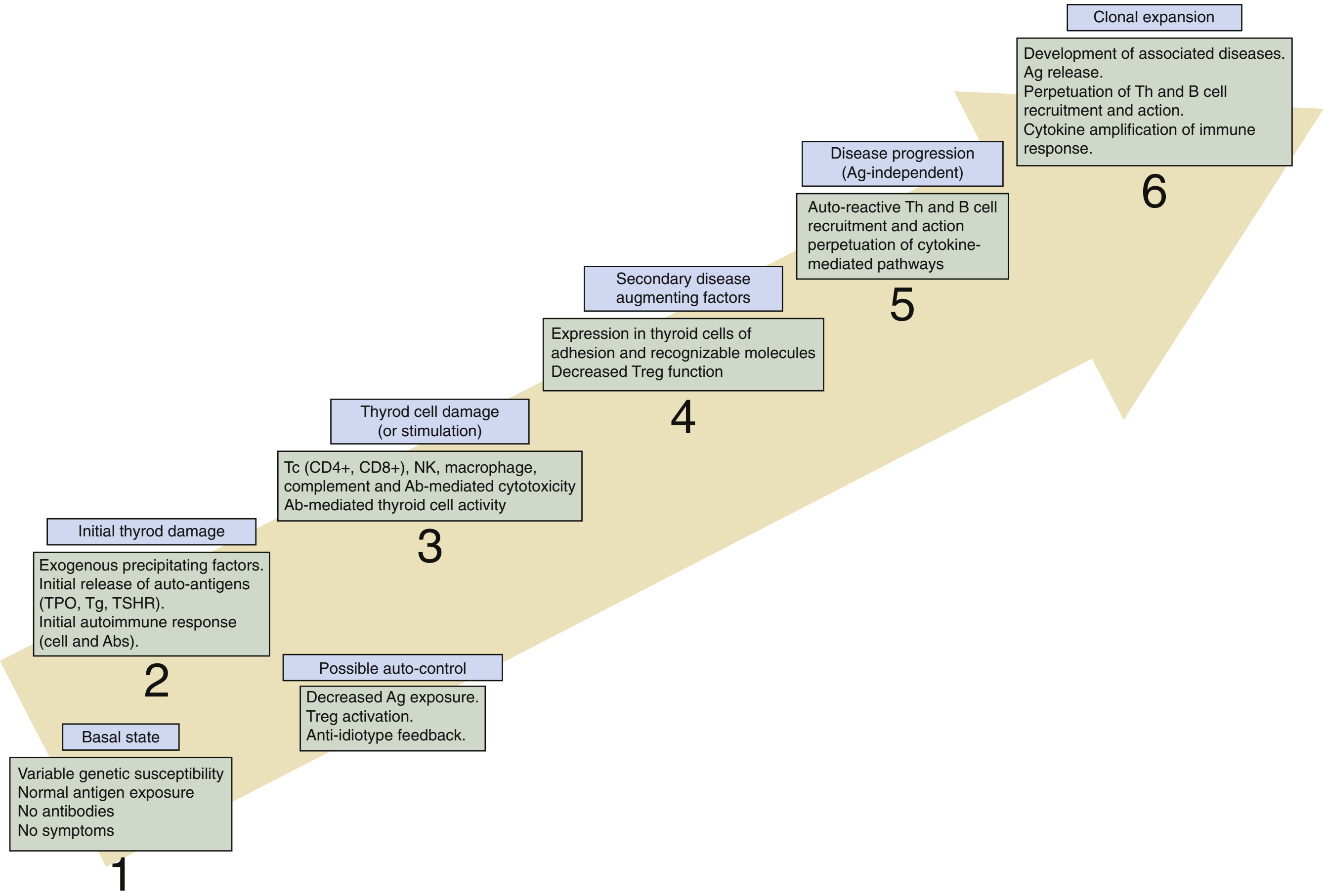

Clonal expansion
Development of associated diseases.
Ag release.
Perpetuation of Th and B cell recruitment and action.
Cytokine amplification of immune response.
6
Disease progression (Ag-independent)
Auto-reactive Th and B cell recruitment and action perpetuation of cytokine-mediated pathways
5
Secondary disease augmenting factors
Expression in thyroid cells of adhesion and recognizable molecules
Decreased Treg function
4
Thyrod cell damage (or stimulation)
Tc (CD4+, CD8+), NK, macrophage, complement and Ab-mediated cytotoxicity
Ab-mediated thyroid cell activity
3
Initial thyrod damage
Exogenous precipitating factors.
Initial release of auto-antigens (TPO, Tg, TSHR).
Initial autoimmune response (cell and Abs).
2
Possible auto-control
Decreased Ag exposure.
Treg activation.
Anti-idiotype feedback.
Basal state
Variable genetic susceptibility
Normal antigen exposure
No antibodies
No symptoms
1

Fig. 73.4 **Flow diagram to illustrate the stages inthe autoimmune process leading to thyroid dysfunction.** Starting at a first basal state, in which there is a variable genetic susceptibility, with normal antigen exposure, no antibodies, and no symptoms, exogenous precipitating factors may initiate the autoimmune response and thyroid damage. The possibility of autocontrol and returning point exists, but perpetuation of the autoimmune response will lead to secondary disease, augmenting factors, disease progression, and clonal expansion. *TPO,* Thyroid peroxidase; *Tg,* thyroglobulin; *TSHR,* thyroid-stimulating hormone receptor; *Ab,* antibody; *Ag,* antigen; *Tc,* cytotoxic T cell; *NK,* natural killer cell; *Treg,* regulatory T cell; *Th,* T helper cell.

autoimmunity-related thyroidites have a higher risk of later developing HT. In fact, some authors have reported that silent thyroiditis behaves as the starting point of subsequent HT. Furthermore, individuals with adequately controlled HT may develop some sort of silent thyroiditis during their lifetime; for instance, young women with HT may develop postpartum thyroiditis, demonstrating the complex spectrum of forms of AITD.[1-3]

Patients with untreated GD sometimes progress to hypothyroidism because of autoimmune destruction, in a similar way to what occurs in other types of thyroiditis, but also because of thyroid receptor-blocking antibodies. In fact, around 10% of thyrotoxic GD patients have been found to have both stimulating and blocking TRAbs, resulting in a trend for hyperthyroidism when stimulating antibodies prevail but progression to hypothyroidism when blocking antibodies are dominant. Overall, patients' clinical features will depend on the balance between stimulating versus blocking antibodies, and between the destructive aspects of humoral and cellular immunity. However, once thyroid cells are destroyed and overtaken by fibrosis, stimulating factors will be ineffective, limiting the occurrence of hyperthyroidism. This becomes relevant in clinical practice because patients with HT on adequate replacement therapy are not exempt from developing hyperthyroidism because of GD if there is viable residual thyroid tissue that can be targeted by stimulating TRAbs.

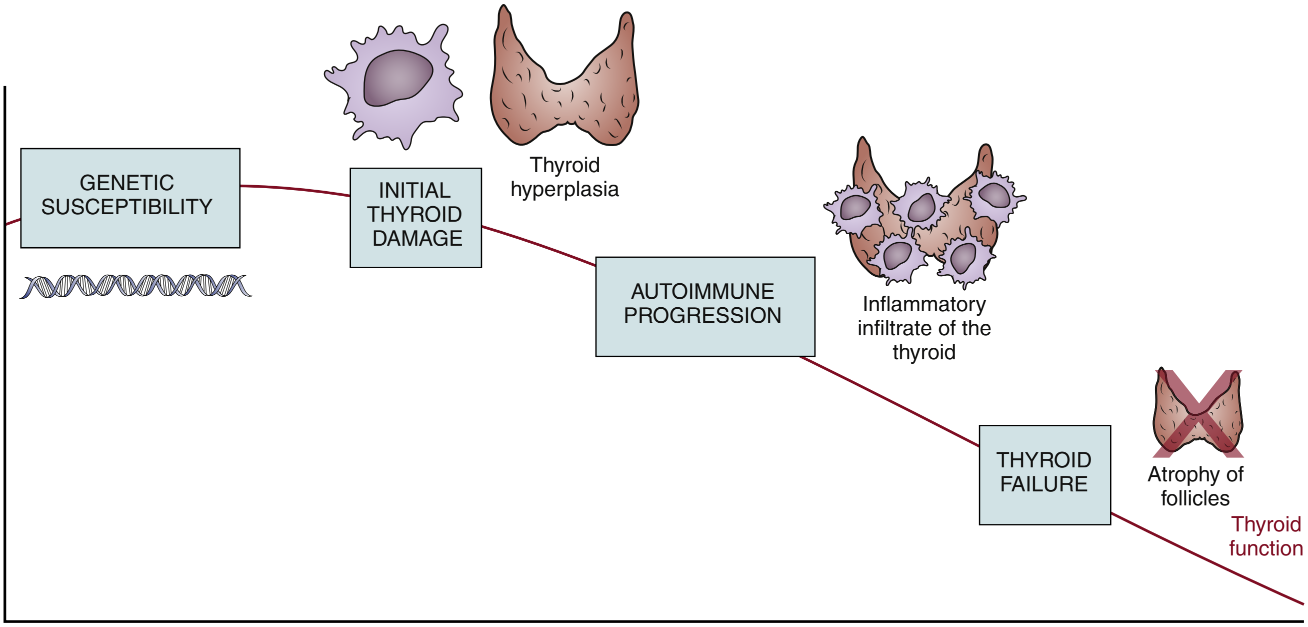
HT has sometimes been subclassified into Ig4 thyroiditis and non-IgG4 thyroiditis, although not without debate (see later). The controversy arises from the observations that some key histologic features of the IgG4 subtype may not be present, concomitant synchronous extrathyroidal involvement is frequently absent, and serum IgG4 is not always specifically directed against Tg. Thus, HT could also be regarded as a condition in which tissue and serum IgG4 may be detected, but with uncertainty regarding its significance.

Relationship With Other Autoimmune Diseases

AITD is associated with organ-specific and systemic autoimmune disease, including T1DM, autoimmune adrenalitis, autoimmune hypoparathyroidism, autoimmune hypophysitis, vitiligo, myasthenia gravis, thrombocytopenic purpura, alopecia, Sjögren syndrome, rheumatoid arthritis, systemic lupus erythematosus, and progressive systemic sclerosis. In addition, there is an increased prevalence of celiac disease in patients with HT.[57-60] These issues are extensively reviewed in the corresponding chapter.

KEY POINTS

- Autoimmune thyroid disease may be associated with organ-specific and systemic autoimmune disease, and to development of thyroid malignancy.



Natural history of Hashimoto's thyroiditis

Fig. 73.5 Schematic representation of the natural history of Hashimoto's thyroiditis. Certain individuals may exhibit a genetic susceptibility to thyroid damage owing to the combined action of endogenous and exogenous factors, leading to thyroid hyperplasia. Progression of autoimmunity leads to an inflammatory infiltrate of the thyroid, with subsequent eventual gland failure attributed to follicle atrophy. The curve represents thyroid function over the course of time. The thyroid diagrams represent the changes the gland exhibits.

HASHIMOTO'S THYROIDITIS AND THYROID MALIGNANCY

Epidemiological studies have reported an average coexistence rate between HT and papillary thyroid cancer (PTC) ranging from 5% to 85%.[61-63] This cooccurrence suggests an underlying immunological connection between both entities, as has been reported in several experimental studies. In line with this observation, the incidence of thyroid nodule malignancy is increased in the setting of positive autoimmunity and/or high TSH levels, in comparison with nonautoimmune nodular goiter. In addition, HT was more frequent in patients with PTC than in those with benign diseases.[64] Interestingly, however, HT has also been suggested to play a protective role against the progression of PTC.[62] In this regard, in a small metaanalysis, PTC patients with coexisting HT were found to have less aggressive clinicopathologic characteristics (i.e., lower rates of extrathyroidal extension and lymph node metastasis) and longer recurrence-free survival.[62] A further and more recent, larger metaanalysis, with a predefined subgroup analysis, proved the association of HT with better PTC outcomes, regardless of tumor size and regional iodine status.[65] Again, the underlying mechanism is not fully understood. A potential explanation concerns the increased expression of the RET/PTC-RAS-BRAF cascade in nonneoplastic follicular cells; thus, in HT, there may be overlap between these molecular mechanisms involved in early tumor development and inflammation.[66] In any case, the connection between HT and PTC may be used as a prognostic factor in the clinical setting of PTC.

Autoimmune thyroiditis is also associated with B cell lymphoma. In this regard, the similarity in the sequence of clonal bands between HT and thyroid lymphoma suggests that the latter may derive from the former in around 40% to 80% of cases.[67] This condition primarily affects females in their sixth decade or older, and diagnosis is frequently motivated by rapid enlargement of a neck mass in the setting of previously known HT being treated with hormone replacement therapy. Fine-needle biopsy is the key for diagnosis. Remission is achieved in the majority of cases with chemotherapy and radiotherapy.

KEY POINTS

- The hallmark for the diagnosis of Hashimoto's thyoiditis is presence of thyroid auto-antibodies, particularly TPOAb.

Diagnosis

Physical examination of patients with HT may reveal a firm, bumpy, symmetric, painless goiter as the initial clue for diagnosis. However, around 10% of patients with chronic autoimmune thyroiditis may exhibit an atrophic gland, presumably as a surrogate marker of thyroid failure in the final stages of HT.[1,2]

The presence of circulating thyroid antibodies, particularly TPOAb, is the hallmark of HT. In fact, 90% and 20% to 50% of HT patients exhibit high levels of TPOAb and TgAb, respectively.[1-3] In this regard, clinical observation of diffuse or nodular goiter and serologic tests with positive results for TPOAb and/or TgAb are usually enough for reaching a definite diagnosis of HT. A minority of patients, however, also have thyroid receptor-blocking and stimulating immunoglobulins. On the contrary, evaluation of thyroid hormones (TSH, FT4, and FT3) is helpful if hypothyroidism is evidenced, but it must be taken into account that a significant proportion of patients may exhibit only subclinical dysfunction. In fact, no differences in the titers of thyroid autoantibodies have been observed between euthyroid and hypothyroid HT patients. Thyroid antibodies may be detected in a nonnegligible 10% of individuals in the general population, without any other clinical sign or symptom of autoimmune thyroiditis. These people should

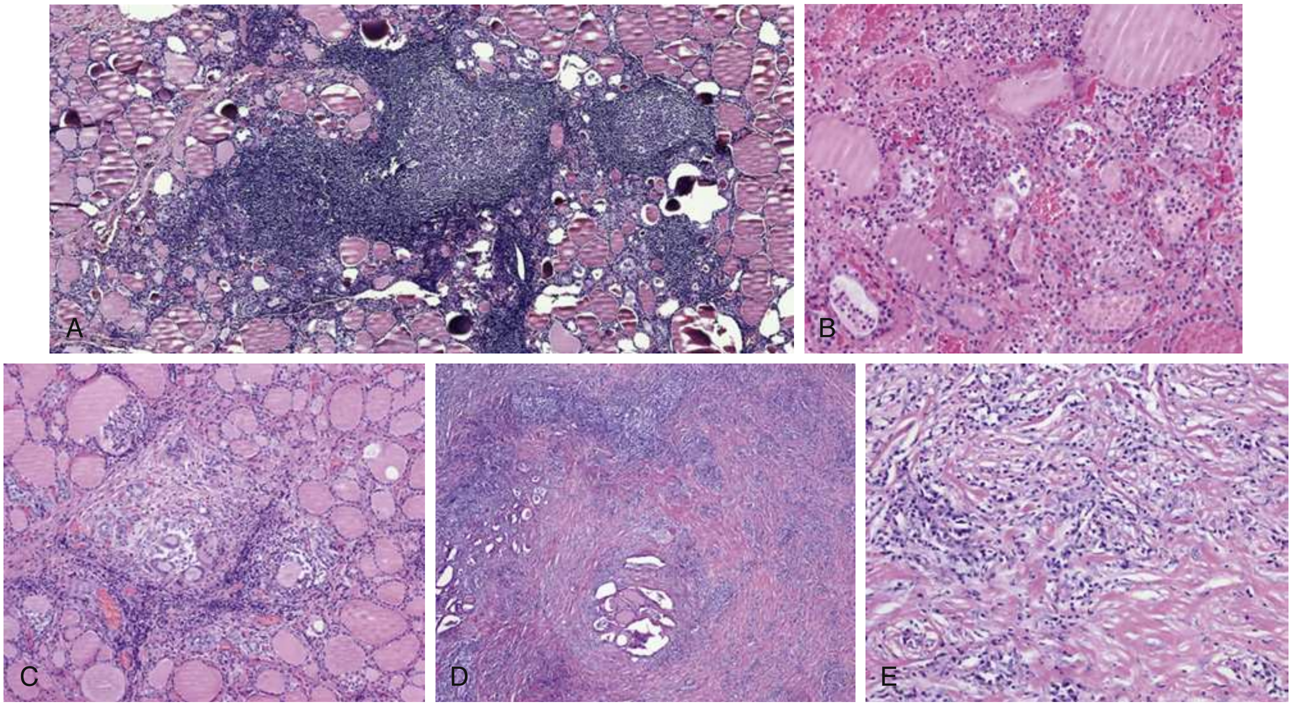


Fig. 73.6 Microscopy images of hematoxylin and eosin (H&E)-stained slides of Hashimoto's thyroiditis (H&E, 10×): Note the diffuse lymphoplasmacytic infiltration and lymphoid follicles with germinal center *(center of image, dyed in intense purple)* (**A**); acute thyroiditis (H&E, 20×): Note the increased presence of polymorphonuclear cells amongst the normal thyroid follicles (**B**); ,Subacute thyroiditis (H&E, 10×): Alteration and destruction of thyroid follicles, with interstitial edema, and with numerous histiocytes surrounding the remaining colloid, resembling giant cells, forming noncaseating granulomas with giant and inflammatory cells *(center of the image)* (**C**); Riedel's thyroiditis (H&E, 4× and 20×,): Extensive follicular and cell destruction, with scarring of the thyroid gland, and replacement of thyroid tissue by a dense fibrous tissue (**D**), and lymphohistiocytic inflammatory infiltrates causing obliteration of thyroid follicles (**E**). (Images kindly provided by Dr. Magdalena Adrados, from the Department of Pathology of Hospital Universitario La Princesa, Madrid, Spain.)

be viewed as having subclinical AITD, and monitoring of thyroid hormones is suggested.[12,13] In addition, primary hypothyroidism may be detected despite negative thyroid antibodies; in these cases, imaging techniques may aid in the differential diagnosis.

Given the potential hyperthyroidism that can occur in the initiation or follow-up of a patient with HT, differential diagnosis with GD, SAT, and toxic nodular goiter may be necessary. In this regard, duration of hyperthyroidism, evaluation of TRAb, serum T3/T4 ratio, ultrasonography, and radioactive iodine uptake (RAIU) may be helpful. Specifically, GD will exhibit longer and more severe hyperthyroidism than thyrotoxicosis, elevated TRAb titers, increased blood flow on ultrasonography, and increased RAIU (see chapter 71).

Thyroid ultrasonography is helpful when physical examination reveals abnormal thyroid palpation. Typical findings in a patient with HT comprise various degrees and heterogeneously distributed zones of parenchymal hypoechogenicity, with a thyroid volume that may range from normal-increased to frankly increased. In cases of atrophic thyroiditis, ultrasonography will show a reduced thyroid volume and a large number of hyperechoic fibrous streaks superimposed on a hypoechoic background. RAIU, however, will not be as helpful.[2]

Fine-needle aspiration cytology is required, following the same criteria for thyroid nodules, but it is sometimes performed in pseudonodular regions to confirm the diagnosis of chronic autoimmune thyroiditis and/or to rule out the presence of malignancy. In this latter clinical setting, findings will reveal lymphocytic infiltration, germinal centers, small follicles with oxyphilic cells, fibrosis, and atrophic areas (Fig. 73.6A).[2]

KEY POINTS

- First-choice treatment for Hashimoto's thyroiditis is levothyroxine replacement.
- There is no effective targeted treatment to reverse the chronic autoimmune process.

Treatment and Follow-Up

To date, there is no targeted effective treatment that reverses the occurrence or prevents the progression of chronic autoimmune thyroiditis. However, once overt hypothyroidism is present, replacement with L-T4 is the first choice of treatment for HT patients. L-T4 is also recommended in certain cases of subclinical hypothyroidism with positive autoantibodies, to anticipate progression to overt hypothyroidism and/or aid in the management of concomitant hyperlipidemia and atherosclerotic heart disease[68] (see chapter 74). In any case, the goal of treatment is achievement of euthyroidism with normal TSH values. Indeed, although no significant changes are observed in repeated fine-needle aspiration cytology after thyroid hormone normalization, a significant decrease in TPOAb and TgAb and reduction in thyroid volume may be noted after 1 year of treatment.[54] Daily replacement doses are usually calculated in accordance with body weight, around 1.2 to 2 mcg/kg/day. For young healthy adults, treatment can be started at full replacement doses; however, older patients or those with coexisting cardiovascular problems need lower starting doses (see chapter 74). Periodic monitoring of TSH values will allow a correct dose adjustment. Thyroid suppressive therapy will not prevent the occurrence of episodes of destructive thyroiditis, but, in repeated cases of painful events, ablative I-131 or surgical treatment may be an effective alternative. In fact, a recent randomized trial observed that total thyroidectomy was more effective than medical treatment in improving health-related quality of life and fatigue, possibly in the setting of concomitant elimination of serum TPOAb.[69]

There is some evidence for a potential role for micronutrients, namely iron, selenium, and vitamin D, in the treatment of HT.[70] Although the role of iodine in thyroid disease has been popularly discussed, most people with thyroid disturbances in iodine-replete areas have some form of AITD. Studies on the incidence of AITD are only available for a few developed countries, but it is true that AITD is more common in iodine-replete areas than in iodine-deficient ones, and even small increases in iodine intake are associated with an increased prevalence of thyroid autoimmunity.[1] With these data and recent evidence, high iodine supplementation is discouraged in HT, as it has not demonstrated any clear benefit and may even be associated with some potential harm. However, this should not preclude appropriate iodine supplementation during pregnancy.

Studies of the efficacy of selenium supplementation in HT patients have yielded conflicting results. On one hand, some metaanalyses confirmed a suppressive effect of selenium supplementation on serum TPOAb and TgAb levels.[71] However, another metaanalysis in the Cochrane library could not support or refute its efficacy.[72] There are several plausible explanations for these discrepant results, including heterogeneity of populations, differences in duration of HT and duration of the studies, the form in which selenium serum levels were measured, or the compound used as supplement. In this regard, selenium supplementation in the form of selenomethionine could be beneficial for HT patients with selenium deficiency and adequate iodine intake, but chronic intake of large quantities may have adverse health effects.

Regarding vitamin D, given the association of its deficiency with the pathogenesis of AITD, as well as the low cost and minimal side effects of vitamin D supplementation, screening for vitamin D deficiency and careful supplementation and monitoring may be recommended in HT.[73]

The association of HT with celiac disease advocates for a gluten-free diet to prevent the clinical course, progression, and complications potentially involved with HT and concomitant autoimmune diseases. However, this regime is not easy to maintain, and it is unclear whether it can truly prevent the onset of autoimmune diseases.[74]

KEY POINTS

- Hashimoto's encephalopathy is a rare autoimmune neuroendocrine syndrome that mimics various neurological and psychiatric disorders and responds to steroids.

Hashimoto's Encephalopathy

Recent clinical studies have suggested an individual susceptibility to psychopathology in patients with HT, regardless of thyroid function. The suggested mechanisms involve induction of neuroinflammation and alterations in serotonin signaling in the frontal cortex[75-78] and are different from the neurological manifestations that can appear in association with late-stage severe hypothyroidism. Moreover, vasculitis, autoantibodies targeting cells in the central nervous system, and dysregulation of hormones have also been postulated to be associated pathogenic mechanisms.

Hashimoto's encephalopathy (HE) is a rare neuroendocrine syndrome, perhaps an autoimmune disease of the brain, also known as steroid-reactive, or steroid-responsive, progressive or relapsing encephalopathy, that is associated with autoimmune thyroiditis. In fact, treatment with glucocorticoids seems to have a positive effect in the majority of cases. Despite its rarity, it is most common in middle-aged female patients.

Symptoms usually mimic a variety of neurological and psychiatric disorders, such as cognitive dysfunction, tremor, consciousness disturbances, transient aphasia, seizures, myoclonus, gait disorder, or even focal deficits mimicking stroke episodes, whereas psychotic symptoms may include paranoia, hallucinations, and mood disturbances.[79] Interestingly, the majority of patients are euthyroid or only slightly hypothyroid at the time of diagnosis, and up to one-third of patients may exhibit another concomitant autoimmune disease.

Diagnosis should be suspected in the setting of autoimmune thyroiditis with evident neurological or psychiatric symptoms, especially cerebellar disorders, seizures, and hallucinations, regardless of thyroid function. It requires exclusion of other more frequent entities. In this regard, brain imaging is nonspecific, and electroencephalogram may not be particularly informative, and only depict a regional slowdown of the frontal intermittent rhythmic delta activity, three-phase waves, sometimes epileptiform activity, and photoparoxysmal and photomyogenic phenomena. Elevated TPOAb is one of the criteria considered for diagnosis,[80] and there are some reports showing TPOAb in cerebrospinal fluid, presumably owing to intrathecal production.[17] Similar observations have been made regarding TgAb, but in lower proportions. In addition, some researchers consider that the increase in the level of antithyroid autoantibodies is proportional to the activity of the disease, and that their levels decrease after treatment with corticosteroids.[79] Proteins and pleocytosis are also frequent in cerebrospinal fluid evaluation. Antibody to alpha-enolase has been identified in some patients but is not specific to HE.

Treatment approaches may be similar to those used in any other autoimmune disease. For instance, short-term, high-dose glucocorticoids may be used, with subsequent tapering. Outcomes are generally favorable, but relapses may occur.

KEY POINTS

- Bacterial infections are responsible for the majority of acute, suppurative infectious thyroiditis. Its frequency is rather low. The clinical picture includes the sudden onset of local manifestations.

ACUTE/INFECTIOUS THYROIDITIS

Definition and Incidence

Acute, suppurative, infectious thyroiditis is usually related to bacterial infection (either gram-positive or gram-negative organisms, mainly *Staphylococcus* and *Streptococcus*), but fungal, mycobacterial, pneumocystic, or parasitic infections may also occur.[81,82] The normal thyroid gland is usually protected from infectious diseases owing to its encapsulation, high iodine content, rich blood supply, and lymphatic drainage[83]; thus, the overall frequency of this type of thyroiditis is usually low. However, in the presence of thyroid abnormalities, such as thyroid cancer, HT, multinodular goiter, or congenital anomalies, such as a pyriform sinus fistula adjacent to the larynx, or in patients with an altered and debilitated immune system, there may be a greater probability of developing opportunistic infections, via hematogenous spread.[84]

Clinical and Diagnostic Features

The main clinical manifestations of suppurative acute thyroiditis include the sudden onset of local manifestations such as dysphagia, dysphonia, anterior neck pain and erythema, and a tender thyroid mas, which may be unilateral and fluctuating, accompanied by general signs and symptoms of infection, such as fever, chills, and overall poor well-being. The onset of the clinical picture frequently occurs after an acute upper respiratory infection, which usually provides clinical guidance in addition to the clinical and analytical alterations. In the particular case of patients with acquired immune deficiency syndrome, opportunistic thyroid infections are more frequently insidious and chronic.[84]

Analytical data usually reveal findings consistent with a generalized infection, including elevation of leukocyte counts and erythrocyte sedimentation rates, while thyroid function tests are usually normal, although both thyrotoxicosis and hypothyroidism have also been reported.[83]

Regarding imaging findings, iodine uptake is usually "cold" (reduced) in suppurative areas. Thyroid ultrasound may aid in the differentiation between other types of thyroiditis if an abscess is identified. Fine-needle aspiration biopsy with gram staining and culture is the diagnostic test of choice to confirm the clinical suspicion if fluid collection with microbes on bacteriologic examination and culture are identified, together with an increased presence of polymorphonuclear cells amongst the normal thyroid follicles (Fig. 73.6B).[84]

Treatment and Follow-Up

Treatment of acute infectious thyroiditis relies on the proper administration of agent-directed antibiotics, preferably intravenously, as well as physical drainage of any abscess, if necessary. In cases of more chronic infections with bilateral disease, thyroid pain and tenderness may be less relevant, but may be accompanied by a reduced thyroid function, thus entailing the need for substitutive treatment with L-T4.

With a correct and prompt targeted therapy, prognosis is usually excellent, and recovery of the thyroid gland is usually complete. However, diagnostic and therapeutic delay may entail fatal consequences.

Postinfection follow-up depends on the specific pathogen, and initial thyroid function monitoring would be useful to rule out persistent alterations. Further radiologic studies, together with an otolaryngology consult and relevant head and neck examination, may be useful to identify fistulae or tracts contributing to the onset of infectious thyroid disease, to avoid recurrence.[84]

KEY POINTS

- Painful subacute De Quervain's thyroiditis is the main cause of thyroid pain. Its clinical features frequently involve the three-phase picture (inflammatory - recovery - stabilization).

SUBACUTE THYROIDITIS

Definition and Incidence

Painful de Quervain's SAT (nonsuppurative, giant-cell, granulomatous, pseudogranulomatous thyroiditis) is a self-limited thyroid inflammatory disorder that may occur in up to 5% of patients with clinical thyroid disease. Its incidence varies between 12 to 35 cases per 100,000/year, it is more frequent in young and middle-aged females, and decreases in frequency with increasing age.[1,85]

Several epidemiologic studies have described its higher incidence following an upper respiratory tract infection, typically 2 to 8 weeks before. An increased incidence has also been observed during the summer, suggesting seasonal incidence but also in relation to an increased peak incidence of enterovirus, coxsackievirus, mumps, measles, and adenovirus, amongst others, and thus a strong potential association with a viral cause, although viral inclusion bodies have not been observed in thyroid tissue.[85]

In fact, thyroid autoimmunity seems to have a less relevant role than in other forms of thyroiditis, even though a strong association with HLA-B35 has been described in many ethnic groups. In fact, the suggested hypothesis regarding this molecule concerns the identification of an antigen either derived from a subclinical viral infection itself or resulting from virus-induced host tissue damage. This antigen would bind to HLA-B35 molecules on macrophages and activate the cascade towards Tc cell differentiation and stimulation, which would cause tissue damage, because follicular cells have some structural similarities with the infection-derived antigen. As a result, a cytotoxic immune reaction directed to the thyroid gland occurs, but this is usually self-limited, unlike the perpetuating process occurring in HT.[34,86]

In any case, thyroid inflammation results in thyroid follicle damage, proteolysis of the accumulated Tg, and subsequent unrestrained release of thyroid hormones into the circulation, resulting a clinical and analytically evident hyperthyroidism. Normal thyroid function ceases during this thyroid alteration because of thyroid damage and to negative feedback of elevated T4 and T3 on TSH stimulation, halting thyroid hormone production. Thus, once the preformed hormone stores have been depleted, thyroid hormone levels may be restored, and, as inflammation wanes and thyroid follicular cells recover, thyroid hormone synthesis and secretion is restored. The velocity with which this process occurs may vary, making the transition from euthyroidism to hypothyroidism more or less relevant, as well as the degree of hypothyroidism, depending on how quickly the thyroid gland is able to regenerate and allow recovery of normal thyroid homeostasis.[1,85]

Clinical Features and Phases: Inflammatory, Recovery, Stabilization

SAT is the most common cause of thyroid pain. The clinical picture usually begins with a prodromal phase of generalized myalgia, pharyngitis, low-grade fever, and fatigue, resembling flulike symptoms. Patients then usually develop higher fever and severe neck pain and/or swelling; the latter may initially be confined to one lobe, but commonly spreads rapidly to involve the whole gland, a circumstance that is sometimes termed "creeping thyroiditis," and even radiate to the jaw and ears. Most patients also exhibit generalized symptoms to a variable degree and extent, including malaise, myalgia with or without arthralgia, mild feverishness, or high fever and anorexia. At this time, around half of patients may present hyperthyroid symptoms in relation to biochemical thyrotoxicosis because of thyroid hormone release form the inflamed gland. This onset of symptoms typically extends over 1 to 2 weeks and may continue with fluctuating intensity for 3 to 12 weeks, although it may last a year. Then, patients usually convert to a hypothyroid phase before returning to normal, as explained earlier.[1,85,87,88]

Diagnosis: Clinical, Laboratory, and Imaging Findings and Differential Diagnosis

The painful hyperthyroid phase may last for a couple of weeks or a few months. During this initial phase, thyroid palpation may reveal a smooth, firm enlargement, frequently accompanied by an exquisite tenderness. With time or treatment, thyroid tenderness wanes, and thyroid size generally returns to normal within several weeks or months.

Laboratory findings depend on the moment of evaluation. The typical course involves an initial phase of clinical and analytical hyperthyroidism, resulting from the initial release of thyroid hormone from the inflamed tissue during the thyrotoxic phase, with a suppressed TSH and elevation of total T4 and T3 levels (with a T3 ng/dL to T4 ug/dL ratio of typically <20). After a couple of weeks or months, a transient or permanent hypothyroid phase because of the lack of circulating thyroid hormones. Erythrocyte sedimentation rate is characteristically elevated, as well as C-reactive protein and Tg. In fact, plasma Tg may remain elevated long after all other evidence of the inflammatory process has dwindled. White blood cell count may be elevated in half of patients, so its normality does not exclude the diagnosis. Other reactants such as serum ferritin and IL-6 may be also increased during the inflammatory phase, reflecting the multisystemic nature of the process. In contrast, the antithyroid antibodies TPOAb and TgAb, as well as TRAb, may be absent or only present in low titers.

RAIU and scan typically reveals low uptake and poor visualization of the thyroid gland, as in all forms of thyroiditis. Thyroid ultrasound stereotypically shows a heterogeneously hypoechoic pattern and a suppressed vascular pattern. Application of newer technologies, such as sonoelastography, may provide further details regarding elasticity, which is decreased in SAT. In addition, T1 and T2 diffusion-weighted magnetic resonance imaging and (^{18}F)-fluorodeoxyglucose–positron emission tomography/computed tomography may aid in the differential diagnosis with GD.[87-88]

Diagnosis of SAT is mainly clinical. Differential diagnosis is usually required during the tender hyperthyroid phase, mainly with other forms of thyrotoxicosis, including GD, toxic nodular goiter, and acute thyroiditis, although basic clinical, analytical, and imaging data are generally enough to help guide the decision-making process (Table 73.2).

Pathology

Fine-needle aspiration reveals the presence of inflammatory cells, clusters of deteriorated follicular cells, and masses of colloid. A large-needle biopsy, or even a small, open surgical biopsy of the thyroid gland may be sometimes necessary to establish a definitive diagnosis. If histologic examination were always performed, a considerable inflammatory infiltrate would be easily identified, mainly composed of neutrophils, lymphocytes, and histiocytes, as well as the characteristic epithelioid granulomas, multinucleated giant cells, and follicular cells with intravacuolar granules. In addition, thyroid follicular cells would exhibit degeneration abnormalities, such as disruption, collapse, and necrosis (Fig. 73.6C). Although some fibrosis may also be evident, the eventual return of thyroid function to normal in the majority of cases is a surrogate for pathological thyroid recuperation.

Treatment and Follow-Up

In the majority of patients, SAT resolves spontaneously, but symptomatic treatment is usually prescribed to help accelerate the process. In fact, more than 90% of patients achieve complete resolution of pain and thyroid function.

In this regard, analgesic treatment with nonsteroidal antiinflammatory agents is frequently sufficient for pain relief; however, in cases with severe pain that do not respond to this initial step, a cycle of prednisone may be necessary, with daily doses of 20 to 40 mg, or equivalent, for 1 to 2 weeks, subsequently tapered down progressively over a period of around 6 weeks as symptoms wane. Occasionally, the clinical picture may be exacerbated when antiinflammatory therapy is decreased, requiring a return to a higher dose. The underlying pathogenic process cannot be altered with treatment, but the inflammatory response may be relatively tapered, allowing progression to a subclinical course.[1,87]

For thyrotoxic symptoms, beta-blocking agents are usually effective. Antithyroid drugs, on the contrary, are useless in the management of known, established SAT, as excess circulating thyroid hormone levels result from the liberation of preexisting hormones in the inflamed gland. Oral cholecystographic agents have sometimes been used to approach the hyperthyroid state when previous attempts are not effective. After this initial hyperthyroid phase of variable duration, L-T4 replacement therapy may be necessary, at least transiently, for more than low-grade hypothyroidism, but with a minimal dose to allow potential thyroid recovery. Long-term L-T4 therapy is more likely to be necessary in those patients who received corticosteroid therapy.[88]

The surgical approach is seldom required; in fact, despite the fact that it may be safe in expert hands, it is rarely used, and mainly in limited cases of persistent pain or refractory hyperthyroidism. Histologic examination of the inflamed thyroid glands has revealed an increased risk of associated papillary cancer.

Despite recovery from a clinical and analytical point of view, thyroid parenchyma may display residual scarring between thyroid follicles, resulting in residual patchy thyroid function. Indeed, around 10% of patients may suffer from residual permanent hypothyroidism and require replacement therapy, a circumstance that may occur more frequently with higher thyroid antibody levels.

KEY POINTS

- Silent thyroiditis is defined by painless thyroid enlargement, positive thyroid auto-antibodies and the three-phase archetypal stages of thyroid function test abnormalities. Diagnosis is often motivated by an incidental finding of altered thyroid hormone levels. Long-term follow-up is essential.

SILENT THYROIDITIS

Definition and Incidence

Given the somewhat similar pathological findings, painless thyroiditis has been considered a variant of HT, in the spectrum of AITD. In fact, many patients with silent thyroiditis exhibit high thyroid antibody titers and a frequent family background of AITD, and some even develop chronic autoimmune thyroiditis during follow-up, with subsequent thyroid function deficiency. In this regard, it seems that silent thyroiditis is caused by exacerbation of autoimmune thyroiditis. As with most thyroid disorders, its incidence is clearly higher in middle-aged women.[85]

Silent autoimmune thyroiditis may occur spontaneously or following pregnancy, with the latter being referred to as “postpartum thyroiditis” (see Chapter 75). It is usually defined by the combination of painless thyroid enlargement, positive antithyroid antibodies, and the typical course of thyroid function test abnormalities described for SAT (Fig. 73.1), over a 9 to 12 month course.

HLA-DR3 has been associated with the development of thyroid thyroiditis, although with a weaker relationship than in the case of HLA-B35 and SAT. Excess iodine intake and several drugs (lithium, amiodarone, tyrosin-kinase inhibitors, immune checkpoint inhibitors) and cytokines (IFN-α, IL-2) have been described as potential precipitating factors of an exacerbated immune response after some sort of underlying thyroid aberration.[89] In addition, painless thyroiditis has been associated with other immune-related circumstances, such as after glucocorticoid withdrawal, adrenalectomy for Cushing disease, lymphocytic hypophysitis, systemic lupus erythematosus, and immune thrombocytopenia, although a clear causal relationship has not been proven.

Clinical Features

Regardless of the precipitating factor determining thyroid inflammation and damage, rupture of thyroid follicles and Tg stores results in thyroid hormone release to circulation, with subsequent clinical hyperthyroidism. Then, around 5% to 20% of patients will have the archetypal sequence of hyperthyroidism, hypothyroidism, and recovery (Fig. 73.1), with the accompanying corresponding clinical symptoms, although to a milder extent. Some patients may even show additional repeated episodes over time throughout their whole life, unlike in SAT. Eventually, up to 30% of patients may develop chronic thyroiditis with permanent hypothyroidism, and some may even exhibit some degree of goiter.

Physical examination will not reveal a painful thyroid gland, but it may show some minimal diffuse enlargement.

KEY POINTS

- Radioactive iodine uptake (RAIU) is a mainstay for the differential diagnosis between thyroiditis and other causes of hyperthyroidism.

Diagnosis: Clinical, Laboratory, and Imaging Findings

Diagnosis is sometimes motivated by an incidental finding of altered levels of thyroid hormones. In such a circumstance, anamnesis may reveal that, in fact, some patients retrospectively exhibited a variable degree of symptoms that were overlooked at that time. Suspicion of painless thyroiditis should also be considered, for instance, in a middle-aged individual with short-term symptoms and no thyroid enlargement, as well as in a hyperthyroid patient who has been receiving therapy with immune checkpoint inhibitors, IFN-α, IL-2, or tyrosine-kinase inhibitors.

Laboratory findings concerning thyroid hormone levels are typically similar to those of SAT, with the hyperhypothyroidism sequence. In fact, many cases are diagnosed because of an altered TSH level, although it must be considered that changes in this parameter frequently run behind changes in free thyroid hormones. In any case, during the hyperthyroid phase, analytical perturbations may not be as intense as in GD, and only mild free hormone elevations with a relatively higher proportion of T4 than T3, or even only low serum TSH concentrations (subclinical hyperthyroidism) may be evident. During the hypothyroid phase, free hormone levels may be low, but also TSH levels, because of the time required for its recovery after the initial hyperthyroid phase. In later stages, TSH levels may be only mildly elevated, showing a state of subclinical hypothyroidism.

Acute inflammation reactant levels are not typically as elevated; in fact, the white blood cell count is usually normal, and the erythrocyte sedimentation rate and C-reactive protein levels are relatively normal or only mildly increased. However, and in a similar way to SAT, serum Tg levels are high, reflecting the proteolysis and secretion of preformed follicle stores, and this elevation may persist beyond thyroid recovery.

Antithyroid antibodies (TPOAb and TgAb), however, are typically elevated, as well as in postpartum thyroiditis, although not as much as in established HT, whilst TRAbs are usually absent or low, serving as a differential diagnosis with GD.

RAIU is low and is a mainstay for the differential diagnosis with other causes of hyperthyroidism. However, because recent exposure to large amounts of iodine may also cause low scintigraphy uptake, measurement of random urine iodine concentrations may sometimes be helpful to ensure that an elevated total body iodine content is not responsible for low thyroid uptake, to avoid misleading interpretations. Thyroid ultrasound is not usually required for the evaluation of patients with suspected thyroiditis, except for cases with concomitant goiter or palpable nodules; but, when performed, findings reveal the typical heterogeneous hypoechoic pattern with low vascularization, as observed in chronic autoimmune thyroiditis.

Similarly, although fine-needle aspiration is not part of the diagnostic routine in these patients, if performed, evaluation will show lymphocytes and macrophages, normal thyroid epithelial cells, damaged thyroid follicles, and masses of colloid. Then, during recovery, there would still be evidence of lymphocytic infiltration and maybe even some fibrosis, but more normalized thyroid follicles. In this way, the main differential characteristic from HT is that there would be a greater amount of damaged follicles, fewer lymphocytes and germinal centers, and less fibrosis.[1,85]

Once again, the differential diagnosis is usually required during the tender hyperthyroid phase, with other forms of thyrotoxicosis, and also with other causes of low RAIU thyrotoxicosis, such as iatrogenic or factitious thyrotoxicosis, ectopic thyroid hormone production, or Jod-Basedow effect because of iodine contamination (Table 73.2).

Treatment

Studies that evaluate the most optimal specific treatment for painless thyroiditis are somewhat scarce and are based on observational data and clinical experience. Thus, many patients can go without specific treatment during the different phases of thyroid dysfunction, given its frequent mild and transient signs. When a symptomatic approach is required, beta-blockers will serve to alleviate the hyperthyroid state or, in rare cases, glucocorticoid therapy may be necessary if thyroid function is severely altered. Antithyroid drugs, however, will not be useful, nor will radioactive iodine, because hyperthyroidism is not caused by increased thyroid hormone synthesis, as explained earlier.

Thyroid function should be frequently monitored, for example every 4 to 8 weeks, to anticipate the evolution to subsequent phases and eventual hypothyroidism. In the event of significant symptoms of hypothyroidism or significantly elevated levels of TSH (for example, >10 uU/mL), L-T4 replacement therapy may be useful. However, a trial of withdrawing this replacement therapy could be attempted after 3 to 6 months to evaluate potential thyroid gland recovery. In this regard, L-T4 dosage could be halved, and thyroid function tests reevaluated after 4 to 6 weeks; if TSH levels rise again, the previous dosage should be resumed, but if TSH levels remain normal, a further reduction could be probably attempted.[1,54,85]

Patients should be followed-up regarding thyroid function during the long-term, because of the possibility of recurrence and because of the potential advent of permanent hypothyroidism.

KEY POINTS

- Riedel's thyroiditis is a rare form of thyroiditis characterized by insidious and progressive fibrosis of the thyroid and adjacent structures. Clinical features include a hard and petrous enlargement of the gland and hypothyroidism.

RIEDEL'S THYROIDITIS

Definition and Incidence

Riedel's thyroiditis is a spectacular, although exceedingly rare, form of chronic fibrous thyroiditis that may occur in middle-aged women.

Etiology

The etiology of Riedel's thyroiditis has not been fully elucidated. In fact, there has long been controversy over whether it is a primary inflammatory disorder in the spectrum of advanced HT, or a local manifestation of a systemic fibrosing disease. In this regard, other fibrosing lesions, such as mediastinal fibrosis, retroperitoneal fibrosis, sclerosing cholangitis, and pseudotumor of the orbit, have been described to be also involved in this "multifocal idiopathic fibroesclerosis" systemic process.[90,91] Histopathologic examination reveals the presence of eosinophils, suggesting the role of an underlying autoimmune response to the fibrous tissue. In addition, a potential association with IgG4-related systemic disease has been hypothesized, given the observation of numerous IgG4+ plasma cells by immunohistochemistry[90] (see later).

Clinical Features

The main characteristic of this type of thyroiditis is fibrosis, which may occur in the thyroid gland itself and in adjacent structures. The thyroid gland is initially normal in size but may be considerably enlarged. Chronic inflamed fibrous tissue in the thyroid itself, and also in the adjacent skeletal muscle, nerves, blood vessels, trachea, and other distant organs, are typical findings. Indeed, involvement of the thyroid could even be regarded as part of the alteration of soft tissues in the neck, in the setting of the previously mentioned generalized fibrous state.[91,92]

The onset of symptoms is usually insidious; thyroid morphological disturbances may remain relatively stable for many years, but may slowly evolve to the onset of hypothyroidism in 25% to 30% of cases. In any case, symptoms are mainly related to compression of adjacent structures, such as the trachea, the esophagus, or the recurrent laryngeal nerve. In this regard, dyspnea, hoarseness, cough, and dysphagia may be frequent findings, and if there is enough pressure on both recurrent laryngeal nerves, there may even be stridor. Concomitantly, patients can frequently experience feelings of suffocation and local neck heaviness. If local fibrosis extends to the parathyroid glands, biochemical or even clinical hypoparathyroidism may develop. However, signs and symptoms of systemic inflammation do not occur, so patients do not show an increase in body temperature or pulse.

Physical examination reveals a remarkably distinctive characteristic: the hard and petrous consistency of the thyroid gland, which is usually enlarged, either in a symmetric or asymmetric way. This petrous consistency and invasion of surrounding structures may occur to such an extent that malignancy is the initial suspicion; however, in Riedel's thyroiditis, the absence of enlarged peripheral lymph nodes helps in the differential diagnosis.[92,93]

Diagnosis: Laboratory and Imaging Findings

Thyroid function is usually preserved, but some patients may develop hypothyroidism, or even hypoparathyroidism, owing to excessive fibrosis and impairment of normal cell function. Thus, RAIU may be normal or low. Thyroid autoantibodies are seldom positive, but, if they are, they are usually less elevated than in HT.

Ultrasonography can also be useful to guide diagnostic suspicion. In this case, Riedel's thyroiditis exhibits a homogeneously hypoechoic pattern, while elastography shows significant stiffness. Computed tomography, may display hypodensity, and magnetic resonance imaging will exhibit hypointensity on T1-weighted and T2-weighted images. Regardless of fibrous invasion in nearby soft tissue, if extrathyroidal invasion and/or vascular encasement is observed, the diagnosis of Riedel's thyroiditis is strongly suggested. Positron emission tomography reveals metabolic activity in the thyroid gland and may be useful for identification of other extrathyroidal areas of fibrosis, especially in the follow-up of patients.

Pathology

The ultimate diagnosis comes from histologic examination. Dahlgren et al. proposed the following histologic criteria for Riedel's thyroiditis: (1) fibroinflammatory process involving all or a portion of the thyroid gland; (2) presence of fibrous extension beyond the thyroid capsule into adjacent anatomic structures; (3) infiltration of inflammatory cells without giant cells, lymphoid follicles, oncocytes, or granulomas; (4) evidence of occlusive phlebitis; and (5) absence of a neoplasm.[90] Hürthle cell metaplasia, extrathyroidal extension of fibrosis, and obliterative phlebitis can aid in the differential diagnosis with the fibrotic variant of HT (Fig. 73.6D and E).

Treatment

In the absence of treatment, Riedel's thyroiditis usually develops slowly and continuously, but stabilization or even spontaneous regression have also been reported. This insidious progression entails a frequent delay in its diagnosis, with time lapses from the onset of observable signs of goiter that may range from 10 months to 2 years.

Surgical treatment may be hazardous and prone to complications, so it is not usually the mainstay of management. In fact, it is mainly indicated in two clinical scenarios: 1) to exclude malignancy; and 2) to relieve compression of adjacent structures, namely the trachea or the esophagus. Indeed, removal of the isthmus may alleviate symptoms if there is extensive invasive fibrosis, but neck dissection is not often performed, given the lack of resection planes and the risk of injuring adjacent structures. Replacement therapy with L-T4 is recommended when hypothyroidism develops, but has no effect on the pathophysiology or the progression of fibrosis.

To treat the underlying cause, immunosuppressive treatment, such as several forms of chemotherapy, glucocorticoids, tamoxifen, or the combination of glucocorticoids and tamoxifen, have been attempted in some patients, but efficacy has been low, with only sporadic improvement in some cases.[91,93,94] Awareness of other extrathyroidal fibrosing manifestations is essential in the follow-up of these patients.

KEY POINTS

- Riedel's thyroiditis is a rare form of fibrous thyroiditis whose etiology has not been fully elucidated.
- Histopathologic examination in Riedel's thyroiditis reveals a fibroinflammatory process, fibrous extension beyond the thyroid capsule into adjacent anatomic structures, infiltration of inflammatory cells without giant cells, lymphoid follicles, oncocytes or granulomas, phlebitis, and eosinophils.
- Malignancy is the main issue in the differential diagnosis of fibrous thyroiditis.
- A potential association of Riedel's thyroiditis with IgG4-related systemic disease has been hypothesized.
- IgG4 thyroiditis is characterized by an intense lymphoplasmacytic infiltration, dense interfollicular pattern of fibrosis, marked follicular cell degeneration, elevated serum IgG4, IgG4-positive plasma cells, and rapid progression to organ failure.

IgG4 THYROIDITIS

Definition, Incidence, and Relationships in the Setting of Other Subtypes of Thyroiditis

Since the first description of high serum IgG4 levels in the context of autoimmune pancreatitis and the subsequent acknowledgment that extrapancreatic organs could also be affected, IgG4-related disease (IgG4-RD), a fibroinflammatory, immune-mediated, systemic condition, has been reported to potentially occur in any organ.[95,96] Thus, the 2012 consensus guideline[97] proposed the term IgG4-RD to unify the terminology for the spectrum of these disorders (Table 73.4).

The observations that hypothyroidism and positive titers of TPOAb and TgAb are highly frequent in patients with autoimmune pancreatitis, and that the thyroid is one of the most commonly affected organs in autoimmune diseases, pointed to a potential link between IgG4-RD and the thyroid itself. In this regard, we now consider that IgG4-thyroid diseases comprise some cases of HT, its fibrotic variant, Riedel's thyroiditis, and even some peculiar cases of GD.

IgG4 thyroiditis is characterized by intense lymphoplasmacytic infiltration, a dense interfollicular pattern of fibrosis, and marked

TABLE 73.4 Diagnostic Criteria for IgG4-Related Disease (IgG4-RD)

1. Clinical examination	Clinical history, physical examination, imaging
2. Immunological examination	IgG4 in serum >135 mg/dL or elevated IgG4/IgG ratio; possibly accompanied by other laboratory tests, such as IgE, γ-globulin, or a complement
3. Histopathologic examination	Lymphoplasmocytic infiltration with storiform fibrosis and obliterative phlebitis, infiltration by IgG4-positive plasma cells (IgG4+/IgG+ >40%)
1 + 2 = possible IgG4-RD; 1 + 3 = probable IgG4-RD; 1 + 2 + 3 = definite IgG4-RD	

(Modified from Kottahachchi D, Topliss D. Immunoglobulin G4-related thyroid disease. *Eur Thyroid J.* 2016;5:231–239.)

follicular cell degeneration, in addition to other features common to IgG4-RD in other organs, such as elevated serum IgG4, mass-forming fibrosis with increased IgG4-positive plasma cells, and rapid progression to organ failure. Serum IgG4 presumably derives from IgG4+ plasma cells within the germinal center in the thyroid. In fact, circulating levels of IgG4 decrease after thyroidectomy.

The peculiar and rare fibrotic variant of HT, which was initially considered a late stage of HT, is actually considered a distinct clinicopathological entity consisting of severe compressive symptoms and firm malignant-like thyroid. It entails marked fibrotic replacement and typical inflammatory features in a lobular pattern, without extracapsular extension. The associated finding of increased IgG4+ cells and serum IgG4/IgG ratio suggests that this fibrotic variant, rather than HT as a whole, could actually be part of the spectrum of IgG4-RD.[91]

The role of IgG4 in GD is not straightforward, and the prevalence of IgG4-RD in patients with GD is presumably very low. Some particular characteristics include an older age at diagnosis, relative male predominance, a better response to antithyroid drugs, and a relatively higher rate of development of Graves' ophtalmopathy.[98,99] There is yet no strong evidence for the existence of IgG4 GD, but rather GD with a potential concomitant association with elevated IgG4.

IgG4-RD is a rather novel entity, entailing a potential underestimation of its real prevalence. For IgG4 thyroiditis, a heterogeneous prevalence has been reported, ranging from 5% to 27%. Several reasons may explain the discrepancies observed. First, there is the variability of histologic criteria and thresholds used to define increased IgG4-positive plasma cells and/or IgG4/IgG ratios across the different studies. In addition, the number of IgG4+ plasma cells depends on disease duration; specifically, these cells are characteristically increased at the onset of the disease but decrease in favor of fibrosis in more advanced stages. Furthermore, the low need for thyroidectomy in patients with thyroiditis may jeopardize the availability of a histologic definitive diagnosis. A rather surprising issue is its predominance in males, in contrast to the higher proportion of women affected in autoimmune diseases.[95,100,101]

KEY POINTS

- IgG4 thyroiditis usually entails the development of thyroid failure, tumefactive enlargement, compressive symptoms and suspicion of malignancy.

Clinical Aspects

IgG4-RD may mimic malignant disease owing to its subacute onset and the frequent development of tumefactive lesions. In the specific setting of the thyroid, it may seem similar to Riedel's thyroiditis, with a trend for more severe compressive symptoms and/or suspicion of malignancy, and subsequent relatively rapid progression to thyroid function failure owing to chronic autoimmune inflammation. The presence of pain may also serve as an indicator of the potential existence of IgG4 thyroiditis. In addition, patients may exhibit synchronous or metachronous lesions in extrathyroidal organs.[90]

Diagnosis

High levels of thyroid antibodies are frequently observed in patients with IgG4 thyroiditis. Ultrasonography usually reveals a hypoechogenic pattern.

Diagnosis relies on several key histologic features: dense lymphoplasmacytic infiltrates rich in IgG4 plasma cells, fibrosis with a storiform (irregularly whorled) pattern, obliterative phlebitis, and eosinophilic infiltration. Phlebitis without lumen obliteration and an increased number of eosinophils are considered minor nonspecific features.[97,102] IgG4 immunostaining is also important for the diagnosis.[102] Indeed, there should be positive IgG4 staining within the involved tissue, for example, thyroid or orbital tissue, for a definitive diagnosis of IgG4-induced thyroiditis or orbitopathy/exophthalmos. All other parameters are not specific or definitive for an accurate diagnosis. In this regard, the role of high circulating levels of IgG4 does not seem to be solely sufficient for a definitive diagnosis; in fact, other conditions, such as infections, connective tissue diseases, and immunodeficient states, may also exhibit high IgG4 levels, and multiple studies have reported conflicting results regarding its predictive sensitivity.

In spite of these uncertainties, today, the term "IgG4 thyroiditis" is used both when the histologic diagnosis (gold standard) is confirmed and when elevated serum IgG4 levels are present in a patient with thyroiditis, regardless of histologic confirmation, a circumstance that is rather frequent, because not many patients with thyroiditis require thyroidectomy as treatment. In an attempt to overcome this difficulty, levels of thyroid antigen–specific IgG4, rather than total serum IgG4, have been suggested to be more helpful for establishing a correct diagnosis.

Clinical, serological, and histological features need to be considered as a whole and collated with the proposed 2012 criteria to decide if a given case deserves the diagnosis of IgG4 thyroiditis.[100,101]

Treatment and Follow-Up

First-line therapy for IgG4-RD usually involves glucocorticoids, but azathioprine, mycophenolate mofetil, and methotrexate have also been commonly used as an alternative, or to maintain remission after glucocorticoid administration. In cases of recurrent or refractory disease, the anti-CD20 monoclonal antibody rituximab may be attempted with the aim of depleting B cells, as it is sometimes successful in treating Riedel's thyroiditis. Extensive and established fibrosis will encumber effectiveness of any of the attempted medical treatments, resulting in the probable need for surgery.

SUMMARY AND FUTURE DIRECTIONS

Thyroiditis comprise several forms of thyroid inflammation with variable repercussions for thyroid function. Their etiology, epidemiology, pathology, and clinical impact differ, and specific terminology has been developed accordingly. The most common forms of thyroiditis include HT, acute/infectious thyroiditis, de Quervain's SAT, silent thyroiditis, Riedel's thyroiditis, and IgG4 thyroiditis. Autoimmunity is the main etiopathogenic mechanism involved in thyroid disease, and its quintessential example is HT. Epidemiological data suggest an interaction between genetic susceptibility, existential factors, and exogenous environmental triggers in this breakdown of immune tolerance. Humoral and cellular immune mechanisms are closely related and cross-linked in AITD. Once they are triggered, they undergo subsequent feedback circuits, with combined amplification and inhibition processes, that reflect the complexity of these underlying pathways in the onset, development, and perpetuation of AITD, and that still deserve further investigation.

The various forms of thyroiditis may be interrelated, and studies have confirmed the association of AITD and other specific or systemic autoimmune disorders and thyroid malignancy. Long-term monitoring of AITD patients is required for adequate management. Further studies to clarify potential preventive or therapeutic approaches, beyond mere replacement therapy, will aid in identifying the optimal approach to treating AITD.

REFERENCES

1. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. *N Engl J Med*. 2003;348:2646.
2. Caturegli P, De Remigis A, Rose NR. Hashimoto's thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev*. 2014;13:391–397.
3. Ragusa F, Fallahi P, Elia G, et al. Hashimoto's Thyroiditis: epidemiology, pathogenesis, clinic and therapy. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101367.
4. Cogni G, Chiovato L. An overview of the pathogenesis of thyroidautoimmunity. *Hormones (Basel)*. 2013;12:19–29.
5. Ramos-Levi AM, Marazuela M. Pathogenesis of thyroid autoimmune disease: the role of cellular mechanisms. *Endocrinol Nutr*. 2016;63:421–429.
6. Antonelli A, Ferrari SM, Corrado A, et al. Autoimmune thyroid disorders. *Autoimmu Rev*. 2015;14:174–180.
7. Shukla SK, Singh G, Ahmad S, et al. Infections, genetic and environmental factors in pathogenesis of autoimmune thyroid diseases. *Microb Pathog*. 2018;116:279–288.
8. Rodríguez-Muñoz A, Martínez-Hernández R, Ramos-Leví AM, et al. Circulating microvesicles regulate Treg and Th17 differentiation in human autoimmune thyroid disorders. *J Clin Endocrinol Metab*. 2015;100:E1531–E1539.
9. Martínez-Hernández R, Sampedro-Núñez M, Serrano-Somavilla A, et al. A microRNA signature for evaluation of risk and severity of autoimmune thyroid diseases. *J Clin Endocrinol Metab*. 2018;103:1139–1150.
10. Martínez-Hernández R, Serrano-Somavilla A, Ramos-Leví A, et al. Integrated miRNA and mRNA expression profiling identifies novel targets and pathological mechanisms in autoimmune thyroid diseases. *E Bio Med*. 2019;50:329–342.
11. Wang Y, Xu F, Zhong JY, et al. Exosomes as mediators of cell-to-cell communication in thyroid disease. *Int J Endocrinol*. 2020:4378345.
12. Marcocci C, Marino M. Thyroid-directed antibodies. In: Braverman LE, Utiger R, eds. *Part II Laboratory Assessment of Thyroid Function*. Philadelphia: Lippincott Williams and Wilkins; 2005:360–372.
13. Fröhlich E, Wahl R. Thyroid autoimmunity: role of anti-thyroid antibodies in thyroid and extra-thyroidal diseases. *Front Immunol*. 2017;8:521.
14. Wiersinga WM. Thyroid autoimmunity. *Endocr Dev*. 2014;26:139–157.
15. de Carvalho G, Perez C, Ward L. The clinical use of thyroid function tests. *Arq Bras Endocrinol Metabol*. 2013;57:193–204.
16. Nielsen CH, Brix TH, Gardas A, et al. Epitope recognition patterns of thyroid peroxidase autoantibodies in healthy individuals and patients with Hashimoto's thyroiditis. *Clin Endocrinol*. 2008;69:664–668.
17. Ferracci F, Moretto G, Candeago RM, et al. Antithyroid antibodies in the CSF: their role in the pathogenesis of Hashimoto's encephalopathy. *Neurology*. 2003;60:712–714.
18. Menconi F, Huber A, Osman R, et al. Tg.2098 is a major human thyroglobulin T-cell epitope. *J Autoimmun*. 2010;35:45–51.
19. Ban Y, Greenberg DA, Concepcion E, et al. Amino acid substitutions in the thyroglobulin gene are associated with susceptibility to human and murine autoimmune thyroid disease. *Proc Natl Acad Sci USA*. 2003;100:15119–15124.
20. Weetman AP. Cellular immune responses in autoimmune thyroid disease. *Clin Endocrinol (Oxf)*. 2004;61:405–413.
21. Rapoport B, McLachlan SM. Graves' hyperthyroidism is antibody-mediated but is predominantly a Th1-type cytokine disease. *J Clin Endocrinol Metab*. 2014;99:4060–4061.
22. Rydzewska M, Jaromin M, Pasierowska IE, et al. Role of the T and B lymphocytes in pathogenesis of autoimmune thyroid diseases. *Thyroid Res*. 2018;11:2.
23. Gonzalez-Amaro R, Marazuela M. T regulatory (Treg) and T helper 17 (Th17) lymphocytes in thyroid autoimmunity. *Endocrine*. 2016;52:30–38.
24. Leskela S, Rodríguez-Muñoz A, de la Fuente H, et al. Plasmacytoid dendritic cells in patients with autoimmune thyroid disease. *J Clin Endocrinol Metab*. 2013;98:2822–2833.
25. Figueroa-Vega N, Alfonso-Pérez M, Benedicto I, et al. Increased circulating pro-inflammatory cytokines and Th17 lymphocytes in Hashimoto's thyroiditis. *J Clin Endocrinol Metab*. 2010;95:953–962.
26. Glick AB, Wodzinski A, Fu P, et al. Impairment of regulatory T-cell function in autoimmune thyroid disease. *Thyroid*. 2013;23:871–878.
27. Rodríguez-Muñoz A, Vitales-Noyola M, Ramos-Levi A, et al. Levels of regulatory T cells CD69+NKG2D+IL-10+ are increased inpatients with autoimmune thyroid disorders. *Endocrine*. 2016;51:478–489.
28. Li D, Cai W, Gu R, et al. Th17 cell plays a role in the pathogenesis of Hashimoto's thyroiditis inpatients. *Clin Immunol*. 2013;149:411–420.
29. Basdeo SA, Moran B, Cluxton D, et al. Polyfunctional, pathogenic CD161+ Th17 lineage cells are resistant to regulatory T cell-mediated suppression in the context of autoimmunity. *J Immunol*. 2015;195:528–540.
30. Li MO, Wan YY, Flavell RA. T cell-produced transforming growthfactor-beta1 controls T cell tolerance and regulates Th1- and Th17-cell differentiation. *Immunity*. 2007;26:579–591.
31. Marazuela M, García-López MA, Figueroa-Vega N, et al. Regulatory T cells in human autoimmune thyroid disease. *J Clin Endocrinol Metab*. 2006;91:3639–3646.
32. Sakaguchi S, Vignali DA, Rudensky AY, et al. The plasticity and stability of regulatory T cells. *Nat Rev Immunol*. 2013;13:461–467.
33. Vitales-Noyola M, Serrano-Somavilla A, Martínez-Hernández R, et al. Patients with autoimmune thyroiditis show diminished levels and defective suppressive function of Tr1 regulatory lymphocytes. *J Clin Endocrinol Metab*. 2018;103:3359–3367.
34. Ajjan RA, Weetman AP. The pathogenesis of Hashimoto's thyroiditis: further developments in our understanding. *Horm Metab Res*. 2015;47:702–710.
35. Li Q, Wang B, Mu K, et al. The pathogenesis of thyroid autoimmune diseases: new T lymphocytes - cytokines circuits beyond the Th1-Th2 paradigm. *J Cell Physiol*. 2019;234:2204–2216.
36. Effraimidis G, Wiersinga WM. Mechanisms in endocrinology: autoimmune thyroid disease: old and new players. *Eur J Endocrinol*. 2014;170:R241–R252.
37. Köhrle J. Selenium and the thyroid. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:392–401.
38. Winther KH, Rayman MP, Bonnema SJ, et al. Selenium in thyroid disorders - essential knowledge for clinicians. *Nat Rev Endocrinol*. 2020;16:165–176.
39. Rayman MP. Multiple nutritional factors and thyroid disease, with particular reference to autoimmune thyroid disease. *Proc Nutr Soc*. 2019;78:34–44.
40. Bellan M, Andreoli L, Mele C, et al. Pathophysiological role and therapeutic implications of vitamin D in autoimmunity: focus on chronic autoimmune diseases. *Nutrients*. 2020;12:789.
41. Chahardoli R, Saboor-Yaraghi AA, Amouzegar A, et al. Can supplementation with vitamin D modify thyroid autoantibodies (Anti-TPO Ab, Anti-Tg Ab) and thyroid profile (T3, T4, TSH) in Hashimoto's thyroiditis? A double blind, Randomized clinical trial. *Horm Metab Res*. 2019;51:296–301.
42. Bozkurt NC, Karbek B, Ucan B, et al. The association between severity of vitamin D deficiency and Hashimoto's thyroiditis. *Endocr Pract*. 2013;19:479–484.

43. Kim D. Low vitamin D status is associated with hypothyroid Hashimoto's thyroiditis. *Hormones (Basel)*. 2016;15:385–393.
44. Koehler VF, Filmann N, Mann WA. Vitamin D status and thyroid autoantibodies in autoimmune thyroiditis. *Horm Metab Res*. 2019;51: 792–797.
45. Guastamacchia E, Giagulli VA, Licchelli B, et al. Selenium and iodine in autoimmune thyroiditis. *Endocr Metab Immune Disord - Drug Targets*. 2015;15:288–292.
46. Tomer Y, Davies TF. Searching for the autoimmune thyroid disease susceptibility genes: from gene mapping to gene function. *Endocr Rev*. 2003;24:694–717.
47. Lee HJ, Li CW, Hammerstad SS, et al. Immunogenetics of autoimmune thyroid diseases: a comprehensive review. *J Autoimmun*. 2015;64:82–90.
48. Hansen PS, Brix TH, Iachine I, et al. The relative importance of genetic and environmental effects for the early stages of thyroid autoimmunity: a study of healthy Danish twins. *Eur J Endocrinol*. 2006;154:29–38.
49. Brix TH, Hegedus L. Twin studies as a model for exploring the aetiology of autoimmune thyroid disease. *Clin Endocrinol*. 2012;76:457–464.
50. McLeod DS, Cooper DS. The incidence and prevalence of thyroid autoimmunity. *Endocrine*. 2012;42:252–265.
51. McLeod DS, Cooper DS, Ladenson PW, et al. Race/Ethnicity and the prevalence of thyrotoxicosis in young Americans. *Thyroid*. 2015;25:621–628.
52. Simmonds MJ. GWAS in autoimmune thyroid disease: redefining our understanding of pathogenesis. *Nat Rev Endocrinol*. 2013;9:277–287.
53. Krassas GE, Tziomalos K, Pontikides N, et al. Seasonality of month of birth of patients with Graves' and Hashimoto's diseases differ from that in the general population. *Eur J Endocrinol*. 2007;156:631–636.
54. Aksoy DY, Kerimoglu U, Okur H, et al. Effects of prophylactic thyroid hormone replacement in euthyroid Hashimoto's thyroiditis. *Endocr J*. 2005;52:337–343.
55. Radetti G, Gottardi E, Bona G, et al. The natural history of euthyroid hashimoto's thyroiditis in children. Study group for thyroid diseases of the Italian society for pediatric endocrinology and diabetes. *J Pediatr*. 2006;149:827–832.
56. Wang SY, Tung YC, Tsai WY, et al. Long-term outcome of hormonal status in Taiwanese children with Hashimoto's thyroiditis. *Eur J Pediatr*. 2006;165:481–483.
57. Hadithi M, de Boer H, Meijer JW, et al. Coeliac disease in Dutch patients with Hashimoto's thyroiditis and vice versa. *World J Gastroenterol*. 2007;13:1715–1722.
58. Kahaly GJ, Frommer L, Schuppan D. Celiac disease and glandular autoimmunity. *Nutrients*. 2018;10:814.
59. Minelli R, Gaiani F, Kayali S, et al. Thyroid and celiac disease in pediatric age: a literature review. *Acta Biomed*. 2018;89:11–16.
60. Sibilla R, Santaguida MG, Virili C, et al. Chronic unexplained anaemia in isolated autoimmune thyroid disease or associated with autoimmune related disorders. *Clin Endocrinol*. 2008;68:640–645.
61. Fiore E, Rago T, Latrofa F, et al. Hashimoto's thyroiditis is associated with papillary thyroid carcinoma: role of TSH and of treatment with L-thyroxine. *Endocr Relat Cancer*. 2011;18:429–437.
62. Lee JH, Kim Y, Choi JW, et al. The association between papillary thyroid carcinoma and histologically proven Hashimoto's thyroiditis: a meta-analysis. *Eur J Endocrinol*. 2013;168:343–349.
63. Silva de Morais N, Stuart J, Guan H, et al. The impact of Hashimoto's thyroiditis on thyroid nodule cytology and risk of thyroid cancer. *J Endocr Soc*. 2019;3:791–800.
64. Lai X, Xia Y, Zhang B, et al. A meta-analysis of Hashimoto's thyroiditis and papillary thyroid carcinoma risk. *Oncotarget*. 2017;8:62414–62424.
65. Moon S, Chung HS, Yu JM, et al. Associations between Hashimoto's thyroiditis and clinical outcomes of papillary thyroid cancer: a meta-analysis of observational studies. *Endocrinol Metab (Seoul)*. 2018;33:473–484.
66. Rhoden KJ, Unger K, Salvatore G, et al. RET/papillary thyroid cancer rearrangement in nonneoplastic thyrocytes: follicular cells of Hashimoto's thyroiditis share low-level recombination events with a subset of papillary carcinoma. *J Clin Endocrinol Metab*. 2006;91:2414–2423.
67. Moshynska OV, Saxena A. Clonal relationship between Hashimoto's thyroiditis and thyroid lymphoma. *J Clin Pathol*. 2008;61:438–444.
68. Feller M, Snel M, Moutzouri E, et al. Association of thyroid hormone therapy with quality of life and thyroid-related symptoms in patients with subclinical hypothyroidism: a systematic review and meta-analysis. *J Am Med Assoc*. 2018;320:1349–1359.
69. Guldvog I, Reitsma LC, Johnsen L, et al. Thyroidectomy versus medical management for euthyroid patients with hashimoto's disease and persisting symptoms: a randomized trial. *Ann Intern Med*. 2019;170:453. 446.
70. Liontiris MI, Mazokopakis EE. A concise review of Hashimoto's thyroiditis (HT) and the importance of iodine, selenium, vitamin D and gluten on the autoimmunity and dietary management of HT patients. Points that need more investigation. *Hell J Nucl Med*. 2017;20:51–56.
71. Wichman J, Winther KH, Bonnema SJ, et al. Selenium supplementation significantly reduces thyroid autoantibody levels in patients with chronic autoimmune thyroiditis: a systematic review and meta-analysis. *Thyroid*. 2016;26:1681–1692.
72. van Zuuren EJ, Albusta AY, Fedorowicz Z, et al. Selenium supplementation for Hashimoto's thyroiditis: summary of a Cochrane systematic review. *Eur Thyroid J*. 2014;3:25–31.
73. Mazokopakis EE, Kotsiris DA. Hashimoto's autoimmune thyroiditis and vitamin D deficiency. Current aspects. *Hell J Nucl Med*. 2014;17:37–40.
74. Lundin KE, Wijmenga C. Coeliac disease and autoimmune diseasegenetic overlap and screening. *Nat Rev Gastroenterol Hepatol*. 2015;12: 507–515.
75. Bauer M, Goetz T, Glenn T, et al. The thyroid-brain interaction in thyroid disorders and mood disorders. *J Neuroendocrinol*. 2008;20:1101–1114.
76. Yalcin MM, Altinova AE, Cavnar B, et al. Is thyroid autoimmunity itself associated with psychological well-being in euthyroid Hashimoto's thyroiditis? *Endocr J*. 2017;64:425–429.
77. Cai YJ, Wang F, Chen ZX, et al. Hashimoto's thyroiditis induces neuroinflammation and emotional alterations in euthyroid mice. *J Neuroinflammation*. 2018;15:299.
78. Leyhe T, Müssig K. Cognitive and affective dysfunctions in autoimmune thyroiditis. *Brain Behav Immun*. 2014;41:261–266.
79. Churilov LP, Sobolevskaia PA, Stroev YI. Thyroid gland and brain: enigma of Hashimoto's encephalopathy. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101364.
80. Castillo P, Woodruff B, Caselli R, et al. Steroid-responsive encephalopathy associated with autoimmune thyroiditis. *Arch Neurol*. 2006;63:197–202.
81. Raman L, Murray J, Banka R. Primary tuberculosis of the thyroid gland: an unexpected cause of thyrotoxicosis. *BMJ Case Rep*. 2014;2014.
82. McAninch EA, Xu C, Lagari VS, et al. Coccidiomycosis thyroiditis in an immunocompromised host post-transplant: case report and literature review. *J Clin Endocrinol Metab*. 2014;99:1537–1542.
83. Farwell AP. Infectious thyroiditis. In: Braverman LE, Utiger RD, eds. *Werner & Ingbar's the Thyroid: A Fundamental and Clinical Text*. 8th ed. Philadelphia: Lippincott Williams & Wilkins; 2000:1044–1050.
84. Paes JE, Burman KD, Cohen J, et al. Acute bacterial suppurative thyroiditis: a clinical review and expert opinion. *Thyroid*. 2010;20:247–255.
85. Samuels MH. Subacute, silent, and postpartum thyroiditis. *Med Clin North Am*. 2012;96:223–233.
86. Stasiak M, Tymoniuk B, Michalak R, et al. Subacute thyroiditis is associated with HLA-B*18:01, -DRB1*01 and -C*04:01-the significance of the new molecular background. *J Clin Med*. 2020;16:534.
87. Görges J, Ulrich J, Keck C, et al. Long-term outcome of subacute thyroiditis. *Exp Clin Endocrinol Diabetes*. 2020;128:703–708.
88. Fatourechi V, Aniszewski J, Fatourechi G, et al. Clinical features and outcome of subacute thyroiditis in an incidence cohort: olmsted County, Minnesota study. *J Clin Endocrinol Metab*. 2003;88:2100–2105.
89. Iyer PC, Cabanillas ME, Waguespack SG, et al. Immune-related thyroiditis with immune check-point-inhibitors. *Thyroid*. 2018;28: 1243–1251.
90. Dahlgren M, Khosroshahi A, Nielsen GP, et al. Riedel's thyroiditis and multifocal fibrosclerosis are part of the IgG4-related systemic disease spectrum. *Arthritis Care Res*. 2010;62:1312–1328.

91. Stan MN, Sonawane V, Sebot TJ, et al. Riedel's thyroiditis association with IgG4-related disease. *Clin Endocrinol*. 2017;86:425–430.
92. Hennessey JV. Clinical review: Riedel's thyroiditis: a clinical review. *J Clin Endocrinol Metab*. 2011;96:3031–3041.
93. Falhammar H, Juhlin CC, Barner C, et al. Riedel's thyroiditis: clinical presentation, treatment and outcomes. *Endocrine*. 2018;60:185–192.
94. Soh SB, Pham A, O'Hehir RE, et al. Novel use of rituximab in a case of Riedel's thyroiditis refractory to glucocorticoids and tamoxifen. *J Clin Endocrinol Metab*. 2013;98:3543–3549.
95. Brito-Zeron P, Ramos-Casals M, Bosch X, et al. The clinical spectrum of IgG4-related disease. *Autoimmun Rev*. 2014;13:1203–1210.
96. Stone JH, Zen Y, Deshpande V. IgG4-related disease. *N Engl J Med*. 2012;366:539–551.
97. Deshpande V, Zen Y, Chan JK, et al. Consensus statement on the pathology of IgG4-related disease. *Mod Pathol*. 2012;25:1181–1192.
98. Takeshima K, Inaba H, Furukawa Y, et al. Elevated serum immunoglobulin G4 levels in patients with Graves' disease and their clinical implications. *Thyroid*. 2014;24:736–743.
99. Bozkirli E, Bakiner OS, Ersozlu Bozkirli ED, et al. Serum immunoglobulin G4 levels are elevated in patients with Graves' ophthalmopathy. *Clin Endocrinol*. 2015;83:962–967.
100. Li Y, Zhou G, Ozaki T, et al. Distinct histopathological features of Hashimoto's thyroiditis with respect to IgG4-related disease. *Mod Pathol*. 2012;25:1086–1097.
101. Rotondi M, Carbone A, Coperchini F, et al. IgG4-related thyroid autoimmune disease. *Eur J Endocrinol*. 2019;180:R175–R183.
102. Kottahachchi D, Topliss D. Immunoglobulin G4-related thyroid disease. *Eur Thyroid J*. 2016;5:231–239.

74

Hypothyroidism

Jacqueline Jonklaas

OUTLINE

Brief History, 1234
Hypothalamic-Pituitary-Thyroid Axis, 1235
Thyroid Physiology and Thyroid Hormone Action, 1235
Epidemiology, 1236
Clinical Manifestations, 1236
  Symptoms, 1236
  Signs, 1237
Etiology, 1237
  Primary Hypothyroidism, 1237
  Central Hypothyroidism, 1238
  Peripheral Causes of Hypothyroidism, 1239
Diagnosis, 1239
  Reference Intervals, 1239
  Serum Thyroid-Stimulating Hormone, 1239
  Log-Linear Relationship, 1240
  Free Thyroxine Levels, 1240
  Triiodothyronine Levels, 1240
  Thyroid Antibody Testing, 1240
  Detection of Central Hypothyroidism, 1240
  Subclinical Versus Overt Hypothyroidism, 1240
  Screening, 1241
  Normal Thyroid-Stimulating Hormone With Symptoms Overlapping With Hypothyroidism, 1241
Treatment, 1241
  Levothyroxine Therapy, 1241
  Therapy Other Than Levothyroxine, 1246
Summary and Areas of Future Research, 1248

KEY POINTS

- Thyroid hormone is important for the function of all organ systems of the body; signs and symptoms of hypothyroidism are thus diverse.
- Hypothyroidism is most commonly caused by progressive autoimmune destruction of the thyroid gland. It most frequently occurs in women and older individuals and results in a lifelong requirement for thyroid hormone replacement.
- Although a constellation of symptoms characterize the hypothyroid state, the lack of pathognomonic symptoms may delay recognition of the need for testing for hypothyroidism.
- The laboratory hallmark of primary hypothyroidism is an elevated serum thyroid-stimulating hormone concentration.
- Treatment with synthetic thyroid hormone in the form of levothyroxine lowers serum thyroid-stimulating hormone concentrations to the normal range and reverses signs and symptoms of hypothyroidism in most patients.
- Optimization of levothyroxine treatment by facilitating adherence and avoiding overtreatment and undertreatment remains a challenge.
- Combination therapy with levothyroxine and liothyronine is being actively investigated as a potential therapy for patients in whom levothyroxine alone does not reverse their symptoms.
- Future advances may allow researchers and clinicians to prevent the development of autoimmune destruction of the thyroid gland or provide therapy in the form of directing progenitor cells to form functioning thyroid follicles *in vivo*.

BRIEF HISTORY

Thyroid diseases and their treatments have been documented in both written and pictorial form throughout history. Sixth-century Chinese scholars were apparently aware of the ability of seaweeds and thyroid glands from deer and sheep to treat goiters that occurred in mountainous region of the county,[1] the latter perhaps representing one of the earliest examples of "organotherapy." The facies depicted in artwork as diverse as the Adena pipe figurine from before the first century AD[2] and the Mona Lisa from the early 1500s[3] have been attributed to hypothyroidism. Although descriptions of myxedema existed, it was not until thyroidectomy was performed in Switzerland for goiter that resection of the thyroid was clearly linked to the subsequent development of myxedema.[4] Starting during the 1890s, various modes of thyroid hormone replacement have been attempted for treating hypothyroidism, including grafting of sheep thyroid glands in patients, intravenous injection of thyroid extract in dogs, and subcutaneous administration of sheep thyroid extract in patients.[5]

Oral administration of sheep thyroid extract was then used, with extraordinary improvement in the hypothyroid symptoms and signs of the patients, and also some descriptions of the toxicity of overreplacement. Desiccated thyroid extract (DTE) then became the standard of treatment for hypothyroidism for many years. Although thyroxine (T4) itself was isolated in 1914 and synthesized in 1927, it was not synthesized for commercial purposes until 1949. Identification, isolation and synthesis of triiodothyronine (T3) occurred from 1952 to 1953.

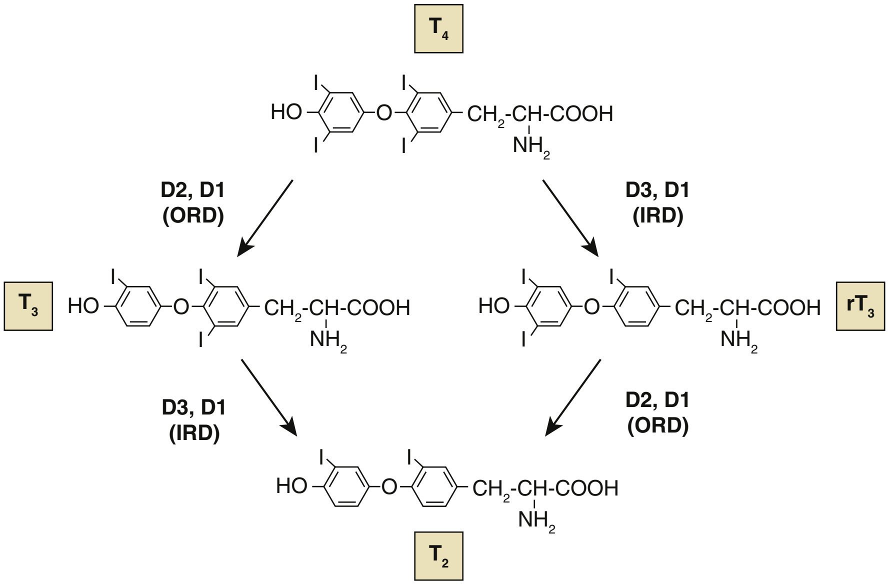


Fig. 74.1 **Deiodination of thyroid hormone isotopes.** (From van der Spek AH, Fliers E, Boelen A. The classic pathways of thyroid hormone metabolism. *Mol Cell Endocrinol*. 2017;458:29–38.)

Starting in the 1950s, levothyroxine (LT4) gradually replaced DTE as standard therapy. Although LT4 remains the most prescribed therapy for hypothyroidism, use of DTE continues, and interest in such therapy has been increasing in recent years.[6] This chapter discusses the diagnosis and treatment of hypothyroidism in adults.

HYPOTHALAMIC-PITUITARY-THYROID AXIS

The hypothalamic-pituitary-thyroid axis is a classic negative feedback loop involving four hormones. Hypothalamic thyrotropin-releasing hormone (TRH) stimulates the release of thyroid-stimulating hormone (TSH) from the pituitary gland. TSH supports the synthetic machinery within the thyroid gland itself and stimulates the resorption of thyroglobulin from within the lumen of thyroid follicles. Both T4 and T3 are then released from the thyroid gland into the circulation in a ratio of approximately 14:1.[7] There is a reciprocal relationship between TSH and free T4 (FT4), such that, as the concentration of FT4 increases or decreases, the response is an exponential decrease or increase in TSH concentration. Thus, an inverse relationship is seen when TSH concentrations are plotted on a logarithmic scale versus FT4 concentrations on a linear scale. However, this relationship may be more complex than a simple linear one, with the slope of this relationship being affected by sex, age, genetics, and use of LT4. The relationship may actually be curvilinear, as opposed to linear, and each individual may have TSH-FT4 set points that differ from the population as a whole.[8] The development of primary hypothyroidism can thus be detected when the serum TSH concentration rises above the upper limit of its reference interval.

THYROID PHYSIOLOGY AND THYROID HORMONE ACTION

Dietary iodine in the form of iodide or iodate is absorbed by the gastrointestinal tract and distributed in the extracellular fluid. Circulating iodide is actively transported into the thyrocyte by the sodium-iodide symporter located within the basolateral membrane. Once iodide has translocated to the apical membrane, its efflux is mediated by three apical iodide channels. A hydrogen peroxide generation system is present at the apical surface of the thyrocyte. Molecules of the enzyme thyroid peroxidase, also present at the apical surface of the thyrocyte, are activated by hydrogen peroxide. Oxidized thyroid peroxidase can then in turn oxidize iodide. The reactive iodonium ion intermediate formed is then covalently bound to tyrosyl residues present in thyroglobulin to generate monoiodotyrosine and diiodotyrosine residues through a process known as organification. Thyroid peroxidase also catalyzes the coupling of the monoiodotyrosine and diiodotyrosine residues to generate T4 and T3 residues in thyroglobulin, which is secreted into the follicular lumen. As needed, thyroglobulin is then pinocytosed at the apical membrane, and T4 and T3 are secreted after proteolysis of thyroglobulin. Pharmacologic amounts of iodine inhibit iodide trapping, organification, and release of the thyroid hormones.

Thyroid hormones are necessary for the development and metabolic homeostasis of all the tissues and organ systems of the body. Depending on the various estimates, the intact thyroid gland produces approximately 85 to 100 mcg per day of T4 and 5 to 6.5 mcg per day of T3. This results in the direct thyroidal production of T4:T3 in the previously mentioned approximately 14:1 ratio. The deiodinase enzymes, which are selenoproteins, are involved in activating and deactivating thyroid hormones. Type 1 and 2 deiodinases convert the precursor T4 into T3, the active form of thyroid hormone, by outer ring deiodination, producing another 26.5 mcg of T3 daily. Type 3 deiodinase is responsible for converting T4 and T3 into their inactive forms of reverse T3 and 3,3′-diiodothyronine, respectively, via inner ring deiodination[9] (Fig. 74.1).

Although thyroid hormones are lipophilic, they nevertheless require specific thyroid hormone transporters to gain entry into cells. Three transporters from two transporter families have high specificity for thyroid hormones. These are monocarboxylate transporters 8 and 10 (MCT8 and MCT10) and organic anion transporting polypeptide 1C1.[10] Deiodinases and thyroid hormone transporters are differentially expressed in different tissues and are thus essential for the tissue-specific actions of thyroid hormones. Once it has gained entry into the nucleus of the target cell, the active ligand T3 binds to thyroid hormone receptors, which in turn bind, mainly as heterodimers with retinoid-X receptors, to thyroid hormone response elements in target genes[11] (Fig. 74.2).

Two genes, *THRA* and *THRB*, encode the thyroid hormone receptor proteins (TRα and TRβ), of which there are several splice products or isoforms. TRα1 is predominately expressed in brain, heart, and skeletal muscle. TRβ1 is widely expressed, while TRβ2 is mostly expressed in the brain, pituitary, retina, and inner ear, and TRβ3 is mostly expressed in kidney, liver, and lung. This differential expression of TRα and TRβ is also essential for the tissue-specific actions of thyroid hormones. Following the binding of the heterodimer complex to the thyroid hormone response element with in the target gene, there is ensuing RNA transcription and protein synthesis to generate the cellular response to thyroid hormone. Most genes are positively regulated by T3, although some are negatively regulated.[12]

EPIDEMIOLOGY

Throughout the world, the prevalence of hypothyroidism varies from 0.25% to 4.2%.[13] Prevalence rates differ in iodine-sufficient countries versus iodine-deficient countries. For example, in an iodine-deficient country (Denmark), the overall prevalence of overt hypothyroidism was 0.37%, compared with an overall prevalence of overt hypothyroidism in an iodine-sufficient area (Norway) of 0.7%. The incidence of hypothyroidism generally increases following introduction of iodine fortification programs. Prevalence rates also differ by sex and age, with more hypothyroidism in women and in older age groups. To illustrate the effect of sex, in the 1995 report of the 20-year follow-up analysis of the Whickham population the prevalence of hypothyroidism was 1.3% and 9.3% in men and women, respectively. To illustrate the effect of age, the prevalence of hypothyroidism ranges from 1% to 2% in some studies, compared with 7% in a study of individuals aged between 85 and 89 years. Cases of subclinical hypothyroidism (SCH) exceed those of overt hypothyroidism. For example, in a study using National Health and Nutrition Examination Survey III data the overall prevalence of hypothyroidism was 4.6%, with 0.3% being overt hypothyroidism and 4.3% being subclinical disease.[14] With respect to overt hypothyroidism, the prevalence in the general population depends on the definition used and population studied and ranges from 0.2% to 5.3% in Europe and from 0.3% to 3.7% in the United States.[13]

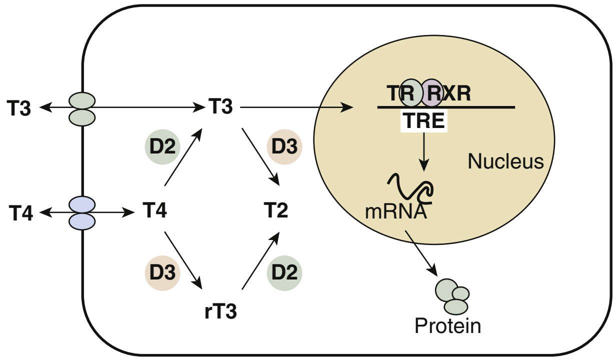


Fig. 74.2 Thyroid hormone transport, metabolism, and action in a thyroid hormone target cell. *TRE*, T3-reponsive elements; *TR*, nuclear thyroid hormone receptor; *RXR*, retinoid-X receptor. (From Groeneweg S, van Geest FS, Peeters RP, et al. Thyroid hormone transporters. *Endocr Rev*. 2020;41: bnz008.)

CLINICAL MANIFESTATIONS

The clinical manifestations of hypothyroidism are diverse and potentially emanate from the effects of thyroid hormone deficiency in any organ system of the body[15] (Table 74.1). Both signs and symptoms can occur across a wide spectrum of severity, ranging from subtle to profound.

Symptoms

The symptoms of hypothyroidism reflect the effect of low levels of thyroid hormone throughout the body. Symptoms may be mild and barely perceptible or can be severe. In addition, symptoms of hypothyroidism are nonspecific and can overlap with symptoms of nonthyroid origin. For example, weight gain can be associated with untreated hypothyroidism, but can also occur in euthyroid individuals. The same statement can be made concerning the presence of fatigue, constipation, memory problems, and dry skin. To provide some examples, in one study, dry skin was reported in 71% of those with hypothyroidism versus 54% of those who were euthyroid, muscle weakness was described in 21% of those with hypothyroidism versus 21% of those

TABLE 74.1 Symptoms and Signs of Hypothyroidism

	Presentation	Signs and Implications
General metabolism	Weight gain, cold intolerance, fatigue	Increase in body mass index, low metabolic rate, myxedema,[a] hypothermia[a]
Cardiovascular	Fatigue on exertion, shortness of breath	Dyslipidaemia, bradycardia, hypertension, endothelial dysfunction or increased intima media thickness,[a] diastolic dysfunction,[a] pericardial effusion,[a] hyperhomocysteinemia,[a] electrocardiogram changes[a]
Neurosensory	Hoarseness of voice, decreased taste, vision, or hearing	Neuropathy, cochlear dysfunction, decreased olfactory and gustatory sensitivity
Neurological and psychiatric	Impaired memory, paresthesia, mood impairment	Impaired cognitive function, delayed relaxation of tendon reflexes, depression,[a] dementia,[a] ataxia,[a] carpal tunnel syndrome and other nerve entrapment syndromes,[a] myxedema coma[a]
Gastrointestinal	Constipation	Reduced esophageal motility, nonalcoholic fatty liver disease,[a] ascites (very rare)
Endocrine	Infertility and subfertility, menstrual disturbance, galactorrhoea	Goiter, glucose metabolism dysregulation, infertility, sexual dysfunction, increased prolactin, pituitary hyperplasia[a]
Musculoskeletal	Muscle weakness, muscle cramps, arthralgia	Creatine phosphokinase elevation, Hoffman syndrome,[a] osteoporotic fracture[a] (most probably caused by overtreatment)
Hemostasis and hematological	Bleeding, fatigue	Mild anemia, acquired von Willebrand disease,[a] decreased protein C and S,[a] increased red cell distribution width,[a] increased mean platelet volume[a]
Skin and hair	Dry skin, hair loss	Coarse skin, loss of lateral eyebrows,[a] yellow palms of the hands,[a] alopecia areata
Electrolytes and kidney function	Deterioration of kidney function	Decreased estimated glomerular filtration rate, hyponatraemia[a]

[a] Uncommon presentation.
(From Chaker L, Bianco AC, Jonklaas J, et al. Hypothyroidism. *Lancet*. 2017;390:1550–1562.)

who were euthyroid, and poor memory was noted in 18% of those with hypothyroidism versus 16% of those who were euthyroid.[16] This has the dual consequences that individuals with these symptoms may be more likely to be suspected of having hypothyroidism, and that individuals with these symptoms may be diagnosed with hypothyroidism because they were screened, even if hypothyroidism was not the proximate cause of their symptoms. In addition to those symptoms mentioned earlier, other symptoms may include cold intolerance, hair loss, brittle hair and nails, muscle cramps, weakness, inability to concentrate, depression, and menometrorrhagia. In the same study as the one providing the percentages earlier, a change in symptoms in those with hypothyroidism compared with those who were euthyroid (such as being colder than the previous year or being more constipated than the previous year) had a higher likelihood ratio for hypothyroidism.[16] High likelihood ratios were evident for having a hoarser voice than the previous year (likelihood ratio 5.2; 95% confidence interval [CI]: 2.1–12.6) and finding math more difficult than the previous year (likelihood ratio 5.4, 95% CI: 2.2–13.1).

Interestingly, these types of symptoms are prevalent in those with Hashimoto hypothyroidism, even in the setting of euthyroidism. Furthermore, higher titers of thyroid peroxidase antibodies (TPO Abs) are associated with greater symptomatology.[17] For example, comparing euthyroid individuals with TPO Ab titers greater than or less than 121 IU/mL revealed statistically significant greater rates of chronic fatigue (66% vs. 49%), lack of concentration (32% vs. 19%), and nervousness (68% vs. 39%) in patients with higher titers.

Signs

Mild degrees of thyroid hormone deficiency may not result in obvious signs, except perhaps a firm goiter because of lymphocytic infiltration of the thyroid gland parenchyma. Over time, and with more decrement in thyroid hormone levels, cool and dry skin, coarse hair, loss of body hair, hoarse voice, coarse facial features, facial edema, generalized edema, bradycardia, and delayed relaxation phase of the deep tendon reflexes may be seen (Fig. 74.3). Extreme manifestations of these findings may be seen in myxedema coma.[18] Profound nail changes with horizontal ridges and marked hair loss may even occur with comparatively modest TSH elevations of 30.5 mIU/L and then be reversed with restoration of euthyroidism.[19] Manifestations of other autoimmune diseases such as vitiligo or hyperpigmentation may coexist with those of hypothyroidism.

KEY POINTS: Clinical Manifestations

- Signs and symptoms of hypothyroidism emanate from all the organ systems of the body.
- Symptoms of hypothyroidism overlap with symptoms of other conditions.
- Sign and symptoms of hypothyroidism extend across a wide spectrum from barely discernible to severe.

ETIOLOGY

The most common cause of hypothyroidism, in which the thyroid gland itself fails, is referred to as primary hypothyroidism. Primary hypothyroidism is generally considered "overt" when the TSH is elevated and the FT4 is low. Milder degrees of hypothyroidism, also known as SCH, are defined by an elevated TSH accompanied by a FT4 concentration that has not yet fallen below the normal range. Secondary or central hypothyroidism occurs when there is deficient production of TSH and is a much less common cause of hypothyroidism. Rarer causes of hypothyroidism also exist, such as TRH deficiency, thyroid dysgenesis, and "consumptive" hypothyroidism[15] (Table 74.2).

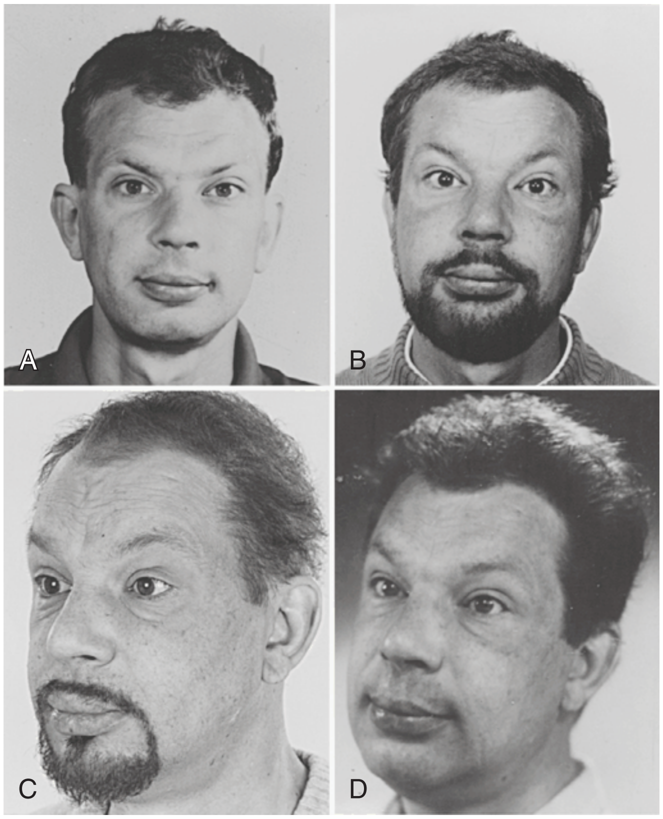


Fig. 74.3 Appearance of a 47-year-old man 12 years (A), 5 years (B), and 3 years (C) before hypothyroidism secondary to atrophic myxedema (D) was diagnosed. Note the typical myxedema face characterized by puffy, nonpitting swelling of the skin and coarse facial features.

Primary Hypothyroidism

Autoimmune Destruction. Among the causes of primary hypothyroidism, autoimmune (Hashimoto) thyroiditis is the most common cause in iodine-sufficient areas.[20] This is heralded by the presence of circulating TPO Abs, lymphocytic infiltration of the thyroid gland, and increased levels of TSH. Although the presence of TPO Abs is associated with an increased risk of progression to hypothyroidism, not all individuals with such positivity develop hypothyroidism. The pathology of Hashimoto thyroiditis involves infiltration of the interstitium around thyrocytes with lymphocytes, plasma cells, and macrophages.[20] The lymphocytes organize into lymphoid follicles, and lymphocytes that come into close contact with the thyrocytes are believed to mediate their destruction. There is also ongoing fibrosis, and the thyroid gland has a firm consistency upon palpation. Hashimoto thyroiditis can occur alone or in association with other autoimmune disorders, such as type 1 diabetes and pernicious anemia. Hashimoto thyroiditis is eight times more common in women than men, with a peak age of onset of 40 to 60 years. It occurs more frequently in White and Asian Americans than in Black Americans.

Drug-Induced. The list of medications that are associated with the development of hypothyroidism is ever-expanding. A partial list is included in Table 74.2.[15] Some of these drugs can also alter the requirement for thyroid hormone in individuals already receiving LT4 replacement therapy. Generally the effect is to cause an increased requirement. With respect to inducing the *de novo* development of hypothyroidism, several mechanisms may be responsible, such as altering thyroid hormone metabolism, altered thyroid-binding proteins, causing damage

TABLE 74.2 Causes of Hypothyroidism

Primary Hypothyroidism

- Chronic autoimmune thyroiditis (also known as Hashimoto thyroiditis)
- Iodine—severe iodine deficiency, mild and severe iodine excess
- Drugs—for example, amiodarone, lithium, tyrosine kinase inhibitors, interferon-α, thalidomide, monoclonal antibodies (e.g., ipilimumab and nivolumab), antiepileptic drugs (e.g., valproate), drugs for second-line treatment of multidrug-resistant tuberculosis
- Iatrogenic—radioiodine treatment (e.g., for Graves disease or toxic nodular disease), hemithyroidectomy, radiotherapy or surgery in the neck or head region
- Transient thyroiditis—viral (de Quervain syndrome), postpartum, silent thyroiditis, destructive thyroiditis
- Thyroid gland infiltration[a] —infectious (e.g., mycoplasma), malignant (e.g., thyroid malignancy, lymphoma, metastasis of malignancy elsewhere), autoimmune (e.g., sarcoidosis), inflammatory (e.g., Riedel thyroiditis)
- Genetic[a] —autoimmunity-related genes (e.g., *HLA* class 1 region, *PTPN22*, *SH2B3*, and *VAV3*), general and thyroid-specific genes (e.g., *F0XE1*, *ATXN2,* and *PDE86*)

Central Hypothyroidism

- Pituitary tumors (secreting or nonsecreting)
- Pituitary dysfunction (e.g., Sheehan syndrome)
- Hypothalamic dysfunction (e.g., posttraumatic)
- Resistance to thyroid-stimulating hormone (TSH) or thyrotropin-releasing hormone
- Drugs (e.g., dopamine, somatostatins, glucocorticosteroids, and retinoid X receptor-selective ligands)
- Increased TSH concentration because of leptin stimulation[b]

Peripheral (Extrathyroidal) Hypothyroidism

- Consumptive hypothyroidism
- Tissue-specific hypothyroidism owing to decreased sensitivity to thyroid hormone (e.g., mutations in *MCT8* [also known as *SLC16A2*], *SEQSBP2*, *TH RA*, *TH RB*)

[a] Rare cause of primary hypothyroidism.

[b] Evidence mainly from animal models.

(From Chaker L, Bianco AC, Jonklaas J, et al. Hypothyroidism. *Lancet*. 2017;390:1550–1562.)

to thyroid gland parenchyma or vasculature, decreasing TSH production, enhancing autoimmunity, or supplying iodine or agents with an iodine-like action. One class of drugs that is frequently associated with the development of hypothyroidism is anti-cancer drugs. Agents associated with the development of hypothyroidism include immune checkpoint inhibitors, tyrosine kinase inhibitors, interferon-α, and interleukin-2.[21]

Other Processes Resulting in Destruction of the Thyroid Gland. There are several iatrogenic means by which the thyroid gland can be destroyed. These include surgery, radioactive iodine given as treatment for hyperthyroidism or thyroid cancer, and radiotherapy for head and neck malignancies. In the case of a hemithyroidectomy, the relative risk of developing hypothyroidism is approximately 30%.[22] Infiltration of the thyroid gland associated with a spectrum of disease processes, such as sarcoidosis, hemochromatosis, hematologic malignancies, solid malignancies, and infectious agents can impair functioning to a sufficient degree to cause hypothyroidism.

Thyroiditis. Thyroiditis can be transient and cause temporary hypothyroidism, as the second phase of a three-phase process (hyperthyroidism, hypothyroidism, euthyroidism). Causes include granulomatous, postpartum, and silent thyroiditis. Alternatively, the thyroiditis can be more widespread and not associated with recovery, as may occur in silent thyroiditis in those who already have lymphocytic infiltration of their thyroid gland, and in Riedel thyroiditis.

Iodine Deficiency or Excess. The effect of iodine on thyroid function is complex. Not only is there a U-shaped curve, with insufficient iodine causing hypothyroidism and excessive iodine causing hyperthyroidism, but also introduction of iodine supplementation programs can enhance autoimmunity and change the pattern of thyroid diseases in a country.[23,24] Iodine deficiency is associated with elevation of serum TSH levels and development of diffuse and nodular goiter as the thyroid axis attempts to maintain adequate thyroid functioning by increasing iodine uptake, accompanied also by enhanced iodine recycling. As the degree of iodine deficiency worsens it can lead to hypothyroidism caused by failure of the compensation mechanisms leading to lack of the substrate needed for thyroid hormone synthesis. Addition of iodine to salt is an inexpensive and effective means of providing adequate dietary iodine. Because of iodine fortification programs, the areas of the world affected by iodine deficiency have decreased.

Exposure to iodine, in addition to reversing hypothyroidism as described earlier, can also induce hypothyroidism in those with predisposing risk factors such as prior episodes of thyroiditis or Hashimoto thyroiditis.[24,25] One way in which iodine exposure can occur is in the form of iodinated contrast media.[26] With the exposure to iodine, a normal thyroid gland will exhibit decreased thyroid hormone synthesis (Wolff–Chaikoff effect), followed by resumption of normal thyroid hormone synthesis (escape from the Wolff–Chaikoff effect). However, a gland affected by Hashimoto thyroiditis may not resume normal thyroid functioning (failure to escape from the Wolff–Chaikoff effect). In the former case of the normal gland, decreased sodium-iodine symporter expression, and possibly increased pendrin-mediated iodine efflux,[27] appear to relieve the gland of the inhibitory effects of the high intrathyroidal iodine concentration, whereas in the latter case of the gland affected by autoimmune thyroiditis, these counterregulatory mechanisms may not deploy.

Central Hypothyroidism

Secondary Hypothyroidism. Secondary hypothyroidism is characterized by insufficient TSH stimulation of a normal thyroid gland. This condition may be inherited because of several genetic defects, for which a number of candidates genes have been identified.[28] The acquired forms of secondary hypothyroidism include lesions in the sella turcica region (pituitary adenomas, craniopharyngioma, gliomas, etc.), pituitary damage owing to surgery or radiation, head injury, vascular accidents, drugs, and infiltrative processes (lymphocytic hypophysitis, sarcoidosis, hemochromatosis, infectious agents).[29]

Tertiary Hypothyroidism. In addition to pituitary dysfunction, hypothalamic dysfunction can also result in central hypothyroidism. Some of the candidate's genes associated with TSH deficiency may also be associated with TRH deficiency. Similar to the case with secondary hypothyroidism, tertiary hypothyroidism may be the consequence of hypothalamic lesions or hypothalamic injury because of surgery or radiation, head injury, vascular accidents, drugs, and infiltrative processes.[29]

Peripheral Causes of Hypothyroidism

Several rare conditions can be associated with impaired sensitivity to thyroid hormone. This was originally described as occurring with alterations in the *THRA* and *THRB* genes, but the discovery of genetic mutations and polymorphisms causing alterations in cell membrane transport (e.g., MCT8) and metabolism (e.g., selenocysteine insertion sequence–binding protein 2 and type 2 deiodinase) of thyroid hormone have led to a broader definition of impaired sensitivity to thyroid hormone that includes many defects that could interfere with the activity of thyroid hormone. Because of the different functions and tissue-specific expression of these genes, affected patients exhibit highly variable phenotypes. Some of them are characterized by a tissue hypothyroidism with well-recognizable alterations in the thyroid function tests, but others display a combination of hypothyroid and hyperthyroid manifestations.[30]

In addition to mutations causing tissue-specific hypothyroidism, an interesting and also rare cause of peripheral hypothyroidism is "consumptive hypothyroidism." This is a paraneoplastic syndrome in which a tumor produces excessive levels of the type 3 deiodinase, leading to markedly increased degradation of T4 and resultant low T3 levels, high reverse T3 levels, and elevated TSH. Large doses of LT4 are generally required until the tumor can be resected. Although originally described in tumors of vascular origin (hepatic hemangiomatosis) in infants, cases have now been described in adults because of both vascular and nonvascular tumors.[31]

KEY POINTS: Etiology

- The commonest cause of hypothyroidism is autoimmune destruction of the thyroid gland.
- Processes interfering with hypothalamic or pituitary function are less common causes of hypothyroidism.

DIAGNOSIS

Diagnosis of thyroid "disease" is based on a thyroid parameter (TSH, FT4, T3, etc.) confirmed as being outside its reference interval, combined with the signs and symptoms that would be anticipated for these "out of range" values.

Reference Intervals

As for any other analyte, the reference interval for a thyroid analyte is derived from the 95% confidence intervals for that particular analyte established in at least 120 normal volunteers.[32] These reference intervals, or decision limits, are provided to try to aid in interpretation of test results. Laboratories typically generate their own reference interval based on their local population. Theoretically, this should be a healthy population, and a minimum of 120 individuals should contribute to the database. Ensuring a healthy population without thyroid disease would involve, for example, excluding individuals with TPO Abs, those taking interfering medications, and those with goiters. If a different reference interval were deemed appropriate for a particular subpopulation, for example men versus women, this would require 120 individuals for each sex.

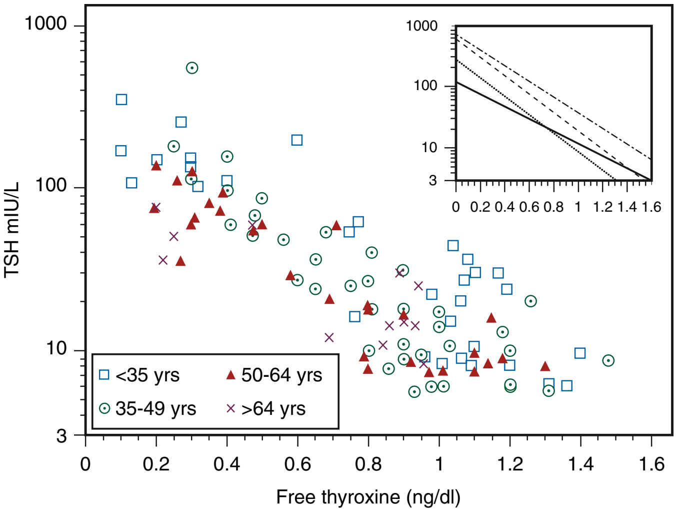


Fig. 74.4 Relationship between free thyroxine and thyroid-stimulating hormone (*TSH*) in a group of patients with hypothyroidism according to age. For all patients together the Spearman correlation coefficient = –0.795, $P < .0001$. The insert shows the slopes within the four age groups: dots and dashes <35 years, dashes 35–49 years, dots 50–64 years, solid line >64 years. *FT4*, Free thyroxine. (From Over R, Mannan S, Nsouli-Maktabi H, et al. Age and the thyrotropin response to hypothyroxinemia. *J Clin Endocrinol Metab.* 2010;95:3675–3683.)

These reference intervals are then used by various agencies to make recommendations about categorization or decisions based on the reference intervals. Given the consequences that may occur depending on the exact limits of these reference intervals, it is critical for clinicians to appreciate the factors that can influence these reference intervals. The bounds of a reference interval can, for example, result in making diagnoses that have minimal consequences or missing diagnoses in which intervention would be of benefit.[33] Natural variations in thyroid parameters such as circadian rhythms, body mass index or weight, season, and climate can potentially affect reference intervals. In addition, age, sex, thyroid hormone assays, race and ethnicity, and iodine status can also affect reference intervals.

Serum Thyroid-Stimulating Hormone

The most commonly measured thyroid analyte is the serum TSH concentration. Hypothyroidism and hyperthyroidism are diagnosed when the TSH is confirmed to be above and below the limits of the reference interval, respectively. The effect of age on the reference interval is a particularly important issue for TSH. Serum TSH concentrations tend to rise with age, such that the upper limit of the reference interval is higher in older individuals. This is a very important consideration, as the benefits of thyroid hormone treatment for mild TSH elevations are not well established in older age groups. TSH assays, however, may detect abnormal TSH isoforms with different amounts of glycosylation, such as those that may be associated with TSH-secreting pituitary tumors or normal aging.[34] Altered glycosylation can cause a mismatch between TSH concentrations and TSH bioactivity. Several factors can interfere with measurement of TSH. For example, consumption of biotin can cause TSH values to be falsely measured as low, whereas heterophilic antibodies can cause TSH values to be falsely measured as high.[35]

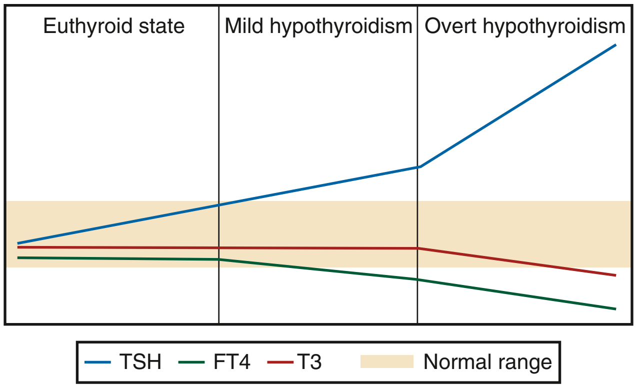


Fig. 74.5 The progression of hypothyroidism. *TSH,* Thyroid-stimulating hormone; *FT4*, free thyroxine; T3, triiodothyronine.

Log-Linear Relationship

Measurement of serum TSH is considered an ideal screening test for thyroid dysfunction, as there is an inverse log/linear relationship between TSH and FT4. This means that a small alteration in FT4 will produce a far larger change in serum TSH, thereby signaling even mild perturbations in FT4. Thus, even though TSH measurement is an indirect assessment of FT4, it is generally more sensitive for assessing thyroid dysfunction than is measuring FT4. Reliance on TSH as an indicator of thyroid status assumes that the pituitary is functional, that the individual is not suffering from a nonthyroidal illness, and that thyroid status is stable. When thyroid status is unstable, the serum TSH concentrations may lag behind the clinical picture. As previously mentioned, the TSH–FT4 relationship may be more complex than a simple linear one, with the slope being affected by sex, age, genetics, and use of LT4. For example, younger individuals may mount a more vigorous TSH response to low FT4 values than older individuals[36] (Fig. 74.4).

Free Thyroxine Levels

FT4 measurements by immunoassay are affected by protein binding and may be less accurate at extremes of binding protein concentrations, such as may occur during pregnancy. Results of FT4 measurement may be anomalous in the presence of albumin variants (e.g., familial dysalbuminemic hyperthyroxinemia), in the presence of medications that displace thyroid hormone from thyroxine-binding globulin (e.g., phenytoin or carbamazepine), or in the context of critical illness. FT4 may also be measured after free hormone has been separated from bound hormone using equilibrium dialysis or ultrafiltration. Such methods are more time-consuming and generally are less readily available. Other forms of interference include interfering antibodies, a problem that can be bypassed by using tandem mass spectrometry–based assays rather than immunoassays.[37] *In vitro* drug interference can also occur, such as the in the case of heparin causing *in vitro* generation of free fatty acids, displacement of thyroid hormone from binding proteins, and false elevation of FT4 values.[35]

FT4 levels below and above the limits of the reference limit are consistent with hypothyroidism and hyperthyroidism, respectively. Although the gold standard for assessing thyroid status has been considered to be serum TSH, some studies suggest that thyroid hormone levels may be important too. A recent systematic review found that clinical parameters representing the cardiac, bone, and metabolism systems were more associated with FT4 levels than TSH levels, suggesting that further research is needed to understand the importance of thyroid hormone levels as an indicator of thyroid status.[38]

Triiodothyronine Levels

If measuring T3, it is generally better to measure total T3 rather than free T3 because of the low concentrations of free hormone present and the lesser performance of immunoassays at these lower concentrations. Concentrations of total T3 are affected by binding protein abnormalities, as might occur with estrogen therapy and during pregnancy, and can also be affected by autoantibodies to thyroid hormone. An accurately measured elevated total T3 that is above the upper limit of the reference interval is consistent with hyperthyroidism. A total T3 that is below the lower limit of the reference interval is not necessarily sensitive or specific for hypothyroidism (Fig. 74.5). This is because total T3 levels fall later than FT4 levels in the course of development of hypothyroidism because of enhanced T4 to T3 conversion, and also because T3 levels are lowered by illness, poor nutrition, and several drugs such as beta-blockers and glucocorticoids.

Thyroid Antibody Testing

With the previously described Hashimoto disease, TPO Abs and thyroglobulin antibodies are generally present in the serum, precede the development of thyroid dysfunction, and signal the ongoing cellular damage to the thyroid gland that is occurring as hypothyroidism develops. Occasionally patients may have autoimmune thyroid disease with negative antibody results. TPO Abs are a risk factor for progression into hypothyroidism over time, and also for developing hypothyroidism after treatment with agents such as amiodarone, lithium, and interferon-α. Although changes in TPO Ab titers may reflect a change in disease activity, they are not useful for monitoring treatment for hypothyroidism. The presence of TPO Abs, however, is helpful for predicting the development of hypothyroidism during pregnancy, as well as the risk of miscarriage and the failure of *in vitro* fertilization. The presence of TSH receptor antibodies is useful for the evaluation of patients with hyperthyroidism.

Detection of Central Hypothyroidism

If central hypothyroidism caused by pituitary or hypothalamic dysfunction is suspected, reliance on the serum TSH values for diagnostic purposes is no longer indicated. A low or low normal TSH in the setting of a low FT4 should raise the possibility of central hypothyroidism. Further evaluation should include a careful patient history, the assessment of other pituitary axes, and imaging of the pituitary and sella area.[39]

Subclinical Versus Overt Hypothyroidism

SCH has generally been defined as an elevation in the serum TSH level above the upper limit of the reference interval, with a FT4 concentration that remains in the normal range. When making a determination of whether SCH is present, the reference interval being used clearly impacts the diagnosis. In the literature, upper limits of normal that have been used include TSH values of 3, 4, 5, and 6 mIU/L. Clearly, these chosen cutoffs are based on the reference interval for the particular TSH assay used, which is in turn defined by the laboratory that is providing the assay. The reference population employed by the laboratory may not be fully described, and standardized collection procedures are often not used.[40] The statistical methods used also affect the results.[41] The particular cutoff value for a normal TSH used in any one study will determine the prevalence of SCH that is identified. This choice has consequences in terms of both detection and classification of thyroid disease. Lower TSH cutoffs will have the downstream effect of identifying a greater number of individuals with SCH in whom potential treatment may be considered. An additional very important consideration before diagnosing SCH is confirming the TSH elevation.

Many above-range TSH values revert to normal upon follow-up, with the rate of normalization being inversely proportional to the degree of the TSH elevation above normal.[33]

Screening

The US Preventive Services Task Force does not recommend routine screening of adults for thyroid disease.[42] Other organizations differ as to whether they recommend routine screening or case-finding in older, asymptomatic individuals. However, with respect to criteria for population screening, hypothyroidism is prevalent and an important health problem. Moreover, diagnosis is simple and accurate, and treatment is efficacious, cost-effective, and safe. Confounding the issue are the possibilities that screening and early diagnosis may either only detect degrees of hypothyroidism that are mild and for which the benefits of treatment are less well documented or may detect disease for which treatment results in significant benefit. Although routine screening for thyroid dysfunction has not been shown to be of benefit for asymptomatic nonpregnant adults, mounting evidence suggests metabolic, cardiovascular, and skeletal risks in populations with TSH values above the normal range.[43,44] Although the benefits of treatment of overt hypothyroidism are undisputed, positive results from randomized controlled trials of treating SCH would bolster the case for screening.[45]

Normal Thyroid-Stimulating Hormone With Symptoms Overlapping With Hypothyroidism

Because symptoms of hypothyroidism are nonspecific and overlap with symptoms of other conditions, even if an individual is found to have hypothyroidism following screening based on symptoms, the symptoms may or may not resolve with initiation of LT4. Thus, the symptoms may have led to the diagnosis of hypothyroidism, despite the fact that hypothyroidism was not the cause. Individuals with a normal serum TSH who have symptoms overlapping with symptoms of hypothyroidism do not have resolution of these symptoms with LT4 therapy.[46]

KEY POINTS: Diagnosis

- Primary hypothyroidism is diagnosed based on the exponential rise in thyroid stimulating hormone that signals declining thyroid hormone levels.
- Accurate measurement of thyroid analytes is essential for correct diagnosis of hypothyroidism.
- Mild disease may apparently have relatively few consequences if undiagnosed.
- If overt hypothyroidism remains undiagnosed, significant adverse health consequence can ensue.

TREATMENT

Although there are some causes of hypothyroidism that are associated with recovery from the hypothyroid condition (thyroiditis, drug-induced hypothyroidism, iodine deficiency), most cases of hypothyroidism are associated with a lifelong need for thyroid hormone. Moreover, despite the interest in alternative therapies such as diets and nutritional supplements, the only known treatment for hypothyroidism is provision of thyroid hormone.

Levothyroxine Therapy

Use of LT4 is, in many respects, an ideal form of drug therapy. LT4 is inexpensive, chemically similar to endogenous T4 and absorbed well when administered orally and has a long enough duration of action to permit once-daily administration. With very rare exceptions, side effects seem to be related to use of inadequate or excessive doses of LT4.

Initiation of Therapy. At the time of their diagnosis with hypothyroidism, individuals may have varying degrees of residual thyroid function. Based on the relationship between declining thyroid hormone levels and stimulation of TSH production, those with SCH would still have substantial endogenous thyroid reserve, whereas those with overt hypothyroidism would have remaining endogenous thyroid function ranging from a complete absence to modest amounts. With respect to initiation of LT4 therapy, this can generally be started at a full replacement dose, so that biochemical derangements are more quickly reversed. This method does not appear to be associated with adverse cardiac consequences in those without underlying cardiac disease.[47] Recommended exceptions to immediate full replacement are noted later under “Specific Situations.”

At least two general approaches can be taken to starting doses of LT4 for replacement, both generally leading to similar estimates. One approach is to select the starting dose based on the degree of TSH elevation. Several formulae are available for guidance, should that approach be taken.[48,49] The more common approach, however, is to calculate the starting dose using a weight-based formula. Assuming little or no residual thyroid function, weight-based replacement doses are on the order of 1.6 to 1.8 mcg/kg.[7] The purest example of this would be a patient who is athyreotic as the result of a thyroidectomy.[50] More accurate estimates may be obtained if ideal body weight or lean body mass are used, rather than actual body weight.[7,51] The consequence of this may be that, if actual body weight is used in an overweight or obese individual, then the dose selected may be too high. The required dose may also be affected by the TSH goal of the treatment, with higher doses needed to achieve TSH values in the lower half of the normal range.[7] The effect of age is generally to decrease a patient’s anticipated requirement, and this phenomenon may be mediated through body weight.[52] Pregnancy will also lead to an increased requirement for LT4, such that dosage increases are needed to maintain a normal serum TSH.

Adjustment of Therapy. LT4 has a half-life of approximately 1 week, such that a new steady state with respect to thyroid parameters (TSH, FT4, T3) is generally achieved by 6 weeks after a dosage initiation or a dosage change. Based on this reassessment of thyroid status, the LT4 can be adjusted up or down and thyroid parameters rechecked in a further 6 weeks. If the TSH is outside of the desired range by a modest amount, an adjustment by one dosage increment (e.g., from 100 mcg to 112 mcg) may be sufficient. For larger deviations from the desired range, an increase of several increments may be needed. Although simple dosage regimens are preferable to facilitate adherence, some patients may require alternating doses (such as 137 mcg alternating with 150 mcg) or may need small adjustments that can be achieved by taking a half tablet or one-and-a-half tablets on one day of the week, while taking the usual one tablet on the other days of the week. Once a patient has achieved the desired serum TSH level, a confirmation that the TSH remains normal at 3 or 6 months may be indicated. However, once a correct dose has been confirmed, usually monitoring of serum TSH and FT4 on an annual basis is sufficient. Additional or earlier laboratory testing to confirm the euthyroid state may be needed if a patient has gained or lost a significant amount of body weight, or if the patient has new medical conditions or new medications that might affect their requirement for LT4.

Although the gold standard for assessing thyroid status has been considered to be the serum TSH, some studies suggest that thyroid hormone levels may be important too. As previously mentioned, a recent systematic review found that clinical parameters representing the cardiac, bone, and metabolism systems were more associated with FT4 levels than TSH levels, suggesting that further research is needed to understand the importance of monitoring and potentially adjusting therapy based on thyroid hormone levels as well as TSH levels.

Goals of Therapy

Standard-of-Care Goals. The goals of treating hypothyroidism are to restore the patient to a euthyroid status by reversing the signs and symptoms of hypothyroidism and restoring the patient's TSH and thyroid hormone levels to normal. The biochemical goals are readily achieved, although T3 levels, if assessed, may be in the lower part of the normal range, particularly in athyreotic patients.[53] With respect to signs of hypothyroidism, although weight gain is associated with the hypothyroid state, reversal of hypothyroidism, although achieving a median TSH that was decreased from 18.3 mIU/L to 2.3 mIU/L, was not associated with any weight loss.[54] Even treatment of more severe hypothyroidism, despite lowering the TSH from 102 mIU/L to 2.2 mIU/L, was associated only with a weight loss of 4.3 kg. This was not caused by a decrease in fat mass, but a decrease in the mass of the lean body compartment only.[55] Although symptoms may improve following LT4 therapy, they may not reach the level of improvement anticipated if the control population is used as the comparison. For example, in a study in which quality of life was assessed using a thyroid specific instrument (ThyPRO), after 6 months of therapy, during which the median TSH was decreased from 8.1 mIU/L to 2.6 mIU/L, improvement was noted in 11 out of 13 subscales compared with baseline, but was only improved compared with the general population in seven out of 13 subscales.[56] Additional goals are to avoid iatrogenic hypothyroidism or iatrogenic hyperthyroidism, which unfortunately occurs in up to 40% of patients being treated for hypothyroidism.[57,58]

Avoidance of long-term iatrogenic hypothyroidism or hyperthyroidism is desirable in all patients treated for hypothyroidism. However, this is particularly important in vulnerable populations such as patients who are pregnant, elderly, or have cardiac disease and other medical conditions. Thus, the ideal TSH value for a particular patient may depend on the patient's age, coexisting medical conditions, and any benefits that may be associated with specific TSH values. A particular TSH target can be achieved by serial adjustment of a patient's LT4 dose. Maintenance of that TSH should then be possible with continuation of that same dose if no other influences act to perturb the equilibrium. Despite such ease of adjustment, many patients with hypothyroidism have out-of-range TSH values. Of a population attending a health fair, 18% were receiving inadequate LT4 therapy, and 22% were receiving excessive LT4 therapy.[58] Other studies show similar percentages of patients with abnormal TSH values.[7] This is particularly a problem in those over 65 years of age, where iatrogenic thyrotoxicosis is especially common.[57] This is of particular concern in patients with comorbidities, which primarily affect elderly patients. One risk of overtreatment with LT4 is adverse effects on the heart, which manifest as decreased cardiac reserve and impaired exercise capacity, and, based on data from populations with both exogenous and endogenous hyperthyroidism, an increased risk of atrial fibrillation. The other major adverse consequence is accelerated bone loss, which may be compounded by hypogonadism in the elderly. Undertreatment with LT4 is also associated with adverse outcomes. Whenever possible, therapy should be adjusted to avoidance iatrogenic hypothyroidism or hyperthyroidism. Serum TSH should be checked annually, and perhaps more frequently in those who have a history of dose instability. Alterations in weight, as well as new medical comorbidities, medications, or symptoms should also trigger consideration of whether repeat laboratory assessment is indicated.

Additional Goals. Other goals of therapy should also be considered. In two studies funded by a pharmaceutical company, frequent LT4 dosage changes were associated with lost time and wages for patients, with reduced patient satisfaction. Based on these studies, it would seem to be important to ensure stability of treatment by minimizing influences that may lead to changing dose requirements.[59,60]

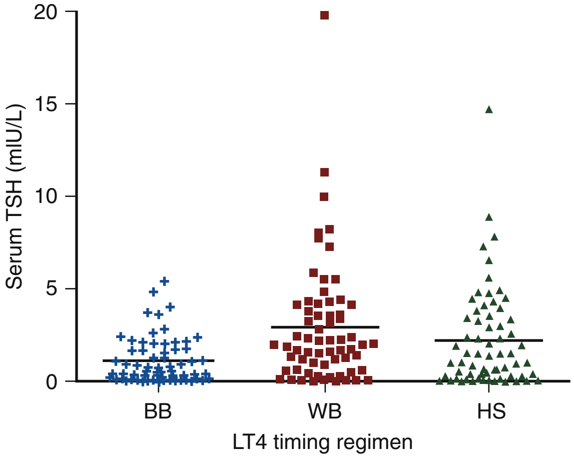


Fig. 74.6 **The effect of the timing of levothyroxine (LT4) administration on serum thyroid-stimulating hormone (*TSH*).** (From Bach-Huynh T, Nayak B, Loh J, et al. Timing of levothyroxine administration affects serum thyrotropin concentration. *J Clin Endocrinol Metabol.* 2009;94:3905–3912.)

Optimization of Therapy. A consistent schedule for daily administration of LT4 seems desirable to facilitate adherence.

Scheduling of Levothyroxine. More important than the particular time that is identified for taking LT4 is maintaining a regular schedule so that absorption varies as little as possible and so that doses are not omitted.[7] Potential times for LT4 administration are 60 or 30 minutes before breakfast, with breakfast, and at bedtime. A wealth of studies have compared these regimens, and in general the schedules that are most associated with maintaining a TSH that is consistently within the normal range are 60 minutes before breakfast and at bedtime[61] (Fig. 74.6). Taking LT4 with breakfast results in the most variation in serum TSH. A consistent schedule should be adopted that is convenient to the patient and the least disruptive to their daily routine. Regimens that involve the patient awakening early from sleep do not seem optimum. The use of a prefilled tablet container is encouraged, particularly for those individuals who are taking multiple medications. If an LT4 dose is omitted, the options include taking the dose are soon as the omission is noted or taking two doses the following day.

Absorption of Levothyroxine. LT4 absorption occurs mostly in the jejunum and ileum. The maximum absorption of LT4 is approximately 75% to 80%, and this occurs when the patient consumes LT4 in a fasting state. The presence of an acidic environment in the stomach also seems to promote absorption.[7] Any food that is coadministered with LT4 can potentially reduce its absorption. Soy products, fiber, and milk, for example, can reduce absorption. Beverages such as Italian espresso can also impair absorption. The manifestation of impaired absorption is an elevated TSH. Medications and supplements are also known to reduce absorption, with common culprits being calcium carbonate, iron, multivitamins, and proton pump inhibitors[62-64] (Table 74.3). Calcium carbonate and iron seem to impair LT4 absorption by adsorbing to LT4 in the stomach. Proton pump inhibitors may exert their effects by reducing stomach acidity. Vitamin C stands alone as the one product that increases LT4 absorption and may have this effect by virtue of increasing stomach acidity. A particular example where the reduced LT4 absorption could have adverse effects for the patient is that seen when LT4 is given concurrently with enteral tube feeds in a hospitalized patient.[65] The impaired absorption associated with foods, beverages, and medications may be overcome by changing the time of the LT4 administration, or a higher dose of LT4 may be administered to compensate for the effect of the decreased absorption.

TABLE 74.3 Medical Conditions, Foods, and Drugs Affecting Levothyroxine (LT4) Absorption

Factor	Strength of Evidence	Result of Interaction (Clinical Significance)	Recommendation
Disorder			
Celiac disease	Very good	↑↑ need for LT4 (clinically significant)	Caution during treatment because of iatrogenic hyperthyroidism
Lactose intolerance	Very good	↑↑ need for LT4 (clinically significant)	Caution during treatment because of iatrogenic hyperthyroidism
Atrophic gastritis and *Helicobacter pylori*	Very good	↑↑ need for LT4 (clinically significant)	Caution during treatment because of iatrogenic hyperthyroidism
Giardiasis	Poor	↑↑ need for LT4 (clinically significant)	Caution during treatment because of iatrogenic hyperthyroidism
Bariatric surgery*	Controversial	↑/↓ need for LT4 (clinically significant)	Monitor thyroid hormones after surgery
Food			
Fiber	Controversial	↑ need for LT4 (clinically significant)	Avoid concomitant use
Soybeans	Very good	↑↑ need for LT4 (clinically significant)	Avoid concomitant use
Coffee*	Very good	↑↑ need for LT4 (clinically significant)	Avoid concomitant use
Grapefruit	Good	↑ need for LT4 (clinically significant)	Avoid concomitant use
Vitamin C[†]	Very good	↓↓↓ need for LT4 (clinically significant)	Can improve absorption
Drug			
Aluminum hydroxide	Very good	↑↑ need for LT4 (D)	Avoid concomitant use for >2 h
Calcium salts	Very good	↑↑ need for LT4 (D)	Avoid concomitant use for at least 2 h
Iron preparations	Very good	↑↑ need for LT4 (D)	Avoid concomitant use as much as possible
Lanthanum	Very good	↑↑ need for LT4 (D)	Avoid concomitant use
Orlistat	Controversial	↑↑↑ need for LT4 (D)	Avoid concomitant use for at least 2 h
Polystyrene sulfonate	Poor	↑↑↑ need for LT4 (D; moderate)	Avoid concomitant use
Raloxifene	Good	↑↑ need for LT4 (D; minor)	Avoid concomitant use as much as possible
Bile acid sequestrants	Very good	↑↑↑ need for LT4 (D; moderate)	Avoid concomitant use as much as possible
Sevelamer	Very good	↑↑↑ need for LT4 (D; moderate)	Avoid concomitant use
Simethicone	Poor	↑↑ need for LT4 (D; moderate)	Avoid concomitant use
Ciprofloxacin	Very good	↑↑↑ need for LT4 (C; moderate)	Avoid concomitant use
Rifampicin	Very good	↑/↓ need for LT4 (C; moderate)	Monitor thyroid hormones
Sucralfate	Controversial	↑↑↑ need for LT4 (C; moderate)	Avoid concomitant use
Chromium	Very good	↑ need for LT4 (B; moderate)	Monitor thyroid hormones
H_2-receptor antagonists	Controversial	↑ need for LT4 (B)	Monitor thyroid hormones
Proton pump inhibitors[a]	Good	↑↑ need for LT4 (B; moderate)	Monitor thyroid hormones

↑ = increase; ↓ = decrease.

Interaction was refuted for soft gel capsules and liquid form of LT4.

[a] Patients with impaired absorption.

(From Skelin M, Lucijanić T, Amidžić Klarić D, et al. Factors affecting gastrointestinal absorption of levothyroxine: a review. *Clin Ther.* 2017;39:378–403.)

In addition to decreased LT4 absorption being associated with the food and medications listed earlier, several gastrointestinal conditions may also be associated with a decrement in LT4 absorption. These include celiac disease, lactose intolerance, ulcerative colitis, and atrophic gastritis associated with *Helicobacter pylori* infection.[62,66] Patients with these conditions may require greater than the predicted weight-based LT4 dose before their treatment, but with resolution of the higher requirement if the condition is successfully treated. The balance of the competing effects of weight loss and malabsorption after bariatric surgery most commonly seems to result in patients requiring a reduced LT4 dose after surgery.[67]

Drugs Affecting Levothyroxine Dose. In addition to the drugs and conditions mentioned earlier that alter LT4 dose requirement by affecting, and usually reducing, absorption, a number of drugs may also alter LT4 dose requirement through their effects on LT4 transport and metabolism (Table 74.4). The classic examples of drugs affecting LT4 transport that increase and reduce the requirement for LT4, respectively, are estrogen and testosterone.[7,63] These influences occur because of the effects of these sex steroids to increase (estrogen) and decrease (testosterone) the concentration of thyroxine-binding globulin. LT4 is metabolized by deiodination, glucuronidation, and sulfation, and many drugs act to increase its metabolism. The conjugation with glucuronates and sulfates occurs in the liver, and several drugs act as hepatic enzyme inducers and enhance LT4 metabolism. Examples of such drugs are phenobarbital, phenytoin, carbamazepine, rifampin, and nicardipine.[7,68] Other drugs such as tyrosine kinase inhibitors may act by altering deiodinase activity. The increased TSH that occurs in previously euthyroid patients taking LT4 when drugs such as sunitinib and sorafenib are initiated appears to be caused by increased activity of

TABLE 74.4 Examples of Medications Causing Altered Transport or Metabolism of Levothyroxine (LT4)

Alteration in LT4 transport	Medication	Effect
Increased TBG	Estrogen	Increased LT4 requirement
	Capecitabine	Increased LT4 requirement
Decreased TBG	Androgens	Decreased LT4 requirement
Increased TBG or increased transport	Estrogen, raloxifene, tamoxifen, heroin, methadone, 5-fluorouracil, mitotane, clofibrate	May not cause thyroid dysfunction
Decreased TBG or decreased transport	Androgens, anabolic steroids, glucocorticoids, nicotinic acid, salicylates, furosemide, heparin, NSAIDs, phenytoin, carbamezepine	May not cause thyroid dysfunction
Alteration in LT4 Metabolism	**Medication**	**Effect**
Increased hepatic metabolism via induction of P450	Phenobarbital, rifampin, phenytoin, sertraline, carbamazepine	Can increase LT4 requirement or may not cause thyroid dysfunction
Decreased activation of T4 to T3 via inhibition of type 2 deiodinase	Beta-blockers, steroids, PTU, radiographic contrast agents, amiodarone	Can increase LT4 requirement or may not cause thyroid dysfunction
Increased deactivation of T4 and T3 via acceleration of type 3 deiodinase	Tyrosine kinase inhibitors (imatinib, sorafenib, motesanib, sunitinib)	Can increase LT4 requirement

TBG, thyroxine-binding globulin; *NSAID,* nonsteroidal antiinflammatory drugs; *PTU,* propylthiouracil.

the type 3 deiodinase, as evidenced by decreased T4 and T3 levels and increased reverse T3 levels.[69]

Levothyroxine Preparations. LT4 is available as tablet preparations and also as a liquid preparation, or as a liquid contained within a gel capsule. The various LT4 tablets include brand-name and generic preparations. The tablet preparations are all similar in the content of their active ingredient (LT4 itself) but differ with respect to their inert ingredients in the form of excipients (mixers, fillers, and colorants). The product that is most affordable to a patient may vary according to the patient's insurance company and pharmacy, and also may not stay constant over time. There are reports of significant changes in serum TSH when a patient's LT4 is changed either from one brand name to another, or from a brand name to generic product, or from one generic to another.[7] The altered TSH is thought to occur because the various products may have different absorption because of the different excipients. Because of their different absorption, each of these products has slightly different bioequivalence.

Bioequivalence is a measure of absorption or bioavailability that the US Food and Drug Administration (FDA) uses to predict therapeutic efficacy. Bioavailability is determined by studying the pharmacokinetics of large oral doses of LT4 in volunteers who have normal endogenous thyroid function.[7] Pharmacokinetic parameters including the maximum serum concentration (Cmax), time to Cmax, and area under the concentration-time curve (AUC) are used to compare different products. If the two 90% confidence intervals from the natural logarithms of the AUC at 48 hours and the Cmax are both within the 80% to 125% range, the products are deemed bioequivalent by the FDA and can be substituted for each other. Measures of systemic exposure (AUC, Cmax) are used in this testing, rather than clinically relevant or biochemical endpoints such as serum TSH levels. As such bioequivalence methodology does not detect LT4 doses with approximately 12.5% difference from each other (e.g., 100-mcg and 112-mcg tablets), it is always advisable to recheck serum TSH in individuals who have been switched from one preparation to another to be sure that iatrogenic hyperthyroidism or hypothyroidism has not developed and that the serum TSH is still in the target range.[7] Furthermore, where possible it is best to maintain a patient on the same "identifiable" product. The name of the manufacturer is on the label of the bottle, so patients can request LT4 made by the same manufacturer when obtaining refills.

Liquid LT4 preparations are available in the United States in which the LT4 is dissolved in glycerol and contained either within a gelatin capsule or in a solution. Liquid preparations are reported by the manufacturer to have 100% absorption from the gastrointestinal tract, compared with approximately 75% absorption in the fasting state for other LT4 preparations. The dissolution of the gelatin capsule preparation appears to be relatively unaffected by pH. Therefore, its absorption, in theory, might be less affected by the altered pH of the gastric environment associated with food consumption, atrophic gastritis, and use of proton pump inhibitors than other LT4 products. There is, in fact, a growing body of evidence that liquid preparations may have better absorption in situations where patients are receiving proton pump inhibitors[70] or tube feeds[71] or have other causes of malabsorption.[72] It is also plausible that gelatin capsules may be better tolerated than tablets in individuals with allergies to the excipients contained within tablet preparations.

Levothyroxine Therapy in Specific Situations

Pregnant Patients. During pregnancy it is recommended that patients being treated for hypothyroidism be frequently monitored so that their TSH values can be maintained in the lower half of the trimester-specific reference range. When such reference ranges are not available, a TSH below 2.5 mIU/L is a reasonable goal.[73] Women being treated for hypothyroidism typically require a 20% to 30% increase in their LT4 dose early in the first trimester of pregnancy. The magnitude of the increase is greater in thyroidectomized patients than in those with some residual thyroid function.

Pediatric Population. Treatment of hypothyroidism in the pediatric population is discussed elsewhere. The management of hypothyroidism in children is similar to its management in adults. However, there are special considerations based on the requirement of normal thyroid function for neurocognitive development as well as growth and development. As a reflection of their unique physiology, newborns, children, and adolescents typically require higher LT4 doses than adults. As an individual advances from childhood into adulthood, LT4 replacement doses decrease, with newborns typically requiring 10 mcg/kg/day, 1-year-old children 4 to 6 mcg/kg/day, and adolescents 2 to 4 mcg/kg/day, with transition to the average adult dose of 1.6 mcg/kg/day once endocrine maturation is complete.

Nonadherent Patients. If patients take a consistent dose of LT4, their serum TSH levels should remain within a fairly narrow range. If patients experience unexpected fluctuations in their serum TSH, or persistently elevated TSH concentrations despite the administration of large doses of LT4, factors affecting LT4 formulation, absorption, and metabolism should be investigated as potential causes. If, however, such factors do not appear to be responsible, variable adherence or nonadherence to LT4 therapy should be considered. Although one survey of patients prescribed LT4 found a self-reported nonadherence rate of 22%, a study of a large insurance claims database indicated nonadherence rates of 40% to 52%. Using another database, adherent patients had lower overall healthcare costs and less diagnosis of other diseases than nonadherent patients. However, it should be noted that the latter two studies were funded by a pharmaceutical company. Because LT4 is generally a lifelong medication, it is important for patients to identify a medication schedule that facilitates adherence. Patients may be taking multiple other medications, including those that require specific administration conditions (e.g., bisphosphonates) or those that impair absorption. It may be necessary to choose a schedule of medication administration that is practical, even if absorption is affected, if this ensures that doses are not omitted.

If a combination of a high TSH and a high-normal or high T4 concentration is documented by laboratory testing, this pattern could be consistent with the syndrome of thyroid hormone resistance, a TSH-secreting pituitary adenoma, or recent resumption of LT4 intake before a scheduled blood test. Intuitively, patient education regarding the benefits of euthyroidism and the risks of iatrogenic thyroid disease would seem to a logical approach to reducing nonadherence. However, a study providing education in the form of booklets mailed to patients' homes did not affect serum TSH, which was used as a surrogate marker of adherence. If efforts to encourage regular daily consumption of LT4 are unsuccessful, options include observed therapy, including twice weekly or weekly therapy. Weekly therapy is associated with supratherapeutic concentrations of FT4 for about 24 hours and elevated TSH levels prior to the administration of the weekly dose, although FT3 levels remain within the normal range, and patients appear not to report side effects. Parenteral administration of LT4 is also possible.

Central Hypothyroidism. When treating patients with hypothyroidism because of TSH deficiency, a parameter other than serum TSH is needed to guide LT4 replacement therapy. Based on a randomized trial comparing two doses of LT4, it has been recommended that FT4 levels be kept in the upper half of the normal range.[7] Slightly lower FT4 levels, perhaps in the mid-normal range, have been suggested for frail or older individuals.

Patients with Thyroid Cancer. A subgroup of patients who are intentionally kept with TSH values below the normal range are those with intermediate or high-risk differentiated thyroid cancer. Studies show either increased survival or increased relapse-free survival in these risk categories with TSH suppression, but not in patients with low-risk thyroid cancer, who can be maintained with a serum TSH in the lower half of the normal range. In higher-risk patients, the benefits of TSH suppression and the risks, such as potential bone loss and adverse cardiac effects, have to be balanced against potential benefits.

Older Patients and Patients With Medical Conditions. Patients older than 60 to 65 years of age, patients who have severe, longstanding hypothyroidism, or those with coronary artery disease need a gradual approach to LT4 replacement. In patients known to have ischemic heart disease, treatment should be initiated with lower doses of LT4 such as 25 mcg daily. In other patients at risk for coronary artery disease, but without documentation of such disease, a conservative starting dose of approximately 50 mcg per day may be advisable. This caution would apply to patients who are in older age groups or who have had longstanding severe hypothyroidism. Older individuals may have other chronic medical conditions that may require altered doses of LT4 to maintain a normal TSH. In addition, euthyroidism should be carefully maintained to avoid iatrogenic thyroid disease that may exacerbate the patient's underlying medical condition. For unclear reasons, sick patients older than 65 years who are taking other medications for a variety of comorbidities in addition to LT4 require lower doses of LT4 to normalize their serum TSH than do healthy controls of a similar age who taking only LT4. The difference in LT4 dose requirement persisted even after accounting for body weight. As other examples of medical conditions, patients with glomerular disease may require higher LT4 doses at times when their urinary protein losses are greatest,[74] and those with cirrhosis may also have an increased requirement for LT4, possibly because of increased levels of thyroxine-binding globulin.

Hospitalized Patients. When patients with hypothyroidism are hospitalized, their LT4 therapy should continue uninterrupted. If patients are unable to take their LT4 by mouth, it should be provided by other enteral routes such as via feeding tubes or the intravenous route, as appropriate. However, in a review of LT4 replacement therapy in an intensive care unit, it was found that 17% did not have their LT4 prescribed for more than 7 days, and 21% did not have their LT4 administered when enteral feeding was instituted.[75] Patients with TSH levels either above or below the reference range had longer median hospital lengths of stay (22 days) than those with normal TSH values (15 days).[75]

Coexistent Adrenal Insufficiency. If a patient has both hypothyroidism and suspected or documented adrenal insufficiency, cortisol replacement should be started concurrently with LT4 to avoid restoration of normal renal function increasing cortisol disposal, or normalization of metabolic rate increasing the requirement for cortisol, and thus precipitating an adrenal crisis.[7]

Subclinical Hypothyroidism. SCH is the one of the most prevalent thyroid diseases in developed countries, with a reported overall prevalence rate of 4.3% in a study of 17,353 individuals in the United States.[14] This prevalence was higher, at 9%, when individuals attending a health fair in the United States were studied.[58] Other studies of the general population reported prevalence rates of 4.9% and 3% in women and men, respectively, in the Netherlands, 7.5% and 2.8% in women and men respectively, in the United Kingdom, and 11% and 7.3% in women and men, respectively, in China.[13] However, different prevalence rates can be observed depending on the specific subpopulation studied. For example, when examining individuals over 60 to 65 years of age, the prevalences for women and men, respectively, were 7.1% and 2.7% in the United States, 11.6% and 2.9% in the United Kingdom, and 6.7% and 6.1% in Brazil.[13]

Each of the studies cited earlier, as well as other studies examining the prevalence of SCH, used a variety of different upper limits of normal for the TSH reference interval that was used to define SCH. Arguments have been made both in favor of lowering the upper end of the reference interval and in favor of maintaining it. The proponents of lowering the upper limit of the normal TSH range suggest that patients may have mild symptoms and abnormalities in lipids and cardiovascular function that can be improved with LT4 therapy. The concern would be labeling a larger population with having SCH, with treatment potentially being undertaken, without proven benefits of lowering TSH values that fall within this range. Indeed, it has been suggested that treatment decisions should not be guided so much by reference intervals, but by data showing benefit of treatment for particular degrees of TSH elevation. Moreover, reference intervals themselves may be better defined by the risk of disease rather than on population-based ranges.

There is an ongoing debate regarding the risks and benefits of treating SCH.[33] Many reviews regarding SCH conclude that there are less adverse consequences of mild SCH, compared with greater degrees of

SCH, in most populations, and that this is particularly true in specific populations such as the elderly.[76] This discussion has to be qualified with a consideration of whether specific populations can be defined as having SCH if a TSH reference interval that is derived from that particular population has not been used. For example, does a 75-year-old with a persistent TSH value of 5.5 mIU/L have SCH? This TSH value may well fall within the normal range if an age-adjusted reference interval is employed. Does a 25-year-old African American presenting with a TSH of 4.9 mIU/L not have SCH, when the reference intervals specific to his/her race and age are considered? This TSH value may be above the 97.5th percentile when race-adjusted reference ranges are employed.[77] Formulae for adjusting the TSH reference interval for age, ethnicity, and sex within the US population have been devised.[78] However, not all experts believe that there is a need for specific reference intervals for these subpopulations.

Even when there is agreement about the existence of SCH, benefits of treating generally have not been shown, except possibly in specific age groups.[79] On the other hand, there are some preliminary data that even serum TSH values that fall within the upper end of the TSH reference interval are associated with adverse outcomes.[43] A metaanalysis studied the risk of developing adverse outcomes in individuals with serum TSH values in the upper end of the normal reference range compared with the lower end of the range. Cardiovascular, metabolic, and skeletal outcomes were examined. The odds ratio for developing adverse cardiovascular and metabolic outcomes was significant at 1.21 (95% CI: 1.15–1.27) and 1.37 (95% CI: 1.27–1.48), respectively, whereas the odds ratio for adverse skeletal outcomes was significantly reduced at 0.55 (95% CI: 0.41–0.72).[43] Should this be confirmed, this might suggest that trials of treatment of SCH should enroll patients with even milder degrees of TSH elevation to better determine the risks and benefits of such treatment. A recent clinical practice guideline suggested that treatment of SCH should not be considered unless the TSH is above 20 mIU/L, except in certain subgroups.[80] This conclusion was reached despite the fact that only two of the trials in their associated metanalysis included patients with TSH values above 10 mIU/L. An alternative approach is to carefully consider the clinical context when deciding whether treatment of SCH is of benefit.[81]

Two recent studies have addressed the effect of LT4 on hypothyroid symptoms in older patients with SCH. Those over 65 years of age were studied in the TRUST trial,[82] and individuals over 80 years of age were studied in the IEMO80-plus Thyroid Trial.[83] Both of these trials enrolled patients with persistent TSH elevations of over 4.6 mIU/L. In the TRUST trial the mean TSH values at enrollment were 6.41 and 6.38 mIU/L, with declines to 3.63 and 5.48 mIU/L, in the LT4 group and the placebo group, respectively. In the study of those over 80 years (which included some individuals from the IEMO80-plus Thyroid Trial and some from the TRUST trial), the TSH values at enrollment were 6.50 and 6.20 mIU/L, with reduction to 3.69 and 5.49 mIU/L, in the LT4 and placebo groups, respectively. Neither of these trials showed any difference in their primary outcomes of hypothyroid symptoms or tiredness at 1 year. There was no effect of LT4 on secondary outcomes either. One could speculate that these trials would potentially have had different results if the TSH reference interval used was age-adjusted and resulted in a different definition of SCH, with a higher TSH value such as greater than 7 to 10 mIU/L being used for enrollment. Of note, a recent survey of clinicians found that 11% to 29% of those responding to the survey would treat a 80-year-old patient with tiredness and a TSH of 6.5 to 6.8 mIU/L with LT4.[84]

Myxedema Coma. Severe hypothyroidism can culminate in myxedema coma, a life-threatening condition characterized by hypothermia, bradycardia, hypotension, altered mental status, and multisystem organ failure. Risk factors include advanced age, poor access to healthcare, and other underlying major organ system diseases. Patients may present more frequently in the winter months, and most patients have severe and longstanding thyroid hormone deficiency. Treatment should include T4 (1.6–1.8 mcg/kg/day, with or without a 200- to 500-mcg loading dose). The rationale for the loading dose is to reoccupy the empty binding sites on thyroxine-binding globulin, so that administered LT4 can contribute to raising FT4 levels. Some experts advocate coadministration of LT3 in divided doses to compensate for impaired conversion of T4 to T3. No controlled trials have been performed to evaluate the relative benefits and risks of these different approaches. Glucocorticoids should be administered in stress doses until the results of a cosyntropin stimulation test performed to check for evidence of concomitant adrenal insufficiency are available. Care should be taken to avoid exposure to potent sedative or analgesic agents that may exacerbate altered mental status. Hypothermia should be treated with external warming to reduce the risk of circulatory collapse.

KEY POINTS: Levothyroxine Therapy

- Hypothyroidism generally requires life-long therapy with thyroid hormone replacement.
- There are well established protocols for initiation of levothyroxine therapy
- Levothyroxine therapy must be continually monitored as dose requirements may change with changes in a patient's physiological and medical situation.
- Both overtreatment and undertreatment of hypothyroidism can have adverse effects and monitoring is necessary to maintain euthyroidism.
- Specific treatment goals may need to be considered for specific patient populations.
- The majority of patients are believed to feel well while taking levothyroxine.
- A subset of patients do not feel well while taking levothyroxine, despite the best attempts to optimize therapy.

Therapy Other Than Levothyroxine

The standard of care therapy for hypothyroidism is LT4. The majority of patients, when they are treated to achieve a normal TSH, feel well while taking LT4. However, there are a subset of patients who do not feel fully restored to health and report reduced quality of life despite being biochemically euthyroid. This has led to studies to try and understand the underpinning of this reduced quality of life and also to investigate whether alternative therapies for hypothyroidism might yield greater levels of satisfaction.

Residual Symptoms While Taking Levothyroxine and Potential Underpinnings. Studies that have assessed well-being using the thyroid symptom questionnaire identified dissatisfaction in both the patients treated for hypothyroidism and also in age- and sex-matched patients without hypothyroidism. Approximately 14% more patients with hypothyroidism were dissatisfied compared with the control group.[85] The hypothyroid group did have more coexistent medical conditions, but the difference between the groups remained significant even after adjusting for these. Another study, in which subjects were recruited by letter and then subsequently took part in cognitive testing and completed questionnaires assessing well-being, also showed lower scores in both these areas in the euthyroid patients taking LT4 compared with standard reference values.[86]

Several factors have been proposed as contributing to the decreased quality of life in patients who are apparently euthyroid. These factors include failure to optimize serum TSH, coexistent medical conditions, low serum T3 levels during therapy, the presence of deiodinase or thyroid hormone transporter polymorphisms, and the presence of TPO Abs.[87] With respect to fine-tuning the TSH value in treated patients, most studies do not show improvement in patients' symptoms with lower TSH values. For example, a recent trial in which patients were randomly assigned to receive LT4 doses that resulted in three different ranges of TSH values showed that these different TSH targets had no effect on quality of life, mood, or cognition.[88] Patients with hypothyroidism do appear to have more chronic medical conditions than other populations,[89] but whether this accounts for their decreased quality of life is not clear.

LT4 therapy is associated with steady serum levels of T3 but is also associated with a higher T4/T3 ratio than is seen in endogenous euthyroidism. Many studies have examined the T3 levels achieved during LT4 monotherapy. These studies have had different methodology, including with respect to the group used as a comparator. In a study using patients as their own controls, T3 levels were better maintained.[53] However, most of these studies show that T3 levels are lower in a proportion of LT4-treated patients[7] and may even be below the T3 reference range in some studies.[90] In a small study of thyroidectomized patients, lower T3 levels were associated with being either homozygous or heterozygous for the Thr92Ala polymorphism of the type 2 deiodinase.[91] There are some limited data about the effect of type 2 deiodinase polymorphism status on either response to LT4 therapy or preference for combination therapy. One study performed a retrospective analysis of data from a study of combination therapy by Saravanan et al.[92] Participants underwent genotyping for the presence of the mutation encoding the Thr92Ala polymorphism and their scores on the General Health Questionnaire, both while taking LT4 and while taking combined LT4/LT3, were analyzed according to their polymorphism status. Ala/Ala homozygotes had worse General Health Questionnaire scores while taking LT4.[93] Moreover, they also had a better response to combination therapy based on their General Health Questionnaire scores.[93] Different TSH values could not be implicated, as TSH values did not differ between patients with the same genotype receiving combination therapy or monotherapy with LT4. However, other studies have not shown any association between scores of quality of life and cognition in LT4-treated patients according to their Thr92Ala status.[94,95] One small study found that patients with both the Thr92Ala polymorphism and a polymorphism in one of the thyroid hormone transporters preferred combination therapy with both LT4 and LT3.[96] There are data suggesting that the presence of TPO Abs, even in patients who have maintained endogenous euthyroidism, may be associated with more symptoms than are reported by euthyroid patients without thyroid autoimmunity. One study found more symptoms and lower quality of life in those with elevated titers of thyroid peroxidase antibodies, whereas another study reported more depression.

Combination Therapy With Levothyroxine and Liothyronine

Synthetic Combination Therapy. A relative T3 deficiency in those being treated with traditional LT4 monotherapy in the setting of some patient dissatisfaction with this therapy may prompt consideration of combination therapy with both LT4 and liothyronine (LT3). If the intention were to replicate the molar ratio of T4:T3 achieved by endogenous thyroid functioning in humans, the prescribed ratio would be approximately 14:1 to 15:1.[7] Thus, this would be best achieved using synthetic LT3, with adjustment of the LT3 dose by the prescribing physician. However, it is not well-understood what parameters should be used to make adjustments, given the short half-life of LT3.

Fourteen randomized trials examining the issue of combination therapy for patients with hypothyroidism have been performed.[7,97] Following an initial study in 1999, 12 additional studies were performed between 2002 and 2010, with a subsequent study in 2016. These studies had variable designs with respect to the outcomes studied, the use of crossover or parallel groups, blinding methodology, and the ratio of T4 to T3 employed. However, autoimmune hypothyroidism was the predominant diagnosis in most studies. With some exceptions (see later), these studies failed to demonstrate a benefit of LT3 combination therapy. Four metaanalyses or reviews that incorporated the trials completed at the time of the analysis also concluded that there is no advantage to combination therapy.[7] However, these metaanalyses were made difficult by the heterogeneity of the trials with respect to causes of hypothyroidism, different dosing regimens, different outcome measures, different duration of treatment, and different TSH and T3 levels achieved in the combination therapy groups. Other design issues included nonvalidated outcome measures, carryover effects in some studies, overtreatment in some studies, and failure to study therapy in men and the elderly. Recent reviews and guidelines have stressed the need for larger, better-designed studies of longer duration.[7,98]

Health-related quality of life or mood was studied in 13 trials, with very heterogeneous results.[7] There was superiority of combination therapy on multiple measures in two trials, one by Bunevicius et al.[99] and one by Nygaard et al.[100] These two trials included patients with thyroid cancer and low TSH values, and used a large single dose of LT3 (20 mcg), respectively. Superiority of combination therapy for a minority of measures was seen in two trials.[92,101] In one of these trials the benefit was seen at 3 months, but was no longer seen at 12 months.[92] Of note, the latter is the only combination therapy study of 12 months' duration, with the next longest trials being 4 months and 6 months in duration. The remaining nine trials did not show any superiority of combination therapy with respect to improving quality of life.

Neurocognitive functioning was studied in ten trials, again with heterogeneous results.[7] Superiority of combination therapy on multiple measures was found in one trial.[99] Superiority of combination therapy for a minority of measures was also found in one trial.[102] The remaining eight trials did not show a superior effect of combination therapy on neurocognitive functioning. An example of a negative trial with respect to both quality of life and cognitive functioning is a 4-month study by Clyde et al.[103] In this study, LT4 alone was compared with combination therapy by substituting 50 mcg of LT4 with 7.5 mcg of LT3. This study demonstrated no beneficial changes in body weight, serum lipid levels, hypothyroid symptoms as measured by a health-related quality of life questionnaire, and standard measures of cognitive performance.

Patient preference has been studied in five blinded crossover design trials and two blinded, parallel design trials of combination therapy.[7] With respect to the crossover design trials, the combination therapy was preferred in four trials, which, when combined, total 128 patients.[99,100,102,104] Another trial of 101 patients did not demonstrate a preference for combination therapy.[105] Of the two parallel design trials, there was a preference for combination therapy in one trial of 130 patients[106] and no preference for combination therapy in another trial of 573 patients.[92] The former trial was characterized by some degree of overtreatment, and preference for combination therapy was associated with weight loss. The latter trial was the largest and longest-duration trial of combination therapy conducted thus far.

Natural Combination Therapy. There is considerable patient interest in "natural" combination therapy in the form of DTE. One small trial of DTE randomized patients to either LT4 or DTE and then switched them to the other therapy after 16 weeks, for a further 16 weeks of therapy.[107] During the DTE treatment arm, patients had significantly higher serum levels of T3 and lower levels of FT4. Multiple different parameters of quality of life were assessed during the trial, and these parameters did not differ between the two groups. However, 49% of patients preferred the DTE, compared with 19% preferring LT4 and 33% having no preference.[107] Preference for DTE was found to be associated with weight loss. A recent online survey of patients with hypothyroidism found that those responding to the survey had a higher level of satisfaction with DTE than those receiving other therapies.[108] Reported satisfaction level on a scale of 1 to 10 was 7 (interquartile range 4, 8) for DTE, compared with 6 (3, 8) for synthetic combination therapy with LT4 and LT3 and 5 (3, 7) for LT4 alone. These data need to be evaluated with the self-selected nature of the respondents being considered.

Monotherapy With Liothyronine. Monotherapy with LT3 has rarely been used, in part because of the need to take LT3 multiple times a day to maintain a reasonably stable TSH. Preference for LT3 alone compared with LT4 alone has not been studied in a controlled fashion.

Reduction of Thyroid Peroxidase Antibodies. Whether TPO Abs themselves affect patient quality of life, including in the setting of a normal TSH, has been investigated. In a recent randomized trial, patients with continued symptoms while receiving LT4, who also had significantly elevated TPO Ab titers of greater than 1000 IU/mL (median 2232 IU/mL), were randomized to thyroidectomy or continued medical management. In those who underwent thyroidectomy, TPO Abs declined to a median of 152 IU/mL. Health-related quality of life and fatigue improved in the thyroidectomized patients and was sustained at 12 to 18 months.[109] There was no sham-thyroidectomized group. An earlier nonrandomized study of thyroidectomy for hypothyroid symptoms did not identify a similar benefit for those without TPO Abs.[110] Other therapies tested as a means of reducing TPO Abs, such as specific diets, have not been sufficiently studied.

KEY POINTS: Therapy Other Than Levothyroxine

- If patients have unresolved symptoms with levothyroxine therapy, potential causes of these symptoms should be fully investigated.
- If no cause of patient dissatisfaction is identified, a trial of personalized combination therapy with addition of liothyronine can be contemplated, and can be continued as long as patient benefit and safety is maintained.
- Trials of combination therapy with levothyroxine and liothyronine have not shown patient benefit, but these trials were largely underpowered, of short duration and used once daily liothyronine dosing.
- Future combination therapy trials that address some of the weaknesses of prior trials, and also utilize a sustained release liothyronine preparation, may show different results.
- A future goal is to prevent the cascade of events that leads to autoimmune destruction of thyroid tissue.

SUMMARY AND AREAS OF FUTURE RESEARCH

Despite the significant advances that have been made in our understanding of thyroid hormone therapy, some of these insights have not been easy to translate into practical applications. Additionally, some of the challenges that continue to face physicians treating patients with hypothyroidism concern basic fundamentals. With respect to the former issue, although we recognize that LT4 does not recapitulate all aspects of normal thyroid physiology, we still do not have a "physiologic" replacement therapy. As an example of the latter issue, a significant number of patients with hypothyroidism are either undertreated or overtreated. This is a complex problem, as out-of-range TSH values may be attributed to multiple factors as diverse as confounding medications, interfering conditions, or intrinsic patient variables. The patient variables may be as basic as adhering to prescribed therapy. Such problems, with multiple etiologies, may require creative and sustained efforts in order for them to be addressed.

These challenges are also combined with exciting opportunities for understanding the nuances of fully reversing the consequences of hypothyroidism. Advances in our understanding of genetic diversity in the function of deiodinases, and possibly thyroid hormone transporters, may ultimately enable us to provide therapy tailored to individual patients. Future prospects include understanding the genetic underpinning of thyroid hormone delivery to tissues, developing more physiologic, sustained release or combination thyroid hormone therapies incorporating T3, and facilitating achievement of adequate thyroid hormone levels in all tissues. Although so far only shown in animal models, another hope for the future is the regeneration of functional thyroid follicles from embryonic or pluripotent stem cells. Such cells, once differentiated into thyroid follicular cells, have been shown to be capable of expressing thyroid-specific genes, responding to TSH stimulation, actively transporting iodine, forming three-dimensional follicles, and expressing thyroglobulin. Should such advances be extended to humans, this would not only provide the opportunity for a better understanding of thyroid physiology and disease but may also pave the way for regenerating the full hormonal profile of a normally functioning thyroid gland and thereby restoring euthyroidism in its entirety.

REFERENCES

1. Slater S. The discovery of thyroid replacement therapy. Part 1: In the beginning. *J R Soc Med*. 2011;104:15–18.
2. Bauduer F, Tankersley KB. Evidence of an ancient (2000 years ago) goiter attributed to iodine deficiency in North America. *Med Hypotheses*. 2018;118:6–8.
3. Mehra MR, Campbell HR. The Mona Lisa decrypted: allure of an imperfect reality. *Mayo Clin Proc*. 2018;93:1325–1327.
4. Slater S. The discovery of thyroid replacement therapy. Part 2: the critical 19th century. *J R Soc Med*. 2011;104:59–63.
5. Slater S. The discovery of thyroid replacement therapy. Part 3: A complete transformation. *J R Soc Med*. 2011;104:100–106.
6. McAninch EA, Bianco AC. The history and future of treatment of hypothyroidism. *Ann Intern Med*. 2016;164:50–56.
7. Jonklaas J, Bianco AC, Bauer AJ, et al. Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association Task Force on thyroid hormone replacement. *Thyroid*. 2014;24:1670–1751.
8. Hoermann R, Midgley JEM, Larisch R, et al. Recent advances in thyroid hormone regulation: toward a new paradigm for optimal diagnosis and treatment. *Front Endocrinol (Lausanne)*. 2017;8:364.
9. van der Spek AH, Fliers E, Boelen A. The classic pathways of thyroid hormone metabolism. *Mol Cell Endocrinol*. 2017;458:29–38.
10. Groeneweg S, van Geest FS, Peeters RP, et al. Thyroid hormone transporters. *Endocr Rev*. 2020;41:bnz008.
11. Mendoza A, Hollenberg AN. New insights into thyroid hormone action. *Pharmacol Ther*. 2017;173:135–145.
12. Astapova I. Role of co-regulators in metabolic and transcriptional actions of thyroid hormone. *J Mol Endocrinol*. 2016;56:73–97.
13. Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*. 2018;14:301–316.
14. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*. 2002;87:489–499.
15. Chaker L, Bianco AC, Jonklaas J, et al. Hypothyroidism. *Lancet*. 2017;390:1550–1562.
16. Canaris GJ, Steiner JF, Ridgway EC. Do traditional symptoms of hypothyroidism correlate with biochemical disease? *J Gen Intern Med*. 1997;12:544–550.
17. Ott J, Promberger R, Kober F, et al. Hashimoto's thyroiditis affects symptom load and quality of life unrelated to hypothyroidism: a prospective case-control study in women undergoing thyroidectomy for benign goiter. *Thyroid*. 2011;21:161–167.
18. Kim J. Images in clinical medicine. Myxedema. *N Engl J Med*. 2015;372:764.
19. Taguchi T. Brittle nails and hair loss in hypothyroidism. *N Engl J Med*. 2018;379:1363.

20. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev*. 2014;13:391–397.
21. Bhattacharya S, Goyal A, Kaur P, et al. Anticancer drug-induced thyroid dysfunction. *Eur Endocrinol*. 2020;16:32–39.
22. Li Z, Qiu Y, Fei Y, et al. Prevalence of and risk factors for hypothyroidism after hemithyroidectomy: a systematic review and meta-analysis. *Endocrine*. 2020;70:243–255.
23. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*. 2015;3:286–295.
24. Duntas LH. The catalytic role of iodine excess in loss of homeostasis in autoimmune thyroiditis. *Curr Opin Endocrinol Diabetes Obes*. 2018;25:347–352.
25. Leung AM, Braverman LE. Iodine-induced thyroid dysfunction. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:414–419.
26. Lee SY, Rhee CM, Leung AM, et al. A review: Radiographic iodinated contrast media-induced thyroid dysfunction. *J Clin Endocrinol Metab*. 2015;100:376–383.
27. Calil-Silveira J, Serrano-Nascimento C, Kopp PA, et al. Iodide excess regulates its own efflux: a possible involvement of pendrin. *Am J Physiol Cell Physiol*. 2016;310:C576–C582.
28. Wassner AJ. Unraveling the genetics of congenital hypothyroidism: challenges and opportunities. *J Clin Endocrinol Metab*. 2020;105:dgaa454.
29. Persani L, Cangiano B, Bonomi M. The diagnosis and management of central hypothyroidism in 2018. *Endocr Connect*. 2019;8:R44–R54.
30. Rurale G, Di Cicco E, Dentice M, et al. Thyroid hormone hyposensitivity: from genotype to phenotype and back. *Front Endocrinol (Lausanne)*. 2019;10:912.
31. Luongo C, Trivisano L, Alfano F, et al. Type 3 deiodinase and consumptive hypothyroidism: a common mechanism for a rare disease. *Front Endocrinol (Lausanne)*. 2013;4:115.
32. Baloch Z, Carayon P, Conte-Devolx B, et al. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid*. 2003;13:3–126.
33. Jonklaas J, Razvi S. Reference intervals in the diagnosis of thyroid dysfunction: treating patients not numbers. *Lancet Diabetes Endocrinol*. 2019;7:473–483.
34. Estrada JM, Soldin D, Buckey TM, et al. Thyrotropin isoforms: implications for thyrotropin analysis and clinical practice. *Thyroid*. 2014;24:411–423.
35. Favresse J, Burlacu MC, Maiter D, et al. Interferences with thyroid function immunoassays: clinical implications and detection algorithm. *Endocr Rev*. 2018;39:830–850.
36. Over R, Mannan S, Nsouli-Maktabi H, et al. Age and the thyrotropin response to hypothyroxinemia. *J Clin Endocrinol Metab*. 2010;95: 3675–3683.
37. van Deventer HE, Soldin SJ. The expanding role of tandem mass spectrometry in optimizing diagnosis and treatment of thyroid disease. *Adv Clin Chem*. 2013;61:127–152.
38. Fitzgerald SP, Bean NG, Falhammar H, et al. Clinical parameters are more likely to be associated with thyroid hormone levels than with thyrotropin levels: a systematic review and meta-analysis. *Thyroid*. 2020;30:1695–1709.
39. Beck-Peccoz P, Rodari G, Giavoli C, et al. Central hypothyroidism - a neglected thyroid disorder. *Nat Rev Endocrinol*. 2017;13:588–598.
40. Barth JH, Spencer JD, Goodall SR, et al. Reference intervals for thyroid hormones on Advia Centaur derived from three reference populations and a review of the literature. *Ann Clin Biochem*. 2016;53:385–389.
41. Strich D, Karavani G, Levin S, et al. Normal limits for serum thyrotropin vary greatly depending on method. *Clin Endocrinol (Oxf)*. 2016;85:110–115.
42. LeFevre ML, US Preventative Task Force. Screening for thyroid dysfunction: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2015;162:641–650.
43. Taylor PN, Razvi S, Pearce SH, et al. Clinical review: A review of the clinical consequences of variation in thyroid function within the reference range. *J Clin Endocrinol Metab*. 2013;98:3562–3571.
44. Thayakaran R, Adderley NJ, Sainsbury C, et al. Thyroid replacement therapy, thyroid stimulating hormone concentrations, and long term health outcomes in patients with hypothyroidism: longitudinal study. *BMJ*. 2019;366:l4892.
45. Cappola AR, Cooper DS. Screening and treating subclinical thyroid disease: getting past the impasse. *Ann Intern Med*. 2015;162:664–665.
46. Pollock MA, Sturrock A, Marshall K, et al. Thyroxine treatment in patients with symptoms of hypothyroidism but thyroid function tests within the reference range: randomised double blind placebo controlled crossover trial. *BMJ*. 2001;323:891–895.
47. Roos A, Linn-Rasker SP, van Domburg RT, et al. The starting dose of levothyroxine in primary hypothyroidism treatment: a prospective, randomized, double-blind trial. *Arch Intern Med*. 2005;165(15):1714–1720.
48. Kabadi UM. Optimal daily levothyroxine dose in primary hypothyroidism. Its relation to pretreatment thyroid hormone indexes. *Arch Intern Med*. 1989;149:2209–2212.
49. Kabadi UM, Jackson T. Serum thyrotropin in primary hypothyroidism. A possible predictor of optimal daily levothyroxine dose in primary hypothyroidism. *Arch Intern Med*. 1995;155:1046–1048.
50. Mistry D, Atkin S, Atkinson H, et al. Predicting thyroxine requirements following total thyroidectomy. *Clin Endocrinol (Oxf)*. 2011;74:384–387.
51. Santini F, Pinchera A, Marsili A, et al. Lean body mass is a major determinant of levothyroxine dosage in the treatment of thyroid diseases. *J Clin Endocrinol Metab*. 2005;90:124–127.
52. Younis IR, Ahmed MA, Burman KD, et al. Stable isotope pharmacokinetic studies provide insight into effects of age, sex, and weight on levothyroxine metabolism. *Thyroid*. 2018;28:41–49.
53. Jonklaas J, Davidson B, Bhagat S, et al. Triiodothyronine levels in athyreotic individuals during levothyroxine therapy. *JAMA*. 2008;299:769–777.
54. Lee SY, Braverman LE, Pearce EN. Changes in body weight after treatment of primary hypothyroidism with levothyroxine. *Endocr Pract*. 2014;20:1122–1128.
55. Karmisholt J, Andersen S, Laurberg P. Weight loss after therapy of hypothyroidism is mainly caused by excretion of excess body water associated with myxoedema. *J Clin Endocrinol Metab*. 2011;96:E99–E103.
56. Winther KH, Cramon P, Watt T, et al. Disease-specific as well as generic quality of life is widely impacted in autoimmune hypothyroidism and improves during the first six months of levothyroxine therapy. *PLoS One*. 2016;11:e0156925.
57. Somwaru LL, Arnold AM, Joshi N, et al. High frequency of and factors associated with thyroid hormone over-replacement and under-replacement in men and women aged 65 and over. *J Clin Endocrinol Metab*. 2009;94:1342–1345.
58. Canaris GJ, Manowitz NR, Mayor G, et al. The Colorado thyroid disease prevalence study. *Arch Intern Med*. 2000;160:526–534.
59. Ernst FR, Barr P, Elmor R, et al. The economic impact of levothyroxine dose adjustments: the CONTROL HE study. *Clin Drug Investig*. 2017;37:71–83.
60. McMillan M, Sandulli W, Engelken D, et al. Levothyroxine dose changes and hypothyroid patient satisfaction-results of the CONTROL TS study. *Annals Thyroid Res*. 2017;3:109–114.
61. Bach-Huynh T, Nayak B, Loh J, et al. Timing of levothyroxine administration affects serum thyrotropin concentration. *J Clin Endocrinol Metab*. 2009;94:3905–3912.
62. Skelin M, Lucijanić T, Amidžić Klarić D, et al. Factors affecting gastrointestinal absorption of levothyroxine: a review. *Clin Ther*. 2017;39:378–403.
63. Irving SA, Vadiveloo T, Leese GP. Drugs that interact with levothyroxine: an observational study from the Thyroid Epidemiology, Audit and Research Study (TEARS). *Clin Endocrinol (Oxf)*. 2015;82:136–141.
64. Burch HB. Drug effects on the thyroid. *N Engl J Med*. 2019;381:749–761.
65. Dickerson RN, Maish GO, Minard G, et al. Clinical relevancy of the levothyroxine-continuous enteral nutrition interaction. *Nutr Clin Pract*. 2010;25:646–652.
66. Virili C, Stramazzo I, Santaguida MG, et al. Ulcerative colitis as a novel cause of increased need for levothyroxine. *Front Endocrinol (Lausanne)*. 2019;10:233.
67. Gadiraju S, Lee CJ, Cooper DS. Levothyroxine dosing following bariatric surgery. *Obes Surg*. 2016;26:2538–2542.
68. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.

69. Torino F, Barnabei A, Paragliola R, et al. Thyroid dysfunction as an unintended side effect of anticancer drugs. *Thyroid*. 2013;23:1345–1366.
70. Vita R, Saraceno G, Trimarchi F, et al. Switching levothyroxine from the tablet to the oral solution formulation corrects the impaired absorption of levothyroxine induced by proton-pump inhibitors. *J Clin Endocrinol Metab*. 2014;99:4481–4486.
71. Pirola I, Daffini L, Gandossi E, et al. Comparison between liquid and tablet levothyroxine formulations in patients treated through enteral feeding tube. *J Endocrinol Invest*. 2014;37:583–587.
72. Pirola I, Formenti AM, Gandossi E, et al. Oral liquid L-thyroxine (L-t4) may be better absorbed compared to L-T4 tablets following bariatric surgery. *Obes Surg*. 2013;23:1493–1496.
73. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27:315–389.
74. Benvenga S, Vita R, Di Bari F, et al. Do not forget nephrotic syndrome as a cause of increased requirement of levothyroxine replacement therapy. *Eur Thyroid J*. 2015;4:138–142.
75. Barrett NA, Jones A, Whiteley C, et al. Management of long-term hypothyroidism: a potential marker of quality of medicines reconciliation in the intensive care unit. *Int J Pharm Pract*. 2012;20:303–306.
76. Baumgartner C, Blum MR, Rodondi N. Subclinical hypothyroidism: summary of evidence in 2014. *Swiss Med Wkly*. 2014;144:w14058.
77. Boucai L, Surks MI. Reference limits of serum TSH and free T4 are significantly influenced by race and age in an urban outpatient medical practice. *Clin Endocrinol (Oxf)*. 2009;70:788–793.
78. Boucai L, Hollowell JG, Surks MI. An approach for development of age-, gender-, and ethnicity-specific thyrotropin reference limits. *Thyroid*. 2011;21:5–11.
79. Razvi S, Weaver JU, Butler TJ, et al. Levothyroxine treatment of subclinical hypothyroidism, fatal and nonfatal cardiovascular events, and mortality. *Arch Intern Med*. 2012;172:811–817.
80. Bekkering GE, Agoritsas T, Lytvyn L, et al. Thyroid hormones treatment for subclinical hypothyroidism: a clinical practice guideline. *BMJ*. 2019;365:l2006.
81. Sawka AM, Cappola AR, Peeters RP, et al. Patient context and thyrotropin levels are important when considering treatment of subclinical hypothyroidism. *Thyroid*. 2019;29:1359–1363.
82. Stott DJ, Rodondi N, Bauer DC, et al. Thyroid hormone therapy for older adults with subclinical hypothyroidism. *N Engl J Med*. 2017;377:e20.
83. Mooijaart SP, Du Puy RS, Stott DJ, et al. Association between levothyroxine treatment and thyroid-related symptoms among adults aged 80 years and older with subclinical hypothyroidism. *JAMA*. 2019;322:1977–1986.
84. Razvi S, Arnott B, Teare D, et al. Multinational survey of treatment practices of clinicians managing subclinical hypothyroidism in older people in 2019. *Eur Thyroid J*. 2021;10:330–338.
85. Saravanan P, Chau WF, Roberts N, et al. Psychological well-being in patients on 'adequate' doses of l-thyroxine: results of a large, controlled community-based questionnaire study. *Clin Endocrinol (Oxf)*. 2002;57:577–585.
86. Wekking EM, Appelhof BC, Fliers E, et al. Cognitive functioning and well-being in euthyroid patients on thyroxine replacement therapy for primary hypothyroidism. *Eur J Endocrinol*. 2005;153:747–753.
87. Jonklaas J. Persistent hypothyroid symptoms in a patient with a normal thyroid stimulating hormone level. *Curr Opin Endocrinol Diabetes Obes*. 2017;24:356–363.
88. Samuels MH, Kolobova I, Niederhausen M, et al. Effects of altering levothyroxine (L-T4) doses on quality of life, mood, and cognition in L-T4 treated subjects. *J Clin Endocrinol Metab*. 2018;103:1997–2008.
89. Peterson SJ, McAninch EA, Bianco AC. Is a normal TSH synonymous with "euthyroidism" in levothyroxine monotherapy? *J Clin Endocrinol Metab*. 2016;101:4964–4973.
90. Gullo D, Latina A, Frasca F, et al. Levothyroxine monotherapy cannot guarantee euthyroidism in all athyreotic patients. *PLoS One*. 2011;6:e22552.
91. Castagna MG, Dentice M, Cantara S, et al. DIO2 Thr92Ala reduces deiodinase-2 activity and serum-T3 levels in thyroid-deficient patients. *J Clin Endocrinol Metab*. 2017;102:1623–1630.
92. Saravanan P, Simmons DJ, Greenwood R, et al. Partial substitution of thyroxine (T4) with tri-iodothyronine in patients on T4 replacement therapy: results of a large community-based randomized controlled trial. *J Clin Endocrinol Metab*. 2005;90:805–812.
93. Panicker V, Saravanan P, Vaidya B, et al. Common variation in the DIO2 gene predicts baseline psychological well-being and response to combination thyroxine plus triiodothyronine therapy in hypothyroid patients. *J Clin Endocrinol Metab*. 2009;94:1623–1629.
94. Wouters HJ, van Loon HC, van der Klauw MM, et al. No effect of the Thr92Ala polymorphism of deiodinase-2 on thyroid hormone parameters, health-related quality of life, and cognitive functioning in a large population-based cohort study. *Thyroid*. 2017;27:147–155.
95. Young Cho Y, Jeong Kim H, Won Jang H, et al. The relationship of 19 functional polymorphisms in iodothyronine deiodinase and psychological well-being in hypothyroid patients. *Endocrine*. 2017;57:115–124.
96. Carlé A, Faber J, Steffensen R, et al. Hypothyroid patients encoding combined MCT10 and DIO2 gene polymorphisms may prefer L-T3 + L-T4 combination treatment - data using a blind, randomized, clinical study. *Eur Thyroid J*. 2017;6:143–151.
97. DiStefano J, Jonklaas J. Predicting optimal combination LT4 + LT3 therapy for hypothyroidism based on residual thyroid function. *Front Endocrinol (Lausanne)*. 2019;10:746.
98. Jonklaas J, Cappola AR, Celi FS. Editorial: combination therapy for hypothyroidism: the journey from bench to bedside. *Front Endocrinol (Lausanne)*. 2020;11:422.
99. Bunevicius R, Kazanavicius G, Zalinkevicius R, et al. Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *N Engl J Med*. 1999;340:424–429.
100. Nygaard B, Jensen EW, Kvetny J, et al. Effect of combination therapy with thyroxine (T4) and 3,5,3′-triiodothyronine versus T4 monotherapy in patients with hypothyroidism, a double-blind, randomised cross-over study. *Eur J Endocrinol*. 2009;161:895–902.
101. Valizadeh M, Seyyed-Majidi MR, Hajibeigloo H, et al. Efficacy of combined levothyroxine and liothyronine as compared with levothyroxine monotherapy in primary hypothyroidism: a randomized controlled trial. *Endocr Res*. 2009;34:80–89.
102. Escobar-Morreale HF, Botella-Carretero JI, Gómez-Bueno M, et al. Thyroid hormone replacement therapy in primary hypothyroidism: a randomized trial comparing L-thyroxine plus liothyronine with L-thyroxine alone. *Ann Intern Med*. 2005;142:412–424.
103. Clyde PW, Harari AE, Getka EJ, et al. Combined levothyroxine plus liothyronine compared with levothyroxine alone in primary hypothyroidism: a randomized controlled trial. *JAMA*. 2003;290:2952–2958.
104. Bunevicius R, Jakuboniene N, Jakubonien N, et al. Thyroxine vs thyroxine plus triiodothyronine in treatment of hypothyroidism after thyroidectomy for Graves' disease. *Endocrine*. 2002;18:129–133.
105. Walsh JP, Shiels L, Lim EM, et al. Combined thyroxine/liothyronine treatment does not improve well-being, quality of life, or cognitive function compared to thyroxine alone: a randomized controlled trial in patients with primary hypothyroidism. *J Clin Endocrinol Metab*. 2003;88:4543–4550.
106. Appelhof BC, Fliers E, Wekking EM, et al. Combined therapy with levothyroxine and liothyronine in two ratios, compared with levothyroxine monotherapy in primary hypothyroidism: a double-blind, randomized, controlled clinical trial. *J Clin Endocrinol Metab*. 2005;90:2666–2674.
107. Hoang TD, Olsen CH, Mai VQ, et al. Desiccated thyroid extract compared with levothyroxine in the treatment of hypothyroidism: a randomized, double-blind, crossover study. *J Clin Endocrinol Metab*. 2013;98:1982–1990.
108. Peterson SJ, Cappola AR, Castro MR, et al. An online survey of hypothyroid patients demonstrates prominent dissatisfaction. *Thyroid*. 2018;28:707–721.
109. Guldvog I, Reitsma LC, Johnsen L, et al. Thyroidectomy versus medical management for euthyroid patients with hashimoto disease and persisting symptoms: a randomized trial. *Ann Intern Med*. 2019;170:453–464.
110. Promberger R, Hermann M, Pallikunnel SJ, et al. Quality of life after thyroid surgery in women with benign euthyroid goiter: influencing factors including Hashimoto's thyroiditis. *Am J Surg*. 2014;207:974–979.

75

Thyroid and Pregnancy

Tim I.M. Korevaar and Elizabeth N. Pearce

OUTLINE

Thyroid Disease and Fertility, 1251
  Epidemiology, 1251
  Physiology, 1251
  Physiology During Assisted Reproduction Technology, 1252
  Clinical Outcomes and Management, 1252
Maternal and Fetal Physiology in Pregnancy, 1254
Thyroid Function Tests in Pregnancy, 1254
Hyperthyroidism in Pregnancy, 1254
  Diagnosis, 1255
  Treatment and Monitoring, 1255
  Lactation, 1257
Hypothyroidism and Hypothyroxinemia In Pregnancy, 1257
  Overt Hypothyroidism, 1257
  Subclinical Hypothyroidism, 1257
  Isolated Hypothyroxinemia, 1258
Thyroid Autoimmunity in Pregnancy, 1258
  Associated Obstetric and Fetal Outcomes, 1259
  Potential Mechanisms for Associations, 1259
  Implications for Therapy and Monitoring, 1259
Screening for Thyroid Dysfunction in Pregnancy, 1260
Thyroid Nodules and Cancer During Pregnancy, 1260
Thyroid Disorders in the Postpartum Period, 1261
  Postpartum Thyroiditis, 1261
  Considerations in Lactating Women, 1262
Summary and Future Directions, 1262

THYROID DISEASE AND FERTILITY

Adequate thyroid hormone availability is important for female reproduction. Thyroid hormone regulates key processes in female reproduction, both indirectly and directly. Various studies have shown that thyroid disease, including subclinical thyroid dysfunction and thyroid autoimmunity, is associated with infertility and suboptimal fertility treatment outcomes. However, the majority of these studies have been limited by their retrospective nature, small sample size, or selected population (typically women undergoing assisted reproduction). The overrepresentation of studies in subfertile women necessitates generalization and extrapolation of results for decision-making in fertile women. Furthermore, clinical decision-making strategies are also based on our understanding of thyroid physiology and pathophysiology, as well as efforts aimed at restoring homeostasis. Diagnosis and interpretation of thyroid function test abnormalities should not only focus on the preconception period, but also anticipate effects on gestation and fetal development.

Epidemiology

Thyroid disorders are relatively common among women of reproductive age because of the sex and age distributions of Hashimoto thyroiditis and Graves hyperthyroidism, as well as the increased detection of subclinical disease entities because of preconception screening. Data on the true prevalence of thyroid disease and thyroid autoimmunity in women of reproductive age are scarce because the majority of studies focus only on subclinical disease or thyroid autoimmunity in populations of subfertile or infertile (used synonymously hereafter) women. Furthermore, instead of normal nonpregnancy reference ranges, many studies have used an upper limit for thyroid-stimulating hormone (TSH) of 2.50 mIU/L, which is an incorrect cutoff that leads to considerable overdiagnosis and overtreatment in up to 20% of otherwise healthy women. Of women referred to a fertility center or with a previous miscarriage, almost 5% will have a thyroid function test abnormality. Overt hypothyroidism or overt hyperthyroidism occur only in approximately 0.2% to 0.3%, whereas subclinical hypothyroidism occurs in approximately 2.4% and subclinical hyperthyroidism in approximately 1.3% of women. Furthermore, among women referred to a fertility center and/or with a previous miscarriage, 9.5% to 12.3% are thyroid peroxidase antibody (TPOAb)–positive, which is probably slightly higher than in otherwise healthy pregnant women (mean 7.5%; range 4.7%–15.0%).[1-3] Similar to the general population, risk factors for thyroid disease and autoimmunity in women of reproductive age include ethnicity (e.g., a higher risk of Hashimoto thyroiditis in Whites but a higher risk of Graves hyperthyroidism in Black and Asian women), increased age, the presence of other autoimmune disorders, smoking (Graves hyperthyroidism), and insufficient or excessive iodine intakes.

Physiology

Thyroid hormone can affect female reproduction either directly, through regulation of key cellular processes, or indirectly, through interference with other hormonal axes. Thyroid hormone receptors are abundantly present throughout the human female reproductive tract, including in granulosa cells and oocytes.[4-8] Preclinical studies show that thyroid hormone regulates the stimulatory effects of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) on granulosa cells, follicular growth, and oocyte development.[9,10] These effects are mediated via thyroid hormone regulation of 3,β-hydroxysteroid dehydrogenase (the enzyme that facilitates the final step in progesterone formation), aromatase (the enzyme that facilitates the final step in estradiol formation), and LH/human chorionic gonadotropin (hCG) receptor expression. Indirect effects of thyroid hormone on female reproduction can be extrapolated from clinical observations of menstrual and ovulatory disturbances in both hypo- and hyperthyroidism. Hypothyroidism leads to a blunted FSH and LH response to gonadotropin-releasing hormone (GnRH), hyperprolactinemia, and lower sex hormone–binding globulin (SHBG) and estradiol concentrations, the clinical consequences of which can be reversed by reestablishing euthyroidism.[11] On the other hand, hyperthyroidism is associated with a greater LH and FSH response to GnRH, higher SHBG and estradiol concentrations, and high testosterone and androstenedione concentrations.[11] Although it is likely that similar physiological changes occur in subclinical thyroid disease, studies remain scarce.

Physiology During Assisted Reproduction Technology

Thyroid function test results remain stable throughout a normal menstrual cycle but can change during controlled ovarian hyperstimulation (COH; as employed for assisted reproduction). The rapid increase in estradiol during COH increases thyroxine-binding globulin (TBG) concentrations and type 3 deiodinase (D3) gene transcription.[12] Therefore, an increase in thyroid hormone production is required to ensure adequate thyroid hormone availability (Fig. 75.1). hCG administration may slightly increase free thyroxine (FT4) concentrations (see Fig. 75.1). The increased demand for thyroid hormone during COH is also demonstrated by the changes in TSH and FT4 concentrations in women treated with a fixed dose of levothyroxine. Furthermore, this is supported by data from TPOAb-positive women, who likely have a reduced thyroid functional capacity, as TPOAb-positive women typically exhibit a considerable and sustained increase in TSH following COH (Fig. 75.2).

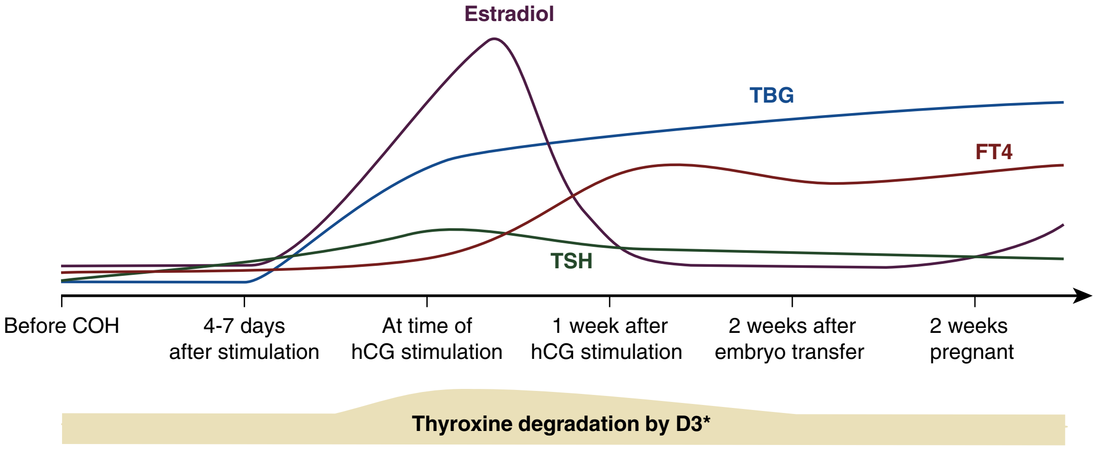
Clinical Outcomes and Management

Data on the association of preconception thyroid function with fertility or pregnancy outcomes in apparently healthy women are sparse, of low quality, and conflicting. For example, a large population-based study[13] performed in China showed that women with a high-normal or increased preconception TSH had a slightly higher risk of abortion and preterm birth, but this was not replicated in a cohort of healthy Australian women.[14] There is a clear indication to treat preconception overt thyroid disease from the view of fertility, conception, and pregnancy. Women with adequately treated preconception hypothyroidism have a similar risk of adverse pregnancy outcomes as euthyroid women. However, suboptimal levothyroxine treatment in women with overt hypothyroidism during early pregnancy is associated with a 1.8- to 4-fold higher risk of miscarriage[15] and slightly higher risks of other adverse pregnancy outcomes such as preterm birth and preeclampsia. Overt hypothyroidism requires thyroid hormone replacement with a preconception or pre-assisted reproduction technology (ART) target TSH of less than 2.5 mIU/L. Women should be counseled to contact their healthcare provider before starting ART or at the time of a positive pregnancy test so that their thyroid function can be checked and the levothyroxine dosage can be increased by roughly 30% as soon as they are pregnant. Any thyroid hormone formulations containing liothyronine should be avoided preconception and during pregnancy, because triiodothyronine (T3) is unlikely to reach the fetal brain, and it can decrease maternal serum thyroxine (T4) concentrations. For women with hyperthyroidism, especially Graves disease, a permanent preconception treatment modality is recommended either before ART or pregnancy (see later).

In euthyroid TPOAb-positive women, clinical follow-up throughout COH and gestation is advised, owing to the high risk of thyroid function deterioration to the point of a treatment indication during pregnancy or during COH (Fig. 75.2). Because of a lack of high-quality data, the decision about whether or not to start levothyroxine treatment for TPOAb-negative subclinical hypothyroidism should be tailored to each specific patient and should take into account data from studies in women undergoing infertility treatment (see later), as well as other risk factors for adverse outcomes such as a prior history of miscarriage or preterm birth, age, body mass index (BMI), and comorbidities (Fig. 75.3).

The majority of studies thus far have focused on the associations of preconception subclinical hypothyroidism and thyroid autoimmunity (mainly TPOAb positivity) in infertile women with early *in vitro* fertilization (IVF) outcomes, miscarriage, and live birth rates. In these heterogeneous study populations, it remains difficult to establish whether subclinical hypothyroidism is a risk factor for adverse outcomes in women undergoing ART.[16] A metaanalysis of three small randomized controlled trials (RCTs) has indicated that levothyroxine treatment in women with subclinical hypothyroidism may increase the number of fertilized oocytes, and reduce the relative risk of miscarriages by 31%.[17]

Data regarding thyroid autoimmunity and the risk of adverse ART outcomes are more consistent. TPOAb positivity is a more specific risk factor for adverse fertility or pregnancy outcomes than thyroglobulin antibody (TgAb) positivity. Although TPOAb positivity is



* Depiction of expected increase in thyroxine degradation by D3 based on estradiol stimulated DIO3 gene transcription

Fig. 75.1 Thyroid physiology throughout controlled ovarian hyperstimulation. *COH,* Controlled ovarian hyperstimulation; *hCG,* human chorionic gonadotropin; *D3,* type 3 deiodinase; *TSH,* thyroid-stimulating hormone; *TBG,* thyroxine-binding globulin; *FT4,* free thyroxine.

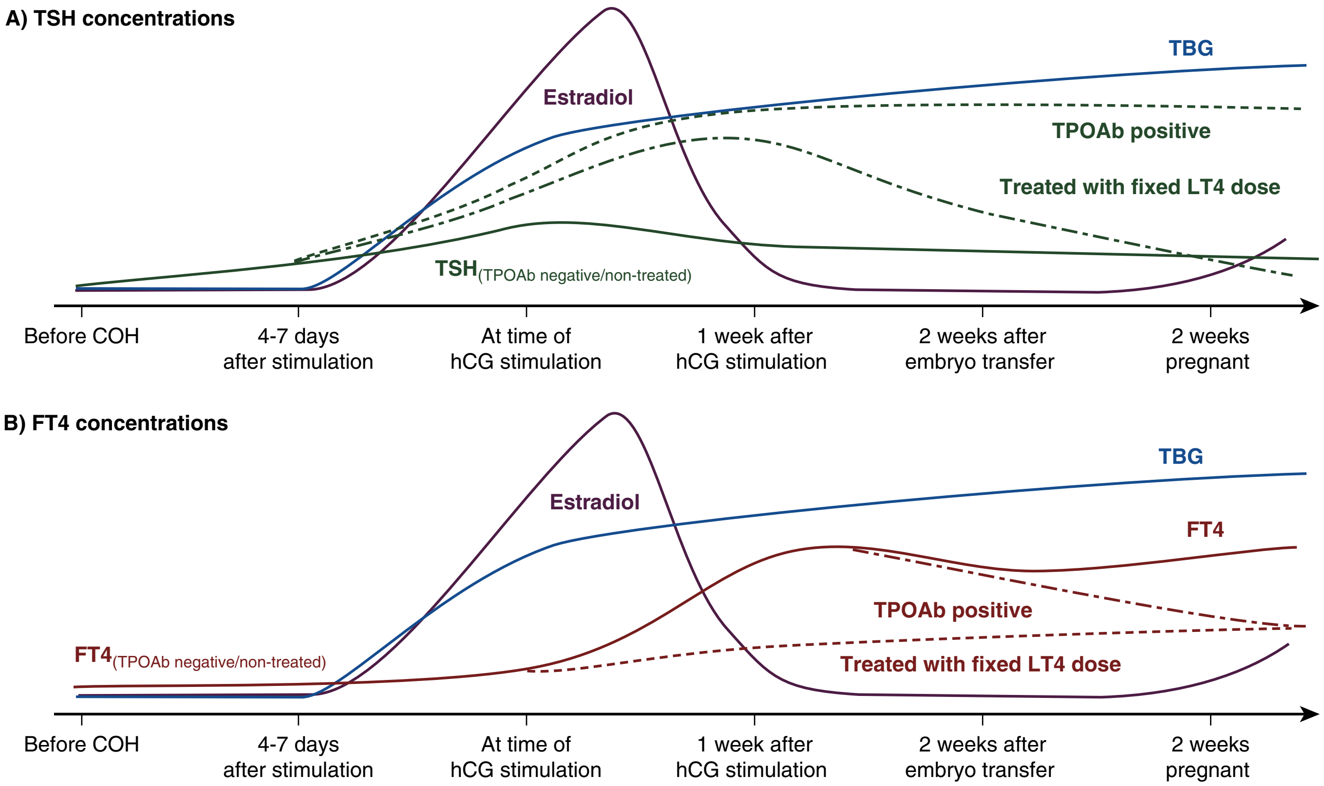
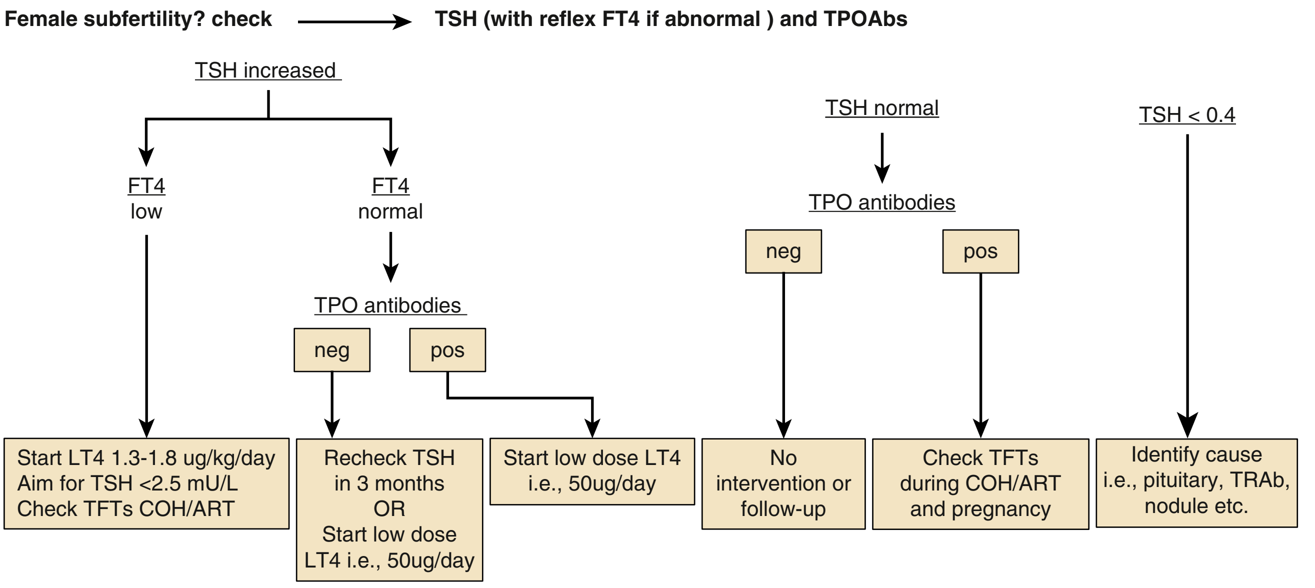


Fig. 75.2 Thyroid function throughout controlled ovarian hyperstimulation (*COH*) in thyroid peroxidase antibody (*TPOAb*)-positive women and women treated with levothyroxine. *hCG,* Human chorionic gonadotropin; *TBG,* thyroxine-binding globulin; *FT4,* free thyroxine; *LT4,* levothyroxine.


Female subfertility? check
TSH (with reflex FT4 if abnormal) and TPOAbs
TSH increased
FT4 low
FT4 normal
TPO antibodies
neg
pos
TSH normal
TPO antibodies
neg
pos
TSH < 0.4
Start LT4 1.3-1.8 ug/kg/day Aim for TSH <2.5 mU/L Check TFTs COH/ART
Recheck TSH in 3 months OR Start low dose LT4 i.e., 50ug/day
Start low dose LT4 i.e., 50ug/day
No intervention or follow-up
Check TFTs during COH/ART and pregnancy
Identify cause i.e., pituitary, TRAb, nodule etc.

[a] Consider if TSH >10 mU/L consider checking macro TSH or heterophilic antibodies, if negative there is an indication for treatment

[b] Consider treatment if for example: TSH >7 mU/L, undergoing ART, or if previous gestational thyroid disease, other risk factors for hypothyroidism related adverse pregnancy outcomes.

Fig. 75.3 Treatment decision flowchart for thyroid function testing in women with subfertility. *TSH,* Thyroid-stimulating hormone; *FT4,* free thyroxine; *TPO,* thyroid peroxidase; *TPOAb,* thyroid peroxidase antibody; *LT4,* levothyroxine; *COH,* controlled ovarian hyperstimulation; *ART,* assisted reproductive technology; *TRAb,* thyroid-stimulating hormone receptor antibody; *TFT,* thyroid function tests.

not associated with early IVF outcomes such as oocyte quality/yield or implantation rate, observational studies show that TPOAb positivity is associated with an approximately 44% higher relative risk of miscarriage and, consequently, with a 35% lower relative chance of live birth.[18] Two randomized trials have indicated that, for euthyroid TPOAb-positive patients with a prior miscarriage or undergoing fertility treatment, levothyroxine treatment has no beneficial effects. One randomized trial compared 226 euthyroid TPOAb-positive women who received levothyroxine (25–50 µg/day, depending on their TSH concentration) to an equally sized control group and found similar rates of miscarriage (10.6% vs. 10.3%) and live births (32.3% vs. 31.7%).[3] Interestingly, subanalyses indicated a beneficial effect of levothyroxine treatment on the risk of miscarriage specifically in cases of a female cause of infertility, as opposed to a male or idiopathic cause of infertility (levothyroxine 7.3% vs. control 16.1%), although the study was not powered to show statistical significance for those analyses.[3] Another randomized trial compared 476 euthyroid TPOAb-positive women who received a fixed TSH dosage of 50 μg/day to an equally sized control group and found similar rates of miscarriage (28.2% vs. 29.6%) and live births (37.4% vs. 37.9%).[19] These studies provide good evidence for a lack of benefit from levothyroxine treatment for euthyroid TPOAb-positive women, even if the TSH is above 2.5 mIU/L. Nonetheless, TPOAb positivity is an indication for follow-up thyroid function tests during COH, pregnancy, and postpartum.

TPOAb positivity with a preconception TSH above the reference range is considered to be an indication to start levothyroxine treatment, particularly when other risk factors for adverse ART or pregnancy outcomes are present, including a prior history of miscarriage or failed ART and age greater than 35 years (Fig. 75.3). The main argument for this is that these women are highly unlikely to normalize their thyroid function during COH or pregnancy and will thus eventually have a treatment indication. If levothyroxine therapy is elected, treatment should start with a low dose (25–50 μg/day), with subsequent dose titration to target a serum TSH of less than 2.5 mIU/L.

KEY POINTS

- Thyroid dysfunction and thyroid autoimmunity are common among women referred for fertility treatment. Overt thyroid dysfunction should be screened for and treated preconception in women with subfertility. Decisions about the treatment of subclinical hypothyroidism in the preconception setting must be individualized.

MATERNAL AND FETAL PHYSIOLOGY IN PREGNANCY

The maternal supply of thyroid hormone to the fetus, an increase in TBG concentrations under the influence of high maternal estrogen, an increase in renal iodine clearance, and the degradation of thyroid hormone by placental D3 together necessitate an increase in thyroid hormone production to ensure adequate availability for mother and fetus.[11] This increased demand for thyroid hormone production is normally met through additional thyroid stimulation by high concentrations of the pregnancy hormone hCG, a weak agonist of the TSH receptor (Fig. 75.4).[20] These pregnancy-specific changes and increased demand may expose preexisting mild thyroid dysfunction or cause new hypothyroidism.

Thyroid hormone is crucial for normal fetal brain development because it regulates the migration, proliferation, and differentiation of fetal neuronal cells, as well as synaptogenesis and myelination. In human fetuses, early neurogenesis begins at approximately 5 weeks postconception, and thyroid hormone receptors have been detected in the fetal brain at as early as 8 weeks' gestation. Because the fetal thyroid gland is not functionally mature before week 18 to 20 of pregnancy, the fetus largely depends on the supply of maternal T4 during the early stages of intrauterine brain development. Later in pregnancy, maternal T4 continues to contribute to fetal thyroid hormone availability (Fig. 75.4). Thyroid autoimmunity, generally reflected by TPOAb positivity, is the most important risk factor for thyroid dysfunction during pregnancy. Recent studies show that the thyroidal response to hCG stimulation is severely impaired in TPOAb-positive women, leading to a relative thyroid hormone shortfall during early pregnancy.[21]

KEY POINTS

- Maternal thyroid hormone synthesis normally increases by 50% in pregnancy. Maternal thyroid hormone is critical for fetal brain development during early gestation, before the onset of fetal thyroid function.

THYROID FUNCTION TESTS IN PREGNANCY

The major changes in thyroid physiology necessitate the use of pregnancy-specific reference ranges to diagnose thyroid disease during gestation. Although many centers previously used fixed upper limits for TSH of 2.5 mIU/L in the first trimester and 3.0 mIU/L in the second or third trimester, it has been shown that these cutoffs are too low and cause considerable overdiagnosis. Ideally, a stepwise approach should be used to define the optimal reference range for each specific institution. Lab-specific, gestational age-specific TSH and FT4 reference ranges are the gold standard. A lab-specific reference range for TSH and FT4 can be calculated as the 2.5th to 97.5th percentiles for a group of at least 400 women who are free of thyroid function-interfering factors (major disease, severe iodine deficiency, TPOAb positivity, thyroid [interfering] drug use, higher physiological hCG [e.g., twin pregnancies and/or IVF treatment]).[22] If it is not possible to obtain lab-specific reference ranges, reference ranges from other centers using the same assay and with a similar patient population can be adopted. Finally, if proper reference ranges cannot be obtained via either of the previous steps, a fixed upper limit for TSH of 4.0 mIU/L may be used.[23]

FT4 is required to distinguish between overt and subclinical thyroid disease. However, it is not possible to identify an evidence-based fixed lower limit for FT4, because published reference ranges differ widely and are highly assay-specific. Therefore, the lower range for FT4 is best defined using a laboratory-specific or "adopted" reference range approach. Commonly used FT4 immunoassays have been reported to be inaccurate because of a shift in thyroid hormone-binding proteins during pregnancy.[24] This shift affects FT4 measurements during the late second and third trimester and is therefore less likely to affect thyroid function tests in the first half of pregnancy. Furthermore, because between-assay correlation for FT4 concentrations is very high,[25-27] commercial assays can be used to adequately identify women with abnormal FT4 if pregnancy-specific and assay-specific reference ranges are used, as shown in metaanalyses.[28,29] Total T4 is a suboptimal marker for thyroid function during pregnancy, as it represents a very crude estimate of the fraction of available thyroid hormone, is highly dependent on changes in TBG, and there is no association with adverse outcomes. For the same reason, T3 concentrations are not used for defining abnormal thyroid function during pregnancy, and the general quality of free T3 assays limits their use both outside of and during pregnancy.

KEY POINTS

- Thyroid function shifts over the course of gestation. Assay- and gestational age-specific reference ranges for TSH and free T4 levels should be used when available.

HYPERTHYROIDISM IN PREGNANCY

In otherwise healthy women, a biochemical diagnosis of overt hyperthyroidism can be made in 0.7% to 0.9%, and another 1.0% to 2.1% of women can be diagnosed with subclinical hyperthyroidism.[30] However, from a clinical perspective, two major subtypes of overt

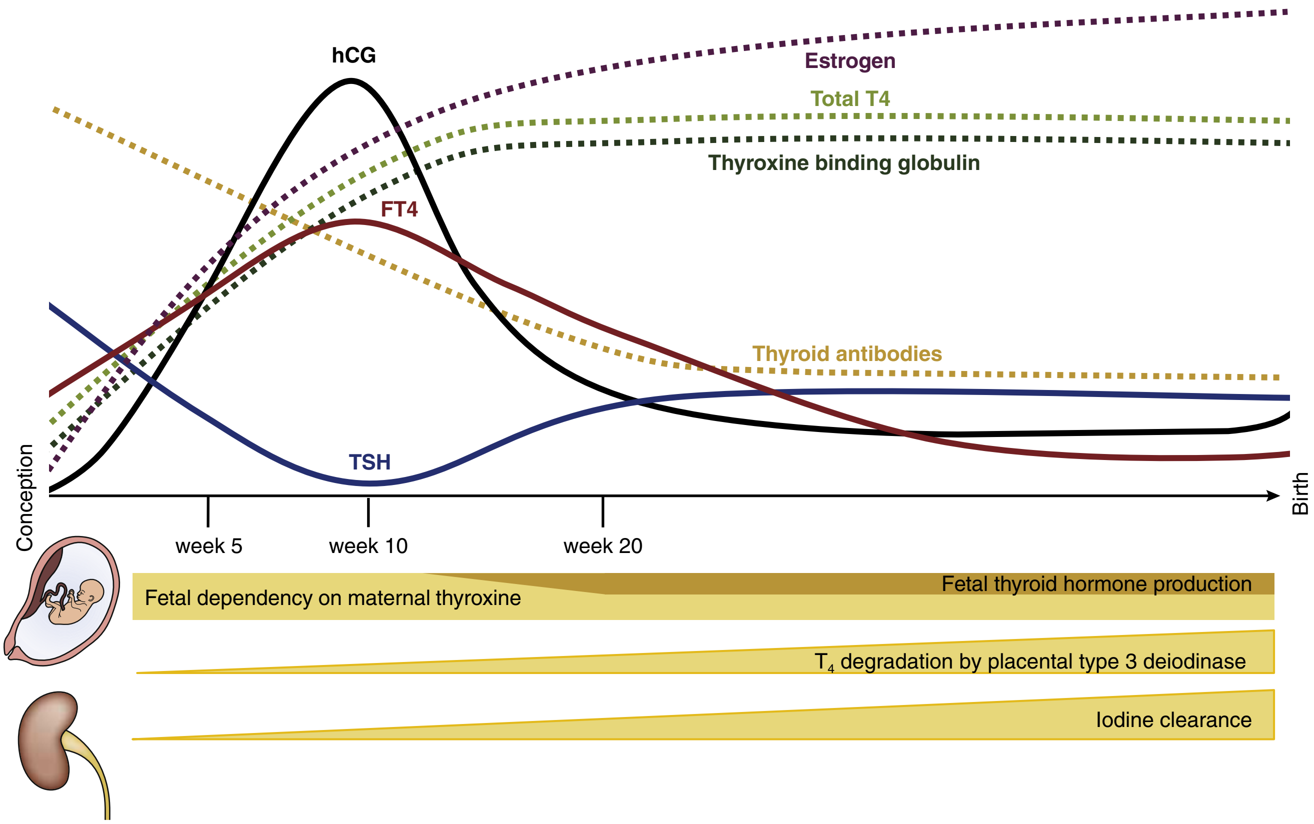


Fig. 75.4 **Thyroid physiology throughout pregnancy.** *TSH,* Thyroid-stimulating hormone; *hCG,* human chorionic gonadotropin; *FT4,* free thyroxine.

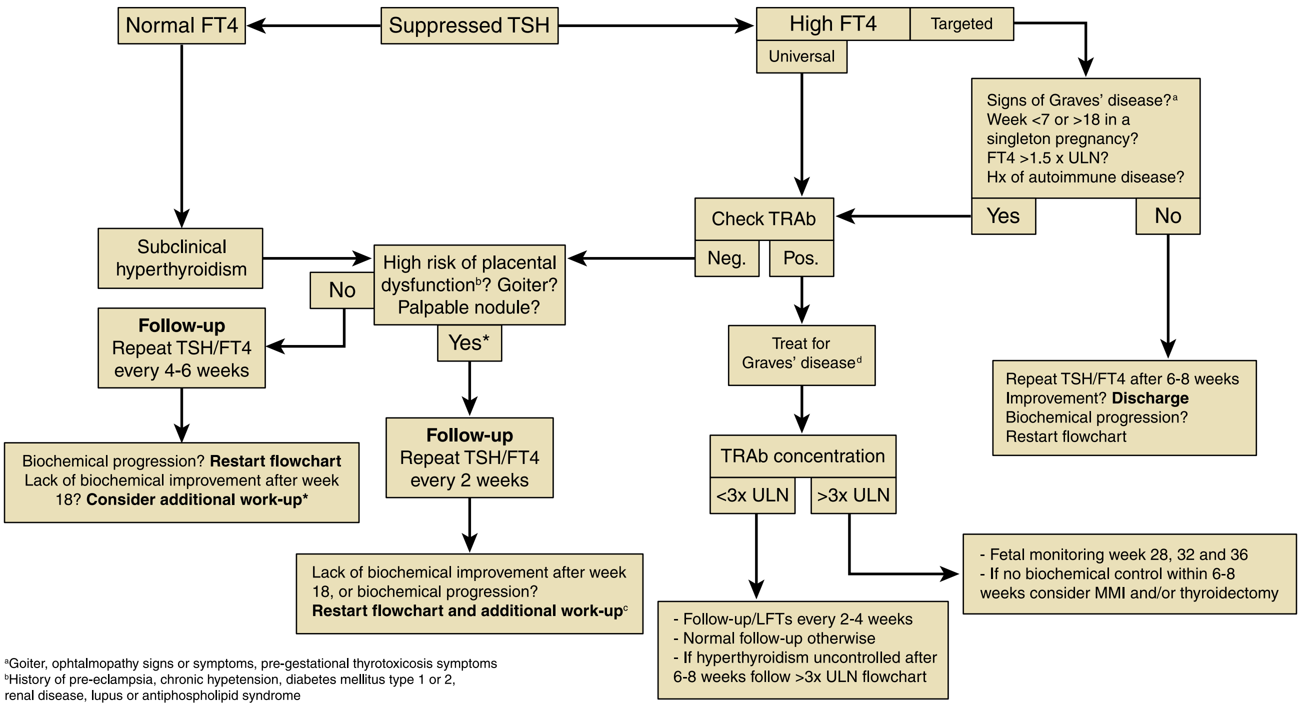
hyperthyroidism should be distinguished. First, the majority of women with gestational hyperthyroidism will present with a physiological form of transient biochemical hyperthyroidism (also referred to as gestational transient thyrotoxicosis). This is caused by high hCG concentrations, which typically peak between 8 to 12 weeks of pregnancy and are higher in IVF and twin pregnancies (Fig. 75.4). Most of these women have no or mild (transient) thyrotoxic symptoms and rarely require symptomatic treatment (i.e., propranolol). Trophoblastic diseases, partial and complete hydatidiform moles, and choriocarcinoma are other rare causes of hCG-mediated (but pathological) hyperthyroidism in pregnancy. Non-hCG-mediated hyperthyroidism, the second major type of overt hyperthyroidism in pregnancy, is caused by Graves disease or, less commonly, autonomous thyroid hormone production (toxic adenoma or multinodular toxic goiter). Non-hCG-mediated hyperthyroidism during pregnancy is rare and can be further categorized as preexisting Graves disease (estimated prevalence 0.5%), new-onset Graves disease (0.05%), and autonomous thyroid hormone production (0.1%).[31,32] When untreated, the vast majority of women with Graves disease will present with overt biochemical abnormalities and thyrotoxic symptoms. Overt hyperthyroidism from Graves disease is associated with a high risk of adverse outcomes such as preeclampsia, preterm birth, low birth weight, and maternal heart failure.[31,33] The clinical presentation of women with hyperthyroidism because of a toxic nodule or toxic multinodular goiter is more heterogeneous. In some cases, the additional hCG stimulation of pregnancy causes only transient hyperthyroidism, while preconception hyperthyroidism because of nodular disease may be more severe and in some cases necessitate treatment with antithyroid drugs, surgery, or radioablation therapy.

Diagnosis

Although some clinical clues, including the personal and family past medical history, preconception onset of symptoms, goiter, or ophthalmopathy, can raise suspicion for Graves hyperthyroidism, biochemical evaluation is the key to diagnosis (Fig. 75.5). Typically, Graves hyperthyroidism in pregnancy presents with a FT4 concentration greater than 1.5 times the upper limit of normal and relatively high T3 concentrations but, most importantly, markers of autoimmunity in the form of positive serum thyroid-stimulating immunoglobulin or TSH receptor antibody (TRAb). Because new-onset Graves disease during pregnancy is very rare, some physicians prefer to perform a targeted approach to TRAb testing, while others prefer a universal testing approach in women with hyperthyroidism during pregnancy. In rare cases of a suspected nonpalpable toxic (T3-producing) adenoma, thyroid ultrasound may be useful, although there is a high baseline risk of thyroid incidentaloma of up to 33%.[34] Thyroid ultrasound cannot be used to distinguish physiological, hCG-mediated hyperthyroidism from Graves hyperthyroidism.

Treatment and Monitoring

Gestational thyrotoxicosis and subclinical hyperthyroidism typically do not require treatment. A conservative approach, for example monitoring thyroid function every 4 to 6 weeks until FT4 concentrations decrease, suffices for the vast majority of cases. Many women with gestational thyrotoxicosis have severe morning sickness or hyperemesis gravidarum and may need supportive care with antiemetics or intravenous hydration. For toxic adenomas and Graves disease, biochemical control of hyperthyroidism is essential to prevent adverse maternal, fetal, and neonatal complications. Maternal


Normal FT4
Suppressed TSH
High FT4
Targeted
Universal
Signs of Graves' disease?ᵃ
Week <7 or >18 in a singleton pregnancy?
FT4 >1.5 x ULN?
Hx of autoimmune disease?
Yes
No
Check TRAb
Neg.
Pos.
Subclinical hyperthyroidism
High risk of placental dysfunctionᵇ? Goiter? Palpable nodule?
No
Yes*
Follow-up
Repeat TSH/FT4 every 4-6 weeks
Treat for Graves' diseaseᵈ
Repeat TSH/FT4 after 6-8 weeks
Improvement? Discharge
Biochemical progression?
Restart flowchart
Follow-up
Repeat TSH/FT4 every 2 weeks
Biochemical progression? Restart flowchart
Lack of biochemical improvement after week 18? Consider additional work-up*
TRAb concentration
<3x ULN
>3x ULN
- Fetal monitoring week 28, 32 and 36
- If no biochemical control within 6-8 weeks consider MMI and/or thyroidectomy
Lack of biochemical improvement after week 18, or biochemical progression?
Restart flowchart and additional work-upᶜ
- Follow-up/LFTs every 2-4 weeks
- Normal follow-up otherwise
- If hyperthyroidism uncontrolled after 6-8 weeks follow >3x ULN flowchart

[a] Goiter, ophtalmopathy signs or symptoms, pre-gestational thyrotoxicosis symptoms
[b] History of pre-eclampsia, chronic hypetension, diabetes mellitus type 1 or 2, renal disease, lupus or antiphospholipid syndrome
[c] For example: T3 measurement, thyroid ultrasound, postpartum diagnostic imaging
in case of a thyroid nodule, consider treatment if either persisting hyperthyroidism at 18 weeks, or FT4 > 1.2X ULN on two occasions before 18 weeks start low-dose PTU
[d] Prefer PTU over MMI titration therapy

Fig. 75.5 Treatment decision flowchart for hyperthyroidism during pregnancy. *TSH,* Thyroid-stimulating hormone; *FT4,* free thyroxine; *TRAb,* thyroid-stimulating hormone receptor antibody; *MMI,* methimazole; *ULN,* upper limit of normal; *TFT,* thyroid function tests; *Hx,* history.

FT4 concentrations should be used to monitor treatment, because TSH can remain suppressed and (F)T3 concentrations do not adequately reflect fetal thyroid hormone availability. The treatment target should be an FT4 concentration in the upper third of the reference range, because the fetal thyroid is more sensitive to antithyroid drugs (ATDs), and lower maternal FT4 concentrations are associated with a higher risk of fetal hypothyroidism (upper third: 10% vs. 36% for lower two-thirds).[43]

When hyperthyroidism is diagnosed preconception, various treatment modalities can be used to establish biochemical control. Of note, there is an important role for prepregnancy counseling and shared decision-making, because toxic adenomas and the vast majority of Graves disease complicating pregnancy will predate conception. One option that can be considered is radioactive iodine, which is contraindicated during pregnancy and lactation. When considered preconception, women should be counseled to avoid conception for 6 months post-radioactive iodine therapy because of risk of teratogenicity from the radiation and the need to establish euthyroidism on levothyroxine. TRAb titers typically increase for approximately 1 year following radioactive iodine administration,[35] and women with high baseline TRAb titers should be counseled that this may increase the risk for fetal and neonatal complications. Another option is thyroidectomy, which results in relatively swift biochemical control and normalization of serum TRAb concentrations but confers a risk of surgical complications and requires lifelong thyroid hormone therapy. Finally, treatment with ATDs is effective in establishing biochemical control and normalization of serum TRAb concentrations but, when used in the first trimester, is associated with an additional 3% to 5% absolute risk increase in fetal birth defects. That is in addition to the well-known, but very rare, side effects of hepatic failure for propylthiouracil (PTU) or agranulocytosis for both ATDs.[36-38] PTU is the preferred drug during organogenesis in the first trimester, because large population studies indicate that the use of PTU during early pregnancy is associated with a slightly lower risk of birth defects (absolute risk increase: 3% vs. 5% for methimazole [MMI]) but also less severe fetal anomalies as compared with MMI.[36,37,38]

To limit the potential adverse outcomes associated with ATDs, several pregnancy-specific ATD strategies can be applied. First, in women of childbearing age with hyperthyroidism, the options for definitive treatment and the patient's preference should be discussed preconception. Second, if the patient prefers ATDs or is already pregnant while on ATDs, so-called block and replace therapy (a high dose of ATD with levothyroxine titration) should be avoided, and therapy should consist of ATD monotherapy with titration to euthyroidism with the lowest possible dose of PTU. Women using MMI preconception who are likely to require continuing therapy during pregnancy (i.e., those with more severe disease) should be switched to PTU preconception. It remains unknown whether switching from MMI to PTU in the first trimester of pregnancy reduces the risk of fetal birth defects. Switching strategies increases the risk of biochemical uncontrolled disease, and the single study that compared women who switched from PTU to MMI in the first trimester with those who continued MMI therapy found no difference in the risk of fetal birth defects.[38] Owing to immune tolerance and the subsequent alterations in maternal immune response during pregnancy, the dose of ATDs should typically be reduced from the second half of pregnancy onwards, guided by maternal FT4 levels, to avoid overtreatment and fetal hypothyroidism.[39-40] Third, in women who are biochemically controlled with a stable low dose of MMI (5–10 mg) or PTU (100–200 mg) before pregnancy, the chance of relapse is lower than the overall 30% to 50% risk, and relapse would most likely occur after a few weeks to months.[41] Therefore, stopping ATD treatment immediately at the time of a positive pregnancy test and monitoring thyroid function every 2 weeks can be considered to prevent the risks related to ATD exposure.[23]

In women with a history of active Graves disease, measurement of TRAbs should be performed in early pregnancy (Fig. 75.5).

If biochemical control is not established before the second half of pregnancy, or if TRAbs are increased more than three times the upper limit of normal, additional maternal and fetal monitoring is required. Additional monitoring to determine the risk of hyperthyroidism (from maternal TRAb) or hypothyroidism (from ATD) in the fetus can start from midpregnancy (i.e., at 28, 32, and 36 weeks) via assessment of the fetal heart rate (normal <160/min), fetal growth, fetal thyroid size, and signs of increased or delayed bone maturation. Maternal thyroid function monitoring at the same time points is also informative, as the maternal thyroid function reflects fetal thyroid hormone availability.[42,43] For rare cases of uncontrollable Graves hyperthyroidism during pregnancy, treatment using high-dose potassium iodide, prednisolone, or thyroidectomy should be considered in the setting of a multidisciplinary team.[44]

Lactation

Both PTU and MMI are safe for use during breastfeeding, and thyroid function testing of infants is not required, as only a small fraction (<0.2%) of the maternal dosage will reach the breast milk.[45,46] The neonate depends on maternal iodine, but not thyroid hormone, transfer via breast milk. Nonetheless, maintaining maternal euthyroidism postdelivery remains important, as hypothyroidism can cause poor lactation. Although case reports have described measurable TRAb concentrations in breast milk, these are unlikely to cause neonatal Graves disease because these will be degraded in the gastrointestinal tract of the neonate.

KEY POINTS

- When hyperthyroidism occurs in early pregnancy, gestational transient thyrotoxicosis (which requires only monitoring and supportive care) must be differentiated from overt Graves hyperthyroidism (which requires treatment). Given the potential teratogenicity of anti-thyroid drugs, preconception counseling is essential for women with Graves disease. Treatment decisions in pregnancy must consider both maternal and fetal effects of therapy.

HYPOTHYROIDISM AND HYPOTHYROXINEMIA IN PREGNANCY

Overt Hypothyroidism

Overt maternal hypothyroidism (increased TSH with low FT4) during pregnancy occurs in roughly 0.5% of women. Thyroid autoimmunity and iodine deficiency (in some parts of the world) are the most important risk factors. Gestational overt hypothyroidism, as well as suboptimally treated preexisting overt hypothyroidism, are associated with a higher risk of adverse pregnancy complications, including miscarriage, premature delivery, low birth weight, miscarriage, and preeclampsia, as well as an up to 7-point lower child intelligence quotient (IQ).[15,47] Although randomized trials are unavailable for ethical reasons, in view of the detrimental effects of untreated overt hypothyroidism and evidence of a normalized risk of pregnancy complications in women with adequately treated hypothyroidism, there is general consensus that overt hypothyroidism during pregnancy should be treated as early as possible. Levothyroxine is the treatment of choice, with a target TSH of less than 2.5 mIU/L. For overt hypothyroidism, the levothyroxine dosage used outside of pregnancy (1.3–1.8 μg/kg/day) should be increased by approximately 30%–50% to meet the increased thyroid hormone demand of pregnancy.[48] To prevent overtreatment, the dose increase should be titrated, and thyroid function should be monitored every 4 to 6 weeks.

KEY POINTS

- Untreated overt hypothyroidism has adverse effects on pregnancy outcomes and fetal brain development. Most women will require increased levothyroxine doses in pregnancy, with close monitoring of serum TSH levels to achieve a goal TSH <2.5 mIU/L.

Subclinical Hypothyroidism

Subclinical hypothyroidism, defined as an elevated serum TSH in the setting of normal peripheral thyroid hormone levels, is relatively common in pregnancy. Prevalence estimates have varied based on the use of differing TSH thresholds for diagnosis, differences in the timing of thyroid function assessment across pregnancy cohorts, and regional variation in iodine intakes. A recent metaanalysis using the 97.5th percentile as the TSH upper limit estimated that 3.5% of pregnant women are subclinically hypothyroid.[30] In iodine-sufficient regions, subclinical hypothyroidism is most frequently caused by autoimmunity. Maternal subclinical hypothyroidism has been linked to preterm birth, placental abruption, pregnancy loss, small-for-gestational-age infants, low birth weight, and neonatal death.[2,49,50] Whether or not subclinical hypothyroidism in pregnancy requires treatment with levothyroxine remains controversial.

Multiple prospective observational studies have examined associations between subclinical hypothyroidism and adverse obstetric outcomes. In a systematic review that included 18 cohorts, maternal subclinical hypothyroidism was associated with increased risk for pregnancy loss (relative risk [RR] 2.01, 95% confidence interval [CI]: 1.66–2.44), premature rupture of membranes (RR 1.43, 95% CI: 1.04–1.95), placental abruption (RR 2.14, 95% CI: 1.23–3.70), and neonatal death (RR 2.58, 95% CI: 1.41–4.73) compared with euthyroidism.[49] An individual participant–level metaanalysis that included data from over 48,000 mother-child pairs concluded that maternal subclinical hypothyroidism is associated with increased risk for having a small-for-gestational-age infant (odds ratio [OR] 1.24, 95% CI: 1.04–1.48) and lower mean birth weight compared with euthyroidism.[50] Having both subclinical hypothyroidism and TPOAb positivity seems to confer greater risk for miscarriage and premature delivery than having either risk factor alone. A prospective cohort study in 3315 Chinese women found that women with both subclinical hypothyroidism and thyroid autoimmunity were at higher risk for miscarriage compared with euthyroid, antibody-negative women (OR 9.56, 95% CI: 3.76–24.28) than those with only subclinical hypothyroidism (OR 3.4, 95% CI: 1.62–7.15) or those with only thyroid autoimmunity (OR 2.71, 95% CI: 1.43–5.12).[51] Similarly, an individual participant–level metaanalysis including over 47,000 pregnant women demonstrated that the risk of preterm birth was higher for subclinically hypothyroid women than euthyroid women (OR 1.29, 95% CI: 1.01–1.64) and still higher (OR 1.55, 95% CI: 1.05–2.27) when both TPOAb positivity and TSH were present.[2] Overall, although individual observational cohorts have reported variable results, likely because of heterogeneity in the definition of subclinical hypothyroidism, adequacy of study power, timing of assessment in pregnancy, and whether or not TPOAb-positive women were excluded, a growing body of evidence from large studies demonstrates that maternal subclinical hypothyroidism, particularly in the presence of TPOAb, is associated with worsened obstetric outcomes.

Multiple studies have assessed associations between maternal subclinical hypothyroidism and child developmental outcomes. A metaanalysis that included 11 studies concluded that maternal subclinical hypothyroidism is significantly associated with the risk of intellectual impairment in children (OR 2.14, 95% CI: 1.20–3.83).[52] However, an individual-level metaanalysis that included 9036 mother–child pairs from three prospective population-based birth cohorts demonstrated that, while maternal FT4 levels were associated with child IQ when assessed between 1.5 and 8 years of age, no independent associations were observed between maternal TSH and child IQ.[29] Overall, while low maternal thyroid hormone availability clearly seems to be associated with adverse effects on child neurodevelopment, maternal FT4 levels appear to be a much stronger predictor than maternal TSH.

Only a few interventional trials to date have examined the effects of levothyroxine therapy on obstetric and child developmental outcomes in subclinically hypothyroid women. A trial conducted in Iran randomized 366 women who were TPOAb-negative and had serum TSH values in the range of 2.5 to 10.0 mIU/L with normal FT4 to levothyroxine dosed at 1 μg/kg/d versus no therapy starting at 12 weeks' gestation.[53] The primary outcome, preterm delivery, did not differ between the treated and untreated women in the overall analysis, but there was a significant reduction in preterm delivery (RR 0.39, 95% CI: 0.15–0.98) when analyses were restricted to only the women with baseline TSH values in the 4- to 10-mIU/L range. Two trials have examined the effects of levothyroxine treatment for subclinically hypothyroid pregnant women on child developmental outcomes. The Controlled Antenatal Screening Study (CATS) was conducted in in Italy and the United Kingdom.[54] At 3 years of age, mean IQ did not differ between the children of 390 subclinically hypothyroid or hypothyroxinemic pregnant women treated with 150 μg/day levothyroxine compared with the children of 404 untreated mothers. This trial has been criticized for initiation of therapy at 13 weeks' gestation, which may have missed a critical window for neurodevelopment in the first trimester. In addition, in light of data suggesting possible adverse effects on child IQ of both too little and too much maternal FT4,[58] the levothyroxine dose employed may have been too high. A subsequent multicenter US trial randomized 677 subclinically hypothyroid women to either levothyroxine treatment (starting at 100 μg/day) or placebo initiated at a median of 17 weeks and saw a nonsignificant 3-point increase in child IQ at age 5 years with treatment, and no difference in any obstetric outcomes, including miscarriage, preterm birth, placental abruption, and neonatal death.[55] As for the CATS study, concerns have been raised that treatment initiation in the US trial may have been too late in gestation to see effects on child development.

Given the uncertainty regarding whether levothyroxine therapy for subclinically hypothyroid pregnant women improves obstetric or child developmental outcomes, current treatment recommendations are variable. Guidelines from the American College of Obstetricians and Gynecologists recommend against treatment.[56] By contrast, the European Thyroid Association recommends treatment whenever subclinical hypothyroidism is diagnosed before conception or during gestation.[57] Because of the observed increased obstetric risks in women with both subclinical hypothyroidism and TPOAb positivity, the American Thyroid Association (ATA) recommends a stepwise approach, taking into account both the degree of TSH elevation and the patient's TPOAb status. Levothyroxine is recommended in TPOAb-positive women with subclinical hypothyroidism and in TPOAb-negative women when the serum TSH is greater than 10 mIU/L.[23] For TPOAb-negative women with TSH in the 4- to 10-mIU/L range or for TPOAb-positive women with a high-normal TSH of 2.5 to 4.0 mIU/L, levothyroxine is not routinely recommended, but may be considered. If treatment is not elected in these women, serum TSH should be monitored. Levothyroxine treatment in women with subclinical hypothyroidism should be initiated at a low dose (typically 50 μg daily) to prevent overtreatment. The goal of therapy is to maintain a serum TSH level of less than 2.5 mIU/L.

KEY POINTS

- Although subclinical hypothyroidism in pregnancy has been associated with adverse pregnancy outcomes, randomized clinical trials have not shown a clear benefit of levothyroxine treatment. Guideline recommendations are variable, but close monitoring of thyroid function should be employed if levothyroxine is not elected for subclinically hypothyroid pregnant women.

Isolated Hypothyroxinemia

Hypothyroxinemia, defined as normal serum TSH with a decreased serum FT4, reflects a multifactorial and pregnancy-specific pathophysiology likely involving iodine and iron status, angiogenic factors, and behavioral and anthropometric characteristics. Isolated hypothyroxinemia has been associated with preterm birth,[2] adverse neurobehavioral outcomes in offspring such as language delay, autistic symptoms, schizophrenia, attention deficit hyperactivity disorder, a slower reaction time, suboptimal school performance, and lower gray matter and cortex volume.[58] It remains unknown whether levothyroxine treatment for isolated hypothyroxinemia has beneficial effects. On one hand, there are observational studies that show a higher risk of various adverse outcomes. On the other hand, two randomized trials investigating the effects of treatment on offspring IQ did not show any beneficial effects. However, these randomized trials are not considered conclusive because of suboptimal study design (late treatment start, overtreatment, and a lack of statistical power). As noted earlier, the CATS trial and a subsequent US trial demonstrated no difference in mean IQ between the children of mothers treated for subclinical hypothyroidism or hypothyroxinemia and the children of untreated mothers, although the study methodology has been criticized.[54,55] One argument for the lack of effect in CATS was the relatively high levothyroxine dosage of 150 μg, which may have contributed to the lack of a net beneficial effect, as recent evidence shows that high, as well as low, FT4 concentrations are associated with suboptimal offspring neurodevelopment.[59,60] Adequately powered RCTs are still required to examine the potential effects of levothyroxine treatment in women with isolated hypothyroxinemia on offspring neurodevelopment. In light of current knowledge, it is not possible to advise for or against levothyroxine treatment for women with hypothyroxinemia. Individual risk factors such as a history of pregnancy complications or other comorbidities, as well as the potential risk for overtreatment, should be taken into account during individualized clinical decision-making.

THYROID AUTOIMMUNITY IN PREGNANCY

TPOAbs are present in 5% to 14% of pregnant women, and TgAb in 3% to 18%.[61] Risk factors for thyroid autoimmunity include older age, having a family history of autoimmune disorders, inadequate or excessive dietary iodine intake, BMI 35 kg/m^2 or higher, and European ancestry.[61,62] TPOAb and TgAb titers tend to wane over the course of gestation, likely because of the immune tolerance that develops during pregnancy. The definition of thyroid antibody positivity is typically based on assay manufacturers' recommended threshold values, but this may be problematic. An individual participant-level metaanalysis of three Dutch pregnancy cohorts demonstrated that manufacturer cutoffs for TPOAb positivity failed to identify 29% of women with TPOAb concentrations high enough to affect maternal thyroid function and pregnancy outcomes, suggesting that pregnancy-specific, population-derived cutoff values may be needed.[63] Thyroid antibodies can cross the placenta, but TPOAb and TgAb do not appear to affect fetal thyroid function.

Associated Obstetric and Fetal Outcomes

Maternal TPOAb positivity has been linked to miscarriage risk in multiple studies over the past three decades. A 2011 metaanalysis of 19 cohort studies and 12 case-control studies found a 2- to 3-fold increase in the odds of miscarriage in women with thyroid autoantibodies (OR 3.90, 95% CI: 2.48–6.12 for the cohorts and OR 1.80, 95% CI: 1.25–2.60 for the case-control studies).[64] This finding has been confirmed in a subsequent prospective Chinese cohort, with a similar increased risk for miscarriage seen in euthyroid, thyroid antibody–positive women compared with euthyroid women without antibodies (OR 2.71, 95% CI: 1.43–5.12).[51] The association between maternal autoimmunity and recurrent miscarriage, defined as three or more pregnancy losses, has been somewhat variable across studies, but a recent metaanalysis of 22 studies concluded that TPOAb positivity was associated with increased risk (random effects OR 1.76; 95% CI: 1.00–3.10).[65]

Preterm delivery has also been associated with TPOAb positivity in numerous studies. A 2011 metaanalysis that included five cohorts, together comprising 12,566 women, concluded that the risk for preterm delivery was increased 2-fold in women with thyroid autoimmunity (OR 2.07, 95% CI: 1.17–3.68).[64] A larger 2012 metaanalysis of 11 prospective cohort studies comprising 35,467 women also reported an increase in the risk for premature delivery in euthyroid women with TPOAb compared with antibody-negative women (RR 1.69, 95% CI: 1.19–2.41) but no increase in risk of prematurity for TgAb-positive women (RR 0.88, 95% CI: 0.60–1.29).[66] Most recently, an individual participant-level metaanalysis including 47,045 participants from 19 cohorts found that, compared with TPOAb-negative women, TPOAb-positive women were at increased risk for both preterm birth (OR 1.33, 95% CI: 1.15–1.56) and very preterm birth (OR 2.45, 95% CI: 1.81–3.32).[2] In this analysis, TgAb positivity was not associated with increased preterm delivery risk (OR 0.88, 95% CI: 0.64–1.20).

Few studies have assessed associations of maternal TPOAb and risk for obstetric outcomes other than miscarriage and prematurity. Based on current evidence, there is likely no association of maternal TPOAb status with preeclampsia or gestational hypertension.[62] Two studies to date suggest a possible link between maternal TPOAb positivity and risk for placental abruption, two studies have documented an association with premature rupture of membranes, and one study has found an association between TPOAb status and risk for polyhydramnios, while studies examining associations with gestational diabetes have been inconclusive.[62] Equivocal evidence suggests a possible link between maternal TPOAb status and child neurodevelopmental outcomes, independent of maternal thyroid function. An analysis of 1268 Chinese mother–child pairs found that maternal TPOAb positivity in gestation was associated with below-average child mean intelligence (OR 6.69, 95%: CI 2.34–19.15) and motor scores (OR 8.25, 95% CI: 3.31–20.53) at 25 to 30 months of age.[67] An analysis of associations between maternal TPOAb positivity during gestation and child IQ, assessed at 5 to 10 years of age, in two European pregnancy cohorts found that TPOAb positivity was associated with a 2.0 ± 0.9-point decrement in mean child IQ in one of the cohorts ($P = .03$), but found no association of maternal TPOAb status with child IQ in the other.[68] These results were not altered by adjustment for maternal thyroid function. Additional research will be needed to better understand the effects of TPOAb positivity on obstetric outcomes other than pregnancy loss and on child neurodevelopmental outcomes.

Potential Mechanisms for Associations

The mechanism for the observed associations between thyroid autoimmunity and pregnancy loss is unknown, although several theories have been advanced. It has been argued that the associations reported between thyroid autoimmunity and pregnancy loss in some cohort studies may simply be attributed to confounding by factors such as older maternal age or higher TSH values within the reference range in the antibody-positive women, although independent associations have been seen in some studies that have controlled for these variables. Women with TPOAb positivity do not experience the normal physiologic hCG-mediated rise in FT4 and decline in serum TSH levels that are observed early in pregnancy in women without thyroid autoimmunity.[69] They tend to have higher serum TSH levels throughout gestation than women who are TPOAb-negative and are at increased risk for the development of hypothyroidism as pregnancy progresses.[63,70] It is, therefore, possible that the association between TPOAb positivity and increased risk for pregnancy loss is mediated by mildly decreased thyroid hormone levels. However, it is also plausible that this association has nothing to do with maternal thyroid status, and that TPOAbs are instead simply markers for an autoimmune milieu. Maternal thyroid autoimmunity is known to be associated with impaired cellular and humoral immune responses and with an increased prevalence of anticardiolipin and other non-organ-specific autoantibodies.[61] Finally, TSH receptor–blocking antibodies are found in up to 10% of patients with Hashimoto thyroiditis. Because of cross-reactivity between TSH receptor–blocking antibodies and the receptors for hCG, these blocking antibodies may inhibit the LH/hCG-receptor in the corpus luteum, causing early pregnancy loss via decreased progesterone and estrogen production.[71]

Implications for Therapy and Monitoring

In the absence of a clear understanding of the mechanism for the observed association between maternal TPOAb positivity and pregnancy loss, it is difficult to be certain of the appropriate intervention, if any. Small studies have investigated treatments designed to reduce autoimmunity. Results of three nonrandomized studies investigating the use of intravenous immunoglobulin therapy to reduce rates have been inconclusive, and such treatment is currently not recommended.[23] Two small randomized clinical trials suggested an improvement in clinical pregnancy and live birth rates in TPOAb-positive women who received prednisolone before undergoing ART, but current guidelines recommend against glucocorticoid therapy for autoantibody-positive euthyroid women undergoing ART in light of uncertainty regarding both risks and benefit.[23]

Several trials have now explored whether levothyroxine therapy for euthyroid, TPOAb-positive pregnant women improves pregnancy outcomes. In a 2006 study, Negro and colleagues randomized 57 TPOAb-positive euthyroid women to treatment with levothyroxine starting at 10 weeks' gestation.[72] Compared with 58 women randomized to no treatment, there was a significant decrease in risk for both miscarriage (RR 1.72, 95%: CI 1.13–2.25) and preterm delivery (RR 1.66, 95% CI: 1.18–2.34). Subsequent trials have failed to replicate those results. Nazarpour and colleagues randomized 131 TPOAb-positive pregnant women with serum TSH values 10 mIU/L or less to levothyroxine (using the same dosing protocol as Negro et al.) versus no therapy and found a lower risk for preterm delivery in the treated versus the untreated group (RR 0.30, 95%: CI 0.1–0.85).[73] However, in sensitivity analyses, the benefit in this trial accrued only to those women with a baseline serum TSH greater than 4.0 mIU/L. Two larger trials have investigated the effects of levothyroxine in TPOAb-positive women when initiated preconception. The POSTAL study randomized TPOAb-positive, euthyroid women with infertility to 25 to 50 μg levothyroxine daily versus no treatment starting before IVF and continued through delivery.[3] Treatment was not associated with any effect on rates of miscarriage (rate difference [RD] –0.34%, 95% CI: –8.65 to 8.12) or live birth (RD –0.67%, 95% CI: –8.09 to 6.77). Most recently, the TABLET trial enrolled 952 euthyroid TPOAb-positive women with a history of infertility or miscarriage who were trying to conceive either naturally or through ART.[19] Women were randomized to 50 μg levothyroxine daily versus placebo, started preconception and continued until delivery. Levothyroxine therapy was not associated with any difference in live birth rates (RR 0.97, 95% CI: 0.83–1.14).

No therapy can currently be definitively recommended for women who are TPOAb-positive and euthyroid either in pregnancy or preconception. It is, however, important to remember that those women are at risk for the development of hypothyroidism during gestation. Guidelines currently recommend measurement of serum TSH when pregnancy is diagnosed and every 4 weeks through midpregnancy in women who are known to be TPOAb-positive.[23]

KEY POINTS

- Maternal TPOAb positivity has been linked to increased risks for miscarriage and preterm delivery, but mechanisms for these associations are uncertain. No therapy to date has demonstrated a benefit for TPOAb positive euthyroid women. TPOAb positive women are at increased risk for the development of hypothyroidism during gestation, and their serum TSH levels should be monitored.

SCREENING FOR THYROID DYSFUNCTION IN PREGNANCY

Whether or not all pregnant women should be routinely screened for thyroid dysfunction has been highly controversial.[74] Screening would largely be targeted toward the detection of hypothyroidism, as women with overt hyperthyroidism are likely to present with symptoms, and women who are subclinically hyperthyroid in pregnancy do not require treatment. It is estimated that universal screening may detect overt hypothyroidism in approximately 0.5% of pregnant women,[30] for whom intervention is likely to improve both obstetric and child developmental outcomes. One study has reported that thyroid screening is cost effective on the basis of the detection of overt hypothyroidism alone.[75] However, universal screening would also detect approximately 3.5% of women with subclinical hypothyroidism and 2.5% of women with isolated hypothyroxinemia (if testing includes FT4), for whom a benefit of therapy is uncertain. Universal screening would increase medical costs, could increase patient anxiety, and may lead to misinterpretation of thyroid function testing. Clinical practice guidelines regarding thyroid function testing in pregnancy are inconsistent, with some organizations advocating for universal screening, some for broadly targeted case-finding, and some for only very limited thyroid testing. The ATA does not recommend either for or against universal screening, but does endorse TSH testing for pregnant women considered to be at high risk for hypothyroidism, including those with suggestive symptoms, a personal or family history of thyroid disease or autoimmunity, a history of miscarriage or infertility, morbid obesity, age over 30 years, or two or more prior pregnancies.[23] A recent retrospective analysis in a single center concluded that these current case-finding criteria would detect fewer cases of subclinical hypothyroidism than a universal screening strategy, but that the proportion of women with overt hypothyroidism identified by each strategy did not differ.[76]

THYROID NODULES AND CANCER DURING PREGNANCY

The prevalence of thyroid nodules in pregnancy varies based on population iodine intakes. Thyroid nodules have been reported in up to one-third of pregnant women.[34] Stimulation of the thyroidal TSH receptor by hCG in early pregnancy and depletion of intrathyroidal iodine stores over the course of gestation in mildly to moderately iodine-deficient regions may be drivers of nodule growth in pregnant women. Among 221 Chinese pregnant women followed prospectively with ultrasonography during pregnancy, 15% had nodules at least 2 mm in size during the first trimester, with nodules being more frequent in older women and those with higher gravidity.[77] Nodule volumes increased over the course of gestation, and sizes remained increased at 3 months postpartum. Eleven percent of participants developed new nodules during pregnancy. Thyroid cancer is also relatively common in women of childbearing age. A retrospective analysis including 4,846,505 deliveries that occurred in California from 1991 to 1999 found that 3.3/10,000 women had thyroid cancer diagnosed within 9 months before pregnancy, 0.3/100,000 had cancer diagnosed at the time of delivery, and 10.8/100,000 were diagnosed within 1 year postpartum.[78]

Decisions about the need for thyroid fine-needle aspiration (FNA) biopsy in pregnant women should be based on the same ultrasonographic patterns that are used outside the pregnancy setting. FNA biopsy is safe during gestation and should generally be performed for suspicious nodules when serum TSH is not suppressed. Radioactive iodine scanning is contraindicated in pregnant women, so when low serum TSH values persist beyond week 16 of pregnancy it is reasonable to defer FNA biopsy until after delivery, when radionuclide scanning can be used to determine whether nodules are hyperfunctioning. When benign FNA biopsy results are obtained before or during gestation, pregnancy should not alter the usual nodule surveillance strategy. For nodules with indeterminate FNA cytopathology, in most cases surgery or repeat biopsy can be safely deferred until after delivery. Testing with molecular markers is not recommended, because these tests have not been evaluated in pregnancy, and it is biologically plausible that gene expression, and thus test validity, may be altered during gestation.[23] Some women prefer to defer FNA biopsy or to defer surgery for suspicious nodules until after delivery. This appears safe based on retrospective data. A cohort that included 61 women whose thyroid cancer was diagnosed during pregnancy found that long-term (median 22.5 years) outcomes did not differ among the 14 women who opted for thyroid surgery during pregnancy, the 47 who deferred surgery until after delivery, and 528 age-matched controls who had not been pregnant at the time of their cancer diagnosis.[79] Another retrospective series followed 19 patients who were diagnosed with papillary thyroid cancer just before or during the early stages of pregnancy and elected to defer surgery until after delivery.[80] The median tumor size at diagnosis was 0.9 cm, and the median duration from diagnosis to surgery was 11.9 months. Serial ultrasonography during pregnancy demonstrated no clinically meaningful change in tumor size and no development of lymph node metastases. Finally, among 50 Japanese women undergoing active surveillance for low-risk papillary carcinomas who experienced 51 pregnancies from 1993 to 2013, pregnancy was associated with an increase in tumor size in only 8%, and none of the patients developed nodal metastases.[81]

In most cases, when papillary thyroid cancer is detected in early pregnancy surgery may be deferred until after delivery.[23] Serial ultrasound should be performed, and surgery during pregnancy may be considered if there is substantial growth before 24 to 26 weeks' gestation or if cytologically malignant cervical lymph nodes are present. When surgery is deferred, levothyroxine may be initiated to maintain the serum TSH between 0.3 and 2.0 mIU/L for the remainder of gestation. It is unclear whether pregnancy alters tumor aggressiveness. Stimulation of TSH receptor by hCG may promote tumor growth, and some differentiated thyroid cancers express estrogen receptor, which may also predispose them to growth during gestation.[82] A retrospective study of 123 Italian women with differentiated thyroid cancer found that diagnosis during pregnancy or in the first year postpartum was associated with higher risk for persistent or recurrent disease (60%) than diagnosis more than a year postpartum (4%) or diagnosis in women who had never been pregnant (13%). Disease persistence/recurrence was primarily defined by elevated serum thyroglobulin levels, and it is possible that pregnancy may have influenced the completeness of the initial thyroid surgery, and therefore confounded the study results. Another retrospective Italian study included 340 women with differentiated thyroid cancer and similarly reported higher risk for persistent or recurrent disease when women were diagnosed during or within 2 years following pregnancy (10.5%) compared with women who were diagnosed at least 2 years after delivery (1.3%) or nulliparous women (4.7%).[83] Tumor persistence/recurrence was not clearly defined, but likely was based on serum thyroglobulin. In both Italian studies, delaying surgery until after delivery did not appear to influence risk. Although only limited data are available regarding the course of anaplastic and medullary thyroid cancer in pregnancy, surgery should not be delayed when these cancers are diagnosed in pregnant women. For women who require thyroid surgery in pregnancy, the second trimester is safest, although the risk of complications of thyroid surgery may still be increased.[84]

A prior history of treated cancer should not be a deterrent to pregnancy plans. Tumor surveillance strategies during gestation should be guided by the ATA disease response category.[85] A single-center retrospective analysis examined 235 women with differentiated thyroid cancer who had a term pregnancy following initial cancer treatment between 1997 and 2015.[86] When evaluated 3 to 12 months after delivery, none of the 197 women who had an ATA excellent, indeterminate, or biochemical incomplete response to therapy before pregnancy had developed structural disease. However, among the 38 women with a structural incomplete response to therapy prepregnancy, worsened structural disease (a ≥3-mm increase in the size of known disease or new metastases) was identified after delivery in 11 (29%), of whom three (8%) required additional treatment. Therefore, additional tumor surveillance is not required during pregnancy for women with an excellent or indeterminate response to therapy, but additional ultrasound and thyroglobulin monitoring are reasonable in women with a known incomplete response to therapy.

Numerous studies have investigated whether a history of radioactive iodine treatment is associated with impaired fertility or adverse pregnancy outcomes. A recent systematic review found that that menstrual irregularities occur in the first year after treatment in up to 31% of patients, that anti-Mullerian hormone levels are significantly lower 1 year after treatment compared with baseline (mean difference 1.50 ng/mL, 95% CI: 0.82–2.17), and that menopause may occur at a slightly younger age in treated compared with untreated women (49.5 years vs. 51 years; $P < .001$).[87] However, in the same paper, a metaanalysis of four studies showed that women with prior radioactive iodine treatment were not at lower risk of becoming pregnant compared with controls (OR 0.98, 95% CI: 0.72–1.33).[87] A review of 54 studies found no evidence for an increase in miscarriage, congenital malformations, or premature delivery in women who had received radioactive iodine treatment before pregnancy, although it noted that transient menstrual irregularities can be expected in approximately 30% of women following radioactive iodine therapy.[88] More recently, a Korean population-based cohort study included 111,459 women of childbearing age who underwent thyroidectomy with or without radioactive iodine treatment for thyroid cancer.[89] Among the 10,842 women who became pregnant, miscarriage, preterm delivery, and congenital malformation rates did not differ between those who received radioactive iodine treatment compared with those who did not (32.1% vs. 30.7% for miscarriage, 12.9% vs. 12.8% for preterm delivery, and 9.0% vs. 8.9% for congenital malformations). However, in a subgroup analysis examining the time interval between radioactive iodine therapy and conception, 18 cases of congenital malformation were observed among women who became pregnant less than 6 months after radioactive iodine treatment, and the odds ratio for congenital malformation (adjusted for maternal age and cumulative radioactive iodine dose) was 1.74 (95% CI: 1.01–2.97; $P = .04$) for women who conceived up to 5 months after treatment compared with those who conceived 12 to 23 months after treatment. Overall, it appears that, although there may be transient menstrual irregularities following radioactive iodine treatment, radioactive iodine therapy does not substantially impair either fertility or subsequent pregnancy outcomes as long as pregnancy is postponed at least 6 months after ablation therapy.

KEY POINTS

- In most instances, surgery for thyroid cancer diagnosed during pregnancy can be safely deferred until after delivery. A prior history of thyroid cancer should not be a deterrent to pregnancy, but tumor surveillance should be intensified during pregnancy for women with known persistent or residual tumor.

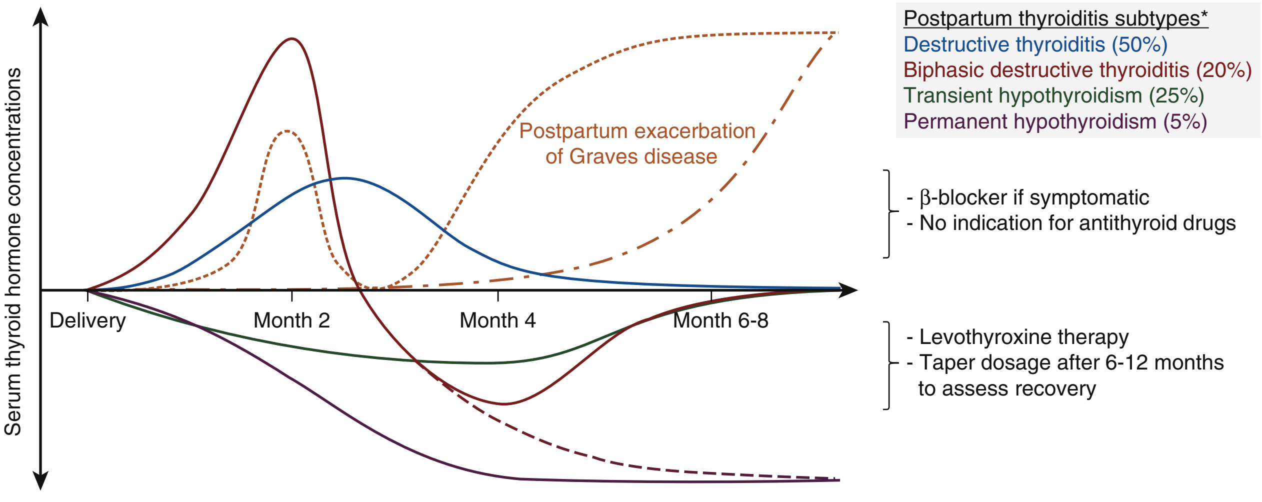
THYROID DISORDERS IN THE POSTPARTUM PERIOD

Postpartum Thyroiditis

Postpartum thyroiditis (PPT) is destructive thyroiditis occurring during the first 12 months postpartum. It likely results from a rebound of autoimmunity following the normal state of immune tolerance induced by pregnancy. It occurs in approximately 5% to 8% of women and is more frequent in women with TPOAb positivity in early pregnancy and in those with a personal or family history of autoimmune disorders.[90,91] PPT may occur following miscarriage. It can also occur in women receiving levothyroxine for mild preexisting hypothyroidism.[92] The typical sequence is a 4- to 8-week period of thyrotoxicosis caused by leakage of preformed hormone from the thyroid gland into the bloodstream, followed by a 4- to 6-month period of transient hypothyroidism (Fig. 75.6). However, only the thyrotoxic phase is seen in 50% of women, and only hypothyroidism in 25%. While most women will become euthyroid again within a year, up to 20% will be left with chronic hypothyroidism, and up to half will develop chronic hypothyroidism after an initial recovery.[90] PPT recurs after 70% of subsequent pregnancies.

New or recurrent Graves hyperthyroidism may present after delivery and must be distinguished from PPT in the postpartum setting. Antibody testing may help to establish the etiology of thyrotoxicosis, as TPOAbs are positive in most women with PPT, but the presence of TRAbs is highly sensitive and specific for Graves disease.[93] The ratio of serum T3 to T4 is generally less than 20:1 in PPT but is often higher in women with Graves hyperthyroidism.[94] Most women with PPT have only mild thyrotoxic symptoms, so the presence of severe thyrotoxic symptoms favors a diagnosis of Graves disease. Timing of onset may also provide some clues, with thyrotoxicosis from PPT presenting most frequently from 1 to 4 months postpartum, whereas Graves disease usually presents at 6 to 12 months postpartum or later.[91] Although rarely employed since the advent of sensitive and specific TRAb assays, the 24-hour radioactive iodine uptake test can be used to differentiate between Graves disease, in which uptake is elevated, and PPT, in which the uptake is low or undetectable.

Palpitations and other symptoms of the thyrotoxic phase of PPT can be managed with a β-adrenergic antagonist such as metoprolol or propranolol that is safe in lactation. ATDs do not have a role in destructive thyroiditis and should not be used. Levothyroxine should



*Up to 20% of women who initially return to a euthyroid state will develope chronic hypothyroidism later in life.

Fig. 75.6 Trajectories of postpartum thyroiditis subtypes.

be initiated in the hypothyroid phase of PPT if women are symptomatic or are actively trying to conceive. Because hypothyroidism is usually transient, levothyroxine doses can be tapered after 6 to 12 months of therapy to determine whether underlying thyroid function has recovered. Women with a history of PPT should have annual TSH monitoring, because they remain at high risk for the development of chronic hypothyroidism. Trials of iodine supplementation and levothyroxine for TPOAb-positive pregnant women have not shown any benefit for the prevention of PPT.[61] Trials of selenium supplementation for the prevention of PPT in TPOAb-positive women have had mixed results, and, in light of potential side effects, such supplementation is not currently recommended.[23,95]

KEY POINTS

- Graves disease must be differentiated from postpartum thyroiditis when thyrotoxicosis presents in the postpartum setting. Symptoms in the thyrotoxic phase of postpartum thyroiditis can be managed with a β-adrenergic antagonist, and levothyroxine may be required in the hypothyroid phase if symptoms are present or conception is desired.

Considerations in Lactating Women

Case reports and animal studies suggest that maternal hypothyroidism will suppress milk production in lactating women; the effects of maternal hyperthyroidism on milk production are largely unknown. TSH testing is advocated in women who have difficulty with lactation in the absence of other known causes, and levothyroxine should be initiated if hypothyroidism is identified.[23,91] Only trace amounts of maternal thyroid hormone are present in breast milk, so maternal thyroid dysfunction or levothyroxine ingestion does not alter the thyroid status of breastfed infants (for effects of maternal ATD ingestion, see earlier).

Sodium-iodide symporter (NIS) is expressed in the lactating breast and actively concentrates iodine in breast milk, where it serves as a critical nutritional source for the breastfed infant.[96] Therefore, women who are lactating have increased iodine requirements, and iodine supplementation is recommended for lactating women in regions where iodine intake is inadequate.[23] Because radioactive iodine is concentrated in breast milk, the use of ^{131}I is absolutely contraindicated in women who are lactating owing to concerns about transmission to the breastfed infant and radiation-induced damage to breast tissue.[23] It is recommended that women who require radioactive iodine therapy for hyperthyroidism or thyroid cancer should cease lactation at least 6 weeks, and optimally at least 3 months, before treatment to allow regression of lactation-induced NIS expression.[97] ^{123}I may safely be used for scanning and uptake testing in lactating women if breast milk is pumped and discarded for 3 to 4 days. Similarly, use of Tc-99m pertechnetate requires breast milk to be pumped and discarded for 24 hours after scanning.[23]

SUMMARY AND FUTURE DIRECTIONS

Approaches to the diagnosis and treatment of thyroid disorders may differ in women during preconception, pregnancy, and the postpartum period. Thyroid function tests must be interpreted using gestation-specific reference ranges. In most regions, additional studies are needed to derive gestation-specific, assay-specific laboratory reference ranges for TSH and FT4 and to determine optimal assay-specific thresholds for TPOAb positivity. Thyroid dysfunction is associated with an increased risk for infertility; however, prospective studies are needed to better understand the impact of preconception thyroid function on both fertility and pregnancy outcomes. Overt hypothyroidism in pregnancy requires levothyroxine treatment to prevent adverse obstetric and child neurodevelopmental outcomes. Whether to treat subclinical hypothyroidism in pregnancy remains controversial. Adequately powered trials to assess whether initiation of levothyroxine in early gestation are needed but may be challenging to conduct. Recent large registry studies have demonstrated teratogenicity of both MMI and PTU when used in the first trimester of pregnancy. This has led to an increased emphasis on preconception counseling of women with Graves disease, with consideration of definitive therapy. When ATDs are used in pregnancy, close monitoring and dose titration are essential. Prospective data are needed to inform optimal strategies for ATD use in pregnancy, particularly whether to change from PTU to MMI after organogenesis is complete. TPOAb positivity is associated with an increased risk for miscarriage and preterm delivery even in euthyroid women, but particularly when cooccurring with subclinical hypothyroidism. Future studies will be needed to better understand associations of maternal TPOAb with obstetric outcomes other than pregnancy loss and with child neurodevelopmental outcomes. Current data do not support levothyroxine treatment for euthyroid, TPOAb-positive women. Studies to better understand the mechanisms behind the observed associations between TPOAb positivity and pregnancy loss will be needed to inform new interventional strategies. The workup of thyroid nodules in pregnancy must take into account patient preferences regarding the timing of biopsy and surgery, as well as ultrasonographic patterns. In women who are thyroid cancer survivors, tumor surveillance strategies during gestation can be guided by the ATA disease response category and should be more aggressive in those with a known structural incomplete response to therapy. Finally, PPT is common, particularly among TPOAb-positive women. It is important to monitor TSH periodically in women with a history of PPT because they are at high risk for the development of chronic hypothyroidism.

REFERENCES

1. Dhillon-Smith RK, Tobias A, Smith PP, et al. The prevalence of thyroid dysfunction and autoimmunity in women with history of miscarriage or subfertility. *J Clin Endocrinol Metab*. 2020;105:dgaa302.
2. Korevaar TIM, Derakhshan A, Taylor PN, et al. Association of thyroid function test abnormalities and thyroid autoimmunity with preterm birth: a systematic review and meta-analysis. *JAMA*. 2019;322:632–641.
3. Wang H, Gao H, Chi H, et al. Effect of levothyroxine on miscarriage among women with normal thyroid function and thyroid autoimmunity undergoing in vitro fertilization and embryo transfer: a randomized clinical trial. *JAMA*. 2017;318:2190–2198.
4. Aghajanova L, Lindeberg M, Carlsson IB, et al. Receptors for thyroid-stimulating hormone and thyroid hormones in human ovarian tissue. *Reprod Biomed Online*. 2009;18:337–347.
5. Maruo T, Matsuo H, Mochizuki M. Thyroid hormone as a biological amplifier of differentiated trophoblast function in early pregnancy. *Acta Endocrinol (Copenh)*. 1991;125:58–66.
6. Maruo T, Katayama K, Barnea ER, et al. A role for thyroid hormone in the induction of ovulation and corpus luteum function. *Horm Res*. 1992;37:12–18.
7. Cecconi S, Rucci N, Scaldaferri ML, et al. Thyroid hormone effects on mouse oocyte maturation and granulosa cell aromatase activity. *Endocrinology*. 1999;140:1783–1788.
8. Wakim AN, Polizotto SL, Buffo MJ, et al. Thyroid hormones in human follicular fluid and thyroid hormone receptors in human granulosa cells. *Fertil Steril*. 1993;59:1187–1190.
9. Zhang C, Xia G, Tsang BK. Interactions of thyroid hormone and FSH in the regulation of rat granulosa cell apoptosis. *Front Biosci (Elite Ed)*. 2011;3:1401–1413.
10. Zhang C, Wang X, Wang Z, et al. Effect of different culture systems and 3, 5, 3'-triiodothyronine/follicle-stimulating hormone on preantral follicle development in mice. *PLoS One*. 2013;8:e61947.
11. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*. 2010;31:702–755.
12. Huang SA, Mulcahey MA, Crescenzi A, et al. Transforming growth factor-beta promotes inactivation of extracellular thyroid hormones via transcriptional stimulation of type 3 iodothyronine deiodinase. *Mol Endocrinol*. 2005;19:3126–3136.
13. Chen S, Zhou X, Zhu H, et al. Preconception TSH and pregnancy outcomes: a population-based cohort study in 184 611 women. *Clin Endocrinol (Oxf)*. 2017;86:816–824.
14. Plowden TC, Schisterman EF, Sjaarda LA, et al. Subclinical hypothyroidism and thyroid autoimmunity are not associated with fecundity, pregnancy loss, or live birth. *J Clin Endocrinol Metab*. 2016;101:2358–2365.
15. Taylor PN, Minassian C, Rehman A, et al. TSH levels and risk of miscarriage in women on long-term levothyroxine: a community-based study. *J Clin Endocrinol Metab*. 2014;99:3895–3902.
16. Zhao T, Chen BM, Zhao XM, et al. Meta-analysis of ART outcomes in women with different preconception TSH levels. *Reprod Biol Endocrinol*. 2018;16:111.
17. Velkeniers B, Van Meerhaeghe A, Poppe K, et al. Levothyroxine treatment and pregnancy outcome in women with subclinical hypothyroidism undergoing assisted reproduction technologies: systematic review and meta-analysis of RCTs. *Hum Reprod Update*. 2013;19:251–258.
18. Busnelli A, Paffoni A, Fedele L, et al. The impact of thyroid autoimmunity on IVF/ICSI outcome: a systematic review and meta-analysis. *Hum Reprod Update*. 2016;22:775–790.
19. Dhillon-Smith RK, Middleton LJ, Sunner KK, et al. Levothyroxine in women with thyroid peroxidase antibodies before conception. *N Engl J Med*. 2019;380:1316–1325.
20. Hershman JM. The role of human chorionic gonadotropin as a thyroid stimulator in normal pregnancy. *J Clin Endocrinol Metab*. 2008;93:3305–3306.
21. Korevaar TIM, Steegers EA, Pop VJ, et al. Thyroid autoimmunity impairs the thyroidal response to human chorionic gonadotropin: Two population-based prospective cohort studies. *J Clin Endocrinol Metab*. 2017;102:69–77.
22. Medici M, Korevaar TI, Visser WE, et al. Thyroid function in pregnancy: what is normal? *Clin Chem*. 2015;61:704–713.
23. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27:315–389.
24. Lee RH, Spencer CA, Mestman JH, et al. Free T4 immunoassays are flawed during pregnancy. *Am J Obstet Gynecol*. 2009;200:260.e1–e6.
25. Anckaert E, Poppe K, Van Uytfanghe K, et al. FT4 immunoassays may display a pattern during pregnancy similar to the equilibrium dialysis ID–LC/tandem MS candidate reference measurement procedure in spite of susceptibility towards binding protein alterations. *Clinica Chimica Acta*. 2010;411:1348–1353.
26. Sapin R, d'Herbomez M. Free thyroxine measured by equilibrium dialysis and nine immunoassays in sera with various serum thyroxine-binding capacities. *Clin Chem*. 2003;49:1531–1535.
27. Berta E, Samson L, Lenkey A, et al. Evaluation of the thyroid function of healthy pregnant women by five different hormone assays. *Pharmazie*. 2010;65:436–439.
28. Bliddal S, Feldt-Rasmussen U, Boas M, et al. Gestational age-specific reference ranges from different laboratories misclassify pregnant women's thyroid status: comparison of two longitudinal prospective cohort studies. *Eur J Endocrinol*. 2014;170:329–339.
29. Levie D, Korevaar TIM, Bath SC, et al. Thyroid function in early pregnancy, child IQ, and autistic traits: a meta-analysis of individual participant data. *J Clin Endocrinol Metab*. 2018;103:2967–2979.
30. Dong AC, Stagnaro-Green A. Differences in diagnostic criteria mask the true prevalence of thyroid disease in pregnancy: a systematic review and meta-analysis. *Thyroid*. 2019:278–289.
31. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:238–249.
32. Carle A, Pedersen IB, Knudsen N, et al. Epidemiology of subtypes of hyperthyroidism in Denmark: a population-based study. *Eur J Endocrinol*. 2011;164:801–809.
33. Aggarawal N, Suri V, Singla R, et al. Pregnancy outcome in hyperthyroidism: A case control study. *Gynecol Obstet Invest*. 2014;77:94–99.
34. Ollero MD, Toni M, Pineda JJ, et al. Thyroid function reference values in healthy iodine-sufficient pregnant women and influence of thyroid nodules on thyrotropin and free thyroxine values. *Thyroid*. 2019;29:421–429.
35. Laurberg P, Wallin G, Tallstedt L, et al. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*. 2008;158:69–75.
36. Laurberg P, Andersen SL. Antithyroid drug use in pregnancy and birth defects: why some studies find clear associations, and some studies report none. *Thyroid*. 2015;25:1185–1190.
37. Andersen SL, Olsen J, Wu CS, et al. Birth defects after early pregnancy use of antithyroid drugs: a Danish nationwide study. *J Clin Endocrinol Metab*. 2013;98:4373–4381.
38. Seo GH, Kim TH, Chung JH. Antithyroid drugs and congenital malformations: A nationwide Korean cohort study. *Ann Intern Med*. 2018;168:405–413.

39. Zakarija M, McKenzie JM. Pregnancy-associated changes in the thyroid-stimulating antibody of Graves' disease and the relationship to neonatal hyperthyroidism. *J Clin Endocrinol Metab*. 1983;57:1036–1040.
40. Barbesino G, Tomer Y. Clinical review: clinical utility of TSH receptor antibodies. *J Clin Endocrinol Metab*. 2013;98:2247–2255.
41. Nedrebo BG, Holm PI, Uhlving S, et al. Predictors of outcome and comparison of different drug regimens for the prevention of relapse in patients with Graves' disease. *Eur J Endocrinol*. 2002;147:583–589.
42. Momotani N, Noh J, Oyanagi H, et al. Antithyroid drug therapy for Graves' disease during pregnancy. Optimal regimen for fetal thyroid status. *N Engl J Med*. 1986;315:24–28.
43. Korevaar TI, Chaker L, Jaddoe VW, et al. Maternal and birth characteristics are determinants of offspring thyroid function. *J Clin Endocrinol Metab*. 2016;101:206–213.
44. Yoshihara A, Noh JY, Watanabe N, et al. Substituting potassium iodide for methimazole as the treatment for graves' disease during the first trimester may reduce the incidence of congenital anomalies: a retrospective study at a single medical institution in Japan. *Thyroid*. 2015;25:1155–1161.
45. Kampmann JP, Johansen K, Hansen JM, et al. Propylthiouracil in human milk. Revision of a dogma. *Lancet*. 1980;1:736–737.
46. Johansen K, Andersen AN, Kampmann JP, et al. Excretion of methimazole in human milk. *Eur J Clin Pharmacol*. 1982;23:339–341.
47. Haddow JE, Palomaki GE, Allan WC, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*. 1999;341:549–555.
48. Alexander EK, Marqusee E, Lawrence J, et al. Timing and magnitude of increases in levothyroxine requirements during pregnancy in women with hypothyroidism. *N Engl J Med*. 2004;351:241–249.
49. Maraka S, Ospina NM, O'Keeffe DT, et al. Subclinical hypothyroidism in pregnancy: a systematic review and meta-analysis. *Thyroid*. 2016;26:580–590.
50. Derakhshan A, Peeters RP, Taylor PN, et al. Association of maternal thyroid function with birthweight: a systematic review and individual-participant data meta-analysis. *Lancet Diabetes Endocrinol*. 2020;8:501–510.
51. Liu H, Shan Z, Li C, et al. Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid*. 2014;24:1642–1649.
52. Thompson W, Russell G, Baragwanath G, et al. Maternal thyroid hormone insufficiency during pregnancy and risk of neurodevelopmental disorders in offspring: A systematic review and meta-analysis. *Clin Endocrinol (Oxf)*. 2018;88:575–584.
53. Nazarpour S, Ramezani Tehrani F, et al. Effects of levothyroxine on pregnant women with subclinical hypothyroidism, negative for thyroid peroxidase antibodies. *J Clin Endocrinol Metab*. 2018;103:926–935.
54. Lazarus JH, Bestwick JP, Channon S, et al. Antenatal thyroid screening and childhood cognitive function. *N Engl J Med*. 2012;366:493–501.
55. Casey BM, Thom EA, Peaceman AM, et al. Treatment of subclinical hypothyroidism or hypothyroxinemia in pregnancy. *N Engl J Med*. 2017;376:815–825.
56. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Obstetrics. Thyroid disease in pregnancy: ACOG Practice Bulletin, Number 223. *Obstet Gynecol*. 2020;135(6):e261–e274.
57. Lazarus J, Brown RS, Daumerie C, et al. 2014 European thyroid association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J*. 2014;3:76–94.
58. Korevaar TIM, Tiemeier H, Peeters RP. Clinical associations of maternal thyroid function with foetal brain development: Epidemiological interpretation and overview of available evidence. *Clin Endocrinol (Oxf)*. 2018;89:129–138.
59. Korevaar TI, Muetzel R, Medici M, et al. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*. 2016;4:35–43.
60. Andersen SL, Andersen S, Vestergaard P, et al. maternal thyroid function in early pregnancy and child neurodevelopmental disorders: a Danish nationwide case-cohort study. *Thyroid*. 2018;28:537–546.
61. De Leo S, Pearce EN. Autoimmune thyroid disease during pregnancy. *Lancet Diabetes Endocrinol*. 2018;6:575–586.
62. Dhillon-Smith RK, Coomarasamy A. TPO antibody positivity and adverse pregnancy outcomes. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101433.
63. Korevaar TIM, Pop VJ, Chaker L, et al. Dose dependency and a functional cutoff for tpo-antibody positivity during pregnancy. *J Clin Endocrinol Metab*. 2018;103:778–789.
64. Thangaratinam S, Tan A, Knox E, et al. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. *BMJ*. 2011;342:d2616.
65. Xie J, Jiang L, Sadhukhan A, et al. Effect of antithyroid antibodies on women with recurrent miscarriage: a meta-analysis. *Am J Reprod Immunol*. 2020;83:e13238.
66. He X, Wang P, Wang Z, et al. Thyroid antibodies and risk of preterm delivery: a meta-analysis of prospective cohort studies. *Eur J Endocrinol*. 2012;167:455–464.
67. Li Y, Shan Z, Teng W, et al. Abnormalities of maternal thyroid function during pregnancy affect neuropsychological development of their children at 25-30 months. *Clin Endocrinol (Oxf)*. 2010;72:825–829.
68. Derakhshan A, Korevaar TIM, Taylor PN, et al. The association of maternal thyroid autoimmunity during pregnancy with child IQ. *J Clin Endocrinol Metab*. 2018;103:3729–3736.
69. Korevaar TIM, Steegers EA, Pop VJ, et al. Thyroid autoimmunity impairs the thyroidal response to human chorionic gonadotropin: two population-based prospective cohort studies. *J Clin Endocrinol Metab*. 2017;102:69–77.
70. Glinoer D, Riahi M, Grün JP, et al. Risk of subclinical hypothyroidism in pregnant women with asymptomatic autoimmune thyroid disorders. *J Clin Endocrinol Metab*. 1994;79:197–204.
71. Toulis KA, Goulis DG, Venetis CA, et al. Thyroid autoimmunity and miscarriages: the corpus luteum hypothesis. *Med Hypotheses*. 2009;73:1060–1062.
72. Negro R, Formoso G, Mangieri T, et al. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. *J Clin Endocrinol Metab*. 2006;91:2587–2591.
73. Nazarpour S, Ramezani Tehrani F, et al. Effects of levothyroxine treatment on pregnancy outcomes in pregnant women with autoimmune thyroid disease. *Eur J Endocrinol*. 2017;176:253–265.
74. Taylor PN, Zouras S, Min T, et al. Thyroid screening in early pregnancy: pros and cons. *Front Endocrinol (Lausanne)*. 2018;9:626.
75. Dosiou C, Barnes J, Schwartz A, et al. Cost-effectiveness of universal and risk-based screening for autoimmune thyroid disease in pregnant women. *J Clin Endocrinol Metab*. 2012;97:1536–1546.
76. Sitoris G, Veltri F, Kleynen P, et al. Screening for thyroid dysfunction in pregnancy with targeted high-risk case finding: can it be improved? *J Clin Endocrinol Metab*. 2019;104:2346–2354.
77. Kung AW, Chau MT, Lao TT, et al. The effect of pregnancy on thyroid nodule formation. *J Clin Endocrinol Metab*. 2002;87:1010–1014.
78. Smith LH, Danielsen B, Allen ME, et al. Cancer associated with obstetric delivery: Results of linkage with California cancer registry. *Am J Obstet Gynecol*. 2003;189:1128–1135.
79. Moosa M, Mazzaferri EL. Outcome of differentiated thyroid cancer diagnosed in pregnant women. *J Clin Endocrinol Metab*. 1997;82:2862–2866.
80. Oh HS, Kim WG, Park S, et al. Serial neck ultrasonographic evaluation of changes in papillary thyroid carcinoma during pregnancy. *Thyroid*. 2017;27:773–777.
81. Ito Y, Miyauchi A, Kudo T, et al. Effects of pregnancy on papillary microcarcinomas of the thyroid re-evaluated in the entire patient series at Kuma Hospital. *Thyroid*. 2016;26:156–160.
82. Vannucchi G, Perrino M, Rossi S, et al. Clinical and molecular features of differentiated thyroid cancer diagnosed during pregnancy. *Eur J Endocrinol*. 2010;162:145–151.
83. Messuti I, Corvisieri S, Bardesono F, et al. Impact of pregnancy on prognosis of differentiated thyroid cancer: clinical and molecular features. *Eur J Endocrinol*. 2014;170:659–666.
84. Kuy S, Roman SA, Desai R, et al. Outcomes following thyroid and parathyroid surgery in pregnant women. *Arch Surg*. 2009;144:399–406.

85. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
86. Rakhlin L, Fish S, Tuttle RM. Response to therapy status is an excellent predictor of pregnancy-associated structural disease progression in patients previously treated for differentiated thyroid cancer. *Thyroid*. 2017;27:396–401.
87. Piek MW, Postma EL, van Leeuwaarde R, et al. The effect of radioactive iodine therapy on ovarian function and fertility in female thyroid cancer patients: a systematic review and meta-analysis. *Thyroid*. 2020;31:658–668.
88. Sioka C, Fotopoulos A. Effects of I-131 therapy on gonads and pregnancy outcome in patients with thyroid cancer. *Fertil Steril*. 2011;95:1552–1559.
89. Kim HO, Lee K, Lee SM, et al. Association between pregnancy outcomes and radioactive iodine treatment after thyroidectomy among women with thyroid cancer. *JAMA Intern Med*. 2019;180:54–61.
90. Pearce EN. Management of thyrotoxicosis: preconception, pregnancy, and the postpartum period. *Endocr Pract*. 2019;25:62–68.
91. Amino N, Arata N. Thyroid dysfunction following pregnancy and implications for breastfeeding. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101438.
92. Moleti M, Mauro MD, Alibrandi A, et al. Postpartum thyroiditis in women with euthyroid and hypothyroid Hashimoto's thyroiditis antedating pregnancy. *J Clin Endocrinol Metab*. 2020;105:dgaa197.
93. Ide A, Amino N, Kang S, et al. Differentiation of postpartum Graves' thyrotoxicosis from postpartum destructive thyrotoxicosis using antithyrotropin receptor antibodies and thyroid blood flow. *Thyroid*. 2014;24:1027–1031.
94. Amino N, Yabu Y, Miki T, et al. Serum ratio of triiodothyronine to thyroxine, and thyroxine-binding globulin and calcitonin concentrations in Graves' disease and destruction-induced thyrotoxicosis. *J Clin Endocrinol Metab*. 1981;53:113–116.
95. Hubalewska-Dydejczyk A, Duntas L, Gilis-Januszewska A. Pregnancy, thyroid, and the potential use of selenium. *Hormones (Athens)*. 2020;19:47–53.
96. Portulano C, Paroder-Belenitsky M, Carrasco N. The Na+/I- symporter (NIS): mechanism and medical impact. *Endocr Rev*. 2014;35:106–149.
97. Sisson JC, Freitas J, McDougall IR, et al. Radiation safety in the treatment of patients with thyroid diseases by radioiodine 131I: practice recommendations of the American Thyroid Association. *Thyroid*. 2011;21:335–346.

76

Iodine Nutrition and the Thyroid

Michael B. Zimmermann

OUTLINE

Dietary Sources Of Iodine, 1266
  Absorption and Bioavailability of Iodine, 1266
  Thyroidal Iodine Uptake, 1266
  Goitrogens, 1266
Metabolism, 1267
  Thyroidal Adaptation to Iodine Deficiency, 1267
  Iodine Requirements, 1267
Methods To Assess Iodine Status, 1267
  Urinary Iodine Concentration, 1267
  Thyroid Size, 1268
  Thyroid-Stimulating Hormone, 1268
  Thyroglobulin, 1268
Effects of Iodine Deficiency on the Thyroid and Brain, 1269
  Goiter, 1269
  Pregnancy and Infancy, 1269
  Childhood, 1270
  Adulthood, 1271
  Epidemiology of Iodine Deficiency, 1271
Treatment And Prevention of Iodine Deficiency, 1272
  Salt Fortification with Iodine, 1272
  Iodine Supplementation, 1272
Iodine Excess, 1273

Bernard Courtois, a French chemist, discovered iodine (atomic weight 126.9) in 1811. He gave the purple crystals to Gay-Lussac, who identified it as a new element and named it “iode,” from the Greek *ιώδης* (i.e., *violet*).[1] In 1819, Jean-François Coindet, in Geneva, reported treating goiter with tincture of iodine. The connection between goiter, myxedema, and iodine was established in 1896, when Baumann and Roos discovered iodine in the thyroid.[1] In the early twentieth century, groundbreaking studies by Swiss and American scientists showed the efficacy of iodine prophylaxis in the prevention of goiter and cretinism. The first use of salt iodization as a public health measure was in northern Switzerland in the early 1920s.[1]

DIETARY SOURCES OF IODINE

Absorption and Bioavailability of Iodine

Iodine (as iodide) is widely but unequally distributed in the Earth’s environment.[2] In many regions, leaching from flooding and erosion have exhausted soils of iodide, and crops grown in these soils will be low in iodine, and humans and animals consuming food grown in these soils become iodine-deficient. Iodine deficiency in these areas will persevere until iodine enters the food chain through dietary diversification using foods produced outside the iodine-deficient area or fortification of iodine in foods (e.g., iodization of salt). The native iodine content of most foods and beverages is small; commonly-consumed foods provide only 3 to 80 μg per serving.[3] Key dietary sources of iodine in industrialized countries like the United States and Switzerland are milk and bread (containing iodized salt).[3,4] In many countries, iodized salt used in the household is the major dietary source. Dietary supplements frequently contain potassium iodide at amounts in the range of 50 to 250 μg.[5]

Iodide is rapidly and nearly entirely absorbed (>90%) in the stomach and duodenum.[5] Iodide absorption in the gut is thought to be largely passive, but active absorption may also occur through the sodium/iodide symporter (NIS)[6] and the sodium/multivitamin transporter[7]; both are present in the gut mucosa. Organically bound iodine is usually digested and the freed iodide absorbed, but some forms may be absorbed intact; for example, approximately 70% of an oral dose of thyroxine (T_4), is absorbed intact.[8]

Thyroidal Iodine Uptake

Iodine is cleared from the bloodstream mostly by the thyroid and kidney, and while renal iodine clearance is fairly constant, thyroid clearance varies strikingly depending on the chronic intake of iodine. In conditions of sufficient dietary iodine supply, the thyroid normally takes up less than 20% of absorbed iodine. In chronic iodine deficiency, this portion can exceed 80%.[9] Iodine in the blood is turned over quickly; normally, plasma iodide has a half-life of approximately 10 hours, but this is reduced in iodine deficiency or hyperthyroidism.[9] The body of a healthy adult contains up to 20 mg of iodine, of which 70% to 80% is in the thyroid.[9] In chronic iodine deficiency, the amount of iodine in the thyroid may fall to less than 20 μg. In iodine-sufficient areas, the adult thyroid traps approximately 60 μg of iodine per day to balance losses and support thyroid hormone synthesis. The NIS, a protein in the basolateral thyrocyte membrane, moves iodide into the thyroid at a concentration gradient 20 to 50 times that of plasma.[10] It concentrates iodine by active transport that uses energy released by the inward translocation of sodium down its electrochemical gradient to drive the simultaneous inward translocation of iodine against its electrochemical gradient.[10] Metabolism of circulating thyroid hormone in peripheral tissues liberates iodine that joins the plasma iodine pool and can be cleared by the thyroid or kidney. Greater than 90% of ingested iodine is eventually excreted in the urine, with only a minor amount appearing in the feces and sweat.[9]

Goitrogens

Although iodine deficiency is typically the underlying cause of endemic goiter, a range of naturally occurring compounds and environmental

TABLE 76.1 Recommendations for Iodine Intake (µg/day) by Age or Population Group

Population group	US Institute of Medicine[7]		Population Group	World Health Organization[22]
	EAR	AI* or RDA		RNI
Infants 0–12 mo	—	110–130*	Children 0–5 y	90
Children 1–8 y	65	90	Children 6–12 y	120
Children 9–13 y	73	120		
Adults ≥14 y	95	150	Adults >12 y	150
Pregnancy	160	220	Pregnancy	250
Lactation	200	290	Lactation	250

EAR, Estimated average requirement; *AI*, adequate intake; *RDA*, recommended daily allowance; *RNI*, recommended nutrient intake.

pollutants may disrupt thyroid function and be goitrogenic.[11] Cruciferous vegetables, including cabbage, kale, cauliflower, broccoli, turnips, and rapeseed, contain glucosinolates; their metabolites compete with iodine for thyroidal uptake. Cassava, lima beans, linseed, sorghum, and sweet potato contain cyanogenic glucosides; these may be metabolized to thiocyanates that also compete with iodine for thyroidal uptake. An example is linamarin, a thioglycoside found in cassava, a staple food in many developing countries. If cassava is not sufficiently soaked or cooked to eliminate the linamarin, it is hydrolyzed in the gut to release cyanide, which is metabolized to thiocyanate. Soy and millet contain flavonoids that may impair activity of the thyroid peroxidase enzyme. Nearly all goitrogens are clinically relevant only if dietary iodine supply is low and/or goitrogen intake is high and prolonged. Dietary deficiencies of selenium, iron, and vitamin A aggravate the effects of iodine deficiency. Glutathione peroxidase and the deiodinases are selenium-dependent enzymes. In selenium deficiency, accumulated peroxides may damage the thyroid, and deiodinase deficiency reduces thyroid hormone synthesis.[12] Iron deficiency limits heme-dependent thyroid peroxidase activity and impairs thyroid hormone synthesis. In areas of endemic goiter, iron deficiency anemia diminishes the efficacy of iodine prophylaxis, while iron supplementation improves the effectiveness of iodized oil and iodized salt.[13] Vitamin A deficiency in iodine-deficient children triggers thyroid-stimulating hormone (TSH) hyperstimulation of the thyroid and increased risk for goiter through decreased vitamin A-mediated suppression of the pituitary *TSHβ* gene.[14]

KEY POINTS

- Most foods have low amounts of iodine, and diets in many countries cannot provide adequate iodine without iodine fortification of salt.

METABOLISM

Thyroidal Adaptation to Iodine Deficiency

The thyroid adjusts to low intakes of dietary iodine by distinct changes in its activity. In most adults, if iodine intake drops below approximately 100 µg/day, TSH secretion is augmented. This increases plasma iodide clearance by the thyroid through stimulation of NIS expression.[15] As the thyroid clears a larger fraction of circulating iodide, there is a gradual reduction in renal iodide excretion, and the urinary iodine concentration falls. However, in most healthy adults, as long as daily iodine intake is at least approximately 60 µg, despite a reduction in circulating plasma inorganic iodine and urinary iodine, absolute uptake of iodine by the thyroid and thyroidal iodine stores remain sufficient. Below approximately 60 µg, despite maximal fractional clearance of circulating iodine by the thyroid, absolute uptake falls, the iodine content of the thyroid is exhausted, and many people develop goiter.[9]

Goiter and thyroid hyperplasia produced by iodine deficiency are accompanied by altered thyroid hormone synthesis. Poorly iodinated thyroglobulin in the colloid leads to an increase in monoiodotyrosine and T_3 and a decrease in diiodotyrosine and T_4. The typical pattern of circulating thyroid hormones in children in areas of moderate-to-severe iodine deficiency is a variably elevated TSH, a normal or high-normal serum T_3, and a low serum T_4; this pattern is also seen in adults, but less predictably.[16] The serum thyroglobulin (Tg) concentration is usually elevated. The increased serum T_3/T_4 ratio is important in the adaptation to iodine deficiency because T_3 has approximately four times the metabolic potency of T_4 but requires only 75% as much iodine for its synthesis. Thyroid failure and cretinism typically develop only in regions of chronic, severe iodine deficiency, where most individuals have low circulating T_4 and T_3 and dramatically elevated TSH.[12,16]

Iodine Requirements

Recommended daily intakes for iodine are shown in Table 76.1. The World Health Organization (WHO)[17] uses the Recommended Nutrient Intake (RNI) to define iodine requirements; the RNI is the intake estimated to cover the needs of "nearly all" healthy individuals in the specified life stage. The US Institute of Medicine uses the Recommended Dietary Allowance (RDA), which is the average daily intake sufficient to meet the iodine requirement of 97% to 98% of healthy individuals in a life stage.[5]

The iodine requirement during pregnancy is increased due to: (a) an increase in maternal T_4 synthesis to preserve maternal euthyroidism and transfer thyroid hormone to the fetus early in the first trimester, before the fetal thyroid is active, (b) iodine transfer to the fetus, chiefly in later gestation, and (c) an increase in renal iodine clearance.[18] The WHO recommends a daily iodine intake of 250 µg/day for pregnant women.[17] A recent study in healthy infants suggested that the iodine requirement in 2- to 5-month-old infants is 70 µg/d; this would correspond to an RDA of 80 µg.[19] Thus, the current adequate intake of 110 to 130 µg/day for infants[5] may exceed actual requirements.

KEY POINTS

- The thyroid can adapt to moderately-low intakes of iodine by increasing plasma iodide clearance. Iodine requirements vary by age and life stage.

METHODS TO ASSESS IODINE STATUS

Urinary Iodine Concentration

Because greater than 90% of dietary iodine ultimately appears in the urine,[9] urinary iodine concentration (UIC) is an excellent indicator of recent iodine intake. UIC can be expressed as a concentration (µg/L), relative to creatinine excretion (µg iodine/g creatinine), or as 24-hour

TABLE 76.2 Epidemiologic Criteria for Assessment of Iodine Nutrition in a Population Based on Median or Range of Urinary Iodine Concentrations

	Iodine Intake	Iodine Nutrition
School-aged children		
<20 μg/L	Insufficient	Severe iodine deficiency
20–49 μg/L	Insufficient	Moderate iodine deficiency
50–99 μg/L	Insufficient	Mild iodine deficiency
100–299 μg/L	Adequate	Optimum
>300 μg/L	Excessive	Risk of adverse health consequences (iodine-induced hyperthyroidism, autoimmune thyroid disease)
Pregnant women		
<150 μg/L	Insufficient	
150–249 μg/L	Adequate	
250–499 μg/L	More than adequate	
≥500 μg/L[a]	Excessive	
Lactating women[b]		
<100 μg/L	Insufficient	
≥100 μg/L	Adequate	
Children less than 2 years of age		
<100 μg/L	Insufficient	
≥100 μg/L	Adequate	

[a] The term excessive means in excess of the amount needed to prevent and control iodine deficiency.

[b] In lactating women, the numbers for median urinary iodine are lower than the iodine requirements, because of the iodine excreted in breast milk.

(From World Health Organization, United Nations Children's Fund, International Council for the Control of Iodine Deficiency Disorders. Assessment of iodine deficiency disorders and monitoring their elimination. A guide for programme managers. 3rd ed. Geneva: World Health Organization; 2007; Zimmermann MB, Aeberli I, Andersson M, et al. Thyroglobulin is a sensitive measure of both deficient and excess iodine intakes in children and indicates no adverse effects on thyroid function in the UIC range of 100-299 μg/L: a UNICEF/ICCIDD study group report. *J Clin Endocrinol Metab.* 2013;98:1271–1280; WHO Secretariat, Andersson M, de Benoist B, et al. Prevention and control of iodine deficiency in pregnant and lactating women and in children less than 2-years-old: conclusions and recommendations of the Technical Consultation. *Public Health Nutr.* 2007;10:1606–1611.)

excretion (μg/day). For population surveys, UIC should be measured in spot urine specimens from a representative sample of the target group and expressed as the median, in μg/L[17] (Table 76.2). Although the median UIC (mUIC) does not provide specific information on thyroid function, a low value implies that a population is at greater risk of developing thyroid disorders.[20] The mUIC is a reliable indicator of iodine status in populations, but its value for gauging individual status is reduced by high day-to-day variability in iodine intakes.[9] In European adults, intraindividual variation in UIC is 35% for both 24-hour collections and spot samples, so that urine samples from approximately 10 different days are required to assess individual iodine status with 20% precision.[21]

Thyroid Size

Two methods are available for determining if a goiter is present: neck inspection and palpation, and thyroid ultrasonography. By palpation, a thyroid is considered goitrous when each lateral lobe has a volume greater than the terminal phalanx of the thumbs of the subject being examined.[17] Goiter surveys are frequently done in school-age children. However, palpation of goiter in areas of mild iodine deficiency has poor sensitivity and specificity; in such areas, measurement of thyroid volume by ultrasound is superior. Thyroid ultrasound is quickly done and feasible in remote areas using compact equipment, and references for thyroid volume are available from iodine-sufficient children.[17] In areas of endemic goiter, although thyroid size predictably declines in response to increases in iodine intake, thyroid size may not return to normal for months or years after correction of iodine deficiency. During this transition period, the goiter rate is problematic to interpret, because it reflects both a population's history of iodine nutrition and its present status. A sustained salt iodization program will reduce the goiter rate to less than 5% in school-age children, and this denotes disappearance of iodine deficiency as a significant public health problem.[22]

Thyroid-Stimulating Hormone

Circulating TSH is an indirect indicator of iodine nutrition because its concentration in blood is determined largely by circulating thyroid hormone, which in turn reflects iodine intake. However, in older children and adults, although serum TSH may be variably increased by iodine deficiency, values frequently remain within the normal range. Thus, an elevated TSH is an insensitive biomarker of iodine nutrition in adults. It may, however, be a sensitive indicator of iodine status in the newborn period.[17] Serum TSH concentrations are increased in iodine-deficient infants for the first few weeks of life. As demonstrated in national data from Switzerland,[23] if less than 3% of newborn TSH values are above 5 mU/L in whole blood collected 3 to 4 days after birth, this suggests iodine sufficiency in the population.[17] Otherwise, thyroid hormone concentrations are generally poor indicators of iodine status in children and adults, except in severe deficiency, when hypothyroidism develops. In mild-to-moderately deficient populations, serum T_3 increases or remains unchanged, and serum T_4 may fall, but these changes are typically within the normal range.[24]

Thyroglobulin

Tg is a thyroid-specific protein. In iodine sufficiency, only small amounts of Tg are secreted into the circulation, and serum Tg is normally less than 10 μg/L.[25] In regions of endemic goiter, serum Tg increases due to larger thyroid cell mass and higher TSH stimulation. Serum Tg is well correlated with the severity of iodine deficiency as measured by UIC.[26] Intervention studies exploring the potential of Tg as a biomarker of response to iodine prophylaxis have shown that Tg decreases rapidly with iodine repletion.[26] An assay for Tg has been developed for dried blood spots taken by a finger prick,[26] simplifying collection and transport; reference ranges are available in iodine-sufficient schoolchildren[27] and pregnant women.[28] The value of Tg to measure response to introduction of iodized salt was confirmed in a national study in Denmark, where correction of mild-to-moderate iodine deficiency decreased the prevalence of elevated Tg (>40 μg/L) from 11.3% to 3.7%.[29]

KEY POINTS

- Four methods are recommended for assessment of iodine nutrition in populations: UIC, goiter rate, TSH, and Tg. These indicators are complementary, in that UIC is a sensitive indicator of recent iodine intake (days), and Tg shows an intermediate response (weeks to months), whereas changes in the goiter rate reflect long-term iodine nutrition (months to years).

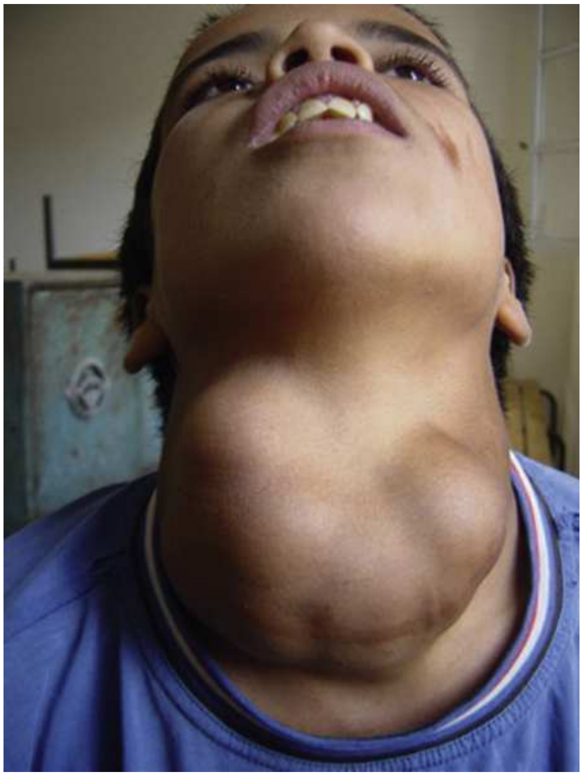


Fig. 76.1 Nodular goiter in a 14-year-old boy photographed in an area of severe iodine deficiency disorders in northern Morocco, with tracheal and esophageal compression and hoarseness, likely due to damage to the recurrent laryngeal nerves. (Photograph courtesy of M.B. Zimmermann, Zurich, Switzerland.)

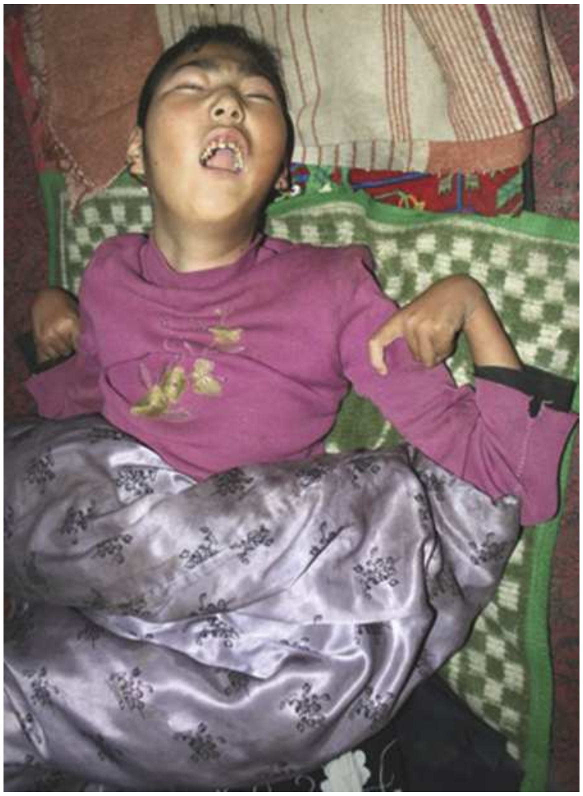


Fig. 76.2 Neurologic cretinism in a 9-year-old girl from western China with the characteristic features: severe mental deficiency together with squint, deaf mutism, and motor spasticity of the arms and legs. (Photograph courtesy of M.B. Zimmermann, Zurich, Switzerland.)

EFFECTS OF IODINE DEFICIENCY ON THE THYROID AND BRAIN

Goiter

Goiter is the classic manifestation of iodine deficiency. Iodine-deficiency goiters are initially characterized by diffuse, homogeneous enlargement, but over time nodules often develop (Fig. 76.1). Iodine deficiency sharply increases risk for multinodular toxic goiter, primarily in women older than 50 years.[9] Many thyroid nodules derive from a somatic mutation and are of monoclonal origin[30]; the mutations appear to be more likely in nodules under the influence of a growth promoter, such as iodine deficiency. In mild iodine deficiency, there is frequently only mild enlargement of the thyroid, and the characteristic histologic findings are hyperplasia with abundant parenchyma, high follicular epithelium, and rare colloid. In chronic severe deficiency, repeated episodes of hyperplasia followed by involution and atrophy result in a thyroid with a mixture of nodules, zones of hyperplasia, and sections of involution, degeneration, and repair.

Pregnancy and Infancy

In regions of chronic iodine deficiency, women enter pregnancy with already depleted iodine stores, which, combined with low dietary iodine, reduces maternal and fetal thyroid hormone production.[31] The adverse effects of hypothyroidism *in utero* depend upon its timing and severity. T_3 is the active form of thyroid hormone in the brain, and approximately 80% of brain T_3 is generated locally from T_4.[32] T_3 plays an essential role in brain myelination and cell migration, differentiation, and maturation.[32] Fetal brain development may be sensitive to maternal hypothyroidism as early as the first trimester; nuclear thyroid hormone receptors are found in the fetal brain by 9 weeks, before full development of the fetal thyroid. Fetal thyroid function only becomes important later in pregnancy; onset of fetal thyroid hormone secretion occurs late in the second trimester, at 18 to 22 weeks of gestation.[32] Thus, maternal thyroid hormone is essential for fetal brain development during the first half of pregnancy, and severe maternal hypothyroxinemia during this period is thought to be the cause of cretinism.[33]

Cretinism is the most severe manifestation of iodine deficiency during pregnancy. The term “cretin” was originally used in Diderot’s *Encyclopedia* in 1754 to refer to “an imbecile who is deaf and dumb, with a goiter hanging down to the waist”; at that time, the condition was common in the Alpine regions of central Europe. In McCarrison’s singular description of cretinism in northern India in 1908,[34] he described a neurologic form, with predominantly neuromotor defects, and a myxedematous form, marked by severe hypothyroidism and short stature.

The three characteristic features of neurologic cretinism are severe mental retardation with squint, deaf mutism, and motor spasticity.[35] There is proximal rigidity of both upper and lower extremities and the trunk, and corresponding proximal spasticity with exaggerated deep tendon reflexes at the knees, adductors, and biceps. In addition to frank cretinism, a greater percentage of an affected population has lesser degrees of mental impairment, elevated hearing thresholds, and coordination defects[35] (Fig. 76.2).

The typical myxedematous cretin has less severe mental impairment than the neurologic cretin. Myxedematous cretins have the features of severe hypothyroidism present since early life, as in untreated sporadic congenital hypothyroidism. These include severe growth retardation,

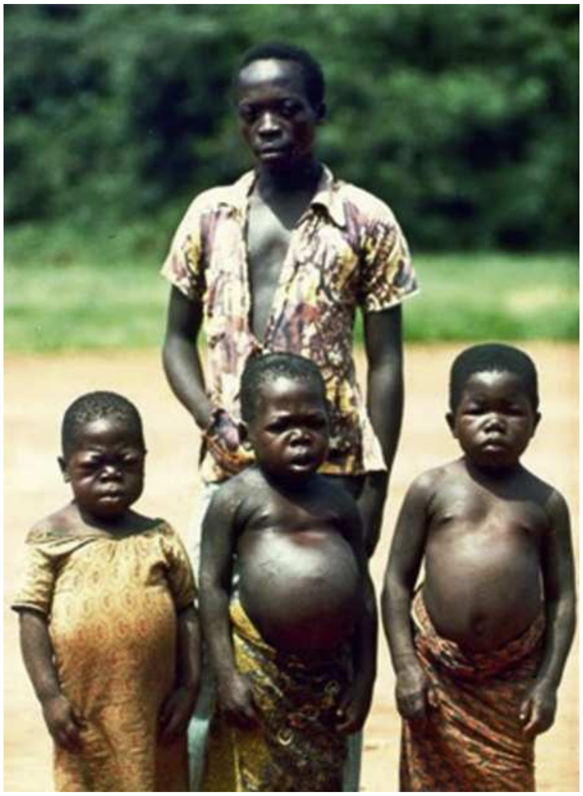


Fig. 76.3 Myxedematous cretinism in the Democratic Republic of Congo. Four inhabitants aged 15 to 20 years: a normal male and three female cretins with severe hypothyroidism with dwarfism, retarded sexual development, puffy features, dry skin and hair, and severe mental retardation. (Photograph courtesy of Professor F. Delange, Brussels, Belgium.)

incomplete maturation of the face, myxedematous, thickened and dry skin, dry and rare hair, and delayed sexual maturation.[9] Circulating T_4 and T_3 are extremely low, often undetectable, and TSH is strongly elevated. Thyroid scans typically demonstrate a thyroid in the normal location but of small size with heterogeneous and patchy distribution of the tracer. It may be difficult to clearly differentiate between neurologic and myxedematous forms of cretinism; cretinism may present as a mixed form with features of both (Fig. 76.3).

In a landmark trial in an area of severe iodine deficiency in Papua New Guinea,[36,37] alternate families received saline (control) or iodized oil injection. The primary outcome was the prevalence of cretinism at 4- and 10-year follow-up. Iodine significantly reduced the prevalence of endemic cretinism: At 4 years of age, the relative risk (95% confidence interval [CI]) was 0.27 (0.12–0.60) and at 10 years of age, the relative risk (95% CI) was 0.17 (0.05–0.58). In a study in western China, an area of severe iodine deficiency and endemic cretinism, participants were groups of children from birth to 3 years and women at each trimester of pregnancy.[33] Untreated children 1 to 3 years of age, studied when first seen, served as controls. The intervention was oral iodized oil, and treated children and the babies born to the treated women were followed for 2 years. The prevalence of moderate or severe neurologic abnormalities among the infants whose mothers received iodine in the first or second trimester was 2%, as compared with 9% among the infants who received iodine during the third trimester (through the treatment of their mothers) or after birth.[33] Treatment at the end of the first trimester improved neurologic outcomes. Treatment in the third trimester of pregnancy or after delivery did not improve neurologic outcomes, but developmental quotients improved slightly. The prevalence of microcephaly was 27% in the untreated children compared with 11% in the treated children. The mean (± standard deviation [SD]) developmental quotient at 2 years of age was higher in the treated than in the untreated children (90 ± 14 vs. 75 ± 18).[38]

The cognitive deficits associated with iodine deficiency may not be limited to severely iodine-deficient areas. Two observational studies in high-income countries reported lower IQ and poorer school performance in children born to mildly iodine-deficient mothers[39,40]; others comparing different doses of iodine supplements did not.[41] In randomized controlled trials of iodine supplementation in mild-to-moderately iodine-deficient pregnant women in Europe, iodine treatment reduced maternal and newborn thyroid size, and, in some, decreased maternal TSH.[31] However, none of the trials found an effect on maternal and newborn total or free thyroid hormone concentrations, which are likely the best surrogate predictors of infant development.[42] A recent multicenter trial randomized pregnant women in India and Thailand ($n = 832$; mUIC 131 μg/L) to receive daily 200 μg oral iodine or placebo until term.[43] Supplementation had no effect on child neurodevelopment measured at 1, 2, and 5.4 years. These studies are reassuring in that they suggest that, in areas of mild-to-moderate iodine deficiency, the maternal thyroid is able to adapt to meet the increased thyroid hormone requirements of pregnancy.[31] The American[44] and European Thyroid Associations[45] recommend women take an iodine supplement during pregnancy in areas of even mild deficiency. In contrast, WHO does not recommend maternal iodine supplements in countries with well-functioning iodized salt programs,[46] concluding that, if women are iodine-replete before conception, they can cover their requirements during pregnancy by increasing fractional clearance of plasma iodide and drawing from thyroidal iodine stores.

In areas of severe iodine deficiency, lower maternal T_4 concentrations predict higher infant mortality, and the infant mortality rate may decrease if iodine deficiency is corrected before or during pregnancy.[47] Delong et al.[47] added potassium iodate to irrigation water in three areas of severe iodine deficiency in China and found a large reduction in both neonatal and infant mortality in the following 2 to 3 years compared with areas that did not receive iodine. In Algeria, compared with untreated women, rates of spontaneous abortion, stillbirth, and prematurity were significantly lower in women treated with oral iodized oil during pregnancy.[48] A randomized, placebo-controlled trial of oral iodized oil (100 mg iodine) was conducted in Indonesia to evaluate the effect on infant mortality.[49] The iodine or placebo was given with oral poliovirus vaccine; infants ($n = 617$) were treated at approximately 6 weeks of age and were followed for 6 months. There was a 72% decrease in risk of infant death during the first 2 months of follow-up.[49]

Childhood

Two recent systematic reviews[50,51] have confirmed the benefits of correcting iodine deficiency. The first looked at 89 studies that provided iodized salt to pediatric and adult populations and found a significant 72% to 76% reduction in risk for low intelligence (defined as IQ <70) and an 8.2- to 10.5-point overall increase in IQ.[50] The second review similarly reported that iodine-sufficient children have a 6.9- to 10.2-point higher IQ than iodine-deficient children.[51] Although the overall quality of the studies included in these analyses was low, they imply that individuals, and particularly children, with chronic, severe iodine deficiency experience a mean reduction in IQ of 7 to 11 points. Two high-quality randomized trials in school-aged children have assessed the effect of iodine treatment on cognition.[52,53] A 6-month intervention trial in moderately iodine-deficient 10- to 12-year-old children in Albania given either 400 mg of iodine as oral iodized oil or placebo[52] reported that iodine increased mean circulating T_4 by approximately 40% and improved several cognitive and motor tests. A controlled trial in mildly iodine-deficient 10- to 13-year-old children in New Zealand[53] gave a daily tablet containing 150 μg iodine as potassium iodide (KI) or placebo for 28 weeks. The overall cognitive score of the iodine group was 0.19 SDs higher than that of the placebo group. Therefore, available data suggest that in children raised in areas of mild-to-moderate iodine deficiency, cognitive impairment is at least partially reversible by iodine repletion.

Adulthood

Iodine status is an important determinant of thyroid diseases in adult populations.[20] Iodine deficiency predictably increases risk for diffuse goiter and thyroid nodules.[20] Severe iodine deficiency causes hypothyroidism because, despite an increase in thyroid activity to maximize iodine uptake and recycling, iodine intakes are simply too low to maintain thyroid hormone production. In mild-to-moderate iodine deficiency, the thyroid gland is able to compensate for low dietary intake by increasing thyroid activity, which maintains thyroid hormone production, but at a price: in many individuals, chronic stimulation of the thyroid leads to thyroid nodularity and autonomy.[20] This increase in nodularity subsequently increases risk of hyperthyroidism if iodine intakes are increased by iodized salt.[20] However, this increased risk of hyperthyroidism in the population is transient, as iodine sufficiency normalizes thyroid activity; in the long term, this reduces nodularity and autonomy.[20,54] The small increase in mild subclinical hypothyroidism in mainly younger adults during a shift from low to high iodine intakes may be due to increased thyroid autoimmunity and might also be transient, but more long-term studies are needed. In a survey across China 20 years after the implementation of salt iodization, iodine deficiency was significantly associated with most thyroid disorders, while excessive iodine intake was only associated with an increased prevalence of subclinical hypothyroidism and clinical hyperthyroidism.[55]

Available evidence from animal and epidemiological studies suggests iodine deficiency is a risk factor for follicular thyroid cancer, and possibly for anaplastic thyroid cancer.[56] However, the overall incidence of thyroid cancer in populations does not seem to be affected by the usual range of normal to high iodine intakes from dietary sources.[57] A pooled analysis of 13 case-control studies reported a significant decrease in thyroid cancer risk with high fish consumption (a surrogate for iodine intake) in regions with endemic goiter (odds ratio 0.65, 95% CI: 0.48–0.88) but not in iodine-rich regions (odds ratio 1.1, 95% CI: 0.85–1.5).[58] Although several ecological studies have suggested an increase in papillary thyroid cancer after the introduction of iodized salt to populations, confounding environmental factors and increasing diagnostic intensity limit their conclusions.[56,59] Finally, a recent metaanalysis of eight case-control studies, including data from China (n = 4974), found that more than adequate or excess iodine intake (>300 mg/d) was associated with a 23% decrease in risk for thyroid cancer.[60] Over the past three to four decades in the United States, the prevalence of thyroid cancer has been steadily increasing, but iodine intakes during the same period have decreased by over 50%.[56] Thus, a higher intake of dietary iodine may be a protective factor against thyroid cancer,[56] but more high-quality case-control studies are needed in this area.

KEY POINTS

- Iodine deficiency has multiple adverse effects on growth and development; these result from inadequate thyroid hormone production due to lack of sufficient iodine. During pregnancy, iodine requirements increase sharply, and iodine deficiency during pregnancy can impair cognitive development of the offspring.

Epidemiology of Iodine Deficiency

The WHO's first estimate of the global prevalence of goiter in 1960 suggested that 20% to 60% of the world's population was affected by iodine deficiency.[61] The International Child Development Steering Group subsequently highlighted iodine deficiency as one of four key global risk factors for impaired child development.[62] Programs against iodine deficiency are appealing for national governments, because the health and economic consequences can be prevented by salt iodization, a low-cost and sustainable intervention. Between 1942 and 2020, 123 countries introduced mandatory legislation on salt iodization. In 2021, 124 countries have legislation for mandatory salt iodization, and 21 have legislation allowing voluntary salt iodization.[63] Salt is considered adequately iodized when the fortification level is 15 to 40 ppm.[64] The United Nations Children's Fund estimated that, in 2018, 88% of the global population used iodized salt.[65]

The Iodine Global Network (IGN) assembles data from iodine surveys conducted throughout the world and continually monitors global iodine status.[66] In population iodine monitoring, the WHO recommends that iodine status in school-age children serve as a proxy for the general population[17]; therefore, in the IGN Scorecard, emphasis is given to studies in school-age children. Cross-sectional studies measuring UICs have been conducted in 152 out of 194 countries in the past 15 years: in 133 countries, the studies were nationally representative.[66] The number of countries with adequate iodine intake doubled over the past 20 years and increased from 67 in 2003 to 105 in 2011 and to 117 in 2021,[66] proving the effectiveness of salt iodization. However, in 2021, 21 countries still have insufficient iodine in their diets (Fig. 76.4). Iodine intake is lowest in Madagascar, where the mandatory salt iodization program recently fell apart due to political instability. In Vietnam, weakening of the previous mandatory standard allowed the introduction of noniodized salt, and iodine status deteriorated. Several large countries have incomplete nationwide coverage and strong regional variations in iodine status, e.g., Sudan and Russia. Iodine intake is low in several countries with strong health systems and otherwise successful public health programs (Norway, Germany, and Finland); in these countries, a major challenge is the low use of iodized salt in the production of processed foods. On the other hand, 14 countries have excessive iodine intake, as defined by a mUIC greater than 300 μg/L.[66] Excess iodine intakes in populations can result from iodine-rich groundwater (e.g., Somalia) and/or diets that are naturally high in iodine (e.g., South Korea). Iodine excess can also occur when the level of iodine added to salt is too high considering per capita salt intakes (e.g., Colombia). The WHO recommends setting the salt iodization level based on estimated per-capita salt intakes: in populations where mean salt intake per capita is 5 and 10 g/day, salt should be iodized at 40 and 20 ppm, respectively.[67] In situations where monitoring suggests excess intakes, the salt iodization level should be reduced to bring the population mUIC to less than 300 μg/L. A good example is Brazil, where, in response to surveys showing excess iodine intakes, the fortification level of iodine in salt has been reduced to 15 to 45 mg.[68] Thus, regular monitoring of salt iodization programs and population iodine status can help prevent both iodine deficiency and iodine excess.

KEY POINTS

- The number of countries with adequate iodine intake has increased from 67 in 2003 to 117 in 2021, but 21 countries still have insufficient iodine intake, while 14 countries have excessive iodine intake, either due to groundwater iodine or overiodized salt.

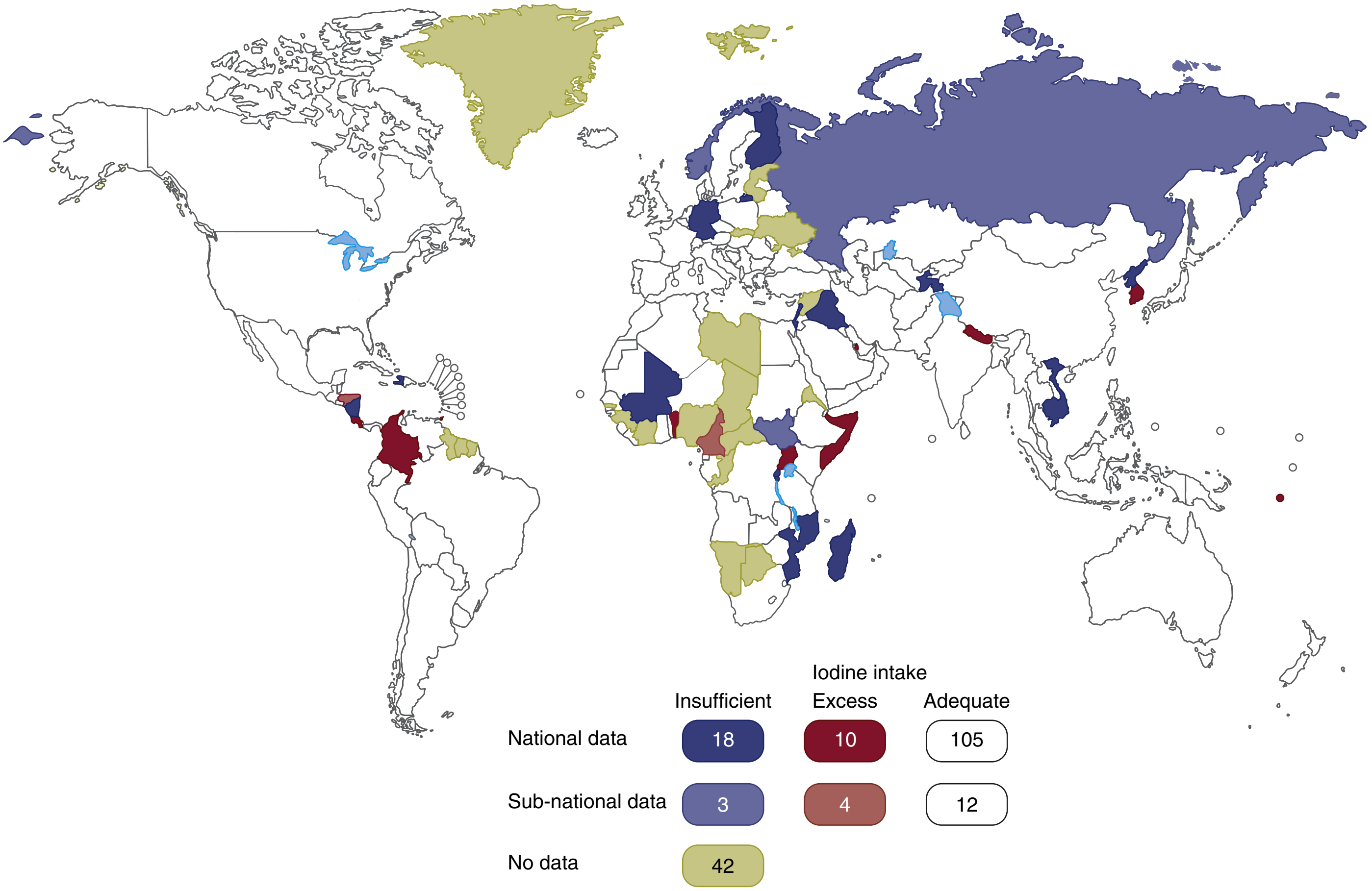


Fig. 76.4 World map showing national iodine status based on the median urinary iodine concentration in school-age children in 2021. (Adapted from The Iodine Global Network. Global scorecard of iodine nutrition in 2021 in the general population based on school-age children (SAC). IGN: Ottawa, Canada. 2021. Available at: https://www.ign.org/cm_data/IGN_Global_Scorecard_MAP_2021_SAC_-_7_May_2021.pdf. Accessed on: Oct 26, 2021.)

TREATMENT AND PREVENTION OF IODINE DEFICIENCY

Salt Fortification with Iodine

In regions affected by iodine deficiency, the most cost-effective and sustainable method to control iodine deficiency is through salt iodization.[17] Salt iodization is the recommended strategy for control of iodine deficiency disorders, because salt is one of few foodstuffs consumed by virtually everyone, and salt intake is fairly consistent through the year. The addition of iodine to salt does not affect its color or taste. In many countries, salt production/importation is limited to a few sources, and iodization technology is simple and inexpensive. Finally, the quantity of iodine in salt can be simply and inexpensively monitored at the production, retail, and household levels.[17]

Iodine can be added to salt in the form of KI or potassium iodate (KIO_3). Because KIO_3 has better stability than KI in the presence of salt impurities, humidity, and poor packaging,[69,70] it is the recommended form in tropical countries with low-grade salt. Potassium iodate has been granted "generally recognized as safe" status by the US Food and Drug Administration and has been used for decades as an additive to salt and bread.[71] Salt iodization is cost-effective.[72,73] Worldwide, the annual costs of salt iodization are US $0.02 to 0.05 per child covered, and the costs per disability-adjusted life years (DALYs) gained are US $34 to 36.[74]

The original concept of salt iodization was for it to be universal; that is, the iodization of all edible salt used in households and food processing.[17] However, many national programs have focused mainly on only household salt being iodized. The consumption of salt through processed foods is increasing in many countries, and in many middle- and high-income countries most salt is consumed through processed foods.[75] If the salt in such foods is well iodized, it can be an important source of iodine.[76] National salt iodization programs should encourage and monitor the use of iodized salt in processed foods. Iodine coming from processed foods likely explains why, in some countries, the mUIC indicates adequate iodine intakes despite low coverage of households with iodized salt.[77] There is growing worldwide interest in reducing salt intake to prevent hypertension and other noncommunicable diseases. Concerns have been raised that programs to reduce dietary salt could adversely impact programs to prevent iodine deficiency, and, conversely, that iodizing salt could encourage higher salt intakes.[78] The WHO has emphasized that these programs are compatible but need to be carefully aligned.[67] Both programs have common elements, including program monitoring via urine collections and interactions with the food and salt industries; this provides an opportunity to share and leverage resources and approaches to be more effective and efficient.[67,78] As salt intakes fall, iodine levels can be increased in salt to adjust for the recommended reduction in dietary salt to less than 5 g/day.[78]

Iodine Supplementation

In some areas, iodization of salt may not be practical for control of iodine deficiency, and iodine supplements can be given.[17] Iodized oil supplements are prepared by esterification of the unsaturated fatty

TABLE 76.3 Recommendations for Iodine Supplementation in Pregnancy and Infancy in Areas Where <90% of Households are Using Iodized Salt and the Median Urinary Iodine Concentration is <100 μg/L in Schoolchildren

Women of child-bearing age	A single annual oral dose of 400 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the RNI of 150 μg/d of iodine
Women who are pregnant or lactating	A single annual oral dose of 400 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the RNI of 250 μg/d iodine. Iodized oil should not be given to a woman who has already been given iodized oil during her current pregnancy or up to 3 mo before her current pregnancy started
Children aged 0–6 mo	A single oral dose of 100 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the of 90 μg/d of iodine Should be given iodized oil only if the mother was not supplemented during pregnancy or if the child is not being breastfed
Children aged 7–24 mo	A single annual oral dose of 200 mg of iodine as iodized oil as soon as possible after reaching 7 mo of age OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the RNI of 90 μg/d of iodine

RNI, Recommended nutrient intake.

(From WHO Secretariat, Andersson M, de Benoist B, et al. Prevention and control of iodine deficiency in pregnant and lactating women and in children less than 2-years-old: conclusions and recommendations of the Technical Consultation. *Public Health Nutr.* 2007;10:1606–1611.)

TABLE 76.4 Recommended Safe Upper Intake Levels for Iodine

Age Group	IOM (μg/day)[7]	SCF (μg/day)[136]	WHO (μg/kg/day)[137,138]
Premature infants	ND	ND	100
0–6 months	ND	ND	150
7–12 months	ND	ND	140
1–3 years	200	200	50
4–6 years	300 (4–8 years)	250	
7–10 years	600 (9–13 years)	300	50 (7–12 years)
11–14 years		450	30 (>13 years)
15–17 years	900 (14–18 years)	500	
Adults	1100	600	
Pregnant and lactating women	1100	600	40

IOM, United States Institute of Medicine; *SCF*, European Union Scientific Committee on Foods; *WHO*, World Health Organization; *ND*, not determined.

acids in seed or vegetable oils and addition of iodine to the double bonds.[79] It can be given orally or by intramuscular injection[17,80]; usual doses are 200 to 400 mg iodine/year.[17] Iodine can also be given as KI or KIO$_3$ as drops or tablets. Single oral doses of potassium iodide monthly (30 mg) or biweekly (8 mg) can provide adequate iodine for school-age children.[81] In iodine-deficient regions with weak iodized salt distribution, the WHO recommends that supplements be given to pregnant women, lactating women, and infants, according to the strategy shown in Table 76.3.[82] Lactating women who receive one dose of 400 mg iodine as oral iodized oil soon after delivery can provide adequate iodine to their infants through breast milk for at least 6 months.[83]

KEY POINTS

- In areas where iodized salt does not provide sufficient iodine, oral iodine supplementation can be given, either as long-lasting iodized oil or as daily iodine supplements.

IODINE EXCESS

In animals and humans, the thyroid gland has intrinsic autoregulatory mechanisms to effectively adjust to excess iodide intake.[9] The acute, transient inhibitory effect of iodide excess on thyroid iodide organification is termed the acute Wolff–Chaikoff effect, and it appears that adaptation to or escape from the Wolff–Chaikoff effect is due to a decrease in NIS protein,[84] thereby lowering intrathyroidal iodide and allowing normal hormone synthesis to resume. Since the initial description by Coindet in 1821 and the subsequent definition by Breuer and Kocher in 1904, iodide-induced thyrotoxicosis has been reported in patients with a variety of underlying thyroid diseases. Hyperthyroidism after iodine administration is termed the Jod–Basedow effect, which is named for the German word for iodine, "jod," and the name of Karl Adolph von Basedow, a German physician who described the effect.[85] Iodine-induced thyrotoxicosis may occur in patients with iodide-deficiency goiter, in patients with multinodular goiters who reside in areas of iodide sufficiency or deficiency, in various thyroid disorders, and in people with no evidence of underlying thyroid disease.[85,86] The incidence of iodine-induced thyrotoxicosis during the introduction of salt iodization in areas previously iodide deficient is variable, but is usually predicted by the severity of iodine deficiency before prophylaxis and to the increase in iodine intake.[87] In Zimbabwe, following the introduction of iodized salt at 30 to 90 mg iodine/kg, there was a 3-fold increase in thyrotoxicosis.[88] Thyrotoxicosis occurs when iodide repletion permits the autonomous thyroid tissue to synthesize and release excess quantities of thyroid hormone. In Denmark, in a population with moderate iodide deficiency, the prevalence of thyrotoxicosis slightly increased after iodine intake was increased by only 50 μg/day[89]; most patients had multinodular goiters and were elderly. This increase led to a transient increase in therapeutic use of antithyroid drugs in the population, which fell to a rate lower than before salt iodization after about 10 years.[90]

The lowest observed adverse effect level (LOAEL) proposed for iodine intakes is 1700 to 1800 μg/day. The tolerable upper intake level (UL) for iodine is the highest average daily intake that is likely to pose no risk of adverse health effects in the general population[5] (Table 76.4). The United States Institute of Medicine recommends a UL of 1100 μg/day,[5] whereas the European Union Scientific Committee on Foods recommends a UL of 600 μg/day[91]; these differences are due to differences in interpretation of the LOAEL effect. In 1988, the joint Food and Agriculture Organization of the United Nations and WHO Expert Committee on Food Additives suggested a UL from all iodine sources of 1 mg/day would be safe for most of the population.[92,93] For population monitoring of iodine excess, the WHO recommends using the mUIC; the upper threshold for iodine excess is a mUIC greater than or equal to 300 μg/L in 6- to 12-year-old children,[17,94] or greater than or equal to 500 μg/L in pregnant women.[17]

Though most foods are poor in native iodine, certain seaweeds, including kelp and kombu, are iodide-rich and can cause iodine excess.[95,96] Unprocessed meat is generally low in iodine, but accidental thyroid tissue found in ground beef has caused thyrotoxicosis (hamburger thyrotoxicosis).[97] Groundwater usually has an iodine content of only 1 to 10 μg/L, but iodine-rich drinking water is a source of excess iodide in some areas of China, where water has an iodine concentration of 300 to 500 μg/L.[98] Iodide readily crosses the placenta and is concentrated by the fetal thyroid. Large quantities of iodide administered to pregnant women may result in goiter in the newborn, probably because the fetal thyroid is highly sensitive to the inhibitory effect of iodide on hormone synthesis.[99] During lactation, iodide is also actively transported from the bloodstream into the breast and secreted into breast milk. Perinatal hypothyroidism and high infant urinary iodine levels were reported in newborns whose mothers consumed a large amount of seaweed during pregnancy and the postpartum period.[100]

KEY POINTS

- Both acute and chronic excess iodine intakes can have adverse effects on thyroid function.

REFERENCES

1. Zimmermann MB. Research on iodine deficiency and goiter in the 19th and early 20th centuries. *J Nutr*. 2008;138:2060–2063.
2. Küpper FC, Feiters MC, Olofsson B, et al. Commemorating two centuries of iodine research: an interdisciplinary overview of current research. *Angew Chem Int Ed Engl*. 2011;50:11598–11620.
3. Pearce EN, Pino S, He X, et al. Sources of dietary iodine: bread, cows' milk, and infant formula in the Boston area. *J Clin Endocrinol Metab*. 2004;89:3421–3424.
4. van der Reijden OL, Zimmermann MB, Galetti V. Iodine in dairy milk: sources, concentrations and importance to human health. *Best Pract Res Clin Endocrinol Metab*. 2017;31:385–395.
5. Institute of Medicine, Academy of Sciences, USA. *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium and Zinc*. Washington DC: National Academy Press; 2001.
6. Nicola JP, Basquin C, Portulano C, et al. The Na+/I– symporter mediates active iodide uptake in the intestine. *Am J Physiol Cell Physiol*. 2009;296:C654–C662.
7. de Carvalho FD, Quick M. Surprising substrate versatility in SLC5A6: Na+-coupled I– transport by the human Na+/multivitamin transporter (hSMVT). *J Biol Chem*. 2011;286:131–137.
8. Hays MT, Nielsen KRK. Human thyroxine absorption: age effects and methodological analyses. *Thyroid*. 1994;4:55–64.
9. Zimmermann MB. Iodine deficiency. *Endocr Rev*. 2009;30:376–408.
10. Eskandari S, Loo DDF, Dai G, et al. Thyroid Na+/I– symporter. Mechanism, stoichiometry, and specificity. *J Biol Chem*. 1997;272:27230–27238.
11. Gaitan E. Goitrogens in food and water. *Annu Rev Nutr*. 1990;10:21–39.
12. Zimmermann MB, Køhrle J. The impact of iron and selenium deficiencies on iodine and thyroid metabolism: biochemistry and relevance to public health. *Thyroid*. 2002;12:867–878.
13. Zimmermann MB. The influence of iron status on iodine utilization and thyroid function. *Annu Rev Nutr*. 2006;26:367–389.
14. Zimmermann MB, Jooste PL, Mabapa NS, et al. Vitamin A supplementation in iodine-deficient African children decreases thyrotropin stimulation of the thyroid and reduces the goiter rate. *Am J Clin Nutr*. 2007;86:1040–1044.
15. Brent GA, Taki K, Kogai T, et al. A thyroid-specific far-upstream enhancer in the human sodium/iodide symporter gene requires Pax-8 binding and cyclic adenosine 3',5'-monophosphate response element-like sequence binding proteins for full activity and is differentially regulated in normal and thyroid cancer cells. *Mol Endocrinol*. 2002;16:2266–2282.
16. Dumont JE, Ermans AM, Bastenie PA. Thyroid function in a goiter endemic V. Mechanism of thyroid failure in Uele endemic cretins. *J Clin Endocr Metab*. 1963;23:847–860.
17. World Health Organization, United Nations Children's Fund, International Council for the Control of Iodine Deficiency Disorders. Assessment of iodine deficiency disorders and monitoring their elimination. 3rd ed. Geneva: World Health Organization; 2007.
18. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocr Rev*. 1997;18:404–433.
19. Dold S, Zimmermann MB, Baumgartner J, et al. A dose-response crossover iodine balance study to determine iodine requirements in early infancy. *Am J Clin Nutr*. 2016;104:620–628.
20. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*. 2015;3:286–295.
21. König F, Andersson M, Hotz K, et al. Ten repeat collections for urinary iodine 1 from spot samples or 24-h samples are needed to reliably estimate individual iodine status in women. *J Nutr*. 2011;141:2049–2054.
22. Zimmermann MB, Hess SY, Adou P, et al. Thyroid size and goiter prevalence after introduction of iodized salt: a 5-y prospective study in schoolchildren in Côte d'Ivoire. *Am J Clin Nutr*. 2003;77:663–667.
23. Zimmermann MB, Aeberli I, Torresani T, et al. Increasing the iodine concentration in the Swiss iodized salt program markedly improved iodine status in pregnant women and children: a 5-y prospective national study. *Am J Clin Nutr*. 2005;82:388–392.
24. Delange F, Camus M, Ermans AM. Circulating thyroid hormones in endemic goiter. *J Clin Endocrinol Metab*. 1972;34:891–895.
25. Spencer CA, Wang CC. Thyroglobulin measurement. Techniques, clinical benefits, and pitfalls. *Endocrinol Metab Clin North Am*. 1995;24:841–863.
26. Zimmermann MB, Moretti D, Chaouki N, et al. Development of a dried whole-blood spot thyroglobulin assay and its evaluation as an indicator of thyroid status in goitrous children receiving iodized salt. *Am J Clin Nutr*. 2003;77:1453–1458.
27. Zimmermann MB, de Benoist B, Corigliano S, et al. Assessment of iodine status using dried blood spot thyroglobulin: development of reference material and establishment of an international reference range in iodine-sufficient children. *J Clin Endocrinol Metab*. 2006;91:4881–4887.
28. Stinca S, Andersson M, Weibel S, et al. Dried blood spot thyroglobulin as a biomarker of iodine status in pregnant women. *J Clin Endocrinol Metab*. 2017;102:23–32.
29. Vejbjerg P, Knudsen N, Perrild H, et al. Thyroglobulin as a marker of iodine nutrition status in the general population. *Eur J Endocrinol*. 2009;161:475–481.
30. Kopp P, Kimura ET, Aeschimann S, et al. Polyclonal and monoclonal thyroid-nodules coexist within human multinodular goiters. *J Clin Endocrinol Metab*. 1994;79:134–139.
31. Zimmermann MB. Iodine deficiency in pregnancy and the effects of maternal iodine supplementation on the offspring: a review. *Am J Clin Nutr*. 2009;89:668s–672s.
32. de Escobar GM, Obregon MJ, del Rey FE. Maternal thyroid hormones early in pregnancy and fetal brain development. *Best Pract Res Clin Endocrinol Metab*. 2004;18:225–248.
33. Cao XY, Jiang XM, Dou ZH, et al. Timing of vulnerability of the brain to iodine deficiency in endemic cretinism. *N Engl J Med*. 1994;331:1739–1744.
34. McCarrison R. Observations on endemic cretinism in the chitral and gilgit valleys. *Lancet*. 1908;2:1275–1280.
35. DeLong R. Neurological involvement in iodine deficiency disorders. In: Hetzel BS, Dunn JT, Stanbury JB, eds. *The Prevention and Control of Iodine Deficiency Disorders*. Amsterdam: Elsevier; 1987:49–63.
36. Pharoah PO, Buttfield IH, Hetzel BS. Neurological damage to fetus resulting from severe iodine deficiency during pregnancy. *Lancet*. 1971;1:308–310.

37. Pharoah POD, Connolly KJ. A controlled trial of iodinated oil for the prevention of endemic cretinism: a long-term follow-up. *Int J Epidemiol*. 1987;16:68–73.
38. O'Donnell KJ, Rakeman MA, Zhi-Hong D, et al. Effects of iodine supplementation during pregnancy on child growth and development at school age. *Dev Med Child Neurol*. 2002;44:76–81.
39. Bath SC, Steer CD, Golding J, et al. Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: results from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Lancet*. 2013;382:331–337.
40. Hynes KL, Otahal P, Hay I, et al. Mild iodine deficiency during pregnancy is associated with reduced educational outcomes in the offspring: 9-year follow-up of the gestational iodine cohort. *J Clin Endocrinol Metab*. 2013;98:1954–1962.
41. Rebagliato M, Murcia M, Alvarez-Pedrerol M, et al. Iodine supplementation during pregnancy and infant neuropsychological development: INMA mother and child cohort study. *Am J Epidemiol*. 2013;177:944–953.
42. Pop VJ, Kuijpens JL, van Baar AL, et al. Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy. *Clin Endocrinol (Oxf)*. 1999;50:149–155.
43. Gowachirapant S, Jaiswal N, Melse-Boonstra A, et al. Effect of iodine supplementation in pregnant women on child neurodevelopment: a randomised, double-blind, placebo-controlled trial. *Lancet Diabetes Endocrinol*. 2017;5:853–863.
44. Alexander EK, Pearce EN, Brent GA, et al. 2017 guidelines of the American thyroid association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27:315–389.
45. Lazarus J, Brown RS, Daumerie C, et al. 2014 European Thyroid Association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J*. 2014;3:76–94.
46. WHO Secretariat on behalf of the participants to the Consultation, Andersson M, de Benoist B, et al. Prevention and control of iodine deficiency in pregnant and lactating women and in children less than 2-years-old: conclusions and recommendations of the Technical Consultation. *Public Health Nutr*. 2007;10:1606–1611.
47. DeLong GR, Leslie PW, Wang SH, et al. Effect on infant mortality of iodination of irrigation water in a severely iodine-deficient area of China. *Lancet*. 1997;350:771–773.
48. Chaouki ML, Benmiloud M. Prevention of iodine deficiency disorders by oral administration of lipiodol during pregnancy. *Eur J Endocrinol*. 1994;130:547–551.
49. Cobra C, Muhilal RK, et al. Infant survival is improved by oral iodine supplementation. *J Nutr*. 1997;127:574–578.
50. Aburto N, Abudou M, Candeias V, et al. Effect and safety of salt iodization to prevent iodine deficiency disorders: a systematic review with meta-analyses. *WHO eLibrary of Evidence for Nutrition Actions*. Geneva: World Health Organization; 2016.
51. Bougma K, Aboud FE, Harding KB, et al. Iodine and mental development of children 5 years old and under: a systematic review and meta-analysis. *Nutrients*. 2013;5:1384–1416.
52. Zimmermann MB, Connolly K, Bozo M, et al. Iodine supplementation improves cognition in iodine-deficient schoolchildren in Albania: a randomized, controlled, double-blind study. *Am J Clin Nutr*. 2006;83:108–114.
53. Gordon RC, Rose MC, Skeaff SA, et al. Iodine supplementation improves cognition in mildly iodine-deficient children. *Am J Clin Nutr*. 2009;90:1264–1271.
54. Bu¨rgi H, Kohler M, Morselli B. Thyrotoxicosis incidence in Switzerland and benefit of improved iodine supply. *Lancet*. 1998;352:1034.
55. LI, Y, Teng D, Ba, J, et al. Efficacy and safety of long-term universal salt iodization on thyroid disorders: epidemiological evidence from 31 provinces of mainland China. *Thyroid*. 2020;30:568–579.
56. Zimmermann MB, Galetti V. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. *Thyroid Res*. 2015;8:8.
57. Peterson E, De P, Nuttall R. BMI, diet and female reproductive factors as risks for thyroid cancer: a systematic review. *PLoS One*. 2012;7:e29177.
58. Bosetti C, Kolonel L, Negri E, et al. A pooled analysis of case-control studies of thyroid cancer. VI. Fish and shellfish consumption. *Cancer Causes Control*. 2001;12:375–382.
59. Lee JH, Hwang Y, Song RY, et al. Relationship between iodine levels and papillary thyroid carcinoma: a systematic review and meta-analysis. *Head Neck*. 2017;39:1711–1718.
60. Cao LZ, Peng XD, Xie JP, et al. The relationship between iodine intake and the risk of thyroid cancer: a meta-analysis. *Medicine (Baltim)*. 2017;96:e6734.
61. Kelly FC, Snedden WW. *Prevalence and Geographical Distribution of Endemic Goiter. Monograph Series, World Health Organization*. Geneva: World Health Organization; 1960.
62. Walker SP, Wachs TD, Gardner JM, et al. Child development: risk factors for adverse outcomes in developing countries. *Lancet*. 2007;369:145–157.
63. Global Fortification Data Exchange (GFDx). Available at http://www.fortificationdata.org. Accessed on: February 26, 2021.
64. UNICEF. *Guidance on the Monitoring of Salt Iodization Programmes and Determination of Population Iodine Status*. New York: United Nations Children's Fund; 2018.
65. UNICEF. Global database, 2019, Based on Multiple Indicator Cluster Surveys (MICS), Demographic and Health Surveys (DHS) and Other Nationally Representative Household Surveys, 2013-2018. Percentage of Households Consuming Salt with Any Iodine, 2018 2019. Available at: https://data.unicef.org/topic/nutrition/iodine/. Accessed on: February 26, 2021.
66. The Iodine Global Network. Global Scorecard of Iodine Nutrition in 2021 in the General Population Based on School-Age Children. IGN: Ottawa, Canada. Available at: https://www.ign.org/scorecard.htm. Accessed on: March 8, 2021.
67. WHO. *Salt Reduction and Iodine Fortification Strategies in Public Health: Report of a Joint Technical Meeting*. Geneva: World Health Organization; 2014.
68. Cesar JA, Santos IS, Black RE, et al. Iodine status of Brazilian school-age children: a national cross-sectional survey. *Nutrients*. 2020;12:1077.
69. Diosady LL, Alberti JO, Mannar MGV, et al. Stability of iodine in iodized salt used for correction of iodine-deficiency disorders, I. *Food Nutr Bull*. 1997;18:388–396.
70. Diosady LL, Alberti JO, Mannar MGV, et al. Stability of iodine in iodized salt used for correction of iodine-deficiency disorders, II. *Food Nutr Bull*. 1998;19:240–250.
71. Burgi H, Schaffner T, Seiler JP. The toxicology of iodate: a review of the literature. *Thyroid*. 2001;11:449–456.
72. Engle PL, Black MM, Behrman JR, et al. Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet*. 2007;369:229–242.
73. Horton S. The economics of food fortification. *J Nutr*. 2006;136:1068–1071.
74. Le C, Richard SA, Rivera JA, et al. Stunting, wasting, and micronutrient deficiency disorders 2006. In: Jamison DT, Breman JG, Measham AR, Aet al., eds. *Disease Control Priorities in Developing Countries*. 2nd ed. New York: Oxford University Press; 2006:551–568.
75. Bhat S, Marklund M, Henry ME, et al. A systematic review of the sources of dietary salt around the world. *Adv Nutr*. 2020;11:677–686.
76. Spohrer R, Garrett GS, Timmer A, et al. Processed foods as an integral part of universal salt iodization programs: a review of global experience and analyses of Bangladesh and Pakistan. *Food Nutr Bull*. 2012;33:S272–S280.
77. Abizari AR, Dold S, Kupka R, et al. More than two-thirds of dietary iodine in children in northern Ghana is obtained from bouillon cubes containing iodized salt. *Publ Health Nutr*. 2017;20:1107–1113.
78. Campbell N, Dary O, Cappuccio FP, et al. Collaboration to optimize dietary intakes of salt and iodine: a critical but overlooked public health issue. *Bull World Health Organ*. 2012;90:73–74.
79. Untoro J, Schultink W, West CE, et al. Efficacy of oral iodized peanut oil is greater than that of iodized poppy seed oil among Indonesian schoolchildren. *Am J Clin Nutr*. 2006;84:1208–1214.
80. Delange F. Administration of iodized oil during pregnancy: a summary of the published evidence. *Bull World Health Organ*. 1996;74:101–108.

81. Todd CH, Dunn JT. Intermittent oral administration of potassium iodide solution for the correction of iodine deficiency. *Am J Clin Nutr*. 1998;67:1279–1283.
82. WHO Secretariat, Andersson M, de Benoist B, et al. Prevention and control of iodine deficiency in pregnant and lactating women and in children less than 2-years-old: conclusions and recommendations of the Technical Consultation. *Public Health Nutr*. 2007;10:1606–1611.
83. Bouhouch RR, Bouhouch S, Cherkaoui M, et al. Direct iodine supplementation of infants versus supplementation of their breastfeeding mothers: a double-blind, randomised, placebo-controlled trial. *Lancet Diabetes Endocrinol*. 2014;2:197–209.
84. Eng PHK, Cardona GR, Fang S-L, et al. Escape from the acute Wolff–Chaikoff effect is associated with a decrease in thyroid/iodide symporter mRNA and protein. *Endocrinology*. 1999;140:3404–3410.
85. Vagenakis AG, Braverman LE. Adverse effects of iodides on thyroid function. *Med Clin North Am*. 1975;59:1075–1088.
86. Fradkin JE, Wolff J. Iodide-induced thyrotoxicosis. *Medicine (Baltim)*. 1983;62:1–20.
87. Stanbury JB, Ermans AE, Bourdoux P, et al. Iodine-induced hyperthyroidism: occurrence and epidemiology. *Thyroid*. 1998;8:83–100.
88. Todd CH, Allain T, Gomo ZAR, et al. Increase in thyrotoxicosis associated with iodide supplements in Zimbabwe. *Lancet*. 1995;346:1563–1564.
89. Bulow Pedersen I, Knudsen N, Jorgensen T, et al. Large differences in incidences of overt hyper-and hypothyroidism associated with a small difference in iodide intake: a prospective comparative register-based population survey. *J Clin Endocrinol Metab*. 2002;87:4462–4469.
90. Møllehave LT, Linneberg A, Skaaby T, et al. Trends in treatments of thyroid disease following iodine fortification in Denmark: a nationwide register-based study. *Clin Epidemiol*. 2018;10:763–770.
91. European Food Safety Authority. *Tolerable Upper Intake Levels for Vitamins and Minerals*. Parma: EFSA; 2006.
92. World Health Organization & Food and Agriculture Organization of the United Nations. *Vitamin and Mineral Requirements in Human Nutrition. Geneva*; 2004.
93. World Health Organization. *Toxicological Evaluation of Certain Food Additives and Contaminants. WHO Food Additives Series 24. Prepared by: The 33rd Meeting of the Joint FAO/WHO Expert Committee on Food Additives (JECFA)*. Geneva: World Health Organization; 1989.
94. Zimmermann MB, Aeberli I, Andersson M, et al. Thyroglobulin is a sensitive measure of both deficient and excess iodine intakes in children and indicates no adverse effects on thyroid function in the UIC range of 100-299 mug/L: a UNICEF/ICCIDD study group report. *J Clin Endocrinol Metab*. 2013;98:1271–1280.
95. Rhee SS, Braverman LE, Pino S, et al. High iodine content of Korean seaweed soup: a health risk for lactating women and their infants? *Thyroid*. 2011;21:927–928.
96. Nagataki S. The average of dietary iodine intake due to the ingestion of seaweeds is 1.2 mg/day in Japan. *Thyroid*. 2008;18:667–668.
97. Parmar MS, Sturge C. Recurrent hamburger thyrotoxicosis. *CMAJ (Can Med Assoc J)*. 2003;169:415–417.
98. Shen H, Liu S, Sun D, et al. Geographical distribution of drinking-water with high iodine level and association between high iodine level in drinking-water and goitre: a Chinese national investigation. *Br J Nutr*. 2011;106:243–247.
99. Theodoropoulos T, Braverman LE, Vagenakis AG. Iodide-induced hypothyroidism: a potential hazard during perinatal life. *Science*. 1979;205:502–503.
100. Emder PJ, Marion M. Iodine-induced neonatal hypothyroidism secondary to maternal seaweed consumption: a common practice in some Asian cultures to promote breast milk supply. *J Paediatr Child Health*. 2011;47:750–752.

77

Euthyroid and Hyperthyroid Nodules and Goiter

Ralf Paschke, Markus Eszlinger, and Peter Kopp

OUTLINE

Basic Aspects, 1277
  Definition and Clinical Manifestation, 1277
  Environment and Heredity, 1277
  Molecular Mechanisms for Euthyroid, Cold, and Hot Thyroid Nodules, 1278
  Hot Thyroid Nodules, 1280
  Cold Thyroid Nodules, 1281
  Euthyroid Nodules, 1281
Familial and Sporadic Nonautoimmune Hyperthyroidism, 1282
Iodine-Induced Hyperthyroidism, 1282
Epidemiology, Occurrence, and Natural History of Euthyroid and Hot Thyroid Nodules, 1283
Thyroid Nodule Diagnostic Pathway to Distinguish Benign and Malignant Thyroid Nodules, 1284
  Clinical Examination, 1284
  Laboratory Investigations, 1285
  Diagnosis of Thyroid Autonomy, 1285
  Malignancy Risk Stratification using Thyroid Ultrasound, 1287
  Fine Needle Aspiration Cytology, 1288
  Molecular Diagnostics of Indeterminate Thyroid Nodule Cytologies, 1288
Treatment of Autonomous Thyroid Nodules, 1290
  Surgery, 1290
  Radioiodine Therapy, 1290
  Follow-Up, 1290
Treatment of Euthyroid Nodules, 1292
  Iodine Supplementation, 1292
  Thyroid Hormone Treatment, 1292
  Surgery, 1292
  Radioiodine Therapy, 1292
  Nonsurgical Thermal Ablation of Thyroid Nodules, 1293

BASIC ASPECTS

Definition and Clinical Manifestation

Benign nodular thyroid includes a heterogeneous spectrum of alterations that are highly prevalent in iodine-deficient areas. Very broadly, it can be divided into solitary nodular and multinodular thyroid disease. Histologically, benign thyroid nodules are distinguished by morphologic criteria according to the World Health Organization (WHO) classification as encapsulated lesions (true adenomas) or adenomatous nodules, which lack a capsule.[1] Functionally, nodules are classified as "cold," "normal," or "hot" depending on whether they show decreased, normal, or increased uptake, respectively, on scintigraphy. Approximately 50% to 85% of all nodules are "cold," up to 40% are scintigraphically indeterminate, and approximately 10% are "hot," although the relative frequencies will vary geographically as a function of iodine intake.

Clonal analysis of thyroid nodules indicates that most of them are "true" monoclonal neoplasias rather than polyclonal hyperplastic lesions. Strict histologic criteria for an adenoma and its differentiation from hyperplastic thyroid nodules or adenomatous nodules (without a capsule) are difficult. Therefore, distinction of hyperplastic thyroid nodules from neoplasias should, ideally, take into account their clonal composition.[1] Because many thyroid nodules without a capsule (adenomatous nodules) are monoclonal, a mixed functional and molecular definition of true thyroid tumors, as outlined in Fig. 77.1, appears more objective and consistent. In contrast to solitary nodular thyroid disease, which has a more uniform clinical, pathologic, and molecular picture, multinodular goiter (MNG) and toxic multinodular goiter (TMNG) are more complex entities. Functionally, they often include a mixture of hyper-, hypo-, or normally functioning thyroid lesions within the same gland. The overall balance of functional properties of individual thyroid nodules within a multinodular goiter ultimately determines the functional status in the individual patient, who may be euthyroid, be subclinically hyperthyroid, or have overt hyperthyroidism. The term MNG is applied in the first scenario, while TMNG refers to the latter situations. It is important to emphasize that this functional characterization is not stationary, and hyperthyroidism may develop over time, because patients with TMNG usually have a history of longstanding MNG.[2] In general, development of MNG proceeds in two phases: global activation of thyroid epithelial cell proliferation (e.g., as the result of iodine deficiency or other goitrogenic stimuli) leading to goiter formation, and a focal increase of thyroid epithelial cell proliferation causing thyroid nodules. So far, the most common stimulus for focal proliferation is a somatic mutation, as detailed in the following section.

Environment and Heredity

Historically, nontoxic goiter was mainly regarded as a consequence of iodine deficiency. However, a number of goitrogens and cigarette smoking are important environmental risk factors in the etiology of nontoxic goiter.[3] The impact of smoking, most likely mediated by thiocyanate, which can competitively inhibit iodide transport into the thyroid, has been extensively studied.[3] Additional important etiological factors include sex,[4] age,[5] and increased body mass index (BMI).[5] The effect of a certain goitrogen is influenced by the degree of iodine intake and varies regionally and interindividually. However, it is most likely

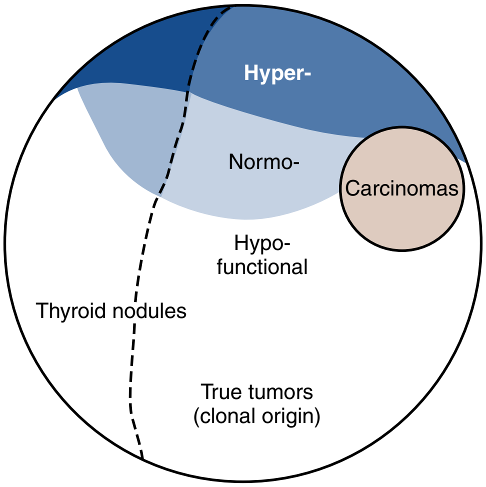


Fig. 77.1 Clonality, histologic, and scintigraphic classification of thyroid tumors, and their respective overlap. Clonality studies imply that the majority of thyroid nodules are "true" thyroid tumors. Polyclonal hyperplastic nodules lack a capsule and a growth pattern that is different from the surrounding normal parenchyma. The clonality and histologic characteristics of benign thyroid nodules overlap differentially with the scintigraphic characteristics, whereas primary thyroid carcinomas are mostly clonal tumors with mostly hypofunctional appearance on scintigraphy.

that interactions between environmental factors and individual genetic determinants ultimately determine goiter development.[5-7] Nontoxic goiter appearing early in life, often clustering in families, suggests strong genetic susceptibility, whereas the environmental determinants are more likely to have additive or triggering effects.

In contrast to sporadic goiters, most cases of familial goiter present an autosomal dominant pattern of inheritance, indicating an important genetic component.[7] Gene–gene interactions or various polygenic mechanisms (i.e., synergistic effects of several variants or polymorphisms) are likely to contribute to the complex pathogenesis of nontoxic goiter and also offer an explanation for its genetic heterogeneity. A strong genetic predisposition is indicated by family and twin studies. Children of parents with goiter have a significantly higher risk for developing goiter compared with children of nongoitrous parents. The high incidence in females and the higher concordance in monozygotic than in dizygotic twins also suggest a genetic predisposition. Moreover, there is preliminary evidence of a positive family history for thyroid diseases in those who have postoperative relapse of goiter, which can occur months to years after surgery.

The development of nontoxic goiter is most likely a continuous process that starts with thyroid hyperplasia. Therefore, defects in genes that play an important role in thyroid physiology and thyroid hormone synthesis could predispose to goiter development, especially in the case of borderline or overt iodine deficiency. Such defects may be associated with partial dyshormonogenesis and the hyperplasia, as well as the subsequent nodular transformation, may reflect an adaptive response to meet the requirements for hormone synthesis. Genes that encode proteins involved in thyroid hormone synthesis (sodium iodide symporter [*SLC5A5*], thyroblobuline [*TG*], thyroid peroxidase [*TPO*], Pendred syndrome gene [*SLC26A4*], solute carrier SLC26A7 [*SLC26A7*], iodotyrosine deiodinase [*DEHAL 1*], dual oxidase 2 and dual oxidase maturation factor [*DUOX2, DUOXA2*], thyroid-stimulating hormone [TSH] receptor [TSHR]) are convincing candidate genes in familial euthyroid goiter. Originally, several mutations in these genes were identified in patients with congenital hypothyroidism. However, in cases of less severe functional impairment, which can still be functionally compensated, a contribution of variants of these genes in the etiology of nontoxic goiter is possible. Moreover, individuals with biallelic mutations in the *TG* or *SLC26* gene can, for example, develop nontoxic goiters.

Linkage analyses have been performed to identify novel susceptibility loci to gain further insights into the genetic mechanisms that contribute to the etiology of nontoxic familial goiter. Linkage studies have identified several candidate loci (*MNG*, *MNG2*, 2q, 3p, 7q, and 8p). Remarkably, for the first time, a more prevalent putative risk locus, present in 20% of the families investigated, has been identified.[8] The recognition of genetic loci conferring susceptibility for goiter development, together with established environmental factors, indicates that nontoxic goiter can be defined as a complex multifactorial disease.

Whereas linkage studies are suitable for the detection of candidate genes with a strong effect, it is possible to miss weak genetic defects of first-line candidate gene variants or of novel genes, which are more likely detected by association studies. A genome-wide association study identified the *FGF7* and the *CAPZB* genes, which are involved in the regulation of thyroid epithelial cell proliferation and colloid engulfment during cellular uptake of thyroglobulin as susceptibility loci for goiter.[9] One to two percent of adult-onset thyroid nodules harbor a hotspot *DICER1* mutation, and three quarters of these mutations are accompanied by a second hit, presumably with loss of function in *DICER1*. Some of these loss of function variants are probably germline in origin, thus representing previously undiagnosed *DICER1* syndrome patients.[10] Importantly, it is conceivable that the sum of several low-penetrance alleles in different genomic regions could lead to goiter predisposition. Therefore, the widely accepted risk factors such as iodine deficiency, smoking, old age, and female sex are likely to interact with and/or trigger the genetic susceptibility. Further narrowing down the candidate regions and performing association studies with single nucleotide polymorphisms in additional families, especially in case-control association studies, and the confirmation of association study results in other populations are necessary to further corroborate the role of currently identified candidate genes for hereditary nontoxic goiter.

Molecular Mechanisms for Euthyroid, Cold, and Hot Thyroid Nodules

Most goiters become nodular with time. Thyroid nodules can present as solitary lesion or as multinodular thyroid or goiter. Functionally, they cover a wide spectrum from autonomous hyperfunctioning adenomas, so-called toxic adenomas (TAs), to nodules that have lost the ability to concentrate iodine, and hence, produce thyroid hormones. From animal models of hyperplasia caused by iodine depletion, it has become apparent that there is an increase in functional activity, as well as a tremendous increase in thyroid cell number. These two events may be associated with an increased rate of mutational events. Thyroid hormone synthesis requires the generation of hydrogen peroxide (H2O2), which may damage genomic DNA and cause mutations. Besides being important as an oxidizing agent for iodine during hormone synthesis, H2O2 can be a major source of free radicals and reactive oxygen species (ROS). Because these molecules can cause substantial damage to a cell and impair normal function, thyroid epithelial cells are likely to have a potent defense mechanism to counterbalance potential damage mediated by free radicals. It has been shown that antioxidant enzymes, such as glutathione peroxidases (GPXs) or thyroid peroxidase, are upregulated during thyroid hormone synthesis. GPX3 has also been suggested to directly interfere with thyroid hormone synthesis by affecting the concentration of H2O2. Together with a higher spontaneous mutation rate, a higher replication rate will more often prevent mutation repair and increase the mutation load of the thyroid, thereby randomly

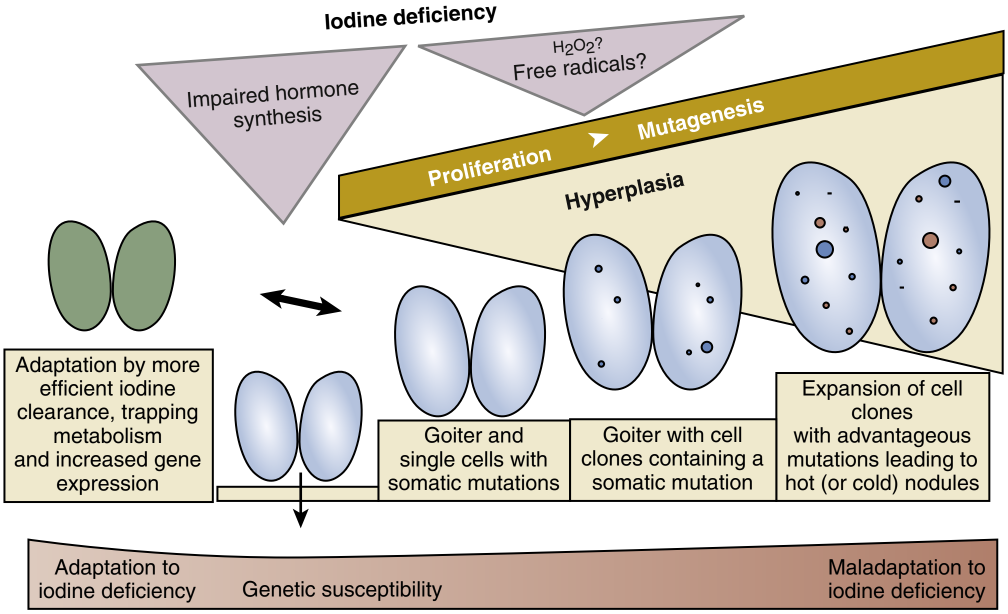

Iodine deficiency
Impaired hormone synthesis
H_2O_2?
Free radicals?
Proliferation
Mutagenesis
Hyperplasia
Adaptation by more efficient iodine clearance, trapping metabolism and increased gene expression
Goiter and single cells with somatic mutations
Goiter with cell clones containing a somatic mutation
Expansion of cell clones with advantageous mutations leading to hot (or cold) nodules
Adaptation to iodine deficiency
Genetic susceptibility
Maladaptation to iodine deficiency

Fig. 77.2 Hypothesis for thyroid nodular transformation. The starting point for the development of multinodular goiter is hyperplasia induced by goitrogenic stimuli (e.g., iodine deficiency). Iodine deficiency increases mutagenesis directly (production of hydrogen peroxide/free radicals) or indirectly (proliferation and increased number of cell divisions). Subsequently, hyperplasia forms cell clones. Some of them contain somatic mutations of the thyroid-stimulating hormone receptor leading to autonomously functioning thyroid nodules *(red circles)* or contain mutations that lead to dedifferentiation and therefore cold thyroid nodules or cold adenomas *(blue circles)*.

affecting genes crucial for thyrocyte physiology. Mutations that confer a growth advantage (e.g., activating TSHR or Gsα mutations) very likely initiate focal growth. Hence, autonomously functioning thyroid nodules (AFTNs) are thought to develop from small cell clones that contain advantageous somatic mutations as shown for the TSHR in “hot” microscopic regions of euthyroid goiters.[11]

However, thyroid nodules also occur in thyroid glands of normal size. Interestingly, a recent study reported that the growth of established thyroid nodules is a process limited to the thyroid nodules, but not associated with growth of the surrounding nonnodular thyroid tissue.[12] This condition should be distinguished from nodular goiter and is better characterized as a uni- or multinodular thyroid gland.

Epidemiologic studies, animal models, and molecular/genetic data outline a general theory of nodular transformation. Based on the identification of somatic mutations and the predominant clonal origin of AFTNs and cold thyroid nodules, the following sequence of events could lead to thyroid nodular transformation in three steps, as outlined in Fig. 77.2.[5] First, iodine deficiency, nutritional goitrogens, or autoimmunity cause diffuse thyroid hyperplasia. Then, at this stage of thyroid hyperplasia, increased proliferation, together with possible DNA damage associated with H2O2 generation, causes a higher mutation load (i.e., a higher number of cells bearing mutations). Some of these spontaneous mutations confer constitutive activation of the cyclic adenosine monophosphate (cAMP) cascade (e.g., TSHR and Gsα mutations), which stimulates growth and function. Finally, in a proliferating thyroid, growth factor expression (e.g., insulin-like growth factor 1 [IGF-1], transforming growth factor β1 [TGF-β1], or epidermal growth factor [EGF]) is increased. As a result of growth factor costimulation, most cells divide and form small clones. After increased growth factor expression ceases, small clones with activating mutations will further proliferate if they maintain an intrinsic proliferative advantage. They can thus form small foci, which develop into thyroid nodules. This mechanism could explain AFTNs by advantageous mutations that both initiate growth and function of the affected thyroid cells, as well as cold thyroid nodules by mutations that stimulate proliferation only (e.g., RAS mutations or other mutations in the RAS/RAF/MEK/ERK/MAP cascade).

Because of the slow proliferation rate of thyroid epithelial cells, a long period (decades) between initiation of the neoplasia and appearance of MNG seems likely. Therefore, the prevalence of thyroid neoplasias can be potentially explained by a high frequency of tumor initiation and/or enhanced thyroid epithelial cell proliferation. The origin of tumor formation in the thyroid could therefore be induced by a high mutation rate and aberrant growth stimulation in the thyroid gland. The latter very likely involves endogenous growth factors and/or exogenous goitrogenic substances.

In addition to the generally high mutation rate in the normal thyroid gland, environmental and lifestyle factors could add to the pool of DNA damage and increase mutagenesis and tumor initiation. For example, smoking could lead to intracellular iodine deficiency through thiocyanate-induced blocking of the iodine transport into thyrocytes.[13]

As outlined earlier, H2O2-mediated ROS generation is very likely to be an important starting point for thyroid tumor development (Fig. 77.3). Because H2O2 is essential for the oxidation of iodide in thyroid hormone synthesis, changes in iodine concentration are associated with the cellular H2O2 concentrations. In fact, generation of H2O2 is inhibited by iodide *in vivo* and *in vitro*. Moreover, H2O2 generation is stimulated by higher concentrations of TSH (predominantly by the protein kinase C pathway rather than the protein kinase A pathway). Low iodine levels and markers of increased thyroid functionality suggest activation of H2O2 generation, which could result in DNA damage and somatic mutations. Low iodine and high H2O2 levels should therefore activate antioxidative defense, which should be detectable in the cellular regulation of enzymes involved in defense against oxidative stress.

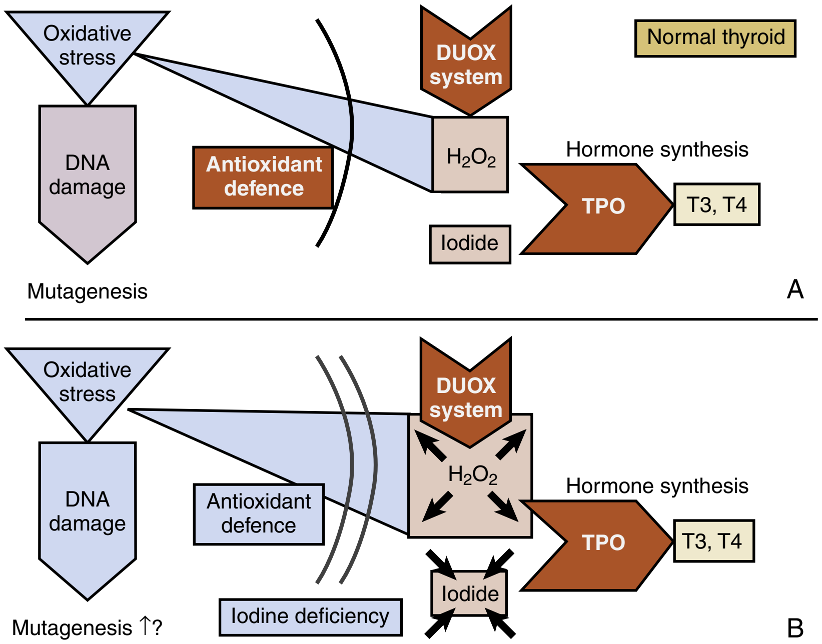


Fig. 77.3 Mechanisms that may cause mutagenesis in the thyroid gland. The figure shows the key molecules involved in those parts of thyroid hormone synthesis, which—in conditions of iodine and most likely also selenium deficiency—lead to oxidative stress, DNA damage, and possibly mutagenesis. **A,** In the normal thyroid gland, the enzymes DUOX1 and DUOX2 generate H2O2, and TPO transfers oxidized iodine to tyrosyl residues of thyroglobulin, the precursor for T3 and T4 synthesis. H2O2 is, however, a source of ROS, which—with other oxidative stress—can cause DNA damage. Normally, antioxidant defense could prevent oxidative stress and DNA damage. Selenoproteins like glutathion peroxidase 3 are part of the defense. **B,** Conditions of iodine deficiency increase levels of H2O2 and may increase the amount of oxidative stress and DNA damage. Selenium deficiency decreases selenoproteins and could thereby weaken antioxidative defense, which exacerbates oxidative stress and DNA damage. *ROS,* Reactive oxygen species; *DUOX system,* thyroid oxidase system; *TPO,* thyroid peroxidase; *H2O2*, hydrogen peroxidase; *T3*, triiodothyronine; *T4*, thyroxine..

Differential expression of additional antioxidant enzymes illustrates the importance of the antioxidant defense in the iodine-deficient thyroid gland.[14] Of note, glutathione peroxidases are selenium proteins. Selenium deficiency can further impair the antioxidant defense and thereby exacerbate oxidative stress (see Fig. 77.3).

In summary, the early molecular prerequisites for nodular/neoplastic transformation in the thyroid gland appear to consist of a sequence of molecular events that include oxidative stress and DNA damage as the trigger for somatic mutations. The oxidative burden is already detectable in the normal thyroid gland and is very likely to be linked to hormone synthesis and H2O2 production. Additionally, environmental conditions (e.g., iodine deficiency) have the potential to deteriorate this situation. In general, any external factor (e.g., smoking) that increases oxidative stress, causes DNA damage, or increases the somatic mutational rate aggravates the risk for tumorigenesis. Also, any factor that increases proliferation (e.g., goitrogens) can accelerate the development of a thyroid tumor.

Hot Thyroid Nodules

A TA is a monoclonal AFTN that synthesizes thyroxine (T4) and/or triiodothyronine (T3), resulting in eventual suppression of serum TSH and in subclinical or overt hyperthyroidism.[15] AFTNs harbor somatic gain-of-function mutations in the TSHR,[16] and the mutations may differ among TAs within the same multinodular goiter. Thyroid dysfunction is readily diagnosed with thyroid function tests, and, in the case of evidence for excessive hormone production, zones of autonomy can be detected through scintigraphy, if indicated. Somatic point mutations that constitutively activate the TSHR explain the clinical phenotype of hot thyroid nodules through their activation of the G_s or the G_s and $G_{q/11}$ pathways, resulting in a stimulation of proliferation, differentiation, and increased thyroid hormone synthesis. However, in different studies the prevalence of TSHR and Gs α mutations in hot thyroid nodules has been reported to vary from 8% to 82% and 8% to 75%, respectively. Available studies differ in the extent of mutation detection and the screening methods. A comprehensive study using the more sensitive denaturing gradient gel electrophoresis technique revealed a frequency of 57% TSHR mutations and 3% GNAS (the gene coding for the Gs α protein) mutations in 75 consecutive hot thyroid nodules.[17] These results raise the question of the molecular etiology of TSHR and GNAS mutation-negative nodules. In order to determine whether other gene(s) may be responsible for the remaining TSHR and GNAS mutation negative cases of hot thyroid nodules, whole-exome sequencing (WES) was performed for 13 TSHR and GNAS mutation-negative samples.[18] While no further causative mutation could be identified, WES revealed a recurrent hot spot mutation (c.1712A>G; p.Gln571Arg) in enhancer of zeste homolog 1 (*EZH1*). Targeted screening in an independent cohort confirmed a high frequency (27%) of this mutation in hot thyroid nodules. *EZH1* mutations were strongly associated with known (*TSHR, GNAS*) or presumed (*ADCY9*) alterations in genes involved in the cAMP pathway. Functional studies revealed that the *EZH1* mutation causes increased histone H3 methylation, as well as increased proliferation of thyroid cells. Hence, this study revealed a hot spot mutation in *EZH1* as the second most frequent genetic alteration in hot thyroid nodules.[18] The association between *EZH1* and *TSHR* mutations suggests at least a two-hit model for the pathogenesis of these tumors, whereby constitutive activation of the cAMP pathway and *EZH1* mutations cooperate to induce hyperproliferation of thyroid cells. Because WES did not reveal a further causative mutation for hot thyroid nodules, a very sensitive and comprehensive targeted next-generation sequencing panel has been developed that covers the whole coding sequence of the *TSHR,* as well as the *GNAS* hot spot loci c. 201 and c. 227.[19] This panel allowed further evaluation of the impact of *TSHR* and *GNAS* mutations on the molecular etiology of hot thyroid nodules in a large cohort of hot thyroid nodules.[20,21] Due to the higher sensitivity of the targeted next-generation sequencing panel compared with previously used methodologies, *TSHR* or *GNAS* mutations could be detected in 85% of hot thyroid nodules. Moreover, in an optimal sample set with multiple samples per nodule, the prevalence of *TSHR* and *GNAS* mutations rose to 96%. These results, in addition to the fact that no further causative driver mutations explaining the phenotype of hot thyroid nodules could be identified by previous WES, strongly support the hypothesis that *TSHR* and *GNAS* mutations are presumably the only somatic mutations leading to hot thyroid nodules.[19]

In addition to the intracellular signaling network that is connected to the TSHR, the extracellular action of different growth factors enhances the complexity of the signal flux into the thyroid cell. Growth factors like IGF-1, EGF, TGF-β1, and fibroblast growth factor stimulate growth and dedifferentiation of thyroid epithelial cells. Studies focused on insulin and IGF show a permissive effect of insulin and IGF-1 on TSH signaling and a cooperative interaction of TSH and insulin/IGF-1. Other studies suggest inactivation of TGF-β1 signaling in hot thyroid nodules due to constitutively activated TSHR (e.g., resulting from *TSHR* mutations).[22] This assumption is supported by the finding of a decreased expression of TGF-β1 messenger RNA (mRNA) after TSH stimulation of thyrocytes. Because TGF-β1 has been shown to inhibit iodine uptake, iodine organification, and thyroglobulin expression,[23] as well as cell proliferation in different cell culture systems, these findings suggest that inactivation of TGF-β1 signaling is a major prerequisite for increased proliferation in hot thyroid nodules.[24] Signal modulation of the TSHR that appears to define the etiology of hot thyroid nodules and the clinical phenotype could therefore take part at a number of stages and very likely involves genetic/epigenetic, sex-related, and environmental factors.

In vivo characteristics of TA include increased iodide uptake and thyroid peroxidase mRNA and protein content, as well as increased thyroid hormone secretion.[25] Upregulation of sodium-iodide symporter (NIS) and pendrin contributing to the augmented iodine-trapping characteristics of TA have been reported by other investigators.[25] Basal inositol phosphate generation was not found to be increased in toxic thyroid nodules, compatible with a predominant constitutive activation of the cAMP–protein kinase A cascade through TSHR or Gs_{α} protein mutations in TA and TMNG.[24] Stimulation of growth has been investigated in TA using the proliferation markers Ki67 and proliferating cell nuclear antigen.[25] However, the increase in the proliferation index in TA was found to be relatively low (2- to 3-fold) compared with the surrounding normal thyroid. These data are in agreement with the slow progression seen in TA and TMNG; however, they are difficult to reconcile with the evolution of a macroscopic TA from a single mutated cell (estimated number of 30 cell divisions to yield a nodule volume of ~1 mL),[24] and thus may more likely reflect the end stage rather than the initiation of disease.

Cold Thyroid Nodules

The term *cold* indicates reduced uptake of iodide or pertechnetate on scintigraphy. Because a histologic diagnosis is typically employed to exclude thyroid malignancy, many investigations of thyroid nodules only refer to the histologic diagnosis of thyroid adenoma. This histologic entity should not be confounded with the scintigraphically characterized entity *cold nodule,* which, like hot or warm (scintigraphically normofunctional) thyroid nodules can histologically appear as thyroid adenomas or adenomatous nodules according to the WHO classification. In contrast, focal hyperplasia is not very well explained at the molecular level and has been discussed in detail elsewhere as the cause of thyroid tumors.[26] A monoclonal origin has been detected for the majority of cold thyroid nodules, which implies nodular development from a single mutated thyroid cell (Fig. 77.1).[27]

With reference to their functional status (i.e., reduced iodine uptake), failure in the iodide transport system or failure of the organic binding of iodide had been detected as functional aberrations of cold thyroid nodules long before the molecular components of the iodine metabolism were known. Subsequently, a decreased expression or membrane localization of the NIS and, more likely, a defective cell membrane targeting of the NIS protein in thyroid carcinomas and benign cold thyroid nodules were suggested as the molecular mechanism underlying the failure of the iodide transport or uptake in cold thyroid nodules (reviewed in[28]).

A study of 40 cold thyroid adenomas and adenomatous nodules detected *RAS* mutations in only a single case.[27] Moreover, in the same set of cold thyroid nodules, no point mutations in the mutational hot spots of the *BRAF* gene were detected.[29] This is in line with the lack of *BRAF* mutations in benign follicular adenomas in other studies.[30]

Increased expression of histone mRNAs and of cell cycle–associated genes like *cyclin D1, cyclin H/cyclin dependent kinase (CDK) 7,* and *cyclin B* most likely reflect a molecular background that is permissive for increased cellular proliferation in cold thyroid nodules.[31] In line with the low prevalence of *RAS* mutations in cold thyroid nodules,[27] a reduced expression of RAS-MAPK cascade-associated genes was found, which may suggest a minor importance of this signaling cascade.

Euthyroid Nodules

Most of the literature on the molecular etiology of euthyroid nodules is focused on cold thyroid nodules. However, scintigraphically cold nodules are difficult to differentiate from scintigraphically isocaptant nodules.[32] It is therefore important to note that the vast literature on benign thyroid nodules comprises scintigraphically cold and scintigraphically isocaptant nodules, which make up approximately 50% to 85% and up to 40% of all nodules, respectively. *RAS* mutations are the most prevalent mutations reported in benign scintigraphically cold/isocaptant nodules (mostly designated as follicular adenoma), with a prevalence of 20% to 25%. In addition to benign thyroid nodules, *RAS* mutations have been described in approximately 30% to 45% of follicular thyroid carcinomas, 30% to 45% of follicular variant papillary thyroid carcinomas, 20% to 40% of poorly differentiated thyroid carcinomas, and 20% to 30% of anaplastic thyroid carcinomas.[33] The finding of *RAS* mutations in benign thyroid nodules, as well as in follicular patterned thyroid carcinomas such as follicular thyroid and follicular variant papillary thyroid carcinomas, suggests that *RAS*-positive benign nodules (follicular adenomas) could serve as a precursor for *RAS*-positive carcinomas.[33] However, additional genetic alterations such as *PTEN* deletions or *PIK3CA* mutations[34] are required to transform a benign thyroid nodule into a follicular thyroid carcinoma. Similarly to *RAS* mutations, the *PAX8/PPARG* fusion oncogene has been reported in both benign thyroid nodules and follicular thyroid carcinomas.[35] Thus, the same diagnostic problem, i.e., a large variation in the prevalence of for *PAX8/PPARG* gene fusions in benign tumors and malignant tumors, is present, akin to the scenario for *RAS* mutations. A comprehensive analysis of benign thyroid nodules based on whole-exome and transcriptome sequencing revealed mutually exclusive SPOP^{P94R}, EZH1^{Q571R}, and ZNF148 mutations in 24% of benign nodules.[36] However, a further study showed that the mutational profiles of benign thyroid nodules (follicular adenomas) and minimally invasive follicular carcinomas are very similar, and no mutational cluster could be identified separating the two tumor entities. Moreover, the comparison of gene expression profiles of benign thyroid nodules (follicular adenoma) and minimally invasive follicular carcinoma confirmed that these two neoplasias do not have significant transcriptional differences.[37] Finally, there is high interobserver variation for the respective histologic diagnoses where capsular and/or vascular invasion is the only criterion that allows one to differentiate follicular thyroid carcinoma from benign thyroid nodules.

KEY POINTS

- Iodine deficiency leads to increased oxidative stress. Somatic mutations induced by oxidative stress are a likely cause for benign clonal thyroid nodules.
- The etiology of multinodular goiter is based on a complex interaction between genetic susceptibility and environmental factors.

FAMILIAL AND SPORADIC NONAUTOIMMUNE HYPERTHYROIDISM

Autosomal dominant familial nonautoimmune hyperthyroidism was recognized as a separate entity in 1982 by Thomas et al.[38] At that time, the underlying etiology was unknown. After the cloning of the TSHR, it then became apparent that gain-of-function mutations can lead to several pathological entities: 1) somatic mutations can cause TAs[39] (see also earlier discussion); 2) germline mutations can be inherited in an autosomal dominant fashion and lead to familial nonautoimmune hyperthyroidism[40]; and 3) *de novo* germline mutations can cause sporadic nonautoimmune hyperthyroidism.[41] In contrast, biallelic loss-of-function mutations lead to compensated or overt hypothyroidism because of resistance to TSH.[42]

Patients with activating germline mutations in the TSHR typically have a goiter, but the characteristic signs associated with autoimmune hyperthyroidism, i.e., the presence of antibodies against the TSHR, endocrine ophthalmopathy, myxedema, and lymphocytic infiltration of the thyroid gland, are absent.[38] In the inherited form, the family history reveals familial clustering consistent with an autosomal disorder. In these families, the age of onset of hyperthyroidism is variable among affected individuals and depends, in part, on the activity of the mutated allele, as well as on other factors such as iodine intake and genetic background. In infants with neonatal hyperthyroidism, it is important to distinguish the more frequent and transient autoimmune form due to transplacental passage of stimulating TSHR autoantibodies from a mother with a history of Graves disease from sporadic nonautoimmune hyperthyroidism.[16] The latter patients tend to have germline gain-of-function mutations of the TSHR[43,44]; congenital hyperthyroidism due to a TA harboring a somatic *TSHR* mutation has been reported as an unusual variant.[16] Importantly, several of the children with severe neonatal hyperthyroidism presented with mild mental retardation. This could be due to a direct effect of excessive thyroid hormone levels on brain development or premature closure of the cranial sutures, which can involve exclusively the squamosal sutures.[41]

A definitive diagnosis of nonautoimmune hyperthyroidism can now be easily confirmed by direct sequencing of the coding region of the *TSHR* gene. The known mutations are accessible in the continuously updated TSHR database that includes more than 400 gain-of-function mutations and close to 200 loss-of-function mutations (tsh-receptor-mutation-database.org).[44] So far, 29 families and 20 children with sporadic occurrence of *TSHR* germline mutations have been reported, mainly in Europe.[44] Thyroid ablation is advocated as first-line treatment (surgery and/or radioiodine) for preventing relapses. Molecular analysis for germline *TSHR* mutations offers the possibility for family screening, preclinical diagnosis, and genetic counseling.[45] If the mutation is unknown, functional *in vitro* analyses are needed to demonstrate that the mutated allele confers constitutive activity to the receptor.

The therapy of patients with nonautoimmune hyperthyroidism can be challenging. The entity is not always readily recognized, and can be confused with Graves disease. This may result in inadequate therapeutic control and frequent relapses in patients treated with thionamides or insufficient amounts of radioiodine. For a definitive cure, it is paramount to destroy all thyroid tissue, either by thyroidectomy followed by radioiodine therapy, or by radioiodine therapy alone. Particularly in infants and small children, the use of thyreostatic therapy is a valid option before considering a definitive therapeutic modality later in life.

IODINE-INDUCED HYPERTHYROIDISM

In individuals with underlying thyroid disease, excessive dietary iodine intake or the administration of iodine-containing compounds can result in iodine-induced thyrotoxicosis (IIT) or hypothyroidism.[15,46] IIT is also referred to as the *Jod-Basedow* phenomenon (Jod meaning iodine in German; Karl von Basedow, the German physician who described the signs of thyrotoxicosis associated with exophthalmos and goiter, i.e. Graves disease). Of note, Graves disease is typically referred to as Basedow disease on the European continent. The condition was reported as early as 1821 by Coindet, who realized that the administration of potassium iodide can lead to reduction in goiter volume but may induce toxic effects with a clinical presentation that reflects IIT.

For adults, the recommended daily allowance for iodine is 150 mcg per day. The tolerable upper intake level for adults has been set to 1100 mcg/day.[47] Normally, the thyroid is protected against iodine excess by the *Wolff–Chaikoff* effect, which results in an immediate reduction in iodide uptake, organification, and thyroid hormone biosynthesis and secretion.[48] Impressively, most individuals with a normal thyroid gland tolerate even a chronic excess of 30 mg up to 2 g iodide per day without clinical symptoms. However, the autoregulatory mechanisms mediated by the Wolff–Chaikoff effect are no longer functional or are impaired in diseased thyroid tissue, in which exposure to even modest amounts of excessive iodine can induce IIT in susceptible individuals. IIT is particularly prevalent in patients with multinodular goiter that contain nodules with an autonomy of function, but can also occur in individuals with autoimmune thyroid disease. In regions with a history of iodine deficiency, where multinodular goiters are particularly prevalent, the implementation of iodination programs has therefore often been associated with an increase in IIT.[15]

The possible sources of excess iodine or iodide are innumerable and include nutrients, radiological contrast agents, disinfectants, expectorants, and drugs.[15]

A particularly common problem in clinical practice consists in amiodarone-induced thyrotoxicosis (AIT) (see Chapter 84).[49] By weight, amiodarone consists of 37% iodine; hence, a 200 mg tablet contains approximately 75 mg iodine. AIT is subdivided into two distinct entities: AIT1 is caused by iodine-induced hyperthyroidism in patients with nodular goiters or underlying Graves disease, and AIT2 is caused by a destructive thyroiditis resulting in the release of thyroid hormones. Mixed forms of AIT1 and AIT2 do exist, and it can be difficult to distinguish the two forms. With improved nutritional iodine intake, the prevalence of AIT2 seems to increase relative to AIT1. AIT1 is treated with thionamides (sometimes in combination with perchlorate in countries where the drug is available), and AIT2 is treated with glucocorticoids. If the etiologic mechanism cannot be clearly defined, which is frequently the case in clinical practice, thionamides are frequently combined with glucocorticoid therapy. In patients with AIT that is difficult to control, thyroidectomy may be indicated.

Radiographic iodinated contrast media (ICM) are a common source of high amounts of iodine, with a typical iodine content of 320 to 370 mg/mL.[47] In a retrospective case-control study evaluating the association between ICM use and thyroid dysfunction, the risk for developing either hyperthyroidism or hypothyroidism after exposure has been estimated to be increased by 2- to 3-fold.[50] In contrast, in a systematic review/metaanalysis, the overall estimated prevalence of overt hyperthyroidism after ICM exposure was extremely low (0.1%, 95% confidence interval: 0–0.6), suggesting that the concern for IIT is often overestimated.[51] Symptomatic patients can be treated transiently with a beta-blocker and, if needed, thionamides. In patients at high risk for thyrotoxicosis, prophylactic administration of thionamides can be considered before the administration of ICM for the imaging study.[47]

In TA and TMNG, a sudden increase in function can be induced by the administration of excessive amounts of iodine, e.g., in the form of contrast media used for angiography and computed tomography (CT) scans or by iodine-containing drugs, e.g., amiodarone. In the European Study Group of Thyrotoxicosis, iodine contamination was found in 36.8% of patients from iodine-deficient areas with thyrotoxicosis.[52] Severity of iodine deficiency, autonomous thyroid cell mass, and older age have been proposed as risk factors for the development of iodine-induced thyrotoxicosis (see Chapter 84).

EPIDEMIOLOGY, OCCURRENCE, AND NATURAL HISTORY OF EUTHYROID AND HOT THYROID NODULES

The natural history of multinodular goiter, with respect to goiter growth and function, varies and is difficult to predict in a given patient. The spontaneous growth rate in selected populations has been estimated to be up to 20% yearly[53] but is usually much lower. No specific parameter exists that can predict the growth potential of multinodular goiter, which can be accurately assessed by serial yearly measurements of the size of the goiter and individual nodules by ultrasonography.

Painful nodules are usually the result of hemorrhage into a nodule or a cyst in the goiter. The diagnosis is readily made by ultrasonographic examination and fine-needle aspiration cytology (FNAC). Multinodular goiter is not usually associated with a significantly increased risk for the development of thyroid malignancy. The risk for malignancy in thyroid nodules occurring within a multinodular goiter has not been completely clarified, but most authors find a similar frequency in uni- and multinodular goiters, of approximately 5%.

Thyroid autonomy is widely viewed as a manifestation of benign disease, and current guidelines recommend that hot nodules in patients with suppressed thyrotropin levels be excluded from further malignancy risk stratification. However, 61 case reports indicating the presence of thyroid carcinoma within a hot nodule were published from 1972 to 2018. Detailed pathologic analyses of reported “toxic thyroid cancer” cases (0.7%–6% in different series) have shown that, if classified correctly, the cancer (a) usually occurred coincidentally within the same or even in the contralateral thyroid lobe, (b) mostly represented a papillary thyroid microcarcinoma (<1 cm), and (c) was rarely located within the hot nodule itself. A recent study demonstrated equivalent malignancy rates in hyperthyroid hot nodules and euthyroid hot nodules, with an overall malignancy rate of 8.5%.[54] Furthermore, a recent animal study observed the development of large papillary thyroid cancer in nearly all mice with a constitutively active TSHR mutation knock-in, the most common etiology of hot nodules and other nonautoimmune hyperthyroidism.[55] Eighteen cases of hyperfunctioning thyroid carcinomas with somatic constitutively activating TSH mutations have been reported, highlighting the need for elucidating the underlying mechanism.[44]

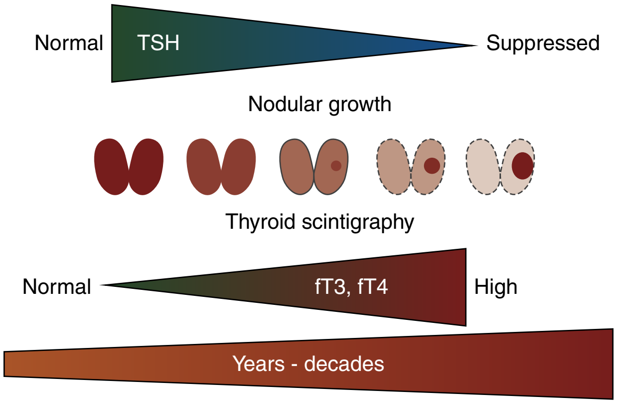


Fig. 77.4 Evolution of a toxic nodule from a scintigraphically undetectable clone of thyroid cells with a somatic thyroid-stimulating hormone (TSH) receptor mutation in a euthyroid thyroid gland to a hot nodule with increasing suppression of TSH.

Patients with nontoxic multinodular goiter can become hyperthyroid or, less commonly, hypothyroid. Hyperthyroidism in such patients often develops insidiously, in contrast to that associated withs Graves disease. It often begins with a prolonged period of subclinical hyperthyroidism characterized by low serum TSH and normal serum free T4 and T3 concentrations. This hyperthyroid state is the consequence of goiter growth and an associated increase in the mass of autonomously hormone-producing thyroid cells. Long-term studies on patients with TA have shown that the natural course tends to be slow, requiring years or decades (see Figs. 77.2 and 77.4). Sandrock et al. reported an overall 4.1% annual incidence of thyrotoxicosis in a group of 375 untreated euthyroid TA patients in Germany who were followed for a mean of 53 months.[56] Hamburger highlighted the correlation between nodule size and development of thyrotoxicosis in a group of 349 patients with TA: 93.5% of patients with overt thyrotoxicosis had TA greater than 3 cm in size, and patients with a euthyroid TA of greater than 3 cm in size had a 20% risk of developing thyrotoxicosis during a 6-year follow-up period, as opposed to a 2% to 5% risk in patients with nodules less than 2.5 cm in size (see Figs. 77.2 and 77.4).[57]

Data on progression from euthyroid to subclinical and overt thyrotoxicosis in patients with multinodular goiter are scarce. An incidence of 9% to 10% of overt thyrotoxicosis was proposed in two studies during a follow-up period of 7 to 12 years.[13,57,58]

In a large population-based cross-sectional study in an iodine-deficient area, nodular autonomy increased with age and reached 15% in older adults. It appears from a few longitudinal studies that, within 5 years, hyperthyroidism will emerge in approximately 10% of patients with a nodular goiter.

Iodine deficiency is still the most frequent single cause of multinodular endemic goiter worldwide (see Chapter 76). Considerable regional variation exists even in nonendemic goiter areas. Illustrating the influence of iodine intake on the epidemiology of sporadic goiter, 31 of 423 (7.3%) 68-year-olds had goiter in Jutland, Denmark (low iodine intake area), versus 2 of 100 (2%) in Reykjavik, Iceland (high iodine intake area).[59] In the Danish population, the goiter prevalence decreased 4 years after iodization of salt, while the prevalence of thyroid nodules remained unchanged. As with nontoxic goiter, iodine intake is of paramount importance for the occurrence of hyperthyroidism. In Denmark, a country with a borderline sufficient iodine intake, multinodular toxic goiter accounts for 44% of patients with hyperthyroidism,[59] whereas in Iceland, with a high iodine intake, there is a lower proportion of multinodular goiter (6%) and more cases of Graves disease.[59]

Thyroid nodules and cystic lesions are extremely common in the general population; their prevalence increases with age, female sex, and adiposity.[60-62] When the thyroid is evaluated through neck palpation, the prevalence is approximately 5% in iodine-sufficient regions. However, the widespread use of neck sonography and other imaging techniques such as CT, magnetic resonance imaging, and positron emission tomography-CT (PET-CT) has shown that their prevalence is as high as 65%.[63,64] Many of these *thyroid incidentalomas* are small and benign and have no further clinical relevance. Hence, the challenge consists in identifying the nodules that are associated with thyroid dysfunction, clinically significant thyroid carcinomas, or nodules that result in the compression of adjacent structures, while avoiding unnecessary testing and surveillance that result in significant economic and personal burden.

The vast majority (90%–95%) of thyroid nodules are benign, although this frequency is influenced by the clinical presentation, personal medical and family history, and the mode of detection. Risk factors for malignancy include radiation exposure,[60] a family history of thyroid cancer, and certain hereditary syndromes such as Cowden syndrome, Gardner syndrome, familial adenomatous polyposis, and, for medullary thyroid cancer, multiple endocrine neoplasia type 2.[60] In nodules detected by PET-CT, increased standardized uptake values correlate with a higher risk of malignancy[65].

The characterization of nodules associated with a normal thyroid function leans heavily on the use of ultrasound, which enables prediction of malignancy risk.[64,66-68] Although nodule growth is often considered as a risk factor for malignancy, it is a parameter with poor predictive value.[68]

Screening for thyroid cancer is not recommended in asymptomatic individuals.[69] Active screening, together with the advent of novel imaging modalities, has contributed to the excessive detection of large number of clinically irrelevant small papillary thyroid carcinomas[70] and to the phenomenon of *overdiagnosis,* resulting in an epidemic of thyroid cancer.[71]

THYROID NODULE DIAGNOSTIC PATHWAY TO DISTINGUISH BENIGN AND MALIGNANT THYROID NODULES

Higher levels of healthcare access, both sociodemographic and age-based, are associated with higher papillary thyroid cancer incidence rates. However, despite the 3-fold increase in incidence over three decades, the mortality rate has remained unchanged. Thus, together with the large subclinical reservoir of occult papillary thyroid cancers, there is evidence for the widespread overdiagnosis of papillary thyroid cancers.[72] The further workup of thyroid nodules after clinical and laboratory investigations will not only be determined by the outcome of the clinical and laboratory investigations, but also, most importantly, by local healthcare settings. The applicability and relevance of published evidence and best practice summarized in the respective guidelines[66,73,74] and in the chapters on imaging and ultrasound (see Chapter 69) and on FNAC (see Chapter 70), vary and depend, in part, on the framework of the healthcare system for thyroid nodule and thyroid cancer diagnostics. It is best to organize and optimize thyroid nodule and thyroid cancer diagnostics as a pathway. The performance and outcomes of each step in this pathway will have impact on all further steps. Deficiencies for one of the diagnostic steps or expertise cannot be compensated by (over)emphasizing any other step or expertise. For example, insufficient ultrasound malignancy risk stratification or inconsistent terminology for the description of ultrasound findings will result in higher FNAC rates with subsequent increases in false positives for FNAC and low malignancy yield for diagnostic lobectomies. Insufficient cytology expertise or nonadherence to cytology classifications cannot be compensated for by molecular diagnostics. Lack of preoperative FNAC application will lead to low yield of malignancies after diagnostic lobectomies, with unnecessary surgical complications and postoperative thyroid hormone replacement.[75] The initial requirement for a thyroid nodule and thyroid cancer diagnostic pathway is the characterization of a nodule using an ultrasound-based risk stratification scheme.[76] If indicated, this is followed by a biopsy and a cytological evaluation adhering to an internationally accepted convention (i.e., the Bethesda system).[77] Based on the cytology results and the clinical presentation, the indication for or against surgery can then be established. The representation of the respective medical expertise by different medical disciplines and the availability of respective quality insurance data will depend on the level of heathcare (primary, secondary, or tertiary) and the respective healthcare system with its respective incentives and restrictions, and most importantly its respective different degrees of integration or fragmentation. Also, different medical training and specialization systems will determine if the ultrasound expertise is mostly restricted to radiology or if fellowship training for cytopathology with resulting broad cytology expertise is available or not, and so on.

Because most thyroid nodules are incidental findings on imaging studies, implementation of a structured pathway is key for thyroid nodule workup.

The available clinical endocrine, ultrasound, cytology, and surgical expertise, their extent of integration in a diagnostic pathway, the implementation and healthcare quality insurance of the diagnostic pathway (Fig. 77.5A), and its integration with the primary care setting (Fig. 77.5B) will determine the diagnostic outcomes and the quality of care for patients with thyroid nodules or thyroid cancer. It is therefore self-evident that local process quality and local diagnostic outcomes and the local quality of care for patients with thyroid nodules can only be determined by local healthcare quality insurance data. Most of the evaluations of thyroid nodule ultrasound malignancy risk stratification systems or determinations of malignancy rates for the different Bethesda thyroid nodule FNAC categories were performed in select secondary or tertiary care settings with high-level FNAC operator and cytology expertise, and may not necessarily be applicable to a nonspecialized environment.[78]

Clinical Examination

The majority of patients with one or several thyroid nodules are asymptomatic. The medical history may reveal compressive symptoms such as dysphagia, dysphonia, hoarseness, respiratory complaints, or stridor. Pertinent clinical signs and symptoms are listed in Table 77.1, and diagnostic aspects are summarized in Table 77.2. Occasionally, hemorrhage into a nodule or cyst can lead to acute compressive symptoms. Pain can occur in patients with inflammatory or infectious processes in the thyroid. The Pemberton sign, defined by facial congestion and respiratory difficulties after elevating both arms, is suggestive for compression of the thoracic inlet. It can be associated with multinodular goiters with extension into the upper mediastinum, as well as with other masses causing a superior vena cava syndrome.[79]

The clinical examination includes inspection of the neck for the presence of a goiter and palpation of the thyroid and the neck for the detection of nodules and cervical lymph nodes. Although palpation is useful for the identification of larger nodules and tenderness (e.g., in the context of a [sub]acute thyroiditis), it has a low sensitivity and specificity. Many nodules are relatively small and cannot be detected if located in the posterior parts of the gland. Similarly, many pathological lymph nodes escape detection by palpation. The widespread availability of ultrasound, which provides detailed structural information of the thyroid and allows stratifying the malignancy risk of thyroid nodules and lateral neck lymph nodes,[64,66,67] has fundamentally transformed clinical practice. Clinical features in a patient with thyroid autonomy can be attributed to (a) symptoms of thyrotoxicosis and (b) the nodular goiter.

The clinical presentation of thyrotoxicosis varies with age, and in the elderly it is often not symptomatic ("apathetic hyperthyroidism").[13] In a series of 84 French patients with overt thyrotoxicosis, atrial fibrillation and anorexia dominated in the older age group (≥70 years), while classical signs of thyrotoxicosis, e.g., nervousness, weight loss despite increased appetite, palpitations, tremor, and heat intolerance were more frequently observed in younger patients (≤50 years).[80] Alternatively, a patient may present with a lump or disfigurement of the neck, intolerance of a tight collar, or increase in collar size. Dysphagia or breathing difficulties due to esophageal or tracheal compression may be apparent, especially with large TMNG.[80]

Subclinical thyrotoxicosis, defined by low or suppressed serum TSH with normal free T4 and free T3 levels is more commonly observed in older patients with TMNG.[2,80] Moreover, the incidental finding of low or suppressed TSH levels on routine investigation for other indications is frequently the first evidence for the presence of thyroid autonomy, particularly in regions with iodine deficiency. Subclinical thyrotoxicosis may be associated with atrial fibrillation and a reduced bone density.

Laboratory Investigations

Thyroid function should be determined in all patients with structural abnormalities of the thyroid, including thyroid nodules.[66] The initial step consists in the measurement of a serum TSH concentration. In patients with suppressed TSH levels, the peripheral hormones (T4 and T3) should also be determined. If subclinical or overt hyperthyroidism is confirmed, thyroid scintigraphy can be performed in order to detect AFTNs. In patients with elevated TSH levels, the measurement of T4 levels and thyroid autoantibodies permit diagnosis of autoimmune thyroid disease. Of note, Hashimoto thyroiditis typically has a very characteristic pattern on ultrasonography.

The measurement of thyroglobulin is not recommended in the evaluation of patients with nodular thyroid disease because it does not allow discrimination between benign and malignant conditions.[66] In patients with an established diagnosis of thyroid cancer originating from thyroid follicular cells, very high preoperative thyroglobulin levels may suggest locoregional or distant metastases, but thyroglobulin is not measured routinely by all experts in this setting.[66]

Calcitonin is a marker for medullary thyroid cancer. Although some experts suggest that measuring calcitonin is useful in the workup of thyroid nodules, this approach remains controversial.[81] Major guidelines do not support measuring calcitonin unless there is a family history for medullary thyroid cancer or suspicious cytologic characteristics because of the low yield and an unfavorable cost-benefit ratio.[66]

Diagnosis of Thyroid Autonomy

Diagnosis of thyroid autonomy is based on three characteristics:

1. Confirmation of clinically suspected thyrotoxicosis by an abnormal thyroid function test. Typically, there is overt thyrotoxicosis (low TSH, high free thyroid hormones). However, dependent on the autonomous cell mass, subclinical thyrotoxicosis (TSH low, free T3 and free T4 normal) or even euthyroidism may be present.
2. Presence of palpable or sonographic nodule(s).
3. Increased radionuclide uptake in the nodule(s) concomitant with a decreased uptake in the surrounding extranodular thyroid tissue. The most commonly used isotopes are ^{99}mTc, ^{123}I, and ^{131}I, which are transported into the thyrocytes by the NIS, but only iodine is organified. Several studies have revealed a comparable diagnostic value of the these isotopes. However, discrepancies due to trapping only nodules can be found in approximately 5% of hot nodules, mostly with normal serum TSH. The advantage of ^{99}mTc, which is more frequently used in Europe, is its low cost.[82]

If thyroid autonomy is suspected in an euthyroid patient, a "suppression" scan can be performed after administration of

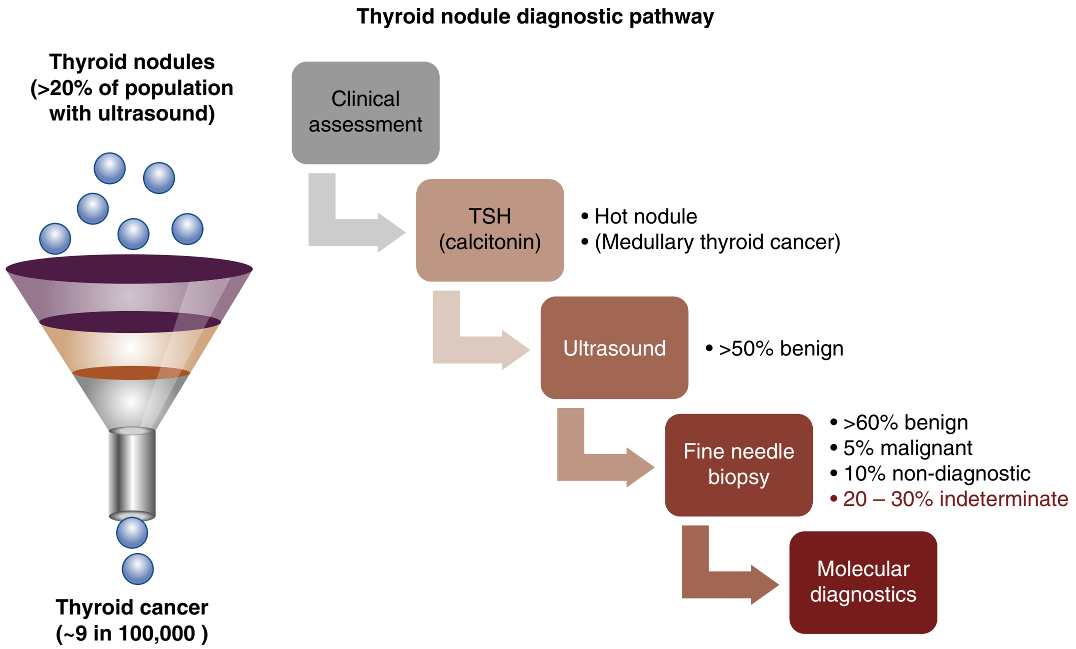
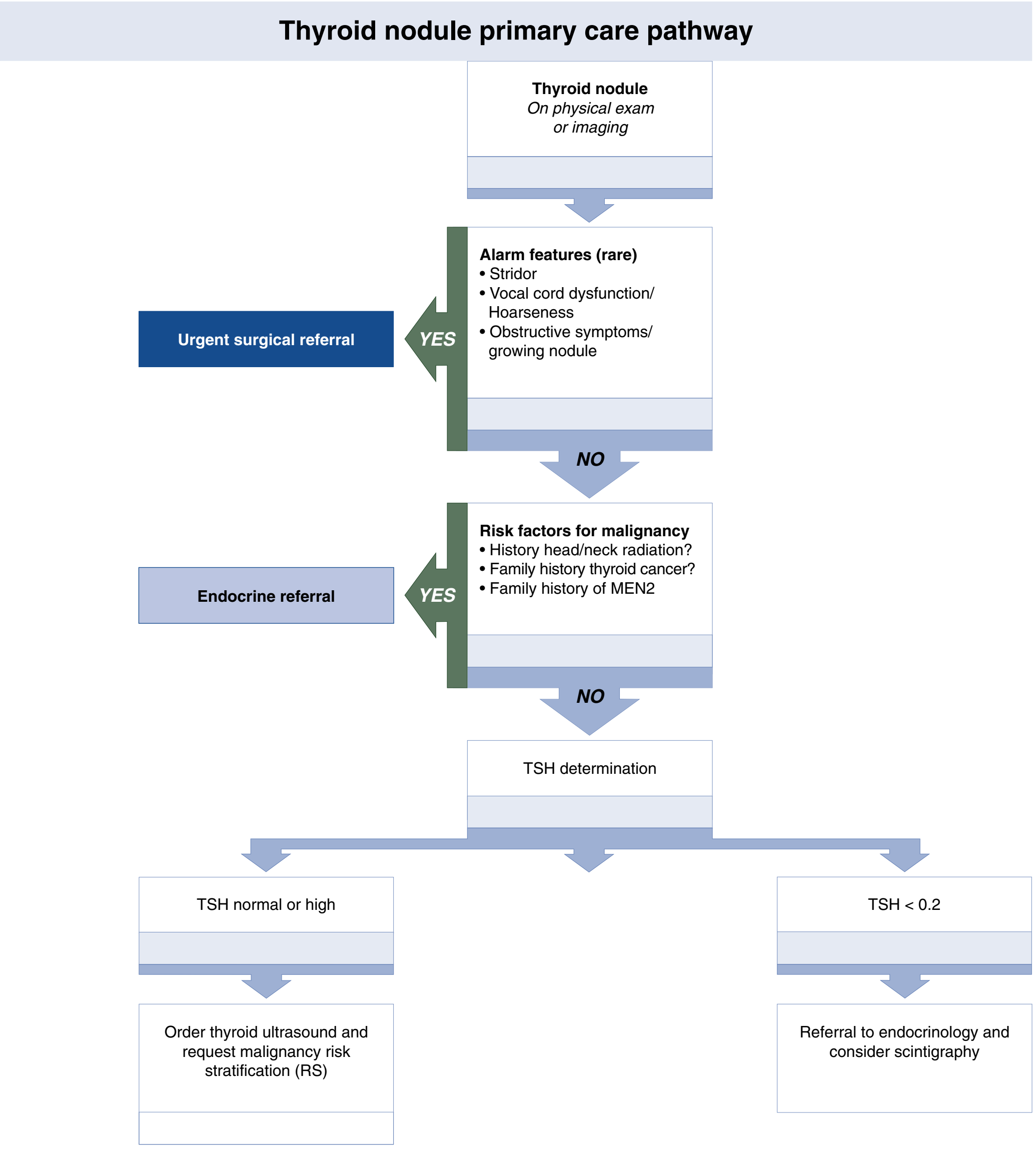


Fig. 77.5 A, Diagnostic pathway for thyroid nodules.

Continued


Thyroid nodule primary care pathway
Thyroid nodule
On physical exam or imaging
Alarm features (rare)
• Stridor
• Vocal cord dysfunction/ Hoarseness
• Obstructive symptoms/ growing nodule
YES
Urgent surgical referral
NO
Risk factors for malignancy
• History head/neck radiation?
• Family history thyroid cancer?
• Family history of MEN2
YES
Endocrine referral
NO
TSH determination
TSH normal or high
TSH < 0.2
Order thyroid ultrasound and request malignancy risk stratification (RS)
Referral to endocrinology and consider scintigraphy

Further management as outlined by malignancy risk stratification

ATA U/S Pattern	ATA risk of cancer	Recommended action
High suspicion	> 70% - 90%	> 1 cm: Biopsy + Endocrinology referral < 1 cm: Endocrinology referral
Intermediate suspicion	10% - 20%	> 1 cm: Biopsy + Endocrinology referral < 1 cm: U/S follow-up in 1 - 2 years
Low suspicion	5% - 10%	> 1.5 cm: Endocrinology referral < 1.5 cm: U/S follow-up in 1 - 2 years
Very low suspicion	< 3%	Clinical follow-up and U/S in 2 years

Table 6, figure 2 in reference 76

Fig. 77.5, cont'd B, Pathway for the workup of thyroid nodules.

thyroid hormone (i.e., 75 µg/day T4 for 2 weeks followed by 150 µg/day for 2 weeks), after which uptake in all nonautonomous tissue will be suppressed, and thyroid autonomy is unmasked. Measurement of thyroid autoantibodies is not routinely performed in TA and TMNG. However, in iodine-deficient areas, distinction between Graves disease and TMNG can be difficult if extrathyroidal manifestations of the former are absent and diagnostic findings are "atypical," e.g., presence of thyroid nodules (~27%–34% by ultrasonography) in suspected Graves disease patients and patchy rather than diffuse uptake on scintigraphy. In these conditions, determination of TSHR antibodies is helpful to establish the correct diagnosis.[83] Urinary iodine excretion can be measured in case

TABLE 77.1 Clinical Signs and Symptoms of Multinodular Goiter

- Slowly growing nodular anterior neck mass
- Tracheal deviation or compression, upper airway obstruction, dyspnea, dysphonia
- Occasional cough and dysphagia, globus sensation
- Sudden pain or enlargement secondary to hemorrhage
- Superior vena cava obstruction syndrome
- Pemberton's sign: obstruction of the thoracic inlet by extending the arms over the head
- Gradually developing hyperthyroidism
- Iodide-induced thyrotoxicosis
- Recurrent nerve palsy (rare)
- Phrenic nerve palsy (rare)
- Horner syndrome (rare)
- Enlargement during pregnancy

TABLE 77.2 Diagnosis of Multinodular Goiter

- Multinodularity on examination
- Asymmetry, tracheal deviation
- Thyroid-stimulating hormone normal or decreased, free thyroxine and free triiodothyronine normal or increased, thyroglobulin elevated
- Calcitonin normal
- Thyroid antibodies negative in approximately 90%
- Scintigraphy with hot and cold areas
- Ultrasound finding of nodularity (nonhomogeneity); cysts and calcifications are common
- Computed tomography, magnetic resonance, positron emission tomography/computed tomography imaging demonstrating a nonhomogeneous mass
- Lung function testing may demonstrate impaired inspiratory capacity
- Benign cytology by fine-needle aspiration of dominant nodules

of suspected iodine contamination. CT and magnetic resonance imaging techniques are not routinely indicated for the diagnosis of thyroid autonomy, but may be used for presurgical planning in patients with large and partly intrathoracic goiters.

Scintigraphy. The American Thyroid Association (ATA) guidelines recommend that a radionuclide thyroid scan should be performed if the serum TSH is subnormal.[66] However, the American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME) thyroid nodule guidelines recommend performing thyroid scintigraphy only in the presence of a thyroid nodule or MNG in combination with a suppressed TSH level, or when ectopic thyroid tissue or a retrosternal goiter is suspected. In iodine-deficient regions, thyroid scintigraphy is recommended to exclude autonomy of a thyroid nodule or MNG even when the TSH level is low-normal.[73] The latter is supported by a meta-analysis of eight studies with 2761 hot thyroid nodules that reported a pooled prevalence of hot nodules with normal TSH levels of 50%.[84] Moreover, thyroid nodules are not only classified as either "cold," or "hot," but may also show "normal" uptake on scintigraphy. In Europe, approximately 50% to 85% of all nodules are "cold," up to 40% are scintigraphically indifferent, and approximately 10% are "hot."[85,86] Moreover, there is some geographic variation that is influenced by iodine intake, as well as the clinical setting. Smaller thyroid nodules (<1 cm) may not be detectable by planar scintigraphy, and hot thyroid nodules in the posterior part of the thyroid may be obscured by overlying normal tissue, thus requiring single-photon emission CT analysis for their detection. The goal of scintigraphy is not only to identify a hot nodule, but also to exclude hot nodules from further malignancy risk stratification based on their very low risk for malignancy.

Malignancy Risk Stratification using Thyroid Ultrasound

Malignancy risk stratification of thyroid nodules with ultrasound requires a structured evaluation and consistent description of all pertinent characteristics using a common dictionary of adjectives followed by a stratified interpretation.[87] This is essential for a meaningful use of any of the currently available Thyroid Imaging Reporting and Data Systems (TI-RADS).[64,66,67] The European Thyroid Association and the American College of Radiology have developed specific lexicons for their risk-assessment systems. Definitions of nodule characteristics like extrathyroidal extension, echogenicity, shape, or microcalcifications can significantly impact the diagnostic performance of the respective classification system. The interobserver agreement for ultrasound-based risk-stratification systems is only fair to moderate (kappa, 0.34–0.44). Specific training is required to increase the operators' ability to recognize and describe these features.[88] An overview of the pertinent nodule characteristics evaluated by such systems, FNAC indications, and follow-up recommendations of four international practice guidelines is summarized in Chapter 69. The nodule characteristics considered in the risk scores are very similar for these systems—structural composition, echogenicity, shape, margins, and echogenic foci—and risk classes are defined by two or more nodule features. The exception to this is the American College of Radiology (ACR) TI-RADS classification, which assesses the same characteristics individually.[67] Each is rated with a numerical score, and the sum of the five scores determines the risk class to which the nodule is assigned. There is some degree of heterogeneity across systems regarding the definitions of certain nodule features, the relative weight assigned to individual features, and the size criteria used for FNAC recommendations. For each risk class, size cutoffs are used to identify lesions whose FNAC can safely be deferred. Prospective evaluations of the recommendations by the British Thyroid Association, the ATA, and the AACE/ACE/AME concluded that the respective classification systems had similar predictive values of malignancy in high-risk classes. The ATA and AACE/ACE/AME systems were effective for ruling out the need for FNAC in ultrasound low-risk nodules. Substantial interobserver agreement was noted for the respective classifications.[76] A further prospective evaluation of the ATA, AACE/ACE/AME, and ACR TI-RADS systems concluded that the ACR TI-RADS classification system has the highest area under the receiver operating characteristic curve for the identification of cytological high-risk nodules. The ATA classification left "unclassified" some nodules with a relatively high risk of malignancy.[89] There is also substantial variability of the ability to decrease the number of unnecessary FNACs in the different stratification systems. In a prospective study, the ACR TI-RADS outperformed four other stratification systems for the reduction of the number of biopsies performed on nodules later diagnosed as benign.[90] The ACR system classified more than one half of the FNACs as deferrable, with a false-negative rate of only 2.2%. All these evaluations were performed at referral centers; thus, the validity of these results in the primary care setting needs further validation. Moreover, it is important to consider that the thyroid nodule malignancy risk stratification systems were developed mainly for the detection of papillary thyroid cancers. Their ability to detect follicular thyroid or medullary thyroid carcinomas requires further study. Current efforts focus on establishing an integrated international thyroid nodule ultrasound malignancy risk classification system based on a validated lexicon and incorporating state-of-the-art techniques,[91] as well as electronic algorithmic tools[92] and machine learning approaches (artificial intelligence) with the goal of reducing the number of unnecessary FNAC evaluations. Of note, depending on the respective referral setting and the operator training, approximately half of the thyroid nodules evaluated with one of the thyroid nodule ultrasound malignancy risk stratification systems can be characterized as very likely benign, with a very low 0.3% false-negative rate.[64] The remaining thyroid nodules with suspicious ultrasound malignancy characteristics should be subjected to FNAC based on stratified guideline-based recommendations.

Fine Needle Aspiration Cytology

FNAC is performed under ultrasound guidance by the following two commonly used techniques. In the parallel technique, the needle enters adjacent to the probe footprint along its long axis and is visualized from the skin entry point as it advances into the thyroid nodule. For the perpendicular technique, the needle is inserted perpendicular to the probe footprint with an angle determined by the thyroid nodule depth. Only the tip of the needle is visualized when it enters the nodule. The FNAC sample can be obtained either by applying gentle suction or without suction, the so-called capillary technique. The further preparation of the cytology material may either use the classic smearing on glass slides or liquid preparations, for which the choice is made together with the respective cytopathologist. Complications of FNAC are exceedingly rare. FNAC as the next step in the thyroid nodule diagnostic pathway should be reserved for thyroid nodules with a FNAC indication according to a stratified thyroid nodule ultrasound malignancy risk assessment (see Chapter 69) or clinical risk factors. Like the thyroid nodule ultrasound results, the FNAC results play a key role in the differential diagnostic pathway for the further workup and therapy of thyroid nodules.

The updated version of the Bethesda System for Reporting Thyroid Cytopathology (BSRTC) provides a robust diagnostic framework and guidance for the evaluation of the cytologic material.[77] Its recommended diagnostic categories are summarized in Chapter 70. Whereas the BSRTC provides evidence-based guidance for the malignancy risks in the respective cytology classes summarized in Chapter 70, the published variability for the percentages of assignment of the six cytology categories and their respective malignancy risks vary greatly. A metaanalysis reported cytologically benign results for 39% to 74% of the nodules tested and found that 2% to 16% of thyroid FNAC samples are cytologically malignant. An atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) result was observed in 1% to 27%, a follicular neoplasm/suspicious for follicular neoplasm (FN/SFN) result in 1% to 25%, a suspicious-for-malignancy (SFM) result in 1% to 6%, and a nondiagnostic result in 2% to 24% of samples.[93] Due to the high interobserver variations in AUS/FLUS and FN/SFN diagnoses, consensus cytopathology is an effective way to reduce discrepant or indeterminate FNAC results. Recently, because of its indolent nature, it was proposed that the entity "noninvasive encapsulated follicular variant of papillary thyroid cancer" be reclassified as "noninvasive follicular thyroid neoplasm with papillary-like nuclear features" (NIFTP).[94] Of note, the diagnosis of NIFTP is established on histological grounds. The risk of malignancy in most BSRTC categories, most significantly the AUS/FLUS, FN/SFN, and SFM categories (see Chapter 70), will thus depend on the local prevalence of NIFTP. Most importantly, determination of local malignancy risks for the different BSRTC classes and the respective percentages for their local use is required by local evaluation and local quality insurance investigation, which will also depend on the local practice setting and resulting local prevalence data.[78] According to the BSRTC, most thyroid nodules subjected to FNAC are benign based on cytology and have an overall low malignancy risk that is less than 5%. Follow-up for an average of 8.5 years for initially benign thyroid nodule FNACs results in a false-negative rate of 1% and confers negligible mortality risk for these false positives during long-term follow-up. Because these malignancies were adequately treated despite detection at a mean 4.5 years after a falsely benign cytology, these data support a recommendation for repeat thyroid nodule evaluation 2 to 4 years after the initial benign FNAC, depending on ultrasound malignancy risk assessment. However, local outcome and quality insurance data are key for the application of this recommendation. Most papillary thyroid cancers and half of medullary thyroid cancers (MTC) can be detected by FNAC. Unfortunately, FNAC was only used in one fifth of patients undergoing surgery for a single thyroid nodule in Germany,[95] and similarly low frequencies of preoperative FNAC were reported in Belgium and France.[75] The deficiency that only half of MTCs are detected by FNAC[96,97] including the use of calcitonin immunocytology can be compensated for by preoperative calcitonin determination in suspicious cases without a family history for multiple endocrine neoplasia type 2 and using assay-, sex-, age-, and BMI-specific reference values for calcitonin assays.[98] The preoperative MTC diagnosis is important for enabling the correct first-line surgery with central compartment lymph node dissection in respective cases according to current guidelines.[99] Moreover, thyroid nodule FNAC has the inherent limitation of not being able to distinguish between benign thyroid adenomas or adenomatous nodules and follicular thyroid carcinomas (FTCs), because this differential diagnosis relies on the histologic criteria of capsular and/or vascular invasion. Therefore, around 20% of FNACs are indeterminate and classified as AUS/FLUS or FN/SFN with malignancy risks of 10% to 30% and 25% to 40% according to BSRTC (Chapter 70). For AUS/FLUS or FN/SFN, a repeat FNAC can be useful to clarify the cytologic diagnosis. However, a second FNAC provides a definitive diagnosis for only 40% of nondiagnostic and AUS/FLUS cytologies.[100] Moreover, if the risk stratification of nodules is well performed by ultrasound, the utility of this approach may be further reduced. Traditionally, diagnostic lobectomy has been required to allow a definitive histologic diagnosis for indeterminate thyroid nodules. This is obviously associated with risks and cost, and if the histology demonstrates the presence of a malignant nodule after lobectomy, this may require subsequent completion thyroidectomy. Histologic outcomes of patients stratified by FNAC for a diagnostic lobectomy suggest that up to 60% of patients were over- or undertreated with an initial surgery. Forty-two percent had potentially been overtreated with a hemithyroidectomy for benign disease when no other indication was recorded. Conversely, 17% had potentially been undertreated by the initial surgery, with the need for completion thyroidectomy.[101] Therefore, molecular diagnostics of indeterminate (AUS/FLUS and FN/SFN) FNAC has been recommended,[66,73] but these tests are currently not available or covered in many healthcare systems.

Molecular Diagnostics of Indeterminate Thyroid Nodule Cytologies

No molecular diagnostic test can replace or resolve deficiencies of a multidisciplinary approach to the management of patients with thyroid nodules. As part of an integrated approach of careful clinical, ultrasound, and FNAC assessment with local outcome data, molecular testing may be able to improve diagnostic outcomes for thyroid nodules by identifying patients with an indeterminate FNAC as most likely benign with less than 5% false negatives, thus obviating diagnostic

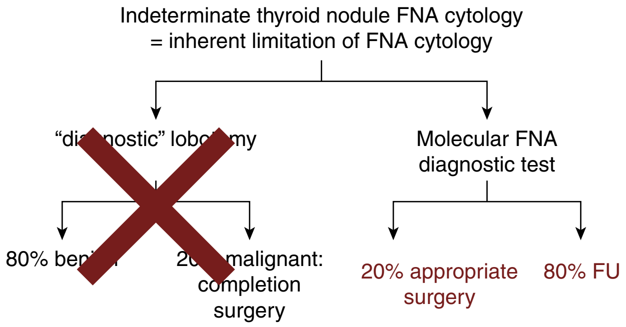


Fig. 77.6 **Diagnostic impact of molecular diagnostics of indeterminate thyroid nodule cytology as part of a diagnostic pathway.** A molecular fine-needle aspiration cytology test with high positive predictive value and >95% negative predictive value based on local data may replace diagnostic lobectomy for Bethesda System for Reporting Thyroid Cytopathology atypia of undetermined significance/follicular lesion of undetermined significance or follicular neoplasm/suspicious for follicular neoplasm cytology by appropriate surgery (lobectomy for mutations with low malignancy risk mutations or total thyroidectomy for mutations with high malignancy risk mutations) or follow-up.

lobectomy, or by identifying patients with malignant lesions or a high likelihood for malignancy for total thyroidectomy as definitive surgical treatment (Fig. 77.6).

Currently available molecular FNAC tests are based on testing for somatic mutations, gene expression evaluation, and microRNA (miRNA)-based classifiers.[102-107] The ThyroSeq test (version 3) uses targeted next-generation sequencing analysis of 112 cancer-related genes to identify point mutations, gene fusions, copy number alterations, or abnormal gene expression. Validation in 257 cytologically indeterminate nodules with surgical pathology reported a sensitivity of 94% and a specificity of 82%, with a negative predictive value of 97% and a positive predictive value of 66%. The test may eliminate the need for diagnostic surgery in up to 61% of patients with indeterminate nodules.[103] The Afirma Genomic Sequencing Classifier uses RNA sequencing and consists of 12 classifiers composed of 10,196 genes. It has a reported specificity of 68% and a positive predictive value of 47% with a sensitivity of 91% and a negative predictive value of 96%.[103] It is expected to reduce the frequency of diagnostic surgeries.[108] Thus, the ThyroSeq and Afirma assays reported positive and negative predictive values that make them suitable for use in both rule-in and rule-out testing. The validation studies for both tests reported sufficient sensitivities and specificities to be used as rule-in or rule-out tests for indeterminate FNACs. However, the validation study cohort for the Afirma assay was not representative of the populations in whom the assay has been used, thus calling into question its reported diagnostic performance, including its negative predictive value.[109] In the case of availability of similar postmarketing studies for the ThyroSeq test, a similar problem would likely also surface for the ThyroSeq test, because the validation study was mostly performed in tertiary referral centers. These findings emphasize once again the requirement for local evaluation of test performance as part of a diagnostic pathway. Currently, the further remaining problems are a lack of independent validation studies, a lack of long-term outcome studies for ThyroSeq and Afirma benign tests, and the high cost that currently limits their application outside the United States. Although longer follow-up is needed, recent postvalidation and clinical experience studies have demonstrated improved performance of both tests for indeterminate FNAC with Hürthle cell-dominant cytology,[110] in spite of the predominance of widespread chromosomal losses and mitochondrial DNA alterations in Hürthle cell carcinomas.[111,112]

ThyroPrint, a gene expression classifier based on interrogation of only 10 miRNAs, has also displayed good performance (sensitivity 96%, specificity 87%, and positive and negative predictive values of 78% and 98%, respectively).[113] Promising results have also been reported for a combination assay using miRNA classification (ThyraMIR) and next-generation sequencing mutation analysis (ThyGeNEXT) (positive predictive value 74%, negative predictive value 94%).[104] However, the assay has yet to be subjected to independent validation. ThyroSPEC is a cost-efficient matrix-assisted laser desorption/ionization–time of flight mass spectrometry-based mutation detection panel that detects the most prevalent 117 point mutations and 23 gene fusions in thyroid cancer.[107] All molecular diagnostic test except the ThyroSPEC molecular test require additional liquid FNAC material generated by dedicated FNAC passes in preserving solutions, whereas ThyroSPEC analyzes the residual liquid material or the FNAC material of a representative FNAC glass slide smear after routine FNAC.[107]

A systematic review of studies examining *RAS* mutations reported a range from 0% to 48% for the prevalence of *RAS* mutation in benign lesions, 0% to 68% for *RET/PTC* rearrangements, and 0% to 55% for *PAX8/PPARG* rearrangements in benign lesions.[114] Therefore, the presence of these biomarkers and the tremendous variation in reports of their prevalence in benign lesions suggests the need for caution when including these markers in diagnostic decisions. Further characterization of the importance of these markers, as well as newly discovered markers of thyroid malignancy, is required in order to avoid overtreatment of patients with benign thyroid tumors. However, the interpretation of these results is further complicated by the large interobserver for differentiating adenomatous nodules, thyroid adenomas, and minimally invasive FTCs.[84] This high rate of histopathological discordance hampers translational studies assessing the value of molecular markers that aim to improve presurgical diagnosis, because the histological reference for these studies can be ambiguous or discrepant. Therefore, it is likely that the current histology-based classification of thyroid tumors[1] will ultimately be complemented by molecular/genetic characteristics for a more accurate classification, akin to other endocrine tumors or other cancers.

The cost-effectiveness of molecular FNAC tests varies dramatically depending on the healthcare setting (for example, in a multidisciplinary tertiary care setting vs. a primary care setting), the cost of the test, the cost of a lobectomy, the risk of malignancy for the respective FNAC diagnosis, and the sensitivity, specificity, and negative and positive predictive values and accuracy and impact on clinical decision-making, and the quality of evidence reported for the respective molecular test. Consequently, the cost-effectiveness of molecular FNAC tests needs to be evaluated for each specific interdisciplinary diagnostic pathway setting, and end users need to perform independent studies on the basis of regional costs and population demographics and ultrasound and FNAC performance.[115] This was recently demonstrated by the first large patient cohort study before and after ThyroSeq implementation with 773 consecutive patients. It reported a doubling of AUS/FLUS and FN/SFN cytologies and an increase of the overall cost of care for patients with thyroid nodules.[116] This study did not include any information on ultrasound malignancy risk stratification and emphasizes again the importance of local outcome evaluations taking into account all building blocks of the diagnostic pathway for the characterization of a nodule.

KEY POINTS

- Thyroid nodule and thyroid cancer diagnostics should be organized as a pathway. Deficiencies for one of the diagnostic steps or expertise cannot be compensated for by (over)emphasizing any other step or expertise.
- As part of an integrated approach of careful clinical, ultrasound, and fine-needle aspiration cytology assessment with local outcome data, molecular testing may be able to improve diagnostic outcomes for thyroid nodules.

TREATMENT OF AUTONOMOUS THYROID NODULES

Antithyroid drugs, usually in combination with beta-blocking drugs, are the first-line treatment in all patients with overt thyrotoxicosis. Depending on the type of antithyroid drug, an initial dosage of 10 to 30 mg/day of methimazole, 20 to 60 mg/day of carbimazole, or 150 to 300 mg/day of propylthiouracil is recommended. Higher doses are associated with more frequent adverse effects (3%–12%) and will only result in marginally faster resolution of thyrotoxicosis. Furthermore, a trial of low-dose drug therapy (5–10 mg methimazole per day) may be justified in selected patients with symptomatic subclinical thyrotoxicosis.

While the purpose of antithyroid drug therapy is to render the patient euthyroid, there is, contrary to Graves disease, practically no spontaneous resolution of thyroid autonomy. This implies that, once thyroid autonomy becomes clinically manifest, one needs to either consider definitive treatment or choose long-term therapy with a thionamide in selected patients. However, benefits and risks of such “long-term” drug therapy have to be weighed against the potential advantages and limitations of definitive treatment.[83] Several ablative treatment options are available for TA and TMNG: thyroid surgery, radioiodine therapy, and thermal ablation techniques.

Surgery

The purpose of thyroid surgery is to cure thyrotoxicosis by removing all autonomously functioning thyroid tissue and other macroscopically visible nodular thyroid tissue. The extent of surgery is determined by preoperative ultrasound and, importantly, intraoperative morphologic inspection. For TA, hemithyroidectomy is usually adequate. In the case of TMNG, a subtotal, near-total, or total thyroidectomy is indicated. The advantages of surgery (removal of all nodular tissue, rapid and permanent resolution of thyrotoxicosis, and definite histologic diagnosis) have to be weighed against general (risk of anesthesia) and thyroid-specific side effects, the latter of which are largely dependent on the surgeon’s training and expertise (vocal cord paralysis ~1%, permanent hypoparathyroidism <1% for an experienced endocrine surgeon) (see Chapter 82). The incidence of postoperative hypothyroidism depends on the extent of thyroid resection.

While thyroid surgery is usually performed after euthyroidism has been achieved with antithyroid drug treatment, surgery is also advocated in patients with overt thyrotoxicosis who, prior to surgery, have had adverse side effects of a drug, or in patients with thyrotoxic storm who are refractory. Rapid control of severe thyrotoxicosis can be achieved by the administration of glucocorticoids, beta-blockers, and iodide in patients allergic to antithyroid drugs (see also Chapter 84).[83]

Radioiodine Therapy

Radioiodine therapy is widely used for treatment of thyroid autonomy and is highly effective in terms of eradicating thyrotoxicosis and reducing thyroid gland volume. The success rate for therapy has been reported to range between 85% and 100% in TA and up to 90% in TMNG. An average thyroid and/or nodule volume reduction of approximately 40% can generally be anticipated.[117]

To reduce the risk of exacerbation of hyperthyroidism, it has been recommended to render the patient euthyroid with antithyroid drugs prior to treatment. Usually, the antithyroid drug is discontinued at least 4 days before radioiodine therapy and resumed no sooner than 3 days afterward.[118] With large goiters (>80–100 mL), and especially with suspicion of malignancy, thyroid surgery is recommended. Different protocols have been suggested for therapy in benign thyroid disease. Some investigators prefer to administer a standard dose, e.g., 10 or 20 mCi (370–740 MBq) per thyroid gland, while others apply a certain activity per gram of thyroid tissue. Different algorithms exist for dose calculation, e.g., in Germany, dosage is mostly calculated according to the Marinelli formula, which takes into account the maximum radioiodine uptake. Advantages of radioiodine are its simplicity and, in many countries, its outpatient-based applicability. Disadvantages are the “time to euthyroidism” period (6 weeks to >3 months) during which drug therapy has to be continued and thyroid function monitored at 3- to 6-week intervals (Table 77. 3). Radioiodine treatment is contraindicated in pregnancy, and contraception is recommended for at least 6 months after receiving therapy.

Side effects of radioiodine treatment include transient local pain and tenderness. Significant exacerbation of thyrotoxicosis due to destructive thyroiditis is uncommon. In general, most studies have not shown an increased risk of thyroid cancer, leukemia, or other malignancies, reproductive abnormalities, or congenital defects in the offspring, in adults treated with radioiodine.[119] However, a more recent study has suggested a modest positive association with the risk of death from solid carcinomas, including breast cancer, in patients treated with radioiodine for hyperthyroidism,[120] and the safety in children remains controversial.[121]

Hypothyroidism after therapy for TMNG and TA usually develops insidiously. The prevalence of hypothyroidism depends on the extent of TSH suppression prior to therapy and the protocol applied, and, importantly, increases with the duration of follow-up. In a retrospective study of 346 patients with hyperfunctioning thyroid nodules, the occurrence of hypothyroidism was 7.6% at 1 year, 28% at 5 years, 46% at 10 years, and 60% at 20 years of follow-up.[122] These data emphasize the requirement of long-term monitoring of thyroid function in all patients receiving therapy.

Radioiodine treatment is considered safe and appropriate in nearly all types of hyperthyroidism, especially in older adult patients.[123,124] Generally, radioiodine therapy is thought to carry a lower rate of complications and a lower cost than surgery. This fact has led a number of centers to offer radioiodine therapy as the first choice of therapy in the majority of patients. In contrast to surgery, which cures nearly all patients and normalizes hyperthyroidism within a few days,[125] only 50% become euthyroid within 3 months of therapy, given that antithyroid drugs are not administered.[126] Twenty to forty percent need additional therapy, and even up to five treatment sessions will not cure all patients. In contrast, persistence of hyperthyroidism after surgery is rare.[125]

A summary of the advantages and disadvantages of the different treatments for TA and TMNG is shown in Table 77.4. For patients who refuse or are not suitable for surgery or radioiodine therapy, radiofrequency or laser ablation of the hot nodule is increasingly accepted as an alternative treatment option.[127]

Follow-Up

The long-term management of TA and TMNG patients is aimed at the detection and adequate treatment of thyroid dysfunction, prevention,

TABLE 77.3 Treatment Options for Patients with Toxic Multinodular Goiter

	Advantages	Disadvantages	Comments
Surgery	Significant goiter reduction Rapid achievement of euthyroidism Allows pathologic examination	Not all patients eligible Surgical mortality and morbidity a little higher than in nontoxic goiter Persistence or recurrence of hyperthyroidism Hypothyroidism*	Standard therapy for a large goiter or when rapid decompression is required Treatment of choice if radioiodine is not feasible
Antithyroid drugs	Easiest treatment option Mainly used at preparation for radioiodine or surgery	Lifelong treatment needed Very little chance of remission Adverse effects in approximately 5% Continuous goiter growth	Major indications are before thyroid surgery and before and after radioiodine, particularly in older adult patients and those with concurrent health problems Long-term treatment recommended only when surgery or radioiodine cannot be used
Radioiodine	Effective in rendering patients euthyroid and in reducing thyroid volume Most often outpatient Few subjective side effects	Only gradual reversal of hyperthyroidism Gradual reduction of the goiter Small risk for acute goiter enlargement Risk for thyroiditis: 3% Risk for transition into Graves disease: 5% Five-year risk for hypothyroidism: 15% Small risk for radiation-induced ophthalmopathy Treatment needs to be repeated in some	Standard therapy unless the goiter is very large (>100 mL). High-dose radioiodine (inpatient treatment) may be an option in those with a very large goiter not fit for or who decline surgery
Thermal ablation - Radiofrequency - Laser ablation	Alternative options to surgical treatment or radioiodine	In experienced hands, serious side effects are very rare	Mostly used for euthyroid nodules

*The percentage of patients affected depends on the extent of surgery.

TABLE 77.4 Treatment Options for Patients with Nontoxic Multinodular Goiter

Treatment	Advantages	Disadvantages	Comments
Surgery	Significant goiter reduction Rapid decompression of vital structures Allows pathologic examination	Not all patients eligible Postoperative hemorrhage (1%) Recurrent laryngeal nerve injury (1%–2%) Hypoparathyroidism (0.5%–5%) Hypothyroidism* and goiter recurrence* Postoperative tracheomalacia (rare) Risk rates are slightly increased in cases of large goiter, intrathoracic extension, or reoperation	Standard therapy for large goiters or when rapid decompression is required Total thyroidectomy should be considered to avoid goiter recurrence
Thyroxine	Outpatient Low cost May help prevent new nodule formation or goiter growth	Low efficacy Mainly impact on the perinodular volume Treatment, if lifelong and aimed at TSH suppression, could induce side effects caused by subclinical hyperthyroidism and inadvertent effects on bone and the heart	Its role on the wane due to possible adverse effects and low efficacy in reducing size Not recommended by the authors Low-dose treatment may help prevent growth
Radioiodine	Halving of thyroid volume within 1 year Improves inspiratory capacity in long term Most often outpatient Can be repeated successfully Few subjective side effects	Gradual reduction of the goiter Decreasing effect with increasing size Small risk for acute goiter enlargement Risk for thyroiditis: 3% Risk for transition into Graves disease: 5% One-year risk for hypothyroidism: 15%–20% Small risk for radiation-induced ophthalmopathy Treatment needs to be repeated in some Risk for radiation-induced malignancy unsettled	Has replaced surgery as the standard therapy in some European countries. Should be contemplated instead of reoperation and in those who decline or are not fit for surgery, also in case of a very large goiter (high-dose radioiodine) For data in relation to rhTSH-stimulated therapy, please see the text
Thermal ablation - Radiofrequency - Laser ablation	Alternative options to surgical treatment or observation alone for solid nodules, purely or predominantly cystic nodules that recur after aspiration and ethanol ablation, and those that have a residual solid component after such therapy	In experienced hands, serious side effects are very rare.	Only 20% and 38% show significant regrowth in the long term after radiofrequency ablation or laser ablation, respectively

*The percentage of patients affected depends on the extent of surgery.
TSH, Thyroid-stimulating hormone; *rhTSH,* recombinant human thyroid-stimulating hormone.

and detection of novel nodular thyroid disease, and, in the case of surgery, detection and treatment of postsurgical hypoparathyroidism. With radioiodine therapy, long-term follow-up for the development of hypothyroidism is mandatory.

TREATMENT OF EUTHYROID NODULES (SEE TABLE 77.3)

Nodular thyroid disease is very common, but most of these goiters do not cause significant symptoms and are best left untreated. Treatment may be indicated in the case of the following:

1. Large goiter or progressive growth of the entire gland or individual nodules
2. Signs of cervical compression
3. Overt or subclinical hyperthyroidism
4. Marked neck disfigurement or cosmetic complaints
5. Suspicion of thyroid cancer

In a number of clinical situations, a discrepancy is seen between clinical findings and patient complaints. In this context, the decision whether to treat can be more difficult.

There is no ideal treatment for goiter, and practice patterns differ quite widely. This is, for example, reflected by the fact that one third of clinicians would refrain from treating a patient with moderate discomfort due to a multinodular nontoxic goiter of 50 to 80 g, where malignancy has been ruled out.[13] Comparative studies of available options are sparse. Thus, treatment is not only a matter of goiter reduction. Patient satisfaction, the risks for hypothyroidism and goiter recurrence, and fear of overlooking a thyroid cancer, age, and comorbidities are all important factors that should be taken into account. It follows that the optimal treatment of toxic and nontoxic multinodular goiter is controversial, and the therapeutic approach must be individualized.

Iodine Supplementation

Some clinicians, particularly in Europe, use iodine supplementation for treatment of goiter.[13] In nodular goiter, iodine is no better than levothyroxine (LT4) suppression therapy for goiter reduction in comparative trials. In addition, the major hindrance for the use of iodine consists in the fact that a sudden increase of iodine intake may induce thyrotoxicosis in individuals with underlying thyroid autonomy.

Thyroid Hormone Treatment

Thyroid hormone therapy for suppression of pituitary TSH secretion has been widely used in nontoxic multinodular goiter. Although LT4 suppressive therapy is effective in reducing the volume of diffuse nontoxic goiters by up to 30%, there are few controlled studies in nontoxic multinodular goiter, in which sonography has been employed for objective size monitoring. In one study, 58% of patients had a significant (>13%) decrease in thyroid volume, but regrowth was seen after discontinuation of therapy.[128] In a randomized trial, Wesche and coworkers[129] found a median reduction of goiter volume in the radioiodine treated group of 38% and 44% after 1 and 2 years, while the corresponding values in the LT4-treated group were 7% and 1%, respectively, and nonsignificant. In a German study, LT4 in combination with elementary iodine supplementation resulted in a goiter volume reduction of only 7.9%, and a nodule volume reduction of 17.3% within 1 year, as compared with placebo.[130]

For volume reduction, the LT4 dose is often adjusted to reach a partly suppressed serum TSH level.[130] The consequence is subclinical hyperthyroidism, which may adversely affect the skeleton and the cardiovascular system.[131] Because lifelong therapy is probably needed to avoid goiter recurrence, and the natural history of the disease is progression toward hyperthyroidism due to autonomous function of the thyroid nodules, LT4 treatment may not be appropriate in many patients.[132] Based on the aforementioned considerations, LT4 treatment has been abandoned by many endocrinologists for this indication.[13]

Surgery

For a comprehensive overview of thyroid surgery, see Chapter 82. The goal of surgery is removal of all thyroid tissue with a nodular appearance, usually by a unilateral hemithyroidectomy and subtotal resection of the contralateral lobe. A bilateral subtotal resection cannot be recommended. Only extremely rarely is a thoracic approach necessary. Further resection is not usually recommended if the final pathologic evaluation incidentally reveals a unilateral cancer less than 1 cm in size. This rather frequent finding accounts for most cancers found in surgical series, the majority of which are of little, if any, clinical significance.[133] Macroscopically normal perinodular tissue often harbors microscopic growth foci, which explains the relatively high risk for recurrence in these patients.[134] In the case of a toxic nodular goiter, thyroid function is more rapidly normalized after surgery than after antithyroid drug therapy. Surgery leads to rapid decompression resulting in improved respiratory function, if affected presurgically.[135]

Not all patients are surgical candidates, but among those undergoing surgery, the surgical mortality rate is less than 1% in experienced centers. Disadvantages include the general risks and side effects of a surgical procedure. Specific risks include transient (6%) or permanent (2%) vocal cord paralysis, transient (6%) or permanent (5%) hypoparathyroidism, and postoperative bleeding (1%).[135] Others have found lower figures.[136] Complications are related to increasing goiter size and extent of the resection.[137] Novel techniques may reduce the operation time, postoperative pain, and length of hospital stay. Postoperative tracheomalacia, necessitating intubation, can occur after the removal of large goiters but is a rare complication. A matter of concern is the apparently high prevalence (7%–17%) of thyroid carcinomas in substernal goiters,[138] but this seemingly high frequency is probably influenced by selection bias.[139]

The long-term risk for hypothyroidism after subtotal resection of multinodular goiters is insufficiently described, but is approximately 10% to 20%, as reported for toxic multinodular goiters, and related to the extent of the resection. Recurrence of the nontoxic multinodular goiter is seen in 15% to 40% of patients with long-term follow-up.[13] Postoperative use of LT4 is preferred by many clinicians to avoid recurrence; but based on results from randomized trials, use of LT4 is generally not recommended, and neither is iodine.[13,132] A reoperation for the recurrent goiter results in a 3-fold to 10-fold increase in risk for permanent vocal cord paralysis or hypoparathyroidism.[136,137] Goiter recurrence can be completely avoided if a total thyroidectomy is carried out initially, an approach that does not have a higher complication compared with subtotal thyroidectomy if the intervention is performed by high-volume surgeons.[140]

Radioiodine Therapy

Besides being able to cure hyperthyroidism, radioiodine therapy can also be used for volume reduction. During the last 25 years, radioiodine therapy for symptomatic multinodular nontoxic goiter has therefore been introduced in a number of mainly European centers as a nonsurgical alternative to LT4 therapy. Thyroid volume reduction is of the same magnitude in toxic and nontoxic multinodular goiters, that is, approximately 40% after 1 year[126,129,141,142] and 50% to 60% after 2 years, without further reduction.[141,143] Sixty percent of this decrease is seen within 3 months of therapy.[141,142] In addition to a relief of compressive symptoms, radioiodine therapy also results in decreased tracheal compression, which can improve pulmonary function, particularly the inspiratory component.[142,144] If used for very large goiters, normalization of thyroid volume, as seen in diffuse toxic and nontoxic goiters, is rarely achieved, but symptoms are in most cases considerably improved, and patient satisfaction is high.[144] If a secondary increase in thyroid volume is seen, this should raise the suspicion of malignancy. Generally, therapy doses of 100 μCi (3.7 MBq) per gram of thyroid tissue corrected for 100% 24-hour therapy uptake have been given.[144,145] However, it has been questioned whether such dose adjustment is worthwhile,[123,145] and fixed doses are given in a number of centers if nuclear medicine regulations allow to do so.[123] The treatment can be repeated if further goiter reduction in a euthyroid patient is required.[141]

Radiation thyroiditis is seen in 3% within the first months of radioiodine therapy[124] and is easily treated with salicylates or corticosteroids. Another possible complication consists in "Graves-like autoimmune hyperthyroidism," which is seen in up to 5%. Rare cases of radioiodine therapy-induced Graves ophthalmopathy have also been reported. The presence of anti-TPO antibodies prior to radioiodine therapy is associated with an increased risk for this complication, which is most likely triggered by therapy related release of thyroid antigens and the appearance of TSHR antibodies typically 3 to 6 months after radioiodine therapy. This phenomenon can also be seen after surgical manipulation of the thyroid or after subacute thyroiditis. The condition is often self-limiting, but may necessitate therapy.

Although goiter enlargement caused by radioiodine therapy may be seen initially after treatment, there is usually no significant acute thyroid enlargement.[124,142,146] The risk for permanent hypothyroidism after therapy in multinodular goiters ranges from 14% to 58% within 5 to 8 years.[124] It occurs more commonly in patients with a small goiter and when anti-TPO antibodies are present. Radioiodine therapy, given for Graves disease for decades, is not thought to be followed by any clinically significant increased risk for cancer deaths,[119,124,147] although one study questions this view (see earlier).[148] Most patients treated are older than 50 years of age and are less susceptible to radiation-induced malignancy than young individuals. Data regarding radioiodine therapy in multinodular goiter are sparse, and in the case of nontoxic goiter, nonexistent. In a study by Ron and coworkers,[147] 1089 patients were treated for a toxic nodular goiter, and these individuals had a 31% increase in overall cancer mortality, nearly exclusively attributable to thyroid malignancy. However, a similar pattern was seen in patients not treated with radioiodine. Hence, the detection of a thyroid cancer in a nodular goiter after radioiodine therapy raises the question of whether malignancy in a nodule had been overlooked prior to radioiodine therapy.

In some European countries, radioiodine therapy has replaced surgery as the treatment of choice in most patients.[13,124] However, the optimal treatment modality remains to be established, ideally through comparative randomized trials, including data on effect, side effects, cost, and patient satisfaction.

The efficacy of radioiodine therapy in multinodular goiter is hampered by the irregular therapy uptake in the gland, and the relative goiter reduction is inversely correlated with the initial goiter size.[124,142] A high dietary iodine intake diminishes efficacy. Recombinant human TSH (rhTSH) has the potential of increasing the 24-hour radioiodine therapy uptake more than 4-fold, and the effect is inversely correlated to the initial thyroid radioiodine therapy uptake.[124,149-152] Moreover, pretreatment with rhTSH causes a more homogeneous distribution of radioiodine therapy within the nodular gland.[153] These properties of rhTSH are ideal in the context of therapy of multinodular goiter. Several studies, including two randomized double-blind trials,[124] have shown that rhTSH, in doses from 0.1 mg to 0.9 mg and administered 24 hours before radioiodine therapy, improves goiter reduction by 35% to 55% within 1 year, compared with radioiodine therapy without stimulation. RhTSH-augmented radioiodine therapy also results in reduced tracheal compression and an enhancement of the inspiratory reserve, as shown in a randomized trial.[154] Long-term follow-up studies have documented that the enhanced goiter volume reduction obtained with rhTSH-augmented therapy improves the reduction in goiter-related symptoms and reduces the need for additional therapy.[124] The administration of rhTSH may also permit the successful use of lower activities of radioiodine to achieve goiter reduction.[117,155] Indeed, in a randomized double-blind study,[155] it was demonstrated that an equivalent goiter reduction after radioiodine therapy can be obtained by a much lower thyroid dose if rhTSH prestimulation is used.[155,156] Although rhTSH stimulation has several attractive aspects for volume reduction in selected patients, it is important to emphasize that it forms an "off-label" treatment for multinodular goiter. RhTSH-augmented radioiodine therapy is generally well tolerated, particularly when using a dose of 0.1 mg or lower. A drawback of rhTSH-augmented radioiodine therapy is the fact that, in parallel with the improved goiter reduction, a higher frequency of hypothyroidism is encountered.[124] If rhTSH stimulation is used in order to lower the amount of the radioactive dose delivered to the thyroid gland, the risk for permanent hypothyroidism is comparable to that seen after radioiodine therapy alone.[155]

Nonsurgical Thermal Ablation of Thyroid Nodules

Thermal ablation has shown efficacy in the treatment of a spectrum of benign thyroid nodules in multiple studies. In benign nodules that cause pressure symptoms and/or cosmetic concerns, when balancing efficacy, side effects, and cost, both laser ablation and radiofrequency ablation are alternative options to surgical treatment or observation for solid nodules, purely or predominantly cystic nodules that recur after aspiration and ethanol ablation, and those that have a residual solid component after such therapy.[127,157] Only 20% and 38% show significant regrowth in the long term after radiofrequency ablation or laser ablation, respectively. In experienced hands, serious side effects are very rare. Thermal ablation for properly selected patients is widely used in some countries, but currently still more rarely used in others. Prior to thermal ablation, FNAC is warranted to characterize the nature of the nodule. Size and location, as well as operator experience, influence outcomes. Lesions that are located in the posterior and medial part of the thyroid may not be amenable to thermal ablation because of the proximity to the recurrent laryngeal nerve and the esophagus ("danger triangle"). Proper identification of patients who are most likely benefit from these localized therapies is key. There is no reason for any therapeutic intervention (thermal ablation or surgery) for sonographically visible nodules *per se* without local symptoms or cosmetic reasons.

Percutaneous ethanol injection therapy (PEIT) has been used for two decades in solitary hot, toxic, and even cold thyroid nodules.[158] The most convincing effect is seen in solitary thyroid cysts.[127] In centers with access to thermal ablation techniques, PEIT now generally plays a secondary role.

KEY POINTS

- Surgery remains the primary therapy in symptomatic patients with benign nontoxic goiter, and should be as complete as possible to avoid recurrence.
- Radioiodine can be used as a nonsurgical therapy option in benign nontoxic nodular goiter. Efficacy can be augmented by recombinant human thyroid-stimulating hormone prestimulation of iodine uptake.
- Thermal ablation has repeatedly shown efficacy in the treatment of a spectrum of benign thyroid nodules.

REFERENCES

1. Lloyd RV, Osamura RY, Klöppel G, et al., eds. WHO Classification of Tumours of Endocrine Organs. *WHO Classification of Tumours*, 4th Ed. 2017:10.
2. Berghout A, Wiersinga WM, Smits NJ, et al. Interrelationships between age, thyroid volume, thyroid nodularity, and thyroid function in patients with sporadic nontoxic goiter. *Am J Med*. 1990;89:602–608.
3. Knudsen N, Laurberg P, Perrild H, et al. Risk factors for goiter and thyroid nodules. *Thyroid*. 2002;12:879–888.
4. Knudsen N, Bulow I, Laurberg P, et al. Parity is associated with increased thyroid volume solely among smokers in an area with moderate to mild iodine deficiency. *Euro J Endocrinol*. 2002;146:39–43.
5. Krohn K, Fuhrer D, Bayer Y, et al. Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocrine Rev*. 2005;26:504–524.
6. Brix TH, Kyvik KO, Hegedus L. Major role of genes in the etiology of simple goiter in females: a population-based twin study. *J Clin Endocrinol Metabol*. 1999;84:3071–3075.
7. Paschke R. Molecular pathogenesis of nodular goiter. *Langenbeck's Archiv Surg*. 2011;396:1127–1136.
8. Bayer Y, Neumann S, Meyer B, et al. Genome-wide linkage analysis reveals evidence for four new susceptibility loci for familial euthyroid goiter. *J Clin Endocrinol Metabol*. 2004;89:4044–4052.
9. Teumer A, Rawal R, Homuth G, et al. Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. *Am J Human Genetic*. 2011;88:664–673.
10. Chong AS, Nikiforov YE, Condello V, et al. Prevalence and spectrum of DICER1 mutations in adult-onset thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab*. 2021;106:968–977.
11. Krohn K, Wohlgemuth S, Gerber H, et al. Hot microscopic areas of iodine-deficient euthyroid goitres contain constitutively activating TSH receptor mutations. *J Pathol*. 2000;192:37–42.
12. Grani G, Bruno R, Lucisano G, et al. Temporal changes in thyroid nodule volume: lack of effect on paranodular thyroid tissue volume. *Thyroid*. 2017;27:1378–1384.
13. Hegedus L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocrine Rev*. 2003;24:102–132.
14. Krohn K, Maier J, Paschke R. Mechanisms of disease: hydrogen peroxide, DNA damage and mutagenesis in the development of thyroid tumors. *Nat Clin Pract Endocrinol Metabol*. 2007;3:713–720.
15. Kopp P. Thyrotoxicosis of other etiologies. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext*. South Dartmouth; 2000. https://www.endotext.org.
16. Kopp P. The TSH receptor and its role in thyroid disease. *Cell Mol Life Sci*. 2001;58:1301–1322.
17. Trulzsch B, Krohn K, Wonerow P, et al. Detection of thyroid-stimulating hormone receptor and Gsalpha mutations: in 75 toxic thyroid nodules by denaturing gradient gel electrophoresis. *J Mol Med*. 2001;78:684–691.
18. Calebiro D, Grassi ES, Eszlinger M, et al. Recurrent EZH1 mutations are a second hit in autonomous thyroid adenomas. *J Clin Investigat*. 2016;126:3383–3388.
19. Stephenson A, Eszlinger M, Stewardson P, et al. Sensitive sequencing analysis suggests thyrotropin receptor and guanine nucleotide-binding protein G subunit alpha as sole driver mutations in hot thyroid nodules. *Thyroid*. 2020;30:1482–1489.
20. Gozu HI, Bircan R, Krohn K, et al. Similar prevalence of somatic TSH receptor and Gsα mutations in toxic thyroid nodules in geographical regions with different iodine supply in Turkey. *Euro J Endocrinol*. 2006;155:535–545.
21. Trulzsch B, Krohn K, Wonerow P, et al. Detection of thyroid-stimulating hormone receptor and Gs mutations: in 75 toxic thyroid nodules by denaturing gradient gel electrophoresis. *J Mol Med*. 2000;12:684–691.
22. Eszlinger M, Krohn K, Frenzel R, et al. Gene expression analysis reveals evidence for inactivation of the TGF-beta signaling cascade in autonomously functioning thyroid nodules. *Oncogene*. 2004;23:795–804.
23. Taton M, Lamy F, Roger PP, et al. General inhibition by transforming growth factor beta 1 of thyrotropin and cAMP responses in human thyroid cells in primary culture. *Mol Cellular Endocrinol*. 1993;95:13–21.
24. Krohn K, Emmrich P, Ott N, et al. Increased thyroid epithelial cell proliferation in toxic thyroid nodules. *Thyroid*. 1999;9:241–246.
25. Deleu S, Allory Y, Radulescu A, et al. Characterization of autonomous thyroid adenoma: metabolism, gene expression, and pathology. *Thyroid*. 2000;10:131–140.
26. Studer H, Peter HJ, Gerber H. Natural heterogeneity of thyroid cells: the basis for understanding thyroid function and nodular goiter growth. *Endocrine Rev*. 1989;10:125–135.
27. Krohn K, Reske A, Ackermann F, et al. Ras mutations are rare in solitary cold and toxic thyroid nodules. *Clin Endocrinol*. 2001;55:241–248.
28. Dohan O, De LV, Paroder V, et al. The sodium/iodide symporter (NIS): characterization, regulation, and medical significance. *Endocrine Rev*. 2003;24:48–77.
29. Krohn K, Paschke R. BRAF mutations are not an alternative explanation for the molecular etiology of ras-mutation negative cold thyroid nodules. *Thyroid*. 2004;14:359–361.
30. Kimura ET, Nikiforova MN, Zhu Z, et al. High prevalence of BRAF mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res*. 2003;63:1454–1457.
31. Krohn K, Stricker I, Emmrich P, et al. Cold thyroid nodules show a marked increase in proliferation markers. *Thyroid*. 2003;13:569–575.
32. Knudsen N, Bulow I, Jorgensen T, et al. Comparative study of thyroid function and types of thyroid dysfunction in two areas in Denmark with slightly different iodine status. *Eur J Endocrinol*. 2000;143:485–491.
33. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer*. 2013;13:184–199.
34. Sponziello M, Lavarone E, Pegolo E, et al. Molecular differences between human thyroid follicular adenoma and carcinoma revealed by analysis of a murine model of thyroid cancer. *Endocrinology*. 2013;154:3043–3053.
35. Nikiforova MN, Biddinger PW, Caudill CM, et al. PAX8-PPARgamma rearrangement in thyroid tumors: RT-PCR and immunohistochemical analyses. *Am J Surg Pathol*. 2002;26:1016–1023.
36. Ye L, Zhou X, Huang F, et al. The genetic landscape of benign thyroid nodules revealed by whole exome and transcriptome sequencing. *Nat Commun*. 2017;8:15533.
37. Yoo SK, Lee S, Kim SJ, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet*. 2016;12:e1006239.
38. Thomas JL, Leclere J, Hartemann P, et al. Familial hyperthyroidism without evidence of autoimmunity. *Acta Endocrinol-Cop*. 1982;100:512–518.
39. Parma J, Duprez L, Van Sande J, et al. Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas. *Nature*. 1993;365:649–651.
40. Duprez L, Parma J, Van Sande J, et al. Germline mutations in the thyrotropin receptor gene cause non- autoimmune autosomal dominant hyperthyroidism. *Nat Genet*. 1994;7:396–401.
41. Kopp P, Vansande J, Parma J, et al. Brief report - congenital hyperthyroidism caused by a mutation in the thyrotropin-receptor gene. *N Engl J Med*. 1995;332:150–154.
42. Sunthornthepvarakul T, Gottschalk ME, Hayashi Y, et al. Brief report: resistance to thyrotropin caused by mutations in the thyrotropin-receptor gene. *N Engl J Med*. 1995;332:155–160.
43. Hébrant A, van Staveren WCG, Maenhaut C, et al. Genetic hyperthyroidism: hyperthyroidism due to activating TSHR mutations. *Euro J Endocrinol*. 2011;164:1–9.
44. Stephenson A, Lau L, Eszlinger M, et al. The thyrotropin receptor mutation database update. *Thyroid*. 2020;30:931–935.
45. Paschke R, Niedziela M, Vaidya B, et al. 2012 European thyroid association guidelines for the management of familial and persistent sporadic non-autoimmune hyperthyroidism caused by thyroid-stimulating hormone receptor germline mutations. *Eur Thyroid J*. 2012;1:142–147.
46. Burgi H. Iodine excess. *Best Pract Res Clin Endocrinol Metab*. 2010;24:107–115.
47. Lee SY, Rhee CM, Leung AM, et al. A review: radiographic iodinated contrast media-induced thyroid dysfunction. *J Clin Endocrinol Metab*. 2015;100:376–383.

48. Wolff J, Chaikoff IL, Goldberg RC, et al. The temporary nature of the inhibitory action of excess iodine on organic iodine synthesis in the normal thyroid. *Endocrinology*. 1949;45:504–513.
49. Bartalena L, Bogazzi F, Chiovato L, et al. 2018 European Thyroid Association (ETA) guidelines for the management of amiodarone-associated thyroid dysfunction. *Eur Thyroid J*. 2018;7:55–66.
50. Rhee CM, Bhan I, Alexander EK, et al. Association between iodinated contrast media exposure and incident hyperthyroidism and hypothyroidism. *Archiv Internal Med*. 2012;172:153–159.
51. Bervini S, Trelle S, Kopp PA, et al. Prevalence of iodine-induced hyperthyroidism after administration of iodinated contrast during radiographic procedures: a systematic review and meta-analysis of the literature. *Thyroid*. 2021;31:1020–1029.
52. Reinwein D, Benker G, Konig MP, et al. The different types of hyperthyroidism in Europe. Results of a prospective survey of 924 patients. *J Endocrinol Invest*. 1988;11:193–200.
53. Vejbjerg P, Knudsen N, Perrild H, et al. Effect of a mandatory iodization program on thyroid gland volume based on individuals' age, gender, and preceding severity of dietary iodine deficiency: a prospective, population-based study. *J Clin Endocrinol Metabol*. 2007;92:1397–1401.
54. Dirikoc A, Polat SB, Kandemir Z, et al. Comparison of ultrasonography features and malignancy rate of toxic and nontoxic autonomous nodules: a preliminary study. *Ann Nucl Med*. 2015;29:883–889.
55. Jaeschke H, Undeutsch H, Patyra K, et al. Hyperthyroidism and papillary thyroid carcinoma in thyrotropin receptor D633H mutant mice. *Thyroid*. 2018;28:1372–1386.
56. Sandrock D, Olbricht T, Emrich D, et al. Long-term follow-up in patients with autonomous thyroid adenoma. *Acta Endocrinol (Copenh)*. 1993;128:51–55.
57. Hamburger JI. Evolution of toxicity in solitary nontoxic autonomously functioning thyroid nodules. *J Clin Endocrinol Metabol*. 1980;50:5.
58. Sandrock D, Olbricht T, Emrich D, et al. Long-term follow-up in patients with autonomous thyroid adenoma. *Acta Endocrinologica*. 1993;128:51–55.
59. Laurberg P, Pedersen KM, Vestergaard H, et al. High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J Internal Med*. 1991;229:415–420.
60. Durante C, Grani G, Lamartina L, et al. The diagnosis and management of thyroid nodules: a review. *J Am Med Assoc*. 2018;319:914–924.
61. Kwong N, Medici M, Angell TE, et al. The influence of patient age on thyroid nodule formation, multinodularity, and thyroid cancer risk. *J Clin Endocrinol Metab*. 2015;100:4434–4440.
62. Volzke H, Ludemann J, Robinson DM, et al. The prevalence of undiagnosed thyroid disorders in a previously iodine-deficient area. *Thyroid*. 2003;13:803–810.
63. Guth S, Theune U, Aberle J, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest*. 2009;39:699–706.
64. Russ G, Bonnema SJ, Erdogan MF, et al. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J*. 2017;6:225–237.
65. Sharma SD, Jacques T, Smith S, et al. Diagnosis of incidental thyroid nodules on 18F-fluorodeoxyglucose positron emission tomography imaging: are these significant? *J Laryngol Otol*. 2015;129:53–56.
66. Haugen BR, Alexander EK, Bible KC, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines Task Force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
67. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting And Data System (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol*. 2017;14:587–595.
68. Nakamura H, Hirokawa M, Ota H, et al. Is an increase in thyroid nodule volume a risk factor for malignancy? *Thyroid*. 2015;25:804–811.
69. US Preventive Services Task Force; Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for thyroid cancer: US preventive Services Task Force recommendation statement. *JAMA*. 2017;317:1882–1887.
70. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"—screening and overdiagnosis. *N Engl J Med*. 2014;371:1765–1767.
71. Vaccarella S, Franceschi S, Bray F, et al. Worldwide thyroid-cancer epidemic? the increasing impact of overdiagnosis. *N Engl J Med*. 2016;375:614–617.
72. Morris LG, Sikora AG, Tosteson TD, et al. The increasing incidence of thyroid cancer: the influence of access to care. *Thyroid*. 2013;23:885–891.
73. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract*. 2016;22:622–639.
74. Paschke R, Cantara S, Crescenzi A, et al. European Thyroid Association guidelines regarding thyroid nodule molecular fine-needle aspiration cytology diagnostics. *EurThyroid J*. 2017;6:115–129.
75. Van den Bruel A, Francart J, Dubois C, et al. Regional variation in thyroid cancer incidence in Belgium is associated with variation in thyroid imaging and thyroid disease management. *J Clin Endocrinol Metab*. 2013;98:4063–4071.
76. Persichetti A, Di Stasio E, Guglielmi R, et al. Predictive value of malignancy of thyroid nodule ultrasound classification systems: a prospective study. *J Clin Endocrinol Metab*. 2018;103:1359–1368.
77. Cibas ES, Ali SZ. The 2017 Bethesda System for reporting thyroid cytopathology. *Thyroid*. 2017;27:1341–1346.
78. Eszlinger M, Ullmann M, Ruschenburg I, et al. Low malignancy rates in fine-needle aspiration cytologies in a primary care setting in Germany. *Thyroid*. 2017;27:1385–1392.
79. Basaria S, Salvatori R. Images in clinical medicine. Pemberton's sign. *N Engl J Med*. 2004;350.
80. Trivalle C, Doucet J, Chassagne P, et al. Differences in the signs and symptoms of hyperthyroidism in older and younger patients. *J Am Geriatr Soc*. 1996;44:50–53.
81. Costante G, Durante C, Francis Z, et al. Determination of calcitonin level in C-cell disease: clinical interest and potential pitfalls. *Nat Clin Pract Endocrinol Metab*. 2009;5:35–44.
82. Nygaard B, Hegedus L, Nielsen KG, et al. Long-term effect of radioactive iodine on thyroid function and size in patients with solitary autonomously functioning toxic thyroid nodules. *Clin Endocrinol (Oxf)*. 1999;50:197–202.
83. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.
84. Franc B, de la Salmoniere P, Lange F, et al. Interobserver and intraobserver reproducibility in the histopathology of follicular thyroid carcinoma. *Hum Pathol*. 2003;34:1092–1100.
85. Belfiore A, La Rosa GL, La Porta GA, et al. Cancer risk in patients with cold thyroid nodules: relevance of iodine intake, sex, age, and multinodularity. *Am J Med*. 1992;93:363–369.
86. Knudsen N, Perrild H, Christiansen E, et al. Thyroid structure and size and two-year follow-up of solitary cold thyroid nodules in an unselected population with borderline iodine deficiency. *Euro J Endocrinol*. 2000;142:224–230.
87. Symonds CJ, Seal P, Ghaznavi S, et al. Thyroid nodule ultrasound reports in routine clinical practice provide insufficient information to estimate risk of malignancy. *Endocrine*. 2018;61:303–307.
88. Seifert P, Gorges R, Zimny M, et al. Interobserver agreement and efficacy of consensus reading in Kwak-, EU-, and ACR-thyroid imaging recording and data systems and ATA guidelines for the ultrasound risk stratification of thyroid nodules. *Endocrine*. 2020;67:143–154.
89. Pantano AL, Maddaloni E, Briganti SI, et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. *Eur J Endocrinol*. 2018;178:595–603.
90. Grani G, Lamartina L, Ascoli V, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "right" TIRADS. *J Clin Endocrinol Metab*. 2019;104:95–102.
91. Tessler FN. Thyroid nodules and real estate: location matters. *Thyroid*. 2020;30:349–350.

92. Garber JR, Papini E, Frasoldati A, et al. American Association of Clinical Endocrinology And Associazione Medici Endocrinologi Thyroid Nodule Algorithmic Tool. *Endocr Pract*. 2021;27:649–660.
93. Bongiovanni M, Spitale A, Faquin WC, et al. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol*. 2012;56: 333–339.
94. Nikiforov YE. Ramifications of new terminology for encapsulated follicular variant of papillary thyroid carcinoma-reply. *JAMA Oncol*. 2016;2:1098–1099.
95. Wienhold R, Scholz M, Adler JR, et al. The management of thyroid nodules: a retrospective analysis of health insurance data. *Dtsch Arztebl Int*. 2013;110:827–834.
96. Trimboli P, Treglia G, Guidobaldi L, et al. Detection rate of FNA cytology in medullary thyroid carcinoma: a meta-analysis. *Clin Endocrinol (Oxf)*. 2015;82:280–285.
97. Essig Jr GF, Porter K, Schneider D, et al. Fine needle aspiration and medullary thyroid carcinoma: the risk of inadequate preoperative evaluation and initial surgery when relying upon FNAB cytology alone. *Endocr Pract*. 2013;19:920–927.
98. d'Herbomez M, Caron P, Bauters C, et al. Reference range of serum calcitonin levels in humans: influence of calcitonin assays, sex, age, and cigarette smoking. *Eur J Endocrinol*. 2007;157:749–755.
99. Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
100. Allen L, Al Afif A, Rigby MH, et al. The role of repeat fine needle aspiration in managing indeterminate thyroid nodules. *J Otolaryngol Head Neck Surg*. 2019;48:16.
101. Stewart R, Leang YJ, Bhatt CR, et al. Quantifying the differences in surgical management of patients with definitive and indeterminate thyroid nodule cytology. *Eur J Surg Oncol*. 2020;46:252–257.
102. Steward DL, Carty SE, Sippel RS. Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: a prospective blinded multicenter study. *JAMA Oncol*. 2019;1:204–212.
103. Patel KN, Angell TE, Babiarz J, et al. Performance of a genomic sequencing classifier for the preoperative diagnosis of cytologically indeterminate thyroid nodules. *JAMA Surg*. 2018;153:817–824.
104. Labourier E, Shifrin A, Busseniers AE, et al. Molecular testing for miRNA, mRNA and DNA on fine needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab*. 2015;100:2743–2750.
105. Gonzalez HE, Martínez JR, Vargas-Salas S, et al. A 10-gene classifier for indeterminate thyroid nodules: development and multicenter accuracy study. *Thyroid*. 2017;27:1058–1067.
106. Eszlinger M, Bohme K, Ullmann M, et al. Evaluation of a two-year routine application of molecular testing of thyroid fine-needle aspirations using a seven-gene panel in a primary referral setting in Germany. *Thyroid*. 2017;27:402–411.
107. Stewardson P, Khalil M, Ghaznavi S, et al. *Prospective Evaluation of the ThyroSPECTM Mutation Panel for the Diagnosis of Indeterminate Thyroid fine Needle Aspiration Cytologies (FNAC) in the Southern Alberta Healthcare Region*. Chicago: 89th annual meeting of the American Thyroid Association; 2019.
108. Angell TE, Heller HT, Cibas ES, et al. Independent comparison of the Afirma genomic sequencing classifier and gene expression classifier for cytologically indeterminate thyroid nodules. *Thyroid*. 2019;29:650–656.
109. Valderrabano P, Hallanger-Johnson JE, Thapa R, et al. Comparison of postmarketing findings vs the initial clinical validation findings of a thyroid nodule gene expression classifier: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*. 2019;145:783–792.
110. Endo M, Nabhan F, Angell TE, et al. Letter to the editor: Use of molecular diagnostic tests in thyroid nodules with hurthle cell-dominant cytology. *Thyroid*. 2020;30:1390–1392.
111. Ganly I, Makarov V, Deraje S, et al. Integrated genomic analysis of Hurthle cell cancer reveals oncogenic drivers, recurrent mitochondrial mutations, and unique chromosomal landscapes. *Cancer Cell*. 2018;34:256–270.e5.
112. Gopal RK, Kubler K, Calvo SE, et al. Widespread chromosomal losses and mitochondrial DNA alterations as genetic drivers in Hurthle cell carcinoma. *Cancer Cell*. 2018;34:242–255.e5.
113. Zafereo M, McIver B, Vargas-Salas S, et al. A thyroid genetic classifier correctly predicts benign nodules with indeterminate cytology: two independent, multicenter, prospective validation trials. *Thyroid*. 2020;30:704–712.
114. Najafian A, Noureldine S, Azar F, et al. RAS mutations, and RET/PTC and PAX8/PPAR-gamma chromosomal rearrangements are also prevalent in benign thyroid lesions: implications thereof and a systematic review. *Thyroid*. 2017;27:39–48.
115. Eszlinger M, Lau L, Ghaznavi S, et al. Molecular profiling of thyroid nodule fine-needle aspiration cytology. *Nat Rev Endocrinol*. 2017;13:415–424.
116. Fazeli SR, Zehr B, Amraei R, et al. ThyroSeq v2 testing: impact on cytologic diagnosis, management, and cost of care in patients with thyroid nodule. *Thyroid*. 2020;30:1528–1534.
117. Nieuwlaat WA, Huysmans DA, van den Bosch HC, et al. Pretreatment with a single, low dose of recombinant human thyrotropin allows dose reduction of radioiodine therapy in patients with nodular goiter. *J Clin Endocrinol Metab*. 2003;88:3121–3129.
118. Bonnema SJ, Bennedbaek FN, Gram J, et al. Resumption of methimazole after 131I therapy of hyperthyroid diseases: effect on thyroid function and volume evaluated by a randomized clinical trial. *Euro J Endocrinol*. 2003;149:485–492.
119. Franklyn JA, Maisonneuve P, Sheppard M, et al. Cancer incidence and mortality after radioiodine treatment for hyperthyroidism: a population-based cohort study. *Lancet*. 1999;353:2111–2115.
120. Kitahara CM, Berrington de Gonzalez A, Bouville A, et al. Association of radioactive iodine treatment with cancer mortality in patients with hyperthyroidism. *JAMA Intern Med*. 2019;179:1034–1042.
121. American Thyroid Association Taskforce On Radioiodine Safety; Sisson JC, Freitas J, McDougall IR, et al. Radiation safety in the treatment of patients with thyroid diseases by radioiodine 131I : practice recommendations of the American Thyroid Association. *Thyroid*. 2011;21:335–346.
122. Ceccarelli C, Bencivelli W, Vitti P, et al. Outcome of radioiodine-131 therapy in hyperfunctioning thyroid nodules: a 20 years' retrospective study. *Clin Endocrinol (Oxf)*. 2005;62:331–335.
123. Weetman AP. Radioiodine treatment for benign thyroid diseases. *Clin Endocrinol*. 2007;66:757–764.
124. Bonnema SJ, Hegedus L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*. 2012;33:920–980.
125. Porterfield Jr JR, Thompson GB, Farley DR, et al. Evidence-based management of toxic multinodular goiter (Plummer's disease). *World J Surg*. 2008;32:1278–1284.
126. Nygaard B, Hegedus L, Ulriksen P, et al. Radioiodine therapy for multinodular toxic goiter. *Archiv Internal Med*. 1999;159:1364–1368.
127. Papini E, Monpeyssen H, Frasoldati A, et al. 2020 European Thyroid Association clinical practice guideline for the use of image-guided ablation in benign thyroid nodules. *Eur Thyroid J*. 2020;9:172–185.
128. Berghout A, Wiersinga WM, Drexhage HA, et al. Comparison of placebo with L-thyroxine alone or with carbimazole for treatment of sporadic non-toxic goitre. *Lancet*. 1990;336:193–197.
129. Wesche MF, Tiel V, Lips P, et al. A randomized trial comparing levothyroxine with radioactive iodine in the treatment of sporadic nontoxic goiter. *J Clin Endocrinol Metabol*. 2001;86:998–1005.
130. Grussendorf M, Reiners C, Paschke R, et al. Reduction of thyroid nodule volume by levothyroxine and iodine alone and in combination: a randomized, placebo-controlled trial. *J Clin Endocrinol Metabol*. 2011;96:2786–2795.
131. Surks MI, Ortiz E, Daniels GH, et al. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *J Am Med Assoc*. 2004;291:228–238.
132. Fast S, Bonnema SJ, Hegedus L. The majority of Danish nontoxic goitre patients are ineligible for Levothyroxine suppressive therapy. *Clin Endocrinol*. 2008;69:653–658.
133. Ito Y, Uruno T, Nakano K, et al. An observation trial without surgical treatment in patients with papillary microcarcinoma of the thyroid. *Thyroid*. 2003;13:381–387.

134. Hegedus L, Nygaard B, Hansen JM. Is routine thyroxine treatment to hinder postoperative recurrence of nontoxic goiter justified? *J Clin Endocrinol Metabol*. 1999;84:756–760.
135. Erickson D, Gharib H, Li H, et al. Treatment of patients with toxic multinodular goiter. *Thyroid*. 1998;8:277–282.
136. al-Suliman NN, Ryttov NF, Qvist N, et al. Experience in a specialist thyroid surgery unit: a demographic study, surgical complications, and outcome. *Euro J Surg*. 1997;163:13–20.
137. Thomusch O, Machens A, Sekulla C, et al. Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg*. 2000;24:1335–1341.
138. Torre G, Borgonovo G, Amato A, et al. Surgical management of substernal goiter: analysis of 237 patients. *Am Surg*. 1995;61:826–831.
139. Hegedus L, Bonnema SJ. Approach to management of the patient with primary or secondary intrathoracic goiter. *J Clin Endocrinol Metab*. 2010;95:5155–5162.
140. Pappalardo G, Guadalaxara A, Frattaroli FM, et al. Total compared with subtotal thyroidectomy in benign nodular disease: personal series and review of published reports. *Eur J Surg*. 1998;164:501–506.
141. Nygaard B, Hegedus L, Gervil M, et al. Radioiodine treatment of multinodular non-toxic goitre. *BMJ*. 1993;307:828–832.
142. Bonnema SJ, Bertelsen H, Mortensen J, et al. The feasibility of high dose iodine 131 treatment as an alternative to surgery in patients with a very large goiter: effect on thyroid function and size and pulmonary function. *J Clin Endocrinol Metab*. 1999;84:3636–3641.
143. Parle JV, Maisonneuve P, Sheppard MC, et al. Prediction of all-cause and cardiovascular mortality in elderly people from one low serum thyrotropin result: a 10-year cohort study. *Lancet*. 2001;358:861–865.
144. Huysmans DA, Hermus AR, Corstens FH, et al. Large, compressive goiters treated with radioiodine. *Ann Intern Med*. 1994;121:757–762.
145. Jarlov AE, Hegedus L, Kristensen LO, et al. Is calculation of the dose in radioiodine therapy of hyperthyroidism worth while? *Clin Endocrinol (Oxf)*. 1995;43:325–329.
146. Nygaard B, Faber J, Hegedus L. Acute changes in thyroid volume and function following 131I therapy of multinodular goitre. *Clin Endocrinol (Oxf)*. 1994;41:715–718.
147. Ron E, Doody MM, Becker DV, et al. Cancer mortality following treatment for adult hyperthyroidism. Cooperative thyrotoxicosis therapy follow-up study group. *J Am Med Assoc*. 1998;280:347–355.
148. Metso S, Jaatinen P, Huhtala H, et al. Increased cardiovascular and cancer mortality after radioiodine treatment for hyperthyroidism. *J Clin Endocrinol Metab*. 2007;92:2190–2196.
149. Huysmans DA, Nieuwlaat WA, Erdtsieck RJ, et al. Administration of a single low dose of recombinant human thyrotropin significantly enhances thyroid radioiodide uptake in nontoxic nodular goiter. *J Clin Endocrinol Metab*. 2000;85:3592–3596.
150. Nielsen VE, Bonnema SJ, Boel-Jorgensen H, et al. Recombinant human thyrotropin markedly changes the 131I kinetics during 131I therapy of patients with nodular goiter: an evaluation by a randomized double-blinded trial. *J Clin Endocrinol Metab*. 2005;90:79–83.
151. Fast S, Nielsen VE, Grupe P, et al. Optimizing 131I uptake after rhTSH stimulation in patients with nontoxic multinodular goiter: evidence from a prospective, randomized, double-blind study. *J Nucl Med*. 2009;50:732–737.
152. Albino CC, Mesa Jr CO, Olandoski M, et al. Recombinant human thyrotropin as adjuvant in the treatment of multinodular goiters with radioiodine. *J Clin Endocrinol Metab*. 2005;90:2775–2780.
153. Nieuwlaat WA, Hermus AR, Sivro-Prndelj F, et al. Pretreatment with recombinant human TSH changes the regional distribution of radioiodine on thyroid scintigrams of nodular goiters. *J Clin Endocrinol Metab*. 2001;86:5330–5336.
154. Bonnema SJ, Nielsen VE, Boel-Jorgensen H, et al. Recombinant human thyrotropin-stimulated radioiodine therapy of large nodular goiters facilitates tracheal decompression and improves inspiration. *J Clin Endocrinol Metab*. 2008;93:3981–3984.
155. Nielsen VE, Bonnema SJ, Hegedus L. Effects of 0.9 mg recombinant human thyrotropin on thyroid size and function in normal subjects: a randomized, double-blind, cross-over trial. *J Clin Endocrinol Metab*. 2004;89:2242–2247.
156. Fast S, Hegedus L, Grupe P, et al. Recombinant human thyrotropin-stimulated radioiodine therapy of nodular goiter allows major reduction of the radiation burden with retained efficacy. *J Clin Endocrinol Metab*. 2010;95:3719–3725.
157. Hegedus L, Miyauchi A, Tuttle RM. Nonsurgical thermal ablation of thyroid nodules: not if, but why, when, and how? *Thyroid*. 2020;30:1691–1694.
158. Hegedus L. Thyroid ultrasound. *Endocrinol Metabol Clinic North Am*. 2001;30:339–360, viii–ix.

78

Differentiated Thyroid Cancer – Streamlining Diagnosis and Optimizing Management

Amanda La Greca and Bryan R. Haugen

OUTLINE

Epidemiology, 1298
Pathology, 1298
  Papillary Thyroid Carcinoma Variants, 1299
Follicular Thyroid Carcinoma, 1300
Hurthle Cell (Oncocytic) Tumors, 1301
Poorly Differentiated Thyroid Carcinoma, 1301
Causes, 1301
  Somatic Gene Alterations, 1301
  Germline Genetic Factors, 1303
  Ionizing Radiation, 1304
Diagnosis, Clinical Features, And Course, 1304
Staging, Risk Of Recurrence, And Response To Therapy, 1306
  Response to Therapy, 1306
Prognostic Factors, 1308
  Age and Gender, 1308
  Histopathological Factors, 1308
  Size of the Primary Tumor and Extrathyroidal Extension, 1308
  Extrathyroidal Extension, 1308
  Lymph Node Metastases, 1308
Therapy And Monitoring, 1308
  Radioiodine Therapy, 1309
  Levothyroxine Therapy, 1309
  Monitoring for Disease Recurrence, 1310
  Treatment of Metastatic Disease, 1310

EPIDEMIOLOGY

Thyroid cancer is the fifth most common cancer in women, contributing 6.3% of the total number of new cases diagnosed in 2018. In the United States, the estimated incidence of new cases of thyroid carcinoma for 2020 was 52,890, representing 2.9% of all new cancer cases. The rate of death from thyroid cancer has been stable at 0.4%, suggesting that the increased incidence is primarily the result of diagnosis of subclinical, indolent tumors. The median age at diagnosis is approximately 45 to 50 years old, and differentiated thyroid carcinoma (DTC) is diagnosed two to four times more frequently in women than in men.[1]

The incidence of thyroid cancer was relatively stable until the early 1990s, after which it increased substantially.[2] Since 1975, the incidence of thyroid cancer has nearly tripled, from 4.9 to 14.3 per 100,000 individuals (absolute increase, 9.4 per 100,000). Almost the entire increase was attributable to papillary thyroid cancer (PTC), from 3.4 to 12.5 per 100,000 (absolute increase, 9.1 per 100,000),[2] suggesting that the increase is the result of diagnosis of subclinical, indolent tumors. The size distribution of detected thyroid cancer has shifted toward smaller lesions. In 1988 to 1989, 25% of detected thyroid cancers were 1 cm or smaller. In the most recent data (2008–2009), 39% were 1 cm or smaller.[2]

Autopsy series have shown that up to 6% of thyroid glands in autopsied adults in the United States, and over 20% in Japan, also harbor microscopically detectable foci of thyroid carcinoma, probably of no clinically significance.[3,4]

Compared with other malignancies, thyroid cancer has a high rate of cure with very high long-term survival rates, such as 98.3% 5-year relative survival.[1] For PTC, the most frequent thyroid malignancy, the 5-year relative survival rate for localized PTC is nearly 100%, for regional PTC is 99%, for patients who present with distant is metastasis is 76%, and for all Surveillance, Epidemiology, and End Results (SEER) stages combined is near 100%.[5]

KEY POINTS

- The incidence of thyroid cancer has nearly tripled, from 4.9 to 14.3 per 100,000 individuals. Almost the entire increase was attributable to small papillary thyroid cancer. The rate of death from thyroid cancer has been stable at 0.4%, suggesting that the increased incidence is primarily the result of diagnosis of subclinical, indolent tumors.
- The median age at diagnosis of thyroid cancer is approximately 45 to 50 years old, and differentiated thyroid cancer is diagnosed two to four times more frequently in women than in men.
- Compared with other malignancies, thyroid cancer has a high rate of cure with very high long-term survival rates, such as 98.3% 5-year relative survival.

PATHOLOGY

Papillary and follicular carcinomas are the two most common subtypes of thyroid cancer, referred to as DTC.

The fourth edition World Health Organization (WHO) Classification of Tumors of Endocrine Organs was published in 2017, in which the new thyroid tumor classification was included.[6] Several modifications to follicular cell tumors were made (Table 78.1). The major revisions were: i) the introduction of borderline tumors (uncertain malignant potential [UMP] and noninvasive follicular thyroid neoplasm with papillary-like nuclear features [NIFTP] and hyalinizing trabecular tumors) in the thyroid tumor classification; ii) PTC comprising 15 variants, and a new (hobnail) histologic variant being included; iii) follicular thyroid carcinoma (FTC) being subdivided into three prognostic groups: minimally invasive (capsule invasion only), encapsulated angioinvasive, and widely invasive; iv) Hurthle/oncocytic cell tumors being separated from follicular adenoma (FA)/FTC in an independent chapter, as it has a different genetic profile from those of the other types of thyroid cancer; v) the diagnostic criteria of poorly differentiated carcinoma being more precisely defined and adopting the Turin consensus criteria; vi) emphasizing the prognostic value of genetic markers, such as BRAFV600E and telomerase reverse transcriptase (*TERT*) promoter mutation, in thyroid carcinoma of follicular cell origin; and vii) emphasizing the examination of the entire capsule to identify or exclude invasiveness in encapsulated thyroid tumors.[7]

Papillary Thyroid Carcinoma Variants

Fifteen variants of PTC were listed in the most recent WHO classification of thyroid tumors (Table 78.2).[6]

Conventional/Classic Type. PTC is a malignant epithelial tumor demonstrating evidence of follicular cell differentiation and a set of distinct nuclear features (nuclear folding, increased mitoses, and hyperchromasia of the nucleus; the cell nuclei have a ground-glass or "cat's eye" appearance and intracellular inclusions are common) (Fig. 78.2).

TABLE 78.1 Modified from WHO Classification of thyroid tumors (4th edition – 2017)

1. Papillary thyroid carcinoma (PTC) variants:
    - Conventional / classic variant
    - Papillary microcarcinoma
    - Encapsulated PTC – applicable to fully encapsulated classic PTC
    - Diffuse sclerosing variant
    - Tall cell variant
    - Columnar cell variant
    - Cribriform-morular variant
    - Hobnail variant – new entity

PTC
- PTC with fibromatosis/fascitiis-like stroma
- Solid/trabecular variant
- Oncocytic variant
- Spindle cell variant
- Clear cell variant
- Warthin like variant

(note that variants diffuse sclerosing, tall cell, columnar cell, hobnail and solid/trabecular are so-called aggressive variants).

2. Follicular thyroid carcinoma (FTC) groups:
    - Minimally invasive FTC – capsular invasion only
    - Angioinvasive FTC
    - Widely invasive FTC – grossly invasive
3. Hürthle cell (oncocytic) tumors as a separate entity:
    Include Hürthle cell adenoma and Hürthle cell carcinoma, previously classified as oncocytic variant of follicular adenoma and FTC. The term "Hürthle" is favored over "oncocytic"
4. Poorly differentiated thyroid carcinoma:
    Turin criteria were adopted

PTC usually does not have a true capsule (invades surrounding normal tissue, invades blood and lymph channels). Papillae, invasion, or cytological features of PTC are required according to the 2017 WHO classification (fourth edition).[6] After publication of the 2017 WHO classification, the identification of definite invasion, as well as the demonstration of well-developed nuclear features and true papillae, have become essential components in the diagnosis of PTC.[6] These criteria have become stricter to reduce overdiagnosis of indolent tumors.[2]

Papillary cancer is often multicentric and bilateral.[8,9] Lymphatic spread within the thyroid is the likely reason for the high frequency of multifocality of the tumor. Vascular invasion and distant metastases (most often to the lung and bone) are rare and account for 2% to 5% of cases.

Papillary Microcarcinomas. Papillary microcarcinomas (PMCs) are tumors less than or equal to 1 cm in diameter. In some series, PMC comprise more than half of the surgically treated patients.[2,10] They may have features of a classic papillary carcinoma, or they may appear as unencapsulated sclerotic nodules of a few millimeters, infiltrating the surrounding thyroid. Most of these small tumors are not biologically significant, considering the much lower incidence of clinically recognized cancer. Less than 1% of patients with PMC have more aggressive disease identified by the presence of distant metastases at presentation.

Encapsulated Papillary Thyroid Cancer—Fully Encapsulated Classic Papillary Thyroid Cancer. This variant is classic PTC completely surrounded by a fibrous capsule that may be intact or only focally infiltrated by the carcinoma.[6] The prognosis is excellent, with close to a 100% survival rate, even though regional lymph node metastases may be present.[11]

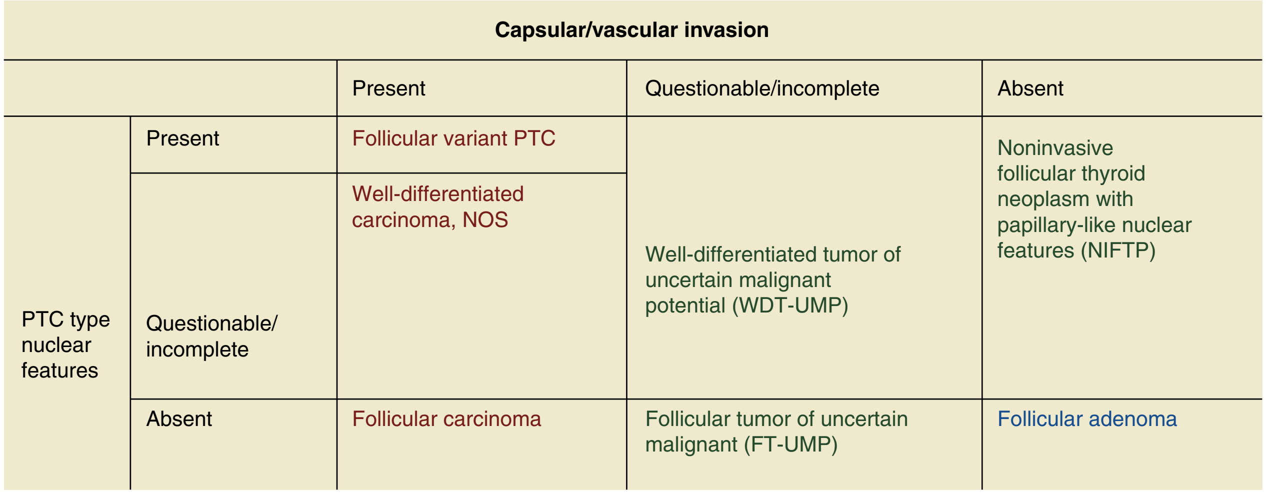
Follicular Variant Papillary Thyroid Cancer. Follicular variant papillary thyroid cancer (FVPTC) has the basic histologic structure of follicular cancers, but the characteristic nuclei of papillary cancers. These constitute the "follicular variant" of papillary cancer, and behave more or less as do other papillary cancers (low-grade tumors).[12] The variant was subclassified into encapsulated and nonencapsulated (infiltrative) variants. The encapsulated FVPTCs were further divided into invasive and noninvasive. The noninvasive encapsulated follicular variant of PTC was reclassified from carcinoma to noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP).[13] Those tumors with definite invasion remain in the category of invasive encapsulated FVPTC.

A new definitions of FVPTC, NIFTP, and well-differentiated tumor (WDT)-UMP were incorporated into the fourth edition of the WHO classification.

TABLE 78.2 Inheritance, Gene Involved, and Risks of Familiar Cancer Syndromes for Developing Thyroid Cancer

Disease	Histological Type	Gene Mutation	Location	Incidence of Thyroid Cancer	Pathological Variant of PTC
FAP and Gardner syndrome	PTC	*APC* tumor suppressor gene	5q21	2%–12%	Cribiform-morular classical variant
Cowden syndrome	FTC, PTC, C-cell hyperplasia	*PTEN* tumor suppressor gene	10q23.2	>10%	
Carney complex	FTC, PTC	*PRKAR1-x*	2p16 17q22-24	60% 4%	
Werner syndrome	FTC, PTC, ATC	*WRN* gene	Spl1-pi 2	18%	

PTC, Papillary thyroid cancer; *FAP,* familial adenomatous polyposis; *APC,* adenomatous polyposis; *FTC,* follicular thyroid cancer; *PTEN,* phosphate and tensin; *PRKAR1-x,* protein kinase A regulatory subunit type 1-alpha; *ATC,* anaplastic thyroid cancer; *WRN,* Werner.
From Son EJ, Nosé V. Familial follicular cell-derived thyroid carcinoma. *Front Endocrinol (Lausanne).* 2012;3:61.



		Capsular/vascular invasion		
		Present	Questionable/incomplete	Absent
PTC type nuclear features	Present	Follicular variant PTC	Well-differentiated tumor of uncertain malignant potential (WDT-UMP)	Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP)
	Questionable/ incomplete	Well-differentiated carcinoma, NOS		
	Absent	Follicular carcinoma	Follicular tumor of uncertain malignant (FT-UMP)	Follicular adenoma

Fig. 78.1 Red indicates malignant tumors, green for borderline tumors, and light blue for benign follicular adenoma.

Noninvasive variants of FVPTC and WDT-UMP were reclassified as NIFTP. Questionable/incomplete capsular invasion of FVPTC was merged into WDT-UMP (Fig. 78.1).[14]

Diffuse Sclerosing Variant Papillary Thyroid Cancer. This variant is rare and is estimated to account for up to 6% of all PTCs.[15] The typical histopathological characteristics include classic nuclear features of PTC, diffuse stromal fibrosis, dense lymphoid infiltration, squamous metaplasia, and abundant psammoma bodies.[16]

Diffuse sclerosing variant papillary thyroid cancer (DSVPTC) is found most often in children and young adults. Tumors are often multifocal and have more extrathyroidal extension (ETE) compared with classical PTC. Extensive lymphocytic infiltration of the thyroid gland is often found, and lymph node metastasis is present in up to 100% of cases. Disease recurrence is also more common in patients with DSVPTC than in those with PTC.[17] In a large multicenter study, the risk of recurrence was significantly higher in DSVPTC patients than for all other PTC patients (hazard ratio [HR] 8.5 [5.2; 13.9], $P < 0.0001$), but not when compared with high-risk PTC (HR 1.1 [0.6; 2.2], $P = 0.5$). Overall, the cancer-related death rate was lower in non-DSVPTC patients than in DSVPTC patients (1.8% vs. 10%, $P = 0.05$).[18]

Tall Cell Variant Papillary Thyroid Carcinoma. These tumors have a papillary pattern (similar nuclear features of PTC), and the cells are tall and columnar in shape, with a granular, eosinophilic cytoplasm.[19] According to the WHO classification, tall cell variant papillary thyroid carcinoma (TCVPTC) is a PTC composed of cancer cells that are two to three times taller than they are wide. The most recent WHO classification also defined that tall cells must account for more than 30% of all tumor cells for a diagnosis of the tall-cell variant (revised from the previous definition of more than 50% by the 2004 WHO classification).[6]

A review of the literature by Silver at al. found that the incidence of TCVPTC ranged between 5% and 11% of PTCs.[20] Several studies have shown that TCVPTC tumors had more aggressive features with advanced tumor stage at diagnosis, including higher incidence of ETE, vascular invasion, larger primary tumor size, lymph node metastases, and older patient age.[21-22] Also, the prevalence of *BRAF* and *TERT* mutations in TCVPTC is significantly higher than that in classical PTC. In a study performed by Dettmer et al., 7% of TCVPTCs had a *TERT* promoter mutation, whereas 61% demonstrated a *BRAF* mutation.[23] Patients had poorer prognosis, with higher overall mortality at 10 years (up to 22% for patients with TCVPTC, compared with 8.6% for PTC).

A large population-based study by Kazaure et al. found that rates of distant metastases in patients with TCVPTC were 2.6% higher than those for patients with PTC. Disease-specific mortality rates at 5 years were also higher for TCVPTC, at 6.4% compared with 2.3% for patients with PTC ($P < 0.001$).[24]

Columnar Cell Variant. Columnar cell variant is a rare subtype of PTC accounting for 0.15% to 0.2% of all PTCs.[25] It has tall-cell morphology without the typical nuclear cell features of PTC.[26] The prognosis is poor when it exhibits ETE, but if the tumor is encapsulated, it has indolent clinical behavior.[27]

Cribiform-Morular Variant. This variant can be sporadic, but more than half of cases occur in patients with familial adenomatous polyposis (FAP) syndrome and *APC* gene alteration.[28] The American Thyroid Association (ATA) guidelines recommend that pathologists specifically note this variant, which could lead clinicians to consider genetic counseling referral.[29] This variant usually develops multiple encapsulated nodules in the thyroid gland, and in general it has a better prognosis, with lower frequency of lymph node metastases, recurrence rates, and distant metastases.[30]

Hobnail Variant. This variant is a new entity first introduced in 2017 to the WHO classification of thyroid tumors.[6] It is a rare subtype, with a frequency of 0.15% to 0.2% of all PTCs. It is a moderately differentiated PTC with aggressive behavior and poor prognosis.[31] This variant consists of complex papillary structures covered by follicular cells that contain eosinophilic voluminous cytoplasm and large apical nuclei ("hobnails"). Hobnail variant PTC was defined with a cutoff of greater than 30%, as some hobnail features could be seen in classic PTC and DSVPTC.[32]

Solid/Trabecular Variant. This variant is diagnosed when all or nearly all of the tumor has a solid, trabecular or insular growth pattern with nuclear features of PTC.[6] This variant was frequently seen in pediatric patients following the Chernobyl nuclear accident,[33] and it was referred to as a nonaggressive variant in children. However, adults with this same variant have a slightly higher risk of recurrence and cause-specific mortality (10% at 10 years).[34]

FOLLICULAR THYROID CARCINOMA

Follicular carcinoma is typically a solitary, encapsulated nodule in the thyroid. The distinction between follicular adenoma and follicular carcinoma depends on the presence of capsular/vascular invasion. The diagnosis of malignancy is based on the demonstration of unequivocal vascular invasion and/or capsular invasion and the lack of typical nuclear features of PTC. Examination of the full capsule of the tumor is necessary to exclude the presence of invasion, therefore FTC cannot be distinguished from benign adenoma on fine-needle aspiration cytology only (Fig. 78.5). The major driver mutation is in the *RAS* gene (RAS-like tumors).[35] The 2017 WHO classification subdivided FTCs, depending on the degree of invasiveness, into three groups: (i) minimally invasive FTC (capsular invasion only), which represent more than 50% of cases, (ii) angioinvasive FTC, and (iii) widely invasive FTC (grossly invasive).[6] This classification highlights the importance of vascular classification in risk stratification of thyroid carcinomas.[36] FTC invades blood vessels but rarely invades lymphatics. Metastases are spread hematogeneously to the lungs, bones, and, less commonly, the brain and liver.[37]

HURTHLE CELL (ONCOCYTIC) TUMORS

The 2017 WHO classification has described these tumors as a separate entity, given their unique biological and clinical features and different genetic profile from FTC.[38] This entity includes Hurthle cell adenoma and Hurthle cell carcinoma (HCC). These tumors are composed of cells derived from the follicular epithelium and characterized by large size with abundant granular, eosinophilic cytoplasm resulting from the large number of mitochondria inside the cells. The cells have large nuclei and prominent nucleoli. Hurthle cells can be found in papillary carcinomas and in other benign conditions such as Hashimoto thyroiditis, benign nodules, and hyperthyroidism.

A diagnosis of HCC is based on the presence of capsular and vascular invasion, similar to FTC.[39]

POORLY DIFFERENTIATED THYROID CARCINOMA

In the new WHO classification, poorly differentiated thyroid carcinoma (PDTC) was more precisely defined with the Turin consensus criteria: (i) a diagnosis of carcinoma of follicular cell origin by conventional criteria (presence of invasion and/or metastasis); (ii) the presence of solid/trabecular/insular growth over more than 50% of the tumor area; (iii) the absence of PTC-N (nuclear features); and (iv) at least one of the following three features: convoluted nuclei (dedifferentiated nuclear feature of PTC), mitotic activity more than three per 10 high-power fields, or tumor necrosis.[40] PDTC occupies an intermediate position between differentiated (PTC and FTC) and anaplastic thyroid cancer, both histologically and clinically. Patients usually present with advanced stage at diagnosis, low radioactive iodine avidity, and frequent metastatic spread.[41]

KEY POINTS

- The fourth edition of the World Health Organization (WHO) Classification of Tumors of Endocrine Organs was published in 2017 and listed 15 variants of papillary thyroid carcinoma (PTC).
- PTC is often multicentric and bilateral. Lymphatic spread is common, while vascular invasion and distant metastases are rare.
- The noninvasive encapsulated follicular variant of PTC was reclassified from carcinoma to noninvasive follicular thyroid neoplasm with papillary-like nuclear features, and is now considered a precancerous lesion.
- The prevalence of *BRAF* and *TERT* mutations in tall cell variant PTC is significantly higher than that in classical PTC, conferring poorer prognosis with higher overall mortality.
- The distinction between follicular adenoma and follicular carcinoma depends on the presence of capsular/vascular invasion. Examination of the full capsule of the tumor is necessary to exclude the presence of invasion; therefore, follicular thyroid carcinoma (FTC) cannot be distinguished from benign adenoma on fine-needle aspiration cytology only. The major driver mutation is in the *RAS* gene (RAS-like tumors). FTC invades blood vessels but rarely invades lymphatics.
- The 2017 WHO classification has described these tumors as a separate entity, given their unique biological and clinical features and different genetic profile from FTC.

CAUSES

Somatic Gene Alterations

The main genes that play a role in cancer are oncogenes and tumor suppressor genes. Protooncogenes are genes that normally help cells grow when growth is needed. When a protooncogene mutates, whether by rearrangements, insertions/deletions, point mutations, or copy number alterations, it becomes an active oncogene and confers a growth advantage to the cell, eventually leading to malignant transformation. Another possible cause is the loss of a tumor suppressor gene. In this case, the malignant transformation occurs when these genes are inactivated.

Thyroid carcinoma is a genetically simple neoplasm with a low number of genetic alterations. It is one of better-characterized malignancies where more than 90% of cases have a driver mutation or gene fusion identified.[42] In well-differentiated thyroid cancer, driver gene mutations are generally mutually exclusive (one per tumor). Less well-differentiated thyroid cancers acquire additional genetic alterations as late events.[43] The most common initiating alterations in thyroid cancer are *BRAF* and *RAS* point mutations and *RET/PTC*, *NTRK*, and *PAX8/PPARg* chromosomal rearrangements, while *TP53* inactivation, *TERT* promoter mutations, and alterations in the PI3 kinase (PI3K) pathway are later events.

A comprehensive investigation of the genomic landscape of PTCs, the most common thyroid malignancy, was reported by The Cancer Genome Atlas Network (TCGA Network). Investigators of this group interrogated nearly 500 PTCs comprising mostly well-differentiated, low- to intermediate-risk cases.[44] These well-differentiated tumors were found to have a low frequency of somatic alterations, with the majority harboring mutually exclusive activating mutations in *BRAF* (60%) and *RAS*-family genes (13%), as well as fusion oncoproteins, primarily involving receptor tyrosine kinases (RTKs) such as *RET* (6%), *NTRK1* or *NTRK3* (2%), and *ALK* (1%). Distinct signaling and transcriptomic consequences were observed between BRAFV600E-like tumors, which showed higher mitogen-activated protein kinase (MAPK) transcriptional output and lower expression of genes involved in iodine metabolism, and RAS-like tumors, which had lower MAPK signaling and comparatively preserved expression of iodine-related genes.

BRAF-like tumors demonstrate a papillary growth pattern (conventional and tall cell variants), and have BRAFV600E mutations and rearrangement of the *BRAF*, *RET*, and *MET* genes. RAS-like tumors exhibit a follicular growth pattern with a distinct tumor capsule in more than 80% of cases, and have mutations in the *RAS*, *EIF1AX*, and *PTEN* genes, and rearrangement of the *PPARG*, *FGFR2*, and *THADA* genes. *TERT* promoter mutations at low frequency (<10%) and gene fusions of *ALK*, *NTRK1*, and *NTRK3* occur in both molecular groups.[43]

MAPK and PI3K-AKT are two main signaling pathways involved in thyroid carcinogenesis. The MAPK pathway, which is primarily involved in development of PTC, is commonly activated through point mutations in *BRAF* or *RAS*, as well as *RET/PTC* rearrangements. The PI3K-AKT pathway, which is associated with development of FTC, is activated through point mutations in *RAS*, *PIK3CA*, *AKT1*, and *PTEN*. Concurrent activation of both pathways is more frequent in more aggressive tumors.[45]

The most frequent mutational event in DTC is point mutations in the *BRAF* gene. This mutation is associated specifically with PTC, with a frequency of nearly 46%.[44] The presence of this mutation is associated with somewhat more aggressive disease (ETE, lymph node metastases), older age, and classic PTC or TCVPTC histology.

The *RAS* oncogene plays a fundamental role in human tumorigenesis. Activating mutations in its three protooncogenes—*HRAS*, *KRAS*, and *NRAS*—are found in nearly all human cancers, including thyroid cancer. Mutations in *RAS* genes occur, on average, in 30% to 45% of FTCs, 30%

to 45% of FVPTCs, 20% to 40% of PDTCs, and rarely in classical PTC. *RAS* mutations also occur in 20% to 25% of follicular thyroid adenomas. DTCs harboring *RAS* mutations alone have an excellent prognosis. Tumor aggressiveness is higher for thyroid cancers harboring *RAS* mutations and additional oncogenic mutations such as *TERT* promoter mutations, alterations of the PI3K pathway, and TP53 inactivation.[46]

The *RET* protooncogene encodes a cell membrane receptor tyrosine kinase. *RET* is highly expressed in thyroid parafollicular or C cells but not in follicular cells, where it can be activated by a chromosomal rearrangement known as the *RET/PTC* rearrangement. The two most common rearrangement types, *RET/PTC1* and *RET/PTC3*, account for most rearrangements found in thyroid cancer.[47] To date, there are more than 50 known rearrangements of *RET* in thyroid, lung, and other cancers. Although *RET/PTC* has been detected in some adenomas and other benign thyroid lesions, clonal *RET/PTC* (i.e., rearrangement that is found in a significant proportion of cells within the tumor) is reasonably specific for PTC and it is found in approximately 10% to 20% of adult sporadic papillary carcinomas. *RET/PTC1* is the most common rearrangement type (comprising up to 60%–70% of all positive cases). Tumors harboring *RET/PTC1* have a very low probability of progression to poorly differentiated and anaplastic carcinomas.[48] *RET/PETC3* is the most prevalent rearrangement in a cohort of papillary carcinomas that developed in children within 10 years of exposure to radiation after the Chernobyl accident. *RET/PTC* rearrangements occur with higher incidence in patients with a history of radiation exposure (50%–80%) and in papillary carcinomas in children and young adults (40%–70%). Papillary carcinomas with *RET/PTC* rearrangements typically present at younger age and have classic papillary histology and a high rate of lymph node metastases.[49]

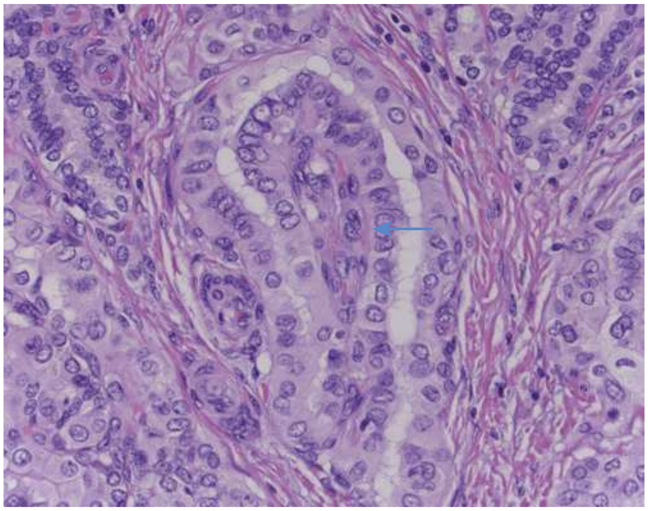


Fig. 78.2 **Classical papillary thyroid carcinoma (PTC).** A true papillary structure is present (fibrovascular core indicated with arrow) and the nuclei show enlargement, nuclear crowding and overlapping, and open pale chromatin. (Courtesy of Dr. Carrie Marshall)

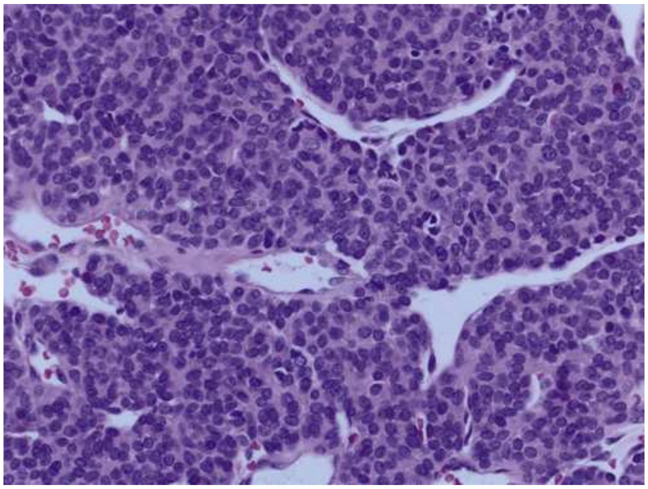


Fig. 78.4 **Poorly differentiated thyroid carcinoma.** The tumor is demonstrating a solid architectural pattern (nested/insular and trabecular architecture are also seen in poorly differentiated thyroid carcinoma) with nuclear crowding and overlapping. The tumor nuclei do not show the open and pale chromatin of papillary thyroid carcinoma. Tumors can have either round and uniform nuclei as demonstrated in this image, or have convoluted nuclear membranes. Mitotic figures and areas of necrosis are features of this diagnosis (not shown in this image). (Courtesy of Dr. Carrie Marshall)

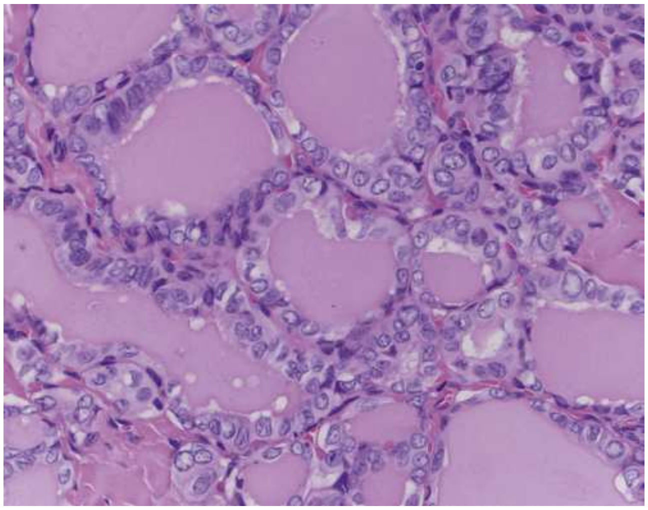


Fig. 78.3 **Follicular variant papillary thyroid cancer.** The tumor cells are arranged in follicles surrounding colloid, and no papillary structure are present. The nuclei show features that are diagnostic of papillary thyroid carcinoma (nuclear enlargement, nuclear crowding and overlapping, and open pale chromatin). (Courtesy of Dr. Carrie Marshall)

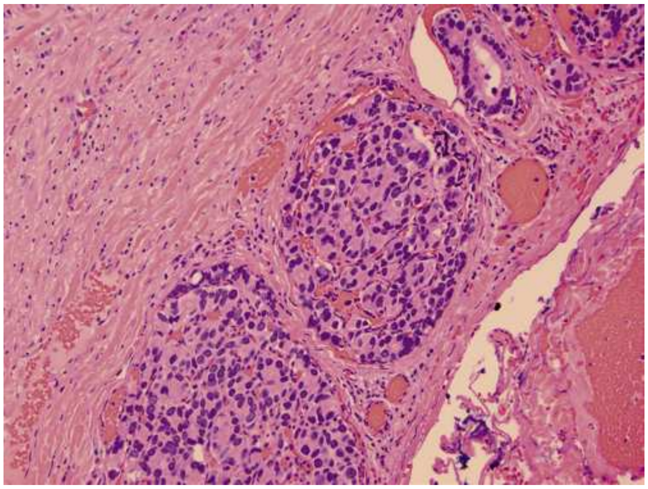


Fig. 78.5 **Follicular thyroid carcinoma with vascular invasion.** The rounded aggregates of microfollicles demonstrate thin-walled vascular spaces invaded by follicular thyroid carcinoma. The nuclei lack the characteristic changes of papillary thyroid carcinoma (pale and open chromatin, nuclear grooves, or nuclear pseudoinclusions). (Courtesy of Dr. Carrie Marshall)

The *PAX8/PPARγ* rearrangement is found in 30% to 40% of conventional-type follicular carcinomas and with lower prevalence in oncocytic carcinomas. Tumors harboring *PAX8/PPARγ* tend to present at a younger age and be smaller in size, and more frequently have vascular invasion. This rearrangement is also found in the follicular variant of papillary carcinoma, with a reported prevalence of less than 5% in most populations. This rearrangement can also be found in a small proportion (2%–13%) of follicular adenomas and Hurthle cell tumors. Detection of *PAX8/PPARγ* rearrangement in a follicular lesion is not fully diagnostic for malignancy by itself, but if exhaustive examination is performed by the pathologist, vascular or capsular invasion is frequently seen.[50]

Other somatic alterations found in PTC include *NTRK* rearrangement. It is found in fewer than 5% of papillary carcinomas in most studies. At least three types of the rearrangement exist, formed by fusion of the *NTRK1* gene to different partners.[51]

EIF1AX is mutated in only 1% of PTCs, largely occurring in a mutually exclusive manner with *BRAF* and *RAS*. *EIF1AX* mutations are predictive of worse survival in PDTCs.[43]

Progression of DTC to higher-grade cancer is associated with the accumulation of genetic alterations in other genes, such as *TP53*, *PIK3CA*, *AKT1*, *CDKN2A/CDKN2B*, and the *TERT* promoter[52] (Fig. 78.6).

TERT promoter mutations are highly prevalent in advanced thyroid cancers, particularly those harboring *BRAF* or *RAS* mutations.[52] Acquisition of a *TERT* promoter mutation could extend survival of BRAF- or RAS-driven clones and enable accumulation of additional genetic alterations leading to disease progression.[53] BRAF mutation alone is a weak independent predictor of tumor aggressiveness, and *TERT* promoter mutation alone carries a modest increased risk of recurrence and reduced survival, while both mutations together carry a significant risk of poor outcome.[54,55] An elegant mechanism of increased MAPK signaling through the *BRAF* mutation causing increased expression of GA-binding protein B, together with increased binding of GA-binding protein A to the mutated *TERT* promoter, leads to increased *TERT* expression and tumor aggressiveness, known as the "oncogenic duet."[56]

HCCs were not included in the CGA thyroid cancer study, which focused solely on PTCs. Although HCC traditionally has been classified as a subtype of FTC, emerging genomic data have validated the long-held idea that HCCs are distinct from other DTCs. A study by Ganly et al. revealed that the mutational profile of HCCs was very different from that of other types of well-differentiated thyroid cancer, such as PTC and FTC. There were no *BRAF* mutations (present in 62% of PTC), and the frequency of *NRAS* mutation was much lower than that of FTC. Genes important in PDTCs, such as TP53, *TERT* promoter, *PTEN*, *PIK3CA*, and *ATM*, were identified in the HCC cohort, but at a lower frequency.[57] This study also showed that HCCs are characterized by a high rate of somatic mitochondrial mutations (71% of HCCs) that may underlie the mitochondrial dysfunction that is the hallmark of this malignancy.

In summary, the most common oncogenic drivers in thyroid cancer include *BRAF* and *RAS* point mutations and *RET/PTC* and *PAX8/PPARγ* rearrangements. These nonoverlapping genetic alterations are found in more than 70% of PTCs and FTCs. These molecular alterations can be detected in surgically resected samples and fine-needle aspiration samples from thyroid nodules and can be of significant diagnostic, prognostic, and therapeutic utility.[47]

Germline Genetic Factors

Well-differentiated thyroid cancer accounts for 95% of thyroid malignancies, and approximately 5% of these patients will have familial disease. Such familial follicular cell-derived carcinomas are divided into two clinicopathological groups. The large majority of familial cancers (primarily papillary) occur as isolated, nonsyndromic, PTC, in which no predisposing oncogene has been detected. Familial nonmedullary thyroid carcinoma (FNMTC) syndrome is diagnosed when three or more family members have nonmedullary thyroid cancer in the absence of other known associated syndromes. This form of familial cancer has been reported in 3% to 10% of patients in different series.[58] Papillary thyroid carcinoma in FNMTC has a well-documented predisposition

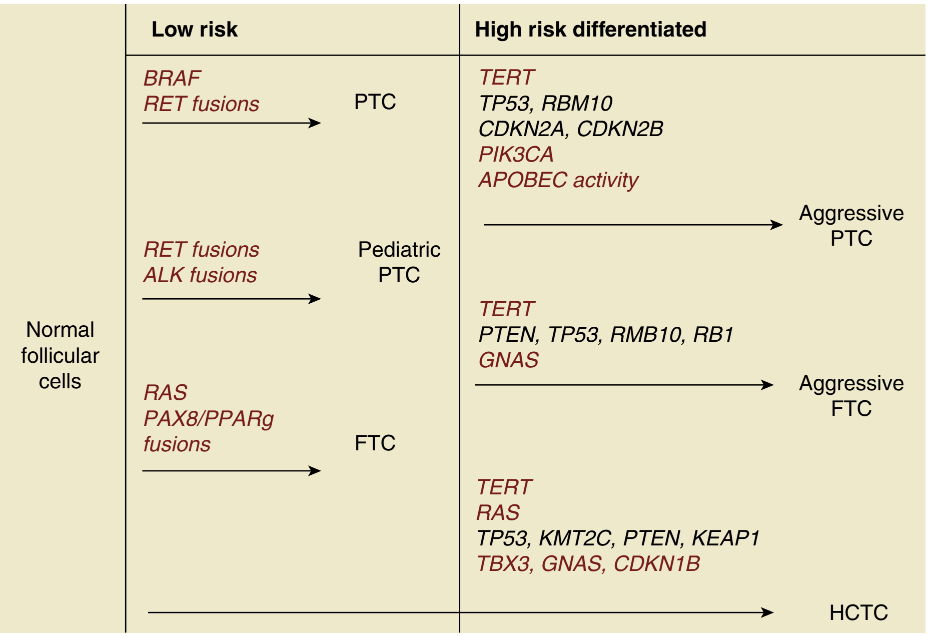


Fig. 78.6 Genetic evolution of thyroid cancer. Genetic alterations causing an increase in protein activity (activating point mutations, fusions and gene amplifications) are highlighted in red while loss-of-function mutation and copy number losses are shown in black. *PTC,* papillary thyroid carcinoma; *FTC,* follicular thyroid carcinoma; *HCTC,* Huthle cell thyroid carcinoma.

to multicentric disease, bilateral disease, local invasion, ETE, lymph node metastases, recurrence, and specific histology.

The syndromic-associated group includes FAP/Gardner syndrome, Cowden syndrome (PTEN-hamartoma tumor syndrome), Carney complex type 1, Pendred syndrome, and Werner syndrome.[59] The inheritance, gene involved, and risks of familial cancer syndromes for developing thyroid cancer are shown in Table 78.2.[60]

Familial adenomatous polyposis is inherited as an autosomal dominant trait caused by a germline mutation in the adenomatous polyposis coli (*APC*) gene. Patients typically have thousands of adenomas that are primarily located in the colon and rectum. PTC is one of the extracolonic manifestations of FAP. Young women with FAP are at particular risk of developing thyroid cancer, with a risk approximately 160 times higher than that of individuals without FAP. Among patients with FAP who have synchronous PTC, over 90% of these cases have been reported to cribriform-morular variant. Patients with cribriform-morular variant PTC should be evaluated for FAP. The ATA recommends that pathologists specifically comment on the cribiform-morular variant to aid in identifying these potential FAP patients.[29] This form of PTC is typically bilateral, presents at a younger age, and is 10 times more common in female patients with FAP. In contrast to conventional PTC, cribriform-morular variant PTC rarely metastasizes and carries a relatively benign prognosis.[62] Cowden syndrome is an autosomal dominant disorder caused by a germline mutation in *PTEN* (phosphatase and tensin homolog, deleted on chromosome 10) and characterized by the development of multiple hamartomas and carcinomas of the thyroid, breast, and uterus. Two thirds of Cowden syndrome patients develop thyroid tumors, including multinodular goiter, multiple adenomatous nodules, follicular adenoma, follicular carcinoma, and, less frequently, PTC.[61]

Follicular carcinoma is a major criterion and an important feature in *PTEN*-hamartoma tumor syndrome. These tumors are more frequently multicentric. PTC is rarely associated with this entity. Multiple adenomatous nodules in a background of lymphocytic thyroiditis and/or C-cell hyperplasia are distinctive findings in this syndrome. The Carney complex is a dominantly inherited syndrome characterized by spotty skin pigmentation, endocrine hormone overactivity, and myxomas. The presence of thyroid nodules is very common, but the risk of thyroid cancer is low (15%), and both PTC and FTC are present.[61] Werner syndrome is a rare autosomal recessive disease characterized by premature aging syndrome that typically begins in the third decade. Malignancy and cardiac disease are the most common causes of death in these patients. Patients have an approximately 3-fold increased risk for follicular carcinoma and six times increased risk for developing anaplastic thyroid carcinoma.[62]

Ionizing Radiation

A significant factor leading to thyroid cancer is ionizing radiation, particularly for papillary carcinoma, the most common thyroid cancer type. Both external radiation (x-ray and γ-radiation) and internal exposure to radioiodines through inhalation or ingestion carry an increased cancer risk.[63]

External radiation to the neck during childhood years increases the risk of PTC. The latency period between exposure and diagnosis is usually 5 years, and is maximal at approximately 20 years. A linear dose-response relationship is found between external irradiation and thyroid cancer, starting with radiation doses as low as 10 cGy and progressing up to 1500 cGy. A major risk factor is young age at the time of irradiation; after the age of 15 or 20 years, the risk is much reduced. In children exposed to a dose of 1 Gy (100 rad) to the thyroid, excess risk for thyroid cancer is 7.7-fold[64] (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2847020/).

Accidental exposure to γ-radiation and radioiodine as a result of nuclear weapon explosions or nuclear reactor accidents is another factor that increases the risk of thyroid cancer. The Chernobyl nuclear power accident in 1986 resulted in the development of cancer in more than 4000 individuals who were children or adolescents at the time of the accident.[65]

The dominant genetic alteration in radiation-induced tumors is chromosomal rearrangements, such as those leading to the formation of *RET/PTC, TRK,* and *BRAF/AKAP9* fusion oncogenes. Of note, *RET/PTC* rearrangements were found in up to 80% of radiation-induced papillary carcinomas.[66] This has been observed after accidental radiation exposure from Chernobyl, in patients treated by external beam radiation, and more recently in atomic bomb survivors in Japan. A very different mutation mechanism, i.e., point mutation, is the most prevalent in sporadic tumors, which typically carry point mutations in the *BRAF* and *RAS* genes.

Therapeutic administration of ^{131}I for hyperthyroidism (mean dose 506 MBq, mean follow-up of 15 years) was not associated with an increased incidence of thyroid cancer.[67] Kitahara and colleagues recently found that, in radioactive iodine therapy (RAI)-treated patients with hyperthyroidism, greater organ-absorbed doses appeared to be modestly positively associated with risk of death from solid cancer, including breast cancer.[68] It is postulated that the increased risk of cancer observed in this study may be related to the hyperthyroidism, and not necessarily the relatively low dose of RAI, although the two could be synergistic.

KEY POINTS

- Thyroid carcinoma is a genetically simple neoplasm with a low number of genetic alterations. It is one of the better characterized malignancies, where more than 90% of cases have a driver mutation or gene fusion identified.
- A comprehensive investigation of the genomic landscape of papillary thyroid carcinomas was reported by The Cancer Genome Atlas Network. These well-differentiated tumors were found to have a low frequency of somatic alterations, with the majority harboring mutually exclusive activating mutations in *BRAF* (60%) and *RAS*-family genes (13%).
- The most common oncogenic drivers in thyroid cancer include *BRAF* and *RAS* point mutations and *RET/PTC* and *PAX8/PPARγ* rearrangements. These nonoverlapping genetic alterations are found in more than 70% of papillary and follicular thyroid carcinomas. These molecular alterations can be of significant diagnostic, prognostic, and therapeutic utility.
- Progression of differentiated thyroid cancer to higher-grade cancer is associated with the accumulation of alterations in other genes, such as *TP53, PIK3CA, AKT1, CDKN2A/CDKN2B,* and the *TERT* promoter.
- Ionizing radiation can lead to thyroid cancer, primarily through chromosomal rearrangements.

DIAGNOSIS, CLINICAL FEATURES, AND COURSE

The most common presentation of thyroid cancer is as an incidental detection of small tumors, with widespread use of radiology studies of the head and neck. The observation that supports the subclinical disease detection hypothesis is the fact that the majority of the increase has been observed in small cancers, which cannot be palpated on exam even for most experienced clinicians. Small cancers are less likely to cause symptoms at the time of their discovery.[69]

Once the diagnosis of DTC is confirmed, the workup begins with a comprehensive history and physical examination. Neck pain, hoarseness, dysphagia, and dyspnea are rare symptoms, but these findings are suggestive of advanced thyroid cancer. Carcinoma should be suspected when the nodule is firm, when the nodule is not mobile with swallowing, when a nodule is found in children and adolescents, or if it is associated with ipsilateral enlarged lymph nodes. Rarely, distant metastases in the lung or the bone from thyroid cancer may be the initial symptom.

Measurement of thyroid-stimulating hormone (TSH) is necessary to rule out hyperfunctioning thyroid nodules, which are almost always benign nodules. Detection of thyroid autoantibodies suggests the presence of an underlying autoimmune disease. This is important to know preoperatively, as thyroglobulin (Tg) autoantibodies (TgAbs) have the potential to interfere with measurement of Tg protein, which is a valuable tumor marker for patients with DTC. Approximately 25% of DTC patients have TgAb detected in their circulation, as compared with 11% in the general population.[70]

Preoperative evaluation in patients with fine-needle aspiration revealing thyroid cancer is of extreme importance in planning initial treatment, which is mainly surgical. The ATA Guidelines published in 2015 strongly recommend preoperative neck ultrasound (US) for cervical (central and especially lateral neck compartments) lymph nodes for all patients undergoing thyroidectomy for malignant or suspicious for malignancy cytologic or molecular findings.[29] Patients with papillary cancer are expected to have clinically significant nodal disease in one third of cases. Computed tomography (CT) of the neck is particularly useful in identifying and delineating the full extent of any ETE or deep cervical lymphadenopathy. Neck CT should be considered for patients with significant adenopathy, or clinical signs or symptoms of more advanced disease (very large primary tumor, dysphagia, or hoarseness). The best CT assessment is achieved with the administration of intravenous contrast.[71] The use of contrast would delay the treatment with postoperative radioactive iodine for up to 6 to 8 weeks. This is usually the amount of time between the standard preoperative evaluation, thyroid surgery, and follow-up evaluation to determine if radioiodine therapy, adjuvant therapy, or remnant ablation is indicated. Contrast-enhanced CT is particularly helpful in nodal disease in the parapharyngeal, retrolaryngeal, retrotracheal, and retrosternal regions. This radiographic combination of US and contrast CT produces an effective preoperative nodal map that enables surgeons to direct the nodal dissection at the time of surgery.[71] Magnetic resonance imaging (MRI) can also be used to evaluate the thyroid gland and surrounding soft tissues of the neck with great resolution, although the cost of MRI is higher than that of CT.

Vocal cord paralysis or paresis may be present without significant vocal symptoms in patients with DTC. Subtle voice changes are sometimes not recognized by patients and could be difficult for clinicians to detect. Conversely, recurrent laryngeal nerve injury is one the most common causes of postoperative voice changes, and it occurs more often in surgeons who perform a low number of thyroid surgeries per year.[72] All these reasons provide a rationale for the inclusion of vocal cord examination in many or most patients. A recent American Head and Neck Society (AHNS) consensus statement recommended preoperative laryngoscopy in all patients with thyroid malignancy,[73] although the ATA guidelines recommended more selective use of the laryngeal exam for patients with preoperative voice abnormalities, history of previous cervical or upper chest surgery, or known thyroid cancer with posterior ETE or extensive cervical metastases.[29] The AHNS has concluded that flexible transnasal laryngoscopy is the optimal laryngeal examination technique.[73]

Papillary carcinomas represent the largest group of well-differentiated thyroid carcinomas with several histologic variants. Papillary carcinomas occur at any age, with highest frequency in the fourth or fifth decades of life. Clinically apparent cervical lymph node metastases are found in 12% to 14% of patients; approximately 10% have extrathyroidal invasion, and 3% to 6% have distant metastases. The most common site of ETE is the surrounding muscle (8%), recurrent laryngeal nerve (6%), and trachea (5%).[72] PTC has good prognosis in most patients and rarely causes death in patients younger than 55 years. Positive cervical nodes do not seem to carry an adverse risk in young individuals, but they do carry a higher risk of recurrence. Small-volume subclinical microscopic nodal disease conveys a much smaller risk of recurrence than large-volume, macroscopic clinically apparent locoregional metastases. The number of positive nodes and the presence of extranodal extension (median risk of recurrence 24%) are also associated with higher risk of recurrence.[74] The most common site of distant metastases is the lungs.[37,75] Pulmonary metastases most commonly present as micronodules (<1 cm), initially located at the lung bases. Children experience lymph node or pulmonary metastases more often than do adults, but mortality from thyroid cancer is extremely low in children.[76] Distant metastases occur in 1% to 25% of PTC patients, representing the lowest rate of all the well-differentiated thyroid carcinomas.[77] The American Cancer Society (ACS) estimates the overall 5-year survival for patients with PTC at greater than 99%, 99% for those with regional neck metastases, and 78% for those with distant metastases.[78]

Follicular thyroid cancers are the second most prevalent type, and they account for 10% to 15% of all thyroid cancers. They occur at an older age, with peak incidence in the fifth decade of life. They commonly present as a slowly growing thyroid mass, with direct invasion of the strap muscles and trachea occurring in up to 25% of patients. They involve cervical lymph nodes in 5% to 10% of cases, although many of these may be FVPTC. FTC has a higher rate of hematogenous spread than PTC and can present with distant metastases in 10% to 20% of patients. The most common sites of distant metastases are lungs and bone. FTC is a slightly more aggressive well-differentiated thyroid cancer than PTC. The ACS estimates 5-year survival at 98% overall, and 96% and 63% for those with locoregional disease and distant metastatic FTC, respectively.[78]

In a study from Shaha et al., adverse prognostic factors included age older than 45 years, Hurthle cell subtype, ETE, tumor size exceeding 4 cm, or the presence of distant metastasis.[79]

HCCs used to be considered a subtype of FTC, but they are now considered a separate entity, given their unique biological and clinical features and different genetic profile than FTC.[57] These tumors have a higher rate of lymph node metastases than FTC. They also have the highest incidence of distant metastases of all the well-differentiated thyroid carcinomas. In a cohort of 988 patients with well-differentiated thyroid cancer, 34% of patients with HCC had distant metastases.[80]

KEY POINTS

- The most common presentation of thyroid cancer is as an incidental detection of small thyroid cancers, with widespread use of radiology studies of the head and neck.
- Preoperative evaluation in patients with fine-needle aspiration is very important for planning initial treatment, which is primarily surgical.
- Vocal cord paralysis or paresis may be present without significant vocal symptoms in patients with differentiated thyroid carcinoma.
- Positive cervical nodes do not seem to carry an adverse risk in young individuals, but they do carry a higher risk of recurrence.
- Hurthle cell carcinomas are now considered a separate entity, given their unique biological and clinical features and different genetic profile than follicular thyroid carcinoma. These tumors have the highest incidence of distant metastases of all the well-differentiated thyroid carcinomas.

STAGING, RISK OF RECURRENCE, AND RESPONSE TO THERAPY

The staging system most often used for thyroid cancer is the American Joint Committee on Cancer (AJCC) TNM system, which is based on three key pieces of information: T (tumor) for extent (size and ETE) of the tumor, N (lymph nodes) for cervical lymph node involvement, and M for distant metastases.[81] This system is designed to predict survival in patients with cancer (see https://www.facs.org/quality-programs/cancer/ajcc for AJCC classification). Most of the changes in the eighth edition of AJCC staging (effective January 2018) were made to downstage a significant number of patients into lower stages that more accurately reflect their low risk of dying from thyroid cancer.

The main changes consisted in raising the age cutoff from 45 to 55 years of age at diagnosis. This change downstaged a large number of patients into stage I, without significantly altering the mortality associated with various stages. It is increasingly recognized that mortality increases progressively with advancing age and is considered as a continuous variable, beginning at approximately 35 years of age.[82]

The other main changes were removing microscopic ETE from the definition of T3 disease and moving regional lymph node metastases from stage III to stage II in older patients. The eighth edition also reemphasizes the critical importance of gross ETE as an unfavorable prognostic factor, while minimizing the significance of minor extension through the thyroid capsule, which is identified only on histological examination.[82] All data for the staging of a tumor should be accumulated preoperatively, intraoperatively, and during the early follow-up period after thyroid surgery. This would impact the initial therapeutic decision-making and the follow-up management strategies.

Over the past few years, the approach to patients with DTC has been changing from intense initial treatment (surgery, radioactive iodine treatment, TSH suppressive therapy) to less intensive management for lower-risk disease.

Recognizing that the commonly used clinicopathologic staging systems were designed to predict death and not recurrence, the 2009 ATA guidelines proposed the first risk stratification system to predict risk of recurrent/persistent disease in DTC, which was updated in 2015.[29,83] The risk of recurrence was divided into three categories (low, intermediate, and high), based on tumor-related parameters (TNM and histologic variant) integrated with other clinical features (the first post-therapy radioiodine whole-body scan and serum Tg measurement). Low-risk patients were defined as having intrathyroidal DTC with no evidence of ETE, vascular invasion, or metastases. Intermediate-risk patients demonstrated microscopic ETE, cervical lymph node metastases, RAI-avid disease in the neck outside the thyroid bed, vascular invasion, or aggressive tumor histology. High-risk patients had gross ETE, incomplete tumor resection, distant metastases, or inappropriate postoperative serum Tg values (Table 78.3).

In 2015, the ATA guidelines incorporated additional prognostic variables (such as the extent of lymph node involvement, mutational status, and/or the degree of vascular invasion in FTC) not included in the 2009 ATA initial risk stratification system.[29] This updated risk stratification scheme for thyroid cancer increased the number of low-risk thyroid cancer, mainly due to low-volume lymph nodal metastases being considered low risk. The other key 2015 ATA changes included that lobectomy could be considered as the initial surgical approach for follicular cell-derived thyroid cancers from 1 to 4 cm in size and surveillance following primary treatment of small PTC. In summary, the 2015 ATA guidelines emphasized the more individualized, and often less aggressive, management of thyroid nodules and DTC.

TABLE 78.3 Initial American Thyroid Association Risk of Recurrence Classification (2009 Guidelines)

Low Risk	Intermediate Risk	High Risk
All the following are present	*Any of the following is present*	*Any of the following is present*
No local or distant metastases	Microscopic invasion into the perithyroidal soft tissues	Macroscopic tumor invasion
All macroscopic tumor has been resected	Cervical lymph node metastases or ^{131}I uptake outside the thyroid bed on the post-treatment scan done after thyroid remnant ablation	Incomplete tumor resection with gross residual disease
No invasion of locoregional tissues	Tumor with aggressive histology or vascular invasion (e.g., tall cell, insular, columnar cell carcinoma, Hurthle cell carcinoma, follicular thyroid cancer)	Distant metastases
Tumor does not have aggressive histology (e.g., tall cell, insular, columnar cell carcinoma, Hurthle cell carcinoma, follicular thyroid cancer)		
No vascular invasion		
No ^{131}I uptake outside the thyroid bed on the posttreatment scan, if done		

From American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2009;19:1167–1214. Erratum in: *Thyroid.* 2010;20:942. Erratum in: Thyroid. 2010;20:674–675.

Response to Therapy

While these initial staging systems provide an important starting point for risk assessment, they are static representations of the patient at the time of initial therapy and are not designed to be modified over time based on the clinical course of the disease. In 2010, Tuttle et al. developed a new concept of "ongoing risk stratification," which better defines the patient risk based on the results of the initial treatment.[84] These concepts integrate the initial risk stratification (at the time of diagnosis) with the clinical, radiologic, and laboratory data becoming available during follow-up. In this study, which consisted of a retrospective review of 588 consecutive thyroid cancer patients followed for a median of 7 years at a single tertiary-care referral center, the ATA risk stratification system for prediction of early recurrence of disease was validated, and it was demonstrated how these initial risk estimates can be refined by incorporating response to therapy variables. The effect of therapy was integrated into the risk assessments, in which clinical data obtained during the first 2 years of follow-up were used to categorize response to therapy as excellent, acceptable, or incomplete.[84]

When patients were restratified at 6 to 12 months after initial treatment, a significant number of patients who were initially considered high-risk were reclassified as low-risk, and a large number of these patients continued to be in apparent remission up to the end of follow-up. The ongoing risk stratification allowed better modulation of the subsequent follow-up, excluding significant number of intermediate-/high-risk patients from unnecessary workup or treatment. Furthermore, the initial ATA risk estimates can be significantly refined based on the assessment of response to initial therapy, providing a dynamic risk assessment that can be used effectively for ongoing follow-up recommendations.

In 2011, the same group further subdivided the group of incomplete response to therapy into either biochemical (Tg elevations in the absence of structurally identifiable disease to widespread) or structural evidence of disease (widespread, progressive structural disease). This subdivision of the incomplete response to therapy category more precisely predicted clinical outcomes. A structural incomplete response to initial therapy was associated with significantly worse clinical outcome than a biochemical incomplete response to therapy.[85] This response to therapy classification was endorsed by the 2015 ATA guidelines, given that the risk of recurrence and disease-specific mortality can change over time as a function of the clinical course of the disease and the response to therapy.[29]

The response to therapy assessment was initially validated for dynamic risk stratification in patients with DTC treated with total thyroidectomy (TT) and RAI. A subsequent study was conducted to validate the response to therapy assessment in patients with DCT treated only with lobectomy or TT without RAI. The study demonstrated that the 2015 ATA risk stratification system effectively defined short-term risk of having recurrent/persistent disease in patients with DTC treated with total lobectomy alone or TT without RAI.[86]

Table 78. 4 summarizes the initial risk stratification of thyroid cancer, which requires the information obtained from the histologic description of the primary tumor and metastatic foci removed at the time of initial surgery and other important preoperative and intraoperative findings.[87]

KEY POINTS

- Most of the changes in the eight edition of American Joint Committee on Cancer staging were made to downstage a significant number of patients into lower stages that more accurately reflect their low risk of dying from thyroid cancer.
- Recognizing that the commonly used clinicopathologic staging systems were designed to predict death and not recurrence, the 2009 American Thyroid Association guidelines proposed the first risk stratification system to predict risk of recurrent/persistent disease in differentiated thyroid cancer, which was updated in 2015.
- The concept of "ongoing risk stratification" integrates the initial risk stratification (at the time of diagnosis) with the clinical, radiologic, and laboratory data becoming available during follow-up. This allowed better modulation of the subsequent monitoring because the risk of recurrence and disease-specific mortality can change over time as a function of the clinical course of the disease and the response to therapy.

TABLE 78.4 Response to Therapy Definition Based on Initial Therapy

	Total Thyroidectomy and RAI Ablation	Total Thyroidectomy without RAI Ablation	Lobectomy
Excellent response	Nonstimulated Tg level <0.2 ng/mL or Stimulated Tg level <1 ng/mL and Undetectable TgAb and Negative imaging	Nonstimulated Tg level <0.2 ng/mL or Stimulated Tg level <1 ng/mL and Undetectable TgAb And Negative imaging	Stable, nonstimulated Tg level <30 ng/mL and Undetectable TgAb and Negative imaging
Biochemical incomplete Response	Nonstimulated Tg level >1 ng/mL or Stimulated Tg level >10 ng mL and Increasing TgAb and negative imaging	Nonstimulated Tg level >5 ng/mL or Stimulated Tg level >10 ng mL and Increasing Tg values over time with similar TSH levels or Increasing TgAb levels and Negative imaging	Nonstimulated Tg level >30 ng/mL or Increasing Tg values over time with similar TSH levels or Increasing TgAb levels and Negative imaging
Structural incomplete response	Structural or functional evidence of disease regardless Tg or TgAb	Structural or functional evidence of disease regardless of Tg or TgAb	Structural or functional evidence of disease regardless of Tg or TgAb
Indeterminate response	Nonspecific findings on imaging studies or Faint uptake in thyroid bed on RAI scanning or Nonstimulated Tg level 0.2–1 ng/mL or Stimulated Tg level 1–10 ng/mL or TgAb levels stable or declining in the absence of structural or functional disease	Nonspecific findings on imaging studies or Faint uptake in thyroid bed on RAI scanning or Nonstimulated Tg level 0.2–5 ng/mL or Stimulated Tg level 2–10 ng/mL or TgAb levels stable or declining in the absence of structural or functional disease	Nonspecific findings on imaging studies or TgAb levels stable or declining in the absence of structural or functional disease

RAI, Radioactive iodine therapy; *Tg*, thyroglobulin; *TgAb*, thyroglobulin antibody; *TSH*, thyroid-stimulating hormone.
From Momesso DP, Tuttle RM. Update on differentiated thyroid cancer staging. *Endocrinol Metab Clin North Am*. 2014;43:401–421.

PROGNOSTIC FACTORS

Most patients with DTC have high cure rates after initial treatment, but a significant minority are at risk for recurrent disease, and a much smaller minority are at risk for cancer-related death.

Age and Gender

Age and gender are two important prognostic factors to be considered in the evaluation of thyroid cancer patients. In children and adolescents, despite a high risk of metastases and recurrent disease (up to 40%), there is an excellent prognosis, which is greater than 98% at 10 years. A study by Elisei et al., where over 4000 patients with thyroid carcinoma were studied, concluded that male gender and older age presented a significant negative influence on survival rate at univariate analysis, but in multivariate analysis, only an age older than 60 years was an independent poor prognostic factor.[88] Jonklaas et al. examined gender differences and found that women experienced a better prognosis only in univariate analysis. When the analysis also included age and disease stage, both genders had the same prognosis. The same trend was confirmed when analysis was performed distinguishing between age at diagnosis. Women diagnosed before 55 years old showed a better prognosis than men. There were no differences between sexes after age 55 years.[89] Older patients tend to have more locally aggressive tumors, a higher incidence of distant metastases at diagnosis, and more aggressive histologic variants.

Histopathological Factors

TCVPTC is considered a more aggressive variant of PTC, with a poor prognosis. This is largely due to the tendency for TCVPTC to present at an older age and with ETE, but even when the major prognostic factors for thyroid cancer are controlled for, including age and ETE, tall-cell histology alone remains a significant prognostic factor for disease-specific death.[90]

DSVPTC also exhibits aggressive biological behavior. A recent metanalysis that included 585 DSVPTC and 64,611 classical PTC patients found that, relative to patients with classical PTC, patients with DSVPTC exhibited a higher risk of ETE and lymph node and distant metastases. The risk of persistent/recurrent disease was three times higher in patients with DSVPTC than it was in classical PTC patients. The risk of cancer-related mortality was not different between DSVPTC and classical PTC.[91]

A good prognosis is generally found with the encapsulated and follicular variants of PTC.[92]

Widely invasive follicular cancers present a less favorable prognosis than minimally invasive tumors do. A recent population-based study found that the 5-year and 10-year cancer-specific survival rates for Hurthle cell carcinomas were 95.4% and 92.6%, respectively.[93] Male gender, multifocal tumors, and extended tumors were associated with increased risk of late-stage HCC. Age over 45 years, distant SEER stage, and late T-stage were independent risk factors for mortality in HCC.[93]

Size of the Primary Tumor and Extrathyroidal Extension

Papillary thyroid microcarcinoma, defined as a tumor with a larger diameter less than or equal to 1 cm, is considered to have an indolent course and an excellent prognosis, whatever the extent of the primary surgical treatment. The 2015 ATA guidelines included this entity in the low-risk category, not recommending postoperative RAI. These guidelines also recommended active surveillance management as an alternative to surgery in patients without evidence of lymph node metastasis or local invasion.[29] Several series have reported a progressive increase in risk for recurrence and tumor-specific mortality with increasing size of the primary tumor.

Extrathyroidal Extension

ETE is present in 5% to 10% of papillary tumors (approximately 10% have minimal ETE, and 4% to 5% have extensive ETE) and in 3% to 5% of follicular tumors; it exposes the patient to higher rates of local recurrence and distant metastases, as well as to a higher percentage of tumor-related death. Sosa and colleagues[94] studied more than 240,000 patients with DTC and evidence of ETE. After adjustment, minimal ETE (HR 1.13; $P < 0.01$) and extensive ETE (HR 1.74; $P < 0.01$) were associated with compromised survival for patients with DTC.

Lymph Node Metastases

Lymph node metastases are present in 10% to 80% in different series of papillary carcinoma and much less often (nearly 5% to 10%) in the follicular carcinoma. Local node involvement could be found in large tumors as well as in microcarcinomas. A study of the SEER database of 9904 patients with PTC found that the survival of patients with well-differentiated thyroid cancer is adversely affected by lymph node metastases, with an overall survival at 14 years of 82% for node-negative and 79% for node-positive patients ($P < 0.05$).[95] A more recent analysis of the National Cancer Database and SEER showed a small but significantly increased risk of death for patients younger than 45 years with lymph node metastases compared with younger patients without involved lymph nodes, and that increasing number of metastatic lymph nodes up to six involved nodes confers additional mortality risk in this age group.[96] Characteristics of the lymph node metastases can further discriminate the risk of recurrence to the patient, especially in those with small-volume microscopic disease, which conveys a much lower risk of recurrence than clinically evident locoregional metastasis, multiple, larger, and/or extracapsular nodal extension.[97]

KEY POINTS

- Age is an important prognostic factor in the evaluation of thyroid cancer patients. In children and adolescents, despite a high risk of metastases and recurrent disease, there is an excellent prognosis. Older patients tend to have more locally aggressive tumors, a higher incidence of distant metastases at diagnosis and more aggressive histologic variants. Papillary thyroid microcarcinoma is considered to have an indolent course and an excellent prognosis, whatever the extent of the primary surgical treatment. There is a progressive increase in risk for recurrence and tumor-specific mortality with increasing size of the primary tumor.
- The tall-cell variant of papillary thyroid carcinoma (PTC) is considered a more aggressive variant of PTC, with a poor prognosis.
- Lymph node metastases are present in 10% to 80% of patients with PTC. Characteristics of the lymph node metastases can further discriminate the risk of recurrence to the patient, with greater risk of recurrence when there are clinically evident locoregional metastasis, multiple, larger, and/or with extracapsular nodal extension.

THERAPY AND MONITORING

The primary goal of DTC treatment and monitoring is to ensure the most effective and least invasive treatment and monitoring plan for a disease that is often cured with surgery alone, or even active surveillance for some patients, and is rarely fatal. The 2015 ATA guidelines have summarized the basic goals of initial therapy for patients with DTC in: 1) improving the overall and disease-specific survival, 2) reducing the risk of persistent/recurrent disease and associated morbidity, and 3) permitting accurate disease staging and risk stratification, while minimizing treatment-related morbidity and unnecessary therapy.[29]

The most important initial treatment in tumors greater than 1 cm is the surgical removal of the primary tumor, disease that has extended beyond the thyroid capsule, and clinically significant lymph node metastases. Preoperative evaluation and eventual completeness of surgical resection is an important determinant of outcome, because residual metastatic lymph nodes represent the most common site of disease persistence/recurrence.[97]

When considering the extent of surgery in a patient with biopsy-proven thyroid cancer, several high-risk features (clinical N1 disease, gross ETE, recurrent laryngeal nerve involvement) should be considered. Previous guidelines have endorsed TT as the primary initial surgical treatment option for nearly all DTCs greater than 1 cm with or without evidence of locoregional or distant metastases, based on studies that suggested that TT would improve survival and decrease recurrence rates and would facilitate follow-up in low- to intermediate-risk patients by receiving RAI ablation.[83] More recent studies hat have demonstrated that, in properly selected lower-risk patients (tumor diameter ≤40 mm, no clinical lymph node metastasis or extrathyroidal invasion), clinical outcomes are very similar following unilateral or bilateral thyroid surgery.[98-103] Furthermore, the current, more selective, approach to RAI ablation in low- to intermediate-risk patients would not necessarily indicate the need for bilateral surgery in many of those patients.

The 2015 ATA guidelines recommend a near-total thyroidectomy or TT and gross removal of all primary tumor for patients with thyroid cancer greater than 4 cm, or with gross ETE (clinical T4), or clinically apparent metastatic disease to nodes (clinical N1) or distant sites (clinical M1). For patients with thyroid cancer greater than 1 cm and less than 4 cm without ETE, and without clinical evidence of any lymph node metastases (cN0), the initial surgical procedure can be either a bilateral procedure (near total thyroidectomy or TT) or a unilateral procedure (lobectomy). Thyroid lobectomy alone is likely sufficient initial treatment for low-risk papillary and follicular carcinomas. In many low-risk patients, the presence of the remaining lobe of the gland may obviate the lifelong need for exogenous thyroid hormone therapy. It is also important to note that even high-volume surgeons (>100 cases/year) have a higher overall postoperative complication rate when performing TT compared with lobectomy.[98]

If surgery is chosen for patients with thyroid cancer less than 1 cm without ETE and cN0, the initial surgical procedure should be a thyroid lobectomy, unless there are clear indications to remove the contralateral lobe (absence of prior head and neck radiation, familial thyroid carcinoma).[29]

There is no evidence that prophylactic central node dissection, without clinical evidence of nodal disease, improves recurrence or mortality rate. Microscopic nodal disease does not carry the recurrence risk of macroscopic clinically detectable disease.[97] The 2015 ATA guidelines recommend considering prophylactic central neck dissection only when patients have evidence of advanced primary tumors (T4) or clinically involved lateral neck nodes (cN1b), or if the information will be used to plan further steps in therapy.[29] Thyroidectomy or lobectomy without prophylactic central neck dissection is appropriate for small (T1 or T2), noninvasive, clinically node-negative PTC (cN0) and for most follicular cancers. TT should be accompanied by therapeutic central-compartment (level VI) neck dissection in patients with clinically involved central nodes, and therapeutic lateral neck compartmental lymph node dissection should be performed for patients with biopsy-proven metastatic lateral cervical lymphadenopathy.[29]

Radioiodine Therapy

When one considers the use of radioiodine after surgery for DTC, it is helpful to consider three basic categories of treatment: 1) therapy—the use of radioiodine to treat known residual or metastatic disease with the intent of improving survival and/or disease progression, 2) adjuvant therapy—treating a patient who will be undergoing surgery with curative intent who is at higher risk for recurrence, with the intent of lowering that risk of recurrence, and 3) remnant ablation—the intent of ablating residual normal thyroid tissue to improve the ability for disease monitoring. For the majority of DTC patients, the effect of RAI on decreasing locoregional recurrence and improving disease-specific and overall survival remains unclear. There are no reported prospective, randomized trials that evaluate the outcome of RAI in different recurrence risk categories, although there are currently two ongoing trials in lower-risk patients (IoN [NCT01398085] and ESTIMABL2 [NCT01837745]). In the last several years, increased emphasis has been placed on using individual estimates of risk to guide radioiodine use and follow-up in DTC patients. Recurrence risk stratification is a good starting point for initial decision-making.[84] Several studies have shown significant improvement in overall and disease-specific mortality, as well as disease-free survival, in stage III and IV patients; therefore, radioiodine adjuvant therapy is indicated in most high-risk patients. In a majority of low-risk patients, radioiodine ablation it is not recommended, as it is unclear if its administration has any significant benefit after complete surgical resection. Most patients can be easily monitored for disease recurrence with neck US and serum Tg, even if remnant ablation has not been used. In patients at intermediate risk, radioiodine adjuvant therapy may be considered, but the decision must be individualized.[29]

Effective treatment with radioiodine requires adequate stimulation by TSH. This may be achieved by thyroid hormone withdrawal (THW) or after recombinant human TSH (rhTSH) administration. The latter procedure is often the method of choice based on several reports demonstrating comparable efficacy compared with THW and better patient acceptance.[99] Guidelines also recommend a low iodine diet for approximately 1 to 2 weeks before RAI treatment.[29]

There is no consensus regarding the optimal administered activity RAI. Postoperative dosing regimens vary from low “ablation” doses (1.1 GBq or 30 mCi) to high “treatment” (5.5 GBq or 150 mCi) doses. The publication of two large randomized controlled trials (ESTIMABL1 and HiLo) of low-risk patients, in which the rate of successful remnant ablation was the same with either 1.1 GBq or 3.7 GBq dose activities of ^{131}I, and with either THW or with the use of rhTSH, suggested that low doses after preparation with rhTSH are a reasonable approach for remnant ablation when used in low- to intermediate-risk patients.[104,105] When considering adjuvant therapy in intermediate- to higher-risk patients, limited database studies suggest that 50 to 150 mCi ^{131}I should be considered based on individual patient and tumor characteristics.

Levothyroxine Therapy

Immediately after surgery, thyroid hormone therapy is initiated with two aims: to replace thyroid hormone and to suppress the potential growth stimulus of TSH on tumor cells (TSH suppressive therapy) in a subset of patients with higher-risk disease. Retrospective and prospective studies have demonstrated that TSH suppressive treatment to below 0.1 mU/L may improve outcomes in high-risk thyroid cancer patients reducing thyroid cancer mortality,[106,107] though no such evidence of benefit has been documented in low-risk patients. The results of these studies suggest that more aggressive TSH suppression with levothyroxine (LT4) is important in patients with high-risk disease or recurrent tumor, whereas less aggressive TSH suppression, or no suppression at all, is reasonable in low-risk patients. There are no studies that specifically examine intermediate-risk patients, because many of the older studies did not apply risk-of-recurrence stratification.

The 2015 ATA guidelines recommend TSH suppression to below 0.1 mU/L for high-risk thyroid cancer patients, TSH suppression to 0.1 to 0.5 mU/L for intermediate-risk patients, and TSH to be maintained at the lower end of the reference range (0.5– 2 mU/L) in low-risk patients. For low-risk patients who have undergone lobectomy, TSH may be maintained in the mid to lower reference range (0.5–2 mU/L). Thyroid hormone therapy may not be needed if patients can maintain their serum TSH in this target range.[29]

Monitoring for Disease Recurrence

The majority of recurrences are detected in the first 5 years after diagnosis. Late recurrences occur in a minority of cases, even 20 years after the initial treatment; however, with the use of high-resolution neck US and sensitive Tg assays for monitoring, these late recurrences are becoming less frequent. During initial follow-up, a 6- to 12-month physical examination, neck US, basal serum Tg measurement, and mandatory TgAb measurement are the most useful tools to monitor for the presence of persistent or recurrent disease. Serum Tg and TgAbs ideally should be measured using the same assay for each patient. Newer methods for serum Tg measurement with functional sensitivity at or below 0.1 ng/mL are currently available, which have replaced the need for Tg stimulation by THW or rhTSH, because stimulation does not provide additional useful information in most patients.[108]

Subsequent follow-up of patients who are considered to have no evidence of disease (excellent response) at the time of their first follow-up should consist of physical examination, basal serum Tg measurement on LT4 therapy, and neck US. Over time, the cancer-monitoring interval may be increased from yearly to every 5 years as long as the patient continues to have an excellent response to therapy.

Cervical US is highly sensitive in the detection of cervical metastases in patients with DTC. US is an operator-dependent procedure that should be performed by those with experience in identifying normal and abnormal neck lymph nodes. Following surgery, the 2015 ATA guidelines recommend that a cervical US to evaluate the thyroid bed and central and lateral cervical nodal compartments should be performed at 6 to 12 months and then periodically, depending on the patient's risk for recurrent disease and response to therapy. If a positive result would change management, ultrasonographically suspicious lymph nodes of 8 to 10 mm in the smallest diameter should be biopsied for cytology, with Tg measurement in the needle washout fluid. Suspicious lymph nodes less than 8 to 10 mm in the smallest diameter may be followed without biopsy, with consideration for fine-needle aspiration or intervention if there is growth or if the node threatens vital structures, generally in the central neck.[29] Diagnostic whole-body scan (WBS), either following THW or rhTSH, 6 to 12 months after adjuvant RAI therapy can be useful in the follow-up of patients with high or intermediate risk (higher-risk features) of persistent disease. Single-photon emission CT/CT RAI imaging is preferred over planar imaging in patients with uptake on planar imaging to better anatomically localize the RAI uptake and distinguish between likely tumors and nonspecific uptake.[29]

(^{18}F)-fluorodeoxyglucose–positron emission tomography (FDG-PET) scanning should be considered in the evaluation of high-risk DTC patients with elevated serum Tg (generally >10–20 ng/mL on LT4 therapy) with negative RAI imaging, as diagnostic of metastatic disease and also as a prognostic tool. Patients with a negative ^{131}I-WBS and positive FDG-PET have a more aggressive and less differentiated disease, with worse prognosis.[109]

CT scan of chest without intravenous (IV) contrast (imaging pulmonary parenchyma) or with IV contrast (to include the mediastinum) should be considered in high-risk DTC patients with elevated serum Tg (generally >10 ng/ mL) or rising TgAbs. Imaging of other organs including MRI brain, magnetic resonance skeletal survey, and/or CT or MRI of the abdomen should be considered in high-risk DTC patients with elevated serum Tg (generally >10 ng/mL) and negative neck and chest imaging.[29]

In Table 78.5, we summarize the use of ongoing risk stratification, modified from 2015 ATA guidelines.[29]

TABLE 78.5 Ongoing Risk Stratification—Modified from 2015 American Thyroid Association Guidelines

	Excellent	Indeterminate	Biochemical Incomplete	Structural Incomplete
TSH target	0.5–2	0.1–0.5	<0.1	<0.1
Serum Tg	Yearly	Yearly	6–12 months	3–6 months
Neck US	5 years	1–3 years	6–12 months	3–6 months
Stim Tg	No	Maybe	Consider	Consider
WBS	No	Maybe	Consider	Consider
Cross-sectional imaging	No	No	Consider	Yes

TSH, Thyroid-stimulating hormone; *Tg*, thyroglobulin; *US*, ultrasound; *WBS*, whole-body scan.

Modified from Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016;26:1–133. doi: 10.1089/thy.2015.0020.

Treatment of Metastatic Disease

Distant Metastases. Distant metastases are observed in approximately 10% of DTC patients, and half of them are detected at presentation. They are usually located in the lungs (50%), bones (25%), or lungs and bones (20%), or at other sites (5%). Distant metastases at diagnosis confer an increase in morbidity and mortality in patients with PTC or FTC, but individual prognosis depends on several factors. In a study of patients with lung metastases from DTC who were observed for 40 years, young age at diagnosis and ^{131}I uptake by metastases were the most important factors positively affecting survival time. RAI could also lead to longer survival time or complete recovery.[110] Another retrospective study showed that age of 45 years or more, symptoms, site other than lung only or bone only, and no RAI for the metastasis were predictors of poor outcome, with 13%, 11%, 16%, and 12% 10-year disease-specific survival, respectively.[112] Improved survival in patients with distant metastases is also associated with responsiveness to directed therapy (external beam radiation therapy, thermal ablation, surgery, etc.) and/or RAI.[113]

Surgery. After primary surgery, recurrences in the neck may develop in the thyroid bed and in surrounding soft tissues or in the regional lymph nodes. Any clinically detectable local biopsy-proven persistent or recurrent disease should be treated by surgery if possible, although reoperations involving central neck dissection are difficult and increase the risk of complications to the parathyroid glands and recurrent laryngeal nerve. Recurrent disease in the lateral cervical nodes is easier to treat surgically, because the operative field has not been dissected previously. The preferred surgical procedure is a modified radical neck dissection.[114] Most studies suggest that surgery results in a high clearance rate of structural disease in over 80% of patients.

Lung metastases that do not concentrate radioactive iodine could be treated surgically in very selected cases, such as a single macronodular lesion or more than one in the same lobe. Bone metastasis surgery is generally palliative, as required in pathologic fractures or to ameliorate neurologic symptoms resulting from spinal cord compression by vertebral metastases.

Radioactive Iodine. RAI could be used for regional nodal metastases discovered on diagnostic WBS, and could be employed in patients with low-volume disease or in combination with surgery. Significant variation exists in regard to RAI use, irrespective of the degree of disease or risk of recurrence.

Micronodular diffuse lung metastases shown by WBS in the absence of radiographic changes have the greatest chance of remission when treated with RAI.[115] This is important in children, who often have a diffuse pattern of metastatic pulmonary spread and do exceptionally well with RAI. Adult patients should be treated with RAI repeatedly every 6 to 12 months as long as disease continues to concentrate RAI and respond clinically; a cumulative administered activity of 600 mCi ^{131}I is a limit in many patients. Children should be treated at longer intervals, 24 to 36 months if possible, and lower cumulative administered activities. In adult patients, the treatment dose is usually 100 to 200 mCi (or 100–150 mCi for patients ≥70 years old). Lower doses (approximately 1 mCi/kg body weight) should be used in children with lung metastases, particularly of the diffuse type, to avoid the risk for radiation-induced pulmonary fibrosis.[116]

Radioiodine-avid lung macronodules may benefit from RAI, but the ability to achieve and excellent response is very low.

RAI of iodine-avid bone metastases has been associated with improved survival and should be employed, although RAI is rarely curative.[117] The treatment dose can be given empirically (100–200 mCi) or determined by dosimetry.

Radioactive-iodine refractory thyroid cancer is defined in four ways: (i) the malignant metastatic tissue does not ever concentrate RAI (no uptake outside the thyroid bed at the first therapeutic WBS), (ii) the tumor tissue loses the ability to concentrate RAI after previous evidence of RAI-avid disease (in the absence of stable iodine contamination), (iii) RAI is concentrated in some lesions but not in others; and (iv) metastatic disease progresses despite significant concentration of RAI.[118] When a patient with DTC is classified as refractory to RAI, there is generally no indication for further RAI treatment.

Directed Therapy. Several local treatment modalities to induce tumor control, other than surgery, may be used to treat distant metastatic disease in the brain, lung, liver, and bone lesions from thyroid carcinoma. The most commonly used techniques include surgery, thermal ablation (radiofrequency ablation or cryotherapy), stereotactic body radiation therapy, or intensity-modulated radiotherapy. The main principle of these techniques is to selectively treat the lesion, especially when patients are symptomatic or at risk for complications, be minimally invasive, and be well tolerated with few side effects.

Systemic Therapy. In 80% of PTCs, constitutive activation of the MAPK signaling pathway is the initiating event in the carcinogenesis of PTC. This includes *RET/PTC* rearrangement and point mutations in *RAS* and *BRAF*, with no overlap between these mutations in primary tumors. The PI3K pathway may also be activated in a few papillary and follicular cancers. Acquisition of additional mutations and gene amplifications that activate this pathway may be a common event in PDTCs. Angiogenesis is another potential molecular target for therapy. Various vascular endothelial growth factors (VEGF) and VEGF receptors (VEGFR-1 [FLT1] and VEGFR-2 [KDR]), as well as receptors for the fibroblast growth factor and for the platelet-derived growth factor, are often overexpressed in the vascular endothelium of thyroid cancer tissues, and they also trigger the MAPK signaling pathway.

Molecules that block kinase activity at distal steps in the MAPK pathway have been identified as drugs for refractory thyroid cancer. These tyrosine kinase inhibitors are multikinase inhibitors that share the ability to inhibit VEGF receptor and other receptor tyrosine kinases. Kinase inhibitor therapy should be considered in RAI-refractory DTC patients with metastatic, rapidly progressive, symptomatic, and/or imminently threatening disease not otherwise amenable to local control using other approaches.[29] Benefit has been demonstrated in the form of improved progression-free survival (delay in time to disease progression or death) in three randomized, double-blinded, placebo-controlled clinical trials of vandetanib,[119] sorafenib,[120] and lenvatinib.[121] On this basis, sorafenib and lenvatinib were approved for use in the United States and the European Union for patients with advanced RAI-refractory DTC. To date, no clinical trial has demonstrated an overall survival advantage or improved quality of life from use of any therapy in RAI-refractory DTC.[120,121]

Advances in understanding the genomic and functional alterations contributing to the pathogenesis of thyroid cancers have opened new therapeutic opportunities and are beginning to improve patient outcomes. In addition to the multiple kinase inhibitor drug approvals for each subtype of thyroid cancer, there are several US Food and Drug Administration indications that are mutation-specific: dabrafenib/trametinib for *BRAF*-mutated anaplastic thyroid cancer, larotrectinib and entrectinib for *NTRK*-fusion thyroid cancer, and selpercatinib and pralsetinib for *RET*-mutated thyroid cancer. Furthermore, other mutation-specific drugs, immunotherapies, and novel strategies for advanced thyroid cancer are under investigation[122] (Table 78.6).

KEY POINTS

- Preoperative evaluation and completeness of surgical resection are important determinants of outcome, because residual metastatic lymph nodes represent the most common site of disease persistence/recurrence.
- In properly selected lower-risk patients, clinical outcomes are very similar following unilateral or bilateral thyroid surgery.
- Radioactive iodine therapy (RAI) adjuvant therapy is indicated in most high-risk patients, given significant improvement in overall and disease-specific mortality as well as disease-free survival. In a majority of low-risk patients, RAI ablation is not recommended, as its administration has no additional benefit after complete surgical resection. In patients at intermediate risk, RAI adjuvant therapy may be considered, but the decision must be individualized.
- The large majority of recurrences are detected in the first 5 years after diagnosis.
- Patients with a negative ^{131}I–whole-body scan and a positive (^{18}F)-fluorodeoxyglucose–positron emission tomography have more aggressive and less differentiated disease, with a worse prognosis.
- Advances in understanding the genomic and functional alterations contributing to the pathogenesis of thyroid cancers have opened new therapeutic opportunities and are beginning to improve patient outcomes. These include multiple kinase inhibitor drugs for each subtype of thyroid cancer, as well as the newest therapies, which are mutation-specific.

TABLE 78.6 Molecular Markers for Therapeutic of Advanced Thyroid Cancer

Genetic Alteration	Tumor Type	Available Targeted Therapeutics
BRAFV600E	PTC	Vemurafenib, dabrafenib + trametinib
HRAS	PTC, FTC, PDTC	Farnesyltransferase inhibitor tipifamib
PAXS/PPARG	FTC	Pioglitazone
ALK fusions	PTC, PDTC	Crizotinib, cerjtinib
NTRK1/2/3 fusions	PTC, PDTC	Entrectinib, larotrectinib
RET	PTC	Vandetanib, cabozantinib, selpercatinib, pralsetinib

*Approved by the US Food and Drug Administration.

PTC, Papillary thyroid carcinoma; *FTC*, follicular thyroid carcinoma; *PDTC*, poorly differentiated thyroid carcinoma.

Modified from Cabanillas ME, Ryder M, Jimenez C. Targeted therapy for advanced thyroid cancer: kinase inhibitors and beyond. *Endocr Rev*. 2019;40:1573–1604.

REFERENCES

1. https://seer.cancer.gov/statfacts/html/thyro.html.
2. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg*. 2014;140:317–322.
3. Mortensen J, Woolner L, Bennett W. Gross and microscopic findings in clinically normal thyroid glands. *J Clin Endocrinol Metab*. 1955;15:1270.
4. Silverberg S, Vidone R. Carcinoma of the thyroid in surgical and post-mortem material. Analysis of 300 cases at autopsy and literature review. *Ann Surg*. 1966;164:291.
5. https://www.cancer.org/content/dam/CRC/PDF/Public/8855.00.pdf.
6. Lloyd RV, Osamura RY, Kloppel G, et al., eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon: IARC; 2017.
7. Kakudo K, Bychkov A, Bai Y, et al. The new 4th edition World Health Organization classification for thyroid tumors, Asian perspectives. *Pathol Int*. 2018;68:641–664.
8. Black B, Kirk TJ, Woolner L. Multicentricity of papillary adenocarcinoma of the thyroid. Influence on treatment. *J Clin Endocrinol Metab*. 1960;20:130.
9. Iida F, Yonekura M, Miyakawa M. Study of intraglandular dissemination of thyroid cancer. *Cancer*. 1969;24:764.
10. Miyauchi A. Clinical trials of active surveillance of papillary microcarcinoma of the thyroid. *World J Surg*. 2016;40:516–522.
11. Rivera M, Tuttle RM, Patel S, et al. Encapsulated papillary thyroid carcinoma: a clinico-pathologic study of 106 cases with emphasis on its morphologic subtypes (histologic growth pattern). *Thyroid*. 2009;19:119–127.
12. Passler C, Prager G, Scheuba C, etal. Follicular variant of papillary thyroid carcinoma: a long-term follow-up. *Arch Surg*. 2003;138:1362–1366.
13. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol*. 2016;2:1023–1029.
14. Kakudo K, El-Naggar AK, Hodak SP, et al. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) in thyroid tumor classification. *Pathol Int*. 2018;68:327–333.
15. Carling T, Ocal I, Udelsman R. Special variants of differentiated thyroid cancer: does it alter the extent of surgery versus well differentiated thyroid cancer? *World J Surg*. 2007;31:916–923.
16. Koo JS, Hong S, Park CS. Diffuse sclerosing variant is a major subtype of papillary thyroid carcinoma in the young. *Thyroid*. 2009;19:1225–1231.
17. Silver C, Owen R, Rodrigo J, et al. Aggressive variants of papillary thyroid carcinoma. *Head Neck*. 2011;7:1052–1059.
18. Chereau N, Giudicelli X, Pattou F, et al. Diffuse sclerosing variant of papillary thyroid carcinoma is associated with aggressive histopathological features and a poor outcome: results of a large multicentric study. *J Clin Endocrinol Metab*. 2016;101:4603–4610.
19. Ghossein RA, Leboeuf L, Patel KN, et al. Tall cell variant of papillary thyroid carcinoma without extrathyroid extension: biologic behavior and clinical implications. *Thyroid*. 2007;17:655–661.
20. Silver C, Owen R, Rodrigo J, et al. Aggressive variants of papillary thyroid carcinoma. *Head Neck*. 2011;7:1052–1059.
21. Ito Y, Hirokawa M, Fukushima M, et al. Prevalence and prognostic significance of poor differentiation and tall cell variant in papillary carcinoma in Japan. *World J Surg*. 2008;32:1535–1543.
22. Morris LG, Shaha AR, Tuttle RM, et al. Tall-cell variant of papillary thyroid carcinoma: a matched-pair analysis of survival. *Thyroid*. 2010;20:153–158.
23. Dettmer MS, Schmitt A, Steinert H, et al. Tall cell papillary thyroid carcinoma: new diagnostic criteria and mutations in BRAF and TERT. *Endocr Relat Cancer*. 2015;22:419–429.
24. Kazaure HS, Roman SA, Sosa JA. Aggressive variants of papillary thyroid cancer: incidence, characteristics and predictors of survival among 43,738 patients. *Ann Surg Oncol*. 2012;19:1874–1880.
25. Sywak M, Pasieka JL, Ogilvie T. A review of thyroid cancer with intermediate differentiation. *J Surg Oncol*. 2004;86:44–54.
26. Bongiovanni M, Mermod M, Canberk S, et al. Columnar cell variant of papillary thyroid carcinoma: cytomorphological characteristics of 11 cases with histological correlation and literature review. *Cancer Cytopathol*. 2017;125:389–397.
27. Song E, Jeon MJ, Oh HS, et al. Do aggressive variants of papillary thyroid carcinoma have worse clinical outcome than classic papillary thyroid carcinoma? *Eur J Endocrinol*. 2018;179:135–142.
28. Uchino S, Ishikawa H, Miyauchi A, et al. Age- and gender-specific risk of thyroid cancer in patients with familial adenomatous polyposis. *J Clin Endocrinol Metab*. 2016;101:4611–4617.
29. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
30. Lam AK, Saremi N. Cribriform-morular variant of papillary thyroid carcinoma: a distinctive type of thyroid cancer. *Endocr Relat Cancer*. 2017;24:R109–R121.
31. Asioli S, Erickson LA, Righi A, et al. Papillary thyroid carcinoma with hobnail features: histopathologic criteria to predict aggressive behavior. *Hum Pathol*. 2013;44:320–328.
32. Liu Z, Kakudo K, Bai Y, et al. Loss of cellular polarity/cohesiveness in the invasive front of papillary thyroid carcinoma, a novel predictor for lymph node metastasis; possible morphological indicator of epithelial mesenchymal transition. *J Clin Pathol*. 2011;64:325–329.
33. Thomas GA, Bunnell H, Cook HA, et al. High prevalence of RET/PTC rearrangements in Ukrainian and Belarussian post Chernobyl thyroid papillary carcinomas: a strong correlation between RET/PTC3 and the solid-follicular variant. *J Clin Endocrinol Metab*. 1999;84:4232–4238.
34. Nikiforov YE, Erickson LA, Nikiforova MN, et al. Solid variant of papillary thyroid carcinoma: incidence, clinical-pathologic characteristics, molecular analysis, and biologic behavior. *Am J Surg Pathol*. 2001;25:1478–1484.
35. Pozdeyev N, Gay LM, Sokol ES, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. *Clin Cancer Res*. 2018;24:3059–3068.
36. Ito Y, Hirokawa M, Masuoka H, et al. Prognostic factors of minimally invasive follicular thyroid carcinoma: extensive vascular invasion significantly affects patient prognosis. *Endocr J*. 2013;60:637–640.
37. Hugen N, Sloot YJE, Netea-Maier RT, et al. Divergent metastatic patterns between subtypes of thyroid carcinoma results from the nationwide Dutch pathology registry. *JCEM*. 2020;105:e299–e306.
38. Ganly I, Ricarte Filho J, Eng S, et al. Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid € malignancy. *J Clin Endocrinol Metab*. 2013;98:E962–E972.

39. Erickson LA, Jin L, Goellner JR, et al. Pathologic features, proliferative activity, and cyclin D1 expression in Hurthle cell neoplasms of the thyroid. *Mod Pathol*. 2000;13:186–192.
40. Volante M, Collini P, Nikiforov YE, et al. Poorly differentiated thyroid carcinoma: the Turin proposal for the use of uniform diagnostic criteria and algorithmic diagnostic approach. *Am J Surg Pathol*. 2007;31:1256–1264.
41. de la Fouchardiere C, Decaussin-Petrucci M, Berthiller J, et al. Predictive factors of outcome in poorly differentiated thyroid carcinomas. *Eur J Cancer*. 2018;92:40–47.
42. Agrawal N, Akbani R, Aksoy BA, et al. For the cancer genome atlas research network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. 2014;159:676–690.
43. Landa I, Ibrahimpasic T, Boucai L, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016;126:1052–1066.
44. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. 2014;159:676–690.
45. Xing M. Clinical utility of RAS mutations in thyroid cancer: a blurred picture now emerging clearer. *BMC Med*. 2016;14:12.
46. French CA, Alexander EK, Cibas ES, et al. Genetic and biological subgroups of low-stage follicular thyroid cancer. *Am J Pathol*. 2003;162:1053–1060.
47. Nikiforov YE. Molecular diagnostics of thyroid tumors. *Arch Pathol Lab Med*. 2011;135:569–577.
48. Tallini G, Santoro M, Helie M, et al. RET/PTC oncogene activation defines a subset of papillary thyroid carcinomas lacking evidence of progression to poorly differentiated or undifferentiated tumor phenotypes. *Clin Cancer Res*. 1998;4:287–294.
49. Zhu Z, Ciampi R, Nikiforova MN, et al. Prevalence of RET/PTC rearrangements in thyroid papillary carcinomas: effects of the detection methods and genetic heterogeneity. *J Clin Endocrinol Metab*. 2006;91:3603–3610.
50. Wood WM, Sharma V, Bauerle KT, et al. PPARγ promotes growth and invasion of thyroid cancer cells. *PPAR Res*. 2011;2011:171765.
51. Musholt TJ, Musholt PB, Khaladj N, et al. Prognostic significance of RET and NTRK1 rearrangements in sporadic papillary thyroid carcinoma. *Surgery*. 2000;128:984–993.
52. Pozdeyev N, Gay LM, Sokol ES, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. *Clin Cancer Res*. 2018;24:3059–3068.
53. Landa I, Ganly I, Chan TA, et al. Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. *J Clin Endocrinol Metab*. 2013;98:E1562–E1566.
54. Vuong HG, Altibi AM, Duong UN, et al. Role of molecular markers to predict distant metastasis in papillary thyroid carcinoma: promising value of TERT promoter mutations and insignificant role of BRAF mutations-a meta-analysis. *Tumour Biol*. 2017;39:1010428317713913.
55. Moon S, Song YS, Kim YA, et al. Effects of coexistent BRAF^{V600E} and TERT promoter mutations on poor clinical outcomes in papillary thyroid cancer: a meta-analysis. *Thyroid*. 2017;27:651–660.
56. Liu R, Zhang T, Zhu G, et al. Regulation of mutant TERT by BRAF V600E/MAP kinase pathway through FOS/GABP in human cancer. *Nat Commun*. 2018;9:579.
57. Ganly I, Makarov V, Deraje S, et al. Integrated genomic analysis of Hürthle cell cancer reveals oncogenic drivers, recurrent mitochondrial mutations, and unique chromosomal landscapes. *Cancer Cell*. 2018;34:256–270.e5.
58. Charkes ND. On the prevalence of familial nonmedullary thyroid cancer in multiply affected kindreds. *Thyroid*. 2006;16:181–186. Erratum in: Thyroid. 2006;16:520.
59. Vriens MR, Suh I, Moses W, et al. Clinical features and genetic predisposition to hereditary nonmedullary thyroid cancer. *Thyroid*. 2009;19:1343–1349.
60. Son EJ, Nosé V. Familial follicular cell-derived thyroid carcinoma. *Front Endocrinol (Lausanne)*. 2012;3:61.
61. Zambrano E, Holm I, Glickman J, et al. Abnormal distribution and hyperplasia of thyroid C-cells in PTEN-associated diseases. *Endocr Pathol*. 2004;15:55–64.
62. Ishikawa Y, Sugano H, Matsumoto T, et al. Unusual features of thyroid carcinomas in Japanese patients with Werner syndrome and possible genotype-phenotype relations to cell type and race. *Cancer Sci*. 1999;85:1345–1352.
63. Ron E, Lubin JH, Shore RE, et al. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res*. 1995;141:259–277.
64. Nikiforov YE. Is ionizing radiation responsible for the increasing incidence of thyroid cancer? *Cancer*. 2010;116:1626–1628.
65. Rabes HM, Demidchik EP, Sidorow JD, et al. Pattern of radiation-induced RET and NTRK1 rearrangements in 191 post-chernobyl papillary thyroid carcinomas: biological, phenotypic, and clinical implications. *Clin Cancer Res*. 2000;6:1093–1103.
66. Ciampi R, Knauf JA, Kerler R, et al. Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest*. 2005;115:94–101.
67. Holm LE, Hall P, Wiklund K, et al. Cancer risk after iodine-131 therapy for hyperthyroidism. *J Natl Cancer Inst*. 1991;83:1072–1077.
68. Kitahara CM, Preston DL, Sosa JA, et al. Association of radioactive iodine, antithyroid drug, and surgical treatments with solid cancer mortality in patients with hyperthyroidism. *JAMA Netw Open*. 2020;3:e209660.
69. Davies L, Morris LG, Haymart M, et al. American Association of Clinical Endocrinologists and American College of Endocrinology disease state clinical review: the increasing incidence of thyroid cancer. *Endocr Pract*. 2015;21:686–696.
70. Demers LM, Spencer CA. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid*. 2003;13:45–56.
71. Slough CM, Randolph GW. Workup of well-differentiated thyroid carcinoma. *Cancer Control*. 2006;99–105.
72. Fundakowski CE, Hales NW, Agrawal N, et al. Surgical management of the recurrent laryngeal nerve in thyroidectomy: American Head and Neck Society Consensus Statement. *Head Neck*. 2018;40:663–675.
73. Sinclair CF, Bumpous JM, Haugen BR, et al. Laryngeal examination in thyroid and parathyroid surgery: an American Head and Neck Society consensus statement: AHNS Consensus Statement. *Head Neck*. 2016;38:811–819.
74. Randolph GW, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*. 2012;22:1144–1152.
75. Durante C, Haddy N, Baudin E, et al. Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy. *J Clin Endocrinol Metab*. 2006;91:2892–2899.
76. Grigsby PW, Gal-or A, Michalski JM, et al. Childhood and adolescent thyroid carcinoma. *Cancer*. 2002;95:724–729.
77. Sebastian SO, Gonzalez JMR, Paricio PP, et al. Papillary thyroid carcinoma: prognostic index for survival including the histological variety. *Arch Surg*. 2000;135:272–277.
78. https://www.cancer.org/cancer.html#.
79. Shaha AR, Loree TR, Shah JP. Prognostic factors and risk group analysis in follicular carcinoma of the thyroid. *Surgery*. 1995;118:1131–1138.
80. Ruegemer JJ, Hay ID, Bergstralh EJ, et al. Distant metastases in differentiated thyroid carcinoma: a multivariate analysis of prognostic variables. *J Clin Endocrinol Metab*. 1988;67:501–508.
81. Tuttle M, Morris LF, Haugen B, et al. Thyroid-differentiated and anaplastic carcinoma. In: Amin MB, Edge SB, Greene F, et al., eds. *AJCC Cancer Staging Manual*. 8th ed. New York: Springer International Publishing; 2017.
82. Tuttle RM, Haugen B, Perrier ND. Updated American Joint Committee on cancer/tumor-node-metastasis staging system for differentiated and anaplastic thyroid cancer (eighth edition): what changed and why? *Thyroid*. 2017;27:751–756.

83. American Thyroid Association (ATA). Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–1214. Erratum in: Thyroid. 2010;20:942. Erratum in: Thyroid. 2010;20:674–675.
84. Tuttle RM, Tala H, Shah J, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. *Thyroid*. 2010;20:1341–1349.
85. Vaisman F, Tala H, Grewal R, et al. In differentiated thyroid cancer, an incomplete structural response to therapy is associated with significantly worse clinical outcomes than only an incomplete thyroglobulin response. *Thyroid*. 2011;21:1317–1322.
86. Momesso DP, Vaisman F, Yang SP, et al. Dynamic risk stratification in patients with differentiated thyroid cancer treated without radioactive iodine. *J Clin Endocrinol Metab*. 2016;101:2692–2700.
87. Momesso DP, Tuttle RM. Update on differentiated thyroid cancer staging. *Endocrinol Metab Clin North Am*. 2014;43:401–421.
88. Elisei R, Molinaro E, Agate L, et al. Are the clinical and pathological features of differentiated thyroid carcinoma really changed over the last 35 years? Study on 4187 patients from a single Italian institution to answer this question. *JCEM*. 2010;95:1516–1527.
89. Jonklaas J, Nogueras-Gonzalez G, Munsell M, et al. The impact of age and gender on papillary thyroid cancer survival. *J Clin Endocrinol Metab*. 2012;97:E878–E887.
90. Morris LG, Shaha AR, Tuttle RM, et al. Tall-cell variant of papillary thyroid carcinoma: a matched-pair analysis of survival. *Thyroid*. 2010;20:153–158.
91. Malandrino P, Russo M, Regalbuto C, et al. Outcome of the diffuse sclerosing variant of papillary thyroid cancer: a meta-analysis. *Thyroid*. 2016;26:1285–1292.
92. Rivera M, Tuttle RM, Patel S, et al. Encapsulated papillary thyroid carcinoma: a clinico-pathologic study of 106 cases with emphasis on its morphologic subtypes (histologic growth pattern). *Thyroid*. 2009;19:119–127.
93. Zhou X, Zheng Z, Chen C, et al. Clinical characteristics and prognostic factors of Hurthle cell carcinoma: a population based study. *BMC Cancer*. 2020;20:407.
94. Youngwirth LM, Adam MA, Scheri RP, et al. Extrathyroidal extension is associated with compromised survival in patients with thyroid cancer. *Thyroid*. 2017;27:626–631.
95. Podnos YD, Smith D, Wagman LD, et al. The implication of lymph node metastasis on survival in patients with well-differentiated thyroid cancer. *Am Surg*. 2005;71:731–734.
96. Adam MA, Pura J, Goffredo P, et al. Presence and number of lymph node metastases are associated with compromised survival for patients younger than age 45 years with papillary thyroid cancer. *J Clin Oncol*. 2015;33:2370–2375.
97. Randolph GW, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*. 2012;22:1144–1152.
98. Adam MA, Pura J, Gu L, et al. Extent of surgery for papillary thyroid cancer is not associated with survival: an analysis of 61,775 patients. *Ann Surg*. 2014;260:601–605;discussion 605–607.
99. Matsuzu K, Sugino K, Masudo K, et al. Thyroid lobectomy for papillary thyroid cancer: long-term follow-up study of 1,088 cases. *World J Surg*. 2014;38:68–79.
100. Barney BM, Hitchcock YJ, Sharma P, et al. Overall and cause-specific survival for patients undergoing lobectomy, near-total, or total thyroidectomy for differentiated thyroid cancer. *Head Neck*. 2011;33:645–649.
101. Nixon IJ, Ganly I, Patel SG, et al. Thyroid lobectomy for treatment of well differentiated intrathyroid malignancy. *Surgery*. 2012;151:571–579.
102. Mendelsohn AH, Elashoff DA, Abemayor E, et al. Surgery for papillary thyroid carcinoma: is lobectomy enough? *Arch Otolaryngol Head Neck Surg*. 2010;136:1055–1061.
103. Haigh PI, Urbach DR, Rotstein LE. Extent of thyroidectomy is not a major determinant of survival in low- or high-risk papillary thyroid cancer. *Ann Surg Oncol*. 2005;12:81–89.
104. Schlumberger M, Catargi B, Borget I, et al. Strategies of radioiodine ablation in patients with low-risk thyroid cancer. *N Engl J Med*. 2012;366:1663–1673.
105. Mallick U, Harmer C, Hackshaw A. The HiLo trial: a multicentre randomised trial of high- versus low-dose radioiodine, with or without recombinant human thyroid stimulating hormone, for remnant ablation after surgery for differentiated thyroid cancer. *Clin Oncol (R Coll Radiol)*. 2008;20:325–326.
106. Cooper DS, Specker B, Ho M, et al. Thyrotropin suppression and disease progression in patients with differentiated thyroid cancer: results from the National Thyroid Cancer Treatment Cooperative Registry. *Thyroid*. 1998;8:737–744.
107. Pujol P, Daures JP, Nsakala N, et al. Degree of thyrotropin suppression as a prognostic determinant in differentiated thyroid cancer. *J Clin Endocrinol Metab*. 1996;81:4318–4323.
108. Malandrino P, Latina A, Marescalco S, et al. Risk-adapted management of differentiated thyroid cancer assessed by a sensitive measurement of basal serum thyroglobulin. *J Clin Endocrinol Metab*. 2011;96: 1703–1709.
109. Robbins RJ, Wan Q, Grewal RK, et al. Real-time prognosis for metastatic thyroid carcinoma based on 2-[18F]fluoro-2-deoxy-D-glucose-positron emission tomography scanning. *J Clin Endocrinol Metab*. 2006;91: 498–505.
110. Eskander A, Merdad M, Freeman JL, et al. Pattern of spread to the lateral neck in metastatic well-differentiated thyroid cancer: a systematic review and meta-analysis. *Thyroid*. 2013;23:583–592.
111. Shoup M, Stojadinovic A, Nissan A, et al. Prognostic indicators of outcomes in patients with distant metastases from differentiated thyroid carcinoma. *J Am Coll Surg*. 2003;197:191–197.
112. Zettinig G, Fueger BJ, Passler C, et al. Long-term follow-up of patients with bone metastases from differentiated thyroid carcinoma -- surgery or conventional therapy? *Clin Endocrinol (Oxf)*. 2002;56:377–382.
113. Shah MD, Hall FT, Eski SJ, et al. Clinical course of thyroid carcinoma after neck dissection. *Laryngoscope*. 2003;113:2102–2107.
114. Steward DL. Update in utility of secondary node dissection for papillary thyroid cancer. *J Clin Endocrinol Metab*. 2012;97:3393–3398.
115. Ronga G, Filesi M, Montesano T, et al. Lung metastases from differentiated thyroid carcinoma. A 40 years' experience. *Q J Nucl Med Mol Imaging*. 2004;48:12–19.
116. Hebestreit H, Biko J, Drozd V, et al. Pulmonary fibrosis in youth treated with radioiodine for juvenile thyroid cancer and lung metastases after Chernobyl. *Eur J Nucl Med Mol Imaging*. 2011;38:1683–1690.
117. Bernier MO, Leenhardt L, Hoang C, et al. Survival and therapeutic modalities in patients with bone metastases of differentiated thyroid carcinomas. *J Clin Endocrinol Metab*. 2001;86:1568–1573.
118. Schlumberger M, Brose M, Elisei R, et al. Definition and management of radioactive iodine-refractory differentiated thyroid cancer. *Lancet Diabetes Endocrinol*. 2014;2:356–358.
119. Leboulleux S, Bastholt L, Krause T, et al. Vandetanib in locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 2 trial. *Lancet Oncol*. 2012;13:897–905.
120. Brose MS, Nutting CM, Jarzab B, et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet*. 2014;384:319–328.
121. Schlumberger M, Tahara M, Wirth LJ, et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *N Engl J Med*. 2015;372: 621–630.
122. Cabanillas ME, Ryder M, Jimenez C. Targeted therapy for advanced thyroid cancer: kinase inhibitors and beyond. *Endocr Rev*. 2019;40: 1573–1604.

79

Medullary Thyroid Cancer

Rossella Elisei, Cristina Romei, and Antonio Matrone

OUTLINE

Introduction, 1315
Clinical Presentation And Diagnosis, 1315
  Sporadic Medullary Thyroid Cancer, 1315
  Hereditary Forms (Multiple Endocrine Neoplasia Type II), 1317
  Gene Carriers, 1318
Pathology And Pathogenesis, 1319
  Histology, 1319
  Genetics, 1319
Initial Therapy, 1320
Postsurgical Follow-Up, 1321
  Follow-Up of Cured Medullary Thyroid Cancer Patients, 1321
  Follow-Up of Patients with Biochemical Persistent Disease, 1321
  Follow-Up of Patients with Structural Disease, 1322
Other Therapies, 1322
  Local Treatment of Recurrent or Persistent Neck Disease, 1322
Conclusions, 1324

INTRODUCTION

Two different types of cells are present in the thyroid gland: the follicular cells that secrete the thyroid hormones thyroxine (T4) and triiodothyronine (T3), and the parafollicular C cells that produce calcitonin (Ct). It is well recognized that these two thyroid cell types have unique embryonic origins, although details of their differentiation and development during organogenesis remain somewhat controversial.[1,2] Parafollicular C cells represent only 2% to 4% of the thyroid cells, which explains the low frequency of medullary thyroid carcinoma (MTC) among all thyroid cancer types.[3] However, tumors arising from the two cell types are clearly distinct entities, with different treatment and prognosis. MTC is the unique well-differentiated thyroid tumor deriving from C cells and belongs to the group of neuroendocrine tumors. In contrast, several histotypes, such as papillary, follicular, insular, poorly differentiated, and anaplastic thyroid cancer, originate from follicular epithelial cells.[4]

Because of the absence of a dedicated registry, the incidence of MTC is not defined. Its prevalence is estimated at only around 5% to 10% of all thyroid cancers and 0.4% to 1.4% of all thyroid nodules, and it is found in less than 1% of thyroids at autopsy.[5] Unlike papillary and follicular carcinomas, which are highly prevalent in females, no difference in gender distribution has been reported in MTC. The clinical appearance is mainly in the fourth and fifth decades, but there is a wide range of ages at onset.[6]

No risk factors, whether environmental or genetic, for the development of MTC have been identified, although associations with prior thyroid diseases and other unrelated disorders such as hypertension, allergies, and gallbladder disease have been reported in a metanalysis of epidemiological studies.[7]

In 1993, the pathogenesis of MTC was reported to be activation, mainly due to point mutations or small insertion/deletions, of the *RET* protooncogene.[8,9] However, it remains unclear what causes *RET* activation, which in turn sets in motion the development of MTC. *RET* activation can be responsible for both the sporadic form, which accounts for approximately 75% of cases, and the hereditary or familial form of MTC, which accounts for the remaining 25%. In sporadic cases the *RET* alteration is present only in the tumor tissue, at the somatic level, while the *RET* protooncogene alteration is transmitted in an autosomal dominant manner at the germline level.[10]

The biological behavior of MTC is much less favorable when compared with that of the other well-differentiated thyroid carcinomas, although it is not as unfavorable as that of anaplastic carcinoma.[11] While in older series a 10-year survival of approximately 50% was frequently reported for patients with MTC,[12] most recent series, including ours, show better overall survival, with 80% to 85% of patients surviving longer than 10 years, likely thanks to improvements in both diagnostic procedures and care (Fig. 79.1). Survival is dependent on several factors, such as age at diagnosis,[13] but the stage at diagnosis remains the most relevant clinical prognostic factor for survival in these patients. When an early diagnosis is made, and the tumor is still intrathyroidal, 90% of patients can survive up to 35 years.[12,14]

KEY POINTS

- MTC is a rare endocrine tumor that can be either sporadic (75%) or familial (25%). Males and females are equally affected. Children are affected only in the familial forms. No risk factors are known. An early diagnosis is desirable, since only intrathyroidal MTC can be completely cured with surgery.

CLINICAL PRESENTATION AND DIAGNOSIS

Sporadic Medullary Thyroid Cancer

The most common clinical presentation of sporadic MTC is a thyroid nodule, either as a single nodule or inside a multinodular goiter. Only in very advanced cases, diarrhea and/or flushing syndrome, due to high levels of serum Ct, can represent the first symptoms.[15] In 5% to 10% of very advanced cases, symptoms due to ectopic adrenocorticotropic hormone (ACTH) production and a consequent paraneoplastic Cushing syndrome with severe hypercortisolism can be present.[16] The simultaneous presence of a (multi)nodular thyroid and a lump in one of the laterocervical regions of the neck may lead the clinician to suspect a thyroid malignancy, but not specifically MTC.

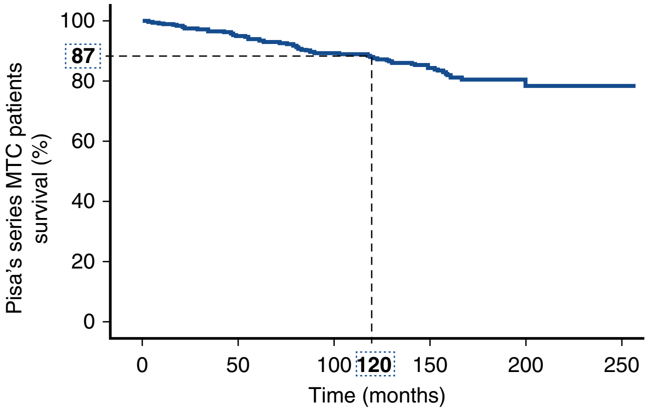


Fig. 79.1 Survival curve of 538 medullary thyroid cancer (MTC) patients followed from 2000–2018 at the University Hospital of Pisa, Italy. In this relatively recent series the survival rate at 10 years from diagnosis was 87%.

TABLE 79.1 Procedures for Calcium Stimulation Test for Serum Calcitonin in Patients with Mildly Elevated Values (Between 20 and 100 pg/mL)

Steps:

1) Patient must be weighed (do not trust what the patient says)
2) 2.5 mg of calcium element (or 25 mg of calcium gluconate) per 1 kg or 2.2 lbs of weight should be taken in a syringe
3) The volume of calcium collected in step 2 must be diluted up to 50 mL with standard saline solution
4) The intravenous infusion of the 50 mL containing calcium must be rapidly administered (in 5 minutes)
5) Blood samples must be taken at baseline before the infusion and then 2, 5, 15 minutes after the complete infusion of the 50 mL containing calcium for calcitonin measurements

NOTES Be careful to collect samples for calcitonin measurement from an intravenous access other than that used for the infusion, or aspirate at least 10 mL of liquid to be drained and then aspirate the next sample for calcitonin measurement to avoid dilution of the sample due to the liquid present in the drip tube.

A classical work-up for thyroid nodular disease is then performed.[17] Neck ultrasound (US) is the first imaging to be performed to describe thyroid nodule features. MTC nodules are usually solid and hypoechoic, sometimes with microcalcifications. A few studies have clearly demonstrated that US can only identify some malignant features and suggest a general risk of malignancy without any specificity for MTC.[18,19] Thyroid scintigraphy, either with ^{131}I or Tc99, is commonly performed only if there is a large multinodular goiter, to map the entire thyroid and the several nodules, or if the thyrotropin-stimulating hormone level is low, suggesting hyperfunction of the gland/nodule. If performed, MTC appears as a cold nodule at neck scintigraphy.

Fine-needle aspiration cytology (FNAC) is the gold standard procedure for determining whether thyroid nodules are benign or malignant. In a typical MTC cytological smear, cells are usually isolated, with shape varying from oval to round, large polygonal or spindled. Cytoplasm may be abundant or scant and usually contains acidophilic granulation visible with specific stains (May-Grunwald-Giemsa). Nuclei are frequently multiple (i.e., two or three), preferentially round, and eccentrically localized.[17] Amyloid is frequently detectable as clumps of amorphous material and revealed by Red Congo staining[20]: when present, it is highly suggestive of MTC. Immunocytochemistry for Ct and/or chromogranin, another neuropeptide secreted by MTC, is not routinely performed, but can be carried out if there is suspicion for MTC.[21,22] Despite this apparently typical MTC cytological pattern, several series have shown a high rate of failure in making a cytological presurgical diagnosis.[23,24] Missed presurgical diagnosis of MTC can lead to incorrect initial surgical treatment, which in principle should be total thyroidectomy and central neck node dissection. Recently, a multicentric international study involving 12 different referral centers across seven different continents and four nations showed that only 46% of 245 MTC cases received a correct cytological diagnosis of MTC, demonstrating that this is a worldwide pathology issue.[25]

Among other explanations, FNAC-negative results might be due to the fact that MTC could be present in a nodule not subjected to FNAC, especially when multinodular goiter is the clinical diagnosis. In these cases, serum Ct measurement is more reliable, because this hormone is elevated even in the presence of MTC microfoci.[23,26] However, there is some controversy in the different guidelines regarding routine measurement of serum Ct in all patients with thyroid nodule(s). The British Thyroid Association[27] and the National Comprehensive Cancer Network[28] share the view that serum Ct should be measured any time a presurgical diagnosis of MTC is suspected due to positive FNAC, familial history, or the simultaneous presence of other endocrine neoplasia. The American Thyroid Association guidelines for the management of thyroid nodules[29] do not recommend for or against routine serum Ct testing. In contrast, the European Thyroid Association (ETA) consensus[30] clearly states that this procedure should be applied in the work-up of all thyroid nodules, because it is more sensitive than FNAC. In 2010, a new document signed by the American Association of Clinical Endocrinologists, the Associazione Medici Endocrinologici, and the ETA was published, declaring that routine serum Ct may be useful, especially before surgery, and strongly recommending Ct measurement in certain high-risk groups.[31]

The major concerns about the routine measurement of serum Ct in all patients with thyroid nodules are related to both the possible "false-positive" cases and the "cost-benefit" analysis. There is a real risk that mildly elevated serum Ct, between 20 and 100 pg/mL, could be due to C cell hyperplasia (CCH) or other neuroendocrine tumors, or even to assay artifact rather than authentic MTC.[32,33] However, serum Ct values in this "grey zone" do raise suspicion for MTC, even though they are not diagnostic, and other diagnostic tools should be used to rule out this diagnostic doubt. Subjects with mildly elevated basal serum Ct could undergo a calcium stimulation test[34,35] following a well-defined procedure, as reported in Table 79.1. The degree of Ct increase can distinguish different clinical scenarios, because only serum Ct produced by C cells will significantly increase after stimulation, while serum Ct due to artifacts, such as heterophilic antibodies, or produced by other neuroendocrine tumors does not respond to the stimulus.[36,37] Another useful diagnostic tool to apply when serum Ct is mildly elevated is measuring Ct in the washout fluid from the needle used for FNAC.[38] This approach is of diagnostic utility to ascertain the nature not only of the thyroid nodule, if MTC is suspected, but also of neck lymph nodes, especially before thyroidectomy, to plan the surgical approach or the most appropriate therapeutic strategy. Moreover, in the presence of mildly elevated serum Ct, the pathologist can be alerted and asked to perform immunocytochemistry for Ct, which, as previously mentioned, is not a routine procedure. Although the routine measurement of serum Ct in all subjects with thyroid nodules is still controversial,[39] there is evidence that this approach allows early diagnosis and treatment, thus significantly improving the outcome of this potentially lethal disease.[40]

KEY POINTS

- Both sporadic and familial MTC commonly appears as a thyroid nodule; the cytological diagnosis of MTC is feasible in only 50% of cases. The measurement of serum calcitonin is recommended in the presence of a thyroid nodule. Mildly elevated serum calcitonin is not diagnostic of MTC but should raise the suspicion and induce the use of other diagnostic tools.

MTC, like many other neuroendocrine tumors, can produce other peptides than Ct, such as chromogranin, somatostatin, gastrin-releasing peptide, vasoactive intestinal peptide, neuron-specific enolase, serotonin, prostaglandins, and other neuroendocrine substances, but none of them are useful for presurgical diagnosis.[41] When elevated, they may all contribute to flushing and diarrhea syndrome.[42]

Serum carcinoembryonic antigen (CEA) is usually elevated when the disease is diffuse and distant metastases are present.[43] In the few cases of dedifferentiated MTC that are unable to secrete Ct, CEA becomes the serum marker of the disease.[44] Nevertheless, it is more useful in monitoring disease progression because its level increases when the disease becomes rapidly progressive.

Because at least 5% to 7% of apparently sporadic cases of MTC are indeed hereditary cases,[45,46] preoperative evaluation of both the adrenal and parathyroid morphology and function should always be performed. The familial history should also be carefully considered, with particular regard to the occurrence of pheochromocytoma (PHEO) and hyperparathyroidism (HyperPTH) in other family members. Nowadays, all subjects with MTC, independently ascertained to be sporadic or familial, must undergo *RET* genetic screening, which can be easily performed in a genetic laboratory by analyzing the sequence of *RET* gene in DNA extracted from the patient's blood or saliva.[10]

KEY POINTS

- All MTC patients must undergo the RET genetic screening, since about 8% of apparently sporadic cases are indeed hereditary. All first-degree relatives of a MTC patient carrying a germline RET mutation should be submitted to the RET genetic screening.

Hereditary Forms (Multiple Endocrine Neoplasia Type II)

In approximately 25% of cases, MTC is one of the components of multiple endocrine neoplasia type II (MEN II), which is an autosomal dominant inherited syndrome with a variable degree of expressivity and an age-related penetrance. Three different hereditary syndromes can be distinguished, depending on the involved glands (Fig. 79.2): 1) multiple endocrine neoplasia type IIA (MEN IIA), a syndrome consisting of MTC, PHEO, and HyperPTH due to parathyroid multiple adenomatosis/hyperplasia[47]; 2) multiple endocrine neoplasia type IIB (MEN IIB), a syndrome consisting of MTC, PHEO, mucosal neuromas, ganglioneuromatosis, marfanoid features, skeletal alterations, and megacolon[48]; and 3) familial MTC (FMTC), which is characterized by the presence of heritable MTC with no apparent association with other endocrine neoplasia.[49] After the introduction of the *RET* genetic screening, the relative prevalence of FMTC syndrome was found to be much higher than expected (accounting for 10%–50% of all MEN syndromes). The increase in FMTC is mainly due to the high number of apparently sporadic MTCs demonstrated to be familial cases by *RET* mutation analysis.[46]

The clinical course of MTC varies considerably in the three syndromes. It is very aggressive and almost invariably unfavorable in MEN IIB, with affected patients rarely surviving past adolescence.[48] In contrast, MTC is almost indolent in most patients with the FMTC form, and shows variable degrees of aggressiveness in patients with MEN IIA. Different types of *RET* gene mutations account for different biological behavior, and separate therapeutic protocols have been defined for the treatment of MTC occurring in the three different syndromes.[50] Between 10% and 30% of patients with MEN IIA develop HyperPTH during the third or fourth decade of life. The clinical findings are superimposable on those of the sporadic form of HyperPTH, and very often no specific symptoms are noted. Unlike the sporadic form, multiple hyperplasia or adenomatosis of parathyroid glands is most commonly found.[47,51] HyperPTH has only occasionally been reported in patients with MEN IIB.[52] Very recently, a peculiar case of simultaneous MEN IIA and IIB clinical features has been described.[53]

Approximately 50% of MEN IIA and 40% to 45% of MEN IIB patients develop PHEO, which shares the same characteristics in both syndromes. Contrary to the sporadic form of PHEO, the adrenal

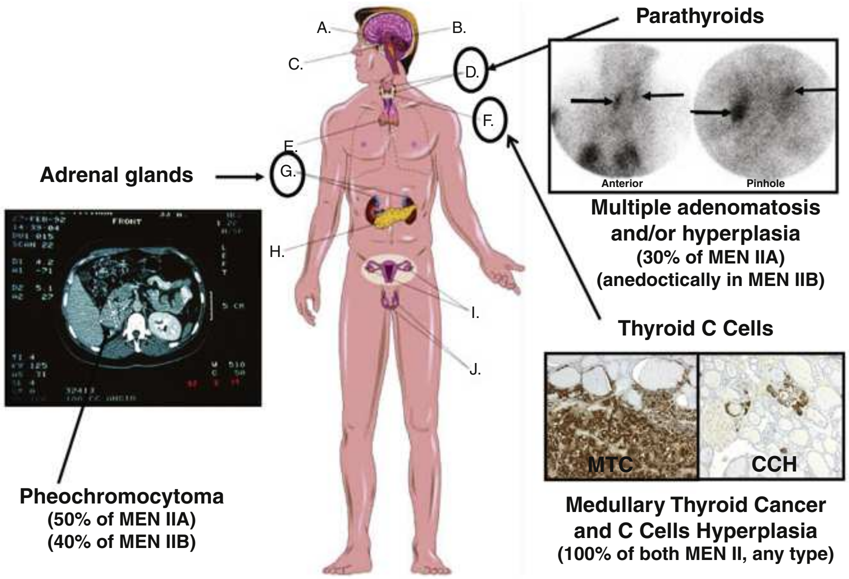


Fig. 79.2 Different endocrine glands' involvement in the three different phenotypes of multiple endocrine neoplasia type II (MEN II).

tumors seen in MEN syndromes are usually bilateral and multicentric. However, the two adrenal glands are rarely simultaneously involved, and a mean period of 10 years usually elapses between the development of the tumor in the two glands.

MEN IIB is characterized by association with mucosal neuromas, which are mainly located on the distal tongue and subconjunctival areas, and ganglioneuromatosis affecting the gastrointestinal tract, which is responsible for the very commonly seen symptom of megacolon. MEN IIB patients should be easily recognized at physical examination by the typical marfanoid habitus characterized by thin and inappropriately long extremities and pectus excavatum.[48] Thick lips and thick eyelids are frequently observed in the presence of mucosal neuromas and are usually clearly evident when the eyes and mouth are explored. Gastrointestinal disorders due to intestinal neuromas throughout the intestinal tract, including obstructive symptoms, cramping, and diarrhea, are frequently observed in early childhood. However, despite these very pathognomonic features, the diagnosis is usually late, when MTC has already spread to lung and/or liver.[54]

An association with cutaneous lichen amyloidosis (CLA), a characteristic pigmented and itchy skin lesion specifically localized in the interscapular region of the back, has been reported in less than 10% of MEN IIA families.[55] The development of CLA may precede the development of MTC. For this reason, when present it is almost invariably diagnostic of MEN IIA and may be considered a predictor of the syndrome. Four specific *RET* mutations in exon 10 (i.e., C620, C618, C611, and C609), collectively referred to as "Janus," are responsible for both MEN IIA and Hirschsprung disease. This latter syndrome is commonly due to *RET* mutations located in the first exons of the gene that inactivate the gene. Historically these mutations were believed to be able to act simultaneously as both gain-of-function mutations, responsible for the MEN IIA phenotype, and loss-of-function mutations, responsible for the intestinal aganglionosis.[56] Recently, however, it has been demonstrated that *RET* overstimulation alone is enough to cause both phenotypes.[57]

The hereditary nature of MTC may be suspected based on a positive family history (other members already affected), or based on the presence of other endocrine neoplasia (PHEO and/or HyperPTH) or other disorders (neuromas, marfanoid features, CLA). However, at least in the Pisa series,[10] index cases of 40% of all hereditary MTC kindreds presented as apparently sporadic cases, and their hereditary origin was only discovered through *RET* screening.

The clinical appearance of MTC in MEN II syndromes is that of a thyroid nodular disease, similar to that of the sporadic form, with the exception that it is usually bilateral and multicentric and is almost invariably associated with CCH.[58] Evaluation of the thyroid nodule in the hereditary form follows the same work-up as that recommended for sporadic cases, but the hereditary forms require mandatory simultaneous examination of adrenal and parathyroid glands. With a few exceptions,[59] development and diagnosis of PHEO usually follow development and diagnosis of MTC. Symptoms of PHEO are not specific and may be confused with those caused by anxiety. Hypertension is rare, especially at the beginning of the disease. The first alteration of catecholamine production is represented by an elevation of the daily urinary excretion of epinephrine, while norepinephrine usually increases only later. Nowadays the measurement of urinary metanephrines, the O-methylated metabolites of catecholamines, is used for the early diagnosis of PHEO, because they are much more stable than catecholamines.[60] Tests for metanephrines are more specific and sensitive than those for catecholamines: while normal concentrations of metanephrines exclude the diagnosis of PHEO, normal concentrations of catecholamines do not.[61] Once the biochemical suspicion of PHEO has arisen, an abdomen US with or without computed tomography (CT) and/or magnetic resonance imaging (MRI) may be useful for localization of the adrenal mass. If there is no demonstrable adrenal mass by CT or MRI, patients should undergo functional imaging, typically positron emission tomography (PET), which can be performed with different tracers (i.e., (^{18}F)-fluorodeoxyglucose, 68Ga-DOTATATE, or (^{18}F)-dihydroxyphenylalanine [DOPA]).[62]

Parathyroid glands may also be involved in hereditary MTC, particularly in MEN IIA. Both adenomas and hyperplasia may be associated with an increase in parathyroid hormone (PTH) secretion, resulting in hypercalcemia and hypercalciuria in more advanced cases.[63] The earliest serum abnormality detected is a moderately elevated level of serum PTH, with normal-high levels of serum calcium levels. A deficit of vitamin D could also be responsible for PTH increase, and for this reason its level should always be verified to exclude secondary HyperPTH due to vitamin D deficit.[64]

Once the MTC case has been proven to be hereditary, based on the presence of a germline *RET* mutation, all first-degree relatives should undergo testing to search for the same genetic alteration.[65] This procedure is fundamental for the early discovery of gene carriers who should undergo clinical and biochemical evaluation immediately after discovery of the mutation in their DNA.

Gene Carriers

By definition, gene carriers are family members of an index patient with hereditary MTC that are potentially already affected or may develop MTC in the future. The diagnosis is performed through *RET* screening of all first-degree relatives of the index case. The mutational analysis can be limited to the *RET* mutation found in the index case: informed consent must be provided for the genetic analysis, and in the case of children either both parents or guardians should sign the consent form and agree that they will inform the child at a more appropriate age about their positive or negative status as a gene carrier.[65]

All first-degree adult relatives should undergo *RET* genetic screening as soon as possible, while the timing of testing children can vary based on the type of *RET* mutation: as soon as possible if the mutation is the M918T mutation typical of MEN IIB, before 5 years of age if the mutation involves a cysteine residue, particularly at codon 634, and even later in cases with any other type of non-cysteine *RET* mutations.[50]

Once the gene carrier has been identified, the diagnostic procedure involves a thyroid neck US and a serum Ct measurement. Some family members will already be unknowingly affected and must be immediately treated, while others will still be negative for both parameters. The timing of total thyroidectomy in these latter cases should be personalized, taking into account the type of *RET* mutation and also the levels of serum Ct.[50,66]

In the 1990s, soon after the discovery of the pathogenic relationship between the presence of a germline *RET* mutation and the development of MTC, prophylactic thyroidectomy was suggested for all gene carriers of any age.[67] Over the years, increasing knowledge of the aggressiveness of different mutations and the awareness that serum Ct highly correlates with the presence of MTC led to rethinking of this approach and the development of a more relaxed algorithm for treatment of these subjects, who can safely undergo early, rather than prophylactic, thyroidectomy.[65] Annual evaluation of basal serum Ct and a neck US should be performed, taking into account that only microMTC and/or CCH, with no evidence of metastatic lymph nodes, have been reported for serum Ct levels less than 30 to 40 pg/mL in large series of gene carriers.[66,68] In this regard, it is worth noting that central neck compartment dissection can be avoided in gene carriers with serum Ct les than 30 to 40 pg/mL who undergo thyroidectomy, especially children, thus reducing the risk of postsurgical hypoparathyroidism.[50]

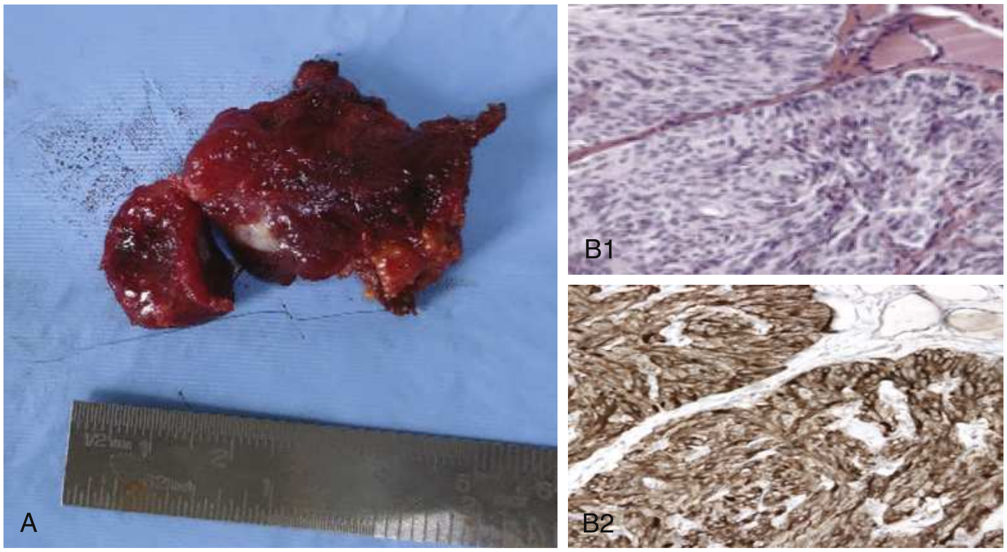


Fig. 79.3 Macroscopic (**A**) and microscopic (**B1–2**) histology of medullary thyroid cancer. **A,** The tumor appears as a solid and compact, red mass. **B1,** Hematoxylin-eosin microscopic appearance of medullary thyroid cancer. **B2,** The same section after calcitonin immunohistochemistry.

KEY POINTS

- Prophylactic thyroidectomy are not mandatory in gene carriers with no evidence of MTC, especially in children. Serum calcitonin and the type of RET mutation should guide the timing of thyroidectomy.

PATHOLOGY AND PATHOGENESIS

Histology

The macroscopic histological appearance of MTC is that of a solid and compact, red or calcareous-white nodule (Fig. 79.3A). Microscopically it is pleiomorphic, with spindle-shaped or rounded cells typically organized in a nested pattern (Fig. 79.3B1–2). Nuclei are uniform, mitoses are rare, and the cytoplasm is characterized by the presence of secretory granules. Amyloid substance can be present among MTC cells, and when present, it allows the diagnosis. Sometimes MTC can be erroneously diagnosed as anaplastic thyroid carcinoma, Huerthle cell carcinoma, or insular differentiated thyroid carcinoma, especially if pseudopapillary aspects or giant cells are present. Positive immunohistochemistry for Ct, and also for chromogranin A and CEA, is undoubtedly diagnostic of MTC.[69] The histological description of MTC should include the number and distribution of tumoral foci. CCH, either diffuse or focal, should also be reported because, although it is the histological hallmark of the hereditary forms, it is also present in approximately 13% of sporadic cases.[70] With the exception of the prevalence of CCH, no other major histological differences have been described between the sporadic and the hereditary forms of MTC.

Genetics

In 1994, after discovering the transforming role of the *RET* oncogene, the correlation between the presence of a germline activating mutation of *RET* and the hereditary form of MTC was demonstrated in almost 500 kindreds affected by MEN II or FMTC.[71] In the same study, somatic *RET* mutations were reported in approximately 40% of sporadic MTCs. Thereafter, both *in vitro* and *in vivo* studies confirmed the driver role of *RET* mutations in the development of MTC.[72,73]

Germline *RET* point mutations are mainly localized in exons 10 to 11 and 13 to 16, but other rare mutations have been reported in other exons, such as 5 and 8, and all of these exons must be analyzed as part of the genetic screening procedure.[65] A significant genotype–phenotype correlation has been reported over the years, with some *RET* mutations being almost exclusively associated with a specific MEN II syndrome.[74] With a few exceptions, known as "variants of unknown significance,"[75] the penetrance of *RET* mutations is complete, in that all gene carriers will develop the disease sooner or later. In contrast, the level of expressivity is different, in that MTC due to different *RET* mutations can have different degree of aggressiveness based on the transforming ability, which is the highest for the M918T mutation.[75] According to these differences, the *RET* mutations have been distinguished in three risk levels (i.e., moderate, high, highest) of aggressiveness, which should be taken into account to identify the right time for the surgical treatment of *RET* gene carriers.[76]

Somatic *RET* mutations are mainly concentrated in exon 16, and have been reported in other exons in only a few cases. Moreover, *RET* somatic mutation prevalence is significantly higher in larger MTC tumors than in smaller MTC tumors,[77] and is particularly high in advanced and progressive MTC.[78] Although not yet introduced into routine clinical practice, it would be valuable to know if a patient with sporadic MTC is carrying a *RET* somatic mutation, as mutated cases have a worse prognosis[79]; thus, a more aggressive therapeutic strategy or a more stringent follow-up could be provided for these patients. Moreover, new *RET*-targeted therapies are under development and require the characterization of the molecular profile of the MTC.[80,81]

RAS oncogene somatic mutations, in particular in *H-RAS* and *K-RAS*, are present in approximately 10% to 20% of cases of sporadic MTC, and they are mutually exclusive with *RET* mutations.[82] No other driver mutations have been discovered in *RET*- and *RAS*-negative MTCs, even when tissues were analyzed with next-generation sequencing techniques.[83,84] These studies did identify very rare mutations, but their prevalence was so low that they are considered "private" mutations that are likely relevant only in that specific case.[85] A similar finding has been reported for hereditary cases: the *ESR2* gene, which encodes the beta subunit of the estrogen receptor, has been found to be mutated in two individuals with FMTC but not in their unaffected relatives. However, the same *ESR2* mutation was not found in other

RET-negative FMTC kindreds, thus suggesting that it was a "private mutation" of the first family analyzed.[86] The next challenge is to identify the driver mutations in both *RET*-negative familial cases and *RET*- and *RAS*-negative sporadic cases.

KEY POINTS

- With a few exceptions, hereditary MTC are positive for a germline RET mutation. Almost 50% of sporadic MTC are positive for RET mutation at somatic level. Another 30% are positive for H- or K-RAS somatic mutations. The remaining 20% of sporadic cases are negative for any other known oncogene.

INITIAL THERAPY

Total thyroidectomy and central node compartment dissection is the minimal surgical treatment, even in intrathyroidal cases. The need for total thyroidectomy is related to the multicentricity and bilaterality of the MTC, which occur in 17% and 5.6%, respectively, of cases with the sporadic form.[87] Moreover, CCH, which is a preneoplastic lesion, is almost invariably associated with the hereditary form of MTC (75%) and, to a lesser extent, with the sporadic form (13%).[70] An additional factor in favor of total thyroidectomy is the fact that, as mentioned above, 5% to 7% of apparently sporadic cases are in fact hereditary forms and almost invariably have bilateral disease.[45,46]

Total thyroidectomy and central neck dissection will be extended to the laterocervical compartment(s) if metastatic lymph nodes are recognized in these regions. For this reason, an accurate neck US should be performed in order to identify suspicious metastatic lymph nodes to be subjected to FNAC, as well as measurement of Ct in the washout fluid from the needle used.[38] However, not all authors agree with this strategy, and some prefer to plan extension of surgery based on the presurgical Ct serum level, independent of the results from the neck US.[88] In this regard it is worth noting that, unfortunately, when the MTC has metastasized to neck lymph node, especially to the laterocervical ones, it is very difficult to achieve definitive surgical and biochemical cure of the disease.[89]

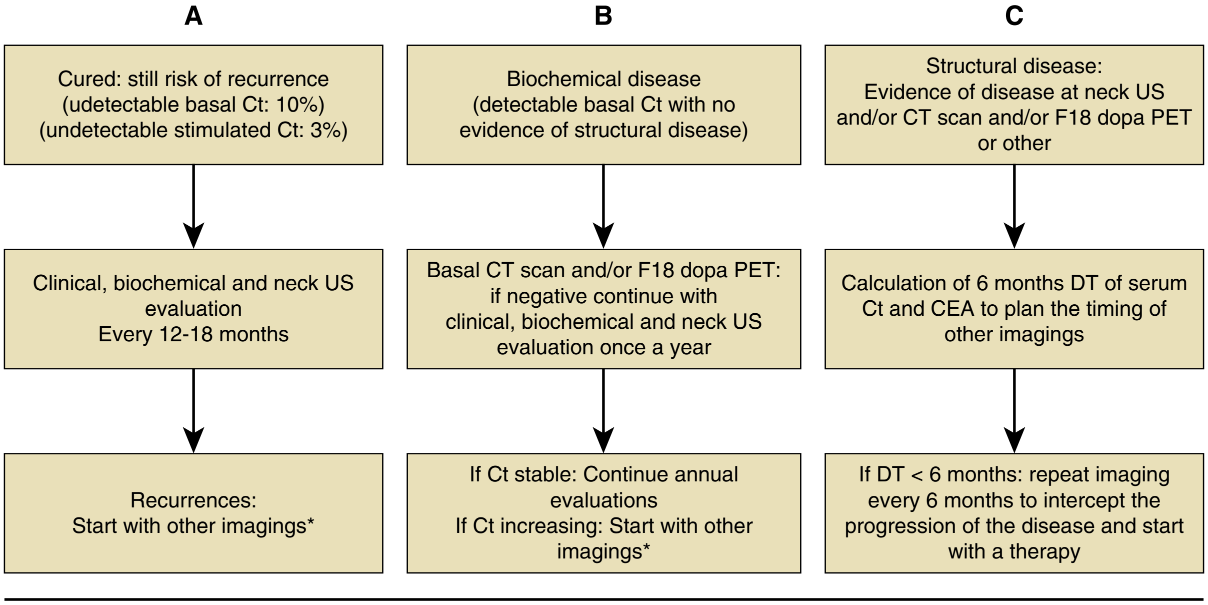
Measurement of serum CEA is also indicated in the preoperative phase, because elevated levels are strongly suggestive of advanced disease. Cases with advanced local disease demonstrated by neck US and associated with elevated serum CEA levels should be studied with computerized tomography (CT) scan to better evaluate the local extension of the tumor and to verify the relationship of the disease with the gross veins, trachea, and esophagus and plan the most appropriate surgical treatment.[50]

KEY POINTS

- The initial therapy consisted of at least total thyroidectomy and lymph node dissection of the central neck compartment. The lymphadenectomy of the latero-cervical compartments should not be prophylactically performed but only when metastatic latero-cervical metastatic lymphnodes are present at the time of the diagnosis or when presurgical serum calcitonin is highly elevated.

Surgical removal of the primary tumor and neck lymph nodes metastases is also recommended in cases with distant metastases already present at diagnosis.[50] In these latter cases, and in cases with tumor invasion of the vital structures of the neck, a multidisciplinary meeting must be held before surgery to plan for additional therapies.[90]

Before surgery, an evaluation of the adrenal gland function to exclude the presence of a PHEO must always be performed, except in genetically documented sporadic MTC. The reason for this procedure is in the need to perform the PHEO excision before the MTC surgery, in cases of the simultaneous presence of both tumors, to avoid unexpected hypertensive crisis during the MTC surgery.


A
B
C
Cured: still risk of recurrence
(udetectable basal Ct: 10%)
(undetectable stimulated Ct: 3%)
Biochemical disease
(detectable basal Ct with no
evidence of structural disease)
Structural disease:
Evidence of disease at neck US
and/or CT scan and/or F18 dopa PET
or other
Clinical, biochemical and neck US
evaluation
Every 12-18 months
Basal CT scan and/or F18 dopa PET:
if negative continue with
clinical, biochemical and neck US
evaluation once a year
Calculation of 6 months DT of serum
Ct and CEA to plan the timing of
other imagings
Recurrences:
Start with other imagings*
If Ct stable: Continue annual
evaluations
If Ct increasing: Start with other
imagings*
If DT < 6 months: repeat imaging
every 6 months to intercept the
progression of the disease and start
with a therapy

* imagings are usually negative until serum Ct < 150 pg/ml

Ct: calcitonin; CEA: carcino embryonic antigen; CT: computerized tomography; PET: positron emission tomography; US: ultrasound; DT: doubling time

Fig. 79.4 Algorithm for follow-up of medullary thyroid cancer (MTC) patients after initial surgical treatment based on the results from the first evaluation. **A,** Follow-up of patients cured of MTC whose risk of recurrence varies from 3% to 10%. **B,** Follow-up of MTC patients with persistent biochemical disease. **C,** Follow-up of MTC patients with persistent structural disease.

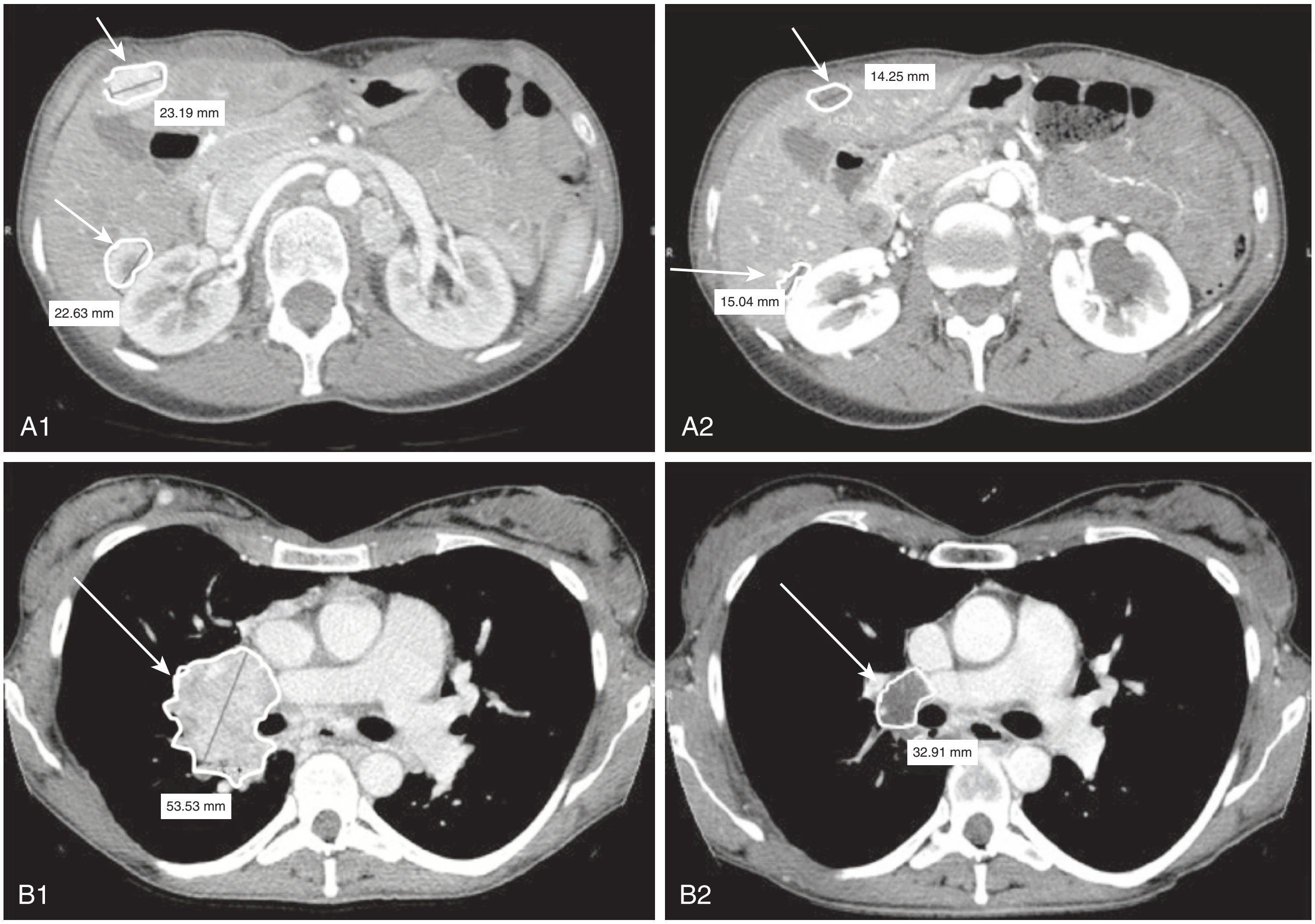


Fig. 79.5 Two cases of medullary thyroid cancer metastatic disease treated with vandetanib (**A**) and cabozantinib (**B**). **A,** Liver metastatic lesions *(arrows)* before vandetanib treatment (**A1**) and after 6 months of therapy (**A2**). **B,** Right metastatic hilar lymph node *(arrows)* before cabozantinib treatment (**B1**) and after 2 months of therapy (**B2**). In both cases, significant shrinkage of the lesions is visible, and their devascularization is demonstrated by the significantly lower density of the lesions after the treatments.

POSTSURGICAL FOLLOW-UP

The first evaluation after surgery should be performed 3 months after the surgical treatment, including physical examination, neck US, and measurement of serum free T3 (FT3), free T4 (FT4), thyroid-stimulating hormone (TSH), Ct, and CEA. Measurement of FT3, FT4, and TSH is requested for monitoring levothyroxine replacement therapy. Serum Ct and CEA measurement and neck US are necessary for the follow-up of MTC. Due to the relatively long half-lives, if performed too early, serum values of Ct may be misleading, especially if a high serum concentration was present preoperatively.[91]

According to the results of this first evaluation after the initial therapy, MTC patients may either be classified as cured or still affected by either biochemical disease or structural disease. The timing and the type of the following evaluations depend highly on the results from this first evaluation (Fig. 79.4).

Follow-Up of Cured Medullary Thyroid Cancer Patients

Cured patients have a postsurgical Ct less than 10 pg/mL, a negative neck US, and low-normal CEA values. These patients have a 10% risk of recurrence during follow-up that becomes 3% in those patients who have both postsurgical basal and peak Ct after calcium stimulation of less than 10 pg/mL.[92] Patients with a negative stimulation test should be reevaluated every 12 to 18 months, with the aim of early interception of rare, but possible, recurrence.[50] In patients with undetectable levels of serum Ct, measurement of CEA is not necessary (Fig. 79.3A).

KEY POINTS

- After the initial treatment patients can be cured or persistently affected. In the latter case they can be either biochemically or structurally affected. Serum calcitonin and carcinoembryonic antigen (CEA) doubling times are fundamental for planning the type and timing of the imaging techniques to be performed for following the development of the disease.

Follow-Up of Patients with Biochemical Persistent Disease

Patients with biochemical persistent disease are those who show a postsurgical basal Ct greater than 10 pg/mL in the absence of any structural disease as assessed by neck US, CT scan, and, whenever possible, DOPA PET/CT. These patients should be monitored with periodical measurements of serum Ct and CEA (every 6–12 months), because increasing levels are strongly suggestive of progressive disease.[43,93] In most cases, the challenge is to find the source of production of Ct and CEA. This is particularly true in MTC cases with serum Ct less than

150 pg/mL, which are frequently negative on any type of imaging.[94] An accurate neck US is the first localization technique to be performed, due to the high frequency of local recurrence and cervical node metastases. Total-body CT scan and bone scintigraphy are also suggested in the follow-up of a MTC patient with detectable and increasing values of serum Ct. Other functional imaging, and in particular DOPA PET/CT, may be useful for identifying micrometastases not detectable with other techniques.[95] Because a detectable serum Ct level is compatible with long-term survival, during which Ct and CEA may remain stable with time or slowly increase, the most widely accepted therapeutic strategy in these patients is that of "wait and see," thus postponing any type of therapy until structural disease is detected[96] (Fig. 79.4B).

Follow-Up of Patients with Structural Disease

Approximately 10% of MTC patients already have distant metastases at diagnosis that will not be cured by thyroid surgery. Another 10% of those with initial biochemical disease will sooner or later develop local or distant metastases. These patients must be followed up with 6-month evaluations of serum Ct and CEA and calculation of their doubling time. A rapid increase in serum Ct and/or CEA is a poor prognostic factor both for recurrence and death, especially if the doubling time is less than 0.5 to 1 year.[97] A poor prognosis for survival has been recently demonstrated for serum carbohydrate antigen 19.9, particularly in advanced MTC.[98] In this event, imaging evaluations must be intensified to verify the progression of the disease and to decide if it is time to start active therapy.[99]

OTHER THERAPIES

Patients with metastatic disease can be treated with other therapies, both for symptomatic reasons and/or for the progression of the disease. In this latter case the progression should be calculated according to the Response Evaluation Criteria in Solid Tumors (RECIST)[100]: if progression of the disease is documented according to these criteria, therapeutic intervention should be started. However, if progression is present but with a slow growth rate, the indication is to wait and see. If the progression is of a single lesion or of multiple lesions in the same organ, local therapy can be applied, while if the progression is of several lesions and in different organs, systemic therapy should be initiated.[101]

Local Treatment of Recurrent or Persistent Neck Disease

Surgery. The regional lymph nodes of the neck and mediastinum are the most frequent sites of persistent or recurrent disease. In these cases, a second surgical treatment with curative intent is desirable, and to this purpose an extensive modified neck dissection involving microdissection of all node-bearing compartments is recommended. Although only less than 10% of patients affected by MTC with extrathyroidal invasion can be cured by a second surgical treatment, cervical reexploration is safe, and in selected patients may limit MTC progression.[102]

A second surgery on the neck could be also indicated for palliative, other than curative, intent, for example in patients with compressive symptoms who can benefit from surgical debulking.[103] Similarly, a second surgical treatment might be performed for symptomatic lesions or when their growth may cause significant morbidity, as may happen for lymph nodes of the mediastinum adjacent to the great vessels, tracheoesophageal groove, carotid sheath, and brachial plexus.

External Radiotherapy. External radiation therapy (ERT) after thyroidectomy and node dissection is not generally recommended on a prophylactic basis in MTC patients, and the procedure should be reserved for patients with local aggressive disease that is either recurrent or persistent after the first surgery. Although MTC has very low sensitivity to ERT, there is evidence of a potential benefit from radiotherapy in terms of a lower risk (from 2- to 4-fold) of further progression in patients with residual disease.[104] In patients who have undergone a less aggressive primary resection, ERT should be postponed to a second surgical treatment.

TABLE 79.2 **Possible Types of Local Treatments to Be Considered in Advanced Cases of Medullary Thyroid Cancer Before Starting Systemic Therapy and to be Chosen on the Basis of Both the Localization of the Growing Lesion(s) and Local Expertise**

Example of Growing Lesions	Type of Local Treatment
Isolated growing lesions in the lung, bone, liver	Surgery or thermal ablation: the latter can be performed by radiofrequency ablation, laser ablation, or cryoablation, according to local experience
Vegetating tumor tissue inside trachea or bronchus	Endotracheal/bronchial laser removal
Single, small brain metastases	Radiosurgery or stereotaxic ablation
Multiple growing brain metastases	Whole-brain irradiation for stabilization
Painful bone metastases	Antalgic external radiotherapy
Growing neck and/or mediastinal lymphnode metastases not more treatable with surgery	Palliative external radiotherapy
Multiple liver metastases	Intraarterial chemotherapy or intraarterial radioembolization

Radiofrequency and Laser Thermal Ablation. While the safety and efficacy of both laser ablation and radiofrequency thermal ablation (RFA) have been widely demonstrated in the treatment of benign thyroid nodules, evidence regarding local tumor control of thyroid malignancies is still limited. However, preliminary results are encouraging, and image-guided thermal ablation techniques can be considered a valid alternative to surgery for the treatment for recurrent thyroid cancers.

Thermal ablation is based on the use of electromagnetic waves or lasers that produce a high degree of heat. A needle-like RFA probe is introduced in the tumoral mass, and the radiofrequency waves passing through the probe increase the temperature within tumor tissue, resulting in destruction of the tumoral tissue.[105,106] Similarly, when the laser light hits the target, a local increase in temperature occurs, causing permanent damage such as coagulative necrosis and tissue carbonization/vaporization. Both techniques are performed under US guidance, which allows real-time monitoring of the procedure. Thermal ablation is of particular utility when the lesion to be treated is the only site of disease or the only one, among others, that is growing, and surgical treatment cannot be performed, as frequently happens in local para- or retrotracheal recurrence of MTC.

Treatment of Distant Metastases. Local treatments, as described earlier, are also indicated for single, growing, distant metastatic lesions or, if multiple, when only one of them represents a clinical problem, such as pain, local compression of other organs, risk of fracture, or risk of blood invasion. The possibility of local treatment should be always considered before starting systemic therapy, which should be reserved for cases with multiple metastatic lesions that involve multiple organs and are actively growing.[107] As shown in Table 79.2, local treatments can vary according with the site and the number of the metastatic lesions, and decision-making requires the involvement of a multidisciplinary team.

TABLE 79.3 More Frequent Adverse Events Reported in Clinical Trials of the Currently Available Tyrosine Kinase Inhibitors Used for the Treatment of Advanced and Progressive Multimetastatic Medullary Thyroid Cancer

Adverse Events (All Grade)#	Vandetanib (%)	Cabozantinib (%)	Selpercatinib (%)	Pralsetinib (%)
Hypertension	32	33	43	40
Diarrhea	56	63	38	34
Skin rash*	45	19	NR^	24
Anorexia	21	46	NR	15
Fatigue	24	41	38	38
Nausea	33	43	35	17
Weight loss	10	48	NR	NR
QT prolongation	14	NR	19	NR
Hand-foot syndrome	NR	50	NR	NR
Increased SGOT	NR	86	57	69
Increased SGPT	NR	86	51	43

[#] For selpercatinib and pralsetinib, the majority of adverse events were grade 1 or 2.

*Rash includes dermatitis, dermatitis acneiform, eczema, palmar-plantar, erythrodysaesthesia syndrome, erythematous rash, macular rash, maculo-papular rash, papular rash, and pustular rash.

NR, Not reported; *SGOT*, serum glutamate oxaloacetate transaminase; *SGPT*, serum glutamate-pyruvate transaminase.

KEY POINTS

- Advanced MTC should be treated when progressing according to RECIST or according to clinical judgment. If the disease progression is related to one single lesion or to several lesions but belonging to the same organ, local therapies should be preferred.

Systemic Targeted Therapies. Once generalized disease progression has been documented according to RECIST, systemic therapy must be started. Traditional chemotherapy for advanced and progressing MTC has shown limited and temporary response rates in several small series published to date.[108] Chemotherapy is no longer used in progressing MTC, and, if ever used, is reserved for those MTC patients who, for any reason, cannot be treated with tyrosine kinase inhibitors (TKIs), which represent the first choice for systemic therapy.[101]

At the present, there are two available oral drugs belonging to the TKI family: vandetanib and cabozantinib. They were approved by both the US Food and Drug Administration (FDA) and the European Medical Agency (EMA) after the results of the phase III clinical studies ZETA and EXAM,[109,110] which both demonstrated a significant increase in progression-free survival (PFS) compared with placebo.

Both vandetanib and cabozantinib are small molecules able to block, with different activities and different patterns, multiple tyrosine kinases.[111] Among these anti–tyrosine kinase activities, both of them are able to block the constitutively activated Ret receptor. The drugs should be initiated with evidence of disease progression according to RECIST or, in very advanced cases, based on clinical judgment. Although the two drugs are both citostatic more than citotoxic, a significant srinkage of the metastatic lesions can be observed (Fig. 79.5). Vandetanib can be used successfully also in symptomatic patients and those with ectopic Cushing syndrome.[112,113]

The choice to start with one drug or another is strictly dependent on the local availability of the drug, as not all countries have approved or reimburse the use of both drugs. However, if both vandetanib and cabozantinib can be prescribed, the choice must be dictated by the patient's clinical features, location of metastatic lesions, and drug characteristics. According to the results of the phase III studies,[109,110] cabozantinib exhibits more rapid action but is associated with more severe adverse events (AEs). Taking into account these factors, cabozantinib should be preferred when rapid shrinkage of the metastatic lesion is needed, and when this need is sufficient to run the risk of AEs. In the phase III EXAM study, patients could be enrolled even if previously treated with other TKIs, and the results showed that cabozantinib works similarly both as first- and second-line treatment in terms of prolongation of PFS. Based on this result, vandetanib should be used as first-line therapy to reserve the possibility of using cabozantinib as second-line therapy when, for any reason, vandetanib needs to be stopped. Vandetanib, but not cabozantinib, has been successfully tested in children affected by advanced and metastatic MTC in MEN II, mainly MEN IIB.[114] The outcomes in these children demonstrated that vandetanib can be safely and satisfactorily used.[115] Moreover, several reports have shown that ectopic ACTH secretion and paraneoplastic Cushing syndrome are completely reverted and cured by vandetanib.[112,113,116] Vandetanib, but not cabozantinib, can result in prolonged QTc (>440 ms in adult males and >460 ms in adult females)[117]; this adverse effect can be life-threatening, thus cabozantinib is the first-choice drug in patient with QTc values that are already prolonged before starting therapy.

Both drugs can induce very similar AEs, but the prevalence varies (Table 79.3), and this must be considered, as well patient co-morbidities, when deciding which drug should be started first. Although both drugs result in a a significant prolongation in PFS, to date there is no evidence for an increase in overall survival. However, exploratory analyses suggest that patients with *RET* M918T–positive MTC may benefit more than those with *RET* M918T–negative from treatment with cabozantinib.[118]

As with all TKIs, both vandetanib and cabozantinib should be continued until there is evidence of clinical benefit. The reason for this is that they are cytostatic and not cytotoxic, meaning they can block cell proliferation and growth but cannot kill tumoral cells. Nevertheless, two studies and real-life experience have demonstrated that the objective response of metastatic lesions can be significant, although complete response has never been observed.[109,110]

Both drugs can induce several AEs, sometimes of grade 3, and their management is fundamental for success of the treatment. They must be early recognized and immediately treated either with other drugs if possible (i.e., hypertension) or by reducing the daily dose of the drug (i.e., fatigue or anorexia). A multidisciplinary team at a referral center should take care of these patients during all treatment follow-up.

The development of drug resistance and the consequent escape phenomenon are the major concerns associated with this treatment. Once the disease shows progression under treatment, clinicians must decide to continue or to stop the drug. Further studies analyzing the possibility of using the two drugs in an alternating modality, or combined with each other or with other drugs, are an important challenge for the future.

A new generation of TKIs that are much more specific for *RET* mutations has been developed in recent years. LOXO-292, also known as selpercatinib, and BLU-667, also known as pralsetinib, have been successfully tested in phase I/II clinical trials and, after publication of the results,[80,81] both were recently approved by the FDA and are under evaluation from the EMA. They are both potent KDR/VEGF2-sparing *RET* inhibitors with preclinical specificity for *RET* alterations, both gene fusions and point mutations, including the V804M mutation, which showed *in vitro* resistance to vandetanib.[119] These results were further confirmed in animal models and clinical studies,[120,121] in which relevant tumor shrinkage was shown. Interestingly, in mouse models, selpercatinib showed antineoplastic activity against brain metastasis too.[120] Both drugs have a very good safety profile, with few AEs that are generally of low grade (Table 79.3). Whenever possible, these drugs should be the first choice, but their prescription is determined by the presence of a *RET* mutation, either somatic or germline; thus, 15% of cases of advanced MTC cannot be treated with any of these drugs.

When no other drugs are available, other TKIs that have shown positive results in phase II clinical trials[122-127] or in real life[128,129] can prescribed "off label." They may be used also as second- or third-line treatments after cabozantinib and/or vandetanib, especially in MTC with no evidence of *RET* mutations.

KEY POINTS

- Nowadays two tyrosine multikinase inhibitors, vandetanib and cabozantinib, are available for the systemic therapy. In the near future a second generation of RET-specific inhibitors, selpercatinib and pralsetinib, will likely reach clinical practice.

Radionuclide Systemic Therapies. Treatment with radioactive elements has been widely explored in the past, including treatment with radioiodine (^{131}I) linked to meta-iodobenzylguanidine[130] and anti-CEA targeted radioimmunotherapy, but with very limited results, and nowadays these approaches are not used or are only used in cases that are refractory to any other type of systemic therapy.

MTC is a neuroendocrine tumor, and 30% to 50% of cases express somatostatin (SMS) receptors, as documented by octreoscan and by 68Ga-DOTATATE PET/CT.[131,132] A few studies have been reported with promising, but not enthusiastic, results.[133] However, a hypothesis to explain this failure could be the too-advanced disease status at the time of the treatment, thus suggesting of that further studies should be performed with less-advanced cases. Several radiolabeled SMS analogs have been explored as potential therapeutic agents. Most studies of radionuclide therapy in MTC have investigated the use of SMS analogs labeled with the radionuclide(s) 90 yttrium (^{90}Y) and/or 177 lutetium (^{177}Lu).[133-136] The use of 111 indium (^{111}In) was also evaluated.[137] A greater than or equal to 50% decrease in Ct values compared with baseline in one third of patients treated with ^{90}Y[135] and ^{177}Lu[133] has been reported, while others demonstrated a relevant reduction in tumor growth rate. Despite these promising results, these treatments are not yet part of clinical practice.

At present, radionuclide therapy is reserved for those cases that, for any reason, do not respond to or cannot be treated with targeted therapy. However, several ongoing clinical trials (NCT00002947, NCT03647657, NCT02088645, and NCT04106843) should provide more reliable data in the near future.

Other Systemic Therapies. Over the years, different types of octreotide, from the native to the long-acting analogs, have been tested as potential therapeutic agents for MTC. In many cases, a significant reduction in serum Ct has been demonstrated,[138] but no evidence of a parallel relevant reduction in structural disease has been shown. In patients with severe diarrhea that is uncontrolled with any other conventional drug, the use of SMS long-acting analogs can be considered.

Although there are no specific studies of MTC series, treatment with either bisphosphonates or the receptor activator of nuclear factor kappa-B ligand (RANKL) inhibitor denosumab has been recognized as valid in the treatment of bone metastases.[76] This therapeutic procedure has been demonstrated to be effective in controlling bone pain and delaying the occurrence of skeletal-related events in patients with bone metastases due to differentiated thyroid cancer.[139] Both pamidronate and zoledronate are administered intravenously on a monthly basis, while denosumab is administered subcutaneously every 4 weeks. The side effects of these potent antiresorptive agents, although rare, include osteonecrosis of the jaw,[140] atypical subtrochanteric fractures,[141] and hypocalcemia, and must be carefully taken into consideration before starting the therapy and during long-term treatment.

CONCLUSIONS

Medullary thyroid cancer is a rare cancer with a relatively poor prognosis, and the only possibility of achieving a definitive cure is an early diagnosis when the tumor is still intrathyroidal. The routine measurement of serum Ct in all patients with thyroid nodules is still debated.

In 25% of cases, MTC is inherited as an autosomal dominant disease, and children can be affected. *RET* genetic screening is recommended in this latter form to identify gene carriers early and adequately plan the timing of surgery.

Initial therapy is complete removal of the thyroid gland and central node neck dissection, but only one third of patients will be definitively cured by this procedure. Patients must be followed up over the years for the possibility of recurrences, especially when serum Ct is still detectable after surgery. Several modalities of local and systemic therapies are available for metastatic progressive lesions. Nowadays, several drugs belonging to the TKI family are available, and others are under development.

MTC patients should be followed in referral centers and by a multidisciplinary team that should be led by an expert endocrinologist, especially when the disease involves other endocrine glands, as occurs in the familial forms.

REFERENCES

1. Nilsson M, Williams D. On the Origin of Cells and Derivation of Thyroid Cancer: C cell story revisited. *Eur Thyroid J*. 2016;5:79–93.
2. Johansson E, Andersson L, Ornros J, et al. Revising the embryonic origin of thyroid C cells in mice and humans. *Development*. 2015;142:3519–3528.

3. Cote GJ, Grubbs EG, Hofmann MC. Thyroid C-cell biology and oncogenic transformation. *Recent Results Cancer Res.* 2015;204:1–39.
4. Lloyd RVO, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon, France: IARC; 2017.
5. Valle LA, Kloos RT. The prevalence of occult medullary thyroid carcinoma at autopsy. *J Clin Endocrinol Metab*. 2011;96:E109–E113.
6. Roman S, Lin R, Sosa JA. Prognosis of medullary thyroid carcinoma: demographic, clinical, and pathologic predictors of survival in 1252 cases. *Cancer*. 2006;107:2134–2142.
7. Negri E, Ron E, Franceschi S, et al. Risk factors for medullary thyroid carcinoma: a pooled analysis. *Cancer Causes Control*. 2002;13:365–372.
8. Mulligan LM, Kwok JB, Healey CS, et al. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature*. 1993;363:458–460.
9. Donis-Keller H, Dou S, Chi D, et al. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum Mol Genet*. 1993;2:851–856.
10. Elisei R, Tacito A, Ramone T, et al. Twenty-five years experience on RET genetic screening on hereditary MTC: an update on the prevalence of germline RET mutations. *Genes (Basel)*. 2019;10:698–710.
11. Elisei R, Pinchera A. Advances in the follow-up of differentiated or medullary thyroid cancer. *Nat Rev Endocrinol*. 2012;8:466–475.
12. Gharib H, McConahey WM, Tiegs RD, et al. Medullary thyroid carcinoma: clinicopathologic features and long-term follow-up of 65 patients treated during 1946 through 1970. *Mayo Clin Proc.* 1992;67:934–940.
13. Matrone A, Gambale C, Prete A, et al. Impact of advanced age on the clinical presentation and outcome of sporadic medullary thyroid carcinoma. *Cancers (Basel)*. 2020;13:94–97.
14. Kebebew E, Ituarte PH, Siperstein AE, et al. Medullary thyroid carcinoma: clinical characteristics, treatment, prognostic factors, and a comparison of staging systems. *Cancer*. 2000;88:1139–1148.
15. Sadeghian A, Rouhana H, Oswald-Stumpf B, et al. Etiologies and management of cutaneous flushing: Malignant causes. *J Am Acad Dermatol*. 2017;77:405–414.
16. Barbosa SL, Rodien P, Leboulleux S, et al. Ectopic adrenocorticotropic hormone-syndrome in medullary carcinoma of the thyroid: a retrospective analysis and review of the literature. *Thyroid*. 2005;15:618–623.
17. Thomas CM, Asa SL, Ezzat S, et al. Diagnosis and pathologic characteristics of medullary thyroid carcinoma-review of current guidelines. *Curr Oncol*. 2019;26:338–344.
18. Valderrabano P, Klippenstein DL, Tourtelot JB, et al. New American Thyroid Association sonographic patterns for thyroid nodules perform well in medullary thyroid carcinoma: institutional experience, systematic review, and meta-analysis. *Thyroid*. 2016;26:1093–1100.
19. Zhao J, Yang F, Wei X, et al. Ultrasound features value in the diagnosis and prognosis of medullary thyroid carcinoma. *Endocrine*. 2021;72(3):727–734.
20. Yakupova EI, Bobyleva LG, Vikhlyantsev IM, et al. Congo Red and amyloids: history and relationship. *Biosci Rep*. 2019:39–61.
21. Takeichi N, Ito H, Okamoto H, et al. The significance of immunochemically staining calcitonin and CEA in fine-needle aspiration biopsy materials from medullary carcinoma of the thyroid. *Jpn J Surg*. 1989;19:674–678.
22. Papaparaskeva K, Nagel H, Droese M. Cytologic diagnosis of medullary carcinoma of the thyroid gland. *Diagn Cytopathol*. 2000;22:351–358.
23. Pacini F, Fontanelli M, Fugazzola L, et al. Routine measurement of serum calcitonin in nodular thyroid diseases allows the preoperative diagnosis of unsuspected sporadic medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 1994;78:826–829.
24. Vierhapper H, Niederle B, Bieglmayer C, et al. Early diagnosis and curative therapy of medullary thyroid carcinoma by routine measurement of serum calcitonin in patients with thyroid disorders. *Thyroid*. 2005;15:1267–1272.
25. Essig Jr GF, Porter K, Schneider D, et al. Fine needle aspiration and medullary thyroid carcinoma: the risk of inadequate preoperative evaluation and initial surgery when relying upon FNAB cytology alone. *Endocr Pract*. 2013;19:920–927.
26. Vierhapper H, Raber W, Bieglmayer C, et al. Routine measurement of plasma calcitonin in nodular thyroid diseases. *J Clin Endocrinol Metab*. 1997;82:1589–1593.
27. Perros P, Colley S, Boelaert K, et al. Guidelines for the management of thyroid cancer, third edition. British Thyroid Association. Clin. Endocrin. 2014;81(1)41–48.
28. Haddad RI, Nasr C, Bischoff L, et al. NCCN Guidelines® Insights: Thyroid Carcinoma, Version 2.2018. Featured Updates to the NCCN Guidelines. *JNCCN*. 2018;16(12).
29. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016;26:1–133.
30. Pacini F, Schlumberger M, Dralle H, et al. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur J Endocrinol*. 2006;154:787–803.
31. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *J Endocrinol Invest*. 2010;33:51–56.
32. Costante G, Meringolo D, Durante C, et al. Predictive value of serum calcitonin levels for preoperative diagnosis of medullary thyroid carcinoma in a cohort of 5817 consecutive patients with thyroid nodules. *J Clin Endocrinol Metab*. 2007;92:450–455.
33. Elisei R. Routine serum calcitonin measurement in the evaluation of thyroid nodules. *Best Pract Res Clin Endocrinol Metab*. 2008;22:941–953.
34. Mian C, Perrino M, Colombo C, et al. Refining calcium test for the diagnosis of medullary thyroid cancer: cutoffs, procedures, and safety. *J Clin Endocrinol Metab*. 2014;99:1656–1664.
35. Fugazzola L, Di Stefano M, Censi S, et al. Basal and stimulated calcitonin for the diagnosis of medullary thyroid cancer: updated thresholds and safety assessment. *J Endocrinol Invest*. 2021;44:587–597.
36. Demers LM, Spencer CA. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. *Clin Endocrinol (Oxf)*. 2003;58:138–140.
37. Karanikas G, Moameni A, Poetzi C, et al. Frequency and relevance of elevated calcitonin levels in patients with neoplastic and nonneoplastic thyroid disease and in healthy subjects. *J Clin Endocrinol Metab*. 2004;89:515–519.
38. Boi F, Maurelli I, Pinna G, et al. Calcitonin measurement in wash-out fluid from fine needle aspiration of neck masses in patients with primary and metastatic medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2007;92:2115–2118.
39. Verbeek HH, de Groot JWB, Sluiter WJ, et al. Calcitonin testing for detection of medullary thyroid cancer in people with thyroid nodules. *Cochrane Database Syst Rev*. 2020;3:CD010159.
40. Elisei R, Bottici V, Luchetti F, et al. Impact of routine measurement of serum calcitonin on the diagnosis and outcome of medullary thyroid cancer: experience in 10,864 patients with nodular thyroid disorders. *J Clin Endocrinol Metab*. 2004;89:163–168.
41. Pacini F, Basolo F, Elisei R, et al. Medullary thyroid cancer. An immunohistochemical and humoral study using six separate antigens. *Am J Clin Pathol*. 1991;95:300–308.
42. Wyon Y, Frisk J, Lundeberg T, et al. Postmenopausal women with vasomotor symptoms have increased urinary excretion of calcitonin gene-related peptide. *Maturitas*. 1998;30:289–294.
43. Busnardo B, Girelli ME, Simioni N, et al. Nonparallel patterns of calcitonin and carcinoembryonic antigen levels in the follow-up of medullary thyroid carcinoma. *Cancer*. 1984;53:278–285.
44. Gambardella C, Offi C, Clarizia G, et al. Medullary thyroid carcinoma with double negative calcitonin and CEA: a case report and update of literature review. *BMC Endocr Disord*. 2019;19:103.
45. Eng C, Mulligan LM, Smith DP, et al. Low frequency of germline mutations in the RET proto-oncogene in patients with apparently sporadic medullary thyroid carcinoma. *Clin Endocrinol (Oxf)*. 1995;43:123–127.

46. Wiench M, Wygoda Z, Gubala E, et al. Estimation of risk of inherited medullary thyroid carcinoma in apparent sporadic patients. *J Clin Oncol*. 2001;19:1374–1380.
47. Keiser HR, Beaven MA, Doppman J, et al. Sipple's syndrome: medullary thyroid carcinoma, pheochromocytoma, and parathyroid disease. Studies in a large family. NIH conference. *Ann Intern Med*. 1973;78:561–579.
48. Castinetti F, Waguespack SG, Machens A, et al. Natural history, treatment, and long-term follow up of patients with multiple endocrine neoplasia type 2B: an international, multicentre, retrospective study. *Lancet Diabetes Endocrinol*. 2019;7:213–220.
49. Farndon JR, Leight GS, Dilley WG, et al. Familial medullary thyroid carcinoma without associated endocrinopathies: a distinct clinical entity. *Br J Surg*. 1986;73:278–281.
50. Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
51. Carney JA, Roth SI, Heath 3rd H, et al. The parathyroid glands in multiple endocrine neoplasia type 2b. *Am J Pathol*. 1980;99:387–398.
52. Cunliffe WJ, Hudgson P, Fulthorpe JJ, et al. A calcitonin-secreting medullary thyroid carcinoma associated with mucosal neuromas, marfanoid features, myopathy and pigmentation. *Am J Med*. 1970;48:120–126.
53. Giani C, Ramone T, Romei C, et al. A new MEN2 syndrome with clinical features of both MEN2A and MEN2B associated with a new RET germline deletion. *Case Rep Endocrinol*. 2020;2020:4147097.
54. Elisei R, Matrone A, Valerio L, et al. Fifty years after the first description, MEN 2B syndrome diagnosis is still late: descriptions of two recent cases. *J Clin Endocrinol Metab*. 2019;104:2520–2526.
55. Ceccherini I, Romei C, Barone V, et al. Identification of the Cys634--> Tyr mutation of the RET proto-oncogene in a pedigree with multiple endocrine neoplasia type 2A and localized cutaneous lichen amyloidosis. *J Endocrinol Invest*. 1994;17:201–204.
56. Coyle D, Friedmacher F, Puri P. The association between Hirschsprung's disease and multiple endocrine neoplasia type 2a: a systematic review. *Pediatr Surg Int*. 2014;30:751–756.
57. Nagy N, Guyer RA, Hotta R, et al. RET overactivation leads to concurrent Hirschsprung disease and intestinal ganglioneuromas. *Development*. 2020;147(21):1–10.
58. Wolfe HJ, Delellis RA. Familial medullary thyroid carcinoma and C cell hyperplasia. *Clin Endocrinol Metab*. 1981;10:351–365.
59. Jadoul M, Leo JR, Berends MJ, et al. Pheochromocytoma-induced hypertensive encephalopathy revealing MEN-IIa syndrome in a 13-year old boy. Implications for screening procedures and surgery. *Horm Metab Res Suppl*. 1989;21:46–49.
60. Lenders JW, Pacak K, Walther MM, et al. Biochemical diagnosis of pheochromocytoma: which test is best? *JAMA*. 2002;287:1427–1434.
61. Eisenhofer G, Huynh TT, Hiroi M, et al. Understanding catecholamine metabolism as a guide to the biochemical diagnosis of pheochromocytoma. *Rev Endocr Metab Disord*. 2001;2:297–311.
62. Taieb D, Hicks RJ, Hindie E, et al. European Association of Nuclear Medicine Practice Guideline/Society of Nuclear Medicine and Molecular Imaging Procedure Standard 2019 for radionuclide imaging of phaeochromocytoma and paraganglioma. *Eur J Nucl Med Mol Imaging*. 2019;46:2112–2137.
63. Silverberg SJ, Shane E, Jacobs TP, et al. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. *N Engl J Med*. 1999;341:1249–1255.
64. Lips P, Wiersinga A, van Ginkel FC, et al. The effect of vitamin D supplementation on vitamin D status and parathyroid function in elderly subjects. *J Clin Endocrinol Metab*. 1988;67:644–650.
65. Elisei R, Alevizaki M, Conte-Devolx B, et al. 2012 European thyroid association guidelines for genetic testing and its clinical consequences in medullary thyroid cancer. *Eur Thyroid J*. 2013;1:216–231.
66. Elisei R, Romei C, Renzini G, et al. The timing of total thyroidectomy in RET gene mutation carriers could be personalized and safely planned on the basis of serum calcitonin: 18 years experience at one single center. *J Clin Endocrinol Metab*. 2012;97:426–435.
67. Brandi ML, Gagel RF, Angeli A, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab*. 2001;86:5658–5671.
68. Rohmer V, Vidal-Trecan G, Bourdelot A, et al. Prognostic factors of disease-free survival after thyroidectomy in 170 young patients with a RET germline mutation: a multicenter study of the Groupe Francais d'Etude des Tumeurs Endocrines. *J Clin Endocrinol Metab*. 2011;96:E509–E518.
69. Lloyd RV, Sisson JC, Marangos PJ. Calcitonin, carcinoembryonic antigen and neuron-specific enolase in medullary thyroid carcinoma. *Cancer*. 1983;51:2234–2239.
70. Yadav M, Agrawal V, Pani KC, et al. C-cell hyperplasia in sporadic and familial medullary thyroid carcinoma. *Indian J Pathol Microbiol*. 2018;61:485–488.
71. Eng C, Clayton D, Schuffenecker I, et al. The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2. International RET mutation consortium analysis. *Jama*. 1996;276:1575–1579.
72. Acton DS, Velthuyzen D, Lips CJ, et al. Multiple endocrine neoplasia type 2B mutation in human RET oncogene induces medullary thyroid carcinoma in transgenic mice. *Oncogene*. 2000;19:3121–3125.
73. Michiels FM, Chappuis S, Caillou B, et al. Development of medullary thyroid carcinoma in transgenic mice expressing the RET protooncogene altered by a multiple endocrine neoplasia type 2A mutation. *Proc Natl Acad Sci U S A*. 1997;94:3330–3335.
74. Niccoli-Sire P, Murat A, Rohmer V, et al. Familial medullary thyroid carcinoma with noncysteine ret mutations: phenotype-genotype relationship in a large series of patients. *J Clin Endocrinol Metab*. 2001;86:3746–3753.
75. Cosci B, Vivaldi A, Romei C, et al. In silico and in vitro analysis of rare germline allelic variants of RET oncogene associated with medullary thyroid cancer. *Endocr Relat Cancer*. 2011;18(5):603–612.
76. Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
77. Romei C, Ugolini C, Cosci B, et al. Low prevalence of the somatic M918T RET mutation in micro-medullary thyroid cancer. *Thyroid*. 2012;22:476–481.
78. Romei C, Casella F, Tacito A, et al. New insights in the molecular signature of advanced medullary thyroid cancer: evidence of a bad outcome of cases with double RET mutations. *J Med Genet*. 2016;53(11):729–734.
79. Elisei R, Cosci B, Romei C, et al. Prognostic significance of somatic RET oncogene mutations in sporadic medullary thyroid cancer: a 10-year follow-up study. *J Clin Endocrinol Metab*. 2008;93:682–687.
80. Wirth LJ, Sherman E, Robinson B, et al. Efficacy of selpercatinib in RET-altered thyroid cancers. *N Engl J Med*. 2020;383:825–835.
81. Hu M, Subbiah M, Wirth LJ, et al. Results from the registrational phase I/II ARROW trial of pralsetinib (BLU-667) in patients (pts) with advanced RET mutation-positive medullary thyroid cancer (RET+ MTC). *Ann Oncol*. 2020;31:S1026–S1033.
82. Ciampi R, Mian C, Fugazzola L, et al. Evidence of a low prevalence of RAS mutations in a large medullary thyroid cancer series. *Thyroid*. 2013;23:50–57.
83. Agrawal N, Jiao Y, Sausen M, et al. Exomic sequencing of medullary thyroid cancer reveals dominant and mutually exclusive oncogenic mutations in RET and RAS. *J Clin Endocrinol Metab*. 2013;98:E364–E369.
84. Ciampi R, Romei C, Ramone T, et al. Genetic landscape of somatic mutations in a large cohort of sporadic medullary thyroid carcinomas studied by next-generation targeted sequencing. *iScience*. 2019;20:324–336.
85. Heilmann AM, Subbiah V, Wang K, et al. Comprehensive genomic profiling of clinically advanced medullary thyroid carcinoma. *Oncology*. 2016;90:339–346.
86. Smith J, Read ML, Hoffman J, et al. Germline ESR2 mutation predisposes to medullary thyroid carcinoma and causes up-regulation of RET expression. *Hum Mol Genet*. 2016;25:1836–1845.
87. Essig Jr GF, Porter K, Schneider D, et al. Multifocality in sporadic medullary thyroid carcinoma: an international multicenter study. *Thyroid*. 2016;26:1563–1572.

88. Machens A, Dralle H. Biomarker-based risk stratification for previously untreated medullary thyroid cancer. *J Clin Endocrinol Metab*. 2010;95:2655–2663.
89. Modigliani E, Cohen R, Campos JM, et al. Prognostic factors for survival and for biochemical cure in medullary thyroid carcinoma: results in 899 patients. The GETC Study Group. Groupe d'etude des tumeurs a calcitonine. *Clin Endocrinol (Oxf)*. 1998;48:265–273.
90. Viola D, Elisei R. Management of medullary thyroid cancer. *Endocrinol Metab Clin North Am*. 2019;48:285–301.
91. Fugazzola L, Pinchera A, Luchetti F, et al. Disappearance rate of serum calcitonin after total thyroidectomy for medullary thyroid carcinoma. *Int J Biol Markers*. 1994;9:21–24.
92. Franc S, Niccoli-Sire P, Cohen R, et al. Complete surgical lymph node resection does not prevent authentic recurrences of medullary thyroid carcinoma. *Clin Endocrinol (Oxf)*. 2001;55:403–409.
93. Rougier P, Calmettes C, Laplanche A, et al. The values of calcitonin and carcinoembryonic antigen in the treatment and management of nonfamilial medullary thyroid carcinoma. *Cancer*. 1983;51:855–862.
94. Raue F, Frank-Raue K. Long-term follow-up in medullary thyroid carcinoma. *Recent Results Cancer Res*. 2015;204:207–225.
95. Terroir M, Caramella C, Borget I, et al. F-18-Dopa positron emission tomography/computed tomography is more sensitive than whole-body magnetic resonance imaging for the localization of persistent/recurrent disease of medullary thyroid cancer patients. *Thyroid*. 2019;29:1457–1464.
96. Rajan N, Khanal T, Ringel MD. Progression and dormancy in metastatic thyroid cancer: concepts and clinical implications. *Endocrine*. 2020;70:24–35.
97. Meijer JA, le Cessie S, van den Hout WB, et al. Calcitonin and carcinoembryonic antigen doubling times as prognostic factors in medullary thyroid carcinoma: a structured meta-analysis. *Clin Endocrinol (Oxf)*. 2010;72:534–542.
98. Elisei R, Lorusso L, Piaggi P, et al. Elevated level of serum carbohydrate antigen 19.9 as predictor of mortality in patients with advanced medullary thyroid cancer. *Eur J Endocrinol*. 2015;173:297–304.
99. Schlumberger M, Bastholt L, Dralle H, et al. 2012 European thyroid association guidelines for metastatic medullary thyroid cancer. *Eur Thyroid J*. 2012;1:5–14.
100. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–247.
101. Hadoux J, Pacini F, Tuttle RM, et al. Management of advanced medullary thyroid cancer. *Lancet Diabetes Endocrinol*. 2016;4:64–71.
102. Kebebew E, Kikuchi S, Duh QY, et al. Long-term results of reoperation and localizing studies in patients with persistent or recurrent medullary thyroid cancer. *Arch Surg*. 2000;135:895–901.
103. Chen H, Roberts JR, Ball DW, et al. Effective long-term palliation of symptomatic, incurable metastatic medullary thyroid cancer by operative resection. *Annals of Surgery*. 1998;227:887–895.
104. Schwartz DL, Rana V, Shaw S, et al. Postoperative radiotherapy for advanced medullary thyroid cancer–local disease control in the modern era. *Head Neck*. 2008;30:883–888.
105. Mazzeo S, Cervelli R, Elisei R, et al. mRECIST criteria to assess recurrent thyroid carcinoma treatment response after radiofrequency ablation: a prospective study. *J Endocrinol Invest*. 2018;41:1389–1399.
106. Mauri G, Gennaro N, Lee MK, et al. Laser and radiofrequency ablations for benign and malignant thyroid tumors. *Int J Hyperthermia*. 2019;36:13–20.
107. Schlumberger MBL, Dralle H, Jarzab B, et al. 2012 European Thyroid Association guidelines for metastatic medullary thyroid cancer. *Eur Thyroid J*. 2012;1:5–14.
108. Orlandi F, Caraci P, Mussa A, et al. Treatment of medullary thyroid carcinoma: an update. *Endocr Relat Cancer*. 2001;8:135–147.
109. Wells Jr SA, Robinson BG, Gagel RF, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol*. 2012;30:134–141.
110. Elisei R, Schlumberger MJ, Muller SP, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol*. 2013;31:3639–3646.
111. Matrone A, Valerio L, Pieruzzi L, et al. Protein kinase inhibitors for the treatment of advanced and progressive radiorefractory thyroid tumors: From the clinical trials to the real life. *Best Pract Res Clin Endocrinol Metab*. 2017;31:319–334.
112. Paepegaey AC, Cochand-Priollet B, Louiset E, et al. Long-term control of hypercortisolism by vandetanib in a case of medullary thyroid carcinoma with a somatic RET mutation. *Thyroid*. 2017;27:587–590.
113. Pitoia F, Bueno F, Schmidt A, et al. Rapid response of hypercortisolism to vandetanib treatment in a patient with advanced medullary thyroid cancer and ectopic Cushing syndrome. *Arch Endocrinol Metab*. 2015;59:343–346.
114. Fox E, Widemann BC, Chuk MK, et al. Vandetanib in children and adolescents with multiple endocrine neoplasia type 2B associated medullary thyroid carcinoma. *Clin Cancer Res*. 2013;19:4239–4248.
115. Kraft IL, Akshintala S, Zhu YJ, et al. Outcomes of children and adolescents with advanced hereditary medullary thyroid carcinoma treated with vandetanib. *Clin Cancer Res*. 2018;24(4):753–765.
116. Nella AA, Lodish MB, Fox E, et al. Vandetanib successfully controls medullary thyroid cancer-related Cushing syndrome in an adolescent patient. *J Clin Endocrinol Metab*. 2014;99:3055–3059.
117. American College of Cardiology/American Heart Association Task Force on Clinical Data S, Buxton AE, Calkins H, et al. ACC/AHA/HRS 2006 key data elements and definitions for electrophysiological studies and procedures: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (ACC/AHA/HRS Writing Committee to Develop Data Standards on Electrophysiology). *Circulation*. 2006;114(23):2534–2570.
118. Schlumberger M, Elisei R, Muller S, et al. Overall survival analysis of EXAM, a phase III trial of cabozantinib in patients with radiographically progressive medullary thyroid carcinoma. *Ann Oncol*. 2017;28:2813–2819.
119. Carlomagno F, Guida T, Anaganti S, et al. Disease associated mutations at valine 804 in the RET receptor tyrosine kinase confer resistance to selective kinase inhibitors. *Oncogene*. 2004;23:6056–6063.
120. Subbiah V, Velcheti V, Tuch BB, et al. Selective RET kinase inhibition for patients with RET-altered cancers. *Ann Oncol*. 2018;29:1869–1876.
121. Subbiah V, Gainor JF, Rahal R, et al. Precision targeted therapy with BLU-667 for RET-driven cancers. *Cancer Discov*. 2018;8:836–849.
122. Schlumberger M, Jarzab B, Cabanillas ME, et al. A phase II trial of the multitargeted tyrosine kinase inhibitor lenvatinib (E7080) in advanced medullary thyroid cancer. *Clin Cancer Res*. 2016;22:44–53.
123. Ravaud A, de la Fouchardiere C, Caron P, et al. A multicenter phase II study of sunitinib in patients with locally advanced or metastatic differentiated, anaplastic or medullary thyroid carcinomas: mature data from the THYSU study. *Eur J Cancer*. 2017;76:110–117.
124. Lam ET, Ringel MD, Kloos RT, et al. Phase II clinical trial of sorafenib in metastatic medullary thyroid cancer. *J Clin Oncol*. 2010;28:2323–2330.
125. Locati LD, Licitra L, Agate L, et al. Treatment of advanced thyroid cancer with axitinib: Phase 2 study with pharmacokinetic/pharmacodynamic and quality-of-life assessments. *Cancer*. 2014;120:2694–2703.
126. Bible KC, Suman VJ, Molina JR, et al. A multicenter phase 2 trial of pazopanib in metastatic and progressive medullary thyroid carcinoma: MC057H. *J Clin Endocrinol Metab*. 2014;99:1687–1693.
127. Schlumberger MJ, Elisei R, Bastholt L, et al. Phase II study of safety and efficacy of motesanib in patients with progressive or symptomatic, advanced or metastatic medullary thyroid cancer. *J Clin Oncol*. 2009;27:3794–3801.
128. Matrone A, Prete A, Nervo A, et al. Lenvatinib as a salvage therapy for advanced metastatic medullary thyroid cancer. *J Endocrinol Invest*. 2021;44(10):2139–2151.
129. Masaki C, Sugino K, Saito N, et al. Lenvatinib induces early tumor shrinkage in patients with advanced thyroid carcinoma. *Endocr J*. 2017;64:819–826.
130. Maiza JC, Grunenwald S, Otal P, et al. Use of 131 I-MIBG therapy in MIBG-positive metastatic medullary thyroid carcinoma. *Thyroid*. 2012;22:654–655.
131. Baudin E, Lumbroso J, Schlumberger M, et al. Comparison of octreotide scintigraphy and conventional imaging in medullary thyroid carcinoma. *J Nucl Med*. 1996;37:912–916.

132. Tuncel M, Kilickap S, Suslu N. Clinical impact of (68)Ga-DOTATATE PET-CT imaging in patients with medullary thyroid cancer. *Ann Nucl Med*. 2020;34:663–674.
133. Bodei L, Handkiewicz-Junak D, Grana C, et al. Receptor radionuclide therapy with 90Y-DOTATOC in patients with medullary thyroid carcinomas. *Cancer Biother Radiopharm*. 2004;19:65–71.
134. Satapathy S, Mittal BR, Sood A, et al. Efficacy and safety of concomitant 177Lu-DOTATATE and low-dose capecitabine in advanced medullary thyroid carcinoma: a single-centre experience. *Nucl Med Commun*. 2020;41:629–635.
135. Parghane RV, Naik C, Talole S, et al. Clinical utility of (177) Lu-DOTATATE PRRT in somatostatin receptor-positive metastatic medullary carcinoma of thyroid patients with assessment of efficacy, survival analysis, prognostic variables, and toxicity. *Head Neck*. 2020;42:401–416.
136. Paganelli G, Zoboli S, Cremonesi M, et al. Receptor-mediated radiotherapy with 90Y-DOTA-D-Phe1-Tyr3-octreotide. *Eur J Nucl Med*. 2001;28:426–434.
137. Valkema R, De Jong M, Bakker WH, et al. Phase I study of peptide receptor radionuclide therapy with [In-DTPA]octreotide: the Rotterdam experience. *Semin Nucl Med*. 2002;32:110–122.
138. Lupoli GA, Fonderico F, Fittipaldi MR, et al. The role of somatostatin analogs in the management of medullary thyroid carcinoma. *J Endocrinol Invest*. 2003;26:72–74.
139. Orita Y, Sugitani I, Takao S, et al. Prospective evaluation of zoledronic acid in the treatment of bone metastases from differentiated thyroid carcinoma. *Ann Surg Oncol*. 2015;22:4008–4013.
140. Khosla S, Burr D, Cauley J, et al. Bisphosphonate-associated osteonecrosis of the jaw: report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res.* 2007;22:1479–1491.
141. Abrahamsen B, Eiken P, Eastell R. Subtrochanteric and diaphyseal femur fractures in patients treated with alendronate: a register-based national cohort study. *J Bone Miner Res.* 2009;24:1095–1102.

80

Anaplastic (Undifferentiated) Thyroid Carcinoma

Robert C. Smallridge and Sarika N. Rao

OUTLINE

Presentation, 1329
  Epidemiology, 1329
  Clinical Characteristics, 1329
Diagnosis, 1329
  Biopsy, 1329
  Histopathology Subtypes, 1329
  Differential Diagnoses, 1330
  Molecular Genetics, 1330
Evaluation, 1331
  Clinical and Labs, 1331
  Imaging, 1331
  Staging, 1331
Establish Goals of Care, 1332
  Multidisciplinary Team, 1332
Prognosis, 1332
  Predictors of Survival, 1332
  Median and Overall Survival, 1333
Therapy, 1333
  Treatment Strategies by Stage, 1333
  Surgery, 1333
  Radiation Therapy, 1333
  Multimodal Therapy, 1333
Future Directions, 1335
Conclusions, 1335

PRESENTATION

Epidemiology

Undifferentiated (i.e., anaplastic) thyroid carcinoma (ATC) comprises a small percentage of all thyroid malignancies worldwide. In a review of frequency in 15 countries, the median percentage of ATC out of all thyroid cancer cases was 3.6% (range: 0.6%–9.8%), with a decrease in several countries after dietary iodide was increased.[1,2] In the United States, the estimated number of all new thyroid cancer cases in 2021 will be 44,280 (32,130 female and 12,150 male) and estimated deaths to be 2,200 (1,150 female; 1,050 male). ATC occurs in only 1% to 2% of thyroid cancers in the United States,[3] but accounts for almost half of thyroid cancer-related deaths, reflecting the extremely aggressive behavior of ATC. The incidence varies slightly among racial/ethnic groups in the United States, occurring in 0.6% of all thyroid cancers in Hispanic Whites, 1.0% in Asian-Pacific Islanders, 1.1% in non-Hispanic Whites, 1.3% in Native Americans, and 1.8% in Black patients.[1]

Clinical Characteristics

In a review of 44 series in 2016 comprising 2881 patients, 1898 (66%) were female, and 983 (34%) were male.[1] Although patients ranged in age from 15 to 98 years, ATC is primarily a disease of older persons, with mean age of 66.5 years (n =19 studies) and median age of 69 years (n = 26 studies). The median size of the primary tumor was 6.8 cm (range: 0.5–25 cm). ATC could arise *de novo* but was more likely to develop in patients who had a longstanding goiter or, in 30% of cases (range: 3%–58%), coexist with differentiated thyroid cancer. Distant metastases were present at initial diagnosis in 42% of patients (range: 11%–90%). Five of the 44 reports identified the causes of death: local disease in 15% (range: 5%–37%) of patients, distant metastases in 59% (range: 12%–68%) or both local and distant disease in 26% (range: 25%–51%). Since that review, ten additional studies (comprising 1953 patients) have reported similar results, mirroring the characteristics observed in the earlier reports.[4-13]

Clinically, the most frequent initial symptom is a rapidly enlarging neck mass, observed in 86% of studies (range: 46%–100%).[1] The mass is often invasive, leading to destruction to nearby structures, including the trachea, esophagus, cervical vessels, and recurrent laryngeal nerves, resulting in hoarseness or dysphonia (33% of cases), dysphagia (38%), dyspnea (27%), pain (16%), and, less frequently, cough, hemoptysis, or superior vena cava syndrome (10% or less). Distant metastases at presentation are detected most often in the lungs (37%), followed by mediastinum, liver, bone, brain, kidneys, heart, and adrenal glands. At autopsy, distant metastases are noted at two to four times the clinical detection rate.[1]

DIAGNOSIS

Biopsy

KEY POINTS

- Anaplastic thyroid cancer is rare, accounting for approximately 3.6% of all thyroid cancers, and is extremely aggressive.
- Median age of diagnosis is approximately at 66 years with median tumor size of 6.8 cm.
- Patients most commonly present with a rapidly enlarging neck mass, with symptoms related to extent of disease involvement.

Fine-needle aspiration biopsy and cytology can be attempted first but has a higher nondiagnostic rate than in other thyroid malignancies because of the undifferentiated nature of the cells, as well as extensive necrosis and infiltration of tumor macrophages. If unsuccessful, then expeditious core tissue biopsy or possibly even surgical biopsy may be necessary. ATC has no histologic features of thyroid tissue, although it may coexist with differentiated thyroid cancer. Several other tumors should be considered in the differential diagnosis, however.

Histopathology Subtypes

The major subtypes of ATC include sarcomatoid (i.e., spindle cell), pleomorphic giant cell, and epithelial (i.e., squamoid) (Fig. 80.1).

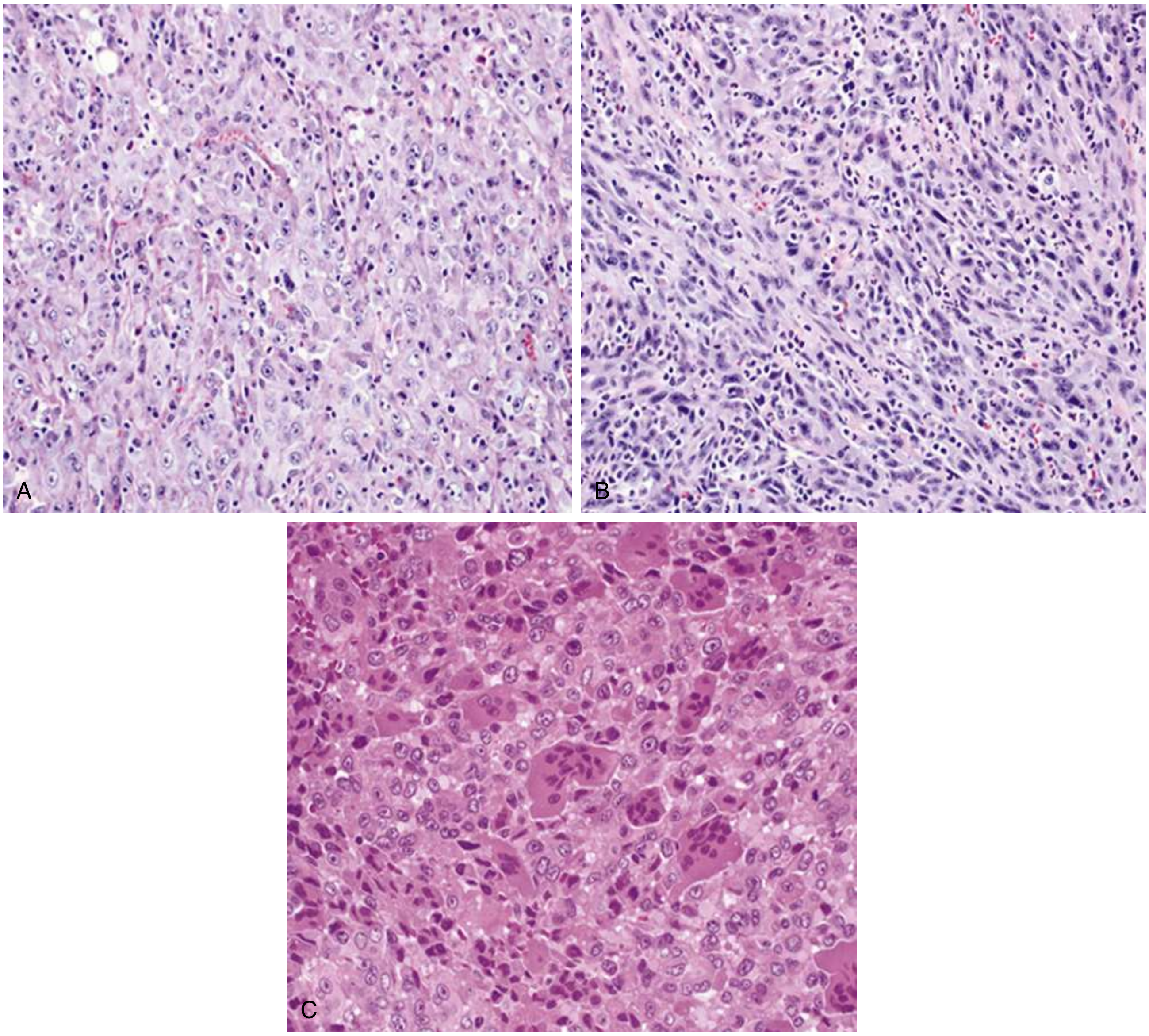


Fig. 80.1 Histologic subtypes of anaplastic thyroid carcinoma. Spindle cell variant (**A**), epithelioid variant (**B**), and giant cell variant (**C**).

These subtypes are often mixed. Other rare variants include paucicellular, rhabdoid, osteoclastic, angiomatoid, lymphoepithelioma-like, and small cell.[9,14-18] The specific histotype may influence survival, as long-term survival (>1 year) is significantly more common in tumors with epithelial or squamous cell component, and short-term survival (<3 months) is more likely with rhabdoid cells.[19]

Differential Diagnoses

Given the undifferentiated features of ATC, a number of other poorly differentiated tumors need to be excluded. Tumors to be considered include primary thyroid lymphoma, sarcoma (either primary or metastatic to the thyroid), squamous cell carcinoma (either primary or metastatic from nearby head/neck tumors), medullary thyroid cancer, intrathyroidal thymic carcinoma, spindle epithelial tumor with thymus-like differentiation, and other metastases (e.g., melanoma, poorly differentiated renal cell carcinoma).[14-16,20]

Immunohistochemistry stains are very helpful in identifying the origin of the thyroid malignancy.[14] Pankeratins (AE1/AE3) are positive in most ATCs; PAX-8 staining is also frequently positive,[21] as is thyroid transcription factor 1,[22] whereas thyroglobulin is never positive[9] unless there is coexisting differentiated (mostly papillary) thyroid cancer.

P53 is positive, and nuclear B-catenin may be detected. Squamous cell carcinomas are immunopositive for pankeratin, E-cadherin, and sometimes high molecular weight keratins. Lymphomas are positive for CD45 and may be positive for PAX-8, P53, and nuclear B-catenin, while medullary thyroid carcinoma is positive for calcitonin, carcinoembryonic antigen, chromogranin, and synaptophysin. Melanoma is positive for Melan-A.

Molecular Genetics

The molecular landscape of ATC is remarkably complex, with chromosomal aneuploidy, multiple gene mutations (in contrast to differentiated thyroid cancer, which has a low mutation rate), upregulated and downregulated expression of hundreds of genes, up- and downregulation of microRNAs (miRNAs), and epigenetic abnormalities.[23] Table 80.1 lists the most common mutations reported in 18 studies. *BRAF-V600E*, *RAS*, and *PIK3CA* are three oncogenic drivers commonly seen in thyroid malignancies, and the addition of *TERT-promoter* and *TP53*

TABLE 80.1 Gene Mutations in Anaplastic Thyroid Carcinoma

Mutated Gene (# Articles)	Median Prevalence (Range) %
TERT-p (13)	61 (32–82)
TP53 (18)	60.5 (25–78)
BRAF-V600E (18)	3 (6–91)
RAS (17)	24 (9–44)
CDKN2A (4)	22.5 (22–32)
PIK3CA (18)	14.5 (0–44)
PTEN (10)	11 (6–19)
NF-1 (5)	9 (9–16)

mutations (as well as other gene mutations) are responsible for the evolution of differentiated thyroid tumors to the anaplastic phenotype. Numerous other mutations have been described in one or more series, including mutations in *EIF1AX*, *NF-2*, *IDH1*, *KMT2D*, *EGF-Receptor*, *NUP93*, and other genes.[24-31] In one study, age influenced which mutations were present, with patients over 70 years old having *BRAF*, *CDKN2A*, *PIK3CA*, and *JAK2* mutations, whereas patients 50 years and younger had *NRAS*, *PTEN*, and *MYC* mutations.[32,33] Yoo et al.[33] reported that, in 27 ATC patients, *CDKN2A* loss was associated with increased disease-specific mortality, that *BRAF-mutant* tumors were associated with increased *VEGF* and *NOTCH* pathway expression and that *RAS* mutations were associated with *MAPK* pathway activity and focal adhesion. Gene fusions have also been detected in ATCs,[34] including *RET/PTC3*, *STRN-ALK*, *NUT-BRD4,* and *KIAA1549-BRAF.*[35]

Differential gene expression has been reported in multiple studies. Many hub genes are upregulated and have oncogenic properties (Table 80.2).[36-39] These genes are involved in multiple areas of cell proliferation, including mitosis, cell cycle progression, mitotic spindle checkpoint, DNA replication and repair, and telomere stability, as well as migration, invasion, epithelial-to-mesenchymal transition, and tumor angiogenesis.[36-39]

Epigenetic influences can alter gene expression by multiple mechanisms, including DNA methylation, histone modification (e.g., deacetylases; methyltransferases), chromatin remodeling, and deregulation of noncoding RNAs. The latter include both miRNAs and long noncoding RNAs (lncRNAs).[40] Multiple miRNAs are dysregulated and can affect cell function. For example, miR-205 suppresses angiogenesis, migration and invasion, and tumor growth.[41] LncRNAs also are overexpressed and participate in gene expression regulation.[42,43]

EVALUATION

Clinical and Labs

Patients should have a complete history and physical examination, as well as baseline laboratory studies. Careful attention should be paid to the head and neck examination, as this tumor is rapidly expansile and invasive. Patients frequently have impaired vocal cord function and may have direct tracheal invasion, so either mirror or fiber optic direct laryngeal examination should be performed.[44,45]

Laboratory studies should include a complete blood count with differential and platelet count, a chemistry profile to include electrolytes, blood urea nitrogen/creatinine, glucose, liver function tests, calcium and albumin to assess nutritional status, thyroid-stimulating hormone, and blood type/cross-match (if surgery is being considered).[14] Limited molecular testing should also be performed because the presence of the *BRAF-V600E* mutation could direct the choice of systemic therapy. Although rare, *ALK*, mTOR pathway, and *TRK* mutations and *RET* fusions also are targetable and could be tested.

TABLE 80.2 Hub Genes Overexpressed in Anaplastic Thyroid Carcinoma

TOP2A	*AURKA*
CDK1	*AURKB*
CCNB1	*BUB2*
VEGFA	*TYMS*
BIRC5	*RACGAP1*
MAPK1	*FEN1*
CCNA2	*PRC1*
MAD2L1	*UBE2C*
CDC20	*MMP3*
BUB1	*MMP13*

(From Hu S, Liao Y, Chen L. Identification of key pathways and genes in anaplastic thyroid carcinoma via integrated bioinformatics analysis. *Med Sci Monit.* 2018;24:6438–6448; Ma Y, Cang S, Li G, et al. Integrated analysis of transcriptome data revealed MMP3 and MMP13 as critical genes in anaplastic thyroid cancer progression. *J Cell Physiol.* 2019;234:22260–22271; Pan Z, Li L, Fang Q, et al. Integrated bioinformatics analysis of master regulators in anaplastic thyroid carcinoma. *Biomed Res Int.* 2019;2019:9734576; and Gao X, Wang J, Zhang S. Integrated bioinformatics analysis of hub genes and pathways in anaplastic thyroid carcinomas. *Int J Endocrinol.* 2019;2019:9651380.)

Imaging

Baseline imaging should be extensive and done quickly, to accurately stage the patient and allow for discussion of appropriate treatment options. A neck ultrasound can be done at time of presentation and provides initial information on extent of tumor within the neck, as well as the presence or absence of cervical lymphadenopathy. A contrast-enhanced computed tomography (CT) scan of the neck is also essential, as the presence of tracheal or esophageal invasion, extension into the prevertebral space, or carotid artery encasement are generally considered contraindications for extensive surgical resection. Patients should have either a positron emission tomography/CT scan from head to midthigh (preferable) or a CT of the chest/abdomen/pelvis with a bone scan, to search for systemic metastases, as well as magnetic resonance imaging (MRI) of the brain; MRI of the spine may also be indicated.[14,46-48] Thorough imaging may detect unexpected lesions and alter management in up to 50% of cases.[46]

Staging

KEY POINTS

- All anaplastic thyroid cancer are stage 4 due to its aggressive nature and risk of early metastases. However, staging is sub-categorized as IVA, B or C based on extent of disease.
- Rapid baseline clinical, lab and imaging assessment is needed, including screening for mutations involving BRAF-V600E.

Unlike differentiated thyroid malignancies, which have four stages, all ATCs are stage IV, although they are further divided into stages IVA, IVB, and IVC. In the tumor-node-metastasis (TNM) staging system, all ATC tumors were previously classified as T4 (T4a if intrathyroidal and T4b if there was gross extrathyroidal extension). In the recent American Joint Committee on Cancer 8th edition revision, the T classification was changed to be consistent with differentiated thyroid cancer, such that now ATC is classified as T1 to T4 based on the extent of the primary tumor.[49] Stage IVA is defined as a primary tumor of any size but confined to the thyroid with no lymph node involvement or distant metastases; stage IVB disease has gross extrathyroidal extension,

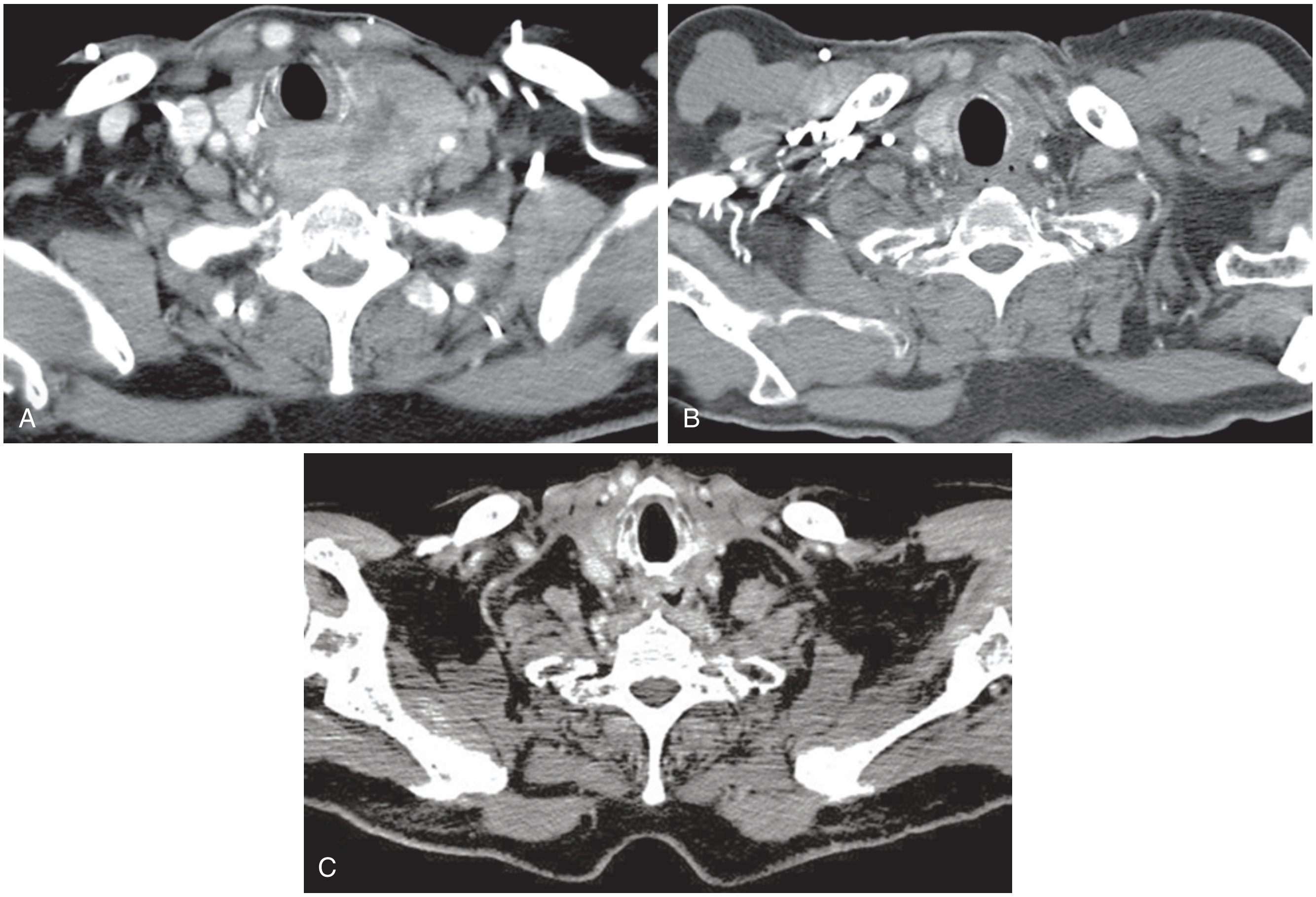


Fig. 80.2 Axial computed tomography neck imaging of a patient with stage IVB anaplastic thyroid carcinoma at diagnosis (**A**), 8 months after intensity-modulated radiation therapy and systemic chemotherapy (**B**), and 10 years later (**C**).

local regional lymph node involvement, or both, but is without distant metastases (Fig. 80.2); stage IVC disease could have a primary tumor of any size, with or without lymph nodes, but distant disease is present (Fig. 80.3). The median percentages of patients with stages IVA, B, and C disease are 10.6%, 48.3%, and 40.7%, respectively, as depicted in Table 80.3.[1,4-6,8,12,13,50-52]

ESTABLISH GOALS OF CARE

Multidisciplinary Team

Patients may present to their endocrinologist because ATC often arises after a history of longstanding goiter or prior papillary thyroid cancer. Given the rapid onset of an expanding neck mass, they may present to the emergency room with breathing difficulty and first see a surgeon. Other practitioners who should be consulted early include a radiation oncologist and a medical oncologist, and a pain/palliative care physician, social worker, psychologist or psychiatrist, and hospice physician should be strongly considered. A coordinated plan should be rapidly formulated by this team as soon as the staging is complete, and the realistic options then presented to the patient and his or her family/caregiver.[14,44,45]

It is important to establish that the patient has adequate decision-making capacity and is capable of providing informed consent. If the patient is unable to do so, then he or she should have a surrogate decision maker to guarantee adequate counseling. The risks and benefits of treatment options, which range from aggressive multimodal care to palliative care and hospice, should be disclosed. Included should be discussion of airway management,[44,45] nutrition, pain control, psychosocial issues, and advanced directives.[14,20]

KEY POINTS

- A multi-disciplinary team should be established to review all available treatment options, though also taking into account the patient's and family's wishes.
- Decision on goals of care should be determined early, though remains a fluid discussion.

PROGNOSIS

Predictors of Survival

A 2016 review determined that distant metastases carried a worse prognosis in 86% of patients enrolled in 14 studies, more extensive surgery improved survival in 78% of patients enrolled in 23 studies, radiotherapy (RT) improved survival in 74% of patients enrolled in 19 studies. In addition, tumor size was reported to influence survival in eight of 12 studies. Furthermore, age influenced survival in 56% of patients, ATC was associated with differentiated thyroid cancer in 50% of patients, and gender had no effect on survival.[53] More recently, numerous studies have shown the benefit of bimodal or trimodal therapy, high-dose RT (≥59 Gy), absence of distant metastases, younger age, smaller tumor size, and positive tumor margin.[4,6-8,11-13,20,50-52,54-67] Low poverty level and median family income have predicted survival,[55,59] as have a low lymphocyte-to-monocyte ratio[68] and leukocytosis.[62] In a review of 15 studies, surgery plus RT versus RT alone reduced risk of death (hazard ratio [HR] = 0.56), and in four reports of stage IVA patients, surgery plus RT improved survival (HR = 0.36).[54] Lai et al.[69] noted significantly shorter survival in patients whose tumors harbored a *RAS* or *PIK3CA* mutation.

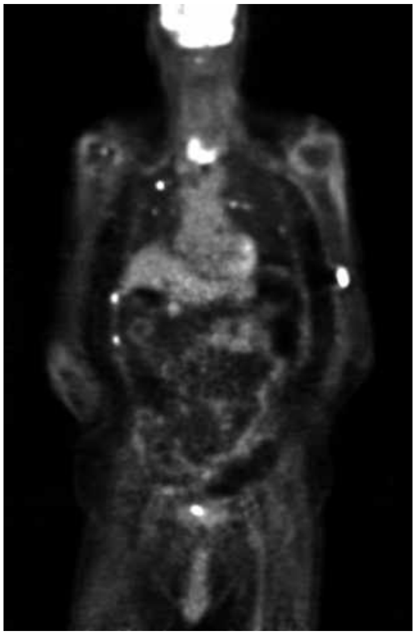


Fig. 80.3 Positron emission tomography/computed tomography imaging of patient with stage IVC anaplastic thyroid carcinoma and distant metastases to the lungs, liver, and skeleton.

TABLE 80.3 Tumor-Node-Metastasis Staging in Anaplastic Thyroid Carcinoma (16 series; *n* = 1907 patients)

Stage	Median Prevalence at Diagnosis % (Range)
IVA	10.6 (3–20)
IVB	48.3 (29–83)
IVC	40.7 (0–68)

(From references 1, 4-6, 8, 12, 13, 50-52.)

Median and Overall Survival

In a 2016 review,[53] one study reported a median survival of 33.5, 6.1, and 2.5 months in stage IVA, IVB, and IVC patients, respectively, whereas 1-year overall survival (OS) rates in these patient groups were 72.7%, 24.8%, and 8.2%, respectively. A second study included in the same review found a median survival of 9 months if tumor was confined to the thyroid, 6 months if adjacent structures were invaded, and 3 months if there was further extension or distant metastases. The 1-year survival rates for these groups were 50.0%, 27.6%, and 7.4%, respectively.

Studies performed from 2017 to 2020 have reported similar results, although recently the use of aggressive multimodal therapy, the introduction of tyrosine kinase inhibitors directed at targetable oncogenic mutations, and perhaps immunotherapy have shown some impressive improvements in response rates and duration. The median survival in these reports was 5.3 months (range: 2.0–10.6 months).[6,12,13,50,55,63,65,70-72] Only three studies provided results according to TNM stage, with median survivals in stages IVA, IVB, and IVC of 14.1, 8.2, and 5.3 months,[71] not reached (NR), 12.3, and 7.5 months,[5] and 69, 8, and 3.9 months, respectively.[65]

In a compilation of recent studies, the percentage of OS at 1 year was 27.9% (range: 0%–72%). Patient groups were heterogeneous, with many including all stages, and a variety of therapies were applied, with some studies including only patients receiving palliative care [4-6,8,12,13,51,52,58,62,65,70,71,73]; in one study grouping patients with incidental anaplastic foci, OS increased to 72%.[62] In a study of eight patients with less than 10% of their tumor being ATC (most were associated with papillary thyroid carcinoma), the 1-year OS rate was 88%, and the 2-year OS rate was 43%.[74] None of the tumors were stage IVA. In the study by Park et al.,[65] OS at 1 year was 29%, with 80% survival in stage IVA, 25% in stage IVB, and 19% in stage IVC patients. While multiple factors influenced length of survival, the most striking factor affecting duration of survival was tumor stage at time of diagnosis.

THERAPY

Treatment Strategies by Stage

Algorithms for initiating and monitoring therapy for each stage of ATC were provided by the American Thyroid Association in 2021,[14] The National Comprehensive Cancer Network also publishes treatment strategies and algorithms annually.[75] Several other recent approaches have incorporated molecular profiling to identify actionable tumor mutations for targeted therapies.[16,20] Rapid, accurate staging is critical, followed by establishing goals of care with each patient, as depicted in Fig. 80.4. These strategies have begun to show some promise in improving survival statistics, as will be reviewed in this section.

Surgery

A review of 1567 cases in the Surveillance, Epidemiology, and End Results database determined that total thyroidectomy down to complete or microscopic dissection, and when combined with other therapies, prolonged OS compared with other treatments.[55] Presumably patients undergoing extensive surgery had smaller tumors and were less likely to have systemic metastases, both of which would portend a more favorable outcome.

Radiation Therapy

A second United States registry, the National Cancer Database (NCDB), includes more than 79% of all new cancer cases seen at more than 1500 Commission on Cancer-accredited facilities. Using this database, Glaser et al.[56] reviewed outcomes of 3552 patients and found that patients had better survival if they received more than 59.4 Gy of radiation. A smaller subset of NCDB patients consisted of 496 who received nonpalliative surgery; 375 received adjuvant external beam radiation therapy (EBRT), of which 198 also received chemotherapy. In those patients who received EBRT after surgery versus those who did not, the OS increased from 9.1 to 12.3 months (P = .004). In those who received with EBRT with or without chemotherapy, the OS increased from 9.1 to 14.0 months.[57]

In a study of 27 patients, the median survival was 3.6 months if they received palliative RT but 11.8 months if they received definitive RT.[50] Durable locoregional control was achieved in 70% of patients, and there were five long-term survivors, whose survival ranged from 18.5 to 157 months. Park et al.[65] treated 41 patients and found that the overall median progression-free survival (PFS) increased from 2.5 to 4.5 months with RT and to 5.9 months if patients had surgery also. Median PFS was 69, 8, and 3.9 months for patients with stage IVA, IVB, and IVC, respectively. In the latter study, median OS increased from 3.9 to 7.2 months after RT and to 10.7 months if patients also underwent surgery. OS was 80%, 25%, and 19% in stages IVA, IVB, and IVC, respectively. Independent factors for survival were stage IVA, surgery, and intensity-modulated radiation therapy (better than three-dimensional conformal radiation therapy).

Multimodal Therapy

Multimodal therapy can be either bimodal (surgery + radiation, surgery + systemic, or radiation + systemic) or trimodal (surgery, radiation, and systemic therapy); systemic therapy includes cytotoxic agents, multikinase inhibitors, and immunotherapeutics.[76] For reference,

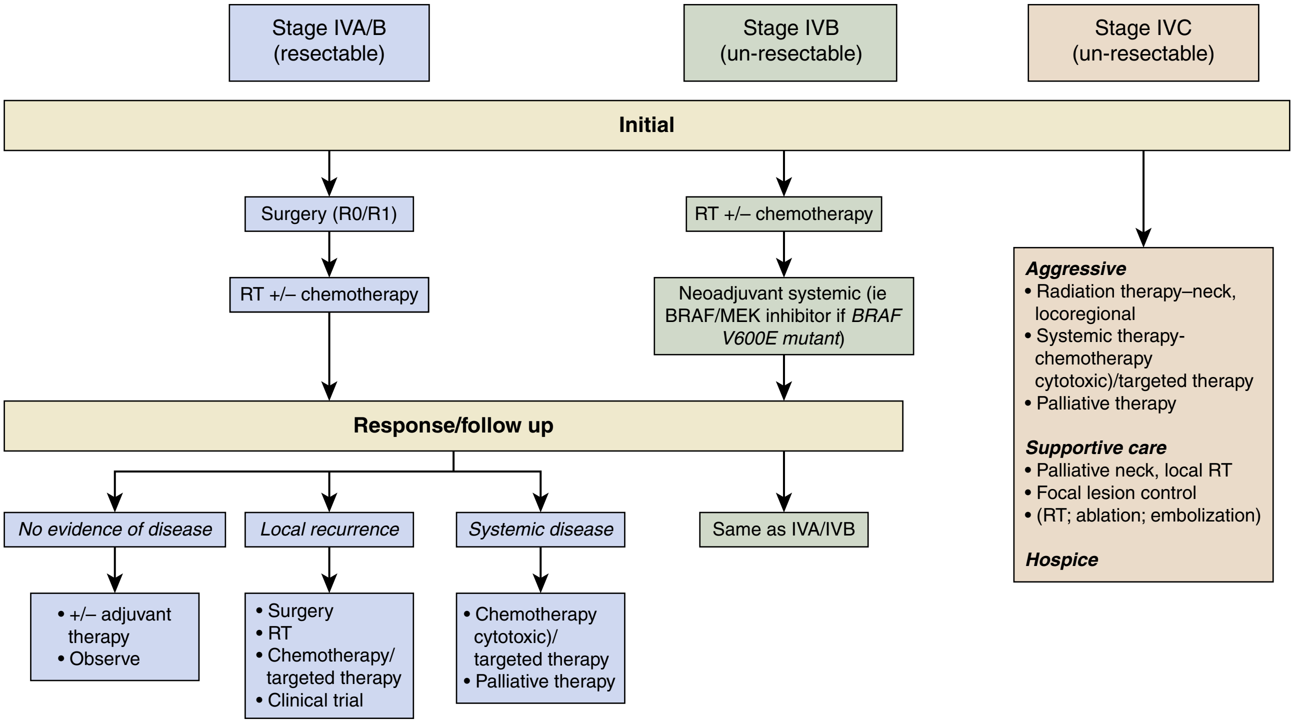

Stage IVA/B
(resectable)
Stage IVB
(un-resectable)
Stage IVC
(un-resectable)
Initial
Surgery (R0/R1)
RT +/– chemotherapy
RT +/– chemotherapy
Neoadjuvant systemic (ie BRAF/MEK inhibitor if BRAF V600E mutant)
Aggressive
• Radiation therapy–neck, locoregional
• Systemic therapy-chemotherapy cytotoxic)/targeted therapy
• Palliative therapy
Supportive care
• Palliative neck, local RT
• Focal lesion control
• (RT; ablation; embolization)
Hospice
Response/follow up
No evidence of disease
Local recurrence
Systemic disease
Same as IVA/IVB
• +/– adjuvant therapy
• Observe
• Surgery
• RT
• Chemotherapy/ targeted therapy
• Clinical trial
• Chemotherapy cytotoxic)/ targeted therapy
• Palliative therapy

Fig. 80.4 Algorithms for Anaplastic Thyroid Carcinoma Management by American Joint Committee on Cancer. *R0,* complete resection; *R1*, microscopic margins; *RT*, radiotherapy; *chemo*, chemotherapy; *NED*, no evidence of disease.

there are several reviews of the variety of drugs available and being tested,[77,78] as well as one on ABC transporters and chemoresistance.[79] In a review of 17 phase I trials (16 single-arm, one randomized), the authors concluded that the outcomes of targeted therapy were "generally disappointing" but may be improved by genetic profiling.[80]

Brignardello et al.[81] treated 55 stage IVB to stage IVC patients with "maximal debulking" (R0, R1, or R2 resections with minimal residual macroscopic disease), with prolonged survival and improved quality of life in both stage IVB and IVC patients. However, a review of 680 ATC patients in the NCDB, of whom 335 underwent surgery, recommended surgical management for intrathyroidal disease, but not necessarily in stage IVB and IVC patients.[82]

Prasongsook et al.[6] treated 48 patients (30 multimodal and 18 palliative). In the former group, 90% had surgery, average radiation was 66 Gy (range: 46–70 Gy), and cytotoxic therapies included doxorubicin, a platin, and a taxane, either as single agent or as a two-drug combination. The multimodal patients had a median PFS of 8.3 months versus 3 months in the palliative patients, and median OS was prolonged from 3.9 to 21 months. Patients with stage IVB disease had a 1-year OS of 68% with multimodal therapy versus 0% if palliative. Toxicities led to hospitalization in 60% and feeding tubes in 60%, emphasizing the importance of identifying patients able to withstand such aggressive treatment.

Rao et al.[5] reported that, in 54 patients (50% of whom were stage IVC), trimodal therapy increased median OS from 6.5 to 22.1 months versus dual therapy (EBRT/chemotherapy). A follow-up study of 479 patients from the same institution[52] found that 1-year OS was 35% from 2000 to 2013 but increased to 47% between 2014 and 2016 with the institution of a Facilitating ATC Specialized Treatment program and to 59% when targeted therapies and immunotherapeutics were introduced. Median OS increased from 7.6 to 15.7 months (HR = 0.49) with targeted/immunotherapy. In a subset of stage IVC patients, 1-year OS increased from 22% to 39% to 52%.

In another analysis of the NCDB, 858 patients were propensity score-matched to compare chemoradiotherapy (CRT) to RT.[58] The 1-year survival was 25.5% in the CRT group versus 14.3% in the RT group (HR = 0.66; $P < .001$). Corrigan et al.[51] reported that, in 28 patients, 1-year OS was 0% without EBRT but 17.9% with EBRT and increased further to 25% if surgery was also performed. They concluded that EBRT plus surgery improved survival. A study by Lee et al.[13] found that multimodal therapy prolonged median survival from 88 to 225 days.

The introduction of targeted therapy, based on molecular characterization of a patient's tumor, has recently shown promise in ATC. Monotherapy with dabrafenib in two patients showed short-term responses.[83] Most encouraging has been the dramatic improvement in OS with the combination of dabrafenib (a BRAF inhibitor) and trametinib (a MEK inhibitor) in patients with a *BRAF-V600E* mutation. Agarwal et al.[84] reported a CT imaging response for 6 months in a single patient, and a small case series of six patients (stage IVB in four; stage IVC in two) used this combination in the neoadjuvant setting, which allowed complete surgical resection and adjuvant chemoradiation.[85] OS at 1 year was 83%. Iyer et al.[86] treated ten patients with lenvatinib and 16 with dabrafenib and trametinib. The median PFS was 2.7 versus 5.2 months, and the median OS was 3.9 versus 9.3 months for lenvatinib versus dabrafenib/trametinib, respectively. Subbiah et al.[87] performed a phase II open-label trial of dabrafenib/trametinib in 16 patients; all had received prior RT and/or surgery, and six had received prior systemic therapy. While median PFS and OS were not achieved, 12-month estimates were 79% and 80%, respectively. The overall response rate was 69% (11/16). Common adverse events included fatigue (38%), pyrexia (37%), and nausea (35%). This trial led to US Food and Drug Administration approval for ATC. An earlier case report using the BRAF inhibitor vemurafenib produced a 61-week response before tumor progression.[88]

Lenvatinib is a multikinase inhibitor (targets include VEGF receptors 1–3; FGFR1–4; PDGFR-α; RET; and KIT). Iwasaki and colleagues[63] treated 16 patients with lenvatinib and 16 with palliative weekly paclitaxel. The median OS was increased from 2.0 to 4.2 months, and clinical partial responses (PRs) were achieved in five patients, but side effects were consequential, with seven developing fistulas and two exhibiting massive tumor necrosis and bleeding. In a report of 14 patients given lenvatinib (eight in the neoadjuvant setting), PRs were obtained in four patients and stable disease in nine, with a median PFS of 5.7 months and a median OS of 6.7 months.[72] Adverse events included tracheal perforation and pneumothorax/pneumomediastinum in two, as well as one fatality because of rapid tumor shrinkage. The authors concluded that there was "limited clinical benefit and there was doubtful survival benefit." The largest study of lenvatinib, from Japan, included 124 ATC patients with a median age of 73 years (range: 43–85 years); 36% were women, 53% had prior surgery, and 19% had prior systemic therapy.[10] Median OS was 3.3 months, and the 1-year survival rate was 15.6%. Responses included complete response in 3%, PR in 41%, and stable disease in 34%. Fatal adverse events were reported in 4.8%, and 8.1% had either carotid artery or venous hemorrhage, or tumor hemorrhage associated with necrosis or shrinkage.

Everolimus is an mTOR kinase inhibitor, and two small case series have been performed in ATC. In one report of seven patients, one had stable disease for 31 months and had mutations of mTOR, PIK3CA, and NF1,[89] and one patient whose tumor had PIK3CA and BRAF mutations had a "response." Both cases had TSC2 mutations. In a second study of five patients (two with stage IVB and three with stage IVC) who received everolimus, one patient had a PR for 27.9 months (this patient had activating mutations of *PIK3CA* and *BRAF*).[90]

Genetic abnormalities have directed therapy in several case reports, including one patient with a CCDC6-RET fusion who had lung and brain metastases and a greater than 19-month PR while on a trial with LOXO-292, targeting RET-fusion tumors.[91] A second patient with an *ALK* rearrangement had a 36-month response to crizotinib.[92] Rebiopsy showed a STRN-ALK fusion, and the patient then had a 16-month response to ceritinib; when cerebral progression developed, brigatinib led to a rapid PR.[93]

Finally, immunotherapeutics have been tried in animal models and in several patients. Kollipara et al.[94] had one patient with stage IVC disease and whose tumor expressed PD-L1 (3+) and PD-1 (1+); nivolumab therapy produced a complete response for 20 months. Iyer et al.[95] showed that five of 12 patients had a PR when pembrolizumab was added after progression on tyrosine kinase therapy, with median OS and PFS of 6.93 and 2.96 months, respectively.

Spalart et al.[96] reported one patient with stage IVC disease and PD-L1 immunopositivity in 60% of cells who responded to pembrolizumab for 16 months before progression. Cabanillas et al.[97] had one patient who received neoadjuvant dabrafenib and trametinib and pembrolimuzab, which permitted subsequent surgery. In another patient with stage IVC disease whose tumor was 60% positive for PD-L1, there was a dramatic response to spartalizumab (a PD-1 receptor monoclonal antibody), three patients (7%) had a complete response, and five (12%) had a PR; all responders expressed PD-L1. One-year OS was 52% in the patients expressing PD-L1.[73]

KEY POINTS

- Multi-modal therapy involving surgery (if possible to achieve complete resection), radiation and systemic therapy is the optimal approach in the management of anaplastic thyroid cancer.
- Dabrafenib and trametinib are tyrosine kinase inhibitors that specifically target BRAF-V600E mutation, and is approved for use in anaplastic thyroid cancer patients whose tumors harbor this mutation.

FUTURE DIRECTIONS

Multiple preclinical and clinical studies support the hope that additional strategies will be forthcoming. In the laboratory, newer models include patient-derived xenografts[98] and tumor organoids.[99] Additional therapeutic targets include miRNAs, lncRNAs,[42,43] histone deacetylases,[100] heat shock proteins,[101] CD47 expression,[102] to enhance tumor phagocytosis by macrophages, and CRISPR/Cas9 editing of the gene encoding the epidermal growth factor receptor.[103] Chimeric antigen receptor T cell therapy is also being tried for solid tumors, including the possibility of ATC and other refractory thyroid cancers.[104] Clearly, further progress against this deadly malignancy will benefit from novel approaches.

CONCLUSIONS

ATC is one of the most aggressive and lethal of all human malignancies. Considerable information regarding its presentation and clinical course has been accumulated over many decades, with only minimal improvement in clinical outcomes. Recently, the employment of aggressive multi/trimodal therapy with cytotoxic drugs and tyrosine kinase inhibitors, especially a BRAF/MEK inhibitor combination, has shown benefit in OS. The interrogation of patient tumors is yielding important information on the remarkable genetic and epigenetic heterogeneity of these tumors and has provided insights into targeted therapies based on an individual patient's tumor. It is hopeful that multiple agents directed at disparate oncogenic drivers will further improve patient outcomes in the not-too-distant future.

ACKNOWLEDGMENTS

Thank you to Aziza Nassar, MD and Michael Rivera, MD for providing pathologic imaging.

REFERENCES

1. Smallridge RC, Abate EG. Anaplastic thyroid cancer: clinical aspects. In: Wartofsky L, Van Nostrand D, eds. *Thyroid Cancer*. 3rd ed. New York: Springer; 2016:929–932.
2. Besic N, Gazic B. Dietary iodine intake, therapy with radioiodine, and anaplastic thyroid carcinoma. *Radiol Oncol*. 2020;54:187–193.
3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin*. 2021;71(1):7–33.
4. Lee DY, Won JK, Lee SH, et al. Changes of clinicopathologic characteristics and survival outcomes of anaplastic and poorly differentiated thyroid carcinoma. *Thyroid*. 2016;26:404–413.
5. Rao SN, Zafereo M, Dadu R, et al. Patterns of treatment failure in anaplastic thyroid carcinoma. *Thyroid*. 2017;27:672–681.
6. Prasongsook N, Kumar A, Chintakuntlawar AV, et al. Survival in response to multimodal therapy in anaplastic thyroid cancer. *J Clin Endocrinol Metab*. 2017;102:4506–4514.
7. Pradhan R, Agarwal A, Lal P, et al. Clinico-pathological profile of anaplastic thyroid carcinoma in an endemic goiter area. *Indian J Endocrinol Metab*. 2018;22:793–797.
8. Hvilsom GB, Londero SC, Hahn CH, et al. Anaplastic thyroid carcinoma in Denmark 1996-2012: A national prospective study of 219 patients. *Cancer Epidemiol*. 2018;53:65–71.
9. Xu B, Fuchs T, Dogan S, et al. Dissecting anaplastic thyroid carcinoma: a comprehensive clinical, histologic, immunophenotypic, and molecular study of 360 cases. *Thyroid*. 2020;30(10):1505–1517.
10. Takahashi S, Kiyota N, Yamazaki T, et al. A Phase II study of the safety and efficacy of lenvatinib in patients with advanced thyroid cancer. *Future Oncol*. 2019;15:717–726.
11. de Ridder M, Nieveen van Dijkum E, et al. Anaplastic thyroid carcinoma: a nationwide cohort study on incidence, treatment and survival in the Netherlands over 3 decades. *Eur J Endocrinol*. 2020;183:203–209.
12. Fan D, Ma J, Bell AC, et al. Outcomes of multimodal therapy in a large series of patients with anaplastic thyroid cancer. *Cancer*. 2020;126:444–452.
13. Lee JH, Ahn HK, Seok JY, et al. Optimal combination of treatment modality to increase survival in patients with anaplastic thyroid carcinoma: a STROBE compliant retrospective study. *Medicine (Baltimore)*. 2018;97:e11037.

14. Bible KC, Kebebew E, Brierley J, Brito JP, Cabanillas ME, Clark TJ Jr., et al. 2021 American Thyroid Association Guidelines for Management of Patients with Anaplastic Thyroid Cancer. *Thyroid*. 2021;31(3):337–386.
15. Talbott I, Wakely Jr PE. Undifferentiated (anaplastic) thyroid carcinoma: Practical immunohistochemistry and cytologic look-alikes. *Semin Diagn Pathol*. 2015;32:305–310.
16. Haugen BR, Bible KC, Smallridge RC. Poorly differentiated thyroid cancer, anaplastic thyroid cancer, and miscellaneous tumors of the thyroid. In: Braverman LE, Cooper DS, Kopp PA., eds. *Werner & Ingbar's The Thyroid: A Fundamental and Clinical Text*. 11th ed. Wolters Kluwer; 2020:796–812.
17. Yang J, Barletta JA. Anaplastic thyroid carcinoma. *Semin Diagn Pathol*. 2020;37(5):248–256.
18. El-Naggar AK, Baloch ZW, Eng C, et al. Anaplastic thyroid carcinoma In: Lloyd RV, Osamura RY, Kloppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon: International Agency for Research on Cancer; 2017:104–106.
19. Hirokawa M, Sugitani I, Kakudo K, et al. Histopathological analysis of anaplastic thyroid carcinoma cases with long-term survival: a report from the Anaplastic Thyroid Carcinoma Research Consortium of Japan. *Endocr J*. 2016;63:441–447.
20. Smallridge RC, Bible KC. Anaplastic thyroid carcinoma. In: Luster M, Duntas LH, Wartofsky L. eds. *The Thyroid and its Diseases*. Cham, Switzerland: Springer; 2019:693–700.
21. Deeken-Draisey A, Yang GY, Gao J, et al. Anaplastic thyroid carcinoma: an epidemiologic, histologic, immunohistochemical, and molecular single-institution study. *Hum Pathol*. 2018;82:140–148.
22. Oishi N, Kondo T, Ebina A, et al. Molecular alterations of coexisting thyroid papillary carcinoma and anaplastic carcinoma: identification of TERT mutation as an independent risk factor for transformation. *Mod Pathol*. 2017;30:1527–1537.
23. Smith N, Nucera C. Personalized therapy in patients with anaplastic thyroid cancer: targeting genetic and epigenetic alterations. *J Clin Endocrinol Metab*. 2015;100:35–42.
24. Kunstman JW, Juhlin CC, Goh G, et al. Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum Mol Genet*. 2015;24:2318–2329.
25. Jeon MJ, Chun SM, Kim D, et al. Genomic alterations of anaplastic thyroid carcinoma detected by targeted massive parallel sequencing in a BRAF(V600E) mutation-prevalent area. *Thyroid*. 2016;26:683–690.
26. Tiedje V, Ting S, Herold T, et al. NGS based identification of mutational hotspots for targeted therapy in anaplastic thyroid carcinoma. *Oncotarget*. 2017;8:42613–42620.
27. Pozdeyev N, Gay LM, Sokol ES, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. *Clin Cancer Res*. 2018;24:3059–3068.
28. Chen H, Luthra R, Routbort MJ, et al. Molecular profile of advanced thyroid carcinomas by next-generation sequencing: characterizing tumors beyond diagnosis for targeted therapy. *Mol Cancer Ther*. 2018;17:1575–1584.
29. Duan H, Li Y, Hu P, et al. Mutational profiling of poorly differentiated and anaplastic thyroid carcinoma by the use of targeted next-generation sequencing. *Histopathology*. 2019;75:890–899.
30. Iniguez-Ariza NM, Jasim S, Ryder MM, et al. Foundation one genomic interrogation of thyroid cancers in patients with metastatic disease requiring systemic therapy. *J Clin Endocrinol Metab*. 2020;105(7):e2346–e2357.
31. Landa I, Pozdeyev N, Korch C, et al. Comprehensive genetic characterization of human thyroid cancer cell lines: a validated panel for preclinical studies. *Clin Cancer Res*. 2019;25:3141–3151.
32. Khan SA, Ci B, Xie Y, et al. Unique mutation patterns in anaplastic thyroid cancer identified by comprehensive genomic profiling. *Head Neck*. 2019;41:1928–1934.
33. Yoo SK, Song YS, Lee EK, et al. Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. *Nat Commun*. 2019;10:2764.12 pages
34. Ravi N, Yang M, Gretarsson S, et al. Identification of targetable lesions in anaplastic thyroid cancer by genome profiling. *Cancers (Basel)*. 2019;11(3): 402. (13 pages).
35. Yakushina VD, Lerner LV, Lavrov AV. Gene fusions in thyroid cancer. *Thyroid*. 2018;28:158–167.
36. Hu S, Liao Y, Chen L. Identification of key pathways and genes in anaplastic thyroid carcinoma via integrated bioinformatics analysis. *Med Sci Monit*. 2018;24:6438–6448.
37. Ma Y, Cang S, Li G, et al. Integrated analysis of transcriptome data revealed MMP3 and MMP13 as critical genes in anaplastic thyroid cancer progression. *J Cell Physiol*. 2019;234:22260–22271.
38. Pan Z, Li L, Fang Q, et al. Integrated bioinformatics analysis of master regulators in anaplastic thyroid carcinoma. *Biomed Res Int*. 2019:9734576.(13 pages).
39. Gao X, Wang J, Zhang S. Integrated bioinformatics analysis of hub genes and pathways in anaplastic thyroid carcinomas. *Int J Endocrinol*. 2019:9651380.(9 pages).
40. Sasanakietkul T, Murtha TD, Javid M, et al. Epigenetic modifications in poorly differentiated and anaplastic thyroid cancer. *Mol Cell Endocrinol*. 2018;469:23–37.
41. Vosgha H, Ariana A, Smith RA, et al. miR-205 targets angiogenesis and EMT concurrently in anaplastic thyroid carcinoma. *Endocr Relat Cancer*. 2018;25:323–337.
42. Wang Y, Hardin H, Chu YH, et al. Long non-coding RNA expression in anaplastic thyroid carcinomas. *Endocr Pathol*. 2019;30:262–269.
43. Tan X, Wang P, Lou J, et al. Knockdown of lncRNA NEAT1 suppresses hypoxia-induced migration, invasion and glycolysis in anaplastic thyroid carcinoma cells through regulation of miR-206 and miR-599. *Cancer Cell Int*. 2020;20:132.(12 pages)
44. Mani N, McNamara K, Lowe N, et al. Management of the compromised airway and role of tracheotomy in anaplastic thyroid carcinoma. *Head Neck*. 2016;38:85–88.
45. Moyer KF, Marcadis AR, Shaha AR. Airway management, symptom relief and best supportive care in anaplastic thyroid cancer. *Curr Opin Otolaryngol Head Neck Surg*. 2020;28:74–78.
46. Ciarallo A, Marcus C, Taghipour M, et al. Value of fluorodeoxyglucose PET/computed tomography patient management and outcomes in thyroid cancer. *PET Clin*. 2015;10:265–278.
47. Ahmed S, Ghazarian MP, Cabanillas ME, et al. Imaging of anaplastic thyroid carcinoma. *AJNR Am J Neuroradiol*. 2018;39:547–551.
48. Chintakuntlawar AV, Foote RL, Kasperbauer JL, et al. Diagnosis and management of anaplastic thyroid cancer. *Endocrinol Metab Clin North Am*. 2019;48:269–284.
49. Tuttle RM, Haugen B, Perrier ND. Updated American Joint Committee on Cancer/tumor-node-metastasis staging system for differentiated and anaplastic thyroid cancer (eighth edition): what changed and why? *Thyroid*. 2017;27:751–756.
50. So K, Smith RE, Davis SR. Radiotherapy in anaplastic thyroid carcinoma: an Australian experience. *J Med Imaging Radiat Oncol*. 2017;61:279–287.
51. Corrigan KL, Williamson H, Elliott Range D, et al. Treatment outcomes in anaplastic thyroid cancer. *J Thyroid Res*. 2019;8218949. (11 pages)
52. Maniakas A, Dadu R, Busaidy NL, et al. Evaluation of overall survival in patients with anaplastic thyroid carcinoma, 2000-2019. *JAMA Oncol*. 2020;6(9):1397–1404.
53. Smallridge RC, Abate EG. Anaplastic thyroid carcinoma: prognosis. In: Wartofsky L, Van Nostrand D, eds. *Thyroid Cancer*. 3rd ed. New York: Springer; 2016:953–956.
54. Kwon J, Kim BH, Jung HW, et al. The prognostic impacts of postoperative radiotherapy in the patients with resected anaplastic thyroid carcinoma: a systematic review and meta-analysis. *Eur J Cancer*. 2016;59:34–45.
55. Lin B, Ma H, Ma M, et al. The incidence and survival analysis for anaplastic thyroid cancer: a SEER database analysis. *Am J Transl Res*. 2019;11:5888–5896.
56. Glaser SM, Mandish SF, Gill BS, et al. Anaplastic thyroid cancer: prognostic factors, patterns of care, and overall survival. *Head Neck*. 2016;38:E2083–E2090.
57. Saeed NA, Kelly JR, Deshpande HA, et al. Adjuvant external beam radiotherapy for surgically resected, nonmetastatic anaplastic thyroid cancer. *Head Neck*. 2020;42:1031–1044.
58. Tian S, Switchenko JM, Fei T, et al. Survival advantage of chemoradiotherapy in anaplastic thyroid carcinoma: propensity score matched analysis with multiple subgroups. *Head Neck*. 2020;42:678–687.

59. Li MU, Gu S, Mao R, et al. County median family income is an independent prognostic factor for stage IV anaplastic thyroid cancer. *Anticancer Res*. 2019;39:949–956.
60. Janz TA, Neskey DM, Nguyen SA, et al. Is the incidence of anaplastic thyroid cancer increasing: A population based epidemiology study. *World J Otorhinolaryngol Head Neck Surg*. 2019;5:34–40.
61. Qiu B, Li HQ, Chang QG, et al. Nomograms predict survival in patients with anaplastic thyroid carcinoma. *Med Sci Monit*. 2019;25:8447–8456.
62. Sugitani I, Onoda N, Ito KI, et al. Management of anaplastic thyroid carcinoma: the fruits from the ATC research consortium of Japan. *J Nippon Med Sch*. 2018;85:18–27.
63. Iwasaki H, Toda S, Suganuma N, et al. Lenvatinib vs. palliative therapy for stage IVC anaplastic thyroid cancer. *Mol Clin Oncol*. 2020;12:138–143.
64. Xia Q, Wang W, Xu J, etal. Evidence from an updated meta-analysis of the prognostic impacts of postoperative radiotherapy and chemotherapy in patients with anaplastic thyroid carcinoma. *Onco Targets Ther*. 2018;11:2251–2257.
65. Park JW, Choi SH, Yoon HI, et al. Treatment outcomes of radiotherapy for anaplastic thyroid cancer. *Radiat Oncol J*. 2018;36:103–113.
66. Nachalon Y, Stern-Shavit S, Bachar G, et al. Aggressive palliation and survival in anaplastic thyroid carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2015;141:1128–1132.
67. Salehian B, Liem SY, Mojazi Amiri H, et al. Clinical Trials in management of anaplastic thyroid carcinoma; progressions and set backs: a systematic review. *Int J Endocrinol Metab*. 2019;17:e67759. (10 pages).
68. Ahn J, Song E, Oh HS, et al. Low lymphocyte-to-monocyte ratios are associated with poor overall survival in anaplastic thyroid carcinoma patients. 2019; 29:824–829.
69. Lai WA, Liu CY, Lin SY, et al. Characterization of driver mutations in anaplastic thyroid carcinoma identifies RAS and PIK3CA mutations as negative survival predictors. *Cancers (Basel).* 2020;12(7):1973 (13 pages).
70. Sosa JA, Elisei R, Jarzab B, et al. Randomized safety and efficacy study of fosbretabulin with paclitaxel/carboplatin against anaplastic thyroid carcinoma. *Thyroid*. 2014;24:232–240.
71. Onoda N, Sugino K, Higashiyama T, et al. The safety and efficacy of weekly paclitaxel administration for anaplastic thyroid cancer patients: a nationwide prospective study. *Thyroid*. 2016;26:1293–1299.
72. Kim M, Ahn J, Song DE, et al. Real-world experience of lenvatinib in patients with advanced anaplastic thyroid cancer. *Endocrine*. 2021;71:427–433.
73. Capdevila J, Wirth LJ, Ernst T, et al. PD-1 blockade in anaplastic thyroid carcinoma. *J Clin Oncol*. 2020;38:2620–2627.
74. Wong KS, Lorch JH, Alexander EK, et al. Clinical outcome of anaplastic thyroid carcinoma with a minor anaplastic component. *Endocr Pathol*. 2020;31:283–290.
75. National Comprehensive Cancer Network. Thyroid Cancer (Version 3.2021). https://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf. Accessed on November 3, 2021.
76. Abate EG, Smallridge RC. Unravelling the best combination of therapies to treat anaplastic thyroid cancer. *Expert Rev Endocrinol Metab*. 2016;11:235–237.
77. Li ZJ, Zhang Y, Wang RN. Genetic alterations in anaplastic thyroid carcinoma and targeted therapies. *Exp Ther Med*. 2019;18:2369–2377.
78. Cabanillas ME, Ryder M, Jimenez C. Targeted therapy for advanced thyroid cancer: kinase inhibitors and beyond. *Endocr Rev*. 2019;40:1573–1604.
79. Abbasifarid E, Sajjadi-Jazi SM, Beheshtian M, et al. The role of ATP-binding cassette transporters in the chemoresistance of anaplastic thyroid cancer: a systematic review. *Endocrinology*. 2019;160:2015–2023.
80. Ljubas J, Ovesen T, Rusan M. A systematic review of phase II targeted therapy clinical trials in anaplastic thyroid cancer. *Cancers (Basel)*. 2019;11:943 (17 pages).
81. Brignardello E, Palestini N, Felicetti F, et al. Early surgery and survival of patients with anaplastic thyroid carcinoma: analysis of a case series referred to a single institution between 1999 and 2012. *Thyroid*. 2014;24: 1600–1606.
82. Goffredo P, Thomas SM, Adam MA, et al. Impact of timeliness of resection and thyroidectomy margin status on survival for patients with anaplastic thyroid cancer: an analysis of 335 cases. *Ann Surg Oncol*. 2015;22:4166–4174.
83. Lim AM, Taylor GR, Fellowes A, et al. BRAF inhibition in BRAFV600E-positive anaplastic thyroid carcinoma. *J Natl Compr Canc Netw*. 2016;14:249–254.
84. Agarwal R, Wang J, Wilson K, et al. Response to targeted therapy in BRAF mutant anaplastic thyroid cancer. *J Natl Compr Canc Netw*. 2016;14:1203–1207.
85. Wang JR, Zafereo ME, Dadu R, et al. Complete surgical resection following neoadjuvant dabrafenib plus trametinib in BRAF(V600E)-mutated anaplastic thyroid carcinoma. *Thyroid*. 2019;29:1036–1043.
86. Iyer PC, Dadu R, Ferrarotto R, et al. Real-world experience with targeted therapy for the treatment of anaplastic thyroid carcinoma. *Thyroid*. 2018;28:79–87.
87. Subbiah V, Kreitman RJ, Wainberg ZA, et al. Dabrafenib and trametinib treatment in patients with locally advanced or metastatic BRAF V600-mutant anaplastic thyroid cancer. *J Clin Oncol*. 2018;36:7–13.
88. Prager GW, Koperek O, Mayerhoefer ME, et al. Sustained response to vemurafenib in a BRAF(V600E)-mutated anaplastic thyroid carcinoma patient. *Thyroid*. 2016;26:1515–1516.
89. Hanna GJ, Busaidy NL, Chau NG, et al. Genomic correlates of response to everolimus in aggressive radioiodine-refractory thyroid cancer: a phase II study. *Clin Cancer Res*. 2018;24:1546–1553.
90. Harris EJ, Hanna GJ, Chau N, et al. Everolimus in anaplastic thyroid cancer: a case series. *Front Oncol*. 2019;9:106. (6 pages)
91. Dias-Santagata D, Lennerz JK, Sadow PM, et al. Response to RET-specific therapy in RET fusion-positive anaplastic thyroid carcinoma. *Thyroid*. 2020;30:1384–1389.
92. Godbert Y, Henriques de Figueiredo B, Bonichon F, et al. Remarkable response to crizotinib in woman with anaplastic lymphoma kinase-rearranged anaplastic thyroid carcinoma. *J Clin Oncol*. 2015;33:e84–e87.
93. Leroy L, Bonhomme B, Le Moulec S, et al. Remarkable response to ceritinib and brigatinib in an anaplastic lymphoma kinase-rearranged anaplastic thyroid carcinoma previously treated with crizotinib. *Thyroid*. 2020;30:343–344.
94. Kollipara R, Schneider B, Radovich M, et al. Exceptional response with immunotherapy in a patient with anaplastic thyroid cancer. *Oncologist*. 2017;22:1149–1151.
95. Iyer PC, Dadu R, Gule-Monroe M, et al. Salvage pembrolizumab added to kinase inhibitor therapy for the treatment of anaplastic thyroid carcinoma. *J Immunother Cancer*. 2018;6:68. (10 pages).
96. Spalart V, Legius B, Segers K, et al. Dramatic response to first line single agent pembrolizumab in anaplastic thyroid carcinoma. *Case Rep Endocrinol*. 2019;9095753. (4 pages).
97. Cabanillas ME, Ferrarotto R, Garden AS, et al. Neoadjuvant BRAF- and immune-directed therapy for anaplastic thyroid carcinoma. *Thyroid*. 2018;28:945–951.
98. Marlow LA, Rohl SD, Miller JL, et al. Methodology, criteria, and characterization of patient-matched thyroid cell lines and patient-derived tumor xenografts. *J Clin Endocrinol Metab*. 2018;103:3169–3182.
99. Weeber F, Ooft SN, Dijkstra KK, et al. Tumor organoids as a pre-clinical cancer model for drug discovery. *Cell Chem Biol*. 2017;24:1092–1100.
100. Spartalis E, Athanasiadis DI, Chrysikos D, et al. Histone deacetylase inhibitors and anaplastic thyroid carcinoma. *Anticancer Res*. 2019;39:1119–1127.
101. Lettini G, Pietrafesa M, Lepore S, et al. Heat shock proteins in thyroid malignancies: potential therapeutic targets for poorly-differentiated and anaplastic tumours? *Mol Cell Endocrinol*. 2020;502:110676. (9 pages).
102. Schurch CM, Roelli MA, Forster S, et al. Targeting CD47 in anaplastic thyroid carcinoma enhances tumor phagocytosis by macrophages and is a promising therapeutic strategy. *Thyroid*. 2019;29:979–992.
103. Huang LC, Tam KW, Liu WN, et al. CRISPR/Cas9 genome editing of epidermal growth factor receptor sufficiently abolished oncogenicity in anaplastic thyroid cancer. *Dis Markers*. 2018;3835783. (14 pages).
104. Vedvyas Y, McCloskey JE, Yang Y, et al. Manufacturing and preclinical validation of CAR T cells targeting ICAM-1 for advanced thyroid cancer therapy. *Sci Rep*. 2019;9:10634. (14 pages)

81

Resistance to Thyroid Hormone and Genetic Defects of the Thyroid

Carla Moran, Nadia Schoenmakers, Maura Agostini, Mark Gurnell, and Krishna Chatterjee

OUTLINE

Thyroid Gland Development, Hormone Biosynthesis, and Action, 1338
Congenital Primary Hypothyroidism, 1339
Epidemiology and Screening, 1339
Genetic Causes of Thyroid Dysgenesis, 1341
Genetic Causes of Dyshormonogenesis, 1343
Classification of Disorders of Thyroid Hormone Action, 1346
Differential Diagnosis of Elevated Thyroxine and/or Triiodothyronine with Nonsuppressed Thyroid-Stimulating Hormone, 1347
Resistance to Thyroid Hormone β, 1347
Clinical Features, 1347
Differential Diagnosis, 1348
Molecular Genetics, 1349
Management, 1351
Resistance to Thyroid Hormone Alpha, 1352
Background, 1352
Clinical Features, 1352
Molecular Genetics, 1353
Management, 1354

THYROID GLAND DEVELOPMENT, HORMONE BIOSYNTHESIS, AND ACTION

Thyroid development: Human thyroid development begins with the specification of a thyroid primordium on the median surface of the developing phyaryngeal floor, which evaginates from the endodermal epithelium at the base of the tongue around embryonic day (E) 24 and migrates caudally, initially in contact with the pharyngeal endoderm via the thyroglossal duct, which later degenerates. Throughout migration, the developing thyroid follicular cells proliferate and expand in close proximity to the aortic sac, and around E44 the ultimobranchial bodies deliver the parafollicular C cells, which fuse with the developing lateral lobes; although originally thought to derive from neural crest, C cells are now also recognized to be endodermal in origin. Once the thyroid reaches its final position anterior to the trachea, the lateral lobes enlarge, and synthesis of thyroglobulin (TG) begins. During the process of folliculogenesis, cellular functional differentiation and expression of key thyroidal genes including thyroid-stimulating hormone (TSH) receptor (*TSHR*) and thyroid peroxidase (*TPO*), followed by the sodium iodide symporter, (*SLC5A5, NIS*), prepare the fetal thyroid for thyroid hormone (TH) biosynthesis. NIS accumulation in the basolateral membrane is thought to represent the final step in thyrocyte differentiation. Fetal THs are detectable in the thyroid from approximately gestational week 11 and in fetal blood from gestational week 12.[1,2]

Thyroid transcription factors: Developing thyroid follicular cells are characterized by the combined expression of four key transcription factors from early on in development: NK2 homeobox 1 (NKX2-1, previously known as TTF1), paired box 8 (PAX8), haemopoetically expressed homeobox (HHEX), and forkhead box E1 (FOXE1, previously known as TTF-2). All interact with the promoter regions of key genes involved in thyroid differentiation (e.g., *TG, TPO,* and *NIS*), and their role in thyroid development has since been interrogated using rodent and cell culture models. NKX2-1 and PAX8 regulate expression of key genes (e.g., *TSHR, TG,* and *TPO*) involved in thyroid differentiation and play a role in murine thyroid precursor cell survival, as well as in maintenance of adult thyroid follicular cell differentiation and architecture. FOXE1 is required for differentiation and survival of the developing thyroid gland, as well as maintenance of the mature thyroid, and murine studies also support a role for FOXE1 in thyroid migration because *Foxe1-/-* mice exhibit either athyreosis or ectopy. HHEX is expressed ubiquitously throughout the foregut endoderm and is required both for thyroid morphogenesis and for maintaining the expression of PAX8 and FOXE1 in the developing murine thyroid.[1,2]

Thyroid hormone biosynthesis: Thyroid hormone biosynthesis is a complex process requiring adequate circulating iodide, appropriate thyroidal iodide handling, and sufficient TG substrate for iodination (Fig. 81.1). The sodium-iodide symporter (NIS, SLC5A5) mediates the uptake of iodide through the basolateral membrane of thyroid follicular cells (thyrocytes) by the electrogenic symport of two sodium ions for one iodide ion, down an electrochemical gradient generated by the Na^+/K^+ ATPase. Iodide efflux into the follicular lumen is facilitated by specific transporter molecules, e.g., pendrin (SLC26A4) and anoctamin-1, and iodide is oxidized in the presence of hydrogen peroxide (H_2O_2) and incorporated into tyrosyl residues on the surface of TG to form mono- and diiodotyrosyl (MIT and DIT), which are coupled to produce THs (thyroxine [T4] and triiodothyronine [T3]). TPO catalyzes the H_2O_2-dependent oxidation of I^- into I^+, the iodination of tyrosyl residues, and the coupling reactions, and H_2O_2 is predominantly synthesized by DUOX2 (an NADPH-oxidase enzyme) and its accessory protein, DUOXA2. TG-bound T3 and T4 are endocytosed back into the thyroid follicular cell then cleaved and secreted into the circulation; transporters including monocarboxylate transporter 8

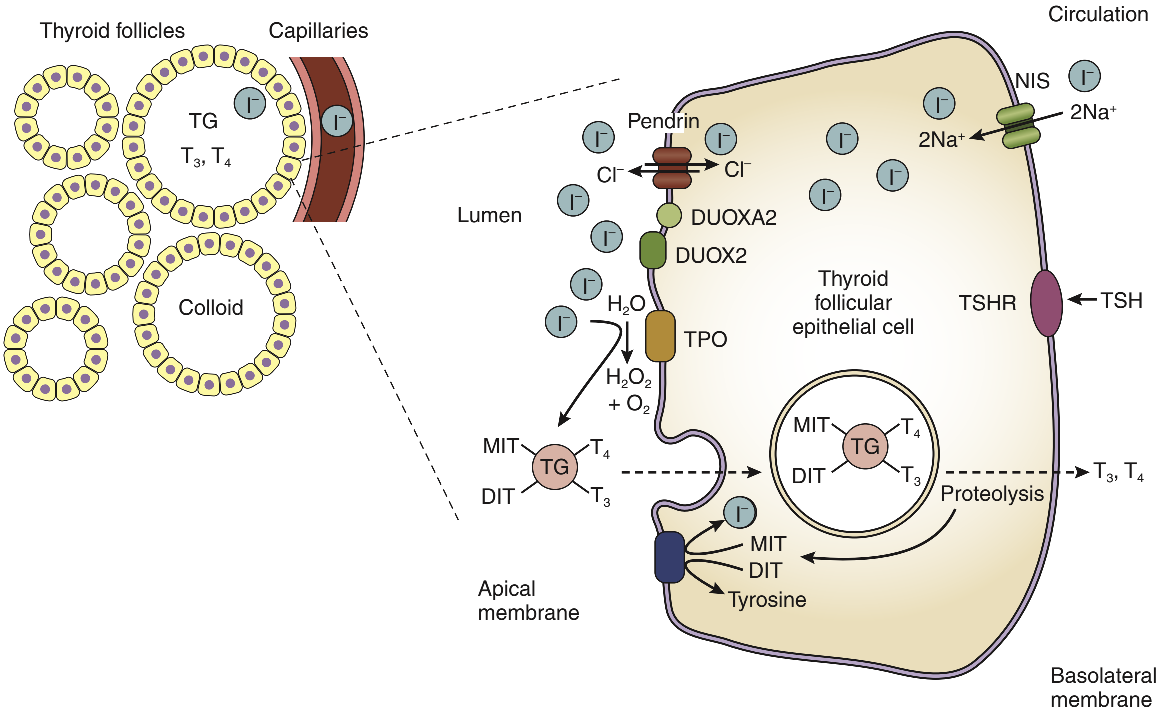


Fig. 81.1 Diagrammatic representation of a thyroid follicular epithelial cell and components of the thyroid hormone biosynthesis machinery in which mutations result in dyshormonogenic congenital hypothyroidism. (Created with Biorender.com)

(MCT8) mediate TH efflux from the thyroid gland, and iodotyrosine deiodinase (IYD) recycles unused iodide moieties.[3]

Hormone action: It is now recognized that entry of THs (T4 and T3) into cells is not a passive process. MCT8, a membrane protein, has been shown to mediate cellular TH transport, particularly in the central nervous system (CNS). Intracellularly, deiodinase enzymes (DIOs) mediate hormone metabolism, with a high-affinity type 2 enzyme (DIO2) mediating T4 to T3 conversion in the CNS, including the pituitary gland and hypothalamus, type I deiodinase (DIO1) in peripheral tissues generating T3, and type 3 deiodinase (DIO3) mediating catabolism of THs to inactive metabolites. The effects of THs on physiologic processes are mediated principally by a receptor protein, the thyroid hormone receptor (TR), belonging to the steroid/nuclear receptor superfamily of ligand-inducible transcription factors, which modulates target gene expression in different tissues (Fig. 81.2; also see Chapter 67). TR binds preferentially to regulatory DNA sequences (thyroid response elements [TREs]) in target gene promoters as a heterodimer with the retinoid X receptor (RXR), although the receptor can bind some TREs as a homodimer or monomer. In the absence of hormone, unliganded receptor homodimers/heterodimers recruit corepressors (e.g., nuclear receptor corepressor [NCoR] and silencing mediator for retinoic acid and thyroid receptors [SMRT]) to repress or "silence" gene transcription (Fig 81.3). Hormone binding results in corepressor dissociation and relief of repression, together with ligand-dependent transcriptional activation, mediated by a complex of coactivators (e.g., steroid receptor coactivator 1 [SRC-1], CREB-binding protein [CBP], and CBP-associated factor [pCAF]).[4] In humans, two highly homologous TH receptors, TRα and TRβ, are encoded by genes (*THRA*, *THRB*) on chromosomes 17 and 3, respectively. Two different proteins are generated from the TRα gene locus by alternate splicing: TRα1 is a ubiquitously expressed receptor isoform, with particular abundance in the CNS, myocardium, skeletal muscle, gastrointestinal tract, and bone; TRα2, which exhibits a modified C-terminal region such that it is unable to bind hormone, is expressed in a variety of tissues (e.g., brain and testis), and its function is not understood. The TRβ gene generates two major receptor isoforms, TRβ1 and TRβ2, which differ in their N-terminal regions: TRβ1, which is widely expressed, is the predominant isoform in liver and kidney, while TRβ2 expression is limited principally to the hypothalamus, pituitary, inner ear, and retina.[5]

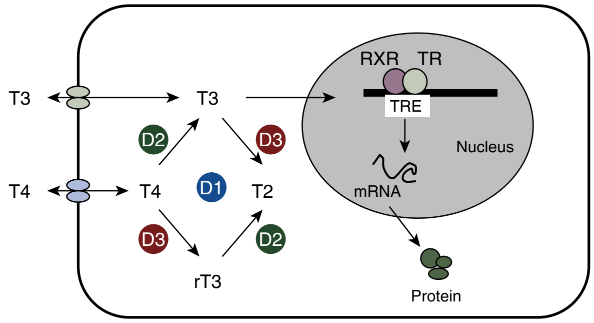


Fig. 81.2 Transport, deiodination, and nuclear action of thyroid hormones. Transporters are required for passage of triiodothyronine (T3) and thyroxine (T4) across the plasma membrane, facilitating hormone uptake, efflux, or both. Deiodinases catalyze conversion of T4 to T3 (DIO1, DIO2) or inactivation of T4 to reverse T3 and T3 to T2 (DIO3). T3 interaction with its nuclear receptor (TR), usually part of a heterodimer with retinoid X receptor, modulates target gene transcription and protein synthesis.

CONGENITAL PRIMARY HYPOTHYROIDISM

Epidemiology and Screening

Primary congenital hypothyroidism (CH) is the most common neonatal endocrine disorder, and is due either to defective thyroid

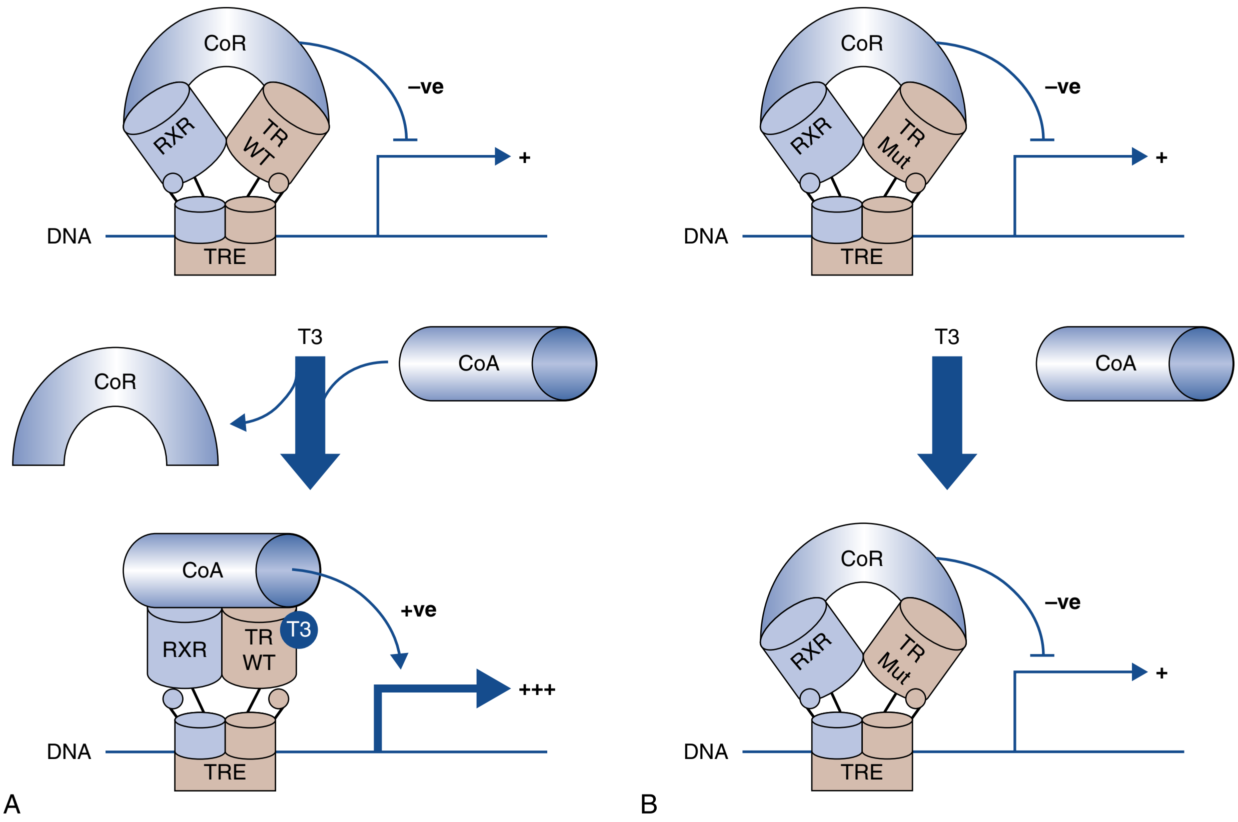


Fig. 81.3 A model for dominant-negative inhibition by mutant receptors in both resistance to thyroid hormone (RTH) β and RTHα. The left panel (**A**) depicts current understanding of wild-type thyroid hormone receptor (TR) action on target genes. The unliganded TR–retinoid X receptor heterodimer or homodimer (not shown) recruits a corepressor complex (CoR) to inhibit or silence basal gene transcription. Receptor occupancy by triiodothyronine promotes corepressor dissociation and derepression, followed by binding of coactivators (CoA), which leads to target gene activation. The right panel (**B**) shows RTH receptor mutant action. In comparison to wild-type TR, the primary defect in mutant receptors may be impaired hormone-dependent corepressor dissociation and coactivator recruitment. For the majority of receptor mutants, this functional alteration is a consequence of their reduced ability to bind ligand. However, a subset of mutants exhibit enhanced corepressor binding or delayed corepressor release or impaired coactivator recruitment *per se*, with relative preservation of hormone binding. Mutant receptor–CoR complexes compete with their wild-type counterparts for occupancy of promoter thyroid response elements (TREs), resulting in inhibition of target gene expression. In this model, DNA binding, dimerization, and corepressor interaction are functional properties that are preserved in mutant receptors and required for their dominant-negative activity.

gland development or impaired TH biosynthesis. Delayed diagnosis and treatment of CH may be devastating for childhood neurodevelopment and growth; therefore, most industrialized countries operate neonatal CH screening programs, facilitating detection of CH before signs of hypothyroidism develop. Screening strategies and biochemical diagnostic thresholds vary, worldwide and within individual nations, thereby influencing the local epidemiology of CH.

Historically, most screening programs initially measured T4 levels, with concomitant or subsequent quantitation of TSH if T4 levels were subnormal. However, increased sensitivity of TSH assays prompted a change in practice to primary TSH measurement, which improved the sensitivity of detection of primary CH. Most programs now use primary TSH-based strategies, with the remainder measuring T4 either in combination with TSH, or with secondary measurement of TSH if T4 is subnormal. In the Netherlands, a more complex algorithm includes stepwise measurement of T4 followed by thyroxine-binding globulin and TSH, enabling indirect quantitation of free T4 (FT4). There is a general consensus that detection of primary CH (for which TSH is the most sensitive measure) should be the priority for neonatal screening; however, central CH will only be detected if T4 is included in the primary screening method. Since the introduction of primary TSH-based screening strategies, TSH screening cut points have decreased, due both to improved sensitivity of TSH assays and the refinement of screening strategies to minimize false-negative results. Additionally, many screening programs now assess a second specimen in populations likely to evade detection on the initial screening test. Infants who may not generate an adequate TSH response in the first few weeks of life include those born prematurely, with very low birth weight, or with significant intercurrent neonatal illness.[6]

Primary CH has traditionally been subclassified as either thyroid dysgenesis (TD) or dyshormonogenesis (DH). TD refers to a defective thyroid gland development, most commonly thyroid ectopy, where the thyroid is situated in an ectopic site (usually sublingually) and is usually small. Twenty to thirty percent of TD cases are due to athyreosis (complete absence of the thyroid gland), and a small minority exhibit a normally located but hypoplastic thyroid. In contrast, individuals with DH exhibit inadequate TH production despite a normally located, often goitrous, thyroid gland, due to disruption of the molecular pathway for TH biosynthesis.[7]

Prior to the advent of CH newborn screening, the incidence of primary CH was estimated at 1:7000. Studies performed shortly after the introduction of newborn screening reported the actual incidence to be 1:3000 to 1:4000, likely due to the detection of individuals with delayed or more subtle clinical features of hypothyroidism. These early studies attributed approximately 75% to 85% of CH cases to TD and the remainder to DH. More recent studies have reported a doubling of the incidence of CH to approximately 1:2000. The incidence of TD has generally remained stable, and these findings are partly due to increased ascertainment of cases with a normally-located gland *in situ* (GIS CH), and milder CH, as a result of lower screening TSH diagnostic cutoffs. The proportion of cases with transient CH (neonatal hypothyroidism that resolves on retesting around 3 years of age, enabling cessation of levothyroxine treatment) has also increased. However, altered TSH cut points do not explain similar increases in incidence of CH in regions where diagnostic thresholds have remained constant, or where severe and mild GIS CH have increased.[7,8] CH incidence varies with ethnicity; therefore, altered population ethnicity may drive changes in occurrence of CH, e.g., in New Zealand, where a rise in incidence of DH was explained by an increase in percentage of Asian and Pacific Islander births.[9] The relative contributions of environmental factors such as iodine deficiency, as well as altered population demographics, including increases in preterm births, have not been ascertained.[7]

KEY POINTS

- The majority of congenital hypothyroidism (CH) is detected on neonatal screening; therefore, local screening strategies and diagnostic thresholds influence detection rate. The incidence of CH is increasing, especially that of mild or transient gland *in situ* CH; decreased thyroid-stimulating hormone screening cut points are a major contributing factor, but do not provide the sole explanation, which may be multifactorial.

Genetic Causes of Thyroid Dysgenesis

The traditional subdivision of CH into TD and DH has facilitated the identification of genetic causes of CH, with genetic mutations in TD generally involving genes with a role in thyroid development, whereas DH-associated mutations involve known components of the TH biosynthesis machinery. TD is generally considered to be a sporadic disease, with less than 5% cases attributable to a mutation in a known gene involved in thyroid differentiation, migration, and growth. In contrast, DH is usually a familial disease, with an identifiable genetic basis.

Mutations in *NKX2-1, PAX8,* and *FOXE1* are well-recognized causes of TD. Although combined expression of these transcription factors defines developing thyroid follicular cells, individually they exhibit unique extrathyroidal tissue expression patterns; therefore, mutations may be associated with characteristic, multisystem developmental defects. Mutations in GLIS3, which belongs to the GLI-similar 1-3 (GLIS1-3) subfamily of Krüppel-like zinc finger protein transcription factors, are also associated with CH. Studies in murine models have suggested that GLIS3 acts downstream of TSH and the TSHR and is essential for TSH/TSHR-mediated proliferation of thyroid follicular cells, expression of NIS and pendrin, and biosynthesis of TH. Zebrafish models have suggested that Glis3 plays a role in specification, determining the number of endodermal cells that commit to a thyroid fate. Besides transcription factor mutations, loss-of-function mutations in *TSHR* may cause TD. TSHR is a G protein-coupled receptor to which TSH binds, stimulating both thyroid growth and TH biosynthesis.[7,10,11] Genetic causes of TD are summarized in Table 81.1.

NKX2-1: The transcription factor most commonly mutated in CH is that encoded by *NKX2-1*, in which more than 100 disease-causing mutations have been reported (in the Human Gene Mutation Database [HGMD] Professional 2021.1). NKX2-1 contains two transcriptional activation domains, one on either side of the homeodomain, and mutations are detected throughout the gene, but missense mutations almost exclusively localized to the homeodomain. Haploinsufficiency is usually assumed to account for the associated phenotype, but a minority of variants exhibit dominant negative effects, which may be promoter- or tissue-specific.[12] NKX2-1 is also implicated in development of the lungs and pulmonary surfactant production, as well as the development of the ventral forebrain, in particular the basal ganglia and the hypothalamus. *NKX2-1* mutations may therefore result in cerebral and pulmonary dysfunction in addition to CH, with around half of those affected exhibiting a complete triad, so-called "brain-lung-thyroid" syndrome. More than 90% of individuals harboring *NKX2-1* mutations exhibit neurological features, typically a benign hereditary chorea manifesting in childhood that may be preceded by hypotonia, ataxia, and developmental delay. Approximately 70% exhibit hypothyroidism, which is usually mild or subclinical with normal thyroid morphology (GIS CH), although gland hypoplasia, hemiagenesis, or athyreosis are also described. Pulmonary compromise is the least common phenotype, affecting approximately 50% cases, but carries an associated mortality of up to 16% and may include severe surfactant deficiency at birth and infant respiratory distress syndrome, with recurrent infections and interstitial lung disease after the neonatal period; pulmonary involvement carries an associated mortality of up to 16%.[12,13] Rarely, pituitary abnormalities or hypothalamic dysfunction have also been reported.[14] There is poor genotype-phenotype correlation and variable expressivity and penetrance of *NKX2-1* mutations.

Key clinical considerations include the observation that *NKX2-1* mutations frequently occur *de novo*; therefore, confirmatory investigations may reassure parents that the risk to future offspring is minimal. Additionally, deletions both involving *NKX2-1* and proximal to *NKX2-1* may cause brain-lung-thyroid syndrome; therefore, exclusion of copy number variants is essential in this context.[12,13]

PAX8: PAX8 mutations are a well-recognized but uncommon cause of CH, with 43 heterozygous mutations defined as "disease-causing," as reported in HGMD. Mutations most commonly comprise substitutions affecting the DNA binding domain, although truncating mutations affecting the C-terminus and a promoter region mutation have also been reported. *PAX8* haploinsufficiency can cause CH, although some mutations exhibit dominant negative activity. Penetrance and expressivity are variable; however, truncating mutations may be associated with milder phenotypes.[15,16] Around 50% of patients harboring pathogenic *PAX8* mutations exhibit thyroid hypoplasia; however, almost a third have a normal sized, normally-located GIS CH, and a minority exhibit thyroid ectopy or athyreosis.[7] Most (but not all) affected patients exhibit CH; however, thyroid dysfunction may also be transient or subclinical and manifest in later childhood or adulthood. PAX8 is expressed in the nephrogenic mesenchyme, and although most mutations cause isolated CH, a spectrum of associated urogenital tract abnormalities have been reported, including disrupted Mullerian duct development in Mayer–Rokitansky–Kuster–Hauser syndrome.[17,18]

FOXE1: FOXE1 is a forkhead domain transcription factor that is expressed in oropharynx, esophagus, choanae, and hair follicles as well as in the developing thyroid. In contrast to *PAX8* and *NKX2-1* mutations, *FOXE1* mutations are biallelic and extremely rare, with only five definitively pathogenic missense mutations reported, all within the forkhead DNA binding domain. A highly penetrant triad of developmental abnormalities reflecting the tissue expression pattern of FOXE1 is associated with these mutations: athyreosis or severe thyroid

TABLE 81.1 A Summary of the Gene Defects and Associated Clinical Features Identified in Thyroid Dysgenesis with a Genetic Basis

Gene	M/B	Extrathyroidal Features	Biochemical Severity of Congenital Hypothyroidism	Thyroid Morphology
NKX2-1	M, often *de novo*	Neurological (BHC, developmental delay, hypotonia), respiratory (IRDS, recurrent infections)	Usually mild/subclinical ranges from euthyroid to severe	Normal GIS (majority), hypoplasia/hemiagenesis, athyreosis reported
PAX8	M	Urogenital tract malformations (rare)	Euthyroid/subclinical-moderate-severe	Typically hypoplasia (~50%), GIS (~30%), athyreosis and ectopy also reported
FOXE1	B	Cleft palate, spiky hair (universal), choanal atresia, bifid epiglottis	Severe CH	Athyreosis
GLIS3	B	Permanent neonatal diabetes, facial dysmorphia renal cystic dysplasia, hepatic fibrosis/cholestasis, glaucoma, learning difficulties, skeletal abnormalities	Severe CH, may have TSH and TG elevation despite normal FT4 on L-T4	GIS (architecture may be abnormal), athyreosis
CDCA8	M/B	Usually isolated thyroid disease, CHD in one case	Euthyroid-severe CH	Ectopy/athyreosis, in CH cases, nodules/asymmetry in euthyroid cases
JAG1	M	Alagille syndrome/ CHD, may be isolated CH	Euthyroid/ mild hypothyroidism-severe CH	Usually orthotopic hypoplasia, GIS and ectopy also reported
NTN1	M	CHD, arthogryposis	Severe CH	Ectopy
TUBB1	M/B	Macroplatelets, platelet hyperaggregation in some cases	Euthyroid to severe CH	Mainly ectopy, but hypoplasia, hemiagenesis, mild asymmetry reported in euthyroid/mild CH cases
TSHR	M/B	Nil	Compensated raised TSH-severe CH, may be difficult to normalize TSH without elevated FT4 levels	Spectrum from GIS to severe hypoplasia

M, Monoallelic; *B*, bilallelic; *GIS*, thyroid gland *in situ*; *BHC*, benign hereditary chorea; *IRDS*, infant respiratory distress syndrome; *CHD*, congenital heart disease; *L-T4*, levothyroxine; *TSH*, thyroid-stimulating hormone; *FT4*, free thyroxine; *CH*, congenital hypothyroidism; *TG*, thyroglobulin.
Only one case has been reported with *NTN1* mutations, with additional variants: 47, XYY karyotype; and atypical 22q11 deletion.

hypoplasia (one individual harboring *FOXE1* p.R102C had apparent eutopic thyroid tissue on ultrasound, but ^{131}I uptake and serum TG levels were negligible, consistent with lack of function), cleft palate, and spiky hair (Bamforth–Lazarus syndrome), with choanal atresia and bifid epiglottis also reported in some patients. *FOXE1* mutations usually impair both DNA binding and transcriptional activity, either completely (p.R102C, p.A65V, and p.F137S) or partially (p.S57N), although a gain-of-function mutation, p.R73S, has also been described, with a clinically indistinguishable phenotype.[19,20,21] Additionally, FOXE1 contains a long polyalanine tract, the length of which may influence risk of TD, with larger studies generally showing that the longer tract length (16A) may be protective for TD development.[20]

GLIS3: In recent years, biallelic *GLIS3* mutations have become a well-established, albeit very rare, cause of CH. GLIS3 contains a DNA binding domain comprising tandem repeats of five zinc fingers and a transactivation domain at the C-terminus. It has pleiotropic extrathyroidal roles, including regulation of pancreatic β cell generation, maturation, and insulin gene expression, spermatogenesis, and maintenance of normal kidney function.[10] Biallelic *GLIS3* mutations result in a variably penetrant multisystem phenotype that consistently includes permanent neonatal diabetes and CH. Additional developmental defects may include renal cystic dysplasia, congenital glaucoma, hepatic cholestasis, progressive liver fibrosis, osteopenia, and minor facial dysmorphisms. Reported mutations include exonic or larger deletions, truncating mutations resulting in loss of the GLIS3 activation domain, and missense mutations within the DNA binding domain or zinc finger motifs. Thyroid morphology in infants harboring *GLIS3* mutations ranges from apparently normal to thyroid hypoplasia or athyreosis. However, one ultrasonographically "normal" gland showed a paucity of colloid microscopically and extensive fibrosis. In some cases, TSH and TG levels remain elevated during treatment with conventional doses of levothyroxine therapy despite normalization of free T4.[22]

TSHR: TSHR mutations may be mono- or biallelic and result in a spectrum of TSH resistance, the severity of which depends both on the deleteriousness of the mutation and the number of mutated *TSHR* alleles. Complete TSH resistance (usually due to biallelic, nonfunctional *TSHR* alleles) results in orthotopic gland hypoplasia and severe CH. There may appear to be athyreosis on imaging; however, because early thyroid gland development and TG synthesis are TSH-independent, athyreosis in this context is only ever apparent, as evidenced by measurable serum TG confirming the presence of thyroid tissue. At the milder end of the spectrum (usually due to monoallelic, hypomorphic *TSHR* alleles), TSH resistance is partial, and the thyroid may be a normal size, with isolated hyperthyrotropinaemia compensating adequately for the defective TSHR, resulting in preserved TH biosynthesis and normal circulating TH levels.[7,23]

Inactivating *TSHR* mutations, especially those conferring partial TSH resistance, may be moderately common, although the incidence is influenced by the ethnicity of the population screened, with founder mutations reported, e.g., *TSHR* p.R450H in individuals from East Asia. *TSHR* mutations resulting in partial TSH resistance have been identified in up to 29% of pediatric Italian patients with nonautoimmune hyperthyrotropinaemia. There is no single "hotspot" for mutations, which may be distributed throughout the coding sequence for the receptor.[7,23]

TSHR predominantly couples to Gs and the cyclic AMP (cAMP) cascade in thyroid but signals via Gq and the phospholipase C and inositol phosphates/calcium pathway at higher TSH levels. Gq signaling is thought to stimulate thyrocyte proliferation, H_2O_2 synthesis, and apical iodide efflux, without promoting thyrocyte iodide uptake via NIS, whereas Gs/cAMP signalling may promote all of these processes. *TSHR* mutations usually result in impaired Gs/cAMP signal transduction, often due to defective membrane expression; however, mutations have been reported (e.g., p.L653V, p.R450H) that preferentially impact Gq signaling, resulting in increased radioiodine uptake due to compensatory, enhanced Gs/cAMP signaling.[24,25]

Patients with complete TSH resistance and severe CH unequivocally require levothyroxine replacement. However, whether or not individuals with partial TSH resistance and normal circulating TH concentrations require treatment remains controversial, because elevated TSH levels in these individuals may compensate for the TSHR defect. Pituitary thyrotrophs and peripheral tissues exhibit normal sensitivity to exogenous T3 administration, and the mechanisms mediating ongoing elevated TSH secretion despite normal circulating TH levels are unclear, probably reflecting a reset TSH axis setpoint in affected individuals.[26]

Some individuals may require excessive levothyroxine doses to normalize TSH levels due to such "resetting" of the hypothalamic-pituitary-thyroid axis, and the consequent supraphysiological FT4 concentrations may cause thyrotoxic symptoms. There are reports of normal growth, development, and pituitary size in untreated cases despite chronically raised TSH, and studies from Israel and Italy have investigated the need for levothyroxine treatment in this context. Cases with heterozygous *TSHR* mutations without additional risk factors for hypothyroidism generally exhibited stable TH biochemistry (isolated hyperthyrotropinaemia) during follow-up periods of up to 15 years. In the Italian study, growth and neuropsychological development were normal, irrespective of whether levothyroxine treatment was given, supporting the notion that this is a compensated state that may not require TH replacement. However, in both studies, thyroid dysfunction did progress in individuals with biallelic *TSHR* mutations, or with heterozygous mutations and additional risk factors for thyroid disease (e.g., positive autoantibodies). Therefore, although there is growing consensus that levothyroxine replacement therapy is often not indicated in partial TSH resistance, individualized treatment decisions are required, and levothyroxine should be considered in some settings.[27,28]

Additional Genes Associated with Thyroid Dysgenesis. *NKX2-5*: NKX2-5 belongs to the NK-2 family of homeodomain-containing transcription factors and plays a role in murine thyroid development, with *Nkx2-5* null mice exhibiting thyroid bud hypoplasia. This observation, together with identification of variably penetrant, heterozygous, loss-of-function *NKX2-5* mutations in patients with thyroid ectopy or athyreosis, seemed to implicate *NKX2-5* mutations in TD. However, subsequent studies have detected pathogenic mutations at higher than expected frequencies in healthy populations (p.R25C, minor allele frequency ≥1%) or in individuals with isolated congenital heart disease, mandating that, although *NKX2-5* variants may contribute to TD risk, coexisting factors/genes must also play a significant role in the pathogenesis of TD, modulating their penetrance and expressivity.[29]

JAG1: Jagged1 (JAG1) is a Notch receptor ligand expressed in thyroid that plays a role in thyroid development in zebrafish. In humans, variably penetrant, heterozygous, loss-of-function *JAG1* mutations are associated with Alagille syndrome, a rare developmental disorder with multisystem involvement. When thyroid function was assessed in a series of patients with Alagille syndrome, up to 28% exhibited mild, nonautoimmune hypothyroidism. Additionally, in a cohort of CH cases without associated Alagille syndrome, 4% harbored heterozygous loss-of-function *JAG1* mutations, substantiating the notion that JAG1 may contribute to the pathogenesis of CH, usually with orthotopic gland hypoplasia.[30]

CDCA8 (BOREALIN): CDCA8, a member of the chromosomal passenger complex with roles in the processes of chromosome segregation and cytokinesis, is a recently-identified gene associated with TD. Mono- and biallelic *CDCA8* loss-of-function mutations with variable expressivity were reported in three unrelated families in association with CH with thyroid ectopy or athyreosis, or variable thyroid structural abnormalities (asymmetry, nodules, hemiagenesis) with euthyroid biochemistry. One patient also had papillary thyroid cancer. Expression of mutant *CDCA8* in a thyroid cell line affected cellular migration and adhesion.[31]

TUBB1: TUBB1 (tubulin, beta 1 class VI) encodes a b-tubulin protein family member that heterodimerizes with a-tubulin to form cytoskeletal microtubules. *Tubb1* null mice are hypothyroid, with defective embryonic thyrocyte proliferation and migration, disrupted thyroid follicular architecture, and impaired TH secretion. Human *TUBB1* mutations have previously been associated with congenital macrothrombocytopenia; however, homozygous missense and heterozygous truncating *TUBB1* mutations were recently reported in three unrelated kindreds in association with CH, predominantly due to thyroid ectopy. These patients did not exhibit thyrombocytopenia; however, platelet size was increased, and some individuals demonstrated platelet hyperaggregation.[32]

Netrin 1 (NTN1): NTN1 is part of a family of laminin-related secreted proteins and acts as an axon guidance molecule during neural development, with additional roles in cell migration and survival and possibly in the development of pharyngeal vessels. Netrin 1-deficient zebrafish embryos exhibit defective aortic arch artery formation and abnormal thyroid morphogenesis, likely due to lack of guidance cues from dysplastic vasculature. A single patient has been reported in whom an NTN1 defect is associated with a congenital heart defect and TD (ectopy), although the contribution of *NTN1* mutations to thyroid and cardiac congenital defects in the wider population remains unclear.[33]

KEY POINTS

- Transcription factor (TF) mutations are a rare monogenic cause of thyroid dysgenesis, but may be associated with characteristic syndromic features reflecting the extrathyroidal expression of the TF involved. Genetic ascertainment in such cases may guide multidisciplinary management and permit accurate genetic counseling. Mutations in the thyroid-stimulating hormone (TSH) receptor (TSHR) may be associated with a spectrum of TSH resistance ranging from euthyroid hyperthyrotropinaemia with gland *in situ* congenital hypothyroidism (CH) to severe CH with apparent athyreosis. Isolated hyperthyrotropinaemia associated with heterozygous TSHR mutations may represent a compensated state that does not require treatment.

Genetic Causes of Dyshormonogenesis

In contrast to TD, when CH occurs due to DH, an underlying monogenic basis can be identified in the majority of cases, involving genes encoding any of the key components of the thyrocyte TH biosynthesis machinery. Genes implicated in dyshormonogenic CH include *TG, TPO, SLC26A4* (pendrin), *SLC5A5* (NIS), *DUOX2, DUOXA2, IYD,* and *SLC26A7.* Each genetic defect is associated with key biochemical and radiological hallmarks, including differences in circulating TG levels, thyroidal iodide uptake, and perchlorate discharge test results, as summarized in Table 81.2. Perchlorate discharge testing enables the discharge of accumulated intrathyroidal radioiodine to be quantified following administration of perchlorate, and provides an indirect measure of efficiency iodide organification, with greater than 90% radioiodine retained within the thyroid if organification is intact.

TABLE 81.2 A Summary of the Gene Defects Mediating Congenital Hypothyroidism with Dyshormonogenesis and Typical Associated Clinical Features

Gene	M/B	Additional Clinical Features	Congenital Hypthyroidism	Radiological Features
TG	B	Inappropriately low TG	Spectrum from euthyroidism to severe CH	GIS/goiter, iodide organification usually normal
TPO	B*		Usually severe CH, although milder phenotypes may occur, especially with monoallelic variants	GIS/goiter TIOD, rarely PIOD
DUOX2	M/B	Delayed TSH rise	Mild-moderate permanent, or transient CH	GIS/goiter PIOD
DUOXA2	M/B		Mild permanent or transient CH	GIS/goiter PIOD
Pendrin, SLC26A4	M	Sensorineural hearing loss with EVA, Mondini cochlear defect in some cases	Euthyroid if iodine replete/mild hypothyroidism	GIS/goiter PIOD
NIS, SLC5A5	M	Delayed TSH rise, delayed presentation	Spectrum from euthyroidism to severe CH	GIS/goiter, severely impaired thyroid ^{123}I/Tc uptake
IYD	M/B	Raised urinary MIT and DIT, delayed TSH rise, delayed presentation	Spectrum from euthyroidism to severe CH	Goiter, normal organification of iodide
SLC26A7	B	Delayed TSH rise, delayed presentation	Moderate-severe permanent CH	Goiter/GIS, PIOD

*Occasional monoallelic cases reported.

M, Monoallelic; *B*, bilallelic; *GIS*, thyroid gland *in situ*; *EVA*, enlarged vestibular aqueduct; *PIOD*, partial iodide organification defect; *TIOD*, total iodide organification defect; *MIT*, monoiodotyrosine; *DIT*, diiodotyrosine; *Tc*, pertechnetate; *CH*, congenital hypothyroidism; *TSH*, thyroid-stimulating hormone; *TG*, thyroglobulin.

However, genetic subtypes of DH are increasingly recognized to show a more variable and broader phenotypes than initially appreciated. Goiter may be associated with all DH-associated genetic defects, but is not universal. The mechanisms underlying variable goitrogenesis in affected individuals are poorly understood, but may include iodine status; early levothyroxine treatment may also mitigate compensatory thyroid enlargement. Rarely, fetal goiter may complicate DH, and has been reported in association with *TG, TPO, NIS,* and *DUOXA2* mutations. Circulating TG levels are elevated at diagnosis in all forms of DH except that associated with *TG* mutations.[7]

Thyroglobuin: TG is the protein on which TH biosynthesis and storage occurs through the iodination of specific tyrosine residues. CH due to *TG* mutations is a relatively common cause of DH, with an estimated frequency of at least 1:100,000 births. *TG* mutations are usually biallelic and associated thyroid dysfunction can range from severe CH to euthyroidism, with frequent occurrence of goiter. The biochemical hallmark of CH due to *TG* mutations comprises an inappropriately low serum TG level despite elevated TSH concentrations and/or goiter, or failure of exogenous TSH to stimulate a rise in serum TG. Thyroidal iodide uptake is enhanced, and iodide organification is usually preserved. T3 levels may be paradoxically normal or even mildly elevated, with low or low-normal T4, resulting in a high T3/T4 ratio; this pattern may reflect increased thyroidal type 2 iodothyronine DIO activity.[3,7] There are more than 220 CH-associated *TG* mutations, which predominantly result in TG misfolding and retention in the endoplasmic reticulum (ER), triggering ER stress.[34] Recent studies suggest that extrusion of ER-stressed, dying thyrocytes into the follicular lumen in individuals with TG mutations may release TG substrate for TH synthesis, explaining the residual TH production observed in some affected individuals.[35] Recent resolution of the full length three-dimensional structure of TG by cryo–electron microscopy, which may yield further insights into the pathogenicity of CH-associated mutations.[36]

TPO: TPO catalyzes the H_2O_2-dependent iodination of tyrosyl residues on TG to form MIT and DIT, and their subsequent coupling to form T3 and T4. *TPO* mutations frequently underlie DH in European and Pakistani cases but may occur less commonly in East Asian individuals.[37-39] Mutations are usually biallelic, and associated CH is frequently severe, often with a goitrous thyroid gland. Thyroidal iodide uptake is usually enhanced, and *TPO* mutations represent the commonest cause of total iodide organification defect (TIOD, release of >90% accumulated intrathyroidal radiodine during a perchlorate discharge test).[37,40] Heterozygous *TPO* mutations and mutations with residual catalytic function have rarely been associated with mild hypothyroidism and partial iodide organification defects (PIOD), sometimes presenting after the neonatal period. In isolated cases, heterozygous *TPO* mutations have been associated with TIOD, perhaps due to monoallelic intrathyroidal expression of TPO.[41,42] More than 100 TPO mutations have been reported that may disrupt folding and membrane insertion and/or catalytic activity, although few have undergone molecular characterization.

DUOX2 and DUOXA2: DUOX2 is a thyroidal NADPH oxidase that generates the H_2O_2 required by TPO as the final electron acceptor for both iodination of TG and the coupling reaction during TH biosynthesis. Its accessory protein DUOXA2 facilitates the delivery of DUOX2 from the ER to the plasma membrane and subsequently forms a stable complex with DUOX2 that is required for its enzymatic activity. *DUOX2* mutations occur frequently, especially in East Asian individuals,[39] and the allele frequency of some DUOX2 mutations approaches 1% in certain populations (e.g., p.Q570L in South Asians) suggesting that *DUOX2* mutations may contribute to CH even more frequently than currently appreciated. *DUOXA2* mutations are a rare cause of DH, with less than 20 mono- or biallelic loss-of-function variants reported, although some populations harbor recurrent mutations.[7,34,39,43]

Both biallelic and monoallelic *DUOX2* mutations cause CH, which is usually transient or mild permanent CH. In transient CH, the H_2O_2 synthesis defect cannot be fully compensated in the neonate, where TH biosynthesis is at its peak, resulting in neonatal hypothyroidism. However, when TH requirements decline in later childhood, other H_2O_2-synthesizing enzymes may compensate fully for the defect, resulting in resolution of hypothyroidism and enabling cessation of levothyroxine treatment. Thyroidal iodide uptake is usually normal or increased with *DUOX2* mutations, and patients exhibit a PIOD.[43,44]

Clinical data are sparse for individuals with *DUOXA2* mutations, but reported cases also demonstrate both mild permanent and transient CH with or without goiter, and PIOD.[39,43,44]

DUOX2 mutations exhibit highly variable penetrance and expressivity, and even biallelic truncating mutations, as well as heterozygous mutations, may be associated with both transient and permanent CH. It is speculated that variants in other H_2O_2-synthesizing enzymes may modulate CH severity, e.g., *DUOX1*, which encodes an additional NADPH-oxidase, is contiguous with *DUOX2* on chromosome 15, and is expressed at lower levels than DUOX2 in thyroid.[43] In support of this hypothesis, cases with likely complete DUOX isoenzyme deficiency exhibited unusually severe CH consistent with a compensatory role for DUOX1.[45] A series of UK cases harboring *DUOX2 and DUOXA2* mutations exhibited borderline neonatal bloodspot screening TSH levels followed by a delayed but significant TSH rise with subnormal venous FT4 measurements, suggesting that clinically actionable CH associated with *DUOX2 and DUOXA2* mutations could be missed on neonatal screening.[44] DUOX2 is also expressed in respiratory epithelium and throughout the gastrointestinal tract, and rare loss-of-function mutations increase risk of inflammatory bowel disease.[46]

IYD (previously known as DEHAL1): IYD is a DIO, catalyzing the deiodination of MIT and DIT residues, thereby releasing thyroidal iodide and tyrosine for further hormonogenesis. *IYD* mutations have only been formally characterized in a few families, although the description of goiter with impaired iodide recycling preceded the first molecular diagnosis by many years. Monoallelic and biallelic *IYD* mutations can both cause goitrous hypothyroidism although heterozygotes may be asymptomatic. Affected individuals exhibit rapid thyroidal uptake of iodide with a normal perchlorate discharge test and characteristically increased urinary concentrations of MIT and DIT. Hypothyroidism may manifest in childhood following a normal neonatal CH screening test result, resulting in neurodevelopmental delay if diagnosis is delayed.[40]

SLC5A5 (NIS): NIS mediates the basolateral uptake of iodide into of thyrocytes by the electrogenic symport of two sodium ions for one iodide ion, down an electrochemical gradient generated by the Na^+/K^+ ATPase. Biallelic *NIS* mutations are a rare cause of DH, and thyroid dysfunction ranges from severe CH to euthyroidism, especially if dietary iodine content is high. NIS transports both iodide and pertechnetate; therefore, mutations cause diminished or absent thyroidal radioiodine or pertechnetate uptake despite clinical or ultrasonographic evidence of a normally located, usually enlarged, thyroid gland. Tracer uptake by other NIS-expressing tissues, e.g., gastric parietal cells and salivary glands, is also reduced, and following injection of radioiodine, the ratio of salivary:plasma radioiodine is decreased in affected patients.[40] NIS-mediated CH may not manifest at the time of neonatal screening and may present with delayed TSH rise and hypothyroidism in childhood, potentially resulting in neurodevelopmental delay.[47] Because NIS mediates iodide accumulation in breast milk, prophylactic iodine supplementation of affected females may be required to prevent neonatal iodide deficiency in exclusively breastfed offspring.[48]

NIS mutations may result in loss of function due to impaired transporter function despite normal plasma membrane expression, or alternatively, mutant NIS proteins with either preserved or abrogated symporter activity may be retained intracellularly.[34,47]

SLC26A4 (*pendrin*): SLC26A4 (pendrin) is a multifunctional anion transporter that mediates both chloride-bicarbonate and chloride-iodide exchange. Biallelic mutations in pendrin cause Pendred syndrome, comprising congenital sensorineural hearing impairment with enlargement of the vestibular aqueduct, and diffuse or multinodular goiter with PIOD. These phenotypes reflect a role for pendrin in chloride-bicarbonate transport in the inner ear and in maintaining endolymphatic fluid acid-base homeostasis, as well as mediating iodine efflux at the apical membrane of the thyrocyte. A subset of cases also harbor a Mondini cochlear malformation (incomplete partition of the cochlea). Only approximately 50% patients with Pendred syndrome exhibit subclinical or overt biochemical hypothyroidism, and thyroid dysfunction or goiter are rare before puberty.[49]

SLC26A7: Biallelic loss of function mutations in *SLC26A7* represent the most recently discovered genetic cause of DH, with 19 cases now reported worldwide. Of those cases evaluated, 12 exhibit goiter, and all have permanent CH, although one case from Japan presented with hypothyroidism in childhood following a normal newborn screening TSH measurement. SLC26A7 belongs to the same family of anion transporters as pendrin and has previously been shown to function as a chloride-bicarbonate exchanger or chloride channel, mediating murine renal bicarbonate resorption and gastric acid secretion. The SlC26A7 transcript is highly expressed in human and murine thyroid, and all mutations thus far are truncating mutations, which disrupt the highly-conserved STAS domain and C-terminus of the translated protein, affecting plasma membrane localization if the truncated protein is expressed.

Individuals who have undergone further evaluation exhibit normal thyroidal uptake of radioiodine with PIOD. *Slc26a7* null mice also exhibit goitrous hypothyroidism with defective hormone biosynthesis. However, in contrast to affected humans, thyroidal radioiodine uptake is reduced in null mice with normal perchlorate discharge; additionally, null mice exhibit distal renal tubular acidosis, which was not recapitulated in affected humans.[50,51]

The role of SLC26A7 in thyroid and its location in thyroid follicular cells have not yet been conclusively ascertained. *In vitro* studies have yielded varying results, with one study reporting a role for SLC26A7 in iodide efflux, whereas another study did not observe this effect.[50,51] Future studies in thyrocytes are required to address its localization and role in thyroidal iodide metabolism, including its putative role as a direct iodide transporter. Because SLC26A7 is thought to regulate acid-base balance in the renal tubule and gastric epithelium, it has also been hypothesized that it may have an analogous function in thyroid follicular cells, where optimal ionic milieu and pH are critical for enzyme (TPO, DUOX2) and anion exchanger activities. Species differences in thyroid cell electrophysiology may then mediate divergent effects of SLC26A7 deficiency on iodide metabolism in humans and mice.

KEY POINTS

- Most individuals with dyshormonogenesis (DH) will have a mutation affecting genes involved in thyroid hormone biosynthesis, all of which cause isolated DH, except for Pendred syndrome. There is a broad spectrum of congenital hypothyroidism (CH) biochemical severity and goitrogenesis, but specific genetic mutations may be associated with specific biochemical or radiological hallmarks. Genetic testing may help predict transiency of CH or potential for delayed thyroid-stimulating hormone rise and inform genetic counseling regarding recurrence risk in future offspring.

Genetic syndromes that may include CH: CH, or hyperthyrotropinaemia, may also be a feature of multisystem involvement in other genetic syndromes, including pseudohypoparathyroidism and related disorders where TSHR signaling is impaired, without direct disruption to the *TSHR* gene, e.g., due to *GNAS* defects or monoallelic defects in *PRKAR1A* or *PDE4D*. Depending on the molecular defect, individuals may exhibit characteristic physical features (e.g., short stature, skeletal defects, obesity) ,and endocrine abnormalities including resistance to TSH and parathyroid hormone. CH or nonautoimmune hypothyroidism may also be associated with di George syndrome, for which *TBX1* is a candidate gene. Neonates and children with trisomy 21 (Down syndrome) often exhibit nonautoimmune hyperthyrotropinaemia or mild primary CH, for which the underlying mechanism is poorly understood but may involve *DYRK1A*. Say–Barber–Biesecker–Young–Simpson syndrome (*KAT6B*) and Johanson–Blizzard syndrome (*UBR1*) may also include CH, although the underlying mechanism has not been fully elucidated.[7]

Recent advances in identifying genetic determinants of CH: Historically, despite the low yield of genetic mutations identified in TD, evaluation of CH cohorts has supported a genetic basis in a larger subset of CH than is currently recognized. A large national survey of CH in France found that 2% of individuals with TD have an affected relative, which is 15-fold greater than predicted by chance alone. Moreover, thyroidal developmental abnormalities were more common in euthyroid first-degree relatives of CH cases than in controls. TD has been found to occur more frequently in consanguineous or less genetically diverse populations, and individuals with CH have an increased risk of extrathyroidal developmental malformations.[52-54]

Conversely, the greater than 90% discordance between monozygotic twins with CH, the female preponderance of thyroid ectopy, and the low frequency of causative mutations in TD seem incompatible with simple Mendelian inheritance and have been used to support the argument that TD is a sporadic disease, for which the underlying basis may include somatic mutations restricted to the thyroid or epigenetic events.[55,56] Attempts to reconcile these observations with data supporting a role for underlying genetic factors include the "two hit" hypothesis, in which a predisposing germline mutation occurs in conjunction with an additional genetic or epigenetic alteration within the thyroid gland or surrounding structures.[57] However, the only study to investigate this showed that, although ectopic thyroid exhibited differing gene expression compared with normally sited tissue, this was not due to differences in somatic gene methylation patterns. Additionally, whole-exome sequencing in lymphocyte DNA from monozygotic twins discordant for TD did not demonstrate frequent somatic mutations.[58,59] Some thyroidal genes do appear to exhibit autosomal monoallelic expression; however, there is only one confirmed case of monoallelic expression of a mutant allele (*TPO*) in association with dyshormonogenic CH.[41,60]

More recently, data from next-generation sequencing (NGS) studies have facilitated the identification of monogenic causes for CH in a greater proportion of cases. Targeted NGS of genes previously known to be implicated in TD or DH, in a hypothesis-free manner, has identified likely pathogenic variants in at least 50% individuals with GIS CH, while demonstrating significant overlap of genetic etiologies in GIS CH and TD morphological subgroups.[61] Mutations in genes characteristically associated with TD (e.g., biallelic *FOXE1* mutations) have been reported in association with GIS CH without extrathyroidal features. Conversely, *pendrin* mutations have been reported in TD, perhaps due to secondary atrophy of the thyroid due to increased oxidative stress, and *TPO* mutations have been reported in association with thyroid hypoplasia.[62,63] N-terminal DUOX2 mutations have also been reported in individuals with thyroid ectopy, raising the possibility of a role for DUOX2 and H_2O_2 production in thyroid development.[64] Crossover between classical TD-associated and DH-associated genes in the pathogenesis of both morphological subtypes of CH may partly explain the previously low mutation detection rate in phenotype-driven targeted genetic studies. Additionally, there has been ongoing discovery of novel causative genes for CH in human studies (e.g., *SLC26A7, TUBB1, CDCA8*).

A further major insight from NGS studies has been the confirmation of the role of oligogenicity in CH. In particular, NGS evaluation of 11 CH-associated genes in 177 Italian patients with GIS CH and TD identified a likely pathogenic variant in more than one gene in 25% of cases, across both morphological subgroups, that cosegregated with CH or nonautoimmune hypothyroidism in several pedigrees.[61] Similar NGS studies have implicated oligogenicity in the pathogenesis of CH across a range of ethnicities. This suggests that CH may occur due to the additive effect of multiple rare variants influencing both thyroid morphogenesis or hormonogenesis, which exert only modest effects on thyroid function when expressed in isolation. Such oligogenic inheritance may go some way towards explaining both the apparently sporadic occurrence of CH and the variable expressivity and penetrance of causal mutations.[65]

Clinical significance of genetics in congenital hypothyroidism: Molecular genetic testing, where available, may be helpful in cases of GIS CH where DH is suspected or in CH where there is a strong family history or other system involvement, and is recommended in these contexts.[6] A genetic diagnosis may assist genetic counseling and enable tailored treatment where multisystem involvement is anticipated, or where levothyroxine treatment may not be required lifelong (e.g., where hypothyroidism is likely to be transient, such as with *DUOX2/DUOXA2* mutations, and with heterozygous TSHR mutations where treatment may not be required). Additionally, for etiologies known to be associated with a delayed TSH rise, such that hypothyroidism may present in childhood despite a normal neonatal screening TSH (*NIS, DUOX2, DUOXA2, SLC26A7,* or *IYD* mutations), establishing the genetic etiology may enable prompt diagnosis in affected siblings to prevent neurodevelopmental delay. An NGS-based approach to genetic testing in CH is most likely to yield positive results.

Future studies: The advent of NGS has improved genetic diagnosis in CH, both by identifying known and novel monogenic causes in a nonhypothesis driven manner and by identifying oligogenic causes for CH. However, conclusions from NGS studies have been limited by lack of functional characterization of the rare, likely pathogenic variants identified, and future studies are required both to substantiate the functional consequences of these variants and to investigate the mechanisms by which they result in CH. Whole-exome or genome sequencing has the potential to identify further novel genetic causes of CH and to delineate novel genes involved in thyroid development or TH biosynthesis. The recent success in differentiating murine and human pluripotent stem cell-derived endodermal precursors into functional thyroid follicular structures, as well as technological advances in thyrocyte culture, present opportunities for validating the roles of such genes in the future. Additionally, the contribution of environmental factors, e.g., iodine and micronutrient status and the role of endocrine disruptors in the pathogenesis of CH, have yet to be established.

KEY POINTS

- Although monogenic causes for thyroid dysgenesis (TD) remain rare, next-generation sequencing has played a significant role in identifying genetic causes for congenital hypothyroidism (CH) by i.) highlighting a role for oligogenicity in both TD and gland *in situ* CH; ii.) confirming overlap of genetic causes across different morphological subgroups of CH; iii.) identifying novel causative genes for CH; and iv.) identifying mutations in known CH-associated genes in individuals with atypical phenotypes. Future studies are required to substantiate the functional consequences of novel variants predicted to be pathogenic *in silico* and to investigate the mechanism by which novel genes and oligogenic variants cause CH.

CLASSIFICATION OF DISORDERS OF THYROID HORMONE ACTION

Resistance to TH (RTH), first described in 1967 and characterized by elevated circulating THs (T4, T3) and nonsuppressed TSH levels

TABLE 81.3 Genetic Disorders Associated with Elevated Thyroid Hormones

Disorder	Dystranstheyretinaemic Hyperthyroxinemia*	Familial Dysalbuminaemic Hyperthyroxinaemia (FDH)	Resistance to Thyroid Hormone beta (RTHbeta)	Resistance to Thyroid Hormone alpha (RTHalpha)	Allan–Herndon–Dudley Syndrome	SBP2 Deficiency**
Gene	*TTR*	*ALB*	*THRB*	*THRA*	*MCT8*	*SBP2*
Free T4	Raised	Raised	Raised	Low-normal or low	Normal or low	Raised
Free T3	Normal or raised	Normal or raised	Raised	High-normal or raised	Raised	Normal or low
TSH	Normal	Normal	Normal (or raised)	Normal (or raised)	Normal (or raised)	Normal (or raised)
Reverse T3	Raised	Raised	Raised	Low or low-normal	Low	Raised
SHBG	Normal	Normal	Normal	Normal or raised	Raised	Raised

*Elevated thyroid hormone measurements can be intermittent.
**Low circulating selenium levels are also characteristic of this disorder.
SHBG, sex hormone-binding globulin; *T3*, triiodothyronine; *T4*, thyroxine; *TSH*, thyroid-stimulating hormone.

together with variable tissue refractoriness to hormone action, is mediated by defects in TRβ.[66] However, it is now recognized that other disorders, due to defective transport of TH into cells, intracellular metabolism of T4 to generate T3 (see Chapter 66), or hormone action via TRα, are also characterized by altered tissue sensitivity to TH. Accordingly, a consensus statement now defines RTH more broadly, to encompass all defects that can interfere with the biological action of TH secreted in normal or excessive amounts.[67]

DIFFERENTIAL DIAGNOSIS OF ELEVATED THYROXINE AND/OR TRIIODOTHYRONINE WITH NONSUPPRESSED THYROID-STIMULATING HORMONE

A number of inherited and acquired disorders are associated with differing patterns of elevated TH (T4 and/or T3) together with nonsuppressed TSH levels (see Table 9.1 in Chapter 9). The first step in making a diagnosis is to verify the validity of hormone measurements: confirmation of elevated free TH levels in two-step or equilibrium dialysis assays excludes abnormal circulating binding proteins or interfering antibodies; preservation of linearity when TSH is assayed in dilution suggests that this measurement is not artefactual. Many causes (nonthyroidal illness, psychiatric disorder, neonatal period, drugs) can be excluded by clinical context.

Genetic disorders associated with elevated TH levels and reduced sensitivity to hormone action can also be distinguished on the basis of different thyroid function test and metabolite patterns (Table 81.3).

RESISTANCE TO THYROID HORMONE β

Clinical Features

Resistance to TH due to defective TRβ (RTHβ) was first described in two siblings who were clinically euthyroid despite high circulating TH levels and who exhibited several other abnormalities, including deaf-mutism, stippled femoral epiphyses with delayed bone maturation and short stature, dysmorphic facies, winging of the scapulae, and pectus carinatum.[66] It is now clear that some of these features are unique to this kindred, in which the disorder was recessively inherited. The majority of RTHβ cases that have been subsequently described are dominantly inherited, with a highly variable clinical phenotype. Affected subjects are either asymptomatic or have nonspecific symptoms and may be noted to have a goiter, prompting thyroid function tests that suggest the diagnosis. In these individuals, classified as exhibiting generalized RTH (GRTH), the high TH levels are thought to compensate for ubiquitous tissue resistance, resulting in a euthyroid state. In contrast, a smaller number of individuals (≈15%) who share the same biochemical phenotype exhibit clinical features of thyrotoxicosis. In adults these can include weight loss, tremor, palpitations, insomnia, and heat intolerance; in children, failure to thrive, abnormal growth (either accelerated, or more commonly retarded), and hyperkinetic behavior have also been noted. When this clinical entity was first described, patients were thought to exhibit "selective" pituitary RTH (PRTH) action with preservation of normal hormonal responses in peripheral tissues.[68] However, it is now recognized that peripheral resistance (typically hepatic) to hormone action is present even in these subjects. Less commonly, hypothyroid features such as growth retardation, delayed dentition and bone age in children, or asthenia and hypercholesterolemia in adults have been observed in RTHβ and may even coexist with thyrotoxic symptoms in the same individual. Taken together, these observations suggest that the clinical features of RTHβ are variable, being governed by either resistance or retention of sensitivity to high circulating levels of free THs in different tissues.

The estimated prevalence of RTHβ is 1 in 40,000 to 50,000 live births, and the disorder can be diagnosed neonatally by screening with a combination of TSH and free T4 measurements.[69,70] Over 900 cases of RTHβ (from more than 400 families) have now been described worldwide, enabling clinical characteristics of this disorder to be defined more precisely.

Goiter. A palpable goiter has been documented in 65% of individuals, particularly adult females. The enlargement is usually diffuse, but benign multinodular change is recognized, particularly in goiters that recur following partial thyroidectomy. Development of toxic multinodular goiter on the background of RTHβ has been documented in a single case.[71] Interestingly, it has been noted that fewer children with RTHβ born to affected mothers exhibit thyroid enlargement (35%) than offspring of unaffected mothers (87%), suggesting that fetal exposure to maternal hyperthyroxinemia may protect against goiter formation.[72] The bioactivity of circulating TSH has been shown to be significantly enhanced in RTHβ, perhaps accounting for the goiter and markedly elevated serum THs observed in many cases, despite normal immunoreactive TSH levels.[73] Rarely, RTHβ cases with coexistent thyroid cancer (usually small foci of papillary carcinoma, but one metastatic) have been described[74,75]; following thyroid ablation, despite inability to fully suppress TSH levels with T4 therapy due to hormone resistance, clinical outcomes are not unfavorable.[76]

Cardiovascular system. Palpitations and resting tachycardia have been reported in approximately 75% of cases with GRTH and almost all cases of PRTH, and atrial fibrillation has been reported in 6% of patients with mutations.[72] The incidence of cardiac symptoms is notably higher in RTHβ patients than in unaffected relatives or in the general population, although still less frequent compared with patients with classic hyperthyroidism.[77] In one study, 30% of RTHβ subjects showed echocardiographic features of increased myocardial contractility and impaired diastolic relaxation, with a greater incidence of mitral valve prolapse.[72] In a prospective study of cardiovascular involvement in a large cohort of children and adults with RTHβ, resting heart rate was significantly higher, and some indices of cardiac systolic and diastolic function (e.g., stroke volume, cardiac output, maximal aortic flow velocity) were intermediate between values in normal and hyperthyroid subjects, whereas other parameters (e.g., ejection and shortening fractions of the left ventricle, systolic diameter, and left ventricle wall thickness) were not different, indicating a partially hyperthyroid response of the heart in this disorder.[77] Clinically significant mixed dyslipidaemia (total cholesterol >5 mmol/L and triglyceride >1.7 mmol/L) has been documented in 49% of an RTHβ patient cohort, together with increased hepatic and intramyocellular lipid and systemic insulin resistance, but normal carotid intima media thickness[78]; however, the hypercoagulable state associated with conventional hyperthyroidism is not present.[79]

Musculoskeletal system. Stippled epiphyses and winged scapulae were noted in the original RTHβ kindred, but have not been observed in other cases, such that these features may represent a specific manifestation of the known gene deletion (*THRB*) or an unrelated genetic abnormality in this consanguineous kindred.[66] In contrast, growth retardation and delayed skeletal maturation are more common in childhood RTHβ patients, with height below the fifth percentile in 18% and delayed bone age (>2 standard deviations) in 29%,[72] with no significant differences between GRTH and PRTH cases. However, despite these abnormalities, final adult height is often not affected.[80]

Bearing in mind the known adverse effects of untreated hyperthyroidism on bone mineralization, we conducted a cross-sectional survey of approximately 80 adult subjects with RTHβ and observed a reduction in bone mineral density in the femoral neck (mean Z score –0.71) and lumbar spine (mean Z score –0.73) but with normal markers of bone turnover (Moran & Chatterjee, unpublished observations).

Energy intake and expenditure. We have documented raised resting energy expenditure (REE) (metabolic rate), mediated by uncoupling of oxidative phosphorylation and mitochondrial ATP synthesis, in skeletal muscle of RTHβ patients. Energy intake is also increased, particularly in childhood.[81] Indeed, an imbalance between such hyperphagia and increased total energy expenditure (reflecting raised REE and increased physical activity due to attention deficit hyperactivity disorder [ADHD]; see later) may account for the abnormally low body mass index seen in approximately one third of children.

Central nervous system. Two studies have documented neuropsychologic abnormalities in RTHβ. First, a history of ADHD in childhood was elicited more frequently in patients with RTHβ (75%) compared with their unaffected relatives (15%).[82] A second study showed that both children and adults with RTHβ exhibited problems with language development, manifested by poor reading skills and problems with articulation, including speech delay and stuttering.[83] However, frank mental retardation (intelligence quotient [IQ] <60) is relatively uncommon (3%), although 30% of patients show mild learning disability (IQ < 85), probably due to CNS resistance to TH action.[84] A direct comparison of individuals with ADHD and RTHβ versus ADHD alone indicates an association with lower nonverbal intelligence and academic achievement in the former group.[85] However, two different surveys of unselected children with ADHD failed to detect any cases of RTHβ by biochemical screening, suggesting that this thyroid disorder is unlikely to be a common cause of hyperactivity.[86] Although magnetic resonance imaging shows that anomalies of the sylvian fissure or Heschl's gyri are more frequent in RTHβ, these features do not correlate with ADHD.[87]

Hearing and vision. Significant hearing loss has been documented in 21% of RTHβ cases, similar to the prevalence reported in CH.[72] In the majority, audiometric tests indicated a conductive defect, probably related to the increased incidence of recurrent ear infections in childhood RTH (67% in RTH vs. 28% in normal controls). Abnormal otoacoustic emissions, consistent with cochlear dysfunction, have also been documented in those with hearing deficit,[72,88] and cochlear expression of TRβ has been shown.[89] The single kindred with deafmutism and recessively inherited RTHβ harbored a complete deletion of the TRβ gene.[66] Overtly abnormal color vision has only been reported in the rare human kindred with recessively inherited RTHβ and a complete *TRβ* gene deletion,[66] or a case with compound heterozygous *TRβ* mutations.[90] Impaired color perception, together with reduced cone and rod function, have been documented in heterozygous RTHβ patients.[91] Together with knowledge that TRβ knockout mice are deaf[92] and that TRb2 null mice have loss of M-cone photoreceptors and abnormal color vision,[93] these observations underscore the important role of TRβ in the audiovisual system.

Other associated disorders. Coexistence of RTHβ with autoimmune thyroid disease has been reported[94]; thyroid autoantibodies are more prevalent in RTHβ subjects than in unaffected relatives, suggesting a possible pathogenic link between these disorders.[95] Coexistence of Pendred syndrome[96] or TD[97] and RTHβ has been documented. Pituitary enlargement, as a consequence of impaired negative feedback regulation of TSH secretion, is another potential association with RTHβ. While pituitary hyperplasia has been reported in a single case, it occurred in the context of massively elevated TSH levels with suboptimal T4 replacement therapy following inappropriate thyroid ablation, and regressed once TSH levels normalized.[98] Only a few cases of RTHβ associated with pituitary adenomas have been described,[99,100] suggesting that pituitary hyperplasia or adenoma formation are not common clinical sequelae in RTHβ, provided that the altered set point of the hypothalamic-pituitary-thyroid axis is not perturbed. A greater frequency of recurrent upper respiratory tract and pulmonary infections has been reported in RTHβ, and affected individuals have reduced serum immunoglobulin levels.[72] A retrospective study of a large Azorean kindred has shown a higher rate of miscarriage in mothers with RTHβ, with fewer unaffected offspring than expected being born and exhibiting lower birth weight and transient suppression of TSH levels neonatally, suggesting that intrauterine exposure to high maternal TH levels does have adverse fetal effects.[101]

Differential Diagnosis

Differentiation of RTHβ, particularly the form associated with hyperthyroid features, from a TSH-secreting pituitary tumor can be difficult. Similar abnormalities in thyroid function tests in first-degree relatives strongly suggest RTHβ, together with normal pituitary imaging and serum α-glycoprotein subunit levels. Other factors that are helpful in making this differential diagnosis, such as dynamic testing and clinical/biochemical features, are discussed in greater detail in Chapter 9.

The development of Graves disease in RTHβ is suggested by atypical features such as ophthalmopathy, severe thyrotoxic symptoms, and a further rise in THs leading to a subnormal or suppressed TSH. Following antithyroid drug treatment, TSH levels become elevated despite normalization of THs. Similarly, requirement of T4 replacement in supraphysiologic dosage to normalize TSH in autoimmune hypothyroidism and associated with elevated FT4 and free T3 (FT3) levels suggests coexistent RTHβ.

TABLE 81.4 Tissue Indices of Thyroid Hormone Action

Pituitary:	Thyroid-stimulating hormone (TSH)
General:	Basal metabolic rate (BMR)
Hepatic:	Sex hormone-binding globulin (SHBG), ferritin, cholesterol
Muscle:	Creatine kinase, ankle jerk relaxation time
Bone:	Height, bone age, bone density, osteocalcin, alkaline phosphatase, pyridinium crosslinks, type 1 collagen telopeptide
Cardiac:	Sleeping pulse rate, systolic time interval, diastolic isovolumic relaxation time
Hematologic:	Soluble interleukin-2 receptor (sIL-2R)
Lung:	Angiotensin-converting enzyme (ACE)

In addition to clinical features, the measurement of various tissue markers of TH action has been suggested to be a useful method for evaluating the differing responses of various target organs and tissues to elevated circulating THs (Table 81.3). These measurements are most useful in assessing the tissue effects of marked TH excess (as typically found in overt thyrotoxicosis) but may be less discriminatory in borderline hyperthyroidism or in hypothyroidism. To improve the sensitivity and specificity of these parameters, it has been suggested that individuals with RTHβ are assessed by measuring tissue responses dynamically, following the administration of graded supraphysiologic doses of T3 (50, 100, and 200 µg/day, each given for a period of 3 days), with comparison of any change in indices from baseline values to those observed in normal subjects.[102]

KEY POINTS

- Thyroid hormone receptor (TR) β2 plays a key role in negative feedback regulation within the hypothalamic-pituitary-thyroid axis, as well as the audiovisual system, likely accounting for the biochemical pattern of elevated thyroid hormones, with nonsuppressed thyroid-stimulating hormone, that defines resistance to thyroid hormone β, together with auditory and visual abnormalities. Predominance of TRβ1 in the liver mediates dyslipidemia and normal hepatic markers of thyroid hormone action. Conversely, hyperthyroid features of the disorder are due to actions of high circulating thyroid hormones in tissues expressing normal TRα1.

Molecular Genetics

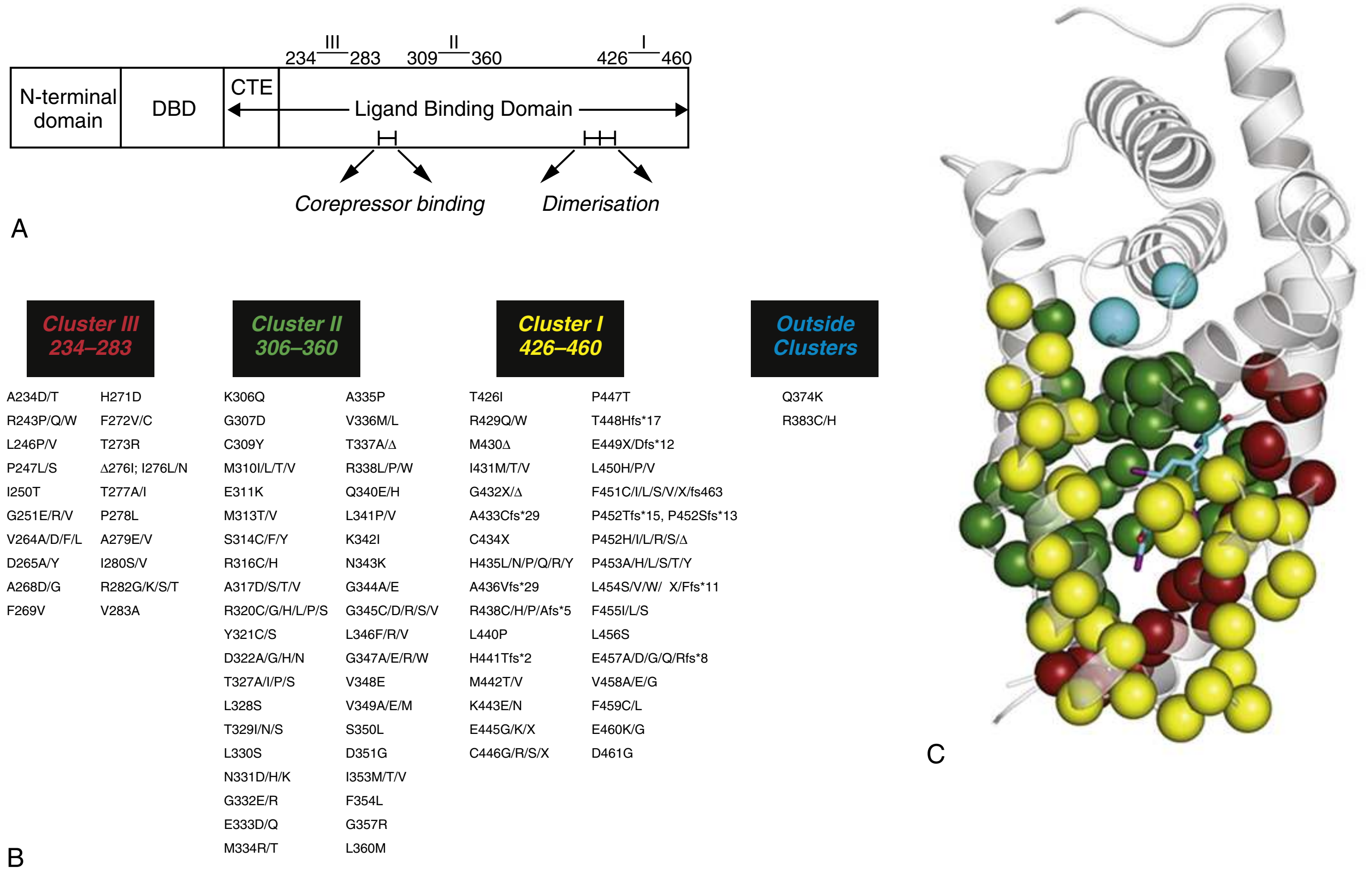
TRβ RTH. Following the cloning of *TRα* and *TRβ*, RTHβ was shown to be tightly linked to the *TRβ* gene locus in a single family.[103] This prompted analysis of the *TRβ* gene in other cases, and a large number of receptor defects have since been associated with this disorder. Eighty percent of RTH is familial, dominantly inherited, and associated with heterozygous mutations in the TRβ gene[84,104-106]; *de novo* receptor mutations occur in the remaining 20% of sporadic cases. Over 218 different defects, including point mutations, in-frame deletions, and frame-shift insertions, have been documented to date, which localize to three mutation clusters within the ligand-binding domain of the receptor (Fig. 81.4). Within each cluster, some codon changes (e.g., R243W, R338W, R438H), representing transitions in mutation-prone CpG dinucleotides, occur more frequently and are overrepresented.[107]

Based on the supposition that PRTH was associated with selective pituitary resistance, it had been hypothesized that this disorder may be associated with defects in DIO2 or the TRβ2 isoform. However, a number of reports have documented *TRβ* mutations in PRTH.[105,108] Receptor mutations found in individuals with PRTH have also been associated with GRTH in unrelated kindreds. Furthermore, even within a single family the same receptor mutation can be associated with abnormal thyroid function and thyrotoxic features consistent with PRTH in some individuals, but similar biochemical abnormalities and a lack of symptoms indicative of GRTH in other members. Overall, these findings indicate that GRTH and PRTH represent the phenotypic spectrum associated with a single genetic entity.

Non-TRβ RTH. In a small but significant number of cases (approximately 10%–15%), clear-cut biochemical evidence of RTHβ is not associated with a mutation in the coding region of *TRβ*—giving rise to the term non-TRβ RTH. However, one explanation for this is somatic mosaicism, with occurrence of a *TRβ* mutation whose expression is selective, being detectable in some tissues but not peripheral blood leukocyte DNA.[109] Alternatively, documentation of sporadic or familial with elevated TH and resistance to T3 administration in whom genetic studies have excluded THRB defects[110,111] raises the possibility that mutations in other factors that are required for TH-dependent gene transcription (e.g., RXR, corepressors, coactivators) could manifest as RTHβ, although no defects in these genes have been identified as yet.

Properties of Mutant Receptors. Consonant with their location in the hormone-binding domain, the majority of receptor mutants identified in RTHβ exhibit moderate or markedly reduced T3 binding; consequently, their ability to activate or repress target gene expression is impaired.[112,113] A subset of RTHβ mutations associated with markedly abnormal thyroid function *in vivo* and altered transcriptional function *in vitro*, but little impairment in ligand binding, have been described. Such natural mutations involve residues that mediate receptor interaction with transcriptional coactivators.[106,114] In the first RTHβ family described,[66] with the recessively inherited form of the disorder, the two affected siblings were found to be homozygous for a complete deletion of both alleles of the *TRβ* gene.[115] Importantly, the obligate heterozygotes in this family, harboring a deletion of one *TRβ* allele, were completely normal, with no evidence of thyroid dysfunction. This suggested that simple deficiency of functional beta receptor, as a consequence of the single deleted *TRβ* allele, was insufficient to mediate the hormone resistance phenotype. This led to the hypothesis that, in dominantly inherited RTHβ, heterozygous mutant receptors were not simply functionally impaired, but also capable of inhibiting the action of wild-type receptor. Studies confirmed that, when coexpressed, the mutant proteins are able to inhibit the function of their wild-type counterparts in a dominant-negative manner.[116,117] Further clinical and genetic evidence supporting this notion has been provided by rare examples of RTHβ. In the first, a childhood case, severe resistance with marked developmental delay and growth retardation associated with cardiac hyperthyroidism was ultimately fatal due to heart failure following septicemia; this individual was homozygous for a mutation (Δ337T) in both alleles of the *TRβ* gene.[118] Several additional cases with a severe clinical phenotype and either homozygous or hemizygous for *TRβ* defects have been described.[119] Presumably, the extreme phenotype observed in such cases reflected not only the absence of normally functioning TRβ, but the added dominant-negative inhibitory effect of mutant β receptors.

Functional studies of mutant receptors indicate that, although they are transcriptionally impaired and dominant-negative inhibitors, their ability to bind DNA and form heterodimers with RXR is preserved.[112,113] Conversely, it has been shown that the introduction of additional artificial mutations that abolish DNA binding or heterodimer formation abrogates the dominant-negative activity of mutant receptors *in vitro*.[113,120,121] Mice heterozygous for a *TRβ* mutation that results in a lack of DNA binding do not exhibit RTH.[122] It has also been suggested that the ability of mutant receptors in RTHβ to repress



Cluster III 234–283		Cluster II 306–360		Cluster I 426–460		Outside Clusters
A234D/T	H271D	K306Q	A335P	T426I	P447T	Q374K
R243P/Q/W	F272V/C	G307D	V336M/L	R429Q/W	T448Hfs*17	R383C/H
L246P/V	T273R	C309Y	T337A/Δ	M430Δ	E449X/Dfs*12	
P247L/S	Δ276I; I276L/N	M310I/L/T/V	R338L/P/W	I431M/T/V	L450H/P/V	
I250T	T277A/I	E311K	Q340E/H	G432X/Δ	F451C/I/L/S/V/X/fs463	
G251E/R/V	P278L	M313T/V	L341P/V	A433Cfs*29	P452Tfs*15, P452Sfs*13	
V264A/D/F/L	A279E/V	S314C/F/Y	K342I	C434X	P452H/I/L/R/S/Δ	
D265A/Y	I280S/V	R316C/H	N343K	H435L/N/P/Q/R/Y	P453A/H/L/S/T/Y	
A268D/G	R282G/K/S/T	A317D/S/T/V	G344A/E	A436Vfs*29	L454S/V/W/ X/Ffs*11	
F269V	V283A	R320C/G/H/L/P/S	G345C/D/R/S/V	R438C/H/P/Afs*5	F455I/L/S	
		Y321C/S	L346F/R/V	L440P	L456S	
		D322A/G/H/N	G347A/E/R/W	H441Tfs*2	E457A/D/G/Q/Rfs*8	
		T327A/I/P/S	V348E	M442T/V	V458A/E/G	
		L328S	V349A/E/M	K443E/N	F459C/L	
		T329I/N/S	S350L	E445G/K/X	E460K/G	
		L330S	D351G	C446G/R/S/X	D461G	
		N331D/H/K	I353M/T/V			
		G332E/R	F354L			
		E333D/Q	G357R			
		M334R/T	L360M			

Fig. 81.4 **A,** A schematic representation of the domains of TRβ showing that, with two exceptions (Q374K, R383C/H), resistance to thyroid hormone (RTH) β receptor mutations localize to three clusters within the ligand-binding domain (LBD). The receptor defects described hitherto include different missense substitutions at each codon, in-frame codon deletions (Δ), premature termination codons (X), and frame-shift mutations (*). The mutations shown include those listed in a public database (Human Gene Mutation Database) together with our unpublished data. No RTH receptor mutations have been described in the zinc-finger DNA-binding domain (DBD) or its carboxy-terminal extension (CTE), which together mediate interaction with DNA, or regions in the LBD that are important for corepressor binding or dimerization with retinoid X receptor. **B,** The crystal structure of the thyroid hormone receptor β LBD (Protein Data Bank accession no. 1BSX) composed of 12 α-helices (grey) is shown, with the location of missense mutations associated with RTH (Cluster I yellow, Cluster II green, Cluster III red, R383 and Q374 cyan) superimposed. As predicted from their functional properties, the majority of mutations involve residues which surround the ligand (triiodothyronine cyan) binding cavity. (See Color Plate.)

or "silence" basal gene transcription is likely to be an important factor contributing to their dominant-negative potency. Non-T3-binding receptor mutants exhibit constitutive silencing function, particularly when bound to DNA as homodimers, which cannot be relieved by ligand; conversely, RTHβ mutants with impaired homodimerization properties are weaker dominant-negative inhibitors.[123] With the identification of corepressors, these observations have been extended to show that some RTHβ mutants either bind corepressor more avidly when unliganded or fail to dissociate fully from corepressor upon T3 binding.[124] Furthermore, artificial mutations that abolish corepressor binding abrogate the dominant-negative activity of RTHβ receptor mutants.[124] An unusual RTHβ receptor mutant (R383H) exhibits both delayed T3-dependent corepressor release and impaired hormone-dependent negative transcriptional regulation.[125] Given the pivotal role of negatively regulated target genes in the pathogenesis of RTH, aberrant corepressor recruitment or release may well prove to be the critical receptor abnormality in this disorder.

Together, these observations allow a model to be constructed (Fig. 81.3) in which occupancy of target gene binding sites by mutant receptor–corepressor complexes mediates dominant-negative inhibition by RTH mutants. Mapping of the three clusters of RTHβ receptor mutations identified hitherto onto the crystal structure of the ligand-binding domain of TRβ[126] provides insights into structure-function relationships in TR (Fig. 81.4). As expected from their impaired ligand-binding properties, most TRβ mutations are located around the hormone-binding cavity; in contrast, receptor regions mediating DNA binding, dimerization and corepressor interaction are devoid of naturally occurring mutations, possibly because they lack dominant-negative activity and are therefore biochemically and clinically silent.

This model also has important implications in interpreting the significance of *THRB* variants identified by NGS. Specifically, *in silico* prediction algorithms may not always reliably predict the pathogenicity of variants not previously reported to be associated with RTHβ.[127] The functional significance of new *THRB* variants is best determined by assessing transcriptional function and dominant-negative activity of TRβ mutants *in vitro* or by showing that the mutant genotype cosegregates with abnormal thyroid function in families.[128]

Pathogenesis of Variable Resistance. Genetic and functional evidence suggests that the ability to exert a dominant-negative effect on target genes within the hypothalamic-pituitary-thyroid axis is a fundamental property of RTHβ receptor mutants, generating the abnormal thyroid function characteristic of the disorder. Indeed, studies indicate that, for a subset of RTHβ mutants, there is a correlation between their functional impairment *in vitro* and the degree of hormone resistance *in vivo*, quantified by the magnitude of elevation in serum FT4 or cholesterol in patients.[78,129] On this background, the heterogeneous clinical phenotype may be due to differing degrees of peripheral resistance in different individuals, as well as variable resistance in different tissues within a single subject. A number of factors may contribute to such variable tissue resistance.

One important contributory element may be the differing tissue distributions of receptor isoforms. As the hypothalamus, pituitary, and liver express predominantly TRβ2 and TRβ1 receptors, respectively, and TRα1 is the major species detected in myocardium, mutations in the *TRβ* gene are likely to be associated with pituitary and liver resistance, as exemplified by nonsuppressed TSH levels and normal sex hormone-binding globulin seen in patients, whereas the tachycardia and cardiac hyperthyroidism in some cases may represent retention of myocardial sensitivity to high circulating THs acting via a normal alpha receptor. Another factor that may influence the degree of tissue resistance is the relative expression of mutant versus wild-type *TRβ* alleles. While one study has suggested that both alleles are equally expressed,[130] another showed marked differences in the relative levels of wild-type and mutant receptor messenger RNA in skin fibroblasts from two RTHβ cases.[131] In one of these individuals, a temporal variation in expression of the mutant allele in fibroblasts appeared to correlate with the degree of skeletal tissue resistance. The dominant-negative inhibitory potency of mutant receptors has been shown to differ depending on target gene promoter context[113,132] and is a further variable that may influence the degree of resistance. Finally, non-*TRβ*-related factors might influence the phenotype. For example, a deleterious R316H mutation was associated with normal TH levels in some members of one kindred[133] but clearly abnormal thyroid function in an unrelated family,[105] suggesting that other genetic variables can modulate the effect of receptor mutations.

While the absence or presence of overt thyrotoxic features allows patients to be classified as either GRTH or PRTH—a clinical definition that will probably remain useful as a guide to the most appropriate form of treatment—studies indicate that there is some overlap of features between the two forms of the disorder. For example, there are no significant differences in age, sex ratio, frequency of goiter, thyroid function, or clinical features between patients with GRTH or PRTH.[134] Importantly, features such as tachycardia, hyperkinetic behavior, and emotional disturbance have been documented in individuals with GRTH.[134] Conversely, serum sex hormone-binding globulin, a hepatic index of TH action, is almost invariably normal in patients with PRTH, suggesting that hormone resistance is not solely confined to the hypothalamic-pituitary-thyroid axis in this group of patients.[135]

Attempts to correlate the phenotype of RTHβ with the nature of the underlying *TRβ* mutation have been confounded by imprecision of criteria used to classify cases as GRTH and PRTH, spontaneous temporal variation in hyperthyroid features in some RTHβ patients,[134] and the relatively small number of patients with any given mutation that have been identified hitherto. Nevertheless, some interesting correlations have emerged from the published literature. The first patient reported to have PRTH[68] was found to harbor an R338W receptor mutation,[108] and the same phenotype has been described in the majority of individuals with this or similar substitutions at this codon.[108] Interestingly, when tested *in vitro*, this mutant exhibits dominant-negative activity with the negatively regulated pituitary TSHα subunit gene promoter but is a relatively poor inhibitor of wild-type receptor action in other promoter contexts.[113] Furthermore, when introduced into other RTH receptor mutant backgrounds, this mutation weakens their dominant-negative potency on positively regulated reporter genes.[136] More recently, it has been suggested that a noncoding polymorphism in the *TRb2* promoter of the mutant allele amplifies pituitary expression of R338W in this patient, further augmenting pituitary dominant-negative activity.[137] A patient with the R383H receptor defect, which is impaired mainly in regulation of thryotropin-releasing hormone and *TSH* genes, exhibited predominant central resistance following T3 administration,[138] and the R429Q mutation, with similar functional properties, may also occur more frequently in association with PRTH. Some receptor mutants (R338W or L, V349M, R429Q, I431T) associated with PRTH are either more deleterious[139] or exert a greater dominant-negative inhibitory effect in a TRβ2 than TRβ1 context.[140] A receptor mutation that selectively fails to bind NCoR but not SMRT is associated with PRTH.[141]

KEY POINTS

- The disorder is usually mediated by heterozygous, loss-of-function thyroid hormone receptor (TR) β mutants that inhibit the function of their wild-type receptor counterparts in a dominant-negative manner. Homozygous, but not heterozygous, *TRβ* haploinsufficency causes resistance to thyroid hormone (RTH) β; a severe phenotype is associated compound heterozygosity or homozygosity for dominant-negative *TRβ* mutations. The genetic basis of 10%–15% of RTH cases that are devoid of THRB mutations remains to be elucidated.
- Factors contributing to variable hormone resistance include differing dominant-negative activity of THRB mutants in different target gene, TRβ2 vs. TRβ1 isoform, or tissue contexts.

Management

The management of RTHβ is difficult, because variable resistance makes it difficult to maintain euthyroidism in all tissues. However, the majority of patients do not require specific treatment, and in general the presence or absence of hyperthyroid features is a useful guide to the need for therapy. In most individuals, the receptor defect is compensated by high circulating TH levels, leading to a euthyroid state, which is not associated with abnormalities other than a small goiter. Attempts to treat the biochemical abnormality with surgery or radioiodine are usually unsuccessful, with recrudescence of the goiter (often nodular in nature) and disruption of the thyroid axis.[84] Certain circumstances, such as hypercholesterolemia in adults or developmental delay and growth retardation in young children, may warrant the administration of supraphysiologic doses of levothyroxine to overcome a higher degree of resistance in certain tissues. Although successful in some cases,[84] such therapy needs careful monitoring of indices of TH action (e.g., sex hormone-binding globulin, heart rate, basal metabolic rate, bone markers) to avoid the adverse cardiac effects or excess catabolism associated with T4 overtreatment. Inappropriate thyroid ablation also renders the RTHβ patient hypothyroid, with elevated TSH levels and risk of thyrotroph hyperplasia,[98] and is another context in which supraphysiologic T4 replacement is indicated. Alternate day administration of liothyronine in supraphysiologic dosage led to significant regression of goiter without inducing thyrotoxic symptoms in one case.[142]

In contrast, a general reduction in TH levels may be of benefit in the management of patients with thyrotoxic symptoms. However, the administration of conventional antithyroid drugs usually causes a further rise in serum TSH levels, which stimulates thyroid enlargement and may also induce thyrotroph hyperplasia, with a theoretical risk of developing autonomous neoplasms at either site. Accordingly, agents that inhibit pituitary TSH secretion, yet are relatively devoid of peripheral thyromimetic effects, are administered to reduce TH levels. The most widely used example is the TH analog 3,5,3'-triiodothyroacetic acid (TRIAC), which has been shown to be beneficial in both childhood and adult cases.[143-146] This compound has a number of interesting properties that make it an attractive therapeutic option in RTHβ: (1) it exerts predominantly pituitary and hepatic thyromimetic effects *in vivo*,[147] target tissues that are relatively refractory to THs in RTHβ; (2) it has greater affinity, potency, and activity than T3 for TRβ[148]; (3) it exhibits a higher affinity for TRβ than TRα *in vitro*[149]; and (4) it can activate some TRβ mutants *in vitro*.[149] A daily dose of 1.4 to 2.8 mg is generally used, and twice daily administration enhances TSH inhibition.[150] TRIAC has been successfully used in pregnancy to control maternal thyrotoxic symptoms, but may have induced fetal goiter.[151] With known spontaneous variation in thyrotoxic symptoms in RTH, periodic cessation of TRIAC therapy and reevaluation of patient clinical status is advisable. Treatment with TRIAC is not always effective,[152] and dextrothyroxine is another agent that has been shown to be useful in some cases.[153,154] However, past experience indicates that TSH secretion escapes from the inhibitory effects of bromocriptine,[143,155] as well as somatostatin analog.[156] Use of antithyroid drugs in combination with TRIAC was beneficial in homozygous RTHβ associated with hyperthyroid cardiomyopathy.[157] Thyroid ablation followed by subphysiologic T4 replacement may also be warranted in cases of RTHβ associated with life-threatening thyrotoxic cardiac failure.

The treatment of thyrotoxic features (e.g., failure to thrive) in childhood RTHβ also requires careful monitoring to ensure that any reduction in TH levels is not associated with growth retardation or adverse neurologic sequelae. Indeed, control of cardiac and sympathomimetic manifestations with β-blockade may be the safest course in this context. Counterintuitively, one study reported that liothyronine therapy improved hyperactivity in nine children with ADHD and RTH, including three individuals who were unresponsive to methylphenidate, and it is conceivable that lower circulating T4 levels may have mediated this.[158] Supporting this notion, TRIAC therapy ameliorated ADHD in one childhood case.[159]

Lipid-lowering therapy may be warranted in patients with hypercholesterolaemia and other cardiac risk factors. Lifestyle changes (e.g., increased physical activity, weight loss) can be recommended in individuals with features of metabolic syndrome (central adiposity, hypertriglyceridaemia, low high-density lipoprotein cholesterol, insulin resistance). Resmetirom, a TRβ-selective agonist that lowers hepatic fat, circulating low-density lipoprotein cholesterol, and triglyceride levels in patients with nonalcoholic steatohepatitis,[160] may represent future rational therapy in patients with mixed dyslipidemia and increased hepatic lipid. Lifestyle measures (weight-bearing exercise, calcium and vitamin D supplementation) may also be advised in patients with reduced bone mineral density; whether bisphosphonate therapy is beneficial in RTHβ is unknown.

On the basis that maternal hyperthyroxinaemia may increase loss of wild-type fetuses,[161] in women with RTHβ and a history of recurrent miscarriage, antithyroid drug treatment from early pregnancy could be considered.[162] During pregnancy, if maternal FT4 does not exceed 50% above the upper limit of normal for gestational age, and fetal growth and heart rate remain normal, this may provide reassurance that significant fetal or neonatal hyperthyroidism is unlikely.[163]

KEY POINTS

- In asymptomatic resistance to thyroid hormone β patients, surveillance for dyslipidemia, metabolic syndrome, and low bone mineral density is recommended, as well as considering lipid-lowering therapy and lifestyle measures (e.g., weight loss, weight-bearing exercise, calcium and vitamin D supplementation) to prevent adverse metabolic and skeletal sequelae.
- β-blockade can ameliorate sympathetic and cardiac hyperthyroid symptoms in childhood and adults. Treatment with 3,5,3'-triiodothyroacetic acid, a centrally acting thyromimetic that lowers circulating thyroid hormone levels, is beneficial in patients with generalized hyperthyroid symptoms.
- Thyroid ablation, followed by subphysiological thyroxine replacement, may be warranted, but only in rare cases with life-threatening, thyrotoxic cardiac failure.

RESISTANCE TO THYROID HORMONE ALPHA

Background

Human TRβ and TRα exhibit marked amino acid sequence similarity, including in their hormone binding domains (80%); accordingly, with approximately 218 different receptor mutations known to be associated with RTHβ, the identification of homologous human TRα defects had been anticipated. Supporting this notion, murine transgenic models harboring different heterozygous TRα mutations are viable and exhibit recognizable abnormalities, but with minimal perturbation of thyroid function.[164,165] Such absence of an overt thyroid biochemical phenotype likely explains why the human disorder (RTHα) eluded discovery for a decade.[166,167]

Clinical Features

At birth, some features (e.g., macroglossia, poor feeding, hoarse cry, umbilical hernia) recognized in hypothyroidism have been noted.[168,169] Two patients were investigated in infancy for growth retardation that predominantly affected the lower segment.[166,167] Abnormal physical characteristics in the majority of cases include macrocephaly, broad facies, a flattened nose, prominent tongue, and thick lips.[166-170] Older children and adults have an excessive number of skin tags and melanocytic naevi in the head and neck area, possibly due to the hyperproliferative state of their normal cutaneous cells.[171]

Biochemical. The most consistent pattern of thyroid function tests comprises low or low-normal FT4 and high or high-normal FT3, resulting in an abnormally low T4/T3 ratio; in addition, reverse T3 levels can be subnormal. In a minority of patients, thyroid function tests (TSH, FT4, FT3) can be completely normal. A mild normocytic anemia, with normal hematinics and hemolytic indices, and raised muscle creatine kinase levels are a consistent abnormality.[166,168,169,172]

Skeletal. Radiographic abnormalities in childhood include delayed fontanelle fusion and excessively serpiginous cranial sutures ("wormian bone" appearance), together with delayed dentition.[166,167] Femoral epiphyseal dysgenesis was present in childhood[166] but not in adult life[168]; bone age can be delayed.[166,167] Cranial and cortical hyperostosis in long bones, together with increased bone mineral density, is present in most cases, especially in adults.

Neurocognitive. In childhood, patients showed delayed milestones (motor, speech). Slow initiation of motor movement, together with fine and gross motor incoordination, manifesting as dyspraxia or a broad-based, ataxic gait and slow, dysarthric speech are a consistent feature. IQ is variably reduced, being markedly subnormal in some cases but with other patients attaining educational qualifications.[168] In severe cases magnetic resonance imaging has documented cortical microencephaly, reduced cerebellar volume, and white matter tract

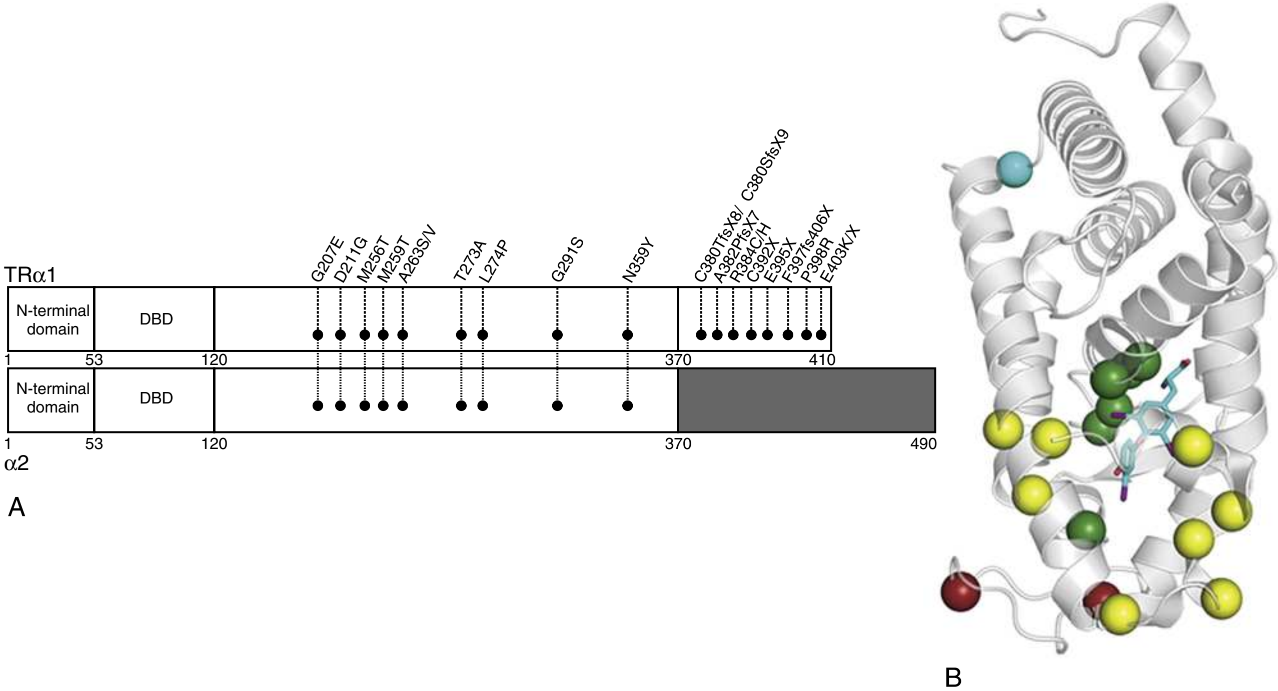


Fig. 81.5 **A,** A schematic representation of the domains of thyroid hormone receptor (TR) α1 and non-hormone-binding α2 proteins. Missense, frameshift, and premature stop mutations that only involve TRα1 or missense mutations that are common to both TRα1 and variant α2 proteins are shown. **B,** The crystal structure of the ligand-binding domain of TRα1 (Protein Data Bank Accession no. 2H77), showing the position of missense, frameshift, and premature stop mutations, color-coded according to the resistance to thyroid hormone b mutation clusters (Fig. 81.3) within which equivalent amino acids in TRβ reside.

density.[173] Genetic studies of a large autism cohort identified a pathogenic *THRA* mutation in one case,[174] and several *THRA* variants whose functional significance is not known have been found in a smaller cohort.[175]

Gastrointestinal. Reduced frequency of bowel movements is a common finding, with severe constipation being a significant problem in two cases.[166,168]

Cardiovascular. Bradycardia[166,168,169] is typical, with abnormal sympathovagal balance and indices of cardiac contractility in the hypothyroid range.[168]

Metabolic and endocrine. REE (metabolic rate) was low in most pati ents.[166,168,169] Both male- and female-to-offspring transmission of TRα defects has been recorded,[167,169] suggesting that fertility is not severely compromised.

KEY POINTS

- Resistance to thyroid hormone α, a disorder characterized by many features of hypothyroidism, including childhood neurodevelopmental delay, growth retardation, and skeletal dysplasia together with constipation and low metabolic rate, is likely to be underdiagnosed because it is associated with near-normal thyroid function tests.

Molecular Genetics

Twenty-one different heterozygous *THRA* mutations (frameshift/premature stop, missense), inherited from either parent or occurring *de novo*, have been documented in 36 cases (Fig. 81.4). Frameshift/premature stop TRa1 mutants fail to bind ligand, are devoid of transcriptional activity, and are associated with a severe clinical phenotype.[166-168] Similar to TRβ mutations in RTHβ, TRα1 mutants inhibit the function of their wild-type receptor counterparts in a dominant-negative manner when they are coexpressed; as has been delineated in RTHβ, constitutive binding of mutant TR to corepressors, with failure of corepressor dissociation and coactivator recruitment following T3 occupancy, likely mediates such dominant-negative inhibition (Fig. 81.3). Expression of TH-responsive target genes in mutation-containing patient peripheral blood mononuclear cells is blunted, suggesting that such dominant-negative inhibition occurs *in vivo.*[166,168]

Many RTHα cases involve *THRA* mutations that affect both TRa1 and TRa2 isoforms.[169,176-178] When studied in the TRα2 protein background, these mutations exhibit no added gain or loss of function, which correlates with absence of any discernible additional clinical phenotype attributable to mutant TRa2, in these patients.[169,176-178] In a single case,[179] highly unusual clinical features (micrognathia, clavicular agenesis, syndactyly) were not seen in a transgenic mouse model[180] and may be unrelated to the *THRA* defect.

Due to the absence of a pathognomonic diagnostic phenotype, highly deleterious receptor defects are likely overrepresented amongst current, disease-associated *THRA* mutations. Consonant with their known homology of TRα and TRβ proteins, most *THRA* mutations that cause RTHα are homologous to known mutations of the equivalent amino acid in *THRB*, causing RTHβ.[181]

Pathogenesis. Many clinical features (e.g., skeletal dysplasia, reduced intestinal motility, bradycardia) in RTHα patients are typical of untreated congenital or adult hypothyroidism. Skeletal abnormalities (growth retardation, delayed tooth eruption, patent cranial sutures, epiphyseal dysgenesis) and intestinal dysmotility in human RTHα are recapitulated in TRα1 mutant mouse models.[164,165] Abnormal thyroid function (low T4/T3 ratio, reduced reverse T3) in RTHα may reflect altered TH metabolism: higher hepatic DIO1 expression in

TRα1 mutant mice[165] may mediate increased T4 to T3 conversion and reverse T3 clearance; alternatively, decreased levels of DIO3 (which is known to be TRα1-regulated) could mediate altered TH metabolism. Differentiation of erythroid progenitors from RTHα patients or transgenic mice is impaired, with dysregulation of Gata-1 being implicated in the animal model.[182,183] Cortical microencephaly seen in adults with severe RTHα correlates with altered proliferation of cortical neural progenitors derived from patient's stem cells.[173] Idiopathic epilepsy seen in human cases[168,169] could be linked to aberrant development of GABAergic inhibitory interneurons in a murine model.[184]

KEY POINTS

- The disorder is mediated by heterozygous, dominant-negative thyroid hormone receptor (TR) α mutants, often homologous to known THRB mutations causing resistance to thyroid hormone β. Mutations involving both TRα1 and TRα2 proteins are not associated with any discernible added phenotype in patients.

Management

T4 treatment in RTHα is beneficial, improving growth (total height and subischial leg length), increasing basal metabolic rate (and thereby limiting weight gain), lowering elevated creatine kinase levels, and enhancing well-being.[168,169,172,178,181] Addition of growth hormone to T4 therapy in childhood does not result in further improvement in growth.[172] In cases harboring mutant TRα1 whose dysfunction is reversible at higher TH levels, treatment from early childhood may have ameliorated their phenotype; in adult life these individuals report improved dyspraxia, constipation, and self-confidence.[169] In virtually all cases, T4 treatment does not improve anemia.

Following T4 treatment in physiological dosage, TSH levels suppress readily, with elevation of FT3 to supraphysiological levels, consonant with preserved TH sensitivity within the hypothalamic-pituitary-thyroid axis; serum sex hormone-binding globulin rises slightly from high-normal baseline levels[168]; conversely, cardiac parameters, REE, and muscle creatine kinase levels are less responsive.[168,169] Overall, these observations are consonant with hormone resistance in organs (e.g., myocardium, skeletal muscle, gastrointestinal tract) expressing predominantly TRα1, with preservation of TH sensitivity in TRβ-expressing tissues (hypothalamus, pituitary, liver).

Increasing use of NGS to identify a genetic basis for phenotypes that are known to be RTHα-associated will likely increase ascertainment of the disorder, as well as identify new clinical manifestations associated with receptor defects. Future identification of specific biomarkers in RTHα have the potential to improve its diagnosis, enabling earlier commencement of T4 therapy to improve outcome.

KEY POINTS

- Thyroxine therapy is beneficial, improving linear growth and limiting excess weight gain in childhood, alleviating constipation and improving well-being.

REFERENCES

1. Nilsson M, Fagman H. Development of the thyroid gland. *Development*. 2017;144:2123–2140.
2. Fernández LP, López-Márquez A, Santisteban P. Thyroid transcription factors in development, differentiation and disease. *Nat Rev Endocrinol*. 2015;11:29–42.
3. Citterio CE, Targovnik HM, Arvan P. The role of thyroglobulin in thyroid hormonogenesis. *Nat Rev Endocrinol*. 2019;15:323–338.
4. Horlein AJ, Heinzel T, Rosenfeld MG. Gene regulation by thyroid hormone receptors. *Curr Opin Endocrinol Diabetes*. 1996;3:412–416.
5. Lazar MA. Thyroid hormone receptors: multiple forms, multiple possibilities. *Endocr Rev*. 1993;14:184–193.
6. van Trotsenburg P, Stoupa A, Léger J, et al. Congenital hypothyroidism: a 2020-2021 consensus guidelines update-an ENDO-European reference Network initiative endorsed by the European Society for pediatric Endocrinology and the European Society for Endocrinology. *Thyroid*. 2021;31:387–419.
7. Peters C, van Trotsenburg ASP, Schoenmakers N. Diagnosis of endocrine disease: congenital hypothyroidism: update and perspectives. *Eur J Endocrinol*. 2018;179:R297–R317.
8. Barry Y, Bonaldi C, Goulet V, et al. Increased incidence of congenital hypothyroidism in France from 1982 to 2012: a nationwide multicenter analysis. *Ann Epidemiol*. 2016;26:100–105.
9. Albert BB, Cutfield WS, Webster D, et al. Etiology of increasing incidence of congenital hypothyroidism in New Zealand from 1993-2010. *J Clin Endocrinol Metab*. 2012;97:3155–3160.
10. Jetten AM. GLIS1-3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. *Cell Mol Life Sci*. 2018;75:3473–3494.
11. Rurale G, Marelli F, Duminuco P, et al. Glis3 as a critical regulator of thyroid primordium specification. *Thyroid*. 2020;30:277–289.
12. Carré A, Szinnai G, Castanet M, et al. Five new TTF1/NKX2.1 mutations in brain-lung-thyroid syndrome: rescue by PAX8 synergism in one case. *Hum Mol Genet*. 2009;18:2266–2276.
13. Thorwarth A, Schnittert-Hübener S, Schrumpf P, et al. Comprehensive genotyping and clinical characterisation reveal 27 novel NKX2-1 mutations and expand the phenotypic spectrum. *J Med Genet*. 2014;51:375–387.
14. Prasad R, Nicholas AK, Schoenmakers N, et al. Haploinsufficiency of NKX2-1 in brain-lung-thyroid syndrome with additional multiple pituitary dysfunction. *Horm Res Paediatr*. 2019;92:340–344.
15. Hermanns P, Grasberger H, Cohen R, et al. Two cases of thyroid dysgenesis caused by different novel PAX8 mutations in the DNA-binding region: in vitro studies reveal different pathogenic mechanisms. *Thyroid*. 2013;23:791–796.
16. Iwahashi-Odano M, Nagasaki K, Fukami M, et al. Congenital hypothyroidism due to truncating PAX8 mutations: a case series and molecular function studies. *J Clin Endocrinol Metab*. 2020;105:dgaa584.
17. Carvalho A, Hermanns P, Rodrigues AL, et al. A new PAX8 mutation causing congenital hypothyroidism in three generations of a family is associated with abnormalities in the urogenital tract. *Thyroid*. 2013;23:1074–1078.
18. Chen N, Zhao S, Jolly A, et al. Perturbations of genes essential for Mullerian duct and Wolffian duct development in Mayer-Rokitansky-Kuster-Hauser syndrome. *Am J Hum Genet*. 2021;108:337–345.
19. Castanet M, Mallya U, Agostini M, et al. Maternal isodisomy for chromosome 9 causing homozygosity for a novel FOXE1 mutation in syndromic congenital hypothyroidism. *J Clin Endocrinol Metab*. 2010;95:4031–4036.
20. Castanet M, Polak M. Spectrum of human Foxe1/TTF2 mutations. *Horm Res Paediatr*. 2010;73:423–429.
21. Carré A, Hamza RT, Kariyawasam D, et al. A novel FOXE1 mutation (R73S) in Bamforth-Lazarus syndrome causing increased thyroidal gene expression. *Thyroid*. 2014;24:649–654.
22. Dimitri P, Habeb AM, Gurbuz F, et al. Expanding the clinical spectrum associated with GLIS3 mutations. *J Clin Endocrinol Metab*. 2015;100:E1362–E1369.
23. Cassio A, Nicoletti A, Rizzello A, et al. Loss-of-function mutations in the thyrotropin receptor gene: when to investigate, clinical effects, and treatment. *J Clin Res Pediatr Endocrinol*. 2013;5(suppl 1):29–39.
24. Narumi S, Nagasaki K, Ishii T, et al. Nonclassic TSH resistance: TSHR mutation carriers with discrepantly high thyroidal iodine uptake. *J Clin Endocrinol Metab*. 2011;96:E1340–E1345.

25. Grasberger H, Van Sande J, Hag-Dahood Mahameed A, et al. A familial thyrotropin (TSH) receptor mutation provides in vivo evidence that the inositol phosphates/Ca2+ cascade mediates TSH action on thyroid hormone synthesis. *J Clin Endocrinol Metab*. 2007;92:2816–2820.
26. Sunthornthepvarakul T, Gottschalk ME, Hayashi Y, et al. Brief report: resistance to thyrotropin caused by mutations in the thyrotropin-receptor gene. *N Engl J Med*. 1995;332:155–160.
27. Vigone MC, Di Frenna M, Guizzardi F, et al. Mild TSH resistance: clinical and hormonal features in childhood and adulthood. *Clin Endocrinol*. 2017;87:587–596.
28. Tenenbaum-Rakover Y, Almashanu S, Hess O, et al. Long-term outcome of loss-of-function mutations in thyrotropin receptor gene. *Thyroid*. 2015;25:292–299.
29. van Engelen K, Mommersteeg MT, Baars MJ, et al. The ambiguous role of NKX2-5 mutations in thyroid dysgenesis. *PLoS One*. 2012;7:e52685.
30. de Filippis T, Marelli F, Nebbia G, et al. JAG1 loss-of-function variations as a novel predisposing event in the pathogenesis of congenital thyroid defects. *J Clin Endocrinol Metab*. 2016;101:861–870.
31. Carré A, Stoupa A, Kariyawasam D, et al. Mutations in BOREALIN cause thyroid dysgenesis. *Hum Mol Genet*. 2017;26:599–610.
32. Stoupa A, Adam F, Kariyawasam D, et al. TUBB1 mutations cause thyroid dysgenesis associated with abnormal platelet physiology. *EMBO Mol Med*. 2018;10:e9569.
33. Opitz R, Hitz MP, Vandernoot I, et al. Functional zebrafish studies based on human genotyping point to netrin-1 as a link between aberrant cardiovascular development and thyroid dysgenesis. *Endocrinology*. 2015;156:377–388.
34. Targovnik HM, Scheps KG, Rivolta CM. Defects in protein folding in congenital hypothyroidism. *Mol Cell Endocrinol*. 2020;501:110638.
35. Zhang X, Kellogg AP, Citterio CE, et al. Thyroid hormone synthesis continues despite biallelic thyroglobulin mutation with cell death. *JCI Insight*. 2021;6(11):148496.
36. Coscia F, Taler-Verčič A, Chang VT, et al. The structure of human thyroglobulin. *Nature*. 2020;578:627–630.
37. Bakker B, Bikker H, Vulsma T, et al. Two decades of screening for congenital hypothyroidism in The Netherlands: TPO gene mutations in total iodide organification defects (an update). *J Clin Endocrinol Metab*. 2000;85:3708–3712.
38. Cangul H, Aycan Z, Olivera-Nappa A, et al. Thyroid dyshormonogenesis is mainly caused by TPO mutations in consanguineous community. *Clin Endocrinol*. 2013;79:275–281.
39. Park KJ, Park HK, Kim YJ, et al. DUOX2 mutations are frequently associated with Congenital Hypothyroidism in the Korean population. *Ann Lab Med*. 2016;36:145–153.
40. Targovnik HM, Citterio CE, Rivolta CM. Iodide handling disorders (NIS, TPO, TG, IYD). *Best Pract Res Clin Endocrinol Metab*. 2017;31:195–212.
41. Fugazzola L, Cerutti N, Mannavola D, et al. Monoallelic expression of mutant thyroid peroxidase allele causing total iodide organification defect. *J Clin Endocrinol Metab*. 2003;88:3264–3271.
42. Zhang J, Han R, Shen L, et al. Mild TPO deficiency characterized by progressive goiter and normal serum TSH level. *Endocrine*. 2020;68:599–606.
43. Muzza M, Fugazzola L. Disorders of H2O2 generation. *Best Pract Res Clin Endocrinol Metab*. 2017;31:225–240.
44. Peters C, Nicholas AK, Schoenmakers E, et al. DUOX2/DUOXA2 mutations frequently cause Congenital Hypothyroidism that evades detection on newborn screening in the United Kingdom. *Thyroid*. 2019;29:790–801.
45. Aycan Z, Cangul H, Muzza M, et al. Digenic DUOX1 and DUOX2 mutations in cases with congenital hypothyroidism. *J Clin Endocrinol Metab*. 2017;102:3085–3090.
46. Grasberger H, Magis AT, Sheng E, et al. DUOX2 variants associate with preclinical disturbances in microbiota-immune homeostasis and increased inflammatory bowel disease risk. *J Clin Invest*. 2021;131:e141676.
47. Szinnai G, Kosugi S, Derrien C, et al. Extending the clinical heterogeneity of iodide transport defect (ITD): a novel mutation R124H of the sodium/iodide symporter gene and review of genotype-phenotype correlations in ITD. *J Clin Endocrinol Metab*. 2006;91:1199–1204.
48. Mizokami T, Fukata S, Hishinuma A, et al. Iodide transport defect and breast milk iodine. *Eur Thyroid J*. 2016;5:145–148.
49. Wémeau JL, Kopp P. Pendred syndrome. *Best Pract Res Clin Endocrinol Metab*. 2017;31:213–224.
50. Cangul H, Liao XH, Schoenmakers E, et al. Homozygous loss-of-function mutations in SLC26A7 cause goitrous congenital hypothyroidism. *JCI Insight*. 2018;3:e99631.
51. Ishii J, Suzuki A, Kimura T, et al. Congenital goitrous hypothyroidism is caused by dysfunction of the iodide transporter SLC26A7. *Commun Biol*. 2019;2:270.
52. Castanet M, Lyonnet S, Bonaïti-Pellié C, et al. Familial forms of thyroid dysgenesis among infants with congenital hypothyroidism. *N Engl J Med*. 2000;343:441–442.
53. Léger J, Marinovic D, Garel C, et al. Thyroid developmental anomalies in first degree relatives of children with congenital hypothyroidism. *J Clin Endocrinol Metab*. 2002;87:575–580.
54. Kreisner E, Neto EC, Gross JL. High prevalence of extrathyroid malformations in a cohort of Brazilian patients with permanent primary congenital hypothyroidism. *Thyroid*. 2005;15:165–169.
55. Perry R, Heinrichs C, Bourdoux P, et al. Discordance of monozygotic twins for thyroid dysgenesis: implications for screening and for molecular pathophysiology. *J Clin Endocrinol Metab*. 2002;87:4072–4077.
56. Devos H, Rodd C, Gagné N, et al. A search for the possible molecular mechanisms of thyroid dysgenesis: sex ratios and associated malformations. *J Clin Endocrinol Metab*. 1999;84:2502–2506.
57. Van Vliet G, Deladoëy J. Sublingual thyroid ectopy: similarities and differences with Kallmann syndrome. *F1000Prime Rep*. 2015;7:20.
58. Abu-Khudir R, Paquette J, Lefort A, et al. Transcriptome, methylome and genomic variations analysis of ectopic thyroid glands. *PLoS One*. 2010;5:e13420.
59. Magne F, Serpa R, Van Vliet G, et al. Somatic mutations are not observed by exome sequencing of lymphocyte DNA from monozygotic twins discordant for congenital hypothyroidism due to thyroid dysgenesis. *Horm Res Paediatr*. 2015;83:79–85.
60. Magne F, Ge B, Larrivée-Vanier S, et al. Demonstration of autosomal monoallelic expression in thyroid tissue assessed by whole-exome and bulk RNA sequencing. *Thyroid*. 2016;26:852–859.
61. de Filippis T, Gelmini G, Paraboschi E, et al. A frequent oligogenic involvement in congenital hypothyroidism. *Hum Mol Genet*. 2017;26:2507–2514.
62. Kühnen P, Turan S, Fröhler S, et al. Identification of PENDRIN (SLC26A4) mutations in patients with congenital hypothyroidism and "apparent" thyroid dysgenesis. *J Clin Endocrinol Metab*. 2014;99:E169–E176.
63. Stoupa A, Chaabane R, Guériouz M, et al. Thyroid hypoplasia in congenital hypothyroidism associated with thyroid peroxidase mutations. *Thyroid*. 2018;28:941–944.
64. Kizys MML, Louzada RA, Mitne-Neto M, et al. DUOX2 mutations are associated with Congenital Hypothyroidism with ectopic thyroid gland. *J Clin Endocrinol Metab*. 2017;102:4060–4071.
65. Stoupa A, Kariyawasam D, Muzza M, et al. New genetics in congenital hypothyroidism. *Endocrine*. 2021;71:696–705.
66. Refetoff S, De Wind LT, De Groot LJ. Familial syndrome combining deaf-mutism, stippled epiphyses, goiter and abnormally high PBI: possible target organ refractoriness to thyroid hormone. *J Clin Endocrinol Metab*. 1967;27:279–294.
67. Refetoff S, Bassett JHD, Beck-Peccoz P, et al. Classification and proposed nomenclature for inherited defects of thyroid hormone action, cell transport, and metabolism. *J Clin Endocrinol Metab*. 2014;99:768–770.
68. Gershengorn MC, Weintraub BD. Thyrotropin-induced hyperthyroidism caused by selective pituitary resistance to thyroid hormone. A new syndrome of inappropriate secretion of TSH. *J Clin Invest*. 1975;56:633–642.
69. Snyder D, Sesser D, Skeels M, et al. Thyroid disorders in newborn infants with elevated screening T4. *Thyroid*. 1997;2:S1–S29.
70. Tajima T, Jo W, Fujikura K, et al. Elevated free thyroxine levels detected by a neonatal screening system. *Pediatr Res*. 2009;66:312–316.
71. Taniyama M, Ishikawa N, Momotani N, et al. Toxic multinodular goiter in a patient with generalized resistance to thyroid hormone who har-

bours the R429Q mutation in the thyroid hormone receptor β gene. *Clin Endocrinol*. 2001;54:121–124.

72. Brucker-Davis F, Skarulis MC, Grace MB, et al. Genetic and clinical features of 42 kindreds with resistance to thyroid hormone. *Ann Intern Med*. 1995;123:572–583.
73. Persani L, Asteria C, Tonacchera M, et al. Evidence for the secretion of thyrotopin with enhanced bioactivity in syndromes of thyroid hormone resistance. *J Clin Endocrinol Metab*. 1994;78:1034–1039.
74. Vinagre J, Borges F, Costa A, et al. Differentiated thyroid cancer in patients with resistance to thyroid hormone syndrome. A novel case and a review of the literature. *Front Mol Biosci*. 2014;1:10.
75. Igata M, Tsuruzoe K, Kawashima J, et al. Coexistence of resistance to thyroid hormone and papillary thyroid carcinoma. *Endocrinol Diabetes Metab Case Rep*. 2016;2016:160003.
76. Unluturk U, Sriphrapradang C, Erdogan MF, et al. Management of differentiated thyroid cancer in the presence of resistance to thyroid hormone and TSH-secreting adenomas: a report of four cases and review of the literature. *J Clin Endocrinol Metab*. 2013;98:2210–2217.
77. Kahaly JG, Matthews CH, Mohr-Kahaly S, et al. Cardiac involvement in thyroid hormone resistance. *J Clin Endocrinol Metab*. 2002;87:204–212.
78. Moran C, McEniery CM, Schoenmakers N, et al. Dyslipidaemia, insulin resistance, ectopic lipid accumulation, and vascular function in resistance to thyroid hormone ß. *J Clin Endocrinol Metab*. 2021;106:e2005–e2014.
79. Elbers LPB, Moran C, Gerdes VEA, et al. The hypercoagulable state in hyperthyroidism is mediated via the thyroid hormone ß receptor pathway. *Eur J Endocrinol*. 2016;174:755–762.
80. Weiss RE, Refetoff S. Effect of thyroid hormone on growth: Lessons from the syndrome of resistance to thyroid hormone. *Endocrinol Metab Clin North Am*. 1996;25:719–730.
81. Mitchell CS, Savage DB, Dufour S, et al. Resistance to thyroid hormone is associated with raised energy expenditure, muscle mitochondrial uncoupling, and hyperphagia. *J Clin Invest*. 2010;120:1345–1354.
82. Hauser P, Zametkin AJ, Martinez P, et al. Attention deficit-hyperactivity disorder in people with generalized resistance to thyroid hormone. *N Engl J Med*. 1993;328:997–1001.
83. Mixson AJ, Parrilla R, Ransom SC, et al. Correlation of language abnormalities with localization of mutations in the b-thyroid hormone receptor in 13 kindreds with generalized resistance to thyroid hormone: identification of four new mutations. *J Clin Endocrinol Metab*. 1992;75:1039–1045.
84. Refetoff S, Weiss RE, Usala SJ. The syndromes of resistance to thyroid hormone. *Endocr Rev*. 1993;14:348–399.
85. Stein MA, Weiss RE, Refetoff S. Neurocognitive characteristics of individuals with resistance to thyroid hormone: Comparisons with individuals with attention-deficit hyperactivity disorder. *J Dev Behav Pediatr*. 1995;16:406–411.
86. Valentine J, Rossi E, O'Leary P, et al. Thyroid function in a population of children with attention deficit hyperactivity disorder. *J Paediatr Child Health*. 1997;33:117–120.
87. Leonard CM, Martinez P, Weintraub BD, et al. Magnetic resonance imaging of cerebral anomalies in subjects with resistance to thyroid hormone. *Am J Med Genet*. 1995;60:238–243.
88. Brucker-Davis F, Skarulis MC, Pikus A, et al. Prevalence and mechanism of hearing loss in patients with resistance to thyroid hormone. *J Clin Endocrinol Metab*. 1996;81:2768–2772.
89. Bradley DJ, Towle HC, Young WS. α and β thyroid hormone receptor (TR) gene expression during auditory neurogenesis: evidence for TR isoform specific transcriptional regulation in vivo. *Proc Natl Acad Sci USA*. 1994;91:439–443.
90. Weiss AH, Kelly JP, Bisset D, et al. Reduced L- and M- and increased S-cone functions in an infant with thyroid hormone resistance due to mutations in the THRβ2 gene. *Ophthal Genet*. 2012;33:187–195.
91. Campi I, Cammarata G, Bianchi Marzoli S, et al. Retinal photoreceptor functions are compromised in patients with resistance to thyroid hormone syndrome (RTHß). *J Clin Endocrinol Metab*. 2017;102:2620–2627.
92. Forrest D, Erway LC, Ng L, et al. Thyroid hormone receptor β is essential for development of auditory function. *Nat Genet*. 1996;13:354–357.
93. Ng L, Hurley JB, Bierks B, et al. A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat Genet*. 2001;27:94–98.
94. Fukata S, Brent GA, Sugawara M. Resistance to thyroid hormone in Hashimoto's thyroiditis. *N Engl J Med*. 2005;352:517–518.
95. Barkoff MS, Kocherginsky M, Anselmo J, et al. Autoimmunity in patients with resistance to thyroid hormone. *J Clin Endocrinol Metab*. 2010;95:3189–3193.
96. Borck G, Seewi O, Jung A, et al. Genetic causes of goiter and deafness: Pendred syndrome in a girl and cooccurrence of Pendred syndrome and resistance to thyroid hormone in her sister. *J Clin Endocrinol Metab*. 2009;94:2106–2109.
97. Grasberger H, Ringkananont U, Croxson M, et al. Resistance to thyroid hormone in a patient with thyroid dysgenesis. *Thyroid*. 2005;15:730–733.
98. Gurnell M, Rajanayagam O, Barbar I, et al. Reversible pituitary enlargement in the syndrome of resistance to thyroid hormone. *Thyroid*. 1998;8:679–682.
99. Teng X, Ting J, Brent GA, et al. A patient with a thyrotropin-secreting pituitary microdenoma and Resistance to Thyroid Hormone (P453T). *J Clin Endocrinol Metab*. 2015;100:2511–2514.
100. Safer JD, Colan SD, Fraser LM, et al. A pituitary tumour in a patient with thyroid hormone resistance: a diagnostic dilemma. *Thyroid*. 2001;11:281–291.
101. Anselmo J, Cao D, Karrison T, Weiss RE, et al. Fetal loss associated with excess thyroid hormone exposure. *J Am Med Assoc*. 2004;292:691–695.
102. Sarne DH, Refetoff S, Rosenfield RL, et al. Sex hormone-binding globulin in the diagnosis of peripheral tissue resistance to thyroid hormone: the value of changes after short term triiodothyronine administration. *J Clin Endocrinol Metab*. 1988;66:740–746.
103. Usala SJ, Bale AE, Gesundheit N, et al. Tight linkage between the syndrome of generalized thyroid hormone resistance and the human c-erbA b gene. *Mol Endocrinol*. 1988;2:1217–1220.
104. Parrilla R, Mixson AJ, McPherson JA, et al. Characterization of seven novel mutations of the c-erbAb gene in unrelated kindreds with generalized thyroid hormone resistance: evidence for two "hot spot" regions of the ligand binding domain. *J Clin Invest*. 1991;88:2123–2130.
105. Adams M, Matthews CH, Collingwood TN, et al. Genetic analysis of twenty-nine kindreds with generalised and pituitary resistance to thyroid hormone. *J Clin Invest*. 1994;94:506–515.
106. Collingwood TN, Wagner R, Matthews CH, et al. A role of helix 3 of the TRβ ligand binding domain in coactivator recruitment identified by characterization of a third cluster of mutations in resistance to thyroid hormone. *EMBO J*. 1998;17:4760–4770.
107. Weiss RE, Weinberg M, Refetoff S. Identical mutations in unrelated families with generalized resistance to thyroid hormone occur in cytosine-guanine-rich areas of the thyroid hormone receptor β gene. *J Clin Invest*. 1993;91:2408–2415.
108. Mixson AJ, Renault JC, Ransom S, et al. Identification of a novel mutation in the gene encoding the β-triiodothyronine receptor in a patient with apparent selective pituitary resistance to thyroid hormone. *Clin Endocrinol*. 1993;38:227–234.
109. Mamanasiri S, Yesil S, Dumitrescu AM, et al. Mosaicism of a thyroid hormone receptor-beta gene mutation in resistance to thyroid hormone. *J Clin Endocrinol Metab*. 2006;91:3471–3477.
110. Weiss RE, Hayashi Y, Nagaya T, et al. Dominant inheritance of resistance to thyroid hormone not linked to defects in the thyroid hormone receptor α or β genes may be due to a defective cofactor. *J Clin Endocrinol Metab*. 1996;81:4196–4203.
111. Pohlenz J, Weiss RE, Macchia PE, et al. Five new families with resistance to thyroid hormone not caused by mutations in the thyroid hormone receptor β gene. *J Clin Endocrinol Metab*. 1999;84:3919–3928.
112. Meier CA, Dickstein BM, Ashizawa K, et al. Variable transcriptional activity and ligand binding of mutant β1 3,5,3'-triiodothyronine receptors from four families with generalised resistance to thyroid hormone. *Mol Endocrinol*. 1992;6:248–258.
113. Collingwood TN, Adams M, Tone Y, et al. Spectrum of transcriptional dimerization and dominant negative properties of twenty different

mutant thyroid hormone β receptors in thyroid hormone resistance syndrome. *Mol Endocrinol*. 1994;8:1262–1277.

114. Collingwood TN, Rajanayagam O, Adams M, et al. A natural transactivation mutation in the thyroid hormone β receptor: impaired interaction with putative transcriptional mediators. *Proc Natl Acad Sci USA*. 1997;94:248–253.
115. Takeda K, Sakurai A, De Groot LJ, et al. Recessive inheritance of thyroid hormone resistance caused by complete deletion of the protein-coding region of the thyroid hormone receptor-β gene. *J Clin Endocrinol Metab*. 1992;74:49–55.
116. Sakurai A, Miyamoto T, Refetoff S, et al. Dominant negative transcriptional regulation by a mutant thyroid hormone receptor β in a family with generalised resistance to thyroid hormone. *Mol Endocrinol*. 1990;4:1988–1994.
117. Chatterjee VKK, Nagaya T, Madison LD, et al. Thyroid hormone resistance syndrome: inhibition of normal receptor function by mutant thyroid receptors. *J Clin Invest*. 1991;87:1977–1984.
118. Ono S, Schwartz ID, Mueller OT, et al. Homozygosity for a dominant negative thyroid hormone receptor gene responsible for generalized resistance to thyroid hormone. *J Clin Endocrinol Metab*. 1991;73:990–994.
119. Ferrara AM, Onigata K, Ercan O, et al. Homozygous thyroid hormone receptor β-gene mutations in resistance to thyroid hormone: three new cases and review of the literature. *J Clin Endocrinol Metab*. 2012;97:1328–1336.
120. Nagaya T, Madison LD, Jameson JL. Thyroid hormone receptor mutants that cause resistance to thyroid hormone: evidence for receptor competition for DNA sequences in target genes. *J Biol Chem*. 1992;267:13014–13019.
121. Nagaya T, Jameson JL. Thyroid hormone receptor dimerization is required for dominant negative inhibition by mutations that cause thyroid hormone resistance. *J Biol Chem*. 1993;268:15766–15771.
122. Shibusawa N, Hashimoto K, Nikrodhanond AA, et al. Thyroid hormone action in the absence of thyroid hormone receptor DNA-binding in vivo. *J Clin Invest*. 2003;112:588–597.
123. Kitajima K, Nagaya T, Jameson JL. Dominant negative and DNA-binding properties of mutant thyroid hormone receptors that are defective in homodimerization but not heterodimerization. *Thyroid*. 1995;5:343–353.
124. Yoh SM, Chatterjee VKK, Privalsky ML. Thyroid hormone resistance syndrome manifests as an aberrant interaction between mutant T3 receptors and transcriptional corepressors. *Mol Endocrinol*. 1997;11:470–480.
125. Clifton-Bligh RJ, de Zegher F, Wagner RL, et al. A novel TRβ mutation (R383H) in resistance to thyroid hormone predominantly impairs corepressor release and negative transcriptional regulation. *Mol Endocrinol*. 1998;12:609–621.
126. Wagner RL, Huber BR, Shiau AK, et al. Hormone selectivity in thyroid hormone receptors. *Mol Endocrinol*. 2001;15:398–410.
127. Korwutthikulrangsri M, Dosiou C, Dumitrescu AM, et al. A novel G385E variant in the cold region of the T3-binding domain of Thyroid Hormone Receptor Beta gene and investigations to assess its clinical significance. *Eur Thyroid J*. 2019;8:293–297.
128. Refetoff S. Suggested guidelines for preparation of brief reports on new cases of resistance to thyroid hormone (RTH). *Thyroid*. 2009;4:139.
129. Hayashi Y, Weiss RE, Sarne DH, et al. Do clinical manifestations of resistance to thyroid hormone correlate with the functional alteration of the corresponding mutant thyroid hormone β receptors? *J Clin Endocrinol Metab*. 1995;80:3246–3256.
130. Hayashi Y, Janssen OE, Weiss RE, et al. The relative expression of mutant and normal thyroid hormone receptor genes in patients with generalized resistance to thyroid hormone determined by estimation of their specific messenger ribonucleic acid products. *J Clin Endocrinol Metab*. 1993;76:64–69.
131. Mixson AJ, Hauser P, Tennyson G, et al. Differential expression of mutant and normal β T3 receptor alleles in kindreds with generalized resistance to thyroid hormone. *J Clin Invest*. 1993;91:2296–2300.
132. Zavacki AM, Harney JW, Brent GA, et al. Dominant negative inhibition by mutant thyroid hormone receptors is thyroid response element and receptor isoform specific. *Mol Endocrinol*. 1993;7:1319–1330.
133. Geffner ME, Su F, Ross NS, et al. An arginine to histidine mutation in codon 311 of the c-erbAb gene results in a mutant thyroid hormone receptor that does not mediate a dominant negative phenotype. *J Clin Invest*. 1993;91:538–546.
134. Beck-Peccoz P, Chatterjee VKK. The variable clinical phenotype in thyroid hormone resistance syndrome. *Thyroid*. 1994;4:225–232.
135. Beck-Peccoz P, Roncoroni R, Mariotti S, et al. Sex hormone-binding globulin measurement in patients with inappropriate secretion of thyrotropin (IST): evidence against selective pituitary thyroid hormone resistance in nonneoplastic IST. *J Clin Endocrinol Metab*. 1990;71:19–25.
136. Ando S, Nakamura H, Sasaki S, et al. Introducing a point mutation identified in a patient with pituitary resistance to thyroid hormone (Arg 338 to Trp) into other mutant thyroid hormone receptors weakens their dominant negative activities. *J Endocrinol*. 1996;151:293–300.
137. Alberobello AT, Congedo V, Liu H, et al. An intronic SNP in the thyroid hormone receptor ß gene is associated with pituitary cell-specific overexpression of a mutant thyroid hormone receptor ß2 (R338W) in the index case of pituitary-selective resistance to thyroid hormone. *J Transl Med*. 2011;9:144.
138. Safer JD, O'Connor MG, Colan SD, et al. The thyroid hormone receptor-β gene mutation R383H is associated with isolated central resistance to thyroid hormone. *J Clin Endocrinol Metab*. 1999;84:3099–3109.
139. Wan W, Farboud B, Privalsky ML. Pituitary resistance to thyroid hormone syndrome is associated with T3 receptor mutants that selectively impair beta2 isoform function. *Mol Endocrinol*. 2005;19:1529–1542.
140. Safer JD, Langlois MF, Cohen R, et al. Isoform variable action among thyroid hormone receptor mutants provides insight into pituitary resistance to thyroid hormone. *Mol Endocrinol*. 1997;11:16–26.
141. Wu SY, Cohen RN, Simsek E, et al. A novel thyroid hormone receptor-beta mutation that fails to bind nuclear receptor corepressor in a patient as an apparent cause of severe, predominantly pituitary resistance to thyroid hormone. *J Clin Endocrinol Metab*. 2006;91:1887–1895.
142. Anselmo J, Refetoff S. Regression of a large goiter in a patient with resistance to thyroid hormone by every other day treatment with triiodothyronine. *Thyroid*. 2004;14:71–74.
143. Beck Peccoz P, Piscitelli G, Cattaneo MG, et al. Successful treatment of hyperthyroidism due to nonneoplastic pituitary TSH hypersecretion with 3,5,3'-triiodothyroacetic acid (TRIAC). *J Endocrinol Invest*. 1983;6:217–223.
144. Crino A, Borrelli P, Salvatori R, et al. Anti-iodothyronine autoantibodies in a girl with hyperthyroidism due to pituitary resistance to thyroid hormones. *J Endocrinol Invest*. 1992;15:113–120.
145. Radetti G, Persani L, Molinaro G, et al. Clinical and hormonal outcome after two years of TRIAC treatment in a child with thyroid hormone resistance. *Thyroid*. 1997;7:775–778.
146. Groeneweg S, Peeters RP, Visser TJ, et al. Therapeutic applications of thyroid hormone analogues in resistance to thyroid hormone (RTH) syndromes. *Mol Cell Endocrinol*. 2017;458:82–90.
147. Bracco D, Morin O, Schutz Y, et al. Comparison of the metabolic and endocrine effects of 3,5,3'-triiodothyroacetic acid and thyroxine. *J Clin Endocrinol Metab*. 1993;77:221–228.
148. Schueler PA, Schwartz HL, Strait KA, et al. Binding of 3,5,3'-triiodothyronine (T3) and its analogues to the in vitro translational products of c-erba protooncogenes: differences in the affinity of the alpha forms and beta forms for the acetic-acid analogue and failure of thehuman testis and kidney alpha-2 products to bind T3. *Mol Endocrinol*. 1990;4:227–234.
149. Takeda T, Suzuki S, Liu R-T, et al. Triiodothyroacetic acid has unique potential for therapy of resistance to thyroid hormone. *J Clin Endocrinol Metab*. 1995;80:2033–2040.
150. Ueda S, Takamatsu J, Fukata S, et al. Differences in response of thyrotropin to 3,5,3'-triiodothyronine and 3,5,3'-triiodothyroacetic acid in patients with resistance to thyroid hormone. *Thyroid*. 1996;6:563–570.
151. Asteria C, Rajanayagam O, Collingwood TN, et al. Prenatal diagnosis of thyroid hormone resistance. *J Clin Endocrinol Metab*. 1999;84:405–410.
152. Kunitake JM, Hartman N, Henson LC, et al. 3,5,3'-triiodothyroacetic acid therapy for thyroid hormone resistance. *J Clin Endocrinol Metab*. 1989;69:461–466.

153. Hamon P, Bovier-LaPierre M, Robert M, et al. Hyperthyroidism due to selective pituitary resistance to thyroid hormones in 15-month-old boy: Efficacy of D-thyroxine therapy. *J Clin Endocrinol Metab*. 1988;67:1089–1093.
154. Dorey F, Strauch G, Gayno JP. Thyrotoxicosis due to pituitary resistance to thyroid hormones. Successful control with D-thyroxine; a study in three patients. *Clin Endocrinol*. 1990;32:221–227.
155. Dulgeroff AJ, Geffner ME, Koyal SN, et al. Bromocriptine and TRIAC therapy for hyperthyroidism due to pituitary resistance to thyroid hormone. *J Clin Endocrinol Metab*. 1992;75:1071–1075.
156. Beck-Peccoz P, Mariotti S, Guillausseau PJ, et al. Treatment of hyperthyroidism due to inappropriate secretion of thyrotropin with the somatostatin analog SMS 201-995. *J Clin Endocrinol Metab*. 1989;68:208–214.
157. Moran C, Habeb AM, Kahaly GJ, et al. Homozygous resistance to thyroid hormone ß: can combined antithyroid drug and triiodothyroacetic acid treatment prevent cardiac failure? *J Endocr Soc*. 2017;1:1203–1212.
158. Weiss RE, Stein MA, Refetoff S. Behavioral effects of liothyronine (L-T3) in children with attention deficit hyperactivity disorder in the presence and absence of resistance to thyroid hormone. *Thyroid*. 1997;7:389–393.
159. Anzai R, Adachi M, Sho N, et al. Long-term 3,5,3'-triiodothyroacetic acid therapy in a child with hyperthyroidism caused by thyroid hormone resistance: pharmacological study and therapeutic recommendations. *Thyroid*. 2012;22:1069–1075.
160. Harrison SA, Bashir MR, Guy CD, et al. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet*. 2019;394:2012–2024.
161. Anselmo J, Cao D, Karrison T, et al. Fetal loss associated with excess thyroid hormone exposure. *JAMA*. 2004;292:691–695.
162. Dhingra S, Owen PJ, Lazarus JH, et al. Resistance to thyroid hormone in pregnancy. *Obstet Gynaecol*. 2008;112:501–503.
163. Pappa T, Anselmo J, Mamanasiri S, et al. Prenatal diagnosis of resistance to thyroid hormone and its clinical implications. *J Clin Endocrinol Metab*. 2017;102:3775–3782.
164. Tinnikov A, Nordstrom K, Thoren P, et al. Retardation of post-natal development caused by a negatively acting thyroid hormone receptor α1. *EMBO J*. 2002;21:5079–5087.
165. Kaneshige M, Suzuki H, Kaneshige K, et al. A targeted dominant negative mutation of the thyroid hormone α1 receptor causes increased mortality, infertility, and dwarfism in mice. *Proc Natl Acad Sci USA*. 2001;98:15095–15100.
166. Bochukova E, Schoenmakers N, Agostini M, et al. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*. 2012;366:243–249.
167. van Mullem A, van Heerebeek R, Chrysis D, et al. Clinical phenotype and mutant TRα1. *N Engl J Med*. 2012;366:1451–1453.
168. Moran C, Schoenmakers N, Agostini M, et al. An adult female with resistance to thyroid hormone mediated by defective thyroid hormone receptor α. *J Clin Endocrinol Metab*. 2013;98:4254–4261.
169. Moran C, Agostini M, Visser E, et al. Resistance to thyroid hormone caused by a mutation in thyroid hormone receptor (TR) alpha1 and alpha2: clinical, biochemical and genetic analyses of three related patients. *Lancet Diabetes Endocrinol*. 2014;2:619–626.
170. Tylki-Szymanska A, Acuna-Hidalgo R, Krajewska-Walasek M, et al. Thyroid hormone resistance syndrome due to mutations in the thyroid hormone receptor α gene (THRA). *J Med Gen*. 2015;52:312–316.
171. Di Cicco E, Moran C, Visser WE, et al. Germ line mutations in the thyroid hormone receptor alpha gene predispose to cutaneous tags and melanocytic nevi. *Thyroid*. 2021;31:1114–1126.
172. van Mullem AA, Chrysis D, Eythimiadou A, et al. Clinical phenotype of a new type of thyroid hormone resistance caused by a mutation of the TRα1 receptor: consequences of LT4 treatment. *J Clin Endocrinol Metab*. 2013;98:3029–3038.
173. Krieger T, Moran CM, Frangini A, et al. Mutations in thyroid hormone receptor a1 cause premature neurogenesis and progenitor cell depletion in human cortical development. *Proc Natl Acad Sci USA*. 2019;116:22754–22763.
174. Yuen RKC, Thiruvahindrapuram B, Merico D, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med*. 2015;21:185–191.
175. Kalikiri MK, Mamidala MP, Rao AN, et al. Analysis and functional characterization of sequence variations in ligand binding domain of thyroid hormone receptors in autism spectrum disorder (ASD) patients. *Autism Res*. 2017;10:1919–1928.
176. Van Gucht AL, Meima ME, Zwaveling-Soonawala N, et al. Resistance to thyroid hormone alpha in an 18-month-old girl: clinical, therapeutic and molecular characteristics. *Thyroid*. 2016;26:338–346.
177. Demir K, van Gucht A, Buyukinan M, et al. Diverse genotypes and phenotypes of three novel thyroid hormone receptor alpha mutations. *J Clin Endocrinol Metab*. 2016;101:2945–2954.
178. Moran C, Agostini M, McGowan A, et al. Contrasting phenotypes in resistance to thyroid hormone alpha correlate with divergent properties of thyroid receptor a1 mutant proteins. *Thyroid*. 2017;27:973–982.
179. Espiard S, Savagner F, Flamant F, et al. A novel mutation in THRA gene associated with an atypical phenotype of resistance to thyroid hormone. *J Clin Endocrinol Metab*. 2015;100:2841–2848.
180. Markossian S, Guyot R, Richard S, et al. CRISPR/Cas9 editing of the mouse Thra gene produces models with variable resistance to thyroid hormone. *Thyroid*. 2018;28:139–150.
181. Furman AE, Dumitrescu AM, Refetoff S, et al. Early diagnosis and treatment of an infant with a novel thyroid hormone receptor α gene (pC380SfsX9) mutation. *Thyroid*. 2020;31–1003–1005.
182. van Gucht ALM, Meima ME, Moran C, et al. Anemia in patients with resistance to thyroid hormone α: a role for thyroid hormone receptor α in human erythropoiesis. *J Clin Endocrinol Metab*. 2017;102:3517–3525.
183. Park S, Han CR, Park JW, et al. Defective erythropoiesis caused by mutations of the thyroid hormone receptor α gene. *PLoS Genet*. 2017;13:e1006991.
184. Wallis K, Sjogren M, van Hogerlinden M, et al. Locomotor deficiencies and aberrant development of subtype-specific GABAergic interneurons caused by an unliganded thyroid hormone receptor α1. *J Neurosci*. 2008;28:1904–1915.

82

Surgical Management of Thyroid Disease

Pia Pace-Asciak, Jonathan O. Russell, Christopher R. Razavi, Vaninder K. Dhillon, Ved Tanavde, Mohammad Shaear, and Ralph P. Tufano

OUTLINE

History of Thyroid Surgery, 1359
Relevant Surgical Anatomy, 1360
- The Recurrent Laryngeal Nerve, 1360
- The External Branch of the Superior Laryngeal Nerve, 1360
- Parathyroid Glands, 1361
- Lymphatics, 1361

Imaging For Surgical Planning, 1361
- Ultrasound, 1361
- Axial Imaging, 1361

Molecular Testing and Its Use For Surgical Planning, 1362
- Indeterminate Nodules, 1362
- Planning the Extent of Surgery, 1363

Diseases Of The Thyroid and Surgical Procedures To Consider, 1363
- Graves Disease, 1363
- Hashimoto Thyroiditis, 1363

Developmental Abnormalities of the Thyroid Gland And Their Surgical Management, 1363
- Thyroglossal Duct Cyst and Surgical Excision, 1363
- Thyroid Hemiagenesis and Agenesis, 1363
- Ectopic Thyroid and Surgical Excision, 1364
- Lingual Thyroid, 1364
- Suprahyoid and Infrahyoid Thyroid, 1364
- Substernal Goiter, 1364

Thyroid Cancer And The Surgical Management, 1365
- Papillary Thyroid Cancer, 1365
- Follicular Thyroid Cancer, 1365
- Noninvasive Follicular Thyroid Neoplasm with Papillary-Like Nuclear Features, 1365
- Hurthle Cell Thyroid Cancer, 1365
- Medullary Thyroid Cancer, 1366
- Anaplastic Thyroid Cancer, 1366
- Radioactive Iodine Therapy, 1367

Indications For Surgery of the Thyroid Gland, 1367
- Indications for Lobectomy versus Total Thyroidectomy, 1367
- Indications for Active Surveillance, 1367
- Indications for Remote Access Surgery, 1367
- Indications for Prophylactic Central Neck Dissection, 1367
- Indications for Therapeutic Central Neck Dissection, 1367
- Indications for Lateral Neck Dissection, 1367

Transoral Approach To Thyroidectomy, 1368
Open Surgical Techniques, 1368
- Lobectomy and Total Thyroidectomy, 1368
- Retrograde Medial Approach to the Thyroid Dissection, 1369
- Central Neck Dissection, 1369
- Lateral Neck Dissection, 1370

Minimally Invasive Techniques, 1370
- Radiofrequency Ablation Technique, 1370

Surgical Complications, 1370
- Recurrent Laryngeal Nerve Injury, 1370
- Hypoparathyroidism, 1371
- Spinal Accessory Nerve Injury, 1371
- Greater Auricular Nerve Injury and Sensory Cervical Rootlets, 1371
- Other Cranial Neuropathies, 1371
- Chyle Leak, 1371
- Cosmesis, 1371
- Complications Associated with Transoral Surgery, 1371
- Complications Associated with Radiofrequency Ablation, 1372

HISTORY OF THYROID SURGERY

The history of thyroid surgery dates back thousands of years, to as early as 952 AD, when al-Zahrawi (Albucasis) described the first thyroidectomy for resection of a large goiter.[1] Even though surgical treatment of a goiter was recognized, the developments that led to the contemporary era of thyroid surgery only began a century ago. Initially, thyroid surgery was fraught with complications, including hemorrhage, infection, and injury to the surrounding structures. With a mortality rate of over 40%, the French Academy of Medicine banned thyroid operations in 1850.[1] A turning point was reached at the end of the nineteenth century when Swiss surgeon Theodor Kocher focused on ligation for hemostasis, knowledge of the surrounding anatomy, and antisepsis.

Thyroid surgery was performed well before the physiology of the thyroid and parathyroid glands were understood. Kocher, a pupil of Theodor Billroth, was instrumental in bridging the physiological importance of the thyroid with surgical technique. He recognized that patients developed "cachexia strumipriva," also known as hypothyroidism, after total thyroidectomy, which led Kocher to perform mostly thyroid lobectomies to avoid patients developing this condition and "sending patients with a goiter into a vegetative existence."[1]

For his contributions to the field of thyroid surgery, Kocher was awarded the Nobel Prize in 1909, with morbidity dropping to less than 1% during his career.[1] Although thyroid surgery has undergone minor modifications, modern thyroid surgery has used much the same approach and technique since the late nineteenth century. Today, conventional transcervical thyroidectomy, when completed by an experienced, high-volume thyroid surgeon, has an exceedingly low mortality rate, with relatively little morbidity.[2]

As the rate of thyroid cancer continues to grow worldwide, clinicians forge ahead to find innovative ways to treat thyroid disease in a minimally invasive and personalized way. Understanding the way in which thyroid disease behaves at a molecular level can help clinicians to effectively target the disease, more or less aggressively, depending on the patient's risk of recurrence. In an era of "personalized medicine," patient care is tailored to meet the needs of the individual patient, rather than the "one size fits all" traditional approach. Options may include surveillance, the traditional midline incision approach that Kocher fathered, or no cutaneous incision at all through remote access surgery using an endoscope or by radiofrequency ablation (RFA) without removing the thyroid gland at all. With improved imaging, biochemical testing, and molecular testing, as well as surgical instrumentation, the pendulum has swung from treating all thyroid disease with one type of surgical approach to a more selective approach that risk-stratifies the disease, as well as the patient's needs.

RELEVANT SURGICAL ANATOMY

The Recurrent Laryngeal Nerve

Surgical Landmarks. Various challenges can be encountered intraoperatively, such as a bulky goiter with multiple nodules, a recurrent laryngeal nerve (RLN) that is encased by tumor, bleeding, or a congenital variation such as a nonrecurrent laryngeal nerve (NRLN). Stable landmarks can help guide the surgeon to avoid inadvertent injury to the RLN. Additionally, intraoperative nerve monitoring is a valuable adjunct for confirming the identification and function of the RLN in these situations, and is used by most high-volume thyroid surgeons in the Western world.[3]

Berry's Ligament. Berry's ligament (BL) and the tracheoesophageal groove (TEG) can help with identification of the RLN during surgery.[4] Up to 80% of the time, the RLN is located superficial to BL, and much less commonly it is found deep to BL (15%).[5] This relationship between BL and the RLN tends to be a symmetrical on either thyroid lobe. Care should be taken when dissecting BL before identification of the RLN, for less common instances when the RLN pierces BL.[6] It is common for anterior and posterior branches of the nerve to bifurcate within BL. In such cases, the anterior or medial branch is most likely to be the motor branch of the nerve, while the posterior branch is sensory.

Tubercle of Zuckerkandl. The tubercle of Zuckerkandl is a pyramidal extension of the thyroid lobe that is located more laterally and posterior on the thyroid.[7] It is a thickening of the lateral aspect of the thyroid gland and is closely connected to the superior parathyroid gland. Surgically, the tubercle of Zuckerkandl is significant, as it may represent a landmark for identifying the RLN, which runs beneath the projection as it enters the cricothyroid joint.[7]

Cricothyroid Joint. The cricothyroid joint is one of the most consistent and fixed landmarks for identification of the RLN entry into the larynx. Palpation of the inferior horn of the thyroid cartilage indicates the depth of the RLN, which is reliable regardless of thyroid pathology or the presence of congenital variations.[8] Variability in the angle of the RLN can be found as it courses toward the cricothyroid junction; however, identification of BL and the cricothyroid joint will yield the RLN just posterior to these structures.[8,9]

Inferior Thyroid Artery. Commonly, the posterior division of the inferior thyroid artery (ITA) has been found to accompany the RLN.[10] The relationship of the ITA to the RLN tends to be variable and asymmetric when the right and left sides are compared. The most common presentation on the left side is for the RLN to be posterior to the ITA, and on the right side, the RLN is anterior to the ITA.[11]

Factors that Make the Recurrent Laryngeal Nerve Vulnerable to Injury

Nonrecurrent Laryngeal Nerve. A NRLN is a rare finding with an incidence of 0.3% to 0.8% on the right and 0.004% on the left side.[12] An NRLN on the right side is almost always associated with a vascular anomaly. The presence of an aberrant right subclavian artery (arteria lusoria) from the descending aorta is associated with the presence of a right NRLN. A left NRLN is extremely rare, and when present, it associated with *situs inversus* or a right-sided aortic arch.[12]

The Proximity of the Recurrent Laryngeal Nerve to the Thyroid Gland. The RLN can be found in the TEG approximately 65% of the time. Less commonly, it may be lateral or anterior to the TEG when large nodules or substernal goiters are present.[5] Sometimes the thyroid parenchyma may envelope the RLN. Again, nerve identification is considered necessary by most high-volume surgeons. In a landmark paper, Lahey reported a significantly lower incidence of RLN injury following thyroidectomy with clear visualization of the RLN than after operation without nerve identification.[13]

Deformities from Large Thyroid Nodules. In the presence of large nodules and substernal goiters, the laryngeal nerves may be pushed more anteriorly than expected or stretched from compression of a large goiter. Careful blunt dissection is recommended for a safe approach until the nerve is identified.

Extralaryngeal Nerve Branching of the Recurrent Laryngeal Nerve. Variations in the branching of the RLN are common and can increase the risk of transient injury 2-fold to 15.8% compared with a nonbranched RLN (8%).[14] The RLN may branch before the nerve enters into the larynx, near BL in approximately 30% to 78% of cases.[15] The branching may differ from side to side, but most often the motor innervation comes from the anterior branch, and the sensory innervation is posteriorly.[15] If the sensory division is ligated, the patient may experience postoperative dysphagia; however, if the anterior branch is ligated, the patient is likely to experience vocal fold paralysis.

Inferior Thyroid Artery. Given the variable location of the ITA in relation to the RLN, it is recommended that the nerve is identified first prior to ligating its branches.

The External Branch of the Superior Laryngeal Nerve

The right and left superior laryngeal nerve (SLN) exit from the vagus and course with the superior thyroid artery until approximately 1 cm before the superior pole, where they enter the capsule of the superior pole of the thyroid.[16,17] The SLN gives off two branches. The first is the external branch, which is primarily motor in function and innervates the cricothyroid muscles.[16,17] The cricothyroid muscle causes the thyroid cartilage to tilt, which tenses the vocal cord and raises the pitch or projection of the voice. The second is the internal branch, which is sensory to the larynx (subglottis) and enters the thyrohyoid membrane, which can result in dysphagia if the sensory branch is injured.[16]

Factors that Make the External Branch of the Superior Laryngeal Nerve Vulnerable to Injury. In most cases, the external branch of the SLN (EBSLN) lies close to the vascular pedicle of the superior thyroid vessels. In some patients, the EBSLN lies on the anterior surface of the thyroid lobe, making the possibility of damage during thyroidectomy even greater.[17] Cernea et al. proposed a classification system for describing the proximity of the EBSLN with the superior thyroid vessels.[18] In the type 1 configuration, the EBSLN crosses the superior thyroid vessels 1 cm or more above the horizontal plan, passing the upper border of the superior thyroid pole. In type 2a, the EBSLN crosses the superior thyroid vessels less than 1 cm above the plane, and in type 2b, the EBSLN crosses the superior thyroid vessels below the plane. The highest-risk configuration is the type 2b nerve, making the nerve vulnerable to iatrogenic injury, particularly for enlarged goiters, because it travels close to the thyroid gland. Even placement of a tie close to the thyroid pole could put the nerve in jeopardy.[18] Thus, care should be taken when ligating the superior vascular pedicle of the thyroid by aiming to dissect close to the thyroid capsule after the EBSLN has been identified, as injury can occur in up to 30% of cases.[16,17] Another way to minimize injury is to visualize the cricothyroid muscle so identification and protection of the EBSLN is possible.

Parathyroid Glands

Even though normal parathyroid glands are typically the size of a grain of rice, they hold the utmost physiological importance in their regulation of calcium hemostasis through the secretion of parathyroid hormone (PTH). There are typically four glands present, two on each side. Rarely, supernumerary glands have been described.[19] Each gland normally weighs 30 to 40 mg. The delicate blood supply of these glands and their small size makes them vulnerable to accidental removal or devascularization during thyroidectomy.[19]

Typically, the inferior parathyroid glands are located below the ITA and anterior to the RLN along the posterior side of the thyroid lobe,[20] whereas the upper parathyroid glands are found above the ITA and usually adjacent to the posterior surface of the middle part of the thyroid lobe, near the RLN as it enters toward the cricothyroid joint within 1 cm of the RLN insertion.[20] The lower parathyroid glands embryologically arise from the third pharyngeal pouch, and their distribution in the neck may be variable, whereas the upper parathyroid glands are derived from the fourth pharyngeal pouch.

Parathyroid glands have a tan or salmon color. With surgical experience, these glands become easier to differentiate from lymph nodes, fat, or thyroid nodules. Frozen section examination during surgery can be helpful in their identification. Additionally, novel instruments that rely on the inherent autofluorescent properties of parathyroid tissue are becoming more commonplace in thyroid surgery.[21]

Lymphatics

The central neck (levels VI and VII) is the primary drainage site for metastatic spread of thyroid carcinoma. Approximately 35% of patients presenting with PTC have cN1 disease.[15] However, up to 80% of patients presenting with cN0 disease may harbor microscopically positive lymph nodes.[15] The lateral neck compartments (levels I–V) can be involved, but less frequently than the central neck. Generally, thyroid cancer spreads sequentially from the central neck to the lymph nodes of the lateral neck. Less commonly, metastasis is found in the retropharyngeal or retroesophageal nodes. Rarely, involvement is seen in the mediastinal lymph nodes.[15] Of note, up to 18% of patients can have skip metastasis to levels II and III of the lateral neck without central neck metastasis, arising from a superior pole cancer.[22] Only clinically detectable lateral nodal metastasis are significant.

Central Neck. The central compartment (level VI) is bounded superiorly by the hyoid bone, laterally by the carotid arteries, inferiorly by the sternal notch and innominate artery, anteriorly by the superficial layer of the deep cervical fascia, and posteriorly by the deep layer of the deep cervical fascia or prevertebral fascia. The central neck is divided further into subcompartments, which can be independently dissected during central neck dissection (CND) for PTC and is based on anatomical location. These subcompartments include: (1) the prelaryngeal (Delphian) lymph nodes, (2) the pretracheal lymph nodes, (3) the right paratracheal lymph nodes, and (4) the left paratracheal lymph nodes.[15,23] Level VII lymph nodes are associated with the brachiocephalic vein and innominate artery.[15] The lowest point in the CND is defined by the innominate artery on the right (at its point of tracheal crossing) and the corresponding axial plane on the left.

The right and left paratracheal regions differ anatomically and in terms of the lymph node distribution around the RLN. The lymphatic tissue in the right paratracheal region is more anterior and posterior to the right RLN due to the presence of the innominate artery, which crosses anterior to the trachea.[15,24] Thus, to completely remove the lymphatic tissue in the right paratracheal compartment, the surgeon needs to transpose the RLN in order to dissect both anterior and deep to the nerve to avoid disease recurrence in this unfavorable location.[24,25] On the other hand, the left RLN runs in a more cranial to caudal direction in the TEG versus a more oblique orientation found on the right. As a result, the lymphatic tissue in the left paratracheal region is located anterior to the RLN and the esophagus but medial to the carotid artery.[15]

Lateral Neck. A constant group of lymph nodes (the lateral neck compartment) lies along the internal jugular vein (IJV) on each side of the neck, and may be divided into levels I to V, corresponding with the submandibular and submental nodes (level I); upper, middle, and lower jugular nodes (levels II, III, and IV); and posterior triangle nodes (level V).[23] The boundaries of the level I compartment are defined by the mandible superiorly, the stylohyoid muscle posteriorly, and the anterior belly of the digastric muscle anteriorly. This level is further subdivided into level Ia, the submental triangle (bounded by the anterior bellies of the digastric muscle and the hyoid bone) and level Ib, containing the submandibular gland and its nodes.[23] Level II, which contains the jugulodigastric lymph nodes, is located in the upper third of the jugular vein, with the stylohyoid muscle as the anterior border and the sternocleidomastoid muscle (SCM) as the posterior border. The spinal accessory nerve runs obliquely and subdivides level IIa, which is anterior to the nerve, from level IIb, which is posterior to the nerve. Level III contains the midjugular nodes, with the hyoid defining the superior border and the cricoid cartilage as the inferior boundary. The anterior limit is the sternohyoid muscle, and the posterior boundary is the SCM. Level VI, the lower jugular nodes, are bounded by the inferior aspect of the cricoid cartilage superiorly and the clavicle inferiorly. The anterior limit is the sternohyoid muscle, and the posterior limit is the SCM. Level V, the posterior compartment, is bounded anteriorly by the SCM and posteriorly by the trapezius muscle. The plane that divides the two compartments is the inferior border of the cricoid cartilage, which subdivides it into level Va, containing the nodes from the spinal accessory nerve and located superiorly; whereas level Vb, located inferiorly, contains the transverse cervical and supraclavicular nodes.[23]

KEY POINTS

- Various surgical landmarks (Berry's ligament, Tubercle of Zuckerhandl, Cricothyroid joint, and the Inferior Thyroid artery) can help surgeons identify the RLN or the EBSLN (Cricothyroid muscle and Superior Thyroid artery).

IMAGING FOR SURGICAL PLANNING

Ultrasound

Ultrasound is the first-line imaging modality for evaluating known or suspected thyroid cancer, as well as overall evaluation of the thyroid gland.[26] Knowing the sonographic pattern of the nodule is key in establishing whether to investigate further (Table 82.1).

Axial Imaging

Cross-sectional imaging such as computed tomography (CT) scan or magnetic resonance imaging (MRI) may have a supplemental role in preoperative imaging for thyroid cancer in the minority of cases and has its advantages over thyroid ultrasonography.

TABLE 82.1 Comparison of ATA with ACR-TIRADS Classification

2015 ATA Classification	2017 ACR TI-RADS Classification
High suspicion	**TR5 (high suspicion)**
70%–90% risk of malignancy	
Solid hypoechoic thyroid nodules with one or more of following:	
Irregular margin	
Microcalcification	
Taller than wide	
Rim calcification with small, extrusive soft-tissue component	>20% risk of malignancy
Extrathyroidal extension	TI-RADS score of 7 points or more
Intermediate suspicion	**TR4 (moderate suspicion)**
10%–20% risk of malignancy	5%–20% risk of malignancy
Hypoechoic nodule with smooth margin	TI-RADS score of 4–6
Low suspicion	**TR3 (mild suspicion)**
5%–10% risk of malignancy	5% risk of malignancy
Isoechoic or hyperechoic solid nodule or partially cystic nodule with eccentric solid areas	TI-RADS score of 3
Very low suspicion	**TR2 (not suspicious)**
<3% risk of malignancy	≤2% risk of malignancy
Spongiform or partially cystic nodules	TI-RADS score of 2
Benign	**TR1 (benign)**
Benign: pure cyst	≤2% risk of malignancy
	TI-RADS score of 0

ACR, American College of Radiology; *ATA*, American Thyroid Association; *TI-RADS*, Thyroid Imaging Reporting and Data System; *TR1*, TI-RADS level 1; *TR2*, TI-RADS level 2; *TR3*, TI-RADS level 3; *TR4*, TI-RADS level 4; *TR5*, TI-RADS level 5.

From Ahmadi S, Oyekunle T, Jian XS, et al. A direct comparison of the ATA and TI-RADS ultrasound scoring systems. *Endocrine Practice.* 2019;25:413–422.

Computed Tomography Scan. CT with intravenous contrast is the most widely accepted form of axial imaging. Of note, the intravenous contrast used for CT scan contains a large load of iodine, which may interfere with radioiodine (RAI) scanning and therapy for 6 to 8 weeks. The use of iodinated contrast may delay postoperative treatment with RAI, but the usefulness in preoperative planning justifies a delay when indicated.[31]

The circumstances where axial imaging is indicated include: (a) clinical or sonographic evidence of an invasive primary tumor, (b) presence of a large primary tumor or bulky nodal disease incompletely imaged with ultrasound, (c) presence or extension of nodal disease outside of the central neck, and (d) absence of sonographic expertise in evaluation of cervical lymph nodes.[32] Cross-sectional imaging also has a role in identification of the substernal or suprasternal extent of benign thyroid disease.

Magnetic Resonance Imaging. MRI has specific applications for staging and surveillance, as well as localization of metastatic disease for thyroid cancer, but tends not to be used for benign disease.[31] Given that CT scan is more sensitive for detecting small metastases to lymph nodes and the lungs, generally, MRI does not play a large role in the initial work-up of thyroid disease and is reserved for more complicated cases. MRI imaging of the neck may be considered in thyroidectomy patients with rising serum thyroglobulin (Tg) levels and a history of RAI therapy and neck dissection.[33]

Positron Emission Tomography Scan. Functional imaging known as (^{18}F)-fluorodeoxyglucose–positron emission tomography (FDG-PET) scan is a nuclear medicine test that detects malignancies based on glucose hypermetabolism. Routine preoperative FDG-PET scanning is not recommended for indeterminate nodules or locoregional thyroid carcinoma, as glucose uptake can be nonspecific when inflammation is present, such as in Hashimoto thyroiditis.[26] Generally, PET scan is best used in more advanced disease if distant metastases are suspected and to help stage a patient before further treatment.[32]

Thyroid incidentalomas detected by FDG-PET scan are infrequent but can be picked up on imaging during work-up for an unrelated problem. The literature reports a detection rate of 0.1% to 4.3% for incidental findings of thyroid focal uptake identified by FDG-PET/CT, with a risk of malignancy of between 10.3% and 80%.[34] Given the wide range for potential malignancy, it is prudent to follow up a positive PET scan with a dedicated thyroid ultrasound to further delineate "incidentalomas."

Thyroid Scintigraphy. When the thyroid-stimulating hormone (TSH) is low (patient is hyperthyroid), thyroid scintigraphy (^{123}I or technetium-99m pertechnetate) or RAI scanning is the appropriate imaging modality to determine if a toxic nodule(s) is present.[35] Scintigraphy provides added information on the functional status of the thyroid gland or nodule, a feature that cross-sectional imaging lacks. Other indications for scintigraphy are for evaluation of a patient suspected of having an ectopic thyroid gland and to determine whether the lingual or substernal tissue is functional.[36] Further, patients who have had thyroid surgery for cancer will have whole-body scintigraphy to define the amount of thyroid remnant, which allows for Tg monitoring for recurrence postoperatively.

KEY POINTS

- Ultrasound is the imaging modality of choice for evaluating thyroid nodules. Axial Imaging (CT scan of MRI) is reserved for primary or recurrent invasive thyroid cancer, lateral neck disease or retrosternal extension.

MOLECULAR TESTING AND ITS USE FOR SURGICAL PLANNING

Indeterminate Nodules

Molecular testing is a topic of debate, as the current paradigm suggests that, among multiple molecular tests, there is no one test that can guarantee carcinoma from a histologically indeterminate nodule. Molecular testing should only be used if it is going to add to the surgical decision-making process for indeterminate Bethesda cytology results. The following scenarios are examples where molecular testing is unlikely to rule out malignancy: a) nodules that yield insufficient material for diagnosis, benign nodules, suspicious for malignancy, or malignant nodules on cytological evaluation; b) nodules with strong clinical or radiological findings suggestive of a thyroid malignancy; c) indeterminate nodules greater than 4 cm, a size at which nodules may harbor an occult thyroid cancer owing to fine-needle aspiration (FNA) sampling error; d) nodules causing compressive symptoms for which surgery would be recommended regardless; and e) patients who defer surgery (i.e., uncomfortable with conservative management of a nodule regardless of the test result.[29] Molecular testing is best suited to patients with indeterminate Bethesda cytology and minimal clinical suspicion for malignancy and if the result would alter decision-making.

Planning the Extent of Surgery

Some centers use molecular testing to anticipate aggressive tumor types, which can in turn be helpful for planning the extent and timing of surgery. The presence of *BRAF* V600E, *RET/PTC1* rearrangement, or *TERT* mutation may make the discussion of surgery easier with a patient who has an indeterminate Bethesda score, knowing that these molecular markers are associated with more aggressive cancers 65% of the time.[43]

KEY POINTS

- Molecular testing should only be used if it is going to add to the surgical decision-making process for indeterminate Bethesda cytology results.

DISEASES OF THE THYROID AND SURGICAL PROCEDURES TO CONSIDER

Graves Disease

In the United States, the prevalence of hyperthyroidism is 1.2%.[35] The most common causes of hyperthyroidism are Graves disease, toxic multinodular goiter (TMNG), and toxic adenoma (TA).[35]

Management of thyrotoxicosis includes beta blocker therapy, whereas Graves disease, TMNG, and TA can be managed with RAI, antithyroid drugs (ATDs), or surgery.[35] Surgery tends to be reserved for patients with a large goiter causing compressive type symptoms or for those who have severe hyperthyroidism that is difficult to control with ATDs or RAI.[44] Patients with an allergy to ATDs, who are of reproductive age and do not wish to pursue RAI, or who have a thyroid nodule that is suspicious for malignancy can also be considered for total thyroidectomy or near-total thyroidectomy.

Malignancy is less likely to occur in hyperfunctioning thyroid nodules. However, an updated systematic review demonstrates that the overall rate of thyroid malignancy in patients with hyperthyroidism (Graves disease, TMNG, or single toxic nodules) is 8.5% (range 0.8%–32.4%).[45] Patients with Graves disease and coexisting thyroid nodules are five times more likely to be diagnosed with thyroid carcinoma than those patients with Graves disease without nodules.[46] Thus, during the work-up for hyperthyroidism, nodules should be evaluated for concurrent thyroid malignancy with FNA biopsy.[46]

Surgical resection involves subtotal thyroidectomy, near-total thyroidectomy, or lobectomy with contralateral subtotal or near-total lobectomy (Dunhill procedure).[36] Evidence shows that, if a remnant of 2 to 2.5 g or more of thyroid tissue is left behind, a recurrence rate of approximately 12% at 10-year follow-up exists.[47] Hence, leaving only a small amount of tissue behind, if at all, is recommended, along with postoperative thyroxine treatment, especially in young patients where the rate of recurrence of thyrotoxicosis tends to be higher. When operating on patients with Graves and severe ophthalmopathy, near-total or total thyroidectomy is the treatment of choice for improvement in eye symptoms.

Hashimoto Thyroiditis

Hypothyroidism is found in 4.6% of the US population (0.3% clinical and 4.3% subclinical).[48] Of the different causes of hypothyroidism, Hashimoto thyroiditis, also known as chronic autoimmune thyroiditis, is one of the most common in iodine-sufficient areas of the world, characterized by antibodies to thyroid peroxidase and Tg.[48]

Hashimoto's itself or hypothyroidism is not an indication for surgery, unless there are suspicious nodules present on ultrasound, pressure symptoms, or concerns of cosmesis. However, Heggie et al. found that patients with Hashimoto's were more likely to have pressure symptoms, even though their glands were significantly lighter compared with those in the larger nonthyroiditis cohort.[49] In their retrospective series, 90% of patients with Hashimoto's had significant improvement in their pressure symptoms postthyroidectomy. Thus, surgery may be considered in patients with significant pressure symptoms with a diagnosis of Hashimoto's, regardless of the size of the goiter.[49]

KEY POINTS

- For symptomatic Graves' disease, surgical resection can involve a subtotal thyroidectomy or near-total thyroidectomy. Although Hashimoto's disease itself is not an indication for surgery, patients with pressure symptoms or suspicious nodules may be considered for surgical resection.

DEVELOPMENTAL ABNORMALITIES OF THE THYROID GLAND AND THEIR SURGICAL MANAGEMENT

Thyroglossal Duct Cyst and Surgical Excision

A thyroglossal duct cyst (TGDC) is a congenital anomaly formed during the embryological development of the thyroid gland. As the thyroid descends, its pathway of descent forms the thyroglossal duct tract, which disappears by the tenth week of gestation. If the tract persists and fails to obliterate anywhere along the pathway of descent, a midline cyst persists known as a TGDC.[50] If a distal remnant of the tract persists, then the pyramidal lobe is formed. TGDC can be seen at birth in about 25% of cases, with approximately one third becoming apparent after 30 years of age.[50] Patients present with a painless, midline cystic neck mass localized next to the hyoid bone (60%), between the hyoid bone and base of tongue (24%), between the hyoid and pyramidal lobe (13%), or intralingually (3%).[51,52] The cyst can become repeatedly infected and at times rupture, developing a persistent fistula or sinus tract to the outside skin or into the oral cavity. Physical exam is diagnostic, as the cyst moves cranially with swallowing or protrusion of the tongue.

In the wall of a TGDC is the second most common place for ectopic thyroid tissue.[53] If a solid component is found in the cyst, this may be the only functional thyroid tissue present, and follow-up TSH levels should be obtained pre- and postsurgery, as well as an ultrasound of the thyroid. The incidence of thyroid carcinoma in a TGDC is less than 1%, which may be less than expected, because excision is often done during childhood before a malignancy has had a chance to form.[53] In adults, there have been reported cases of incidental papillary thyroid carcinoma (PTC) found within the cysts, at a rate as high as 13% to 14%.[54] In a systematic review of the literature, the most common pathologic diagnosis was papillary cancer, at 92.1%, followed by squamous cell carcinoma, at 4.3%.[55] In such cases, there is an associated risk factor present, such as previous low-dose external beam irradiation of the head and neck in infancy or childhood as a trigger.

The treatment of choice for TGDC to avoid recurrence is the Sistrunk procedure. A Sistrunk procedure is the comprehensive excision of the cyst and its tract, which includes resection of the central portion of the hyoid bone as well as the tract up to the base of tongue. This procedure is necessary because the thyroglossal duct is intimately associated with the central part of the hyoid bone. An incision is made on the skin over the mass, and dissection is done bluntly to isolate the cyst and then the hyoid bone. Excision up to the base of tongue results in cure in most cases.[52] If a fistula tract connects to the skin, this should be excised as well to avoid recurrence.

Thyroid Hemiagenesis and Agenesis

Thyroid hemiagenesis (THA) is a rare congenital disorder that is characterized by the absence of one thyroid lobe. If the median thyroid anlage fails to develop, the result is athyreosis. Often THA is picked up incidentally on ultrasound, and less frequently on clinical exam, because most patients are euthyroid and asymptomatic. Left-sided isolated agenesis is the most common presentation (80%); however, combined THA of the right lobe with absence of the isthmus can occur, but less frequently.[56] When operating on a thyroid where hemiagenesis is noted, the parathyroid glands will remain intact, given their separate embryologic origin from the thyroid (third and fourth pharyngeal pouches).

Rarely, the thyroid gland fails to develop altogether, resulting in athyreosis. If left untreated in the newborn, this can lead to cretinism.

Ectopic Thyroid and Surgical Excision

Ectopic thyroid tissue is the most common cause of congenital hypothyroidism.[41] Failure of thyroid descent occurs in 1 out of 200,000 normal subjects and 1 out of 6000 patients with thyroid disease.[57] Ectopic thyroid is functioning thyroid tissue that can be found anywhere along the normal pathway of descent, from the foramen cecum down to the pretracheal region. Ectopic tissue can result in a lingual thyroid in a suprahyoid, infrahyoid, or intratracheal location. Less commonly, the entire thyroid gland can descend into the superior mediastinum behind the sternum, adjacent to the aortic arch or between the aorta and pulmonary trunk, within the upper portion of the pericardium, and even within the interventricular septum of the heart.[36] Care must be taken when surgically removing ectopic thyroid tissue, as it may be the only functioning thyroid tissue present. This is an important aspect to discuss with patients preoperatively, regarding the potential for a lifetime of thyroid hormone replacement therapy.

Lingual Thyroid

A lingual thyroid is a rare finding and is estimated to occur in 1 out of 3000 cases of thyroid disease.[58] However, it represents the most common site for functioning ectopic thyroid tissue. In 75% of patients with a lingual thyroid, it is their only functioning tissue due to the absence of normal cervical thyroid, or it functions poorly, resulting in hypothyroidism.[35] Malignant transformation of the lingual thyroid is rare (1 per 100); however, when it does occur, papillary or follicular carcinoma are the likely subtypes.[59]

The presence of hypothyroidism may cause the mass to enlarge, causing dysphagia, dysphonia, dyspnea, or a sensation of choking.[59] Rarely is a lingual thyroid associated with hyperthyroidism. Flexible nasopharyngoscopy shows the presence of a submucosal mass at the base of the tongue with prominent vascular markings akin to thyroid tissue. Radionuclide scanning can confirm the diagnosis. Up to 70% of patients require thyroid hormone replacement to treat their hypothyroidism, which can potentiate involution of the ectopic tissue and improve symptoms.[60] However, symptomatic patients who fail medical treatment may require surgical excision. Transoral surgical access to the tongue base is challenging to gain appropriate visualization in a narrow working space. Open approaches such as labiotomy, mandibulotomy, tongue-split, or cervical pharyngotomy provide exposure, but at the cost of being invasive.[37-40] Transoral robotic-assisted surgery (TORS) has emerged as a minimally invasive approach to the oropharynx and hypopharynx, with decreased morbidity that provides exceptional 3D visualization and instrument navigation.[40] Prisman et al. describe TORS-assisted excision of an ectopic lingual thyroid gland as safe and feasible.[61]

Suprahyoid and Infrahyoid Thyroid

Similar to a lingual thyroid gland, the descent of thyroid tissue can be arrested in the midline position above or below the hyoid bone. An enlarging mass commonly occurs during infancy or childhood, or later in life. Often, this mass is mistaken for a TGDC because it is in the same anatomic position. Prior to surgical removal, it is recommended that ultrasound of the neck is done to ensure the presence of a normal thyroid gland.

Substernal Goiter

Most substernal goiters originate in the neck and then “fall” or descend into the mediastinum due to gravity, rather than from aberrant thyroid tissue that was embryologically determined. Isolated intrathoracic thyroid masses are rare but occur in 5.8% of all mediastinal masses.[62] Up to 40% of substernal goiters are asymptomatic and found incidentally.[62] The most accepted definition of a retrosternal goiter is a thyroid gland that extends below the thoracic inlet.[62] A mediastinal mass can also be described by its association with the aortic arch or whether it reaches the level of T4 on a chest x-ray, or whether more than 50% of the gland extends below the thoracic inlet.[62]

Most substernal goiters are discovered incidentally on routine chest radiography in patients who are completely asymptomatic. If symptoms are present, the most common symptom when the thyroid extends below the thoracic inlet is dyspnea from tracheal displacement or compression that may or may not be positional.[62] The second most common symptom is dysphagia from esophageal compression, mostly to solids, and can be associated with laryngopharyngeal reflux.[62] Dysphonia is less common but can occur due to compression on the RLN, which can be assessed by flexible nasopharyngoscopy preoperatively. Even less frequently, a substernal goiter can cause superior vena cava syndrome, which often shows the presence of Pemberton sign.[62] A positive sign includes facial plethora, jugular vein distension, and symptoms of dyspnea when the patient lifts their arms above their head for a few minutes.[27]

When patients present in acute airway distress from tracheal compression due to their enlarged goiter, early surgical intervention is the preferred approach.[89] In a systematic review, preoperative positional dyspnea (defined as difficulty breathing at rest but improved by position change) was improved or resolved in 82.4% of cases overall post-thyroidectomy, in 65% of patients with cervical goiter, and in 88% of patients with substernal goiter.[94]

When faced with a large goiter causing tracheal compression and/or deviation, most patients can be easily intubated, because it is often the larynx that is misplaced. The equipment that should be available in the room to manage the difficult airway includes a rigid bronchoscope, fiberoptic intubation equipment, video laryngoscopes, and a laryngeal mask airway (through which a fiberoptic intubation can be performed).[89] The surgeon should be in the room on induction in the event that an emergent tracheotomy or rigid bronchoscopy is needed. Having the patient awake during fiberoptic induction may be helpful when severe tracheal compression is noted, as well as intubation with a small diameter endotracheal tube.[89] Ideally, an endotracheal nerve-monitoring tube would be inserted, but the outer diameter tends to be larger and may pose difficulty during intubation in cases with severe tracheal compression. Consideration can be made for staging the procedure if there is any concern for the functional status of the RLN during resection.

Most substernal goiters are found in the anterior mediastinum, and less frequently in the posterior mediastinum.[62] A CT scan of the neck and chest with contrast should be done to help delineate the extent and the blood supply and for surgical planning. Generally, the blood supply to the substernal thyroid gland originates in the neck and can be readily controlled in this area. Hemostasis can be obtained by control of the superior and inferior thyroid arteries; hence, removal through the neck with a transcervical incision can be an option. Only rarely does an intrathoracic goiter receive its blood supply from mediastinal vessels; however, such a finding favors a developmental cause. In this case, a sternotomy is anticipated, particularly if the goiter extends below the aortic arch. When in doubt, a multidisciplinary approach should be coordinated with preoperative consultation involving thoracic and vascular surgery.[62] Video-assisted thoracoscopic surgery can also be used to mobilize the mediastinal portion of the substernal goiter and aid in its removal in selected cases.[36]

KEY POINTS

- The surgical procedure of choice for a thyroglossal duct cyst is the Sistrunk procedure. A lingual thyroid is the most common site for functioning ectopic thyroid tissue, and has a 1% chance of harboring malignancy.

THYROID CANCER AND THE SURGICAL MANAGEMENT

The incidence of thyroid cancer in the United States and worldwide has increased 300% over the past three decades, due predominantly to an increase in PTC.[63,64] Differentiated thyroid cancer, arising from thyroid follicular cells, includes PTC, follicular thyroid cancer (FTC), and Hurthle cell cancer (HCC). Of the differentiated cancers, PTC comprises approximately 85% to 90% of cases compared with 12% with follicular histology; less frequent is HCC, and even less common is poorly differentiated cancer. PTC has the best prognosis, with an estimated 5-year survival rate of 98.1% overall, 99.9% for localized disease, and 55.5% for distant disease.[64] Recurrence, metastases, and death may occur in 10% to 15% of patients with more aggressive variants of PTC, such as tall cell variant, columnar cell variant, and hobnail variant,[26] whereas a diagnosis of anaplastic thyroid cancer (ATC) tends to be grim.

Papillary Thyroid Cancer

The surgical treatment of PTC is best divided into two groups based on the size, clinical characteristics, and aggressiveness of these lesions: papillary microcarcinoma (PMC) and standard PTC.

Papillary Microcarcinoma. PMC refers to a small papillary cancer, less than 1 cm in diameter. PMC does not show invasion through the thyroid capsule, is not associated with lymph node metastases, and tends to be found in a young person as an occult lesion when thyroidectomy has been performed for another benign condition, or incidentally on high-resolution ultrasound. The options for very low–risk disease are active surveillance (AS) of the lesion with close clinical and sonographic follow-up or thyroid lobectomy. The rationale for offering AS is related to the indolence of PTC, as demonstrated by two landmark clinical trials that reported similar outcomes in patients with biopsy-proven PMC that was not surgically removed and were followed for up to 15 years.[65,66]

Standard Treatment of Most Papillary Carcinomas. The American Thyroid Association (ATA) guidelines outline that surgery is the standard of care for most PTCs greater than 1 cm.[26] The most recent 2015 ATA guidelines have changed the way we approach low-risk thyroid cancer and the extent of surgery patients are offered. Well-differentiated thyroid cancer (WDTC) that is unilateral, is less than 4 cm in size, and has no high-risk features can be managed with lobectomy alone.[26] Additional low-risk features include the absence of aggressive histology (hobnail variant, tall cell variant, columnar cell carcinoma), and no evidence of vascular invasion.[26] Intermediate-risk disease has microscopic invasion into the perithyroidal soft tissue, aggressive histology, vascular invasion, and clinical N1 disease with more than five positive lymph nodes (<3 cm in largest dimension).[26] High-risk disease has gross extrathyroidal extension into the surrounding tissue, incomplete tumor resection, distant metastases, and N1 disease with lymph nodes greater than 3 cm in the largest dimension.[26] Universal total thyroidectomy followed by RAI is no longer the standard approach for all PTCs. The best treatment for PTC is more nuanced and may include lobectomy or total thyroidectomy, with appropriate CND or lateral neck dissection (LND) when nodes are involved or whether RAI therapy is required postoperatively.[26] For high-risk disease, bilateral thyroid resection improves survival and reduces recurrence rates compared with unilateral resection,[26] whereas low-risk disease can often be treated with lobectomy alone.[26]

Follicular Thyroid Cancer

FTC is classified under the umbrella of WDTC and is derived from follicular epithelium. According to the Surveillance, Epidemiology, and End Results Program (SEER) data, the 1980 to 2009 FTC incidence was 1.19 per 100,000 woman-years and 0.55 per 100,000 man-years.[67] This is much lower than the incidence of PTC, at 9.21 per 100,000 woman-years and 3.10 per 100,000 man-years.[67]

FTC differs considerably from PTC in terms of epidemiological, cytological, and ultrasound features, as well as clinical behavior. FTC is less common than PTC, patients tend to present at an older age, and it is more common in females (3:1).[67] Tumor multicentricity and local lymph node spread are less common than in PTC. Metastatic spread of tumor occurs by hematogenous dissemination to the lungs, bones, and other peripheral tissues, compared with the lymphatic spread seen in PTC.

For patients with minimally invasive FTC (which includes tumors with invasion of only the capsule of the tumor without vascular invasion), ipsilateral lobectomy is sufficient.[26,68] However, for most patients with invasive FTC with gross capsular invasion (with or without surrounding soft tissue invasion) or vascular invasion (fewer than four foci present is minimally invasive, and more than four foci present is extensive vascular invasion), the ideal operation is like that for PTC, although the rationale for its performance differs.[26,68] Near-total or total thyroidectomy should be performed, not because of multicentricity but rather to facilitate later treatment of metastatic disease with RAI.[69] Remnants of normal thyroid in the neck are ablated by RAI, and if peripheral metastases are detected they should be treated with high-dose RAI therapy. Although lymph node metastases in the lateral region of the neck are not commonly found, a unilateral CND and a selective LND should be performed if metastatic nodes are identified.

Finally, all patients with papillary or follicular cancer should be treated for life with levothyroxine therapy in sufficient doses to suppress TSH to the appropriate level.[69] Careful follow-up is required to monitor for cardiac or other problems, such as osteoporosis from thyrotoxicosis.

Noninvasive Follicular Thyroid Neoplasm with Papillary-Like Nuclear Features

Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) is a diagnosis that can only be made on surgical pathology, and not on FNA cytology. When thyroid lobectomy is performed, and the result is NIFTP, no further surgery or postoperative treatment is required. In a case series, Rosario et al. reported no evidence of recurrence after 1 to 10 years of follow-up among patients who received lobectomy for tumors greater than 1 cm.[70] The literature demonstrates that, when NIFTP is diagnosed, less than 1% of tumors will recur within the first 15 years, whereas most differentiated carcinomas are at risk of relapsing within the first decade after initial therapy.[42] Thus, no further surgery is indicated. The diagnosis of NIFTP is an important one to make, as it can alleviate the psychological impact of having a cancer diagnosis and can save the healthcare system thousands of dollars per patient ($5000 to $8500 per patient in the United States) by avoiding unnecessary treatment with RAI.[71]

Hurthle Cell Thyroid Cancer

Hurthle cell tumors were once thought to be variants of follicular neoplasms; however, the most recent edition of the World Health Organization classification of tumors designates HCC and FTC as separate entities.[72] Although the incidence of thyroid cancer, namely PTC, is on the rise, the rates of HCC have remained stable and represent approximately 5% of differentiated thyroid carcinomas.[72]

Hurthle cell tumors tend to behave more aggressively, have a higher rate of locoregional recurrence, and are more difficult to treat than the usual follicular neoplasms for several reasons: 1) the incidence of carcinoma varies from 5.3% to 62% in different clinical series; 2) benign-appearing tumors can later metastasize in up to 2.5% of patients; and 3) HCCs are far less likely to concentrate RAI than are the usual follicular carcinomas, which makes treatment of metastatic disease difficult. Given the heterogeneity and unpredictability in presentation and prognosis, the best treatment approach remains controversial. Our group typically advocates a more aggressive surgical approach, including total thyroidectomy with or without compartmental lymphadenectomy.[26,73]

Medullary Thyroid Cancer

Medullary thyroid cancer (MTC) accounts for 5% to 8% of all thyroid cancers. It is a C-cell, calcitonin-producing tumor that contains amyloid or an amyloid-like substance, as well as other peptides and amines such as carcinoembryonic antigen, serotonin, neurotensin, and a high-molecular weight adrenocorticotropic hormone-like peptide. These substances may result in a carcinoid-like syndrome with diarrhea or Cushing syndrome, especially when widely metastatic tumor is present. MTC spreads to the lymph nodes of the neck and mediastinum and later disseminates to the lungs, liver, and elsewhere. The tumor is radioresistant, does not take up RAI, and is not responsive to thyroid hormone suppression. Hence, a thorough surgical approach is mandatory.

Most MTC of the thyroid is sporadic (approximately 75%), but it can also be transmitted in a familial pattern in 25% of cases. This tumor or its precursor, C-cell hyperplasia, occurs as a part of multiple endocrine neoplasia type 2A (MEN2A) and type 2B (MEN2B) syndromes or, more rarely, as part of familial MTC syndrome (see Table 82.2).[74] MEN2 syndromes are transmitted in an autosomal dominant manner caused by mutations of the *RET* oncogene on chromosome 10.[75] All patients with MTC should be screened for hyperparathyroidism and pheochromocytoma.[75] The risk of these two disease states accompanying MTC in MEN2A is greatest in patients with a 630 or 634 *RET* mutation.[75] Family members, especially children, of a patient with MTC of the thyroid should also be screened for MTC if the patient has MEN2 with a *RET* oncogene mutation, but also if the tumor is bilateral or if C-cell hyperplasia is present.[76]

The standard treatment for patients with MTC is a total thyroidectomy with a prophylactic CND. LND is dictated by the presence of biopsy-proven disease in the lateral neck. A thorough surgical approach is paramount to achieving a cure. Recent evidence shows that preoperative calcitonin values can help determine the extent of lymph node dissection.[77] The risk of lymph node metastases is low when calcitonin levels are less than 20 pg/mL (the normal reference range is <10 pg/mL).[77,78] Calcitonin levels higher than 20 pg/mL, 50 pg/mL, 200 pg/mL, and 500 pg/mL are associated with the presence of lymph node metastases in the ipsilateral central and ipsilateral lateral neck, in the contralateral central neck, in the contralateral lateral neck, and in the upper mediastinum, respectively.[77,78]

Prevention of the development of MTC is best achieved by prophylactic thyroidectomy in young children who are found by genetic screening to have a mutated *RET* oncogene. Further reading of a comprehensive set of revised guidelines regarding diagnosis, screening, surgical, and medical treatment, follow-up of MEN2A and 2B, their pheochromocytomas, and hyperparathyroidism, as well as sporadic medullary cancer, is recommended for greater depth on the topic.[79]

Generally, patients with MEN2A syndrome have a better prognosis than do those with a sporadic tumor, whereas patients with MEN2B syndrome have the most aggressive tumors and rarely survive to middle age.[79] In children with MEN2B, screening for *RET* oncogene mutation should be done after birth, and with a mutated *RET* oncogene, total thyroidectomy should be done as early as possible within the first year of life in a tertiary care setting.[79] In MEN2B patients or infants and children older than 1 year of age, a prophylactic level VI neck dissection should be considered as well.[79] A multidisciplinary approach is recommended, with involvement of genetic counseling.

TABLE 82.2 Inherited Subtypes of Medullary Thyroid Cancer (RET oncogene)

MEN2A (*RET* oncogene)	MEN2B (*RET* oncogene)	Familial medullary thyroid cancer syndrome
Medullary carcinoma	Medullary carcinoma	Medullary thyroid carcinoma
Pheochromocytoma	Pheochromocytoma	Variant of MEN2A
Hyperparathyroidism	Hyperparathyroidism - unlikely	
	Ganglioneuroma phenotype	
	Mucosal neuromas	

Patients with recurrent or persistent MTC who may not be candidates for surgery may be eligible for new chemotherapeutic modalities like vandetanib or selpercatinib. A recent trial in patients with previously treated *RET*-mutant MTC showed a promising antitumor response with selpercatinib, whether or not they received previous first-line therapy with vandetanib, cabozantinib, or both[80] (Table 82.2).

Anaplastic Thyroid Cancer

ATC is a rare and lethal form of thyroid cancer that is initially derived from a differentiated tumor such as PTC or FTC. However, over time, ATC accumulates a series of mutations, making it a more undifferentiated phenotype and one of the most aggressive of all cancers in humans. It comprises 1.3% to 9.8% of all thyroid cancers globally.[81] In contrast to other WDTCs, anaplastic cancers are extremely aggressive, with a disease-specific mortality rate approaching 100%. The tumor grows rapidly, with a median survival of 5 to 6 months and a 1-year survival rate of 20%.[36,81] Often, ATC presents at an advanced stage, with symptoms of airway obstruction and vocal cord paralysis on initial evaluation. Up to 46% of newly diagnosed patients present with widely metastatic disease at diagnosis, and up to 90% of patients have invasion to the surrounding soft tissue.[36] Thus, surgical cure is unlikely with ATC, and patients will only benefit from palliative measures, no matter how aggressively surgery is pursued.

For ATC, it is paramount that an efficient multidisciplinary approach is coordinated. Flexible nasopharyngoscopy should be performed for airway assessment, as well as urgent imaging (CT with contrast, with or without MRI), with potential panendoscopy and bronchoscopy to assess for local soft tissue invasion of the trachea, esophagus, and potential distant metastases to the chest or abdomen.[82] A PET-CT scan is also necessary for identifying local and distant metastases. An FNA, core biopsy, or open biopsy is essential for making the diagnosis and to avoid confusion with other entities such as lymphoma. Tracheostomy should generally be avoided in patients with ATC.

Despite the advanced stage in presentation of ATC, there is optimism, with novel immunotherapies that have demonstrated improved outcomes and quality of life for these patients.[83] One group recently reported a *BRAF*-mutated ATC patient with end-stage, locally advanced, unresectable ATC who was effectively treated with combination therapy with dabrafenib (a selective BRAF inhibitor), trametinib (a selective MEK inhibitor), and pembrolizumab. Following induction of therapy with this three-drug combination, the tumor shrunk significantly, allowing for complete surgical resection and chemoradiation with prolonged survival.[83] Large clinical trials are in progress; however, this case report represents a promising treatment modality for *BRAF*-mutated ATC.

Radioactive Iodine Therapy

After surgery for WDTC, RAI therapy with ^{131}I is commonly used in order to ablate any remaining normal thyroid remnant present in the thyroid bed, as adjuvant therapy (to destroy suspected but unproven metastatic disease), or to target known persistent disease.[84] However, it is paramount to achieve complete surgical resection of macroscopic disease rather than relying on RAI to postoperatively "mop up" unresected residual disease for WDTC.

Contrary to previous practice, RAI ablation is more selectively being used postsurgery, especially in low-risk patients. RAI does not tend to be as beneficial in low-risk patients with PMC or in patients with WDTC 1 to 4 cm in size and no other worrisome features.[84] As more evidence emerges, there has been an increasing shift away from the routine use of RAI in low- to intermediate-risk patients and for selective use based on interoperative findings, imaging, and pathology. Even the dose of RAI has been questioned, leading the ATA guidelines to recommend use of a lower dose of RAI in low- and intermediate-risk disease.[26,85] Studies have shown that the use of high-dose RAI (i.e., 100 mCi) has no advantage for remnant ablation over a lower dose of 30 to 50 mCi for low- to intermediate-risk disease and is associated with greater side effects.[84,86]

KEY POINTS

- Active Surveillance can be offered to patients with very low-risk papillary microcarcinoma. Thyroid Lobectomy is an option for well-differentiated thyroid cancer that is unilateral, less than 4 cm in size with no evidence of high-risk features. Follicular and Medullary thyroid cancer are approached with more aggressive surgical intervention.

INDICATIONS FOR SURGERY OF THE THYROID GLAND

Indications for Lobectomy versus Total Thyroidectomy

The indications for lobectomy and total thyroidectomy are derived from the recommendations of the ATA guidelines.[26]

The indications for lobectomy include:

a. Low-risk PTC (WDTC <2 cm, no extrathyroidal invasion, no locoregional metastases, and lack of aggressive histology or aggressive genetic markers)
b. Unifocal WDTC less than 4 cm with no other locoregional disease or clinical risk factors (family history of thyroid cancer, or history of radiotherapy to the head and neck)
c. Benign symptomatic thyroid disease with single-lobe pathology, i.e., compressive symptoms
d. Diagnostic purposes (i.e., follicular neoplasm or undetermined pathology on FNA).

The indications for total thyroidectomy include:

e. Unifocal disease greater than 4 cm, or malignant disease in bilateral lobes
f. Need for postoperative surveillance (with Tg) or further treatments (with RAI) requiring complete excision of thyroid gland
g. Locoregional or metastatic disease
h. Benign symptomatic disease such as Graves disease refractory to medical management, or compression along TEG or major vasculature
i. If the patient has a less than 4-cm malignancy that is considered low-risk and wishes to have a near-total or total thyroidectomy, so they are completely cured of their disease and undergo one operation instead of two, the guidelines endorse this approach.

Indications for Active Surveillance

AS for low-risk WDTC is an alternative to immediate surgical resection for patients with PMC (<1 cm), or indolent disease less than 2 cm. AS of a malignancy involves close follow-up both diagnostically and clinically, in lieu of immediate surgery, with the intention of treatment with curative intent if the disease progresses or if the patient wishes to have surgery.[92] Even though the data for AS demonstrate a safe and effective alternative, most patients would prefer to have surgical resection.[26,65,87] However, not all patients are eligible for surgery, and in the right setting, AS is a reasonable option.

Indications for Remote Access Surgery

While the transcervical approach has excellent safety, neck scarring at the incision site may have a detrimental effect on quality of life. Though most patients are satisfied with the appearance of their postthyroidectomy scar, many younger patients report feeling self-conscious of their scar.[104-106] Furthermore, while most patients experience adequate wound healing, patients with darker skin are significantly more likely to develop keloids or hypertrophic scars. Improved cosmesis was therefore the principal motivation for the development of remote-access techniques for thyroidectomy, with the intention of avoiding a visible scar and ultimately creating "scarless" surgery.

Indications for Prophylactic Central Neck Dissection

For low-risk and intermediate-risk patients with small (T1 or T2), noninvasive, cN0 disease, generally a prophylactic CND is not done unless there is aggressive PTC histology, such as that presented by tall cell, diffuse sclerosing, and insular variants.[15,26] Prophylactic CND may increase the risk of permanent hypoparathyroidism and confer little to no survival advantage to the patient or improvement in long-term outcome.[15,28,88] However, for patients with advanced primary tumors (T3 or T4) or risk factors that may increase the risk of recurrence (such as older or very young age, larger tumor size, multifocal disease, extrathyroidal extension, or known lateral neck metastases) or recommendations by the multidisciplinary team, a prophylactic CND may be considered.[15,26,28]

There is a role for bilateral prophylactic CND with total thyroidectomy in MTC, given the high risk for metastasis in the central neck at the time of presentation, according to the National Comprehensive Cancer Network guidelines and the ATA guidelines.[90,91]

Indications for Therapeutic Central Neck Dissection

The role for therapeutic CND (levels VI and VII) in suspected or biopsy-proven nodal disease (cN1) is well accepted.[26] A distinction should be made from microscopic central nodal disease, which does not carry the same recurrence risk as macroscopic clinically detectable disease.[93] Once lymph node involvement is noted in the lateral neck, over 80% of patients have involvement in the paratracheal region.[15] Thus, therapeutic CND should accompany total thyroidectomy when macroscopic disease is present, either ipsilaterally or bilaterally. A CND can be bilateral or unilateral and refers to comprehensive removal of the pretracheal and prelaryngeal lymph nodes, along with at least one paratracheal nodal basin.[15] Generally, a unilateral CND is preferred when the disease is confined to one paratracheal region, to minimize the risk to bilateral RLNs and parathyroid glands.

Indications for Lateral Neck Dissection

There is a role for therapeutic LND in biopsy-proven nodal disease; however, routine prophylactic LND has not been proven to improve survival.[23] Therapeutic surgical resection of the lateral neck can reduce the rate of recurrence and mortality.[23] Two studies of the large SEER database found a poorer overall survival in patients with follicular carcinoma and PTC over age 45 years, with cervical lymph node metastases or distant metastases, and with larger tumor size.[26,93]

Lateral neck lymph node involvement occurs in up to 40% of patients with thyroid cancer.[26] When macroscopic disease is involved with metastatic differentiated thyroid cancer, the procedure of choice is a selective neck dissection, and rarely is a modified radical neck dissection required. A selective neck dissection involves removal of less than all five nodal levels and is directed by the patterns of lymphatic drainage from the primary tumor, with preservation of the IJV, cranial nerve (CN) XI, and SCM,[23] whereas a modified radical neck dissection describes comprehensive removal of nodal metastases from level I to V, with preservation of one or more of the following nonlymphatic structures: IJV, CN XI, and SCM.[23]

Cervical metastases to the lateral neck compartment generally occur after central neck involvement, and commonly to levels II to V.[23,26] If macroscopic metastases are present, a selective neck dissection of levels IIa, III, IV, and Vb should be performed.[23] Rarely, the submental nodes (level I) are involved in metastatic thyroid cancer. Thus, unless evidence of macroscopic disease is present clinically or on imaging, routine dissection of levels I, IIb, and Va is not recommended.[23] Level Va is only necessary when US evaluation shows suspicious nodes in level V, as there is risk to injury of the accessory nerve with elective dissection of this compartment.[23] Similarly, level 2b resection is required if nodal disease exists in level 2a.[23] While uncommon, skip metastasis to the lateral neck can occur without evidence of central neck disease on imaging or clinical examination. It is important, however, that if lateral neck compartment dissection is considered, the ipsilateral central neck is also considered.[41]

KEY POINTS

- Therapeutic central neck dissection should accompany total thyroidectomy when macroscopic disease is present, either ipsilaterally or bilaterally. Routine prophylactic lateral neck dissection has not proven to improve survival. However, there is a role for therapeutic lateral neck dissection in biopsy-proven nodal disease.

TRANSORAL APPROACH TO THYROIDECTOMY

Although remote-access techniques have become a popular alternative to the transcervical approach to thyroidectomy in Asia, where many of these techniques were pioneered, they have not yet achieved widespread adoption in the United States.[107,108] This trend may shift with the popularization of the transoral endoscopic thyroidectomy vestibular approach (TOETVA), given its favorable outcomes both in Asian and North American cohorts.[108-110]

The early remote-access techniques, such as the anterior breast approach[111] and the axillary approach,[112] evolved to the axillo-breast approach,[113] which offered cosmesis and a low rate of complications but was technically challenging. The postauricular "facelift" approach sought to improve the technical challenges by relying on familiar dissection planes that are used in parotidectomy.[114,115] This approach offered a shorter distance to the thyroid and good postoperative cosmesis; however, the narrow working space impedes exposure of the contralateral thyroid lobe, meaning a second, contralateral postauricular incision is often necessary for total thyroidectomy.[116]

While the aforementioned remote-access techniques have demonstrated good cosmesis and outcomes, their operative difficulty, long routes to the central neck, extensive dissection, and unfamiliar dissection planes have slowed their adoption.[117] With the aim of developing a more minimally invasive approach, transoral thyroidectomy was experimentally performed via sublingual as well as combined sublingual and vestibular access in human cadavers and porcine models.[118,119] A significant risk of the sublingual approach is injury to the hypoglossal and sublingual nerves. To avoid this risk and the restricted freedom of movement through the sublingual port, Richmon et al. developed a transoral approach using three vestibular ports and the da Vinci robot without an incision through the floor of the mouth.[120] Three gingival-buccal incisions are made, at midline and on either side laterally, and carbon dioxide insufflation maintains a working cavity. This transoral vestibular approach offers superior visibility of the central compartment of the neck and, in contrast to other remote-access approaches, provides midline access to the thyroid.

In 2016, Anuwong et al. demonstrated success in performing this approach with conventional laparoscopic instruments and an ultrasonic device. This initial experience of TOETVA had median operative time of 115.5 minutes (range 75–300 minutes). Outcomes were comparable to conventional open thyroidectomy, with minimal complications. Out of 60 procedures, there were two cases of transient RLN injury and three cases of transient hypoparathyroidism. One patient had hematoma occurring 3 days postoperatively. No patients had permanent RLN injury or hypoparathyroidism.[109] Early experience with TOETVA in North America demonstrated similarly good outcomes.[110] Since these early reports, TOETVA has enjoyed widespread adoption globally.[121] To date, the largest North American series comparing TOETVA with the transcervical approach demonstrates similar safety and efficacy regardless of body mass index, with no difference in RLN injury, hypoparathyroidism, seroma, infection, or hematoma.[122]

More recently, transoral thyroidectomy using a robotic system has been explored. In 2018, Razavi et al. reported early experiences with transoral robotic thyroidectomy vestibular approach (TORTVA) using the da Vinci robot and three vestibular ports as initially described by Richmon et al. Comparing seven cases of TORTVA to 20 cases of TOETVA, Razavi et al. found comparable outcomes overall; however, the median operative time was significantly longer for TORTVA (322 vs. 188 min).[123] Similar to TOETVA, transoral robotic thyroidectomy (TORT) utilizes three vestibular incisions, but with an additional unilateral axillary incision for introduction of a trocar to produce countertraction. In their series of 24 cases of TORT, Kim et al. reported a mean operative time of 232 minutes. In three cases, there was permanent mental nerve injury; however, no other complications were reported.[124]

KEY POINTS

- Remote access techniques, such as transoral endoscopic thyroidectomy vestibular approach (TOETVA) demonstrates similar safety and efficacy to the open transcervical approach regardless of body mass index, with no difference in RLN injury, hypoparathyroidism, seroma, infection or hematoma.

OPEN SURGICAL TECHNIQUES

Lobectomy and Total Thyroidectomy

The descriptions of the surgical techniques outlined below are based on the principles employed by the authors at our institution.

In a patient with thyroid carcinoma undergoing a total thyroidectomy, the surgeon should approach the side of concern first, where the largest nodule with a presumed cancer diagnosis is present. Frozen section review of the index thyroid nodule by pathologists is typically unreliable for Bethesda stage 3 and stage 4 lesions and should not be utilized in most cases of thyroid pathology. The goal is to completely resect the pathology in question while protecting important structures, including the RLN and the parathyroid glands on the contralateral (uninvolved) side.

Intraoperative nerve monitoring is often used to assess the integrity of the RLN. It is important to maintain a stable vagal nerve and RLN stimulus throughout the case in order to assure adequate vocal fold function postoperatively. If the signal is weakened intraoperatively, the patient should be counseled preoperatively that staging the procedure is a possibility to prevent the need for tracheostomy. Tracheostomy in this patient population is not without risk, due to the potential for a large mediastinal space and exposure of the innominate artery. In cases where gross nerve integrity has been confirmed but physiologic activity is not noted, staging the contralateral thyroidectomy or CND may be appropriate.

An open thyroidectomy involves a patient placed in supine position with the neck extended. A transcervical incision is made approximately two fingerbreadths superior to the clavicle, within a normal skin crease on the neck for cosmesis. A 3- to 5-cm midline low collar incision is made and carried down through the subcutaneous tissue and platysma muscle, down to the strap musculature (the sternohyoid, sternothyroid, and thyrohyoid). The superior and inferior skin flaps are raised, and the strap muscles are divided along the midline raphe vertically before retracting them laterally to expose each thyroid lobe. The thyroid lobe of interest is dissected first, and blunt instrumentation with or without cautery is used to laterally retract the sternothyroid and sternohyoid musculature away from the thyroid fascia. Once the strap muscles are retracted, the thyroid lobe is bluntly dissected while staying close to the thyroid capsule and within the fascia plane to reduce the risks of inadvertent injury of surrounding structures. The thyroid at the inferior trachea is dissected at midline, in order to identify the trachea as well as to mobilize the lobe inferiorly over the trachea, and then superiorly to dissect the pyramidal lobe and Delphian nodes. The superior pole of the lobe is dissected and mobilized, including the adjacent cricothyroid and inferior constrictor muscle, where the trajectory of the EBSLN can be identified in 85% of cases.[95] The superior pole vessels are ligated adjacent to the thyroid lobe, rather than cephalad to it, to prevent damage to EBSLN. We identify the RLN at the mid thyroid lobe region and then proceed to dissect superiorly (anterograde) toward the nerve insertion point at the cricothyroid region, which serves as a stable landmark. The intraoperative nerve monitor is used as a guide and is meant to help the surgeon confirm and protect the RLN if any question regarding the anatomy or bleeding is encountered. We can also identify the ITA, which courses perpendicular to the nerve. Rather than ablate the artery, we tend to dissect bluntly deep to the ITA to identify the RLN. Once the parathyroid glands are isolated, this also allows for mapping of the RLN, as the superior parathyroid lies deep to the nerve, and the inferior parathyroid superficial to it. Often, a fat pad is identified first and preserved, as it is intimately associated with the parathyroid gland. Additionally, care is taken to leave an adequate blood supply to the parathyroid glands while moving the gland off the thyroid lobe. Once the RLN is preserved, the final step is to mobilize the thyroid off BL medially and rotate the gland over the trachea.

We are careful to identify any intracapsular parathyroid glands that may have been adherent to the thyroid capsule during dissection and/or whose vascular supply was from the thyroid gland itself. If intracapsular dissection of the parathyroid gland off the thyroid capsule was required, and the blood supply is compromised, then the parathyroid gland is placed in normal saline on ice for later autoreimplantation, once frozen section confirmation has been made that it is indeed parathyroid tissue. These glands, if spliced and minced, can be reimplanted into the rich vascular supply of the accessible, adjacent neck musculature, such as the SCM or strap muscle. Any parathyroid gland that appears devascularized can undergo autoreimplantation, particularly if there is concern with the integrity of the other three parathyroid glands during surgery. Generally, once the parathyroid gland reestablishes its new blood supply, the secretory function of the gland is preserved, and PTH function is maintained, depending on the presence of benign or malignant disease.[96]

In a total thyroidectomy, the contralateral lobe is dissected in a similar fashion. It is important to confirm RLN functioning on the ipsilateral side prior to moving to the contralateral lobe to avoid potential bilateral vocal cord immobility and the requirement for a tracheostomy. If weakness of the RLN is noted intraoperatively, staging the remaining lobectomy until RLN recovery is a reasonable option. A near-total thyroidectomy means that a very small amount of thyroid tissue is left on the contralateral side of the cancer to protect the parathyroid glands and RLN. The ultimate goal for the surgeon is completeness of resection, as residual tissue can affect Tg levels during follow-up, as well as the success of radioactive ablation. Just before closing the wound, careful hemostasis and visualization of important anatomic structures are mandatory for success. Some surgeons use energy devices to help with hemostasis, but this is based on surgeon preference and availability of hospital resources.

When closing the wound, the fascia of the strap musculature is reapproximated, leaving a space inferiorly for potential drainage. Surgical drains are not necessary in most cases, but may be used if a large amount of dead space is present. After verifying the absence of any further bleeding, the skin is closed with 3-0 vicryl subdermal sutures. Surgeon preference is the deciding factor for whether stitches or glue alone are used in skin closure. Stitches can be placed in a subcuticular fashion underneath the skin for closure of the wound, and/or tissue adhesive glue may be placed over the skin after everting the edges meticulously for final wound closure.

Retrograde Medial Approach to the Thyroid Dissection

Most often, surgeons approach the RLN through lateral dissection of the thyroid; however, a medial approach can be advantageous in cases with a nonrecurrent RLN, large tubercle of Zuckerkandl, extrathyroidal extension of cancer along the distal RLN segment (with exception of cricothyroid junction involvement), or large goiters.[8] Some of the advantages of this approach include early identification of the RLN, with no thyroid tissue left unresected at the cricothyroid junction.[8] If the tumor is in the isthmus and there is a concern of tumor spillage, then a lateral approach to the thyroid is advised.

Central Neck Dissection

The central neck compartment (levels VI and VII) can be accessed through the thyroidectomy incision. Once the thyroid gland has been exposed, and the carotid sheath contents have been lateralized, we begin by dissecting over the carotid artery inferiorly to the level of the innominate artery and superiorly to the level of the cricoid on the respective side. After identification of the RLN, careful dissection continues along the nerve, removing fibrofatty tissue between the course of the nerve and the trachea. On the right-hand side, lymph node tissue lies both superficial and deep to the nerve, and therefore the right RLN should be transposed for the fibrofatty tissue to be resected medially to the trachea. The posterior boundary of the compartment is the esophageal muscularis, and it is important to clean all fibrofatty tissue off this structure. In the rare case of invasive disease, the muscularis itself can be removed carefully to avoid creating a perforation through the underlying mucosa. The preoperative placement of a Maloney dilator or gastric tube may facilitate identification of the esophagus. Esophageal defects should be closed primarily while inverting the mucosa, such as with a Connell technique. The superior extent of the dissection leads to the ITA, the trunk which gives off blood supply to the superior parathyroid gland. Preservation of this trunk is paramount in order to preserve the vascular integrity to the superior parathyroid tissue. It is expected that the inferior parathyroid glands will be devascularized and likely removed during a central compartment resection, given the unpredictable location of the inferior glands compared to the superior parathyroid glands.[23] As such, they should be identified early if possible. If the parathyroid tissue is inadvertently removed, it can be clipped and removed from the final specimen and then reimplanted at the end of the case. Parathyroid tissue can be stored on sterile ice until the time of reimplantation. Due to the risk of hypoparathyroidism from central compartment resection and the inherent difficulty with identification of these glands once the dissection has begun, one may choose to harvest these inferior glands prophylactically if they are identified early in the dissection for later reimplantation.

Lateral Neck Dissection

LND incorporates levels IIA through VB in patients with evidence of lateral neck disease in these compartments on preoperative evaluation. The incision for our LND is an extension of the thyroidectomy incision, approximately to the level of the mid-SCM. When possible, the incision is placed within a normal skin crease for cosmesis rather than curving the incision superiorly and against relaxed skin tension lines. If level IIB is involved, the incision may be extended higher with a more vertical direction in order to gain adequate exposure.

The subplatysmal planes are raised superiorly to one to two fingerbreadths below the angle of the mandible (to prevent injury to the marginal mandibular nerve) and inferiorly to the level of the clavicle. The subplatysmal fascia is incised at the inferior border of the submandibular gland, and we identify the digastric tunnel in order to follow it laterally to the IJV. We can then also identify the hypoglossal nerve just posterior to the digastric tendon. The SCM is then incised medially so that fibrofatty tissue is elevated medially. We identify level IIA by tracing out the accessory nerve in a lateral to medial trajectory towards the IJV. This defines the boundaries of levels IIA and IIB, and we then proceed to dissect down to the floor of the neck using a hemostatic scalpel blade or other dissecting instrumentation. Gentle retraction of the accessory nerve can help facilitate the level IIB dissection if necessary. Once this compartment has been cleared of disease, the dissection proceeds in an inferior direction along the medial edge of the SCM to dissect the compartments in level II to V. The contents of the floor of the neck, including the phrenic nerve, the brachial plexus (located in level IV), and the cervical rootlets, are protected during dissection. Next, we identify and skeletonize the omohyoid, which transverses in a diagonal direction at the level III–level IV junction. The exposure allows us to safely retract the omohyoid during the rest of the inferior dissection and expose the inferior aspect of level IV. The transverse cervical vessels are preserved, and, as we move medially toward the IJV, the thoracic duct is identified if we are on the left side of the neck and the lymphatic duct on the right side. An energy device, ties, or clips can be used to seal the lymphatic channels associated with the thoracic duct that are close to the IJV to prevent a potential chyle leak. Clips may make subsequent imaging and follow-up more difficult, and we generally avoid their use. Once the IJV is identified inferiorly, we elevate the fibrofatty tissue medially over the carotid artery, vagus nerve, and IJV. Once medialized, we stay over the anterior edge of the internal jugular and remove levels IIA/B, as well as excise the tissue over the strap musculature to cleanly resect our specimen *en bloc*. It is important to evaluate the retrocarotid location for any occult lymphadenopathy (much more common in thyroid carcinomas) and to dissect any fibrofatty tissue in level VB behind the SCM. If level VA is involved, dissection of the accessory nerve through the compartment from the SCM to the trapezius is completed. Although this may create traction on the accessory nerve, careful dissection and gentle retraction will protect the nerve and keep it intact. We perform a Valsalva at the end of the procedure in order to intraoperatively identify a chyle leak or bleeding from a major vessel. We stimulate all nerves for final confirmation of physiologic function, including the vagus, accessory nerve, hypoglossal nerve, and phrenic, and then proceed to close the neck. We place a single 7FR Blake drain into the lateral neck for every dissection, sutured to the skin. Subdermal closure is achieved using 3-0 vicryl interrupted sutures, and the skin can be closed with monocryl in a subcuticular fashion. Dermal glue can be placed over the incision.

MINIMALLY INVASIVE TECHNIQUES

Radiofrequency Ablation Technique

Image-guided thyroid ablation procedures can be performed using either thermal energy (RFA, laser ablation [LA], high-intensity focused ultrasound, and microwave ablation) or chemical methods such as ethanol instillation.[125] All ablative techniques cause coagulation necrosis to the desired nodule; however, each technique differs in the type of energy applied and how the tissue is targeted. Compared with other thermal ablation techniques, RFA and LA are accepted worldwide and have been found to be effective.[126-128] Our group is most familiar with RFA, which is an outpatient image-guided thermal ablative procedure that is a potential alternative to surgery for treating symptomatic benign nodules,[129-143] autonomously functioning thyroid nodules,[144-147] early primary thyroid cancer,[129,148-151] or recurrent thyroid cancer[129,152-155] that is not amenable to surgery. This approach is attractive, because it eliminates the need for a general anesthetic, an incision, or removal of the thyroid lobe at all. Compared with surgery, RFA takes significantly less time, costs less, and is associated with shorter hospitalizations.[148,156,157] After treatment with RFA, the size of a nodule is reduced over time by 33% to 58% after 1 month, by 51% to 85% after 6 months, and by 93% after 4 years.[150]

At the time of writing this chapter, the application of thermal ablative techniques is rapidly evolving and starting to infiltrate practices in North America.

KEY POINTS

- Thermal ablation is gaining traction worldwide as an alternative to surgery for treating symptomatic benign nodules, autonomously functioning thyroid nodules, early low risk primary thyroid cancer or recurrent thyroid cancer that is not amenable to surgery.

SURGICAL COMPLICATIONS

Recurrent Laryngeal Nerve Injury

Generally, the risk of injury to the RLN is low (temporary paralysis 9.8%, permanent 2.3%) in experienced hands, but can be variable depending on whether malignant or benign disease is present.[98,99] Furthermore, there is no evidence to suggest that the addition of a CND to a total thyroidectomy for thyroid cancer increases the permanent nerve injury rates when performed in a high-volume center.[162] However, in the presence of a NRLN, the chance of injury increases to 13%.[97] In-office ultrasound or preoperative axial imaging during the preoperative work-up of a thyroid nodule can help exclude an aberrant subclavian artery to anticipate a NRLN. Additionally, attempted early stimulation of the distal vagus nerve can alert the surgeon to the presence or absence of a NRLN: if the vagus nerve does not stimulate beyond the midpoint of the ipsilateral thyroid lobe, surgeons should suspect early branching of the RLN.

Unilateral RLN injury results in unilateral vocal fold paralysis with glottic insufficiency. The voice becomes breathy, and there is risk for aspiration of liquids, given incomplete glottic closure. Temporizing measures such as in-office injection laryngoplasty can help close the glottic gap if the injury is temporary. If RLN injury is permanent (lasts >9–12 months after onset of injury), then considerations are made for medialization thyroplasty, which medializes the paralyzed vocal fold permanently. Bilateral RLN injury is more worrisome and can lead to airway difficulty if both vocal folds assume a midline position, leading to a narrow glottic inlet. This may, in turn, lead to tracheostomy placement or measures to support breathing.

Hypoparathyroidism

The incidence of hypocalcemia due to parathyroid gland devascularization or removal after total thyroidectomy is 3% to 52% for transient and 0.4% to 13% for permanent postthyroidectomy hypocalcemia.[100,101] Because of their small size, delicate blood supply, and intimate anatomic position with the thyroid gland, parathyroid glands are at risk of accidental removal, trauma, or being devascularized during thyroidectomy.[102] If a parathyroid gland is recognized to be ischemic or nonviable, then it may be morselized and autotransplanted into the ipsilateral SCM. The rate of temporary hypoparathyroidism is much higher with CND, occurring in 14% to 40% of cases, with permanent hypoparathyroidism in 1.2%.[23,161,162]

Symptoms of hypocalcemia include circumoral numbness, tingling of the fingers and toes, and intense anxiety occurring soon after surgery. Chvostek's sign, which is carpopedal spasm, can appear early on physical exam. These symptoms occur with calcium levels of less than 7.5 to 8 mg/dL.

The diagnosis of postoperative hypoparathyroidism can be made in the early postoperative period (1–6 hours postoperatively) by checking PTH and serum calcium levels.[101] The pre- and postoperative trends can be tracked to determine the gradient difference in serum PTH and calcium.[101] If serum calcium levels are less than 8 mg/dL, oral calcium carbonate supplementation (with 500 mg–1 g of elemental calcium) of 1250 to 2500 mg four times per day with or without active vitamin D supplementation (calcitriol) is recommended. The addition of 1,25-dihydroxyvitamin D (calcitriol) is preferred, because it is the active metabolite of vitamin D and facilitates absorption of oral calcium. Patients with severe symptomatic hypocalcemia are treated in hospital with 1 g (10 mL) of 10% intravenous calcium gluconate immediately and infused over several minutes. Then, if necessary, 5 g of this calcium solution is added to 500 mL of intravenous saline to run continuously, at 30 mL/hour. Patients must be admitted for monitoring of serum calcium levels and for any cardiac QT abnormalities with electrocardiogram. Most often, postthyroidectomy hypocalcemia is transient, and intravenous calcium can be tapered and transitioned to oral therapy as patients become asymptomatic.

Spinal Accessory Nerve Injury

Injury to the eleventh CN is a complication of LND, particularly when level IIb or Va is dissected. Rarely is injury caused by transection of the nerve; rather, up to 20% of patients experience temporary or permanent injury from retraction or handling of the nerve during dissection.[158] Unfortunately, such an injury causes significant morbidity, such that patients experience a debilitating shoulder droop and are unable to raise their arm above their shoulder.[23] Some patients experience "shoulder syndrome," where shoulder movement is limited, painful, and disabling.[23]

Greater Auricular Nerve Injury and Sensory Cervical Rootlets

The most frequent permanent complication from LND, particularly of level II, is numbness of the lateral neck and ear.[23] With time, the area of numbness can reduce in size, but tends to be permanent around the ear lobe.[23]

Other Cranial Neuropathies

During LND, the marginal mandibular branch is at risk if dissecting level I and even the upper region of level II. This branch can lie low in the neck as part of the investing fascia of the submandibular gland.[23] Weakness is most often temporary, but can be permanent if transected, and results in an asymmetrical smile with difficulty drinking from a glass.[23]

Other less common injuries encountered during LND in the level III to level IV region include: phrenic nerve injury, brachial plexus injury, cervical sympathetic trunk injury, and thoracic duct injury.[23] Care should be taken while dissecting near the anterior and middle scalene muscles, as the phrenic nerve (which controls the diaphragm) lies on the scalene muscle deep to the carotid sheath. In between the anterior and middle scalene muscles is the entrance of the brachial plexus (which controls motor and sensory innervation to the upper extremity) into the neck, which is invested by the deep cervical fascia and is rarely injured during neck dissection. Injury to the cervical sympathetic nerves can result in Horner syndrome, characterized by ipsilateral ptosis, miosis, and anhidrosis.[159] Care should be taken when dissecting in the carotid sheath, as the sympathetic nerves can be found posterior to the carotid artery at the level of the C1 to C4 vertebrae.

Chyle Leak

Care should be taken when dissecting the lower lateral compartment of the neck to avoid inadvertent injury to the left thoracic duct and the right lymphatic duct. The left thoracic duct is larger than the right lymphatic duct and is responsible for lymph drainage from the entire body, except for the right sides of the head and neck. The right lymphatic duct drains the right side of the thorax and the right upper extremity.[103] The thoracic duct and lymphatic duct carry chyle and transport lymph back into the bloodstream. On the left side, the thoracic duct ascends into the superior mediastinum, which appears grossly as a lymph vessel that passes forward and laterally to drain into the left subclavian vein or the IJV. After dissection in this region, careful observation during a Valsalva maneuver is done to rule out a potential leak within this area so that a leak, if present, can be dealt with at the time of neck dissection.

Even though a chyle leak related to LND for thyroid carcinoma is uncommon, one series describes a chyle leak in 5.2% of patients intraoperatively and 8.3%of patients postoperatively.[103] A chyle leak should be treated appropriately, or it can lead to severe fluid and electrolyte imbalances. A low-output chyle leak (<500 mL/day) can be managed conservatively with fasting or a low-fat diet, whereas high-output leaks may also require octreotide therapy and possibly surgical ligation of the thoracic duct.[103]

Cosmesis

Cosmesis, or the appearance of the scar, after transcervical thyroid surgery is an important component that can influence a patient's quality of life. The authors' goal for cosmesis is to plan all incisions within a normal skin crease so the scar blends in with the natural skin lines. If the patient is known to develop hypertrophic scars and keloids or would like to avoid a neck scar altogether, transoral endoscopic thyroid surgery can be offered. For patients undergoing neck dissections for more extensive disease, the thyroid incision can be extended into the lateral neck crease in order to ensure good cosmesis.

Complications Associated with Transoral Surgery

There are complications associated with the transoral vestibular approach to thyroidectomy that may not be present with open surgery. Temporary or permanent (less likely) mental nerve injury occurs in

0% to 3.5% of cases, causing a sensory deficit in the lower lip and chin region. Other complications that can occur while developing the vestibular skin flap include temporary bruising or skin dimpling over the midline lower chin, perforation of the skin flap during vestibular dissection if not bluntly approached, and tearing of the inferior labial frenulum if there is excessive midline port movement. Rarely is a carbon dioxide embolism encountered; however, it is possible if there is venous injury or if excessive carbon dioxide insufflation pressure is used.[160]

Complications Associated with Radiofrequency Ablation

For treatment of benign nodules, the overall complication rate is 2.11% (95% confidence interval [CI]: 1.15–3.06), whereas the overall complication rate for recurrent thyroid cancers is higher, at 10.98% (95% CI: 4.82–17.15).[155] No life-threatening complications are reported, and minor complications include pain, skin burn, and hematoma.[129] The most common major complication noted is voice change, with an incidence of 0.94% for benign nodules, and there is an exceedingly low incidence of permanent changes.[155] Nodule rupture is the second most common major complication after RFA treatment (0.17%, 4/2421).[155] Patients may present with sudden neck bulging and pain at the RFA site due to delayed hemorrhage; however, the incidence remains low.

KEY POINTS

- The potential complications that occur with the open transcervical approach include infection, bleeding, scarring, temporary, or less commonly permanent RLN injury, hypoparathyroidism, or other cranial neuropathies (greater auricular nerve or spinal accessory nerve injury) if a lateral neck dissection is performed.

REFERENCES

1. Sarkar S, Banerjee S, Sarkar R, et al. A review on the history of ‘thyroid surgery’. *Indian J Surg*. 2016;78:32–36.
2. Loyo M, Tufano RP, Gourin CG. National trends in thyroid surgery and the effect of volume on short-term outcomes. *Laryngoscope*. 2013;123:2056–2063.
3. Randolph GW, Dralle H, Abdullah H, et al. International Intraoperative Monitoring Study Group. Electrophysiologic recurrent laryngeal nerve monitoring during thyroid and parathyroid surgery: international standards guideline statement. *Laryngoscope*. 2011;121:S1–S16.
4. Asgharpour E, Maranillo E, Sanudo J, et al. Recurrent laryngeal nerve landmarks revisited. *Head Neck*. 2020;34:1240–1246.
5. Henry BM, Sanna B, Graves MJ, et al. The reliability of the tracheoesophageal groove and the ligament of Berry as landmarks for identifying the recurrent laryngeal nerve: a cadaveric study and meta-analysis. *BioMed Res Int*. 2017;4:1–11.
6. Yalcin B, Ozan H. Detailed investigation of the relationship between the inferior laryngeal nerve including laryngeal branches and ligaments of Berry. *J Am Coll Surg*. 2006;202:292–296.
7. Pelizzo MR, Toniato A, Gemo G. Zuckerkandle tuberculum: an arrow pointing to the recurrent laryngeal nerve (constant anatomical landmark). *J Am Coll Surg*. 1998;187:333–336.
8. Butskiy O, Chang BA, Luu K, et al. A systematic approach to the recurrent laryngeal nerve dissection at the cricothyroid junction. *J Otolaryngol Head Neck Sur*. 2018;47:57.
9. Shindo ML, Wu JC, Park EE. Surgical anatomy of the recurrent laryngeal nerve revisited. *Otolaryngol Head Neck Surg*. 2005;133:514–519.
10. Issa KF, Sidiki D, Naouma C, et al. Superior approach of recurrent laryngeal nerve: review of the literature. *J Thyroid Research*. 2019;2019. Article ID 5671816.
11. Henry BM, Vikse J, Graves MJ, et al. Variable relationship of the recurrent laryngeal nerve to the inferior thyroid artery: meta-analysis and surgical implications. *Head Neck*. 2017;39:177–186.
12. Bakalinis E, Makris I, Demesticha T, Tsakotos G, Skandalakis P, Filippou D. Non-recurrent laryngeal nerve and concurrent vascular variants: a review. *Acta Medica Academica*. 2018;47:186–192.
13. Lahey FH. Routine dissection and demonstration of recurrent laryngeal nerves in subtotal thyroidectomy. *Surg Gynecol Obstet*. 1938;66:775–777.
14. Sancho JJ, Pascual-Damieta M, Pereira JA, et al. Risk factors for transient vocal cord palsy after thyroidectomy. *Br J Surgery*. 2008;95:961–967.
15. Agrawal N, Evasovich MR. Indications and extent of central neck dissection for papillary thyroid cancer: an American Head and Neck Society Consensus Statement. *Head Neck*. 2017;39:1269–1279.
16. Cernea CR, Ferraz AR, Nishio S, Dutra A Jr, Hojaij FC, Medina Dos Santos LR. Surgical anatomy of the external branch of the superior laryngeal nerve. *Head Neck*. 1992;14:380–383.
17. Morton RP, Whitfield P, Al-Ali S. Anatomical and surgical considerations of the external branch of the superior laryngeal nerve: a systematic review. *Clin Otolaryngol*. 2006;31:368–374.
18. Cernea CR, Ferraz AR, Furlani J, et al. Identification of the external branch of the superior laryngeal nerve during thyroidectomy. *Am J Surg*. 1992;164:634–639.
19. Kaplan EL, Sugimoto J, Yang H, et al. Postoperative hypoparathyroidism: diagnosis and management. In: Kaplan EL, ed. *Surgery of the Thyroid and Parathyroid Glands*. New York: Churchill Livingstone; 1983:262–274.
20. Caron NR, Sturgeon C, Clark OH. Persistent and recurrent hyperparathyroidism. *Curr Treat Options Oncol*. 2004;5:335–345.
21. Kose E, Rudin AV, Kahramangil B, et al. Autofluorescence imaging of parathyroid glands: an assessment of potential indications. *Surgery*. 2020;167:173–179.
22. Park JH, Lee YS, Kim BW, et al. Skip lateral neck node metastases in papillary thyroid carcinoma. *World J Surg*. 2012;36:743–747.
23. Stack BC Jr, Ferris RL, Goldenberg D, et al. Review and statement regarding the anatomy, terminology and rationale for lateral neck dissection in differentiated thyroid cancer. *Thyroid*. 2012;22:501–508.
24. Tufano RP, Bishop J, Wu G. Reoperative central compartment dissection for patients with recurrent/persistent papillary thyroid cancer: efficacy, safety, and the association of the BRAF mutation. *Laryngoscope*. 2012;122:1634–1640.
25. Pai SI, Tufano RP. Central compartment neck dissection for thyroid cancer. Technical considerations. *ORL J Otorhinolaryngol Relat Spec*. 2008;70:292–297.
26. Haugen BR, Alexander EK, Bible KC, et al. American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2015;26:1. 2016.
27. Detweiler K, Elfenbein DM, Mayers D. Evaluation of thyroid nodules. *Surg Clin*. 2019;99:571–586.
28. Carty SE, Cooper DS, Doherty GM, et al. Consensus statement on the terminology and classification of central neck dissection for thyroid cancer. *Thyroid*. 2009;19:1153–1158.
29. Noureldine S, Olson MT, Agrawal N, et al. Effect of gene expression classifier molecular testing on the surgical decision making process for patients with thyroid nodules. *JAMA Otolaryngol Head and Neck Surg*. 2015;141:1082–1088.
30. Ahmadi S, Oyekunle T, Jian XS, et al. A direct comparison of the ATA and TI-RADS ultrasound scoring systems. *Endocr Pract*. 2019;25:413–422.
31. Santhanam P, Ladenson PW. Surveillance for differentiated thyroid cancer recurrence. *Endocrinol Metab Clin N Am*. 2019;48:239–252.
32. Yeh MW, Bauer AJ, Bernet VA, et al. American Thyroid Association statement on preoperative imaging for thyroid cancer surgery. *Thyroid*. 2015;25:3–14.
33. Kaplan SL, Mandel SJ, Muller R, et al. The role of MR imaging in detecting nodal disease in thyroidectomy patients with rising thyroglobulin levels. *Am J Neuroradiol*. 2009;30:608–612.
34. Hagenimana N, Dallaire J, Vallee E, et al. Thyroid incidentalomas on 18FDG-PET/CT: a metaolico-pathological correlation. *J Otolaryngology Head Neck Surg*. 2017;46:22.
35. Ross DS, Burch HB, Cooper DS, et al. American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.
36. Jameson JL, De Groot LJ, De Kretser, DM, et al. (Eds.). *Endocrinology: Adult and Pediatric*. 7th ed, Philadelphia, PA: Elsevier/Saunders; 2016.

37. Terris DJ, Seybt MW, Vaughters III RB. A new minimally invasive lingual thyroidectomy technique. *Thyroid*. 2010;20:1367–1369.
38. Vairaktaris E, Semergidis T, Christopoulou P, et al. Lingual thyroid: a new surgical approach–a case report. *J Cranio-Maxillo-Fac Surg*. 1994;22:307–310.
39. Atiyeh BS, Abdelnour A, Haddad FF, et al. Lingual thyroid: tongue–splitting incision for transoral excision. *J Laryngol Otol*. 1995;109: 520–524.
40. Genden EM, Kotz T, Tong CC, et al. Transoral robotic resection and reconstruction for head and neck cancer. *Laryngoscope*. 2011;121:1668–1674.
41. Haugen BR, Sawka AM, Alexander EK, et al. American thyroid association guidelines on the management of thyroid nodules and differentiated thyroid cancer task force review and recommendations on the proposed renaming of encapsulated follicular variant papillary thyroid carcinoma without invasion to noninvasive follicular thyroid neoplasm with papillary-like nuclear features. *Thyroid*. 2017;27:481–483.
42. Nikiforov Y, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma. A paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol*. 2016;2:1023–1029.
43. Krasner JR, Alyouha N, Pusztaszeri M, et al. Molecular mutations as a possible factor for determining extent of thyroid surgery. *J Otolaryngol Head Neck Surg*. 2019;51:1–7.
44. Burch HB, Burman KD, Cooper DS. 2011 survey of clinical practice patterns in the management of Graves' Disease. *J Clin Endocrinol Metab*. 2012;97:4549–4558.
45. Varadharajan K, Choudhury N. A systematic review of the incidence of thyroid carcinoma in patients undergoing thyroidectomy for thyrotoxicosis. *Clin Otolaryngol*. 2020;45:538–544.
46. Staniforth JUL, Edirimanne S, Eslick GD. Thyroid carcinoma in Graves' disease: a meta-analysis. *Int J Surg*. 2016;27:118–125.
47. Sridama V, Reilly M, Kaplan EL, et al. Long-term follow up study of compensated low dose. *N Engl J Med*. 1984;311:426.
48. Hollowell JG, Staehling NW, Flanders D, et al. Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): national health and nutrition examination survey (NHANES III). *J Clin Endocrinol Metab*. 2002;87(2):489–499.
49. Heggie K, Yeung M, Grodski S, et al. Total thyroidectomy for pressure symptoms in patients with Hashimoto's thyroiditis. *ANZ J Surgery*. 2018;88:359–362.
50. LaRiviere CA, Waldhausen JHT. Congenital cervical cysts, sinuses, and fistulae in pediatric surgery. *Surg Clin N America*. 2012;92:583–597.
51. Allard RH. The thyroglossal cyst. *Head Neck Surg*. 1982;5:134–146.
52. Foley DS, Fallat ME. Thyroglossal duct and other congenital midline cervical anomalies. *Semin Pediatr Surg*. 2006;15(2):70–75.
53. Peretz A, Leiberman E, Kapelushnik J, et al. Thyroglossal duct carcinoma in children: case presentation and review of the literature. *Thyroid*. 2004;14:777–785.
54. Sturniolo G, Moleti M, Violi MA, et al. Prevalence of thyroglossal duct cyst carcinoma in adults having surgery for thyroglossal duct cysts. *Thyroid*. 2012;22:1191–1192.
55. Rayess HM, Monk I, Svider PF, et al. Thyroglossal duct cyst carcinoma: a systematic review of clinical features and outcomes. *Otolaryngol Head Neck Surg*. 2017;156:794–802.
56. Szczepanek-Parulska, Zybek-Kocik A, Wartofsky L, Ruchala M. Thyroid hemiagenesis: incidence, clinical significance and genetic background. *J Clin Endocrinol Metab*. 2017;102:3124–3137.
57. Di Benedetto V. Ectopic thyroid gland in the submandibular region simulating a thyroglossal duct cyst: a case report. *J Pediatr Surg*. 1997;32:1745.
58. Cheng TC, Yu-Hsuan L. Lingual thyroid. *Ear Nose Throat J*. 2020;99:122–123.
59. Huang NS, Wei WJ, Qu N, Wang YL, Wang Y, Ji QH. Lingual ectopic papillary thyroid carcinoma: two case reports and review of the literature. *Oral Oncol*. 2019;88:186–189.
60. Kang HC. Lingual thyroid: marked response to suppressive therapy. *Thyroid*. 2004;14:401–402.
61. Prisman E, Patsias A, Genden EM. Transoral robotic excision of ectopic lingual thyroid: case series and literature review. *Head Neck*. 2015;37:E88–E91.
62. Pace-Asciak P, Higgins K. The management of an intrathoracic goiter. *Medical Research Archives*. 2018;6:1–6.
63. Seib CD, Sosa JA. Evolving understanding of the epidemiology of thyroid cancer. *Endocrinol Metab Clin N Am*. 2019;48:23–35.
64. Surveillance Research Program NCI. Fast stats: an interactive tool for access to SEER cancer statistics. Available at: https://seer.cancer.gov/faststats. Accessed October 19, 2020.
65. Ito Y, Miyauchi A, Kihara M, et al. Patient age is significantly related to the progression of papillary microcarcinoma of the thyroid under observation. *Thyroid*. 2014;24:27–34.
66. Sugitani I, Toda K, Yamada K, et al. Three distinctly different kinds of papillary thyroid microcarcinoma should be recognized: our treatment strategies and outcomes. *World J Surg*. 2010;34:1222–1231.
67. Aschebrook-Kilfoy B, Grogan RH, Ward MH, et al. Follicular thyroid cancer incidence patterns in the United States, 1980-2009. *Thyroid*. 2013;23:1015.
68. Lloyd RV, Osamura RY, Kloppel G, et al. *WHO Classification of Tumours: Pathology and Genetics of Tumours of Endocrine Organs*. 4th ed. Lyon: IARC; 2017.
69. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. *Am J Med*. 1994;97:418–428.
70. Rosario PW, Mourao GF, Nunes MB, et al. Noninvasive follicular thyroid neoplasm with papillary like nuclear features. *Endocr Relat Cancer*. 2016;23(12):893–897.
71. Goffredo P, Thomas SM, Dinan MA, et al. Patterns of use and cost for inappropriate radioactive iodine treatment for thyroid cancer in the United States: use and misuse. *JAMA Intern Med*. 2015;175:638–640.
72. WHO. *WHO Classification of Tumors of Endocrine Organs*. 4th ed. Geneva: World Health Organization; 2017.
73. Bishop J, Wu G, Tufano RP, et al. Histological patterns of locoregional recurrence in Hurthle cell carcinoma of the thyroid gland. *Thyroid*. 2012;22:690–694.
74. Sizemore GW, van Heerden JA, Carney JA. In: Kaplan EL, ed. *Medullary Carcinoma of the Thyroid Gland and the Multiple Endocrine Neoplasia Type 2 Syndrome*. Edinburgh: Churchill Livingstone; 1983:75.
75. Kloos RT, Eng C, Evans DB, et al. Medullary thyroid cancer: management guidelines of the American thyroid association. *Thyroid*. 2009;19:1295.
76. Machens A, Niccoli-Sire P, Hoegel J, et al. Early malignant progression of hereditary medullary thyroid cancer. *New Eng J Med*. 2003;349:1517–1525.
77. Viola D, Elisei R. Management of medullary thyroid cancer. *Endocrinol Metab Clin North Am*. 2019;48:285–301.
78. Machens A, Dralle H. Biomarker-based risk stratification for previously untreated medullary thyroid cancer. *J Clin Endocrinol Metab*. 2010;95:2655–2663.
79. Wells SA Jr, Asa SL, Dralle H, et al. Revised American thyroid association guidelines for the management of medullary thyroid carcinoma prepared by the American thyroid association guidelines task force on medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
80. Wirth L, Sherman E, Robinson B, et al. Efficacy of selpercatinib in RET-altered thyroid cancers. *N Engl J Med*. 2020;383:825–835.
81. SEER Cancer Statistics Review 1975-2009. http://seer.cancer.gov/csr/1975_2009_pops09/results_merged/sect_26_thyroid.pdf. Accessed on February 12, 2013.
82. Smallridge RC, Ain KB, Asa SL, et al. American Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid*. 2012;22:1104–1139.
83. Cabanillas ME, Ferrarotto R, Garden AS, et al. Neoadjuvant BRAF- and Immunie-directed therapy for anaplastic thyroid carcinoma. *Thyroid*. 2018;28:945–951.
84. Sia Y, Dave RV, Nour D, et al. Radioactive iodine ablation post differentiated thyroid cancer surgery: an analysis of use and impact of the American Thyroid Association guidelines. *ANX J Surg*. 2019;89:E502–E506.
85. Lepoutre-Lussey C, Deandreis D, Leboulleaux S, et al. Postoperative radioactive iodine administration for differentiated thyroid cancer patients. *Curr Opin Endocrinol Diabetes Obes*. 2014;21:363–371.

86. Maepaa HO, Keikkonen J, Vaalavirta L, et al. Low vs. high radioiodine activity to ablate the thyroid after thyroidectomy for cancer: a randomized study. *PLoS One*. 2008;3:e1885.
87. Sakai T, Sugitani I, Ebina A, et al. Active surveillance for T1bN0M0 Papillary thyroid carcinoma. *Thyroid*. 2019;29:59–63.
88. Carling T, Long WD, Udelsman R. Controversy surrounding the role for routine central lymph node dissection for differentiated thyroid cancer. *Curr Opin Oncol*. 2010;22:30–34.
89. Chen AY, Bernet VJ, Carty SE, et al. American Thyroid Association statement on optimal surgical management of goiter. *Thyroid*. 2014;24:181–189.
90. Wells SA, Asa SL, Dralle H, et al. Revised American thyroid association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
91. National Comprehensive Cancer Network. http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed October 30, 2021.
92. Sawka AM, Ghai S, Tomlinson G, et al. A protocol for a Canadian prospective observational study of decision-making on active surveillance or surgery for low risk papillary thyroid cancer. *BMJ Open*. 2018;8:e020298.
93. Randolph GW, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*. 2012;22:1144–1152.
94. Stang MT, Armstrong MJ, Ogilvie JB, et al. Positional dyspnea and tracheal compression as indications for goiter resection. *Arch Surg*. 2012;147:621–626.
95. Sinclair CF, Bumpous JM, Haugen BR, et al. Laryngeal examination in thyroid and parathyroid surgery: an American Head and Neck Society consensus statement. *Head Neck*. 2016;38:811–819.
96. Noureldine SI, Genther DJ, Lopez M, et al. Early predictors of hypocalcemia after total thyroidectomy: an analysis of 304 patients using a short-stay monitoring protocol. *JAMA Otolaryngol Head Neck Surg*. 2014;140:1006–1013.
97. Toniato A, Mazzarotto R, Piotto A, Bernante P, Pagetta C, Pelizzo MR. Identification of the non-recurrent laryngeal nerve during thyroid surgery: 20 year experience. *World J Surgery*. 2004;28:659–661.
98. Friedrich T, Hansch U, Eichfeld U, et al. Therapeutic management of postoperatively diagnosed bilateral recurrent laryngeal nerve paralysis. *Zentralblatt fur Chirurgie*. 2000;125:137–143.
99. Barczynski M, Konturek A, Cichon S. Randomized clinical trial of visualization versus neuromonitoring of recurrent laryngeal nerves during thyroidectomy. *Br J Surg*. 2009;96:240–246.
100. Vescan A, Witterick I, Freeman J. Parathyroid hormone as a predictor of hypocalciemia after thyroidectomy. *Laryngoscope*. 2005;115:2105–2108.
101. Al Khadem MG, Rettig EM, Dhillon VK, et al. Postoperative IPTH compared with IPTH gradient as predictors of post-thyroidectomy hypocalcemia. *Laryngoscope*. 2018;128:769–774.
102. Kaplan EL, Sugimoto J, Yang H, et al. Postoperative hypoparathyroidism: diagnosis and management. In: *Surgery of the Thyroid and Parathyroid Glands*. New York: Churchill Livingstone; 1983:262–274.
103. Roh JL, Kim DH, Park C. Prospective identification of chyle leakage in patients undergoing lateral neck dissection for metastatic thyroid cancer. *Ann Surg Oncol*. 2008;15:424–429.
104. Best AR, Shipchandler TZ, Cordes SR. Midcervical scar satisfaction in thyroidectomy patients. *Laryngoscope*. 2017;127:1247–1252.
105. Felix C, Russell JO, Juman S, et al. Cervical scar satisfaction post conventional thyroidectomy. *Gland Surg*. 2019;8:723.
106. Choi Y, Lee JH, Kim YH, et al. Impact of postthyroidectomy scar on the quality of life of thyroid cancer patients. *Ann Dermatol*. 2014;26:693–699.
107. Patel KN, Yip L, Lubitz CC, et al. The American Association of Endocrine Surgeons Guidelines for the definitive surgical management of thyroid disease in adults. *Ann Surg*. 2020;271:e21–e93.
108. Russell JO, Razavi CR, Garstka ME, et al. Remote-access thyroidectomy: a multi-institutional north American experience with transaxillary, robotic facelift, and transoral endoscopic vestibular approaches. *J Am Coll Surg*. 2019;228:516–522.
109. Anuwong A. Transoral endoscopic thyroidectomy vestibular approach: a series of the first 60 human cases. *World J Surg*. 2016;40:491–497.
110. Russell JO, Clark J, Noureldine SI, et al. Transoral thyroidectomy and parathyroidectomy–A North American series of robotic and endoscopic transoral approaches to the central neck. *Oral Oncol*. 2017;71:75–80.
111. Ohgami M, Ishii S, Arisawa Y, et al. Scarless endoscopic thyroidectomy: breast approach for better cosmesis. Surg Laparosc Endosc Percutaneous Tech. 2000;10:1–4.
112. Ikeda Y, Takami H, Niimi M, et al. Endoscopic thyroidectomy by the axillary approach. *Surg Endosc*. 2001;15:1362–1364.
113. Shimazu K, Shiba E, Tamaki Y, et al. Endoscopic thyroid surgery through the axillo-bilateral-breast approach. *Surg Laparosc Endosc Percutaneous Tech*. 2003;13:196–201.
114. Terris DJ, Singer MC, Seybt MW. Robotic facelift thyroidectomy: II. Clinical feasibility and safety. *Laryngoscope*. 2011;121:1636–1641.
115. Terris DJ, Haus BM, Nettar K, et al. Prospective evaluation of endoscopic approaches to the thyroid compartment. *Laryngoscope*. 2004;114:1377–1382.
116. Alabbas H, Ali DB, Kandil E. Robotic retroauricular thyroid surgery. *Gland Surg*. 2016;5:603.
117. Russell JO, Sahli ZT, Shaear M, et al. Transoral thyroid and parathyroid surgery via the vestibular approach—a 2020 update. *Gland Surg*. 2020;9:409.
118. Witzel K, Von Rahden BHA, Kaminski C, et al. Transoral access for endoscopic thyroid resection. *Surg Endosc*. 2008;22:1871–1875.
119. Wilhelm T, Harlaar J, Kerver A, et al. Transorale endoskopische thyreoidektomie. *Chirurg*. 2010;81:50–55.
120. Richmon JD, Holsinger FC, Kandil E, et al. Transoral robotic-assisted thyroidectomy with central neck dissection: preclinical cadaver feasibility study and proposed surgical technique. *J Robot Surg*. 2011;5:279–282.
121. Dionigi G, Chai YJ, Tufano RP, et al. Transoral endoscopic thyroidectomy via a vestibular approach: why and how? *Endocrine*. 2018;59: 275–279.
122. Russell JO, Razavi CR, Shaear M, et al. Transoral thyroidectomy: safety and outcomes of 200 consecutive North American cases. *World J Surg*. 2021;45:774–781.
123. Razavi CR, Khadem MGA, Fondong A, et al. Early outcomes in transoral vestibular thyroidectomy: robotic versus endoscopic techniques. *Head Neck*. 2018;40:2246–2253.
124. Kim HY, Chai YJ, Dionigi G, et al. Transoral robotic thyroidectomy: lessons learned from an initial consecutive series of 24 patients. *Surg Endosc*. 2018;32:688–694.
125. Mauri G, Pacella CM, Papini E, et al. Image-guided thyroid ablation: proposal for standardization of terminology and reporting criteria. *Thyroid*. 2019;29:611–618.
126. Ha EJ, Baek JH, Kim KW, et al. Comparative efficacy of radiofrequency and laser ablation for the treatment of benign thyroid nodules: systematic review including traditional pooling and bayesian network meta-analysis. *J Clin Endocrinol Metab*. 2015;100:1903–1911.
127. Trimboli P, Castellana M, Sconfienza LM, et al. Efficacy of thermal ablation in benign non-functioning solid thyroid nodule: a systematic review and meta-analysis. *Endocrine*. 2020;67:35–43.
128. Papini E, Monpeyssen H, Frasoldati A, et al. European thyroid association clinical guideline for the use of image-guided ablation in benign thyroid nodules. *Eur Thyroid J*. 2020;9:172–185.
129. Kim JH, Baek JH, Lim HK, et al. Thyroid radiofrequency ablation guideline: Korean society of thyroid Radiology. *Korean J Radiol*. 2018;19:632–655.
130. Lee M, Baek JH, Suh CH, et al. Clinical practice guidelines for radiofrequency ablation of benign thyroid nodules: a systematic review. *Ultrasonography*. 2021;40:256–264.
131. Garberoglio R, Aliberti C, Appetecchia M, et al. Radiofrequency ablation for thyroid nodules: which indications? The first Italian opinion statement. *J Ultrasound*. 2015;18:423–430.
132. National Institute for Health and Care Excellence. *Ultrasound Guided Percutaneous Radiogrequency Ablation for Benign Thyroid Nodules*Interventional procedure guidance [IPT562] [Internet] London: National Institute for Health and Care Excellence; 2016. Available from: https://www.nice.org.uk/guidance/ipg562.
133. Dobnig H, Zechmann W, Hermann M, et al. Radiofrequency ablation of thyroid nodules: “good clinical practice recommendations” for Austria : an interdisciplinary statement from the following professional

associations: Austrian thyroid association (OSDG), Austrian society for nuclear medicine and molecular imaging (OGNMB), Austrian society for endocrinology and metabolism (OGES), surgical endocrinology working group (ACE) of the Austrian surgical society (OEGCH). *Wien Med Wochenschr*. 2020;170:6–14.

134. Papini E, Monpeyssen H, Frasoldati A, et al. European thyroid association clinical guideline for the use of image-guided ablation in benign thyroid nodules. *European Thyroid Journal*. 2020;9:172–185.
135. Cho SJ, Baek JH, Chung SR, et al. Long-term results of thermal ablation of benign thyroid nodules: a systematic review and meta-analysis. *Endocrinol Metab (Seoul)*. 2020;35:339–350.
136. Lee GM, You JY, Kim HY, et al. Successful radiofrequency ablation strategies for benign thyroid nodules. *Endocrine*. 2019;64:316–321.
137. Jung SL, Baek JH, Lee JH, et al. Efficacy and safety of radiofrequency ablation for benign thyroid nodules: a prospective multicenter study. *Korean J Radiol*. 2018;19:167–174.
138. Chen F, Tian G, Kong D, et al. Radiofrequency ablation for treatment of benign thyroid nodules. A PRISMA-compliant systematic review and meta-analysis of outcomes. *Medicine (Baltim)*. 2016;95:e4659.
139. Cesareo R, Pasqualini V, Simeoni C, et al. Prospective study of effectiveness of ultrasound-guided radiofrequency ablation versus control group in patients affected by benign thyroid nodules. *J Clin Endocrinol Metab*. 2015;100:460–466.
140. Trimboli P, Castellana M, Sconfienza LM, et al. Efficacy of thermal ablation in benign non-functioning solid thyroid nodule: a systematic review and meta-analysis. *Endocrine*. 2020;67:35–43.
141. Deandrea M, Trimboli P, Garino F, et al. Long-term efficacy of a single session of RFA for benign thyroid nodules: a longitudinal 5-year observational study. *J Clin Endocrinol Metab*. 2019;104:3751–3756.
142. Sung JY, Baek JH, Jung SL, et al. Radiofrequency ablation for autonomously functioning thyroid nodules: a multicenter study. *Thyroid*. 2015;25:112–117.
143. Bernardi S, Giudici F, Cesareo R, et al. Five-year results of radiofrequency and laser ablation of benign thyroid nodules: a multicenter study from the Italian Minimally Invasive treatments of the Thyroid Group. *Thyroid*. 2020;30:1759–1770.
144. Cesareo R, Naciu AM, Iozzino M, et al. Nodule size as predictive factor of efficacy of radiofrequency ablation in treating autonomously functioning thyroid nodules. *Int J Hyperthermia*. 2018;34:617–623.
145. Cappelli C, Franco F, Pirola I, et al. Radiofrequency ablation of functioning and non-functioning thyroid nodules: a single institution 12-month survey. *J Endocrinol Invest*. 2020;43:477–482.
146. Cervelli R, Mazzeo S, Boni G, et al. Comparison between radioiodine therapy and single-session radiofrequency ablation of autonomously functioning thyroid nodules: a retrospective study. *Clin Endocrinol*. 2019;90:608–616.
147. Cesareo R, Palermo A, Benvenuto D, et al. Efficacy of radiofrequency ablation in autonomous functioning thyroid nodules. A systematic review and meta-analysis. *Rev Endocr Metab Disord*. 2019;20:37–44.
148. Zhang M, Tufano R, Russell JO, et al. Ultrasound-guided radiofrequency ablation versus surgery for low-risk papillary thyroid microcarcinoma: results of over 5 years follow up. *Thyroid*. 2020;30:408–417.
149. Zhang M, Luo Y, Zhang Y, et al. Efficacy and safety of ultrasound-guided radiofrequency ablation for treating low-risk papillary thyroid microcarcinoma: a prospective study. *Thyroid*. 2016;26:1581–1587.
150. Lim HK, Lee JH, Ha EJ, et al. Radiofrequency ablation of benign non-functioning thyroid nodules: 4-year follow-up results for 111 patients. *Eur Radiol*. 2013;23:1044–1049.
151. Xiao J, Zhang M, Zhang Y, et al. Efficacy and safety of ultrasonography-guided radiofrequency ablation for the treatment of T1bN0M0 papillary thyroid carcinoma: a retrospective study. *Int J Hyperthermia*. 2020;37:392–398.
152. Zhao Q, Tian G, Kong D, et al. Meta-analysis of radiofrequency ablation for treating the local recurrence of thyroid cancers. *J Endocrinol Invest*. 2016;39:909–916.
153. Chung SR, Baek JH, Choi YJ, et al. Longer-term outcomes of radiofrequency ablation for locally recurrent papillary thyroid cancer. *Eur Radiol*. 2019;29:4897–4903.
154. Park KW, Shin JH, Han BK, et al. Inoperable sympatomatic recurrent thyroid cancers: preliminary results of radiofrequency ablation. *Ann Surg Oncol*. 2011;18:2564–2568.
155. Chung SR, Suh CH, Baek JH, et al. Safety of radiofrequency ablation of benign thyroid nodules and recurrent thyroid cancers: a systematic review and meta-analysis. *Int J Hyperthermia*. 2017;33:920–930.
156. Che Y, Jin S, Shi C, et al. Treatment of benign thyroid nodules: comparison of surgery with radiofrequency ablation. *AJNR Am J Neuroradiol*. 2015;36:1321–1325.
157. Bernardi S, Dobrinja C, Fabris B, et al. Radiofrequency ablation compared to surgery for the treatment of benign thyroid nodules. *Int J Endocrinol*. 2014:934595.
158. Hillel AD, Kroll H, Dormon J, et al. Radical neck dissection: a subjective and objective evaluation of postoperative disability. *J Otolaryngol*. 1989;18:53–61.
159. Prim MP, De Diego JI, Verdaguer JM, et al. Neurological complications following functional neck dissection. *Eur Arch Oto-Rhino-Laryngol*. 2006;263:473–476.
160. Zhang D, Park D, Sun H, et al. Indications, benefits and risks of transoral thyroidectomy. *Best Pract Res Clin Endocrinol Metabol*. 2019;33:1–25.
161. Roh JL, Park JY, Park CI. Total thyroidectomy plus neck dissection in differentiated papillary thyroid carcinoma patients: pattern of nodal metastasis, morbidity, recurrence, and postoperative levels of serum parathyroid hormone. *Ann Surg*. 2007;245:604–610.
162. Chrisholm EJ, Kulinskaya E, Tolley NS. Systematic review and meta-analysis of the adverse effects of thyroidectomy combined with central neck dissection as compared with thyroidectomy alone. *Laryngoscope*. 2009;119:1135–1139.

83

Nonthyroidal Illness Syndrome

Eric Fliers and Anita Boelen

OUTLINE

Low Triiodothyronine States, 1376
Nonthyroidal Illness Syndrome with Low Serum Thyroxine, 1376
Serum Hormone Levels, 1377
  Serum Total and Free Triiodothyronine, 1377
  Serum Reverse Triiodothyronine, 1377
  Serum Total and Free Thyroxine, 1377
  Serum Thyroid-Stimulating Hormone, 1377
Thyroid Hormone Transport, 1378
Thyroid Hormone Metabolism and Action, 1378
Tissue- and Organ-Specific Responses in Nonthyroidal Illness Syndrome, 1379
  Hypothalamus, 1379
  Pituitary, 1379
  Liver, 1379
  Muscle, 1380
  Adipose Tissue, 1381
  Neutrophils and Macrophages, 1381
Mechanism of Altered Thyroid Hormone Metabolism in Nonthyroidal Illness Syndrome, 1381
  Downregulation of Hypothalamic and/or Pituitary Function, 1381
  Cyrokines, 1382
  Nuclear Factor κB, 1383
  Alternative Mechanisms, 1383
  Energy Status, 1383
Clinical Implications and Therapeutic Considerations, 1384
Conclusion, 1385

Since the 1970s, it has been known that serum thyroid hormone (TH) levels drop both during starvation and during illness. In mild illness, this involves only a decrease in serum triiodothyronine (T3) levels. However, as the severity and length of the illness increase there is a drop in both serum T3 and thyroxine (T4). This decrease of serum TH levels is seen in a variety of species and pathologies, including sepsis, myocardial infarction, and in fact probably any severe illness[1-3] (Table 83.1). The condition has been called the *euthyroid sick syndrome* or *sick euthyroid syndrome*. An alternative and probably more appropriate designation, which does not presume the metabolic status of the patient, is *nonthyroidal illness syndrome* (NTIS).

LOW TRIIODOTHYRONINE STATES

Food deprivation has profound effects on all levels of the hypothalamus–pituitary–thyroid (HPT) axis in a variety of species.[4] Starvation, and more precisely carbohydrate deprivation, appears to rapidly inhibit deiodination of T4 to T3 by type 1 iodothyronine deiodinase in the liver, thus inhibiting generation of T3 and preventing metabolism of reverse T3 (rT3).[5] This probably contributes to a drop in serum T3 and elevation of rT3. Because starvation induces a decrease in basal metabolic rate,[6] this decrease in TH levels has been interpreted as an adaptive response by the body to limit catabolism by inducing some degree of hypothyroidism. Likewise, a variety of illnesses induces decreased serum T3. Patients who exhibit only a drop in serum T3, which may occur during acute or mild disease, do not show clinical signs of hypothyroidism. Nor has it been shown that this decrease in serum T3 (in the absence of a drop in T4) has an adverse physiologic effect on the body.[1]

NONTHYROIDAL ILLNESS SYNDROME WITH LOW SERUM THYROXINE

As the severity of illness, and often associated starvation, progresses, there is the gradual development of a more extensive NTIS associated with low T3 and low T4 levels. Generally, thyroid-stimulating hormone (TSH) levels are low or normal despite the low serum hormone levels, and rT3 levels are normal or elevated. A large proportion of patients in an intensive care unit (ICU) setting have various degrees of NTIS, with low T3 and T4. A marked decrease in serum T3 and T4 in NTIS is associated with a high probability of death. For example, when serum T4 levels drop below 4 g/dL, the probability of death is approximately 50%, and with serum T4 levels below 2 g/dL, the probability of death reaches 80%.[7-9] Obviously, such associations do not prove a causal relationship, but they have certainly raised the consideration of TH treatment in the ICU setting.

Two major interpretations of NTIS can be found in the literature:

1. Although patients with NTIS are biochemically hypothyroid according to their serum levels, this reflects a physiologic response and does not exclude sufficient TH availability at the tissue level. Therefore, TH supplementation should not be considered in patients.
2. NTIS in patients with prolonged critical illness is a form of central hypothyroidism, and tissue hypothyroidism is present. As this is probably disadvantageous to the patient, therapy should be initiated if serum T4 levels are markedly decreased.

In this chapter, recent studies addressing changes at the tissue level during NTIS, as well as the mechanisms behind these changes, will be addressed. We will also summarize evidence for or against treatment of NTIS during prolonged critical illness and identify knowledge gaps that may guide the design of future clinical trials in this area.

TABLE 83.1 Illness-Induced Alterations in Serum Triiodothyronine (T3), Thyroxine (T4), Reverse Triiodothyronine (rT3), and Thyroid-Stimulating Hormone (TSH) in Various Species. ↓: decreased, ↑: increased, N: not affected, -: not reported.

Species	T3	T4	rT3	TSH	Ref
Human	↓↓	N or ↓	↑	N or ↓	(1)
Mouse	↓	↓↓	N or ↓	N	(11)(145)(32)
Rat	↓	↓	-	↓	(31)(58)
Rabbit	↓	↓	↓	N	(183)(174)

(From Boelen A, Kwakkel J, Fliers E. Beyond low plasma T3: local thyroid hormone metabolism during inflammation and infection. *Endocr Rev.* 2011;32:670–693.)

SERUM HORMONE LEVELS

Serum Total and Free Triiodothyronine

The majority of reports on NTIS indicate that total serum T3 levels are low.[10-13] Chopra and coworkers reported that free T3 levels were low[14] or, in a second report, often normal.[15] However, it is important to note that, in the second report, the patients with NTIS actually had average serum T4 levels that were above the normal mean and did not have significant NTIS. Sapin and coworkers compared free T3 levels found in patients with NTIS by direct dialysis, microchromatography, analog, two-step immune extraction, and a labeled antibody radioimmunoassay method.[16] Results were significantly below normal by five of the methods used and low in the most severe cases by one method. Faber and colleagues evaluated TH levels in 34 seriously ill patients, most of whom had low T4 and free T4 index values, and found generally normal free T3 and free T4 using an ultrafiltration technique.[17] A point to consider is that some ultrafiltration techniques fail to exclude TH-binding proteins from the filtrate and give spuriously high free hormone values.

Serum Reverse Triiodothyronine

Serum rT3 may be reduced, normal, or elevated during NTIS, and once elevated, rT3 does not increase much more in patients with more severe illness. Routine assessment of rT3 in patients with NTIS is not recommended, although it may help in distinct cases to differentiate between hypothyroidism and NTIS. Peeters and coworkers[18] found that in patients with NTIS, serum TSH, T4, T3, and the T3/rT3 ratio were lower, whereas serum rT3 was higher than in healthy subjects ($P < .0001$). Degradation of rT3 is reduced by decreased function of type 1 deiodinase, the enzyme that generates T3 and clears rT3. However, formation of rT3 is limited by the low level of substrate (T4) in serum and in tissues and perhaps by inhibition of T4 entry into cells. Thus, while it may be expected that rT3 should always be elevated, this is not the case, as it may be within the normal range. In contrast, starvation induces a fall in serum T3 and a rise in serum rT3 within 24 to 36 hours and a return to baseline upon refeeding.[5]

Serum Total and Free Thyroxine

Total serum T4 levels are reduced in NTIS in proportion to the severity and, probably, length of the illness.[10-12,15,19] In acute, short-term trauma such as cardiac bypass,[20] or in short-term starvation,[21] there is no drop in serum T4. However, with increasing severity and duration of trauma, illness, or infection, there is a drop in T4, which may become extreme. As indicated, serum T4 levels below 4 μg/dL are associated with a marked increased risk for death (up to 50%), and once T4 is below 2 μg/dL prognosis becomes extremely guarded. In neonates, low total T4 and TSH are associated with a greater risk for death and severe intraventricular hemorrhage. It has been suggested that TH supplementation might be a potential benefit in infants with the lowest T4 values.[13]

Total serum T4 is reduced, in part because of a reduction in thyroxine-binding globulin (TBG). A rapid drop in TBG to 60% of baseline within 12 hours after bypass surgery has been reported, possibly because of cleavage of TBG by protease, which causes TBG to lose its T4-binding activity.[22] Indeed, a cleaved form of TBG has been observed in the serum of patients with sepsis.[23] The impact of meningococcal sepsis on peripheral TH metabolism and binding proteins was studied in 69 children with meningococcal sepsis. All children had decreased total T3 and total T3/rT3 ratios, without elevated TSH. Lower total T4 levels were related to increased turnover of TBG by elastase. Thus, lowered TBG is a partial explanation for lower total T4 and T3 in NTIS.[24]

The results of free T4 assays in NTIS are method-dependent to some extent. Assays that include an estimate of TBG capacity to estimate free hormone typically return low values for calculated free T4 in NTIS. The free T4 level determined by dialysis varies widely, as does T4 measured by ultrafiltration[11-13,19,25]; the majority of reports are of normal or low values but, in some samples, elevated values.[11,12,26-28] Free fatty acids, if elevated to 2 to 5 mmol/L, can displace T4 binding to TBG and elevate free T4. Free fatty acids almost never reach such levels *in vivo*.[29,30] However, even small quantities of heparin (0.08 units/kg given intravenously, or 5000 units given subcutaneously), commonly given to patients in an ICU, can lead to *in vitro* generation of free fatty acids during extended serum dialysis and falsely augment apparent free hormone levels.[31] This is probably a widespread and serious problem, which explains many instances of apparently elevated free T4 levels in patients with acute illness. Of note, the subcutaneous administration of low molecular weight heparin was also reported to increase FT4 levels by up to 170% in specimens taken 2 to 6 hours after injection, leading to the advice to take specimens for FT4 estimation more than 10 hours postdose in patients receiving enoxaparin.[32] Results obtained using ultrafiltration compared with equilibrium dialysis are also variable but generally below the reference range. Thus, although free T4 is low in most assays that involve a correction for TBG levels, there is still some question as to the true free T4 in patients with NTIS. It is of interest that this problem does not carry over to estimates of free T3, which are depressed in most studies. There may be two reasons for this difference. Firstly, the depression of total T3 is proportionately greater than of total T4. Secondly, factors that affect TH binding are more apt to alter T4 assays than T3, as T4 is normally more tightly bound to TBG than is T3.

Serum Thyroid-Stimulating Hormone

Serum TSH in NTIS is typically normal or reduced and may be markedly low, although usually not less than 0.05 μU/mL.[1,10,11,19,25,33] It should be noted, however, that NTIS may have a major impact on pulsatile TSH secretion, which may fall to a large extent in patients with critical illness, as reported by Van den Berghe et al.[34] Therefore, in view of negative feedback regulation in the HPT axis, TSH levels are inappropriately low for the observed serum T4 and T3. Serum TSH in patients with NTIS may have reduced biological activity, perhaps because of reduced thyrotropin-releasing hormone (TRH) secretion with reduced TSH glycosylation as a result. Of interest, elevation of TSH above normal commonly occurs transiently if patients recover from NTIS.[35] This may indicate that these patients are recovering from a hypothyroid state, during which the ability of the pituitary gland to respond had been temporarily inhibited.

Patients treated in ICUs for a prolonged period of time show dramatically diminished TSH pulsatility, characterized by an absent nocturnal TSH surge and a decreased TSH pulse amplitude. Pulsatile TSH secretion was positively related to a low serum T3 level, in keeping with the notion that the reduced production of THs in prolonged critical illness may have, at least in part, a neuroendocrine origin.[36] Indeed, a postmortem investigation of TRH mRNA expression in the hypothalamic paraventricular nucleus (PVN) of patients with NTIS biochemically documented from serum samples taken less than 24 hours before death showed decreased TRH expression in patients with prolonged critical illness (Fig. 83.1). In fact, the decrease in total TRH mRNA expression in the PVN significantly correlated with the decrease in both serum TSH and T3, but not in T4. These patients had not been treated with glucocorticoids or dopamine.[37]

Further support for a major role for the hypothalamus in the resetting of the HPT axis in NTIS came from the observation that a continuous administration of TRH to patients with prolonged critical illness partially restored the serum concentrations of T3, T4, and TSH, especially when combined with growth hormone secretagogues.[38]

KEY POINTS

- During NTIS, TSH levels are inappropriately low for serum hormone levels reflecting altered HPT axis feedback regulation, partly resulting from reduced hypothalamic secretion of TRH.

THYROID HORMONE TRANSPORT

TH needs to be taken up in the cell by specific TH transporters like monocarboxylate transporter 8 (MCT8), monocarboxylate transporter 10 (MCT10), and organic anion transporting polypeptide family 1C1 (OATP1C1) to exert its action at the cellular level.[39] In the liver, NTIS affects TH transport at the level of the plasma membrane. Many years ago it was shown that decreased liver adenosine triphosphate and increased concentrations of compounds that inhibit T4 transport into the liver (bilirubin, nonesterified fatty acids) may contribute to reduced transport of T4 into the hepatocyte.[40] Using deiodination of T4 as an index of cellular transport of T4 into rat hepatocytes, Lim and coworkers[41] and Vos and colleagues[42] found that serum from patients with NTIS inhibited T4 uptake, elevated nonesterified fatty acids and bilirubin, and reduced albumin were considered to play a role. Inhibition of uptake of T4 into hepatocytes caused by sera of patients with NTIS also was observed by Sarne and Refetoff.[43] Recently, studies have been published about the effect of illness or inflammation on TH transporter expression in specific tissues, and it is clear that tissue TH transporter expression itself is also affected during NTIS. Interestingly, changes in liver TH transporter expression appear to differ between various animal models of NTIS; acute illness lowers Slc16A2 (MCT8) and Slc16A10 (MCT10) expression, whereas chronic inflammation results in decreased Slc16A2 mRNA expression and an increase in Slc16A10 mRNA. By contrast, lethal bacterial sepsis affects hepatic TH transport only marginally.[44] Tissue-specific changes will be discussed in more detail later.

In theory, reduced cellular TH uptake by itself would cause tissue hypothyroidism, which is generally not observed. Therefore, it is likely that reduced cellular uptake, if present, is just one of many factors involved in NTIS.

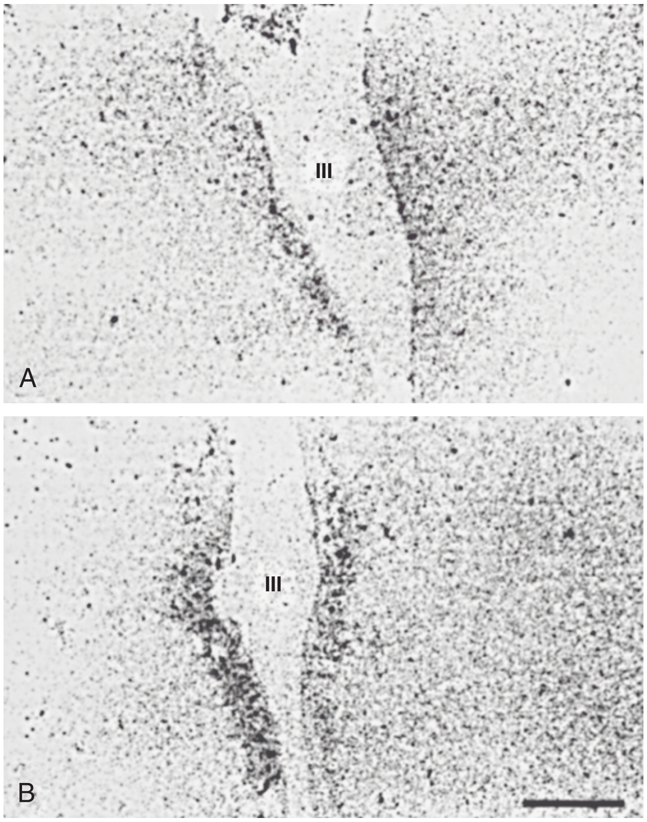


Fig. 83.1 *In situ* hybridization study demonstrating mRNA for TRH in the periventricular nuclei of (A) a subject who died after prolonged illness and NTIS and (B) a subject who experienced acute death. In patients with NTIS, mRNA for TRH is significantly reduced. *mRNA*, Messenger ribonucleic acid; *NTIS*, nonthyroidal illness syndrome; *TRH*, thyrotropin-releasing hormone. (From Lu B, Moser AH, Shigenaga JK, et al. Type II nuclear hormone receptors, coactivator, and target gene repression in adipose tissue in the acute-phase response. *J Lipid Res.* 2006;47:2179–2190.)

THYROID HORMONE METABOLISM AND ACTION

Intracellular TH is metabolized via iodothyronine deiodinases, a group of selenium cysteine-containing enzymes that are able to remove an iodine atom from the inner or outer ring of TH. Three types are known: type 1 (D1) catalyzes inner- and outer-ring monodeiodination, type 2 (D2) catalyzes only outer-ring deiodination (converting its preferred substrate T4 into T3), and type 3 (D3) catalyzes only inner-ring monodeiodination (degrading its preferred substrate T3 into T2, but also T4 into rT3, thereby inactivating TH).[45] Changes in type D1, D2, and D3 expression and activity in a variety or organs and tissues have been described in NTIS.[46]

T3 binds to TH receptors (TR) mediating genomic (canonical) and noncanonical actions.[47,48] Three T3-binding TRs have been described: TRα1, TRβ1, and TRβ2. TRα1 and TRβ1 are important in regulating TH target genes in peripheral tissues and are differentially expressed between organs. TR expression levels are also affected during illness.[46]

KEY POINTS

- Tissue TH concentrations during NTIS do not necessarily reflect serum concentrations and may decrease, remain unaltered, or even increase according to the organ and type of illness studied.

TISSUE- AND ORGAN-SPECIFIC RESPONSES IN NONTHYROIDAL ILLNESS SYNDROME

In contrast to a uniform whole-body response, there are wide variations in responses to TH supply and action in different organs and tissues, and between the response to acute illness and chronic illness. These multifactorial systems involve serum TH levels, TH transporters, deiodinase, and TR responses that are under different regulation in individual tissues.[46,49]

Hypothalamus

Not much is known about the effects of NTIS on hypothalamic TH transporters. Inducing critical illness in experimental animals induced increased MCT10 and OATP1C1 expression, while hypothalamic MCT8 mRNA expression remained unaltered. Because these changes were unrelated to local tissue T3 concentrations, their functional consequences remain unclear at present.[50]

Both D2 and D3 are expressed in the human hypothalamus. D2 is mainly expressed in glial cells of the infundibular nucleus and median eminence and in cells lining the third ventricle, while D3 is present in neurons of the PVN, supraoptic nucleus, and infundibular nucleus.[51] Because these studies were performed in postmortem tissue specimens, no data are available on TH metabolism in the human hypothalamus. Studies in rodents, however, have shown that hypothalamic D2 is a determinant of hypothalamic T3 production.[52] Furthermore, the administration of bacterial endotoxin (or lipopolysaccharide [LPS]), which is an established experimental model of acute NTIS, was reported to increase both D2 mRNA expression[53] and enzyme activity in the hypothalamus, specifically in tanycytes lining the third ventricle[54] (Fig. 83.2). Increased D2 activity is not specific for the acute phase of illness, as hypothalamic D2 activity is upregulated both in mice with bacterial sepsis[55] and in rabbits with critical illness.[50] Hypothalamic D3 is regulated in a reciprocal fashion: acute and chronic inflammation decrease D3 expression.[56,57] The observed D2 increase and D3 decrease probably result in net increased local T3 production, which may contribute to the decreased TRH mRNA expression in the PVN, as observed during illness both in animals and in humans.[57-59] An elegant study using an *in vitro* coculture system confirmed that LPS modulates glial D2 as well as neuronal D3 expression, resulting in altered T3 concentrations.[60] This has been confirmed by a recent study showing that the LPS-induced decrease of D3 in a neuronal cell line resulted in increased cellular T3 concentrations in the same cells.[61]

Hypothalamic TR mRNA expression by itself does not show dramatic changes during NTIS,[50] but changes in specific TR isoform expression in PVN neurons have not been carefully evaluated during NTIS. This could be of interest because the LPS-induced increase in hypothalamic D2 and decrease in hypothalamic D3 is exaggerated in $TR\beta^{-/-}$ mice compared with wild-type (WT) mice,[56] indicating that TRβ modulates the hypothalamic deiodinase responses to acute inflammation. In line with this, $TR\alpha^{0/0}$ mice do not display prominent changes in hypothalamic D2 and D3 responses to LPS.[62]

Very few studies have addressed hypothalamic T3 concentrations during NTIS directly. Mebis et al. showed that prolonged illness in rabbits does not increase hypothalamic T3 tissue concentrations, despite increased D2 expression, but these results should be interpreted carefully given the technical difficulties in selectively harvesting infundibular nucleus and PVN samples for the assessment of T3 concentrations.[50] Of note, in postmortem human hypothalamic specimens, T3 concentrations were even lower in NTIS patients compared with patients with acute death from trauma[63] (Table 83.2).

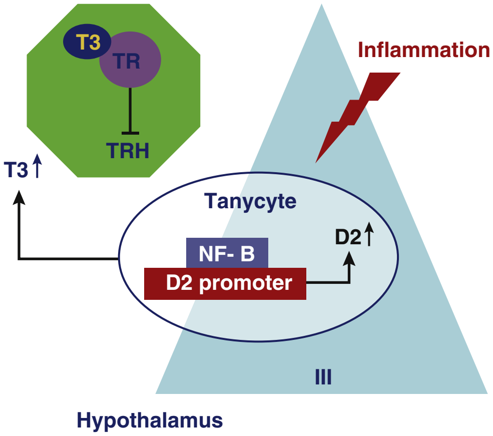


Fig. 83.2 Schematic model of hypothalamic thyroid hormone signaling during inflammation. Inflammation activates the nuclear factor (*NF*)-κB pathway in tanycytes, specialized cells lining the third ventricle (*III*). Tanycytes express type 2 deiodinase (*D2*), the main triiodothyronine (*T3*)-producing enzyme in the brain, the promoter of which contains NF-κB–responsive elements. Binding of NF-κB increases D2 expression and activity, and this stimulates the conversion of the prohormone thyroxine into T3. T3 will enter adjacent neurons and bind to neuronal thyroid hormone receptors, thereby regulating transcriptional activity of thyrotropin-releasing hormone (*TRH*). (Used with permission from Fliers E, Bianco AC, Langouche L, et al. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol.* 2015;3:816–825.)

Pituitary

Data on the pituitary gland are few. D2 levels have not shown consistent changes in animal models, whereas TRβ2 expression is reduced in acute NTIS.[64] In humans with fatal illness, pituitary T3 levels were low.[63] The overall picture is of central downregulation of the hypothalamus and pituitary gland and low levels of tissue T3 despite increments in transporter activity and D2 deiodinase. These findings fit with the observed correction of TSH and TH levels in human NTIS through administration of TRH.

Liver

Hepatic TH metabolism has been extensively studied during NTIS in both humans and experimental animals. The main TH transporters in liver are MCT8 and MCT10, and the expression of these transporters is affected during NTIS. In humans, MCT8 and MCT10 expression has been studied in postmortem liver biopsy samples from critically ill patients and compared with liver biopsy samples from patients after elective abdominal surgery (acute illness). Liver MCT8 expression was markedly increased in patients with prolonged illness compared with acute illness, while MCT10 expression did not differ. However, healthy controls were not available, which makes the study difficult to interpret.[65]

TH transporter expression was also studied in different NTIS animal models and shows differential changes; acute illness in mice lowers MCT8 and MCT10 expression, whereas chronic inflammation results in decreased MCT8 expression and an increase in MCT10. Lethal bacterial sepsis in mice affects hepatic TH transport only marginally. It remains questionable, however, whether the (sometimes small) alterations in liver TH transporter expression were instrumental in affecting liver TH concentrations.[44]

Both D1 and D3 are expressed in liver, although D3 is expressed only at very low levels in a healthy liver. During NTIS, D1 and D3 are heavily affected. Liver D1 is reduced during acute and prolonged

TABLE 83.2 Tissue T3 Concentrations in Nonthyroidal Illness Syndrome (nmol of T3/kg of Wet Weight)

	CONTROL GROUP			NTI GROUP	
Tissue	**Mean**	**SD**	**P**	**Mean**	**SD**
Cerebral cortex	2.2	0.9	<0.05	1.2	1.1
Hypothalamus	3.9	2.2	<0.01	1.4	1.2
Anterior pituitary gland	6.8	2.5	<0.005	3.7	1.1
Liver	3.7	2.3	<0.01	0.9	0.9
Kidney	12.9	4.3	<0.001	3.7	2.8
Lung	1.8	0.8	<0.01	0.8	0.5
Skeletal muscle	2.3	1.2	NS	>10.9	
Heart	4.5	1.5	NS	>16.3	

Most of the NTI patients received dopamine at some time in their ICU stay.

NS, Not significantly different; *NTI*, nonthyroidal illness; *T3*, triiodothyronine; *SD*, standard deviation.

(From Arem R, Wiener GJ, Kaplan SG, et al. Reduced tissue thyroid hormone levels in fatal illness. *Metabolism.* 1993;42:1102–1108.)

illness in both human and experimental animals.[53,66,67] Liver D3 is differentially regulated; D3 activity decreases during acute inflammation in mice[53] but was induced in the liver of severely ill patients.[18] Furthermore, liver D3 activity was positively correlated with serum rT3 concentrations and negatively correlated with serum T3 levels.[68] A rabbit model for prolonged critical illness showed identical clinical, biochemical, and endocrine features as seen in ICU patients,[69,70] and liver D3 activity tends to increase in these animals.[67] The response of D3 to illness may depend on energy status; fasting results in increased liver D3 activity in rodents,[71,72] and prolonged illness is associated with diminished food intake.[73] The meaning of the observed changes in liver deiodinase activity is currently unknown, as experiments performed with D3 knockout (D3KO) mice showed that D1 nor D3 contributed to the observed changes in serum TH concentrations during sepsis: the changes in TH levels were similar in D3KO and WT mice. In addition, basal liver D1 activity was relatively low in D3KO mice and not responsive to sepsis, suggesting that changes in liver D1 were also not critical for the decrease in serum TH concentrations.[55]

Both TRα and TRβ are expressed in liver, and mRNA expression of the TR is differentially affected during illness, depending on the severity and duration. TRα and TRβ decreased during acute illness in rodents[53,74] but did not change during chronic inflammation and bacterial sepsis in mice. Human data are not consistent: TRα and TRβ mRNA expression in liver biopsies from patients with chronic liver disease before liver transplantation both increased in association with low serum T3 and T4 levels.[75] In postmortem liver biopsies from ICU patients, the TRα1/TRα2 ratio increased with the severity of illness.[76] Chamba et al. studied hepatic TRα1, TRα2, and TRβ1 and T3 target gene expression in patients with a variety of chronic liver diseases and found no changes in TR or T3 target gene expression, despite reduced serum TH levels, indicating a euthyroid condition in the liver of these patients.[77] These findings are in line with the chronic inflammation mouse and rat models published to date and represent another example of a dissociation between serum TH levels and TH action at the organ level during NTIS. However, there is a caveat, as proper TR antibodies are lacking, which makes it impossible to study the effect of illness on TR protein expression.

In general, illness results in marked changes in TH transporter expression, deiodinase activity, and TR expression. These changes are associated with reduced metabolic activity in the liver, reduced T3 target gene expression, and lowered hepatic TH concentrations.[44,63,78,79]

Muscle

Muscle is an important T3 target tissue, as a variety of genes expressed in muscle are either positively or negatively regulated by T3. Local muscle thyroid accessibility and availability are regulated by the TH transporters MCT8 and MCT10 and D2 and D3, respectively. TRα1 is the major TR isoform in muscle, but TRβ1 is also expressed.[65,80,81] Local conversion of T4 by D2 is important in muscle function, as 40% of nuclear-bound muscle T3 in rat is derived from local production,[40] and even more in humans.[82,83]

Several studies have reported altered muscle TH metabolism during NTIS, both in acute and more chronic models.[46] In mice, muscle TH transporter expression is low during acute inflammation and bacterial sepsis and did not change during chronic inflammation. The same changes are also observed in patients; MCT8 expression is remarkably low in muscle tissue of patients during acute surgical stress compared with patients with prolonged illness.[65] Muscle MCT8 and MCT10 mRNA expression was also studied in a rabbit model of prolonged illness, and the findings differed from the human samples, as MCT8 mRNA expression did not change, while MCT10 markedly increased. THs are thought to play a role in the regulation of TH transporter status during NTIS because TH treatment in a rabbit model of NTIS returns the transporters to normal.[65]

During acute inflammation in mice, muscle D2 mRNA expression increases, while muscle D3 decreases,[84] and the changes in deiodinase expression coincide with decreased TR expression.[44] D2 mRNA expression and activity were also studied in skeletal muscle tissue of patients with acute surgical stress compared with healthy controls. Serum T3 levels decreased significantly during the surgical procedure, but D2 mRNA expression and activity in muscle did not change compared with controls.[85]

A few studies have been published describing the effect acute illness on T3 signaling in muscle tissue. The expression of *UCP3* (both mRNA and protein), a T3 target gene, increased in the skeletal muscle of mice during acute sepsis[86] and inflammation.[80] In agreement, UCP3 expression increased in the skeletal muscle of NTIS patients, coinciding with increased D2 mRNA expression, although a variety of other T3 target genes (*SERCA*, *MHC,* and *GLUT4*) did not change.[87] The clinical relevance of increased muscle T3 signaling is unclear at present. In contrast, chronic inflammation in an experimental mouse model results in simultaneously increased D2 and D3 activity.[88] This may result in a coordinated breakdown of both T4 and T3, ultimately leading to increased intracellular T2 levels. Both T3 and T2 are known to regulate mitochondrial activity in skeletal muscle.[89,90]

Bacterial sepsis lowers muscle D2 mRNA expression in mice[88] and in patients with septic shock, while D2 activity was hardly detectable in human skeletal muscle and did not differ from the controls undergoing surgery.[81] In contrast, muscle D3 activity increased in septic patients compared with controls, possibly enhancing T3 clearance in muscle. Furthermore, TRβ1 mRNA, as well as RXRγ mRNA, decreased during sepsis. During *Streptococcus pneumoniae* infection in mice, D3 mRNA expression did not change, while D3 activity tended to increase.[88] The combination of decreased D2 expression and unaltered or increased D3, observed in both human and experimental sepsis models, may result in decreased T3 bioavailability in muscle tissue. However, muscle mRNA expression of *UCP3*, a T3 target gene, is unaltered in the muscle tissue of septic patients compared with control patients,[81] and myogenin, another T3-responsive gene, did not change in murine muscle during chronic inflammation and bacterial sepsis.[44] These observations suggest that the downregulation of TR expression may be an adaptation to restore TR occupancy and T3 signaling. To make it even more complicated, muscle T3 and T4 concentrations were lower in all animal NTIS models compared with controls, despite variable alterations in transporter expression and deiodinating enzymes.[44]

TR expression has also been studied in the diaphragm, which is the main respiratory muscle, with high energy demands. Acute inflammation in mice profoundly decreases TRα1 and TRβ1 mRNA expression in the diaphragm, in association with decreased expression of coactivators involved in nuclear receptor activation.[91] The effects of bacterial sepsis and chronic inflammation on the diaphragm have also been studied; sepsis resulted in increased D3 and decreased D2 expression, which was associated with reduced TH signaling. Chronic inflammation, on the other hand resulted, in decreased D3 expression.[92] Thus, the situation in the diaphragm with respect to D2 and D3 expression is identical to that observed in skeletal muscle during bacterial sepsis and chronic inflammation.[44]

The differential regulation of muscle deiodinase expression among acute inflammation, chronic inflammation, and sepsis may relate to the severity of illness and support the concept that the net result of altered muscle deiodinase expression determines muscle function during specific stages of illness.

Adipose Tissue

White adipose tissue (WAT) is a key T3-responsive metabolic organ. WAT expresses multiple genes involved in TH transport and metabolism, such as *MCT8, D2, D3, TRα1, TRβ1,* and a variety of co-activators involved in TR signaling.[93] Illness affects adipose tissue via stimulating lipolysis, which is a partly T3-regulated process.[94] The effect of illness on adipose tissue, however, has not been extensively studied. Acute inflammation induced in experimental animals resulted in decreased mRNA expression of TRα1 and TRβ1, which is associated with a decrease in coactivators and reduced expression of the well-known T3 target genes *ME* and *Spot14*.[93] LPS administration also resulted in increased D2 mRNA expression and a decrease in D3 mRNA expression; the increase in D2 mRNA expression was, however, not accompanied by an increase in D2 enzyme activity.[95]

In contrast, subcutaneous WAT of NTIS patients with septic shock within 5 days after admission to the ICU showed moderately decreased MCT8 and TRβ1/TRα1 mRNA expression, while D2 mRNA expression and D3 activity did not change.[81] Surprisingly, UCP3 expression was increased, suggesting increased T3 signaling. This again shows the dissociation between T3 uptake and TR expression on the one hand and T3 action on the other. As increased T3 signaling in adipose tissue during sepsis is opposite to decreased T3 signaling observed during nonlethal acute inflammation, this again shows the differential response in different metabolic organs during different states of NTIS.

KEY POINTS

- A novel role for TH metabolism in innate immune cells, including neutrophils and monocytes/macrophages, was shown. Its role in susceptibility to infections is unclear at present.

Neutrophils and Macrophages

Innate immune cells are part of the innate immune system, which is responsible for the host defense against invading pathogens via a variety of bacterial killing mechanisms, including phagocytosis or the production of specific inflammatory signals. Important players of the innate immune system are neutrophils and macrophages, and recent studies indicate that the cellular function of these cells is highly dependent on the availability of intracellular TH, regulated via D3 and D2, respectively. As observed in other tissues, D3 and D2 expression alters during inflammation, supporting the role of TH in cellular function.[96]

D3 protein is highly expressed in infiltrating murine neutrophils during both bacterial infection and sterile inflammation[97,98] and is also expressed in human neutrophils.[99] D3 activity was strongly elevated in a subcutaneous abscess containing predominantly activating neutrophils.[97] D3 in human neutrophils was found in intracellular granules involved in bacterial killing and in early stage neutrophil extracellular traps.[99] Mice that lack D3 have impaired bacterial killing upon infection with *S. pneumoniae*.[55] In line with this observation, D3 knockdown in zebrafish embryos with pneumococcal meningitis resulted in increased mortality and reduced neutrophil infiltration during infection. *In vitro* experiments with stimulated neutrophils from female D3KO mice showed that NADPH-oxidase activity is impaired compared with WT neutrophils.[100] These findings strongly support a critical role for the TH inactivating enzyme D3 in neutrophil function during infection.

Macrophages contain several essential elements of intracellular TH metabolism; D2 is the main deiodinase present, and TRα1 the dominant TR.[101] Activation of macrophages by bacterial endotoxin results in the induction of D2, together with TRα1 and MCT10, indicating a shift towards increased TH action during inflammation.[101] Furthermore, D2 knockdown resulted in impaired macrophage phagocytosis and blunted cytokine response to LPS stimulation.[101,102] In zebrafish embryos, D2 knockdown increased mortality during pneumococcal meningitis, and this can be restored by T3 administration during infection.[102] Increased intracellular T3 availability and action thus appear to play a crucial role in macrophage function.

KEY POINTS

- The role of inflammatory pathways in non-systemic changes during NTIS has begun to be clarified.

MECHANISM OF ALTERED THYROID HORMONE METABOLISM IN NONTHYROIDAL ILLNESS SYNDROME

It has gradually become clear that the cause of NTIS is multifactorial and may differ in different groups of patients. Specifically, the changes in liver disease and renal disease are somewhat different from those occurring in other forms of illness. Several mechanisms are known to be involved in the illness-induced alterations in the HPT axis and TH target tissues, and these mechanisms will be discussed later. Inflammatory mediators play an important role and have been studied extensively using a variety of *in vivo* and *in vitro* models. Well-studied NTIS animal models are acute illness, induced by administration of a sublethal dose of LPS[53,54]; chronic inflammation, induced by injection of turpentine in the hind limb, ultimately resulting in the formation of an abscess[97,103]; bacterial sepsis, induced by inoculation with *S. pneumoniae* or intraperitoneal injection with *Escherichia coli*[98,104]; and prolonged critical illness in rabbits induced by burn injury.[105] Characteristic for these models is the downregulation of serum TH concentrations despite the variability of the inflammatory response. The recent availability of a variety of transgenic mice models makes it possible to study the specific contribution of genes involved in the pathogenesis of NTIS. The role of inflammatory mediators involved in altered TH metabolism has been studied extensively *in vitro* using a variety of cell lines that were stimulated with LPS or proinflammatory cytokines with and without specific inhibitors.

Downregulation of Hypothalamic and/or Pituitary Function

Considerable evidence suggests that an alteration in hypothalamic and pituitary function causes the low production of T4, which in turn causes the low production of T3. In rats, starvation reduces hypothalamic mRNA for TRH, reduces portal serum TRH, and lowers pituitary TSH content.[106] One study documents low TRH mRNA in hypothalamic paraventricular nuclei in NTIS patients.[37] In addition, several animal studies show that hypothalamic TRH expression also decreases after acute inflammation,[107] chronic inflammation,[57] and prolonged critical illness in rabbits.[50] Responses to administered TRH vary in different reports, being suppressed or even augmented.[33,108] Administration of TRH has been suggested as an effective means of restoring serum hormone levels to normal in individuals with NTIS. A recent report of great significance by Van den Berghe and coworkers proves that administration of TRH to patients with severe NTIS leads directly to increased TSH levels, increased T4 levels, and increased T3 levels.[38] These data provide strong support (albeit not proof) for the role of diminished hypothalamic function as an important factor causing NTIS.

The illness-induced suppression of TRH in the PVN is thought to be mediated by increased local T3 production via increased D2 expression in tanycytes and decreased D3 expression in neurons.[61] Inflammatory mediators contribute to the illness induced alteration in both D2 and D3; the mechanisms will be discussed later.

An additional factor in the pathogenesis of the illness-induced downregulation of hypothalamic hypophysiotropic TRH neurons may be regulation by afferent pathways projecting from brainstem nuclei to the PVN. LPS is known to modulate autonomic centers in the lower brainstem, in turn activating hypophysiotropic corticotrophin-releasing hormone (CRH) neurons in the PVN.[109,110] Like CRH neurons, hypophysiotropic TRH neurons receive neural input from the brainstem. However, as transsection of these ascending pathways did not affect the LPS-induced decrease of TRH expression in the PVN, neural input from the brainstem appears to be relatively unimportant in this respect.[111]

Corticosterone is also known to affect TRH expression directly and could also contribute to the TRH decrease during illness.[112] The diurnal variation in glucocorticoid levels at least in part controls the normal diurnal variation in TSH levels, perhaps by affecting pituitary responsiveness to TRH.[113] High levels of glucocorticoids in Cushing disease suppress TSH and cause a modest reduction in serum hormone levels.[114] High levels of glucocorticoids are known to suppress the pituitary response to TRH in humans.[115] Stress-induced elevation of glucocorticoids in animals causes suppression of TSH and serum T4 and T3 hormone levels.[116] Thus, stress-induced glucocorticoid elevation may be one factor affecting TRH and TSH production.

TRH is also decreased during starvation, suggesting that part of the observed TRH decrease during illness may be a result of the diminished food intake associated with illness. However, in a mouse model for chronic inflammation, diminished food intake does not play a role in the observed TRH decrease.[57]

Pituitary production of TSH is often diminished despite low serum TH levels, as described in patients[34] and in a wide variety of animal models.[50,53,54,57] The mechanisms involved are incompletely understood, although it has been shown that the illness-induced decrease of TSHβ expression depends on functional TR signaling because TRβ$^{0/0}$ mice show a blunted TSHβ decrease upon LPS stimulation compared with WT mice.[56] It is also hypothesized that intrapituitary conversion of T4 to T3 was increased, which allowed the pituitary gland to remain euthyroid while the rest of the body was actually hypothyroid. While some data supported this idea in a uremic rat model of NTIS,[117] careful studies in both experimental animals and humans yielded conflicting results.

The most obvious possibility is that low TSH stems from diminished TRH production, as previously described. It must also be remembered that the defect in pituitary function is not restricted to TSH, but that LH and FSH are also suppressed in seriously ill patients, and testosterone is reduced, in contrast to the generally augmented glucocorticoid response. Quite possibly, these changes are the effect on the hypothalamus of neural integration of multiple factors, including stress, starvation, glucocorticoids, and cytokines.

Cyrokines

Cytokines have been implicated in the development of NTIS for more than two decades. Interleukin (IL)-6 was found to be negatively correlated with serum T3 concentrations in hospitalized patients.[118] In a series of septic patients studied shortly after admission to an ICU, total T4, free T4, total T3, and TSH were depressed, and IL-1β, soluble interleukin-2 receptor (sIL-2R), IL-6, and tumor necrosis factor (TNF)α were elevated.[119] In mice, acute inflammation is associated with an acute increase of serum IL-6 and TNFα concentrations.[64] A causal role for IL-6 in the development of NTIS in mice was shown, as IL-6 knockout mice show a less pronounced drop in serum T3 during illness.[120] However, acute injection of cytokines failed to induce NTIS-like features in mice, except for injection of interferon gamma (IFNγ), which reduces serum T3 and T4.[64] In humans, subcutaneous administration of IFNγ did not alter TNFα levels but caused a small elevation of IL-6 levels, and thus did not support a role for IFNγ in the pathogenesis of the euthyroid sick syndrome in humans.[121]

Chronic infusions with IL-1 and IL-6 mimic certain symptoms, such as decreased serum T4 and T3 and decreased TRH expression in the hypothalamus in mice,[122] whereas chronic IL-1β infusions mimics certain aspects of NTIS in rats.[123] Van der Poll and colleagues[124] studied the effect of IL-1 receptor blockade in human volunteers to determine if it could alter the NTIS induced by endotoxin. Blockade of IL-1 signaling was achieved by infusing recombinant human IL-1 receptor antagonist, but this did not prevent the drop in T4, free T4, T3, and TSH or the rise in rT3 caused by endotoxin. Stouthard and coworkers[125] gave recombinant human IL-6 chronically to human volunteers. Short-term infusion of IL-6 caused a suppression of TSH, but daily injections over 42 days caused only a modest decrease in T3 and a transient increase in rT3 and free T4 concentrations.

Infusion of recombinant TNFα in man by van der Poll and coworkers[126] produced a decrease in serum T3 and TSH and an increase in rT3. However, blocking TNFα by a recombinant TNF receptor–IgG fusion protein did not prevent the endotoxin-induced NTIS in humans,[127] indicating that TNFα alone did not cause the changes in hormone economy induced by endotoxin.

A variety of studies (reviewed by de Vries and coworkers) clearly show that cytokines, either alone or synergistically, are able to downregulate various components of the TH synthesis pathway in the thyroid via an effect on iodide uptake and thyroglobulin synthesis, ultimately leading to decreased secretion of T4 and T3.[128] Proinflammatory cytokines are also able to affect TH metabolism and action in hepatocytes; IL-1β decreases TRβ, TRα1, and D1 mRNA expression in a human hepatoma cell line (HepG2),[129,130] and TNFα, IL-1, and IL-6 decrease the capacity of T3 binding to the TR.[131]

In summary, the potential interaction between proinflammatory cytokines and the HPT axis is certainly complicated, as cytokines, increased in many infectious, inflammatory, and neoplastic conditions, themselves operate in a network via the stimulation of other proinflammatory and antiinflammatory cytokines. Many studies therefore fail to pinpoint one cytokine as the crucial mediator, but we can be convinced that striking changes in cytokines cooccur during NTIS and probably play a pathogenic role.

Nuclear Factor κB

Recent studies indicate a role for nuclear factor κB (NF-κB) in the pathogenesis of NTIS. The NF-κB protein family is an inflammatory signal transduction pathway consisting of several subunits, of which RelA (the 65-kD subunit of NF-κB) acts as a transcriptional activator able to bind to the Dio2 promoter.[132,133] LPS stimulation of primary tanycyte cell cultures increases D2 and RelA mRNA expression and RelA phosphorylation, confirming activation of this pathway. Inhibition of RelA with chemical inhibitors prevented the LPS-induced D2 increase, indicating that NF-κB signaling is essential for the upregulation of D2 in tanycytes *in vitro*.[134] Conditional elimination of RelA in tanycytes using transgenic mice shows that NF-κB is essential for the LPS-induced hypothalamic D2 increase and TRH decrease *in vivo*. The central changes in the HPT axis are, however, not required for the downregulation of Tshβ expression and serum TH concentrations.[135]

Certainly, one important cause of the drop in serum T3 is a decreased generation of T3 by type 1 iodothyronine deiodinase.[136] NF-κB is also thought to play a role in the illness-induced decrease of D1. TNFα impairs the T3-dependent induction of D1 expression in HepG2 cells via interference of NF-κB with TR function.[137] Kwakkel et al. showed that the IL-1β–induced decrease in TRβ mRNA expression in HepG2 cells depends on NF-kB signaling, and the decreases in D1 on both NF-kB and AP-1 signaling.[129,130]

Alternative Mechanisms

The *DIO1* gene is positively regulated via a TR/retinoid-X receptor (RXR) heterodimer, indicating that decreased *DIO1* expression may result from reduced TR and/or reduced RXR expression.[138] Liver TRβ1 expression decreases during acute inflammation in mice,[53,74,139] and Yu et al. proposed a very elegant mechanism for the TRβ-mediated repression of liver D1 during acute inflammation. They showed both *in vivo* and *in vitro* that adding exogenous coactivator SRC-1 attenuates the illness-induced liver D1 decrease.[140,141] This indicates that competition for limiting amounts of SRC-1, which is a shared coactivator for TR and inflammatory signaling pathways, could be responsible for the illness-induced D1 decrease. Indeed, restoration of liver Dio1 expression by exogenous SRC-1 prevents the fall in serum TH levels after LPS[141] (Fig. 83.3).

An alternative possibility is the availability of a specific cofactor glutathione (GSH), which is required for D1 catalytic activity. Decreased amounts of GSH may play a role in the illness-induced decrease of liver D1 activity, as indicated by a study performed by Wajner et al. showing that addition of N-acetyl-cysteine, an antioxidant that restores intracellular GSH levels, prevents the IL-6–induced suppression of D1.[142]

Although these studies provide mechanisms behind the illness-induced D1 decrease in the liver, at this stage, the importance of decreased liver D1 for the development of NTIS is unclear, as studies in D1/D2 knockout mice showed similar responses to LPS administration with regard to serum T4 and T3 compared with WT littermates.[143] Although the effect of a lack of D1 on illness-induced liver T3 concentrations has not been documented, the suppression of liver D1 activity was correlated with decreased hepatic T3 concentrations in rabbits with critical illness.[79]

Changes in D3 are observed in liver and muscle, depending on the type and the timing of the illness.[18,84,88] These changes may be independent of inflammatory pathways, as neither sepsis nor chronic inflammation induces phosphorylation of the NF-κB and ERK pathways.[88] In cardiomyocytes, D3 is upregulated during myocardial infarction.[144,145] Hypoxia because of decreased tissue perfusion is thought to play a role in the observed increase, as both D3 and hypoxia inducible factor (HIF)1α are upregulated in the hypertrophic heart. HIF1α activity is regulated by prolyl hydroxylases that prime HIF1α for degradation.[146] Under hypoxic conditions, HIF1α stabilizes and translocates to the nucleus, dimerizes with HIF1β, and activates D3 gene transcription.[147] In addition, during inflammation NF-κB also directly enhances *HIF1α* gene transcription, thereby increasing total HIF1α availability.[148] Increased D3 expression during myocardial infarction is also associated with increased expression of a specific set of microRNAs that may enhance the proliferative capacity of cardiomyocytes.[149]

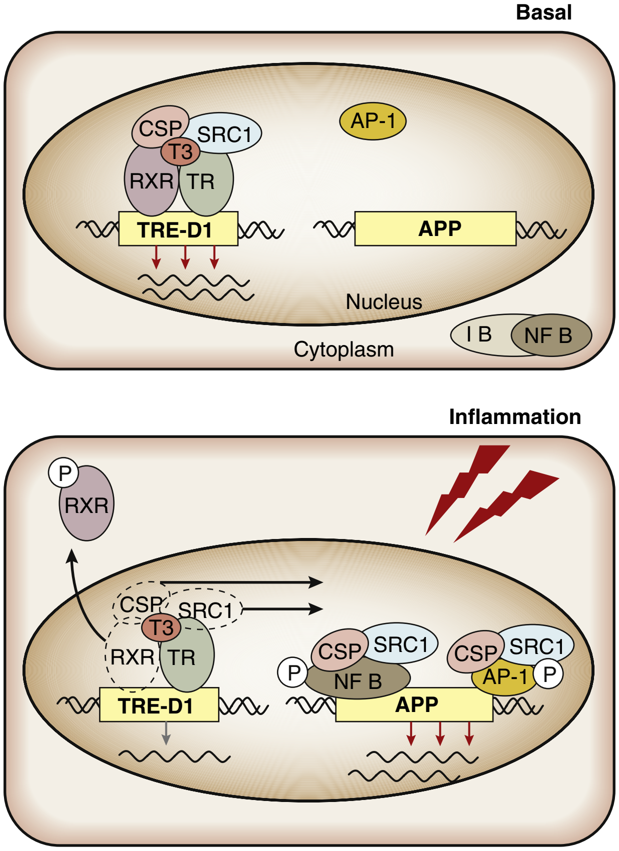


Fig. 83.3 Schematic model of the proposed mechanisms involved in illness induced downregulation of liver type 1 deiodinase (*D1*) expression. In this model, limited amounts of SRC-1, a coactivator for both nuclear factor (*NF*)-κB and thyroid hormone receptor (*TR*)β1, decrease liver D1 mRNA expression as a result of competition between cytokine-induced NF-κB and the TRβ1/D1 complex for SRC-1. Note the role of decreased nuclear retinoid-X receptor α protein by lipopolysaccharide-induced rapid nuclear export via JNK phosphorylation and subsequent proteasomal degradation. (Used with permission from Boelen A, Kwakkel J, Fliers E. Beyond low plasma T3: local thyroid hormone metabolism during inflammation and infection. *Endocr Rev.* 2011;32:670–693.)

KEY POINTS

- Research on NTIS in patient populations should take (parenteral) feeding into account as a major determinant of NTIS as well as clinical outcome.

Energy Status

Critical illness is associated with marked loss of appetite and poor nutritional intake.[150] As food deprivation in healthy subjects induces a NTIS response similar to critically ill patients, decreased caloric intake during illness may be a major determinant of the development of NTIS, in addition to many other factors already discussed (Fig. 83.4). Three clinical studies involving different nutritional strategies indeed indicated that

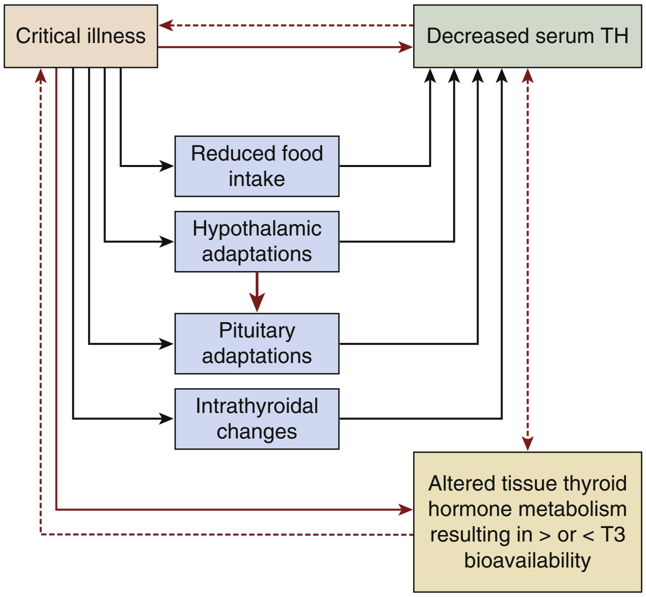


Fig. 83.4 Schematic representation of the variety of changes occurring during critical illness. Solid lines represent a causal relation, while a dashed line represents a probable effect. The scheme is based on both experimental and human studies. The net result of altered tissue thyroid hormone (*TH*) metabolism may be beneficial or maladaptive, dependent on disease duration and severity. *T3*, triiodothyronine. (Used with permission from Fliers E, Bianco AC, Langouche L, et al. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol.* 2015;3:816–825.)

decreased caloric intake during critical illness is associated with a more pronounced NTIS.[151-153] The mechanism by which fasting induces a decrease in serum TH is multifactorial and includes a decrease in serum leptin, contributing to downregulation of hypothalamic hypophysiotropic TRH neurons, allowing persistence of low serum TSH.[154]

Leptin plays a key role in the control of TH levels during starvation in animals. During starvation, circulating leptin levels drop. This induces diminished stimulation of TRH, and thus diminished secretion of TSH and lowered TH levels. Administration of leptin appears to work via the arcuate nucleus of the hypothalamus to induce production of proopiomelanocortin, and thus α melanocyte-stimulating hormone (αMSH) and reduce Agouti-related protein (AgRP). Normally, αMSH stimulates the melanocortin 4 receptor (MC4R), whereas AgRP suppresses it. Presumably through these actions, a lack of leptin during starvation leads to diminished stimulation of the MC4R receptor on the TRH neurons in ventricular nuclear centers, and thus diminished TRH secretion. Administration of leptin partially reverses this sequence.[155] These actions appear to be part of an energy-conserving scheme related to thyroid changes during starvation and are associated with leptin-induced increase in appetite, decreased energy expenditure, and modified neuroendocrine function. The relevance of this to human physiology is as yet unclear, but the data strongly suggest that leptin is involved in the downregulation of thyroid function during acute starvation.[156,157] In clinical trials, stimulation of growth hormone secretion by growth hormone secretagogues led to increased insulin and leptin levels in severely ill ICU patients. To date, studies of leptin levels in patients with NTIS have indicated they are normal or elevated, not low.[38] Liver D3 increases during fasting in mice, and this increase is dependent on the drop of leptin.[71] Increased activity of D3, however, is not supposed to contribute to the clearance of T3 and T4 from plasma during fasting because D3KO and WT mice show the same decreased serum TH concentrations in response to fasting.[158]

CLINICAL IMPLICATIONS AND THERAPEUTIC CONSIDERATIONS

It is straightforward that the typical clinical parameters of severe hypothyroidism are absent in patients with NTIS. The clinical endocrinologist is frequently presented with a severely ill patient in the ICU without a history of pituitary or thyroid disease with a low (free)T4, a low or normal TSH, and, if measured, a very low (free)T3. If T4 is below 4 µg/dL in this setting, the diagnosis of NTIS, associated with a potentially fatal outcome, may be assumed; rT3 may be normal or elevated and is not diagnostic. An elevated TSH suggests the presence of prior hypothyroidism, which should then be treated. Finding positive antithyroid antibody titers supports the diagnosis of primary hypothyroidism but does not prove it. Use of anticonvulsants should be noted, as these agents can lower (free)T4. Dopamine used in the setting of severe illness can decrease serum TSH, contributing to persistent low TH concentrations. Hyperthyroidism is the typical cause of suppression of TSH below 0.01 µU/mL, but it is rarely difficult to exclude this diagnosis in the setting of severely depressed T4 and T3.

The question of whether interventions aimed at normalizing TH concentrations in patients with NTIS in the framework of extended critical illness are beneficial has so far not been satisfactorily answered. Only a small number of randomized controlled trials (RCTs) involving heterogeneous patient groups, ranging from critical illness at the medical ICU to heart surgery in otherwise healthy participants, have assessed the effects of treatment with THs in patients with NTIS. Both T3 and levothyroxine have been used. Overall, the RCTs of T3 or T4 in critical illness have been largely negative in terms of clinical benefit (for review, see Fliers et al.).[159] This may be partly explained by the concept that normalizing TH concentrations in serum does not equal normalizing tissue concentrations of TH. This was clearly shown in a study involving patients who were critically ill and received TH treatment. Hepatic T3 concentrations after TH treatment were disproportionally high compared with the increase in serum and muscle T3 concentrations.[78]

An unresolved and somewhat controversial issue is whether there is a place for TH treatment in patients with heart failure. Although the use of TH treatment in patients with heart failure has not been adequately studied, some trials have reported encouraging findings. A small RCT[160] in patients with dilated cardiomyopathy showed beneficial effects of medium-term (3 months) levothyroxine treatment on cardiac performance, whereas another RCT[161] showed that the TH analogue 3,5 di-iodothyropropionic acid improved some hemodynamic variables in heart failure.

An important question is whether hypothalamic neuropeptides (e.g., growth hormone-releasing hormone, growth hormone-releasing peptide 2, gonadotropin-releasing hormone, and TRH) can be used in patients with long-term critical illness in an attempt to stimulate the anterior pituitary gland, thereby restoring endocrine function in terms of plasma concentrations and hormone pulsatility.[58,162] A number of small clinical studies showed that hypothalamic neuropeptides, especially TRH in combination with growth hormone releasing peptide 2 given intravenously, restored circulating concentrations of TH and TSH pulsatility to a remarkable extent.[163] Moreover, this strategy improved bone markers and anabolic variables. These findings may point to insufficient hypothalamic drive of pituitary TSH release in protracted critical illness, which was supported by a study reporting that deceased patients with NTIS had decreased TRH mRNA expression in the hypothalamic PVN, correlating with decreased antemortem plasma TSH and T3 concentrations.[37] However, the clinical studies using hypothalamic neuropeptides in patients with protracted critical illness involved only small numbers of patients and lacked clinically significant endpoints related to morbidity and mortality.

In sum, no definitive conclusion about the effectiveness of TH treatment in patients in the ICU with NTIS can yet be made. One study in patients with acute renal failure even raised some concern about possible harm, because of a tendency towards increased mortality after treatment with levothyroxine.[164] Clearly, the reported studies so far were not adequately powered to detect clinically meaningful differences. Furthermore, most of these studies used rather high doses of either T4 or T3, probably inducing further suppression of pituitary TSH release and altered tissue deiodinase activities. In this respect, the use of neuropeptides, including TRH, to stimulate the HPT axis could be more promising. Large RCTs in well-defined patient groups will be needed to assess possible positive effects of this approach in terms of clinical outcome. High priority should be given to RCTs comparing the effect of hypothalamic neuropeptides, including TRH, with placebo, because this approach has already been shown to partly normalize concentrations of serum THs and at the same time improve metabolic markers.[165] Another theoretical possibility would be to investigate treatment with recombinant human TSH because this is a physiological stimulus—similar to TRH—for TH release from the thyroid.

KEY POINTS

- There is at present insufficient evidence that TH supplementation during NTIS is beneficial. There is a clear need for adequately powered RCTs with clinically relevant endpoints to fill this knowledge gap.

CONCLUSION

NTIS has traditionally been seen as a syndrome with lower plasma TH concentrations of unknown significance. However, it is currently clear that NTIS represents a complex set of marked and differential changes in TH physiology, both at the level of the HPT axis and at the organ level.

Whether the observed changes in critically ill patients defined as NTIS are beneficial or deleterious is currently unknown and probably depends on disease stage and severity, the need of prolonged vital support, and environmental factors including parenteral nutrition. At present, there is no evidence-based guideline supporting TH treatment of NTIS in critically ill patients. There is a clear need to perform RCTs to define whether pharmacological management of NTIS, (e.g., using neuropeptides, including TRH, may yield benefit in terms of clinical outcome).

REFERENCES

1. Docter R, Krenning EP, de Jong M, et al. The sick euthyroid syndrome: changes in thyroid hormone serum parameters and hormone metabolism. *Clin Endocrinol*. 1993;39:499–518.
2. Wiersinga WM. Nonthyroidal illness. In: Braverman LE, Utiger RD, eds. *The Thyroid*. Philadelphia: Lippincot; 2000:281–295.
3. Wartofsky L, Burman KD. Alterations in thyroid function in patients with systemic illness: the "euthyroid sick syndrome." *Endocr Rev*. 1982;3:164–217.
4. Boelen A, Wiersinga WM, Fliers E. Fasting-induced changes in the hypothalamus-pituitary-thyroid axis. *Thyroid*. 2008;18:123–129.
5. Harris AR, Fang SL, Vagenakis AG, et al. Effect of starvation, nutriment replacement, and hypothyroidism on in vitro hepatic T4 to T3 conversion in the rat. *Metabolism*. 1978;27:1680–1690.
6. Welle SL, Campbell RG. Decrease in resting metabolic rate during rapid weight loss is reversed by low dose thyroid hormone treatment. *Metabolism*. 1986;35:289–291.
7. Maldonado LS, Murata GH, Hershman JM, et al. Do thyroid function tests independently predict survival in the critically ill? *Thyroid*. 1992;2:119–123.
8. Vaughan GM, Mason Jr AD, McManus WF, et al. Alterations of mental status and thyroid hormones after thermal injury. *J Clin Endocrinol Metab*. 1985;60:1221–1225.
9. De Marinis L, Mancini A, Masala R, et al. Evaluation of pituitary-thyroid axis response to acute myocardial infarction. *J Endocrinol Invest*. 1985;8:507–511.
10. Surks MI, Hupart KH, Pan C, et al. Normal free thyroxine in critical nonthyroidal illnesses measured by ultrafiltration of undiluted serum and equilibrium dialysis. *J Clin Endocrinol Metab*. 1988;67:1031–1039.
11. Melmed S, Geola FL, Reed AW, et al. A comparison of methods for assessing thyroid function in nonthyroidal illness. *J Clin Endocrinol Metab*. 1982;54:300–306.
12. Kaptein EM, MacIntyre SS, Weiner JM, et al. Free thyroxine estimates in nonthyroidal illness: comparison of eight methods. *J Clin Endocrinol Metab*. 1981;52:1073–1077.
13. Kantor MJ, Leef KH, Bartoshesky L, et al. Admission thyroid evaluation in very-low-birth-weight infants: association with death and severe intraventricular hemorrhage. *Thyroid*. 2003;13:965–969.
14. Chopra IJ, Taing P, Mikus L. Direct determination of free triiodothyronine (T3) in undiluted serum by equilibrium dialysis/radioimmunoassay (RIA). *Thyroid*. 1996;6:255–259.
15. Chopra IJ. Simultaneous measurement of free thyroxine and free 3,5,3'-triiodothyronine in undiluted serum by direct equilibrium dialysis/radioimmunoassay: evidence that free triiodothyronine and free thyroxine are normal in many patients with the low triiodothyronine syndrome. *Thyroid*. 1998;8:249–257.
16. Sapin R, Schlienger JL, Kaltenbach G, et al. Determination of free triiodothyronine by six different methods in patients with non-thyroidal illness and in patients treated with amiodarone. *Ann Clin Biochem*. 1995;32:314–324.
17. Faber J, Kirkegaard C, Rasmussen B, et al. Pituitary-thyroid axis in critical illness. *J Clin Endocrinol Metab*. 1987;65:315–320.
18. Peeters RP, Wouters PJ, Kaptein E, et al. Reduced activation and increased inactivation of thyroid hormone in tissues of critically ill patients. *J Clin Endocrinol Metab*. 2003;88:3202–3211.
19. Chopra IJ, Solomon DH, Hepner GW, et al. Misleadingly low free thyroxine index and usefulness of reverse triiodothyronine measurement in nonthyroidal illnesses. *Ann Intern Med*. 1979;90:905–912.
20. Klemperer JD, Klein I, Gomez M, et al. Thyroid hormone treatment after coronary-artery bypass surgery. *N Engl J Med*. 1995;333:1522–1527.
21. Osburne RC, Myers EA, Rodbard D, et al. Adaptation to hypocaloric feeding: physiologic significance of the fall in serum T3 as measured by the pulse wave arrival time (QKd). *Metabolism*. 1983;32:9–13.
22. Afandi B, Schussler GC, Arafeh AH, et al. Selective consumption of thyroxine-binding globulin during cardiac bypass surgery. *Metabolism*. 2000;49:270–274.
23. Jirasakuldech B, Schussler GC, Yap MG, et al. A characteristic serpin cleavage product of thyroxine-binding globulin appears in sepsis sera. *J Clin Endocrinol Metab*. 2000;85:3996–3999.
24. den Brinker M, Joosten KF, Visser TJ, et al. Euthyroid sick syndrome in meningococcal sepsis: the impact of peripheral thyroid hormone metabolism and binding proteins. *J Clin Endocrinol Metab*. 2005;90:5613–5620.
25. Bacci V, Schussler GC, Kaplan TB. The relationship between serum triiodothyronine and thyrotropin during systemic illness. *J Clin Endocrinol Metab*. 1982;54:1229–1235.
26. Uchimura H, Nagataki S, Tabuchi T, et al. Measurements of free thyroxine: comparison of per cent of free thyroxine in diluted and undiluted sera. *J Clin Endocrinol Metab*. 1976;42:561–566.
27. Nelson JC, Weiss RM. The effect of serum dilution on free thyroxine (T4) concentration in the low T4 syndrome of nonthyroidal illness. *J Clin Endocrinol Metab*. 1985;61:239–246.
28. Wang R, Nelson JC, Weiss RM, et al. Accuracy of free thyroxine measurements across natural ranges of thyroxine binding to serum proteins. *Thyroid*. 2000;10:31–39.
29. Liewendahl K, Helenius T, Lamberg BA, et al. Free thyroxine, free triiodothyronine, and thyrotropin concentrations in hypothyroid and thyroid carcinoma patients receiving thyroxine therapy. *Acta Endocrinologica*. 1987;116:418–424.
30. Mendel CM, Frost PH, Cavalieri RR. Effect of free fatty acids on the concentration of free thyroxine in human serum: the role of albumin. *J Clin Endocrinol Metab*. 1986;63:1394–1399.
31. Jaume JC, Mendel CM, Frost PH, et al. Extremely low doses of heparin release lipase activity into the plasma and can thereby cause artifactual

elevations in the serum-free thyroxine concentration as measured by equilibrium dialysis. *Thyroid*. 1996;6:79–83.

32. Stevenson HP, Archbold GP, Johnston P, et al. Misleading serum free thyroxine results during low molecular weight heparin treatment. *Clin Chem*. 1998;44:1002–1007.
33. Chopra IJ. Nonthyroidal illness syndrome or euthyroid sick syndrome? *Endocrine Pract*. 1996;2:45–52.
34. Van den Berghe G. Dynamic neuroendocrine responses to critical illness. *Front Neuroendocrinol*. 2002;23:370–391.
35. Brent GA, Hershman JM. Thyroxine therapy in patients with severe nonthyroidal illnesses and low serum thyroxine concentration. *J Clin Endocrinol Metab*. 1986;63:1–8.
36. Langouche L, Van den Berghe G. Hypothalamic-pituitary hormones during critical illness: a dynamic neuroendocrine response. *Handb Clin Neurol*. 2014;124:115–126.
37. Fliers E, Guldenaar SE, Wiersinga WM, et al. Decreased hypothalamic thyrotropin-releasing hormone gene expression in patients with nonthyroidal illness. *J Clin Endocrinol Metab*. 1997;82:4032–4036.
38. Van den Berghe G, de Zegher F, Baxter RC, et al. Neuroendocrinology of prolonged critical illness: effects of exogenous thyrotropin-releasing hormone and its combination with growth hormone secretagogues. *J Clin Endocrinol Metab*. 1998;83:309–319.
39. Visser WE, Friesema EC, Visser TJ. Minireview: thyroid hormone transporters: the knowns and the unknowns. *Mol Endocrinol*. 2011;25:1–14.
40. Hennemann G, Docter R, Friesema EC, et al. Plasma membrane transport of thyroid hormones and its role in thyroid hormone metabolism and bioavailability. *Endocr Rev*. 2001;22:451–476.
41. Lim CF, Docter R, Visser TJ, et al. Inhibition of thyroxine transport into cultured rat hepatocytes by serum of nonuremic critically ill patients: effects of bilirubin and nonesterified fatty acids. *J Clin Endocrinol Metab*. 1993;76:1165–1172.
42. Vos RA, De Jong M, Bernard BF, et al. Impaired thyroxine and 3,5,3'-triiodothyronine handling by rat hepatocytes in the presence of serum of patients with nonthyroidal illness. *J Clin Endocrinol Metab*. 1995;80:2364–2370.
43. Sarne DH, Refetoff S. Measurement of thyroxine uptake from serum by cultured human hepatocytes as an index of thyroid status: reduced thyroxine uptake from serum of patients with nonthyroidal illness. *J Clin Endocrinol Metab*. 1985;61:1046–1052.
44. Boelen A, van der Spek AH, Bloise F, et al. Tissue thyroid hormone metabolism is differentially regulated during illness in mice. *J Endocrinol*. 2017;233:25–36.
45. Bianco AC, Kim BW. Deiodinases: implications of the local control of thyroid hormone action. *J Clin Invest*. 2006;116:2571–2579.
46. Boelen A, Kwakkel J, Fliers E. Beyond low plasma T3: local thyroid hormone metabolism during inflammation and infection. *Endocr Rev*. 2011;32:670–693.
47. Bassett JH, Harvey CB, Williams GR. Mechanisms of thyroid hormone receptor-specific nuclear and extra nuclear actions. *Mol Cell Endocrinol*. 2003;213:1–11.
48. Senese R, Cioffi F, de Lange P, et al. Thyroid: biological actions of 'nonclassical' thyroid hormones. *J Endocrinol*. 2014;221:R1–R12.
49. Mebis L, Van den Berghe G. Thyroid axis function and dysfunction in critical illness. *Best Pract Res Clin Endocrinol Metab*. 2011;25:745–757.
50. Mebis L, Debaveye Y, Ellger B, et al. Changes in the central component of the hypothalamus-pituitary-thyroid axis in a rabbit model of prolonged critical illness. *Crit Care*. 2009;13:R147.
51. Alkemade A, Friesema EC, Unmehopa UA, et al. Neuroanatomical pathways for thyroid hormone feedback in the human hypothalamus. *J Clin Endocrinol Metab*. 2005;90:4322–4334.
52. Coppola A, Liu ZW, Andrews ZB, et al. A central thermogenic-like mechanism in feeding regulation: an interplay between arcuate nucleus T3 and UCP2. *Cell Metab*. 2007;5:21–33.
53. Boelen A, Kwakkel J, Thijssen-Timmer DC, et al. Simultaneous changes in central and peripheral components of the hypothalamus-pituitary-thyroid axis in lipopolysaccharide-induced acute illness in mice. *J Endocrinol*. 2004;182:315–323.
54. Fekete C, Gereben B, Doleschall M, et al. Lipopolysaccharide induces type 2 iodothyronine deiodinase in the mediobasal hypothalamus: implications for the nonthyroidal illness syndrome. *Endocrinol*. 2004;145:1649–1655.
55. Boelen A, Kwakkel J, Wieland CW, et al. Impaired bacterial clearance in type 3 deiodinase-deficient mice infected with Streptococcus pneumoniae. *Endocrinology*. 2009;150:1984–1990.
56. Boelen A, Kwakkel J, Chassande O, et al. Thyroid hormone receptor beta mediates acute illness-induced alterations in central thyroid hormone metabolism. *J Neuroendocrinol*. 2009;21:465–472.
57. Boelen A, Kwakkel J, Wiersinga WM, et al. Chronic local inflammation in mice results in decreased TRH and type 3 deiodinase mRNA expression in the hypothalamic paraventricular nucleus independently of diminished food intake. *J Endocrinol*. 2006;191:707–714.
58. Fliers E, Alkemade A, Wiersinga WM, et al. Hypothalamic thyroid hormone feedback in health and disease. *Prog Brain Res*. 2006;153:189–207.
59. Lechan RM, Fekete C. Feedback regulation of thyrotropin-releasing hormone (TRH): mechanisms for the non-thyroidal illness syndrome. *J Endocrinol Invest*. 2004;27:105–119.
60. Freitas BC, Gereben B, Castillo M, et al. Paracrine signaling by glial cell-derived triiodothyronine activates neuronal gene expression in the rodent brain and human cells. *J Clin Invest*. 2010;120:2206–2217.
61. de Vries EM, Surovtseva O, Vos WG, et al. Downregulation of type 3 deiodinase in the hypothalamus during inflammation. *Thyroid*. 2019;29:1336–1343.
62. Boelen A, Kwakkel G, Chassande O, et al. A possible role for the thyroid hormone receptor alpha in inflammation-induced changes in pituitary thyroid hormone metabolism. *Thyroid*. 2009;19:s68.
63. Arem R, Wiener GJ, Kaplan SG, et al. Reduced tissue thyroid hormone levels in fatal illness. *Metabolism*. 1993;42:1102–1108.
64. Boelen A, Platvoet-ter Schiphorst MC, Bakker O, et al. The role of cytokines in the lipopolysaccharide-induced sick euthyroid syndrome in mice. *J Endocrinol*. 1995;146:475–483.
65. Mebis L, Paletta D, Debaveye Y, et al. Expression of thyroid hormone transporters during critical illness. *Eur J Endocrinol*. 2009;161:243–250.
66. Peeters RP, Kester MH, Wouters PJ, et al. Increased thyroxine sulfate levels in critically ill patients as a result of a decreased hepatic type I deiodinase activity. *J Clin Endocrinol Metab*. 2005;90:6460–6465.
67. Debaveye Y, Ellger B, Mebis L, et al. Regulation of tissue iodothyronine deiodinase activity in a model of prolonged critical illness. *Thyroid*. 2008;18:551–560.
68. Peeters RP, Wouters PJ, van Toor H, et al. Serum rT3 and T3/rT3 are prognostic markers in critically ill patients and are associated with post-mortem tissue deiodinase activities. *J Clin Endocrinol Metab*. 2005;90:4559–4565.
69. Debaveye Y, Ellger B, Mebis L, et al. Tissue deiodinase activity during prolonged critical illness: effects of exogenous thyrotropin-releasing hormone and its combination with growth hormone-releasing peptide-2. *Endocrinology*. 2005;146:5604–5611.
70. Weekers F, Giulietti AP, Michalaki M, et al. Metabolic, endocrine, and immune effects of stress hyperglycemia in a rabbit model of prolonged critical illness. *Endocrinology*. 2003;144:5329–5338.
71. Boelen A, van BM, Vos X, et al. Leptin administration restores the fasting-induced increase of hepatic type 3 deiodinase expression in mice. *Thyroid*. 2012;22:192–199.
72. de Vries EM, van Beeren HC, Ackermans MT, et al. Differential effects of fasting vs food restriction on liver thyroid hormone metabolism in male rats. *J Endocrinol*. 2015;224:25–35.
73. Langouche L, Vander Perre S, Marques M, et al. Impact of early nutrient restriction during critical illness on the nonthyroidal illness syndrome and its relation with outcome: a randomized, controlled clinical study. *J Clin Endocrinol Metab*. 2013;98:1006–1013.
74. Beigneux AP, Moser AH, Shigenaga JK, et al. Sick euthyroid syndrome is associated with decreased TR expression and DNA binding in mouse liver. *Am J Physiol Endocrinol Metab*. 2003;284:E228–E236.
75. Williams GR, Franklyn JA, Neuberger JM, et al. Thyroid hormone receptor expression in the "sick euthyroid" syndrome. *Lancet*. 1989;2:1477–1481.
76. Thijssen-Timmer DC, Peeters RP, Wouters P, et al. Thyroid hormone receptor isoform expression in livers of critically ill patients. *Thyroid*. 2007;17:105–112.
77. Chamba A, Neuberger J, Strain A, et al. Expression and function of thyroid hormone receptor variants in normal and chronically diseased human liver. *J Clin Endocrinol Metab*. 1996;81:360–367.
78. Peeters RP, Van der Geijten S, Wouters PJ, et al. Tissue thyroid hormone levels in critical illness. *J Clin Endocrinol Metab*. 2005;90(12):6498–6507.

79. Debaveye Y, Ellger B, Mebis L, et al. Effects of substitution and high-dose thyroid hormone therapy on deiodination, sulfoconjugation, and tissue thyroid hormone levels in prolonged critically ill rabbits. *Endocrinology*. 2008;149:4218–4228.
80. Yu F, Gothe S, Wikstrom L, et al. Effects of thyroid hormone receptor gene disruption on myosin isoform expression in mouse skeletal muscles. *Am J Physiol Regul Integr Comp Physiol*. 2000;278:R1545–R1554.
81. Rodriguez-Perez A, Palos-Paz F, Kaptein E, et al. Identification of molecular mechanisms related to nonthyroidal illness syndrome in skeletal muscle and adipose tissue from patients with septic shock. *Clin Endocrinol (Oxf)*. 2008;68:821–827.
82. Maia AL, Kim BW, Huang SA, et al. Type 2 iodothyronine deiodinase is the major source of plasma T3 in euthyroid humans. *J Clin Invest*. 2005;115:2524–2533.
83. Gereben B, Zeold A, Dentice M, et al. Activation and inactivation of thyroid hormone by deiodinases: local action with general consequences. *Cell Mol Life Sci*. 2008;65:570–590.
84. Kwakkel J, Chassande O, van Beeren HC, et al. Thyroid hormone receptor {alpha} modulates lipopolysaccharide-induced changes in peripheral thyroid hormone metabolism. *Endocrinology*. 2010;151:1959–1969.
85. Mebis L, Langouche L, Visser TJ, et al. The type II iodothyronine deiodinase is up-regulated in skeletal muscle during prolonged critical illness. *J Clin Endocrinol Metab*. 2007;92:3330–3333.
86. Sun X, Wray C, Tian X, et al. Expression of uncoupling protein 3 is upregulated in skeletal muscle during sepsis. *Am J Physiol Endocrinol Metab*. 2003;285:E512–E520.
87. Lado-Abeal J, Romero A, Castro-Piedras I, et al. Thyroid hormone receptors are down-regulated in skeletal muscle of patients with non-thyroidal illness syndrome secondary to non-septic shock. *Eur J Endocrinol*. 2010;163:765–773.
88. Kwakkel J, van Beeren HC, Ackermans MT, et al. Skeletal muscle deiodinase type 2 regulation during illness in mice. *J Endocrinol*. 2009;203:263–270.
89. Harper ME, Seifert EL. Thyroid hormone effects on mitochondrial energetics. *Thyroid*. 2008;18:145–156.
90. Moreno M, de LP, Lombardi A, et al. Metabolic effects of thyroid hormone derivatives. *Thyroid*. 2008;18:239–253.
91. Feingold KR, Moser A, Patzek SM, et al. Infection decreases fatty acid oxidation and nuclear hormone receptors in the diaphragm. *J Lipid Res*. 2009;50:2055–2063.
92. Bloise FF, van der Spek AH, Surovtseva OV, et al. Differential effects of sepsis and chronic inflammation on diaphragm muscle fiber type, thyroid hormone metabolism, and mitochondrial function. *Thyroid*. 2016;26:600–609.
93. Lu B, Moser AH, Shigenaga JK, et al. Type II nuclear hormone receptors, coactivator, and target gene repression in adipose tissue in the acute-phase response. *J Lipid Res*. 2006;47:2179–2190.
94. Klieverik LP, Coomans CP, Endert E, et al. Thyroid hormone effects on whole-body energy homeostasis and tissue-specific fatty acid uptake in vivo. *Endocrinology*. 2009;150:5639–5648.
95. van Zeijl CJ, Surovtseva OV, Kwakkel J, et al. Thyrostimulin deficiency does not alter peripheral responses to acute inflammation-induced nonthyroidal illness. *Am J Physiol Endocrinol Metab*. 2014;307:E527–E537.
96. van der Spek AH, Fliers E, Boelen A. Thyroid hormone metabolism in innate immune cells. *J Endocrinol*. 2017;232:R67–R81.
97. Boelen A, Kwakkel J, Alkemade A, et al. Induction of type 3 deiodinase activity in inflammatory cells of mice with chronic local inflammation. *Endocrinology*. 2005;146:5128–5134.
98. Boelen A, Boorsma J, Kwakkel J, et al. Type 3 deiodinase is highly expressed in infiltrating neutrophilic granulocytes in response to acute bacterial infection. *Thyroid*. 2008;18:1095–1103.
99. van der Spek AH, Bloise FF, Tigchelaar W, et al. The thyroid hormone inactivating enzyme type 3 deiodinase is present in bactericidal granules and the cytoplasm of human neutrophils. *Endocrinology*. 2016;157:3293–3305.
100. van der Spek AH, Jim KK, Karaczyn A, et al. The thyroid hormone inactivating type 3 deiodinase is essential for optimal neutrophil function: observations from three species. *Endocrinology*. 2018;159:826–835.
101. Kwakkel J, Surovtseva OV, de Vries EM, et al. A novel role for the thyroid hormone-activating enzyme type 2 deiodinase in the inflammatory response of macrophages. *Endocrinology*. 2014;155:2725–2734.
102. van der Spek AH, Surovtseva OV, Jim KK, et al. Regulation of intracellular triiodothyronine is essential for optimal macrophage function. *Endocrinology*. 2018;159:2241–2252.
103. Chopra IJ, Huang TS, Boado R, et al. Evidence against benefit from replacement doses of thyroid hormones in nonthyroidal illness (NTI): studies using turpentine oil-injected rat. *J Endocrinol Invest*. 1987;10:559–564.
104. Knapp S, Leemans JC, Florquin S, et al. Alveolar macrophages have a protective antiinflammatory role during murine pneumococcal pneumonia. *Am J Respir Crit Care Med*. 2003;167:171–179.
105. Weekers F, Van HE, Coopmans W, et al. A novel in vivo rabbit model of hypercatabolic critical illness reveals a biphasic neuroendocrine stress response. *Endocrinology*. 2002;143:764–774.
106. Blake NG, Eckland DJ, Foster OJ, et al. Inhibition of hypothalamic thyrotropin-releasing hormone messenger ribonucleic acid during food deprivation. *Endocrinology*. 1991;129:2714–2718.
107. Kakucska I, Romero LI, Clark BD, et al. Suppression of thyrotropin-releasing hormone gene expression by interleukin-1-beta in the rat: implications for nonthyroidal illness. *Neuroendocrinology*. 1994;59:129–137.
108. Vierhapper H, Laggner A, Waldhausl W, et al. Impaired secretion of TSH in critically ill patients with 'low T4-syndrome.' *Acta Endocrinologica*. 1982;101:542–549.
109. Sagar SM, Price KJ, Kasting NW, et al. Anatomic patterns of Fos immunostaining in rat brain following systemic endotoxin administration. *Brain Res Bull*. 1995;36:381–392.
110. Marvel FA, Chen CC, Badr N, et al. Reversible inactivation of the dorsal vagal complex blocks lipopolysaccharide-induced social withdrawal and c-Fos expression in central autonomic nuclei. *Brain Behav Immun*. 2004;18:123–134.
111. Fekete C, Singru PS, Sarkar S, et al. Ascending brainstem pathways are not involved in lipopolysaccharide-induced suppression of thyrotropin-releasing hormone gene expression in the hypothalamic paraventricular nucleus. *Endocrinology*. 2005;146:1357–1363.
112. Kakucska I, Qi Y, Lechan RM. Changes in adrenal status affect hypothalamic thyrotropin-releasing hormone gene expression in parallel with corticotropin-releasing hormone. *Endocrinology*. 1995;136:2795–2802.
113. Brabant G, Brabant A, Ranft U, et al. Circadian and pulsatile thyrotropin secretion in euthyroid man under the influence of thyroid hormone and glucocorticoid administration. *J Clin Endocrinol Metab*. 1987;65:83–88.
114. Benker G, Raida M, Olbricht T, et al. TSH secretion in Cushing's syndrome: relation to glucocorticoid excess, diabetes, goitre, and the 'sick euthyroid syndrome.' *Clin Endocrinol (Oxf)*. 1990;33:777–786.
115. Nicoloff JT, Fisher DA, Appleman Jr MD. The role of glucocorticoids in the regulation of thyroid function in man. *J Clin Invest*. 1970;49:1922–1929.
116. Bianco AC, Nunes MT, Hell NS, et al. The role of glucocorticoids in the stress-induced reduction of extrathyroidal 3,5,3'-triiodothyronine generation in rats. *Endocrinology*. 1987;120:1033–1038.
117. Lim VS, Passo C, Murata Y, et al. Reduced triiodothyronine content in liver but not pituitary of the uremic rat model: demonstration of changes compatible with thyroid hormone deficiency in liver only. *Endocrinology*. 1984;114:280–286.
118. Boelen A, Platvoet-ter Schiphorst MC, Wiersinga WM. Association between serum interleukin-6 and serum 3,5,3'- triiodothyronine in nonthyroidal illness. *J Clin Endocrinol Metab*. 1993;77:1695–1699.
119. Monig H, Arendt T, Meyer M, et al. Activation of the hypothalamo-pituitary-adrenal axis in response to septic or non-septic diseases--implications for the euthyroid sick syndrome. *Intensive Care Med*. 1999;25:1402–1406.
120. Boelen A, Maas MA, Lowik CW, et al. Induced illness in interleukin-6 (IL-6) knock-out mice: a causal role of IL-6 in the development of the low 3,5,3'-triiodothyronine syndrome. *Endocrinology*. 1996;137:5250–5254.
121. de Metz J, Romijn JA, Gouma DJ, et al. Interferon-gamma administration does not affect human thyroid hormone metabolism in the post-surgical euthyroid sick syndrome. *J Endocrinol Invest*. 2002;25:315–320.
122. van Haasteren GA, Linkels E, Klootwijk W, et al. Starvation-induced changes in the hypothalamic content of prothyrotrophin-releasing hormone (proTRH) mRNA and the hypothalamic release of proTRH-derived peptides: role of the adrenal gland. *J Endocrinol*. 1995;145:143–153.

123. Hermus AR, Sweep CG, Demacker PN, et al. Continuous infusion of interleukin-1 beta in rats induces a profound fall in plasma levels of cholesterol and triglycerides. *Arterioscler Thromb*. 1992;12:1036–1043.
124. van der Poll T, Van Zee KJ, Endert E, et al. Interleukin-1 receptor blockade does not affect endotoxin-induced changes in plasma thyroid hormone and thyrotropin concentrations in man. *J Clin Endocrinol Metab*. 1995;80:1341–1346.
125. Stouthard JM, van der PT, Endert E, et al. Effects of acute and chronic interleukin-6 administration on thyroid hormone metabolism in humans. *J Clin Endocrinol Metab*. 1994;79:1342–1346.
126. van der Poll T, Romijn JA, Wiersinga WM, et al. Tumor necrosis factor: a putative mediator of the sick euthyroid syndrome in man. *J Clin Endocrinol Metab*. 1990;71:1567–1572.
127. van der Poll T, Endert E, Coyle SM, et al. Neutralization of TNF does not influence endotoxininduced changes in thyroid hormone metabolism in humans. *Am J Physiol*. 1999;276:R357–R362.
128. de Vries EM, Fliers E, Boelen A. The molecular basis of the non-thyroidal illness syndrome. *J Endocrinol*. 2015;225:R67–R81.
129. Kwakkel J, Wiersinga WM, Boelen A. Differential involvement of nuclear factor-kappaB and activator protein-1 pathways in the interleukin-1beta-mediated decrease of deiodinase type 1 and thyroid hormone receptor beta1 mRNA. *J Endocrinol*. 2006;189:37–44.
130. Kwakkel J, Wiersinga WM, Boelen A. Interleukin-1beta modulates endogenous thyroid hormone receptor alpha gene transcription in liver cells. *J Endocrinol*. 2007;194:257–265.
131. Jakobs TC, Schmutzler C, Meissner J, et al. The promoter of the human type I 5'-deiodinase gene--mapping of the transcription start site and identification of a DR+4 thyroid-hormone-responsive element. *Eur J Biochem*. 1997;247:288–297.
132. Fekete C, Gereben B, Doleschall M, et al. Lipopolysaccharide induces type 2 iodothyronine deiodinase in the mediobasal hypothalamus: implications for the nonthyroidal illness syndrome. *Endocrinology*. 2004;145:1649–1655.
133. Zeold A, Doleschall M, Haffner MC, et al. Characterization of the nuclear factor-kappa B responsiveness of the human dio2 gene. *Endocrinology*. 2006;147:4419–4429.
134. de Vries EM, Kwakkel J, Eggels L, et al. NFkappaB signaling is essential for the lipopolysaccharide-induced increase of type 2 deiodinase in tanycytes. *Endocrinology*. 2014;155:2000–2008.
135. de Vries EM, Nagel S, Haenold R, et al. The role of hypothalamic NF-kappaB signaling in the response of the HPT-axis to acute inflammation in female mice. *Endocrinology*. 2016;157:2947–2956.
136. Kaplan MM. Subcellular alterations causing reduced hepatic thyroxine-5'-monodeiodinase activity in fasted fats. *Endocrinology*. 1979;104:58–64.
137. Nagaya T, Fujieda M, Otsuka G, et al. A potential role of activated NF-kappa B in the pathogenesis of euthyroid sick syndrome. *J Clin Invest*. 2000;106:393–402.
138. Amma LL, Campos-Barros A, Wang Z, et al. Distinct tissue-specific roles for thyroid hormone receptors beta and alpha1 in regulation of type 1 deiodinase expression. *Mol Endocrinol*. 2001;15:467–475.
139. Beigneux AP, Moser AH, Shigenaga JK, et al. The acute phase response is associated with retinoid X receptor repression in rodent liver. *J Biol Chem*. 2000;275:16390–16399.
140. Yu J, Koenig RJ. Regulation of hepatocyte thyroxine 5'-deiodinase by T3 and nuclear receptor coactivators as a model of the sick euthyroid syndrome. *J Biol Chem*. 2000;275:38296–38301.
141. Yu J, Koenig RJ. Induction of type 1 iodothyronine deiodinase to prevent the nonthyroidal illness syndrome in mice. *Endocrinology*. 2006;147:3580–3585.
142. Wajner SM, Goemann IM, Bueno AL, et al. IL-6 promotes nonthyroidal illness syndrome by blocking thyroxine activation while promoting thyroid hormone inactivation in human cells. *J Clin Invest*. 2011;121:1834–1845.
143. St Germain DL, Galton VA, Hernandez A. Minireview: Defining the roles of the iodothyronine deiodinases: current concepts and challenges. *Endocrinology*. 2009;150:1097–1107.
144. Olivares EL, Marassi MP, Fortunato RS, et al. Thyroid function disturbance and type 3 iodothyronine deiodinase induction after myocardial infarction in rats a time course study. *Endocrinology*. 2007;148:4786–4792.
145. Wassen FW, Schiel AE, Kuiper GG, et al. Induction of thyroid hormone-degrading deiodinase in cardiac hypertrophy and failure. *Endocrinology*. 2002;143:2812–2815.
146. Aragones J, Fraisl P, Baes M, et al. Oxygen sensors at the crossroad of metabolism. *Cell Metab*. 2009;9:11–22.
147. Simonides WS, Mulcahey MA, Redout EM, et al. Hypoxia-inducible factor induces local thyroid hormone inactivation during hypoxic-ischemic disease in rats. *J Clin Invest*. 2008;118:975–983.
148. Oliver L, Olivier C, Marhuenda FB, et al. Hypoxia and the malignant glioma microenvironment: regulation and implications for therapy. *Curr Mol Pharmacol*. 2009;2:263–284.
149. Janssen R, Zuidwijk M, Muller A, et al. Cardiac expression of deiodinase type 3 (Dio3) following myocardial infarction is associated with the induction of a pluripotency microRNA signature from the Dlk1-Dio3 genomic region. *Endocrinology*. 2013;154:1973–1978.
150. Schutz P, Bally M, Stanga Z, et al. Loss of appetite in acutely ill medical inpatients: physiological response or therapeutic target? *Swiss Med Wkly*. 2014;144:w13957.
151. Richmand DA, Molitch ME, O'Donnell TF. Altered thyroid hormone levels in bacterial sepsis: the role of nutritional adequacy. *Metabolism*. 1980;29:936–942.
152. Chourdakis M, Kraus MM, Tzellos T, et al. Effect of early compared with delayed enteral nutrition on endocrine function in patients with traumatic brain injury: an open-labeled randomized trial. *JPEN: J Parenter Enteral*. 2012;36:108–116.
153. Ouchi K, Matsubara S, Matsuno S. Effects of supplementary parenteral nutrition on thyroid hormone patterns in surgical patients with liver cirrhosis. *Nutrition*. 1991;7:189–192.
154. Fekete C, Lechan RM. Central regulation of hypothalamic-pituitary-thyroid axis under physiological and pathophysiological conditions. *Endocr Rev*. 2014;35:159–194.
155. Legradi G, Emerson CH, Ahima RS, et al. Leptin prevents fasting-induced suppression of prothyrotropin-releasing hormone messenger ribonucleic acid in neurons of the hypothalamic paraventricular nucleus. *Endocrinology*. 1997;138:2569–2576.
156. Legradi G, Emerson CH, Ahima RS, et al. Arcuate nucleus ablation prevents fasting-induced suppression of ProTRH mRNA in the hypothalamic paraventricular nucleus. *Neuroendocrinology*. 1998;68:89–97.
157. Flier JS, Harris M, Hollenberg AN. Leptin, nutrition, and the thyroid: the why, the wherefore, and the wiring. *J Clin Invest*. 2000;105:859–861.
158. Galton VA, Hernandez A, St Germain DL. The 5-deiodinases are not essential for the fasting-induced decrease in circulating thyroid hormone levels in male mice: possible roles for the type 3 deiodinase and tissue sequestration of hormone. *Endocrinology*. 2014;155:3172–3181.
159. Fliers E, Bianco AC, Langouche L, et al. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol*. 2015;3:816–825.
160. Moruzzi P, Doria E, Agostoni PG. Medium-term effectiveness of L-thyroxine treatment in idiopathic dilated cardiomyopathy. *Am J Med*. 1996;101:461–467.
161. Goldman S, McCarren M, Morkin E, et al. DITPA (3,5-diiodothyropropionic acid), a thyroid hormone analog to treat heart failure: phase II trial veterans affairs cooperative study. *Circulation*. 2009;119:3093–3100.
162. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. *Thyroid*. 2014;10:1456–1465.
163. Van den Berghe G, Baxter RC, Weekers F, et al. The combined administration of GH-releasing peptide-2 (GHRP-2), TRH and GnRH to men with prolonged critical illness evokes superior endocrine and metabolic effects compared to treatment with GHRP-2 alone. *Clin Endocrinol*. 2002;56:655–669.
164. Acker CG, Singh AR, Flick RP, et al. A trial of thyroxine in acute renal failure. *Kidney Int*. 2000;57:293–298.
165. Van den Berghe G, Wouters P, Bowers CY, et al. Growth hormone-releasing peptide-2 infusion synchronizes growth hormone, thyrotrophin and prolactin release in prolonged critical illness. *Eur J Endocrinol*. 1999;140:17–22.

84

Drugs and Thyroid

Elizabeth M. Bauer, Nicole O. Vietor, and Henry B. Burch

OUTLINE

Introduction, 1389
Drugs Affecting Hypothalamic-Pituitary Control of the Thyroid, 1389
Drugs Affecting Thyroid Hormone Synthesis or Secretion, 1395
Drugs That Enhance Thyroid Autoimmunity, 1396
Drugs Causing Direct Thyroid Damage, 1397
Drugs Affecting Protein Binding of Thyroid Hormone, 1398
  Changes in Binding Proteins, 1398
  Displacement of Thyroid Hormone from Binding Proteins, 1399
Drugs Affecting Thyroid Hormone Activation, Metabolism, and Excretion, 1399
  Thyroxine-to-Triiodothyronine Conversion, 1399
  Thyroid Hormone Metabolism, 1400
  Disposal of Thyroid Hormone, 1400
Drugs Affecting the Absorption of Thyroid Hormone Preparations, 1400
Drugs Causing Abnormal Thyroid Testing in Euthyroid Patients, 1401
The Effect of Thyroid Dysfunction on Drug Metabolism, 1403
Summary and Future Directions, 1403

INTRODUCTION

The production of thyroid hormone is an intricate process requiring multiple coordinated steps. Each step in thyroid hormone synthesis is susceptible to drug interactions, and the effects vary depending on the level of interaction (Fig. 84.1).[1] Medications affect hypothalamic-pituitary control of the thyroid, alter thyroid hormone synthesis and secretion, enhance thyroid autoimmunity, cause direct damage to the thyroid, alter protein binding to thyroid hormone, impact thyroid hormone activation, metabolism, and excretion, and diminish the absorption of thyroid hormone (Table 84.1). In addition, several common medications cause abnormal thyroid tests in euthyroid persons. Some medications interfere at multiple levels of thyroid homeostasis, which can increase the complexity of diagnosis and management. The list of medications that affect thyroid function or the interpretation of thyroid tests has grown substantially and includes prescription drugs and over-the-counter preparations (Table 84.2). Most are unintended consequences, and others have been harnessed therapeutically. This chapter will review currently known medications that interfere with thyroid hormone production, thyroid functionality, and laboratory interpretation.

DRUGS AFFECTING HYPOTHALAMIC-PITUITARY CONTROL OF THE THYROID

Metyrapone, a potent glucocorticoid inhibitor, has been shown to increase thyroid stimulating hormone (TSH), likely related to a reduction in serum cortisol. In a study of 12 healthy volunteers, metyrapone decreased morning serum cortisol by up to 47%, eliminating normal circadian variation of TSH and increasing serum TSH levels by 35% during the day, with no effect on nocturnal TSH levels.[2] TSH response to thyrotropin-releasing hormone (TRH) did not change after metyrapone administration, suggesting possible hypothalamic dysfunction as the etiology.[2]

Lithium is an alkali metal used in the treatment of bipolar affective disorders and has several effects on thyroid hormone physiology. Hypothyroidism occurs at least twice as frequently in those treated with lithium, with an annual rate of 1.5% observed in over 15 years of follow-up.[3] One metaanalysis showed a 6-fold increased risk for developing clinical hypothyroidism in lithium users.[4] The annual incidence of hypothyroidism in lithium users increases from 0.8% without thyroid antibodies to 6.4% in patients with preexisting thyroid antibodies.[3] Thyrotoxicosis also occurs more frequently in those treated with lithium.[3,5,6] This section will review the mechanism by which lithium affects the hypothalamic-pituitary-thyroid (HPT) axis. First, an exaggerated TSH response to TRH is seen in 50% to 100% of patients, and there is a significant rise in basal TSH, reflecting clinical or subclinical hypothyroidism.[5] Second, lithium concentrates in the thyroid and is a potent inhibitor of thyroid hormone release, leading to a compensatory increase in TSH concentration that may be sustained in patients with diminished thyroid reserve. Third, in rat studies lithium appears to inhibit type I (D1) and type II deiodinase 2 (D2) enzyme activity, leading to decreased triiodothyronine (T3) generation within the pituitary, further elevating TSH,[7] as well as decreased peripheral generation of T3.[5] TSH stimulation along with activation of tyrosine kinase and the Wnt/beta-catenin signaling pathways contributes to cell proliferation, resulting in goiter,[5] which is present in 5% to 59% of patients on chronic lithium therapy, depending on the population and method of detection.[8] Clinical inspection and palpation miss half of these cases compared with ultrasound, and treatment with levothyroxine (LT4) reduces the prevalence of goiter in lithium-treated patients.[9] Baseline thyroid function tests and thyroid antibodies should be performed prior to starting lithium.[7] In patients with positive antibodies, close laboratory follow-up is indicated to monitor for hypothyroidism. If

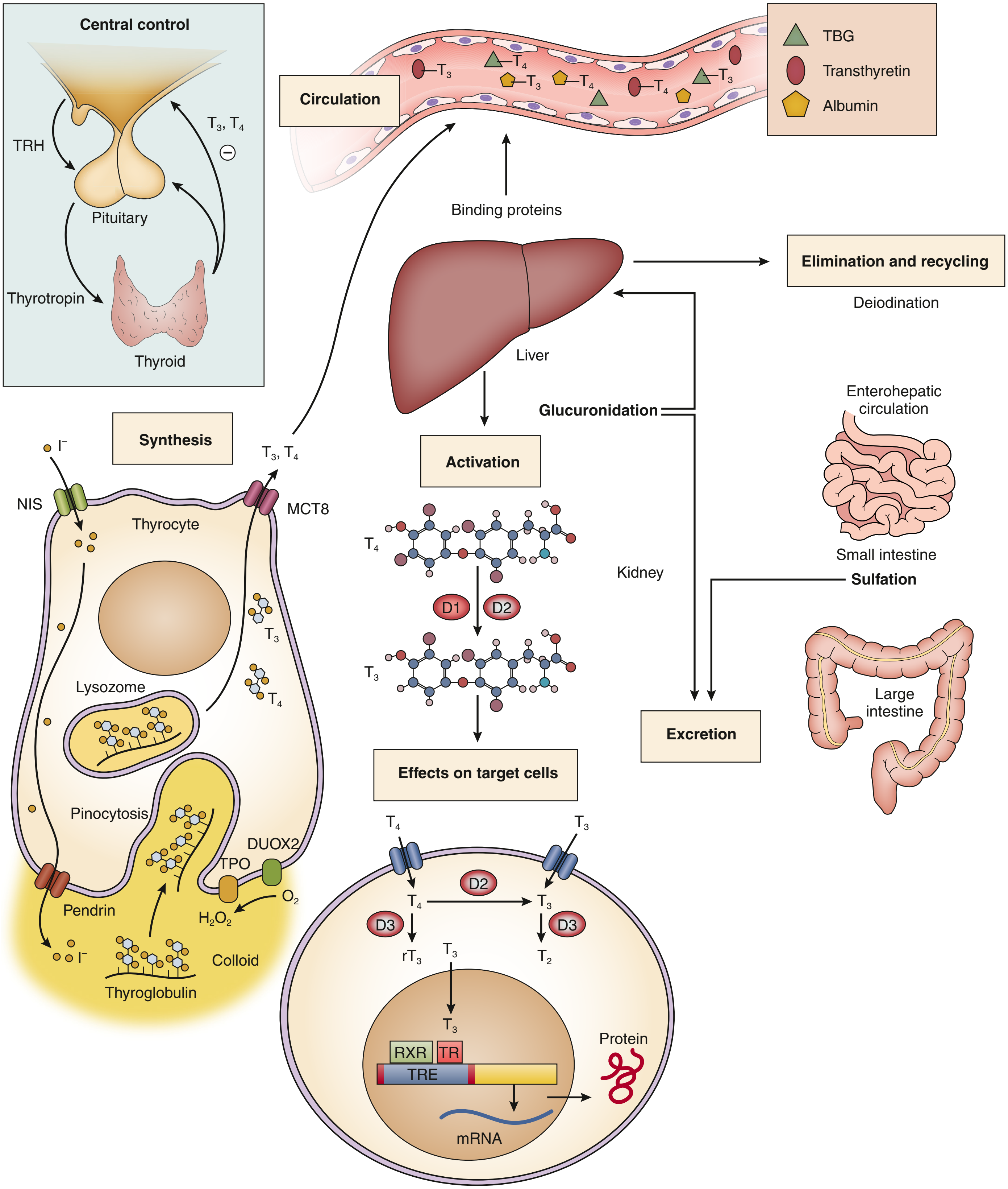

Central control
TRH
T₃, T₄
⊖
Pituitary
Thyrotropin
Thyroid
Circulation
T₃
T₄
T₃
T₄
T₄
T₃
TBG
Transthyretin
Albumin
Binding proteins
Liver
Elimination and recycling
Deiodination
Glucuronidation
Activation
Enterohepatic circulation
Small intestine
Sulfation
Kidney
T₄
D1
D2
T₃
Large intestine
Excretion
Synthesis
I⁻
NIS
T₃, T₄
MCT8
Thyrocyte
T₃
T₄
Lysozome
Pinocytosis
DUOX2
TPO
O₂
H₂O₂
Pendrin
I⁻
Colloid
Thyroglobulin
Effects on target cells
T₄
T₃
D2
T₄
T₃
D3
D3
rT₃
T₃
T₂
T₃
RXR
TR
TRE
Protein
mRNA

Fig. 84.1 Thyroid hormone physiology. Hypothalamic thyrotropin-releasing hormone (TRH) stimulates the release of thyrotropin (TSH) from pituitary thyrotrophs . This in turn stimulates the synthesis of thyroid hormone within thyroid follicles. Thyroid follicles consist of hundreds of thyrocytes surrounding a follicular lumen filled with colloid matrix. Thyroglobulin is produced in the thyrocyte and subsequently secreted into the follicular lumen. Thyroglobulin serves as the backbone for thyroid hormone synthesis in the follicle. Iodide is actively transported into the thyrocyte by the sodium-iodide symporter (NIS) and then into the follicular lumen through the anion exchange protein pendrin. In a hydrogen peroxide-dependent reaction,

thyroid peroxidase (TPO) performs several key reactions that oxidize and organify iodine, linking it to specific tyrosine residues on thyroglobulin. When two tyrosine moieties are coupled, thyroxine (T4) and triiodothyronine (T3) molecules are formed. Pinocytosis of the iodinated thyroglobulin allows lysosomal proteolysis with subsequent release of T4 and T3 into the circulation using the monocarboxylate transporter 8 (MCT8). Circulating thyroid hormone is >99% bound to carrier proteins, with thyroxine-binding globulin (TBG) accounting for approximately 70% of binding, followed by transthyretin and albumin. Local conversion of T4 to T3 at the tissue level occurs through 5′-monodeiodination of T4 by 5′-monodeiodinase type 2 (D2) and type 1 (D1). T3 binds to the thyroid hormone receptors (TR), causing a nuclear effect, which, along with its retinoid X receptor (RXR) can bind to thyroid hormone response elements (TREs). Deactivation (creation of reverse T3 from T4 and T2 from T3) also occurs at the tissue level through the actions of 5′monodeiodination type 3 (D3). Pituitary T3 is obtained from both systemic circulation and intrapituitary D2, which then provides feedback to the hypothalamus and pituitary. Thyroid hormone disposal is primarily performed through successive deiodination and conjugation, including glucuronidation and sulfation in the liver and kidney. These actions increase water solubility and subsequent secretion into the bile. Glucuronidated thyroid hormone can be eliminated in feces or recycled in the enterohepatic circulation.

TABLE 84.1 Classification of Drug Effects on the Thyroid

- Drugs affecting hypothalamic-pituitary control of the thyroid
- Drugs affecting thyroid hormone synthesis or secretion
- Drugs that enhance thyroid autoimmunity
- Drugs causing direct thyroid damage
- Drugs affecting protein binding of thyroid hormone
  - Changes in binding proteins
  - Displacement of thyroid hormone from binding proteins
- Drugs affecting thyroid hormone activation, metabolism, and excretion
  - Thyroxine-to-triiodothyronine conversion
  - Thyroid hormone metabolism
  - Disposal of thyroid hormone
- Drugs affecting the absorption of thyroid hormone preparations
- Drugs causing abnormal thyroid testing in euthyroid patients
- The effect of thyroid dysfunction on drug metabolism

hypothyroidism develops, thyroid supplementation can be initiated without the need to stop lithium.[5,7,8]

Opiates decrease basal TSH and the TSH response to TRH, while T3 and thyroxine (T4) concentrations typically remain unchanged with chronic opiate use.[10]

Immune checkpoint inhibitors (ICIs) regulate immune response by modulating T-cell antitumor activity. There are three main ICI categories, including cytotoxic T-lymphocyte antigen 4 (CTLA-4), programmed cell death-1 (PD-1), and programmed cell death-ligand 1 (PD-L1) inhibitors. CTLA-4 inhibitors are primarily used to treat malignant melanoma, and PD-1 inhibitors are used in the treatment of various solid tumors. Multiple endocrine toxicities are seen with use of ICI medications; here, we will discuss hypophysitis resulting in central thyroid dysfunction. Enhanced thyroid autoimmunity and destructive thyroiditis will be discussed later in the chapter.

Hypophysitis leading to central hypothyroidism is a serious side effect of ICI therapy and occurs most frequently with combination therapy (CTLA-4/PD-1 or PD-L1 blockade). Monotherapy with CTLA-4 inhibitors, specifically ipilimumab, has also been associated with destructive hypophysitis in 10% or more of patients treated.[1] This side effect is rarely seen with PD-1/PD-L1 inhibitor monotherapy, where the incidence is less than 0.6%.[11] The precise mechanism of destructive hypophysitis remains to be elucidated.[12] A review by Byun et al. found an overall incidence of hypophysitis in 9.1% using strict criteria for excluding primary dysfunction, but reports have ranged from 0.4% to 17%, with a higher incidence in men compared with women (15.6% vs. 3.6%).[11] Decreasing TSH levels may be an early sign of impending hypophysitis.[13] Among 184 patients developing hypophysitis during ICI therapy, the incidence of secondary hypothyroidism was 7.6% within the first 4 to 16 weeks of therapy.[11] Thyroid labs will show a normal TSH with low thyroid hormones, which can be confused with non-thyroidal illness (NTI). As a result, clinical and hormonal monitoring are critical prior to and throughout ICI therapy. Immunotherapy is not usually stopped with the discovery of hypophysitis, but has occasionally been delayed.[13] Thyroid function recovers in approximately 50% to 60% of patients after months to years following discontinuation of therapy.[12,13] Recovery rates are highest with longer periods of follow-up.[13]

KEY POINTS: Immune Checkpoint Inhibitors (anti-CTLA-4, PD-1, PD-L1)

- Hypophysitis leading to central hypothyroidism is a serious side effect of immune checkpoint inhibitor (ICI) therapy.
- Decreasing thyroid-stimulating hormone (TSH) levels may be an early sign of impending hypophysitis.
- Thyroid function recovers in approximately 50%–60% of patients after months to years following discontinuation of therapy.
- Primary thyroid dysfunction occurs in 5%–10% of patients treated with CTLA-4 inhibitors, 10%–20% treated with PD-1 inhibitors, and >20% treated with combination therapy.
- Destructive thyroiditis is treated with LT4 without interruption in ICI therapy. Preexisting thyroid antibodies increase the risk of developing thyroid dysfunction.

Dopaminergic agents have an inhibitory effect on TSH secretion.[14] Dopamine infusion at a dose of 4 ug/kg has been shown to suppress mean TSH levels and TSH pulse amplitude by 70%.[15] Studies in critically ill adults requiring dopamine infusions show a suppressive effect on the HPT axis that may be superimposed upon changes due to NTI.[16] Dopamine agonist therapy, with bromocriptine or cabergoline, has also been examined in this regard. In a study of 17 obese women treated with bromocriptine compared to placebo, both pulsatile and nonpulsatile TSH secretion was decreased, but free T4 (FT4) and total T3 (TT3) concentrations did not change.[17] Conversely, in a study of nine patients with prolactinomas treated with cabergoline, significant reductions in T4 and T3 levels were reported, with a slight but insignificant decrease in TSH.[18] Two of the nine patients in this study required LT4 treatment based on symptoms and low T4 levels.[18] Considering these results, the inhibition of TSH secretion with dopamine agonist therapy is variable, often clinically insignificant, and likely related to the dose and specific medication.

Data on the effect of the dopamine antagonist metoclopramide on TSH secretion is conflicting. In a study of 11 healthy volunteers treated with intravenous (IV) metoclopramide, 24-hour mean TSH levels were found to be increased.[19] However, in another study of 17 women treated with oral metoclopramide, plasma TSH and thyroid hormone levels did not change.[20]

TABLE 84.2 **Drugs Influencing Endogenous Thyroid Function**

Representative Drug	Drug Class	Nature of Interaction	Severity Grade[1]	Suggested Monitoring Interval[2] or Comment
Drugs Affecting Hypothalamic-Pituitary Control of the Thyroid				
Metyrapone	11-beta-hydroxylase inhibitor	TSH increase	1	None or annually
Lithium	Element	TSH increase	1–2	Quarterly
Opiates	Opioids	TSH decrease	1	None
Ipilimumab Tremelimumab	Anti-CTLA-4, Immune checkpoint inhibitors	Hypophysitis	3–5	Monthly or with each cycle
Bromocriptine	Dopamine agonists	TSH suppression	1	None or annually
Metoclopromide	Dopamine antagonist	TSH increased or normal	1	None
Octreotide	Somatostatin analogs	TSH suppression	1	None or annually
Dobutamine	Inotrope	TSH suppression	1	None
Bexarotene	Retinoid	Central hypothyroidism	3–4	Monthly
Mitotane	Adrenolytic agent	TSH suppression	2	Quarterly
Dexamethasone Hydrocortisone Methylprednisolone	Glucocorticoids	TRH/TSH suppression	1–2	None or monthly x 1
Metformin	Biguanide	TSH suppression	1	None or annually
Drugs Affecting Thyroid Hormone Synthesis or Secretion				
Iodine	Micronutrient	Decreased TH synthesis Iodine-induced hyperthyroidism	1–4	None or quarterly x 1
Amiodarone	Class III antiarrhythmic	Decreased TH synthesis Iodine-induced hyperthyroidism	2–5	Semiannually
Lithium	Element	Decreased TH release	1–2	Quarterly
Aminoglutethimide	Steroidogenesis inhibitor	Decreased TH synthesis	1–2 (dose-dependent)	Periodically with high doses
Sunitinib	Tyrosine kinase inhibitor	Impairs cellular uptake of TH by MCT8 inhibition	3–5	Quarterly
Drugs that Enhance Thyroid Autoimmunity				
Alemtuzumab	Anti-CD52	*De novo* development of thyroid antibodies	3–4	Semiannually
Ipilimumab Nivolumab	Immune checkpoint inhibitors	Reduction in immune self-tolerance	3–5	Monthly or with each cycle
Interferon-α Interleukin-2	Cytokine	*De novo* development of thyroid antibodies	2	Semiannually
Drugs Causing Direct Thyroid Damage				
Amiodarone	Class III antiarrhythmic	Direct cytotoxicity	3–5	Quarterly to semiannually
Nivolumab Pembrolizumab	Anti-PD-1, Immune checkpoint inhibitors	Immune thyroiditis	3–5	Monthly or with each cycle
Interleukin-2	Cytokine	Immune thyroiditis	2	Semiannually
Interferon-α	Cytokine	Immune thyroiditis	2	Semiannually
Sunitinib Sorafenib	Tyrosine kinase inhibitor	Immune thyroiditis	3–5	Quarterly
Lithium	Element	Thyroiditis, unknown association	1–2	Quarterly
Drugs Affecting Protein Binding of Thyroid Hormone				
Changes in Binding Proteins				
Oral estrogen	Sex steroid	Increased thyroxine-binding globulin	2	None (unless taking thyroid hormone)
Tamoxifen Raloxifene	Selective estrogen receptor modulator	Increased thyroxine-binding globulin	1	None (unless taking thyroid hormone)
Heroin Methadone	Opioids	Increased thyroxine-binding globulin	1	None
Mitotane	Adrenolytic agent	Increased thyroxine-binding globulin	1	None
5-Fluorouracil	Antimetabolite	Increased thyroxine-binding globulin	1	None (unless taking thyroid hormone)

TABLE 84.2 Drugs Influencing Endogenous Thyroid Function—cont'd

Representative Drug	Drug Class	Nature of Interaction	Severity Grade[1]	Suggested Monitoring Interval[2] or Comment
Oral androgen	Sex steroid	Decreased thyroxine-binding globulin	1	None (unless taking thyroid hormone)
Danazol	Synthetic steroid	Decreased thyroxine-binding globulin	1	None
Oral glucocorticoids	Glucocorticoids	Decreased thyroxine-binding globulin	1–2	None or monthly x 1
Niacin	Vitamin B3	Decreased thyroxine-binding globulin	1	None
Displacement of Thyroid Hormone from Binding Proteins				
Phenytoin Carbamazepine	Antiepileptic	Laboratory changes	1–2	Annually
Salsalate Fenclofenac	Nonsteroidal antiinflammatory	Laboratory changes	1	None
Furosemide	Loop diuretic	High dose only	1	None
Drugs Affecting Thyroid Hormone Activation, Metabolism and Excretion				
Drugs that inhibit T4 toT3 conversion				
Amiodarone	Class III antiarrhythmic	Laboratory changes	2 (Note: greater severity in other types of interaction)	Semiannually
Propranolol Metoprolol Atenolol	Beta-adrenergic blocker	Reduced T3 (high dose only)	1	None
Dexamethasone	Glucocorticoid	Reduced T3	2	None or monthly x 1
Drugs that Affect Thyroid Hormone Metabolism				
Phenobarbital Phenytoin Carbamazepine	Antiepileptic	Increased hepatic conjugation	1	Annually
Rifampin	Antimicrobial	Increased hepatic conjugation	1	Annually
Imatinib Motesanib Sorafenib Sunitinib Vandetanib	Tyrosine kinase inhibitor	Increased type 3 deiodinase activity	2	Quarterly
Bexarotene	Retinoid	Increased sulfation	2 (Note: greater severity in other types of interaction)	Monthly
Drugs Affecting Disposal of Thyroid Hormone				
Cholestyramine Colestipol Colesevelam	Bile acid sequestrant	Decreased enterohepatic recycling	1–2	None
Drugs Affecting the Absorption of Thyroid Hormone Preparations				
Iron sulfate	Mineral	Decreased absorption	1–2	Consider repeating labs 4 weeks after starting or stopping
Calcium carbonate	Antacid, supplement	Decreased absorption	1–2	Consider repeating labs 4 weeks after starting or stopping
Aluminum hydroxide	Antacid	Decreased absorption	1–2	Consider repeating labs 4 weeks after starting or stopping
Cholestyramine Colesevelam Colestipol	Bile acid sequestrant	Decreased absorption and reduced enterohepatic recycling	1–2	Consider repeating labs 4 weeks after starting or stopping
Sucralfate	Sucrose sulfate-aluminum complex	Decreased absorption	1–2	Consider repeating labs 4 weeks after starting or stopping

Continued

TABLE 84.2 Drugs Influencing Endogenous Thyroid Function—cont'd

Representative Drug	Drug Class	Nature of Interaction	Severity Grade[1]	Suggested Monitoring Interval[2] or Comment
Raloxifene	Selective estrogen receptor modulator	Decreased absorption	1–2	Consider repeating labs 4 weeks after starting or stopping
Lansoprazole Omeprazole	Protein pump inhibitor	Decreased pill dissolution	1–2	Consider repeating labs 4 weeks after starting or stopping
Drugs Causing Abnormal Thyroid Testing in Euthyroid Patients				
Biotin	Vitamin B7	Laboratory Changes	1	None
Amiodarone	Class III antiarrhythmic	Laboratory changes	1	None
Phenytoin Carbamazepine	Antiepileptic	Laboratory changes	1	None
Salsalate	Nonsteroidal antiinflammatory	Laboratory Changes	1	None
Heparin	Anticoagulant	Laboratory Changes	1	None
The Effect of Thyroid Dysfunction on Drug Metabolism				
Warfarin	Anticoagulant	Hyperthyroidism increases catabolism of vitamin K–dependent clotting factors, prolonging PT	3–5	Dose may need to be decreased in hyperthyroid patients
Digoxin	Cardiac glycoside	Enhanced (hyperthyroidism) or delayed (hypothyroidism) clearance of medication	3–5	Dose may need to be adjusted in hyperthyroid or hypothyroid patients
Acetaminophen	Analgesic	Increased absorption, earlier peak, more rapid decrease with hyperthyroidism	1	Dose may need to be adjusted in hyperthyroid patients
Propranolol	Beta-blocker	Hepatic first-pass metabolism increased (hyperthyroidism) or decreased (hypothyroidism)	1	Dose may need to be adjusted in hyperthyroid patients
Glucocorticoids	Glucocorticoid	Hepatic clearance increased (hyperthyroidism) or decreased (hypothyroidism)	1–4	Increased metabolism of corticosteroids during thyrotoxicosis could precipitate adrenal crisis in a patient dependent on adrenal replacement therapy/ patients with borderline adrenal function
Statin medications	Statin	Increased risk of adverse events in hypothyroidism (myalgias, myositis, rhabdomyolysis)	1–3	Ensure patient is euthyroid prior to starting or increasing statin dose

[1] Severity grade: Adapted from *National Cancer Institute Common Terminology Criteria for Adverse Events v5.0*. 27 Nov. 2017, ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf.
Grade 1, mild: asymptomatic or mild symptoms; clinical or diagnostic observation only; intervention not generally indicated.
Grade 2, moderate: intervention indicated, potentially limiting activities of daily living.
Grade 3, severe: medically significant but generally not immediately life-threatening; potentially requiring hospitalization or prolongation of hospitalization.
Grade 4, life-threatening consequences: urgent intervention required.
Grade 5, death related to adverse event.

[2] Monitoring intervals are based on the timing, frequency, and severity (grade) of the drug interaction. When available, data were derived from clinical practice guidelines.

TSH, Thyroid-stimulating hormone; *CTLA-4*, cytotoxic T-lymphocyte-associated protein; *TRH*, thyrotropin-releasing hormone; *TH*, thyroid hormone; *PD-1*, programmed cell death protein 1; *T3*, triiodothyronine; *T4*, thyroxine; *LT4*, levothyroxine; *MCT8*, monocarboxylate transporter 8.

(From Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.)

Somatostatin analogs bind to multiple receptors on pituitary cells and directly affect hormone secretion. The effects on thyroid function appear to depend on the duration of treatment. Administration of somatostatin to healthy volunteers resulted in decreased serum TSH pulse frequency and amplitude.[21] In a study of 25 patients with acromegaly started on octreotide, transient acute changes occurred, including decreased TSH response to TRH and a subsequent decrease in T3 levels, but these changes reverted to baseline during continued therapy.[22] In summary, somatostatin analogs directly suppress TSH through direct effects on pituitary thyrotrophs, but the effects are transient and do not appear to cause clinically significant central hypothyroidism.

Dobutamine is an inotrope that acts on cardiac β1-adrenergic receptors and weakly on peripheral β2- and α-receptors. Clinically, this agent is used for the treatment of congestive heart failure to increase cardiac output. Dobutamine doses of 5 μg/kg/minute or less do not affect TSH levels,[23] but higher doses are associated with a small but significant decrease in TSH, detectable within 15 minutes and persisting for 15 minutes after drug discontinuation.[24] Dobutamine infusion has also been shown to increase FT3 levels in patients with decompensated heart failure, which may be in part related to a reversal of the acute effects of NTI by treating the underlying disorder. For example, after an improvement in hemodynamic status with dobutamine, all patients had improvement in low T3, with 42% attaining normal T3 levels.[25]

Bexarotene is a vitamin A analog specific to rexinoid receptors and used in the treatment of T-cell lymphomas. Bexarotene causes a rapid suppression of TSH, leading to central hypothyroidism in 40% to 70% of patients, in a dose-related fashion.[26,27] This effect is likely related to direct effects on pituitary thyrotrophs, as well as by enhanced thyroid hormone metabolism through nondeiodinase pathways,[28] which will be discussed later in this chapter. Normal thyroid function generally returns after bexarotene discontinuation.[29] Isotretinoin therapy in acne vulgaris patients has been examined for similar effects, but only minor and clinically insignificant changes in TSH, T3, and T4 levels were observed.[30]

Mitotane is a synthetic derivative of the insecticide dichlorodiphenyltrichloroethane with adrenolytic properties that result in impaired steroidogenesis. It is used to treat adrenocortical carcinoma. Mitotane can cause both primary and secondary thyroid dysfunction. Central hypothyroidism is usually seen in the first 3 months of therapy with a normal TSH, impaired TSH response to TRH, and low FT4. Free T3 (FT3) generally remains normal. FT4 reductions are inversely proportional to mitotane concentration.[31] Long-term LT4 replacement is often required.[32] Mitotane may also increase thyroxine-binding globulin (TBG) and competes with T4 for TBG binding sites. The precise etiology of central hypothyroidism in this setting is currently unknown.[31]

Glucocorticoids have multiple effects on the thyroid gland, which will be discussed throughout this chapter. Here, we will discuss the effects on the HPT axis. TSH secretion is inhibited by both physiologic and therapeutic glucocorticoid administration.[2,33] Higher doses of hydrocortisone eliminate the nocturnal TSH surge, reduce 24-hour TSH secretion, and decrease the pulse amplitude of TSH release, with minimal changes in T4 levels.[33] Dexamethasone has been shown to reduce TSH in a dose-dependent fashion.[34,35] In one study, older individuals required higher doses of dexamethasone to inhibit the TSH response to TRH than that seen in younger research participants.[35] Overall, the potency, dose, and route of administration of glucocorticoid influence the magnitude of the effect on thyrotrophs. Importantly, chronic high-dose glucocorticoid or other causes of Cushing syndrome do not appear to cause clinically evident central hypothyroidism or require thyroid hormone supplementation.[34]

Metformin was found to reduce TSH in patients with overt and subclinical hypothyroidism, but not in euthyroid patients.[36,37] The suppressive effect on TSH is possibly due to inhibitory action of metformin on AMP-activated protein kinase in the hypothalamus, as well as other potential mechanisms.[38] No metformin-related changes in absorption of thyroid hormone have been found.[39] Thyroid hormone levels are unchanged, and although the TSH decrease is statistically significant, it does not suppress to levels that would be concerning for central hypothyroidism.[38]

DRUGS AFFECTING THYROID HORMONE SYNTHESIS OR SECRETION

Iodine is a trace element that is required for thyroid hormone synthesis. A daily intake of 150 μg is recommended by the World Health Organization for all nonpregnant or lactating adults. However, many prescriptions, over-the-counter supplements, topical agents, and foods contain iodine concentrations that far exceed the daily recommended intake. Examples are amiodarone, expectorants/vaginal douches, povidone iodine, and kelp.[40] The physiologic response to iodine excess is inhibition of thyroid peroxidase enzyme activity, which results in decreased synthesis of thyroid hormone. This phenomenon is known as the Wolff–Chaikoff effect. In a normal thyroid gland, regulatory mechanisms exist to maintain a euthyroid state despite wide fluctuations in iodine intake. A healthy thyroid gland will escape the Wolff–Chaikoff effect and resume hormone synthesis within days by decreasing the expression of sodium-iodide symporter, thus limiting the influx of iodine into the thyroid. However, patients with decreased thyroid reserve, as occurs with chronic thyroiditis, partial thyroidectomy, or thyroid irradiation, are unable to escape the Wolff–Chaikoff effect, which can result in a more prolonged period of hypothyroidism. Patients who have underlying thyroid autonomy, such as those with toxic adenoma or toxic multinodular goiter, are susceptible to iodine-induced hyperthyroidism, a phenomenon known as Jod-Basedow.

Amiodarone's effects on the thyroid can be classified as either iodine-related effects or intrinsic drug effects. Intrinsic amiodarone drug effects include disruption of thyroid hormone synthesis, direct damage to thyroid follicular cells, and inhibition of both peripheral and central monodeiodination. Additionally, amiodarone and its metabolites have structural similarities to T3 and T4, allowing them to potentially interact with thyroid hormone receptors.[41] Finally, there are acute and chronic effects on thyroid function laboratory testing in amiodarone-treated patients who remain euthyroid.[42] Intrinsic drug effects will be discussed later in the chapter; here we will focus on the iodine-related effects. Fig. 84.2 shows the direction of thyroid laboratory changes in euthyroid, thyrotoxic, and hypothyroid patients treated with amiodarone.

Amiodarone is classified as a class III antiarrhythmic (a potassium channel blocker), but also has some class I, II, and IV effects by affecting sodium, β-adrenergic, and calcium channels making it a very potent antiarrhythmic.[41] The drug has a long (>100-day) half-life, mainly due to accumulation in adipose tissue, but can also accumulate in the liver, lung, muscle, and thyroid. A 200-mg dose releases 75 mg of organic iodine, leading to the release of approximately 6 to 7 mg of free iodine into the circulation. This is 30 to 45 times greater than the average American daily dietary intake (150–200 mcg).[41,42] As such, amiodarone can cause both iodine-induced hypothyroidism and hyperthyroidism through mechanisms described earlier in this chapter for iodine. Predictors of amiodarone-induced thyroid dysfunction include female sex, preexisting autoimmune thyroiditis, and nodular thyroid disease.[41]

Amiodarone-induced hypothyroidism (AIH) is more common in areas of iodine deficiency than regions of iodine sufficiency, and the reverse is true with amiodarone-induced thyrotoxicosis (AIT). AIH occurred in 22% of amiodarone patients in Worcester, Massachusetts but only 5% of those from the low-iodine region of Tuscany, Italy.[43] Conversely, AIT occurred in 9.6% of amiodarone-treated patients in Tuscany, but only 2% in Worcester.[43] AIH can occur in patients without underlying thyroid disease but is generally seen in patients with diminished thyroid reserve who are unable to escape the Wolff–Chaikoff effect. Patients with AIH present with elevated TSH and a low FT4 levels, signifying primary hypothyroidism. Hypothyroidism in this setting is easily addressed with thyroid hormone replacement therapy, and therefore amiodarone does not need to be discontinued.[41]

Iodine-induced hyperthyroidism in patients taking amiodarone is referred to as type-1 amiodarone-induced thyrotoxicosis (AIT1). Amiodarone can have direct toxic effects on thyrocytes, leading to a destructive thyroiditis, known as type 2 amiodarone-induced thyrotoxicosis (AIT2), which will be discussed later in this chapter. AIT1 is diagnosed in patients presenting with a suppressed TSH and elevated levels of T4 and T3.

Euthyroid	Hyperthyroid	Hypothyroid
↑ Free T4	↑ Free T4	↓ Free T4
↑ TSH (typically <10 mU/L and transient)	↓ TSH (undetectable)	↑ TSH (typically >10)
↓ Total T3	↑ Total T3	↓ Total T3
	+/– Antithyroid antibodies	+ Antithyroid antibodies

Fig. 84.2 **Amiodarone can cause thyroid test abnormalities in a euthyroid person or induce clinical hyperthyroidism (AIT) or hypothyroidism (AIH).** The figure shows the expected thyroid test results in each of these situations. Amiodarone can induce thyrotoxicosis with two classical variants, AIT1 and AIT2. Often, however, these two variants overlap, and diagnostic uncertainty can exist. AIT1 is due to the iodine-related effects of amiodarone and often occurs in patients with underlying thyroid autonomy, such as toxic multinodular goiter or Graves disease. AIT2 occurs due to the intrinsic effects of amiodarone that cause direct damage to the thyroid. This often occurs in patients without a prior history of thyroid abnormality. AIH occurs due to the iodine-related effects of amiodarone and occurs most commonly in individuals with underlying chronic thyroiditis.

Treatment of AIT1 is directed at controlling increased thyroid hormone synthesis using thionamides, and occasionally potassium perchlorate. Definitive therapy for the underlying disease process is recommended after achieving a euthyroid state, particularly in those for whom continued amiodarone therapy is required. In patients with mixed AIT1 and AIT2 or indeterminate features, both glucocorticoids and methimazole are recommended, particularly if underlying cardiovascular compromise is present.[44] In a small subset of patients with rapid clinical deterioration who are unresponsive to medical therapy, total thyroidectomy may be required to achieve a euthyroid status.[41]

Lithium accumulates in thyroid tissue, leading to concentrations three to four times that of plasma. Lithium decreases release of thyroid hormone from the thyroid, most likely through disruption of effective pinocytosis.[5] The ability of lithium to increase iodine retention has been utilized as adjunctive therapy to radioiodine to both maximize delivery of radioactive iodine to the thyroid or thyroid cancer and to limit whole-body radiation dose; however, it is rarely practical to do so. Lithium has also been purported to decrease the risk of aggravation of thyrotoxicosis in the first 1 to 2 weeks after radioiodine therapy for Graves disease.[45] Current American Thyroid Association guidelines state that there is insufficient evidence to support the use of lithium as adjunctive therapy to radioiodine for either benign or malignant thyroid disease.[44,46]

In addition to alteration of iodine retention by the thyroid, lithium affects thyroid hormone synthesis through a number of additional mechanisms, including interference with tyrosine iodination, alteration of thyroglobulin structure, and inhibition of colloid formation.[7,8] As a result of these antithyroid properties, lithium has been utilized in the treatment of thyrotoxicosis.[47] Lithium use in this setting is reserved for thyrotoxic individuals who do not tolerate or respond to thionamides. The typical doses required are relatively low, such as 750 mg per day,[5] but monitoring of lithium levels is required to avoid toxicity.

Aminoglutethimide, a steroidogenesis inhibitor, blocks multiple CYP enzymes, resulting in inhibition of both androgen and corticosteroid synthesis. It has typically been used for treatment of Cushing syndrome, adrenal malignancy, and breast and prostate cancer. Aminoglutethimide was previously used as an anticonvulsant but was withdrawn from the market due to adverse effects, including goiter and clinical hypothyroidism in children. This drug induces hypothyroidism through inhibition of T4 synthesis, with a compensatory increase in TSH.[48] This effect appears to be dose-dependent. Daily doses of 250 mg or less have minimal effects on thyroid function, but doses of 1000 mg/day, as have been used to treat prostate cancer, cause marked TSH elevation in approximately 31% of patients.[49] These patients are clinically hypothyroid and require LT4 replacement therapy.

Tyrosine kinase inhibitors (TKIs) are proteins developed as chemotherapy agents to block ATP-binding sites on tyrosine kinase receptors, causing inhibition of signal transduction. The result is alteration of multiple tumor cell processes, including cell proliferation, survival, invasiveness, and angiogenesis, depending on the target. Although the predominant etiology of TKI-induced thyroid dysfunction occurs through destructive thyroiditis and effects on hepatic metabolism (discussed later in the chapter), these drugs also impair cellular uptake of thyroid hormone by inhibiting monocarboxylate transporter 8 (MCT8). This leads to a reduced supply of T3 to peripheral and central cells.[50]

DRUGS THAT ENHANCE THYROID AUTOIMMUNITY

Alemtuzumab, a monoclonal antibody targeting CD52 cell surface markers on lymphocytes and monocytes, is commonly used in the treatment of multiple sclerosis (MS), leukemia, and rheumatoid arthritis. Interestingly, thyroid dysfunction occurs when alemtuzumab is used in the treatment of MS,[51,52] but not when used in the treatment of leukemia[53] or rheumatoid arthritis.[54] In one study, 73 of 216 (33.8%) alemtuzumab-treated patients with MS developed thyroid dysfunction, and among those 65.8% (22.2% overall) developed Graves disease.[51] At baseline, 11 of 73 (15.1%) were thyroid peroxidase (TPO) antibody-positive, and 40 of 73 (54.8%) developed TPO antibody positivity *de novo*.[51] In another study, 102 of 248 (41.1%) patients with MS treated with alemtuzumab developed thyroid dysfunction, and among those 71.6% (29.4% overall) developed Graves disease.[52] The median interval between the final alemtuzumab dose and the onset of Graves disease was 17 months[52]; however, Graves disease associated with alemtuzumab has occurred up to 9 years following completion of therapy.[51,52] Patients with MS treated with alemtuzumab may less commonly develop destructive thyroiditis and hypothyroidism. Spontaneous resolution of overt thyrotoxicosis due to Graves disease occurred in 36.7% of alemtuzumab-treated patients, while resolution of overt hypothyroidism occurred in 17.4% in one study.[51] Smoking, family history of autoimmune disease, and positive pretreatment TPO antibodies are risk factors for developing thyroid dysfunction in patients with MS treated with alemtuzumab.[51,52]

KEY POINTS: Alemtuzumab

- Thyroid dysfunction has been shown to occur when alemtuzumab is used in the treatment of multiple sclerosis (MS).
- The most common thyroid manifestation is Graves disease, but destructive thyroiditis and hypothyroidism may also occur.
- Smoking, family history of autoimmune disease, and positive pretreatment thyroid peroxidase antibodies are risk factors for developing thyroid dysfunction in patients with MS treated with alemtuzumab.

ICIs, including CTLA-4 inhibitors and PD-1 inhibitors, are designed to promote immune system eradication of cancer cells, but also increase the risk of autoimmunity. Primary thyroid dysfunction during immune checkpoint blockade is believed to result from a reduction in immune self-tolerance caused by these drugs.[55,56] In 40% to 50% of affected patients, TPO or thyroglobulin antibodies are present at the time of thyroid disruption,[57,58] but their pathophysiological role is unknown, as most series do not include pretreatment antibody levels. Some studies show increases in preexisting antibodies[59,60] or antibody development[61] during immune checkpoint blockade. Positive pretreatment thyroid antibodies increase the risk of drug-induced thyroid dysfunction.[59,61] Destructive thyroiditis related to immune checkpoint blockade is discussed later in this chapter.

Cytokine therapy. *De novo* development of thyroid antibodies has occurred during therapy with the cytokines interleukin-2 (IL-2) and interferon-α (IFN-α). IL-2 is a T-cell growth factor that acts to promote CD8 and natural killer cell activity[62] and is used in the treatment of metastatic melanoma and renal cell carcinoma. In a series of 89 patients being treated with IL-2, *de novo* development of thyroid antibodies occurred in 11 of 18 (61%) patients with negative antibodies before treatment, including six who became transiently positive and five who became permanently positive.[63]

Approximately 5% to 15% of patients using IFN-α develop clinical autoimmune thyroid disease, as part of a larger number developing thyroid antibodies without clinical dysfunction.[64] In one systemic review, studies providing sufficient data to calculate the incidence of *de novo* development of thyroid antibodies showed that this occurred in 135 of 825 (16.4%) patients treated with IFN-α.[65] These antibodies can develop during or after completion of therapy and may persist long term. In one report, 72% of patients developing thyroid antibodies during treatment with IFN-α continued to have positive antibodies at a median of 6.2 years of follow-up after completing therapy.[66] Cytokine therapy causing destructive thyroiditis is discussed in the next section of this chapter.

DRUGS CAUSING DIRECT THYROID DAMAGE

Amiodarone also causes destructive thyroiditis in 5% to 10% of individuals without underlying thyroid disease, a complication known as AIT2. AIT2 has been postulated to result from the proapoptotic and cytotoxic effects of amiodarone and its main metabolite desethylamiodarone. As with other forms of thyroiditis, the thyrotoxic phase results from leakage of preformed hormone into circulation after direct follicular cell damage, followed by a hypothyroid phase before full or partial recovery. Those with preexisting diminished thyroid reserve, such as occurs with chronic lymphocytic thyroiditis, may not achieve full recovery and will require thyroid supplementation.[41,42] Differentiating between AIT1 and AIT2 is challenging, and indeterminate forms may have overlapping features.[44] In general, AIT1 occurs in individuals with underlying thyroid disease such as toxic nodules or Graves disease. On exam, a goiter may be present with increased vascularity seen on Doppler flow ultrasonography. Conversely, patients with AIT2 often have normal or decreased blood flow to the thyroid. Iodine uptake measurements may be increased with AIT1, but published results are conflicting. ^{99m}Tc-sestamibi scanning may help to differentiate AIT1 from AIT2, with increased uptake in those with AIT1 and low or absent in AIT2.[67,68] Mixed or indeterminate AIT has been considered in those with persistent uptake or rapid washout of tracer.[41,42]

Under ideal circumstances, AIT1 would be treated with methimazole, and AIT2 would be treated with glucocorticoids. Given frequent diagnostic uncertainty, the following management approach has been recommended by the American Thyroid Association.[44] Patients presenting with AIT should first be evaluated for definite evidence of underlying thyroid disease that would suggest AIT1, such as Graves disease or toxic multinodular goiter. If these etiologies are present, and the patient is without evidence of acute cardiovascular compromise, the patient should be treated initially with methimazole alone. In patients presenting with cardiovascular compromise, combination therapy with methimazole and glucocorticoids should be started immediately in order to avoid delay in the delivery of appropriate therapy resulting from possible misclassification. For most patients without evidence of underlying Graves disease or toxic nodular disease and without initial cardiovascular compromise, glucocorticoid therapy should be the initial therapy. Treatment is targeted at reducing the inflammatory process with moderate doses of oral glucocorticoids such as prednisone 0.5mg/kg/day. Discontinuation of amiodarone is often considered but not absolutely required, particularly if it is the only antiarrhythmic therapy showing efficacy in the patient under evaluation. Rapid improvement in AIT2 can occur within 1 to 2 weeks after starting glucocorticoids, with 60% of patients in one study returning to a euthyroid state within 30 days, although 16% required more than 90 days for resolution.[69] Patients failing to respond rapidly to corticosteroid therapy often have methimazole added to their corticosteroid regimen.[44] In patients who are refractory to all medical therapy, total thyroidectomy is occasionally required to rapidly restore euthyroidism,[41,42] but this group of patients is at increased risk of perioperative complications.[70] Approximately 80% of patients with AIT2 have spontaneous restoration of euthyroidism; the remainder develop prolonged or permanent hypothyroidism.[41,69]

KEY POINTS: Amiodarone

- Amiodarone causes expected changes in thyroid tests in euthyroid persons (Figure 2).
- Amiodarone can induce both thyrotoxicosis (amiodarone-induced thyrotoxicosis type 1 [AIT1] and amiodarone-induced thyrotoxicosis type 1 [AIT2]) and hypothyroidism (amiodarone-induced hypothyroidism [AIH]).
- Amiodarone-induced thyrotoxicosis can be difficult to manage due to frequent diagnostic uncertainty.
- If AIT1 is suspected (underlying Graves disease [GD] or toxic multinodular goiter [TMNG]) and no evidence of cardiovascular compromise: treat initially with antithyroid drugs (ATDs) alone.
- If AIT2 is suspected (no underlying GD or TMNG) and no evidence of cardiovascular compromise: treat initially with glucocorticoid therapy alone.
- In patients presenting with cardiovascular compromise: ATDs and glucocorticoids should be started immediately to avoid delay in therapy from possible misclassification.

ICIs. Thyroid dysfunction is noted in approximately 5% to 10% of patients treated with CTLA-4 inhibitors, 10% to 20% of those treated with PD-1 inhibitors, and more than 20% of patients treated with combination therapy.[1] Patients typically present with painless thyroiditis, manifested by transient thyrotoxicosis followed by hypothyroidism; Graves disease rarely occurs. Thyrotoxicosis commonly occurs 4 to 8 weeks after the start of therapy[71] but may occur in the first 1 to 2 weeks with combination therapy[71-73] or by 18 weeks with CTLA-4 inhibitor monotherapy.[72] The thyrotoxic phase is often mild and may be overlooked, resolving in 3 to 10 weeks,[72] but severe cases have occurred.[74] Hypothyroidism follows at 10 to 20 weeks after the initiation of therapy[58] and is frequently irreversible.[58,75] Immune checkpoint blockade therapy is generally not stopped as a result of adverse thyroid effects,[72] which are managed conventionally.[44,76]

Lithium has been associated with painless thyroiditis.[6] Although the nature of this association is unknown, approximately half of affected patients have positive tests for thyroid antibodies.[6] Conservative treatment with beta-blockers is typically the treatment of choice.

Cytokine Therapies IL-2. Thyroid dysfunction, generally resulting from destructive thyroiditis, is a common complication of IL-2 therapy that is seen in 15% to 50% of patients treated.[1] In a series of 89 patients with solid tumors treated with IL-2, 22% developed thyroid dysfunction, presenting with mild thyrotoxicosis at a median of 7 weeks, followed by hypothyroidism at a median of 11 weeks.[63] TSH receptor antibodies are negative, and radioactive iodine uptake and scan shows a low iodine uptake.[77] Ultrasound shows diffuse hypoechogenicity with normal or reduced vascularity, consistent with destructive thyroiditis.[64] Fine-needle aspiration in such patients has shown infiltration by lymphocytes, plasma cells, and histiocytes, indicative of thyroiditis. Thyrotoxicosis is typically mild and requires only symptomatic treatment. Subsequent hypothyroidism is usually self-limited, with only the minority of patients requiring LT4 supplementation. It is controversial whether thyroid dysfunction occurring during cytokine therapy is associated with improved tumor control and increased overall survival. Such predictive ability was reported in a large retrospective study of 1535 patients being treated with IL-2,[78] but other prospective trials did not find this correlation.[79] The most significant risk factors for the development of thyroid dysfunction are positive thyroid antibodies and female gender. Thyroid function testing should be completed before starting therapy, periodically throughout treatment, and a few months following treatment.[79]

IFN-α therapy has been linked to the development of thyroid dysfunction in reports, varying from 1% to 35%.[64] A series including 1233 patients with hepatitis C who were euthyroid before receiving IFN-α identified thyroid events in 16.7%, including approximately 5% with thyrotoxicosis and 12% with hypothyroidism.[80] LT4 supplementation should be used as required, and IFN-α therapy is not generally interrupted. Graves disease may also occur in patients treated with IFN-α,[81] but this is uncommon. In these cases, discontinuation of IFN-α therapy did not lead to remission of thyrotoxicosis,[81] though this approach has been recommended.[64] The presence of TPO antibodies prior to initiation of IFN-α therapy has a positive predictive value of 67% for the development of thyroid dysfunction.[65] Screening for thyroid antibodies before the start of therapy and continued monitoring during and after treatment are recommended. There have been conflicting reports regarding recovery of thyroid function following discontinuation of IFN-α therapy, with small series showing high rates of recovery and euthyroidism after biphasic thyroiditis.[82] Recurrent thyroiditis can also be seen with repeat courses of IFN-α therapy.[65]

TKIs are frequently associated with thyroid dysfunction due to destructive thyroiditis, likely resulting from vascular damage and ischemia. Thyroid dysfunction is highest with TKIs that target vascular endothelial growth factor receptors 1 to 3 and platelet-derived growth factor receptors. A thyrotoxic phase is seen after a median of 6 weeks, followed by hypothyroidism occurring after a median of 22 weeks, but the ranges of onset for both phases are quite broad.[50] The most widely-reported changes occur in patients receiving sunitinib for metastatic renal cell carcinoma or gastrointestinal stromal tumors.[83,84] Sunitinib is given in 6-week cycles, with chemotherapy being delivered over 4 weeks (ON phase) with 2 weeks off in between (OFF phase). An elevation in TSH is frequently seen during the ON phase that normalizes during the OFF phase but can become persistent after many cycles of therapy. In three large clinical trials involving patients with renal cell carcinoma, hypothyroidism occurred in 14% to 25%.[85-87] Most cases were rated as mild to moderate in severity, but rare patients experienced severe dysfunction. In a trial including 42 participants with normal thyroid function before receiving sunitinib, hypothyroidism occurred in 36%, many with TSH values greater than 20 mIU/L.[84] Forty percent of patients were found to have transient thyrotoxicosis prior to becoming hypothyroid, consistent with destructive thyroiditis. Hypothyroidism may be transient, but persistence after completion of therapy has been reported in up to 37.9% (range 10%–37.9%) of patients.[50] The likelihood that hypothyroidism will occur increases with longer duration of therapy. Overall, thyroid dysfunction with TKIs is treated either symptomatically or with thyroid hormone replacement, without having to discontinue therapy. Multiple surveillance algorithms have been suggested for surveillance of thyroid function depending on the type of TKI being used, but generally monthly testing is recommended during the treatment period.[50,88,89]

KEY POINTS: Tyrosine Kinase Inhibitors

- Tyrosine kinase inhibitor (TKI) therapy is frequently associated with thyroid dysfunction due to destructive thyroiditis.
- Thyroid dysfunction is treated either symptomatically or with thyroid hormone replacement without having to discontinue therapy.
- Monthly thyroid testing is recommended during the treatment period.

DRUGS AFFECTING PROTEIN BINDING OF THYROID HORMONE

Changes in Binding Proteins

More than 99% of T4 and T3 is bound to plasma proteins, including TBG, transthyretin, and albumin. TBG is a glycoprotein made in the liver and the primary transport protein for T4 and T3 molecules, accountable for approximately 70% of protein binding, with the remainder bound by transthyretin and albumin. As these binding proteins act as a natural reservoir, changes in their concentration can affect the balance of free and bound thyroid hormone.[90,91] While the normal thyroid gland accommodates changes in binding proteins with commensurate changes in thyroid hormone production, patients whose thyroid glands are compromised by diminished thyroid reserve or dependent on exogenous thyroid hormone therapy are unable to do so, resulting in hypothyroidism. Numerous medications affect plasma protein concentrations, and thus serum thyroid hormone concentrations.

Estrogen-containing preparations. Estrogen therapy is the most clinically significant cause of increased TBG. Estrogen increases the hepatic sialylation of TBG, resulting in decreased renal clearance rates and a subsequent dose-dependent rise in serum concentration.[92] These increases occur within 2 weeks of starting therapy, and a steady state is achieved within 4 to 8 weeks. Typical doses of ethinyl estradiol (20–35 μg per day) or conjugated estrogen (0.625 mg per day), increase TBG concentrations by 30% to 60%, resulting in increased circulating total T4 (TT4) by 20% to 35%. In patients taking LT4 for hypothyroidism, thyroid function should be monitored, and the dose increased or decreased as needed after starting or stopping estrogen therapy, respectively.[93] Because transdermal estrogen preparations evade first-pass hepatic effects seen with oral therapy, they do not raise TBG, and no LT4 dose adjustment is required with these agents.[94]

Selective estrogen receptor modifiers (SERMs) have estrogen-agonist effects in the liver, resulting in modest TBG elevations. Tamoxifen increases TBG by 24% to 30% in patients with normal thyroid function.[95,96] The SERMs raloxifene and droloxifene increase TBG to a lesser extent than tamoxifen, with no effects on free thyroid hormone levels.[97,98]

Opioids. Studies of heroin and methadone users found an increase in TBG by up to 50%, with corresponding increases in TT4 and TT3 levels.[99,100] The etiology is thought to relate to changes in production at the hepatic level, possibly aggravated by increased estrogen or decreased androgen. There are insufficient data to determine whether patients taking LT4 for hypothyroidism require LT4 dose adjustments when starting opioid medications.[100]

Mitotane, in addition to causing central and primary hypothyroidism, has been shown to increase TBG through unknown mechanisms, thereby increasing TT4 and TT3.[31]

5-Fluorouracil (5-FU) is used in the treatment of several solid malignancies and associated with increases in TT4 and TT3 levels. As patients remain euthyroid with normal TSH and FT4 levels, an effect on thyroid hormone binding proteins is likely. No treatment is necessary, and long-term therapy with 5-FU has not been shown to significantly impair thyroid function.[101]

Androgen therapy results in a decreased concentration of TBG in the serum, with subsequent reduction in of TT4 and TT3, which occur within 2 weeks of starting therapy. A steady state is achieved within 4 to 8 weeks for TBG and 8 to 12 weeks for T4. In patients with normal thyroid function, TSH concentrations decrease slightly, but patients remain euthyroid. In a small group of breast cancer patients with known hypothyroidism on thyroid replacement therapy, the addition of the androgen fluoxymesterone led to increases of FT4, reductions in TSH, and symptoms of thyroid hormone excess, necessitating LT4 dose reduction by 25% to 50%.[102] Changes were reversible with the discontinuation of therapy, and TBG levels returned to baseline within 8 to 10 weeks of discontinuing therapy.[102] A study of transgender males on testosterone therapy revealed a 26% decrease in TT4 and a 14% decrease in TBG after 4 months of testosterone therapy.[103] In studies using nonaromatizable androgens, a greater than 50% decrease in TBG has been observed.[91] It is unknown if this decline is related to reduced synthesis or enhanced clearance of TBG. It is important for clinicians to be aware of these effects in patients receiving thyroid hormone therapy. Monitoring of thyroid function should be done within 6 to 8 weeks after starting androgen therapy in patients with known thyroid disease.[91]

Danazol is a synthetic steroid with antigonadotropic, anabolic, and mildly androgenic properties used mainly in the treatment of endometriosis. It has been shown to decrease T4 but not affect TSH. These effects are thought to be due to decreased TBG resulting from the its androgenic properties.[104]

Glucocorticoids also affect the concentration of thyroid hormone–binding proteins, specifically by decreasing TBG and increasing T4-binding prealbumin.[105,106] Increased TBG was seen irrespective of the mode of administration but was more pronounced in the group treated with oral rather than with IV glucocorticoids.[105]

Niacin (nicotinic acid, B3), at doses of 0.5 to 4.0 g daily to treat hyperlipidemia caused a 23% decrease in serum TBG concentration with a subsequent decrease in serum TT4 by 21% and TT3 by 13% in five patients, after 0.5 to 3.5 years of therapy.[107] FT4 and TSH concentrations remained unchanged. These effects are most prominent with doses greater than 2 g daily and are reversible with discontinuation of the medication.

Displacement of Thyroid Hormone From Binding Proteins

Antiepileptics (phenytoin, carbamazepine). Phenytoin and carbamazepine have been shown to displace T4 and T3 from binding proteins when measured using an ultrafiltration assay in undiluted serum.[108] This displacement results in transient increases in free thyroid hormones and a compensatory decrease in TSH, followed by a decrease in TT4, and a new steady state with lower TT4 but normal levels of FT4 (see also the later section: Drugs Causing Abnormal Thyroid Testing in Euthyroid Patients) and TSH.

Nonsteroidal anti-inflammatory drugs, specifically salsalate and acetylsalicylic acid (aspirin), compete with thyroid hormone binding to predominantly TBG, and to a lesser extent transthyretin and albumin.[109] This results in decreased concentrations of TT4 and TT3 with a transient increase in their free fractions. A transient decrease in TSH ensues before a new equilibrium is reached. Salsalate in doses of greater than 1.5g/day can decrease TT4 by nearly 40%.[109]

Furosemide, a potent loop diuretic, has no effect on thyroid function at usual doses. However, a single large dose can compete with T4 for binding sites on TBG, prealbumin, and albumin.[110] These effects were also noted in critically ill patients with acute renal failure who received large cumulative doses of furosemide.[111] A pronounced decrease in TT4 occurs, which persists while serum concentrations of the drug remain elevated. When compared with other drugs that compete for T4 binding sites (fenclofenac, mefenamic acid, and sodium salicylate), furosemide was the most potent.[110] Effects are dose-dependent, starting at an oral dose of 80 mg or higher.[112] The length of time from drug administration, rate of renal clearance, and serum concentrations of albumin also affect the magnitude of the effect.

DRUGS AFFECTING THYROID HORMONE ACTIVATION, METABOLISM, AND EXCRETION

Thyroxine-to-Triiodothyronine Conversion

Amiodarone inhibits both peripheral and central monodeiodination of T4-to-T3 by D1 and D2, respectively. The manifestations of these effects in euthyroid persons prescribed amiodarone will be discussed later in the chapter.

Propranolol therapy at daily doses of 160 mg or higher leads to elevations in T4 levels, suggesting impaired conversion to T3, in clinically euthyroid persons.[113] Even a single propranolol dose of 40 mg in euthyroid participants was found to lead to modest reductions in T3 levels within the normal range.[114] In one study, hyperthyroid patients treated only with propranolol, alprenolol, atenolol, or metoprolol had reductions in T3 levels of 8% to 20%, which may have therapeutic importance in patients being treated for severe thyrotoxicosis or thyroid storm.[115]

Glucocorticoids inhibit outer ring deiodination, leading to a decreased conversion of T4-to-T3 with reciprocal increase in reverse T3 (rT3). They also result in decreased iodine uptake into the thyroid gland, manifested by increased urinary excretion of iodine.[105] Decreases in serum T3 levels reach a nadir at 30 hours. This effect may last for several days, and is more pronounced when using glucocorticoids with long biological half-lives, such as dexamethasone.[116,117]

Thyroid Hormone Metabolism

Antiepileptic drugs including phenobarbital, phenytoin, and carbamazepine are considered enzyme-inducing, as manifested by their ability to stimulate hepatic cytochrome P-450 enzymes and consequently increase the metabolism of thyroid hormone through glucuronidation and sulfation conjugation pathways.[76] In patients taking thyroid hormone for hypothyroidism, a dose increase may be needed when these antiepileptic drugs are added. These effects are reversible when these drugs are discontinued.[118,119] Oxcarbazepine has less enzyme induction activity compared with carbamazepine, but similar effects can be seen at higher doses.[120] Newer antiepileptics, including lamotrigine, levetriacetam, tiagabine, and vigabatrin, have not consistently been shown to alter thyroid function.[120]

Rifampin, an antibiotic used mostly in the treatment of tuberculosis, also stimulates hepatic thyroid hormone metabolism pathways, as occurs with the antiepileptic drugs noted above. T4 concentrations decrease, with a compensatory increase in TSH. This is important clinically for patients receiving LT4 therapy, who may require dose adjustment. In one study, after initiation of rifampin, an increase in LT4 was needed in 50% of patients on TSH suppressive therapy for thyroid cancer and 26% of patients on replacement therapy.[121] In individuals without baseline thyroid dysfunction, rifampin reduces thyroid hormone levels without significantly elevating TSH levels. Rifampin-related changes in thyroid function resolve after cessation of treatment.[122]

TKIs. In addition to damage to the thyroid, an increased LT4 dose requirement has been observed in athyreotic patients being treated with imatinib, motesanib, sorafenib, sunitinib, and vandetanib.[76] Hypothyroid patients receiving imatinib in one study required average LT4 dose increases of 47%.[123] Increased thyroid hormone metabolism through activation of type 3 deiodinase[124] and inhibitory effects on MCT8-mediated thyroid hormone transport,[125] as well as other mechanisms,[50] have been implicated as potential contributors to this finding.

Bexarotene, as noted above, enhances the peripheral metabolism of thyroid hormone. One study showed parallel decreases in serum T4, T3, and rT3 levels and an increase in the T3/rT3 ratio, suggesting increased metabolism through nondeiodinase pathways.[28]

Disposal of Thyroid Hormone

Bile acid sequestrants (cholestyramine, colestipol, colesevelam) enhance thyroid hormone elimination primarily through interference with enterohepatic circulation. Because of these effects, bile acid sequestrants have been useful in the treatment of both endogenous and exogenous causes of thyrotoxicosis.[126,127] Patients on LT4 replacement therapy who are started on a bile acid sequestrant require monitoring of thyroid function tests, with LT4 dose changes possibly required. Thyroid function is not affected by bile acid sequestrants in persons with normal thyroids.

DRUGS AFFECTING THE ABSORPTION OF THYROID HORMONE PREPARATIONS

LT4 absorption is pH-dependent, and 60% to 80% of an oral dose is absorbed over the first 3 hours after ingestion. Anything taken by mouth, including drugs or food, can significantly reduce the bioavailability of LT4. The extent of this effect is dependent on the timing of LT4 dosing relative to intake of food or medications that impair absorption. Multiple over-the-counter, as well as prescribed, medications impair the absorption of LT4, as has been recently reviewed.[128] Patients should be advised to separate LT4 from these medications by at least 4 hours to avoid potential interaction.

Iron preparations, when taken simultaneously with LT4, have been shown to cause a variable reduction in T4 absorption. In a study of 14 previously well-controlled patients on LT4 initiated on 300 mg ferrous sulfate and advised to take the two medications simultaneously, TSH concentrations increased from a mean of 2.8 mIU/L to 5.4 mIU/L after 12 weeks, with an increase in hypothyroid symptoms but no effect on FT4 index.[129] Although no specific study has examined the optimal separation time between these medications, at least 4 hours' separation is generally advised.

Calcium salts are frequently used over-the-counter medications. Calcium carbonate is known to decrease absorption of LT4, but this may also occur with other calcium preparations. A study of healthy subjects given 1000 mcg of LT4 alone or in combination with one of the three calcium salts (carbonate, citrate, or acetate) found a 20% to 25% reduction in LT4 absorption with each of these preparations.[130]

Aluminum hydroxide is used mostly as an antacid and decreases LT4 absorption through drug–drug complexing. A study of five hypothyroid patients taking stable doses of LT4 before and 2 weeks after starting aluminum hydroxide containing antacids found that mean TSH levels increased from 2.6 mIU/L to 7.2 mIU/L, despite instructing participants to separate the two drugs by at least 2 hours.[131]

Bile acid sequestrants, in addition to interference with enterohepatic recycling of thyroid hormone, physically bind to LT4, resulting in decreased intestinal absorption. Colesevelam 3.75 g, given concurrently with an absorption test dose of 600 mcg LT4, nearly completely obliterated the expected increase in serum concentration of T4 unless LT4 was given 4 hours before giving the colesevelam.[132] Cholestyramine given to normal subjects at intervals from 1 to 5 hours before a tracer dose of radiolabeled LT4 showed significant increases in fecal radioactivity that decreased after a 5-hour separation between the two drugs, suggesting impaired absorption of the radiolabeled drug, although decreased enterohepatic recycling may also have contributed to this finding.[133]

Sucralfate, commonly used in the treatment of duodenal ulcers, has been inconsistently linked to impaired absorption LT4. *In vitro* conditions show strong binding of sucralfate to LT4, which is a proposed mechanism for decreased absorption when these two medications are taken concurrently. One study performing a LT4 absorption test in which 1000 mcg was administered to healthy volunteers found only 23% absorption of LT4 when taken concurrently with 1 g sucralfate, compared with 80% absorption when taken alone.[134] Conversely, a study of 10 patients taking stable doses of LT4 found no significant change in T4 or TSH values after adding in sucralfate 1 g, taken concurrently with LT4 for 6 weeks.[135] Despite the conflicting results, patients taking LT4 and sucralfate should separate the medications by more than 4 hours, and thyroid function tests should be monitored.

Raloxifene has been suggested in case reports to decrease absorption of LT4 when the two are taken concurrently, manifested by a rise in TSH after raloxifene was started. In one such report abnormal results were obtained on a LT4 absorption test, despite the separation of raloxifene from the LT4 by several hours.[136]

Proton pump inhibitors (PPIs) are believed to reduce LT4 absorption by decreasing gastric acidity. Some reports show an increase in TSH after starting PPIs, necessitating an LT4 dose

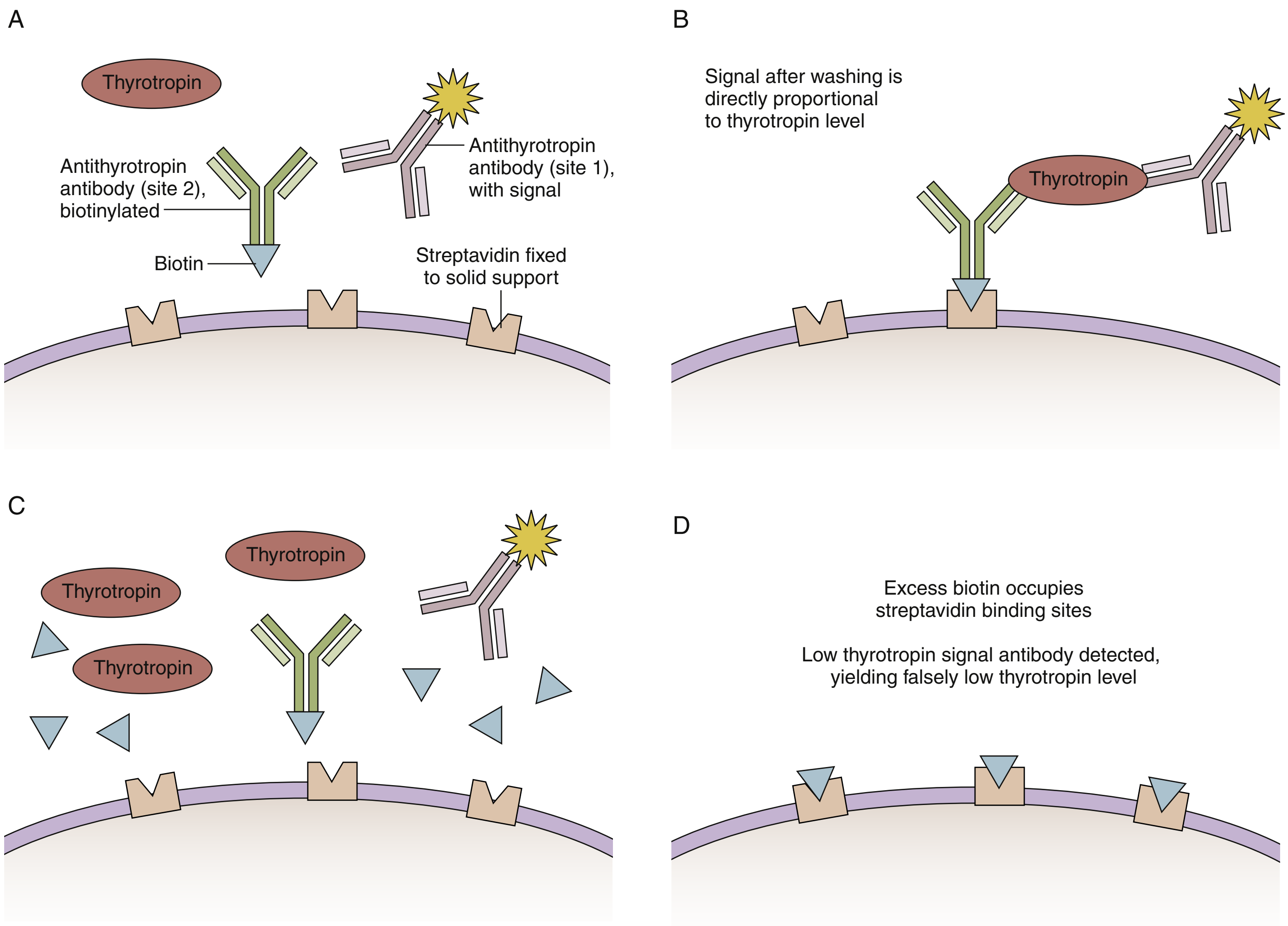

A
Thyrotropin
Antithyrotropin antibody (site 2), biotinylated
Biotin
Antithyrotropin antibody (site 1), with signal
Streptavidin fixed to solid support
B
Signal after washing is directly proportional to thyrotropin level
Thyrotropin
C
Thyrotropin
Thyrotropin
Thyrotropin
D
Excess biotin occupies streptavidin binding sites
Low thyrotropin signal antibody detected, yielding falsely low thyrotropin level

Fig. 84.3 **Biotin interference in two-site (sandwich) thyrotropin assay.** Biotin, when taken in excess, can result in variable interference in assays, such as that for thyrotropin, that utilize biotinylated reagents. Two-site assays include two antibodies: an antithyrotropin antibody containing a signal moiety (site 1) and a biotinylated antithyrotropin antibody (site 2). Streptavidin is fixed to a solid phase such as an enzyme-linked immunosorbent assay and will bind these antibodies forming a complex (**A**). Normally, the complex of thyrotropin bound to the two antithyrotropin antibodies binds to streptavidin. After washing (**B**), the signal is directly proportional to the thyrotropin concentration in the specimen. However, if excess biotin is present, the complex cannot compete effectively for streptavidin binding on the solid phase (**C**), and after washing, a low signal results in a falsely low thyrotropin level (**D**).

increase of approximately 35%,[137] while other studies found no change in TSH after PPI initiation.[138,139] Despite the conflicting results, it is generally recommended that thyroid function tests are monitored after initiation of PPI therapy. Clinicians may also consider liquid LT4 in patients affected by LT4 malabsorption for whom an acidic gastric environment required for tablet dissolution cannot be maintained.[128,137]

DRUGS CAUSING ABNORMAL THYROID TESTING IN EUTHYROID PATIENTS

Biotin, vitamin B7, is prescribed for MS and several rare metabolic disorders but is commonly used as an over-the-counter supplement for hair, skin, and nail enhancement. It is also found in combined B-complex supplements. Biotinylation is utilized in many two-site and competitive immunoassays due to the strong interaction of biotin with streptavidin. In most current TSH assays, two antithyrotropin antibodies (one biotinylated and the other with a signal moiety attached) are used to bind or "sandwich" thyrotropin, forming an antigen–antibody–antigen complex. This complex then binds to the affixed streptavidin and is measured, with signal being directly proportional to the TSH concentration (Fig. 84.3). Excess serum biotin can displace the biotinylated antibody-antigen-antibody complex from streptavidin, resulting in falsely low TSH reading.[1] Conversely, in competitive assays used to measure FT4 and FT3, the signal is inversely proportional to hormone concentrations, and excess biotin in the patient's serum can lead to an overestimation of FT4 and FT3.[1] This is caused by binding of free biotin to the streptavidin sites, thereby blocking the biotinylated antibody and analyte of interest. With nowhere to bind, this complex is washed away, causing a decreased signal, resulting in a falsely increased estimation of FT4 or FT3 (Fig. 84.4).[1,140] Together, falsely high thyroid hormone measurements and falsely low TSH results are easily confused with overt thyrotoxicosis. While intake of 35 to 70 mcg daily does not interfere with these assays, higher doses of 1.5 to 300 mg/day cause significant interference with some, but not all, assays utilizing biotinylation.[141] Patients taking biotin should stop

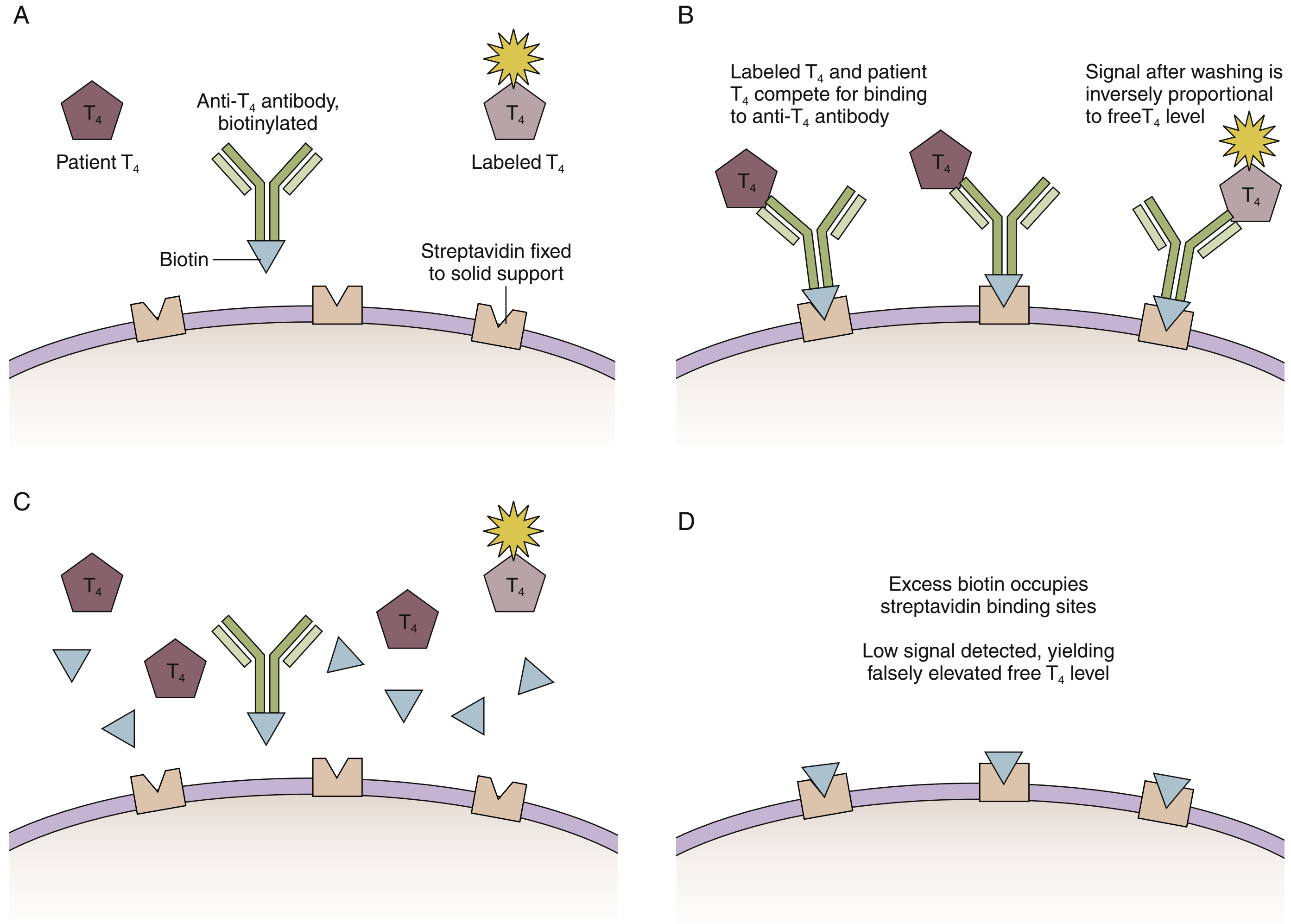

A
Patient T_4
Anti-T_4 antibody, biotinylated
Labeled T_4
T_4
T_4
Biotin
Streptavidin fixed to solid support
B
Labeled T_4 and patient T_4 compete for binding to anti-T_4 antibody
Signal after washing is inversely proportional to freeT_4 level
T_4
T_4
T_4
C
T_4
T_4
T_4
T_4
D
Excess biotin occupies streptavidin binding sites
Low signal detected, yielding falsely elevated free T_4 level

Fig. 84.4 Biotin interference in competitive free thyroxine (T4) immunoassay. Biotin, when taken in excess, can result in variable interference in competitive immunoassays (which includes free T4, free triiodothyronine, and thyrotropin receptor antibodies). Free T4 assay components include labeled T4, a biotinylated anti-T4 antibody, and a streptavidin bound to a solid phase (**A**). Normally, free T4 in serum competes with labeled T4 for binding to the biotinylated anti-T4 antibody. After washing, the amount of signal is inversely proportional to the serum free T4 level (**B**). If excess biotin is present, these molecules occupy the streptavidin binding sites, preventing the anti-T4 antibody from binding (**C**). After washing, a low or absent signal leads to calculation of a falsely elevated free T4 value (**D**).

this supplement for several days before drawing laboratory testing for thyroid function.

KEY POINTS: Biotin

- Biotin has no direct effect on thyroid function but does interfere with biotinylated lab assay resulting in abnormal thyroid test results. The following can be seen based on the type of assay used:
- Two-site ("sandwich") assay ê falsely low TSH reading
- Competitive immunoassay ê falsely high free thyroxine (FT4) and free triiodothyronine (FT3)
- Patients should be advised to stop biotin for several days prior to laboratory testing for thyroid function.

Amiodarone. In addition to causing true hypothyroidism and thyrotoxicosis, amiodarone leads to expected changes in thyroid testing results in euthyroid persons.[1,42,44] Inhibition of peripheral and central T4-to-T3 conversion by amiodarone and its major active metabolite, desethylamiodarone, leads to reductions in circulating and intrapituitary T3 levels, thereby stimulating thyrotropin release through an absence of negative feedback. Increases in thyrotropin prompt thyroidal release of T4, which accumulates further because of inhibited conversion to T3. The net effect of these changes is a serum TSH level that is elevated or at the high end of the normal range, high levels of TT4 and FT4, and a T3 level at the low end of the normal range in a euthyroid patient. A steady state is usually achieved within 3 months with normalization or near-normalization of TSH, but thyroid hormone levels remain elevated.

Phenytoin and **carbamazepine**, as noted earlier, displace thyroid hormone from binding proteins. Patients on chronic therapy with these agents will have lower TT4 and TT3 levels, while maintaining normal TSH values. Free thyroid hormones measured by equilibrium dialysis are also normal. However, in assays that use serial dilution, free thyroid hormone concentrations are underestimated, appearing falsely low, suggesting central hypothyroidism. As a result, clinicians depending on FT4 measurements utilizing serum dilution should base treatment decisions on TSH values or utilize FT4 assays that do not require serum dilution.[108]

Salsalate, by displacing thyroid hormone from binding proteins, reduces TT4 and TT3 levels. Patients on chronic therapy with salsalate will have low total thyroid hormones and a normal TSH, which, as is the case with dilantin and carbamazepine, can be mistaken for central hypothyroidism.[142]

Heparin liberates lipoprotein lipase from the vascular endothelium. Blood samples from a patient receiving heparin have increased lipoprotein lipase activity, which persists *in vitro*. Free thyroid hormone assays with lengthy incubation periods, such as equilibrium dialysis, are most affected, because free fatty acids released by the lipase have longer times to displace T4 and T3 from binding proteins, causing spuriously high values.[143] Similar effects are seen with low molecular weight heparin preparations.[144] Standard competitive free hormone assays and thyrotropin testing are generally unaffected, so the testing can be repeated with the use of these assays. Alternatively, if thyroid dysfunction is a concern, clinicians can draw thyroid before staring heparin therapy or prior to the next heparin dose, or at a minimum 10 hours after injection to avoid these spurious results and unnecessary evaluation.[144,145]

THE EFFECT OF THYROID DYSFUNCTION ON DRUG METABOLISM

Throughout this chapter, we have discussed how different drugs affect thyroid function. Here we will review what is known about the impact of thyroid dysfunction on drug metabolism.

Warfarin, a vitamin K antagonist, is a commonly used anticoagulant. Hyperthyroidism increases the catabolism of vitamin K–dependent clotting factors, prolonging the prothrombin time. Therefore, the warfarin dose commonly needs to be decreased in patients who develop hyperthyroidism while taking this drug, which may seem counterintuitive. The half-life of warfarin is not affected by hyperthyroidism.[146,147]

Digoxin, a cardiac glycoside used for the treatment of atrial fibrillation may require dose adjustments in patients with thyroid dysfunction. Higher doses are often required in thyrotoxic patients, while lower doses are needed in patients with severe hypothyroidism. These effects are thought to be due to enhanced (hyperthyroidism) or delayed (hypothyroidism) clearance of the medication through the renal and biliary systems.[146] Increased cardiac excitability in hyperthyroidism, combined with increased digoxin clearance, may contribute to an increased risk of recurrent atrial fibrillation in a newly thyrotoxic patient previously well-controlled on digoxin therapy.

Acetaminophen demonstrates a more rapid absorption, earlier peaking, and more rapid decrease in patients with untreated thyrotoxicosis, indicating a balance between increased absorption and more rapid metabolism.[148]

Propranolol undergoes a significant hepatic first-pass metabolism, which is accelerated in thyrotoxicosis and reduced in hypothyroidism.[149,150] In addition, increased hepatic blood flow in thyrotoxicosis contributes to a shorter half-life and lower serum steady-state concentrations, while hypothyroidism has the opposite effect. Clinically, this is manifested by higher propranolol dose requirements in thyrotoxic patients compared to those who are euthyroid or hypothyroid.

Glucocorticoids are metabolized by hepatic oxidases, which may be functionally affected by changes in thyroid functional status. Hepatic clearance is delayed in hypothyroidism, resulting in an increased half-life and decreased urinary cortisol metabolite concentrations. Conversely, a decreased glucocorticoid half-life is observed in thyrotoxicosis.[151] These findings are clinically relevant, as increased metabolism of corticosteroids during thyrotoxicosis could precipitate adrenal crisis in a patient dependent on adrenal replacement therapy or in patients with borderline adrenal function at baseline.[152,153]

Statin medication usage has been extensively linked to myalgias, myositis, and rhabdomyolysis, and the risk of these adverse effects is increased with hypothyroidism.[154] The mechanism of this association is unknown but may relate to higher serum statin levels during hypothyroidism.[155] Clinically, ensuring a patient is euthyroid prior to starting or increasing a statin dose is important, because correcting hypothyroidism may both improve the dyslipidemia and decrease the risk of statin-related myopathic effects.

SUMMARY AND FUTURE DIRECTIONS

Drugs may disturb control of the thyroid at the hypothalamic-pituitary level, cause thyroid destruction either directly or through the immune system, initiate or aggravate existing thyroid autoimmunity, and cause hyperthyroidism, hypothyroidism, or both. Drugs affect binding of thyroid hormone to protein carriers and the activation of T4-to-T3, as well as the metabolism, recycling, and elimination of thyroid hormone. Many drugs affect the efficiency of exogenous thyroid hormone therapy, requiring vigilance to prevent both undertreatment while taking them and overtreatment after stopping. Equally concerning, drugs may alter thyroid laboratory results in a manner that artifactually mimics primary hyperthyroidism, central hypothyroidism, and central hyperthyroidism despite the maintenance of a euthyroid state. Being aware of these interactions allows clinicians to monitor for them, intervene when necessary, and avoid needless testing or inappropriate treatment. Each drug, whether newly released or an established mainstay of therapy, should be considered for its potential to interact with the thyroid.

Disclaimer: The views expressed in this manuscript are those of the authors and do not reflect the official policy of the Department of the Army, the Department of Defense, the National Institutes of Health, or the United States Government. One or more authors are military service members or employees of the U.S. Government. This work was prepared as part of our official duties. Title 17 U.S.C. 105 provides the "Copyright protection under this title is not available for any work of the United States Government." Title 17 U.S.C. 101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties. We certify that all individuals who qualify as authors have been listed; each has participated in the conception and design of this work, the analysis of data (when applicable), the writing of the document, and/or the approval of the submission of this version; that the document represents valid work; that if we used information derived from another source, we obtained all necessary approvals to use it and made appropriate acknowledgments in the document; and that each takes public responsibility for it.

REFERENCES

1. Burch HB. Drug effects on the thyroid. *N Engl J Med*. 2019;381:749–761.
2. Samuels MH. Effects of metyrapone administration on thyrotropin secretion in healthy subjects--a clinical research center study. *J Clin Endocrinol Metab*. 2000;85:3049–3052.
3. Bocchetta A, Cocco F, Velluzzi F, et al. Fifteen-year follow-up of thyroid function in lithium patients. *J Endocrinol Invest*. 2007;30:363–366.
4. McKnight RF, Adida M, Budge K, et al. Lithium toxicity profile: a systematic review and meta-analysis. *Lancet*. 2012;379:721–728.
5. Lazarus JH. Lithium and thyroid. *Best Pract Res Clin Endocrinol Metab*. 2009;23:723–733.
6. Miller KK, Daniels GH. Association between lithium use and thyrotoxicosis caused by silent thyroiditis. *Clin Endocrinol*. 2001;55:501–508.
7. Giusti CF, Amorim SR, Guerra RA, et al. Endocrine disturbances related to the use of lithium. *Arq Bras Endocrinol Metabol*. 2012;56:153–158.
8. Kibirige D, Luzinda K, Ssekitoleko R. Spectrum of lithium induced thyroid abnormalities: a current perspective. *Thyroid Res*. 2013;6:3–8.
9. Bauer M, Blumentritt H, Finke R, et al. Using ultrasonography to determine thyroid size and prevalence of goiter in lithium-treated patients with affective disorders. *J Affect Disord*. 2007;104:45–51.
10. Fountas A, Van Uum S, Karavitaki N. Opioid-induced endocrinopathies. *Lancet Diabetes Endocrinol*. 2020;8:68–80.
11. Byun DJ, Wolchok JD, Rosenberg LM, et al. Cancer immunotherapy - immune checkpoint blockade and associated endocrinopathies. *Nat Rev Endocrinol*. 2017;13:195–207.
12. Chang L-S, Barroso-Sousa R, Tolaney SM, et al. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. 2019;40:17–65.
13. Albarel F, Castinetti F, Brue T. Management of endocrine disease: immune check point inhibitors-induced hypophysitis. *Eur J Endocrinol*. 2019;181:R107–R118.
14. Morley JE. Neuroendocrine control of thyrotropin secretion. *Endocr Rev*. 1981;2:396–436.
15. Samuels MH, Henry P, Ridgway EC. Effects of dopamine and somatostatin on pulsatile pituitary glycoprotein secretion. *J Clin Endocrinol Metab*. 1992;74:217–222.
16. Van den Berghe G, de Zegher F, Lauwers P. Dopamine and the sick euthyroid syndrome in critical illness. *Clin Endocrinol*. 1994;41:731–737.
17. Kok P, Roelfsema F, Frolich M, et al. Bromocriptine reduces augmented thyrotropin secretion in obese premenopausal women. *J Clin Endocrinol Metab*. 2009;94:1176–1181.
18. Keogh MA, Wittert GA. Effect of cabergoline on thyroid function in hyperprolactinaemia. *Clin Endocrinol*. 2002;57:699–700.
19. Samuels MH, Kramer P. Effects of metoclopramide on fasting-induced TSH suppression. *Thyroid*. 1996;6:85–89.
20. Kauppila A, Kivinen S, Ylikorkala O. Metoclopramide increases prolactin release and milk secretion in puerperium without stimulating the secretion of thyrotropin and thyroid hormones. *J Clin Endocrinol Metab*. 1981;52:436–439.
21. Haugen BR. Drugs that suppress TSH or cause central hypothyroidism. *Best Pract Res Clin Endocrinol Metab*. 2009;23:793–800.
22. Christensen SE, Weeke J, Orskov H, et al. Long-term efficacy and tolerability of octreotide treatment in acromegaly. *Metabolism*. 1992;41:44–50.
23. Heinen E, Herrmann J, Lippe J, et al. [Effect of dopamine and dobutamine on thyroid hormone concentration]. *Med Welt*. 1983;34:696–699.
24. Lee E, Chen P, Rao H, et al. Effect of acute high dose dobutamine administration on serum thyrotrophin (TSH). *Clin Endocrinol*. 1999;50:487–492.
25. D'Aloia A, Vizzardi E, Bugatti S, et al. Effect of short-term infusive dobutamine therapy on thyroid hormone profile and hemodynamic parameters in patients with acute worsening heart failure and low-triiodothyronine syndrome. *J Invest Med*. 2012;60:907–910.
26. Duvic M, Martin AG, Kim Y, et al. Phase 2 and 3 clinical trial of oral bexarotene (Targretin capsules) for the treatment of refractory or persistent early-stage cutaneous T-cell lymphoma. *Arch Dermatol*. 2001;137:581–593.
27. Sherman SI, Gopal J, Haugen BR, et al. Central hypothyroidism associated with retinoid X receptor-selective ligands. *N Engl J Med*. 1999;340:1075–1079.
28. Smit JW, Stokkel MP, Pereira AM, et al. Bexarotene-induced hypothyroidism: bexarotene stimulates the peripheral metabolism of thyroid hormones. *J Clin Endocrinol Metab*. 2007;92:2496–2499.
29. Graeppi-Dulac J, Vlaeminck-Guillem V, Perier-Muzet M, et al. Endocrine side-effects of anti-cancer drugs: the impact of retinoids on the thyroid axis. *Eur J Endocrinol*. 2014;170:R253–R262.
30. Yıldırım N, Doğan S, Atakan N. Evaluation of thyroid function tests of acne vulgaris patients treated with systemic isotretinoin. *J Dermatolog Treat*. 2017;28:141–144.
31. Russo M, Scollo C, Pellegriti G, et al. Mitotane treatment in patients with adrenocortical cancer causes central hypothyroidism. *Clin Endocrinol*. 2016;84:614–619.
32. Daffara F, De Francia S, Reimondo G, et al. Prospective evaluation of mitotane toxicity in adrenocortical cancer patients treated adjuvantly. *Endocr Relat Cancer*. 2008;15:1043–1053.
33. Samuels MH, Luther M, Henry P, et al. Effects of hydrocortisone on pulsatile pituitary glycoprotein secretion. *J Clin Endocrinol Metab*. 1994;78:211–215.
34. Brabant A, Brabant G, Schuermeyer T, et al. The role of glucocorticoids in the regulation of thyrotropin. *Acta Endocrinologica*. 1989;121:95–100.
35. Iovino M, Steardo L, Monteleone P. Impaired sensitivity of the hypothalamo-pituitary-thyroid axis to the suppressant effect of dexamethasone in elderly subjects. *Psychopharmacology*. 1991;105:481–484.
36. Casteras A, Zafon C, Ciudin A, et al. Are levothyroxine requirements lower in thyroidectomized diabetic patients on metformin treatment? *Thyroid*. 2013;23:1510–1513.
37. Vigersky RA, Filmore-Nassar A, Glass AR. Thyrotropin suppression by metformin. *J Clin Endocrinol Metab*. 2006;91:225–227.
38. Lupoli R, Di Minno A, Tortora A, et al. Effects of treatment with metformin on TSH levels: a meta-analysis of literature studies. *J Clin Endocrinol Metab*. 2014;99:E143–E148.
39. Al-Alusi MA, Du L, Li N, et al. Metformin does not suppress serum thyrotropin by increasing levothyroxine absorption. *Thyroid*. 2015;25:1080–1084.
40. Leung AM, Braverman LE. Iodine-induced thyroid dysfunction. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:414–419.
41. Bogazzi F, Tomisti L, Bartalena L, et al. Amiodarone and the thyroid: a 2012 update. *J Endocrinol Invest*. 2012;35:340–348.
42. Basaria S, Cooper DS. Amiodarone and the thyroid. *Am J Med*. 2005;118:706–714.
43. Martino E, Safran M, Aghini-Lombardi F, et al. Environmental iodine intake and thyroid dysfunction during chronic amiodarone therapy. *Ann Intern Med*. 1984;101:28–34.
44. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26:1343–1421.
45. Bogazzi F, Bartalena L, Brogioni S, et al. Comparison of radioiodine with radioiodine plus lithium in the treatment of Graves' hyperthyroidism. *J Clin Endocrinol Metab*. 1999;84:499–503.
46. Haugen BR, Alexander EK, Bible KC, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1–133.
47. Temple R, Berman M, Carlson HE, et al. The use of lithium in Graves' disease. *Mayo Clin Proc*. 1972;47:872–878.
48. Santen RJ, Wells SA, Cohn N, et al. Compensatory increase in TSH secretion without effect on prolactin secretion in patients treated with aminoglutethimide. *J Clin Endocrinol Metab*. 1977;45:739–746.
49. Figg WD, Thibault A, Sartor AO, et al. Hypothyroidism associated with aminoglutethimide in patients with prostate cancer. *Arch Intern Med*. 1994;154:1023–1025.
50. Jannin A, Penel N, Ladsous M, et al. Tyrosine kinase inhibitors and immune checkpoint inhibitors-induced thyroid disorders. *Crit Rev Oncol Hematol*. 2019;141:23–35.
51. Daniels GH, Vladic A, Brinar V, et al. Alemtuzumab-related thyroid dysfunction in a phase 2 trial of patients with relapsing-remitting multiple sclerosis. *J Clin Endocrinol Metab*. 2014;99:80–89.
52. Pariani N, Willis M, Muller I, et al. Alemtuzumab-induced thyroid dysfunction exhibits distinctive clinical and immunological features. *J Clin Endocrinol Metab*. 2018;103:3010–3018.

53. Ferrajoli A, O'Brien SM, Cortes JE, et al. Phase II study of alemtuzumab in chronic lymphoproliferative disorders. *Cancer*. 2003;98:773–778.
54. Lorenzi AR, Clarke AM, Wooldridge T, et al. Morbidity and mortality in rheumatoid arthritis patients with prolonged therapy-induced lymphopenia: twelve-year outcomes. *Arthritis Rheum*. 2008;58:370–375.
55. Corsello SM, Barnabei A, Marchetti P, et al. Endocrine side effects induced by immune checkpoint inhibitors. *J Clin Endocrinol Metab*. 2013;98:1361–1375.
56. Postow MA, Sidlow R, Hellmann MD. Immune-related adverse events associated with immune checkpoint blockade. *N Engl J Med*. 2018;378:158–168.
57. de Filette J, Jansen Y, Schreuer M, et al. Incidence of thyroid-related adverse events in melanoma patients treated with pembrolizumab. *J Clin Endocrinol Metab*. 2016;101:4431–4439.
58. Iyer PC, Cabanillas ME, Waguespack SG, et al. Immune-related thyroiditis with immune checkpoint inhibitors. *Thyroid*. 2018;28:1243–1251.
59. Kimbara S, Fujiwara Y, Iwama S, et al. Association of antithyroglobulin antibodies with the development of thyroid dysfunction induced by nivolumab. *Cancer Science*. 2018;109:3583–3590.
60. Yamauchi I, Sakane Y, Fukuda Y, et al. Clinical features of nivolumab-induced thyroiditis: a case series study. *Thyroid*. 2017;27:894–901.
61. Osorio JC, Ni A, Chaft JE, et al. Antibody-mediated thyroid dysfunction during T-cell checkpoint blockade in patients with non-small-cell lung cancer. *Ann Oncol*. 2017;28:583–589.
62. Jiang T, Zhou C, Ren S. Role of IL-2 in cancer immunotherapy. *OncoImmunology*. 2016;5:e1163462.
63. Kruit WH, Bolhuis RL, Goey SH, et al. Interleukin-2-induced thyroid dysfunction is correlated with treatment duration but not with tumor response. *J Clin Oncol*. 1993;11:921–924.
64. Carella C, Mazziotti G, Amato G, et al. Clinical review 169: interferon-alpha-related thyroid disease: pathophysiological, epidemiological, and clinical aspects. *J Clin Endocrinol Metab*. 2004;89:3656–3661.
65. Mandac JC, Chaudhry S, Sherman KE, et al. The clinical and physiological spectrum of interferon-alpha induced thyroiditis: toward a new classification. *Hepatology*. 2006;43:661–672.
66. Carella C, Mazziotti G, Morisco F, et al. Long-term outcome of interferon-alpha-induced thyroid autoimmunity and prognostic influence of thyroid autoantibody pattern at the end of treatment. *J Clin Endocrinol Metab*. 2001;86:1925–1929.
67. Censi S, Bodanza V, Manso J, et al. Amiodarone-induced thyrotoxicosis: differential diagnosis using 99mTc-SestaMIBI and target-to-background ratio (TBR). *Clin Nucl Med*. 2018;43:655–662.
68. Patel NR, Tamara LA, Lee H. 99mTc sestamibi thyroid scan in amiodarone-induced thyrotoxicosis type I. *Clin Nucl Med*. 2016;41:566–567.
69. Bogazzi F, Bartalena L, Tomisti L, et al. Glucocorticoid response in amiodarone-induced thyrotoxicosis resulting from destructive thyroiditis is predicted by thyroid volume and serum free thyroid hormone concentrations. *J Clin Endocrinol Metab*. 2007;92:556–562.
70. Kotwal A, Clark J, Lyden M, et al. Thyroidectomy for amiodarone-induced thyrotoxicosis: mayo clinic experience. *Journal of the Endocrine Society*. 2018;2:1226–1235.
71. Lee H, Hodi FS, Giobbie-Hurder A, et al. Characterization of thyroid disorders in patients receiving immune checkpoint inhibition therapy. *Cancer immunology research*. 2017;5:1133–1140.
72. Scott ES, Long GV, Guminski A, et al. The spectrum, incidence, kinetics and management of endocrinopathies with immune checkpoint inhibitors for metastatic melanoma. *Eur J Endocrinol*. 2018;178:175–182.
73. Burch HB. Drug effects on the thyroid. Reply. *N Engl J Med*. 2019;381:1980–1981.
74. McMillen B, Dhillon MS, Yong-Yow S. A rare case of thyroid storm. *BMJ Case Reports*. 2016. https://doi.org/10.1136/bcr-2016-214603.
75. Chang LS, Barroso-Sousa R, Tolaney SM, et al. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. 2019;40:17–65.
76. Jonklaas J, Bianco AC, Bauer AJ, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. *Thyroid*. 2014;24:1670–1751.
77. Vialettes B, Guillerand MA, Viens P, et al. Incidence rate and risk factors for thyroid dysfunction during recombinant interleukin-2 therapy in advanced malignancies. *Acta Endocrinologica*. 1993;129:31–38.
78. Curti B, Daniels GA, McDermott DF, et al. Improved survival and tumor control with Interleukin-2 is associated with the development of immune-related adverse events: data from the PROCLAIM(SM) registry. *J Immunother Cancer*. 2017;5:102.
79. Krouse RS, Royal RE, Heywood G, et al. Thyroid dysfunction in 281 patients with metastatic melanoma or renal carcinoma treated with interleukin-2 alone. *J Immunother Emphas Tumor Immunol*. 1995;18:272–278.
80. Mammen JS, Ghazarian SR, Pulkstenis E, et al. Phenotypes of interferon-alpha-induced thyroid dysfunction among patients treated for hepatitis C are associated with pretreatment serum TSH and female sex. *J Clin Endocrinol Metab*. 2012;97:3270–3276.
81. Wong V, Fu AX, George J, et al. Thyrotoxicosis induced by alpha-interferon therapy in chronic viral hepatitis. *Clin Endocrinol (Oxf)*. 2002;56:793–798.
82. Tran HA, Jones TL, Ianna EA, et al. The natural history of interferon-α induced thyroiditis in chronic hepatitis c patients: a long term study. *Thyroid Res*. 2011;4:2–7.
83. Illouz F, Braun D, Briet C, et al. Endocrine side-effects of anti-cancer drugs: thyroid effects of tyrosine kinase inhibitors. *Eur J Endocrinol*. 2014;171:R91–R99.
84. Desai J, Yassa L, Marqusee E, et al. Hypothyroidism after sunitinib treatment for patients with gastrointestinal stromal tumors. *Ann Intern Med*. 2006;145:660–664.
85. Motzer RJ, Hutson TE, Tomczak P, et al. Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2009;27:3584–3590.
86. Motzer RJ, McCann L, Deen K. Pazopanib versus sunitinib in renal cancer. *N Engl J Med*. 2013;369:1970.
87. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med*. 2018;378:1277–1290.
88. Bianchi L, Rossi L, Tomao F, et al. Thyroid dysfunction and tyrosine kinase inhibitors in renal cell carcinoma. *Endocr Relat Cancer*. 2013;20:R233–R245.
89. Illouz F, Braun D, Briet C, et al. Endocrine side-effects of anti-cancer drugs: thyroid effects of tyrosine kinase inhibitors. *Eur J Endocrinol*. 2014;171:R91–R99.
90. Torre F, Calogero AE, Condorelli RA, et al. Effects of oral contraceptives on thyroid function and vice versa. *J Endocrinol Invest*. 2020;43:1181–1188.
91. Tahboub R, Arafah BM. Sex steroids and the thyroid. *Best Pract Res Clin Endocrinol Metab*. 2009;23:769–780.
92. Ain KB, Mori Y, Refetoff S. Reduced clearance rate of thyroxine-binding globulin (TBG) with increased sialylation: a mechanism for estrogen-induced elevation of serum TBG concentration. *J Clin Endocrinol Metab*. 1987;65:689–696.
93. Arafah BM. Increased need for thyroxine in women with hypothyroidism during estrogen therapy. *N Engl J Med*. 2001;344:1743–1749.
94. Steingold KA, Matt DW, DeZiegler D, et al. Comparison of transdermal to oral estradiol administration on hormonal and hepatic parameters in women with premature ovarian failure. *J Clin Endocrinol Metab*. 1991;73:275–280.
95. Anker GB, Lønning PE, Aakvaag A, et al. Thyroid function in postmenopausal breast cancer patients treated with tamoxifen. *Scand J Clin Lab Invest*. 1998;58:103–107.
96. Mamby CC, Love RR, Lee KE. Thyroid function test changes with adjuvant tamoxifen therapy in postmenopausal women with breast cancer. *J Clin Oncol*. 1995;13:854–857.
97. Duntas LH, Mantzou E, Koutras DA. Lack of substantial effects of raloxifene on thyroxine-binding globulin in postmenopausal women: dependency on thyroid status. *Thyroid*. 2001;11:779–782.
98. Marqusee E, Braverman LE, Lawrence JE, et al. The effect of droloxifene and estrogen on thyroid function in postmenopausal women. *J Clin Endocrinol Metab*. 2000;85:4407–4410.
99. Chan V, Wang C, Yeung RT. Effects of heroin addiction on thyrotrophin, thyroid hormones and porlactin secretion in men. *Clin Endocrinol (Oxf)*. 1979;10:557–565.
100. Azizi F, Vagenakis AG, Portnay GI, et al. Thyroxine transport and metabolism in methadone and heroin addicts. *Ann Intern Med*. 1974;80:194–199.
101. Beex L, Ross A, Smals A, Kloppenborg P. 5-fluorouracil-induced increase of total serum thyroxine and triiodothyronine. *Cancer Treatment Reports*. 1977;61:1291–1295.

102. Arafah BM. Decreased levothyroxine requirement in women with hypothyroidism during androgen therapy for breast cancer. *Ann Intern Med*. 1994;121:247–251.
103. Bisschop PH, Toorians AW, Endert E, et al. The effects of sex-steroid administration on the pituitary-thyroid axis in transsexuals. *Eur J Endocrinol*. 2006;155:11–16.
104. Thorell JI, Dymling JF, Rannevik G. Effect of danazol on thyroid function in women. *Postgrad Med*. 1979;55(suppl 5):33–36.
105. Gamstedt A, Järnerot G, Kågedal B, et al. Corticosteroids and thyroid function. Different effects on plasma volume, thyroid hormones and thyroid hormone-binding proteins after oral and intravenous administration. *Acta Med Scand*. 1979;205:379–383.
106. Oppenheimer JH, Werner SC. Effect of prednisone on thyroxine-binding proteins. *J Clin Endocrinol Metab*. 1966;26:715–721.
107. Shakir KM, Kroll S, Aprill BS, et al. Nicotinic acid decreases serum thyroid hormone levels while maintaining a euthyroid state. *Mayo Clin Proc*. 1995;70:556–558.
108. Surks MI, DeFesi CR. Normal serum free thyroid hormone concentrations in patients treated with phenytoin or carbamazepine. A paradox resolved. *JAMA*. 1996;275:1495–1498.
109. Wang R, Nelson JC, Wilcox RB. Salsalate and salicylate binding to and their displacement of thyroxine from thyroxine-binding globulin, transthyrin, and albumin. *Thyroid*. 1999;9:359–364.
110. Stockigt JR, Lim CF, Barlow JW, et al. Interaction of furosemide with serum thyroxine-binding sites: in vivo and in vitro studies and comparison with other inhibitors. *J Clin Endocrinol Metab*. 1985;60:1025–1031.
111. Stockigt JR, Lim CF, Barlow JW, et al. High concentrations of furosemide inhibit serum binding of thyroxine. *J Clin Endocrinol Metab*. 1984;59:62–66.
112. Newnham HH, Hamblin PS, Long F, et al. Effect of oral frusemide on diagnostic indices of thyroid function. *Clin Endocrinol*. 1987;26:423–431.
113. Cooper DS, Daniels GH, Ladenson PW, et al. Hyperthyroxinemia in patients treated with high-dose propranolol. *Am J Med*. 1982;73:867–871.
114. Lotti G, Delitala G, Devilla L, et al. Reduction of plasma triiodothyronine (T3) induced by propranolol. *Clin Endocrinol (Oxf)*. 1977;6:405–410.
115. Burch HB, Wartofsky L. Life-threatening thyrotoxicosis. Thyroid storm. *Endocrinol Metab Clin N Am*. 1993;22:263–277.
116. Chopra IJ, Williams DE, Orgiazzi J, et al. Opposite effects of dexamethasone on serum concentrations of 3,3',5'-triiodothyronine (reverse T3) and 3,3'5-triiodothyronine (T3). *J Clin Endocrinol Metab*. 1975;41:911–920.
117. Williams DE, Chopra IJ, Orgiazzi J, et al. Acute effects of corticosteroids on thyroid activity in Graves' disease. *J Clin Endocrinol Metab*. 1975;41:354–361.
118. Verrotti A, Scardapane A, Manco R, et al. Antiepileptic drugs and thyroid function. *J Clin Endocrinol Metab*. 2008;21:401–408.
119. Yılmaz U, Yılmaz TS, Akıncı G, et al. The effect of antiepileptic drugs on thyroid function in children. *Seizure*. 2014;23:29–35.
120. Hamed SA. The effect of antiepileptic drugs on thyroid hormonal function: causes and implications. *Expet Rev Clin Pharmacol*. 2015;8:741–750.
121. Kim HI, Kim TH, Kim H, et al. Effect of rifampin on thyroid function test in patients on levothyroxine medication. *PLoS One*. 2017;12:e0169775.
122. Weiss RE, Refetoff S. Thyroid function testing. In: Jameson JL, De Groot LJ, de Kretser DM, et al, eds. *Endocrinology: Adult and Pediatric.* 7th ed. Philadelphia, PA: Elsevier Saunders; 2016:chap 78.
123. de Groot JW, Zonnenberg BA, Plukker JT, et al. Imatinib induces hypothyroidism in patients receiving levothyroxine. *Clin Pharmacol Ther*. 2005;78:433–438.
124. Abdulrahman RM, Verloop H, Hoftijzer H, et al. Sorafenib-induced hypothyroidism is associated with increased type 3 deiodination. *J Clin Endocrinol Metab*. 2010;95:3758–3762.
125. Braun D, Kim TD, le Coutre P, et al. Tyrosine kinase inhibitors noncompetitively inhibit MCT8-mediated iodothyronine transport. *J Clin Endocrinol Metab*. 2012;97:E100–E105.
126. Shakir KM, Michaels RD, Hays JH, et al. The use of bile acid sequestrants to lower serum thyroid hormones in iatrogenic hyperthyroidism. *Ann Intern Med*. 1993;118:112–113.
127. Solomon BL, Wartofsky L, Burman KD. Adjunctive cholestyramine therapy for thyrotoxicosis. *Clin Endocrinol (Oxf)*. 1993;38:39–43.
128. Skelin M, Lucijanić T, Amidžić Klarić D, et al. Factors affecting gastrointestinal absorption of levothyroxine: a review. *Clin Therapeut*. 2017;39:378–403.
129. Campbell NR, Hasinoff BB, Stalts H, et al. Ferrous sulfate reduces thyroxine efficacy in patients with hypothyroidism. *Ann Intern Med*. 1992;117:1010–1013.
130. Zamfirescu I, Carlson HE. Absorption of levothyroxine when coadministered with various calcium formulations. *Thyroid*. 2011;21:483–486.
131. Liel Y, Sperber AD, Shany S. Nonspecific intestinal adsorption of levothyroxine by aluminum hydroxide. *Am J Med*. 1994;97:363–365.
132. Brown KS, Armstrong IC, Wang A, et al. Effect of the bile acid sequestrant colesevelam on the pharmacokinetics of pioglitazone, repaglinide, estrogen estradiol, norethindrone, levothyroxine, and glyburide. *J Clin Pharmacol*. 2010;50:554–565.
133. Northcutt RC, Stiel JN, Hollifield JW, et al. The influence of cholestyramine on thyroxine absorption. *J Am Med Assoc*. 1969;208:1857–1861.
134. Sherman SI, Tielens ET, Ladenson PW. Sucralfate causes malabsorption of L-thyroxine. *Am J Med*. 1994;96:531–535.
135. Khan F, Jeanniton E, Renedo M. Does sucralfate impede levothyroxine therapy? *Ann Intern Med*. 1993;118:317.
136. Siraj ES, Gupta MK, Reddy SSK. Raloxifene causing malabsorption of levothyroxine. *Arch Intern Med*. 2003;163:1367–1370.
137. Vita R, Saraceno G, Trimarchi F, et al. Switching levothyroxine from the tablet to the oral solution formulation corrects the impaired absorption of levothyroxine induced by proton-pump inhibitors. *J Clin Endocrinol Metab*. 2014;99:4481–4486.
138. Ananthakrishnan S, Braverman LE, Levin RM, et al. The effect of famotidine, esomeprazole, and ezetimibe on levothyroxine absorption. *Thyroid*. 2008;18:493–498.
139. Dietrich JW, Gieselbrecht K, Holl RW, et al. Absorption kinetics of levothyroxine is not altered by proton-pump inhibitor therapy. *Horm Metab Res*. 2006;38:57–59.
140. Favresse J, Burlacu MC, Maiter D, et al. Interferences with thyroid function immunoassays: clinical implications and detection algorithm. *Endocr Rev*. 2018;39:830–850.
141. Li D, Radulescu A, Shrestha RT, et al. Association of biotin ingestion with performance of hormone and nonhormone assays in healthy adults. *J Am Med Assoc*. 2017;318:1150–1160.
142. Latham BB, Horst E, Nankin H, et al. Case report: abnormal thyroid function tests in a patient and two normal volunteers treated with salsalate. *Am J Med Sci*. 1993;305:111–113.
143. Mendel CM, Frost PH, Kunitake ST, et al. Mechanism of the heparin-induced increase in the concentration of free thyroxine in plasma. *J Clin Endocrinol Metab*. 1987;65:1259–1264.
144. Stevenson HP, Archbold GP, Johnston P, et al. Misleading serum free thyroxine results during low molecular weight heparin treatment. *Clin Chem*. 1998;44:1002–1007.
145. Laji K, Rhidha B, John R, et al. Abnormal serum free thyroid hormone levels due to heparin administration. *QJM*. 2001;94:471–473.
146. Eichelbaum M. Drug metabolism in thyroid disease. *Clin Pharmacokinet*. 1976;1:339–350.
147. Kurnik D, Loebstein R, Farfel Z, et al. Complex drug-drug-disease interactions between amiodarone, warfarin, and the thyroid gland. *Medicine (Baltim)*. 2004;83:107–113.
148. Forfar JC, Pottage A, Toft AD, et al. Paracetamol pharmacokinetics in thyroid disease. *Eur J Clin Pharmacol*. 1980;18:269–273.
149. Aro A, Anttila M, Korhonen T, et al. Pharmacokinetics of propranolol and sotalol in hyperthyroidism. *Eur J Clin Pharmacol*. 1982;21:373–377.
150. O'Connor P, Feely J. Clinical pharmacokinetics and endocrine disorders. Therapeutic implications. *Clin Pharmacokinet*. 1987;13:345–364.
151. Shenfield GM. Influence of thyroid dysfunction on drug pharmacokinetics. *Clin Pharmacokinet*. 1981;6:275–297.
152. Tsatsoulis A, Johnson EO, Kalogera CH, et al. The effect of thyrotoxicosis on adrenocortical reserve. *Eur J Endocrinol*. 2000;142:231–235.
153. Naik D, Jebasingh KF, Thomas N. Delayed diagnosis of Graves' thyrotoxicoisis presenting as recurrent adrenal crisis in primary adrenal insufficiency. *J Clin Diagn Res*. 2016;10:20–22.
154. Robison CD, Bair TL, Horne BD, et al. Hypothyroidism as a risk factor for statin intolerance. *J Clin Lipidol*. 2014;8:401–407.
155. Lando HM, Burman KD. Two cases of statin-induced myopathy caused by induced hypothyroidism. *Endocr Pract*. 2008;14:726–731.

85

Adrenal Development and Homeostasis

Kaitlin J. Basham and Pierre Val

OUTLINE

Introduction and Historical Background, 1408
Adrenal Structure and Function, 1408
  Anatomy, 1408
  Adrenal Cortex, 1409
Embryonic Adrenal Development, 1411
  Overview, 1411
  Adrenal Fate Determination, 1411
  Fetal Adrenal Development, 1414
Postnatal Adrenal Development, 1414
  Fetal Zone Involution, 1414
  Establishment of Adrenal Cortex Zonation, 1418
Maintenance of Adult Adrenal Cortex Function by Cortical Cell Renewal, 1420
  Homeostatic Renewal, 1420
  Role of Renin-Angiotensin-Aldosterone System and Adrenocorticotropic Hormone/PKA Signaling in Regulation of Renewal, 1423
  Sexual Dimorphism in Renewal, 1423
  Adrenal Cortex Regeneration, 1423
  Tissue Stress Progenitors and Adrenal Aging, 1424
Conclusion and Future Directions, 1424

INTRODUCTION AND HISTORICAL BACKGROUND

The adrenal gland is an essential endocrine organ that functions to produce and secrete steroid and catecholamine hormones. Bartholomeus Eustachius (1520–1574) published the first detailed illustrations of the adrenal as a discrete organ in 1563 and proposed that the gland likely had an important, yet unknown, function.[1] This notion was highly controversial, with many prominent anatomists and physicians at the time arguing that Eustachius had simply observed part of the kidney.[2] Despite its ambiguity, the adrenal gland remained the subject of in-depth study. Cuvier (1769–1832) was the first to recognize the heterogeneous nature of the gland, with an outer portion that is morphologically distinct from the inner. These separate regions were respectively termed the "*cortex*" and "*medulla*" by N. Nagel in 1836.[3] With advancements in compound microscopy techniques, R.A. von Kölliker (1817–1905) provided a detailed microscopic analysis in 1852 and concluded, "the cortical and medullary substances are physiologically distinct and have different functions."[4,5]

Throughout these observational advances, there was widespread conjecture regarding the actual function of the adrenal gland. It was not until 1855, with the work of Thomas Addison, that the adrenal was associated with human disease. Addison was investigating an unusual form of anemia and observed a similar set of symptoms in a series of 11 patients, each of whom had gross destruction of both adrenal glands upon autopsy. In his work "*On the Constitutional and Local Effects of Disease of the Suprarenal Capsule,*" Addison reported the first case series of adrenal insufficiency.[6] This disease, later recognized as Addison's disease, most often results from autoimmune destruction of the glands.[7] In 1856, Charles Brown-Sequard confirmed the vital importance of the adrenal gland by demonstrating that guinea pigs were unable to survive bilateral adrenalectomy.[8] In the twentieth century, the actual substances produced by the adrenal glands were identified. John Jacob Abel and Jōkichi Takamine independently isolated epinephrine, or adrenaline, as the first principle hormone from the adrenal medulla.[9] Edward Calvin Kendall, Tadeus Reichstein, and Philip Showalter Hench later collectively elucidated the structure and biological effects of steroid hormones produced by the adrenal cortex, for which they shared the 1950 Nobel Prize in Physiology or Medicine.[10] Since these seminal advances in the field, the adrenal gland has been shown to play an important role in a wide range of physiological processes throughout the body. Further, adrenal gland dysfunction leads to human diseases with severe morbidity. This chapter will focus on the development and homeostatic maintenance of the gland, which is fundamental to understanding and managing adrenal diseases.

KEY POINTS

- The vital importance of the adrenal gland was not recognized until 1855, when Thomas Addison described the first case series of adrenal insufficiency.
- The adrenal gland is an essential endocrine organ that functions to produce steroid and catecholamine hormones.

ADRENAL STRUCTURE AND FUNCTION

Anatomy

The adrenal glands, which are also known as the suprarenal glands, are located above the upper pole of each kidney. Although both glands function similarly to produce and secrete hormones, the pair displays an asymmetry with respect to size and shape. The right adrenal gland is pyramid-shaped and smaller compared with the left, which is crescent-shaped. Each adult human adrenal gland measures approximately 3 cm in width and 5 cm in length and weighs 4 to 5 grams.[11] Notably, the adrenal is comprised of two distinct endocrine organs, the outer cortex and centrally located medulla, which each have a unique embryonic origin and function. An outermost mesenchymal capsule encases the entire gland (Fig. 85.1).

Adrenal Cortex

The adrenal cortex is derived from intermediate mesoderm and functions to produce steroid hormones. Specifically, the adrenal cortex is responsible for the production and secretion of mineralocorticoids, glucocorticoids, and androgens. Each of these steroids is synthesized from the common precursor cholesterol but is produced in response to a different physiological demand. To enable this type of rapid and precise response to varying organismal needs, the cortex is organized into three concentric regions or "zones" that are functionally specialized. Julius Arnold first described the three subdivisions within the adrenal cortex in 1866. Based entirely on histological observations, Arnold named these the glomerulosa, fasciculata, and reticularis.[12] The nomenclature established by Arnold reflected the unique organization of cells within each zone, but it later became clear that function follows form, with each zone responsible for the production and secretion of specific hormones (Fig. 85.2). This functional specificity within each region is conferred by the expression of a different set of cytochrome P450 steroidogenic enzymes.[13]

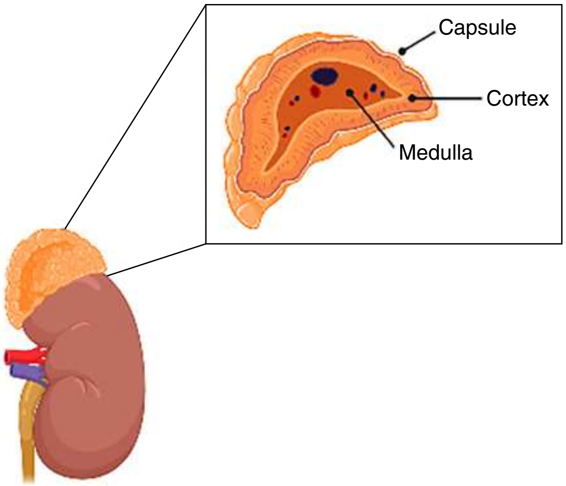


Fig. 85.1 **Adrenal gland structure.** The adrenal gland is located above the upper pole of each kidney. The centrally located medulla (catecholamine-producing) is surrounded by the adrenal cortex (steroid hormone-producing), and an outermost mesenchymal capsule encases the entire gland. Figure was created using BioRender (https://biorender.com/).

Zona Glomerulosa. The zona glomerulosa (zG) is located directly beneath the capsule and makes up approximately 15% of the cortex. The zG is comprised of cells that are organized into round clusters or functional units called glomeruli (Latin *glomus*, ball), which act to produce mineralocorticoids. Aldosterone is the primary mineralocorticoid and is regulated by angiotensin II (Ang II) as part of the renin-angiotensin-aldosterone system (RAAS). Aldosterone plays a central role in controlling blood pressure by regulating sodium balance and thereby influencing water retention and intravascular volume.[14]

Cells within the glomerulosa are interconnected and encased within a laminin B1-rich basal membrane.[15] During postnatal development, zG cells organize into multicellular rosettes of 10 to 15 cells that are connected at a single central point by adherens junction-mediated membrane constriction.[15] Morphogenesis into these mature rosette structures helps facilitate aldosterone production by coordinating the calcium-dependent signaling response following Ang II stimulation.[16] Notably, rosettes contain both CYP11B2-positive cells as well as CYP11B2-negative cells. CYP11B2 (or aldosterone synthase) catalyzes the conversion of deoxycorticosterone (DOC) into aldosterone, which is the final step in the synthesis of aldosterone from cholesterol. Thus, CYP11B2 marks those cells with the capacity to produce aldosterone,

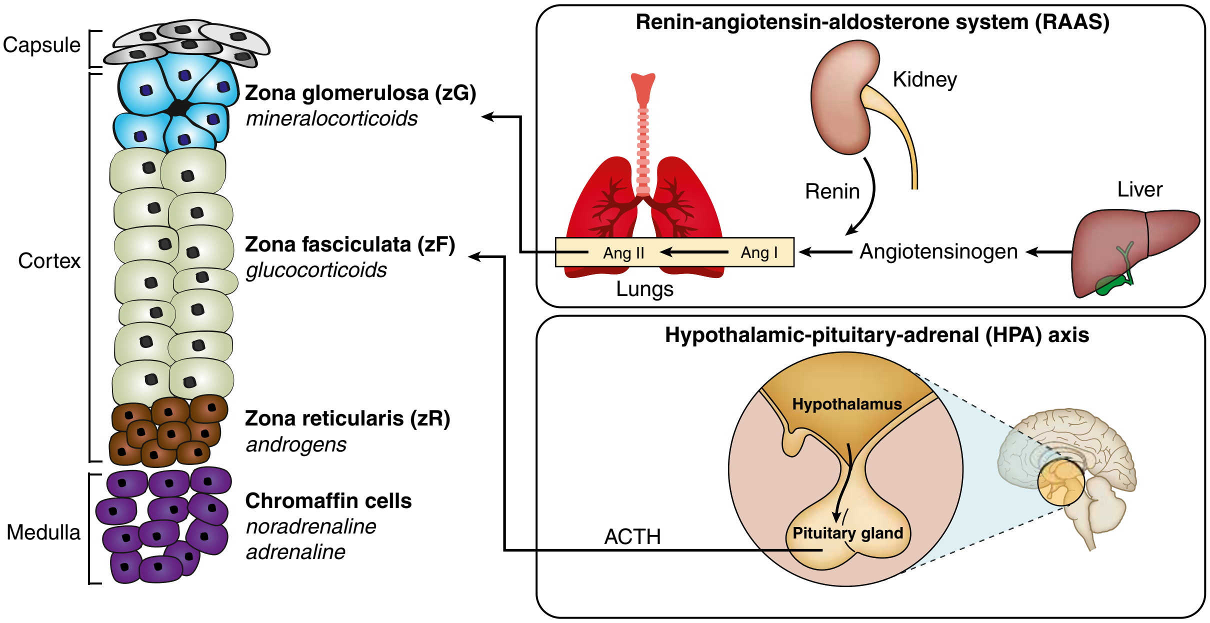


Fig. 85.2 **Adrenal gland function.** The inner medulla contains chromaffin cells that produce catecholamine hormones (noradrenaline and adrenaline). The outer cortex is subdivided into three functionally distinct zones. The zona glomerulosa (zG) is located just beneath the capsule and is composed of cells organized into round glomeruli. The zG functions to produce mineralocorticoids (aldosterone) in response to angiotensin II (Ang II) as part of the renin-angiotensin-aldosterone system (RAAS). The intermediate zona fasciculata (zF) contains cells organized into radial cords or fascicles. The zF functions to produce glucocorticoids (cortisol) in response to ACTH as part of the hypothalamic–pituitary–adrenal (HPA) axis. Finally, the innermost zona reticularis (zR) contains cells spread across a net-like structure and functions to produce androgens (DHEA and DHEA-S). Figure was created using BioRender (https://biorender.com/).

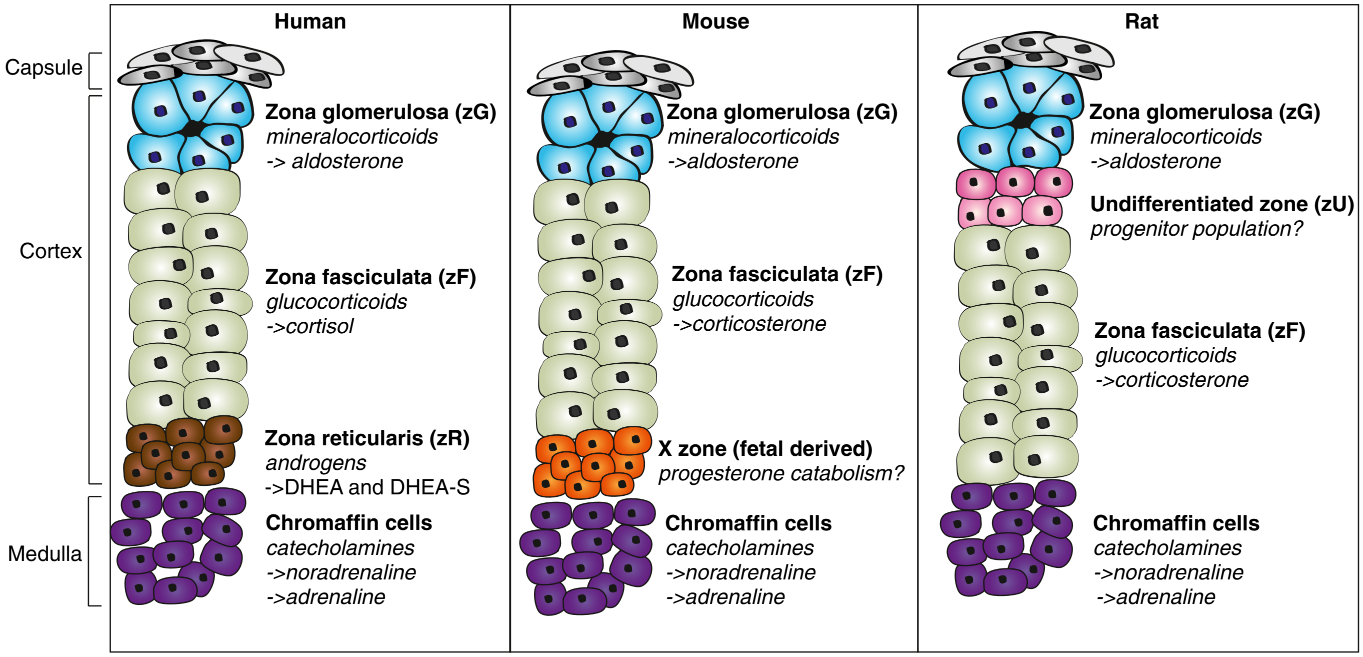

Human
Mouse
Rat
Capsule
Cortex
Medulla
Zona glomerulosa (zG)
mineralocorticoids
-> aldosterone
Zona fasciculata (zF)
glucocorticoids
->cortisol
Zona reticularis (zR)
androgens
->DHEA and DHEA-S
Chromaffin cells
catecholamines
->noradrenaline
->adrenaline
Zona glomerulosa (zG)
mineralocorticoids
->aldosterone
Zona fasciculata (zF)
glucocorticoids
->corticosterone
X zone (fetal derived)
progesterone catabolism?
Chromaffin cells
catecholamines
->noradrenaline
->adrenaline
Zona glomerulosa (zG)
mineralocorticoids
->aldosterone
Undifferentiated zone (zU)
progenitor population?
Zona fasciculata (zF)
glucocorticoids
->corticosterone
Chromaffin cells
catecholamines
->noradrenaline
->adrenaline

Fig. 85.3 **Adrenal gland structure across different species.** The structure of the adrenal gland, and particularly the adrenal cortex, is highly conserved between species. However, there are some key differences in mice and rats as compared with humans. The majority of mouse species *(middle)* do not express CYP17A1 after fetal development. Consequently, these animals synthesize corticosterone rather than cortisol as the main glucocorticoid, and they do not make adrenal androgens (i.e., they lack a functional zona reticularis). The mouse adrenal cortex contains an additional zone called the X zone, which is hypothesized to function in progesterone catabolism. The X zone is derived from the fetal cortex and regresses in males during puberty and in females after pregnancy. The rat adrenal cortex *(right)* contains a unique region between the zona glomerulosa (zG) and zone fasciculata (zF) called the undifferentiated zone (zU). This region is defined by a lack of expression of zG or zF differentiation markers and is proposed to contain progenitor cells that repopulate the cortex.

while CYP11B2-negative cells in the zG constitute a reservoir of cortical progenitor cells.[17]

Zona Fasciculata. Directly adjacent to the zG, the zona fasciculata (zF) is composed of cells organized into radial cords (Latin *fascicle*, bundle) that function to produce glucocorticoids. Cortisol is the main glucocorticoid in humans and is produced in response to adrenocorticotropin hormone (ACTH) as part of the hypothalamic–pituitary–adrenal axis. Cortisol has a wide range of effects throughout the body, including increased blood sugar through stimulation of gluconeogenesis and suppression of the immune system.[18]

Cells within the zF are characteristically polygonal and lipid-rich, with fenestrated capillaries encasing each cord to facilitate rapid hormone circulation. The production of glucocorticoids is driven by pituitary ACTH, which acts through the MC2R receptor in complex with the accessory protein MRAP.[19] The ultimate production of cortisol depends largely on two key enzymes, CYP17A1 and CYP11B1. CYP17A1 hydroxylase activity converts pregnenolone into 17α-OH-pregnenolone, which is then converted into 17α-OH-progesterone by 3βHSD. This is then converted to 11-deoxycortisol by CYP21 and into cortisol by CYP11B1. However, the majority of rodent species do not express *Cyp17a1* in the adrenal cortex after fetal development.[20] As a result, CYP11B1 instead converts DOC into corticosterone, which acts as the main glucocorticoid in these animals.

Zona Reticularis. The innermost zona reticularis (zR) contains cells spread across a mesh or net-like structure (Latin *rete*, net), which function to produce androgens. These include dehydroepiandrosterone (DHEA) and DHEA-sulfate (DHEA-S). DHEA is a weak androgen and estrogen receptor agonist itself but can be converted peripherally into more potent steroids, including testosterone and estradiol, and to a lesser extent, glucocorticoids.[21]

The zR is composed of polyhedral cells and first becomes morphologically distinguishable at 3 years of age in humans. It forms a continuous layer of cells by 6 years, when adrenal androgens synthesis significantly increases with the onset of adrenarche, and further enlarges until 13 years of age.[21,22] The production of DHEA is mediated in large part by CYP17A1. As a result, rodents deficient for adrenal *Cyp17a1* lack a functional zR and do not produce adrenal androgens. While CYP17A1 is expressed in both the zF and zR, lower expression of 3βHSD in the zR and high expression of cytochrome B5 (CYB5) and P450 oxydoreductase (POR) favors CYP17A1 lyase activity, thus rerouting 17α-OH-pregnenolone into DHEA synthesis. zR specific expression of SULT2A1 further allows production of DHEA-S.[21]

Mouse X Zone and Rat Undifferentiated Zone. Mice and rats each contain an additional cortical region not found in the human adrenal gland (Fig. 85.3). The mouse adrenal cortex contains a transient inner zone called the "X zone" that is located at the cortical-medullary boundary.[23] The origin of the X zone was debated for decades, but lineage-tracing studies have now demonstrated that this region is a remnant of fetal cortex.[24] The X zone is initially quite narrow just after birth and becomes most evident by 3 weeks of age before ultimately regressing in males after puberty and in females during pregnancy.[23] The X zone is distinguished based on the expression of specific molecular markers, including 20-α-hydroxysteroid dehydrogenase (20αHSD)[25] and phosphatidylinositol-4-phosphate 3-kinase c2 domain-containing gamma polypeptide (PIK3C2G)[26]. The function of the X zone remains poorly understood, though it has been hypothesized to participate in progesterone catabolism because 20αHSD catalyzes the reduction of progesterone and deoxycorticosterone into their 20α-hydroxylated derivatives.[25] However, the role of this enzymatic activity in the adrenal cortex of rodents remains unclear, as 20αHSD knockout adrenals are not distinguishable from their wild-type counterparts.[25]

The rat adrenal cortex contains a unique region between the zG and zF known as the "undifferentiated zone" (zU).[27] The zU is defined by a lack of zG or zF differentiation markers and contains undifferentiated progenitor cells that may help replenish the adrenal cortex.[28] In mice, stem and progenitor cells are contained within the outer mesenchymal capsule and underlying zG, respectively.[17,29,30]

Adrenal Medulla. The adrenal medulla is derived from ectodermal neural crest cells and functions to produce catecholamines.[31] Unlike steroid hormones produced by the adrenal cortex, catecholamines are derived from the amino acid tyrosine and are water-soluble. Tyrosine is converted through a multistep process into dopamine followed by the predominant catecholamines, noradrenaline, and adrenaline (also called norepinephrine and epinephrine, respectively). Catecholamines are produced in chromaffin cells that are grouped around venous sinusoids, which facilitate their rapid circulation during the "fight-or-flight" response. The release of catecholamines is directly controlled by preganglionic nerve fibers that provide a neural stimulus and connect chromaffin cells to the sympathetic division of the autonomic nervous system. Adrenaline acts as a hormone to enhance cardiac output through increased heart rate and blood pressure, and to increase glucose metabolism in the liver and muscle tissue. Noradrenaline is released in smaller amounts and acts as a neurotransmitter to promote vasoconstriction, thus increasing blood pressure.

Vasculature. To facilitate the rapid delivery of hormones into the bloodstream, the adrenal gland contains a rich vascular network. Blood is supplied from three main arteries: the superior, middle, and inferior adrenal artery, which derive from the inferior phrenic artery, the abdominal aorta, and the renal artery, respectively.[32] These major vessels form a plexus, or branched network of arteries, in the capsule, which then descend centripetally through the cortical layers. In the medulla, arteries form a capillary network that directs blood flow into the medullary vein before ultimately draining into either the inferior vena cava on the right side or the left renal vein. This extensive vascular network ensures that most cells are in close proximity (1–2 cells) to vascular endothelial cells, thus facilitating rapid hormone distribution throughout the body.

Nerve Supply. In contrast to the cortex, where steroid production is controlled in large part by hormonal stimulation, catecholamine production from the medulla is regulated by synaptic stimulation. The medulla is innervated by preganglionic nerve fibers of the sympathetic nervous system that originate in the intermediolateral cell column of the spinal cord.[33] These fibers form the greater splanchnic nerve, from which some fibers directly enter the adrenal gland to synapse at the membrane of chromaffin cells. Other fibers from the greater splanchnic nerve synapse at the celiac ganglion, whose postsynaptic fibers innervate blood vessels in the adrenal gland. Together, this network of nerve fibers enables the medulla to release its neurohormones in response to activation of the sympathetic nervous system.

KEY POINTS

- The adrenal gland is composed of two distinct endocrine organs: the inner medulla derived from ectodermal neural crest cells that produces catecholamine hormones, and the outer cortex derived from intermediate mesoderm that produces steroid hormones.
- The cortex is subdivided into three structurally and functionally distinct zones: the outer zona glomerulosa produces mineralocorticoids in response to angiotensin II as part of the renin-angiotensin-aldosterone system, the intermediate zona fasciculata produces glucocorticoids in response to adrenocorticotropic hormone as part of the hypothalamic-pituitary-adrenal axis, and the innermost zona reticularis produces androgens.
- Much of our understanding about development and homeostasis of the adrenal gland is derived from studies in mice. While the mouse adrenal gland is similar in many respects to the human gland, there are some important differences: (1) the mouse adrenal cortex contains a transient inner zone not found in humans called the "X zone," which is a remnant of fetal cortex, and (2) many strains of mice do not express CYP17A1 after fetal adrenal development. As a result, these animals synthesize corticosterone rather than cortisol as the main glucocorticoid, and they do not produce adrenal androgens.

EMBRYONIC ADRENAL DEVELOPMENT

Overview

The adrenal cortex is derived from the adrenogonadal primordium (AGP), which arises from thickening of the coelomic epithelium. The AGP extends from the dorsal aorta to the dorsal coelomic epithelia of the urogenital ridge.[34] This bipotent cell population serves as a common embryological origin for both the adrenal gland and somatic cells of the gonad. The AGP is identified during development based on the expression of steroidogenic factor 1 (SF-1, encoded by the gene *NR5A1*), which is observed in humans starting at approximately 4 weeks postconception[35,36] and at embryonic day (E) 9 in mice.[37] Cells on the dorsal aorta side progressively separate from the AGP to form the nascent adrenal cortex, while cells on the coelomic epithelial side give rise to the gonadal anlage.[34] AGP separation begins during the fifth week of human gestation or between E10.0 and E10.5 in mice. Following AGP separation, neural crest cells invade the adrenal primordium (~6 weeks human gestation or E13 in mice) to give rise to the adrenal medulla.[38] These cells are initially dispersed but eventually coalesce after birth and differentiate into catecholamine-producing chromaffin cells in the presence of glucocorticoids.[39] Mesenchymal cells encapsulate the gland at approximately 9 weeks of human gestation or E14 to E14.5 in mice,[38,40] and the cortex proceeds with functional zonation and differentiation (Fig. 85.4).

Adrenal Fate Determination

AGP development and specification of the adrenal primordium is dependent upon a number of important factors, including several key nuclear hormone receptors and transcription factors. Because the AGP is a common precursor to both the adrenal and the gonad, disruption of these factors in humans or mice is typically associated with both adrenal and gonadal defects. The nuclear hormone receptors SF-1 and DAX-1 are among the most central and well-studied. Their expression and activity are highly interdependent and must be tightly regulated. In addition, GATA family transcription factors, and particularly the balance between GATA6 and GATA4, helps direct adrenal versus gonadal fate. Following adrenal fate determination, growth factor signaling, including Wnt/β-catenin, sonic hedgehog (SHH), and fibroblast growth factor (FGF), plays a critical role supporting the survival and further progression of the developing gland.

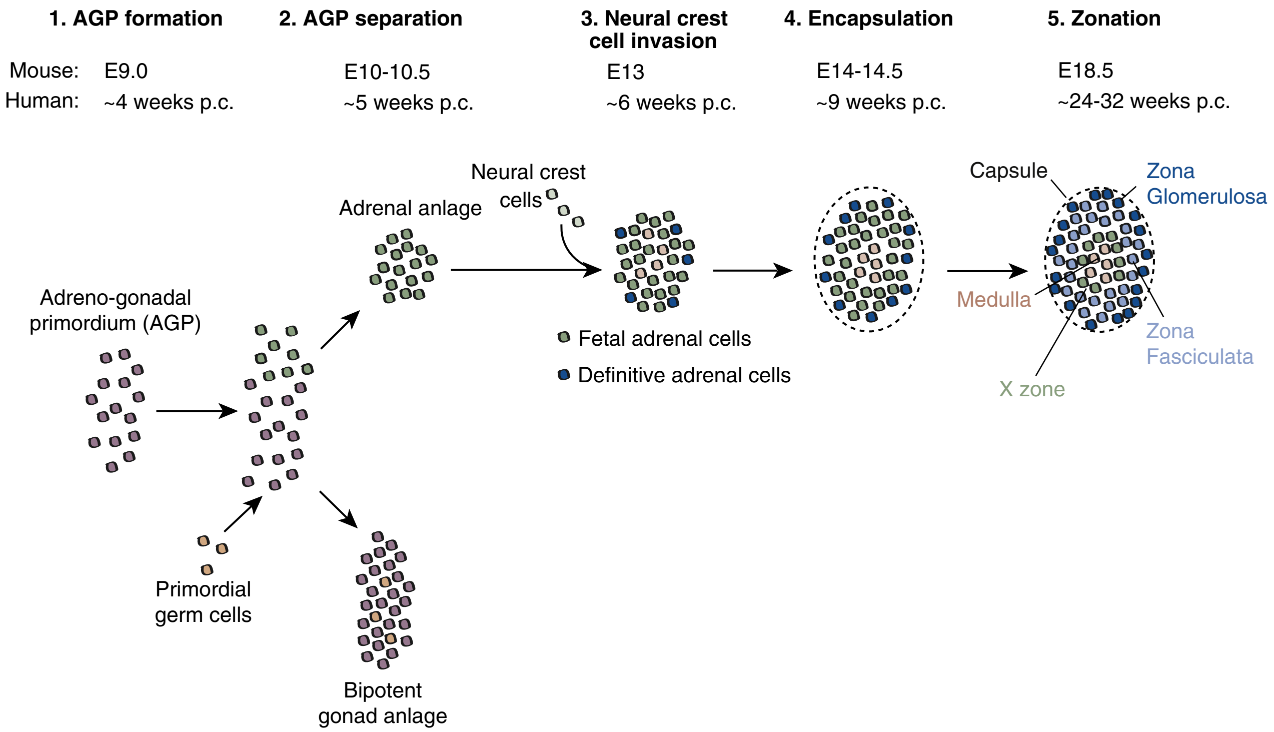


Fig. 85.4 Embryonic adrenal development. The adrenal gland is derived from the adrenogonadal primoridum (AGP). The AGP forms at embryonic day (E) 9 in mouse and ~4 weeks postconception (p.c.) in humans and serves as a common embryological origin for both the adrenal gland and somatic cells of the gonad. Following AGP separation (E10–10.5 in mice, ~5 weeks p.c. in humans), neural crest cells invade the adrenal primordium and ultimately coalescence to form the medulla. The gland is encapsulated by surrounding mesenchyme (E14–14.5 in mice, 9 weeks p.c. in humans) before proceeding with functional zonation and differentiation (E18.5 in mice, ~24–32 weeks p.c. in humans). Fetal adrenal cells give rise to definitive adrenal cells that comprise the adult cortex. In mice, some fetal adrenal cells remain in the cortex and form the X zone, which regresses in males during puberty and in females after pregnancy. (Adapted from Val, P and Swain, A. *Molecular and Cellular Endocrinology.* 2010;323:105–114.)

Steroidogenic Factor 1 (NR5A1). SF-1 (also called adrenal 4 binding protein [Ad4BP]) is a master regulator of steroidogenic cell differentiation. Consistent with its expression early in AGP formation, *Sf-1*-deficient mice fail to develop adrenal glands or gonads and die soon after birth due to adrenal insufficiency.[41,42] Rudimentary structures initially form in *Sf-1* null animals, but adrenal and gonadal tissues are fully extinguished by E13 through apoptotic cell death, suggesting that SF-1 is necessary for the survival of early progenitor cell populations. Importantly, the actual dosage of *Sf-1* gene expression is a critical determinant of adrenocortical growth and differentiation. This is evident in *Sf-1* heterozygous mice, which have significantly smaller adrenal glands compared with controls,[43] as well as impaired compensatory adrenal growth following unilateral adrenalectomy.[44] Consistent with these observations in mouse models, patients with *SF-1* mutations have been found to have a broad spectrum of phenotypes, often including gonadal dysgenesis, and in some cases adrenal insufficiency.[45-48]

Because the dosage of *Sf-1* is critical, tissue-specific enhancers and promoter elements tightly control *Sf-1* activation and expression. Early in AGP formation, a 647-bp fragment in the *Sf-1* promoter is sufficient for its expression.[49] This region contains key Wilm's tumor suppressor (WT1) response elements that are required for WT1-mediated transcriptional activation (see the following section). Following AGP separation, tissue-specific enhancer activity independently maintains *Sf-1* expression in the adrenal and gonad.[50,51] In the adrenal primordium, a regulatory element located in the fourth intron drives expression. This element, termed the fetal adrenal enhancer (FAdE), is highly conserved between mouse and human.[50] However, FAdE shuts off by E14.5 in mouse, and *Sf-1* expression is then maintained in the adrenal cortex by a proposed, but yet to be identified, definitive adrenal enhancer.[52]

Wilm's Tumor Suppressor. Expression of SF-1 in the AGP is directly controlled by WT1.[49] WT1 is a transcriptional regulator expressed early in the urogenital ridge.[29,35,53] Consistent with SF-1 as a direct WT1 target, *Wt1* knockout (KO) mice fail to express SF-1 and display adrenal and gonadal agenesis.[54,55] The ability of WT1 to activate *Sf-1* expression is dependent on a transcription cofactor called Cited2.[56] WT1 and Cited2 act together directly on the *Sf-1* promoter to drive expression above the critical threshold required for adrenal development.[57] However, despite its critical role activating *Sf-1* expression early in AGP formation, WT1 is repressed in the adrenal primordium soon after separation, which is required for further adrenocortical maturation.[29]

While WT1 is not expressed in steroidogenic cells after AGP separation, it is expressed in a high proportion of mesenchymal cells that encapsulate the gland. This population of WT1+ capsular cells was identified experimentally using a tamoxifen-inducible *Wt1:CreERT2* driver combined with the R26R^{mTmG} reporter.[29] This approach permanently labels WT1-expressing cells and their descendants after the administration of tamoxifen. Following tamoxifen at E11.5 and E12.5 (i.e., after AGP separation), a portion of the capsule is green fluorescent protein (GFP)+.[29] Importantly, these capsular WT1+ cells are multipotent and have the capacity to give rise to adrenocortical cells, which can be visualized experimentally using lineage tracing. Tamoxifen injection at E12.5 followed by analysis at E18.5 onwards revealed that GFP-labeled (i.e., WT1-derived) steroidogenic cells (GFP+/SF-1+/WT1-) expand centripetally over time.[29] These capsular WT1+ stem cells contribute at relatively low levels to normal development and homeostasis, but can be activated in response to extreme stress, including gonadectomy[29] (see section on Tissue Stress Progenitors and Adrenal Aging).

DAX-1 (*NR0B1*). One of the most highly studied SF-1-target genes in the developing adrenal is *DAX-1*.[47] This atypical nuclear hormone receptor is expressed in the AGP shortly after the onset of SF-1 expression.[58,59] DAX-1 alters the transcriptional activity of SF-1 in a context-dependent manner, most traditionally through the recruitment of corepressors that function to inhibit steroidogenesis.[60] Mutations in *DAX-1* in humans cause X-linked adrenal hypoplasia congenita, which often presents with adrenocortical failure.[61,62] *Dax-1*-deficient mice initially have a mild phenotype characterized by retention of the X zone.[63] However, loss of *Dax-1* partially rescues *Sf-1* haploinsufficiency,[64] and aged animals display significant adrenocortical dysplasia.[65] These results are consistent with DAX-1 as a key mediator of SF-1-dependent adrenal development and function.

GATA factors. GATA4 and GATA6 are both expressed in the AGP and are essential transcriptional regulators of embryonic development. However, whereas GATA6 expression is maintained in the adrenal cortex throughout development and postpartum, GATA4 expression in steroidogenic cells is extinguished as early as E12.5.[29,66,67] Combined inactivation of both *Gata4* and *Gata6* using *Sf-1:Cre* results in adrenal aplasia.[68] In contrast, mice with isolated inactivation of *Gata6* have adrenal hypoplasia characterized by a thin cytomegalic cortex, overt hypoaldosteronism, and subclinical hypocorticosteronism.[26] A striking observation in these mice is upregulation of GATA4 and gonadal-like markers at the expense of adrenal markers, which is consistent with the idea that GATA6 promotes adrenocortical differentiation, whereas GATA4 drives gonadal differentiation.[69-71] Interestingly, mice lacking the histone methyltransferase EZH2 in steroidogenic cells also show adrenal hypoplasia and glucocorticoid insufficiency.[67] Although GATA6 expression is unaltered in these mice, GATA4 and a number of gonadal differentiation markers (e.g., *Pdgfrα*, *Foxl2*, and *Lhr*) are also markedly upregulated. Altogether, these data demonstrate that the balance between GATA6 and GATA4 expression has a profound impact on adrenal versus gonadal differentiation.

Wnt/β-catenin signaling. Canonical Wnt/β-catenin signaling is one of the main pathways required for early adrenal development. This pathway, which plays an essential role in the development and homeostasis of multiple tissues, relies on extracellular secreted Wnt ligands and the multifunctional protein β-catenin.[72] In the absence of ligands, the cytoplasmic pool of β-catenin is maintained at low levels. This is mediated by phosphorylation of key β-catenin residues by serine/threonine kinases CK1α and GSK3β that are part of the so-called destruction complex, which also comprises the tumor suppressors APC and AXIN. Phosphorylated β-catenin is recognized and ubiquitinylated by the E3 ubiquitin ligase βTrCP, which promotes proteasomal degradation of the protein. This inhibits activation of the pathway by preventing nuclear translocation of β-catenin and stimulation of TCF/LEF-dependent transcriptional activity. Upon Wnt ligand binding to Frizzled (FZD)/LRP5-6 receptor complexes, the destruction complex is sequestered to the plasma membrane, thus preventing phosphorylation and subsequent degradation of β-catenin. Stabilization of β-catenin facilitates its nuclear translocation and interaction with TCF/LEF family members, which triggers expression of canonical Wnt target genes, such as *Axin2*, *Lef1*, or *Apcdd1*.[73-75]

Further levels of fine-tuning of the pathway also exist at the plasma membrane. Specifically, FZD receptors are ubiquitinylated by membrane E3 ubiquitin ligases ZNRF3 and RNF43,[76,77] which are targets of the canonical Wnt pathway.[76] This allows establishment of a negative feedback loop, preventing deregulated pathway activation. This negative regulation mechanism can be counteracted by secreted R-Spondins (RSPOs),[78] which induce membrane clearance of ZNRF3 and RNF43 through both LGR-dependent and LGR-independent mechanisms.[79,80] In addition to the RSPO-ZNRF3/RNF43 signaling module, canonical Wnt/β-catenin activation can also be negatively controlled by other mechanisms, including secreted FZD-related proteins (SFRPs) and (DKKs) glycoproteins.[81]

Activation of canonical Wnt/β-catenin signaling begins early in adrenal development. In mice, immunohistochemistry for β-catenin and TCF/LEF-LacZ reporter activity both demonstrate that signaling is first initiated between E12.5 and E14.5.[82] Pathway activation is highest just beneath the capsule in the underlying zG. Complete inactivation of *Ctnnb1* (the gene encoding β-catenin) using a high efficiency Cre expressed in all SF-1–positive cortical cells (*Sf-1-Cre*^{high})[83] results in severe loss of adrenocortical cells by E14.5 and complete adrenal aplasia by E16.5.[82] These results suggest that canonical Wnt/β-catenin signaling is required for early adrenal development. Interestingly, the converse experiment in which β-catenin is constitutively activated throughout the adrenal cortex early in development results in a similar outcome. Approaches to either directly (*Sf-1-Cre*^{high}*:Catnblox(ex3)* mice) or indirectly (*Sf-1-Cre*^{high}*:Apcfx/fx*) stabilize *Ctnnb1* result in significant adrenal hypoplasia by E16.5.[84,85] These results suggest that the dosage of Wnt/β-catenin activation is critically important for proper adrenal development. Several additional models that alter the level of Wnt activation have been subsequently generated that bypass the early β-catenin-dependent developmental defects and offer insight into the role of Wnt/β-catenin signaling during postnatal adrenal development and homeostasis (see section on Adrenal Cortex Regeneration).

Sonic Hedgehog Signaling. Similar to Wnt/β-catenin signaling, the hedgehog (HH) pathway is a paracrine signaling mechanism driven by secreted factors that is required for the growth and patterning of a wide range of tissues.[86] The three mammalian orthologs, sonic hedgehog (SHH), Indian hedgehog (IHH), and desert hedgehog (DHH), act by binding to the transmembrane receptor Patched1 (PTCH1). In the absence of these ligands, pathway activation is blocked by PTCH1, which represses the transmembrane protein smoothened (SMO). Ligand binding derepresses SMO through PTCH1 inactivation, resulting in intracellular signal transduction and activation of the GLI family of transcription factors. These include GLI1, which is a transcriptional target of active HH signaling and functions in a feedforward loop, as well as GLI2 and GLI3, which are primarily regulated posttranscriptionally. The activation of GLI transcription factors ultimately stimulates a wide range of target genes that promote cell proliferation and help direct tissue patterning.

In the adrenal gland, the HH pathway was first implicated in adrenal development when mutations in *SHH* and other components were identified in patients with holoprosencephaly and Pallister–Hall syndrome.[87-90] These patients have endocrine disorders, including adrenal hypoplasia, which suggested that SHH signaling may be critical for adrenal specification and growth. Consistent with this hypothesis, *in situ* hybridization studies found strong expression of *SHH* mRNA as early as E12.5 throughout the peripheral cortex of the developing gland.[17,91,92] Functional studies targeting *SHH* for deletion throughout the cortex (*Sf-1-Cre*^{high}*:Shhfx/fx*) resulted in significantly smaller adrenal glands in mice, consistent with a critical role for SHH signaling in early adrenal development.[17,92,93] Expanding upon these KO models, lineage-tracing studies further demonstrated that *SHH*-expressing cortical cells in the periphery of the gland give rise to cells of the inner cortex through centripetal migration and differentiation. In these studies, a tamoxifen-inducible *SHH:CreERT2* driver was combined with an enhanced GFP (EGFP)-based lineage reporter to label and track the *SHH* lineage over time.[17] Following the administration of tamoxifen at E14.5, *SHH*-derived steroidogenic cells (EGFP+/SCC+) were observed in the peripheral cortex after 5 days and extended into longer vertical columns by 28 days. These results demonstrate that, during embryonic development, *Shh*-expressing cells help to populate the inner cortex.

Although *SHH* is expressed in the peripheral cortex, its receptor, PTCH1, is expressed in the capsule.[93,94] Consequently, SHH ligands produced in the zG act in a paracrine fashion on capsular cells to activate downstream signaling. Consistent with this paradigm, GLI1, which acts in a feedforward loop to promote active SHH signaling, is also expressed in the capsule.[17,93,94] *Gli1+* capsular cells act as multipotent stem cells to give rise to adrenocortical cells during embryonic development, which has been visualized by lineage tracing experiments. Using *Gli1:CreERT2* combined with an EGFP-based reporter, cortical cells derived from the embryonic Gli1 lineage were tracked over time and found to progressively extend from the capsule into the cortex.[17] Notably, SHH signaling and the GLI1 lineage also significantly contribute to homeostatic maintenance of the adrenal cortex (see section on Adrenal Cortex Regeneration).

Fibroblast Growth Factor Signaling. The FGF signaling pathway is critical for the growth and development of many tissues.[95] The pathway is comprised of 22 secreted FGF ligands and four FGF receptors (FGFRs), with an additional level of complexity through alternative splicing of FGFR1 to 3. Notably, FGFs bind heparan and heparan sulfate, which results in their sequestration to the extracellular matrix. This allows for local release of ligands by heparan sulfate proteoglycans during tissue morphogenesis or in response to injury. Once activated, FGF signaling is generally transduced through activation of RAS/MAP kinase, PI3/AKT, or PLCγ, which promote mitogenic growth, cell survival, and cell migration.

In the developing adrenal gland, all four receptors have been detected,[96,97] along with three primary ligands.[97] These include *Fgf2* and *Fgf9*, which are expressed in the capsule, and *Fgf1*, which is expressed in the cortex. Functional studies have primarily focused on loss of *Fgfr2*, which results in significantly impaired adrenal growth. Mouse models of germline loss of *Fgfr2-IIIb*,[97,98] as well as conditional *Fgfr2* deletion,[15,99,100] consistently show a significant reduction in adrenal size that is often associated with lower rates of proliferation and increased apoptosis. Analysis of glomerulosa structure has further demonstrated that the loss of *Fgfr2* impairs the formation of rosettes and reduces zG function.[15] Together, these studies demonstrate that FGF signaling is critical for adrenocortical cell growth and proliferation, particularly in the early stages of development.

KEY POINTS

- The adrenal cortex is derived from the adrenogonadal primordium (AGP), which separates to give rise to both the adrenal gland and somatic cells of the gonad.
- Following AGP separation, neural crest cells invade the adrenal primordium to give rise to the medulla. These cells progressively coalesce and differentiate into catecholamine-producing chromaffin cells.
- Finally, mesenchymal cells encapsulate the gland and the cortex proceeds with functional zonation and differentiation.
- Proper specification and development of the adrenal gland is dependent upon a number of important factors, including nuclear hormone receptors (e.g., steroidogenic factor 1, Wilm's tumor suppressor, and DAX-1), transcription factors (e.g., GATA4 and GATA6), and growth factor signaling (e.g., Wnt/β-catenin, sonic hedgehog, and fibroblast growth factor).
- The initial adrenal primordium is composed of a large fetal zone that functions to produce high levels of dehydroepiandrosterone, dehydroepiandrosterone sulfate, and cortisol, and a smaller definitive zone that ultimately becomes the adult cortex after fetal zone involution.

Fetal Adrenal Development

Fetal Zone Versus Definitive Zone. The initial adrenal blastema is composed of two distinct zonal compartments: a large region of fetal zone (FZ) cells surrounded by a smaller adult, definitive zone (DZ). In humans, the fetal adrenal cortex grows disproportionately after encapsulation.[38,101] This expansion within the FZ enables the adrenal gland to significantly increase in size (nearly 10-fold), becoming one of the largest organs at term and comparable in size to the kidney. Surrounding the FZ, there is a narrow band of definitive cortex. The DZ initially contains densely packed basophilic cells that are lipid poor. However, these cells are proliferative and begin to accumulate more cytoplasmic lipid after midgestation as they become increasingly steroidogenic. A so-called transitional zone (TZ) takes shape between the DZ and FZ and is capable of producing cortisol. By late gestation, the fetal adrenal begins to resemble the adult cortex, with the DZ and TZ taking on the morphology and function of the zG and zF, respectively. The FZ undergoes involution after birth (see section on Fetal Zone Involution), causing a significant decrease in adrenal weight just after birth.

Function of the Fetal Zone. The FZ is comprised of large eosinophilic cells with high steroidogenic activity. Specifically, FZ cells strongly express CYP17A1, which converts pregnenolone to DHEA. This results in the production of large amounts of DHEA and DHEA-S (up to 200 mg per day) that are then converted by the placenta into estrogens.[101] As the human placenta cannot make estrogens *de novo*, the generation of estrogens through this process is critically important to help maintain normal pregnancy. In addition to DHEA, CYP17A1 also mediates cortisol production in the FZ, which promotes maturation of the fetal organ system. Thus, the FZ primarily functions to produce steroids that help maintain intrauterine development.

Adrenocorticotropic Hormone-Driven Fetal Zone Growth. The principal regulator of fetal adrenocortical development is ACTH. This tropic regulator of the postnatal adrenal cortex is known to promote both hypertrophy and hyperplasia of adrenocortical cells. However, rapid FZ growth predominately results from hypertrophy, whereby the size of each cell rather than the number of cells increases.[101]

Unexpectedly, rather than increasing during FZ expansion, plasma ACTH levels decrease by nearly 50% during prenatal adrenal development.[102] This paradox is partially explained by the progressive sensitization of FZ cells. While not maximally sensitive initially, FZ cells become more ACTH-responsive over time through mechanisms that include increased expression of the ACTH receptor (and thus enhanced ACTH binding capacity).[103,104] This adaptive ACTH response is thought to help prepare the fetus for the stress of birth and postnatal development by increasing the steroid-producing capacity of the FZ.

POSTNATAL ADRENAL DEVELOPMENT

Fetal Zone Involution

Human Adrenal. The first notable event in postnatal adrenals is the rapid involution of the FZ, followed by progressive expansion and zonation of the adult definitive cortex (Fig. 85.5). Whereas the FZ

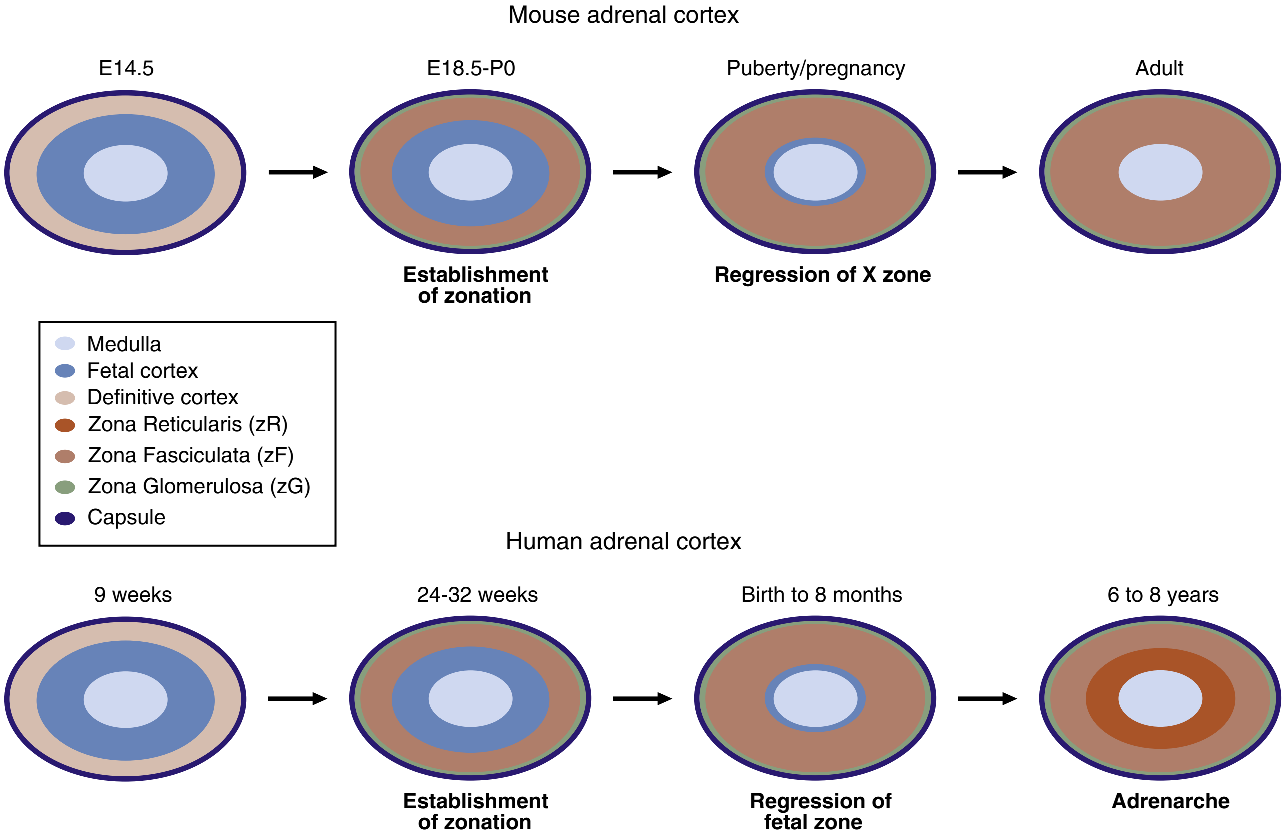


Fig. 85.5 **Pre- and postnatal adrenal development in mice and humans.** Functional zonation of the adrenal cortex starts at embryonic day 18.5 in mouse and between 24–32 weeks in humans. Whereas the fetal zone immediately regresses in humans after birth, its regression is delayed until puberty in male mice or first pregnancy in female mice. In human adrenals, zona reticularis develops during adrenarche (6–8 years), between the zona fasciculata and the medulla. There is no zona reticularis in the mouse adrenal cortex.

represents up to 70% of total adrenal volume at birth (up to 85% of the cortex), it only takes up 3% of the gland after 8 months.[105] FZ involution is evident immediately after birth and can be divided in two phases.[105,106] The first phase is characterized by rapid regression of FZ until the end of the second week postpartum. This phase is associated with hemorrhage and high levels of apoptosis, which is maximal during the second week of life (30% of FZ cells). The second phase is characterized by slower involution and relatively constant levels of apoptosis (20% of FZ cells). The FZ mainly synthesizes androgens, and during fetal life, concentrations of DHEA follow the increase in FZ size in the last two trimesters of pregnancy.[107] Following FZ involution, adrenal androgen levels rapidly decline until the onset of adrenarche at around 6 to 9 years.[108,109] The mechanisms involved in FZ regression in the human adrenal gland remain elusive. Early reports showed that levels of plasma DHEA-S and urinary 3β-OH-5ene steroids remained elevated for many weeks following parturition in preterm babies, consistent with persistence of the FZ in this context. This suggested that FZ involution was controlled by the duration of gestation rather than by parturition itself.[110] However, more recent evidence relying on ultrasonography and measurement of serum cortisol and androstenedione concentrations reported similar adrenal involution, and endocrine activity in preterm and full-term babies.[109] This suggests that parturition itself triggers FZ involution.

A number of factors have been proposed to control human FZ maintenance and involution (Fig. 85.6). Anencephalic fetuses display marked adrenal atrophy, and *in vitro* treatment of human fetal adrenal cells with ACTH stimulates their growth and endocrine activity.[111–115] Although fetal ACTH is required to support normal FZ growth and activity during fetal life, there is no decrease in ACTH at the time of FZ involution,[110,115] suggesting that this process is dependent on other mechanisms. Experimental evidence suggests involvement of transforming growth factor β (TGFβ) signaling, and in particular of activins and inhibins, in this process. These are peptide hormones formed by the dimerization of α, βA, and βB inhibin subunits, encoded by three distinct genes. Activins A or B are formed by homodimerization of βA or βB subunits, respectively, whereas inhibins A or B are formed by heterodimerization of α subunits with either βA or βB subunits, respectively. Activins signal through ACVRI and ACVRII activin receptors, which results in phosphorylation of SMAD2 and SMAD3. These interact with SMAD4 and migrate to the cell nucleus, where they trigger expression of target genes that can be involved in apoptosis. Inhibins, on the other hand, can sequester ACVRII by interacting with betaglycan. This results in inhibition of activin signaling.[116] βA and βB subunits are expressed throughout human adrenal cortex, with predominant expression in zG, whereas the α subunit is mostly found in the zR and in the fetal adrenal.[117-120] ACVRI/II receptors and betaglycan are expressed in the adult and fetal adrenal cortex.[121] Fetal adrenal cells also express SMADs3/4/6/7, suggesting that inhibin/activin signaling may play a role in FZ involution. Consistent with this idea, TGFβ and activin A were shown to specifically inhibit growth and induce apoptosis of human FZ cells in culture.[106,118,122] This suggests that these TGFβ/activin/inhibin family members may play a role in FZ involution. However, this would require induction of activin signaling in the human fetal adrenal after birth. Interestingly, evaluation of

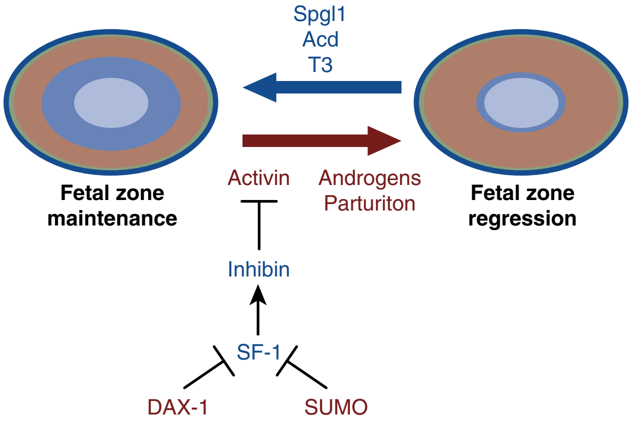


Fig. 85.6 **Factors involved in maintenance or regression of the fetal cortex.** Hormonal manipulations and mouse genetics have shown that activins, androgens, and parturition favor fetal cortex regression. This can be counteracted by inhibin, which can be stimulated by steroidogenic factor 1 (SF-1). This effect of SF-1 is inhibited by its sumoylation and by the antagonistic activity of DAX-1. In contrast, Spgl1, Acd, and triiodothyronine may be involved in maintenance of the fetal cortex. Most of the molecular mechanisms involved in these processes remain elusive.

activins and inhibins in the artery and vein of the umbilical cord of term infants showed that the placenta was the major source of inhibin B, whereas the fetus was the main producer of activin A.[123] It is therefore tempting to speculate that parturition, by removing a major source of inhibin, could alter the activin/inhibin ratio to favor activin signaling and FZ involution. Although these mechanisms are difficult to study in humans, studies in mouse models have provided further insight into potential regulators of FZ homeostasis (Fig. 85.6).

Mouse Adrenal. The mouse X zone (see section on Mouse X Zone and Rat Undifferentiated Zone) is exclusively derived from fetal adrenal cells that express the fetal enhancer of *Sf-1*,[24,50,124] and may thus be the mouse equivalent to the human FZ. However, it does not express CYP17,[20] which does not allow production of adrenal androgens. Rather, the X zone is characterized by expression of 20αHSD,[25] which appears at around 3 weeks of life.[25,125,126] Another key feature that distinguishes X zone from the human FZ is the kinetics of its involution (Fig. 85.5). Indeed, whereas the FZ regresses postnatally in both sexes in humans, X zone involution is observed after around 4 months or during the first pregnancy in females and after puberty in males.[23,25,127]

Role of Sex Hormones and Prolactin. The sexual dimorphism associated with X zone regression suggested an involvement of sex hormones in this phenomenon. Indeed, injection of testosterone propionate or the nonaromatizable androgen dihydrotestosterone (DHT) to nulliparous adult females results in X zone regression, which can be reversed upon removal of DHT after 7 days of treatment. Conversely, gonadectomy in prepubertal males prevents X zone regression and results in growth of a secondary X zone in postpubertal males.[23,25,128,129] This indicates that androgens, but not estrogens, directly induce involution of the X zone in male mice at puberty. These observations have been further confirmed by the targeted ablation of the androgen receptor (AR) in steroidogenic cells of the adrenal cortex, which resulted in a dramatic reduction in apoptosis and X zone maintenance in male adrenals at puberty. Interestingly, AR ablation resulted in maintenance of a larger X zone than castration alone, suggesting that AR may exert androgen independent effects on X zone homeostasis[129] (Fig. 85.6).

The mechanisms associated with X zone regression during pregnancy are a lot more elusive. Concordant observations have shown a rapid decline in 20αHSD activity and morphological X zone during pregnancy, which are not restored after parturition.[23,25,125,127] Ovariectomy is associated with transient X zone thickening,[130] and progesterone injection induces rapid regression of the X zone.[125,131] However, this effect is only observed in females with intact ovaries,[131] suggesting that the effect of progesterone is dependent on its conversion to androgens by the ovary, particularly during pregnancy.[132] Consistent with this, testosterone peaks at 9 days postcoitum (dpc), which is coincident with onset of X zone regression between 8 and 12 dpc in gravid females.[25,133] Induction of pseudopregnancy by mating female mice with vasectomized males has no effect on X zone morphology and 20αHSD activity, compared with nulliparous females.[25] This suggests that the elevation of progesterone and estrogens induced during the first 8 days of pseudopregnancy is not sufficient to trigger the first phases of X zone involution, even though hormonal levels are comparable with pregnancy. Interestingly, removal of pups, which results in decreased prolactin secretion in deprived dams, is associated with partial recovery of 20αHSD activity and partial regrowth of the X zone.[25] This suggests that the rise in prolactin secretion during pregnancy and before weaning may be involved in X zone involution. However, prolactin injection to nulliparous females has no impact on X zone homeostasis. Similarly, inactivation of STAT5 signaling pathway, which plays a central role in prolactin response, has no impact on the X zone of either males or females.[25] Altogether, these data suggest that the prolactin increase induced by suckling may have an impact on X zone involution in primiparous females, but does not significantly contribute to regression during pregnancy or at puberty in males.

Role of Thyroid Hormones. Thyroxin and thyroid extracts induce enlargement of a histological X zone in the adrenal cortex of mice.[134,135] Conversely, hypothyroid *Hyt* mutant mice that harbor a mutation of the thyrotropin receptor, have a poorly developed X zone,[136] suggesting that thyroid hormones and their receptors may play a role in X zone homeostasis. Recent evidence, based on a LacZ reporter inserted at the *Thrb* locus, showed that thyroid hormone receptor beta 1 (TRβ1) is expressed in mouse inner adrenal cortex from E14.5 into the postnatal period. Its expression pattern is reminiscent of and partly overlaps with that of 20αHSD. However, it is more broadly expressed, expanding in the inner zona fasciculata, and although it declines in males by postnatal day 28, it persists after pregnancy in females. Interestingly, hormonal treatments at different time points showed that triiodothyronine (T3) could accelerate appearance of the 20αHSD+ zone after birth, induce hypertrophy of the X zone before puberty, and counteract X zone regression at puberty in males, although it could not induce regeneration after regression.[126] Analysis of *Thrb*-deficient mice showed that these effects of exogenous T3 were dependent on the thyroid hormone receptor. However, *Thrb-/-* adrenals were indistinguishable from wild-type at baseline.[126] Although this may reflect compensation by other thyroid hormone receptors, it suggests that thyroid hormones are not required for normal development of the X zone, but that they may be involved in modulating its homeostasis and function. Whether this has relevance to the human fetal adrenal cortex is currently unknown.

Role of Activin/Inhibin Signaling. Consistent with findings in human adrenals, the activin/inhibin/TGFβ signaling pathway may also control fetal adrenal cell growth in mouse adrenals (Fig. 85.6). Indeed, all four subunits of activin receptors (ACTR-IA, ACTR-IB, ACTR-IIA, and ACTR-IIB) are expressed in the mouse adrenal cortex, and ACTR-IA, ACTR-IIB, and SMAD2, an essential component of the TGFβ signaling pathway, are expressed more specifically in the X zone.[137] Interestingly, activin treatment has a specific growth inhibitory effect on primary cultures of adrenals containing X zone cells and is able to induce apoptosis in the presumptive X zone of whole adrenals in culture.[137] The potential role of activin in controlling X zone homeostasis is further supported by accelerated X zone regression in inhibin-deficient mice that show increased levels of gonadal and adrenal activin secretion.[137,138] This effect is counteracted by gonadectomy, which induces unrestricted growth of presumptive X zone cells and SMAD3-dependent adrenal cortex tumor development.[137,139] However, there is no reported adrenal gland phenotype in mice deficient for activin.[140]

Genetic Models Associated with Failure to Develop an X Zone.

A number of genetically engineered or mutant mice display X zone aplasia. The adrenocortical dysplasia gene (*ACD*) encodes TPP1, which plays a central role in telomere protection within the TRF1 complex.[141] Mice harboring a spontaneous homozygous splice donor mutation in the *Acd* gene display multiple developmental defects, including cytomegalic adrenocortical dysplasia and glucocorticoid insufficiency. They are also characterized by X zone aplasia.[142,143] Adrenal aplasia is associated with massive induction of cellular senescence and can be partly reversed by genetic inactivation of *Tp53.*[144] However, it is unclear if reduction of senescence following *Tp53* inactivation also rescues X zone differentiation.[144]

Sgpl1 mutant mice also show X zone aplasia.[145] *Sgpl1* encodes the enzyme responsible for the last step of sphingosine 1 phosphate (S1P) catabolism, which works as a signaling molecule. Interestingly, SGPL1 is expressed in the human fetal adrenal (from Carnegie stage 19 onwards) in both the FZ and DZ. In the adult adrenal, SGPL1 expression is found throughout the cortex, but is more concentrated in the zR. Besides X zone aplasia, *Sgpl1-/-* mice also show altered cortical zonation and atrophic zF, associated with lower CYP11A1 and CYP11B2 expression. This is consistent with the primary adrenal insufficiency observed in patients with homozygous inactivating mutations in *SGPL1.*[145,146] The molecular mechanisms underpinning adrenal insufficiency and X zone aplasia in this context are unclear. However, *Sgpl1* inactivation results in accumulation of S1P,[145,147] which can be converted back to sphingosine by S1P phosphatases. Whereas S1P has been shown to stimulate steroidogenesis in H295R cells in culture,[148] sphingosine is an inhibitory ligand for SF-1 and acts to maintain it in an inactive conformation.[149,150] This suggests that alterations of SF-1 activity may be involved in X zone homeostasis (Fig. 85.6).

Genetic Models Associated with Persistent X Zone. Activity of the nuclear receptor SF-1, the master regulator of steroidogenic tissues development and differentiation,[151] can be regulated by a number of posttranslational modifications, including phosphorylation,[152] acetylation,[153] and sumoylation.[154-156] This modification, which consists in the reversible addition of SUMO modules on two lysine residues of SF-1 (K119 and K194), does not have a global effect on SF-1-mediated transcription, but rather dampens recognition of a subset of SF-1 target genes, coined SUMO-sensitive targets, which are essentially transcribed when SF-1 is hyposumoylated.[156] *In vivo*, the replacement of the two lysine residues by nonsumoylatable arginine residues, through homologous recombination (SF-1^{2KR}), results in abnormal adrenal development. This is characterized by aberrant expression of gonadal markers (*Sox9*, *Amhr2*, *Aldh1a1*), upregulation of HH signaling, hypoaldosteronism, and persistence of a 20αHSD-positive X zone after puberty, even though it eventually regresses later in life.[157,158] Although the molecular mechanisms underlying persistence of the FZ were not evaluated in this model, previous data had shown that inhibinα is a SUMO-sensitive target of SF-1, the expression of which is upregulated by nonsumoylatable SF-1 mutants.[156] This may account for an imbalance in activin/inhibin signaling that would prevent normal apoptosis of the X zone (Fig. 85.6).

DAX-1 is an atypical non-DNA-binding nuclear receptor that has been shown to repress SF-1 transcriptional activity[159-161] (see section on DAX-1 (*NR0B1*)). DAX-1 mutations in patients are associated with X-linked adrenal hypoplasia congenita (AHC).[61,62] Patients with AHC present with neonatal adrenal insufficiency associated with cytomegalic hypoplastic adrenals.[162] Surprisingly, mice with *Dax-1* gene ablation have no signs of adrenal insufficiency in early life, but show temporary persistence of the X zone after puberty.[63,158] Interestingly, analysis of the fetal adrenal lineage using the fetal adrenal enhancer of *Sf-1* to drive expression of a LacZ reporter gene (*FAdE:LacZ*) shows persistence of FAdE activity in 20αHSD+ cells of both *Sf-12KR* and *Dax-1* KO adrenals, confirming their fetal rather than adult origin.[158]

Altogether, these data suggest that stimulation of SF-1 transcriptional activity, either by inhibiting its sumoylation or by inactivation of a key repressor, can result in temporary persistence of the X zone. Regression of the X zone may thus be mediated by inhibition of SF-1 activity (Fig. 85.6). In cell culture, sumoylation of SF-1 can increase interactions between SF-1 and DAX-1, resulting in almost complete inactivation of *Sf-1* fetal enhancer activity.[158] This may provide a mechanism that would explain extinction of *Sf-1* expression in fetal adrenal cells, resulting in their death, consistent with the role of SF-1 in the survival of steroidogenic cells.[41,163,164] The recent observation that PKA signaling inhibits sumoylation in the adrenal cortex may add another level of control to these processes.[165]

KEY POINTS

- The fetal cortex regresses after birth in both mouse and human. Whereas human fetal cortex regresses immediately after birth, regression is delayed until puberty in male mice and after the first pregnancy in female mice.
- Involution of the fetal cortex involves apoptosis and is mostly controlled by activin/inhibin/transforming growth factor β signaling. This may also involve sex hormones, particularly in mice.
- Genetically engineered mouse models show that tampering with SF-1 dosage has a key role in controlling homeostasis of the fetal zone.
- Lineage-tracing studies in mice have shown that the definitive cortex, which takes over the entire gland after fetal zone regression, originates from cells of the fetal cortex during embryonic development.

Origin of the Definitive Cortex. Lineage-tracing studies relying on FAdE have shown that most of the definitive cortex is derived from FAdE+ fetal adrenal cells at E11.5. However, the capacity of FAdE+ cells to give rise to adult cells is lost after E14.5.[24,50] Stronger expression of DAX-1 in cells of the definitive cortex is thought to play a role in the extinction of FAdE in this lineage. In these cells, *Sf-1* expression is potentially maintained through mobilization of an adult enhancer that has yet to be identified.[24,50] In addition, analysis of the role of SHH signaling in the adrenal cortex showed that SF-1-/GLI1+ capsular cells could give rise to some definitive cortical cells at E14.5 and onwards, albeit with low efficiency.[17,93] Further lineage-tracing studies suggest that a very small subset of SF-1-/GLI1+ cells (~7%) may be derived from FAdE+ cells.[166] However, such capsular lineage is not found with *Sf-1-Cre*, suggesting that this may be the result of ectopic expression of the short FAdE construct. Altogether, these observations suggest that both fetal SF-1+/FAdE+ and capsular SF-1-/GLI1+ cells can give rise to the definitive cortex. Whether similar mechanisms also exist in human adrenals is unknown, but the high conservation of FAdE sequences in humans is in support of this hypothesis.

Establishment of Adrenal Cortex Zonation

Functional zonation is established in cells of the definitive cortex in the perinatal period. It manifests as the establishment of three (primates) or two (most rodents) concentric zonae with distinct histological and functional features, that undertake the endocrine activity of the gland[167] (see section on Adrenal Cortex). In rodents' adrenals, zonal restriction of steroidogenic enzymes expression is established during embryonic development. CYP11B1 (zF) is expressed in the entire cortex from E13.5 onwards. In contrast, CYP11B2 and angiotensin receptor AT1b (zG) are expressed at the periphery of the gland after E18.5 (Fig. 85.5). At this stage, CYP11B1 and CYP11B2 are expressed in mutually exclusive areas of the cortex.[17,168-170] In human adrenals, steroidogenic enzymes responsible for glucocorticoids and androgens synthesis are readily detected during the first trimester, mostly in the fetal and transition zones. Their expression markedly increases during the second trimester, concomitant with increased production of glucocorticoids.[171-173] Consistent with the lack of aldosterone production during the first and second trimester,[172] CYP11B2 appears in the DZ at 24 weeks of gestation and is maximal at 32 weeks[172,174] (Fig. 85.5). Establishment of these zonal expression patterns is the basis for functional adult adrenal cortex zonation, which is maintained throughout life.[169,175-177] Genetic and cell lineage tracing studies in mouse models have allowed understanding of the molecular mechanisms involved in establishing and maintaining these differentiation patterns at homeostasis and in response to hormonal stimuli and cell renewal.

Differentiation of Zona Glomerulosa. Ang II and plasma potassium concentrations are the main regulators of zG endocrine activity.[178] Activation of the renin-angiotensin system following hyponatremia or hyperkaliemia induces a rapid expansion of zG, an increase in the number of CYP11B2+ cells, and an induction of CYP11B2 expression, allowing for increased aldosterone secretion. Conversely, inhibition of the renin-angiotensin system using drugs such as Captopril results in a rapid atrophy of zG.[179-181] However, mice deficient for the two type I receptors of Ang II (*Agtr1a-/-;Agtr1b-/-*) have normal adrenal histology.[182] This indicates that Ang II signaling is not involved in the establishment and initial differentiation of zG, even though it plays a central role in remodeling the zone to modulate the supply of mineralocorticoids. Consistent with this idea, ectopic chemogenetic activation of Gαq signaling, which is downstream of angiotensin receptors, results in ectopic expression of CYP11B2 within the zF and increased renin-independent aldosterone production.[183]

Over the last decade, analysis of genetically engineered mouse models has allowed identification of the canonical Wnt/β-catenin signaling pathway as one of the key regulators of zG differentiation and homeostasis[15,85,184-186] (Fig. 85.7). Whereas β-catenin is activated throughout the developing adrenal cortex at E12.5, its activity is highest in the presumptive zG by E18.5.[82] In the adult, β-catenin is expressed and accumulated at high levels throughout the zG.[82,85,184,185] This area is also characterized by accumulation of the Wnt targets LEF1 and AXIN2 and by expression of Wnt reporter constructs.[82,85,184,185] This suggests that strong Wnt pathway activation in zG cells may play a role in the differentiation of this zone.

Consistent with this idea, conditional inactivation of *Ctnnb1* (the gene encoding β-catenin) within zG cells, using *Cyp11b2-Cre* as the driver, results in smaller and morphologically altered glomeruli.[15] Conversely, constitutive activation of β-catenin within CYP11B2+ cells (*Cyp11b2-Cre* driver) results in continuous expansion of zG, increased rosette frequency, and a significant increase in plasma aldosterone (albeit without renin suppression).[15,186] This is further confirmed by stochastic activation of β-catenin throughout the entire cortex (*Akr1b7-Cre*), which results in zG expansion, ectopic differentiation of scattered zG cells within the zF, and elevated plasma aldosterone levels.[85,187] The observation of primary aldosteronism in these models suggested that deregulated Wnt pathway activity could be involved in hyperaldosteronism in patients. Consistent with this idea, 33/47 patients with aldosterone-producing adenomas (APA) had constitutive Wnt pathway activation.[188] This can result from activating mutations of *CTNNB1*, found in approximately 5% of APA patients[189,190] or from downregulation of Wnt inhibitors such as SFRP2.[188] Altogether, these data show that β-catenin signaling plays a central role in zG differentiation. Mechanistic investigations have shown that β-catenin, either directly or indirectly through NURR1 and NUR77 nuclear receptors, stimulates transcription of *CYP21*, *CYP11B2,* and *AT1R.*[188] This provides an explanation for increased aldosterone production following Wnt pathway activation. However, this does not account for increased rosette formation. Interestingly, the growth factor receptor FGFR2 is upregulated in response to constitutive β-catenin activation and its genetic inactivation in CYP11B2+ cells disrupts rosette morphology and frequency.[15] This suggests that FGF signaling may play a role downstream of Wnt signaling to establish the specific three-dimensional structure of the zG.

How is Wnt/β-catenin activated to allow for zG differentiation? A number of Wnt ligands, including WNT2b, WNT4, WNT5a, WNT5b, WNT9a, and WNT11, are expressed at the periphery of the adrenal cortex during embryonic development and/or in the adult adrenal.[85,184,185,191] However, only the role of WNT4 has been extensively studied so far (Fig. 85.7). It is expressed in the developing adrenal glands of mice as early as E11.5 and shows restricted expression at the periphery of the cortex by E14.5[192] and within zG in the adult adrenal,[85] suggesting that it may be involved in zG differentiation. Indeed, full-body ablation of *Wnt4* is associated with reduced zG differentiation and decreased plasma aldosterone concentration.[192] This is further confirmed by conditional ablation of *Wnt4* in adrenal steroidogenic cells using *Sf-1-Cre.*[85,184] A role for WNT4 in human adrenals is suggested by the observation that patients with mutations in *WNT4* associated with SERKAL (SEx Reversion, Kidneys, Adrenal and Lung dysgenesis) syndrome have signs of adrenal dysgenesis.[193] Conversely, overexpression of WNT4 in human adrenocortical H295R cell lines induces CYP11B2 expression and aldosterone production.[194]

Another important aspect of Wnt pathway control is modulation of FZD receptor availability at the plasma membrane. This involves the membrane E3 ubiquitin ligases ZNRF3 and RNF43 that target FZDs for degradation[76,77] and the RSPOs (RSPO1–4) that antagonize ZNRF3 and RNF43 by inducing their membrane clearance.[79,80,195] *Rspo3*, and to a lesser extent *Rspo1,* are expressed in the adrenal capsule,[184,185] and genetic ablation of *Rspo3* in the capsule results in a marked reduction of Wnt pathway activity and zG differentiation, associated with decreased expression of WNT4.[184] This suggests that RSPO3 from the capsule plays a key paracrine function to stimulate Wnt signaling in the underlying zG (Fig. 85.7). Whereas RNF43 is barely detectable in the adrenal, ZNRF3 is expressed throughout the cortex.[185] Consistent with its low level of expression, adrenal-specific ablation of *Rnf43* does not alter adrenal homeostasis. In contrast, mice with *Znrf3* inactivation show robust hyperproliferation and hyperplasia, resulting in a nearly 8.5-fold increase in adrenal size by 6 weeks. This phenotype can be significantly reversed by ablation of Porcupine (*Porcn*), which is an enzyme required for Wnt ligand secretion, indicating that the hyperplasia is at least in part Wnt-dependent. However, in contrast to mice bearing constitutive activation of β-catenin, there is no alteration of zonation patterns, and the hyperplasia is focused on zF rather than

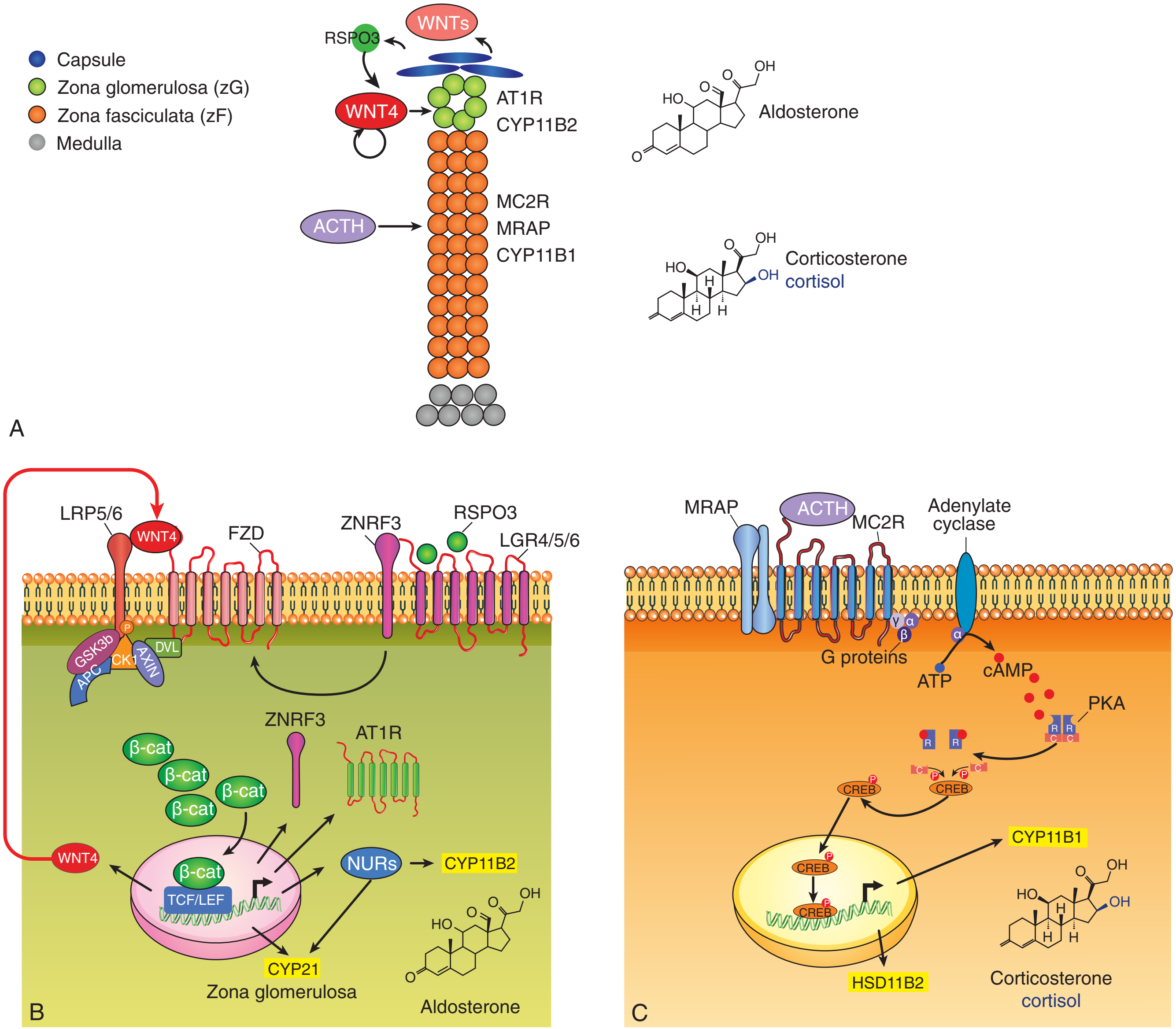

Capsule
Zona glomerulosa (zG)
Zona fasciculata (zF)
Medulla
RSPO3
WNTs
WNT4
AT1R
CYP11B2
ACTH
MC2R
MRAP
CYP11B1
Aldosterone
Corticosterone
cortisol
A
LRP5/6
WNT4
FZD
ZNRF3
RSPO3
LGR4/5/6
P
GSK3b
CK1
DVL
AXIN
APC
ZNRF3
AT1R
β-cat
TCF/LEF
WNT4
NURs
CYP11B2
CYP21
Zona glomerulosa
Aldosterone
B
MRAP
ACTH
MC2R
Adenylate cyclase
γ α β
G proteins
α
ATP
cAMP
PKA
R
C
CREB
P
CYP11B1
HSD11B2
Corticosterone
cortisol
C

Fig. 85.7 **Pathways involved in functional zonation of the adrenal cortex. A,** WNT signaling and adrenocorticotropic hormone (ACTH) signaling control differentiation of the zona glomerulosa (zG) and zona fasciculata (zF), respectively. Schematic representation of the mouse adrenal cortex. Canonical WNT signaling pathway activated by WNT4 (produced by zG cells) and RSPO3 (produced by capsule cells) stimulates differentiation of the zG and production of aldosterone. ACTH/cAMP/PKA signaling pathway stimulates differentiation of the zF and production of corticosterone in rodents or cortisol in humans. AT1R and CYP11B2 are specifically expressed in the zG, whereas MC2R, MRAP, and CYP11B1 are specifically expressed in the zF. **B,** Molecular mechanisms of zG differentiation. In the zG, canonical WNT signaling is activated by RSPO3, which inhibits the membrane E3 ubiquitin ligase ZNRF3, preventing degradation of Frizzled (FZD) receptors. This allows stimulation of FZD by WNT4 produced within the zG, which results in accumulation and nuclear translocation of β-catenin (β-cat). β-catenin, by interacting with LEF/TCF transcription factors, stimulates expression of its target genes, including WNT4. This results in reinforcement of WNT pathway activation within zG cells. β-catenin also stimulates expression of AT1R, allowing response to Angiotensin II, of CYP21 and of the zG-specific CYP11B2 to allow production of aldosterone. This is achieved directly (AT1R and CYP21) and indirectly (CYP11B2 and CYP21) through stimulation of NURR1 and NUR77 expression (NURs). ZNRF3 is also a direct target of β-catenin, which ensures a negative feedback on WNT pathway activation to maintain tightly controlled levels of activation. **C,** Molecular mechanisms of zF differentiation. Expression of MC2R and the accessory protein MRAP allows response of zF cells to pituitary ACTH. This induces recruitment of the Gα protein to the adenylate cyclase, which converts ATP to cAMP. By binding to the regulatory subunits of PKA (R), cAMP allows release of the catalytic subunits (C), which phosphorylate the transcription factor CREB. Phosphorylated CREB migrates in the nucleus, where it stimulates transcription of a number of steroidogenic enzymes, among which the zF-specific CYP11B1, which allows production of corticosterone (mouse) or cortisol (human).

zG.[185] This is correlated with normal, high levels of Wnt activation in the zG. *Znrf3-/-* adrenals instead display sustained low Wnt pathway activity throughout the zF. Moreover, heterozygous inactivation of *Ctnnb1* partially rescues the hyperplastic phenotype. These results suggest that, whereas high levels of canonical Wnt signaling achieved by direct β-catenin stabilization primarily drive zG differentiation, and to a lesser extent hyperproliferation, moderate levels resulting from *Znrf3* inactivation promote proliferation and are not sufficient for glomerulosa differentiation. This is consistent with the observation of moderate Wnt activation and high levels of proliferation in the normal outer zF.[185]

Differentiation of Zona Fasciculata. The pituitary hormone ACTH is a key regulator of zF differentiation. It acts through MC2R and the accessory protein MRAP, which are expressed at higher levels in the zF compared with the zG, to stimulate cAMP/PKA signaling.[196-198] In turn, PKA through direct phosphorylation of CREB, and indirect effects on SF-1 and some other transcription factors, stimulates expression of zF-specific genes such as *CYP11B1*, *Akr1b7*, or *HSD11B2*[199-201] (Fig. 85.7). Consistent with a specific role of ACTH/cAMP/PKA signaling in zF differentiation, chronic administration of ACTH induces hypertrophy of zF and increases glucocorticoid secretion, whereas dexamethasone administration induces zF atrophy and inhibition of glucocorticoid production, without alteration of zG.[124,202-204] In further support of this concept, mutations in either *MC2R* or *MRAP* in patients are associated with zF degeneration and unresponsiveness to ACTH, resulting in familial glucocorticoid deficiency.[197,205,206] The key role of the ACTH/cAMP/PKA pathway in zF differentiation is further confirmed by massive hypoplasia of the zF and failure to produce glucocorticoids in *Pomc*, *Mc2r,* and *Mrap* KO mice.[207-210] Although *Pomc* and *Mc2r* KO mice also have mild mineralocorticoid insufficiency, consistent with a minor role of ACTH in acutely controlling aldosterone secretion,[211] *Mrap* KO mice have normal circulating aldosterone levels.[210] This suggests that MRAP plays a more zF-specific role in the ACTH response. Isolated glucocorticoid insufficiency and aberrant zF differentiation are also observed in mice with adrenal-specific ablation of the histone-methyltransferase EZH2, which results in overexpression of phosphodiesterases PDE1B, 3A, and 7B, and a subsequent reduction in cAMP/PKA signaling.[67] At the other end of the spectrum, patients with inactivating mutations of the regulatory subunit RIα of PKA (encoded by *PRKAR1A*)[212] or of the phosphodiesterases PDE11A[213] and PDE8B[214] have increased cAMP/PKA signaling, which is associated with hypercortisolism and ACTH-independent Cushing syndrome.[215] This is also the case for patients with activating mutations of the α-subunit of the Gs protein (encoded by *GNAS1*)[216] or of the catalytic subunit of PKA (encoded by *PRKACA*).[217] In most cases, the endocrine defect is restricted to hyperproduction of cortisol, confirming the prominent role of PKA signaling in zF differentiation and homeostasis. Consistent with this idea, mice with conditional ablation of *Prkar1a* in the adrenal cortex have expansion of zF and autonomous corticosterone production.[85,128,218] Similarly, *Pde8b* KO mice have elevated urinary corticosterone, as a result of hypersensitivity to ACTH stimulation.[219]

Differentiation of Zona Reticularis. Adrenarche, which occurs between 6 and 8 years in humans, corresponds to a marked increase in DHEA and DHEA-S production by cells of the zR, which is morphologically apparent by 3 years of age but only reaches its maximum size at around 13 years.[22,220,221] The molecular mechanisms that control differentiation of zR remain elusive. Interestingly, constitutive activation of PKA, resulting from conditional inactivation of *Prkar1a* in cells of the definitive cortex, results in development of a reticularis-like zone in genetically engineered mice.[128] Indeed, in contrast to wild-type mice, which do not develop any zR features, *Cyp11b2-Cre;Prkar1afl/fl* mice express high levels of CYP17 in cells located between the zF and medulla and produce cortisol, DHEA, and DHEA-S, which are normally not found in mice. As in humans, this is associated with high levels of CYB5 and low levels of HSD3B expression in the reticularis-like zone.[128] This suggests that increased PKA activity may be sufficient to trigger zR development. This is consistent with data showing that dexamethasone can suppress adrenal androgens production[222-224] and failure of adrenarche in children with ACTH receptor deficiency.[225] However, there is no evidence of a concomitant rise in DHEA and ACTH levels during adrenarche,[226-228] suggesting that some other, unidentified factors may trigger zR differentiation.

KEY POINTS

- Functional zonation of the adrenal cortex is established perinatally by zone-restricted expression of Angiotensin II receptor (zona glomerulosa), MC2R/MRAP (zona fasciculata), CYP11B2 (zona glomerulosa), and CYP11B1 (zona fasciculata).
- Zona glomerulosa differentiation is dependent on canonical WNT/β-catenin signaling and is modulated by Angiotensin II.
- Zona fasciculata differentiation is dependent on adrenocorticotropic hormone/cAMP/PKA signaling.
- The mechanisms involved in zona reticularis differentiation remain elusive, although the process may involve PKA signaling.

MAINTENANCE OF ADULT ADRENAL CORTEX FUNCTION BY CORTICAL CELL RENEWAL

Function of the adult adrenal cortex is maintained throughout life by a constant renewal process relying on stem and progenitor cells in the outer capsule and peripheral cortex, centripetal cell migration, and sequential differentiation of migrating cells as a function of their position along the corticomedullary axis[167] (Fig. 85.8). Some of the first evidence for the presence of progenitor cells within the capsule and outer cortex was provided by experiments showing that, following rat adrenal enucleation, remaining tissues were able to form a fully differentiated and hormonally active cortex after autotransplantation to the ovary.[229,230] This suggested that adrenal cortex progenitor cells were present in the capsular/subcapsular compartment. Consistent with this, most proliferation was observed in the outer cortex and could be strongly stimulated by unilateral adrenalectomy to allow for compensatory growth of the remaining adrenal.[44,230,231] Labeling of cells with a single injection of BrdU, a synthetic nucleoside incorporated during DNA replication, followed by different periods of chase allowed demonstration of a continuous stream of BrdU+ cells from the outer to the inner cortex, associated with cell death as they reached the medulla.[232-236] Interestingly, whereas BrdU+ cells at the periphery did not express steroidogenic differentiation markers, BrdU+ cells within the cortex were differentiated.[236] Altogether, these experiments suggested that undifferentiated progenitor cells in the capsule and outer cortex could proliferate and migrate centripetally towards the medulla to reconstitute pools of differentiated cells throughout the cortex. This theory was demonstrated by lineage-tracing and genetic studies that have identified a number of stem and progenitor cell populations and the molecular pathways involved in their recruitment.[17,29,124]

Homeostatic Renewal

The first formal evidence for a pool of capsular stem and subcapsular progenitor cells came from analyses of the HH signaling pathway[17]

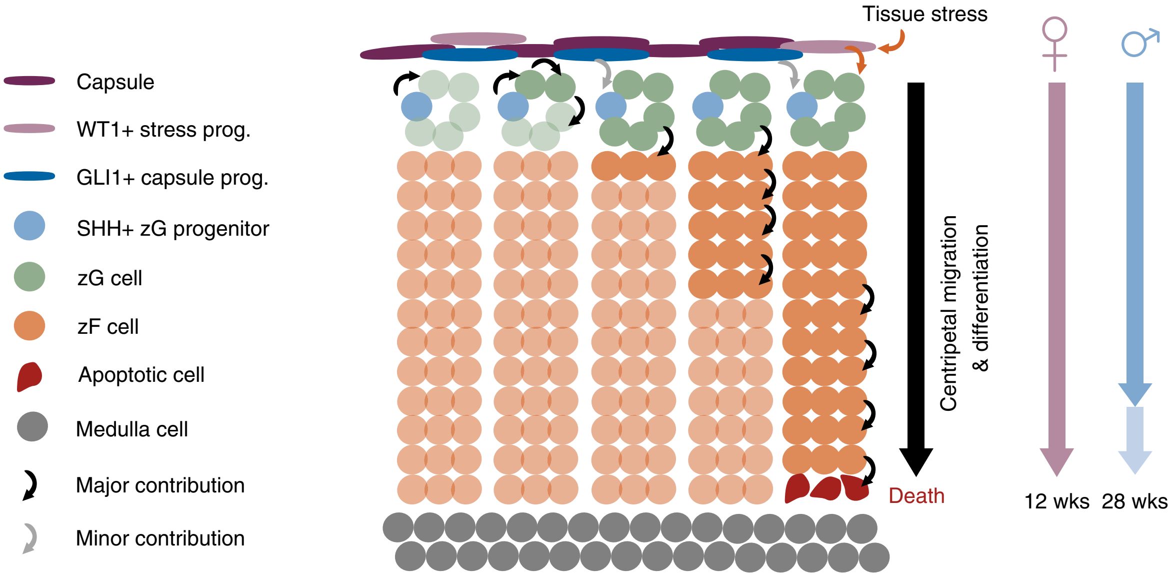


Fig. 85.8 Adrenal cortex renewal. Adrenal cortex homeostasis is maintained by constant renewal. This involves stem/progenitor cells located within the capsule and the zona glomerulosa (zG). Sonic hedgehog–positive (SHH+) zG cells ensure the majority of cortical renewal in the normal postpubertal adrenal cortex. After recruitment, SHH+ cells and their progeny initially differentiate as CYP11B2+ zG cells. They subsequently adopt a CYP11B1+ zF cell phenotype along their centripetal displacement within the cortex. When reaching the corticomedullary boundary, migrating cells enter apoptosis and are eliminated. GLI1+ cells can also contribute, albeit with low efficiency, to adrenal cortex renewal in both males and females before puberty. However, their contribution ceases in postpubertal males and is maintained in females. This sexual dimorphism is also reflected in the speed of renewal: full cortical replacement takes approximately 12 weeks in females and more than 7 months in males. WT1+ capsular cells are only mobilized following tissue stress, such as after gonadectomy. They may be mobilized in aging adrenals following the natural decline in sex steroids following menopause and andropause.

(see section on Sonic Hedgehog Signaling). The only HH ligand in the adrenal cortex is SHH, which is expressed in SF-1+ cells within zG rosettes.[17] Elegant lineage-tracing studies using *Shh-gfpCre* and *Shh-CreERT2* showed that SHH+ cells and their progeny migrate centripetally throughout the cortex. They differentiate as zG (CYP11B2+/SF-1+) cells and zF (CYP11B1+/SF-1+) cells. This showed that SHH+ cells are progenitors for both zG and zF[17] (Fig. 85.8). Lineage tracing also showed that SHH-responsive undifferentiated GLI1+/SF-1- capsular cells could also contribute to all steroidogenic lineages of the cortex and to a subset of SHH+ progenitors. However, their contribution to the different lineages is rather low during homeostasis and is sexually dimorphic[17,237] (see section on Sexual Dimorphism in Renewal). Early experimental evidence relying on trypan blue injections in the capsular and subcapsular space had shown that labeled cells could transit from the zG to the zF.[238] This, together with the observation that GLI1+ and SHH+ cells could contribute to both terminally differentiated lineages, suggested that cortical renewal occurs by sequential differentiation resulting from conversion of capsular cells to zG cells that subsequently adopt a zF phenotype during their centripetal displacement in the cortex. This was formally demonstrated by using the *Cyp11b2-Cre* driver and the mTmG reporter for lineage tracing. Indeed, whereas EGFP+ CYP11B2-derived cells were initially restricted to the zG (CYP11B2+/CYP11B1-), by 5 weeks, descendants of the CYP11B2 lineage were found in the zF (CYP11B1+/CYP11B2-). By 12 weeks in females, the entire zF was recombined, and all the cells in the zF had lost CYP11B2 expression and gained CYP11B1 expression.[124] Therefore, although it is technically challenging to prove that descendants of CYP11B2+ cells are also descendants of SHH+ and/or GLI1+ cells, these data demonstrate that the majority of adrenal cortex renewal occurs through sequential transdifferentiation from zG to zF, during migration within the cortex (Fig. 85.8). However, while this is the major homeostatic pathway, cells of the inner cortex can be derived independent of the CYP11B2 lineage under some extreme conditions.[124]

These findings raise the question of the mechanisms involved in transdifferentiation from zG to zF (Fig. 85.9). Analysis of models of constitutive activation of PKA throughout the cortex showed that this pathway plays an essential role in blocking Wnt pathway activation in adrenal steroidogenic cells, to allow conversion from a zG to zF phenotype. This inhibitory mechanism relies on inhibition of WNT4 expression and destabilization of β-catenin, through direct phosphorylation and indirect activation of GSK3β within the destruction complex.[85] Conversely, Wnt pathway activation was found to inhibit PKA signaling. Indeed, pharmacological activation of Wnt signaling in adrenocortical cells in culture resulted in repression of *MC2R* and *CYP11B1* expression.[239] This was further confirmed in mouse models of constitutive β-catenin activation in which expression of *CYP11B1* and AKR1B7 was inhibited[187,188] and conversion from fasciculata to glomerulosa was blocked, resulting in hyperaldosteronism.[186] This may involve overexpression of the phosphodiesterase PDE2A, a direct target of WNT/β-catenin signaling in the adrenal cortex,[186] but also inhibition of both SF-1 expression and transcriptional activity, presumably through titration and displacement of SF-1 from the promoters of steroidogenic genes, by β-catenin.[239]

Altogether, these findings strongly suggest that, after subcapsular SHH+ progenitor cells are recruited, they initially differentiate as zG cells under the influence of Wnt signaling, activated by WNT4 and further potentiated by RSPO3. This stimulates expression of AT1R and CYP11B2 and blocks expression of MC2R and CYP11B1. As cells migrate within the cortex, they progressively escape the positive

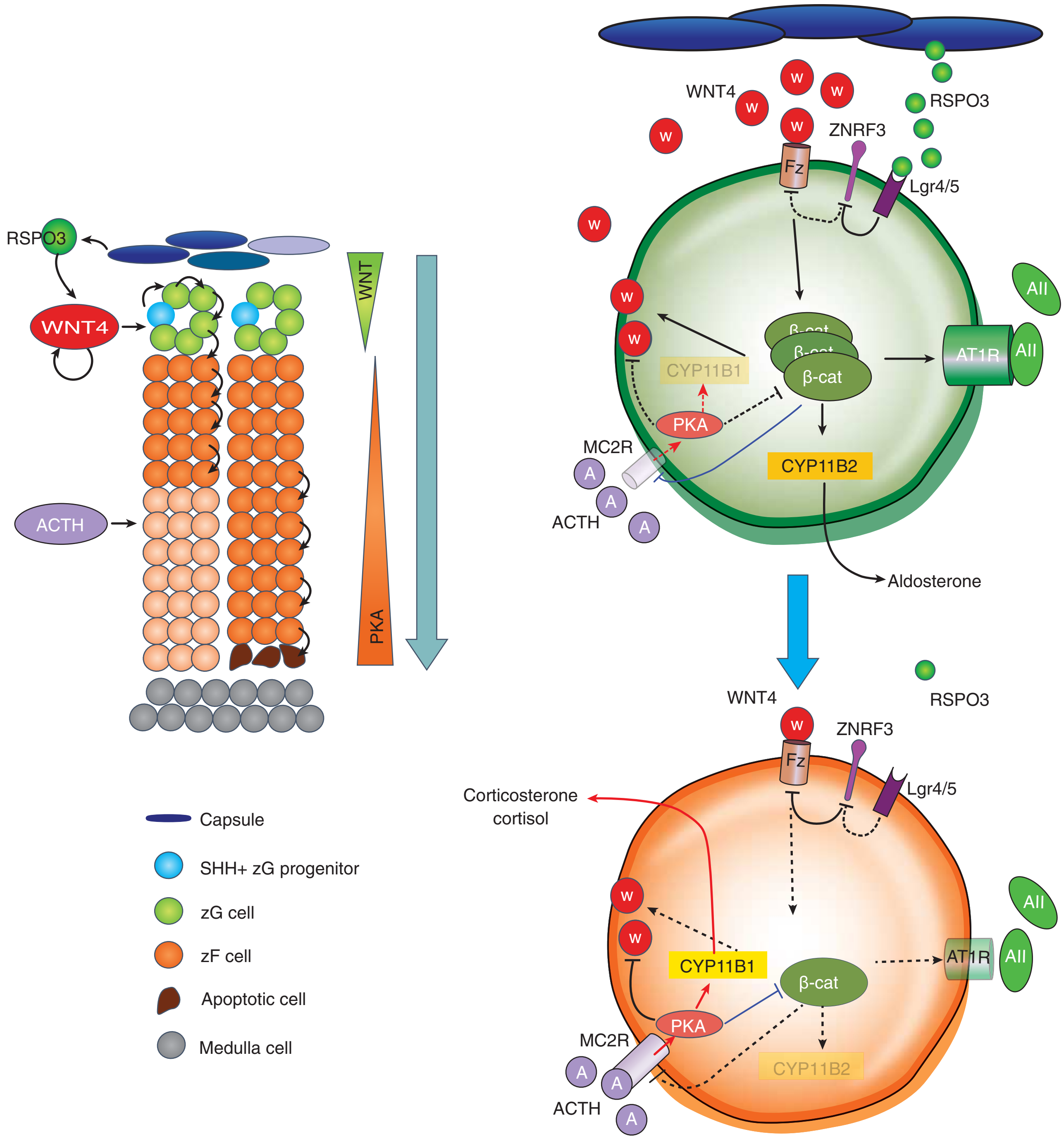

RSPO3
WNT4
ACTH
WNT
PKA
Capsule
SHH+ zG progenitor
zG cell
zF cell
Apoptotic cell
Medulla cell
WNT4
w
RSPO3
ZNRF3
Fz
Lgr4/5
All
AT1R
β-cat
CYP11B1
PKA
CYP11B2
MC2R
A
ACTH
Aldosterone
WNT4
RSPO3
ZNRF3
Fz
Lgr4/5
Corticosterone
cortisol
All
AT1R
CYP11B1
β-cat
PKA
MC2R
CYP11B2
ACTH

Fig. 85.9 Antagonistic interaction between WNT and PKA signaling pathways control zonal differentiation during renewal. After recruitment, sonic hedgehog–positive progenitor cells initially differentiate as zona glomerulosa cells under the influence of Wnt signaling, activated by WNT4 and further potentiated by RSPO3, which inactivates ZNRF3, allowing accumulation of Frizzled (FZD; Wnt receptor). This is reinforced by induction of WNT4 expression by β-catenin, which also stimulates expression of AT1R and CYP11B2 and blocks expression of MC2R and CYP11B1. Aldosterone is then produced under the control of Angiotensin II, acting through AT1R. As cells migrate within the cortex, they progressively escape the positive influence of RSPO3/WNT4. This prevents Wnt pathway activation by allowing ZNRF3-mediated clearance of FZD, resulting in a further decrease in WNT4 expression. This reduction in Wnt activation allows expression of MC2R, which in turn stimulates the PKA signaling pathway. This further inhibits expression of WNT4, prevents β-catenin accumulation, and stimulates expression of CYP11B1, allowing zona fasciculata differentiation and production of glucocorticoids.

influence of RSPO3/WNT4, which is further suppressed by ZNRF3-mediated downregulation of Wnt receptors.[185] This reduction in Wnt activation allows expression of MC2R, which in turn stimulates the PKA signaling pathway. This further inhibits expression of WNT4, prevents β-catenin accumulation, and stimulates expression of CYP11B1, allowing zF differentiation (Fig. 85.9).

Whether these mechanisms are also involved in the human adrenal cortex is unknown. However, consistent with mouse adrenals, proliferation in the human adrenal is also more concentrated in the outer cortex, whereas apoptosis is localized at the corticomedullary boundary.[240] There is also more intense staining for SHH at the periphery of the cortex and GLI1 expression in the capsule.[241,242] Interestingly, analysis of telomere length within the different zones of human adrenal cortices showed that telomeres were shorter in zF than in zG in subjects between 22 and 68 years of age, whereas they were identical in younger individuals.[243] This age- and zone-dependent telomeric attrition suggests that the human adrenal cortex also undergoes constant renewal, and that zF cells are older than zG cells, consistent with the sequential differentiation model. However, zR cells were found to have longer telomeres than either zG or zF.[243] This suggests that the zR may be established and/or maintained independently from the zG and zF.

Role of Renin-Angiotensin-Aldosterone System and Adrenocorticotropic Hormone/PKA Signaling in Regulation of Renewal

The RAAS and ACTH/PKA signaling pathways, which control endocrine activity and proliferation in the adrenal cortex, also play a role in progenitors' homeostasis. Indeed, in the context of genetic ablation of *Cyp11b2*, which results in a 5-fold increase in renin-angiotensin activation, lineage conversion from zG to zF is significantly accelerated, which is correlated with increased proliferation and expansion of the zG. This results in complete zF colonization by the descendants of zG cells after 5 weeks, compared with 12 weeks in wild-type mice.[124,186] This effect can be partially reversed by treatment with candesartan, indicating that cortical cell renewal can be stimulated by increased renin-angiotensin activity.[186] Interestingly, cortical cell renewal can also be accelerated by PKA signaling. Indeed, whereas 2-week treatment with dexamethasone to suppress pituitary ACTH secretion results in inhibition of zG-to-zF conversion,[124] constitutive activation of PKA using *Cyp11b2-Cre* is associated with a marked acceleration of cortical cell renewal.[128] Altogether, these data show that the major adrenal regulators of adrenal endocrine activity are able to stimulate adrenal cortex renewal to allow for sustained response to potentially chronic stimuli. How these hormones can stimulate differentiation, proliferation, and renewal remains unclear.

Sexual Dimorphism in Renewal

The adrenal cortex is a sexually dimorphic tissue. In mice, female adrenals are larger after puberty,[244-246] and sex-specific transcriptomic signatures have been described.[247,248] Interestingly, analysis of proliferation using BrdU injections showed that mature female adrenals had a much higher proliferation rate (6-fold) than male adrenals. This was partially compensated by a 2-fold increase in apoptosis at the corticomedullary junction, suggesting that female adrenals undergo a faster cell turnover than male adrenals.[237] Consistent with this hypothesis, lineage tracing using *Cyp11b2-Cre*, *Axin2-CreERT2,* or *Wnt4-CreERT2* showed full replacement of the adult female cortex by zG-derived cells after 3 months, whereas only 42% and 75% of cortical cells were derived from the zG after 3 and 7 months, respectively, in males (Fig. 85.8).[128,237] Lineage tracing using *Gli1-CreERT2* to identify the descendants of capsular progenitors at 3 weeks showed a similar contribution of these cells to differentiated cortical cells in males in females. However, when similar experiments were conducted after puberty, the contribution of GLI11+ cells to the differentiated cortex almost ceased in males, whereas it was maintained in females.[237] This shows that, whereas both prepubertal male and female adrenals use GLI11+ and SHH+ progenitors for cortical renewal, GLI11+ progenitors are only used in females after puberty. This suggests that male adrenals may be more sensitive to aging than female adrenals. Consistent with this idea, human aging male zF cells have shorter telomeres than age-matched female zF cells.[249]

How are these sexually dimorphic processes regulated? To evaluate a potential contribution of chromosomal sex, Grabek et al. used models of sex reversal consisting of XX males (overexpressing SOX9 under the control of *Wt1* regulatory regions) and XY females (bearing *Sf-1-Cre* mediated inactivation of SOX9). These experiments showed that XY females had increased capsular proliferation compared with control males, whereas XX males had decreased proliferation compared with control females. In contrast to control XX females, GLI1+ cells did not contribute to the pool of steroidogenic cells, which was reminiscent of XY males.[237] This suggested that sex-specific proliferation and recruitment were not dependent on sex chromosomes, but were rather controlled by sex hormones. Indeed, gonadectomy in male mice induced a significant increase in capsular proliferation and a 3-fold induction in recruitment of capsular cells to the steroidogenic lineage. Conversely, administration of the nonaromatizable androgen DHT to female mice inhibited proliferation and almost completely blocked GLI1+ progenitor recruitment.[237] Similar experiments using *Cyp11b2-Cre* to identify descendants of zG cells also showed that androgen deprivation in male mice could dramatically accelerate renewal, which was reversed by supplementation with DHT.[128] Altogether, these experiments show an all-encompassing role of androgens in inhibition of capsular progenitor cell recruitment and overall cortical cell renewal in male mice. Although the molecular mechanisms underlying this effect of androgens are not completely understood, transcriptomic analyses of castrated and supplemented mice suggest that they may involve stimulation of Wnt/β-catenin signaling, which would in turn counteract the prorenewal effects of PKA signaling.[128] These sexually dimorphic proliferation and renewal capacities may account for the female prevalence of adrenal diseases such as Cushing syndrome[250] and adrenocortical carcinomas.[251]

Adrenal Cortex Regeneration

Chronic administration of glucocorticoids is used for treatment of a number of medical conditions. By suppressing pituitary ACTH, these treatments result in adrenal cortex atrophy, which can be reversed by a regenerative process, following withdrawal of corticoids.[252,253] Recent experimental evidence in mice has shown that this process involves mobilization of adrenal progenitors in response to both HH and Wnt signaling pathways. Indeed, following chronic dexamethasone treatment removal, *SHH* expression is increased, and SHH+ and GLI1+ cells are mobilized to reconstitute a functional cortex.[204] This process can be efficiently blocked using GANT61, a specific inhibitor of GLI1 and GLI2 DNA-binding activity. Conversely, constitutive activation of HH signaling within GLI1+ cells using a SmoM2 mutant facilitates regeneration.[204] Wnt signaling is also actively engaged during adrenal cortex regeneration and inactivation of β-catenin in AXIN2+ Wnt-responsive cells inhibits the regeneration process.[204] This may involve direct stimulation of *SHH* expression by canonical Wnt signaling to trigger progenitor cell recruitment.[184,191,204] Failure of these processes may result in persistent adrenal insufficiency despite glucocorticoid withdrawal.[252,253]

Tissue Stress Progenitors and Adrenal Aging

Although SHH+ cells are the key progenitors for adrenal homeostasis, lineage-tracing studies have identified some other populations of cells that can be mobilized to respond to tissue stress. In the adult, lineage tracing with *Wt1-CreERT2* showed that WT1+ cells can give rise to fully differentiated steroidogenic cells, albeit at very low frequency (Fig. 85.8).[29] Interestingly, though, gonadectomy increases recruitment and differentiation of WT1-derived cells into the steroidogenic lineage, in both males and females. However, these cells are not fully differentiated and express features of gonadal steroidogenic cells, such as CYP17 and LHR, that are not found in normal adrenals.[29] Gonadectomy also induces subcapsular accumulation of GLI1-derived aberrantly differentiated steroidogenic cells that express the gonadal markers PDGFRα, FOXL2, and GATA4.[254] This gonadectomy-induced phenomenon is further amplified in the context of loss of adrenal inhibin (*Inha*-/-) and results in the development of adrenocortical tumors characterized by ovarian identity.[255] Further work has demonstrated that adrenal inhibin normally acts in an LH-dependent manner to oppose GATA4-mediated gonadal differentiation by antagonizing TGFβ2.[256] Interestingly, GLI1-derived subcapsular cells also accumulate in intact aging mice of the B6D2F2 strain,[254] which is reminiscent of the GATA4-positive subcapsular hyperplasia observed following gonadectomy or in intact aging mice.[67,257] On the basis of these observations, it is tempting to speculate that these GLI1/WT1-derived GATA4+ cells could accumulate in the adrenal, following the natural decline of sex steroids after menopause and andropause. Whether GLI1+, WT1+, and/or GATA4+ cells could contribute to suboptimal adrenal cortex renewal following late exhaustion of SHH+ progenitors remains to be investigated.

KEY POINTS

- Steroidogenic cells of the adrenal cortex are continuously renewed by the sexually dimorphic recruitment of stem and progenitor cells located in the capsule and within the zona glomerulosa (zG).
- Both GLI1+ capsular and sonic hedgehog–positive (SHH+) zG cells can contribute to cortex renewal in males and females before puberty, but GLI1+ contribution ceases in postpubertal males.
- SHH+ progenitors are the major source of steroidogenic cells during homeostasis in the adult.
- The vast majority of progenitor cells undergo sequential differentiation as zG cells and subsequent transdifferentiation into zone fasciculata cells during their centripetal migration within the cortex.
- Recruitment and sequential differentiation are tightly controlled by antagonistic interactions between Wnt/β-catenin and PKA signaling pathways.
- Progenitors are mobilized to allow adrenal regeneration following withdrawal from chronic glucocorticoid treatment.
- Some distinct Wilm's tumor suppressor-positive progenitors can be mobilized in response to abnormal tissue stress and may participate in renewal of the aging adrenal cortex.

CONCLUSION AND FUTURE DIRECTIONS

Since the first description of the adrenal gland by Bartholomeus Eustachius in 1563, significant progress has been made in understanding its role as a central regulator of the stress response, through secretion of corticoids and catecholamines. Over the last 25 years, a combination of human and mouse genetics has identified the key molecular mechanisms controlling the development, differentiation, and renewal of the adrenal cortex. This has translated into a better understanding of adrenal diseases ranging from insufficiency to aggressive tumor development. However, these novel findings have raised many questions that will need to be addressed to further our understanding of adrenal pathophysiology. Indeed, we still do not know how deregulation of adrenal cortex renewal, in particular in the aging gland, may play a role in development of adrenal diseases. In this respect, the appearance of abnormally differentiated subcapsular cell clusters such as aldosterone-producing cell clusters and DLK1+ clusters in the aging human adrenal is particularly intriguing.[258-261] Understanding the molecular underpinnings of the sexual dimorphism of adrenal cortex function and homeostasis will also undoubtedly shed light on the dimorphic nature of many adrenal conditions. Another key aspect that has been underestimated so far, is the multiple interactions between steroidogenic cells and their tissue microenvironment. The key role of such interactions has recently been illustrated by the demonstration of an important role of the neuropeptide substance P in controlling aldosterone secretion in human adrenals.[262] Answering these questions will help development of novel therapeutic approaches, such as regenerative therapies to treat adrenal insufficiency or targeted therapies to treat adrenocortical cancer patients.

REFERENCES

1. Batholemeus Eustachius. Tabular Anatomica. 1563.
2. Lenard A. The history of research on the adrenals; 1563-1900. *J Hist Med Allied Sci*. 1951;6:496–505.
3. Nagel N. *Ueber die Structur der Nevernieren*; 1836.
4. Leoutsakos B, Leoutsakos A. The adrenal glands: a brief historical perspective. *Hormones (Basel)*. 2008;7:334–336.
5. Kölliker A. *Handbuch der Braunschweig Gewebelehre der Menschen*; 1852.
6. Addison T. *On the Constitutional and Local Effects of Disease of the Suprarenal Capsules*; 1855.
7. Saverino S, Falorni A. Autoimmune Addison's disease. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101379.
8. Brown-Sequard E. Recherches experimentales sur la physiologie et la pathologie des capsules surrénales. *Arch General De Med*. 1856;8:572–598.
9. Abel PJ. Takamine, and the isolation of epinephrine. *J Allergy Clin Immunol*. 2010;125:514–517.
10. Raju TN. The Nobel chronicles. 1950: Edward Calvin Kendall (1886-1972); Philip Showalter Hench (1896-1965); and Tadeus Reichstein (1897-1996). *Lancet (London, England)*. 1999;353:1370.
11. Keene MF. Observations on the development of the human suprarenal gland. *J Anat*. 1927;61:302–324.
12. Arnold J. Ein Beitrag zu der feineren Structur und dem Chemismus der Nebennieren. *Virchows Archiv Pathologie Anatomie Physiol*. 1866;39:64–117.
13. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32:81–151.
14. Seccia TM, Caroccia B, Gomez-Sanchez EP, et al. The biology of normal zona glomerulosa and aldosterone-producing adenoma: pathological implications. *Endocr Rev*. 2018;39:1029–1056.
15. Leng S, et al. β-Catenin and FGFR2 regulate postnatal rosette-based adrenocortical morphogenesis. *Nat Commun*. 2020;11:1680.
16. Guagliardo NA, et al. Angiotensin II induces coordinated calcium bursts in aldosterone-producing adrenal rosettes. *Nat Commun*. 2020 Apr 3; 11(1):1679.
17. King P, Paul A, Laufer E. Shh signaling regulates adrenocortical development and identifies progenitors of steroidogenic lineages. *Proceed Nat Acad Sci U S A*. 2009;106:21185–21190.
18. Miller WL. The hypothalamic-pituitary-adrenal axis: a brief history. *Hormone Res Paediatr*. 2018;89:212–223.
19. Angelousi A, Margioris AN, Tsatsanis C. ACTH action on the adrenals. In: Feingold KR, et al., ed. *Endotext*. 2000; MDText.com, Inc.
20. Keeney DS, Jenkins CM, Waterman MR. Developmentally regulated expression of adrenal 17 alpha-hydroxylase cytochrome P450 in the mouse embryo. *Endocrinology*. 1995;136:4872–4879.
21. Turcu A, Smith JM, Auchus R, et al. Adrenal androgens and androgen precursors-definition, synthesis, regulation and physiologic actions. *Compr Physiol*. 2014;4:1369–1381.

22. Nakamura Y, Gang HX, Suzuki T, et al. Adrenal changes associated with adrenarche. *Rev Endocr Metab Disord*. 2009;10:19–26.
23. Howard-Miller E. A transitory zone in the adrenal cortex which shows age and sex relationships. *Am J Anat*. 1927;40:251–293.
24. Zubair M, Parker KL, Morohashi K. Developmental links between the fetal and adult zones of the adrenal cortex revealed by lineage tracing. *Mol Cell Biol*. 2008;28:7030–7040.
25. Hershkovitz L, Beuschlein F, Klammer S, et al. Adrenal 20alpha-hydroxysteroid dehydrogenase in the mouse catabolizes progesterone and 11-deoxycorticosterone and is restricted to the X-zone. *Endocrinology*. 2007;148:976–988.
26. Pihlajoki M, et al. Conditional mutagenesis of Gata6 in SF1-positive cells causes gonadal-like differentiation in the adrenal cortex of mice. *Endocrinology*. 2013;154:1754–1767.
27. Mitani F, et al. A novel cell layer without corticosteroid-synthesizing enzymes in rat adrenal cortex: histochemical detection and possible physiological role. *Endocrinology*. 1994;135:431–438.
28. Guasti L, Paul A, Laufer E, et al. Localization of Sonic hedgehog secreting and receiving cells in the developing and adult rat adrenal cortex. *Mol Cell Endocrinol*. 2011;336:117–122.
29. Bandiera R, et al. WT1 maintains adrenal-gonadal primordium identity and marks a population of AGP-like progenitors within the adrenal gland. *Dev Cell*. 2013;27:5–18.
30. Hammer GD, Basham KJ. Stem cell function and plasticity in the normal physiology of the adrenal cortex. *Mol Cell Endocrinol*. 2020;519:111043.
31. Kastriti ME, Kameneva P, Adameyko I. Stem cells, evolutionary aspects and pathology of the adrenal medulla: a new developmental paradigm. *Mol Cell Endocrinol*. 2020;518:110998.
32. Megha R, Wehrle CJ, Kashyap S, et al. Anatomy, abdomen and pelvis, adrenal glands (suprarenal glands). In: *StatPearls*. StatPearls Publishing LLC, Treasure Island, Flordia; 2020.
33. Parker TL, Kesse WK, Mohamed AA, et al. The innervation of the mammalian adrenal gland. *J Anat*. 1993;183:265–276.
34. Hatano O, Takakusu A, Nomura M, et al. Identical origin of adrenal cortex and gonad revealed by expression profiles of Ad4BP/SF-1. *Gene Cell*. 1996;1:663–671.
35. Hanley NA, et al. Expression of steroidogenic factor 1 and Wilms' tumour 1 during early human gonadal development and sex determination. *Mech Dev*. 1999;87:175–180.
36. Hanley NA, Rainey WE, Wilson DI, et al. Expression profiles of SF-1, DAX1, and CYP17 in the human fetal adrenal gland: potential interactions in gene regulation. *Mol Endocrinol*. 2001;15:57–68.
37. Ikeda Y, Shen WH, Ingraham HA, et al. Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Mol Endocrinol*. 1994;8:654–662.
38. Mesiano S, Jaffe RB. Developmental and functional biology of the primate fetal adrenal cortex. *Endocr Rev*. 1997;18:378–403.
39. Anderson DJ, Axel R. A bipotential neuroendocrine precursor whose choice of cell fate is determined by NGF and glucocorticoids. *Cell*. 1986;47:1079–1090.
40. Uotila UU. The early embryological development of the fetal and permanent adrenal cortex in man. *Am J Anat*. 1940;6:259–390.
41. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*. 1994;77:481–490.
42. Sadovsky Y, et al. Mice deficient in the orphan receptor steroidogenic factor 1 lack adrenal glands and gonads but express P450 side-chain-cleavage enzyme in the placenta and have normal embryonic serum levels of corticosteroids. *Proceed Nat Acad Sci USA*. 1995;92:10939–10943.
43. Bland ML, et al. Haploinsufficiency of steroidogenic factor-1 in mice disrupts adrenal development leading to an impaired stress response. *Proceed Nat Acad Sci USA*. 2000;97:14488–14493.
44. Beuschlein F, et al. Steroidogenic factor-1 is essential for compensatory adrenal growth following unilateral adrenalectomy. *Endocrinology*. 2002;143:3122–3135.
45. Ozisik G, Achermann JC, Jameson JL. The role of SF1 in adrenal and reproductive function: insight from naturally occurring mutations in humans. *Mol Genet Metabol*. 2002;76:85–91.
46. El-Khairi R, Achermann JC. Steroidogenic factor-1 and human disease. *Semin Reprod Med*. 2012;30:374–381.
47. Suntharalingham JP, Buonocore F, Duncan AJ, et al. DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract Res Clin Endocrinol Metab*. 2015;29:607–619.
48. Buonocore F, Achermann JC. Primary adrenal insufficiency: new genetic causes and their long-term consequences. *Clin Endocrinol*. 2020;92:11–20.
49. Wilhelm D, Englert C. The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes Dev*. 2002;16:1839–1851.
50. Zubair M, Ishihara S, Oka S, et al. Two-step regulation of Ad4BP/SF-1 gene transcription during fetal adrenal development: initiation by a Hox-Pbx1-Prep1 complex and maintenance via autoregulation by Ad4BP/SF-1. *Mol Cell Biol*. 2006;26:4111–4121.
51. Shima Y, et al. Identification of an enhancer in the Ad4BP/SF-1 gene specific for fetal Leydig cells. *Endocrinology*. 2012;153:417–425.
52. Morohashi K, Zubair M. The fetal and adult adrenal cortex. *Mol Cell Endocrinol*. 2011;336:193–197.
53. Moore AW, et al. YAC transgenic analysis reveals Wilms' tumour 1 gene activity in the proliferating coelomic epithelium, developing diaphragm and limb. *Mech Dev*. 1998;79:169–184.
54. Kreidberg JA, et al. WT-1 is required for early kidney development. *Cell*. 1993;74:679–691.
55. Moore AW, McInnes L, Kreidberg J, et al. YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development (Cambridge, England)*. 1999;126:1845–1857.
56. Val P, Martinez-Barbera JP, Swain A. Adrenal development is initiated by Cited2 and Wt1 through modulation of Sf-1 dosage. *Development (Cambridge, England)*. 2007;134:2349–2358.
57. Val P, Swain A. Gene dosage effects and transcriptional regulation of early mammalian adrenal cortex development. *Mol Cell Endocrinol*. 2010;323:105–114.
58. Ikeda Y, et al. Steroidogenic factor 1 and Dax-1 colocalize in multiple cell lineages: potential links in endocrine development. *Mol Endocrinol*. 1996;10:1261–1272.
59. Swain A, Zanaria E, Hacker A, et al. Mouse Dax1 expression is consistent with a role in sex determination as well as in adrenal and hypothalamus function. *Nat Genet*. 1996;12:404–409.
60. Lalli E, Sassone-Corsi P. DAX-1, an unusual orphan receptor at the crossroads of steroidogenic function and sexual differentiation. *Mol Endocrinol*. 2003;17:1445–1453.
61. Muscatelli F, et al. Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature*. 1994;372:672–676.
62. Zanaria E, et al. An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature*. 1994;372:635–641.
63. Yu RN, Ito M, Saunders TL, et al. Role of Ahch in gonadal development and gametogenesis. *Nat Genet*. 1998;20:353–357.
64. Babu PS, et al. Interaction between Dax-1 and steroidogenic factor-1 in vivo: increased adrenal responsiveness to ACTH in the absence of Dax-1. *Endocrinology*. 2002;143:665–673.
65. Scheys JO, Heaton JH, Hammer GD. Evidence of adrenal failure in aging Dax1-deficient mice. *Endocrinology*. 2011;152:3430–3439.
66. Kiiveri S, et al. Transcription factors GATA-4 and GATA-6 during mouse and human adrenocortical development. *Endocr Res*. 2002;28:647–650.
67. Mathieu M, et al. Steroidogenic differentiation and PKA signaling are programmed by histone methyltransferase EZH2 in the adrenal cortex. *Proc Natl Acad Sci U S A*. 2018;115:E12265–E12274.
68. Tevosian SG, et al. Adrenal development in mice requires GATA4 and GATA6 transcription factors. *Endocrinology*. 2015;156:2503–2517.
69. Bielinska M, et al. Gonadectomy-induced adrenocortical neoplasia in the domestic ferret (Mustela putorius furo) and laboratory mouse. *Vet Pathol*. 2006;43:97–117.
70. Looyenga BD, Hammer GD. Origin and identity of adrenocortical tumors in inhibin knockout mice: implications for cellular plasticity in the adrenal cortex. *Mol Endocrinol*. 2006;20:2848–2863.

71. Pihlajoki M, Färkkilä A, Soini T, et al. GATA factors in endocrine neoplasia. *Mol Cell Endocrinol*. 2016;421:2–17.
72. Nusse R, Clevers H. Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. *Cell*. 2017;169:985–999.
73. Jho EH, et al. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol Cell Biol*. 2002;22:1172–1183.
74. Lustig B, et al. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol Cell Biol*. 2002;22:1184–1193.
75. Ramakrishnan AB, Cadigan KM. Wnt target genes and where to find them. *F1000Res*. 2017;6:746.
76. Hao HX, et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature*. 2012;485:195–200.
77. Koo BK, et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature*. 2012;488:665–669.
78. de Lau WB, Snel B, Clevers HC. The R-spondin protein family. *Genome Biol*. 2012;13:242.
79. Lebensohn AM. Rohatgi RR-spondins can potentiate WNT signaling without LGRs. *Elife*. 2018;7:e33126. http://dx.doi.10.7554/eLife.33126.
80. Szenker-Ravi E, et al. RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6. *Nature*. 2018;557:564–569.
81. Cruciat CM, Niehrs C. Secreted and transmembrane wnt inhibitors and activators. *Cold Spring Harb Perspect Biol*. 2013;5:a015081.
82. Kim AC, et al. Targeted disruption of beta-catenin in Sf1-expressing cells impairs development and maintenance of the adrenal cortex. *Development (Cambridge, England)*. 2008;135:2593–2602.
83. Bingham NC, Verma-Kurvari S, ParadaLF, et al. Development of a steroidogenic factor 1/Cre transgenic mouse line. *Genesis*. 2006;44:419–424.
84. Heaton JH, et al. Progression to adrenocortical tumorigenesis in mice and humans through insulin-like growth factor 2 and beta-catenin. *Am J Pathol*. 2012;181:1017–1033.
85. Drelon C, et al. PKA inhibits WNT signalling in adrenal cortex zonation and prevents malignant tumour development. *Nat Commun*. 2016;7:12751.
86. Wu F, Zhang Y, Sun B, et al. Hedgehog signaling: from basic biology to cancer therapy. *Cell Chem Biol*. 2017;24:252–280.
87. Begleiter ML, Harris DJ. Holoprosencephaly and endocrine dysgenesis in brothers. *Am J Med Genet*. 1980;7:315–318.
88. Kang S, Graham JM, Olney AH, Biesecker LG. GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet*. 1997;15:266–268.
89. Dubourg C, et al. Holoprosencephaly. *Orphanet J Rare Dis*. 2007;2:8.
90. King PJ, Guasti L& Laufer E. Hedgehog signalling in endocrine development and disease. *J Endocrinol*. 2008;198:439–450.
91. Bitgood MJ, McMahon AP. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev Biol*. 1995;172:126–138.
92. Ching S, Vilain E. Targeted disruption of Sonic Hedgehog in the mouse adrenal leads to adrenocortical hypoplasia. *Genesis*. 2009;47:628–637.
93. Huang CC, Miyagawa S, Matsumaru D, et al. Progenitor cell expansion and organ size of mouse adrenal is regulated by sonic hedgehog. *Endocrinology*. 2010;151:1119–1128.
94. Finco I, LaPensee CR, Krill KT, et al. Hedgehog signaling and steroidogenesis. *Annu Rev Physiol*. 2015;77:105–129.
95. Brewer JR, Mazot P, Soriano P. Genetic insights into the mechanisms of Fgf signaling. *Genes Dev*. 2016;30:751–771.
96. Stark KL, McMahon JA, McMahon AP. FGFR-4, a new member of the fibroblast growth factor receptor family, expressed in the definitive endoderm and skeletal muscle lineages of the mouse. *Development (Cambridge, England)*. 1991;113:641–651.
97. Guasti L, Candy Sze WC, McKay T, et al. FGF signalling through Fgfr2 isoform IIIb regulates adrenal cortex development. *Mol Cell Endocrinol*. 2013;371:182–188.
98. Revest JM, et al. Fibroblast growth factor receptor 2-IIIb acts upstream of Shh and Fgf4 and is required for limb bud maintenance but not for the induction of Fgf8, Fgf10, Msx1, or Bmp4. *Dev Biol*. 2001;231:47–62.
99. Kim Y, et al. Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104:16558–16563.
100. Hafner R, Bohnenpoll T, Rudat C, et al. Fgfr2 is required for the expansion of the early adrenocortical primordium. *Mol Cell Endocrinol*. 2015;413:168–177.
101. Ishimoto H, Jaffe RB. Development and function of the human fetal adrenal cortex: a key component in the feto-placental unit. *Endocr Rev*. 2011;32:317–355.
102. Winters AJ, Oliver C, Colston C, et al. Plasma ACTH levels in the human fetus and neonate as related to age and parturition. *J Clin Endocrinol Metab*. 1974;39:269–273.
103. Mesiano S, et al. Localization and regulation of corticotropin receptor expression in the midgestation human fetal adrenal cortex: implications for in utero homeostasis. *J Clin Endocrinol Metab*. 1996;81:340–345.
104. Rehman KS, Sirianni R, Parker CR, et al. The regulation of adrenocorticotrophic hormone receptor by corticotropin-releasing hormone in human fetal adrenal definitive/transitional zone cells. *Reprod Sci*. 2007;14:578–587.
105. Bocian-Sobkowska J, Woźniak W, Malendowicz LK. Postnatal involution of the human adrenal fetal zone: stereologic description and apoptosis. *Endocr Res*. 1998;24:969–973.
106. Spencer SJ, Mesiano S, Lee JY, et al. Proliferation and apoptosis in the human adrenal cortex during the fetal and perinatal periods: implications for growth and remodeling. *J Clin Endocrinol Metab*. 1999;84:1110–1115.
107. Parker CR, Leveno K, Carr BR, et al. Umbilical cord plasma levels of dehydroepiandrosterone sulfate during human gestation. *J Clin Endocrinol Metab*. 1982;54:1216–1220.
108. Grueters A, Korth-Schutz S. Longitudinal study of plasma dehydroepiandrosterone sulfate in preterm and fullterm infants. *J Clin Endocrinol Metab*. 1982;55:314–320.
109. Ben-David S, et al. Parturition itself is the basis for fetal adrenal involution. *J Clin Endocrinol Metab*. 2007;92:93–97.
110. Midgley PC, et al. Adrenal function in preterm infants: ACTH may not be the sole regulator of the fetal zone. *Pediatr Res*. 1998;44:887–893.
111. Kahri AI, Huhtaniemi I, Salmenperä M. Steroid formation and differentiation of cortical cells in tissue culture of human fetal adrenals in the presence and absence of ACTH. *Endocrinology*. 1976;98:33–41.
112. Gray ES, Abramovich DR. Morphologic features of the anencephlaic adrenal gland in early pregnancy. *Am J Obstet Gynecol*. 1980;137:491–495.
113. Simonian MH, Gill GN. Regulation of the fetal human adrenal cortex: effects of adrenocorticotropin on growth and function of monolayer cultures of fetal and definitive zone cells. *Endocrinology*. 1981;108:1769–1779.
114. Fujieda K, Faiman C, Feyes FI, et al. The control of steroidogenesis by human fetal adrenal cells in tissue culture. IV. The effect of exposure to placental steroids. *J Clin Endocrinol Metab*. 1982;54:89–94.
115. Winter JS, Smail PJ. Effects of ACTH and estradiol on steroid production by cultured adrenal cells from an anencephalic fetus and from normal adults. *Steroids*. 1983;42:677–685.
116. Hofland J, de Jong FH. Inhibins and activins: their roles in the adrenal gland and the development of adrenocortical tumors. *Mol Cell Endocrinol*. 2012;359:92–100.
117. Voutilainen R, Erämaa M, Ritvos O. Hormonally regulated inhibin gene expression in human fetal and adult adrenals. *J Clin Endocrinol Metab*. 1991;73:1026–1030.
118. Spencer SJ, Rabinovici J, Mesiano S, et al. Activin and inhibin in the human adrenal gland. Regulation and differential effects in fetal and adult cells. *J Clin Invest*. 1992;90:142–149.
119. Munro LM, Kennedy A, McNicol AM. The expression of inhibin/activin subunits in the human adrenal cortex and its tumours. *J Endocrinol*. 1999;161:341–347.
120. Vänttinen T, Kuulasmaa T, Liu J, et al. Expression of activin/inhibin receptor and binding protein genes and regulation of activin/inhibin peptide secretion in human adrenocortical cells. *J Clin Endocrinol Metab*. 2002;87:4257–4263.

121. Vänttinen T, Liu J, Kuulasmaa T, et al. Expression of activin/inhibin signaling components in the human adrenal gland and the effects of activins and inhibins on adrenocortical steroidogenesis and apoptosis. *J Endocrinol*. 2003;178:479–489.
122. Riopel L, Branchaud CL, Goodyer CG, et al. Growth-inhibitory effect of TGF-B on human fetal adrenal cells in primary monolayer culture. *J Cell Physiol*. 1989;140:233–238.
123. Florio P, et al. Inhibin A, inhibin B and activin A concentrations in umbilical cord artery and vein. *Gynecol Endocrinol*. 2003;17:181–185.
124. Freedman BD, et al. Adrenocortical zonation results from lineage conversion of differentiated zona glomerulosa cells. *Dev Cell*. 2013;26:666–673.
125. Ungar F, Stabler TA. 20 alpha-Hydroxysteroid dehydrogenase activity and the X-zone of the female mouse adrenal. *J Steroid Biochem*. 1980;13:23–28.
126. Huang CCJ, Kraft C, Moy N, et al. A novel population of inner cortical cells in the adrenal gland that displays sexually dimorphic expression of thyroid hormone receptor-β1. *Endocrinology*. 2015;156:2338–2348.
127. Holmes PV, Dickson AD. X-zone degeneration in the adrenal glands of adult and immature female mice. *J Anat*. 1971;108:159–168.
128. Dumontet T, et al. PKA signaling drives reticularis differentiation and sexually dimorphic adrenal cortex renewal. *JCI Insight*. 2018;3:e98394.
129. Gannon AL, et al. Androgen receptor signalling in the male adrenal facilitates X-zone regression, cell turnover and protects against adrenal degeneration during ageing. *Sci Rep*. 2019;9:10457.
130. Matsuura S, Suzuki K. Morphological changes in the submandibular glands and in the X zone of the adrenal gland following ovariectomy in mice. *Cell Tissue Res*. 1986;246:549–556.
131. Asari M, Fukaya K, Eguchi Y, et al. [Effect of testosterone and progesterone on the adrenal X-zone in female mice (author's transl)]. *Nihon Juigaku Zasshi*. 1979;41:61–67.
132. Vinson GP, Jones IC. The conversion of progesterone into testosterone by the mouse ovary and its relationship to adrenal X zone degeneration. *J Endocrinol*. 1963;26:407–414.
133. Barkley MS, Geschwind II , Bradford GE. The gestational pattern of estradiol, testosterone and progesterone secretion in selected strains of mice. *Biol Reprod*. 1979;20:733–738.
134. Preston MI. Effects of thyroxin injections on the suprarenal glands of the mouse. *Endocrinology*. 1928;12:323–334.
135. Gersh I, Grollman A. The nature of the X-zone of the adrenal gland of the mouse. *Anat Rec*. 1939;75:131–153.
136. Shire JG, Beamer WG. Adrenal changes in genetically hypothyroid mice. *J Endocrinol*. 1984;102:277–280.
137. Beuschlein F, et al. Activin induces x-zone apoptosis that inhibits luteinizing hormone-dependent adrenocortical tumor formation in inhibin-deficient mice. *Mol Cell Biol*. 2003;23:3951–3964.
138. Matzuk MM, et al. Development of cancer cachexia-like syndrome and adrenal tumors in inhibin-deficient mice. *Proc Natl Acad Sci USA*. 1994;91:8817–8821.
139. Looyenga BD, Hammer GD. Genetic removal of Smad3 from inhibin-null mice attenuates tumor progression by uncoupling extracellular mitogenic signals from the cell cycle machinery. *Mol Endocrinol*. 2007;21:2440–2457.
140. Matzuk MM, et al. Functional analysis of activins during mammalian development. *Nature*. 1995;374:354–356.
141. Else T, et al. Tpp1/Acd maintains genomic stability through a complex role in telomere protection. *Chromosome Res*. 2007;15:1001–1013.
142. Beamer WG, et al. Adrenocortical dysplasia: a mouse model system for adrenocortical insufficiency. *J Endocrinol*. 1994;141:33–43.
143. Keegan CE, et al. Urogenital and caudal dysgenesis in adrenocortical dysplasia (acd) mice is caused by a splicing mutation in a novel telomeric regulator. *Hum Mol Genet*. 2005;14:113–123.
144. Else T, et al. Genetic p53 deficiency partially rescues the adrenocortical dysplasia phenotype at the expense of increased tumorigenesis. *Cancer Cell*. 2009;15:465–476.
145. Prasad R, et al. Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest*. 2017;127:942–953.
146. Lovric S, et al. Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest*. 2017;127:912–928.
147. Janecke AR, et al. Deficiency of the sphingosine-1-phosphate lyase SGPL1 is associated with congenital nephrotic syndrome and congenital adrenal calcifications. *Hum Mutat*. 2017;38:365–372.
148. Lucki NC, Li D, Sewer MB. Sphingosine-1-phosphate rapidly increases cortisol biosynthesis and the expression of genes involved in cholesterol uptake and transport in H295R adrenocortical cells. *Mol Cell Endocrinol*. 2012;348:165–175.
149. Urs AN, Dammer E, Sewer MB. Sphingosine regulates the transcription of CYP17 by binding to steroidogenic factor-1. *Endocrinology*. 2006;147:5249–5258.
150. Urs AN, et al. Steroidogenic factor-1 is a sphingolipid binding protein. *Mol Cell Endocrinol*. 2007;265–266:174–178.
151. Val P, Lefrancois Martinez AM, Veyssiere G, et al. SF-1 a key player in the development and differentiation of steroidogenic tissues. *Nucl Recept*. 2003;1:8.
152. Hammer GD, et al. Phosphorylation of the nuclear receptor SF-1 modulates cofactor recruitment: integration of hormone signaling in reproduction and stress. *Mol Cell*. 1999;3:521–526.
153. Chen WY, Juan LJ, Chung BC. SF-1 (nuclear receptor 5A1) activity is activated by cyclic AMP via p300-mediated recruitment to active foci, acetylation, and increased DNA binding. *Mol Cell Biol*. 2005;25:10442–10453.
154. Chen WY, Lee WC, Hsu NC, et al. SUMO modification of repression domains modulates function of nuclear receptor 5A1 (steroidogenic factor-1). *J Biol Chem*. 2004;279:38730–38735.
155. Komatsu T, et al. Small ubiquitin-like modifier 1 (SUMO-1) modification of the synergy control motif of Ad4 binding protein/steroidogenic factor 1 (Ad4BP/SF-1) regulates synergistic transcription between Ad4BP/SF-1 and Sox9. *Mol Endocrinol*. 2004;18:2451–2462.
156. Campbell LA, et al. Decreased recognition of SUMO-sensitive target genes following modification of SF-1 (NR5A1). *Mol Cell Biol*. 2008;28:7476–7486.
157. Lee FY, et al. Eliminating SF-1 (NR5A1) sumoylation in vivo results in ectopic hedgehog signaling and disruption of endocrine development. *Dev Cell*. 2011;21:315–327.
158. Xing Y, Morohashi K, Ingraham HA, et al. Timing of adrenal regression controlled by synergistic interaction between Sf1 SUMOylation and Dax1. *Development*. 2017;144:3798–3807.
159. Lalli E, et al. A transcriptional silencing domain in DAX-1 whose mutation causes adrenal hypoplasia congenita. *Mol Endocrinol*. 1997;11:195019–195060.
160. Lalli E, Melner MH, Stocco DM, et al. DAX-1 blocks steroid production at multiple levels. *Endocrinology*. 1998;139:4237–4243.
161. Nachtigal MW, et al. Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell*. 1998;93:445–454.
162. Reutens AT, et al. Clinical and functional effects of mutations in the DAX-1 gene in patients with adrenal hypoplasia congenita. *J Clin Endocrinol Metab*. 1999;84:504–511.
163. Doghman M, et al. Nephroblastoma overexpressed/cysteine-rich protein 61/connective tissue growth factor/nephroblastoma overexpressed gene-3 (NOV/CCN3), a selective adrenocortical cell proapoptotic factor, is down-regulated in childhood adrenocortical tumors. *J Clin Endocrinol Metab*. 2007;92:3253–3260.
164. Lalli E, Doghman M, Latre de Late P, et al. Beyond steroidogenesis: novel target genes for SF-1 discovered by genomics. *Mol Cell Endocrinol*. 2013;371:154–159.
165. Dumontet T, et al. Hormonal and spatial control of SUMOylation in the human and mouse adrenal cortex. *FASEB J*. 2019;33:10218–10230.
166. Wood MA, et al. Fetal adrenal capsular cells serve as progenitor cells for steroidogenic and stromal adrenocortical cell lineages in M. musculus. *Development (Cambridge, England)*. 2013;140:4522–4532.
167. Yates R, et al. Adrenocortical development, maintenance, and disease. *Curr Top Dev Biol*. 2013;106:239–312.
168. Mitani F, Mukai K, Ogawa T, et al. Expression of cytochromes P450aldo and P45011 beta in rat adrenal gland during late gestational and neonatal stages. *Steroids*. 1997;62:57–61.

169. Mitani F, Mukai K, Miyamoto H, et al. Development of functional zonation in the rat adrenal cortex. *Endocrinology*. 1999;140:3342–3353.
170. Wotus C, Levay-Young BK, Rogers LM, et al. Development of adrenal zonation in fetal rats defined by expression of aldosterone synthase and 11b-hydroxylase. *Endocrinology*. 1998;139:4397–4403.
171. Coulter CL, Jaffe RB. Functional maturation of the primate fetal adrenal in vivo: 3. Specific zonal localization and developmental regulation of CYP21A2 (P450c21) and CYP11B1/CYP11B2 (P450c11/aldosterone synthase) lead to integrated concept of zonal and temporal steroid biosynthesis. *Endocrinology*. 1998;139:5144–5150.
172. Johnston ZC, et al. The human fetal adrenal produces cortisol but no detectable aldosterone throughout the second trimester. *BMC Med*. 2018;16:23.
173. Melau C, et al. Characterization of human adrenal steroidogenesis during fetal development. *J Clin Endocrinol Metab*. 2019;104:1802–1812.
174. Naccache A, et al. Temporal and spatial distribution of mast cells and steroidogenic enzymes in the human fetal adrenal. *Mol Cell Endocrinol*. 2016;434:69–80.
175. Mitani F, Miyamoto H, Mukai K, et al. Effects of long term stimulation of ACTH and angiotensin II-secretions on the rat adrenal cortex. *Endocr Res*. 1996;22:421–431.
176. Ulrich-Lai YM, et al. Chronic stress induces adrenal hyperplasia and hypertrophy in a subregion-specific manner. *Am J Physiol Endocrinol Metab*. 2006;291:E965–E973.
177. Laufer E, Kesper D, Vortkamp A, et al. Sonic hedgehog signaling during adrenal development. *Mol Cell Endocrinol*. 2012;351:19–27.
178. Hattangady NG, Olala LO, Bollag WB, et al. Acute and chronic regulation of aldosterone production. *Mol Cell Endocrinol*. 2012;350:151–162.
179. Pignatelli D, et al. Captopril effects on the rat adrenal cortex. *Endocr Res*. 2000;26:965–972.
180. Guasti L, et al. Dlk1 up-regulates Gli1 expression in male rat adrenal capsule cells through the activation of β1 integrin and ERK1/2. *Endocrinology*. 2013;154:4675–4684.
181. Nishimoto K, Harris RBS, Rainey WE, et al. Sodium deficiency regulates rat adrenal zona glomerulosa gene expression. *Endocrinology*. 2014;155:1363–1372.
182. Oliverio MI, et al. Reduced growth, abnormal kidney structure, and type 2 (AT2) angiotensin receptor-mediated blood pressure regulation in mice lacking both AT1A and AT1B receptors for angiotensin II. *Proc Natl Acad Sci U S A*. 1998;95:15496–15501.
183. Taylor MJ, et al. Chemogenetic activation of adrenocortical Gq signaling causes hyperaldosteronism and disrupts functional zonation. *J Clin Invest*. 2020;130:83–93.
184. Vidal V, et al. The adrenal capsule is a signaling center controlling cell renewal and zonation through Rspo3. *Genes Dev*. 2016;30:1389–1394.
185. Basham KJ, et al. A ZNRF3-dependent Wnt/β-catenin signaling gradient is required for adrenal homeostasis. *Genes Dev*. 2019;33:209–220.
186. Pignatti E, et al. Beta-catenin causes adrenal hyperplasia by blocking zonal transdifferentiation. *Cell Rep*. 2020;31:107524.
187. Berthon A, et al. Constitutive {beta}-catenin activation induces adrenal hyperplasia and promotes adrenal cancer development. *Hum Mol Genet*. 2010;19:1561–1576.
188. Berthon A, et al. WNT/β-catenin signalling is activated in aldosterone-producing adenomas and controls aldosterone production. *Hum Mol Genet*. 2014;23:889–905.
189. Åkerström T, et al. Activating mutations in *CTNNB1* in aldosterone producing adenomas. *Sci Rep*. 2016;6:19546.
190. Wu V-C, et al. The prevalence of CTNNB1 mutations in primary aldosteronism and consequences for clinical outcomes. *Sci Rep*. 2017;7:39121.
191. Drelon C, Berthon A, Mathieu M, et al. Adrenal cortex tissue homeostasis and zonation: a WNT perspective. *Mol Cell Endocrinol*. 2015;408:156–164.
192. Heikkila M, et al. Wnt-4 deficiency alters mouse adrenal cortex function, reducing aldosterone production. *Endocrinology*. 2002;143:4358–4365.
193. Mandel H, et al. SERKAL syndrome: an autosomal-recessive disorder caused by a loss-of-function mutation in WNT4. *Am J Hum Genet*. 2008;82:39–47.
194. Chen M, Hornsby PJ. Adenovirus-delivered DKK3/WNT4 and steroidogenesis in primary cultures of adrenocortical cells. *Horm Metab Res*. 2006;38:549–555.
195. de Lau W, Peng WC, Gros P, et al. The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength. *Genes Dev*. 2014;28:305–316.
196. Mountjoy KG, Robbins LS, Mortrud MT, et al. The cloning of a family of genes that encode the melanocortin receptors. *Science*. 1992;257:1248–1251.
197. Metherell LA, et al. Mutations in MRAP, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2. *Nat Genet*. 2005;37:166–170.
198. Gorrigan RJ, Guasti L, King P, et al. Localisation of the melanocortin-2-receptor and its accessory proteins in the developing and adult adrenal gland. *J Mol Endocrinol*. 2011;46:227–232.
199. Wang XL, et al. Transcriptional regulation of human 11beta-hydroxylase (hCYP11B1). *Endocrinology*. 2000;141:3587–3594.
200. Aigueperse C, et al. SF-1 (steroidogenic factor-1), C/EBPbeta (CCAAT/enhancer binding protein), and ubiquitous transcription factors NF1 (nuclear factor 1) and Sp1 (selective promoter factor 1) are required for regulation of the mouse aldose reductase-like gene (AKR1B7) expression in adrenocortical cells. *Mol Endocrinol*. 2001;15:93–111.
201. Nishimoto K, et al. Transcriptome analysis reveals differentially expressed transcripts in rat adrenal zona glomerulosa and zona fasciculata. *Endocrinology*. 2012;153:1755–1763.
202. Dallman MF. Control of adrenocortical growth in vivo. *Endocr Res*. 1984;10:213–242.
203. Thomas M, Keramidas M, Monchaux E, et al. Dual hormonal regulation of endocrine tissue mass and vasculature by adrenocorticotropin in the adrenal cortex. *Endocrinology*. 2004;145:4320–4329.
204. Finco I, Lerario AM, Hammer GD. Sonic hedgehog and WNT signaling promote adrenal gland regeneration in male mice. *Endocrinology*. 2018;159:579–596.
205. Clark AJ, McLoughlin L, Grossman A. Familial glucocorticoid deficiency associated with point mutation in the adrenocorticotropin receptor. *Lancet*. 1993;341:461–462.
206. Novoselova TV, et al. ACTH signalling and adrenal development: lessons from mouse models. *Endocrine Connections*. 2019;8:R122–R130.
207. Yaswen L, Diehl N, Brennan MB, et al. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nat Med*. 1999;5:1066–1070.
208. Coll AP, et al. The effects of proopiomelanocortin deficiency on murine adrenal development and responsiveness to adrenocorticotropin. *Endocrinology*. 2004;145:4721–4727.
209. Chida D, et al. Melanocortin 2 receptor is required for adrenal gland development, steroidogenesis, and neonatal gluconeogenesis. *Proc Natl Acad Sci U S A*. 2007;104:18205–18210.
210. Novoselova TV, et al. MRAP deficiency impairs adrenal progenitor cell differentiation and gland zonation. *Faseb J*. 2018;32:6186–6196.
211. El Ghorayeb N, Bourdeau I, Lacroix A. Role of ACTH and other hormones in the regulation of aldosterone production in primary aldosteronism. *Front Endocrinol*. 2016;7:72.
212. Kirschner LS, et al. Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. *Nat Genet*. 2000;26:89–92.
213. Horvath A, et al. A genome-wide scan identifies mutations in the gene encoding phosphodiesterase 11A4 (PDE11A) in individuals with adrenocortical hyperplasia. *Nat Genet*. 2006;38:794–800.
214. Horvath A, et al. A cAMP-specific phosphodiesterase (PDE8B) that is mutated in adrenal hyperplasia is expressed widely in human and mouse tissues: a novel PDE8B isoform in human adrenal cortex. *Eur J Hum Genet*. 2008;16:1245–1253.
215. Bonnet-Serrano F, Bertherat J. Genetics of tumors of the adrenal cortex. *Endocr Relat Cancer*. 2018;25:R131–R152.
216. Weinstein LS, et al. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*. 1991;325:1688–1695.
217. Beuschlein F, et al. Constitutive activation of PKA catalytic subunit in adrenal Cushing's syndrome. *N Engl J Med*. 2014;370:1019–1028.

218. Sahut-Barnola I, et al. Cushing's syndrome and foetal features resurgence in adrenal cortex-specific Prkar1a knockout mice. *PLoS Genet*. 2010;6:e1000980.
219. Tsai LCL, Shimizu-Albergine M, Beavo JA. The high-affinity cAMP-specific phosphodiesterase 8B controls steroidogenesis in the mouse adrenal gland. *Mol Pharmacol*. 2011;79:639–648.
220. Dhom G. The prepuberal and puberal growth of the adrenal (adrenarche). *Beiträge zur Pathologie*. 1973;150:357–377.
221. Endoh A, Kristiansen SB, Casson PR, et al. The zona reticularis is the site of biosynthesis of dehydroepiandrosterone and dehydroepiandrosterone sulfate in the adult human adrenal cortex resulting from its low expression of 3 beta-hydroxysteroid dehydrogenase. *J Clin Endocrinol Metab*. 1996;81:3558–3565.
222. Abraham GE. Ovarian and adrenal contribution to peripheral androgens during the menstrual cycle. *J Clin Endocrinol Metab*. 1974;39:340–346.
223. Kim MH, Hosseinian AH, Dupon C. Plasma levels of estrogens, androgens and progesterone during normal and dexamethasone-treated cycles. *J Clin Endocrinol Metab*. 1974;39:706–712.
224. Rich BH, Rosenfield RL, Lucky AW, et al. Adrenarche: changing adrenal response to adrenocorticotropin. *J Clin Endocrinol Metab*. 1981;52:1129–1136.
225. Weber A, Clark AJ, Perry LA, et al. Diminished adrenal androgen secretion in familial glucocorticoid deficiency implicates a significant role for ACTH in the induction of adrenarche. *Clin Endocrinol*. 1997;46:431–437.
226. Mellon SH, Shively JE, Miller WL. Human proopiomelanocortin-(79–96), a proposed androgen stimulatory hormone, does not affect steroidogenesis in cultured human fetal adrenal cells. *J Clin Endocrinol Metab*. 1991;72:19–22.
227. Penhoat A, et al. Human proopiomelanocortin-(79–96), a proposed cortical androgen-stimulating hormone, does not affect steroidogenesis in cultured human adult adrenal cells. *J Clin Endocrinol Metab*. 1991;72:23–26.
228. Rege J, Rainey WE. The steroid metabolome of adrenarche. *J Endocrinol*. 2012;214:133–143.
229. Ingle DJ, Higgins GM. Autotransplantation and regeneration of the adrenal gland. *Endocrinology*. 1938;22:458–464.
230. Ennen WB, Levay-Young BK, Engeland WC. Zone-specific cell proliferation during adrenocortical regeneration after enucleation in rats. *Am J Physiol Endocrinol Metab*. 2005;289:E883–E891.
231. Mitani F. Functional zonation of the rat adrenal cortex: the development and maintenance. *Proc Jpn Acad Ser B Phys Biol Sci*. 2014;90:163–183.
232. Bertholet JY. Proliferative activity and cell migration in the adrenal cortex of fetal and neonatal rats: an utoradiographic study. *J Endocrinol*. 1980;87:1–9.
233. Zajicek G, Ariel I, Arber N. The streaming adrenal cortex : direct evidence of centripetal migration of adrenocytes by estimation of cell turnover rate. *J Endocrinol*. 1986;111:477–482.
234. McNicol AM, Duffy AE. A study of cell migration in the adrenal cortex of the rat using bromodeoxyuridine. *Cell Prolif*. 1987;20:519–526.
235. Carsia RV, Macdonald GJ, Gibney JA, et al. Apoptotic cell death in the rat adrenal gland: an in vivo and in vitro investigation. *Cell Tissue Res*. 1996;283:247–254.
236. Chang SP, et al. Cell proliferation, movement and differentiation during maintenance of the adult mouse adrenal cortex. *PLoS One*. 2013;8:e81865.
237. Grabek A, et al. The adult adrenal cortex undergoes rapid tissue renewal in a sex-specific manner. *Cell Stem Cell*. 2019;25:290–296.e2.
238. Salmon T, Zwemer R. A study of the life history of cortico-adrenal gland cells of the rat by means of trypan blue injections. *Anat Rec*. 1941;80:421–429.
239. Walczak EM, et al. Wnt-signaling inhibits adrenal steroidogenesis by cell-autonomous and non-cell-autonomous mechanisms. *Mol Endocrinol*. 2014;28:1471–1486.
240. Sasano H, Imatani A, Shizawa S, et al. Cell proliferation and apoptosis in normal and pathologic human adrenal. *Mod Pathol*. 1995;8:11–17.
241. Boulkroun S, et al. Aldosterone-producing adenoma formation in the adrenal cortex involves expression of stem/progenitor cell markers. *Endocrinology*. 2011;152:4753–4763.
242. Gomes DC, et al. Sonic hedgehog signaling is active in human adrenal cortex development and deregulated in adrenocortical tumors. *J Clin Endocrinol Metab*. 2014;99:E1209–E1216.
243. Nonaka K, et al. Correlation between differentiation of adrenocortical zones and telomere lengths measured by Q-FISH. *J Clin Endocrinol Metab*. 2019;104:5642–5650.
244. Moog F, Bennett CJ, Dean CM. Growth and cytochemistry of the adrenal gland of the mouse from birth to maturity. *Anat Rec*. 1954;120:873–891.
245. Badr FM, Spickett SG. Genetic variation in adrenal weight in young adult mice. *J Endocrinol*. 1971;49:105–111.
246. Bielohuby M, et al. Growth analysis of the mouse adrenal gland from weaning to adulthood: time- and gender-dependent alterations of cell size and number in the cortical compartment. *Am J Physiol Endocrinol Metab*. 2007;293:E139–E146.
247. El Wakil A, Mari B, Barhanin J, et al. Genomic analysis of sexual dimorphism of gene expression in the mouse adrenal gland. *Horm Metab Res*. 2013;45:870–873.
248. Jopek K, et al. Transcriptome profile of rat adrenal evoked by gonadectomy and testosterone or estradiol replacement. *Front Endocrinol*. 2017;8:26.
249. Nonaka K, et al. Correlation between telomere attrition of zona fasciculata and adrenal weight reduction in older men. *J Clin Endocrinol Metab*. 2020;105:dgz214.
250. Steffensen C, Bak AM, Rubeck KZ, et al. Epidemiology of Cushing's syndrome. *Neuroendocrinology*. 2010;92(suppl 1):1–5.
251. Audenet F, Méjean A, Chartier-Kastler E, et al. Adrenal tumours are more predominant in females regardless of their histological subtype: a review. *World J Urol*. 2013;31:1037–1043.
252. Arlt W, Allolio B. Adrenal insufficiency. *Lancet*. 2003;361:1881–1893.
253. Broersen LHA, Pereira AM, Jørgensen JOL, et al. Adrenal insufficiency in corticosteroids use: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2015;100:2171–2180.
254. Dörner J, et al. GLI1(+) progenitor cells in the adrenal capsule of the adult mouse give rise to heterotopic gonadal-like tissue. *Mol Cell Endocrinol*. 2017;441:164–175.
255. Looyenga BD, Hammer GD. Origin and identity of adrenocortical tumors in inhibin knockout mice: implications for cellular plasticity in the adrenal cortex. *Mol Endocrinol*. 2006;20:2848–2863.
256. Looyenga BD, Wiater E, Vale W, et al. Inhibin-A antagonizes TGFbeta2 signaling by down-regulating cell surface expression of the TGFbeta coreceptor betaglycan. *Mol Endocrinol*. 2010;24:608–620.
257. Röhrig T, et al. Toying with fate: Redirecting the differentiation of adrenocortical progenitor cells into gonadal-like tissue. *Mol Cell Endocrinol*. 2015;408:165–177.
258. Nishimoto K, et al. Aldosterone-stimulating somatic gene mutations are common in normal adrenal glands. *Proc Natl Acad Sci U S A*. 2015;112:E4591–E4599.
259. Lim JS, Rainey WE. The potential role of aldosterone-producing cell clusters in adrenal disease. *Horm Metab Res*. 2020;52:427–434.
260. Nishimoto K, et al. Adrenocortical zonation in humans under normal and pathological conditions. *J Clin Endocrinol Metab*. 2010;95:2296–2305.
261. Hadjidemetriou I, et al. DLK1/PREF1 marks a novel cell population in the human adrenal cortex. *J Steroid Biochem Mol Biol*. 2019;193:105422.
262. Wils J, et al. The neuropeptide substance P regulates aldosterone secretion in human adrenals. *Nat Commun*. 2020;11:2673.

Genetic Disorders of the Adrenal Cortex

Christopher J. Smith, Avinaash Maharaj, Louise A. Metherell, and Leonardo Guasti

OUTLINE

Primary Adrenal Insufficiency: An Overview, 1430
Genetic Disorders of the Adrenal Cortex, 1431
Steroidogenesis, 1431
Hypoplasia, 1437
Familial Glucocorticoid Deficiency-Like (FGD-Like), 1439
Other Causes Of Syndromic Adrenal Insufficiency, 1441
Conclusion, 1443
Further Resources, 1443

KEY POINTS

- Primary adrenal insufficiency is a result of adrenal dysfunction and is characterized by low cortisol and aldosterone levels, as well as high adrenocorticotropic hormone.
- Conversely, secondary and tertiary adrenal insufficiency are results of disordered pituitary or hypothalamic function.
- Glucocorticoids, mineralocorticoids, and androgens are produced from cholesterol by enzyme cascades in the adrenal cortex.

PRIMARY ADRENAL INSUFFICIENCY: AN OVERVIEW

Primary adrenal insufficiency (PAI) is the result of disordered adrenal function and is characterized by low cortisol and aldosterone levels, as well as high plasma adrenocorticotropic hormone (ACTH). This differs from secondary and tertiary adrenal insufficiency (SAI), resulting from disordered pituitary or hypothalamic function, which presents with low cortisol levels and ACTH deficiency. While cortisol and aldosterone are both deficient in PAI, in SAI, where the adrenals remain functional, it is only cortisol levels that are reduced. This is due to aldosterone being regulated through the renin-angiotensin system (RAS) and explains the relatively different clinical presentation of PAI and SAI.

Glucocorticoid production is discussed in detail elsewhere in this text, but in brief, ACTH, produced in the pituitary gland under the control of the hypothalamus–pituitary–adrenal axis, induces cortisol production through both transcription-independent stimulation and increased gene expression of steroidogenic enzymes via melanocortin 2 (MC2R) receptors in the adrenocortical cells of the zona fasciculata.[1] Upon ACTH activation, cholesterol is transported to the inner mitochondrial membrane (IMM) by steroidogenic acute regulatory protein (STAR) and converted to pregnenolone by cytochrome P450 side chain cleavage enzyme (CYP11A1) (Fig. 86.1). Pregnenolone is subsequently transported out of the mitochondria and hydroxylated to 17-OH-pregnenolone by P450c17 (CYP17A1) before being converted to 17-OH-progesterone (17-OHP) through the action of 3β-hydroxysteroid dehydrogenase (HSD3B2) at the endoplasmic reticulum. 17-OHP is next hydroxylated by P450c21 (CYP21A2) to produce 11-deoxycortisol, which is then returned to the mitochondria, and P450c11β (CYP11B1) completes the cortisol synthesis pathway.[2] Glucocorticoids regulate ACTH secretion by a negative feedback loop,[3,4] maintain cardiac function,[5-7] and are essential for hepatic glycogen deposition.[4,8] Consequently, glucocorticoid deficiency typically presents with hypoglycemia, hypotension, reduced stroke volume, and ACTH-mediated hyperpigmentation (via activation of MC1R in melanocytes).

Aldosterone production takes place in the zona glomerulosa, the outermost layer of the adrenal cortex, where its synthesis is controlled by the RAS. Briefly, a drop in blood pressure results in the release of renin from juxtaglomerular cells in the kidneys, which is then proteolytically cleaved by angiotensinogen from the liver to produce angiotensin I. Angiotensin-converting enzyme produced by the lungs converts angiotensin I to angiotensin II, which stimulates the release of aldosterone from the adrenal cortex via binding to angiotensin II receptor type I.[9] Upon angiotensin II activation, cholesterol is processed in the same manner as cortisol synthesis to produce pregnenolone. With the absence of P450c17 in the zona glomerulosa, pregnenolone is solely converted to progesterone through HSDB2. Progesterone is subsequently oxidised by P450c21 to 11-deoxycorticosterone (DOC), substrate for zona glomerulosa-specific P450c11AS (CYP11B2), which produces corticosterone, a glucocorticoid. Corticosterone is then further processed through the action of P450c11AS to produce 18-hydroxycorticosterone and finally aldosterone.[2] Mineralocorticoids regulate sodium, potassium and hydrogen ion levels in the kidneys, and therefore a deficiency of these can result in hypotension, hyponatremia, hyperkalemia, and metabolic acidosis.[10]

The functions of corticosteroids and the effects of their absence due to autoimmune and infectious adrenal insufficiency are explored extensively elsewhere in the text. This chapter will provide an overview of the clinical phenotype, comorbidities, and mechanisms of action (where known) for genetic disorders of the adrenal cortex resulting in PAI.

KEY POINTS

- Primary adrenal insufficiency can be caused by hereditary mutations that affect steroidogenic enzymes, development and growth of the adrenal gland, resistance to adrenocorticotropic hormone action or adrenal destruction.

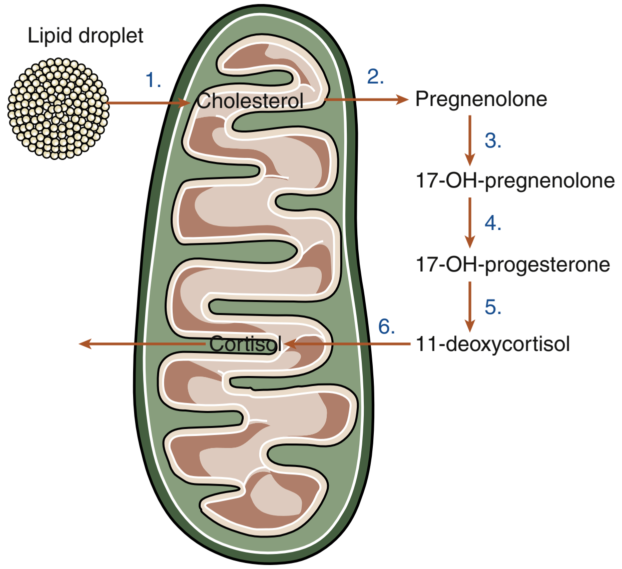


Figure 86.1 **Cortisol production pathway in the zona fasciculata.** CYP11A1 conversion of cholesterol to pregnenolone is the rate-limiting step of steroidogenesis, performed at the inner mitochondrial membrane. Pregnenolone is transported to the endoplasmic reticulum, where a cascade of enzyme activity results in the production of 11-deoxycortisol, which is subsequently returned to the mitochondria and converted to cortisol. The genes encoding the enzymes involved at each step of the pathway are as follows: 1. *STAR*, 2. *CYP11A1*, 3. *CYP17A1*, 4. *HSD3B2*, 5. *CYP21A2* 6. *CYP11B1*.

GENETIC DISORDERS OF THE ADRENAL CORTEX

PAI encompasses a heterogeneous group of diseases distinguished by decreased production of cortisol and steroid hormones by the adrenal cortex and arises from anomalies in the development of the adrenal gland, impaired steroidogenesis, resistance to ACTH action (familial glucocorticoid deficiency [FGD]) or adrenal destruction. In adults, most cases are caused by adrenal destruction due to autoimmune Addison disease, whereas in childhood the cause is usually hereditary and monogenic in origin and may be related to considerable morbidity and mortality. The International Classification of Pediatric Endocrine Diagnoses Consortium (http://www.icped.org/) classifies them into congenital adrenal hyperplasias (CAH), adrenal hypoplasia congenita (AHC), and FGD-like (http://www.icped.org/revisions/0/2015/diagnoses/#!/8), with CAH and AHC leading to adrenal insufficiencies of both gluco- and mineralocorticoids, whereas in FGD-like disease patients chiefly lack glucocorticoids (Table 86.1).

CAH accounts for most of PAI cases in childhood, with the most common type due to mutations in *CYP21A2*, but defects in *CYP11B1*, *CYP17A1*, *CYP11A1*, and *HSD3B2* can also result in CAH.[11] This results in a compensatory increase in ACTH that drives the adrenal hyperplastic response due to its trophic action on adrenocortical cells.[12,13] Congenital adrenal hypoplasia is characterized by underdevelopment of the adrenal glands and can be caused by defects in nuclear receptor subfamily 5, group A, member 1 (NR5A1) or the nuclear receptor subfamily 0, group B, member 1 (NR0B1) but is also seen as a constituent part of complex multisystem growth disorders caused by mutations in genes such as cyclin-dependent kinase inhibitor 1C (*CDKN1C*) (Table 86.1).

Finally, FGD-like or ACTH resistance syndromes[14] are characterized by isolated glucocorticoid deficiency in a patient who retains normal mineralocorticoid production. FGD-like syndromes have considerable phenotypic variation, dependent on the gene defect, but normally present at an early age with neonatal hypoglycemia, seizures, hyperpigmentation, and increased susceptibility to infection. The first mutations were identified in MC2R and its accessory protein MRAP; however, with advances in next-generation sequencing and genetic diagnosis tools, further gene mutations have been discovered that give rise to disorders with isolated cortisol deficiency through defects in cell proliferation (mini-chromosome maintenance complex component 4 [*MCM4*]) and cellular redox balance (nicotinamide nucleotide transhydrogenase [*NNT*] and thioredoxin reductase 2 [*TXNRD2*]).

KEY POINTS

- Mutations in genes that encode adrenal steroidogenic enzymes are common causes of congenital adrenal hyperplasia.
- Congenital adrenal hyperplasia can present differently depending on the dysfunctional enzyme, and patients can have varied glucocorticoid, mineralocorticoid, and androgen production.
- Mutations in *CYP11A1* and *STAR* can also cause congenital lipoid adrenal hyperplasia, a disease where lack of hormone production results in increased accumulation of cholesterol, and consequently mitochondrial oxidative damage.

Steroidogenesis

Defects in enzymes within the adrenal steroidogenic cascade (Fig. 86.2) are common causes of CAH. As a complex network of enzymes, mutations in different genes can cause variable disease phenotypes due to deficiency of glucocorticoid, mineralocorticoid, or both, as well as possible gonadal insufficiency.

Cytochrome P450 Family 21 Subfamily A Member 2 (*CYP21A2*). *CYP21A2* encodes steroid 21-hydroxylase, the steroidogenic enzyme that converts progesterone to deoxycortisone and 17-OHP to deoxycortisol (Fig. 86.2). Mutations in *CYP21A2* cause 90% of CAH and result in reduced mineralocorticoid and glucocorticoid production, as well as androgen excess, leading to hypergonadism. CAH due to steroid 21-hydroxylase deficiency presents with three major phenotypes: classical salt-wasting, classical simple virilizing, and nonclassical. The absence of aldosterone results in the inability to retain sodium, which is consequently lost in urine, a process known as salt wasting.[15] Classical salt-wasting is the more severe of the two classical forms and constitutes 75% of cases.[16] Due to mutations resulting in less than 2% enzyme activity, undiagnosed cases can lead to life-threatening adrenal crisis in the first 2 weeks of life.[17,18] In addition, 46, XY patients are born with normal genitalia, putting them most at risk, while 46, XX patients are commonly born with genital ambiguity, prompting further evaluation and treatment.[18] For classical salt-wasting cases, both mineralocorticoid and glucocorticoid therapy are required. In the simple virilizing cases, which possess greater than 2% enzyme activity, sufficient aldosterone is produced to prevent salt wasting. For both disease variants, patients undergoing postnatal development may develop signs of androgen excess, including precocious pubarche, adrenarche, acne, and advanced bone age leading to reduced final height.[19,20] Furthermore, 46, XY patients are likely to develop testicular adrenal rest tumors, benign testicular tumors that morphologically and functionally resemble adrenal tissue.[21] Patients with the nonclassical form of the disease retain up to 50% of enzyme activity. This allows sufficient hormone production to prevent adrenal insufficiency and genital ambiguity in 46, XX patients, resulting in many patients going undiagnosed.[18,22] The primary method to diagnose steroid 21-hydroxylase deficiency is to measure serum levels of 17-OHP, the substrate for steroid 21-hydroxylase, during newborn screening.[18]

TABLE 86.1 **Monogenetic causes of primary adrenal insufficiency and associated phenotype(s) in humans.**

GENE NAME	GENE ID	OMIM DISEASE	OMIM GENE	PHENOTYPE	GENOMIC LOCATION	GENOME COORDINATES (GRCH38)	SUBCELLULAR LOCATION OF PROTEIN	MODE OF INHERITANCE	MOUSE MODEL IMPC	MOUSE MODEL OTHER	MOUSE REF	HUMAN REF
Cytochrome P450 family 21 subfamily A member 2	*CYP21A2*	201910	613815	Congenital adrenal hyperplasia, ambiguous genitalia, virilisation	6p21.33	6:32,038,414-32,041,643	ER	AR	Yes (Cyp21a1) preweaning lethality (in)complete penetrance	No	N/A	11
Steroidogenic acute regulatory protein	*STAR*	201710	600617	Lipoid congenital adrenal hyperplasia	8p11.23	8:38,142,699-38,150,951	Mitochondria	AR	No	Yes, model for CLAH	31	27
Cytochrome P450, family 11, subfamily A member 1	*CYP11A1*	613743	118485	Adrenal insufficiency, 46,XY sex reversal	15q24.1	15:74,337,761-74,367,645	Mitochondria	AR	Yes, Preweaning lethality complete penetrance	Yes, model for p450scc	41	34
Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	*HSD3B2*	201810	613890	Congenital adrenal hyperplasia, male pseudohermaphroditism	1p12	1:119,414,930-119,423,033	Mitochondria, ER	AR	No	No	N/A	45
Cytochrome P450 family 17, subfamily A, polypeptide 1	*CYP17A1*	202110	609300	Congenital adrenal hyperplasia, ambiguous genitalia, hypokalemic hypertension, delayed puberty	10q24.32	10:102,830,530-102,837,412	ER	AR	Yes, no adrenal phenotype, decreased bone mineral density	Yes, embryonic lethal, no adrenal phenotype tested	56	51
Cytochrome P450 Oxidoreductase	*POR*	613571	124015	Congenital adrenal hyperplasia, virilization in females, Antley–Bixler syndrome	7q11.23	7:75,915,154-75,986,854	ER, cytosol	AR	No	Embryonic lethal incomplete penetrance, hypomorph with no adrenal phenotype noted	63	59
Nuclear receptor subfamily 0 group B member 1	*NR0B1*	300200	300473	Adrenal hypoplasia congenital, Hypogonadotrophic hypogonadism	Xp21.2	X:30,304,205-30,309,389	Nucleus	XR	Yes, No adrenal phenotype, abnormal bone mineralization	Yes, gonadal phenotype in males, adrenal phenotype on aging	68	69

TABLE 86.1—cont’d

GENE NAME	GENE ID	OMIM DISEASE	OMIM GENE	PHENOTYPE	GENOMIC LOCATION	GENOME COORDINATES (GRCH38)	SUBCELLULAR LOCATION OF PROTEIN	MODE OF INHERITANCE	MOUSE MODEL IMPC	MOUSE MODEL OTHER	MOUSE REF	HUMAN REF
Nuclear receptor subfamily 5A1	*NR5A1*	612965	184757	46,XY sex reversal, partial or complete with or without adrenal failure	9q33.3	9:124,481,235-124,507,398	Nucleus	AD	No	Yes, adrenal and gonadal phenotype	78	82
Wnt family member 4	*WNT4*	611812	603490		1p36.12	1:22,117,307-22,143,980	Extracellular, ER, membrane, Golgi	AR	Yes, mice die within 24 h of birth	No	87,88,89, 90,91	86
Cyclin-dependent kinase inhibitor 1C	*CDKN1C*	614732	600856	IMAGE Syndrome - intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies	11p15.4	11:2,883,217-2,885,774	Nucleus	AD	No	No, models for Beckwith–Wiedemann and Silver–Russell syndromes but no adrenal phenotype	97	93
Sterile alpha motif domain containing 9	*SAMD9*	617053	610456		7q21.2	7:93,099,517-93,117,978	Cytosol	AD	No	No	N/A	98
Polymerase epsilon 1	*POLE*	N/A	174762		12q24.33	12:132,623, 757-132,687, 518	Nucleus	AR	No	Yes, no adrenal phenotype	104	101
Melanocortin 2 receptor	*MC2R*	202200	607397	Isolated GC deficiency	18p11.21	18:13,882,041-13,915,706	Membrane	AR	No	Yes, recapitulates FGD	113	107
Melanocortin 2 receptor accessory protein	*MRAP*	607398	609196	Isolated GC deficiency	21q22	21:32,291,812-32,314,783	Membrane	AR	No	Yes, recapitulates FGD	124	116
Nicotinamide nucleotide transhydrogenase	*NNT*	614736	607878	Isolated GC deficiency (TART, DCM)	5p12	5:43,601,091-43,707,395	Mitochondria	AR	Yes, adrenal phenotype not noted, decreased bone mineral density	Yes, 50% reduction in corticosterone	126	126
Thioredoxin reductase 2	*TXNRD2*	617825	606448	Isolated GC deficiency (cardiac defects)	22q11	22:19,875, 521-19, 941,817	Mitochondria, cytosol	AR	No	Yes, embryonic lethal, no adrenal phenotype tested	136	136
Aladin WD repeat nucleoporin	*AAAS*	231550	605378	Hypoadrenalism, alacrima, achalasia, neurological dysfunction	12q13.13	12:53,307,455-53,321,609	Cytoskeleton, nucleus, cytosol	AR	No	Yes, no adrenal phenotype	145	140

Continued

TABLE 86.1—cont'd

GENE NAME	*GENE ID*	*OMIM DISEASE*	*OMIM GENE*	*PHENOTYPE*	*GENOMIC LOCATION*	*GENOME COORDINATES (GRCH38)*	*SUBCELLULAR LOCATION OF PROTEIN*	*MODE OF INHERITANCE*	*MOUSE MODEL IMPC*	*MOUSE MODEL OTHER*	*MOUSE REF*	*HUMAN REF*
Mini-chromosome maintenance complex component 4	*MCM4*	607398	602638	Adrenal failure, Natural Killer cell deficiency, short stature, chromosomal instability	8q11.21	8:47,960,940-47,978,159	Nucleus	AR	Yes, embryonic/pre-weaning lethality complete penetrance	Yes, hypomorph is small with adrenal phenotype	147,149	147
Sphingosine-1 phosphate lyase	*SGPL1*	617575	603729	Primary adrenal insufficiency, Steroid resistant nephrotic syndrome, ichthyosis, primary hypothyroidism, primary gonadal failure, dysplidemia, neurological abnormalities, lymphopenia	10q22.1	10:70,815,919-70,881,183	ER	AR	Yes, preweaning lethality complete penetrance	Yes, recapitulates main characteristics of SGPL1 deficiency	155	155
Autoimmune regulator	*AIRE*	240300	607358	Autoimmune polyendocrine syndrome type I Hypoparathyroidism, immune deficiency, chronic candidiasis, autoimmune adrenal insufficiency	21q22.3	21:44,285,875-44,298,647	Nucleus	AR, AD	Yes, embryonic lethality (complete?)	Yes, good model for APS1	167	165
Peroxisomal Biogenesis Factor 6	*PEX6*	614862	601498		6p21.1	6:42,963,864-42,980,223	Peroxisome, cytosol	AR, AD	No	No	N/A	184
Peroxisomal Biogenesis Factor 1	*PEX1*	214100	602136		7q21.2	7:92,487,022-92,528,519	Peroxisome, cytosol	AR	Yes, no adrenal phenotype	Yes, mild Zellweger phenotype	186	185
ATP Binding Cassette Subfamily D Member 1	*ABCD1*	300100	300371		Xq28	X:153,724,850-153,744,754	Mitochondria, peroxisome, lysosome ER, cytosol	XR	Yes, no adrenal phenotype	Yes, model for adrenoleukodystrophy	179	202
Lysosomal Acid Lipase	*LIPA*	278000	613497		10q.23.31	10:89,213,570-89,251,925	Lysosome, nucleus	AR	Yes, develop accumulations of triglycerides and cholesterol esters	No	201	203

IMPC, International Mouse Phenotyping Consortium; *ER*, endoplasmic reticulum; *CLAH*, congenital lipoid adrenal hyperplasia; *AR*, autosomal recessive; *AD*, autosomal dominant; *XR*, X-linked recessive; *OMIM*, online Mendelian inheritance in man.

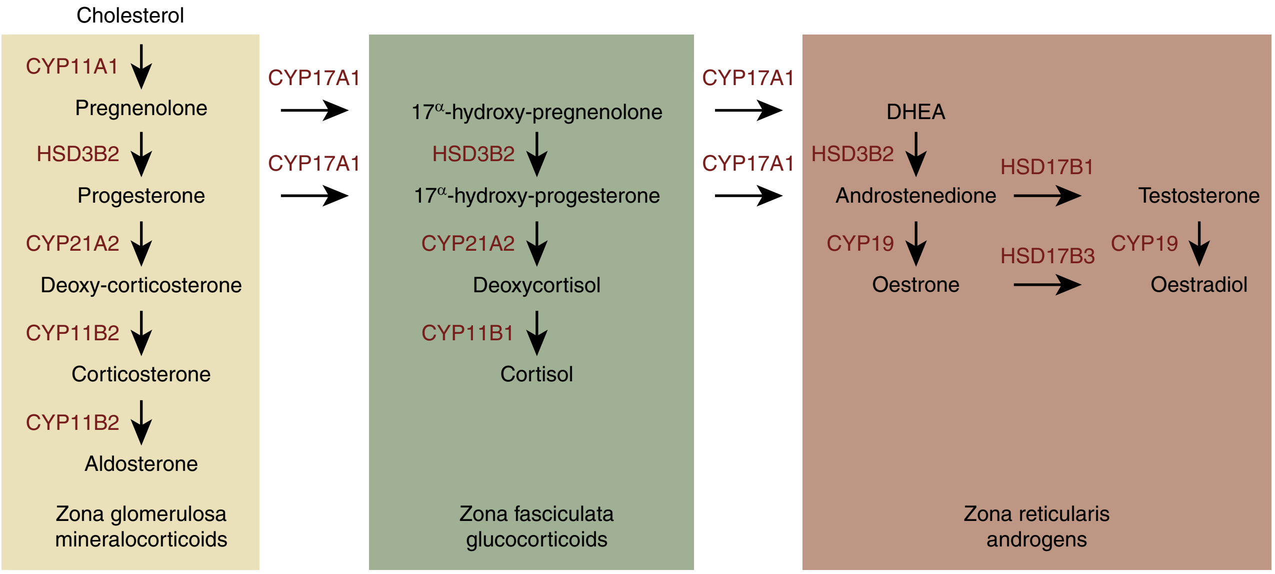


Figure 86.2 **Steroidogenic cascade of the adrenal gland.** The adrenal gland produces mineralocorticoids, glucocorticoids, and androgens in different zones of the cortex. Mutations in several of the *CYP* genes result in congenital adrenal hyperplasia, leading to a compensatory increase in adrenocorticotropic hormone that drives the adrenal hyperplastic response due to its trophic action on adrenocortical cells. *DHEA,* dehydroepiandrosterone.

A naturally occurring mouse strain, H-2aw18, with deletion of *cyp21* and complement component c4 due to meiotic recombination, dies at an early postnatal stage.[23] The loss of *cyp21* leads to hyperproduction of ACTH with adrenocortical hyperplasia and accumulation of precursor steroids.[24] Because mouse adrenals lack steroid 17-alpha-hydroxylase/17,20 lyase, the accumulated precursor is progesterone, and no androgenization occurs. The adrenals of these mice present with lack of zonation, hypertrophy, and hyperplasia of adrenocortical mitochondria, although adenovirus-mediated transfer of human steroid 21-hyrdoxylase into these mice has been shown to transiently restore adrenal steroidogenesis.[24]

Steroidogenic Acute Regulatory Protein (*STAR*). STAR is part of a complex responsible for the transportation of cholesterol from the outer mitochondrial membrane to the IMM, where it is processed to pregnenolone, reported as the rate-limiting step of steroidogenesis (Fig. 86.1). Along with mitochondrial cholesterol side-chain cleavage enzyme (CYP11A1), it is one of the key factors in hormone production in both the adrenals and the gonads.[25] Mutations in *STAR* therefore inhibit the delivery of cholesterol to the steroidogenic enzyme cascade and prevent glucocorticoid and mineralocorticoid production.[26] This shutdown of steroidogenesis due to mutations in *STAR* is known as congenital lipoid adrenal hyperplasia (CLAH), first characterized in 1996.[27] Patients with this condition present in the first month of life with salt-wasting adrenal and gonadal insufficiency.[28] Due to dysfunctional gonadal steroidogenesis, CLAH infants with a 46, XY karyotype fail to produce testosterone *in utero* and have female genitalia, while infants with a 46, XX karyotype present with variable ovarian insufficiency at adolescence.

A two-hit model has been proposed to account for the pathogenesis of CLAH, with the first "hit" being lack of *STAR* and the second being cholesterol-mediated oxidative damage.[27] Upon ACTH stimulation, a steroidogenic cell gathers cholesterol from endogenous synthesis, stored lipid droplets or low-density lipoprotein (LDL) receptor-mediated endocytosis, and STAR subsequently facilitates transport through the mitochondria. In cells with defective STAR transport, only minimal residual STAR-independent steroid biosynthesis occurs, resulting in low levels of steroidogenesis (first hit). Inadequate steroid levels drive an increase in steroidogenic stimuli and culminate in the accumulation of cholesterol within the cells. This progressive lipid accumulation leads to mitochondrial oxidative damage, which eventually culminates in reduced steroidogenic capacity and enlarged adrenal size (second hit).[27]

Beyond the "classic" disease of CLAH, partial defects in *STAR* can cause nonclassic CLAH (NCLAH) in 5% to 10% of patients.[29,30] This disease predominantly affects glucocorticoid synthesis and masquerades as FGD, as patients have normal genitalia and may be fertile. Variants pR188C and pR192C reduce STAR's interactions with the 3–βOH group of cholesterol and cause partial loss of function (LOF).[29,30] Long-term reproductive follow-up is required for NCLAH patients to ensure that correct pubertal development occurs, and appropriate counseling and sperm banking may be offered.[28]

Star knockout mice are born normally, indicating that STAR is not required for survival *in utero*; however, both genetic male and genetic female mice have female external genitalia.[31] Most mice die within 1 to 2 days of birth, with none surviving beyond 16 days without treatment, most likely due to respiratory distress secondary to adrenal insufficiency.[31] This was supported by the survival of pups that underwent corticosteroid replacement. Furthermore, levels of corticosterone and aldosterone were low compared to littermates, while levels of ACTH were elevated. Finally, although adrenal glands were smaller in knockout mice, they showed high quantities of lipid deposition, supporting the second hit of the two-hit model of CLAH.[31]

Cytochrome P450, Family 11, Subfamily A Member 1 (*CYP11A1*). *CYP11A1* encodes the mitochondrial cholesterol side-chain cleavage enzyme that performs the rate-limiting step of steroidogenesis, cholesterol conversion to pregnenolone, at the beginning of the steroidogenic enzyme cascade (Fig. 86.2). Pregnenolone is subsequently processed to produce cortisol, aldosterone, and androgens in different pathways and, consequently, loss-of-function mutations can cause CAH associated with glucocorticoid and mineralocorticoid deficiency, as well as disorders of sex development. Absence of the enzyme was previously believed to be incompatible with term gestation; however, several patients with side-chain cleavage enzyme deficiency have been reported, initially in 2001.[32-34]

Mutations in *CYP11A1* classically result in mitochondrial cholesterol side-chain cleavage enzyme syndrome characterized by primary adrenal failure in combination with 46, XY sex reversal, although, unlike STAR-deficient subjects, adrenal size is unaffected. Similar to *STAR,* partial loss-of-function mutations (retention of up to 20% of wild-type function) can produce a nonclassic form of the disease with a phenotype of predominant glucocorticoid deficiency with variable degrees of mineralocorticoid deficiency and sexual development, depending on the residual activity of the mutant proteins.[35-38] Patients can present at different ages during infancy, typically with hyperpigmentation, hypoglycemia, or prolonged illness with infections.[28]

Two recent studies have found that missplicing of *CYP11A1* mRNA results in a truncated version of the protein that can cause adrenal insufficiency.[39,40] Genetic analysis revealed that many families of European heritage with PAI are compound heterozygous for a c.940G>A variant (rs6161) on one allele of *CYP11A1,* with a second disruptive change on the other allele.[39] The *in vitro* splicing assays showed that the predicted benign change of c.940G>A (p.E314K) resulted in missplicing and a truncated protein consistent with cleavage at the predicted amino acid change.[39,40] This variant is carried by 1:140 people of European descent but has only been shown to cause PAI when inherited with a rare disruptive change on the other allele.[28]

A study investigating *Cyp11a1* null mice showed that they do not survive weaning and present with corticosterone and aldosterone deficiency compared with wild-type littermates. However, when treated with corticosteroids they survive to adulthood. *Cyp11a1*$^{-/-}$ XY males exhibit sex reversal with feminization of external genitalia and disorganization of internal genitalia. The adrenals of the null mice are reduced in size, although they do show progressive lipid accumulation similar to *Star* knockout mice.[41]

Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- And Steroid Delta-Isomerase 2 (*HSD3B2*). Mutations in *HSD3B2* result in a very rare variant of CAH causing less than 0.5% of cases.[42] HSD3B2 catalyzes the conversion of Δ5-3β-hydroxysteroids (pregnenolone, 17-hydroxypregnenolone and dehydroepiandrosterone [DHEA]) to Δ4-3-ketosteroids (progesterone, 17-OHP and androstenedione) (Fig. 86.2). HSD3B2 deficiency therefore impairs the synthesis of aldosterone, cortisol, and testosterone and results in high levels of renin, ACTH, and DHEA.[43] DHEA can be converted to androstenedione by extraadrenal HSD3B1, resulting in high levels of testosterone in females, although this fails to compensate for absence of the hormone in males.[43,44]

Severe impairment of steroid synthesis was first mapped to *HSD3B2* through identification of nonsense and frameshift mutations in 1992.[45] In general, frameshift and nonsense mutations present with a salt-wasting phenotype, while missense mutations are associated with residual enzyme activity and a non-salt-wasting phenotype.[15,46] Further presenting symptoms include incomplete masculinization in 46, XY patients, virilization in 46, XX patients, gynecomastia in males, hypospadias, and palpable testis.[47-49] This form of CAH is best diagnosed by high levels of 17-hydroxypregnenolone in the blood plasma and is treated with glucocorticoid, mineralocorticoid, and possibly sex hormone replacement therapy.

Cytochrome P450 Family 17, Subfamily A, Polypeptide 1 (*CYP17A1*). Due to its location in the steroidogenic cascade (Fig. 86.2), steroid 17-alpha-hydroxylase/17,20 lyase (CYP17A1) acts as the exclusive gateway to androgen production, with severe mutations resulting in impaired adrenal and gonadal sex steroid production, causing sexual infantilism and pubertal failure.[44,50] Steroid 17-alpha-hydroxylase/17,20 lyase functions in the conversion of pregnenolone to 17-hydroxypregnenolone and onwards to DHEA, as well as that of progesterone to 17-OHP and onwards to androstenedione, and therefore deficiency of the enzyme results in low cortisol and androgens but high levels of DOC and corticosterone (Fig. 86.2). Pubertal failure is therefore a prominent feature of steroid 17-alpha-hydroxylase/17,20 lyase deficiency. Importantly, however, due to the high levels of DOC and corticosterone, which possess glucocorticoid properties, mutations in *CYP17A1* do not result in glucocorticoid deficiency and adrenal crisis.[44] High levels of DOC, however, cause sodium retention, leading to hypertension and consequently suppression of aldosterone production.

CYP17A1 deficiency was first identified in 1988 in a patient with a four-base duplication.[51] 46, XY patients with *CYP17A1* mutations are usually diagnosed at birth due to ambiguous external genitalia.[52] Conversely, as patients do not undergo adrenal crisis at a young age, many 46, XX patients are not diagnosed until adolescence. Although DOC accumulates, manifestations of mineralocorticoid excess tend not to occur until adolescence, where it presents with hypertension and hypokalemia.[50] At puberty, lack of DHEA prevents adrenarche and development of pubic and axillary hair, and 46, XX patients can present with amenorrhea without hypertension and ovarian cysts.[53-55] At present there are no published examples of successful pregnancy in which either partner has 17-hydroxylase deficiency. The deficiency is treated with the appropriate androgen replacement therapy and partial glucocorticoid replacement. While glucocorticoid therapy is not essential, due to corticosterone substitution, it functions to prevent adrenal axis suppression and normalize blood pressure, and therefore a partial replacement is used as a compromise.[50] In one study, *Cyp17a1*$^{-/-}$ mice died around embryonic day 7 and could not be rescued through maternal administration of DHEA, androstenedione, and 17-OH pregnenolone.[56] The International Mouse Phenotyping Consortium, however, presents data indicating that mice reach adulthood but display varying phenotypes, including increased circulating cholesterol level and absent male genitalia.

Cytochrome P450 Oxidoreductase (*POR*). Cytochrome P450 oxidoreductase (POR) is a flavoprotein that plays a key role in electron transport, donating electrons to mitochondrial P450 enzymes including steroid 17-alpha-hydroxylase/17,20 lyase and steroid 21-hydroxylase and aromatase (CYP19A1).[44] Therefore, in POR deficiency (PORD), a reduction in steroidogenesis is due to impairment of key steroidogenic enzymes resulting in a lack of glucocorticoids and androgens. As POR acts an electron donor for all microsomal cytochrome P450 enzymes, there are multiple comorbidities associated with PORD. These can, but do not always, include changes in drug metabolism, pathogenesis of skeletal dysplasia, and other skeletal manifestations that resemble Antley–Bixler syndrome.[57-59]

The first mutations in POR linked with disordered steroidogenesis were identified in 2004.[59] Most POR mutations retain some enzymatic function, resulting in high levels of pregnenolone, progesterone, 17-OHP, DOC, and corticosterone, low levels of DHEA and androstenedione, and variable levels of cortisol and aldosterone.[44,60] Affected individuals can be detected *in utero,* as fetoplacental deficiency of POR can lead to virilization of the mother due to defective conversion of androgen precursors to estrogen.[59] Once born, affected 46, XX patients present with ambiguous genitalia, but circulating androgens remain at a low level, and virilization does not progress.[61] Urinary steroid metabolite profiling also provides a further useful noninvasive method for diagnosis.[44] PORD presentation varies greatly, from amenorrhea and polycystic ovaries in 46, XX patients and androgen deficiency in 46, XY patients to severe hormone disturbances resulting in atypical genitalia. One study showed that more than 80% of 46, XX patients in their cohort presented with differences in sex development (DSD), whereas DSD was present in only half of 46, XY patients.[62] Patients do not develop mineralocorticoid deficiency, but require permanent or stress-dosed glucocorticoid replacement therapy.

In one study, targeted deletion of the translation start site and membrane-binding domain in *Por*$^{-/-}$ mice resulted in embryonic lethality by day 13.5 of gestation, with embryos possessing multiple defects including abnormalities in the neural tube, heart, eye, and limb or a gross retardation of development.[63] A further study created a hypomorphic mouse with 5% to 26% of normal POR activity that grew to adulthood and presented with decreased weight and decreased heart, lung, and kidney size.[64] Interestingly no adrenal phenotype was noted in these mice.

KEY POINTS

- Mutations in certain genes can cause congenital adrenal hypoplasia, which is underdevelopment or hypotrophy of the adrenal cortex.
- Mutations in the genes encoding the transcription factors NROB1 and NR5A1 affect the transcription of genes involved in steroidogenesis and the development of the adrenal glands themselves.
- Mutations in genes such as *CDKN1C*, *SAMD9,* and *POLE1* can cause complex multisystem growth disorders that contain adrenal hypoplasia as a hallmark of the disease.

Hypoplasia

Adrenal hypoplasia is the underdevelopment or hypotrophy of the adrenal cortex leading to reduced steroid hormone production, and subsequently PAI. For more detail on adrenocortical developmental genetics, refer to Chapter 85.

Nuclear Receptor Subfamily 0 Group B Member 1 (*NROB1*). AHC is a condition that results from the underdevelopment of the adrenal gland. It is caused by LOF mutations to dosage-sensitive sex reversal, adrenal hypoplasia critical region on chromosome X, DAX1, encoded by *NROB1*, and classically presents with PAI and associated salt-wasting, hypogonadotropic hypogonadism (HH), and infertility.[28] DAX1, a transcription factor belonging to the nuclear receptor superfamily, acts as a transcriptional repressor of genes involved in the steroidogenic pathway and the development of the hypothalamus–pituitary–gonadal axis.[65,66] DAX1, in part, serves as a transcriptional repressor of steroidogenic factor 1 (SF-1, also known as NR5A1)-mediated transcriptional activation. Furthermore, DAX1 downregulates anti-Müllerian hormone (AMH), a hormone that, in 46, XY embryos, inhibits folliculogenesis and development of female gonads.[67] Loss of DAX1-dependent inhibition therefore results in abnormal gonad development in 46, XY patients. This is supported by DAX1-deficient mice presenting with testicular abnormality and infertility.[68]

Pathogenic variants in *NROB1* were first reported as a cause of X-linked AHC in 1994.[69] Its association with Duchenne muscular dystrophy and glycerol kinase deficiency allowed for an AHC-critical region of 200 to 500 kb to be physically mapped to the short arm of the X chromosome.[69] Most affected infants present within the first months of life with classic PAI symptoms: low cortisol, high ACTH, low aldosterone, and increased renin resulting in salt wasting.[70] Affected individuals are therefore treated with replacement doses of hydrocortisone and salt supplements, as well as sex hormone replacement to induce puberty if necessary. Patients may also present with paradoxical features such as precocious puberty, making the clinical spectrum of the disease quite variable.[71] Late-onset X-linked AHC has also been well documented. Patients can also present with partial HH and infertility. This is often due to a partial LOF variant in the ligand-like binding domain or an early stop variant leading to translation of a truncated protein.[72-74] Moreover, late-onset adrenal failure in patients with AHC has been preceded by cortisol excess in some patients.[75] While it remains unclear how LOF *NROB1* mutations could lead to these disparate adrenal phenotypes,[65,66] data in mice with genetic LOF mutations in *NROB1* are consistent with premature adrenocortical progenitor cell differentiation at the expense of ultimate adrenal failure.[76]

Nuclear Receptor Subfamily 5 Group A Member 1 (*NR5A1*). SF-1 is a transcription factor belonging to the nuclear receptor superfamily and acts as a master regulator of both adrenal and gonadal development.[77] Expressed in the adrenal cortex, testicular Leydig cells, and ovarian theca cells, SF-1 regulates the transcription of genes encoding key steroidogenic enzymes such as CYP11A1, CYP11B1, CYP21, and STAR, as well as AMH, gonadotrophins, and aromatase.[78] Indeed, SF-1 acts as an active determinant of gender differentiation through upregulation of *AMH.*[79] As demonstrated in mice, loss of Sf-1 can therefore cause adrenal agenesis, gonadal dysgenesis with a subsequent female phenotype, and variable degrees of HH.[78]

Mutations in *NR5A1* were first associated with adrenal insufficiency and XY sex reversal in 1999.[80] Here a 46, XY patient presented with primary adrenal failure as well as phenotypically female genitalia as a result of failure of Müllerian regression due to loss of SF-1–directed AMH production. In 46, XY patients, pathogenic variants are associated with a range of phenotypes, including testicular dysgenesis/dysfunction, severe hypospadias, and infertility.[81] Primarily causing defects in reproductive development, *NR5A1* mutations have rarely been associated with PAI, with only six patients documented to date.[28] Affected individuals with PAI can present with hyperpigmentation, hypotonia, jaundice, weight loss, and failure to thrive.[82]

Wnt Family Member 4 (*WNT4*). WNT4 belongs to the large WNT family of secreted glycoproteins that play a key role in the development of multicellular animals. WNT4 itself is considered the most important molecule for the correct differentiation of female gonadal tissues and is also known to play a role in the development of the kidneys, adrenals, pituitary gland, and mammary tissues.[83,84] Considering that the gonads, kidneys, and adrenals all originate from the urogenital ridge, it can be assumed that WNT4 functions in early embryological development. Mutations in *WNT4* can therefore disrupt the normal development of these organs and cause adrenal insufficiency. WNT4 is also known to repress the steroidogenic enzymes *CYP17A1* and *HSD3B2,* and consequently loss of WNT4 leads to androgen excess and sex reversal.[85]

Adrenal insufficiency due to *WNT4* mutations is rare, and the autosomal recessive syndrome it causes, SEx Reversion, Kidneys, Adrenal and Lung dysgenesis (SERKAL) syndrome, has been identified in three fetuses.[86] This study used a candidate gene approach to identify a homozygous missense mutation in *WNT4* (p.A114V) in three terminated fetuses. The parents of two of these fetuses were second-degree cousins, with the third fetus also belonging to the wider kindred, and the pregnancies were terminated due to renal agenesis determined through ultrasonic scanning. This mutation markedly reduced *WNT4* mRNA levels in transfected ovarian epithelium cells compared with wild-type *WNT4*, resulting in a downregulation of WNT4-dependent inhibition of β-catenin degradation.[86] This change in β-catenin-induced gene expression could result in the developmental defects witnessed in SERKAL syndrome. *Wnt4*-deficient mice die within 24 h of birth, showing signs of kidney dysfunction and improper development of the pituitary gland, as well as partial sex reversal in females.[87-89] In the adrenal cortex, *Wnt4*-deficient mice showed reduced expression of Cyp11b2 in the zona glomerulosa, resulting in significantly lowered aldosterone secretion in newborn mice, suggesting a role of Wnt4 in the development and maintenance of the zona glomerulosa.[90] A mouse model in which Wnt4 was depleted in the adrenal cortex corroborated these results, while also uncovering crucial crosstalk between PKA and Wnt-beta catenin signaling in determining the differentiation and fate of zona glomerulosa cells.[91]

Cyclin Dependent Kinase Inhibitor 1c (*CDKN1C*). A complex multisystem growth disorder that involves PAI is IMAGe syndrome, characterized by Intrauterine growth restriction, Metaphysical dysplasia, Adrenal hypoplasia, and Genitourinary anomalies (in males), and was first described in 1999.[92] The syndrome is normally caused by heterozygous missense variants in cyclin-dependent kinase inhibitor 1C (CDKN1C), a negative regulator of cell proliferation. All identified mutations that cause IMAGe syndrome are clustered in the proliferating cell nuclear antigen (PCNA) binding domain and cause a gain-of-function resulting in repressed growth.[93] The mechanism is still unclear but may involve reduced degradation of CDKN1C and subsequently prolonged cell cycle repression by impairing entry into S phase.[28,94] Conversely, LOF mutations outside of the PCNA binding domain of the *CDKN1C* gene contribute to the overgrowth phenotype of Beckwith–Wiedemann syndrome (including adrenal cytomegaly) due to an imprinting defect of the *IGF2* locus at 11p15 (containing *CDKN1C*).[93]

IMAGe syndrome is clinically variable, with some patients presenting without adrenal insufficiency and with diabetes mellitus.[95] This diabetes phenotype is most likely due to CDKN1C being involved in insulin-producing β cells. *CDKN1C* is paternally imprinted on chromosome 11 and is consequently only expressed from the maternal allele; while this inheritance can mimic an X-linked disease, IMAGe syndrome patients can be male or female.[96] Expression of IMAGe-causing mutations (mutations in the PCNA binding domain) in a human embryonic kidney (HEK) cell line demonstrated that, in IMAGe syndrome CDKN1C becomes more stable and does not undergo normal cell cycle–related oscillation, inhibiting progression from the G1 phase.[94] Overexpression of wild-type CDKN1C and Beckwith–Wiedemann syndrome mutant CDKN1C both inhibited cell proliferation, but not to the same extent as the IMAGe mutant. No mouse line has been created with an IMAGe-mutant *CDKN1C,* but a mouse containing a Beckwith–Wiedemann mutant *CDKN1C* presented with no adrenal phenotype.[97]

Sterile Alpha Motif Domain Containing 9 (*SAMD9*). A further complex multisystem growth restriction disorder that causes adrenal hypoplasia is MIRAGE syndrome, named according to it symptoms: Myelodysplasia, Infections, Restricted growth, Adrenal hypoplasia, Gonadal anomalies, and Enteropathy.[98] The syndrome is a result of gain-of-function mutations in the growth suppressor gene *SAMD9,* located on the long arm of chromosome 7 (7q21.2).[98,99] Conversely, LOF mutations have been reported in tumors, including normophosphatemic familial tumoral calcinosis.[100] Patients with MIRAGE syndrome commonly die at a young age (<2 years) due to invasive infection, are born preterm, and have variable intellectual development, and patients who do survive show limited pubertal progression.[28,98] They can also present with further symptoms, including respiratory distress, anemia, and thrombocytopenia.[28]

SAMD9 was first discovered to cause MIRAGE syndrome by a study performed in Japan that involved whole-exome sequencing of six patients presenting with adrenal hypoplasia but no mutations in other known genes.[98] Three patients were found to have variants in *SAMD9,* and subsequent Sanger sequencing of other cohort members identified a further eight patients, with eight different variants in total, almost exclusively *de novo*.[98] Postmortem analysis of two 46, XX patients who died before 2 years of age showed very small and disorganized adrenal glands, with the medulla only partially surrounded by cortex.[98] The same study transfected HEK cells with wild-type *SAMD9* and the different variants of *SAMD9*. Their results displayed markedly increased growth inhibition in the gain-of-function variants compared with wild-type, a finding that supported by a separate study with seven further variants.[98,99] This restriction in cell growth could be due to abnormalities in the endosome system and impaired extracellular-signal regulated kinase and epidermal growth factor receptor recycling.[98]

Patients from one cohort who survived the first 3 months of life all showed signs of a postzygotic loss-of-heterozygosity event, such as monosomy, deletions of 7q, or secondary somatic LOF mutations in hematopoietic cell lines, all of which have been associated with increased survival by rescuing the blood phenotype.[99] In this study, four patients possessed somatic nonsense or frameshift LOF mutations; these variants were then transfected into HEK cells along with the primary gain-of-function variants. This resulted in an increase in proliferation, indicating that the gain-of-function phenotype can be "rescued."[99] However, although monosomy or 7q deletion can lead to increased survival, it has also been shown that patients can develop myelodysplastic syndrome, a cancer that can be lethal.[98,99]

Polymerase Epsilon-1 (*POLE1*). *POLE1* encodes the essential Pol ε, the major leading-strand DNA polymerase involved in DNA replication that also possesses exonuclease ability for genetic proofreading. Somatic and germline mutations in *POLE1* have been associated with colon and endometrial cancer, but biallelic LOF mutations can also cause an IMAGe-like syndrome.[101] One study that performed whole-genome sequencing of 48 individuals with microcephalic primordial dwarfism, a set of extreme growth disorders, identified heterozygous LOF *POLE1* mutations in three patients.[101] The cohort was increased to 15 patients by identifying previously published IMAGe patients who were mutation-negative for *CDKN1C*. Of the cohort, 12 patients presented with adrenal insufficiency with variable salt wasting, and all males had genitourinary abnormalities. Interestingly, all patients possessed the same intronic variant (c.1686+32C>G) as part of a common haplotype, in combination with different LOF mutations in *trans* (11 different mutations in the 15-patient cohort).

Experiments in patient-derived fibroblasts demonstrated that the intronic variant alters splicing, and part of intron 15 is retained. This inclusion of an extra 47 bp causes a frameshift that results in an early stop codon (p.Asn563Valfs*16). This mutation, along with a LOF mutation on the other allele, significantly reduces the protein levels of POLE1 in patient fibroblasts, leading to delayed S-phase progression and, consequently, restricted cell growth.[101]

A single homozygous intronic variant (c.4444+3A>G) in *POLE1* has been associated with immunodeficiency, lymphopenia, and short stature.[102,103] Immunodeficiency was also identified in seven of the 15 patients from the discussed study, and therefore the spectrum of biallelic *POLE* mutations is greater than that of IMAGe syndrome caused by *CDKN1C* mutations.[101] A further study investigated the effect of a knock-in mutation in mice to selectively inactivate the proofreading exonuclease activity of the enzyme while preserving its normal polymerase activity.[104] This study demonstrated that the

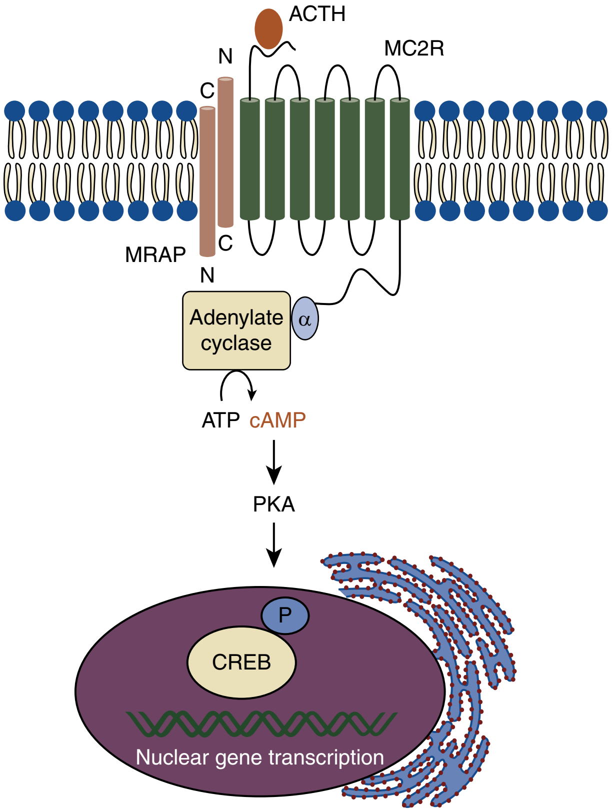


Figure 86.3 **Adrenocorticotropic hormone (ACTH) signaling in adrenocortical cells.** Ligand binding results in generation of increased levels of cAMP, which activates protein kinase A (PKA) and subsequently phosphorylates cAMP response element binding protein (CREB). CREB is directly involved in the transcriptional regulation of steroidogenic enzymes.

impaired enzyme resulted in a strong mutator phenotype with a high incidence of neoplasms.

KEY POINTS

- Familial glucocorticoid deficiency (FGD) is characterized by the lack of response to adrenocorticotropic hormone (ACTH) and results in glucocorticoid deficiency but unaffected plasma renin and aldosterone levels.
- Mutations in *MC2R* and *MRAP* cause FGD through the disruption of ACTH signaling and consequential inactivation of the steroidogenic cascade.
- *NNT* and *TXNRD2* mutations cause FGD through uncontrolled oxidative stress in the cells of the zona fasciculata, which disrupt steroidogenesis.

Familial Glucocorticoid Deficiency-Like (FGD-Like)

FGD, a hereditary unresponsiveness to ACTH, is characterized by glucocorticoid deficiency in a patient who retains normal plasma renin and aldosterone levels. It can be caused by several genes, and disease phenotypes are therefore often variable.

Melanocortin Type 2 Receptor (*MC2R*). *MC2R* encodes a G protein–coupled receptor that belongs to the melanocortin receptor family. Binding of its ligand ACTH triggers the activation of adenylate cyclase, which increases intracellular cAMP, subsequently activating PKA and the glucocorticoid steroidogenic cascade (Fig. 86.3). Mutations in *MC2R* account for the largest number of FGD cases, and the majority are nonconservative missense mutations that culminate in a failure to traffic the receptor to the cell surface, resulting in 20% to 100% reduced membrane expression in an *in vitro* study.[105,106] *MC2R* mutations were first associated with FGD via a candidate gene sequencing approach of a family with the condition in 1993.[107] More than 50 have now been described, and, interestingly, regional variants exist in *MC2R*; for example, p.S74I is the most common in the United Kingdom, while c.560delT is a recurrent mutation in Turkey.[108] Mutations in *MC2R* therefore inhibit the initiation of steroidogenesis and disrupt the negative feedback loop, resulting in high ACTH plasma levels. It is important to note that, while *MC2R* mutations most typically cause FGD, in severe LOF cases due to homozygous frameshift mutations, minor disturbances to the renin–angiotensin–aldosterone axis have been documented, therefore resulting in a form of PAI with transient mineralocorticoid deficiency.[109,110]

Patients with *MC2R* mutations typically present with symptoms of glucocorticoid deficiency, including jaundice, hypoglycemia, failure to thrive, sepsis, and hyperpigmentation.[111] Hyperpigmentation of the skin is due to supraphysiologic levels of ACTH causing overactivation of melanocortin 1 receptor (MC1R), culminating in the production of excess melanin. The age of onset is highly variable (range 0–16 years; median 2 years), with some patients presenting in later childhood with mild disease.[106] Patients present with a mean height standard deviation score of +1.76 ± 1.52,[106] believed to be due to increased ACTH-level activation of melanocortin receptors expressed in bone and cartilaginous growth plates.[112] This is supported by evidence from adrenalectomized mice showing that the melanocortin system plays a role in linear growth.[112] Interestingly, however, *Mc2r* knockout mice do not exhibit any significant difference in body length compared with wild-type mice.[113] Clinical observations indicate that the advanced growth rate in humans can be normalized with glucocorticoid replacement therapy.[114,115]

Mc2r Accessory Protein (*MRAP*). *MRAP* encodes a single transmembrane protein that acts as an essential cofactor in the trafficking of MC2R from the endoplasmic reticulum to the cell membrane to enable its functional expression.[116] It exists as a homodimer with a unique antiparallel dual topology conformation essential for its function.[117,118] Mutations in *MRAP* therefore prevent trafficking of MC2R to the plasma membrane and consequently inhibit the initiation of steroidogenesis through ACTH signaling. Mutations in the gene were first associated with FGD following linkage studies and homozygosity mapping using single nucleotide polymorphism array genotyping of a consanguineous family.[116] Unlike the majority of *MC2R* mutations, which are missense and may result in some residual function, *MRAP* mutations are typically splice-site or nonsense mutations.[106] These are predicted to result in a truncated protein that lacks the highly conserved transmembrane domain. This therefore inhibits its function of trafficking MC2R from the endoplasmic reticulum and results in degradation of the receptor.[105,116,119]

Patients with *MRAP* mutations present with the same symptoms as those with *MC2R* mutations. However, due to a complete breakdown of ACTH signaling, the disease can often present with earlier onset. One study showed that the median age of onset for *MRAP*-induced FGD was 0.08 years (compared with 2 years for *MC2R* mutations), and the range was significantly reduced, at 0 to 1.6 years.[106] Considering the earlier onset and consequential earlier glucocorticoid replacement in these patients, the taller stature seen in *MC2R* patients is largely absent in the *MRAP* patient cohort.[120] There are, however, exceptions to the rule, and sequencing of two affected families revealed homozygous missense mutations c.175T>G (pY59D) and c.76T>C (p.V26A), both of which presented with a later onset and subtler phenotype.[121] Beyond the trafficking of MC2R, MRAP has been shown to interact

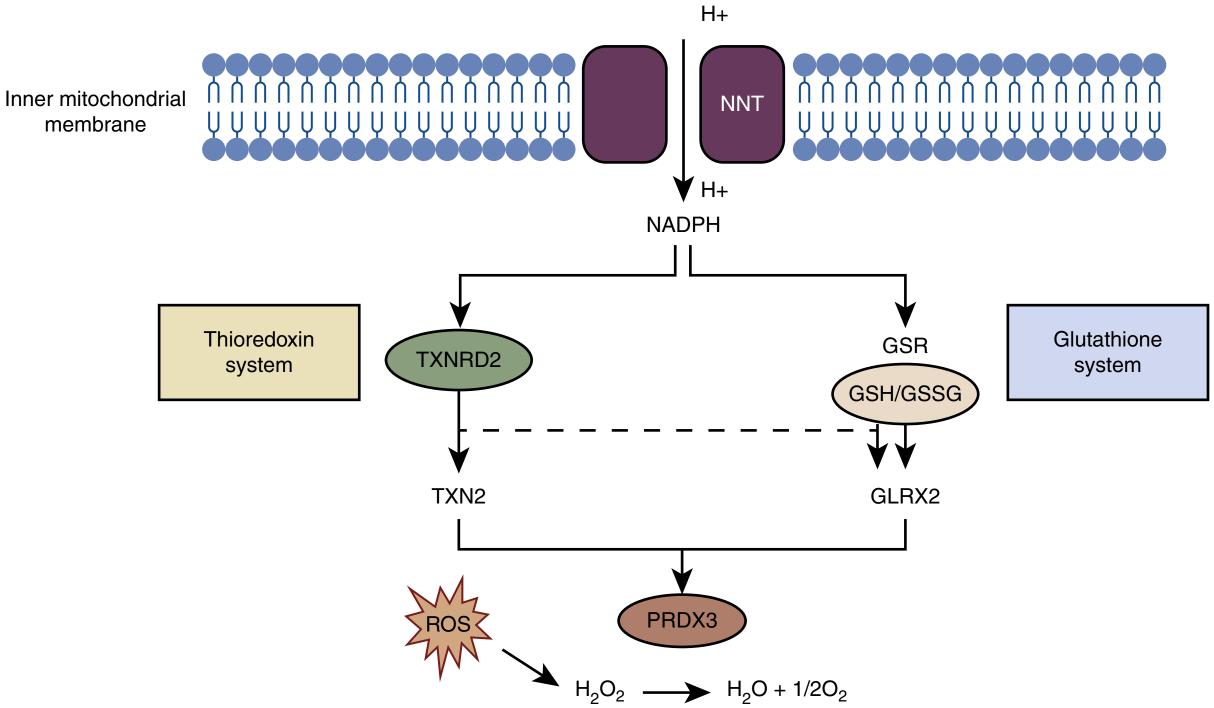


Figure 86.4 **The thioredoxin and glutathione redox homeostatic pathways.** NADPH is the electron donor that drives the thioredoxin and glutathione pathways, with NNT being the primary regenerator of NADPH levels. TXNRD2 reduces both TXN2 and GLRX2 (glutathione reductase and GSH play a further role in reduction of GLRX2), which subsequently reduces PRDX3. PRDX3, a highly conserved thioredoxin-dependent peroxide reductase, protects the cellular environment from oxidative stress due to its antioxidant properties.

with each of the melanocortin receptors, and it is postulated that its facilitation of MC4R expression could account for the obesity observed in one patient.[122,123] This remains speculative, however, and a study comparing the phenotypic differences between *MC2R* and *MRAP* patients concluded that there was no evidence that MRAP deficiency impacted any other physiological function beyond that seen with a defective MC2R.[106]

A recent study showed that *Mrap*$^{-/-}$ mice have almost complete penetrance of neonatal lethality (322 out of 325 mice) due to incomplete lung development, although this phenotype was rescuable by administration of corticosterone to pregnant dams. Mice that survived weaning showed the characteristic symptoms of MRAP deficiency: isolated glucocorticoid deficiency with normal mineralocorticoid function. Analysis of adrenal histology of knockout mice showed altered zonation, a thicker capsule, and lower mRNA expression of steroidogenic genes.[124]

Nicotinamide Nucleotide Transhydrogenase (*NNT*). NNT is the major enzymatic source of NADPH in the mitochondria, contributing 45% of the total supply.[125] Existing as a dimer, it is expressed ubiquitously throughout the body, where it spans the IMM orchestrating H^+ movement and supplying NADPH for the detoxification of reactive oxygen species (ROS) through the glutathione and thioredoxin pathways (Fig. 86.4). This role in redox homeostasis and removal of ROS contributes to limiting oxidative stress.

NNT was initially linked to PAI through single nucleotide polymorphism genotyping of consanguineous families with cases of isolated glucocorticoid deficiency, followed by targeted exome sequencing that identified many causal variants in the gene in different families.[126] This study documented 15 mutations in different kindreds that presented with a classical FGD phenotype, indistinguishable from mutations in *MC2R* and *MRAP*, with patients presenting with hypocortisolemia and elevated ACTH with normal renin and aldosterone levels. Furthermore, this study demonstrated that *NNT* mutations cause adrenal dysfunction in humans, primarily affecting cells of the zona fasciculata responsible for cortisol production, and when NNT expression was stably knocked down in adrenocortical carcinoma cells *in vitro*, increased mitochondrial ROS generation with lowered reduced glutathione to oxidized glutathione was observed.[126]

A second study reported that FGD patients with *NNT* mutations present at a median age of 12 months, with a range of 3 days to 39 months, and 53% display hyperpigmentation.[127] In a cohort of 23 patients, four presented with coexisting mineralocorticoid deficiency and showed evidence of hyponatremia, hyperkalemia, and Addisonian crises.

Further studies have added comorbidities to the classical FGD symptoms, with a cardiac phenotype being linked to *NNT* mutations. Heterozygous LOF *NNT* mutation patients have been linked to left ventricular noncompaction,[128] while a homozygous mutation (p.R379*) identified in a single patient resulted in features of glucocorticoid insufficiency with superimposed progressive left ventricular hypertrophy.[129] Furthermore, testicular adrenal rest tumors have also been documented with *NNT* mutations, similar to CAH patients.[129,130] These extra comorbidities make early screening in affected families essential.

The mouse strain C57BL/6J carries an intrinsic inactivating mutation in *Nnt* that results in an untranslated protein. Analysis of the adrenal glands of 3-month-old mice of this strain shows slightly disorganized zona fasciculata with higher levels of apoptosis than the C57BL/6N strain without this mutation.[126] A further study reported reduced steroidogenic capacity (50% reduction in corticosterone levels, perhaps due to reduction in *Cyp11a1/Cyp11b1* expression) of the adrenals due to the inability of other antioxidant enzymes to compensate for the loss of NNT, leading to oxidative stress.[131] Interestingly, C57BL/6J mice do not exhibit a cardiac phenotype, despite lacking NNT.[132]

Thioredoxin Reductase 2 (*TXNRD2*). TXNRD2 is a dimeric NADPH-dependent flavin adenine dinucleotide-containing enzyme that reduces thioredoxin 2 and other substrates.[133] It achieves this by utilizing electrons supplied by NNT (Fig. 86.4). The role of this protein, along with the glutathione system, is to eliminate ROS, which are generated in high quantities by the steroidogenic pathway and impede adrenocortical cellular function. The final step of cortisol production, performed by CYP11B1, accounts for approximately 40% of total electron flow from NADPH to ROS during steroidogenesis.[134] This may explain the susceptibility of the zona fasciculata to oxidative stress and why *NNT* and *TXNRD2* patients develop primarily glucocorticoid deficiency.

Knockdown of *TXNRD2* in adrenocortical carcinoma cells results in an increase in mitochondrial ROS levels, as it is unable to keep PRDX3 in its reduced form. This is only partially compensated for by the glutathione system.[135] Mutations in *TXNRD2* therefore can lead to oxidative stress-induced damage to mitochondria and inhibit steroidogenesis.

Thus far only one mutation of *TXNRD2* has been linked to PAI, a stop-gain mutation (p.Y447X) discovered in a Kashmiri kindred with seven affected individuals.[135] These patients presented with a wide spectrum of disease with later onset in a number of family members. All clinically affected family members demonstrated a poor cortisol response to ACTH with normal mineralocorticoid levels, and only one patient presented with a comorbidity (cardiac failure).

Txnrd2$^{-/-}$ mice exhibit embryonic lethality due to severe growth restriction, cardiac anomalies, and hematopoietic dysregulation; however, no adrenal phenotype was noted.[136] Electron microscopy of neonatal hearts showed severe swelling and destruction of mitochondrial cristae in cardiomyocytes, highlighting the damage that can be caused to mitochondria by loss of Txnrd2. This is supported by a further study that showed mitochondria isolated from *Txnrd2* knockout cardiomyocytes were unable to utilize oxygen efficiently, and consequently generated high levels of ROS.[137] Interestingly, two separate heterozygous mutations (p.G375R and p.A59T) have been identified in individuals with dilated cardiomyopathy, supporting a link between *TXNRD2* and heart failure in humans.[138]

Aladin Wd Repeat Nucleoporin (*AAAS*). Triple A syndrome (Allgrove syndrome) is a rare autosomal recessive disease caused by mutations in *AAAS*, characterized by alacrima, achalasia, and adrenal failure. It is associated with a range of progressive motor, sensory, and autonomic neurological defects.[139,140] *AAAS* encodes the nuclear pore complex protein alacrima-achalasia-adrenal insufficiency neurological disorder (ALADIN), which is expressed ubiquitously with predominance in the adrenal and central nervous system and was discovered to cause triple A syndrome in 2001.[140,141] The clinical phenotype and age of onset of the disease are highly variable, even with members of the same family, with associated features including pupil and cranial nerve abnormalities, optic atrophy, and hypoglycemic seizures.[142]

Knockdown of *AAAS* in adrenocortical carcinoma cells results in increased susceptibility to oxidative stress, along with alteration of redox homeostasis leading to apoptosis.[143,144] Furthermore, loss of AAAS has been shown to cause reduced expression of key steroidogenic enzymes STAR, mitochondrial cytochrome P450 11B1, steroid 17-alpha-hydroxylase/17,20 lyase, and steroid 21-hydroxylase, which results in a significant reduction in cortisol production in human cell lines.[143,144] Improvements in both cell viability and cortisol production were demonstrated by treatment with N-acetylcysteine, which replenishes stores of reduced glutathione and acts as a direct scavenger of free radicals.[143] This highlights the importance of the glutathione antioxidant pathway and promotes N-acetylcysteine as a possible therapeutic adjunct for triple A syndrome. Interestingly, while *Aaas* knockout mice show female infertility, they fail to develop a phenotype similar to human triple A syndrome.[145]

Mini-Chromosome Maintenance Complex Component 4 (*MCM4*). MCM4 is a protein essential for the initiation of eukaryotic genome replication. It possesses intrinsic DNA helicase activity and forms part of a hexameric complex with other MCMs. PAI appears at a higher prevalence in the Irish Traveller community, a genetically isolated population with higher levels of consanguinity than in the wider population (1 in 2506 compared with 1 in 201,898).[146] Homozygosity mapping of a cohort of Irish Travellers who were *MC2R*-, *MRAP*-, and *STAR*-mutation negative highlighted a single variant in *MCM4* (c.71-1insG) that is predicted to produce a truncated protein (p.Pro24Argfs*4).[147] This variant is unique to the Irish Traveller community and is associated with adrenal failure, growth retardation, and natural killer cell deficiency (also known as immunodeficiency 54).[147,148]

MCM4 patients present with a milder adrenal phenotype than other FGD syndromes, with onset of hypocortisolemia in childhood, sometimes following a period of normal adrenal function.[147] This is possibly explained by the presence of a shorter isoform of the protein containing all necessary functional domains and partially compensating for the loss of the full-length protein.[148] Due to its role in genome replication, *MCM4* patients suffer from chromosome fragility, increased susceptibility to viral infections, and an increased risk of developing neoplasms.[148] Mice carrying a hypomorphic allele of *Mcm4* present with capsular/subcapsular hyperplasia with nonsteroidogenic (*Cyp11b1*- and *Cyp11b2*-negative) spindle-like cells invading the cortex. This consequently reduces the number of steroidogenic cells in the zona fasciculata and is likely to reduce glucocorticoid output of the gland.[147,149]

KEY POINTS

- Primary adrenal insufficiency can also be caused by genes that do not fall into any of the above categories.
- Mechanisms include accumulation of antisteroidogenic sphingolipids (*SGPL1*), autoantibodies against steroidogenic enzymes (*AIRE*), and accumulation of toxic very long chain fatty acids (*ABCD1*, *PEX1,* and *PEX6*).

Other Causes Of Syndromic Adrenal Insufficiency

Sphingosine-1-Phosphate Lyase 1 (*SGPL1*). SGPL1 performs the final degradative step in sphingolipid metabolism. A 5’-phosphate-dependent aldehyde-lyase, it is localized to the endoplasmic reticulum, where it cleaves sphingosine 1-phosphate (S1P) and degrades it to phosphoethanolamine and hexadecanal. Loss of the enzyme can result in an accumulation of sphingolipids and proapoptotic ceramides, a feature of sphingolipidoses (Gaucher disease and Niemann–Pick disease). Conversely, in healthy conditions, its degradation of S1P makes it the major modulator of S1P signaling, which has been linked to pathways involved in angiogenesis, immune cell trafficking, stem cell differentiation, and programmed cell death.[150]

Sphingolipid intermediates including ceramide, sphingosine, and S1P are known modulators of the steroidogenic pathway.[151] In *SGPL1* patient-derived fibroblasts, elevated levels of sphingosine and ceramides result in a reduced level of cortisol production.[152] Sphingosine achieves this by acting as an endogenous antagonist of SF1, maintaining it in an inactive conformation.[153] Conversely, S1P increases cortisol secretion through promoting the expression of multiple genes involved in steroidogenesis, such as *CYP17A1* and *STAR*.[151,154] Mass spectroscopic analysis of sphingolipid intermediates in the plasma of an *SGPL1* patient and heterozygous parents revealed elevated levels of ceramide and sphingosine, and to a lesser extent S1P.[155] A further study observed a similar increase in ceramide, compared with S1P, in conditioned media from patient fibroblasts.[156] This accumulation of antisteroidogenic intermediates could explain the reduced cortisol expression seen in *SGPL1* patients.

Patients with *SGPL1* mutations are very rare, with most cases being lethal *in utero*. Those who do survive, present in infancy with multiple comorbidities, often including isolated glucocorticoid deficiency, (although mineralocorticoid deficiency has also been reported), and steroid-resistant nephrotic syndrome.[155,156] Other comorbidities include primary hypothyroidism, ichthyosis, neurodevelopmental delay, primary gonadal failure, lymphopenia, dyslipidemia, and axonal degeneration.[155-157] Cytosolic accumulations of sphingolipid intermediates, along with mitochondrial dysfunction, could be the underlying pathology behind these manifestations.[152,155]

Sgpl1$^{-/-}$ mice are born at normal Mendelian ratios, but most die within the first 2 weeks of life.[155] In mice that survive weaning, cortical zonation is compromised, with less definition between the zona glomerulosa and zona fasciculata, and with cells of the zona fasciculata appearing smaller, with fewer lipid droplets. The adrenals of the knockout mice showed lower expression of *Cyp11a1*, which supported the disruption of steroidogenesis that was observed.[155]

Autoimmune Regulator (*AIRE*). AIRE is a nuclear protein that regulates self-tolerance of T-cells, with highest expression in the thymus, and is also expressed to a lesser extent in peripheral lymphoid tissues.[158,159] Mutations in the gene cause autoimmune-poly-endocrinopathy-candidiasis–ectodermal-dystrophy syndrome (APECED) (also known as autoimmune polyendocrine syndrome), a rare monogenic autosomal recessive disease that is most common in Finnish and Iranian Jewish populations.[160] The classic criteria for diagnosis of APECED are the presence of any two of the following: chronic mucocutaneous candidiasis, chronic hypoparathyroidism, or PAI. However, the latest criteria are based on genetic and immunological features of the syndrome.[160] Dysfunctional AIRE results in the production of autoantibodies against self-antigens, including mitochondrial cholesterol side-chain cleavage enzyme, steroid 17-alpha-hydroxylase/17,20 lyase, and steroid 21-hydroxylase inhibiting steroidogenesis in the adrenals and gonads.[161] Further studies have found that almost all patients possess antibodies to interferons and interleukins, which lead to a severe autoimmune disease affecting multiple endocrine organs and other tissues.[162,163]

Initially identified in 1997 to cause APECED, to date, approximately 115 mutations have been found to affect *AIRE,* with most of these variants being either nonsense or frameshift mutations resulting in a truncated protein, or a single amino acid missense variant.[160,164,165] Due to the nature of autoimmune disease, AIRE deficiency presents with many different manifestations in a highly patient-variable manner, making the disease unpredictable. These manifestations can include mucocutaneous candidiasis, chronic hypoparathyroidism, and PAI, as well as ectodermal dystrophy, hypogonadism, vitiligo, alopecia, malabsorption, and cataracts. The most common causes of death from APECED include hypocalcemia and adrenal crisis, oral and esophageal squamous cell carcinoma, and acute hepatitis.[166] *Aire*$^{-/-}$ mice develop normally in terms of weight, size, and maturation but present with autoimmune features such as exaggerated immune response, multiorgan lymphocytic infiltration, circulating autoantibodies, and infertility, as well as undetectable adrenal glands in some mice.[167]

ATP Binding Cassette Subfamily D Member 1 (*ABCD1*). *ABCD1*, located on chromosome Xq28, encodes a protein that transports very long chain fatty acids (VLCFAs) into the peroxisome, a small membrane-enclosed organelle that contains enzyme pathways for lipid metabolism and detoxification processes.[168] Mutations in *ABCD1,* commonly missense, nonsense, or splicing-defect, result in failure to traffic VLCFAs into the peroxisome and their subsequent accumulation in plasma and tissues. VLCFAs are toxic to adrenocortical cells and also alter the structure of the cell membrane, suppressing the availability of MC2R and consequently inhibiting steroidogenesis.[169,170] Furthermore, VLCFAs are also toxic to myelin-producing oligodendrocytes, where they cause a disturbance in calcium homeostasis and mitochondrial dysfunction.[171] This toxicity caused by VLCFAs through ABCD1 dysfunction results in the genetic metabolic disorder adrenoleukodystrophy (ALD).[172]

ALD is an X-linked inheritance peroxisomal disease with two main phenotypes: cerebral demyelinating form and adrenomyeloneuropathy (AMN).[173] Sufferers of the cerebral demyelinating form of the disease appear asymptomatic at birth but start to present with cognitive dysfunction, through declining school performance and behavioral problems, between 4 and 8 years of age. The disease then progresses rapidly, with central demyelination resulting in seizures, dementia, cortical blindness, coma, and death, most frequently within 2 years of presentation and before the age of puberty.[174] The other form of the disease, AMN, presents in men in their 20s or above and is characterized by slow progressive mixed motor and sensory peripheral neuropathy, with 10% of cases developing cerebral demyelination.[175] Although all 46, XY patients with ALD are eventually affected by AMN, the age of onset and rate of progression are highly variable.[173] Virtually all 46, XY patients will develop adrenocortical insufficiency, with approximately 80% doing so before adulthood.[176]

Although ALD is deemed an X-linked disease, adrenal insufficiency can be present in 46, XX heterozygous carriers of *ABCD1* mutations.[177] In these carriers, 63% have myelopathy, whereas 57% have peripheral neuropathy, and this increases to 80% for patients over the age of 60 years.[178] This could be due to X-inactivation of the nonmutant allele, but this explanation remains contested and is not supported by this particular study.[178] One study found that *Abcd1*$^{-/-}$ mice appear normal for the first 12 months of life and show an increase in VLCFA but no demyelination in the brain, spinal cord, or peripheral nerves, indicating that VLCFAs alone may not be sufficient to demyelinate the nervous system.[179] A further study showed no statistical difference between wildtype and *Abcd1*$^{-/-}$ mice in corticosterone production or oxidative stress, while knockout of *Abcd2* resulted in the acceleration of ceroid production (lipogenic pigmentation) in the adrenal gland, the end product of oxidative damage.[180]

Peroxisomal Biogenesis Factors (*PEX* Genes). Peroxisomal diseases are divided into two main groups: disorders with a deficient enzyme or transporter (ALD), and peroxisome biogenesis disorders with multiple metabolic abnormalities, such as Zellweger spectrum disorders (ZSDs). ZSDs are autosomal recessive diseases with a range of severity due to mutations in *PEX* genes resulting in the reduction or absence of functional peroxisomes.[181] The phenotype of ZSDs is variable, depending on the *PEX* gene mutated, but the most common features are liver dysfunction, developmental delay, adrenal insufficiency, and hearing/vision impairment.[182] Adrenal insufficiency is caused by mutations in *PEX1* and *PEX6* and could be due to accumulation of VLCFAs; these are toxic to adrenocortical cells and suppress steroidogenesis, causing adrenal insufficiency in approximately 30% of patients.[169,170,183,184] Furthermore, VLCFAs are also toxic to myelin-producing oligodendrocytes, resulting in hypomyelination and neurological defects.[171]

Of the 15 *PEX* genes, mutations in *PEX1* are the most common cause of Zellweger syndrome, the most severe form of ZSDs.[185] Variants within the gene can dictate the severity of the disease. p.G843D leads to a less severe form with long survival, whereas p.Pro970* results in a nonfunctional protein, a more severe form of the disease, and shorter survival.[186-188] Indeed, nonsense mutations have been shown to result in more severe cases than missense variants.[189]

Further symptoms can include epileptic seizures, craniofacial dysmorphism, leukodystrophy, and cerebellar ataxia.[190] Zellweger disease, the most severe form, can itself be divided into three categories—neonatal-infantile, childhood, and adolescent-adult—with severity of disease decreasing as age of presentation increases. There is no curative therapy for patients, and therefore treatment is supportive and symptomatic. Hypomorphic *Pex1*-G843D mice recapitulate many of the human traits, such as adrenal insufficiency, growth retardation, and fatty liver with cholestasis.[186]

Lysosomal Acid Lipase (*LIPA*). *LIPA* encodes an enzyme that hydrolyses cholesteryl esters (CE) and triglycerides (TG) in lysosomes, producing free cholesterol and free fatty acids (FFA).[191] Free cholesterol and FFAs function as key structural components of cell membranes, energy sources, and signaling molecules. Mutations in *LIPA* result in accumulation of CE and TG all over the body, particularly the liver, spleen, lymph nodes, bone marrow, and macrophages.[191] This manifests as an autosomal recessive lysosomal storage disease with two phenotypes, presentation of which is dependent on the quantity of functional enzyme the patient possesses. Wolman disease is the early-onset form of LIPA deficiency; it presents in newborns who have no or almost no enzyme activity.[192] Sufferers of this disease rarely survive beyond infancy and present with hepatomegaly, calcification of the adrenal glands, malabsorption, and hepatic failure.[191,192] Patients who retain partial LIPA function (up to 12%) are diagnosed with CE storage disease (CESD), present later in life, if at all, and have variable clinical phenotypes.[193] Patients can present with hepatomegaly, abdominal distension, and dyslipidemia, as well as, in severe cases, multiorgan failure and death .[191] Interestingly, however, large adrenal calcium deposits are rare, although fine punctate calcifications are often detected. CESD is rare and is often misdiagnosed as nonalcoholic fatty liver disease or other forms of hypercholesterolemia.[194]

Mutations in *LIPA* have also been linked to coronary heart disease through genome-wide association studies.[195,196] This could be due to accumulation of TG and CE leading to foam cell formation and consequently atherosclerosis.[197] High levels of oxidised LDL, a result of LIPA deficiency, have been shown to result in increased levels of intracellular calcium in cultured adrenal cells, which could explain the adrenal calcification observed in Wolman disease.[198] This calcification, combined with the reduced abundance of free cholesterol, could inhibit steroidogenesis and result in a form of adrenal insufficiency in these patients.[193] LIPA deficiency can be treated by bone marrow transplant, to normalize cholesterol levels, and enzyme replacement therapy.[199,200]

Lipa$^{-/-}$ mice are born at Mendelian ratios and appear normal at birth. However, within 21 days, TG and CE accumulations appear in several organs at levels up to 30-fold higher than those found in normal mice, and the liver appears one and a half to two times larger. Adrenal glands present with an expanded adrenal cortex due to lipid accumulations. Heterozygotes have approximately 50% enzyme functionality and do not show lipid accumulation.[201]

CONCLUSION

The number of genes identified to cause PAI continues to increase, both those involved in isolated glucocorticoid deficiency and those involved in syndromic forms. With many dysfunctional genes causing similar and overlapping phenotypes, it is of vital importance to attain a genetic diagnosis for patient treatment. This process has been eased significantly with the dawn of next-generation sequencing, which allows sequencing of all genes concomitantly rather than taking a candidate gene approach. Many PAI patients remain without a genetic diagnosis, and researchers are working to identify these causative genes.

FURTHER RESOURCES

The following resources contain more information about the genes, mutations, and mouse models discussed in this chapter.

ClinVar: Information about genomic variation and its relationship to human health; https://www.ncbi.nlm.nih.gov/clinvar/

Human Gene Mutation Database: A database for identifying disease-causing mutations; https://mastermind.genomenon.com/

International Mouse Phenotyping Consortium: An international scientific endeavor to create and characterize the phenotype of 20,000 knockout mouse strains; https://www.mousephenotype.org/

Leiden Open Variation Database: An open-source DNA variation database; www.lovd.nl

Mouse Genome Informatics: International database resource for the laboratory mouse; http://www.informatics.jax.org/

Online Mendelian Inheritance in Man: An online catalogue of human genes and genetic disorders; https://omim.org/

REFERENCES

1. Gallo-Payet N, Payet MD. Mechanism of action of ACTH: beyond cAMP. *Microsc Res Tech*. 2003;61:275–287.
2. Miller WL. Mechanisms in endocrinology: rare defects in adrenal steroidogenesis. *Eur J Endocrinol*. 2018;179:R125–R141.
3. Spiegel RJ, Vigersky RA, Oliff AI, et al. Adrenal suppression after short-term corticosteroid therapy. *Lancet*. 1979;1:630–633.
4. Olefsky JM. Effect of dexamethasone on insulin binding, glucose transport, and glucose oxidation of isolated rat adipocytes. *J Clin Invest*. 1975;56:1499–1508.
5. Reidenberg MM, Ohler EA, Sevy RW, et al. Hemodynamic changes in adrenalectomized dogs. *Endocrinology*. 1963;72:918–923.
6. Clarke PW, Cleghorn RA, Ferguson JK, et al. Factors concerned in the circulatory failure of adrenal insufficiency. *J Clin Invest*. 1947;26:359–363.
7. Lefer AM, Verrier RL, Carson WW. Cardiac performance in experimental adrenal insufficiency in cats. *Circ Res*. 1968;22:817–827.
8. Livingston JN, Lockwood DH. Effect of glucocorticoids on the glucose transport system of isolated fat cells. *J Biol Chem*. 1975;250:8353–8360.
9. Nehme A, Zouein FA, Zayeri ZD, et al. An update on the tissue renin angiotensin system and its role in physiology and pathology. *J Cardiovasc Dev Dis*. 2019;6:75–91.
10. Marver D, Kokko JP. Renal target sites and the mechanism of action of aldosterone. *Miner Electrolyte Metab*. 1983;9:1–18.
11. White PC, New MI, Dupont B. Congenital adrenal hyperplasia. (1). *N Engl J Med*. 1987;316:1519–1524.
12. New MI. Diagnosis and management of congenital adrenal hyperplasia. *Annu Rev Med*. 1998;49:311–328.
13. Merke DP, Bornstein SR. Congenital adrenal hyperplasia. *Lancet*. 2005;365:2125–2136.
14. Maharaj A, Maudhoo A, Chan LF, et al. Isolated glucocorticoid deficiency: genetic causes and animal models. *J Steroid Biochem Mol Biol*. 2019;189:73–80.
15. Morel Y, Mébarki F, Rhéaume E, et al. Structure-function relationships of 3 beta-hydroxysteroid dehydrogenase: contribution made by the molecular genetics of 3 beta-hydroxysteroid dehydrogenase deficiency. *Steroids*. 1997;62:176–184.
16. Speiser PW, White PC. Congenital adrenal hyperplasia. *N Engl J Med*. 2003;349:776–788.

17. Speiser PW, Azziz R, Baskin LS, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95:4133–4160.
18. Parsa AA, New MI. Steroid 21-hydroxylase deficiency in congenital adrenal hyperplasia. *J Steroid Biochem Mol Biol*. 2017;165:2–11.
19. Lin-Su K, Harbison MD, Lekarev O, et al. Final adult height in children with congenital adrenal hyperplasia treated with growth hormone. *J Clin Endocrinol Metab*. 2011;96:1710–1717.
20. Quintos JB, Vogiatzi MG, Harbison MD, et al. Growth hormone therapy alone or in combination with gonadotropin-releasing hormone analog therapy to improve the height deficit in children with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2001;86:1511–1517.
21. Claahsen-van der Grinten HL, Dehzad F, Kamphuis-van Ulzen K, et al. Increased prevalence of testicular adrenal rest tumours during adolescence in congenital adrenal hyperplasia. *Horm Res Paediatr*. 2014;82:238–244.
22. Bidet M, Bellanné-Chantelot C, Galand-Portier MB, et al. Clinical and molecular characterization of a cohort of 161 unrelated women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency and 330 family members. *J Clin Endocrinol Metab*. 2009;94:1570–1578.
23. Shiroishi T, Sagai T, Natsuume-Sakai S, et al. Lethal deletion of the complement component C4 and steroid 21-hydroxylase genes in the mouse H-2 class III region, caused by meiotic recombination. *Proc Natl Acad Sci USA*. 1987;84:2819–2823.
24. Tajima T, Okada T, Ma XM, et al. Restoration of adrenal steroidogenesis by adenovirus-mediated transfer of human cytochromeP450 21-hydroxylase into the adrenal gland of21-hydroxylase-deficient mice. *Gene Ther*. 1999;6:1898–1903.
25. Miller WL. Disorders in the initial steps of steroid hormone synthesis. *J Steroid Biochem Mol Biol*. 2017;165:18–37.
26. Lin D, Sugawara T, Strauss 3rd JF, et al. Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science*. 1995;267:1828–1831.
27. Bose HS, Sugawara T, Strauss 3rd JF, et al. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. *N Engl J Med*. 1996;335:1870–1878.
28. Buonocore F, Achermann JC. Primary adrenal insufficiency: New genetic causes and their long-term consequences. *Clin Endocrinol (Oxf)*. 2020;92:11–20.
29. Baker BY, Lin L, Kim CJ, et al. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*. 2006;91:4781–4785.
30. Metherell LA, Naville D, Halaby G, et al. Nonclassic lipoid congenital adrenal hyperplasia masquerading as familial glucocorticoid deficiency. *J Clin Endocrinol Metab*. 2009;94:3865–3871.
31. Caron KM, Soo S-C, Wetsel WC, et al. Targeted disruption of the mouse gene encoding steroidogenic acute regulatory protein provides insights into congenital lipoid adrenalhyperplasia. *Proc Natl Acad Sci*. 1997;94:11540–11545.
32. Guran T, Buonocore F, Saka N, et al. Rare causes of primary adrenal insufficiency: genetic and clinical characterization of a large nationwide cohort. *J Clin Endocrinol Metab*. 2016;101:284–292.
33. Chan LF, Campbell DC, Novoselova TV, et al. Whole-exome sequencing in the differential diagnosis of primary adrenal insufficiency in children. *Front Endocrinol (Lausanne)*. 2015;6:113.
34. Tajima T, Fujieda K, Kouda N, et al. Heterozygous mutation in the cholesterol side chain cleavage enzyme (p450scc) gene in a patient with 46,XY sex reversal and adrenal insufficiency. *J Clin Endocrinol Metab*. 2001;86:3820–3825.
35. Parajes S, Kamrath C, Rose IT, et al. A novel entity of clinically isolated adrenal insufficiency caused by a partially inactivating mutation of the gene encoding for P450 side chain cleavage enzyme (CYP11A1). *J Clin Endocrinol Metab*. 2011;96:E1798–E1806.
36. Sahakitrungruang T, Tee MK, Blackett PR, et al. Partial defect in the cholesterol side-chain cleavage enzyme P450scc (CYP11A1) resembling nonclassic congenital lipoid adrenal hyperplasia. *J Clin Endocrinol Metab*. 2011;96:792–798.
37. Parajes S, Chan AO, But WM, et al. Delayed diagnosis of adrenal insufficiency in a patient with severe penoscrotal hypospadias due to two novel P450 side-change cleavage enzyme (CYP11A1) mutations (p.R360W; p.R405X). *Eur J Endocrinol*. 2012;167:881–885.
38. Tee MK, Abramsohn M, Loewenthal N, et al. Varied clinical presentations of seven patients with mutations in CYP11A1 encoding the cholesterol side-chain cleavage enzyme, P450scc. *J Clin Endocrinol Metab*. 2013;98:713–720.
39. Maharaj A, Buonocore F, Meimaridou E, et al. Predicted benign and synonymous variants in CYP11A1 cause primary adrenal insufficiency through missplicing. *J Endocr Soc*. 2019;3:201–221.
40. Goursaud C, Mallet D, Janin A, et al. Aberrant splicing is the pathogenicity mechanism of the p.Glu314Lys variant in CYP11A1 gene. *Front Endocrinol (Lausanne)*. 2018;9:491.
41. Hu MC, Hsu NC, El Hadj NB, et al. Steroid deficiency syndromes in mice with targeted disruption of Cyp11a1. *Mol Endocrinol*. 2002;16:1943–1950.
42. Arlt W, Willis DS, Wild SH, et al. Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 patients. *J Clin Endocrinol Metab*. 2010;95:5110–5121.
43. Al Alawi AM, Nordenström A, Falhammar H. Clinical perspectives in congenital adrenal hyperplasia due to 3β-hydroxysteroid dehydrogenase type 2 deficiency. *Endocrine*. 2019;63:407–421.
44. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390:2194–2210.
45. Rhéaume E, Simard J, Morel Y, et al. Congenital adrenal hyperplasia due to point mutations in the type II 3β–hydroxysteroid dehydrogenase gene. *Nat Genet*. 1992;1:239–245.
46. Araújo VG, Oliveira RS, Gameleira KP, et al. 3β-hydroxysteroid dehydrogenase type II deficiency on newborn screening test. *Arq Bras Endocrinol Metabol*. 2014;58:650–655.
47. Martin F, Perheentupa J, Adlercreutz H. Plasma and urinary androgens and oestrogens in a pubertal boy with 3 beta-hydroxysteroid dehydrogenase deficiency. *J Steroid Biochem*. 1980;13:197–201.
48. Morel Y, Roucher F, Plotton I, et al. 3B-Hydroxysteroid Dehydrogenase Deficiency. *Genet Steroid Disord*. 2014:99–110.
49. Bongiovanni AM. The adrenogenital syndrome with deficiency of 3β-hydroxysteroid dehydrogenase. *J Clin Invest*. 1962;41:2086–2092.
50. Auchus RJ. Steroid 17-hydroxylase and 17,20-lyase deficiencies, genetic and pharmacologic. *J Steroid Biochem Mol Biol*. 2017;165:71–78.
51. Kagimoto M, Winter JS, Kagimoto K, et al. Structural characterization of normal and mutant human steroid 17 alpha-hydroxylase genes: molecular basis of one example of combined 17 alpha-hydroxylase/17,20 lyase deficiency. *Mol Endocrinol*. 1988;2:564–570.
52. Van Den Akker EL, Koper JW, Boehmer AL, et al. Differential inhibition of 17alpha-hydroxylase and 17,20-lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. *J Clin Endocrinol Metab*. 2002;87:5714–5721.
53. Roger M, Merceron RE, Girard F, et al. Dexamethasone-suppressible hypercorticosteronism in two 46,XX subjects with ambiguous genitalia and ovarian cysts. Partial defect of 17 alpha-hydroxylase or 17-20-desmolase. *Horm Res*. 1982;16:23–31.
54. ten Kate-Booij MJ, Cobbaert C, Koper JW, et al. Deficiency of 17,20-lyase causing giant ovarian cysts in a girl and a female phenotype in her 46,XY sister: case report. *Hum Reprod*. 2004;19:456–459.
55. Simsek E, Ozdemir I, Lin L, et al. Isolated 17,20-lyase (desmolase) deficiency in a 46,XX female presenting with delayed puberty. *Fertil Steril*. 2005;83:1548–1551.
56. Bair SR, Mellon SH. Deletion of the mouse P450c17 gene causes early embryonic lethality. *Mol Cell Biol*. 2004;24:5383–5390.
57. Tomalik-Scharte D, Maiter D, Kirchheiner J, et al. Impaired hepatic drug and steroid metabolism in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *Eur J Endocrinol*. 2010;163:919–924.
58. Laue K, Pogoda HM, Daniel PB, et al. Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. *Am J Hum Genet*. 2011;89:595–606.

59. Flück CE, Tajima T, Pandey AV, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat Genet*. 2004;36:228–230.
60. Otto DM, Henderson CJ, Carrie D, et al. Identification of novel roles of the cytochrome p450 system in early embryogenesis: effects on vasculogenesis and retinoic acid homeostasis. *Mol Cell Biol*. 2003;23:6103–6116.
61. Arlt W, Walker EA, Draper N, et al. Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. *Lancet*. 2004;363:2128–2135.
62. Krone N, Reisch N, Idkowiak J, et al. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2012;97:E257–E267.
63. Shen AL, O'Leary KA, Kasper CB. Association of multiple developmental defects and embryonic lethality with loss of microsomal NADPH-cytochrome P450 oxidoreductase. *J Biol Chem*. 2002;277:6536–6541.
64. Wu L, Gu J, Cui H, et al. Transgenic mice with a hypomorphic NADPH-cytochrome P450 reductase gene: effects on development, reproduction, and microsomal cytochrome P450. *J Pharmacol Exp Ther*. 2005;312:35–43.
65. Suthiworachai C, Tammachote R, Srichomthong C, et al. Identification and functional analysis of six DAX1 mutations in patients with X-linked adrenal hypoplasia congenita. *J Endocr Soc*. 2019;3:171–180.
66. Ito M, Yu R, Jameson JL. DAX-1 inhibits SF-1-mediated transactivation via a carboxy-terminal domain that is deleted in adrenal hypoplasia congenita. *Mol Cell Biol*. 1997;17:1476–1483.
67. Münsterberg A, Lovell-Badge R. Expression of the mouse anti-müllerian hormone gene suggests a role in both male and female sexual differentiation. *Development*. 1991;113:613–624.
68. Jeffs B, Ito M, Yu RN, et al. Sertoli cell-specific rescue of fertility, but not testicular pathology, in Dax1 (Ahch)-deficient male mice. *Endocrinology*. 2001;142:2481–2488.
69. Muscatelli F, Strom TM, Walker AP, et al. Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature*. 1994;372:672–676.
70. Suntharalingham JP, Buonocore F, Duncan AJ, et al. DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract Res Clin Endocrinol Metab*. 2015;29:607–619.
71. Landau Z, Hanukoglu A, Sack J, et al. Clinical and genetic heterogeneity of congenital adrenal hypoplasia due to NR0B1 gene mutations. *Clin Endocrinol (Oxf)*. 2010;72:448–454.
72. Kyriakakis N, Shonibare T, Kyaw-Tun J, et al. Late-onset X-linked adrenal hypoplasia (DAX-1, NR0B1): two new adult-onset cases from a single center. *Pituitary*. 2017;20:585–593.
73. Tabarin A, Achermann JC, Recan D, et al. A novel mutation in DAX1 causes delayed-onset adrenal insufficiency and incomplete hypogonadotropic hypogonadism. *J Clin Invest*. 2000;105:321–328.
74. Mantovani G, Ozisik G, Achermann JC, et al. Hypogonadotropic hypogonadism as a presenting feature of late-onset X-linked adrenal hypoplasia congenita. *J Clin Endocrinol Metab*. 2002;87:44–48.
75. Xing Y, Achermann JC, Hammer GD. Chapter 2 - Adrenal Development. In: New MI, Lekarev O, Parsa A, et al., eds. *Genetic Steroid Disorders*. San Diego: Academic Press; 2014:5–27.
76. Scheys JO, Heaton JH, Hammer GD. Evidence of adrenal failure in aging Dax1-deficient mice. *Endocrinology*. 2011;152:3430–3439.
77. Mariniello K, Ruiz-Babot G, McGaugh EC, et al. Stem cells, self-renewal, and lineage commitment in the endocrine system. *Front Endocrinol*. 2019;772: Article 10.
78. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*. 1994;77:481–490.
79. Shen WH, Moore CC, Ikeda Y, et al. Nuclear receptor steroidogenic factor 1 regulates the müllerian inhibiting substance gene: a link to the sex determination cascade. *Cell*. 1994;77:651–661.
80. Achermann JC, Ito M, Ito M, et al. A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat Genet*. 1999;22:125–126.
81. Köhler B, Lin L, Mazen I, et al. The spectrum of phenotypes associated with mutations in steroidogenic factor 1 (SF-1, NR5A1, Ad4BP) includes severe penoscrotal hypospadias in 46,XY males without adrenal insufficiency. *Eur J Endocrinol*. 2009;161:237–242.
82. Achermann JC, Ozisik G, Ito M, et al. Gonadal determination and adrenal development are regulated by the orphan nuclear receptor steroidogenic factor-1, in a dose-dependent manner. *J Clin Endocrinol Metab*. 2002;87:1829–1833.
83. Kim Y, Capel B. Balancing the bipotential gonad between alternative organ fates: a new perspective on an old problem. *Dev Dyn*. 2006;235:2292–2300.
84. Bernard P, Harley VR. Wnt4 action in gonadal development and sex determination. *Int J Biochem Cell Biol*. 2007;39:31–43.
85. Biason-Lauber A, De Filippo G, Konrad D, et al. WNT4 deficiency--a clinical phenotype distinct from the classic Mayer-Rokitansky-Kuster-Hauser syndrome: a case report. *Hum Reprod*. 2007;22:224–229.
86. Mandel H, Shemer R, Borochowitz ZU, et al. SERKAL syndrome: an autosomal-recessive disorder caused by a loss-of-function mutation in WNT4. *Am J Hum Genet*. 2008;82:39–47.
87. Stark K, Vainio S, Vassileva G, et al. Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature*. 1994;372:679–683.
88. Treier M, Gleiberman AS, O'Connell SM, et al. Multistep signaling requirements for pituitary organogenesis in vivo. *Genes dev*. 1998;12:1691–1704.
89. Vainio S, Heikkilä M, Kispert A, et al. Female development in mammals is regulated by Wnt-4 signalling. *Nature*. 1999;397:405–409.
90. Heikkila¨ M, Peltoketo H, Leppäluoto J, et al. Wnt-4 deficiency alters mouse adrenal cortex function, reducing aldosterone production. *Endocrinology*. 2002;143:4358–4365.
91. Drelon C, Berthon A, Sahut-Barnola I, et al. PKA inhibits WNT signalling in adrenal cortex zonation and prevents malignant tumour development. *Nat Comm*. 2016;7:12751.
92. Vilain E, Le Merrer M, Lecointre C, et al. IMAGe, a new clinical association of intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies. *J Clin Endocrinol Metab*. 1999;84:4335–4340.
93. Arboleda VA, Lee H, Parnaik R, et al. Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. *Nat Genet*. 2012;44:788–792.
94. Borges KS, Arboleda VA, Vilain E. Mutations in the PCNA-binding site of CDKN1C inhibit cell proliferation by impairing the entry into S phase. *Cell Div*. 2015;10:2.
95. Kerns SL, Guevara-Aguirre J, Andrew S, et al. A novel variant in CDKN1C is associated with intrauterine growth restriction, short stature, and early-adulthood-onset diabetes. *J Clin Endocrinol Metab*. 2014;99:E2117–E2122.
96. Stampone E, Caldarelli I, Zullo A, et al. Genetic and epigenetic control of CDKN1C expression: importance in cell commitment and differentiation, tissue homeostasis and human diseases. *Int J Mol Sci*. 2018;19:32–55.
97. Zhang P, Liégeois NJ, Wong C, et al. Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome. *Nature*. 1997;387:151–158.
98. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet*. 2016;48:792–797.
99. Buonocore F, Kühnen P, Suntharalingham JP, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. *J Clin Invest*. 2017;127:1700–1713.
100. Topaz O, Indelman M, Chefetz I, et al. A deleterious mutation in SAMD9 causes normophosphatemic familial tumoral calcinosis. *Am J Hum Genet*. 2006;79:759–764.
101. Logan CV, Murray JE, Parry DA, et al. DNA polymerase epsilon deficiency causes image syndrome with variable immunodeficiency. *Am J Hum Genet*. 2018;103:1038–1044.
102. Pachlopnik Schmid J, Lemoine R, Nehme N, et al. Polymerase ε1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). *J Exp Med*. 2012;209:2323–2330.

103. Thiffault I, Saunders C, Jenkins J, et al. A patient with polymerase E1 deficiency (POLE1): clinical features and overlap with DNA breakage/instability syndromes. *BMC Med Genet*. 2015;16:31.
104. Albertson TM, Ogawa M, Bugni JM, et al. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci U S A*. 2009;106:17101–17104.
105. Chung TT, Webb TR, Chan LF, et al. The majority of adrenocorticotropin receptor (melanocortin 2 receptor) mutations found in familial glucocorticoid deficiency type 1 lead to defective trafficking of the receptor to the cell surface. *J Clin Endocrinol Metab*. 2008;93:4948–4954.
106. Chung T-TLL, Chan LF, Metherell LA, et al. Phenotypic characteristics of familial glucocorticoid deficiency (FGD) type 1 and 2. *Clin Endocrinol*. 2010;72:589–594.
107. Clark AJ, McLoughlin L, Grossman A. Familial glucocorticoid deficiency associated with point mutation in the adrenocorticotropin receptor. *Lancet*. 1993;341:461–462.
108. Guran T, Buonocore F, Saka N, et al. Rare causes of primary adrenal insufficiency: genetic and clinical characterization of a large nationwide cohort. *J Clin Endocrinol Metab*. 2016;101:284–292.
109. Chan LF, Metherell LA, Krude H, et al. Homozygous nonsense and frameshift mutations of the ACTH receptor in children with familial glucocorticoid deficiency (FGD) are not associated with long-term mineralocorticoid deficiency. *Clin Endocrinol (Oxf)*. 2009;71:171–175.
110. Lin L, Hindmarsh PC, Metherell LA, et al. Severe loss-of-function mutations in the adrenocorticotropin receptor (ACTHR, MC2R) can be found in patients diagnosed with salt-losing adrenal hypoplasia. *Clin Endocrinol (Oxf)*. 2007;66:205–210.
111. Clark AJ, Metherell LA, Cheetham ME, et al. Inherited ACTH insensitivity illuminates the mechanisms of ACTH action. *Trends Endocrinol Metab*. 2005;16:451–457.
112. Yeh JK, Evans JF, Niu QT, et al. A possible role for melanocortin peptides in longitudinal growth. *J Endocrinol*. 2006;191:677–686.
113. Chida D, Nakagawa S, Nagai S, et al. Melanocortin 2 receptor is required for adrenal gland development, steroidogenesis, and neonatal gluconeogenesis. *Proc Natl Acad Sci USA*. 2007;104:18205–18210.
114. Imamine H, Mizuno H, Sugiyama Y, et al. Possible relationship between elevated plasma ACTH and tall stature in familial glucocorticoid deficiency. *Tohoku J Exp Med*. 2005;205:123–131.
115. Elias LL, Huebner A, Metherell LA, et al. Tall stature in familial glucocorticoid deficiency. *Clin Endocrinol (Oxf)*. 2000;53:423–430.
116. Metherell LA, Chapple JP, Cooray S, et al. Mutations in MRAP, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2. *Nat Genet*. 2005;37:166–170.
117. Cooray SN, Almiro Do Vale I, Leung KY, et al. The melanocortin 2 receptor accessory protein exists as a homodimer and is essential for the function of the melanocortin 2 receptor in the mouse y1 cell line. *Endocrinology*. 2008;149:1935–1941.
118. Sebag JA, Hinkle PM. Melanocortin-2 receptor accessory protein MRAP forms antiparallel homodimers. *Proc Natl Acad Sci USA*. 2007;104:20244–20249.
119. Janovick JA, Maya-Nunez G, Conn PM. Rescue of hypogonadotropic hypogonadism-causing and manufactured GnRH receptor mutants by a specific protein-folding template: misrouted proteins as a novel disease etiology and therapeutic target. *J Clin Endocrinol Metab*. 2002;87:3255–3262.
120. Novoselova TV, Chan LF, Clark AJL. Pathophysiology of melanocortin receptors and their accessory proteins. *Best Pract Res Clin Endocrinol Metab*. 2018;32:93–106.
121. Hughes CR, Chung TT, Habeb AM, et al. Missense mutations in the melanocortin 2 receptor accessory protein that lead to late onset familial glucocorticoid deficiency type 2. *J Clin Endocrinol Metab*. 2010;95:3497–3501.
122. Rumié H, Metherell LA, Clark AJ, et al. Clinical and biological phenotype of a patient with familial glucocorticoid deficiency type 2 caused by a mutation of melanocortin 2 receptor accessory protein. *Eur J Endocrinol*. 2007;157:539–542.
123. Chan LF, Webb TR, Chung TT, et al. MRAP and MRAP2 are bidirectional regulators of the melanocortin receptor family. *Proc Natl Acad Sci USA*. 2009;106:6146–6151.
124. Novoselova TV, Hussain M, King PJ, et al. MRAP deficiency impairs adrenal progenitor cell differentiation and gland zonation. *Faseb J*. 2018;32. fj201701274RR.
125. Nickel AG, von Hardenberg A, Hohl M, et al. Reversal of mitochondrial transhydrogenase causes oxidative stress in heart failure. *Cell Metab*. 2015;22:472–484.
126. Meimaridou E, Kowalczyk J, Guasti L, et al. Mutations in NNT encoding nicotinamide nucleotide transhydrogenase cause familial glucocorticoid deficiency. *Nat Genet*. 2012;44:740–742.
127. Jazayeri O, Liu X, van Diemen CC, et al. A novel homozygous insertion and review of published mutations in the NNT gene causing familial glucocorticoid deficiency (FGD). *Eur J Med Genet*. 2015;58:642–649.
128. Bainbridge MN, Davis EE, Choi WY, et al. Loss of function mutations in NNT are associated with left ventricular noncompaction. *Circ Cardiovasc Genet*. 2015;8:544–552.
129. Roucher-Boulez F, Mallet-Motak D, Samara-Boustani D, et al. NNT mutations: a cause of primary adrenal insufficiency, oxidative stress and extra-adrenal defects. *Eur J Endocrinol*. 2016;175:73–84.
130. Hershkovitz E, Arafat M, Loewenthal N, et al. Combined adrenal failure and testicular adrenal rest tumor in a patient with nicotinamide nucleotide transhydrogenase deficiency. *J Pediatr Endocrinol Metab*. 2015;28:1187.
131. Meimaridou E, Goldsworthy M, Chortis V, et al. NNT is a key regulator of adrenal redox homeostasis and steroidogenesis in male mice. *J Endocrinol*. 2018;236:13–28.
132. Williams JL, Paudyal A, Awad S, et al. Mylk3 null C57BL/6N mice develop cardiomyopathy, whereas Nnt null C57BL/6J mice do not. *Life Sci Alliance*. 2020;3:e201900593.
133. Biterova EI, Turanov AA, Gladyshev VN, et al. Crystal structures of oxidized and reduced mitochondrial thioredoxin reductase provide molecular details of the reaction mechanism. *Proc Natl Acad Sci USA*. 2005;102:15018–15023.
134. Hanukoglu I. Antioxidant protective mechanisms against reactive oxygen species (ROS) generated by mitochondrial P450 systems in steroidogenic cells. *Drug Metab Rev*. 2006;38:171–196.
135. Prasad R, Chan LF, Hughes CR, et al. Thioredoxin reductase 2 (TXNRD2) mutation associated with familial glucocorticoid deficiency (FGD). *J Clin Endocrinol Metab*. 2014;99:E1556–E1563.
136. Conrad M, Jakupoglu C, Moreno SG, et al. Essential role for mitochondrial thioredoxin reductase in hematopoiesis, heart development, and heart function. *Mol Cell Biol*. 2004;24:9414–9423.
137. Kiermayer C, Northrup E, Schrewe A, et al. Heart-specific knockout of the mitochondrial thioredoxin reductase (Txnrd2) induces metabolic and contractile dysfunction in the aging myocardium. *J Am Heart Assoc*. 2015;4:e002153.
138. Sibbing D, Pfeufer A, Perisic T, et al. Mutations in the mitochondrial thioredoxin reductase gene TXNRD2 cause dilated cardiomyopathy. *Eur Heart J*. 2011;32:1121–1133.
139. Allgrove J, Clayden GS, Grant DB, et al. Familial glucocorticoid deficiency with achalasia of the cardia and deficient tear production. *Lancet*. 1978;1:1284–1286.
140. Handschug K, Sperling S, Yoon SJ, et al. Triple A syndrome is caused by mutations in AAAS, a new WD-repeat protein gene. *Human Mole Genet*. 2001;10:283–290.
141. Cho AR, Yang KJ, Bae Y, et al. Tissue-specific expression and subcellular localization of ALADIN, the absence of which causes human triple A syndrome. *Exp Mol Med*. 2009;41:381–386.
142. Houlden H, Smith S, De Carvalho M, et al. Clinical and genetic characterization of families with triple A (Allgrove) syndrome. *Brain*. 2002;125:2681–2690.
143. Prasad R, Metherell LA, Clark AJ, et al. Deficiency of ALADIN impairs redox homeostasis in human adrenal cells and inhibits steroidogenesis. *Endocrinology*. 2013;154:3209–3218.
144. Jühlen R, Idkowiak J, Taylor AE, et al. Role of ALADIN in human adrenocortical cells for oxidative stress response and steroidogenesis. *PLOS ONE*. 2015;10:e0124582.
145. Huebner A, Mann P, Rohde E, et al. Mice lacking the nuclear pore complex protein ALADIN show female infertility but fail to devel-

op a phenotype resembling human triple A syndrome. *Mol Cell Biol*. 2006;26:1879–1887.

146. O'Riordan SM, Lynch SA, Hindmarsh PC, et al. A novel variant of familial glucocorticoid deficiency prevalent among the Irish traveler population. *J Clin Endocrinol Metab*. 2008;93:2896–2899.
147. Hughes CR, Guasti L, Meimaridou E, et al. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. *J Clin Invest*. 2012;122:814–820.
148. Gineau L, Cognet C, Kara N, et al. Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. *J Clin Invest*. 2012;122:821–832.
149. Chuang CH, Wallace MD, Abratte C, et al. Incremental genetic perturbations to MCM2-7 expression and subcellular distribution reveal exquisite sensitivity of mice to DNA replication stress. *PLoS Genet*. 2010;6:e1001110.
150. Maceyka M, Harikumar KB, Milstien S, et al. Sphingosine-1-phosphate signaling and its role in disease. *Trends Cell Biol*. 2012;22:50–60.
151. Lucki NC, Li D, Sewer MB. Sphingosine-1-phosphate rapidly increases cortisol biosynthesis and the expression of genes involved in cholesterol uptake and transport in H295R adrenocortical cells. *Mol Cell Endocrinol*. 2012;348:165–175.
152. Maharaj A, Williams J, Bradshaw T, et al. Sphingosine-1-phosphate lyase (SGPL1) deficiency is associated with mitochondrial dysfunction. *J Steroid Biochem Mol Biol*. 2020;202:105730.
153. Lucki NC, Sewer MB. Nuclear sphingolipid metabolism. *Ann Rev Physiol*. 2012;74:131–151.
154. Ozbay T, Rowan A, Leon A, et al. Cyclic adenosine 5'-monophosphate-dependent sphingosine-1-phosphate biosynthesis induces human CYP17 gene transcription by activating cleavage of sterol regulatory element binding protein 1. *Endocrinology*. 2006;147:1427–1437.
155. Prasad R, Hadjidemetriou I, Maharaj A, et al. Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest*. 2017;127:942–953.
156. Lovric S, Goncalves S, Gee HY, et al. Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest*. 2017;127:912–928.
157. Janecke AR, Xu R, Steichen-Gersdorf E, et al. Deficiency of the sphingosine-1-phosphate lyase SGPL1 is associated with congenital nephrotic syndrome and congenital adrenal calcifications. *Hum Mutat*. 2017;38:365–372.
158. Poliani PL, Kisand K, Marrella V, et al. Human peripheral lymphoid tissues contain autoimmune regulator-expressing dendritic cells. *Am J Pathol*. 2010;176:1104–1112.
159. Suzuki E, Kobayashi Y, Kawano O, et al. Expression of AIRE in thymocytes and peripheral lymphocytes. *Autoimmunity*. 2008;41:133–139.
160. Pellegrino M, Bellacchio E, Dhamo R, et al. A novel homozygous mutation of the AIRE gene in an APECED patient from Pakistan: case report and review of the literature. *Front Immunol*. 2018;9:1835.
161. Söderbergh A, Myhre AG, Ekwall O, et al. Prevalence and clinical associations of 10 defined autoantibodies in autoimmune polyendocrine syndrome type I. *J Clin Endocrinol Metab*. 2004;89:557–562.
162. Meager A, Visvalingam K, Peterson P, et al. Anti-interferon autoantibodies in autoimmune polyendocrinopathy syndrome type 1. *PLoS Med*. 2006;3:e289.
163. Puel A, Döffinger R, Natividad A, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J Exp Med*. 2010;207:291–297.
164. Bruserud Ø, Oftedal BE, Landegren N, et al. A longitudinal follow-up of autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*. 2016;101:2975–2983.
165. Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*. 1997;17:399–403.
166. Borchers J, Pukkala E, Mäkitie O, et al. Patients with APECED have increased early mortality due to endocrine causes, malignancies and infections. *J Clin Endocrinol Metab*. 2020;105:e2207–e2213.
167. Ramsey C, Winqvist O, Puhakka L, et al. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum Mol Genet*. 2002;11:397–409.
168. Wiesinger C, Kunze M, Regelsberger G, et al. Impaired very long-chain acyl-CoA β-oxidation in human X-linked adrenoleukodystrophy fibroblasts is a direct consequence of ABCD1 transporter dysfunction. *J Biol Chem*. 2013;288:19269–19279.
169. Whitcomb RW, Linehan WM, Knazek RA. Effects of long-chain, saturated fatty acids on membrane microviscosity and adrenocorticotropin responsiveness of human adrenocortical cells in vitro. *J Clin Invest*. 1988;81:185–188.
170. Powers JM, Schaumburg HH, Johnson AB, et al. A correlative study of the adrenal cortex in adreno-leukodystrophy--evidence for a fatal intoxication with very long chain saturated fatty acids. *Invest Cell Pathol*. 1980;3:353–376.
171. Hein S, Schönfeld P, Kahlert S, et al. Toxic effects of X-linked adrenoleukodystrophy-associated, very long chain fatty acids on glial cells and neurons from rat hippocampus in culture. *Hum Mol Genet*. 2008;17:1750–1761.
172. Mosser J, Douar AM, Sarde CO, et al. Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature*. 1993;361:726–730.
173. Engelen M, Kemp S, Poll-The BT. X-linked adrenoleukodystrophy: pathogenesis and treatment. *Curr Neurol Neurosci Rep*. 2014;14:486.
174. Kemp S, Berger J, Aubourg P. X-linked adrenoleukodystrophy: clinical, metabolic, genetic and pathophysiological aspects. *Biochim Biophys Acta*. 2012;1822:1465–1474.
175. van Geel BM, Bezman L, Loes DJ, et al. Evolution of phenotypes in adult male patients with X-linked adrenoleukodystrophy. *Ann Neurol*. 2001;49:186–194.
176. Dubey P, Raymond GV, Moser AB, et al. Adrenal insufficiency in asymptomatic adrenoleukodystrophy patients identified by very long-chain fatty acid screening. *J Pediatr*. 2005;146:528–532.
177. el-Deiry SS, Naidu S, Blevins LS, et al. Assessment of adrenal function in women heterozygous for adrenoleukodystrophy. *J Clin Endocrinol Metab*. 1997;82:856–860.
178. Engelen M, Barbier M, Dijkstra IM, et al. X-linked adrenoleukodystrophy in women: a cross-sectional cohort study. *Brain*. 2014;137:693–706.
179. Kobayashi T, Shinnoh N, Kondo A, et al. Adrenoleukodystrophy protein-deficient mice represent abnormality of very long chain fatty acid metabolism. *Biochem Biophys Res Commun*. 1997;232:631–636.
180. Lu J-F, Barron-Casella E, Deering R, et al. The role of peroxisomal ABC transporters in the mouse adrenal gland: the loss of Abcd2 (ALDR), Not Abcd1 (ALD), causes oxidative damage. *Lab Invest*. 2007;87:261–272.
181. Waterham HR, Ebberink MS. Genetics and molecular basis of human peroxisome biogenesis disorders. *Biochim Biophys Acta*. 2012;1822:1430–1441.
182. Bowen P, Lee CS, Zellweger H, et al. A familial syndrome of multiple congenital defects. *Bull Johns Hopkins Hosp*. 1964;114:402–414.
183. Berendse K, Engelen M, Linthorst GE, et al. High prevalence of primary adrenal insufficiency in Zellweger spectrum disorders. *Orphanet J Rare Dis*. 2014;9:133.
184. Matsumoto N, Tamura S, Moser A, et al. The peroxin Pex6p gene is impaired in peroxisomal biogenesis disorders of complementation group 6. *J Hum Genet*. 2001;46:273–277.
185. Reuber BE, Germain-Lee E, Collins CS, et al. Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders. *Nat Genet*. 1997;17:445–448.
186. Hiebler S, Masuda T, Hacia JG, et al. The Pex1-G844D mouse: a model for mild human Zellweger spectrum disorder. *Mol Genet Metab*. 2014;111:522–532.
187. Maxwell MA, Nelson PV, Chin SJ, et al. A common PEX1 frameshift mutation in patients with disorders of peroxisome biogenesis correlates with the severe Zellweger syndrome phenotype. *Hum Genet*. 1999;105:38–44.
188. Maxwell MA, Allen T, Solly PB, et al. Novel PEX1 mutations and genotype–phenotype correlations in Australasian peroxisome biogenesis disorder patients. *Hum Mut*. 2002;20:342–351.

189. Preuss N, Brosius U, Biermanns M, et al. PEX1 mutations in complementation group 1 of Zellweger spectrum patients correlate with severity of disease. *Pediatr Res*. 2002;51:706–714.
190. Klouwer FC, Berendse K, Ferdinandusse S, et al. Zellweger spectrum disorders: clinical overview and management approach. *Orphanet J Rare Dis*. 2015;10:151.
191. Pericleous M, Kelly C, Wang T, et al. Wolman's disease and cholesteryl ester storage disorder: the phenotypic spectrum of lysosomal acid lipase deficiency. *Lancet Gastroenterol Hepatol*. 2017;2:670–679.
192. Abramov A, Schorr S, Wolman M. Generalized xanthomatosis with calcified adrenals. *AMA J Dis Child*. 1956;91:282–286.
193. Aguisanda F, Thorne N, Zheng W. Targeting Wolman disease and cholesteryl ester storage disease: disease pathogenesis and therapeutic development. *Curr Chem Genom Transl Med*. 2017;11:1–18.
194. Zhang B, Porto AF. Cholesteryl ester storage disease: protean presentations of lysosomal acid lipase deficiency. *J Pediatr Gastroenterol Nutr*. 2013;56:682–685.
195. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43:339–344.
196. Wild PS, Zeller T, Schillert A, et al. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet*. 2011;4:403–412.
197. Zschenker O, Illies T, Ameis D. Overexpression of lysosomal acid lipase and other proteins in atherosclerosis. *J Biochem*. 2006;140:23–38.
198. Nègre-Salvayre A, Salvayre R. UV-treated lipoproteins as a model system for the study of the biological effects of lipid peroxides on cultured cells. 4. Calcium is involved in the cytotoxicity of UV-treated LDL on lymphoid cell lines. *Biochim Biophys Acta*. 1992;1123:207–215.
199. Krivit W, Peters C, Dusenbery K, et al. Wolman disease successfully treated by bone marrow transplantation. *Bone Marrow Transplant*. 2000;26:567–570.
200. Valayannopoulos V, Plantaz D, Vara R, et al. Clinical effect of sebelipase alfa on survival and growth in infants with lysosomal acid lipase deficiency (Wolman disease). *Mol Genet Metab*. 2014;111:S108.
201. Du H, Duanmu M, Witte D, et al. Targeted disruption of the mouse lysosomal acid lipase gene: long-term survival with massive cholesteryl ester and triglyceride storage. *Hum Mol Genet*. 1998;7:1347–1354.
202. Cartier N, Sarde CO, Douar AM, et al. Abnormal messenger RNA expression and a missense mutation in patients with X-linked adrenoleukodystrophy. *Hum Mol Genet*. 1993;2:1949–1951.
203. Aslanidis C, Ries S, Fehringer P, et al. Genetic and biochemical evidence that CESD and Wolman disease are distinguished by residual lysosomal acid lipase activity. *Genomics*. 1996;33:85–93.

87

Enzymes and Pathways of Human Steroidogenesis

Richard J. Auchus and John D.C. Newell-Price

OUTLINE

Overview of the Human Steroidogenic Enzymes and Steroidogenesis, 1449
  Cytochrome P450 Enzymes, 1450
  Hydroxysteroid Dehydrogenases and Reductases, 1450
  Acute Regulation of Steroidogenesis, 1451
  Chronic Maintenance of the Steroidogenic Machinery, 1452
Human Steroidogenic Cytochrome P450 Enzymes, 1454
  P450scc (Encoded by *CYP11A1*), 1454
  P450c17 (Encoded by *CYP17A1*), 1454
  P450c21 (Encoded by *CYP21A2*), 1456
  P450c11β and P450c11AS (Encoded by *CYP11B1* and *CYP11B2* in Humans), 1456
  P450aro (Aromatase), 1458
Redox Partner Proteins, 1458
  Ferredoxin, 1458
  Ferredoxin Reductase, 1459
  P450 Oxidoreductase, 1459
  Cytochrome b5, 1460
Steroidogenic Dehydrogenases and Reductases, 1460
  3β-Hydroxysteroid Dehydrogenase/$\Delta^5 \rightarrow \Delta^4$-Isomerases, 1460
  17β-Hydroxysteroid Dehydrogenases, 1461
  Steroid 5α-Reductases, 1461
  3α-Hydroxysteroid Dehydrogenases, 1462
  11β-Hydroxysteroid Dehydrogenases, 1462
Steroid Sulfonation, 1463
Pathways, 1464
  Adrenal Steroidogenic Pathways, 1464
  Gonadal Steroidogenic Pathways, 1464
  The Backdoor Pathway to Dihydrotestosterone, 1464
  Pathways to 11-Oxygenated Androgens, 1464
  Androgen Synthesis in Prostate Cancer, 1468

OVERVIEW OF THE HUMAN STEROIDOGENIC ENZYMES AND STEROIDOGENESIS

All steroid hormones derive from cholesterol in a process that can be conceptualized as having six distinct components:

1. *The conversion of cholesterol to pregnenolone.* Although viewed superficially as a single chemical transformation, the mobilization of cholesterol into the steroidogenic pathways is a complex event that serves as a key locus of regulation and also conventionally defines a tissue as "steroidogenic." In humans, only the cells of the adrenal cortex, testicular Leydig cells, ovarian granulosa cells, and trophoblast cells of the placenta possess high capacity to cleave cholesterol into pregnenolone (the C21 precursor of all active steroid hormones) and isocaproaldehyde, while skin and specific glial and neuronal cells of the brain have lower capacity for local synthesis. The differences in how this process is regulated and in how pregnenolone is subsequently metabolized define the roles of the various steroidogenic cells and tissues in human physiology. Unlike peptide-secreting glands, steroidogenic cells do not store steroid hormones and intermediates; it is the activation of this first step that enables the rapid synthesis and release of steroids in response to hormonal and environmental stimuli.
2. *The transformation of pregnenolone to active hormones, intermediates, and secreted steroid derivatives.* The repertoire of enzymes and cofactor proteins present in a given steroidogenic cell is responsible for the characteristic steroid profile of that cell type, and the coordinate regulation of their expression promotes the completion of all steps of a given pathway. Thus, these enzymes determine qualitatively what steroids are made, but since these steps are not kinetically limiting, it is step 1 that quantitatively regulates how much steroid is made at a given moment.
3. *Peripheral metabolism of hormones and precursors.* Although not "steroidogenic" as defined by being able to convert cholesterol to pregnenolone, some organs, such as the liver, possess tremendous capacity to transform various steroids. For example, 70% to 80% of circulating testosterone in normally cycling women derives from the conversion of adrenal dehydroepiandrosterone (DHEA). Steroids can be activated in target tissues, such as the conversion of testosterone to dihydrotestosterone (DHT) in the prostate. In contrast, active androgens and estrogens are inactivated in the uterus and in other peripheral tissues.
4. *Multiple layers of regulation.* The regulation of steroidogenesis by tropic hormones, such as adrenocorticotropic hormone (ACTH), angiotensin II, and luteinizing hormone (LH), is well known; less well appreciated are the multiple levels at which such regulation may occur. These include the transcriptional regulation of the genes encoding steroidogenic enzymes and cofactors; regulation of mitochondrial protein import; transfer of electrons from nicotinamide adenine dinucleotide phosphate (NADPH); posttranslational modification and proteosomal degradation of steroidogenic enzymes; and subcellular localization and/or targeting.
5. *Catabolism and unproductive metabolism.* A panopoly of steroids can be isolated from human plasma and tissues, many of which have negligible biological activity. Most inactive by-products derive from hepatic transformations (e.g., 6β-hydroxylation and

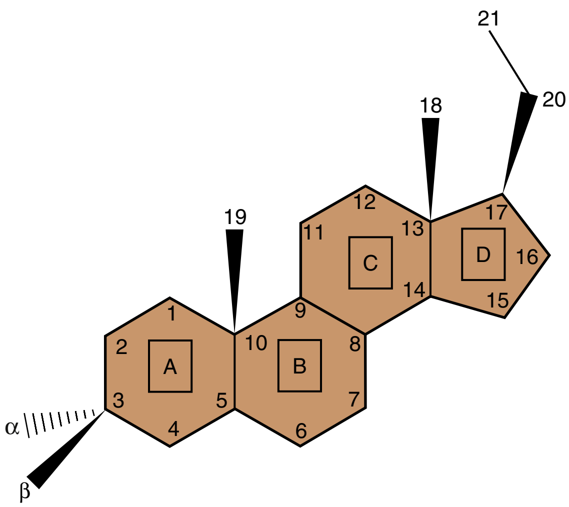


Figure 87.1 Cycloperhydropentanophenanthrene steroid nucleus. Steroid rings are identified with *boxed capital letters,* and carbon atoms are *numbered.* Substituents and hydrogens are labeled as α or β if they are positioned behind or in front of the plane of the page, respectively.

5β-reduction of C21 and C19 steroids, and 4-hydroxylation of estrogens), which promote renal excretion of these steroids.

6. *More than one pathway to specific steroids.* Particularly for the terminal steps of long pathways, such as the synthesis of DHT or estradiol, two or more routes may yield the same final product. Often the different routes utilize different enzymes, and these enzymes may be found in different tissues under different regulatory mechanisms. The relative importance of these pathways differ with age, sex, and physiologic state.

Steroids are molecules derived from the cyclopentanoperhydrophenanthrene four-ring hydrocarbon nucleus (Fig. 87.1). Most enzymes involved in steroid biosynthesis are either cytochrome P450 enzymes or hydroxysteroid dehydrogenases (Fig. 87.2).[1] All P450-mediated hydroxylations and carbon–carbon bond cleavage reactions are mechanistically and physiologically irreversible. Hydroxysteroid dehydrogenase reactions are mechanistically reversible and can run in either direction under certain conditions *in vitro*, but each hydroxysteroid dehydrogenase drives steroid flux predominantly in either the oxidative or reductive mode *in vivo*.[2] However, two or more hydroxysteroid dehydrogenases drive the flux of a steroid pair in opposite directions, some favoring ketosteroid reduction and others favoring hydroxysteroid oxidation.

KEY POINTS

- The conversion of cholesterol to pregnenolone is the quantitative, rate-limiting step of steroidogenesis and the site of acute regulation.
- Steroidogenesis follows a specific sequence with some branch points and redundancies, and each step is either irreversible or has a strong directional preference.
- Steroid formation features multiple layers of regulation, redundancy for some pathways, and multiple activities for some key enzymes.

Cytochrome P450 Enzymes

Mammalian cytochrome P450 enzymes fall into two broad classes: type 1 and type 2.[1] Type 1 enzymes and their electron transfer proteins reside in the mitochondria (Table 87.1) of eukaryotes; almost all bacterial P450s are also type 1 enzymes. Human type 1 P450 enzymes include the cholesterol side-chain cleavage enzyme (P450scc); the two isozymes of 11-hydroxylase (P450c11β and P450c11AS); and three of the principal enzymes in vitamin D metabolism (1α-hydroxylase, 24-hydroxylase, and 25-hydroxylase). Type 1 enzymes receive electrons from the reduced form of NADPH via ferredoxin (FDX1), a small, soluble, iron-sulfur protein. FDX1 does not oxidize NADPH directly, however, but receives the two electrons from NADPH via the flavoprotein ferredoxin reductase (FDXR, Fig. 87.3). Type 2 enzymes, in contrast, receive electrons from NADPH via the flavin adenine dinucleotide (FAD)-flavin mononucleotide (FMN) two-flavin protein, P450 oxidoreductase (POR). Type 2 enzymes are exclusively located in the smooth endoplasmic reticulum and constitute 50 of the 57 human P450 enzymes.

P450 enzymes activate molecular oxygen using their heme center and electrons from NADPH. Substrate binding is required prior to heme reduction with one electron, which enables oxygen binding, the second one-electron transfer, and formation of the iron-oxygen complex, which then oxygenates the substrate. Thus, P450 reactions on steroids are limited to oxygen insertion (hydroxylation) reactions and, in a few notable cases, oxidative carbon–carbon bond cleavage reactions (Table 87.2).

Hydroxysteroid Dehydrogenases and Reductases

All hydroxysteroid dehydrogenases (HSDs) and related enzymes use nicotinamide cofactors either to reduce or to oxidize the steroid by two electrons through a hydride transfer mechanism.[1] Most examples involve the conversion of a secondary alcohol to a ketone or vice versa, and in the case of the 3β-hydroxysteroid dehydrogenase/$\Delta^5 \rightarrow \Delta^4$-isomerases, the dehydrogenation is accompanied by the isomerization of the adjacent carbon–carbon double bond from the Δ^5 (between carbons 5 and 6) to the Δ^4 positions (see Figs. 87.1 and 87.2). The human steroid 5α-reductases types 1 and 2, which are included with the HSDs for convenience, reduce olefinic carbon–carbon double bonds to the saturated state rather than acting on carbon centers bonded to oxygen.

The HSDs can be categorized according to either structural or functional classification schemes. *Structurally*, HSDs are members of either the short-chain dehydrogenase reductase (SDR) or aldo-keto reductase (AKR) families.[2] The SDR enzymes are β-α-β proteins, where up to seven parallel β-strands fan across the center of the molecule, forming the "Rossman fold" characteristic of oxidation/reduction enzymes that use nicotinamide cofactors. The AKR enzymes are soluble proteins that contain a beta-barrel or triose-phosphate isomerase (TIM-barrel) motif in which eight parallel β-strands lie in a slanted circular distribution like the staves of a barrel. In both cases, the active site contains a critical tyrosine and lysine pair of residues involved in proton transfer from or to the steroid alcohol during catalysis. *Functionally,* HSDs act either as true dehydrogenases, using NAD^+ as a cofactor to convert hydroxysteroids to ketosteroids; or as ketosteroid reductases, utilizing predominantly NADPH to reduce ketosteroids. Many HSDs catalyze either oxidation or reduction *in vitro* based on the pH and cofactor concentrations, but these enzymes, when expressed in intact mammalian cells, drive steroid flux primarily in one direction.[2] These directional preferences derive primarily from the relative abundance of the oxidized and reduced form of cofactors and the relative affinity of each enzyme for NAD(H) versus NADP(H) because cofactor concentrations exceed steroid concentrations by many orders of magnitude.[2] Consequently, the directional preference of some "reductive" enzymes can be reduced or reversed by depleting cells of NADPH or by mutations that attenuate the preferential binding of NADPH over NAD(H).[3,4]

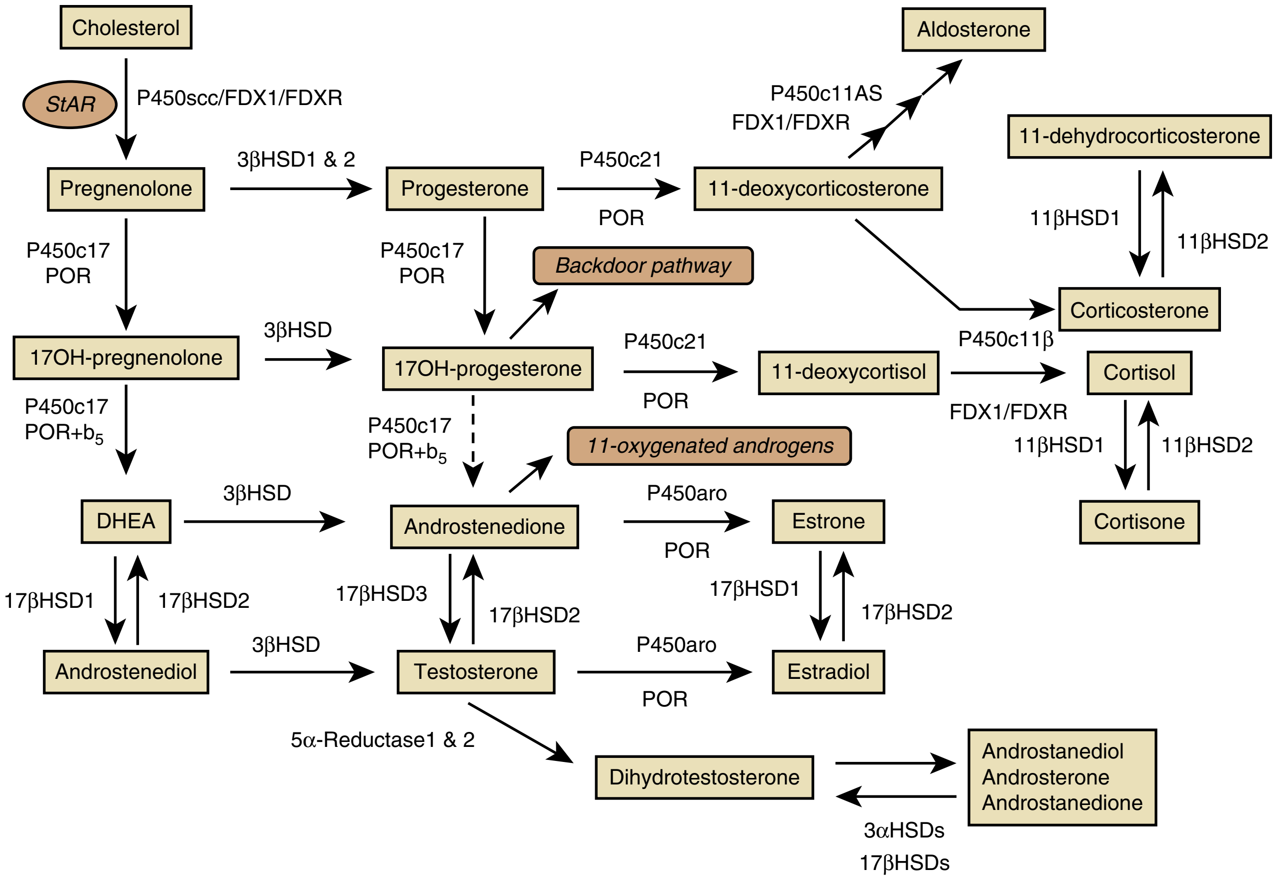

Cholesterol
StAR
P450scc/FDX1/FDXR
Pregnenolone
3βHSD1 & 2
Progesterone
P450c21
POR
11-deoxycorticosterone
P450c11AS
FDX1/FDXR
Aldosterone
11-dehydrocorticosterone
11βHSD1
11βHSD2
Corticosterone
P450c17
POR
P450c17
POR
Backdoor pathway
17OH-pregnenolone
3βHSD
17OH-progesterone
P450c21
POR
11-deoxycortisol
P450c11β
FDX1/FDXR
Cortisol
11βHSD1
11βHSD2
Cortisone
P450c17
POR+b$_5$
P450c17
POR+b$_5$
11-oxygenated androgens
DHEA
3βHSD
Androstenedione
P450aro
POR
Estrone
17βHSD1
17βHSD2
17βHSD3
17βHSD2
17βHSD1
17βHSD2
Androstenediol
3βHSD
Testosterone
P450aro
POR
Estradiol
5α-Reductase1 & 2
Dihydrotestosterone
Androstanediol
Androsterone
Androstanedione
3αHSDs
17βHSDs

Figure 87.2 **Major human steroidogenic pathways.** Key enzymes and cofactor proteins are shown near *arrows* indicating chemical reactions. The steroidogenic acute regulatory protein (StAR) protein *(oval)* mobilizes cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane, where P450scc cleaves cholesterol to pregnenolone, the first committed intermediate in steroid biosynthesis. The steroids in the *first column* are Δ^5-steroids, which constitute the preferred pathway to C19 steroids in humans, and the *dashed arrow* indicates poor flux from 17α-hydroxyprogesterone to androstenedione. Steroids in the *second column and farther right* are Δ^4-steroids, except the C18 estrogens (estrone and estradiol) and 5α-reduced steroids, including the potent androgen dihydrotestosterone and other androstanes *(bottom row)*. *Diagonal arrows* indicate additional pathways from 17α-hydroxyprogesterone (backdoor pathway) and androstenedione (11-oxygenated androgens) illustrated in Fig. 87.7. Not all intermediate steroids, pathways, and enzymes are shown.

KEY POINTS

- The two major classes of steroid biosynthetic enzymes are cytochrome P450 enzymes and the hydroxysteroid dehydrogenases/reductases.
- Of the steroidogenic cytochrome P450 enzymes, P450scc, P450c11β, and P450c11AS are mitochondrial and use ferredoxin/ferredoxin reductase as electron transfer proteins; P450c17, P450c21, and P450aro reside in the endoplasmic reticulum and use P450-oxidoreductase as their electron transfer protein.
- The hydroxysteroid dehydrogenases catalyze mechanistically reversible oxidation/reduction reactions using NAD[P][H] cofactors; some enzymes strongly favor hydroxysteroid oxidation, and others strongly favor ketosteroid reduction in intact cells.

Acute Regulation of Steroidogenesis

Every time a pulse of ACTH reaches the adrenal cortex, or a pulse of LH reaches the gonad, a subsequent pulse of steroid hormone production is observed within minutes. Although it has long been known that the loss of trophic hormones from the pituitary gland leads to adrenal and gonadal atrophy, the action of ACTH and LH to promote organ survival and to maintain steroidogenic capacity occurs at three distinct levels. First, as seen in long-term exposure to ACTH (e.g., in Cushing disease), ACTH promotes adrenal growth. This growth occurs primarily by ACTH stimulating the production of cyclic adenosine monophosphate (cAMP), which, in turn, promotes the synthesis of insulin-like growth factor 2 (IGF-2), basic fibroblast growth factor,

TABLE 87.1 Intracellular Location of Steroidogenic Proteins

Mitochondria	Cytoplasm	Endoplasmic Reticulum
P450scc		P450c17
P450c11β		P450c21
P450c11AS		P450aro
Ferredoxin reductase		P450-oxidoreductase
Ferredoxin	17βHSD1	17βHSD1, 2, 3, 6*
StAR	StAR	Cytochrome b5
3βHSD1 and 2	3βHSD1 and 2	3βHSD1 and 2
	AKR1C1-4	11βHSD1 and 2

HSD, Hydroxysteroid dehydrogenase.
*17βHSD4 is located in peroxisomes.

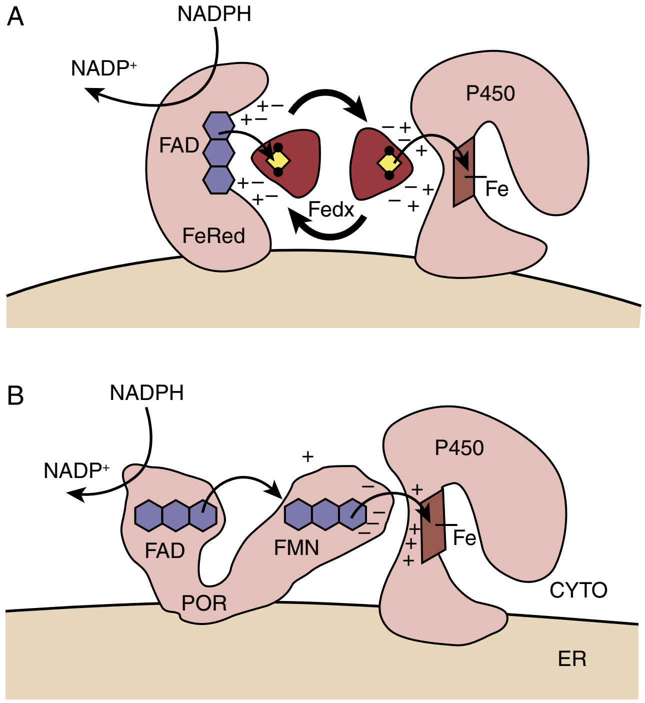


Figure 87.3 Electron transfer pathways for steroidogenic cytochrome P450 enzymes. A, In type I (mitochondrial) enzymes, the two electrons from the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) pass from the flavin (FAD) of ferredoxin reductase (FeRed) to the iron-sulfur (Fe2S2, diamond with dots) cluster of ferredoxin (Fedx) and then to the heme of the P450 *(square with Fe).* Negatively charged residues in Fedx (–) guide docking and electron transfer with positively charged residues (+) in both Fedx and the P450. **B,** In type II (microsomal) enzymes, the flavoprotein P450-oxidoreductase (POR) receives electrons from NADPH to its FAD moiety, transfers electrons to its FMN moiety, and after a conformational rearrangement, directly transfers electrons from the FMN to the P450. Negative charges on POR (–) and positive charges (+) on the P450 guide the interaction as with the type I P450; with some P450 enzymes, phosphorylation and cytochrome b5 also regulate electron transfer and catalysis. Heme of P450 is indicated by a *square* with iron atom (Fe).

and epidermal growth factor. Together, these growth factors stimulate adrenal cellular hypertrophy and hyperplasia. Second, ACTH acts long-term through cAMP, and angiotensin II acts through the calcium/calmodulin pathway to promote the transcription of genes encoding various steroidogenic enzymes and electron-donating cofactor proteins. Third, ACTH fosters the increased flow of cholesterol into the mitochondria, where it becomes substrate for the first and rate-limiting enzyme, P450scc. This acute response occurs within minutes and is inhibited by inhibitors of protein synthesis (e.g., puromycin or cycloheximide), indicating that a short-lived protein species mediates this process. Although other proteins are involved in the chronic replenishment of mitochondrial cholesterol, abundant biochemical, clinical, and genetic evidence implicates the steroidogenic acute regulatory protein (StAR) as this labile protein mediator.[5]

StAR is a 37-kD phosphoprotein that is cleaved to a 30-kD form when it enters the mitochondrion. Overexpression of mouse StAR in mouse Leydig MA-10 cells increased their basal steroidogenic rate,[5] and cotransfection of expression vectors for both StAR and the P450scc system in nonsteroidogenic COS-1 cells augmented pregnenolone synthesis above that obtained with the P450scc system alone.[6] Mutations in StAR cause congenital lipoid adrenal hyperplasia,[6,7] in which very little steroid is made; and targeted disruption of the *Star* gene in the mouse causes a similar phenotype.[8]

The mechanism of StAR's action is not completely understood. After other proteins transport cholesterol to the mitochondria, StAR acts on the outer mitochondrial membrane (OMM),[9] and its activity in promoting steroidogenesis is proportional to its residency time on the OMM.[10] When expressed in cytoplasm or added to mitochondria *in vitro*, both the 37-kD "precursor" and the 30-kD "mature form" of StAR are equally active, but StAR is inactive in the mitochondrial intramembranous space or matrix.[10] Thus, it is StAR's cellular localization, not its cleavage, that determines whether or not it is active. StAR has a sterol-binding pocket that accommodates a single molecule of cholesterol.[11] The interaction of StAR with the OMM involves conformational changes[12] that are necessary for StAR to accept and discharge cholesterol molecules. Although StAR can transfer cholesterol between synthetic membranes *in vitro*, suggesting that other protein molecules are not needed for its action, this activity can also be seen with the inactive mutant R182L, which causes lipoid congenital adrenal hyperplasia.[13] Thus StAR's action to promote steroidogenesis can be dissociated from its cholesterol-transfer activity. StAR appears to interact with a complex of OMM proteins that includes the peripheral benzodiazepine receptor (PBR, also termed the *translocator protein* [TSPO]), voltage-dependent anion channel 1 (VDAC1), the acyl-CoA-binding-domain proteins ACBD-1 and ACBD-3, and possibly others. Each molecule of StAR appears to be recycled, moving hundreds of molecules of cholesterol before the cleavage/inactivation event.[14] Although StAR is required for the acute steroidogenic response, steroidogenesis will persist in the absence of StAR at approximately 14% of the StAR-induced rate, accounting for the steroidogenic capacity of tissues that lack StAR (e.g., the placenta and the brain). TSPO ligands stimulate testosterone production in rats,[15] yet mice with targeted deletion of the *Tspo* gene in Leydig cells have normal stimulated testosterone production and are fertile.[16] These paradoxical findings suggest that more than one protein complex or cholesterol transfer mechanism can contribute to steroidogenesis in specific cells.

Chronic Maintenance of the Steroidogenic Machinery

The acute regulation of steroidogenesis is determined by access of cholesterol to the P450scc enzyme, and P450scc is the rate-limiting enzymatic step in steroidogenesis. Thus, the chronic regulation of steroidogenesis is quantitatively (how much) determined by *CYP11A1* (encoding P450scc) expression[17] and qualitatively (which steroids) determined by the expression and activities of downstream enzymes. The episodic bursts of cAMP resulting from the binding of ACTH and LH to their respective receptors are necessary but not sufficient for the continued expression of the steroidogenic enzymes and the production of steroids. Patients with inactivating mutations in the ACTH receptor (melanocortin type 2 receptor, MC2R)[18] or LH receptor[19] make negligible amounts of steroids from the affected glands. Conversely, activating mutations of the Gsα protein, which couples receptor binding to cAMP generation, and activating mutations of the LH receptor cause hypersecretion of steroids.[20] Indeed, cAMP-responsive elements have been identified in the genes for most of the human steroidogenic P450 enzymes, but this mechanism alone does not account for the diversity of steroid production observed in the various zones of the adrenal cortex, the gonads of both sexes, the placenta, and the brain.

Other transcription factors (e.g., AP-2, SP-1, SP-3, NF1C, NR4A1, NR4A2, GATA4, and GATA6) aid in defining the basal- and cAMP-stimulated transcription of each gene, which is also regulated in a tissue-specific manner by the regulatory elements unique to each gene. Among these factors, steroidogenic factor-1 (SF-1, NR5A1), an orphan nuclear receptor, coordinates the expression of steroidogenic enzymes

TABLE 87.2 Key Human Steroidogenic Enzymes and Cofactor Proteins

Protein	Gene	Gene Size (kb)	Chromosomal Locus	Location	Principal Substrates	Major Activities	Deficiency Syndromes
P450scc	*CYP11A1*	>20	15q23-q24	ZG/ZF/ZR, gonads (L,G), placenta, brain	Cholesterol Hydroxysterols	22R-Hydroxylase 20R-Hydroxylase 20,22-Lyase	Deficiency of all steroids
P450c17	*CYP17A1*	6.6	10q24.3	ZF/ZR, gonads (L,T), brain	Preg, 17OH-Preg, Prog [17OH-Prog]* 5α-reduced C21-steroids	17α-Hydroxylase 17,20-Lyase [16α-Hydroxylase] [Δ^{16}-Synthase]	17-Hydroxylase deficiency Isolated 17,20-lyase deficiency
P450c21	*CYP21A2*	3.4	6p21.1	ZG/ZF/ZR	Prog, 17OH-Prog	21-Hydroxylase	21-Hydroxylase deficiency
P450c11β	*CYP11B1*	9.5	8q21-q22	ZF/ZR, brain	11-Deoxycortisol 11-DOC [cortisol]	11β-Hydroxylase [18-Hydroxylase]	11-Hydroxylase deficiency
P450c11AS	*CYP11B2*	9.5	8q21-q22	ZG, brain, heart	Corticosterone 11-DOC [cortisol] 18OH-corticosterone	11β-Hydroxylase 18-Hydroxylase 18-Oxidase	Aldosterone synthase deficiencies
P450aro	*CYP19A1*	>52	15q21.1	Gonads (L,G), placenta, brain, bone, fat	Androstenedione Testosterone	19-Hydroxylase 19-Oxidase aromatization	Aromatase deficiency
3βHSD1	*HSD3B1*	7.8	1p13.1	Placenta, liver, brain	Preg, 17OH-Preg DHEA, Δ^5-A	3β-Dehydrogenase $\Delta^{5/4}$-Isomerase	Not described
3βHSD2	*HSD3B2*	7.8	1p13.1	ZG/ZF>ZR Gonad (L,T)	Preg, 17OH-Preg DHEA, Δ^5-A	3β-Dehydrogenase $\Delta^{5/4}$-Isomerase	3βHSD deficiency
17βHSD1	*HSD17B1*	3.3	17q11-q21	Gonad (G), placenta, breast	Estrone, [DHEA]	17β-Ketosteroid reductase	Not described
17βHSD2	*HSD17B2*	>40	16q24.1-q24.2	Endometrium, broadly	Testosterone, estradiol, DHT	17β-Hydroxysteroid dehydrogenase	Not described
17βHSD3	*HSD17B3*	>60	9q22	Gonad (L)	Androstenedione, 5α-A, 5α/3α-A [DHEA]	17β-Ketosteroid reductase	17-Ketosteroid Reductase deficiency
Reductive 3αHSDs	*AKR1C1-4*	13-25 each	10p14-p15	Liver, broadly	DHT, 5α-A 5α-reduced C21-steroids [17β-HSD: DHEA Androstenedione, 5α/3α-A]	3α-Ketosteroid reductase 17β-Ketosteroid reductase	DSD resembling isolated 17,20-lyase deficiency
Oxidative 3α-HSD	*HSD17B6*	>23	12q13	Liver, prostate, broadly	Adiol, 5α/3α-A, 5α/3α-reduced C21 steroids, (products)	3α-Hydroxysteroid dehydrogenase (3α-, 3β-, and 17β- ketosteroid reductase)	Not described
5α-Reductase 1	*SRD5A1*	>35	5p15	Liver, brain, skin	Testosterone, Δ^4/C21-steroids	5α-Reductase	Not described
5α-Reductase 2	*SRD5A2*	>35	2p23	Prostate, genital skin	Testosterone, Δ^4/C21-steroids	5α-Reductase	5α-Reductase deficiency
11βHSD1	*HSD11B1*	9	1q32-q41	Liver, brain, placenta, fat, broadly	Cortisone, 11-Dehydrocorticosterone [products]	11β-Ketosteroid reductase	Cortisone reductase deficiency
11βHSD2	*HSD11B2*	6.2	16q22	Kidney, gut, placenta	Cortisol, corticosterone	11β-Hydroxysteroid dehydrogenase	Syndrome of apparent mineralocorticoid excess
Ferredoxin	*FDX1*	>30	11q22	Ubiquitous	Mitochondrial P450s	Electron transfer	Not described
Ferredoxin Reductase	*FDXR*	11	17q24-q25	Ubiquitous	Ferredoxin	Electron transfer	Not described
StAR	*STAR*	8	8p11.2	ZF/ZG/ZR, gonad (L, T)	Cholesterol flux within mitochondria	Sterol delivery to P450scc	Lipoid CAH
P450-Oxido-Reductase	*POR*	73	7q11.2	Ubiquitous	Microsomal P450s	Electron transfer	Multiple steroidogenic defects ± ABS

Continued

TABLE 87.2 **Key Human Steroidogenic Enzymes and Cofactor Proteins—cont'd**

Protein	Gene	Gene Size (kb)	Chromosomal Locus	Location	Principal Substrates	Major Activities	Deficiency Syndromes
Cytochrome b5	*CYB5A*	32	18q23	ZR>ZG/ZF, gonads, liver, erythrocytes	Allosteric factor with P450c17	Augments 17,20-lyase activity	Isolated 17,20-lyase deficiency ± methemoglobinemia
H6PDH	*H6PD*	36.5	1p36	Fat, liver	Glucose-6-phosphate + $NADP^+$	NADP+ reduction	Cortisone reductase deficiency
SULT2A1	*SULT2A1*	17	19q13.3	ZR	DHEA, other $\Delta5$-steroids	Sulfonation	Not described
PAPSS2	*PAPSS2*	85	10q24	ZR, cartilage liver	ATP + sulfate APS + ATP	ATP sulfurylase APS kinase	Premature adrenarche with skeletal abnormalities

ZG/ZF/ZR, Adrenal zona glomerulosa/fasciculata/reticularis, respectively; *L,* Leydig cells; *T,* theca cells; *G,* granulosa cells; *17OH-Preg,* 17α-hydroxypregnenolone; *17OH-Prog,* 17α-hydroxyprogesterone; *DHEA,* dehydroepiandrosterone; *DHT,* dihydrotestosterone; *DOC,* deoxycorticosterone; Adiol, 5α-androstane-3α,17β-diol; *Δ^5-A,* androsta-5-ene-3β,17β-diol; *5α-A,* 5α-androstane-3,20-dione; *5α/3α-A,* androsterone; *DSD,* disorder of sex development; *CAH,* congenital adrenal hyperplasia; *ABS,* Antley–Bixler syndrome, *APS,* adenosine phosphosulfate.
*Steroids in brackets are poor substrates.

in adrenal and gonadal cells.[21] By contrast, steroidogenesis in the brain and placenta[22] is independent of SF-1. Targeted disruption of SF-1 in the mouse not only disrupts steroid biosynthesis but also blocks the development of the adrenal glands, gonads, and the ventromedial hypothalamus in homozygous animals.[23] Furthermore, SF-1 does not act in isolation, but its action is modified by other transcription factors (e.g., WT-1 and DAX-1)[24] or by phosphorylation.[25] Thus, the development of steroidogenic organs is intimately related to the capacity to produce steroids, and multiple factors acting on the genes for steroidogenic enzymes yield both common features and diversity among the steroidogenic tissues.

Most steroidogenic enzymes derive from a single mRNA species. The most prominent exception to this paradigm is aromatase, whose gene has four different promoters that enable vastly different regulation of expression of the same aromatase protein in many different tissues.[26]

KEY POINTS

- The acute regulation of steroidogenesis is determined by access of cholesterol to the P450scc enzyme, and P450scc is the rate-limiting enzymatic step in steroidogenesis.
- Adrenocorticotropic hormone in the adrenal cortex, and luteinizing hormone in the gonad, stimulate this first step by the influx of cholesterol, mediated by steroidogenic acute regulatory protein.

HUMAN STEROIDOGENIC CYTOCHROME P450 ENZYMES

P450scc (Encoded by *CYP11A1*)

P450scc consumes three equivalents of NADPH and molecular oxygen during the conversion of cholesterol to pregnenolone. Although the enzyme is named for the cleavage of the cholesterol side chain, this process consists of three discrete steps: 1) the 22-hydroxylation of cholesterol; 2) the 20-hydroxylation of 22(R)-hydroxycholesterol; and 3) the oxidative scission of the C20–C22 bond of 20(R), 22(R)-dihydroxycholesterol—the side-chain cleavage event. The enzyme will utilize free hydroxysterol intermediates as substrates for the side-chain cleavage reaction, a tool that is used experimentally because the hydroxysterols are much more soluble than cholesterol and because their access to P450scc is independent of StAR.[6] *In vivo*, however, little of these free intermediates probably accumulate, because their k_{cat}/K_m ratios are much higher than for cholesterol,[27] and the high K_D for pregnenolone (approximately 3000 nM) drives product dissociation. This complex process is the rate-limiting step in steroidogenesis, with turnover numbers of only approximately 20 molecules of cholesterol per molecule P450scc per minute,[27] and all three reactions utilize the iron oxene (compound I) species without incorporation of dioxygen atoms into pregnenolone or isocaproaldehyde during the third reaction.[28] The x-ray crystal structure of the P450scc–ferredoxin complex shows that cholesterol substrate binds with the steroid nucleus angled to approximately 45 degrees with the heme ring, with the side chain extended across the heme,[29] which is consistent with the known order of reactions. P450scc will also cleave the side chain of other hydroxysterols (e.g., 7-dehydrocholesterol), and 20- and 22-hydroxylates vitamin D.

The *CYP11A1* gene (Table 87.2) encodes an mRNA of 2 kb. Following translation, a 39-amino acid mitochondrial leader peptide that targets P450scc to the mitochondria is then proteolytically removed to yield a 482-amino acid protein. Forms of P450scc engineered to lack the mitochondrial leader are inactive,[30] demonstrating that the mitochondrial environment is required for activity. Expression of P450scc is induced in the adrenal zonae fasciculata and reticularis,[31] testis, and ovary by cAMP; and in the zona glomerulosa by intracellular calcium/protein kinase C.[32] In contrast, placental P450scc expression is constitutive and is caused at least in part by the long terminal repeat-binding protein family of transcription factors.[22] Side-chain cleavage activity and pregnenolone biosynthesis have been demonstrated in the rat and human brain[33]; and abundant P450scc expression is found in the rodent brain, especially in fetal life. Deletion of the gene for P450scc has been described in rabbits and mice,[34] abrogating all steroidogenesis and thus proving that P450scc is the only enzyme that can convert cholesterol to pregnenolone. While homozygous mutations in P450scc might be expected to be embryonic lethal by eliminating placental progesterone synthesis, patients have been described having P450scc mutations that typically retain partial enzymatic activity.[35]

P450c17 (Encoded by *CYP17A1*)

P450c17 catalyzes both the 17α-hydroxylase and 17,20-lyase reactions,[36] with additional regulatory mechanisms for the 17,20-lyase reaction.

The *CYP17A1* gene (Table 87.2) encoding P450c17 is expressed in the adrenals and gonads,[37] and leads to a single 2.1-kb mRNA species yielding a 57-kD protein in these tissues; mutations in this gene produce a spectrum of deficiencies in 17-hydroxysteroids and C19 steroids.[38]

Human P450c17 17-hydroxylates both pregnenolone and progesterone with approximately equal efficiency, but all other reactions show prominent differences among substrates. The 17,20-lyase activity is approximately 50 times more efficient for the 17α-hydroxypregnenolone-to-DHEA reaction than for the 17α-hydroxyprogesterone-to-androstenedione reaction.[39] Although the rate of the lyase reaction can be increased more than 10-fold by the addition of a molar equivalent of cytochrome b5,[39] the Δ^5 preference persists, and the lyase rate never quite reaches the rate of the hydroxylase reactions. Residue N202 of P450c17 serves as a hydrogen bond donor (NH) to the 3-keto oxygen of Δ^4 steroids or acceptor (CO) for the 3-hydroxyl hydrogen of Δ^5 steroids,[40] and mutation N202S eliminates the Δ^5 preference for the 17,20-lyase reaction.[41] In addition, human P450c17 16α-hydroxylates progesterone but not pregnenolone,[39] and, in the presence of cytochrome b5, diverts approximately 10% of pregnenolone metabolism to a Δ^{16} andiene product that is also formed by this pathway in pigs and that acts as a porcine pheromone. Although experiments to study the chemistry of human P450c17 often require manipulations that could be considered nonphysiologic, the remarkable consistency for substrate preferences and kinetic constants observed for the modified, solubilized P450c17 expressed in *Escherichia coli* and native P450c17 expressed in yeast microsomes, or intact COS-1 cells, or that obtained from human tissues and cells, serve to verify these conclusions.

In addition, the 5α-reduced C21 steroids dihydroprogesterone (5α-pregnane-3,20-dione) and allopregnanolone (5α-pregnan-3α-ol-20-one) are excellent substrates for the 17α-hydroxylase activity of P450c17.[42] Furthermore, 17α-hydroxylated allopregnanolone (5α-pregnane-3α,17α-diol-20-one) is the most efficient substrate yet identified for the 17,20-lyase activity of human P450c17, and its cleavage to androsterone is minimally dependent on cytochrome b5,[42] unlike 17α-hydroxypregnenolone metabolism to DHEA.[39] The conversion of 5α-pregnane-3α,17α-diol-20-one to androsterone by the 17,20-lyase activity of P450c17, first described in the testes of tammar wallaby pouch young,[43] provides an alternative or "backdoor" pathway to DHT (discussed later), by which DHT is produced without utilizing DHEA, androstenedione, and testosterone as intermediates. Consequently, the presence of 5α-reductases in steroidogenic cells does not preclude the production of C19 steroids, but rather paradoxically enhances the production of DHT by directing flux to 5α-reduced precursors of DHT. Human enzymes catalyze all of the other reactions required to complete this alternate route to DHT,[44] and the fetal adrenal gland produces 5α-reduced androgens, at least in some pathologic states. The backdoor pathway provides a major route to DHT in pathologic states in which 17α-hydroxyprogesterone accumulates, including 21-hydroxylase deficiency and P450 oxidoreductase deficiency.[45] Androgen production by the backdoor pathway probably contributes to the severe virilization in newborn girls with 21- and 11-hydroxlase deficiencies, while those with 3β-HSD2 deficiency, whose adrenal glands cannot make 17α-hydroxyprogesterone, are minimally virilized.[46] Furthermore, 11-oxygenated pregnanes are also substrates for human P450c17, which can generate 11-oxygenated androgens.[47]

The chemistry of P450c17-mediated hydroxylations is believed to proceed via the common iron oxene species and "oxygen rebound" mechanism proposed for prototypical P450 hydroxylations.[48] The mechanism(s) of the 17,20-lyase reaction involving a carbon–carbon bond cleavage, however, is not resolved, despite considerable study. The capacity of chemically formed iron oxene,[48] but not hydrogen peroxide alone, to support catalysis suggest that the same heme-oxygen complex might participate in both hydroxylations and the 17,20-lyase reaction, but spectroscopic studies suggest the participation of a peroxy-anion intermediate.[49] The sequential reactions follow distributive rather than processive kinetics,[50] meaning that the intermediate 17-hydroxysteroids are released from the active site prior to the 17,20-lyase reaction.

One consequence of the Δ^5 preference of the human enzyme for the 17,20-lyase reaction is that most human C19- and C18-steroids derive from DHEA as an intermediate. This Δ^5 preference allows for the phenomenon of adrenarche to occur in humans, an event that only takes place in large primates.[51] However, Δ^5-lyase activity is not sufficient for adrenarche to occur, because some monkeys (e.g., rhesus macaques) produce high amounts of DHEA throughout life, but most mammals (e.g. cattle, dogs, cats) never produce much DHEA.[51] The biochemistry of P450c17, with its differential regulation of 17α-hydroxylase and 17,20-lyase activities, provides clues to the genesis of this enigmatic process of adrenarche. P450c17 is a phosphoprotein, and phosphorylation selectively enhances the 17,20-lyase activity.[52] Mitogen-associated kinase p38α (MAPK14), can phosphorylate P450c17 and confer 17,20-lyase activity to recombinant P450c17 *in vitro* and in immortalized cell systems,[52] and this kinase activity is counterbalanced by protein phosphatase 2A, which in turn is regulated by cAMP via phosphoprotein SET.[53] Cytochrome b5 also augments 17,20-lyase activity, and high expression of b5 in the zona reticularis of monkeys and humans[54] suggests that the developmentally regulated expression of cytochrome b5 might be a key event. The transcriptional regulation of cytochrome b5 in the adrenal gland is similar to that of P450c17,[55] and the mechanisms enabling zone-specific expression have not been elucidated. Finally, limiting steroid flux to the Δ^5 pathway by low 3β-HSD activity in the zona reticularis (where most DHEA derives) potentiates the effect of increased 17,20-lyase activity.[54,56]

The initial description of 17α-hydroxylase deficiency was a case in which both 17α-hydroxylase and 17,20-lyase products were absent.[57] When the gene for human P450c17 was cloned, patients with 17α-hydroxylase deficiency were found to harbor mutations in the *CYP17A1* gene, and approximately 100 mutations scattered throughout the *CYP17A1* gene have been characterized.[38] Mutations in W406R and R362C accounted for the high prevalence of 17α-hydroxylase deficiency in Brazil,[58] and triple deletion of D487, S488, F489, and Y329fs accounted for most cases of 17α-hydroxylase deficiency in China and Southeast Asia.[38]

Isolated 17,20-lyase deficiency is extremely rare, mainly caused by mutations in arginines 347 and 358.[59,60] Computer modeling studies demonstrate that R347H and R358Q neutralize positive charges in the redox-partner binding site.[59] Biochemical studies confirm that mutations R347H and R358Q impair interactions of P450c17 with its electron donor POR and with cytochrome b5.[61] Therefore, these cases of isolated 17,20-lyase deficiency are not caused by an inability of the mutant enzymes to bind the intermediate 17α-hydroxypregnenolone, but rather the uncoupling of electron transfer with catalysis.[59,61,62] In contrast, mutation E305G has been shown to cause 17,20-lyase deficiency by selectively disrupting binding of 17α-hydroxypregnenolone and DHEA synthesis despite enhanced conversion of 17α-hydroxyprogesterone to androstenedione.[63] This unusual variant of isolated 17,20-lyase deficiency provides further genetic evidence that the flux of androgens derived from conversion of 17α-hydroxyprogesterone to androstenedione in the minor Δ4 pathway is not sufficient to form normal male external genitalia. Mutation V366M is thought to selectively impair 17,20-lyase activity by interfering with substrate binding and egress.[64] One of the first patients reported to have isolated 17,20-lyase deficiency bears a homozygous mutation in P450 oxidoreductase (G539R), further emphasizing the crucial role of efficient electron transfer in the 17,20-lyase reaction.[65] Apparent isolated 17,20-lyase deficiency may also be caused by specific mutations in cytochrome b5.[66,67]

The P450c17 inhibitors ketoconazole and abiraterone are used to treat castration-resistant prostate cancer (CRPC) by suppressing testosterone production beyond limits achieved with medical or surgical castration.[68] The x-ray crystal structures of modified human P450c17 with bound abiraterone or galeterone (TOK-001) show the azole nitrogens bound to the heme iron with the α-face of the steroid facing the heme and the A-ring draped against the I-helix.[69] These structures explain the 17α-hydroxylation chemistry of P450c17 and suggest that the 17,20-lyase reaction involves oxidation from the α-face as well.

KEY POINTS

- The 17-hydroxylase and 17,20-lyase activities of P450c17 qualitatively regulate the type of steroid hormones produced in a cell.
- Cytochrome b_5 augments the 17,20-lyase activity of P450c17 10-fold by increasing the coupling of electron transfer to substrate oxidation.

P450c21 (Encoded by *CYP21A2*)

Microsomal P450c21 performs adrenal 21-hydroxylation of the Δ4 steroids 17α-hydroxyprogesterone and progesterone, an essential step in the biosynthesis of both mineralocorticoids and glucocorticoids (see Fig. 87.2). The human P450c21 protein is found only in the adrenal glands; the extra-adrenal 21-hydroxylase activity found in other organs such as the liver and the aorta is not catalyzed by P450c21 but appears to be catalyzed by CYP2C9, CYP3A4, and possibly CYP2C19 and other enzymes as well.[70]

The locus containing the *CYP21* genes is among the most complex in the human genome and explains why 21-hydroxylase deficiency (affecting approximately 1 in 14,000 live births) is one of the most common autosomal-recessive diseases. The *CYP21A2* gene and the *CYP21A1P* pseudogene lie on chromosomal locus 6p21.1 in the midst of the human leukocyte antigen (HLA) locus. Because the HLA locus is highly recombinogenic, exchange between the *CYP21A1P* and *CYP21A2* loci is common. Thus, approximately 95% of 21-hydroxylase mutations derive from gene micro- or macroconversion events where some or all of the *CYP21A1P* pseudogene replaces the corresponding area of the *CYP21A2* gene, thus reducing the expression of the encoded P450c21 protein and/or impairing its activity.[71] Most of the remaining mutations causing 21-hydroxylase deficiency are *CYP21A2* gene deletions, again caused by recombination events; only approximately 1% of *CYP21A2* mutations do not occur via recombination. In addition, at least eight additional genes lie in this locus (Fig. 87.4), including the liver-specific C4A and C4B genes and the ubiquitously expressed tenascin X or *TNXB* gene, the disruption of which is one cause of Ehlers–Danlos syndrome (EDS). Occasionally, a patient with 21-hydroxylase deficiency and EDS will have a contiguous gene syndrome with tenascin-X deficiency as well.[72] Heterozygosity for TNXB causes a mild form of EDS characterized by joint hypermobility, and up to 7% of patients with classic CAH also have this form of EDS.[73]

It is clear that genotype consistently predicts phenotype in very severe and very mild cases of 21-hydroxylase deficiency. In contrast, patients with P450c21 variants (e.g., the common P30L and V281L mutations and the less common R339H and P453S mutations), which have 5% to 20% of wild-type activity, can have various phenotypes, implying additional factors that can modify the clinical manifestations of 21-hydroxylase deficiency.[71]

Unlike P450c17, P450c21 is not very sensitive to the abundance of POR or regulated via cytochrome b5. The x-ray crystal structure of bovine P450c21 contains two molecules of bound steroid, one with the steroid nucleus perpendicular to the heme ring, so positioning the C-21 hydrogen atoms directly above the heme iron for hydroxylation.[74] The second steroid molecule lies toward the periphery of the enzyme and is unlikely to undergo hydroxylation, but it may contribute to enzyme structure or allostery. The structure of human P450c21 contains only the progesterone molecule in the active site with V359 holding the steroid in place.[75] Mutation of V359 to the smaller residues alanine and glycine increases progesterone mobility and confers 16α-hydroxylase activity to the enzyme.[76] The 21-hydroxylase reaction of human P450c21 shows large intra- and intermolecular kinetic isotope effects,[77,78] meaning that C-H bond breaking is substantially rate-limiting. Impairments of substrate binding, catalysis, heme incorporation, and thermal stability are observed for the common mutations causing varying degrees of 21-hydroxylase deficiency.[79]

KEY POINTS

- The *CYP21* gene locus lies within the is highly human leukocyte antigen complex, and recombination exchange between the *CYP21A1P* and *CYP21A2* loci is common.
- 21-hydroxylase deficiency is one of the most common autosomal recessive disorders.
- Genotype predicts phenotype in very severe and very mild cases of 21-hydroxylase deficiency.

P450c11β and P450c11AS (Encoded by *CYP11B1* and *CYP11B2* in Humans)

The classical descriptions of distinct deficiencies in 11β-hydroxylase, 18-hydroxylase (previously called corticosterone methyl oxidase I or CMOI), and 18-oxidase (CMOII) suggested that three enzymes executed these three respective transformations. Analogous to the scenario for P450c17, a single enzyme[80] and corresponding gene were found in bovine adrenal glands that possessed all three activities. In contrast, humans have two genes named *CYP11B1* and *CYP11B2*[81] that encode the mitochondrial enzymes 11β-hydroxylase (P450c11β) and aldosterone synthase (P450c11AS), respectively, and rats, but not mice, have three functional *Cyp11b* genes. Although P450c11β and P450c11AS both possess 11β-hydroxylase activities, P450c11AS also performs the two oxygenations at C18 required for aldosterone biosynthesis.[82] P450c11β catalyzes a small amount of 18-hydroxylase but no 18-oxidase activity.[83] Mutations in *CYP11B1* cause 11β-hydroxylase deficiency,[84] whereas defects in *CYP11B2* cause variable aldosterone synthase deficiency, with salt wasting, failure to thrive, and hypotension.[85] Severe defects can impair all P450c11AS activities, leading to the clinical phenotype of CMOI deficiency,[82] while P450c11β provides 11β-hydroxylase activity in the zona fasciculata. Fortuitously, mutations R181W plus V386A, which mainly impair 18-oxidase activity, are found in patients and cause CMOII deficiency.[86]

Unlike the multistep distributive kinetics of P450scc and P450c17, P450c11AS uses a highly processive mechanism to synthesize aldosterone from deoxycorticosterone (DOC) without release of the intermediates.[87] Consequently, more aldosterone is formed and at a faster rate using DOC as the substrate than corticosterone, even though one additional step is required. Furthermore, only a trace of aldosterone is formed from 18-hydroxycorticosterone, presumably due to formation of a cyclic acetal in solution with unfavorable geometry, which may explain why a processive mechanism from DOC is essential for efficient aldosterone synthesis.[87]

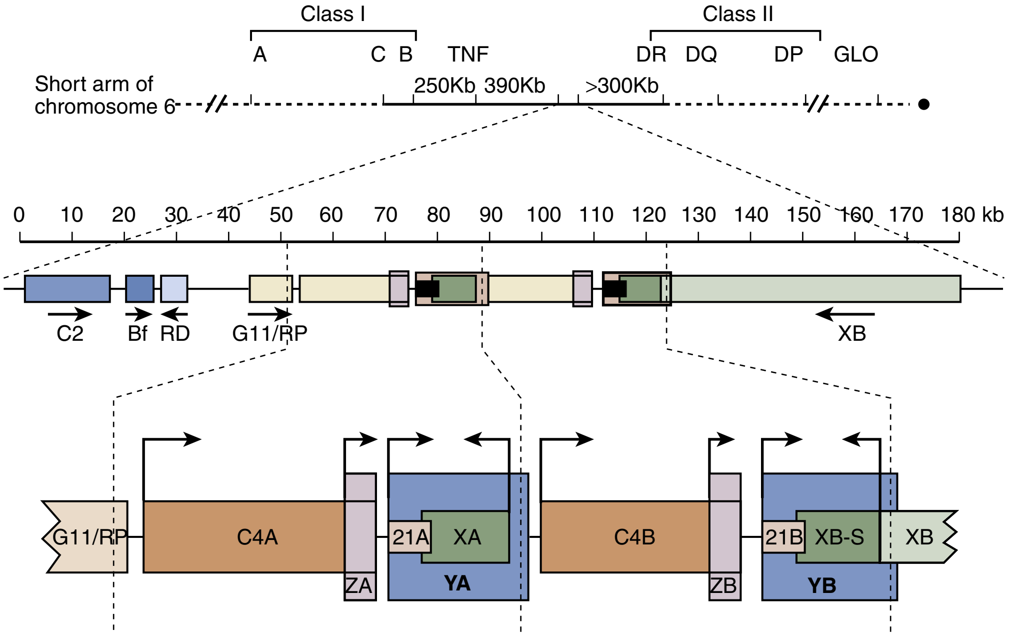


Figure 87.4 Genetic map of the human leukocyte antigen (HLA) locus containing the genes for P450c21. The *top line* shows the chromosome 6p21.1 region, with the telomere to the left and the centromere to the right. Most HLA genes are found in the class I and class II regions; the class III region containing the *CYP21* genes lies between these two. The *second line* shows the scale (in kb) for the diagram immediately below, showing *(from left to right)* the genes for complement factor C2, properdin factor Bf, and the RD and G11/RP genes of unknown function; *arrows* indicate transcriptional orientation. The *bottom line* shows the 21-hydroxylase locus on an expanded scale, including the *C4A* and *C4B* genes for the fourth component of complement, the *CYP21A1P* pseudogene (21A), and the active *CYP21A2* gene (21B) that encodes P450c21. XA, YA, and YB are adrenal-specific transcripts that lack open reading frames. The *TNXB* gene (XB) encodes the extracellular matrix protein tenascin-X; XB-S encodes a truncated adrenal-specific form of the tenascin-X protein whose function is unknown. ZA and ZB are adrenal-specific transcripts that arise within the C4 genes and have open reading frames, but it is not known if they are translated into protein; however, the promoter elements of these transcripts are essential components of the *CYP21A1P* and *CYP21A2* promoters. The *arrows* indicate transcriptional orientation. The *vertical dotted lines* designate the boundaries of the genetic duplication event that led to the presence of A and B regions.

The coding regions of the *CYP11B1* and *CYP11B2* genes share 93% amino acid identity and the same exonic gene structure found in all mitochondrial P450 genes. Despite the sequence similarities of these tandem genes, located within 40 kb on chromosome 8q24.3, the expression of P450c11AS is restricted to the adrenal zona glomerulosa, whereas P450c11β is found in the zona fasciculata and zona reticularis, emphasizing activity of zona-specific gene promoters. The regulation of P450c11β is driven mainly by cAMP in response to ACTH, whereas P450c11AS expression derives from potassium and angiotensin II activation of the protein kinase C pathway.[88] Thus, under normal circumstances, 18-hydroxylase and 18-oxidase activities are restricted to the zona glomerulosa, where 17-hydroxylase activity is low, limiting the repertoire of steroids that can undergo 18-oxygenation.

Although the organization of two highly homologous adjacent *CYP11B1* and *CYP11B2* genes on chromosome 8 is reminiscent of the *CYP21A1P* and *CYP21A2* genes, gene conversion in the *CYP11B* locus occurs rarely. Glucocorticoid-remediable aldosteronism (GRA) arises, however, when an unequal crossing over of the *CYP11B1* and *CYP11B2* genes creates a third, hybrid gene in which the ACTH-regulated promoter of *CYP11B1* drives expression of a chimeric protein with aldosterone synthase activity.[89] As a result, 18-hydroxylase and 18-oxidase activities are ectopically expressed in the zona fasciculata, leading to elevated renin-independent production of aldosterone, as well as 18-oxygenated metabolites of cortisol. The expression of this gene is suppressed by blunting ACTH production with glucocorticoids such as dexamethasone, which is used for diagnosis and treatment.[90]

The genetics of GRA has assisted in the precise identification of residues in P450c11AS that enable 18-oxygenase activities: residues 288, 296, 301, 302, 325, and, perhaps most importantly, 320 are critical for 18-oxygenase activities.[91,92] Therefore, crossovers 3′ to codon 320 do not enable aldosterone synthase activity. These key residues lie in or near the I-helix, which contains the catalytically important threonine residue implicated in oxygen activation for almost all P450s; thus, these mutations would be expected to alter active site geometry.

The x-ray crystal structure of modified human P450c11AS with bound DOC reveals the β-face of the steroid nucleus facing the heme iron, which enables oxygenation at C-11β, to form corticosterone, and subsequently at C-18, to form 18-hydroxycorticosterone and aldosterone.[93] In the x-ray crystal structure of P450c11AS with the inhibitor fadrozole bound to the active-site heme, the cyanophenyl ring of the *R*-fadrozole enantiomer forms hydrophobic interactions with the I-helix.[93] In the corresponding structure of the P450c11β-fadrozole complex, *S*-fadrozole adopts an opposite orientation pointed away from the I-helix.[94] The marked and unexpected differences for these two enzymes may contribute to their catalytic differences and enable the design of selective inhibitors. Osilodrostat, which is structurally related to *R*-fadrozole, potently inhibits both P450c11β and P450c11AS and is approved for the treatment of Cushing disease in the United States and Europe.[95] The activity of osilodrostat inhibiting P450c11β, and so lowering circulating cortisol levels, precluded its clinical development as an antihypertensive agent acting to reduce aldosterone secretion by inhibiting P450c11AS. Compounds that selectively inhibit P450c11AS but not P450c11β have begun to appear[96] and are being advanced into clinical trials.

In addition to 21-carbon steroids, both P450c11β and P450c11AS efficiently 11β-hydroxylate 19-carbon steroids, including androstenedione and testosterone.[93] In adrenal vein effluent, 11β-hydroxyandrostenedione is the third most abundant steroid after DHEAS and cortisol and is stimulated by ACTH.[97] In peripheral tissues, 11β-hydroxyandrostenedione is metabolized to 11-ketotestosterone, which is nearly as potent an agonist of the human androgen receptor as testosterone[98] and a dominant androgen in teleost fish and some reptiles. Accumulating evidence has demonstrated important contributions of these 11-oxygenated androgens to human physiology and disease[99] (see Pathways to 11-Oxygenated Androgens later).

KEY POINTS

- Expression of P450c11AS is restricted to the adrenal zona glomerulosa, whereas P450c11β is found in the zona fasciculata and zona reticularis.
- Regulation of P450c11β is driven mainly by cyclic adenosine monophosphate in response to adrenocorticotropic hormone (ACTH), whereas P450c11AS expression derives from potassium and angiotensin II activation of the protein kinase C pathway.
- Glucocorticoid-remediable aldosteronism arises from an unequal crossing over of the *CYP11B1* and *CYP11B2* genes to creates a hybrid gene in which the ACTH-regulated promoter of *CYP11B1* drives expression of a chimeric protein with aldosterone synthase activity.

P450aro (Aromatase)

The oxidative demethylation of C19 steroids, mainly androstenedione and testosterone, consumes three equivalents of molecular oxygen and NADPH, yielding formic acid and C18 steroids with an aromatic A-ring; hence the common name for this enzyme, *aromatase*. As is the case for P450scc, each subsequent oxygenation proceeds with greater efficiency with a distributive mechanism,[100] aiding in the completion of this transformation that is essential for estrogen biosynthesis in all animals.[26] The mechanism of this aromatization must account for the incorporation of the final oxygen atom from molecular oxygen into the formic acid by-product. The data are consistent with hydroxylation at C2 of 19-oxo-androstenedione via the iron oxene,[101] followed by an enzyme-assisted rearrangement and tautomerization of the intermediate dienone to the phenolic A-ring.[101] The x-ray structure of human aromatase with bound androstenedione supports this model, showing the steroid positioned with the β-face, C-19 methyl group, and A-ring adjacent to the heme iron.[102] Because all estrogen synthesis requires aromatase activity, the potent aromatase inhibitors letrozole, anastrozole, and exemestane are now first-line treatment for metastatic estrogen receptor-positive breast cancer in postmenopausal women.

P450aro is expressed in steroidogenic tissues (ovarian granuloma cells, placenta), in brain, and in nonsteroidogenic tissues, especially fat and bone.[26] The *CYP19A1* gene for P450aro spans over 75 kb (Table 87.2) and contains five different transcriptional start sites[26] with individual promoters that permit the tissue-specific regulation of expression in diverse tissues. P450aro is a glycoprotein, but glycosylation *per se* does not appear to affect activity.

Studies of patients with aromatase deficiency confirm that biologically significant estrogen synthesis derives entirely from this enzyme.[103] In fetuses homozygous for aromatase deficiency, the principal manifestation results from its deficiency in the placenta,[103] because ovarian steroidogenesis is quiescent during fetal life. Although huge amounts of estriol and estradiol are produced by the fetoplacental unit, estrogens are not needed for fetal development, the maintenance of pregnancy, or the onset of parturition; all of these processes proceed normally in fetuses lacking StAR, P450c17, or aromatase, or even in fetuses wholly lacking adrenal because of mutations in *SF-1* or *DAX-1*. However, in the absence of placental aromatase activity, androgenic C19 steroids derived from the fetal adrenal are passed into the maternal circulation, causing marked virilization of the mother.[103]

Furthermore, in pregnancies in which the mother has poorly treated 21-hydroxylase deficiency, maternal testosterone values can exceed 300 ng/dL (a midpubertal value for males), yet the fetus is not virilized,[104] because the maternal testosterone is efficiently metabolized to estradiol by placental aromatase. Thus, placental aromatase is a key enzyme in protecting both the fetus and mother from unwanted androgen exposure. After birth, individuals with aromatase deficiency grow normally and continue linear growth after completion of puberty, with males producing normal amounts of testosterone. However, when treated with estrogens, aromatase-deficient subjects fuse their epiphyses and cease linear growth.[105] These observations provide powerful evidence that bony maturation and epiphyseal fusion in children are mediated by estrogens, not androgens, even in males. These observations have led to the experimental use of aromatase inhibitors in various disorders of accelerated bone maturation.

REDOX PARTNER PROTEINS

The proteins collectively referred to as *redox partners* channel reducing equivalents from NADPH to the heme centers of P450 enzymes.[1] Many studies, however, suggest that these proteins promote catalysis by more than just their electron transfer properties. Because of this, the precise nature of the interactions of the P450s with their redox partners is of considerable importance.

Ferredoxin

Ferredoxin (FDX1) is encoded by a gene on chromosome 11q22 that spans over 30 kb. FDX1 is a small (14-kD), soluble Fe_2S_2 electron shuttle protein that resides either free in the mitochondrial matrix or is loosely bound to the inner mitochondrial membrane.[106] FDX1 is expressed in many tissues, and its expression in steroidogenic tissues is induced by cAMP in parallel with P450scc.[107]

Bovine FDX1 consists of two domains,[108] a core region and an interaction domain. The core region contains residues 1 to 55 and 91 to end (bovine numbering), including the four cysteines whose sulfur atoms tether the Fe_2S_2 cluster to the protein. Residues 56 to 90 form the interaction domain, which is a hairpin containing a helix at its periphery that includes acidic residues critical for the interaction of FDX1 with P450scc,[109] including aspartates 72, 76, and 79, plus glutamate 73). Similarly, the carboxylate group of D76 in FDX1 is cross-linked *in vitro* with lysine amino groups substituted for R370 or R366 of P450c11β and P450c11AS, respectively,[110] which is consistent with direct electrostatic pairings. A hydrophobic interaction with the adjacent L80 of FDX1 is also critical for binding and catalysis.[110] The Fe_2S_2 cluster lies in a protuberance in the molecule at the junction of its two domains. The charged residues of FDX1 cluster in the interaction domain, giving the molecule a highly negatively charged surface above the Fe_2S_2 cluster (see Fig. 87.3A). This description of the FDX1 molecule concurs with earlier studies that showed that overlapping sets of negative charges on FDX1 drive its interactions with positive charges on both P450scc and FDXR, which reduces oxidized FDX1. Because a preponderance of the evidence favors a model in which the same surface of FDX1 shuttles between FDXR and the P450 to transport electrons,[110,111] a model of how FDX1 interacts with FDXR should suggest how mitochondrial P450s interact with FDX1.

Ferredoxin Reductase

Like FDX1, ferredoxin reductase is widely expressed in human tissues, but its expression is two orders of magnitude higher in steroidogenic tissues.[112] The primary RNA transcript from the 11-kb *FDXR* gene (Table 87.2) on chromosome 17q24-q25 is alternatively spliced, generating two mRNA species that differ by only 18 bp, but only the protein encoded by the shorter mRNA is active in steroidogenesis.[113] Unlike most steroidogenic genes, the promoter for FDXR contains six copies of GGGCGGG sequences, which is the canonical binding site for the transcription factor SP-1 typically found in "housekeeping" genes. Accordingly, cAMP does not regulate transcription of the *FDXR* gene, as is the case for FDX1 and P450scc,[112] implying that FDXR plays additional roles in human physiology beyond steroidogenesis. Given their essential roles in the conversion of cholesterol to pregnenolone, no null mutations in FDXR or FDX1 have been described in humans.

Bovine FDXR also consists of two domains, each comprising a β-sheet core surrounded by α-helices.[114] The NADP(H)-binding domain is a compact region composed of residues 106 to 331 (bovine numbering), whereas the more open FAD domain, formed by the remaining N- and C-terminal residues, binds the dinucleotide portion of FAD across a Rossman fold with the redox-active flavin isoalloxazine ring abutting the NADP(H) domain. By analogy to related structures including glutathione and thioredoxin reductases, the nicotinamide ring of NADPH is modeled to lie adjacent to the flavin ring in position to transfer its two electrons to the FAD. Thus, intramolecular electron transfer occurs in the cleft formed by the angled apposition of these two domains. Within this cleft, basic residues abound, including arginines 240 and 244, which are important for interactions with FDX1.[113] Hypothetical docking of the two structures suggests that the negative surface of FDX1 fits elegantly into the positive surface of FDXR, even with NADP(H) bound.[114] These studies illustrate the similarities in the electrostatic interactions of FDX1 with both FDXR and the mitochondrial P450s.

P450 Oxidoreductase

The flavoprotein POR is expressed widely in human tissues and serves as the sole electron-transfer protein for all microsomal P450s, including xenobiotic-metabolizing hepatic P450s, steroidogenic P450s, and P450s found in other tissues such as the kidney and brain.[115] Crystallographic studies show that POR contains two lobes, one binding FAD and the other binding FMN, and a flexible N-terminus that tethers it to the endoplasmic reticulum.[116] NADPH is bound to the cofactor-binding domain above the FAD in a β-sheet–rich FAD domain, and an α-helical connecting domain joins the FAD and the FMN domains. A disordered "hinge" of approximately 25 residues lies between the FMN domain and the connecting domain, suggesting that the FMN and FAD domains can move substantially relative to each other. In the x-ray structure of rat liver POR, the FMN and FAD lie at the base of a cleft formed by the butterfly-shaped apposition of the FAD and FMN domains, reminiscent of the electron transfer surface of FDXR.[114] Nuclear magnetic resonance (NMR) and small-angle x-ray scattering studies[117] and ion mobility spectrometry confirm that, once the electrons reach the FMN domain, POR "opens up" via its hinge so that the FMN domain may dock with the redox-partner binding site of the P450 by electrostatic interactions (see Fig. 87.3). The surface of the electron-donating FMN domain is dominated by acidic residues, whereas the redox-partner binding site of P450 enzymes contain numerous basic residues. Negative charges of the FMN domain guide interactions with positive charges on the P450. The FMN approaches no closer than 18 Å from the heme, similar to the 16-Å distance of FAD from the Fe_2S_2 cluster in the modeled FDXR-FDX1 complex, and presumably similar to the distance of the heme from the Fe_2S_2 cluster in the P450-FDX1 complex.[114] These distances are too far for electrons to "jump" directly to the heme; rather, electron transfer apparently uses the polypeptide chain as a conduit.[118] Basic residues in the redox-partner-binding surface are crucial for interactions with POR and for electron transfer, and these positive charges in human P450c17 are critical for maximal 17,20-lyase activity.[61] Thus, these structures demonstrate several key principles of the electron transfer proteins involved in human steroidogenesis: NADPH and prosthetic groups lie at the interfaces of protein domains in which electron transfer occurs; the electron transfer surfaces are negatively charged to pair with positive charges on the P450s; the terminal electron transfer moiety (FMN domain or FDX1) must be mobile or soluble to pass electrons on to the P450; and electrons flow from the FMN or Fe_2S_2 cluster along the adjacent polypeptide chain to the heme.

Cytochrome P450 Oxidoreductase Deficiency: A Disorder Affecting Multiple P450 Enzymes. Beginning with a clinical report in 1985,[119] several patients have been described with clinical and hormonal findings, suggesting partial deficiencies of both P450c17 and P450c21. Some of these individuals were born to mothers who had become virilized during pregnancy, suggesting fetoplacental aromatase deficiency, and many also had the Antley–Bixler skeletal malformation syndrome characterized by craniosynostosis and radioulnar synostosis. Approximately half of patients with Antley–Bixler syndrome have normal steroidogenesis and normal genitalia; these subjects have dominant, gain-of-function mutations in the gene for fibroblast growth factor type 2 receptor (*FGFR2*); however, patients with Antley–Bixler syndrome who also have genital anomalies and disordered steroidogenesis do not have *FGFR2* mutations. The initial report described three patients with Antley–Bixler syndrome, genital ambiguity, and hormonal findings, suggesting partial deficiencies of 17α-hydroxylase and 21-hydroxylase, as well as a fourth patient who was phenotypically normal but had a similar hormonal profile; all had recessive, loss-of-function amino acid replacement mutations in POR.[120] One of these patients was born to a woman who had become virilized during the pregnancy, suggesting partial fetoplacental aromatase deficiency.

In vitro biochemical assays of the recombinant mutant POR proteins showed that the mutations in the Antley–Bixler subjects had severely impaired, but not totally absent, activity, whereas the mutations found in a phenotypically normal subject with amenorrhea were less severe.[120] Examination of the *POR* and *FGFR2* genes in a series of 32 patients established that the recessive *POR* mutations and the dominant *FGFR2* mutations segregate completely.[121] Approximately 100 POR-deficient patients have been described.[122] It appears unlikely that subjects will be found who are homozygous for null *POR* alleles, because knockouts of POR in mice cause embryonic lethality. POR is required for the activities of all hepatic drug-metabolizing P450 enzymes, and liver-specific ablation of the *Por* gene in mice results in phenotypically and reproductively normal animals with profoundly impaired drug metabolism.[123] Furthermore, the human *POR* mutations identified in patients interfere with *in vitro* drug metabolism by most hepatic P450 enzymes to varying degrees, and at least one report describes impaired metabolism of test drugs by a POR-deficient patient.[124] Consequently, it seems likely that some patients with POR deficiency may also metabolize drugs poorly.

The human *POR* gene, located on chromosome 7, consists of 16 exons (Table 87.2). The sequence of this gene in 842 normal persons from four ethnic groups revealed a high degree of polymorphism.[125] The coding sequence variant A503V, found on approximately 28% of all alleles, reduced the 17α-hydroxylase and 17,20-lyase activities of P450c17 to approximately 60% of normal[121,125] and some, but not all drug metabolism by CYP3A4 and CYP2D6, but had no measurable effect on the activities of P450c21 or hepatic CYP1A2 or CYP2C19. The POR promoter polymorphism –152C→A, found on 13% of Caucasian alleles, reduced *POR* gene transcription by about half in both adrenal and liver cells in culture, but the potential clinical significance of this is unknown.

Cytochrome b5

The small (12- to 17-kD) hemoprotein cytochrome b5 (CYB5A, b5) is found in two forms: the full-length protein is membrane-bound in the liver, and a soluble form lacking the C-terminal membrane anchor is found in erythrocytes. Full-length cytochrome b5 is expressed in both the adrenal and the gonads, where it can interact with P450c17; the adrenal expression is largely confined to the zona reticularis and may contribute to the genesis of adrenarche.[54] While cytochrome b5 appears to transfer the second electron for some P450 reactions,[126] apo-b5, which does not transfer electrons, also stimulates the 17,20-lyase activity of human P450c17.[39] These experiments suggest that cytochrome b5 does not act alone as an electron donor, but rather functions in concert with POR to somehow aid catalysis, yet b5 might substitute for POR in other P450 systems.

The soluble form of bovine b5 was one of the first proteins studied by x-ray crystallography, and a wealth of structural data for b5 have been acquired using molecular dynamics and NMR spectroscopy for both the holo- and apo-b5.[127,128] Analogous to FDX1, b5 consists of two domains, a heme-liganding core 1 domain (residues 40 to 65, bovine numbering); and a structural core 2 domain, from which the C-terminal membrane-anchoring helix extends. The heme extends more to the periphery of cytochrome b5 than does the Fe_2S_2 cluster of FDX1, and the entire surface is dominated by negatively charged residues rather than just one cluster of negative charges near the heme. In addition, the core 1 domain acquires considerable conformational flexibility in apo-b5, whereas the core 2 domain remains folded as in holo-b5.[127] Finally, the C-terminal membrane-spanning helix (exiting the core 2 domain) is required to stimulate the 17,20-lyase activity of human P450c17, but the signal peptide is not.[129] Genetic and biochemical studies have implicated basic residues in P450c17, including R347, R358, R449, and K89, as important for its interaction with cytochrome b5,[59,129,130] while E48 and E49 of cytochrome b5 are required for high 17,20-lyase activity.[131] The major effect of cytochrome b5 on the P450c17-POR complex to augment 17,20-lyase activity is to increase the coupling of electron transfer to product formation, from 5% to almost 50%.[62]

KEY POINTS

- Redox partner proteins play and essential role in transferring reducing equivalents from nicotinamide adenine dinucleotide phosphate to the heme centers of P450 enzymes.
- Ferredoxin reductase contains one flavin prosthetic group and transfers electrons to the iron-sulfur protein ferredoxin, the ultimate electron-transfer protein to the mitochondrial P450 enzymes.
- P450-oxidoreductase (POR) contains two flavins and is the sole electron-transfer protein for all microsomal P450 enzymes; mutations in POR impair multiple P450 activities.
- Cytochrome b5 augments the 17,20-lyase activity of the P450c17-POR complex.

STEROIDOGENIC DEHYDROGENASES AND REDUCTASES

3β-Hydroxysteroid Dehydrogenase/$\Delta^5\rightarrow\Delta^4$-Isomerases

Conversion of Δ^5 steroids into their Δ^4 congeners, a step required for the production of progestins, mineralocorticoids, glucocorticoids, and sex steroids, consists of two chemical transformations, both performed by the 3β-hydroxysteroid dehydrogenase/$\Delta^5\rightarrow\Delta^4$-isomerase (3βHSD) enzymes. The first reaction is the oxidation of the 3β-hydroxyl group to the ketone, and during this process NAD^+ is converted to NADH. The intermediate Δ^5-3-ketosteroid remains tightly bound to the enzyme with nascent NADH, and the presence of NADH in the cofactor-binding site activates the enzyme's second activity, the $\Delta^5\rightarrow\Delta^4$-isomerase activity.[132] Competition experiments have shown that the dehydrogenase and isomerase activities reside in a single active site,[133] yet these enzymes are often referred to by their dehydrogenase activity alone.

Although rodents contain multiple 3βHSD isoforms, human beings have only two active genes. The type 1 enzyme (3βHSD1) is expressed in the placenta, liver, brain, and some other tissues.[134] This isoform is required for placental progesterone production during pregnancy, which may explain why a deficiency of 3βHSD1 has never been described. In contrast, the type 2 enzyme (3βHSD2) is by far the principal isoform in the adrenal glands and the gonads.[135] Deficiency of 3βHSD2 causes the rare form of congenital adrenal hyperplasia known as *3βHSD deficiency*.[46] The presence of the type 1 isozyme in these patients helps to explain the paradox of why 46,XX individuals born with severe 3βHSD2 deficiency can virilize slightly *in utero* and can have elevated 17-hydroxyprogesterone (17OHP) in newborn screening tests. The 3βHSD block in the adrenal gland diverts Δ^5-steroids away from cortisol and toward DHEA; extraadrenal 3βHSD1 enables 17OHP and testosterone synthesis despite absent 3βHSD2 in the adrenal. The N367T allele of 3βHSD1 is variably present in all populations, and this variant has equivalent catalytic activity as wild-type 3βHSD1; however, the N-to-T substitution disrupts a ubiquitination site. In prostate cancer cells, the 367T allele is resistant to degradation and provides a selective advantage by maintaining the 3βHSD activity required to form active androgens from DHEA.[136]

The types 1 and 2 enzymes share 93.5% amino acid identity, and all biochemical studies comparing the two enzymes yield very similar results. The enzymes are strongly inhibited by Δ^4 products[137] and by synthetic Δ^4 steroids such as medroxyprogesterone acetate and the 3-keto-Δ^4-metabolite of abiraterone.[138] Compared to the type 1 enzyme, the type 2 enzyme is higher affinity/lower capacity (lower K_m, and V_{max}) and thus efficiently converts Δ^5 steroids pregnenolone, 17α-hydroxypregnenolone, DHEA, and androst-5-ene-3β,17β-diol to their Δ^4 congeners in the adrenals and gonads.[133] The enzymes are primarily membrane-bound and are found both in the microsomal and mitochondrial fractions during subcellular fractionation.[133]

Considerable evidence suggests that 3βHSD activity is an important factor in regulating adrenal production of DHEAS. The human fetal adrenal, which produces vast amounts of DHEAS, contains little 3βHSD immunoreactivity.[139] Furthermore, the expression of 3βHSD in the innermost regions of the adrenal cortex declines as the zona reticularis develops in childhood, and 3βHSD immunoreactivity is low in the zona reticularis of the adult rhesus macaque and of humans.[54] Thus, the development of an adrenal cell type (reticularis) that is relatively deficient in 3βHSD activity is a necessary component of adrenarche, in which adrenal production of the Δ^5 steroids DHEA and DHEAS rises exponentially.[140]

17β-Hydroxysteroid Dehydrogenases

There are at least 14 human 17β-hydroxysteroid dehydrogenase (17βHSD) isoforms; these isoforms vary widely in size, structure, substrate specificity, cofactor utilization, and physiologic functions.[141] This section focuses on the human isoforms that possess significant, rather than gratuitous, 17βHSD activity.

17β-Hydroxysteroid Dehydrogenase Type 1. In the late 1980s, three independent groups reported the cloning of 17βHSD1 cDNA, the first of any human HSD.[142] Located on chromosome 17q25 adjacent to a pseudogene, the *HSD17B1* gene (Table 87.2) encodes a 34-kD protein subunit that is expressed primarily in the placenta and in ovarian granulosa cells of developing follicles.[141] The enzyme, which is active only as a dimer, accepts mainly estrogens such as estrone, although it also has low catalytic activity for the conversion of androstenedione to testosterone and DHEA to androst-5-ene-3β,17β-diol.[143] Although the enzyme can oxidize 17β-hydroxysteroids in the presence of NAD+ *in vitro* at a high pH, the enzyme functions *in vivo* to reduce estrone to estradiol and 16α-hydroxyestrone to estriol.[143]

Sequence alignments with other members of the SDR family identified a Y-X-X-X-K active-site motif in residues 155 to 159, which was confirmed crystallographically.[144] The structure demonstrates that cofactor lies across the β-sheet core of the protein in a Rossman fold characteristic of all SDR enzymes. Steroid appears to dangle from the top of the enzyme almost perpendicular to the cofactor, with a hydrophobic pocket holding the body of the steroid in place while the 3-hydroxyl forms hydrogen bonds with H221 and E282. At the place where the steroid and cofactor meet, S142, Y155, and K159 help to form a proton-relay system that drives catalysis.

Because steroid flux to estrogens preferentially occurs via the aromatization of androstenedione to estrone, 17βHSD1 appears to be required for the conversion of estrone to biologically active estradiol in the ovary and placenta. This role has not been proven unequivocally, because no cases of human 17βHSD1 deficiency have been reported, although female mice with targeted disruption of the *Hsd17b1* gene show reduced fertility.[145] Such a disease is theoretically compatible with life, because fetuses with aromatase deficiency[103] and estrogen insensitivity (ERα mutations) are viable. Nevertheless, this enzyme is probably critical for ovulation and may be important in the pathogenesis and progression of estrogen-dependent breast cancers.[146]

17β-Hydroxysteroid Dehydrogenase Type 2. In contrast to the "activating" role of 17βHSD1 in the placenta and ovary, human endometrium inactivates estradiol by the conversion to estrone. The 17βHSD1 mRNA is not detected in the human uterus, but a related cDNA encoding microsomal *HSD17B2* was cloned[147] and found to be expressed in endometrium, placenta, and other tissues.[148] 17βHSD2 converts estradiol to estrone and oxidizes testosterone and DHT to their inactive 17-ketosteroid homologs androstenedione and 5α-androstanedione, respectively. The widespread tissue distribution and broad substrate specificity of 17βHSD2 suggests that its role in human physiology is to protect tissues from excessive exposure to active steroid hormones by oxidation to inactive 17-ketosteroids.[141] Human deficiency of this most active human inactivating (oxidizing) 17βHSD has not been described, but targeted disruption of the mouse *Hsd17b2* gene results in placental defects and embryonic lethality,[149] which suggests a broad and fundamental importance of this isoenzyme. The type 2 enzyme also oxidizes 20α-dihydroprogesterone to progesterone, but this activity is low relative to its 17βHSD activity.[147]

17β-Hydroxysteroid Dehydrogenase Type 3. A form of 46,XY disordered sex development (DSD) lacking the enzyme capable of reducing androstenedione to testosterone, was reported in the 1970s.[150] When the large, complex *HSD17B3* gene encoding 17βHSD3 was cloned, patients with the androgenic "17-ketosteroid reductase deficiency" were found to harbor mutations in this gene,[151,152] proving the central role of this enzyme in male sexual differentiation and marking 17βHSD3 as the only 17βHSD enzyme whose role in human physiology is genetically established by a deficiency syndrome. Nonetheless, patients with 17βHSD3 deficiency produce enough testosterone to virilize incompletely at puberty, suggesting that one or more additional human 17βHSD enzymes convert androstenedione to testosterone (see later). Similarly, the human ovary exports some testosterone despite an absence of 17βHSD3 expression, and women with 17βHSD3 deficiency produce normal amounts of androgens and estrogens.[153]

Unlike 17βHSD1, which has been the subject of intense biochemical study, relatively little is known about 17βHSD3 enzymology. This knowledge gap is at least in part caused by the very hydrophobic nature of the encoded 310-amino acid protein, hampering the expression of this enzyme in bacteria. From experiments in transiently transfected HEK-293 cells, 17βHSD3 reduces all of the C19 17-ketosteroids that serve as precursors of testosterone and DHT in human beings, including DHEA, 5α-androstanedione, and androsterone.[152] The conversion of DHEA to androst-5-ene-3β,17β-diol by 17βHSD3 probably contributes significantly to testicular testosterone synthesis. Estrogens such as estrone are poor substrates for human 17βHSD3.[143]

17β-Hydroxysteroid Dehydrogenase Type 4. Many additional HSD isoforms have been described in rodents and in humans, but the activities of these isoforms for steroids are generally poor. For example, the type 4 enzyme is a trifunctional protein located in peroxisomes, but its (oxidative) HSD activity toward estradiol is 106 times slower than its 3-hydroxyacyl-coenzyme A dehydrogenase activity.[154] Deficiency of the type 4 enzyme causes Zellweger syndrome, in which bile acid synthesis is disturbed but steroidogenesis is not affected. Thus, this enzyme has 17βHSD activity as one of its repertoire of transformations, but steroidogenesis is not its principal physiologic function.

17β-Hydroxysteroid Dehydrogenase Type 5 (Aldo-Keto Reductase Family 1 Member C3). Unlike 17βHSD types 1 to 4, which are SDR enzymes, the type 5 enzyme (AKR1C3) is an AKR enzyme that is expressed both in steroidogenic and nonsteroidogenic tissues.[155] Originally described as hepatic 3αHSD type 2 for its ability to reduce DHT to 5α-androstane-3α,17β-diol, this protein was later found also to have 17βHSD activity.[155] This enzyme, now known as AKR1C3 (see 3α-HSDs), may account for much of the extratesticular androstenedione-to-testosterone conversion, although its catalytic efficiency as a 17βHSD is poor[156] compared with its 20α-HSD activity with progesterone and DOC or its prostaglandin dehydrogenase activity, reducing PGH2 to PGF2α. Nevertheless, AKR1C3 is more highly expressed in the human fetal adrenal during the time of sexual differentiation than 17βHSD3[44] and may participate in androgen production, particularly in virilizing congenital adrenal hyperplasias. The postnatal adrenal also expresses low levels of AKR1C3, accounting for the small amount of testosterone secreted directly by the adrenal.[157]

17β-Hydroxysteroid Dehydrogenase Type 6. This enzyme is discussed under the 3α-hydroxysteroid dehydrogenases below.

Steroid 5α-Reductases

The conversion of testosterone to DHT in target tissues was described in the 1960s,[158] and studies using fibroblasts suggested that at least two human enzymes with different pH optima and genetics performed these transformations.[159] These initial results were confirmed when the genes encoding the type 1 and type 2 enzymes were cloned, and patients with clinical 5α-reductase deficiency were found to have deletions or mutations in the *SRD5A2* gene.[160] The two isoforms are very hydrophobic 30-kD microsomal proteins that share 50% identity. The type 1 enzyme is expressed in adult nongenital skin and liver and is not expressed significantly in fetal peripheral tissues, but the enzyme is expressed in the fetal testis and fetal adrenal very early in gestation.[44] Consequently, deficiency of the type 2 enzyme is not compensated for by the type 1 enzyme in the fetus.[161] The type 2 enzyme remains the predominant enzyme in genital skin, male accessory sex glands, and prostate, whereas the type 1 enzyme accounts for most of the hepatic 5α-reduction.

Although 5α-reductase activity is generally discussed in the context of male genital differentiation and androgen action, both isozymes reduce a variety of steroids in what are believed to be degradative pathways in humans. Progesterone, 17OHP, and related C21 steroids are the best substrates for both 5α-reductases, particularly the type 1, and cortisol, cortisone, corticosterone, and related compounds are also good substrates.[162] The 5α- (and 5β-) reduced steroids may be metabolized further and conjugated for excretion in the urine. Given the importance of 5α-reductase type 2 in prostate growth, inhibitors of the type 2 enzyme have been developed for the treatment of prostatic hyperplasia and the prevention of its recurrence after surgery. Finasteride selectively inhibits human 5α-reductase type 2, whereas dutasteride inhibits both the type 1 and type 2 isoenzymes. Both drugs are approved for treatment of prostatic hyperplasia in the United States.

Although the function of 5α-reductase type 2 is firmly established from the studies of its deficiency, which is a form of 46,XY DSD, the role of the type 1 isoform in humans is less clear, but its expression in the fetal adrenal and testis and the now-established role of the backdoor pathway in human male sexual development suggest that the type 1 enzyme might participate in normal and abnormal sexual development. Furthermore, the type 1 isoenzyme appears to be the dominant 5α-reductase expressed in most prostate cancers and thus controls the pathways of precursor flux to DHT.[163] Given the abundant expression of the type 1 isoform in liver, and its high activity with C21 steroids, this enzyme has been ascribed a role of degrading circulating C21 steroids in preparation for excretion in the urine. However, disruption of the *Srd5a1* gene in mice results in delayed parturition, a defect that can be rescued with 5α-androstane-3α,17β-diol. In immature mice, 5α-reductase type 1 is expressed both in the ovary and in the Leydig cells, and this enzyme participates in the testicular synthesis of 5α-androstane-3α,17β-diol via two pathways.[164] Whether 5α-reductases are expressed in postnatal human adrenal or gonads in normal physiology or in pathologic states is not known.

3α-Hydroxysteroid Dehydrogenases

The four major human 3α-hydroxysteroid dehydrogenases (3αHSDs) are AKR enzymes with reductive preferences that belong to the AKR1C family (AKR1C1, 2, 3, and 4). The 3αHSDs types 1, 2, 3, and 4 are trivial names for AKR1C4, 1C3, 1C2, and 1C1, respectively, which are clustered on chromosome 10p14-p15. Each enzyme has its characteristic tissue distribution[165] and repertoire of catalytic activities.[2,156] AKR1C3 also performs the 17βHSD reaction with androstenedione and is known also as 17βHSD5, and all of these AKR1C isoforms catalyze additional reactions, such as the 20α-reduction of pregnanes. In the brain, 3αHSDs reduce 5α-dihydroprogesterone to tetrahydroprogesterone (allopregnanolone), which is an allosteric activator of the GABA_A receptor-chloride channel complex with a nanomolar affinity[166] and is approved for the treatment of postpartum depression in the United States.[167] AKR1C4 is abundant in liver but has been found in adrenal and gonad; AKR1C3 was cloned from liver, prostate, and brain; AKR1C2 is found in the prostate and brain; and AKR1C1 is abundant in the uterus. The amino acid sequences of the type 2 and 3 isoenzymes differ by only a few residues and have common allelic variations. These minor differences in sequences, however, cannot be neglected, because these differences might alter substrate utilization.

In the central nervous system, selective serotonin reuptake inhibitors (e.g., the antidepressant drugs fluoxetine and paroxetine) directly lower the *K*m of rat brain type 2 3αHSD for 5α-dihydroprogesterone by almost 10-fold,[168] which explains why these drugs augment brain allopregnanolone concentrations, perhaps contributing to their antidepressant activity. In addition, x-ray crystallography has shown that the β subunit of the mammalian voltage-gated potassium channel is a tetrameric structure[169] in which each subunit closely resembles the rat liver 3αHSD (AKR1C9)[170] and even contains bound NADP^+ with high occupancy. Although the broader implications of this work are not yet known, these studies suggest a role of HSDs in coupling intracellular redox state to membrane excitation.

The 3αHSDs differ from the 11βHSDs, 3βHSDs, and 17βHSDs types 1 through 4 in several respects, because all reductive 3αHSDs are AKR enzymes rather than SDR enzymes.[2] As AKR enzymes, they function as monomers with a TIM-barrel structure; they bind cofactor with the nicotinamide ring draped across the mouth of the “barrel” rather than lying on a Rossman fold; and their kinetic mechanisms are highly ordered, with cofactor dissociation the final and rate-limiting step.[171] Tight NADP(H) binding derives from interaction of R276 with the 2′-phosphate, and mutation of R276 eliminates a conformational change associated with tight binding[172] and attenuates or reverses the preference for ketosteroid reduction in intact cells.[3] As shown in the structure of AKR1C9,[170] their active sites also contain tyrosine and lysine residues to facilitate proton transfer during catalysis, but these residues are distantly located in linear sequence rather than confined to the Y-X-X-X-K motif as in SDR enzymes.

In contrast to the reductive 3αHSDs, the oxidative 3αHSDs belong to the SDR family and show greatest similarity to the retinol dehydrogenase or cis-retinol/androgen dehydrogenase (RoDH/CRAD) subfamily.[173] Although several of these RoDH/CRAD enzymes show some 3αHSD activity, the most active enzyme has been called RoDH, the microsomal 3αHSD, 3(α→β)-hydroxysteroid epimerase, or formally 17βHSD6, whose *HSD17B6* cDNA was first cloned from prostate.[174] This enzyme converts the inactive C19 steroid 5α-androstane-3α,17β-diol to DHT, and thus likely catalyzes the final step in the backdoor pathway from 17OHP to DHT via androsterone.[44] Prolonged incubation of 3α-hydroxysteroids with cells transfected with the cDNA for 17βHSD6 or with microsomes containing the recombinant enzyme, however, yields subsequent 3-ketosteroid metabolites, including both 3α- and 3β-hydroxysteroids and 17β-hydroxysteroids.[175] Hence, this enzyme has complex catalytic flexibility, and may serve a variety of biological functions.

11β-Hydroxysteroid Dehydrogenases

The 11β-hydroxysteroid dehydrogenases (11βHSD) regulate the bioactivity of endogenous and synthetic glucocorticoids, and a comparison of the types 1 and 2 enzymes exemplifies some key principles of HSD enzymology (Table 87.3). Both enzymes are hydrophobic, membrane-bound proteins that bind cortisol/cortisone and corticosterone/11-dehydrocorticosterone, but otherwise their properties and physiologic roles differ substantially[176] (see Table 87.3). The type 2 enzyme shares only 21% sequence identity with 11βHSD1, whereas 11βHSD2 and 17βHSD2 share 37% identity and favor steroid oxidation *in vivo*. Thus, 11βHSD1 and 11βHSD2 are only distantly related members of the SDR family, yet they perform opposite functions in specific tissues in human physiology and pharmacology.

TABLE 87.3 Comparison of 11β-Hydroxysteroid Dehydrogenases Types 1 and 2

Property	Type 1	Type 2
Size	34 kD	41 kD
Orientation in ER	Luminal	Cytoplasmic
Expression	Liver, decidua, lung, gonad, pituitary, brain, fat, bone	Kidney, placenta, colon, salivary gland
Principal reaction	Reduction	Oxidation
Cofactor preference	NADPH via H6PDH	Cytoplasmic NAD^+
Substrate binding	Low affinity (K_m 0.2-2 μM)	High affinity (K_m 0.01-0.1 μM)
Inhibition by carbenoxolone	Moderate	Strong
Deficiency state	CRD (H6PDH > HSD11B1)	AME

ER, Endoplasmic reticulum; *NADPH,* nicotinamide adenine dinucleotide phosphate, reduced form; *NAD+*, nicotinamide adenine dinucleotide, oxidized form; *H6PDH,* hexose-6-phosphate dehydrogenase; *CRD,* cortisone reductase deficiency; *AME,* apparent mineralocorticoid excess.

The 34-kD type 1 enzyme (11βHSD1) is expressed in the liver, testis, lung, fat, and proximal convoluted tubule.[177] The type 1 enzyme catalyzes both the oxidation of cortisol using $NADP^+$ as a cofactor (Km 1–2 μM) and the reduction of cortisone using NADPH as a cofactor (Km 0.1–0.3 μM), with cortisone reduction being the dominant reaction in transfected cells.[176] Some synthetic glucocorticoids such as prednisone are 11-ketosteroids that must be reduced to their 11β-hydroxy derivatives to attain biologic activity, and these transformations are performed mainly in the liver by 11βHSD1. In contrast, when recombinant 11βHSD1 is studied *in vitro*, cortisol oxidation with $NADP^+$ is most efficient, and cortisone reduction is only achieved if $NADP^+$ is scrupulously removed with an enzymatic NADPH regeneration system.[178] Thus, the net flux of steroid driven by 11βHSD1 depends on the relative concentrations of available NADPH and $NADP^+$, which usually favors reduction in cells, especially given the high Km of the enzyme for cortisol.[178]

The mechanism for the discrepancy between the prominent oxidative preference *in vitro* and the reductive dominance *in vivo*, however, is more complex and derives from the localization of 11βHSD1 in the lumen of the endoplasmic reticulum.[179] In this compartment, the ratio of NADPH to $NADP^+$ is not maintained by the cytoplasmic $NADP^+$-coupled dehydrogenases (mainly glucose-6-phosphate dehydrogenase), but by hexose-6-phosphate dehydrogenase (H6PDH). Indeed, most patients with apparent cortisone reductase deficiency (CRD), which is characterized by high ratios of cortisone to cortisol and of their respective metabolites in blood and urine,[180] bear loss-of-function mutations in *H6PDH*[181] rather than *HSD11B1*. These mutations impair regeneration of the NADPH within the endoplasmic reticulum, which serves as cofactor for the 11βHSD1 reaction. The genetics and pathophysiology of CRD provide an excellent example of the critical role of nicotinamide cofactors in HSD function and biology and the molecular complexity of steroidogenesis.

In contrast, the 41-kD type 2 enzyme catalyzes the oxidation of cortisol and corticosterone using NAD^+, and, although this enzyme has a high affinity for its steroid substrates (K_m 0.01–0.1 μM),[182] catalysis of the reductive reactions by 11βHSD2 has not been conclusively demonstrated. In addition, 11βHSD2 also oxidizes 11β-hydroxyandrostenedione and 11β-testosterone to their 11-ketosteroids, as described in the 11-oxyandrogen pathway section later. Cortisol is a potent agonist at the mineralocorticoid (glucocorticoid type 2) receptor in the distal nephron, but its oxidized 11-keto derivative, cortisone, is not a mineralocorticoid. The reason cortisol does not act as a mineralocorticoid *in vivo*, even though cortisol concentrations can exceed aldosterone concentrations by three orders of magnitude, is because cortisol is enzymatically converted to cortisone in the cells lining the cortical and medullary collecting ducts. Thus, the type 2 enzyme inactivates the mineralocorticoid activity of cortisol in the kidney tubule,[183] and inactivating mutations in the type 2 enzyme cause a syndrome of apparent mineralocorticoid excess[184] (see Chapter 88). This enzyme system may be overwhelmed in cases of prodigious excess cortisol production, as seen in the ectopic ACTH syndrome, where cortisol acting at the mineralocorticoid receptor causes hypokalemia and hypertension. The presence of the type 2 enzyme in the placenta[182] also inactivates endogenous and synthetic corticosteroids such as prednisolone, allowing the use of these agents during pregnancy without affecting the fetus. In contrast, 9α-fluorinated steroids such as dexamethasone are minimally inactivated by the type 2 enzyme, primarily because of a shift in the oxidation/reduction preference rather than a reduction in affinity for the enzyme. It is this resistance to inactivation by placental 11βHSD2 that is essential for synthetic glucocorticoids to “cross the placenta” and to exert a pharmacologic effect on the fetus. Furthermore, the relatively high placental concentrations of $NADP^+$ may also favor the oxidative action of 11βHSD1, so that both placental enzymes protect the fetus from the high maternal concentrations of cortisol that occur during pregnancy.[176]

STEROID SULFONATION

The steroid sulfotransferase (SULT) enzymes are cytosolic enzymes that use 3′-phosphoadenine-5′-phosphosulfate (PAPS) as a sulfate donor. The SULT1E1 isoenzyme sulfonates the phenolic 3-hydroxyl group of estrogens, and the SULT2A1 isoenzyme sulfonates the Δ^5 steroids pregnenolone, 17-hydroxypregnenolone, androst-5-ene-3β, 17β-diol, and most importantly DHEA. SULT2A1 is abundant in the adrenal zona reticularis,[185] which efficiently converts most of the nascent DHEA to DHEAS. The sulfate donor PAPS derives from two molecules of ATP and sulfate, through the ATP sulfurylase and adenosine phosphosulfate (APS) kinases, the PAPS synthases types 1 and 2 (PAPSS1, PAPSS2). Whereas PAPSS1 is ubiquitously expressed, PAPSS2 expression is high in cartilage, adrenal, and liver. Loss-of-function mutations in PAPSS2 deplete the adrenal gland of PAPS and increase production of unconjugated DHEA. This circulating DHEA is a substrate for peripheral 3βHSD1, which yields adrenal-derived androgens. Rare cases of incomplete PAPSS2 deficiency have been described in girls with premature pubarche, advanced bone age, acne, hirsutism, and subsequent secondary amenorrhea.[186] These children have mild to moderate skeletal deformities, whereas complete PAPSS2 deficiency causes autosomal recessive brachylomia and severe bone disease, primarily affecting the spine.

KEY POINTS

- Conversion of Δ^5 steroids into their Δ^4 congeners, a step required for the production of progestins, mineralocorticoids, glucocorticoids, and sex steroids, consists of two chemical transformations, both performed by the 3β-hydroxysteroid dehydrogenase/$\Delta^5 \rightarrow \Delta^4$-isomerase enzymes.
- 17β-hydroxysteroid dehydrogenase isoforms play critical roles in the production and degradation of sex steroids.
- 3α-hydroxysteroid dehydrogenases are involved in the “backdoor pathway” to DHT.
- 11β-hydroxysteroid dehydrogenases catalyze the interconversion of cortisol/cortisone, and corticosterone/11-dehydrocorticosterone; 11βHSD2 also oxidizes 11β-hydroxyandrostenedione and 11β-testosterone to their 11-ketosteroids in the 11-oxyandrogen pathway.

PATHWAYS

Adrenal Steroidogenic Pathways

Diagrams of steroidogenic pathways, such as that shown in Fig. 87.2, typically combine the pathways from multiple cell types to give a comprehensive and inclusive illustration; however, such diagrams may be misleading, as dominant pathways by which specific steroids are synthesized differ in each steroidogenic cell type. The three major pathways of steroidogenesis in each zone of the human adrenal cortex are shown in Fig. 87.6. The adrenal zona glomerulosa is characterized by three features: its expression of angiotensin II receptors, its unique expression of P450c11AS, and its absence of P450c17 expression. This combination of factors permits the zona glomerulosa to produce aldosterone under regulation by the renin/angiotensin system. By contrast, the adrenal zona fasciculata does not express angiotensin II receptors or P450c11AS, but instead expresses the melanocortin type 2 receptor (MC2R, the receptor for ACTH) and P450c11β, which cannot convert 18-hydroxycorticosterone to aldosterone and has minimal capacity to convert corticosterone to 18-hydroxycorticosterone.[187] Both the zona glomerulosa and zona fasciculata express P450c21, but the zona fasciculata also expresses P450c17, which allows cortisol synthesis. The zona fasciculata, however, expresses very little b5[54]; consequently, P450c17 in the zona fasciculata catalyzes 17α-hydroxylation but much less 17,20-lyase activity. Thus, the zona fasciculata produces two glucocorticoids, mainly cortisol and smaller amounts of corticosterone, under the influence of ACTH. Patients with severe mutations in P450c17 cannot synthesize cortisol but instead increase corticosterone production (as do rodent adrenal glands, which normally lack P450c17), explaining why they are not glucocorticoid-deficient, despite the lack of cortisol (see Fig. 87.5). The adrenal zona reticularis also expresses MC2R but very little P450c21, P450c11β, or 3βHSD2, and as a result, the zona reticularis produces minimal amounts of cortisol. By contrast, high b5 expression in the zona reticularis[54] maximizes the 17,20-lyase activity of P450c17,[39] so that DHEA is produced and sulfated by SULT2A1 to DHEAS. As DHEA accumulates, small amounts are converted to androstenedione, and very small amounts of this androstenedione are converted to testosterone, probably by AKR1C3/17βHSD5. Thus, the pattern of steroid products secreted by each adrenal zone is determined by the enzymes produced in that zone and may be logically deduced from an understanding of their specific enzymatic properties.

Gonadal Steroidogenic Pathways

Testicular synthesis of testosterone follows a pathway that is similar to C19-steroid production in the adrenal zona reticularis, with the notable exceptions that Leydig cells express 3βHSD2 and 17βHSD3 but no SULT2A1, and the stimulus for steroidogenesis is transduced by the LH receptor rather than MC2R. Consequently, DHEA produced under LH stimulation is not sulfated but is readily converted to androstenedione and then testosterone, or via androst-5-ene-3β,17β-diol as the intermediate (Fig. 87.6). As in the adrenal gland, the principal pathway to C19 steroids is via Δ^5 steroids to DHEA; the Δ^4 pathway from 17OHP to androstenedione makes a minimal contribution.[39,63] By contrast, ovarian steroidogenesis is more complex, as the enzymatic steps are partitioned between the granulosa and theca cells, which surround the oocyte and form a follicle. In addition, the patterns of steroidogenesis vary during the cycle, directed mainly to estradiol in the follicular phase and to progesterone in the luteal phase (see Fig. 87.6). The key point in ovarian steroidogenesis is that granulosa cells do not express P450c17. Thus, in general, steroidogenesis is initiated in granulosa cells under the influence of LH, which, via cAMP, stimulates the expression of P450scc.[17] Pregnenolone and progesterone from granulosa cells diffuse into adjacent theca cells, where P450c17 and 3βHSD2 can act upon them to produce androstenedione. Some androstenedione is secreted or converted to testosterone (via AKR1C3/17βHSD5), and a portion returns to the granulosa cells, where androstenedione is converted to estrone and then to estradiol by P450aro and 17βHSD1, respectively. Thus, as with the three zones of the adrenal gland, the patterns of gonadal steroidogenesis are dictated by the cell-specific expression of specific steroidogenic enzymes.

The Backdoor Pathway to Dihydrotestosterone

An alternative pathway from 17OHP affords dihydrotestosterone when 17OHP is sequentially 5α- and 3α-reduced to 5α-pregnane-3α,17α-diol-20-one in steroidogenic cells expressing P450c17 (Fig. 87.7*A*).[43] Not only is 5α-pregnane-3α,17α-diol-20-one the most efficient substrate known for the 17,20-lyase activity of human P450c17, but unlike the conversion of 17α-hydroxypregnenolone to DHEA, its conversion to androsterone occurs efficiently without b5.[42] The resulting androsterone may be reduced to 5α-androstane-3α,17β-diol and then 3α-oxidized to dihydrotestosterone, probably by RoDH (17βHSD6). Thus this alternative pathway is a "backdoor pathway" to dihydrotestosterone, which is produced without the intermediacy of DHEA, androstenedione, and testosterone. The backdoor pathway enables production of potent androgens from 17OHP, despite the poor 17,20-lyase activity of human P450c17 for 17OHP, by using 5α-pregnane-3α,17α-diol-20-one as the substrate for the 17,20-lyase reaction.

The human backdoor pathway is relevant to normal physiology and to pathologic states. The only enzymatic activities unique to the backdoor pathway that are not found in "conventional" steroidogenic pathways are the oxidative and reductive 3αHSD activities, but at least one isoenzyme capable of catalyzing each requisite reaction is expressed in appropriate tissues of the human fetus.[44] Mass spectrometric analyses of urinary steroids in patients with POR deficiency confirm that all the steroidal intermediates in the backdoor pathway are produced in patients with POR deficiency.[188] Similar studies have shown that infants, children, and adults with 21-hydroxylase deficiency who overproduce 17OHP, the "gateway steroid" of the backdoor pathway, excrete all the expected urinary metabolites of the backdoor pathway steroids, including high amounts of androsterone (5α-reduced C19 steroid) relative to etiocholanolone (5β-reduced C19 steroid).[45] This androgen production by the backdoor pathway helps to explain why newborn girls with 21- and 11β-hydroxylase deficiencies can be severely virilized, while those with 3βHSD2 deficiency, who cannot generate much intraadrenal 17OHP, are minimally virilized.[46]

Pathways to 11-Oxygenated Androgens

In the testes of teleost fish, nascent testosterone is subject to 11β-hydroxylase and oxidative 11β-hydroxysteroid dehydrogenase activities, yielding 11-ketotestosterone as the major androgen product. The human Leydig cell expresses neither P450c11β nor 11βHSD2, but the human adrenal cortex expresses P450c11β in the zona fasciculata and zona reticularis, where androstenedione can be produced and 11β-hydroxylated. In fact, the concentration of 11β-hydroxyandrostenedione in adrenal vein effluent is higher than androstenedione,[97] and circulating concentrations of 11β-hydroxyandrostenedione exceed those of androstenedione by a factor of 1.5 to 3 in males and females throughout life.[98,189] Circulating 11β-hydroxyandrostenedione is efficiently oxidized via 11βHSD2 in the kidney,[190] yet little 11-ketoandrostenedione accumulates. Instead, 11-ketoandrostenedione is a much better substrate for AKR1C3 than androstenedione by roughly a factor of eight,[191] and thus generates 11-ketotestosteorne via two conversions of the adrenal precursor in peripheral tissues (Fig. 87.7*B*). Furthermore, 11-ketotestosterone is a potent agonist of the human androgen receptor, nearly as active as testosterone.[98]

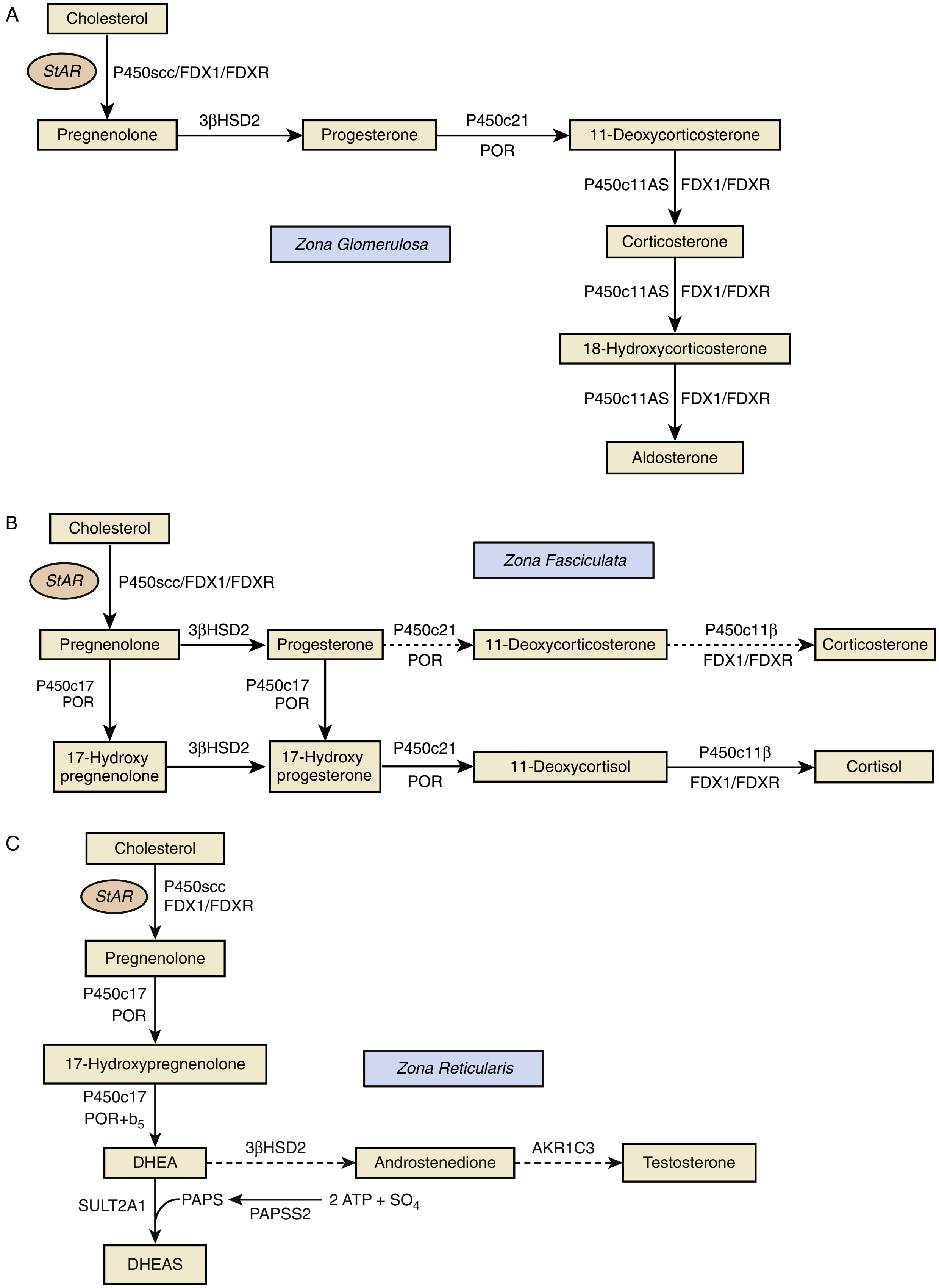

A
Cholesterol
StAR
P450scc/FDX1/FDXR
Pregnenolone
3βHSD2
Progesterone
P450c21
POR
11-Deoxycorticosterone
P450c11AS
FDX1/FDXR
Zona Glomerulosa
Corticosterone
P450c11AS
FDX1/FDXR
18-Hydroxycorticosterone
P450c11AS
FDX1/FDXR
Aldosterone
B
Cholesterol
StAR
P450scc/FDX1/FDXR
Zona Fasciculata
Pregnenolone
3βHSD2
Progesterone
P450c21
POR
11-Deoxycorticosterone
P450c11β
FDX1/FDXR
Corticosterone
P450c17
POR
P450c17
POR
17-Hydroxy pregnenolone
3βHSD2
17-Hydroxy progesterone
P450c21
POR
11-Deoxycortisol
P450c11β
FDX1/FDXR
Cortisol
C
Cholesterol
StAR
P450scc
FDX1/FDXR
Pregnenolone
P450c17
POR
17-Hydroxypregnenolone
Zona Reticularis
P450c17
POR+b_5
DHEA
3βHSD2
Androstenedione
AKR1C3
Testosterone
SULT2A1
PAPS
PAPSS2
2 ATP + SO_4
DHEAS

Figure 87.5 **Major steroidogenic pathways of the human adrenal cortex.** The conversion of cholesterol to pregnenolone is common to all three zones. **A,** In the zona glomerulosa, 3βHSD2 converts pregnenolone to progesterone. P450c17 is absent, but P450c21 produces 11-deoxycorticosterone, which is a substrate for P450c11AS (aldosterone synthase), the enzyme catalyzing 11β-hydroxylation and two 18-oxygenations to complete aldosterone synthesis. **B,** The zona fasciculata expresses P450c17 but little cytochrome b5, so pregnenolone is hydroxylated to 17α-hydroxypregnenolone but not cleaved to DHEA. Instead, 3βHSD2 yields 17α-hydroxyprogesterone, the preferred substrate for P450c21, which produces 11-deoxycortisol. P450c11β, which is unique to the zona fasciculata, completes the synthesis of cortisol. Corticosterone is normally a minor product *(dashed arrows)* derived from a parallel pathway without the action of P450c17. **C,** The zona reticularis has high P450c17 and cytochrome b5 but low 3βHSD2, so pregnenolone is sequentially oxidized to 17α-hydroxypregnenolone and then DHEA. SULT2A1 sulfates DHEA, and DHEAS is exported to the circulation.

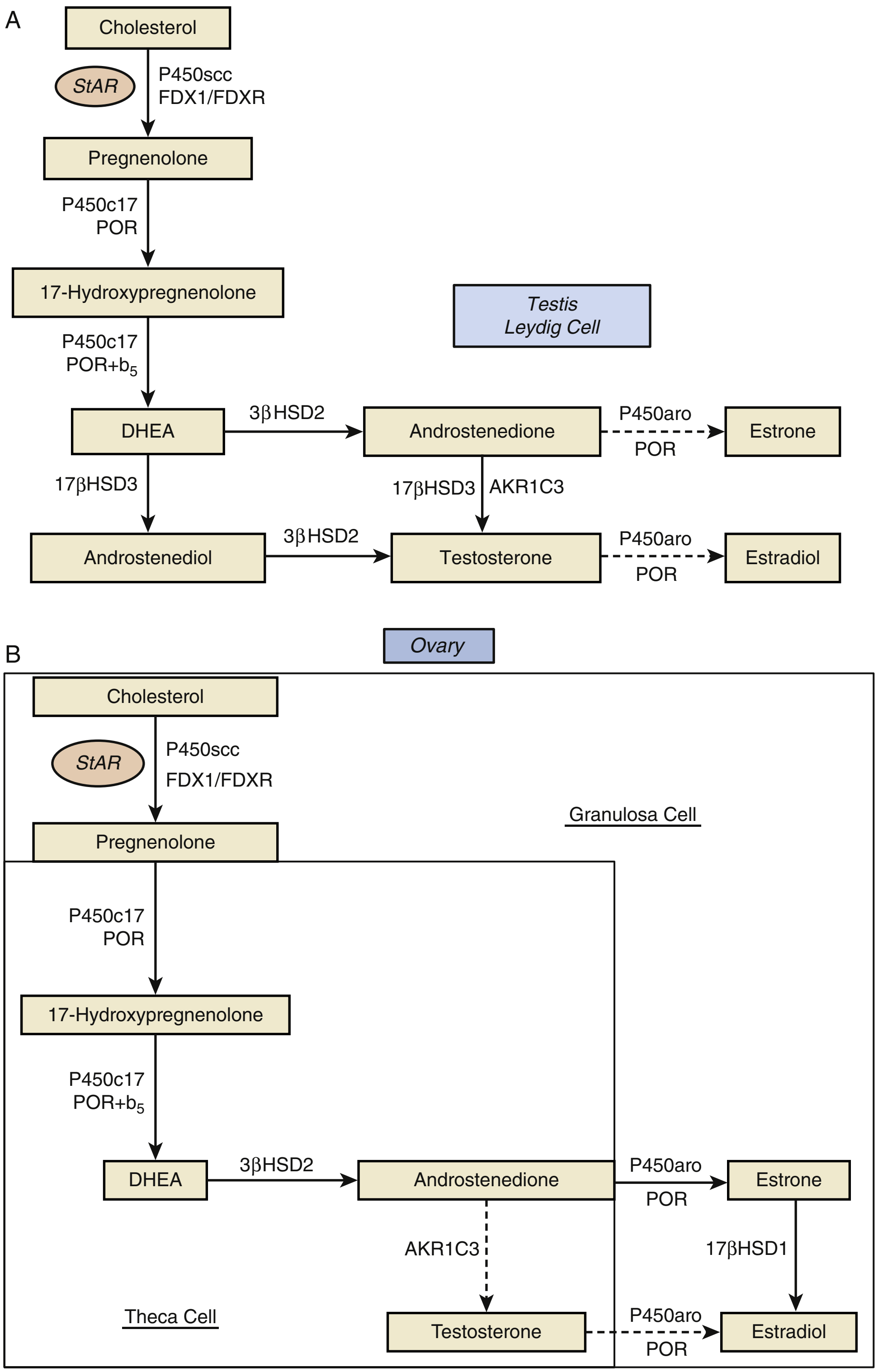

A
Cholesterol
StAR
P450scc
FDX1/FDXR
Pregnenolone
P450c17
POR
17-Hydroxypregnenolone
P450c17
POR+b5
DHEA
3βHSD2
Androstenedione
P450aro
POR
Estrone
17βHSD3
17βHSD3
AKR1C3
Androstenediol
3βHSD2
Testosterone
P450aro
POR
Estradiol
Testis
Leydig Cell
B
Ovary
Cholesterol
StAR
P450scc
FDX1/FDXR
Granulosa Cell
Pregnenolone
P450c17
POR
17-Hydroxypregnenolone
P450c17
POR+b5
DHEA
3βHSD2
Androstenedione
P450aro
POR
Estrone
AKR1C3
17βHSD1
Theca Cell
Testosterone
P450aro
POR
Estradiol

Figure 87.6 Major steroidogenic pathways of the gonads. Both the Leydig cell of the testis and the granulosa cell of the ovary convert cholesterol to pregnenolone, and the Leydig and theca cells have the capacity to synthesize androstenedione and testosterone. The major difference in their enzymatic machinery is abundant 17βHSD3 expression in Leydig cells (**A**), which efficiently completes testosterone synthesis in the testis, whereas androstenedione is the major C19 product of the theca cell (**B**). The granulosa cells of the ovary contain abundant aromatase and 17βHSD1, which complete the biosynthesis of estradiol. Minor pathways are shown with *dashed arrows.*

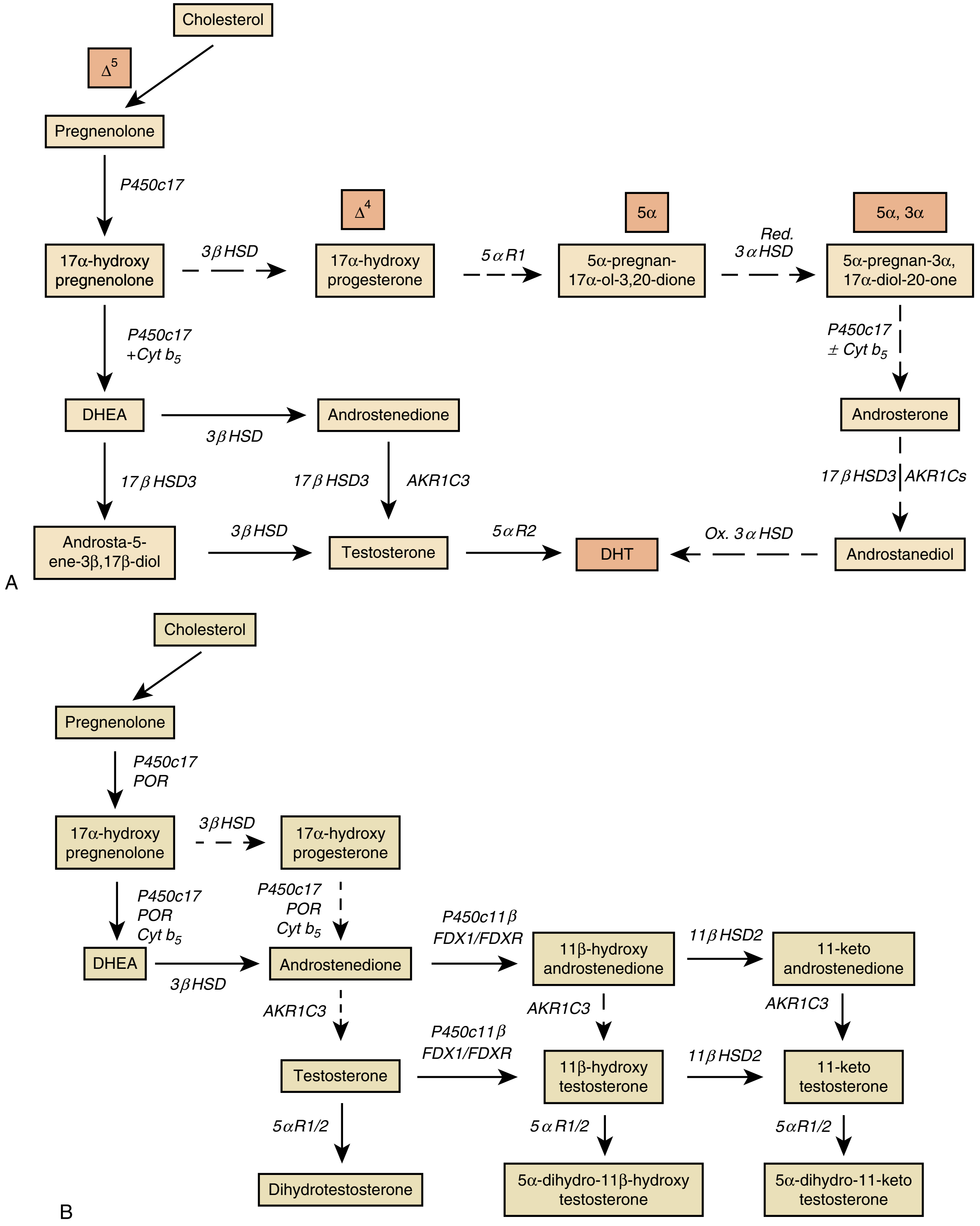

Cholesterol
Δ5
Pregnenolone
P450c17
Δ4
5α
5α, 3α
17α-hydroxy pregnenolone
3βHSD
17α-hydroxy progesterone
5αR1
5α-pregnan-17α-ol-3,20-dione
Red. 3αHSD
5α-pregnan-3α, 17α-diol-20-one
P450c17 +Cyt b5
P450c17 ± Cyt b5
DHEA
3βHSD
Androstenedione
Androsterone
17βHSD3
17βHSD3
AKR1C3
17βHSD3
AKR1Cs
Androsta-5-ene-3β,17β-diol
3βHSD
Testosterone
5αR2
DHT
Ox. 3αHSD
Androstanediol
A
Cholesterol
Pregnenolone
P450c17 POR
17α-hydroxy pregnenolone
3βHSD
17α-hydroxy progesterone
P450c17 POR Cyt b5
P450c17 POR Cyt b5
P450c11β FDX1/FDXR
DHEA
3βHSD
Androstenedione
11β-hydroxy androstenedione
11βHSD2
11-keto androstenedione
AKR1C3
AKR1C3
AKR1C3
P450c11β FDX1/FDXR
Testosterone
11β-hydroxy testosterone
11βHSD2
11-keto testosterone
5αR1/2
5αR1/2
5αR1/2
Dihydrotestosterone
5α-dihydro-11β-hydroxy testosterone
5α-dihydro-11-keto testosterone
B

Figure 87.7 Additional important pathways in human androgen biosynthesis. **A,** The two pathways to DHT using the different 17,20-lyase activities of human P450c17. In the conventional or Δ5-pathway *(solid arrows),* the 17,20-lyase activity of P450c17 requires cytochrome b_5 to efficiently convert 17α-hydroxyprogesterone to DHEA, and testosterone is reduced in target tissues by 5α-reductase 2 (5αR2) to DHT. In the "backdoor" or 5α,3α-pathway *(broken arrows)*, 5α-reduction by 5αR1 and 3α-reduction of C21 steroids occurs in the steroidogenic tissue prior to the 17,20-lyase reaction. In the best characterized pathway, 5α-pregnane-3α,17α-diol-20-one is cleaved to androsterone without requiring cytochrome b5 and reduced to 5α-androstane-3α,17β-diol. The 5α-androstane-3α,17β-diol is exported from the testis and metabolized to DHT by oxidative 3αHSDs (Ox. 3αHSD, probably 17βHSD6). Note that testosterone is not an intermediate in the backdoor pathway to DHT, that different isoforms of 5α-reductase appear to be involved in the two pathways, and that both reductive and oxidative 3αHSDs are required for the backdoor pathway. **B,** The 11-oxygenated androgen pathway. The major substrate is androstenedione, which is 11β-hydroxylated via P450c11β to 11β-hydroxyandrostenedione in the adrenal cortex. Primarily in peripheral tissues, the enzyme 11βHSD2 oxidizes this adrenal product to 11-ketoandrostenedione, which is an excellent substrate for AKR1C3. The 17β-reduction via AKR1C3 yields 11-ketotestosterone, a potent androgen, which can be further 5α-reduced to 5α-dihydro-11-ketotestosterone. Additional minor pathways are also shown.

Continued

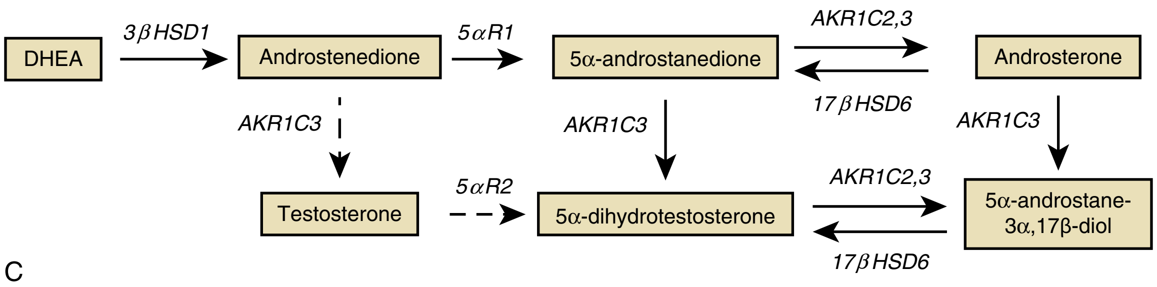


Figure 87.7, cont'd C, DHEA metabolism in castration-resistant prostate cancer cells. DHEA is converted to androstenedione via 3βHSD1, which is then 5α-reduced by 5α-reductase type 1 (5αR1) to 5α-androstanedione. AKR1C3 completes the pathway to dihydrotestosterone, and AKR1C enzymes plus 17βHSD6 reversibly convert these 5α-reduced steroids to other side products.

In children with premature adrenarche[98] and premature pubarche,[192] circulating 11-ketotestosterone is elevated compared to age-matched controls and 2- to 3-fold higher than testosterone, regardless if DHEAS is elevated. Thus, while DHEAS might be the traditional biomarker of adrenarche, it is 11-ketotestosterone that is the dominant circulating and biologically active androgen that yields the physical manifestations of adrenarche. Circulating concentrations of both 11β-hydroxyandrostenedione and 11-ketotestosterone exceed those of androstenedione and testosterone, respectively, in conditions of adrenal-derived androgen excess with preserved 11β-hydroxylase activity, most prominently undertreated 21-hydroxylase deficiency.[193,194] Note that 11-oxygenated androgen production is not elevated in 11β-hydroxylase deficiency or 3β-hydroxysteroid dehydrogenase deficiency, because 11β-hydroxylase activity is low or androstenedione is generated outside the adrenal cortex, respectively.

In addition, circulating concentrations of 11β-hydroxyandrostenedione exceed those of androstenedione by at least a factor of two in most men and women,[189] whereas 11-ketotestosterone is the dominant circulating androgen in prepubertal children[98] and is slightly higher than testosterone most women.[189] Furthermore, while circulating concentrations of DHEAS, androstenedione, and testosterone all decline with age in men and women after early adulthood, those of 11β-hydroxyandrostenedione and 11-ketotestosterone are preserved without much change in healthy adults until the ninth decade of life.[189] These observations are paralleled during aging by changes in the different zones of the adrenal cortex, with an expansion of the zona fasciculata, despite reduction in zona reticularis and the zona glomerulosa (in men), so allowing preservation of 11β-hydroxylase activity that may then account for the preservation of 11β-hydroxyandrostenedione and 11-ketotestosterone levels.[195] Additional studies demonstrate that the 11-oxygenated androgens are substrates for the 5α-reductases (Fig. 87.7*B*), which amplifies the activity of 11-ketotestosterone.[196] Thus, the 11-oxyandrogens are not curiosities or minor adrenal side-products, but rather important contributors to human physiology in children and women and to disease states, including 21-hydroxylase deficiency, polycystic ovary syndrome, prostate cancer, and possibly to other conditions related to advanced age.[99]

Androgen Synthesis in Prostate Cancer

The classical paradigms of steroidogenesis reflect the enzymes and pathways present in endocrine glands, which generate enough steroids to maintain circulating concentrations high enough to activate their cognate receptors in various tissues. Much less flux is required for a cell to generate biologically significant amounts of hormones—either from cholesterol or from circulating intermediates—to activate the receptors in the same tissues, in an autocrine or intracrine manner. Under these circumstances, minor reactions, different enzymes, and modified pathways may dominate. CRPC is an example of this scenario, in which minute amounts of testosterone or dihydrotestosterone are sufficient to fuel disease progression. Most of the cognate mRNAs for the enzymes and proteins necessary to synthesize dihydrotestosterone from cholesterol are upregulated in CRPC metastases relative to primary tumors, and androgen concentrations in the metastases are preserved.[197] Given that the abundance of most of the enzymes in the early steps is still quite low, however, it is not clear whether the majority of androgens in CRPC cells derive entirely from de novo synthesis or from metabolism of circulating intermediates.[68] Abundant circulating DHEA and DHEAS could serve as the precursor, requiring only two to four steps to testosterone and dihydrotestosterone. The conversion of DHEA to dihydrotestosterone has been well documented in CRPC samples and in prostate cancer cell lines derived from such tumors.[198] In these cells, the dominant pathway is oxidation of DHEA to androstenedione via 3βHSD1 (not 3βHSD2), then 5α-reduction of androstenedione to 5α-androstanedione via 5α-reductase type 1 (not type 2), followed by reduction of 5α-androstanedione to dihydrotestosterone via one or more 17βHSDs (Fig. 87.8*C*).[163] In addition, dihydrotestosterone production via the "backdoor" pathway appears to occur,[68] and oxidation of 5α-androstane-3α,17β-diol to dihydrotestosterone has been demonstrated.[199] To complicate matters, AKR1C3, which may catalyze the critical (activating) 17β-reduction of 17-ketosteroids to testosterone and dihydrotestosterone, also catalyzes the (inactivating) 3α-reduction of dihydrotestosterone, as does AKR1C2.[199] Consequently, it is difficult to predict how changes in enzyme expression translate to shifts in the delicate balance between intracellular androgen synthesis and degradation. The evolving biology of androgen synthesis in CRPC continues to revise our traditional concept of steroidogenic pathways and to identify additional potential therapeutic targets.

KEY POINTS

- In the adrenal, zone-specific expression of differing steroidogenic enzymes and receptors to adrenocorticotropic hormone and angiotensin II determine the specific repertoire and regulation of the steroids synthesized.
- The human backdoor pathway is relevant to normal physiology and to pathologic states.
- 11-Oxygenated androgens are important in disease states, including 21-hydroxylase deficiency, polycystic ovary syndrome, prostate cancer, and possibly to other conditions related to advanced age.

REFERENCES

1. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32:81–151.
2. Penning TM, Wangtrakuldee P, Auchus RJ. Structural and functional biology of aldo-keto reductase steroid-transforming enzymes. *Endocr Rev*. 2019;40:447–475.
3. Papari-Zareei M, Brandmaier A, Auchus RJ. Arginine 276 controls the directional preference of AKR1C9 (rat liver 3α-hydroxysteroid dehydrogenase) in human embryonic kidney 293 cells. *Endocrinology*. 2006;147. 1591-1517.
4. Sherbet DP, Guryev OL, Papari-Zareei M, et al. Biochemical factors governing the steady-state estrone/estradiol ratios catalyzed by human 17β-hydroxysteroid dehydrogenases types 1 and 2 in HEK-293 cells. *Endocrinology*. 2009;150:4154–4162.
5. Clark BJ, Wells J, King SR, et al. The purification, cloning and expression of a novel luteinizing hormone-induced mitochondrial protein in MA-10 mouse Leydig tumor cells. Characterization of the steroidogenic acute regulatory protein (StAR). *J Biol Chem*. 1994;269:28314–28322.
6. Lin D, Sugawara T, Strauss III JF, et al. Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science*. 1995;267:1828–1831.
7. Bose HS, Sugawara T, Strauss III JF, et al. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. *N Engl J Med*. 1996;335:1870–1878.
8. Caron K, Soo S-C, Wetsel W, et al. Targeted disruption of the mouse gene encoding steroidogenic acute regulatory protein provides insights into congenital lipoid adrenal hyperplasia. *Proc Natl Acad Sci USA*. 1997;94:11540–11545.
9. Arakane F, Sugawara T, Nishino H, et al. Steroidogenic acute regulatory protein (StAR) retains activity in the absence of its mitochondrial targeting sequence: implications for the mechanism of StAR action. *Proc Natl Acad Sci USA*. 1996;93:13731–13736.
10. Bose H, Lingappa VR, Miller WL. Rapid regulation of steroidogenesis by mitochondrial protein import. *Nature*. 2002;417:87–91.
11. Tsujishita Y, Hurley JH. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Biol*. 2000;7:408–414.
12. Baker BY, Yaworsky DC, Miller WL. A pH-dependent molten globule transition is required for activity of the steroidogenic acute regulatory protein, StAR. *J Biol Chem*. 2005;280:41753–41760.
13. Baker BY, Epand RF, Epand RM, et al. Cholesterol binding does not predict activity of the steroidogenic acute regulatory protein, StAR. *J Biol Chem*. 2007;282:10223–10232.
14. Artemenko IP, Zhao D, Hales DB, et al. Mitochondrial processing of newly synthesized steroidogenic acute regulatory protein (StAR), but not total StAR, mediates cholesterol transfer to cytochrome P450 side chain cleavage enzyme in adrenal cells. *J Biol Chem*. 2001;276:46583–46596.
15. Chung JY, Chen H, Midzak A, et al. Drug ligand-induced activation of translocator protein (TSPO) stimulates steroid production by aged brown Norway rat Leydig cells. *Endocrinology*. 2013;154:2156–2165.
16. Morohaku K, Pelton SH, Daugherty DJ, et al. Translocator protein/peripheral benzodiazepine receptor Is not required for steroid hormone biosynthesis. *Endocrinology*. 2014;155:89–97.
17. Voutilainen R, Tapanainen J, Chung B, et al. Hormonal regulation of P450scc (20,22-desmolase) and P450c17 (17α-hydroxylase/17,20-lyase) in cultured human granulosa cells. *J Clin Endocrinol Metab*. 1986;63:202–207.
18. Tsigos C, Arai K, Hung W, et al. Hereditary isolated glucocorticoid deficiency is associated with abnormalities of the adrenocorticotropin receptor gene. *J Clin Invest*. 1993;92:2458–2461.
19. Martens JW, Verhoef-Post M, Abelin N, et al. A homozygous mutation in the luteinizing hormone receptor causes partial Leydig cell hypoplasia: correlation between receptor activity and phenotype. *Mol Endocrinol*. 1998;12:775–784.
20. Shenker A. G protein-coupled receptor structure and function: the impact of disease-causing mutations. *Baillieres Clin Endocrinol Metab*. 1995;9:427–451.
21. Parker KL, Schimmer BP. Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocr Rev*. 1997;18:361–377.
22. Huang N, Miller WL. Cloning of factors related to HIV-inducible LBP proteins that regulate steroidogenic factor-1-independent human placental transcription of the cholesterol side-chain cleavage enzyme, P450scc. *J Biol Chem*. 2000;275:2852–2858.
23. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*. 1994;77:481–490.
24. Nachtigal MW, Hirokawa Y, Enyeart-VanHouten DL, et al. Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell*. 1998;93:445–454.
25. Hammer GD, Krylova I, Zhang Y, et al. Phosphorylation of the nuclear receptor SF-1 modulates cofactor recruitment: integration of hormone signaling in reproduction and stress. *Mol Cell*. 1999;3:521–526.
26. Santen RJ, Brodie H, Simpson ER, et al. History of aromatase: saga of an important biological mediator and therapeutic target. *Endocr Rev*. 2009;30:343–375.
27. Tuckey RC, Cameron KJ. Catalytic properties of cytochrome P-450scc purified from the human placenta: comparison to bovine cytochrome P-450scc. *Arch Biochem Biophys*. 1993;1163:185–194.
28. Yoshimoto FK, Jung IJ, Goyal S, et al. Isotope-labeling studies support the electrophilic compound I iron active species, FeO(3+), for the carbon-carbon bond cleavage reaction of the cholesterol side-chain cleavage enzyme, cytochrome P450 11A1. *J Am Chem Soc*. 2016;138:12124–12141.
29. Strushkevich N, MacKenzie F, Cherkesova T, et al. Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system. *Proc Natl Acad Sci USA*. 2011;108:10139–10143.
30. Black SM, Harikrishna JA, Szklarz GD, et al. The mitochondrial environment is required for activity of the cholesterol side-chain cleavage enzyme, cytochrome P450scc. *Proc Natl Acad Sci USA*. 1994;91:7247–7251.
31. John ME, John MC, Boggaram V, et al. Transcriptional regulation of steroid hydroxylase genes by corticotropin. *Proc Natl Acad Sci USA*. 1986;83:4715–4719.
32. Barrett PQ, Bollag WB, Isales CM, et al. The role of calcium in angiotensin II-mediated aldosterone secretion. *Endocrin Rev*. 1989;10:496–518.
33. Baulieu EE. Neurosteroids: of the nervous system, by the nervous system, for the nervous system. *Recent Prog Horm Res*. 1997;52:1–32.
34. Hu MC, Hsu NC, El Hadj NB, et al. Steroid deficiency syndromes in mice with targeted disruption of Cyp11a1. *Mol Endocrinol*. 2002;16:1943–1950.
35. Tee MK, Abramsohn M, Loewenthal N, et al. Varied clinical presentations of seven patients with mutations in CYP11A1 encoding the cholesterol side-chain cleavage enzyme, P450scc. *J Clin Endocrinol Metab*. 2013;98:713–720.
36. Zuber MX, Simpson ER, Waterman MR. Expression of bovine 17α-hydroxylase cytochrome P450 cDNA in non-steroidogenic (COS-1) cells. *Science*. 1986;234:1258–1261.
37. Chung BC, Picado-Leonard J, Haniu M, et al. Cytochrome P450c17 (steroid 17α-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. *Proc Natl Acad Sci U S A*. 1987;84:407–411.
38. Auchus RJ. Steroid 17-hydroxylase and 17,20-lyase deficiencies, genetic and pharmacologic. *J Steroid Biochem Mol Biol*. 2017;165:71–78.
39. Auchus RJ, Lee TC, Miller WL. Cytochrome b_5 augments the 17,20 lyase activity of human P450c17 without direct electron transfer. *J Biol Chem*. 1998;273:3158–3165.
40. Petrunak EM, DeVore NM, Porubsky PR, et al. Structures of human steroidogenic cytochrome P450 17A1 with substrates. *J Biol Chem*. 2014;289:32952–32964.
41. Gregory MC, Mak PJ, Khatri Y, et al. Human P450 CYP17A1: control of substrate preference by asparagine 202. *Biochemistry*. 2018;57:764–771.
42. Gupta MK, Guryev OL, Auchus RJ. 5α-reduced C_{21} steroids are substrates for human cytochrome P450c17. *Arch Biochem Biophys*. 2003;418:151–160.
43. Wilson JD, Auchus RJ, Leihy MW, et al. 5α-androstane-3α,17β-diol is formed in tammar wallaby pouch young testes by a pathway involving

5α-pregnane-3α,17α-diol-20-one as a key intermediate. *Endocrinology*. 2003;144:575–580.

44. Reisch N, Taylor AE, Nogueira EF, et al. Alternative pathway androgen biosynthesis and human fetal female virilization. *Proc Natl Acad Sci USA*. 2019;116:22294–22299.
45. Kamrath C, Hochberg Z, Hartmann MF, et al. Increased activation of the alternative "backdoor" pathway in patients with 21-hydroxylase deficiency: evidence from urinary steroid hormone analysis. *J Clin Endocrinol Metab*. 2012;97:E367–E375.
46. Moisan AM, Ricketts ML, Tardy V, et al. New insight into the molecular basis of 3β-hydroxysteroid dehydrogenase deficiency: identification of eight mutations in the HSD3B2 gene in eleven patients from seven new families and comparison of the functional properties of twenty-five mutant enzymes. *J Clin Endocrinol Metab*. 1999;84:4410–4425.
47. van Rooyen D, Yadav R, Scott EE, et al. CYP17A1 exhibits 17α-hydroxylase/17, 20-lyase activity towards 11β-hydroxyprogesterone and 11-ketoprogesterone metabolites in the C11-oxy backdoor pathway. *J Steroid Biochem Mol Biol*. 2020;199:105614.
48. Yoshimoto FK, Gonzalez E, Auchus RJ, et al. Mechanism of 17α, 20-Lyase and New Hydroxylation Reactions of Human Cytochrome P450 17A1: 18O labeling and oxygen surrogate evidence for a role of a perferryl oxygen. *J Biol Chem*. 2016;291:17143–17164.
49. Mak PJ, Gregory MC, Denisov IG, et al. Unveiling the crucial intermediates in androgen production. *Proc Natl Acad Sci U S A*. 2015;112:15856–15861.
50. Gonzalez E, Guengerich FP. Kinetic processivity of the two-step oxidations of progesterone and pregnenolone to androgens by human cytochrome P450 17A1. *J Biol Chem*. 2017;292:13168–13185.
51. Cutler GB, Glenn M, Bush M, et al. Adrenarche: a survey of rodents, domestic animals and primates. *Endocrinology*. 1978;103:2112–2118.
52. Tee MK, Miller WL. Phosphorylation of human cytochrome P450c17 by p38α selectively increases 17, 20 lyase activity and androgen biosynthesis. *J Biol Chem*. 2013;288:23903–23913.
53. Pandey AV, Mellon SH, Miller WL. Protein phosphatase 2A and phosphoprotein SET regulate androgen production by P450c17. *J Biol Chem*. 2003;278:2837–2844.
54. Suzuki T, Sasano H, Takeyama J, et al. Developmental changes in steroidogenic enzymes in human postnatal adrenal cortex: immunohistochemical studies. *Clin Endocrinol (Oxf)*. 2000;53:739–747.
55. Huang N, Miller WL. Regulation of cytochrome b_5 gene transcription by Sp3, GATA-6, and steroidogenic factor 1 in human adrenal NCI-H295A cells. *Mol Endocrinol*. 2005;19:2020–2034.
56. Gell JS, Carr BR, Sasano H, et al. Adrenarche results from development of a 3β-hydroxysteroid dehydrogenase-deficient adrenal reticularis. *J Clin Endocrinol Metab*. 1998;83:3695–3701.
57. Biglieri EG, Herron MA, Brust N. 17α-hydroxylation deficiency in man. *J Clin Invest*. 1966;15:1945–1954.
58. Costa-Santos M, Kater CE, Auchus RJ. Two prevalent CYP17 mutations and genotype-phenotype correlations in 24 Brazilian patients with 17-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2004;89:49–60.
59. Geller DH, Auchus RJ, Mendonça BB, et al. The genetic and functional basis of isolated 17, 20 lyase deficiency. *Nature Genet*. 1997;17:201–205.
60. Van Den Akker EL, Koper JW, Boehmer AL, et al. Differential inhibition of 17α-hydroxylase and 17, 20-lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. *J Clin Endocrinol Metab*. 2002;87:5714–5721.
61. Geller DH, Auchus RJ, Miller WL. P450c17 mutations R347H and R358Q selectively disrupt 17, 20-lyase activity by disrupting interactions with P450 oxidoreductase and cytochrome b_5. *Mol Endocrinol*. 1999;13:167–175.
62. Peng HM, Im SC, Pearl NM, et al. Cytochrome b5 activates the 17,20-lyase activity of human cytochrome P450 17A1 by increasing the coupling of NADPH consumption to androgen production. *Biochemistry*. 2016;55:4356–4365.
63. Sherbet DP, Tiosano D, Kwist KM, et al. CYP17 mutation E305G causes isolated 17,20-lyase deficiency by selectively altering substrate binding. *J Biol Chem*. 2003;278:48563–48569.
64. Fernandez-Cancio M, Camats N, Fluck CE, et al. mechanism of the dual activities of human CYP17A1 and binding to anti-prostate cancer drug abiraterone revealed by a novel V366M mutation causing 17,20 lyase deficiency. *Pharmaceuticals (Basel)*. 2018;11:37.
65. Hershkovitz E, Parvari R, Wudy SA, et al. Homozygous mutation G539R in the gene for P450 oxidoreductase in a family previously diagnosed as having 17,20-lyase deficiency. *J Clin Endocrinol Metab*. 2008;93:3584–3588.
66. Idkowiak J, Randell T, Dhir V, et al. A missense mutation in the human cytochrome b_5 gene causes 46,XY disorder of sex development due to true isolated 17,20 lyase deficiency. *J Clin Endocrinol Metab*. 2012;97:E465–E475.
67. Kok RC, Timmerman MA, Wolffenbuttel KP, et al. Isolated 17,20-lyase deficiency due to the cytochrome b_5 mutation W27X. *J Clin Endocrinol Metab*. 2010;95:994–999.
68. Attard G, Reid AHM, Auchus RJ, et al. Clinical and biochemical consequences of CYP17A1 inhibition with abiraterone given with and without exogenous glucocorticoids in castrate men with advanced prostate cancer. *J Clin Endocrinol Metab*. 2012;97:507–516.
69. DeVore NM, Scott EE. Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. *Nature*. 2012;482:116–119.
70. Gomes LG, Huang N, Agrawal V, et al. Extraadrenal 21-hydroxylation by CYP2C19 and CYP3A4: effect on 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2009;94:89–95.
71. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*. 2000;21:245–291.
72. Burch GH, Gong Y, Liu W, et al. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet*. 1997;17:104–108.
73. Merke DP, Chen W, Morissette R, et al. Tenascin-X haploinsufficiency associated with Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2013;98:E379–E387.
74. Zhao B, Lei L, Kagawa N, et al. A three-dimensional structure of steroid 21-hydroxylase (Cytochrome P450 21A2) with two substrates reveals locations of disease-associated variants. *J Biol Chem*. 2012;287:10613–10622.
75. Pallan PS, Nagy LD, Lei L, et al. Structural and kinetic basis of steroid 17α,20-lyase activity in teleost fish cytochrome P450 17A1 and its absence in cytochrome P450 17A2. *J Biol Chem*. 2015;290:3248–3268.
76. Mizrachi D, Wang Z, Sharma KK, et al. Why human cytochrome P450c21 is a progesterone 21-hydroxylase. *Biochemistry*. 2011;50:3968–3974.
77. Pallan PS, Wang C, Lei L, et al. Human Cytochrome P450 21A2, the Major Steroid 21-Hydroxylase: structure of the enzyme progesterone substrate complex and rate-limiting C-H bond cleavage. *J Biol Chem*. 2015;290:13128–13143.
78. Yoshimoto FK, Zhou Y, Peng HM, et al. Minor activities and transition state properties of the human steroid hydroxylases cytochromes P450c17 and P450c21, from reactions observed with deuterium-labeled substrates. *Biochemistry*. 2012;51:7064–7077.
79. Wang C, Pallan PS, Zhang W, et al. Functional analysis of human cytochrome P450 21A2 variants involved in congenital adrenal hyperplasia. *J Biol Chem*. 2017;292:10767–10778.
80. Yanagibashi K, Haniu M, Shively JE, et al. The synthesis of aldosterone by the adrenal cortex: two zones (fasciculata and glomerulosa) possess one enzyme for 11-, 18-hydroxylation, and aldehyde synthesis. *J Biol Chem*. 1986;261:3556–3562.
81. Mornet E, Dupont J, Vitek A, et al. Characterization of two genes encoding human steroid 11β-hydroxylase (P45011β). *J Biol Chem*. 1989;264:20961–20967.
82. Carnow KM, Tusie-Lung M, Pascoe L, et al. The product of the CYP11B2 gene is required for aldosterone biosynthesis in the human adrenal cortex. *Mol Endocrinol*. 1991;5:1513–1522.
83. Mulatero P, Curnow KM, Aupetit-Faisant B, et al. Recombinant CYP11B genes encode enzymes that can catalyze conversion of 11-deoxycortisol to cortisol, 18-hydroxycortisol, and 18-oxocortisol. *J Clin Endocrinol Metab*. 1998;83:3996–4001.
84. White PC, Dupont J, New MI, et al. A mutation in CYP11B1 (Arg 448→His) associated with steroid 11β-hydroxylase deficiency in Jews of Moroccan origin. *J Clin Invest*. 1991;87:1664–1667.

85. White PC, Curnow KM, Pascoe L. Disorders of steroid 11β-hydroxylase isozymes. *Endocr Rev*. 1994;15:421–438.
86. Pascoe L, Curnow K, Slutsker L, et al. Mutations in the human CYP11B2 (aldosterone synthase) gene causing corticosterone methlyoxidase II deficiency. *Proc Natl Acad Sci USA*. 1992;89:4996–5000.
87. Reddish MJ, Guengerich FP. Human cytochrome P450 11B2 produces aldosterone by a processive mechanism due to the lactol form of the intermediate 18-hydroxycorticosterone. *J Biol Chem*. 2019;294:12975–12991.
88. Clyne CD, Zhang Y, Slutsker L, et al. Angiotensin II and potassium regulate human CYP11B2 transcription through common cis-elements. *Mol Endocrinol*. 1997;11:638–649.
89. Lifton R, Dluhy RG, Powers M, et al. A chimaeric 11β-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*. 1992;335:262–265.
90. Litchfield WR, New MI, Coolidge C, et al. Evaluation of the dexamethasone suppression test for the diagnosis of glucocorticoid-remediable aldosteronism. *J Clin Endocrinol Metab*. 1997;82:3570–3573.
91. Bottner B, Denner K, Bernhardt R. Conferring aldosterone synthesis to human CYP11B1 by replacing key amino acid residues with CYP11B2-specific ones. *Eur J Biochem*. 1998;252:458–466.
92. Curnow KM, Mulatero P, Emeric-Blanchouin N, et al. The amino acid substitutions Ser288Gly and Val320Ala convert the cortisol producing enzyme, CYP11B1, into an aldosterone producing enzyme [letter]. *Nat Struct Biol*. 1997;4:32–35.
93. Strushkevich N, Gilep AA, Shen L, et al. Structural insights into aldosterone synthase substrate specificity and targeted inhibition. *Mol Endocrinol*. 2013;27:315–324.
94. Brixius-Anderko S, Scott EE. Structure of human cortisol-producing cytochrome P450 11B1 bound to the breast cancer drug fadrozole provides insights for drug design. *J Biol Chem*. 2019;294:453–460.
95. Pivonello R, Fleseriu M, Newell-Price J, et al. Efficacy and safety of osilodrostat in patients with Cushing's disease (LINC 3): a multicentre phase III study with a doubleblind, randomised withdrawal phase. *Lancet Diabet Endocrinol*. 2020;8:748–761.
96. Bogman K, Schwab D, Delporte ML, et al. Preclinical and early clinical profile of a highly selective and potent oral inhibitor of aldosterone synthase (CYP11B2). *Hypertension*. 2017;69:189–196.
97. Rege J, Nakamura Y, Satoh F, et al. Liquid chromatography-tandem mass spectrometry analysis of human adrenal vein 19-carbon steroids before and after ACTH stimulation. *J Clin Endocrinol Metab*. 2013;98:1182–1188.
98. Rege J, Turcu A, Kasa-Vubu JZ, et al. 11-ketotestosterone is the dominant circulating bioactive androgen during normal and premature adrenarche. *J Clin Endocrinol Metab*. 2018;103:4589–4598.
99. Turcu AF, Rege J, Auchus RJ, et al. 11-Oxygenated androgens in health and disease. *Nat Rev Endocrinol*. 2020;16:284–296.
100. Sohl CD, Guengerich FP. Kinetic analysis of the three-step steroid aromatase reaction of human cytochrome P450 19A1. *J Biol Chem*. 2010;285:17734–17743.
101. Yoshimoto FK, Guengerich FP. Mechanism of the third oxidative step in the conversion of androgens to estrogens by cytochrome P450 19A1 steroid aromatase. *J Am Chem Soc*. 2014;136:15016–15025.
102. Ghosh D, Griswold J, Erman M, et al. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature*. 2009;457:219–223.
103. Conte FA, Grumbach MM, Ito Y, et al. A syndrome of female pseudohermaphrodism, hypergonadotropic hypogonadism, and multicystic ovaries associated with missense mutations in the gene encoding aromatase (P450arom). *J Clin Endocrinol Metab*. 1994;78:1287–1292.
104. Lo JC, Schwitzgebel VM, Tyrrell JB, et al. Normal female infants born of mothers with classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 1999;84:930–936.
105. Morishima A, Grumbach MM, Simpson ER, et al. Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens. *J Clin Endocrinol Metab*. 1995;80:3689–3698.
106. Hanukoglu I, Suh BS, Himmelhoch S, et al. Induction and mitochondrial localization of cytochrome P450scc system enzymes in normal and transformed ovarian granulosa cells. *J Cell Biol*. 1990;111:1373–1381.
107. Voutilainen R, Picado-Leonard J, DiBlasio AM, et al. Hormonal and developmental regulation of human adrenodoxin mRNA in steroidogenic tissues. *J Clin Endocrinol Metab*. 1988;66:383–388.
108. Müller A, Müller JJ, Müller YA, et al. New aspects of electron transfer revealed by the crystal structure of a truncated bovine adrenodoxin, Adx(4-108). *Structure*. 1998;6:269–280.
109. Coghlan VM, Vickery LE. Site-specific mutations in human ferredoxin that affect binding to ferredoxin reductase and cytochrome P450scc. *J Biol Chem*. 1991;266:18606–18612.
110. Peng HM, Auchus RJ. Molecular recognition in mitochondrial cytochromes P450 that catalyze the terminal steps of corticosteroid biosynthesis. *Biochemistry*. 2017;56:2282–2293.
111. Wada A, Waterman MR. Identification by site-directed mutagenesis of two lysine residues in cholesterol side chain cleavage cytochrome P450 that are essential for adrenodoxin binding. *J Biol Chem*. 1992;267:22877–22882.
112. Brentano ST, Black SM, Lin D, et al. cAMP post-transcriptionally diminishes the abundance of adrenodoxin reductase mRNA. *Proc Natl Acad Sci USA*. 1992;89:4099–4103.
113. Brandt ME, Vickery LE. Charge pair interactions stabilizing ferredoxin-ferredoxin reductase complexes. Identification by complementary site-specific mutations. *J Biol Chem*. 1993;268:17126–17130.
114. Ziegler GA, Vonrhein C, Hanukoglu I, et al. The structure of adrenodoxin reductase of mitochondrial P450 systems: electron transfer for steroid biosynthesis. *J Mol Biol*. 1999;289:981–990.
115. Pandey AV, Fluck CE. NADPH P450 oxidoreductase: structure, function, and pathology of diseases. *Pharmacol Ther*. 2013;138:229–254.
116. Wang M, Roberts DL, Paschke R, et al. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci USA*. 1997;94:8411–8416.
117. Ellis J, Gutierrez A, Barsukov IL, et al. Domain motion in cytochrome P450 reductase: conformational equilibria revealed by NMR and small-angle x-ray scattering. *J Biol Chem*. 2009;284:36628–36637.
118. Sevrioukova I, Huiying L, Zhong H, et al. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci (USA)*. 1999;96:1863–1868.
119. Peterson RE, Imperato-McGinley J, Gautier T, et al. Male pseudohermaphroditism due to multiple defects in steroid-biosynthetic microsomal mixed-function oxidases. A new variant of congenital adrenal hyperplasia. *N Engl J Med*. 1985;313:1182–1191.
120. Flück CE, Tajima T, Pandey AV, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat Genet*. 2004;36:228–230.
121. Huang N, Pandey AV, Agrawal V, et al. Diversity and function of mutations in P450 oxidoreductase in patients with Antley-Bixler syndrome and disordered steroidogenesis. *Am J Hum Genet*. 2005;76:729–749.
122. Krone N, Reisch N, Idkowiak J, et al. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2012;97:E257–E267.
123. Henderson CJ, Otto DM, Carrie D, et al. Inactivation of the hepatic cytochrome P450 system by conditional deletion of hepatic cytochrome P450 reductase. *J Biol Chem*. 2003;278:13480–13486.
124. Tomalik-Scharte D, Maiter D, Kirchheiner J, et al. Impaired hepatic drug and steroid metabolism in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *Eur J Endocrinol*. 2010;163:919–924.
125. Huang N, Agrawal V, Giacomini KM, et al. Genetics of P450 oxidoreductase: sequence variation in 842 individuals of four ethnicities and activities of 15 missense mutations. *Proc Natl Acad Sci U S A*. 2008;105:1733–1738.
126. Bridges A, Gruenke L, Chang YT, et al. Identification of the binding site on cytochrome P450 2B4 for cytochrome b_5 and cytochrome P450 reductase. *J Biol Chem*. 1998;273:17036–17049.
127. Falzone CJ, Mayer MR, Whiteman EL, et al. Design challenges for hemoproteins: the solution structure of apocytochrome b_5. *Biochemistry*. 1996;35:6519–6526.
128. Muskett FW, Kelly GP, Whitford D. The solution structure of bovine ferricytochrome b_5 determined using heteronuclear NMR methods. *J Mol Biol*. 1996;258:172–189.

129. Lee-Robichaud P, Kaderbhai MA, Kaderbhai N, et al. Interaction of human CYP17 (P-450$_{17}$α, 17α-hydroxylase-17,20-lyase) with cytochrome b_5: importance of the orientation of the hydrophobic domain of cytochrome b_5. *Biochem J*. 1997;321:857–863.
130. Peng HM, Liu J, Forsberg SE, et al. Catalytically relevant electrostatic interactions of cytochrome P450c17 (CYP17A1) and cytochrome b_5. *J Biol Chem*. 2014;289:33838–33849.
131. Naffin-Olivos JL, Auchus RJ. Human cytochrome b_5 requires residues E48 and E49 to stimulate the 17, 20-lyase activity of cytochrome P450c17. *Biochemistry*. 2006;45:755–762.
132. Thomas JL, Frieden C, Nash WE, et al. An NADH-induced conformational change that mediates the sequential 3β-hydroxysteroid dehydrogenase/isomerase activities is supported by affinity labeling and the time-dependent activation of isomerase. *J Biol Chem*. 1995;270:21003–21008.
133. Thomas JL, Myers RP, Strickler RC. Human placental 3β-hydroxy-5-ene-steroid dehydrogenase and steroid 5/4-ene-isomerase: purification from mitochondria and kinetic profiles, biophysical characterization of the purified mitochondrial and microsomal enzymes. *J Steroid Biochem*. 1989;33:209–217.
134. Luu-The V, Lechance Y, Labrie C, et al. Full length cDNA structure and deduced amino acid sequence of human 3β-hydroxy-5-ene steroid dehydrogenase. *Mol Endocrinol*. 1989;3:1310–1312.
135. Rhéaume E, Lachance Y, Zhao HL, et al. Structure and expression of a new complementary DNA encoding the almost exclusive 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerase in human adrenals and gonads. *Mol Endocrinol*. 1991;5:1147–1157.
136. Chang KH, Li R, Kuri B, et al. A gain-of-function mutation in DHT synthesis in castration-resistant prostate cancer. *Cell*. 2013;154:1074–1084.
137. Thomas JL, Berko EA, Faustino A, et al. Human placental 3β-hydroxy-5-ene-steroid dehydrogenase and steroid 5/4-ene-isomerase: purification from microsomes, substrate kinetics, and inhibition by product steroids. *J Steroid Biochem*. 1988;31:785–793.
138. Li Z, Bishop AC, Alyamani M, et al. Conversion of abiraterone to D4A drives anti-tumour activity in prostate cancer. *Nature*. 2015;523:347–351.
139. Mesiano S, Coulter CL, Jaffe RB. Localization of cytochrome P450 cholesterol side-chain cleavage, cytochrome P450 17α-hydroxylase/17,20 lyase, and 3β-hydroxysteroid dehydrogenase-isomerase steroidogenic enzymes in human and rhesus monkey fetal adrenal glands: reappraisal of functional zonation. *J Clin Endocrinol Metab*. 1993;77:1184–1189.
140. Remer T, Boye KR, Hartmann MF, et al. Urinary markers of adrenarche: reference values in healthy subjects, aged 3-18 years. *J Clin Endocrinol Metab*. 2005;90:2015–2021.
141. Peltoketo H, Luu-The V, Simard J, et al. 17β-hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. *J Mol Endocrinol*. 1999;23:1–11.
142. Peltoketo H, Isomaa V, Mäenlavsta O, et al. Complete amino acid sequence of human placental 17β-hydroxysteroid dehydrogenase deduced from cDNA. *FEBS Lett*. 1988;239:73–77.
143. Luu-The V, Zhang Y, Poirier D, et al. Characteristics of human types 1, 2 and 3 17β-hydroxysteroid dehydrogenase activities: oxidation/reduction and inhibition. *J Steroid Biochem Mol Biol*. 1995;55:581–587.
144. Ghosh D, Pletnev VZ, Zhu DW, et al. Structure of human estrogenic 17β-hydroxysteroid dehydrogenase at 2.20 Å resolution. *Structure*. 1995;3:503–513.
145. Hakkarainen J, Jokela H, Pakarinen P, et al. Hydroxysteroid (17β)-dehydrogenase 1-deficient female mice present with normal puberty onset but are severely subfertile due to a defect in luteinization and progesterone production. *FASEB J*. 2015;29:3806–3816.
146. Sasano H, Frost AR, Saitoh R, et al. Aromatase and 17β-hydroxysteroid dehydrogenase type 1 in human breast carcinoma. *J Clin Endocrinol Metab*. 1996;81:4042–4046.
147. Wu L, Einstein M, Geissler WM, et al. Expression cloning and characterization of human 17β-hydrosteroid dehydrogenase Type 2, a microsomal enzyme possessing 20α-hydroxysteroid dehydrogenase activity. *J Biol Chem*. 1993;268:12964–12969.
148. Casey ML, MacDonald PC, Andersson S. 17β-Hydroxysteroid dehydrogenase type 2: chromosomal assignment and progestin regulation of gene expression in human endometrium. *J Clin Invest*. 1994;94:2135–2141.
149. Rantakari P, Strauss L, Kiviranta R, et al. Placenta defects and embryonic lethality resulting from disruption of mouse hydroxysteroid (17-β) dehydrogenase 2 gene. *Mol Endocrinol*. 2008;22:665–675.
150. Saez JM, De Peretti E, Morera AM, et al. Familial male pseudohermaphroditism with gynecomastia due to a testicular 17-ketosteroid reductase defect. I. Studies in vivo. *J Clin Endocrinol Metab*. 1971;32:604–610.
151. Geissler WM, Davis DL, Wu L, et al. Male pseudohermaphroiditism caused by mutations of testicular 17β-hydroxysteroid dehydrogenase 3. *Nat Genet*. 1994;7:34–39.
152. Moghrabi N, Hughes IA, Dunaif A, et al. Deleterious missense mutations and silent polymorphism in the human 17β-hydroxysteroid dehydrogenase 3 gene (HSD17B3). *J Clin Endocrinol Metab*. 1998;83:2855–2860.
153. Mendonca BB, Arnhold IJ, Bloise W, et al. 17β-hydroxysteroid dehydrogenase 3 deficiency in women. *J Clin Endocrinol Metab*. 1999;84:802–804.
154. Qin YM, Poutanen MH, Helander HM, et al. Peroxisomal multifunctional enzyme of beta-oxidation metabolizing D-3- hydroxyacyl-CoA esters in rat liver: molecular cloning, expression and characterization. *Biochem J*. 1997;321:21–28.
155. Lin HK, Jez JM, Schlegel BP, et al. Expression and characterization of recombinant type 2 3α-hydroxysteroid dehydrogenase (HSD) from human prostate: demonstration of bifunctional 3α/17β-HSD activity and cellular distribution. *Mol Endocrinol*. 1997;11:1971–1984.
156. Penning TM, Burczynski ME, Jez JM, et al. Human 3α-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem J*. 2000;351:67–77.
157. Nakamura Y, Hornsby PJ, Casson P, et al. Type 5 17β-hydroxysteroid dehydrogenase (AKR1C3) contributes to testosterone production in the adrenal reticularis. *J Clin Endocrinol Metab*. 2009;94:2192–2198.
158. Bruchovsky N, Wilson JD. The conversion of testosterone to 5α-androstan-17β-ol-3-one by rat prostate *in vivo* and *in vitro*. *J Biol Chem*. 1968;243:2012–2021.
159. Moore RJ, Wilson JD. Steroid 5α-reductase in cultured human fibroblasts. Biochemical and genetic evidence for two distinct enzyme activities. *J Biol Chem*. 1976;251:5895–5900.
160. Andersson S, Berman DM, Jenkins EP, et al. Deletion of a steroid 5α-reductase 2 gene in male pseudohermaphroditism. *Nature*. 1991;354:159–161.
161. Thigpen AE, Davis DL, Milatovich A, et al. Molecular genetics of steroid 5α-reductase 2 deficiency. *J Clin Invest*. 1992;90:799–811.
162. Frederiksen DW, Wilson JD. Partial characterization of the nuclear reduced nicotinamide adenine dinucleotide phosphate: Δ^4-3-ketosteroid 5α-oxidoreductase of rat prostate. *J Biol Chem*. 1971;246:2584–2593.
163. Chang KH, Li R, Papari-Zareei M, et al. Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *Proc Natl Acad Sci U S A*. 2011;108:13728–13733.
164. Mahendroo MS, Wilson JD, Richardson JA, et al. Steroid 5α-reductase 1 promotes 5α-androstane-3α,17β-diol synthesis in immature mouse testes by two pathways. *Mol Cell Endocrinol*. 2004;222:113–120.
165. Khanna M, Qin KN, Wang RW, et al. Substrate specificity, gene structure and tissue-specific distribution of multiple human 3α-hydroxysteroid dehydrogenases. *J Biol Chem*. 1995;270:20162–20168.
166. Rupprecht R, Hauser CA, Trapp T, et al. Neurosteroids: molecular mechanisms of action and psychopharmacological significance. *J Steroid Biochem Mol Biol*. 1996;56:163–168.
167. Meltzer-Brody S, Colquhoun H, Riesenberg R, et al. Brexanolone injection in post-partum depression: two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. *Lancet*. 2018;392:1058–1070.
168. Griffin LD, Mellon SH. Selective serotonin reuptake inhibitors directly alter activity of neurosteroidogenic enzymes. *Proc Natl Acad Sci USA*. 1999;96:13512–13517.
169. Gulbis JM, Mann S, MacKinnon R. Structure of a voltage-dependent K$^+$ channel beta subunit. *Cell*. 1999;97:943–952.

170. Hoog SS, Pawlowski JE, Alzari PM, et al. Three-dimensional structure of rat liver 3α-hydroxysteroid/dihydrodiol dehydrogenase: a member of the aldo-keto reductase superfamily. *Proc Natl Acad Sci USA*. 1994;91:2517–2521.
171. Cooper WC, Jin Y, Penning TM. Elucidation of a complete kinetic mechanism for a mammalian hydroxysteroid dehydrogenase (HSD) and identification of all enzyme forms on the reaction coordinate: the example of rat liver 3α-HSD (AKR1C9). *J Biol Chem*. 2007;282:33484–33493.
172. Ratnam K, Ma H, Penning TM. The arginine 276 anchor for NADP(H) dictates fluorescence kinetic transients in 3α-hydroxysteroid dehydrogenase, a representative aldo-keto reductase. *Biochemistry*. 1999;38:7856–7864.
173. Napoli JL. 17β-Hydroxysteroid dehydrogenase type 9 and other short-chain dehydrogenases/reductases that catalyze retinoid, 17β- and 3α-hydroxysteroid metabolism. *Mol Cell Endocrinol*. 2001;171:103–109.
174. Biswas MG, Russell DW. Expression cloning and characterization of oxidative 17β- and 3α-hydroxysteroid dehydrogenases from rat and human prostate. *J Biol Chem*. 1997;272:15959–15966.
175. Chetyrkin SV, Hu J, Gough WH, et al. Further characterization of human microsomal 3α-hydroxysteroid dehydrogenase. *Arch Biochem Biophys*. 2001;386:1–10.
176. White PC, Mune T, Agarwal AK. 11β-hydroxysteroid dehydrogenase and the syndrome of apparent mineralocorticoid excess. *Endocr Rev*. 1997;18:135–156.
177. Tannin GM, Agarwal AK, Monder C, et al. The human gene for 11β-hydroxysteroid dehydrogenase. Structure, tissue distribution, and chromosomal localization. *J Biol Chem*. 1991;266:16653–16658.
178. Walker EA, Clark AM, Hewison M, et al. Functional expression, characterization, and purification of the catalytic domain of human 11β-hydroxysteroid dehydrogenase type 1. *J Biol Chem*. 2001;276:21343–21350.
179. Odermatt A, Arnold P, Stauffer A, et al. The N-terminal anchor sequences of 11β-hydroxysteroid dehydrogenases determine their orientation in the endoplasmic reticulum membrane. *J Biol Chem*. 1999;274:28762–28770.
180. Jamieson A, Wallace AM, Andrew R, et al. Apparent cortisone reductase deficiency: a functional defect in 11β-hydroxysteroid dehydrogenase type 1. *J Clin Endocrinol Metab*. 1999;84:3570–3574.
181. Lavery GG, Walker EA, Tiganescu A, et al. Steroid biomarkers and genetic studies reveal inactivating mutations in hexose-6-phosphate dehydrogenase in patients with cortisone reductase deficiency. *J Clin Endocrinol Metab*. 2008;93:3827–3832.
182. Brown RW, Chapman KE, Edwards CRW, et al. Human placental 11β-hydroxysteroid dehydrogenase: evidence for and partial purification of a distinct NAD$^+$-dependent isoform. *Endocrinology*. 1993;132:2614–2621.
183. Funder JW, Pearce PT, Smith R, et al. Mineralocorticoid action: target tissue specificity is enzyme, not receptor, mediated. *Science*. 1988;242:583–585.
184. Mune T, Rogerson FM, Nikkila H, et al. Human hypertension caused by mutations in the kidney isozyme of 11β-hydroxysteroid dehydrogenase. *Nat Genet*. 1995;10:394–399.
185. Seely J, Amigh KS, Suzuki T, et al. Transcriptional regulation of dehydroepiandrosterone sulfotransferase (SULT2A1) by estrogen-related receptor alpha. *Endocrinology*. 2005;146:3605–3613.
186. Noordam C, Dhir V, McNelis JC, et al. Inactivating PAPSS2 mutations in a patient with premature pubarche. *N Engl J Med*. 2009;360:2310–2318.
187. Mulatero P, Curnow KM, Aupetit-Faisant B, et al. Recombinant CYP11B genes encode enzymes that can catalyze conversion of 11-deoxycortisol to cortisol, 18-hydroxycortisol, and 18-oxocortisol. *J Clin Endocrinol Metab*. 1998;83:3996–4001.
188. Homma K, Hasegawa T, Nagai T, et al. Urine steroid hormone profile analysis in cytochrome P450 oxidoreductase deficiency: implication for the backdoor pathway to dihydrotestosterone. *J Clin Endocrinol Metab*. 2006;91:2643–2649.
189. Davio A, Woolcock H, Nanba AT, et al. Sex differences in 11-oxygenated androgen patterns across adulthood. *J Clin Endocrinol Metab*. 2020;105:e2921–e2929.
190. Gent R, du Toit T, Bloem LM, et al. The 11β-hydroxysteroid dehydrogenase isoforms: pivotal catalytic activities yield potent C11-oxy C19 steroids with 11βHSD2 favouring 11-ketotestosterone, 11-ketoandrostenedione and 11-ketoprogesterone biosynthesis. *J Steroid Biochem Mol Biol*. 2019;189:116–126.
191. Barnard M, Quanson JL, Mostaghel E, et al. 11-Oxygenated androgen precursors are the preferred substrates for aldo-keto reductase 1C3 (AKR1C3): implications for castration resistant prostate cancer. *J Steroid Biochem Mol Biol*. 2018;183:192–201.
192. Wise-Oringer BK, Burghard AC, O'Day P, et al. The unique role of 11-oxygenated C19 steroids in both premature adrenarche and premature pubarche. *Horm Res Paediatr*. 2020;93:460–469.
193. Turcu AF, El-Maouche D, Zhao L, et al. Androgen excess and diagnostic steroid biomarkers for nonclassic 21-hydroxylase deficiency without cosyntropin stimulation. *Eur J Endocrinol*. 2020;183:63–71.
194. Turcu AF, Nanba AT, Chomic R, et al. Adrenal-derived 11-oxygenated 19-carbon steroids are the dominant androgens in classic 21-hydroxylase deficiency. *Eur J Endocrinol*. 2016;174:601–609.
195. Tezuka Y, Atsumi N, Blinder AR, et al. The age-dependent changes of the human adrenal cortical zones are not congruent. *J Clin Endocrinol Metab*. 2021;106:1389–1397.
196. Storbeck KH, Bloem LM, Africander D, et al. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: a putative role in castration resistant prostate cancer? *Mol Cell Endocrinol*. 2013;377:135–146.
197. Montgomery RB, Mostaghel EA, Vessella R, et al. Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. *Cancer Res*. 2008;68:4447–4454.
198. Evaul K, Li R, Papari-Zareei M, et al. 3β-hydroxysteroid dehydrogenase is a possible pharmacological target in the treatment of castration-resistant prostate cancer. *Endocrinology*. 2010;151:3514–3520.
199. Bauman DR, Steckelbroeck S, Williams MV, et al. Identification of the major oxidative 3α-hydroxysteroid dehydrogenase in human prostate that converts 5α-androstane-3α,17β-diol to 5α-dihydrotestosterone: a potential therapeutic target for androgen-dependent disease. *Mol Endocrinol*. 2006;20:444–458.

88

Mineralocorticoids: Physiology, Metabolism, Receptors, and Resistance

Morag J. Young and Peter J. Fuller

OUTLINE

Feedback Control of Aldosterone Secretion, 1474
Sodium Homeostasis, 1475
Potassium Homeostasis, 1476
Specificity-Conferring Enzymes, 1476
  11β-Hydroxysteroid Dehydrogenase Type 2, 1476
Mineralocorticoid Receptors, 1476
  Evolution, 1476
  Structure, 1477
  Genomic Versus Nongenomic Aldosterone Actions, 1478
Cellular Mechanisms of Aldosterone Action, 1478
  Sodium Transport, 1478
  Epithelial Sodium Channels, 1478
  Na+/K+-ATPase, 1479
  Potassium Transport, 1479
  Hydrogen Ion Transport, 1480
Nonepithelial Tissues, 1480
  Cardiovascular System, 1480
  Central Nervous System, 1481
Summary and Future Directions, 1481

The steroid hormone aldosterone first appeared in vertebrate evolution with the emergence of terrestrial life and the consequent need to conserve sodium and water.[1] The primary and best-characterized actions of aldosterone are those that stimulate sodium retention in transporting epithelia, particularly the distal nephron, distal colon, and salivary glands.[1] In these epithelia, the conservation of sodium is associated with an increased secretion of both potassium and hydrogen ions. Aldosterone also has actions in the cardiovascular system, in inflammatory cells, in adipose tissue, and in the central nervous system (CNS).

The existence of an adrenocorticoid natriuretic factor, distinct from the other adrenal steroid hormones, had been suspected for several years before its isolation by Simpson and colleagues in 1953.[2] Using the toad urinary bladder as a model system, Crabbé[3] was the first to show that aldosterone increases sodium transport *in vitro*. Subsequent studies demonstrated the presence of binding sites for aldosterone in the toad bladder and in other target tissues, particularly the principal cells of the cortical collecting duct of the kidney.[4]

FEEDBACK CONTROL OF ALDOSTERONE SECRETION

Both serum sodium concentration and total body sodium are maintained within a narrow range by a complex set of endocrine feedback loops (Fig. 88.1). The most important of these involves the renin-angiotensin-aldosterone system (RAAS), which responds to volume status. The feedback loops involved with potassium homeostasis operate in parallel and overlap those for sodium.

The renin-secreting juxtaglomerular cells of the kidney sense volume status. When the sodium status (and hence, intravascular volume) is low, renin will be secreted. Renin is an aspartyl protease that is synthesized as inactive prorenin and subsequently activated by the action of a protease. Renin release from the juxtaglomerular cells is influenced by several factors (Table 88.1). Stimulatory regulators include renal perfusion pressure, sympathetic nervous system, and prostaglandins, while inhibitory factors include dopamine, atrial natriuretic peptide (ANP) and angiotensin II (AngII). Renin cleaves angiotensinogen to release the decapeptide angiotensin I, which in turn is subject to further proteolysis by angiotensin-converting enzyme 1, primarily in the pulmonary vascular bed, to yield the octapeptide AngII. Ang II acts via its specific G protein–coupled receptor, the type 1 AngII receptor (AT1R), in the vasculature as a potent vasoconstrictor (thereby defending plasma volume and blood pressure) and on the adrenal glomerulosa cells to stimulate aldosterone synthesis.[5] The latter response promotes sodium retention, with a consequent increase in intravascular volume. Aldosterone biosynthesis in the zona glomerulosa of the adrenal cortex is regulated by transcription of the aldosterone synthase gene (*CYP11B2*), which encodes the last and rate-limiting step in aldosterone biosynthesis (see Chapter 87). Although AngII is important in the regulation of aldosterone synthase activity, a response to low-salt or high-potassium diets is also seen in mice in which the angiotensinogen gene has been deleted.[6] In these mice, the regulation of aldosterone is primarily directed by serum potassium levels. Both AT1R and potassium interact with potassium channels in the glomerulosa cell membrane. These channels maintain the hyperpolarization of the cell membrane; depolarization of the membrane through changes in ion flux through these channels results in an increased calcium flux, with consequent stimulation of aldosterone secretion. Experimental and clinical studies have identified channels involved in this coupling, including TWIK-related acid sensitive channels, GIRK4 (G protein–activated inward rectifier potassium channel 4), an inward-rectifying potassium channel encoded by the *KCNJ5* gene, the α1 subunit of Na^+/K^+-ATPase encoded by the ATP1A1 gene, PMCA3 (plasma membrane Ca^{2+}-ATPase type 3) encoded by the ATP2B3 gene, Cav1.3 (a voltage-dependent calcium channel) encoded by the *CACNAID* gene, and

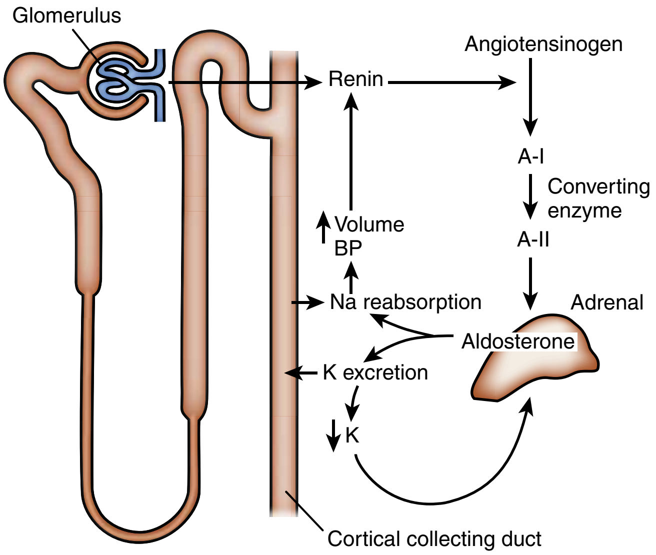


Fig. 88.1 Interacting feedback loops controlling aldosterone secretion. Volume is regulated through the renin-angiotensin system, and potassium is regulated through direct feedback. *A-I*, Angiotensin I; *A-II*, angiotensin II; *BP*, blood pressure; *Na*, sodium; *K*, potassium.

TABLE 88.1 Factors Regulating Renin Release

Stimulatory	Inhibitory
Decreased perfusion pressure	Increased chloride delivery at the macula densa
PGI_2	Angiotensin II
ACTH	Atrial natriuretic factor
	Vasopressin
β-adrenergic stimulation	α-adrenergic stimulation
	Dopamine

ACTH, Adrenocorticotropic hormone; *PGI_2*, prostacyclin.

ClC-2 (chloride channel protein 2) encoded by the CLCN2 gene.[7] Somatic mutations in the *KCNJ5*, *ATP1A1*, *ATP2B3*, or *CACNAID* genes are found in approximately 88% of aldosterone-producing adenomas,[7] as discussed in Chapter 96.

The secretion of aldosterone stimulated by potassium forms one arm of a negative feedback loop for potassium and aldosterone that this independent of AngII levels. Small increases in extracellular potassium levels stimulate calcium influx via depolarization of the glomerulosa cell plasma membrane. Plasma potassium levels also alter the sensitivity of the adrenal to AngII.

Aldosterone also affects acid–base balance by increasing the exchange of hydrogen ions for sodium. Therefore, the net effect of an increase in aldosterone, as may result from an aldosterone-producing tumor or exogenous mineralocorticoid administration (e.g., 9α-fludrocortisone), is sodium resorption, with consequent volume expansion, hypertension, and suppression of plasma renin activity, hypokalemia, and metabolic alkalosis (Chapter 96).

The third principal regulator of aldosterone synthesis is adrenocorticotrophic hormone (ACTH), whose primary role is the regulation of cortisol synthesis in the adrenal fasciculata in the feedback loop between the pituitary corticotropes and the adrenal cortex, the hypothalamo-pituitary-adrenal axis (Chapter 85). ACTH acutely increases aldosterone synthesis, but chronic ACTH exposure results in a decrease in the levels of CYP11B2, and hence aldosterone synthesis[8]; however, regulation of aldosterone synthesis is normal in patients with hypopituitarism.[9]

TABLE 88.2 Factors Regulating Aldosterone Secretion

Factor	Stimulatory	Inhibitory
Peptides	Angiotensin II	Atrial natriuretic peptide
	Angiotensin III	Somatostatin
	ACTH	
	Vasopressin	
	Endothelin	
Ions	Plasma potassium	
Other	Serotonin	Dopamine
	Substance P	Ouabain

ACTH, Adrenocorticotropic hormone.

Other substances that have been shown to regulate aldosterone secretion are shown in Table 88.2. For example, intraadrenal regulation of aldosterone secretion has been proposed for substance P–positive nerve fibers. In the adrenal zona glomerulosa, a stimulatory tone for aldosterone secretion maintained by substance P acting via the neurokinin type 1 receptor on zona glomerulosa cells.[10] Serotonin is exclusively synthesized and produced by adrenal subcapsular mast cells; it stimulates aldosterone secretion through a paracrine mechanism involving the serotonin type 4 receptor.[11] The physiological and pathophysiological significance of these local regulators of aldosterone secretion is the subject of ongoing investigation. Aldosterone secretion is also subject to negative regulation. ANP is a potent inhibitor of aldosterone secretion, consistent with its role in promoting natriuresis. Dopamine is also a well-characterized inhibitor of aldosterone secretion. Other inhibitors have been described, but their physiologic relevance is even less clear.

KEY POINTS

- Aldosterone synthesis/secretion is primarily regulated by angiotensin II and potassium in two parallel feedback loops.

SODIUM HOMEOSTASIS

Under normal physiological conditions, over 99% of the sodium filtered through the glomerulus is reabsorbed along the nephron: approximately 70% by the proximal tubules, 20% to 30% by the thick ascending limb of Henle, 5% to7% by the distal convoluted tubules (DCTs), and finally by the aldosterone-sensitive connecting tubules (and cortical collecting ducts.[12] As the final arbiter of sodium excretion, these aldosterone-sensitive segments are thus pivotal. The RAAS contributes to an integrated homeostatic response to fluctuations in dietary sodium intake.[13] Sodium deficiency increases adrenal sensitivity to AngII through time; the converse is true of the vasopressor response. Aldosterone-induced sodium retention restores intravascular volume status by maintaining the balance between volume and vascular capacity.[12,13]

The response of the individual to aldosterone-mediated sodium retention is self-limiting, in that, after 3 to 4 days, the expansion of extracellular volume plateaus and sodium secretion returns to control levels. This process is termed *escape*.[14] It should be noted that the kaliuretic effect persists despite the escape of sodium retention. Intrarenal regulators, particularly prostaglandins, are probably the critical mediators of the escape, although other factors (e.g., ANP) may play a role.[15]

Other epithelia, particularly the distal colon, contribute to aldosterone-mediated sodium homeostasis through mechanisms that parallel those in the distal nephron. In addition, nuclei within the brain contribute to the regulation of salt appetite.[16]

A local RAAS has been reported to operate in a number of tissues, although the relative physiologic versus pathophysiological importance of these local systems remains unresolved.[5]

POTASSIUM HOMEOSTASIS

Aldosterone is primarily involved in the chronic regulation of plasma potassium levels.[17] Acute regulation involves nonrenal mechanisms such as those mediated by insulin and β-adrenergic agonists. Aldosterone regulates potassium homeostasis through direct effects on transporting epithelia. Potassium flux occurs in the principal cells of the collecting duct in response to aldosterone-mediated sodium reabsorption, with the resulting electrochemical gradient favoring potassium excretion at the basolateral membrane.[17a] Aldosterone also regulates potassium transport in the distal nephron independent of the sodium flux. The collecting duct contains intercalated cells, which have well-recognized roles in potassium and acid–base homeostasis.[18]

SPECIFICITY-CONFERRING ENZYMES

Prereceptor metabolism (activation or inactivation) is seen for several steroid hormones, including the corticosteroids. The physiological glucocorticoid, cortisol (corticosterone in rodents), is synthesized in the adrenal fasciculata, and also in the periphery, particularly the liver, by the conversion of cortisone (11-dehydrocorticosterone in rodents) to the active glucocorticoid by the enzyme 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1), a member of the short-chain dehydrogenase/reductase family of microsomal enzymes. The reverse reaction, mediated by the 11β-HSD type 2 isoform (11β-HSD2), mediates the conversion of cortisol to the inactive metabolite, cortisone, thereby enabling the mineralocorticoid actions of aldosterone.[19]

11β-Hydroxysteroid Dehydrogenase Type 2

Although the affinity of the mineralocorticoid receptor *in vitro*[20] is equivalent for aldosterone, corticosterone, and cortisol, *in vivo* cortisol is excluded from such receptors in the kidney, parotid, and colon (but not in the hippocampus).[19] This reflects the activity of 11β-HSD2, which is responsible for the conversion of cortisol to its inactive metabolite, cortisone. In the kidney, this is reflected in the reduced cortisol:cortisone ratio in human renal venous blood. Unlike 11β-HSD1, 11β-HSD2 colocalizes with MR in renal distal tubular elements, colon, sweat, and salivary glands; it is also expressed at high abundance in the placenta and in select nuclei in the rat brain.[16] It has a low Km (i.e., high affinity) for both corticosterone (~5 nM) and cortisol (~50 nM), unlike the micromolar Km of 11β-HSD1; *in vivo*, it appears operationally unidirectional, acting uniquely as a dehydrogenase, whereas 11β-HSD1 acts predominantly as a reductase.[19] Deletion of the gene encoding 11β-HSD2 in mice yields a phenotype[19] consistent with the clinical syndrome of apparent mineralocorticoid excess.[21,22]

Liquorice ingestion has long been known to cause sodium retention, hypokalemia, and hypertension; the mechanism of its action was elucidated by elegant studies[23,24] in human volunteers, in whom ingestion of 250 g of liquorice per day for 10 days produced a clinical picture equivalent to a mild form of apparent mineralocorticoid excess. When rats were administered glycyrrhetinic acid (the active principal of liquorice) or carbenoxolone (glycyrrhetinic acid hemisuccinate)[25] to block 11β-HSD, the normal aldosterone selectivity of epithelial MR was abolished. These studies demonstrated the crucial specificity-conferring role in mineralocorticoid target tissues for 11β-HSD2. Aldosterone is not a substrate for 11β-HSD2, because its 11β-OH group is stabilized as an 11,18-hemiketal by cyclization with the highly reactive aldehyde group unique to aldosterone at C18.

KEY POINTS

- Aldosterone selectivity of the MR in epithelial tissues is conferred by 11β-hydroxysteroid dehydrogenase type II.

MINERALOCORTICOID RECEPTORS

The classic actions of aldosterone in the distal nephron and distal colon involve epithelial cells that mediate sodium flux. As with other steroid hormones, the principal mode of action, at least in sodium transport, involves an intracellular receptor that, when activated by ligand binding, regulates gene transcription, a so-called genomic mechanism of action.

High-affinity cytosol and nuclear binding of ^{3}H-aldosterone was first described in classic mineralocorticoid target tissues such as kidney and parotid gland.[4] Spironolactone was shown to block aldosterone binding and action on urinary electrolytes in parallel,[26] providing evidence that these sites are physiologic MRs. In contrast to glucocorticoid receptors (GRs), which are expressed ubiquitously, MRs have a tissue-specific pattern of expression, with the highest levels observed in the distal nephron,[20,27] distal colon,[28] and hippocampus.[20] Lower levels of expression are observed elsewhere in the gastrointestinal tract, in cardiovascular tissues, in regions of the brain, and in a range of other tissues that are both epithelial and nonepithelial.[20,27,28] In these nonepithelial tissues, a physiologic effect of MR activation on sodium homeostasis is unlikely (e.g., hippocampus).

The MR is unique among the steroid hormone receptors in that it is a receptor for two physiologic agonists, aldosterone and cortisol; indeed, the MR has a higher affinity for cortisol than does the GR.[20] Progesterone is a physiological competitive antagonist of the human MR.[29]

KEY POINTS

- Aldosterone activates the mineralocorticoid receptor (MR), a member of the nuclear receptor superfamily.

Evolution

The MR is found only in vertebrates; it diverged with the GR as a gene duplication from the ancestral corticoid receptor gene over 450 million years ago.[30] This ancestral corticoid receptor first appeared in cartilaginous fish and is found in lamprey and hagfish (jawless fish). Aldosterone first appeared as an active steroid in amphibians. It has been postulated that the MR originated in fish to regulate ion balance under the control of glucocorticoids, and that amphibians evolved mineralocorticoids to appropriate the MR for this function. In most tetrapods (including human, rodent, alligator, and *Xenopus*) progesterone is an antagonist of MR, as is its pharmacologic derivative spironolactone.[29] Although the MR is present in teleost fish that lack aldosterone synthesis, the teleost MR responds to aldosterone, as well as to cortisol and 11-deoxycorticosterone, consistent with the MR being highly conserved throughout vertebrate evolution.[29,31] Characterization of various fish MRs found that, in contrast to terrestrial vertebrate, teleost MRs were activated by both progesterone and spironolactone.[29] These findings suggest that progesterone may be a physiological agonist for the MR in teleosts. Curiously, this evolutionary switch of the MR response to progesterone from agonist to antagonist reflects a single amino acid change in the MR[31] that occurred at the time of the appearance of aldosterone synthesis in terrestrial vertebrates. The physiological significance of the MR as a progesterone receptor (PR), or perhaps as a receptor for a progesterone metabolite, has not been fully characterized.[32]

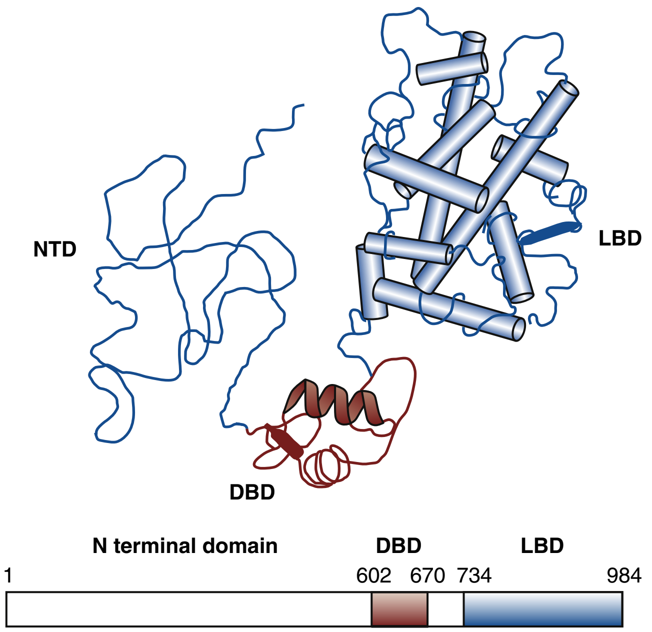


Fig. 88.2 Domain structure of the mineralocorticoid receptor showing the three principal functional domains, including the N-terminal domain (*NTD*), the DNA-binding domain (*DBD*), and the ligand-binding domain (*LBD*). A predicted tertiary structure is shown above the linear representation.

KEY POINTS

- The MR is unique, being the receptor for two physiologic ligands: aldosterone and cortisol.

Structure

The human MR is a protein of 984 amino acids,[20] which, with the GR, PR, and androgen receptor (AR), forms a distinct subfamily within the steroid/thyroid/retinoid/orphan receptor superfamily. A central cysteine-rich DNA-binding domain (DBD) defines this receptor superfamily. At the C-terminus is the ligand-binding domain (LBD), which has a highly conserved tertiary structure. The N-terminal domain has little or no homology between receptors. Within the MR/GR/PR/AR subfamily, MR and GR are closely related, with 94% amino-acid identity in the central cysteine-rich DBD and 57% identity in the C-terminal LBD (Fig. 88.2). The genes encoding MR and GR are, however, located on different chromosomes (MR on 4q31.2; GR on 5q31).[33,34]

The cysteine residues of the DBD complex around two zinc atoms to form two α-helices, one of which lies in the major groove and binds with a common consensus sequence in the DNA, the hormone response element (HRE).[35] This element is an inverted palindrome that appears to be similar, if not identical, to the well-characterized GR response element (GRE). Several groups have sought to characterize the genome-wide distribution of the HRE for the MR, the MR cistrome, using chromatin immunoprecipitation with massive parallel sequencing, with differing outcomes. A study using a transformed human renal cell line identified 974 MR binding sites, the majority of which, somewhat surprisingly, did not include canonical GRE but did contain recognition motifs for other transcription factors,[36] whereas in a murine epithelial cell line 1113 binding sites were identified, with many containing the consensus GRE sequence.[37]A total of 918 unique MR binding sites were identified in rat hippocampus that included the well-characterized canonical GRE sequence; the MR-specific binding sites included a motif known to bind the NeuroD family of transcription factors.[38]

The LBD consists of 11 α-helices (designated helix 1 to 12 based on the estrogen receptor structure: helix 2 is unstructured in the MR) that form a three-layered antiparallel structure with the ligand-binding pocket buried in the lower third of the struture.[39-41] The N-terminal domain, by contrast, is relatively unstructured, allowing a diversity of interactions.[42] The N-terminal domains are not conserved between steroid receptors, although in several (including the MR) a functional interaction has been described between the N- and C-termini.[43]

The unliganded receptor is predominantly cytoplasmic, being complexed with the heat shock proteins 70 and 90 and their cochaperones.[44] This configuration maintains the receptor in a transcriptionally inactive high-affinity binding state. The interaction of the LBD with this complex is an important determinant of ligand-binding affinity and specificity. The MR antagonists spironolactone and eplerenone appear to be accommodated into the ligand-binding pocket without distortion, suggesting that the mechanism of their antagonism differs from that of the estrogen receptor antagonists such as tamoxifen and raloxifene.[40] In nonepithelial tissues, the response of MR to cortisol/corticosterone and aldosterone is often not equivalent, leading to speculation about the relationship of epithelial to nonepithelial MR. Most evidence to date would suggest that, although transcription of the gene encoding MR is activated through multiple, tissue-specific promoters,[45] the coding region is unaltered between tissues, with the possible exception of some minor isoforms.[46] The explanation for such differences in MR target gene activation by the various ligands may lie in the nature of the conformation that the MR adopts after ligand binding.[47] Such conformational differences may alter some, but not all, transactivation functions, such that tissue-specific receptor coactivators or corepressors may mediate different responses in different tissues. In contrast to the other steroid receptors, such interactions are only now being characterized for the MR. Several MR coactivator molecules have now been described, including some with ligand specificity.[48] The interaction of these coregulators with the LBD, as with other nuclear receptors, involves a surface pocket that interacts with an LxxLL motif (where L is leucine and x is any amino acid) in the coactivator molecule.[41]

Other factors also appear to modulate MR signaling, including Rac1 (Ras-related C3 botulinum toxin substrate1), a small RhoGTPase, which may play a role in regulating nuclear translocation of transcription factors and has been reported to activate MR through both ligand-dependent and ligand-independent mechanisms.[49]

In addition to its effects on ion flux in classic mineralocorticoid target tissues, aldosterone has been shown act via the MR to mediate a range of responses, unrelated to sodium homeostasis, in other tissues. Aldosterone elevates blood pressure in the rat when infused into the cerebral ventricles[50]; this effect is clearly via MR unprotected by11β-HSD2, because it is blocked by simultaneous infusion of low doses of corticosterone. Thus, corticosterone in the anteroventral third ventricle region acts as an aldosterone antagonist on MR, in contrast with the kidney and other epithelia, where its action is to mimic aldosterone. An additional difference between epithelial and nonepithelial tissues is that, in the former, activation of GR has been shown to mimic that of MR, whereas in nonepithelial tissues this is clearly not the case.

Mice homozygous for inactivating mutations in the gene encoding MR[51] (MR knockout; MRKO) showed classic features of aldosterone deficiency—salt wasting, hyperkalemia, and dehydration—but had marked hyperaldosteronism; these features are also seen in the syndrome of pseudohypoaldosteronism (PHA) type 1. In patients with PHA1, neonatal salt wasting results in hypovolemia with high renin and aldosterone levels, hyperkalemia, and resistance to exogenous mineralocorticoids (9α-fludrocortisone). MRKO mice are born at the expected frequency from heterozygote mating; untreated, they begin to deteriorate from day 5, and die between day 8 and 11; salt supplementation allows survival and growth. Mutations of the gene encoding MR have also been reported in the autosomal dominant form of PHA,[52] which thus appears to be equivalent to mice heterozygous for the MR gene knockout.

Genomic Versus Nongenomic Aldosterone Actions

As with other steroid hormone receptors, the principal mode of action involves the ligand-activated MR binding to specific responses elements in the genes targeted for transcriptional control. Mice homozygous for an inactivating mutation in the DBD of the MR phenocopy MRKO mice, demonstrating the essential role of DNA binding for MR signaling, at least in sodium-transporting epithelial tissues.[52]

For several other steroid hormone receptors there is an alternate mode of action, again involving regulation of gene expression, but where the steroid receptor interacts with other transcription factors, including other nuclear receptors, to modulate their signaling, a mechanism described as transrepression or tethering.[53] The relevance of this mode of signaling to the MR remains to be clearly defined.

There has also been considerable interest with respect to steroid hormone action as to whether all responses are mediated through the classic nuclear receptor with direct regulation of gene expression, or whether other pathways involving novel cell membrane receptors exist. The evidence for novel receptors is focused on the seven-transmembrane G protein–coupled receptor GRP30, also known as the G protein estrogen receptor.[54] There is clear evidence both *in vitro* and *in vivo* for rapid, nongenomic signaling. This can involve activation of intracellular kinases such as Src kinase associated with the epidermal growth factor receptor, with consequent downstream signaling through the MAP kinase pathway; the signaling appears to require only the LBD of the MR.[55] Rapid effects of aldosterone have also been reported to be associated both with protein kinase signaling[56] in isolated cell patches from cardiomyocytes[57] and in vascular cells[58] and with glutamate release from the pyramidal neurons of the hippocampus.[59] In each case the receptor involved is the classic MR. The relative contribution of this signaling to the mineralocorticoid response has not yet been evaluated, although it is speculated that these rapid responses may prime the transcriptional response or alter the dynamic range of the response.

CELLULAR MECHANISMS OF ALDOSTERONE ACTION

As the term *mineralocorticoid* implies, the classic effects of aldosterone reflect its role in the regulation of electrolyte flux across transporting epithelia. The molecular basis of the regulation of sodium transport in the distal nephrons is relatively well-characterized. The molecular mechanisms that mediate the more recently defined roles of aldosterone in the central regulation of blood pressure,[50] in salt appetite,[16] and in the pathogenesis of cardiac fibrosis[60] are less well understood. Emerging evidence points strongly to a proinflammatory, profibrotic effect of MR activation in cardiovascular tissues consistent with the clinical observation that hypertension secondary to hyperaldosteronism is associated with worse cardiovascular outcomes than idiopathic hypertension.[60]

Sodium Transport

Aldosterone increases transepithelial sodium flux in a number of target tissues, of which the amphibian bladder, the mammalian distal nephron, and the mammalian distal colon are the best characterized.[1] The temporal pattern of response comprises a lag period (latent phase) of 30 to 60 minutes, followed by an early phase in which preexisting pumps and channels are activated, followed by a so-called late phase.[61] This late phase starts 3 to 6 hours after steroid exposure and is characterized by an increase in the number of pumps and channels; with longer exposure, morphologic as well as functional changes are observed.[62] The latent phase is consistent with the concept that the effects of aldosterone are primarily genomic, with the genes regulated by the MR either mediating sodium transport *per se* or modulating components of the transport pathway. In the classic model of aldosterone-induced transport, sodium entry at the apical membrane is through an amiloride-sensitive electrogenic sodium channel, with efflux at the basolateral membrane via the sodium pump, with adenosine triphosphate (ATP) generation needed to drive the process (Fig. 88.3).[1]

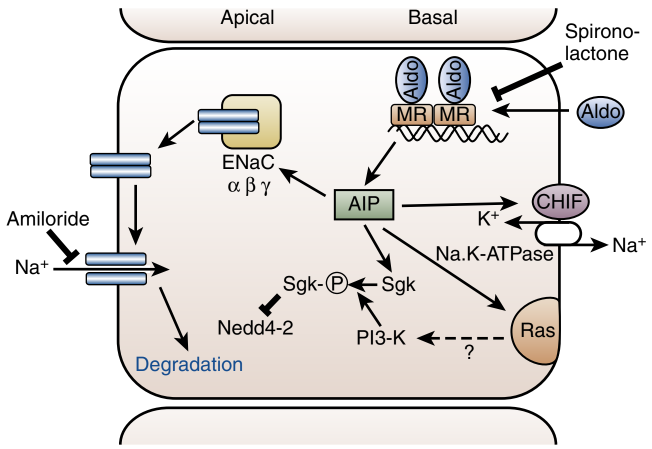


Fig. 88.3 Aldosterone (*Aldo*) acts on epithelial cells of the distal nephron and the distal colon via the mineralocorticoid receptor to induce aldosterone-induced proteins (*AIPs*). *ENaC*, Epithelial sodium channels; *sgk*, serum and glucocorticoid-induced kinase; *CHIF*, channel-inducing factor; *PI3-K*, phosphoinositide 3-kinase; *Na+*, sodium; *K+*, potassium.

KEY POINTS

- Aldosterone-induced epithelial sodium transport involves the epithelial sodium channel at the apical membrane and Na+/K+-ATPase at the basolateral membrane.

Epithelial Sodium Channels

Although a variety of amiloride-sensitive sodium channels have been identified in epithelia, only one has been clearly shown to be directly involved in aldosterone-dependent sodium transport. The epithelial sodium channel (ENaC) genes encode three homologous subunits (α, β, and γ).[63] Each subunit consists of two transmembrane domains with intracellular N- and C-termini. When the α subunit is expressed in *Xenopus* oocytes, a weak amiloride-sensitive sodium flux can be demonstrated; expression of the other subunits alone is without activity, and coexpression of all three subunits is required for maximal amiloride-sensitive sodium transport.[63] The subunits exhibit a heterotrimeric complex of the three subunits, α, β, and γ, arranged in a counter-clockwise manner.[64]

The genes for the ENaC subunits are members of the degenerin (DEG)/ENaC superfamily of sodium channels; the DEGs mediate mechanosensory transduction in *Caenorhabditis elegans*.[63] The ENaCs are unique among the DEG/ENaC channels in being activated by proteolysis of the extracellular domain to release inhibitory peptides. These serine proteases act synergistically with aldosterone to increase channel opening probability.[63]

Insights into the structural determinants of function of the ENaC subunits has come from a combination of both naturally occurring and engineered mutations. Deletion of the *αENaC* gene in transgenic mice results in early neonatal death caused by defective clearance of lung liquid[66]; partial genetic restoration of *αENaC* expression in these mice results in a phenotype similar to that seen in PHA. Mice with the *βENaC* or *γENaC* gene deleted show a less severe pulmonary phenotype but die between 24 and 36 hours with salt wasting and profound hyperkalaemia, again reminiscent of PHA[67,68]; low levels of *βENaC* subunit gene expression in mice in which the *βENaC* gene has been disrupted also result in a milder phenotype analogous to PHA.[69] Autosomal recessive PHA is indeed associated with inactivating mutations of genes encoding all three of the ENaC subunits.[70]

The molecular characterization of Liddle syndrome[71] also provided important insights into the function of the C-terminal intracellular domain and emphasized the central role of the channel in amiloride-sensitive sodium transport. Liddle syndrome, or pseudoaldosteronism, has the reciprocal pathophysiology to PHA,[71] with apparent aldosterone excess (i.e., hypertension, hypokalemia, and suppressed plasma renin activity) but low aldosterone levels, and no response to spironolactone but a response to amiloride. A number of studies have identified mutations in either the *β*- or *γENaC* genes. The mutations are either nonsense or missense mutations localized to the C-terminus; the latter are particularly informative, because they identify a key motif, proline-proline-proline-X-tyrosine (PY), which is conserved across the subunits and is disrupted in all cases of Liddle syndrome.[72] ENaC is a relatively short-lived protein that is ubiquitinated on residues in the N-terminus of the α and γ but not the β subunits; the PY motif interacts with Nedd4-2, a ubiquitin protein-ligase whose role is to target the channels for proteosomal degradation.[73]

These studies demonstrated the central role of the ENaC in mediating aldosterone-induced epithelial sodium transport. ENaC subunit gene expression is regulated by aldosterone.[1] Both β- and γENaC subunit mRNA levels are increased in the colon by sodium depletion and by dexamethasone or aldosterone treatment; in the kidney, however, levels remained unaltered, although when separate regions of the kidney were distinguished, an increase in αENaC mRNA levels was seen in the inner medulla. Cultured medullary collecting duct cells show increased α- but not β- or γENaC mRNA levels after 3 hours of aldosterone exposure; in cultured cortical collecting duct cells, a γENaC subunit response to aldosterone required 24 hours of treatment.[74] Although it is clear from the aforementioned studies that aldosterone can induce ENaC synthesis (at least in the late phase), a transcriptional effect of the MR in the early phase (i.e., a primary response) has not clearly been demonstrated.

Aldosterone treatment *in vivo* rapidly increases the levels of the serine threonine kinase serum and glucocorticoid-regulated kinase (Sgk) in the kidney and colon, with a time course consistent with an effect of aldosterone on transcription.[75,76] Sgk directly interacts with Nedd4-2 to block its binding with ENaC and, as a consequence, slows ENaC degradation.[77,78] Regulation of αENaC subunit gene expression involves Sgk-1 through relief of Dotla-Af9-mediated transcriptional repression.[79] Nedd4-2 is also regulated by Usp2-45, a deubiquitinylation enzyme that is itself regulated by aldosterone.[80] Sgk requires phosphorylation by the phosphatidylinosital 3-kinase (PI3 kinase) pathway for full activity. This may be a point at which the signaling of nuclear receptors and membrane-associated receptors such as the insulin receptor are integrated.[81] PI3 kinase may be activated by small monomeric G proteins, including Ras. *K-ras 2A* has been identified as an aldosterone-induced gene in the rodent distal colon.[82] The glucocorticoid-induced leucine zipper protein is also aldosterone-induced; it interacts with and inhibits Raf-1, leading to repression of ERK signaling, a negative regulator of ENaC, as well as directly interacting with Sgk and Nedd 4-2.[83] A further aldosterone-induced protein, connector enhancer of kinase repressor of Ras3, appears to serve as a scaffold protein in the assembly of the ENaC regulatory complex.[74]

Na^+/K^+-ATPase

Active extrusion of sodium from the cell reflects sodium pump activity in the basolateral cell membrane. Na^+/K^+-ATPase activity is increased in epithelia by aldosterone. In toad bladder and in A6 cells, there is evidence that Na^+/K^+-ATPase α and β subunit gene expression is significantly increased in the early phase of the response to aldosterone,[61] although this does not appear to be the case in mammalian systems.[84] Na^+/K^+-ATPase activity is sensitive to intracellular sodium concentration, and in isolated cortical tubules the early Na^+/K^+-ATPase response to aldosterone is blocked by amiloride, suggesting that the increased activity is secondary to sodium influx at the apical membrane.[85] In the late phase of the aldosterone response, Na^+/K^+-ATPase mRNA, protein, and activity levels are all increased.[86] Channel-inducing factor (CHIF) was first identified as a novel corticosteroid-induced gene in the rat distal colon. CHIF is a member of the FXYD family of small transmembrane proteins, which includes the γ-subunit of Na^+/K^+-ATPase.[87] Under experimental conditions, CHIF is upregulated in the distal colon in response to aldosterone[88,89] through a primary transcriptional response. CHIF increases the affinity of Na^+/K^+-ATPase for sodium.[87] The early aldosterone-induced increase in Na^+/K^+-ATPase activity is mediated by CHIF.

Potassium Transport

Aldosterone also regulates potassium transport, independent of its effects on the electroneutral apical sodium transport coupled to the Na^+/K^+-ATPase-mediated basolateral membrane exchange of sodium for potassium. Aldosterone also increases the apical membrane channel density of renal outer medullary potassium channel (ROMK), an ATP-regulated potassium channel with weak inward rectification in the principal cells of the cortical collecting duct.[90] Sgk-1 has been reported to increase the channel density of ROMK through direct phosphorylation, through Nedd4-2,[91] or perhaps via the with-no-lysine (WNK) kinases.[92] Insights into the roles of ROMK and the channels regulated by the WNK kinases have come from studies of Bartter, Gitelman, and Gordon syndromes; however, whether the responses in the DCT are mediated via aldosterone directly or via its effects elsewhere in the nephron remains controversial. Other inwardly rectifying potassium channels expressed on the basolateral membrane act as sensors of extracellular potassium to play a central role in potassium and sodium homeostasis.[90] This aldosterone-independent pathway enables extracellular potassium to act through the type 2 mTOR complex to activate Sgk-1 and stimulate ENaC to enhance ENaC-mediated sodium transport and, indirectly, potassium excretion.[93]

The intercalated cells within the collecting duct consist of subtypes α, β, and non-α/non-β.[94] Intercalated cells express H^+-ATPase on the apical or basolateral plasma membrane, depending on whether the subtype mediates hydrogen ion or bicarbonate secretion. The β-intercalated cells express the chloride/bicarbonate ion exchanger pendrin on the apical plasma membrane and H^+-ATPase at the basolateral plasma membrane, such that increased activity will promote potassium retention. The response of the β-intercalated cell to aldosterone is modulated by phosphorylation of a specific serine (serine 843) in the LBD of the MR, which then precludes aldosterone binding.[95] Diminished plasma potassium levels promote dephosphorylation of this serine, allowing aldosterone-induced increases in both pendrin and H^+-ATPase. Dephosphorylation was also described in response to AngII; this has been suggested to provide a mechanism by which the AngII-stimulated, aldosterone-mediated response to volume depletion can result in sodium reabsorption with limited potassium depletion.

The H^+/K^+-ATPase in the apical membrane of β-intercalated cells in the collecting duct is composed of catalytic α and regulatory β subunits, with expression of the α_2 subunit being regulated by dietary potassium and corticosteroids in both the outer medullary collecting duct and the distal colon.[96]

Bartter and Gitelman syndromes are inherited conditions of hypokalemia induced by secondary aldosteronism in response to a chloride leak in the thick ascending limb and the DCT, respectively.[97] In contrast to primary aldosteronism, they are associated with hypotension rather than hypertension. The syndromes can be viewed as being pharmacologically mimicked by loop diuretics such as frusemide and thiazide diuretics, respectively. Bartter syndrome is associated with a low calcium and a normal magnesium being both uncommon and severe, whereas Giteman syndrome is associated with a normal calcium and a low magnesium, often with a milder and later onset. There are various subcategories of Bartter syndrome, but the predominant mutations are in ROMK or the sodium-potassium-chloride cotransporter, whereas mutations in the thiazide-sensitive sodium-chloride cotransporter (NCCT) (or less commonly CLCNKB—a member of CLC family of voltage-gated chloride channels) are seen in Gitleman syndrome, usually as a compound heterozygote. Gordon syndrome (pseudohypoaldosteronism type 2 or familial hyperkalemia and hypertension syndrome) can be viewed as the reciprocal phenotype of Gitleman syndrome, with low-renin hypertension, hyperkalemia, and metabolic acidosis; the condition responds to thiazide diuretics. It is a monogenic syndrome resulting from mutations in four genes, including those encoding the WNK kinases, that directly or indirectly regulate the NCCT.[98]

KEY POINTS

- Aldosterone-induced epithelial potassium and hydrogen ion transport involves sodium-dependent and -independent mechanisms.

Hydrogen Ion Transport

As for potassium, aldosterone also produces effects on proton excretion beyond a simple cation exchange for sodium across the epithelium to maintain electroneutrality. The targets for this effect are carbonic anhydrase–rich cells, particularly the α-intercalated cells within the outer medullary collecting ducts. Aldosterone-induced transport across the apical membrane is through an H+-ATPase activity, coupled with increased activity of the basolateral chloride/bicarbonate exchanger (which is upregulated by aldosterone).[99] These are largely sodium-independent effects, as demonstrated by a lack of an effect of amiloride. Aldosterone also produces an effect on the Na^+/H^+ antiporter in a range of tissues; in some tissues the response is rapid and represents a nongenomic action.[100]

NONEPITHELIAL TISSUES

The presence of the MR in a range of tissues that may not be related to electrolyte homeostasis suggests additional actions for aldosterone ion physiology and pathophysiology. In some tissues, such as the vasculature[101] and specific regions of the CNS where 11β-HSD2 is coexpressed with the MR, the MR response is likely to be attributed to activation by aldosterone. In other tissues such as cardiomyocytes and monocyte/macrophages where the MR is expressed, but 11β-HSD2 is not, rather than that of aldosterone, cortisol is the predominant ligand that activates MR, except perhaps in pathological states of aldosterone excess.[102] The MR is also expressed in a range of other tissues, including adipocytes, immune cells, and reproductive tissues, where the physiological roles remain to be determined, although in most tissues, in the absence of 11β-HSD2, they are likely to be responding to cortisol.

Cardiovascular System

In recent decades, the pathophysiology of MR in the cardiovascular systems has been defined by the use of pharmacological, genetic, and cell-selective knockout animal models to reveal cell-, ligand-, and disease-specific roles for the MR. The MR plays an important physiological role in this system to control the response to volume loss. Aldosterone-directed MR activation in the myocardium regulates the ionotropic and chronotropic activity of cardiomyocytes, resulting in increased cardiac output. MR in the vascular smooth muscle cells (VSMCs), for example, contributes to vasoconstriction, which, together with the endothelial response, has been argued to reflect aldosterone-mediated protection of vascular integrity to maintain volume homeostasis.[103] Enhanced contraction of the peripheral vasculature, via direct MR actions or via enhanced signaling of vasoactive peptides, serves to increase peripheral resistance, which, together with an increased cardiac output, will raise blood pressure and increase organ perfusion. Recent studies using transgenic mice null for the MR in VSMCs have also identified a role for the MR in vascular stiffness and blood pressure control in aging.[101]

Single-cell studies using patch-clamp to analyze single channels or ion exchangers in rabbit cardiomyocytes identified ligand-selective actions of the MR on the sodium-potassium-chloride cotransporter. Whereas acute aldosterone acutely induced flux, cortisol did not. However, under oxidative stress, cortisol was an equivalent agonist to aldosterone.[104] These studies support the increasing body of evidence that the MR in the heart has ligand-selective responses. These ligand-specific responses are important, given that the cardiac MR it is not protected by 11β-HSD2 and is thus continuously occupied by cortisol under physiological conditions.[105] Indeed, several studies have shown that cortisol can oppose aldosterone agonist effects. The question remains as to the physiological role of the MR in the heart in the absence of volume depletion and an adequate sodium/potassium intake. Very recent studies have identified the circadian molecular clock as a target of the MR in the heart and potentially in other tissues (i.e., kidney); this may represent a physiological action of the cortisol-activated MR.[106]

In the vasculature, where the MR is coexpressed with 11β-HSD2, aldosterone regulates a range of genes associated with inflammation and fibrosis, resulting in vascular remodeling. In vascular endothelial cells, chronic elevation of aldosterone increases transcription and cell surface expression of intracellular adhesion molecule-1 and vascular adhesion molecule-1 to promote recruitment of mononuclear cells as part of the profibrotic response.[103] These effects occur independent of the aforementioned increases in vascular reactivity.

In the heart (a nonepithelial tissue expressing MR), *in vivo* studies show that levels of aldosterone inappropriately high for the sodium status of the rat produce diffuse perivascular and interstitial fibrosis, an effect that can be antagonized by either corticosterone or spironolactone.[107] The clinical correlation of these observations was found in large trials showing a benefit of the addition of a mineralocorticoid antagonist to the conventional regimen in the treatment of severe cardiac failure, with respect to both morbidity and mortality.[108-110] Studies in transgenic mice using tissue-specific deletion of the MR are starting to clarify the specific roles and mechanisms of signaling through the MR in these tissues.[111]

KEY POINTS

- Activation of the MR in nonepithelial tissue is associated with a proinflammatory, profibrotic response that drives adverse cardiovascular outcome, as seen with chronic aldosterone excess.

Central Nervous System

Similarly, the MR is expressed extensively in the CNS; indeed, the hippocampus arguably has the highest abundance of MR of any tissue except for the distal colon.[28] The function of the MR in the CNS has been the subject of extensive investigation; it appears to mediate diverse behavioral responses, including memory and affect.[112] In the majority of brain regions, including the hippocampus, the MR is likely functioning as a receptor for cortisol, given that 11β-HSD2 is not coexpressed with the MR, and penetration of the blood-brain barrier by aldosterone is substantially less than that of most steroid hormones.[16] There is, however, coexpression of 11β-HSD2 and the MR in a small number of nuclei in the nucleus of the solitary tract (NTS); in these nuclei the role of the MR is to regulate salt appetite in response to aldosterone, and therefore sodium balance.[113] They occupy a subregion of the NTS with a diminished blood-brain barrier that may afford exposure to circulating aldosterone. These neurons express the AT1R, and MR activation appears to interact synergistically with AngII to promote salt appetite.[114]

SUMMARY AND FUTURE DIRECTIONS

Aldosterone acts through the MR to act as the final critical arbiter of sodium, potassium, and hydrogen ion homeostasis in the distal nephron and the distal colon. Aldosterone also promotes sodium retention, and hence volume homeostasis, through effects on salt appetite and on vascular integrity. The role of the MR in nonepithelial tissues remains to be fully elucidated, including the relative importance of cortisol and aldosterone as the physiological ligand in these tissues. The significance of progesterone as an MR antagonist in terrestrial vertebrates and as an MR agonist in fish suggests yet unappreciated roles for aldosterone and the MR.

REFERENCES

1. Rossier BC, Baker ME, Studer RA. Epithelial sodium transport and its control by aldosterone: the story of our internal environment revisited. *Physiol Rev*. 2015;95:297–340.
2. Simpson SA, Tait JF, Wettstein A, et al. [Isolierung eines neuen kristallisierten hormons aus nebennieren mit besonders hoher wirksamkeit auf den mineral-stoffwechsel.] *Experientia*. 1953;9:333–335.
3. Crabbé J. Stimulation of active sodium transport across toad bladder with aldosterone in vitro. *J Clin Invest*. 1961;76:2103–2110.
4. Feldman D, Funder JW, Edelman IS. Subcellular mechanisms in the action of adrenal steroids. *Am J Med*. 1972;53:545–560.
5. Paul M, Poyan Mehr A, Kreutz R. Physiology of local renin-angiotensin systems. *Physiol Rev*. 2006;86:747–803.
6. Okubo S, Niimura F, Nishimura H, et al. Angiotensin-independent mechanism for aldosterone synthesis during chronic extracellular fluid volume depletion. *J Clin Invest*. 1997;99:855–860.
7. De Sousa K, Boulkroun S, Baron S, et al. Genetic, cellular, and molecular heterogeneity in adrenals with aldosterone-producing adenoma. *Hypertension*. 2020;75:1034–1044.
8. Hattangady NG, Olala LO, Bollag WB, et al. Acute and chronic regulation of aldosterone production. *Mol Cell Endocrinol*. 2012;350:151–162.
9. Williams GH, Rose LI, Dluhy RG, et al. Aldosterone response to sodium restriction and ACTH stimulation in panhypopituitarism. *J Clin Endocrinol Metab*. 1971;32:27–35.
10. Wils J, Duparc C, Cailleux A-F, et al. The neuropeptide substance P regulates aldosterone secretion in human adrenals. *Nat Commun*. 2020;11:2673.
11. Lefebvre H, Duparc C, Naccache A, et al. Paracrine regulation of aldosterone secretion in physiological and pathophysiological conditions. *Vitam Horm*. 2019;109:303–339.
12. Nishimoto M, Fujita T. Renal mechanisms of salt-sensitive hypertension: contribution of two steroid receptor-associated pathways. *Am J Physiol Renal Physiol*. 2015;308:F377–F387.
13. Kotchen TA, Cowley Jr AW, Frohlich ED. Medical progress: salt in health and disease—A delicate balance. *New Engl J Med*. 2013;368:1229–1237.
14. Knox F, Burnett JJ, Kohan D, et al. Escape from the sodium-retaining effects of mineralocorticoids. *Kidney Int*. 1980;17:263–276.
15. Zimmerman RS, Edwards BS, Schwab TR, et al. Atrial natriuretic peptide during mineralocorticoid escape in the human. *J Clin Endocrinol Metab*. 1987;64:624–627.
16. Geerling JC, Loewy AD. Aldosterone in the brain. *Am J Physiol Renal Physiol*. 2009;297:F559–F576.
17. Young DB. Quantitative analysis of aldosterone's role in potassium regulation. *Am J Physiol*. 1988;55:F811–F822.

17a Palmer BF. Regulation of potassium homeostasis. *Clin J Am Soc Nephrol*. 2015;10:1050–1060.

18. Wall SM, Verlander JW, Romero CA. The renal physiology of pendrin-positive intercalated cells. *Physiol Rev*. 2020;100:1119–1147.
19. Odermatt A, Kratschmar DV. Tissue-specific modulation of mineralocorticoid receptor function by 11β-hydroxysteroid dehydrogenases: an overview. *Mol Cell Endocrinol*. 2012;350:168–186.
20. Arriza JL, Weinberger C, Cerelli G, et al. Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science*. 1987;237:268–275.
21. Wilson RC, Krozowski ZS, Obeyesekere VR, et al. A mutation in the HSD11B2 gene in a family with apparent mineralocorticoid excess. *J Clin Endocrinol Metab*. 1995;80:2263–2266.
22. Mune T, Rogerson FM, Nikkila H, et al. Human hypertension caused by mutations in the kidney isozyme of 11 beta-hydroxysteroid dehydrogenase. *Nature Genet*. 1995;10:394–399.
23. Stewart PM, Valentino R, Wallace AM, et al. Mineralocorticoid activity of licorice: 11β-hydroxysteroid dehydrogenase activity comes of age. *Lancet*. 1987;2:821–824.
24. Edwards CRW, Stewart PM, Burt D, et al. Localization of 11β-hydroxysteroid dehydrogenase: tissue-specific protector of the mineralocorticoid receptor. *Lancet*. 1988;2:986–989.
25. Funder JW, Pearce PT, Smith R, et al. Mineralocorticoid action: target-tissue specificity is enzyme, not receptor-mediated. *Science*. 1988;242:583–585.
26. Marver D, Stewart J, Funder JW, et al. Renal aldosterone receptors: studies with (^{3}H)aldosterone and the anti-mineralocorticoid (^{3}H)spirolactone (SC-26304). *Proc Natl Acad Sci U S A*. 1974;71:1431–1435.
27. Todd-Turla MD, Schnermann J, Fejes-Toth G, et al. Distribution of mineralocorticoid glucocorticoid receptor mRNA along the nephron. *Am J Physiol*. 1993;264:F781–F791.
28. Fuller PJ, Verity K. Mineralocorticoid receptor gene expression in the gastrointestinal tract: distribution and ontogeny. *J Steroid Biochem*. 1990;36:263–267.
29. Baker ME, Katsu Y. Progesterone: an enigmatic ligand for the mineralocorticoid receptor. *Biochem Pharmacol*. 2020;177:113976.
30. Bridgham JT, Carroll SM, Thornton JW. Evolution of hormone-receptor complexity by molecular exploitation. *Science*. 2006;312:97–101.
31. Fuller PJ, Yao YZ, Jin R, et al. Molecular evolution of the switch for progesterone and spironolactone from mineralocorticoid receptor agonist to antagonist. *Proc Natl Acad Sci U S A*. 2019;116:18578–18583.
32. Kiilerich P, Geffroy G, Valotaire C, et al. Endogenous regulation of 11-deoxycorticosterone (DOC) and corticosteroid receptors (CRs) during rainbow trout early development and the effects of corticosteroids on hatching. *Gen Comp Endocrinol*. 2018;265:22–30.
33. Fan Y-S, Eddy RL, Byers LL, et al. The human mineralocorticoid receptor gene (MLR) is located on chromosome 4 at q31.2. *Cytogenet Cell Genet*. 1989;52:83–84.
34. Theriault A, Boyd E, Harrap SB, et al. Regional chromosomal assignment of the human glucocorticoid receptor gene to 5q31. *Hum Genet*. 1989;83:289–291.
35. Hudson WH, Youn C, Ortlund EA. Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA. *PLoS One*. 2014;9:e107000.

36. Le Billan F, Khan JA, Lamribet K, et al. Cistrome of the aldosterone-activated mineralocorticoid receptor in human renal cells. *FASEB J*. 2015;29:3977–3989.
37. Ueda K, Fujiki K, Shirahige K, et al. Genome-wide analysis of murine renal distal convoluted tubular cells for the target genes of mineralocorticoid receptor. *Biochem Biophys Res Commun*. 2014;445:132–137.
38. van Weert LTCM, Buurstede JC, Mahfouz A, et al. NeuroD factors discriminate mineralocorticoid from glucocorticoid receptor DNA binding in the male rat brain. *Endocrinology*. 2017;158:1511–1522.
39. Fagart J, Huyet J, Pinon GM, et al. Crystal structure of a mutant mineralocorticoid receptor responsible for hypertension. *Nature Struct Mol Biol*. 2005;12:554–555.
40. Bledsoe RK, Madauss KP, Holt JA, et al. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *J Biol Chem*. 2005;280:31283–31293.
41. Li Y, Suino K, Daugherty J, et al. Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol Cell*. 2005;19:367–280.
42. Kumar R, McEwan IJ. Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. *Endoc Rev*. 2012;33:271–299.
43. Pippal JB, Yao Y, Rogerson FM, et al. Structural and functional characterization of the interdomain interaction in the mineralocorticoid receptor. *Mol Endocrinol*. 2009;23:1360–1370.
44. Bruner KL, Derfoul A, Robertson NM, et al. The unliganded mineralocorticoid receptor is associated with heat shock proteins 70 and 90 and the immunophilin FKBP-52. *Recept Signal Transduct*. 1997;7:85–98.
45. Zennaro M-C, Keightley MC, Kotelevtsev Y, et al. Human mineralocorticoid receptor genomic structure and identification of expressed isoforms. *J Biol Chem*. 1995;270:21016–21020.
46. Zennaro MC, Souque A, Viengchareun S, et al. A new human MR splice variant is a ligand-independent transactivator modulating corticosteroid action. *Mol Endocrinol*. 2001;15:1586–1598.
47. Fuller PJ, Yang J, Young MJ. Mechanisms of mineralocorticoid receptor signaling. *Vitam Horm*. 2019;109:37–68.
48. Fuller PJ, Yang J, Young MJ. Coregulators as mediators of mineralocorticoid receptor signalling diversity. *J Endocrinol*. 2017;234:T23–T34.
49. Nishimoto M, Fujita T. Renal mechanisms of salt-sensitive hypertension: contribution of two steroid receptor-associated pathways. *Am J Physiol Renal Physiol*. 2015;308:F377–F387.
50. Gomez-Sanchez EP, Venkataraman MT, Thwaites D, et al. Intracerebroventricular infusion of corticosterone antagonizes ICV-aldosterone hypertension. *Am J Physiol*. 1990;258:E649–E653.
51. Berger S, Bleich M, Schmid W, et al. Mineralocorticoid receptor knockout mice: pathophysiology of Na+ metabolism. *Proc Natl Acad Sci U S A*. 1998;95:9424–9429.
52. Zennaro MC, Fernandes-Rosa F. Mineralocorticoid receptor mutations. *J Endocrinol*. 2017;234:T93–T106.
52. Cole TJ, Terella L, Morgan J, et al. Aldosterone-mediated renal sodium transport requires intact mineralocorticoid receptor DNA-binding in the mouse. *Endocrinology*. 2015;156:2958–2968.
53. De Bosscher K, Desmet SJ, Clarisse D, et al. Nuclear receptor crosstalk - defining the mechanisms for therapeutic innovation. *Nat Rev Endocrinol*. 2020;16:363–377.
54. Feldman RD, Limbird LE. GPER (GPR30): a nongenomic receptor (GPCR) for steroid hormones with implications for cardiovascular disease and cancer. *Annu Rev Pharmacol Toxicol*. 2017;57:567–584.
55. Grossmann C, Freudinger R, Mildenbergr S, et al. EF-domains are sufficient for nongenomic mineralocorticoid receptor actions. *J Biol Chem*. 2008;283:7109–7116.
56. McEneaney V, Harvey BJ, Thomas W. Aldosterone regulates rapid trafficking of ENaC subunits in renal cortical collecting duct cells via protein kinase D activation. *Mol Endocrinol*. 2008;22:881–882.
57. Mihailidou AS, Mardini M, Funder JW. Rapid, nongenomic effects of aldosterone in the heart mediated by epsilon protein kinase C. *Endocrinology*. 2003;145:773–780.
58. Alzamora R, Marusic ET, Gonzalez M, et al. Nongenomic effect of aldosterone on Na+,K+-adenosine triphosphatase in arterial vessels. *Endocrinology*. 2003;144:1266–1272.
59. Karst H, Berger S, Turiault M, et al. Mineralocorticoid receptors are indispensable for nongenomic modulation of hippocampal glutamate transmission by corticosterone. *Proc Natl Acad Sci USA*. 2005;102:19204–19207.
60. Young MJ, Adler GK. Aldosterone, the mineralocorticoid receptor and mechanisms of cardiovascular disease. *Vitam Horm*. 2019;109:361–385.
61. Verrey F. Early aldosterone action: toward filling the gap between transcription and transport. *Am J Physiol*. 1999;277:F319–F327.
62. Wade JB, Stanton BA, Field MJ, et al. Morphological and physiological responses to aldosterone: time course and sodium dependence. *Am J Physiol*. 1990;259:F88–F94.
63. Canessa CM, Schild L, Buell G, et al. Amiloride-sensitive epithelial Na+ channel is made of three homologous subunits. *Nature*. 1994;367:463–467.
64. Noreng S, Bharadwaj A, Posert R, et al. Structure of the human epithelial sodium channel by cryo-electron microscopy. *Elife*. 2018;7:Ge39340.
65. Vuagniaux G, Vallet V, Jaeger NF, et al. Synergistic activation of ENaC by three membrane-bound channel-activating serine proteases (mCAP1, mCAP2, and mCAP3) and serum- and glucocorticoid-regulated kinase (Sgk1) in Xenopus oocytes. *J Gen Physiol*. 2002;120:191–201.
66. Hummler E, Barker P, Gatzy J, et al. Early death due to defective neonatal lung liquid clearance in αENaC-deficient mice. *Nat Genet*. 1996;12:325–328.
67. Barker PM, Nguyen MS, Gatzy JT, et al. Role of gamma ENaC subunit in lung liquid clearance and electrolyte balance in newborn mice. Insights into perinatal adaptation and pseudohypoaldosteronism. *J Clin Invest*. 1998;102:1634–1640.
68. McDonald FJ, Yang B, Hrstka RF, et al. Disruption of the beta subunit of the epithelial Na+ channel in mice: hyperkalemia and neonatal death associated with a pseudohypoaldosteronism phenotype. *Proc Natl Acad Sci U S A*. 1999;96:1727–1731.
69. Pradervand S, Barker PM, Wang Q, et al. Salt restriction induces pseudohypoaldosteronism type 1 in mice expressing low levels of the beta-subunit of the amiloride-sensitive epithelial sodium channel. *Proc Natl Acad Sci U S A*. 1999;96:1732–1737.
70. Chang SS, Grunder S, Hanukoglu A, et al. Mutations in subunits of the epithelial sodium channel cause salt wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1. *Nat Genet*. 1996;12:248–253.
71. Shimkets RA, Warnock DG, Bositis CM, et al. Liddle's syndrome: heritable human hypertension caused by mutations in the β subunit of the epithelial sodium channel. *Cell*. 1994;79:407–414.
72. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. *Cell*. 2001;104:545–556.
73. Staub O, Gautschi I, Ishikawa T, et al. Regulation of stability and function of the epithelial Na+ channel ENaC by ubiquitination. *EMBO J*. 1997;16:6325–6336.
74. Soundararajan R, Pearce D, Ziera T. The role of the ENaC-regulatory complex in aldosterone-mediated sodium transport. *Mol Cell Endocrinol*. 2012;350:242–247.
75. Brennan FE, Fuller PJ. Rapid up-regulation of serum and glucocorticoid-regulated kinase (sgk) gene expression by corticosteroids in vivo. *Mol Cell Endocrinol*. 2000;30:129–136.
76. Bhargava A, Fullerton MJ, Myles K, et al. The serum- and glucocorticoid-induced kinase is a physiological mediator of aldosterone action. *Endocrinology*. 2001;142:1587–1594.
77. Debonneville C, Flores SY, Kamynina E, et al. Phosphorylation of Nedd4-2 by Sgk1 regulates epithelial Na+ channel cell surface expression. *EMBO J*. 2001;20:7052–7059.
78. Snyder PM, Olson DR, Thomas BC. Serum and glucocorticoid-regulated kinase modulates Nedd4-2-mediated inhibition of the epithelial Na+ channel. *J Biol Chem*. 2002;277:5–8.
79. Zhang W, Xia X, Reisenauer MR, et al. Aldosterone-induced Sgk1 relieves Dot1a-Af9-mediated transcriptional repression of epithelial Na+ channel α. *J Clin Invest*. 2007;117:773–783.
80. Fakitsas P, Adam G, Daidié D, et al. Early aldosterone-induced gene product regulates the epithelial sodium channel by deubiquitylation. *J Am Soc Nephrol*. 2007;18:1084–1092.

81. Wang J, Barbry P, Maiyar AC, et al. SGK integrates insulin and mineralocorticoid regulation of epithelial sodium transport. *Am J Physiol*. 2001;280:F303–F313.
82. Brennan FE, Fuller PJ. Mammalian K-ras2 is a corticosteroid-induced gene in vivo. *Endocrinology*. 2006;147:2809–2816.
83. Rashmi P, Colussi G, Ng M, et al. Glucocorticoid-induced leucine zipper protein regulates sodium and potassium balance in the distal nephron. *Kidney Int*. 2017;91:1159–1177.
84. Fuller PJ, Verity K. Colonic sodium-potassium adenosine triphosphate subunit gene expression: ontogeny and regulation by adrenocortical steroids. *Endocrinology*. 1990;127:32–38.
85. Petty KJ, Kokko JP, Marver D. Secondary effect of aldosterone on Na-KATPase activity in the rabbit cortical collecting tubule. *J Clin Invest*. 1981;68:1514–1521.
86. Sansom SC, O'Neil RG. Mineralocorticoid regulation of apical cell membrane Na+ and K+ transport of the cortical collecting duct. *Am J Physiol*. 1985;248:F858–F868.
87. Beguin P, Crambert G, Guennoun S, et al. CHIF, a member of the FXYD protein family, is a regulator of Na,K-ATPase distinct from the gamma-subunit. *EMBO J*. 2001;20:3993–4002.
88. Wald H, Goldstein O, Asher C, et al. Aldosterone induction and epithelial distribution of CHIF. *Am J Physiol*. 1996;271:F322–F329.
89. Brennan FE, Fuller PJ. Acute regulation by corticosteroids of CHIF mRNA in the distal colon. *Endocrinology*. 1999;140:1213–1218.
90. Manis AD, Hodges MR, Staruschenko A, et al. Expression, localization, and functional properties of inwardly rectifying K+ channels in the kidney. *Am J Physiol Renal Physiol*. 2020;318:F332–F337.
91. Yoo D, Kim BY, Campo C, et al. Cell surface expression of the ROMK (Kir 1.1) channel is regulated by the aldosterone-induced kinase, sgk-1, and protein kinase A. *J Bio Chem*. 2003;278:23066–23075.
92. Glover M, O'Shaughnessy KM. Molecular insights from dysregulation of the thiazide-sensitive WNK/SPAK/NCC pathway in the kidney: Gordon syndrome and thiazide-induced hyponatraemia. *Clin Exp Pharmacol Physiol*. 2013;40:876–884.
93. Sørensen MV, Saha B, Jensen IS, et al. Potassium acts through mTOR to regulate its own secretion. *JCI Insight*. 2019;5:e126910.
94. Wall SM, Verlander JW, Romero CA. The renal physiology of pendrin-positive intercalated cells. *Physiol Rev*. 2020;100:1119–1147.
95. Shibata S, Rinehart J, Zhang J, et al. Mineralocorticoid receptor phosphorylation regulates ligand binding and renal response to volume depletion and hyperkalemia. *Cell Metab*. 2013;18:660–671.
96. Greenlee MM, Lynch IJ, Gumz ML, et al. Mineralocorticoids stimulate the activity and expression of renal H+, K+-ATPases. *J Am Soc Nephrol*. 2011;22:49–58.
97. Seyberth HW, Weber S, Kömhoff M. Bartter's and Gitelman's syndrome. *Curr Opin Pediatr*. 2017;29:179–186.
98. Mabillard H, Sayer JA. The molecular genetics of Gordon syndrome. *Genes (Basel)*. 2019;10:986.
99. Hays TR. Mineralocorticoid modulation of apical and basolateralmembrane H+/OH-/HCO3 transport processes in the rabbit inner stripe of outer medullary collecting duct. *J Clin Invest*. 1992;90:1–8.
100. Dooley R, Harvey BJ, Thomas W. Non-genomic actions of aldosterone: from receptors and signals to membrane targets. *Mol Cell Endocrinol*. 2012;350:223–234.
101. McCurley A, Pires PW, Bender SB, et al. Direct regulation of blood pressure by smooth muscle cell mineralocorticoid receptors. *Nat Med*. 2012;18:1429–1433.
102. Iqbal J, Andrew R, Cruden NL, et al. Displacement of cortisol from human heart by acute administration of a mineralocorticoid receptor antagonist. *J Clin Endocrinol Metab*. 2014;99:915–922.
103. Biwer LA, Wallingford MC, Jaffe IZ. vascular mineralocorticoid receptor: evolutionary mediator of wound healing turned harmful by our modern lifestyle. *Am J Hypertens*. 2019;32:123–134.
104. Mihailidou AS, Tzakos AG, Ashton AW. Non-genomic effects of aldosterone. *Vitam Horm*. 2019;109:133–149.
105. Iqbal J, Andrew R, Cruden NL, et al. Displacement of cortisol from human heart by acute administration of a mineralocorticoid receptor antagonist. *J Clin Endocrinol Metab*. 2014;99:915–922.
106. Fletcher EK, Morgan J, Kennaway DR, et al. Deoxycorticosterone/salt-mediated cardiac inflammation and fibrosis are dependent on functional CLOCK signaling in male mice. *Endocrinology*. 2017;158:2906–2917.
107. Young MJ, Funder JW. Mineralocorticoid receptors and pathophysiological roles for aldosterone in the cardiovascular system. *J Hypertens*. 2002;20:1465–1468.
108. Pitt B, Zannad F, Remme WJ, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999;341:709–717.
109. Pitt B, Remme W, Zannad F, et al. Eplerenone post-acute myocardial infarction heart failure efficacy and survival study investigators. *N Engl J Med*. 2003;348:1309–1321.
110. Zannad F, McMurray JJ, Krum H, et al. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med*. 2011;364:11–21.
111. Cole TJ, Young MJ. Mineralocorticoid receptor null mice: informing cell-type-specific roles. *J Endocrinol*. 2017;234:T83–T92.
112. de Kloet ER, Joëls M. Brain mineralocorticoid receptor function in control of salt balance and stress-adaptation. *Physiol Behav*. 2017;178:13–20.
113. Gasparini S, Resch JM, Narayan SV, et al. Aldosterone-sensitive HSD2 neurons in mice. *Brain Struct Funct*. 2019;224:387–417.
114. Resch JM, Fenselau H, Madara JC, et al. Aldosterone-sensing neurons in the NTS exhibit state-dependent pacemaker activity and drive sodium appetite via synergy with angiotensin II signaling. *Neuron*. 2017;96:190–206.

89

Glucocorticoid Receptors: Mechanisms of Action in Health and Disease

Robert H. Oakley, Stephen R. Hammes, Donald B. DeFranco, and John A. Cidlowski

OUTLINE

Introduction, 1484
Glucocorticoid Receptor Domain Organization, 1484
Glucocorticoid Receptor Regulation of Transcription, 1484
Glucocorticoid Receptor Polymorphisms, 1486
Glucocorticoid Receptor Isoforms Derived From Single Gene, 1486
Glucocorticoid Receptor Posttranslational Modifications, 1488
Glucocorticoid Receptor Physiology, 1489
- Immune System, 1489
- Cardiovascular System, 1490
- Central Nervous System, 1490
- Hepatic Carbohydrate and Lipid Metabolism, 1490

Glucocorticoid Receptor Ligands and Pulsatility, 1491
Glucocorticoid Receptor Resistance, 1492
Summary and Future Directions, 1492

INTRODUCTION

Glucocorticoids are primary stress hormones necessary for life that affect virtually every aspect of human physiology. They are synthesized by and released from the adrenal glands following activation of the hypothalamic-pituitary-adrenal (HPA) axis in a circadian manner and in response to physical and emotional stress. Synthetic derivatives of glucocorticoids are one of the most commonly prescribed classes of drugs in the world today.[1] The physiological actions of glucocorticoids include highly effective antiinflammatory and immunomodulatory actions that are exploited for the treatment of diseases such as asthma, arthritis, allergic rhinitis, ulcerative colitis, multiple sclerosis, and leukemia/lymphoma.[2,3] In addition, glucocorticoids have important roles in lung development, skeletal growth, behavior, reproduction, metabolism, and cardiovascular function. Ultimately, glucocorticoids act to maintain homeostasis in the face of stressful stimuli. The broad actions of glucocorticoids account for the serious side effects commonly experienced with chronic glucocorticoid treatment. These include the development of glucocorticoid resistance in diseased tissues, osteoporosis, growth retardation in children, hypertension, muscle atrophy, and metabolic syndrome.[1,4] The physiological and pharmacological actions of glucocorticoids are mediated by the glucocorticoid receptor (GR), a member of the nuclear receptor superfamily of proteins that regulate gene transcription in a ligand-dependent manner.

GLUCOCORTICOID RECEPTOR DOMAIN ORGANIZATION

Like other members of the nuclear receptor superfamily, GR is a modular protein comprised of an N-terminal transactivation domain (NTD), a central DNA-binding domain (DBD), and a C-terminal ligand-binding domain (LBD) (Fig. 89.1).[5] Separating the DBD and LBD is a flexible region of the protein termed the hinge region. The DBD is the most conserved region across the nuclear receptor superfamily. It contains two zinc finger motifs that recognize and bind target DNA sequences, termed glucocorticoid-responsive elements (GREs). The NTD is the most variable region among family members and contains a strong transcriptional activation function (AF1) that interacts with coregulators and the basal transcription machinery. The LBD forms a hydrophobic pocket for binding glucocorticoids and contains a second transcriptional activation function (AF2) that interacts with coregulators in a ligand-dependent manner. The LBD also possesses a nuclear localization signal and sites for interaction with chaperone proteins and other transcription factors. An additional nuclear localization signal spans the junction of the DBD and hinge region. The GR protein is a substrate for various types of posttranslational modifications, including phosphorylation, ubiquitination, sumoylation, and acetylation that regulate GR function (Fig. 89.1).

GLUCOCORTICOID RECEPTOR REGULATION OF TRANSCRIPTION

Unliganded GR is located in the cytoplasm of cells as part of a multiprotein complex that includes chaperone proteins (HSP90, HSP70, and P23) and immunophilins of the FK506 family (FKBP51 and FKBP52) (Fig. 89.2).[6] The chaperone proteins maintain the receptor in a transcriptionally inactive state that favors high-affinity ligand binding. Upon binding glucocorticoids, GR undergoes a conformational change resulting in the dissociation of the associated proteins, exposure of the nuclear localization signals, and translocation of the receptor into the nucleus via nuclear pores. The receptor then induces or represses gene transcription by binding directly to GREs and/or by interacting with other transcription factors and modulating their activity.

The consensus GRE sequence is an imperfect palindrome (GGAACAnnnTGTTCT) that is comprised of two 6-bp half sites separated by a three-nucleotide spacer. GR binds this element as a homodimer, with each half site occupied by one receptor subunit (Fig. 89.2A). The interaction of ligand-bound GR with the GRE results in the enhanced expression of numerous genes. For this reason, the GRE is often referred to as an activating or positive GRE. However, GR occupancy of the canonical GRE can also lead to the repression of target genes, suggesting a critical role for factors outside the GRE sequence, such as epigenetic regulators and chromatin context, in determining the polarity of the transcriptional response.[7] A negative GRE (nGRE)

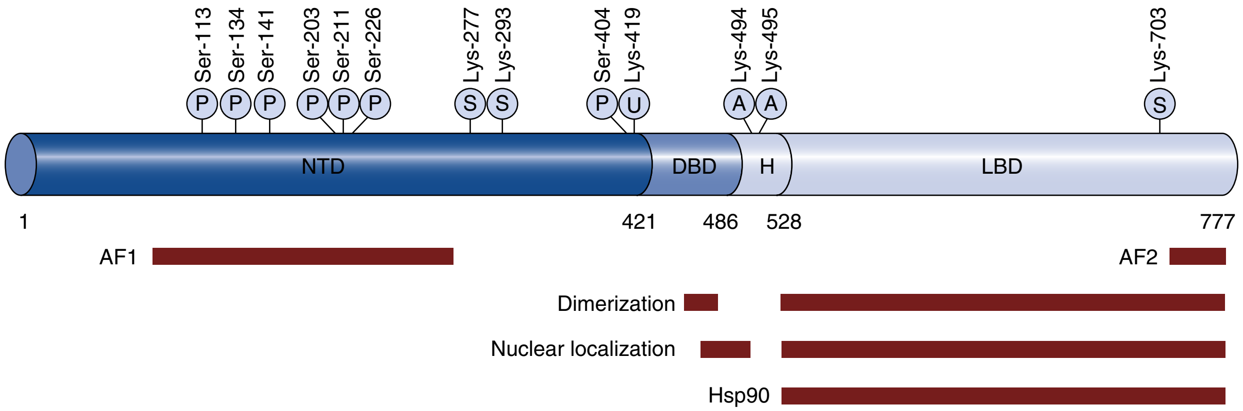

Ser-113
Ser-134
Ser-141
Ser-203
Ser-211
Ser-226
Lys-277
Lys-293
Ser-404
Lys-419
Lys-494
Lys-495
Lys-703
P
P
P
P
P
P
S
S
P
U
A
A
S
NTD
DBD
H
LBD
1
421
486
528
777
AF1
AF2
Dimerization
Nuclear localization
Hsp90

Figure 89.1 **Glucocorticoid receptor (GR) domain structure and sites of posttranslational modification.** GR is composed of an N-terminal transactivation domain (NTD), DNA-binding domain (DBD), hinge region (H), and ligand-binding domain (LBD). Regions involved in transactivation (AF1 and AF2), dimerization, nuclear localization, and Hsp90 binding are indicated. Also shown are some of the amino acid residues posttranslationally modified by phosphorylation (P), sumoylation (S), ubiquitination (U), and acetylation (A). Numbers are for the human GR.

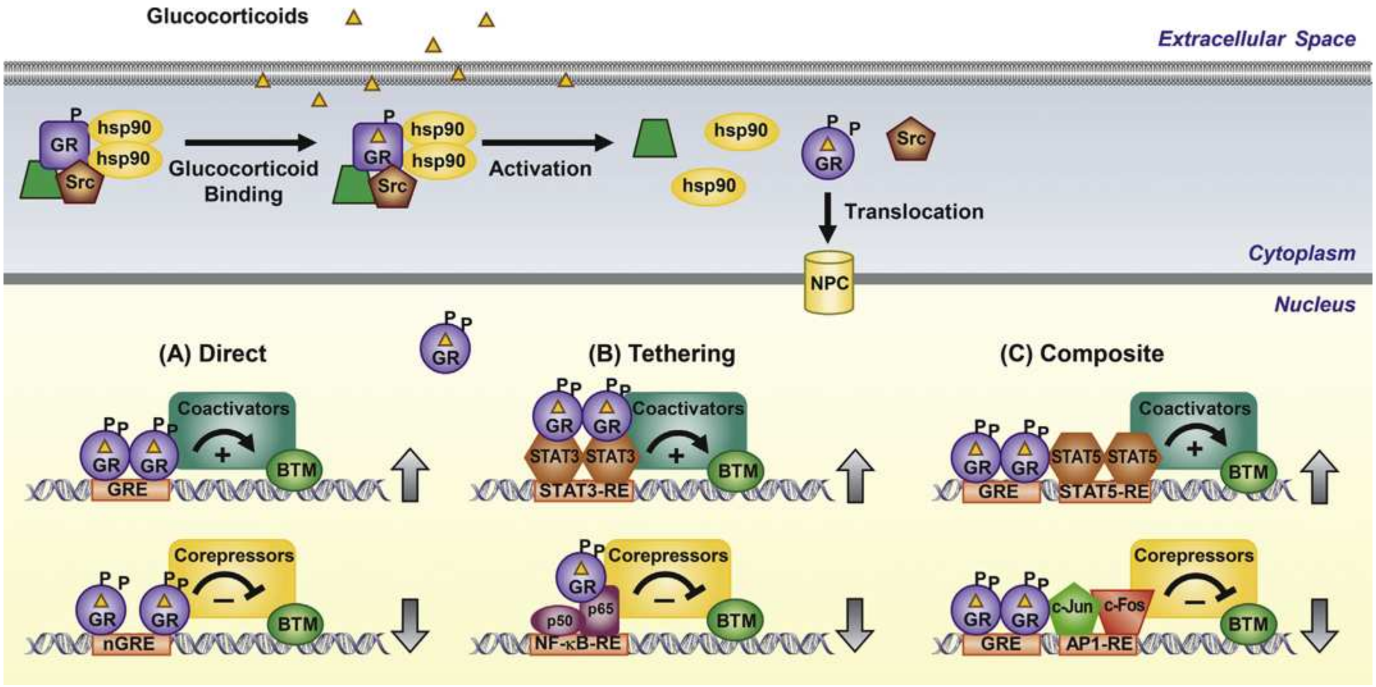

Glucocorticoids
Extracellular Space
GR
hsp90
hsp90
Src
P
Glucocorticoid Binding
Activation
Src
Translocation
Cytoplasm
NPC
Nucleus
(A) Direct
(B) Tethering
(C) Composite
Coactivators
GR
GRE
BTM
STAT3
STAT3-RE
STAT5
STAT5-RE
Corepressors
nGRE
p50
p65
NF-κB-RE
c-Jun
c-Fos
AP1-RE

Figure 89.2 **Glucocorticoid receptor (GR) signaling pathways.** The unliganded GR resides in the cytoplasm of cells in a complex with other proteins. Upon binding glucocorticoids, the receptor undergoes a change in conformation (activation), dissociates from accessory proteins, and translocates into nucleus via the nuclear pore complex (NPC). Nuclear GR regulates gene expression in three primary ways: binding directly to DNA (A), tethering itself to other DNA-bound transcription factors (B), or binding directly to DNA and interacting with neighboring DNA-bound transcription factors (C). *BTM*, Basal transcription machinery.

that differs in sequence from the classic GRE has also been described that mediates glucocorticoid-dependent repression of specific genes (Fig. 89.2A).[8] The consensus nGRE sequence, CTCC(n)$_{0\text{-}2}$GGAGA, is palindromic but has a variable spacer and is occupied by two individual GR monomers rather than a homodimer.[9] Binding sites for GR are found not only within promoters of glucocorticoid-responsive genes but also in intragenic and intergenic regions that can be far removed from the transcription start site.[10] For example, glucocorticoids have been shown to induce the expression of the β-arrestin-1 gene (*ARRB1*) via an intron 1 GRE and to repress the expression of the β-arrestin-2 gene (*ARRB2*) via an intron 11 nGRE.[11] Glucocorticoid regulation of the β-arrestin-1/β-arrestin-2 ratio alters the signaling profiles of G protein-coupled receptors and may account for the superior clinical efficacy of the glucocorticoid/beta-2-adrenergic receptor agonist combination therapies used to treat asthma and chronic obstructive pulmonary disease. Another example of clinical importance is that GR occupancy of an nGRE located in intron 6 of the *GR* gene itself mediates homologous downregulation of GR expression,[12] a mechanism of glucocorticoid resistance that limits the therapeutic response to glucocorticoids.

Many GREs are found in the genome, yet only a small fraction of these sites is actually bound by the receptor. The specific sites of

GR occupancy vary in a tissue-specific manner due to differences in chromatin landscape that regulate the accessibility of the GRE.[13] Glucocorticoid concentration can also govern GR binding to specific GREs and subsequent regulation of target gene expression.[14] For some GREs, GR occupancy occurs at very low concentrations of hormone. Other GREs, however, require high doses of glucocorticoids for GR binding to take place. Distinguishing hypersensitive from hyposensitive GREs is an important endeavor because low-dose glucocorticoid therapy targeting hypersensitive sets may avoid the harmful side effects of higher glucocorticoid doses frequently used in the clinic. The GR–GRE interaction is short-lived, as the receptor rapidly cycles on and off target sites every few seconds.[15] Upon binding DNA, GR undergoes a conformational change that leads to the recruitment of coregulators and chromatin remodeling complexes that regulate RNA polymerase II activity and thereby alter the transcription rates of target genes.[16-18] Steroid receptor coactivators (SRC1–3), the histone acetyltransferases CBP/p300, and the nuclear methylase coactivator-associated arginine methyltransferase (CARM1) are coregulators that mediate transcriptional activation, whereas nuclear receptor corepressor 1 (NCoR) and silencing mediator of retinoid and thyroid hormone receptors (SMRT) are corepressors that mediate transcriptional repression. The specific cofactors that are assembled and their activity on glucocorticoid-responsive genes depends on the ligand bound to GR, as well as the specific GRE sequence occupied by the receptor.[19,20]

In response to glucocorticoids, GR can also regulate gene expression by interacting with other transcription factors (Fig. 89.2B,C). The two most studied examples of this form of regulation involve the transcription factors activator protein 1 (AP-1) and nuclear factor-κB (NF-κB). These two proteins are central mediators of the inflammatory and immune response, and their inhibition by GR is thought to underlie the major antiinflammatory and immunosuppressive actions of glucocorticoids.[5] When activated by stress signals such as proinflammatory cytokines, bacterial and viral infectious agents, or proapoptotic stimuli, AP-1 and NF-κB bind their cognate response elements and induce the expression of many proinflammatory genes, including those encoding cytokines, cell adhesion molecules, and enzymes involved in tissue destruction. GR represses the activity of AP-1 and NF-κB through a direct interaction with the c-Jun subunit of AP-1 and the p65 subunit of NF-κB. For some genes, the inhibition is accomplished by GR tethering itself to these DNA-bound proteins without the receptor binding to DNA (Fig. 89.2B). For other genes, the repression involves GR acting in a composite manner in which the receptor both binds a GRE and physically associates with AP-1 or NF-κB bound to neighboring DNA sites (Fig. 89.2C). Glucocorticoids can also indirectly antagonize the actions of AP-1 and NF-κB by inducing the expression of antiinflammatory genes encoding factors such as IκB protein, which sequesters NF-κB in the cytoplasm,[21] MAPK phosphatase 1, which dephosphorylates c-Jun N-terminal kinase to prevent activation of AP-1,[22] and tristetraprolin, which destabilizes the mRNA of many AP-1- and NF-κB-induced genes.[23] In contrast to its repressive actions on AP-1 and NF-κB, GR interacts with specific members of the signal transducer and activator of transcription (STAT) family of transcription factors, either apart from (tethering action) or in conjunction with (composite action) GRE binding, and enhances their activity on responsive genes (Fig. 89.2B,C).[24] STAT transcription factors are activated by a range of cytokines through induction of the Janus kinase pathway (JAK) and tyrosine phosphorylation. Upon binding their cognate response elements, STATs regulate genes involved in the immune response, differentiation, survival, and apoptosis.

Global gene expression analyses indicate that up to 10% to 20% of the genome is regulated by glucocorticoids.[5] Many of these genes, termed primary glucocorticoid-responsive genes, are regulated by GR directly binding to DNA (GREs and nGREs) and/or interacting with other DNA-bound transcription factors that control the expression of the specific gene. However, GR can also have a profound influence on the genome in an indirect manner. These secondary gene changes occur more slowly and are mediated by the protein products of the primary glucocorticoid-responsive genes. This second wave of gene expression changes can be accomplished by transcription factors that are direct targets of GR regulation. For example, in the heart, glucocorticoid-activated GR increases the expression of the zinc finger DNA-binding transcription factor Kruppel like factor 13 (KLF13) via binding a GRE located in intron 1 of the *KLF13* gene. The upregulated KLF13, in turn, regulates a large cohort of genes that protect cardiomyocytes from cell death induced by the antineoplastic drug doxorubicin.[25]

KEY POINTS

- Upon glucocorticoid binding, cytoplasmic glucocorticoid receptor (GR) translocates to the nucleus.
- GR binds directly to glucocorticoid-responsive elements (GREs) or negative GREs to regulate target gene expression.
- GR also interacts with other transcription factors, such as AP-1 and NF-κB, to regulate target gene expression.

GLUCOCORTICOID RECEPTOR POLYMORPHISMS

Several polymorphisms in the *GR* gene have been described that change the amino acid sequence of the encoded receptor and alter its transcriptional activity. The single nucleotide polymorphism N363S, which results in an asparagine to serine substitution in the NTD of GR, occurs in approximately 4% of the population and is associated with glucocorticoid hypersensitivity.[26,27] N363S carriers have been reported to have an increased risk for metabolic syndrome, type 2 diabetes, decreased bone mineral density, and cardiovascular disease. Genome-wide microarrays on isolated macrophages and cultured osteosarcoma cells have revealed unique gene regulatory profiles for the N363S polymorphism compared with wild-type GR in response to glucocorticoid treatment.[28,29] The ER22/23EK polymorphism occurs in approximately 3% of the population and results in an arginine to lysine change at position 23 within the NTD of GR.[26,27] The ER22/23EK polymorphism, in contrast to N363S, exhibits reduced transcriptional activity on target genes and has been associated with glucocorticoid insensitivity. Carriers of this polymorphism have a more favorable metabolic profile due to their decreased sensitivity to glucocorticoids and exhibit a lower tendency to develop type 2 diabetes and cardiovascular disease. The mechanism underlying the glucocorticoid resistance in ER22/23EK carriers may involve alterations in the expressed complement of GR isoforms (discussed later).[30] These GR polymorphisms may account, at least in part, for the heterogeneous responses observed among individuals on glucocorticoid therapy.

GLUCOCORTICOID RECEPTOR ISOFORMS DERIVED FROM SINGLE GENE

GR is expressed throughout the body, yet there is considerable diversity among cell types in terms of their responses to glucocorticoids. The discovery of multiple GR isoforms has provided a molecular mechanism contributing to the specificity of glucocorticoid action. The human *GR* gene is located on chromosome 5q31-32 and is comprised of nine exons (Fig. 89.3). The GR NTD is encoded primarily by exon 2, the DBD is encoded by exons 3 and 4, and the hinge region and LBD are encoded by exons 5 to 9. Alternative splicing in exon 9

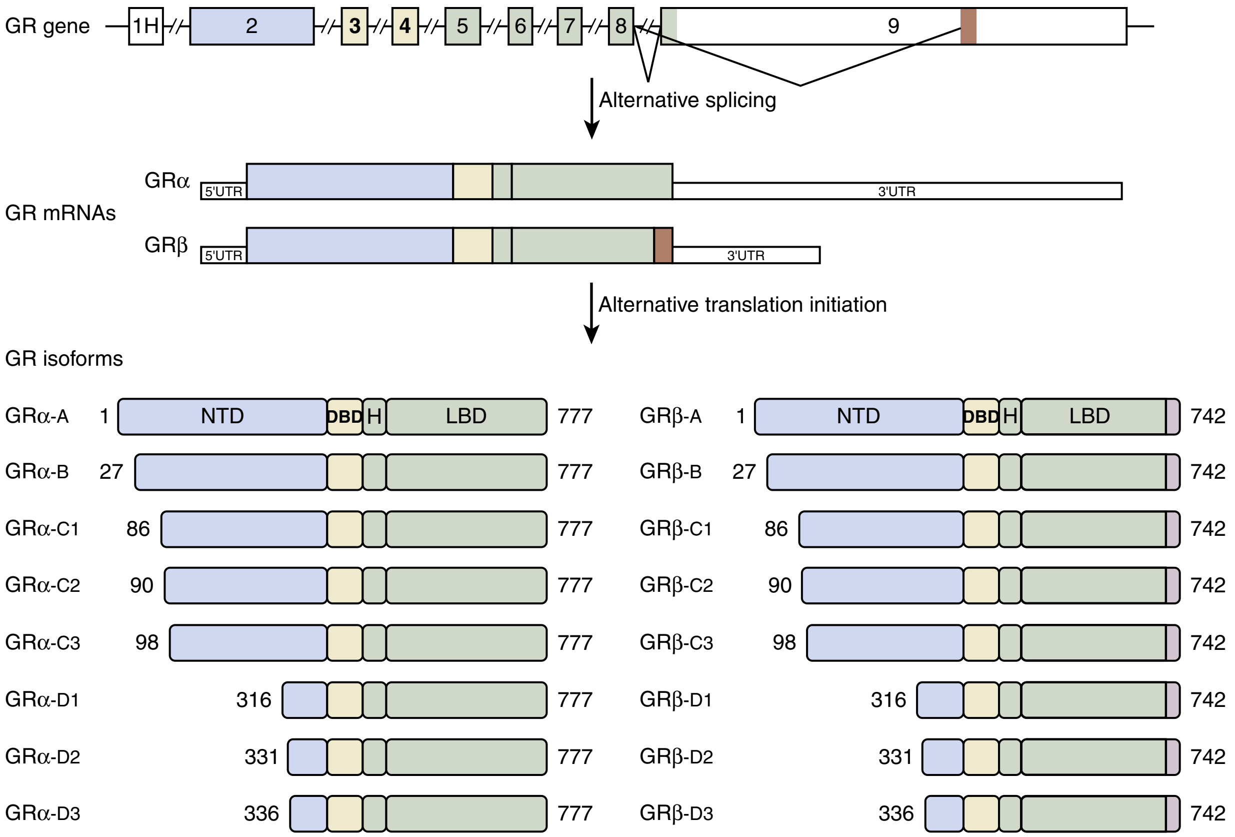


Figure 89.3 **Multiple glucocorticoid receptor (GR) isoforms generated from single gene.** The human *GR* gene is comprised of nine exons. Alternative splicing at the 3′ end of the primary transcript generates GRα and GRβ mRNAs, which encode GRα and GRβ proteins differing only at their C-terminus. Alternative translation initiation from eight different AUG start codons derived from exon 2 generates additional protein isoforms with progressively shorter N-terminal domains. Numbers shown denote the first and last residues for the human GR isoforms. For simplicity, only the most proximal of nine alternate exon 1s (1H) is shown.

generates two receptor isoforms, termed GRα and GRβ, that are identical through amino acid 727 but then diverge at their C-termini (Fig. 89.3).[5,31] The classic, full-length GRα contains an additional 50 amino acids, whereas the splice variant GRβ has an additional, nonhomologous 15 amino acids. Because of its unique C-terminus, GRβ does not bind glucocorticoids and resides constitutively in the nucleus of cells. GRβ has been shown to function as a dominant negative inhibitor and repress the transcriptional activity of GRα[32,33]; therefore, alterations in GRβ expression may contribute to changes in glucocorticoid responsiveness. Indeed, expression of GRβ is selectively increased over GRα in cells exposed to proinflammatory cytokines and microbial superantigens leading to reduced sensitivity to glucocorticoids.[5,31] Glucocorticoid-resistant forms of inflammatory diseases, such as asthma, rheumatoid arthritis, and ulcerative colitis, have also been associated with elevated expression of GRβ. Furthermore, upregulation of GRβ and diminished GRα signaling is observed in erythroid cells from patients with polycythemia vera. Conversely, methotrexate, an effective drug for treating autoimmune and inflammatory diseases, promotes a selective increase in GRα at the expense of GRβ and improves the glucocorticoid sensitivity of lymphocytes. The mechanisms underlying the dominant negative activity of GRβ are unclear but may involve competition for GRE binding and/or formation of GRα/GRβ heterodimers that are transcriptionally deficient.

The splicing factor SRp30c has been shown to play a critical role in the selective increase in GRβ expression.[34] In addition, elevated GRβ levels result from the naturally occurring ATTTA to GTTTA polymorphism (A3669G) in the 3’ untranslated region of GRβ. This nucleotide substitution disrupts an mRNA destabilization motif and leads to increased stability of the GRβ mRNA and enhanced protein expression.[35,36] Persons harboring the A3669G allele have an elevated risk for pathologies with inflammatory underpinnings, such as autoimmune disease, coronary artery disease, myocardial infarction, and heart failure, suggesting the rise in GRβ limits the beneficial immunosuppressive actions of GRα.[37-39] Another polymorphism in the GRβ 3’ untranslated region (G3134T) has also been shown to increase GRβ expression by mRNA stabilization.[40] Carriers of the G3134T polymorphism have elevated levels of endogenous glucocorticoids, display glucocorticoid resistance, and exhibit alterations in the glucocorticoid-regulated transcriptome in isolated macrophages.

A broader role for GRβ in cell signaling has recently emerged with the demonstration that this isoform can modulate gene expression apart from its effects on GRα. In a genome-wide microarray analysis performed in human osteosarcoma cells selectively expressing GRβ, the isoform was found to alter the expression of over 5000 genes.[41] Less than 20% of the genes were commonly regulated by ligand-activated GRα, indicating that GRβ possesses its own unique gene-regulatory profile. GRβ was also shown to bind the glucocorticoid antagonist mifepristone (RU486), and binding of this ligand abolished most of the GRβ-mediated changes in gene expression.[41] Interestingly, structural analysis of the GRβ/RU486 complex suggests GRβ is found in a conformation that favors corepressor binding.[42] To explore the function of GRβ *in vivo*, the splice variant was expressed in the liver of wild-type mice and in GR liver knockout mice using an adeno-associated virus.[43] GRβ antagonized the function of GRα and attenuated hepatic gluconeogenesis. Moreover, GRβ regulated many genes by GRα-independent mechanisms. For example, GRβ upregulated the expression of the *STAT1* gene in the livers of both wild-type and GR liver knockout mice. This induction involved GRβ binding to an intergenic GRE adjacent to the *STAT1* gene and was inhibited by RU486 treatment. As a *bona fide* transcription factor, GRβ may contribute to alterations in glucocorticoid responsiveness in healthy and diseased tissues by genomic effects apart from its dominant negative activity on GRα.

Alternative translation initiation of the GR mRNA produces an additional cohort of receptor subtypes (Fig. 89.3).[44,45] Eight GR translational isoforms with progressively shorter NTDs are generated from one GRα mRNA transcript via different AUG start codons: GRα-A, GRα-B, GRα-C1, GRα-C2, GRα-C3, GRα-D1, GRα-D2, and GRα-D3. The eight initiation sites are located in exon 2 and are highly conserved across species. GRα-A is the classic, full-length 777-amino acid human receptor that is generated from the first initiator AUG codon. The GRβ mRNA also contains the identical start codons and would be expected to give rise to a similar complement of subtypes.

The GRα isoforms, all of which retain an intact DBD and LBD, bind glucocorticoids with similar affinity and show a similar capacity to interact with GREs.[46] Additionally, all eight isoforms occupy the nucleus of cells following glucocorticoid treatment. In the absence of hormone, however, the subcellular distribution of the subtypes differ with the GRα-D isoforms residing predominantly in the nucleus and the others predominantly in the cytoplasm.[44,46] Marked differences have been reported in the transcriptional properties of the GRα translational isoforms.[46,47] Global microarray analysis of cells selectively expressing individual GRα isoforms revealed that each receptor subtype displays a unique gene-regulatory profile. Less than 10% of the glucocorticoid-responsive transcriptome was commonly regulated by all the isoforms, indicating that the vast majority of genes were regulated in an isoform-specific manner. The GRα-C3 isoform is the most transcriptionally active subtype, whereas GRα-D isoforms are the least active. The heightened activity of GRα-C3 has been linked to a motif in the NTD that sterically hinders the larger isoforms.[48] The isoform-unique gene-regulatory profiles were further shown to produce functional differences in cellular responsiveness to glucocorticoids, as the GRα translational isoforms exhibited distinct capabilities to induce apoptosis.[46,47] Cells expressing GRα-C3 were the most sensitive to the apoptosis-inducing actions of glucocorticoids, whereas cells expressing GRα-D3 were the most resistant. Unlike other GRα translational isoforms, GRα-D3 did not repress NF-κB activity and failed to inhibit the transcription of certain antiapoptotic genes.[49]

The GRα translational isoforms show a widespread tissue distribution; however, their relative levels vary across cell types. GRα-A and GRα-B are the most highly expressed isoforms in many cells. However, GRα-C and GRα-D are abundant in trabecular meshwork cells,[50] and the GRα-D isoform predominates in immature dendritic cells.[51] In rodents, GRα-C isoforms are expressed at high levels in the pancreas and colon, whereas GRα-D isoforms are abundant in the spleen and lungs.[44] In the human brain, the composition of GRα translational isoforms changes during development and aging.[52] Additionally, a selective increase in the GRα-D isoform has been observed in the brains from patients with schizophrenia and bipolar disorder.[53,54] Differential expression of GRα translational isoforms may be a major mechanism behind the diverse tissue sensitivity and responses to glucocorticoids. *In vivo* support for this conclusion comes from knockin mice that exclusively express the GRα-C3 isoform.[55] The GRα-C3 knockin mice die at birth due to respiratory distress but can be rescued by antenatal glucocorticoid administration. Rescued GRα-C3 mice exhibited alterations in sensitivity to lipopolysaccharide (LPS)-induced inflammation, and gene expression studies revealed a deficiency in the ability of GRα-C3 to regulate a large cohort of immune and inflammatory response genes in the spleen.

KEY POINTS

- The single glucocorticoid receptor (GR) gene gives rise to multiple protein isoforms.
- Alternative slicing in exon 9 of the GR gene produces the GRβ splice variant.
- Alternative translation initiation of the GR mRNA produces eight GRα translational isoforms with progressively shorter N-terminal domains.
- The various GR isoforms, with their unique expression and gene-regulatory profiles, contribute to tissue specificity in glucocorticoid responsiveness.

GLUCOCORTICOID RECEPTOR POSTTRANSLATIONAL MODIFICATIONS

In addition to the diversity in cellular response to glucocorticoids driven by changes in the coding sequence of GR (i.e., via polymorphisms, splicing, or translational isoforms), posttranslational modifications to the receptor can have a profound effect on glucocorticoid-regulated transcription.[56-59] Phosphorylation is the most well characterized, but not the only, posttranslational modification of GR and was initially described in the mid-1970s. It was proven to occur in the late 1980s and early 1990s with the identification of specific phosphorylated amino acid residues within the receptor. The application of sophisticated mass spectrometry methods and assessment of distinct model systems have led to more recent identification of new GR phosphorylation sites previously unrecognized. Currently, 16 highly conserved amino acids within the GR (15 within the amino terminal domain and one within the hinge region) have been shown to be phosphorylated, with 15/16 occurring within serine and/or threonine residues and only one at a tyrosine residue. Many of the GR phosphorylation sites are regulated by ligand binding to the receptor, with two recently identified phosphorylation sites induced by brain-derived neurotrophic factor (BDNF).

Studies using pharmacological approaches to modulate GR phosphorylation first pointed to a role for site-specific GR phosphorylation in directing distinct transcriptional responses of the receptor. The application of site-directed mutagenesis, as well as biochemical and genomic approaches, revealed the mechanistic basis for distinct transcriptional responses directed by GR with site-specific phosphorylation. Thus, each phosphorylation site on GR presents a unique surface that impacts the binding affinity of a wide variety of cofactors.[60-62] These selective phospho-GR/cofactor complexes are likely to be differentially stabilized at direct or indirect GR binding sites on the genome, ultimately leading to selective modulation of transcriptional output from linked glucocorticoid-responsive promoters. Furthermore, this modulation of GR transcriptional activity will be responsive to many physiological signals or drugs that impact the activity of intracellular protein kinases and phosphatases that directly act on the receptor. For example, long-acting beta-2-adrenergic agonists, which are used in combination with inhaled corticosteroids to treat moderate or serve persistent asthma, may enhance the antiinflammatory effects of glucocorticoids through modulation of site-specific GR phosphorylation.[63,64]

In many cases, site-directed mutagenesis has confirmed the biological effects of specific GR phosphorylation sites. For example, GR phosphorylation at serine 134 is elevated in triple negative breast cancer

(TNBC) relative to other breast cancers.[65] Furthermore, *in vitro* experiments with TNBC cells established the role for serine 134 phosphorylation in the migratory, invasive, and stem cell properties of these cells, demonstrating that GR phosphorylated at serine 134 generates a unique transcriptome. The relevance of these *in vitro* results was suggested from analysis of gene expression profiles of breast cancer patient samples, which reveal poorer survival outcomes in those stratified patients with relatively higher expression levels of the GR phosphoserine-134 gene signature. In mice, BDNF-regulated GR phosphorylation plays a role in the retention of newly learned motor skills through a mechanism that involves the stabilization of newly formed postsynaptic dendritic spines in the motor cortex.[66] The impact of BDNF on multiple brain circuits regulating resilience to stress could be driven in part by regulation of specific GR phosphorylation events.[59] Assessment of site-specific phosphorylation of GR may be a useful clinical biomarker for diseases influenced by altered GC signaling, as evidenced by studies associating changes in site-specific receptor phosphorylation with severe asthma[67] and type 2 diabetes.[68]

In addition to dynamic changes in GR phosphorylation in response to physiologic or therapeutic inputs, receptor phosphorylation is also likely influenced by cell- and tissue-specific factors. GR phosphorylation is prominent in the central nervous system,[69] and many studies in animal models have identified site-specific modulations of receptor phosphorylation within select brain regions in response to acute and chronic insults or various types of stress.[70-72] Some of the same GR phosphorylation sites implicated in rodent stress responses are altered in peripheral blood mononuclear cells (PBMCs) of humans with neuropsychiatric conditions, in some cases in a sex-specific manner.[73,74] Thus, a comprehensive analysis of GR phosphorylation in PBMCs may provide some insights into changes in GC signaling in the brain that are associated with select neuropsychiatric disorders.

Other posttranslational modifications in addition to phosphorylation have been shown to modulate GR function. Degradation of the GR through the ubiquitin-proteasome pathway requires modification of a specific receptor lysine residue (lysine 419 in human GR) by ubiquitin.[75] This posttranslational modification involves the concerted action of three specific enzymes culminating in the covalent addition of the 79-amino acid ubiquitin moiety on GR by an E3 ligase enzyme. A number of E3 ubiquitin ligases have been identified that catalyze the ubiquitylation and subsequent degradation of GR,[76-78] and in some cases, the signaling pathways that regulate GR ubiquitylation and turnover have been identified and biological effects revealed.[79] Small ubiquitin-like modifiers are small proteins that are also covalently attached to GR using an analogous multienzyme pathway related to the ubiquitin pathway. However, rather than target the receptor for degradation, sumoylation of GR influences its promoter-specific transcriptional regulatory activity.[80-82] Finally, GR has been shown to be subjected to nitration[83] and acetylation,[84,85] which may regulate the receptor's action in antiinflammatory and circadian pathways, respectively.

KEY POINTS

- Posttranslational modifications of glucocorticoid receptor (GR) include phosphorylation, ubiquitination, sumoylation, nitration, and acetylation.
- Site-specific phosphorylation of GR directs unique transcriptional responses by regulating the receptor's interaction with various cofactors.
- Site-specific phosphorylation of GR has specific physiological and pathological effects.
- Site-specific ubiquitination of GR alters the degradation properties of the receptor.

GLUCOCORTICOID RECEPTOR PHYSIOLOGY

Glucocorticoids have major effects on the organ systems of the body. Understanding the physiological actions of glucocorticoids has been aided by the development of novel mouse strains with conditional knockout of GR in specific cell types.[86] The following sections summarize recent progress elucidating the role of glucocorticoids and GR signaling in the immune system, cardiovascular system, central nervous system, and metabolism.

Immune System

Proper modulation of immune system activity and inflammation is critical to normal human function. A blunted immune response leaves the door open for potentially fatal infections, whereas an overstimulated immune response can result in autoimmune activity that damages organs. Glucocorticoids are widely used in the clinic for their ability to suppress the immune system and inhibit inflammation.[87,88] The glucocorticoid-mediated immunosuppression impacts both the innate and adaptive immune response and involves multiple mechanisms, including the stimulation of antiinflammatory genes, repression of proinflammatory genes (primarily by inhibiting NF-κB and AP-1 activity), and regulation of genes involved in apoptosis. The innate immune system is the first line of defense in response to pathogen or injury and is mediated by leukocytes, macrophages, and dendritic cells, among others. The adaptive immune response is a second line of defense mediated by T and B lymphocytes. Glucocorticoids achieve their immunosuppressive effects by acting on nearly all cell types of the immune system.

Macrophages play a central role in innate immunity via the expression of pattern recognition receptors that sense infectious agents and harmful signals. This leads to activation of the inflammasome complex and the release of proinflammatory cytokines. GR signaling has been shown to suppress the immune response in macrophages as mice lacking GR specifically in macrophages exhibit greater mortality and increased cytokine production in response to LPS than wild-type mice.[89] A critical role for macrophage GR in the downregulation of inflammatory cytokines and chemokines in contact dermatitis has also been demonstrated in macrophage GR knockout mice.[90] Interestingly, glucocorticoids have also been shown to augment the proinflammatory response in macrophages by inducing expression of inflammasome components and specific Toll-like receptors and by promoting macrophage migration via the upregulation of exopeptidase dipeptidyl peptidase-4.[88,91] Thus, macrophage GR signaling may function initially to prime the immune system to respond to injury and then subsequently to limit the inflammatory response and restore homeostasis.

Dendritic cells are antigen-presenting cells that coordinate both innate and adaptive immunity. Glucocorticoids suppress the maturation and migration of dendritic cells and stimulate their death by apoptosis.[92] Mice lacking GR in dendritic cells have revealed that glucocorticoids protect against LPS-induced sepsis by suppressing the expression of the proinflammatory cytokine IL-12 by dendritic cells.[93]

Proinflammatory T-cells of the adaptive immune response undergo apoptosis in response to glucocorticoids.[92] Mice with conditional knockout of GR in T-cells have revealed that glucocorticoid signaling limits potentially lethal cytokine production by T-cells in response to immune activation.[94] In addition, studies on T-cell-specific GR-deficient mice have shown that the beneficial actions of glucocorticoids in models of rheumatoid arthritis, multiple sclerosis, sepsis, and helminth infection are mediated by GR signaling events that inhibit T-cell function, alter T-cell migration, and/or lead to thymocyte cell death.[95-99] Furthermore, it has been demonstrated in these conditional knockout mice that GR signaling promotes T-cell selection.[100] The function of B-cells is also modulated by glucocorticoids. Glucocorticoids regulate B-cell activation, survival, proliferation, and differentiation.[101] Moreover, it was recently demonstrated in mice with B-cell–specific deletion of GR that glucocorticoid signaling is required for proper migration of mature B-cells from blood to bone marrow.[102]

Many immune genes are regulated by GR in a sex-specific manner.[103] In rats, glucocorticoids were shown to regulate a greater number of inflammatory genes and elicit a greater antiinflammatory response in males compared with females.[104] Sexual dimorphism in the immunosuppressive actions of glucocorticoids not only provides a mechanistic basis for the predisposition of females to develop autoimmune diseases but also suggests that glucocorticoid dosages used in the clinic should be optimized based on the sex of the patient.

Cardiovascular System

Glucocorticoids can have both positive and negative effects on the heart.[105] The contractile performance of the heart is improved by glucocorticoid treatment, and cardiomyocyte cell death triggered by various insults can be blocked by glucocorticoids. However, elevated glucocorticoids can also lead to bradycardia, cardiac hypertrophy, and heart failure. Moreover, major risk factors for cardiovascular disease, including metabolic syndrome and hypertension, are associated with increased levels of glucocorticoids. Differences in physiological context and the magnitude and duration of the glucocorticoid exposure likely play a role in whether the actions of glucocorticoids on the cardiovascular system are beneficial or adverse. The cardiovascular outcomes may also depend on whether glucocorticoids are acting directly on the heart or indirectly via their systemic effects on blood pressure and metabolism.

New insights into the direct actions that glucocorticoids have on the heart have come from transgenic mouse models with altered expression of GR in cardiomyocytes.[105] Mice overexpressing GR in cardiomyocytes have conduction defects, including bradycardia and atrioventricular block.[106] Prenatal studies performed on mice with global knockout of GR or selective knockout of GR in cardiomyocytes revealed a critical role for GR in the maturation of the fetal heart.[107,108] Glucocorticoid signaling in fetal cardiomyocytes was necessary to maintain appropriate expression levels of genes important for cardiomyocyte architecture and function. Using adult mice bearing a conditional deletion of GR in cardiomyocytes, several laboratories have demonstrated that GR signaling is required for normal heart physiology and function.[109-111] Mice lacking cardiomyocyte GR spontaneously develop heart disease characterized by pathological cardiac hypertrophy, left ventricular systolic dysfunction, heart dilatation, and fibrosis, and these knockout mice die prematurely from heart failure. Mechanistically, the absence of cardiomyocyte GR signaling leads to the dysregulation of key genes involved in calcium handling. Interestingly, cardiomyocyte GR has been shown to act in a sexually dimorphic manner, as male mice lacking the receptor were more sensitive to heart disease than their female counterparts.[112] In addition, deleterious signaling by the closely related mineralocorticoid receptor (MR), which can be bound and activated by glucocorticoids, was found to exacerbate the pathology in hearts lacking GR, suggesting that balanced cardiomyocyte GR and MR signaling is critical for normal heart homeostasis.[109]

Central Nervous System

Receptors for glucocorticoids are widely expressed in the central nervous system (CNS) and thereby influence various behaviors, emotional responses, learning and memory, and central control of metabolic and autonomic function. The expression of MR, which is activated by basal levels of natural glucocorticoids, such as cortisol in humans, is much more restricted in the brain than GR and is limited primarily to structures within the limbic system (i.e., hippocampus, amygdala, and prefrontal cortex).[113] Furthermore, MR is responsive to cortisol in the brain, because it lacks 11β-hydroxysteroid dehydrogenase-2, the cortisol inactivating enzyme highly expressed in kidney that blocks MR activation by cortisol in kidney. Extensive studies in humans and animal models established a role for MR in various aspects of memory and executive function, as well as cognitive responses to stress.[114] Both genomic[115] and nongenomic[116] actions of MR are likely to mediate its complex regulation of memory, cognition, and emotional state and driven by dynamic changes in connectivity and neurotransmission that accompany a central response to an emerging stress. As GRs are also eventually engaged in the central response to stress, the dynamic shift from MR- to GR-driven outputs[117] aids in the recognition, processing, and response to variable stresses that is adaptive. Chronic stress, long-term exposure to synthetic glucocorticoids, or disruption of the ultradian and/or circadian rhythm of endogenous glucocorticoid accumulation can alter the dynamic balance between MR and GR signaling in the CNS, leading to maladaptive responses to stress that can contribute to affective disorders.[118]

GRs in the CNS regulate behavior, cognition, and affective functions; however, they also control energy metabolism,[119] acting through specific neuroendocrine cells within the hypothalamus.[120] Regulation of the HPA axis is another function of central GRs, although negative feedback of adrenal glucocorticoid synthesis and secretion also operates at the level of corticotrophs within the anterior pituitary. Glucocorticoid suppression of corticotrophin-releasing hormone (CRH) synthesis and release from neuroendocrine cells within the paraventricular nucleus in the hypothalamus involves numerous pathways and circuits regulated directly and indirectly by central GRs.[121] In the paraventricular nucleus (PVN), rapid inhibition of CRH release is due to glucocorticoids acting indirectly via a nongenomic mechanism to reduce glutamatergic drive to CRH neurons.[121] Specifically, activation of classical GR located within the plasma membrane stimulates the production of endocannabinoids, which act on cannabinoid receptor type 1 within presynaptic terminals that synapse onto CRH neurons to reduce glutamate release.[122,123] As in other examples of nongenomic glucocorticoid actions in the brain,[124] rapid stimulation of endocannabinoid production by glucocorticoids require nuclear GR localized at the plasma membrane.[125] Multiple intracellular pathways are activated by ligand-bound plasma membrane-associated nuclear GR, although mobilization of intracellular calcium stores is one of the more prominent features of rapid GR signaling in neuronal systems.[124,126] In contrast to the inhibitory effects of nongenomic GR signaling on neurotransmission in the PVN, stress-induced activation of plasma membrane GR and MR via glucocorticoids in limbic areas increases glutamate release and is likely driven by intracellular signaling pathways distinct from those that modulate glucocorticoid effects on endocannabinoid production.[127,128]

The completion of central negative feedback control of the HPA axis requires GR action within limbic structures, including the prefrontal cortex and hippocampus.[121,129] While the importance of "delayed feedback" from limbic areas to regulate HPA axis function had been recognized for many years based upon lesion and pharmacologic studies,[129,130] identification of the specific regions and cell types that utilize transcriptional responses of GR to modulate inhibitory inputs into CRH neurons was firmly established using mouse models with cell type–specific deletions of the receptor.[86,131,132]

Hepatic Carbohydrate and Lipid Metabolism

Glucocorticoid effects on carbohydrate and lipid metabolism are widely recognized.[133-135] The homeostatic effects of these hormones on systemic metabolism are due to their promotion of gluconeogenesis in the liver and the mobilization of substrates in peripheral tissues, such as white adipose and skeletal muscle. They also limit glucose update in peripheral tissues and promote glycogen storage in liver. Glucocorticoids also promote gluconeogenesis and lipid metabolism through "permissive" effects on other hormones, including glucagon and catecholamines.[133-135]

The mechanisms responsible for glucocorticoid stimulation of gluconeogenesis in liver have been extensively studied and rely primarily, but not exclusively, on the transcriptional induction by GR of many genes in the gluconeogenic pathway.[135,136] This transcriptional response is essential to physiologic regulation of glucose homeostasis, as evidenced by hypoglycemia caused by adrenal insufficiency. Much of the current understanding of the molecular mechanisms responsible for hepatocyte-specific transcriptional regulation of gluconeogenesis comes from extensive molecular studies of the phosphoenolpyruvate carboxykinase (*PCK1*) gene, which revealed one of the earliest examples of "composite" GREs that include GR and other tissue- or cell type–enriched transcription factors required for a functional glucocorticoid response.[136-138] Members of the forkhead box transcriptional factor family play a prominent role in both glucocorticoid and insulin regulation of many genes in the gluconeogenesis pathway in hepatocytes.[135,139-141]

In mouse liver the C/EBPβ protein is required to "preprogram" chromatin accessible sites to allow for GR binding and an ensuing transcriptional response.[142] However, this is only one level of complexity in GR regulation of gluconeogenesis, as the dynamics of GR cooperation with other transcription factors can change in response to the duration of a fasting state.[143] Specifically, during a short-term fast, GR enhances the recruitment of the cyclic AMP response element-binding protein 1 (CREB1) transcription factor to gluconeogenic gene enhancers, leading to synergistic activation of transcription initially mobilized following glucagon-induced site-specific phosphorylation and activation of CREB1.[144] In contrast, in response to a long-term fast, GR promotes the transcription of ketogenic and fatty acid oxidation pathway genes through the induction of a major regulator of these pathways, the peroxisome proliferator-activated receptor alpha (PPARα).[144]

Glucocorticoids also regulate lipogenesis in hepatocytes and increase triglyceride levels primarily through the inhibition of lipolysis and induction of triglyceride synthesis.[136,145] The impact of glucocorticoids on lipid metabolism in the liver is not limited to direct action of the GR on genes in the lipogenesis pathway, as other signaling pathways affecting glucose metabolism and insulin sensitivity influence the effect of glucocorticoids on lipid metabolism.[136,143,146] For example, the gene encoding the transcriptional repressor Hes1 is a direct GR target gene, and repression of its transcription by glucocorticoids is essential not only for the regulation of lipid metabolism in hepatocytes,[147] but also for many systemic responses to glucocorticoids.[148]

GLUCOCORTICOID RECEPTOR LIGANDS AND PULSATILITY

The physiological response to glucocorticoids occurs in the context of their rhythmic production and secretion from the adrenal gland. Specifically, superimposed upon the daily fluctuations in circulating glucocorticoid levels, which peaks in the early morning and reaches its nadir around midnight in humans, are ultradian rhythms of glucocorticoid secretion that have a periodicity of 60 to 90 minutes.[149] Different physiological pathways underlie the complex regulation of daily versus ultradian rhythms of circulating glucocorticoids, with circadian pacemakers in the hypothalamus predominantly regulating the daily rhythm of glucocorticoid secretion from the adrenal, while ultradian rhythms are mediated by the balance between feedforward control of adrenal glucocorticoid production by pituitary ACTH and negative feedback by glucocorticoids on ACTH production.[150] Various approaches have been used in animal models to reveal an impact of altered glucocorticoid rhythms on metabolism, inflammation, and multiple behavioral outcomes. Alterations in the daily rhythms of glucocorticoids modulate their action on regulators of circadian rhythms (e.g., PER and REV-ERBA) in peripheral organs contributing to disruptions in carbohydrate and lipid metabolism.[151,152] In humans, chronic elevations of endogenous or exogenous glucocorticoids in Cushing syndrome trigger high morbidity and mortality associated with pathophysiological effects on metabolism, cardiovascular function, immune responses, and behavior.[149] These adverse effects may be exacerbated by alterations in circadian clock gene expression, which coordinate many external outputs to optimize systemic metabolic activity. This has led to glucocorticoid replacement therapy regimens for patients with adrenal insufficiency that mimic the natural daily and ultradian rhythm of glucocorticoid secretion (i.e., "chronopharmacotherapy"), which have demonstrated clinical benefit.[153] Molecular changes in cellular response to glucocorticoid chronotherapy have also been reported. Specifically, analysis of peripheral blood mononuclear cells from patients in the Dual Release Hydrocortisone versus Conventional Glucocorticoid Replacement Treatment in Hypocortisolism (DREAM) trial revealed physiological patterns of *CLOCK* gene expression by modified release hydrocortisone but not conventional multiple daily doses.[154] Modified release patients also had better body weight reduction, normal immune profiles, and improved quality of life. Furthermore, in healthy subjects with pharmacological suppression of glucocorticoid synthesis, the restoration of normal daily and ultradian rhythms of cortisol was most effective at maintaining positive mood and "resting state neural dynamics," as revealed by functional neuroimaging.[155] Finally, the administration of synthetic glucocorticoids (i.e., two daily doses 24 hours apart) to pregnant women "out of phase" with their natural daily rhythm of circulating cortisol can have long-lasting effects on the stress responsivity of their infants at 5 years of age.[156] Therefore, glucocorticoid chronotherapy that takes into account the natural rhythm of cortisol may have long-lasting clinical benefit in patients requiring chronic or even acute glucocorticoid treatment.

How do the patterns of glucocorticoid rhythms affect the cellular response of the GR? Synthetic glucocorticoids used therapeutically typically exhibit a higher affinity for GR than cortisol and dissociate much more slowly from the receptor. These differences in the kinetics of ligand occupancy of the receptor modulate the dynamics of GR binding to specific target sites in the genome. Cycles of GR occupancy on target genes are further modified by ATP and chaperone proteins operating in the nucleus.[157,158] At the molecular level, the dynamics of GR binding to genomic sites in cells occurs at a much more rapid time scale (i.e., seconds) than ultradian pulses of natural glucocorticoids.[15] This inherent property of the GR and other transcription factors may aid in the stochastic sampling of preassembled macromolecular complexes within the genome and facilitate the selectivity of transcriptional responses of the receptor. Finally, dynamic chromatin occupancy of the GR in cells and animal models driven by pulsatile exposure to natural glucocorticoids that mimic ultradian rhythms can impact gene-specific responses to glucocorticoids.[159] Therefore, enhanced clinical efficacy of glucocorticoid chronotherapy may be driven by natural cycles in the genomic structure and organization of GR-responsive genes that require the appropriate assembly of receptor-interacting factors for an efficient transcriptional response to glucocorticoid exposure.

KEY POINTS

- Glucocorticoids are secreted by the adrenal gland in circadian and ultradian rhythms.
- Glucocorticoid chronotherapy that mimics the natural rhythm of circulating cortisol has demonstrated clinical benefit.

GLUCOCORTICOID RECEPTOR RESISTANCE

Generalized glucocorticoid resistance (GGR) or "Chrousos syndrome" is a rare disorder characterized by ACTH-dependent hypercortisolism and the absence of clinical features of glucocorticoid excess.[160,161] The glucocorticoid sensitivity of all tissues is decreased in GGR. Physiologically normal serum cortisol concentrations are therefore insufficient to suppress CRH and ACTH secretion from the hypothalamus and anterior pituitary, respectively. Consequently, ACTH levels are elevated and stimulate the adrenal glands to produce greater than normal amounts of cortisol, adrenal androgens, and mineralocorticoids. In GGR, glucocorticoid resistance extends to peripheral tissues, although sensitivity to androgens and mineralocorticoids is normal. Hence, the clinical findings are not those of glucocorticoid excess, but rather those of mineralocorticoid excess (hypertension and hypokalemia) or hyperandrogenism. Treatment of Chrousos syndrome consists of suppressing the excess ACTH secretion through administration of mineralocorticoid-sparing glucocorticoids such as dexamethasone.

Loss of function mutations within the *GR* gene have been identified in a majority of individuals with GGR, with the LBD as the most common site of mutation.[161,162] While many LBD mutations in GGR reduce hormone-binding affinity, the transcriptional regulatory activity of GR can also be affected by LBD mutations that do not impact ligand binding.[163] Loss-of-function mutations have also been found in other GR domains in GGR individuals leading to reduced DNA binding, nuclear translocation, or expression.[161] In some individuals with GGR, *GR* gene mutations have not been identified, suggesting that other components of the glucocorticoid signaling pathway in GGR may be uniquely sensitive to specific gene mutations that do not affect other endocrine systems.[164,165]

Acquired glucocorticoid resistance is much more common than GGR and is often characterized by tissue-selective resistance of GR to endogenous or synthetic glucocorticoids.[166] This form of resistance is often associated with chronic inflammatory conditions,[167,168] corticotropinomas,[169,170] or hematological malignancies.[166,171] This resistance reduces the therapeutic benefit of glucocorticoids to treat leukemia, lymphoma, and myeloma, as well as many chronic inflammatory conditions. Although uncommon, somatic GR mutations have been identified that are responsible for malignant cell resistance to glucocorticoid therapy, although they are not associated with GGR, as expected from their somatic origin.[166] More common mechanisms responsible for glucocorticoid resistance in malignant cells are reductions in GR expression and the generation of alternative GR mRNA splice isoforms, such as GRβ discussed earlier. Recently, genome-wide analysis identified novel genes and pathways contributing to glucocorticoid resistance in acute lymphoblastic leukemia.[172] Such comprehensive genomic analysis of glucocorticoid resistance is likely to generate novel therapeutic strategies and drug combinations that could mitigate malignant cell resistance to glucocorticoid therapy.[172]

Most severe asthma patients do not respond well to glucocorticoid treatment and are designated as having glucocorticoid-unresponsive, refractory, or resistant asthma.[173-175] The molecular basis of glucocorticoid-resistant asthma has been extensively studied, revealing several molecular and cellular mechanisms that may contribute to the failure of this typically highly effective class of agents to control asthma symptoms and exacerbations.[168,173,175] Mutations or polymorphisms within the *GR* gene are unlikely to be a major contributor to glucocorticoid-resistant asthma, although some rare cases in specific ethnic populations may exist.[176] However, genome-wide association studies have identified a functional variant in a glucocorticoid-regulated gene (i.e., the glucocorticoid-induced transcript 1) that is associated with reduced lung function response to inhaled glucocorticoids.[177] Additional rare genetic variants that contribute to the development of glucocorticoid-resistant asthma may be uncovered with more extensive pharmacogenetic and pharmacogenomic analyses in distinct populations.[178]

As mentioned earlier for other forms of glucocorticoid resistance, an increase in the GRβ/GRα ratio plays a significant role in glucocorticoid-resistant asthma.[179-182] The induction of *GRβ* gene expression in airway epithelial cells that could generate glucocorticoid resistance is promoted by specific T-helper-type 17 cytokines.[183,184] Activation of the p38 mitogen-activated protein kinase (MAPK) by proinflammatory cytokines is one intracellular pathway that contributes to upregulation of GRβ expression and the development of glucocorticoid-resistant asthma.[185,186] The impact of proinflammatory cytokines and activated MAPK pathways on glucocorticoid-resistant asthma are not limited to the regulation of GRβ expression and extend to inhibitory effects on various components of the GRα intracellular signaling, including DNA binding,[187] nuclear translocation,[188] and interaction with transcription factor or coregulator partners.[187,189,190] Some of these effects may be mediated by direct posttranslational modification of GRα (e.g., phosphorylation[191]) or alterations in receptor or partner protein oxidation as a result of oxidative or nitrative stress.[192]

KEY POINTS

- Generalized glucocorticoid resistance affects all tissues and is most often caused by mutations in the glucocorticoid receptor (GR) ligand-binding domain, resulting in reduced hormone-binding affinity.
- Acquired glucocorticoid resistance is tissue-specific and is commonly caused by reductions in GR expression and/or increased expression of the dominant-negative splice variant GRβ.

SUMMARY AND FUTURE DIRECTIONS

Glucocorticoids act through GR to regulate numerous physiological processes, and synthetic derivatives of these hormones are widely prescribed for their antiinflammatory and immunosuppressive actions. Tissue responses to glucocorticoids are markedly different, and this heterogeneity in glucocorticoid signaling presents a challenge to clinicians in managing patients on glucocorticoid therapy. The diversity in glucocorticoid signaling is achieved by multiple processes, including polymorphisms in the *GR* gene, the generation of various GR isoforms by alternative splicing and alternative translation initiation, and posttranslational modifications of the receptor protein. The capacity of a cell to generate dozens of GR isoforms with unique gene-regulatory profiles provides enormous potential for variations in GR signaling. Moreover, the glucocorticoid response is impacted by the glucocorticoid secretion rhythm and by both general and acquired glucocorticoid resistance. A critical goal of future studies will be to use genetic approaches to assess the contribution that an individual GR isoform makes to the actions of glucocorticoids in the whole animal. Additionally, it will be important to determine whether changes in the tissue complement of GR isoforms underlie pathologies characterized by glucocorticoid resistance and/or the severe side effects that accompany glucocorticoid treatment. Finally, it will be crucial to determine the transcriptome for individual GR isoforms and to assess how it changes in response to receptor posttranslational modifications. A greater understanding of the factors and molecular processes that control tissue-specific responses to glucocorticoids should aid in the development of safer and more effective glucocorticoid therapies.

REFERENCES

1. Rhen T, Cidlowski JA. Antiinflammatory action of glucocorticoids--new mechanisms for old drugs. *N Engl J Med*. 2005;353:1711–1723.
2. Barnes PJ. Anti-inflammatory actions of glucocorticoids: molecular mechanisms. *Clin Sci (Lond)*. 1998;94:557–572.
3. Sapolsky RM, Romero LM, Munck AU. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr Rev*. 2000;21:55–89.
4. Miner JN, Hong MH, Negro-Vilar A. New and improved glucocorticoid receptor ligands. *Expert Opin Investig Drugs*. 2005;14:1527–1545.
5. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *The Journal of Allergy and Clinical Immunology*. 2013;132:1033–1044.
6. Grad I, Picard D. The glucocorticoid responses are shaped by molecular chaperones. *Mol Cell Endocrinol*. 2007;275:2–12.
7. Uhlenhaut NH, Barish GD, Yu RT, et al. Insights into negative regulation by the glucocorticoid receptor from genome-wide profiling of inflammatory cistromes. *Molecular Cell*. 2013;49:158–171.
8. Surjit M, Ganti KP, Mukherji A, et al. Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell*. 2011;145:224–241.
9. Hudson WH, Youn C, Ortlund EA. The structural basis of direct glucocorticoid-mediated transrepression. *Nature Structural & Molecular Biology*. 2013;20:53–58.
10. Burd CJ, Archer TK. Chromatin architecture defines the glucocorticoid response. *Mol Cell Endocrinol*. 2013;380:25–31.
11. Oakley RH, Revollo J, Cidlowski JA. Glucocorticoids regulate arrestin gene expression and redirect the signaling profile of G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109:17591–17596.
12. Ramamoorthy S, Cidlowski JA. Ligand-induced repression of the glucocorticoid receptor gene is mediated by an NCoR1 repression complex formed by long-range chromatin interactions with intragenic glucocorticoid response elements. *Molecular and Cellular Biology*. 2013;33:1711–1722.
13. John S, Sabo PJ, Thurman RE, et al. Chromatin accessibility predetermines glucocorticoid receptor binding patterns. *Nature Genetics*. 2011;43:264–268.
14. Reddy TE, Gertz J, Crawford GE, et al. The hypersensitive glucocorticoid response specifically regulates period 1 and expression of circadian genes. *Molecular and Cellular Biology*. 2012;32:3756–3767.
15. McNally JG, Muller WG, Walker D, et al. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*. 2000;287:1262–1265.
16. Jenkins BD, Pullen CB, Darimont BD. Novel glucocorticoid receptor coactivator effector mechanisms. *Trends Endocrinol Metab*. 2001;12:122–126.
17. Lonard DM, O'Malley BW. Expanding functional diversity of the coactivators. *Trends Biochem Sci*. 2005;30:126–132.
18. Rosenfeld MG, Glass CK. Coregulator codes of transcriptional regulation by nuclear receptors. *J Biol Chem*. 2001;276:36865–36868.
19. Meijsing SH, Pufall MA, So AY, et al. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science*. 2009;324:407–410.
20. Ronacher K, Hadley K, Avenant C, et al. Ligand-selective transactivation and transrepression via the glucocorticoid receptor: role of cofactor interaction. *Mol Cell Endocrinol*. 2009;299:219–231.
21. De Bosscher K, Vanden Berghe W, Haegeman G. The interplay between the glucocorticoid receptor and nuclear factor-kappaB or activator protein-1: molecular mechanisms for gene repression. *Endocr Rev*. 2003;24:488–522.
22. Caelles C, Gonzalez-Sancho JM, Munoz A. Nuclear hormone receptor antagonism with AP-1 by inhibition of the JNK pathway. *Genes Dev*. 1997;11:3351–3364.
23. Smoak K, Cidlowski JA. Glucocorticoids regulate tristetraprolin synthesis and posttranscriptionally regulate tumor necrosis factor alpha inflammatory signaling. *Molecular and Cellular Biology*. 2006;26:9126–9135.
24. Rogatsky I, Ivashkiv LB. Glucocorticoid modulation of cytokine signaling. *Tissue Antigens*. 2006;68:1–12.
25. Cruz-Topete D, He B, Xu X, et al. Kruppel-like factor 13 is a major mediator of glucocorticoid receptor signaling in cardiomyocytes and protects these cells from DNA damage and death. *J Biol Chem*. 2016;291:19374–19386.
26. Gross KL, Cidlowski JA. Tissue-specific glucocorticoid action: a family affair. *Trends Endocrinol Metab*. 2008;19:331–339.
27. Manenschijn L, van den Akker EL, Lamberts SW, et al. Clinical features associated with glucocorticoid receptor polymorphisms. An overview. *Ann N Y Acad Sci*. 2009;1179:179–198.
28. Jewell CM, Cidlowski JA. Molecular evidence for a link between the N363S glucocorticoid receptor polymorphism and altered gene expression. *The Journal of Clinical Endocrinology and Metabolism*. 2007;92:3268–3277.
29. Jewell CM, Katen KS, Barber LM, et al. Healthy glucocorticoid receptor N363S carriers dysregulate gene expression associated with metabolic syndrome. *Am J Physiol Endocrinol Metab*. 2016;311:E741–E748.
30. Russcher H, van Rossum EF, de Jong FH, et al. Increased expression of the glucocorticoid receptor-A translational isoform as a result of the ER22/23EK polymorphism. *Mol Endocrinol*. 2005;19:1687–1696.
31. Lewis-Tuffin LJ, Cidlowski JA. The physiology of human glucocorticoid receptor beta (hGRbeta) and glucocorticoid resistance. *Ann N Y Acad Sci*. 2006;1069:1–9.
32. Bamberger CM, Bamberger AM, de Castro M, et al. Glucocorticoid receptor beta, a potential endogenous inhibitor of glucocorticoid action in humans. *J Clin Invest*. 1995;95:2435–2441.
33. Oakley RH, Sar M, Cidlowski JA. The human glucocorticoid receptor beta isoform. Expression, biochemical properties, and putative function. *J Biol Chem*. 1996;271:9550–9559.
34. Jain A, Wordinger RJ, Yorio T, et al. Spliceosome protein (SRp) regulation of glucocorticoid receptor isoforms and glucocorticoid response in human trabecular meshwork cells. *Investigative Ophthalmology & Visual Science*. 2012;53:857–866.
35. Derijk RH, Schaaf MJ, Turner G, et al. A human glucocorticoid receptor gene variant that increases the stability of the glucocorticoid receptor beta-isoform mRNA is associated with rheumatoid arthritis. *J Rheumatol*. 2001;28:2383–2388.
36. Schaaf MJ, Cidlowski JA. AUUUA motifs in the 3'UTR of human glucocorticoid receptor alpha and beta mRNA destabilize mRNA and decrease receptor protein expression. *Steroids*. 2002;67:627–636.
37. Geelhoed JJ, van Duijn C, van Osch-Gevers L, et al. Glucocorticoid receptor-9beta polymorphism is associated with systolic blood pressure and heart growth during early childhood. The Generation R Study. *Early Human Development*. 2011;87:97–102.
38. Otte C, Wust S, Zhao S, et al. Glucocorticoid receptor gene, low-grade inflammation, and heart failure: the Heart and Soul study. *The Journal of Clinical Endocrinology and Metabolism*. 2010;95:2885–2891.
39. van den Akker EL, Koper JW, van Rossum EF, et al. Glucocorticoid receptor gene and risk of cardiovascular disease. *Arch Intern Med*. 2008;168:33–39.
40. Whirledge SD, Jewell CM, Barber LM, et al. Generating diversity in human glucocorticoid signaling through a racially diverse polymorphism in the beta isoform of the glucocorticoid receptor. *Lab Invest*. 2017;97:1282–1295.
41. Lewis-Tuffin LJ, Jewell CM, Bienstock RJ, et al. The human glucocorticoid receptor {beta} (hGR{beta}) binds RU-486 and is transcriptionally active. *Molecular and Cellular Biology*. 2007;27:2266–2282.
42. Min J, Perera L, Krahn JM, et al. Probing dominant negative behavior of glucocorticoid receptor beta through a hybrid structural and biochemical approach. *Molecular and Cellular Biology*. 2018;38:1–18.
43. He B, Cruz-Topete D, Oakley RH, et al. Human glucocorticoid receptor beta regulates gluconeogenesis and inflammation in mouse liver. *Molecular and Cellular Biology*. 2015;36:714–730.

44. Lu NZ, Cidlowski JA. Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes. *Molecular Cell*. 2005;18:331–342.
45. Yudt MR, Cidlowski JA. Molecular identification and characterization of a and b forms of the glucocorticoid receptor. *Mol Endocrinol*. 2001;15:1093–1103.
46. Lu NZ, Collins JB, Grissom SF, et al. Selective regulation of bone cell apoptosis by translational isoforms of the glucocorticoid receptor. *Molecular and Cellular Biology*. 2007;27:7143–7160.
47. Wu I, Shin SC, Cao Y, et al. Selective glucocorticoid receptor translational isoforms reveal glucocorticoid-induced apoptotic transcriptomes. *Cell Death & Disease*. 2013;4:e453.
48. Bender IK, Cao Y, Lu NZ. Determinants of the heightened activity of glucocorticoid receptor translational isoforms. *Mol Endocrinol*. 2013;27:1577–1587.
49. Gross KL, Oakley RH, Scoltock AB, et al. Glucocorticoid receptor {alpha} isoform-selective regulation of antiapoptotic genes in osteosarcoma cells: a new mechanism for glucocorticoid resistance. *Mol Endocrinol*. 2011;25:1087–1099.
50. Nehme A, Lobenhofer EK, Stamer WD, et al. Glucocorticoids with different chemical structures but similar glucocorticoid receptor potency regulate subsets of common and unique genes in human trabecular meshwork cells. *BMC Med Genomics*. 2009;2:58.
51. Cao Y, Bender IK, Konstantinidis AK, et al. Glucocorticoid receptor translational isoforms underlie maturational stage-specific glucocorticoid sensitivities of dendritic cells in mice and humans. *Blood*. 2013;121:1553–1562.
52. Sinclair D, Webster MJ, Wong J, et al. Dynamic molecular and anatomical changes in the glucocorticoid receptor in human cortical development. *Mol Psychiatry*. 2010;16:1–12.
53. Sinclair D, Tsai SY, Woon HG, et al. Abnormal glucocorticoid receptor mRNA and protein isoform expression in the prefrontal cortex in psychiatric illness. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2011;36:2698–2709.
54. Sinclair D, Webster MJ, Fullerton JM, et al. Glucocorticoid receptor mRNA and protein isoform alterations in the orbitofrontal cortex in schizophrenia and bipolar disorder. *BMC Psychiatry*. 2012;12:84.
55. Oakley RH, Ramamoorthy S, Foley JF, et al. Glucocorticoid receptor isoform-specific regulation of development, circadian rhythm, and inflammation in mice. *FASEB J*. 2018;32:5258–5271.
56. Galliher-Beckley AJ, Cidlowski JA. Emerging roles of glucocorticoid receptor phosphorylation in modulating glucocorticoid hormone action in health and disease. *IUBMB Life*. 2009;61:979–986.
57. Duma D, Cidlowski JA. Generating diversity in glucocorticoid receptor signaling: mechanisms, receptor isoforms, and post-translational modifications. *Horm Mol Biol Clin Investig*. 2010;3:319–328.
58. Ramamoorthy S, Cidlowski JA. Corticosteroids: mechanisms of action in health and disease. *Rheum Dis Clin North Am*. 2016;42:15–31,vii.
59. Jeanneteau F, Borie A, Chao MV, et al. Bridging the gap between brain-derived neurotrophic factor and glucocorticoid effects on brain networks. *Neuroendocrinology*. 2019;109:277–284.
60. Chen W, Dang T, Blind RD, et al. Glucocorticoid receptor phosphorylation differentially affects target gene expression. *Mol Endocrinol*. 2008;22:1754–1766.
61. Kino T. GR-regulating serine/threonine kinases: new physiologic and pathologic implications. *Trends Endocrinol Metab*. 2018;29:260–270.
62. Lambert WM, Xu CF, Neubert TA, et al. Brain-derived neurotrophic factor signaling rewrites the glucocorticoid transcriptome via glucocorticoid receptor phosphorylation. *Molecular and Cellular Biology*. 2013;33:3700–3714.
63. Kobayashi Y, Mercado N, Miller-Larsson A, et al. Increased corticosteroid sensitivity by a long acting beta2 agonist formoterol via beta2 adrenoceptor independent protein phosphatase 2A activation. *Pulm Pharmacol Ther*. 2012;25:201–207.
64. Rider CF, Altonsy MO, Mostafa MM, et al. Long-acting beta2-adrenoceptor agonists enhance glucocorticoid receptor (GR)-mediated transcription by gene-specific mechanisms rather than generic effects via GR. *Mol Pharmacol*. 2018;94:1031–1046.
65. Perez Kerkvliet C, Dwyer AR, Diep CH, et al. Glucocorticoid receptors are required effectors of TGFbeta1-induced p38 MAPK signaling to advanced cancer phenotypes in triple-negative breast cancer. *Breast Cancer Res*. 2020;22:39.
66. Arango-Lievano M, Borie AM, Dromard Y, et al. Persistence of learning-induced synapses depends on neurotrophic priming of glucocorticoid receptors. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116:13097–13106.
67. Kobayashi Y, Ito K, Kanda A, et al. Impaired dual-specificity protein phosphatase DUSP4 reduces corticosteroid sensitivity. *Mol Pharmacol*. 2017;91:475–481.
68. Panagiotou C, Lambadiari V, Maratou E, et al. Insufficient glucocorticoid receptor signaling and flattened salivary cortisol profile are associated with metabolic and inflammatory indices in type 2 diabetes. *J Endocrinol Invest*. 2021;44:37–48.
69. Garabedian MJ, Harris CA, Jeanneteau F. Glucocorticoid receptor action in metabolic and neuronal function. *F1000Res*. 2017;6:1208.
70. Guidotti G, Calabrese F, Anacker C, et al. Glucocorticoid receptor and FKBP5 expression is altered following exposure to chronic stress: modulation by antidepressant treatment. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2013;38:616–627.
71. Li H, Su P, Lai TK, et al. The glucocorticoid receptor-FKBP51 complex contributes to fear conditioning and posttraumatic stress disorder. *J Clin Invest*. 2020;130:877–889.
72. Ponce-Lina R, Serafin N, Carranza M, et al. Differential phosphorylation of the glucocorticoid receptor in hippocampal subregions induced by contextual fear conditioning training. *Front Behav Neurosci*. 2020;14:12.
73. Simic I, Maric NP, Mitic M, et al. Phosphorylation of leukocyte glucocorticoid receptor in patients with current episode of major depressive disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;40:281–285.
74. Simic I, Adzic M, Maric N, et al. A preliminary evaluation of leukocyte phospho-glucocorticoid receptor as a potential biomarker of depressogenic vulnerability in healthy adults. *Psychiatry Res*. 2013;209: 658–664.
75. Wallace AD, Cao Y, Chandramouleeswaran S, et al. Lysine 419 targets human glucocorticoid receptor for proteasomal degradation. *Steroids*. 2010;75:1016–1023.
76. Sengupta S, Wasylyk B. Ligand-dependent interaction of the glucocorticoid receptor with p53 enhances their degradation by Hdm2. *Genes Dev*. 2001;15:2367–2380.
77. Wang X, DeFranco DB. Alternative effects of the ubiquitin-proteasome pathway on glucocorticoid receptor down-regulation and transactivation are mediated by CHIP, an E3 ligase. *Mol Endocrinol*. 2005;19:1474–1482.
78. Verma S, Ismail A, Gao X, et al. The ubiquitin-conjugating enzyme UBCH7 acts as a coactivator for steroid hormone receptors. *Molecular and Cellular Biology*. 2004;24:8716–8726.
79. Petrillo MG, Oakley RH, Cidlowski JA. beta-Arrestin-1 inhibits glucocorticoid receptor turnover and alters glucocorticoid signaling. *J Biol Chem*. 2019;294:11225–11239.
80. Holmstrom SR, Chupreta S, So AY, et al. SUMO-mediated inhibition of glucocorticoid receptor synergistic activity depends on stable assembly at the promoter but not on DAXX. *Mol Endocrinol*. 2008;22:2061–2075.
81. Paakinaho V, Kaikkonen S, Makkonen H, et al. SUMOylation regulates the chromatin occupancy and anti-proliferative gene programs of glucocorticoid receptor. *Nucleic Acids Res*. 2014;42:1575–1592.
82. Hua G, Ganti KP, Chambon P. Glucocorticoid-induced tethered transrepression requires SUMOylation of GR and formation of a SUMO-SMRT/NCoR1-HDAC3 repressing complex. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113:E635–E643.
83. Paul-Clark MJ, Roviezzo F, Flower RJ, et al. Glucocorticoid receptor nitration leads to enhanced anti-inflammatory effects of novel steroid ligands. *J Immunol*. 2003;171:3245–3252.
84. Ito K, Yamamura S, Essilfie-Quaye S, et al. Histone deacetylase 2-mediated deacetylation of the glucocorticoid receptor enables NF-kappaB suppression. *J Exp Med*. 2006;203:7–13.
85. Nader N, Chrousos GP, Kino T. Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine cluster: potential physiological implications. *FASEB J*. 2009;23:1572–1583.

86. Whirledge S, DeFranco DB. Glucocorticoid signaling in health and disease: insights from tissue-specific GR knockout mice. *Endocrinology*. 2018;159:46–64.
87. Cain DW, Cidlowski JA. Immune regulation by glucocorticoids. *Nat Rev Immunol*. 2017;17:233–247.
88. Cruz-Topete D, Cidlowski JA. One hormone, two actions: anti- and pro-inflammatory effects of glucocorticoids. *Neuroimmunomodulation*. 2015;22:20–32.
89. Bhattacharyya S, Brown DE, Brewer JA, et al. Macrophage glucocorticoid receptors regulate Toll-like receptor 4-mediated inflammatory responses by selective inhibition of p38 MAP kinase. *Blood*. 2007;109:4313–4319.
90. Tuckermann JP, Kleiman A, Moriggl R, et al. Macrophages and neutrophils are the targets for immune suppression by glucocorticoids in contact allergy. *J Clin Invest*. 2007;117:1381–1390.
91. Diaz-Jimenez D, Petrillo MG, Busada JT, et al. Glucocorticoids mobilize macrophages by transcriptionally up-regulating the exopeptidase DPP4. *J Biol Chem*. 2020;295:3213–3227.
92. Baschant U, Tuckermann J. The role of the glucocorticoid receptor in inflammation and immunity. *J Steroid Biochem Mol Biol*. 2010;120:69–75.
93. Li CC, Munitic I, Mittelstadt PR, et al. Suppression of dendritic cell-derived IL-12 by endogenous glucocorticoids is protective in LPS-induced sepsis. *PLoS Biol*. 2015;13:e1002269.
94. Brewer JA, Khor B, Vogt SK, et al. T-cell glucocorticoid receptor is required to suppress COX-2-mediated lethal immune activation. *Nat Med*. 2003;9:1318–1322.
95. Baschant U, Frappart L, Rauchhaus U, et al. Glucocorticoid therapy of antigen-induced arthritis depends on the dimerized glucocorticoid receptor in T cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108:19317–19322.
96. Guo L, Zheng Z, Ai J, et al. Scavenger receptor BI and high-density lipoprotein regulate thymocyte apoptosis in sepsis. *Arterioscler Thromb Vasc Biol*. 2014;34:966–975.
97. Kugler DG, Mittelstadt PR, Ashwell JD, et al. CD4+ T cells are trigger and target of the glucocorticoid response that prevents lethal immunopathology in toxoplasma infection. *J Exp Med*. 2013;210:1919–1927.
98. Schweingruber N, Fischer HJ, Fischer L, et al. Chemokine-mediated redirection of T cells constitutes a critical mechanism of glucocorticoid therapy in autoimmune CNS responses. *Acta Neuropathol*. 2014;127:713–729.
99. Wust S, van den Brandt J, Tischner D, et al. Peripheral T cells are the therapeutic targets of glucocorticoids in experimental autoimmune encephalomyelitis. *J Immunol*. 2008;180:8434–8443.
100. Mittelstadt PR, Monteiro JP, Ashwell JD. Thymocyte responsiveness to endogenous glucocorticoids is required for immunological fitness. *J Clin Invest*. 2012;122:2384–2394.
101. Zen M, Canova M, Campana C, et al. The kaleidoscope of glucorticoid effects on immune system. *Autoimmun Rev*. 2011;10:305–310.
102. Cain DW, Bortner CD, Diaz-Jimenez D, et al. Murine glucocorticoid receptors orchestrate B cell migration selectively between bone marrow and blood. *J Immunol*. 2020;205:619–629.
103. Quinn MA, Cidlowski JA. Endogenous hepatic glucocorticoid receptor signaling coordinates sex-biased inflammatory gene expression. *FASEB J*. 2016;30:971–982.
104. Duma D, Collins JB, Chou JW, et al. Sexually dimorphic actions of glucocorticoids provide a link to inflammatory diseases with gender differences in prevalence. *Sci Signal*. 2010;3:ra74.
105. Oakley RH, Cidlowski JA. Glucocorticoid signaling in the heart: A cardiomyocyte perspective. *J Steroid Biochem Mol Biol*. 2015;153:27–34.
106. Sainte-Marie Y, Nguyen Dinh Cat A, Perrier R, et al. Conditional glucocorticoid receptor expression in the heart induces atrio-ventricular block. *Faseb J*. 2007;21:3133–3141.
107. Rog-Zielinska EA, Craig MA, Manning JR, et al. Glucocorticoids promote structural and functional maturation of foetal cardiomyocytes: a role for PGC-1alpha. *Cell Death and Differentiation*. 2014;22:1106–1116.
108. Rog-Zielinska EA, Thomson A, Kenyon CJ, et al. Glucocorticoid receptor is required for foetal heart maturation. *Human Molecular Genetics*. 2013;22:3269–3282.
109. Oakley RH, Cruz-Topete D, He B, et al. Cardiomyocyte glucocorticoid and mineralocorticoid receptors directly and antagonistically regulate heart disease in mice. *Sci Signal*. 2019;12:eaau9685.
110. Oakley RH, Ren R, Cruz-Topete D, et al. Essential role of stress hormone signaling in cardiomyocytes for the prevention of heart disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110:17035–17040.
111. Richardson RV, Batchen EJ, Thomson AJ, et al. Glucocorticoid receptor alters isovolumetric contraction and restrains cardiac fibrosis. *J Endocrinol*. 2017;232:437–450.
112. Cruz-Topete D, Oakley RH, Carroll NG, et al. Deletion of the cardiomyocyte glucocorticoid receptor leads to sexually dimorphic changes in cardiac gene expression and progression to heart failure. *J Am Heart Assoc*. 2019;8:e011012.
113. Reul JM, de Kloet ER. Two receptor systems for corticosterone in rat brain: microdistribution and differential occupation. *Endocrinology*. 1985;117:2505–2511.
114. de Kloet ER, Otte C, Kumsta R, et al. Stress and depression: a crucial role of the mineralocorticoid receptor. *J Neuroendocrinol*. 2016;28:1–12.
115. Mifsud KR, Reul J. Mineralocorticoid and glucocorticoid receptor-mediated control of genomic responses to stress in the brain. *Stress*. 2018;21:389–402.
116. Karst H, Berger S, Turiault M, et al. Mineralocorticoid receptors are indispensable for nongenomic modulation of hippocampal glutamate transmission by corticosterone. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:19204–19207.
117. de Kloet ER. From receptor balance to rational glucocorticoid therapy. *Endocrinology*. 2014;155:2754–2769.
118. McEwen BS, Akil H. Revisiting the stress concept: implications for affective disorders. *J Neurosci*. 2020;40:12–21.
119. Kellendonk C, Eiden S, Kretz O, et al. Inactivation of the GR in the nervous system affects energy accumulation. *Endocrinology*. 2002;143:2333–2340.
120. Shibata M, Banno R, Sugiyama M, et al. AgRP neuron-specific deletion of glucocorticoid receptor leads to increased energy expenditure and decreased body weight in female mice on a high-fat diet. *Endocrinology*. 2016;157:1457–1466.
121. Herman JP, McKlveen JM, Ghosal S, et al. Regulation of the hypothalamic-pituitary-adrenocortical stress response. *Compr Physiol*. 2016;6:603–621.
122. Di S, Malcher-Lopes R, Halmos KC, et al. Nongenomic glucocorticoid inhibition via endocannabinoid release in the hypothalamus: a fast feedback mechanism. *J Neurosci*. 2003;23:4850–4857.
123. Evanson NK, Tasker JG, Hill MN, et al. Fast feedback inhibition of the HPA axis by glucocorticoids is mediated by endocannabinoid signaling. *Endocrinology*. 2010;151:4811–4819.
124. Samarasinghe RA, Di Maio R, Volonte D, et al. Nongenomic glucocorticoid receptor action regulates gap junction intercellular communication and neural progenitor cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108:16657–16662.
125. Nahar J, Haam J, Chen C, et al. Rapid nongenomic glucocorticoid actions in male mouse hypothalamic neuroendocrine cells are dependent on the nuclear glucocorticoid receptor. *Endocrinology*. 2015;156:2831–2842.
126. Harris C, Weiss GL, Di S, Tasker JG. Cell signaling dependence of rapid glucocorticoid-induced endocannabinoid synthesis in hypothalamic neuroendocrine cells. *Neurobiol Stress*. 2019;10:100158.
127. Gray JD, Kogan JF, Marrocco J, et al. Genomic and epigenomic mechanisms of glucocorticoids in the brain. *Nat Rev Endocrinol*. 2017;13:661–673.
128. Popoli M, Yan Z, McEwen BS, et al. The stressed synapse: the impact of stress and glucocorticoids on glutamate transmission. *Nat Rev Neurosci*. 2011;13:22–37.
129. Diorio D, Viau V, Meaney MJ. The role of the medial prefrontal cortex (cingulate gyrus) in the regulation of hypothalamic-pituitary-adrenal responses to stress. *J Neurosci*. 1993;13:3839–3847.
130. Keller-Wood ME, Dallman MF. Corticosteroid inhibition of ACTH secretion. *Endocr Rev*. 1984;5:1–24.

131. Boyle MP, Kolber BJ, Vogt SK, et al. Forebrain glucocorticoid receptors modulate anxiety-associated locomotor activation and adrenal responsiveness. *J Neurosci*. 2006;26:1971–1978.
132. Solomon MB, Furay AR, Jones K, et al. Deletion of forebrain glucocorticoid receptors impairs neuroendocrine stress responses and induces depression-like behavior in males but not females. *Neuroscience*. 2012;203:135–143.
133. Exton JH. Regulation of gluconeogenesis by glucocorticoids. *Monogr Endocrinol*. 1979;12:535–546.
134. Kraus-Friedmann N. Hormonal regulation of hepatic gluconeogenesis. *Physiol Rev*. 1984;64:170–259.
135. Kuo T, McQueen A, Chen TC, et al. Regulation of Glucose Homeostasis by Glucocorticoids. *Adv Exp Med Biol*. 2015;872:99–126.
136. Patel R, Williams-Dautovich J, Cummins CL. Minireview: new molecular mediators of glucocorticoid receptor activity in metabolic tissues. *Mol Endocrinol*. 2014;28:999–1011.
137. Imai E, Stromstedt PE, Quinn PG, et al. Characterization of a complex glucocorticoid response unit in the phosphoenolpyruvate carboxykinase gene. *Molecular and Cellular Biology*. 1990;10:4712–4719.
138. Imai E, Miner JN, Mitchell JA, et al. Glucocorticoid receptor-cAMP response element-binding protein interaction and the response of the phosphoenolpyruvate carboxykinase gene to glucocorticoids. *J Biol Chem*. 1993;268:5353–5356.
139. Nakae J, Kitamura T, Silver DL, et al. The forkhead transcription factor Foxo1 (Fkhr) confers insulin sensitivity onto glucose-6-phosphatase expression. *J Clin Invest*. 2001;108:1359–1367.
140. Wolfrum C, Asilmaz E, Luca E, et al. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. *Nature*. 2004;432:1027–1032.
141. Puigserver P, Rhee J, Donovan J, et al. Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. *Nature*. 2003;423:550–555.
142. Grontved L, John S, Baek S, et al. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *EMBO J*. 2013;32:1568–1583.
143. Xiao Y, Kim M, Lazar MA. Nuclear receptors and transcriptional regulation in non-alcoholic fatty liver disease. *Mol Metab*. 2020;101119.
144. Goldstein I, Baek S, Presman DM, et al. Transcription factor assisted loading and enhancer dynamics dictate the hepatic fasting response. *Genome Res*. 2017;27:427–439.
145. Dolinsky VW, Douglas DN, Lehner R, et al. Regulation of the enzymes of hepatic microsomal triacylglycerol lipolysis and re-esterification by the glucocorticoid dexamethasone. *Biochem J*. 2004;378:967–974.
146. Patel R, Patel M, Tsai R, et al. LXRbeta is required for glucocorticoid-induced hyperglycemia and hepatosteatosis in mice. *J Clin Invest*. 2011;121:431–441.
147. Lemke U, Krones-Herzig A, Berriel Diaz M, et al. The glucocorticoid receptor controls hepatic dyslipidemia through Hes1. *Cell Metab*. 2008;8:212–223.
148. Revollo JR, Oakley RH, Lu NZ, et al. HES1 is a master regulator of glucocorticoid receptor-dependent gene expression. *Sci Signal*. 2013;6:ra103.
149. Lightman SL, Wiles CC, Atkinson HC, et al. The significance of glucocorticoid pulsatility. *Eur J Pharmacol*. 2008;583:255–262.
150. Lightman SL, Birnie MT, Conway-Campbell BL. Dynamics of ACTH and cortisol secretion and implications for disease. *Endocr Rev*. 2020;41:470–490.
151. Quagliarini F, Mir AA, Balazs K, et al. Cistromic reprogramming of the diurnal glucocorticoid hormone response by high-fat diet. *Molecular Cell*. 2019;76:531–545. e535.
152. Bahrami-Nejad Z, Zhao ML, Tholen S, et al. A transcriptional circuit filters oscillating circadian hormonal inputs to regulate fat cell differentiation. *Cell. Metab*. 2018;27:854–868. e858.
153. Minnetti M, Hasenmajer V, Pofi R, et al. Fixing the broken clock in adrenal disorders: focus on glucocorticoids and chronotherapy. *J Endocrinol*. 2020;246:R13–R31.
154. Venneri MA, Hasenmajer V, Fiore D, et al. Circadian rhythm of glucocorticoid administration entrains clock genes in immune cells: a DREAM trial ancillary study. *The Journal of Clinical Endocrinology and Metabolism*. 2018;103:2998–3009.
155. Kalafatakis K, Russell GM, Ferguson SG, et al. Glucocorticoid ultradian rhythmicity differentially regulates mood and resting state networks in the human brain: a randomised controlled clinical trial. *Psychoneuroendocrinology*. 2021;124:105096.
156. Astiz M, Heyde I, Fortmann MI, et al. The circadian phase of antenatal glucocorticoid treatment affects the risk of behavioral disorders. *Nat Commun*. 2020;11:3593.
157. Elbi C, Walker DA, Romero G, et al. Molecular chaperones function as steroid receptor nuclear mobility factors. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101:2876–2881.
158. Conway-Campbell BL, George CL, Pooley JR, et al. The HSP90 molecular chaperone cycle regulates cyclical transcriptional dynamics of the glucocorticoid receptor and its coregulatory molecules CBP/p300 during ultradian ligand treatment. *Mol Endocrinol*. 2011;25:944–954.
159. Stavreva DA, Wiench M, John S, et al. Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat Cell Biol*. 2009;11:1093–1102.
160. Chrousos GP, Vingerhoeds A, Brandon D, et al. Primary cortisol resistance in man. A glucocorticoid receptor-mediated disease. *J Clin Invest*. 1982;69:1261–1269.
161. Martins CS, de Castro M. Generalized and tissue specific glucocorticoid resistance. *Mol Cell Endocrinol*. 2021;530:111277.
162. Charmandari E, Kino T, Souvatzoglou E, et al. Natural glucocorticoid receptor mutants causing generalized glucocorticoid resistance: molecular genotype, genetic transmission, and clinical phenotype. *The Journal of Clinical Endocrinology and Metabolism*. 2004;89:1939–1949.
163. Charmandari E, Raji A, Kino T, et al. A novel point mutation in the ligand-binding domain (LBD) of the human glucocorticoid receptor (hGR) causing generalized glucocorticoid resistance: the importance of the C terminus of hGR LBD in conferring transactivational activity. *The Journal of Clinical Endocrinology and Metabolism*. 2005;90:3696–3705.
164. Huizenga NA, de Lange P, Koper JW, et al. Five patients with biochemical and/or clinical generalized glucocorticoid resistance without alterations in the glucocorticoid receptor gene. *The Journal of Clinical Endocrinology and Metabolism*. 2000;85:2076–2081.
165. Nicolaides NC, Roberts ML, Kino T, et al. A novel point mutation of the human glucocorticoid receptor gene causes primary generalized glucocorticoid resistance through impaired interaction with the LXXLL motif of the p160 coactivators: dissociation of the transactivating and transrepressive activities. *The Journal of Clinical Endocrinology and Metabolism*. 2014;99:E902–E907.
166. Gross KL, Lu NZ, Cidlowski JA. Molecular mechanisms regulating glucocorticoid sensitivity and resistance. *Mol Cell Endocrinol*. 2009;300:7–16.
167. Vandewalle J, Luypaert A, De Bosscher K, et al. Therapeutic mechanisms of glucocorticoids. *Trends Endocrinol Metab*. 2018;29:42–54.
168. Barnes PJ, Adcock IM. Glucocorticoid resistance in inflammatory diseases. *Lancet*. 2009;373:1905–1917.
169. Antonini SR, Latronico AC, Elias LL, et al. Glucocorticoid receptor gene polymorphisms in ACTH-secreting pituitary tumours. *Clin Endocrinol (Oxf)*. 2002;57:657–662.
170. Ciato D, Albani A. Molecular mechanisms of glucocorticoid resistance in corticotropinomas: new developments and drug targets. *Front Endocrinol (Lausanne)*. 2020;11:21.
171. Renner K, Ausserlechner MJ, Kofler R. A conceptual view on glucocorticoid-induced apoptosis, cell cycle arrest and glucocorticoid resistance in lymphoblastic leukemia. *Curr Mol Med*. 2003;3:707–717.
172. Autry RJ, Paugh SW, Carter R, et al. Integrative genomic analyses reveal mechanisms of glucocorticoid resistance in acute lymphoblastic leukemia. *Nat Cancer*. 2020;1:329–344.
173. Barnes PJ, Greening AP, Crompton GK. Glucocorticoid resistance in asthma. *Am J Respir Crit Care Med*. 1995;152:S125–S140.
174. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med*. 2012;18:716–725.
175. Henderson I, Caiazzo E, McSharry C, et al. Why do some asthma patients respond poorly to glucocorticoid therapy? *Pharmacol Res*. 2020;160:105189.
176. Zhao F, Zhou G, Ouyang H, et al. Association of the glucocorticoid receptor D641V variant with steroid-resistant asthma: a case-control study. *Pharmacogenet Genomics*. 2015;25:289–295.

177. Tantisira KG, Lasky-Su J, Harada M, et al. Genomewide association between GLCCI1 and response to glucocorticoid therapy in asthma. *N Engl J Med*. 2011;365:1173–1183.
178. Edris A, de Roos EW, McGeachie MJ, et al. Pharmacogenetics of inhaled corticosteroids and exacerbation risk in adults with asthma. *Clin Exp Allergy*. 2021 http://dx.doi.10.1111/cea.13829.
179. Hamid QA, Wenzel SE, Hauk PJ, et al. Increased glucocorticoid receptor beta in airway cells of glucocorticoid-insensitive asthma. *Am J Respir Crit Care Med*. 1999;159:1600–1604.
180. Leung DY, Hamid Q, Vottero A, et al. Association of glucocorticoid insensitivity with increased expression of glucocorticoid receptor beta. *J Exp Med*. 1997;186:1567–1574.
181. Sousa AR, Lane SJ, Cidlowski JA, et al. Glucocorticoid resistance in asthma is associated with elevated in vivo expression of the glucocorticoid receptor beta-isoform. *The Journal of Allergy and Clinical Immunology*. 2000;105:943–950.
182. Goleva E, Li LB, Eves PT, et al. Increased glucocorticoid receptor beta alters steroid response in glucocorticoid-insensitive asthma. *Am J Respir Crit Care Med*. 2006;173:607–616.
183. Al Heialy S, Gaudet M, Ramakrishnan RK, et al. Contribution of IL-17 in steroid hyporesponsiveness in obese asthmatics through dysregulation of glucocorticoid receptors alpha and beta. *Front Immunol*. 2020;11:1724.
184. Vazquez-Tello A, Semlali A, Chakir J, et al. Induction of glucocorticoid receptor-beta expression in epithelial cells of asthmatic airways by T-helper type 17 cytokines. *Clin Exp Allergy*. 2010;40:1312–1322.
185. Bhavsar P, Khorasani N, Hew M, et al. Effect of p38 MAPK inhibition on corticosteroid suppression of cytokine release in severe asthma. *Eur Respir J*. 2010;35:750–756.
186. Mercado N, To Y, Kobayashi Y, et al. p38 mitogen-activated protein kinase-gamma inhibition by long-acting beta2 adrenergic agonists reversed steroid insensitivity in severe asthma. *Mol Pharmacol*. 2011;80:1128–1135.
187. Adcock IM, Lane SJ, Brown CR, et al. Differences in binding of glucocorticoid receptor to DNA in steroid-resistant asthma. *J Immunol*. 1995;154:3500–3505.
188. Chang PJ, Michaeloudes C, Zhu J, et al. Impaired nuclear translocation of the glucocorticoid receptor in corticosteroid-insensitive airway smooth muscle in severe asthma. *Am J Respir Crit Care Med*. 2015;191:54–62.
189. Adcock IM, Ito K, Barnes PJ. Histone deacetylation: an important mechanism in inflammatory lung diseases. *COPD*. 2005;2:445–455.
190. Qiu W, Guo F, Glass K, et al. Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma. *The Journal of Allergy and Clinical Immunology*. 2018;141:1250–1258.
191. Irusen E, Matthews JG, Takahashi A, et al. p38 Mitogen-activated protein kinase-induced glucocorticoid receptor phosphorylation reduces its activity: role in steroid-insensitive asthma. *The Journal of Allergy and Clinical Immunology*. 2002;109:649–657.
192. Barnes PJ. Mechanisms and resistance in glucocorticoid control of inflammation. *J Steroid Biochem Mol Biol*. 2010;120:76–85.

90

Adrenal Androgens, Adrenarche, and Adrenopause

Selma Feldman Witchel, Anne Claire Burghard, and Sharon E. Oberfield

OUTLINE

Adrenarche, 1498
The Adrenal Glands, 1498
The Fetal Adrenal Gland, 1498
Postnatal Adrenal Morphology and Function, 1499
  Clinical Facets of Adrenarche, 1500
  Adrenal Steroidogenesis, 1501
Dehydroepiandrosterone and *in Vitro* Fertilization, 1502
  Adrenopause, 1503
  Epidemiology/Associations with Dehydroepiandrosterone Sulfate Concentrations, Aging, and Cardiovascular Risks, 1503
Neurobiology of Dehydroepiandrosterone and Dehydroepiandrosterone Sulfate, 1504
  Clinical Implications of Dehydroepiandrosterone Sulfate/Dehydroepiandrosterone, 1505
  Dehydroepiandrosterone Sulfate/Dehydroepiandrosterone and "The Fountain of Youth", 1505
Conclusion, 1505

ADRENARCHE

The morphologic and functional characteristics of the adrenal glands change dramatically through the different stages of prenatal and postnatal development. *Adrenarche* is defined as the increased production of dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEAS) by the zona reticularis of the adrenal glands between 6 and 8 years of age. However, with newer data indicating that urinary excretion of DHEA and its metabolites begins to increase at as early as 3 years of age, the age at onset of adrenarche is being reexamined.[1]

The phenomenon of adrenarche is observed in humans and some nonhuman primates. Current data suggest that the great apes and other nonhuman primates may also experience adrenarche. Yet, sustained postnatal increased DHEAS concentrations are limited to humans, chimpanzees, and bonobos.[2,3] It has been speculated that one function of adrenarche is to modify the neural, behavioral, and psychosocial development as part of the human transition from child to adult.[4] Clarifying the physiologic roles of the C19 adrenal steroid hormones, especially DHEAS, during adulthood is an ongoing process. Investigating the mechanisms regulating changes in DHEA and DHEAS production will improve understanding of zona reticularis function in health and disease.

KEY POINTS

- *Adrenarche* is defined as the rise in adrenal androgens that occurs in mid-childhood and is responsible for the appearance of pubic hair.
- The term "adrenal androgens" refers to 19-carbon steroids (C19 steroids) including dehydroepiandrosterone, dehydroepiandrosterone sulfate, and androstenedione. These steroids do not bind to the androgen receptor. Rather, these steroids are converted to more potent sex steroids.

THE ADRENAL GLANDS

The adrenal cortex synthesizes and secretes diverse C21 and C19 steroid hormones beginning *in utero* and ending with the demise of the individual. The Italian anatomist and physician Bartolomeo Eustachio was the first to describe and provide anatomic depictions of the adrenal glands, in 1543.[5] Approximately 300 years later, Thomas Addison described the phenotype associated with adrenal insufficiency.[6] Subsequently, Charles Edouard Brown-Séquard confirmed the importance of the adrenal glands; he demonstrated that bilateral adrenalectomy led to rapid death attributed to loss of vital hormones.[7] In 1950, the Nobel Prize in Physiology and Medicine was awarded to Edward Calvin Kendall, Tadeusz Reichstein, and Philip Showalter Hench for their discoveries of the adrenal cortical hormones and their biological effects.

THE FETAL ADRENAL GLAND

The adrenal glands consist of two separate components, the cortex and the medulla, which have distinct embryonic origins. In the human fetus, the cortex develops around 28 to 30 days postconception from cells in the intermediate mesoderm located between the coelomic epithelium and the dorsal aorta. These cells develop into the adrenogonadal primordium. Around 32 to 40 days postconception, the adrenogonadal primordium separates. Cells that migrate retroperitoneally to the cranial pole of the mesonephros develop into the adrenal primordium. Subsequently, neural crest derived cells invade the adrenal primordium; these cells develop into the catecholamine secreting chromaffin cells of the adrenal medulla.[8]

An outer mesenchymal capsule surrounds the entire adrenal gland. In the fetus, two distinct concentric zones, the outer definitive and inner fetal zones, can be identified. Beginning at approximately 8 weeks of gestation, the fetal adrenal is an active endocrine organ. The clear demarcation in adrenal weights for normal infants compared to those with anencephaly and congenital adrenal hyperplasia highlights the important role of adrenocorticotropic hormone (ACTH) as a tropic factor influencing the growth of the fetal adrenal gland (Fig. 90.1).[9]

Due to minimal CYP17A1 activity, the human placenta cannot synthesize estrogen *de novo*. Hence, one major activity of the human

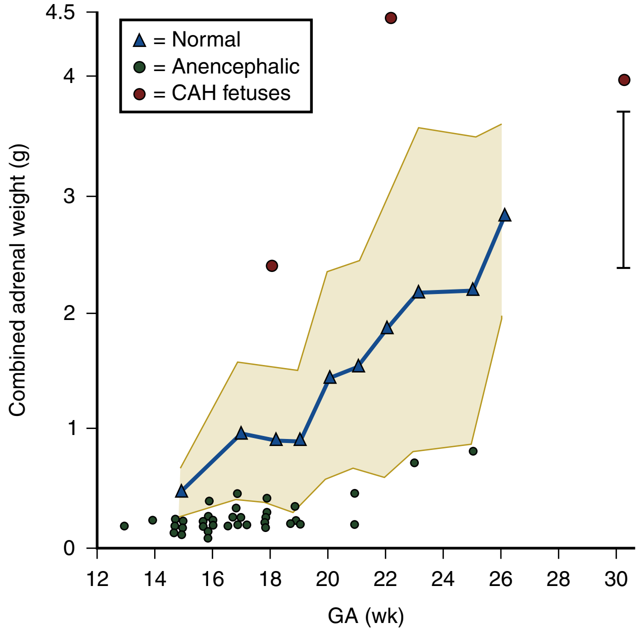


Figure 90.1 **Combined fetal adrenal weights according to gestational age.** *CAH*, Congenital adrenal hyperplasia; *GA*, gestational age. (Data from Young MC, Laurence KM, Hughes IA. Relationship between fetal adrenal morphology and anterior pituitary function, *Horm Res*. 1989;32:130–135.)

fetal adrenal is to provide substrates to the placenta, with its robust aromatase activity, for estrogen biosynthesis. During a brief period at around 8 to 12 weeks of gestation, the enzyme 3β hydroxysteroid dehydrogenase type 2 (HSD3B2) is expressed, leading to cortisol synthesis to presumably prevent virilization of female fetuses.[10] In concert with the morphologic changes in the fetal adrenal gland, DHEA, pregnenolone, 17OH-pregnenolone, and their respective sulfated conjugates are secreted. The steroid output amounts to approximately 200 mg/day, of which 60% is DHEAS.[11] As noted above, DHEAS is a major substrate for estrogen biosynthesis via the fetoplacental unit (Fig. 90.2). The fetal adrenal gland provides approximately 60% of the DHEAS for placental estradiol synthesis.[12] Of note, 16α-hydroxyl-DHEAS, metabolized by the fetal liver, provides the substrate for placental estriol synthesis.

Enzymes expressed in the fetal zone include steroidogenic acute regulatory protein (StAR), P450c11A1, P450c17A1, P450c21A2, P450c11B1/B2, and sulfotransferase family 2A member 1 (SULT2A1). Two enzymes, 3β-HSD2 and aldo-keto reductase family 1 member C3 (ARK1C3), are constitutively expressed throughout the first and second trimesters of gestation.[13] The outer definitive/adult zone is largely inactive during the first two trimesters of gestation and begins to produce cortisol during the third trimester.

After 24 to 28 weeks of gestation, the fetal adrenal directly secretes cortisol. Evaluation of second-trimester human adrenals showed the presence of cortisol. Intraadrenal steroid content, measured in second-trimester human fetal adrenal glands, showed that pregnenolone was the most abundant steroid; 17-OHP, progesterone, DHEAS, and cortisol were detectable, whereas aldosterone was undetectable.[14]

Cortisol promotes terminal differentiation of specific tissues such as lung, thyroid, and liver in anticipation of postnatal life.[15] Regarding the lung, cortisol promotes structural maturation of the lung and stimulates fetal lung surfactant synthesis.[16] At birth, the adrenals weigh 8 to 9 g, which is comparable to the weight of adult adrenals. Over the few months following birth, the cells in the fetal zone undergo apoptosis, leading to involution of the fetal adrenal zone.[17] The timing of this involution appears to be more closely tied to gestational age rather than the birth of the infant.[18]

POSTNATAL ADRENAL MORPHOLOGY AND FUNCTION

Postnatally, a precipitous fall in adrenal weight by approximately 50% occurs due to the involution of the fetal zone. The definitive zone persists and develops into the adult adrenal cortex. The adult adrenal cortex is organized into three functionally distinct concentric zones, which were initially described in 1866.[19] The names of the zones reflect the organization of the cells within each region. Cells of the outer zone, the zona glomerulosa, are organized into round clusters (Latin *glomus*, ball). Cells of the middle zone, the zona fasciculata, are arranged in radial pillars or cords (Latin *fascicle*, bundle). Cells of the innermost zone, zona reticularis, form a mesh-like structure (Latin *rete*, net).[20] Lineage-tracing

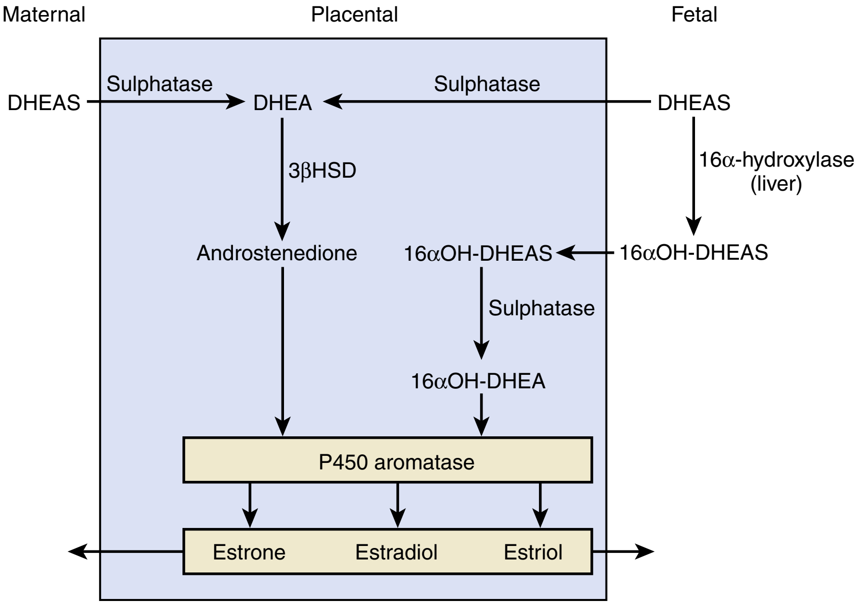


Figure 90.2 **The fetal-placental-maternal steroid unit.** *DHEA*, Dehydroepiandrosterone; *DHEAS*, dehydroepiandrosterone sulfate.

studies have confirmed that adrenocortical cells originate in the outer portion of the cortex and progressively move towards the medulla. These functional status of these cells changes during this process.[21]

The outer zone, the zona glomerulosa, secretes aldosterone. Zona glomerulosa function is regulated by the renin-angiotensin system and serum potassium concentrations. The enzyme aldosterone synthase encoded by the aldosterone synthase (*CYP11B2*) gene is only expressed in the zona glomerulosa.

The zona fasciculata secretes cortisol in response to ACTH acting through the ACTH receptor encoded by the melanocortin-2 receptor (*MC2R*) gene. ACTH has chronic and acute effects. Chronic or tonic ACTH stimulation is essential to promote transcription and translation of adrenal steroidogenic enzymes. ACTH acts acutely to stimulate cortisol secretion. Cortisol induces negative feedback inhibition at the hypothalamus and pituitary. Activity of the hypothalamic–pituitary–adrenal (HPA) axis is organized into circadian and ultradian rhythms, with the highest ACTH and cortisol secretion in the early morning hours prior to awakening.[22] Due to its lipophilic nature, cortisol must be synthesized *de novo* because it cannot be stored in vesicles.

The inner zone, the zona reticularis, secretes DHEA, DHEAS, and androstenedione. Despite being labeled as "adrenal androgens," these C19 steroid hormones do not activate the androgen receptor. Rather, these hormones can be considered to be "prohormones" that can be converted to androgens and estrogens in peripheral tissues. DHEAS, the steroid hormone circulating in the greatest abundance, shows no diurnal variation. The 7- to 10-hour half-life of DHEAS likely contributes to this lack of diurnal variation.[23] Extraadrenal conversion between DHEA and DHEAS through reversible sulfoconjugation by SULTA2 (adrenal and liver) and steroid sulfatase (peripheral tissues) activity. These reversible enzyme activities enable tissue-specific regulation of steroid hormone exposure, also known as intracrinology.[24,25]

The human adrenals are the primary source of the 11-oxygenated metabolites of testosterone and androstenedione, 11β-hydroxyandrostenedione (11OHA4) and 11β-hydroxytestosterone (11OHT). In peripheral tissues, these hormones are converted to 11-ketoandrostenedione (11KA4), 11-ketotestosterone (11KT), and 11-ketodihydrotestosterone (11KDHT). Collectively, these 19-carbon steroids containing an oxygen atom on carbon 11 are referred to as the 11-oxyandrogens (11oAs). The androgenic potencies of 11KT and 11KDHT are comparable to those of testosterone and dihydrotestosterone, respectively.[26-28]

KEY POINTS

- Postnatally, the fetal zone involutes leaving the definitive zone to develop into the three distinct zones of the adult adrenal cortex: the zona glomerulosa, the zona fasciculata, and the zona reticularis, which secrete aldosterone, cortisol, and C19 steroid hormones, respectively.

Clinical Facets of Adrenarche

Adrenarche represents the peripubertal maturation of the adrenal cortex. It has been arbitrarily defined as the rise in adrenal androgens, DHEA, DHEAS, and androstenedione that occurs between 6 and 8 years of age. Despite this commonly used definition, the maturation of the adrenal cortex may truly be a more gradual process; urinary steroid markers of adrenarche show a continuous rise starting at 3 to 4 years of age.[1] This concept is supported by a report regarding a small group of Finnish children in whom DHEAS concentrations were measurable at 1 year of age and correlated with DHEAS concentrations at 6 years of age.[29] Despite much investigation, the factors initiating the onset of adrenarche and subsequently regulating adrenal C19 steroid secretion remain unknown. Studies involving monozygotic and dizygotic twins indicate heritability for onset of adrenarche to be 58% to 61%.[30,31] Although small case-control studies assessed for genetic markers associated with age at adrenarche, no distinct genetic markers were consistently identified and validated.[32,33]

Pubarche is the clinical manifestation of adrenarche. Pubarche is characterized by the development of pubic or axillary hair, adult apocrine odor, acne, and increased oiliness of the skin and hair. In girls, the pubic hair growth typically starts on the labia. Among boys, both adrenal and testicular C19 steroids promote pubic and axillary hair development. The most recent large cross-sectional study, the Third National Health and Nutrition Study conducted between 1988 and 1994, reported mean ages for pubic hair development among American youth. For girls, the mean ages for pubic hair development were 9.5 years for non-Hispanic Blacks, 10.3 years for Mexican Americans, and 10.5 years for non-Hispanic Whites. For boys, mean ages for pubic hair development were 11.1 years for non-Hispanic Blacks, 12.3 years for Mexican Americans, and 12 years for non-Hispanic Whites.[34]

Premature pubarche is arbitrarily defined as occurring when pubic hair growth appears before 8 years of age in girls and before 9 years of age in boys. The most common cause of premature pubarche is premature adrenarche. Because premature adrenarche is a diagnosis of exclusion, other disorders associated with androgen excess need to be excluded. The differential diagnosis includes congenital adrenal hyperplasia, androgen-secreting tumors, cortisone reductase deficiency, apparent cortisone reductase deficiency, 3'-phospho-adenosine 5'-phosphosulfate synthase deficiency, glucocorticoid resistance, Cushing syndrome, and exposure to exogenous androgens (Table 90.1).

Premature adrenarche is more common in girls than in boys.[35] Children with premature adrenarche tend to be taller. Skeletal maturation is normal or minimally advanced. Skeletal maturation is assessed by an x-ray of the left hand and wrist to evaluate the ossification centers.[36] When the skeletal maturation is significantly advanced, greater than two standard deviations (SDs), concerns arise regarding the potential for adult short stature. Despite conflicting outcome studies, most children achieve an adult height appropriate for their predicted midparental height.[37-40] The mini growth spurt in normal children that also occurs at around 6 to 8 years of age is probably not a function of adrenarche.[41] Importantly, the finding of a significantly advanced bone age (>2 SDs) in a child with premature adrenarche should prompt evaluation for other disorders associated with premature pubarche, such as nonclassic congenital adrenal hyperplasia and androgen-secreting tumors.[42] Because all children have an increase in adrenal androgen production between 6 and 8 years of age, why hair development occurs in a small minority of children remains puzzling. In addition to DHEA, DHEAS, and androstenedione, the 11-oxygenated C19 steroid adrenal androgens play a role in the clinical signs of adrenarche.

TABLE 90.1 Differential Diagnosis of Premature Pubarche

Premature adrenarche
Congenital adrenal hyperplasia
Androgen-secreting tumor (adrenal, ovary, or testis)
Human chorionic gonadotropin-secreting tumor
Apparent cortisone reductase deficiency
Cortisone reductase deficiency
Apparent dehydroepiandrosterone sulfate sulfotransferase deficiency
Cushing syndrome
Familial glucocorticoid resistance
Exposure to exogenous androgens

KEY POINTS

- The factors governing onset of adrenarche remain unclear.
- The diagnosis of premature adrenarche can be made by the presence of pubic hair before the age of 8 years in girls and 9 years in boys.
- Premature adrenarche should be differentiated from more serious conditions such as congenital adrenal hyperplasia and virilizing tumors.

Associations reported between low birth weight and premature adrenarche have been inconsistent.[43] Sexual dimorphism is not apparent in normal children, in whom a relationship between lower birth weight and higher adrenal androgen levels is present in both sexes.[44] Androgens and estrogens affect skeletal maturation and bone health through their actions on osteoblast, osteocytes, and osteoclasts.[45] In a small study of girls with premature adrenarche, increased linear growth was already evident in the first 2 years of life and associated with higher IGF-1 levels.[46] Obesity can further accelerate the rate of skeletal maturation in premature adrenarche.

Levels of 11oAs have been measured in several disorders, including castration-resistant prostate cancer,[47] 21-hydroxylase deficiency,[48-50] and polycystic ovarian syndrome (PCOS).[26,51] A recent study showed that 11oAs correlate with surrogate markers of metabolic risk in a large cohort of women with PCOS.[51] The 11oAs also appear to significantly contribute to the clinical signs observed in adrenarche,[26,27,47] including premature adrenarche.[52] Recent findings suggest that elevated 11oAs—rather than DHEA, DHEAS, and A4—may be of particular diagnostic and clinical significance in cases of clinical signs of adrenarche (premature pubarche) without the biochemical diagnosis of premature adrenarche based on DHEAS levels (for premature adrenarche definition, see "Clinical Facets of Adrenarche" later).[53] Wise-Oringer et al. have proposed that measurement of serum 11oAs may be a more accurate way to screen and classify children with premature adrenarche than measurement of serum DHEAS.[53] These findings could potentially also be broadly applied to other disease states of androgen excess.

Obesity, hyperinsulinemia, and/or insulin resistance have been observed in children with premature adrenarche, and the molecular basis of these findings is likely multifactorial. Obesity exacerbates insulin resistance and hyperinsulinemia, which can promote increased adrenal and gonadal steroid secretion. Adiposity may influence the association between elevated adrenal androgens and signs of metabolic syndrome. Children evaluated for premature adrenarche should be clinically followed for signs of metabolic syndrome, although these relationships remain to be fully elucidated.[54-59] These studies report associations without establishing causality.

The question regarding the risk for girls with premature adrenarche to develop PCOS in later life was initially based on findings in a cohort of Catalan girls. Of these girls who had been previously diagnosed with premature pubarche, 45% developed functional ovarian hyperandrogenism.[60] Additional features of dyslipidemia, increased visceral fat, hyperinsulinemia, inflammatory marker changes, and early menarche were also identified in this cohort.[61] Daughters of mothers with PCOS are at higher risk for developing this syndrome, and a significant number exhibit exaggerated adrenarche.[62] Studies investigating the relationship between premature adrenarche and PCOS have led to a proposed developmental sequence beginning with reduced fetal growth, followed by a postnatal catch-up in height and weight, premature adrenarche, and finally clinical androgen excess and PCOS, with hyperinsulinemia and increased hepatovisceral fat stores playing a key role.[63-66]

Still, among retrospective studies, the development of PCOS in girls with premature adrenarche varies.[67-69] Not all girls with premature adrenarche develop PCOS, and conversely, not all girls with PCOS have a history of premature adrenarche.[70] Inquiry into the relationship of premature adrenarche to PCOS can be a challenge, because the diagnostic criteria for PCOS include normal features for peripubertal development during adolescence, such as irregular menses, polycystic ovary morphology, and mild hyperandrogenism.[71] Anti-Müllerian hormone (AMH) concentrations are typically elevated in women with PCOS, and AMH has been assessed as a possible tool to predict PCOS in girls with premature adrenarche. Whereas one study reported that AMH concentrations were significantly higher in girls with premature adrenarche and correlated with maternal history of PCOS,[72] another study suggests that AMH may not have adequate predictive value for the diagnosis of PCOS.[73] Currently, AMH should not be used to predict and diagnose PCOS.

Currently, no blinded randomized controlled trials exist to provide evidence-based data regarding the use of insulin sensitizers in a child presenting with typical clinical and biochemical features of premature adrenarche. It is appropriate to counsel parents that some girls with premature adrenarche may later experience menstrual disturbances in adolescence. Early intervention after menarche with insulin sensitizers may prevent progression to PCOS and reduce the risk for long-term cardiovascular problems.[74] Counseling regarding healthy lifestyle choices is also important.

KEY POINTS

- Associations between premature adrenarche and bone age advancement, obesity, hyperinsulinemia, insulin resistance, and polycystic ovary syndrome have been described.

Adrenal Steroidogenesis

The initial step in steroidogenesis is the conversion of cholesterol to pregnenolone. The P450c17 enzyme functions as the qualitative regulator of steroidogenesis.[75] This enzyme uniquely catalyzes both 17α-hydroxylase and 17,20-lyase activities. It is the latter enzyme activity that preferentially ensures C19 steroid production and, in particular, DHEA synthesis and adrenarche. A number of factors enhance the differential activities of P450c17. These include posttranslational regulation of P450c17 by phosphorylation on serine/threonine residues and the role of electron transfer proteins.[76-78] The principal electron donor to microsomal P450 enzymes, including P450c17, is nicotinamide adenine dinucleotide phosphate cytochrome P450 oxidoreductase. This enzyme cofactor augments both 17α-hydroxylase and 17,20-lyase activities but is not the principal determinant of increased DHEA synthesis. Enhancement of 17,20-lyase activity is linked specifically to the action of cytochrome b_5 in the presence of adequate P450 oxidoreductase.[75] This effect is far more substantial with Δ^5 substrates such as 17OH-pregnenolone as compared with a Δ^4 substrate such as 17OH-progesterone, hence the predominance of DHEA synthesis. Coupled to the promotion of DHEA synthesis is a relative deficiency of 3β-hydroxysteroid dehydrogenase (HSDB2) in the fetal adrenal and postnatal zona reticularis. This enzyme would normally compete with P450c17 to convert Δ^5 to Δ^4 steroids.[79]

Mutations in the gene encoding P450 oxidoreductase lead to a biochemical profile reminiscent of combined partial 17α-hydroxylase and 21-hydroxylase deficiencies with virilization in affected girls and undervirilization in affected boys.[80,81] Antley–Bixler syndrome, a condition characterized by severe skeletal abnormalities, genital anomalies, and adrenal dysfunction, is also a manifestation of this enzyme deficiency.[82]

More than 99% of DHEA is sulfated to DHEAS through the action of dehydroepiandrosterone sulfatase (SULT2A1).[83] Sulfated steroids are unavailable as substrates for HSDB2 activity, thus maintaining the high levels of DHEAS. Immunohistochemical

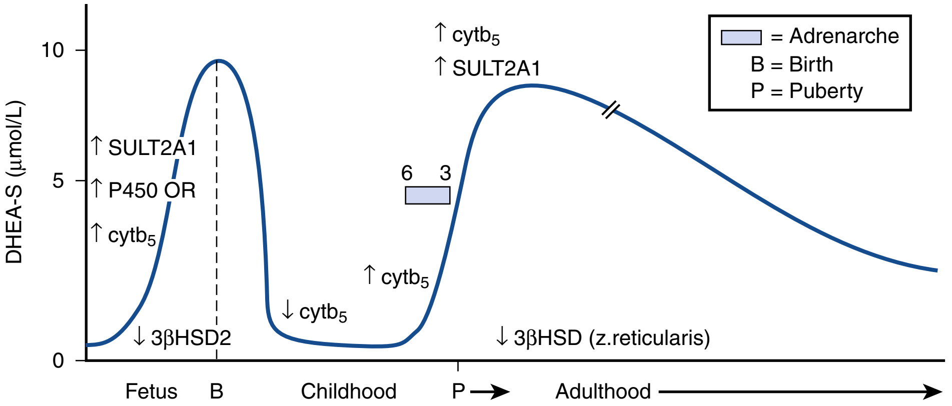


Figure 90.3 Schematic of dehydroepiandrosterone sulfate (DHEAS) production from fetal to adult life. The *arrows* indicate temporal changes in factors controlling DHEAS.

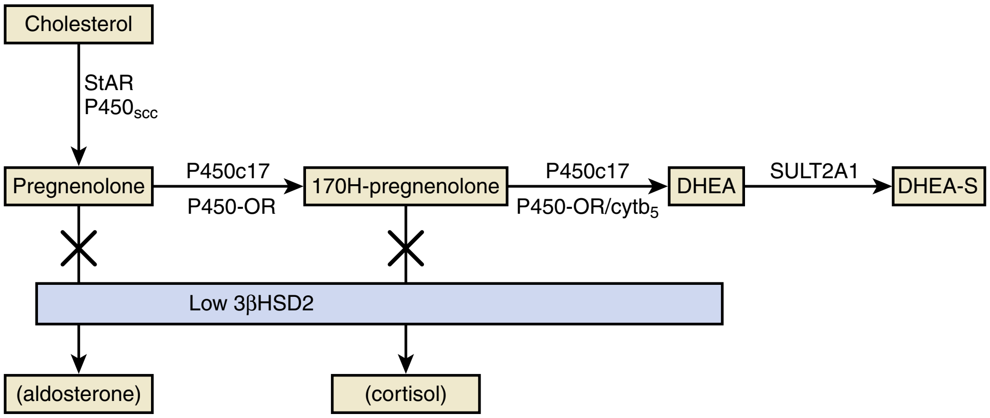


Figure 90.4 Pathway of dehydroepiandrosterone sulfate (DHEAS) synthesis in fetal adrenal and adult zona reticularis. Also depicted is the effect of low expression of 3β-hydroxysteroid dehydrogenase 2 (3β-HSD2) in these tissues.

studies in the fetal adrenal gland and in age-related postnatal adrenal glands have demonstrated a developmental pattern of enzyme and cofactor expression that may explain the intraadrenal control of adrenarche.[84-86] In summary, the pattern of increased DHEAS secretion is associated with increased expression of P450c17, P450 oxidoreductase, cytochrome b_5, SULT2A1, and decreased expression of HSDB2 in the fetal adrenal gland and again in the postnatal zona reticularis after 5 years of age. In contrast, a reversal of this pattern occurs in the postnatal adrenal gland between infancy and 5 years of age. Expression of cytochrome b_5 and SULT2A1 is maintained beyond the adult age, when DHEAS levels start to decline. The zona reticularis becomes smaller with increasing age, suggesting that the decline in DHEAS levels is associated with a decrease in cell number rather than a change in enzyme content. What factors stimulate the changes in expression of key proteins controlling DHEAS synthesis remain to be determined. HSDB2 has the important function of determining the ratio of Δ^5 and Δ^4 steroid production. A family of orphan nuclear receptors (nerve growth factor-induced clone B) is expressed in parallel with HSDB2 and directly acts on the promoter to upregulate the *HSD3B2* gene (Figs. 90.3 and 90.4).[87]

DEHYDROEPIANDROSTERONE AND *IN VITRO* FERTILIZATION

DHEA has been used as a supplemental treatment for women undergoing *in vitro* fertilization, particularly in patients who respond poorly to gonadotropin stimulation. Regarding the definition of poor ovarian response (POR), The European Society of Human Reproduction and Embryology has recommended that at least two of the following three criteria should be present in identifying a patient as a potential poor responder to ovarian stimulation: (i) advanced maternal age or any other risk factor for POR; (ii) a previous POR; and (iii) an abnormal ovarian reserve test.[88] However, defining POR is still unclear, as randomized trials on POR have often used differing definitions.[89]

The use of DHEA supplementation for POR was first suggested in 2000 in a case series.[90] A worldwide survey found that 25.8% of *in vitro* fertilization (IVF) clinicians in 45 different countries use DHEA to supplement IVF treatment protocols in women with POR.[91] Studies have shown that DHEA treatment may increase the number of oocytes, increase embryo numbers and quality, increase pregnancy rate, increase live birth rate, and decrease miscarriage rate.[92-96] In a systematic review and metaanalysis of six randomized controlled trials (RCTs) of DHEA supplementation in women undergoing IVF (five studies in women with POR, one in women without POR), the pooled estimates suggested that DHEA supplementation was associated with a significant increase in clinical pregnancy, endometrial thickness, and retrieved oocytes. No relationship was identified to an improvement in live birth rate, embryos transferred, or an increase of estradiol on human chorionic gonadotropin day.[97] Additionally, several studies have found no significant differences in the parameters mentioned; one RCT, for example, failed to show that DHEA supplementation significantly improved IVF outcomes.[98] Another study found that, in women undergoing DHEA supplementation, the markers of very early (AMH) and early differentiated (inhibin B) follicles were not different compared with controls, whereas antral follicle count increased significantly, suggesting that DHEA specifically targets the 2- to 10-mm antral cohort of follicles that have escaped atresia for that particular cycle, improving follicular survival.[99]

The presumed major mechanism of DHEA in improving reproductive outcomes is the increase in androgen levels, because DHEA serves as a precursor to estrogen and testosterone. Androgen receptors are expressed in the granulosa cells, theca cells, and oocytes as follicles mature.[100] In this situation, DHEA may act through intracrine mechanisms, allowing for tissue-specific modifications compared with androgens such as testosterone. This may explain the more precise actions and fewer adverse effects of DHEA supplementation compared with testosterone supplementation.[24] Studies have suggested that DHEA could have mitochondrial effects, increasing mitochondrial mass, reducing mitochondrial apoptosis, and restoring mitochondrial homeostasis in cumulus cells and granulosa cells.[101,102] The exact mechanism of DHEA supplementation in augmenting IVF in poor responders is therefore likely multifaceted and remains to be fully understood.[103] More studies, particularly randomized controlled trials, are needed to establish a clearer picture of the mechanisms of DHEA on improving reproductive outcomes in women undergoing IVF and the efficacy and safety of this intervention.

KEY POINTS

- Dehydroepiandrosterone has been used as supplementation for women undergoing *in vitro* fertilization, possibly working through intracrine mechanisms as a precursor to estrogen and testosterone. More randomized controlled trials are needed to determine the efficacy and safety of this intervention.

Adrenopause

The term “adrenopause” refers the observation that DHEA and DHEAS concentrations decline following peak DHEA and DHEAS during the mid-20s. DHEA and DHEAS concentrations at 80 years of age are only 20% of those at age 25 years.[104] Despite decreasing DHEA and DHEAS concentrations, glucocorticoid and mineralocorticoid secretion are largely unchanged. In a longitudinal study of women (Study of Women’s Health Across the Nation), a modest perimenopausal increase was noted in DHEAS concentrations; this was accompanied by stable testosterone concentrations and minimal decreases in androstenedione concentrations.[105] During the perimenopausal period, DHEAS concentrations were lower in Black women than in White women.[106]

In a study involving 100 premenopausal and 100 postmenopausal women, sulfated steroids, i.e., DHEAS, 17-hydroxypregnenolone, and androstenediol-3-sulfate, showed a negative correlation with age.[107] Yet, despite the decline in DHEA and DHEAS concentrations, the 11oAs did not decline among aging women. Among these women, no significant associations between steroid concentrations, body mass index (BMI), hypertension, dyslipidemia, hyperglycemia, or bone density were identified. With expansion of this dataset to include more women and men, 11OHA4 and 11OHT concentrations were noted to slightly increase in women while remaining stable in men across adult ages. In this population, 11KT concentrations showed positive correlation with 11KA4, 11OHT, and 11OHA4 in both sexes.[108] After adjusting for age, androstenedione concentrations declined and were inversely related to BMI in both sexes. Of the 11oAs, 11OHT increased with BMI in women, whereas 11KT increased with higher BMI in men.[108] As would be anticipated, testosterone concentrations were higher in men compared with women; 11OHA4, 11KA4, and 11KT concentrations were minimally higher among men.[108] The factors responsible for the decreased DHEA/DHEAS production and the dissociation of DHEA/DHEAS production from the 11-oxygenated C19 steroids are indeterminate.

This decline in DHEA and DHEAS concentrations is accompanied by immunohistochemical changes in the adrenal cortex of older adults. Specifically, using human adrenal tissues obtained from reproductive-aged adults and older adults, expression of both CYB5A2 and SULT2A1 in the adrenal glands from older individuals was reduced. Similar to the findings in adrenal glands obtained from older women, adrenals from older men demonstrated a less distinct zona fasciculata/zona reticularis zonation and more commonly displayed areas where cells expressing HSD3B2 and CYB5A intermingled.[108] The molecular basis for these changes in zona reticularis function is unknown. The small sample size, cross-sectional nature, and incomplete clinical information limit conclusions from these studies.

KEY POINTS

- Adrenopause refers to the fall in adrenal production of dehydroepiandrosterone and dehydroepiandrosterone sulfate that occurs in later life.

Epidemiology/Associations with Dehydroepiandrosterone Sulfate Concentrations, Aging, and Cardiovascular Risks

Cross-sectional epidemiologic studies have queried the relationship between declining DHEAS concentrations and various aspects of aging. A 1986 study showed DHEAS concentrations to be inversely related to death from any cause and to death from cardiovascular disease in men over 50 years of age.[109] These findings were not replicated in adult women in the Rancho Bernando Study.[110] The Osteoporotic Fractures in Men study in Sweden was a 5-year longitudinal study that involved 2416 men between 69 to 81 years of age; this study reported that low DHEA and DHEAS concentrations predicted an increased risk for coronary heart disease but not cerebrovascular disease.[111] Results from the Women’s Ischemia Syndrome Evaluation reported a greater than 2-fold increased risk of cardiovascular disease mortality among women in the lowest tertile of DHEAS compared to those in higher categories (hazard ratio = 2.55; 95% confidence interval: 1.19–5.45).[112] A metaanalysis involving 18 studies with a total of 92,489 patients concluded that the prognosis was poorer for patients with cardiovascular disease and lower DHEAS levels compared with those with higher DHEAS concentrations.[113] Hence, current data suggest that lower DHEAS concentrations are associated with an increased risk for cardiovascular disease.

The Atherosclerosis Risk in Communities (ARIC) study assessed some features of the cardiovascular disease. In the ARIC study, lower DHEAS concentrations were associated with increased risk for subclinical myocardial injury, heart failure hospitalization, and death in men and women.[114] White participants tended to have lower DHEAS concentrations compared with Black participants. Within this cohort, men who experienced a greater decline in DHEAS concentrations showed an increased risk for hospitalization for heart failure. Their data suggest a threshold effect in which low DHEAS concentrations are associated with increased risk, but higher concentrations do not modify the risk for heart failure and death.

Among postmenopausal women, DHEAS concentrations were weakly but significantly related to endothelial function independent of other coronary risk factors, suggesting a protective effect of DHEA on the endothelium with improved endothelial cell function.[115] To elucidate the basis for the association of low DHEAS concentrations with myocardial injury, Jia and colleagues utilized a proteomics assay and Ingenuity Pathway Analysis. Significant associations between DHEAS concentrations and proteins with known functions pathways involved in the extracellular matrix and cardiac remodeling were identified. The association of DHEAS concentrations with POMC and Agouti-related protein, which are melanocortin receptor agonists and antagonists, respectively, suggests that DHEAS concentrations may indicate HPA axis dysfunction in older adults.[114] Low DHEAS concentrations appear to predict risk for heart failure in older adults.[114] However, whether DHEAS has a direct, indirect, or passive bystander role in myocardial remodeling remains to be determined.

DHEAS concentrations appear to be associated with risk for ischemic strokes. In the Nurses' Health Study, lower DHEAS concentrations were associated with increased risk for ischemic strokes.[116] Potential mechanisms for DHEAS actions could include inhibiting migration and proliferation of cells within the vascular wall, thereby increasing vascular smooth muscle apoptosis and decreasing vascular remodeling following injury.[117] In addition, animal models suggest that DHEAS may influence insulin resistance by affecting hepatocyte glucose production and improving peripheral tissue glucose clearance.[118,119] Nevertheless, the question as to whether DHEAS plays an active role in the metabolic pathways or is merely a marker remains to be answered.

KEY POINTS

- Adrenopause is accompanied by immunohistochemical changes in the adrenal cortex but is not accompanied by changes in glucocorticoid, mineralocorticoid, or 11-oxyandrogen production.
- Declining dehydroepiandrosterone sulfate concentrations may be associated with various aspects of aging, including increased risk for cardiovascular disease and ischemic strokes.

NEUROBIOLOGY OF DEHYDROEPIANDROSTERONE AND DEHYDROEPIANDROSTERONE SULFATE

DHEA and DHEAS concentrations are higher in the brain than in peripheral circulation.[120] One likely explanation for this finding is that both DHEA and DHEAS are synthesized *de novo* in the brain in astrocytes and neurons. For this reason, DHEA and DHEAS are considered to be "neurosteroids." Another source of brain DHEA is peripheral DHEA, which can readily cross blood–brain barrier due to its lipophilic nature. In contrast, brain DHEAS is more likely synthesized in the brain because its hydrophilic nature interferes with crossing the blood–brain barrier.[121]

Several mechanisms are likely responsible for DHEA/DHEAS actions. The most obvious mechanism is conversion of DHEA/DHEAS to more potent sex steroids, followed by activation of androgen or estrogen receptors in specific target tissues.[24] In addition, DHEA/DHEAS can bind to the tropomyosin-related kinase (TrkA), γ-aminobutyric acid type A ($GABA_A$), sigma subtype 1, and NMDA receptors.[122] DHEA/DHEAS appear to have inhibitory actions at the $GABA_A$ receptor, whereas they act as agonists at the sigma 1 receptor.[123] DHEA has been described as a modulator of neurotransmission because it can affect serotonin, γ-amino butyric acid (GABA), glutamate, and dopamine levels. In the central nervous system, it can be metabolized to ADIOL, which can bind to ERβ, negatively regulating proinflammatory gene expression.[124]

Neurobiological actions of DHEA/DHEAS include neuroprotection, neurite growth, neurogenesis, and neuronal survival. At low nanomolar concentrations, in primary cultures of mouse embryonic neocortical neurons, DHEA promoted growth of Tau-immunopositive axons; in contrast, DHEAS had no effect on axonal growth and stimulated dendritic growth.[123] In human fetal cortical neural stem cells, DHEA promoted neurogenesis and neuronal survival; this effect was blocked by an NMDA receptor antagonist and sigma 1 receptor antagonists.[125] In addition, DHEA has antiglucocorticoid and neuroprotective effects. Animal studies, both *in vivo* and *in vitro*, have demonstrated that DHEA protects against the neurotoxic effects of corticosterone.[126] Neurotrophins and their tyrosine kinase receptors influence neuronal differentiation, growth, and survival. DHEA binds to TrkA and TrkC receptors and induces receptor phosphorylation. The functional significance of this finding in human biology is unclear.[127] In a mouse model involving lipopolysaccharide stimulation, DHEA appeared to act as an endogenous regulator of neuroinflammatory microglial responses. The proposed mechanism involves interactions with the tropomyosin-related kinase receptor (TrkA), which activated the Akt1/Akts2-CREB pathway, ultimately resulting in enhanced expression of the histone 3 lysine 27 (H3K27) demethylase Jumonji d3 (Jmjd3).[128]

Preclinical findings have led to clinical investigations assessing DHEA/DHEAS concentrations and potential for therapeutic interventions. For example, researchers in traumatic stress have been intrigued by the potential relationship between post-traumatic stress disorder (PTSD) and DHEA/DHEAS. Elevated DHEA and low cortisol/DHEA or low cortisol/DHEAS concentrations have been described in some studies. A metaanalysis reviewing a modest number of publications found no consistent relationship between PTSD and DHEA/DHEAS concentrations. The heterogeneity of the results and use of only a single hormone determination without regard to time of day/diurnal rhythm confound interpretation of these studies. Nevertheless, there appeared to be an association between increased basal DHEA/DHEAS concentrations with a trauma exposure irrespective of PTSD development.[129]

Measuring hormone samples is hair samples provides a mechanism to circumvent the methodological issues associated with using a single hormone determination to characterize a dynamic hormone. In a study involving self-reports on psychopathology and coping in 92 Palestinian girls aged 11 to 16 years affected by the Israeli–Palestinian conflict, trauma-exposed girls had significantly higher DHEA and lower cortisol/DHEA ratios in their hair than non-trauma-exposed girls.[130]

DHEA has been found to have effect on immune and endothelial function. Available data are accumulating indicating that the cortisol-to-DHEAS ratio may be a useful indicator of chronic stress in humans. Indeed, this ratio may provide an indicator of chronic stress in non-human animal species and may be useful for animal welfare studies.[131]

In animal models for multiple sclerosis, DHEA and DHEAS may be beneficial. DHEA suppresses Th17 cell responses, reduces microglia-mediated neuroinflammation, and expands IL-10-producing regulatory T-cell populations in an ERβ-dependent manner. DHEAS decreases demyelination and axonal loss.[124]

KEY POINTS

- Dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEAS) are synthesized *de novo* in the brain, and their actions are regulated by several mechanisms.
- DHEA/DHEAS can contribute to neuroprotection, neurite growth, neurogenesis, and neuron survival and have been shown to be increased in traumatic stress.

Clinical Implications of Dehydroepiandrosterone Sulfate/Dehydroepiandrosterone

Data were obtained in the English Longitudinal Study of Aging using a community-based population to assess for an association between DHEAS concentrations and depression; data were obtained at baseline and at 4 years. Although no associations were apparent between DHEAS concentrations and depression at baseline, low DHEAS concentrations obtained at baseline predicted future depressive symptoms independent of age, gender, economic circumstances, and health behaviors and characteristics.[132] DHEAS concentrations fall after acute injury/trauma. Whereas cortisol concentrations typically increase during septic shock, DHEAS concentrations fall.[133] Among patients with schizophrenia, increased serum DHEA and lower cortisol/DHEA ratios were reported and associated with an inverse relationship between hippocampal volume and cortisol/DHEA ratio.[134]

Dehydroepiandrosterone Sulfate/Dehydroepiandrosterone and "The Fountain of Youth"

The observation that DHEA/DHEAS concentrations declined concomitant with onset of aging led to the hypothesis that provision of supplemental DHEA/DHEAS could restore physical health and protect the brain from the degenerative changes due to age. This hypothesis was supported by association studies linking decreased DHEA and DHEAS concentrations to decreased bone mineral density, increased risk of fracture, decrease cognitive function, decreased quality of life, and increased risk of cardiovascular disease.[135] Based on this hypothesis, DHEA/DHEAS replacement therapy was tried in older adults.

In the study by Nair et al., administering DHEA for 2 years had no effect on body composition, physical performance, or insulin sensitivity.[136] No effects were observed on insulin secretion, insulin action, and suppression of systemic lipolysis following mixed meal or glucose tolerance tests.[137,138] Most clinical studies of DHEAS administration in Alzheimer disease enrolled more females than males; no improvement in cognition was seen with DHEAS treatment.[139]

No consistent data supporting a benefit of DHEA supplementation on body composition, cognition, mood, or sexual function in older men.[140] RCTs of DHEA supplementation have yielded inconsistent results regarding clinical benefits. Improved arterial elasticity was observed among a heterogeneous group of elderly men and women treated with DHEA (50 mg/day) for 1 year.[141] One potential clinical indication for DHEA treatment is vaginal dryness; daily intravaginal 0.05% DHEA improved vaginal dryness and decreased dyspareunia.[142]

Hence, no evidence-based data exist substantiating that DHEA and DHEAS are the proverbial "fountain of youth."[143] However, the US Food and Drug Administration classifies these steroids as dietary supplements rather than drugs (DHEA Uses, Benefits, and Side Effects; verywellhealth.com). This means that distribution of these medications is unregulated and does not require a prescription. Hence, DHEA is available for personal use; these preparations have been found characterized to contain 0–150% of the amount stated on the package.[144]

KEY POINTS

- There is limited evidence that supplemental dehydroepiandrosterone can improve quality of life.

CONCLUSION

Despite being the most abundant circulating hormones in the circulation, the precise physiologic actions of DHEA and DHEAS remain mysteries. DHEA and DHEAS are secreted by the adrenal cortex and synthesized *de novo* in the brain. These hormones play a major role as substrates for placental estrogen synthesis in the fetus. DHEAS production escalates at adrenarche, peaking in early adulthood. Yet, the factors influencing adrenocortical DHEA/DHEAS secretion remain indeterminate. Low DHEAS concentrations appear to be associated with an increased risk for cardiovascular disease in older adults. Investigation into DHEA/DHEAS actions are uncovering effects on endothelial cells, immune function, and neuronal development. Much, however, remains to be elucidated regarding the physiologic roles and regulation of DHEA and DHEAS.

REFERENCES

1. Remer T, Boye KR, Hartmann MF, et al. Urinary markers of adrenarche: reference values in healthy subjects, aged 3-18 years. *J Clin Endocrinol Metab*. 2005;90:2015–2021.
2. Sabbi KH, Muller MN, Machanda ZP, et al. Human-like adrenal development in wild chimpanzees: a longitudinal study of urinary dehydroepiandrosterone-sulfate and cortisol. *Am J Primatol*. 2020;82:e23064.
3. Behringer V, Hohmann G, Stevens JM, et al. Adrenarche in bonobos (Pan paniscus): evidence from ontogenetic changes in urinary dehydroepiandrosterone-sulfate levels. *J Endocrinol*. 2012;214:55–65.
4. Campbell BC. Adrenarche and middle childhood. *Hum Nat*. 2011;22:327–349.
5. Eustacio B. *Opuscula Anatomicae*. 1543.
6. Addison T. On the constitutional and local effects of disease of the supra-renal capsules. *Br Foreign Med Chir Rev*. 1856;18:404–413.
7. Brown-Séquard C-E. Recherches expérimentales sur la physiologie, et al Pathologie des capsules surrénales. 1856.
8. Lumb R, Schwarz Q. Sympathoadrenal neural crest cells: the known, unknown and forgotten? *Dev Growth Differ*. 2015;57:146–157.
9. Young MC, Laurence KM, Hughes IA. Relationship between fetal adrenal morphology and anterior pituitary function. *Horm Res*. 1989;32:130–135.
10. Goto M, Piper Hanley K, Marcos J, et al. In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development. *J Clin Invest*. 2006;116:953–960.
11. Rehman KS, Carr BR, Rainey WE. Profiling the steroidogenic pathway in human fetal and adult adrenals. *J Soc Gynecol Investig*. 2003;10:372–380.
12. Kaludjerovic J, Ward WE. The interplay between estrogen and fetal adrenal cortex. *J Nutr Metab*. 2012;2012:837901.
13. Melau C, Nielsen JE, Frederiksen H, et al. Characterization of human adrenal steroidogenesis during fetal development. *J Clin Endocrinol Metab*. 2019;104:1802–1812.
14. Johnston ZC, Bellingham M, Filis P, et al. The human fetal adrenal produces cortisol but no detectable aldosterone throughout the second trimester. *BMC Med*. 2018;16:23.
15. Fowden AL, Li J, Forhead AJ. Glucocorticoids and the preparation for life after birth: are there long-term consequences of the life insurance? *Proc Nutr Soc*. 1998;57:113–122.
16. Garbrecht MR, Klein JM, Schmidt TJ, et al. Glucocorticoid metabolism in the human fetal lung: implications for lung development and the pulmonary surfactant system. *Biol Neonate*. 2006;89:109–119.
17. Ishimoto H, Jaffe RB. Development and function of the human fetal adrenal cortex: a key component in the feto-placental unit. *Endocr Rev*. 2011;32:317–355.
18. Heckmann M, Hartmann MF, Kampschulte B, et al. Persistent high activity of the fetal adrenal cortex in preterm infants: is there a clinical significance? *J Pediatr Endocrinol Metab*. 2006;19:1303–1312.
19. Arnold J. Ein Beitrag zu der feineren Structur und dem Chemismus der Nebennieren. Virchows Archiv Pathologie. *Anatomie Physiologie*. 1866;39:64–117.
20. Hammer GD, Basham KJ. Stem cell function and plasticity in the normal physiology of the adrenal cortex. *Mol Cell Endocrinol*. 2021;519:111043.
21. Freedman BD, Kempna PB, Carlone DL, et al. Adrenocortical zonation results from lineage conversion of differentiated zona glomerulosa cells. *Dev Cell*. 2013;26:666–673.

22. Lightman SL, Birnie MT, Conway-Campbell BL. Dynamics of ACTH and cortisol secretion and implications for disease. *Endocr Rev*. 2020;41: 470–490.
23. Bird CE, Murphy J, Boroomand K, et al. Dehydroepiandrosterone: kinetics of metabolism in normal men and women. *J Clin Endocrinol Metab*. 1978;47:818–822.
24. Labrie F, Martel C, Bélanger A, et al. Androgens in women are essentially made from DHEA in each peripheral tissue according to intracrinology. *J Steroid Biochem Mol Biol*. 2017;168:9–18.
25. Schiffer L, Arlt W, Storbeck KH. Intracrine androgen biosynthesis, metabolism and action revisited. *Mol Cell Endocrinol*. 2018;465:4–26.
26. Pretorius E, Arlt W, Storbeck KH. A new dawn for androgens: novel lessons from 11-oxygenated C19 steroids. *Mol Cell Endocrinol*. 2017;441:76–85.
27. Rege J, Nakamura Y, Satoh F, et al. Liquid chromatography-tandem mass spectrometry analysis of human adrenal vein 19-carbon steroids before and after ACTH stimulation. *J Clin Endocrinol Metab*. 2013;98:1182–1188.
28. Turcu AF, Rege J, Auchus RJ, et al. 11-Oxygenated androgens in health and disease. *Nat Rev Endocrinol*. 2020;16:284–296.
29. Liimatta J, Jääskeläinen J, Karvonen AM, et al. Tracking of serum DHEAS concentrations from age 1 to 6 years: a prospective cohort study. *J Endocr Soc*. 2020;4:bvaa012.
30. Pratt JH, Manatunga AK, Li W. Familial influences on the adrenal androgen excretion rate during the adrenarche. *Metabolism*. 1994;43:186–189.
31. Li H, Ji C, Yang L, et al. Heritability of serum dehydroepiandrosterone sulphate levels and pubertal development in 6~18-year-old girls: a twin study. *Ann Hum Biol*. 2017;44:325–331.
32. Utriainen P, Laakso S, Liimatta J, et al. Premature adrenarche—a common condition with variable presentation. *Horm Res Paediatr*. 2015;83:221–231.
33. Roldan MB, White C, Witchel SF. Association of the GAA1013-->GAG polymorphism of the insulin-like growth factor-1 receptor (IGF1R) gene with premature pubarche. *Fertil Steril*. 2007;88:410–417.
34. Sun SS, Schubert CM, Chumlea WC, et al. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics*. 2002;110:911–919.
35. Oberfield SE, Tao RH, Witchel SF. Present knowledge on the etiology and treatment of adrenarche. *Pediatr Endocrinol Rev*. 2018;15:244–254.
36. Creo AL, Schwenk 2nd WF. Bone age: a handy tool for pediatric providers. *Pediatrics*. 2017;140:e20171486.
37. Gurnurkar S, Arheart KL, Messiah SE, et al. Skeletal maturation and predicted adult height in children with premature adrenarche. *J Pediatr Endocrinol Metab*. 2014;27:69–74.
38. Diaz A, Bhandari S, Sison C, et al. Characteristics of children with premature pubarche in the New York Metropolitan Area. *Horm Res*. 2008;70:150–154.
39. DeSalvo DJ, Mehra R, Vaidyanathan P, et al. In children with PA, bone age advancement by 2 or more years in common and generally benign. *J Pediatr Endocrinol Metab*. 2013;26:215–221.
40. Oron T, Lebenthal Y, de Vries L, et al. Interrelationship of extent of precocious adrenarche in appropriate for gestational age girls with clinical outcome. *J Pediatr*. 2012;160:308–313.
41. Remer T, Manz F. The midgrowth spurt in healthy children is not caused by adrenarche. *J Clin Endocrinol Metab*. 2001;86:4183–4186.
42. Bizzarri C, Crea F, Marini R, et al. Clinical features suggestive of non-classical 21-hydroxylase deficiency in children presenting with precocious pubarche. *J Pediatr Endocrinol Metab*. 2012;25:1059–1064.
43. Charkaluk ML, Trivin C, Brauner R. Premature pubarche as an indicator of how body weight influences the onset of adrenarche. *Eur J Pediatr*. 2004;163:89–93.
44. Ong KK, Potau N, Petry CJ, et al. Opposing influences of prenatal and postnatal weight gain on adrenarche in normal boys and girls. *J Clin Endocrinol Metab*. 2004;89:2647–2651.
45. Almeida M, Laurent MR, Dubois V, et al. Estrogens and androgens in skeletal physiology and pathophysiology. *Physiol Rev*. 2017;97:135–187.
46. Utriainen P, Voutilainen R, Jääskeläinen J. Girls with premature adrenarche have accelerated early childhood growth. *J Pediatr*. 2009;154: 882–887.
47. du Toit T, Bloem LM, Quanson JL, et al. Profiling adrenal 11beta-hydroxyandrostenedione metabolites in prostate cancer cells, tissue and plasma: UPC(2)-MS/MS quantification of 11beta-hydroxytestosterone, 11keto-testosterone and 11keto-dihydrotestosterone. *J Steroid Biochem Mol Biol*. 2017;166:54–67.
48. Turcu AF, Nanba AT, Chomic R, et al. Adrenal-derived 11-oxygenated 19-carbon steroids are the dominant androgens in classic 21-hydroxylase deficiency. *Eur J Endocrinol*. 2016;174:601–609.
49. Kamrath C, Wettstaedt L, Boettcher C, et al. Androgen excess is due to elevated 11-oxygenated androgens in treated children with congenital adrenal hyperplasia. *J Steroid Biochem Mol Biol*. 2018;178:221–228.
50. Kamrath C, Wettstaedt L, Hartmann MF, et al. Height velocity defined metabolic control in children with congenital adrenal hyperplasia using urinary GC-MS analysis. *J Clin Endocrinol Metab*. 2019;jc.2019–20100438.
51. O'Reilly MW, Kempegowda P, Jenkinson C, et al. 11-oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2017;102:840–848.
52. Rege J, Turcu AF, Kasa-Vubu JZ, et al. 11-ketotestosterone is the dominant circulating bioactive androgen during normal and premature adrenarche. *J Clin Endocrinol Metab*. 2018;103:4589–4598.
53. Wise-Oringer BK, Burghard AC, O'Day P, et al. The unique role of 11-oxygenated C19 steroids in both premature adrenarche and premature pubarche. *Horm Res Paediatr*. 2020;93:460–469.
54. Kaya G, Yavas Abali Z, Bas F, et al. Body mass index at the presentation of premature adrenarche is associated with components of metabolic syndrome at puberty. *Eur J Pediatr*. 2018;177:1593–1601.
55. Liimatta J, Utriainen P, Laitinen T, et al. Cardiometabolic risk profile among young adult females with a history of premature adrenarche. *J Endocr Soc*. 2019;3:1771–1783.
56. Williams KM, Oberfield SE, Zhang C, et al. The relationship of metabolic syndrome and body composition in children with premature adrenarche: is it age related? *Horm Res Paediatr*. 2015;84:401–407.
57. Silfen ME, Manibo AM, Ferin M, et al. Elevated free IGF-1 levels in prepubertal Hispanic girls with PA: relationship with hyperandrogenism and insulin sensitivity. *J Clin Endocrinol Metab*. 2002;87:398–403.
58. Ibanez L, Potau N, Francois I, et al. Precocious pubarche, hyperinsulinism, and ovarian hyperandrogenism in girls: relation to reduced fetal growth. *J Clin Endocrinol*. 1998;83:3558–3562.
59. Mantyselka A, Jaaskelainen J, Lindi V, et al. The presentation of adrenarche is sexually dimorphic and modified by body adiposity. *J Clin Endocrinol Metab*. 2014;99:3889–3894.
60. Ibañez L, Potau N, Virdis R, et al. Postpubertal outcome in girls diagnosed of premature pubarche during childhood: increased frequency of functional ovarian hyperandrogenism. *J Clin Endocrinol Metab*. 1993;76:1599–1603.
61. Ibáñez L, Díaz R, López-Bermejo A, et al. Clinical spectrum of premature pubarche: links to metabolic syndrome and ovarian hyperandrogenism. *Rev Endocr Metab Disord*. 2009;10:63–76.
62. Maliqueo M, Sir Petermann T, Perez V, et al. Adrenal function during childhood and puberty in daughters of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2009;94:1923–1930.
63. Ibáñez L, Potau N, Francois I, et al. Precocious pubarche, hyperinsulinism, and ovarian hyperandrogenism in girls: relation to reduced fetal growth. *J Clin Endocrinol Metab*. 1998;83:3558–3562.
64. Ibáñez L, Ong K, Dunger DB, et al. Early development of adiposity and insulin resistance after catch-up weight gain in small-for-gestational-age children. *J Clin Endocrinol Metab*. 2006;91:2153–2158.
65. de Zegher F, López-Bermejo A, Ibáñez L. Central obesity, faster maturation, and 'PCOS' in girls. *Trends Endocrinol Metab*. 2018;29:815–818.
66. Ibáñez L, Del Río L, Díaz M, et al. Normalizing ovulation rate by preferential reduction of hepato-visceral fat in adolescent girls with polycystic ovary syndrome. *J Adolesc Health*. 2017;61:446–453.
67. Meas T, Chevenne D, Thibaud E, et al. Endocrine consequences of premature pubarche in post-pubertal Caucasian girls. *Clin Endocrinol (Ofx)*. 2002;57:101–106.
68. Ribeiro FA, Resende EAMR, Silva APD, et al. Metabolic and hormonal assessment of adolescent and young adult women with prior premature adrenarche. *Clinics (Sao Paulo)*. 2019;74:e836.

69. Livadas S, Bothou C, Kanaka-Gantenbein C, et al. Unfavorable hormonal and psychologic profile in adult women with a history of premature adrenarche and pubarche, compared to women with polycystic ovary syndrome. *Horm Metab Res*. 2020;52:179–185.
70. Witchel SF, Pinto B, Burghard AC, et al. Update on adrenarche. *Curr Opin Pediatr*. 2020;32:574–581.
71. Peña AS, Witchel SF, Hoeger KM, et al. Adolescent polycystic ovary syndrome according to the international evidence-based guideline. *BMC Med*. 2020;18:72.
72. Efthymiadou A, Bogiatzidou M, Kritikou D, et al. Anti-Mullerian hormone in girls with premature adrenarche: the impact of polycystic ovary syndrome history in their mothers. *J Pediatr*. 2019;205:190–194.
73. Brar PC, Dingle E, Ovadia D, et al. Interpretation of androgen and anti-Mullerian hormone profiles in a Hispanic cohort of 5- to 8-year-old girls with premature adrenarche. *Ann Pediatr Endocrinol Metab*. 2018;23:210–214.
74. Ibanez L, Ferrer A, Ong K, et al. Insulin sensitisation early after menarche prevents progression from precocious pubarche to polycystic ovary syndrome. *J Pediatr*. 2004;144:4–6.
75. Miller WL. Androgen biosynthesis from cholesterol to DHEA. *Mol Cell Endocrinol*. 2002;198:7–14.
76. Zhang L, Rodriguez H, Ohno S, et al. Serine phosphorylation of human P450c17 increases 17,20 lyase activity: implications for adrenarche and for the polycystic ovary syndrome. *Proc Natl Acad Sci USA*. 1995;92:10619–10623.
77. Lin D, Black SM, Nagahama Y, et al. Steroid 17a-hydroxylase and 17,20-lyase activities of P450c17: contributions of serine 106 and P450 reductase. *Endocrinology*. 1993;132:2496–2506.
78. Lee-Robichaud P, Wright JN, Akhtar ME, et al. Modulation of the activity of human 17a-hydroxylase-17,20lyase (CYP17) by cytochrome b5: endocrinological and mechanistic implications. *Biochem J*. 1995;308:901–908.
79. Conley AJ, Birm IM. The role of cytochrome P450 17α-hydroxylase and 3β-hydroxysteroid dehydrogenase in the integration of gonadal and adrenal steroidogenesis via the Δ5 and Δ4 pathway of steroidogenesis in mammals. *Biol Reprod*. 1997;56:789–799.
80. Arlt W, Walker EA, Draper N, et al. Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. *Lancet*. 2004;363:2128–2135.
81. Miller WL. P450 oxidoreductase deficiency: a new disorder of steroidogenesis with multiple clinical manifestations. *Trends Endocrinol Metab*. 2004;15:311–315.
82. Flück CE, Tajima T, Pandey AV, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat Genet*. 2004;36:228–230.
83. Strott CA. Steroid sulfotransferases. *Endocr Rev*. 1996;17:670–697.
84. Suzuki T, Sasano H, Takeyama J, et al. Developmental changes in steroidogenic enzymes in human postnatal adrenal cortex: immunohistochemical studies. *Clin Endocrinol*. 2000;53:739–747.
85. Rainey WE, Carr BR, Sasano H, et al. Dissecting human adrenal androgen production. *Trends Endocrinol Metab*. 2002;13:234–239.
86. Auchus RJ, Rainey WE. Adrenarche-physiology, biochemistry and human disease. *Clin Endocrinol*. 2004;60:288–296.
87. Bassett MH, Suzuki T, Sasano H, et al. The orphan nuclear receptor NGF1B regulates transcription of 3 β-hydroxysteroid dehydrogenase: implications for the control of adrenal functional zonation. *J Biol Chem*. 2004;279:37622–37630.
88. Ferraretti AP, La Marca A, Fauser BC, et al. ESHRE working group on poor ovarian response definition. ESHRE consensus on the definition of ‘poor response’ to ovarian stimulation for in vitro fertilization: the Bologna criteria. *Hum Reprod*. 2011;26:1616–1624.
89. Polyzos NP, Devroey P. A systematic review of randomized trials for the treatment of poor ovarian responders: is there any light at the end of the tunnel? *Fertil Steril*. 2011;96:1058–1061.e7.
90. Casson PR, Lindsay MS, Pisarska MD, et al. Dehydroepiandrosterone supplementation augments ovarian stimulation in poor responders: a case series. *Hum Reprod*. 2000;15:2129–2132.
91. Patrizio P, Vaiarelli A, Levi Setti PE, et al. How to define, diagnose and treat poor responders? Responses from a worldwide survey of IVF clinics. *Reprod Biomed Online*. 2015;30:581–592.
92. Barad D, Gleicher N. Effect of dehydroepiandrosterone on oocyte and embryo yields, embryo grade and cell number in IVF. *Hum Reprod*. 2006;21:2845–2849.
93. Kotb MM, Hassan AM, AwadAllah AM. Does dehydroepiandrosterone improve pregnancy rate in women undergoing IVF/ICSI with expected poor ovarian response according to the Bologna criteria? A randomized controlled trial. *Eur J Obstet Gynecol Reprod Biol*. 2016;200:11–15.
94. Chern CU, Tsui KH, Vitale SG, et al. Dehydroepiandrosterone (DHEA) supplementation improves in vitro fertilization outcomes of poor ovarian responders, especially in women with low serum concentration of DHEA-S: a retrospective cohort study. *Reprod Biol Endocrinol*. 2018;16:90.
95. Tartagni M, Cicinelli MV, Baldini D, et al. Dehydroepiandrosterone decreases the age-related decline of the in vitro fertilization outcome in women younger than 40 years old. *Reprod Biol Endocrinol*. 2015;13:18.
96. Gleicher N, Ryan E, Weghofer A, et al. Miscarriage rates after dehydroepiandrosterone (DHEA) supplementation in women with diminished ovarian reserve: a case control study. *Reprod Biol Endocrinol*. 2009;7:108.
97. Liu Y, Hu L, Fan L, et al. Efficacy of dehydroepiandrosterone (DHEA) supplementation for in vitro fertilization and embryo transfer cycles: a systematic review and meta-analysis. *Gynecol Endocrinol*. 2018;34:178–183.
98. Kara M, Aydin T, Aran T, et al. Does dehydroepiandrosterone supplementation really affect IVF-ICSI outcome in women with poor ovarian reserve? *Eur J Obstet Gynecol Reprod Biol*. 2014;173:63–65.
99. Hyman JH, Margalioth EJ, Rabinowitz R, et al. DHEA supplementation may improve IVF outcome in poor responders: a proposed mechanism. *Eur J Obstet Gynecol Reprod Biol*. 2013;168:49–53.
100. Astapova O, Minor BMN, Hammes SR. Physiological and pathological androgen actions in the ovary. *Endocrinology*. 2019;160:1166–1174.
101. Tsui KH, Wang PH, Lin LT, et al. DHEA protects mitochondria against dual modes of apoptosis and necroptosis in human granulosa HO23 cells. *Reproduction*. 2017;154:101–110.
102. Li CJ, Chen SN, Lin LT, et al. Dehydroepiandrosterone ameliorates abnormal mitochondrial dynamics and mitophagy of cumulus cells in poor ovarian responders. *J Clin Med*. 2018;7:293.
103. Fouany MR, Sharara FI. Is there a role for DHEA supplementation in women with diminished ovarian reserve? *J Assist Reprod Genet*. 2013;30:1239–1244.
104. Kroboth PD, Salek FS, Pittenger AL, et al. DHEA and DHEA-S: a review. *J Clin Pharmacol*. 1999;39:327–348.
105. Kim C, Harlow SD, Zheng H, et al. Changes in androstenedione, dehydroepiandrosterone, testosterone, estradiol, and estrone over the menopausal transition. *Womens Midlife Health*. 2017;3:9.
106. Crawford S, Santoro N, Laughlin GA, et al. Circulating dehydroepiandrosterone sulfate concentrations during the menopausal transition. *J Clin Endocrinol Metab*. 2009;94:2945–2951.
107. Nanba AT, Rege J, Ren J, et al. 11-Oxygenated C19 Steroids Do Not Decline With Age in Women. *J Clin Endocrinol Metab*. 2019;104:2615–2622.
108. Davio A, Woolcock H, Nanba AT, et al. Sex differences in 11-oxygenated androgen patterns across adulthood. *J Clin Endocrinol Metab*. 2020;105:e2921–e2929.
109. Barrett-Connor E, Khaw KT, Yen SS. A prospective study of dehydroepiandrosterone sulfate, mortality, and cardiovascular disease. *N Engl J Med*. 1986;315:1519–1524.
110. Barrett-Connor E, Goodman-Gruen D. Dehydroepiandrosterone sulfate does not predict cardiovascular death in postmenopausal women. The Rancho Bernardo Study. *Circulation*. 1995;91:1757–1760.
111. Tivesten Å, Vandenput L, Carlzon D, et al. Dehydroepiandrosterone and its sulfate predict the 5-year risk of coronary heart disease events in elderly men. *J Am Coll Cardiol*. 2014;64:1801–1810.
112. Shufelt C, Bretsky P, Almeida CM, et al. DHEA-S levels and cardiovascular disease mortality in postmenopausal women: results from the National Institutes of Health--National Heart, Lung, and Blood Institute

(NHLBI)-sponsored Women's Ischemia Syndrome Evaluation (WISE). *J Clin Endocrinol Metab*. 2010;95:4985–4992.
113. Wu TT, Chen Y, Zhou Y, et al. Prognostic value of dehydroepiandrosterone sulfate for patients with cardiovascular disease: a systematic review and meta-analysis. *J Am Heart Assoc*. 2017;6:e004896.
114. Jia X, Sun C, Tang O, et al. Plasma dehydroepiandrosterone sulfate and cardiovascular disease risk in older men and women. *J Clin Endocrinol Metab*. 2020;105:e4304–e4327.
115. Akishita M, Hashimoto M, Ohike Y, et al. Association of plasma dehydroepiandrosterone-sulfate levels with endothelial function in postmenopausal women with coronary risk factors. *Hypertens Res*. 2008;31:69–74.
116. Jiménez MC, Sun Q, Schürks M, et al. Low dehydroepiandrosterone sulfate is associated with increased risk of ischemic stroke among women. *Stroke*. 2013;44:1784–1789.
117. Bonnet S, Paulin R, Sutendra G, et al. Dehydroepiandrosterone reverses systemic vascular remodeling through the inhibition of the Akt/GSK3-{beta}/NFAT axis. *Circulation*. 2009;120:1231–1240.
118. Aoki K, Taniguchi H, Ito Y, et al. Dehydroepiandrosterone decreases elevated hepatic glucose production in C57BL/KsJ-db/db mice. *Life Sci*. 2004;74:3075–3084.
119. Hansen PA, Han DH, Nolte LA, et al. DHEA protects against visceral obesity and muscle insulin resistance in rats fed a high-fat diet. *Am J Physiol*. 1997;273:R1704–R1708.
120. Maggio M, De Vita F, Fisichella A, et al. DHEA and cognitive function in the elderly. *J Steroid Biochem Mol Biol*. 2015;145:281–292.
121. Maninger N, Wolkowitz OM, Reus VI, et al. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS). *Front Neuroendocrinol*. 2009;30:65–91.
122. Prough RA, Clark BJ, Klinge CM. Novel mechanisms for DHEA action. *J Mol Endocrinol*. 2016;56:R139–R155.
123. Compagnone NA, Mellon SH. Dehydroepiandrosterone: a potential signalling molecule for neocortical organization during development. *Proc Natl Acad Sci U S A*. 1998;95:4678–4683.
124. Yilmaz C, Karali K, Fodelianaki G, et al. Neurosteroids as regulators of neuroinflammation. *Front Neuroendocrinol*. 2019;55:100788.
125. Suzuki M, Wright LS, Marwah P, et al. Mitotic and neurogenic effects of dehydroepiandrosterone (DHEA) on human neural stem cell cultures derived from the fetal cortex. *Proc Natl Acad Sci USA*. 2004;101:3202–3207.
126. Karishma KK, Herbert J. Dehydroepiandrosterone (DHEA) stimulates neurogenesis in the hippocampus of the rat, promotes survival of newly formed neurons and prevents corticosterone-induced suppression. *Eur J Neurosci*. 2002;16:445–453.
127. Pediaditakis I, Iliopoulos I, Theologidis I, et al. Dehydroepiandrosterone: an ancestral ligand of neurotrophin receptors. *Endocrinology*. 2015;156:16–23.
128. Alexaki VI, Fodelianaki G, Neuwirth A, et al. DHEA inhibits acute microglia-mediated inflammation through activation of the TrkA-Akt1/2-CREB-Jmjd3 pathway. *Mol Psychiatry*. 2018;23:1410–1420.
129. van Zuiden M, Haverkort SQ, Tan Z, Daams J, et al. DHEA and DHEA-S levels in posttraumatic stress disorder: a meta-analytic review. *Psychoneuroendocrinology*. 2017;84:76–82.
130. Schindler L, Shaheen M, Saar-Ashkenazy R, et al. Victims of war: dehydroepiandrosterone concentrations in hair and their associations with trauma sequelae in Palestinian adolescents living in the West Bank. *Brain Sci*. 2019;9:20.
131. Whitham JC, Bryant JL, Miller LJ. Beyond glucocorticoids: integrating dehydroepiandrosterone (DHEA) into animal welfare research. *Animals (Basel)*. 2020;10:1381.
132. Souza-Teodoro LH, de Oliveira C, Walters K, et al. Higher serum dehydroepiandrosterone sulfate protects against the onset of depression in the elderly: findings from the English Longitudinal Study of Aging (ELSA). *Psychoneuroendocrinology*. 2016;64:40–46.
133. Arlt W, Hammer F, Sanning P, et al. Dissociation of serum dehydroepiandrosterone and dehydroepiandrosterone sulfate in septic shock. *J Clin Endocrinol Metab*. 2006;91:2548–2554.
134. Ji E, Weickert CS, Purves-Tyson T, et al. Cortisol-dehydroepiandrosterone ratios are inversely associated with hippocampal and prefrontal brain volume in schizophrenia. *Psychoneuroendocrinology*. 2021;123:104916.
135. Arlt W. Dehydroepiandrosterone and ageing. *Best Pract Res Clin Endocrinol Metab*. 2004;18:363–380.
136. Nair KS, Rizza RA, O'Brien P, et al. DHEA in elderly women and DHEA or testosterone in elderly men. *N Engl J Med*. 2006;355:1647–1659.
137. Basu R, Dalla Man C, Campioni M, et al. Two years of treatment with dehydroepiandrosterone does not improve insulin secretion, insulin action, or postprandial glucose turnover in elderly men or women. *Diabetes*. 2007;56:753–766. *Erratum in: Diabetes*. 2007;56:1486.
138. Espinosa De Ycaza AE, Rizza RA, Nair KS, et al. Effect of dehydroepiandrosterone and testosterone supplementation on systemic lipolysis. *J Clin Endocrinol Metab*. 2016;101:1719–1728.
139. Wolkowitz OM, Kramer JH, Reus VI, et al. DHEA-Alzheimer's Disease Collaborative Research. DHEA treatment of Alzheimer's disease: a randomized, double-blind, placebo-controlled study. *Neurology*. 2003;60:1071–1076.
140. Walther A, Seuffert J. Testosterone and dehydroepiandrosterone treatment in ageing men: are we all set? *World J Mens Health*. 2020;38:178–190.
141. Weiss EP, Villareal DT, Ehsani AA, et al. Dehydroepiandrosterone replacement therapy in older adults improves indices of arterial stiffness. *Aging Cell*. 2012;11:876–884.
142. Labrie F, Archer DF, Koltun W, et al. Efficacy of intravaginal dehydroepiandrosterone (DHEA) on moderate to severe dyspareunia and vaginal dryness, symptoms of vulvovaginal atrophy, and of the genitourinary syndrome of menopause. *Menopause*. 2018;25:1339–1353.
143. Stewart PM. Aging and fountain-of-youth hormones. *N Engl J Med*. 2006;355:1724–1726.
144. Parasrampuria J, Schwartz K, Petesch R. Quality control of dehydroepiandrosterone dietary supplement products. *J Am Med Assoc*. 1998;280:1565.

91

Adrenal Pathology

Ozgur Mete and Lori A. Erickson

OUTLINE

Anatomic Pathology in Developmental Abnormalities of the Adrenal Gland, 1509
  Adrenal Aplasia or Hypoplasia, 1509
  Accessory or Heterotopic Adrenal Tissue, 1509
Tumor-Like Lesions of the Adrenal Gland, 1510
  Adrenal Myelolipoma, 1510
  Ectopic Thyroid Tissue, 1510
  Ovarian Thecal Metaplasia, 1510
  Adrenal Amyloidosis, 1511
  Adrenal Calcification, 1511
  Adrenal Cysts, 1511
  Adrenal Infection and Abscess, 1512
  Congenital Adrenal Hyperplasia, 1513
  Adrenal Cytomegaly, 1513
  Inflammatory Adrenalitis, 1513
Functional Approach in Adrenal Pathology, 1513
  Pathological Correlates of Primary Aldosteronism, 1514
  Pathological Correlates of Adrenal Cushing Syndrome, 1516
  Pathological Correlates of Virilism/Feminization, 1517
  Pathological Correlates of Catecholamine Excess, 1518
Pathology in the Distinction of Adrenocortical Carcinoma, 1519
Diagnostic and Predictive Immunohistochemical Biomarkers in Adrenal Pathology, 1523
  Confirmation of Adrenocortical and Medullary Origin, 1523
  Confirmation of Functional Sites in Primary Aldosteronism, 1523
  Diagnostic and Predictive Biomarkers of Adrenocortical Carcinoma, 1523
  Molecular Immunohistochemistry in Pheochromocytoma, 1524
Synoptic Reporting in Adrenocortical Carcinoma and Pheochromocytoma, 1525

ANATOMIC PATHOLOGY IN DEVELOPMENTAL ABNORMALITIES OF THE ADRENAL GLAND

The adrenal cortex consists of steroidogenic cells that originate from the intermediate mesoderm.[1] A web of factors, including transcriptional factors (e.g., DAX1 encoded by *NR0B1*, SF1/Ad4BP encoded by *NR5A1,* and GATA4/6 encoded by *GATA4* and *GATA6*), play a crucial role during the organogenesis and functional differentiation of adrenocortical cells.[1,2,3] This complex developmental process stems from the formation of adrenogonadal primordia within the urogenital ridge and is followed by the separation of surrounding mesenchyme and formation of an encapsulated adrenal anlagen (fetal adrenal zone), which later develops into the adult adrenal cortex via the expansion and maturation of the definitive adrenocortical zones.[1]

The adrenal medulla consists of chromaffin cells that are surrounded by mesodermal steroidogenic cortical cells. Recent evidence suggests that only a small fraction of the adrenal medulla originates directly from the migratory neural crest cells, whereas preganglionic nerve fiber–related pluripotent Schwann cell precursors at the adrenal anlage provide the major path for the development of chromaffin cells.[4,5]

Disruption in various steps of the organogenesis results in developmental abnormalities. This section provides a brief summary of developmental conditions related to anatomic pathology.

Adrenal Aplasia or Hypoplasia

Adrenal aplasia (agenesis) is a very rare condition that refers to the absence of the adrenal gland, whereas adrenal hypoplasia refers to adrenocortical atrophy. Anencephaly is a well-known cause of neonatal adrenal hypoplasia or aplasia.[6] In humans, *NR5A1* mutations are rare, but they have been reported in bilateral adrenal aplasia or hypoplasia.[2,7] Patients with Pollister–Hall syndrome harbor *GLI3* mutations and can exhibit adrenal hypoplasia.[2] Morphologically, hypoplastic adrenal glands appear to have a prominent medulla due to significantly reduced thickness of the cortex without normal cortical zonation. Several etiologies can result in adrenal hypoplasia (e.g., exogenous glucocorticoid administration, insufficiency related to the hypothalamic-pituitary-adrenal axis, including but not limited to adrenocorticotropic hormone [ACTH] resistance and familial glucocorticoid deficiency syndromes). Congenital adrenal hypoplasia is also linked to *NR0B1*-related pathogenesis. The latter can either manifest with hemizygous pathogenic *NR0B1* variants that cause X-linked adrenal hypoplasia congenita with hypogonadotropic hypogonadism or Xp21 deletion, which results in deletion of *NR0B1* and *GK* (complex glycerol kinase deficiency), and sometimes *DMD* (Duchenne muscular dystrophy).[8] Cytomegalic cells with lipid depletion pattern have been reported in hypoplastic adrenal glands.

KEY POINTS

- Defects in primary and secondary developmental regulators of the adrenal cortex can lead to congenital adrenal aplasia or hypoplasia.

Accessory or Heterotopic Adrenal Tissue

Accessory or heterotopic adrenocortical tissue is not an uncommon finding. These consist of isolated adrenocortical tissue rests and may have associated adrenal medullary cells (chromaffin cells), as seen in an orthotopic adrenal gland. The occurrence of chromaffin cells in heterotopic adrenal tissue also underscores the crosstalk between adrenal cortex and medulla during embryogenesis. Delamination in the adrenogonadal primordia leads to the separation and differential migration of embryonic cells towards the gonadal anlagen and the adrenal anlagen. Therefore, failure in encapsulation of steroidogenic factor 1 (SF1)-expressing developmental cells that are responsible for the formation of the fetal adrenal zone may explain the origin of adrenocortical tissue rests that are identified in the abdomen (e.g., along the celiac axis, kidney and kidney capsule, periadrenal, spleen and splenic hilum, mesentery), as well as in various anatomic locations that are aligned with the developmental tract of the gonads (e.g., spermatic cord, scrotum, broad ligament, uterus, ovarian and spermatic vessels, gonads).[9,10] However, the occurrence of adrenocortical rests in unusual locations such as brain, pituitary, spinal nerves, lung, liver, and pancreas also expands the spectrum of adrenocortical tissue rests that cannot be explained from an embryological point of view.[9,10,11] The term "adrenocortical choristoma" is applied to unusual heterotopic adrenocortical rests (e.g., adrenocortical choristoma in a silent corticotroph tumor).[11] Failure to recognize these atypical locations may lead to diagnostic dilemmas.

Mechanisms that alter with the successful separation of the developmental periadrenal mesenchyme may result in adrenal fusions (union) or adhesions.

Adrenal fusions (unions) occur between the adrenal gland and parenchymal elements of the liver or kidney (e.g., adrenohepatic fusion, adrenorenal fusion) (Fig. 91.1). Morphologically, in the setting of an adrenal fusion, there is no intervening connective tissue capsule separating cellular elements of the organ that is united with the adrenal gland. Therefore, adrenocortical elements are intermingled with cellular components of the other organ. Midline adrenal fusions also occur between both adrenal glands. The latter leads to the gross appearance of a single horseshoe-shaped midline adrenal gland. Other developmental defects, including those of central nervous system, can be identified in patients with midline adrenal unions.[9]

In adrenal adhesions, there is an intact intervening connective tissue capsule separating parenchymal elements of the adrenal gland and the other organ (e.g., adrenohepatic adhesion). As a consequence, the adrenal gland displays a normal medulla and preserved cortical zonation. Failure to recognize the spectrum of developmental abnormalities such as adrenohepatic fusion or adrenohepatic adhesion, and adrenocortical rests in the liver can result in misinterpretation of a benign developmental process as invasion or metastasis of an adrenocortical carcinoma. The latter may be diagnostically challenging, especially in small biopsy specimens.

TUMOR-LIKE LESIONS OF THE ADRENAL GLAND

KEY POINTS

- Tumor-like lesions of the adrenal gland encompass a large spectrum of lesions, including adrenal myelolipoma, ectopic thyroid tissue, ovarian thecal metaplasia, amyloid deposition, adrenal hematoma and hemorrhage, adrenal calcification, adrenal cysts, infection with or without abscess, congenital adrenal hyperplasia, and adrenal cytomegaly with or without Beckwith–Wiedemann syndrome.

Adrenal Myelolipoma

Myelolipoma consists of a mature adipose tissue admixed with variable number of hematopoietic cells, including megakaryocytes (Fig. 91.2). Myelolipomas with extensive hemorrhage can simulate an adrenal hematoma or malignant process. Microscopic foci of myelolipomatous change can occur in adrenocortical proliferations. Bilateral adrenal gland involvement was recorded in association with congenital adrenal hyperplasia (CAH) and primary bilateral macronodular adrenocortical disease (also known as primary bilateral macronodular adrenocortical hyperplasia).[12,13] Giant adrenal myelolipomas exceeding 30 cm have been defined in patients with CAH.[12] Increased ACTH- and androgen-mediated stimulus and overexpression of melanocortin 2 receptor and androgen receptor have been postulated in the developmental pathogenesis of some adrenal myelolipomas, especially in the setting of CAH.[14]

Ectopic Thyroid Tissue

The presence of thyroid tissue within the adrenal gland expands the unusual spectrum of ectopic thyroid tissue outside the usual migratory path of the thyroid anlage that typically lies between the foramen cecum and mediastinum.[15,16] An abnormal descent of the thyroid anlage can be the source of ectopic thyroid tissue that can be identified in adrenals (Fig. 91.3). While diagnosticians are still required to exclude the possibility of metastatic differentiated thyroid carcinoma in any unusual location, failure to recognize the occurrence of ectopic thyroid tissue in abdominal organs may result in misdiagnosis.[15]

Ovarian Thecal Metaplasia

Ovarian thecal metaplasia is a relatively rare tumor-like lesion of the adrenal gland. Originally described in postmenopausal women with markedly elevated gonadotropins, these abnormal cellular nests were defined as wedge-shaped cortical lesions that are composed of ovarian stroma-like elements. However, these lesions have been defined in both women and men. Subsequent evidence also suggested that these

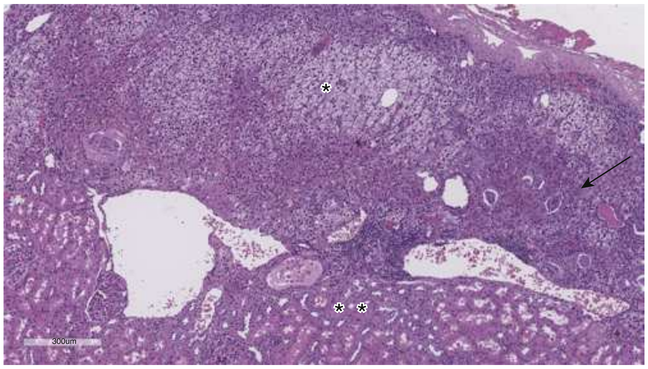


Fig. 91.1 Adrenorenal fusion. The adrenal cortex (*) is in continuity with the renal parenchyma (**). There is no limiting connective tissue between adrenal gland and kidney. The arrow illustrates renal elements in the zona reticularis.

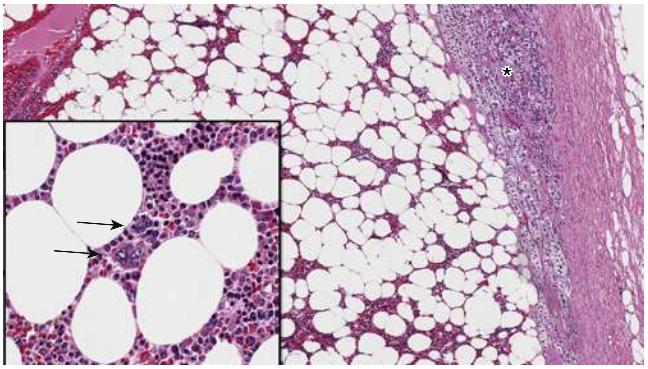


Fig. 91.2 Adrenal myelolipoma. Myelolipoma consists of a mature adipose tissue admixed with variable number of hematopoietic cells including megakaryocytes. Asterisk indicates the nonlesional adrenal cortex; arrows indicate multinucleated megakaryocytes.

lesions rarely contain true ovarian stromal elements, and are composed of fibroblastic and/or myofibroblastic mesenchymal cells. Some cases featuring radial scar-like appearance have also been defined.[17]

Adrenal Amyloidosis

Adrenal amyloidosis is a rare cause of adrenal insufficiency.[9] Affected individuals may exhibit adrenal enlargement, depending on the degree of the involvement. Histologically, amyloid deposition is characterized by a homogenous acellular pink hyaline material. It typically occurs in the wall of vascular structures, but also in the extracellular stroma between adrenocortical cells. The later results in diffuse replacement of cortical cells, depending on the extent of the disease. Special histochemical stains are frequently used to distinguish amyloid from fibrosis or stromal hyalinization, which can be identified in cortical nodular disease as well as in autoimmune adrenalitis.

Adrenal Calcification

Adrenocortical neoplasms and medullary tumors may contain dystrophic calcifications and even metaplastic bone formation. The nonneoplastic correlates of adrenal calcifications can be seen in the setting of infectious disease (e.g., histoplasma), adrenal cysts (see later), Wolman disease (lysosomal acid lipase deficiency) and GM1 gangliosidosis (β-galactosidase deficiency).[9,10,18,19] Wolman disease manifests with hepatosplenomegaly and infantile bilateral adrenal enlargement with dystrophic calcifications.[9,18] Necrosis and fibrosis can also occur. GM1 gangliosidosis presents with bilateral adrenal calcifications restricted to the medulla, given the accumulation of ganglioside in neuronal tissues.[10,19]

Adrenal Cysts

Primary adrenal cysts have four histologic variants, including (i) vascular (endothelial) adrenal cysts, (ii) epithelial adrenal cysts, (iii) parasitic adrenal cysts, and (iv) adrenal pseudocysts.[9,10,20] Most primary adrenal cysts are unilateral disease. Pseudocysts and vascular (endothelial) adrenal cysts are the most common subtypes that can simulate adrenocortical or medullary neoplasm.

Absence of epithelial or endothelial lining distinguishes adrenal pseudocysts from other types of adrenal cysts[9,20] (Fig. 91.4). These often have hemorrhagic, fibrinoid, or necrotic cyst content that can be mistaken for an adrenocortical carcinoma on imaging studies.[9,20] Vascular adrenal cysts are typically lined by endothelial cells that are commonly of lymphatic endothelial cell origin (Fig. 91.5). Vascular cysts may mimic pheochromocytomas. Epithelial cysts are lined by keratin-expressing cells of mesothelial origin (Fig. 91.6). Parasitic cysts occur in the course of parasitic adrenal infections (e.g., hydatid disease). Adrenal cysts should be distinguished from other adrenal lesions that can manifest with cystic degeneration, including adrenocortical neoplasms, pheochromocytomas, and neuroblastic tumors, as well as some metastatic carcinomas. Beckwith–Wiedeman syndrome can also show hemorrhagic multiloculated adrenal cysts in the background of adrenal enlargement.

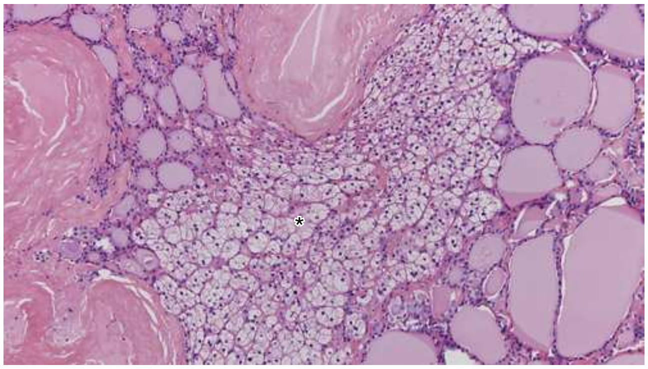


Fig. 91.3 Ectopic thyroid tissue in the adrenal. Adrenal gland can host ectopic thyroid tissue. The lack of cytological and architectural atypia warrants a diagnosis of benign ectopic thyroid tissue. The asterisk indicates the nonlesional adrenocortical cells in the zona fasciculata.

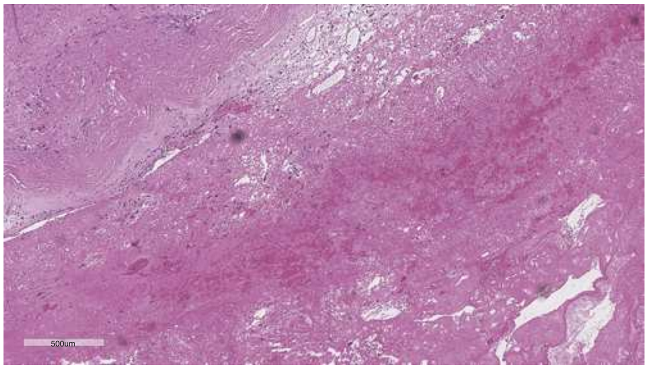


Fig. 91.4 Adrenal pseudocyst. Adrenal pseudocysts have no lining and are composed of hemorrhagic fibrinoid material.

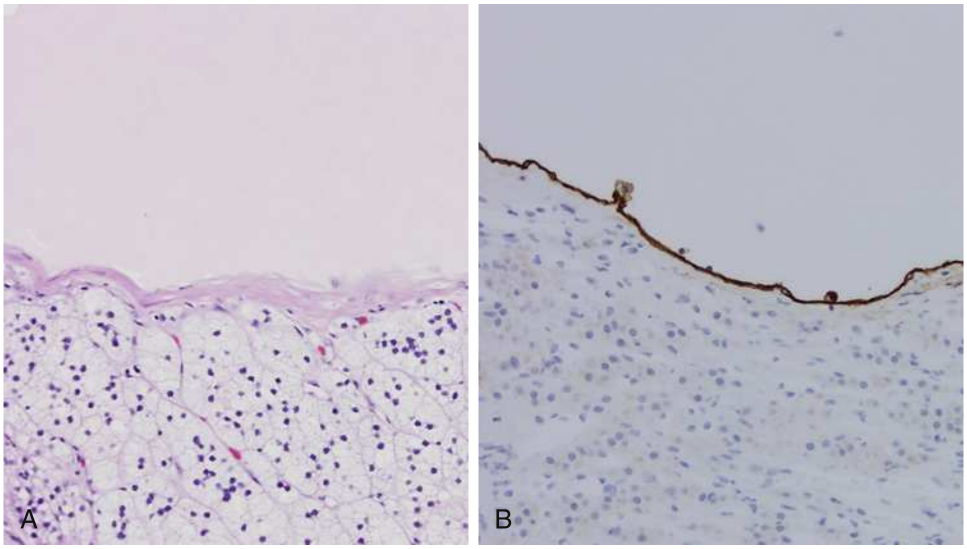


Fig. 91.5 Vascular (endothelial) cyst of the adrenal. This composite photomicrograph illustrates a vascular cyst surrounded by nonlesional adrenocortical cells (A). The lumen is acellular (A). CD31 (not illustrated herein) and D2-40 stain the vascular endothelial lining of this cyst (B).

Adrenal Infection and Abscess

Primary adrenal infection is a rare clinical presentation, because most develop as a result of widespread or locoregional infections.[9,10,21-24] Immunocompromised individuals have an increased risk of adrenal infection. On computed tomography (CT) scans, infection-related adrenal enlargement can be mistaken for an adrenal tumor because of a mass-like appearance or diffuse enlargement with variable heterogeneity.[21,22] Adrenal insufficiency (Addison disease) occurs, depending on the extent of tissue destruction.

Various microorganisms can cause adrenal infections. Parasitic infestations caused by *Echinococcus,* amebic species, *Trypanosoma, Leishmania,* and *Microsporidia* have been reported in adrenals.[9,23] Parasites can manifest with adrenal cystic disease (e.g., hydatid disease).

Among fungal pathogens, *Cryptococcus* and *Histoplasma* are the most common pathogens in nontropical geography, whereas *Paracoccidioides* predominates tropic geographic zones.[9,10,23,25] Histologically, necrotizing granulomatous inflammation is common in fungal infections. Enlargement of the adrenal glands with extensive adrenal destruction in association with extensive caseous necrosis and calcification is a feature of disseminated *Histoplasma* infection[9,23] (Fig. 91.7). In addition to cultures, histochemical stains (periodic acid–Schiff and Grocott's methenamine silver stains) are used to confirm fungal elements.

Immunocompromised hosts are more likely to develop viral adrenal infections. Among these, cytomegalovirus (CMV)-related adrenalitis is well-recognized; however, other viruses can also involve adrenal glands.[9,23] Unlike fungal and mycobacterial infections, adrenals show no marked enlargement in viral infections. Mixed inflammation in the corticomedullary junction and/or hemorrhagic necrosis (more frequent in *Herpesviridae*) can be seen.[9,10,24] Demonstration of viral cytopathic changes (e.g., Cowdry type A inclusions in CMV) confirms the diagnosis of a viral adrenal infection. Immunohistochemical antibodies against viral antigens are also very useful in the detection of viral antigens.

Bacterial adrenal infections typically occur either due to mycobacteria infections (e.g., immunocompromised or tuberculosis reactivation) or in the context of hematogenous spread in bacterial

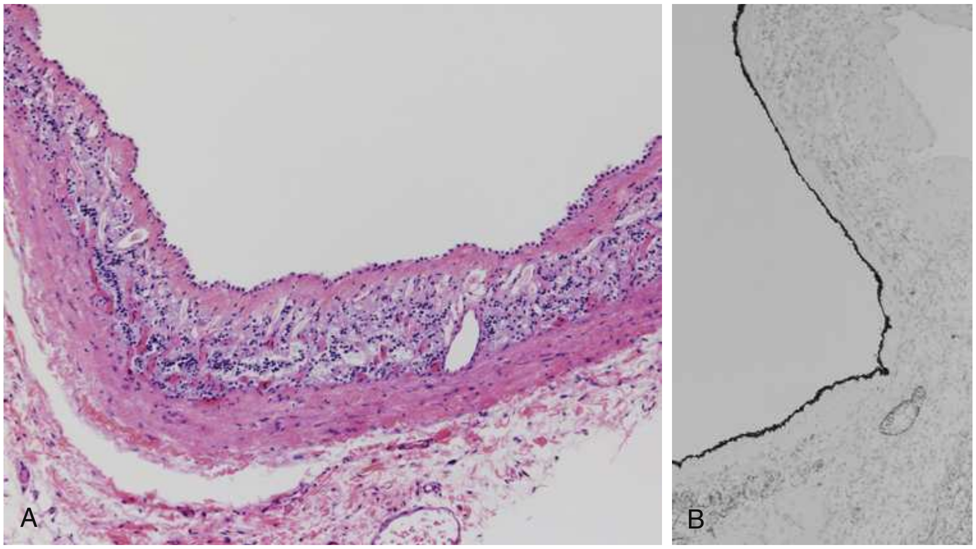


Fig. 91.6 **Epithelial cyst of the adrenal.** This composite photomicrograph illustrates an epithelial cyst surrounded by nonlesional adrenocortical cells (**A**). Pankeratin stains the lining of this cyst, as well as some cortical cells (**B**).

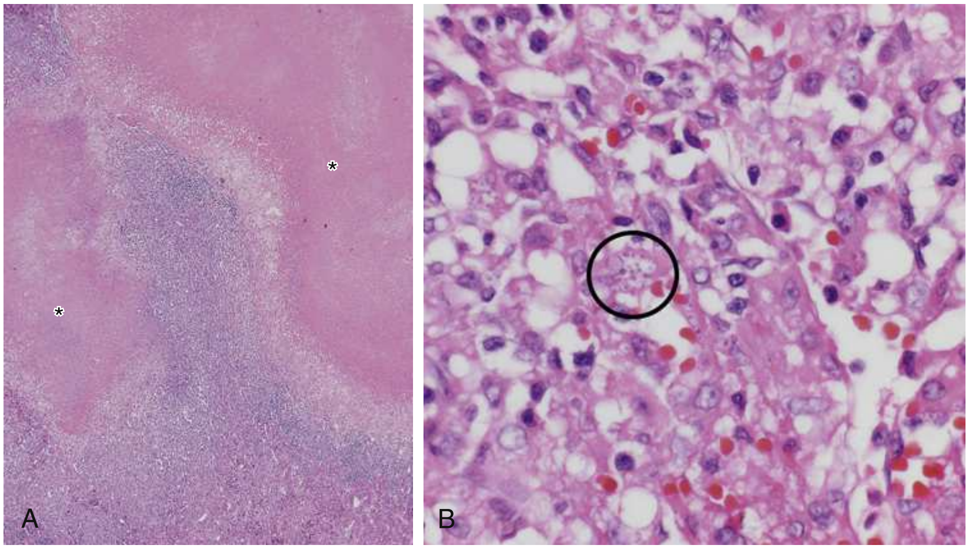


Fig. 91.7 **Histoplasma infection in the adrenal.** This composite photomicrograph illustrates a necrotizing granulomatous inflammation (**A**). Asterisks indicate the necrotic center of granulomatous inflammation. Some macrophages in the lesion contain multiple fungi, consistent with *Histoplasma* (**B**). Circled area represents fungal elements on hematoxylin and eosin–stained sections.

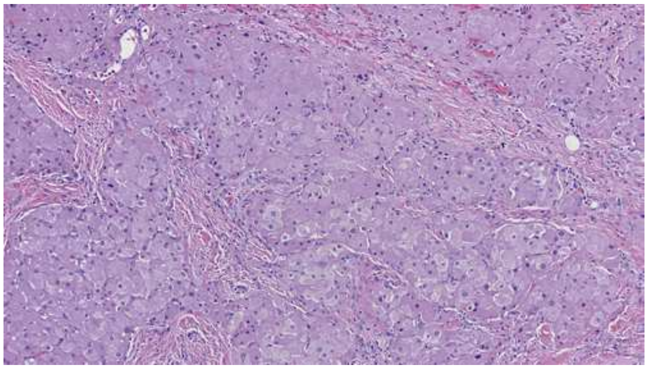


Fig. 91.8 **Congenital adrenal hyperplasia.** Bilateral compact (lipid-poor) cell adrenocortical hyperplasia is a characteristic morphologic feature of this disorder.

sepsis (e.g., *Streptococcus, Haemophilus, Staphylococcus, Meningococcus, Escherichia, Pseudomonas,* and *Klebsiella*).[9] Bacterial adrenal infections can manifest with adrenal enlargement, hemorrhage, necrosis, abscess, and malakoplakia.[9] Abscess occurs as a result of hematogenous spread of bacterial pathogens. Unilateral adrenal abscess can be mistaken for a tumor. Adrenal malakoplakia is a rare tumor-like inflammatory condition caused by gram-negative bacteria. Histologically, large collections of foamy macrophages with intracytoplasmic basophilic inclusions (Michaelis–Gutmann bodies) are characteristics features of malakoplakia. Waterhouse–Friderichsen syndrome refers to bilateral adrenal hemorrhage and necrosis that is caused by bacterial sepsis. Historically, mycobacterium tuberculosis has been recognized to account an important etiology of Addison disease.[9,10] Bilateral marked adrenal enlargement with variable degree of necrotizing granulomatous inflammation is a frequent finding in mycobacterial adrenal infections. Special histochemical stains (e.g., Ziehl–Neelsen) are required to visualize acid-fast bacilli in the cytoplasm of macrophages. However, histochemical stains fail to detect a very low number of bacteria; therefore, tissue cultures, PCR studies, and interferon-gamma release assays such as QuantiFERON have been frequently used as the golden standards in the distinction of mycobacterial infection (including the latent infection), as well as for the accurate characterization of mycobacteria.

Congenital Adrenal Hyperplasia

Defects in the enzymes involved steroid hormone biosynthesis cause CAH. Most common forms have been linked to *CYP21*-related 21-hydroxylase deficiency.[9,10,24] Patients who are not adequately treated with exogenous steroids manifest with marked bilateral adrenal enlargement due to ACTH-driven bilateral diffuse compact cell adrenocortical hyperplasia[9,10,24] (Fig. 91.8). Heterotopic adrenocortical tissue can become hyperplastic in patients with CAH. These hyperplastic tissues often simulate neoplastic disease in various locations.

Adrenal Cytomegaly

Adrenal cytomegaly is characterized by variable amount of enlarged adrenocortical cells with pleomorphic hyperchromatic nuclei with or without pseudoinclusions. Adrenal cytomegaly is distinguished from viral cytopathic changes of adrenal CMV infections by the presence of a single eosinophilic nuclear inclusion that is typically surrounded by a clear halo (Cowdry A) in infected cortical cells.[9] Premature infants and neonates can display areas of adrenal cytomegaly.[9,24] This finding does not qualify as a precursor or malignant process, and adrenal cytomegaly often resolves within months or years. However, focal adrenal cytomegaly rarely persists in other age groups.

Bilateral adrenal gland enlargement due to cortical proliferations (usually adrenocortical hyperplasia) with prominent adrenal cytomegaly, hemorrhagic cortical cysts, and prominence of chromaffin cells of the adrenal medulla are characteristic features of Beckwith–Wiedemann syndrome (Fig. 91.9). This congenital overgrowth syndrome is caused by altered imprinting of genes in chromosome *11p15.5* and has also been regarded as a predisposing condition for adrenocortical carcinoma.[26]

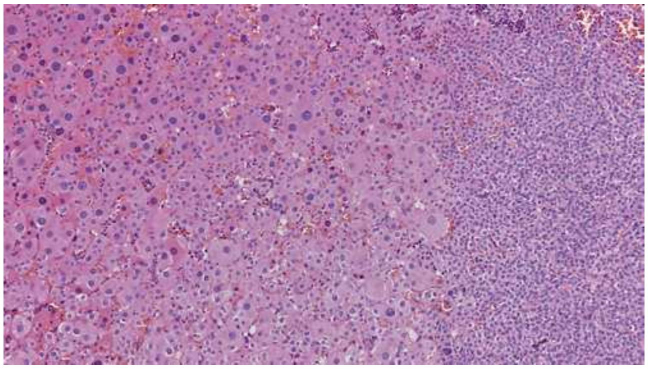


Fig. 91.9 **Adrenal cytomegaly in Beckwith–Wiedemann syndrome.** Bilateral adrenal gland enlargement due to cortical proliferations (usually adrenocortical hyperplasia) with prominent adrenal cytomegaly (cortical cells with enlarged nuclei) is one of the characteristic features of Beckwith–Wiedemann syndrome.

Inflammatory Adrenalitis

Focal adrenalitis is not an uncommon histologic finding, and it is often of no clinical significance. Noninfectious inflammatory changes are often seen in the context of autoimmune adrenalitis, which is typically seen in association with autoimmune polyglandular endocrinopathy types I and II[9,10] (Fig. 91.10). Depending of the disease stage, the cortex of both adrenals shows marked mixed chronic inflammation composed of lymphocytes, plasma cells, and macrophages.[9,10] The inflammatory adrenocortical destruction results in fibrohyalinization and small adrenal glands. The distinction of adrenalitis from a lymphoproliferative disorder is required when inflammation predominates the histology. Ancillary tools (e.g., immunohistochemistry and molecular pathology techniques) can be used to distinguish a lymphoproliferative disease from adrenalitis.

FUNCTIONAL APPROACH IN ADRENAL PATHOLOGY

The morphological spectrum of adrenal lesions includes hyperplastic and neoplastic disorders that are associated with hormonally active or nonfunctional "silent" clinicopathologic manifestations. Hormonally active adrenocortical lesions can cause mineralocorticoid excess (primary aldosteronism), glucocorticoid excess (adrenal Cushing syndrome and subclinical adrenal Cushing syndrome), sex hormone excess (virilism and feminization), and a combination of the these hormones excess syndromes.[9,27] Adrenal medullary lesions are often functional and result in catecholamine excess–related clinical signs and symptoms.

KEY POINTS

- Adrenocortical lesions are classified into three diagnostic morphological categories: (1) adrenocortical hyperplasia, (2) benign adrenocortical neoplasm, which includes adrenocortical adenomas and adrenocortical nodular disease, and (3) adrenocortical carcinoma. Micronodular and macronodular forms of adrenocortical nodular disease and diffuse adrenocortical hyperplasia are almost always functional disorders involving both adrenal glands. Nonfunctional lesions of the adrenal gland are more frequent in a subset of adrenocortical neoplasms that are enriched in Wnt pathway alterations.[28,29]

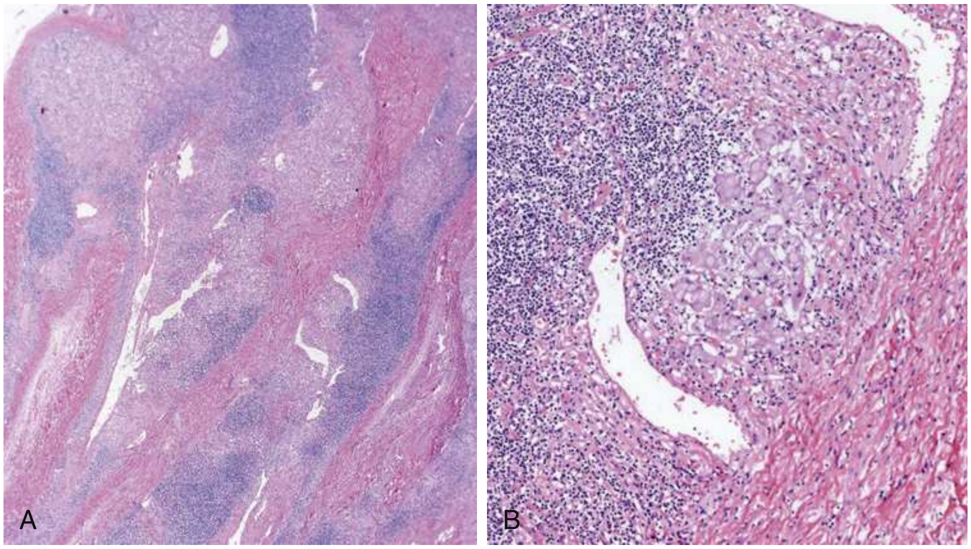


Fig. 91.10 **Inflammatory adrenalitis.** This composite photomicrograph illustrates marked mixed chronic inflammation composed of lymphocytes, plasma cells, and macrophages (A). The inflammatory adrenocortical destruction results in fibrohyalinization (A, B).

KEY POINTS

- The endocrine lesions of the adrenal medulla can be divided into two morphological categories: (1) adrenal medullary hyperplasia and (2) pheochromocytoma (adrenal paraganglioma). Pheochromocytomas are frequently associated with mature (adrenergic or mixed adrenergic and noradrenergic); however, immature (dopaminergic, noradrenergic, or mixed dopaminergic and noradrenergic) secretory biochemical phenotypes can also be identified.[30]

This section will provide a brief overview of the morphological entities that are reflected in functional adrenal disease. The details on molecular correlates of these lesions[29] are discussed in several other chapters of this book.

Pathological Correlates of Primary Aldosteronism

Discovery of ion channel mutations that result in the transcription of *CYP11B2* (encoding aldosterone synthase [CYP11B2]), occurrence of aldosterone-producing microscopic adenoma(s) that cannot be identified on CT imaging studies from patients with uniglandular disease, and application of CYP11B2 immunohistochemistry in adrenals from patients with primary aldosteronism have helped diagnosticians to better appreciate the clinicopathologic manifestations of primary aldosteronism.[29,31-34] The current evidence suggests that aldosterone-producing cortical lesions encompass a wide spectrum of lesions that range from microscopic to gross disease that can manifest with unifocal or multifocal disease with synchronous or asynchronous contralateral gland involvement.[29]

Traditionally, the most common clinical manifestation of primary aldosteronism is thought to represent bilateral hyperplasia of the zona glomerulosa[27,33]; however, the latter also includes bilateral microscopic aldosterone-producing adrenocortical adenomas that are not detected on imaging studies.[29] Bilateral diffuse zona glomerulosa hyperplasia is uncommon in the pathologist's practice, because affected patients often do not require bilateral adrenalectomy, but instead are treated with lifelong antimineralocorticoid therapy. Among neoplastic correlates of this disorder, aldosterone-producing carcinomas are extremely rare. As a consequence, aldosterone-producing benign cortical neoplasms account for most of the diagnoses of primary aldosteronism in the pathologist's practice (Fig. 91.11).

Aldosterone-producing adrenocortical adenomas show cytomorphological and epidemiological heterogeneity with respect to their underlying somatic mutations (Table 91.1).[29,32] For instance, aldosterone-producing adenomas that harbor *KCNJ5* mutations are enriched in lipid-rich zona fasciculata–type cortical cells. These tumors are more frequent in females, and are often solitary large tumors.[29,32] In contrast, adenomas that are associated with a nonmutated *KCNJ5* gene are enriched in compact cortical cells in the setting of multifocal disease with smaller tumor sizes when compared with *KCNJ5*-related disease. These tumors are also reported to be more frequent in males.[29,32]

The preoperative prediction of bilateral disease is still an ongoing challenge in the clinical management of patients with primary aldosteronism. Despite continued efforts to effectively preoperatively lateralize unilateral aldosterone excess via adrenal venous sampling, biochemical recurrence can still occur in such cases. Recently, genotype–phenotype correlations (Table 91.1) have provided critical insights into the genetic and cellular features underlying unilateral versus bilateral disease. Furthermore, CYP11B2 immunohistochemical features underscore morphological findings that may help predict the potential risk for bilateral disease.[29,31,34] As a consequence, the modern endocrine pathology practice of primary aldosteronism requires a thorough cytomorphological examination of the entire adrenal gland. For example, in patients with preoperative spironolactone treatment, functional histologic foci can display variable amount of spironolactone bodies that may provide clues to multifocality and bilaterality of disease.[27,33] However, the accurate distinction of functional sites requires the use of CYP11B2 immunohistochemistry in adrenalectomy specimens.[31,32,34]

Mixed aldosterone and glucocorticoid secretion from an adrenocortical adenoma can occur. This manifestation is more frequent in a subset of adrenocortical neoplasms with *KCNJ5* mutations.[35,36] Unlike aldosterone-producing benign cortical disease, aldosterone-producing adrenocortical carcinomas lack ion channel mutations but display genetic alterations as seen in other adrenocortical carcinomas.[37]

KEY POINTS

- The HISTALDO classification[34] has been introduced by a group of international disease experts to provide a unified diagnostic nomenclature relevant to reporting of surgical pathological findings in adrenalectomy specimens from patients with unilateral primary aldosteronism.[34] The conventional histopathological examination and CYP11B2 staining patterns identified in adrenal nodules and the adjacent cortex form the basis of this classification system.

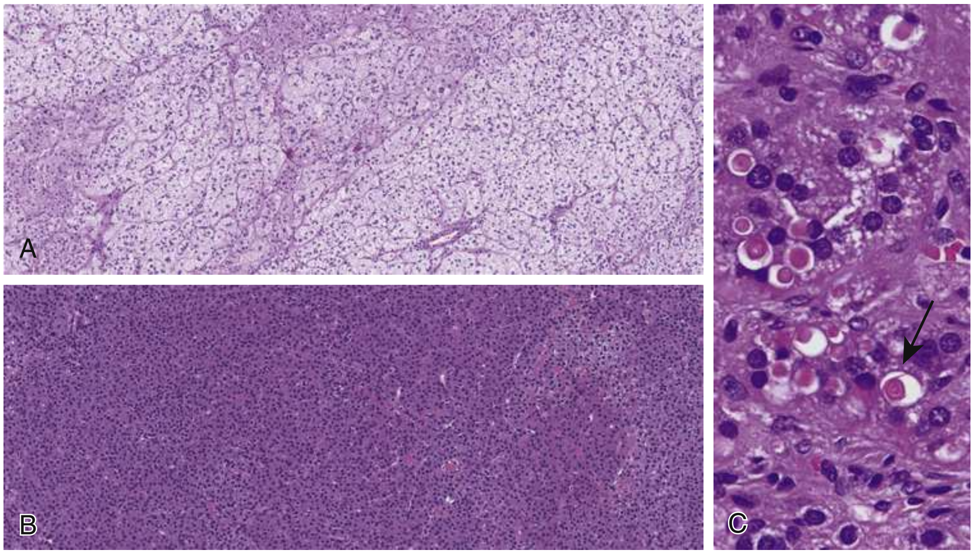


Fig. 91.11 **Aldosterone-producing adrenocortical adenoma (Conn syndrome).** The morphological heterogeneity in primary aldosteronism is reflected in this composite photomicrograph. Aldosterone-producing adenomas that harbor *KCNJ5* mutations are enriched in zona fasciculata–like clear cells (**A**), whereas *KCNJ5*–wild-type adenomas are enriched in zona reticularis– and zona glomerulosa–like cells (**B**). Patients treated with spironolactone often show tumor cells with eosinophilic lamellated inclusions that are also known as spironolactone bodies (**C**). The arrow indicates one of the many inclusions.

TABLE 91.1 Genotype–Phenotype Correlations in Primary Aldosteronism[29,32]

Molecular Subgroups	Gene(s)	Characteristics	Germline Variants*
***KCNJ5*-related PA**	*KCNJ5*	- Most frequent cause of PA - More frequent in females - Pronounced PA - Often earlier age of disease onset - Often unilateral disease - Often solitary or dominant adenoma - Enriched in ZF-like cortical cells - Higher *CYP17A1* and *CYP11B1* transcripts compared with *KCNJ5*–wild-type PA - Can feature cortisol cosecretion	FHA type III
***KCNJ5*–wild-type PA**	*ATP1A1* *ATP2B3* *ATP2B4*** *CACNA1H* *CACNA1D* *CLCN2*	- More frequent in males - Often less pronounced PA - Often later age of disease onset - Germline pathogenic *CACNA1H* and *CACNA1D* variants manifest with early age of disease onset - Smaller tumor size; increased frequency of multifocal disease - Increased risk of recurrence - Enriched in ZG- and ZR-like compact cells - Increased CYP11B2 to CYP11B1 ratio - With the exception of *CLCN2*-mutant tumors, lower levels of *CYP17A1* and *CYP11B1* expression - Higher levels of CYP11B2 mRNA transcripts compared with *KCNJ5*-related disease	FHA type II: *CLCN2* FHA type IV: *CACNA1H* PASNA syndrome: *CACNA1D* *ATP2B4* variants
***CTTNB1*-related PA**	*CTTNB1*	- Rare cause of PA (~5% of PA) - More frequent in females - Often later age of disease onset - Higher risk of recurrence - Heterogenous cytomorphology - Heterogenous CYP11B2 expression - Nuclear and cytoplasmic beta-catenin	

*Germline pathogenic variants that can be seen in this molecular subgroup.

**Rare germline variants.

PA, Primary aldosteronism; ZF, zona fasciculata; ZG, zona glomerulosa; ZR, zona reticularis; FHA, familial hyperaldosteronism; PASNA syndrome, early onset of primary aldosteronism with seizure and neurological abnormalities.

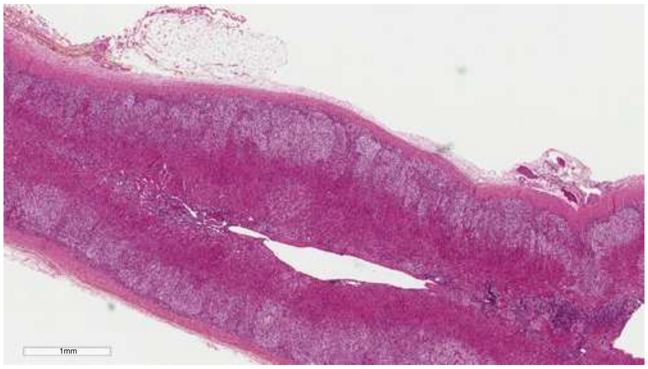


Fig. 91.12 **Pituitary ACTH-related diffuse adrenocortical hyperplasia.** Pituitary ACTH-related diffuse adrenocortical hyperplasia is associated with bilateral adrenal gland enlargement. While zona reticularis is slightly expanded, all three cortical zones are preserved.

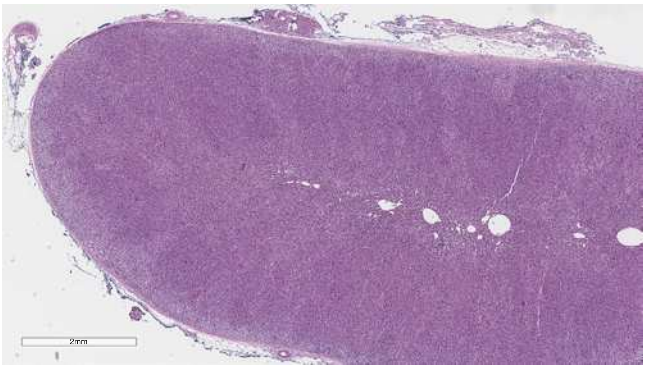


Fig. 91.13 **Ectopic adrenocorticotropic hormone (ACTH)-related diffuse adrenocortical hyperplasia.** Unlike the pituitary ACTH-related pathogenesis, the adrenocortical zonation is not preserved. The adrenal cortex in both adrenals show a lipid depletion pattern.

HISTALDO Classification

Aldosterone-producing carcinoma (APACC)

CYP11B2-expressing adrenocortical carcinoma

Aldosterone-producing adenoma (APA)

CYP11B2-expressing solitary benign adrenocortical neoplasm (≥1.0 cm)

Aldosterone-producing nodule (APN)

CYP11B2-expressing subcentimeter benign adrenocortical nodular disease; CYP11B2 staining is often stronger at the periphery

Aldosterone-producing micronodule (APM)

CYP11B2-expressing subcentimeter microscopic cortical proliferation consisting of zona glomerulosa–like cells that tend to display a gradient staining as seen in APN

Multiple aldosterone-producing nodules (MAPN) or multiple aldosterone-producing micronodules (MAPM)

Aldosterone-producing diffuse hyperplasia (APDH)

CYP11B2-expressing nonnodular broad or continuous zona glomerulosa–layer hyperplasia that accounts for >50% of the zona glomerulosa layer

The pathological correlates of adrenal Cushing syndrome

(v) **Diffuse adrenocortical hyperplasia**
  a) ACTH dependent (Pituitary, ectopic)
  b) CRH dependent (Hypothalamic, ectopic)
(vi) **Bilateral nodular adrenocortical disease**
  c) Micronodular form
  d) Macronodular form
(vii) **Adrenocortical adenoma**
(viii) **Adrenocortical carcinoma**

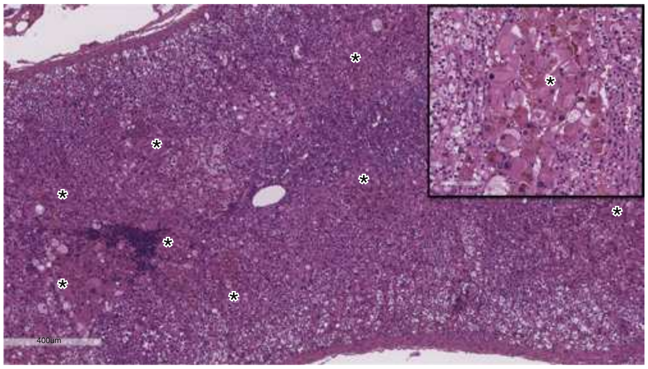


Fig. 91.14 **Primary bilateral micronodular adrenocortical disease.** Both adrenal glands are slightly enlarged and contain compact cell–rich micronodules that are variably pigmented. The illustration is from an individual with primary pigmented nodular adrenocortical disease. Asterisks indicate micronodules.

Pathological Correlates of Adrenal Cushing Syndrome

Diffuse adrenocortical hyperplasia is the only “true hyperplastic” disorder of the adrenal cortex in patients with Cushing syndrome (Fig. 91.12). The latter also accounts for the most frequent cause of endogenous Cushing syndrome.[32,38] From a pathogenic perspective, ACTH-dependent disease is due to functional pituitary corticotroph tumor in the vast majority of patients.[9,32,38] However, ectopic ACTH or corticotropin-releasing hormone (CRH) production in neuroendocrine neoplasms,[27,38] and CRH-producing hypothalamic tumors[39] can also result in diffuse adrenocortical hyperplasia. At the morphologic level, the adrenal cortex in both adrenal glands is expanded with preserved adrenocortical zonation in the setting of pituitary ACTH-dependent pathogenesis. Endocrine pathologists can distinguish pituitary ACTH-dependent diffuse hyperplasia from ectopic ACTH-related diffuse hyperplasia, because in the latter the lack of cortical zonation and the presence of prominent zona reticularis–like compact cell hyperplasia result in lipid depletion pattern that gives an appearance that is often referred to “pink adrenal cortices”[9,27] (Fig. 91.13).

Bilateral nodular adrenocortical disease consists of two clinicopathologically significant forms, including (i) primary bilateral micronodular adrenocortical disease (Fig. 91.14) and (ii) primary bilateral macronodular adrenocortical disease (Fig. 91.15) (Table 91.2). At the morphological level, the distinction between the two forms is made based on the size of the cortical nodules. The micronodular forms consists of bilateral small cortical nodular proliferations (measuring less than 1.0 cm; usually ranging in size from 0.1 to 0.4 cm). The micronodules are enriched in compact (lipid-poor) cortical cells that can be pigmented in some cases (e.g., primary pigmented nodular adrenocortical disease) (Fig. 91.14). These are often located in the deep zona fasciculata and/or zona reticularis.[27,29,32] In contrast to micronodular forms of this disease, the macronodular forms are almost always

KEY POINTS

- Traditional diagnostic terminologies such as primary micronodular or macronodular hyperplasia of the adrenal cortex are no longer credible nomenclatures for primary bilateral nodular forms of adrenocortical disease, because those nodules are composed of genetically transformed adrenocortical cells; therefore, modern endocrine pathology practices classify them within the spectrum of adrenocortical neoplasia, and they are referred to as bilateral nodular adrenocortical disease.[27,29]

associated with significantly enlarged adrenals that are composed of several nodules that generously exceed 1.0 cm in size. The nodular proliferations are enriched in zona fasciculata–like clear cells, but variations in cytomorphological findings can occur[27,32] (Fig. 91.15).

The distinction of primary bilateral nodular adrenocortical disease in Cushing syndrome is of clinical significance given the increased frequency of germline susceptibility in both micro- and macronodular forms of this manifestation.[27,29,32,38]

Cortisol-producing adrenocortical adenomas are often unilateral and solitary neoplasms. At the morphological level, they can display cytomorphological heterogeneity; however, most are composed predominantly of zona fasciculata–like lipid rich cortical cells.[9,27,32,38] Oncocytic change and pigmentation can feature in some examples.[9,32] Cortisol-producing adenomas are distinguished from ***cortisol-producing adrenocortical carcinomas*** using multiparameter scoring systems and ancillary tools[9,27] that are summarized later in this chapter.

At a molecular level, autonomous overt cortisol excess from adrenal neoplasms (e.g., cortisol-producing adrenocortical adenomas and bilateral micronodular adrenocortical disease) are often associated with protein kinase A pathway activation,[27,29,32,38] which has also been correlated with high expression of the *DLK1-MEG3* miRNA in a recent genomic studies.[28] Most patients with primary macronodular adrenocortical disease have been linked to *ARMC5*-driven pathogenesis[27-29,41] that is also correlated with global CpG hypermethylation status and an ovarian gene expression signature in a recent genomic studies[28]; however, rare examples of primary macronodular disease have been reported in association with *MEN1*, *FH*, *GNAS*, *PDE8B*, *PDE11A*, and *PRKACA* variants.[29,32,38] Adrenocortical adenomas causing mild autonomous cortisol excess (subclinical Cushing syndrome) frequently harbor *CTNNB1* mutations and chromosome 17q losses.[28]

KEY POINTS

- From a functional histopathology approach, the nontumorous adrenocortical atrophy is a reliable sign of autonomous glucocorticoid secretion from an adrenocortical neoplasm in patients with no history of exogenous glucocorticoid administration (Fig. 91.16). Therefore, assessment of the nontumorous adrenal cortex is a critical responsibility of pathologists to prevent an Addisonian crisis, especially in patients with unrecognized mild autonomous glucocorticoid-secreting adrenocortical neoplasms.[9,27,40]

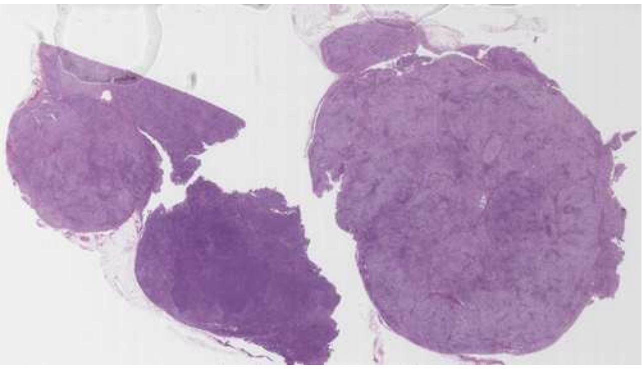


Fig. 91.15 Primary bilateral macronodular adrenocortical disease. Both adrenal glands are significantly enlarged due to multiple benign cortical nodules; each nodule exceeds 1 cm.

Pathological Correlates of Virilism/Feminization

KEY POINTS

- In both adult and pediatric age groups, the identification of virilism/feminization due to an adrenal mass requires exclusion of a functional adrenocortical carcinoma,[9,27] most of which tend to cosecrete glucocorticoids.

Rare examples of testosterone-producing adrenocortical adenomas have been reported in adults.[42] Unlike adults, this phenomenon has been relatively more frequently reported in pediatric adrenocortical adenomas,[43-46] including some examples, especially oncocytic adrenocortical adenomas, that feature synchronous virilization and adrenal Cushing syndrome.[44,45] Given the well-known challenges in the distinction of malignant behavior in adrenocortical neoplasms[47] and the fact that this functional group is more frequent in adrenocortical

TABLE 91.2 Clinicopathological Features of Primary Bilateral Nodular Adrenocortical Disease[27-29,32]

Pathology	Gene Alterations or Variants	Characteristics
Primary bilateral micronodular adrenal cortical disease - PPNAD - MAD	*PRKAR1A* *CNC2 locus/2p16* *PDE11A* *PDE8B* *PRKACA* *PRKACB* *ARCM5**	- Bilateral multiple compact cell small nodules ranging in size from 0.1 to 0.4 cm - Nodules are located in the zona fasciculata layer or in the intersection of zona fasciculata and reticularis layers - PPNAD is distinguished histologically from MAD based on the presence of pigmented micronodules - PPNAD may be associated with Carney complex - Bead-like appearance of micronodules on imaging studies - Internodular cortex shows atrophy in PPNAD - *PRKAR1A*-related pathogenesis is associated with an earlier age disease onset compared to *CNC2* locus-related pathogenesis - Paradoxical response on Liddle's test
Primary bilateral macronodular adrenal cortical disease	*ARMC5* *MEN1*** *FH*** *APC*** *PDE8*** *PDE11A*** *PRKAR1A*** *PRKACA***	- Marked bilateral enlargement of adrenals - Multiple clear cell–rich clonal nodules exceeding 1.0 cm - Childhood-onset disease shows more pronounced Cushing syndrome compared with adult-onset disease - *ARCM5*-related disease is more common in adults - *ARMC5*-related disease shows significantly more enlargement of adrenals with more nodularity and hypercortisolism compared with *ARMC5*–wild-type disease - Ovarian gene expression signature (*FOXL2, CYP19A1, PTHLH*) and global CpG hypermethylation

*Rare missense variant.

**Rare events; can be more frequent in childhood manifestations.

PPNAD, Primary pigmented micronodular adrenocortical disease; *MAD*, micronodular adrenocortical disease.

carcinomas,[48] one should still be careful when rendering a diagnosis of adenoma in this setting. The morphological criteria that help distinguish adrenocortical carcinoma and ancillary tools available to diagnosticians are discussed later in this chapter.

Pathological Correlates of Catecholamine Excess

KEY POINTS

- *Pheochromocytoma* (adrenal paraganglioma) is the most common pathological correlate of catecholamine excess from the adrenal gland.

At the morphological level, these tumors are no different than their extraadrenal counterparts, which are termed as paragangliomas. They display heterogenous cytology (e.g., conventional, clear cell, oncocytic) and growth patterns (e.g., Zellballen-like small nests, solid diffuse, trabecular, large confluent nests, cystic) (Figs. 91.17 and 91.18).

The term *composite pheochromocytoma* is applied to pheochromocytomas that are admixed with composite elements, including neuroblastoma, ganglioneuroblastoma, ganglioneuroma, and peripheral nerve sheath tumor (Fig. 91.19). Pathologists are required to determine the presence/absence of composite elements.

Adrenal medullary hyperplasia is restricted to diffuse expansion of the adrenal medulla that exceeds one third of the adrenal gland thickness or when adrenal medulla is identified in the tail of the adrenal gland.[49-51] In the past, microscopic adrenal medullary nodules measuring less than 1.0 cm were referred to as nodular medullary hyperplasia; however, the latter is no longer accepted. Any distinct microscopic nodules (<1.0 cm) in the adrenal medulla are now considered clonal disease; thus, the term "micropheochromocytoma" is applied[49-51] (Fig. 91.20).

The distinction of pheochromocytoma from adrenocortical pathologies is of clinical significance. This tumor family, paraganglioma and pheochromocytoma, has the highest germline susceptibility among all tumors.[50,51] The biochemical status of the tumor, the cytomorphology of the tumor and its surrounding stroma, the presence or absence of associated composite elements or adrenal medullary hyperplasia, and molecular immunohistochemistry (see below the section on diagnostic and predictive immunohistochemical biomarkers in adrenal pathology) help endocrine pathologists in the distinction of various forms of inherited pheochromocytomas.[51] For instance, *VHL*-related pheochromocytomas are often encapsulated and show clear cell change with stromal degeneration and increased microvasculature[50-53] (Fig. 91.21). Succinate dehydrogenase subunit (*SDHx*)-mutant tumors often show tumors cells with variable cytoplasmic vacuoles[50,51,54] (Fig. 91.22). Multifocal and/or bilateral pheochromocytomas are almost always a sign of germline disease, and the latter occurs more frequently in the setting of germline *RET* and *VHL* disease.[51,53] Composite elements

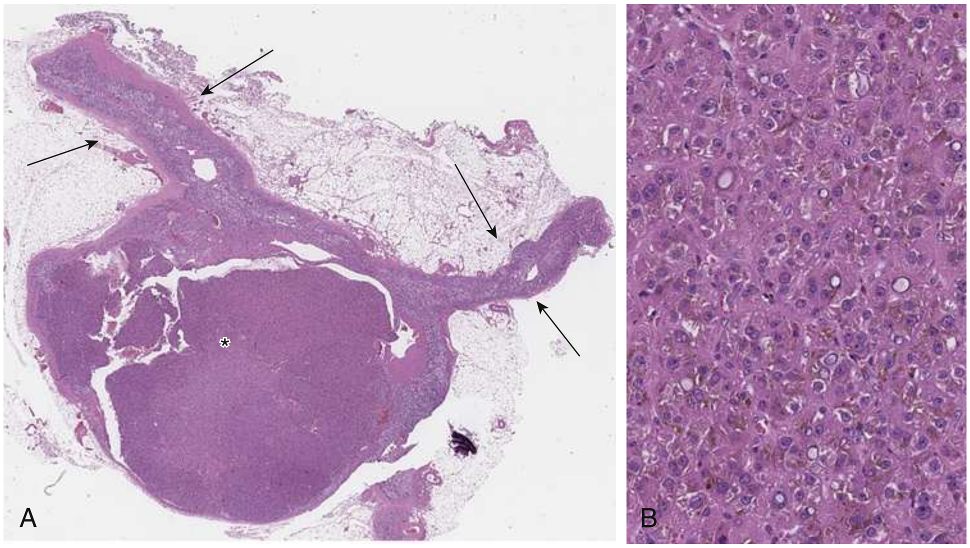


Fig. 91.16 **Cortisol-producing adrenocortical adenoma.** This composite photomicrograph illustrates a cortisol-producing adrenocortical adenoma (**A**, **B**). The tumor cells are variably pigmented (**B**). The asterisk indicates the tumor substance (**A**). There is marked atrophy of the nontumorous cortex *(arrows)*.

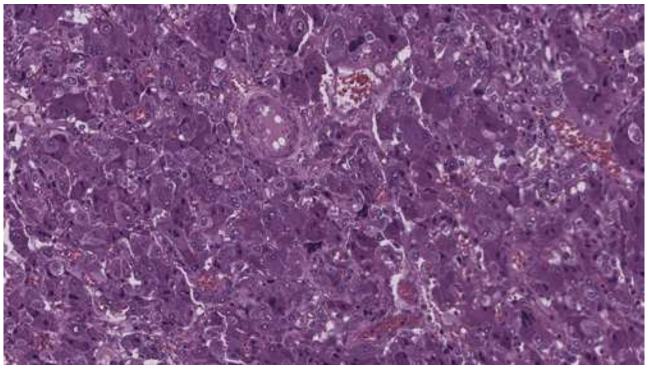


Fig. 91.17 **Pheochromocytoma.** This figure represents a conventional pheochromocytoma that is composed of tumor cells with basophilic granularity due to mature catecholamine granules.

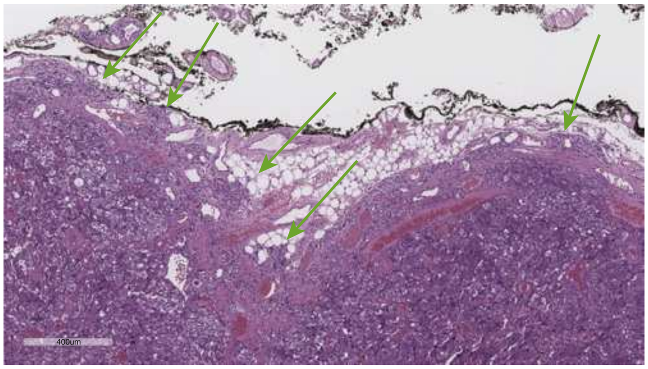


Fig. 91.18 **Invasive growth in pheochromocytoma.** Pheochromocytomas are not benign neoplasms, and they harbor lifelong metastatic potential. The status of invasive growth can be appreciated in a fraction of these tumors. The arrows illustrate multifocal microscopic tumor extension into the periadrenal adipose tissue.

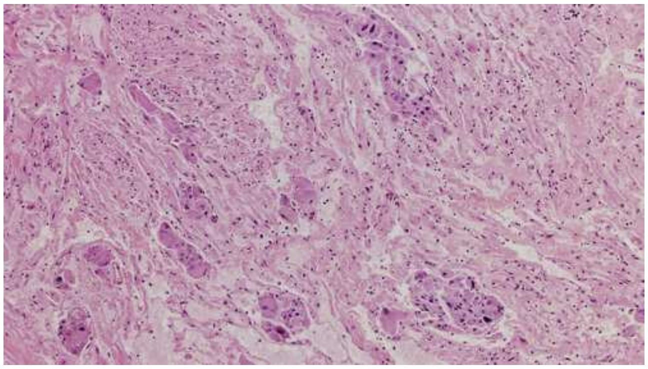


Fig. 91.19 Composite pheochromocytoma. This composite pheochromocytoma consists of ganglioneuroma and pheochromocytoma.

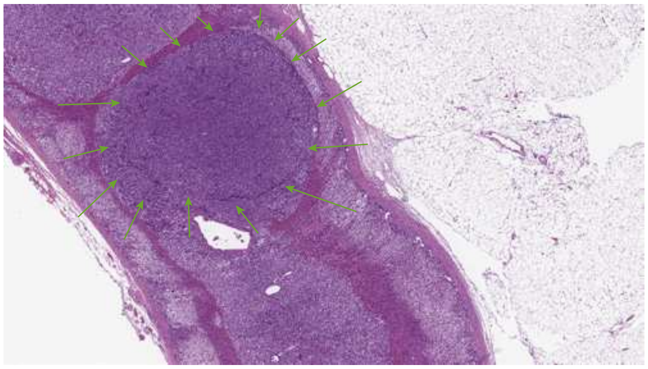


Fig. 91.20 Micropheochromocytoma. This photomicrograph illustrates a micropheochromocytoma *(arrows)* characterized by micronodular proliferation in the background of germline *RET*-driven diffuse adrenal medulla hyperplasia. Nodular hyperplasia of the adrenal medulla is not accepted terminology for distinct adrenal medullary micronodules.

can be seen in *NF1*- and *MAX*-related disease.[51 55] Adrenal medullary hyperplasia may be a harbinger of germline *RET, TMEM127, MAX,* and *SDHx*-related pathogenesis.[49-51,56-58]

KEY POINTS

- All paragangliomas and pheochromocytomas are malignant non-epithelial neuroendocrine neoplasms. These tumors have metastatic potential depending on their risk category.

Several multiparametric scoring systems (e.g., Pheochromocytoma of the Adrenal gland Scaled Score [PASS], Grading system for Adrenal Pheochromocytoma and Paraganglioma [GAPP], modified GAPP, and COmposite Pheochromocytoma/Paraganglioma Prognostic Score [COPPS] systems)[59,60-62] have been introduced in an attempt to predict the risk of metastasis. However, no single system is highly reproducible, and none have been endorsed by the World Health Organization (WHO) classification system of endocrine tumors. A recent metaanalysis series underscored the negative predictive value of low scores in the distinction of the low-risk disease.[63] Nevertheless, a dynamic risk stratification that combines several clinicopathologic parameters is more commonly adopted in modern endocrine oncology practices. The large tumor size (often >5.0 cm), immature secretory phenotype (e.g., dopaminergic, noradrenergic, or mixed dopaminergic and noradrenergic), tumor necrosis, increased proliferation (e.g., Ki67 labeling index >3%), and molecular characteristics (e.g., somatic *MAML3* fusions or *CSDE1* alterations related to altered Wnt-pathway clusters, *SDHB*-related pseudohypoxic pathway disease, select alterations in the kinase signaling pathway, and somatic *SETD2*, *ATRX,* and *TERT* alterations) are considered important risk factors in dynamic risk stratification.[60,64-68]

Although molecular testing for germline screening is now standard care in the management of patients with pheochromocytomas and paragangliomas, routine molecular testing for somatic alterations is still not the standard of endocrine pathology practice, even in most advanced practices.

PATHOLOGY IN THE DISTINCTION OF ADRENOCORTICAL CARCINOMA

Adrenocortical carcinoma is defined as a malignant epithelial tumor of adrenocortical cells.[69] These are rare tumors comprising less than 3% of endocrine tumors and affect all ages, but are more common in those 40 to 50 years of age. Adrenocortical carcinomas affect females more commonly than males. Approximately half of these tumors are functional, and functional tumors are more commonly seen in females.

Adrenocortical carcinomas are usually large tumors, often 10 cm or greater in size, but the size can vary greatly from 1 cm up to 30 cm or more. Size alone is a significant feature in adrenocortical neoplasms. Adrenocortical neoplasms weighing 50 g or more are of concern, and those weighing 100 g or more are of very significant concern by size in weight alone. However, adrenocortical carcinomas weighing less than 50 g can occur, and the size and weight alone are not sufficient to differentiate benign from malignant adrenocortical neoplasms.

Adrenocortical carcinomas are usually solid and may show necrosis, but cystic adrenocortical carcinomas have been reported.[70] Adrenocortical carcinomas can also be myxoid (Fig. 91.23), and the behavior of these tumors may be more difficult to predict than in conventional adrenocortical neoplasms (Fig. 91.24).[71] The behavior of oncocytic adrenocortical tumors may also be difficult to predict by conventional classification; thus, a separate classification (the Lin–Weiss–Biscgelia system) is used for these tumors.[72,73] A separate classification system is also utilized in diagnosis of malignancy in adrenocortical neoplasms in children.[74]

Distinction of adrenocortical adenoma from carcinoma is generally straightforward (Figs. 91.23, 91.24, and 91.25). However, some cases can be extraordinarily difficult, as reflected in the numerous classification systems that have been proposed for the diagnosis of adrenocortical carcinoma.[75-78] In 1984, Dr. Weiss and colleagues evaluated 43 adult adrenocortical neoplasms with 5 years of follow-up for nine histologic parameters.[76] Eighteen of 19 tumors with a score of 4 or more were associated with recurrence or metastasis. All 24 tumors with a score of 2 or less were associated with benign outcome. They found no single feature that was diagnostic of malignancy. The most significant features were mitotic rate of greater than five per 50 high-power fields (10 mm^2), atypical mitoses, and vascular invasion (venous angioinvasion). In 1989, the Weiss score was lowered from 4 to 3, as a patient with a score of 3 had recurrent disease and died of disease.[77] Weiss and colleagues found that only mitotic rate had a significant association with outcome. Tumors with mitotic rate of greater than 20 per 50 high-power fields were regarded as high-grade and associated with a median survival of 14 months. Tumors with a mitotic rate of less than or equal to 20 per

Weiss system*

1. Fuhrman nuclear grade III or IV
2. Mitotic activity >5 per 50 high-power fields
3. Atypical mitotic figures
4. Clear cells ≤25% of the tumor volume
5. Diffuse architecture >30% of the tumor volume
6. Tumor necrosis
7. Venous angioinvasion
8. Lymphatic (sinusoidal) invasion
9. Capsular invasion

**Malignant: Score ≥3 out of 9 Applied only to nononcocytic and adult-onset tumors*

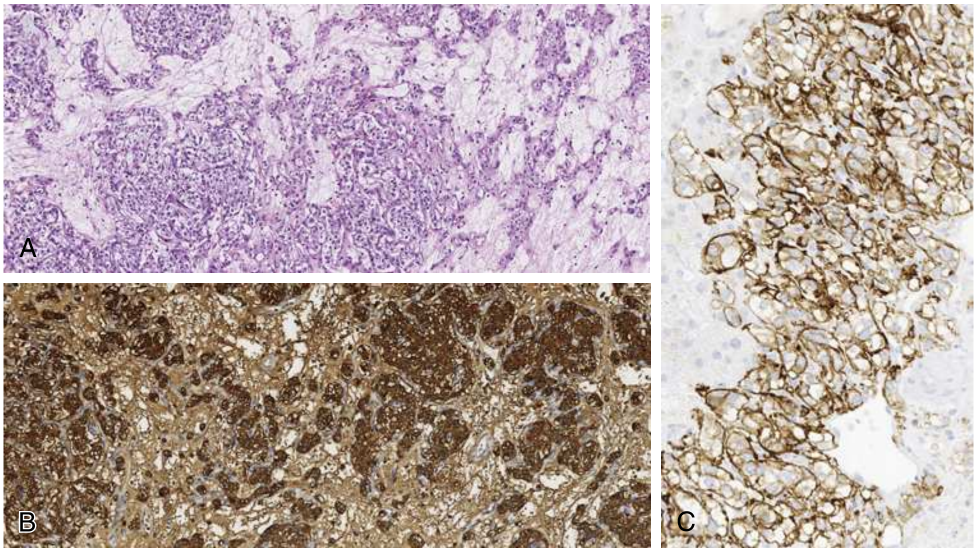


Fig. 91.21 ***VHL*-related pheochromocytoma.** VHL-related pheochromocytomas are often encapsulated tumors enriched in clear cells (**A**). Positivity for neuroendocrine markers (not illustrated herein) and tyrosine hydroxylase (**B**) distinguishes this tumor from a clear cell–rich adrenocortical neoplasm. Membranous carbonic anhydrase IX (**C**) expression is a feature of *VHL*-related pathogenesis.

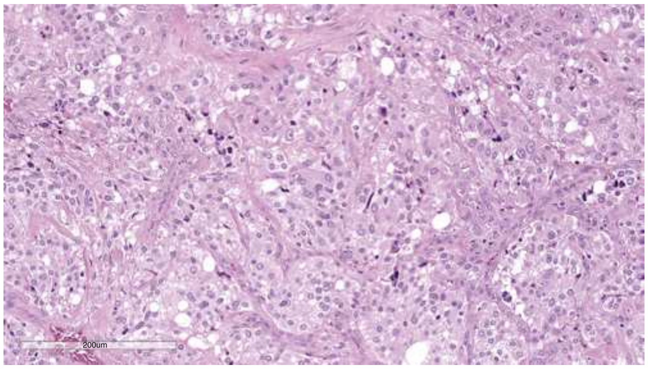


Fig. 91.22 **Morphology of *SDHx*-related pheochromocytoma.** *SDHx*-related tumors are composed of tumor cells with variable cytoplasmic eosinophilia and cytoplasmic vacuoles.

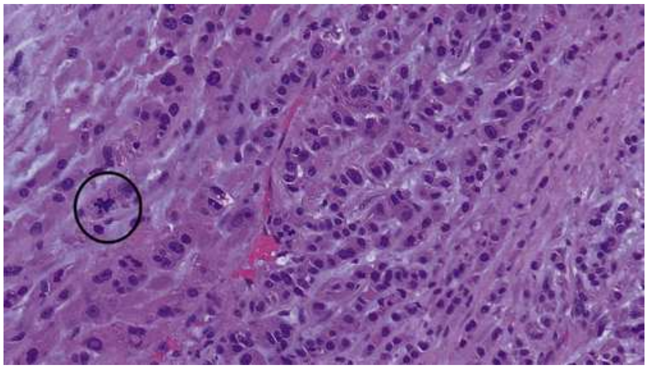


Fig. 91.23 **Myxoid change in adrenocortical carcinoma.** Myxoid change can obscure signs of malignancy; however, a detailed morphological assessment often helps in the distinction of carcinoma. Adrenocortical carcinomas are driven by proliferation. The circled area represents an atypical mitotic figure. This tumor had increased mitotic activity (>5 per 50 high-power fields), including several atypical mitoses.

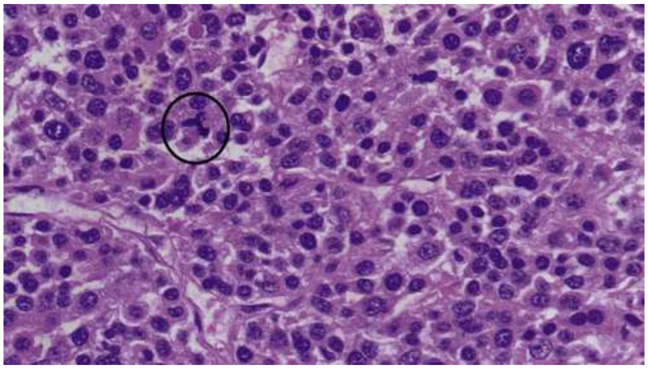


Fig. 91.24 **Conventional adrenocortical carcinoma.** Most adrenocortical carcinomas are composed of lipid poor cells that are arranged in sheets. A tripolar atypical mitotic figure (circled area) is illustrated in this photomicrograph.

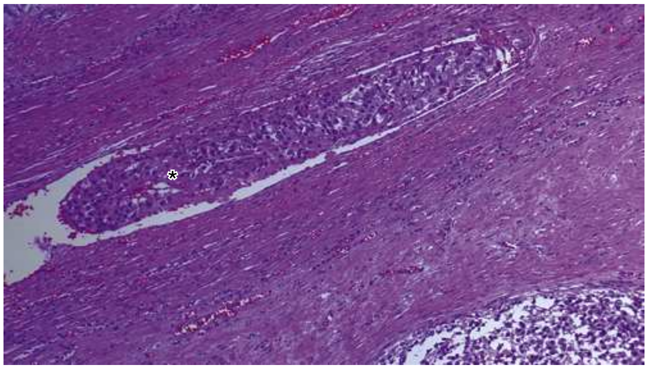


Fig. 91.25 **Vascular invasion in adrenocortical carcinoma.** The identification of a *bona fide* vascular invasion (venous angioinvasion) is not only a sign of malignancy but is also a prognostic factor for metastatic spread. The presence of intravascular tumor cells admixed with fibrin thrombus is illustrated *(asterisk)*.

50 high-power fields were regarded as low-grade, with median survival of 58 months.

Numerous other systems have been suggested previously and subsequently, but the system proposed by Weiss in 1984 and its subsequent modification in 1989 with three or more of the nine features as malignant remains the gold standard today.[77,79] In 2002, another modified Weiss system, or "Weiss revisited" system, suggested by matching functional status of 24 adrenocortical carcinomas with distant metastasis, gross local invasion, or recurrence with 25 benign tumors.[80] This is a modified system based on the most reliable criteria of the prior Weiss studies from 1984 and 1989, including mitotic rate of greater than five per 50 high-power fields, atypical mitotic figures, clear cells in less than or equal to 25% of the tumor, necrosis, and capsular invasion. A score of 3 or more out 7 in this study was considered diagnostic of malignancy. Still, the 1989 Weiss system remains the diagnostic standard for adult nononcocytic adrenocortical neoplasms.

Modified Weiss System*

- mitotic rate (score: 2)
- compact cytoplasm (score: 2)
- atypical mitoses (score: 1)
- necrosis (score: 1)
- capsular invasion (score: 1)

*Malignant: score ≥3 out 7

The "reticulin algorithm" evaluated disruption of the reticular network (highlighted by reticulin staining) (Fig. 91.26) and one or more of the following criteria (mitoses greater than five per 50 high-power fields, necrosis, or vascular invasion) in 92 carcinomas in 47 adenomas and found sensitivity and specificity of 100%.[81] Volante and colleagues found tumor stage III or IV and mitoses greater than nine per 50 high-power fields had a strong adverse impact on disease-free and overall survival, and were able to divide the tumors into three risk groups. A subsequent multicenter study in 2013 by Duregon and colleagues validated the reticulin algorithm in a study of 184 adrenocortical carcinomas and 61 adenomas.[82] The reticulin algorithm may be helpful in the evaluation of myxoid and oncocytic adrenocortical tumors. In a study by Mete and colleagues, loss of the reticulin framework, rather than qualitative alterations of the reticulin framework, was found to differ between adrenocortical carcinoma and adenomas.[83]

Reticulin Algorithm

Altered reticulin framework in association with one of the following parameters:

- Vascular invasion
- Necrosis
- Mitoses exceeding 5 per 50 high-power fields

The Helsinki score was proposed in 2015 for predicting metastasis in adrenocortical carcinomas. Adrenocortical neoplasms from 177 adults with 5 years of follow-up were evaluated with the Weiss score, Weiss revisited score, and proliferation index by computer-assisted image analysis.[84] A weighted calculation of three parameters (mitotic activity, necrosis, and Ki67-labeling index) had 100% sensitivity and 99.4% specificity for diagnosing metastatic adrenocortical carcinoma, with a cutoff of 8.5. In a validation study the following year of 225 adrenocortical carcinomas, the Helsinki and Weiss scores were found to be predictive of poor prognosis.[85] Reports suggest that the Helsinki scoring system may be helpful in the evaluation of myxoid and oncocytic adrenocortical tumors.[85,86]

Helsinki Scoring System*

- Mitoses >5 per 50 high-power fields: score 3
- Necrosis: score 5
- Ki67 labeling index in hot spots

*Malignant: score ≥8.5

Myxoid adrenocortical neoplasms are uncommon, with most cases published as small series or case reports. These tumors are often functional. In a study from Mayo Clinic of 14 cases of adrenocortical neoplasms with a myxoid component comprising 10% to 95% of the tumor, all six characterized as adenomas were associated with benign behavior, while four of the eight carcinomas were associated with a fatal outcome, with two alive with disease and one alive without

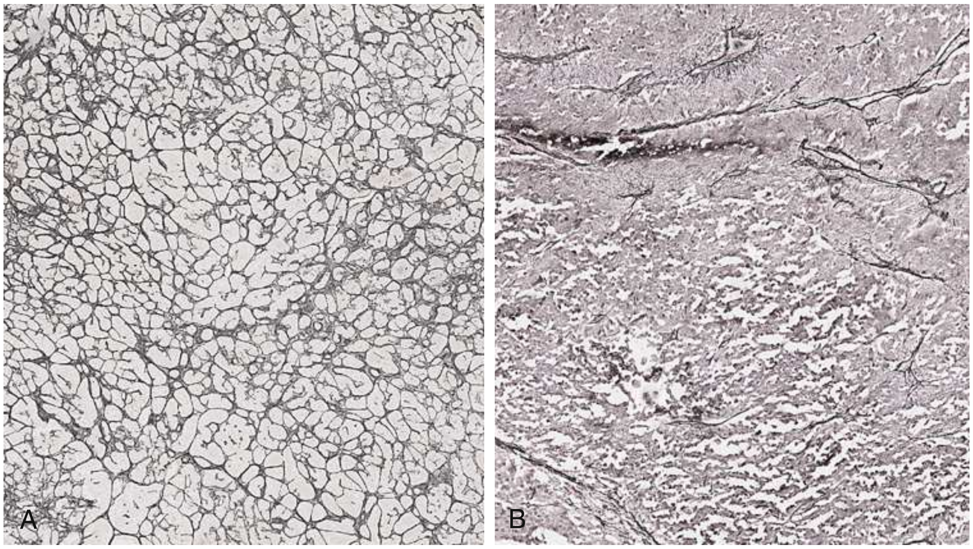


Fig. 91.26 Reticulin histochemistry in adrenocortical tumors. Adrenocortical adenomas are associated with retained acinar reticulin framework (**A**), whereas most adrenocortical carcinomas show loss of reticulin framework (**B**). The latter forms the basis of the reticulin algorithm. However, the identification of altered reticulin framework is often coupled with one of the three criteria proposed on this algorithm.

disease.[87] The immunohistochemical findings are similar to other adrenocortical neoplasms. A subsequent study from Italy by Papotti et al. evaluated four conventional adrenocortical carcinomas with focal myxoid change (5%–20% of the tumor), and found one of the four patients died of disease, and three were alive without disease at the end of the study.[71] They also evaluated 10 myxoid adrenocortical carcinomas with small cells, mild atypia, cords, and microcysts, and six of the 10 died of disease, one was alive with disease, and three were alive with no disease at follow-up. The most striking part of this study is that one of the two tumors with a Weiss score of 1 had a fatal outcome. Thus, the behavior of myxoid adrenocortical tumors is difficult to predict with standard morphological criteria. However, myxoid change does not make an adrenocortical neoplasm as malignant, but the behavior is difficult to predict with standard diagnostic systems.

Oncocytic adrenocortical tumors may be overdiagnosed as malignant by applying the standard Weiss systems. In 1998 Lin et al. evaluated seven oncocytic adrenocortical tumors (5–13.5 cm), all showing nuclear atypia and increased Ki67 proliferative index, but none had invasion or metastasis, and all behaved in a benign manner.[73] In 2004, Bisceglia and colleagues evaluated 10 oncocytic adrenocortical tumors and found that two criteria (clear cells less than or equal to 25% of the tumor and high nuclear grade) were present in all of the oncocytic tumors studied, and diffuse architecture was present in nine of 10.[72] They proposed an algorithm including the remaining six Weiss parameters and an additional criterion of "large size and huge weight" (greater than 10 cm and/or greater than 200 g). The current system used to evaluate oncocytic adrenocortical carcinoma is referred to as the "Lin–Weiss–Bisceglia system," which is comprised of major criteria and minor criteria. The presence of one or more major criteria (high mitotic rate, atypical mitoses, or vascular invasion) is diagnostic of malignancy. One to four minor criteria (large size and huge weight, necrosis, capsular invasion, or sinusoidal invasion) indicates a tumor of uncertain malignant potential. Tumors with no major or minor criteria are regarded as benign.[72] In 2011, a subsequent study of 13 oncocytic adrenocortical neoplasms and all published oncocytic adrenocortical neoplasms confirm the Lin–Weiss–Bisceglia system.[88] They also found that 30% of these tumors were functional, and these tumors are associated with median survival of 58 months (which is comparatively longer than for conventional adrenocortical carcinomas). A recent multicenter retrospective study from France matching oncocytic adrenocortical tumors with conventional adrenocortical carcinomas for age, sex, European Network for the Study of Adrenal Tumors (ENSAT) stage, and resection status also showed oncocytic adrenocortical carcinomas to be associated with better overall survival than conventional adrenocortical carcinomas.[86] Similar to all adrenocortical neoplasms, extensive sampling of oncocytic adrenocortical neoplasms is required so that the correct criteria are utilized, as the behavior does vary with the amount of oncocytic change. What is regarded as a pure oncocytic tumor requires greater than 90% of the tumor be oncocytic, requiring extensive sampling of an adrenocortical neoplasm. Mixed oncocytic adrenocortical tumors are characterized as clear cell component of 10% to 50% of cells, and in an adrenocortical tumor with focal oncocytic change the oncocytic component comprises less than 50% of the tumor.[27]

Lin–Weiss–Bisceglia Criteria for Oncocytic Adrenocortical Tumors*

Major Criteria:

- Mitoses >5 per 50 high-power fields
- Atypical mitoses
- Vascular invasion

Minor Criteria:

- Large tumor size (>10 cm) and weight >200 g
- Capsular invasion
- Necrosis
- Sinusoidal invasion

Malignant: at least one major criterion
Uncertain malignant potential: at least one minor criterion
Benign: absence of major and minor criteria

Sarcomatoid adrenocortical carcinomas are extremely rare and highly aggressive tumors that occur over a wide age range and are generally nonfunctioning.[89] Sarcomatoid adrenocortical carcinomas may have a component of more conventional adrenocortical carcinoma or can be completely sarcomatoid in appearance. These tumors must be differentiated from retroperitoneal sarcomas. Most sarcomatoid adrenocortical carcinomas exist in the literature as case reports. A study of six cases, three of which had been previously reported, evaluated the pathologic and immunohistochemical features of these tumors and had targeted next-generation sequencing analysis performed.[89] Among the six sarcomatoid adrenocortical carcinomas, the authors identified 16 morphologically distinct tumor components, eight of which were epithelial, and eight of which were sarcomatoid.[89] Thus, the morphologic heterogeneity of these tumors is quite remarkable and may account for some of the difficulty in differentiating these tumors from retroperitoneal sarcoma or other tumors. In the series above, patients ranged in age from 23 to 79 years, with a mean age of approximately 54 years. The tumors were large in size, averaging 14 cm (range: 6.5–24 cm). Some of the tumors had spindle cell components, with some having foci of osteosarcomatous differentiation, rhabdomyosarcomatous differentiation, or even PNET-like areas. Purely mesenchymal components such as osteosarcomatous or rhabdomyosarcomatous were less common. Keratin (AE1/AE3) showed focal positivity in the epithelial component of one tumor and in the sarcomatoid components of two tumors, but was otherwise negative. SF1 was present in epithelial component of all cases but negative in the sarcomatoid areas. Interestingly, β-catenin showed aberrant nuclear staining in the sarcomatoid component of five of the six cases. They also found *TP53* and *CTNNB1* mutations in both epithelial and sarcomatoid components, which they note supports a common clonal origin in at least a subset of these tumors.[89]

Wieneke Criteria for Pediatric Adrenocortical Tumors*

1. Tumor weight >400 g
2. Tumor size >10.5 cm
3. Mitoses >15 per 20 high-power fields
4. Atypical mitotic figures
5. Vena cava invasion
6. Capsular invasion
7. Vascular invasion
8. Extension into adrenal soft tissue or adjacent organs
9. Necrosis

Malignant/poor outcome: *score ≥4*
Uncertain malignant potential: *score 3*
Benign behavior: *score ≤2*

Pediatric adrenocortical neoplasms are not evaluated with the same classification systems as adult adrenocortical neoplasms. Pediatric tumors may have features that in an adult adrenocortical tumor would be very worrisome for malignancy. Thus, the Wieneke system is often utilized for these challenging tumors.[74]

DIAGNOSTIC AND PREDICTIVE IMMUNOHISTOCHEMICAL BIOMARKERS IN ADRENAL PATHOLOGY

Confirmation of Adrenocortical and Medullary Origin

KEY POINTS

- Biomarkers are frequently used in the distinction of nonfunctional adrenocortical neoplasms from other tumors. As adrenocortical carcinoma is an exceptionally rare tumor, a frankly malignant neoplasm in the adrenal gland would more likely be a metastasis from another site than a primary adrenocortical carcinoma. Immunohistochemical studies to characterize the tumor as a primary adrenocortical tumor are critical.

Adrenocortical neoplasms will often show variable staining for keratins; however, significant staining for keratin is not always identified, rather it is often focal and weak. SF1 is the most reliable biomarker that should be used to confirm the adrenocortical origin[31,90] (Fig. 91.27). However, several other somewhat less specific immunohistochemical stains (melan-A, synaptophysin, alpha-inhibin, and calretinin) are often used in traditional practices.[91,92] Nevertheless, one should recognize that synaptophysin[31] and alpha-inhibin[93] can be expressed in pheochromocytomas.

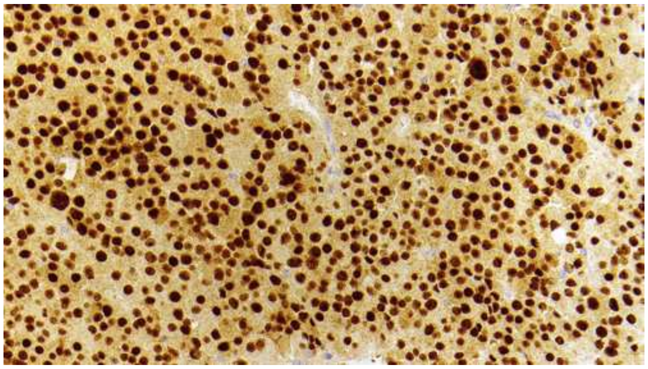


Fig. 91.27 **Confirmation of the adrenocortical origin.** Steroidogenic factor 1 is the most reliable biomarker that should be used to confirm the adrenocortical origin.

Pheochromocytomas are often positive for neuroendocrine differentiation markers (e.g., chromogranin-A, synaptophysin, and INSM1) and negative for keratins. When the distinction from other neuroendocrine neoplasms is indicated, the application of tyrosine hydroxylase (rate-limiting enzyme in catecholamine biosynthesis) or dopamine β-hydroxylase, along with GATA3 (variable), helps diagnosticians confirm the diagnosis of pheochromocytoma.[94-96] Most pheochromocytomas also show variable amounts of intratumoral sustentacular cells that are typically positive for SOX10 and S100. Evidence also suggests that sustentacular cells are not neoplastic in these tumors.[97] This may be helpful in distinguishing multifocal primary paragangliomas from metastasis in some anatomic sites such as liver and lung, where both primary and metastatic disease can occur.

Confirmation of Functional Sites in Primary Aldosteronism

Commercially available antibodies against steroidogenic enzymes have gained interest in the workup of functional sites in adrenalectomy specimens. The use of CYP11B2 immunohistochemistry adds great value to the workup of adrenalectomy specimens with primary aldosteronism.[31,34] Although assessment for this biomarker is generally not routinely performed in pathology laboratories, the use of this biomarker may be especially useful when dealing with multiple adrenal nodules or when no nodules are noted in the gross examination.

Diagnostic and Predictive Biomarkers of Adrenocortical Carcinoma

Numerous biomarkers have been evaluated diagnostically and prognostically in adrenal pathology.[31,83] Proliferation markers such as Ki67 (MIB1), proteins (IGF2, CDKN1C/p57, H19) associated with genes on chromosome 11p15 (which is also involved in Beckwith–Wiedemann syndrome); p53, for which *TP53* (17q13) mutations occur in Li–Fraumeni syndrome, and somatic *TP53* mutations and allelic loss of 17p13 may be seen in sporadic adrenocortical carcinomas; and alterations involving the Wnt/β-catenin-signaling pathway, among others.[9,29,31,83]

Advances in our understanding of the molecular pathogenesis of adrenocortical carcinoma underscored the importance of p53 and β-catenin as prognostic biomarkers, because most carcinomas with adverse molecular clusters tend to display nuclear β-catenin expression and/or abnormal p53 staining (overexpression or global loss) (Fig. 91.28). For this reason, endocrine pathologists are increasingly using

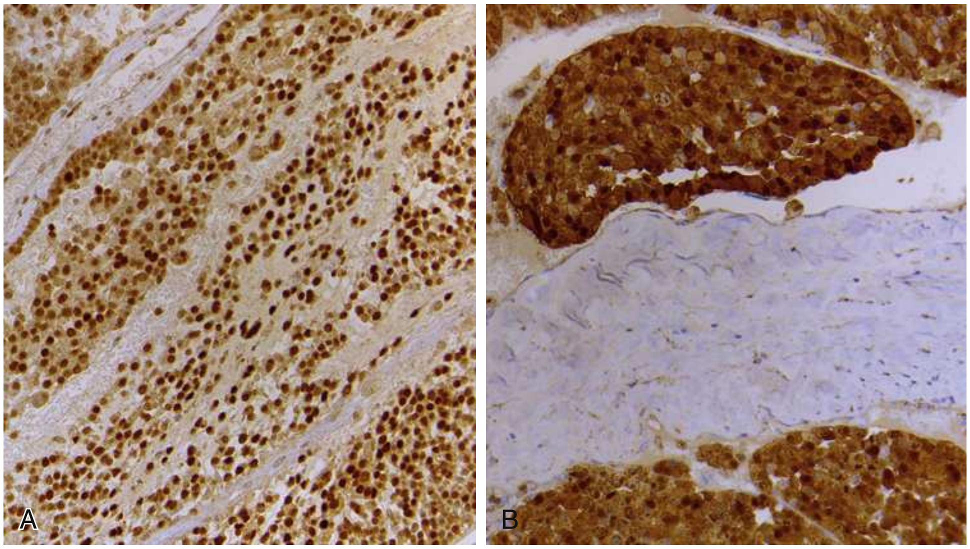


Fig. 91.28 **p53 overexpression and nuclear β-catenin expression.** p53 overexpression is seen in a subset of adrenocortical carcinomas that are often associated with high-grade proliferative disease (**A**). Diffuse nuclear β-catenin is also seen in some adrenocortical carcinomas, particularly in adverse molecular clusters (**B**).

both p53 and β-catenin immunohistochemistry in the workup of adrenocortical carcinomas.[27,29,31,83]

KEY POINTS

- IGF2 immunohistochemistry stands out as an important diagnostic biomarker of adrenocortical carcinomas.

Mete et al. demonstrated IGF2 immunohistochemistry optimized to detect paranuclear dot-like reactivity in adrenocortical carcinomas in adults[83] (Fig. 91.29).

Proliferative activity in adrenocortical neoplasms is a continually important theme. The mitotic rate is present in numerous diagnostic algorithms and is considered by some as the most important criterion of malignancy.[98] Proliferative activity as measured with a Ki67 labeling index of greater than 5% has been suggested as a sensitive and specific marker for malignancy[83,99,100] (Fig. 91.30). A combination of Ki67 proliferative index and IGF2 expression has also been proposed to be useful in separating adrenocortical carcinomas from adenomas.[83,100,101] The widely adopted Ki67 prognostic groups of less than 10%, 10% to 19%, and greater than 20% are also suggested by the WHO classification.[69] Given the critical value of Ki67 labeling index, eyeballing is less accurate. The use of pathologist-driven automated image analysis nuclear algorithms or manual counting is helpful to more reliably determine the Ki67 proliferation index that can also be reproduced by other users[27,106,107] (Fig. 91.31).

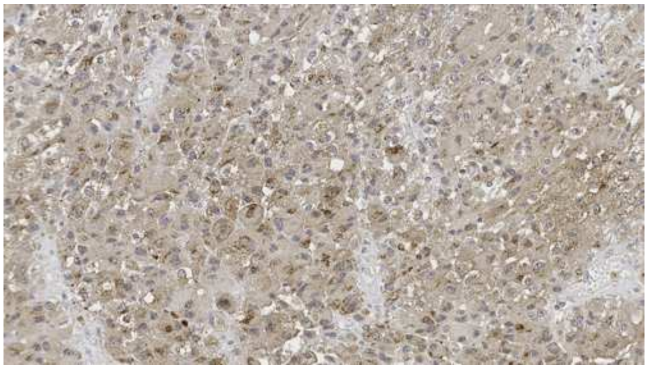


Fig. 91.29 IGF2 immunohistochemistry in the distinction of adrenocortical carcinoma. Irrespective of tumor cytomorphology, proliferative grade, or functional status of the tumor, paranuclear IGF2 reactivity is a reliable diagnostic biomarker of adrenocortical carcinoma.

KEY POINTS

- Mitotic rate and proliferative activity as measured by Ki67 have been used in the grading and prognosis of adrenocortical carcinomas over time.[81,83,102-105]

SF1 immunohistochemistry not only confirms the cortical origin of an adrenal mass, but may also have prognostic impact.[31,90] It has been reported that high expression of SF1 (N1665 antibody) positively correlated with high mitotic count, high Ki67 index, high stage, and decreased survival in a study of 75 adrenocortical carcinomas that also included myxoid in oncocytic variants.[90]

Immunohistochemistry for MMR proteins (MLH1, MSH2, MSH6, and PMS2) has been frequently used by endocrine pathologists,[31] as adrenocortical carcinoma is one of the extracolonic manifestations of Lynch syndrome[108,109] and may help determine eligibility for treatment options in affected patients.

Molecular Immunohistochemistry in Pheochromocytoma

In the evaluation of pheochromocytomas and paragangliomas, there are several biomarkers that can help predict disease pathogenesis and germline susceptibility. Commercially available antibodies against SDHB have revolutionized the screening process for more than a decade in endocrine pathology practice.[110] Given the genotype–phenotype correlations, pheochromocytomas lacking a mature secretory phenotype (absence of adrenergic secretion) is usually assessed for SDHB immunohistochemistry.[51] Germline DNA from patients with SDH-deficient tumors should be screened for all *SDHx* (*SDHA*, *SDHB*, *SDHC*, *SDHD*, and *SDHF2*) pathogenic variants, because a small fraction of SDH-deficient tumors may also be related to somatic/epigenetic alterations.[111] Antibodies against SDHA (e.g., loss of SDHA expression suggests SDHA-related pathogenesis) have also shown promise and are being used in clinical practice.[110,111]

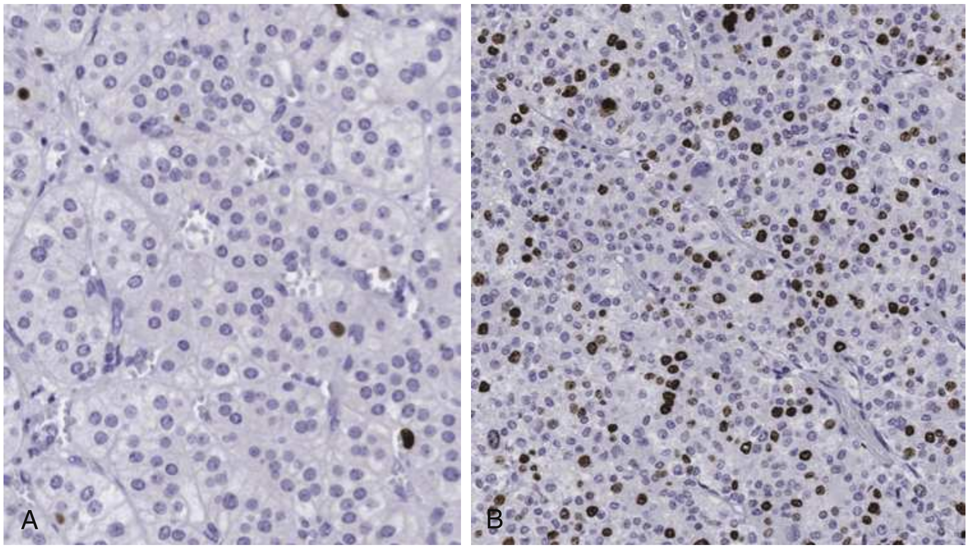


Fig. 91.30 Ki67 proliferation index in the distinction of adrenocortical carcinoma. Most adrenocortical adenomas have a Ki67 labeling index of less than 5% (**A**), whereas adrenocortical carcinomas tend to show higher labeling indices (**B**).

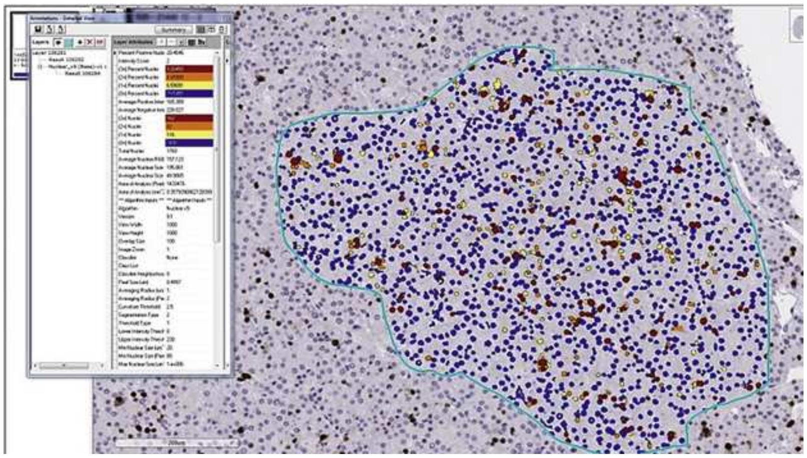


Fig. 91.31 Manual or automated counting of Ki67 labeling index. The current standard of practices requires meticulous assessment of the Ki67 labeling index in adrenocortical carcinomas. Eyeballing is no longer recommended. This photomicrograph illustrates a pathologist-driven automated image analysis nuclear algorithm in a patient with adrenocortical carcinoma.

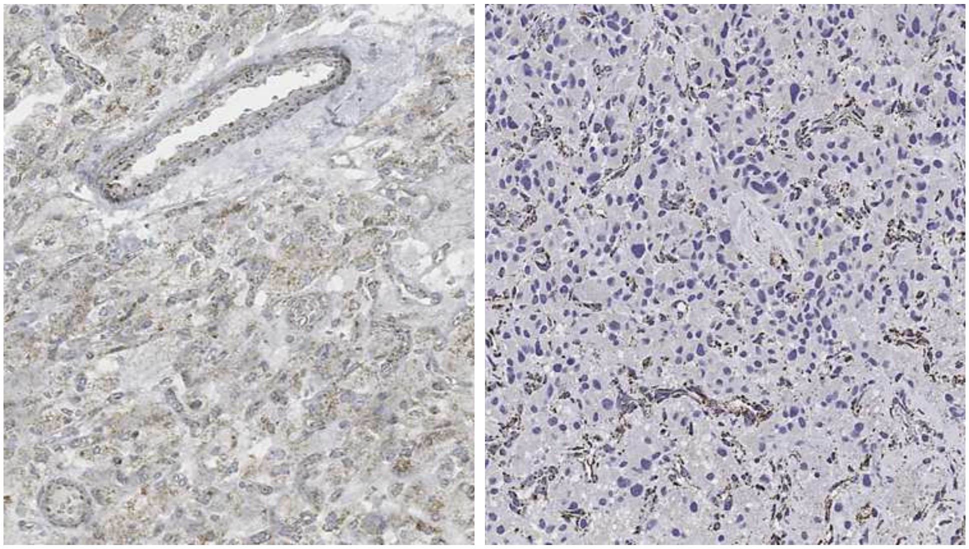


Fig. 91.32 SDHB immunohistochemistry in pheochromocytoma. This composite photomicrograph illustrates two pheochromocytomas. The tumor on the left shows normal SDHB expression, which argues against *SDHx*-related pathogenesis. The tumor on the right tumor displays loss of cytoplasmic granular SDHB expression in the tumor, while the nontumorous endothelial cells still display preserved granular reactivity. This finding is consistent with a SDH-deficient pheochromocytoma and requires germline testing of all *SDH* genes.

KEY POINTS

- The term SDH-deficient pheochromocytoma is applied to a pheochromocytoma that shows loss of cytoplasmic granular SDHB reactivity, while nontumorous elements preserve SDHB reactivity (Fig. 91.32).

KEY POINTS

- SDHB immunohistochemistry is also of clinical significance during the interpretation of *SDHx* variants of uncertain significance (VUS), because loss of SDHB expression by immunohistochemistry would favor a pathogenic process in the setting of VUS.[51]

Other molecular immunohistochemistry tools available to pathologists include antibodies against MAX and FH/2-SC that correlates with *MAX*- and *FH*-related pathogenesis.[112,113] *VHL*-related tumors often show membranous carbonic anhydrase IX staining[93,114,115] (Fig. 91.21). Recently, alpha-inhibin immunohistochemistry has been shown to predict pseudohypoxia-driven pathogenesis, including *SDHx*- and *VHL*-related pheochromocytomas[93] (Fig. 91.33).

SYNOPTIC REPORTING IN ADRENOCORTICAL CARCINOMA AND PHEOCHROMOCYTOMA

The synoptic templates of the College of American Pathologists and International Collaboration on Cancer Reporting (ICCR)

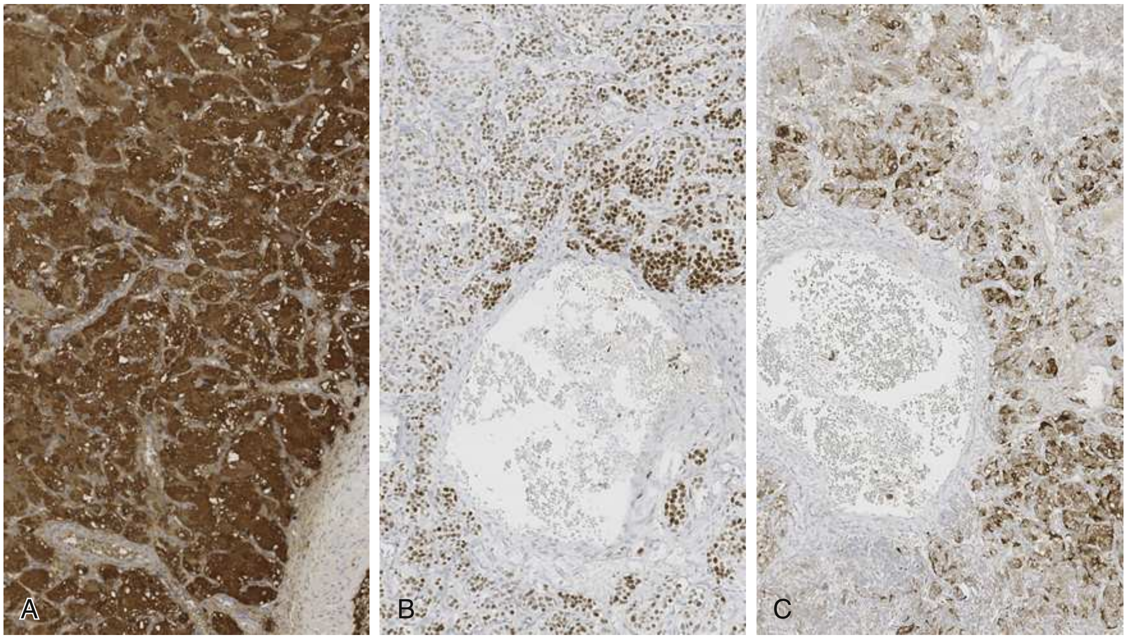


Fig. 91.33 **Alpha-inhibin expands biomarkers of pseudohypoxia-driven pathogenesis.** This composite photomicrograph illustrates a case of SDH-deficient pheochromocytoma. By immunohistochemistry, the tumor is positive for tyrosine hydroxylase (**A**) and GATA3 (**B**), confirming the diagnosis. SDHB immunohistochemistry is not illustrated. This tumor is also variably positive for alpha-inhibin (**C**). It is important to recognize this pitfall, as inhibin expression has been regarded as a marker of adrenal cortex, but recent evidence clearly confirms that this marker cannot be used in this distinction.

have been available for unified reporting of adrenocortical carcinomas.[107,116] The synoptic reports often integrate the pathological tumor-node-metastasis (TNM) data at the time of diagnosis.

European Network for the Study of Adrenal Tumors (ENSAT)

Stage I: Tumor confined to adrenal gland and <5 cm in diameter
Stage II: Tumor confined to adrenal gland and >5 cm in diameter
Stage III: Tumor involves surrounding tissue (not confined to the adrenal gland) or involvement of regional lymph node(s) or regional veins
Stage IV: Distant metastases

KEY POINTS

- Essential parameters that are reported in pheochromocytomas include (i) tumor size, (ii) presence/absence of extraadrenal invasion, (iii) presence/absence of composite elements, (iv) vascular invasion, (v) lymphatic invasion, (vi) mitotic activity (mitoses per 2 mm^2 based on 10 mm^2 from hot spots), (vii) tumor necrosis, (viii) status of resection margins, (ix) status of adrenal medullary hyperplasia, (x) Ki67 proliferation index, (xi) histologically confirmed nodal metastasis, and (xii) molecular immunohistochemistry (at least SDHB immunohistochemistry in all tumors with no mature secretory phenotype). It is desirable to integrate the findings on biochemical and functional imaging studies into a pathological report.

Until the eight edition of the American Joint Committee on Cancer (AJCC) TNM staging system, there was no formal TNM staging for pheochromocytomas. The current AJCC TNM staging introduced three distinct pT stages based on size (the cut-off of 5 cm) and presence of extra-adrenal invasion into surrounding tissues (see https://www.facs.org/quality-programs/cancer/ajcc for AJCC staging information). Because pheochromocytomas/paragangliomas are neuroendocrine neoplasms with metastatic potential, these tumors are now staged and are required to be reported using synoptic templates that are available to pathologists. The very first multidisciplinary international expert consensus synoptic template for pheochromocytomas and paragangliomas was introduced back in 2014[110]; and this was followed by additional reporting guides including but not limited to the ICCR template.[117]

KEY POINTS

- In adrenalectomy specimens, essential histological parameters that should be covered in adrenocortical carcinomas include (i) tumor size, (ii) presence of extraadrenal invasion, (iii) gross invasion into a large vein, (iv) microscopic vascular invasion, (v) microscopic lymphatic invasion, (vi) mitotic activity, reported as per number of mitoses per 50 high-power fields, (vii) atypical mitoses, (viii)tumor mitotic grade (low or high grade, based on the cutoff of 20 per 50 high-power fields), (ix) tumor necrosis, (x) Ki67 proliferation index, (xi) status of resection margins, and (xii) status of nodal disease (if lymph nodes are excised).

REFERENCES

1. Xing Y, Lerario AM, Rainey W, et al. Development of adrenal cortex zonation. *Endocrinol Metab Clin North Am*. 2015;44:243–274.
2. Kempná P, Flück CE. Adrenal gland development and defects. *Best Pract Res Clin Endocrinol Metab*. 2008;22:77–93.
3. Pihlajoki M, Färkkilä A, Soini T, et al. GATA factors in endocrine neoplasia. *Mol Cell Endocrinol*. 2016;421:2–17.
4. Kastriti ME, Kameneva P, Adameyko I. Stem cells, evolutionary aspects and pathology of the adrenal medulla: a new developmental paradigm. *Mol Cell Endocrinol*. 2020;518:110998.
5. Kastriti ME, Kameneva P, Kamenev D, et al. Schwann cell precursors generate the majority of chromaffin cells in Zuckerkandl organ and some sympathetic neurons in paraganglia. *Front Mol Neurosci*. 2019;12:6.
6. Folligan K, Roume J, Razavi F, et al. [L'hypoplasie surrénale congénitale de type anencéphalique: deux cas avec anomalies hypophysaires, sans mutation de DAX-1 et SF-1 et revue de la littérature]. *Morphologie*. 2011;95:26–33.
7. Biason-Lauber A, Schoenle EJ. Apparently normal ovarian differentiation in a prepubertal girl with transcriptionally inactive steroidogenic factor 1 (NR5A1/SF-1) and adrenocortical insufficiency. *Am J Hum Genet*. 2000;67:1563–1568.
8. Achermann JC, Vilain EJ. *NR0B1*-related adrenal hypoplasia congenita. In: Adam MP, Ardinger HH, Pagon RA, eds. *GeneReviews [Internet]*. Seattle: University of Washington; 2001:1993–2020.
9. Lloyd RV, Douglas BR, Young WF. *Endocrine Diseases. Atlas of Nontumor Pathology. First Series, Fascicle 1*. Washington, DC: American Registry of Pathology and the Armed Forces Institute of Pathology ARP Press; 2002.
10. Duan K, Giordano TJ, Mete O. Adrenal cortex. In: Mete O, Asa SL, eds. *Endocrine Pathology*. Cambridge University Press; 2016:588–627.
11. Mete O, Ng T, Christie-David D, et al. Silent corticotroph adenoma with adrenocortical choristoma: a rare but distinct morphological entity. *Endocr Pathol*. 2013;24:162–166.
12. Kale G, Pelley EM, Davis DB. Giant myelolipomas and inadvertent bilateral adrenalectomy in classic congenital adrenal hyperplasia. *Endocrinol Diabetes Metab Case Rep*. 2015;2015:150079.
13. Larose S, Bondaz L, Mermejo LM, et al. Coexistence of myelolipoma and primary bilateral macronodular adrenal hyperplasia with GIP-dependent Cushing's syndrome. *Front Endocrinol*. 2019;10:618.
14. Almeida MQ, Kaupert LC, Brito LP, et al. Increased expression of ACTH (MC2R) and androgen (AR) receptors in giant bilateral myelolipomas from patients with congenital adrenal hyperplasia. *BMC Endocr Disord*. 2014;14:42.
15. Cassol CA, Noria D, Asa SL. Ectopic thyroid tissue within the gallbladder: case report and brief review of the literature. *Endocr Pathol*. 2010;21:263–265.
16. Gourmaud J, Bongiovanni M, Triponez F, et al. Ectopic thyroid tissue in the adrenal gland. *Endocr Pathol*. 2014;25:353–355.
17. Mete O, Raphael S, Pirzada A, et al. Is adrenal ovarian thecal metaplasia a misnomer? Report of three cases of radial scar-like spindle cell myofibroblastic nodule of the adrenal gland. *Endocr Pathol*. 2011;22:222–225.
18. Porto AF. Lysosomal acid lipase deficiency: diagnosis and treatment of Wolman and cholesteryl ester storage diseases. *Pediatr Endocrinol Rev*. 2014;12(suppl 1):125–132.
19. Nada R, Gupta K, Lal SB, et al. An autopsy case of infantile GM1 gangliosidosis with adrenal calcification. *Metab Brain Dis*. 2011;26:307–310.
20. Erbil Y, Salmaslioğlu A, Barbaros U, et al. Clinical and radiological features of adrenal cysts. *Urol Int*. 2008;80:31–36.
21. Johnson PT, Horton KM, Fishman EK. Adrenal imaging with MDCT: nonneoplastic disease. *AJR Am J Roentgenol*. 2009;193:1128–1135.
22. Huebener KH, Treugut H. Adrenal cortex dysfunction: CT findings. *Radiology*. 1984;150:195–199.
23. Paolo WF, Nosanchuk JD. Adrenal infections. *Int J Infect Dis*. 2006;10:343–353.
24. Lack EE. *AFIP Atlas of Tumor Pathology, Fourth Series, Fascicle 8. Tumors of the Adrenal Glands and Extraadrenal Paraganglia*. Washington, DC: ARP Press; 2007.
25. Hermsen IG, Polak MP, Haak HR. Disappearing adrenal masses. *Endocrine*. 2010;38:153–157.
26. Jouinot A, Bertherat J. Diseases predisposing to adrenocortical malignancy (Li-Fraumeni syndrome, Beckwith-Wiedemann syndrome, and Carney complex). *Exp Suppl*. 2019;111:149–169.
27. Hodgson A, Pakbaz S, Mete O. A diagnostic approach to adrenocortical tumors. *Surg Pathol Clin*. 2019;12:967–995.
28. Faillot S, Foulonneau T, Néou M, et al. Genomic classification of benign adrenocortical lesions. *Endocr Relat Cancer*. 2021;28:79–95.
29. Juhlin CC, Bertherat J, Giordano TJ, et al. What did we learn from the molecular biology of adrenocortical neoplasia? From histopathology to translational genomics. *Endocr Pathol*. 2021;32:102–133.
30. Furnica RM, Dusoruth MM, Persu A, et al. Influence of secretory phenotype and preoperative preparation on surgical outcome in pheochromocytoma. *Endocr Connect*. 2021;10:92–101.
31. Mete O, Asa SL, Giordano TJ, et al. Immunohistochemical biomarkers of adrenocortical neoplasms. *Endocr Pathol*. 2018;29:137–149.
32. Mete O, Duan K. The many faces of primary aldosteronism and Cushing syndrome: a reflection of adrenocortical tumor heterogeneity. *Front Med*. 2018;12(5):54.
33. Duan K, Mete O. Clinicopathologic correlates of primary aldosteronism. *Arch Pathol Lab Med*. 2015;139:948–954.
34. Williams TA, Gomez-Sanchez CE, Rainey WE, et al. International histopathology consensus for unilateral primary aldosteronism. *J Clin Endocrinol Metab*. 2021;106:42–54.
35. Yamada M, Nakajima Y, Taguchi R, et al. KCNJ5 mutations in aldosterone- and cortisol-co-secreting adrenal adenomas. *Endocr J*. 2012;59:735–741.
36. Tang L, Li X, Wang B, et al. Clinical characteristics of aldosterone- and cortisol-coproducing adrenal adenoma in primary aldosteronism. *Int J Endocrinol*. 2018;2018:4920841.
37. Mouat IC, Omata K, McDaniel AS, et al. Somatic mutations in adrenocortical carcinoma with primary aldosteronism or hyperreninemic hyperaldosteronism. *Endocr Relat Cancer*. 2019;26:217–225.
38. Duan K, Gomez Hernandez K, Mete O. Clinicopathological correlates of adrenal Cushing's syndrome. *J Clin Pathol*. 2015;68:175–186.
39. Asa SL, Mete O. Hypothalamic endocrine tumors: an update. *J Clin Med*. 2019;8:1741.
40. Mete O, Asa SL. Morphological distinction of cortisol-producing and aldosterone-producing adrenocortical adenomas: not only possible but a critical clinical responsibility. *Histopathology*. 2012;60:1015–1016. author reply 1016–1017.
41. Assié G, Libé R, Espiard S, et al. ARMC5 mutations in macronodular adrenal hyperplasia with Cushing's syndrome. *N Engl J Med*. 2013;369:2105–2114.
42. Zhou WB, Chen N, Li CJ. A rare case of pure testosterone-secreting adrenal adenoma in a postmenopausal elderly woman. *BMC Endocr Disord*. 2019;19:14.
43. Balakumar T, Perry LA, Savage MO. Adrenocortical adenoma–an unusual presentation with hypersecretion of oestradiol, androgens and cortisol. *J Pediatr Endocrinol Metab*. 1997;10:227–229.
44. Hong Y, Hao Y, Hu J, et al. Adrenocortical oncocytoma: 11 Case reports and review of the literature. *Medicine (Baltim)*. 2017;96:e8750.
45. Al Badi MK, Al-Alwan I, Al-Dubayee M, et al. Testosterone- and cortisol-secreting oncocytic adrenocortical adenoma in the pediatric age-group. *Pediatr Dev Pathol*. 2018;21:568–573.
46. Hsiao HP, Chao MC, Lin CY, et al. Feminizing adrenocortical adenoma presenting as heterosexual precocious puberty: report of one case. *Acta Paediatr Taiwan*. 2005;46:97–100. Erratum in: Acta Paediatr Taiwan. 2005;46:392.
47. Mardi K. Virilizing adrenocortical oncocytic neoplasm with uncertain malignant potential in a child: a rare case report. *Indian J Cancer*. 2016;53:203–204.
48. Wang Z, Liu G, Sun H, et al. Clinical characteristics and prognosis of adrenocortical tumors in children. *Pediatr Surg Int*. 2019;35:365–371.
49. Mete O, Asa SL. Precursor lesions of endocrine system neoplasms. *Pathology*. 2013;45:316–330.
50. Duan K, Mete O. Hereditary endocrine tumor syndromes: the clinical and predictive role of molecular histopathology. *AJSP Rev Rep*. 2017;22:246–268.
51. Mete O, Hannah-Shmouni F, Kim R, et al. Inherited neuroendocrine neoplasms. In: Asa SL, La Rosa S, Mete O, eds. *The Spectrum of Neuroendocrine Neoplasia*. Springer; 2021:409–459.

52. Koch CA, Mauro D, Walther MM, et al. Pheochromocytoma in von Hippel-Lindau disease: distinct histopathologic phenotype compared to pheochromocytoma in multiple endocrine neoplasia type 2. *Endocr Pathol*. 2002;13:17–27.
53. Cassol C, Mete O. Endocrine manifestations of von Hippel-Lindau disease. *Arch Pathol Lab Med*. 2015;139:263–268.
54. Turchini J, Gill AJ. Morphologic clues to succinate dehydrogenase (SDH) deficiency in pheochromocytomas and paragangliomas. *Am J Surg Pathol*. 2020;44:422–424.
55. Seabrook AJ, Harris JE, Velosa SB, et al. Multiple endocrine tumors associated with germline MAX mutations: multiple endocrine neoplasia type 5? *J Clin Endocrinol Metab*. 2021;106:1163–1182.
56. Romanet P, Guerin C, Pedini P, et al. Pathological and genetic characterization of bilateral adrenomedullary hyperplasia in a patient with germline MAX mutation. *Endocr Pathol*. 2017;28:302–307.
57. Hernandez KG, Ezzat S, Morel CF, et al. Familial pheochromocytoma and renal cell carcinoma syndrome: TMEM127 as a novel candidate gene for the association. *Virchows Arch*. 2015;466:727–732.
58. Grogan RH, Pacak K, Pasche L, et al. Bilateral adrenal medullary hyperplasia associated with an SDHB mutation. *J Clin Oncol*. 2011;29:e200–e202.
59. Thompson LD. Pheochromocytoma of the Adrenal gland Scaled Score (PASS) to separate benign from malignant neoplasms: a clinicopathologic and immunophenotypic study of 100 cases. *Am J Surg Pathol*. 2002;26:551–566.
60. Kimura N, Takayanagi R, Takizawa N, et al. Phaeochromocytoma Study Group in Japan. Pathological grading for predicting metastasis in phaeochromocytoma and paraganglioma. *Endocr Relat Cancer*. 2014;21:405–414.
61. Koh JM, Ahn SH, Kim H, et al. Validation of pathological grading systems for predicting metastatic potential in pheochromocytoma and paraganglioma. *PLoS One*. 2017;12:e0187398.
62. Pierre C, Agopiantz M, Brunaud L, et al. COPPS, a composite score integrating pathological features, PS100 and SDHB losses, predicts the risk of metastasis and progression-free survival in pheochromocytomas/paragangliomas. *Virchows Arch*. 2019;474:721–734.
63. Stenman A, Zedenius J, Juhlin CC. The value of histological algorithms to predict the malignancy potential of pheochromocytomas and abdominal paragangliomas-a meta-analysis and systematic review of the literature. *Cancers*. 2019;11:225.
64. Job S, Draskovic I, Burnichon N, et al. Telomerase activation and ATRX mutations are independent risk factors for metastatic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2019;25:760–770.
65. Suh YJ, Choe JY, Park HJ. Malignancy in pheochromocytoma or paraganglioma: integrative analysis of 176 cases in TCGA. *Endocr Pathol*. 2017;28:159–164.
66. Fishbein L, Leshchiner I, Walter V, et al. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell*. 2017;31:181–193.
67. Jochmanova I, Abcede AMT, Guerrero RJS, et al. Clinical characteristics and outcomes of SDHB-related pheochromocytoma and paraganglioma in children and adolescents. *J Cancer Res Clin Oncol*. 2020;146:1051–1063.
68. Juhlin CC. Challenges in paragangliomas and pheochromocytomas: from histology to molecular immunohistochemistry. *Endocr Pathol*. 2021;32:228–244.
69. Llyod RV, Rosai J, Kloppel G, et al., eds. *World Health Organization Classification of Tumours of Endocrine Organs*. 4th ed. Lyon: International Agency for Research on Cancer (IARC); 2017.
70. Erickson LA, Lloyd RV, Hartman R, et al. Cystic adrenal neoplasms. *Cancer*. 2004;101:1537–1544.
71. Papotti M, Volante M, Duregon E, et al. Adrenocortical tumors with myxoid features: a distinct morphologic and phenotypical variant exhibiting malignant behavior. *Am J Surg Pathol*. 2010;34:973–983.
72. Bisceglia M, Ludovico O, Di Mattia A, et al. Adrenocortical oncocytic tumors: report of 10 cases and review of the literature. *Int J Surg Pathol*. 2004;12:231–243.
73. Lin BT, Bonsib SM, Mierau GW, et al. Oncocytic adrenocortical neoplasms: a report of seven cases and review of the literature. *Am J Surg Pathol*. 1998;22:603–614.
74. Wieneke JA, Thompson LD, Heffess CS. Adrenocortical neoplasms in the pediatric population: a clinicopathologic and immunophenotypic analysis of 83 patients. *Am J Surg Pathol*. 2003;27:867–881.
75. Hough AJ, Hollifield JW, Page DL, et al. Prognostic factors in adrenocortical tumors. A mathematical analysis of clinical and morphologic data. *Am J Clin Pathol*. 1979;72:390–399.
76. Weiss LM. Comparative histologic study of 43 metastasizing and nonmetastasizing adrenocortical tumors. *Am J Surg Pathol*. 1984;8:163–169.
77. Weiss LM, Medeiros LJ, Vickery Jr AL. Pathologic features of prognostic significance in adrenocortical carcinoma. *Am J Surg Pathol*. 1989;13:202–206.
78. van Slooten H, Schaberg A, Smeenk D, et al. Morphologic characteristics of benign and malignant adrenocortical tumors. *Cancer*. 1985;55:766–773.
79. Weiss M, Ford VL. Differences in steroidogenesis by the subcellular fractions of adrenocortical special zone and cortex proper of the female possum (Trichosurus vulpecula). *J Steroid Biochem*. 1984;21:701–707.
80. Aubert S, Wacrenier A, Leroy X, et al. Weiss system revisited: a clinicopathologic and immunohistochemical study of 49 adrenocortical tumors. *Am J Surg Pathol*. 202;26:1612–1619.
81. Volante M, Bollito E, Sperone P, et al. Clinicopathological study of a series of 92 adrenocortical carcinomas: from a proposal of simplified diagnostic algorithm to prognostic stratification. *Histopathology*. 200;55:535–543.
82. Duregon E, Fassina A, Volante M, et al. The reticulin algorithm for adrenocortical tumor diagnosis: a multicentric validation study on 245 unpublished cases. *Am J Surg Pathol*. 2013;37:1433–1440.
83. Mete O, Gucer H, Kefeli M, et al. Diagnostic and prognostic biomarkers of adrenocortical carcinoma. *Am J Surg Pathol*. 2018;42:201–213.
84. Pennanen M, Heiskanen I, Sane T, et al. Helsinki score-a novel model for prediction of metastases in adrenocortical carcinomas. *Hum Pathol*. 2015;46:404–410.
85. Duregon E, Cappellesso R, Maffeis V, et al. Validation of the prognostic role of the "Helsinki Score" in 225 cases of adrenocortical carcinoma. *Hum Pathol*. 2017;62:1–7.
86. Renaudin K, Smati S, Wargny M, et al. Clinicopathological description of 43 oncocytic adrenocortical tumors: importance of Ki-67 in histoprognostic evaluation. *Mod Pathol*. 2018;31:1708–1716.
87. Brown FM, Gaffey TA, Wold LE, et al. Myxoid neoplasms of the adrenal cortex: a rare histologic variant. *Am J Surg Pathol*. 2000;24:396–401.
88. Wong DD, Spagnolo DV, Bisceglia M, et al. Oncocytic adrenocortical neoplasms–a clinicopathologic study of 13 new cases emphasizing the importance of their recognition. *Hum Pathol*. 2011;42:489–499.
89. Papathomas TG, Duregon E, Korpershoek E, et al. Sarcomatoid adrenocortical carcinoma: a comprehensive pathological, immunohistochemical, and targeted next-generation sequencing analysis. *Hum Pathol*. 2016;58:113–122.
90. Duregon E, Volante M, Giorcelli J, et al. Diagnostic and prognostic role of steroidogenic factor 1 in adrenocortical carcinoma: a validation study focusing on clinical and pathologic correlates. *Hum Pathol*. 2013;44:822–828.
91. Renshaw AA, Granter SR. A comparison of A103 and inhibin reactivity in adrenocortical tumors: distinction from hepatocellular carcinoma and renal tumors. *Mod Pathol*. 1998;11:1160–1164.
92. Loy TS, Phillips RW, Linder CL. A103 immunostaining in the diagnosis of adrenocortical tumors: an immunohistochemical study of 316 cases. *Arch Pathol Lab Med*. 2002;126:170–172.
93. Mete O, Pakbaz S, Lerario AM, et al. The significance of alpha-inhibin expression in pheochromocytomas and paragangliomas. *Am J Surg Pathol*. 2021 (in press).
94. Asa SL, Ezzat S, Mete O. The diagnosis and clinical significance of paragangliomas in unusual locations. *J Clin Med*. 2018;7:280.
95. Kimura N. Dopamine β-hydroxylase: an essential and optimal immunohistochemical marker for pheochromocytoma and sympathetic paraganglioma. *Endocr Pathol*. 2021;32:258–261.
96. Kimura N, Miura Y, Nagatsu I, et al. Catecholamine synthesizing enzymes in 70 cases of functioning and non-functioning phaeochromocytoma and extra-adrenal paraganglioma. *Virchows Arch A Pathol Anat Histopathol*. 1992;421:25–32.
97. Powers JF, Tischler AS. Immunohistochemical staining for SOX10 and SDHB in SDH-deficient paragangliomas indicates that sustentacular cells are not neoplastic. *Endocr Pathol*. 2020;31:307–309.
98. Blanes A, Diaz-Cano SJ. Histologic criteria for adrenocortical proliferative lesions: value of mitotic figure variability. *Am J Clin Pathol*. 2007;127:398–408.

99. Wachenfeld C, Beuschlein F, Zwermann O, et al. Discerning malignancy in adrenocortical tumors: are molecular markers useful? *Eur J Endocrinol*. 2001;145:335–341.
100. Schmitt A, Saremaslani P, Schmid S, et al. IGFII and MIB1 immunohistochemistry is helpful for the differentiation of benign from malignant adrenocortical tumours. *Histopathology*. 2006;49:298–307.
101. Soon PS, Gill AJ, Benn DE, et al. Microarray gene expression and immunohistochemistry analyses of adrenocortical tumors identify IGF2 and Ki-67 as useful in differentiating carcinomas from adenomas. *Endocr Relat Cancer*. 2009;16:573–583.
102. Erickson LA, Jin L, Sebo TJ, et al. Pathologic features and expression of insulin-like growth factor-2 in adrenocortical neoplasms. *Endocr Pathol*. 2001;12:429–435.
103. Picard C, Orbach D, Carton M, et al. Revisiting the role of the pathological grading in pediatric adrenocortical tumors: results from a national cohort study with pathological review. *Mod Pathol*. 2019;32:546–559.
104. Guntiboina VA, Sengupta M, Islam N, et al. Diagnostic and prognostic utility of SF1, IGF2 and p57 immunoexpression in pediatric adrenocortical tumors. *J Pediatr Surg*. 2019;54:1906–1912.
105. Martins-Filho SN, Almeida MQ, et al. Clinical Impact of Pathological Features Including the Ki-67 labeling index on diagnosis and prognosis of adult and pediatric adrenocortical tumors. *Endocr Pathol*. 2021;32(2):288–300.
106. Papathomas TG, Pucci E, Giordano TJ, et al. An international Ki67 reproducibility study in adrenocortical carcinoma. *Am J Surg Pathol*. 2016;40:569–576.
107. Giordano TJ, Berney D, de Krijger RR, et al. Data set for reporting of carcinoma of the adrenal cortex: explanations and recommendations of the guidelines from the International Collaboration on Cancer Reporting. *Hum Pathol*. 2021;110:50–61.
108. Raymond VM, Everett JN, Furtado LV, et al. Adrenocortical carcinoma is a lynch syndrome-associated cancer. *J Clin Oncol*. 2013;31:3012–3018.
109. Brondani VB, Montenegro L, Lacombe AMF, et al. High prevalence of alterations in DNA mismatch repair genes of Lynch syndrome in pediatric patients with adrenocortical tumors carrying a germline mutation on TP53. *Cancers (Basel)*. 2020;12:621.
110. Mete O, Tischler AS, de Krijger R, et al. Protocol for the examination of specimens from patients with pheochromocytomas and extra-adrenal paragangliomas. *Arch Pathol Lab Med*. 2014;138:182–188.
111. Papathomas TG, Oudijk L, Persu A, et al. SDHB/SDHA immunohistochemistry in pheochromocytomas and paragangliomas: a multicenter interobserver variation analysis using virtual microscopy: a Multinational Study of the European Network for the Study of Adrenal Tumors (ENS@T). *Mod Pathol*. 2015;28:807–821.
112. NGS in PPGL (NGSnPPGL) Study Group, Toledo RA, Burnichon N, et al. Consensus statement on next-generation-sequencing-based diagnostic testing of hereditary phaeochromocytomas and paragangliomas. *Nat Rev Endocrinol*. 2017;13:233–247.
113. Udager AM, Magers MJ, Goerke DM, et al. The utility of SDHB and FH immunohistochemistry in patients evaluated for hereditary paraganglioma-pheochromocytoma syndromes. *Hum Pathol*. 2018;71:47–54.
114. Pinato DJ, Ramachandran R, Toussi ST, et al. Immunohistochemical markers of the hypoxic response can identify malignancy in phaeochromocytomas and paragangliomas and optimize the detection of tumours with VHL germline mutations. *Br J Cancer*. 2013;108:429–437.
115. Favier J, Meatchi T, Robidel E, et al. Carbonic anhydrase 9 immunohistochemistry as a tool to predict or validate germline and somatic VHL mutations in pheochromocytoma and paraganglioma-a retrospective and prospective study. *Mod Pathol*. 2020;33:57–64.
116. Mete O, David JL, Rudzinski ER, et al. *The College of American pathologists. Protocol for the examination of specimens from patients with carcinoma of the adrenal gland*; 2020. Version: 4.1.0.0. Available at: https://www.cap.org/protocols-and-guidelines/cancer-reporting-tools.
117. Thompson LDR, Gill AJ, Asa SL, et al. Data set for the reporting of pheochromocytoma and paraganglioma: explanations and recommendations of the guidelines from the International Collaboration on Cancer Reporting. *Hum Pathol*. 2021;110:83–97.

92

Adrenal Gland Imaging

David S. Lin, Ka Kit Wong, Elaine M. Caoili, James Shields,
Anca M. Avram, and Benjamin L. Viglianti

OUTLINE

Introduction, 1530
Adrenal Gland Anatomy, 1530
Adrenal Gland Imaging, 1530
  Arteriography and Adrenal Vein Sampling, 1531
Clinical Utility, 1532
  Adenoma or Metastases, 1532
  Benign Adrenal Lesions, 1535
  Neoplastic Adrenal Lesions, 1537
  Interpretation of Adrenal Vein Sampling assays in Primary Hyperaldosteronism, 1547
  Functional Imaging in Primary Hyperaldosteronism, 1548
Summary, 1548

INTRODUCTION

A variety of different pathologies can affect the adrenal glands. High-resolution anatomic imaging with computed tomography (CT) and magnetic resonance imaging (MRI) can demonstrate subtle irregularities of the adrenal contours and can identify lesions millimeters in size. When combined with sensitive biochemical tests of adrenocortical and medullary hormones, diagnosis of adrenal gland dysfunction can be made at very early stages of disease. Certain adrenal pathologies can be confidently diagnosed based on imaging features alone and can potentially circumvent the need for percutaneous tissue sampling. Additionally, functional imaging with nuclear medicine scintigraphy and positron emission tomography (PET) add significant clinical utility, particularly in the evaluation of biochemically functional tumors as well as distant extra-adrenal metastatic disease.

ADRENAL GLAND ANATOMY

The adrenal glands are located in the retroperitoneum in the superior aspect of Gerota's fascia in the perirenal space. These paired glands typically lie adjacent and lateral to the first lumbar vertebral body. The adrenal glands are suprarenal structures with the right gland lying more cranial relative to the left. They demonstrate an inverted-Y or inverted-V morphology with posteromedial and posterolateral limbs. The left gland tends to be slightly larger, with a width measuring up to 8 mm, whereas the width of the right gland measures up to approximately 6 mm (Fig 92.1).[1]

Histologically, there are two distinct components of the adrenal glands, an outer cortex and an inner medulla, which together can have a trilaminar appearance, although this is often not well depicted on imaging. The cortex is further subdivided into three zones, each with its own dedicated function: the zona glomerulosa, the zona fasciculata, and the zona reticularis. The zona glomerulosa is the outer most component of the cortex, secretes aldosterone, and is involved with electrolyte balance. The zona fasciculata, located just deep to the zona glomerulosa, synthesizes glucocorticoids and is involved with glucose metabolism. Lastly, the zona reticularis, the innermost component of the cortex, is involved with sex hormone production. The medulla is involved in the production of catecholamines, epinephrine, and norepinephrine, and regulates the stress response. The adrenal gland is surrounded by a collagenous capsule that is not seen on cross-sectional imaging.

The arterial supply to each adrenal gland comes from three sources: a superior adrenal artery arising from the inferior phrenic artery, the middle adrenal from the aorta, and the inferior adrenal from the renal artery. One or more central veins drain the cortex and the medulla and form the adrenal vein. The adrenal vein often freely communicates with one or more retroperitoneal and/or hepatic veins. The right adrenal vein is shorter than the left and usually enters the right posterolateral surface of the inferior vena cava (IVC) at a level between the eleventh and twelfth ribs. The central vein of the left adrenal gland joins the left inferior phrenic vein to form the phrenicoadrenal trunk, which then enters the superior aspect of the left renal vein. Collateral communications with renal capsular, azygous, and retroperitoneal veins are common. These descriptions are of classic venous anatomy and there are numerous described anatomic variants.

KEY POINTS

- The adrenal gland is made of a cortex and medulla, each having specific endocrine function and related pathology.

ADRENAL GLAND IMAGING

Ultrasound, while it is widely available and does not involve ionizing radiation, has limited diagnostic utility in the evaluation of the adrenal glands in adult patients due to size and location. It is, however, the imaging modality of choice for adrenal gland

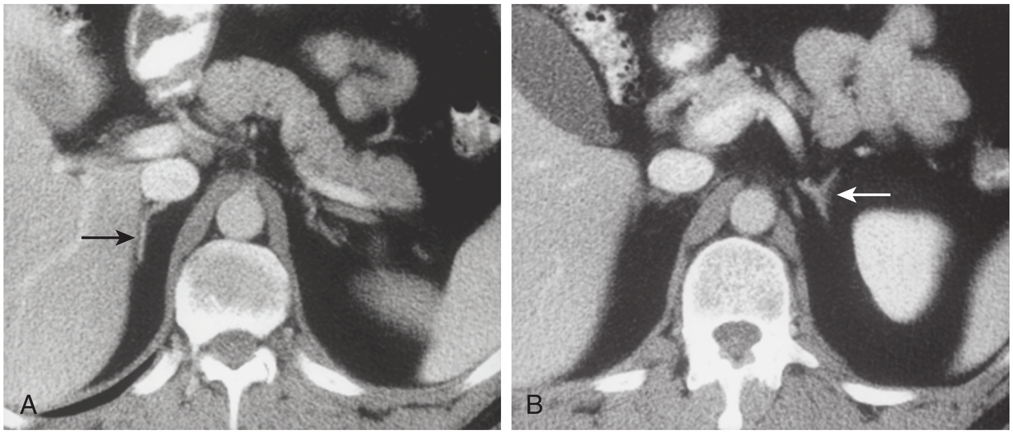


Fig. 92.1 Computed tomography appearance of normal adrenal glands. **A,** The cephalad portion of the right adrenal *(arrow)* is seen. **B,** The three limbs of the left adrenal *(arrow)* are shown at this level. (From Korobkin M, Francis IR. Adrenal imaging. *Semin Ultrasound CT MR.* 1995;16:317–330.)

evaluation in the pediatric population and is particularly useful for the assessment of adrenal hemorrhage and masses such as neuroblastoma (NB). The adrenal glands are normally hypoechoic or have echogenicity that is similar to that of adjacent perirenal fat. A hyperechoic center can sometimes be seen, corresponding with the medulla. Ultrasound may also have a role in percutaneous image-guided biopsies of adrenal masses in both pediatric and adult patients.

CT is currently the primary method of anatomic imaging of the adrenal gland and can visualize the glands in virtually all cases. The attenuation and enhancement of the adrenal glands are similar to soft tissue. Adrenal abnormalities are often identified incidentally (e.g., "incidentaloma") during evaluation of a separate indication other than adrenal mass detection. Density measurements can be helpful for lesion analysis. A dedicated adrenal protocol CT can be especially beneficial for further characterization, particularly in distinguishing between a benign adenoma from metastatic disease.

MRI, in comparison with CT, demonstrates improved tissue contrast and utility in patients with renal insufficiency or iodinated contrast allergy. On MRI, the adrenal glands are isointense to the spleen on T1-weighted sequences and iso- to hypointense to the spleen on T2-weighted sequences. Fat-saturation sequences to detect macroscopic fat and in- and out-of-phase sequences to detect intracellular lipid provide additional utility for lesion characterization. Recent advances in MRI technology have made MRI a competitive method with CT for imaging suspected adrenal disease.

Positron emission tomography (PET)/CT using (^{18}F)-fluorodeoxyglucose (FDG) is now widely used for imaging of hematological and solid malignancies that have characteristic overexpression of insulin-independent glucose transporters (GLUT1, GLUT3) leading to uptake and intracellular trapping of FDG in cancer cells. FDG-PET/CT allows characterization of the "molecular signature" of an adrenal mass and can identify adrenal metastases with excellent sensitivity and high specificity.[2,3]

Radiopharmaceuticals for adrenocortical imaging have been available for more than 30 years, since the mid 1970s, with the radiocholesterol analog ^{131}I-6β-iodomethylnorcholesterol (NP-59) being the first in commercial use[4] and newer radiopharmaceutical inhibitors of 11β-hydroxylase, ^{11}C-etomidate and ^{11}C-, ^{18}F-, and ^{123}I-metomidate developed for imaging the normal adrenal cortex and various adrenocortical tumors.[5] The first clinically successful adrenal medulla imaging agent, metaiodobenzylguanidine (MIBG), labelled with ^{123}I or ^{131}I, was developed in the early 1980s and continues to be used clinically for imaging pheochromocytoma, paraganglioma, and NB.[6] MIBG accumulation in neuronal tissues involves the norepinephrine transporter reuptake mechanism, and subsequent packaging by the vesicular monoamine transporter into catecholamine storage vesicles of adrenergic nerves and adrenomedullary cells. Other PET radiotracers for adrenomedullary imaging of pheochromocytoma are being investigated, including ^{18}F-6-fluorodopamine (18F-DA), ^{18}F-fluorodihydroxyphenylalanine (18F-DOPA), and ^{11}C-hydroxyephedrine (HED).[7-9]

^{68}Ga-1,4,7,10–tetraazacyclododecane–1,4,7,10–tetraacetic acid (DOTA)–labeled somatostatin analogs have been developed for PET imaging of neuroendocrine tumors (NETs), affording image acquisition within sixty minutes postinjection and higher spatial resolution than radiotracers labeled for single-photon emission computed tomography (SPECT), supplanting the ^{111}In-DTPA-octreotide scintiscan. ^{68}Ga-DOTA peptide PET has been used to depict pheochromocytoma and paraganglioma successfully.[10]

KEY POINTS

- Several modalities are used to image the adrenal glands based on anatomy or function, the most common of which is computed tomography.

Arteriography and Adrenal Vein Sampling

Arteriography is used only in a few selected adrenal lesions or in the case of extraadrenal pheochromocytoma, when additional information about tumor resectability and vascular anatomy is needed. Adrenal venous sampling (AVS) is performed in patients with proven primary hyperaldosteronism to distinguish hypersecretory unilateral adenoma from bilateral adrenal hyperplasia (Fig 92.2). Rarely, AVS is performed to determine whether autonomous hormone excess is unilateral or bilateral in Cushing syndrome. Indications, specificity, and sensitivity, as well as data interpretations and treatment recommendations based on AVS, are discussed in another chapter in this textbook.

The specific risk with adrenal arteriography is rupture of the adrenal arteries with contrast extravasation. Potential complications of adrenal vein catheterization include intraadrenal extravasation of contrast medium, bleeding, and adrenal vein thrombosis. Adrenal hemorrhage can cause severe pain and fever, as well as destruction of adrenal glandular function and hypoadrenalism if both adrenal glands are damaged.

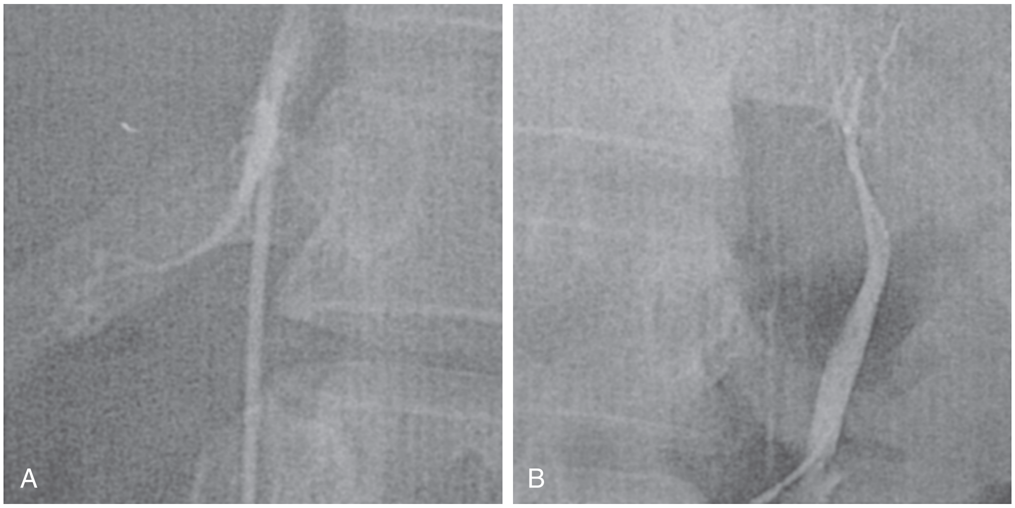


Fig. 92.2 Bilateral adrenal venograms in a patient with primary hyperaldosteronism. **A,** Right adrenal venogram in the left anterior oblique projection. The intrarenal branches are filled poorly. The main vein enters the posterolateral aspect of the inferior vena cava. **B,** Left adrenal venogram in the anteroposterior projection with the injection of contrast medium via the 3F microcatheter introduced through the 5F reverse hook catheter positioned near the junction of the main adrenal vein and the left renal vein.

KEY POINTS

- Adrenal vein sampling (AVS) allows direct measure of adrenal hormone synthesis.
- AVS is invasive and technically challenging.

CLINICAL UTILITY

Adenoma or Metastases

The increasing availability and use of anatomic imaging, particularly CT, performed for indications unrelated to the adrenal gland have led to a growing number of incidentally found adrenal lesions, also termed incidentalomas. The discovery of an adrenal mass presents a diagnostic and, at times, a therapeutic challenge. The evaluation of these incidentalomas could require identification of malignancy and would subsequently impact clinical management. The clinical history of a preexisting cancer diagnosis is important due to significant differences in prevalence of adrenal malignancy in oncologic versus nononcologic patients. In addition, it is necessary to distinguish if the adrenal mass is functional or hormonally active by biochemical testing of secretory function (screening for mineralocorticoid, glucocorticoid, and catecholamine secretion).

Because the overwhelming majority of incidentalomas are benign, an aggressive approach is often not indicated. The prevalence of adrenal adenomas is high, ranging from 4% to 10% of patients studied with CT or MRI for indications other than suspected adrenal disease.[11] Small, homogeneous adrenal masses discovered incidentally are likely to be adenomas. Unenhanced CT densitometry can be used to accurately differentiate adrenal adenomas from metastases. Most adenomas, which are lipid-rich, have unenhanced CT attenuation values lower than those of metastases. In a histologic/radiologic study of a small number of resected adrenal adenomas that had undergone presurgical unenhanced CT, chemical shift MRI, or both, a good inverse linear correlation was seen between the estimated number of lipid-rich cells and the unenhanced CT attenuation values, and a good linear correlation was noted with the relative change in signal intensity on opposed-phase MRI (Figs. 92.3 and 92.4).[12]

In the scenario of an oncologic patient with an unsuspected adrenal mass and no other evidence of metastatic disease, the goal of noninvasive diagnostic adrenal imaging is to identify an adrenal adenoma with high specificity, and thus exclude adrenal metastases. With the use of pooled data from multiple published studies, it has been shown that the most optimal combination of a sensitivity (71%) and specificity (98%) for the diagnosis of lipid-rich adrenal adenoma is achieved when a threshold attenuation value of <10 Hounsfield units (HU) is used on unenhanced CT (Fig. 92.5).[13] However, lipid-poor adenomas tend to have attenuation values greater than 10 HU on unenhanced CT imaging and are thus more difficult to distinguish from metastases. Additionally, postcontrast attenuation values for all adenomas, whether lipid-rich or lipid-poor, show considerable overlap with metastases. This poses a diagnostic dilemma on routine single-phase contrast-enhanced CTs of the abdomen and pelvis.

For indeterminate adrenal lesions (e.g., those with unenhanced attenuation values >10 HU or lesions identified on a routine contrast-enhanced CT), a dedicated adrenal protocol CT is beneficial for distinguishing between benign adenomas from metastases. Adrenal adenomas, whether lipid-rich or lipid-poor, demonstrate rapid washout of contrast. A dedicated adrenal protocol CT typically includes precontrast, 60 to 75-second postcontrast, and 15-minute delayed imaging. Absolute washout is calculated as follows: [(postcontrast HU – delayed HU)/(postcontrast HU – precontrast HU)] x 100. In cases where precontrast imaging is not obtained, a relative washout can be calculated as follows: [(postcontrast HU – delayed HU)/postcontrast HU)] x 100. Adenoma can be diagnosed with absolute washout greater than or equal to 60% or relative washout greater than or equal to 40% (Fig. 92.6). In a prospective study of 166 adrenal masses with unenhanced CT, those with attenuation values greater than 10 HU underwent contrast-enhanced and 15-minute delayed enhanced CT. This protocol correctly characterized 160 of the 166 masses (96%). Exclusion of primary adrenal neoplasms increased sensitivity and specificity for characterizing a mass as adenoma versus metastasis to 98% (124/127) and 97% (33/34), respectively.[14] Similar results were subsequently confirmed with similar protocols.[15,16] Therefore, a dedicated adrenal protocol CT to calculate washout is being used in many centers to characterize incidentally discovered adrenal masses.

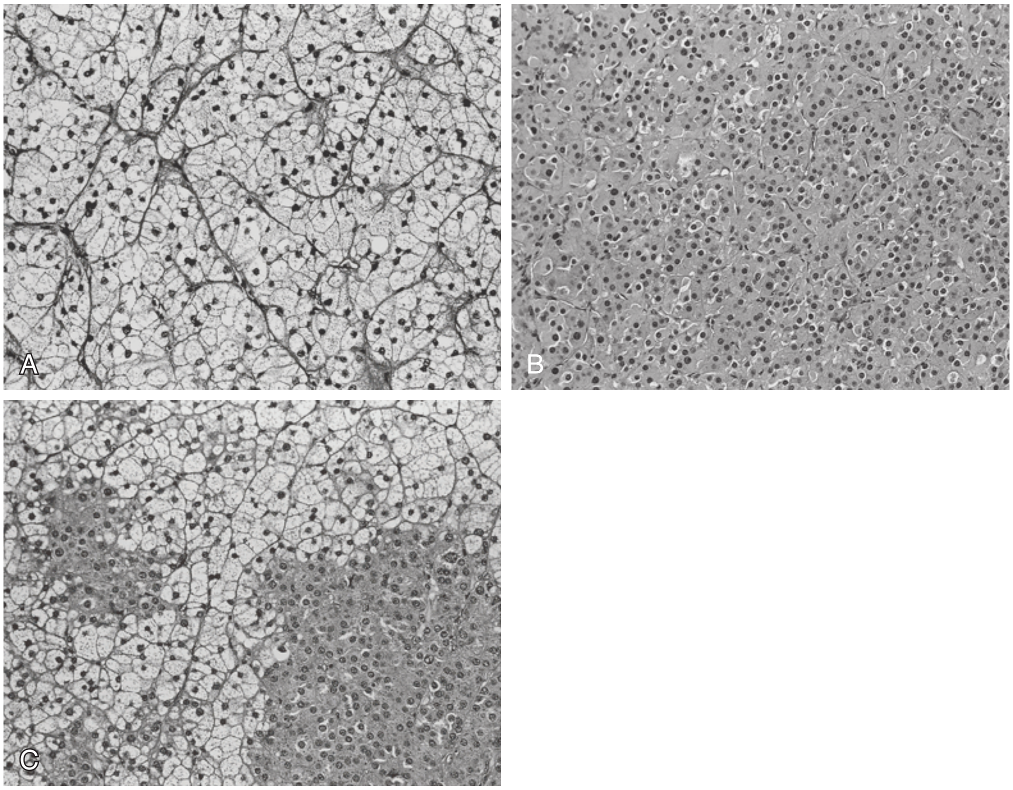


Fig. 92.3 **Histologic specimens from resected adrenal adenomas. A,** Primarily of lipid-rich clear cells. **B,** Primarily of lipid-poor clear cells. **C,** Of an admixture of clear and compact cortical cells. (Hematoxylin and eosin stain, original magnification ×200.) (From Korobkin M, Giordano TJ, Brodeur FJ, et al. Adrenal adenomas: relationship between histologic lipid and CT and MR findings. *Radiology*. 1996;200:743–747.)

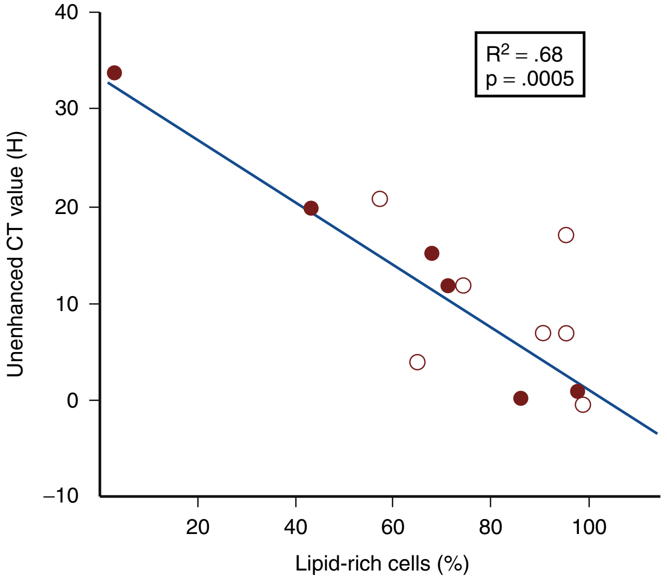


Fig. 92.4 **Plot of unenhanced computed tomography attenuation number versus the percentage of lipid-rich cells in 13 surgically resected adrenal adenomas.** *HU*, Hounsfield units. (From Korobkin M, Giordano TJ, Brodeur FJ, et al. Adrenal adenomas: Relationship between histologic lipid and CT and MR findings. *Radiology*. 1996;200:743–747.)

Notably, the assessment of enhancement washout curves for adrenal masses is valid only for lesions with relatively homogeneous attenuation after contrast enhancement; the diagnosis of adrenal adenoma cannot be made in masses that contain prominent regions of necrosis or hemorrhage.

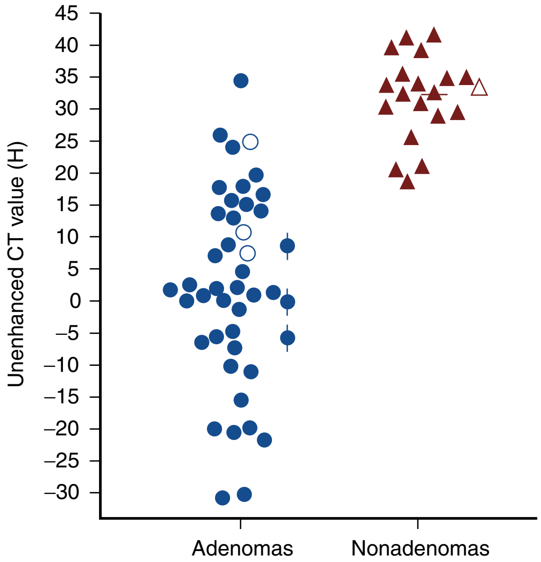


Fig. 92.5 **Scattergram of attenuation values on unenhanced compute tomography of adrenal adenomas and nonadenomas.** All masses with H values of less than 18 were adenomas. ●, Nonhyperfunctioning adenomas; O, Cushing adenomas; ●, primary aldosteronism adenomas; ▲, metastases; Δ, pheochromocytomas; ▲, cortical carcinomas. (From Korobkin M, Brodeur FJ, Yutzy GG, et al. Differentiation of adrenal adenomas from nonadenomas using CT attenuation values. *Am J Roentgenol*. 1996;166:531–536.)

On MRI, chemical shift sequences can be used to identify intracytoplasmic lipid and can therefore distinguish adenomas.[17,18] Taking advantage of the different precession rates for hydrogen atoms in water

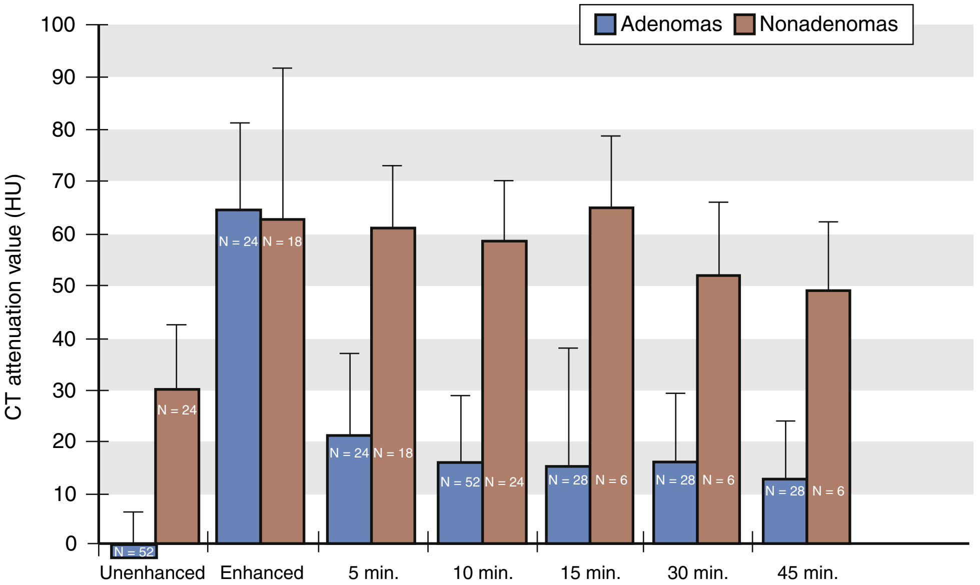


Fig. 92.6 Bar graph shows mean computed tomography attenuation value plus 1 standard deviation for adrenal adenomas *(black)* and nonadenomas *(white)* on unenhanced, enhanced, and delayed enhanced scans. (From Korobkin M, Brodeur FJ, Francis IR, et al. CT-time-attenuation washout curves of adrenal adenomas and nonadenomas. *Am J Roentgenol.* 1998;170:747–752.)

and lipid molecules, chemical shift MRI results in a decrease in signal intensity on opposed-phase versus in-phase imaging in tissue containing intracellular lipid that is often present in adenomas and absent in most metastases. Assessment of the chemical shift change can be made via simple visual analysis or by quantitative methods using standard region-of-interest measurements of the mass (Fig 92.7). The two most commonly used formulae for quantitative analysis of chemical shift change are the signal intensity index (SII) and adrenal-to-spleen chemical shift ratio (ASR). SII is calculated by $[(IPI_a - OPI_a)/IPI_a] \times 100$ and ASR is calculated by $[(OPI_a/OPI_s)/(IPI_a/IPI_s)] \times 100$, where IPI_a represents in-phase intensity in the adrenal, OPI_a represents opposed-phase intensity in the adrenal, IPI_s represents in-phase intensity in the spleen, and OPI_s represents opposed-phase intensity in the spleen. An SII of 16.5% (or 20%) and an ASR of less than 0.71 are considered to be highly accurate for the characterization of lipid-containing adrenal adenomas.[18-21]

The advent of FDG-PET/CT for determination of the metabolic signature of adrenal masses to predict malignant potential has replaced earlier work using functional adrenal imaging such as NP-59 or MIBG scintigraphy to characterize specific molecular processes of adrenal tumors.[22] A metaanalysis of 1217 patients from 21 studies reported that FDG-PET had a pooled sensitivity of 97% and specificity of 91% for distinguishing malignant from benign adrenal masses.[23] Another metaanalysis reported similar results, with pooled sensitivity 91% and specificity 91% for both FDG-PET and PET/CT, although significant heterogeneity of the included studies was noted.[24] In subjects without preexisting cancer history, FDG-PET/CT had sensitivity 86.7% and specificity 86.1% to distinguish malignant from benign etiology of adrenal incidentaloma.[25]

The accuracy of FDG-PET/CT appears to be similar using either qualitative or quantitative assessment of the adrenal mass uptake compared with adjacent liver background. In a recent study of metabolic parameters for characterization of adrenal incidentaloma in noncancer patients, a standardized uptake value (SUV) max cutoff greater than 3.1 and adrenal:liver ratio greater than 1.8 had 100% sensitivity and 67% specificity for identifying cancer.[26] Similar findings were reported in 59 patients (malignant lesions in 52 using histological gold standard), that adrenal tumor SUV threshold greater than 3.5 had an accuracy of 87% for predicting malignancy.[27] FDG-PET/CT metabolic measurements have been combined with CT histograms and unenhanced CT parameters (HU, size) on hybrid PET/CT cameras to improve diagnostic accuracy.[28,29]

Current evidence suggests a complementary role for FDG-PET/CT to CT for the assessment of indeterminate adrenal masses characterized by CT or MRI. In a systematic review of 353 indeterminate adrenal lesions measuring greater than 10 HU, CT showed a sensitivity 100% with a specificity of 33%, while 89 patients with FDG-PET/CT having a tumor:liver ratio greater than 1.8 had a sensitivity of 87% and a specificity 84% for malignancy. Therefore, FDG-PET is helpful in further characterization of indeterminate lesions, improving specificity.[30]

Despite excellent sensitivity, FDG-PET/CT false negatives can occur with small tumors (<5 mm), hemorrhagic or necrotic tumors, and tumors with low FDG avidity postchemotherapy.[31] Adrenal metastases secondary to pulmonary carcinoid or bronchioloalveolar lung carcinoma have produced false-negative PET results due to low FDG avidity in these cancers. False positives are reported in inflammatory/infectious processes, benign pheochromocytoma, adrenocortical adenoma, adrenal nodular cortical hyperplasia, myelolipoma, hemorrhagic nodules, adrenal endothelial cysts, and perirenal brown fat and following chemotherapy.[32] A case of a false-positive FDG-PET due to adrenal angiomyolipoma with a SUV of 18.8 was attributed to the effect of colony-stimulating factor on the tumor's hemopoietic components.[33] Interestingly, FDG-PET has been shown to allow metabolic assessment of response to mammalian target of rapamycin (mTOR) therapy in another case of adrenal angiomyolipoma.[34] Inflammation during immunotherapy due to ipilimumab-induced adrenalitis was reported to have uptake on PET.[35] Benign bilateral adrenal FDG uptake can occur with Cushing syndrome due to ectopic adrenocorticotropic hormone (ACTH),[36] and with ACTH-secreting esthesioneuroblastoma.[37] Bilateral adrenal FDG uptake has also been described with a vasovagal stress reaction.[38,39] The coexistence of an adrenal adenoma and a metastasis (collision adrenal tumor) has led to the apparently increased FDG activity in a benign-appearing lesion; careful coregistration of PET and CT data, as well as meticulous analysis of anatomic

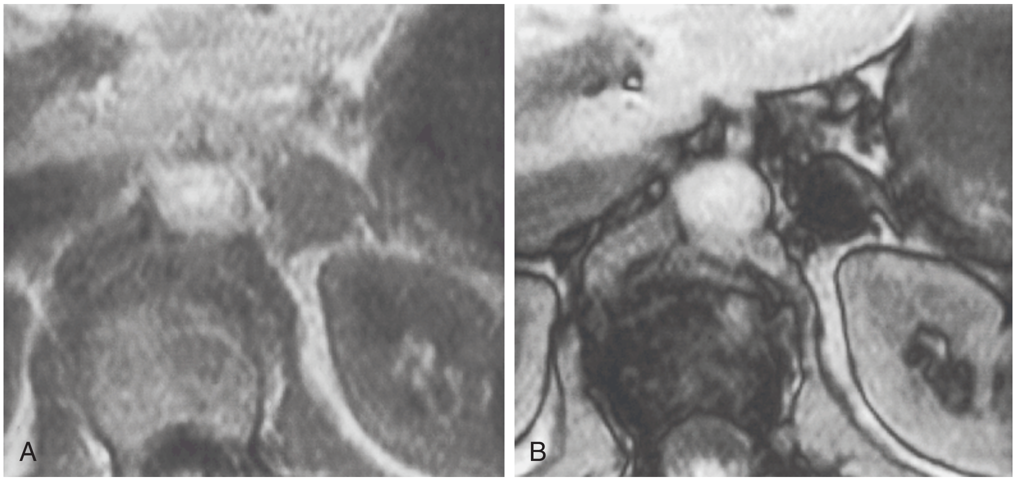


Fig. 92.7 **Adenoma.** **A,** Left adrenal gland mass is seen on this in-phase magnetic resonance image (MRI). **B,** A significant decrease in signal is seen on the out-of-phase MRI. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol.* 2002;179:559–568.)

details on CT, are essential for accurate diagnosis when there are two different pathologic processes within the adrenal gland.

KEY POINTS

- The primary use and goal for imaging is to differentiate a benign lesion from metastasis.
- Computed tomography, magnetic resonance imaging, and more recently, positron emission tomography, are useful for this distinction.
- Primary adrenal neoplasms (benign or malignant) are more challenging to evaluate.

Benign Adrenal Lesions

Adenoma. Adenoma, the most common adrenal tumor, occurs in 1.4% to 8.7% of postmortem examinations. The incidence is even higher amongst patients with hypertension or diabetes. Because most adenomas contain intracytoplasmic lipid, many have a low CT density, often approximating water on unenhanced CT (Fig. 92.8). Chemical shift MRI can also be used to distinguish many adenomas by identifying the intracytoplasmic lipid. Calcification is rare. Adenomas show early enhancement after intravenous contrast administration (Fig. 92.9). Although the degree of enhancement is not significantly different from that of other adrenal tumors, adenomas show more rapid washout of contrast than do adrenal malignancies.

Lipid-poor adenomas (LPAs), as mentioned earlier, can be problematic to characterize with CT and MRI. These comprise one third of adrenal adenomas and measure greater than 10 HU on precontrast imaging, having an absolute washout greater than 60% and relative washout greater than 40%. A study of 57 LPAs, 37 with FDG-PET scan, reported that surgery was more likely to be performed with larger tumors, masses with higher HU, and those with higher FDG uptake. Notably, oncocytic features were seen in 70% of these LPAs; therefore, a number of LPAs may be oncocytomas.[40]

It is not fully understood why some benign adrenal tumors demonstrate increased FDG uptake; however, the functional status of the tumor is presumed to be a factor, with a higher degree of metabolic activity occurring in functioning adenomas. A cohort of patients with FDG-PET scan were evaluated for differences in the metabolic signature between groups of adrenal lesions that included nonfunctioning tumors (n = 4) and functioning tumors, 11 with Cushing syndrome and four with Conn syndrome.[41] Notably, cortisol-secreting tumors had a significantly higher average SUVmax of 5.9 compared with nonfunctioning tumors (SUVmax of 4.2) and aldosterone-secreting tumors (SUVmax of 3.2). The underlying mechanism proposed on histological examination was that GLUT3 expression was higher in patients with Cushing syndrome compared with those with nonfunctioning tumors or with Conns syndrome. Another histopathology study reported that intensity of FDG uptake in adrenal adenoma was related to GLUT transporter, clear cell ratio, and higher unenhanced CT density.[42]

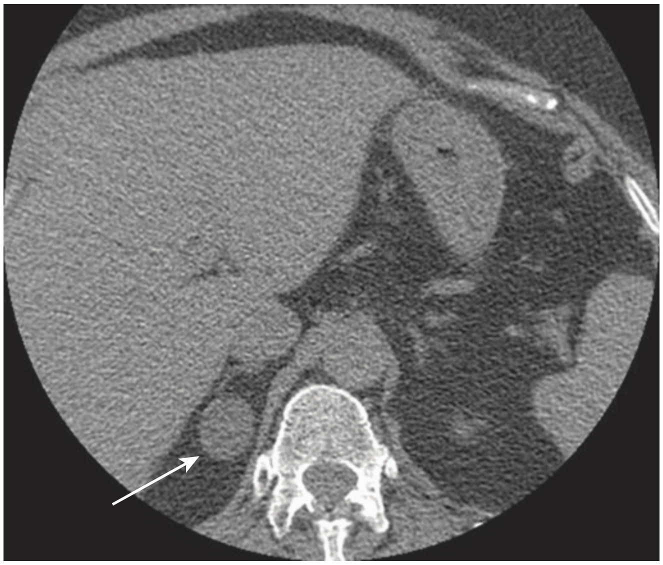


Fig. 92.8 **Lipid-rich adenoma.** Right adrenal mass (3 cm) is shown on this unenhanced computed tomography scan. Attenuation value of –4 Hounsfield units allows confident diagnosis of benign lesion, either cyst or lipid-rich adenoma.

Myelolipoma. A myelolipoma is a rare incidental benign tumor composed of bone marrow elements, accounting for 2% to 6% of primary adrenal lesions. Occasionally, large tumors or those undergoing tumor necrosis or spontaneous hemorrhage may cause flank pain.[43,44] Because these tumors contain large amounts of fat, most myelolipomas

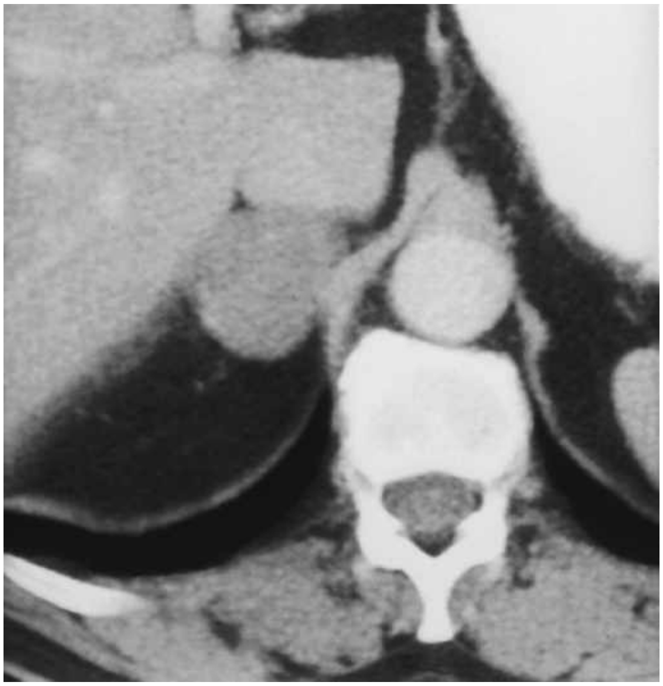


Fig. 92.9 **Adenoma.** Right adrenal gland mass (4 cm) is shown on this contrast-enhanced computed tomography examination. Attenuation value on enhanced scans is not sufficiently characteristic to distinguish benign from malignant causes. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol*. 2002;179:559–568.)

are easily recognized on CT (Fig. 92.10). The fatty components are intermixed with elements of soft-tissue density and calcification (seen in up to 20% of cases).[45,46]

The appearance of a myelolipoma on MRI also reflects the fat and bone marrow elements in the tumor. Fat exhibits high signal intensity on both non-fat-suppressed T1- and T2-weighted sequences, with loss of signal intensity on fat-suppressed images. Bone marrow elements have a low signal intensity on T1-weighted images and a moderate signal intensity on T2-weighted images. On in-phase and opposed-phase imaging, an India ink artifact can be seen at the periphery between the lesion and normal adrenal tissue, which is suggestive of a myelolipoma. Because a confident diagnosis of myelolipoma can be usually made with CT or MRI, most are treated conservatively.

Myelolipomas are not metabolically active on FDG-PET/CT (being predominately composed of fat, although mild FDG uptake can occur within the nonfatty portions of the lesions).[47]

Hemorrhage. Adrenal hemorrhage can be bilateral or unilateral. When bilateral, they are associated with anticoagulation therapy or a blood dyscrasia, and less commonly surgical stress, sepsis, large surface burns, hypotension, or rarely, trauma.[48] Unilateral adrenal hemorrhage usually is caused by blunt force abdominal trauma or has been seen occasionally after liver transplantation or by hemorrhage into a preexisting adrenal neoplasm. Adrenal vein thrombosis may also cause unilateral adrenal hemorrhage and may be a complication of adrenal vein catheterization.[49] Unilateral hemorrhage is seen more often affecting the right gland than the left.[50]

Acute or subacute adrenal hemorrhage typically has unenhanced attenuation values of 55 to 90 HU (Fig. 92.11). Follow-up studies show a diminution in size and a gradual decrease in attenuation values. The high attenuation value of a recent adrenal hemorrhage is usually readily apparent on unenhanced CT, but is indistinguishable from a solid adrenal neoplasm on contrast-enhanced CT. An adrenal mass detected on contrast-enhanced CT after trauma usually is assumed to be a hematoma; however, an unrelated adrenal neoplasm can only be excluded by serial follow-up imaging. Similarly, MRI may suggest hemorrhage depending on the age of the bleed. For example, a subacute adrenal hemorrhage can be identified by high signal intensity on T1- and T2-weighted sequences, reflecting the presence of methemoglobin.[51]

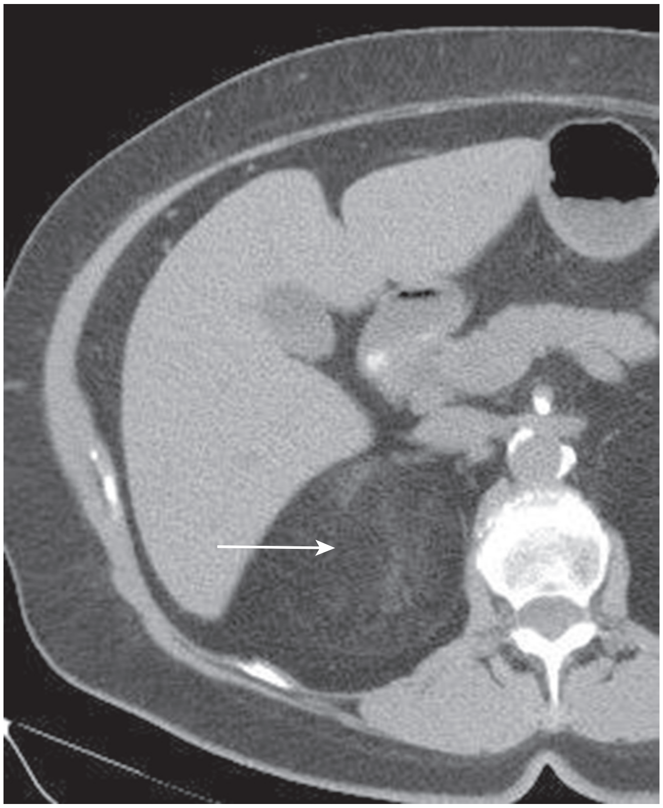


Fig. 92.10 **Myelolipoma.** Large, fatty right adrenal mass (M) is seen on this contrast-enhanced computed tomography examination.

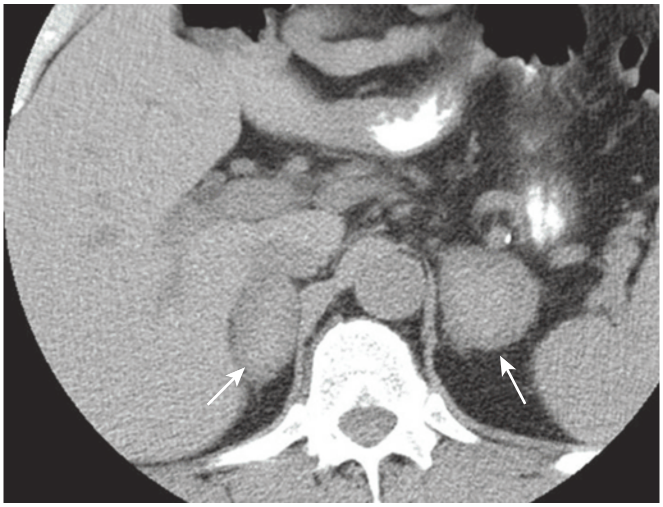


Fig. 92.11 **Adrenal insufficiency secondary to bilateral adrenal hemorrhage.** Unenhanced computed tomography shows a high attenuation mass in both adrenals *(arrows)* characteristic of acute or subacute hematoma.

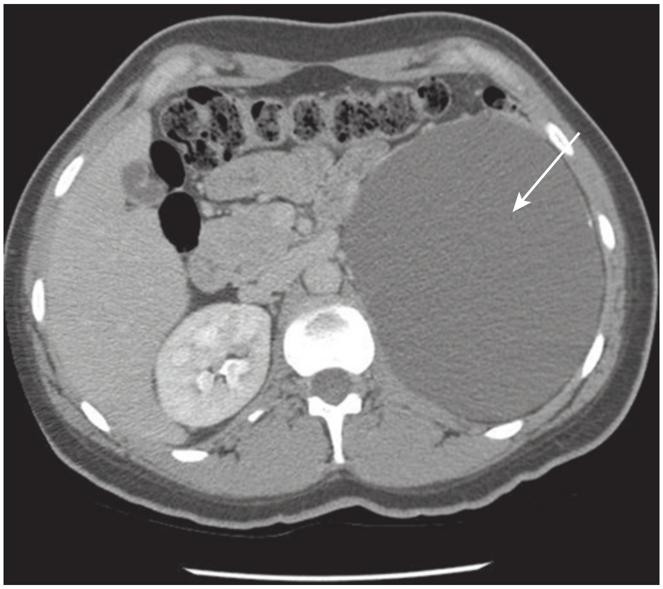


Fig. 92.12 **Adrenal cyst.** This very large homogeneous left adrenal mass measured near-water density (8 Hounsfield units). The wall is thickened but smooth and less than 3 mm.

Cyst. Adrenal cysts are uncommon lesions with a 3:1 female predilection. There are four types identified on the basis of pathologic classification: endothelial, epithelial, parasitic, and posttraumatic pseudocysts, with the endothelial cysts being the most common, accounting for 45% of adrenal cysts.[45,46,52,53] Pseudocysts are usually a sequela of prior hemorrhage and account for approximately 40% of adrenal cysts.[54]

Cysts are easily diagnosed by their imaging characteristics. On CT, cysts demonstrate a homogeneous water density (Fig. 92.12). On MRI, they tend to be T2 hyperintense and T1 hypointense and demonstrate no enhancement. Peripheral calcification is common in pseudocysts, but is rarely seen in endothelial cysts. Pseudocysts can have a complex MRI appearance, with septations, residual blood products, thicker walls, and often central hypointensity. If the cyst demonstrates soft tissue components, other causes, such as cystic metastasis and cystic pheochromocytoma, should be considered.[54]

Granulomatous Disease. Tuberculosis, histoplasmosis, and other granulomatous diseases are usually bilateral but often asymmetric. CT findings are nonspecific and include soft-tissue masses, cystic changes, and/or calcification (Fig. 92.13).[55] Although these are uncommon adrenal lesions, they should be considered in the differential diagnosis of incidental bilateral adrenal masses in the absence of a primary neoplasm or coagulopathy. Biopsy can confirm the diagnosis and identify the responsible organism.

Infectious and inflammatory conditions affecting the adrenals are another cause of false positive findings on FDG-PET due to lymphocytes and neutrophils mediating the PET uptake. Infectious FDG activity in the adrenals has been reported with cases of cryptococcus,[56,57] histoplasmosis,[58,59] adrenal tuberculosis,[60-62] and candidiasis.[63]

Hemangioma. An adrenal hemangioma is a rare, benign tumor. Hemangiomas are comprised of closely adjacent vascular channels lined with a single layer of endothelium. On CT, hemangiomas are seen as large, well-defined masses of soft tissue density with calcifications due to phleboliths or prior hemorrhage, and demonstrate heterogeneous contrast enhancement.[51,64]

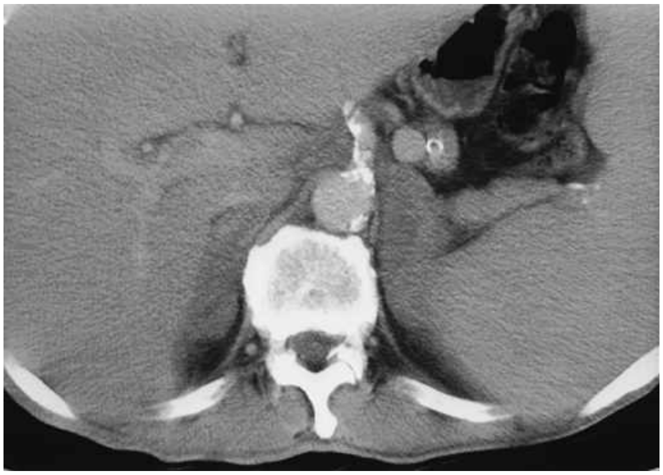


Fig. 92.13 **Histoplasmosis.** Both adrenal glands are enlarged but maintain adreniform configuration. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol.* 2002;179:559–568.)

MRI findings include a hypointense appearance relative to the liver on T1-weighted sequences, interspersed with areas of high signal intensity due to hemorrhage. On T2-weighted sequences, hemangiomas are hyperintense. Peripheral nodular enhancement that persists on delayed phases is characteristic. Tumors are usually removed because of the increased risk for hemorrhage and the inability to exclude malignancy.[65] Adrenal hemangioma will have absent uptake on FDG-PET.[66]

Ganglioneuroma. A ganglioneuroma is a benign tumor composed of Schwann cells and ganglion cells. These tumors occur anywhere along the paravertebral sympathetic plexus, and approximately 20% to 30% arise in the adrenal medulla.[67] Because they do not secrete hormones, most ganglioneuromas are detected as an incidental finding. On CT, ganglioneuromas appear as a solid adrenal mass measuring up to 11 cm. Clinically silent, extraadrenal retroperitoneal tumors may be even larger. Ganglioneuromas are typically negative with FDG-PET due to their low proliferative rate.[68]

Hypoadrenalism. Adrenal insufficiency is most commonly due to autoimmune atrophy, and imaging usually is not useful for diagnosis; the glands are small and atrophic (Fig. 92.14). In patients with adrenal insufficiency due to granulomatous infection from tuberculosis, histoplasmosis, or blastomycosis, the acute and subacute phases of disease are associated with bilateral adrenal enlargement. Heterogeneous low attenuation within the adrenal mass on enhanced CT or heterogenous high T2 signal intensity on MRI is due to caseous necrosis.[69] Unlike in patients with adrenal metastases, the enlarged glands typically retain their normal shape.

Adrenal insufficiency can also be associated with bilateral adrenal hemorrhage, and the detection of bilateral hyperattenuating masses on unenhanced CT can sometimes be the first clue to this diagnosis (Fig. 92.11). Adrenal insufficiency can also be caused by several systemic diseases, such as acquired immunodeficiency syndrome and antiphospholipid antibody syndrome. The appearance of the adrenal glands in these diseases can be variable.[70,71]

Neoplastic Adrenal Lesions

Adrenocortical Carcinoma. Adrenocortical carcinomas (ACCs) are rare tumors with a reported incidence of two cases per million.[72,73] Patients may present with abdominal pain, a palpable mass, or Cushing

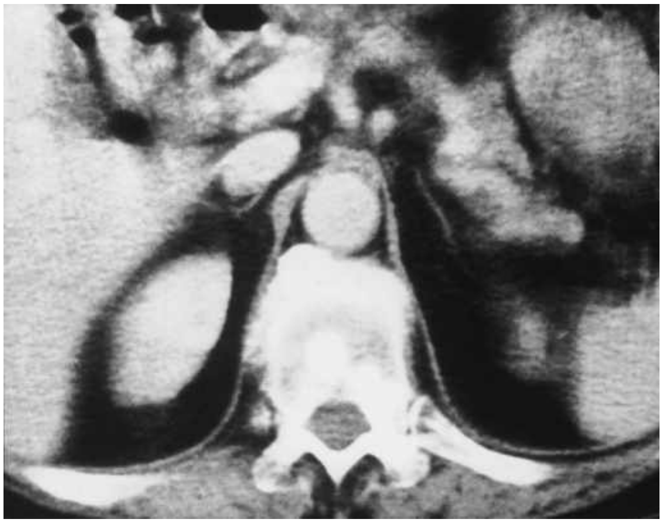


Fig. 92.14 Computed tomography shows markedly atrophic adrenal glands in a patient with adrenal insufficiency secondary to idiopathic autoimmune atrophy. (From Korobkin M, Francis IR. Adrenal imaging. *Semin Ultrasound CT MR.* 1995;16:317–330.)

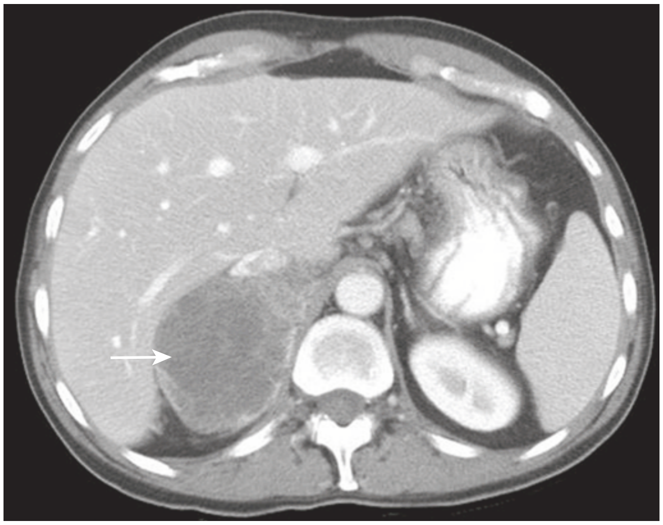


Fig. 92.15 Adrenal cortical carcinoma causing Cushing syndrome. Computed tomography scan shows a large, inhomogeneous, right adrenal mass (M).

syndrome, as approximately 50% of these tumors are functional and secrete cortisol or other steroid hormone metabolites. Other endocrine manifestations of ACC include aldosteronism, virilization, and feminization, but these are very rare. Tumors tend to be very large at the time of presentation. For this reason, any adrenal lesion larger than 4 cm in size, regardless of imaging characteristics, requires surgical consultation due to the risk of a possible underlying ACC.

The CT and MRI appearance of an ACC is a large, heterogeneously enhancing mass, commonly demonstrating central necrosis (Fig. 92.15). Calcification can be seen in 20% to 30% of cases.[72,74] Only a small number of patients with ACC have been studied with CT densitometry or chemical shift MRI. In one series that included 11 ACCs, absolute and relative enhancement loss on 10-minute delayed imaging was significantly lower than for the adenomas and similar for metastases and pheochromocytomas. In another series of seven patients with ACC, delayed enhanced images at approximately 20 minutes showed a relative percentage of enhancement washout of less than 40%, consistent with malignancy.[75]

Venous extension of tumor into the renal veins or the IVC can be identified on contrast-enhanced studies (Fig. 92.16). It is important to precisely define the cephalad extent of the intravenous tumor, as this defines the point where the surgeon can gain vascular control of the tumor.[76]

It is important to remember that a large adrenal metastasis, or even a pheochromocytoma, can have an imaging appearance identical to that of an ACC. In the absence of metastases, differential diagnosis is usually carcinoma versus adenoma. The larger the mass, the greater the likelihood of carcinoma, although on rare occasions adenomas can be greater than 5 cm and can have large regions of hemorrhage, necrosis, and calcification.

ACC has been consistently demonstrated to display high metabolic activity on FDG-PET imaging. Due to its poor prognosis, marked by high incidence of disseminated disease (in 50% of patients) at presentation, and the high rate of recurrence after initial complete resection (73%–86% at 2 years), patients undergo extensive and serial imaging evaluation for staging and restaging of disease. In a study of 68 patients with adrenal malignancy that included ACC, pheochromocytoma, NB, and lymphoma, the authors reported that FDG-PET had better accuracy than CT (93.4% vs. 75%) for ACC, and that SUVmax was a significant predictor of outcomes in these malignant adrenal masses.[77]

In 48 patients with confirmed ACC recurrence, FDG-PET had lower sensitivity than CT for liver and lung recurrence, although it had higher specificity than CT and better positive predictive value.[78] Therefore, FDG-PET is a useful complementary tool in managing surveillance of ACC after primary surgery and management. As mentioned earlier, ACC is a malignant highly aggressive neoplasm with propensity for tumor invasion into the IVC, causing pulmonary embolism, which can be depicted on FDG-PET/CT.[79] With respect to prognosis, Satoh and colleagues evaluated 30 patients with ACC using metabolic tumor volume (MTV), total lesion glycolysis (TLG), and SUVmax and found that patients with high MTV (>87 mls), TLG (>229), or SUVmax (>8.8) had worse outcomes for both progression-free and overall survival.[80]

Overexpression of chemokine receptor 4 (CXCR4) in ACC is a potential diagnostic and therapeutic receptor target.[81] FDG-PET was compared to ^{68}Ga-pentixafor PET, with both studies performed within a week of each other. In 30 patients with disease, the PET studies in five (17%) were complementary and in 10 (33%) were comparable. In 13 (43%), more lesions were found with FDG-PET than with ^{68}Ga-pentixafor PET. Overall, 70% of patients with advanced metastases from the ACC may be suitable for CXCR4-directed therapy (radiotherapy or pharmacotherapy). CXCR4 positivity in ACC correlates with expression and quantification of metastases.[82]

Pheochromocytoma. Adrenal pheochromocytomas and paragangliomas, arising from extraadrenal tissues, are rare tumors derived from the sympathomedullary cells of postganglionic sympathetic neurons. The majority arise within the adrenal medulla or in an extraadrenal location near the celiac axis or aortic bifurcation (organ of Zuckerkandl). Those that are catecholamine-secreting result in the classic signs and symptoms of hypertension, tachycardia, headaches, palpitations, diaphoresis, and chest pain. Although many patients present with some or all of these manifestations of excess catecholamine production, approximately 8% to 10% of tumors are silent and incidentally detected by imaging studies. A significant proportion of these tumors are found to be malignant once evidence of metastatic spread is documented, usually with imaging.

Although pheochromocytomas were traditionally taught as following the rule of 10s (10% being multiple, 10% being inherited, 10% being extraadrenal in location, and 10% being malignant), more

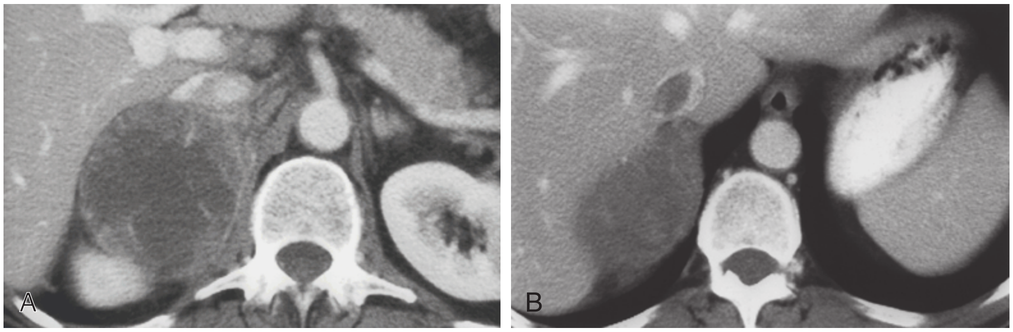


Fig. 92.16 **Adrenal cortical carcinoma. A,** Contrast-enhanced computed tomography examination shows a 9-cm right adrenal mass. Irregular wall and low-density center indicate necrosis. **B,** More cephalad image in same examination reveals tumor extension into inferior vena cava. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol.* 2002;179:559–568.)

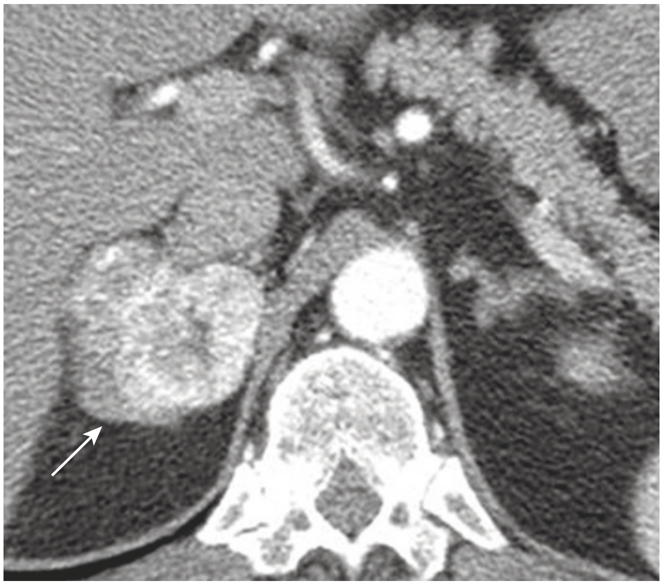


Fig. 92.17 Intravenous contrast-enhanced computed tomography shows a right adrenal pheochromocytoma (P) with areas of necrosis.

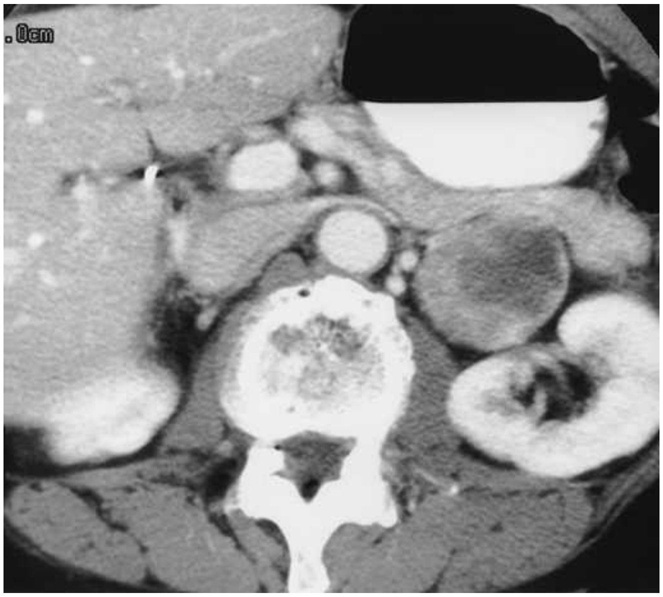


Fig. 92.18 **Retroperitoneal extraadrenal pheochromocytoma.** Enhanced computed tomography shows an inhomogeneous enhancing mass medial to the left kidney. A normal left adrenal was demonstrated on scans cephalad to this mass.

contemporary understanding of genetic predisposition has identified closer to 40% of cases as being inherited and associated with disorders including multiple endocrine neoplasia (MEN) syndromes, neurofibromatosis, von Hippel–Lindau (VHL), or familial pheochromocytoma.[83]

The diagnosis of pheochromocytoma can be confirmed by elevated blood or urine levels of the catecholamines epinephrine and norepinephrine or their metabolites. Localization of a pheochromocytoma is essential, because surgical resection of the tumor can be curative. On CT, pheochromocytomas typically range from 2 to 5 cm and demonstrate heterogeneous attenuation (usually >10 HU on unenhanced images) due to necrosis or hemorrhage, particularly when they are larger in size. Imaging of pheochromocytomas with contrast-enhanced CT has historically been associated with contrast-induced hypertensive crisis; however, use of nonionic contrast agents has eliminated this concern. On contrast-enhanced CT, pheochromocytomas typically demonstrate heterogeneous enhancement similar to that seen in ACC or metastases (Fig. 92.17). Oral contrast opacification of the gastrointestinal tract can be occasionally helpful for the detection of extraadrenal pheochromocytomas/paragangliomas, because unopacified bowel may rarely be mistaken for a "mass" and may mimic these tumors (Fig. 92.18).[84] Unusual locations of paragangliomas have also been reported, such as the rare, intrapericardial paraganglioma adjacent to or involving the left atrium (Fig. 92.19).

On MRI, most pheochromocytomas are hypointense on T1-weighted images and markedly hyperintense on T2-weighted images. There is considerable overlap of signal on T2-weighted images with other neoplasms, including ACC, in up to 33% of cases. This overlap on T2-weighted imaging may be related to necrotic or cystic areas that are commonly seen in these tumors. Contrast-enhanced MRI is useful for demonstrating the presence or absence of intracaval extension of adrenal pheochromocytoma. MRI is also useful for the evaluation of extraadrenal paragangliomas that can be seen from the neck to the bladder.

^{123}I-MIBG can be used to image pheochromocytoma and paraganglioma with high specificity, thus confirming their origin from the neural crest, with the additional benefit of whole-body screening to identify multifocal and/or remote metastatic disease (Fig. 92.20). A metaanalysis found MIBG scintigraphy to have high sensitivity (94%) and specificity (95%) for imaging pheochromocytoma/paraganglioma.[85] MIBG scintigraphy is clinically useful in depicting diagnostically elusive and often small paragangliomas because of their close relationship to other structures and remote locations (extraadrenal sites), screening from the base of the skull to the pelvis (Fig. 92.21). In malignant pheochromocytoma/paraganglioma studied with MIBG, the most common sites of metastatic disease are the skeleton, lymph nodes, lung, and peritoneum. Importantly, MIBG imaging is not particularly useful in head and neck paragangliomas, as these are tumors are of parasympathetic origin that lack chromaffin cells and are nonsecretory; these neoplasms are better imaged with ^{68}Ga-DOTAtate or ^{18}F-DOPA PET.[86]

A range of PET radioisotopes such as ^{18}F-FDG, ^{18}F-DA, ^{18}F-DOPA, and ^{11}C-HED (Fig. 92.24) have been successfully used to image pheochromocytoma and paraganglioma and can be useful in cases when the disease does not show strong MIBG uptake.[87] Most pheochromocytomas, whether benign or malignant, can be imaged with FDG-PET, although focal FDG uptake is found in greater percentage of malignant than benign pheochromocytomas. Shulkin et al. reported that 22 of 29 patients with pheochromocytoma could be identified with FDG-PET, and that MIBG-negative pheochromocytomas were depicted with FDG, a finding that has been confirmed by others.[88] Timmers et al. found that FDG-PET had a sensitivity of 77% and a specificity of 90% in 216 patients with suspected pheochromocytoma/paraganglioma and outperformed both MIBG and CT/MRI.[89] In one study, the combination of FDG-PET and apparent diffusion coefficient on MRI was useful to separate 11 pheochromocytomas from 22 other benign tumors.[90] Preoperative PET may find extra lesions prior to surgical resection of pheochromocytoma.[91]

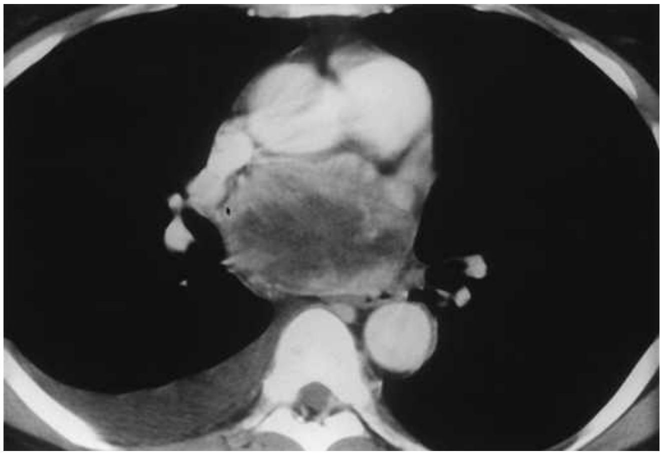


Fig. 92.19 Intrapericardial extraadrenal pheochromocytoma. Enhanced computed tomography shows an inhomogeneous mass in the expected location of the left atrium. Metaiodobenzylguanidine scan was positive in this location. (From Hamilton BH, Francis IR, Gross BH, et al. Intrapericardial paragangliomas [pheochromocytomas]: Imaging features. *Am J Roentgenol.* 1997;168:109–113.)

Metabolic subtyping of pheochromocytoma with FDG-PET has been reported.[92] Metabolic uptake is higher for malignant than benign pheochromocytoma and for extraadrenal paraganglioma than adrenal sites. The degree of FDG uptake in pheochromocytoma depends on genetic mutation. In genetic syndromes, cluster 1 genes (*SDHx, VHL*) have higher FDG uptake than cluster 2 genes (*RET, NF1*).[93] The use of FDG-PET for screening for *SDHx* mutation carriers has high sensitivity and intermediate specificity,[94] and FDG-PET use has been explored for patients with VHL, patients with MEN2, and for screening for pheochromocytoma in NF1 patients.[95]

Fottner et al. reported an ^{18}F-DOPA-PET sensitivity of 98% (62/64 lesions) and specificity of 100%, much higher compared to an MIBG sensitivity of only 53% (34/64 lesions) and specificity of 91%.[96] ^{18}F-DOPA-PET was reported to be markedly avid in a more recently described genetic predisposition due to MYC-associated factor X (MAX) pheochromocytoma.[97]

Studies of ^{18}F-DA-PET for imaging pheochromocytoma/paraganglioma show high sensitivity and specificity.[98] Discordant results with MIBG, ^{18}F-FDG, ^{18}F-DA, and ^{18}F-DOPA imaging appear to depend on tumor phenotype and are most likely related to the specific underlying mutation. In one comparative study of different molecular imaging

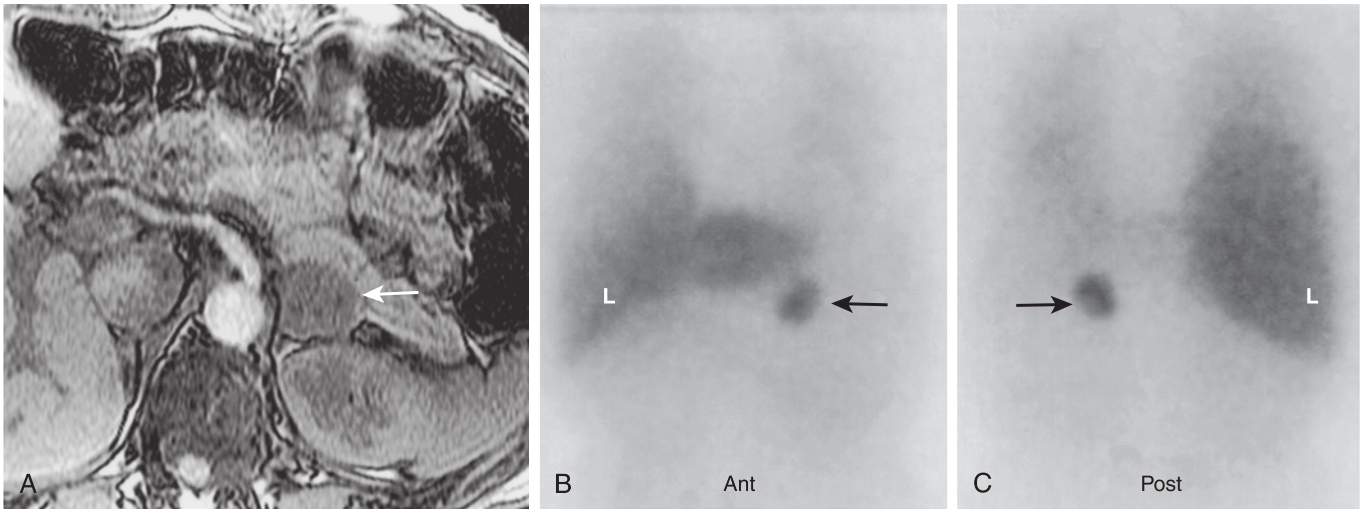


Fig. 92.20 **Left adrenal pheochromocytoma demonstrated by ^{123}I-metaiodobenzylguanidine (MIBG) and magnetic resonance imaging (MRI).** A 50-year-old female with hypertension, a history of neurofibromatosis, elevated plasma catecholamines, and a 3-cm left adrenal mass on MRI. **A,** Abdominal MRI, transverse section, with left adrenal pheochromocytoma *(white arrow)*. **B,** Anterior abdominal ^{123}I-MIBG scan with normal liver (L) uptake. The *black arrow* indicates intense, focal tracer uptake in the region of the left adrenal gland. **C,** Posterior abdominal ^{123}I-MIBG scan. The *black arrow* indicates left adrenal pheochromocytoma. (From Rubello D, Bui C, Casara D, et al. Functional scintigraphy of the adrenal gland. *Eur J Endocrinol.* 2002;l47:13–28.)

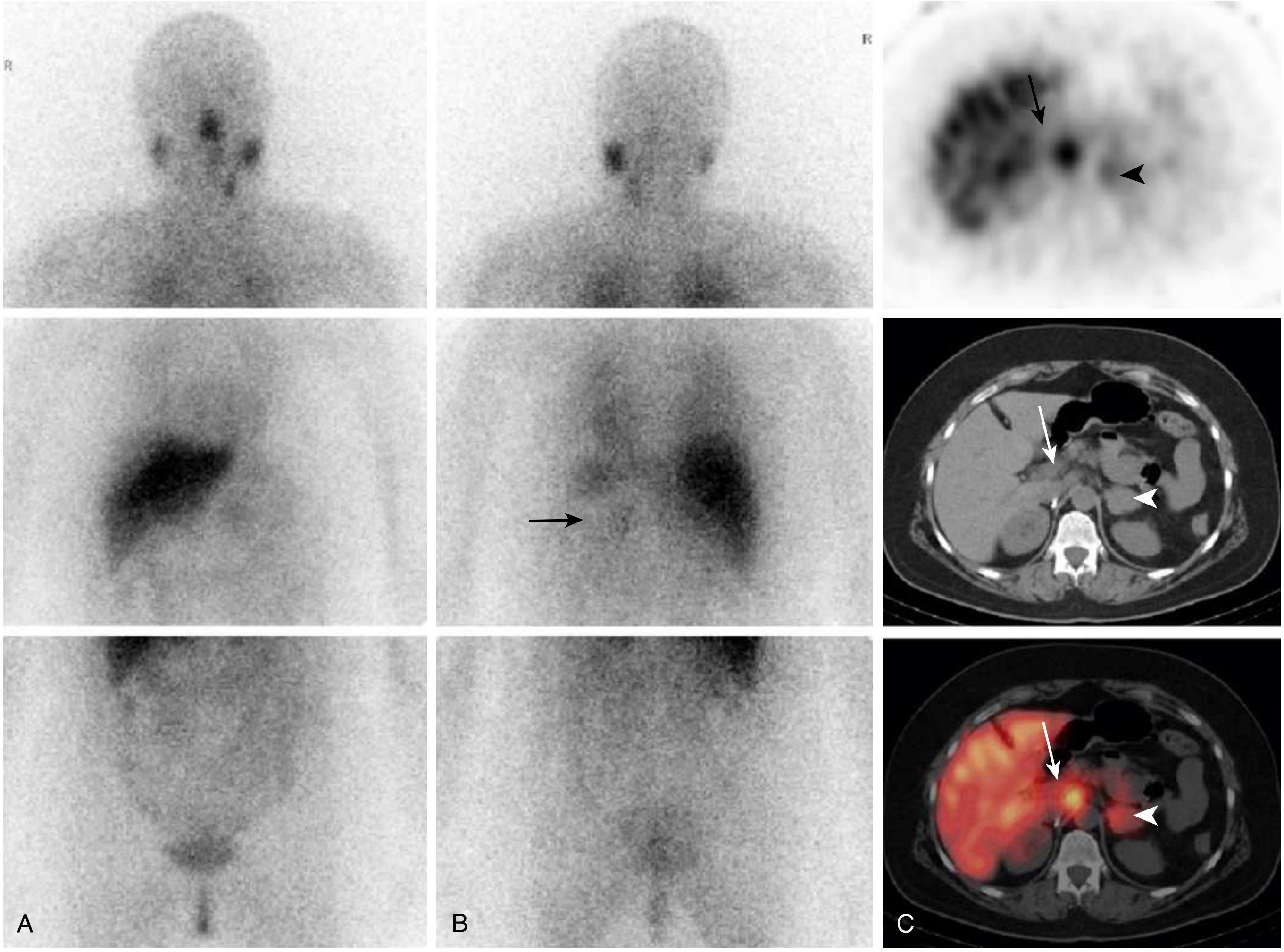


Fig. 92.21 **A 53-year-old man with uncontrolled hypertension and catecholamine hypersecretion, suspected for pheochromocytoma.** ^{123}I-metaiodobenzylguanidine (MIBG) composite anterior (**A**) and posterior (**B**) whole-body images at 24 hours postinjection show focal intense activity in the region of the left adrenal gland *(arrow)*. **C,** Axial computed tomography (CT), and fused single-photon emission CT (SPECT)/CT images confirm a left adrenal pheochromocytoma with intense ^{123}I-MIBG uptake *(arrowheads)*. There is also preaortic soft tissue nodule with intense ^{123}I-MIBG uptake *(arrows)* on the SPECT/CT that may represent an extraadrenal paraganglioma or, less likely, a metastatic retroperitoneal lymph node indicating malignant disease. (From Youssef E, Wong KK. Hybrid SPECT-CT endocrine scintigraphy: a pictorial review. AUR, Annual Meeting in San Antonio, Texas, 2011.)

radiotracers, Timmers et al. investigated 30 patients with the succinate dehydrogenase subunit B (*SDHB*) germline mutation and metastatic pheochromocytoma/paraganglioma.[89] They found that ^{18}F-FDG-PET had a sensitivity of 100% that exceeded ^{18}F-DOPA (88%), MIBG (80%), and somatostatin receptor scintigraphy (SRS) (81%). At least 90% of the lesions that were negative on ^{18}F-DOPA and MIBG imaging were depicted with ^{18}F-FDG, making it the imaging agent of choice in patients with *SDHB* germline mutations.

PET imaging of ^{68}Ga-DOTA-conjugate peptides such as ^{68}Ga-DOTAtoc (DOTA^0-1NaI^3-octreotide), ^{68}Ga-DOTAnoc (DOTA^0-Tyr^3-octreotide) and ^{68}Ga-DOTAtate (DOTA^0-Tyr^3-octreotate) has brought dramatic improvement in sensitivity and spatial resolution and is preferred for NET localization and staging studies. Further advantages are same-day imaging protocols and availability of whole-body nondiagnostic CT component, use of newer-generation DOTA chelator with improved affinity for SSTR2, and ability to predict efficacy of ^{177}Lu or ^{90}Y-radionuclide labeled somatostatin analogs. ^{68}Ga-DOTANOC-PET imaging of pheochromocytoma has been reported and can successfully image metastatic pheochromocytoma/paraganglioma in MIBG-negative patients (Fig. 92.22).[10,99] Widespread availability of ^{68}Ga-DOTANOC-PET since 2016 in North America has resulted in it largely replacing ^{111}In-octreotide for imaging of NETs. It has high sensitivity for pheochromocytoma/paraganglioma; however, specificity is limited, as many nonendocrine tumors (e.g., breast cancer, meningioma, lymphoma), granulomatous and inflammatory diseases (e.g., sarcoidosis, tuberculosis), and autoimmune diseases (rheumatoid arthritis) express type 2 somatostatin receptors. Somatostatin receptor PET imaging is most useful when MIBG scanning is negative in metastatic pheochromocytoma/paraganglioma or when other neuroendocrine neoplasms are suspected (Fig 92.23).[95,100,101]

Lymphoma. Primary lymphoma of the adrenal glands is rare, and secondary involvement of the adrenals is seen more often in patients with non-Hodgkin lymphoma than in Hodgkin lymphoma.[102] In secondary lymphoma, adrenal involvement is often bilateral, and other retroperitoneal disease is usually present. On CT, there is homogeneous

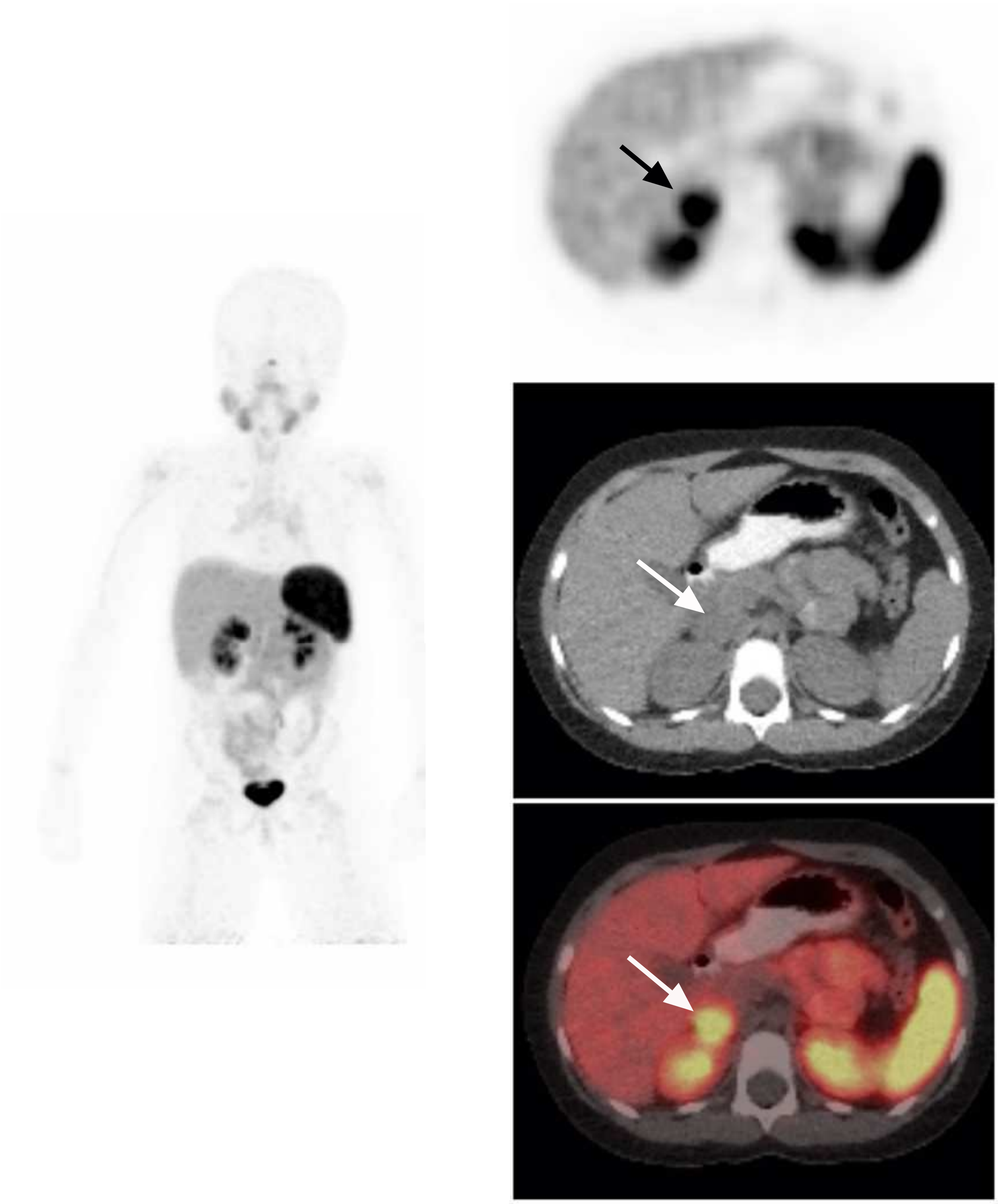


Fig. 92.22 **^{68}Ga-DOTA-tate positron emission tomography/computed tomography scan in a patient with elevated catecholamines and a right-sided adrenal mass.** Pheochromocytoma was confirmed on adrenalectomy specimen.

enlargement of the adrenal gland with variable enhancement.[103] On MRI, lymphoma is characterized by hypointense signal on T1-weighted images and heterogeneous hyperintensity on T2-weighted images with minimal postcontrast enhancement.

FDG-PET/CT is a useful imaging modality for both secondary and primary adrenal lymphoma.[104-108] In secondary lymphoma, the adrenal masses typically have limited washout and high SUV values.[109] They are most often due to diffuse large B cell lymphoma and are well-defined masses with preserved adreniform shape, having mean SUV 18.6 and size greater than 3 cm. FDG-PET was reported in 28 patients with primary adrenal lymphoma, with disease usually manifesting as intensely hypermetabolic adrenal masses. FDG-PET provides prognostic information for primary adrenal lymphoma,[110] with poor overall survival of 61.9% at 2 years despite chemotherapy (Fig 92.25).[111]

Metastases. The adrenal glands are a common site of metastatic disease, which is found in approximately 27% of postmortem examinations of patients with malignant neoplasms of epithelial origin.[112] The most common neoplasms that metastasize to the adrenal glands are carcinomas of the lung and breast, and melanoma. Small metastases

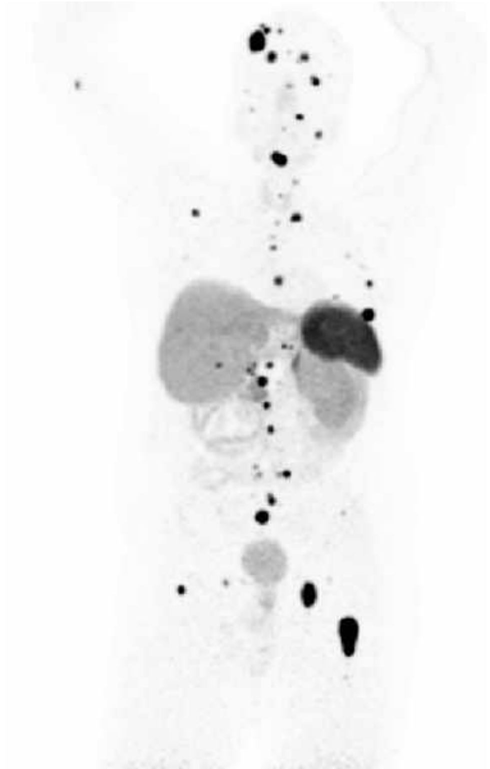


Fig. 92.23 ^{68}Ga-DOTA-tate positron emission tomography maximum intensity projection images in a 25-year-old man with metastatic paraganglioma, demonstrated intensely avid uptake in multiple osseous lesions scattered in the skeleton.

(Figs. 92.26 and 92.27) are often homogenous on contrast-enhanced CT or MRI, whereas larger metastases often appear heterogeneous due to necrosis and/or hemorrhage. Calcification is rare in adrenal metastases. Intracellular lipid can be seen in metastases from clear cell renal cell carcinoma and hepatocellular carcinoma. In addition, adrenal washout characteristics need to be interpreted with caution in patients with suspected metastases from hepatocellular carcinoma, renal cell carcinoma, and NET, as these lesions may demonstrate washout characteristics similar to that of benign adrenal adenomas.[113]

In a metaanalysis for staging/restaging lung cancer that included 707 patients from nine studies, metabolic imaging with FDG-PET had a pooled sensitivity of 88% and specificity of 90.8% for detection of metastatic disease.[114] It is recognized that in cancer patients with oligometastatic disease to the adrenal (i.e., solitary metastasis to the adrenal) adrenalectomy can have a survival advantage.[115,116] In a retrospective study of 289 patients undergoing adrenalectomy, 39 (13%) had suspected oligometastatic disease due to a positive adrenal FDG-PET.[117] In 28/39 (71.8%) of these patients, the oligometastasis was confirmed at adrenalectomy; history of primary lung cancer and SUVmax greater than 2.65 increased the odds ratio of metastatic disease. However, notably in this series, ten benign cortical adenomas and one nonfunctioning pheochromocytoma had positive PET findings, leading to a 28.2% false-positive rate in this setting (Figs. 92.28 and Fig 92.29).

Neuroblastoma. Neuroblastoma (NB), a malignant tumor composed of primitive neuroblasts, is the most common extracranial solid malignancy of children, representing 8% to 10% of pediatric cancers. The mean age of diagnosis is 15 months, and the majority occur in infancy and early childhood.[118] While these tumors may occur anywhere along the sympathetic plexus, the most common site for primary NB is the adrenal gland, accounting for approximately 35% to 40% of cases.[119] Unfortunately, metastases are present in over half of pediatric patients at diagnosis.

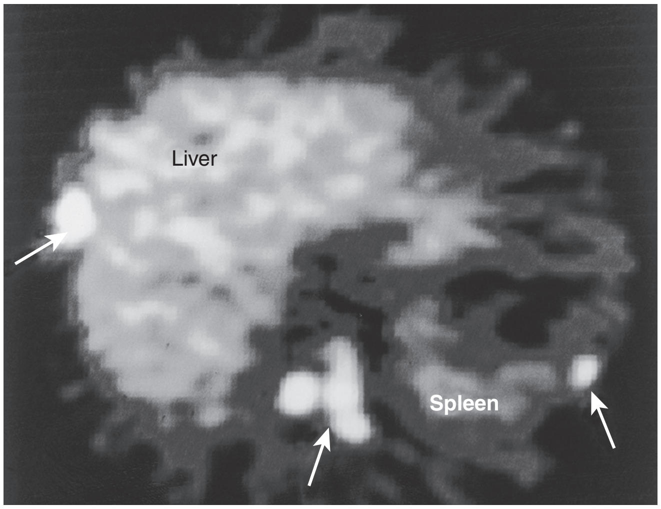


Fig. 92.24 ^{11}C-Hydroxyephedrine (^{11}C-HED) positron emission tomography scan in a patient with malignant pheochromocytoma. Multiple areas of increased uptake are seen in the liver, abdomen, and spleen *(arrows)*. (From Gross MD, Shapiro B. Adrenal scintigraphy. In: Khalkhali I, Maublant J, Goldsmith S, eds. *Nuclear Oncology*. Philadelphia: Lippincott, Williams & Wilkins; 2000:472.)

NB in adults, though rare, usually have imaging findings that are similar to those in children, with disseminated disease at presentation being common. On CT, NB appears as a large, irregular, and heterogeneous mass due to necrosis and/or hemorrhage. Calcification is common and can be seen in 80% to 90% of cases.[120] On MRI, NB typically demonstrates heterogeneous hypointense signal on T1-weighted images and hyperintense signal on T2-weighted images with variable and heterogeneous enhancement. NB can be aggressive and infiltrative, causing narrowing and encasement of adjacent vessels. Invasion of adjacent soft tissues and organs can also occur.[121]

^{123}I-MIBG diagnostic imaging is an established modality for diagnosis, staging, and treatment response assessment of NB, as well as for determining suitability of targeting of metastatic disease for subsequent ^{131}I-MIBG treatment.[122] ^{123}I-MIBG is more accurate than bone scan for detection of osseous metastatic disease in NB.[123] More recently, ^{68}Ga-DOTATATE-PET/CT has been found to be positive in just over two thirds of patients with refractory NB, in six of eight patients in each of two small pilot studies.[124,125] Furthermore, ^{68}Ga-DOTATATE-PET/CT detected additional sites of disease compared with ^{123}I-MIBG in three out of eight patients.[125] ^{68}Ga-DOTATATE-PET/CT imaging has that potential to inform on suitable targeting of radionuclide treatment of NB with 177Lu-DOTA peptides.[124]

Cushing Syndrome. Cushing syndrome results from excessive glucocorticoids, resulting in characteristic clinical signs and symptoms. Cushing syndrome refers to the clinical and metabolic disorder regardless of the underlying cause. The most common etiology is iatrogenic steroid administration. Endogenous Cushing syndrome is due to overproduction of cortisol by the adrenal cortex, caused by (1) excess production of ACTH by a pituitary tumor; (2) benign or malignant steroid-producing adrenocortical neoplasm; or (3) bilateral adrenal hyperplasia secondary to an ectopic source of ACTH production. Very rare causes of ACTH-independent bilateral hyperplasia have also been described. Strictly speaking, Cushing disease refers only to

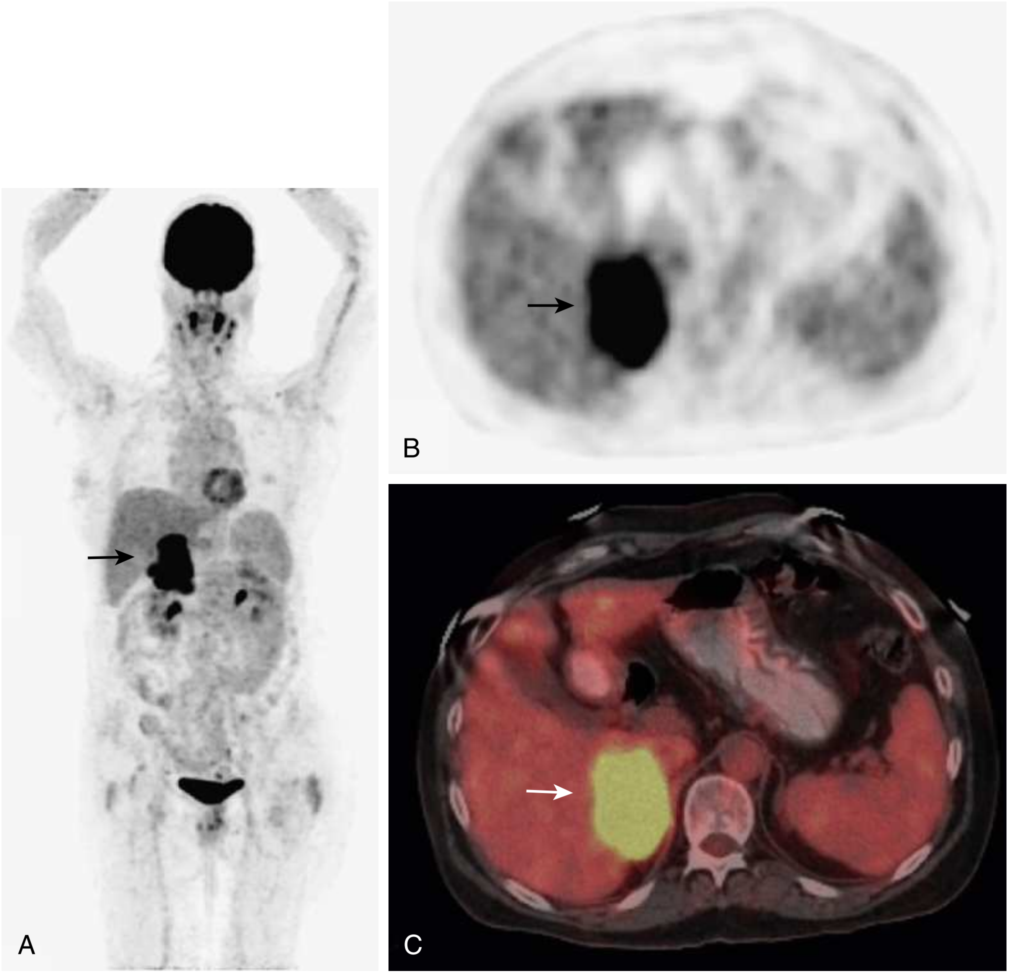


Fig. 92.25 **(^{18}F)-fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT) in a 53-year-old woman with chronic myeloid leukemia, status post–bone marrow transplantation.** Maximum intensity projection image (**A**), axial PET (**B**), and fusion PET/CT (**C**) demonstrate a 6.5 x 5.5 cm right suprarenal mass *(arrows)* with intensely increased FDG uptake (maximum standard uptake value = 17.1) above liver background (liver mean standard uptake value = 2.8). Histopathology specimen of the resected mass was compatible with a hematolymphoid malignancy, an extramedullary myeloid tumor. (From Wong KK, Arabi M, Bou-Assaly W, et al. Evaluation of incidentally discovered adrenal masses with PET and PET/CT. *Eur J Radiol.* 2012;81:441–450.)

bilateral adrenal hyperplasia due to ACTH hypersecretion by a pituitary adenoma.

Most cases of Cushing syndrome (up to 85%) are due to excess ACTH production from pituitary or ectopic sources. The adrenal glands may be normal or show diffuse bilateral gland thickening and nodularity on CT, a finding indicative of hyperplasia. A small percentage of patients (12%–15%) with Cushing disease demonstrate multiple or, less commonly, single macronodules, varying in size from several millimeters to as much as 7 cm in diameter.[126] If macronodular adrenal hyperplasia is characterized by a single dominant nodule, this entity may be confused with a unilateral autonomous adrenal adenoma, leading to inappropriate unilateral adrenalectomy. Additional small nodules or diffuse overall enlargement of both glands are usually present; however, correlation with the biochemical findings allows the correct diagnosis to be made.

A much rarer form of macronodular adrenal hyperplasia is ACTH-independent disease. Two unique CT and MRI features of this disorder are the large mass of cortical tissue and the size of individual nodules (Fig. 92.30).[127] The size of the nodules can suggest alternate diagnoses such as bilateral metastases or bilateral adenomas, but in the presence of Cushing syndrome, the appearance is practically pathognomonic of ACTH-independent macronodular adrenal hyperplasia, and bilateral adrenalectomy is indicated as optimal therapy, on the basis of clinical and CT findings.

Approximately 15% of cases of ACTH-dependent Cushing syndrome are due to ectopic ACTH secretion from a nonpituitary source. The most common sources of ectopic ACTH-producing tumors are small cell lung carcinomas and bronchial carcinoids. However, the diagnosis of patients with occult sources of ACTH secretion presents a difficult diagnostic challenge. In one study, five of eight bronchial carcinoid tumors measured between 4 and 10 mm in diameter, and thin-section CT was required to detect these small tumors. Less common sources of ectopic ACTH production include elusive islet cell tumors of the pancreas, pheochromocytomas, medullary thyroid carcinomas, and thymic carcinoids. Another study that included molecular imaging found that the source of ectopic ACTH-secreting tumors in descending order was the lung (55.3%), pancreas (8.5%), mediastinum thymus (7.9%), adrenals (6.4%), and gastrointestinal tract (5.4%). Bronchial NETs were the most common (54.8%) etiology. However, imaging was only moderately successful in localizing

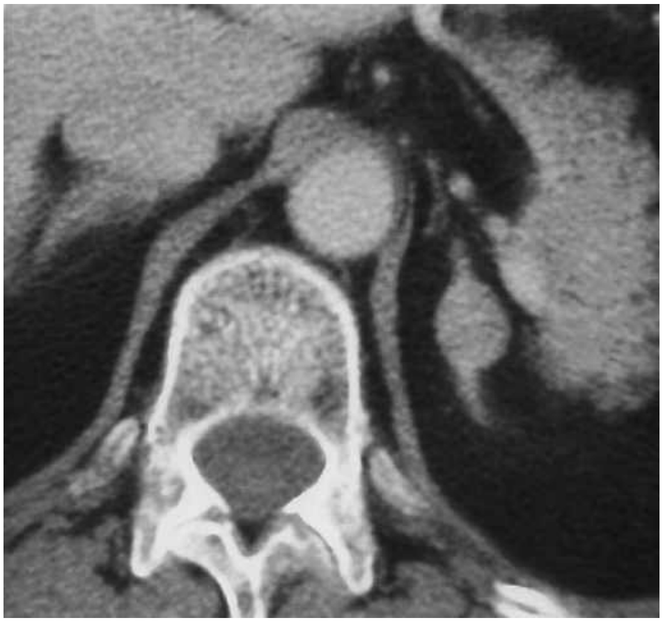


Fig. 92.26 Metastasis from renal carcinoma in a 31-year-old woman. Small, homogeneous left adrenal mass is seen on this contrast-enhanced computed tomography examination. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol.* 2002;179:559–568.)

the tumor site (CT 66.2%, MRI 51.5%, FDG-PET 51.7%, 18F-DOPA 57.1%, and MIBG 30.8%).[128] The best imaging modality for bronchial carcinoid and other NETs causing ectopic ACTH secretion was ^{68}Ga-DOTATATE-PET, with a sensitivity of 81.8%,[129] whereas FDG-PET performs better for detection of small cell lung cancer.[130]

Up to 30% of cases of Cushing syndrome are due to an ACTH-independent adrenocortical neoplasm; approximately two-thirds of these are due to adrenal adenomas, and the other one-third to ACCs. These tumors are easily detectable on both CT and MRI. Adrenocortical adenomas are nearly always less than 5 cm in diameter, typically 2 to 2.5 cm in size, and have a nonspecific morphologic appearance (Fig. 92.31).

Primary Aldosteronism. Primary aldosteronism (PA) is characterized by moderate to severe hypertension caused by unregulated secretion of aldosterone with elevated levels of serum and urinary aldosterone, hypokalemia, and suppressed plasma renin activity. A solitary aldosterone-producing adenoma (APA) is present in approximately 70% of patients, and surgical or laparoscopic adrenalectomy corrects hypertension and hypokalemia in 75% to 90% of cases (Fig. 92.32).[131] The majority of the remainder of patients have idiopathic hyperaldosteronism (IHA) from bilateral adrenal hyperplasia. Unlike in patients with unilateral APA, adrenalectomy rarely cures hypertension and biochemical abnormalities, and as a result these patients are managed medically. It is essential, therefore, to distinguish APA from IHA at the outset.[132]

CT evaluation is often problematic, because unilateral aldosteronomas may be associated with ipsilateral or contralateral nonfunctioning adenomas, resulting in misdiagnosis of "nodular" adrenal hyperplasia.[133] In addition, very rarely, bilateral hyperplasia may have a predominant unilateral macronodule and contralateral gland thickening that may lead to an erroneous diagnosis of a unilateral aldosteronoma.

Doppman and colleagues[134] demonstrated that CT cannot reliably differentiate APA from IHA whenever bilateral adrenal nodules are identified. In their series, six of 21 patients with APA were incorrectly diagnosed on CT as having IHA due to the presence of non-aldosterone-secreting nodules (nonfunctioning adenomas), in addition to an aldosteronoma. Sam and colleagues demonstrated similar results in 342 patients: upwards of 22% were not correctly identified by CT.[135] Most patients with PA and a unilateral adrenal mass can proceed to unilateral adrenalectomy with successful treatment of PA. Patients with bilateral adrenal nodules, as well as those with confounding CT and biochemical evaluations, should undergo bilateral, selective AVS for aldosterone levels to identify lateralization of, and thus the source or sources of, aldosterone secretion.[136]

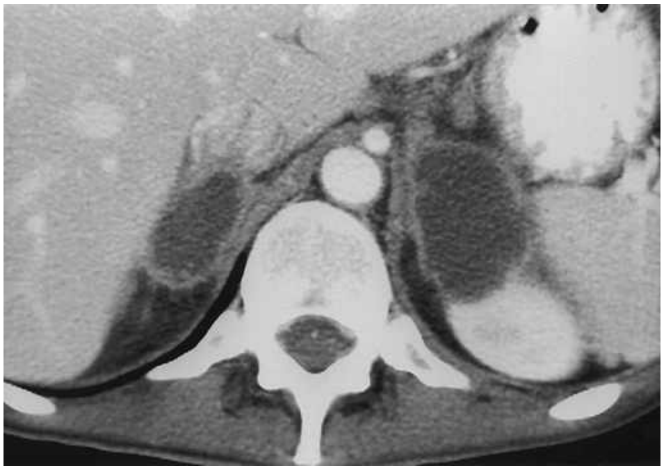


Fig. 92.27 Necrotic metastasis from adenocarcinoma of the lung in a 34-year-old woman. Bilateral adrenal masses with areas of central necrosis are seen on this contrast-enhanced computed tomography examination. (From Dunnick NR, Korobkin M. Imaging of adrenal incidentalomas: current status. *Am J Roentgenol.* 2002;179:559–568.)

Although most cross-sectional studies of patients with PA have used CT, smaller series have suggested that MRI may be able to differentiate APA from IHA.[137] Among 20 patients who underwent MRI, ten (50%) had APA, and ten (50%) had IHA. In the detection of APA, MRI had a sensitivity (70%), specificity (100%), and accuracy (85%) comparable to CT.[138,139] Traditionally, diagnosis of IHA by CT or MRI has been made by excluding the presence of an "adenoma." One CT-based study suggested that diagnosis of hyperplasia can be made by measuring adrenal gland limb width, because adrenal glands were significantly larger in bilateral hyperplasia than in APA or in normal subjects. A sensitivity for identification of IHA of 100% was achieved when a mean limb width exceeded 3 mm and a specificity of 100% when the mean limb width was greater than 5 mm. The authors proposed restricting adrenal venous sampling to situations in which mean adrenal limb width fell between 3 and 5 mm.[132]

Catheterization for venography and AVS can help identify a cause for primary hyperaldosteronism. AVS is usually performed with a percutaneous right transfemoral venous approach. Tandem sheaths are utilized, with one large enough to allow for simultaneous aspiration of iliac venous blood that represents IVC sampling. Shaped catheters are directed into right and left adrenal veins to allow for simultaneous sampling. Aliquots are collected and sent for cortisol and aldosterone assays. If there is doubt about the catheter position, rapid cortisol assay may be performed. Rapid aldosterone assay is not yet available in the United States. Gentle hand injection into the adrenal vein with simultaneous digital subtraction imaging is usually sufficient to confirm successful adrenal venous catheterization. Samples are obtained from both adrenal veins and the iliac venous or IVC sheath. After baseline samples are obtained, adrenal cortical hormonal stimulation with Cortrosyn (0.25 mg) (synthetic analogue

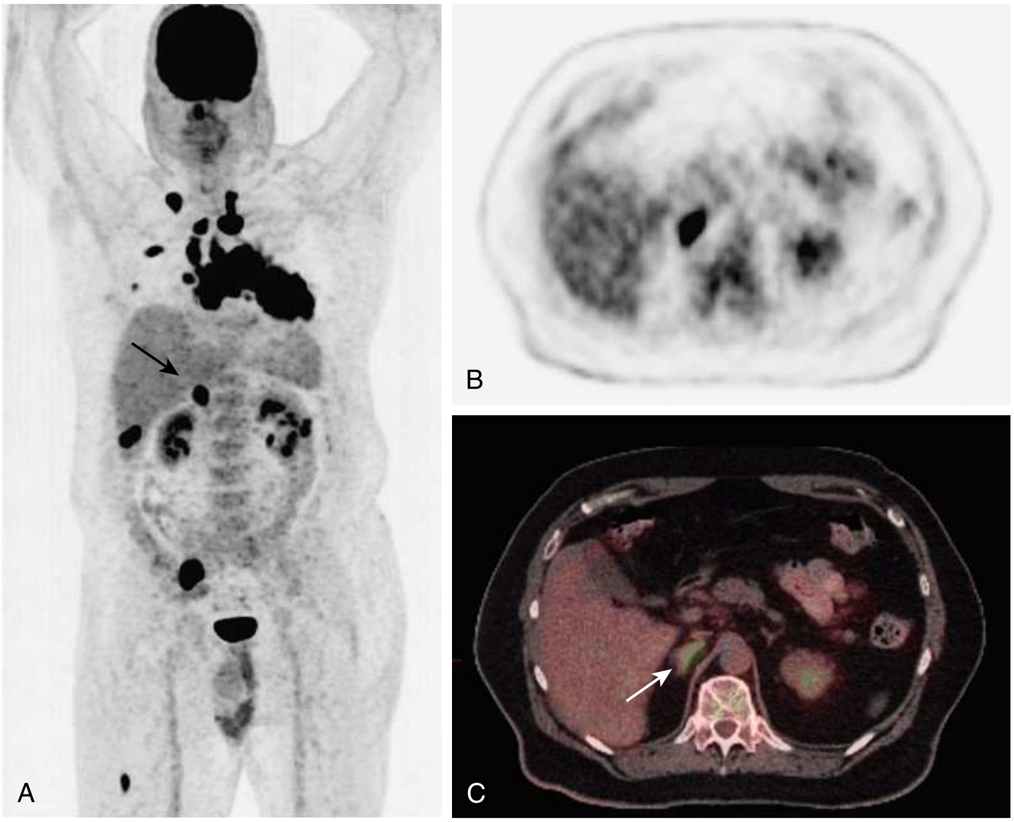


Fig. 92.28 **(^{18}F)-fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT) study for staging of lung cancer.** Maximum intensity projection (**A**), axial PET (**B**), and fused PET/CT (**C**) images show thickening of the right adrenal gland *(arrows)* with an increased FDG uptake of similar intensity to the primary left lower lobe tumor. The adrenal maximum standard uptake value (SUV) of 28 compared with the liver background SUV mean of 2.6 is compatible with metastatic disease to the adrenal gland, in the presence of widespread metastases. (From Wong KK, Arabi M, Zerizer I, et al. Role of PET/CT in adrenal and neuroendocrine tumors: FDG and non-FDG tracers. *Nuc Med Comm.* 2011;32:764–781.)

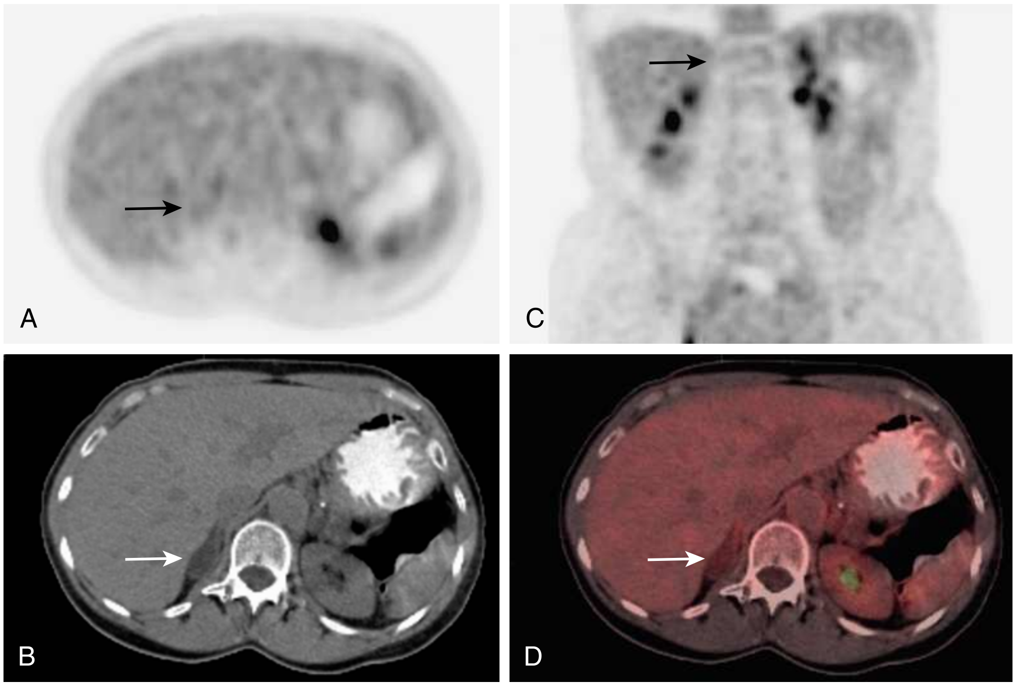


Fig. 92.29 **(^{18}F)-fluorodeoxyglucose positron emission tomography/computed tomography (PET/CT) study in a 74-year-old man with lung cancer.** Restaging PET/CT images show a 3-cm right adrenal lesion *(arrows)* with maximum standard uptake value (SUV) of 4.4 as compared with the normal liver background SUV mean of 2.4 with lesion/liver maximum SUV ratio of 1.5, meeting the metabolic criteria for adrenal nonadenoma. However, this lesion displayed low CT attenuation with <10 Hounsfield units and was stable at 2 years imaging follow-up compatible with an adrenocortical adenoma. There is overlap in SUV uptake of benign versus malignant adrenal lesions. (From Wong KK, Arabi M, Bou-Assaly W, et al. Evaluation of incidentally discovered adrenal masses with PET and PET/CT. *Eur J Radiol.* 2012;81:441–450.)

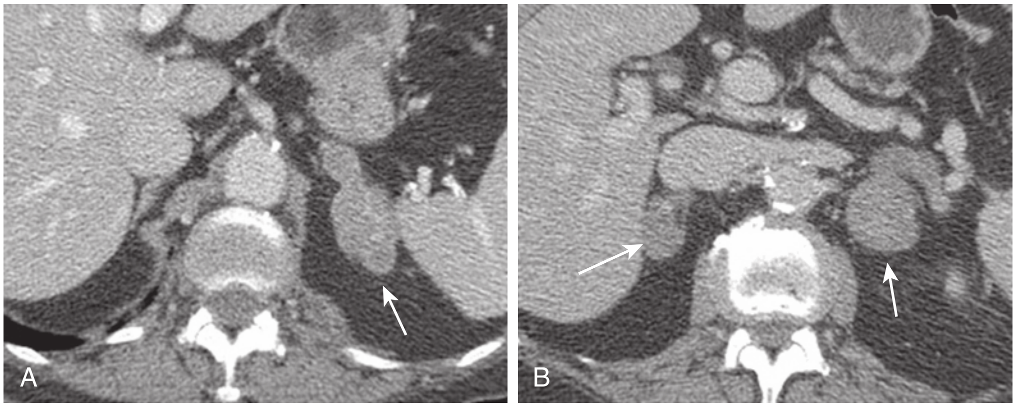


Fig. 92.30 ACTH-independent macronodular adrenal hyperplasia. A and B show large nodules bilaterally superimposed on markedly thickened limbs. *ACTH*, Adrenocorticotropic hormone.

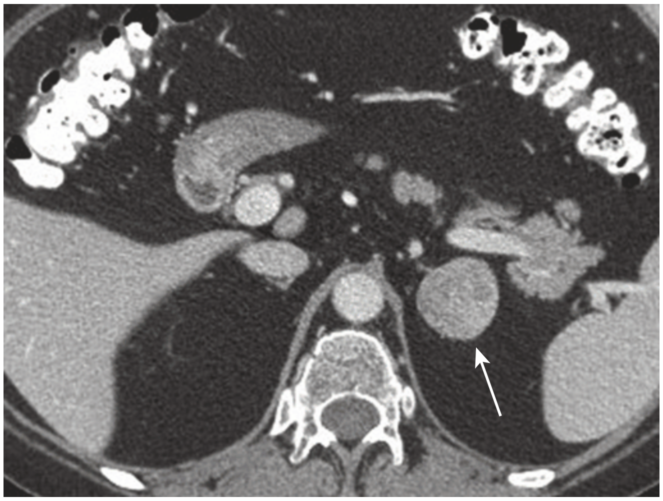


Fig. 92.31 Adrenal adenoma causing Cushing syndrome. Homogeneous 4-cm left adrenal mass *(arrow)* has a nonspecific computed tomography appearance. *L*, Left. (From Korobkin M, Francis I, Kloos R, et al. The incidental adrenal mass. *RCNA*. 1996;34:1037–1054.)

of ACTH) is initiated, followed by poststimulation sampling after 5 to 30 minutes. After sampling, catheters and sheaths are removed, and hemostasis is achieved by manual groin pressure, and patients are discharged to home after a brief recovery period. Potential complications of AVS include venous infarction of the adrenal gland and adrenal hemorrhage. These complications may cause pain, fever, and damage to adrenal gland function. Note that performance of AVS can be quite technically challenging, and success depends upon operator skill and experience.[140,141]

Interpretation of Adrenal Vein Sampling assays in Primary Hyperaldosteronism

Determination of aldosterone and cortisol concentrations in adrenal venous blood can accurately localize the site of aldosterone excess and thereby distinguish unilateral from bilateral adrenal disease. Cortisol assays provide a means of compensating for differences in hormone concentration resulting from catheter placement and dilution of adrenal venous effluent. Calculation of the selectivity index (adrenal vein cortisol/IVC cortisol) is performed for adrenal venous and IVC sampling obtained before and after Cortrosyn stimulation. Adequate sampling is reflected by a selectivity index of greater than 2 before stimulation and greater than 5 after stimulation.

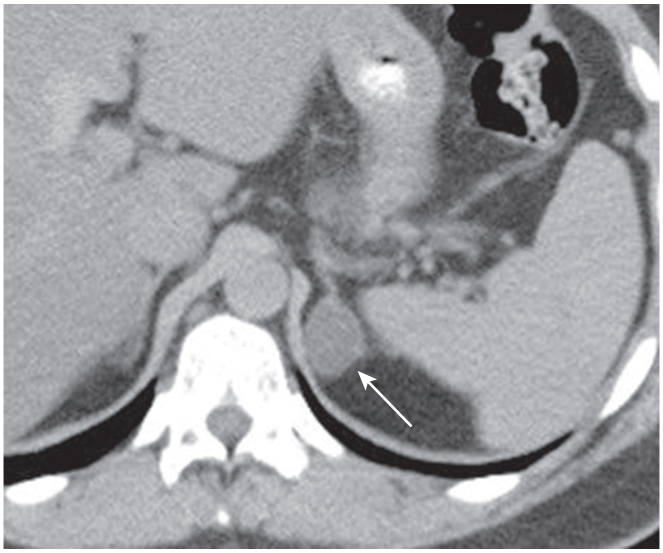


Fig. 92.32 Computed tomography demonstration of a 1-cm aldosterone-producing left adrenal adenoma.

The sensitivity index is calculated using the ratio of aldosterone/cortisol (A/C) concentrations from each aliquot. Right and left adrenal vein and IVC samples before stimulation are compared. The A/C ratio of the side with the higher value is used as a numerator, and that of the lesser value as the denominator and a new ratio is calculated, comparing the two adrenal glands. This value is the sensitivity index. If this value is greater than 3 prestimulation or greater than 5 poststimulation, unilateral hypersecretion is confirmed, provided that the selectivity index previously calculated indicates adequate and valid sampling. Added confidence is present when the A/C ratio of the nondominant side is less than the A/C ratio of the IVC obtained in the concomitant sampling. This indicates "suppression" of aldosterone secretion by the nondominant side.

PA guidelines recommend adrenalectomy that can cure cases of hypertension and hypokalemia when the etiology is secondary to unilateral adrenal secretory masses.[142] Adrenal CT and MRI remain the mainstay with AVS, with care required for the occasional ACC. False positive imaging findings (nonsecretory adenoma) and false negatives (unilateral adrenal hyperplasia) can both occur. Currently, there is a need for functional imaging to guide management of secretory adrenal tumors.[143]

Functional Imaging in Primary Hyperaldosteronism

NP-59 adrenocortical scintigraphy has been well established for localization of functional adrenal masses for both Conn syndrome and PA, although it currently has limited availability in North America. For PA, the addition of dexamethasone suppression of pituitary ACTH secretion leads to decreased accumulation of radiocholesterol by the inner adrenal cortex and allows accurate depiction of aldosterone producing adenoma.[2] NP-59 imaging has been improved with combined SPECT/CT imaging technique for lateralizing the source of hyperaldosteronism and distinguishing adenoma from bilateral adrenal hyperplasia.[144,145] In a single-site retrospective study in 81 patients, NP-59 scintigraphy was reported to have sensitivity of greater than 83% and a positive predictive value of greater than 92%.[146]

An alternative to adrenocortical imaging has been development of a radiotracer, ^{11}C-metomidate (MTO), a ligand of CYP11B1 and CYP11B2, which has a potential role in the characterization of adrenocortical adenomas and for lateralization of functional adenomas.[147] ^{11}C-MTO–PET can aid in differentiating tumors of adrenocortical origin (including nonhypersecreting and hypersecreting cortical adenomas, ACCs, and macronodular hyperplasia) from nonadrenocortical tumors (benign and malignant pheochromocytoma, and metastases to the adrenal),[148,149] with a sensitivity of 89% and a specificity of 96% reported in 212 patients.[150] ^{11}C-MTO–PET has been successfully used for localizing unilateral APAs when AVS is ambiguous or cannot be performed and using dexamethasone suppression can depict PA due to adrenal adenoma.[151] Compared with AVS, ^{11}C-MTO–PET had a reported sensitivity of 76% and specificity of 87% for aldosteronoma.[152] Another small study of 15 patients with PA confirmed ^{11}C-MTO–PET as a useful adjunct to AVS and CT.[153] Radioiodine labeling of metomidate (^{123}I-MTO) allows SPECT/CT imaging of hypersecreting adrenocortical masses with an efficacy similar to PET. A fluorinated version of a CYP11B1/2 radiotracer ligand promises wider availability of PET imaging.[154]

A novel approach to hyperaldosteronism is the use of chemokine receptor 4 (CXCR4) PET imaging, with aldosterone-producing adrenal adenoma demonstrated to have intense uptake on ^{68}Ga-pentixafor–PET imaging.[155] In 36 patients with PA, Heinze et al. reported that ^{68}Ga-pentixafor–PET imaging had 100% sensitivity, 78.6% specificity, and 92.3% accuracy for APA.[156] They found that CXCR4 was expressed in aldosterone-producing tumors but, importantly, not in nonfunctioning adenomas. In nine patients with hyperaldosteronism, ^{68}Ga-pentixafor PET findings were able to lateralize the side of disease, which is important for predicting surgical cure.[157]

KEY POINTS

- Both benign and malignant adrenal lesions can have varied appearances, some with specific diagnostic characteristics depending on the lesion and/or the imaging modality chosen.
- Correlation with patient history can be helpful in narrowing down differential diagnoses.
- Functional imaging can help delineate pathology type and location.
- Functional status (hormone production) is difficult to determine with imaging.
- Although functional status can be difficult to determine in certain pathologies, the presence of hormones on AVS can help with diagnosis and quantifying severity of disease.

SUMMARY

The challenge of adrenal imaging lies in understanding the unique and complementary interplay of the anatomic- and functionally based imaging modalities. The presence of an adrenal mass does not invariably confirm abnormal adrenal function; nor does the presence of normal adrenal anatomy necessarily exclude abnormal adrenal function. Adrenal physiology and pathophysiology must be considered so that appropriate follow-up studies are properly performed and correctly interpreted. Regardless of whether the imaging modality is anatomic- or functionally based, the imaging approach to a given patient must be flexible and at times specifically tailored for each particular disease entity. When conducted in the proper clinical setting and sequence (reasonable suspicion based on clinical features and biochemical studies), CT, MRI, catheter-directed hormone sampling, and radiopharmaceutical imaging using hybrid PET/CT and/or SPECT/CT can be expected to have their greatest clinical utility and diagnostic success.

REFERENCES

1. Vincent JM, Morrison ID, Armstrong P, et al. The size of normal adrenal glands on computed tomography. *Clin Radiol*. 1994;49:453–455.
2. Gross MD, Avram A, Fig LM, et al. PET in the diagnostic evaluation of adrenal tumors. *Q J Nucl Med Mol Imaging*. 2007;51:272–283.
3. Metser U, Miller E, Lerman H, et al. 18F-FDG PET/CT in the evaluation of adrenal masses. *J Nucl Med*. 2006;47:32–37.
4. Avram AM, Fig LM, Gross MD. Adrenal gland scintigraphy. *Sem Nucl Med*. 2006;36:212–227.
5. Eriksson B, Bergstrom M, Sundin A, et al. The role of PET in localization of neuroendocrine and adrenocortical tumors. *Ann N-Y Acad Sci*. 2002;970:159–169.
6. Shapiro B, Copp JE, Sisson JC, et al. Iodine-131 metaiodobenzylguanidine for the locating of suspected pheochromocytoma: experience in 400 cases. *J Nucl Med*. 1985;26:576–585.
7. Hoegerle S, Nitzsche E, Altehoefer C, et al. Pheochromocytomas: detection with 18F DOPA whole body PET--initial results. *Radiology*. 2002;222:507–512.
8. Jager PL, Chirakal R, Marriott CJ, et al. 6-L-18F-fluorodihydroxyphen ylalanine PET in neuroendocrine tumors: basic aspects and emerging clinical applications. *J Nucl Med*. 2008;49:573–586.
9. Pacak K, Eisenhofer G, Carrasquillo JA, et al. 6-[18F]fluorodopamine positron emission tomographic (PET) scanning for diagnostic localization of pheochromocytoma. *Hypertension*. 2001;38:6–8.
10. Win Z, Al-Nahhas A, Towey D, et al. 68Ga-DOTATATE PET in neuroectodermal tumours: first experience. *Nucl Med Comm*. 2007;28:359–363.
11. Kloos RT, Gross MD, Francis IR, et al. Incidentally discovered adrenal masses. *Endocr Rev*. 1995;16:460–484.
12. Korobkin M, Giordano TJ, Brodeur FJ, et al. Adrenal adenomas: relationship between histologic lipid and CT and MR findings. *Radiology*. 1996;200:743–747.
13. Boland GW, Lee MJ, Gazelle GS, et al. Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. *AJR Am J Roentgenol*. 1998;171:201–204.
14. Caoili EM, Korobkin M, Francis IR, et al. Adrenal masses: characterization with combined unenhanced and delayed enhanced CT. *Radiology*. 2002;222:629–633.
15. Blake MA, Kalra MK, Sweeney AT, et al. Distinguishing benign from malignant adrenal masses: multi-detector row CT protocol with 10-minute delay. *Radiology*. 2006;238:578–585.
16. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental indeterminate adrenal mass on CT (> 10 H) in patients without cancer: is further imaging necessary? Follow-up of 321 consecutive indeterminate adrenal masses. *AJR Am J Roentgenol*. 2007;189:1119–1123.
17. Mayo-Smith WW, Lee MJ, McNicholas MM, et al. Characterization of adrenal masses (< 5 cm) by use of chemical shift MR imaging: observer performance versus quantitative measures. *AJR Am J Roentgenol*. 1995;165:91–95.
18. Mitchell DG, Crovello M, Matteucci T, et al. Benign adrenocortical masses: diagnosis with chemical shift MR imaging. *Radiology*. 1992;185:345–351.

19. Fujiyoshi F, Nakajo M, Fukukura Y, et al. Characterization of adrenal tumors by chemical shift fast low-angle shot MR imaging: comparison of four methods of quantitative evaluation. *AJR Am J Roentgenol*. 2003;180:1649–1657.
20. Israel GM, Korobkin M, Wang C, et al. Comparison of unenhanced CT and chemical shift MRI in evaluating lipid-rich adrenal adenomas. *AJR Am J Roentgenol*. 2004;183:215–219.
21. Korobkin M, Lombardi TJ, Aisen AM, et al. Characterization of adrenal masses with chemical shift and gadolinium-enhanced MR imaging. *Radiology*. 1995;197:411–418.
22. Maurea S, Klain M, Mainolfi C, et al. The diagnostic role of radionuclide imaging in evaluation of patients with nonhypersecreting adrenal masses. *J Nucl Med*. 2001;42:884–892.
23. Boland GW, Dwamena BA, Jagtiani Sangwaiya M, et al. Characterization of adrenal masses by using FDG PET: a systematic review and meta-analysis of diagnostic test performance. *Radiology*. 2011;259:117–126.
24. Kim SJ, Lee SW, Pak K, et al. Diagnostic accuracy of (18)F-FDG PET or PET/CT for the characterization of adrenal masses: a systematic review and meta-analysis. *Br J Radiol*. 2018;91:20170520.
25. Guerin C, Pattou F, Brunaud L, et al. Performance of 18F-FDG PET/CT in the characterization of adrenal masses in noncancer patients: a prospective study. *J Clin Endocrinol Metab*. 2017;102:2465–2472.
26. Pavón-Paz I, Rosado-Sierra JA, Balsa-Bretón M, et al. Use of 18F-FDG-positron emission tomography in presurgical evaluation of nonspecific or suspectious adrenal masses in non-oncologic patients. *Med Clin*. 2016;146:104–107.
27. Altinmakas E, Hobbs BP, Ye H, et al. Diagnostic performance of (18-)F-FDG-PET-CT in adrenal lesions using histopathology as reference standard. *Abdom Radiol*. 2017;42:577–584.
28. Ozcan Kara P, Kara T, Kara Gedik G, et al. The role of fluorodeoxyglucose-positron emission tomography/computed tomography in differentiating between benign and malignant adrenal lesions. *Nucl Med Comm*. 2011;32:106–112.
29. Perri M, Erba P, Volterrani D, et al. Adrenal masses in patients with cancer: PET/CT characterization with combined CT histogram and standardized uptake value PET analysis. *AJR Am J Roentgenol*. 2011;197:209–216.
30. Delivanis DA, Bancos I, Atwell TD, et al. Diagnostic performance of unenhanced computed tomography and (18) F-fluorodeoxyglucose positron emission tomography in indeterminate adrenal tumours. *Clin Endocrinol*. 2018;88:30–36.
31. Chen CC, Carrasquillo JA. Molecular imaging of adrenal neoplasms. *J Surg Oncol*. 2012;106:532–542.
32. Leboulleux S, Deandreis D, Escourrou C, et al. Fluorodesoxyglucose uptake in the remaining adrenal glands during the follow-up of patients with adrenocortical carcinoma: do not consider it as malignancy. *Eur J Endocrinol*. 2011;164:89–94.
33. Li W, Pang H, Cao Y, et al. High 18F-fluorodeoxyglucose uptake in adrenal angiomyolipoma: case report and review of literature. *Medicine*. 2015;94:e900.
34. Anwar H, Sachpekidis C, Schwarzbach M, et al. Fluorine-18-FDG PET/CT in a patient with angiomyolipoma: response to mammalian target of rapamycin inhibitor therapy. *Hellenic J Nucl Med*. 2017;20:169–171.
35. Bacanovic S, Burger IA, Stolzmann P, et al. Ipilimumab-induced adrenalitis: a possible pitfall in 18F-FDG-PET/CT. *Clin Nucl Med*. 2015;40:e518–e519.
36. Aktas GE, Soyluoglu Demir S, Sarikaya A. Bilateral symmetrical adrenal hypermetabolism on FDG PET/CT due to Cushing syndrome in well differentiated neuroendocrine carcinoma. *Revista Espanola de Med Nucl e Imagen Mol*. 2016;35:118–120.
37. Matsusaka Y, Nakahara T, Kurihara I, et al. Bilateral adrenal hyperplasia due to adrenocorticotropic hormone-secreting esthesioneuroblastoma on FDG PET and 111In-pentetreotide scintigraphy. *Clin Nucl Med*. 2018;43:452–453.
38. Jinguji M, Nakajo M, Nakajo M, et al. Vasovagal-related stress immediately before FDG injection may increase bilateral adrenal FDG uptake. *Br J Radiol*. 2016;89:20150950.
39. Otomi Y, Shinya T, Otsuka H, et al. Increased (18)F-fluorodeoxyglucose accumulation in bilateral adrenal glands of the patients suffering from vasovagal reaction due to blood vessel puncture. *Ann Nucl Med*. 2016;30:501–505.
40. De Leo A, Mosconi C, Zavatta G, et al. Radiologically defined lipid-poor adrenal adenomas: histopathological characteristics. *J Endocrinol Invest*. 2020;43(9):1197-1204.
41. Patel D, Gara SK, Ellis RJ, et al. FDG PET/CT scan and functional adrenal tumors: a pilot study for lateralization. *World J Surg*. 2016;40:683–689.
42. Murayama R, Nishie A, Hida T, et al. Uptake of 18F-FDG in adrenal adenomas is associated with unenhanced CT value and constituent cells. *Clin Nucl Med*. 2019;44:943–948.
43. Goldman HB, Howard RC, Patterson AL. Spontaneous retroperitoneal hemorrhage from a giant adrenal myelolipoma. *J Urol*. 1996;155:639.
44. Russell C, Goodacre BW, vanSonnenberg E, et al. Spontaneous rupture of adrenal myelolipoma: spiral CT appearance. *Abdom Imaging*. 2000;25:431–434.
45. Rao P, Kenney PJ, Wagner BJ, et al. Imaging and pathologic features of myelolipoma. *Radiographics*. 1997;17:1373–1385.
46. Sneiders A, Zhang G, Gordon BE. Extra-adrenal perirenal myelolipoma. *J Urol*. 1993;150:1496–1497.
47. Rowe SP, Javadi MS, Solnes LB, et al. Appearance of adrenal myelolipomas on 2-deoxy-2-((18)F) fluoro-D-glucose positron emission tomography-computed tomography. *World J Nucl Med*. 2017;16:271–274.
48. Burks DW, Mirvis SE, Shanmuganathan K. Acute adrenal injury after blunt abdominal trauma: CT findings. *AJR Am J Roentgenol*. 1992;158:503–507.
49. Bookstein JJ, Conn J, Reuter SR. Intra-adrenal hemorrhage as a complication of adrenal venography in primary aldosteronism. *Radiology*. 1968;90:778–779.
50. Ling D, Korobkin M, Silverman PM, et al. CT demonstration of bilateral adrenal hemorrhage. *AJR Am J Roentgenol*. 1983;141:307–308.
51. Kawashima A, Sandler CM, Ernst RD, et al. Imaging of nontraumatic hemorrhage of the adrenal gland. *Radiographics*. 1999;19:949–963.
52. Cheema P, Cartagena R, Staubitz W. Adrenal cysts: diagnosis and treatment. *J Urol*. 1981;126:396–399.
53. Rozenblit A, Morehouse HT, Amis Jr ES. Cystic adrenal lesions: CT features. *Radiology*. 1996;201:541–548.
54. Mendiratta-Lala M, Avram A, Turcu AF, et al. Adrenal imaging. *Endocrinol Metab Clin North Am*. 2017;46:741–759.
55. Wilson DA, Muchmore HG, Tisdal RG, et al. Histoplasmosis of the adrenal glands studied by CT. *Radiology*. 1984;150:779–783.
56. Zhou J, Lv J, Pan Y, et al. Unilateral adrenal cryptococcosis on FDG PET/CT. *Clin Nucl Med*. 2017;42:565–566.
57. Papadakis GZ, Holland SM, Quezado M, et al. Adrenal cryptococcosis in an immunosuppressed patient showing intensely increased metabolic activity on (18)F-FDG PET/CT. *Endocrine*. 2016;54:834–836.
58. Gajendra S, Sharma R, Goel S, et al. Adrenal histoplasmosis in immunocompetent patients presenting as adrenal insufficiency. *Turk patoloji dergisi*. 2016;32:105–111.
59. Kalathoorakath RR, Sharma A, Sood A, et al. 18)F-FDG PET/CT imaging and PET-guided biopsy in evaluation and treatment decision in adrenal histoplasmosis. *BJR Case Rep*. 2016;2:20150451.
60. Arambewela M, Ross R, Pirzada O, et al. Tuberculosis as a differential for bilateral adrenal masses in the UK. *BMJ Case Rep*. 2019;12(5).
61. Gorla AK, Gupta K, Sood A, et al. Adrenal tuberculosis masquerading as disseminated malignancy: a pitfall of (18)F-FDG PET/CT Imaging. *Rev Esp Med Nucl Imagen Mol*. 2016;35:257–259.
62. Koh SA. Addison's disease due to bilateral adrenal tuberculosis on 18F-fluorodeoxyglucose positron emission tomography computed tomography. *Infect Dis Rep*. 2018;10:7773.
63. Altinmakas E, Guo M, Kundu UR, et al. Computed tomography and (18) F-fluorodeoxyglucose positron emission tomography/computed tomography findings in adrenal candidiasis and histoplasmosis: two cases. *Clin Imaging*. 2015;39:1115–1118.

64. Honig SC, Klavans MS, Hyde C, et al. Adrenal hemangioma: an unusual adrenal mass delineated with magnetic resonance imaging. *J Urol*. 1991;146:400–402.
65. Krebs TL, Wagner BJ. MR imaging of the adrenal gland: radiologic-pathologic correlation. *Radiographics*. 1998;18:1425–1440.
66. Wilson B, Becker A, Estes T, et al. Adrenal hemangioma definite diagnosis on CT, MRI, and FDG PET in a patient with primary lung cancer. *Clin Nucl Med*. 2018;43:e192–e194.
67. Radin R, David CL, Goldfarb H, et al. Adrenal and extra-adrenal retroperitoneal ganglioneuroma: imaging findings in 13 adults. *Radiology*. 1997;202:703–707.
68. Mylonas KS, Schizas D, Economopoulos KP. Adrenal ganglioneuroma: What you need to know. *World J Clin Cases*. 2017;5(10):373–377.
69. McMurry Jr JF, Long D, McClure R, et al. Addison's disease with adrenal enlargement on computed tomographic scanning. Report on two cases of tuberculosis and review of the literature. *Am J Med*. 1984;77:365–368.
70. Caron P, Chabannier MH, Cambus JP, et al. Definitive adrenal insufficiency due to bilateral adrenal hemorrhage and primary antiphospholipid syndrome. *J Clin Endocrinol Metab*. 1998;83:1437–1439.
71. Piédrola G, Casado JL, López E, et al. Clinical features of adrenal insufficiency in patients with acquired immunodeficiency syndrome. *Clin Endocrinol*. 1996;45:97–101.
72. Dunnick NR, Heaston D, Halvorsen R, et al. CT appearance of adrenal cortical carcinoma. *J Comp Assist Tomogr*. 1982;6:978–982.
73. Hedican SP, Marshall FF. Adrenocortical carcinoma with intracaval extension. *J Urol*. 1997;158:2056–2061.
74. Schlund JF, Kenney PJ, Brown ED, et al. Adrenocortical carcinoma: MR imaging appearance with current techniques. *JMRI*. 1995;5:171–174.
75. Szolar DH, Korobkin M, Reittner P, et al. Adrenocortical carcinomas and adrenal pheochromocytomas: mass and enhancement loss evaluation at delayed contrast-enhanced CT. *Radiology*. 2005;234:479–485.
76. Dunnick NR, Doppman JL, Geelhoed GW. Intravenous extension of endocrine tumors. *AJR Am J Roentgenol*. 1980;135:471–476.
77. Cistaro A, Niccoli Asabella A, Coppolino P, et al. Diagnostic and prognostic value of 18F-FDG PET/CT in comparison with morphological imaging in primary adrenal gland malignancies - a multicenter experience. *Hellenic J Nucl Med*. 2015;18:97–102.
78. Ardito A, Massaglia C, Pelosi E, et al. 18F-FDG PET/CT in the postoperative monitoring of patients with adrenocortical carcinoma. *Eur J Endocrinol*. 2015;173:749–756.
79. Shekhar S, Gubbi S, Papadakis GZ, et al. Adrenocortical carcinoma and pulmonary embolism from tumoral extension. *Endocrinol Diabet Metab Case Rep*. 2019;19(95):1-4.
80. Satoh K, Patel D, Dieckmann W, et al. Whole body metabolic tumor volume and total lesion glycolysis predict survival in patients with adrenocortical carcinoma. *Ann Surg Oncol*. 2015;22:S714–S720.
81. Bluemel C, Hahner S, Heinze B, et al. Investigating the chemokine receptor 4 as potential theranostic target in adrenocortical cancer patients. *Clin Nucl Med*. 2017;42:e29–e34.
82. Weiss ID, Huff LM, Evbuomwan MO, et al. Screening of cancer tissue arrays identifies CXCR4 on adrenocortical carcinoma: correlates with expression and quantification on metastases using (64)Cu-plerixafor PET. *Oncotarget*. 2017;8:73387–73406.
83. Crona J, Taïeb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification. *Endocr Rev*. 2017;38(6):489–515.
84. Lucon AM, Pereira MA, Mendonça BB, et al. Pheochromocytoma: study of 50 cases. *J Urol*. 1997;157:1208–1212.
85. Jacobson AF, Deng H, Lombard J, et al. 123I-meta-iodobenzylguanidine scintigraphy for the detection of neuroblastoma and pheochromocytoma: results of a meta-analysis. *J Clin Endocrinol Metab*. 2010;95:2596–2606.
86. Gabriel S, Blanchet EM, Sebag F, et al. Functional characterization of nonmetastatic paraganglioma and pheochromocytoma by (18) F-FDOPA PET: focus on missed lesions. *Clin Endocrinol*. 2013;79:170–177.
87. Ilias I, Chen CC, Carrasquillo JA, et al. Comparison of 6-18F-fluorodopamine PET with 123I-metaiodobenzylguanidine and 111in-pentetreotide scintigraphy in localization of nonmetastatic and metastatic pheochromocytoma. *J Nucl Med*. 2008;49:1613–1619.
88. Shulkin BL, Koeppe RA, Francis IR, et al. Pheochromocytomas that do not accumulate metaiodobenzylguanidine: localization with PET and administration of FDG. *Radiology*. 1993;186:711–715.
89. Timmers HJ, Kozupa A, Chen CC, et al. Superiority of fluorodeoxyglucose positron emission tomography to other functional imaging techniques in the evaluation of metastatic SDHB-associated pheochromocytoma and paraganglioma. *J Clin Oncol*. 2007;25:2262–2269.
90. Nakajo M, Nakajo M, Fukukura Y, et al. Diagnostic performances of FDG-PET/CT and diffusion-weighted imaging indices for differentiating benign pheochromocytoma from other benign adrenal tumors. *Abdom Imaging*. 2015;40:1655–1665.
91. Nockel P, El Lakis M, Gaitanidis A, et al. Preoperative 18F-FDG PET/CT in pheochromocytomas and paragangliomas allows for precision surgery. *Ann Surg*. 2019;269:741–747.
92. van Berkel A, Vriens D, Visser EP, et al. Metabolic subtyping of pheochromocytoma and paraganglioma by (18)F-FDG pharmacokinetics using dynamic PET/CT scanning. *J Nucl Med*. 2019;60:745–751.
93. Tiwari A, Shah N, Sarathi V, et al. Genetic status determines (18) F-FDG uptake in pheochromocytoma/paraganglioma. *J Med Imaging Radiat Oncol*. 2017;61:745–752.
94. Lepoutre-Lussey C, Caramella C, Bidault F, et al. Screening in asymptomatic SDHx mutation carriers: added value of ^{18}F-FDG PET/CT at initial diagnosis and 1-year follow-up. *Eur J Nucl Med Mol Imaging*. 2015;42:868–876.
95. Képénékian L, Mognetti T, Lifante JC, et al. Interest of systematic screening of pheochromocytoma in patients with neurofibromatosis type 1. *Eur J Endocrinol*. 2016;175:335–344.
96. Fottner C, Helisch A, Anlauf M, et al. 6-18F-fluoro-L-dihydroxy phenylalanine positron emission tomography is superior to 123I-metaiodobenzyl-guanidine scintigraphy in the detection of extraadrenal and hereditary pheochromocytomas and paragangliomas: correlation with vesicular monoamine transporter expression. *J Clin Endocrinol Metab*. 2010;95:2800–2810.
97. Taïeb D, Jha A, Guerin C, et al. 18F-FDOPA PET/CT imaging of MAX-related pheochromocytoma. *J Clin Endocrinol Metab*. 2018;103:1574–1582.
98. Timmers HJ, Chen CC, Carrasquillo JA, et al. Comparison of 18F-fluoro-L-DOPA, 18F-fluoro-deoxyglucose, and 18F-fluorodopamine PET and 123I-MIBG scintigraphy in the localization of pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2009;94:4757–4767.
99. Kroiss A, Putzer D, Uprimny C, et al. Functional imaging in phaeochromocytoma and neuroblastoma with 68Ga-DOTA-Tyr 3-octreotide positron emission tomography and 123I-metaiodobenzylguanidine. *Eur J Nucl Med Mol Imaging*. 2011;38:865–873.
100. Shell J, Tirosh A, Millo C, et al. The utility of (68)Gallium-DOTATATE PET/CT in the detection of von Hippel-Lindau disease associated tumors. *Eur J Radiol*. 2019;112:130–135.
101. Tuzcu SA, Pekkolay Z. Multiple endocrine neoplasia type 2A syndrome (MEN2A) and usefulness of 68Ga-DOTATATE PET/CT in this syndrome. *Ann Ital Chir*. 2019;90:497–503.
102. Paling MR, Williamson BR. Adrenal involvement in non-Hodgkin lymphoma. *AJR Am J Roentgenol*. 1983;141:303–305.
103. Falchook FS, Allard JC. CT of primary adrenal lymphoma. *J Comput Ass Tomogr*. 1991;15:1048–1050.
104. An P, Chen K, Yang GQ, et al. Diffuse large B cell lymphoma with bilateral adrenal and hypothalamic involvement: a case report and literature review. *World J Clin Cases*. 2019;7:4075–4083.
105. Dong P, Wang L, Shen G, et al. Primary adrenal extranasal NK/T cell lymphoma with subcutaneous involvement demonstrated on FDG PET/CT: a clinical case report. *Medicine*. 2019;98:e14818.
106. Doroudinia A, Bakhshayesh Karam M, Ranjbar M, et al. Mantle cell lymphoma presenting as bilateral adrenal huge masses. *BMJ Case Rep*. 2018;bcr2017223247.
107. Li W, Lin W, Ma C, et al. A case of intravascular large B-cell lymphoma in the left adrenal and another tumor in the right adrenal detected by (18)F-FDG PET/CT. *Hellenic J Nucl Med*. 2016;19:57–59.

108. Zhou J, Zhao Y, Gou Z. High 18F-fluorodeoxyglucose uptake in primary bilateral adrenal diffuse large B-cell lymphomas with nongerminal center B-cell phenotype: a case report. *Medicine*. 2018;97:e0480.
109. Altinmakas E, Üçışık-Keser FE, Medeiros LJ, et al. CT and (18)F- FDG-PET-CT findings in secondary adrenal lymphoma with pathologic correlation. *Acad Radiol*. 2019;26:e108–e114.
110. Wang M, Xu H, Xiao L, et al. Prognostic value of functional parameters of (18)F-FDG-PET images in patients with primary renal/adrenal lymphoma. *Contrast Media Mol Imaging*. 2019;2019:2641627.
111. Laurent C, Casasnovas O, Martin L, et al. Adrenal lymphoma: presentation, management and prognosis. *QJM*. 2017;110:103–109.
112. Abrams HL, Spiro R, Goldstein N. Metastases in carcinoma; analysis of 1000 autopsied cases. *Cancer*. 1950;3:74–85.
113. Caoili EM, Korobkin M, Francis IR, et al. Delayed enhanced CT of lipid-poor adrenal adenomas. *AJR Am J Roentgenol*. 2000;175:1411–1415.
114. Wu Q, Luo W, Zhao Y, et al. The utility of 18F-FDG PET/CT for the diagnosis of adrenal metastasis in lung cancer: a PRISMA-compliant meta-analysis. *Nucl Med Comm*. 2017;38:1117–1124.
115. Kanaya N, Noma K, Okada T, et al. A case of long-term survival after surgical resection for solitary adrenal recurrence of esophageal squamous carcinoma. *Surg Case Rep*. 2017;3:61.
116. Tsuchida K, Watanabe H, Kameda Y, et al. [A case of solitary adrenal metastasis from Rectal cancer treated by adrenalectomy after pre-operative chemotherapy]. *Gan To Kagaku Ryoho Cancer Chemother*. 2016;43:1751–1753.
117. Lang BH, Cowling BJ, Li JY, et al. High false positivity in positron emission tomography is a potential diagnostic pitfall in patients with suspected adrenal metastasis. *World J Surg*. 2015;39:1902–1908.
118. Carrasquillo JA, Pandit-Taskar N, Chen CC. Radionuclide therapy of adrenal tumors. *J Surg Oncol*. 2012;106:632–642.
119. Papaioannou G, McHugh K. Neuroblastoma in childhood: review and radiological findings. *Cancer Imag*. 2005;5:116–127.
120. Abramson SJ. Adrenal neoplasms in children. *Radiologic Clin North Am*. 1997;35:1415–1453.
121. Lonergan GJ, Schwab CM, Suarez ES, et al. Neuroblastoma, ganglioneuroblastoma, and ganglioneuroma: radiologic-pathologic correlation. *Radiographics*. 2002;22:911–934.
122. Shulkin BL, Shapiro B, Hutchinson RJ. Iodine-131-metaiodobenzylguanidine and bone scintigraphy for the detection of neuroblastoma. *J Nucl Med*. 1992;33:1735–1740.
123. Sharp SE, Trout AT, Weiss BD, et al. MIBG in Neuroblastoma diagnostic imaging and therapy. *Radiographics*. 2016;36:258–278.
124. Gains JE, Bomanji JB, Fersht NL, et al. 177Lu-DOTATATE molecular radiotherapy for childhood neuroblastoma. *J Nucl Med*. 2011;52:1041–1047.
125. Kong G, Hofman MS, Murray WK, et al. Initial experience with gallium-68 DOTA-octreotate PET/CT and peptide receptor radionuclide therapy for pediatric patients with refractory metastatic neuroblastoma. *J Pediatr Hematology/Oncol*. 2016;38:87–96.
126. Doppman JL, Miller DL, Dwyer AJ, et al. Macronodular adrenal hyperplasia in Cushing disease. *Radiology*. 1988;166:347–352.
127. Doppman JL, Chrousos GP, Papanicolaou DA, et al. Adrenocorticotropin-independent macronodular adrenal hyperplasia: an uncommon cause of primary adrenal hypercortisolism. *Radiology*. 2000;216:797–802.
128. Isidori AM, Sbardella E, Zatelli MC, et al. Conventional and nuclear medicine imaging in ectopic Cushing's syndrome: a systematic review. *J Clin Endocrinol Metab*. 2015;100:3231–3244.
129. Santhanam P, Taieb D, Giovanella L, et al. PET imaging in ectopic Cushing syndrome: a systematic review. *Endocrine*. 2015;50:297–305.
130. Parihar AS, Mittal BR, Vadi SK, et al. Ectopic Cushing Syndrome (ECS): 68Ga-DOTANOC PET/CT localizes the site of ectopic adrenocorticotropic hormone production. *Clin Nucl Med*. 2018;43:769–770.
131. Korobkin M, Francis IR. Adrenal imaging. *Semin Ultrasound CT MR*. 1995;16:317–330.
132. Lingam RK, Sohaib SA, Vlahos I, et al. CT of primary hyperaldosteronism (Conn's syndrome): the value of measuring the adrenal gland. *AJR Am J Roentgenol*. 2003;181:843–849.
133. Hollak CE, Prummel MF, Tiel-van Buul MM. Bilateral adrenal tumours in primary aldosteronism: localization of a unilateral aldosteronoma by dexamethasone suppression scan. *J Int Med*. 1991;229:545–548.
134. Doppman JL, Gill Jr JR, Miller DL, et al. Distinction between hyperaldosteronism due to bilateral hyperplasia and unilateral aldosteronoma: reliability of CT. *Radiology*. 1992;184:677–682.
135. Sam D, Kline GA, So B, et al. Discordance between imaging and adrenal vein sampling in primary aldosteronism irrespective of interpretation criteria. *J Clin Endocrinol Metab*. 2019;104:1900–1906.
136. Blevins Jr LS, Wand GS. Primary aldosteronism: an endocrine perspective. *Radiology*. 1992;184:599–600.
137. Sohaib SA, Peppercorn PD, Allan C, et al. Primary hyperaldosteronism (Conn syndrome): MR imaging findings. *Radiology*. 2000;214:527–531.
138. Dunnick NR, Leight Jr GS, Roubidoux MA, et al. CT in the diagnosis of primary aldosteronism: sensitivity in 29 patients. *AJR Am J Roentgenol*. 1993;160:321–324.
139. Sheaves R, Goldin J, Reznek RH, et al. Relative value of computed tomography scanning and venous sampling in establishing the cause of primary hyperaldosteronism. *Eur J Endocrinol*. 1996;134:308–313.
140. Betz MJ, Degenhart C, Fischer E, et al. Adrenal vein sampling using rapid cortisol assays in primary aldosteronism is useful in centers with low success rates. *Eur J Endocrinol*. 2011;165:301–306.
141. Daunt N. Adrenal vein sampling: how to make it quick, easy, and successful. *Radiographics*. 2005;25:S143–S158.
142. Bardet S, Chamontin B, Douillard C, et al. SFE/SFHTA/AFCE consensus on primary aldosteronism, part 4: Subtype diagnosis. *Ann Endocrinol*. 2016;77:208–213.
143. Naruse M, Umakoshi H, Tsuiki M, et al. The latest developments of functional molecular imaging in the diagnosis of primary aldosteronism. *Horm Metab Res*. 2017;49(12):929–935. doi: 10.1055/s-0043-.
144. Wong KK, Komissarova M, Avram AM, et al. Adrenal cortical imaging with I-131 NP-59 SPECT-CT. *Clin Nucl Med*. 2010;35:865–869.
145. Yen RF, Wu VC, Liu KL, et al. 131I-6beta-iodomethyl-19-norcholesterol SPECT/CT for primary aldosteronism patients with inconclusive adrenal venous sampling and CT results. *J Nucl Med*. 2009;50:1631–1637.
146. Wu MH, Liu FH, Lin KJ, et al. Diagnostic value of adrenal iodine-131 6-beta-iodomethyl-19-norcholesterol scintigraphy for primary aldosteronism: a retrospective study at a medical center in North Taiwan. *Nucl Med Comm*. 2019;40:568–575.
147. Mendichovszky IA, Powlson AS, Manavaki R, et al. Targeted molecular imaging in adrenal disease-an emerging role for metomidate PET-CT. *Diagnostics*. 2016;6(4):42.
148. Minn H, Salonen A, Friberg J, et al. Imaging of adrenal incidentalomas with PET using (11)C-metomidate and (18)F-FDG. *J Nucl Med*. 2004;45:972–979.
149. Zettinig G, Mitterhauser M, Wadsak W, et al. Positron emission tomography imaging of adrenal masses: (18)F-fluorodeoxyglucose and the 11beta-hydroxylase tracer (11)C-metomidate. *Eur J Nucl Med Mol Imaging*. 2004;31:1224–1230.
150. Hennings J, Lindhe O, Bergström M, et al. [11C]metomidate positron emission tomography of adrenocortical tumors in correlation with histopathological findings. *J Clin Endocrinol Metab*. 2006;91:1410–1414.
151. Powlson AS, Gurnell M, Brown MJ. Nuclear imaging in the diagnosis of primary aldosteronism. *Curr Opin Endocrinol Diabet Obesity*. 2015;22:150–156.
152. Burton TJ, Mackenzie IS, Balan K, et al. Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *J Clin Endocrinol Metab*. 2012;97:100–109.
153. O'Shea PM, O'Donoghue D, Bashari W, et al. 11) C-Metomidate PET/CT is a useful adjunct for lateralization of primary aldosteronism in routine clinical practice. *Clin Endocrinol*. 2019;90:670–679.
154. Bongarzone S, Basagni F, Sementa T, et al. Development of [(18)F]FAMTO: a novel fluorine-18 labelled positron emission tomography (PET) radiotracer for imaging CYP11B1 and CYP11B2 enzymes in adrenal glands. *Nucl Med Bio*. 2019;68–69:14–21.

155. Ding J, Tong A, Zhang Y, et al. Intense 68Ga-pentixafor activity in aldosterone-producing adrenal adenomas. *Clin Nucl Med*. 2020;45:336–339.

156. Heinze B, Fuss CT, Mulatero P, et al. Targeting CXCR4 (CXC chemokine receptor type 4) for molecular imaging of aldosterone-producing adenoma. *Hypertension*. 2018;71:317–325.

157. Ding J, Zhang Y, Wen J, et al. Imaging CXCR4 expression in patients with suspected primary hyperaldosteronism. *Eur J Nucl Med Mol Imaging*. 2020;47(11):2656-2665.

93

Adrenal Insufficiency

David J. Torpy, Irina Bancos, and Eystein Husebye

OUTLINE

Brief Historical Perspective: the Changing Spectrum of Adrenal Insufficiency Etiology, 1553
Epidemiology of Adrenal Insufficiency, 1554
Pathogenesis of Adrenal Insufficiency, 1554
Clinical Features of Adrenal Insufficiency, 1554
Etiology, 1555
  Primary Adrenal Insufficiency, 1555
Critical Illness-Related Corticosteroid Insufficiency, 1560
Diagnosis of Adrenal Insufficiency, 1560
  Adrenocorticotropic Hormone Stimulation Testing, 1560
  Overnight Metyrapone Test, 1561
  Insulin Tolerance Test, 1561
  Adrenal Autoantibody Tests, 1561
  Serum 17-Hydroxyprogesterone, 1561
  Very Long Chain Fatty Acids, 1561
  Adrenal Imaging, 1561
Treatment of Adrenal Insufficiency, 1561
  Replacement of Dehydroepiandrosterone, 1562
  Quality of Life in Adrenal Insufficiency, 1563
  Adrenal Suppression, 1563
Adrenal Crisis, 1563
  Treatment of Adrenal Crisis, 1563
  Prevention of Adrenal Crises, 1564
  The Need for Novel Treatments for Adrenal Insufficiency, 1566

BRIEF HISTORICAL PERSPECTIVE: THE CHANGING SPECTRUM OF ADRENAL INSUFFICIENCY ETIOLOGY

Adrenal insufficiency (AI) is defined by inadequate secretion of cortisol and/or aldosterone by the adrenal cortex to maintain homeostasis. Bartolomeo Eustachi, in 1563, first recognized the adrenal glands as organs distinct from the kidneys.[1] Addison disease, eponymously named after Thomas Addison, refers to primary AI (PAI), that is, resulting from direct adrenal pathology. Thomas Addison described many diseases through correlation of clinical presentation with autopsy examination. Addison's description of AI remains as accurate today as it was in 1855: "anaemia, general languor and debility, remarkable weakness of the heart's action, irritability of the stomach, and a peculiar change of colour in the skin." Addison was studying pernicious anemia when he discovered adrenal pathology: tuberculous adrenal lesions in some patients with tuberculosis and adrenal atrophy in others, some of which had vitiligo and pernicious anemia. Of 11 patients described, five had bilateral adrenal tuberculosis, one had unilateral adrenal tuberculosis, three had carcinomatous adrenal involvement, one had adrenal hemorrhage, and one showed atrophy and fibrosis. Addison's findings were quickly confirmed by Brown-Sequard,[2] who verified Addison's hypothesis by showing that, in several laboratory animals, bilateral adrenalectomy was uniformly fatal. The clinical syndrome was named after Addison by Trousseau.[3] Osler attempted unsuccessfully to treat a young patient with Addison disease, employing a glycerine extract of fresh porcine adrenals given orally,[4-6] and isolated and characterized cortisone and cortisol in the 1930s, and Sarett devised a partial synthesis for cortisone from deoxycholic acid in 1945. The clinical effects of cortisone soon were made apparent by the work of Hench and coworkers[7] in the treatment of rheumatoid arthritis and by Thorn and Forsham[8] in the treatment of AI. Cushing largely clarified the role of the pituitary gland in regulating adrenal function, and Harris[9] in the 1940s clarified the role of the hypothalamus in regulating pituitary function via the hypothalamo-hypophyseal portal circulation. Adrenocorticotropic hormone (ACTH) was isolated and characterized by Li,[10] and corticotrophin-releasing hormone (CRH), in turn, was characterized by Vale[11]. Sampson, in 1961, first recognized the syndrome of acute AI in a surgical patient who had atrophic adrenal glands secondary to longstanding glucocorticoid treatment.

The etiological spectrum of AI has evolved considerably. The majority of cases now result from the use of supraphysiologic, anti-inflammatory doses of glucocorticoids in the many diseases that respond well. Among children, genetic disorders (e.g., developmental pituitary defects or congenital adrenal hyperplasia [CAH]) are frequent causes of AI, as well as childhood cerebral tumors or their treatment, now often successful. Autoimmunity is a common cause of PAI, but newer drugs, particularly those used in oncology to increase immune activity, the CTLA-4, PD-1, and PD-L1 antibodies, may produce hypophysitis/secondary AI (SAI).[12] In addition, opioids cause SAI, especially in high cumulative doses.[13] Human immunodeficiency virus infection of the adrenals or attendant adrenal mycotic infections and/or antifungal inhibition of adrenal steroidogenesis may lead to AI, an outcome less frequent with antiretroviral therapies.[14] Observations of salutory pressor effects of hydrocortisone in septic shock and their possible relation to impaired responsiveness of cortisol to ACTH stimulation have led to the notion of relative AI, now referred to as critical illness-related corticosteroid insufficiency (CIRCI), an area that requires better endocrine definition.[15] These observations highlight the importance of considering clinical context in AI differential diagnosis and considering newer causes as they appear in the community.

EPIDEMIOLOGY OF ADRENAL INSUFFICIENCY

Primary AI has an annual incidence of 4 to 6 cases per million and a prevalence of 100 to 220 cases per million, which may be increasing SAI is considered to be twice as common as PAI. The exact prevalence of exogenous adrenal suppression is not known. Adrenal crises are more frequent in PAI than in SAI (estimated in one study to be 5.2 and 3.6 adrenal crisis (AC) per 100 person-years (PY), respectively).[16]

PATHOGENESIS OF ADRENAL INSUFFICIENCY

The hypothalamic-pituitary-adrenal (HPA) axis comprises a classic negative feedback loop (see Chapter 132).

Many pathological processes can lead to loss of adrenal or hypothalamic-pituitary function, and each of these have intrinsic implications.[7] Endocrine consequences arise from loss of adrenal hormones, most importantly cortisol and aldosterone. Although not routinely replaced, there is a loss of production of adrenal androgens, DHEA(-S) and androstenedione, which may result in reduced androgen-dependent hair, especially in females, where adrenal androgens represent approximately 50% of circulating androgens. A lack of adrenal cortisol secretion markedly reduces adrenomedullary epinephrine secretion through depriving the phenylethanolamine-N-methyltransferase enzyme of its stimulatory drive; the extent of practical pathological implications is unclear, but may theoretically include a blunted response to hypoglycemia and an altered sympathoadrenal (epinephrine) response to stress. Low plasma epinephrine and metanephrine levels and low urinary epinephrine excretion can be observed CAH patients with 21-hydroxylase deficiency, which, when severe, can result in frequent hospitalizations for hypoglycemia and hypotension.[18]

Primary AI, most often autoimmune where tuberculosis is not endemic, results in loss of glomerulosa cells and fasciculata cells, leading to lack of aldosterone and cortisol. Typical sequential hormone deficits include increased serum renin, reduced serum aldosterone, and increased serum ACTH, followed by reduced serum cortisol.[19] Usually, AI patients are diagnosed in the most symptomatic low-cortisol state. Isolated loss of cortisol secretion may be seen in autoimmune disease but more commonly follows other insults such as adrenal hemorrhage.

In general, PAI involves deficiency of aldosterone and cortisol, as well as ACTH hypersecretion. The loss of aldosterone secretion and ACTH hypersecretion are not seen in SAI; hence, only PAI patients typically have hyperkalemia from aldosterone deficiency and hyperpigmentation from ACTH and other proopiomelanocortin (POMC) peptides such as α-MSH that have melanocyte-stimulating properties via melanocortin-1 receptors.

SAI is caused by deficits in ACTH secretion, because of loss of hypothalamic CRH production or direct pituitary pathology. The result is low cortisol secretion and normal or low serum ACTH concentration. Common pituitary/hypothalamic pathologies are pituitary adenomas, suprasellar or hypothalamic tumors such as craniopharyngioma, inflammatory disease of either organ, or iatrogenic damage from surgery or radiotherapy. These pathologies may have distinct clinical features or influence the rate of onset of AI. Hypothalamic-pituitary lesions produce cortisol deficiency, but aldosterone is normal because it is regulated primarily by the renin-angiotensin system and serum potassium.

The actions and mechanisms of action of glucocorticoids and mineralocorticoids are discussed in detail elsewhere in this text (see Chapters 88-89). Cortisol has protean effects, including: immunomodulation with an initial immune enhancing effect on infection or injury followed by immune response containment to prevent healthy tissue damage; substrate mobilization increasing circulating glucose, fatty acids, and amino acids to maintain nutrition of key tissues in times of stress or starvation; cardiovascular support acting to maintain cardiac contractility; modulation of the vascular response to the β-adrenoceptor agonists; and neurocognitive effects so as to enhance attention and awareness, thereby facilitating an optimum response to stress. Aldosterone regulates exchange of sodium, potassium, and hydrogen ions at the renal tubule, the gut, and the salivary glands, promoting sodium reabsorption at the expense of potassium and hydrogen excretion.

CLINICAL FEATURES OF ADRENAL INSUFFICIENCY

Primary adrenal glucocorticoid insufficiency is manifest by weakness, fatigue, nausea, loss of appetite, vomiting, and weight loss, ACTH-mediated hyperpigmentation (typically of the palmar creases, scars, knuckles, and oral mucosa) (Fig. 93.1), hypotension, often postural, with tachycardia, because of reduced stroke volume, and decreased peripheral vascular resistance. Some cases may present with hypoglycemia. Depressed mood and reduced ability to concentrate are common. Back and leg pains may develop. Concomitant mineralocorticoid deficiency is manifest by hyponatremia, hyperkalemia, and metabolic acidosis. Hyperpigmentation is evident in palmar creases, scars, knuckles, and oral mucosa.

In SAI, isolated glucocorticoid insufficiency lead to the effects listed earlier for PAI. Although hyperkalemia does not occur, hyponatremia develops because of arginine vasopressin antidiuretic hormone ADH-mediated water retention as a physiological response to hypotension/volume depletion with exacerbation from the vasopressin-stimulating effect of nausea.

Clinical features of AI may arise acutely or chronically. The insidious, nonspecific evolution of symptoms in most patients leads to delayed diagnosis, sometimes in excess of 1 year. Adrenal crisis may supervene before diagnosis of AI, with an attendant mortality risk. Prediagnosis considerations sometimes include mood disorders, anorexia nervosa, or malignancy.

Cortisol deficiency leads to derepression of inflammatory cytokines, leading to increased body temperature with various patterns of pyrexia and the sickness syndrome.[20] In females, decreased axillary and pubic hair, loss of libido because of loss of adrenal androgens, and amenorrhea may occur because of weight loss.

Routine laboratory tests may show hyponatremia, mild metabolic acidosis, and prerenal azotemia, but hyperkalemia (in PAI) is more specific and facilitates diagnosis. Although hyperkalemia is typical, many patients have normal potassium at least in part because of frequent vomiting. Less common laboratory abnormalities include hypoglycemia (more frequent in children) and hypercalcemia. Hematologic abnormalities associated with AI include normochromic, normocytic anemia, relative lymphocytosis, and eosinophilia.

Adrenal crisis is a medical emergency with generally severe features of AI and an acute deterioration in health associated with absolute hypotension (systolic blood pressure <100 mm Hg) or relative hypotension (systolic blood pressure at least 20 mmHg lower than usual); features resolve within 1 to 2 hours after parenteral glucocorticoid administration (i.e., a marked resolution of hypotension within 1 hour and improvement in clinical symptoms over a period of 2 hours).[20]

KEY POINTS

- Adrenal insufficiency (AI) is uncommon and arises from lesions of the adrenal gland, pituitary, or hypothalamus.
- Pathologies such as autoimmune disease, tumors, surgery, genetic disorders, drugs, or exogenous glucocorticoids account for the majority of cases.
- This diversity of causes implies that AI may develop in any clinical setting. Relatively nonspecific clinical presentations in unfamiliar settings often leads to hazardous diagnostic delay.

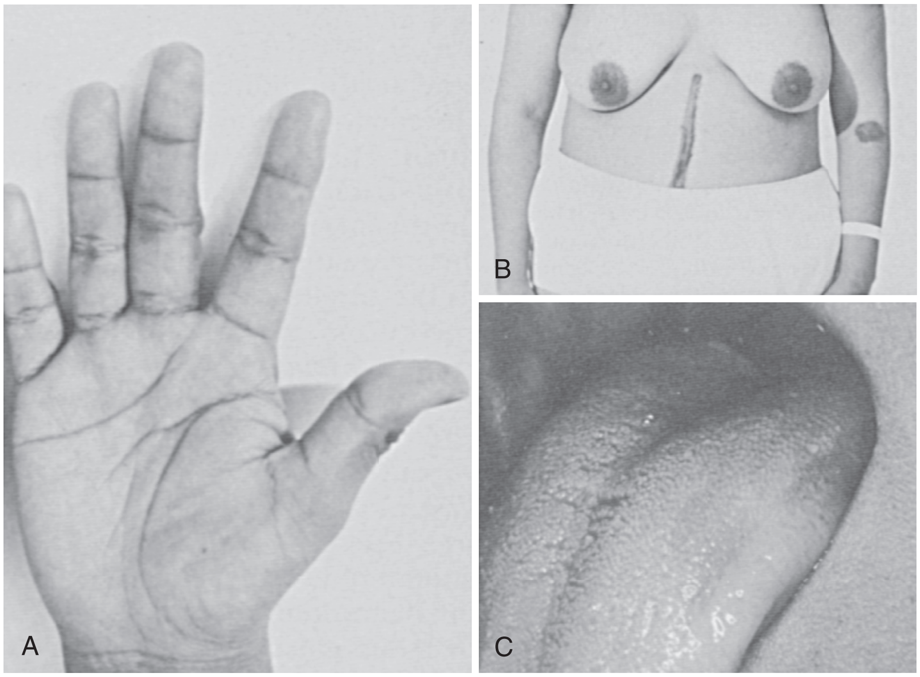


Fig. 93.1 Hyperpigmented hand (A), scar and areolae (B), and buccal mucosa (C). (From Loriaux L, Cutler Jr G. Disease of the adrenal gland. In: Kohler P, ed. *Clinical Endocrinology*. New York: Churchill-Livingstone; 1986:211.)

ETIOLOGY

Primary Adrenal Insufficiency

Primary AI diverse etiologies, as listed in Table 93.1.

Autoimmune Adrenal Insufficiency. In high-income countries, 80% to 90% of patients with PAI have autoimmune adrenalitis, which can be isolated, or seen as part of an autoimmune polyendocrine syndrome, with autoimmune thyroid disease the most frequent association.[21] Autoimmune Addison disease is associated with antibodies to adrenal biosynthetic enzymes, most often to the 21-hydroxylase enzyme, detectable in the serum of more than 90% of recent-onset patients. The cumulative risk of developing autoimmune Addison disease in the presence of 21-hydoxylase antibodies was 39%, somewhat higher in males, with median follow-up of 10 years.

Autoimmune polyendocrine syndrome type II (APS-II). Approximately 50% of patients with Addison disease exhibit an autoimmune polyendocrinopathy. APS-II was previously defined as autoimmune AI combined with either autoimmune thyroid disease (Schmidt syndrome) and/or type 1 diabetes mellitus (T1DM). A wider definition is now applied incorporating numerous other organ-specific autoimmune diseases such as vitiligo (Fig. 93.2), alopecia, autoimmune gastritis with iron and vitamin B12 deficiency, premature ovarian insufficiency, celiac disease, and, more rarely, myasthenia gravis, central diabetes insipidus, and lymphocytic hypophysitis (Table 93.2).[22] APS-II shows familial aggregation, with a wide variation in number and type of components between family members. Autoimmune disease presents mainly in the third or fourth decade, with a female-to-male ratio of 1 to 1.8:1. APS-II is the most common immunoendocrinopathy, with an estimated prevalence of 5 to 14 cases per 100,000 population.

Autoimmune Addison disease is polygenic in origin. The strongest genetic association is with specific haplotypes of human leukocyte antigen class 2 (HLA class II), including DR3-DQ2 and DR4*0404-DQ8 and the cytotoxic T cell antigen (CTLA-4).[22] Other genes associated with autoimmune Addison disease include *BACH2* and *PTPN22*. Many of these genetic associations overlap with related disease such Hashimoto thyroiditis, Graves disease, T1DM, and celiac disease, reflecting the high prevalence of APS-II.

The environmental factors precipitating a break in tolerance to the adrenal cortex are largely unknown. Herpes virus infection has been suggested as a potential factor, as the virus may infect the adrenal and lead to adrenal dysfunction.[23]

Autoimmune polyendocrine syndrome type I (APS-I) is also known as autoimmune polyendocrinopathy, mucocutaneous candidiasis, and ectodermal dystrophy. This is a rare, monogenic, autosomal recessive disease that is most prevalent in certain stable populations, including Finns and Persian Jews. The gene that causes this syndrome is located on chromosome 21q22 and encodes a protein known as autoimmune regulator (molecular pathogenesis covered in Chapter 132). To date, more than 120 mutations have been reported, some of which are dominant and give rise to milder phenotypes.

The typical presentation is persistent *Candida* infection of the skin and mucous membranes, without severe systemic infection, in childhood, followed by hypoparathyroidism and then adrenal failure.[24] Almost all patients display autoantibodies against interferon omega and alpha.[25]

X-linked polyendocrinopathy immune dysfunction and diarrhea is a rare inborn error of immune regulation that presents as neonatal diabetes and is often fatal. The disorder is also known as X-linked autoimmunity and allergic dysregulation and immune dysfunction, polyendocrinopathy, and enteropathy, X-linked (IPEX). IPEX is caused by mutations in *FOXP3*, which is a critical determinant of CD4+ and CD25+ T regulatory cell development and function. See Chapter 132.

Infectious Adrenalitis. Infectious diseases represent the most common cause of primary adrenal failure worldwide, with generalized tuberculosis being the most frequent single cause.[23] On abdominal computed tomography (CT), an enlarged adrenal with necrotic areas

TABLE 93.1 Causes of Primary Adrenal Insufficiency

Diagnosis	Pathogenesis
Autoimmune Adrenal Insufficiency	
Autoimmune Addison disease	Antibodies against 21-hydroxylase most common, other steroidogenic enzymes such as 17α-hydroxylase, side-chain cleavage enzyme, and *CTLA-4*
APS type I or APCED	Mutations in *AIRE* (autoimmune regulator 1)
APS type II	Multiple genes including *HLA DR3/4, DRBI, DQAI, DQBI, BACH2,* and *CTLA-4*
X-linked polyendocrinopathy	*FOXP3*
Infectious	
Tuberculosis	Tuberculous adrenalitis
Systemic fungal infections	Histoplasmosis, cryptococcosis, blastomycosis
AIDS	Opportunistic infection with CMV, bacteria, protozoa, Kaposi sarcoma
Adrenal infiltration	Metastases, lymphoma, sarcoidosis, amyloidosis, hemochromatosis, histiocytosis (Erdheim–Chester disease)
Bilateral adrenalectomy	Indications include ACTH-independent Cushing syndrome or persistent ACTH-dependent Cushing syndrome, bilateral pheochromocytoma
Medication-induced	Mitotane, ketoconazole, and etomidate
Bilateral adrenal hemorrhage	Meningococcal meningitis, antiphospholipid syndrome, anticoagulation therapy, coagulation disorder, septic shock
Genetic Disorders	
Congenital adrenal hyperplasia	Mutations in the genes encoding 21-hydroxylase (*CYP21A2*), 11β-hydroxylase (*CYB11B1*), 17-hydroxylase (*CYP17A1*), or 3β-hydroxysteroid dehydrogense type 2 (*HSD3B2*), PORD
Adrenoleukodystrophy/adrenomyeloneuropathy	Excessive accumulation of very long chain fatty acids in brain and adrenal cortex; mutations in the *ABCD-1* gene
Familial Glucocorticoid Deficiency	
Type 1	Melanocortin 2 receptor (*MC2R*) mutation
Type 2	Melanocortin 2 accessory protein (*MRAP*) gene mutation
Type 3	Genetic origin unknown
Congenital lipoid adrenal hyperplasia	Mutations in the gene encode the steroidogenic acute regulatory protein (*StAR*)
Triple A syndrome	Mutations in *AAAS* encoding ALADIN
Congenital adrenal hypoplasia	Mutation in the orphan nuclear receptor DAX-I (*NROB1*), *SF-I*, or other unidentified genes

ACTH, Adrenocorticotropic hormone; *AIDS*, acquired immunodeficiency syndrome; *APCED*, autoimmune polyendocrinopathy, mucocutaneous candidiasis, and ectodermal dystrophy; *APS*, autoimmune polyendocrine syndrome; *CMV*, cytomegalovirus; *PORD,* cytochrome P450 oxidoreductase deficiency.

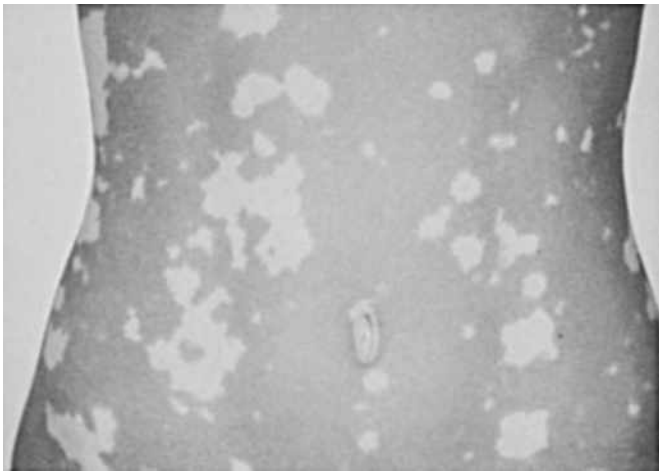


Fig. 93.2 Vitiligo. (From Loriaux L, Cutler Jr G. Disease of the adrenal gland. In: Kohler P, ed. *Clinical Endocrinology*. New York: Churchill-Livingstone; 1986:211.)

can be seen in the early stages, while later stages are characterized by adrenal calcification (Fig. 93.3). All the clinically important fungi except *Candida* can also cause AI. The most common is histoplasmosis, which is particularly prominent in the Ohio and Tennessee River Valleys and along the Piedmont Plateau of the Middle Atlantic States, as well as in Southern India. South American blastomycosis is the next most common fungal cause of AI, followed by North American blastomycosis, coccidioidomycosis, and cryptococcosis, although all are rare causes of adrenal destruction. The pathophysiology of this process is much like tuberculosis, with early adrenal enlargement, healing, calcification, and adrenal atrophy or resumption of normal volume.

Acquired immunodeficiency syndrome (AIDS) can be associated with AI in its late stages. The adrenals are involved with infection or tumor in well over half of autopsy cases, although less than 50% of the adrenal gland is destroyed in 97% of cases. This explains the rarity of overt symptoms. Cytomegalovirus infection of the adrenal glands is common in AIDS, as is infection with *Mycobacterium avium intracellulare* and fungi that can colonize and destroy the adrenal glands. The plasma cortisol response to ACTH administration, however, is abnormal in only 10% to 15% of affected cases.

Medications and Adrenal Insufficiency. Although fluconazole, ketoconazole, phenytoin sodium, rifampicin, and barbiturates may precipitate AI, they are rarely the cause, except in the case of high-dose azole antifungals, where the intention is to inhibit adrenal steroidogenesis to treat endogenous Cushing syndrome. One exception is the anesthetic agent etomidate, an inhibitor of the adrenal steroidogenic enzyme 11β-hydroxylase, which can induce AI, with an attendant increase in mortality, as initially noted in trauma victims.[26]

Adrenal Infiltration. Metastases to the adrenal gland can be found in as many as 60% of patients with disseminated breast or lung cancer

TABLE 93.2 **Features of Autoimmune Polyendocrine Syndrome (APS)-I, APS-II, and X-Linked Polyendocrinopathy**

Features	APS-I	APS-II	X-Linked Polyendocrinopathy Immune Dysfunction and Diarrhea
Prevalence	Rare	Common	Very rare
Time of onset	Infancy	Adulthood	Neonatal period
Genetic and inheritance	Monogenic, *AIRE*	Polygenic	*FOXP2*, X-linked
Immunodeficiency	Asplenism, susceptibility to candidiasis	None	Overwhelming autoimmunity Loss of regulatory T-cells
Adrenal insufficiency	60%–70%	40%–50%	+
Diabetes mellitus	<20%	50%–60%	80%
Autoimmune thyroid disease	10%	70%–75%	+
Premature ovarian failure		Up to 21%	
Hypoparathyroidism	80%–85%	0%–5%	–
Mucocutaneous candidiasis	70%–80%	Nil	–
Male hypogonadism	12%	Rare	–
Hypopituitarism	0%–2%	<0.1%	–
Chronic active hepatitis	+	Rare	+
Pernicious anemia	+	0.1%–10%	–
Skin manifestation	Vitiligo, alopecia 10%–15%	Vitiligo, 5%–10%	Eczema, psoriasis, or atopic dermatitis
Gastrointestinal	Diarrhea, constipation	Celiac, 4%–9%	Enteropathy, malabsorption

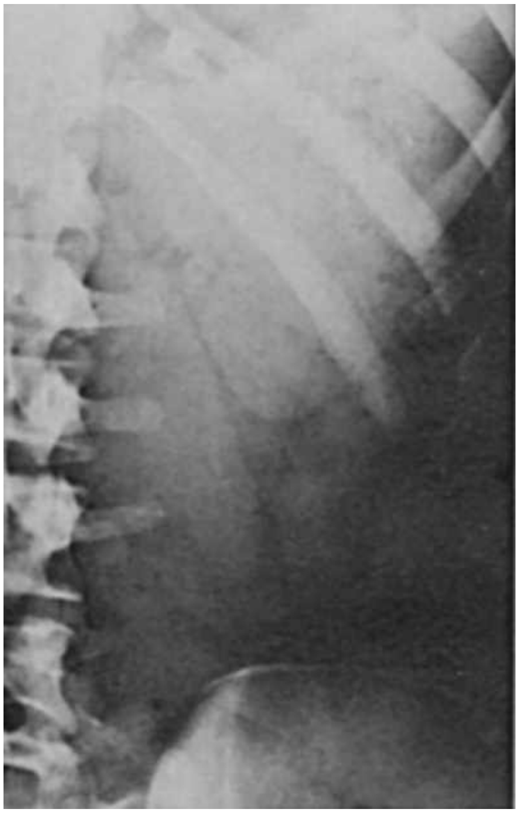


Fig. 93.3 Adrenal calcification on "flat plate" of the abdomen. (From Loriaux L, Cutler Jr G. Disease of the adrenal gland. In: Kohler P, ed. *Clinical Endocrinology*. New York: Churchill-Livingstone; 1986:211.)

and are often bilateral. Although generally considered uncommon on clinical grounds, PAI was reported in 12.4% of patients with bilateral adrenal metastases, and in 20% of patients with bilateral metastases and adrenal mass size greater than 4 cm. Tumors that are commonly associated with AI are cancers of the lung, kidney, stomach, colon, and breast, as well as melanoma and lymphoma.

A further occasional cause of adrenal failure is amyloidosis,[27] which is often underdiagnosed and masked by other clinical manifestations of the disease. Likewise, severe untreated hemochromatosis has been reported to cause AI, but is now rarely seen. In addition, Erdheim–Chester disease can cause PAI as a result of adrenal infiltration.

Bilateral Adrenalectomy. This procedure may be used to treat persistent ACTH-dependent Cushing disease or bilateral micronodular or macronodular hyperplasia with ACTH-independent Cushing syndrome, or for bilateral pheochromocytomas, and results in abrupt-onset PAI.

Adrenal Hemorrhage. With the advent of the abdominal CT, adrenal hemorrhage is recognized more frequently as a cause of AI. Early reports of adrenal hemorrhage and its consequences followed cases of meningococcal septicemia, particularly in children and young adults (Waterhouse–Friedrichsen syndrome). The usual setting is a stressed individual anticoagulated for the prevention of pulmonary emboli or other thrombotic phenomena, frequently after surgery. Other scenarios include trauma, hemorrhagic shock, sepsis, extensive burns, or during adrenal vein sampling for investigation of primary aldosteronism. Adrenal hemorrhage may be a manifestation of the antiphospholipid syndrome. Typically, the patient will complain of back pain, which is followed in a few days by the acute onset of the first signs and symptoms of AI, without the usual hyperpigmentation. These patients rarely recover adrenal function.

Genetic Disorders. *CAH* is a family of inborn errors of steroidogenesis, each characterized by a specific enzyme deficiency that leads to a range of adrenal hormone abnormalities depending on the specific enzyme affected. Adrenal steroid abnormalities include impairment of cortisol and aldosterone synthesis from the adrenal, excess or reduced sex steroid synthesis, and excess mineralocorticoid precursors. These autosomal recessive diseases involve impaired enzymatic function arising from gene mutations affecting are the cytochrome P450 enzymes 21-hydroxylase, 11β-hydroxylase, and 3β-hydroxylase dehydrogenase and, less often, 17α-hydroxylase/17,20-lyase, cholesterol desmolase, and the more recently described mutations of P450 oxidoreductase, the electron donor for all CYP450 enzymes.[28] Over 95% of cases of CAH have 21-hydroxylase deficiency with varying degrees of deficiency of cortisol and aldosterone synthesis, elevations of ACTH, and consequent excess adrenal androgens. 21-hydroxylase deficiency is the most common cause of AI in children. Detailed descriptions of the CAH variants and other genetic forms of altered adrenal steroidogenesis are found in Chapter 94.

Congenital lipoid adrenal hyperplasia is a rare form of adrenal steroidogenic defect and is inherited in an autosomal recessive pattern. The disease results from mutations in the gene on chromosome 8p11 that encodes the steroidogenic acute regulatory protein, which regulates cholesterol uptake into the mitochondria in readiness for conversion to pregnenolone, the first step of steroidogenesis.[29]

Adrenoleukodystrophy (ALD) and *adrenomyeloneuropathy (AMN)* are two clinical presentations of the same disorder that may exhibit a broad phenotype. ALD, also known as *Brown Schilder disease* (*brown* being an adjective describing the hyperpigmentation of the skin) or *sudanophilic leukodystrophy,* is typically a disease of children characterized by rapidly progressive central demyelination that eventuates in seizures, dementia, cortical blindness, coma, and death, usually before puberty. However, clinical manifestations can be highly variable, and long survival may occur. Adrenomyeloneuropathy is a disease of young adults that is characterized by a slowly progressive mixed motor and sensory peripheral neuropathy associated with an upper motor neuropathy, leading to an ascending spastic paraparesis. Both forms are associated with progressive failure of steroid-secreting cells leading to adrenal and gonadal failure, but adrenal failure may occur in isolation. Usually ALD is an X-linked disorder, but it may be underrecognized as a cause of adrenal failure. In one large series, it accounted for a significant proportion (8%) of young adult males believed to have autoimmune Addison disease.[30] The metabolic marker for these diseases is an elevated circulating level of very long chain fatty acids (VLCFAs), C26 and greater in length, that rises in response to the primary abnormality, which is a peroxisomal defect in VLCFA metabolism. Peroxisomes are small, intracytoplasmic, membrane-enclosed bodies that contain the enzyme pathways for a number of metabolic and detoxification processes. The gene underlying ALD, identified in 1993 on chromosome Xq28, encodes a half-transporter of the adenosine triphosphate (ATP)-binding cassette membrane transporter class. The protein includes six transmembrane domains and an ATP-binding site and probably functions as a heterodimer with another half-transporter to regulate the uptake of VLCFA into peroxisomes. Many missense, nonsense, and splicing defects in the *ALD* gene have been described, and accumulation of VLCFA in the adrenals seems to be the pathogenic mechanism, although the neuronal pathogenic process may differ. Several treatments have been tried, including Lorenzo's oil containing a mixture of oleic and erucic fatty acids, but only autologous bone marrow transplantation appears to be effective.

Familial glucocorticoid deficiency (FGD) is an autosomal recessive syndrome characterized by cortisol deficiency despite high plasma ACTH and a preserved renin-aldosterone axis.[31] There is considerable phenotypic variation. Usually, patients present in early childhood with clinical AI and undetectable circulating cortisol, although some pass unrecognized until later childhood, and the diagnosis is then made after an ACTH stimulation test. Children tend to be tall for unknown reasons. Patients tend to be highly resistant to glucocorticoid suppression of hyperpigmentation, perhaps as a result of impaired short-loop ACTH feedback. The first molecular abnormality demonstrated was a mutation of the gene encoding ACTH receptor, located on chromosome 18p11, although it is now known that these cases, known as FGD type 1, represent only some 25% of cases of FGD. FGD type 2 is caused by a mutation of the chromosome 21 gene encoding the melanocortin 2 receptor accessory protein. This accounts for approximately 20% of all FGD cases, implying that approximately half of all FGD cases result from other genes. Two additional causative genetic mutations involving minichromosome maintenance-deficient 4 homolog, nicotinamide nucleotide transhydrogenase, and thioredoxin reductase 2 have been identified in FDG cohorts.

Triple A syndrome (Allgrove syndrome) is the association of alacrima, achalasia, and Addison disease. It is associated with a variety of progressive motor, sensory, and autonomic neurologic defects plus mineralocorticoid insufficiency. This recessive condition results from abnormalities in *AAAS* on chromosome 12q13, which encodes a putative nuclear pore complex protein, ALADIN. Clinical features are highly variable, even within families with the same mutation.

Congenital adrenal hypoplasia (also known as *adrenal hypoplasia congenita*) most often occurs as an X-linked disorder caused by mutation in the nuclear receptor DAX1 (NR0B1),[32] which is expressed in the adrenal cortex, gonads, and gonadotrophs, and the ventromedial nucleus of the hypothalamus.[33] Boys with this condition tend to present with salt-losing PAI in infancy (60%) or childhood (40%), followed by hypogonadotrophic hypogonadism secondary to abnormal development of both the hypothalamus and the pituitary. In addition, these boys show disordered tubular structures and Leydig cell hyperplasia. Recent evidence points to the existence of a milder form, with onset in adult life presenting with mild AI and partial hypogonadotropic hypogonadism. The genetic basis for the recessive form of congenital adrenal hypoplasia remains unknown, but homozygous and heterozygous mutations have been described in the *SF-1* gene that produce primary adrenal failure and a female phenotype complete with uterus in an XY male.

Secondary Adrenal Insufficiency. SAI implies a pituitary etiology; in common use it often incorporates hypothalamic (formally tertiary) etiologies. Adrenal insufficiency may develop after prolonged use of exogenous glucocorticoids, reflecting incomplete escape from negative feedback suppression of the HPA axis. This may be termed secondary/tertiary AI, but its unique characteristics (potential reversibility over time with glucocorticoid weaning and its often incomplete extent) suggest it may be considered separately using terms such as adrenal suppression or exogenous AI. The causes of SAI are listed in Table 93.3. The most common cause is HPA axis suppression of CRH and ACTH secretion from prolonged (at least 2–4 weeks) exogenous glucocorticoid administration with a dose exceeding endogenous cortisol production, equating to standard adult replacement doses such as hydrocortisone 15 to 25 mg daily, prednisolone 3 to 5 mg daily, or dexamethasone 0.3 to 0.5 mg daily. Approximately 3% of adults are using glucocorticoids for indications other than adrenal hormone replacement. Adrenal suppression has characteristics that differ from other etiologies of SAI. These include potential reversibility with weaning of exogenous glucocorticoid, high degree of interindividual variability in suppression extent and duration, and a lack of other hypothalamic and pituitary deficits. The likelihood of adrenal suppression, its extent, and time to recovery depend on exogenous glucocorticoid dose, duration, and administration schedule. High-dose glucocorticoid administration for prolonged periods may lead to adrenal suppression lasting 12 to 18 months. Adrenal suppression is most often iatrogenic but may also develop after treatment of endogenous hypercortisolism, following successful surgical or medical treatment of endogenous Cushing syndrome of pituitary, adrenal, or ectopic ACTH tumoral origin (see Chapter 7, 97). Rarely, topical steroid cream can cause adrenal suppression, as can high-dose progesterone such as megestrol therapy. SAI is a risk for patients with inhaled high-dose potent inhaled glucocorticoids

TABLE 93.3 **Causes of Secondary Adrenal Insufficiency**

1. Hypothalamic-pituitary-adrenal suppression
    - A. Exogenous
        - i. Glucocorticoids
        - ii. Opioids
    - B. Endogenous: Cushing syndrome, ACTH-dependent, or ACTH-independent
2. Lesions of the hypothalamus or pituitary gland
    - A. Neoplasm
        - i. Pituitary tumor
        - ii. Metastatic tumor
        - iii. Craniopharyngioma
        - iv. Meningioma
    - B. Craniopharyngioma
    - C. Immune hypophysitis
    - D. Infection
        - i. Tuberculosis
        - ii. Actinomycosis
        - iii. *Nocardia*
    - E. Immune checkpoint inhibitors
3. Sarcoid
4. Head trauma
5. Isolated deficiency of ACTH

ACTH, Adrenocorticotropic hormone.

used in the treatment of asthma.[34] The use of perioperative stress-dose glucocorticoids was instituted after early reports of adrenal crises in relation to surgery; however, traditional doses may be excessive, and newer regimens are likely to produce fewer side effects.[35] Overall, hospitalization with AI or adrenal crisis in patients using exogenous glucocorticoids is rare.[36]

Opioid-Induced Hypothalamic-Pituitary-Adrenal Axis Suppression. Widespread use of opioids for chronic noncancer pain has led to reports of hypoandrogenism, more frequently in men. Suppression of the HPA axis may follow opioid receptor stimulation affecting hypothalamic CRH production and/or superadded peripheral mechanisms; naloxone leads to prompt sequential ACTH and cortisol release when used in high doses, suggesting that relatively low-affinity opioid receptors are involved.[37] Low cortisol and sex steroids potentially add to the disability associated with long-term opioid use. Adrenal insufficiency appears to be related to the cumulative opioid exposure (dose times and duration) and may be diagnosed in 9% to 15% of patients taking chronic opioids.[13] Adrenal crisis has been reported in opioid-induced AI.[38] Recovery of adrenal function may occur following the cessation of opioids.

Other Causes of Secondary Adrenal Insufficiency. Apart from adrenal suppression, other causes of SAI require investigation with a dedicated pituitary magnetic resonance imaging (MRI), with gadolinium unless contraindicated. CT scans and head MRI generally provide less anatomical definition, although CT is sometimes used to reliably delineate calcification. The majority of cases of SAI are structural, such as tumors and granulomatous disease, the result of surgery or radiotherapy in the region, or hypophysitis resulting mostly from autoimmune or immune stimulatory therapies for cancer (see Table 93.3). However, there are many causes, and the imaging characteristics and clinical context are crucial in differential diagnosis. Occasionally, biopsy is required for pathological diagnosis, but often, the risks of biopsy exceed the benefits, and observation with serial imaging is preferred.

The most common tumor etiology of SAI is pituitary adenomas, which can damage pituicytes, generally when in the macroadenoma size range (>10mm), or following pituitary surgery or radiotherapy aimed at controlling pituitary adenomas. Pituitary adenomas may become evident although hypersecretory phenomena resulting from excess prolactin, growth hormone, or ACTH and rarely, thyroid-stimulating hormone (TSH). Hypersecretion of gonadotropins or their fragments is usually asymptomatic. Pituitary tumors of all types often present with structural complications from impingement of adjacent structures such as the optic chiasm/nerves (visual field defects classically bitemporal hemianopia) and extraocular nerves within the cavernous sinus (diplopia). Pituitary tumors may present with SAI and loss of other pituitary hormones. Such losses are often time-sequenced (in order: growth hormone, gonadotropins, TSH, ACTH), so that SAI tends to develop late in the process of compressive hypopituitarism.

Other tumors in the sella turcica and parasellar region may lead to SAI. These include metastases that may progress rapidly, unlike the usual slow-growing pituitary adenomas. Rathke's cleft/pouch cysts may occasionally cause ACTH deficiency and may be complicated with proteinaceous material or hemorrhage so they may be indistinguishable from adenomas on MRI. Craniopharyngiomas arise from craniopharyngeal duct epithelium and may be located in the sella turcica, in the suprasellar region, or in the third ventricle floor. They are divided into adamantinomatous and papillary histological types arising from tumor mutations of the *β-catenin* and BRAFV600E genes, respectively.[39] Treatment with surgery and radiotherapy may lead to both pituitary (hormone deficits) and hypothalamic (appetite, weight, temperature, neurocognitive deficits) injury.

Nontumor pathologies in the pituitary can lead to SAI. These include inflammation of lymphocytic type, considered likely to be autoimmune, which may present with an acute inflammatory mass, especially in relation to pregnancy or following IgG4-related disease.[40] Silent autoimmunity may produce hypopituitarism, including isolated ACTH deficiency, and hence SAI. Infections in the region include bacterial pituitary abscess or chronic infections such as actinomycosis and nocardiosis. Granulomatous disorders affecting the pituitary include sarcoid, amyloid, and Langerhan's cell histiocytosis, all of which represent systemic disorders that may produce disease outside the pituitary or may be isolated to the pituitary. Haemochromatosis can result in ACTH deficiency, believed to be secondary to hypothalamic injury. Sheehan syndrome, or hypopituitarism after childbirth, generally follows a postpartum hemorrhage with shock and hypoperfusion to an enlarged hypervascular pituitary with consequent infarction, classically presenting with loss of lactation. Head trauma may also lead to SAI and other pituitary hormone deficits.[41]

Checkpoint inhibitor therapy, used for melanoma and other malignancies, includes the CTLA-4 inhibitor ipilimumab, which has a dose-related risk of hypophysitis, generally isolated ACTH deficiency that is not reversible, in up to 20% of treated cases and may rarely produce PAI (1% risk); PD-1 antibodies and PD-L1 antibodies incur lower rates of AI.[12]

Inherited forms of isolated ACTH deficiency present in childhood.[42] Some of these patients demonstrate mutations in the *T-Pit* (*TBX-19*) gene, which encodes an essential transcription factor for *POMC* gene expression in corticotroph cells. Other genetic forms of SAI include mutations of key pituitary transcription factors, *PROP1*, *PIT-1,* and the *POMC* gene itself. SAI may also follow gene defects involving hypothalamic function such as those of the *CRH* gene or the gene encoding the CRH receptor. Several developmental disorders of the brain may be associated with hypopituitarism, likely from hypothalamic dysfunction, including septooptic dysplasia and optic nerve dysplasia.

CRITICAL ILLNESS-RELATED CORTICOSTEROID INSUFFICIENCY

Patients with critical illness are subject to profound physiological stress, with multiple stimuli such as fever, elevation of inflammatory cytokines, pain, and hypovolemia leading to activation of the HPA axis. Cortisol has a pivotal role in maintenance of homeostasis and in immunomodulation in critical illness. Routine measurements of serum cortisol may not reflect end-organ glucocorticoid sufficiency because of tissue resistance to cortisol and reduced concentrations of the principal carrier proteins for cortisol, corticosteroid-binding globulin (CBG) and albumin. Free cortisol correlates more closely with illness severity. Septic shock, a common state and important cause of mortality, which has a 30% to 40% mortality rate despite standard antibiotics, fluids, and catecholaminergic vasopressors, has been studied extensively over the past two decades from the point of view of glucocorticoid therapy. Earlier studies of high-dose short-term synthetic glucocorticoids for septic shock showed increased mortality. Continuing clinical observations of benefit from hydrocortisone in septic shock led to a succession of small trials revealing salutary effects such as a pressor response[43] and reduced inflammatory cytokines,[44] ultimately leading to progressively larger multicenter randomized controlled trials of "low-dose" hydrocortisone. The general conclusion is that randomly administered hydrocortisone, generally 200 mg daily, leads to reduced time-to-withdrawal of vasopressor support (1–2 days) and a possible small benefit in mortality.[45-47] Sicker patients and, curiously, those treated with fludrocortisone may benefit more in terms of mortality. It may be that a subset of patients benefit more from hydrocortisone. A state of reduced cortisol response to cosyntropin (ACTH 1–24) 250 μg (<9 μg/dL [<248 nmol/L] at 60 mins) may identify a state of relative AI, later referred to as CIRCI, although routine ACTH stimulation testing is not supported on a consensus basis.[15] Mechanistically, CIRCI has been considered to potentially result from adrenal damage, inhibition of steroidogenesis, reduced hypothalamic-pituitary drive, reduced cortisol transport and delivery—inhibition of CBG synthesis, reduced albumin, CBG cleavage-mediated exhaustion of CBG, reduced cortisol metabolism (5α reductase in the liver, 11β-hydroxysteroid dehydrogenase in the kidney), and glucocorticoid resistance—and excessive NF-κB activation inhibits GR-α activity.[48] Initial studies suggested that CIRCI predicted mortality benefit from parenteral hydrocortisone, but this was not confirmed in larger trials. Early studies in septic shock with higher rates of CIRCI appear to have been affected by the use of etomidate[49] or compromised by insufficient statistical power. The two largest trials in septic shock have shown benefits in required duration of vasopressor treatment, but only the study with the higher mortality rate showed a mortality benefit.[45,47]

At present, hydrocortisone 200 mg daily, 50 mg four times daily, or by intravenous infusion for 7 days without taper, perhaps with fludrocortisone 0.05 mg enterally via nasogastric tube daily, is considered indicated in patients with hypotension despite fluid resuscitation and vasopressor treatment.[15] The effects of cortisol concentrations achieved by these doses (~3000 nmol/L, 109 μg/dL) include attenuation of the inflammatory cytokines IL-6 and IL-8; soluble E-selectin, a factor important in endothelial-granulocyte interactions; expression of the neutrophil activation markers CD11b and CD64; and the antiinflammatory cytokines IL-10 and soluble TNF receptors I and II. Markers of monocyte function such as HLA-DR antigen expression and *in vitro* phagocytosis are not affected, nor are the Th1 function markers IL-12 and IFN-γ.[50] Overall, these hydrocortisone responses are considered antiinflammatory rather than immunosuppressive.[50] Other critical illnesses where hydrocortisone therapy may be clinically indicated include hospitalized community-acquired pneumonia, acute respiratory distress syndrome, and bacterial meningitis.[48] A place for hydrocortisone in severe trauma is not established.

DIAGNOSIS OF ADRENAL INSUFFICIENCY

In times of clinical urgency, serum cortisol is the most important test in diagnosing AI. Very low or clearly normal concentrations are useful, but circadian and ultradian pulsatile cortisol secretion make random cortisol testing unreliable. A morning plasma cortisol below 3 μg/dL (83 nmol/L) is indicative of AI, whereas cortisol above 19 μg/dL (525 nmol/L) rules out the disorder. Morning serum cortisol concentrations within 1 hour of waking are generally high; hence, low cortisol values at this time are suspicious for AI. The specificity of morning cortisol concentrations in ruling out AI, obviating the need for ACTH stimulation testing, depends on the cortisol level chosen—for example, in one study a cortisol, a cutoff of greater than 236 nmol/L (8.6 nmol/L) was 71% specific for ruling out AI, and a cutoff of 375 nmol/L (13.6 μg/dL) was 95% specific.[51] Basal cortisol should be taken as soon as possible after waking because serum cortisol concentrations fall by 30 nmol/L (1.2 μg/dL) per hour between 7:00 AM and 12:00 PM.[52] In clinical practice the extent to which AI needs to be ruled out will take into account clinical factors that increase or decrease the degree of clinical suspicion. For example, the presence of hypertension makes AI very unlikely. In patients with typical clinical features of AI, the presence of a very low serum cortisol is likely to be diagnostic and should be met with a rapid, often within hours, recovery of well-being and function that is itself confirmatory. Similarly, values less than 7 μg/dL (200 nmol/L) in severely unwell individuals with features of AI are suspicious for AI. In the appropriate context, given a clear clinical picture, it is inappropriate to delay glucocorticoid treatment to allow the more definitive testing. Cosyntropin, if required, can be administered after appropriate resuscitation. Common HPA axis pharmacologic stimuli include cosyntropin, insulin hypoglycemia, or metyrapone.[53] The cortisol responses to these tests have generally been compared among tests in patients and healthy controls, rather than related to the clinical effects of glucocorticoid administration.

Adrenocorticotropic Hormone Stimulation Testing

Stimulation of the adrenals with synthetic ACTH, comprising N-terminal amino acids 1–24 of the complete 39-amino acid peptide, in a dose of 250 μg delivered intravenously (IV) or intramuscularly (IM) at any time of day, with a serum cortisol at baseline and at 30, and/or 60 minutes, has long been the standard of care test for cortisol deficiency. The test depends on the fact that the normal adrenal gland, under maximal and acute ACTH stimulation, produces a peak plasma cortisol concentration of approximately 18 μg/dL (500 nmol/L) (assay-dependent) at the lower boundary of the normal range. Cortisol assays vary widely, and it is important to refer to ranges supplied by assay manufacturers.[54] Compared with liquid chromatography tandem mass spectroscopy (LCMS-MS), immunoassays generally overestimate cortisol concentration. Cosyntropin stimulation is rapid, insensitive to interference from diet or medication, simple to interpret, and reliable, and it can be applied to people of all ages without fear of untoward effects, although anaphylaxis has been reported rarely. Similar times and criteria have been demonstrated after IM administration.

A normal cortisol response to cosyntropin renders AI extremely unlikely. However, some factors need to be considered. First, this is a test of *adrenal* function; therefore, if it is used as an assessment of *pituitary*-adrenal function, the indirect nature of the test must be appreciated. For example, although hypothalamic-pituitary failure eventually will lead to adrenal hypofunction, and a close agreement between tests of HPA function has been demonstrated, there are exceptions. In particular, a patient who has been recently rendered hypopituitary after pituitary surgery may show a maintained response to cosyntropin stimulation for 4 to 6 weeks while being profoundly ACTH and cortisol deficient.[55] In addition, patients using oral estrogen have higher total cortisol levels because of increased synthesis of CBG and can be misinterpreted as normal despite having AI. Finally, some patients with chronic fatigue syndrome or fibromyalgia may show marginally low cortisol responses to ACTH but use of glucocorticoid therapy is not recommended.

The 250-μg cosyntropin dose is the only dose commercially available. Cosyntropin represents ACTH 1–24, the bioactive N-terminal fragment of full-length ACTH (a 39-amino acid peptide). There is considerable clinical experience with its use, although ACTH concentrations after injection are clearly supraphysiologic. Some studies have suggested possible increased sensitivity of ACTH stimulation testing in patients with SAI when a low dose of 1 μg of ACTH is employed, this dose being closer to the peak of the dose-response curve for synthetic ACTH. However, a systematic review and metaanalysis of ACTH stimulation studies for diagnosis of SAI (n = 30 studies) and PAI (n = 5 studies) showed similar diagnostic accuracy of the two ACTH doses (1 μg and 250 μg).[56] The dilution of the prepared 1 μg/mL solution of ACTH used clinically may vary substantially, which introduces uncertainty to the test result.

Plasma ACTH separates patients with PAI from those with SAI. In primary disease, plasma ACTH usually is greatly increased, with values greater than 100 pg/mL (22.0 pmol/L).[21]

Overnight Metyrapone Test

Metyrapone is administered at midnight, as a single dose test at 30 mg/kg body weight. Measurements of the 11-deoxycortisol and cortisol are obtained the following morning. Metyrapone blocks the conversion of 11-deoxycortisol to cortisol by 11β-hydroxylase, inducing a rapid decrease of cortisol and an increase in its precursor 11-deoxycortisol because of the increased ACTH secretion. A failure of 11-deoxycortisol levels to rise following the overnight administration of metyrapone indicates either ACTH deficiency or primary adrenal disease. The overnight metyrapone test is mainly used to diagnose partial defects in ACTH secretion. A decline in 8:00 AM. serum cortisol to less than 7 μg/dL (172 nmol/L) confirms compliance with metyrapone administration, and an 8:00 AM serum 11-deoxycortisol of less than 7 to 10 μg/dL, or less than 7 μg/dL (202 nmol/L), is diagnostic of AI.[57]

Insulin Tolerance Test

Insulin-induced hypoglycemia is a well-known test of cortisol reserve.[58] This stimulus depends on intact hypothalamic stimulation of pituitary ACTH secretion and on the ability of the adrenal gland to respond to ACTH with the secretion of cortisol. Thus, a normal test indicates a functioning HPA axis. An abnormal test implies a lesion in this system that can be anywhere between the hypothalamus and the adrenal gland. The usual test is performed by administering 0.15 unit/kg of regular insulin IV as a bolus and measuring blood glucose and cortisol at 15-minute intervals throughout the subsequent 2 hours. The blood glucose level must fall below 45 mg/mL (<2.2 mmol/L) to ensure adequate stimulation for interpretation of the test. The normal response is a plasma cortisol of greater than 20 μg/dL (550 nmol/L, assay dependent) at any time during the test. The test requires close observation to avoid severe prolonged hypoglycemia but can be done safely in selected patients at low risk of coronary ischemia or seizures.

Adrenal Autoantibody Tests

As mentioned above, antibodies against the adrenal biosynthetic enzyme 21-hydroxylase are present in more than 90% of patients with recent-onset autoimmune adrenalitis and are recommended in all cases of PAI, unless the cause is obvious. Antibodies against the adrenal side-chain cleavage enzyme are present less frequently and associated with premature ovarian failure.

Serum 17-Hydroxyprogesterone

Serum 17-hydroxyprogesterone is the cornerstone of diagnosis of 21-hydroxylase deficiency. In many regions, it is measured in newborns to reduce risk of adrenal crises in the first week of life, particularly in boys, where genital ambiguity is not present at birth. This testing is followed by ACTH stimulation testing for confirmatory purposes. Further clarification may be obtained through *CYP21A2* gene sequencing, a measure that may be useful in genetic counseling. Serum 17-hydroxyprogesterone concentration diagnostic cutoffs are assay-specific.

Very Long Chain Fatty Acids

Males with isolated AI, without evidence of autoimmunity, should be tested for VLCFAs to exclude X-linked ALD or AMN. Up to 20% of males with elevated VLCFA levels may have PAI, and neurological features may not be evident for up to 2 years.[59]

Adrenal Imaging

Imaging is not essential for diagnosis of autoimmune AI, and an atrophic gland may be difficult to visualize. In patients with negative 21-hydroxylase antibodies or where clinical suspicion of a nonautoimmune etiology such as hemorrhage, infection, infiltration, or neoplastic disease exists, abdominal CT should be performed.

KEY POINTS

- Cortisol deficiency is the hallmark of adrenal insufficiency.
- Adrenocorticotropic hormone (ACTH) stimulation testing is the primary diagnostic test for cortisol deficiency and is more reliable than random serum cortisol.
- Elevated serum ACTH concentration establishes primary adrenal insufficiency and excludes pituitary/hypothalamic etiology.
- Measurement of serum aldosterone and renin concentrations determines if adrenal aldosterone secretion is intact.
- 21-hydroxylase antibodies are a useful adjunct to determine an autoimmune etiology in primary adrenal insufficiency. Elevated serum 17-hydroxyprogesterone concentrations are a key diagnostic feature of 21-hydroxylase deficiency.
- Imaging of the adrenals is needed in cases of primary adrenal insufficiency where a nonautoimmune etiology is considered likely, and imaging of the pituitary/hypothalamus is generally required in secondary adrenal insufficiency.

TREATMENT OF ADRENAL INSUFFICIENCY

Treatment for AI consists of replacement of the missing steroid hormones: cortisol in SAI and cortisol and aldosterone in PAI. The glucocorticoids employed include naturally occurring hydrocortisone (cortisol) and cortisone acetate and the synthetic steroids prednisolone and the less frequently used dexamethasone, although the latter is not recommended, as patients are easily overtreated, leading to Cushingoid side effects. These glucocorticoids differ in potency, with equivalent replacement daily doses and serum disappearance half-lives ranging from 15 to 25 mg and 90 min for hydrocortisone, 20 to 30 mg and 90 min for cortisone acetate, 3 to 5 mg and 120 min for prednisolone, and 0.3 to 0.5 mg and 190 min for dexamethasone. All have mineralocorticoid activity except dexamethasone. For example, hydrocortisone has 1/400th the mineralocorticoid potency of fludrocortisone by virtue of the 9α fluorine atom that confers resistance to metabolism by 11β-hydroxylase type 2, so that 40 mg hydrocortisone has the mineralocorticoid activity of 0.1 mg fludrocortisone.

Hydrocortisone is currently the main glucocorticoid recommended in AI. In previous decades relatively high doses of hydrocortisone and other glucocorticoids were used, based on erroneous estimates of cortisol production rates. Subsequent studies estimated lower daily cortisol production rates in adults and children at approximately 5.7 to 6.8 mg/m^2/day, or approximately 10 mg per day, equating to oral hydrocortisone dosing of 15 to 25 mg/day, or prednisolone 3 to 5 mg/day because of estimated pharmacokinetics and metabolism.[21]

Retrospective studies showed lower bone mineral density, greater body weight, and higher blood pressure in glucocorticoid-treated patients,[60] with a negative effect of glucocorticoid dose, although factors other than dose may have contributed. A limitation of retrospective studies in this area is that the indication(s) for use of higher hydrocortisone equivalent doses, generally above 30 mg daily, is unclear.

Many dosing schedules for hydrocortisone have been used. A double-blind crossover study of two-dose versus four-dose daily hydrocortisone showed a more physiological serum diurnal cortisol pattern with four doses per day and a patient preference for that regimen.[61] Systematic analyses designed to obviate such biases have not shown any effect of glucocorticoid dosing schedule on well-being and found that bone density data are inconsistent. Dose regimens vary depending on the relative priorities of pursuing diurnal cortisol patterns—requiring more frequent dosing in an effort to improve well-being, or less frequent dosing to improve medication adherence and convenience. One common regimen may be hydrocortisone 10 mg on rising, 5 mg at lunchtime, and 5 mg in the early evening. The final dose regimen is then modulated according to clinical response and patient preference. Observation for side effects includes growth velocity monitoring in children and measurements of weight and waist circumference and examination for features of incipient Cushing syndrome in adults, as well as resolution of clinical features of AI. No objective measures are available to assist in glucocorticoid dose adjustment—the use of ACTH is unreliable, as the disturbance in cortisol/ACTH feedback in PAI leads to elevated ACTH in apparently well-treated PAI patients, and ACTH is absent or unrelated to glucocorticoid dose in SAI. Pregnancy is associated with elevated cortisol secretion and increases in total and free cortisol in serum, reaching approximately 2.5-fold normal in late third trimester, driven by placental CRH.[62] No studies guide glucocorticoid dose adjustment, but an increased dose of 20% to 40% is suggested after 24 weeks.[21] Use of dexamethasone is not recommended, as it is not metabolized in the placenta, so fetal glucocorticoid exposure may be excessive.

Hydrocortisone metabolism is highly variable, such that individuals taking 0.2 to 0.3 mg/kg/day have cortisol levels that substantially overlap those of patients taking 0.4 to 0.6 mg/kg/day.[63] In associated analyses, the higher hydrocortisone dose was associated with better quality of life,[64] higher blood pressure (BP) (by 5/2 mm Hg), and no negative effects on cognition.[65] Other studies in PAI and CAH have suggested that lower doses of short-acting glucocorticoids may be associated with an increased risk of adrenal crises.[66,67] The use of prednisolone with its longer serum half-life compared with hydrocortisone (150 vs. 90 minutes) may make once-daily dosing more feasible if required.

Plenadren contains immediate- and delayed-release (dual-release) hydrocortisone given on awakening. Cortisol levels are quite physiological, but there is no early morning prewakening cortisol rise as seen in normal physiology. A 12-week open trial showed weight reduction and improved glucose tolerance, blood pressure, and quality of life. A further single-blind study that took place over 24 weeks showed a lower weight difference (4 kg) between the groups receiving Plenadren and rapid-release hydrocortisone, with 1.9-kg gain in the rapid-release group. However, glucocorticoid exposure may have been lower in the Plenadren group despite the same dose, because of altered pharmacokinetics.[68] Chronocort is a twice-daily hydrocortisone tablet designed to be taken in the morning and at bedtime where favorable circadian cortisol concentrations are apparent—there may be advantages in control of androgens in CAH, but these have not been demonstrated in a larger multicenter trial.[69] Longer-acting glucocorticoids (plasma half-life 150 mins prednisolone vs. 90 mins for hydrocortisone) are more convenient than hydrocortisone, but at present there are no reported studies of the relative benefits of prednisolone. The provision of circadian cortisol concentrations in a double-blind study of subcutaneously infused hydrocortisone showed no effect on well-being,[70] although a positive effect was seen in an open crossover study.[71] To date, no double-blind study of delayed or dual-release tablets has been performed to address the well-being question definitively.

The only preparation available to specifically replace aldosterone is fludrocortisone, at a once-daily maintenance oral dose of 50 to 300 μg/day. A typical starting dose in adults and children is 100 μg daily. Response is monitored in terms of serum sodium, potassium, and renin levels, taken in the seated position with reference to blood pressure. Peripheral edema is a common problem on instituting fludrocortisone in new patients with PAI and may make low doses necessary for a period of time. No parenteral preparation of fludrocortisone is available.

KEY POINTS

- Adrenal hormone replacement consists of glucocorticoid tablets, taken at a dose and frequency of administration consistent with physiological cortisol secretion, with reference to the potency and elimination half-life/tissue effect of the steroid employed.
- Typically, hydrocortisone doses are 15 to 25 mg per day, or prednisolone 4 to 6 mg per day, but some patients do require higher doses, as glucocorticoid metabolism and tissue sensitivity display considerable interindividual variability. Dose adjustment requires careful clinical assessment and judgment, as no pathology tests are helpful.
- Mineralocorticoid replacement is often needed in primary adrenal insufficiency and involves fludrocortisone, typically 0.1 to 0.3 mg. Dose adjustment primarily involves control of symptoms of aldosterone deficiency and normalization of serum potassium and renin.

Replacement of Dehydroepiandrosterone

Given that dehydroepiandrosterone (DHEA) and its sulfated entity DHEA-S are the most abundant steroids secreted by the adrenal, attention has focused on the value of replacement of DHEA, which is generally present in very low concentrations in AI patients relative to healthy individuals. The role of DHEA in pregnancy, where secretion by the fetal adrenal, providing substrate for placental estriol production, is well-established, as is as its role in producing axillary and pubic hair in adolescents, a phenomenon known as adrenarche. It cannot be said that DHEA has other roles mandating replacement in adults, although it assists in maintaining terminal hair in the pubic and axillary areas, where the skin has high androgen sensitivity, in women. In AI, DHEA has been employed to help combat the known reduced quality of life compared with normal individuals despite adequate glucocorticoid and mineralocorticoid replacement. Several short-term trials in PAI and SAI have shown that DHEA replacement potentially improves mood and well-being, but beneficial effects have not been reported by all. A metaanalysis of these trials in women with PAI or SAI showed small (perhaps trivial) effects on quality of life and depression, but no effect on anxiety or sexual well-being, with the evidence appearing insufficient to support routine use of DHEA.[72] There are no data regarding the long-term safety of DHEA administration in AI. DHEA can be trialed for patients whose well-being is severely impaired despite optimal glucocorticoid and mineralocorticoid replacement. A daily dose of 10 to 25 mg DHEA taken in the morning with monitoring of therapy with serum DHEA to target the middle of the normal range can be trialed.

KEY POINTS

- Although circulating dehydroepiandrosterone (DHEA[-S]) concentrations are often low in adrenal insufficiency, replacement of DHEA is not routinely required, as consistent benefits have not been demonstrated in short-term trials.
- DHEA treatment may be considered in selected patients and should be delivered as a trial, monitoring clinical benefits and side effects (if any) while aiming for normal serum DHEA-S concentrations.

Quality of Life in Adrenal Insufficiency

A substantial minority of patients with PAI complain of persistent poor quality of life with reduced work capacity and increased use of social welfare, as systematically verified in Scandinavian studies.[60] It is not known why patients with AI have reduced quality of life despite appropriate adrenal hormone replacement, and it is likely that individual factors are important. Possibilities include periods of inadequate hydrocortisone exposure, low glucose concentrations at time of cortisol deficiency, lack of DHEA, and synergy from associated hormone deficits in autoimmune polyglandular syndromes. Patients often find glucocorticoid replacement regimens a burden, to an extent that seems to correlate with their illness severity perception.[20]

Adrenal Suppression

Management of adrenal suppression involves a graded withdrawal of exogenous glucocorticoid so that the endogenous HPA axis function can recover, a process that may take 6 to 18 months. Adrenal suppression is often partial. Patients taking 5 to 10 mg prednisolone daily comprise approximately one-third with unsuppressed cortisol response to ACTH 1–24, one-third with complete suppression of cortisol response to ACTH, and one-third partially suppressed.

Weaning of glucocorticoids requires three considerations:

1. The possibility of recrudescence of the underlying disorder, most often inflammatory, that prompted use of chronic supraphysiologic glucocorticoids.
2. The appearance of glucocorticoid withdrawal syndrome, a symptom complex that also occurs with withdrawal of other hormones. Symptoms often overlap with those of AI, despite a glucocorticoid dose exceeding the physiological range.
3. True AI developing when the glucocorticoid (GC) dose is below the physiological replacement range.

In general, as long as there is no disease recrudescence, glucocorticoid doses can be weaned quite quickly above the physiological range, but need to be withdrawn at a much slower pace when withdrawing within the physiological range. For example, with prednisolone, doses can be reduced by as much as 20% per week above 5 mg daily. The rate of glucocorticoid dose reduction needs to be individualized, with AI symptoms and morning serum cortisol, before morning prednisolone dosing, acting as arbiters of the rate of prednisolone withdrawal. In general, morning cortisol levels above 400 nmol/L suggest that rapid withdrawal or cessation of prednisolone is indicated; cortisol below 200 nmol/L suggests that withdrawal may need to be stopped until cortisol levels rise; and levels of 200 to 400 nmol/L suggest that slow withdrawal may be continued as long as the patient is asymptomatic. Recovery can also be gauged by the response of plasma cortisol to an IV injection of cosyntropin, with recovery being indicated by a plasma cortisol level greater than 18 μg/dL (500 nmol/L) at intervals based on clinical judgment when there is a need to verify adrenal secretory status. Stress dosing of glucocorticoids may be necessary for up to 1 year after glucocorticoid withdrawal to avoid adrenal crisis.

KEY POINTS

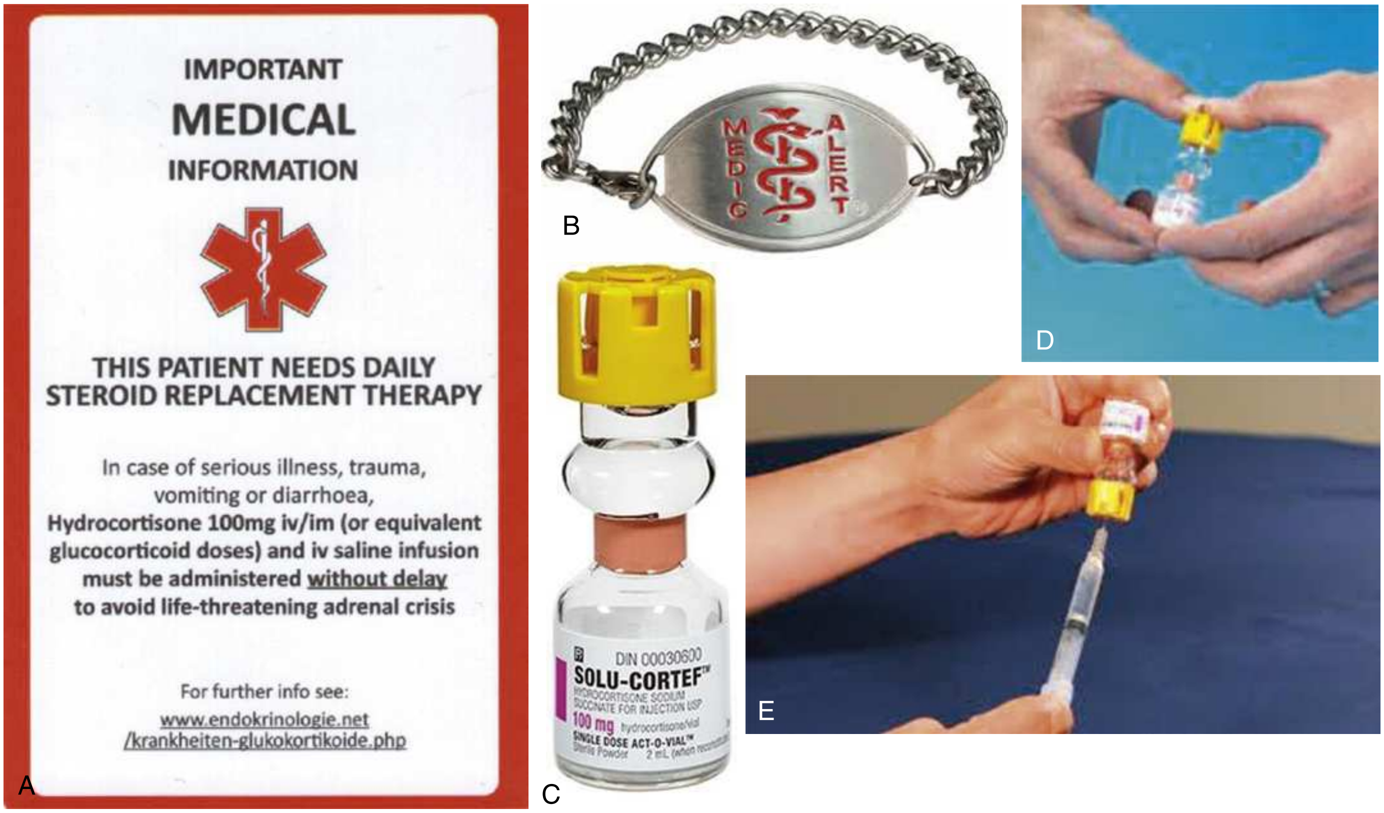
- Quality-of-life impairments are common in treated adrenal insufficiency; there are many possible causes. Optimization of adrenal hormone replacement and consideration of extraadrenal causes are necessary first steps in management. Despite standard assiduous management, there remain a proportion of patients with poor quality of life—an area where systematic study is needed to help provide new therapeutic approaches.
- Adrenal suppression resulting from the therapeutic use of supraphysiologic glucocorticoid doses for inflammatory and other disorders is unique in its reversibility, requiring gradual stepwise withdrawal of glucocorticoid replacement while observing progress of the underlying disease, Cushingoid features, and the presence of relative or absolute adrenal insufficiency. Morning cortisol and/or adrenocorticotropic hormone stimulation may help guide the rapidity of glucocorticoid withdrawal.

ADRENAL CRISIS

Adrenal crises may be defined as an acute deterioration in health status associated with hypotension—absolute hypotension (systolic blood pressure <100 mm Hg in an adult) or relative (systolic blood pressure at least 20 mm Hg lower than usual)—with features that resolve within 1 to 2 hours of parenteral glucocorticoid administration (i.e., a marked resolution of hypotension within an hour and improvement of clinical symptoms over 2 hours).[20] Approximately 6% to 8% of patients with AI will have an single incident of adrenal crisis per year, and rates appear to be rising with lower-dose, short-acting glucocorticoid regimens.[20] Adrenal crises are more common in patients with PAI than in those with SAI, probably owing to partial preservation of cortisol secretion in some patients with SAI and the absence of mineralocorticoid secretion in PAI. Some AI patients experience multiple episodes, while others have few or no episodes.[73] Susceptibility to adrenal crises varies among patients with hypoadrenalism, depending on a number of factors, including age, prior adrenal crises, the presence of APS-I, diabetes mellitus, and nonendocrine comorbidities, such as asthma and cardiac disease, which increase risk.[74] Adrenal crisis mortality may be as high as 6% per event.[73] Adrenal crisis has been reported in patients without an established diagnosis of AI, although typical symptoms may have been present in retrospect.[67]

Treatment of Adrenal Crisis

Treatment includes administration of intravenous hydrocortisone, 100 mg IV, followed by 200 mg/24 hours as a continuous infusion or as IV (or IM) hydrocortisone 50 mg every 6 hours, with subsequent does tailored to clinical response. Hydrocortisone can be substituted with methylprednisolone 40 mg/24 hours or dexamethasone 4 mg every 24 hours if necessary—use of dexamethasone would make it necessary to give fludrocortisone, which is unnecessary for high doses of hydrocortisone or



IMPORTANT
MEDICAL
INFORMATION

THIS PATIENT NEEDS DAILY
STEROID REPLACEMENT THERAPY

In case of serious illness, trauma,
vomiting or diarrhoea,
**Hydrocortisone 100mg iv/im (or equivalent
glucocorticoid doses) and iv saline infusion
must be administered without delay
to avoid life-threatening adrenal crisis**

For further info see:
www.endokrinologie.net
/krankheiten-glukokortikoide.php

Fig. 93.4 A–E, Strategies used to prevent adrenal crises. Initially, use of stress dosing in illness is crucial. For patients unable to take glucocorticoid tablets, injections of hydrocortisone, most conveniently stored and delivered using the Solucortef Actovial device are very useful. For patients unable to communicate, the use of medical jewelry such as MedicAlert and carrying a card with details regarding the imperative for parenteral hydrocortisone may also be life saving.

prednisolone, as these steroids have substantial mineralocorticoid activity. Replacement of fluid with normal saline is needed, according to standard resuscitation guidelines; typically volumes of 2 to 3 liters are required, but this may be higher in cases with fluid loss related to the adrenal crisis precipitant, such as gastroenteritis. Rapid clinical improvement is expected, often within 1 hour; if such an improvement is not seen, especially with respect to blood pressure, another cause of shock should be sought.[21]

Prevention of Adrenal Crises

The effectiveness of preventive measures, largely patient-initiated, is incomplete, with intensive education not apparently affecting adrenal crisis risk in one prospective study.[73] Patients with PAI, chronic symptoms suggestive of AI perhaps reflecting underreplacement, poor well-being, or previous episodes of adrenal crises are more likely to develop adrenal crises.[75] The three standard measures are:

1. “Stress dosing” with intercurrent illness, especially infections with fever, where the replacement dose is increased 3-fold for 3 days (“3 × 3 rule”) or until the illness subsides.
2. Use of injectable hydrocortisone (Solucortef Actovial 100 mg) (Fig. 93.4), which can be delivered IM (or subcutaneously if needed)[76] in unwell patients by the patient or relatives when unable to take or reliably absorb tablets, for example in gastroenteritis.
3. The provision of a communicative device to inform emergency medical staff of the need for parenteral hydrocortisone when the patients is unable to effectively communicate this. Such devices include MedicAlert jewelry and/or emergency measures in the form of a “steroid card” carried by the patient. A European Emergency card has been introduced that is translated into numerous languages (see Fig. 93.4).

Studies have shown that subcutaneous hydrocortisone is well-absorbed.[77] Although studies of subcutaneous hydrocortisone in adrenal crisis (AC) prevention in actual or incipient crisis have not been performed; such administration is surely superior to no parenteral dosing at all, recognizing that many patients have difficulty with self-administration of IM injections.

Crucial in adrenal crisis prevention is the management of perioperative patients. In general, AI patients all require ongoing glucocorticoids with surgery, and, as surgical stress is associated with a physiological rise in cortisol, it is customary to recommend parenteral (or sometimes oral) glucocorticoids in a dose commensurate with the extent of the surgery. Many schema have been produced and are approved on a consensus basis, and examples are freely available, such as: www.addisons.org.uk/surgery (Table 93.4).

KEY POINTS

- Adrenal crises are life-threatening medical emergencies.
- Prevention is possible using a combination of patient education, patient-driven stress dosing of glucocorticoid, home parenteral hydrocortisone, and use of MedicAlert jewelry/patient information cards to advise clinicians to promptly administer hydrocortisone in patients unable to do so themselves.
- Prompt delivery of parenteral hydrocortisone is needed on patient presentation at medical facilities to avert worsening of the clinical features of incipient or actual adrenal crisis.
- Treatment of adrenal crises involves parenteral glucocorticoids and standard fluid resuscitation, as well as addressing any evident medical precipitant.

TABLE 93.4 **Dosing Strategies for the Perioperative Management of Patients With Adrenal Insufficiency**

	Steroid-Dependent Patient Requires Continuous/Parenteral Steroid Cover	**See Surgical Guidelines: www.addisons.org.uk/surgery**
TYPE OF PROCEDURE	***PREOPERATIVE AND OPERATIVE NEEDS (SEE NOTES 1, 2)***	***POSTOPERATIVE NEEDS (SEE NOTES 6, 8, 9)***
Lengthy, major surgery with long recovery time, e.g., open heart surgery, major bowel surgery	100 mg hydrocortisone IM or IV just before anesthesia (see Notes 2, 3, 7), immediately followed by: • 100 mg IM or IV 6 hourly or • continuous infusion 200 mg/24 hours	100 mg IM or IV every 6 hours or continuous IV infusion 200 mg/24 hours (see Notes 3, 5) or until able to eat and drink normally (discharged from ITU). If well, then double oral dose for 48+ hours. Then taper the return to normal dose.
Major surgery with rapid recovery, e.g., caesarean section, joint replacement	100 mg hydrocortisone IM or IV just before anesthesia (see Notes 2, 6, 7), immediately followed by: • 100 mg IM or IV 6 hourly or • continuous infusion 200 mg/24 hours	100 mg IM or IV or continuous infusion 200 mg/24 hours for 24–48 hours (see Notes 3, 5) for 24–48 hours (or until able to eat and drink normally). If well, then double oral dose for 24–48 hours. Then return to normal dose.
Labor and vaginal birth	100 mg hydrocortisone IM or IV at onset of active labor (see Notes 4–7), immediately followed by continuous IV infusion 200 mg/24 hours or 100 mg IM or IV 6 hourly until delivery	Double oral dose for 24–48 hours after delivery. If well, then return to normal dose.
Minor surgery, e.g., cataract surgery, hernia repairs, laparoscopy with local anesthetic	100 mg hydrocortisone IM just before anesthesia (see Note 6)	Double oral dose for 24 hours. Then return to normal dose.
Minor procedure, e.g., skin mole removal with local anesthetic	Take an extra oral dose 60 minutes ahead of the procedure	An extra dose 60 minutes after the procedure. Then return to normal dose.
Invasive bowel procedures requiring laxatives, e.g., colonoscopy, barium enema	Hospital admission overnight with IV fluids and 100 mg hydrocortisone: • IM during preparation (see Notes 3, 5, 6) • 100 mg hydrocortisone IM at commencement (see Notes 1, 6)	Double dose oral medication for 24 hours. Then return to normal dose.
Other invasive procedures, e.g., endoscopy, gastroscopy	100 mg hydrocortisone IM just before commencing	Double-dose oral medication for 24 hours. Then return to normal dose.
Major dental surgery, e.g., dental extraction(s) with local or general anesthetic	100 mg hydrocortisone IM just before anesthesia (see Notes 6, 7, 8)	Double-dose oral medication for 24 hours. Then return to normal dose.
Dental surgery, e.g., root canal work with local anesthetic	Double oral dose (up to 20 mg hydrocortisone) 1 hour before surgery	Double-dose oral medication for 24 hours. Then return to normal dose.
Minor dental procedure, e.g., replace filling, scale and polish	Take an extra oral dose. 60 minutes ahead of the procedure	An extra dose where hypoadrenal symptoms occur afterwards. Then return to normal dose.

NOTES

1. Give the steroid-dependent patient first-on-the list status (alongside insulin-dependent diabetes) to minimize the risks of dehydration.
2. For any nil-by-mouth regimen, arrange an IV saline infusion (0.9% saline or equivalent) to prevent dehydration and maintain mineralcorticoid stability, e.g., 1000 mL every 8 hours if >50 kg.
3. Continuous IV hydrocortisone infusion is preferable to 6 hourly IM or IV injection, as it gives more stable cover. Arrange as 100-mg bolus followed by 8.33 mg per hour or 200 mg per 24 hours.
4. Active labor is cervical dilation >4 cm.
5. Arrange continuous IV infusion cover for steroid-dependent patients taking CYP3A4 accelerants, e.g., anticonvulsants, rifampicin, and antifungal drugs, to minimize the risk of decompensation.
6. IM hydrocortisone is preferable to IV injection for its more sustained duration.
7. Administer bolus hydrocortisone over a minimum of 10 minutes to prevent vascular damage.
8. Hydrocortisone acetate cannot be used because of its slow-release, microcrystalline formulation. Ensure parenteral drug is hydrocortisone sodium phosphate or hydrocortisone sodium succinate, 100 mg.
9. Monitor electrolytes and blood pressure postoperatively for all procedures requiring parenteral steroid cover. If the patient becomes hypotensive, drowsy, or peripherally shut down, administer 100 mg hydrocortisone IV or IM bolus immediately.
10. If any postoperative complications arise, e.g., fever, delay the return to normal dose.
11. Ensure back-up supplies of oral and injectable hydrocortisone are available for resuscitation before commencing surgery. Even at full steroid cover, postoperative resuscitation may occasionally be required.
12. A preassessment meeting with the anesthetist is advisable for all steroid-dependent patients to ensure any comorbidities and potential drug interactions are taken into account.
13. Patients who have been taking 5 mg prednisolone or more long term should be regarded as potentially suppressed and managed with perioperative supplemental steroid cover, on a precautionary basis.

IM, Intramuscular*; IV,* intravenous*; ITU,* Intensive therapy unit.

The Need for Novel Treatments for Adrenal Insufficiency

The major problems with current adrenal hormone replacement therapy are the poor well-being in a substantial minority of patients and the ever-present risk of adrenal crisis. At present no test accurately assesses the adequacy of glucocorticoid replacement. A test that assessed tissue glucocorticoid sufficiency would be a significant advance and would allow individualized glucocorticoid dose optimization. Some forms of AI have specific difficulties with current therapy. For example, CAH resulting from 21-hydroxylase deficiency has the dilemma that replacement-level glucocorticoids fail to suppress androgen production, and higher doses may control androgens but have Cushingoid side effects (see Chapter 94).

In principle, CAH patients may achieve restoration of adrenocortical function using gene transfer to obviate enzyme deficiencies, as described in rodent models.[78] For patients with autoimmune-based Addison disease, attempts have been made to determine if remnant adrenocortical cells from the autoimmune process, analogous to surviving β cells in type 1 diabetes, may respond to regenerative approaches. Adrenocococrtical stem cells, which have a subcapsular location, proliferate and differentiate into the adrenal steroid secretory cells under the influence of ACTH.[79]

In one study of 13 PAI patients who received 20 weeks of depot ACTH 1–24, two patients achieved stimulated serum cortisol concentrations of greater than 400nmol/L: one was independent of exogenous glucocorticoid for 28 weeks, and one for 28 months (end of reported observation).[80] A further study used the B-lymphocyte–depleting agent rituximab and 12 weeks of depot ACTH 1–24 in 13 PAI patients. Ten of these patients had detectable cortisol at study entry (peak 26–265 nmol/L), and 54% had higher cortisol responses, but no patient achieved a normal response to ACTH.[81] Although these studies do not point to regenerative therapies being clinically viable at present, they suggest that a form of regenerative therapy with immune modulation may control Addison disease in the future.

Finally, PAI may respond to cell-based replacement therapies designed to replace adrenocortical cells. Most PAI cases are autoimmune; hence, such therapies would be most applicable in already immunosuppressed individuals such as those who have had a renal transplant[82-84] or if the adrenal cell allograft was provided in an encapsulated form where the immune system can be excluded, obviating the need for immunosuppression—this has been done in animal models, such as a xenograft model in rats.[85]

Acknowledgments

We thank Prof. Stefan Bornstein and Dr. Andreas Barthel for writing the previous version of this chapter, which we have updated.

REFERENCES

1. Eustachi B, Fallopio G. Opuscula anatomica. Venice: M.A. Ulmus; 1563.
2. Brown-Sequard CE. [Recherches experimentales sur la physiologie et la pathologie des capsules surrenales.]. *Acad Sci Paris*. 1856;43:422–425.
3. Trousseau A. Bronze Addison's disease. *Arch Gen Med*. 1856;8:478.
4. Osler W. Case of Addison's disease: death during treatment with the supravital extra. *Bull Johns Hopkins Hosp*. 1896;7:208–209.
5. Wintersteiner OP, Pfiffner J. Chemical studies on the adrenal cortex: II. *J Biol Chem*. 1935;111:599–612.
6. Kendall EC. A chemical and physiological investigation of the suprarenal cortex. *Symp Quant Biol*. 1937;5:299.
7. Hench PS, Slocumb CH, Barnes AR, et al. The effects of the adrenal cortical hormone, compound E, on the acute phase of rheumatic fever: a preliminary report. *Proc Mayo Clin*. 1949;24:277–297.
8. Thorn GW, Forsham PH. The treatment of adrenal insufficiency. *Rec Prog Horm Res*. 1949;4:229.
9. Harris GW. Neural control of the pituitary gland. *Physiol Rev*. 1948;28:139.
10. Li CH, Dixon JS, Chung D. The structure of bovine corticotropin. *J Am Chem Soc*. 1958;80. 2787–2596.
11. Vale W, Spiers J, Rivier C, et al. Characterization of a 41 residue ovine hypothalamic peptide that stimulates secretion of corticotropin and beta-endorphin. *Science*. 1981;213:585–587.
12. Lu J, Li L, Lan Y, Liang Y, et al. Immune checkpoint inhibitor adrenal dysfunction: a systematic review and meta-analysis. *Cancer Med*. 2019;8:7503–7515.
13. Li T, Cunningham JL, Gilliam WP, et al. Prevalence of opioid-induced adrenal insufficiency in patients taking chronic opioids. *J Clin Endocrinol Metab*. 2020;105:dgaa499.
14. Mirza FS, Luthra P, Chirch L. Endocrinological aspects of HIV infection. *J Endocrinol Invest*. 2018;41:881–899.
15. Annane D, Pastores SM, Arlt W, et al. Critical illness-related corticosteroid insufficiency: a narrative review from a multispeciality task force of the society of critical care medicine (SCCM) and the European society of intensive care medicine (ESICM). *Crit Care Med*. 2017;45:2089–2098.
16. Smans LC, Van der Valk ES, Hermus AR, et al. Incidence of adrenal crisis in patients with adrenal insufficiency. *Clin Endocrinol (Oxf)*. 2016;84:17–22.
17. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. *Lancet*. 2014;383:2152–2167.
18. Merke DP, Chrousos GP, Eisenhofer G, et al. Adrenomedullary dysplasia and hypofunction in patients with classic 21-hydroxylase deficiency. *N Engl J Med*. 2000;343:1362–1368.
19. Betterle C, Garelli S, Presotto F, et al. From appearance of adrenal autoantibodies to clinical symptoms of Addison's disease: natural history. *Front Horm Res*. 2016;46:133–145.
20. Rushworth RL, Torpy DJ, Falhammar H. Adrenal crisis. *NEJM*. 2019;381:852–861.
21. Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:364–389.
22. Hellesen A, Bratland E, Husebye ES. Autoimmune Addison's disease: an update on pathogenesis. *Ann Endocrinol (Paris)*. 2018;79:157–163.
23. Paolo Jr WF, Nosanchuk JD. Adrenal infections. *Int J Infect Dis*. 2006;10:343–353.
24. Perheentupa J. APS1/APECED: the clinical disease and therapy. *Endocrinol Metab Clinic North Am*. 2002;31:295–320.
25. Husebye ES, Anderson MS, Kämpe O. Autoimmune polyendocrine syndromes. *N Engl J Med*. 2018;378:2543–2544.
26. Watt I, Ledingham IM. Mortality amongst multiple trauma patients admitted to an intensive therapy unit. *Anaesthesia*. 1984;39:973–981.
27. Erdkamp FL, Gams RO, Hoorntje SJ. Endocrine organ failure due to systemic AA-amyloidosis. *Neth J Med*. 1991;38:24–28.
28. New MI. Diagnosis and management of congenital adrenal hyperplasia. *Annu Rev Med*. 1998;49:311–328.
29. Lin D, Sugawara T, Strauss JF, et al. Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science*. 1995;267:1828–1831.
30. Laureti S, Casucci G, Santeusanio F, et al. X-linked adrenoleukodystrophy is a frequent cause of idiopathic Addison disease in young adult male patients. *J Clin Endocrinol Metab*. 1996;81:470–474.
31. Maharaj A, Maudhoo A, Chan LF, et al. Isolated glucocorticoid deficiency: genetic causes and animal models. *Steroid Biochem Mol Biol*. 2019;189:73–80.
32. Phelan JK, McCabe ER. Mutations in NR0B1 (DAX1) and NR5A1 (SF1) responsible for adrenal hypoplasia congenital. *Hum Mutat*. 2001;18:472–487.
33. Reutens AT, Achermann JC, Ito M, et al. Clinical and functional effects of mutations in DAX-1 gene in patients with adrenal hypoplasia congenital. *J Clin Endocrinol Metab*. 1999;84:504–511.
34. Woods CP, Argese N, Chapman M, et al. Adrenal suppression in patients taking inhaled glucocorticoids is highly prevalent and management can be guided by morning cortisol. *Eur J Endocrinol*. 2015;173:633–642.
35. Salem M, Tainsh Jr RE, Bromberg J, et al. Perioperative glucocorticoid coverage. A reassessment of 42 years after emergence of a problem. *Ann Surg*. 1994;219:416–425.

36. Rushworth RL, Chrisp GL, Torpy DJ. Glucocorticoid-induced adrenal insufficiency: a study of the incidence in hospital patients and a review of peri-operative management. *Endocrine Practice*. 2018;24:437–445.
37. Torpy DJ, Grice JE, Hockings GI, et al. Alprazolam blocks the naloxone-stimulated hypothalamo-pituitary-adrenal axis in man. *J Clin Endocrinol Metab*. 1993;76:388–391.
38. Donegan D. Opioid induced adrenal insufficiency: what is new? *Curr Opin Endocrinol Diabetes Obes*. 2019;26:133–138.
39. Erfurth E-V. Diagnosis, background and treatment of hypothalamic damage in craniopharyngiona. *Neuroendocrinology*. 2020;110:767–779.
40. Angelousi A, Alexandraki K, Tsoli M, et al. Hypophysitis (including IgG4 and immunotherapy). *Neuroendocrinology*. 2020;110:822–835.
41. Gilis-Januszewska A, Kluczyński Ł, Hubalewska-Dydejczyk A. Traumatic brain injuries induced pituitary dysfunction: a call for algorithms. *Endocr Connect*. 2020;9:R112–R123.
42. Xatzipsalti M, Voutetakis A, Stamoyannou L, et al. Congenital hypopituitarism: various genes, various phenotypes. *Horm Metab Res*. 2019;51:81–90.
43. Bellissant E, Annane D. Effect of hydrocortisone on phenylephrine-mean arterial pressure dose-response relationship in septic shock. *Clin Pharmacol Ther*. 2000;68:293–303.
44. Oppert M, Schindler R, Husung C, et al. Low dose hydrocortisone improves shock reversal and reduces cytokine levels in early hyperdynamic septic shock. *Crit Care Med*. 2005;33:2457–2464.
45. Annane D, Renault C, Brun-Buisson B, et al. Hydrocortisone plus fludrocortisone for adults with septic shock. *N Engl J Med*. 2018;378:809–818.
46. Annane D, Sebille V, Charpentier C, et al. Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA*. 2002;283:1038–1045.
47. Venkatesh B, Finfer S, Cohen J, et al. ADRENAL trial investigators and the Australian-New Zealand Intensive Care Society Trials Group: adjunctive glucocorticoid therapy in patients with septic shock. *N Engl J Med*. 2018;378:797–808.
48. Annane D, Pastores SM, Rochwerg B, et al. Guidelines for the diagnosis and management of critical illness related corticosteroid insufficiency (CIRCI) in critically ill patients (Part 1): Society of critical care medicine (SCCM) and European Society of Intensive Care Medicine (ESICM) 2017. *Crit Care Med*. 2017;45:2078–2088.
49. Annane D, Sebille V, Charpentier C, et al. Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA*. 2002;288:862–871.
50. Keh D, Boehnke T, Weber-Cartens S, et al. Immunologic and hemodynamic effects of "low-dose" hydrocortisone in septic shock: a double-blind, randomized, placebo-controlled, cross-over study. *Am J Respir Crit Care Med*. 2003;167:512–520.
51. Yo WS, Toh I-M, Brown S, et al. How good is a morning cortisol in predicting an adequate response to intramuscular synacthen stimulation? *Clin Endocrinol (Oxf)*. 2014;81:19–24.
52. Brown S, Hadlow N, Badshaw I, et al. A time-adjusted cortisol cut-off can reduce referral rate for synacthen stimulation test whilst maintaining diagnostic performance. *Clin Endocrinol (Oxf)*. 2017;87:418–424.
53. Giordano R, Picu A, Bonelli L, et al. Hypothalamus-pituitary-adrenal axis evaluation in patients with hypothalamo-pituitary disorders: comparison of different provocative tests. *Clin Endocrinol (Oxf)*. 2008;68:935–941.
54. El-Farhan N, Picket A, Ducroq D, et al. Method-specific serum cortisol responses to the adrenocorticotrophin test: comparison of gas chromatography-mass spectrometry and five automated immunoassays. *Clin Endocrinol (Oxf)*. 2013;78:673–680.
55. Mukherjee JJ, Jacome de Castro J, Kaltsas G, et al. A comparison of insulin tolerance/glucagon tests with the short ACTH stimulation test for assessment of hypothalamo-pituitary-adrenal axis in the early post-operative period after hypophysectomy. *Clin Endocrinol (Oxf)*. 1997;47:51–60.
56. Ospina NS, Nofal AA, Bancos I, et al. ACTH stimulation tests for the diagnosis of adrenal insufficiency; systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2016;101:427–434.
57. English K, Inder WJ, Weedon Z, et al. Prospective evaluation of a week one overnight metyrapone test with subsequent dynamic assessments of hypothalamic-pituitary-adrenal axis function after pituitary surgery. *Clin Endocrinol (Oxf)*. 2017;87:312.
58. Streeten DH, Anderson Jr GH, Dalakos TG, et al. Normal and abnormal function of the hypothalamic-pituitary-adrenocortical system in man. *Endocr Rev*. 1984;5:371–394.
59. Dubey P, Raymond GV, Moser AB, et al. Adrenal insufficiency in asymptomatic adrenoleukodystrophy patients identified by very long-chain fatty acid screening. *J Pediatr*. 2005;146:528–532.
60. Lovas K, Loge JH, Husebye ES. Subjective health status in Norwegian patients with Addison's disease. *Clin Endocrinol (Oxf)*. 2002;56:581–588.
61. Ekman B, Bachrach-Lindstrom M, Lindstrom T, et al. A randomized double-blind, crossover study comparing two and four-dose hydrocortisone regimen with regard to quality of life, cortisol and ACTH profiles in patients with primary adrenal insufficiency. *Clin Endocrinol (Oxf)*. 2012;77:18–25.
62. Jung C, Ho JT, Torpy DJ, et al. A longitudinal study of plasma and urinary cortisol in pregnancy and postpartum. *J Clin Endocrinol Matab*. 2011;96:1533–1540.
63. Werumeus Buning J, Touw DJ, Brummelman P, et al. Pharmacokinetics of oral hydrocortisone - Results and implications from a randomized controlled trial. *Metabolism: Clin Experimental*. 2017;71:7–16.
64. Werumeus Buning J, Brummelman P, Koerts J, et al. Hydrocortisone dose influences pain, depressive symptoms and perceived health in adrenal insufficiency: a randomized controlled trial. *Neuroendocrinology*. 2016;103:771–778.
65. Werumeus Buning J, Brummelman P, Koerts J, et al. The effects of two different doses of hydrocortisone on cognition in patients with secondary adrenal insufficiency-results from a randomized controlled trial. *Psychoneuroendocrinology*. 2015;55:36–47.
66. El-Maouche D, Hargreaves CJ, Sinaii N, et al. Longitudinal assessment of illnesses, stress dosing, and illness sequelae in patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metabol*. 2018;103:2336–2345.
67. Rushworth RL, Torpy DJ. Modern hydrocortisone replacement regimens in adrenal insufficiency patients and the risk of adrenal crisis. *Hormone Metabol Res*. 2015;47:637–642.
68. Isidori AM, Venneri MA, Graziadio C, et al. Effect of once-daily, modified-release hydrocortisone versus standard glucocorticoid therapy on metabolism and innate immunity in patients with adrenal insufficiency (DREAM): a single-blind, randomised controlled trial. *Lancet Diabetes Endocrinol*. 2018;6:173–185.
69. Jones CM, Mallappa A, Reisch N, et al. Modified-release and conventional glucocorticoids and diurnal androgen excretion in congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2017;102:1797–1806.
70. Gagliardi L, Nenke MA, Thynne TR, et al. Continuous subcutaneous hydrocortisone infusion therapy in Addison's disease: a randomized, placebo-controlled clinical trial. *J Clin Endocrinol Metab*. 2014;99:4149–4157.
71. Oknes M, Bjornsdottir S, Isaksson M, et al. Continuous subcutaneous hydrocortisone infusion versus oral hydrocortisone replacement for treatment of Addison's disease: a randomised clinical trial. *J Clin Endocrinol Metab*. 2014;99:1665–1674.
72. Alkatib AA, Cosma M, Elamin MB, et al. A systematic review and meta-analysis of randomized placebo-controlled trials of DHEA treatment effects on quality of life in women with adrenal insufficiency. *J Clin Endocrinol Metab*. 2009;94:3676–3781.
73. Hahner S, Spinnler C, Fassnacht M, et al. High incidence of adrenal crisis in educated patients with chronic adrenal insufficiency: a prospective study. *J Clin Endocrinol Metab*. 2015;100:407–416.
74. Rushworth RL, Torpy DJ. Adrenal insufficiency in Australia: is it possible that the use of lower dose, short-acting glucocorticoids has increased the risk of adrenal crises? *Horm Metab Res*. 2015;47:427–432.
75. Meyer G, Koch M, Herrmann E, et al. Longitudinal AddiQol scores may identify higher risk for adrenal crises in Addison's disease. *Endocrine*. 2018;60:355–361.
76. Hahner S, Burger-Stritt S, Allolio B. Subcutaneous hydrocortisone administration for emergency use in adrenal insufficiency. *Eur J Endocrinol*. 2013;169:147–154.

77. Hahner S, Burger-Stritt S, Allolio B. Subcutaneous hydrocortisone administration for emergency use in adrenal insufficiency. *Eur J Endoc inol*. 2013;169:147–154.
78. Tajima T, Okada T, Ma XM, et al. Restoration of adrenal steroidogenesis by adenovirus-mediated transfer of human cytochrome P450 21-hydroxylase into the adrenal gland of 21-hydroxylase-deficient mice. *Gene Ther*. 1999;6:1898–1903.
79. Lerario AM, Finco I, LaPensee C, et al. Molecular mechanisms of stem/progenitor cell maintenance in the adrenal cortex. *Front Endocrinol (Lausanne)*. 2017;8:52.
80. Gan EH, MacArthur K, Mitchell AL, et al. Residual adrenal function in autoimmune Addison's disease: improvement after tetracosactide (ACTH 1-24) treatment. *J Clin Endocrinol Metab*. 2014;99:111–118.
81. Napier C, Gan EH, Mitchell AL, et al. Residual adrenal function in autoimmune Addison's disease - effect of dual therapy with rituximab and depot tetracosactide. *J Clin Endocrinol Metab*. 2019;105:e1250–e1259.
82. Dubernard JM, Cloix P, Tajra LC, et al. Simultaneous adrenal gland and kidney allotransplantation after synchronous bilateral renal cell carcinoma: a case report. *Transplant Proc*. 1995;27:1320–1321.
83. Grodstein E, Hardy MA, Goldstein MJ. A case of human intramuscular adrenal gland transplantation as a cure for chronic adrenal insufficiency. *Am J Transplant*. 2010;10:431–433.
84. Vouillarmet J, Buron F, Houzard C, et al. The first simultaneous kidney-adrenal gland-pancreas transplantation: outcome at 1 year. *Am J Transplant*. 2013;13:1905–1909.
85. Balyura M, Gelfgat E, Ehrhart-Bornstein M, et al. Bovine alginate encapsulated adrenal cells in rats (peritoneum); function and survival. *Proc Natl Acad Sci U S A*. 2015;112:2527–2532.
86. Addison T (1855) On the constitutional and local effects of disease of the suprarenal capsules. London.
87. Al Nofal A, Bancos I, Benkhadra K, et al. Glucocorticoid replacement regimens in chronic adrenal insufficiency: a systematic review and meta-analysis. *Endocr Pract*. 2017;23:17–31.
88. Olafsson AS, Sigurjonsdottir HA. Increasing prevalence of Addison disease: results from a nationwide study. *Endocr Pract*. 2016;22:30–35.
89. Dean B, Chrisp GL, Quartararo M, et al. P450 oxidoreductase deficiency: a systematic review and meta-analysis of genotypes, phenotypes, and their relationships. *J Clin Endocrinol Metab*. 2020;105:dgz255.
90. Calis M, Gokee C, Ates F, et al. Investigation of the hypothalamic-pituitary-adrenal axis (HPA) by 1 microg ACTH test and metyrapone test in primary fibromyalgia syndrome. *J Endocrinol Invest*. 2006;27:42–46.
91. Vuong C, Van Uum SHM, O'Dell LE, et al. The effects of opioids and opioid analogs on animal and human endocrine systems. *Endocr Rev*. 2010;31:98–132.
92. Dittmar M, Kahaly GJ. Extensive personal experience. Polyglandular autoimmune syndromes: immunogenetics and long-term follow up. *J Clin Endocrinol Metab*. 2003;88:2983–2992.
93. Esteban NV, Loughlin T, Yergey AL, et al. Daily cortisol production rate in man determined by stable isotope dilution/mass spectrometry. *J Clin Endocrinol Metab*. 1991;72:39–45.
94. Friderichsen C. [Nebennierenapoplexie bei kleinen Kindern.]. *Jahrbuch für Kinderheilkunde und Physische Erziehung*. 1918;87:109–125.
95. Diri H, Karaca Z, Tanriverdi F, et al. Sheehan's syndrome: new insights into an old disease. *Endocrine*. 2016;51:22–31.
96. Jitske T, Cornelie DA, Alberto MP, et al. Patients with adrenal insufficiency hate their medication: concerns and stronger beliefs about the necessity of hydrocortisone intake are associated with more negative illness perceptions. *J Clin Endocrinol Metab*. 2014;99:3668–3676.
97. Herndon J, Nadeau AM, Davidge-Pitts CJ, et al. Primary adrenal insufficiency due to bilateral infiltrative disease. *Endocrine*. 2018;62:721–728.
98. Naletto L, Frigo AC, Ceccato F, et al. The natural history of autoimmune Addison's disease from the detection of autoantibodies to development of the disease: a long-term follow-up study on 143 patients. *Eur J Endocrinol*. 2019;180:223–234.
99. Meyer EJ, Nenke MA, Rankin W, et al. Total and high affinity corticosteroid-binding globulin depletion in septic shock is associated with mortality. *Clin Endocrinol (Oxf)*. 2019;90:232–240.
100. Pazderska A, Oftedal BE, Napier CM, et al. A variant in the BACH2 gene is associated with susceptibility to autoimmune addison's disease in humans. *J Clin Endocrinol Metab*. 2016;101:3865–3869.
101. Speiser PW, Arlt W, Auchus RJ, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:1–46.
102. Sarett L. The synthesis of hydrocortisone from desoxycholic acid. *J Biol Chem*. 1946;162:601.
103. Udelsman R, Chrousos GP, Loriaux DL. Adaptation during surgical stress. *J Clin Invest*. 1986;77:1377–1381.
104. Waterhouse R. A case of suprarenal apoplexy. *Lancet*. 1911;4566:577–578.
105. Zuckerman-Levin N, Tiosano D, Eisenhofer G, et al. The importance of adrenocortical glucocorticoids for adrenomedullary and physiological response to stress: a study in isolated glucocorticoid deficiency. *J Clin Endocrinol Metab*. 2001;86:5920–5924.

Defects of Adrenal Steroidogenesis

Deborah P. Merke, Phyllis W. Speiser, and Perrin C. White

OUTLINE

Virilizing Forms of Congenital Adrenal Hyperplasia, 1572
  21-Hydroxylase Deficiency, 1572
  11β-Hydroxylase Deficiency, 1576
  P450 Oxidoreductase Deficiency, 1578
  3β-Hydroxysteroid Dehydrogenase Type 2 Deficiency, 1578
Nonvirilizing Forms of Congenital Adrenal Hyperplasia, 1579
  17α-Hydroxylase/17,20 Lyase Deficiency, 1579
  Diagnosis of 17α-Hydroxylase/17,20 Lyase Deficiency, 1580
  Lipoid Adrenal Hyperplasia and Side Chain Cleavage (CYP11A1) Deficiency, 1580
Other Conditions Affecting Adrenal Steroid Synthesis, 1581
  Aldosterone Synthase Deficiency, 1581
  Glucocorticoid-Remediable Aldosteronism, 1581
  Smith–Lemli–Opitz Syndrome, 1582
Differential Diagnosis, 1582
Medical Therapy of Congenital Adrenal Hyperplasia, 1583
  Glucocorticoid Therapy, 1583
  Mineralocorticoid Therapy, 1584
  Sex Steroid Replacement, 1584
  Antihypertensive Therapy, 1584
  Management of Atypical Genitalia, 1584
Future Directions and Novel Therapies, 1585
Conclusion, 1586

The adrenal cortex, a steroid hormone-producing gland, is made up of three distinct functional zones. The outermost zona glomerulosa produces mineralocorticoids, principally aldosterone, that are responsible for sodium and fluid balance. Aldosterone is secreted in response to stimulation from the renin-angiotensin axis and ambient potassium levels. The zona fasciculata produces glucocorticoids, in humans, mainly cortisol, which is required to maintain glucose homeostasis and vascular tone. Adrenocorticotropic hormone (ACTH) is the main regulator of cortisol synthesis. The innermost zona reticularis produces sex steroids, mainly androgen precursors that support secondary sexual characteristics, such as pubic hair. The regulation of the latter class of steroids is less well understood, but they are also ACTH-responsive. A more detailed discussion of basic adrenal anatomy and physiology may be found in Chapters 85–90.

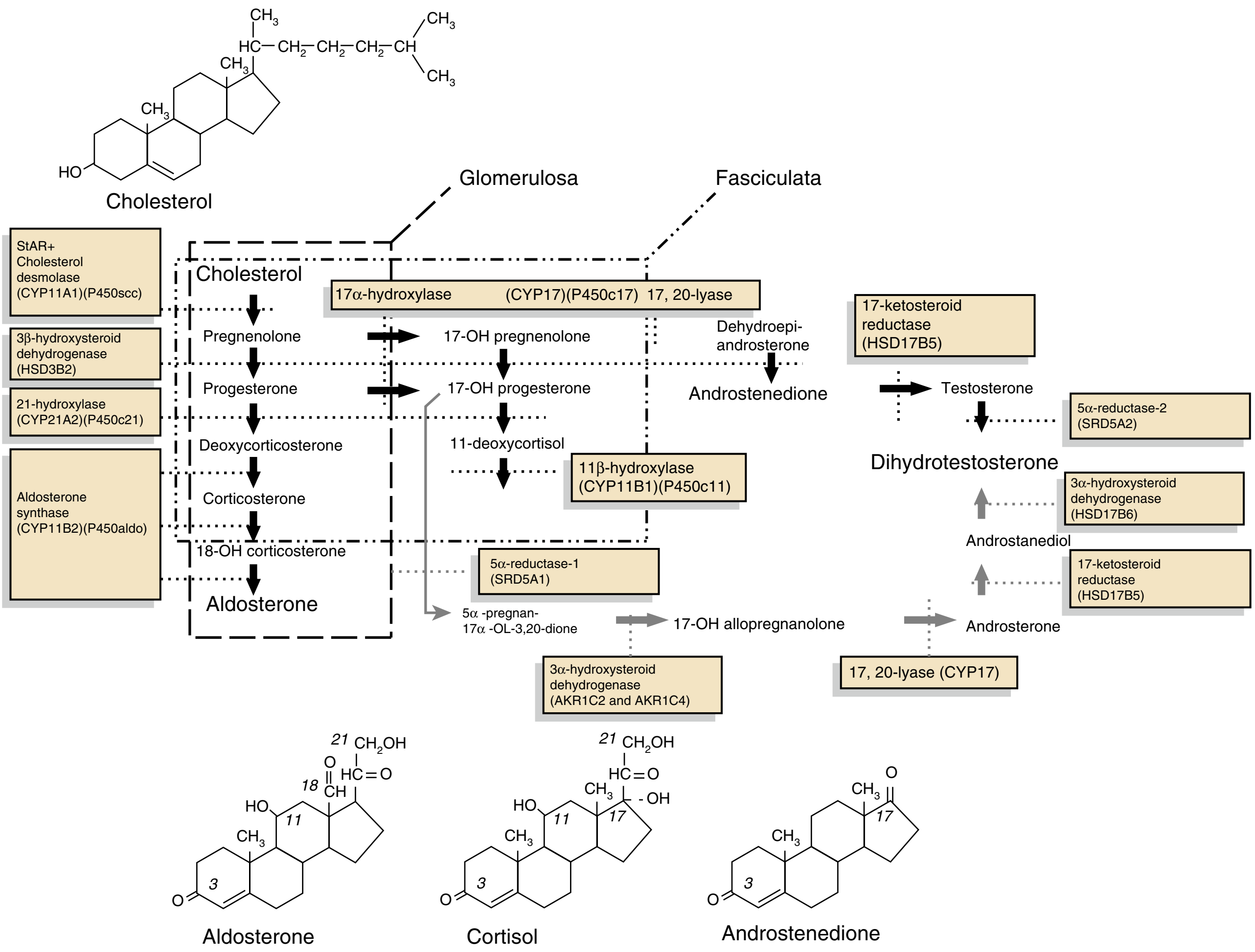
This chapter will address several disorders of adrenal steroidogenesis resulting in overproduction or underproduction of these corticosteroids, depending on the enzymatic defect. The key first step in synthesis of all steroids is importing cholesterol from the outer to the inner mitochondrial membrane, facilitated by steroidogenic acute regulatory protein (StAR).[1] The side chain of cholesterol is then cleaved to yield pregnenolone. A series of enzymatic reactions thereafter lead to aldosterone, cortisol, or androstenedione, representing the three major categories of adrenal cortical products (Fig. 94.1).

The most common adrenal steroidogenic defects are enzymatic deficiencies affecting cortisol production, collectively referred to as *congenital adrenal hyperplasia* (CAH). All are transmitted as autosomal recessive traits. Impaired cortisol synthesis leads to chronic elevation of ACTH via loss of the negative feedback inhibition of the hypothalamic–pituitary–adrenal (HPA) axis. This overstimulation of the adrenal cortex by ACTH results in adrenal hyperplasia. Specific forms of CAH are identified by the abnormal patterns of glucocorticoid, mineralocorticoid, and sex steroid secretion, and by accompanying clinical manifestations, including genital atypia, disturbances in sodium homeostasis and blood pressure regulation, and the postnatal consequences of sex steroid imbalance affecting somatic growth and fertility. A summary of the clinical, hormonal, and genetic features of these steroidogenic defects appears in Tables 94.1 to 94.3.

The different types of CAH may be divided into virilizing and nonvirilizing disorders. In the first group of enzyme deficiencies, sex steroids, especially androgen precursors, are overproduced. Affected 46,XX newborns are often identified by genital atypia. Newborn screening for the most common form of CAH caused by 21-hydroxylase deficiency (21-OHD) has resulted in expedited diagnosis and treatment. Screening programs have greatly reduced the incidence of late diagnosis and progressive signs of androgen excess in patients of either sex. In nonvirilizing forms of CAH, 46,XY individuals show incomplete genital development. Additional nongenital phenotypic abnormalities may be present in most types of CAH, including musculoskeletal, cardiovascular, and metabolic problems. Individualized treatment is required to optimize overall health and quality of life.

KEY POINTS

- The adrenal gland produces glucocorticoids, mineralocorticoids, and sex steroids.
- The most common defects of adrenal steroidogenesis are due to autosomal recessive enzymatic deficiencies affecting cortisol production, collectively referred to as congenital adrenal hyperplasia (CAH).
- Cortisol deficiency results in elevation of adrenocorticotropic hormone (ACTH), and chronic ACTH stimulation of the adrenal glands results in abnormal patterns of glucocorticoid, mineralocorticoid, and sex steroid secretion.
- The clinical manifestation of CAH is closely related to the type and severity of the impairment, with types divided into two major categories: virilizing and nonvirilizing forms.



Cholesterol
Glomerulosa
Fasciculata
StAR+ Cholesterol desmolase (CYP11A1)(P450scc)
3β-hydroxysteroid dehydrogenase (HSD3B2)
21-hydroxylase (CYP21A2)(P450c21)
Aldosterone synthase (CYP11B2)(P450aldo)
Cholesterol
Pregnenolone
Progesterone
Deoxycorticosterone
Corticosterone
18-OH corticosterone
Aldosterone
17α-hydroxylase (CYP17)(P450c17) 17, 20-lyase
17-OH pregnenolone
17-OH progesterone
11-deoxycortisol
11β-hydroxylase (CYP11B1)(P450c11)
Dehydroepi-androsterone
Androstenedione
17-ketosteroid reductase (HSD17B5)
Testosterone
5α-reductase-2 (SRD5A2)
Dihydrotestosterone
3α-hydroxysteroid dehydrogenase (HSD17B6)
Androstanediol
17-ketosteroid reductase (HSD17B5)
Androsterone
5α-reductase-1 (SRD5A1)
5α -pregnan-17α -OL-3,20-dione
17-OH allopregnanolone
3α-hydroxysteroid dehydrogenase (AKR1C2 and AKR1C4)
17, 20-lyase (CYP17)
Aldosterone
Cortisol
Androstenedione

Fig. 94.1 Adrenal steroidogenesis. Biosynthetic steps occurring in the zonae glomerulosa and fasciculata are indicated. Enzymes are boxed with dotted lines extending to arrows denoting each enzymatic conversion; two enzymes, CYP11B2 and CYP17, catalyze several successive enzymatic conversions. Planar structures of cholesterol, aldosterone, cortisol, and androstenedione are illustrated. Additional conversions inside and outside the adrenal cortex generate more potent androgens such as testosterone, dihydrotestosterone, and 11-oxo androgens. The "backdoor," or alternative, pathway from 17-OH progesterone to dihydrotestosterone is illustrated, with the involved intermediates and enzymes printing in italics (also see Chapter 87).

TABLE 94.1 Genetic Causes and Clinical Features of the Various Forms of Congenital Adrenal Hyperplasia

Type of CAH	21-Hydroxylase Deficiency	11β-Hydroxylase Deficiency	17α-Hydroxylase/17,20-Lyase Deficiency	3β-Hydroxy-Steroid Dehydrogenase Type 2 Deficiency	P450 Oxido-reductase Deficiency	Lipoid Adrenal Hyperplasia	Cholesterol Side Chain Cleavage Enzyme Deficiency
Genetic cause (affected gene, OMIM#)	*CYP21A2* #201910	*CYP11B1*, #202010	*CYP17A1*, #202110	*HSD3B2*, #201810	*POR*, #201750	*StAR*, #600617	*CYP11A1*, #118485
Incidence	Classic: 1:10,000 to 1:20,000 NC: 1:200 to 1:1,000	1:100,000 1:6,000 Moroccan Jews[1] NC: unknown	1:50,000 Increased frequency in Brazil[1]	Rare	Rare, 130 cases from 11 countries reported	Rare, mostly Japanese, Korean, and Palestinian populations	Rare, <30 patients, mostly from Eastern Turkey
Affected Organs	Adrenal	Adrenal	Adrenal, gonads	Adrenal, gonads	Adrenal, gonads, liver, skeletal	Adrenal, gonads	Adrenal, gonads

TABLE 94.1 Genetic Causes and Clinical Features of the Various Forms of Congenital Adrenal Hyperplasia—cont'd

Type of CAH	21-Hydroxylase Deficiency	11β-Hydroxylase Deficiency	17α-Hydroxylase/17,20-Lyase Deficiency	3β-Hydroxy-Steroid Dehydrogenase Type 2 Deficiency	P450 Oxido-reductase Deficiency	Lipoid Adrenal Hyperplasia	Cholesterol Side Chain Cleavage Enzyme Deficiency
Genital atypia	Classic: 46,XX NC: No	Classic: 46,XX NC: No	46,XY	Classic: 46,XX, 46,XY NC: No	46,XX, 46,XY (variable)	46,XY NC: 46,XY (variable)	46,XY NC: 46,XY (variable)
Salt-wasting	Classic: Yes NC: No	No	No	Yes	Variable	Classic: yes NC: Minimal to none	Classic: yes NC: Minimal to none
Hypertension[2]	No	Yes	Yes	No	Yes	No	No
Postnatal virilization	Classic: yes NC: yes	Classic: Yes NC: Yes	No	Classic: 46,XX NC: 46,XX	No	No	No
Sex steroid deficiency	No	No	Yes	Classic: yes NC: no	Yes	Yes NC: variable	Yes NC: variable
Fertility	Subfertility	Subfertility	infertility	Classic: infertility NC: subfertility	Subfertility	Infertility	Infertility
Other					± Skeletal malformations ± Maternal virilization		

[1] Due to presence of a founder mutation.

[2] Intrinsic to the disorder. Treatment-related or other cause hypertension may develop in salt-wasting forms of CAH.

CAH, Congenital adrenal hyperplasia; *NC*, nonclassic.

(Adapted from El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390:2198–210.)

TABLE 94.2 Clinical Presentation and Biochemical Findings

		BIOCHEMICAL FINDINGS*					
Type/ Enzyme	**Clinical Presentation**	***Cortisol***	***Corticosterone***	***Aldosterone***	***Renin***	***Androgens***	***Other***
21-hydroxylase deficiency	Atypical genitalia (46,XX only) Virilization <4 years old Neonatal salt-wasting (75%) NC: Precocious pubarche, hirsutism, amenorrhea, female subfertility	↓	↓	↓	↑	↑	↑ 17OHProg
11β-hydroxylase deficiency	Classic: atypical genitalia (46,XX), virilization <4 years old (46,XY), hypertension, hypokalemia NC: Precocious pubarche, hirsutism, amenorrhea, female infertility, ± hypertension	↓	↓	Suppressed	↓	↑	↑ DOC, 11-deoxycortisol
17α-hydroxylase/ 17,20-lyase deficiency	Adolescent female with absence of secondary sexual characteristics, hypertension, hypokalemia	↓	↑	Suppressed	↓	↓	↑ DOC
3β-hydroxysteroid dehydrogenase type 2 deficiency	Atypical genitalia (46,XX- mild; 46,XY), neonatal salt-wasting NC: Precocious pubarche, hirsutism, amenorrhea (46,XX); atypical genitalia (46,XY—mild)	↓	↓	↓	↑	↑DHEA ↓others	↑17OHPreg ↓17OHProg
P450 oxidoreductase deficiency	Ambiguous genitalia, ± skeletal manifestation (Antley–Bixler), ± maternal virilization	Varies	Varies	Varies	Varies	Varies	↑: Pregnenolone, progesterone, 17OHProg, DOC
Lipoid adrenal hyperplasia or SCC enzyme deficiency	Classic: Female genitalia, neonatal salt-wasting NC: Adrenal insufficiency (2 years–adulthood), variable gonadal function, variable genitalia (46,XY-mild)	↓	↓	↓	↑	↓	

*Variable biochemical findings in the nonclassic forms. Administration of standard dose of 250 mcg cosyntropin with serum hormone levels measured at 60 min often needed for diagnosis.

NC, Nonclassic; *17OHP*, 17-hydroxyprogesterone; *DHEA*, dehydroepiandrosterone; *DOC*, deoxycorticosterone; *17OHPreg*, 17-hydroxypregnenolone; *SCC*, side chain cleavage.

(Adapted from El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390:2199–210.)

TABLE 94.3 Steroid profiles in Congenital Adrenal Hyperplasia (CAH)

Type/Enzyme	Cosyntropin stimulation testing*	Other Testing
21-hydroxylase deficiency		
Classic	17-OHP generally >10,000 ng/dL (150 nmol/L)	
Nonclassic	17OHP >1000 ng/dL (30 nmol/L)	Early-morning follicular phase serum 17OHP <200 ng/dL (<6 nmol/L) excludes NC CAH
11β-hydroxylase deficiency	11-deoxycortisol generally >3500 ng/dL for classic and >twice normal for NC	
17α-hydroxylase/17,20-lyase deficiency	Poor response of 17OHPreg and 17OHP Elevated ratio of: DOC/cortisol, corticosterone/sex steroids	
3β-hydroxysteroid dehydrogenase type 2 deficiency	17OHPreg >6600 ng/dL (200 nmol/L) Markedly elevated ratios 17OHPreg/cortisol (generally >180 basal and >480 stimulated)	Poor testosterone response hCG stimulation in infancy
P450 oxidoreductase deficiency	Variable 17OHP response, variable cortisol response (often inadequate)	Urine or serum steroid metabolite profile shows evidence of both 21-hydroxylase and 17-hydroxylase deficiencies¥
Lipoid adrenal hyperplasia or SCC enzyme deficiency	Minimal to no response NC: variable response, ↓cortisol common	Classic: Minimal response hCG stimulation Genetic testing needed to differentiate lipoid CAH and SCC deficiency

*Administration of standard dose of 250 mcg cosyntropin with serum hormone levels measured at 60 min.

¥Urine: Ratio of progesterone metabolite pregnanediol over cortisol metabolites (tetrahydrocortisone [THE] + tetrahydrocortisol [THF] + 5αTHF) is elevated.[56]

NC, Nonclassic; *17OHP*, 17-hydroxyprogesterone; *DHEA*, dehydroepiandrosterone; *DOC*, deoxycorticosterone; *17OHPreg*, 17-hydroxypregnenolone; *hCG*, human chorionic gonadotropin; *SCC*, side chain cleavage.

(Adapted from El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390:2199–210.)

VIRILIZING FORMS OF CONGENITAL ADRENAL HYPERPLASIA

21-Hydroxylase Deficiency

Approximately 90% to 95% of cases of CAH are caused by 21-OHD.[2] The literature has historically described classic and nonclassic forms of this disorder, although current thinking views *CYP21A2* allelic variants and their phenotypic manifestations as a continuum. The classic form is defined by significantly reduced enzyme activity with impaired cortisol production manifesting clinically at birth. Nonclassic forms are caused by less severe enzyme defects, and symptoms are generally milder with absence of genital atypia in affected female newborns.

Clinical Features of Classic 21-Hydroxylase Deficiency. This is the most common cause of atypical external genital development in 46,XX individuals. The prevalence ranges from approximately 1 in 14,000 to 1 in 18,000 live births in most countries that undertake screening, with higher prevalence in regions with consanguineous families.[3] The higher disease prevalence among inbred populations is attributable to the fact that all adrenal steroidogenic enzyme deficiencies are inherited by autosomal recessive transmission, and inbred populations promote propagation of pathogenic alleles by so-called founder effects.

Adrenal steroid 21-hydroxylase (CYP21, P450c21) is required for both glucocorticoid and mineralocorticoid production. It accepts electrons from an NADPH-dependent P450 oxidoreductase (POR; see later) and converts 17-hydroxyprogesterone (17-OHP) to 11-deoxycortisol and progesterone to deoxycorticosterone (DOC), respective precursors for cortisol and aldosterone. When this enzyme is defective, insufficient negative feedback by cortisol stimulates ACTH secretion, causing adrenal hyperplasia with diversion of precursor steroids to the sex hormone pathway, which does not require 21-hydroxylase. Surplus androgens produced in this manner virilize 46,XX fetuses' external genitalia (Fig. 94.1).

Gonadal differentiation proceeds normally during gestation. Because there are no testes to secrete anti-Müllerian hormone, Müllerian duct development in the affected female is also normal, forming the uterus and fallopian tubes. A urogenital sinus, consisting of a conjoined urethra and vagina, may originate near the uterus or more distally just proximal to the perineum. The clitoris is enlarged to a variable degree at birth and may be indistinguishable from a penis. The urethral meatus may open on the peritoneum, on the shaft of the phallus, or rarely at the tip of the glans in the normal male position. Descriptive scoring systems have been developed to characterize genital appearance[4] (Fig. 94.2). With normal ovaries and uterus, childbearing potential exists in females, even though their external genitalia are virilized.

Wolffian duct differentiation is normal in males but generally absent in females, as ductal development in males is supported in a paracrine manner by testosterone secreted by the testes. Elevated systemic levels of adrenal androgens are not usually adequate to support this process. Penile size is normal in newborn males with CAH. However, both males and females who are not diagnosed and treated continuously from early infancy can undergo progressive virilization of the external genitalia.

In approximately 75% of cases, males and females with this enzyme deficiency not only suffer from cortisol deficiency, but also synthesize insufficient aldosterone to maintain sodium homeostasis. Before the era of near-universal newborn screening, infants with this severe, "salt-wasting" type of CAH, especially males who lacked the physical signs of genital ambiguity, often presented with failure to thrive culminating in hypovolemic shock with hypoglycemia, hyponatremia, and hyperkalemia, termed salt-wasting adrenal crisis. Hyperpigmentation of the scrotum, areola, and mucus membranes may be present but is not a reliable diagnostic sign. The remaining 25% of classic CAH-affected infants produce slightly more cortisol and aldosterone, are less likely to suffer acute episodes of adrenal insufficiency, and are said to have "simple virilizing" disease. Such patients may nevertheless benefit from

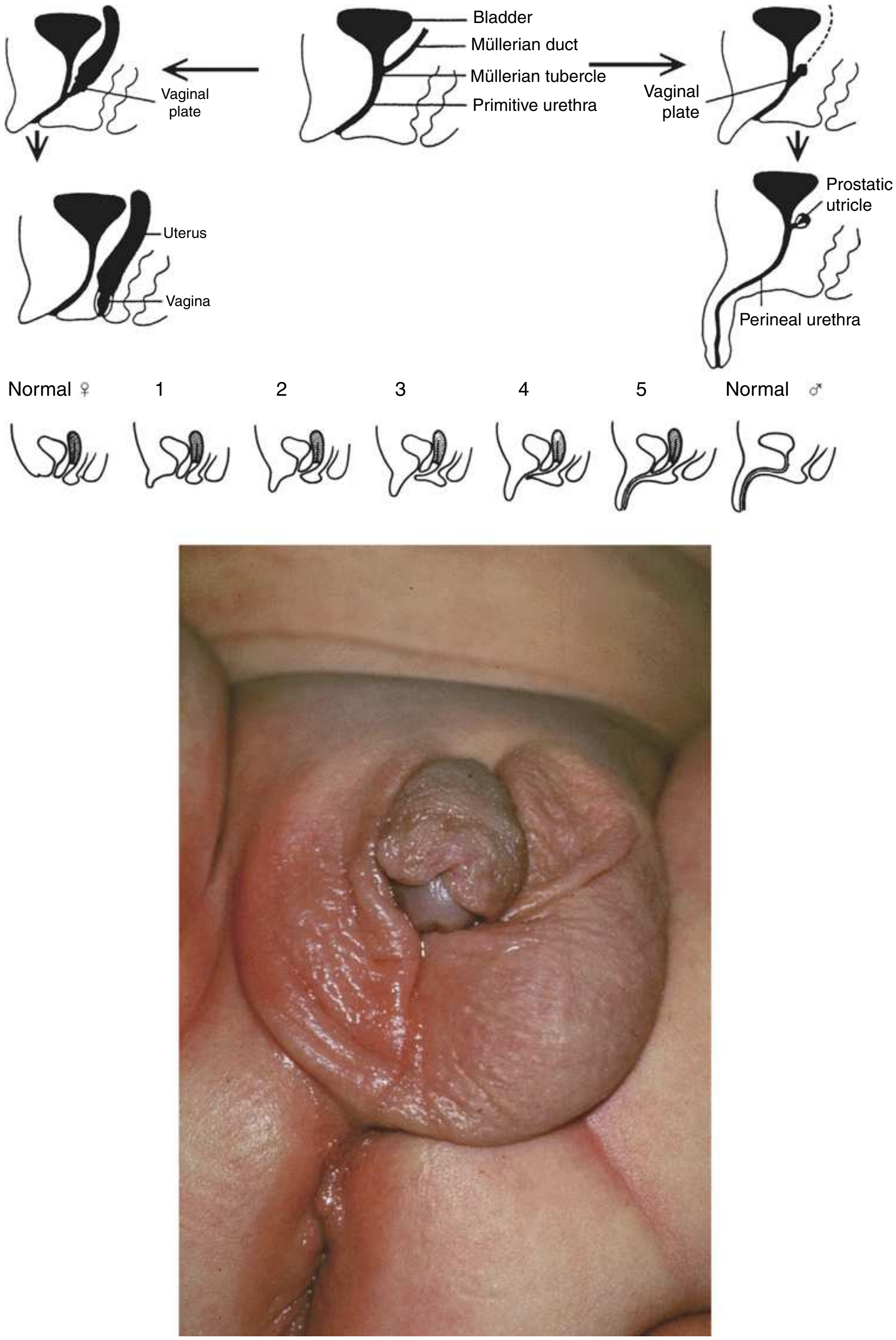

Bladder
Müllerian duct
Müllerian tubercle
Primitive urethra
Vaginal plate
Vaginal plate
Uterus
Vagina
Prostatic utricle
Perineal urethra
Normal ♀
1
2
3
4
5
Normal ♂

Fig. 94.2 *Top,* Schematic representations of normal female and male anatomy flank a series of schematics illustrating different degrees of virilization of females. The uterus persists in virilized females even when the external genitalia have a completely masculine appearance. *Bottom,* An example of virilized external genitalia in a newborn girl with 21-hydroxlase deficiency. There were no gonads in the labioscrotal structure.

low-dose mineralocorticoid treatment in addition to standard glucocorticoid replacement (see later), especially early in life due to relative renal tubular resistance to the salt-retaining effects of aldosterone in early infancy and the low sodium content of infant diets.

Growth Sequelae. Sequelae of adrenal sex hormone excess are recognized in CAH patients, but in some cases are difficult to differentiate from the adverse effects of chronic exogenous glucocorticoid therapy.[5] First among these complications in children are perturbations of growth and puberty. A systematic review and metaanalysis for over 1000 patients showed adult height to be somewhat shorter than average for midparental heights (-1.03 standard deviations, corresponding to ~7 cm).[6] Many of these published reports were based on children diagnosed clinically before implementation of newborn screening, with its resultant early initiation of treatment. Since this degree of height compromise is compatible with short normal stature in many cases, growth-enhancing drugs have not been incorporated into standard practice guidelines.[3] Patients who are not diagnosed early and treated may experience rapid somatic growth and bone maturation, with precocious pubarche. Prevention of hyperandrogenism is important, and early diagnosis and treatment of classic 21-OHD have been associated with improved height outcomes compared with those diagnosed late.[7] If there is marked bone age advancement compared with chronologic age, central precocious puberty may ensue, prompting the need for gonadotropin releasing hormone analog (GnRHa) therapy. Such treatment is not recommended for patients with less than or equal to 2 standard deviations bone age readings and no clinical signs of precocious puberty, nor would GnRHa be effective in patients at or near adult height. There have been few long-term controlled clinical trials of puberty-modulating therapies in CAH, discussed in further detail later.

Reproductive Function. In adolescents and young adults, concerns about pubertal, sexual, and reproductive function are prominent. In addition to precocious puberty, elevated levels of adrenal sex hormones interfere with the hypothalamic–pituitary–gonadal axis. Young women with well-controlled CAH usually experience normal menarche.[8] Those whose condition is less well-controlled often experience acne, female hirsutism, male pattern hair loss, altered body habitus, irregular menses, and ultimately decreased fertility.[9] In up to 50% of poorly controlled males, benign testicular adrenal rest tumors (TARTs) may be detected on ultrasound, or rarely by palpation.[10] Adrenal rest tissue is thought to originate from ether ACTH-stimulated gonadal progenitors or aberrant cells of adrenocortical origin that migrate with primitive gonadal cells during fetal development when the adrenal and gonadal cells separate from the urogenital ridge. Although most commonly found in the testes, adrenal rest tumors have been described in the ovaries, broad ligament, and other locations as well.[11] TARTs are usually bilateral and arise from the rete testes, and if such tumors expand, normal testicular parenchyma can be replaced, impairing testosterone production and causing oligospermia. Treatment for these issues is discussed later.

Screening by testicular ultrasound should begin in adolescence when first appearance often occurs. Optimization of glucocorticoid therapy can shrink TARTs and prevent progressive enlargement and eventual infertility; however, nonreversible fibrotic changes can occur with time. Semen cryopreservation could be offered to patients with TARTs, because irreversible testicular damage may occur. Surgery can be used to treat orchialgia when large masses are present, but has not been able to restore fertility.[12]

Metabolic Sequelae. Alterations in the metabolome are seen in CAH patients treated with glucocorticoids.[13] Conceivably, such changes may mediate some of the long-term adverse outcomes in this disease. Both children and adults are prone to overweight and obesity. Cardiovascular changes including elevated blood pressure and increased carotid intimal thickness have been detected even in the pediatric population, progressing in frequency among adults (reviewed in[14]).

Counseling about healthy lifestyle choices and early attention to cardiometabolic risk factors is indicated in patients receiving chronic glucocorticoid therapy. A metaanalysis of 20 studies of cardiovascular risk factors in glucocorticoid treated patients with CAH found higher prevalence of increased blood pressure, insulin resistance, and carotid intima thickness compared with controls.[14] Furthermore, a Swedish national cohort study of 588 patients with CAH showed an increased prevalence of hypertension, hyperlipidemia, diabetes, and venous thromboembolism in adults with classic CAH compared with controls.[15]

Bone Health. Endogenous adrenal sex hormone excess and exogenous glucocorticoid exposure represent competing influences on bone mineral accrual. A large retrospective European multicenter survey indicated that higher glucocorticoid doses were correlated with low bone density among young women, but fractures were infrequent.[16] The authors concluded that CAH patients may be at increased risk for osteopenia and osteoporosis later in life.

Overall, adults with 21-OHD have lower bone mineral density than control populations, with the prevalence of osteoporosis or osteopenia ranging from 37% to 81%,[17,18] and few studies report increased fracture rate.[19]

Mental Health. With respect to cognitive function, a relatively large and well-controlled Swedish study showed that patients with CAH generally have normal intelligence but may have discretely impaired executive functions.[20] Severe salt-wasting disease with null CAH genotypes was associated with poorer general cognitive capacity. Behavioral investigations point to male-pattern behaviors in 46,XX individuals, although gender identity is largely female.[21,22] Anxiety and emotional problems were no different in children affected with CAH compared with controls,[23] but higher prevalence of anxiety, depression, and alcohol misuse compared with controls were reported in adults in population-based studies.[23-25] Differences have been identified in brain structure relating to visuospatial and memory task performance in young adult CAH patients.[26] Patients who had antenatal exposure to dexamethasone in the first trimester not only showed alterations in brain structure but also had epigenetic changes reflected in altered gene methylation patterns.[27]

Quality of life in adults has been studied extensively; however, results are far from uniform (reviewed in[3]). As with many chronic illnesses, multiple factors determine mental health and social functioning. Although hormonal influences may contribute to mental health, sexual well-being, especially in females born with atypical genitalia, is a key element affecting quality of life. Nonetheless, most are heterosexual and do not have gender identity disorders.[28,29]

Clinical Features of Nonclassic 21-Hydroxylase Deficiency. In the typical nonclassic patient, cortisol and aldosterone insufficiency are subclinical, and androgen excess is less apparent compared with the classic patient.[30] This form of CAH is quite common, but seldom presents in the newborn period and is not the primary target of newborn screening programs, hence the alternative term “late-onset” CAH. The estimated prevalence is approximately 1 in 200 in the White American population.[31] Beyond the newborn period, progressive virilization may be observed in young children, associated with unusually rapid growth, axillary or pubic hair development, acne, and radiographic evidence of accelerated skeletal maturation. The most common clinical presentation is an adolescent or young adult woman with hirsutism, oligomenorrhea, severe acne, and subfertility.[32] Adolescent and adult males, if they are clinically identified, may have severe acne. Rare individuals with nonclassic CAH have been reported with TARTs.[33] In those instances, the *CYP21A2* genotype included one Pro-30-Leu allele, intermediate in severity between nonclassic and classic forms of CAH. Similarly, subfertility is rare in men with nonclassic CAH. A retrospective survey of over 200 Israeli adult men with abnormal sperm analysis found no cases of nonclassic CAH based on cosyntropin stimulation tests.[34] A separate Swedish population-based cohort analysis found no evidence of infertility in young men with nonclassic CAH, although classic CAH was associated with male infertility.[35]

Frank evidence of hypovolemic shock does not usually occur in individuals with nonclassic CAH, unless chronically treated with glucocorticoids, subjected to severe stress (e.g., surgery, febrile illness, or prolonged gastroenteritis), and deprived of sufficient glucocorticoid

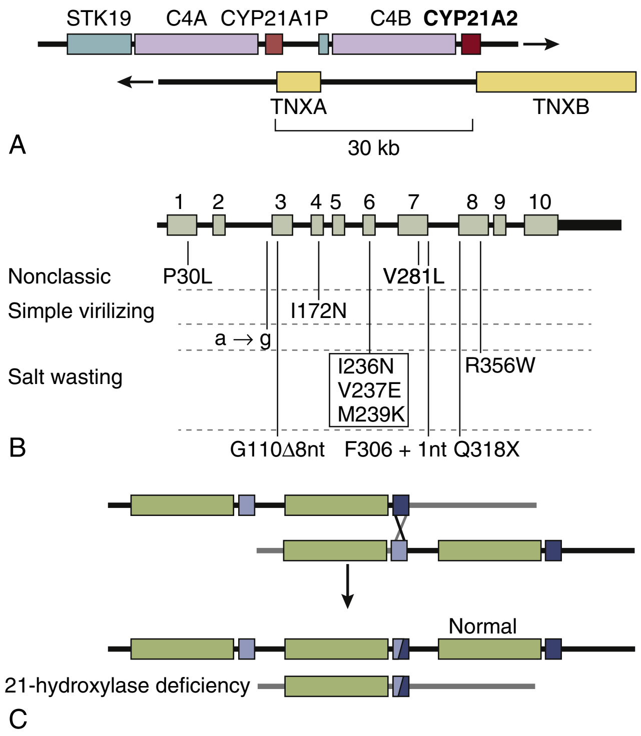


Fig. 94.3 The *CYP21* genes. **A,** The genetic region on chromosome 6p21.3. The genes are described in text, and a 30,000-base pair (30-kb) scale is included. **B,** Structure of the *CYP21* genes. Exons are numbered. Mutations affecting enzymatic function that are normally present in the *CYP21A1P* pseudogene are shown. They are positioned vertically to show the severity of congenital adrenal hyperplasia they cause when transferred to *CYP21A2* in gene conversion events. **C,** An illustration of unequal meiotic crossing-over generating a deletion representing a salt-wasting 21-hydroxylase deficiency allele. The other chromosome has three copies of the tandem and is not associated with disease. For simplicity, only the *C4* and *CYP21* genes are illustrated.

replacement. To our knowledge, there are no cases of previously untreated patients with nonclassic CAH experiencing an adrenal crisis. Additionally, hyperandrogenic signs and symptoms are much more variable and often subtle, such that many individuals diagnosed with nonclassic CAH can avoid glucocorticoid treatment or be treated only when symptomatic. The treatment of CAH is discussed in more detail below.

Molecular Genetics. The 21-hydroxylase enzyme (CYP21, P450c21) is a microsomal cytochrome P450 enzyme (i.e., it is located in the endoplasmic reticulum). Its expression is regulated by trophic stimuli, including ACTH in the zona fasciculata and angiotensin II in the zona glomerulosa (see Chapter 87). The structural gene encoding CYP21, *CYP21A2,* is located on the short arm of chromosome 6 (6p21.3), within a 30-kb tandem duplication (Fig. 94.3) that includes a highly homologous pseudogene, *CYP21A1P.*[2] The duplicated region also contains the *C4B* and *C4A* genes, encoding two isoforms of the fourth component of serum complement, and a small part of *STK19,* which encodes a serine-threonine kinase. On the opposite DNA strand, this region contains part of the *TNXB* gene and the *TNXA* gene, which encode tenascin-X, a connective tissue protein, which, when deleted or mutated, causes hypermobility-type Ehlers–Danlos syndrome (EDS).[36] The combination of CAH and EDS, termed CAH-X syndrome, is due to a contiguous gene deletion and is found in 10% to 15% of CAH patients.[37,38]

The nucleotide sequences of *CYP21A2* and *CYP21A1P* are 98% identical in exons and approximately 96% identical in introns, but *CYP21A1P* carries 10 mutations that prevent synthesis of a functional enzyme, including frameshifts, nonsense (stop codon) and missense (amino acid substitutions) mutations, and mutations affecting splicing of pre-mRNA. Approximately 90% to 95% of 21-OHD alleles are the result of homologous recombination events between *CYP21A2* and *CYP21A1P*. During meiosis, misalignment of homologous chromosomes leads to unequal crossing over, resulting in 30-kb deletions. If the breakpoint of the crossover is within *CYP21A2*, this will prevent synthesis of a functional enzyme. Deletions account for approximately 20% of mutant alleles in most populations. Gene conversion can result from mitotic or meiotic events that transfer one or more mutations normally present in *CYP21A1P* to *CYP21A2*. In general, the functional consequence of each DNA alteration corresponds with the clinical severity of the inherited disorder. Therefore, *CYP21A2* mutations are classified into three general groups in descending order of enzymatic compromise. Deletions, nonsense mutations, frameshifts, and missense mutations that totally abolish enzymatic activity are salt-wasting classic alleles, comprising mutation group A (divided in some analyses into “null” and group A categories). A single-nucleotide mutation that alters splicing of intron 2 (“intron 2 G” mutation) is particularly common, comprising approximately 25% of mutant alleles in most populations. It has been seen in both salt-wasting and simple virilizing patients, suggesting that there is a small amount of normally-spliced mRNA; it is sometimes considered the only member of “mutation group A2.” A nonconservative amino acid substitution, isoleucine-172 to asparagine (I172N), has been associated with the simple virilizing form of the disorder. It reduces enzymatic activity to approximately 1% of normal (mutation group B). Finally, missense mutations such as valine-281 to leucine (V281L) and proline-30 to leucine (P30L) reduce enzyme activity to approximately 20% to 50% of normal (mutation group C) and are associated with nonclassic disease. Although enzyme function *in vitro* appears to be similar,[39] clinical observations suggest that patients carrying the P30L allele are somewhat more symptomatic, straddling the border between simple virilizing and nonclassic forms of CAH.[40] In addition to alleles arising from deletions or gene conversions, there are at least 230 other mutations causing 21-OHD, and close to 700 additional sequence variants of unknown significance have been reported.[41] Because 21-OHD is a recessive disorder, the phenotype of each patient can generally be predicted from the less severely affected allele; e.g., a patient carrying a group A mutation (null enzyme activity) on one allele and a group C mutation (partial enzyme impairment) on the other will usually have nonclassic disease. Nevertheless, such compound heterozygotes may have a somewhat more severe phenotype than patients carrying two group C alleles.[40,42,43]

It is not well understood how other factors—such as individual polymorphisms in the androgen receptor or variations in the activities of other enzymes required for metabolism of steroid precursors—influence phenotype.

Diagnosis of 21-Hydroxylase Deficiency. CAH owing to 21-OHD is diagnosed on the basis of 1) clinical findings, 2) biochemical and hormonal measurements, and 3) when necessary, genotyping *CYP21A2* to identify disease-specific mutations. Phenotypic features have been discussed above. Hormone levels in blood or urine can establish the diagnosis of an enzyme defect by demonstrating an increased precursor-to-product ratio. In classic 21-OHD CAH, serum cortisol levels are usually low to low-normal, inappropriate for the markedly elevated levels of ACTH and 17-OHP; moreover, cortisol measurements by immunoassay may be falsely normal, rather than low, owing to cross-reaction with 21-deoxycortisol. The latter steroid has been suggested as a possible diagnostic alternative to measuring 17-OHP.[44] The typical analyte used for diagnosis of CAH is blood or serum 17-OHP, the primary substrate for 21-hydroxylase, most accurately measured by liquid chromatography followed by liquid chromatography tandem mass spectrometry (LC-MS/MS).

The standard confirmatory diagnostic procedure for all forms of CAH is the 60-minute cosyntropin (ACTH 1–24) stimulation test, generally done after the first 24 to 48 hours of life, or later in very small infants. Blood is drawn before and then 60 minutes after intravenous or intramuscular injection of a supraphysiologic 0.25-mg bolus of ACTH; in low-birth weight, premature infants, the dose may be reduced to 0.125 mg. Hormonal diagnosis should be correlated with clinical and genetic findings. Clinicians should be aware that blood filter paper and serum measurements for 17-OHP are not directly comparable. Ideally, a full adrenal cortical hormone profile should be obtained to exclude other types of CAH; such profiles are available in endocrine specialty laboratories. Aldosterone, plasma renin activity (PRA), serum electrolytes, and glucose should also be measured. Newborns often have elevated levels of PRA, relative hyponatremia, and mild hyperkalemia due to renal immaturity,[45] making it challenging to identify infants as having permanent salt-wasting CAH as opposed to more transient abnormalities of electrolyte homeostasis.

Patients with classic CAH usually have markedly elevated baseline levels of 17-OHP, and treatment of a critically ill infant should not be delayed to perform cosyntropin stimulation testing. In nonclassic CAH, because of diurnal variation beyond infancy, random measurements of baseline serum 17-OHP may be normal. If one obtains early morning (8:00 a.m.) samples, 17-OHP levels are more likely to be elevated. Confirmation of nonclassic CAH often requires cosyntropin stimulation; performing this test also permits evaluation of peak cortisol production and analysis of other adrenal steroids.

Blood levels of 17-OHP from heel-stick filter paper specimens, detected by automated immunoassays, form the basis for CAH newborn screening. Now conducted in all 50 states in the United States, as well as in many other countries, newborn screening has excellent sensitivity for classic CAH, markedly reducing the likelihood of salt-wasting crises owing to delayed diagnosis (especially in males). However, the positive predictive value is low (as low as 1% in many programs), leading to many recalls for retesting, particularly in premature, low-birth weight, or stressed neonates. The positive predictive value is improved by gestational-age adjusted cutoffs and second-tier screening using liquid chromatography followed by LC-MS/MS. Besides measurement of blood 17-OHP, additional data should be collected in the evaluation of a newborn with atypical genitalia potentially indicative of CAH. Karyotype along with fluorescence *in situ* hybridization for *SRY* is done to establish genetic sex. Pelvic and abdominal ultrasound can identify the presence or absence of the uterus, although the nature of gonadal tissue cannot be definitively ascertained by imaging. Because infants with atypical genitalia may also have renal malformations, the kidneys should also be examined by ultrasound.[46] The overall diagnostic approach to genital atypia is discussed in Chapter 101.

Molecular genetic analysis for *CYP21A2* gene mutations is widely available in commercial laboratories; however, it is important to first establish the diagnosis by hormone measurements. Because genotyping remains an expensive test with uncertain insurance reimbursement in the United States, and results are often not reported for several weeks, genotyping should not be the sole determinant for implementing and continuing medical management. Knowing the genotype cannot reliably determine phenotype and whether a given patient has a propensity to salt-wasting.

11β-Hydroxylase Deficiency

It was recognized by the early 1950s that rare patients with CAH develop hypertension rather than mineralocorticoid deficiency, and that the hypertension in such patients responds to glucocorticoid replacement. Most of these patients suffer from 11β-hydroxylase deficiency (11-OHD). This is perhaps the second or third most common form of CAH, comprising approximately 5% of cases or approximately 1 in 100,000 live births in most populations, but is more prevalent in North Africa and the Middle East.[47] Steroid 11β-hydroxylase activity is seen in the zona fasciculata and is required for conversion of 11-deoxycortisol to cortisol, and DOC to corticosterone.[48]

When these hormones are deficient, decreased production of cortisol and corticosterone reduces feedback inhibition on the hypothalamus and anterior pituitary, increasing secretion of ACTH. This stimulates the zona fasciculata of the adrenal cortex to overproduce steroid precursors proximal to the blocked 11β-hydroxylase step.

One of these precursors, DOC, is an active mineralocorticoid. Approximately two thirds of patients with the severe, "classic" form of 11-OHD have high blood pressure,[49,50] often beginning in the first few years of life. Although the hypertension is usually of mild to moderate severity, left ventricular hypertrophy and/or retinopathy have been observed in up to one third of patients, deaths from cerebrovascular accidents have been reported, and hypertensive nephropathy requiring renal transplantation may occur rarely. Other signs of mineralocorticoid excess such as hypokalemia and muscle weakness or paralysis occur in a minority of patients and are not well correlated with blood pressure. Plasma renin activity is usually suppressed in older children, and levels of aldosterone are consequently low, even though the ability to synthesize aldosterone is actually unimpaired.

Patients with 11-OHD often exhibit signs of androgen excess similar to those seen in 21-OHD. Affected females are born with some degree of masculinization of their external genitalia. This includes clitoromegaly and partial or complete fusion of the labioscrotal folds. The genitalia in affected 46,XX infants can be difficult to distinguish from those of a 46,XY cryptorchid infant.[49] Nevertheless, the gonads and the internal genital structures are normal, and affected 46,XX infants have reproductive potential.

As with 21-OHD, other signs of androgen excess that occur postnatally in both sexes include rapid somatic growth in childhood and accelerated skeletal maturation leading to premature closure of the epiphyses and short adult stature. Additionally, patients may have premature adrenarche and acne. Androgens may affect the hypothalamic–pituitary–gonadal axis, leading to amenorrhea or oligomenorrhea in females and true precocious puberty or, conversely, poor spermatogenesis in males.

Because severely affected infants with 11-OHD generally do not have salt wasting crises—in contrast to patients with 21-OHD—severely virilized females are less likely to be diagnosed early and are thus more likely to be raised as males than 21-OHD patients. This happened more frequently in the past, especially in regions with poor access to specialty medical care. Long-term sequelae of this disorder are similar to those described above for classic 21-OHD, including male infertility due to TARTs, although they are less thoroughly studied.

In addition to the classic form of this disorder, there exists a milder nonclassic type of 11-OHD, representing *CYP11B1* allelic variants, analogous to classic and nonclassic 21-OHD. Females with the nonclassic form of 11-OHD typically have normal genitalia at birth. With time they may develop hirsutism, acne, and primary or secondary amenorrhea. Children of both sexes with the nonclassic form of this condition have early pubarche, advanced bone maturation, and adult short stature. Hypertension is not a consistent feature of nonclassic 11-OHD.[51]

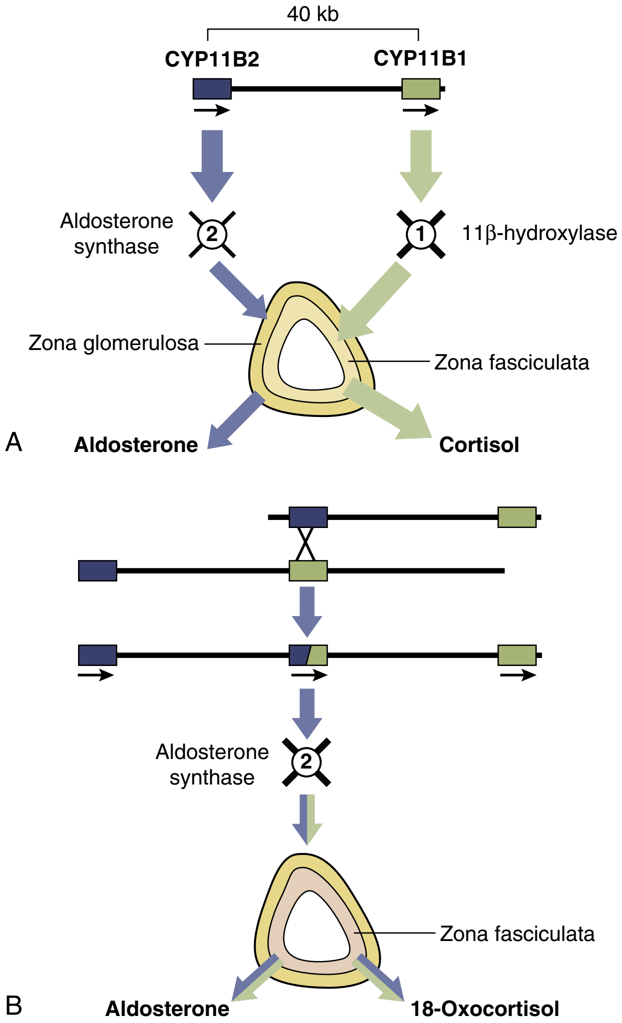


Fig. 94.4 The *CYP11B* genes, located on chromosome 8q24. A, *CYP11B2* and *CYP11B1* are expressed in the zonae glomerulosa and fasculata respectively; mutations in these genes cause aldosterone synthase and 11-hydroxylase deficiencies, respectively. B, Unequal crossing-over generates a chromosome with a 40-kb deletion (not shown) and another chromosome with three genes, the middle one of which has aldosterone synthase activity like CYP11B2 but is expressed in the zona fasciculata like CYP11B1. This causes glucocorticoid-remediable aldosteronism.

Physiology and Molecular Genetics. Humans have two 11β-hydroxylase isozymes that are respectively responsible for cortisol and aldosterone biosynthesis, CYP11B1 (P450c11) and CYP11B2 (P450aldo, P450c11AS) (Fig. 94.4). These isozymes, which are 93% identical in amino acid sequence, are mitochondrial cytochromes P450s located in the inner membrane on the matrix side. They accept electrons from an NAPDH-dependent electron-transport system consisting of two proteins, adrenodoxin reductase and adrenodoxin.

The CYP11B1 isozyme 11β-hydroxylates 11-deoxycortisol to cortisol. It can also convert DOC to either corticosterone (the predominant product) or 18-hydroxy,11-deoxycorticosterone, but it 18-hydroxylates corticosterone poorly. It cannot convert corticosterone into aldosterone. In contrast, CYP11B2 has somewhat weaker 11β-hydroxylase activity than CYP11B1, and it is normally expressed at much lower levels. Therefore, in general, "11β-hydroxylase" refers exclusively to CYP11B1. CYP11B2 can 18-hydroxylate and then 18-oxidize corticosterone to aldosterone, and is therefore commonly termed "aldosterone synthase." Old terminology refers to these latter conversions as being catalyzed by "corticosterone methyloxidase" I and II, respectively, but these terms are no longer useful, given that these reactions are catalyzed by a single enzyme.

CYP11B2 can also convert 11-deoxycortisol to 18-oxocortisol. This reaction normally does not occur at significant levels *in vivo* because CYP11B2 is not expressed in the zona fasciculata, but it becomes important in patients with glucocorticoid-suppressible hyperaldosteronism (see later).

In humans, CYP11B1 and CYP11B2 are encoded by two genes that are located 30 kb apart on chromosome 8q24.3. The coding sequences of these genes are 95% identical. *CYP11B2* is on the left (if the genes are pictured as being transcribed left to right). These genes are regulated in a manner consistent with their respective roles in cortisol and aldosterone biosynthesis; *CYP11B1* is expressed at high levels in normal adrenal glands that are appropriately regulated by ACTH.[52] *CYP11B2* is expressed only in the zona glomerulosa, and is regulated by angiotensin II and potassium levels (also see Chapter 88).

At least 100 mutations in *CYP11B1* have been identified in patients with classic 11-OHD.[53] In Moroccan Jews, a group that has a high prevalence of 11-OHD, almost all affected alleles carry the same mutation, arginine-448 to His (R448H).[54] This presumably represents a founder effect. This and most other missense mutations are in regions of known functional importance and severely inhibit enzymatic activity. Splice site, nonsense, or frameshift mutations also abolish enzymatic activity. Large deletions have rarely been reported, some of which are the products of unequal meiotic crossing-over with the adjacent *CYP11B2* gene, the same process that generates deletions in the *CYP21A2* gene causing 21-OHD. Each patient with mild, nonclassic disease carries at least one mutation that reduces (to ~10%–30% of normal) but does not destroy activity; the other mutation may be either mild or severe.

Although classic patients apparently completely or almost completely lack 11β-hydroxylase activity, they differ significantly in the severity of the various signs and symptoms of their disease. There is no strong correlation between severity of hypertension and biochemical parameters such as plasma levels of the 11β-hydroxylase substrates, deoxycortisol and DOC, or in urinary excretion of tetrahydrodeoxycortisol. Moreover, there is no consistent correlation between the severity of hypertension and degree of virilization.[55] These phenotypic variations must be governed by factors outside the *CYP11B1* locus.

Mutations in *CYP11B2* cause aldosterone synthase deficiency, and recombinations between *CYP11B1* and *CYP11B2* can result in glucocorticoid remediable aldosteronism; these conditions are discussed later in this chapter.

Diagnosis of 11β-Hydroxylase Deficiency. The basic diagnostic procedures are similar to those described for 21-OHD. The steroid profile for 11-OHD is distinctive for high levels of DOC and 11-deoxycortisol, as well as high levels of adrenal androgen metabolites. Because this enzyme deficiency inhibits steroid metabolism one step after steroid 21-hydroxylase, newborn screening may in some cases detect this type of CAH, as it can be associated with elevated 17-OHP levels. For most newborn screening programs, steroidogenic enzyme deficiencies other than 21-OHD CAH are not the primary targets. To differentiate 11-OHD from other types of CAH, detailed steroid profiling, usually in the context of a cosyntropin stimulation test, is performed (see earlier under Diagnosis of 21-Hydroxylase Deficiency). The serum electrolyte profile will occasionally show hypokalemia; plasma renin activity is suppressed in untreated patients. *CYP11B1* genotyping can aid in genetic counseling.

P450 Oxidoreductase Deficiency

POR is an electron donor for all cytochrome P450 enzymes in the endoplasmic reticulum. Thus, POR deficiency (PORD) affects CYP21 and CYP17 activity and presents with signs and symptoms that may resemble 21-OHD, 17α-hydroxylase/17,20 lyase deficiency, or a combination of the two enzyme deficiencies.[56,57] The precise prevalence of this condition is unknown but seems highest in Japan; over 100 cases have been reported in total. Most patients have normal baseline levels of cortisol, but mildly blunted responses to cosyntropin stimulation are frequently seen. Hypertension occasionally occurs, as more frequently happens in patients with 17-hydroxylase deficiency owing to CYP17 mutations. As with that disorder, hypertension is presumably a consequence of elevated levels of DOC resulting from impaired cortisol biosynthesis.

The clinical presentation of PORD usually includes disorders of sex development (DSD) in both XY and XX individuals, from minor degrees of undervirilization in males with micropenis or hypospadias to near-complete virilization of females. The latter fact seems paradoxical, because patients with actual mutations in CYP17 are unable to synthesize sex steroids. This difference arises because PORD preferentially inhibits the 17,20-lyase activity of CYP17, with 17-hydroxylase activity less affected. Thus, deficient 21-hydroxylase activity results in elevated 17-OH pregnenolone and 17-OH progesterone levels. Although, in contrast to 21-OHD, these are not efficiently converted to dehydroepiandrosterone (DHEA) and androstenedione because of low 17,20-lyase activity, 17-OH progesterone can be converted to the potent androgen, dihydrotestosterone, in a series of enzymatic reactions termed the "backdoor," or alternative, pathway[58] (Fig. 94.1; also see Chapter 87).

Maternal virilization has also been observed during pregnancy with a PORD fetus.[59] This is attributed to decreased activity of placental aromatase (CYP19), which is dependent on POR. Consequently, maternal screening reveals low serum and urine estriol.

A unique feature of PORD versus other forms of CAH, occurring in approximately 80% of cases, is a skeletal disorder known as Antley–Bixler syndrome.[57] This may affect the face (midface hypoplasia), cranium (craniosynostosis), hands and feet (arachnodactyly, talipes), large joints (radiohumeral synostosis), femurs (bowing, fractures), and other areas (e.g., scoliosis, pectus excavatum). It is probably a consequence of reduced activity of lanosterol 14α-demethylase (CYP51A1, a P450 enzyme required for cholesterol biosynthesis) and squalene monooxygenase (also called squalene epoxidase, a rate limiting enzyme for sterol biosynthesis), both of which are dependent on POR as an electron donor. The phenotype may also be influenced by abnormal levels of retinoic acid resulting from its reduced metabolism by CYP26 family isozymes.

It is possible to establish genotype–phenotype relationships because two mutations, arginine-457 to histidine (R457H) and alanine-287 to proline (A287P), occur relatively frequently (each represents ~25% of all mutant alleles, but the former is characteristically seen in Japanese patients). Patients have not been found with homozygous null mutations, which would presumably be lethal owing to abolition of activity of all P450 enzymes in the endoplasmic reticulum. Four genotype groups have been established: group B is comprised of patients homozygous for R457H; group C is R457H or A287P on one chromosome and any other missense mutation on the other; group D (the largest genotype group) is R457H or A287P on one chromosome and a null mutation on the other; group A comprises patients with any two other missense mutations. Skeletal malformations are seen in only 20% of patients in group A but invariably occur in group D.[56]

Maternal virilization is most often seen in the mothers of fetuses that carry R457H. In contrast, the R457H mutation is associated with less impaired cortisol secretion than other mutations.[60] These differences suggest that different missense mutations differentially affect molecular interactions with different P450 enzymes. The R457H mutation in POR may affect CYP17 (17-hydroxylase/17,20-lyase) less, resulting in more DHEA and androstenedione being synthesized, and decrease activity of CYP19 (aromatase) more, allowing these steroids to cross the placenta and virilize the mother. In contrast, glycine-539 to arginine (G539R) affects only the 17,20-lyase activity of CYP17A1, thus causing an isolated deficiency of sex hormones.

Diagnosis of P450 Oxidoreductase Deficiency. Diagnosis of PORD employs steroid profiling in urine samples by gas chromatography-mass spectrometry or LC-MS/MS in blood samples, which demonstrate high levels of pregnenolone, progesterone, and 17-OHP and decreased cortisol and androgen metabolites.[61] Nonetheless, PORD patients are seldom identified by newborn screening measuring 17-OHP. Genotyping of *POR* may be done once the hormonal diagnosis is confirmed.

3β-Hydroxysteroid Dehydrogenase Type 2 Deficiency

There are two homologous isozymes of 3β-hydroxysteroid dehydrogenase (3β-HSD).[62] Both oxidize the 3β-hydroxyl of the steroid to a ketone and isomerize the Δ^5 double bond to the Δ^4 position. Thus, the enzymes are formally termed *3β-HSD/*Δ^5>Δ^4 *isomerases*. The type 1 isozyme, encoded by the *HSD3B1* gene, is expressed in placenta and in peripheral tissues including small intestine, skin, colon, and liver. The type 2 isozyme, encoded by *HSD3B2,* is expressed in the adrenals and gonads. Mutations in the latter isozyme cause a rare form of CAH, accounting for less than 1% of cases, or less than 1:10^6 births.

In CAH due to 3β-HSD type 2 deficiency, steroid synthesis is blocked early in the pathway from cholesterol to end-products in both the adrenal cortex and gonads. The steroid profile is notable for elevated ratios of Δ^5 steroids (pregnenolone, 17-hydroxypregnenolone, DHEA) to Δ^4 steroids (progesterone, 17-OH progesterone, androstenedione) with relatively low levels of hormones dependent on this enzyme, including aldosterone, cortisol, and sex steroids. Salt-wasting adrenal crisis may occur and may be fatal if untreated. Male and female infants are born with variable degrees of genital atypia. XY individuals are incompletely masculinized owing to low levels of testosterone and dihydrotestosterone. The abnormalities typically involve the external genitalia (perineal hypospadias), but internal Wolffian structures (e.g., vas deferens) are unaffected. This is similar to findings in other disorders of androgen biosynthesis such as 17β-HSD type 3 (17-ketosteroid reductase) deficiency, as well as 5α-reductase type 2 deficiency. Despite low levels of sex steroid production, DHEA (itself not an active androgen) may be converted to more potent androgens via the unaffected 3β-HSD type 1 isozyme, causing mild virilization in girls. High levels of 17-hydroxypregnenolone may undergo extraadrenal conversion to 17-OHP, leading to positive results on newborn screening tests.[63]

Diagnosis of 3β-Hydroxysteroid Dehydrogenase Deficiency. The early diagnosis of 3β-HSD deficiency is somewhat less straightforward than other forms of CAH, unless frank salt-wasting and adrenal crisis are observed. It is rare, whereas hypospadias and cryptorchidism from other causes occur relatively frequently in newborn males, and affected females have only mildly virilized external genitalia.

Moreover, 3β-HSD type 2 activity is highly regulated and varies widely during development and childhood (also see Chapter 87). It is normally low during the second trimester of gestation, rises near the end of gestation (thus increasing fetal cortisol secretion, which helps to trigger parturition), remains high during childhood, and decreases at adrenarche (thus facilitating synthesis of adrenal androgens).[64] In light of these developmental changes in enzyme activity, premature infants may be misdiagnosed as having 3β-HSD type 2 deficiency, just as they are liable to fail the newborn screen for 21-OHD. Premature adrenarche in children represents a similar diagnostic pitfall. Moreover, there is a high frequency of elevated Δ^5/Δ^4-steroid ratios in women with hirsutism, but these are very rarely associated with mutations in *HSD3B2*.[62] It is therefore important to confirm the diagnosis by cosyntropin stimulation testing and genotyping *HSD3B2*.[65] Basal or ACTH-stimulated serum levels of 17-OH pregnenolone above 6600 ng/dL (200 nmol/L) raise suspicion for this diagnosis; detailed age-related hormonal diagnostic criteria, from newborns to adults, have been determined in genotype-confirmed cases.[66,67]

Molecular Genetics. The HSD3B1 and HSD3B2 genes are approximately 93% identical in amino acid sequence and are located in tandem 100 kb apart on chromosome 1p13.1. Humans also have five pseudogenes in the same genetic region; two are located between the two active genes. Although this genetic arrangement is reminiscent of the duplicated *CYP21A1P* and *CYP21A2* genes (recombinations between which cause the vast majority of 21-OHD alleles), recombinations between *HSD3B2* and the adjacent pseudogenes apparently do not generate 3β-HSD deficiency alleles (deletions or gene conversions), presumably reflecting differences in recombination frequencies in the different chromosomal regions in which these gene clusters are located.

Genotype and phenotype are generally highly correlated in cases of classic 3β-HSD type 2 deficiency.[62,68] No functional 3β-HSD type 2 enzyme is expressed in the adrenals and gonads of patients suffering from the severe salt-wasting form, either owing to null (nonsense or frameshift) mutations in *HSD3B2*, or missense mutations that completely destroy enzymatic activity. Non-salt-losing classic 3β-HSD type 2 deficiency results from missense mutations that impair enzyme activity but leave sufficient activity to prevent salt wasting. In many cases these mutations change the kinetic behavior of the enzyme, but some seem to affect protein stability. Individuals diagnosed with putative mild, "nonclassic" 3β-HSD deficiency rarely carry mutations in *HSD3B2,* and heterozygous carriers of *HSD3B2* mutations are asymptomatic.

KEY POINTS

- Mutations in adrenal steroidogenic pathway genes result in four types of virilizing forms of congenital adrenal hyperplasia (CAH): 21-hydroxylase deficiency (*CYP21A2* gene), 11β-hydroxylase deficiency (*CYP11B1* gene), P450 oxidoreductase deficiency (*POR* gene), and 3β-hydroxysteroid dehydrogenase deficiency (*3βHSD2* gene).
- 21-hydroxylase deficiency accounts for approximately 95% of cases of CAH and is the most common cause of 46,XX atypical genitalia.
- The clinical features of the virilizing forms of CAH include genital atypia, abnormal puberty and growth patterns, reduced fertility, and variable degrees of salt-wasting and hypertension, depending on the form.
- There is a continuum of disease severity and phenotypic variability that mostly correlates with genotype.
- Nonclassic 21-hydroxylase is common and can be subclinical, whereas nonclassic forms of the other types of virilizing CAH are rare.
- Diagnosis is based on clinical findings, biochemical and hormonal measurements, and, when necessary, genotyping.
- Newborn screening for 21-hydroxylase deficiency is in place in the United States and many other countries, is based on 17-hydroxyprogesterone level from a heel-stick filter paper specimen, and can sometimes detect other forms of virilizing CAH.

NONVIRILIZING FORMS OF CONGENITAL ADRENAL HYPERPLASIA

17α-Hydroxylase/17,20 Lyase Deficiency

The conversions of pregnenolone to 17-OH pregenolone, and progesterone to 17-OH progesterone, are catalyzed by the 17-hydroxylase activity of CYP17A1 (P450c17), a P450 enzyme in the endoplasmic reticulum. The 17,20-lyase activity of the same enzyme further catalyzes the conversion of 17-OH pregnenolone to dehydroepiandrosterone, an essential step for all sex steroid synthesis (the direct conversion of 17-OH progesterone to androstenedione is not efficiently catalyzed by the human enzyme, see Chapter 87). These reactions utilize electrons donated by POR. Efficient redox coupling of NADPH oxidation to the 17,20 lyase reaction also requires cytochrome b5.[69]

CAH caused by mutations in *CYP17A1* is thought to be either the second or third most common form of CAH, after 21-hydroxylase and 11β-hydroxylase deficiencies. The exact worldwide prevalence is unknown, but clusters of cases have been identified in several locales (see later).

Most cases have impaired activities of both 17α-hydroxylase and 17,20-lyase. Although these deficiencies impair cortisol biosynthesis, affected patients do not suffer from adrenal insufficiency, in contradistinction to the presentation of 21-OHD. This is because synthesis of corticosterone, which is not a 17-hydroxylated steroid, is unaffected. Corticosterone is a glucocorticoid agonist (indeed, it is the main glucocorticoid in rodents and birds, which do not express CYP17A1 in the adrenal gland), but it is considerably (~3-fold) less potent than cortisol because the lack of a 17-hydroxyl group weakens its ability to bind the human glucocorticoid receptor. Thus, much more corticosterone must be secreted in patients with CYP17A1 deficiency than normal cortisol secretion in unaffected individuals. This increases secretion of corticosterone precursors, particularly DOC. As discussed previously, DOC is a potent mineralocorticoid. As a consequence, and similar to 11-OHD, patients with 17α-hydroxylase/17,20 lyase deficiency (17-OHD) frequently suffer from sodium-sensitive, low-renin hypertension and hypokalemia.

In contrast to the virilizing forms of CAH, 17-OHD results in underdeveloped or atypical external genitalia in males and amenorrhea in females.[65,69] Both males and females fail to undergo normal pubertal development and have sparse pubic and body hair. Male gynecomastia, though incompletely explained, may occur due to testosterone deficiency, with consequent relative estrogen excess produced by alternative enzymatic pathways.[70]

Over 100 mutations in CYP17A1 have been identified in patients with combined 17-OHD. Particular mutations prevail in specific locales owing to founder effects. The first mutation characterized was a duplication of four nucleotides in proline-480 (P480Hfs) found in Dutch Frieslanders and in Canadian Mennonites originally from this region. In Brazil, 17-OHD may be the second most common cause of CAH after 21-OHD, owing to two mutations, arginine-362 to cysteine (R362C) and tryptophan-406 to arginine (W406R). In China, a frameshift involving tyrosine-329 (Y329Kfs) and a deletion of glutamate-487 through phenylalanine-489 (D487-F489del) account for approximately 60% and approximately 20% of mutant alleles, respectively, and histidine-373 to leucine (H373L) has been identified in Korean, Chinese, and Japanese patients.

Rarely, the 17,20-lyase activity of CYP17A1 is impaired, with 17-hydroxylase activity relatively or completely preserved. Such patients have isolated defects of sex steroid synthesis but do not have adrenal hyperplasia or hypertension. In some cases, this is the result of mutations that impair interactions of CYP17A1 with POR or cytochrome b5. Examples include arginine-347 to histidine or cysteine (R347H or R347C) and arginine-358 to glutamine (R358Q), which involve residues on the enzyme surface that dock with accessory proteins, and glutamate-305 to glycine (E305G) in the active site. Mutations in accessory proteins can produce the identical phenotype. These include the unusual POR mutation glycine-539 to arginine (G539R) (most POR mutations affect more P450 enzymes) and *CYB5A* (the gene for cytochrome b5) mutations tryptophan-27 to stop (W27X) and histidine-44 to leucine (H44L).

The steroid profile for 17-OHD is distinctive for high levels of DOC, corticosterone (and their 18-hydroxy metabolites), pregnenolone, and progesterone, as well as low levels of cortisol, androgen, and estrogen precursors. In addition to treatment with glucocorticoids, these patients require gender-specific sex hormone replacement therapy to support puberty. Hypertension unresponsive to glucocorticoid treatment often requires separate treatment in 17-OHD, as in 11-OHD.

Mild, nonclassic cases of 17-OHD appear to be quite rare but may be difficult to ascertain. Females with mild 17-OHD might have irregular menses or decreased fertility, whereas affected males might have mildly low testosterone levels with mildly elevated gonadotropins or perhaps oligospermia. One child with mild undervirilization, normal blood pressure, and biochemical evidence of partial 17-OHD was a compound heterozygote for a frameshift mutation in arginine-36 (R36fs) and a missense mutation, tryptophan-121 to arginine (W121R), which partially impairs enzymatic activity. One single-nucleotide polymorphism in *CYP17A1* (rs1004467) has been identified as a susceptibility allele for hypertension, but it does not change the coding region of the enzyme and is thus of uncertain functional significance.

Diagnosis of 17α-Hydroxylase/17,20 Lyase Deficiency

The diagnosis of 17-OHD depends on finding elevated precursor-to-product ratios with cosyntropin stimulation. In general, the diagnosis relies on high levels of DOC (>100 ng/dL [>3 nmol/L]) and corticosterone (>4000 ng/dL [>116 nmol/L]) with low cortisol (<5 mcg/dL [<138 nmol/L]), androgens, and estrogens. Progesterone is also elevated, while aldosterone and renin are suppressed.[71] Gonadotropins and corticotropin (ACTH) are elevated, even in children. A characteristic feature of lyase deficiency is an elevated ratio of 17-OHP/androstenedione (>50), with low levels of sex steroids. As with other types of CAH, genotyping may clarify the diagnosis when hormone measurements are inconclusive.

Lipoid Adrenal Hyperplasia and Side Chain Cleavage (CYP11A1) Deficiency

Lipoid congenital adrenal hyperplasia (LCAH) is a rare, severe form of CAH that affects conversion of cholesterol to pregnenolone and thus impairs synthesis of all adrenal and gonadal steroids.[72]

In classic patients, mineralocorticoid and glucocorticoid deficiencies can cause salt-wasting crises and vascular collapse. Patients may die in infancy. 46,XY males have female-appearing external genitalia, but no internal Müllerian structures. In older children, serum anti-Müllerian hormone and inhibin levels are low.[73] Karyotypic females have normal internal and external genital structures.

LCAH is caused by defects in the StAR protein,[74] which normally facilitates cholesterol transport from the outer to the inner mitochondrial membrane, where the side chain of cholesterol is cleaved to produce pregnenolone, catalyzed by the cholesterol side chain cleavage enzyme (cholesterol desmolase, CYP11A1, P450scc).[75]

Mutations in StAR affect steroidogenesis through two "hits." First, because importation of cholesterol into mitochondria is the rate-limiting step of steroidogenesis, deficient StAR activity directly impairs steroid biosynthesis in affected tissues by approximately 90%. Second, because cholesterol cannot be utilized, cholesteryl esters accumulate in lipid droplets and eventually induce apoptosis in steroidogenic cells. Prenatally these include cells of the adrenal cortex and testicular Leydig cells. However, ovaries do not synthesize steroids prenatally. Thus, affected 46,XX females can develop secondary sexual characteristics and even menstruate in adolescence, because ovarian theca and granulosa cells do not accumulate abnormal levels of lipids until they become hormonally active. There is a single dominant ovarian follicle during each menstrual cycle. Sufficient estrogen is produced by that follicle during the follicular phase of the menstrual cycle to maintain secondary sexual characteristics (and to trigger estrogen-withdrawal bleeding at the end of the cycle) but is usually insufficient to trigger ovulation, and progesterone secretion later in the cycle is blunted owing to accumulating intracellular lipids. Assisted reproductive technologies can sometimes facilitate fertility in such patients.

Steroidogenesis in the placenta is not dependent on StAR, and thus the placenta of a fetus with StAR deficiency is able to produce the progesterone and estrogen necessary to sustain the pregnancy.

Although still rare, LCAH occurs at a somewhat higher frequency—approximately 1 in 300,000—in Japanese and Korean patients, most of whom carry the same nonsense mutation, glutamine-258 to stop (Q258X). A characteristic mutation is also seen in Palestinian Arabs, arginine-182 to leucine (R182L).

A nonclassic form of LCAH is recognized, caused by mutations, particularly arginine-188 to cysteine (R188C), that retain 20% to 30% of normal activity. Patients typically experience the gradual onset of adrenal insufficiency between 2 and 4 years of age but may not present until they are middle-aged adults. Mineralocorticoid deficiency may be evident only as a mild elevation of renin activity in patients on a low-salt diet. Males have normal external genitalia but may have primary gonadal failure and azoospermia.[76]

Before the discovery of StAR, it was assumed that LCAH was caused by mutations in *CYP11A1* (encoding the actual side chain cleavage enzyme). *CYP11A1* mutations do indeed occur, albeit less frequently than *STAR* mutations. Most of the reported mutations do not completely destroy enzymatic activity, suggesting that *CYP11A1* null mutations tend to be lethal prenatally. The reason for this is that, unlike *STAR* mutations, *CYP11A1* mutations interfere with placental steroid synthesis. Miscarriages and preterm births have been reported. The phenotype is otherwise similar to that produced by *STAR* mutations, but without intracellular accumulation of lipid droplets. The nonclassic form of CYP11A1 deficiency has variable age of onset of adrenal insufficiency and a range of gonadal effects including 46,XY hypospadias, cryptorchidism, or normal male genitalia, but with eventual gonadal failure and azoospermia.[77]

CYP11A1 is a mitochondrial P450 enzyme (see Chapter 87), and like the CYP11B isozymes, it receives electrons from an NADPH-dependent transport chain consisting of adrenodoxin reductase and adrenodoxin. It might be predicted that mutations in either of these latter proteins would yield a phenotype similar to CYP11A1 deficiency, by analogy with POR mutations, which affect activity of "microsomal" P450 enzymes in the endoplasmic reticulum (see the section on P450 Oxidoreductase Deficiency). However, mutations in the genes encoding adrenodoxin reductase and adrenodoxin have not been detected, suggesting that they are lethal prenatally.[72]

Diagnosis of LCAH and side chain cleavage deficiencies: Patients with classic forms of LCAH and CYP11A deficiency fail to respond to acute or chronic stimulation with either ACTH or human chorionic gonadotropin. Due to deficiency of all adrenal and gonadal steroids, levels of ACTH, gonadotropins, and renin are elevated. Hypoglycemia, hyponatremia, and hyperkalemia may be observed. The diagnosis may be confirmed by sequence analysis of *STAR* and *CYP11A1*.

KEY POINTS

- The non-virilizing forms of congenital adrenal hyperplasia (CAH; 17α-hydroxylase/17,20-lyase deficiency, lipoid adrenal hyperplasia, and side chain cleavage enzyme deficiency) have sex steroid deficiency, resulting in underdeveloped or atypical external genitalia in 46,XY individuals and lack of normal pubertal development in both sexes.
- Unlike other forms of CAH, patients with 17α-hydroxylase/17,20-lyase deficiency do not suffer from adrenal insufficiency because corticosterone, a glucocorticoid agonist, is produced in excess.
- 17α-hydroxylase/17,20-lyase deficiency caused by mutations in *CYP17A1* is thought to be either the second or third most common form of CAH, with increased prevalence in various locations, including Brazil.
- Lipoid CAH, caused by mutations in the gene encoding steroidogenic acute regulatory protein, and deficiency of side chain cleavage enzyme encoded by the *CYP11A1* gene, have similar clinical and biochemical profiles, necessitating genotyping to distinguish between the two.
- A spectrum of phenotypes of the nonvirilizing forms of CAH have been described.

OTHER CONDITIONS AFFECTING ADRENAL STEROID SYNTHESIS

Aldosterone Synthase Deficiency

The most frequent defect of aldosterone biosynthesis is congenital adrenal hyperplasia due to steroid 21-OHD (see earlier). Signs of androgen excess are prominent in such patients. Rare patients have aldosterone deficiency in the context of entirely normal cortisol and sex steroid synthesis. This is most often caused by deficient activity of the aldosterone synthase enzyme (corticosterone methyloxidase, CYP11B2, P450aldo, P450c11AS, P450cmo). This mitochondrial P450 enzyme, expressed in the adrenal zona glomerulosa, catalyzes a series of three conversions: DOC to corticosterone, then to 18-hydroxycorticosterone, and finally, to aldosterone[48] (also see Chapter 88).

Mineralocorticoid deficiency leads to excessive sodium excretion and potassium retention in the renal distal tubule and cortical collecting duct, causing hyperkalemia, hyponatremia, and hypovolemia. The clinical presentation of aldosterone synthase deficiency varies with age.[78] Infants may develop signs and symptoms of mineralocorticoid deficiency at a few days to weeks of age. These include vomiting and dehydration leading to hypovolemia and, occasionally, frank shock. Serum sodium is usually in the range of 120 to 130 mEq/L, whereas serum potassium ranges from 6.0 to 8.5 mEq/L. Although fatalities have occurred, the morbidity of aldosterone synthase deficiency is usually not as severe as that engendered by the salt-wasting forms of congenital adrenal hyperplasia. This presumably reflects normal synthesis of DOC, corticosterone, and cortisol in aldosterone synthase deficiency, which ameliorate the development of hyponatremia and shock. Some children are diagnosed in early childhood with failure to thrive, anorexia, mild dehydration, and electrolyte abnormalities. Although electrolytes usually normalize by a few years of age (especially on the typical Western high-salt diet), growth retardation may persist throughout childhood. Adults are usually asymptomatic but occasionally tolerate severe salt loss less well than unaffected individuals. Plasma renin activity is markedly elevated in affected infants and young children, but it may be normal in adults.

Aldosterone synthase deficiency was previously divided into two subtypes termed corticosterone methyloxidase (i.e., aldosterone synthase) deficiency types I and II.[78] These syndromes have identical clinical features but differ in profiles of secreted steroids; thus, the distinction is not very useful clinically. Levels of DOC are increased, and urinary excretion of corticosterone metabolites is elevated in both type I and type II deficiencies relative to excretion of cortisol metabolites. Whereas excretion of 18-hydroxycorticosterone (the intermediate between corticosterone and aldosterone) is mildly decreased in type I deficiency, urinary and serum levels of this steroid are dramatically increased in patients with type II deficiency. Aldosterone and its metabolites are undetectable in patients with type I deficiency, whereas urinary excretion is mildly decreased in type II deficiency, and serum levels of aldosterone are usually within normal limits, particularly in older children and adults. Type II deficiency may thus be readily diagnosed by marked (often 100-fold) elevation of the ratio of 18-hydroxycorticosterone to aldosterone in either urine or serum; the ratio does not vary with age in affected individuals and may be the sole biochemical abnormality in adults. Indeed, asymptomatic adults have occasionally been ascertained in family studies because of this ratio. However, this ratio is not useful for diagnosis of type I deficiency, because the very low levels of aldosterone render it meaningless.[78]

Genetic Analysis. Patients with aldosterone synthase type I deficiency all carry mutations in *CYP11B2*, such as frameshift or nonsense mutations, or nonconservative missense mutations that destroy enzymatic activity.[79] The converse is not always true; some patients with such mutations may have type II deficiency.

Iranian Jewish patients with type II deficiency are all homozygous for two mutations, R181W and V386A.[80] Genetic studies suggest that 0.4% of normal 18-oxidase activity is not rate-limiting for aldosterone biosynthesis, but 0.2% is. This is consistent with CAH due to 21-OHD, in which patients hemizygous for the I172N mutation with approximately 1% of wild-type 21-hydroxylase activity are also able to synthesize normal amounts of aldosterone.

Some infants with isolated hyperreninemic hypoaldosteronism do not carry mutations in *CYP11B2.* In some male infants, this may represent adrenal hypoplasia congenita due to NR0B1 (DAX1) mutations, which sometimes presents with isolated hypoaldosteronism, with cortisol deficiency developing later[81,82] (also see Chapter 86). In other kindreds, this may imply the existence of a distinct inherited form of hyperreninemic hypoaldosteronism, which has been termed familial hyperreninemic hypoaldosteronism-2 (FHHA2), distinct from aldosterone synthase deficiency (thus termed FHHA1).[83]

Glucocorticoid-Remediable Aldosteronism

Glucocorticoid remediable aldosteronism (GRA; also called dexamethasone- or glucocorticoid-suppressible hyperaldosteronism) is a form of hypertension inherited in an autosomal dominant manner with high penetrance. It is characterized by moderate hypersecretion of aldosterone, suppressed plasma renin activity, and rapid reversal of these abnormalities after administration of glucocorticoids. This feature distinguishes it from other familial forms of primary aldosteronism.[84,85] The clinical hallmark of GRA is early-onset hypertension, even in infancy or early childhood, but usually prior to age 20 years. The family history is often notable for several similarly affected generations in a kindred, some of whom may have died before age 40 years with cerebrovascular accidents precipitated by hypertensive crisis. In general, the blood pressure is approximately 30 mm Hg higher than in unaffected family members. Thus, not all affected individuals actually have hypertension if family members otherwise have low blood pressure.[86]

All patients with glucocorticoid-suppressible hyperaldosteronism have the same type of mutation, a chromosome that carries three *CYP11B* genes instead of the normal two[87,88] (Fig. 94.4). The middle gene on this chromosome is a chimera, with 5′ and 3′ ends corresponding to *CYP11B1* (encoding 11β-hydroxylase) and *CYP11B2* (encoding aldosterone synthase), respectively. The chimeric disease-associated gene is flanked by normal *CYP11B2* and *CYP11B1* genes. The chimeric gene is regulated like *CYP11B1* (expressed at high levels in the zona fasciculata and regulated primarily by ACTH) because it has transcriptional regulatory sequences identical to those of *CYP11B1.* If the chimeric gene has enzymatic activity similar to that of *CYP11B2,* a single copy of such an abnormally regulated gene should be sufficient to cause the disorder, consistent with the known autosomal dominant mode of inheritance of this syndrome.

Hypokalemia occurs often but not invariably; it is exacerbated by thiazide diuretics. Chronic hypokalemia may lead to polyuria, nocturia, enuresis, and polydipsia. Muscle weakness and discomfort, intermittent paralysis, fatigue, and growth failure affect children with severe hypokalemia.

Serum pH and the carbon dioxide and sodium concentrations may be elevated, and the serum chloride and magnesium levels decreased. Serum levels of calcium are normal. Plasma levels of aldosterone may be normal or elevated. Aldosterone concentrations in 24-hour urine collections are always increased. Plasma levels of renin are consistently low. The ratio of plasma aldosterone concentration to renin activity is always high, and this represents a cost-effective screening test for primary aldosteronism.[89] However, both renin and aldosterone levels may vary by time of day, posture, and sodium intake, making it difficult to establish consistent reference ranges. Therefore, a consistent sampling protocol should be used—for example, midmorning after the patient has been sitting for 15 min. If possible, antihypertensive drugs or other medications that can affect aldosterone or renin secretion should be avoided for several weeks prior to testing, including diuretics, β-blockers, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, clonidine, and nonsteroidal antiinflammatory agents. Calcium channel blockers have smaller effects on the biochemical measurements.

Urinary and plasma levels of 18-oxocortisol and 18-hydroxycortisol—17-hydroxylated homologs of aldosterone and 18-hydroxycorticosterone, respectively—are markedly increased in GRA, and to a lesser extent in other forms of primary aldosteronism.[90] This is a consequence of chimeric gene expression in the zona fasciculata, where the same steroid molecule can be exposed to both 17-hydroxylase and aldosterone synthase activities.

GRA should be distinguished from other forms of primary aldosteronism, which are treated differently. An autosomal dominant pattern of inheritance should raise suspicion for GRA, which is diagnosed by dexamethasone suppression tests in conjunction with genotyping.

Smith–Lemli–Opitz Syndrome

Defects in sterol delta-7-reductase encoded by the *DHCR7* gene are responsible for Smith–Lemli–Optiz (SLO) syndrome. During gestation, low maternal estriol levels and decreased fetal movements may signal the diagnosis. The phenotype for this enzyme deficiency includes breech presentation, low birth weight, short stature, developmental delays, microcephaly, seizures, behavioral problems, ptosis, hypertelorism, cataracts, micrognathia, low-set ears, cleft palate, cardiac defects, hypoplastic lungs, intestinal malrotation, renal defects, skeletal defects (especially polydactyly), and varying degrees of undervirilization (46,XY DSD).[91] The biochemical diagnosis may be made by finding elevated serum 7-dehydrocholesterol (DHC) and 8-DHC with low cholesterol. These assays may be complemented by genotyping. Somewhat paradoxically, deficiency of adrenal steroid hormones is rare in SLO syndrome. The HPA axis functions normally during stress, so that corticosteroid supplementation is usually unnecessary in all but the most severe cases.[92] SLO treatment consists of a high-cholesterol diet, which may have salutary effects on behavior. Statins have also been used, sometimes combined with dietary cholesterol supplements, to reduce DHC accumulation that is thought to be neurotoxic.[93]

Other syndromes affecting adrenal development and function, such as adrenal hypoplasia congenita, IMAGE syndrome, MIRAGE syndrome, and familial glucocorticoid deficiency, are discussed in Chapter 86.

KEY POINTS

- Mineralocorticoid deficiency in the context of normal cortisol and sex steroid synthesis is most often caused by deficiency of aldosterone synthase encoded by the *CYP11B2* gene.
- The clinical manifestations of aldosterone synthase deficiency vary with age and range from hypovolemia, hyperkalemia, and dehydration in the newborn to failure to thrive in early childhood, while adults can be asymptomatic.
- Autosomal dominant mineralocorticoid excess due to glucocorticoid remediable aldosteronism is due to formation of a chimeric *CYP11B1/CYP11B2* gene that is regulated by adrenocorticotropic hormone.
- Glucocorticoid-remediable aldosteronism is characterized by early-onset hypertension with elevated aldosterone, suppressed plasma renin activity, and rapid reversal of these abnormalities following glucocorticoid administration.
- Defects in sterol delta-7-reductase encoded by the *DHCR7* gene cause Smith–Lemli–Opitz syndrome and low cholesterol with varying degrees of 46,XY undervirilization, but usually normal cortisol production.

DIFFERENTIAL DIAGNOSIS

The diagnosis of a disorder of steroid biosynthesis is generally considered in one or more of several clinical contexts. The clinician should be aware of alternative diagnoses (especially when they are equally or more common than the autosomal recessive steroidogenic defects reviewed in this chapter) and assess for them as necessary. Space does not permit a detailed discussion of differential diagnosis, and the reader is referred to the appropriate chapters for further information.

The first situation is an infant presenting with adrenal insufficiency, with hypotension, hyponatremia, and often hypoglycemia and hyperkalemia (see Chapter 93). Although CAH owing to 21-OHD is the most common genetic cause of primary adrenal insufficiency, in males X-linked adrenal hypoplasia congenital owing to NR0B1 (DAX1) mutations is at least equally common. However, the latter condition is more difficult to ascertain, because there are no elevated steroid precursors, and thus it cannot be detected by newborn screening targeting elevated 17-OHP. The clinician should always consider the possibility of an acquired cause of adrenal insufficiency such as sepsis or hemorrhage. Finally, in the absence of hyperkalemia, the possibility of secondary adrenal insufficiency should be considered, particularly in infants suffering from central nervous system insults such as head trauma or hypoxic-ischemic encephalopathy.

A second situation is the infant with atypical genitalia (Chapter 101). In a phenotypic female infant with a uterus, the most common alternative etiology to 21-OHD is mixed gonadal dysgenesis, which occurs in approximately 1:20,000 births. Phenotypic male infants with isolated hypospadias rarely have an identifiable genetic or metabolic etiology. More severely undervirilized XY males, including XY phenotypic females, most often have complete or partial androgen insensitivity, with nonvirilizing forms of congenital adrenal hyperplasia such as 17-hydroxylase deficiency occurring much less frequently.

Finally, nonvirilizing forms of CAH are often considered as etiologies for lack of development of secondary sexual characteristics or primary amenorrhea in adolescent phenotypic females (see Chapter 103). Turner syndrome is far more common than CAH in this clinical context.

KEY POINTS

- The diagnosis of a defect of adrenal steroidogenesis needs to be considered within the context of the clinical situation.
- Although congenital adrenal hyperplasia (CAH) due to 21-hydroxyase deficiency is the most common cause of primary adrenal insufficiency in children and genital atypia in the newborn, other genetic and nongenetic causes should be considered in the evaluation of these presentations.
- Nonvirilizing forms of CAH should be considered along with more common diagnoses when there is lack of development of secondary sexual characteristics or primary amenorrhea in an adolescent phenotypic female.

MEDICAL THERAPY OF CONGENITAL ADRENAL HYPERPLASIA

Glucocorticoid Therapy

The fundamental aim of therapy for disorders of steroidogenesis is to provide replacement of deficient hormones, while balancing hormone excess. Glucocorticoid therapy is the cornerstone of treatment for all forms of cortisol deficiency. The goal of glucocorticoid therapy is 2-fold: to replace deficient cortisol and diminish risk of adrenal crisis,[94] and to control ACTH-driven steroid excess.[95] Higher doses of glucocorticoid are typically needed to achieve adequate suppression of ACTH-driven androgens in the virilizing forms of CAH than doses given for cortisol replacement when all steroids are deficient. Thus, it is important to also avoid the adverse effects of glucocorticoid excess leading to short stature, obesity, osteoporosis, and an adverse metabolic profile (see later).[3,14,96] The need for glucocorticoid therapy is variable in the nonclassic forms of CAH and in POR deficiency.

Glucocorticoid in infants and children with classic CAH is usually administered as hydrocortisone 10 to 18 mg/m^2 per day divided into three doses, with lower doses (8–12 mg/m^2 per day) used for nonclassic or mild forms of CAH. Adults may receive hydrocortisone 15 to 40 mg, or rarely 60 mg/day divided into two or three doses, or receive long-acting glucocorticoids or both. It is important to use the lowest effective glucocorticoid dose to avoid the side effects of excessive dosing.

Laboratory evaluation should guide but not define management of glucocorticoid dosing. The laboratory profile in monitoring 21-OHD typically includes consistently timed blood measurements for 17-hydroxyprogesterone, androstenedione, testosterone, plasma renin activity, and electrolytes. Suggested target ranges are mildly to moderately elevated 17-hydroxyprogesterone and normal age- and sex-specific androstenedione and testosterone levels. Normal levels of 17-hydroxyprogesterone generally indicate excessive dosing of glucocorticoid. Plasma renin activity and electrolytes should be in the normal range for age. During infancy laboratory tests are performed about every 2 to 3 months, in conjunction with growth tracking and blood pressure measurements. Older children and adolescents may be monitored every 4 to 6 months, and clinically stable adults once or twice yearly. The steroid profile over time, along with blood pressure, rate of growth, and weight gain, as well as bone maturation and physical stigmata of androgen excess, are all used in medication titration.

There are several available glucocorticoid preparations. Short-acting hydrocortisone is the glucocorticoid of choice in children because of potential growth suppressive effects of longer-acting glucocorticoids.[97] The growth-suppressive effect of glucocorticoid therapy is dose-dependent and is especially observed when doses exceed 15 to 20 mg/m^2 per day during infancy and puberty.[98,99] In 2018, an immediate-release granule formulation of hydrocortisone in capsules for opening for direct oral dosing or dosing mixed with food (Infacort/Alkindi, Diurnal Limited, UK) became available in the European Union with pediatric doses (0.5, 1, 2, and 5 mg).[100] Dividing hydrocortisone 5 mg tablets is often used in the United States for pediatric dosing, but this practice is particularly problematic in the European Union and other countries, where the smallest licensed dose of hydrocortisone is 10 mg. The use of compounded hydrocortisone or liquid suspension by individual pharmacies is not recommended because of reports of variable dose accuracy in compounded preparations.[3,101,102]

Practice is highly variable in the treatment of adults. Hydrocortisone remains preferred, because it is associated with a higher bone mineral density and better metabolic and cardiovascular outcome compared with longer-acting glucocorticoid preparations.[103,104] However, due to its short half-life of 4 to 6 hours, it often needs to be given three times daily and may not adequately suppress the overnight rise of ACTH driving excess adrenal androgens in the virilizing forms of CAH and DOC in the hypertensive forms of CAH. Therefore, long-acting glucocorticoids are often used. In adults with virilizing CAH, a nighttime dose of prednisolone or prednisone can be added at bedtime.[105] A twice-daily regimen with prednisolone or prednisone (4.0–8.0 mg per day) or once- to twice-daily dose of dexamethasone (0.25–0.5 mg per day) may be used instead. Dexamethasone, in contrast to prednisolone, prednisone, or hydrocortisone, is not inactivated by placental 11β-hydroxysteroid dehydrogenase type 2. Therefore, dexamethasone should not be used in women of reproductive age who are heterosexually active, and should be avoided during pregnancy.[3] In 2011, a modified-release hydrocortisone formulation (Plenadren, Shire Services BVBA) became available in Europe. It provides immediate release in the morning and is used once daily in patients with adrenal insufficiency, with positive effects on weight, body mass index, and hemoglobin 1Ac compared with conventional hydrocortisone replacement.[106] However, Plenadren fails to control the early-morning rise in ACTH, and therefore is not used to treat disorders of steroidogenesis with ACTH-driven hormone excess such as the virilizing forms of CAH. A modified- and delayed-release preparation of hydrocortisone (Chronocort, Diurnal, UK) that mimics the circadian rhythm of cortisol release is currently in phase III clinical trials.[107,108] This formulation aims to control the overnight ACTH-driven hormone excess (see Future Directions and Novel Therapies later).

Patients with nonclassic CAH do not necessarily need glucocorticoid therapy. They are treated according to their symptoms.[3] Accordingly, children with nonclassic 21-OHD are treated if they have progressive signs and symptoms of virilization with advanced skeletal maturation. Women with nonclassic 21-OHD with signs of androgen excess can sometimes be treated with oral contraceptives or in combination with spironolactone. Glucocorticoid therapy is used for female infertility in nonclassic 21-OHD, and time to conception[109] and miscarriage rate are reduced with glucocorticoid treatment.[110,111]

Glucocorticoid therapy is also used to treat infertility in males with TARTs.[3,112] Glucocorticoid optimization and intensification can shrink TART and prevent progressive enlargement and eventual infertility; however, nonreversible fibrotic changes can occur with time. Semen cryopreservation could be offered to patients with TARTs, because irreversible testicular damage may occur.[12] Surgery can be used to treat orchialgia when large masses are present but has not been able to restore fertility.[113]

All patients with disorders of steroidogenesis who are receiving glucocorticoid therapy are at risk for increased mortality due to adrenal crisis and require glucocorticoid stress dosing for illness, trauma, and major surgery (see Adrenal Insufficiency, Chapter 93). Patients receiving glucocorticoid therapy should be educated regarding stress dosing during illness, trauma, and major surgery. Prevention of adrenal crisis is best accomplished through repeated structured patient education, and all patients should wear medical alert jewelry or have an emergency card indicating adrenal insufficiency.[114] Life-threatening adrenal crises are most often triggered by infectious illnesses[115] and are reported as the leading cause of death in patients with adrenal insufficiency due to CAH.[116,117]

Mineralocorticoid Therapy

In the salt-wasting forms of CAH, mineralocorticoid replacement in the form of fludrocortisone is given in doses ranging between 0.05 and 0.2 mg once or twice daily to achieve a plasma renin activity in the normal range. All patients with classic 21-hydroxylase have relative aldosterone insufficiency[118] and are initially given fludrocortisone, including those with the simple virilizing form. Mineralocorticoid use during childhood in patients with classic 21-OHD is associated with improved height outcome, because mineralocorticoid replacement ensures euvolemia and diminishes ACTH-driven androgen production, allowing for reduced glucocorticoid dosing.[6]

Mineralocorticoid requirements are independent of body surface area. Physiological mineralocorticoid resistance is present during infancy, requiring salt supplementation (1–2 g NaCl given in divided doses throughout the day) and higher doses of fludrocortisone. Salt supplementation may not be needed if high doses of fludrocortisone are used[119]; regardless, an increased renal sensitivity to mineralocorticoid occurs around 3 to 4 months of life requiring reduction of salt supplementation (if used) and fludrocortisone, with close monitoring to prevent hypertension. In patients with the salt-wasting forms of CAH, patients may benefit from additional salt intake during hot summer months or on days of sports activities. A seasonal increase of fludrocortisone may be warranted for extremely hot climates.[120]

Sex Steroid Replacement

Disorders of steroidogenesis that impair gonadal function (3βHSD type 2 deficiency, LCAH, cholesterol side chain cleavage deficiency, CYP17A1 deficiency, POR deficiency) require pubertal induction and sex steroid replacement. For girls, puberty is induced with low-dose estrogen given orally or transdermally. Initial doses are below those required to induce menstruation, gradually uptitrated. The addition of intermittent progesterone (as micronized progesterone for those using transdermal estrogen, or as combined oral contraceptive pills) for cyclical withdrawal bleeding is required for 46,XX females with an intact uterus. For males, short-term testosterone may be given to stimulate penile growth in childhood. Testosterone therapy is then given at the expected time of pubarche and continued during adulthood. Gonadal function and the need for sex steroid replacement are variable for the nonclassic forms of CAH that impair sex steroid synthesis.

Antihypertensive Therapy

ACTH-dependent mineralocorticoid excess occurs with CAH due to 11-OHD and CYP17A1 deficiency. Glucocorticoid treatment mitigates the effects of mineralocorticoid excess, but DOC and mineralocorticoid metabolites often persist, and additional antihypertensive therapy may be needed. Spironolactone (50–200 mg/day in one or two divided doses) blocks the mineralocorticoid receptor and is commonly used as first-line therapy. Eplerenone and calcium channel blockers can also be used.[69]

Management of Atypical Genitalia

Prenatal Therapy. Starting in the 1980s, dexamethasone was offered to pregnant women at risk of having a child with classic virilizing CAH with the aim to prevent prenatal virilization of the external genitalia in an affected 46,XX fetus.[121,122] Dexamethasone crosses the placental barrier without inactivation, reaches the fetus, and suppresses fetal adrenal androgen production. Dexamethasone treatment is effective in reducing the virilization of affected 46,XX fetuses if started early in the first trimester, but the diagnosis of CAH cannot be made until the end of the first trimester, most often with chorionic villus sampling obtained in week 10 to 11. Because CAH is autosomal recessive, the risk of having an affected girl is 1 in 8 when both parents are carriers. Therefore, with this treatment approach, seven out of eight fetuses (all boys and CAH-unaffected girls) are treated in early pregnancy with no direct benefit until prenatal testing can determine the sex and the diagnostic status, potentially creating an unfavorable risk-benefit balance. CAH-affected girls are treated until term.

Prenatal dexamethasone is no longer routinely recommended, due to unknown longer-term safety risks.[3] In follow-up studies, early prenatal dexamethasone exposure in children not affected by CAH was shown to have an effect on cognitive function, especially poor visual spatial working memory in girls.[123] Metabolic effects, such as negative effects on glucose and lipid metabolism, have been reported in young adults who received prenatal dexamethasone therapy.[124] Multiple international groups, including medical societies in the United States and Europe, have stated that prenatal dexamethasone treatment is experimental and should only be performed in centers with full disclosure of potential risks and benefits and taking part in long-term follow-up studies.[125,126]

Genital Surgery. Historically, physicians have commonly recommended genital surgery in the first year of life for atypical genitalia. In the past two decades, advocacy groups and some physicians have recommended delaying surgery so that patients can participate in decision-making. However, the psychosocial stress of not doing early surgery has not been studied. The Endocrine Society Clinical Practice Guideline and a European Expert Consensus statement recommend that parents be informed about all surgical options, including delayed surgery and observation until the child can participate in decision-making regarding surgical intervention.[3,127] Families should be educated on the known risks and benefits of having and not having surgery, as well as the complexities of gender identity and gender assignments; this often requires the support of an interdisciplinary team of experts.[128] Long-term data on outcomes in individuals who have and have not had surgery are needed to aid in decision-making.

In the virilizing forms of CAH, vaginoplasty and repair of the urogenital sinus allow for penetrative vaginal intercourse and the potential for fertility, but concerns have emerged regarding sexual function, cosmetic outcomes, and consideration of possible male gender identity, especially in fully virilized 46,XX individuals.[129] Although prenatal and postnatal exposure to androgens may influence behavior, and male-typical behavior is common among 46,XX patients with classic 21-OHD, gender dysphoria is rare, and the usual sex assignment of 46,XX patients with virilizing forms of CAH is female.[28,29] If surgery is deferred, an examination to determine whether anatomy allows for adequate menstrual flow is needed prior to puberty. Lifelong testosterone therapy would be indicated starting at puberty if there is male gender identity in a 46,XX individual with virilizing CAH. In these cases, cryopreservation of ovarian tissue should be discussed. When the diagnosis is late and the child is being raised as a boy, the option has been suggested to suppress puberty with a GnRH agonist until a choice of gender identity is made by the individual.[130] The cultural environment is an important factor in such decisions.

Surgical reconstruction of the 46,XY atypical genitalia is complex and challenging. Surgery for repair of chordee and distal hypospadias have high success rates, but proximal hypospadias repair has higher complication and reoperation rates.[131] The complications include meatal stenosis, urethral stricture, and fistula formation. If the diagnosis of a disorder of steroidogenesis is made late in a 46,XY prepubertal child raised female, puberty can be blocked until the individual's gender identity is clear. In general, gonadectomy is recommended in 46,XY patients with a female phenotype being raised as female, but the optimal timing is unknown. Although the risk of gonadal malignancy is thought to be low, early gonadal neoplastic changes were observed histologically in a 1-year-old 46,XY patient with classic LCAH.[132]

KEY POINTS

- The main goal of medical therapy is to replace deficient hormones and control adrenocorticotropic hormone-driven steroid excess.
- Short-acting hydrocortisone is the glucocorticoid of choice for disorders with cortisol deficiency, because it is associated with fewer adverse outcomes compared with long-acting glucocorticoids.
- Management of glucocorticoid therapy is based on steroid profiles as well as clinical findings of both under- and overreplacement.
- The need for glucocorticoid therapy is variable in the nonclassic forms of congenital adrenal hyperplasia and in P450 oxidoreductase deficiency.
- Both female and male patients with the hypogonadal disorders of steroidogenesis require pubertal induction and sex steroid replacement.

FUTURE DIRECTIONS AND NOVEL THERAPIES

Emerging treatments for CAH that target various aspects of steroid production are being developed (Fig. 94.5). Circadian cortisol replacement is one approach being studied. Modified- and delayed-release hydrocortisone preparations were discussed previously. Continuous subcutaneous hydrocortisone infusion mimicking cortisol circadian rhythm and delivered via an insulin pump also has shown promise, with improved disease control, improved quality-of-life, and decreased fatigue compared with similar doses administered orally.[133] Longstanding comorbidities remain to be studied with long-term use of physiologic cortisol replacement.

Strategies for reducing ACTH are being investigated as a way to control the ACTH-driven excess androgen production and ultimately allow for glucocorticoid dose reduction. A corticotropin-releasing hormone type I receptor antagonist was studied in a single-dose phase I study and achieved delayed and attenuated morning rise in ACTH and 17- hydroxyprogesterone in six of eight females with classic 21-OHD.[134] Two additional agents of this drug class are currently being studied in clinical trials in adults with classic 21-OHD (NCT03257462; NCT03687242; NCT04045145). An ACTH receptor (MC2R) antagonist[135] and a long-acting neutralizing monoclonal ACTH antibody have been tested in animal models[136] as ways to attenuate ACTH-dependent steroidogenesis.

Adrenal sex hormone-lowering agents and antagonists are being studied in prepubertal children and women receiving sex hormone replacement therapy as a way to allow for lower-dose glucocorticoid treatment in the virilizing forms of CAH. The use of an antiandrogen and aromatase inhibitor in combination with lower-dose hydrocortisone and fludrocortisone was successful in normalizing growth over 2 years in 28 prepubertal children with classic 21-OHD.[137] Abiraterone acetate, a potent inhibitor of CYP17A1, lowered androstenedione concentrations in five of six women after 6 days of treatment.[138] A study of abiraterone acetate in prepubertal children with CAH has been initiated (NCT02574910). A case report describes the use of mitotane, an adrenal enzyme inhibitor with adrenolytic properties, in the treatment of TARTs, with restoration of fertility, although potential toxic side effects limit the use of mitotane for this indication.[139] Nevanimibe (ATR-101), another inhibitor of adrenal enzyme production with potential adrenolytic properties, decreased 17-hydroxyprogesterone levels in patients with classic 21-OHD within 2 weeks of treatment in a dose-escalation study.[140] Further study of longer duration and with higher doses is needed to determine safety and efficacy.

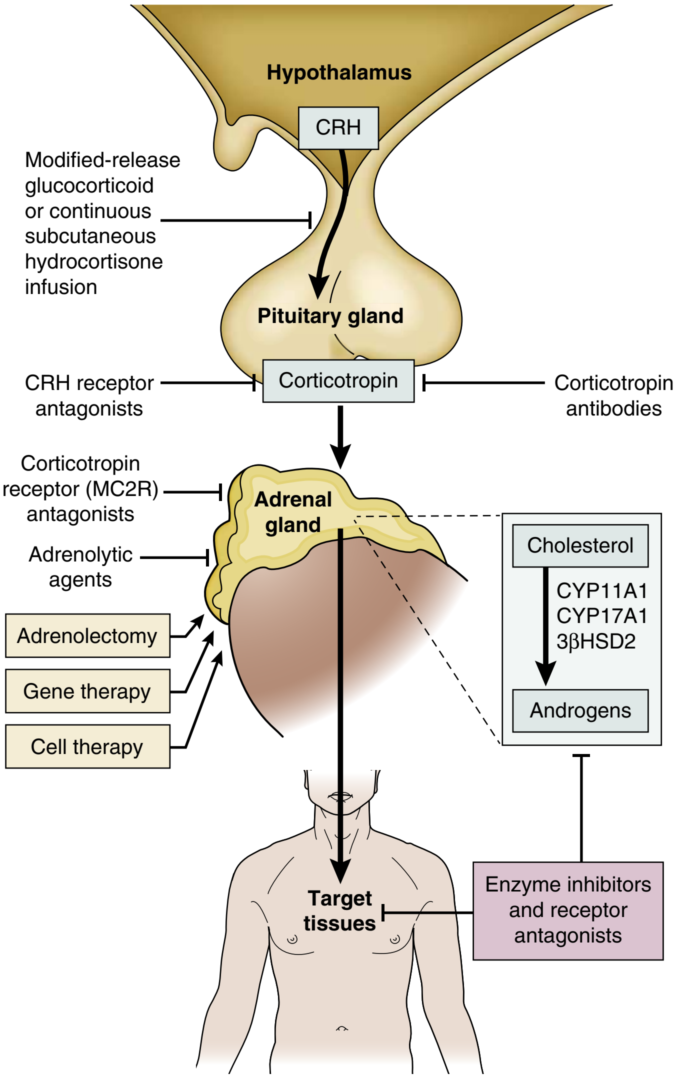


Fig. 94.5 New therapies for congenital adrenal hyperplasia. New therapeutic approaches target different aspects of the pathophysiology of congenital adrenal hyperplasia (CAH). Circadian cortisol replacement through the use of a modified-release glucocorticoid or subcutaneous hydrocortisone infusion aim to control corticotropin-driven hyperandrogenism by replacing cortisol in a physiological manner. Other approaches to reduce androgen production without chronic supraphysiological glucocorticoid exposure include corticotropin-releasing hormone receptor-1 antagonists, adrenocorticotropic hormone (corticotropin; ACTH) antibodies, adrenocorticotropic hormone receptor (MC2R) antagonists, adrenolytic agents, adrenalectomy, and pharmacological inhibition of steroidogenic enzymes or steroid receptors in the adrenal or peripheral tissues. Because CAH is a monogenic disorder, gene therapy with cell-based and gene-editing technologies may be able to restore defective steroidogenesis. CRH denotes corticotropin-releasing hormone (sometimes referred to as corticotropin-releasing factor [CRF]). (From Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *NEJM.* 2020;83:1258. Reprinted with permission.)

Isolated and uncontrolled proof-of-concept trials have used growth hormone, with or without GnRHa,[141] or, in two patients with classic 11-OHD CAH, an aromatase inhibitor.[142,143] In the absence of more detailed studies, routine use of growth hormone or aromatase inhibitors is also not recommended.

Surgical adrenalectomy has been used to restore fertility in women with classic CAH and for select patients with excessive virilization who cannot be successfully managed with conventional medical therapy without concomitant iatrogenic Cushing syndrome. In a metaanalysis of 48 patients with CAH who had undergone bilateral adrenalectomy, the majority of patients reported symptomatic improvement postoperatively with successful reduction of both glucocorticoid dose and androgen levels, but adverse outcomes included adrenal crisis and the development of adrenal rest tumors.[144]

Disorders of steroidogenesis are monogenic diseases that may be amenable to gene therapy. With advancements in cell-based and gene editing technologies and the increase in safety of gene therapy using recombinant adeno-associated virus vectors, the development of gene therapies for patients with CAH are underway in preclinical programs but will be of lower priority for clinical implementation than treatments for diseases (such as adrenoleukodystrophy) for which no satisfactory treatment exists. An adeno-associated virus CYP21 vector restored 21-hydroxylase activity in CAH mice for more than 15 weeks.[145] In another approach, regenerative therapies are being studied including the transformation of pluripotent stem cells to steroidogenic cells and the use of immune-isolating biomaterials for adrenal transplantation.[146]

KEY POINTS

- Novel therapies being developed for congenital adrenal hyperplasia target various aspects of steroid production and include mimicking the circadian rhythm of cortisol secretion, reducing adrenocorticotropic hormone, administering adrenal sex hormone-lowering agents and antagonists, and applying gene-modifying therapies.

CONCLUSION

Disorders of steroidogenesis are a group of rare diseases with potential mortality and high morbidity if left undiagnosed and untreated. Advances in our understanding of the genetics and pathophysiology of this spectrum of disorders has improved diagnosis and disease management, but additional advances are needed to improve outcomes. The overarching goal of therapy is to provide replacement of deficient hormones, while balancing hormone excess and reducing comorbidities. Numerous novel approaches aim to improve patient outcomes for these rare genetic diseases of adrenal steroidogenesis.

REFERENCES

1. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32:81–151.
2. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*. 2000;21:245–291.
3. Speiser PW, Arlt W, Auchus RJ, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:4043–4088.
4. van der Straaten S, Springer A, Zecic A, et al. The external genitalia Score (EGS): a European multicenter Validation study. *J Clin Endocrinol Metab*. 2020;105:e222–e230.
5. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *N Engl J Med*. 2020;383:1248–1261.
6. Muthusamy K, Elamin MB, Smushkin G, et al. Clinical review: adult height in patients with congenital adrenal hyperplasia: a systematic review and metaanalysis. *J Clin Endocrinol Metab*. 2010;95:4161–4172.
7. Manoli I, Kanaka-Gantenbein C, Voutetakis A, et al. Early growth, pubertal development, body mass index and final height of patients with congenital adrenal hyperplasia: factors influencing the outcome. *Clin Endocrinol*. 2002;57:669–676.
8. Völkl TM, Öhl L, Rauh M, et al. Adrenarche and puberty in children with classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Horm Res Paediatr*. 2011;76:400–410.
9. Gomes LG, Bachega TA, Mendonca BB. Classic congenital adrenal hyperplasia and its impact on reproduction. *Fertil Steril*. 2019;111:7–12.
10. Mendes-Dos-Santos CT, Martins DL, Guerra-Júnior G, et al. Prevalence of testicular adrenal rest tumor and factors associated with its development in congenital adrenal hyperplasia. *Horm Res Paediatr*. 2018;90:161–168.
11. Crocker MK, Barak S, Millo CM, et al. Use of PET/CT with cosyntropin stimulation to identify and localize adrenal rest tissue following adrenalectomy in a woman with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2012;97:E2084–E2089.
12. Engels M, Span PN, van Herwaarden AE, et al. Testicular adrenal rest tumors: current insights on prevalence, characteristics, origin, and treatment. *Endocr Rev*. 2019;40:973–987.
13. Nguyen LS, Prifti E, Ichou F, et al. Effect of congenital adrenal hyperplasia treated by glucocorticoids on plasma metabolome: a machine-learning-based analysis. *Sci Rep*. 2020;10:8859.
14. Tamhane S, Rodriguez-Gutierrez R, Iqbal AM, et al. Cardiovascular and metabolic outcomes in congenital adrenal hyperplasia: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2018;103:4097–4103.
15. Falhammar H, Frisen L, Hirschberg AL, et al. Increased cardiovascular and metabolic morbidity in patients with 21-hydroxylase deficiency: a Swedish population-based national cohort study. *J Clin Endocrinol Metab*. 2015;100:3520–3528.
16. Riehl G, Reisch N, Roehle R, et al. Bone mineral density and fractures in congenital adrenal hyperplasia: findings from the dsd-LIFE study. *Clin Endocrinol*. 2020;92:284–294.
17. Finkielstain GP, Kim MS, Sinaii N, et al. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2012;97:4429–4438.
18. Falhammar H, Filipsson Nystrom H, Wedell A, et al. Bone mineral density, bone markers, and fractures in adult males with congenital adrenal hyperplasia. *Eur J Endocrinol*. 2013;168:331–341.
19. Rangaswamaiah S, Gangathimmaiah V, Nordenstrom A, et al. Bone mineral density in adults with congenital adrenal hyperplasia: a systematic review and meta-analysis. *Front Endocrinol*. 2020;11. https://doi.org/10.3389/fendo.2020.00493.
20. Karlsson L, Gezelius A, Nordenstrom A, et al. Cognitive impairment in adolescents and adults with congenital adrenal hyperplasia. *Clin Endocrinol*. 2017;87:651–659.
21. Berenbaum SA, Beltz AM, Bryk K, et al. Gendered peer involvement in girls with congenital adrenal hyperplasia: effects of prenatal androgens, gendered activities, and gender cognitions. *Arch Sex Behav*. 2018;47:915–929.
22. Endendijk JJ, Beltz AM, McHale SM, et al. Linking prenatal androgens to gender-related attitudes, identity, and activities: evidence from girls with congenital adrenal hyperplasia. *Arch Sex Behav*. 2016;45:1807–1815.
23. Messina V, Hirvikoski T, Karlsson L, et al. Good overall behavioural adjustment in children and adolescents with classic congenital adrenal hyperplasia. *Endocrine*. 2020;68:427–437.
24. Engberg H, Butwicka A, Nordenstrom A, et al. Congenital adrenal hyperplasia and risk for psychiatric disorders in girls and women born between 1915 and 2010: a total population study. *Psychoneuroendocrinology*. 2015;60:195–205.
25. Falhammar H, Butwicka A, Landen M, et al. Increased psychiatric morbidity in men with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2014;99:E554–E560.

26. Van't Westeinde A, Karlsson L, Thomsen Sandberg M, et al. Altered gray matter structure and white matter microstructure in patients with congenital adrenal hyperplasia: relevance for working memory performance. *Cereb Cortex*. 2020;30:2777–2788.
27. Van't Westeinde A, Karlsson L, Nordenström A, et al. First trimester prenatal dexamethasone treatment is associated with alterations in brain structure at adult age. *J Clin Endocrinol Metab*. 2020;105:2575–2586.
28. Meyer-Bahlburg HF, Dolezal C, Baker SW, et al. Sexual orientation in women with classical or non-classical congenital adrenal hyperplasia as a function of degree of prenatal androgen excess. *Arch Sex Behav*. 2008;37:85–99.
29. Frisen L, Nordenstrom A, Falhammar H, et al. Gender role behavior, sexuality, and psychosocial adaptation in women with congenital adrenal hyperplasia due to CYP21A2 deficiency. *J Clin Endocrinol Metab*. 2009;94:3432–3439.
30. Kamenický P, Blanchard A, Lamaziere A, et al. Cortisol and aldosterone responses to hypoglycemia and Na depletion in women with non-classic 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2020;105:55–64.
31. Hannah-Shmouni F, Morissette R, Sinaii N, et al. Revisiting the prevalence of nonclassic congenital adrenal hyperplasia in US Ashkenazi Jews and Caucasians. *Genet Med*. 2017;19:1276–1279.
32. Falhammar H, Nordenstrom A. Nonclassic congenital adrenal hyperplasia due to 21-hydroxylase deficiency: clinical presentation, diagnosis, treatment, and outcome. *Endocrine*. 2015;50:32–50.
33. Engels M, Gehrmann K, Falhammar H, et al. Gonadal function in adult male patients with congenital adrenal hyperplasia. *Eur J Endocrinol*. 2018;178:285–294.
34. Pinkas H, Fuchs S, Klipper-Aurbach Y, et al. Non-classical 21-hydroxylase deficiency: prevalence in males with unexplained abnormal sperm analysis. *Fertil Steril*. 2010;93:1887–1891.
35. Falhammar H, Frisén L, Norrby C, et al. Reduced frequency of biological and increased frequency of adopted children in males with 21-hydroxylase deficiency: a Swedish population-based national cohort study. *J Clin Endocrinol Metab*. 2017;102:4191–4199.
36. Miller WL, Merke DP. Tenascin-X, congenital adrenal hyperplasia, and the CAH-X syndrome. *Horm Res Paediatr*. 2018;89:352–361.
37. Lao Q, Brookner B, Merke DP. High-throughput screening for CYP21A1P-TNXA/TNXB chimeric genes responsible for Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. *J Mol Diagn*. 2019;21:924–931.
38. Gao Y, Lu L, Yu B, et al. The prevalence of the chimeric TNXA/TNXB gene and clinical symptoms of Ehlers-Danlos syndrome with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2020;105:2288–2299.
39. Wang C, Pallan PS, Zhang W, et al. Functional analysis of human cytochrome P450 21A2 variants involved in congenital adrenal hyperplasia. *J Biol Chem*. 2017;292:10767–10778.
40. New MI, Abraham M, Gonzalez B, et al. Genotype-phenotype correlation in 1,507 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Proc Natl Acad Sci USA*. 2013;110:2611–2616.
41. Simonetti L, Bruque CD, Fernandez CS, et al. CYP21A2 mutation update: Comprehensive analysis of databases and published genetic variants. *Hum Mutat*. 2017;39:5–22.
42. Speiser PW, Dupont J, Zhu D, et al. Disease expression and molecular genotype in congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Invest*. 1992;90:584–595.
43. Krone N, Braun A, Roscher AA, et al. Predicting phenotype in steroid 21-hydroxylase deficiency? Comprehensive genotyping in 155 unrelated, well defined patients from southern Germany. *J Clin Endocrinol Metab*. 2000;85:1059–1065.
44. Miller WL. Congenital adrenal hyperplasia: time to replace 17OHP with 21-deoxycortisol. *Horm Res Paediatr*. 2019;91:416–420.
45. Martinerie L, Pussard E, Foix-L'Hélias L, et al. Physiological partial aldosterone resistance in human newborns. *Pediatr Res*. 2009;66:323–328.
46. Nabhan ZM, Eugster EA. Upper-tract genitourinary malformations in girls with congenital adrenal hyperplasia. *Pediatrics*. 2007;120:e304–e307.
47. White PC. Congenital adrenal hyperplasia owing to 11beta-hydroxylase deficiency. *Adv Exp Med Biol*. 2011;707:7–8.
48. White PC, Curnow KM, Pascoe L. Disorders of steroid 11 beta-hydroxylase isozymes. *Endocr Rev*. 1994;15:421–438.
49. Rosler A, Leiberman E, Cohen T. High frequency of congenital adrenal hyperplasia (classic 11 beta-hydroxylase deficiency) among Jews from Morocco. *Am J Med Genet*. 1992;42:827–834.
50. Kandemir N, Yordam N. Congenital adrenal hyperplasia in Turkey: a review of 273 patients. *Acta Paediatr*. 1997;86:22–25.
51. Joehrer K, Geley S, Strasser-Wozak EM, et al. CYP11B1 mutations causing non-classic adrenal hyperplasia due to 11 beta-hydroxylase deficiency. *Hum Mol Genet*. 1997;6:1829–1834.
52. Curnow KM, Tusie-Luna MT, Pascoe L, et al. The product of the CYP11B2 gene is required for aldosterone biosynthesis in the human adrenal cortex. *Mol Endocrinol*. 1991;5:1513–1522.
53. Khattab A, Haider S, Kumar A, et al. Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11β-hydroxylase deficiency. *Proc Natl Acad Sci USA*. 2017;114:E1933–E1940.
54. White PC, Dupont J, New MI, et al. A mutation in CYP11B1 (Arg-448---His) associated with steroid 11 beta-hydroxylase deficiency in Jews of Moroccan origin. *J Clin Invest*. 1991;87:1664–1667.
55. White PC. Steroid 11B-hydroxylase deficiency and related disorders. In: New MI, Lekarev O, Parsa A, et al, eds. Genetic steriod disorders. London: Elsevier; 2014:71–85.
56. Krone N, Reisch N, Idkowiak J, et al. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2012;97:E257–E267.
57. Dean B, Chrisp GL, Quartararo M, et al. P450 oxidoreductase deficiency: a systematic review and meta-analysis of genotypes, phenotypes, and their relationships. *J Clin Endocrinol Metab*. 2020:105:e42–e52.
58. Miller WL, Auchus RJ. The "backdoor pathway" of androgen synthesis in human male sexual development. *PLoS Biol*. 2019;17:e3000198.
59. Reisch N, Idkowiak J, Hughes BA, et al. Prenatal diagnosis of congenital adrenal hyperplasia caused by P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2013;98:E528–E536.
60. Fukami M, Nishimura G, Homma K, et al. Cytochrome P450 oxidoreductase deficiency: identification and characterization of biallelic mutations and genotype-phenotype correlations in 35 Japanese patients. *J Clin Endocrinol Metab*. 2009;94:1723–1731.
61. Storbeck KH, Schiffer L, Baranowski ES, et al. Steroid metabolome analysis in disorders of adrenal steroid biosynthesis and metabolism. *Endocr Rev*. 2019;40:1605–1625.
62. Simard J, Ricketts ML, Gingras S, et al. Molecular biology of the 3beta-hydroxysteroid dehydrogenase/delta5-delta4 isomerase gene family. *Endocr Rev*. 2005;26:525–582.
63. Jeandron DD, Sahakitrungruang T. A novel homozygous Q334X mutation in the HSD3B2 gene causing classic 3β-hydroxysteroid dehydrogenase deficiency: an unexpected diagnosis after a positive newborn screen for 21-hydroxylase deficiency. *Horm Res Paediatr*. 2012;77:334–338.
64. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocr Rev*. 2004;25:947–970.
65. Miller WL. Mechanisms in Endocrinology: rare defects in adrenal steroidogenesis. *Eur J Endocrinol*. 2018;179:R125–R141.
66. Lutfallah C, Wang W, Mason JI, et al. Newly proposed hormonal criteria via genotypic proof for type II 3beta- hydroxysteroid dehydrogenase deficiency. *J Clin Endocrinol Metab*. 2002;87:2611–2622.
67. Mermejo LM, Elias LL, Marui S, et al. Refining hormonal diagnosis of type II 3beta-hydroxysteroid dehydrogenase deficiency in patients with premature pubarche and hirsutism based on HSD3B2 genotyping. *J Clin Endocrinol Metab*. 2005;90:1287–1293.
68. Al Alawi AM, Nordenstrom A, Falhammar H. Clinical perspectives in congenital adrenal hyperplasia due to 3beta-hydroxysteroid dehydrogenase type 2 deficiency. *Endocrine*. 2019;63:407–421.
69. Auchus RJ. Steroid 17-hydroxylase and 17,20-lyase deficiencies, genetic and pharmacologic. *J Steroid Biochem Mol Biol*. 2017;165:71–78.
70. Paris F, Gaspari L, Mbou F, et al. Endocrine and molecular investigations in a cohort of 25 adolescent males with prominent/persistent pubertal gynecomastia. *Andrology*. 2016;4:263–269.

71. Neres MS, Auchus RJ, Shackleton CH, et al. Distinctive profile of the 17-hydroxylase and 17,20-lyase activities revealed by urinary steroid metabolomes of patients with CYP17 deficiency. *Arq Bras Endocrinol Metabol*. 2010;54:826–832.
72. Miller WL. Disorders in the initial steps of steroid hormone synthesis. *J Steroid Biochem Mol Biol*. 2017;165:18–37.
73. Zhao X, Su Z, Liu X, et al. Long-term follow-up in a Chinese child with congenital lipoid adrenal hyperplasia due to a StAR gene mutation. *BMC Endocr Disord*. 2018;18:78.
74. Stocco DM. StAR protein and the regulation of steroid hormone biosynthesis. *Annu Rev Physiol*. 2001;63:193–213.
75. Bose HS, Lingappa VR, Miller WL. The steroidogenic acute regulatory protein, StAR, works only at the outer mitochondrial membrane. *Endocr Res*. 2002;28:295–308.
76. Baker BY, Lin L, Kim CJ, et al. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*. 2006;91:4781–4785.
77. Kolli V, Kim H, Torky A, et al. Characterization of the CYP11A1 nonsynonymous variant p.E314K in children presenting with adrenal insufficiency. *J Clin Endocrinol Metab*. 2019;104:269–276.
78. Ulick S, Wang JZ, Morton DH. The biochemical phenotypes of two inborn errors in the biosynthesis of aldosterone. *J Clin Endocrinol Metab*. 1992;74:1415–1420.
79. Lages AS, Vale B, Oliveira P, et al. Congenital hyperreninemic hypoaldosteronism due to aldosterone synthase deficiency type I in a Portuguese patient - case report and review of literature. *Arch Endocrinol Metab*. 2019;63:84–88.
80. Pascoe L, Curnow KM, Slutsker L, et al. Mutations in the human CYP11B2 (aldosterone synthase) gene causing corticosterone methyloxidase II deficiency. *Proc Natl Acad Sci USA*. 1992;89:4996–5000.
81. Peter M, Viemann M, Partsch CJ, et al. Congenital adrenal hypoplasia: clinical spectrum, experience with hormonal diagnosis, and report on new point mutations of the DAX-1 gene. *J Clin Endocrinol Metab*. 1998;83:2666–2674.
82. Iughetti L, Lucaccioni L, Bruzzi P, et al. Isolated hypoaldosteronism as first sign of X-linked adrenal hypoplasia congenita caused by a novel mutation in NR0B1/DAX-1 gene: a case report. *BMC Med Genet*. 2019; 20(98):1–8.
83. Kayes-Wandover KM, Tannin GM, Shulman D, et al. Congenital hyperreninemic hypoaldosteronism unlinked to the aldosterone synthase (CYP11B2) gene. *Journal of Clinical Endocrinology Metabolism*. 2001;86:5379–5382.
84. Zennaro MC, Jeunemaitre X. SFE/SFHTA/AFCE consensus on primary aldosteronism, part 5: genetic diagnosis of primary aldosteronism. *Ann Endocrinol*. 2016;77:214–219.
85. Lenzini L, Prisco S, Caroccia B, et al. Saga of familial hyperaldosteronism: yet a new channel. *Hypertension*. 2018;71:1010–1014.
86. Gates LJ, MacConnachie AA, Lifton RP, et al. Variation of phenotype in patients with glucocorticoid remediable aldosteronism. *J Med Genet*. 1996;33:25–28.
87. Lifton RP, Dluhy RG, Powers M, et al. A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*. 1992;355:262–265.
88. Pascoe L, Curnow KM, Slutsker L, et al. Glucocorticoid-suppressible hyperaldosteronism results from hybrid genes created by unequal crossovers between CYP11B1 and CYP11B2. *Proc Natl Acad Sci USA*. 1992;89:8327–8331.
89. Stowasser M, Ahmed AH, Pimenta E, et al. Factors affecting the aldosterone/renin ratio. *Horm Metab Res*. 2012;44:170–176.
90. Dluhy RG, Lifton RP. Glucocorticoid-remediable aldosteronism. [Review]. *Endocrinol Metab Clin North Am*. 1994;23:285–297.
91. Bianconi SE, Cross JL, Wassif CA, et al. Pathogenesis, epidemiology, diagnosis and clinical aspects of Smith-Lemli-Opitz syndrome. *Expert Opin Orphan Drugs*. 2015;3:267–280.
92. Bianconi SE, Conley SK, Keil MF, et al. Adrenal function in Smith-Lemli-Opitz syndrome. *Am J Med Genet*. 2011;155A:2732–2738.
93. Svoboda MD, Christie JM, Eroglu Y, et al. Treatment of Smith-Lemli-Opitz syndrome and other sterol disorders. *Am J Med Genet C Semin Med Genet*. 2012;160C:285–294.
94. Rushworth RL, Torpy DJ, Falhammar H. Adrenal crisis. *N Engl J Med*. 2019;381:852–861.
95. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390:2194–2210.
96. Han TS, Walker BR, Arlt W, et al. Treatment and health outcomes in adults with congenital adrenal hyperplasia. *Nat Rev Endocrinol*. 2014;10:115–124.
97. Bonfig W, Bechtold S, Schmidt H, et al. Reduced final height outcome in congenital adrenal hyperplasia under prednisone treatment: deceleration of growth velocity during puberty. *J Clin Endocrinol Metab*. 2007;92:1635–1639.
98. Bonfig W, Pozza SB, Schmidt H, et al. Hydrocortisone dosing during puberty in patients with classical congenital adrenal hyperplasia: an evidence-based recommendation. *J Clin Endocrinol Metab*. 2009;94:3882–3888.
99. Balsamo A, Cicognani A, Baldazzi L, et al. CYP21 genotype, adult height, and pubertal development in 55 patients treated for 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2003;88:5680–5688.
100. Porter J, Withe M, Ross RJ. Immediate-release granule formulation of hydrocortisone, Alkindi(R), for treatment of paediatric adrenal insufficiency (Infacort development programme). *Expert Rev Endocrinol Metab*. 2018;13:119–124.
101. Neumann U, Burau D, Spielmann S, et al. Quality of compounded hydrocortisone capsules used in the treatment of children. *Eur J Endocrinol*. 2017;177:239–242.
102. Barillas JE, Eichner D, Van Wagoner R, et al. Iatrogenic Cushing syndrome in a child with congenital adrenal hyperplasia: erroneous compounding of hydrocortisone. *J Clin Endocrinol Metab*. 2018;103:7–11.
103. Paizoni L, Auer MK, Schmidt H, et al. Effect of androgen excess and glucocorticoid exposure on metabolic risk profiles in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Steroid Biochem Mol Biol*. 2020;197:105540.
104. Whittle E, Falhammar H. Glucocorticoid regimens in the treatment of congenital adrenal hyperplasia: a systematic review and meta-analysis. *J Endocr Soc*. 2019;3:1227–1245.
105. Auchus RJ. Management considerations for the adult with congenital adrenal hyperplasia. *Mol Cell Endocrinol*. 2015;408:190–197.
106. Johannsson G, Nilsson AG, Bergthorsdottir R, et al. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab*. 2012;97:473–481.
107. Mallappa A, Sinaii N, Kumar P, et al. A phase 2 study of Chronocort, a modified-release formulation of hydrocortisone, in the treatment of adults with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2015;100:1137–1145.
108. Debono M, Ghobadi C, Rostami-Hodjegan A, et al. Modified-release hydrocortisone to provide circadian cortisol profiles. *J Clin Endocrinol Metab*. 2009;94:1548–1554.
109. Eyal O, Ayalon-Dangur I, Segev-Becker A, et al. Pregnancy in women with non-classic congenital adrenal hyperplasia: time to conceive and outcome. *Clin Endocrinol (Oxf)*. 2017;87:552–556.
110. Moran C, Azziz R, Weintrob N, et al. Reproductive outcome of women with 21-hydroxylase-deficient nonclassic adrenal hyperplasia. *J Clin Endocrinol Metab*. 2006;91:3451–3456.
111. Bidet M, Bellanne-Chantelot C, Galand-Portier MB, et al. Fertility in women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2010;95:1182–1190.
112. Reisch N, Flade L, Scherr M, et al. High prevalence of reduced fecundity in men with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2009;94:1665–1670.
113. Claahsen-van der Grinten HL, Otten BJ, Hermus AR, et al. Testicular adrenal rest tumors in patients with congenital adrenal hyperplasia can cause severe testicular damage. *Fertil Steril*. 2008;89:597–601.

114. Repping-Wuts HJ, Stikkelbroeck NM, Noordzij A, et al. A glucocorticoid education group meeting: an effective strategy for improving self-management to prevent adrenal crisis. *Eur J Endocrinol*. 2013;169: 17–22.
115. Yang M, White PC. Risk factors for hospitalization of children with congenital adrenal hyperplasia. *Clin Endocrinol*. 2017;86:669–673.
116. Falhammar H, Frisen L, Norrby C, et al. Increased mortality in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2014;99:E2715–E2721.
117. Ali SR, Bryce J, Haghpanahan H, et al. Real world estimates of adrenal insufficiency related adverse events in children with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2020;106:e192–e203.
118. Frisch H, Battelino T, Schober E, et al. Salt wasting in simple virilizing congenital adrenal hyperplasia. *J Pediatr Endocrinol Metab*. 2001;14:1649–1655.
119. Bonfig W, Roehl F, Riedl S, et al. Sodium chloride supplementation is not routinely performed in the majority of German and Austrian infants with classic salt-wasting congenital adrenal hyperplasia and has no effect on linear growth and hydrocortisone or fludrocortisone dose. *Horm Res Paediatr*. 2018;89:7–12.
120. Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:364–389.
121. Forest MG, David M, Morel Y. Prenatal diagnosis and treatment of 21-hydroxylase deficiency. *J Steroid Biochem Mol Biol*. 1993;45:75–82.
122. Carlson AD, Obeid JS, Kanellopoulou N, et al. Congenital adrenal hyperplasia: update on prenatal diagnosis and treatment. *J Steroid Biochem Mol Biol*. 1999;69:19–29.
123. Wallensteen L, Zimmermann M, Thomsen SM, et al. Sex-dimorphic effects of prenatal treatment with dexamethasone. *J Clin Endocrinol Metab*. 2016;101:3838–3846.
124. Wallensteen L, Karlsson L, Messina V, et al. Perturbed beta-cell function and lipid profile after early prenatal dexamethasone exposure in individuals without CAH. *J Clin Endocrinol Metab*. 2020;105:e2439–e2448.
125. Joint LWPES/ESPE CAH Working Group. Consensus statement on 21-hydroxylase deficiency from the Lawson Wilkins pediatric endocrine society and the European society for paediatric Endocrinology. *J Clin Endocrinol Metab*. 2002;87:4048–4053.
126. Hirvikoski T, Nordenstrom A, Wedell A, et al. Prenatal dexamethasone treatment of children at risk for congenital adrenal hyperplasia: the Swedish experience and standpoint. *J Clin Endocrinol Metab*. 2012;97:1881–1883.
127. Cools M, Nordenstrom A, Robeva R, et al. Caring for individuals with a difference of sex development (DSD): a Consensus Statement. *Nat Rev Endocrinol*. 2018;14:415–429.
128. Fisher AD, Ristori J, Fanni E, et al. Gender identity, gender assignment and reassignment in individuals with disorders of sex development: a major of dilemma. *J Endocrinol Invest*. 2016;39:1207–1224.
129. Mieszczak J, Houk CP, Lee PA. Assignment of the sex of rearing in the neonate with a disorder of sex development. *Curr Opin Pediatr*. 2009;21:541–547.
130. Mouriquand PD, Gorduza DB, Gay CL, et al. Surgery in disorders of sex development (DSD) with a gender issue: if (why), when, and how? *J Pediatr Urol*. 2016;12:139–149.
131. DiSandro M, Merke DP, Rink RC. Review of current surgical techniques and medical management considerations in the treatment of pediatric patients with disorders of sex development. *Horm Metab Res*. 2015;47:321–328.
132. Abdulhadi-Atwan M, Jean A, Chung WK, et al. Role of a founder c.201_202delCT mutation and new phenotypic features of congenital lipoid adrenal hyperplasia in Palestinians. *J Clin Endocrinol Metab*. 2007;92:4000–4008.
133. Mallappa A, Nella AA, Sinaii N, et al. Long-term use of continuous subcutaneous hydrocortisone infusion therapy in patients with congenital adrenal hyperplasia. *Clin Endocrinol*. 2018;89:399–407.
134. Turcu AF, Spencer-Segal JL, Farber RH, et al. Single-dose study of a corticotropin-releasing factor receptor-1 antagonist in women with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2016;101:1174–1180.
135. Sanders K, Mol JA, Kooistra HS, et al. Melanocortin 2 receptor antagonists in canine pituitary-dependent hypercortisolism: in vitro studies. *Vet Res Commun*. 2018;42:283–288.
136. Gehrand AL, Phillips J, Malott K, et al. A long-acting neutralizing monoclonal ACTH antibody blocks corticosterone and adrenal gene responses in neonatal rats. *Endocrinology*. 2019;160:1719–1730.
137. Merke DP, Keil MF, Jones JV, et al. Flutamide, testolactone, and reduced hydrocortisone dose maintain normal growth velocity and bone maturation despite elevated androgen levels in children with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2000;85:1114–1120.
138. Auchus RJ, Buschur EO, Chang AY, et al. Abiraterone acetate to lower androgens in women with classic 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2014. jc20141258.
139. Bry-Gauillard H, Cartes A, Young J. Mitotane for 21-hydroxylase deficiency in an infertile man. *N Engl J Med*. 2014;371:2042–2044.
140. El-Maouche D, Merke DP, Vogiatzi MG, et al. A phase 2, multicenter study of nevanimibe for the treatment of congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2020;105:2771–2778.
141. Lin-Su K, Harbison MD, Lekarev O, Vogiatzi MG, New MI. Final adult height in children with congenital adrenal hyperplasia treated with growth hormone. *J Clin Endocrinol Metab*. 2011;96:1710–1717.
142. Nour MA, Pacaud D. Height augmentation in 11β-hydroxylase deficiency congenital adrenal hyperplasia. *Int J Pediatr Endocrinol*. 2015;2015:12.
143. Hawton K, Walton-Betancourth S, Rumsby G, et al. Growth hormone with aromatase inhibitor may improve height in CYP11B1 congenital adrenal hyperplasia. *Pediatrics*. 2017;139(2):e20160730.
144. MacKay D, Nordenstrom A, Falhammar H. Bilateral adrenalectomy in congenital adrenal hyperplasia: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2018;103:1767–1778.
145. Perdomini M, Dos SC, Goumeaux C, et al. An AAVrh10-CAG-CYP21-HA vector allows persistent correction of 21-hydroxylase deficiency in a Cyp21-/- mouse model. *Gene Ther*. 2017;24:275–281.
146. Bornstein SR, Malyukov M, Heller C, et al. New horizons: novel adrenal regenerative therapies. *J Clin Endocrinol Metab*. 2020:105;3103–3107.

95

Adrenal Genomics II: Familial and Sporadic Neoplasia

Dipika R. Mohan, Guillaume Assie, and Antonio M. Lerario

OUTLINE

Introduction, 1590
Genetic Causes of Adrenal Neoplasia, 1590
Adrenocortical Tumors, 1590
Pheochromocytoma, 1594
Summary and Future Directions, 1597

INTRODUCTION

KEY POINTS

- Hereditary and sporadic neoplasms of the adrenal cortex and medulla are characterized by a myriad of driver genetic events that alter flux through a few core cell signaling and metabolic pathways.

Primary adrenal tumors are a phenotypically heterogeneous group of diseases arising from distinct cell populations of the adrenal cortex or medulla. Disease heterogeneity also emerges from a variety of additional features, including age of onset, malignant potential, functional status (hormonally silent or active), and genetic context (sporadic or inherited). As a group, adrenal neoplasia is common in older individuals (>60 years), in whom the prevalence reaches up to 9%.[1] Most neoplasms are benign adrenocortical tumors like adrenocortical adenomas (ACA); exceedingly rare forms of adrenal neoplasia include malignant adrenocortical carcinomas (ACC), adrenal medullary pheochromocytomas (Pheo), and extraadrenal paragangliomas (PGL; often associated with inherited/syndromic Pheo).[1] Benign adrenal neoplasia also encompasses rare forms of diffuse and multinodular bilateral hyperplasia.[1] Hormonally active adrenocortical tumors can produce any class of steroid hormones or precursors, including glucocorticoids, androgens, mineralocorticoids, and estrogens. While ACA and hyperplasia usually produce a single class of steroid hormones (typically aldosterone or cortisol), ACC commonly secrete several types of steroids and precursors.[2,3] Hormonally active Pheo is also heterogeneous; epinephrine production is an exclusive feature of adrenal Pheo, norepinephrine production is more common in PGL, and elevated dopamine and production of its metabolite methoxytyramine is a feature of malignant Pheo.[4,5] In terms of the genetic context, unilateral and sporadic forms of ACA, ACC, and Pheo are more common.[6-8] However, pediatric/early-onset, bilateral, and multifocal tumors occur and are clinically suggestive of inherited forms.[6,7]

Consistent with the extraordinary clinical heterogeneity of adrenal neoplasia, the genetics of these diseases is complex. To date, approximately 20 genes are associated with inherited forms of adrenocortical neoplasia, including different types of bilateral adrenocortical hyperplasia, ACA, and ACC.[9-11] Likewise, approximately 20 genes are currently associated with inherited forms of Pheo/PGL.[7] Interestingly, these genes converge on a few signaling pathways. Genes associated with inherited adrenocortical neoplasia primarily converge on programs coordinating cell proliferation, stemness, and zone-specific hormone production. Specifically, frequently altered genes coordinate cell cycle machinery itself or signaling through IGF, Wnt/β-catenin, protein kinase A (PKA), and calcium/calmodulin signaling pathways. Genes associated with inherited Pheo/PGL primarily converge on the mitogen-activated protein kinase (MAPK) pathway, energy metabolism, and hypoxia-sensing pathways. Clinical aspects of adrenal tumor syndromes including manifestations, diagnosis, and management are discussed in other chapters.

The genetics of sporadic adrenal tumors is also complex. Candidate gene approaches revealed that genes and pathways targeted in inherited adrenal tumors are also recurrently altered in different subtypes of sporadic tumors.[9] More recently, pangenomic approaches have illuminated a spectrum of novel recurrent somatic events converging on the same signaling pathways spanned by genes associated with inherited syndromes.[12-14] Furthermore, integration of somatic alteration profiles with other "omics" such as transcriptomics, methylomics, miRnomics, and metabolomics has enabled identification of core molecular subtypes. This large array of somatic alterations has added complexity to our understanding of the molecular basis of primary adrenal tumors, while simultaneously reconciling the broad spectrum of clinical manifestations. In this chapter, we will briefly summarize the genomics of inherited and sporadic primary adrenal tumors from a pangenomic perspective.

GENETIC CAUSES OF ADRENAL NEOPLASIA

Adrenocortical Tumors

KEY POINTS

- Genetic events in benign adrenocortical neoplasia target endocrine and paracrine signaling pathways that serve to integrate homeostatic maintenance and organ function.
- Affected pathways include the Wnt/β-catenin signaling pathway, the protein kinase A pathway, and the calcium/calmodulin signaling pathway, which regulate the stem/progenitor cell compartment of the adrenal cortex, cortisol secretion in response to adrenocorticotropic hormone, and aldosterone secretion, respectively.
- Inherited and sporadic adrenocortical carcinomas possess genetic alterations in these pathways, in addition to recurrent alterations in growth factor signaling, cell cycle checkpoints, DNA mismatch-repair genes, and epigenetic regulators.

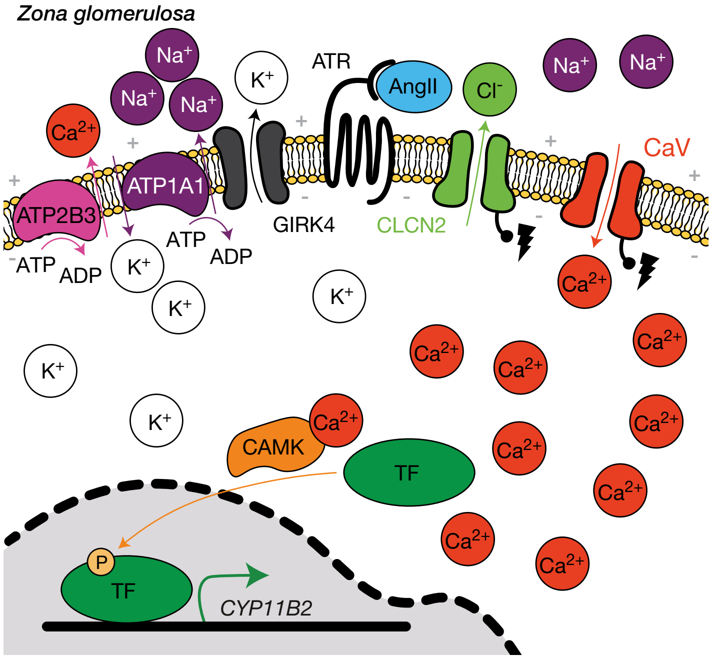


Fig. 95.1 Calcium/calmodulin signaling pathway of the adrenocortical zona glomerulosa. High levels of serum potassium and/or angiotensin II (Ang II) binding to its receptor (ATR, e.g., the receptor encoded by *AGTR1*) initiates zG membrane depolarization by inhibiting the actions of Na/K ATPase ATP1A1 and decreasing permeability of potassium channel GIRK4 (encoded by *KCNJ5*). Consequent changes in transmembrane potential trigger opening of voltage-gated chlorine channels (CLCN2) and voltage-gated calcium channels (CaV, such as those encoded by *CACNA1D* and *CACNA1H*), calcium influx and cytoplasmic accumulation, and activation of calcium/calmodulin-dependent protein kinase (CAMK). Active CAMK directly or indirectly phosphorylates and/or upregulates immediate early response transcription factors (TFs) such as the NR4A, ATF, or CREB family members, leading to transcription of aldosterone synthase (encoded by *CYP11B2*). Resting membrane potential is restored by GIRK4, ATP1A1, and calcium efflux pump ATP2B3. Inherited or somatic mutations in several components of this pathway, leading to constitutive membrane depolarization, cytoplasmic calcium accumulation, and/or transcription of *CYP11B2*, cause primary aldosteronism.

Inherited disorders associated with adrenal hyperfunction and/or overgrowth may affect the cortex or medulla. These rare diseases are typically inherited in an autosomal dominant manner and are caused by alterations in genes coordinating developmental and homeostatic pathways. Frequently altered adrenocortical signaling pathways include the Wnt/β-catenin signaling pathway, the PKA pathway, and the calcium/calmodulin signaling pathway. Disease-causing mutations in these pathways disrupt adrenocortical-specific mechanisms of integration of endocrine-paracrine signals that govern adrenal homeostasis. In addition, other pathways that are associated with broader mechanisms of tumorigenesis, including epigenetic regulation, fetal growth, and cell cycle control, are also targeted by mutations that cause adrenal cortex tumors.

Calcium/calmodulin signaling pathway: zG aldosterone production is controlled by angiotensin II (AngII) through its receptor, the angiotensin II receptor (ATR), such as the G protein-coupled receptor (GPCR) encoded by *AGTR1*. AngII binding to ATR triggers membrane depolarization, release of intracellular calcium stores, and activation of the calcium/calmodulin signaling pathway. This pathway is also activated in an ATR-independent manner, by membrane depolarization initiated by increased serum potassium. Pathway activation culminates in the rapid transcription of target genes essential for aldosterone production, including *CYP11B2*.[15-17] Germline mutations in genes that affect zG cell membrane polarization, including *CLCN2* and *KCNJ5*, and genes that regulate intracellular calcium homeostasis, including *CACNA1D* and *CACNA1H*, cause familial hyperaldosteronism type II, III, and IV and PASNA (primary aldosteronism, seizures, and neurological abnormalities). These syndromes are characterized by early-onset primary hyperaldosteronism and bilateral zG hyperplasia. Sporadic aldosterone-producing ACA bear recurrent mutations in the same classes of genes, including *KCNJ5*, *CLCN2*, *CACNA1D*, and *CACNA1H*.[10] Another class of genes recurrently mutated in sporadic aldosterone-producing ACA encode ion-transporting ATPases, including ATP1A1 and ATP2B3, also enabling constitutive zG cell membrane depolarization.[10] The aforementioned mutations are collectively present in approximately 90% of aldosterone-producing ACA.[18] This pathway is depicted in Fig. 95.1.

Wnt/β-catenin signaling pathway: The Wnt/β-catenin signaling pathway is required for adrenocortical development and homeostasis. It is a key regulator of the progenitor cell pool and shapes adrenocortical functional and anatomic zonation, presumably through tissue-specific roles.[19] The Wnt/β-catenin signaling pathway is tightly regulated by cell-autonomous and nonautonomous mechanisms (Fig. 95.2). Inactivating germline mutations in *APC,* encoding a negative regulator of β-catenin stability, causes familial adenomatous polyposis (FAP). FAP is an autosomal dominant disorder characterized by predisposition to colon cancer and other tumors such as lipomas, connective tissue sarcomas, osteomas, and bilateral adrenal tumors.[9] Subsequent studies identified somatic activating mutations in *CTNNB1*, encoding β-catenin, in up to a third of sporadic ACC and ACA, and somatic *APC* inactivation in a small subset of sporadic ACC.[9] More recently, whole-exome sequencing (WES) studies on large cohorts of ACC identified frequent (~20%) recurrent somatic biallelic deletions in *ZNRF3*, encoding an E3 membrane-bound ubiquitin ligase that also functions

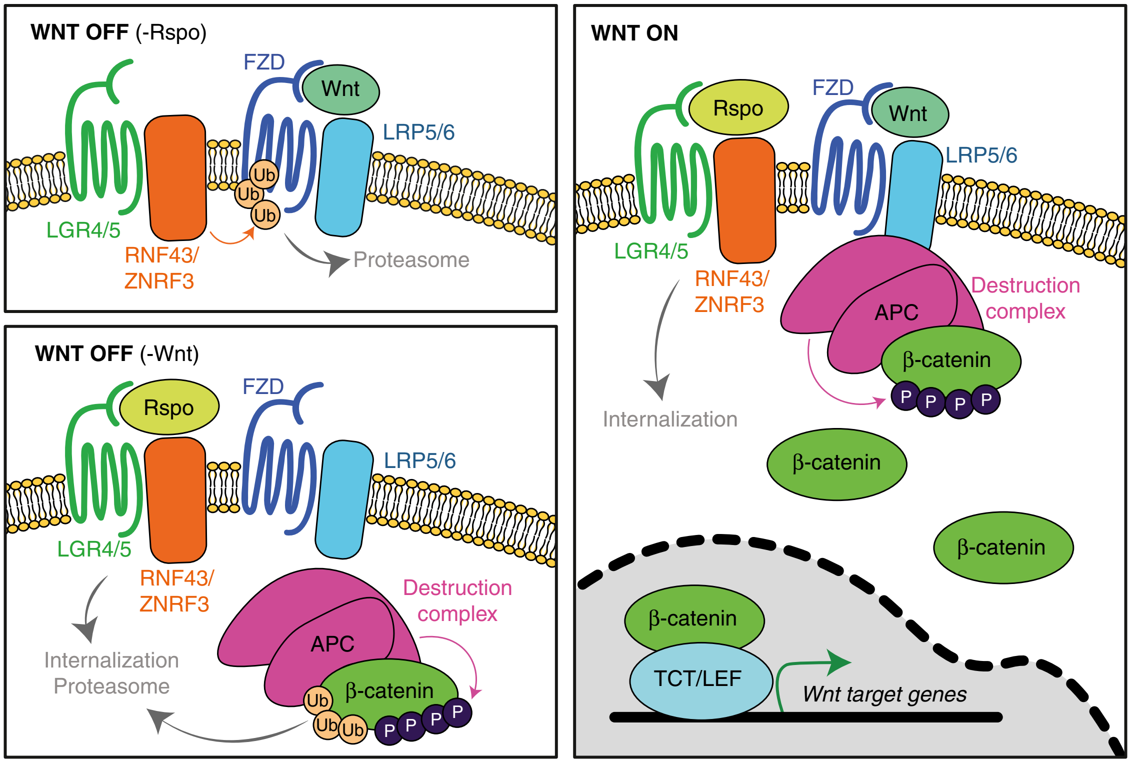


Fig. 95.2 **Wnt/β-catenin signaling pathway.** The Wnt/β-catenin signaling pathway is required for stem cell maintenance and homeostasis of a variety of organs, including the adrenal gland. The final effect of pathway activation by secreted Wnt ligands is to turn on expression of Wnt target genes through β-catenin-dependent coactivation of TCF/LEF-driven transcription. Pathway activation is controlled at several levels, primarily through regulation of stability/compartmentalization of Wnt receptors (Frizzled receptors [FZD]) and β-catenin. Membrane localization of FZDs is controlled by an additional secreted ligand, R-spondin (Rspo). Shown in the upper left ("WNT OFF (-Rspo)"), in the absence of Rspo, FZDs are targeted for degradation by membrane ubiquitin ligase RNF43 or ZNRF3. Shown in bottom left ("WNT OFF (-Wnt)"), when Rspo binds to its receptor LGR4/5, the Rspo/LGR/RNF43 or ZNRF3 complex is internalized, increasing FZD availability at the membrane. In the absence of Wnt ligands, however, β-catenin stability is regulated by a large, cytoplasmic multiprotein complex termed the "destruction complex," which includes the protein APC. In the "WNT OFF" state (as depicted bottom left), this complex binds β-catenin, phosphorylates it, and targets it for ubiquitination and proteasome-mediated destruction. Finally, when Wnt ligands bind to FZD receptors and Wnt signaling is turned on ("WNT ON," right), the destruction complex is sequestered at the cell membrane and incapable of targeting β-catenin for degradation. This allows β-catenin to accumulate in the cytoplasm and translocate to the nucleus to turn on TCF/LEF-dependent transcription. Benign and malignant adrenocortical lesions frequently harbor activating mutations in *CTNNB1*, encoding β-catenin. These mutations are often located at key residues for phosphorylation of β-catenin by the destruction complex, promoting constitutive accumulation of the protein and TCF/LEF-dependent transcription. Adrenocortical carcinomas also harbor inactivating mutations/deletions in *APC* and *ZNRF3*.

as a negative regulator of Wnt/β-catenin signaling.[12,13,20] Importantly, somatic alterations in Wnt pathway components are prevalent (nearly 40% of all ACC) and mutually exclusive, demonstrating the selective advantage conferred by constitutive Wnt/β-catenin signaling as a core driver mechanism of adrenocortical tumorigenesis.

The mechanisms by which Wnt/β-catenin signaling induces adrenocortical tumorigenesis are not fully understood. Observations from mouse models and the clinical spectrum of these tumors suggest that this mechanism is highly context-dependent. In human patients, germline *APC* mutations are associated with nonfunctional hyperplasia, and more rarely, ACA and ACC.[9] *CTNNB1* mutations are prevalent in up to 20% of adenomas and carcinomas.[9] In ACA, tumors bearing *CTNNB1* mutations are typically large and nonfunctional, exhibit mild autonomous cortisol secretion, or, rarely, exhibit aldosterone production.[21,22] In ACC, tumors bearing *CTNNB1* mutations are associated with cortisol production.[13] Pathological mutations and deletions in *ZNRF3* are exclusive to ACC. In mouse models, constitutive Wnt/β-catenin pathway activation through biallelic deletion of *Apc* or deletion of exon 3 of *Ctnnb1* leads to disrupted zonation, progressive zG hyperplasia, and coincident aldosterone overproduction.[23-25] In contrast, biallelic adrenal deletion of *Znrf3* leads to zF-like hyperplasia with massive adrenal enlargement.[26] These observations suggest constitutive Wnt/β-catenin pathway activation confers a context-dependent selective advantage to either zG or zF cells of the adrenal cortex. Combined constitutive Wnt/β-catenin pathway activation with other alterations targeting *Igf2* (encoding IGF2) or *Trp53* (encoding p53) leads to ACC and glucocorticoid excess.[23,25]

Protein kinase A: While Wnt/β-catenin signaling is an essential part of a paracrine relay that establishes and maintains the adrenocortical stem cell compartment, the PKA pathway is the effector arm of the endocrine system (adrenocorticotropic hormone [ACTH]) that induces differentiation and steroid production in the zF and zR. In

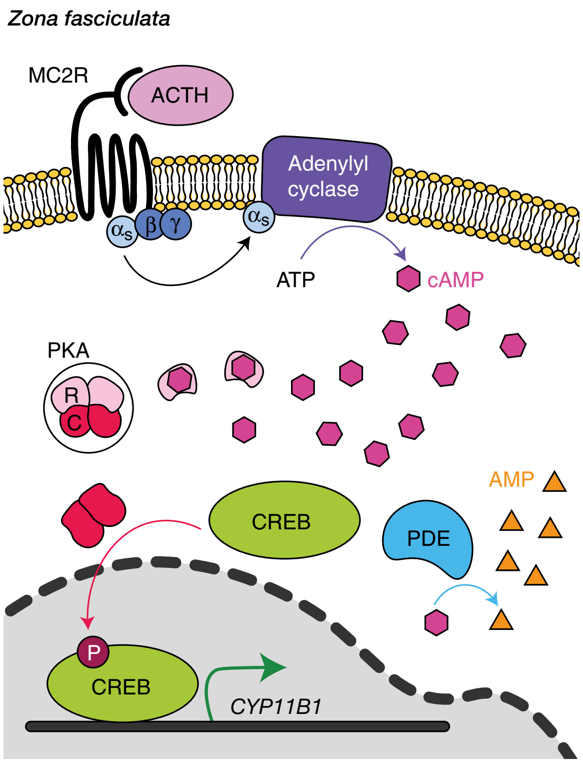


Fig. 95.3 Adrenocorticotropic hormone (ACTH)/protein kinase A (PKA) signaling pathway of the adrenocortical zona fasciculata. ACTH binds to its cell surface receptor, G protein-coupled receptor (GPCR) MC2R to activate the protein kinase A (PKA) signaling pathway. ACTH binding to MC2R triggers dissociation of the Gα,s subunit from the associated Gs protein. Gα,s binds to and activates adenylyl cyclase. Adenylyl cyclase then catalyzes the conversion of cellular ATP to cyclic AMP (cAMP). PKA is a heterotetramer, comprised of two regulatory subunits (R) and two catalytic subunits (C). cAMP binds the regulatory subunits, liberating the catalytic subunits of PKA which then catalyze phosphorylation and nuclear translocation of transcription factor CREB. CREB drives transcription of a variety of genes, including *CYP11B1*, required for glucocorticoid synthesis in the zF. Pathway activation is extinguished by intracellular phosphodiesterases (PDEs), which catalyze the conversion of cAMP to AMP, enabling reassembly of the PKA tetramer. Genetic alterations in components of this pathway leading to constitutive PKA activation and/or *CYP11B1* transcription are major causes of primary hypercortisolism in benign lesions of the adrenal cortex. Recurrent loss of function mutations in genes encoding the R subunit of PKA (e.g., *PRKAR1A*) have also been identified in adrenocortical carcinoma.

the adrenal, ACTH binding to its receptor, the GPCR MC2R, leads to cAMP accumulation and PKA activation. Catalytically active PKA in turn phosphorylates CREB family transcription factors, which translocate into the nucleus to drive expression of genes required for glucocorticoid production (e.g., *CYP11B1*). ACTH-driven PKA activation results in cell proliferation and zF differentiation with increased output of cortisol. PKA signaling is rapidly terminated by the actions of phosphodiestarases, which dephosphorylate the PKA regulatory subunits, favoring inhibitory binding to the catalytic subunit[11,27,28] (Fig. 95.3).

Several genetic syndromes presenting with adrenocortical hyperplasia and hyperfunction are caused by germline defects in genes that encode members of the PKA pathway, including *GNAS*, *PRKAR1A*, *PDE11A*, *PDE8B*, and *PRKACA*.[11] These genetic defects lead to constitutive PKA activation, rendering adrenocortical cells in a permanent state of activation characterized by cortisol overproduction and hyperplasia. Somatic mutations in PKA pathway members, including *GNAS*, *PRKACA*, *PRKAR1A*, *PRKAR1B*, and *PDE11A*, are commonly found in cortisol-producing adenomas.[29] Recurrent somatic mutations of *PRKAR1A* have been recently identified in ACC, implicating constitutive PKA signaling in malignant programming.[13]

Growth factor signaling: IGF2 is a growth factor that is widely expressed during fetal life and is one of the major regulators of intrauterine growth. Growth-promoting effects of IGF2 are mediated by the insulin-like growth factor 1 receptor (IGF1R), a receptor tyrosine kinase that activates the MAPK pathway. In physiological somatic tissues, the *IGF2* locus on 11p15.5 is imprinted and exclusively expressed by the paternal allele. The imprinting of *IGF2* is regulated by the *H19/IGF2*-imprinting control region (ICR1), located between *H19*, a long noncoding RNA with growth inhibitory functions expressed exclusively from the maternal allele, and *IGF2*. Patients with germline mutations or deletions in imprinted genes located on 11p15.5, including *CDKN1C*, *H19*, and *KCNQ1OT1*, and hypermethylation of ICR1 develop Beckwith–Wiedemann syndrome, an overgrowth disorder characterized by abdominal wall defects, abnormal growth, macroglossia, visceromegaly (including adrenocortical cytomegaly), and embryonal malignancies such as Wilms tumor, hepatoblastoma, and ACC. Interestingly, *IGF2* overexpression is present in 90% of sporadic ACC, usually associated with loss of heterozygosity (LOH) at 11p15.5 (and consequent loss of ICR1), suggesting a critical role for dysregulation of the MAPK pathway in ACC tumorigenesis.[9]

Cell cycle regulation: Germline mutations in *TP53* cause Li–Fraumeni syndrome (LFS). The prevalence of ACC in patients with LFS is approximately 5%.[30] Interestingly, the prevalence of germline *TP53* mutations among pediatric ACC cases is 50% to 90%. Somatic mutations and frequent LOH of 17p13 have often been described in sporadic ACC. The prevalence of somatic *TP53* mutations is approximately 20%. In addition to alterations in *TP53*, recurrent somatic events targeting other checkpoint regulators include inactivating mutations/deletions of *CDKN2A* and *RB1* and amplification of *CDK4* and *CCNE1*.[9,12,13,31]

DNA mismatch repair: Germline mutations in mismatch-repair genes cause Lynch syndrome. This disease is characterized by predisposition to several types of neoplasia, including gastrointestinal and gynecological tumors. Characteristically, these tumors present with microsatellite instability. ACC has also been described as part of the spectrum of Lynch syndrome.[32] Somatic mutations in mismatch-repair genes have been recently described in ACC by The Cancer Genome Atlas (TCGA), and up to 30% of ACC exhibit a mismatch repair-deficient signature, but their role in tumorigenesis is unclear.[9,33] Beyond Lynch syndrome genes, variants in other genes implicated in DNA repair, such as *MutYH*, seem to predispose to ACC, with a specific mutational signature.[34]

Other mechanisms: Germline and somatic alterations targeting other general cellular intrinsic regulatory mechanisms such as chromatin remodeling and telomere maintenance may also play a role in ACC tumorigenesis. Germline mutations in *MEN1* cause multiple endocrine neoplasia type 1 (MEN1), an autosomal dominant disorder characterized by multiple endocrine tumors, including gastrinomas, insulinomas, and parathyroid and pituitary adenomas. *MEN1* encodes menin, a protein best characterized as a scaffold protein for chromatin remodelers. Precisely how menin contributes to tissue-specific tumorigenesis is still not understood. Furthermore, up to 40% of patients with MEN1 develop nonfunctional bilateral adrenal macronodular hyperplasia that may rarely evolve to ACC.[9] Recently, recurrent somatic mutations in *MEN1* have been described in ACC, suggesting a driver role for mutations in this gene in adrenocortical carcinogenesis.[13] In addition to bearing mutations in *MEN1*, sporadic ACC may also possess alterations in genes encoding other chromatin remodeling and associated proteins, including *MLL* family members, *ATRX*, *DAXX*, *TET1*, *SETD2*, and *SMARCA4*.[12,13]

KEY POINTS

- Integrative multiplatform analyses of adrenocortical tumors have enabled the identification of distinct molecular subtypes, characterized by the coocurrence of specific molecular features such as somatic mutations targeting a given signaling pathway, activation of specific transcriptional programs, DNA methylation patterns, and chromosome copy-number alteration profiles.
- In adrenocortical carcinoma, at least three molecular subtypes have been identified, and they are strongly associate with clinical and histological metrics such as hormone secretion, mitotic counts, metastatic potential, and clinical outcomes.

Integrated molecular classification of adrenocortical adenomas: By integrating tumor genome sequencing data with gene expression (microarray, RNA-seq, and miRNA-seq), DNA methylation (Illumina arrays), chromosome alteration profiles (single nucleotide polymorphism [SNP] arrays), and proteomics, unbiased and multiplatform studies further characterized the heterogeneity of adrenal tumors and helped illuminate the molecular consequences of driver events for tumorigenesis.[12,13] Furthermore, while these studies have limited immediate translational or clinical applicability, they offer an excellent opportunity for identification of prognostic and predictive biomarkers, providing the rationale for preclinical studies directing new agents and existing targeted therapies to specific molecular subtypes.

Few studies have performed pangenomic analyses of ACA. Di Dalmazi et al.[21] performed RNA-seq in a large cohort of ACA, including samples from patients with overt Cushing syndrome, mild autonomous cortisol secretion, endocrine inactive adenomas, and low-grade low-stage ACC. Unsupervised analyses have identified three different types of ACA that presented with specific clinical and demographical features, as well as distinct somatic mutational profiles. Samples bearing somatic mutations in genes encoding PKA components including *PRKACA* and *GNAS* formed an individual cluster of tumors characterized by overt Cushing syndrome presentation, increased frequency of female sex, and smaller size. On the other hand, ACA with *CTNNB1* mutations formed another cluster of tumors characterized by mild autonomous cortisol secretion or endocrine silent tumors, older age at presentation, and no gender predilection. Finally, a small subgroup of ACA with *CTNNB1* mutations formed a cluster that exhibit transcriptional features resembling low-grade ACC. These results are consistent with other transcriptome studies of non-aldosterone-producing ACA, which identified two subtypes of tumors by unsupervised analysis that segregated cortisol-producing adenomas from those that are endocrine-silent or exhibit mild cortisol production.[22,35]

Two integrated multiplatform studies have performed molecular profiling of a large number of ACC samples.[12,13] These studies have identified discrete molecular subtypes of ACC that exhibit distinct clinical features and prognosis. De Reynies et al. identified two major transcriptional classes of ACC, C1A and C1B,[36] with striking differences in prognosis (C1A is associated with poor prognosis, and C1B is associated with good prognosis, often cured by surgical resection). Using a multiplatform approach that included microarray, methylation, microRNA-seq, SNP-array, and WES, Assie et al. expanded the characterization of the two previously identified molecular subtypes of ACC. C1A ACC exhibited a higher mutation burden, as well as frequent somatic events in driver genes, including *CTNNB1, ZNRF3, APC, TP53, RB1, CDKN2A,* and *DAXX*.[12] In addition, approximately two-thirds of C1A tumors exhibited abnormal CpG island hypermethylation (CIMP), a phenomenon previously associated with poor outcomes in ACC.[37] On the other hand, C1B exhibited a lower mutation burden, and several tumors in this class did not bear an identifiable somatic "driver." However, approximately two thirds of C1B tumors exhibited a profound deregulation of a cluster of microRNAs located in 14q32 (*DLK1-MEG3* cluster), coupled with LOH of this region, suggesting this may be the molecular event driving these tumors.

As part of the TCGA consortium, Zheng et al.[13] expanded on these studies to provide a broader multiplatform classification of ACC (ACC-TCGA). ACC-TCGA identified three unique classes of ACC (COC1, COC2, COC3), driven by differential regulation of steroidogenesis, cell cycle, and immune infiltration. Notably, each type of ACC also possessed a unique somatic copy number alteration (SCNA) profile and DNA methylation landscape. Strikingly, each molecular subtype of ACC is associated with a distinct prognostic outcome: COC1 tumors exhibited good prognosis (progression rate [PR] of 7%; median event-free survival [EFS] not reached), COC2 tumors exhibited intermediate prognosis (PR 56%; EFS 38 months), and COC3 tumors exhibited dismal prognosis (PR 96%; EFS 8 months).[13]

The vast majority of ACC is "pure," that is, free of stromal and other infiltrating immune cells, with average purity approaching 90%; however, while COC1 tumors still exhibit high purity levels in comparison to other TCGA cancers, they are characterized by low expression of steroidogenic enzymes and high expression of immune genes consistent with a higher degree of immune infiltration.[13] Like C1B tumors identified by de Reynies et al., COC1 tumors also did not possess recurrent somatic alterations. COC1 tumors had low levels of genome-wide CpG island hypermethylation (CIMP-low) and an SCNA profile termed "chromosomal," characterized by whole-chromosomal gains and losses and copy-neutral LOH, sometimes accompanied by whole-genome doubling. COC2 tumors were characterized by largely absent immune infiltration, high expression of steroidogenic enzymes, and recurrent somatic alterations targeting the Wnt/β-catenin pathway. COC2 ACC had intermediate levels of genome-wide CpG island hypermethylation (CIMP-intermediate) and a chromosomal SCNA profile. Finally, COC3 tumors were characterized by largely absent immune infiltration, high expression of steroidogenic enzymes and cell cycle machinery, and recurrent somatic alterations targeting the Wnt/β-catenin pathway and cell cycle. COC3 tumors had high levels of genome-wide CpG island hypermethylation (CIMP-high) and a SCNA profile termed "noisy," characterized by focal arm-level gains and losses and numerous breakpoints, sometimes accompanied by whole-genome doubling (Fig. 95.4).

The completion of ACC-TCGA enabled incorporation of ACC into larger pancancer pangenomic datasets. These studies have identified several unique characteristics of ACC, with therapeutic implications. Up to a third of ACC exhibit a somatic mismatch repair-deficient signature,[33] and most ACC is immune-poor, with minimal lymphocyte infiltration,[38] suggesting these tumors are as a majority resistant to immunotherapy, consistent with recent clinical studies.[39] On an epigenetic level, the accessible chromatin landscape of ACC is largely dominated by the adrenocortical transcription factor SF1,[40] suggesting that potentially efficacious epigenetic therapies would need to unwire this highly tissue-specific program.

Pheochromocytoma

Up to 40% of Pheo/PGL are associated with inherited syndromes, and to date there are approximately 20 well-documented susceptibility genes.[7] Clinical aspects of these syndromes and strategies for genetic testing and clinical management are beyond the scope of this chapter and are discussed elsewhere. Broadly, two signaling pathways are most affected by these germline mutations: cellular hypoxia and activated kinase signaling.[41,42] Activation of the cellular hypoxia pathway

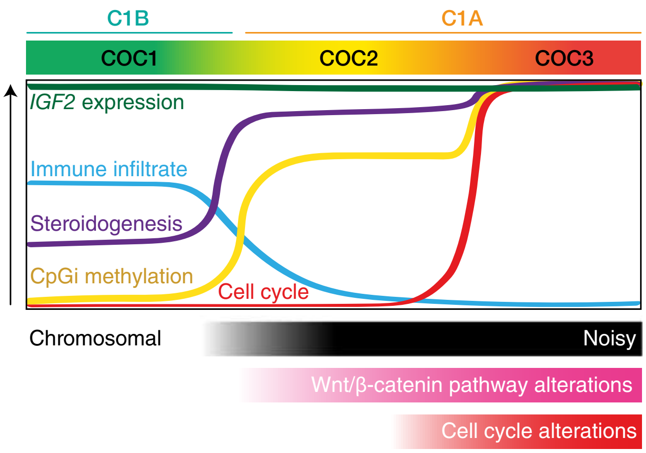


Fig. 95.4 **Molecular subtyping of adrenocortical carcinoma (ACC).** Recent pangenomic studies (namely, The Cancer Genome Atlas study on ACC [ACC-TCGA]) have revealed that ACC is comprised of three distinct molecular subtypes, COC1, COC2, and COC3. Not shown in this figure, each subtype is associated with a unique clinical outcome where patients with COC1 tumors have the best outcomes (longer progression-free survival), patients with COC2 tumors have intermediate outcomes, and patients with COC3 tumors have the worst outcomes (shortened progression-free survival with near uniform likelihood of relapse within 2 years). Strikingly, ACC-TCGA revealed that each class of tumors is defined by a combination of core transcriptional programs, DNA methylation profiles, and somatic alterations. All ACC possess high levels of *IGF2* expression, secondary to loss of imprinting of this locus. COC1 tumors possess the lowest expression of steroidogenic enzymes, the lowest degree of cell cycle activation, the lowest levels of promoter CpG island (CpGi) methylation (CIMP-low), and few to no recurrent somatic alterations. COC1 tumors bear a somatic copy number alteration profile characterized by whole arm/whole chromosome loss of heterozygosity (LOH), called "chromosomal." COC1 tumors also possess a moderate degree of immune infiltration (approaching 50% in some tumors). In contrast, COC2 and COC3 tumors bear high expression of steroidogenic enzymes, higher levels of CpGi methylation (CIMP-intermediate and CIMP-high), and frequent Wnt/β-catenin pathway alterations. These tumors are also characterized by a low to near-absent immune infiltration. COC3 tumors possess the highest expression of steroidogenic enzymes, highest degree of cell cycle activation, and highest degree of CpGi methylation (CIMP-high). In contrast to COC1–2 tumors, COC3 possess a "noisy" somatic copy number alteration profile, characterized by numerous focal arm-level copy number gains and losses throughout the genome. Finally, as shown at the top of the plot, the transcriptional programs driving COC1 and COC2–3 tumors resemble the C1B and C1A signatures originally identified by de Reynies et al. in a seminal unbiased study using microarray to characterize transcriptional programs driving adrenocortical tumors.

is driven by the hypoxia-inducible factor (HIF) family of transcription factors. Physiologically, this pathway is activated in response to cellular hypoxia, a state that is recognized by hypoxia-sensing proteins, which trigger a downstream signaling cascade that leads to the accumulation of HIF family members. HIF induces the expression of a broad spectrum of genes that promote erythropoiesis, angiogenesis, cell survival, cell proliferation, and glycolysis. A central protein in this pathway is the von Hippel–Lindau protein (VHL), encoded by *VHL*. Under normal oxygen concentrations, VHL is a member of a multiprotein complex that promotes ubiquitin/proteasome-mediated degradation of HIF transcription factors. Hydroxylation of proline residues in HIF by the prolyl-hydroxylase PHD2 (EGLN1) is critical for VHL docking, a step crucial for ubiquitination.[43]

Inactivating germline mutations in *VHL* cause von Hippel–Lindau syndrome, an autosomal dominant disorder characterized by increased risk for several tumor types, including retinal, brain, and spinal cord hemangioblastomas, pancreatic neuroendocrine tumors, clear cell renal carcinomas, angiolymphatic sac tumors, and Pheo/PGL. Complete loss of VHL is usually a consequence of a somatic LOH event at 3p25.3. Consistent with increased angiogenesis, the tumors associated with VHL are rich in blood vessels. Interestingly, rare germline mutations in VHL that lead to mild defects in the protein function cause an autosomal recessive form of congenital erythrocytosis. This is presumably through global low-level activation of HIF-induced transcription, which leads to inappropriately high levels of erythropoietin, thereby inducing otherwise hypoxia-dependent erythrocytosis.[43] Recurrent somatic mutations in *VHL* have also been described in Pheo/PGL.[14]

In addition to *VHL*, mutations targeting another gene of the hypoxia pathway have been described in patients with syndromic forms of Pheo/PGL syndromes, including the polycythemia-paraganglioma-somatostatinoma syndrome. Genetically, this syndrome is characterized by a somatic mosaicism involving activating mutations of *EPAS1*, also known as *HIF2A*. These mutations also lead to constitutive activation of HIF-dependent transcriptional programs, leading to erythropoietin overproduction and tumor formation in target tissues.[44] Recurrent somatic mutations in *EPAS1* have also been described in sporadic Pheo/PGL.[14] Finally, mutations in genes encoding energy metabolism enzymes, including *SDH* family members (*SDHx*, encompassing *SDHA*, *SDHAF2*, *SDHB*, *SDHC*, and *SDHD*), *FH*, and *MDH2* lead to a metabolic switch that results in abnormal accumulation of metabolites (succinate or fumarate), which disrupt cell metabolism through the TCA cycle by inhibiting iron(II)- and α-ketoglutarate-dependent (α-KG) dioxygenases. These enzymes coordinate a broad spectrum of cellular programs. In particular, prolyl hydroxylases such as EGLN1, which are negative regulators of HIF stability, are inhibited by succinate and fumarate accumulation. TET enzymes, which coordinate DNA demethylation, are also inhibited by abnormal metabolite accumulation. TET inhibition leads to global epigenetic remodeling, including acquisition of the CIMP signature, thought to play a major role in tumorigenesis and in malignant transformation.[45,46] At the somatic level, the hotspot *IDH1* p.R132C mutation, which is known to dysregulate the TCA cycle through oncometabolite accumulation,[47] has also been described in a subset of sporadic Pheo/PGL.[14,48]

The other broad signaling pathway that is targeted by germline mutations in Pheo/PGL, activated kinase signaling,[41,42] is a growth-promoting pathway that promotes activation of translation and cell proliferation through MAPK. Transcriptional effectors of this signaling pathway include the MYC/MAX complex.[49] In Pheo/PGL, germline mutations associated with increased kinase activity include activating mutations of *RET*, inactivating mutations of *NF1*, and inactivating mutations in *TMEM127*, which cause multiple endocrine neoplasia type 2, neurofibromatosis type 1, and familial pheochromocytoma, respectively.[42] Mutations in *RET* occur at specific hotspots, including C634 and M918, and lead to the constitutive activation of its tyrosine kinase activity. Mutations in *NF1* are scattered throughout the gene and include large deletions. The protein NF1 is a negative regulator of the MAPK signaling pathway. Recurrent somatic mutations in *RET* and *NF1* have also been described in spontaneous Pheo/PGL, with *RET* mutations occurring at restricted hotspots.[14,50,51] Finally, germline mutations in *MAX*, which encodes MYC-associated factor X (MAX protein), are another cause of familial Pheo/PGL. Disruption of MAX promotes oncogenic effects of MYC. Recurrent somatic mutations in *MAX* have also been described in sporadic Pheo/PGL.[52] Constitutive activation of the MAPK pathway can also be triggered by mutations

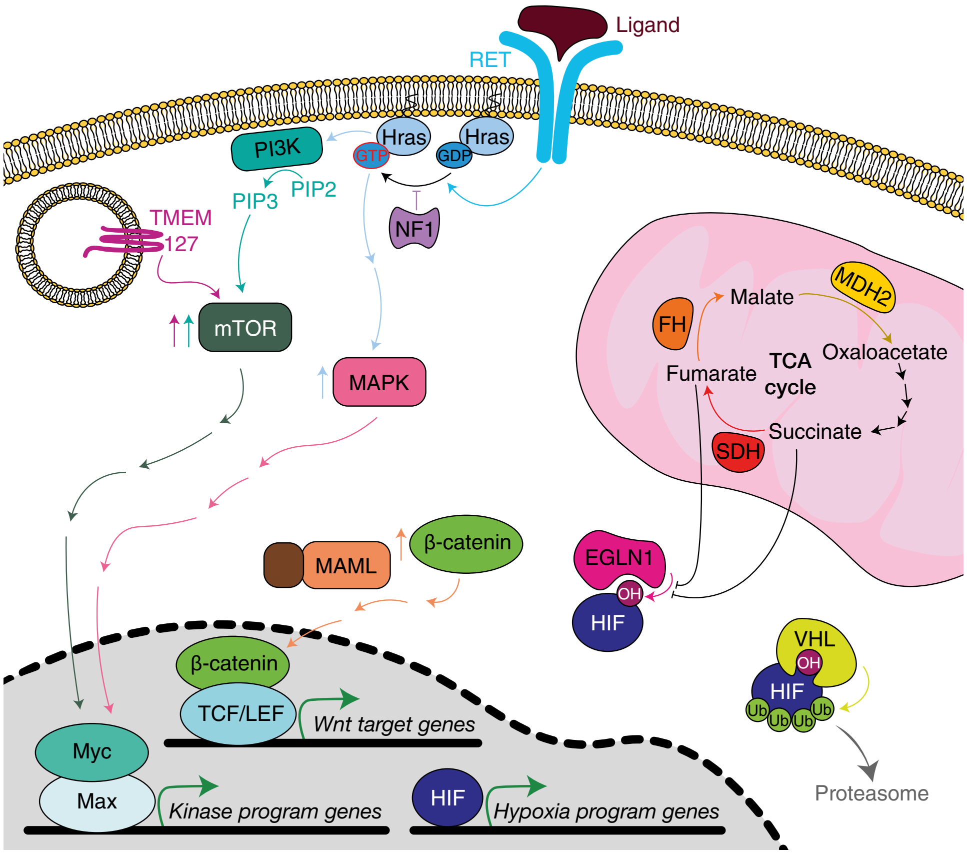


Fig. 95.5 Recurrent genetic alterations driving pheochromocytomas and paragangliomas (Pheo/PGL). Recent pangenomic studies of Pheo /PGL have revealed that these tumors are defined by activation of key cellular programs: cellular hypoxia signaling and activated kinase signaling. This is accomplished through alterations in genes encoding components of the TCA cycle (SDHx, FH, MDH2), effectors/putative modulators of receptor tyrosine kinase and mTOR/MAPK signaling (RET, NF1, HRAS, TMEM127, MAX), and critical regulators of cellular hypoxia (genes encoding hypoxia-inducible transcription factors, von Hippel–Lindau, and EGLN1). Sporadic Pheo also possess recurrent fusions targeting the *MAML3* locus, leading to expression of a chimeric fusion transcript and upregulation of Wnt/β-catenin–dependent transcription. How these different pathway components coordinate signaling cascades believed to drive expression of oncogenic programs is depicted in this figure. Notably, mutations in enzymes coordinating TCA cycle are frequently loss of function, driving the hypoxia program through aberrant accumulation of TCA metabolites (succinate and fumarate). Partially depicted, accumulation of succinate and fumarate inhibit several classes of enzymes including EGLN1 and TET enzymes that erase DNA methylation. Consequently, tumors bearing TCA cycle mutations also possess a DNA hypermethylation signature.

that only occur somatically in sporadic Pheo/PGL, such as hotspot mutations in *HRAS*.[14,53]

More recently, a study that performed multiplatform molecular profiling of a large number of Pheo/PGL identified novel recurrent somatic events, including mutations in *CSDE1* and gene fusions involving *MAML3*.[14] *CSDE1* encodes cold shock domain-containing E1, an RNA-binding protein involved in RNA stability, translation initiation, and neuronal development. Somatic mutations in *CSDE1* were present in 2% of the samples, usually affecting splicing sites, resulting in loss of function. Fusions involving *MAML3* were present in 5% of Pheo/PGL and involve the promoter regions of either *UBTF* or *TCF4* and the coding region of *MAML3*, resulting in in-frame chimeric transcripts. Because these resultant chimeric transcripts are overexpressed, *MAML3* fusions are presumably gain-of-function. Additional fusions were identified involving other genes, such as *KIAA1737-NGFR*, *RUNDC1-BRAF*, and *NF1-RAB11F1P4*. The first two result in overexpression of the kinase transcripts, and therefore are thought to be gain-of-function; the last results in disruption of *NF1* expression.

Other recurrent somatic mutations that reached statistical significance, according to this study, were the previously identified *HRAS*, *NF1*, *EPAS1*, and *RET* (present in 10%, 9%, 5%, and 3% of samples, respectively). In addition to these, somatic hotspot and cancer-relevant genes mutations in *BRAF*, *IDH1* (p.R132C), *FGFR1*, *VHL*, *ATRX*, *TP53*, *SETD2*, and *ARNT* were also detected. Importantly, unlike other tumor types that may have mutations in more than one driver gene, a striking mutual exclusivity was detected among somatic mutations and pathogenic germline variants detected in this study, indicating that tumors typically require only one mutation in this class of genes, except for *ATRX* mutations which usually cooccur with germline *SDHB* mutations, a previously described association.[54] Pathways relevant for Pheo/PGL biology are depicted in Fig. 95.5.

Integrated molecular classification of pheochromocytomas/paragangliomas: Multiomics profiling of Pheo/PGL have confirmed previous observations that prominent somatic and germline driver alterations converge on specific molecular subtypes.[41,42,51] Unsupervised consensus clustering analysis on the transcriptome data revealed that Pheo/PGL consists of four molecular subtypes, designated "kinase signaling," "pseudohypoxia," "Wnt-altered," and "cortical-admixture."[14] Consistent with previous studies, the "kinase-signaling" subtype was observed predominantly in epinephrine-producing adrenal pheochromocytoma of the "adrenergic" phenotype (see other chapters for a detailed discussion). This subtype was enriched for germline and somatic mutations targeting *NF1, RET, TMEM127,* and *HRAS,* as well as rarer fusion events involving *NF1, HRAS,* and *NGFR*. Furthermore, specific arm-level chromosomal deletions involving 1p, 3q, and 17q were also enriched in this subtype. Also consistent with previous observations, the "pseudohypoxia" subtype is enriched for germline mutations and somatic mutations in *SDHx, VHL, EPAS1,* and *EGLN1* (which were exclusive of this subtype) and non-epinephrine-producing Pheo/PGL. Other distinctive molecular features of the "pseudohypoxia" cluster include increased DNA methylation on CpG islands secondary to metabolite-dependent inhibition of DNA demethylases. Finally, whole-genome doubling events, characterized by extra copies of all chromosomes, were significantly enriched in the "pseudohypoxia" subtype, frequently in association with *VHL* and *EPAS1* mutations. The previously unappreciated "Wnt-altered" cluster exclusively consisted of sporadic adrenal pheochromocytomas, and was significantly enriched for *MAML3*-containing fusion transcripts and *CSDE1* mutations. This subtype features increased expression of targets of Wnt/β-catenin and Sonic Hedgehog pathways. Finally, the "cortical-admixture" subtype is characterized by increased expression of several adrenocortical genes, including *STAR, CYP11B2,* and *CYP21A2,* consistent with significant infiltration with cortical tissue. The presence of *MAX* mutations in this group, which is often associated with multifocal local disease, suggests that coalescing multiple tumor foci may "trap" cortical cells interspersed within the tumor.

SUMMARY AND FUTURE DIRECTIONS

Adrenal tumors are a heterogeneous group of diseases with diverse clinical phenotypes. This complexity is mirrored at the molecular level. A large number of genetic drivers targeting recurrent biological processes or molecular pathways have been described in sporadic and familial forms of adrenal tumors and are usually shared between the inherited and equivalent sporadic form. Within a similar tumor type, these genetic alterations cluster around molecular subtypes, which are biologically correlated groups of tumors based on unbiased multiomics classification. These molecular subtypes closely explain important aspects of the clinical variability, such as the endocrine manifestations, growth patterns, and disease prognosis. However, classifying a new tumor sample in a given molecular subtype outside the scope of a comprehensive multiomics study can be challenging. This task can be accomplished by the use of a biomarker, which is a single measurement of an analyte that can be used as a surrogate to identify a specific disease state. In adrenal tumors, these biomarkers can be hormones, mutations, or other molecular features, such as patterns of expression of specific genes, and targeted DNA methylation.[2,3,36,55-57] Examples of biomarkers for adrenal tumors include catecholamine/metanephrine levels to establish a diagnosis of Pheo/PGL, steroid profiling and *IGF2* expression to distinguish between ACA and ACC, steroid profiling and catecholamine/metanephrine levels to diagnose recurrence, expression levels of *BUB1B, PINK1,* and *G0S2,* and methylation levels of *G0S2* to diagnose highly recurrent/aggressive (C1A and ACC-TCGA COC3) subtypes of ACC. However, implementing biomarkers into clinical practice to aid into the decision-making process requires extensive validation, which is the goal of ongoing clinical studies.

Molecular subtypes, in particular, offer several meaningful insights that can be exploited therapeutically under the umbrella of personalized medicine as an alternative to the prevailing blanket therapy approaches. Molecular subtypes are particularly relevant for targeted therapies, such as kinase inhibitors, cell cycle inhibitors, epigenetic drugs, and immunotherapy. Early clinical trials using targeted therapies for adrenal tumors were disappointing.[58,59] However, these trials were performed at a time when the molecular subtypes were not fully characterized. In fact, small subgroups of patients exhibited measurable and long-term benefits from "failed" agents. Prospective identification of the potential responders could have changed the history of these trials, as molecular subtyping may more effectively narrow down likely responders according to the therapy's mechanism of action. In this setting, the use of molecular biomarkers is essential. However, the potential advantages should be extensively demonstrated and validated in preclinical studies and clinical trials in order to support this paradigmatic shift and justify the increase in cost and complexity of healthcare.

REFERENCES

1. Sherlock M, Scarsbrook A, Abbas A, et al. Adrenal incidentaloma. *Endocr Rev*. 2020;41(6):775–820.
2. Wajchenberg BL, Albergaria Pereira MA, Medonca BB, et al. Adrenocortical carcinoma: clinical and laboratory observations. *Cancer*. 2000;88:711–736.
3. Arlt W, Biehl M, Taylor AE, et al. Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*. 2011;96:3775–3784.
4. Eisenhofer G, Lenders JW, Siegert G, et al. Plasma methoxytyramine: a novel biomarker of metastatic pheochromocytoma and paraganglioma in relation to established risk factors of tumour size, location and SDHB mutation status. *Eur J Cancer*. 2012;48:1739–1749.
5. Crona J, Lamarca A, Ghosal S, et al. Genotype-phenotype correlations in pheochromocytoma and paraganglioma: a systematic review and individual patient meta-analysis. *Endocr Relat Cancer*. 2019;26:539–550.
6. Else T, Kim A, Sabolch A, et al. Adrenocortical carcinoma. *Endocr Rev*. 2014;35(2):282–326.
7. Neumann HPH, Young WF, Eng C. Pheochromocytoma and paraganglioma. *N Engl J Med*. 2019;381:552–565.
8. Lenders JW, Eisenhofer G, Mannelli M, et al. Phaeochromocytoma. *Lancet*. 2005;366:665–675.
9. Lerario AM, Moraitis A, Hammer GD. Genetics and epigenetics of adrenocortical tumors. *Mol Cell Endocrinol*. 2014;386:67–84.
10. Boulkroun S, Fernandes-Rosa FL, Zennaro MC. Old and new genes in primary aldosteronism. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101375.
11. Berthon A, Bertherat J. Update of genetic and molecular causes of adrenocortical hyperplasias causing Cushing syndrome. *Horm Metab Res*. 2020;52:598–606.
12. Assié G, Letouzé E, Fassnacht M, et al. Integrated genomic characterization of adrenocortical carcinoma. *Nat Genet*. 2014;46(6):607–612.
13. Zheng S, Cherniack AD, Dewal N, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016;30:363.
14. Fishbein L, Leshchiner I, Walter V, et al. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell*. 2017;31:181–193.
15. Bandulik S, Penton D, Barhanin J, et al. TASK1 and TASK3 potassium channels: determinants of aldosterone secretion and adrenocortical zonation. *Horm Metab Res*. 2010;42:450–457.
16. Bassett MH, White PC, Rainey WE. The regulation of aldosterone synthase expression. *Mol Cell Endocrinol*. 2004;217:67–74.
17. Szekeres M, Turu G, Orient A, et al. Mechanisms of angiotensin II-mediated regulation of aldosterone synthase expression in H295R human adrenocortical and rat adrenal glomerulosa cells. *Mol Cell Endocrinol*. 2009;302:244–253.
18. Nanba K, Omata K, Else T, et al. Targeted molecular characterization of aldosterone-producing adenomas in white Americans. *J Clin Endocrinol Metab*. 2018;103:3869–3876.

19. Little DW, Dumontet T, LaPensee CR, et al. β-catenin in adrenal zonation and disease. *Mol Cell Endocrinol*. 2021;522:111120.
20. Juhlin CC, Goh G, Healy JM, et al. Whole-exome sequencing characterizes the landscape of somatic mutations and copy number alterations in adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2015;100:E493–E502.
21. Di Dalmazi G, Altieri B, Scholz C, et al. RNA sequencing and somatic mutation status of adrenocortical tumors: novel pathogenetic insights. *J Clin Endocrinol Metab*. 2020;105(12):e4459–e4473.
22. Faillot S, Foulonneau T, Néou M, et al. Genomic classification of benign adrenocortical lesions. *Endocr Relat Cancer*. 2021;28:79–95.
23. Heaton JH, Wood MA, Kim AC, et al. Progression to adrenocortical tumorigenesis in mice and humans through insulin-like growth factor 2 and beta-catenin. *Am J Pathol*. 2012;181:1017–1033.
24. Berthon A, Sahut-Barnola I, Lambert-Langlais S, et al. Constitutive beta-catenin activation induces adrenal hyperplasia and promotes adrenal cancer development. *Human Mol Genet*. 2010;19(8):1561–1576.
25. Borges KS, Pignatti E, Leng S, et al. Wnt/β-catenin activation cooperates with loss of p53 to cause adrenocortical carcinoma in mice. *Oncogene*. 2020;39:5282–5291.
26. Basham KJ, Rodriguez S, Turcu AF, et al. A ZNRF3-dependent Wnt/β-catenin signaling gradient is required for adrenal homeostasis. *Genes Dev*. 2019;33:209–220.
27. Ruggiero C, Lalli E. Impact of ACTH signaling on transcriptional regulation of steroidogenic genes. *Front Endocrinol (Lausanne)*. 2016;7:24.
28. Azevedo MF, Faucz FR, Bimpaki E, et al. Clinical and molecular genetics of the phosphodiesterases (PDEs). *Endocr Rev*. 2014;35:195–233.
29. Zennaro M-C, Boulkroun S, Fernandes-Rosa F. Genetic causes of functional adrenocortical adenomas. *Endocr Rev*. 2017;38:516–537.
30. Raymond VM, Else T, Everett JN, et al. Prevalence of germline TP53 mutations in a prospective series of unselected patients with adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2013;98:E119–E125.
31. Ragazzon B, Libé R, Assié G, et al. Mass-array screening of frequent mutations in cancers reveals RB1 alterations in aggressive adrenocortical carcinomas. *Eur J Endocrinol*. 2014;170:385–391.
32. Raymond VM, Everett JN, Furtado LV, et al. Adrenocortical carcinoma is a lynch syndrome-associated cancer. *J Clin Oncol*. 2013;31:3012–3018.
33. Knijnenburg TA, Wang L, Zimmermann MT, et al. Genomic and molecular landscape of DNA damage repair deficiency across the Cancer Genome Atlas. *Cell Rep*. 2018;23:239–254.e6.
34. Pilati C, Shinde J, Alexandrov LB, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol*. 2017;242:10–15.
35. Wilmot Roussel H, Vezzosi D, Rizk-Rabin M, et al. Identification of gene expression profiles associated with cortisol secretion in adrenocortical adenomas. *J Clin Endocrinol Metab*. 2013;98:E1109–E1121.
36. de Reyniès A, Assié G, Rickman DS, et al. Gene expression profiling reveals a new classification of adrenocortical tumors and identifies molecular predictors of malignancy and survival. *J Clin Oncol*. 2009;27:1108–1115.
37. Barreau O, Assié G, Wilmot-Roussel H, et al. Identification of a CpG island methylator phenotype in adrenocortical carcinomas. *J Clin Endocrinol Metab*. 2013;98:E174–E184.
38. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2019;51:411–412.
39. Raj N, Zheng Y, Kelly V, et al. PD-1 blockade in advanced adrenocortical carcinoma. *J Clin Oncol*. 2020;38(1):71–80.
40. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape of primary human cancers. *Science*. 2018;362(6413):eaav1898.
41. Dahia PL, Ross KN, Wright ME, et al. A HIF1alpha regulatory loop links hypoxia and mitochondrial signals in pheochromocytomas. *PLoS Genet*. 2005;1:72–80.
42. Castro-Vega LJ, Letouzé E, Burnichon N, et al. Multi-omics analysis defines core genomic alterations in pheochromocytomas and paragangliomas. *Nat Commun*. 2015;6:6044.
43. Kaelin WG. The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer. *Nat Rev Cancer*. 2008;8:865–873.
44. Yang C, Hong CS, Prchal JT, et al. Somatic mosaicism of EPAS1 mutations in the syndrome of paraganglioma and somatostatinoma associated with polycythemia. *Hum Genome Var*. 2015;2:15053.
45. Letouzé E, Martinelli C, Loriot C, et al. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*. 2013;23:739–752.
46. Xiao M, Yang H, Xu W, et al. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev*. 2012;26:1326–1338.
47. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462:739–744.
48. Gaal J, Burnichon N, Korpershoek E, et al. Isocitrate dehydrogenase mutations are rare in pheochromocytomas and paragangliomas. *J Clin Endocrinol Metab*. 2010;95:1274–1278.
49. Zhu J, Blenis J, Yuan J. Activation of PI3K/Akt and MAPK pathways regulates Myc-mediated transcription by phosphorylating and promoting the degradation of Mad1. *Proc Natl Acad Sci USA*. 2008;105:6584–6589.
50. Burnichon N, Buffet A, Parfait B, et al. Somatic NF1 inactivation is a frequent event in sporadic pheochromocytoma. *Hum Mol Genet*. 2012;21:5397–5405.
51. Burnichon N, Vescovo L, Amar L, et al. Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. *Hum Mol Genet*. 2011;20:3974–3985.
52. Burnichon N, Cascón A, Schiavi F, et al. MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2012;18:2828–2837.
53. Stenman A, Welander J, Gustavsson I, et al. HRAS mutation prevalence and associated expression patterns in pheochromocytoma. *Genes Chromosomes Cancer*. 2016;55:452–459.
54. Fishbein L, Khare S, Wubbenhorst B, et al. Whole-exome sequencing identifies somatic ATRX mutations in pheochromocytomas and paragangliomas. *Nat Commun*. 2015;6:6140.
55. Garinet S, Néou M, de La Villéon B, et al. Calling chromosome alterations, DNA methylation statuses, and mutations in tumors by simple targeted next-generation sequencing: a solution for transferring integrated pangenomic studies into routine practice? *J Mol Diagn*. 2017;19:776–787.
56. Mohan DR, Lerario AM, Else T, et al. Targeted assessment of *G0S2* methylation identifies a rapidly recurrent, routinely fatal molecular subtype of adrenocortical carcinoma. *Clin Cancer Res*. 2019;25:3276–3288.
57. Gicquel C, Bertagna X, Gaston V, et al. Molecular markers and long-term recurrences in a large cohort of patients with sporadic adrenocortical tumors. *Cancer Res*. 2001;61:6762–6767.
58. Lerario AM, Worden FP, Ramm CA, et al. The combination of insulin-like growth factor receptor 1 (IGF1R) antibody cixutumumab and mitotane as a first-line therapy for patients with recurrent/metastatic adrenocortical carcinoma: a multi-institutional NCI-sponsored trial. *Horm Cancer*. 2014;5:232–239.
59. Fassnacht M, Berruti A, Baudin E, et al. Linsitinib (OSI-906) versus placebo for patients with locally advanced or metastatic adrenocortical carcinoma: a double-blind, randomised, phase 3 study. *Lancet Oncol*. 2015;16:426–435.

96

Primary Aldosteronism

Anand Vaidya, Maria-Christina Zennaro, and Michael Stowasser

OUTLINE

Pathophysiology, 1599
Pathogenesis And Genetics, 1599
Familial Forms of Primary Aldosteronism, 1600
Somatic Mutations in Unilateral Primary Aldosteronism, 1601
Somatic Mutations in Bilateral Adrenal Hyperplasia, 1602
New Diagnostic and Therapeutic Opportunities Associated with Genetic Knowledge, 1603
Prevalence, 1603
Clinical Diagnosis, 1603
Symptoms and Signs of Primary Aldosteronism, 1603
Assays Used in the Diagnostic Workup of Primary Aldosteronism, 1604
Diagnostic Workup for Primary Aldosteronism, 1604
Treatment, 1608

PATHOPHYSIOLOGY

Primary aldosteronism is a syndrome of inappropriate and relatively nonsuppressible renin-independent aldosterone production that persists despite expansion of effective arterial circulating volume and hypokalemia.

This autonomous production of aldosterone results in excessive activation of the mineralocorticoid receptor (MR) in the principal cell in the collecting duct, thereby inducing increased epithelial sodium channel mediated sodium reabsorption and consequent volume expansion. To maintain electroneutrality, the reabsorption of sodium in the distal nephron is paired with an obligate excretion of either potassium or hydrogen ions. This pathophysiology begets a vicious cycle: increased distal nephron sodium reabsorption contributes to increases in effective arterial circulating volume expansion, in turn increasing blood pressure and glomerular filtration, thereby resulting in suppression of renin and angiotensin II, which begets further increases in distal nephron sodium delivery, hence further increasing distal nephron sodium reabsorption.[1,2]

The cardinal features of primary aldosteronism (PA) include a suppression of baseline renin secretion, an inability to stimulate renin normally (despite provocation by upright posture, volume contraction or sodium depletion, or angiotensin-converting enzyme [ACE] inhibition), and inappropriate and nonsuppressible aldosterone production despite renin suppression.[3,4] Importantly, hypertension and hypokalemia are not obligate characteristics of PA; rather, they are dependent features that occur depending on the degree of intraarterial volume expansion and distal nephron sodium delivery.

The pathophysiology of PA extends beyond mechanisms in the kidney alone. The MR is expressed in the myocardium, vascular endothelium, smooth muscle, and other tissues. Animal experiments have elegantly demonstrated that, in contrast to the renin-dependent aldosteronism that is vital for the physiologic response to volume depletion, activation of the cardiovascular MR in the context of volume expansion and/or sodium-avid conditions results in myocardial fibrosis, necrosis, and clinical features of cardiovascular disease.[5-10] The molecular mechanism for this paradox remains unresolved: why is activation of the MR by aldosterone in volume/sodium-depleted states physiologic, whereas activation of the MR by aldosterone in volume/sodium-expanded states is pathophysiologic? However, observations from human studies have echoed these basic findings. Clinical trials in patients with heart failure with reduced ejection fraction (a state of increased left ventricular volume/pressure) have shown that MR antagonists dramatically reduce adverse clinical events and mortality.[11,12] Observational studies have suggested that patients with PA, compared with patients with essential hypertension, have a higher risk for coronary artery disease, congestive heart failure, left ventricular hypertrophy, atrial fibrillation, stroke, diabetes mellitus, metabolic syndrome, decreased bone density, and kidney disease.[2] Importantly, many of these observational studies demonstrated that the increased cardiometabolic risk associated with untreated PA, compared with essential hypertension, persists even when patients are matched for blood pressure and duration of hypertension.

PATHOGENESIS AND GENETICS

Primary aldosteronism results from autonomous production of aldosterone in the adrenal cortex. Under normal circumstances, aldosterone biosynthesis in the adrenal cortex is tightly regulated by the renin-angiotensin system and extracellular potassium concentrations to maintain blood volume, sodium, potassium and hydrogen homeostasis. The zona glomerulosa expresses the enzyme aldosterone synthase, encoded by *CYP11B2*, which catalyzes the three final enzymatic reactions leading to aldosterone biosynthesis.[13] Due to the expression of a large number of potassium channels[14] and the Na^+/K^+-ATPase, unstimulated zona glomerulosa cells show a strongly negative membrane potential, which parallels the potassium equilibrium potential. Angiotensin II and increased plasma potassium lead to cell membrane depolarization, followed by opening of voltage-gated calcium channels and increased intracellular calcium concentrations. Angiotensin II also acts by binding to its membrane G protein-coupled angiotensin II type 1 receptor that activates signaling via the inositol triphosphate pathway, increasing intracellular calcium levels from intracellular calcium stores. The increase in cytosolic calcium activates calcium signaling via calmodulin and calcium/calmodulin-dependent protein kinases,

TABLE 96.1 Genetic Forms of Primary Aldosteronism

Disease	Age of Onset	Specific Features	Gene	Transmission	Specific Treatment
Familial hyperaldosteronism type I	In most cases before age 20 yrs	Increased risk of cerebrovascular events before age 30 yrs; aldosterone suppressible by dexamethasone; increased urinary levels of hybrid steroids 18-oxocortisol and 18-hydroxycortisol	Chimeric *CYP11B1/B2*	AD	Glucocorticoids, MRA
Familial hyperaldosteronism type II	Variable, mostly in childhood	Usually mild, bilateral hyperaldosteronism	*CLCN2*	AD	MRA
Familial hyperaldosteronism type III	Before age 20 yrs Variable in mild forms	Massive bilateral adrenal hyperplasia in severe cases	*KCNJ5*	AD	MRA Bilateral adrenalectomy in severe cases
Familial hyperaldosteronism type IV	Variable, often in childhood	Developmental disorder in some cases	*CACNA1H*	AD	MRA
PASNA	Childhood	Seizures and neurological abnormalities	*CACNA1D*	*de novo* mutations	Calcium channel blockers, MRA
Familial hyperaldosteronism: genetic basis undefined*	Variable	None; diagnosis is based on two or more affected family members	To be determined*	Variable	Unknown

*The definition familial hyperaldosteronism reflects the clinical diagnosis. Genetic analysis allows identifying patients with chimeric *CYP11B1/B2, CLCN2, KCNJ5,* or *CACNA1H* mutations and reclassifying them as FH-I, -II, -II, or -IV, respectively. Other genes may be involved in cases without identified gene mutation.

PASNA, Primary Aldosteronism, Seizures and Neurologic Abnormalities; *AD*, autosomal dominant; *MRA*, mineralocorticoid receptor antagonists.

which is the major regulator of *CYP11B2* expression and aldosterone production.[15] In PA, regulatory mechanisms of aldosterone biosynthesis are bypassed by the chronic activation of intracellular calcium signaling resulting from different genetic defects.

Different genetic abnormalities have been identified in familial forms of PA and in aldosterone-producing adenoma (APA). While they occur as somatic mutations in APA, they are germline in familial cases and transmitted as an autosomal dominant trait. It is worth noting however, that somatic mutations also occur in the normal adrenal cortex in so-called aldosterone-producing cell clusters (APCCs) and in secondary aldosterone-producing nodules from adrenals with APA.

Familial Forms of Primary Aldosteronism

At least four different Mendelian forms of familial hyperaldosteronism (FH) have been reported (Table 96.1). They are defined as FH-I to FH-IV, depending on the underlying genetic defect. Familial forms occur in approximately 6% of patients with PA[16] and are transmitted as an autosomal dominant trait. Although these forms are rare, genetic screening should be offered to patients with early-onset PA before the age of 20 years and/or with a family history of PA, given the low cost and the benefits of early diagnosis in affected family members,[17] as deleterious cardiovascular effects due to aldosterone excess may occur well before patients become hypertensive.[18]

FH-I, also called glucocorticoid-remediable aldosteronism (OMIM #103900), is characterized by early and severe hypertension, usually before 20 years of age, biochemical abnormalities of PA, excessive production of the hybrid steroids 18-hydroxycortisol and 18-oxocortisol, and bilateral hyperplasia, in rare cases associated with adrenal nodules; remarkably, symptoms are relieved by treatment with dexamethasone.[19] FH-I is due to the presence of a chimeric gene resulting from an unequal crossing over between *CYP11B2* and the highly homologous *CYP11B1* gene coding 11β-hydroxylase, which is responsible for the last steps of cortisol biosynthesis in the adrenal zona fasciculata. The chimeric gene contains the regulatory regions of *CYP11B1* juxtaposed to *CYP11B2*-specific coding sequences. Therefore, aldosterone synthase is expressed ectopically throughout the adrenal cortex under the control of adrenocorticotropic hormone (ACTH) rather than angiotensin II.[20] Exogenous glucocorticoids are effective in reducing aldosterone synthase expression and aldosterone production. Low-dose glucocorticoids have been shown to be sufficient to normalize blood pressure and plasma potassium levels without affecting hormonal parameters, and may provide prolonged control of hypertension.[21] Mineralocorticoid antagonists or amiloride may be added in order to lower doses of glucocorticoids if required. In children, eplerenone is preferable, to avoid the side effects of glucocorticoids (retarded growth) or spironolactone (antiandrogen effects).[22] The prevalence of FH-I has been estimated to be approximately 0.5% to 1.0% of PA in the hypertensive adult population,[16,23] but its frequency appears to be higher in hypertensive children, where a prevalence of 3% of the chimeric *CYP11B1/CYP11B2* gene has been reported.[24] Although FH-I is a rare disease, it is associated with high morbidity and mortality before age 40 years from early-onset hemorrhagic stroke and ruptured intracranial aneurysms,[25] criteria that may orient diagnosis in affected patients. Abnormalities in cardiac structure and function may appear before the rise in blood pressure in carriers of the chimeric gene.[18] The genetic diagnosis is usually made by long-range polymerase chain reaction (PCR) or Southern blot, while urinary levels of 18-oxocortisol and 18-hydroxycortisol and the dexamethasone suppression test do not show as high sensitivity and specificity.[17]

FH-II (OMIM #605635), first described in 1991,[26] was originally defined as PA occurring in two or more family members (i.e., FH) that is not glucocorticoid-remediable and is not associated with the hybrid gene mutation. By that definition, it was reported to be the most frequent form of FH, accounting for approximately 1.2% to 6% of cases,[16,27] with a presentation undistinguishable from that of sporadic PA. Patients showed heterogeneous clinical presentations with varying aldosterone response on postural test and to angiotensin II and different subtypes of PA even within the same family.[28] Recently, germline mutations in *CLCN2*, coding for the chloride channel ClC-2, have been identified in a large Australian family with FH-II and in early-onset PA.[29,30] A recurrent germline *CLCN2* mutation was found in the first reported FH-II family, as well as in four additional unrelated patients,[29] while other *CLCN2* mutations were found in unrelated patients with early-onset PA.[29,30] This has led to the term "FH-II" being restricted to PA caused by germline *CLCN2* mutations. *CLCN2* encodes a voltage-gated chloride channel composed of 18 transmembrane helices and an intracellular N-terminus and C-terminus. Mutations are mostly located in particular regions of the protein and induce a loss of the voltage gating of the channel. Thus, the channel is constitutively open at the zona glomerulosa resting potential, leading to sustained chloride efflux, followed by cell membrane depolarization, activation of calcium signaling, and increased expression of aldosterone synthase and autonomous aldosterone production.[29,30] In addition to *CLCN2* mutations, some patients previously diagnosed with FH-II harbor mutations in genes identified in other familial forms and are therefore reclassified after genetic diagnosis. In particular, germline *KCNJ5* mutations are found in patients with a moderate phenotype resembling FH-II,[31,32] as well as mutations in *CACNA1H*.[33] Somatic mutations of *KCNJ5* were also reported in APA from patients diagnosed with FH-II according to the clinical definition.[32] Hence, the phenotype of FH includes different subtypes of FH, which, in the absence of discriminating clinical features, require genetic testing for classification. In addition, there may be some cases in which FH results from familial aggregation of sporadic PA, given the high frequency of the condition in patients with hypertension.

FH-III (OMIM #613677) is a rare form of FH characterized in most patients by severe hypertension early in childhood that is refractory to treatment and associated with severe hypokalemia and massive bilateral adrenal hyperplasia.[34] The hybrid steroids 18-oxocortisol and 18-hydroxycortisol in urine are increased, but aldosterone production is not suppressed by dexamethasone.[34] FH-III is due to recurrent mutations of the *KCNJ5* gene coding for the G protein-activated inward rectifier potassium channel GIRK4.[35] Some heterogeneity in the clinical presentation, particularly in the severity of the disease, has been described, which appears to be associated with the type of mutation.[31,32] Some patients present a particularly mild form of PA, diagnosed in young adulthood, with or without hypokalemia with no evident signs of hyperplasia at imaging. This form responds well to pharmacological treatment with MR antagonists and had often been previously diagnosed as FH-II.[31,32] Almost all *KCNJ5* mutations are located within or near the selectivity filter of the GIRK4 potassium channel. They induce a change in ion selectivity, with loss of potassium selectivity and increased sodium conductance. This leads to increased sodium influx into the cell, followed by opening of voltage-gated Ca^{2+} channels, activation of Ca^{2+} signaling, and stimulation of aldosterone biosynthesis.[35,36] Pathological examination of the adrenal gland in two affected individuals after adrenalectomy has shown adrenal enlargement with loss of zonation and expression of aldosterone synthase throughout the adrenal cortex.[37] Coexpression of 11β-hydroxylase or 17α-hydroxylase with aldosterone synthase in some adrenal areas explains the increased levels of hybrid steroids in these patients. In one patient carrying a *KCNJ5* mutation, cosecretion of aldosterone and cortisol has been reported.[38]

A fourth form of FH, FH-IV (OMIM #617027), has been discovered by whole-exome sequencing of patients with early-onset PA before the age of 10 years.[33,39] A recurrent mutation in the *CACNA1H* gene was identified in five patients, with autosomal dominant transmission in four families and a *de novo* mutation in one case.[39] In parallel, different *CACNA1H* mutations were reported in patients with early-onset PA or with mild PA diagnosed as FH-II.[33] In some cases, *CACNA1H* mutations are associated with developmental disorders or attention deficit.[33,39] In most cases, no adrenal abnormalities are observed on imaging, but in one case, the pathological examination of the resected adrenal has shown micronodular adrenal hyperplasia.[39] *CACNA1H* encodes the pore-forming α1 subunit of the T-type voltage-dependent calcium channel Cav3.2. The channel is composed of four repeat domains (I–IV), each consisting of six transmembrane segments (S1–S6), and mediates voltage-dependent calcium influx into the cell upon membrane depolarization. Cav3.2 mutations affect highly conserved amino acids of the protein, modifying important channel properties. In particular, they shift the voltage dependence of the channels towards more negative membrane potentials, inducing channel opening at less depolarized membrane potentials, and/or delay the voltage-dependent inactivation of the channel. The functional consequence is sustained calcium influx into the cell, activating calcium signaling and aldosterone biosynthesis.[33,39]

A particular form of PA that is not inherited, but rather due to germline mutations affecting another calcium channel, is PASNA (Primary Aldosteronism, Seizures, and Neurologic Abnormalities; OMIM #615474). In this condition, mutations affect the *CACNA1D* gene, coding for the voltage-dependent L-type calcium channel subunit alpha-1D.[40] Those patients show early onset PA in the context of a complex neurological disorder, including cerebral palsy and epileptic seizures. Patients have severe hypertension and hypokalemia, without adrenal hyperplasia on imaging, and carry *de novo* germline mutations of the *CACNA1D* gene. The disease has been described in only four patients so far, in one case associated with congenital hyperinsulinism, and its prevalence is currently unknown.[40-42] The pathogenic mechanism whereby *CACNA1D* mutations lead to PA is similar to FH-IV: the mutations affect the channel properties, leading to opening of the channel at less depolarized membrane potentials and activation of calcium signaling and aldosterone biosynthesis.

Somatic Mutations in Unilateral Primary Aldosteronism

The majority of cases of PA occur sporadically and are due to a unilateral APA or bilateral adrenal hyperaplasia. From a clinical point of view, the distinction is made based on unilateral versus bilateral aldosterone secretion defined by adrenal venous sampling (AVS). From a pathological point of view, it has become apparent that unilateral forms encompass a large variety of morphological alterations, ranging from an isolated APA to multinodular adrenals. Recurrent somatic mutations are found in a majority of APAs, as well as aldosterone-producing structures in adrenal glands from patients with PA.[43-46] They affect genes involved in familial forms of the disease (*KCNJ5*, *CACNA1H*, *CLCN2*), as well as genes coding for ATPases (*ATP1A1,* coding for the α1 subunit of the Na^+,K^+-ATPase, and *ATP2B3,* coding for the plasma membrane calcium-transporting ATPase 3 [PMCA3]) and *CACNA1D*, coding for the α1 subunit of the L-type Ca^{2+} channel Cav1.3. In certain cases, somatic mutations have been identified in genes mainly involved in adrenal disorders associated with cortisol excess. Indeed, 2% to 5% of APAs carry somatic mutations in *CTNNB1*, coding for β-catenin.[47-49] Somatic mutations affecting *PRKACA* (coding for catalytic alpha subunit of the cAMP-dependent protein kinase) have been described in rare cases as well, and may be associated with cortisol excess.[50]

Somatic mutations affect specific conserved amino acids of the proteins and lead to major functional modifications. They can be classified in two major categories: those that lead to cell membrane depolarization by altering intracellular ion homeostasis (*KCNJ5*, *ATP1A1*, *CLCN2*), and those that directly increase intracellular calcium levels by affecting calcium channels and pumps (*ATP2B3*, *CACNA1D*, *CACNA1H*) (Fig. 96.1).

As described for FH-III, the most prevalent mutations in *KCNJ5* lie within or near the selectivity filter of GIRK4 and modify the ion selectivity of the channel, rendering it permeable to sodium.[35,36] Mutations in *ATP1A1* decrease the activity of the sodium potassium ATPase and reduce the affinity for K^+,[51] while mutations in

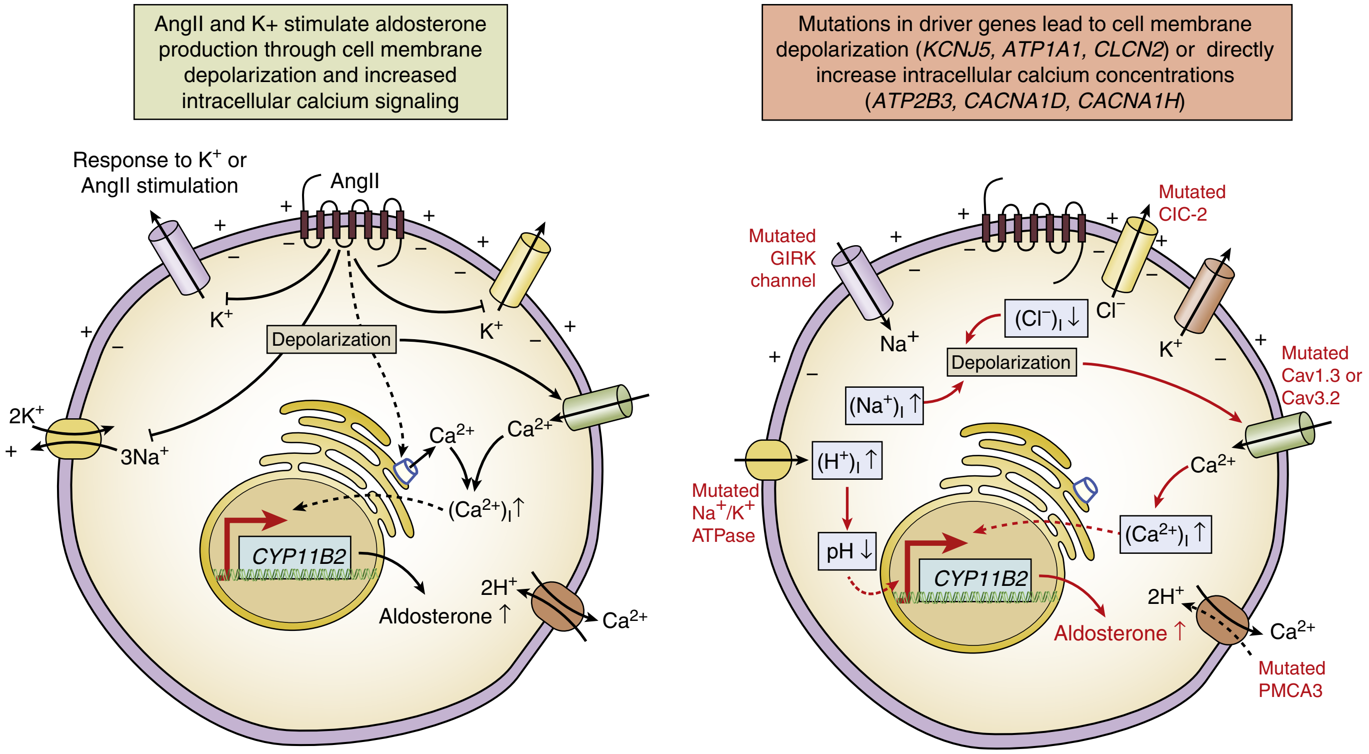

AngII and K+ stimulate aldosterone production through cell membrane depolarization and increased intracellular calcium signaling
Mutations in driver genes lead to cell membrane depolarization (KCNJ5, ATP1A1, CLCN2) or directly increase intracellular calcium concentrations (ATP2B3, CACNA1D, CACNA1H)
Response to K+ or AngII stimulation
AngII
K+
K+
Depolarization
2K+
3Na+
Ca2+
Ca2+
(Ca2+)i ↑
CYP11B2
2H+
Aldosterone ↑
Ca2+
Mutated GIRK channel
Mutated ClC-2
Na+
(Cl−)i ↓
Cl−
K+
Depolarization
(Na+)i ↑
Mutated Cav1.3 or Cav3.2
(H+)i ↑
Ca2+
Mutated Na+/K+ ATPase
pH ↓
(Ca2+)i ↑
CYP11B2
2H+
Aldosterone ↑
Ca2+
Mutated PMCA3

Fig. 96.1 Different mechanisms lead to APA formation. Following increase of extracellular potassium or AngII binding to its AT1R receptor, inhibition of potassium channels leads to cell membrane depolarization and opening of voltage gated calcium channels. Increased intracellular calcium concentrations lead to activation of calcium signaling, which induces aldosterone biosynthesis. Binding of AngII to AT1R also inhibits Na+/K+-ATPase, inducing cell membrane depolarization, while AT1R signaling through Gαq promotes the release of calcium from intracellular calcium stores (left panel). Mutations in *KCNJ5*, coding for the potassium channel GIRK4, change the channel's ion selectivity, promoting sodium entry into the cell instead of potassium efflux, leading to cell membrane depolarization and opening of voltage gated calcium channels. Mutations in *ATP1A1*, coding for the α1 subunit of the Na+/K+-ATPase, lead to increased aldosterone production by increasing intracellular Na+ and H+ concentrations, resulting in cell membrane depolarization and decrease of intracellular pH. Mutations in *CLCN2*, coding for the chloride channel ClC-2, lead to increased chloride efflux from the cell, similarly leading to cell membrane depolarization. Mutations in *CACNA1D*, encoding the Cav1.3 calcium channel, *CACNA1H*, encoding the Cav3.2 calcium channel, and *ATP2B3*, coding for the Ca2+ pump PMCA3 directly affect intracellular calcium concentrations and thus aldosterone biosynthesis (right panel). Ca2+, calcium; K+, potassium; Na+, sodium; Cl-, chloride. (Modified from Zennaro, et al. *Endocr Rev* **38**, 516-537 (2017).)

CLCN2 act by increasing chloride efflux as in FH-II.[52] On the other hand, *ATP2B3* mutations decrease pump activity of the PMCA3, impairing calcium export. In addition, *ATP2B3* mutations may also induce a sodium leak, inducing cell membrane depolarization as well as a calcium leak through mutated pumps.[53] *CACNA1D* and *CACNA1H* mutations have similar functional consequences. They shift the voltage dependency of the mutated calcium channels towards more negative potentials, therefore leading to the channels opening at less depolarized potentials upon stimulation, or delay their voltage-dependent inactivation.[33,40,54] *In fine*, both these mechanisms increase intracellular calcium concentrations and lead to activation of calcium signaling, activating molecular cascades triggering increased expression of aldosterone synthase and autonomous aldosterone production.

The most frequent somatic mutations found in APA are *KCNJ5* mutations. They account for around 40% of adenomas,[55,56] with higher frequencies in patients from East Asia.[56] These figures were not changed by the introduction of CYP11B2 immunohistochemistry-guided next-generation sequencing.[43-46] In contrast, the prevalence of somatic mutations in other genes increased significantly following next-generation sequencing, which allows extended coverage of genes. *CACNA1D* mutations are now identified in 14% to 42% of APAs, *ATP1A1* mutations in 5% to 17% of APAs, and *ATP2B3* mutations in 4% to 10% of patients, depending on the population.[43-46] *CACNA1H* and *CLCN2* mutations are found in rare cases.[52,57]

Somatic Mutations in Bilateral Adrenal Hyperplasia

Idiopathic hyperaldosteronism, or bilateral adrenal hyperplasia, is diagnosed on the basis of bilateral secretion of aldosterone at AVS, and treatment with MR antagonists is the recommended therapeutic approach. A recent study had the opportunity to investigate 15 adrenals from patients with idiopathic hyperaldosteronism who had undergone adrenalectomy for exceptional reasons, including difficult-to-control hypertension or hypokalemia and unwillingness for long-term medical treatment. An increased number of so-called APCCs of the adrenal cortex was identified, which carried somatic mutations in *CACNA1D*, as well as a *KCNJ5* mutation in a micronodule.[58] APCCs are clusters of cells located in the zona glomerulosa, express CYP11B2, and are composed of an outer part of compact cells expressing disabled-2, a marker of the zona glomerulosa, and an inner part of larger cells not expressing disabled-2, suggesting that they may be intermediate structures between the zona glomerulosa and the zona fasciculata.[59,60,70,72] In normal adrenals, the number of APCCs increases with age and progressively replaces the continuous zona glomerulosa observed in younger individuals; this is associated with an age-related dysregulation of aldosterone production.[61] The peculiar feature of APCCs is that they carry somatic mutations even in normal adrenal glands, and therefore have been suggested to represent a possible precursor of APA.[60] Whether or not an accumulation of APCCs carrying somatic mutations underlies the development of idiopathic hyperaldosteronism in general, or rather is a contributor to autonomous aldosterone production in hyperplastic glands, requires further studies. Indeed, APCCs are also present in adrenals with APA and may carry different somatic mutations in the same adrenal gland.[45]

New Diagnostic and Therapeutic Opportunities Associated with Genetic Knowledge

One important question is whether genetic knowledge may improve diagnosis and treatment of patients with PA. There is no doubt that genetic diagnosis in familial forms allows for earlier detection of affected family members and better prevention of cardiovascular complications that may occur before hypertension onset.[18] On the other hand, identification of somatic mutations in resected adrenal glands has not provided benefit to patients so far. Identification of surrogate biomarkers of the presence of somatic mutations may eventually allow simplification of the diagnostic procedure by identifying patients likely to carry a unilateral APA to be selected for AVS. Some attempts have indicated that steroid profiling in peripheral vein samples may provide good information on the presence of somatic mutations. Indeed, levels of the hybrid steroids 18-oxocortisol and 18-hydroxycortisol are significantly higher in patients with *KCNJ5* mutations compared with other mutational groups,[62,63,45] and steroid fingerprints have been applied for distinguishing different genetic groups.[62] Similarly, the pharmacological properties of mutated GIRK4 channels are currently used to test the hypothesis that macrolides may be used for diagnostic purposes.[64] These approaches may be used together with peculiar imaging characteristics of APA carrying *KCNJ5* mutations on computed tomography, which are characterized by larger tumor size and lower precontrast Hounsfield units,[48] as well as the high prevalence in young patients and in women,[55,56] to elaborate a decision tree guiding diagnostic assessment. It is therefore the purpose of ongoing genetic studies not only to understand the development, but also to allow for personalized approaches to diagnosis and management of patients with PA.

KEY POINTS

- Mutations in different genes coding for ion channels (*KCNJ5*, *CACNA1D*, *CACNA1H*, *CLCN2*) and ATPases (*ATP1A1*, *ATP2B3*) are found in familial hyperaldosteronism and in aldosterone-producing adenoma. They lead to cell membrane depolarization, followed by opening of voltage-gated calcium channels, or directly increase intracellular calcium concentrations, with activation of calcium signaling leading to increased aldosterone biosynthesis.

PREVALENCE

Iterative population-based and convenience sample-based studies have shown that the prevalence of PA is high and largely underestimated. The challenge in establishing accurate prevalence estimates has been the lack of a histopathologic or other gold-standard reference diagnostic. Rather, the diagnosis of PA is made using a variety of arbitrary screening and confirmatory diagnostic thresholds; variations in these diagnostic thresholds can easily modulate prevalence estimates.

Population-based studies that used the aldosterone-to-renin ratio (ARR) have suggested that 7% to 12% of the hypertensive population may have PA,[65,66] and that higher aldosterone levels in the context of suppressed renin activity were associated with higher risk for incident coronary artery calcification and all-cause mortality.[67] Many international studies have used a two-tiered approach to estimate prevalence, wherein the ARR is first used to identify a "positive" screen, which is then confirmed using a variety of physiologic or dynamic confirmatory tests designed to evaluate aldosterone suppressibility.[68-76] Collectively, these studies have shown that overt PA can be detected in 2% to 19% of hypertensive individuals, with the prevalence paralleling the severity of hypertension. In contrast, some studies have estimated prevalence by conducting dynamic aldosterone suppression tests in hypertensive individuals without antecedent ARR screening and have reported much higher prevalence estimates, presumably because many ARRs are considered to be falsely "negative" owing to the variability of aldosterone production and the arbitrary nature of screening thresholds.[77] For example, 28% to 30% of participants with hypertension were confirmed to have PA when undergoing direct dexamethasone-fludrocortisone suppression testing,[78] 29% of patients with resistant hypertension were confirmed to have PA when undergoing direct saline suppression testing,[79] and 24% of patients with resistant hypertension were confirmed to have PA when undergoing direct oral sodium suppression testing.[80] Beyond these crude prevalence estimates born out of reliance on arbitrary thresholds for diagnosis, studies have shown a continuum of renin-independent aldosterone production in hypertension that is associated with the severity of hypertension and kaliuresis,[80,81]; thus, suggesting that the prevalence of PA captured using binary diagnostics likely underestimates the true burden of pathogenic aldosteronism.

The clinical relevance of this high prevalence of PA and the continuum of renin-independent aldosterone is the association of this biochemical phenotype with incident cardiometabolic disease that can be mitigated by MR antagonists[82-84] or, in patients with unilateral PA, unilateral adrenalectomy. Intervention studies have consistently shown that, the greater the magnitude of the renin-independent aldosterone production, the greater the efficacy of MR antagonists in lowering blood pressure.

Finally, although clinical indications for PA screening generally focus on individuals with hypertension, accruing evidence indicates that a PA phenotype can be detected even among those with normal blood pressure (sometimes referred to as subclinical or nonclassical PA[85]). Multiple studies that have used dynamic confirmatory testing have observed that approximately 9% to 14% of normotensive participants have nonsuppressible aldosterone production that meets conventional definitions for diagnosing PA.[80,86-90] In parallel with these observations, population-based studies have shown that this biochemical phenotype of renin-independent aldosterone production in normotensive people is associated with an increased risk for developing incident hypertension,[65,91-93] thereby suggesting a pathologic continuum of PA that may have its origins in normotension and may contribute to the pathology of hypertension.

CLINICAL DIAGNOSIS

Symptoms and Signs of Primary Aldosteronism

The great majority of patients with PA are hypertensive. Within some families with inheritable forms, blood pressure levels in members testing positive for the causative mutation and demonstrating biochemical evidence of PA have varied from normal to severely elevated.[94,95] Occasional normotensive patients with apparently nonfamilial PA, usually heralded by hypokalemia rather than hypertension, have also been described. Hypokalemia is now known to present in a minority (approximately 20%–25%) of patients, and hence its absence does not exclude PA.[96,97] When it is present, associated symptoms may include muscle weakness, cramping, tetany, paresthesias, palpitations, polyuria, and nocturia. Nocturia often occurs even when plasma potassium is normal and may reflect the rise in atrial natriuretic peptide levels that has been associated with PA or low intracellular renal levels of potassium resulting from increased release in an attempt to maintain normokalemia. Other common symptoms in patients with PA include lethargy, impaired mental concentration, and mood disturbances.[98-101]

Hypertension and symptoms may be masked during pregnancy[102,103] due to the high circulating levels of placental progesterone (which is an MR antagonist). In other women, development of PA due to APA during pregnancy has been linked to the presence of somatic activating mutations in *CTNNB1* that encodes β-catenin in the Wnt cell differentiation pathway (possibly explaining tumor development) and increased expression of *LHCGR* (encoding lutotropin-choriongonadotropic hormone receptor) and *GNRHR* (encoding gonadotropin-releasing hormone receptor), suggesting that high levels of human chorionic hormone during pregnancy may result in stimulation of aldosterone production by these APAs via aberrant receptor activation.[47]

Once thought not to occur in PA, malignant hypertension has been reported by several investigators.[104,105] The intrarenal ischemia that occurs in this may lead to unsuppression of renin, which could result in the diagnosis of PA being missed.

Assays Used in the Diagnostic Workup of Primary Aldosteronism

Highly reproducible assays are essential for all stages of the diagnostic workup of PA, including screening, confirmatory suppression testing, and AVS.

Aldosterone Assays. Accurate and reproducible measurement of plasma or serum aldosterone is a challenge for clinical laboratories, because it circulates at relatively low (picomolar) concentrations. Although radioimmunassays were until recently the most widely used method of measuring aldosterone, these have now been replaced in many laboratories by faster and less labor-intensive automated platforms using immunometric assay technology. Concerns have been raised regarding the accuracy of immunoassays in general and the need for high-affinity, highly specific antibodies and meticulous laboratory technique. Lack of specificity probably explains why newer, highly accurate and reproducible methods of measuring aldosterone within the clinically relevant range using high-performance liquid chromatography and tandem mass spectrometry (LC-MS/MS) generate aldosterone levels that are substantially lower than those measured by immunoassays within the same plasma samples,[106,107] necessitating adjustment of diagnostic thresholds such as the ARR and confirmatory suppression testing.[108,109] These methods, previously laborious and time-consuming, now employ semiautomated technology, which allows high sample throughput and faster generation of assay results, making them suitable for the clinical setting. While running costs are relatively low, the initial cost outlay associated with purchasing the equipment represents a barrier to wide application of LC-MS/MS, a situation that may change with time.

Renin Assays. The most commonly utilized renin assay methods are the measurement of plasma renin activity (PRA) or of direct (or active) renin concentration (DRC). Each has its advantages and disadvantages. PRA measures (usually by radioimmunoassay, but in some laboratories by LC-MS/MS) the amount of angiotensin I generated by the action of renin on its substrate, angiotensinogen, in a defined volume of plasma per unit of time, and hence is a measure of renin enzyme activity. For samples with levels less than 1 ng/mL/min, reproducibility of the assay can be maintained by extending the incubation time from the recommended 90 minutes to 18 hours.[110]

The principal advantage of DRC, which measures the plasma concentration of the active (cleaved) form of renin, is that it can be undertaken by automated immunometric assay,[111] which is much faster and less labor-intensive than PRA. Some laboratories offer combined aldosterone and renin measurement on the same sample using automated immunometric technology. However, concern has been raised over the lack of reliability and reproducibility of these assays at the lower end of the human reference range, which is especially pertinent for the diagnostic workup of PA, where renin levels are suppressed. In this situation, small changes in renin can result in large changes of the ARR,[112] and it could therefore be argued that it is more important to measure renin accurately than aldosterone when screening for PA.

A further concern for DRC measurement is that it may not reflect the true activity of the renin-angiotensin system as well as PRA. Renin does not affect blood pressure or aldosterone synthesis directly, but only through angiotensin II, concentrations of which may be better reflected by PRA. For example, with administration of exogenous estrogen (in oral contraceptive or hormone replacement preparations), the concentration of angiotensinogen (substrate) increases, and DRC falls to maintain angiotensin II concentrations in the normal range, whereas PRA remains relatively unaffected.[113,114] On the other hand, advocates for automated immunometric renin assays have reported that they appear to have better interlaboratory agreement.[115]

A potential alternative to the measurement of renin in the diagnostic workup of PA is recently reported mass spectrometric methods of assaying angiotensin I (generated from angiotensinogen as per the PRA approach) or, more promising still, angiotensin II (measured during a period of equilibrium of formation versus metabolism). Unlike renin, angiotensin II directly regulates aldosterone synthesis via AT1 receptors on adrenal zona glomerulosa cells, and its measurement is therefore arguably more physiologically relevant in clinically assessing patients for the presence of angiotensin II-independent aldosterone overproduction.[116]

Diagnostic Workup for Primary Aldosteronism

This is divided into three main phases: screening, confirmatory testing, and determining the subtype of PA. All require expertise for optimal performance and interpretation of results and are the subject of considerable ongoing research efforts towards refinement.

Screening. Screening of at-risk populations for PA is primarily geared towards identifying those patients who do not have this condition, and therefore are not required to undergo confirmatory testing and subtype differentiation (both of which involve procedures that are relatively invasive and expensive). It is therefore important that the screening test is highly sensitive, while retaining a reasonable degree of specificity.

Most patients with hypertension, and especially those with moderate to severe forms, should be regarded as candidates for screening. In addition, screening is recommended for patients with hypertension and spontaneous or diuretic induced hypokalemia, adrenal incidentaloma, sleep apnea, or a family history of early-onset hypertension or cerebrovascular accident at a young age (<40 years), and all first-degree relatives of patients with PA.[22] Some investigators suggest screening of all hypertensives, given the high prevalence of this condition among the hypertensive population and the reported short- and long-term benefits from specific surgical or medical treatment.[117] Screening before commencement of antihypertensive medications avoids potentially confounding effects of these agents on renin and aldosterone levels and permits earlier institution of specific treatment, which has been shown to be a major determinant of benefit achieved (including the chance of cure of hypertension following unilateral adrenalectomy for unilateral forms of PA).[118] Contrary to previous thinking, it is no longer considered appropriate to restrict screening to hypertensive patients with hypokalemia, as this would lead to as many as 80% of patients with PA being missed.[96]

Currently the most popular method of screening for PA involves the measurement of the plasma ARR. This is based on the lack of sensitivity of the demonstration of hypokalemia or of frankly elevated plasma aldosterone levels,[97,119] and also the lack of specificity of suppressed renin levels (which are also observed in patients receiving β-adrenergic or nonsteroidal antiinflammatory agents, ingesting a high dietary salt intake, of advanced age, or with chronic kidney disease or other varieties of low renin hypertension such as Liddle syndrome, hypertensive forms of congenital adrenal hyperplasia, familial hypertension and hyperkalemia, the syndrome of apparent mineralocorticoid excess, ectopic ACTH syndrome, deoxycortisone-secreting tumor, or glucocorticoid resistance states). Normal aldosterone levels could be viewed as "inappropriately normal" in the face of renin suppression, because individuals without PA should exhibit low plasma aldosterone levels in such situations. Raised aldosterone levels also lack specificity, as they may be secondary to activation of the renin-angiotensin system (secondary hyperaldosteronism) due, for example, to treatment with diuretics or renovascular hypertension. The ARR is much more sensitive for detection of PA than plasma potassium or aldosterone measured in isolation, and more specific than isolated renin, in that aldosterone is also suppressed in most other salt-dependent low-renin forms of hypertension, and hence the ARR is not elevated.

False-positive and false-negative ARRs can result from a variety of physiological and pharmacological factors (Tables 96.2 and 96.3), an appreciation of which aids substantially in the optimal performance of ARR testing and interpretation of results. The most common of these are medications, with β-adrenoceptor blockers, clonidine, α-methyldopa, nonsteroidal antiinflammatory drugs and estrogen-containing oral contraceptive agents or hormone replacement therapy prone to cause false positives, and diuretics (including potassium-sparing), ACE inhibitors, angiotensin receptor blockers (ARBs), dihydropyridine-type calcium channel blockers, and selective serotonin receptor inhibitors having the potential to cause false negatives.[120,121] Where possible, antihypertensives known to affect the ARR should be withdrawn (for at least 4 weeks for diuretics and at least 2 weeks for other agents) prior to screening. Antihypertensive medications that appear to have minimal, if any, effects on the ARR and can be used to control hypertension (where necessary) during screening and subsequent steps in the diagnostic workup of PA include verapamil SR, hydralazine (best used in combination with verapamil SR to avoid reflex tachycardia), prazosin, doxazosin, and moxonidine.[121] ARR testing while patients are still on interfering medications can still be informative. For example, a raised ARR with suppressed renin in a patient receiving an ACE inhibitor, ARB, or diuretic would be highly suggestive of PA, whereas a normal ARR in a patient on a β-blocker would make PA highly unlikely.

Other circumstances in which false positives can be encountered include the luteal phase of the menstrual cycle in premenopausal women, advancing age, chronic kidney disease, a high dietary salt intake, and the syndrome of familial hyperkalemic hypertension (Gordon syndrome). False negatives may occur during pregnancy, dietary salt restriction, vomiting or diarrhea, or uncorrected hypokalemia, and in patients with malignant hypertension or concomitant renovascular hypertension.[121]

The ARR has greater sensitivity if measured in the morning in an upright (e.g., seated) posture. Controlling for the above factors or at least taking them into account enhances the usefulness of ARR testing in selecting patients for further diagnostic workup. The ARR also shows good within-patient reproducibility under such conditions.[122]

One limitation of the ARR is that, in the presence of extremely low renin levels (for example, a DRC of 1.5 mU/L or PRA 0.1 ng/mL/h or less), the ARR may be elevated and thereby raise the possibility of PA even when plasma aldosterone is also very low (for example, 110 pmol/L or 4 ng/dL) and clearly not consistent with PA. While some investigators have suggested the inclusion of a minimum plasma aldosterone concentration (for example, 415 pmol/L or 15 ng/dL) to accompany an elevated ARR within the screening criteria in an attempt to overcome this problem,[123] others have found this to exclude many patients who have subsequently been confirmed as having PA.[119]

The ratio should be regarded as a screening test only, and should be measured more than once before deciding whether or not to go on to confirmatory testing for PA.

Confirmatory Testing. Because the ARR can be falsely positive even under ideal testing conditions, confirmatory testing is required to definitively confirm or exclude the diagnosis of PA. This important step allows selection of patients with confirmed PA for further testing, including AVS, to determine the subtype. Conversely, exclusion of PA spares patients from having to undergo this invasive and relatively expensive and difficult procedure.

Confirmatory testing usually involves measuring aldosterone during maneuvers designed to bring about complete suppression of circulating renin, and therefore of aldosterone's normal chronic regulator angiotensin II. Failure of aldosterone to suppress confirms aldosterone production that is relatively autonomous of angiotensin II, the *sine qua non* of PA.

Of the various approaches described, the most common confirmatory tests in use are the oral sodium loading test, the saline infusion testing, and the captopril challenge test. Fludrocortisone suppression testing (FST) is regarded by some investigators to be the most reliable, but is relatively labor-intensive and requires a 5-day admission to hospital, during which the plasma aldosterone response during 4 days of oral administration of fludrocortisone (0.1 mg 6 hourly) and oral salt loading (by high sodium diet and administration of slow-release sodium chloride tablets, 1800 mg three times daily with meals) is assessed.[124] The FST is considered diagnostic of PA if upright (seated) 1000-h plasma aldosterone fails to suppress to less than 6 ng/100 mL (165 pmol/L) at the conclusion of the test, provided that renin has been sufficiently suppressed, plasma potassium has been maintained in the normal range by adequate slow-release KCl supplementation (given 6 hourly), and there has been no rise in cortisol during the morning of the final day to suggest a rise in ACTH that might have prevented suppression of aldosterone.

For the oral salt loading test, patients are instructed to consume sufficient dietary salt to achieve a urine sodium excretion of over 200 mmol/day and are given adequate KCl supplementation to maintain normokalemia. A 24-h urinary aldosterone concentration of over 12 μg/day (33 nmol/day) on the third day is regarded as diagnostic.[123]

The saline infusion test (or saline suppression test [SST]) involves intravenous administration of 0.9% saline (usually 2 L over 4 hours) with measurement of plasma aldosterone at the end of the infusion, traditionally while patients are maintained in the recumbent position (unlike FST, in which the diagnostic plasma aldosterone level is collected while the patient is seated following at least 2 h of upright posture). Concentrations regarded as diagnostic for PA have varied from greater than 5 to greater than 10 ng/100 mL (>140 to >280 pmol/L).[125,126] A major advantage of this approach over FST is that it can be performed as an outpatient procedure. However, concerns have been raised over its lack of sensitivity, especially in patients in whom plasma aldosterone demonstrates responsiveness to the assumption of upright posture, as their recumbent aldosterone levels are lower than those in posture-unresponsive patients.[127] A recently reported approach has been to perform SST in the seated position, resulting in substantially improved sensitivity for detecting PA while retaining high specificity.[128] By this method, diagnostic cutoffs for normal

TABLE 96.2 Factors Causing False-Positive Aldosterone-to-Renin Ratios

	Effect on Aldosterone	Effect on Renin	Mechanism
β-blockers	↓	↓↓	Blockade of JG cell beta-adrenergic receptors
Clonidine α-methyldopa	↓	↓↓	Suppression of CNS sympathetic output
NSAIDs	↔	↓↓	Tendency to Na^+ retention and suppression of renal prostaglandins (which normally enhance release of renin) leads to renin suppression Tendency to K^+ retention prevents fall in aldosterone
Estrogen- and progestogen-containing agents*	↑	↔	Progestogenic component induces natriuresis, which stimulates renin and aldosterone Estrogenic component stimulates renin substrate (angiotensinogen) and thus angiotensin production, resulting in greater negative feedback on renin (and hence DRC), thereby opposing the stimulatory effects of the progestogenic component
Premenopausal women in luteal menstrual phase*	↑	↔	Progesterone induces natriuresis which stimulates renin and aldosterone Estrogen stimulates renin substrate and thus angiotensin production, resulting in greater negative feedback on renin
Advancing age	↔	↓↓	Reducing nephron (and JG cell) number and tendency to Na^+ retention leads to renin suppression Tendency to K^+ retention prevents fall in aldosterone
Chronic kidney disease	↔	↓↓	Reducing nephron number and tendency to Na^+ retention leads to renin suppression Tendency to K^+ retention prevents fall in aldosterone
High dietary Na^+	↓	↓↓	Suppression of renin (to a greater degree than aldosterone)
Familial hyperkalemic hypertension	↔	↓↓	Mutations in *WNK1*, *WNK4*, *CUL3* or *KLH3L* cause upregulation of NCC. Resulting Na^+ retention leads to renin suppression and K^+ retention prevents fall in aldosterone

*False positives only occur if renin is measured as direct renin concentration (and not plasma renin activity).

JG, Juxtaglomerular; *CNS*, central nervous system; *NSAIDs*, nonsteroidal antiinflammatory drugs; *NCC*, sodium chloride cotransporter; *DRC,* direct renin concentration.

TABLE 96.3 Factors Causing False-Negative Aldosterone-To-Renin Ratios

	Effect on Aldosterone	Effect on Renin	Mechanism
Diuretics (K^+ wasting or sparing)	↑↔	↑↑	Volume contraction and sympathetic nervous system stimulation stimulate renin Fall in plasma K^+ with K^+ wasting diuretics lowers aldosterone
ACE inhibitors	↓	↑↑	Reduced production of angiotensin II relieves negative feedback on JG cells, leading to marked increase in renin while also reducing angiotensin II stimulation of aldosterone production by ZG cells
ARBs	↓	↑↑	Blockade of angiotensin II receptors relieves negative feedback on JG cells, leading to marked increase in renin while also reducing angiotensin II stimulation of aldosterone production by ZG cells
Dihydropyridine CCBs	↔↑	↑	Reflex sympathetic stimulation as BP falls, natriuretic effects, and direct stimulation of calcium-dependent renin regulatory pathways Interference with aldosterone biosynthesis by blocking intracellular calcium-dependent steps
SSRIs	↑	↑↑	Uncertain—serotonin may stimulate renin via the 5-HT_2 subtype of 5-HT receptors
Pregnancy	↑	↑↑	Progesterone induces natriuresis, which stimulates renin and aldosterone
Low dietary Na^+ Vomiting and diarrhea	↑	↑↑	Volume contraction and sympathetic nervous system stimulation stimulate renin
Malignant hypertension Concomitant renovascular hypertension	↑	↑↑	Intrarenal ischemia leads to renin production
Hypokalemia	↓	↔↑	Reduced stimulation of aldosterone production

ACE, Angiotensin-converting enzyme; *ARB*, angiotensin II receptor blocker; *ZG*, zona glomerulosa; *JG*, juxtaglomerular; *CCB*, calcium channel blocker; *SSRI*, selective serotonin receptor inhibitor; *BP,* blood pressure.

suppression of plasma aldosterone were reported to be less than 162 pmol/L if measured by mass spectroscopy[128] and less than 171 pmol/L (but with a grey zone up to 217 pmol/L) if measured by chemiluminescence immunoassay.[109]

Unlike the above approaches, the captopril challenge test does not involve salt loading, but instead measures aldosterone and renin responses to 25 to 50 mg oral captopril, an ACE inhibitor, administered after sitting or standing for at least 1 h.[129] Blood samples are drawn basally and at 1 or 2 h after challenge, with the patient remaining seated throughout. In subjects without PA, plasma aldosterone is suppressed (>30%) and PRA stimulated by captopril, and, as a result, the ARR falls. In patients with PA aldosterone fails to suppress, PRA remains suppressed, and the ARR remains elevated. In one commonly used protocol, the cutoff for normal suppression of plasma aldosterone is less than 8.5 ng/dL (240 pmol/L) and for the ARR is less than 30 (with aldosterone expressed as ng/dL and PRA as ng/mL/h) 2 h after receiving 50 mg of captopril. Because it does not involve salt loading, the captopril challenge testing avoids the potential for inducing fluid overload in patients at risk because of compromised cardiac or renal function and hypokalemia. However, it has been reported to have poor discriminatory power[130,131] and may be associated with symptomatic hypotension due to the blood pressure-lowering effect of captopril in upright patients.

Because of the enhanced risk of potentially dangerous fluid overload, it may be necessary to avoid confirmatory testing involving salt loading in patients with severe hypertension or compromised cardiac status (e.g., reduced left ventricular systolic function or marked left ventricular hypertrophy) or renal dysfunction. Alternatively, consideration may be given to either a period of medical treatment with an agent that blocks aldosterone action as an attempt to establish better hypertension control and improve cardiac function in order to reduce the risk of confirmatory testing to a more acceptable level, or to proceeding directly to subtype differentiation (including AVS).

Confirmatory testing (followed, if positive by AVS) could also be avoided in favor of an empiric trial of medical treatment for PA in patients for whom unilateral adrenalectomy is not likely to be a viable treatment option, either because they are poor candidates for surgery or do not desire surgical intervention.

Subtype Differentiation. As described earlier in this chapter, the availability of immunohistochemical staining using antibodies directed against CYP11B2 has enabled investigators to identify the true source of aldosterone excess within adrenals removed from patients with PA, and this in turn has led to the realization that the morphology associated with PA is diverse and far more complex than previously suspected. From a clinical standpoint, however, rather than attempting to distinguish the many pathological variants from each other, subtype differentiation aims to assist selection of the optimal management approach. This is achieved by:

- genetic testing when a familial form of PA is suspected;
- imaging to enable detection of large adrenal lesions that may warrant consideration of removal because of their malignant potential; and
- determining whether autonomous aldosterone production is unilateral (confined to one adrenal), and therefore potentially curable by unilateral adrenalectomy, or bilateral, in which case management usually involves treatment with medications that antagonize aldosterone action.

Genetic testing for FH-I involves the use of either Southern blotting or a faster and less expensive PCR-based method to detect the causative hybrid *CYP11B1*/*CYP11B2* mutation within peripheral blood DNA.[132,133] The principal reasons for identifying individuals with this subtype of PA are that (1) patients with FH-I demonstrate excellent clinical responses to glucocorticoid medications given in small doses that do not cause Cushingoid side effects, making them the treatment of choice for this (but no other) subtype of PA; and (2) it heralds the need for genetic testing of other family members to detect additional individuals with FH-I similarly at risk of hypertension and early death from hemorrhagic stroke, but who will also benefit greatly from specific glucocorticoid treatment. The recommendation by the Endocrine Society is to test patients with onset of confirmed PA earlier than at 20 years of age and those who have a family history of PA or of strokes at young age (<40 years).[22] The case for screening for germline mutations in *CLCN2* (causing FH-II), *KCNJ5* (FH-III), and *CACNA1H* (FH-IV) is not as strong as that for FH-I, being even rarer than hybrid gene mutations and (unlike FH-I) with no specific form of treatment yet available. At present, given that early age of onset of hypertension is almost universal among these individuals, it would seem prudent to restrict genetic testing for these mutations to very young patients with PA. As with FH-I, genetic testing for these mutations identifies individuals whose family members should be genetically screened, thereby facilitating timely institution of specific medical or surgical treatment in those who test positive.

Adrenal imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is essential for the detection of larger (e.g., >2.5 cm) mass lesions that have greater potential for malignant behavior and may warrant consideration for removal on that basis alone, or at the very least deserve careful follow-up.

While the above facets of subtype differentiation are relatively straightforward, distinguishing unilateral from bilateral PA is much more complex. It is nevertheless an important aspect of the diagnostic workup of PA, as the clinical responses to unilateral adrenalectomy in patients with unilateral forms of PA are generally quite marked, with cure of hypertension achieved in 40% to 70%, and greater reductions in pill burden, better cardiovascular outcomes, and greater improvements in quality of life compared with patients with PA treated medically with drugs that antagonize aldosterone action.[99,101,134,135] For several decades, AVS has served as the mainstay of differentiating these two PA subtypes, as it is the most reliable. It is also the most invasive and expensive and requires considerable expertise. As a result, many alternative approaches have been explored, but with inefficient reliability to replace AVS. Although unilateral PA (including APA) is generally associated with more florid features of PA than bilateral, there is considerable overlap, and clinical prediction algorithms that have incorporated such parameters as plasma sodium and potassium, total carbon dioxide, aldosterone, renin, and the ARR have shown variable (and not very reproducible) performance characteristics.[136] Adrenal imaging by CT or MRI frequently misses APAs and can be frankly misleading by detecting nonfunctioning adrenal lesions.[97,137] Scintigraphy lacks sufficient resolution to detect smaller APAs. Posture stimulation studies (comparing recumbent with upright plasma aldosterone concentrations) more often showed posture unresponsiveness and plasma or urinary levels of "hybrid steroids" (18-hydroxy- and 18-oxocortisol) are more often elevated in patients with APA compared with bilateral PA,[138] but again, considerable overlap exists. The diversity in phenotype among patients with APA regarding these two parameters may in part be genetically determined: patients with APAs associated with somatic *KCNJ5* mutations usually show posture-unresponsiveness and elevated hybrid steroid levels, whereas the reverse holds true for those with somatic *CACNA1D* (and other) mutations.[139] Despite the shortcomings of these alternatives to AVS, they can nevertheless provide useful ancillary information that may help in the decision-making process, especially when AVS has been declined for some reason or results have been inconclusive (e.g., because of failure to cannulate one or both adrenal veins). For example, the presence of an obvious unilateral mass lesion on CT in a young patients with florid, hypokalemic, posture-unresponsive PA and elevated hybrid steroid levels (if available) would make APA highly likely. Nuclear imaging using positron emission tomography–CT with (11)C-metomidate has been reported to be a viable alternative or adjunct to AVS in difficult cases.[140,141] However, the short half-life and lack of specificity of the isotope for CYP11B2 (being a ligand also for CYP11B1) limits the current widespread application and reliability of this technique, though alternatives are being investigated.

The Endocrine Society guideline[22] recommends that AVS be offered to all patients with PA who desire adrenalectomy and for whom surgery is not contraindicated, regardless of biochemical or CT findings. Some investigators advocate performing AVS in the morning following overnight recumbency in order to avoid the confounding effects of changes in posture on aldosterone concentrations and to capitalize on the higher endogenous ACTH levels (compared with later in the day), which ensure that samples are collected during maximal ACTH-induced stimulation of aldosterone production (unless stimulation with exogenous ACTH is employed).[124] The likelihood of achieving successful cannulation of the adrenal veins is maximized when CT scanning is performed prior to AVS to assist in localizing the adrenal veins,[142] a radiologist highly skilled in AVS performs the procedure, at least two samples are collected from each side, and rapid point-of care cortisol testing is used to confirm correct catheter placement. Examination of the ratio or "gradient" between adrenal and peripheral venous cortisol concentrations permits confirmation (or otherwise) of cannulation success, with gradients of at least two or three (or at least five if exogenous ACTH stimulation is used) most commonly used to indicate adequacy of AVS. The right adrenal vein tends to be harder to cannulate than the left because it is usually smaller and shorter and, unlike the left adrenal vein, usually empties into the inferior vena cava rather than the renal vein.[143] Calculating the aldosterone/cortisol ratio for each adrenal and peripheral venous sample corrects for differences in dilution with nonadrenal venous blood. Criteria for lateralization vary widely from one institution to another,[144] with some comparing adrenal venous aldosterone/cortisol ratios on each side with the simultaneous peripheral venous ratio (with lateralization defined as a the ratio being at least 2.5 times peripheral on the affected side and no higher than peripheral on the other side), while others rely on the comparison of aldosterone/cortisol ratios on one side versus the other (the so-called "lateralization index"), with lateralization defined as the ratio on the higher side being at least two to four times, depending on the center, that on the lower side. Several groups employ exogenous ACTH stimulation during AVS to (1) maximize adrenal/peripheral venous cortisol gradients, (2) reduce fluctuations in steroid secretion during AVS and (3) stimulate aldosterone production by APAs, and thus avoid sampling during a period of secretion "quiescence."[144,145] The risk of adrenal hemorrhage associated with AVS is low (less than 2% of all procedures) in experienced hands.[142]

Recent advances and changes in thinking regarding AVS have included: (1) "superselective" sampling of branch adrenal veins, enabling more precise localization of the source (or sources) of aldosterone excess within a gland, which is particularly relevant in situations where subtotal adrenalectomy is considered desirable (for example, if the patient has only one functioning adrenal or is thought to have bilateral APAs)[146]; (2) the development of point-of-care plasma cortisol measurement, which permits semiquantitative determination of adrenal venous cortisol levels, and thus an assessment of the success or otherwise of cannulation within minutes of collection[147]; and (3) normalization of aldosterone levels with hormones other than cortisol, either to improve apparent rates of successful cannulation or where the patient is suspected of having an adrenal tumor that is concomitantly autonomously secreting both cortisol and aldosterone.[107,148]

KEY POINTS

- Detection and diagnostic workup for primary aldosteronism involves screening by measurement of the plasma aldosterone/renin ratio, confirmatory suppression testing, and determining the subtype by genetic testing (when familial forms are suspected), adrenal imaging, and adrenal venous sampling to differentiate unilateral from bilateral forms.

TREATMENT

Early recognition of PA is paramount to facilitate the opportunity to initiate targeted treatments. The objective of targeted treatments is to reduce the risk for incident cardiovascular and kidney disease. The traditional dogma for decades has been to prescribe surgical adrenalectomy or lifelong MR antagonist therapy, depending on the subtype of PA. It is important to note that there have been no robust, randomized, and controlled trials to evaluate the validity of this dogma or the comparative efficacy of these options. For example, whether MR antagonists are superior to conventional (nontargeted) antihypertensive medications, or whether surgical adrenalectomy is superior to MR antagonist therapy, has not been validated in prospective and controlled studies. Rather, these approaches have been recommended based on their pathophysiologic mechanisms of action, their provenance, and supportive observational human studies.

Herein, treatments for PA are discussed in two groupings: surgical adrenalectomy and medical therapy. Surgical adrenalectomy may involve curative unilateral adrenalectomy, noncurative unilateral adrenalectomy, and bilateral adrenalectomy. Medical therapy may involve MR antagonists, dietary sodium restriction, and epithelial sodium channel inhibitors.

Surgical Adrenalectomy. The treatment of choice for unilateral PA is curative unilateral laparoscopic adrenalectomy when patients are willing and able to undergo this operation. Laparoscopic and retroperitoneoscopic approaches are the gold standard and allow surgical resection with a minimally invasive approach. The unique attribute of adrenalectomy for unilateral PA is the opportunity to completely cure the disease by eliminating the source of inappropriate aldosterone production, curing hypokalemia, and either curing hypertension or dramatically improving blood pressure control.

The definition of cure following unilateral adrenalectomy has been defined by the Primary Aldosteronism Surgery Outcomes (PASO) study,[149] where a "complete clinical cure" was defined as postoperative normalization of blood pressure without the use of any antihypertensive medication use. A "complete biochemical cure" was defined by normalization of serum potassium without the need for supplements, along with a decrease in circulating aldosterone and/or an increase in renin such that there was normalization of the ARR. Notably, the PASO study highlighted that, whereas the majority of patients achieve a biochemical cure following adrenalectomy for unilateral PA (94%), only a minority (37%) of patients achieve a complete clinical cure, in part owing to the vasculopathy induced by longstanding hypertension and aldosteronism.

Biochemical changes induced by curative unilateral adrenalectomy include normalization of serum potassium, an increase in serum creatinine and decrease in estimated glomerular filtration rate, a dramatic decline in plasma aldosterone, an increase in or normalization of renin (from suppressed to unsuppressed/detectable), and consequently a decline in the ARR.[150] These biochemical changes all reflect the expected physiology induced by removing inappropriate and excessive aldosterone production: as intravascular volume contracts, renin and angiotensin II increase, glomerular hyperfiltration decreases, delivery of sodium ion to the distal nephron decreases, and excretion of urinary potassium and hydrogen ion decrease. The increase in serum creatinine following curative adrenalectomy is usually not a marker of kidney injury, but rather represents a reflection of the true underlying kidney function once hyperfiltration is corrected. In this regard, many patients with PA may be found to have some degree of chronic kidney disease postadrenalectomy that was previously not recognized.

The clinical changes induced by curative adrenalectomy may include a cure of hypertension, reductions in cardiometabolic risk, and reduction in the risk of death. Cure of hypertension is observed in the minority of cases and more commonly among younger patients with a shorter duration of hypertension. More commonly, hypertension is not cured, but blood pressure control is improved and/or the number of antihypertensive medications required is decreased.

Early observational data from a small sample size of 54 patients comparing long-term cardiovascular outcomes between PA patients treated with adrenalectomy, PA patients treated with MR antagonist therapy, and a comparator group of patients with essential hypertension suggested no difference in long-term cardiovascular risk among these groups.[151] However, more recent observational data from large studies[152] have shown that, despite similar blood pressure control between the groups, patients with PA treated with MR antagonists had a 3-fold higher risk for incident cardiovascular events compared with patients with PA treated with curative adrenalectomy and a 2-fold higher risk compared with patients with essential hypertension. In contrast, patients with PA treated with curative adrenalectomy had a 40% lower risk for incident cardiovascular events compared with patients with essential hypertension.[152-154] However, when MR antagonist therapy resulted in a substantial increase in renin activity, the risk for incident cardiovascular events was observed to be no different than comparable patients with essential hypertension and similar to those who underwent curative adrenalectomy. These observational findings underscore the efficacy of surgical adrenalectomy, but also suggest that, when MR antagonist therapy can be titrated to achieve a similar physiology to adrenalectomy, adverse outcome rates may approximate, but not quite capitulate, the effects of adrenalectomy.[150]

Curative adrenalectomy in PA has been shown to have a beneficial effect in terms of preventing incident atrial fibrillation compared with MR antagonist therapy and in preventing incident diabetes when compared to essential hypertension. Several studies have shown that curative adrenalectomy is associated with a preservation of kidney function when compared to MR antagonist therapy, where declines in glomerular filtration rate and the incidence of chronic kidney disease are greater. Both curative adrenalectomy and medical therapy have been shown to improve quality of life; however, the improvements in quality-of-life metrics following curative adrenalectomy have been more substantial.[150]

Although surgical adrenalectomy is most commonly recommended as a curative procedure for unilateral PA, there is also evidence to support the efficacy of noncurative unilateral adrenalectomy in bilateral PA. Importantly, there is no high-grade clinical trial evidence comparing noncurative unilateral adrenalectomy with lifelong MR antagonist therapy in bilateral PA; rather, supportive evidence stems from observational studies and anecdotal patient series. In a case series of 12 patients who underwent noncurative unilateral adrenalectomy for bilateral PA, patients who had noncurative adrenalectomy had a substantial and durable reduction in blood pressure, a modest decrease in aldosterone levels, a substantial increase in renin that paralleled the rise in renin seen in those who underwent curative unilateral adrenalectomy, and near-normalization of serum potassium.[155] In a larger study of 51 patients who underwent noncurative adrenalectomy for bilateral PA, the operation cured hypertension in 15% of patients, improved blood pressure control in 40%, decreased aldosterone levels, substantially increased renin levels (and consequently decreased the ARR), normalized serum potassium, and increased serum creatinine.[156] Notably, the adrenalectomy in this series was performed to target the morphologically larger adrenal and/or where adrenal venous sampling suggested greater lateralization. These findings, in combination with anecdotal clinical experiences, suggest that unilateral noncurative adrenalectomy for bilateral PA could be considered as an adjunct method to induce improvements in pathophysiology and clinical parameters in selected patients, particularly when guided by some degree of lateralization and/or greater morphologic abnormalities. In some centers, the use of super-selective adrenal venous sampling to detect specific foci of aldosterone production, as well as thermal and arterial ablations, have also been shown to be effective targeted therapies.

Bilateral adrenalectomy for bilateral PA is rarely employed for sporadic cases of PA, and is not recommended or supported by any reliable studies. Because bilateral adrenalectomy can cure the aldosteronism, it may be used as an option in extreme instances where unilateral adrenalectomy and/or medical therapy are unable to improve clinical and biochemical parameters, and when the risk for incident cardiometabolic disease is considered to be more severe than the risk for lifelong primary adrenal insufficiency. Detailed discussions regarding the risks, benefits, and ethics of such a decision should occur in conjunction with an experienced endocrinologist with knowledge in managing adrenal insufficiency and PA.

In summary, surgical adrenalectomy is effective when utilized to unilateral PA, but sometimes also when used to attenuate the severity of the excess aldosterone production in bilateral PA. The hallmark characteristics suggestive of efficacy include improvements (or cure) of hypertension, reductions in aldosterone and increases in renin (such that the ARR is markedly reduced), normalization of serum potassium, and increases in serum creatinine indicative of a resolution in glomerular hyperfiltration. These features are associated with a reduction in the incident risk for cardiometabolic disease and death.

Medical Therapy. Medical therapy, here defined as any nonsurgical therapy, is the approach of choice when PA is deemed to be bilateral in origin by adrenal venous sampling, or when patients are unable or unwilling to undergo surgical adrenalectomy. Prior to initiating targeted MR antagonists, the critical importance of dietary sodium restriction should be underscored for patients, as it is a foundational intervention for all patients with hypertension. The efficacy of dietary sodium restriction for PA is high, and this is particularly true when combined with MR antagonists. Sodium is the substrate that fuels the pathophysiology of PA. Dietary sodium restriction results in reductions in effective arterial circulating volume, decreases in glomerular filtration, reductions in distal nephron sodium delivery, and consequently reductions in MR-mediated sodium reabsorption and potassium and hydrogen ion excretion. Most patients in industrialized societies consume more than 150 to 200 mmol of sodium per day. In contrast, the American Heart Association recommends consuming less than 65 mmol of sodium per day for patients with hypertension. In one study, restricting dietary sodium intake to less than 50 mmol per day in patients with PA dramatically lowered blood pressure, substantially increased renin, and normalized the ARR in over 50% of participants.[157] While the efficacy of dietary sodium restriction is established, the ability for most patients to adhere to these degrees of sodium restriction is a major challenge. Therefore, involving professional dieticians and nutrition support can be a valuable adjunct to pharmacologic therapy.

In addition to dietary sodium restriction, MR antagonists are the mainstay of pharmacologic medical therapy. By blocking the MR, MR antagonists reduce sodium reabsorption and volume expansion, and consequently potassium and hydrogen ion excretion decline. In this manner, MR antagonists are capable of improving blood pressure, hypokalemia, and alkalosis. The main risk of MR antagonist therapy is hyperkalemia, which is more likely to occur when patients have concomitant reductions in glomerular filtration rate or use renin-angiotensin-aldosterone system inhibitors. Spironolactone and eplerenone are the most well-known MR antagonists; both have a steroidal structure that permits competitive inhibition of aldosterone. Spironolactone is generally regarded as the first choice, whereas eplerenone should be considered for patients who cannot tolerate spironolactone. Spironolactone is approximately twice as potent as eplerenone and also inhibits the androgen receptor and activates the progesterone receptor; therefore, the most common adverse reaction leading to intolerance is gynecomastia in adult men, with menstrual irregularities in women being less common. The starting dose of spironolactone is usually 12.5 to 25 mg daily, but it can be titrated up to 200 mg daily. Eplerenone is usually initiated at 25 to 50 mg daily and can be titrated up to 200 to 400 mg daily. Beyond these conventional MR antagonists, new agents in development include finerenone and esaxerenone. Both finerenone and esaxerenone are nonsteroidal MR antagonists that been shown to reduce the progression of kidney disease. Esaxerenone has shown efficacy in lowering blood pressure, and finerenone is undergoing further study in heart failure.

It should be noted that epithelial sodium channel antagonists, such as amiloride and triamterene, are also effective for blood pressure and potassium control in PA; however, these agents also carry the risk for hyperkalemia when kidney function is impaired, and further, they do not block the MR. Therefore, whether long-term use of amiloride can mitigate MR-mediated cardiovascular disease beyond blood pressure control remains unknown.

The primary goals of medical therapy (dietary sodium restriction in combination with MR antagonists and/or epithelial sodium channel inhibitors) are normalization of blood pressure and serum potassium. The secondary goals are to optimize biomarkers associated with a decreased risk for incident cardiometabolic adverse outcomes. MR antagonists should be titrated to adequately block the renal and extrarenal MR. Optimized medical therapy is ideally defined as normal blood pressure, normal or high-normal serum potassium, and an increase in renin activity or concentration.[150] The achievement of these clinical and biochemical milestones is associated with similar cardiovascular risk reduction achieved by surgical adrenalectomy. When MR antagonist therapy is unable to optimize these clinical and biochemical parameters, amiloride can be added, or noncurative unilateral adrenalectomy considered.

KEY POINTS

- There are multiple modalities to effectively treat primary aldosteronism. When possible, curative surgical adrenalectomy is the preferred method to improve blood pressure and risk for future cardiovascular events. Alternatively, medical therapy that combines dietary sodium restriction and mineralocorticoid receptor antagonists can also impart similar reductions in blood pressure and cardiovascular risk. In some select cases, the use of epithelial sodium channel inhibitors or noncurative unilateral adrenalectomy may be beneficial.

REFERENCES

1. Stowasser M, Gordon RD. Primary aldosteronism: changing definitions and new concepts of physiology and pathophysiology both inside and outside the kidney. *Physiol Rev*. 2016;96:1327–1384.
2. Vaidya A, Mulatero P, Baudrand R, et al. The expanding spectrum of primary aldosteronism: implications for diagnosis, pathogenesis, and treatment. *Endocr Rev*. 2018;39:1057–1088.
3. Conn JW, Cohen EL, Rovner DR. Landmark article Oct 19, 1964: suppression of plasma renin activity in primary aldosteronism. Distinguishing primary from secondary aldosteronism in hypertensive disease. By J.W. Conn, E.L. Cohen and D.R. Rovner *J Am Med Assoc*. 1985;253:558–566.
4. Vaidya A, Carey RM. The evolution of the primary aldosteronism syndrome: updating the approach. *J Clin Endocrinol Metab*. 2020;105(12):3771-3783.
5. Martinez DV, Rocha R, Matsumura M, et al. Cardiac damage prevention by eplerenone: comparison with low sodium diet or potassium loading. *Hypertension*. 2002;39:614–618.
6. Rocha R, Stier Jr CT, Kifor I, et al. Aldosterone: a mediator of myocardial necrosis and renal arteriopathy. *Endocrinology*. 2000;141:3871–3878.
7. Jia G, Habibi J, Aroor AR, et al. Endothelial mineralocorticoid receptor mediates diet-induced aortic stiffness in females. *Circ Res*. 2016;118:935–943.
8. Kim SK, McCurley AT, DuPont JJ, et al. Smooth muscle cell-mineralocorticoid receptor as a mediator of cardiovascular stiffness with aging. *Hypertension*. 2018;71(4):609-621.
9. McCurley A, Pires PW, Bender SB, et al. Direct regulation of blood pressure by smooth muscle cell mineralocorticoid receptors. *Nat Med*. 2012;18:1429–1433.
10. Mueller KB, Bender SB, Hong K, et al. Endothelial mineralocorticoid receptors differentially contribute to coronary and mesenteric vascular function without modulating blood pressure. *Hypertension*. 2015;66:988–997.
11. Pitt B, Zannad F, Remme WJ, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999;341:709–717.
12. Pfeffer MA, Claggett B, Assmann SF, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation*. 2015;131:34–42.
13. Connell JM, Davies E. The new biology of aldosterone. *J Endocrinol*. 2005;186:1–20.
14. Bandulik S, Penton D, Barhanin J, et al. TASK1 and TASK3 potassium channels: determinants of aldosterone secretion and adrenocortical zonation. *Horm Metab Res*. 2010;42:450–457.
15. Spat A, Hunyady L. Control of aldosterone secretion: a model for convergence in cellular signaling pathways. *Physiol Rev*. 2004;84:489–539.
16. Mulatero P, Tizzani D, Viola A, et al. Prevalence and characteristics of familial hyperaldosteronism: the PATOGEN study (Primary Aldosteronism in TOrino-GENetic forms). *Hypertension*. 2011;58:797–803.
17. Mulatero P, Monticone S, Deinum J, et al. Genetics, prevalence, screening and confirmation of primary aldosteronism: a position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J Hypertens*. 2020;38:1919–1928.
18. Stowasser M, Sharman J, Leano R, et al. Evidence for abnormal left ventricular structure and function in normotensive individuals with familial hyperaldosteronism type I. *J Clin Endocrinol Metab*. 2005;90:5070–5076.
19. Sutherland DJ, Ruse JL, Laidlaw JC. Hypertension, increased aldosterone secretion and low plasma renin activity relieved by dexamethasone. *Can Med Assoc J*. 1966;95:1109–1119.
20. Lifton RP, Dluhy RG, Powers M, et al. A chimaeric 11beta-hydroxylase aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*. 1992;355:262–265.
21. Stowasser M, Bachmann AW, Huggard PR, et al. Treatment of familial hyperaldosteronism type I: only partial suppression of adrenocorticotropin required to correct hypertension. *J Clin Endocrinol Metab*. 2000;85:3313–3318.

22. Funder JW, Carey RM, Mantero F, et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:1889–1916.
23. Pallauf A, Schirpenbach C, Zwermann O, et al. The prevalence of familial hyperaldosteronism in apparently sporadic primary aldosteronism in Germany: a single center experience. *Horm Metab Res*. 2012;44:215–220.
24. Aglony M, Martinez-Aguayo A, Carvajal CA, et al. Frequency of familial hyperaldosteronism type 1 in a hypertensive pediatric population: clinical and biochemical presentation. *Hypertension*. 2011;57:1117–1121.
25. Litchfield WR, Anderson BF, Weiss RJ, et al. Intracranial aneurysm and hemorrhagic stroke in glucocorticoid-remediable aldosteronism. *Hypertension*. 1998;31:445–450.
26. Gordon RD, Stowasser M, Tunny TJ, et al. Clinical and pathological diversity of primary aldosteronism, including a new familial variety. *Clin Exp Pharmacol Physiol*. 1991;18:283–286.
27. Stowasser M, Gordon RD. Familial hyperaldosteronism. *J Steroid Biochem Mol Biol*. 2001;78:215–229.
28. Stowasser M, Gordon RD, Tunny TJ, et al. Familial hyperaldosteronism type II: five families with a new variety of primary aldosteronism. *Clin Exp Pharmacol Physiol*. 1992;19:319–322.
29. Scholl UI, Stolting G, Schewe J, et al. CLCN2 chloride channel mutations in familial hyperaldosteronism type II. *Nat Genet*. 2018;50:349–354.
30. Fernandes-Rosa FL, Daniil G, Orozco IJ, et al. A gain-of-function mutation in the CLCN2 chloride channel gene causes primary aldosteronism. *Nat Genet*. 2018;50:355–361.
31. Scholl UI, Nelson-Williams C, Yue P, et al. Hypertension with or without adrenal hyperplasia due to different inherited mutations in the potassium channel KCNJ5. *Proc Natl Acad Sci USA*. 2012;109:2533–2538.
32. Mulatero P, Tauber P, Zennaro MC, et al. KCNJ5 mutations in European families with nonglucocorticoid remediable familial hyperaldosteronism. *Hypertension*. 2012;59:235–240.
33. Daniil G, Fernandes-Rosa FL, Chemin J, et al. CACNA1H mutations are associated with different forms of primary aldosteronism. *EBioMedicine*. 2016;13:225–236.
34. Geller DS, Zhang J, Wisgerhof MV, et al. A novel form of human mendelian hypertension featuring nonglucocorticoid-remediable aldosteronism. *J Clin Endocrinol Metab*. 2008;93:3117–3123.
35. Choi M, Scholl UI, Yue P, et al. K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science*. 2011;331:768–772.
36. Oki K, Plonczynski MW, Luis Lam M, et al. Potassium channel mutant KCNJ5 T158A expression in HAC-15 cells increases aldosterone synthesis. *Endocrinology*. 2012;153:1774–1782.
37. Gomez-Sanchez CE, Qi X, Gomez-Sanchez EP, et al. Disordered zonal and cellular CYP11B2 enzyme expression in familial hyperaldosteronism type 3. *Mol Cell Endocrinol*. 2017;439:74–80.
38. Tong A, Liu G, Wang F, et al. A novel phenotype of familial hyperaldosteronism type III: concurrence of aldosteronism and Cushing's syndrome. *J Clin Endocrinol Metab*. 2016;101:4290–4297.
39. Scholl UI, Stolting G, Nelson-Williams C, et al. Recurrent gain of function mutation in calcium channel CACNA1H causes early-onset hypertension with primary aldosteronism. *Elife*. 2015;4:e06315.
40. Scholl UI, Goh G, Stolting G, et al. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nat Genet*. 2013;45:1050–1054.
41. Semenova NA, Ryzhkova OR, Strokova TV, et al. [The third case report a patient with primary aldosteronism, seizures, and neurologic abnormalities (PASNA) syndrome de novo variant mutations in the CACNA1D gene]. *Zh Nevrol Psikhiatr Im S S Korsakova*. 2018;118:49–52.
42. De Mingo Alemany MC, Mifsud Grau L, Moreno Macian F, et al. A de novo CACNA1D missense mutation in a patient with congenital hyperinsulinism, primary hyperaldosteronism and hypotonia. *Channels*. 2020;14:175–180.
43. Nanba K, Omata K, Else T, et al. Targeted molecular characterization of aldosterone-producing adenomas in White Americans. *J Clin Endocrinol Metab*. 2018;103:3869–3876.
44. Nanba K, Omata K, Gomez-Sanchez CE, et al. Genetic characteristics of aldosterone-producing adenomas in Blacks. *Hypertension*. 2019;73:885–892.
45. De Sousa K, Boulkroun S, Baron S, et al. Genetic, cellular, and molecular heterogeneity in adrenals with aldosterone-producing adenoma. *Hypertension*. 2020;75:1034–1044.
46. Nanba K, Yamazaki Y, Bick N, et al. Prevalence of somatic mutations in aldosterone-producing adenomas in Japanese patients. *J Clin Endocrinol Metab*. 2020;105:e4066–e4073.
47. Teo AE, Garg S, Shaikh LH, et al. Pregnancy, primary aldosteronism, and adrenal CTNNB1 mutations. *N Engl J Med*. 2015;373:1429–1436.
48. Scholl UI, Healy JM, Thiel A, et al. Novel somatic mutations in primary hyperaldosteronism are related to the clinical, radiological and pathological phenotype. *Clin Endocrinol*. 2015;83:779–789.
49. Akerstrom T, Maharjan R, Sven Willenberg H, et al. Activating mutations in CTNNB1 in aldosterone producing adenomas. *Sci Rep*. 2016;6:19546.
50. Rhayem Y, Perez-Rivas LG, Dietz A, et al. PRKACA somatic mutations are rare findings in aldosterone-producing adenomas. *J Clin Endocrinol Metab*. 2016;101:3010–3017.
51. Stindl J, Tauber P, Sterner C, et al. Pathogenesis of adrenal aldosterone-producing adenomas carrying mutations of the Na(+)/K(+)-ATPase. *Endocrinology*. 2015;156:4582–4591.
52. Rege J, Nanba K, Blinder AR, et al. Identification of somatic mutations in CLCN2 in aldosterone-producing adenomas. *J Endocr Soc*. 2020;4:bvaa123.
53. Tauber P, Aichinger B, Christ C, et al. Cellular pathophysiology of an adrenal adenoma-associated mutant of the plasma membrane Ca(2+)-ATPase ATP2B3. *Endocrinology*. 2016;157:2489–2499.
54. Azizan EA, Poulsen H, Tuluc P, et al. Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. *Nat Genet*. 2013;45:1055–1060.
55. Fernandes-Rosa FL, Williams TA, Riester A, et al. Genetic spectrum and clinical correlates of somatic mutations in aldosterone-producing adenoma. *Hypertension*. 2014;64:354–361.
56. Lenzini L, Rossitto G, Maiolino G, et al. A meta-analysis of somatic KCNJ5 K(+) channel mutations in 1636 patients with an aldosterone-producing adenoma. *J Clin Endocrinol Metab*. 2015;100:E1089–E1095.
57. Dutta RK, Arnesen T, Heie A, et al. Case report: a somatic mutation in CLCN2 identified in a sporadic aldosterone producing adenoma. *Eur J Endocrinol*. 2019;181:K37–K41.
58. Omata K, Satoh F, Morimoto R, et al. Cellular and genetic causes of idiopathic hyperaldosteronism. *Hypertension*. 2018;72:874–880.
59. Boulkroun S, Samson-Couterie B, Dzib JF, et al. Adrenal cortex remodeling and functional zona glomerulosa hyperplasia in primary aldosteronism. *Hypertension*. 2010;56:885–892.
60. Nishimoto K, Nakagawa K, Li D, et al. Adrenocortical zonation in humans under normal and pathological conditions. *J Clin Endocrinol Metab*. 2010;95:2296–2305.
61. Nanba K, Vaidya A, Williams GH, et al. Age-related autonomous aldosteronism. *Circulation*. 2017;136:347–355.
62. Williams TA, Peitzsch M, Dietz AS, et al. Genotype-specific steroid profiles associated with aldosterone-producing adenomas. *Hypertension*. 2016;67:139–145.
63. Guo Z, Nanba K, Udager A, et al. Biochemical, histopathological, and genetic characterization of posture-responsive and unresponsive APAs. *J Clin Endocrinol Metab*. 2020;105:e3224–e3235.
64. Maiolino G, Ceolotto G, Battistel M, et al. Macrolides for KCNJ5-Mutated Aldosterone-Producing Adenoma (MAPA): Design of a Study for Personalized Diagnosis of Primary Aldosteronism. Vol. 27. Blood Press; 2018:200–205.
65. Newton-Cheh C, Guo CY, Gona P, et al. Clinical and genetic correlates of aldosterone-to-renin ratio and relations to blood pressure in a community sample. *Hypertension*. 2007;49:846–856.
66. Hannemann A, Bidlingmaier M, Friedrich N, et al. Screening for primary aldosteronism in hypertensive subjects: results from two German epidemiological studies. *Eur J Endocrinol*. 2012;167:7–15.

67. Inoue K, Goldwater D, Allison M, et al. Serum aldosterone concentration, blood pressure, and coronary artery calcium: the multi-ethnic study of atherosclerosis. *Hypertension*. 2020;76:113–120.
68. Monticone S, Burrello J, Tizzani D, et al. Prevalence and clinical manifestations of primary aldosteronism encountered in primary care practice. *J Am Coll Cardiol*. 2017;69:1811–1820.
69. Mosso L, Carvajal C, Gonzalez A, et al. Primary aldosteronism and hypertensive disease. *Hypertension*. 2003;42:161–165.
70. Williams JS, Williams GH, Raji A, et al. Prevalence of primary hyperaldosteronism in mild to moderate hypertension without hypokalaemia. *J Hum Hypertens*. 2006;20:129–136.
71. Omura M, Saito J, Yamaguchi K, et al. Prospective study on the prevalence of secondary hypertension among hypertensive patients visiting a general outpatient clinic in Japan. *Hypertens Res*. 2004;27:193–202.
72. Rossi GP, Bernini G, Caliumi C, et al. A prospective study of the prevalence of primary aldosteronism in 1,125 hypertensive patients. *J Am Coll Cardiol*. 2006;48:2293–2300.
73. Rossi E, Regolisti G, Negro A, et al. High prevalence of primary aldosteronism using postcaptopril plasma aldosterone to renin ratio as a screening test among Italian hypertensives. *Am J Hypertens*. 2002;15:896–902.
74. Strauch B, Zelinka T, Hampf M, et al. Prevalence of primary hyperaldosteronism in moderate to severe hypertension in the Central Europe region. *J Hum Hypertens*. 2003;17:349–352.
75. Xu Z, Yang J, Hu J, et al. Primary aldosteronism in patients in China with recently detected hypertension. *J Am Coll Cardiol*. 2020;75:1913–1922.
76. Hu Y, Zhang J, Liu W, et al. Determining the prevalence of primary aldosteronism in patients with new-onset type 2 diabetes and hypertension. *J Clin Endocrinol Metab*. 2020:105;(4):1079-1085.
77. Yozamp N, Hundemer GL, Moussa M, et al. Intraindividual variability of aldosterone concentrations in primary aldosteronism: implications for case detection. *Hypertension*. 2020;77(3):891-899.
78. Tsiavos V, Markou A, Papanastasiou L, et al. A new highly sensitive and specific overnight combined screening and diagnostic test for primary aldosteronism. *Eur J Endocrinol*. 2016;175:21–28.
79. Parasiliti-Caprino M, Lopez C, Prencipe N, et al. Prevalence of primary aldosteronism and association with cardiovascular complications in patients with resistant and refractory hypertension. *J Hypertens*. 2020;38(9):1841–1848.
80. Brown JM, Siddiqui M, Carey R, et al. The unrecognized prevalence of primary aldosteronism. *Ann Int Med*. 2020;173(1):10–20.
81. Funder JW. Primary aldosteronism: at the tipping point. *Ann Intern Med*. 2020.
82. Calhoun DA, Nishizaka MK, Zaman MA, et al. Hyperaldosteronism among black and white subjects with resistant hypertension. *Hypertension*. 2002;40:892–896.
83. Williams B, Macdonald TM, Morant S, et al. Spironolactone versus placebo, bisoprolol, and doxazosin to determine the optimal treatment for drug-resistant hypertension (PATHWAY-2): a randomised, double-blind, crossover trial. *Lancet*. 2015;386:2059–2068.
84. Williams B, MacDonald TM, Morant SV, et al. Endocrine and haemodynamic changes in resistant hypertension, and blood pressure responses to spironolactone or amiloride: the PATHWAY-2 mechanisms substudies. *Lancet Diabet Endocrinol*. 2018;(6):464–475.
85. Adlin EV. Subclinical primary aldosteronism. *Ann Int Med*. 2017;167(9):673–674.
86. Markou A, Pappa T, Kaltsas G, et al. Evidence of primary aldosteronism in a predominantly female cohort of normotensive individuals: a very high odds ratio for progression into arterial hypertension. *J Clin Endocrinol Metab*. 2013;98:1409–1416.
87. Piaditis G, Markou A, Papanastasiou L, et al. Progress in aldosteronism: a review of the prevalence of primary aldosteronism in pre-hypertension and hypertension. *Eur J Endocrinol*. 2015;172:R191–R203.
88. Baudrand R, Guarda FJ, Fardella CE, et al. Continuum of Renin-independent Aldosteronism in Normotension. *Hypertension*. 2017;69(5):950–956.
89. Carey RM, Ayers CR, Vaughan Jr ED, et al. Activity of [des-aspartyl1]-angiotensin II in primary aldosteronism. *J Clin Invest*. 1979;63:718–726.
90. Wisgerhof M, Brown RD. Increased adrenal sensitivity to angiotensin II in low-renin essential hypertension. *J Clin Invest*. 1978;61:1456–1462.
91. Vasan RS, Evans JC, Larson MG, et al. Serum aldosterone and the incidence of hypertension in nonhypertensive persons. *N Engl J Med*. 2004;351:33–41.
92. Perschel FH, Schemer R, Seiler L, et al. Rapid screening test for primary hyperaldosteronism: ratio of plasma aldosterone to renin concentration determined by fully automated chemiluminescence immunoassays. *Clin Chem*. 2004;50:1650–1655.
93. Brown JM, Robinson-Cohen C, Luque-Fernandez MA, et al. The spectrum of subclinical primary aldosteronism and incident hypertension: a cohort study. *Ann Int Med*. 2017;167(9):630–641.
94. Rich GM, Ulick S, Cook S, et al. Glucocorticoid-remediable aldosteronism in a large kindred: clinical spectrum and diagnosis using a characteristic biochemical phenotype. *Ann Intern Med*. 1992;116:813–820.
95. Stowasser M, Huggard PR, Rossetti TR, et al. Biochemical evidence of aldosterone overproduction and abnormal regulation in normotensive individuals with familial hyperaldosteronism type I. *J Clin Endocrinol Metab*. 1999;84:4031–4036.
96. Mulatero P, Stowasser M, Loh KC, et al. Increased diagnosis of primary aldosteronism, including surgically correctable forms, in centers from five continents. *J Clin Endocrinol Metab*. 2004;89:1045–1050.
97. Stowasser M, Gordon RD, Gunasekera TG, et al. High rate of detection of primary aldosteronism, including surgically treatable forms, after 'non-selective' screening of hypertensive patients. *J Hypertens*. 2003;21:2149–2157.
98. Sonino N, Tomba E, Genesia ML, et al. Psychological assessment of primary aldosteronism: a controlled study. *J Clin Endocrinol Metab*. 2011;96:E878–E883.
99. Ahmed AH, Gordon RD, Sukor N, et al. Quality of life in patients with bilateral primary aldosteronism before and during treatment with spironolactone and/or amiloride, including a comparison with our previously published results in those with unilateral disease treated surgically. *J Clin Endocrinol Metab*. 2011;96:2904–2911.
100. Sonino N, Fallo F, Fava GA. Psychological aspects of primary aldosteronism. *Psychother Psychosom*. 2006;75:327–330.
101. Sukor N, Kogovsek C, Gordon RD, et al. Improved quality of life, blood pressure, and biochemical status following laparoscopic adrenalectomy for unilateral primary aldosteronism. *J Clin Endocrinol Metab*. 2010;95:1360–1364.
102. Gordon RD, Tunny TJ. Aldosterone-producing-adenoma (A-P-A): effect of pregnancy. *Clin Exp Hypertens*. 1982;4:1685–1693.
103. Ronconi V, Turchi F, Zennaro MC, et al. Progesterone increase counteracts aldosterone action in a pregnant woman with primary aldosteronism. *Clin Endocrinol*. 2011;74:278–279.
104. Murphy BF, Whitworth JA, Kincaid-Smith P. Malignant hypertension due to an aldosterone producing adrenal adenoma. *Clin Exp Hypertens*. 1985;7:939–950.
105. Zarifis J, Lip GY, Leatherdale B, et al. Malignant Hypertension in Association with Primary Aldosteronism. Vol. 5. Blood Press; 1996:250–254.
106. Taylor PJ, Cooper DP, Gordon RD, et al. Measurement of aldosterone in human plasma by semiautomated HPLC-tandem mass spectrometry. *Clin Chem*. 2009;55:1155–1162.
107. Peitzsch M, Dekkers T, Haase M, et al. An LC-MS/MS method for steroid profiling during adrenal venous sampling for investigation of primary aldosteronism. *J Steroid Biochem Mol Biol*. 2015;145:75–84.
108. Guo Z, Poglitsch M, McWhinney BC, et al. Aldosterone LC-MS/MS assay-specific threshold values in screening and confirmatory testing for primary aldosteronism. *J Clin Endocrinol Metab*. 2018;103:3965–3973.
109. Thuzar M, Young K, Ahmed AH, et al. Diagnosis of primary aldosteronism by seated saline suppression test-variability between immunoassay and HPLC-MS/MS. *J Clin Endocrinol Metab*. 2020:105:E447–E483.
110. Sealey JE, Laragh JH. Radioimmunoassay of plasma renin activity. *Semin Nucl Med*. 1975;5:189–202.
111. Ferrari P, Shaw SG, Nicod J, et al. Active renin versus plasma renin activity to define aldosterone-to-renin ratio for primary aldosteronism. *J Hypertens*. 2004;22:377–381.
112. Montori VM, Schwartz GL, Chapman AB, et al. Validity of the aldosterone-renin ratio used to screen for primary aldosteronism. *Mayo Clin Proc*. 2001;76:877–882.

113. Campbell DJ, Nussberger J, Stowasser M, et al. Activity assays and immunoassays for plasma renin and prorenin: information provided and precautions necessary for accurate measurement. *Clin Chem*. 2009;55:867–877.
114. Oelkers WK. Effects of estrogens and progestogens on the renin-aldosterone system and blood pressure. *Steroids*. 1996;61:166–171.
115. Morganti A. European study group for the validation of DiaSorin LDRA. A comparative study on inter and intralaboratory reproducibility of renin measurement with a conventional enzymatic method and a new chemiluminescent assay of immunoreactive renin. *J Hypertens*. 2010;28:1307–1312.
116. Guo Z, Poglitsch M, McWhinney BC, et al. Measurement of equilibrium angiotensin II in the diagnosis of primary aldosteronism. *Clin Chem*. 2020;66:483–492.
117. Gordon RD, Stowasser M. Primary aldosteronism: the case for screening. *Nat Clin Pract Nephrol*. 2007;3:582–583.
118. Zarnegar R, Young Jr WF, Lee J, et al. The aldosteronoma resolution score: predicting complete resolution of hypertension after adrenalectomy for aldosteronoma. *Ann Surg*. 2008;247:511–518.
119. Stowasser M, Gordon RD. Primary aldosteronism-careful investigation is essential and rewarding. *Mol Cell Endocrinol*. 2004;217:33–39.
120. Mulatero P, Rabbia F, Milan A, et al. Drug effects on aldosterone/plasma renin activity ratio in primary aldosteronism. *Hypertension*. 2002;40:897–902.
121. Stowasser M, Ahmed AH, Pimenta E, et al. Factors affecting the aldosterone/renin ratio. *Horm Metab Res*. 2012;44:170–176.
122. Rossi GP, Seccia TM, Palumbo G, et al. Within-patient reproducibility of the aldosterone: renin ratio in primary aldosteronism. *Hypertension*. 2010;55:83–89.
123. Young WFJ. Primary aldosteronism: update on diagnosis and treatment. *Endocrinololgist*. 1997;7:213–221.
124. Gordon RD. Primary aldosteronism. *J Endocrinol Invest*. 1995;18:495–511.
125. Kem DC, Weinberger MH, Mayes DC, et al. Saline suppression of plasma aldosterone in hypertension. *Arch Intern Med*. 1971;128:380–386.
126. Holland OB, Brown H, Kuhnert L, et al. Further evaluation of saline infusion for the diagnosis of primary aldosteronism. *Hypertension*. 1984;6:717–723.
127. Ahmed AH, Cowley D, Wolley M, et al. Seated saline suppression testing for the diagnosis of primary aldosteronism: a preliminary study. *J Clin Endocrinol Metab*. 2014;99:2745–2753.
128. Stowasser M, Ahmed AH, Cowley D, et al. Comparison of seated with recumbent saline suppression testing for the diagnosis of primary aldosteronism. *J Clin Endocrinol Metab*. 2018;103:4113–4124.
129. Iwaoka T, Umeda T, Naomi S, et al. The usefulness of the captopril test as a simultaneous screening for primary aldosteronism and renovascular hypertension. *Am J Hypertens*. 1993;6:899–906.
130. Westerdahl C, Bergenfelz A, Isaksson A, et al. Captopril suppression: limitations for confirmation of primary aldosteronism. *J Renin Angiotensin Aldosterone Syst*. 2011;12:326–332.
131. Mulatero P, Bertello C, Garrone C, et al. Captopril test can give misleading results in patients with suspect primary aldosteronism. *Hypertension*. 2007;50:e26–e27.
132. Lifton RP, Dluhy RG, Powers M, et al. A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*. 1992;355:262–265.
133. Jonsson JR, Klemm SA, Tunny TJ, et al. A new genetic test for familial hyperaldosteronism type I aids in the detection of curable hypertension. *Biochem Biophys Res Commun*. 1995;207:565–571.
134. Hundemer GL, Curhan GC, Yozamp N, et al. Incidence of atrial fibrillation and mineralocorticoid receptor activity in patients with medically and surgically treated primary aldosteronism. *JAMA Cardiol*. 2018;3:768–774.
135. Rossi GP, Maiolino G, Flego A, et al. Adrenalectomy lowers incident atrial fibrillation in primary aldosteronism patients at long term. *Hypertension*. 2018;71:585–591.
136. Kupers EM, Amar L, Raynaud A, et al. A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab*. 2012;97:3530–3537.
137. Young WF, Stanson AW, Thompson GB, et al. Role for adrenal venous sampling in primary aldosteronism. *Surgery*. 2004;136:1227–1235.
138. Gordon RD, Gomez-Sanchez CE, Hamlet SM, et al. Angiotensin-responsive aldosterone-producing adenoma masquerades as idiopathic hyperaldosteronism (IHA: adrenal hyperplasia) or low-renin essential hypertension. *J Hypertens Suppl*. 1987;5:S103–S106.
139. Guo Z, Nanba K, Udager A, et al. Biochemical, Histopathological, and genetic characterization of posture-responsive and unresponsive APAs. *J Clin Endocrinol Metab*. 2020;105;e:3224–3335.
140. O'Shea PM, O'Donoghue D, Bashari W, et al. 11) C-Metomidate PET/CT is a useful adjunct for lateralization of primary aldosteronism in routine clinical practice. *Clin Endocrinol*. 2019;90:670–679.
141. Powlson AS, Gurnell M, Brown MJ. Nuclear imaging in the diagnosis of primary aldosteronism. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:150–156.
142. Daunt N. Adrenal vein sampling: how to make it quick, easy, and successful. *Radiographics*. 2005;25:S143–S158.
143. Doppman JL, Gill Jr JR. Hyperaldosteronism: sampling the adrenal veins. *Radiology*. 1996;198:309–312.
144. Rossi GP, Auchus RJ, Brown M, et al. An expert consensus statement on use of adrenal vein sampling for the subtyping of primary aldosteronism. *Hypertension*. 2014;63:151–160.
145. Young WF, Stanson AW. What are the keys to successful adrenal venous sampling (AVS) in patients with primary aldosteronism? *Clin Endocrinol*. 2009;70:14–17.
146. Satoh F, Morimoto R, Seiji K, et al. Is there a role for segmental adrenal venous sampling and adrenal sparing surgery in patients with primary aldosteronism? *Eur J Endocrinol*. 2015;173:465–477.
147. Yoneda T, Karashima S, Kometani M, et al. Impact of new quick gold nanoparticle-based cortisol assay during adrenal vein sampling for primary aldosteronism. *J Clin Endocrinol Metab*. 2016;101:2554–2561.
148. Goupil R, Wolley M, Ungerer J, et al. Use of plasma metanephrine to aid adrenal venous sampling in combined aldosterone and cortisol over-secretion. *Endocrinol Diabetes Metab Case Rep*. 2015;2015:150075.
149. Williams TA, Lenders JWM, Mulatero P, et al. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabet Endocrinol*. 2017;5:689–699.
150. Hundemer GL, Vaidya A. Management of endocrine disease: the role of surgical adrenalectomy in primary aldosteronism. *Eur J Endocrinol*. 2020;183:R185–R196.
151. Catena C, Colussi G, Nadalini E, et al. Cardiovascular outcomes in patients with primary aldosteronism after treatment. *Arch Intern Med*. 2008;168:80–85.
152. Hundemer GL, Curhan GC, Yozamp N, et al. Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diabet Endocrinol*. 2018;6:51–59.
153. Hundemer G, Curhan G, Yozamp N, et al. Incidence of atrial fibrillation and mineralocorticoid receptor activity in patients with medically and surgically treated primary aldosteronism. *JAMA Cardiol*. 2018;3:768–774.
154. Hundemer G, Curhan G, Yozamp N, et al. Renal outcomes in medically treated primary aldosteronism. *Hypertension*. 2018;72:658–666.
155. Groth H, Vetter W, Stimpel M, et al. Adrenalectomy in primary aldosteronism: a long-term follow-up study. *Cardiology*. 1985;72:107–116.
156. Sukor N, Gordon RD, Ku YK, et al. Role of unilateral adrenalectomy in bilateral primary aldosteronism: a 22-year single center experience. *J Clin Endocrinol Metab*. 2009;94:2437–2445.
157. Baudrand R, Guarda FJ, Torrey J, et al. Dietary sodium restriction increases the risk of misinterpreting mild cases of primary aldosteronism. *J Clin Endocrinol Metab*. 2016. jc20161963.

97

Adrenocorticotropic Hormone–Independent Cushing Syndrome

Maria Candida Barisson Villares Fragoso, Annabel Berthon, and Jérôme Bertherat

OUTLINE

Introduction, 1614
Epidemiology, 1614
Pathophysiology—Molecular Genetics and Animal Models, 1615
  Molecular and Genetic Causes of Cortisol-Producing Adrenocortical Lesions, 1615
  Animal Models, 1618
Diagnosis, 1619
  Clinical Manifestations, 1619
  Hormonal Evaluation, 1619
Etiology, 1622
  Unilateral Tumors, 1622
  Adrenocortical Adenoma, 1622
  Adrenocortical Cancer, 1622
  Bilateral Adrenocortical Tumors, 1624
  Micronodular Adrenal Hyperplasia and Primary Pigmented Nodular Adrenal Disease, 1624
  Primary Macronodular Adrenal Hyperplasia, 1625
Management of Adrenal Cushing Syndrome, 1626
  Medical Treatment with Steroidogenesis Inhibitors, 1626
  Surgery, 1627
Perspectives, 1628

INTRODUCTION

KEY POINTS

- Cushing syndrome is a rare disease, characterized by chronic and excessive exposure to glucocorticoids.
- Cushing syndrome is a prototype of metabolic syndrome and results in high morbidity and mortality if not treated.
- During the last few years there have been improvements in our knowledge of the molecular mechanisms involved in the pathogenesis of adrenal Cushing syndrome.

Cushing syndrome (CS) is a rare endocrine condition, firstly described in 1912 by Harvey Cushing,[1] and is characterized by chronic and excessive exposure to glucocorticoids. CS can include significant morbidity and even mortality (if untreated), and its diagnosis and treatment can be challenging.[2] In humans, cortisol is the primary active glucocorticoid produced by the adrenal cortex under the control of pituitary adrenocorticotropic hormone (ACTH), and its actions include mobilization of fats, proteins, and carbohydrates. CS can be secondary to chronic stimulation of the adrenal cortex by ACTH secreted by a pituitary or ectopic ACTH-secreting tumor, or by a primary adrenal dysregulation. In the latter case, the causative adrenal lesions are independent of physiological regulation by pituitary ACTH, as the activity of the pituitary corticotroph cells is typically suppressed by glucocorticoid-mediated negative feedback; this can be classified as "adrenal Cushing." The adrenal lesions responsible for adrenal Cushing can be unilateral or bilateral, and can represent benign or malignant disease.

This chapter will discuss clinical aspects of the diagnosis and management of the various causes of adrenal Cushing and recent advances in our understanding of their molecular genetics.

EPIDEMIOLOGY

KEY POINTS

- Adrenal Cushing syndrome accounts for approximately 20%–25% of all cases of endogenous Cushing syndrome, second only to benign and malignant adrenal disorders.
- Adrenocortical tumors are frequent in the general population (3%–10% of adults), but the subset responsible for overt Cushing are rare, with the incidence being estimated at approximately 1.5/million/year, with a higher prevalence in females.

CS is a rare disease, with an estimated incidence ranging between 1 to 5 per million each year,[3] although the estimations, as is the case for many rare diseases, are imprecise. Approximately 20% to 25% of endogenous CS cases are secondary to adrenal gland disorders including adenomas, carcinomas, and nodular hyperplasias (which may be macronodular and micronodular, based on nodules size [>1 cm or <1 cm in diameter, respectively]).[4-6] The specific incidence of adrenal Cushing has rarely been investigated but may be around 1.5 per million per year.[7] On the other hand, adrenal tumors are quite common, affecting 3% to 10% of the human population. According to autopsy data, one in 10 adults (over the age of 50 years) has at least one adrenocortical nodule measuring up to 1 cm.[8] The majority are small benign or nonoversecreting adrenocortical adenomas (ACAs), but investigation of incidentalomas suggests that a proportion could produce mild cortisol excess and be related to typical features of the metabolic syndrome. The most prevalent somatic molecular defect alters the protein kinase A catalytic alpha-subunit, primarily in adenomas associated with overt CS. Interestingly, no such mutations occur in patients with mild cortisol secretion or with other adrenocortical tumors (ACTs).[5] By contrast, adrenocortical carcinoma (ACC) is a rare malignant disease that accounts for 0.05% to 0.2% of all human cancers, with an estimated incidence of 0.5 to 2 per million per year in adults.[9] The frequency is even lower in children, with only approximately 25 new cases of ACC occurring in the United States each year. Nevertheless, the incidence of pediatric ACT is remarkably high in South and Southeast Brazil, where it is estimated to be 15 times greater than the worldwide incidence, owing to a high frequency of a germline-specific pathogenic variant (p.R337H) of the *TP53* tumor suppressor gene.[10-12] In addition to the age of onset, pediatric ACT differs from adult tumors in many other ways.[13] Women are three times more likely to have either benign or malignant adrenal tumor, and approximately four to five times more likely to have CS associated with an adrenal tumor, but the reasons for this preponderance in women are unknown.[14-17]

All other causes of CS are extremely rare, including CS due to either primary macronodular adrenal hyperplasia (PMAH) or primary micronodular adrenal hyperplasia,[5] also known as primary bilateral macronodular adrenal hyperplasia (PBMAH) and previously called ACTH-independent macronodular adrenal hyperplasia (AIMAH), massive macronodular adrenocortical disease, and giant macronodular adrenal hyperplasia, among other names, is classically considered to represent less than 1% of cases of endogenous CS.[18-21] However, currently PMAH is often diagnosed after investigation of patients with bilateral adrenal incidentalomas associated with mild autonomous cortisol excess; therefore, the true prevalence is likely higher than previously thought.[22] PMAH presents a bimodal age distribution with a slightly female prevalence and a rare subset presenting in the first years of life, particularly in patients with the McCune–Albright syndrome.[23] Most patients with PMAH present in their fifth and sixth decades. The disease can be sporadic or associated with genetic syndromes (multiple endocrine neoplasia type 1 syndrome, familial adenomatous polyposis and hereditary leiomyomatosis, and renal cell cancer disorder) and more frequently occurs in familial form than earlier supposed, mostly due to germline *ARMC5* pathogenic variants.[24,25] Of the micronodular variety, primary pigmented nodular adrenocortical disease (PPNAD) is the most common endocrine manifestation in Carney complex (CNC), a rare autosomal dominant multiple endocrine neoplasia.[1] In approximately 60% of young patients with CNC, PPNAD causes adrenal Cushing, and more than 80% of these patients have germline pathogenic variants in *PRKAR1A*.[26]

Mortality from untreated adrenal CS is high, up to 2-fold that of the general population, mostly due to cardiovascular and cerebrovascular events and infections. After removal of a benign adenoma, survival is comparable to that of the general population; nevertheless, increased risk of infections and cardiovascular events appear to persist.[27] Mortality from malignant adrenocortical cancer is high for stage 3 and 4 of the disease and slightly better for patients with localized disease.[27] In CS due to PPNAD as part of the familial multiple endocrine neoplasia CNC, extraendocrine manifestations such as cardiac myxomas contribute to the excess mortality.[26]

PATHOPHYSIOLOGY—MOLECULAR GENETICS AND ANIMAL MODELS

KEY POINTS

- Mutations in regulatory members of the protein kinase A (PKA) signaling pathway (*PRKAR1A, PRKACA,* and *PDE* genes) lead to Cushing syndrome.
- *ARMC5* mutations are responsible for approximately 40% of primary macronodular adrenal hyperplasias, but the molecular relationship between its inactivation and cortisol hypersecretion remains unclear.
- Activating mutations in *CTNNB1*, which encodes β-catenin, are the second most frequent alteration identified in cortisol-producing adenomas (after *PRKACA* mutations).
- The crucial role of the opposition between PKA and Wnt/β-catenin pathway in maintaining normal adrenocortical zonation has been demonstrated using animal models.

The main molecular characteristic of a cortisol-producing adrenocortical lesion is abnormal activation of protein kinase A (PKA) signaling, which naturally controls cortisol secretion in response to ACTH (Fig. 97.1). To date, various genetic or molecular causes of this activation have been identified, and, interestingly, the adrenocortical features differ depending on the genetic/molecular origin of this activation. We will discuss herein: 1) the genetic/molecular alterations discovered in the different forms of ACTs and hyperplasias; 2) the advances made in our understanding of the adrenocortical physiopathology through the characterization of mouse models.

Molecular and Genetic Causes of Cortisol-Producing Adrenocortical Lesions

The important advances made in "omics" approaches have been great assets in order to discover the molecular alterations that are potentially responsible for cortisol-producing adrenocortical lesions. Interestingly, most of the genetic and molecular alterations identified in both cortisol-producing adenomas (CPAs) and hyperplasias affect regulatory members of the PKA pathway, which is consistent with the central role played by PKA in the regulation of cortisol secretion (Table 97.1). Although genetic alterations are mostly specific to a form of cortisol-producing adrenocortical lesion, some genes are, for instance, altered in both macronodular and micronodular hyperplasias. The known relationships between genetic/molecular alterations identified in patients and their adrenocortical presentation will be discussed herein.

Bilateral Adrenocortical Hyperplasia. Inactivating germline mutations affecting *PDE11A* and *PDE8B* have been identified in CS patients.[28] Both genes encode for phosphodiesterases (PDEs) that hydrolyze cyclic adenosine monophosphate (cAMP) and therefore inhibit the PKA pathway. Interestingly, these genetic alterations have been identified in PMAH, PPNAD, and ACA suggesting that they are directly involved in cortisol hypersecretion, probably through at least partial loss of their inhibitory effect on the cAMP/PKA pathway, which is abnormally activated as part of the different adrenocortical causes of CS.[28] Consistent with this, treatment of the human adrenocortical cell line H295 with a PDE11A inhibitor leads to an increase in cAMP and cortisol secretion. Interestingly, PDE alterations have been identified in association with others genetic defects, supporting the hypothesis that PDE mutations may be a secondary event that amplifies activation of cAMP/PKA signaling, resulting in hypercortisolism.[29]

Primary Macronodular Adrenal Hyperplasia. Mutations in the melanocortin 2 receptor (*MC2R*) gene encoding the ACTH receptor have been described in few cases of PMAH. These mutations lead to activation of the receptor, which increases cytosolic cAMP levels, resulting in release of the catalytic subunit of PKA and therefore inducing steroidogenesis and cortisol secretion (Fig. 97.1). Although MC2R-activating mutations are a rare cause of PMAH, the illegitimate expression of ectopic or eutopic membrane G protein-coupled receptors (GPCRs) mimicking ACTH response under the control of other hormones has for years been considered the main molecular characteristic of PMAH compared with other cortisol-producing tumors. Indeed, aberrant expression in cortisol-producing adrenocortical cells of β-adrenergic receptors (β-AR), vasopressin (V2–V3-vasopressin receptor), serotonin (5-HT7 receptor), and glucose-dependent insulinotropic peptide (GIPR) have been described in PMAH. In some others cases, the hypersecretion is due to altered expression of eutopic receptors like vasopressin (V1-vasopressin receptor), luteinizing hormone/human chorionic gonadotropin (LH/hCG), serotonin (5-HT4 receptor), and leptin (LEPR).[29] As the cause of this abnormal GPCR expression is mainly unknown, it is unclear whether this molecular alteration is a primary

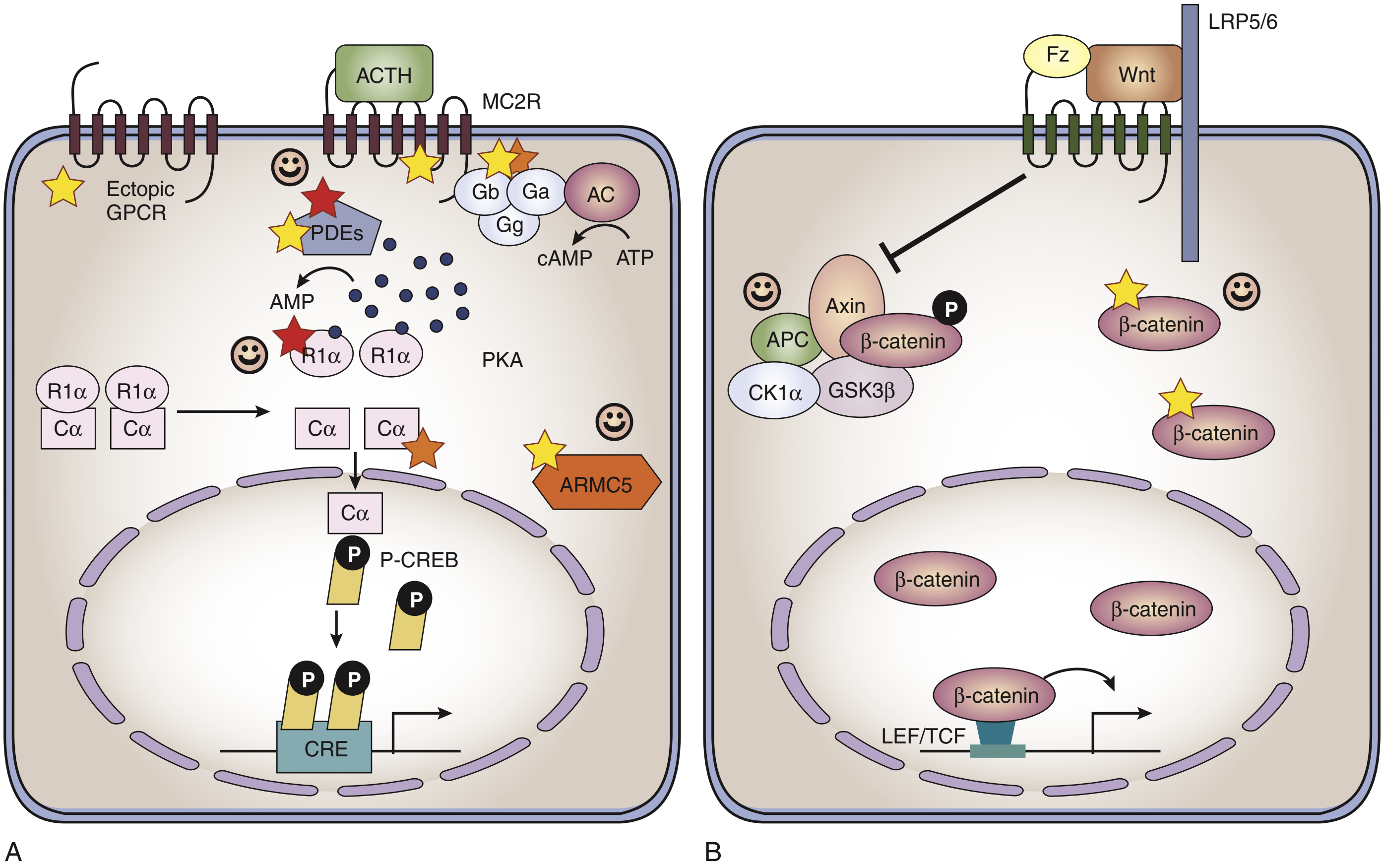


Figure 97.1 **Overview of the most frequently altered signaling pathways in cortisol producing adrenocortical lesions. A,** Protein kinase A (PKA) signaling. In response to adrenocorticotropic hormone binding to melanocortin 2 receptor, adenylate cyclase (AC) increases the cyclic adenosine monophosphate (cAMP) level after activation through Gsα. cAMP then binds to the regulatory (R1α) subunits of PKA that dissociate from the catalytic subunit (Cα). Free calcium phosphorylates cAMP response element binding protein (CREB), which, in turn, regulates the expression of steroidogenic genes. Stars indicate the elements that have been observed to be mutated in primary pigmented nodular adrenocortical disease *(red)*, primary macronodular adrenal hyperplasia (PMAH) *(yellow)*, and cortisol-producing adenoma (CPA) *(orange)*. The smiley face indicates the genetic alterations that have been mimicked in mouse models. **B,** In absence of Wnt ligand, β-catenin is phosphorylated by a destruction complex composed of GSK3β, APC, CK1α, and axin. Phosphorylated β-catenin is then degraded by the proteosome. However, in presence of ligand, the destruction complex is inhibited, allowing the nucleocytoplasmic accumulation of β-catenin, which acts as a cotranscriptional activator and activates genes involved in cell proliferation. The stars highlight the genetic alterations that have been found in PMAH *(yellow)* and CPA *(orange)*.

or a secondary event in PMAH. However, it has been reported that GIPR overexpression is due in a few cases to duplication of chromosome 19q13.32, including the *GIPR* gene.[30]

The bilateral presentation of the macronodules that characterizes PMAH suggests the presence of a germline genetic predisposition. Accordingly, PMAH cases have been reported in patients presenting familial syndromes such as multiple endocrine neoplasia type 1, familial adenomatous polyposis, and hereditary leiomyomatosis and renal cancer syndrome. The presence of PMAH in these patients suggest that mutations in the menin 1 (*MEN1*), adenomatous polyposis coli (*APC*) or fumarate hydratase (*FH*) genes causing these syndromes could lead to adrenocortical hyperplasia, but these alterations only explain a small subset of PMAH cases.[29] Therefore, in order to identify more frequent cause of PMAH development, blood and tumor DNA from 33 PMAH patients have been analyzed using different genomic approaches. Interestingly, the most frequent somatic chromosome alteration and mutation are both located at 16p. Indeed, whereas a loss of heterozygosity at 16p was found in eight out of 33 patient tumors (24%), mutations in a gene called armadillo repeat-containing protein 5 (*ARMC5*) located at 16p11.2 was described in 18 out of 33 patients (55%).[31] Since its initial discovery in 2013, genotyping of *ARMC5* has been performed in several PMAH cohorts in centers worldwide, and it is now estimated that germline *ARMC5* mutations are responsible for approximately 40% of PMAH cases.[32] Interestingly, most *ARMC5* mutations are frameshift or nonsense mutations, leading most probably to a loss of *ARMC5* function, and its germline defect is associated with somatic mutations or loss of heterozygosity in the nodules' DNA.[29] Altogether, these observations support a tumor suppressor role of *ARMC5* in adrenocortical cells. However, as ARMC5 function is still unclear, one important remaining challenge is to determine if germline missense *ARMC5* variants are deleterious to its function. This is especially important, as several familial form of PMAH due to *ARMC5* mutations have been reported.[32] Interestingly, although adrenocortical size and hormonal phenotype vary among members of *ARMC5* families, *ARMC5*-mutated patients have a particular clinical phenotype characterized by overt CS and an increase in adrenal size compared with PMAH patients without *ARMC5* defects. These observations suggest that loss of *ARMC5* function drives the development

TABLE 97.1 Genetic Alterations Identified in Cortisol-Producing Bilateral Micronodular Hyperplasia (Primary Pigmented Nodular Adrenocortical Disease [PPNAD]), Cortisol-Producing Bilateral Macronodular Hyperplasia (Primary Macronodular Adrenal Hyperplasia [PMAH]), and Cortisol-Producing Adenoma (CPA)

Adrenocortical presentation	Genes	Chromosomal location	Genetic Event	Identified Mutations	Molecular Function	References
Reported in PPNAD and PMAH	*PDE8B*	5q13	Inactivating mutations	p.H305P in PPNAD and p.H391A in PMAH	Hydrolysis of cAMP, inhibition of PKA signaling	26
	PDE11A	2q31	Inactivating mutations	p.R307* in PPNAD	Hydrolysis of cAMP and cGMP, inhibition of PKA signaling	26
PMAH	*MC2R*	18p11	Activating mutations	-	ACTH receptor, activation of PKA signaling	
PMAH	*GNAS*	20q13	Postzygotic activating mutations	-	Ga subunit GTPase, activation of PKA signaling	
PMAH	*APC*	5q12-22	Inactivating mutations	-	Inhibition of Wnt/β-catenin pathway	
PMAH	*MEN1*	11q13	Inactivating mutations or deletion	-	Controls gene transcription and cell proliferation	
PMAH	*FH*	1q42	Inactivating mutations	-	Fumarate conversion to malate in Krebs cycle	
PMAH	*ARMC5*	16p11	Inactivating mutations	More than 50 deleterious mutations	Controls steroidogenesis and proliferation	30
PPNAD	*PRKAR1A*	17q24	Inactivating mutations or deletion	More than 100 mutations	Regulatory subunit of PKA, inhibition of PKA signaling	
PPNAD and PMAH	*PRKACA*	19p13	Duplication or triplication	-	Activation of PKA pathway	
CPA	*PRKACA*	19p13	Somatic activating mutations	Hot spot: p.L206R	Catalytic subunit α, activation of PKA signaling	32
CPA	*PRKACB*	1p31	Somatic activating mutations	p.S54L (one case)	Catalytic subunit β, activation of PKA signaling	37
CPA	*GNAS*	20q13	Somatic activating mutations	p.R201 substitutions	GTPase, activation of PKA pathway	38
CPA	*CTNNB1*	3p22	Somatic activating mutations	p.S45 substitutions	Activation of Wnt/β-catenin pathway	38

cAMP, cyclic adenosine monophosphate; *PKA,* protein kinase A; *cGMP,* cyclic guanosine monophosphate; *ACTH,* adrenocorticotropic hormone.

of a subtype of PMAH, even if the second *ARMC5* event occurring in adrenocortical nodules may affect the appearance of disease features as observed in relatives.

Primary Pigmented Nodular Adrenocortical Disease. As PPNAD is the main endocrine manifestation of a familial multiple neoplasia syndrome called CNC, it is mainly due to mutations in *PRKAR1A* that are found in the germline of more than 80% of CNC patients. To date, more than 100 mutations throughout the *PRKAR1A* gene have been identified in at least one CNC family. In most cases (83%), these mutations result in a premature stop codon.[33] Although there are three other regulatory subunits of PKA (PRKAR1B, PRKAR2A, PRKAR2B), this haploinsufficiency of the regulatory subunit R1α of PKA encoded by *PRKAR1A* gene is sufficient to increase the activity of the cAMP/PKA pathway, especially in adrenocortical cells. Moreover, based on the critical role of *PRKAR1A* in the regulation of PKA signaling, it is not surprising that somatic mutations or loss of the *PRKAR1A* locus have also been identified outside the CNC spectrum, in ACA and carcinoma.

More recently, genome-wide copy-number analysis has identified the presence of *PRKACA* copy number gain in three PPNAD patients with histological features similar to the those observed in patients with *PRKAR1A* haploinsufficiency.[34,35] As expected, knowing that *PRKACA* encodes the catalytic subunit α of PKA, functional analyses confirmed that these genetic alterations cause a gain of Cα function and therefore lead to activation of cAMP/PKA signaling. Indeed, *PRKACA* is overexpressed at both the transcript and protein level in the adrenocortical tissues from patients exhibiting *PRKACA* copy number gain. Interestingly, out of these three patients, two presented a *PRKACA* triplication associated with a younger age at diagnosis, suggesting a dose-dependent effect of *PRKACA* on the adrenocortical phenotype. Moreover, two related patients, mother and son, with *PRKACA* duplication have bilateral adrenocortical hyperplasia similar to PMAH and not PPNAD, suggesting that the adrenocortical presentation may vary depending on the *PRKACA* copy number gain.[35]

Unilateral Adrenocortical Adenoma. In 2014, exome sequencing data highlighted for the first time the presence of point mutations in *PRKACA,* which encodes the catalytic subunit α of PKA, in eight out of 10 CPAs.[34] Based on *PRKACA* genotyping data available from large cohorts of patients, the frequency of *PRKACA* mutations in CPA is now estimated to be between 27% and 40%, depending on the study.[36] Interestingly, somatic *PRKACA* mutations are highly specific to patients with overt Cushing caused by a unilateral adenoma, as they are very rare in other adrenocortical lesions.[34,36] These mutations are mainly found in a hot spot region of *PRKACA,* including the most frequent mutation, which is a missense mutation affecting amino acid 206 (p.L206R). *In silico* and *in vitro* functional analysis demonstrates that p.L206R alters the interaction between the catalytic and regulatory subunits, destabilizing the PKA holoenzyme and leading both to continuous activation of mutant Cα[34] and degradation of the regulatory subunit RIIβ (*PRKAR2B*) protein of PKA by a caspase 16-dependent mechanism.[37] Accordingly, CPAs with the *PRKACA* p.L260R mutation belong to the previously described group that exhibits lower RIIβ expression.[37,38]

In addition to p.L260R, six missense variants in *PRKACA* have been identified in at least one CPA, with all except one (p.E32V) located in the hot spot region at the interface with the regulatory subunit.[39] As previously observed for p.L260R, three others mutations (199_200insW, 200_201_insV, and S213R+insIILR) partially abolish the interaction with both RIα and RIIβ, whereas no clear difference was observed after overexpression of p.W197R and E32V. However, Bathon and Colls demonstrated that *PRKACA* mutations not only increase global PKA activity but also favor phosphorylation of specific substrates, such as histone 1 (H1), both *in vitro* and *in vivo*. Indeed, a 12- and 9-fold elevation of H1.4 phosphorylation is observed after overexpression of the p.L205R mutant in HEK293 and in CPA carrying the p.L206R mutation compared with nonmutated CPA, respectively.[39] Altogether, these results suggest that PKA activation due to *PRKACA* mutations could explain cortisol hypersecretion, whereas stimulation of cell proliferation could be linked at least partially to the hyperphosphorylation of new substrates like H1.4. Further studies are required to better characterize the molecular consequences of these mutations.

The genetic alterations affecting genes encoding the catalytic subunit of the PKA are not limited to *PRKACA,* as one mutation has recently been described in *PRKACB,* which encodes the catalytic subunit β of PKA, in a CPA of a patient with severe CS.[40] The missense variant, p.S54L, increases the sensitivity of the tetramer to cAMP and therefore leads to an elevation of PKA activity under basal conditions.[40] Although this is a rare genetic event identified in CPA, it highlights that genetic alterations affecting members of the PKA holoenzyme and favoring PKA activation play an important role in the development of cortisol-producing adrenocortical lesions.

Next-generation whole-exome sequencing was performed on CPAs not mutated at the *PRKACA* hots pot in order to identify others genetic alterations responsible for CPA.[41] This study resulted in the identification of genetic alterations both in known and new genes.[41] Indeed, as previously reported,[42] activating mutations have been identified in the *GNAS* gene encoding subunit α of the G protein. These mutations inactivate GTPase activity and therefore constitutively activate PKA signaling indirectly through activation of the adenylate cyclase.[41] Consistent with the causal role of eight somatic *GNAS* mutations identified in 74 CPAs, adrenal CS has been reported in 7.1% of patients with McCune–Albright syndrome due to postzygotic *GNAS* genetic defects.[29]

Somatic activating mutations in the *CTNNB1* gene encoding β-catenin are the most frequent genetic alteration identified after *PRKACA* mutations in CPA. In contrast to the other genetic alterations described above, constitutive activation of β-catenin does not cause PKA activation, but rather constitutive activation of the Wnt/β-catenin canonical pathway. β-catenin is degraded in the absence of Wnt ligands but accumulates in the cytoplasm and the nucleus of activated cells (Fig. 97.1). This activation, which is essential for normal adrenocortical development and differentiation, is tightly regulated, because abnormal β-catenin activation due, for instance, to *CTNNB1* missense mutations is sufficient to induce adrenocortical tumorigenesis.[43] Indeed, constitutive β-catenin activation characterized by nucleocytoplasmic β-catenin staining is the most frequent molecular alteration identified in both ACA and ACC.[43] Interestingly, the observation of constitutive β-catenin activation in nonsecreting and aldosterone-secreting ACTs indicates important variability in patient phenotypes that may be mediated by unique mutations or additional modifier genes.[43] In addition to *CTNNB1* activating mutations, alterations affecting others regulatory members of this pathway, particularly inactivating mutations of *APC* or *ZNRF3* have been identified in ACTs.[44] It can be hypothesized that the patients' phenotype and the form of the adrenocortical lesions rely on the genetic cause(s) as well as on the β-catenin dosage resulting from this alteration.

Animal Models

Animal models are essential assets for understanding physiology and physiopathology. Indeed, their characterization provides essential clues regarding the signaling pathways and molecular mechanisms responsible, for instance, for tumor growth and glucocorticoid hypersecretion in CS that cannot always be explored *in vitro*. Although a broad range of studies can be performed using animal models, this part of the chapter will focus on studies aiming to 1) elucidate the molecular causes of spontaneous CS development in certain species; 2) phenotype transgenic models carrying genetic alterations identified in CS patients.

Animal Models with Spontaneous Adrenocortical Tumors. Both ferrets and mice develop spontaneous adrenocortical lesions after gonadectomy. In both cases, it is thought that the chronic elevation of LH due to gonadectomy plays an important role in the adrenocortical tumorigenesis. Moreover, molecular studies revealed ectopic expression of LH receptor and CYP17, which are not robustly expressed in the adrenal (and are primarily gonadal proteins) in adrenocortical lesions.[45] This is associated with a switch from expression of the transcription factor *Gata4* to expression of *Gata6,* which normally regulates gonadal-specific gene expression and cell proliferation. Consistent with the gonadal-like cell phenotype identified in these adrenocortical lesions, the vast majority of these tumors in ferrets secrete sex steroids rather than glucocorticoid, even though a case of LH-dependent CS has been reported. Similar ACTs have been observed in other gonadectomized animals, including guinea pigs, hamsters, goats, and cats. Although sequencing of the ferret genome and identification of additional single nucleotide polymorphisms (SNPs) in mice and other model species would facilitate the characterization of genes influencing gonadectomy-induced adrenocortical tumorigenesis, studies have already identified genetic alterations associated with these tumors. For instance, linkage analysis in nonsusceptible (C57BL/6J) and susceptible (DBA/2J) mice already identified a major locus on chromosome 8 including, among others, *Sfrp1*, which encodes a negative regulator of Wnt/β-catenin, as a potential genetic factor for susceptibility. Moreover, the p.A172S substitution in steroidegenic factor 1 (*Sf1*) has been observed in a mouse strain with high tumor susceptibility and modified adrenocortical cell steroidogenic ability. Therefore, based on the role of Sf1 dosage in tumor growth in both human ACC and transgenic mouse models, it can be hypothesized that this polymorphism could lead to an increase in Sf1 expression and therefore promote ACT growth.[45]

CPAs are frequent in dogs, with one to two new cases per 1000 dogs reported per year. They are mostly unilateral, with only approximately 10% being bilateral lesions. In contrast to mice and ferrets, gonadectomy does not modify the incidence of CPAs in dogs; instead, they occur mainly in middle and old age, without sex predilection. As in humans, the diagnosis of adrenal CS is based on the determination of elevated circulating or urinary cortisol concentrations, which is not suppressed in response to a high dose of dexamethasone but is suppressed with suppressed circulating ACTH concentration.[45] Adrenalectomy is the best therapeutic option, but if the tumor is inoperable mitotane remains an option to target ACTs and metastases. A competitive inhibitor of 3β-hydroxysteroid dehydrogenase, trilostane, is used to control the clinical signs of glucocorticoid excess. The canine genome and SNP map, which are now publicly available, along with

SNP array technology could facilitate the use of spontaneous canine ACTs as a model to study adrenocortical tumorigenesis and to test potential treatment of these tumors that share a high degree of similarity with human ACTs.[45]

Xenografted and Genetically Modified Mouse Models. In recent decades, the identification of genetic alterations responsible for the development of ACT using omics approaches has been combined with transgenic techniques in mice to obtain crucial understanding of the molecular mechanism involved in adrenocortical tumorigenesis and hormonal secretion. To date, both xenografted and transgenic mice have been used to investigate the involvement of specific pathways. Although many models did not fully reproduce the human disorders, they provided important clues, such as highlighting the essential role played by PKA and the Wnt/β-catenin pathway, not only in adrenocortical tumorigenesis but also during adrenal cortex development and differentiation. This section provides a brief overview of the main findings made through the characterization of mouse models developed to study PKA signaling. More comprehensive reviews of mouse models developed to study adrenocortical function are already available.[46,47]

Consistent with the frequent genetic alterations affecting PKA regulators discovered in human CPAs, a number of mouse models presenting cAMP/PKA defects develop adrenocortical hyperplasia and cortisol hypersecretion. For instance, *Prkar1a* knockout is lethal in mice at embryonic day 9.5, and adrenocortical lineage-specific knockout has also been investigated. Interestingly, these mice develop a PPNAD-like phenotype characterized by a glucocorticoid hypersecretion associated with ACTH suppression and adrenocortical hyperplasia.[48] In accordance with the negative regulatory function of RIα on PKA activity, the endocrine hyperactivity is due to an abnormal activation of PKA signaling. This study therefore confirms that *PRKAR1A* loss is sufficient to induce adrenocortical hyperplasia and glucocorticoid hypersecretion through PKA activation. Furthermore, this phenotype is observed in females, first demonstrating that PKA effects are influenced by sex, which is consistent with the sexual dimorphism observed in most human adrenocortical disorders. Different approaches have been used in mouse models to elucidate the causes of this sexual dimorphism.[49,50] It appears that adrenal cortex renewal is faster in females than in males, partly because of dihydrotestosterone secretion in males. These studies illustrate the essential advances made in the comprehension of adrenocortical physiopathology through the characterization of animal models.

Altogether, the studies performed in the various models discussed herein and beyond have provided new knowledge and complementary information that can be translated to human disease. Moreover, in addition to the crucial molecular knowledge acquired through their characterization, these models can also be used for preclinical trials in order to test potential new, more targeted treatments for CS, which are needed to improve patients' conditions.

DIAGNOSIS

KEY POINTS

- Before evaluating for possible Cushing syndrome (CS), it is essential to obtain a careful history, excluding exogenous glucocorticoid intake.
- The diagnosis of CS depends on at least two different first-line tests (urinary free cortisol, midnight salivary cortisol, and cortisol after low doses of dexamethasone suppression test).
- The diagnosis of CS involves three steps: suspicion based on signs and symptoms, documentation of the presence of hypercortisolism, and determination of its cause.

Clinical Manifestations

CS is the prototypical metabolic syndrome. It is important to emphasize that, in severe hypercortisolism, catabolic changes are most useful for making the differential diagnosis. The major discriminative symptoms and signs of CS in adults include the following: thinning of the skin, facial plethora, proximal muscle weakness, extremity wasting with increased fat in the abdomen and face, wide purplish striae, and bruising with no apparent trauma (Fig. 97.2).[27] In children/adolescents, there is a retardation of growth velocity or cessation of growth and arrest of pubertal development associated with remarkable obesity (Table 97.2).[51] The less discriminative symptoms and signs comprise: global obesity, supraclavicular fat pads, infertility, hypogonadotropic hypogonadism, diabetes, dyslipidemia, hypertension, osteopenia, obstructive sleep apnea, and more (Table 97.3).

An essential clue to the presence of glucocorticoid excess is the simultaneous development and increasing severity of several of these symptoms. Nevertheless, many of these signs and symptoms can occur in individuals without hypercortisolism, and not all patients with CS present with obvious features.[51]

Clinical manifestations of CS are classified according to the involvement of the different systems: cardiovascular, reproductive, musculoskeletal, dermatologic, and neuropsychiatric. In addition, hypercortisolism is associated with a higher susceptibility to infections and to an immunosuppressive state.[52]

The presence and relative severity of symptoms vary, as well as the degree and duration of hypercortisolism and the presence or absence of androgen excess (because hypercortisolism alone causes neither hirsutism nor pustular acne). Androgen excess occurs only in women with ACC or ACTH-stimulated hyperandrogenism (the adrenal gland being a major source of androgen production in women but not men). ACA causing adrenal Cushing generally secretes only glucocorticoids.[53]

In patients with ACAs, both the degree of hypercortisolemia and many of the clinical manifestations of CS tend to be less severe in patients greater than 50 years of age.[53] Many patients with incidentally discovered ACAs have mild autonomous cortisol excess (MACE) without specific signals and symptoms of CS; however, glucose intolerance, obesity, and hypertension are common.[54]

Menstrual irregularities are common in women with CS. The menstrual abnormalities correlate with increased serum cortisol and decreased serum estradiol concentrations, but not with serum androgen concentrations. Because both LH and follicle-stimulating hormone levels are low, the menstrual irregularities appear to be due to the suppression of the secretion of gonadotropin-releasing hormone by high levels of cortisol.[55]

Women with some types of CS have signs of androgen excess (most commonly ACC). Signs of androgen excess in CS are most common in women with adrenal carcinomas. These tumors often secrete large amounts of androgenic precursors, because they are inefficient at converting cholesterol to cortisol. Androgen excess in affected women can cause hirsutism, oily facial skin and acne on the face, neck, or shoulders, and increased libido.[55] ACC causes virilization, including temporal balding, voice deepening, and clitoral hypertrophy. Prepubertal males/females may develop precocious isosexual or heterosexual puberty, respectively.[56]

Hypertension, low bone mineral density, and pathologic fractures (ribs, feet, and vertebrae are frequently affected sites) are common, and osteopenia in young patients with CS should raise the index of suspicion of PPNAD. More rarely, adenomas may also secrete various steroids (e.g., cortisol and aldosterone), androgens, or estrogens, but this should raise the suspicion of a malignant tumor.[5]

Hormonal Evaluation

The diagnosis of CS involves three steps: suspicion based on signs and symptoms, documentation of the presence of hypercortisolism, and

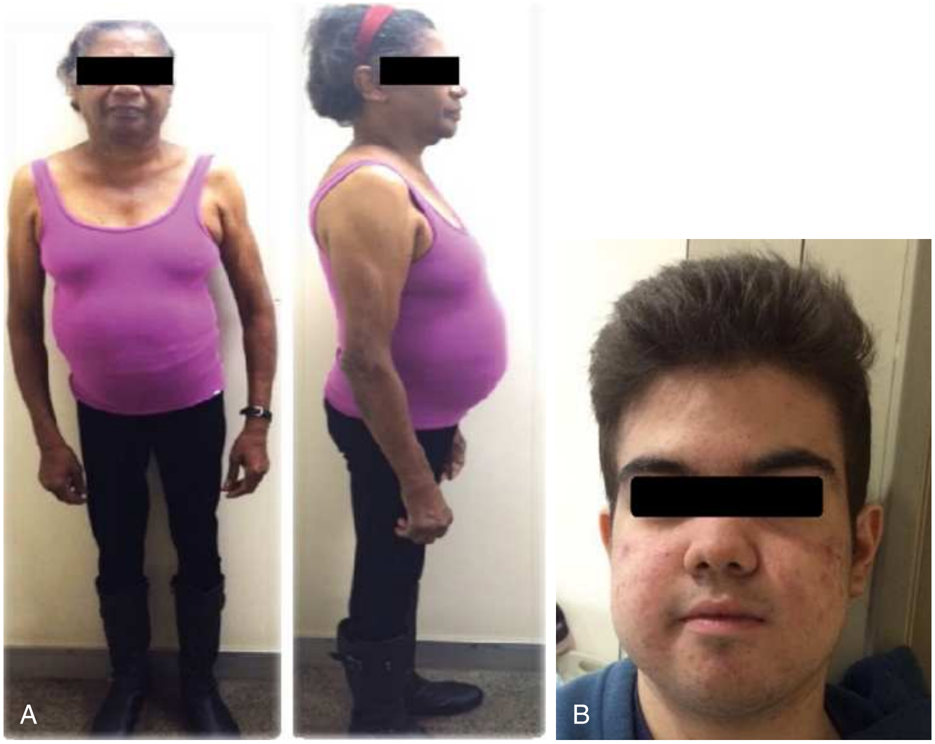

[redacted]
[redacted]
[redacted]

Figure 97.2 Adrenal Cushing syndrome in a patient with primary macronodular adrenal hyperplasia with truncal obesity, facial fullness, supraclavicular fat deposit, and proximal myopathy (**A**). Patient with mixed syndrome with facial roundness and acne, due to tumoral cortisol and androgen secretion caused by adrenocortical cancer (**B**).

TABLE 97.2 More Discriminative Signs and Symptoms of Cushing Syndrome

SIGNS AND SYMPTOMS	
More Discriminative	**Percentage (%)**
Facial plethora	90
Easy bruising	65
Proximal myopathy/weakness	60
Unusual osteopenia or osteoporosis/fracture for age	50
Skin finding*	75
Decreasing growth velocity	70–80
Striae (especially if red or purple and more than 1 cm wide)	70–90

*Skin atrophy, acne, plethora, hypertrichosis, hirsutism, acanthosis nigricans, hyperpigmentation, alopecia, purpura, and many other dermatological conditions.

TABLE 97.3 Less Discriminative Signs and Symptoms of Cushing Syndrome

SIGNS AND SYMPTOMS	
Less Discriminative	**Percentage (%)**
Weight gain or obesity/abdominal obesity	95
Facial fullness	90
Decreased libido/erectile dysfunction	90
Menstrual abnormalities/amenorrhea	80
Arterial hypertension	75
Hirsutism	75
Depression/emotional lability	70
Dyslipidemia	70
Impaired glucose tolerance/diabetes mellitus	60
Dorsocervical fat pad/supraclavicular fullness	50–70
Kidney stones	50
Exophthalmos	45
Polyuria	30
Headache	20–50
Back pain	20–50
Peripheral edema	20–50
Unusual infections/fungal infections	20–50
Hypokalemia	10–70
Acne	<20
Alopecia/female balding	<20
Hyperpigmentation	10

determination of its cause. The literature reports a variety of approaches and test combinations.[57] All tests are biochemical assessments of abnormalities of the hypothalamic–pituitary–adrenal axis. Before evaluating for possible CS, it is essential to obtain a careful history and to exclude exogenous glucocorticoid intake. The initial diagnostic tests for hypercortisolism should be extremely sensitive, even though the diagnosis may be excluded later by more specific tests.[57] The diagnosis of CS depends on at least two different first-line tests being unequivocally abnormal, as described in Fig. 97.3.[57] These include abnormal overnight dexamethasone suppression test with 1 mg (morning cortisol cutoffs <1.8 ug/dL or 50 nmol/L), at least two samples of low or suppressed ACTH, loss of circadian rhythm (increased serum or salivary midnight cortisol levels), and increased urinary free 24-hour cortisol.[58,59] The 1-mg dexamethasone suppression test is considered to be very sensitive for adrenal Cushing. However, when it is the only alteration, and the cortisol level is between 50 and 138 nmol/L, the patient can be considered as having possible autonomous cortisol excess. It is then advised to repeat the test few months later. If the test is still abnormal, and a third investigation remains abnormal, the diagnosis of MACE can be made. Some rare patients have cyclic CS, characterized by excess cortisol episodes alternating with periods of normal cortisol secretion. Hypercortisolism episodes can occur regularly or irregularly, with intercyclic phases ranging from days to months.[57] For patients with suspected cyclic CS, it has been suggested that urinary free cortisol (UFC) or late-night salivary cortisol can facilitate making the diagnosis better than the dexamethasone suppression text. If initial testing is normal, but clinical suspicion is high, follow-up with repeated testing is suggested. True cyclic CS is relatively rare, however, and variability in UFC excretion is common to all forms of CS.[58]

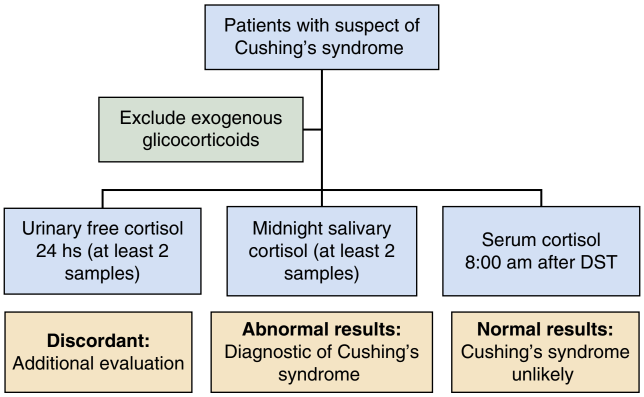

Patients with suspect of Cushing's syndrome
Exclude exogenous glicocorticoids
Urinary free cortisol 24 hs (at least 2 samples)
Midnight salivary cortisol (at least 2 samples)
Serum cortisol 8:00 am after DST
Discordant: Additional evaluation
Abnormal results: Diagnostic of Cushing's syndrome
Normal results: Cushing's syndrome unlikely

Figure 97.3 The Endocrine Society's Clinical Practice Guidelines recommend screening for Cushing syndrome using tests with high accuracy to evaluate hypothalamic-pituitary-adrenal axis integrity: 1) Excess total endogenous cortisol secretion, based on measuring 24-h urinary free cortisol levels; 2) loss of the circadian rhythm of cortisol secretion, based on midnight or late night salivary cortisol levels; 3) loss of normal cortisol feedback, assessed with the low-dose (1-mg) dexamethasone suppression test to identify autonomous cortisol secretion.

Once the diagnosis of CS is established, additional evaluation is done to identify the cause of the hypercortisolism. Measurement of plasma ACTH is therefore critical to differentiating ACTH-dependent from ACTH-independent CS. ACTH is secreted in a pulsatile fashion, dictating the same circadian rhythm to cortisol, and is extremely sensitive to stress. Its half-life in plasma is short due to plasmapeptidase activity. For these reasons, clinicians should collect several blood samples in tubes containing enzymes inhibitors (preferably two or three samples over 30 minutes) and maintain the samples at low temperatures before separation into aliquots.[60] Plasma ACTH concentrations are low in most patients with adrenal CS; however, up to 25% may present values within the normal range. It remains uncertain whether assay methodology contributes to this unexpectedly normal lab value, or whether there is incomplete suppression of pituitary ACTH secretion.[58,59] ACTH concentrations have rarely been frankly nonsuppressed in cortisol-secreting ACC due to the ectopic production and autocrine action of ACTH by the ACC.[51] In addition to patients with adrenal CS presenting measurable or normal ACTH concentrations, 10% of patients with Cushing disease display ACTH values below 10 pg/mL (2 pmol/L), thereby blurring the boundaries between the etiologies. To correctly diagnose patients falling into this gray area, stimulation with corticotropin-releasing hormone can prove useful; patients with pituitary-dependent hypercortisolism will mount an ACTH response to the stimulus, whereas patients with ACTH-independent hypercortisolism will not.[51]

Recently, adrenal steroids other than cortisol have been used for diagnosis of subclinical hypercortisolism, as measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS). This has been further supported by an LC-MS/MS steroid profiling study establishing a panel of steroids that can be used to identify patients with CS and discriminate between those with and without ACTH-dependent subtypes.[27] Other studies have indicated that urinary steroid profiling

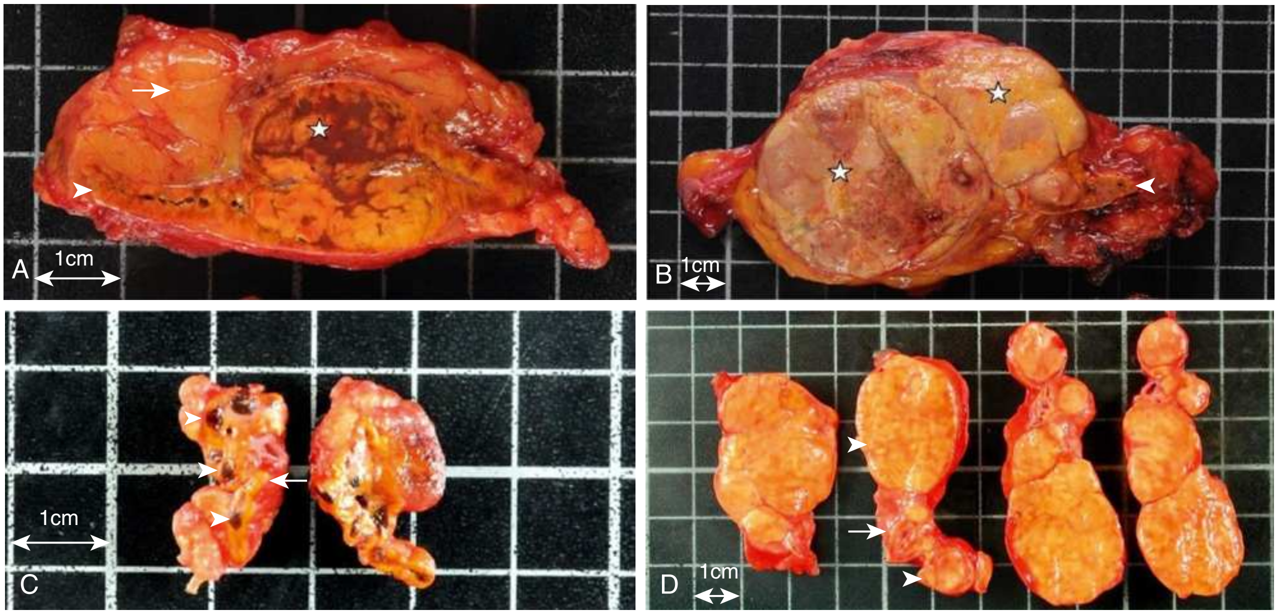


Figure 97.4 Illustration of the adrenal lesions observed in patients with Cushing syndrome: macroscopic analysis. **A,** An adrenocortical adenoma 2 cm in diameter *(star)* with its adjacent adrenal gland and adipose tissue indicated by the *arrowhead* and *arrow*, respectively. **B,** A 7-cm multinodular adrenocortical carcinoma *(star)* and the adjacent adrenal gland *(arrowhead)*. **C,** Primary pigmented nodular adrenocortical disease presenting 0.1–0.3-cm pigmented micronodules *(arrowhead)* and the internodular adrenal *(arrow)*. **D,** Primary macronodular adrenal hyperplasia with macronodules ranging in size from 0.5 to 3 cm *(arrowhead)* and internodular adrenal indicated by the *arrow*. (Courtesy of Dr. Florian Violon and Mathilde Sibony, Cochin Hospital.)

may be useful for differential diagnosis of benign versus malignant ACTs.[61,62] These methods are predicted to improve both diagnostic accuracy and tumoral surveillance of adrenal tumors.

ETIOLOGY

KEY POINTS

- Adrenal Cushing can be caused by bilateral or unilateral neoplasia (hyperplasias or tumors).
- Bilateral tumors are more often associated with genetic origin.
- The evaluation includes clinical, hormonal, and radiological imaging and possible genetic investigation.
- Adrenocortical adenomas is frequent in the general population, frequently diagnosed in patients with adrenal incidentaloma, and much more often responsible for mild autonomous cortisol excess (MACE) than overt Cushing.
- Adrenocortical carcinoma is a rare tumor with a poor prognosis.
- Primary bilateral macronodular adrenal hyperplasia is a rare adrenal cause of overt Cushing but is more frequently responsible for MACE.
- Primary pigmented nodular adrenocortical disease is responsible for overt Cushing and is most often associated with Carney complex.

Adrenal Cushing can be caused by bilateral or unilateral tumors (Fig. 97.4). Unilateral tumors are more frequent than bilateral tumors, and are usually sporadic in adults. In contrast, bilateral tumors are more often of genetic origin. Unilateral tumors can be benign (ACA) or malignant (ACC). ACAs are common in the general population, diagnosed incidentally, do not often secrete excessive cortisol and hence do not frequently result in over Cushing syndrome.[63,64] ACCs are rare tumors with a poor prognosis. Bilateral tumors can be divided in two groups, with clearly different pathology: micronodular adrenal hyperplasia and PMAH. These two entities are clearly different from the pathophysiological, as well as the morphological and clinical, points of view (Fig. 97.4). Clinical and hormonal investigations associated with imaging, and in some specific situations genetic analysis, can in most cases diagnose the cause of adrenal Cushing. However, in operated patients the pathology finding is fundamental, especially for the diagnosis of malignant tumors.

Unilateral Tumors

The clinical, hormonal, and imaging presentation of ACA can differ from those of ACC. However, when the ACC is localized to the adrenal gland (European Network for the Study of Adrenal Tumors [ENSAT] stage 1 or 2), the differential diagnosis can be difficult, and a careful pathological examination is fundamental. The histological diagnosis between a benign and a malignant adult ACT is based on the Weiss score system, and a malignant ACT must meet at least three out of nine criteria.[65,66]

Adrenocortical Adenoma

Cortisol-secreting adrenal adenoma is diagnosed by the investigation of clinical signs of CS or by the investigation of adrenal incidentaloma. In the latter, cortisol autonomous secretion can be mild or even undetectable by routine hormonal investigations in patients with no specific signs of CS. From the clinical perspective these two situations are clearly different and will be presented separately. At pathological investigation, cortisol-producing adenomas are enriched in zona fasciculata-like clear cells with variable amounts of compact cells or pigment deposition. The presence of cortical atrophy in the adjacent adrenal cortex in the case of cortisol excess reflects the decreased levels of circulating ACTH, which has a trophic effect required for maintenance of the adrenal cortex. The Weiss score of an ACA is most often 0 or 1, and less frequently 2.

ACAs responsible for overt CS are usually diagnosed between the age of 35 to 50 years, and more frequently in women.[67] They represent the most frequent adrenal cause of CS. A benign adrenal cortisol secreting tumor typically oversecretes cortisol but no androgens nor mineralocorticoids, in contrast to malignant adrenal tumors. The clinical signs are therefore due to cortisol excess with facial-truncal obesity, hypertension, and diabetes, as well as catabolic signs (muscle weakness, osteoporosis), often starting 2 to 4 years before diagnosis.

In ACA responsible for overt CS, the UFC, as well as midnight salivary or plasma cortisol, is increased. Plasma cortisol is clearly elevated after the 1-mg dexamethasone test, and ACTH levels are usually completely suppressed. Levels of the adrenal androgen dehydroepiandrosterone (DHEA) are low due to the lack of ACTH stimulation. On imaging, these adenomas can have typical characteristics of benign lipid-rich tumors on computed tomography (CT) scan or magnetic resonance imaging (MRI) (Fig. 97.5). However, spontaneous density on CT scan is less often low than in benign adrenal incidentalomas, probably reflecting the increased steroid synthesis of these adenomas.[68] ^{131}I-iodocholesterol (NP-59) scintigraphy typically shows a unilateral uptake of the tracer, with the normal contralateral adrenal being inactive when the circulating ACTH is suppressed.

These days, most ACAs are diagnosed by investigation of an adrenal incidentaloma. In this situation, specific clinical signs of cortisol excess are absent, in contrast to patients with ACA causing overt Cushing. However, diabetes, obesity, or hypertension is present in 30% to 70% of cases, and the potential role of autonomous cortisol secretion in these comorbidities should be discussed in this setting for patient management. In these incidentally found ACAs, the first diagnostic step is imaging. It is crucial to demonstrate whether a new adrenal mass is benign or malignant. Typically, CT scan shows a homogenous mass smaller than 4 cm and with a low spontaneous density (<10 Hounsfield units [HU]) due to the rich lipid content of the tumor, defining the mass as a benign lipid-rich ACA. The chemical shift observed on MRI reflects the same characteristics. When the spontaneous density on CT scan is not low, determination of washout after media contrast injection is important. A relative washout of greater than 40% and an absolute washout of greater than 60% suggests a benign ACA. Autonomous cortisol secretion in a patient with a fortuitously discovered ACA will be screened for by a systematic 1-mg overnight test of dexamethasone suppression of serum cortisol. A postdexamethasone cortisol level of less than or equal to 50 nmol/L (1.8 ug/dL) excludes cortisol dysregulation. It has been accepted that a postdexamethasone serum cortisol level between 51 and 138 nmol/L suggests possible autonomous cortisol secretion, and a postdexamethasone cortisol level greater than 138 nmol/L is evidence of autonomous cortisol secretion (Fig 97.6). In this situation, further hormonal investigations with UFC, midnight salivary cortisol level assay, and ACTH level assay will be done to better assess cortisol excess. In most cases of possible autonomous cortisol secretion (i.e., serum cortisol levels after 1-mg overnight dexamethasone of between 51 and 138 nmol/L), additional hormonal investigations will be normal or show only mild alterations. In cases of mild cortisol alterations, careful analysis of patients' comorbidities is important to guide management. These comorbidities can include hypertension, glucose intolerance or diabetes mellitus, obesity, osteoporosis, dyslipidemia, and psychological patterns.

Adrenocortical Cancer

ACC can be diagnosed by the investigation of endocrine signs of steroid excess or after investigations of an adrenal incidentaloma, and less frequently based on symptoms due to tumor mass or metastasis. Most tumors demonstrate steroid oversecretion, often of both glucocorticoids and androgens.[69] ACC represents on an average 1% of adrenal

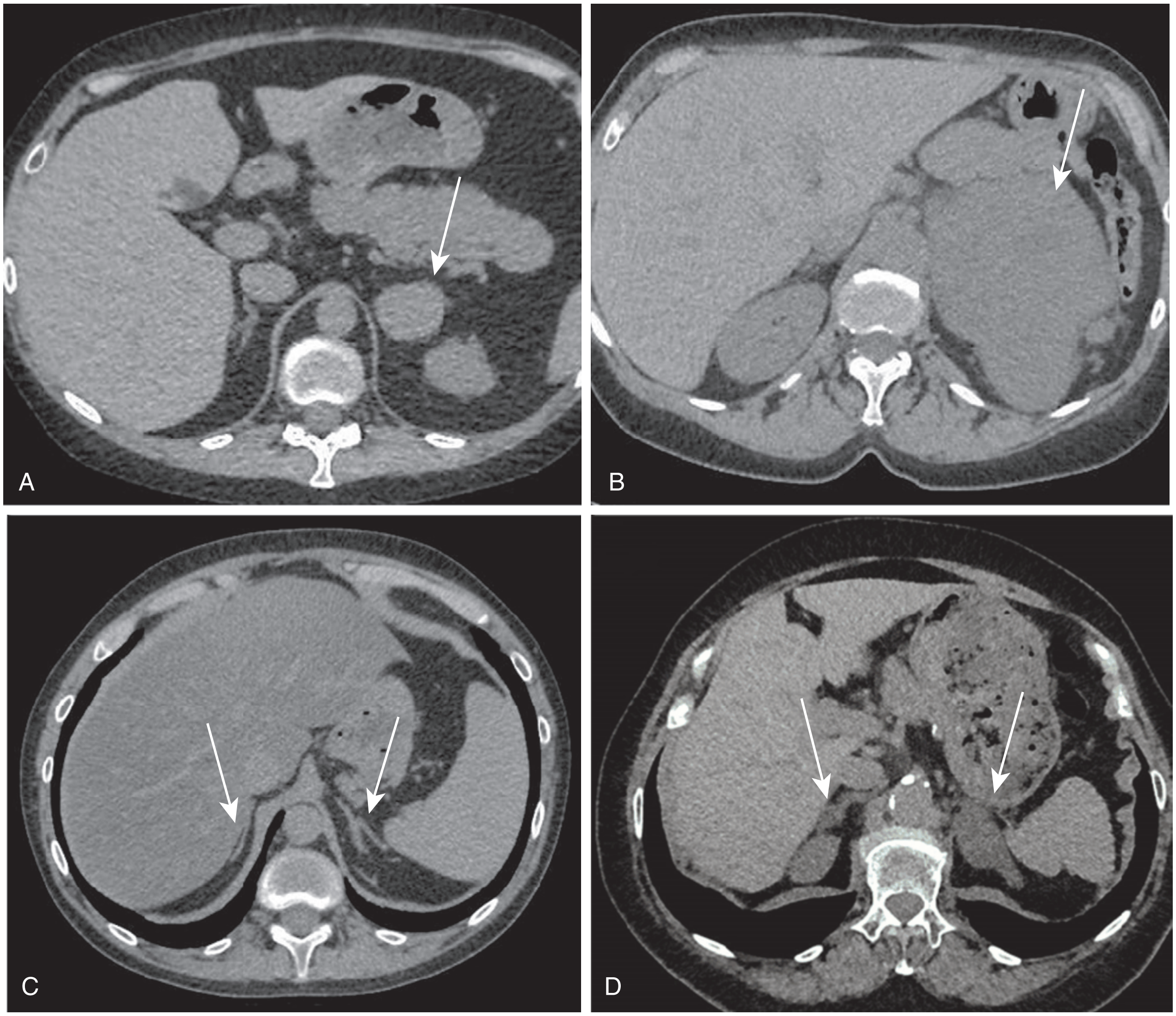


Figure 97.5 Computed tomography scan of the adrenal tumors observed in adrenal Cushing syndrome. A, An adrenocortical adenoma responsible for overt Cushing. **B,** An adrenal cancer responsible for androgen and cortisol excess. **C,** Primary pigmented nodular adrenocortical disease responsible for overt Cushing. **D,** Incidentally discovered asymmetric (L>R) primary macronodular adrenal hyperplasia responsible for moderate autonomous cortisol excess (MACE). The *arrows* point to the tumors.

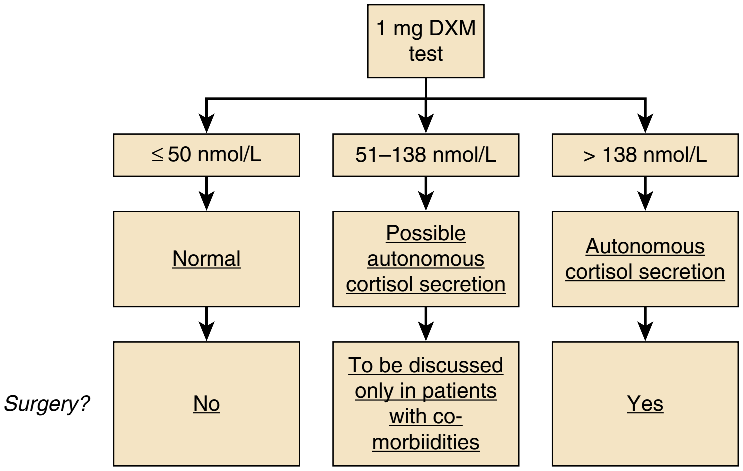

1 mg DXM test
≤ 50 nmol/L
51–138 nmol/L
> 138 nmol/L
Normal
Possible autonomous cortisol secretion
Autonomous cortisol secretion
Surgery?
No
To be discussed only in patients with co-morbiidities
Yes

Figure 97.6 Management of patients with incidentaloma according to autonomous cortisol secretion status. The figure summarizes the algorithm for benign incidentaloma management according to the European guideline for the management of adrenal incidentalomas. (From Fassnacht M, Arlt W, Bancos I, et al. Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*. 2016;175:G1–G34.)

incidentalomas in surgical series.[63] Careful investigation of adrenal incidentalomas with rigorous imaging and stringent criteria is important, as this could help diagnose ACC at an earlier stage. Pathologically, ACCs are usually large tumors with nests intermingled with solid and trabecular areas circumscribed by a fibrous capsule.[66] Cells are usually eosinophilic or vacuolated. Immunostaining for SF1 or melan A can be used to determine the adrenocortical nature of the tumor, along with pathology analysis (both macroscopic and microscopic), and this is especially important to avoid misdiagnosing a nonsecreting tumor. Cytopathological alterations are analyzed and quantitated to determine the Weiss score. When the score is above 3, the adrenal tumor is classified as an ACC. A Weiss score of 3 is a suspicious tumor.[65] Despite an overall low 5-year survival rate (below 30% in most series), some tumors will not show recurrence after surgical removal, and others will have a relatively slow metastatic progression. Prognostication of ACC is not easy, but Ki-67 has demonstrated some prognostic value.[70] Genomic analysis has identified subgroups of ACC with different outcomes, and molecular markers have been suggested.[71,72] Among the various clinical parameters that have been shown to impact ACC prognosis, tumor staging has been demonstrated as one of the most important. The ENSAT staging system is the most commonly used system and relies on surgical findings and extended workup. Four different stages are differentiated with this score.[73] Stage 1 and stage 2 tumors are localized to the adrenal cortex and present a maximum diameter below or above 5 cm, respectively. Locally invasive tumors or tumors with regional lymph node metastases are classified as stage 3, whereas stage 4 consists of tumors with distant metastases. The prognosis of stage 1 and 2 tumors is better than that of stage 3 or 4 tumors. When careful hormonal investigations are performed, it appears that the majority of ACCs are secreting tumors. In contrast to benign ACTs (which usually secrete a single class of steroid), ACCs can secrete various types of steroids.[62] Cosecretion of androgens and cortisol is the most frequent case and is highly suggestive of a malignant ACT. In ACC, cortisol oversecretion will induce the same symptoms as in ACA. By contrast, androgen oversecretion, which is a main feature of steroid excess in ACC, may induce various manifestations in women: hirsutism, menstrual abnormalities, infertility, and eventually frank virilization (alopecia, deepening of the voice, clitoral hypertrophy). ACC can also secrete mineralocorticoids and steroid precursors, leading to hypertension, edema, or hypokalemia. Estrogen oversecretion can be observed in rare cases. Estrogen excess is responsible for gynecomastia in males. Routine hormonal investigations therefore aim to characterize the steroid secretory profile of ACC. Diagnosis of steroid excess is useful to establish the adrenocortical origin of the tumor and can be used for follow-up. ACTH-independent cortisol oversecretion will be screened for as described for ACA. Furthermore, hormonal workup will include plasma 17-OH progesterone, the adrenal androgen DHEA, plasma testosterone in females, and estradiol in males. Other steroids including compound S, DOC, delta 4 androstenedione, and estradiol can be overproduced by the tumor. Secretion of aldosterone by ACC is not frequent and can be detected by plasma aldosterone and renin assays.

Radiological imaging is an essential diagnostic step for ACC, especially in cases of incidentaloma. It is important both for the diagnosis of malignancy of an adrenal mass, but also for the extended workup.[73] CT scan shows a unilateral mass that is most often large (greater than 5–6 cm, typically 10 cm or more), lowering the kidney (Fig. 97.5). Apart from the size of the tumor, the features suggestive of malignancy are the lack of homogeneity, with foci of necrosis and irregular margins and a high spontaneous density observed before contrast media injection during CT scan (above 10 HU), indicating a low-fat content, in contrast to what is often found in ACA. When the spontaneous density is above 10 HU, dynamic measurement of contrast-enhanced densities may provide additional information to help distinguish between benign and malignant lesions. The washout after contrast media injection during CT scan is typically below 50% in ACC. CT scan can also help detect local invasion and distant metastases (liver, lung). positron emission tomography (PET) scan with (^{18}F)-fluorodeoxyglucose (FDG) shows high uptake of the tracer in the primary tumor as well as metastasis, and can be informative for both the diagnosis of malignancy and for extended workup.[74]

Bilateral Adrenocortical Tumors

Bilateral ACTs consist of multiple benign nodules observed in both adrenals, with various patterns of internodular adrenocortical tissues. They are often considered to be a form of nodular adrenal hyperplasia. However, from the molecular point of view, activation of specific oncogenic pathways independently in each adrenal nodule has been demonstrated in some cases, and from the mechanistic point of view, the term hyperplasia is somewhat misleading. The bilateral nature is easily explained in cases for which a hereditary genetic defect has been demonstrated.[75]

Bilateral ACTs can be divided in two groups, with clearly different pathology: micronodular adrenal hyperplasia and PMAH.[18,66] The most classic and frequent form of micronodular adrenal hyperplasia is PPNAD. In micronodular adrenal hyperplasia, multiple small nodules less than 10 mm are observed. In macronodular adrenal hyperplasia, by contrast, some large nodules above 1 cm are observed. PMAH is also referred to as PBMAH and was previously called ACTH-independent macronodular adrenal hyperplasia. The latter term is not used anymore, since the demonstration of the role of locally produced ACTH on cortisol dysregulation.

Micronodular Adrenal Hyperplasia and Primary Pigmented Nodular Adrenal Disease

Micronodular adrenal hyperplasias are multiple small micronodules that are disseminated in both adrenals and are responsible for adrenal Cushing. In very young children with overt Cushing some rare cases of non-pigmented nodules have been reported.[76] However, most patients with micronodular adrenal hyperplasia present with PPNAD.[77] This is a rare cause of ACTH-independent hypercortisolism, most often diagnosed in children and young adults. Its name comes from the particular macroscopic aspect of the adrenals, with pigmented micronodules widespread in the cortex of both adrenals. Adrenal glands from patients with PPNAD are usually normal in size, and their weight is normal or moderately increased (4–15 g). PPNAD pathological characteristics are quite unique already from macroscopic examination: the adrenal cortex is dotted with numerous small pigmented nodules. The multiple pigmented micronodules are usually quite small (0.1–0.3 mm), but some can be slightly larger, although they are rarely larger than 1 cm. Some rare unique pigmented macronodules with a somatic β-catenin mutation have been reported. The adjacent adrenocortical tissue is in most cases atrophic. PPNAD is most often observed in patients with CNC, a familial form of multiple neoplasia predisposition syndrome, responsible for several types of tumor, including other endocrine lesions: somatotroph pituitary adenomas, testicular Sertoli cell calcified tumors, and benign or malignant thyroid lesions, but also nonendocrine manifestations, mainly cardiac myxomas, pigmented skin lesions (lentiginosis and naevi), and melanocytic schwannomas.[78,79] CNC is inherited in an autosomal dominant manner. Inactivating germline mutations of the *PRKAR1A* gene, coding for the regulatory subunit R1A of PKA, located at 17q22-24, are found nowadays in more than 80% of the patients with a classic form of the disease. Interestingly, germline *PRKAR1A* mutation can be found in half of patients with PPNAD as the only manifestation of CNC. The prevalence of PPNAD in patients with CNC is approximately 60% and is the most common endocrine feature seen in this disease.[80] PPNAD is more frequent in females (approximately 70% of patients), with a median age at diagnosis of 34 years. It is rare but possible before the age of 4 years, and it is rarely diagnosed after the age of 45 years, even during systematic screening of patients with known CNC. In autopsy series of CNC patients, histological evidence of PPNAD was present in almost all patients.[78] Diagnosis of CS in PPNAD is often difficult, because hypercortisolism can develop progressively over years. In contrast, in some patients a rapid, significant burst of cortisol excess can be observed that may spontaneously regress. Indeed, in some cases of PPNAD clearly cyclic hypercortisolism has been observed. CS due to PPNAD occurring during pregnancy has also been observed. PPNAD can now also be diagnosed by systematic screening in patients with CNC, who are investigated for other clinical manifestations of the complex or after familial screening. The mean time between appearance of the first symptoms and diagnosis was approximately 4 years in historical series. Despite the unusual time course of CS that can be observed in some patients with PPNAD, clinical signs are quite similar to the observations made in patients presenting with other causes of hypercortisolism. Due to the occurrence of CS in children, short stature was observed in 29% of the 72 patients reported by Carney.[81] However, pediatricians should be aware that growth arrest is not a constant finding in children with PPNAD. This could be explained in some cases by cyclic hypercortisolism. It could be postulated that simultaneous growth hormone dysregulation due to somatotroph hyperplasia or adenoma in patients with CNC may also protect at least partly from the consequences of cortisol excess on growth. Osteoporosis is also frequently observed and can constitute a long-term consequence of CS, especially in patients in whom the diagnosis of PPNAD was delayed. Urinary cortisol is increased in most patients at the time of diagnosis of PPNAD, but its level can be highly variable. The circadian rhythm of cortisol is completely abolished. As for others causes of ACTH-independent CS, plasma levels of ACTH are suppressed in patients with PPNAD. Dexamethasone fails to suppress cortisol, even with administration of high doses. Interestingly, a paradoxical rise in urinary cortisol can be observed after dexamethasone treatment, particularly with administration of high doses (8 mg/day).[82] This could be helpful for the diagnosis of PPNAD, especially in patients with normal basal urinary cortisol levels.[83] An increase in urinary cortisol of at least 100% seems quite specific to PPNAD. This stimulatory effect of glucocorticoid can be observed during primary culture of PPNAD tissues *in vitro*. On CT scan, adrenals can appear normal or moderately enlarged (Fig. 97.5), but higher definition (CT slice thickness <5 mm) allows visualization of small, round, hypodense nodules, usually less than 10 mm. However, adrenal macronodules associated with micronodules can be observed in a subset of rare adult patients.[84] Adrenal NP-59 scintigraphy shows bilateral uptake that can be asymmetric in patients with the presence of unilateral macronodule.

Primary Macronodular Adrenal Hyperplasia

PMAH is typically characterized by the presence of several bilateral adrenal macronodules (diameter >10 mm). PMAH has long been called AIMAH, but the discovery of intraadrenal ACTH synthesis responsible for local stimulation of cortisol production prompted a change of the terminology to PBMAH.[85] Because PMAH more rarely can be also asymmetric, and in some rare cases apparently unilateral, the term PMAH is also used by extension to include the apparently unilateral cases, but this is clearly the same disease group. The ectopic ACTH is secreted by PMAH adrenocortical cells localized in the subcapsular region and in hyperplasic nodules. These cells express proconvertase 1, the cleavage enzyme needed to process the ACTH-precursor proopiomelanocortin to mature ACTH. The regulation of this “ectopic” ACTH differs from the regulation of pituitary ACTH secretion.[86] In operated PMAH, macroscopically, the adrenal glands are classically enlarged, harboring multiple yellowish nonpigmented nodules larger than 1 cm. When patients are operated for overt Cushing syndrome, the adrenal weight is usually between 45 to 100 g, equivalent to 10- to 25-fold the normal adrenal weight. Internodular atrophy is typically observed but not constant, and in some cases hyperplasia of the internodular cortex can be observed, which may then not be a diagnostic criterion. This suggests heterogeneity of this disease that is also illustrated by the various genetic origins. Adrenal nodules are usually composed of two cell types, without mitosis: clear cells with large cytoplasm also known as spongiocytes, with 3β-hydroxysteroid dehydrogenase expression, and compact cells with 17-hydroxylase expression. There is no unifying histopathologic feature to distinguish PMAH from other benign adrenal nodules.

Interestingly, in PMAH cortisol secretion dysregulation is almost always associated with illegitimate membrane receptor expression.[87] This abnormal regulation of cortisol was first investigated in detail in PMAH and later appeared to also be possible in some ACAs. In this situation, cortisol secretion is stimulated by a ligand that physiologically does not control cortisol. Various stimulating ligands binding to protein GPCRs have been described, including glucose-dependent insulinotropic peptide (GIP), which is responsible for food-dependent CS, LH/hCG, which is responsible for CS during pregnancy and after menopause, vasopressin, catecholamines, and serotonin 5-HT. Most of these illegitimate GPCRs stimulate adenylyl cyclase. From a clinical point of view, the rhythm of cortisol secretion can be modified by these illegitimate receptors. For instance, in food-dependent CS, fasting plasma cortisol will be low, while postprandial cortisol will be increased. The upright position may lead to plasma cortisol stimulation in patients with illegitimate adrenergic or vasopressin receptor expression.

PMAH is typically diagnosed in patients between 40 and 65 years of age.[88] As for ACA, PMAH is nowadays diagnosed after investigations of clinical signs of cortisol excess or in patients with bilateral adrenal incidentalomas.[89] PMAH is slightly more prevalent in women. The degree of cortisol autonomous secretion is extremely variable as in ACA, and overall there is a tendency for a positive correlation between adrenal nodule size and cortisol dysregulation. When PMAH is diagnosed in patients with adrenal incidentaloma, cortisol dysregulation is usually mild or absent, and the adrenal volume is moderately increased. It should be noted that 5% to 15% of adrenal incidentalomas are bilateral. In most cases, bilateral adrenal masses represent benign bilateral adrenocortical disease and can be bilateral adenomas or macronodular hyperplasia. As for unilateral incidentalomas, careful imaging analysis is important, and each side needs to be rigorously analyzed to confirm the benign characteristics of the nodules on both sides and that the patient do not present a different type of lesions on each side. In bilateral incidentalomas, if surgery is not performed, thereby allowing pathological diagnosis, it might be difficult in very moderate forms to differentiate PMAH from bilateral adenomas, although careful imaging analysis may help when multiples nodules are visible.

Significant cortisol excess in PMAH leads to frequent overt Cushing syndrome leading to weight gain with android obesity, proximal myopathy, buffalo neck, red stretch marks, skin fragility, and easy bruising, as well as hypertension, diabetes, venous thromboembolism, osteopenia, or osteoporosis, and the clinical picture can in some rare cases be severe. Hormonal investigations usually demonstrate marked alterations, with increased 24-hour UFC, midnight plasma or salivary cortisol, and morning plasma cortisol after 1-mg overnight dexamethasone suppression test, as well as suppressed ACTH. Patients diagnosed after investigation of bilateral adrenal incidentaloma do not present many specific signs of cortisol excess, but they frequently suffer from hypertension, diabetes, or obesity. Interestingly, some data suggest that there is 2.5-fold increase in the presence of vertebral fragility fractures in patients with bilateral adrenal incidentaloma than in those with unilateral adrenal adenoma, even after adjusting for the degree of cortisol hypersecretion and bone mineral density.[90] Hormonal investigations in these moderate forms frequently show an insufficient suppression of plasma cortisol during the 1-mg overnight dexamethasone suppression test. Interestingly, this alteration is also more frequent in bilateral than in unilateral adrenal incidentalomas.[91] Plasma ACTH in these moderate forms with possible autonomous cortisol secretion is usually low but not completely suppressed, and UFC is usually not elevated. In addition, in order to demonstrate the presence of aberrant GPCR expression, the cortisol response to different stimuli could be assessed, as described in Table 97.4. These investigations are interesting for research purposes, but could help in some selected cases to identify potential response to medical therapy.

In PMAH patients, CT scan shows enlarged adrenal glands with at least two nodules larger than 1 cm in diameter (Fig. 97.5). This is clearly different from PPNAD, in which adrenal glands typically appear either normal or moderately and diffusely enlarged. PBMAH nodules usually present similarly to unilateral benign lesions, with a low spontaneous density below 10 HU and rapid washout after intravenous iodine contrast injection. Similar to the observed variability in biochemical presentation, the morphological aspect of the adrenal glands is also highly variable, ranging from a single nodule on each adrenal gland to massively enlarged adrenals, up to 10 cm and larger, with numerous nodules. The internodular adrenal cortex can be hyperplastic or hypoplastic. According to different series, *ARMC5*-mutated index case patients have significantly larger adrenals and more numerous nodules than wild-type patients, correlated with the level of cortisol excess. Abdominal MRI also show characteristics of benign lipid-rich lesions. In case of noncharacteristic imaging features and malignancy suspicion, FDG-PET/CT is often used to assess unilateral adrenal masses. At present, it is difficult to know how to interpret FDG-PET/CT in PMAH, as high FDG uptake has been demonstrated in some cases of PMAH.[90] NP-59 scintigraphy can be useful in demonstrating bilateral autonomous activity in patients with low ACTH levels. It could also provide evidence for asymmetric glucocorticoid output, in which case unilateral adrenalectomy would be an option.

TABLE 97.4 Comorbidities to be Investigated in Patients with Mild Autonomous Cortisol Secretion

Hypertension
Glucose intolerance/diabetes mellitus
Obesity
Dyslipidemia
Osteoporosis

MANAGEMENT OF ADRENAL CUSHING SYNDROME

KEY POINTS

- The surgical procedure used depends on the severity of the hypercortisolism and the malignancy potential.
- Deciding on the right approach for mild autonomous cortisol excess can be challenging, and each case should be individually evaluated based on the progression of metabolic syndrome.
- Steroidogenesis inhibitors can be used in patients with a severe condition as preparation before surgery or for nonoperable tumors.

Management will aim to correct cortisol excess regardless of the etiology, and, in the case of a malignant or suspicious tumor, to achieve tumor control. Management is discussed with a multidisciplinary team, and identifying the cause of adrenal Cushing is a major element in the therapeutic choice. Assessment of the level of autonomous cortisol secretion and its consequences is also a key element in this discussion. While there is no debate about the need to treat overt Cushing, there is more debate regarding benign lesions responsible for possible autonomous cortisol secretion and eventually mild adrenal Cushing. In this situation the treatment plan will be decided on an individual basis based mostly on analysis of the potential consequences of cortisol excess. Medical treatment will be based mostly on steroidogenesis inhibitors and specific treatment of the complications of cortisol excess. Surgery by a laparoscopic or conventional approach, depending on the causative tumor, is always discussed and in many patients can be the single treatment approach used. Also, both medical and surgical treatment can cause transient or definitive adrenal insufficiency that will need specific management.

Medical Treatment with Steroidogenesis Inhibitors (Table 97.5)

The medical approach to controlling steroid excess in adrenal Cushing is mostly discussed in patients with a severe condition as preparation

TABLE 97.5 Level of Action, Usual Dose, Advantages, and Side Effects of Oral Steroidogenesis Inhibitors

Drug	Level of Action	Daily Dose Range	Delay of Action	Side Effects and Potential Problems
Ketoconazole	Inhibits steroidogenic enzymes at multiple levels	200–1200 mg orally two to three times per day	Rapid	Rare hepatotoxicity requiring monitoring of liver enzymes (weekly during first month and monthly for the following 6 months); potential for drug–drug interactions
Metyrapone	Inhibits 11β- hydroxylase and aldosterone synthase	500–6000 mg orally three to four times per day	Rapid	Hirsutism, hypertension, hypokalemia, digestive intolerance
Osilodrostat	Inhibits 11β-hydroxylase and aldosterone synthase	1–30 mg orally two times per day	Rapid	Potential for hirsutism, hypertension, hypokalemia; potential for drug–drug interactions, QT interval monitoring (electrocardiogram)
Mitotane	Inhibits steroidogenic enzymes at multiple levels, cytotoxic and cytolytic effects on the adrenal cortex	0.5–9 g orally three times per day	Delayed (weeks/ months)	Digestive intolerance, increased cholesterol and liver enzyme, neurological intolerance (mostly in case of high blood mitotane levels) Teratogenic; Potential for drug–drug interactions

before surgery or for nonoperable tumors (especially metastatic ACC). It could be also discussed in patients with a mild form of the disease as an alternative to surgery or to determine the role of cortisol dysregulation in its potential complications, but at present there is almost no evidence base directing decision-making in this situation. The drug dose needed to control adrenal Cushing is often lower than that needed for Cushing disease, likely because increased ACTH resulting from lowered cortisol does not occur to the same extent as it does in pituitary Cushing. All steroidogenic inhibitors can cause, as a consequence of their efficacy, varying degrees of adrenal insufficiency. Regular monitoring of cortisol levels is important both for efficacy and safety assessment. For drugs like metyrapone or osilodrostat that increase 11-deoxycortisol, cortisol should be measured by tandem mass spectrometry to avoid the possible cross-reactivity between 11-deoxycortisol and cortisol in immunoassays. Patient should be educated to immediately start glucocorticoid substitution therapy when sign of adrenal insufficiency occur. In cases of demonstrated low cortisol levels suggestive of significant adrenal insufficiency, the drug dosage can be decreased, or a systematic glucocorticoid substitutive treatment utilized (“block and replace” scheme).

Ketoconazole. Ketoconazole was initially used as an antimycotic drug, but can also inhibit steroidogenesis. It acts on several steroidogenic enzymes, such as CYP11A1, CYP17, CYP11B1, and CYP11B2.[92] Ketoconazole is licensed as an orphan drug for CS in Europe but is used “off label” in the management of Cushing patients in the United States. It interacts with medications metabolized through or modulating the activity of the CYP3A4 enzyme, and interaction with concomitant treatments should be taken in account. Hepatic toxicity should be monitored, as mild elevation of transaminase can occur, mostly in the first months of treatment. Pending regular monitoring, it is generally asymptomatic and reversible on drug discontinuation. The drug is given twice or thrice daily orally, and the starting dose is usually 400 mg/day, which can be increased to up to 1200 mg/day. More recently, a stereoenantiomer of ketoconazole, levoketoconazole, has been used to treat CS.[93]

Metyrapone. Metyrapone decreases cortisol synthesis by inhibiting 11β-hydroxylase (CYP11B1 and CYP11B2), the last enzyme involved in cortisol biosynthesis. It is licensed for CS in Europe but is used “off label” in the management of Cushing patients in the United States. The half-life is short, requiring multiple (generally three to four) doses per day, taken orally. The starting dose is 500 to 750 mg/day, and the maximum dose is 4000 mg/day.[92] Metyrapone can cause adrenal insufficiency, dizziness, nausea, and dyspepsia, which are often reversible on dose reduction. Taking the medication with food may reduce gastrointestinal symptoms. Accumulation of adrenal steroid precursors with mineralocorticoid activity may occur, but probably less than in Cushing disease, leading to hypertension, edema, or hypokalemia.

Osilodrostat. Osilodrostat decreases cortisol biosynthesis by inhibiting 11β-hydroxylase (CYP11B1 and CYP11B2), similar to metyrapone. Its affinity for the enzyme is higher, and its half-life is longer. Osilodrostat has been recently approved in the United States and in Europe for the treatment of CS. Experience in adrenal Cushing is limited, as most trials have only involved patients with Cushing disease.[94] The starting dose for osilodrostat is 2 mg, twice daily, taken orally, and can be increased up to 30 mg two times a day. Osilodrostat can interact with drugs metabolized by CYP3A4, CYP1A2, CYP2C19, and CYP2D6, and interaction with concomitant treatments should be taken in account. Furthermore, medications that are associated with QT prolongation are not recommended, because minor QT prolongation has been observed.[92]

Mitotane. o,p’-DDD (ortho, para’, dichloro-, diphenyl-, dichloro ethane; mitotane) is approved for the treatment of ACC. It inhibits steroid synthesis by acting on multiple steroidogenic enzymes, including CYP11A1, 3β-HSD, CYP11B1, and CYP11B2. Interestingly, o,p’-DDD also has a cytotoxic effect on adrenocortical cells that is important for its use in ACC. It is usually effective for controlling steroid excess in patients with secreting ACCs. The treatment is given orally, thrice daily, at a dose of 1.5 to 9 g/day. Tolerance and mitotane blood levels need to be carefully monitored to adjust the daily dose. A mitotane blood level of at least 14 mg/L seems to improve the tumor response rate.[95] However, the side effects of mitotane (mainly digestive and neurologic) often limit the ability to reach this suggested optimal level. The daily mitotane dose required to achieve this 14 mg/L level varies from patient to patient. Therefore, close monitoring of mitotane blood levels is very helpful to remain in the narrow range between 14 to 20 mg/L, considered by most authors to be the therapeutic range for mitotane in ACC. Because o,p’-DDD can induce adrenal insufficiency, glucocorticoid substitution is systematic, and mineralocorticoid therapy should be discussed. Because mitotane induces CYP3A4, it can inactivate hydrocortisone substitution therapy, and the substitutive dose is usually much higher than that used in other causes of adrenal insufficiency.[96]

Other Medical Treatments for Adrenal Cushing Syndrome. In rare patients with severe adrenal Cushing that is difficult to control, which is mostly observed in the context of ACC, the glucocorticoid receptor antagonist mifepristone has been used. For patients who are unable to take an oral drug and whose condition is severe, intravenous etomidate has been used.[73]

Patients with CS comorbidities including hypertension or diabetes need specific treatment for their comorbidities along with treatment of cortisol excess. In cases of mild autonomous cortisol secretion, as discussed below, treatment of comorbidities can be a first step before deciding whether or not to formally treat adrenal cortisol excess. In patients with overt CS, even though cortisol control will lead to improvement of comorbidities in most patients, persistence of these comorbidities is also frequent, and therefore long-term monitoring and specific treatment are important to reduce long-term morbidity and mortality.[67]

Surgery

Surgery is the main treatment for unilateral tumors causing CS.[97] For benign adenomas this is achieved by a laparoscopic approach, with a low complication rate.[98] For adrenal cancer it is performed by a conventional transabdominal approach, although for small ACCs laparoscopic surgery by an expert adrenal surgeon may be an option,[99] always keeping in mind a R0 resection is crucial to improve the prognosis of such an aggressive cancer. For bilateral tumors, laparoscopic adrenalectomy is also discussed in overt Cushing. In patients with PPNAD, adrenalectomy is in most cases bilateral, as good control of CS has rarely been reported after unilateral adrenalectomy.[100] In PMAH, despite the disease being bilateral in most cases, the possibility of unilateral adrenalectomy is discussed to avoid permanent adrenal insufficiency.[101] In ACA and PMAH, the therapeutic choice needs to take in account the level of cortisol excess. There is no doubt that patients with overt Cushing need specific treatment to control cortisol excess. However, in patients with possible MACE, which is most frequently diagnosed in patients with ACA or PMAH, the treatment is discussed on an individual basis.[63] When unilateral ACAs are not treated surgically, worsening of autonomous cortisol secretion or tumor growth during follow-up is rare.[102] However, diabetes, obesity, hypertension, and dyslipidemia seem to worsen with time. Most retrospective studies suggest that surgery for ACA causing mild autonomous cortisol secretion can improve diabetes or hypertension.[103] In this situation, the comorbidities that may be a consequence of cortisol excess are important to consider. Patient history, age, and individual factors that could play a role in the comorbidities and the persistence of hormonal alterations are important parameters to take in account in the surgical indication. For PMAH, the consensus is that bilateral adrenalectomy should not be offered to patients with MACE.[63] However, unilateral adrenalectomy could be discussed. Normalization of urinary cortisol and response to dexamethasone suppression test has been reported after unilateral adrenalectomy in PMAH, and corticotrophin deficiency can also be observed, and is in most cases transient.[104] In PMAH patients with overt Cushing, bilateral adrenalectomy can be discussed, and in small series has proven to be efficient in the long term.[105] However, unilateral adrenalectomy in PMAH with overt Cushing has also been reported in small series to improve cortisol dysregulation, at least transiently, with the advantage of avoiding definite and complete adrenal insufficiency. In patients with asymmetric adrenal volume, removal of the largest adrenal has been suggested, and the results appear to be associated with the characteristics of the remaining adrenal mass.[104]

Surgery in adrenal Cushing should take the patient's medical condition into account in order to prevent perioperative complications. Diabetes, hypertension, or hypokalemia will require specific treatment to improve patient's condition before surgery and during the perioperative period.[106] Diabetes and cortisol excess increase the risk of infection and can delay wound healing. *Pneumocystis jiroveci* infections can be observed during treatment of severe hypercortisolism, and prophylactic trimethoprimsulfamethoxazole can be administered.[2] In cases of very severe cortisol excess with major clinical complications, steroidogenesis inhibitors may be given in preparation for surgery. However, in patients with ACC, presurgical administration of steroidogenesis inhibitors should not significantly delay removal of the malignant tumor. Cushing patients are at higher risk of venous thrombosis, therefore perioperative thromboembolic prophylaxis is very important.

Successful adrenal surgery for adrenal CS will almost always induce immediately adrenal insufficiency. This is obvious for bilateral adrenalectomy, but will also occur almost in all cases after unilateral adrenalectomy. In this situation, hypercortisolism suppresses the hypothalamic–pituitary–adrenal axis, and at the time of adrenal tumor removal endogenous cortisol levels become low or undetectable. It is therefore mandatory to start glucocorticoid substitution by intravenous administration of hydrocortisone at the time of surgery, and as soon as the patient can take oral medications, switch to an orally administered stress dose that is tapered down to a physiologic dose before cortisol levels are assessed. Steroid withdrawal symptoms can occur after correction of cortisol excess. Increasing the dose of hydrocortisone may reduce these symptoms; however, this should be avoided in the long term, because glucocorticoid overdosage will limit improvement after Cushing treatment and impair the possibility of recovery from adrenal insufficiency.

PERSPECTIVES

Investigation of patients at high risk (metabolic syndrome, psychiatric disorders, low bone density in young adult patients, and so on) is important to rule out CS. Although the Endocrine Society recommends the first-line screening test, the choice is dependent on expert understanding and on the patient's clinical presentation. Improvements in the sensitivity and specificity of screening tests are fundamental for the diagnosis. Structurally based assays, such as LC-MS/MS, have only recently become available for clinical practice and are probably the most accurate methodologies.

Although the precise sequence of events and the molecular mechanisms underlying the role of PKA pathway in causing adrenal CS are not fully understood, the central role of abnormal PKA activation in cortisol hypersecretion has been well established through both cellular and murine models. Accordingly, genetic alterations affecting key regulatory members of the PKA signaling pathway, such as *PRKACA*, *PRKAR1A,* and the *PDE* genes, are the most frequent mutations identified in CPA and PPNAD. However, the molecular relationship between the inactivating *ARMC5* mutations identified in 40% of cases of PMAH and the more moderate activation of PKA signaling observed in these hyperplasias remains to be elucidated. Altogether, genomic studies and omics approaches have then allowed the identification of genetic causes of more than 50% of cortisol-producing ACTs, but it is important to understand the genetic origin of the remaining tumors. Indeed, studies of the genes affected in these tumors through the characterization of cellular and mouse models containing these alterations have drastically improved our understanding of the physiopathology of ACTs, but also of the control of adrenocortical homeostasis. Finally, based on these data, one important perspective is to develop new treatments for cortisol-producing ACTs that could first be tested in mouse models that best reproduced human adrenocortical disease.

The management of patients with MACE and the benefit of adrenal surgery need to be demonstrated in prospective studies. The potential role of medical treatment in inhibiting cortisol secretion or antagonizing the glucocorticoid receptor need to be investigated in MACE due to unilateral or bilateral ACTs.

The role of MACE in the development of comorbidities need to be better understood and demonstrated, as well as the potential genetic factors that could explain an increased susceptibility to the consequences of cortisol dysregulation. This would help to identify patients who would benefit from treatment of MACE.

Prognostication for adrenal cancer and novel therapies for progressing tumors are urgently needed. The genetic causes of bilateral adrenal tumors remain to be established for most PMAH patients and some PPNAD patients. The development of specific medical treatments targeting genetic defects in signaling pathways would be appealing.

The benefit of unilateral versus bilateral adrenalectomy in PMAH needs to be demonstrated, and the potential role of medical treatments needs to be investigated in clinical trials.

REFERENCES

1. Cushing H. The basophil adenomas of the pituitary body. *Ann R Coll Surg Engl*. 1969;44:180–181.
2. Nieman LK, Biller BMK, Findling JW, et al. Treatment of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2015;100:2807–2831.
3. Wengander S, Trimpou P, Papakokkinou E, et al. The incidence of endogenous Cushing's syndrome in the modern era. *Clin Endocrinol (Oxf)*. 2019;91:263–270.
4. Nieman LK, Biller BMK, Findling JW, et al. The diagnosis of cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1526–1540.
5. Hannah-Shmouni F, Stratakis CA. A gene-based classification of primary adrenocortical hyperplasias. *Horm Metab Res*. 2020;52:133–141.
6. Bonnet-Serrano F, Bertherat J. Genetics of tumors of the adrenal cortex. *Endocr Relat Cancer*. 2018;25:R131–R152.
7. Ahn CH, Kim JH, Park MY, et al. Epidemiology and comorbidity of adrenal Cushing syndrome: a nationwide cohort study. *J Clin Endocrinol Metab*. 2021;106:e1362–e1372.

8. Whalen DH. Subcategorical phonetic mismatches and lexical access. *Percept Psychophys*. 1991;50:351–360.
9. Allolio B, Fassnacht M. Clinical review: adrenocortical carcinoma: clinical update. *J Clin Endocrinol Metab*. 2006;91:2027–2037.
10. Latronico AC, Pinto EM, Domenice S, et al. An inherited mutation outside the highly conserved DNA-binding domain of the p53 tumor suppressor protein in children and adults with sporadic adrenocortical tumors. *J Clin Endocrinol Metab*. 2001;86:4970–4973.
11. Ribeiro RC, Sandrini F, Figueiredo B, et al. An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. *Proc Natl Acad Sci USA*. 2001;98:9330–9335.
12. Sandrini R, Ribeiro RC, DeLacerda L. Childhood adrenocortical tumors. *J Clin Endocrinol Metab*. 1997;82:2027–2031.
13. Almeida MQ, Latronico AC. The molecular pathogenesis of childhood adrenocortical tumors. *Horm Metab Res*. 2007;39:461–466.
14. Carpenter PC. Diagnostic evaluation of Cushing's syndrome. *Endocrinol Metab Clin North Am*. 1988;17:445–472.
15. Hutter AM, Kayhoe DE. Adrenal cortical carcinoma. Clinical features of 138 patients. *Am J Med*. 1966;41:572–580.
16. Luton JP, Cerdas S, Billaud L, et al. Clinical features of adrenocortical carcinoma, prognostic factors, and the effect of mitotane therapy. *N Engl J Med*. 1990;322:1195–1201.
17. Kasperlik-Załuska AA, Migdalska BM, Zgliczyński S, et al. Adrenocortical carcinoma. A clinical study and treatment results of 52 patients. *Cancer*. 1995;75:2587–2591.
18. Bourdeau I, Parisien-La Salle S, Lacroix A. Adrenocortical hyperplasia: a multifaceted disease. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101386.
19. De Venanzi A, Alencar GA, Bourdeau I, et al. Primary bilateral macronodular adrenal hyperplasia. *Curr Opin Endocrinol Diabetes Obes*. 2014;21:177–184.
20. Fragoso MCBV, Alencar GA, Lerario AM, et al. Genetics of primary macronodular adrenal hyperplasia. *J Endocrinol*. 2015;224:R31–R43.
21. Morani AC, Jensen CT, Habra MA, et al. Adrenocortical hyperplasia: a review of clinical presentation and imaging. *Abdom Radiol (NY)*. 2020;45:917–927.
22. Mariani BM, de P, Nishi MY, Wanichi IQ, et al. Allelic variants of ARMC5 in patients with adrenal incidentalomas and in patients with Cushing's syndrome associated with bilateral adrenal nodules. *Front Endocrinol (Lausanne)*. 2020;11:36.
23. Kirk JM, Brain CE, Carson DJ, et al. Cushing's syndrome caused by nodular adrenal hyperplasia in children with McCune-Albright syndrome. *J Pediatr*. 1999;134:789–792.
24. Lieberman SA, Eccleshall TR, Feldman D. ACTH-independent massive bilateral adrenal disease (AIMBAD): a subtype of Cushing's syndrome with major diagnostic and therapeutic implications. *Eur J Endocrinol*. 1994;131:67–73.
25. Stratakis CA, Kirschner LS. Clinical and genetic analysis of primary bilateral adrenal diseases (micro- and macronodular disease) leading to Cushing syndrome. *Horm Metab Res*. 1998;30:456–463.
26. Kamilaris CDC, Faucz FR, Voutetakis A, et al. Carney complex. *Exp Clin Endocrinol Diabetes*. 2019;127:156–164.
27. Rubinstein G, Osswald A, Hoster E, et al. Time to diagnosis in Cushing's syndrome: a meta-analysis based on 5367 patients. *J Clin Endocrinol Metab*. 2020;105:e12–e22.
28. Szarek E, Stratakis CA. Phosphodiesterases and adrenal Cushing in mice and humans. *Horm Metab Res*. 2014;46:863–868.
29. Berthon A, Bertherat J. Update of genetic and molecular causes of adrenocortical hyperplasias causing Cushing syndrome. *Horm Metab Res*. 2020;52:598–606.
30. Lecoq AL, Stratakis CA, Viengchareun S, et al. Adrenal GIPR expression and chromosome 19q13 microduplications in GIP-dependent Cushing's syndrome *JCI Insight*. 2017;2(18):e92184.
31. Assié G, Libé R, Espiard S, et al. ARMC5 mutations in macronodular adrenal hyperplasia with Cushing's syndrome. *N Engl J Med*. 2013;369:2105–2114.
32. Stratakis CA, Berthon A. Molecular mechanisms of ARMC5 mutations in adrenal pathophysiology. *Curr Opin Endocr Metab Res*. 2019;8:104–111.
33. Horvath A, Bertherat J, Groussin L, et al. Mutations and polymorphisms in the gene encoding regulatory subunit type 1-alpha of protein kinase A (PRKAR1A): an update. *Hum Mutat*. 2010;31:369–379.
34. Beuschlein F, Fassnacht M, Assié G, et al. Constitutive activation of PKA catalytic subunit in adrenal Cushing's syndrome. *N Engl J Med*. 2014;370:1019–1028.
35. Carney JA, Lyssikatos C, Lodish MB, et al. Germline PRKACA amplification leads to Cushing syndrome caused by 3 adrenocortical pathologic phenotypes. *Hum Pathol*. 2015;46:40–49.
36. Dalmazi GD, Beuschlein F. PRKACA mutations in adrenal adenomas: genotype/phenotype correlations. *Horm Metab Res*. 2017;49:301–306.
37. Weigand I, Ronchi CL, Vanselow JT, et al. PKA Cα subunit mutation triggers caspase-dependent RIIβ subunit degradation via Ser114 phosphorylation. *Sci Adv*. 2021;7(8):eabd4176.
38. Vincent-Dejean C, Cazabat L, Groussin L, et al. Identification of a clinically homogenous subgroup of benign cortisol-secreting adrenocortical tumors characterized by alterations of the protein kinase A (PKA) subunits and high PKA activity. *Eur J Endocrinol*. 2008;158:829–839.
39. Bathon K, Weigand I, Vanselow JT, et al. Alterations in protein kinase a substrate specificity as a potential cause of Cushing syndrome. *Endocrinology*. 2019;160:447–459.
40. Espiard S, Knape MJ, Bathon K, et al. Activating PRKACB somatic mutation in cortisol-producing adenomas. *JCI Insight*. 2018;3(8):e98296.
41. Ronchi CL, Di Dalmazi G, Faillot S, et al. Genetic landscape of sporadic unilateral adrenocortical adenomas without PRKACA p.Leu206Arg mutation. *J Clin Endocrinol Metab*. 2016;101:3526–3538.
42. Fragoso MCBV, Domenice S, Latronico AC, et al. Cushing's syndrome secondary to adrenocorticotropin-independent macronodular adrenocortical hyperplasia due to activating mutations of GNAS1 gene. *J Clin Endocrinol Metab*. 2003;88:2147–2151.
43. Little DW, Dumontet T, LaPensee CR, et al. β-catenin in adrenal zonation and disease. *Mol Cell Endocrinol*. 2021;522:111120.
44. Ragazzon B, Libé R, Gaujoux S, et al. Transcriptome analysis reveals that p53 and {beta}-catenin alterations occur in a group of aggressive adrenocortical cancers. *Cancer Res*. 2010;70:8276–8281.
45. Beuschlein F, Galac S, Wilson DB. Animal models of adrenocortical tumorigenesis. *Mol Cell Endocrinol*. 2012;351:78–86.
46. Basham KJ, Hung HA, Lerario AM, et al. Mouse models of adrenocortical tumors. *Mol Cell Endocrinol*. 2016;421:82–97.
47. Leccia F, Batisse-Lignier M, Sahut-Barnola I, et al. Mouse models recapitulating human adrenocortical tumors: what is lacking? *Front Endocrinol (Lausanne)*. 2016;7:93.
48. Sahut-Barnola I, de Joussineau C, Val P, et al. Cushing's syndrome and fetal features resurgence in adrenal cortex-specific Prkar1a knockout mice. *PLoS Genet*. 2010;6:e1000980.
49. Grabek A, Dolfi B, Klein B, et al. The adult adrenal cortex undergoes rapid tissue renewal in a sex-specific manner. *Cell Stem Cell*. 2019;25:290–296.
50. Dumontet T, Sahut-Barnola I, Septier A, et al. PKA signaling drives reticularis differentiation and sexually dimorphic adrenal cortex renewal. *JCI Insight*. 2018;33(2):e98394.
51. Machado MC, Fragoso MCBV, Moreira AC, et al. A review of cushing's disease treatment by the department of neuroendocrinology of the Brazilian society of Endocrinology and metabolism. *Arch Endocrinol Metab*. 2018;62:87–105.
52. Elenkov IJ. Neurohormonal-cytokine interactions: implications for inflammation, common human diseases and well-being. *Neurochem Int*. 2008;52:40–51.
53. Esteban NV, Loughlin T, Yergey AL, et al. Daily cortisol production rate in man determined by stable isotope dilution/mass spectrometry. *J Clin Endocrinol Metab*. 1991;72:39–45.
54. Nestler JE, McClanahan MA. Diabetes and adrenal disease. *Baillieres Clin Endocrinol Metab*. 1992;6:829–847.
55. Lindsay JR, Jonklaas J, Oldfield EH, et al. Cushing's syndrome during pregnancy: personal experience and review of the literature. *J Clin Endocrinol Metab*. 2005;90:3077–3083.
56. Lanna CMM, Paula FJA, Montenegro RM, et al. Parathyroid hormone secretion in chronic human endogenous hypercortisolism. *Braz J Med Biol Res*. 2002;35:229–236.

57. Newell-Price J, Trainer P, Besser M, et al. The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. *Endocr Rev*. 1998;19:647–672.
58. Pecori Giraldi F, Ambrogio AG, De Martin M, et al. Specificity of first-line tests for the diagnosis of Cushing's syndrome: assessment in a large series. *J Clin Endocrinol Metab*. 2007;92:4123–4129.
59. Reimondo G, Pia A, Bovio S, et al. Laboratory differentiation of Cushing's syndrome. *Clin Chim Acta*. 2008;388:5–14.
60. Martin NM, Dhillo WS, Banerjee A, et al. Comparison of the dexamethasone-suppressed corticotropin-releasing hormone test and low-dose dexamethasone suppression test in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 2006;91:2582–2586.
61. Chortis V, Bancos I, Nijman T, et al. Urine steroid metabolomics as a novel tool for detection of recurrent adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2020;105(3):e307–e318.
62. Arlt W, Biehl M, Taylor AE, et al. Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*. 2011;96:3775–3784.
63. Fassnacht M, Arlt W, Bancos I, et al. Management of adrenal incidentalomas: European Society of Endocrinology clinical practice guideline in collaboration with the European Network for the study of adrenal tumors. *Eur J Endocrinol*. 2016;175:G1–G34.
64. Cawood TJ, Hunt PJ, O'Shea D, et al. Recommended evaluation of adrenal incidentalomas is costly, has high false-positive rates and confers a risk of fatal cancer that is similar to the risk of the adrenal lesion becoming malignant; time for a rethink? *Eur J Endocrinol*. 2009;161:513–527.
65. Lau SK, Weiss LM. The Weiss system for evaluating adrenocortical neoplasms: 25 years later. *Hum Pathol*. 2009;40:757–768.
66. Juhlin CC, Bertherat J, Giordano TJ, et al. What did we learn from the molecular biology of adrenal cortical neoplasia? From histopathology to translational genomics. *Endocr Pathol*. 2021;32:102–133.
67. Ahn CH, Kim JH, Park MY, et al. Epidemiology and comorbidity of adrenal Cushing syndrome: a nationwide cohort study. *J Clin Endocrinol Metab*. 2021;106:e1362–e1372.
68. Chambre C, McMurray E, Baudry C, et al. The 10 Hounsfield units unenhanced computed tomography attenuation threshold does not apply to cortisol secreting adrenocortical adenomas. *Eur J Endocrinol*. 2015;173:325–332.
69. Abiven G, Coste J, Groussin L, et al. Clinical and biological features in the prognosis of adrenocortical cancer: poor outcome of cortisol-secreting tumors in a series of 202 consecutive patients. *J Clin Endocrinol Metab*. 2006;91:2650–2655.
70. Beuschlein F, Weigel J, Saeger W, et al. Major prognostic role of Ki67 in localized adrenocortical carcinoma after complete resection. *J Clin Endocrinol Metab*. 2015;100:841–849.
71. Zheng S, Cherniack AD, Dewal N, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016;30:363.
72. Jouinot A, Bertherat J. Management of endocrine disease: adrenocortical carcinoma: differentiating the good from the poor prognosis tumors. *Eur J Endocrinol*. 2018;178:R215–R230.
73. Fassnacht M, Dekkers O, Else T, et al. European society of Endocrinology clinical practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European Network for the study of adrenal tumors. *Eur J Endocrinol*. 2018;179:G1–G46.
74. Groussin L, Bonardel G, Silvéra S, et al. 18F-Fluorodeoxyglucose positron emission tomography for the diagnosis of adrenocortical tumors: a prospective study in 77 operated patients. *J Clin Endocrinol Metab*. 2009;94:1713–1722.
75. Lodish M, Stratakis CA. A genetic and molecular update on adrenocortical causes of Cushing syndrome. *Nat Rev Endocrinol*. 2016;12:255–262.
76. Gunther DF, Bourdeau I, Matyakhina L, et al. Cyclical Cushing syndrome presenting in infancy: an early form of primary pigmented nodular adrenocortical disease, or a new entity? *J Clin Endocrinol Metab*. 2004;89:3173–3182.
77. Shenoy BV, Carpenter PC, Carney JA. Bilateral primary pigmented nodular adrenocortical disease. Rare cause of the Cushing syndrome. *Am J Surg Pathol*. 1984;8:335–344.
78. Carney JA, Gordon H, Carpenter PC, et al. The complex of myxomas, spotty pigmentation, and endocrine overactivity. *Medicine (Baltim)*. 1985;64:270–283.
79. Bouys L, Bertherat J. Management of endocrine disease: carney complex: clinical and genetic update 20 years after the identification of the CNC1 (PRKAR1A) gene. *Eur J Endocrinol*. 2021;184:R99–R109.
80. Bertherat J, Horvath A, Groussin L, et al. Mutations in regulatory subunit type 1A of cyclic adenosine 5'-monophosphate-dependent protein kinase (PRKAR1A): phenotype analysis in 353 patients and 80 different genotypes. *J Clin Endocrinol Metab*. 2009;94:2085–2091.
81. Carney JA, Young WFJ. Primary pigmented nodular adrenocortical disease and its associated conditions. *Endocrinololgist*. 1992;2:6.
82. Stratakis CA, Sarlis N, Kirschner LS, et al. Paradoxical response to dexamethasone in the diagnosis of primary pigmented nodular adrenocortical disease. *Ann Intern Med*. 1999;131:585–591.
83. Bourdeau I, Lacroix A, Schürch W, et al. Primary pigmented nodular adrenocortical disease: paradoxical responses of cortisol secretion to dexamethasone occur in vitro and are associated with increased expression of the glucocorticoid receptor. *J Clin Endocrinol Metab*. 2003;88:3931–3937.
84. Vezzosi D, Tenenbaum F, Cazabat L, et al. Hormonal, radiological, NP-59 scintigraphy, and pathological correlations in patients with Cushing's syndrome due to primary pigmented nodular adrenocortical disease (PPNAD). *J Clin Endocrinol Metab*. 2015;100:4332–4338.
85. Lacroix A. Heredity and cortisol regulation in bilateral macronodular adrenal hyperplasia. *N Engl J Med*. 2013;369:2147–2149.
86. Louiset E, Duparc C, Young J, et al. Intraadrenal corticotropin in bilateral macronodular adrenal hyperplasia. *N Engl J Med*. 2013;369:2115–2125.
87. El Ghorayeb N, Bourdeau I, Lacroix A. Multiple aberrant hormone receptors in Cushing's syndrome. *Eur J Endocrinol*. 2015;173:M45–M60.
88. Lacroix A. ACTH-independent macronodular adrenal hyperplasia. *Best Pract Res Clin Endocrinol Metab*. 2009;23:245–259.
89. Bourdeau I, El Ghorayeb N, Gagnon N, et al. Management of endocrine disease: differential diagnosis, investigation and therapy of bilateral adrenal incidentalomas. *Eur J Endocrinol*. 2018;179:R57–R67.
90. Bouys L, Chiodini I, Arlt W, et al. Update on primary bilateral macronodular adrenal hyperplasia (PBMAH). *Endocrine*. 2021;71(3):595–603.
91. Paschou SA, Kandaraki E, Dimitropoulou F, et al. Subclinical Cushing's syndrome in patients with bilateral compared to unilateral adrenal incidentalomas: a systematic review and meta-analysis. *Endocrine*. 2016;51:225–235.
92. Tritos NA. Adrenally directed medical therapies for Cushing syndrome. *J Clin Endocrinol Metab*. 2021;106:16–25.
93. Creemers SG, Feelders RA, de Jong FH, et al. Levoketoconazole, the 2S,4R enantiomer of ketoconazole, a new steroidogenesis inhibitor for Cushing's syndrome treatment. *J Clin Endocrinol Metab*. 2021;106:e1618–e1630.
94. Varlamov EV, Han AJ, Fleseriu M. Updates in adrenal steroidogenesis inhibitors for Cushing's syndrome - a practical guide. *Best Pract Res Clin Endocrinol Metab*. 2021;35(1):101490.
95. Libè R, Fratticci A, Bertherat J. Adrenocortical cancer: pathophysiology and clinical management. *Endocr Relat Cancer*. 2007;14:13–28.
96. Chortis V, Taylor AE, Schneider P, et al. Mitotane therapy in adrenocortical cancer induces CYP3A4 and inhibits 5α-reductase, explaining the need for personalized glucocorticoid and androgen replacement. *J Clin Endocrinol Metab*. 2013;98:161–171.
97. Di Dalmazi G, Reincke M. Adrenal surgery for Cushing's syndrome: an update. *Endocrinol Metab Clin North Am*. 2018;47:385–394.
98. Gaujoux S, Bonnet S, Leconte M, et al. Risk factors for conversion and complications after unilateral laparoscopic adrenalectomy. *Br J Surg*. 2011;98:1392–1399.
99. Gaujoux S, Mihai R, joint working group of ESES and ENSAT. European Society of Endocrine Surgeons (ESES) and European Network for the Study of Adrenal Tumours (ENSAT) Recommendations for the Surgical Management of Adrenocortical Carcinoma. *Br J Surg*. 2017;104:358–376.
100. Sarlis NJ, Chrousos GP, Doppman JL, et al. Primary pigmented nodular adrenocortical disease: reevaluation of a patient with carney complex 27 years after unilateral adrenalectomy. *J Clin Endocrinol Metab*. 1997;82:1274–1278.

101. Debillon E, Velayoudom-Cephise F-L, Salenave S, et al. Unilateral adrenalectomy as a first-line treatment of Cushing's syndrome in patients with primary bilateral macronodular adrenal hyperplasia. *J Clin Endocrinol Metab*. 2015;100:4417–4424.
102. Elhassan YS, Alahdab F, Prete A, et al. Natural history of adrenal incidentalomas with and without mild autonomous cortisol excess: a systematic review and meta-analysis. *Ann Intern Med*. 2019;171:107–116.
103. Bancos I, Alahdab F, Crowley RK, et al. Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and subclinical Cushing's syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175: R283–R295.
104. Sheikh-Ahmad M, Dickstein G, Matter I, et al. Unilateral adrenalectomy for primary bilateral macronodular adrenal hyperplasia: analysis of 71 cases. *Exp Clin Endocrinol Diabetes*. 2020;128:827–834.
105. Osswald A, Quinkler M, Di Dalmazi G, et al. Long-term outcome of primary bilateral macronodular adrenocortical hyperplasia after unilateral adrenalectomy. *J Clin Endocrinol Metab*. 2019;104:2985–2993.
106. Schreiner F, Anand G, Beuschlein F. Perioperative management of endocrine active adrenal tumors. *Exp Clin Endocrinol Diabetes*. 2019;127:137–146.
107. Fassnacht M, Arlt W, Bancos I, et al. Management of adrenal incidentalomas: European Society of Endocrinology clinical practice guideline in collaboration with the European Network for the study of adrenal tumors. *Eur J Endocrinol*. 2016;175:G1–G34.

98

Adrenocortical Carcinoma

Mouhammed Amir Habra, Tobias Else, and Martin Fassnacht

OUTLINE

Epidemiology, 1632
Risk Factors and Genetics of Adrenocortical Carcinoma, 1632
Clinical Presentation, 1632
Diagnosis, 1633
  Hormonal Evaluation, 1633
  Imaging, 1633
  Histopathology, 1634
Staging, 1634
Therapy, 1635
  Surgery, 1635
  Radiation Therapy, 1637
Medical Therapy, 1637
  Mitotane, 1637
  Cytotoxic Chemotherapy, 1638
  Targeted Therapies, 1638
  Immunotherapy, 1639
Follow-Up, 1639

EPIDEMIOLOGY

KEY POINTS

- Adrenocortical carcinoma (ACC) is a rare malignancy that can occur at any age, with a maximum incidence around the fourth to fifth decade.
- In adults, ACC rarely presents as a part of an inherited tumor syndrome, in contrast to children, in whom ACC frequently arises from germline mutations (e.g., *TP53* mutation in Li–Fraumeni syndrome).

Benign adrenal tumors are one of the most common human neoplasias. with a prevalence of greater than 4% in computed tomography (CT) studies.[1] In contrast, adrenocortical carcinoma (ACC) is a rare malignancy, with an incidence of 0.7 to 2 per million population per year.[2,3] An unusually high incidence of ACC has been found in children in southern Brazil (3.4–4.2 per 1 million children vs. an estimated worldwide incidence of 0.3 per 1 million children) and is related to the high prevalence (0.27%) of a specific germline *TP53* pathogenic variant (p.R337H) that is likely to have low penetrance.[4] ACC is more frequent in women than in men (ratio: 1:1.5). The incidence peaks around the fourth and fifth decade, but the tumor can appear at any age.[5]

RISK FACTORS AND GENETICS OF ADRENOCORTICAL CARCINOMA

There are no established environmental risk factors associated with an increased risk of developing ACC. There is a predilection for the female sex, which does not influence any clinical decision-making in terms of prevention, diagnosis, or therapy. Several hereditary tumor syndromes are associated with an increased risk for ACC. However, the absolute risk is still rather low. In Li–Fraumeni syndrome[6] the frequency of ACC is as high as 4%,[7] mostly affecting children younger than 18 years of age. Up to 80% of children with a diagnosis of ACC harbor a *TP53* pathogenic germline variant, and 4% of adult patients with ACC carry a germline *TP53* pathogenic variant.[8,9] Lynch syndrome, also known as hereditary nonpolyposis colon cancer, has been found to be predisposing condition in another 4% of cases of adult ACC.[9] Screening for Lynch syndrome can be conducted by staining for the four Lynch syndrome–associated gene products by immunohistochemistry of the tumor and microsatellite instability testing. Patients with multiple endocrine neoplasia 1 (MEN1) have a higher risk for adrenal tumors, including ACC. ACC cases have been reported in the context of familial adenomatous polyposis, providing molecular evidence of an involvement of WNT/β-catenin signaling in ACC pathogenesis. ACC can also be a part of Beckwith–Wiedemann Syndrome and has been described in cases of Carney complex.[10] Overall, a familial cancer syndrome can be observed in up to 10% of patients with ACC, and therefore a detailed family history can further streamline genetic testing, which should be considered in any ACC patient for *TP53* (Li–Fraumeni syndrome) and, depending on other information from tumor analysis or family history, for *MSH2*, *MSH6*, *MLH1*, and *PMS2*, as well as *MEN1*, *APC*, and *PRKAR1A*. Identifying a predisposing germline variant will further inform treatment (e.g., immunotherapy for Lynch syndrome, caution regarding radiation therapy for Li–Fraumeni syndrome), set the basis for family cascade testing, and define future surveillance for cured patients and at-risk family members.

Tumor testing for somatic variants provides insight into potential therapeutic targets. However, there is only limited evidence to recommend this as a diagnostic procedure. However, future individualization of therapy may be based on individual tumor profiles, e.g., CDK4/6 inhibition (for *CDKN2A*-mutated tumors) or immunotherapy for patients with Lynch syndrome, hypermutated, or microsatellite unstable tumors.

CLINICAL PRESENTATION

KEY POINTS

- Approximately 50%–60% of patients present with signs or symptoms related to overproduction of adrenal steroid hormones.
- Patients with hormonally nonfunctioning adrenocortical carcinoma (ACC) report symptoms of abdominal discomfort or back pain due to the large tumor size (the median diameter of ACC at diagnosis is 11 cm).

In approximately 10% of patients, ACC is detected incidentally during imaging for other purposes.

Most adult patients with ACC (50%–60%) present with signs and symptoms of adrenal steroid excess, and this percentage is much higher in children (>85%).[11,12] Rapidly progressing Cushing syndrome with or without virilization is the most frequent presentation. Androgen-secreting ACCs in women present with hirsutism and virilization with male pattern baldness and oligoamenorrhea of recent onset. Estrogen-secreting tumors in males lead to gynecomastia and testicular atrophy. Rare aldosterone-producing adrenocortical carcinomas present with severe hypertension and profound hypokalemia.[13] However, low serum potassium is more often the result of excessive cortisol production leading to incomplete renal inactivation by 11-β-dehydrogenase type 2 with consecutive apparent mineralocorticoid excess. In addition, ACC patients with elevated aldosterone levels and hypokalemia often have a hyperreninemic state suggested to arise from compression of the renal parenchyma and vasculature by the often large adrenal tumor.[14]

Patients with a nonfunctioning ACC usually present with symptoms of abdominal discomfort (nausea, vomiting, abdominal fullness) or back pain caused by the large tumor mass. Tumor invasion into the inferior vena cava is not uncommon and may lead to complaints suggestive of deep vein thrombosis and lower extremity edema. Extension to the level of the right atrium may induce shortness of breath and be paralleled by pulmonary tumor emboli. Due to more frequent and improved abdominal imaging, an increasing percentage of ACC is discovered incidentally. Nonspecific symptoms such as fever, weight loss, and loss of appetite are rare.[15] In fact, patients may carry a large tumor burden without much evidence of systemic disease besides signs and symptoms of hormone excess.

KEY POINTS

- Comprehensive assessment of hormonal activity is mandatory in suspected adrenocortical carcinoma (ACC) for optimal perioperative management and for monitoring during follow-up.
- Cross-sectional imaging using computed tomography (unenhanced and delayed postcontrast imaging) or magnetic resonance imaging (with chemical shift) are frequently able to provide hints to the diagnosis of ACC prior to surgery.
- (^{18}F)-fluorodeoxyglucose positron emission tomography can provide valuable information in patients with indeterminate conventional imaging results or when soft tissue or bone metastases are suspected.
- Weiss score is the most commonly used diagnostic criterion for ACC and must be reported along the other prognostic factors, including resection margin and Ki67 index, to guide further treatment.
- Determination of tumor stage provides the basis for further treatment and carries major prognostic information.

DIAGNOSIS

Hormonal Evaluation

Detailed hormonal evaluation must be performed prior to surgery for ACC to identify tumor markers for long-term follow-up and to guide perioperative treatment strategies (e.g., replacement of glucocorticoids after removal of a cortisol-secreting ACC). Guidance for hormonal evaluation in suspected or established ACC has been provided by the European Society for Endocrinology and the European Network for the Study of Adrenal Tumours (Table 98.1). High testosterone levels in women or high estradiol in men, or cosecretion of glucocorticoids and sex hormones, is indicative of a malignant adrenal lesion. In addition, benign cortisol-producing adrenocortical tumors very often show low dehydroepiandrosterone sulfate (DHEAS) concentrations (as a result of suppressed adrenocorticotropic hormone drive), whereas highly elevated DHEAS is suggestive of ACC. Similarly, high levels of steroid precursors, particularly 11-deoxycortisol, are often observed in seemingly hormonally-inactive ACCs. In our experience, autonomous hormone secretion can be detected in greater than 80% of adult patients and in an even higher percentage of children with ACC.[11,12] When using liquid chromatography tandem mass spectrometry–based methods in almost all patients with ACC, steroid or its metabolites can be detected in plasma or urine.[16] Therefore, in large, apparently nonfunctional tumors, measurement of urinary or plasma metanephrines is mandatory prior to surgery to exclude a pheochromocytoma.

TABLE 98.1 Diagnostic Workup in Patients With Suspected or Proven Adrenocortical Carcinoma*

Hormonal Workup	
• Glucocorticoid excess	- Dexamethasone suppression test (1 mg, 2300 h) - 24-h urine cortisol - Basal ACTH (plasma)
• Sexual steroids and steroid precursors	- DHEAS (serum) - 17-OH-progesterone (serum) - Androstendione (serum) - Testosterone (serum, only in women) - 17-beta-estradiol (serum, only in men and postmenopausal women) - 11-deoxycortisol
• Mineralocorticoid excess	- Potassium (serum) - Aldosterone/renin ratio (only in patients with arterial hypertension and/or hypokalemia)
• Exclusion of a pheochromocytoma	- Fractionated metanephrines in 24-h urine or free plasma metanephrines
Imaging	- CT or MRI of abdomen and pelvis and CT chest - FDG-PET (optional) - Bone scintigraphy or brain imaging (when skeletal or brain metastases are suspected)

ACTH, Adrenocorticotropic hormone; *DHEAS*, dehydroepiandrosterone sulfate; *CT*, computed tomography; *MRI*, magnetic resonance imaging; *FDG-PET*, (^{18}F)-fluorodeoxyglucose positron emission tomography.
*Recommendations of the ACC Working Group of the European Network for the Study of Adrenal Tumors (ENSAT), May 2005.

Imaging

Imaging plays a key role in the diagnostic workup of patients with suspected adrenal malignancy. Together with a careful endocrine workup, modern cross-sectional imaging is able to correctly diagnose an adrenal mass as ACC prior to surgery in the majority of cases. Size is an obvious criterion, and for an adrenal mass greater than 6 cm, the likelihood of ACC increases dramatically.[17] In the German ACC registry (n = 1540), the median tumor size at diagnosis was 10.9 cm (range 2.3–40 cm). Thus, in many patients, size alone is a strong indicator of malignancy. However, stage I ACC with a diameter of less than 5 cm has the best prognosis, and it is therefore important to identify adrenal malignancy as early as possible. Thus, tumors between 3 and 8 cm represent a major diagnostic challenge. In children younger than 15 years of age, every adrenal mass is suspicious of malignancy, although no formal imaging studies on this population have been performed.

Currently, no single imaging method can unequivocally determine that an adrenal mass is ACC. The imaging criteria with best evidence

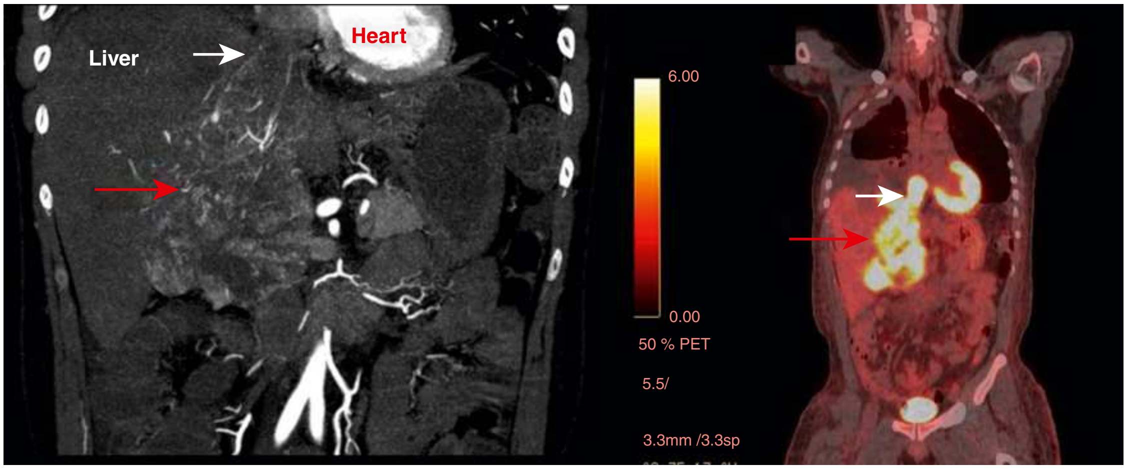


Fig. 98.1 **Coronal images from computed tomography (CT) with intravenous contrast (left) and FDG-PET/CT (right) of a 50-year old male patient with locally advanced right-sided adrenocortical carcinoma *(red arrow)* with extension to the inferior vena cava *(white arrow)* and right atrium.**

is using a threshold of less than or equal to 10 Hounsfield units (HU) in unenhanced CT for the diagnosis of a benign adrenal lesions (for review, see[18,19]).

However, lipid-poor benign adenomas are frequent and show unenhanced HU values greater than 10 [20] and thus are indistinguishable from ACCs. In these cases, delayed postcontrast CT with calculation of the percentage washout for adrenal masses at 10 to 15 minutes or magnetic resonance imaging (MRI) including chemical shift are reasonable "second-line" imaging methods. It is important, however, to recognize that all these imaging characteristics can only be applied to tumors with a homogeneous appearance. Heterogeneous masses and masses that remain indeterminate represent at least 20% of all adrenal tumors (and are even more common amongst larger tumors), necessitating additional diagnostic work-up, such as (^{18}F)-fluorodeoxyglucose positron emission tomography (FDG-PET) imaging.[21,22] In patients with low risk for an ACC, follow-up imaging after 6 to 12 months can be performed.[23]

Multiplanar MRI is particularly suited to separating adrenal masses from surrounding structures like liver, spleen, pancreas, and kidney and often gives a better evaluation for vascular structures, such as adrenal, renal, and cavernous vein. Preoperative CT can be combined with chest imaging for the presence of metastatic disease. Ultrasound techniques (including contrast-enhanced methods) have not been thoroughly investigated but may be particularly valuable for the detection of liver metastases and during follow-up in selected patients. It has been shown that FDG-PET is highly valuable in the differential diagnosis of indeterminate adrenal masses, and while most ACCs show a high FDG signal and most adenomas show a low FDG signal, approximately 10% of tumors in each group show discordance[24,25] (Fig. 98.1). This is particularly important for the positive predictive value, taking into consideration the much higher prevalence of benign adrenal tumors. In addition to 10% of benign adenomas, pheochromocytomas also show uptake of FDG, and in our experience 5% to 10% of primary ACCs are FDG-PET–negative. Imaging with FDG also has the advantage of simultaneously detecting metastases at other sites, but metastatic lesions less than 1 cm (particularly in the lung) can escape detection by FDG,[26,27] indicating that PET cannot substitute for CT imaging.

Fine-needle biopsy as a differential diagnostic tool is rarely helpful, as the diagnostic accuracy is often low.[28,29] In addition, there can be significant complications, as it violates the tumor capsule, which is not in accordance with the usual oncological principle of ACC surgery, which aims to remove tumor in its entirety.

Staging at the time of diagnosis should include high-resolution CT of the chest (in addition to abdomen and pelvis) for early detection of lung metastases. Specific imaging of brain and bones is only needed in case of clinical suspicion. There is some evidence that FDG-PET/CT can improve staging at initial diagnosis or during follow-up.[25-27]

Histopathology

The pathologic diagnosis of ACC can be difficult due to the lack of clear-cut morphologic criteria,[30] and it is recommended to involve a specialized pathologist. In the German ACC registry, the initial histopathologic diagnosis was changed in 13% of cases by the reference pathologist.[31]

Different diagnostic scores have been introduced to diagnose malignancy. The Weiss system[32] is the most widely used and combines nine morphologic parameters related to tumor structure, cytologic features, and tumor invasion. Further morphologic parameters of importance are broad fibrous bands and hemorrhage. An altered reticulin framework is typical for ACC, and together with the necrosis, high mitotic rate, and vascular invasion, it may become a reliable and more reproducible tool for the diagnosis of malignancy than the Weiss score.[33] Careful assessment of the resection status (R0, R1, R2) is also of great importance, as it defines further treatment strategies. For the same reason, violation of the tumor capsule must be reported.

Important information is provided by immunohistochemistry. Here, Ki67 expression carries highly important prognostic information and must be part of the pathological report. In ACC, a Ki67 labeling index of greater than or equal to 20% is associated with poor clinical outcome.[34] Although there are several markers, such as melan A and inhibin-alpha, SF1 has emerged as the single most useful marker to define the adrenocortical origin of the tumor. ACCs are typically negative for chromogranin A and S100.[30,35-37]

STAGING

The most commonly used tumor-node-metastasis classification is the European Network for the Study of Adrenal Tumors (ENSAT) classification (Table 98.2). In this staging system, stage III is defined by tumor infiltration in surrounding tissue or in adjacent organs or tumor thrombus in vena cava/renal vein or positive lymph nodes, whereas stage IV is defined by the presence of distant metastasis.[38] The

ENSAT staging system provides a powerful prognostic tool for predicting both disease-free and disease-specific survival in patients with ACC (Fig. 98.2) and has been confirmed to be superior to the 2004 Union for International Cancer Control classification.[39] In addition to staging, other pathological and clinical factors influence the long-term outcomes and prognosis. In addition to the prognostic role of Ki67%, in patients with stage I, II, or III who had complete surgical resection, cortisol overproduction has been associated with shorter time to recurrence and shorter overall survival compared to non–cortisol–producing ACCs.[40]

GRAS (Grade, Resection status, Age, and Symptoms) parameters have been used to further predict ACC prognosis and outcomes. In patients with resectable ACC (mostly stages I–III), high GRAS score is associated with poor prognosis.[41,42] Similarly, in patients with advanced ACC at diagnosis (stages III and IV), utilizing GRAS score in addition to staging provided a clinically useful tool to predict prognosis.[43]

TABLE 98.2 Staging System for Adrenocortical Carcinoma: ENSAT 2008

Stage	ENSAT 2008
I	T1, N0, M0
II	T2, N0, M0
III	T1–2, N1, M0
	T3–4, N0–1, M0
IV	T1–4, N0–1, M1

ENSAT, European Network for the Study of Adrenal Tumor; *T1,* tumor ≤5 cm; *T2,* tumor >5 cm; *T3,* tumor infiltration of surrounding tissue; *T4,* tumor invasion of adjacent organs, venous tumor thrombus in vena cava/renal vein; *N0,* no positive lymph nodes; *N1,* positive lymph node(s); *M0,* no distant metastases; *M1*, presence of distant metastasis.

From Fassnacht M, Johanssen S, Quinkler M, et al. Limited prognostic value of the 2004 International Union Against Cancer staging classification for adrenocortical carcinoma: proposal for a revised TNM Classification. *Cancer*. 2009;115:243–250.

KEY POINTS

- Surgery for adrenocortical carcinoma (ACC) should be performed by an expert surgeon aiming at an R0 resection in both primary and recurrent ACC.
- Medical treatment with mitotane may be useful both as adjuvant therapy and in advanced disease. It requires drug monitoring (target concentration ≥14 mg/L) and adequate steroid replacement.
- Radiation therapy may be used as an adjuvant measure in selected patients with positive margins after surgery or as a palliative treatment option in select patients with advanced ACC.
- First-line cytotoxic therapy in adults with advanced ACC consists of etoposide, doxorubicin, and cisplatin plus mitotane. No standard protocol is established in children.
- Targeted therapies for ACC have not yet shown activity superior to cytotoxic chemotherapy. Similarly, checkpoint inhibitors have yielded heterogeneous results and are only successful in a subset of patients.

THERAPY (FIG. 98.3)

Surgery

Most patients with ACC are diagnosed at stages I to III (see Fig. 98.2). In these stages, complete tumor removal is the single most important therapeutic measure and offers the best chance for cure. All available data indicate that surgery should be performed by an experienced surgeon who is aware of the pitfalls of surgery for ACC (violation of the tumor capsule, tumor spillage). An R0 resection (resection margins microscopically tumor-free) is of utmost importance to improve long-term prognosis. To achieve this goal, sometimes surgery needs to be extensive, with *en bloc* resection of invaded organs. Tumor spillage during surgery is a key factor in upgrading to stage IV and poor prognosis.[38] The presence of a tumor thrombus is compatible with complete tumor resection but occasionally necessitates a cardiac bypass technique.[44] Several reports have analyzed the value of laparoscopic adrenalectomy in ACC. Considering the retrospective nature of the reports and the potential biases, open adrenalectomy is still regarded as standard treatment for ACC, and laparoscopic surgery

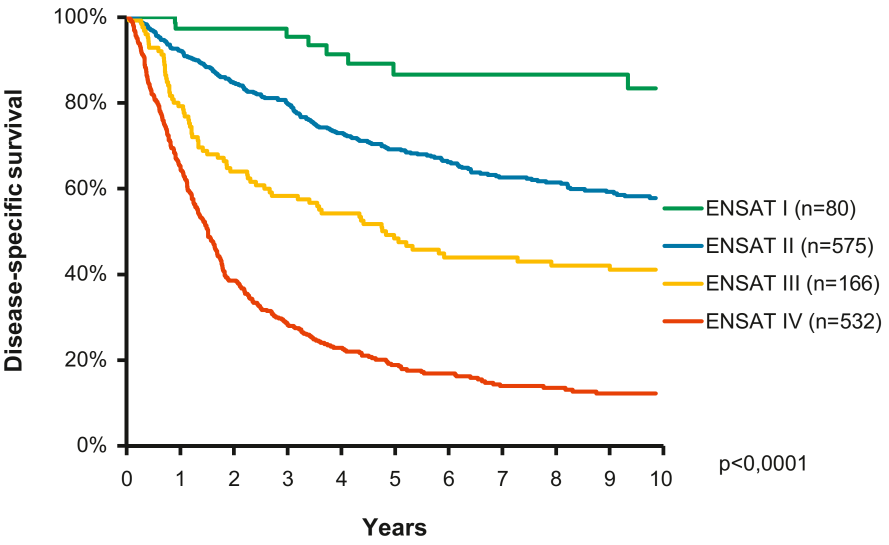


Fig. 98.2 Disease-specific survival according tumor stage (European Network for the Study of Adrenal Tumor classification; see Table 98.2). ((Data derived from the adrenocortical carcinoma center in Würzburg, Germany, March 2021.))

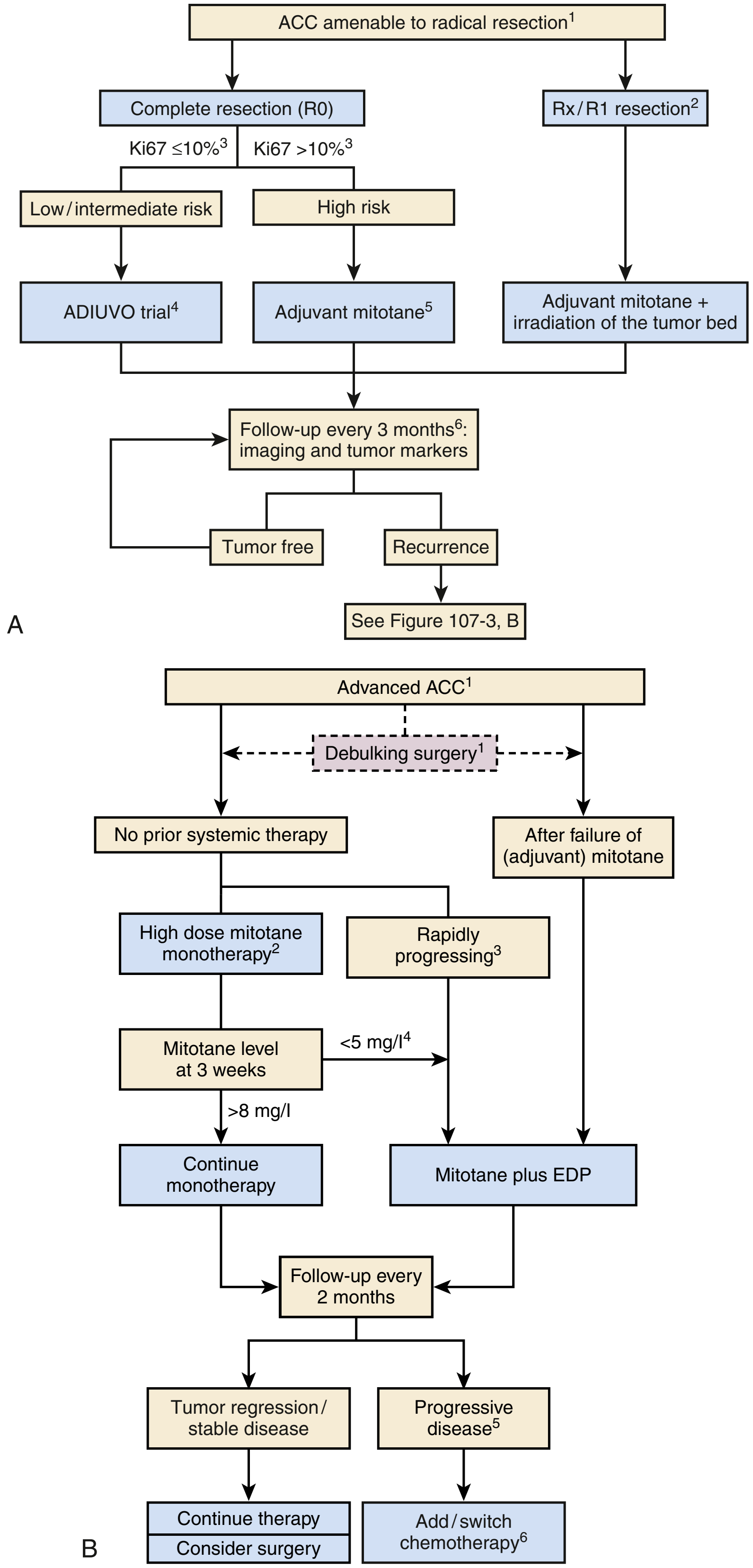

ACC amenable to radical resection^1
Complete resection (R0)
Rx/R1 resection^2
Ki67 ≤10%^{3}
Ki67 >10%^{3}
Low/intermediate risk
High risk
ADIUVO trial^4
Adjuvant mitotane^5
Adjuvant mitotane + irradiation of the tumor bed
Follow-up every 3 months^6: imaging and tumor markers
Tumor free
Recurrence
See Figure 107-3, B
A
Advanced ACC^1
Debulking surgery^1
No prior systemic therapy
After failure of (adjuvant) mitotane
High dose mitotane monotherapy^2
Rapidly progressing^3
Mitotane level at 3 weeks
<5 mg/l^4
>8 mg/l
Continue monotherapy
Mitotane plus EDP
Follow-up every 2 months
Tumor regression/ stable disease
Progressive disease^5
Continue therapy
Consider surgery
Add/switch chemotherapy^6
B

Fig. 98.3 **Treatment flowchart for patients with adrenocortical carcinoma (ACC).** Always consider enrollment of patients in a clinical trial (for an updated list see www.clinicaltrials.gov). **A,** ACC amenable to radical resection. **B,** Advanced ACC. *EDP,* Etoposide, doxorubicin, cisplatin

by an experienced team can be an option when the primary adrenal mass diameter is less than 6 cm without evidence of local invasion or lymphadenopathy.[21,22,45]

The role of localized lymph node dissection (LND) is evolving. The results of a retrospective analysis of 283 patients with stage I to III ACC suggest that locoregional LND can improve both diagnostic accuracy and therapeutic outcome.[46] In multivariate analyses, a significantly reduced risk for tumor recurrence (hazard ratio [HR] 0.65) and disease-related death (HR 0.54) was found in LND patients compared with non-LND patients. In contrast, systematic ipsilateral nephrectomy in the absence of gross local invasion is probably not of benefit.[47]

The role of surgery in recurrent or metastatic disease is ill-defined. Several studies have addressed this issue.[47-52] Based on these studies, surgery for recurrent ACC can be performed in patients with a disease-free interval of preferably more than 12 months, in whom a complete resection is feasible and in the setting of an experienced multidisciplinary team. It is expected that patients with small-volume metastatic disease and adequate performance status will have a survival advantage over patients with unresectable metastases and poor performance status.[53] However, we advocate against surgery if the time between surgery and recurrence is less than 6 months, as these patients often require systemic therapy options. In all other patients, an individualized decision is required, and they may be candidates for a combined approach of systemic therapy and/or ablative procedures (e.g., radiofrequency ablation, chemoembolization).

Radiation Therapy

There is a general assumption that ACC is not a radioresistant malignancy.[54] Over the past two decades, there have been multiple retrospective studies of relatively small number of patients showing conflicting evidence about the use of adjuvant radiation therapy in ACC. In general, adjuvant radiation therapy is viewed as a tool to reduce local recurrence when resection margins are microscopically positive (R1), while the routine use of adjuvant radiation therapy is not recommended when the surgical margins are negative (R0).[21,55] A commonly used regimen that utilizes a fractionation scheme with single doses of 1.8 to 2.0 Gy for 5 days per week for 5 to 6 weeks is recommended. The total dose should not fall below a minimum of 40 Gy, and ideally should reach 50 to 60 Gy.

Palliative radiotherapy in symptomatic metastatic lesions is a well-established treatment option for patients with ACC, in particular for bone metastases. Palliative radiotherapy with limited doses may also be effective in unresectable abdominal recurrences or symptomatic metastases causing pain and vascular or intestinal obstruction. For optimum results of radiotherapy in ACC, an experienced radiotherapist using modern treatment concepts with CT planning, high-voltage radiation, and multiple fields is a prerequisite.

MEDICAL THERAPY

Mitotane

Mitotane (1.1 dichloro-2(o-chlorophenyl)-2-(p-chloro-phenyl) ethane, o,p'-DDD) is an adrenolytic compound with specific adrenolytic activity that was introduced in the treatment of ACC more than 60 years ago. Its precise mechanism of action is not fully understood. Its therapeutic activity depends on metabolic activation in adrenocortical cells. Mitotane is hydroxylated in the mitochondria and transformed into an acyl-chloride. The activated metabolites may directly bind to macromolecules, thereby inhibiting their activity and inducing oxidative damage through production of free radicals.[56] More recently, mitotane adrenolytic activity has been linked to increasing endoplasmic reticulum stress secondary to the inhibition of sterol-O-acyl-transferase 1, leading to the accumulation of toxic intracellular lipids in the adrenal cortex.[57] In addition to its adrenolytic activity, mitotane has an inhibitory effect on steroidogenesis, which adds to its value in the management of patients with hormonally functional ACCs.

Despite the lack of randomized trials, there is evidence that mitotane can induce significant tumor regression in up to 25% of cases with metastatic ACC and achieve control of hypercortisolism in most patients.[58,59] Complete responses have been described, and long-term survival is seen in the occasional patient, but is the exception, not the rule.

In the adjuvant setting, a large retrospective analysis has indicated that mitotane holds great potential to prolong disease-free survival and overall survival in ACC.[60] Nonetheless, it remains a matter of discussion whether adjuvant mitotane is of long-term benefit for all patients.[61] Two randomized studies were launched to explore the role of adjuvant mitotane alone or in combination with chemotherapy in ACC. The ADIUVO trial aims to compare adjuvant mitotane use with a strategy of "watchful waiting" in patients with low or intermediate risk for recurrence (i.e., defined by R0 resection, absence of metastases, and a Ki67 ≤10%). More recently, the ADIUVO-2 study started enrolling high-risk patients, defined as having stage I to III with Ki67 greater than 10%. Patients will be randomized to adjuvant mitotane for 2 years versus adjuvant mitotane for 2 years combined with 3 months of cisplatin and etoposide.

Mitotane is given as 500-mg tablets. Monitoring mitotane blood concentrations is essential for providing optimal efficacy and reducing toxicity. Treatment aims at mitotane concentrations above 14 mg/L for an antitumor response. Higher concentrations (>20–30 mg/dL) can be associated with more adverse effects, whereas a response is less likely at a concentration less than 14 mg/L. Thus, mitotane has a narrow therapeutic window. However, some patients may respond to lower mitotane concentrations. Due to the long half-life of mitotane, target levels are often achieved only after weeks to months of therapy.[62] Treatment is often initiated with 1.5 g/day and rapidly increased to 4.5 to 6 g/day until target concentrations have been reached. While most adverse effects are related to mitotane plasma concentrations, adverse gastrointestinal effects such as diarrhea seem to be more related to the daily dose. During long-term treatment, the dose of mitotane can usually be reduced, and many patients need only 2 to 3 g/day during long-term therapy. Ingestion of mitotane together with a lipid-rich drink or meal may enhance mitotane absorption and decreases gastrointestinal side effects. Due to its adrenolytic activity, mitotane treatment induces adrenal insufficiency. Furthermore, mitotane also increases the metabolic clearance of glucocorticoids by a very strong induction of cytochrome P450 3A4 (CYP3A4) and increases the concentration of cortisol-binding globulin.[63-65] Patients on mitotane often require a higher dose of glucocorticoid replacement (averaging 40–60 mg per day) compared with patients with other forms of adrenal insufficiency (average requirements approximately 15–30 mg per day). Insufficient glucocorticoid replacement enhances mitotane intolerance. Aldosterone secretion is less often affected, as mitotane primarily acts on the zona fasciculata and the zona reticularis. However, aldosterone secretion should be monitored, and replacement with fludrocortisone may become necessary with long-term use of mitotane. In addition, mitotane increases sex hormone–binding globulin[63] and often leads to low free testosterone concentrations in males. With long-term therapy, free thyroid hormone concentrations decline in the presence of low to low-normal thyroid-stimulating hormone, suggesting secondary hypothyroidism. Due to its estrogenic activity, mitotane may facilitate the development of gynecomastia.

TABLE 98.3 Recommended Drug Regimens for Advanced Adrenocortical Carcinoma*

Mitotane Monotherapy

In mitotane-naïve patients in good clinical condition and with presumably less aggressive disease:

- Start with 1.5 g/day and increase daily dose within 4–6 days to 6 g/day.
- Measure mitotane blood level every 3–4 weeks.
- Adapt dosage according to tolerability and blood level (maximum dose 12 g/day; most patients do not tolerate >8 g/day).
- In case of relevant adverse effects, lowering the dosage or even interrupting treatment may be required to allow compliance.

For patients in poor clinical condition or with concomitant cytotoxic drugs, a lower dosage (e.g., 3 g/day) is recommended.

Etoposide, Doxorubicin and Cisplatin (EDP) Plus Mitotane (EDP/M)

Every 28 days:

- day 1 40mg/m^2 D
- day 2 100mg/m^2 E
- day 3+4 100mg/m^2 E + 40mg/m^2 P

Plus oral mitotane aiming at a blood level between 14 and 20 mg/L. However, in most patients we do not administer more than 4–5 g/day of mitotane.

In patients unfit for the EDP-M regimen, P-M may constitute a reasonable alternative.

Second-Line/Salvage Therapies

Gemcitabine plus capecitabine

- 800 mg/m^2 gemcitabine on day 1 and 8 (repeated every 3 weeks)
- 1500 mg capecitabine orally per day in a continuous fashion
- additional administration of mitotane should be consideredStreptozotocin plus Mitotane (Sz/M)
- induction: day 1 to 5: 1g Sz/d
- afterwards 2 g/day Sz every 21 days
- plus oral mitotane aiming at a blood level between 14 and 20 mg/L

*Consider enrollment of patients in clinical trials (www.clinicaltrial.gov). From Fassnacht M, Kroiss M, Allolio B. Update in adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2013;98:4551–4564; Berruti A, Baudin E, Gelderblom H, et al. Adrenal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2012; 23(Suppl 7):vii131–vii38; and Sperone P, Ferrero A, Daffara F, et al. Gemcitabine plus metronomic 5-fluorouracil or capecitabine as a second-/third-line chemotherapy in advanced adrenocortical carcinoma: a multicenter phase II study. *Endocr Relate Cancer*. 2010;17:445–453

Changes in hepatic gamma glutamyl transaminase levels are so frequent that their absence calls into question patient compliance. However, in some cases, serious hepatotoxicity and even liver failure have been observed. Mitotane prolongs bleeding time, and it is advisable to stop mitotane for a minimum of 1 week prior to major surgery. High low-density lipoprotein cholesterol and/or triglyceride concentrations are regularly observed.[66]

Despite the plethora of adverse effects of mitotane, most patients can be managed with acceptable toxicity in the long term.[67] This is particularly important in an adjuvant setting, as treatment must be given for a minimum of 2 years.

The abovementioned effect of mitotane as one of the strongest available inducers of hepatic CYP3A4[64] has a major impact on the care of ACC patients, because mitotane lowers the blood levels of many frequently coadministered drugs (including but not limited to steroids, antihypertensives, and antibiotics)[65] (Table 98.3).

Cytotoxic Chemotherapy

Only two cytotoxic regimens have been investigated in a randomized controlled trial.[68] This trial, FIRM-ACT, included a total of 304 patients and compared etoposide, doxorubicin, cisplatin, and mitotane (EDP-M) with streptozotocin and mitotane (Sz-M), both as first- and second-line cytotoxic treatment. Although overall survival was not significantly different (14.8 months [EDP-M] vs. 12.0 months [Sz-M]; HR 0.79; $P = 0.07$), objective response rate and progression-free survival clearly favored EDP-M (23.2% vs. 9.2%, $P < 0.001$ and 5.0 months vs. 2.1 months; HR 0.55; $P < 0.001$). Most experts now advocate EDP-M as first-line therapy for patients requiring cytotoxic treatment.[69] Because children were not included in the FIRM-ACT trial, there is no proof that EDP-M is also the first treatment of choice in children, and other regimens might be considered (see Fig. 98.3B).[70]

For patients who fail EDP-M chemotherapy, an individualized approach is needed, with a strong preference for receiving therapy through a clinical trial.[21]

Over the past decade, multiple cytotoxic regimens were reported to have limited efficacy in the management of progressive ACC after failing first-line therapy.

The combination of gemcitabine and capecitabine (or metronomic 5-fluorouracil) has been studied as a second-/third-line regimen based on earlier reports showing disease stabilization for at least 6 months in eight of 28 patients (29%) when using this combination.[71] A more recent report of 145 patients assessed retrospectively after having gemcitabine-based chemotherapy (132 in combination with capecitabine) found more limited impact on survival, with partial responses seen in 4.9% and stable disease in 25% of subjects.[72]

In another retrospective report of 28 patients with advanced ACC who were treated with temozolomide (200 mg/m^2/day for 5 consecutive days every 28-day cycle), median progression-free survival was 3.5 months, with six patients (21%) showing evidence of short-lived objective response, though one subject had complete response to therapy.[73]

Other potential salvage therapies might be trofosfamide or thalidomide. However, both have only modest antitumor activity (150 mg/day).[74,75]

Targeted Therapies

The current success with cytotoxic chemotherapy in ACC is not satisfactory, with less than 50% of tumors responding to treatment, and most of them for only a limited time. Small-molecule multikinase inhibitors (MKIs) have transformed the care of multiple malignancies; however, so far, no breakthrough has been accomplished in ACC.

While the experience with MKIs (e.g., sunitinib, sorafenib, dovitinib, axitinib) has been largely disappointing,[76] it is important to consider multiple factors. First, the patients studied in these trials had virtually all received several different cytotoxic treatments prior to their targeted therapy. Thus, targeted therapies were used as salvage treatment in very high-risk cohorts with poor prognoses. Furthermore, most of these patients had received previous or even concomitant mitotane leading to strong induction of CYP3A4, which frequently plays a major role in the degradation of MKIs and other targeted therapies, thereby profoundly reducing levels of these compounds. Accordingly, studies evaluating earlier treatment with targeted therapies in mitotane-naive patients are urgently needed and may generate an entirely different picture.

c-Met signaling appears to be a worthy target, based on the finding of increased expression of c-Met after exposing an ACC cell line *in vitro* to commonly used treatments in ACC (cisplatin, mitotane, radiation). Thus, it is believed that c-Met signaling could be one of the resistance mechanisms in ACC.[77] In a retrospective report of 16 ACC patients who failed prior therapies, the use of cabozantinib resulted in partial response in three, stable disease in five, and progressive disease in eight patients, with a median progress-free survival of approximately 4 months.[78] Two ongoing phase II trials aim to ascertain the efficacy and safety of cabozantinib in ACC (ClinicalTrials.Gov identifiers: NCT03612232 and NCT03370718).

As ACCs express high levels of insulin-like growth factor (IGF)-2, which acts via the IGF-1 receptor, blockade of the IGF-1 receptor has been studied in ACC. A randomized, phase III trial (139 patients) using the small-molecule IGF-1 receptor inhibitor linsitinib (90 on linsitinib and 49 on placebo) was terminated early due to lack of significant improvement in progression-free survival or overall survival, despite a few patients having an objective response to linsitinib.[79] In a phase I study, the combination of an IGF-1 receptor antibody (cixutumumab) with the mTOR inhibitor temsirolimus led to stable disease for at least 6 months in 11 of 26 patients,[80] whereas other studies exploring monotherapy with the mTOR inhibitor everolimus did not show benefit.[81] Using selective radiopharmaceuticals has been explored in ACC. ^{131}I-iodometomidate binds selectively to adrenal CYP11B enzymes as a targeted radionuclide therapy for advanced ACC. ^{131}I-iodometomidate was given on a compassionate use basis to 11 patients with advanced ACC. One patient experienced a partial response lasting for 26 months, and another five patients had stable disease.[82]

Immunotherapy

In the past decade, the use of immune checkpoint inhibitors has transformed cancer care, and many drugs have been approved to treat different malignancies. The experience with immunotherapy in ACC is evolving, and the results of four small clinical trials were recently published. In a phase Ib clinical trial, avelumab (anti–programmed cell death ligand-1 antibody) was studied in 50 ACC patients, out of whom three (6%) had partial response and 21 (42%) had stable disease. It is important to note that half of the study participants were also on mitotane therapy, including two of the patients who had partial response.[83]

In a phase II clinical trial, nivolumab (anti–programmed cell death-1 [PD-1] monoclonal antibody) was used in 10 ACC patients but had minimal efficacy without objective responses, and only two patients showed stable disease.[84]

Pembrolizumab (anti–PD-1 antibody) was studied in two parallel phase II clinical trials and was associated with partial response rates of 14% to 23%, with reasonable tolerance and low levels of severe adverse events.[85,86]

The role of using immunotherapy with mitotane or MKIs is emerging, and so far there are limited retrospective data to suggest the potential for enhancing the activity of immunotherapy when combined with mitotane or anti–vascular endothelial growth factor therapy.[87,88] These early findings are encouraging to propose future clinical trials combining MKIs or mitotane with immune checkpoint inhibitors.

FOLLOW-UP

Close follow-up is particularly important after seemingly complete surgery for ACC stage I to stage III, as surgery for tumor recurrence is an important treatment option. Initially, staging is repeated every 3 months for a minimum of 2 years (CT of abdomen + chest). Even after 2 years without recurrence, there remains a high risk for relapse. Thus, further follow-up is required, but imaging intervals may increase. Steroid metabolomics are currently being investigated for their potential to detect tumor recurrence early and may lead to less frequent imaging and lower radiation exposure. Regular surveillance of the patient should continue for a minimum of 5 years.

REFERENCES

1. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *AJR Am J Roentgenol*. 2008;190:1163–1168.
2. Kebebew E, Reiff E, Duh QY, et al. Extent of disease at presentation and outcome for adrenocortical carcinoma: have we made progress? *World J Surg*. 2006;30:872–878.
3. Kerkhofs TM, Verhoeven RH, Van der Zwan JM, et al. Adrenocortical carcinoma: a population-based study on incidence and survival in The Netherlands since 1993. *Eur J Cancer*. 2013;49:2579–2586.
4. Custodio G, Komechen H, Figueiredo FR, et al. Molecular epidemiology of adrenocortical tumors in southern Brazil. *Mol Cell Endocrinol*. 2012;351:44–51.
5. Fassnacht M, Allolio B. Clinical management of adrenocortical carcinoma. *Best Pract Res Clin Endocrinol Metab*. 2009;23:273–289.
6. Li FP, Fraumeni Jr JF. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med*. 1969;71:747–752.
7. Hisada M, Garber JE, Fung CY, et al. Multiple primary cancers in families with Li-Fraumeni syndrome. *J Natl Cancer Inst*. 1998;90:606–611.
8. Herrmann LJ, Heinze B, Fassnacht M, et al. TP53 germline mutations in adult patients with adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2012;97:E476–E485.
9. Raymond VM, Everett JN, Furtado LV, et al. Adrenocortical carcinoma is a lynch syndrome-associated cancer. *J Clin Oncol*. 2013;31:3012–3018.
10. Maher ER, Reik W. Beckwith-Wiedemann syndrome: imprinting in clusters revisited. *J Clin Invest*. 2000;105:247–252.
11. Michalkiewicz E, Sandrini R, Figueiredo B, et al. Clinical and outcome characteristics of children with adrenocortical tumors: a report from the International Pediatric Adrenocortical Tumor Registry. *J Clin Oncol*. 2004;22:838–845.
12. Hubertus J, Boxberger N, Redlich A, et al. Surgical aspects in the treatment of adrenocortical carcinomas in children: data of the GPOH-MET 97 trial. *Klin Pädiatr*. 2012;224:143–147.
13. Seccia TM, Fassina A, Nussdorfer GG, et al. Aldosterone-producing adrenocortical carcinoma: an unusual cause of Conn's syndrome with an ominous clinical course. *Endocr Relat Cancer*. 2005;12:149–159.
14. Mouat IC, Omata K, McDaniel AS, et al. Somatic mutations in adrenocortical carcinoma with primary aldosteronism or hyperreninemic hyperaldosteronism. *Endocr Relat Cancer*. 2019;26:217–225.
15. Schteingart DE, Giordano TJ, Benitez RS, et al. Overexpression of CXC chemokines by an adrenocortical carcinoma: a novel clinical syndrome. *J Clin Endocrinol Metab*. 2001;86:3968–3974.
16. Bancos I, Taylor AE, Chortis V, et al. Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. *Lancet Diabetes Endocrinol*. 2020;8:773–781.
17. Ross NS, Aron DC. Hormonal evaluation of the patient with an incidentally discovered adrenal mass. *N Engl J Med*. 1990;323:1401–1405.
18. Young Jr WF. Conventional imaging in adrenocortical carcinoma: update and perspectives. *Horm Cancer*. 2011;2:341–347.
19. Blake MA, Cronin CG, Boland GW. Adrenal imaging. *AJR Am J Roentgenol*. 2010;194:1450–1460.
20. Hamrahian AH, Ioachimescu AG, Remer EM, et al. Clinical utility of noncontrast computed tomography attenuation value (hounsfield units) to differentiate adrenal adenomas/hyperplasias from nonadenomas: Cleveland Clinic experience. *J Clin Endocrinol Metab*. 2005;90:871–877.
21. Fassnacht M, Dekkers OM, Else T, et al. European Society of Endocrinology clinical Practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European network for the study of adrenal tumors. *Eur J Endocrinol*. 2018;179:G1–G46.
22. Fassnacht M, Assie G, Baudin E, et al. Adrenocortical carcinomas and malignant phaeochromocytomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31:1476–1490.
23. Fassnacht M, Arlt W, Bancos I, et al. Management of adrenal incidentalomas: European Society of Endocrinology clinical Practice guideline in

collaboration with the European network for the study of adrenal tumors. *Eur J Endocrinol*. 2016;175:G1–G34.

24. Boland GW, Dwamena BA, Jagtiani Sangwaiya M, et al. Characterization of adrenal masses by using FDG PET: a systematic review and meta-analysis of diagnostic test performance. *Radiology*. 2011;259:117–126.
25. Deandreis D, Leboulleux S, Caramella C, et al. FDG PET in the management of patients with adrenal masses and adrenocortical carcinoma. *Horm Cancer*. 2011;2:354–362.
26. Mackie GC, Shulkin BL, Ribeiro RC, et al. Use of [18F]fluorodeoxyglucose positron emission tomography in evaluating locally recurrent and metastatic adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2006;91:2665–2671.
27. Leboulleux S, Dromain C, Bonniaud G, et al. Diagnostic and prognostic value of 18-fluorodeoxyglucose positron emission tomography in adrenocortical carcinoma: a prospective comparison with computed tomography. *J Clin Endocrinol Metab*. 2006;91:920–925.
28. Williams AR, Hammer GD, Else T. Transcutaneous biopsy of adrenocortical carcinoma is rarely helpful in diagnosis, potentially harmful, but does not affect patient outcome. *Eur J Endocrinol*. 2014;170:829–835.
29. Bancos I, Tamhane S, Shah M, et al. Diagnosis of Endocrine Disease: the diagnostic performance of adrenal biopsy: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175:R65–R80.
30. Volante M, Buttigliero C, Greco E, et al. Pathological and molecular features of adrenocortical carcinoma: an update. *J Clin Pathol*. 2008;61:787–793.
31. Johanssen S, Hahner S, Saeger W, et al. Deficits in the management of patients with adrenocortical carcinoma in Germany. *Dtsch Arztebl Int*. 2010;107:885–891.
32. Weiss LM, Medeiros LJ, Vickery Jr AL. Pathologic features of prognostic significance in adrenocortical carcinoma. *Am J Surg Pathol*. 1989;13:202–206.
33. Duregon E, Fassina A, Volante M, et al. The reticulin algorithm for adrenocortical tumor diagnosis: a multicentric validation study on 245 unpublished cases. *Am J Surg Pathol*. 2013;37:1433–1440.
34. Beuschlein F, Weigel J, Saeger W, et al. Major prognostic role of Ki67 in localized adrenocortical carcinoma after complete resection. *J Clin Endocrinol Metab*. 2015;100:841–849.
35. Saeger W, Fassnacht M, Chita R, et al. High diagnostic accuracy of adrenal core biopsy: results of the German and Austrian adrenal network multicenter trial in 220 consecutive patients. *Hum Pathol*. 2003;34:180–186.
36. Saeger W. Histopathological classification of adrenal tumours. *Eur J Clin Invest*. 2000;30:58–62.
37. Sasano H, Suzuki T, Moriya T. Recent advances in histopathology and immunohistochemistry of adrenocortical carcinoma. *Endocr Pathol*. 2006 Winter;17(4):345–354.
38. Fassnacht M, Johanssen S, Quinkler M, et al. Limited prognostic value of the 2004 international union against cancer staging classification for adrenocortical carcinomas. *Cancer*. 2009;115:243–250.
39. Lughezzani G, Sun M, Perrotte P, et al. The European Network for the Study of Adrenal Tumors staging system is prognostically superior to the international union against cancer-staging system: a North American validation. *Eur J Cancer*. 2010;46:713–719.
40. Berruti A, Fassnacht M, Haak H, et al. Prognostic role of overt hypercortisolism in completely operated patients with adrenocortical cancer. *Eur Urol*. 2014;65:832–838.
41. Liang J, Liu Z, Zhou L, et al. The clinical utility of 'GRAS' parameters in stage I-III adrenocortical carcinomas: long-term data from a high-volume institution. *Endocrine*. 2020;67:449–456.
42. Baechle JJ, Marincola Smith P, Solorzano CC, et al. Cumulative GRAS score as a predictor of survival after resection for adrenocortical carcinoma: analysis from the U.S. Adrenocortical Carcinoma Database. *Ann Surg Oncol*. 2021;28(11):6551–6561.
43. Libe R, Borget I, Ronchi CL, et al. Prognostic factors in stage III-IV adrenocortical carcinomas (ACC): a European Network for the Study of Adrenal Tumor (ENSAT) study. *Ann Oncol*. 2015;26:2119–2125.
44. Gaujoux S, Brennan MF. Recommendation for standardized surgical management of primary adrenocortical carcinoma. *Surgery*. 2012;152:123–132.
45. Gaujoux S, Mihai R. European Society of endocrine surgeons (ESES) and European network for the study of adrenal tumours (ENSAT) recommendations for the surgical management of adrenocortical carcinoma. *Br J Surg*. 2017;104:358–376.
46. Reibetanz J, Jurowich C, Erdogan I, et al. Impact of lymphadenectomy on the oncologic outcome of patients with adrenocortical carcinoma. *Ann Surg*. 2012;255:363–369.
47. Gaujoux S, Al-Ahmadie H, Allen PJ, et al. Resection of adrenocortical carcinoma liver metastasis: is it justified? *Ann Surg Oncol*. 2012;19:2643–2651.
48. op den Winkel J, Pfannschmidt J, Muley T, et al. Metastatic adrenocortical carcinoma: results of 56 pulmonary metastasectomies in 24 patients. *Ann Thorac Surg*. 2011;92:1965–1970.
49. Kemp CD, Ripley RT, Mathur A, et al. Pulmonary resection for metastatic adrenocortical carcinoma: the National Cancer Institute experience. *Ann Thorac Surg*. 2011;92:1195–1200.
50. Datrice NM, Langan RC, Ripley RT, et al. Operative management for recurrent and metastatic adrenocortical carcinoma. *J Surg Oncol*. 2012;105:709–713.
51. Erdogan I, Deutschbein T, Jurowich C, et al. The role of surgery in the management of recurrent adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2013;98:181–191.
52. Prendergast KM, Smith PM, Tran TB, et al. Features of synchronous versus metachronous metastasectomy in adrenal cortical carcinoma: analysis from the US adrenocortical carcinoma database. *Surgery*. 2020;167:352–357.
53. Lo WM, Kariya CM, Hernandez JM. Operative management of recurrent and metastatic adrenocortical carcinoma: a systematic review. *Am Surg*. 2019;85:23–28.
54. Polat B, Fassnacht M, Pfreundner L, et al. Radiotherapy in adrenocortical carcinoma. *Cancer*. 2009;115:2816–2823.
55. Nelson DW, Chang SC, Bandera BC, et al. Adjuvant radiation is associated with improved survival for select patients with non-metastatic adrenocortical carcinoma. *Ann Surg Oncol*. 2018;25:2060–2066.
56. Hahner S, Fassnacht M. Mitotane for adrenocortical carcinoma treatment. *Curr Opin Investig Drugs*. 2005;6:386–394.
57. Sbiera S, Leich E, Liebisch G, et al. Mitotane inhibits sterol-o-acyl transferase 1 triggering lipid-mediated endoplasmic reticulum stress and apoptosis in adrenocortical carcinoma cells. *Endocrinology*. 2015;156:3895–3908.
58. Megerle F, Herrmann W, Schloetelburg W, et al. Mitotane monotherapy in patients with advanced adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2018;103:1686–1695.
59. Else T, Kim AC, Sabolch A, et al. Adrenocortical carcinoma. *Endocr Rev*. 2014;35:282–326.
60. Terzolo M, Angeli A, Fassnacht M, et al. Adjuvant mitotane treatment for adrenocortical carcinoma. *N Engl J Med*. 2007;356:2372–2380.
61. Huang H, Fojo T. Adjuvant mitotane for adrenocortical cancer--a recurring controversy. *J Clin Endocrinol Metab*. 2008;93:3730–3732.
62. Kerkhofs TM, Baudin E, Terzolo M, et al. Comparison of two mitotane starting dose regimens in patients with advanced adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2013;98:4759–4767.
63. Nader N, Raverot G, Emptoz-Bonneton A, et al. Mitotane has an estrogenic effect on sex hormone-binding globulin and corticosteroid-binding globulin in humans. *J Clin Endocrinol Metab*. 2006;91:2165–2170.
64. Chortis V, Taylor AE, Schneider P, et al. Mitotane therapy in adrenocortical cancer induces CYP3A4 and inhibits 5alpha-reductase, explaining the need for personalized glucocorticoid and androgen replacement. *J Clin Endocrinol Metab*. 2013;98:161–171.
65. Kroiss M, Quinkler M, Lutz WK, et al. Drug interactions with mitotane by induction of CYP3A4 metabolism in the clinical management of adrenocortical carcinoma. *Clin Endocrinol*. 2011;75:585–591.
66. Shawa H, Deniz F, Bazerbashi H, et al. Mitotane-induced hyperlipidemia: a retrospective cohort study. *Int J Endocrinol*. 2013;2013:624962.
67. Basile V, Puglisi S, Calabrese A, et al. Unwanted hormonal and metabolic effects of postoperative adjuvant mitotane treatment for adrenocortical cancer. *Cancers*. 2020;12(9):2615.
68. Fassnacht M, Terzolo M, Allolio B, et al. Combination chemotherapy in advanced adrenocortical carcinoma. *N Engl J Med*. 2012;366:2189–2197.

69. Berruti A, Baudin E, Gelderblom H, et al. Adrenal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2012;23:131–138. vii.
70. Redlich A, Boxberger N, Strugala D, et al. Systemic treatment of adrenocortical carcinoma in children: data from the German GPOH-MET 97 trial. *Klin Pädiatr*. 2012;224:366–371.
71. Sperone P, Ferrero A, Daffara F, et al. Gemcitabine plus metronomic 5-fluorouracil or capecitabine as a second-/third-line chemotherapy in advanced adrenocortical carcinoma: a multicenter phase II study. *Endocr Relat Cancer*. 2010;17:445–453.
72. Henning JEK, Deutschbein T, Altieri B, et al. Gemcitabine-based chemotherapy in adrenocortical carcinoma: a multicenter study of efficacy and predictive factors. *J Clin Endocrinol Metab*. 2017;102:4323–4332.
73. Cosentini D, Badalamenti G, Grisanti S, et al. Activity and safety of temozolomide in advanced adrenocortical carcinoma patients. *Eur J Endocrinol*. 2019;181:681–689.
74. Kroiss M, Deutschbein T, Schlotelburg W, et al. Salvage treatment of adrenocortical carcinoma with trofosfamide. *Horm Cancer*. 2016;7:211–218.
75. Kroiss M, Deutschbein T, Schlotelburg W, et al. Treatment of refractory adrenocortical carcinoma with thalidomide: analysis of 27 patients from the European Network for the Study of Adrenal Tumours Registry. *Exp Clin Endocrinol Diabetes*. 2019;127:578–584.
76. Altieri B, Ronchi CL, Kroiss M, et al. Next-generation therapies for adrenocortical carcinoma. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101434.
77. Phan LM, Fuentes-Mattei E, Wu W, et al. Hepatocyte growth factor/cMET pathway activation enhances cancer hallmarks in adrenocortical carcinoma. *Cancer Res*. 2015;75:4131–4142.
78. Kroiss M, Megerle F, Kurlbaum M, et al. Objective response and prolonged disease control of advanced adrenocortical carcinoma with cabozantinib. *J Clin Endocrinol Metab*. 2020;105(5):1461–1468.
79. Fassnacht M, Berruti A, Baudin E, et al. Linsitinib (OSI-906) versus placebo for patients with locally advanced or metastatic adrenocortical carcinoma: a double-blind, randomised, phase 3 study. *Lancet Oncol*. 2015;16:426–435.
80. Naing A, Lorusso P, Fu S, et al. Insulin growth factor receptor (IGF-1R) antibody cixutumumab combined with the mTOR inhibitor temsirolimus in patients with metastatic adrenocortical carcinoma. *Br J Cancer*. 2013;108:826–830.
81. De Martino MC, Feelders RA, Pivonello C, et al. The role of mTOR pathway as target for treatment in adrenocortical cancer. *Endocr Connect*. 2019;8:R144–R156.
82. Hahner S, Kreissl MC, Fassnacht M, et al. [131I]iodometomidate for targeted radionuclide therapy of advanced adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2012;97:914–922.
83. Le Tourneau C, Hoimes C, Zarwan C, et al. Avelumab in patients with previously treated metastatic adrenocortical carcinoma: phase 1b results from the JAVELIN solid tumor trial. *J Immunother Cancer*. 2018;6(1):111.
84. Carneiro BA, Konda B, Costa RB, et al. Nivolumab in metastatic adrenocortical carcinoma: results of a phase 2 trial. *J Clin Endocrinol Metab*. 2019;104:6193–6200.
85. Habra MA, Stephen B, Campbell M, et al. Phase II clinical trial of pembrolizumab efficacy and safety in advanced adrenocortical carcinoma. *J Immunother Cancer*. 2019;7:253.
86. Raj N, Zheng Y, Kelly V, et al. PD-1 blockade in advanced adrenocortical carcinoma. *J Clin Oncol*. 2020;38(1):71–80.
87. Head L, Kiseljak-Vassiliades K, Clark TJ, et al. Response to immunotherapy in combination with mitotane in patients with metastatic adrenocortical cancer. *J Endocr Soc*. 2019;3:2295–2304.
88. Bedrose S, Miller KC, Altameemi L, et al. Combined lenvatinib and pembrolizumab as salvage therapy in advanced adrenal cortical carcinoma. *J Immunother Cancer*. 2020;8(2):e001009.

99

Pheochromocytoma

Karel Pacak, Henri J.L.M. Timmers, David Taieb, Jacques W.M. Lenders, and Graeme Eisenhofer

OUTLINE

Introduction and History, 1642
Catecholamine Production, Secretion, and Metabolism, 1643
Clinical Presentation, 1643
  Who to Test According to Presentation?, 1644
  Signs and Symptoms, 1644
  Differential Diagnosis, 1644
Pathology, 1645
Genetics of Pheochromocytoma, 1646
  Multiple Endocrine Neoplasia Syndromes, 1646
  Von Hippel–Lindau Syndrome, 1646
  Neurofibromatosis Type 1, 1648
  Succinate Dehydrogenase Gene Family, 1648
  Less Common Hereditary or Other Causes, 1648
Biochemical Diagnosis of Pheochromocytoma, 1649
  Preanalytics and Analytics, 1649
  Interpretative Considerations, 1649
  Follow-Up Biochemical Testing, 1650
Localization of Pheochromocytoma, 1650
  Anatomic Imaging, 1650
  Differential Diagnosis, 1651
  Functional Imaging, 1651
  Metaiodobenzylguanidine Scintigraphy, 1651
  Positron Emission Tomography, 1652
Management of Pheochromocytoma, 1652
  Medical Therapy and Preparation for Surgery, 1653
  Postoperative Management, 1654
Special Presentations and Therapeutic Problems, 1655
  Multifocal and Metastatic Pheochromocytoma/Paraganglioma, 1655
  Pediatric Pheochromocytoma, 1656
  Pheochromocytoma in Pregnancy, 1657
Future Directions, 1657

INTRODUCTION AND HISTORY

KEY POINTS

- Pheochromocytomas and paragangliomas are dangerous neuroendocrine (chromaffin cell) tumors that can have lethal consequences if overlooked but are typically curable by surgery.
- Pheochromocytomas, which arise from adrenal medullary chromaffin tissue, are more common (80%–85%) than paragangliomas (15%–20%), which arise from extraadrenal paraganglia.

The term pheochromocytoma derives from the Greek words *phaios* (“dusky”) and *chroma* (“color”); it refers to the staining that occurs when the tumors are treated with chromium salts. The first diagnosis of pheochromocytoma was made in 1886 by Fränkel, and the first large series of successful surgical removals with pharmacologic blockade was reported by Kvale and colleagues in 1956. Pheochromocytomas are rare but treacherous catecholamine-producing tumors that, if missed or not properly treated, almost invariably prove fatal.[1] Prompt initial diagnosis therefore is essential for effective treatment, usually by surgical resection. The clinical manifestations of the tumor are diverse and can mimic a variety of conditions, often resulting in erroneous diagnoses.[1,2]

Pheochromocytomas comprise 80% to 85% of all chromaffin cell tumors and typically arise from adrenal medullary chromaffin tissue. In approximately 15% to 20% of cases tumors arise from extraadrenal paraganglia and are commonly referred to as paragangliomas. Paragangliomas are divided into two groups: those that arise from parasympathetic paraganglia of the skull base and neck (e.g., glomus: caroticum, jugulare, tympanicum, and vagale) as well as anterior mediastinum and those that arise from sympathetic-associated chromaffin tissue. Sympathetic paragangliomas arise from sympathetic ganglia in the abdomen (90%), less commonly from the pelvis, and rarely from the posterior mediastinum (2%). Many abdominal paragangliomas arise from large sympathetic ganglia located around the origin of the inferior mesenteric artery called the organ of Zuckerkandl.[2] Pheochromocytoma occurs at any age, but most often in the fourth and fifth decades, and it occurs equally in men and women. In this chapter, the term “pheochromocytoma” is used to collectively refer to the tumors at both adrenal (pheochromocytoma) and extraadrenal locations (sympathetic paraganglioma), but excluding those at head and neck, as dealt with separately.

Most pheochromocytomas arise sporadically, but up to 35% are hereditary.[3,4] In contrast to sporadic pheochromocytomas that are usually solitary, hereditary pheochromocytomas are often multifocal and bilateral.[5] An early autopsy series indicated a prevalence of pheochromocytoma of 1:1000; significant numbers of tumors remained undiagnosed until death, and up to 50% of these may have contributed to patient mortality.[6] Since then there have been numerous advances in genetics, biochemical diagnosis, localization, and surgical approaches to pheochromocytoma. According to a Dutch nationwide study,[7] there has been an increase in annual incidence of pheochromocytoma to eight cases per million, this presumably reflecting improved detection.

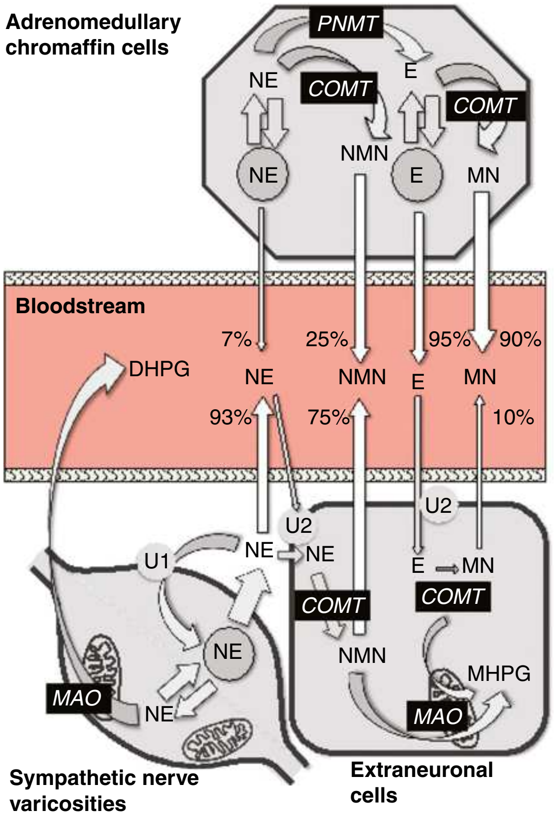


Figure 99.1 Pathways of catecholamine uptake, vesicular storage and leakage, secretion, and metabolism in sympathetic varicosities compared with adrenal medullary chromaffin cells and other extraneuronal cells. *NE*, norepinephrine; *E*, epinephrine; *NMN*, normetanephrine; *MN*, metanephrine; *DHPG*, dihydroxyphenylglycol; *MHPG*, methoxyhydroxyphenylglycol; *PNMT*, phenylethanolamine-N-methyltransferase; *COMT*, catechol-O-methyltransferase; *MAO*, monoamine oxidase; *U1*, uptake 1 (by neuronal transporters); *U2*, uptake 2 (by extraneuronal transporters). Percentages indicate proportions of circulating amines derived from intraadrenal vs. intraneuronal or extraneuronal metabolism. Note: most metabolism occurs initially intraneuronally by deamination of NE leaking from storage vesicles, whereas O-methylation is a relatively minor pathway but important in adrenal chromaffin cells and their tumor derivatives.

CATECHOLAMINE PRODUCTION, SECRETION, AND METABOLISM

KEY POINTS

- The major hormonal products of pheochromocytomas and paragangliomas are the catecholamines, norepinephrine, and epinephrine; these are responsible for most of the clinical signs and symptoms of these tumors.
- Most catecholamines produced by the pheochromocytomas and paragangliomas are metabolized to metanephrines within tumor cells rather than after secretion.

Pheochromocytomas are often described as tumors that secrete catecholamines. The tumors, however, also metabolize catecholamines, and this is a more consistent process than catecholamine secretion (Fig. 99.1).[8] Failure to recognize this key feature probably reflects several misconceptions about the storage and metabolism of catecholamines, perhaps the most pervasive being that metabolism occurs after catecholamine secretion. In fact, substantial amounts of catecholamines are metabolized in the same cells where the amines are synthesized (Fig. 99.1).[9]

Because the enzymes responsible for the metabolism of catecholamines have intracellular locations, the primary mechanism limiting the lifespan of catecholamines in the extracellular space is uptake by active transport, not metabolism by enzymes.[10] Catecholamine metabolism occurs through an array of enzymes, resulting in a variety of metabolites, but in general the most important first step is metabolism within cells of synthesis, and primarily from catecholamines leaking from storage vesicles into the cytoplasm (Fig. 99.1).[9] Most norepinephrine is synthesized within sympathetic neurons; because these neurons contain monoamine oxidase, but not catechol-O-methyltransferase, initial metabolism is by deamination, mainly from leakage of the catecholamine from vesicles and secondly from norepinephrine recaptured after secretion.

In contrast to neurons, chromaffin cells contain catechol-O-methyltransferase so that these cells, unlike neurons, produce methoxytyramine from dopamine, normetanephrine from norepinephrine, and metanephrine from epinephrine (Fig. 99.1).[11] These O-methylated metabolites are produced in small amounts and either only within chromaffin cells after leakage from storage vesicles or in other extraneuronal cells, after uptake following secretion of the catecholamines from sympathetic nerves or the adrenals. The single largest source of O-methylated metabolites represents adrenal chromaffin cells, which account for more than 90% of circulating metanephrine and approximately 25% of circulating normetanephrine (Fig. 99.1).[11] Because these metabolites are similarly produced in chromaffin cell tumors from catecholamines continuously leaking from storage vesicles, a process that is independent of catecholamine secretion, they are superior as diagnostic biomarkers compared with the parent catecholamines.

The final main product of the deamination pathway is vanillylmandelic acid (VMA).[9] Because rates of catecholamine metabolism through the deamination pathway are far larger than for pathways involving extraneuronal and chromaffin cell O-methylation, VMA and other deaminated metabolites are poor biomarkers of catecholamine-producing tumors.

In contrast to norepinephrine and epinephrine, most dopamine in the body is produced in peripheral tissues other than sympathetic nerves and chromaffin cells.[12] Gastrointestinal tissues are a particularly major source of dopamine, large amounts of which are metabolized to sulfate-conjugated derivatives by a specific sulfate-conjugating enzyme confined to these tissues.[13] That same enzyme is also responsible for sulfate conjugation of O-methylated metabolites, which are commonly measured in urine as metanephrines after acid hydrolysis-catalyzed deconjugation; thus, these urinary deconjugated metabolites are largely distinct from the free metanephrines usually measured in plasma. Urinary free dopamine and methoxytyramine are derived from renal uptake and decarboxylation of circulating L-dopa,[14,15] which explains why these metabolites in urine are poor biomarkers of dopamine-producing tumors.

CLINICAL PRESENTATION

KEY POINTS

- The clinical manifestations of a tumor are diverse, variable, and typically non-specific. This can make pheochromocytoma difficult to recognize.
- Signs and symptoms are typically paroxysmal in nature (spells). Spells can occur spontaneously or be triggered by various factors, including certain medications.
- Careful assessment of key signs and symptoms determine the indications for further testing and can assist with interpretation of test results.

TABLE 99.1 Signs and Symptoms of Pheochromocytoma

Symptoms		Signs	
Headaches	++	Hypertension	++++
Palpitations	+++#	Tachycardia or reflex bradycardia	+++=#
Sweating	+++#	Postural hypotension	++
Anxiety/nervousness	++	Hypertension, paroxysmal	++
Tremulousness	++#		
Nausea/emesis	++#	Weight loss	++#
Pain in chest/abdomen	++	Pallor	++#
		Hypermetabolism	++
Weakness/fatigue	+++#	Fasting hyperglycemia	++
Dizziness	+	Tremor	++#
Heat intolerance	+	Increased respiratory rate	++
Paresthesias	+	Decreased gastrointestinal motility	++
Constipation	++#	Psychosis (rare)	very rare
Dyspnea	+	Flushing, paroxysmal (rare)	+
Visual disturbances	+		
Seizures, grand mal	very rare		

[#] indicates a significant difference between patients with and without pheochromocytoma/paraganglioma.

Incidence: ++++, 76%–100%; +++, 51%–75%; ++, 26%–50%; +, 1%–25%.

Adapted from Lenders et al.[1] and Geroula et al.[16]

Who to Test According to Presentation?

The following patients should be evaluated for a pheochromocytoma: 1) anyone with paroxysmal profuse sweating, pallor, tremor, and tachycardia, especially in a subject with episodic hypertension; 2) as part of routine screening for anyone with a known mutation of one of the susceptibility genes and/or a family history of pheochromocytoma or past history of pheochromocytoma; 3) anyone with an incidentally discovered adrenal mass or mass at an extraadrenal location that may represent a paraganglioma; and 4) anyone who has had episodic hypertension, stroke, tachycardia, or an arrhythmia in response to anesthesia, surgery, medications, or food known to precipitate symptoms in patients with pheochromocytoma (Table 99.1). Among ptients evaluated as part of routine screening or because of an incidentaloma, signs and symptoms of catecholamine excess are not always present or, when present, may be overlooked in part due to their nonspecific nature.[16] Nevertheless, for all presentations, there should be careful consideration of the nature of key signs and symptoms to ascertain or pretest likelihood of disease and thereby further interpret results of subsequent biochemical testing.

Signs and Symptoms

The clinical signs and symptoms related to the presence of pheochromocytoma, as listed in the Table 99.1, result from hemodynamic and metabolic actions of excessive circulating catecholamines and rarely of other amines or cosecreted neuropeptides. Most signs and symptoms considered alone are nonspecific; nevertheless, the episodic presence of palpitations, tremor, and sweating in patients with hypertension with or without pallor should arouse immediate suspicion for a pheochromocytoma, because these symptoms and signs are significantly more prevalent among patients with than without pheochromocytoma.[16] Paroxysmal hypertension is one of the most common clinical signs (40%–50%). The extent of the increase and types of catecholamine secreted, as well as the nature of episodic versus continuous secretion, are important determinants of signs and symptoms. Moreover, adrenoceptor downregulation caused by consistently high circulating catecholamines as well as tumor necrosis or hemorrhage may significantly impact pheochromocytoma-associated elevations in blood pressure. Pheochromocytoma also may present with hypotension, particularly in patients with epinephrine-secreting tumors.

Pheochromocytoma-induced attacks (so-called spells) may usually last from a few seconds to 1 hour, with intervals between attacks varying widely; some occur only once every few months. Initially, the episodes may be mild, of short duration, and infrequent. A typical paroxysm is often characterized by a sudden major increase in blood pressure or heart rate associated with a constellation of one or more signs and symptoms: a severe, often pounding headache; profuse sweating over most of the body, especially the trunk; palpitations with tachycardia; anxiety or a sense of doom; skin pallor, especially on the face; nausea, with or without emesis; and pain in the abdomen, the chest, or both. After an episode, patients usually feel drained, exhausted, and weak, and some may urinate more frequently.

Unusual symptoms related to paroxysmal blood pressure elevations during procedures such as endoscopy or administration of anesthesia (particularly during the induction phase of general anesthesia), or associated with ingestion of food or beverages containing tyramine, should arouse immediate suspicion of pheochromocytoma.[17,18] The use of certain drugs such as metoclopramide, droperidol, glucocorticoids, monoamine oxidase inhibitors, tricyclics and other antidepressants, opiates (e.g., morphine, fentanyl), naloxone, glucagon, certain antibiotics (linezolid), drugs for obesity management (phentermine, sibutramine), and chemotherapy may precipitate a hypertensive episode.[17,18]

Causes of episodic catecholamine secretion often remain unclear, but in some situations a surge in catecholamine secretion may be caused by intentional or accidental tumor manipulation with or without an increase in intraabdominal pressure from palpation, defecation, severe stressful situations, or pregnancy. However, approximately 8% to 10% of patients may be completely asymptomatic; this is often the situation for small (<1 cm) tumors associated with insignificant catecholamine production found during screening or as incidentalomas, but is more often associated with tumors that produce and metabolize but do not secrete catecholamines. In such situations, there may be substantial conversion to metanephrines so that the clinical picture is characterized by an asymptomatic presentation associated with increases in plasma or urinary metanephrines but no or little increase in catecholamines. Other situations include tumors, particularly of the skull base and neck, characterized by the absence of catecholamine-synthesizing enzymes and no production or secretion of catecholamines.

It should also be noted that catecholamines suppress intestinal motility; therefore, severe constipation may occur and occasionally may even cause adynamic ileus. Up to 50% of patients may also present with high glucose levels or even with diabetes mellitus type 2 due to inhibition of insulin secretion via α-adrenoceptors, insulin resistance, and increased glucagon secretion.[19]

Differential Diagnosis

The differential diagnosis of pheochromocytoma includes a long list of conditions that present with signs or symptoms that may also suggest the presence of pheochromocytoma (Table 99.2). Some of these conditions can be excluded readily on the basis of a good history and physical examination. The most common mimic is hyperactivation of the sympathoadrenal axis, which is characterized by hypertension, tachycardia, sweating, anxiety, and increased cardiac output. Sympathetic nervous system activation is characteristic of cardiovascular dysfunction due

TABLE 99.2 Differential Diagnosis for Pheochromocytoma Based on Clinical Presentations

Neuroblastoma, ganglioneuroblastoma, ganglioneuroma
Adrenal medullary hyperplasia
Hyperadrenergic essential hypertension
Baroreflex failure
Thyrotoxicosis
Anxiety, panic attacks
Migraine or cluster headaches
Autonomic epilepsy
Chronic use of amphetamines and steroids
Chronic use of cocaine, tricyclic antidepressants
Chronic use of monoamine oxidase inhibitors
Ingestion of tyramine-containing foods or over-the-counter cold and cough preparations
Hypoglycemia, insulin reaction
Paroxysmal tachycardias including postural tachycardia syndrome
Angina pectoris or myocardial infarction
Mitral valve prolapse
Abdominal catastrophic situations/aortic dissection
Renal parenchymal or renal artery diseases (e.g., stenosis)
Intracranial lesions, cerebral vasculitis, and hemorrhage
Menopausal syndrome
Lead poisoning
Toxemia of pregnancy
Pseudopheochromocytoma
Acute intermittent porphyria

TABLE 99.3 Medications that Are Contraindicated in Patients With Suspected or Proven Pheochromocytoma

Dopamine D2 receptor antagonists	(e.g., metoclopramide, antiemetics)
Sympathomimetics	(e.g., ephedrine)
Nonselective β-adrenoceptor blockers	(e.g. propranolol)
Opioid analgesics	
Noradrenaline reuptake inhibitors	(including tricyclics)
Monoamine oxidase inhibitors	
Corticosteroids	(particularly in high doses)
Peptides	(e.g., adrenocorticotropic hormone, glucagon)
Neuromuscular blocking agents	(used for anesthesia)
Serotonin reuptake inhibitors	(rare)

to various pathologies, but can also occur in patients taking amphetamines (Table 99.3). Blockade of norepinephrine reuptake by antidepressants may also be associated with signs and symptoms similar to those observed with sympathetic nervous system activation. Decreased clearance of released catecholamines can thereby lead to increased plasma or urinary metanephrines (usually normetanephrine). The clonidine suppression test can be useful for some but not all of these patients; in particular, the test is inappropriate in patients taking antidepressants or other drugs that block the uptake of released norepinephrine. Other contraindications include patients with compromised cardiac function.

Another problem involves differentiating patients with pheochromocytoma from those with anxiety, panic attacks, or pseudopheochromocytoma. This generally requires a detailed history and examination, together with close observation of the patient during episodes and biochemical testing. Differentiation from an acute myocardial infarction due to obstructive coronary artery disease may be difficult, because angina pectoris and myocardial damage may occur in the absence of obstructive coronary artery disease in a patient with a pheochromocytoma. Many nonspecific electrocardiographic changes have been reported, as well as various supraventricular and ventricular tachycardias, in patients with pheochromocytoma. In the very rare situation in which a strong discordance is noted between symptoms and laboratory findings, one must consider the possibility of factitious administration of catecholamines that can be diagnosed based on an elevated epinephrine without a concomitant increase in its metabolite, metanephrine.

Biochemical test results that are more suggestive of sympathoadrenomedullary activation than of a tumor (e.g., as occurs in hypernoradrenergic hypertension, renovascular hypertension, congestive heart failure, panic disorder, and dumping syndrome) include proportionally larger elevations of plasma norepinephrine or of epinephrine than of plasma normetanephrine or of metanephrine.[20]

Although severe paroxysmal hypertension should always raise suspicion of pheochromocytoma, it may also reflect a clinical entity called *pseudopheochromocytoma.* Pseudopheochromocytoma refers to individuals (often women) with severe paroxysmal hypertension, whether normotensive or hypertensive between episodes, in whom pheochromocytoma has been ruled out and in whom plasma or urinary metanephrines are usually normal.[21] The most common clinical characteristics of this syndrome may be attributable in many cases to short-term activation of the sympathetic nervous system. Paroxysmal hypertension usually is associated with tachycardia, palpitations, nervousness, tremor, weakness, excessive sweating, pounding headache, feeling hot, and facial paleness or (rarely) redness. In contrast to pheochromocytoma, patients with pseudopheochromocytoma more often present with panic attacks or anxiety, flushing, nausea, and polyuria. Pseudopheochromocytoma is sometimes treatable by antihypertensive drugs or psychotherapy.

One of the most difficult situations for both identifying a pheochromocytoma and patient management involves acute medical emergencies, particularly when this requires intensive care support with pressor amines.[22] Such presentations can involve situations that may evolve into hypertensive crisis, multisystem failure, and shock preceded by or associated with cardiomyopathy, pulmonary edema, myocardial infarction, stroke, bowel ischemia, and acute kidney and liver injury. For the majority of cases, which are not caused by pheochromocytoma, there can be profound activation of the sympathoadrenal system that renders biochemical test interpretation difficult. When pheochromocytoma seems possible, it is appropriate to proceed directly to imaging studies. Biochemical testing is best carried out once the patient is stabilized.

PATHOLOGY

KEY POINTS

- In a minority of patients (10%–17%), the tumor may spread (metastasize) to other organs such as the lungs, bones, lymph nodes, or liver.
- There are no reliable pathologic features to predict whether metastases will or will not develop in the future.

Pheochromocytomas are well-circumscribed, encapsulated tumors that exhibit a typical endocrine pattern with chief nests (i.e., Zelballen pattern). The stroma surrounding and separating the nests is mainly composed of fibrovascular tissue. Sustentacular cells may be seen at the periphery of the cell nests. The tumors are characterized by positive chromogranin A staining and sustentacular cells that are positive for the S100 protein.[23,24] Pheochromocytoma is considered intraadrenal in origin if the cortex of the adrenal is found in close relationship with the pheochromocytoma on histopathological review.

Capsular and lymphovascular invasion, coarse nodularity, the presence of atypical nuclei, growth pattern, high proliferative index (Ki67), and the presence of comedonecrosis are often found in malignant pheochromocytoma (last three features having the highest predictive value), but it is impossible to distinguish malignant from benign tumors based on such histopathologic features.[24] Currently, there remains no well-accepted grading system. In the 2017 World Health Organization classification of endocrine tumors, the terminology of benign and malignant was removed, because all pheochromocytomas are now considered to have metastatic potential; thus, the term "malignant" has been replaced by the term "metastatic,"[25,26] now defined by the presence of metastatic lesions in bones and pathologically confirmed in lymph nodes, though lesions in liver or lungs are also useful.

Some patients may present with composite tumors, defined as a tumor that combines morphological features of pheochromocytoma with those of a developmentally related neurogenic tumor, including ganglioneuroma, ganglioneuroblastoma, neuroblastoma, or malignant peripheral sheath tumor.[24] Today, all pheochromocytomas and paragangliomas are covered by a new staging system: T1 = intraadrenal tumor = pheochromocytoma, T2 = extraadrenal tumor = sympathetic paraganglioma (not indicating metastases), T3 = tumor with invasion into surrounding tissue or lymph node metastasis, and T4 = tumor presenting with metastasis to bone or multiple other sites with or without lymph node involvement.[24]

GENETICS OF PHEOCHROMOCYTOMA

KEY POINTS

- Mutations of at least 20 different tumor susceptibility genes account for more than 70% of all pheochromocytomas and paragangliomas.
- An underlying mutation should be considered in all patients, regardless of family history or syndromic features.
- Certain clinical presentations of hereditary pheochromocytoma or paraganglioma depend on the underlying mutation.
- All patients with identified predisposing mutations or hereditary syndromes should receive periodic surveillance for the tumors to which they are susceptible (this also applies to tumors other than pheochromocytoma and paraganglioma).

Pheochromocytoma has the highest degree of heritability among all tumors; up to 40% of cases result from germline mutations.[4,27-29] In a similarly large proportion (30%), somatic mutations have been identified.[4,27,29] Mutations of at least 20 pheochromocytoma susceptibility genes have been identified to date, impacting ten of the most clinically relevant genes (Table 99.4). Pheochromocytoma separates into three main clinically relevant transcriptional cluster groups (pathways): the pseudohypoxia signaling (Cluster 1), kinase signaling (Cluster 2), and c/ Wnt signaling clusters.[4,27] The underlying mutation, as well as its close relationship to a particular cluster, relates to clinical presentation, including metastatic behavior, biochemical phenotype, cluster-specific imaging signature, and to some extent therapeutic response and nature of follow-up.[30-33] Currently, it is recommended that genetic testing should be considered in all patients,[1,34] and it is our recommendation that all patients should be seen by a genetic counselor. Genetic testing is particularly important in patients presenting at a young age, with a family history of pheochromocytoma or with extraadrenal, multiple, recurrent, and metastatic tumors.

Multiple Endocrine Neoplasia Syndromes

Multiple endocrine neoplasia 2 (MEN2) is an autosomal dominant syndrome (Sipple syndrome) affecting approximately one in 40,000 individuals and is characterized by medullary thyroid carcinoma, pheochromocytoma, and parathyroid hyperplasia/adenoma (Table 99.4).[35] The gene responsible for MEN2 is a protooncogene, *RET*, that is expressed in C cells in the thyroid gland and catecholamine-producing chromaffin cells in the adrenal gland. *RET* is located on chromosome 10q11.2 and encodes a receptor tyrosine kinase. As an oncogene, activation of *RET* leads to hyperplasia of target cells *in vivo*. Subsequent secondary events then lead to tumor formation.59 *RET* consists of 21 exons with six so-called "hot-spot exons" (exons 10, 11, 13, 14, 15, and 16) in which mutations are identified in 97% of patients with MEN2.

Pheochromocytomas in MEN 2 develop on a background of adrenomedullary hyperplasia and become manifest (e.g., biochemically or on imaging) in approximately 50% of patients. MEN2-associated pheochromocytomas are almost exclusively benign (with <5% reported to be metastatic) and localized to the adrenals. The peak age is approximately 40 years, but children as young as 10 years can be affected. Patients with MEN2-related pheochromocytoma often lack sustained hypertension or other catecholamine-related symptoms/signs (they occur only in ~50% of patients). MEN2-related pheochromocytomas are almost always intraadrenal, are often multifocal and bilateral (~30% at diagnosis, ~70% lifetime risk). MEN2-related pheochromocytomas produce both epinephrine and norepinephrine; "preferential" stimulation by epinephrine of β-adrenoceptors found on heart or vasculature causes palpitations and tachycardia, respectively.[36] Thus, the detection of MEN2-related pheochromocytoma is invariably based on elevations of plasma or urinary metanephrine, usually with additional elevations of normetanephrine. In addition, as with most epinephrine-producing pheochromocytomas, the hypertension is more likely to be paroxysmal than sustained.[37]

Patients with MEN2B also present with ganglioneuromatosis, multiple mucosal neuromas of eyelids, lips, and tongue, and connective tissue disorders that include marfanoid habitus, scoliosis, kyphosis, pectus excavatum, slipped femoral epiphysis, and pes cavus. This syndrome is caused by germline mutations in the *RET* protooncogene that affect the tyrosine kinase catalytic site of the protein. MEN2B-associated pheochromocytomas are associated with a higher risk of malignancy compared with MEN2A. The screening for MEN-related pheochromocytoma should start between ages of 10 and 16 years.[38]

Von Hippel–Lindau Syndrome

Another syndrome commonly associated with pheochromocytoma is von Hippel–Lindau (VHL), which is caused by mutations of a gene in chromosome 3 (3p25-26) that encodes the *VHL* tumor suppressor gene (Table 99.4).[39] Pheochromocytoma occurs in approximately 15%–20% of patients with *VHL* mutations. Deletions, truncations, or splice-site mutations are associated with low risk, in contrast to missense mutations, which are associated with a high risk of developing pheochromocytoma.[39,40] The mean age at diagnosis is 28 years.

The disease has been divided into two types based on the significant genotype–phenotype correlations observed. Type 1 includes mainly large deletions or disruptive mutations and expresses the full phenotype of vascular lesions of the retina (i.e., angiomas), cysts, or solid tumors in the brain or spinal cord (i.e., hemangioblastomas), pancreatic cysts, renal cell carcinoma (RCC), epididymal cystadenoma, and endolymphatic sac tumors, but low risk for pheochromocytoma. Type

TABLE 99.4 **Pheochromocytoma-Associated Mutations in Clinically Relevant Genes**

Syndrome	Gene	Penetrance	Bilateral/ Multifocal	Metastatic	PHEO or PGL	Biochemical Phenotype	Other Tumors and Important Findings
MEN2[*]	*RET*	50%	50%–80%	Rare	PHEO	EPI	Medullary thyroid cancer, hyperparathyroidism, marphanoid habitus, mucosal ganglioneuromas
VHL[#]	*VHL*	10%–30%	50%	5%–7%	PHEO	NE	Retinal angiomas, CNS hemangioblastomas, clear-cell renal carcinoma, pancreatic cysts, epididymal cystadenoma, endolymphatic sac tumors
NF-1[*]	*NF-1*	1%–2%	16%	5%–10%	PHEO	EPI	Café-au-lait spots, neurofibromas, medullary thyroid cancer, carcinoid tumors, peripheral nerve sheath tumors
PGL1[#]	*SDHD*	90% if paternally transmitted[†]	Inconsistent Reports	15%–20%	PGL; less commonly PHEO	NE and/or DA; some biochemically silent	Clear-cell renal carcinoma, GIST, pituitary adenoma, pulmonary chondroma
PGL2[#]	*SDHAF2/SDHD5*	100% if paternally transmitted	-	-	Head and neck PGL	Unknown	None known
PGL3[#]	*SDHC*	Inconsistent reports	Inconsistent reports	Rare	PGL, less commonly PHEO	NE and/or DA; some biochemically silent	Clear-cell renal carcinoma
PGL4[#]	*SDHB*	30%–50%	Rare	31%–71%	PGL; less commonly PHEO	NE and/or DA; some biochemically silent	Clear-cell renal carcinoma, GIST, pituitary adenoma, breast and thyroid cancer (?), neuroblastoma, pulmonary chondroma
PGL5[#]	*SDHA*	1%–13%	Inconsistent reports	30%–66%	PGL	NE and/or DA; some biochemically silent	Clear-cell renal carcinoma, GIST, pituitary adenoma
-[*]	*TMEM127*	Inconsistent reports	33%	<5%	PHEO	EPI	Clear-cell renal carcinoma
-	*MAX*	Inconsistent reports	67%	<25%	PHEO	Mixed EPI and NE	
[#]	*FH*	Approximately 1%	More reports needed, but usually high	More reports needed, but usually high	PHEO and PGL	NE	Cutaneous and uterine leiomyoma, type 2 papillary RCC
-[#]	*HIF2A (EPAS)*	None, somatic	100%	Over 40%	PHEO/PGL	NE	Duodenal Somatostatinoma, polycythemia
-	*HRAS*	Unknown	Low	Unknown	PHEO; less commonly PGL	EPI	More common in Chinese than Europeans[88]

[#]and * indicate that particularly these hereditary pheochromocytomas belong to cluster 1 or cluster 2, respectively.

CNS, Central nervous system; *GIST*, gastrointestinal stromal tumor; *EPI*, adrenergic; *DA*, dopaminergic; *NE*, noradrenergic; *PHEO*, pheochromocytoma; *PGL*, paraganglioma; *RET*, rearranged during transfection; *VHL*, von Hippel–Lindau; *NF-1*, neurofibromatosis 1; *SDHA–D*, succinate dehydrogenase subunits A, B, C, and D; *SDHAF2*, succinate dehydrogenase assembly factor 2; *TMEM127*, transmembrane protein 127; *MAX*, myc-associated factor X; *FH*, fumarate hydratase; *HIF2A* (*EPAS1*), hypoxia-inducible factor 2α (endothelial PAS domain–containing protein 1); *HRAS*, Harvey rat sarcoma virus oncogene/Hras protooncogene.

2, which is predominantly associated with *VHL* missense mutations, is defined by the occurrence of pheochromocytomas, either alone (type 2C) or in combination with hemangioblastomas (type 2A) or with hemangioblastomas and RCCs (type 2B) (Table 99.4).

Patients are often asymptomatic when they present with other aspects of this disease. This syndrome is variable in terms of the different organ systems involved and the extent of involvement from patient to patient and from family to family. Pheochromocytomas as part of the VHL syndrome have an exclusively noradrenergic phenotype, reflecting the production of norepinephrine.[41] These tumors are located mainly intraadrenally and are bilateral in approximately 50% of patients, with a less than 8% incidence of metastases. In addition, sympathetic paragangliomas are frequent (20%). Tumors are commonly found based on periodic annual screening or during searches for other tumors that are part of this syndrome. Therefore, when detected, these tumors are often small, and approximately 80% of pheochromocytomas found during screening are asymptomatic and are not associated with hypertension. In addition, they may evade detection by functional imaging. The chance of tumor recurrence or development of a second primary tumor resulting from the hereditary nature of VHL supports the need for lifelong surveillance. Screening for VHL-related pheochromocytoma should start around age 4 to 5 years and include measurement of blood pressure and plasma or urine metanephrines.[39,42]

Neurofibromatosis Type 1

Von Recklinghausen neurofibromatosis (NF) is divided into two types: NF-1 includes peripheral neurofibromas of peripheral nerves, whereas NF-2 has central nervous system schwannomas (Table 99.4). NF-1 is inherited in an autosomal dominant pattern. The association with pheochromocytoma is one between a relatively common disease and a rare disease. Current literature suggests that 7% of patients with NF-1 have pheochromocytoma, and approximately 5% of patients with pheochromocytoma have *NF-1* germline mutations.[43] Pheochromocytoma associated with NF-1 is caused by germ line mutations in chromosome 17 (17q11), where the *NF-1* gene encodes for the protein neurofibromin. Pheochromocytoma in patients with NF-1 is rarely seen in children; it usually occurs at a later age (around 50 years). Only 12% of all patients with NF-1 and pheochromocytoma are diagnosed with bilateral and multifocal pheochromocytomas, and less than 6% of patients have metastatic pheochromocytoma.

Previously, routine screening for pheochromocytoma in NF-1 patients was not recommended. Only those patients with hypertension or other symptoms/signs possibly related to catecholamine excess were considered for screening, reflecting the relatively low penetrance of pheochromocytoma in these patients. Continued adherence to this approach ignores improvements in biochemical testing that now allow easier and more efficient diagnosis of pheochromocytoma than formerly possible. A recent study by Kepenekian et al.[44] in 156 patients with NF-1 found that 10 out of 12 patients with pheochromocytoma (6.4% of total patients) were asymptomatic (six biochemically negative and six biochemically positive). Therefore, they recommended performing pheochromocytoma screening every 4 to 5 years in all patients with NF-1 after 40 years of age using both the measurement of metanephrines and anatomical imaging by either computed tomography (CT) or magnetic resonance imaging (MRI). Current guidelines suggest that screening for pheochromocytoma in patients with NF-1 be carried out once every 3 years starting at an age between 10 and 14 years.[45]

Succinate Dehydrogenase Gene Family

Germline mutations of the succinate dehydrogenase (*SDH*) gene family now represent a well-established and important cause of hereditary pheochromocytoma.[29] The *SDH* genes (*SDHA*, *SDHB*, *SDHC*, and *SDHD*) encode the four subunits of complex II of the mitochondrial electron transport chain,[46] which is essential for the generation of adenosine triphosphate (oxidative phosphorylation).

SDHB and *SDHD* mutations can lead to complete loss of SDH enzymatic activity, a phenomenon that has now been well linked to tumorigenesis through upregulation of hypoxic-angiogenetic–responsive genes.[47] The *SDHB*, *SDHC*, and *SDHD* genes are associated with the presence of familial and nonfamilial pheochromocytoma and sympathetic and parasympathetic paragangliomas (Table 99.4).[31,48,49] Frameshift, missense, and nonsense mutations, but also large deletions, were identified for the *SDH* gene family.

Studies have shown that approximately 4% to 12% of apparently sporadic pheochromocytomas and up to 50% of familial pheochromocytomas are associated with *SDH* mutations.[28,50,51] The *SDHB*, *SDHC*, and *SDHD* traits are inherited in an autosomal dominant fashion and give rise to familial paraganglioma syndromes 4, 3, and 1, respectively (see Table 99.4). For *SDHD* and *SDHAF2* genes, the disease almost always occurs only when the mutations are inherited from the father, which is consistent with maternal imprinting. Penetrance is incomplete, but for *SDHB* mutations ranges between 25% and 65% by age 80 and is higher in males than females.[52-55] *SDHB* mutations predispose mainly to sympathetic paragangliomas with a high metastatic potential.[56-59] *SDHD* mutations typically are associated with multifocal parasympathetic head and neck paragangliomas and less commonly with sypathetic paraganglioma and pheochromocytomas.[60] The incidence of metastatic pheochromocytoma/paraganglioma is less common in *SDHD* than *SDHB* mutations.[59] *SDHC* mutations are rare and associated with both sympathetic and parasympathetic paragangliomas and pheochromocytoma.[61] *SDH*-related pheochromocytomas and paragangliomas produce norepinephrine or both norepinephrine and dopamine and their metabolites.[62,63] Therefore, measurement of plasma levels of dopamine or its O-methylated metabolite methoxytyramine are important in *SDH*-related pheochromocytomas.[63]

More recently, *SDHA* gene defects were identified as a rare cause of abdominal, thoracic, and head and neck paragangliomas.[64] Also, mutations of SDH assembly factor 2 (SDHAF2), which is required for the assembly of mitochondrial complex II, were associated with familial head and neck paragangliomas, but not pheochromocytoma.[65]

Careful genetic counseling and genetic testing should be offered to all first-degree family members of patients with *SDH*-related pheochromocytoma, and subsequently it is recommended that periodic clinical, biochemical, and radiologic tumor screening be performed among mutation carriers. However, no evidence-based protocol has been used for tumor screening, and recommendations vary, from annual measurement of plasma or urinary metanephrines to performing a whole-body anatomic imaging study every 2 to 3 years starting between the ages of 5 and 10 years for *SDHB* and 10 to 15 years for *SDHD* mutations carriers.[5] Recent recommendations suggest that carriers of pathogenic SDH variants should be offered surveillance with annual biochemical screening and biannual whole-body MRI, together with clinical examination.[66] Surveillance should start 5 years before the earliest age of onset in the family, and thus only children eligible for surveillance should be offered presymptomatic genetic testing. The surveillance of children younger than 15 years needs to be individually designed. For patients with previous diagnosis of *SDH*-related pheochromocytoma, lifetime follow-up is necessary, with biochemical screening every 6 to 12 months (a shorter interval is usually recommended for *SDHB* mutation patients) and imaging every 1 to 3 years based on the initial diagnosis and surgery.

It should be also noted that pheochromocytoma, especially those related to *SDHx* mutations, also may occur as part of Carney's triad (i.e., gastric leiomyosarcoma, pulmonary chondroma, and paraganglioma). The syndrome is very rare, and only approximately 25% of patients manifest all three parts of the triad. It occurs sporadically, mainly in young women, and is clearly nonfamilial. More recently, a new syndrome called the Stratakis–Carney syndrome, or the "dyad of gastrointestinal stromal tumor and paraganglioma" was identified.[67] Another association with *SDH* mutations is occurrence of pituitary adenoma (so called 3P association: pheochromocytoma/paraganglioma/pituitary adenoma).[68]

Less Common Hereditary or Other Causes

The most recently reported hereditary forms of pheochromocytoma are related to four germline mutations that include transmembrane protein 127 (*TMEM127*),[69] myc-associated factor X (*MAX*),[70] fumarate hydratase (*FH*),[71] and dihydrolipoamide S-succinyltransferase (*DSLT*)[72] (Table 99.4). The *TMEM127* gene encodes a transmembrane protein that appears to be implicated in protein trafficking and/or the protein-recycling machinery. Genetic studies of pheochromocytoma patients carrying *TMEM127* mutations indicate a low prevalence of these variants (~2% of all cases negative for other susceptibility mutations).[73] The *TMEM127*-related phenotype mainly includes single benign pheochromocytomas of the adrenal glands. *MAX* behaves as a classic tumor suppressor gene and is a key member of the MYC-MAX-MXD1 network of basic helix-loop-helix leucine zipper transcription factors that regulate cell growth, proliferation, and stem cell development. The prevalence of germline *MAX* mutations among a large group of apparently sporadic cases of pheochromocytoma or paraganglioma was low (1.12%).[74] *MAX*-related pheochromocytomas are often multifocal, and the estimated prevalence of malignancy is 10%. FH is an enzyme in the Krebs cycle that converts fumarate to malate. Pheochromocytomas related to *FH* mutations present with a noradrenergic biochemical phenotype and, similar to other pseudohypoxia-related pheochromocytomas, are associated with higher rates of multiplicity and metastasis. DSLT is the part of the E2 subunit mitochondrial α-ketoglutarate dehydrogenase complex and facilitates conversion of α-ketoglutarate to succinyl CoA and carbon dioxide. The prevalence of germline *DLST* mutations among sporadic cases of pheochromocytomas is estimated to be around 1%.[72] Individuals with *DLST* mutations present with multiple pheochromocytomas.

Recent discoveries of somatic mutations related to pheochromocytoma include mutations in hypoxia-inducible factor 1α (*HIF2A* or *EPAS1*)[75] and Harvey-ras (*H-RAS*),[76] as outlined in the Table 99.4. *HIF2A* was found to be associated with a new syndrome of multiple pheochromocytomas and duodenal somatostatinomas associated with polycythemia (Pacak–Zhuang syndrome).[77] *HIF2A* somatic mutations were also reported in patients without polycythemia and occasionally may present as germline mutations. Somatic mutations of another two genes, alpha thalassemia/mental retardation syndrome X-linked (*ATRX*) and telomerase reverse transcriptase (*TERT*), are associated with more aggressive and metastatic behavior, but are not considered to be driver mutations.[3,78,79]

BIOCHEMICAL DIAGNOSIS OF PHEOCHROMOCYTOMA

KEY POINTS

- Measurements of plasma or urinary metanephrines provide the preferred biochemical tests for diagnosis of pheochromocytoma and paraganglioma.
- Appropriate preanalytical conditions and reference intervals are crucial for accurate interpretation of laboratory test results.
- Metanephrine test results within normal limits can exclude a diagnosis of pheochromocytoma or paraganglioma in virtually all symptomatic patients.

Although catecholamines are the principle hormones produced by pheochromocytoma, their measurement in urine or plasma is not reliable for detection of tumors. In contrast, because metanephrines are produced continuously from catecholamines within pheochromocytomas by a process independent of catecholamine secretion, these metabolites provide diagnostic advantages over the parent catecholamines.[80] Consequently, the 2014 Endocrine Society Clinical Practice Guideline on pheochromocytomas and paragangliomas recommended that testing for the tumors should include measurement of plasma or urinary metanephrines.[34] Since then, a multicenter prospective study involving mass spectrometry-based measurement of plasma free, urinary free, and urinary deconjugated metanephrines established that the diagnostic efficacy of plasma and urinary free metabolites are equally effective in symptomatic patients whereas the plasma test is superior to both urine tests in patients screened due to hereditary risk, an incidentaloma, or past history of the tumors.[81] The utility of measurements of methoxytyramine for diagnosis of dopamine-producing paragangliomas was further clarified to be restricted to plasma measurements.

Preanalytics and Analytics

False-negative results for measurements of metanephrines can occur but are rare, particularly for plasma measurements[81]; in contrast, there are numerous causes of false-positive results for both plasma and urine tests (Table 99.5). False negatives can occur with use of inappropriately high reference intervals, in small tumors or tumors that do not synthesize catecholamines, including some paragangliomas and particularly those confined to the head and neck. False negatives can also occur for paragangliomas that produce only dopamine when test panels do not include plasma methoxytyramine. For acid hydrolyzed urinary metabolites measured by liquid chromatography with electrochemical detection (LC-ECD), the presence of 3-methoxy-4-hydroxybenzylamine in dietary curry leaves can result in inaccurately low urinary outputs of all metabolites, and potentially false-negative results.[82]

False positives for plasma or urine tests are commonly associated with activation of the sympathoadrenal system or with the use of drugs, such as tricyclic antidepressants, that alter the disposition of norepinephrine (Table 99.5). For the plasma test, sampling blood in the seated position is a common cause of false-positive results.[83,84] Blood for this test should be collected with patients supine for at least 20 to 30 minutes before sampling.[34] If blood cannot be taken in the supine position, the next best test involves measurements of urinary free metanephrines either in a 24-hour collection or in a first morning urine sample.[81,85]

Immunoassays for metanephrines do not provide reliable accuracy for catecholamine-producing tumors.[86] Although methods employing LC-ECD remain in use for urine measurements, liquid chromatography with liquid chromatography tandem mass spectrometry (LC-MS/MS) is preferable for both plasma and urine. This is because LC-MS/MS is minimally impacted by analytical interferences compared with other methods. Nevertheless, analytical interferences, which are assay-specific and not covered here, are minor and highly variable causes of false-positive results compared with sympathoadrenal activation and pharmacophysiological causes of false-positive results.

Interpretative Considerations

Appropriately determined reference intervals for laboratory tests are essential for minimizing false-negative or false-positive results.[34] For plasma free metanephrines, upper cutoffs of reference intervals should be optimized for blood sampling in the supine position.[87] For plasma normetanephrine, age-specific upper cutoffs minimize false-positive results for older patients and false-negative results for younger patients. For measurements of metanephrines in urine, sex-specific reference intervals are appropriate. A related consideration is the use of creatinine for normalizations in spot urines, particularly important for children and patients in whom both sex and body size warrant consideration.

For most patients with pheochromocytoma, the extent and pattern of increases of multiple metabolites in a panel are usually sufficient to indicate a tumor. Assessing the posttest probability of disease should also take into account pretest prevalence according to clinical presentation or reason for testing.[16,81] Follow-up biochemical testing after screening is prudent for patients with milder increases (e.g., <2-fold above upper cutoffs) in solitary metabolites of test panels, where the posttest probability of pheochromocytoma remains low. Before follow-up testing is implemented, consideration should be directed toward sources of false-positive results (see Table 99.5).

Given the high diagnostic accuracy of today's biochemical tests, the need for follow-up pharmacologic procedures is negligible, and in particular provocative tests are obsolete. On the other hand, in some specific patients, suppression tests such as the clonidine suppression test may still be useful.[20] If plasma normetanephrine does not fall below 112 pg/mL (0.61 nmol/L) and by less than 40% 3 hours after administration of 0.3 mg/70 kg clonidine, a catecholamine-producing tumor is highly likely.

TABLE 99.5 **Causes of False-Positive and False-Negative Results for Plasma Free or Urinary Deconjugated Normetanephrine (NMN), Metanephrine (MN), and Methoxytyramine (MTY)**

	PLASMA FREE METANEPHRINES			URINARY DECONJUGATED METANEPHRINES*		
	NMN	**MN**	**MTY**	**NMN**	**MN**	**MTY****
Sympathoadrenal activation						
Exercise or stressful daily activity	FP++	FP+	?	FP++	FP+	?
Medical conditions (e.g., heart failure)	FP++	-	?	FP++	-	?
Medical emergencies	FP+++	FP++	?	FP+++	FP++	?
Seated blood sampling/upright posture	FP++	-	-	+	-	-
Mental stress/anxiety (e.g., needle phobia)	FP++	FP+	?	?	?	?
Cold ambient temperature	FP+	-	-	?	?	?
Pharmacophysiological causes						
NE uptake blockers (e.g., tricyclics)	FP++	-	-	FP++	-	-
Nonselective alpha blockers	FP++	-	-	FP++	-	-
MAO inhibitors (rare)	FP+++	FP+++	FP+++	FP+++	FP+++	FP+++
Sympathomimetics	FP+	-	-	FP+	-	-
Foods or drugs containing catecholamine precursors or related compounds						
Certain fruits, nuts, cereals, and other foods[†]	-	-	FP++	FP+	-	FP++
L-dopa	-	-	FP+++	FP+	-	FP+++
MHBA in curry leaves [§]	na	na	na	FN+++	FN+++	FN+++
Miscellaneous causes						
Collection of urine beyond/less than 24 h	na	na	na	FP+/FN+	FP+/FN+	FP+/FN+
Large/small body size	na	na	na	FP+/FN+	FP+/FN+	FP+/FN+
Dopamine-producing tumor**	na	na	-	na	na	FN++
UC of reference intervals too high/low	FN+/FP+	FN+/FP+	FN+/FP+	FN+/FP+	FN+/FP+	FN+/FP+
Small tumor size (<2 cm)	FN++	FN+	FN+	FN++	FN+	FN+
No tumoral catecholamine synthesis	FN+++	FN+++	FN+++	FN+++	FN+++	FN+++

*Interferences for urine are shown for deconjugated measurements after an acid-hydrolysis step, but except for MHBA are likely similar for urinary free measurements.

**Urine methoxytyramine is a poor marker of tumoral dopamine production.

[†] A wide array of foods (e.g., fruits, nuts, meats) can elevate methoxytyramine in plasma and deconjugated normetanephrine and methoxytyramine in urine.

[§] MHBA is an internal standard used in assays involving liquid chromatography with electrochemical detection, so only causes false-negative results for those assays.

FP, false positive; *FN*, false negative; +, weak influence; ++, mild influence; +++, strong influence; –, no influence; *?*, influence unknown; *na*, not applicable; *NE*, norepinephrine; *MAO*, monoamine oxidase; *MHBA*, 3-methoxy-4-hydroxybenzylamine; *UC*, upper cutoff.

Apart from indicating presence or absence of disease, biochemical test results may also provide information about tumor size and location, an underlying mutation, and the presence of metastases. Pheochromocytomas may produce exclusively norepinephrine or both norepinephrine and epinephrine. In contrast, sympathetic paragangliomas almost invariably produce only norepinephrine, with or without methoxytyramine. Although increases in metanephrine or methoxytyramine relative to normetanephrine can thus indicate respective adrenal versus extraadrenal tumor locations, there are ethnic differences that must be considered.[88] Differences in metabolite profiles more reliably relate to the nature of underlying mutations and can thereby be used to assist interpretation of pathogenicity after next-generation sequencing or interpret biochemical test results in patients with known mutations.[89] Finally, metastatic tumors are often characterized by increased dopamine production, best assessed by measurements of plasma methoxytyramine.[62]

Follow-Up Biochemical Testing

Follow-up biochemical testing during the several weeks after recovery from surgery is essential for documenting potential residual tumor mass or metastatic lesions.[90] Testing and interpretation of test results should be according to the nature of elevated metabolites in test panels before surgery. After return of normal test results, annual biochemical testing is recommended in all patients for at least 10 years after surgery. In patients with high risk of recurrence (paragangliomas, large or multifocal tumors, childhood presentation, or *SDHB* mutations), annual biochemical follow-up should be lifelong.[90] Such tests should ideally include plasma metabolites, including methoxytyramine.

LOCALIZATION OF PHEOCHROMOCYTOMA

KEY POINTS

- Anatomical imaging studies (i.e., computed tomography or magnetic resonance imaging) are first used to locate a pheochromocytoma or paraganglioma. Functional imaging studies, such as positron emission tomography, provide additional utility for confirmation.

Anatomic Imaging

In most institutions, CT of the abdomen provides the initial method of localizing pheochromocytoma; this imaging technique is easy, widely available, and relatively inexpensive (Table 99.6). A dedicated adrenal CT consists of three series: noncontrast, contrast-enhanced

TABLE 99.6 **Proposed Clinical Algorithm for Imaging Investigations in Pheochromocytoma**

	Anatomic Imaging	Functional Imaging
Pheochromocytoma (sporadic)	Adrenal CT or MRI	In selected cases: inconclusive result on anatomic imaging, size >5 cm, borderline elevation of metanephrines or uncertainty regarding their accuracy ^{18}F-FDOPA or ^{123}I-MIBG preferred as first choices ^{68}Ga-DOTA-SSAs as second choice
Hereditary pheochromocytoma (except *SDHx*): *NF1/RET/VHL/MAX*	Thoracoabdominopelvic CT or MRI	In selected cases: inconclusive results on anatomic imaging regarding bilaterality in *NF-1/RET*, in *VHL* (due to multifocality risk) and *MAX* (due to multifocality and malignancy risk) ^{18}F-FDOPA preferred ^{123}I-MIBG as second choice
Extraadrenal sympathetic and/or multifocal and/or metastatic and/or *SDHx* mutation	Thoracoabdominopelvic CT, craniocervical MRI, and temporal bone CT (for TP and JP)	^{68}Ga-DOTA-SSA preferred ^{18}F-FDG as second choice in *SDHx* ^{18}F-FDOPA as second choice in *VHL/HIF2A*
Head and neck paraganglioma (sporadic or unknown genetic background)	Craniocervical MRI (with DCE sequences), temporal bone CT (for TP and JP)	^{68}Ga-DOTA-SSA preferred ^{18}F-FDOPA as second choice

DCE, dynamic contrast enhanced; TP, tympanic paraganglioma; JP, jugular paraganglioma; CT, computed tomography; MRI, magnetic resonance imaging; ^{18}F-FDOPA, (^{18}F)-fluorodihydroxyphenylalanine; ^{123}I-MIBG, ^{123}I-metaiodobenzylguanidine; ^{68}Ga-DOTA-SSA, ^{68}Ga-DOTA–somatostatin analog; ^{18}F-FDG, (^{18}F)-fluorodeoxyglucose.

(at 60 seconds), and delayed images (at 15 minutes) of the abdomen. A contrast-enhanced CT is usually obtained in the presence of expected large and multiple lesions. Previously, there have been concerns regarding the administration of iodinated contrast to patients with pheochromocytoma. However, it is now established that modern contrast agents can be safely administered. Typically, pheochromocytomas have soft tissue attenuation with a spontaneous density measurement greater than 10 Hounsfield units and typically demonstrate avid enhancement.[91] Enhancement can be heterogeneous, or there may be no enhancement due to cystic or degenerated regions within the lesion. Sensitivity is approximately 95%, but specificity is approximately 50%. Sensitivity of CT for the detection of sympathetic paraganglioma is lower than that for pheochromocytoma. For head and neck paraganglioma, CT offers some advantages over MRI (e.g., better spatial resolution and fewer motion artifacts) and enables better evaluation of the temporal bone extension of jugular paraganglioma and tympanic paraganglioma (TP).

MRI is also a very reliable method of detection of pheochromocytoma (sensitivity is ~95%, and specificity is ~70%). MRI does not expose patients to ionizing radiation, which is critical in patient populations such as children, pregnant women, and those subject to lifelong imaging surveillance. Pheochromocytoma is classically described as appearing "light-bulb" bright on T2-weighted imaging. Pheochromocytoma is also usually hypointense on T1-weighted imaging with no signal loss on opposed-phase images and demonstrates avid contrast enhancement following gadolinium injection, similar to enhancement on CT. Many pheochromocytomas may show moderate or low T2-weighted signal intensity (in up to 30% of cases), an inhomogeneous appearance (especially if they are cystic and/or hemorrhagic). Diffusion-weighted MRI can be useful in depicting lymph node and liver metastases. MRI has been reported as superior to CT for the detection of extraadrenal tumors. For head and neck paraganglioma, MRI provides better soft-tissue contrast than CT and thus offers unique information for tumor delineation and grading according to recommended classifications. In this setting, there is a current trend to use four-dimensional time-resolved gadolinium magnetic resonance angiography; this typically shows an early and intense enhancement following gadolinium administration, with a rapid washout. *In vivo* magnetic resonance spectroscopy is also an emerging technique that can be performed in selected cases to assess succinate levels (e.g., *SDH*x mutations of unknown pathogenicity).[92,93]

Differential Diagnosis

Pheochromocytoma can be mistaken for an adrenal cyst, a hematoma, an atypical adenoma, or an adrenocortical carcinoma. The imaging differential for retroperitoneal masses includes primary and secondary malignancies such as lymphoma, primary retroperitoneal sarcoma, malignant nerve sheath tumors, Castleman disease, and nodal metastases from primary malignancies. Differential diagnoses of head and neck paraganglioma are also numerous and include nerve sheath tumors (schwannoma and neurofibroma), lymph node metastasis, miscellaneous lesions, middle ear adenoma (for TP), and other rare neoplasms.

Functional Imaging

Nuclear imaging offers several advantages over anatomic imaging, such as its higher specificity conferred by the use of very specific tracers and the high visual contrast between tumor and healthy tissue, which enables the detection of tumors that could potentially be missed by anatomic imaging. Beyond its localization value, the diversity of radiopharmaceuticals provides unique opportunities for enhanced characterization of these tumors at the molecular level (e.g., catecholamine synthesis, transporter expression, somatostatin receptor expression, glucose metabolism), mirroring histological classification on a whole-body, *in vivo*, scale. These imaging phenotypes also often reflect the diverse genetic landscape of pheochromocytoma (Table 99.6). It should be noted that pheochromocytoma related to *SDH* mutations may be functionally dedifferentiated, leading to an inability to accumulate pheochromocytoma-specific radiopharmaceuticals.

Metaiodobenzylguanidine Scintigraphy

Metaiodobenzylguanidine (MIBG) is an iodinated analog of guanidine, with some structural similarities to norepinephrine. It is taken up by cells via the norepinephrine transporter and stored within the neurosecretory granules via vesicular monoamine transporters 1 and 2.[94] Many drugs modify the uptake and storage of MIBG, such as opioids, tricyclic antidepressants, antipsychotics, and some antihypertensive agents such as labetalol. ^{123}I-MIBG scintigraphy is strongly preferred over ^{131}I-MIBG scintigraphy for various reasons (e.g., better image quality, lower radiation burden, shorter need of thyroid blockade). Thyroid blockade should be performed 1 hour before ^{123}I-MIBG injection. Discontinuation of drugs interfering with MIBG uptake and retention is also required. ^{123}I-MIBG scans are usually obtained 24 hours after tracer injection and consist of planar and a single photon emission computed tomography (SPECT)/CT images. Normal uptake of ^{123}I-MIBG in the adrenals is observed up to 80% of the time and should be considered physiological if it is mild and symmetrical and if the adrenals are not enlarged on CT images.

Pheochromocytomas and sympathetic paragangliomas are detected by intense tumor uptake on ^{123}I-MIBG scan (sensitivity 83%–100%, specificity approaching 100%). Sensitivity is decreased in infracentimeter tumors (due to limited spatial resolution), *SDH*-related and metastatic pheochromocytoma, and head and neck paraganglioma. Despite the lower sensitivity for detecting metastases, a very important advantage of using ^{123}I-MIBG scintigraphy in the setting of metastatic disease is the fact that it can identify patients who are likely to benefit from targeted radionuclide therapy with low- or high-specific-activity ^{131}I-MIBG. In other clinical scenarios, positron emission tomography (PET), if available, is preferred for comprehensive localization of the disease, especially in more complex scenarios such as hereditary and/or metastatic pheochromocytoma.

Positron Emission Tomography

The advantages of PET are that it can be done within minutes or hours after injection of short-lived positron-emitting agents, radiation exposure is low, and sensitivity and spatial resolution is superior to SPECT. Two sympathoneuronal PET imaging agents that target norepinephrine transporter, (18F)-fluorodopamine and 11C-hydroxyephedrine (11C-HED), have been evaluated but are unfortunately not available at most centers and are not approved for use. Moreover, the short half-life of ^{11}C-HED is a major drawback for its routine clinical use.

The target of (^{18}F)-fluorodihydroxyphenylalanine (DOPA) is the L-type amino acid transporters (mainly LAT1 and LAT2) involved in the uptake of amine precursors. DOPA PET is very specific and can also visualize head and neck paraganliomas. Only a few anecdotal case reports of DOPA uptake in nonneuroendocrine malignancies have been reported, such as solid pseudopapillary tumor of the pancreas, thyroid neoplasm of follicular origin (oncocytoma), squamous cell carcinoma (cervix, epiglottis), poorly differentiated adenocarcinoma, and melanoma. Another advantage of using 18F-F DOPA instead of other radiopharmaceuticals for PET/CT is its limited uptake by normal adrenal glands, which allows detection of small pheochromocytomas due to *RET* or *MAX* mutations.[95] A metaanalysis including 11 studies involving 275 patients with pheochromocytomas and paragangliomas showed that the pooled sensitivity and specificity in a lesion-based analysis of DOPA PET/CT were 79% (95% confidence interval [CI]: 76%–81%) and 95% (95% CI: 84%–99%), respectively.[96]

In cases of sporadic pheochromocytoma, DOPA scanning has a sensitivity of nearly 100% and is also very sensitive for detection of pheochromocytoma related to *VHL*, *EPAS1* (*HIF2A*), and *FH* mutations.[97] In contrast, decreased sensitivity is observed for *SDH*-related pheochromocytoma of sympathetic origin, including metastatic cases. In this setting, (^{18}F)-fluorodeoxyglucose (FDG) PET was found to offer improved performance over DOPA PET and ^{123}I-MIBG.[98] The increased glucose metabolism, which characterizes *SDH*-related pheochromocytoma, has been attributed to metabolic reprogramming due to Krebs cycle blockade (pseudohypoxia)[99] and activation of the surrounding stromal tissue.[100] ^{1}FDG PET has however been surpassed by ^{68}Ga-DOTA–somatostatin analogs (SSAs).[101]

In a systematic review and metaanalysis, the pooled detection rate of pheochromocytoma and paraganglioma using ^{68}Ga-DOTA-SSA PET/CT was 93% (95% CI: 91%–95%), which was significantly higher than that of DOPA PET/CT (80% [95% CI: 69%–88%]), FDG PET/CT (74% [95% CI: 46%–91%]), and $^{123/131}$I-MIBG scintigraphy (38% [95% CI: 20%–59%], $P < 0.001$ for all) (Fig. 99.2).[102] This metaanalysis was, however, hampered by the mix of pheochromocytomas of various origin and genotypes. Head-to-head comparison between ^{68}Ga-DOTA-SSAs and DOPA PET/CT has been performed in only seven studies. Overall, ^{68}Ga-DOTA-SSAs is the most sensitive tracer for head and neck paragangliomas and *SDH*-related and metastatic pheochromocytomas and sympathetic paragangliomas, regardless of their genetic background. Physicians should be aware of possible false-positive findings, including uncinate process, duodenal heterotopic pancreatic tissue, accessory spleen (pancreatic tail), splenosis, pancreatic serous cystadenoma, bone hemangioma (vertebral), enchondroma, fibrous dysplasia, active chronic inflammation (e.g., sarcoidosis, tuberculosis, Hashimoto thyroiditis), and other somatostatin receptor-expressing tumors (e.g., meningioma, breast cancer, lymphoma, thyroid neoplasms). In hereditary syndromes (e.g., VHL), uptake can be related to another primary neuroendocrine tumors or other malignancies (e.g., central nervous system hemangioblastoma).

In summary, imaging of patients with confirmed or suspected pheochromocytoma is important in every step of their management. Although CT and MRI have excellent sensitivity, these anatomic imaging approaches may lack the specificity required to identify a mass unequivocally as a pheochromocytoma. The recent excellent results obtained with ^{68}Ga-DOTA-SSAs has simplified the diagnostic approach to pheochromocytoma (Table 99.6) and can also assess their somatostatin receptor expression for peptide receptor radiotherapy.[103]

MANAGEMENT OF PHEOCHROMOCYTOMA

KEY POINTS

- Prompt surgical resection via minimally invasive surgery provides optimal treatment of pheochromocytoma in nearly all patients.
- Surgical intervention in patients with a biochemically positive pheochromocytoma or paraganglioma should be performed after preoperative blockade of the actions of catecholamines (α-adrenoceptor blocker for 7–14 days prior to surgery).
- Postoperative management includes follow-up biochemical testing to confirm success of surgical intervention, with continual follow-up to screen for recurrent or metastatic disease.

An individualized management approach in patients with pheochromocytoma may be expected to assure best outcomes. The optimal therapy for a pheochromocytoma is prompt surgical removal of the tumor; this is because an unresected tumor represents a time bomb waiting to explode with a potentially lethal cardiovascular crisis.[36] In today's experienced teams, most retroperitoneal tumors can be resected safely with a mortality rate approaching zero due to the many improvements in current clinical practice (experienced and coordinated team, preoperative preparation, laparoscopic approach, new anesthetic drugs and monitoring devices).[104,105]

Minimally invasive surgery by laparoscopic adrenalectomy is the first choice technique in all patients with pheochromocytoma, as it is associated with lower operative blood loss and shorter hospital stay than open adrenalectomy.[106,107] Robotic adrenalectomy and a single-port approach are evolving, but so far, the benefits over laparoscopic adrenalectomy have not been established.[108]

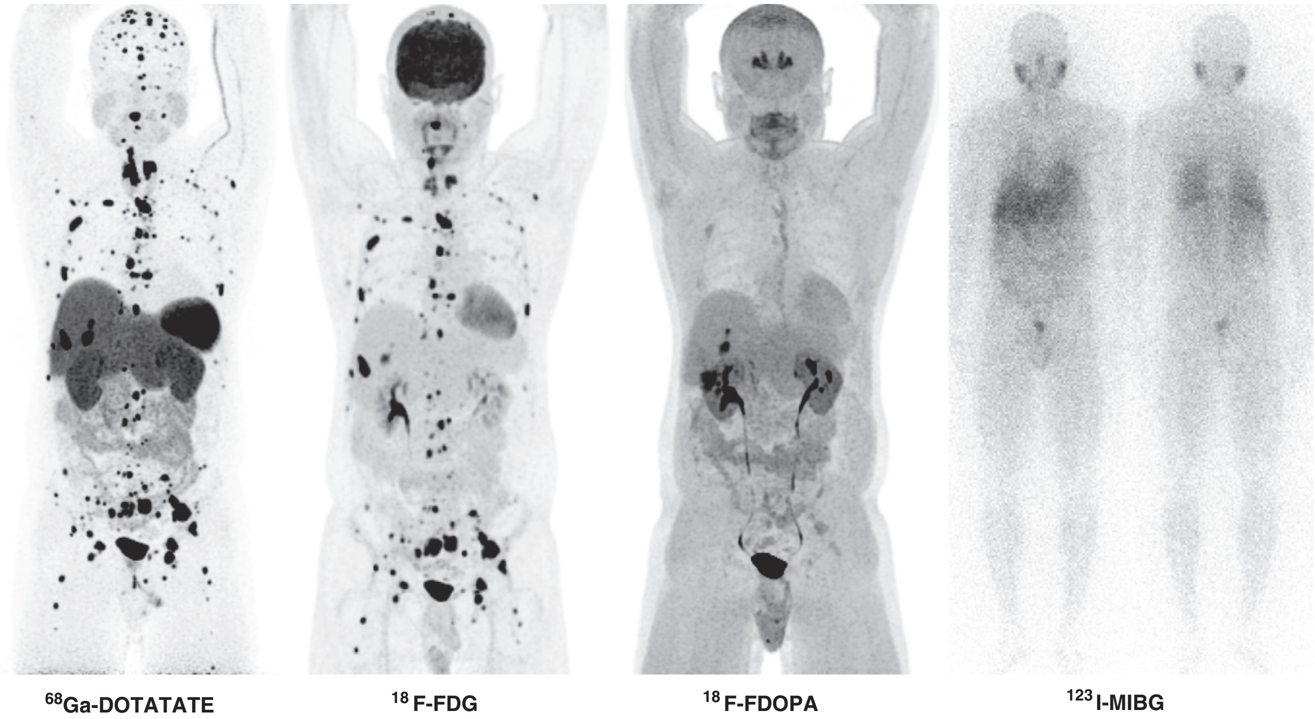


Figure 99.2 **Metastatic *SDHB*-related paraganglioma.** A 41-year-old male with *SDHB*-related metastatic paraganglioma in multiple bones, liver, and pelvic nodes showing variable tumor localization on multimodality molecular imaging. ^{68}Ga-DOTATATE positron emission tomography (PET)/computed tomography (CT) shows the highest number of lesions of any of the imaging modalities, followed by (^{18}F)-fluorodeoxyglucose PET/CT. (^{18}F)-fluorodihydroxyphenylalanine PET/CT and ^{123}I-metaiodobenzylguanidine scintigraphy are suboptimal due to functional dedifferentiation. This imaging phenotype is typical of metastatic *SDHx*-related pheochromocytoma and paraganglioma.

Head and neck paragangliomas often exhibit indolent behavior and rarely secrete catecholamines. Although complete surgical resection is the only curative treatment for these tumors, it may not be advised due to their size and specific anatomical location. Additionally, lower cranial nerve injury represents a major risk in surgical resection. Surgery should therefore be considered in Shamblin class I and II carotid body and most tympanic paragangliomas. *SDHB*- and *SDHD*-related carotid body paragangliomas should be surgically removed when they reach size of 1.5 or 2 cm, respectively. In other cases (e.g., vagal or jugular paragangliomas), observation or various forms of therapeutic radiation and, in some patients with rapidly growing tumors, chemotherapy should be considered as first choices.[60,109]

Medical Therapy and Preparation for Surgery

The main goal of preoperative management of a pheochromocytoma patient is 3-fold: 1) to normalize blood pressure, heart rate, and function of other organs; 2) to restore volume depletion; and 3) to prevent a surgery-induced catecholamine storm, with its consequences on the cardiovascular system and metabolism.[17,18] Therefore, when preoperative adrenoceptor blockade is initiated, the finally used drug doses are usually much higher, more than one drug is often necessary, and targets for cardiovascular indices are much more strict than they would be if no operation were performed (Table 99.7). The minimal duration of pretreatment is usually 7 to 14 days. Panels of experts either under the US Endocrine Society or the European Society of Hypertension recommended that all patients with a biochemically positive pheochromocytoma should receive appropriate preoperative medical management to block the effects of released catecholamines.[1,34] The recommendations on preoperative preparation with α-adrenoceptor blockade are based on optimal care of a pheochromocytoma patient, not only during surgery but also before surgery, when cardiovascular emergency and crisis may occur.

Despite wide-ranging practices, international differences in available or approved therapies, and a scarcity of evidence-based studies comparing different therapies, α-adrenoceptor blockers are the therapy of choice (Table 99.7). Other antihypertensives such as calcium channel blockers may be occasionally used in situations where there is only very mild elevations of blood pressure. For tachyarrhythmias, β-adrenoceptors or the calcium channel blocker verapamil are recommended. Correct use of these drugs, as well as improved surgical and anesthesiologic approaches, advances in medical imaging, and overall perioperative care, resulted in a significant drop in perioperative mortality, which is currently approximately 0% to 3 %.[110]

Maintenance of adequate blood pressure control for at least 1 to 2 weeks before the operation is an important aspect of management. Although no specific consensus on blood pressure and heart rate targets has been established, a seated blood pressure target of less than 130/80 mm Hg and upright systolic blood pressure of greater than 90 mm Hg has been proposed (Table 99.7).[34] Suggested targets for heart rate are 60 to 70 bpm in a seated and 70 to 80 bpm for an upright position, respectively.

Currently, nonselective and selective α-adrenoceptor blockers are available for treatment of patients with biochemically positive pheochromocytoma (Table 99.7).[17,18] Phenoxybenzamine is a noncompetitive nonselective α-adrenoceptor blocker commonly used for preoperative blockade. Phenoxybenzamine has a long-lasting effect that diminishes only after *de novo* α-adrenoceptor synthesis. The initial low dose of long-acting phenoxybenzamine of 10 mg twice daily

TABLE 99.7 Proposed Algorithm for the Preoperative Treatment of Patients With Pheochromocytoma

Start α-adrenoceptor blocker at least 7–14 days before surgery
Oral salt and fluid loading (e.g., keep regular salt diet, may use NaCl tablets/250 mmol = 15 g/one to three tablets a day)
Gradually increase α-adrenoceptor blocker until hemodynamic goals* are met (e.g., until maximum dose of phenoxybenzamine 30–40 mg three times daily or doxazosin 12–15 mg twice daily)
Add β-adrenoceptor blocker after at least 2–4 days of α-adrenoceptor blocker, when significant tachycardia or a catecholamine-induced arrhythmia occurs (e.g., atenolol 12.5–25 mg once or twice daily, metoprolol 25–50 mg once, twice, or three times daily)
Gradually increase β-adrenoceptor blocker until hemodynamic goals* are met
Add calcium-channel antagonist when hemodynamic goals* are not met despite maximum α-adrenoreceptor blockade (e.g., nifedipine or amlodipine)
Administer a saline infusion (e.g., 2 L NaCL 0.9% in 24 hours) 12–24 h before surgery

*Hemodynamic goals: supine systolic/diastolic blood pressure <130/80 mm Hg, systolic blood pressure in the upright position between 90 and 110 mm Hg, heart rate <80 beats per minute in the supine position and <100 beats per minute while standing.
Adapted from Berends et al.[18]

is to prevent early orthostatic hypotension; the dose is increased until clinical manifestations are controlled or intolerable side effects appear. For most patients, a total daily dose of 1 mg/kg is sufficient. Nevertheless, some patients may require much larger doses. Other useful α-adrenoceptor blocking agents are prazosin, terazosin, and the more commonly used doxazosin; all are more readily available and less expensive than phenoxybenzamine. All three drugs are competitive and short-acting α1-adrenoceptor blockers; all three may have the potential for severe postural hypotension immediately after the first dose (especially prazosin), which therefore should be administered just as the patient is ready to go to bed. Thereafter, the dose can be increased as needed.

The efficacy of phenoxybenzamine versus doxazosin on hemodynamic instability during surgery has been recently evaluated in PRESCRIPT, the first randomized controlled trial on presurgical treatment of patients with pheochromocytoma and paraganglioma.[110] It was concluded that although phenoxybenzamine was more effective in preventing intraoperative hemodynamic instability than doxazosin, there was no difference in clinical outcome between drugs. Another study showed that the non-selective α-adrenoceptor blocker, phenoxybenzamine, produced a more significant antihypertensive effect than selective α-adrenoceptor blockade in patients undergoing adrenalectomy.[111] However, no significant difference in intraoperative hemodynamic instability, postoperative recovery, or postoperative complications was observed between groups. Some retrospective studies, however, found that phenoxybenzamine may cause more pronounced postoperative hypotension than selective α-adrenoceptor blockade.[112]

It should be emphasized that, together with α-adrenoceptor blockers, patients are encouraged to consume a normal or high-salt diet and engage in high fluid intake to restore diminished blood volume to normal. This ameliorates the severity of postural hypotension due to α-adrenoceptor blockade.

Labetalol is another drug that has both α- and β-adrenoceptor blocking activity, and has been used in patients with pheochromocytoma, usually at a dosage of 200 to 600 mg twice daily. However, due to its stronger blocking effect on β- than α-adrenoceptors, it can only be used in those who are on another α-adrenoceptor blocker to avoid unopposed α-adrenoceptor-induced vasoconstriction due to loss of β-adrenoceptor-mediated vasodilatation, potentially resulting in a hypertensive crisis.

β-adrenoceptor blockers are needed only when significant tachycardia or a catecholamine-induced arrhythmia occurs (Table 99.7).[36] β-adrenoceptor blockers such as atenolol or metoprolol can be used. β-adrenoceptor blockers, however, should never be used before α-adrenoceptor blockers are administered, because unopposed stimulation of α-adrenoceptors due to loss of β-adrenoceptor-mediated vasodilatation may elicit a serious and life-threatening elevation of blood pressure.[36]

Metyrosine is a drug that can be also used in the treatment of patients with pheochromocytoma, but only as an add-on after appropriate adrenoceptor blockade has been initiated. The drug competitively inhibits tyrosine hydroxylase, the rate-limiting step in catecholamine biosynthesis depleting catecholamine stores. Thus, it facilitates blood pressure control both before and during surgery, especially during the induction of anesthesia and surgical manipulation of large tumors with largely elevated catecholamine/metanephrine levels. Treatment is started at a dosage of 250 mg orally every 6 to 12 hours; thereafter, the dose is increased by 250 to 500 mg every 2 to 3 days. Inhibition of catecholamine synthesis in the brain may often cause sedation and depression, as well as extrapyramidal signs (e.g., Parkinsonism) in older patients.

Hypertensive crises that can manifest as severe headache, visual disturbances, or cerebrovascular accidents can be appropriately treated with an intravenous bolus of 5 mg phentolamine. Phentolamine has a very short half-life; therefore, if necessary, the same dose can be repeated every 2 minutes until hypertension is adequately controlled, or phentolamine can be administered as a continuous infusion (100 mg of phentolamine in 500 mL of 5% dextrose in water). It is contraindicated in patients with coronary artery disease or heart failure, as it can cause severe vasodilation and reflex tachycardia. A continuous intravenous infusion of sodium nitroprusside, urapidil (α_1-adrenoceptor blocker), or calcium channel blockers such as nicardipine, diltiazem, or clevidipine or hydralazine as vasodilators can be used to control hypertensive crisis, as recently extensively reviewed.[36] Attention should also be paid to the possibility that some drugs (e.g., tricyclic antidepressants, metoclopramide, naloxone) can cause hypertensive crisis in patients with pheochromocytoma (Table 99.3). Severe tachyarrhythmia can be treated by intravenous esmolol or metoprolol (β_1-adrenoceptor blocker). A proposed algorithm for preoperative treatment is shown in Table 99.7.

Postoperative Management

Volume replacement is the treatment of choice if hypotension should occur during surgery (i.e., after the tumor is removed) or in the postoperative period.[113] Control of postoperative hypotension is even more imperative if both metyrosine and phenoxybenzamine are used preoperatively because the former inhibits catecholamine synthesis by both the tumor and the sympathetic nervous system, whereas the latter blocks the action of any catecholamines that are synthesized. In this situation, the vascular bed is effectively paralyzed, resulting in a vasodilated state. Therefore, the best way to control blood pressure is using adequate volume replacement. The volume of fluid required is often large (0.5–1.5 times the patient's total blood volume) during the first 24 to 48 hours after removal of the tumor. The amount of fluid replacement needed is readily determined by observation of blood pressure, heart rate, central venous pressure, and urinary output. If hypotension persists despite adequate volume replacement, infusion of norepinephrine or other pressor agents such as vasopressin may be warranted.[113]

Postoperative hypertension may indicate that some tumor tissue was not resected. However, during the first 24 hours after surgery, hypertension is most likely attributed to pain, volume overload, or autonomic instability; these are all readily treated symptomatically. If hypertension persists in a patient even after he or she has returned to dry weight, the coexistence of essential hypertension is still the most likely diagnosis. Any attempt to collect specimens for biochemical evidence of a residual tumor should be delayed at least 3 to 5 days postsurgery to ensure that the large increases in both plasma and urinary catecholamines due to sympathetic activation by surgery have dissipated. Usually, 3 to 6 weeks after surgery, every patient should undergo repeated measurement of plasma metanephrines to detect any residual tumor. It should be noted that at least 25% of patients remain hypertensive, but this is usually easily controlled with medication. Some patients may also present with postoperative hypoglycemia, usually during the first 24 hours due to persisting inhibitory effect of high catecholamines on insulin secretion. A hypoglycemic event needs to be treated with intravenous administration of glucose.[113] Postoperative management of patients with pheochromocytoma has been recently extensively reviewed and summarized by Mamilla et al.[113]

SPECIAL PRESENTATIONS AND THERAPEUTIC PROBLEMS

Multifocal and Metastatic Pheochromocytoma/Paraganglioma

KEY POINTS

- There are currently no effective cures for metastatic pheochromocytoma or paraganglioma. However, palliative relief may be achieved through therapeutic procedures that should be considered according to presentation (e.g., surgical debulking, 131I-metaiodobenzylguanidine radiotherapy, ^{177}Lu-DOTATATE radiotherapy, or chemotherapy).

Approximately 10% to 17% of the patients with a pheochromocytoma have metastatic disease, with prevalence highly dependent on the underling gene mutation. In approximately 20% of cases, metastases are discovered synchronously with initial diagnosis of a primary tumor. In other cases, metastases may develop metachronously after several years or even decades after diagnosis and intended curative treatment of primary tumors. The most frequent metastatic locations are lymph nodes (80%), bones (72%), liver (50%), and lungs (50%).[56,114,115] In rare cases, metastases to the brain, breasts, skin, and ovaries have been described.

A larger than 5 cm diameter, an extraadrenal location of the primary tumor, and high plasma concentrations of methoxytyramine are established determinants of metastatic risk.[57,58] The most prominent molecular marker to predict a malignant course is the presence of a germline *SDHB* mutation.[57,58,116]

The natural course of metastatic pheochromocytoma is highly heterogeneous among patients. The overall 5-year survival rate lies between 34% and 85%.[56,117] Mortality, progression-free survival, and clinical outcome are difficult to predict on an individual basis. Patients may have (very) indolent disease, despite the presence of extensive distant metastases, especially when confined to the skeleton. These patients may have an excellent quality of life and may need little or no therapeutic intervention for a prolonged period of time. In a French series, half of therapy-naive patients with metastatic pheochromocytoma achieved stable disease at 1 year follow-up.[118] On the other hand, some patients suffer from aggressive disease that requires systemic therapy. Several factors have been linked to decreased survival as outlined above. This includes the presence of a germline mutation in *SDHB* with a prevalence of 30% in patients with metastatic pheochromocytoma and up to 90% if the tumor originates from an extraadrenal location.[116]

Currently, there are no formal guidelines for treatment of metastatic pheochromocytoma, though application of the most common treatment modalities may follow certain considerations related to positivity on ^{123}I-MIBG or ^{68}Ga-DOTATATE scans (radionuclide therapy), rapid growth (chemotherapy), and clinically relevant compression (e.g., spinal cord) (surgery). Treatment should be tailored to the individual patient, with appropriate management that requires a multidisciplinary approach, involving specialists in endocrinology, oncology, surgery, radiotherapy, interventional radiology, pathology, and others. The goals of treatment are to prevent and manage catecholamine-related adverse events, reduce tumor size, prevent tumor progression, and palliate symptoms related to tumor burden.

In patients with limited metastatic disease, such as lymph node metastases, extended surgical resection can be considered, but rarely with a curative outcome. Even in the presence of incurable disease, resection (or debulking) of the primary tumor may be of benefit for survival. Other local therapies that can be applied for local symptom relief include external beam irradiation, radiofrequency ablation, tumor embolization, cryotherapy, and percutaneous microwave coagulation. Patients with epidural cord compression may benefit from spine surgery, external beam irradiation, or a combination of both. Stable compression fractures can be safely treated with percutaneous cement injection. Spinal instability can only be treated with surgery. Prior to these interventions patients should receive adrenoceptor blockers, with the only exceptions being those in whom it is clear that disease does not involve production of catecholamines. Such blockade is also essential to control symptoms of catecholamine excess and minimize risk of catecholamine crisis. In some cases, metyrosine might be required to block catecholamine synthesis.

Apart from cardiovascular manifestations, patients with metastatic pheochromocytoma can suffer from severe constipation or even pseudoobstruction due to inhibitory actions of catecholamines, in particular norepinephrine, on bowel motility. Such presentations can benefit from vigorous additional treatment with laxatives.

Targeted radiotherapy with ^{131}I-MIBG can be considered in patients with MIBG-positive metastatic lesions on diagnostic scanning. Both high-dose regimens and fractionated low-dose regimens can be effective in achieving a therapeutic response. A metaanalysis on the effectiveness of different regimens of ^{131}I-MIBG radiotherapy showed a complete response in 3%, a partial response in 27%, and stable disease in 52% of the patients.[119] Biochemical responses were as follows: 11% complete response, 40% partial response, and 21% stable disease. ^{131}I-MIBG treatment is generally well-tolerated, although high-dose regimens may cause bone marrow toxicity. More recently, positive effects have also been reported for ultratrace iobenguane, a high-specific-activity formulation of ^{131}I-MIBG that is US Food and Drug Administration-approved in the United States.[120]

A potentially more effective treatment than ^{131}I-MIBG radiotherapy involves peptide receptor radionuclide therapy (PRRT) with radiolabeled somatostatin analogs, which can be considered in patients with metastases that are positive on somatostatin receptor-targeted PET scanning. Prospective evaluations of this treatment modality are underway, but according to a systematic review and metaanalysis, treatment with PRRT achieved an objective response rate of 25% and a disease control rate of 84%.[121] Clinical and biochemical responses were seen in 61% and 64% of patients, respectively. Similar tumor response rates were noted for ^{90}Y-yttrium– and ^{177}Lu-lutetium-based PRRT regimens. Treatment-related adverse effects such a bone marrow toxicity and nephrotoxicity appear to be limited, especially with ^{177}Lu-DOTA-SSA.

Different chemotherapeutic agents have been used for metastatic pheochromocytoma/paraganglioma, including cyclophosphamide, doxorubicin, vincristine, dacarbazine, temozolomide, 5-fluorouracil, methotrexate, ifosfamide, streptozotocin, and platinum compounds.[122] The best-studied protocol, although in retrospective studies only, is the combination of cyclophosphamide, vincristine, and dacarbazine, usually administered in 21-day cycles. Complete and partial radiological response rates were 11% and 44%, respectively.[123] The median duration of response was 20 months. There was no difference in overall survival between patients whose tumors objectively shrank and those with stable or progressive disease. However, this combination chemotherapy can facilitate control of blood pressure as well as symptoms.[124] On the other hand, quality of life may be negatively impacted by side effects. Patients who are most likely to benefit from chemotherapy are those with rapidly progressive disease and tumor-related symptoms that cannot be easily controlled by supportive therapies. As an oral alternative, temozolomide has been successfully applied in some patients with metastatic pheochromocytoma due to *SDHB* mutations.[32] The role of chemotherapy in a (neo)adjuvant setting of pheochromocytoma remains to be determined.

In a subset of patients, antiangiogenetic treatment with tyrosine kinase inhibitors such as sunitinib may result in a partial response, disease stabilization, and improved blood pressure control.[125] Prospective clinical trials are ongoing.

Pediatric Pheochromocytoma

KEY POINTS

- The prevalence of hereditary cases is higher in children than in adults. Mutations associated with higher prevalences of extraadrenal, multifocal, metastatic, and recurrent pheochromocytomas are particularly higher in children than in adults.

Presentations of pheochromocytoma in children deserve special consideration, particularly due to the much higher prevalence of hereditary cases than in adults, especially cluster 1 mutations that lead to activated pseudohypoxic pathways.[126] More specifically, tumors due to cluster 1 mutations (*VHL*, *SDHB/D*) are twice as prevalent among children than adults, whereas the prevalence of cluster 2 mutations involving *RET*, *NF1*, *TMEM127*, and *MAX* is 3-fold lower in children than in adults. Due to the above differences, children show considerably higher prevalences of extraadrenal, multifocal, metastatic, and recurrent pheochromocytomas than adults. The genes most commonly reported to harbor mutations as the cause of childhood pheochromocytoma include, in decreasing prevalence, *SDHB, VHL,* and *SDHD* and, at a much lower prevalence, *RET*.[127-129]

Considering the high prevalence of underlying germline mutations in children, even in apparently sporadic cases, involvement of a genetic counselor is essential in all cases of pediatric pheochromocytoma. Importantly, most children who present with metastatic pheochromocytoma harbor *SDHB* mutations (metastases, which often develop within 7 years after the initial diagnosis of a primary tumor, are the most common cause of death).[130]

Because of the early age of development of hereditary pheochromocytoma and associated tumors including well-defined syndromes (e.g., retinal hemangioblastoma in VHL syndrome, medullary thyroid cancer in MEN2), it is recommended that screening for mutations in relatives of identified carriers begin in early childhood, as outlined earlier. It should be also noted that several somatic or *de novo* mutations of pheochromocytoma susceptibility genes (*HIF2A*, *NF1*, *VHL*) have been reported in pediatric pheochromocytoma.[77] Thus, physicians should be aware of the fact that some pediatric patients may not present with any positive family history.

Pheochromocytomas are the most common endocrine tumors in children, but only account for 10% to 14% of all pheochromocytomas, with a detected incidence of less than 2 per million per year.[131] Nevertheless, this prevalence is likely a gross underestimate, because many pheochromocytomas that develop in childhood are likely not diagnosed until early adulthood.[126] This can be appreciated from considerations of the generally slow-growing nature of pheochromocytoma; with an average doubling time of 2 years,[132] it can be calculated that it would take 12 years for a 1-cm-diameter tumor (0.52 cubic cm volume) to reach a diameter of 4 cm and a volume of 33 cubic cm. Because the peak age at diagnosis of cluster 1 pheochromocytomas is 30 years,[133] a substantial proportion of such tumors must actually first develop in childhood.

When pheochromocytoma is detected in childhood, the peak age is between 10 and 13 years, with a male predominance (male vs. female, 2:1) before puberty.[134] As in adults, pheochromocytoma in children usually presents with hypertension and symptoms of catecholamine excess. In contrast to adult patients, in whom sustained hypertension is found in only 50% of cases, more than 70% to 90% of children present with sustained hypertension.[135] Pheochromocytoma is the underlying cause in 1% to 2% of cases of pediatric hypertension and should be considered after exclusion of more common causes such as renal disease and renal artery stenosis. Sweating, visual problems, weight loss, polyuria, polydipsia, nausea, and vomiting are common in children.[136] Less frequent clinical manifestations include fever, constipation, headache, anxiety, syncope, cyanotic episodes, and those related to cardiomyopathy.

As in adults, measurements of plasma free metanephrines provide a highly sensitive (100%) and specific (94%) test for the diagnosis of pheochromocytoma in children.[137] However, the biochemical diagnosis of pheochromocytoma can be made difficult by stress-associated false-positive results during blood sampling and incomplete urine collections (when relying on 24-hour urinary excretion of metanephrines). While plasma concentrations of normetanephrine are lower in children than in adults, plasma concentrations of metanephrine are higher in children than in adults, and in boys than in girls. Thus, sex- and age-appropriate reference intervals for plasma metanephrines are also called for in children. This is most important for plasma normetanephrine, which shows a continuous upward trend with aging from 5 years of age onwards through adulthood.[138] The predominance of cluster 1 mutations among children is paralleled by a higher prevalence of noradrenergic tumors in children compared to adults; this is characterized by lack of increases in plasma metanephrine in 93.2% of childhood cases compared with 57.3% of adult cases.[126]

For the localization of pheochromocytoma in children and young adults, MRI is preferred over CT to limit radiation exposure. Additional functional imaging is recommended in children, considering a higher likelihood of multifocal disease. In the differential diagnosis of ^{123}I-MIBG-positive lesions, neuroblastoma should be considered.[139]

For the surgical management of pediatric pheochromocytoma, a laparoscopic approach via either the retroperitoneal or transperitoneal route is preferred.[140] Especially in children, in the setting of bilateral adrenal pheochromocytoma or a high chance of developing contralateral disease, for example in case of MEN2, cortex-sparing partial adrenalectomy should be considered to avoid adrenal insufficiency. The principles of perioperative and anesthesiologic management are similar in children and adults, on condition of age-appropriate use of medications.

Pheochromocytoma in Pregnancy

Pheochromocytoma in pregnancy is extremely rare, at one in 15,000 to 54,000 pregnancies, but carries a high risk of morbidity and mortality if the tumor remains undiagnosed antenatally.[141] Today, maternal and fetal mortality rates have decreased to 8% and 17%, respectively.[142] The clinical presentation of pheochromocytoma during pregnancy is similar to that outside pregnancy, and as gestation progresses, symptoms worsen, eventually culminating in a life-threatening catecholaminergic crisis. This applies in particular to the peripartum period. A clinical presentation with hypertension is usually ascribed to more prevalent pregnancy-related hypertensive syndromes such as (pre) eclampsia, resulting in missing the diagnosis of pheochromocytoma. Other clinical clues that may assist in differentiating pregnancy-related hypertension from pheochromocytoma are acute onset of symptoms, presentation in first 20 weeks of gestation, paroxysmal hypertension, paroxysmal headache, palpitations, and postural hypotension, all which are more likely in pheochromocytoma.[143] Patients known to have a previous history of pheochromocytoma or a mutation in one of the pheochromocytoma-associated susceptibility genes have a high *a priori* risk.

The biochemical diagnosis is not different from that in patients without pheochromocytoma, with measurement of plasma free or urinary metanephrines as the test of choice. If the diagnosis is biochemically confirmed, MRI is the preferred imaging method. Once the tumor is localized, it is recommended to remove the tumor by laparoscopy in the first 24 weeks of pregnancy after proper preoperative α-adrenoceptor blockade. In patients in whom the diagnosis is established in the third trimester, tumor removal is to be done at or after delivery.[141]

Although the optimal mode of delivery is controversial, in most cases elective cesarean section has been used as mode of delivery, as intensive labor during vaginal delivery has been considered a risk factor for cardiovascular complications. However, choice of delivery should be personalized depending on parity, previous cesarean delivery, success of pretreatment, and personal preference of the patient.[144] Therefore, management of this specific group of high-risk patients requires involvement of an experienced multidisciplinary team.

FUTURE DIRECTIONS

Multiomics approaches together with artificial intelligence will soon dominate large-scale national and international studies to better uncover the development, pathogenesis, and molecular biology of this tumor.[4] We also predict that algorithms using artificial intelligence approaches and novel bioinformatics technologies will result in better prediction of malignant potential in pheochromocytoma and an improved decision-making process for appropriate patient-tailored therapeutic options, surveillance, and ultimately its cure. Near future new treatments including demethylating agents, HIF-2α inhibitors, radiolabelled somatostatin receptor 2 antagonists, topoisomerase/PARP inhibitors, glutaminase inhibitors, and immune checkpoint inhibitors, as well as new theranostics approaches, will become reality.[122]

REFERENCES

1. Lenders JWM, Kerstens MN, Amar L, et al. Genetics, diagnosis, management and future directions of research of phaeochromocytoma and paraganglioma: a position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J Hypertens*. 2020;38:1443–1456.
2. Lenders JW, Eisenhofer G, Mannelli M, et al. Phaeochromocytoma. *Lancet*. 2005;366:665–675.
3. Fishbein L, Khare S, Wubbenhorst B, et al. Whole-exome sequencing identifies somatic ATRX mutations in pheochromocytomas and paragangliomas. *Nat Commun*. 2015;6:6140.
4. Crona J, Taieb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification. *Endocr Rev*. 2017;38:489–515.
5. Gupta G, Pacak K, Committee AAS. Precision medicine: an update on genotype/biochemical phenotype relationships in pheochromocytoma/paraganglioma patients. *Endocr Pract*. 2017;23:690–704.
6. Sutton MG, Sheps SG, Lie JT. Prevalence of clinically unsuspected pheochromocytoma. Review of a 50-year autopsy series. *Mayo Clin Proc*. 1981;56:354–360.
7. Berends AMA, Buitenwerf E, de Krijger RR, et al. Incidence of pheochromocytoma and sympathetic paraganglioma in The Netherlands: a nationwide study and systematic review. *Eur J Intern Med*. 2018;51:68–73.
8. Eisenhofer G, Goldstein DS, Kopin IJ, et al. Pheochromocytoma: rediscovery as a catecholamine-metabolizing tumor. *Endocr Pathol*. 2003;14:193–212.
9. Eisenhofer G, Kopin IJ, Goldstein DS. Catecholamine metabolism: a contemporary view with implications for physiology and medicine. *Pharmacol Rev*. 2004;56:331–349.
10. Eisenhofer G. The role of neuronal and extraneuronal plasma membrane transporters in the inactivation of peripheral catecholamines. *Pharmacol Ther*. 2001;91:35–62.
11. Eisenhofer G, Rundqvist B, Aneman A, et al. Regional release and removal of catecholamines and extraneuronal metabolism to metanephrines. *J Clin Endocrinol Metab*. 1995;80:3009–3017.
12. Eisenhofer G, Goldstein DS. Peripheral dopamine systems. In: Robertson D, Low P, Burnstock G, et al., eds. *Primer on the Autonomic Nervous System*. San Diego: Elsevier; 2004.
13. Eisenhofer G, Aneman A, Friberg P, et al. Substantial production of dopamine in the human gastrointestinal tract. *J Clin Endocrinol Metab*. 1997;82:3864–3871.
14. Baines AD, Craan A, Chan W, et al. Tubular secretion and metabolism of dopamine, norepinephrine, methoxytyramine and normetanephrine by the rat kidney. *J Pharmacol Exp Ther*. 1979;208:144–147.
15. Brown MJ, Allison DJ. Renal conversion of plasma DOPA to urine dopamine. *Br J Clin Pharmacol*. 1981;12:251–253.
16. Geroula A, Deutschbein T, Langton K, et al. Pheochromocytoma and paraganglioma: clinical feature-based disease probability in relation to catecholamine biochemistry and reason for disease suspicion. *Eur J Endocrinol*. 2019;181:409–420.
17. Pacak K. Preoperative management of the pheochromocytoma patient. *J Clin Endocrinol Metab*. 2007;92:4069–4079.
18. Berends AMA, Kerstens MN, Lenders JWM, et al. Approach to the patient: perioperative management of the patient with pheochromocytoma or sympathetic paraganglioma. *J Clin Endocrinol Metab*. 2020;105:3088–3102.
19. Mesmar B, Poola-Kella S, Malek R. The physiology behind diabetes mellitus in patients with pheochromocytoma: a review of the literature. *Endocr Pract*. 2017;23:999–1005.
20. Eisenhofer G, Goldstein DS, Walther MM, et al. Biochemical diagnosis of pheochromocytoma: how to distinguish true- from false-positive test results. *J Clin Endocrinol Metab*. 2003;88:2656–2666.
21. Mamilla D, Gonzales MK, Esler MD, et al. Pseudopheochromocytoma. *Endocrinol Metab Clin North Am*. 2019;48:751–764.
22. Eisenhofer G, Januszewicz C, Pamporaki K, et al. Endocrine hypertensive emergencies. In: Matfin GA, ed. *Endocrine and Metabolic Medical Emergencies*. Hoboken, NJ: John Wiley & Sons; 2018.
23. Mamilla D, Manukyan I, Fetsch PA, et al. Immunohistochemical distinction of paragangliomas from epithelial neuroendocrine tumors-gangliocytic duodenal and cauda equina paragangliomas align with epithelial neuroendocrine tumors. *Hum Pathol*. 2020;103:72–82.

24. Thompson LDR, Gill AJ, Asa SL, et al. Data set for the reporting of pheochromocytoma and paraganglioma: explanations and recommendations of the guidelines from the International Collaboration on Cancer Reporting. *Hum Pathol*. 2020;110:83–97.
25. Tischler A, De Krijger RR, Gill AJ, et al. Familial paraganglioma-phaeochromocytoma syndromes caused by SDHB, SDHC, and SDHD mutations. In: Lloyd RV, Osamura RY, Kloppel G, et al., eds. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon: WHO Press; 2017:262–265.
26. Kimura N, Takekoshi K, Naruse M. Risk stratification on pheochromocytoma and paraganglioma from laboratory and clinical medicine. *J Clin Med*. 2018;7:242.
27. Fishbein L, Leshchiner I, Walter V, et al. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell*. 2017;31:181–193.
28. Neumann HP, Bausch B, McWhinney SR, et al. Germ-line mutations in nonsyndromic pheochromocytoma. *N Engl J Med*. 2002;346:1459–1466.
29. Buffet A, Burnichon N, Favier J, et al. An overview of 20 years of genetic studies in pheochromocytoma and paraganglioma. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101416.
30. Nolting S, Ullrich M, Pietzsch J, et al. Current management of pheochromocytoma/paraganglioma: a guide for the practicing clinician in the era of precision medicine. *Cancers (Basel)*. 2019;11:1505.
31. Dahia PL. Pheochromocytoma and paraganglioma pathogenesis: learning from genetic heterogeneity. *Nat Rev Cancer*. 2014;14:108–119.
32. Hadoux J, Favier J, Scoazec JY, et al. SDHB mutations are associated with response to temozolomide in patients with metastatic pheochromocytoma or paraganglioma. *Int J Cancer*. 2014;135:2711–2720.
33. Jimenez C, Fazeli S, Roman-Gonzalez A. Antiangiogenic therapies for pheochromocytoma and paraganglioma. *Endocr Relat Cancer*. 2020;27:R239–R254.
34. Lenders JW, Duh QY, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:1915–1942.
35. Grubbs EG, Halperin DM, Waguespack SG, et al. History of the multiple endocrine neoplasia workshops and overview of MEN2019. *Endocr Relat Cancer*. 2020;27:E1–E5.
36. Nazari MA, Rosenblum JS, Haigney MC, et al. Current view: Pathophysiology and acute management of tachyarrhythmias in pheochromocytoma. *J Am Coll Cardiol*. 2020;74:451–464.
37. Eisenhofer G, Huynh TT, Elkahloun A, et al. Differential expression of the regulated catecholamine secretory pathway in different hereditary forms of pheochromocytoma. *Am J Physiol Endocrinol Metab*. 2008;295:E1223–E1233.
38. Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
39. Rednam SP, Erez A, Druker H, et al. Von Hippel-Lindau and hereditary pheochromocytoma/paraganglioma syndromes: clinical features, genetics, and surveillance recommendations in childhood. *Clin Cancer Res*. 2017;23:e68–e75.
40. Jafri M, Maher ER. The genetics of phaeochromocytoma: using clinical features to guide genetic testing. *Eur J Endocrinol*. 2012; 166:151–158.
41. Eisenhofer G, Lenders JW, Linehan WM, et al. Plasma normetanephrine and metanephrine for detecting pheochromocytoma in von Hippel-Lindau disease and multiple endocrine neoplasia type 2. *N Engl J Med*. 1999;340:1872–1879.
42. Nielsen SM, Rhodes L, Blanco I, et al. Von Hippel-Lindau disease: genetics and role of genetic counseling in a multiple neoplasia syndrome. *J Clin Oncol*. 2016;34:2172–2181.
43. Al-Sharefi A, Javaid U, Perros P, et al. Clinical presentation and outcomes of phaeochromocytomas/paragangliomas in neurofibromatosis type 1. *Eur Endocrinol*. 2019;15:95–100.
44. Kepenekian L, Mognetti T, Lifante JC, et al. Interest of systematic screening of pheochromocytoma in patients with neurofibromatosis type 1. *Eur J Endocrinol*. 2016;175:335–344.
45. Geurts JL, Strong EA, Wang TS, et al. Screening guidelines and recommendations for patients at high risk of developing endocrine cancers. *J Surg Oncol*. 2020;121:975–983.
46. Bezawork-Geleta A, Rohlena J, Dong L, et al. Mitochondrial complex ii: at the crossroads. *Trends Biochem Sci*. 2017;42:312–325.
47. Jochmanova I, Yang C, Zhuang Z, et al. Hypoxia-inducible factor signaling in pheochromocytoma: turning the rudder in the right direction. *J Natl Cancer Inst*. 2013;105:1270–1283.
48. Burnichon N, Rohmer V, Amar L, et al. The succinate dehydrogenase genetic testing in a large prospective series of patients with paragangliomas. *J Clin Endocrinol Metab*. 2009;94:2817–2827.
49. Benn DE, Croxson MS, Tucker K, et al. Novel succinate dehydrogenase subunit B (SDHB) mutations in familial phaeochromocytomas and paragangliomas, but an absence of somatic SDHB mutations in sporadic phaeochromocytomas. *Oncogene*. 2003;22:1358–1364.
50. Amar L, Bertherat J, Baudin E, et al. Genetic testing in pheochromocytoma or functional paraganglioma. *J Clin Oncol*. 2005;23:8812–8818.
51. Pandit R, Khadilkar K, Sarathi V, et al. Germline mutations and genotype-phenotype correlation in Asian Indian patients with pheochromocytoma and paraganglioma. *Eur J Endocrinol*. 2016;175:311–323.
52. Jochmanova I, Wolf KI, King KS, et al. SDHB-related pheochromocytoma and paraganglioma penetrance and genotype-phenotype correlations. *J Cancer Res Clin Oncol*. 2017;143:1421–1435.
53. Rijken JA, Niemeijer ND, Jonker MA, et al. The penetrance of paraganglioma and pheochromocytoma in SDHB germline mutation carriers. *Clin Genet*. 2018;93:60–66.
54. Niemeijer ND, Rijken JA, Eijkelenkamp K, et al. The phenotype of SDHB germline mutation carriers: a nationwide study. *Eur J Endocrinol*. 2017;177:115–125.
55. Andrews KA, Ascher DB, Pires DEV, et al. Tumour risks and genotype-phenotype correlations associated with germline variants in succinate dehydrogenase subunit genes SDHB, SDHC and SDHD. *J Med Genet*. 2018;55:384–394.
56. Turkova H, Prodanov T, Maly M, et al. Characteristics and outcomes of metastatic Sdhb and sporadic pheochromocytoma/paraganglioma: an National Institutes of Health study. *Endocr Pract*. 2016;22:302–314.
57. Crona J, Lamarca A, Ghosal S, et al. Genotype-phenotype correlations in pheochromocytoma and paraganglioma. *Endocr Relat Cancer*. 2019;26:539–550.
58. Hescot S, Curras-Freixes M, Deutschbein T, et al. Prognosis of malignant pheochromocytoma and paraganglioma (MAPP-Prono study): a European network for the study of adrenal tumors retrospective study. *J Clin Endocrinol Metab*. 2019;104:2367–2374.
59. Lee H, Jeong S, Yu Y, et al. Risk of metastatic pheochromocytoma and paraganglioma in SDHx mutation carriers: a systematic review and updated meta-analysis. *J Med Genet*. 2020;57:217–225.
60. Taieb D, Kaliski A, Boedeker CC, et al. Current approaches and recent developments in the management of head and neck paragangliomas. *Endocr Rev*. 2014;35:795–819.
61. Else T, Marvin ML, Everett JN, et al. The clinical phenotype of SDHC-associated hereditary paraganglioma syndrome (PGL3). *J Clin Endocrinol Metab*. 2014;99:E1482–E1486.
62. Eisenhofer G, Lenders JW, Siegert G, et al. Plasma methoxytyramine: a novel biomarker of metastatic pheochromocytoma and paraganglioma in relation to established risk factors of tumour size, location and SDHB mutation status. *Eur J Cancer*. 2012;48:1739–1749.
63. Rao D, Peitzsch M, Prejbisz A, et al. Plasma methoxytyramine: clinical utility with metanephrines for diagnosis of pheochromocytoma and paraganglioma. *Eur J Endocrinol*. 2017;177:103–113.
64. Jha A, de Luna K, Balili CA, et al. Clinical, diagnostic, and treatment characteristics of SDHA-related metastatic pheochromocytoma and paraganglioma. *Front Oncol*. 2019;9:53.
65. Bayley JP, Kunst HP, Cascon A, et al. SDHAF2 mutations in familial and sporadic paraganglioma and phaeochromocytoma. *Lancet Oncol*. 2010;11:366–372.
66. Muth A, Crona J, Gimm O, et al. Genetic testing and surveillance guidelines in hereditary pheochromocytoma and paraganglioma. *J Intern Med*. 2019;285:187–204.
67. Pasini B, McWhinney SR, Bei T, et al. Clinical and molecular genetics of patients with the Carney-Stratakis syndrome and germline mutations of the genes coding for the succinate dehydrogenase subunits SDHB, SDHC, and SDHD. *Eur J Hum Genet*. 2008;16:79–88.

68. Xekouki P, Szarek E, Bullova P, et al. Pituitary adenoma with paraganglioma/pheochromocytoma (3PAs) and succinate dehydrogenase defects in humans and mice. *J Clin Endocrinol Metab*. 2015;100:E710–E719.
69. Qin Y, Yao L, King EE, et al. Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nat Genet*. 2010;42:229–233.
70. Comino-Mendez I, Gracia-Aznarez FJ, Schiavi F, et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet*. 2011;43:663–667.
71. Castro-Vega LJ, Buffet A, De Cubas AA, et al. Germline mutations in FH confer predisposition to malignant pheochromocytomas and paragangliomas. *Hum Mol Genet*. 2014;23:2440–2446.
72. Remacha L, Pirman D, Mahoney CE, et al. Recurrent germline DLST mutations in individuals with multiple pheochromocytomas and paragangliomas. *Am J Hum Genet*. 2019;104:651–664.
73. Yao L, Schiavi F, Cascon A, et al. Spectrum and prevalence of FP/TMEM127 gene mutations in pheochromocytomas and paragangliomas. *JAMA*. 2010;304:2611–2619.
74. Burnichon N, Cascon A, Schiavi F, et al. MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2012;18:2828–2837.
75. Zhuang Z, Yang C, Lorenzo F, et al. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. *N Engl J Med*. 2012;367:922–930.
76. Crona J, Delgado Verdugo A, Maharjan R, et al. Somatic mutations in H-RAS in sporadic pheochromocytoma and paraganglioma identified by exome sequencing. *J Clin Endocrinol Metab*. 2013;98:E1266–E1271.
77. Pacak K, Jochmanova I, Prodanov T, et al. New syndrome of paraganglioma and somatostatinoma associated with polycythemia. *J Clin Oncol*. 2013;31:1690–1698.
78. Job S, Draskovic I, Burnichon N, et al. Telomerase activation and ATRX mutations are independent risk factors for metastatic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2019;25:760–770.
79. Dwight T, Flynn A, Amarasinghe K, et al. TERT structural rearrangements in metastatic pheochromocytomas. *Endocr Relat Cancer*. 2018;25:1–9.
80. Eisenhofer G, Huynh TT, Hiroi M, et al. Understanding catecholamine metabolism as a guide to the biochemical diagnosis of pheochromocytoma. *Rev Endocr Metab Disord*. 2001;2:297–311.
81. Eisenhofer G, Prejbisz A, Peitzsch M, et al. Biochemical diagnosis of chromaffin cell tumors in patients at high and low risk of disease: plasma versus urinary free or deconjugated o-methylated catecholamine metabolites. *Clin Chem*. 2018;64:1646–1656.
82. Madhavaram H, Woollard GA. Interference from Indian diet on the internal standard in a commercial method for the measurement of urinary metanephrines by high-performance liquid chromatography with electrochemical detection. *Ann Clin Biochem*. 2014;51:400–405.
83. Darr R, Pamporaki C, Peitzsch M, et al. Biochemical diagnosis of phaeochromocytoma using plasma-free normetanephrine, metanephrine and methoxytyramine: importance of supine sampling under fasting conditions. *Clin Endocrinol (Oxf)*. 2014;80:478–486.
84. Boyd J, Leung AA, Sadrzadeh HS, et al. A high rate of modestly elevated plasma normetanephrine in a population referred for suspected PPGL when measured in a seated position. *Eur J Endocrinol*. 2019;181:301–309.
85. Peitzsch M, Kaden D, Pamporaki C, et al. Overnight/first-morning urine free metanephrines and methoxytyramine for diagnosis of pheochromocytoma and paraganglioma: is this an option? *Eur J Endocrinol*. 2020;182:499–509.
86. Weismann D, Peitzsch M, Raida A, et al. Measurements of plasma metanephrines by immunoassay vs liquid chromatography with tandem mass spectrometry for diagnosis of pheochromocytoma. *Eur J Endocrinol*. 2015;172:251–260.
87. Eisenhofer G, Peitzsch M, Kaden D, et al. Reference intervals for LC-MS/MS measurements of plasma free, urinary free and urinary acid-hydrolyzed deconjugated normetanephrine, metanephrine and methoxytyramine. *Clin Chim Acta*. 2019;490:46–54.
88. Jiang J, Zhang J, Pang Y, et al. Sino-European differences in the genetic landscape and clinical presentation of pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2020;105:3295–3307.
89. Eisenhofer G, Klink B, Richter S, et al. Metabologenomics of phaeochromocytoma and paraganglioma: an Integrated approach for Personalised biochemical and genetic testing. *Clin Biochem Rev*. 2017;38:69–100.
90. Plouin PF, Amar L, Dekkers OM, et al. European Society of Endocrinology Clinical Practice Guideline for long-term follow-up of patients operated on for a phaeochromocytoma or a paraganglioma. *Eur J Endocrinol*. 2016;174:G1–G10.
91. Buitenwerf E, Berends AMA, van Asselt ADI, et al. Diagnostic accuracy of computed tomography to exclude pheochromocytoma: a systematic review, meta-analysis, and cost analysis. *Mayo Clin Proc*. 2019;94:2040–2052.
92. Lussey-Lepoutre C, Bellucci A, Morin A, et al. In vivo detection of succinate by magnetic resonance spectroscopy as a hallmark of SDHx mutations in paraganglioma. *Clin Cancer Res*. 2016;22:1120–1129.
93. Varoquaux A, le Fur Y, Imperiale A, et al. Magnetic resonance spectroscopy of paragangliomas: new insights into in vivo metabolomics. *Endocr Relat Cancer*. 2015;22:M1–M8.
94. Fottner C, Helisch A, Anlauf M, et al. 6-18F-fluoro-L-dihydroxy phenylalanine positron emission tomography is superior to 123I-metaiodobenzyl-guanidine scintigraphy in the detection of extraadrenal and hereditary pheochromocytomas and paragangliomas: correlation with vesicular monoamine transporter expression. *J Clin Endocrinol Metab*. 2010;95:2800–2810.
95. Taieb D, Jha A, Guerin C, et al. 18F-FDOPA PET/CT imaging of MAX-related pheochromocytoma. *J Clin Endocrinol Metab*. 2018;103:1574–1582.
96. Treglia G, Cocciolillo F, de Waure C, et al. Diagnostic performance of (18)F-dihydroxyphenylalanine positron emission tomography in patients with paraganglioma: a meta-analysis. *Eur J Nucl Med Mol Imaging*. 2012;39:1144–1153.
97. Janssen I, Chen CC, Zhuang Z, et al. Functional imaging signature of patients presenting with polycythemia/paraganglioma syndromes. *J Nucl Med*. 2017;58:1236–1242.
98. Timmers HJ, Chen CC, Carrasquillo JA, et al. Staging and functional characterization of pheochromocytoma and paraganglioma by 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography. *J Natl Cancer Inst*. 2012;104:700–708.
99. Favier J, Briere JJ, Burnichon N, et al. The Warburg effect is genetically determined in inherited pheochromocytomas. *PLoS One*. 2009;4:e7094.
100. Garrigue P, Bodin-Hullin A, Balasse L, et al. The evolving role of succinate in tumor metabolism: an (18)F-FDG-based study. *J Nucl Med*. 2017;58:1749–1755.
101. Kan Y, Zhang S, Wang W, et al. 68)Ga-somatostatin receptor analogs and (18)F-FDG PET/CT in the localization of metastatic pheochromocytomas and paragangliomas with germline mutations: a meta-analysis. *Acta Radiol*. 2018;59:1466–1474.
102. Han S, Suh CH, Woo S, et al. Performance of (68)Ga-DOTA-Conjugated somatostatin receptor-Targeting peptide PET in detection of pheochromocytoma and paraganglioma: a systematic review and metaanalysis. *J Nucl Med*. 2019;60:369–376.
103. Taieb D, Hicks RJ, Hindie E, et al. European association of nuclear medicine practice guideline/society of nuclear medicine and molecular imaging procedure standard 2019 for radionuclide imaging of phaeochromocytoma and paraganglioma. *Eur J Nucl Med Mol Imaging*. 2019;46:2112–2137.
104. Walz MK, Alesina PF, Wenger FA, et al. Laparoscopic and retroperitoneoscopic treatment of pheochromocytomas and retroperitoneal paragangliomas: results of 161 tumors in 126 patients. *World J Surg*. 2006;30:899–908.
105. Shen WT, Grogan R, Vriens M, et al. One hundred two patients with pheochromocytoma treated at a single institution since the introduction of laparoscopic adrenalectomy. *Arch Surg*. 2010;145:893–897.
106. Tiberio GA, Baiocchi GL, Arru L, et al. Prospective randomized comparison of laparoscopic versus open adrenalectomy for sporadic pheochromocytoma. *Surg Endosc*. 2008;22:1435–1439.
107. Stefanidis D, Goldfarb M, Kercher KW, et al. SAGES guidelines for minimally invasive treatment of adrenal pathology. *Surg Endosc*. 2013;27:3960–3980.

108. Alemanno G, Bergamini C, Prosperi P, et al. Adrenalectomy: indications and options for treatment. *Updates Surg*. 2017;69:119–125.
109. Ellis RJ, Patel D, Prodanov T, et al. The presence of SDHB mutations should modify surgical indications for carotid body paragangliomas. *Ann Surg*. 2014;260:158–162.
110. Buitenwerf E, Osinga TE, Timmers H, et al. Efficacy of alpha-blockers on hemodynamic control during pheochromocytoma resection: a randomized controlled trial. *J Clin Endocrinol Metab*. 2020;6:830–838.
111. Liu C, Lv Q, Chen X, et al. Preoperative selective vs non-selective alpha-blockade in PPGL patients undergoing adrenalectomy. *Endocr Connect*. 2017;6:830–838.
112. van der Zee PA, Pheochromocytoma de BA. A review on preoperative treatment with phenoxybenzamine or doxazosin. *Neth J Med*. 2014;72:190–201.
113. Mamilla D, Araque KA, Brofferio A, et al. Postoperative management in patients with pheochromocytoma and paraganglioma. *Cancers (Basel)*. 2019;11:936.
114. Ayala-Ramirez M, Palmer JL, Hofmann MC, et al. Bone metastases and skeletal-related events in patients with malignant pheochromocytoma and sympathetic paraganglioma. *J Clin Endocrinol Metab*. 2013;98:1492–1497.
115. Ayala-Ramirez M, Feng L, Johnson MM, et al. Clinical risk factors for malignancy and overall survival in patients with pheochromocytomas and sympathetic paragangliomas: primary tumor size and primary tumor location as prognostic indicators. *J Clin Endocrinol Metab*. 2011;96:717–725.
116. Amar L, Baudin E, Burnichon N, et al. Succinate dehydrogenase B gene mutations predict survival in patients with malignant pheochromocytomas or paragangliomas. *J Clin Endocrinol Metab*. 2007;92:3822–3828.
117. Plouin PF, Fitzgerald P, Rich T, et al. Metastatic pheochromocytoma and paraganglioma: focus on therapeutics. *Horm Metab Res*. 2012;44:390–399.
118. Hescot S, Leboulleux S, Amar L, et al. One-year progression-free survival of therapy-naive patients with malignant pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2013;98:4006–4012.
119. van Hulsteijn LT, Niemeijer ND, Dekkers OM, et al. 131)I-MIBG therapy for malignant paraganglioma and phaeochromocytoma: systematic review and meta-analysis. *Clin Endocrinol (Oxf)*. 2014;80:487–501.
120. Pryma DA, Chin BB, Noto RB, et al. Efficacy and safety of high-specific-activity (131)I-MIBG therapy in patients with advanced pheochromocytoma or paraganglioma. *J Nucl Med*. 2019;60:623–630.
121. Satapathy S, Mittal BR, Bhansali A. Peptide receptor radionuclide therapy in the management of advanced pheochromocytoma and paraganglioma: a systematic review and meta-analysis. *Clin Endocrinol (Oxf)*. 2019;91:718–727.
122. Nolting S, Grossman A, Pacak K. Metastatic phaeochromocytoma: spinning towards more promising treatment options. *Exp Clin Endocrinol Diabetes*. 2019;127:117–128.
123. Huang H, Abraham J, Hung E, et al. Treatment of malignant pheochromocytoma/paraganglioma with cyclophosphamide, vincristine, and dacarbazine: recommendation from a 22-year follow-up of 18 patients. *Cancer*. 2008;113:2020–2028.
124. Ayala-Ramirez M, Feng L, Habra MA, et al. Clinical benefits of systemic chemotherapy for patients with metastatic pheochromocytomas or sympathetic extra-adrenal paragangliomas: insights from the largest single-institutional experience. *Cancer*. 2012;118:2804–2812.
125. Ayala-Ramirez M, Chougnet CN, Habra MA, et al. Treatment with sunitinib for patients with progressive metastatic pheochromocytomas and sympathetic paragangliomas. *J Clin Endocrinol Metab*. 2012;97:4040–4050.
126. Pamporaki C, Hamplova B, Peitzsch M, et al. Characteristics of pediatric vs adult pheochromocytomas and paragangliomas. *J Clin Endocrinol Metab*. 2017;102:1122–1132.
127. Cascon A, Inglada-Perez L, Comino-Mendez I, et al. Genetics of pheochromocytoma and paraganglioma in Spanish pediatric patients. *Endocr Relat Cancer*. 2013;20:L1. –L6.
128. Bausch B, Wellner U, Bausch D, et al. Long-term prognosis of patients with pediatric pheochromocytoma. *Endocr Relat Cancer*. 2014;21:17–25.
129. de Tersant M, Genere L, Freycon C, et al. Pheochromocytoma and paraganglioma in children and adolescents: experience of the French society of pediatric oncology (SFCE). *J Endocr Soc*. 2020;4:bvaa039.
130. Jochmanova I, Abcede AMT, Guerrero RJS, et al. Clinical characteristics and outcomes of SDHB-related pheochromocytoma and paraganglioma in children and adolescents. *J Cancer Res Clin Oncol*. 2020;146:1051–1063.
131. Waguespack SG, Rich T, Grubbs E, et al. A current review of the etiology, diagnosis, and treatment of pediatric pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2010;95:2023–2037.
132. Olson SW, Yoon S, Baker T, et al. Longitudinal plasma metanephrines preceding pheochromocytoma diagnosis: a retrospective case-control serum repository study. *Eur J Endocrinol*. 2016;174:289–295.
133. Eisenhofer G, Timmers HJ, Lenders JW, et al. Age at diagnosis of pheochromocytoma differs according to catecholamine phenotype and tumor location. *J Clin Endocrinol Metab*. 2011;96:375–384.
134. Ross JH. Pheochromocytoma. Special considerations in children. *Urol Clin North Am*. 2000;27:393–402.
135. Reddy VS, O’Neill Jr JA, Holcomb 3rd GW, et al. Twenty-five-year surgical experience with pheochromocytoma in children. *Am Surg*. 2000;66:1085–1091. ; discussion 1092.
136. Ein SH, Pullerits J, Creighton R, et al. Pediatric pheochromocytoma. A 36-year review. *Pediatr Surg Int*. 1997;12:595–598.
137. Weise M, Merke D, Pacak K, et al. Utility of plasma free metanephrines for detecting childhood pheochromocytoma. *J Clin Endocrinol Metab*. 2002;87:1955–1960.
138. Eisenhofer G, Lattke P, Herberg M, et al. Reference intervals for plasma free metanephrines with an age adjustment for normetanephrine for optimized laboratory testing of phaeochromocytoma. *Ann Clin Biochem*. 2013;50:62–69.
139. Taieb D, Jha A, Treglia G, et al. Molecular imaging and radionuclide therapy of pheochromocytoma and paraganglioma in the era of genomic characterization of disease subgroups. *Endocr Relat Cancer*. 2019;26:R627–R652.
140. Walz MK, Iova LD, Deimel J, et al. Minimally invasive surgery (MIS) in children and adolescents with pheochromocytomas and retroperitoneal paragangliomas: experiences in 42 patients. *World J Surg*. 2018;42:1024–1030.
141. Lenders JWM, Langton K, Langenhuijsen JF, et al. Pheochromocytoma and pregnancy. *Endocrinol Metab Clin North Am*. 2019;48:605–617.
142. Biggar MA, Lennard TW. Systematic review of phaeochromocytoma in pregnancy. *Br J Surg*. 2013;100:182–190.
143. Oliva R, Angelos P, Kaplan E, et al. Pheochromocytoma in pregnancy: a case series and review. *Hypertension*. 2010;55:600–606.
144. van der Weerd K, van Noord C, Loeve M, et al. Endocrinology in Pregnancy: pheochromocytoma in pregnancy: case series and review of literature. *Eur J Endocrinol*. 2017;177:R49–R58.

100

Adrenal Surgery

Barbra S. Miller and Gerard M. Doherty

OUTLINE

Background and History, 1661
Preoperative Evaluation, 1662
  Clinical, 1662
  Familial History and Genetic Testing, 1662
  Biochemical Evaluation, 1663
  Imaging, 1663
  Fine-Needle Aspiration, 1663
Disease-Specific Decision-Making for Surgery, 1664
  Benign Nonfunctional Adrenal Tumors, 1664
  Myelolipomas, 1665
  Adrenal Cysts, 1665
  Biochemically Functional Adrenal Disorders, 1665
  Hypercortisolism, 1665
  Congenital Adrenal Hyperplasia, 1665
  Bilateral Macronodular Adrenal Hyperplasia, 1666
  Primary Aldosteronism, 1666
  Androgen- and Estrogen-Secreting Tumors, 1666
  Pheochromocytoma, 1666
  Adrenocortical Carcinoma, 1667
  Metastatic Disease to the Adrenal Gland, 1667
Surgical Approaches to the Adrenal Gland, 1668
Open Anterior Transabdominal, 1669
  Right, 1669
  Left, 1671
Open Flank–Transabdominal or Retroperitoneal, 1671
Open Posterior (Lumbotomy), 1671
Laparoscopic Anterior Transabdominal, 1671
  Right, 1671
  Left, 1671
Laparoscopic Lateral Transabdominal Adrenalectomy, 1672
  Right, 1672
  Left, 1672
Laparoscopic Hand-Assisted Adrenalectomy, 1673
Lateral Retroperitoneoscopic Adrenalectomy, 1673
Posterior (Dorsal) Retroperitoneoscopic Adrenalectomy, 1673
Robot-Assisted Adrenalectomy, 1674
Thoracoabdominal Approach to Adrenalectomy, 1674
Disease-Specific Considerations and Decision for Extent of Adrenalectomy, 1674
Bilateral Adrenalectomy, 1674
Partial Adrenalectomy, 1675
Disease-Specific Intraoperative Considerations, 1676
  Surgical Complications, 1677
  Intraoperative Complications, 1678
  Postoperative Complications, 1678
  Complications and Operative Approach, 1678
Disease-Specific Postoperative Care and Follow-Up, 1679
  Bilateral Adrenalectomy, 1680
  Nelson Syndrome, 1680
  Additional Considerations, 1680
Conclusion, 1681

BACKGROUND AND HISTORY

Since Perry Sargent performed and reported the first planned adrenalectomy in 1914,[1] surgeons have developed multiple operative approaches. Early on, clinicians rarely knew if a large tumor causing local symptoms in the abdomen and retroperitoneum was of adrenal origin. Steady progress in the field and discovery of adrenal steroids (cortisol by Kendall and Reichstein in the 1940s,[2] aldosterone by Conn,[3] androgens and estrogens and catecholamines[4]) led to advances in surgical technique to account for the unique needs to address hormone excess and neoplasia. In 1926, Cesar Roux removed a pheochromocytoma by a flank approach. Seven months later, Charles Mayo did the same in the United States. In 1931, Broster described adrenalectomy by a thoracoabdominal approach. In 1936, Young described an open posterior approach, removing the eleventh and/or twelfth rib, providing access to the retroperitoneum and adrenal gland without having to traverse the contents of the peritoneal cavity or mobilize the liver or spleen. It was through a bilateral open posterior approach that Baum performed the first operation for primary aldosteronism (Conn syndrome).[5] The open posterior approach grew in popularity for this particular indication as it allowed for nearly direct access to relatively normal-size adrenal glands and a quicker immediate recovery, return of bowel function, and hospital discharge compared with an anterior approach.[6] Knowledge regarding malignancies affected by estrogens and androgens, such as breast and prostate cancer, led to favoring a posterior approach to facilitate removal of bilateral adrenal glands. This was one way to minimize circulating levels of androgens and estrogens prior to the development of medical therapies for this purpose.[7,8] Other modifications of approaches to the adrenal gland were reported in ensuing years.[9,10] More recently, in 1993, Makuuchi described an incision extending from just inferior to the xiphoid to several centimeters above the umbilicus and then extending laterally with the option for coursing between the ninth and tenth ribs to provide access to the pleural cavity if desired.[11] Modifications and extensions of all of these approaches have been described over the years.

In 1992, a major change in approach to adrenalectomy occurred when Gagner and colleagues in Canada and Higashihara in Japan reported adrenalectomy by a laparoscopic approach.[12-16] Since then, additional minimally invasive approaches have been described, including posterior retroperitoneoscopic, hand-assisted laparoscopic, and robot-assisted approaches. Technical aspects of these approaches continue to be refined. Currently, a minimally invasive approach is the gold standard for excision of the majority of benign adrenal lesions.[17,18]

With a significant increase in the number of adrenal abnormalities identified (4%–10% of patients undergoing imaging studies) due to the sharp rise in imaging of the chest and abdomen, proper evaluation and management has never been more important.[19] Additionally, it is recognized that the spectrum of adrenal hormone excess is much greater than previously shown.[20] Even with mild autonomous hormone secretion, detrimental effects result in a significantly increased risk of morbidity and mortality.[21,22] Early identification and treatment of hormone excess will become a focus for clinicians and researchers in the next decade. The most competent adrenal surgeons have a comprehensive knowledge of adrenal physiology and the biology and genetics of the many different adrenal disease processes, and are facile with numerous operative approaches to the adrenal gland.

PREOPERATIVE EVALUATION

A thorough history and physical examination, along with biochemical and imaging evaluation, is critical to correctly diagnose and manage patients with adrenal disorders.[23,24]

Clinical

Patients should be questioned about and examined for evidence of signs and symptoms suggestive of hormone excess, although a significant proportion with biochemical evidence of hormone excess or autonomous production will not have physical signs or symptoms (Table 100.1). In patients with possible adrenal cancer, either adrenocortical carcinoma or malignant pheochromocytoma, evidence consistent with possible metastatic disease should be sought prior to any surgical intervention. Clues to adrenal malignancy may include rapid or progressive weight gain or weight loss and lower extremity swelling due to occlusive tumor thrombus in the vena cava and potentially resulting in Budd–Chiari syndrome. Rare patients with adrenocortical carcinoma have an enlarged supraclavicular (Virchow's) node. Metastatic disease involving the vertebral bodies or impinging on the spinal cord or nerve roots may cause symptomatic fractures or neurologic impairment. Patients with large tumor burden and excess cortisol may have severe weakness, extreme hypertension, hypokalemia, or hyperglycemia. Lymphoma of the adrenal gland(s) may lead to localized symptoms or "B" symptoms (fever, weight loss, etc.). Signs of adrenal insufficiency may be evident if a malignant process (usually not of adrenocortical origin) obliterates the adrenal cortex.[25] This may also occur with infectious and autoimmune processes.

Thorough investigation of the patient's past medical history sheds light on the possibility of malignancy metastatic to the adrenal gland. Prior history of malignancy, family history, and new symptoms suggestive of malignancy should be reviewed, including pursuit of routine health maintenance investigations. While any malignancy theoretically can metastasize to the adrenal glands, malignancies that most commonly metastasize to the adrenal glands include melanoma and primary tumors of the lung, breast, and kidney. Colorectal cancer is less common, although reports are increasing.[26] Lymphoma may arise within the adrenal glands or metastasize to the adrenal glands.[27]

Infectious diseases may localize within the adrenal glands, the most common being tuberculosis, histoplasmosis, cryptococcosis, paracoccidioidomycosis, and cytomegalovirus. Abnormal response to synthetic adrenocorticotropic hormone (ACTH) stimulation tests or frank adrenal insufficiency dependent on the severity of human immunodeficiency virus/acquired immunodeficiency syndrome ranges from 14% to 54%.[28] On rare occasions, surgeons may be consulted to obtain tissue for confirmation when the adrenal gland is inaccessible to interventional image-guided biopsy, and no other sites exist or are amenable to biopsy.

TABLE 100.1 **Clinical Signs and Symptoms of Hormone Excess, Cancer, and Adrenal Insufficiency**

Hormone Excess or Disease Process	Signs and Symptoms
Aldosterone	Hypertension, hypokalemia, hypernatremia, edema
Cortisol	Moon facies, facial plethora, buffalo hump, muscular weakness, sarcopenia, fatigue, striae, hyperglycemia, weight gain, increased visceral adiposity, hypokalemia, hirsutism, hair loss, easy bruising, hypercoagulability, emotional lability
Androgens and estrogens	Change in secondary sexual characteristics
Catecholamines	Palpitations, hypertension, dyspnea, sudden-onset headaches, diaphoresis, hyperglycemia, pallor, tremor, anxiety
Adrenal cancer	Hormone excess, rapid weight gain or weight loss, pain, localized symptoms, lower extremity edema, hyperglycemia, hypertension, hair loss.
Adrenal lymphoma	Pain, adrenal insufficiency, fever, weight loss, and/or other "B" symptoms
Adrenal insufficiency due to gland obliteration	Fatigue, weight loss, nausea, diarrhea, vomiting, hyperpigmentation, hypotension, hyperkalemia, hyponatremia, weakness

Familial History and Genetic Testing

The discovery of genetic mutations leading to adrenal disorders has driven specific biochemical (adrenal vein sampling) and radiologic testing (NP-59, metaiodobenzylguanidine, and now DOTATATE–positron emission tomography [PET] computed tomography [CT]) to identify laterality of disease and inform surgical approach. A thorough family history should be taken at the time of the initial consultation. Genetic testing should be guided by risk, and in some cases, all patients with a certain type of tumor should be referred to a genetic counselor for consideration of genetic testing for a germline mutation, as in the case of pheochromocytoma and adrenocortical carcinoma.[29-32] In some situations, such as patients with pheochromocytoma, knowledge of disease driven by a germline mutation impacts concern for malignancy and surgical decision-making. In those with benign disease driven by a germline mutation (e.g., multiple endocrine neoplasia [MEN] 2, von Hippel–Lindau, neurofibromatosis), the surgeon should consider partial adrenalectomy to preserve cortical function for as long as possible. Each case is different, and the decision regarding extent of surgery must be individualized to the patient.

In addition to testing for germline mutations through bloodwork, investigation of mutations within the tumor itself can also be helpful. In those without a personal or family history concerning for an inherited mutation, testing for mutations within a resected tumor can provide valuable information. Certain mutations, such as those in *MSH* and *MLH* genes, may suggest Lynch syndrome and lead to testing for a germline mutation after surgery.[31] If a somatic mutation alone is noted on final pathology after surgery (given that biopsy preoperatively is to be avoided in general), that knowledge may inform future treatment decisions and therapies. Knowledge of the particular type of *SDH* somatic mutation in patients with pheochromocytomas or paragangliomas can provide clinicians with information as to the potential aggressiveness of a tumor and potential for recurrence and metastasis.[33] As adrenal cancers and the mutations driving them may be different in different areas of a tumor or change over time, repeat tissue sampling over the course of disease may be warranted to guide therapy. Some institutions accomplish this through in-house testing. A number of commercial offerings are also available.

Biochemical Evaluation

Assessment of adrenal hormone production is performed in a consistent and comprehensive fashion. Selection of individual tests based on a patient's clinical presentation can lead to misdiagnosis. In nearly 25% of cases of hormone excess, patients have subclinical disease. With a new understanding of the expanded spectrum of adrenal hormone dysregulation, especially in patients with primary aldosteronism and mild autonomous cortisol secretion, this percentage is likely even higher. Similarly, the suboptimal implementation of screening guidelines also contributes to the underestimation of the prevalence of adrenal hormone dysregulation.

Table 100.2 lists the steroid hormones and biochemical tests routinely ordered to assess hypothalamic-pituitary-adrenal (HPA) axis integrity and adrenal hormone production, including catecholamines from the medulla. Additional tests may be necessary, especially in the case of an indeterminate adrenal nodule possibly representing adrenocortical carcinoma (ACC). Intermediaries along the steroidogenesis pathway can be helpful in predicting malignancy.[34] Along these lines, the breakdown products of adrenal steroids and their intermediaries may also be measured in urine.[35] Urine steroid patterns have been found to aid in differentiating between benign and malignant adrenocortical tumors. Preoperative identification of hormone excess, including intermediaries, is important for postoperative care, and in patients with malignancies of adrenal origin may be used as tumor markers for recurrence or extent of disease.[36]

Imaging

In addition to biochemical data to understand the presence or absence of a functional disorder, the morphology of the adrenal glands is important in understanding adrenal pathology. Glands may be diminutive, enlarged, hyperplastic, have single or multiple nodules, contain cysts, contain calcifications, and/or show evidence of hemorrhage. While there is significant overlap, certain disease processes have certain radiologic characteristics that aid clinicians with diagnosis. Most commonly, adrenal protocol CT provides the most information pertinent to a surgeon[37-39]; however, adrenal protocol magnetic resonance imaging (MRI) also provides useful information.[40,41] From the standpoint of surgical planning, MRI is most often utilized to assess for evidence of vascular invasion of structures adjacent to adrenal cancers and intravenous tumor thrombus, and has greater sensitivity for identifying metastases to the liver. (^{18}F)-fluorodeoxyglucose (FDG)-PET/CT provides additional information in cases of adrenal nodules classified as "indeterminate" by CT or MRI.[42-45] Standard uptake value ratios of the tumor activity to the liver background and/or blood pool activity may be informative. Tumors metastatic to the adrenal gland have useful radiologic FDG avidity patterns; for example, renal cell cancers often have little FDG avidity. ^{11}C-metomidate PET is an additional study available outside the United States that may be utilized in differentiating benign from malignant adrenal tumors of cortical origin.[46,47]

Pheochromocytomas are most often heterogeneous or indeterminate based on intravenous contrast washout patterns on CT scan. Biochemical testing typically leads to the correct diagnosis and avoids the need to pursue additional imaging. FDG-PET/CT can reveal increased FDG uptake, and therefore is not particularly useful for distinguishing between a pheochromocytoma and malignant cortical tumor, or for distinguishing between benign and malignant pheochromocytomas. However, FDG-PET/CT may reveal metastatic disease, which would indicate malignancy. More recently, ^{68}Gallium-DOTATATE-CT has been shown to be the most sensitive of functional imaging studies in identifying tumors of neuroendocrine origin and determining extent of disease.[48-50]

TABLE 100.2 Biochemical evaluation

INITIAL ADRENAL BIOCHEMICAL EVALUATION

Routine	Consider
Complete blood count and metabolic panel	17-OH progesterone
ACTH	17-OH pregnenolone
Aldosterone	11-deoxycortisol
Renin	Androstenedione
1-mg DST with 8:00 AM ACTH, cortisol, dexamethasone level	Estradiol
24-hour urine free cortisol and creatinine	FSH, LH
DHEA-S	Progesterone
Testosterone (total or bioavailable)	Dopamine
Plasma and/or 24-hour urine metanephrine and normetanephrine levels	Chromogranin A

ACTH, Adrenocorticotropic hormone; *DST*, dexamethasone suppression test; *DHEA-S*, dehydroepiandrosterone sulfate; *FSH*, follicle-stimulating hormone; *LH*, luteinizing hormone.

Fine-Needle Aspiration

There is a very limited role for fine-needle aspiration (FNA) of adrenal nodules.[23,51] The ability to discriminate between benign adrenal tissue and ACC is poor, and the undesirable possibility of seeding the needle tract with tumor has been reported. Treatment should be based on biochemical and radiologic imaging characteristics. FNA should only be pursued if it will change management. An indeterminate mass without evidence of metastatic disease will be resected almost regardless of FNA results. The most common circumstances for which FNA may be considered in patients with adrenal masses are: (1) patients who are poor operative candidates and need a diagnosis to guide additional treatment; (2) patients with suspected metastatic disease from a nonadrenal primary tumor in which the adrenal mass is the most easily accessible metastatic tumor deposit; (3) suspected adrenal lymphoma or abnormalities due to an infectious etiology (tuberculosis, histoplasmosis), which are often bilateral; (4) patients with presumed unresectable ACC and without evidence of hormone excess for whom tissue confirmation is necessary to guide neoadjuvant or palliative chemotherapy; and (5) if sarcoma is in the differential diagnosis, and neoadjuvant therapy would be pursued prior to resection. FNA in other situations is not particularly helpful. If FNA of an adrenal mass is pursued, pheochromocytoma must be ruled out before biopsy, as hypertensive crisis may result.[52] Similarly, retroperitoneal lesions along the axis of the body (e.g., along the spine, aorta, vena cava) that could be a paraganglioma should not be biopsied prior to biochemical assessment for elevated catecholamine production.[53]

TABLE 100.3 **Phases of Care in the Perioperative Period**

Preoperative	Intraoperative	Postoperative	Postdischarge
Medical control: blood pressure, hormone excess, electrolytes, glucose	Antibiotics, glucose control, blood pressure control, corticosteroids if indicated	For large incisions, consider antibiotics for 24 h; glucose; blood pressure; corticosteroids if indicated	Encourage mobilization, physical or occupational therapy as needed
Nutrition: emphasize protein intake	Care in positioning: skin, skeletal, nerves	Nutrition: emphasize protein	Nutrition: emphasize protein
Prehabilitation: Emphasize core muscle conditioning, walking, deep breathing; physical or occupational therapy if necessary	Complete resection, minimize trauma to surrounding structures (spleen, liver), intraoperative ultrasonography to help localize gland within retroperitoneal fat	Early mobilization, physical or occupational therapy as needed	Wean or reinstitute antihypertensive and diabetic medications as indicated, manage electrolytes, watch renal function
Select operative approach based on patient disease, surgical history, comorbidities, and anatomic factors	Incision care, minimize hematoma and fluid collections	Provide information before discharge regarding signs of cortisol withdrawal, Addison disease, how to obtain medical alert bracelet	Review pathology
Disease-Specific Considerations			
Cancer: staging, evaluate for metastatic disease **PPG**: genetic testing, preoperative blockade, and volume expansion	**Cortisol excess**: consider additional mechanisms for fascial closure, prevent wound infection **PPG**: communication with Anesthesia service during case to minimize hypertensive episodes, prior to clipping adrenal vein, volume resuscitation, management of hypotension	**Aldosterone**: hold BP medications, reinstitute if warranted, watch potassium and creatinine **PPG**: hold BP medications, reinstitute if needed, watch for hypoglycemia **Cortisol**: test for adrenal insufficiency, determine need for steroids	**Cancer**: genetic testing if ACC, determine need for additional treatment, develop survivorship plan **Aldosterone**: assess degree of clinical and biochemical resolution over first year **Cortisol**: wean corticosteroids as indicated if unilateral adrenalectomy, ensure corticosteroids and fludrocortisone if bilateral adrenalectomy **PPG**: genetic testing if not already completed, annual fractionated metanephrine levels

PPG, Pheochromocytoma paraganglioma; *BP*, blood pressure; *ACC*, adrenocortical carcinoma.

KEY POINTS

- A thorough understanding of relevant adrenal anatomy, physiology, and interpretation of radiologic findings, is critical to facilitate sound decision making regarding adrenalectomy.

DISEASE-SPECIFIC DECISION-MAKING FOR SURGERY

The decision to proceed with adrenalectomy is based on imaging characteristics, size, and evidence of hormone excess. Once a decision for surgery is made, a surgical approach can be chosen. Considerations for the preoperative care of patients are listed in Table 100.3.

Benign Nonfunctional Adrenal Tumors

In general, benign biochemically nonfunctional tumors of the adrenal gland, including enlarged hyperplastic glands, do not need to be removed. The size of the tumor alone should not dictate the need for adrenalectomy if benign and nonfunctional. In the past, tumor size greater than 4 or 6 cm alone was used as a cutoff, given an increase in rate of malignancy[24]; however, this view has changed over time, and currently imaging characteristics (taking into account washout percentages and other imaging characteristics) are considered in combination with size.[23] A purely nonfunctional tumor with benign imaging characteristics that is larger than 4 to 6 cm does not necessarily need resection based on size alone. A decision to pursue resection for benign-appearing nonfunctional tumors reaching 6 to 10 cm may be made, because tumors of this size are still potentially amenable to a minimally invasive approach. With increasing size, an open approach is more likely needed due to difficulty manipulating the tumor, identifying and preserving adjacent structures, avoiding disruption of the tumor capsule and rupture, and simply needing to extract the tumor from the peritoneal cavity. Morcellation of an adrenal tumor is contraindicated, as it obviates the possibility of adequate pathologic examination. Adrenal tumors can grow quite large (10–15 cm) prior to patients developing local symptoms. The true incidence of localized symptoms attributable to adrenal tumors based on size is not well known. These symptoms may not be apparent to the patient until after the tumor is removed and the patient has recovered from surgery.

Patient age and comorbidities should also be taken into consideration to aid in the decision-making process regarding resection of benign nonfunctional tumors. In young patients, continued follow-up over many years adds to medical costs, and the patient may prefer resection sooner rather than later, especially while still healthy and with few comorbidities. The issue of follow-up and nonoperative management of benign nonfunctional tumors, including hyperplastic glands without tumors, has been investigated by many groups. Achieving consensus regarding follow-up is challenging given the short follow-up period of many studies, when many patients are diagnosed with tumors in midlife. The most recent guideline recommendations from Europe suggest that a single follow-up imaging study is sufficient.[23] Two studies specifically addressed follow-up of adrenal tumors incorporating risk of malignancy, age, and cost of follow-up. Suh and colleagues performed a cost–utility analysis examining optimal follow-up from a cost standpoint.[53] A strategy of one-time follow-up in patients with "nonfunctional," radiologically benign tumors was superior to other strategies in patients with adrenal tumors smaller than 4 cm. In patients over 60 years old, one-time or no follow-up was found to be most cost-effective in this model. Limitations to this include lack of long-term follow-up over decades, as the assumptions in this study were based on data with follow-up for progression of 5 years or less. It is also unclear how many patients had full biochemical evaluation, including 1-mg dexamethasone suppression test (DST), to assess for hormone dysregulation. It is known that tumors larger than 2.5 cm can begin to autonomously produce small amounts of cortisol. It is becoming more apparent that even mild autonomous cortisol excess significantly increases risk of cardiovascular morbidity, infection, and death. While in years past it was unclear that there was an adenoma-to-carcinoma sequence in adrenal tumors similar to that associated with other types of malignancies, it has been documented within the past decade that adrenocortical carcinomas can arise from within adenomas.[54] The incidence of this is unknown. It is unclear if this mechanism dominates development of ACC or if ACC arises *de novo*, and what proportion of all ACCs it contributes to. While nonoperative management is most common, it may be reasonable to consider resection if the patient is younger to prevent the question of need for follow-up over a long period of time. If a neoplasm is nearing the limit for safe minimally invasive resection, it may also be reasonable to remove these tumors to prevent the need for a more invasive open resection if growth occurs. Wang et al. examined the cost-effectiveness of surgery for nonfunctional adrenal tumors based on size.[55] This study incorporated risk of malignancy and did not solely examine benign tumors. The study also did not stratify by imaging findings in terms of benign or indeterminate characteristics. They found that surgical resection based on size alone was cost-effective in patients with tumors greater than or equal to 4 cm or in patients under 65 years old. Whether or not this strategy holds for those with radiologically benign tumors is unknown.

Myelolipomas

The overwhelming majority of myelolipomas are benign, nonfunctional tumors. In rare instances, they may secrete excess adrenal hormone (most commonly cortisol).[56] For the most part, myelolipomas do not need to be removed if they are small and biochemically nonfunctional and do not cause bothersome local symptoms. Rare cases of malignant transformation have occurred, but it is not clear that at a population level this warrants changing recommendations for long-term follow-up. Growth may occur over time, and resection may be recommended, but as long as they remain benign and nonfunctional, observation is most often pursued. It is difficult to predict which neoplasms will grow and which will remain stable over time.

Adrenal Cysts

Patients with adrenal cysts need to be considered in a slightly different manner. Pure cysts without evidence of a solid component can often be observed. If small cysts less than 4 to 6 cm are growing slowly, and all other imaging characteristics are otherwise benign, aspiration with injection of ethanol can be considered for ablation. The long-term durability of this approach is approximately 30%, with decreased success rates as cyst size increases.[57] Cysts with a solid component harbor a small risk of malignancy, whereas in those with cystic areas within a larger, more solid tumor, the risk of malignancy is even higher because fluid-filled portions may be due to tumor necrosis or hemorrhage. Unlike pure cysts, the cystic portions of heterogeneous adrenal masses should not undergo aspiration of the fluid-filled portion of the tumor prior to or during surgery to decrease the tumor size, as this violates oncologic principles. Pheochromocytomas may show evidence of necrosis and fluid within the tumor. Biochemical evaluation will usually secure the diagnosis in these patients. Aspiration of any sort in these patients should be avoided as previously discussed.

Biochemically Functional Adrenal Disorders

Those patients with evidence of overt hormone excess due to a unilateral adrenal tumor, hyperplastic gland, or pituitary disorder are most often offered surgery rather than medical management. In patients with mild autonomous hormone secretion, the decision for surgery is more individualized and often dependent on hormone levels, morbidities, and expected duration of survival.

Hypercortisolism

The management of those with biochemical evidence of subclinical hormone excess can be more difficult than overt cases from a decision-making standpoint. Most patients with subclinical hypercortisolism are diagnosed by a DST.[58,59] For those with subsequent morning cortisol levels greater than 5 μg/dL, resection is usually pursued. For those with morning cortisol levels between 3 and 5 μg/dL, resection may be pursued in some.[60,62] Often, clinicians will assess sequelae of disease in this range to assist in the decision to pursue resection; however, there is compelling evidence from a number of studies revealing significantly increased risk of cardiovascular morbidity and mortality in these patients without classic signs of cortisol excess.[21,22] In those with post-DST morning cortisol levels 1.8 to 3 μg/dL, surveillance is most often pursued, but again, evidence of increased morbidity and mortality has also been described for this group of patients. The decision for operation in patients with dysregulated cortisol secretion should be made in the context of the severity of other potential disease processes that may be attributable to hypercortisolism, such as hypertension, diabetes, weight gain, and bone density loss. Studies reporting outcomes of patients with subclinical hypercortisolism do show benefit after adrenalectomy; however, many of these studies have methodologic flaws, and accurately differentiating patients with subclinical disease from those with overt hypercortisolism can be difficult.[62-65] For those with hypercortisolism due to bilateral excess adrenal production (ACTH-dependent or -independent disease and other types of hyperplasia), the decision for unilateral (starting with the largest adrenal gland) versus bilateral adrenalectomy should be made jointly by a multidisciplinary team including an endocrinologist.

For patients with hypercortisolism, exercise and consumption of a high-protein diet should be encouraged before surgery, given muscle deterioration due to the disease process. If time permits, several months of preoperative treatment with a cortisol-suppressing medication can be helpful in better controlling the quantity of cortisol excess. During surgery, patients with overt hypercortisolism should be given stress-dose steroids. If undergoing bilateral adrenalectomy, stress-dose steroids can be given after resection of the first adrenal gland. Alternatively, these may be held until several hours after the operation is completed to allow a decrease in cortisol levels to a more reasonable level in those with severe excess cortisol. A standard stress-dose steroid regimen is 50 to 100 mg of intravenous hydrocortisone every 8 hours. This is then tapered over several days to a defined dose to be given at the time of discharge in conjunction with the input of the managing endocrinologist while awaiting response of the HPA axis and return of autonomous function of the contralateral adrenal gland. Those with suspected subclinical disease may forego intraoperative stress-dose steroids if testing of the HPA axis is planned for the morning after surgery in the form of an early morning cortisol level or cosyntropin stimulation test.

Congenital Adrenal Hyperplasia

Several mutations affecting the presence and function of enzymes in the steroid synthesis pathway lead to congenital adrenal hyperplasia

(CAH).[66] Three phenotypes exist (two classical and one nonclassical), and patients have varying degrees of decreased cortisol, salt-wasting, and hyperandrogenism. Patients are most often identified in childhood but may present as adults. Depending on the severity, glucocorticoids and mineralocorticoids are often provided. Glucocorticoid administration also limits excessive secretion of both corticotropin-releasing hormone and ACTH, thereby minimizing the production of adrenal-derived androgens. These patients develop myelolipomas over time that can become quite large, necessitating resection. Growth can be asymmetric, but most often bilateral adrenalectomy is pursued. These patients are usually already on steroids, and so are used to taking additional medication, eliminating this as a reason to avoid bilateral adrenalectomy. The decision for and timing of surgery often depends on size of the myelolipomas that develop. Timing for removal may also depend on local symptoms, as well as the ability to perform a minimally invasive or hand-assisted approach rather than an open approach.

Bilateral Macronodular Adrenal Hyperplasia

As bilateral macronodular hyperplasia is often driven by a mutation leading to constitutive activation of a G protein–coupled receptor leading to cortisol excess, any remnant tissue will have the same mutation, leading to the same issue of autonomous cortisol secretion.[67] A common approach to surgery in these patients is to remove one adrenal gland and then determine if the remaining degree of excess cortisol production warrants contralateral adrenalectomy. Given that even small amounts of cortisol excess from dysregulated production are associated with significant risk for cardiovascular and neurologic morbidity, as well as death, bilateral adrenalectomy is pursued by some groups.[68,69,70] While cure of hypercortisolism is ensured, assuming all adrenocortical tissue is removed, this results in the need for lifelong steroid replacement. More recently, pursuit of unilateral adrenalectomy with contralateral subtotal adrenalectomy leaving a remnant of at least 30% up to a normal size gland has been reported. Cure of hypercortisolism without recurrence and without need for supplemental steroids was achieved in all patients who had undergone bilateral or subtotal adrenalectomy with a follow-up period of 40 months. Given that tissue remains, and the underlying cause is mediated by a mutation, the long-term durability of this approach for patients to remain eucortisolemic awaits further study. If any cortical tissue remains, persistent autonomous cortisol production will result, but the degree of hypercortisolism is difficult to predict based on the amount of tissue; therefore, understanding how much tissue to remove is also difficult. Residual tissue after subtotal adrenalectomy may hypertrophy and result in clinically relevant sequelae. Should patients develop recurrent hypercortisolism, they may be treated with medication until such time as completion adrenalectomy can be pursued.

Primary Aldosteronism

Patients with primary aldosteronism need further evaluation to determine whether surgical resection is appropriate. Traditionally, an aldosterone:renin ratio greater than 20 suggests the diagnosis in a screening setting (e.g., screening in a hypertension clinic), although the key to diagnosis remains a suppressed renin level. Recent data suggest that current guidelines for testing fail to identify many cases, and that cutoffs for absolute aldosterone levels may need to be lowered,[20] as our understanding of the true spectrum of disease has increased significantly. Most cases undergoing evaluation and surgery at this time can now be considered to be on the severe end of the spectrum, regardless of potassium level or number of hypertension medications. If hypertension and hypokalemia are present, a patient has likely had longstanding disease. Also recently described are aldosterone-producing cell clusters, which develop with age and autonomously secrete aldosterone.[71] Even with mild disease, risk of cardiovascular and renal morbidity and mortality is increased, and this occurs even prior to development of hypertension.[20] The diagnosis is mostly dependent on renin suppression when no other reason exists. Certain medications should be held before testing, as they can affect the renin level. Confirmatory tests such as saline or captopril suppression tests may be pursued. The Endocrine Society published guidelines detailing the biochemical evaluation of these patients.[72] CT imaging has been found to be poor in terms of specificity for unilateral versus bilateral excess aldosterone secretion from the adrenal glands. Specificity averages approximately 40% to 50%.[73] Therefore, once a clear biochemical diagnosis has been made, adrenal vein sampling should be pursued in most patients to determine the unilateral or bilateral nature of the excess hormone production.[72,73] Understanding how to interpret these studies is critical, as is having an interventional radiologist with high success of entering the left adrenal vein and especially the right adrenal vein. Aside from those with hormone excess, the decision for operation should also be based on the physical status of the patient and existing comorbidities. Some patients can be easily medically managed, and this may be more appropriate in patients with significant comorbidities. Scoring systems have been developed to estimate the likelihood of complete clinical cure; however, the generalizability to other cohorts has been questioned. Additionally, it is not simply the improvement in blood pressure and heart rate that is important. A reduction in aldosterone is also extremely important, and has only recently been recognized.[20] Inappropriate activation of the mineralocorticoid receptor leads to adverse effects in multiple organ systems, increasing cardiovascular and neurologic morbidity and mortality with or without hypertension.

For those with primary aldosteronism, improved control of hypertension is often best achieved with spironolactone or eplerenone. Creatinine, renin, sodium and potassium levels should be closely followed while titrating these medications. In addition to potassium supplementation, hypokalemia (which occurs in ~50% of patients) may also be controlled in part with amiloride. If patients are hypokalemic in the immediate preoperative setting, lower doses than normal should be given as replacement if needed, due to the potential for rebound hyperkalemia in the postoperative setting.

Androgen- and Estrogen-Secreting Tumors

No specific preoperative preparation is required prior to removal of these tumors. Of adrenocortical carcinomas secreting hormones, testosterone is the second most common type after cortisol. An adrenal mass with indeterminate imaging characteristics with evidence of excess androgen secretion is presumed to be ACC until assessed by a pathologist. In rare circumstances, large, indeterminate androgen-secreting adrenal masses are found to be benign. Androgen-secreting tumors may be found in the pediatric population, and may be found at an earlier stage if precocious puberty or ambiguous secondary sexual characteristics are noted.

Pheochromocytoma

Preoperative preparation for resection of pheochromocytoma is critical. Due to improved preoperative preparation, intraoperative anesthetic management, and postoperative care, outcomes have dramatically improved since the mid-20th century, when mortality rates were close to 50%.

Patients should be given detailed information regarding the risks of resection of pheochromocytoma and advised that adherence to instructions for blockade and hydration are critically important to follow. Communication with the physician managing this aspect of care is also critical. Blockade may begin when the patient is diagnosed, as it is likely they have been referred for hypertension. For those without

TABLE 100.4 Medications for Preoperative Blockade of Pheochromocytoma and Paraganglioma

Drug	Starting Dose
Alpha Blockade	
Phenoxybenzamine	10 mg BID or TID
Doxazosin	2 mg QHS
Beta Blockade	
Propranolol	20 mg TID
Atenolol	25 mg QD
Calcium Channel Blockade	
Nicardipine	20 mg TID
Nifedipine	30 mg QD (shorter half-life)
Amlodipine	5 mg QD
Catecholamine Synthesis Inhibitor	
Metyrosine	250 mg QID

BID, Twice a day; *TID*, three times a day; *QD*, daily; *QHS*, nightly; *QID*, four times a day.

hypertension, blockade may be delayed until closer to surgery. Rarely, patients may present with dopamine-secreting pheochromocytomas or paragangliomas. If secreting dopamine only, preoperative blockade is not necessary.[74,75] Medications that can be utilized for blockade are listed in Table 100.4.[32] A gradual shift from phenoxybenzamine (PBZ) to doxazosin has occurred over time. In part, this is due to the high cost of PBZ, lack of coverage by some insurance companies, lack of stock in some pharmacies, and more severe side effects. However, many surgeons continue to prefer alpha blockade with PBZ in patients with highly active tumors and severe hypertension. With alpha blockade, patients should liberally hydrate to expand the intravascular volume as vasodilation occurs. Patients should understand that nasal stuffiness, fatigue, and occasional dizziness are expected. Titration of blockade should not stop when these symptoms occur, but should instead continue until an appropriate goal blood pressure is reached. As alpha blockade and vasodilation progress, the heart rate may increase. Fluid administration should decrease the heart rate as euvolemia is reached. If volume does not decrease the heart rate, and additional blood pressure control is needed, a beta-blocker can be instituted. In patients with minimal to no hypertension, a calcium channel blocker is an alternative. This has even fewer side effects than selective alpha blockers, but does not cause any significant vasodilation. The effect on blood pressure is less than that of alpha- and beta-blockers. Intraoperative hypotension after ligation of the adrenal vein can be greater, as vasodilation occurs all at once in the operating room rather than predominantly over a slower period of time in the preoperative setting. Metyrosine is a tyrosine hydroxylase inhibitor that decreases catecholamine production. While its routine use has been described, it is seldom necessary.[76,77] The onset of effect is slow, given that bodily stores of catecholamines need to be depleted before effect is seen. Metyrosine is not commonly stocked in pharmacies and is expensive. It is most often utilized in patients with extremely difficult-to-control pheochromocytomas that require more control than can be gained with standard alpha, beta, and calcium channel blockade.

Blockade goals for systolic pressures below 120 mm Hg are ideal, but may be higher dependent on the situation, patient comorbidities necessitating a higher goal perfusion pressure, and/or degree of catecholamine secretion. Preparation for the operating room should begin no later than 1 to 2 weeks prior to surgery. In those with other comorbidities, especially cardiac, pulmonary and/or renal, where fluid shifts and changes in blood pressure could cause issues, slow blockade for a longer period of time is useful. Some literature exists suggesting that preoperative blockade of patients with pheochromocytomas is not necessary in some patients. The risk–benefit ratio favors routine blockade and is safest for the patient.[32]

Hyperglycemia due to catecholamine excess may be present in some patients with pheochromocytoma paraganglioma (PPG). Catecholamine stimulation of beta-receptors leads to gluconeogenesis, and stimulation of α2 receptors results in decreased release of insulin. Insulin resistance occurs in musculoskeletal tissue. The phenomenon is more often associated with larger tumors and those with greater production of epinephrine compared with norepinephrine.[78] Epinephrine-secreting pheochromocytomas may be more likely to induce glucose intolerance or frank diabetes mellitus due to a preferential affinity for the β receptor. Postoperatively, patients may experience hypoglycemia as excess catecholamine production ceases and rebound hyperinsulinemia results from a reduction in α2 receptor stimulation.[79,80] Improved peripheral insulin sensitivity with increased glucose use by skeletal muscle further contributes to hypoglycemia.[81] Postoperative hypoglycemia after resection of pheochromocytoma was reported in 4.2% of patients in one study.[78] The hypoglycemia is temporary, treated with intravenous dextrose as needed if more than oral consumption of food is needed, and usually resolves in 1 to 2 days at the most.

Adrenocortical Carcinoma

Up to 50% of patients with ACC have evidence of hormone excess. This is likely an underestimation given a high frequency of inadequate biochemical evaluation. Hormone excess should be managed as best as possible to optimize safety for surgical resection and facilitate postoperative recovery.[83] For those with suspected or obvious ACC, a chest CT should be obtained as part of the routine preoperative evaluation to evaluate for metastatic disease.[30] In those with metastatic disease, resection is usually delayed in favor of chemotherapy and mitotane, with a select few being able to undergo resection if all sites of disease can be addressed in some form or fashion (medical or procedural). In part, chemotherapy allows for time to understand the tumor biology and tempo of disease. Regression or stability indicates that resection may be beneficial, whereas progression and development of additional sites of disease indicates that resection of the primary tumor is unlikely to impact overall survival, and the patient will spend 2 to 3 months recovering from surgery if no complications prolong recovery. Those with locally invasive tumors, where an R0 resection would be less likely, may benefit from neoadjuvant chemotherapy and mitotane for at least two cycles, and possibly more depending on response.[83]

Metastatic Disease to the Adrenal Gland

Patients with metastatic disease to the adrenal glands should be carefully evaluated with regard to their overall prognosis and planned future therapy. There is ample evidence to suggest benefit of adrenalectomy for resection of a malignancy metastatic to the adrenal gland if it is the single site of disease.[84] Other indications may be a growing adrenal mass that could result in local invasion of adjacent structures or other issues when other disease is stable or being controlled with another treatment modality. Those with additional metastatic disease elsewhere may be better served by other treatment modalities. Some malignancies are quite radiosensitive, and in the situation of an aggressive cancer with poor prognosis, focused radiation therapy or other types of local intervention may be more appropriate; however, these are not recommended as first- or second-line therapy.

SURGICAL APPROACHES TO THE ADRENAL GLAND

The individual patient's anatomy should be carefully studied when selecting an operative approach. The operative approach is not solely based on size alone, but also includes prior surgical history, concern for malignancy, need for access to other intraabdominal organs, planned concomitant procedures elsewhere in the peritoneal cavity, cardiac and pulmonary status in terms of toleration of a lateral or prone position, body habitus, and position of the adrenal gland with respect to the renal vasculature and upper pole of the kidney from the standpoint of selecting an anterior or posterior approach. Important anatomic features to consider when choosing between a laparoscopic and retroperitoneoscopic approach are listed in Table 100.5. An attempt to control hormone excess and correct electrolyte abnormalities should be pursued if possible. Medications to assist in curbing excess hormone production in advance of surgery are listed in Table 100.6.

The most common indications for selection of an open approach cited in two reviews were contraindications to laparoscopy, tumor size greater than 5 cm, increasing body mass index, and contraindications related to specific disease processes.[17,86,87] Somewhat counterintuitively, Lee et al. reported that patients included in the Veterans' Affairs National Surgical Quality Improvement Program registry undergoing open adrenalectomy had considerably higher preoperative risk than those selected to undergo laparoscopic adrenalectomy.[88] Buell et al. reported a series of 48 patients who underwent bilateral adrenalectomy for hypercortisolism by anterior or posterior open approaches from 1985 to 1997.[89] Operative time, blood loss, and need for transfusion were equal. Acute morbidity was similar. Long-term morbidity was greater in those undergoing an open posterior approach. Patients required more pain medicine for longer, had increased chronic back pain, and took longer to return to work. Table 100.5 compares anterior, posterior, flank, and thoracoabdominal approaches to adrenalectomy.

As indications and approaches to adrenalectomy changed over time, so too did outcomes. Complication rates (increases and decreases) related to changes in treatment (new therapies, new techniques) vary over time.[87] Different treatment approaches bring with them the potential for new or unique complications. This is illustrated in longitudinal studies by Sidhu et al.[90] and Lo et al.[91] Comparisons of various approaches to adrenalectomy can be difficult to carry out due to patient comorbidities and risk, provider and hospital-level differences, and selection bias related to choosing one approach over another. When comparing costs and complications associated with various approaches to adrenalectomy, the data vary widely, with some finding comparable costs and complication rates for various open and minimally invasive approaches and others finding significant differences. Hospital stay is generally longer for open approaches, but can be comparable dependent on the disease process being treated. A significant variation in cost has been found at the level of preoperative evaluation as opposed to the operation itself.[87,92,93]

Adrenalectomy may be accomplished by a number of approaches. Most commonly, for small benign tumors, a laparoscopic transperitoneal approach is preferred. An alternative approach, for those with experience in adrenal surgery, is the posterior retroperitoneoscopic approach.[94] The spleen and liver do not need to be mobilized, and blood loss is often diminished due to increased insufflation pressures

TABLE 100.5 Comparison of Operative Approaches to the Adrenal Gland (Anterior, Posterior, Flank, Thoracoabdominal, Anterior/Posterior Minimally Invasive +/- Robotic Assistance

Approach	Advantages	Disadvantages
Open		
Anterior	Access to entire peritoneal cavity, facilitates multivisceral resections, appropriate for vascular bypass, can make multiple incisions. Approach of choice for malignancies. Access to both adrenal glands	Increased pain, wound infections, hernias, adhesions from prior surgery, ileus, hospital stay
Flank	Can extend between floating ribs, can remain fully retroperitoneal if desired	Difficult to access superior gland if incision is below twelfth rib, poor access to remainder of peritoneal cavity. No access to contralateral gland
Posterior	Avoids scar tissue from prior intraperitoneal operations. Access to both adrenal glands. Faster recovery than open anterior approach. No mobilization of liver or spleen/pancreas	Limited access, difficult to remove large tumors, limited vascular control, no multivisceral resection other than kidney. Possible chronic pain, nerve entrapment, abdominal wall laxity
Thoracoabdominal	Simultaneous access to thorax and abdomen. Advantageous when remobilizing liver, spleen, or pancreas after prior operation. Good access to aorta and vena cava	Largest, most painful incision, possible chronic pain due to nerve entrapment. Entry into chest
Minimally Invasive		
Anterior (laparoscopic transabdominal)	Supine or lateral position. Less pain, shorter hospital stay, fewer wound infections. Faster recovery. Access to most organs of the peritoneal cavity	If lateral positioning, no access to contralateral gland.
Flank (transabdominal or retroperitoneal)	Retroperitoneal approach avoids intraperitoneal scar tissue. Allows for removal of larger tumors than posterior retroperitoneoscopic approach	Limited working space, limited access to peritoneal cavity, limited ability for multivisceral resection or vascular control. Patient in lateral position
Posterior (retroperitoneal)	No lifting restrictions, less pain, fastest recovery, less bleeding, access to both glands without repositioning	Limited space, difficult to resect large tumors. Different view with definite learning curve
Hand-assisted	Allows for removal of larger tumors, pressure on tumor from retraction spread over wider area	Larger incision than laparoscopic approach. Visualization can be impeded by presence of hand
Robot-assisted	Increased degrees of freedom for wristed motion compared with minimally invasive instrumentation. Better ergonomics for surgeon	Often more expensive, longer operations. Less rapid exchange of instruments. Surgeon is not at bedside

TABLE 100.6 Medications to Minimize Adrenal Cortex Hormone Excess

Hormone in Excess	Medication	Mechanism	Starting Dose	Notes
Aldosterone	Spironolactone	MR antagonist	25 mg QD	Titrate slowly to BP, potassium, renin, and creatinine
	Eplerenone	MR antagonist	50 mg QD	Titrate slowly to BP, potassium, renin, and creatinine
	Amiloride	Antikaliuretic-diuretic	5 mg QD	Avoids rise in renin if performing diagnostic testing and/or AVS
Cortisol	Ketoconazole	Inhibits 11β-hydroxylase, 17-hydroxylase, and 17,20-lyase	200 mg QD	Not effective for moderate to severe hypercortisolism
	Metyrapone	Inhibits 11β-hydroxylase	250 mg Q4 hours	Monitor cortisol and 11-deoxycortisol levels
	Osilodrostat	Inhibits 11β-hydroxylase and aldosterone synthase	2 mg BID	Monitor cortisol levels, titrate slowly
	Mifepristone	Cortisol receptor blocker	300 mg QD	Cannot use cortisol level to monitor receptor blockade (leads to increased cortisol levels); monitor by symptoms and glucose
	Etomidate	Inhibits 11β-hydroxylase	0.04 mg/kg per hour intravenously	Requires ICU monitoring
	Mitotane	Inhibits cholesterol side-chain cleavage enzyme, 11β-hydroxylase, 18-hydroxylase, and 3β-hydroxysteroid dehydrogenase	1 g BID	For use in patients with adrenal cancer; also an adrenolytic; derivative of the insecticide DDT
Androgens	Spironolactone	MR antagonist	25 mg QD	Abiraterone and ketoconazole are alternatives
Estrogens	Tamoxifen	Estrogen receptor antagonist	5 mg QD	Raloxifene, anastrozole, letrozole, exemestane are alternatives

MR, Mineralcorticoid receptor; *QD*, every day; *AVS*, adrenal vein sampling; *Q4*, every 4 hours; *BID*, twice daily; *DDT*, dichlorodiphenyltrichloroethane.

and the need for less mobilization of overlying structures. This approach also allows for access to both adrenal glands without repositioning, reprepping, and redraping during bilateral laparoscopic transperitoneal adrenalectomy. Postoperative pain can be less and recovery quicker. However, the retroperitoneoscopic approach does not allow for multivisceral resection or adequate access to the peritoneal cavity. Those with significant renal insufficiency may experience further worsening of their glomerular filtration rate due to the increased insufflation pressures. Robotic assistance may be used for either laparoscopic or retroperitoneoscopic approaches.[95] For those patients with larger tumors not amenable to a purely laparoscopic or laparoscopic hand-assisted approach, those with tumors concerning for primary adrenocortical carcinoma, those with prior surgery in the upper retroperitoneum, or those needing multivisceral resection or concomitant intraperitoneal procedures, an open transabdominal approach using a subcostal or midline incision is usually pursued. Although used frequently in the past, the open posterior approach is rarely used anymore, but remains a useful approach-especially in those who have undergone prior surgery in the upper quadrants. Finally, a thoracoabdominal approach to the adrenal glands may also be used but is rarely needed, except in the situation where access to the chest and abdomen is required or there has been significant prior surgery in the area and an open posterior or posterior retroperitoneoscopic approach is not possible. Most often, these situations involve patients with extensive invasive or recurrent malignant adrenal tumors.

KEY POINTS

- The surgical approach to adrenalectomy should be chosen carefully and customized to the individual patient, taking into consideration the specific disease process and type of hormone excess, radiologic findings, concern for malignancy, patient anatomy and comorbidities.

OPEN ANTERIOR TRANSABDOMINAL

Right

Because open adrenalectomy is most often performed for malignant tumors, this section will describe the operative approach from an oncologic standpoint. Although a midline incision can be made, a wide right subcostal incision extending to the left side is favored because of an improved ability to retract the ribs and mobilize the liver, providing exceptional access to the adrenal gland, kidney, and vena cava. The peritoneal cavity is inspected in a systematic fashion. Intraoperative ultrasound of the liver and evaluation of the adrenal vein and vena cava for tumor thrombus may be performed. The attachments to the liver are divided and the right lobe of the liver rotated medially. A retractor (Omni or Thompson most often) is placed to provide exposure. Through the surface of the posterior peritoneal lining, the kidney and the adrenal gland are able to be appreciated. If the tumor is found to be invading the posterior aspect of the liver, a rim of normal liver may be resected *en bloc* as part of the anterior margin. As delineated in Fig. 100.1A, dissection occurs inferiorly over the superior half of the kidney, taking the posterior peritoneal lining and any fat over the superior half of the kidney to provide an ample inferior margin. This is continued to the plane found between the adrenal gland and kidney. If there is any concern for the invasion of the kidney, removal of a portion of the capsule of the kidney or partial or complete nephrectomy may be performed, depending on the extent of suspected invasion, because one does not want to create a plane between tumor and the surface of the kidney where none exists, allowing for microscopic abrasion and dissemination of tumor cells. Dissection is continued laterally to the abdominal sidewall, and all retroperitoneal fat surrounding the tumor is taken as part of the lateral and posterior margins, with dissection exposing the posterior musculature. Dissection continues superiorly to the diaphragm. Some of the diaphragm may be excised *en bloc* with the tumor if necessary and closed in standard fashion with a running locking suture. Occasionally, mesh is required for closure of a large defect. The medial dissection is then performed. Dissection begins over the anterior aspect of the vena cava. The fat medial to the tumor is then taken as part of the specimen, including any fat or lymph nodes posterior to the renal hilum and vena cava and along the lateral surface of the aorta. Working superior to inferior and inferior to superior, the adrenal vein is eventually identified, ligated, and divided in standard fashion. If invasion of the vena cava is identified, various methods can be used to resect portions of vein to provide negative medial margins or extract tumor thrombus, including resection and

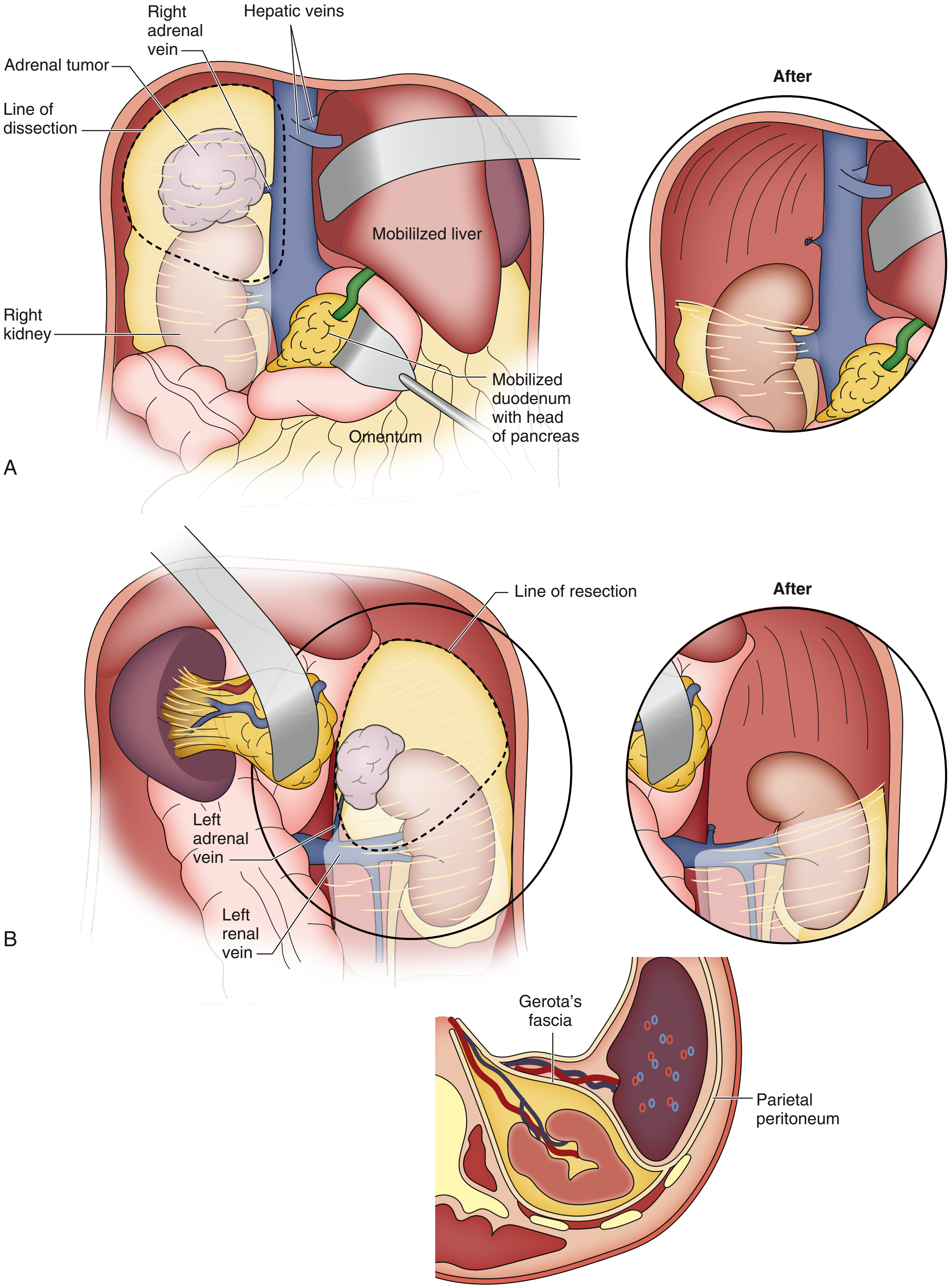

Right adrenal vein
Hepatic veins
Adrenal tumor
Line of dissection
Mobililzed liver
Right kidney
Mobilized duodenum with head of pancreas
Omentum
A
After
Line of resection
After
Left adrenal vein
Left renal vein
B
Gerota's fascia
Parietal peritoneum

Figure 100.1 Lines of resection for right (A) and left (B) open adrenalectomy. This includes the retroperitoneal fat surrounding the adrenal gland.

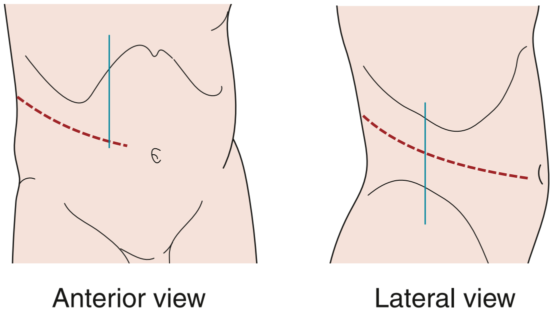


Figure 100.2 Positioning and incision for open flank transabdominal or retroperitoneal adrenalectomy.

replacement of a segment of the vena cava with ringed polytetrafluoroethylene graft. Venovenous or cardiopulmonary bypass may be needed in some cases. The tumor is then removed and marked with sutures. Clips can then be placed if radiation therapy will be considered in the future. The abdominal wall is closed in two fascial layers along with suture to approximate the deep subcutaneous tissue. A running suture or staples can be used for skin closure.

Left

Similar to the right side, a wide subcostal incision is made (Fig. 100.1B). The colon is mobilized along the white line of Toldt. The splenic flexure and spleen are then released from their attachments to the lateral side wall diaphragm to the esophageal hiatus. The stomach, spleen, pancreas, and colon are then retracted medially, exposing the kidney, left adrenal gland, and aorta. Conduct of the operation on the left side is similar to that of the right side, taking all surrounding and retroperitoneal fat in the area with the adrenal gland.

KEY POINTS

- Surgeons need to ensure performance of the critical elements of resection for adrenocortical carcinoma, with surgery best carried out in high volume centers given surgery for benign adrenal disorders is quite different than for adrenocortical carcinoma.

OPEN FLANK–TRANSABDOMINAL OR RETROPERITONEAL

With the patient in a lateral position (full or partial), a transabdominal or retroperitoneal approach may be taken (Fig. 100.2). The incision may extend between the eleventh and twelfth ribs, the twelfth rib may be removed, or the incision may be placed beneath the twelfth rib, although this approaches the adrenal from the most inferior aspect, and the surgeon may struggle to reach the superior aspect of the adrenal gland. The surgeon may traverse the peritoneum from this approach or may choose to remain extraperitoneal, mobilizing the ascending colon on the right or the descending colon, splenic flexure, spleen, and pancreas on the left to access the organs and structures of the retroperitoneum. The adrenal is then removed in similar fashion to the more standard open anterior approach.

OPEN POSTERIOR (LUMBOTOMY)

The patient is placed in a prone position. A curvilinear or angled incision is made as shown in Fig. 100.3, extending from medial to lateral along the course of the twelfth rib. The twelfth rib is identified and removed. The retroperitoneum is entered. It is important to sweep the retroperitoneal fat from superior to inferior and from medial to lateral to identify and preserve the diaphragm and locate the gland. Because of the location, the pleural cavity may inadvertently be entered and may be sutured-closed. Using handheld retractors to retract the ribs and musculature, the kidney is identified and retracted inferiorly with the surgeon's hand, and the adrenal gland identified medially and superior to the renal hilum. Adrenalectomy is performed, albeit from a dorsal approach, in standard fashion. The wound is closed in multiple layers, making sure not to injure the neurovascular bundle of the eleventh and twelfth ribs.

LAPAROSCOPIC ANTERIOR TRANSABDOMINAL

Right

The patient is placed supine, sometimes with a gel roll or rolled blanket under the right side, as shown in Fig. 100.4A. Port sites are triangulated using standard techniques for access. On the right side, four ports (occasionally more) are most often placed, one of which is for retraction of the liver. The triangular and coronary ligaments of the right lobe of the liver are divided, and the liver is retracted medially. An alternative approach is to forego mobilizing the attachments to the lateral and superior aspect of the liver and instead retract the liver anteriorly to dissect around the adrenal gland. This approach is less ideal than the lateral transadominal approach, especially for larger masses, as it limits working space and access to the adrenal gland. The upper pole of the kidney is identified, and the space between it and the adrenal gland is identified. The vena cava is also identified. Intraoperative ultrasound scanning can be useful for identifying the adrenal gland when abundant retroperitoneal fat is present. The attachments of the periadrenal fat are released from the upper pole of the kidney using standard techniques, most often ultrasonic dissection. It is critical to identify and remove the most inferior extent of the adrenal gland, which sometimes overlies the renal vasculature, as even a small amount of hormonally functional tissue can lead to persistent disease. Once the inferior attachments have been released, some of which may include small vessels coming off the renal artery or renal vein, the dissection continues superomedially along the vena cava until the adrenal vein is identified. Options for ligating and dividing the adrenal vein include ultrasonic instruments, clips, and sometimes a stapler if access is difficult or the vein is wide. The more superior attachments medially and laterally are then released, and the adrenal gland is removed. Port sites are closed in standard fashion. The operation is more difficult than when the patient is in a lateral decubitus position due to lack of aid from gravity to help allow the liver to fall away from the adrenal gland to the contralateral side. However, if the use of hand assistance with a gel port is anticipated, or the potential for risk of conversion to a laparoscopic approach is high, this may be a more optimal approach.

Left

Four ports (occasionally more) are most often required for left adrenalectomy (Fig. 100.4B). The splenocolic ligament is divided, and the colon mobilized inferiorly and medially. The remaining splenic attachments are then mobilized to the esophageal hiatus so that the spleen will remain mediatized, often without the need for a retractor. Left adrenalectomy is then carried out in similar fashion to the right side. The left adrenal vein is usually identified in the 7 o'clock position. This may be visualized on preoperative imaging in many cases. A communicating branch may course longitudinally from the adrenal vein to the phrenic vein along the medial aspect of the adrenal gland. Attachments of the adrenal gland to the surrounding tissue are ligated and divided, and the adrenal gland is removed. Similar to the right side, the supine positioning impedes the spleen from falling away to the right side away from the adrenal gland, and a retractor is necessary. An alternative approach is to avoid mobilizing the lateral and superior attachments and instead retract the tail of the pancreas and spleen anteriorly to allow for dissection around the adrenal gland. This approach limits access to the gland and working room.

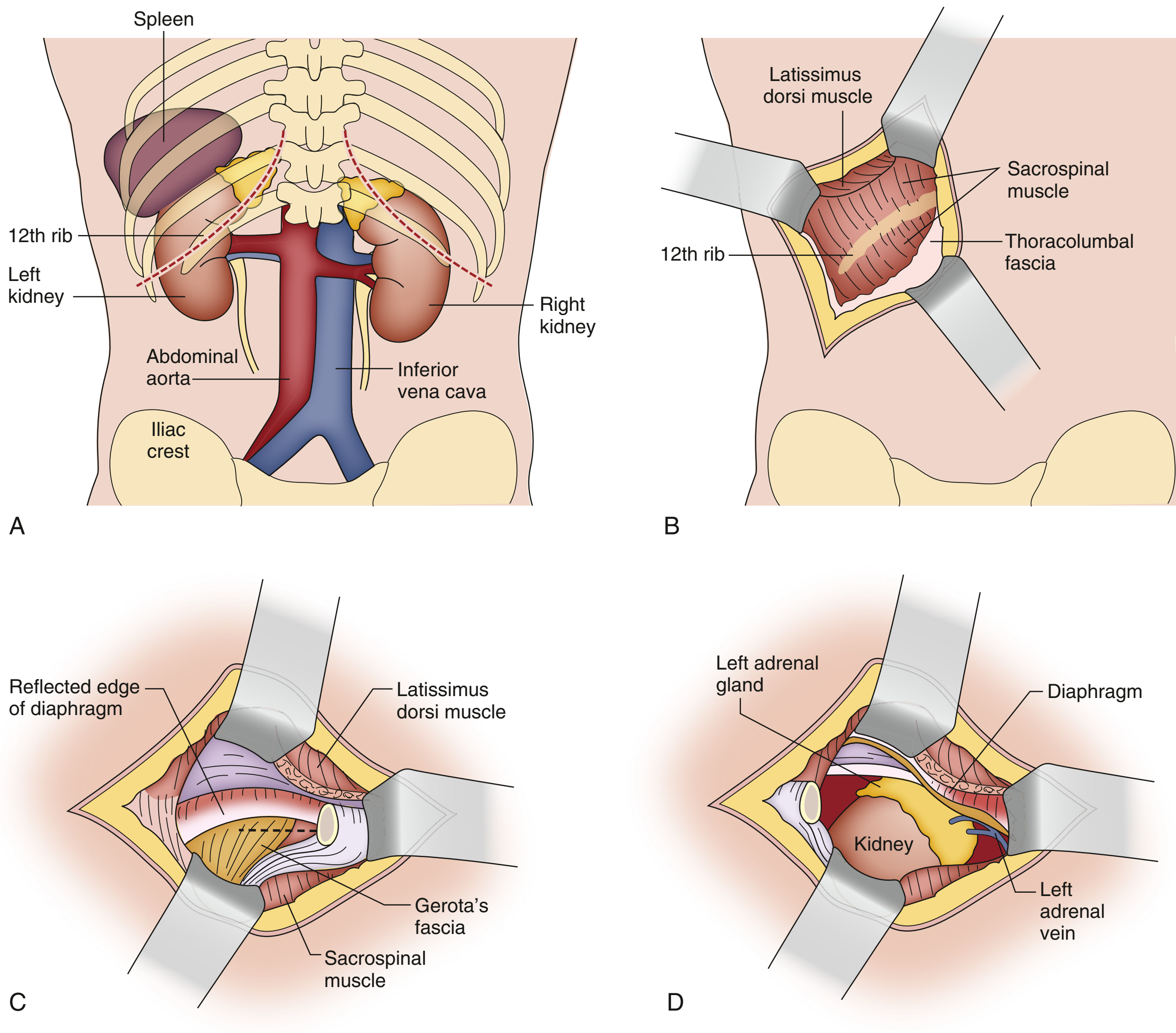

Spleen
12th rib
Left kidney
Abdominal aorta
Iliac crest
Inferior vena cava
Right kidney
A
Latissimus dorsi muscle
Sacrospinal muscle
Thoracolumbal fascia
12th rib
B
Reflected edge of diaphragm
Latissimus dorsi muscle
Gerota's fascia
Sacrospinal muscle
C
Left adrenal gland
Diaphragm
Kidney
Left adrenal vein
D

Figure 100.3 Open posterior adrenalectomy. **A,** Incision. **B,** Exposure of the twelfth rib. **C,** Removal of twelfth rib and exposure of the adrenal gland. **D,** Dissection of the left adrenal gland.

LAPAROSCOPIC LATERAL TRANSABDOMINAL ADRENALECTOMY

Right

The patient is placed in left lateral decubitus position at a 45-degree or greater angle, as shown in Fig. 100.5A, so that the abdomen and flank are accessible if conversion to an open procedure is necessary. The space between the ribs and pelvis is extended. Port sites are triangulated using standard techniques for access. On the right side, four ports (occasionally more) are most often placed, one of which is for retraction of the liver. The triangular and coronary ligaments of the right lobe of the liver are divided, and the liver is retracted medially. The upper pole of the kidney is identified, and the space between it and the adrenal gland is identified. The vena cava is also identified. Intraoperative ultrasound scanning can be useful for identifying the adrenal gland when abundant retroperitoneal fat is present. The attachments of the periadrenal fat are released from the upper pole of the kidney using standard techniques, most often ultrasonic dissection. It is critical to identify and remove the most inferior extent of the adrenal gland, which sometimes overlies the renal vasculature, as even a small amount of hormonally functional tissue can lead to persistent disease. Once the inferior attachments have been released, some of which may include small vessels coming off the renal artery or renal vein, the dissection continues superomedially along the vena cava until the adrenal vein is identified. Options for ligating and dividing the adrenal vein include ultrasonic instruments, clips, and sometimes a stapler if access is difficult or the vein is wide. The more superior attachments medially and laterally are then released, and the adrenal gland is removed. Port sites are closed in standard fashion.

Left

Three ports (occasionally four or more) are most often required for left adrenalectomy (Fig. 100.5B). The splenocolic ligament is divided and the colon mobilized inferiorly and medially. The remaining splenic attachments are then mobilized to the esophageal hiatus so that the spleen will remain medialized, often without the need for a retractor. Left adrenalectomy is then carried out in similar fashion to the right side. The left adrenal vein is usually identified in the 7 o'clock position. This may be visualized on preoperative imaging in many cases. A communicating branch may course longitudinally from the adrenal

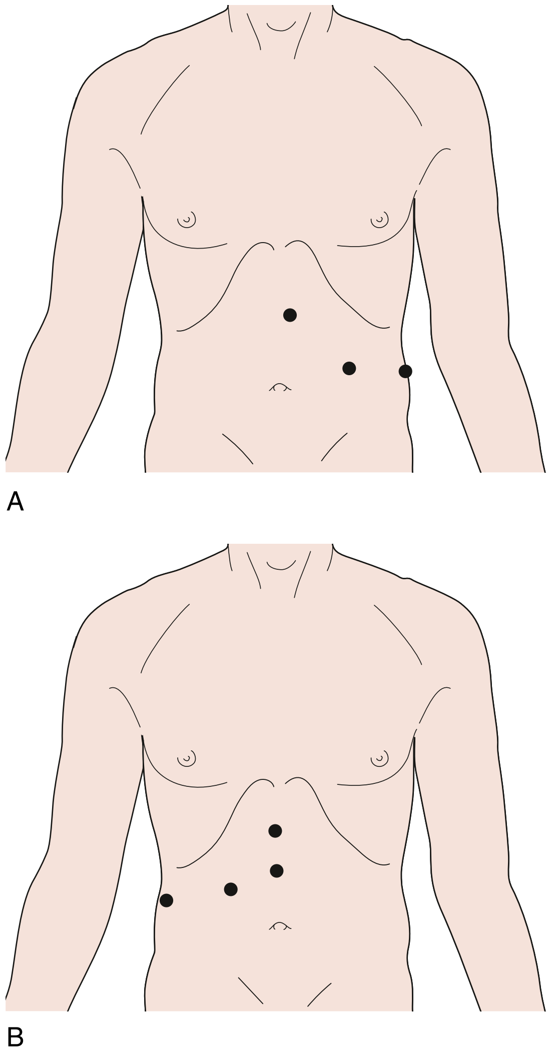


Figure 100.4 Positioning and port placement for right (A) and left (B) laparoscopic anterior transabdominal adrenalectomy.

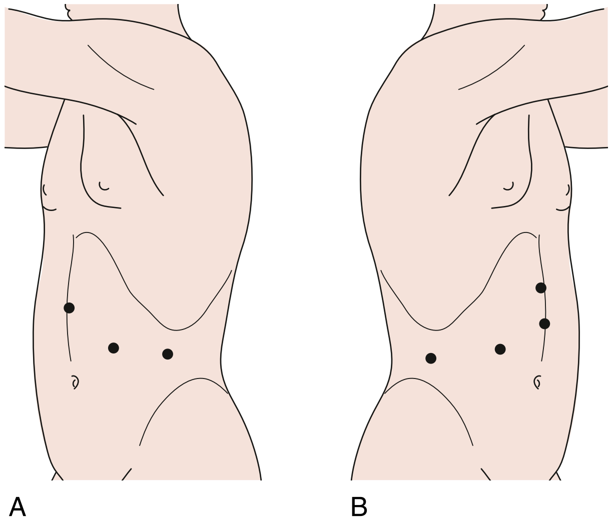


Figure 100.5 Positioning and port placement for left (A) and right (B) laparoscopic lateral transabdominal adrenalectomy.

vein to the phrenic vein along the medial aspect of the adrenal gland. Attachments of the adrenal gland to the surrounding tissue are ligated and divided, and the adrenal gland is removed.

LAPAROSCOPIC HAND-ASSISTED ADRENALECTOMY

In patients with moderately to significantly enlarged benign adrenal tumors, an intermediate option between minimally invasive and standard open surgery is incorporation of hand assistance with laparoscopic visualization and instruments. Attempts to remove large tumors laparoscopically are more prone to tumor capsule disruption. The ability to mobilize the tumor is hampered by difficulty with visualization, limited working space, and at times the sheer weight of the tumor. While an open approach usually allows for excellent access and working space, a larger incision is associated with a higher incidence of wound infection, longer hospital stay, and greater chance of incisional hernia. An intermediate option with growing appeal is a laparoscopic hand-assisted approach utilizing a hand port to allow for manipulation of surrounding organs and the tumor with one hand while the other hand utilizes the usual laparoscopic instrumentation for division of soft tissue and vasculature (Fig. 100.6). Given that morcellation of adrenal tumors is to be avoided, enlargement of at least one incision is required for extraction of the tumor. For tumors greater than 6 to 8 cm, the size of the incision needed for extraction will admit a small- to medium-sized hand. Adrenalectomy performed with hand assistance for moderate to large tumors can often be performed faster than those performed by conventional laparoscopy. Incisions, while larger than those for a pure laparoscopic approach, but smaller than those required for an open approach, are associated with fewer wound infections, less postoperative pain, fewer incisional hernias, and shorter hospital stay.

LATERAL RETROPERITONEOSCOPIC ADRENALECTOMY

In this least often utilized approach for adrenalectomy, the patient is placed in a lateral decubitus position. A 10-mm camera port is placed in the midaxillary line, followed by two 5-mm working ports for additional instruments in the anterior and posterior axillary lines (Fig. 100.7). Because access is obtained less posteriorly than in a dorsal retroperitoneoscopic approach, it is important to avoid entering the peritoneal cavity. The upper pole of the kidney is mobilized and separated from the inferolateral aspect of the kidney, and the rest of the dissection performed similarly to that discussed previously. Dissection medially may be more difficult, especially on the right side, because the gland will be resting on medial structures (vena cava and aorta) more so than compared with anterior, posterior, or lateral transabdominal cases, where the patient is not in full lateral decubitus position. This particular approach has the advantage of permitting a retroperitoneal approach in patients with higher body mass index or tumors of larger size where a dorsal retroperitoneoscopic becomes more difficult.[96] This approach can also be useful in those with dense scar tissue in the upper abdominal quadrants due to prior surgery. Hand assistance may be utilized with this retroperitoneoscopic approach, unlike with the dorsal retroperitoneoscopic approach.[97]

POSTERIOR (DORSAL) RETROPERITONEOSCOPIC ADRENALECTOMY

The patient is placed in a prone position, as shown in Fig. 100.8. An initial incision is made near the tip of the twelfth rib. Blunt dissection into the retroperitoneal space is pursued. The retroperitoneal fat is bluntly swept from the musculature using a finger, allowing for placement of a 5-mm port near the tip of the eleventh rib laterally and a 10-mm port medially at the lateral edge of the erector spinae muscles approximately 30 degrees inferior to the original incision. The retroperitoneum is insufflated to a pressure of 20 to 22 mm Hg. This

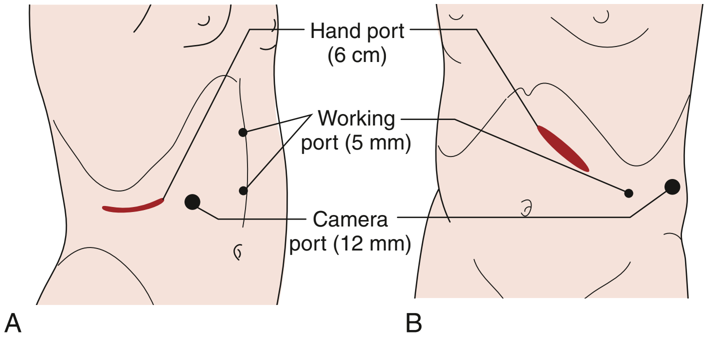


Figure 100.6 Gel port used for hand-assistance: right (**A**) and left (**B**).

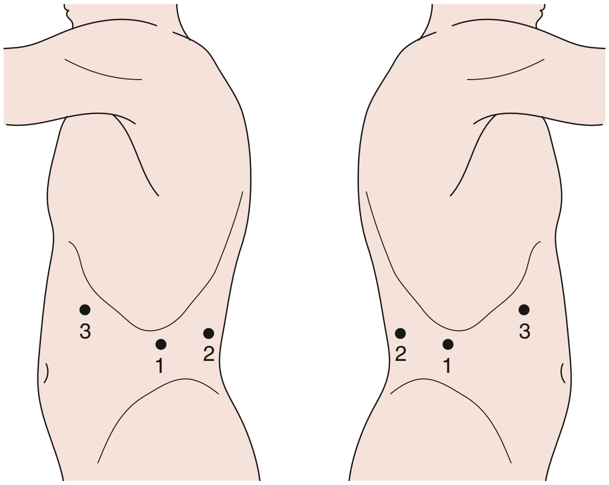


Figure 100.7 Port placement for lateral retroperitoneoscopic adrenalectomy.

can be increased if needed to assist with control of oozing or bleeding up to approximately 30 mm Hg. Areolar connective tissue is bluntly dissected from the posterior musculature, and the dissection carried medially and superiorly. As dissection and exposure proceed, the hue of the liver will be seen superolaterally. It is important not to dissect into the peritoneal cavity, because visualization becomes extremely difficult with loss of insufflation of the retroperitoneal space. Once the adrenal gland is identified, dissection should proceed from an inferior to superior fashion. On the right side, surgeons should proceed with caution during the medial dissection until the vena cava is clearly identified, because it will be compressed due to increased insufflation pressures. Because the right adrenal vein often emanates from the posterolateral aspect of the vena cava, it will be visualized more anteriorly in this approach than in a traditional laparoscopic approach, given the prone positioning. The adrenal vein is ligated and divided, and the remaining attachments are released. The adrenal is removed, the fascia of the middle port site is reapproximated, and the incisions are closed. A similar approach is taken for left retroperitoneoscopic adrenalectomy. The hue of the spleen is identified superolaterally, and dissection is carried out as described. The left adrenal vein can be more difficult to identify than the right adrenal vein due to the proximity of the left renal vein and a more posterior position compared with the traditional transperitoneal approach. With all laparoscopic and retroperitoneoscopic procedures, it is important to avoid disturbing the capsule of the adrenal gland, because this will lead to bothersome oozing during the case, which may obscure visualization. Some literature suggests that partial adrenalectomy is appropriate in select situations. This should be very carefully considered and performed by a high-volume adrenal surgeon, because there are nuances to the decision-making and conduct of the operation.

ROBOT-ASSISTED ADRENALECTOMY

Recently, enthusiasm has been increasing for utilization of computer-aided devices to perform minimally invasive adrenalectomy. The main advantage is that of the wristed motion afforded by certain instrumentation on the various robotic platforms. With computer-aided surgery, the adrenals can be reached from anterior, posterior, and lateral approaches. Collision of robotic “arms” can be an issue, and placement of incisions in close proximity can be an issue in some cases. An additional port for use by the “assistant” remaining at the bedside is utilized by some surgeons to more easily switch out instruments when needed. Costs may be increased, but strategies to mitigate this have been put in place by some surgeons.[98,99]

THORACOABDOMINAL APPROACH TO ADRENALECTOMY

Large tumors, tumors with extension into the thoracic cavity, need for control of the infracardiac suprahepatic vena cava, or patients who have undergone prior right upper quadrant surgery with severe scar tissue limiting access and remobilization and rotation of the liver may benefit from a thoracoabdominal incision. However, for the vast majority of adrenalectomies, including many of the listed situations, a less invasive approach, including more standard open approaches, is more than sufficient. A thoracoabdominal incision is the most painful of all incisions used for adrenalectomy and is associated with prolonged recovery and a greater chance for postoperative morbidities (Fig. 100.9).

DISEASE-SPECIFIC CONSIDERATIONS AND DECISION FOR EXTENT OF ADRENALECTOMY

The most common indications for surgery involve a single tumor or adrenal gland needing removal. In a minority of cases, partial unilateral, unilateral with contralateral partial, or bilateral adrenalectomy may be most suitable for treatment of a patient’s specific condition.

BILATERAL ADRENALECTOMY

The ultimate approach chosen by the surgeon for bilateral adrenalectomy is dependent on multiple factors. Time in the operating room can be lengthy, especially if a laparoscopic transabdominal approach is selected, as patients need to be repositioned, reprepped, and redraped in the middle of the case. An open anterior approach does not require this but is associated with higher wound infection rates, greater pain,

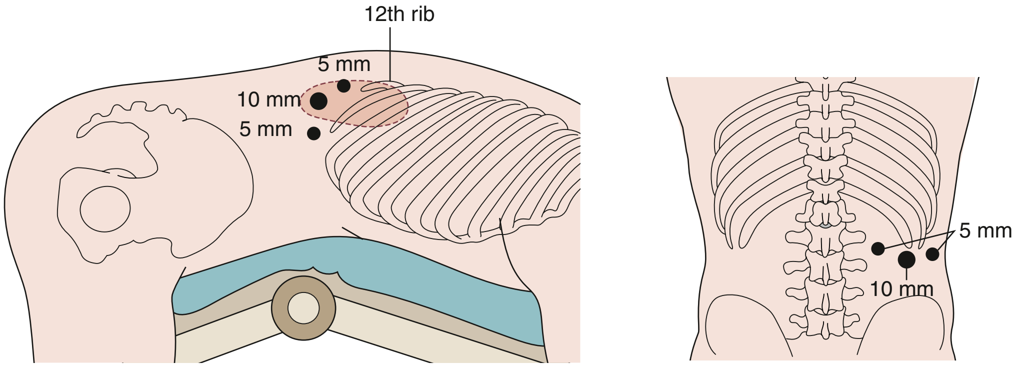


Figure 100.8 Patient positioning and port site placement for posterior retroperitoneoscopic adrenalectomy.

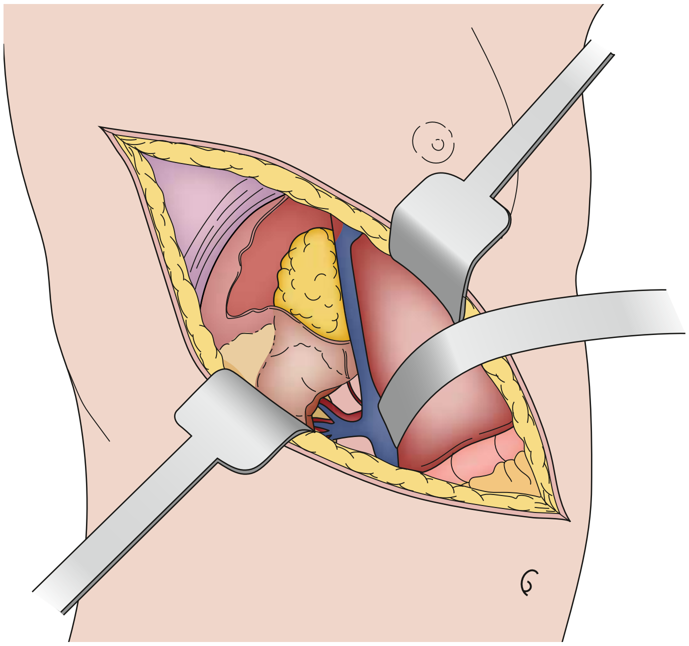


Figure 100.9 Positioning and incision for thoracoabdominal approach to adrenalectomy.

longer hospital stay, and longer recovery. Bilateral open posterior adrenalectomy, the most popular approach for many years prior to development of minimally invasive techniques, has advantages compared with open anterior approaches but is no longer learned or utilized by many surgeons, as it has been supplanted by the retroperitoneoscopic posterior approach. However, the open posterior approach remains useful for those not familiar with the retroperitoneoscopic approach, when prior surgery or complications have disrupted the normal planes in the retroperitoneum, or when conversion to an open approach is needed and the patient is already in a prone position.

PARTIAL ADRENALECTOMY

In the case of adrenal tumors possibly resulting from a genetic mutation, patients should be referred to a genetic counselor. Ideally, those suspected of having a genetic mutation would undergo this testing in the preoperative setting, as knowledge of the specific mutation can impact preoperative testing and conduct of the operation. Adrenal tumors may be found in patients with MEN 1. ACC is identified in patients with Li–Fraumeni syndrome (p53 mutation). Pheochromocytoma is associated with MEN 2A, MEN 2B, neurofibromatosis, and von Hippel–Lindau syndrome. Patients who are potential candidates for partial adrenalectomy should be referred to a surgeon with expertise in the disease process, the decision-making process, and the nuances of partial adrenalectomy. Partial adrenalectomy is most often performed in patients with pheochromocytomas resulting from a genetic syndrome. Given that tumor development is driven by a mutation, many of these patients develop tumors and symptoms at a young age, and tumors can eventually involve both adrenal glands. Given that both glands may become abnormal, rather than remove the entire gland with the tumor and normal adrenal tissue, an attempt can be made to spare adrenal cortical tissue to avoid the need for corticosteroids and the potential for Addisonian crisis. Quality of life is also reduced in patients on lifelong steroid replacement. The surgeon must carefully select patients for cortical-sparing adrenalectomy. The position of the tumor within the gland, the amount of remaining adrenal tissue, the venous outflow that will be disrupted during surgery and that will remain after surgery, as well as what will be anticipated in any future reoperative completion adrenalectomy must be considered (Fig. 100.10). Where the remnant will be in relation to the vena cava, aorta, and/or renal hilar vasculature may impact the approach to the initial surgery to allow for easier future reoperation. Remobilization of the liver after a transabdominal approach can be difficult, especially if the first operation was done as an open approach; a subsequent posterior approach may be easier at

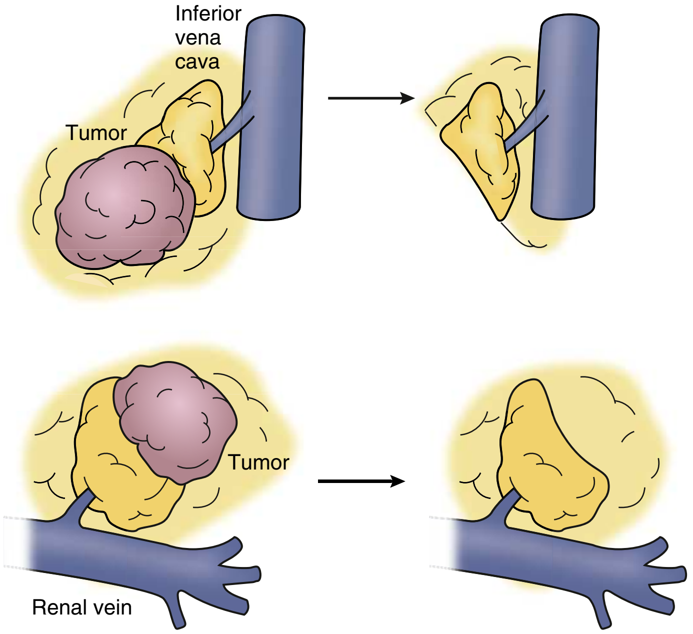


Figure 100.10 **Partial adrenalectomy.** Multiple tumor positions within the adrenal gland are amenable to partial adrenalectomy, and each must be considered individually with regard to vascular inflow and outflow, expected size of the residual remnant, and difficulty of any expected reoperation.

the time of reoperation, especially if the remnant ends up positioned posterior to the vena cava. Similarly, if a remnant is likely to reside near the renal hilum, it may be best to approach the initial operation posteriorly to allow for an anterior approach at the time of reoperation. At least 30% of a normal gland should remain if the hope is for cortical function not requiring supplemental steroids. During dissection, it is extremely important to avoid dissection of the entire periphery of the gland, and instead concentrate on confining dissection to the area of the portion of gland to be removed and avoid disruption of small vessels providing inflow and outflow to the portion of the gland that will remain. Indocyanine green can be utilized to assess perfusion of the remnant.

DISEASE-SPECIFIC INTRAOPERATIVE CONSIDERATIONS

Considerations for the intraoperative care of patients are listed in Table 100.4.

Extent of Resection for Primary Aldosteronism. Despite some reports suggesting that partial adrenalectomy is associated with similar success compared with total adrenalectomy, the recommendation for total unilateral adrenalectomy remains for patients undergoing surgery for primary aldosteronism. Outcomes reported by some groups suggesting equivalence in these studies were of a short-term nature, and the degree of cure according to the Primary Aldosteronism Surgical Outcomes group reporting structure is unclear. Recent work on the molecular side illustrates the concern of performing partial adrenalectomy. Approximately 10% of adenomas identified on CT do not produce aldosterone when stained by immunohistochemistry for Cyp11b2 (aldosterone synthase).[100] It is elsewhere in the gland that excess aldosterone is being produced. Even if the adenoma is producing excess aldosterone, other areas in the gland may also be autonomously producing aldosterone, whether areas of hyperplasia, other microadenomas unable to be appreciated on preoperative imaging, or aldosterone-producing cell clusters. In addition, it is now known that, with age, areas of aldosterone-producing cell clusters can autonomously produce aldosterone. It is impossible at this time to reliably predict in advance the site(s) of production of aldosterone. Some have suggested the possibility of superselective venous sampling. Given the new knowledge of the distribution of autonomous aldosterone production within the adrenal glands and the variable venous outflow from aldosterone-producing areas within the gland, not to mention named and unnamed venous structures visible intraoperatively, less than complete adrenalectomy based on imaging or superselective venous anatomy cannot be recommended at this time.[99,100]

Hypercortisolism. During the preoperative timeout, discussion should center on electrolyte and glucose management, as well as expected blood loss. Positioning with ample padding of pressure points is necessary given the increased risk of skin breakdown due to cortisol excess. A discussion about any necessary steroid administration should occur. As a general rule, patients with overt signs of cortisol excess will require steroid supplementation in the postoperative period. Cortisol may be administered intraoperatively, but in severe cases with significant excess may be held until several hours after surgery given that the half-life of cortisol is 1 to 2 hours (depending on cortisol levels and degree of excess). Surgeons should work in concert with endocrinology colleagues to devise an appropriate plan tailored to the patient.

For patients with mild autonomous cortisol secretion, postoperative supplementation may or may not be needed. In general, intraoperative steroid administration can be avoided to facilitate testing of the HPA axis and need for steroids the morning after surgery. Dexamethasone administration by the anesthesia service to prevent postoperative nausea and vomiting should be avoided in these cases, as the large doses given will potentially interfere with morning cortisol levels depending on the amount given and the timing of surgery.

Pheochromocytoma. Good communication between surgical, anesthesia, and operating room staff is imperative when removing a

TABLE 100.7 Critical Elements of Resection of Adrenocortical Carcinoma

Element	Description	Additional Details
Exploration	Fully explore the peritoneal cavity unless limited by scar tissue	
Complete R0 resection of the primary tumor and periadrenal fat	Resection to microscopically negative margins without disruption of the tumor capsule	The adrenal gland and tumor, surrounding retroperitoneal fat, Gerota's fascia, and overlying peritoneum should be resected *en bloc*
En bloc resection of involved adjacent organs	Resect adjacent organs (fully or partially) if directly invaded by or adherent to the primary tumor	Creation of an artificial plane between the primary tumor and an adherent or invaded structure should not be attempted
En bloc resection of involved vasculature and/or tumor thrombus	Dependent on extent, utilize bypass, liver perfusion, and/or circulatory arrest as needed, with or without reconstruction	Most often refers to the renal veins, renal arteries, and/or vena cava. Tumor thrombus may extend into the right atrium
Therapeutic lymphadenectomy	Abnormal lymph nodes noted on imaging or intraoperatively. Examine periadrenal, perirenal, aortocaval lymph nodes.	An extended prophylactic lymphadenectomy is not recommended. The station(s) of origin of any resected lymph nodes should be specifically noted in the operative report

pheochromocytoma. Preoperative timeouts should include a discussion of goals of fluid administration before and after adrenal vein ligation, potential for blood loss, goals for blood pressure management, and available agents for treating hypertensive and hypotensive episodes. The anesthesia team should be notified, if possible, prior to ligation of the adrenal vein to allow time for transition of blood pressure agents for anticipated hypotension and increased fluid administration. Fluid administration is the first method to increase blood pressure. Vasopressors may be required in some patients to maintain a mean arterial pressure greater than 60 mm Hg. Alpha agonists may not work well, or increased amounts may be required, depending on the degree of preoperative alpha blockade. Vasopressin is an alternative that works well in many cases. Hypotension often resolves by the end of the case, but may persist over the first 24 hours.

Adrenocortical Carcinoma. Given the aggressive tumor biology, resection of ACC differs from resection of benign adrenal masses with regard to the recommended technical components of resection. Aside from removal of the entire tumor or gland depending on the disease process, there are no specific recommendations for benign adrenal processes. Because of the high chance of local tumor recurrence, peritoneal dissemination, and poor prognosis, technical aspects of resection play a primary role in ACC. Table 100.7 lists the technical elements of resection that must be performed. Multivisceral resection is often required, as is management of vascular involvement, necessitating some form of reconstruction. The American College of Surgeons Operative Standards for Cancer Surgery has set the standard for technical aspects of resection and has evaluated several areas of debate, including operative approach, surgery in stage IV patients at the time of presentation, and the utility of prophylactic lymphadenectomy.[101] It is also recommended that in addition to a detailed operative report, a synoptic operative report also be recorded in the medical record to capture aspects of the operation in a consistent format.

On occasion, unsuspected metastatic disease may be noted upon exploration of the peritoneal cavity. In this situation, long-term prognosis is usually quite poor. A decision to abort the operation is most often made, but there could be instances where a decision to proceed is made. This decision hinges on the volume of disease and the ability to resect all disease. Resection of the primary tumor and all metastatic disease followed by early institution of mitotane, and possibly etoposide, doxorubicin, and cisplatin (EDP), could be pursued.

There has been significant debate over the past 10 years regarding the operative approach to adrenalectomy for ACC, and an open approach remains the recommended approach for resection of ACC. If a surgeon were to consider a minimally invasive approach for resection of a small (<6 cm) ACC, the following conditions would need to be satisfied: the individual surgeon has considerable experience with both laparoscopic adrenalectomy and surgical resection of ACC, can ensure performance of the critical elements as described, and can produce acceptable oncologic outcomes.[101] Few surgeons meet these requirements at this time.

Performance of prophylactic lymphadenectomy at the time of resection of the primary tumor is not recommended despite some calls to do so.[101] The vast majority of patients with stage IV disease do not undergo surgery; however, there is a highly select group of patients that may benefit from resection of the primary tumor. Selection is key and requires careful assessment to understand tumor biology, tempo of disease, and other factors to gauge the expected long-term durability of surgical intervention. Often, treatment of patients with stage IV disease with EDP with mitotane for at least two rounds with restaging will provide an understanding of patients in whom surgery will not be of benefit, if progression of metastatic disease has occurred, versus patients who are noted to have regression or stability, in whom surgery could be considered. Performance status also plays a significant role in the decision for surgery in these patients. One must also consider the length of time during which metastatic disease will be unable to be treated with chemotherapy after surgery or if complications occur.

Reoperative Surgery for Adrenocortical Carcinoma. The decision for surgery in patients with recurrent ACC is complex. The disease-free interval, tumor grade, quality of initial resection, location of recurrent tumor, anticipated operation required to remove recurrent disease, ability to treat all recurrent disease, and fitness of the patient to tolerate reoperation all enter into the decision for reoperation. Even if an operation is technically feasible, the tumor biology may dictate avoidance of surgery and pursuit of systemic therapy instead. Recurrence after reoperation for an initial recurrence reaches a rate of 70% to 94%, and many times disease recurs in the same organ.[102,103] Therefore, there is now a greater emphasis on performing tissue-sparing procedures by nonsurgical means, including radiofrequency ablation, cryoablation, regional therapy, external beam radiotherapy, and stereotactic body radiation therapy. Long-term durability of reoperation to control hormone excess is often poor, but can be beneficial in select patients. The usually short increase in survival (3–6 months) may not be significant when length of time to recover from surgery and any surgical complications is considered from a quality of life standpoint.

Surgical Complications

With the volume of adrenalectomies performed having increased over time, the influence of surgeon adrenalectomy volume on complications has been recently reviewed.[104] Just over 80% of patients undergo adrenalectomy by a low-volume surgeon. This same study found that 82% of surgeons performing adrenalectomy completed an average of one adrenalectomy per year.[105] This raises concern regarding the understanding of endocrine physiology, as it can be quite nuanced and affect decision for surgery and the extent of surgery. A significant difference became apparent in complication rates at the level of six adrenalectomies per year.[106] However, this does not indicate that surgeons performing more than six adrenalectomies per year are high-volume adrenal surgeons. While there are no specific data to support a particular cutoff, at least one group suggests that a high-volume surgeon performs at least 30 adrenalectomies per year.[30] Similarly, technical abilities alone do not necessarily indicate familiarity with a particular disease process, type of hormone excess, or ability to have expertise in treating malignant adrenal disease.

Intraoperative Complications

Overall, intraoperative and postoperative surgical complications are few. Intraoperatively, complications may be the result of technical errors, physiologic derangements secondary to underlying hormone excess, and unrelated preexisting comorbidities. Cardiac complications are infrequent. In the past, the morbidity and mortality after resection of pheochromocytomas was extremely high. With improved preoperative care, modern anesthetic and surgical techniques, and postoperative care, cardiac morbidity and mortality are quite rare. Bothersome bleeding from the adrenal gland itself may occur if the capsule is penetrated. Other sources of brisk bleeding can result from laceration or puncture of surrounding solid organs such as the kidney, liver, spleen, pancreas, short gastric vessels, and the occasional small vascular structure coursing through the splenic and hepatic flexures or retroperitoneal fat. Massive bleeding may result from disruption of named vascular structures, including the renal veins and arteries, adrenal vessels, vena cava, aorta, phrenic vessels, hepatic veins, and splenic vessels. Interestingly, posterior retroperitoneoscopic approaches are associated with less blood loss compared with anterior or lateral transperitoneal approaches. This is due to the ability of the increased insufflation pressures utilized to create the retroperitoneal space to overcome the low venous pressure of the vessels at greatest risk.

Perforation of hollow viscus structures may occur due to inadvertent puncture or thermal injury during minimally invasive or open approaches. Similarly, the diaphragm may be penetrated, and the pleural cavity entered. This is usually easily repaired, and most often does not require chest tube placement. During resection of ACC, resection of a portion of the diaphragm may be necessary, and in many cases can make superior dissection easier, also allowing for development of a superomedial plane from which to dissect inferiorly along the vena cava or aorta. Dissection along the aorta and vena cava can lead to lymph and chyle leak. Output can be high, reaching a liter or more daily. Careful ligation of lymphatic structures is imperative.

Minimally invasive operative approaches with use of CO_2 gas for insufflation of the peritoneal cavity or retroperitoneal space can lead to CO_2 embolism. Insertion of a trocar into the liver or a venous structure may also result in CO_2 embolus. Pulmonary embolus from deep vein thrombosis (DVT) rarely occurs intraoperatively. Tumor thrombus in the vena cava in patients with ACC can embolize, but is remarkably stable in many cases. Patients with cortisol excess are at higher risk for DVT and pulmonary embolism, and consideration should be given to postoperative DVT prophylaxis for several months. The increased risk is not simply due to obesity resulting from cortisol excess, but also to the hormone itself.

Postoperative Complications

Complications common in the postoperative period are those common to other surgical procedures, and have been reported to reach a rate of 20%, although this varies according to approach, disease process, and surgeon volume.[104,105] Postoperative bleeding requiring return to the operating room is rare, as is wound infection. Superficial wound infections are less common when a minimally invasive approach is taken compared with an open approach. Patients producing excess cortisol are at higher risk for wound infection and development of incisional hernias. Hernias are more common with open approaches than with minimally invasive approaches.

Change in renal function may result from hypovolemia and inadequate volume resuscitation during or after surgery. In addition, small arterial branches may be inadvertently or unavoidably ligated during dissection along the superior aspect of the renal hilum. If nephrectomy is anticipated in a patient with preexisting renal insufficiency, a renal perfusion scan may be obtained preoperatively to determine if contralateral kidney function will allow for avoidance of hemodialysis. In some cases, this will inform the decision for or against proceeding with surgery versus systemic therapy in ACC patients. An interesting phenomenon occurs after adrenalectomy for primary aldosteronism. Renal function is often overestimated in the preoperative setting, especially in those either not placed on or suboptimally managed on mineralocorticoid receptors, given a hypervolemic state. Once excess aldosterone production is controlled with adrenalectomy and a euvolemic state is achieved, the amount of fluid filtered by the kidneys is less, and the true state of renal function is able to be appreciated. Operative approach can also affect renal function, although this is usually transient. In patients undergoing posterior-approach retroperitoneoscopic adrenalectomy, increased insufflation pressures can limit urine output during the operation and also result in acute renal insufficiency. These changes usually quickly resolve in the first 24 to 48 hours but could be an issue in patients with stage 3 or 4 renal insufficiency.

Hyperkalemia sometimes occurs. This may be noted in the rare case of acute renal insufficiency or failure, but is much more common in patients who have undergone adrenalectomy for primary aldosteronism. If the zona glomerulosa of the contralateral adrenal gland is suppressed due to excess aldosterone production from the gland being removed, the potassium level may rise substantially, even to a level requiring treatment. Measurement of potassium levels in the postoperative setting for at least the first 2 weeks is warranted.[106] A significant percentage of patients requiring temporary fludrocortisone while awaiting resumption of production of normal levels of aldosterone are identified in the first 2 weeks after discharge of the patient from the hospital.

Complications and Operative Approach

Operative approaches have advantages and disadvantages from the standpoint of risk of morbidity. Open approaches are often associated with greater risk of morbidity, but selection bias plays a part, as patients often have larger tumors, require multivisceral resection, or have other reasons for why a minimally invasive approach has not been chosen.

When comparing minimally invasive anterior approaches with posterior approaches, there are some differences in complications and outcomes. Posterior approaches are generally associated with less blood loss. The increased insufflation pressures decrease venous bleeding. In addition, some believe that the increased insufflation pressures may limit outflow of catecholamines during resection of pheochromocytomas and minimize significant intraoperative swings in blood pressure, thus representing another difference in approach leading to differences in outcomes and management. Given the access points and that dissection is confined to the retroperitoneum, development of an incisional hernia is not of major concern when a posterior approach is taken. Lifting restrictions are not required as they are with an anterior approach. An exception is if penetration of the peritoneal cavity is inadvertently made, but bowel herniation would be rare in this situation as well. Finally, many patients experience change in sensation around the incision site(s), there may be a higher incidence of broader changes in sensation when posterior approaches are utilized. The neurovascular bundle of the twelfth rib may be compressed, leading to sensory deficits in the distribution of the entire dermatome as it proceeds anteriorly from the site of the incision.

DISEASE-SPECIFIC POSTOPERATIVE CARE AND FOLLOW-UP

Considerations for the care postoperative care of patients are listed in Table 100.4.

Pheochromocytoma. Patients operated on for pheochromocytoma may experience hypotension over the first 24 hours as the effects of decreased catecholamine secretion and vasodilation are encountered. Fluid should be administered to provide volume repletion. Alpha agonists may be required, but, due to preoperative alpha blockade, often do not result in significant effect or require more than usual to overcome the blockade. Vasopressin is a useful alternative if needed. Hypotension is often avoided or experienced to a lesser degree and for a shorter period of time if preoperative alpha blockade and volume expansion are optimized pre- and intraoperatively.

Hypoglycemia may also be noted in the immediate postoperative period, as previously discussed.[78-80] Pathology should be reviewed. If patients have not undergone genetic testing preoperatively, they should be referred to a genetic counselor for assessment. If invasion into adjacent structures is noted intraoperatively, lymph node involvement is noted or reported in the pathology report, or metastatic disease is identified in a patient without a germline mutation, testing of the tumor itself can be pursued to identify a potential driver mutation. Malignant pheochromocytomas are often more frequent and more aggressive when due to *SDHB* mutations compared with other *SDH* mutations.

Primary Aldosteronism. Patients having undergone adrenalectomy for primary aldosterone often experience changes in volume status and electrolytes in the postoperative setting. The degree of this is in part dependent on the severity of disease and degree of preoperative control of hypervolemia and hypokalemia with medication. Patients should be monitored for changes in potassium levels. The effects of excess aldosterone begin to resolve within the first 24 hours. Dependent on whether or not the patient was cured, whether there is excess contralateral gland aldosterone secretion, and what medications are taken, potassium levels may stabilize, decrease, or increase after surgery. Significant diuresis related to decreased aldosterone levels can result in hypokalemia. Longstanding suppression of aldosterone secretion from the contralateral adrenal gland can result in hyperkalemia.[106] Some patients may require fludrocortisone in the postoperative setting if potassium levels are significantly elevated or hyponatremia or volume depletion is noted. Changes in potassium levels usually become apparent within anywhere from 12 hours to 2 weeks. Intermittent assessment and appropriate handoffs to other providers are warranted.

Blood pressure should also be assessed in the postoperative setting. In most patients, all blood pressure and potassium supplements are usually stopped, aside from clonidine and similar medications (which would result in severe rebound hypertension) and/or beta-blockers if utilized for cardiac-related issues. Blood pressure should be tracked closely in the first few days to assess the need for reinstitution of antihypertensive medications. Different classes of drugs may be required, dependent on the situation and dependent on cure of the primary aldosterone.

Creatinine levels may increase and glomerular filtration rate may decrease acutely after surgery. While intraoperative hypotension or ligation of an accessory renal artery or other complications may result in changes in creatinine, it is more likely that creatinine increases due to a decrease in glomerular filtration as hypervolemia resolves. With excess aldosterone, sodium and water retention ensue. The volume cleared by nephrons decreases with the resolution of excess aldosterone, resulting in decreased glomerular filtration rate and increased creatinine.

The final outcome of surgery can best be determined much longer after surgery than previously thought. It is now recognized that aldosterone levels may change over several months after surgery, as can blood pressure, potassium, and need for medication. Most outcome studies report data based on measures taken within 2 weeks of surgery. It is likely that many outcomes reported in the past are inaccurately reported due to lack of long-term follow-up. In 2016, the Primary Aldosteronism Surgical Outcomes group established recommendations for assessing and reporting outcomes.[107] Outcomes are stratified by clinical and biochemical outcomes and at multiple timepoints, with the last measurement 1 year after surgery. Aldosterone, renin, blood pressure, potassium, and creatinine should be checked on an annual basis at the time of the annual visit with a primary care physician. Future studies will hopefully be able to more accurately report outcomes, as indications for surgery and goals for clinical and biochemical outcomes will likely change.

Hypercortisolism. Steroid replacement in patients with hypocortisolism after surgery is critically important and a lifesaving measure while awaiting return of HPA axis function. In those having undergone bilateral adrenalectomy, steroid replacement is a lifelong necessity. Patients should be counseled prior to discharge regarding when and how to take steroids to replicate the diurnal rhythm as much as possible, with an early morning surge. Doses after 5:00 PM are avoided if possible, as this can impact sleep. Timing of administration may vary depending on the patient's individual situation, occupation, and timing of work shifts. Administration twice daily is most common, but regimens once or three times a day may be pursued, depending on the patient. It is advisable to utilize hydrocortisone in the postoperative setting when weaning will occur instead of prednisone, given the shorter half-life and ability to wean more quickly.

In patients who have undergone unilateral adrenalectomy, temporary full replacement (15–20 mg hydrocortisone total daily, which is then weaned) is usually necessary in patients who presented with overt signs and symptoms of Cushing syndrome, or in those with 24-hour urine cortisol levels above the normal range and suppressed ACTH. Return of HPA axis function may take 1 to 2 years in these patients before they can be weaned off steroids.[108,109]

In those who have mild autonomous cortisol secretion based on a 1-mg overnight dexamethasone suppression test and normal 24-hour urine cortisol and are without overt signs of hypercortisolism, some patients will need steroid supplementation. Postoperative testing is important in these patients. How and when to test are variable, ranging from testing cortisol the morning after surgery to formal stimulation testing with a synthetic derivative of ACTH(1–24) (Cosyntropin) up to several days after surgery.[60,110-113] Signs of cortisol insufficiency include lethargy, fatigue, weakness, nausea, poor appetite, hyperkalemia, hyponatremia, and volume depletion.

Patients with overt hypercortisolism often have significant musculoskeletal weakness and sarcopenia. Physical therapy to aid in restoring patients to their condition prior to developing hypercortisolism is underutilized and should be pursued. Changes related to adverse effects of hypercortisolism are slow to resolve, but do in general over the first 2 years.[108,109,113,114] While the body adjusts to a return to normal cortisol levels and rhythms, subjective reports of fatigue, nausea, and poor appetite are normal but usually resolve. Patients should be counseled preoperatively regarding this expected time course.

Adrenocortical Carcinoma. Pathology should be reviewed, and results, concerns about intraoperative events, or surgery-related morbidities impacting future care should be discussed between providers, preferably as part of a multidisciplinary tumor board.[30] Staging should be completed if necessary depending on the extent of preoperative need for adjuvant therapies, including external beam radiation, mitotane, and/or chemotherapy, surveillance, and survivorship should be addressed. Patients with hormone excess prior to surgery should be treated as discussed above. For patients placed on mitotane after surgery as part of adjuvant therapy, institution of this should start approximately 6 weeks after surgery if the patient has sufficiently recovered. Gastrointestinal side effects are common, and ensuring adequate nutrition is imperative to allow for full recovery in the months after surgery. Thyroid function and liver function must be monitored when patients receive mitotane. In addition, given that mitotane is an adrenolytic, glucocorticoid replacement is necessary.

Bilateral Adrenalectomy

Patients undergoing bilateral adrenalectomy require steroid replacement to prevent death. In the vast majority of patients, both glucocorticoid and mineralocorticoid replacement are required. Mineralocorticoid replacement is usually not required in patients receiving daily doses of glucocorticoids greater than 40 to 50 mg daily of hydrocortisone or an equivalent dose of another type of glucocorticoid.

Postoperatively, hydrocortisone is preferred for replacement over prednisone, given faster recovery of the HPA axis. However, in the case of bilateral adrenalectomy where there is no chance of endogenous cortisol production, prednisone may be prescribed by some providers, although this is not routine. A multidisciplinary approach with input from an endocrinologist should be used on discharge, and patients should be provided information about signs of cortisol withdrawal and adrenal insufficiency or crisis and what to do. Electrolyte levels, blood pressure, and glucose levels should be monitored after discharge. A transition of care from the surgeon to the endocrinologist and primary care physician should occur in a few weeks. Good communication with all health care professionals involved is key to the safety of the patient. Patients should be encouraged to obtain some sort of medical alert bracelet or chain to reflect that adrenalectomy has been performed and that corticosteroids may be lifesaving. This information can also be entered into smartphone emergency contacts which are accessible to rescue crews and other medical personnel.

Nelson Syndrome

Hypertrophy of the pituitary gland may develop after bilateral adrenalectomy.[114] Dependent on the degree of growth, this may result in headaches, vision impairment, and the cessation of menstrual periods in women. ACTH and β-melanocyte–stimulating hormone are elevated. This phenomenon occurs in an estimated 20% (range 0%–47%) of patients undergoing bilateral adrenalectomy.[114] Adenomas of the pituitary may develop, in addition to gland hypertrophy and hyperplasia. These are treated with medication, radiation, and/or surgery.

KEY POINTS

- Goals of care of patients undergoing adrenalectomy can be addressed according to the specific disease, type of hormone excess, and phase of care, including preoperative, intraoperative, postoperative and postdischarge phases.

Additional Considerations

Adrenalectomy in Pregnancy. As with other diagnoses necessitating surgery during pregnancy, it is best if adrenalectomy is delayed until after delivery.[115] Should surgery be required, it is best performed during the second trimester. Most commonly, adrenal disorders encountered during pregnancy include pheochromocytoma, Cushing syndrome, and adrenal cortical carcinoma, all of which are extremely rare. For those with Cushing syndrome, the scant amount of literature available suggests adrenalectomy during the second trimester.[116] While surgery can be pursued during the second trimester for pregnant patients with pheochromocytoma, it is likely best to wait until delivery or afterward and instead manage the patient with good alpha, beta, and/or calcium channel blockade. To maintain adequate placental perfusion pressure, maternal systolic pressure should be maintained at or above 120 mm Hg. Should resection of a pheochromocytoma be pursued during pregnancy, surgical, anesthesia, obstetric, and neonatal teams should work closely together throughout the pregnancy, and especially when surgery is pursued. If gestation is less than 32 weeks, where the fetus will not be delivered, the multidisciplinary team must aggressively treat hypertensive episodes while maintaining placental perfusion pressure. Aggressive volume repletion should be pursued before and during surgery to limit significant hypotension and decreased placental perfusion once the adrenal vein is ligated, along with any other venous outflow. Additional treatment with pressors is warranted if inadequate placental perfusion is questioned. If gestation is greater than 32 weeks, C-section can be performed, with adrenalectomy performed immediately afterward. Evaluating and managing patients with indeterminate adrenal masses where ACC is in the differential, or in those with confirmed ACC, is complicated. The full spectrum of diagnostic studies is not available. Hormones associated with pregnancy can obscure the biochemical evaluation. The common recommendations regarding surgery during pregnancy remain so for ACC during pregnancy. Should significant growth be noted during pregnancy, or if hormone excess is unable to be well controlled, which would compromise the safety of the mother or fetus, surgery during the second trimester may be pursued. Should surgery become more urgent during the third trimester, waiting until at least 32 weeks' gestation is best to allow for production of fetal pulmonary surfactant and greater maturation of the fetal lungs.

Adrenalectomy in the Pediatric Population. The pediatric population sees a different spectrum of adrenal disease than the adult population. Most adrenal disease in pediatric patients is the result of genetic mutations causing dysfunction or neoplasia. The incidence is rare, and the need for adrenalectomy is variable, depending on the disease process. Most often, disorders of adrenal function are recognized in the setting of precocious puberty, ambiguous sex, or change in secondary sexual characteristics. Adrenocortical carcinoma has two peaks in presentation: one in childhood and the other in the fifth and sixth decades of life. Adrenalectomy in children is similar in technique to that in adults. Open and minimally invasive approaches, including anterior and posterior, may be utilized. Use of computer-aided assistance is difficult in children, given the limited working space and resulting collision of the arms of the robot. In older children, it may be more feasible. Timing of and extent of adrenalectomy when performed for benign disease should be carefully considered, and especially when bilateral adrenalectomy is necessary. Balancing the risk of persistent hormone or catecholamine excess should be balanced against the risk of need for lifelong steroid replacement. Compliance should be factored into the decision as well. Partial adrenalectomy, with consideration for the technical aspects of any future surgery, can provide a reasonable alternative in many patients, again dependent on the disease process being treated.

Considerations in Resource-Limited Settings. Third- world, developing, and other resource-limited countries see a different distribution of adrenal disorders compared with more developed countries. Patient presentation and methods for evaluation can be quite different. Access to surgeons familiar with adrenal disorders and hospitals with resources to carry out minimally invasive approaches to adrenalectomy may be limited due to lack of access to adequate imaging, laboratory, and other ancillary resources, laparoscopic and other operating room equipment or mentorship and training to be able to perform minimally invasive adrenalectomy. Organizations such as the World Health Organization, universities with international alliances, surgical organizations, and private organizations can provide mentorship and assist with education in new procedures. Patient presentation, methods for evaluation, and disease processes encountered in these regions may be quite different from those in more developed nations. In a study from India,[117] a developing region with fewer resources than some, but more compared with many countries, ultrasound was the most commonly utilized imaging modality (69.4%), as CT and MRI (20.3%) were not widely available. The mean tumor size was significantly larger than in Western countries. Aldosterone and renin levels, routinely available in developed countries, were rarely measured due to financial constraints. Inflammatory lesions accounted for 9% of presenting adrenal disorders, something not often seen in more developed nations.

CONCLUSION

There have been significant changes in adrenalectomy over the past 120 years. Multiple approaches exist, technology has advanced, and outcomes have improved; yet, many of the original approaches remain relevant today. Each approach has unique advantages and disadvantages compared with other approaches. Surgeons performing adrenalectomy need to be well-versed in both the medical and surgical aspects of adrenal disorders to provide optimal care to patients. In the future, continued efforts to understand the biology of adrenal hormones will change the surgical approach to patients. Exciting discoveries in patients with dysregulation of aldosterone and cortisol production will drive new indications for surgery and refine and/or remove surgery as the treatment of choice in others. Advances in understanding malignant adrenal disorders, coupled with new medical therapies, will allow for refinement of surgical approaches in these patients, too.

KEY POINTS

- A thorough understanding of relevant adrenal anatomy, physiology, and interpretation of radiologic findings is critical to facilitate sound decision-making regarding adrenalectomy.
- The surgical approach to adrenalectomy should be chosen carefully and customized to the individual patient, taking into consideration the specific disease process and type of hormone excess, radiologic findings, concern for malignancy, patient anatomy, and comorbidities.
- Goals of care for patients undergoing adrenalectomy can be addressed according to the specific disease, type of hormone excess, and phase of care, including preoperative, intraoperative, postoperative, and postdischarge phases.
- Surgeons need to ensure performance of the critical elements of resection for adrenocortical carcinoma, with surgery best carried out in high-volume centers, given that surgery for benign adrenal disorders is quite different than that for adrenocortical carcinoma.

REFERENCES

1. Hergus S, Lynn J. Surgical Anatomy and Surgery of the Adrenal Glands. In: Lynn J, Bloom S, eds. *Surgical Endocrinology*. London: Buttersworth/Heinemann; 1993:458–467.
2. Hench PS, Kendall EC, Slocumb CH, et al. The effect of a hormone of the adrenal cortex (17-hydroxy-11-dehydrocorticosterone; compound E) and of pituitary adrenocorticotropic hormone on rheumatoid arthritis. *Proc Staff Meet Mayo Clin*. 1949;24:181–197.
3. Conn JW, Louis LH. Primary aldosteronism, a new clinical entity. *Ann Intern Med*. 1956;44:1–15.
4. Oliver G, Schäfer EA. On the physiological action of extract of the suprarenal capsules. *J Physiol*. 1894;16:I–IV.
5. Ladd MR, Zeiger MA. Who was Dr. William C. Baum? World. *J Surg*. 2018;42:2437–2443.
6. Lapides J. An operative technique for adrenalectomy. *Bull NY Acad Med*. 1958;34:303–310.
7. Cade S. The role of adrenalectomy in cancer of the breast. *Cancer*. 1957;10:777–788.
8. Barker WF, Yuhl ET, Beal Jr JM, et al. Bilateral adrenalectomy in the treatment of advanced neoplastic disease. *West J Surg Obstet Gynecol*. 1953;61:491–498.
9. Tocantins R, Smith DR. The modified Garlock incision for adrenal surgery. *J Urol*. 1961;85:417–419.
10. Hudson PB. An extrapleural thoracolumbar surgical approach for adrenalectomy and for radical retroperitoneal dissection; *Ann Surg*. 1954;139(1):44–51.
11. Ruffolo LI, Nessen MF, Probst CP, et al. Open adrenalectomy through a Makuuchi incision: a single institution's experience. *Surgery*. 2018;164(6):1372–1376.
12. Gagner M, Lacroix A, Bolte E. Laparoscopic adrenalectomy in Cushing's syndrome and pheochromocytoma [letter]. *N Engl J Med*. 1992;327:1033.
13. Gagner M, Lacroix A, Prinz RA, et al. Early experience with laparoscopic approach for adrenalectomy. *Surgery*. 1993;114:1120–1125.
14. Assalia A, Gagner M. Laparoscopic adrenalectomy. *Br J Surg*. 2004;91:1259–1274.
15. Gumbs AA, Gagner M. Laparoscopic adrenalectomy. *Best Pract Res Clin Endocrinol Metab*. 2006;20:483–499.
16. Higashihara E, Tanaka Y, Horie S, et al. A case report of laparoscopic adrenalectomy. *Nippon Hinyokika Gakkai Zasshi*. 1992;83:1130–1133.
17. Lee J, El-Tamer M, Schifftner T, et al. Open and laparoscopic adrenalectomy: analysis of the national surgical quality improvement Program. *J Am Coll Surg*. 2008;206:953–959.
18. Elfenbein DM, Scarborough JE, Speicher PJ, et al. Comparison of laparoscopic versus open adrenalectomy: results from American College of surgeons-national surgery quality improvement project. *J Surg Res*. 2013;184:216–220.
19. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *Am J Roentgenol*. 2008;190: 1163–1168.
20. Vaidya A, Mulatero P, Baudrand R, et al. The expanding spectrum of primary aldosteronism: implications for diagnosis, pathogenesis, and treatment. *Endocr Rev*. 2018;39:1057–1088.
21. Debono M, Bradburn M, Bull M, et al. Cortisol as a marker for increased mortality in patients with incidental adrenocortical adenomas. *J Clin Endocrinol Metab*. 2014;99:4462–4470.
22. Bancos I, Alahdab F, Crowley RK, et al. Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and subclinical Cushing's syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175:R283–R295.
23. Fassnacht M, ArltW, Bancos I, et al. Management of adrenal incidentalomas: European society of endocrinology clinical practice guideline in collaboration with the European network for the study of adrenal tumors. *Eur J Endocrinol*. 2016;175:G1–G34.
24. Zeiger MA, Thompson GB, Duh QY, et al. The American association of clinical endocrinologists and American association of endocrine surgeons medical guidelines for the management of adrenal incidentalomas. *Endocr Pract*. 2009;15:1–20.

25. Tallis PH, Rushworth RL, Torpy DJ, et al. Adrenal insufficiency due to bilateral adrenal metastases - a systematic review and meta-analysis. *Heliyon*. 2019;5:e01783.
26. Tsujimoto A, Ueda T, Kuge H, et al. Long-term survival after adrenal metastasectomy from colorectal cancer: a report of two cases. *Surg Case Rep*. 2019;5:61.
27. Rashidi A, Fisher SI. Primary adrenal lymphoma: a systematic review. *Ann Hematol*. 2013;92:1583–1593.
28. Nassoro DD, Mkhoi ML, Sabi I, et al. Adrenal insufficiency: a forgotten diagnosis in HIV/AIDS patients in developing countries. *Intl J Endocrinol*. 2019; 2019:2342857. doi: 10.1155/2019/2342857.
29. He X, Modi Z, Else T. Hereditary causes of primary aldosteronism and other disorders of apparent excess mineralocorticoid activity. Gland Surg. 202;9:150–158.
30. Fassnacht M, Dekkers OM, Else T, et al. European society of endocrinology clinical practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European network for the study of adrenal tumors. *Eur J Endocrinol*. 2018;179:G1–G46.
31. Petr EJ, Else T. Adrenocortical carcinoma (ACC): when and why should we consider germline testing? *Presse Med*. 2018;47:e119–e125.
32. Lenders JWM, Duh Q-Y, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metabol*. 2014;99:1915–1942.
33. Gupta G, Pacak K. Precision medicine: an update on genotype/biochemical phenotype relationships in pheochromocytoma/paraganglioma patients. *Endocr Pract*. 2017;23:690–704.
34. Taylor DR, Ghataore L, Couchman L, et al. A 13-steroid serum panel based on LC-MS/MS: use in detection of adrenocortical carcinoma. *Clin Chem*. 2017;63:1836–1846.
35. Bancos I, Taylor AE, Chortis V, et al. Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. *Lancet Diabetes Endocrinol*. 2020;8:773–781.
36. Chortis V, Bancos I, Nijman T, et al. Urine steroid metabolomics as a novel tool for detection of recurrent adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2020;105:e307–e318.
37. Caoili EM, Korobkin M, Francis IR, et al. Adrenal masses: characterization with combined unenhanced and delayed enhanced CT. *Radiology*. 2002;222:629–633.
38. Caoili EM, Korobkin M, Francis IR, et al. Delayed enhanced CT of lipid-poor adrenal adenomas. *AJR Am J Roentgenol*. 2000;175:1411–1415.
39. Corwin MT, Navarro SM, Dania G, et al. Differences in Growth Rate on CT of adrenal adenomas and malignant adrenal nodules. *Am J Roentgenol*. 2019;213(3):632–636.
40. Haider MA, Ghai S, Jhaveri K, et al. Chemical shift MR imaging of hyperattenuating (>10 HU) adrenal masses: does it still have a role? *Radiology*. 2004;231:711–716.
41. Elsayes KM, et al. Adrenal masses: MR imaging features with pathologic correlation. *Radiographics*. 2004;24:S73–S86.
42. Caoili EM, Korobkin M, Brown RK, et al. Differentiating adrenal adenomas from nonadenomas using (18)F-FDG PET/CT: quantitative and qualitative evaluation. *Acad Radiol*. 2007;14:468–475.
43. Groussin L, et al. 18F-Fluorodeoxyglucose positron emission tomography for the diagnosis of adrenocortical tumors: a prospective study in 77 operated patients. *J Clin Endocrinol Metab*. 2009;94:1713–1722.
44. Leboulleux S, et al. Diagnostic and prognostic value of 18-fluorodeoxyglucose positron emission tomography in adrenocortical carcinoma: a prospective comparison with computed tomography. *J Clin Endocrinol Metab*. 2006;91:920–925.
45. Khan TS, et al. 11C-metomidate PET imaging of adrenocortical cancer. *Eur J Nucl Med Mol Imaging*. 2003;30:403–410.
46. Hahner S, et al. [123I]Iodometomidate for molecular imaging of adrenocortical cytochrome P450 family 11B enzymes. *J Clin Endocrinol Metab*. 2008;93:2358–2365.
47. Khan TS, et al. 11C--metomidate PET imaging of adrenocortical cancer. *Eur J Nucl Med Mol Imaging*. 2003;30:403–410.
48. Ryder SJ, Love AJ, Duncan EL, et al. PET detectives: molecular imaging for phaeochromocytomas and paragangliomas in the genomics era. *Clin Endocrinol (Oxf)*. 2021;95(1):13–28.
49. Jha A, Ling A, Millo C, et al. Superiority of 68Ga-DOTATATE over 18F-FDG and anatomic imaging in the detection of succinate dehydrogenase mutation (SDHx)-related pheochromocytoma and paraganglioma in the pediatric population. *Eur J Nucl Med Mol Imaging*. 2018;45:787–797.
50. Nockel P, El Lakis M, Gaitanidis A, et al. Preoperative 18F-FDG PET/CT in pheochromocytomas and paragangliomas allows for precision surgery. *Ann Surg*. 2019;269:741–747.
51. Bancos I, Tamhane S, Shah M, et al. Diagnosis of endocrine disease: the diagnostic performance of adrenal biopsy: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175:R65–R80.
52. Vanderveen KA, Thompson SM, Callstrom MR, et al. Biopsy of pheochromocytomas and paragangliomas: potential for disaster. *Surgery*. 2009;146:1158–1166.
53. Chomsky-Higgins K, Seib C, Rochefort H, et al. Less is more: cost-effectiveness analysis of surveillance strategies for small, nonfunctional, radiographically benign adrenal incidentalomas. *Surgery*. 2018;163:197–204.
54. Giordano TJ, Berney D, de Krijger RR, et al. Data set for reporting of carcinoma of the adrenal cortex: explanations and recommendations of the guidelines from the International Collaboration on Cancer Reporting. *Hum Pathol*. 2021;110:50–61.
55. Wang TS, Cheung K, Roman SA, et al. A cost-effectiveness analysis of adrenalectomy for nonfunctional adrenal incidentalomas: is there a size threshold for resection? *Surgery*. 2012;152:1125–1132.
56. Brogna A, Scalisi G, Ferrara R, et al. Giant secreting adrenal myelolipoma in a man: a case report. *J Med Case Reports*. 2011;5:298.
57. Tung GA, Pfister RC, Papanicolaou N, et al. Adrenal cysts: imaging and percutaneous aspiration. *Radiology*. 1989;173:107–110.
58. Nieman LK, Biller BM, Findling JW, et al. The diagnosis of cushing's syndrome: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1526–1540.
59. Nieman LK. Update on subclinical Cushing's syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:180–184.
60. Nieman LK, Biller BM, Findling JW, et al. Treatment of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2015;100(8):2807–2831.
61. Li D, El Kawkgi OM, Henriquez AF, et al. Cardiovascular risk and mortality in patients with active and treated hypercortisolism. *Gland Surg*. 2020;9:1–2.
62. Bancos I, Alahdab F, Crowley RK, et al. Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and subclinical Cushing's syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*. 2016;175:R283–R295.
63. Di Dalmazi G, Vicennati V, Garelli S, et al. Cardiovascular events and mortality in patients with adrenal incidentalomas that are either nonsecreting or associated with intermediate phenotype or subclinical Cushing's syndrome: a 15-year retrospective study. *Lancet Diabetes Endocrinol*. 2014;2:396–405.
64. Petramala L, Cavallaro G, Galassi M, et al. Clinical benefits of unilateral adrenalectomy in patients with subclinical hypercortisolism due to adrenal incidentaloma: results from a single center. *High Blood Press Cardiovasc Prev*. 2017;24:69–75.
65. Raffaelli M, De Crea C, D'Amato G, et al. Outcome of adrenalectomy for subclinical hypercortisolism and Cushing syndrome. *Surgery*. 2017;161:264–271.
66. Speiser PW, Azziz R, Baskin LS, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95(9):4133–4160.
67. St-Jean M, Ghorayeb NE, Bourdeau I, et al. Aberrant G-protein coupled hormone receptor in adrenal diseases. *Best Pract Res Clin Endocrinol Metab*. 2018;32:165–187.
68. Bourdeau I, Parisien-La Salle S, Lacroix A. Adrenocortical hyperplasia: a multifaceted disease. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101386.
69. Guerin C, Taieb D, Treglia G, et al. Bilateral adrenalectomy in the 21st century: when to use it for hypercortisolism? *Endocr Relat Cancer*. 2016;23:R131–R142.
70. Lowery AJ, Seeliger B, Alesina PF, et al. Posterior retroperitoneoscopic adrenal surgery for clinical and subclinical Cushing's syndrome in patients with bilateral adrenal disease. *Langenbeck's Arch Surg*. 2017;402:775–785.

71. Nanba AT, Nanba K, Byrd JB, et al. Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol*. 2017;87:665–672.
72. Funder JW, Carey RM, Mantero F, et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2016;101:1889–1916.
73. Rossi GP, Auchus RJ, Brown M, et al. An expert consensus statement on use of adrenal vein sampling for the subtyping of primary aldosteronism. *Hypertension*. 2014;63:151–160.
74. Proye C, Fossati P, Fontaine P, et al. Dopamine-secretingpheochromocytoma: an unrecognized entity? Classification of pheochromocytomas accordingto their type of secretion. *Surgery*. 1986;100(6):1154–1162.
75. Dubois LA, Gray DK. Dopamine-secreting pheochromocytomas: in search of a syndrome. *World J Surg*. 2005;29:909–913.
76. Wachtel H, Kennedy EH, Zaheer S, et al. Preoperative metyrosine improves cardiovascular outcomes for patients undergoing surgery for pheochromocytoma and paraganglioma. *Ann Surg Oncol*. 2015;22:646–654.
77. Steinsapir J, Carr AA, Prisant LM, et al. Metyrosine and pheochromocytoma. *Arch Intern Med*. 1997;157:901–906.
78. Chen Y, Hodin RA, Pandolfi C, et al. Hypoglycemia after resection of pheochromocytoma. *Surgery*. 2014;156:1404–1408.
79. Akiba M, Kodama T, Ito Y, et al. Hypoglycemia induced by excessive rebound secretion of insulin after removal of pheochromocytoma. *World J Surg*. 1990;14:317–324.
80. Reynolds C, Wilkins GE, Schmidt N, et al. Hyperinsulinism after removal of a pheochromocytoma. *Can Med Assoc J*. 1983;129:349–353.
81. Wiesner TD, Bluher M, Windgassen M, et al. Improvement of insulin sensitivity after adrenalectomy in patients with pheochromocytoma. *J Clin Endocrinol Metab*. 2003;s88:3632–3636.
82. Else T, Williams AR, Sabolch A, et al. Adjuvant therapies and patient and tumor characteristics associated with survival of adult patients with adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2014;99:455–461.
83. Bednarski BK, Habra MA, Phan A, et al. Borderline resectable adrenal cortical carcinoma: a potential role for preoperative chemotherapy. *World J Surg*. 2014;38:1318–1327.
84. Romero Arenas MA, Sui D, Grubbs EG, et al. Adrenal metastectomy is safe in selected patients. *World J Surg*. 2014;38:1336–1342.
85. Chen WC, Baal JD, Baal U, et al. Stereotactic body radiation therapy of adrenal metastases: a pooled meta-analysis and systematic review of 39 studies with 1006 patients. *Int J Radiat Oncol Biol Phys*. 2020;107:48–61.
86. Taffurelli G, Ricci C, Casadei R, et al. Open adrenalectomy in the era of laparoscopic surgery: a review. *Updates Surg*. 2017;69:135–143.
87. Miller BS. 'Old fashioned' open adrenalectomy. *World J Surg*. 2020;44:618–621.
88. Lee J, El-Tamer M, Schifftner T, et al. Open and laparoscopic adrenalectomy: analysis of the national surgical quality improvement program. *J Am Coll Surg*. 2008;206:953–959. Discussion 959–961.
89. Buell JF, Alexander HR, Norton JA, et al. Bilateral adrenalectomy for Cushing's syndrome. Anterior versus posterior surgical approach. *Ann Surg*. 1997;225:63–68.
90. Sidhu S, Bambach C, Pillinger S, et al. Changing pattern of adrenalectomy at a tertiary referral centre 1970-2000. *ANZ J Surg*. 2002;72:463–466.
91. Lo CY, Chan WF. From open to laparoscopic adrenalectomy: a review of 16-year experience. *Chin Med J Engl*. 1999;112:1080–1084.
92. Al-Qurayshi Z, Robins R, Buell J, et al. Surgeon volume impact on outcomes and cost of adrenal surgeries. *Eur J Surg Oncol*. 2016;42(10):1483–1490.
93. Ortega J, Sala C, Garcia S, et al. Cost-effectiveness of laparoscopic vs open adrenalectomy: small savings in an expensive process. *J Laparoendosc Adv Surg Tech*. 2002;12:1–5.
94. Walz MK, Peitgen K, Hoermann R, et al. Posterior retroperitoneoscopy as a new minimally invasive approach for adrenalectomy: results of 30 adrenalectomies in 27 patients. *World J Surg*. 1996;20:769–774.
95. Feng Z, Feng MP, Levine JW, et al. Robotic retroperitoneoscopic adrenalectomy: useful modifications of the described posterior approach. *J Robot Surg*. 2017;11:409–414.
96. Agha A, von Breitenbuch P, Gahli N, et al. Retroperitoneoscopic adrenalectomy: lateral versus dorsal approach. *J Surg Oncol*. 2008;97:90–93.
97. Chen W, Lin W, Han DJ, et al. Lateral retroperitoneoscopic adrenalectomy for complicated adrenal tumor larger than 5 centimeters. *Afr Health Sci*. 2017;17:293–300.
98. Feng Z, Feng MP, Feng DP, et al. A cost-conscious approach to robotic adrenalectomy. *J Robot Surg*. 2018;12:607–611.
99. Nanba AT, Nanba K, Byrd JB, et al. Discordance between imaging and immunohistochemistry in unilateral primary aldosteronism. *Clin Endocrinol*. 2017;87:665–672.
100. Turcu AF, Gomez-Sanchez CE. Segmental adrenal vein sampling in patients with primary aldosteronism: superlative or superfluous? *Hypertension*. 2020;76:662–664.
101. Miller BS, Doherty GM. Adrenocortical Carcinoma. In: Katz MHG, Hunt KK; Veeramachaneni N, Roland C, Crago A, et al., eds. *Operative Standards for Cancer Surgery, Volume 3: Sarcoma, Adrenal, Neuroendocrine, Peritoneal Malignancies, Urothelial, Hepatobiliary*. Wolters Kluwer; 2023: 73–140.
102. Erdogan I, Deutschbein T, Jurowich C, et al. The 378 role of surgery in the management of recurrent adrenocortical carcinoma. *J Clin Endocrinol Metab*. 2013;98:181–191.
103. Glenn JA, Else T, Hughes DT, et al. Longitudinal patterns of recurrence in patients with adrenocortical carcinoma. *Surgery*. 2019;165:186–195.
104. Anderson Jr KL, Thomas SM, Adam MA, et al. Each procedure matters: threshold for surgeon volume to minimize complications and decrease cost associated with adrenalectomy. *Surgery*. 2018;163:157–164.
105. Kazaure HS, Sosa JA. Volume-outcome relationship in adrenal surgery: a review of existing literature. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101296.
106. Shariq OA, Bancos I, Cronin PA, et al. Contralateral suppression of aldosterone at adrenal venous sampling predicts hyperkalemia following adrenalectomy for primary aldosteronism. *Surgery*. 2018;163:183–190.
107. Williams TA, Lenders JWM, Mulatero P, et al. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabetes Endocrinol*. 2017;5:689–699.
108. Sippel RS, Elaraj DM, Kebebew E, et al. Waiting for change: symptom resolution after adrenalectomy for Cushing's syndrome. *Surgery*. 2008;144:1054–1060; discussion 1060–1061.
109. Doherty GM, Nieman LK, Cutler Jr GB, et al. Time to recovery of the hypothalamic-pituitary-adrenal axis after curative resection of adrenal tumors in patients with Cushing's syndrome. *Surgery*. 1990;108:1085–1090.
110. Foster T, Bancos I, McKenzie T, et al. Early assessment of postoperative adrenal function is necessary after adrenalectomy for mild autonomous cortisol secretion. *Surgery*. 2021;169:150–154.
111. Ortiz DI, Findling JW, Carroll TB, et al. Cosyntropin stimulation testing on postoperative day 1 allows for selective glucocorticoid replacement therapy after adrenalectomy for hypercortisolism: results of a novel, multidisciplinary institutional protocol. *Surgery*. 2016;159:259–265.
112. Ortiz DI, Findling JW, Carroll TB, et al. Cosyntropin stimulation testing on postoperative day 1 allows for selective glucocorticoid replacement therapy after adrenalectomy for hypercortisolism: results of a novel, multidisciplinary institutional protocol. *Surgery*. 2016;159:259–265.
113. Miller BS, Auchus RJ. Evaluation and treatment of patients with hypercortisolism: a review. *JAMA Surg*. 2020;155:1152–1159.
114. Ritzel K, Beuschlein F, Mickisch A, et al. Clinical review: outcome of bilateral adrenalectomy in Cushing's syndrome: a systematic review. *J Clin Endocrinol Metab*. 2013;98:3939–3948.
115. Bancos I, Atkinson E, Eng C, et al. Maternal and fetal outcomes in phaeochromocytoma and pregnancy: a multicentre retrospective cohort study and systematic review of literature. *Lancet Diabetes Endocrinol*. 2021;9:13–21.
116. Brue T, Amodru V, Castinetti F. Management of endocrine disease: management of Cushing's syndrome during pregnancy: solved and unsolved questions. *Euro J Endocrinol*. 2018;178:R259–R266.
117. Bhargav PR, Mishra A, Agarwal G, et al. Adrenal incidentalomas: experience in a developing country. *World J Surg*. 2008;32:1802–1808.

101

Differences of Sex Development

Eric Vilain and Kim Shimy

OUTLINE

Introduction to Sexual Development, 1686
Embryology, 1686
  Gonadal Development (Sex Determination), 1686
  Development of Internal and External Genitalia (Sex Differentiation), 1688
Genetic Pathways of Sex Development and Their Variations, 1691
  Genetic Pathways, 1691
  Genetic Variants Associated with Differences of Sex Development, 1691
Differences of Sex Development, 1692
  46,XX Differences of Sex Development, 1693
  46,XY Differences of Sex Development, 1702
  Sex Chromosome Differences of Sex Development, 1707
  Syndromic Differences of Sex Development, 1708
Evaluation and Management of Differences of Sex Development, 1708
  Tools in Differences of Sex Development Evaluation, 1708
  Initial Evaluation of a Newborn with Atypical Genital Appearance, 1710
  Initial Gender Assignment, 1710
  Management of Differences of Sex Development, 1711
Acknowledgements, 1712

INTRODUCTION TO SEXUAL DEVELOPMENT

One of the most defining moments of our lives is when, in the womb, we embark on a male, female, or intermediate path. Variations of typical male or female development result in differences of sex development (DSDs), also referred to as intersex conditions. They occur frequently, in approximately 1% of the human population. DSDs are phenotypically heterogeneous, ranging from minor genital malformations (hypospadias, cryptorchidism, hypertrophy of the clitoris) to genital ambiguity. DSD can result in adverse consequences for mental health, fertility, cancer risk, and quality of life. For families, the birth of a child with a DSD, and the accompanying uncertainty about future psychological and sexual development, is believed to be extraordinarily stressful.

Recently, the debate over clinical management of DSD, in particular gender assignment and genital surgery, has intensified; yet the scientific data on patient outcomes have remained very incomplete. Major obstacles to optimal clinical management of DSD include gaps in understanding of pathophysiology, which impedes precise diagnostic categorization, along with the absence of prospective longitudinal studies of health and quality-of-life outcomes. Understanding the underlying mechanisms of sexual development is therefore crucial to the management of DSD. This chapter explores the embryological, genetic, and endocrine pathways underlying sexual development and their variations leading to DSD.

EMBRYOLOGY

Development of the gonads and genitalia comprises two major phases and has a sexually dimorphic timetable. In the earliest stages of gestation, the bipotential gonads develop into either ovaries or testes (sex determination). Testicular development is completed by approximately 8 weeks' gestation, but ovarian development does not begin until around 10 to 11 weeks' gestation. Development of these reproductive primordia during the first 6 weeks of human gestation is dependent on a network of interacting factors encoded by many genes. In the second stage, sex differentiation, two sets of undifferentiated embryonic internal and external genitalia develop along male or female lines, depending upon the nature of the hormonal secretions from the gonads and tissue responsiveness to those hormones. These processes occur between 8 and 14 weeks' gestation in the male fetus and extend into the middle of the second trimester in the female fetus. Each developmental phase is described separately in the sections that follow.

Gonadal Development (Sex Determination)

From a Bipotential Gonad to a Testis or an Ovary. The bipotential gonads hold a unique place in fetal development as the only organs that can pursue a dual developmental course into either a testis or an ovary. Each cell type within the bipotential gonad exhibits this dual potential, and as such an expression pattern is orchestrated within each cell type to promote the predetermined phenotype and actively repress the alternative pathway.[1,2] In the first few weeks of embryonic development, there are no discernable differences between the gonads and genitalia of 46,XY and 46,XX embryos. The renal, adrenal, and gonadal cell precursors all arise from the same regions of the primitive mesoderm and coelomic epithelium known as the adrenogenital primordium.[3] This region condenses to form a ridge on the dorsal wall of the body cavity, often referred to as the *urogenital ridge*. The most medial aspect of this urogenital ridge is the gonadal primordium (Fig. 101.1).[4]

As development progresses, the primary processes focus on cellular proliferation, and the transcriptional program required for this intense increase in cellular mass is shared by the adrenogenital primordium. Concurrently, between the fifth and sixth weeks of gestation (mouse

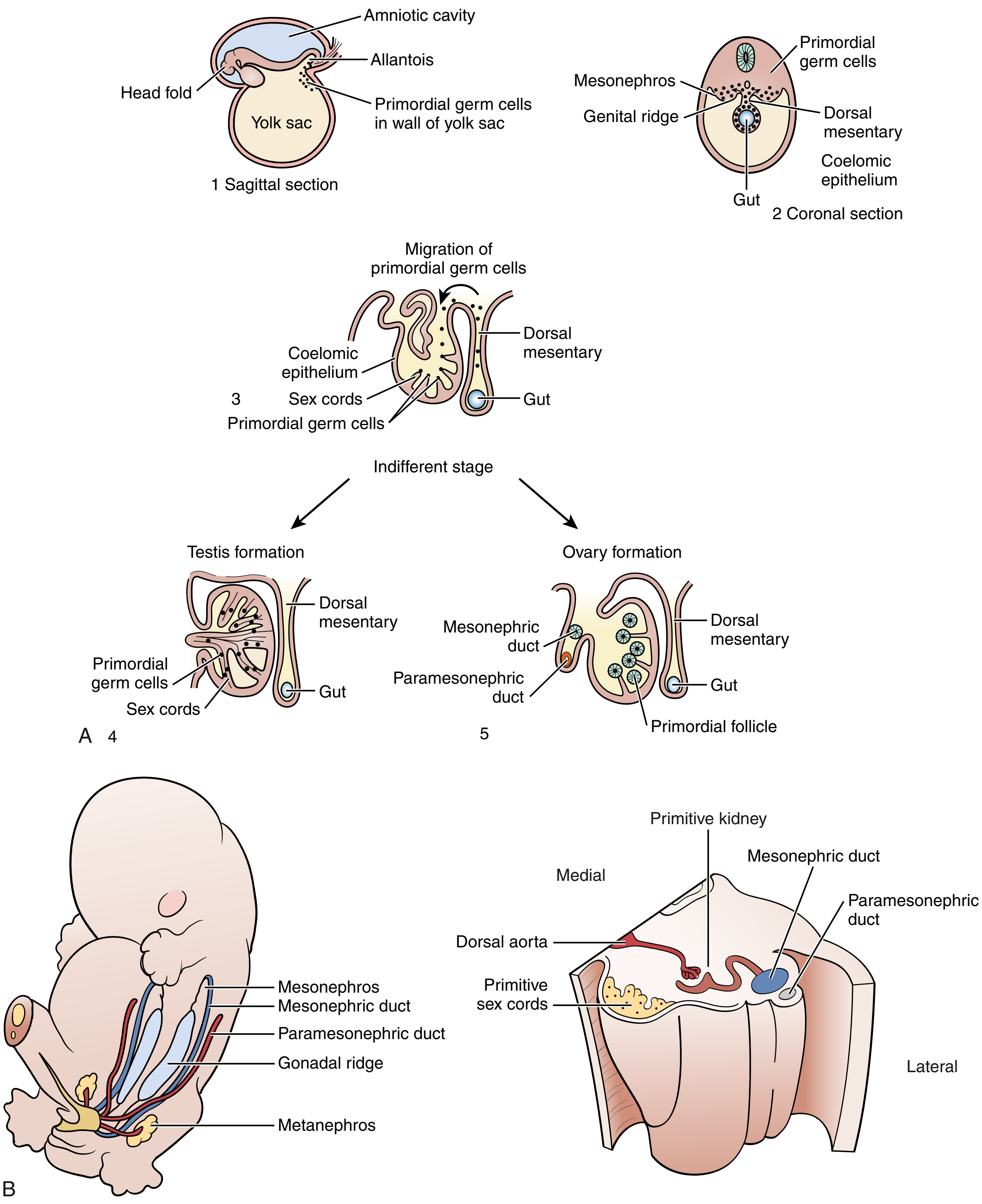

Amniotic cavity
Allantois
Head fold
Yolk sac
Primordial germ cells in wall of yolk sac
1 Sagittal section
Primordial germ cells
Mesonephros
Genital ridge
Dorsal mesentary
Coelomic epithelium
Gut
2 Coronal section
Migration of primordial germ cells
Dorsal mesentary
Coelomic epithelium
3
Sex cords
Gut
Primordial germ cells
Indifferent stage
Testis formation
Ovary formation
Dorsal mesentary
Primordial germ cells
Sex cords
Gut
A
4
Mesonephric duct
Paramesonephric duct
Dorsal mesentary
Gut
Primordial follicle
5
Mesonephros
Mesonephric duct
Paramesonephric duct
Gonadal ridge
Metanephros
B
Primitive kidney
Mesonephric duct
Medial
Paramesonephric duct
Dorsal aorta
Primitive sex cords
Lateral

Figure 101.1 **A,** Embryology of the gonads. *1–2,* Primordial germ cells are located in the wall of the yolk sac in early embryogenesis. *3,* At about the fifth to sixth week of human gestation, germ cells migrate by ameboid movement from the coelomic epithelium, along the dorsal mesentery of the hindgut, to the genital ridge, the site of the future gonad. During formation of the indifferent (bipotential) gonad, the primordial germ cells infiltrate the primary sex cords, which subsequently become the seminiferous tubules of the testis or the primordial follicles of the ovary. *4–5,* Formation of the testis continues after arrival of the germ cells, although these are not required for the process to occur. Probably under the direction of the newly differentiated Sertoli cells, the primitive sex cords differentiate as the seminiferous tubules, in which germ cells are embedded. Morphologic development of the testis is complete by 8 weeks' gestation. Formation of the ovary occurs later than that of the testis, at around week 10 of human gestation, and requires the presence of germ cells. The primary sex cords develop into primordial follicles at approximately week 16 of gestation. **B,** Anatomic relationships between renal and gonadal primordia during embryogenesis. *Left*, Diagram of human fetus at 5 to 6 weeks' gestation. The primordial genital ridges are located just medial to the mesonephros, the primitive fetal kidney. The mesonephros is functional only during early fetal life and is replaced later in gestation by the metanephros (located more caudally in this diagram), which becomes the definitive kidney. The mesonephric duct, which later becomes the Wolffian duct, runs in a craniocaudal direction just lateral to the mesonephros. The paramesonephric duct, which later becomes the Müllerian duct, runs almost parallel to the mesonephric duct, a little farther laterally on the posterior abdominal wall. *Right*, Partial horizontal section of the fetus at 5 to 6 weeks' gestation, showing (from medial to lateral) the primitive sex cords (containing primordial germ cells), primitive kidney, mesonephric duct, and paramesonephric duct.

embryonic day [E]10.5–12), primordial germ cells migrate by ameboid movement from the endoderm of the yolk sac along the dorsal mesentery of the hindgut and into the bipotential gonad, where they invade the developing primary sex cords. At this stage, the gonad comprises an outer cortex and inner medulla, and there is still no morphologic difference between the gonads of male and female fetuses. Following the arrival of germ cells, the gonad begins to differentiate. As development progresses and populations of cells begin to differentiate along specific paths, there is a requirement for a much more focused and precise program of transcription-factor action, which must be actively maintained.[1,2] Despite the earlier belief that once sex determination occurred it was fixed, recent studies in mice have shown that the determination of cell fate within the developing gonad is reversible, and active repression of the alternate pathway is required throughout life.[1,2] However, whether this is true in humans has not been ascertained. In addition to cell types, several other aspects of gonadal development are sexually dimorphic, including blood vessel growth and the initiating factors that push the bipotential gonad toward development as a testis or an ovary.

Development of the Testis. The first discernible event in testis development is the appearance of primordial Sertoli cells, which differentiate from somatic cells of the coelomic epithelium at approximately 7 weeks of human gestation. Soon after their appearance in the gonadal ridge, Sertoli cells proliferate, aggregate around the primitive germ cells, and align into cordlike structures that subsequently become the seminiferous tubules (see Fig. 101.1).[4-6] The seclusion of germ cells within the tubules prevents meiosis and commits the germ cells to spermatogenic development. The prevention of meiosis may be a key event that directs gonadal development away from the ovarian pathway, as it is believed that testis cord formation and germ cell entry into meiosis are competing pathways in gonad development.

Another key event in testis development is the differentiation of peritubular myoid cells—smooth muscle-like cells that ensheath the testis cords to provide structural integrity—thought to derive from the same interstitial cell lineage as Leydig cells. These myoepithelial cells are required for the development of the testis cords, the defining event in testicular organogenesis.[7] Approximately 1 week after Sertoli cell development (around 8 weeks in human embryos), steroidogenic Leydig cells differentiate from primitive interstitial cells of mesonephric origin. This process may be controlled by paracrine influences from Sertoli cells, possibly anti-Müllerian hormone (AMH). Germ cells are not required for the processes of testis development to occur; morphologic testis development occurs in their absence.

Finally, establishment of testis-specific vasculature enables testicular hormones to be transported to the developing fetal internal and external genitalia to direct masculinization. Initially, the primitive vasculature of the genital ridge is similar in XX and XY gonads, until E11.5 in the mouse (approximately equivalent to 6 weeks of gestational age in humans). Between E11.5 and E12.5, XY gonads establish a pattern of blood flow distinct from that of XX gonads, with the development of a large male-specific coelomic vessel and a network of finely branched vessels.[8] Recruitment of vascular cells from the mesonephros is essential for the development of seminiferous tubules[9-11] and does not occur in XX gonads. The process of blood vessel development in the gonads is highly sexually dimorphic. The female gonad recruits vasculature by a classic angiogenic process. However, in the developing testis, vasculature is patterned by a remodeling mechanism that begins with the breakdown of an existing mesonephric vessel, followed by the migration and the reaggregation of XY endothelial cells to form the coelomic vessel.[12] The growth rate of the vasculature is similar between XX and XY gonads, but XY gonads show a distinct arterial specification during early testis development, as demonstrated by the male-specific pattern of gonadal expression of ephrin-B2, a marker of developing arteries.[13]

During the second trimester of human gestation, testicular volume increases exponentially, fueled primarily by doubling of Sertoli cell numbers every 2 weeks throughout this period. This rapid increase in Sertoli cells is accompanied by a parallel increase in germ cell number and an increase in Leydig cells.[7]

Development of the Ovary. In contrast with the testes, normal ovarian differentiation requires the presence of germ cells.[14,15] Without germ-cell seeding of the gonadal primordium, the tissue degenerates into a nonfunctional, mainly fibrous, “streak.” While the process of testis formation is complete by 8 weeks’ gestation, the process of ovarian development has not yet begun.[16] Only after 6 to 8 weeks of human gestation does ovarian structure begin to become distinguishable, with the development of the primary medullary sex cords. Secondary cortical sex cords provide the supporting structure for the germ cells arriving from the gut. Within the sex cords, the primary germ cells undergo vigorous mitotic replication to become oogonia; then the sex cords break up into clusters, becoming the primordial follicles at approximately week 16. The primordial follicles contain the diploid (i.e., 46 chromosomes) primary oocytes, which, after entering into the first stage of meiosis (reduction division), remain quiescent until puberty. Follicular (granulosa) cells arise from the same somatic cell lineage as Sertoli cells, and theca cells represent the ovarian counterpart of Leydig cells. The fact that female genital development is normal even in the absence of the ovary highlights the lack of involvement of the ovary in the processes of normal female sex differentiation. The embryologic development of the testis and ovary are summarized in Fig. 101.1.

In the 46,XX fetus, the cortex develops under the influence of the germ cell lineage, and the medulla regresses; in the 46,XY fetus, the reverse occurs, with development of the medullary sex cords under the influence of the supporting cell (pre-Sertoli) lineage (see Fig. 101.1).[17] The sexually dimorphic development of germ cells is dictated in large part by retinoid signaling.[18] The mesonephros produces large amounts of retinoic acid. In the developing male gonad, the cytochrome p450 enzyme Cyp26b1, which is expressed in somatic cells, degrades retinoids, and the primordial germ cells undergo mitotic arrest until later in development when they resume proliferation. In contrast, the developing female gonad does not express Cyp26b1, allowing retinoids to induce a genetic program (i.e., stimulated by retinoid acid 8) that results in germ cell entry into meiosis, where they arrest in meiosis I until stimulated to resume meiosis during follicle development in the adolescent and adult.

Development of Internal and External Genitalia (Sex Differentiation)

Male Sex Differentiation. Both the urinary and genital components of the urogenital duct systems are derived primarily from the intermediate mesoderm, which becomes segmented into units termed *nephrotomes.* By approximately week 4 of human gestation, the lateral portions of the nephrotomes unite, forming the mesonephric duct (later to become the Wolffian duct). At approximately 6 weeks’ gestation, the paramesonephric duct (forbear of the Müllerian duct) forms by invagination of a tube from the epithelium on the surface of the urogenital ridge, and by 6 weeks’ gestation, both male and female fetuses are endowed with two sets of internal reproductive duct structures[4,19] (see Figs. 101.1B and 101.3).

At week 4 of gestation, the external genitalia of both sexes are represented simply by a midline protuberance—the genital tubercle. By week 6, two medial folds—the urethral folds—flank the urogenital groove, and two larger folds—the labioscrotal folds—are present

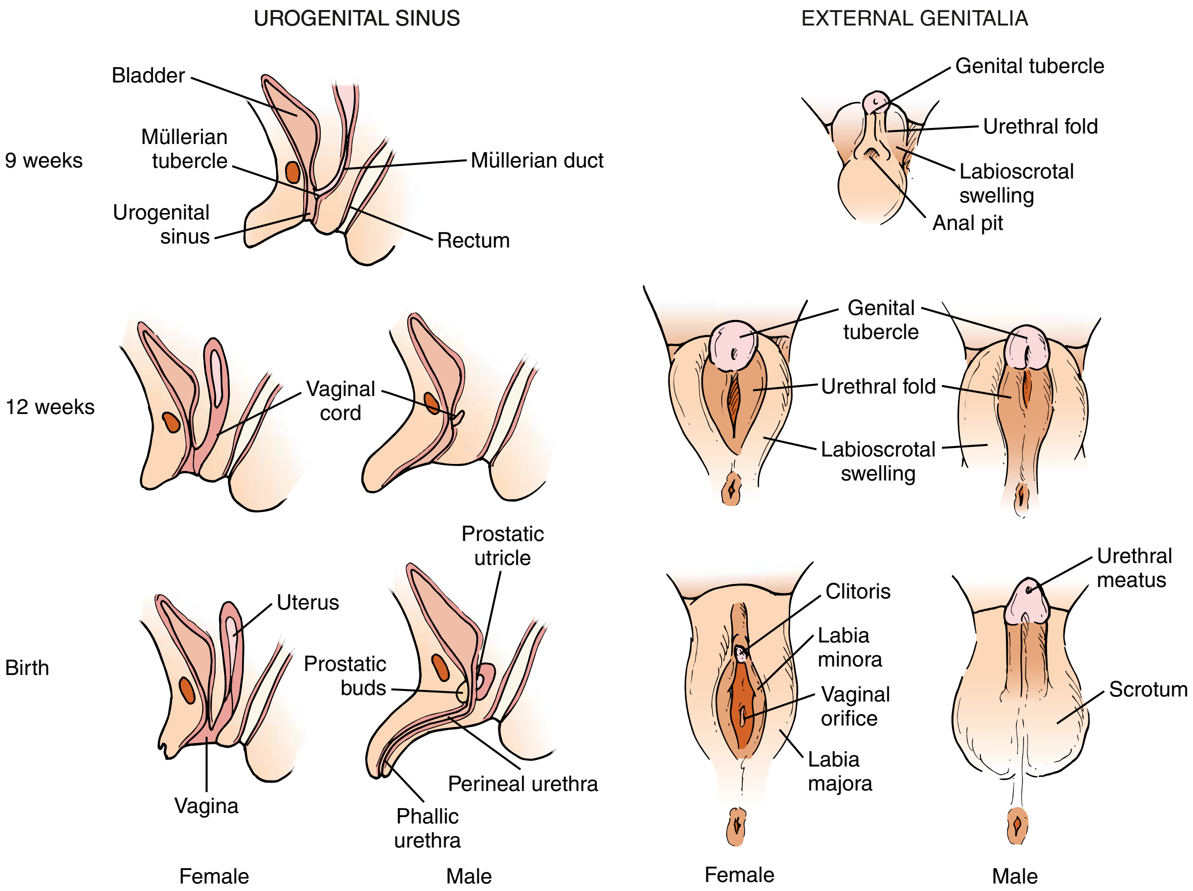

UROGENITAL SINUS
EXTERNAL GENITALIA
9 weeks
Bladder
Müllerian tubercle
Urogenital sinus
Müllerian duct
Rectum
Genital tubercle
Urethral fold
Labioscrotal swelling
Anal pit
12 weeks
Vaginal cord
Genital tubercle
Urethral fold
Labioscrotal swelling
Birth
Uterus
Prostatic utricle
Prostatic buds
Perineal urethra
Phallic urethra
Vagina
Clitoris
Labia minora
Vaginal orifice
Labia majora
Urethral meatus
Scrotum
Female
Male
Female
Male

Figure 101.2 Sex differentiation of urogenital sinus *(left)* and external genitalia *(right)*. (From Rey R, Josso N, Racine C. Sexual Differentiation. 2020 May 27. In: Feingold KR, Anawalt B, Boyce A, et al editors. Endotext [Endotext.org]. South Dartmouth (MA): MDText.com, Inc.; 2000–. PMID: 25905232.)

laterally[19,20] (see Fig. 101.2). Subsequent sex-specific development is described in the sections that follow.

Once function of the testis is established at approximately 8 weeks of human gestation, development of the male sexual phenotype is under control of testicular secretions and the receptors that mediate their action. These include fetal luteinizing hormone (LH), the LH/chorionic gonadotropin (CG) receptor (LHCGR), at least five steroidogenic enzymes involved in testosterone biosynthesis, steroid 5α-reductase 2, required for conversion of testosterone to dihydrotestosterone (DHT), the androgen receptor (AR), AMH, and the AMH receptor.

From 8 weeks' gestation, hormonal secretions from the fetal testes induce masculinization of the internal genital structures.[19,21] Development of Wolffian structures requires testosterone, which is produced by testicular Leydig cells following activation of the cell surface LHCGR. Between 9 and 13 weeks' gestation, they undergo differentiation into the epididymides, vasa deferentia, and seminal vesicles. During the first trimester of gestation, the ligand for this receptor is probably CG, produced in large quantity by the placenta. Subsequently, endogenous fetal LH is the primary ligand. The lack of dependence upon fetal LH during genital morphogenesis is evidenced by the normal penile development of male infants with hypopituitarism; however, the subnormal penile size of these infants highlights the role of LH/testosterone action in the penile growth that normally occurs in the third trimester. Action of CG or LH at the Leydig cell LHCGR stimulates testicular steroidogenesis. Sertoli cells may also help regulate Leydig cell secretion in a paracrine fashion. The biochemistry and molecular biology of the steroidogenic enzymes is described in further detail later. The outcome of this multistep process is a high local (and to a lesser extent, systemic) concentration of testosterone. DHT does not appear to mediate the processes of Wolffian duct development, as the enzyme required for its production (5α-reductase 2) is not expressed in Wolffian tissues at the time of their differentiation.

In parallel to the masculinization of the internal genitalia represented by Wolffian development, a process of "defeminization" of the alternate set of genital ducts occurs as the Müllerian ducts regress under the influence of the locally acting glycoprotein hormone, AMH, secreted by testicular Sertoli cells, acting via a complex of serine/threonine kinase receptors. Action of AMH at its receptor results in involution of Müllerian structures by apoptosis. The Müllerian ducts are obliterated by week 11 of gestation, the only remnant of their existence in the 46,XY fetus being the prostatic utricle. Absence of one testis results in retention of the ipsilateral Müllerian structures and only limited Wolffian development on that side, indicating that the effects of AMH and testosterone on the internal genitalia are mediated largely in a paracrine fashion[16] (see Fig. 101.3).

Meanwhile, under the influence of androgen action (primarily the 5α-reduced derivative of testosterone, DHT), the genital tubercle elongates to form the body of the penis, and the urethral folds fuse ventrally from behind forward to form the penile urethra.[16,19,20] The labioscrotal swellings/folds grow toward each other, fusing in the midline to form the scrotum (see Fig. 101.2). DHT also induces the urogenital sinus to differentiate as the prostate and inhibits the formation of the vesicovaginal septum (described in the section entitled "Sex Differentiation"). These processes are completed by week 12 of gestation. Testosterone and DHT action are mediated by the AR, a

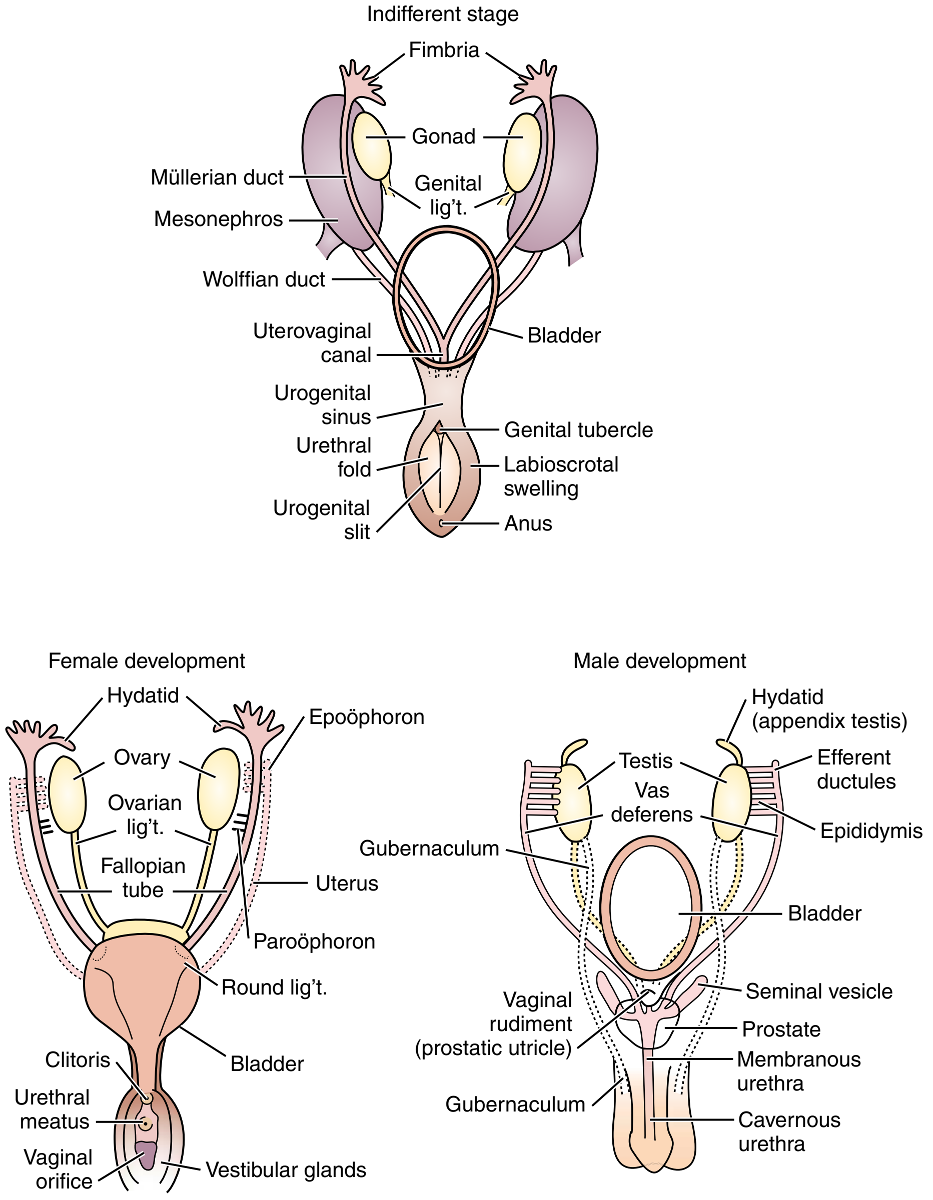

Indifferent stage
Fimbria
Gonad
Müllerian duct
Genital lig't.
Mesonephros
Wolffian duct
Uterovaginal canal
Bladder
Urogenital sinus
Genital tubercle
Urethral fold
Labioscrotal swelling
Urogenital slit
Anus
Female development
Hydatid
Epoöphoron
Ovary
Ovarian lig't.
Fallopian tube
Uterus
Paroöphoron
Round lig't.
Clitoris
Bladder
Urethral meatus
Vaginal orifice
Vestibular glands
Male development
Hydatid (appendix testis)
Testis
Efferent ductules
Vas deferens
Epididymis
Gubernaculum
Bladder
Seminal vesicle
Vaginal rudiment (prostatic utricle)
Prostate
Membranous urethra
Gubernaculum
Cavernous urethra

Figure 101.3 **Embryologic development of the internal genitalia.** *Top,* Indifferent stage. The primitive internal genitalia at 7 weeks' gestation have both paramesonephric or Müllerian (female) and mesonephric or Wolffian (male) duct systems. *Lower left,* Female development. In the 9- to 13-week female fetus, the Wolffian ducts regress *(dotted lines)*, and Müllerian ducts *(shaded)* differentiate and develop. Upper portions of Müllerian ducts form fallopian tubes; lower portions fuse to form the uterus, cervix, and upper part of the vagina. The urogenital sinus remains patent in the absence of androgen action, forming the lower part of the vagina. *Lower right,* Male development. Differentiation of the Wolffian ducts in the male fetus occurs under the influence of testosterone between 8 and 13 weeks' gestation. These develop as the epididymides, vasa deferentia, and seminal vesicles *(shaded)*. By week 11, Müllerian ducts *(dotted lines)* are obliterated by a process of apoptosis that occurs under the direction of anti-Müllerian hormone (AMH) secreted by Sertoli cells.

nuclear receptor/transcription factor found in high concentration in the tissues of the Wolffian ducts and external genitalia (where DHT is the ligand). In the absence of high androgen concentration, as in the female (or androgen-deficient) fetus, there is insufficient activated AR to induce transcription of target genes required for stabilization and development of the Wolffian system and for masculinization of the external genitalia.

Between 12 and 24 weeks' gestation, the testes migrate from their original lumbar location to the level of the internal inguinal ring above the scrotum. Descent of the testes through the inguinal ring and into the scrotum, controlled by testosterone, begins around week 28, and in most infants is completed by term. Testicular descent is also controlled by a Leydig cell–secreted protein, insulin-like protein 3, which belongs to the insulin-like hormone superfamily (including insulin, relaxin, and insulin-like growth factors 1 and 2).

Female Sex Differentiation. In the absence of testicular secretions, as in the 46,XX fetus, a set of genital tract developmental processes inverse to those described for the male occurs.[20,21] The Wolffian ducts regress, a phenomenon long thought to be caused by the absence of androgen, but recently shown to be actively promoted by COUP-TFII, suppressing mesenchyme-epithelium crosstalk responsible for

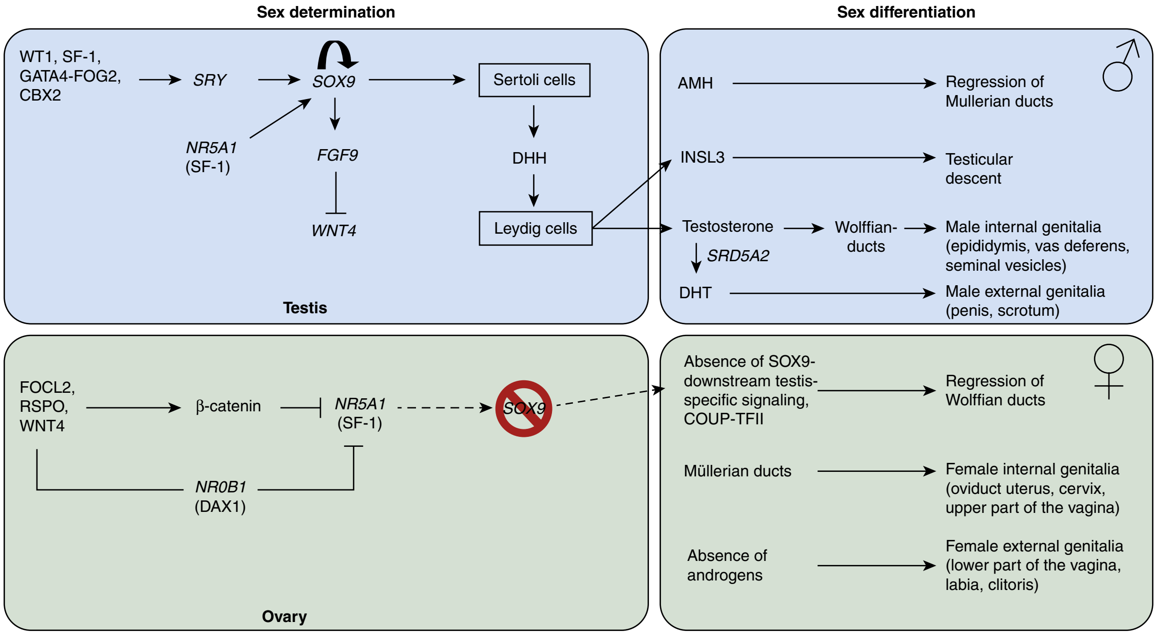

Sex determination
Sex differentiation
WT1, SF-1, GATA4-FOG2, CBX2
SRY
SOX9
Sertoli cells
NR5A1 (SF-1)
FGF9
DHH
WNT4
Leydig cells
Testis
AMH
Regression of Mullerian ducts
INSL3
Testicular descent
Testosterone
Wolffian-ducts
Male internal genitalia (epididymis, vas deferens, seminal vesicles)
SRD5A2
DHT
Male external genitalia (penis, scrotum)
FOCL2, RSPO, WNT4
β-catenin
NR5A1 (SF-1)
SOX9
NR0B1 (DAX1)
Ovary
Absence of SOX9-downstream testis-specific signaling, COUP-TFII
Regression of Wolffian ducts
Müllerian ducts
Female internal genitalia (oviduct uterus, cervix, upper part of the vagina)
Absence of androgens
Female external genitalia (lower part of the vagina, labia, clitoris)

Figure 101.4 Simplified overview of sex determination and sex differentiation in humans. (Adapted from Parivesh A, Barseghyan H, Délot E, et al. Translating genomics to the clinical diagnosis of disorders/differences of sex development. *Curr Top Dev Bio* 2019:317–375.)

Wolffian duct maintenance.[22] In the absence of AMH, the Müllerian ducts are permitted to develop. Their upper portions form the fallopian tubes; the lower sections fuse and differentiate as the uterus and upper part of the vagina (see Fig. 101.3). The lower portion of the vagina derives from the urogenital sinus (the ventral part of the embryonic mammalian cloaca formed by the growth of a fold dividing the cloaca where the gut and allantois meet), which remains patent when not exposed to androgen. External genital development in the female fetus (or androgen-deficient or resistant male) begins with the elongation of the genital tubercle to form the clitoris, and the urogenital sinus remains open (see Fig. 101.2). The vesicovaginal septum (a fold of tissue that develops in the female to separate the posterior wall of the bladder from the anterior wall of the vagina) forms, so that the urethra opens anteriorly and the vagina posteriorly.[16,19,20] The vestibule of the urogenital sinus is bordered laterally by the urethral folds, which do not fuse and instead develop as the labia minora. Further lateral, the labioscrotal swellings enlarge somewhat but also remain unfused, forming the labia majora. There is minor labial fusion posteriorly, forming the posterior commissure, and anteriorly, producing the mons pubis. These events occur from about week 7 to week 12 of gestation.

KEY POINTS

- Sex determination is the initial developmental choice leading to testis or ovary from the undifferentiated, bipotential gonad. Sex differentiation refers to the steps subsequent to sex determination, including the development of the internal and external genitalia.

GENETIC PATHWAYS OF SEX DEVELOPMENT AND THEIR VARIATIONS

Genetic Pathways

The strategy of sexual development is divided into two sequential stages: sex determination and sex differentiation (Fig. 101.4). Sex determination starts with the sex chromosome complement, inherited from the parental gametes, which in turn prompts a path of molecular signaling events in the presence or absence of the Y chromosome, driving the bipotential gonad in the developing embryo towards a male or female fate, respectively. Expression of Y-linked gene *SRY* in the somatic cells of the bipotential gonad of an XY embryo kickstarts a cascade of signaling events that lead to testicular organogenesis. In the absence of *SRY*, female-specific pathways are initiated, triggering ovarian development under the control of genes such as *RSPO1, WNT4*, and *DAX1/NR0B1*.

Sex differentiation follows the development of a testis or an ovary from a bipotential gonad. Under the influence of hormones and signaling molecules secreted by the testicular or ovarian tissues, sex-specific morphogenesis of the reproductive organs (external and internal) occurs and is named sex differentiation. Disruption of these developmental processes and of the enzymatic cascades involved in the synthesis of steroid hormones in the adrenals and gonads, as well as in the central control of these processes, can lead to DSD. The molecular mechanisms involved in this pathways have been reviewed in detail in numerous publications.[23-27] We will focus on the clinical applications of the variations of genetic pathways.

Genetic Variants Associated with Differences of Sex Development

A major issue to the understanding of the role of genes in DSD is to make sense of the immense number of variants identified in genes involved in sexual development. Some of them will be frequent in the general population and therefore likely benign, while others will be rarer and more likely to be pathogenic. We have summarized, in Table 101.1, the variants reported in ClinVar, the repository of clinical genomic variation curated by the National Institutes of Health. Sixty-nine genes associated with a DSD phenotype are presented (commonly present in diagnostic panels). For each gene, the type of variant (missense, nonsense, frameshift, larger INDELs, and rearrangements) is reported.

TABLE 101.1 **Variants Involved in Differences of Sex Development**

Genes	Variants	Variant Category
Sex Determination		
BMP15 (POF4)	3*	Missense, nonsense
CBX2	2	Missense
DHH	9	FS del, nonsense, missense
DMRT1	1	Missense
DMRT2	none	-
FSHR	17	Missense
GATA4	1	Missense
HHAT	none	-
MAP3K1	9	Missense, splice site
NR0B1 (DAX1)	75	FS del/ins, rearrangement, gene deletion, gene duplication, missense, nonsense, splice site
NR5A1 (SF1)	26	Missense, nonsense, in-frame deletion, FS del
RSPO1	3	Exon deletion, FS ins, splice site
SOX3	3	Gene duplication, missense, in-frame duplication
SOX9	14	Promoter del, promoter dup/triplication, FS ins, FS del, missense, nonsense
SRY	27	Missense, FS del, in-frame del, nonsense
STAG3 (POF8)	4	FS del, FS dup, splice site, nonsense
WNT4	4	Missense
WT1	23	Splice site, missense, nonsense, FS del, FS dup
WWOX	none	-
ZFPM2	3	Missense
Sex differentiation		
AKR1C2	4	Missense
AKR1C4	none	-
AMH	4	FS del, FS ins, nonsense
AMHR2	4	FS del, in-frame del, splice site, missense
AR	109	Splice site, missense, FS ins/del, nonsense, multi exon deletion
ARX	2	Nonsense, FS del
ATRX	24	Splice site, Nonsense, FS del, missense
CYP11A1	10	FS del, FS ins, in-frame insertion, missense
CYP17A1	35	Splice site, in-frame deletion, FS del, FS dup, rearrangement, missense, nonsense
CYP19A1	13	fusion, insertion, deletion, missense SNV, intron SNV, missense SNV, nonsense SNV
CYP21A2	49	Deletion, duplication, SNVs, insertion, intron SNV, missense SNV
DHCR7	90	Missense, FS ins, in-frame deletion, nonsense, Splice site
FGFR2	3	Intragenic deletion, missense
FOXL2 (POF 3)	5	FS del, FS ins, in-frame deletion, missense
HSD17B3	23	Missense, splice site, FS del, SNV deleting a stop codon
HSD3B2	11	FS ins, FS del, nonsense, missense
KDM5D	none	-
LHCGR	32	FS del, in-frame deletion, in-frame insertion, missense, nonsense, splice site
MAMLD1	4	Nonsense
POR	14	Splice site, missense, FS ins, FS dup, deletion
SRD5A2	41	FS del, missense, nonsense, splice site, exon loss deletion, in-frame deletion
STAR	33	FS del, FS ins, splice site, missense, nonsense
VAMP7	none	-

FS, frameshift; *ins*, insertion; *del*, deletion; *dup*, duplication; *SNV*, single-nucleotide variation.
Adapted from Parivesh A, Barseghyan H, Délot E, et al. Translating genomics to the clinical diagnosis of disorders/differences of sex development. *Curr Top Dev Bio*. 2019:317–375.

KEY POINTS

- A complex network of genetic pathways contributes to sexual development. Variations in the genes involved in sex determination and sex differentiation contribute to variations in the phenotypes of the gonads and genitalia.

DIFFERENCES OF SEX DEVELOPMENT

The term DSD, defined as "congenital conditions in which development of chromosomal, gonadal, or anatomical sex is atypical,"[28,29] was adopted following the 2005 consensus meeting of the Lawson Wilkins Pediatric Endocrine Society and the European Society for Paediatric Endocrinology. It replaces controversial terms such as hermaphroditism, pseudohermaphroditism, and intersex. This umbrella term covers a wide variety of phenotypes (e.g., hypospadias, discordance between external genitalia and chromosomal sex); hormonal phenotypes (e.g., androgen insensitivity, over-/undersecretion of androgens); gonadal phenotypes (e.g., gonadal dysgenesis, ovotestis); and chromosomal complements (e.g., 46,XY, 46,XX, 47,XXY, 45,X). DSD can occur in isolated or syndromic forms, which are presented in Table 101.2 and reviewed elsewhere.[30,31] There has been considerable debate about the word "disorder" within DSD, with positive aspects described in the Global DSD Update,[32] such as simplified access to insurance coverage and a clearer distinction from conditions such as gender dysphoria, but also negative connotations such as stigma often accompanying the word "disorder." For this chapter and, recognizing the negative social impact of the word "disorder," we have elected to replace it with the word "difference," which has been increasingly used in the literature and by advocacy organizations, and which does not modify the well-recognized acronym DSD.

The classification of DSD is based on the physiopathology and categorized by karyotype as 46,XX DSD, 46,XY DSD, and sex chromosome DSD (Table 101.3).

46,XX Differences of Sex Development

DSD in 46,XX individuals can be broadly categorized as 1) differences of gonadal development (including 46,XX testicular DSD, 46,XX ovotesticular DSD, and 46,XX pure gonadal dysgenesis), 2) 46,XX DSD due to androgen excess (resulting from virilizing congenital adrenal hyperplasias [CAHs], maternal-derived androgens, and aromatase

TABLE 101.2 Syndromic Forms of Differences of Sex Development (DSD)

Syndromes Associated with Differences of Sex Development	Mutated Gene	Phenotype	OMIM
46,XY DSD with congenital heart defect	*GATA4*		615542
Aarskog syndrome	*FGD1*	Short stature, facial dysmorphism, brachydactyly. Scrotal anomalies in XY	305400
Axenfeld–Rieger type 1	PITX2	Hypospadias	180500
Bardet–Biedl syndrome (AR, oligogenic?)	20+ loci	Retinitis pigmentosa, obesity, intellectual disability, renal anomalies, polydactyly, genital anomalies	209900
Blepharophimosis-ptosis-epicanthus inversus syndrome type 1 (AD, AR)	*FOXL2*	With or without premature ovarian failure	110100
Borjeson–Forssman–Lehmann syndrome	Unknown	Severe mental defects, epilepsy, hypogonadism, hypometabolism, marked obesity	301900
Campomelic dysplasia	*SOX9*	Ambiguous genitalia, congenital bowing of long bones, hypoplastic scapulae, thoracic vertebrae pedicles. Sex reversal in XY	114290
CHARGE syndrome (AD)	*CHD7*	Choanal atresia, intellectual disability; inner ear, retinal, and genital abnormalities in XX and XY	214800
Cornelia de Lange syndrome	*NIPBL*	Low anterior hairline, anteverted nares, maxillary prognathism, long philtrum, "carp" mouth in association with prenatal and postnatal growth restriction, intellectual disability. Genital anomalies in XY	122470
Cystic fibrosis	*CFTR*	Congenital absence of vasa deferentia	277180
Denys–Drash syndrome	*WT1*	Early diffuse mesangial sclerosis of kidneys, Wilms' tumor. Genital and gonadal anomalies in XX and XY	194080
DSD with adrenal hypoplasia XY GD	*NR5A1/SF1*	Hypoplasia, XY gonadal dysgenesis	300200
Dubowitz syndrome	Unknown	Short stature, microcephaly, mild intellectual disability with behavior problems, eczema, distinctive facies	223370
Faciogenitopopliteal syndrome	Unknown	Cleft palate, webbing of the intercural pterygium	119500
Frasier syndrome	*WT1*	XY and XY gonadal dysgenesis, early adolescent progressive glomerulopathy, gonadoblastoma	136680
Genitourinary and/or brain malformation syndrome (GUBS) (AD)	*PPP1R12A* (*de novo*)	Seizures, holoprosencephaly, genital anomalies in XY	602021
Goldenhar syndrome (hemifacial microsomia)	Unknown	Unilateral deformity of the external ear and small ipsilateral half of the face, with epibulbar dermoid, vertebral anomalies, and congenital cardiac anomalies	164210
Gonadal dysgenesis with neuropathy	*DHH*	XY GD with associated minifascicular neuropathy (one of four cases)	607080
Hand-Foot-Genital Syndrome	*HOXA13*	Carpal, tarsal, digit anomalies, Müllerian fusion in XX, hypospadias in XY	140000
Holoproscencephaly	Many	Craniofacial dysmorphology	236100
Hyperterlorism with esophageal abnormality and hypospadias	Unknown	Laryngotracheoesophageal cleft; clefts of lip, palate, and uvula; difficulty swallowing, hoarse cry; genitourinary defects, intellectual disability; congenital heart defects	145410
IMAGe syndrome	*CDKN1C* (imprinted)	Intrauterine growth restriction, metaphyseal dysplasia, adrenal hypoplasia congenita, genital anomalies	614732

Continued

TABLE 101.2 Syndromic Forms of Differences of Sex Development (DSD)—cont'd

Syndromes Associated with Differences of Sex Development	Mutated Gene	Phenotype	OMIM
Johanson–Blizzard syndrome	Unknown	Aplasia or hypoplasia of the nasal alae, congenital deafness, hypothyroidism, postnatal growth delay, malabsorption, intellectual disability, midline ectodermal scalp defects, absent permanent teeth	243800
Juberg–Marsidi Syndrome	Unknown	Cleft lip/palate, with abnormal thumbs and microcephaly	216100
Kallmann syndrome	*FGFR-1,* oligogenic + 24 others	Hypogonadotropic hypogonadism and anosmia	308700
Lenz–Majewski hyperstosis syndrome	Unknown	High palate; short, yellow, carious teeth; progeroid appearance; short stature; increased venous pattern of the forehead and thorax	151050
Lowe syndrome (XLR)	*OCRL*	Hydrophtalmia, cataract, intellectual disability, vitamin D–resistant rickets, amino aciduria and cryptorchidism	309000
Malpeuch facial clefting syndrome	Unknown	Short stature, hypertelorism, eye anomalies, facial clefting, hearing loss, urogenital abnormalities, intellectual disabilities	248340
Maturity onset diabetes of the young type 5 (MODY-5)	*HNF1B* (also named *TCF-2*)	Müllerian aplasia and renal dysfunction	137920
McKusick–Kaufman syndrome	Unknown	Hydometrocolpos, congenital heart malformations, postaxial polydactyly	236700
Miller–Dieker Syndrome	17p13.3 deletion	Lissencephaly, growth restriction, intellectual disability, +/- heart, face, kidney anomalies. Genital anomalies in XY	247200
MURCS = Mayer–Rokitansky–Kuster–Hauser Syndrome (MRKH) with developmental anomalies (in XX)	*WNT4* (in subset of MRKH with hyperandrogenism)	Müllerian duct aplasia, renal agenesis, cervicothoracic somite anomalies, deafness.	277000
MURCS syndrome	Unknown	Müllerian duct aplasia, renal aplasia, cervicothoracic somite dysplasia	601076
Najjar syndrome	Unknown	Genital anomaly, intellectual disability, cardiomyopathy	212120
Noonan syndrome	*PTPN11*	Hypertelorism, downward eye slant, and low-set posteriorly rotated ears, short stature, webbed neck, cardiac anomalies. Cryptorchidism in XY	163950
Opitz G/BBB	*MID1*	Hypospadias	300000
Pallister–Hall Syndrome (AD)	*GLI3*	Hypothalamic hamartoblastoma, pituitary dysfunction, postaxial polydactyly, other developmental anomalies. Genital anomalies in XY	146510
Palmoplantar hyperkeratosis with squamous cell carcinoma and XX sex reversal (AR)	*RSPO1*	Congenital bilateral corneal opacities, onychodystrophy, and hearing impairment	610644
Prader–Willi Syndrome	*SNRPN*, chr. 15q11.2 imprinting	Obesity, muscular hypotonia, intellectual disability, short stature, hypogonadotropic hypogonadism in XY; hypoplastic labia, amenorrhea in XX	176270
Robinow syndrome	*ROR2*	Mesomelic limb shortening, facial dysmorphism. Genital anomalies in XX and XY	180700
Rubinstein-Taybi syndrome	*CREBBP*	Intellectual disability, broad thumbs and toes, facial abnormalities. Hypospadias, shawl scrotum in XY	180849
Seckel syndrome	*SCKL1*	Short stature, microcephaly with intellectual disability, characteristic "birdheaded" facial appearance. Hypospadias in XY, clitoromegaly in XX	210600
Smith–Lemil–Opitz syndrome (AR)	*DHCR7*	Multiple congenital malformations and intellectual disability, genital abnormalities in XY	270400
Torticollis, keloids, cryptochidism, renal dysplasia	Unknown	Torticollis, keloids, cryptochidism, renal dysplasia	314300
Ubiquitous glucose-6-phosphatase deficiency	*G6PC3*	Severe congenital neutropenia plus cardiovascular and/or urogenital abnormalities	611045
VACTERL/VACTER	Unknown	Vertebral defects, anal atresia, tracheoesophageal fistula with esophageal atresia, radial dysplasia, limb anomalies	192350
Varadi–Papp syndrome	Unknown	Big toes, hexadactyly, cleft lip/palate or lingual nodule, somatic and psychomotor disability; some show absent olfactory bulbs and tracts, cryptochidism	277170
Velocardiofacial	Del 22q11.2 (*TBX1*)	Hypospadias, Müllerian aplasia, and renal dysfunction	192430
X-linked alpha-thalassemia/intellectual disability	*ATRX*	Hemoglobin H disease, intellectual disability, dysmorphic facies, hypogonadism	301040

XLR, X-linked recessive; *AR*, autosomal recessive; *AD*, autosomal dominant.

TABLE 101.3 Classification of Differences in Sex Development (DSD)

Sex Chromosome DSD	46,XY DSD	46,XX DSD
45,X (Turner Syndrome and variants)	*Differences of gonadal (testicular) development* • complete gonadal dysgenesis (Swyer syndrome) • partial gonadal dysgenesis • gonadal regression • ovotesticular DSD	*Differences of gonadal (ovarian) development:* • ovotesticular DSD • testicular DSD (e.g., *SRY*$^+$, duplicate *SOX9*) • gonadal dysgenesis
47,XXY (Klinefelter syndrome and variants)	*Disorders in androgen synthesis or action* • androgen biosynthesis defect (e.g., 17β-hydroxysteroid dehydrogenase deficiency, 5αRD2 deficiency, StAR mutations) • defect in androgen action (e.g., complete androgen insensitivity syndrome, partial androgen insensitivity syndrome) • luteinizing hormone receptor defects (e.g., Leydig cell hypoplasia, aplasia)	*Androgen excess* • fetal (e.g., 21-hydroxylase deficiency, 11β-hydroxylase deficiency) • fetoplacental (aromatase deficiency, POR [P450 oxidoreductase]) • maternal (luteoma, exogenous, etc.)
45,X/46,XY (Mixed gonadal dysgenesis, ovotesticular DSD)	*Disorders of anti-Müllerian hormone and anti-Müllerian hormone receptor* (persistent Müllerian duct syndrome)	*Other* (e.g., cloacal exstrophy, vaginal atresia, MURCS [Müllerian, renal, cervicothoracic somite abnormalities], other syndromes)
46XX/46,XY (chimeric, ovotesticular DSD)	*Other* (e.g., complex syndromic disorders)	

deficiency), and 3) other (such as disorders of Müllerian development and syndromic disorders).

46,XX Differences in Gonadal Development

46,XX Testicular Differences of Sex Development. This condition, previously called XX male or XX sex reversal, is characterized by testicular development and complete virilization of the internal and external genitalia in subjects who have two X chromosomes and lack a Y chromosome. Approximately one in 20,000 to 30,000 males have a 46,XX karyotype. Most cases are sporadic, and a form of X-Y paternal interchange is suggested when part of the short arm of the Y chromosome carrying *SRY* is transferred to one of the two X chromosomes.

The high sequence identity and identical orientation of *PRKY* to its homologous gene on the X chromosome, *PRKX*, seem to explain the high frequency of abnormal pairing and subsequent ectopic recombination outside the pseudoautosomal region, leading to 46,XX males and 46,XY females (Fig. 101.5).[33] *SRY*(+) XX males usually have a normal male phenotype with no particular feature during childhood. A small proportion (~10%) of XX males lack *SRY* sequences. The underlying pathogenic mechanism has remained elusive for years, but the recent advent of high throughput genetic analyses has allowed the identification of familial forms of *SRY*(-) XX males associated with overexpression of *SOX3*, *SOX9*,[34–42] or *SOX10*[43] resulting from gene duplication. The interpretation should be cautious in *SRY*(-) 46,XX patients with ambiguous genitalia: insufficient sampling of the gonads may miss ovarian tissue.

Before puberty, testicular endocrine function does not seem to be affected in XX males.[44] The condition is often diagnosed at puberty or adulthood owing to small testes, gynecomastia, and sterility, associated with some degree of testosterone deficiency and elevated gonadotropins.[45]

Owing to the presence of two X chromosomes and the lack of Y chromosome genes involved in spermatogenesis, germ cells do not go through meiosis, which explains the small testis size and azoospermia, as is seen in Klinefelter syndrome. However, XX males show shorter stature than Klinefelter patients.[45,46]

Ovotesticular Differences of Sex Development. Ovotesticular (OT)-DSD, formerly known as "true hermaphroditism," a pejorative term that is no longer used but is still present in the older literature, is defined by the presence of both ovarian and testicular tissue in the same individual.

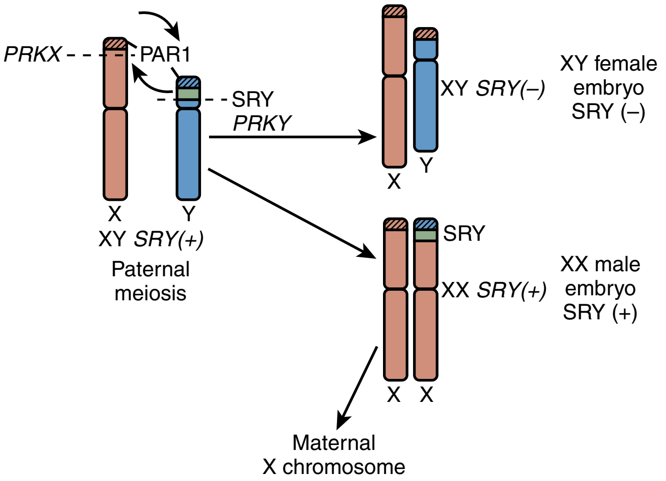


Figure 101.5 **Origin of *SRY*-negative XY females and *SRY*-positive XX males.** During meiosis in the paternal germ cells, an abnormal translocation occurs of *SRY* from the Y chromosome to the X chromosome (for example, at the level of the *PRKX* and *PRKY* genes). If the spermatozoon carrying an X chromosome with *SRY* fertilizes an oocyte, an *SRY*-positive XX embryo results. Conversely, if the spermatozoon carrying the *SRY*-negative Y chromosome fertilizes an oocyte, an *SRY*-negative XY embryo results.

The gonadal distribution of OT-DSD can be referred to as unilateral (ovotestis on one side and testis or ovary on the other), bilateral (ovotestis on each side), or lateral (testis on one side and ovary on the other).[47] The most common form is unilateral, with ovotestis/ovary in 34%, followed by bilateral (ovotestis/ovotestis) in 29% and lateral (ovary/testis) in 25%. The least frequent is unilateral with ovotestis/testis (12%).[48,49]

The degree of genital ambiguity depends on the mass of functional testicular tissue and may range from almost normal female to almost normal male appearance. In the absence of genital ambiguity, the diagnosis is reached at puberty because of virilization in patients reared as females, or gynecomastia and urethral bleeding in those raised male. Ovarian tissue is usually of good quality, and patients reared as females often experience normal pubertal development with menses[50]; pregnancy has been reported in a dozen cases, but only half have resulted in live births.[51] Testicular tissue progressively becomes dysgenetic. Two cases of spermatogenesis in males with OT-DSD have been reported.[48]

The geographical distribution of OT-DSD shows an overrepresentation in Africa, with the number of published cases at 17 per 100 million inhabitants, followed by Europe at 15.3. Asia is underrepresented, at 1.2 cases per 100 million, and South America at 3.6 per 100 million.[48]

The most frequent karyotype is 46,XX (65% of cases), and it is interesting to note that, in most cases, 46,XX patients do not carry the *SRY* gene,[52,53] and the cause remains unexplained genetically for the majority of them. Approximately 10% are XY, and the remaining are complex mosaics containing a Y chromosome.[48,49]

SRY translocated on one X explains approximately 10% of XX OT-DSD, and in some patients it is limited to gonadal mosaicism.[54] A small number of XY patients with OT-DSD carry point mutations in *SRY*.[55] It is likely that *SRY*(+) and *SRY*(-) OT-DSD and testicular DSD (T-DSD) are part of the same clinical spectrum, as OT-DSD and T-DSD can be observed in the same family, and with the same causative genotype, such as the presence of a translocated *SRY* gene.[56,57] The nonrandom inactivation of the *SRY*-bearing X chromosome has been proposed as an explanation for OT-DSD in *SRY*(+) XX OT-DSD.[58] The clinical features of OT-DSD are similar, whether or not the karyotype contains a Y chromosome, apart from the risk for gonadal tumors, which is higher if a Y chromosome is present.[59]

Recently, mutations in the *RSPO-1* gene have been identified in *SRY*(-) XX patients with OT-DSD associated with palmoplantar hyperkeratosis with squamous cell carcinoma of the skin.[60,61]

Partial 22q duplication (containing the *SOX10* gene) with OT-DSD (46,XX) has been described.[43] A case of a *de novo* heterozygous 3-Mb deletion on chromosome 15q26.2 has been reported in a 20-year-old with OT-DSD.[62] Duplication of the region of chromosome 17 containing *SOX9* has been identified as associated with OT-DSD, whether *SOX9* or a regulatory region upstream of *SOX9* is duplicated.[63-66]

There have been examples of specific point mutations in known sex-determining genes (*NR5A1*, encoding steroidogenic factor 1 [SF1], and Wilms' tumor 1 (*WT1*]) that typically cause XY gonadal dysgenesis associated with XX OT-DSD. The heterozygous missense mutation p.Arg92Trp in the part of *NR5A1* encoding an accessory DNA-binding region was found in cases of 46,XX OT-DSD.[67] This particular variant may switch the development pathway from ovarian to testicular tissue. Another example is heterozygous variants affecting the fourth zinc finger (ZF4) of the *WT1* gene. In a series of 78 patients with T-DSD and OT-DSD, all *SRY*(-), seven (approximately 9%) were found to have a variant in the ZF4 domain of WT1.[68] The occurrence of specific point mutations in *SRY*(-) OT-DSD suggests that a diagnostic approach including whole-exome sequencing is indicated.

46,XX Pure Gonadal Dysgenesis (Ovarian Dysgenesis). Ovarian dysgenesis can broadly be defined as conditions resulting from genetic defects in ovarian development, including impairment in early ovarian development (i.e., primordial follicle formation and differentiation) as well as defects in the later stages of follicle development at the time of puberty. Individuals with 46,XX pure gonadal dysgenesis are classically born with normal external female genitalia and diagnosed at puberty due primary amenorrhea, absent breast development, and elevated gonadotropins. It is important to distinguish ovarian dysgenesis from premature ovarian insufficiency (POI), an umbrella term encompassing a variety of conditions associated with impaired ovarian function, including ovarian dysgenesis (also see Chapter 122). Fifty to ninety percent of cases of POI have been historically been considered "idiopathic"; however, multiple candidate genes involved in ovarian development have been described in recent years, though not all have been confirmed in animal models.[69] X chromosome abnormalities (including Turner syndrome [TS], described later) are the most common genetic causes of POI. Premutations in the fragile X mental retardation 1 gene (*FMR1*), located on the long arm of the X chromosome at Xq27.3, have been associated with both sporadic and familial POI.[70] Haploinsufficiency of bone morphogenetic protein 15 (*BMP-15*) at Xp11.2 appears to cause X-linked ovarian dysgenesis in affected females, likely due to impairment of folliculogenesis.[71] Among autosomal genes, mutations in forkhead box L2 (*FOXL2*, encoding a factor essential for primary follicular development) are responsible for a syndromic cause of POI, blepharophimosis ptosis epicanthus inversus syndrome type 1, but have also been implicated in nonsyndromic POI.[72,73] POI has also been reported with mutations in the gene encoding PSMC3-interacting protein (PSMC3IP),[73] a nuclear protein involved in meiotic recombination and as a coactivator of estrogen-driven transcription, in NOBOX oogenesis homeobox (*NOBOX*),[74] an ovary-specific gene involved in signal transduction, and in stromal antigen 3 (*STAG3*),[75] a gene encoding a meiosis-specific subunit of the cohesin ring. Mutations in nuclear receptor subfamily 5 group A member 1 (*NR5A1*, also known as *SF1*), a modulator of ovarian steroidogenesis, have been identified in several cases of familial and idiopathic POI.[76] Mutations in the gonadotropin receptor,[77,78] estrogen receptor alpha,[79]and aromatase [80] genes impair ovarian function but do not appear to induce typical ovarian dysgenesis. POI also occurs in several genetic syndromes,[69] including ataxia telangiectasia, ovarian leukodystrophy, and Perrault syndrome (ovarian dysgenesis with deafness). However, in most cases of syndromic POI, it is unclear that POI has resulted from ovarian dysgenesis.

46,XX Differences of Sex Development due to Androgen Excess. Virilization of a 46,XX fetus occurs due to exposure to androgens in fetal development. Androgen excess causes clitoromegaly and fusion of the labioscrotal folds, leading the external genitalia to appear "ambiguous," or, in severe cases, completely male-appearing but without palpable testes (see Fig. 101.6A and B). The separation of urethra and vagina may also be disrupted, leading to a common urogenital sinus. DHT is the predominant androgen leading to labioscrotal fusion and clitorophallic growth. The degree of virilization depends both on the amount of exposure to DHT, as well as the time frame of exposure (labioscrotal fusion typically occurs when DHT is present before 12 weeks' gestational age, whereas clitorophallic growth may occur due to androgen exposure throughout gestation). The most common source of androgens leading to 46,XX virilization is the fetal adrenal, as a result of fetal virilizing CAH. Less common causes of 46,XX fetal virilization include aromatase deficiency and exposure to maternal-derived androgens.

Virilizing Congenital Adrenal Hyperplasias. CAH is a group of autosomal recessive disorders resulting from adrenal steroidogenic enzymatic deficiencies impacting cortisol production. Details of the incidence, molecular genetics, clinical presentation, and diagnosis of the forms CAH are discussed in detail in Chapter 94 but are briefly described here as well. Due to impaired cortisol synthesis, there is loss of negative feedback inhibition of the hypothalamic pituitary adrenal axis. The resultant increase in adrenocorticotropic hormone (ACTH) secretion overstimulates the adrenal cortex and drives accumulation of cortisol precursors (Fig. 101.7). In the virilizing forms of CAH, cortisol precursors are shunted towards adrenal androgens, leading to masculinization of external genitalia in 46,XX individuals. Affected females have normal ovarian development and, due to the absence of anti-Müllerian hormone, normal Müllerian duct development (Fig. 101.8). The degree of genital virilization in female infants with CAH is variable, ranging from mild clitoromegaly to male-appearing genitalia. Fusion of the labioscrotal

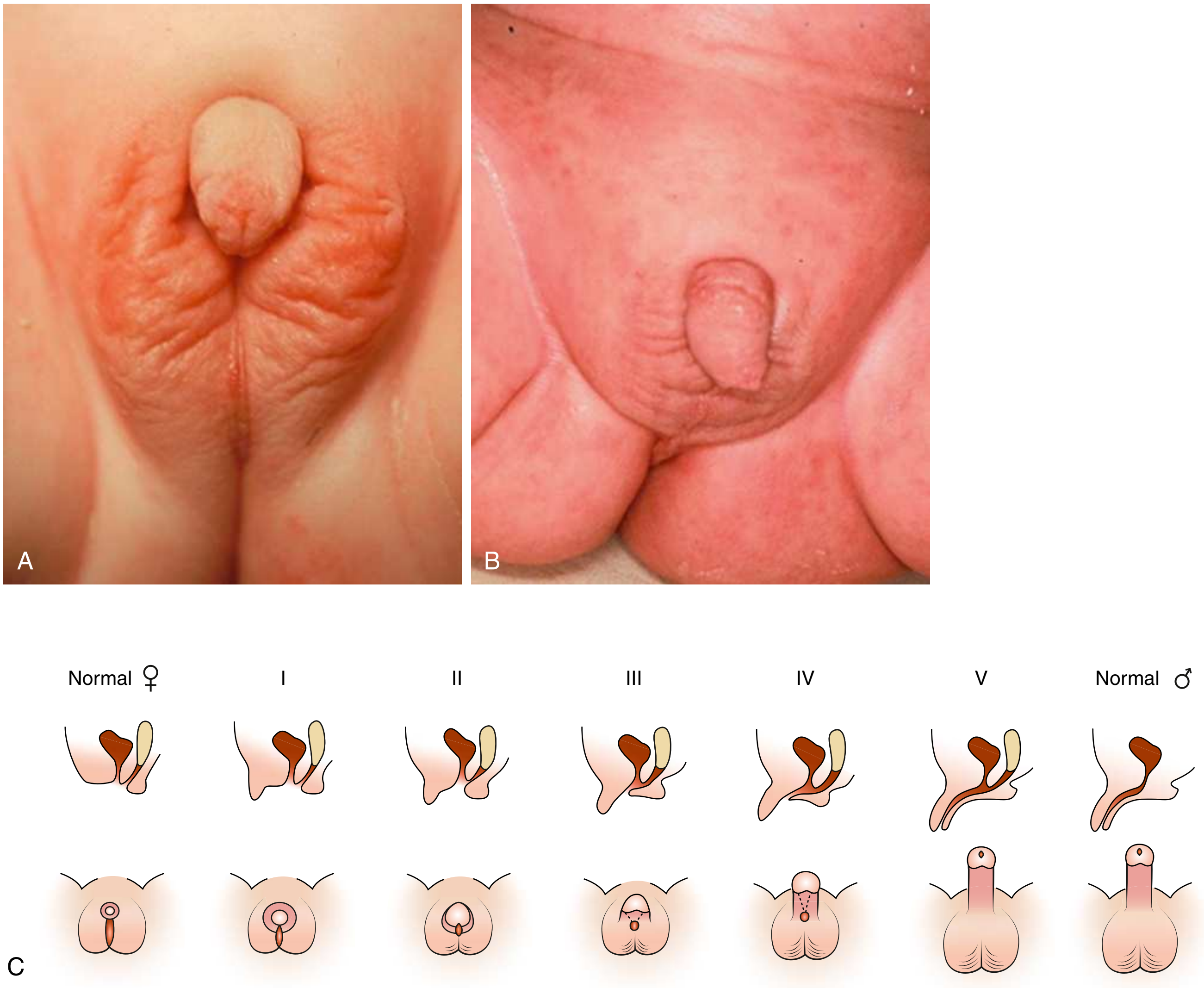


Figure 101.6 **Genitalia of 46,XX patients with congenital adrenal hyperplasia (CAH) resulting from 21-hydroxylase deficiency. A,** Prader stage 4. **B,** Prader stage 5. **C,** Schematic of Prader staging for patients with CAH. (**A** and **B** Kindly provided by Dr. M. Podestá, Buenos Aires. **C** Reprinted from Ogilvy-Stuart AL, Brain CE. Early assessment of ambiguous genitalia. *Arch Dis Child.* 2004;89;401–407, with permission from BMJ Publishing Group Ltd.)

folds may occur, leading to the appearance of male scrotum but without palpable testes. Internal lower urogenital anatomy may also be impacted, with the development of a common urogenital sinus comprising the confluence of the urethra and vagina. Rarely, there can be complete fusion of the labioscrotal folds resulting in a penis-like structure with normally positioned urethral meatus at the tip of the glans. Apart from elevated adrenal androgen precursors, other factors influencing the degree of genital virilization may include the degree of conversion of androgen precursors to more potent products (e.g., DHT) and the concentration and transcriptional activity of androgen receptors.[81]

The severity of virilization of the external genitalia in 46,XX females has typically been described using the five-point Prader scale (Fig. 101.6C); however, other validated tools exist. The degree of labioscrotal fusion can be described in newborn females via the measurement of anogenital distances (ratio of the measurement from the center of the anus to fourchette to the measurement from the center of the anus to the anterior base of the clitoris).[82] More recently, the External Genitalia Score (EGS) has been developed as a gradual scale ranging from female (EGS=0) to male (EGS=12), to be used in both sexes.[83] The EGS incorporates degree of labioscrotal fusion (Fig. 101.9), length of the genital tubercle, position of the urethral meatus, and location of the gonads.

Despite published evidence of masculine gender behaviors in females with CAH, the majority of female-raised individuals with 46,XX CAH report female core gender identity (with around 5% reporting gender dysphoria).[84] Optimal gender assignment in the rare patients with 46,XX CAH and complete virilization (Prader 5, with normally formed penis) is an area of controversy.[81]

21-Hydroxylase Deficiency. The most common cause of 46,XX DSD is a deficiency in 21-hydroxylase, an enzyme that is required for both cortisol and aldosterone production (see Fig. 101.7). As discussed in Chapter 94, classic forms of 21-hydroxylase deficiency (21-OHD) are associated with significantly reduced enzyme activity, leading to accumulation of the adrenal precursor 17-hydroxyprogesterone (17-OHP). Individuals with classic 21-OHD present in infancy with cortisol deficiency and, in salt-wasting forms, aldosterone deficiency. Affected females have genital ambiguity as a result of excess DHT produced from accumulated adrenal androgens (including the conversion of 17-OHP to DHT via the "backdoor" or "alternative" pathway) (see Fig. 101.10);

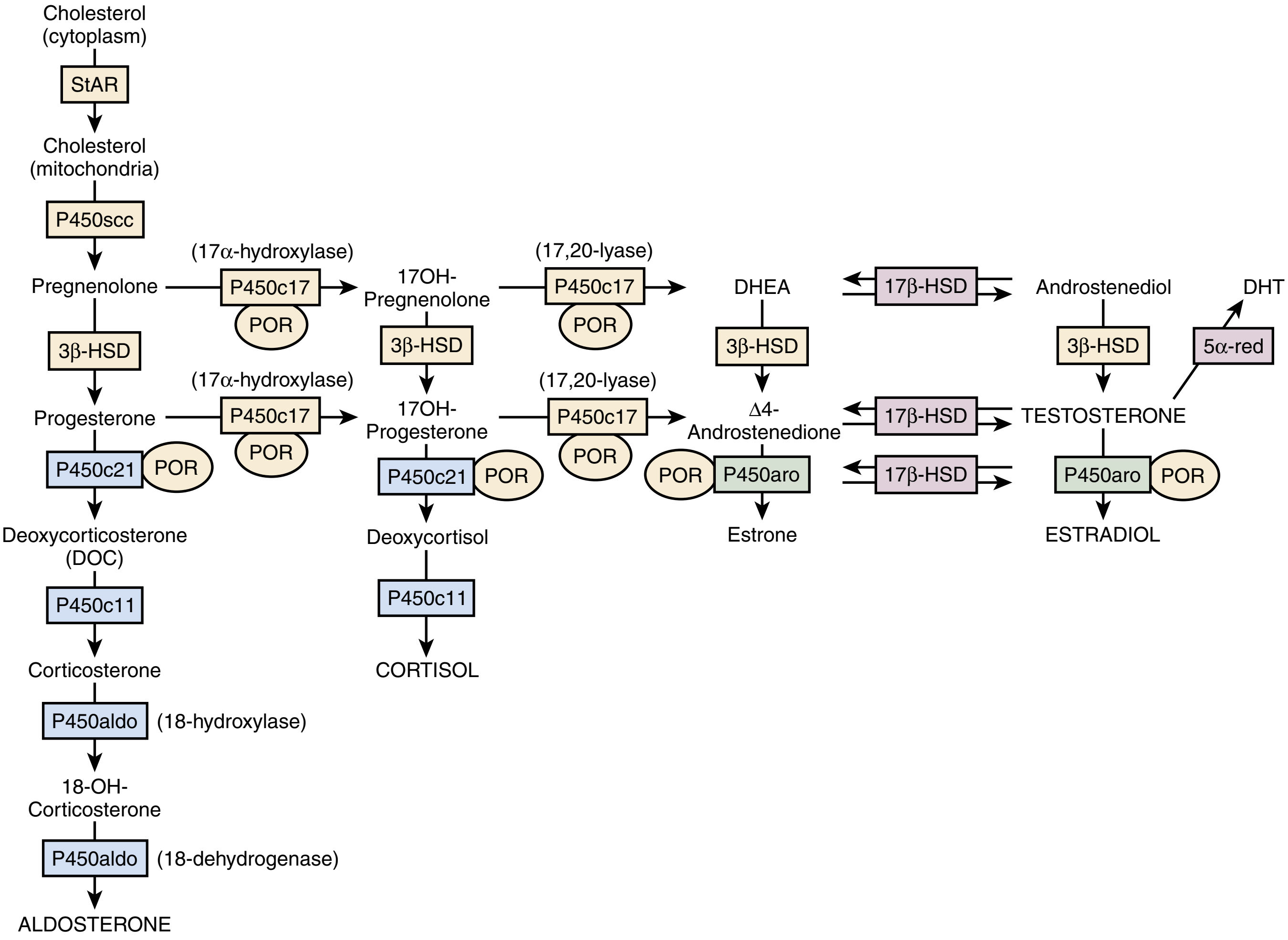

Cholesterol (cytoplasm)
StAR
Cholesterol (mitochondria)
P450scc
(17α-hydroxylase)
(17,20-lyase)
Pregnenolone
P450c17
POR
17OH-Pregnenolone
P450c17
POR
DHEA
17β-HSD
Androstenediol
DHT
3β-HSD
3β-HSD
3β-HSD
3β-HSD
5α-red
(17α-hydroxylase)
(17,20-lyase)
Progesterone
P450c17
POR
17OH-Progesterone
P450c17
POR
Δ4-Androstenedione
17β-HSD
TESTOSTERONE
P450c21
POR
P450c21
POR
POR
P450aro
17β-HSD
P450aro
POR
Deoxycorticosterone (DOC)
Deoxycortisol
Estrone
ESTRADIOL
P450c11
P450c11
Corticosterone
CORTISOL
P450aldo
(18-hydroxylase)
18-OH-Corticosterone
P450aldo
(18-dehydrogenase)
ALDOSTERONE

Figure 101.7 Adrenal and gonadal steroidogenesis. The first four steps take place in both adrenals and gonads *(yellow)*. (1) *StAR,* Steroidogenic acute regulatory protein; (2) *P450scc,* Cholesterol side-chain cleavage; (3) *3β-HSD,* 3β-Hydroxysteroid dehydrogenase type 2; (4) P450c17 for both 17α-hydroxylase and 17,20-lyase (or desmolase). Subsequent steps are organ-specific. In the adrenal cortex *(blue)* P450c21, 21-hydroxylase, and P450c11, 11β-hydroxylase, are active in both the zona glomerulosa and fasciculata. Then P450aldo (aldosterone synthase with its two activities, 18-hydroxylase and 18-dehydrogenase or oxidase) is specific of zona glomerulosa. In the gonads, 17β-HSD, for 17β-hydroxysteroid dehydrogenase type 3, synthesizes testosterone, and P450aro aromatizes androgens to estrogens. Mainly in peripheral androgen-target tissues, 5α-reductase converts testosterone to dihydrotestosterone (DHT). POR (P450 oxidoreductase) is a flavoprotein necessary for most P450 cytochromes.

however, affected males have normal genitalia and penile size. Prompt diagnosis and treatment are necessary to prevent adrenal crisis. Infants born with ambiguous or male-appearing genitalia and nonpalpable gonads should be immediately evaluated for 46,XX DSD due to classic 21-OHD by measurement of 17-OHP; markedly elevated levels greater than 10,000 ng/dL are suggestive of the diagnosis (Table 101.4). Because 17-OHP may not be unequivocally elevated the first day of life, hormonal investigations must be repeated daily unless an alternative diagnosis is found. Meanwhile, studies of salt balance (Na+, K+ in blood and urine) should also be performed, as salt loss usually appears evident from the end of the second week of life. Less severe defects in the 21-hydroxylase enzyme may result in nonclassic or "late-onset" 21-OHD; however, this disorder is generally milder, is associated with subclinical cortisol deficiencies and less severe androgen excess, and does not typically lead to genital ambiguity in female newborns. The 21-hydroxylase enzyme is encoded by the gene *CYP21A2* on chromosome 6 (6p21.3) in a region that includes a homologous pseudogene, *CYP21A1P*. Recombination events between *CYP21A12* and its pseudogene lead to mutant alleles with variable degree of enzymatic compromise (see Chapter 94).

11β-Hydroxylase Deficiency. 11β-hydroxylase deficiency is a less common form of virilizing CAH (see Chapter 94). The hormonal marker of the disorder is the selective elevation of 11-deoxycortisol due to impaired conversion to cortisol (see Table 101.4). Accumulation of another adrenal precursor, deoxycortisone (an active mineralocorticoid), leads to salt retention, hypokalemia, and hypertension. Due to absence of salt wasting, adrenal crisis is rare. Affected females with the severe or "classic" form of 11β-hydroxylase deficiency will exhibit masculinization of the external genitalia, while those with milder or "nonclassic" forms will have normal genitalia at birth but may develop symptoms of hyperandrogenism in childhood or adulthood.

Other Forms of Virilizing Congenital Adrenal Hyperplasia β. CAH due to 3β-hydroxysteroid dehydrogenase deficiency and P450 oxidoreductase deficiency cause virilization in affected females, but also undervirilization in affected males, and therefore are causes of both 46,XX and 46,XY DSD. These conditions are discussed more in the section entitled "*46,XY Disorders in Androgen Synthesis or Action*."

Maternal-Derived Androgens. During normal pregnancy, several measures exist to protect the fetus from maternal-derived androgens. An

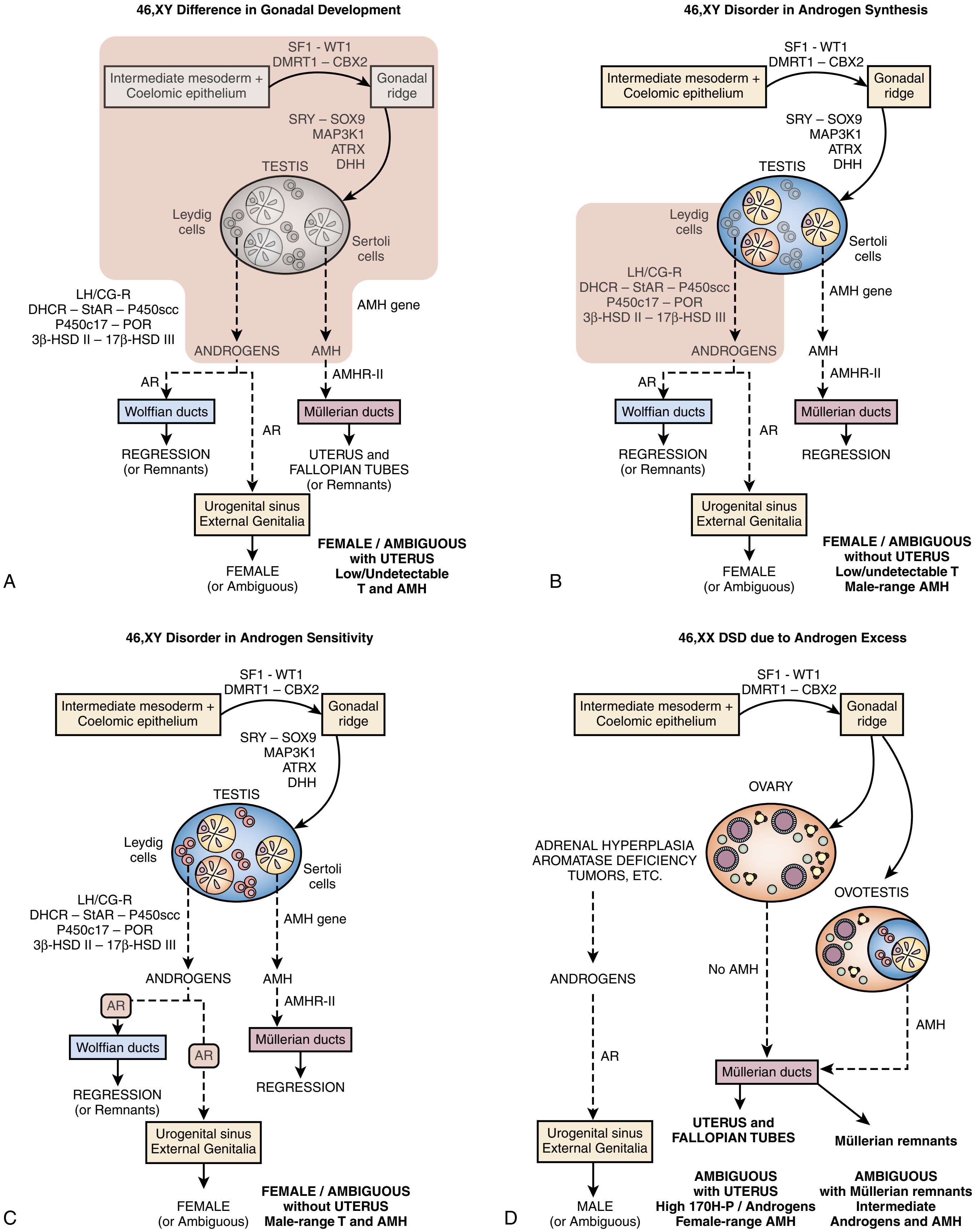

46,XY Difference in Gonadal Development
Intermediate mesoderm + Coelomic epithelium
SF1 - WT1
DMRT1 – CBX2
Gonadal ridge
SRY – SOX9
MAP3K1
ATRX
DHH
TESTIS
Leydig cells
Sertoli cells
LH/CG-R
DHCR – StAR – P450scc
P450c17 – POR
3β-HSD II – 17β-HSD III
AMH gene
ANDROGENS
AMH
AR
AMHR-II
Wolffian ducts
Müllerian ducts
AR
REGRESSION (or Remnants)
UTERUS and FALLOPIAN TUBES (or Remnants)
Urogenital sinus External Genitalia
FEMALE (or Ambiguous)
FEMALE / AMBIGUOUS with UTERUS Low/Undetectable T and AMH
A
46,XY Disorder in Androgen Synthesis
Intermediate mesoderm + Coelomic epithelium
SF1 - WT1
DMRT1 – CBX2
Gonadal ridge
SRY – SOX9
MAP3K1
ATRX
DHH
TESTIS
Leydig cells
Sertoli cells
LH/CG-R
DHCR – StAR – P450scc
P450c17 – POR
3β-HSD II – 17β-HSD III
AMH gene
ANDROGENS
AMH
AR
AMHR-II
Wolffian ducts
Müllerian ducts
AR
REGRESSION (or Remnants)
REGRESSION
Urogenital sinus External Genitalia
FEMALE (or Ambiguous)
FEMALE / AMBIGUOUS without UTERUS Low/undetectable T Male-range AMH
B
46,XY Disorder in Androgen Sensitivity
Intermediate mesoderm + Coelomic epithelium
SF1 - WT1
DMRT1 – CBX2
Gonadal ridge
SRY – SOX9
MAP3K1
ATRX
DHH
TESTIS
Leydig cells
Sertoli cells
LH/CG-R
DHCR – StAR – P450scc
P450c17 – POR
3β-HSD II – 17β-HSD III
AMH gene
ANDROGENS
AMH
AR
AMHR-II
Wolffian ducts
Müllerian ducts
AR
REGRESSION (or Remnants)
REGRESSION
Urogenital sinus External Genitalia
FEMALE (or Ambiguous)
FEMALE / AMBIGUOUS without UTERUS Male-range T and AMH
C
46,XX DSD due to Androgen Excess
Intermediate mesoderm + Coelomic epithelium
SF1 - WT1
DMRT1 – CBX2
Gonadal ridge
OVARY
ADRENAL HYPERPLASIA
AROMATASE DEFICIENCY
TUMORS, ETC.
OVOTESTIS
ANDROGENS
No AMH
AMH
AR
Müllerian ducts
Urogenital sinus External Genitalia
UTERUS and FALLOPIAN TUBES
Müllerian remnants
MALE (or Ambiguous)
AMBIGUOUS with UTERUS High 17OH-P / Androgens Female-range AMH
AMBIGUOUS with Müllerian remnants Intermediate Androgens and AMH
D

Figure 101.8 Pathogenesis of differences of sex development (DSDs). **A,** 46,XY difference in gonadal development. **B,** 46,XY disorder in androgen synthesis. **C,** 46,XY disorder in androgen sensitivity. **D,** 46,XX DSD due to androgen excess. (From Rey R, Josso N, Racine C. Sexual Differentiation. In: Feingold KR, Anawalt B, Boyce A, et al editors. Endotext [Endotext.org]. South Dartmouth (MA): MDText; 2000.com, Inc.; 2000–. PMID: 25905232.)

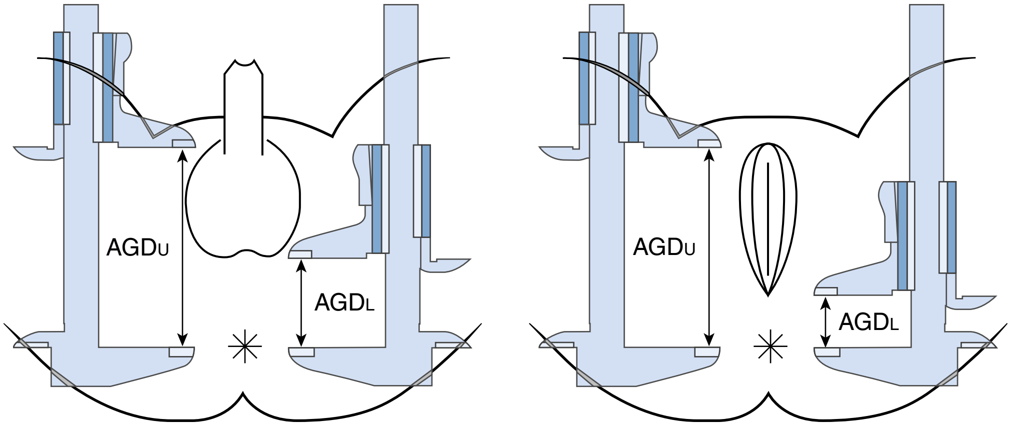


Figure 101.9 **Measurement of anogenital distances using a gender-neutral method suitable for all babies.** Lower/upper anogenital distance ratio (AGD_L/AGD_U) is calculated as a marker of genital virilization. AGD_L, Lower anogenital distance (measured from the center of the anus to the base of the labioscrotal border); AGD_U, upper anogenital distance (measured from the center of the anus to the anterior base of the genital tubercle). (Adapted from van der Straaten S, Springer A, Zecic A, et al. The External Genitalia Score (EGS): A European Multicenter Validation Study. *J Clin Endocrinol Metab.* 2020;105.)

increase in sex hormone-binding globulin (which binds testosterone and DHT) reduces the percentage of free androgen. Additionally, aromatase converts testosterone and androstenedione into estradiol and estrone, respectively. However, if excess maternal androgens are produced in high enough concentrations, placental aromatase activity may be overwhelmed, leading to masculinization of a 46,XX fetus. Fetal virilization has been reported with maternal luteoma of pregnancy, adrenal adenoma, hyperreactio luteinalis, metastatic carcinoma (Krukenberg tumor), CAH, and adrenal corticocarcinoma.[85,86] Maternal familial glucocorticoid resistance (characterized by manifestations of androgen and mineralocorticoid excess, with increased cortisol secretion without clinical signs of hypercortisolism) has also been associated with fetal virilization.[87] However, the overall occurrence of fetal virilization is rare, even in cases of significant maternal virilization. For example, among androgen-secreting luteomas of pregnancy, less than half of affected mothers develop symptomatic hyperandrogenism, and only 50% of female infants born to symptomatic mothers experience virilization.[86] Variability in fetal virilization may relate to proportion and type of androgens secreted, as some steroids are not substrates for aromatase and therefore cannot be converted to estrogens by placental aromatase. Gestational hyperandrogenism due to PCOS has not been shown to cause virilization in female infants, likely due to the protective mechanisms described above.[88]

Apart from endogenous maternal androgens, fetal virilization has been reported following maternal treatment with progestational agents (given to prolong a pregnancy threatened by spontaneous abortion), testosterone, and danazol.[89,90] Effects depend on the medication, dose, and duration. Iatrogenic fetal virilization has become very rare, because hormonal treatment is no longer advocated during pregnancy.

Aromatase Deficiency. The aromatase enzyme, a member of the cytochrome p450 superfamily, is encoded by the *CYP19* gene (located at 15q21) and catalyzes the synthesis of estrogens from androgens. Aromatase deficiency is an autosomal recessive condition due to a genetic defect in the fetal component of the placenta. During pregnancy, placental aromatase protects the fetus from maternally-derived androgens and protects the fetus and mother from high concentrations of androgens synthesized normally by the fetal adrenal. Therefore, deficiency in aromatase (OMIM 613546) results in virilization of a 46,XX fetus, as well as maternal virilization during pregnancy that resolves following delivery. As aromatase deficiency is also a feature of p450 oxidoreductase deficiency (OMIM 613571), a rare form of virilizing CAH (discussed in the section entitled "*46,XY Disorders of Androgen Synthesis or Action*"), it is critical to rule out cortisol deficiency in the fetus if the diagnosis is suspected. Beyond the newborn period, features of aromatase deficiency include pubertal failure, mild virilization, multicystic ovaries (due to hyperstimulation by follicle-stimulating hormone [FSH] and luteinizing hormone [LH] resulting from inability of the ovary to aromatase testosterone and androstenedione to estrogens), and delayed bone age (supporting the notion that estrogens drive fusion of epiphyses during puberty).[91] Biochemical findings include elevated androgens, low estrone and estradiol levels, and elevated gonadotropins (due to decrease in negative feedback from reduced estrogens).[92] Aromatase gene mutations can also produce "nonclassic" phenotypes: low residual aromatase activity may be sufficient for breast and uterine development to occur at puberty, despite significant androgenization *in utero.*[93]

Other Causes of 46,XX Differences of Sex Development

Disorders of Müllerian Development. In the absence of AMH, the Müllerian ducts develop into the fallopian tubes, uterus, and upper part of the vagina in 46,XX individuals. Although abnormal expression of AMH in a 46,XX individual may theoretically result in Müllerian duct anomalies, this condition has not been reported.[94] Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome is a heterogeneous disorder referring to the congenital aplasia or hypoplasia of Müllerian duct derivatives, occurring in one out of 5000 newborn girls.[95] There are several subtypes of MRKH, including those in which the upper vagina, cervix, and uterus are solely affected, and other subtypes in which additional associated malformations occur, including renal, skeletal (including cervicothoracic somite dysplasia or MURCs association, OMIM 601076), cardiac, and hearing anomalies. Several candidate genes have been implicated in MRKH, including mutations in genes encoding transcription factors such as HNF1B (OMIM 137920, associated with renal malformations and maturity onset diabetes of the young type 5) and TBX-6.[96] A number of genes in the *WNT* family have also been associated with MRKH in human and animal models, most notably mutations in *WNT4,* which have occurred in individuals with MRKH and hyperandrogenism.[96] Müllerian anomalies have also been reported in other malformative syndromes (see Table 101.2). Müllerian and renal agenesis have been reported in velocardiofacial syndrome resulting from a deletion of chromosome 22q11.2 (OMIM 192430). Malformations of the vagina, uterus, and bladder occur in females with hand-foot-genital syndrome resulting from *HOXA13*

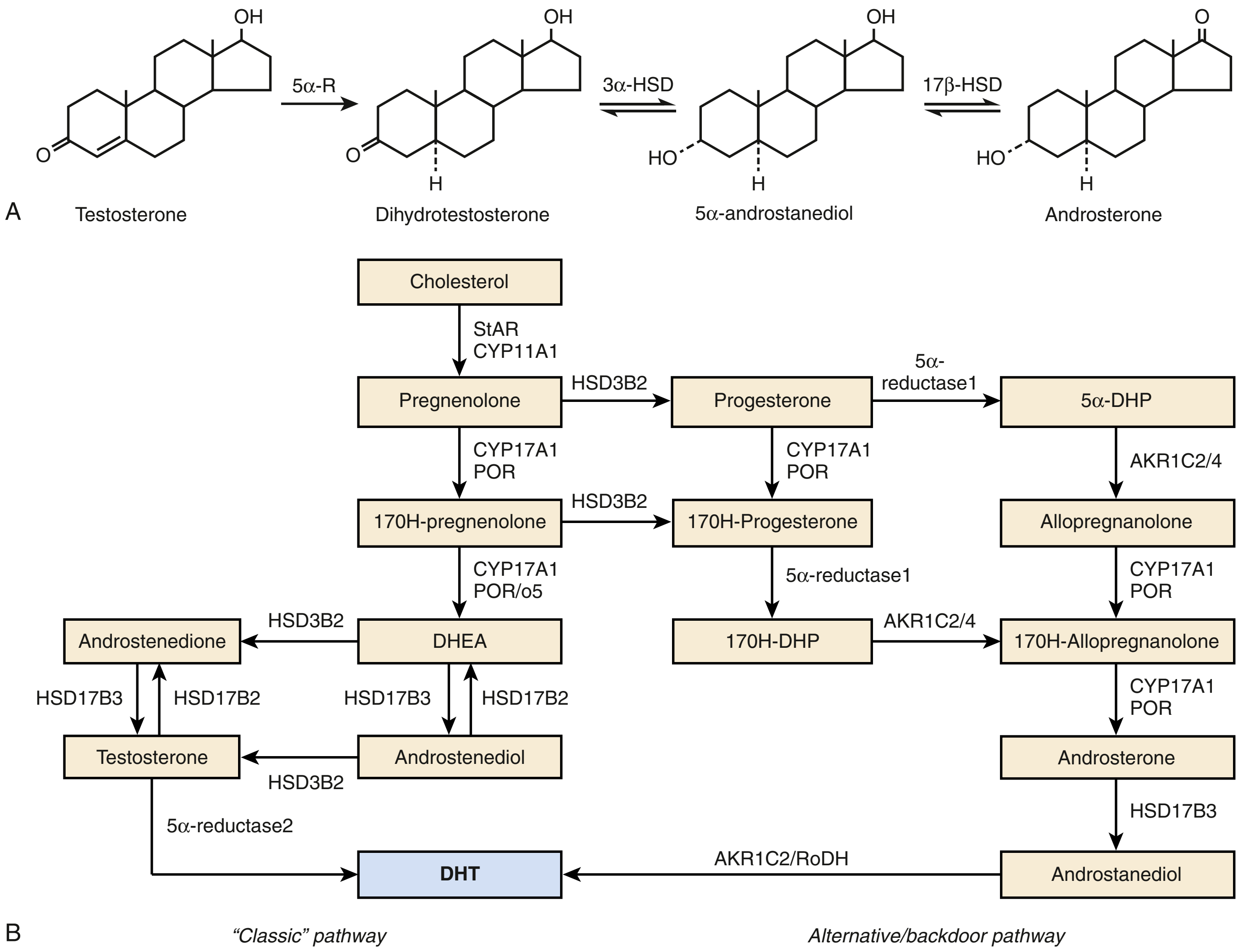

OH
OH
OH
O
5α-R
3α-HSD
17β-HSD
O
O
HO
HO
H
H
H
A
Testosterone
Dihydrotestosterone
5α-androstanediol
Androsterone
Cholesterol
StAR
CYP11A1
Pregnenolone
HSD3B2
Progesterone
5α-reductase1
5α-DHP
CYP17A1
POR
CYP17A1
POR
AKR1C2/4
17OH-pregnenolone
HSD3B2
17OH-Progesterone
Allopregnanolone
CYP17A1
POR/o5
5α-reductase1
CYP17A1
POR
Androstenedione
HSD3B2
DHEA
17OH-DHP
AKR1C2/4
17OH-Allopregnanolone
HSD17B3
HSD17B2
HSD17B3
HSD17B2
CYP17A1
POR
Testosterone
Androstenediol
Androsterone
HSD3B2
HSD17B3
5α-reductase2
DHT
AKR1C2/RoDH
Androstanediol
B
"Classic" pathway
Alternative/backdoor pathway

Figure 101.10 **Synthesis of dihydrotestosterone (DHT) via the classic and alternative (backdoor) pathways. A,** Synthesis of DHT from testosterone (classic) and androsterone (alternative). (From Wilson JD, Shaw G, Leihy ML, et al. The marsupial model for male phenotypic development. *Trends Endocrinol Metab*. 2002;13:78–83, with permission from Elsevier.) **B,** The classic pathway of steroidogenesis leading to DHT *(left)* and the alternative pathway *(right)*. The factors in the classic pathway are CYP11A1 (cholesterol side-chain cleavage enzyme, P450scc), StAR (steroidogenic acute regulatory protein), CYP17A1 (17α-hydroxylase/17,20-lyase, P450c17), HSD3B2 (3β-hydroxysteroid dehydrogenase, type 2), and HSD17B3 (17β-HSD3 [17β-hydroxysteroid dehydrogenase, type 3] and 5α-reductase, type 2 [5α-reductase 2, encoded by SRD5A2]). The alternative pathway is characterized by the presence of additional enzymes: 5α-reductase, type 1 (5α-reductase1, encoded by SRD5A1), AKR1C23 (3α-reductase, type 3) and possibly AKR1C4 (3α-reductase, type 1) and RoDH (3-hydroxyepimerase, encoded by HSD17B6). Most steroids are identified by their trivial names: 17-hydroxy-dihydroprogesterone (17OH-DHP) is 5α-pregnane-17α-ol-3,20 dione; 17-hydroxy-allo-pregnanolone (17OH-allo) is 5α-pregnan-3α,17α-diol-20-one; 5α-dihydroprogesterone (5α-DHP) is 5α-pregnane-3,20-dione, and allo-pregnanolone is 3α-hydroxy-dihydroprogesterone (3α-OH-DHP) or 5α-pregnane-3α-ol-20-one. (Reprinted from Flück CE, Meyer-Böni M, Pandey AV, et al. Why boys will be boys: Two pathways of fetal testicular androgen biosynthesis are needed for male sexual differentiation. *Am J Hum Genet*. 2011;89:201–218, with permission from Elsevier.)

mutations (OMIM 140000). Vaginal atresia has also been described in McKusick–Kaufman syndrome (OMIM 236700), probably caused by mutations in the *MKKS* gene. Another rare malformation syndrome closely related to MRKH is Herlyn–Werner–Wunderlich syndrome, characterized by uterus didelphys, blind hemivagina ,and ipsilateral renal agenesis.[97] Müllerian anomalies also occur in females with anorectal malformations and are commonly seen in those with cloacal exstrophy.[98]

Syndromic Disorders. Syndromic disorders leading to 46,XX DSD are discussed in the section entitled "*Syndromic DSD*."

KEY POINTS

- 46,XX differences in sex development (DSD) occur due to variations in gonadal development, fetal exposure to androgen excess, or abnormalities in Müllerian development, or in association with genetic syndromes. Excessive levels of androgens result in virilization of the external genitalia in a 46,XX individual. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency is the most common cause of 46,XX DSD, and should be suspected in any infant with ambiguous or male-appearing genitalia and nonpalpable gonads.

TABLE 101.4 Main Characteristics of Inborn Errors of Steroid Biosynthesis Leading to Differences in Sex Development (DSD)

Defect	DSD	Cortisol	Androgens	Mineralocorticoid Axis	Biochemical Markers
StAR or P450scc (*CYP11A1*)	XY	↓	↓	Salt loss, ↑↑ PRA	↓↓ All steroids,
Combined 17α-hydroxylase/17,20 lyase (P450c17, *CYP17A1*)	XY	↓*	↓	Hypertension, ↓ PRA	↑ Pregnenolone ↑ Progesterone ↑ DOC ↑ corticosterone
Isolated 17,20-lyase (P450c17, *CYP17A1*)	XY	Normal	↓	Normal	↑ 17OHPreg ↑ 17-OHP ↓ DHEA
17 β-HSD type 3 (HSD17B3)	XY	Normal	↓	Normal	Testosterone/androstenedione <0.8
3β-HSD type 2 (HSD3B2)	XX, XY	↓	↑ DHEA ↓ others	Salt loss, ↑ PRA	↑ DHEA ↑ Pregnenolone ↑ 17OHPreg (>6600 ng/dL or 200 nmol/L) ↓ 17-OHP
POR	XX, XY transient maternal virilization	Variable	Variable	Variable/hypertension, ↑/normal PRA	↑ Progesterone ↑ Pregnenolone ↑ 17-OHP ↑ DOC
21-OH (*CYP21A1*)	XX	↓	↑	Variable salt loss, ↑/normal PRA	↑ 17-OHP (generally >10,000 ng/dL or 300 nmol/L)
11β-OH (*CYP11B1*)	XX	↓	↑	Hypertension, ↓ PRA	↑ 11-deoxycortisol (>3,500 ng/dL) ↑ DOC ↑ 17-OHP
Aromatase (P450aro, *CYP19A1*)	XX transient maternal virilization	Normal	↑ testosterone	Normal	↓ Estradiol ↑ Testosterone ↑ Androstenedione

PRA, Plasma renin activity; *DOC*, deoxycorticosterone; *DHEA*, dehydroepiandrosterone; *17OHP*, 17-hydroxypregnenolone; *17OHPreg, 17-hydroxypregnenolone; StAR,* steroidogenic acute regulatory.

46,XY Differences of Sex Development

Disruption in hormone signaling during the critical period of fetal development may result in 46,XY DSD. Insufficient androgen exposure (either due to impaired androgen synthesis or action) leads to undervirilized external genitalia and Wolffian duct regression; uterus may be present or absent depending on whether AMH production is impacted as well (see Fig. 101.8).

As these conditions are present early in fetal development, the androgen-dependent fusion of the urethral and labioscrotal folds (discussed earlier in the section entitled "*Male Sex Differentiation*") and elongation of the phallus is disrupted, leading to genital ambiguity (e.g., hypospadias and micropenis) or, in extreme cases, near-female–appearing external genitalia. The Quigley scale has been used to describe the appearance of external genitalia in 46,XY DSD (Fig. 101.11).

Isolated defects in AMH production or sensitivity are characterized by the existence of uterus and fallopian tubes in an otherwise normally virilized patient. 46,XY DSD may also result from nonhormonal causes, including intrauterine growth restriction and various genetic syndromes.

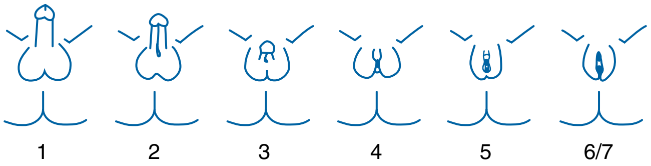


Figure 101.11 Schematic representation of Quigley scale, now widely used for classification of defective genital masculinization in 46,XY difference of sex development. Grades are numbered 1–7 in order of increasing severity (more defective masculinization). Grade 1: normal masculinization in utero; grade 2: male phenotype with mild defect in masculinization (e.g., isolated hypospadias); grade 3: male phenotype with severe defect in masculinization—small penis, perineoscrotal hypospadias, bifid scrotum, and/or cryptorchidism; grade 4: severe genital ambiguity—clitoral-like phallus, labioscrotal folds, single perineal orifice; grade 5: female phenotype with posterior labial fusion and clitoromegaly; grade 6/7: female phenotype (grade 6 if pubic hair present in adulthood, grade 7 if no pubic hair in adulthood). (Adapted from Quigley CA, Bellis ADE, Marschke KB, et al. Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocr Rev.* 1995;16:271–321.)

46,XY Differences in Gonadal Development

Pure Gonadal Dysgenesis. Pure, sometimes called complete, gonadal dysgenesis results in streak gonads that do not secrete sex steroids or AMH. Consequently, these patients have female internal and external genitalia. In 46,XY individuals, the condition is also known as Swyer syndrome or as complete sex reversal (a term that is no longer used). Absence of mammary development and primary amenorrhea are the most frequent presenting feature in a child reared unequivocally as a female. Adults with 46,XY gonadal dysgenesis are taller than those with 46,XX gonadal dysgenesis, which suggests the existence of a Y-specific growth gene that promotes growth independently of gonadal steroids.[99] Serum AMH is undetectable,

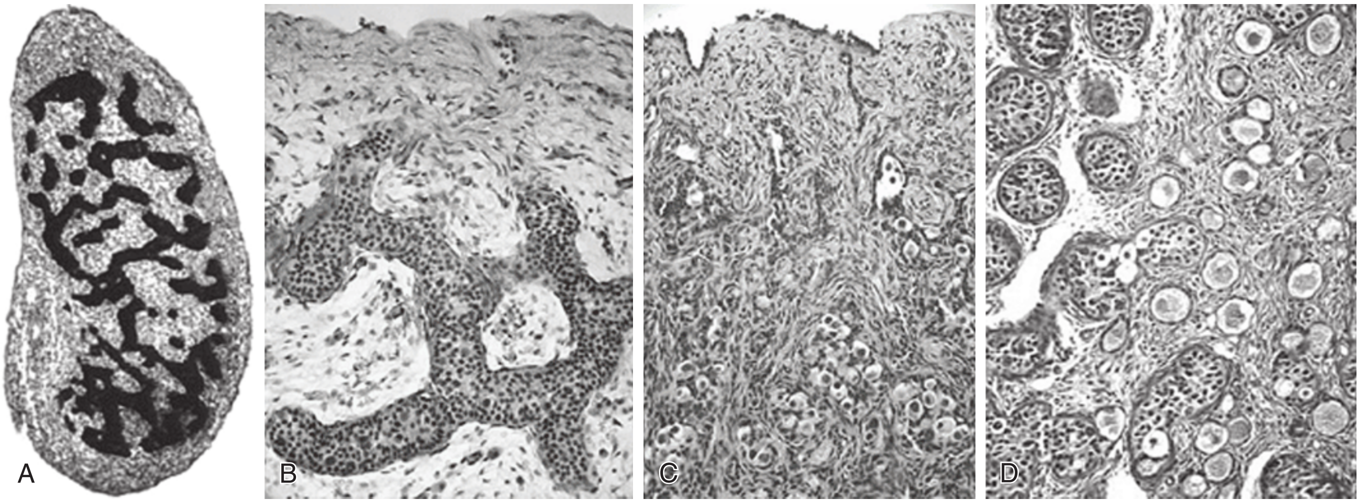


Figure 101.12 **A,** Normal testis of an 11-week-old fetus: anastomosing seminiferous cords, positively stained for anti-Müllerian hormone (AMH), form a typical network. **B,** Dysgenetic testis with abnormal seminiferous cords forming an anastomosing network. **C,** Streak gonad with abnormal germ cells. **D,** Ovotestis, showing seminiferous cords and follicles with oocytes. (Modified with permission from Chemes H, Muzulin PM, Venara MC, et al. Early manifestations of testicular dysgenesis in children: Pathological phenotypes, karyotype correlations and precursor stages of tumour development. *APMIS*. 2003;111:12–24. Copyright APMIS 2003.)

androgens (of adrenal origin) are low and do not respond to human chorionic gonadotropin (hCG) stimulation, and gonadotropin levels are elevated.[100] 46,XY patients are at increased risk for the development of gonadal tumors, gonadoblastoma, and dysgerminoma.[101]

Partial Gonadal Dysgenesis. The phenotypic differences resulting from partial gonadal dysgenesis (previously called dysgenetic male pseudohermaphroditism), which are observed in 46,XY DSD patients, depend on the extent of testicular differentiation. A mixture of Wolffian and Müllerian duct derivatives can be observed. A continuum of phenotypes ranging from almost-female genitalia with clitoromegaly to male genitalia with mild hypospadias and cryptorchidism has been described. In the most severe cases, gonadal tissue may contain areas of poorly differentiated seminiferous tubules in combination with areas of ovary-like stroma. These "streak-testes" should be distinguished from ovotestes, in which ovarian follicles are present.[102] All dysgenetic gonads in Y-chromosome-bearing patients have an increased risk for gonadal tumors.[29,101] Spontaneous breast development may provide the first evidence of an estrogen-secreting gonadal tumor. In milder forms, testes have distinct seminiferous tubules; however, the gonads are small, the albuginea is thin and loosely organized, and the seminiferous tubules are separated by wide intertubular spaces, frequently forming a network that resembles the first stages of sex cord formation in the developing testis and contains scarce germ cells (Fig. 101.12).[102] Ring tubules, calcified psamomma bodies, or a persistent coelomic epithelium can also be observed. A frequent surgical finding is the dissociation between testis and epididymis, joined only by a thin connective membrane. Serum testosterone, AMH, and inhibin B are intermediate between the male and female ranges, and gonadotropins are elevated, in correlation with the degree of dysgenesis.[100,103]

Mild Gonadal Dysgenesis. In the mildest forms, testes are almost normal, except for a reduced number of germ cells. Patients have normal male genitalia and hormonal levels within the expected range for sex and age. In some cases, gonadal dysgenesis is suspected by testicular-epididymal dissociation observed during surgery for cryptorchidism and is confirmed by histologic examination. In other cases, the first manifestation is observed in adults who present with infertility or testicular tumor.[104]

Asymmetric (Mixed) Gonadal Dysgenesis. As suggested by Jost's experiments,[1] the action of fetal testicular hormones is mainly local. This can be observed clearly in asymmetric gonadal differentiation, also called mixed gonadal dysgenesis, characterized by a dysgenetic testis on one side and a streak contralaterally. Patients usually, but not exclusively, present with male ducts and a palpable gonad in a hemiscrotum or inguinal canal on one side and a hemiuterus and an empty labioscrotal fold on the other side. The size of the phallus and the position of the urethral meatus are variable. The most commonly found karyotype in peripheral blood is 45,X/46,XY, but the karyotype may range from normal 46,XY to various mosaicisms. When studied in gonadal tissue, a 46,XY karyotype is more prevalent in the testis, and abnormal karyotypes are found in the streak.[102]

Testicular Regression Syndrome. Testicular regression syndrome (also known as congenital anorchia or vanishing testes syndrome) is a condition in which testes are thought to have been present early in fetal development but subsequently atrophied. The underlying etiology is unclear; bilateral testicular torsion has been theorized as a possible cause, though there may also be a genetic basis, including pathogenic variants of *DHX37*.[105] External genitalia appear typically male, though micropenis may be present due to interruption in androgen-induced phallic growth later in gestation. Uterus is typically absent, as AMH secretion early in fetal development is normal.

46,XY Ovotesticular Differences of Sex Development. Ovotesticular DSD has been described earlier in the section on 46,XX DSD; as noted, a minority of individuals with OT-DSD have a 46,XY karyotype.

46,XY Disorders in Androgen Synthesis or Action. Differences in androgen synthesis or action in a 46,XY individual cause undervirilized (ambiguous or female) external genitalia. As AMH production is normal, regression of Müllerian structures occurs normally, and these individuals have no uterus (see Fig. 101.8). Internal male genitalia may be absent, as testosterone is required for development of Wolffian ducts. Testes are present but may be intraabdominal or incompletely descended.

DSDs in this category include Leydig cell aplasia or hypoplasia, defects in cholesterol synthesis (Smith–Lemli–Opitz syndrome [SLOS]), enzymatic defects in testosterone biosynthesis (including rare forms of CAH impacting both adrenal and testicular steroidogenesis, as well as isolated defects in testicular steroidogenesis such as isolated 17,20-lyase deficiency and 17β-hydroxysteroid dehydrogenase III deficiency), defects in testosterone metabolism (5α-reductase type 2 deficiency), and defects in androgen action (including complete and partial androgen insensitivity syndrome).

Leydig Cell Aplasia or Hypoplasia. Leydig cell aplasia (OMIM 238320) is a rare autosomal recessive form of 46,XY DSD resulting from inactivating mutations of the LHCGR (located on chromosome 2p16.3),which is essential to Leydig cell differentiation and function before and after birth. In 46,XY individuals, impairment of prenatal and postnatal testosterone secretion results in a phenotypic spectrum ranging from predominantly female external genitalia without pubertal virilization in severe forms to micropenis and infertility. Owing to normal secretion of AMH by Sertoli cells, Müllerian derivatives are absent. Basal and hCG-stimulated testosterone levels are low, plasma LH is elevated, and FSH is normal. Testes are inguinal or intraabdominal and contain normal Sertoli cells but few or no mature Leydig cells.[106] In genetic females, the phenotype is usually limited to oligomenorrhea and infertility due to anovulation.

Smith–Lemli–Opitz Syndrome (Cholesterol Synthesis Defect). SLOS (OMIM 270400) is an autosomal recessive disorder caused by a defect in dehydrocholesterol reductase blocks the last step of cholesterol synthesis. SLOS has a wide phenotypic spectrum; however, features may include microcephaly, characteristic facial findings (e.g., low-set ears, short and broad nose, small chin), cleft palate, syndactyly of the second and third toes, cardiac and renal defects, growth retardation, developmental delays, and aggressive or self-injurious behavior. Intrauterine growth retardation can be noted prenatally. Hypospadias or bilateral cryptorchidism occurs in 50% of 46,XY patients, and sexual ambiguity with a rudimentary uterus is observed in severe cases. 46,XX patients usually have normal genitalia. The diagnostic biochemical finding is an elevated 7-dehydrocholesterol, the immediate precursor to cholesterol. SLOS results from mutations in the *DHCR7* gene (located on chromosome 11q13.4). During fetal development, cholesterol is required for steroidogenesis and for processing sonic hedgehog, which plays a major role in pattern formation of the nervous system, face, and limbs.[107] Although there have been several case reports of adrenal insufficiency in neonates with SLOS, abnormal adrenal function testing is not commonly seen in individuals with mild to moderate SLOS.[108]

Enzymatic Defects in Testosterone Biosynthesis. Five enzymatic activities are necessary for the synthesis of testosterone from cholesterol (see Fig. 101.7). Three are common to both the cortisol and testosterone biosynthetic pathways: cholesterol side-chain cleavage (P450scc), 3β-hydroxysteroid dehydrogenase (3β-HSD), and 17α-hydroxylase (P450c17). A deficiency in any of these activities results in both CAH and DSD in 46,XY patients. Two—17,20-lyase (activity contained in P450c17) and 17β-hydroxysteroid dehydrogenase (17β-HSD)—are required only for biosynthesis of sex steroids. P450 oxidoreductase (POR) supplies electrons to the steroidogenic enzymes P450c17, 21-hydroxylase, and aromatase; therefore, deficiency in POR results in CAH as well. The diagnosis of enzymatic defects in testosterone biosynthesis is typically based on biochemical markers and clinical features (see Table 101.4).

Steroidogenic defects leading to CAH (steroidogenic acute regulatory [StAR], p450scc, 3β-HSD, and combined 17α-hydroxylase/17,20-lyase deficiencies) are discussed in detail in Chapter 94 and are reviewed briefly here for completeness.

Lipoid Congenital Adrenal Hyperplasia and Cholesterol Side-Chain Cleavage Enzyme Deficiency. Lipoid CAH (LCAH, OMIM 201710) is a rare and severe form of CAH caused by defects in the StAR protein, which is involved in the very first step of steroidogenesis (see Fig. 101.7). The StAR protein accelerates the transport of cholesterol transport from outer to the inner mitochondrial membrane, following which the cholesterol side chain cleavage enzyme (cholesterol desmolase, CYP11A1, P450scc) removes the side chain of cholesterol to form pregnenolone. Deficiency in StAR protein impairs production of all adrenal and gonadal steroids directly by impacting the rate-limiting step of steroidogenesis, but also indirectly, as the accumulation of unutilized cholesterol esters induces apoptosis of steroidogenic cells. Steroidogenesis in the placenta is not dependent on StAR protein, and therefore adequate progesterone and estrogen is still produced to sustain a pregnancy in a fetus with LCAH. Mutations in cholesterol side chain cleavage enzyme (P450scc deficiency, OMIM 613743) are less common than StAR mutations but produce a similar phenotype to LCAH (except for the accumulation of intracellular cholesterol, which does not occur). In an affected 46,XY fetus, hCG-stimulated testosterone production does not occur, which prevents virilization of the external genitalia. Affected 46,XY newborns with the classic (most severe) form of LCAH or deficiency in cholesterol side chain cleavage enzyme may have phenotypically female external genitalia. 46,XX patients develop normally until puberty but subsequently develop polycystic ovaries without hyperandrogenism. Infants with classic forms may present with glucocorticoid and mineralocorticoid deficiencies leading to adrenal crisis and salt wasting. All steroid hormones in blood or urine are barely measurable at birth and do not respond to trophic hormone stimulation. Levels of plasma ACTH, plasma renin activity, and gonadotropins (at the time of hypothalamic-pituitary-gonadal [HPG] axis activation) are high. Milder, nonclassic forms with late onset, limited to adrenal insufficiency, have also been described.[109]

3β-Hydroxysteroid Dehydrogenase Type II Deficiency. 3β-hydroxysteroid dehydrogenase type II deficiency (3β-HSD2D) is a rare form of CAH that causes both 46,XX and 46,XY DSD. The 3β-HSD enzyme is responsible for the conversion of 3β-5 steroids to the 3-keto-4 configuration during the synthesis of glucocorticoids, mineralocorticoids, and sex steroids (see Fig. 101.7). 3β-HSD exits in two isoforms; however, only deficiencies in the type 2 isoform (3β-HSD2) expressed in adrenals and gonads leads to CAH. 3β-HSD2 expression is detectable in human testicular interstitial cells from 8 weeks' gestation; however, expression in fetal ovaries is not observed until 28 weeks; gestation. XY individuals are undervirilized with varying degrees of hypospadias and micropenis, and testes can usually be palpated in the groin. Although adrenal sex steroid production is impacted by the deficiency in 3β-HSD type 2 enzyme, conversion of dehydroepiandrosterone (DHEA) to more potent androgens by the unaffected 3β-HSD type 1 isoform (expressed in the placenta and peripheral tissues) leads to mild virilization of 46,XX infants. The affected female fetus may also be protected from severe masculinization due to physiologic variations in fetal cortisol synthesis and 3β-HSD2 expression, which peak early on (by 8–9 weeks' gestation), decline in midgestation, and rise again only in the third trimester.[110] Individuals with classic 3β-HSD2 have adrenal insufficiency with or without salt wasting. An elevated basal or ACTH-stimulated ratio of $\Delta 5$ versus $\Delta 4$ steroids or 17-hydroxypregnonolone >6600 ng/dL (200 nmol/L) may suggest the diagnosis (see Table 101.4).

P450 Oxidoreductase Deficiency. POR is a flavoprotein located in the endoplasmic reticulum of steroidogenic cells that transports electrons from nicotinamide adenine dinucleotide phosphate (NADPH) to all microsomal (type II) cytochrome P450 enzymes.

Three steroidogenic enzymes are affected by POR deficiency: 21-hydroxylase, P450c17 (preferentially 17,20-lyase activity with 17α-hydroxylase activity less affected), and aromatase (see Fig. 101.7). Deficiencies in POR may lead to DSD in both XX and XY individuals. In 46,XY POR deficiency, impaired 17,20-lyase activity decreases the synthesis of testosterone by the fetal testis in the second trimester of pregnancy. In 46,XX individuals, placental aromatase deficiency results in overproduction of fetoplacental androgens (see the section entitled "Aromatase Deficiency"). In addition, as a result of 21-hydroxylase deficiency and preferential inhibition of 17,20-lyase activity, there is overproduction of 17-hydroxyprogesterone and 17-hydroxypregnenolone, which subsequently leads to virilization via the "backdoor" pathway of DHT synthesis (see the section entitled "*Virilizing Congenital Adrenal Hyperplasias*" and Fig. 101.10). Due to decreased activity of placental aromatase, transient maternal virilization may also occur during pregnancy with an affected fetus. Affected individuals can experience glucocorticoid deficiency or hypertension (likely due to elevated levels of deoxycorticosterone resulting from 17α-hydroxylase deficiency). Patients with POR deficiency may also present with bone malformations (craniosynostosis, fusion of long bones, midface hypoplasia, choanal stenosis) that are indistinguishable from Antley–Bixler syndrome (ABS, OMIM 201750), possibly due to disruption of enzymes involved in cholesterol and/or retinoic acid metabolism.[111] More than 50 mutations in the *POR* gene (located at 7q11.23) have been identified, with high variability of clinical presentation, though homozygous null mutations tend to be lethal.[111] Milder forms of POR may not be associated with ABS, or even necessarily with genital ambiguity. Biochemical findings include elevated pregnenolone, progesterone, and 17-hydroxyprogesterone and decreased cortisol.

17α-Hydroxylase/17,20 Lyase Deficiency. A single microsomal enzyme, cytochrome P450c17 (encoded by *CYP17A1*, located at 10q24.3) catalyzes two consecutive oxidation reactions, the 17α-hydroxylase and 17,20-lyase reactions required for the adrenal glands and gonads to synthesize glucocorticoids and sex steroids (see Fig. 101.7). The 17α-hydroxylase reaction converts pregnenolone to 17-hydroxypregnenolone and progesterone to 17-hydroxyprogesterone, respectively, with the former being rate-limiting step in androgen biosynthesis. The 17,20-lyase reaction converts 17-hydroxypregnenolone to DHEA and catalyzes the equivalent conversion of 17-hydroxyprogesterone to androstenedione, although this latter reaction occurs much more slowly. Like all microsomal P450 enzymes, P450c17 receives electrons from NADPH via POR. Cytochrome b5, acting as an allosteric factor, promotes the interaction of the two molecules, further increasing 17,20-lyase activity. Deficiency in P450c17 (causing combined 17α-hydroxylase/17,20-lyase deficiency) is a rare cause of CAH, though frequency is higher in specific locales (for example, in Brazil, where it is the second most common cause of CAH after 21-OHD). Combined 17α-hydroxylase/17,20-lyase deficiency is often characterized in both sexes by hypertension and hypokalemia, which result from massive overproduction of the 17-deoxysteroids deoxycorticosterone and corticosterone; however, some patients may be normotensive.[112] Because of high production of corticosterone, which is a weak glucocorticoid, adrenal crisis is rarely observed. Lack of production of both gonadal and adrenal sex steroids results in severe undervirilization of 46,XY patients. Indeed, most XY patients have been raised as females because of a completely female external phenotype at birth, and they keep a female social sex at puberty. 46,XX patients have a normal phenotype at birth but fail to undergo puberty, and they may develop ovarian cysts because of excess gonadotropin stimulation.[113] Biochemical findings in combined 17α-hydroxylase/17,20-lyase deficiency include elevation in steroids proximal to the block (see Table 101.4), low levels of cortisol, androgen and estrogen precursors, elevated deoxycorticosterone and corticosterone, suppressed renin and aldosterone, and elevated FSH and LH concentrations.

Isolated 17,20 Lyase Deficiency. Individuals with 46,XY isolated 17,20-lyase deficiency (ILD) experience variable undervirilization due to exclusive loss of sex steroid synthesis. Due to retention of 17α-hydroxylase activity, other consequences of adrenal CYP17A1 deficiency (discussed earlier) are not typically seen. True isolated 17,20-lyase deficiency is extremely rare. A few cases of confirmed ILD involved mutations in the redox-partner binding surface of the CYP17A1 enzyme.[114] However, others with apparent "ILD" have been later found to have mutations in other genes including those encoding cytochrome b5 (CYB5A) and POR, and AKR1C2/AKR1C4 (3α-hydroxysteroid dehydrogenase enzymes involved in the "backdoor" pathway for DHT production, see Fig. 101.10). [115]

17β-Hydroxysteroid Dehydrogenase III Deficiency. The 17β-hydroxysteroid dehydrogenase III (17β-HSD) enzyme catalyzes the last step in testosterone biosynthesis (i.e., the reduction of androstenedione to testosterone, see Fig. 101.7). 17β-HSD type 3 deficiency (OMIM 264300) involves the testicular isoform of 17β-HSD, resulting in a DSD limited to 46,XY individuals. The condition may often be mistaken for androgen insensitivity at birth and 5α-reductase deficiency at puberty. Patients typically have intraabdominal or inguinal testes, predominantly female external genitalia, well-developed Wolffian structures, a blind vagina, and no uterus or tubes. Development of internal male structures has been attributed to the activity of 17β-HSD isozymes other than 17β-HSD3 converting testis-derived androstenedione to testosterone in the proximity of the Wolffian ducts. Undervirilization of external genitalia is due to insufficient testosterone at the level of the external genital primordia to act as a substrate for 5α-reductase conversion to DHT. Most patients are raised female and seek medical attention at puberty because of amenorrhea, poor breast development, and severe virilization sometimes accompanied by gynecomastia. At puberty, affected 46,XY individuals may develop male body habitus with abundant body and facial hair, enlargement of the penis and testes to normal adult size, and pigmentation and rugation of the labioscrotal folds. Approximately half of those not gonadectomized prepubertally switch to a male gender role, and individuals with the same mutation even within the same family may choose differently.[116] Biological findings include massive elevation in androstenedione compared to testosterone, which may be elicited with an hCG stimulation test in prepubertal children. A testosterone/androstenedione ratio less than 0.8 is considered diagnostic.[116] In newborns and pubertal patients, LH is elevated, and FSH is normal. Plasma levels of ACTH, glucocorticoid, and mineralocorticoid hormones are normal, which is expected because 17β-HSD is not involved in adrenal function. 17β-HSD occurs due to mutations in *HSD17B3* gene, located at chromosome 9q22.32 and inherited as an autosomal recessive trait. 17β-HSD deficiency is a rare disease, though incidence is higher in some regions, including the Gaza Strip. Over 45 mutations in *HSD17B3* have been reported, with the most common mutations being c.277 + 4A>T and p. R80Q.[117]

5α-Reductase Type 2 Deficiency. Testosterone itself is not a potent androgen; its reduced metabolite DHT is the main virilizing agent during male reproductive development. DHT is formed mainly in target tissues through reduction of testosterone by the microsomal enzyme 5α-reductase type 2, coded by the *SRD5A2* gene (located at 2p23.1). It is expressed in androgen target tissues such as the external genitalia, the urogenital sinus, and the prostate, and also the Wolffian ducts after they have differentiated into vasa deferentia and their accessories. Thus, testosterone, not DHT, controls Wolffian early development. Another functional isoenzyme of 5α-reductase, with a

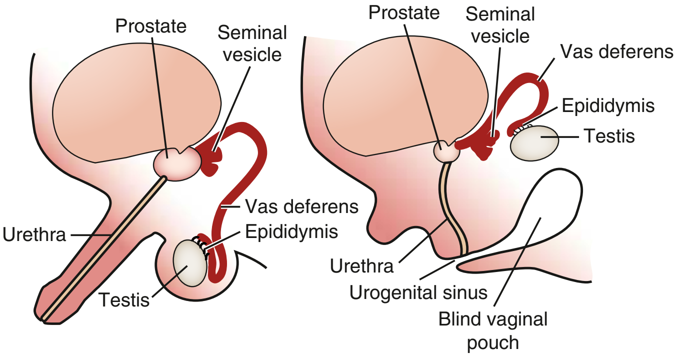


Figure 101.13 **Male genitalia.** Genitalia of a normal male *(left)* and that of a subject with 5α-reductase deficiency *(right),* illustrating the relative roles of testosterone and DHT in male sexual development. (Organs dependent on testosterone are shown in *red,* and those dependent on DHT are in the *lighter shaded* areas). (From Rey and Josso 119-9.Adapted from Imperato-McGinley J, Peterson RE. Male pseudohermaphroditism: The complexities of male phenotypic development. *Am J Med.* 1976;61:251–272.)

different pH optimum, 5α-reductase type 1, 50% homologous to type 2, is expressed after birth in the liver, transiently in neonatal skin, and permanently in skin from the time of puberty,[118] and its emergence probably accounts for pubertal virilization.

Steroid 5α-reductase isoform 2 deficiency (OMIM 264600) is a recessive autosomal disorder that is frequent in communities of Turkey, the Dominican Republic, and Papua New Guinea where consanguinity occurs, but relatively rare in other parts of the world.[119] It is clinically very similar to 17β-HSD3 deficiency, as both conditions are characterized by a poor virilization of the genitalia at birth, often limited to clitoromegaly; however, there is marked variability of the aspect of the genitalia even within the same family.[120] Microphallus associated with hypospadias and a urogenital sinus is not uncommon, and microphallus without genital ambiguity has even been reported.[119] At surgery, normal Wolffian duct derivatives are found that open in a blind vaginal pouch, and there is no uterus or tubes. The testes are usually extraabdominal, in the inguinal canal, the labia, or scrotum. More than two thirds of the patients are raised female, sometimes unequivocally until puberty. Then, if the testes have been left in place, the voice deepens, muscle mass and body hair increase, the phallus enlarges, a typical masculine habitus develops, and some patients transition their gender identity from female to male (Fig. 101.13). Gynecomastia is rare, and the prostate remains small. The sperm count is variable, from azoospermia to normal,[121] and fertility has been reported.[122] Infants with predominantly male phenotype are typically raised as males; however, some advocate for male gender assignment for all infants, given the occurrence of virilization later in life.

Biochemical findings include elevated testosterone/DHT ratio, when testosterone and DHT are expressed in the same units. Basal concentrations of testosterone and DHT are often insufficient to make the diagnosis prior to puberty, and hCG stimulation testing is usually required in prepubertal children. While the testosterone/DHT ratio in affected individuals most often exceeds 20,[120] lower ratios (>8.5) may be seen in the neonatal period.[123] Unlike in androgen insensitivity, where AMH is elevated, AMH serum values are not elevated in 5α-reductase type 2 deficiency, indicating that testosterone does not need to be converted to DHT to inhibit AMH production by Sertoli cells.[124]

More than 50 different mutations in *SRD5A2* have been reported, most commonly point mutations. Homozygous mutations are seen in over a third of patients, consistent with frequent finding of familial consanguinity. Other patients have compound heterozygote mutations.[119]

Androgen Insensitivity. Testosterone and DHT exert their action on androgen-dependent tissues by binding to the AR, which is a 110-kD member of the steroid receptor family. A long N-terminal domain, involved in transactivation, is coded by the first exon and harbors polyglutamine and polyglycine repeats. A conserved cysteine-rich DNA-binding domain (exons 2 and 3) is connected by a short hinge region to the ligand-binding domain (exons 4–8).[125] After binding to DHT or testosterone in the cytoplasm, the receptor is translocated to the nucleus, where it activates androgen-response elements in the promoter of androgen-responsive genes and initiates transcription. Nuclear localization is controlled by a nuclear localization signal spanning part of the DNA-binding domain and the hinge region, competing with an androgen-regulated nuclear export signal in the ligand-binding domain.[126] AR transactivation is enhanced by coregulators.[127,128] Eleven alpha-helices and four short beta-strands are involved in the formation of the ligand-binding pocket and the interaction surface for coactivators.[129]

The *AR* gene maps to Xq12, and its eight exons span approximately 186 kb. Over 1000 *AR* gene mutations have been recorded in the McGill University database (http://www.androgendb.mcgill.ca).[130] Mutations in the AR cause abnormalities in the formation of the steroid receptor–coregulator complex, its nuclear transport, its binding to DNA, or its transcriptional activation ability, causing complete or partial androgen insensitivity syndromes (PAIS). Inheritance is recessive X-linked and consequently causes DSD only in 46,XY individuals. The DSD phenotype in 46,XY individuals is variable, ranging from normal external female genitalia in complete androgen insensitivity syndrome (CAIS) to undervirilized male genitalia in PAIS. Additionally, more mild forms exist in 46,XY individuals associated with infertility and/or gynecomastia but with normal external male genitalia.

Complete Androgen Insensitivity Syndrome. CAIS is the most common cause of DSD in 46,XY completely feminized subjects, with an estimated prevalence of between 1/20,000 and 1/99,000.[131] Patients appear as normal females at birth. They may be diagnosed in infancy because of an inguinal hernia containing a palpable gonad that turns out to be a testis, or prenatally because of discordance between genotype and phenotype. Chorionic villus sampling at 10 to 12 weeks followed by sequencing of the AR in XY fetuses is useful in carrier mothers. Otherwise, patients come to medical attention at puberty because of primary amenorrhea. Feminine body contours and generous breast development are results of unopposed action of estrogens derived from androgen aromatization. Pubic and axillary hair are absent or sparse. The vagina is short and blind-ending, and the uterus is absent. Final height in CAIS is above normal mean female height, likely related to growth-controlling genes on the Y chromosome.[132]

Although Wolffian duct derivatives are not expected to develop in CAIS, Wolffian structures have been described in some cases and are thought to be related to residual activity of the AR that responds to the high local concentrations of androgen.[133] Müllerian remnants may also be found.[134] The testes are usually located in the inguinal canal or the labia majora. Testes were previously removed as early as possible after diagnosis due to concerns for risk of germ cell tumors, necessitating long-term estrogen supplementation. However, as the risk of malignant degeneration appears to be negligible before puberty, and relatively low in adults (estimated risk of 0.8%–2%), it is now recommended to delay gonadectomy to allow for spontaneous puberty through the aromatization of testosterone.[29,135] If gonadectomy is deferred, gonads may be monitored regularly by ultrasound or magnetic resonance imaging (MRI); however, these surveillance methods have not been systematically evaluated.[136] Vaginal dilation may be initiated after puberty or prior to initiating sexual activity, with the majority of patients achieving satisfactory response without requiring vaginoplasty.[137] Because of their apparently normal female phenotype and absence of exposure to male androgens, a female gender assignment is universally recommended.

Serum testosterone level is either within or above the normal range, and LH levels are high, but this pattern is evocative of hormone resistance and develops only after puberty. In the infant at 30 and 60 days of age, plasma testosterone values are below the normal range, and plasma LH values are low.[138] In contrast, the testosterone response to hCG is similar to or greater than that of normal infants. Plasma AMH levels may be normal or above the male reference range, relating to lack of inhibition of Sertoli cell function. Postpuberty, estradiol levels may be normal or slightly elevated for a male patient.[139] FSH levels are usually normal.

CAIS results from *AR* mutations that completely abolish the capacity of the receptor to mediate androgen action, for instance the creation of stop codons and frameshifts leading to premature termination of transcription resulting from insertions, deletions, or splice mutations. Most CAIS mutations occur in the ligand-binding domain, and nearly all mutations in exon 1 lead to CAIS.[130]

Partial Androgen Insensitivity Syndrome. The incidence of PAIS is difficult to determine, because some groups label as PAIS many 46,XY DSD cases of unknown etiology, even if *AR* mutations have not been found.[140] Clinically, PAIS is defined by some degree of virilization, which may range from predominantly female genitalia with slight degree of labial fusion to phenotypic males with hypospadias and micropenis. The aspect at birth is nonspecific, and diagnosis is difficult in the absence of family history. Hormone levels in individuals with PAIS may be similar to those with CAIS.[139] At puberty, scanty pubic and axillary hair and gynecomastia are suggestive of the diagnosis.

Whereas *AR* mutations are found in nearly all CAIS cases with usual sequencing methods, they can be detected in a large proportion of generation methods. Variants in exons 5 to 6 that encode the ligand-binding domain are more frequent in PAIS.[139] There is poor genotype/phenotype correlation, and the same mutation can result in widely different phenotypes even within the same kindred.[139]

Establishing a diagnosis of PAIS may be challenging, and other causes of 46,XY DSD must be considered. A short course of course of testosterone treatment (25 mg monthly for 3 months) may be useful to test androgen responsiveness; however, a positive response in infancy does not predict normal penile development at the time of puberty. Individuals with mild PAIS may present with male infertility and/or pubertal gynecomastia; however, external genitalia are normal. Elevated serum LH and testosterone levels may be found.[139] Most *AR* mutations leading to mild PAIS are found in the region encoding the N-terminal domain and involve single-base substitutions.[130]

Persistent Müllerian Duct Syndrome. Defective AMH production or sensitivity, which is rare, is characterized by the existence of uterus and fallopian tubes in an otherwise normally virilized patient. Regression of Müllerian ducts normally occurs between 8 and 10 weeks of fetal development, under the influence of AMH, a glycoprotein dimer member of the TGF-β family abundantly produced by fetal Sertoli cells. If active AMH is not produced, or if Müllerian ducts are insensitive to its action, Müllerian ducts develop into the uterus, fallopian tubes, and upper vagina notwithstanding normal virilization of external genitalia and urogenital sinus. Persistent Müllerian duct syndrome (PMDS) (OMIM 261550) is a rare, autosomal recessive condition that most commonly results from mutations of the *AMH* or AMH receptor type II genes, though idiopathic forms with no identified molecular cause have also been reported. The presence of Müllerian derivatives is often discovered incidentally; however, PMDS may also present as bilateral cryptorchidism (with the uterus fixed in the pelvis and both testes embedded in the broad ligament in an "ovarian" location), unilateral cryptorchidism, or with both testes in the same scrotal sac (a condition known as "transverse testicular ectopia").[142] Various congenital anomalies may occur with PMDS, including intestinal defects, renal anomalies, and low birth weight. In general, patients are infertile, although some cases of normal spermatogenesis have been reported.[142] Homozygous females are asymptomatic and fertile. AMH levels are low or undetectable in those with *AMH* mutations and normal in those with receptor mutations. Other hormone levels including gonadotropins and testosterone are normal.

Other Causes of 46,XY Differences of Sex Development. Although isolated hypospadias is quite common, evaluation for DSD should be considered in any infant with severe (penoscrotal) hypospadias, or in cases of milder hypospadias accompanied by micropenis or undescended testes. Rarely, 46,XY DSD has been associated with *in utero* exposure to endocrine disruptors, including phenytoin, phenobarbital, and other environmental exposures; however, a causative relationship has not clearly been established in humans.[143,144] Among individuals with 46,XY DSD of unclear etiology, nearly a third had prematurity and intrauterine growth restriction.[103] Although the mechanism for this association between low birth weight and hypospadias is unclear, placental insufficiency has been theorized as a possible cause.[145] 46,XY DSD may also be associated with genetic syndromes (see the section entitled "*Syndromic DSD*").

KEY POINTS

- Disruption in critical hormone signaling during fetal development may result in 46,XY difference in sex development (DSD). Insufficient androgen exposure leads to undervirilized external genitalia, while Müllerian derivatives may develop in the absence of anti-Müllerian hormone (AMH). Causes of 46,XY DSD include differences in gonadal development, disorders in androgen synthesis or action, disorders in AMH synthesis or action, and genetic syndromes.

Sex Chromosome Differences of Sex Development

Sex chromosome aneuploidies usually do not represent a differential diagnosis in patients with ambiguous genitalia; however, they are briefly described here because of their atypical sex chromosome constitution.

Turner Syndrome. TS is a chromosomal disorder that involving phenotypic females with a single intact X chromosome and complete or partial absence of the second sex chromosome, along with characteristic clinical manifestations. TS occurs in approximately 25 to 50 per 100,000 females. Forty to fifty percent of TS cases are caused my monosomy X (46,X); however, a variety of other chromosomal abnormalities may result in the condition, including mosaic forms (e.g., 45,X/46,XX, 45,X/47,XXX, 45,X/46,XX/4,XXX), mixed gonadal dysgenesis (45,X/46,XY), X chromosome deletions (e.g., deletion Xp22.3), ring X chromosomes (46,X,r(X)/46,XX), and isochromosomes made from two parts of the long i(X)q or two parts of the short chromosome segment i(Xp).[146] The syndrome spans a wide spectrum of clinical presentations.[147] The main characteristics are abnormal physical features, gonadal dysgenesis, and short stature. Follicles do not develop in the fetal gonad, which evolve to streaks.[148] Gonadotropin levels are found to be above the normal range[149] but may be normal between 6 and 9 years of age,[150] and AMH is undetectable[151] throughout childhood in girls with a 45,X karyotype. Children with TS may also have a short, thick, and webbed neck, hypertelorism, epicanthus and ptosis of the eyelids, low posterior hairline, congenital lymphedema, broad childlike chest, cubitus valgus and short fourth metacarpals, multiple pigmented nevi, intestinal telangiectasia, recurrent otitis media, and a tendency for keloid formation. TS patients have an increased incidence of celiac disease, cardiovascular defects (especially coarctation of the aorta and nonstenotic bicuspid aorta), kidney anomalies, chronic lymphocytic thyroiditis, diabetes mellitus, rheumatoid arthritis, and inflammatory bowel disease. Bone age is delayed, and bone mineral density may be affected. Compared to individuals with a 45,X karyotype, individuals with mosaicism (45,X/46,XX) tend to have a milder phenotype and may exhibit partial gonadal dysgenesis. These individuals may be diagnosed following spontaneous pubertal onset due to development of primary amenorrhea or premature ovarian failure. TS patients with a Y chromosome have an increased risk of gonadoblastoma, and gonadectomy is recommended in these individuals.

Klinefelter Syndrome. Klinefelter syndrome, the most common cause of male hypogonadism, with a prevalence of approximately 1.5 in 1000 males, was first described in 1942 as the association of eunuchoidism, gynecomastia, small testes, azoospermia, and elevated FSH excretion.[152] Several groups subsequently reported that a high proportion of patients with this syndrome were chromatin-positive, a finding explained in 1959 when the XXY constitution was reported.[153] Patients with the 47,XXY genotype have classic Klinefelter syndrome. Since the mid-1980s, the clinical spectrum of the syndrome has expanded to include patients with pure cell lines or mosaicisms with two or more X chromosomes and a Y chromosome.[154] However, some authors warn that patients with 48,XXYY, 48,XXXY, and 49,XXXXY polysomies present more complex physical, medical, and psychological phenotypes and should not be considered as having Klinefelter syndrome, even though they share features of hypergonadotropic hypogonadism.[155]

Gonadal dysgenesis primarily affects germ cells beginning in early fetal life[156] but does not impair Leydig and Sertoli cell function until midpuberty.[157] Therefore, internal and external genitalia are normally virilized in Klinefelter patients. Generally, at birth, no particular stigmata occur in Klinefelter syndrome. During childhood, no specific signs or symptoms are noted, except for bilateral cryptorchidism, learning disabilities, or behavior disorders, which are frequent.[158] Klinefelter boys are long-legged, most probably as the result of an overdose of *SHOX* gene expression.[46]

Pubertal development usually occurs spontaneously. In all cases, the predominant features are the small size of the testes (100%), which do not increase normally at puberty, and pubertal gynecomastia (60%). At puberty, as the gonadotropins increase, the seminiferous tubules do not enlarge but rather undergo fibrosis and hyalinization, which results in small, firm testes. Obliteration of the seminiferous tubules results in azoospermia. The marked reduction of tubular volume accounts for the apparent hyperplasia of the interstitial tissue. Leydig cells are present in clumps, and although their total mass is normal, they are functionally abnormal, resulting in a reduced testosterone production rate with compensatory LH hypersecretion. Because of the high LH levels, estradiol secretion is stimulated, with an increased estradiol/testosterone ratio, which is believed to account for the development of gynecomastia.[46,154]

47,XYY Males and 47,XXX Females. Although the prevalence of these conditions is approximately one in 1000 live births,[158] they are rarely diagnosed. Most cases are oligo- or asymptomatic. Gonadal development and function are normal in the vast majority of cases, except for a few XXX women who have been reported to develop premature ovarian failure.[159] Children are long-legged, with accelerated growth from infancy resulting in increased height at pubertal onset, and they may show behavior disturbances.[158]

Mosaic 45,X/46,XY Mixed Gonadal Dysgenesis. Mixed (or asymmetric) gonadal dysgenesis has been described in the section entitled "*46, XY Differences in Gonadal Development*."

Syndromic Differences of Sex Development

DSD in 46,XX and 46,XY may be isolated, with only gonadal or genital development affected, but may also be part of a syndrome affecting multiple organs. Syndromes involving a DSD symptom, as well as their genetic etiology, are presented in Table 101.2.

KEY POINTS

- Differences of sex development (DSD) are a spectrum of phenotypes encompassing variations of the development of the sex chromosomes, gonads, or urogenital tract. Pathophysiology of DSD depends on genetic and endocrine causes, which can be classified as 46,XX DSD, 46,XY DSD, and sex chromosome DSD.

EVALUATION AND MANAGEMENT OF DIFFERENCES OF SEX DEVELOPMENT

Management of DSD is complex, as the conditions may be associated with increased infertility, cancer, and gender dysphoria risks, as well as pervasive challenges to psychosocial adaptation for patients and families.[160-165] Correlation between treatment interventions (surgical, hormonal, psychosocial) and their clinical outcomes is poorly understood. This suggests that further research in DSD pathophysiology is needed to provide optimal evidence-based clinical management. Major challenges in DSD management include lack of standardization in descriptions of phenotype and medical/surgical interventions, resulting in variations in diagnostic and treatment practices across medical, surgical, and behavioral aspects of health.[164] Substantial unexplained variance in outcomes is revealed by studies examining the relationship between molecular diagnosis, gender development, and psychosocial trajectories.[166] These challenges complicate the principles of clinical management of DSDs, with disagreements among and between healthcare providers, advocacy groups, and patient communities regarding the definition of optimal care.[167] International efforts are underway to standardize practice and diagnosis.[30,168]

Tools in Differences of Sex Development Evaluation

Genetic Testing. From a diagnostic standpoint too, DSDs are challenging. For example, 46,XY disorders of androgen biosynthesis, which include SRDA2, 17β-HSD, 3β-HSD, StAR, CYP17A1, and POR deficiencies, are often difficult to distinguish clinically. However, an early genetic diagnosis is critical, as these conditions have drastically different natural histories requiring different management (recently reviewed in [24]). The various tools used for genetic evaluation, their clinical availability, and their turnaround time are described in Table 101.5.

Hormonal Testing. Biochemical assessments to evaluate gonadal and adrenal function are useful for diagnosis of DSD. To best evaluate gonadal function in an infant, timing of sex-steroid measurement should correspond to activation of the HPG axis. In boys, serum testosterone may rise for approximately 12 hours after birth, corresponding to a brief physiologic surge in LH after delivery.[169] The HPG axis is

TABLE 101.5 Clinical Testing Approaches to Diagnosis of Differences in Sex Development

Technique	Time Taken	Variants Detected					Resolution	Diagnostic Yield
		Aneuploidies/ Mosaics	Copy-Number Variants (CNVs)	Coding SNVs/ splice variants/indels	Noncoding SNV	Structural Variants (SVs)		
CLINICALLY AVAILABLE								
Karyotype	1–2 weeks	🟦				🟨	>5 Mb	15% (mostly mosaics)
Interphase FISH* *X, Y, and *SRY* markers	<3 days	🟦					N/A	Rapid sex determination
Microarray	2–3 weeks	🟦	⬛				>50 kb	15%
Single gene test/ panel	Up to 6 weeks			🟧			SNV: 1 nucleotide, Indels: <50 bp	Variable/panel-dependent
Exome sequencing	Up to 12 weeks* *<1 week possible if urgent			🟧			SNV: 1 nucleotide, Indels: <50 bp	30%–45%
Whole-genome sequencing	Up to 16 weeks		* ⬛	🟧	🟥	* 🟨	SNV: 1 nucleotide, CNV	At least 30% –45%

*Note: requires validation
SNV, single-nucleotide variant; *FISH*, fluorescence *in situ* hybridization.

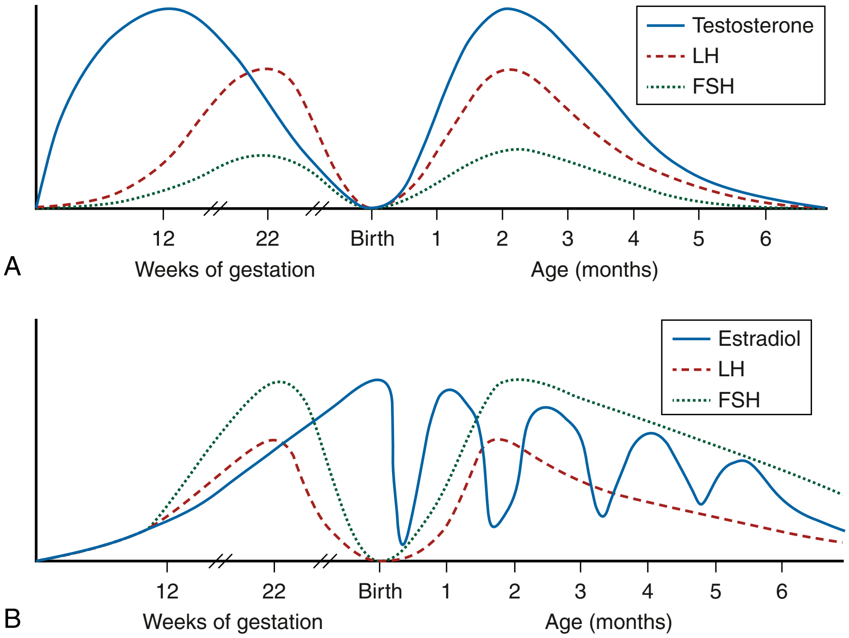


Figure 101.14 Patterns of fetal and postnatal gonadotropins and sex steroids. **A,** Males. **B,** Females. *LH*, Luteinizing hormone; *FSH*, follicle-stimulating hormone. (Adapted from Lanciotti L, Cofini M, Leonardi A, et al. Up-to-date review about minipuberty and overview on hypothalamic-pituitary-gonadal axis activation in fetal and neonatal life. *Front Endocrinol (Lausanne)*. 2018.)

subsequently suppressed but reactivated by around 1 week of life at the start of the "minipuberty" of infancy. Testosterone levels peak between 2 to 10 weeks of life but subsequently decline by 4 to 6 months at the end of the minipuberty. In girls, gonadotropin levels also begin to rise within the first week of life; however, FSH levels are higher and may remain elevated until 3 to 4 years of age. Estradiol levels often fluctuate during the minipuberty and typically decline by the second year of life[169] (Fig. 101.14).

Measurement of LH is useful to confirm activation of the HPG axis; if the LH level is low in the first 2 weeks of life, testing may be repeated at 1 to 3 months of life. AMH and inhibin B, both secreted by Sertoli cells, can be useful markers of testicular tissue even beyond the minipuberty period. Boys have significantly higher AMH values (>30 ng/dL) compared with girls, from infancy to puberty. Evaluation for ovarian tissue is more challenging due to the absence of a reliable marker, particularly in cases where testicular tissue may also be present (i.e., in OT-DSD). Proposed approaches to detect ovarian tissue have included measurement of inhibin A levels[170] and estradiol response to human menopausal gonadotropins[171]; limited evidence exists to support these methods. In older infants and prepubertal children with male-range

AMH, hCG stimulation testing may be utilized to evaluate testicular hormonal function. Protocols for hCG stimulation testing vary widely but often involve multiple doses of hCG given over 2 to 3 days, as well as laboratory measurements (typically testosterone, androstenedione, and DHT) at baseline and poststimulation.[172] The interpretation of hCG stimulation test results is also variable and may depend on testing protocol and other subject characteristics; however, stimulated ratios of testosterone to other androgen metabolites may be informative to evaluate for disorders of androgen biosynthesis or metabolism (e.g., ratio of testosterone to DHT of >10 may suggest 5α-reductase type 2 deficiency and ratio of testosterone to androstenedione <0.8 may suggest 17β-hydroxysteroid dehydrogenase type III deficiency). Given the increasing ability to provide genetic diagnoses for DSD, hCG stimulation may not always be required. Assessment of adrenal function is critical for both for diagnostic purposes and to rule out life-threatening complications of CAH. Serum electrolytes including sodium, potassium, and glucose should be measured in any infant in whom CAH may be possible. Aldosterone and plasma-renin activity may be useful to evaluate for salt-wasting CAH phenotype. 17-OHP is used to evaluate for 21-OHD (the most common cause of 46,XX DSD); basal levels will be markedly elevated in classic 21-OHD. Other adrenal metabolites should be measured when evaluating rarer forms of CAH (see Table 101.4); however, basal levels may not always be sufficient, and often the ACTH stimulation test is required to satisfy the diagnosis (see Chapter 94 for details). Plasma or serum steroid measurements should be performed by liquid chromatography–mass spectroscopy, or alternatively, immunoassays after extraction given the potential for crossreactivity.[172]

Imaging Studies and Other Procedures. Imaging modalities are utilized to detect internal genital structures (i.e., presence or absence of a uterus) and to localize gonads. Ultrasonography is the primary imaging modality in DSD evaluation and is particularly useful in the neonatal period, when the uterus and ovaries are prominent due to maternal hormonal stimulation. MRI can provide more detailed anatomic information and may be more sensitive than ultrasound in detection of intraabdominal gonads.[173] Voiding cystourethrography and genitography may be used to define the internal genital anatomy and to delineate the level of confluence of the vagina and urethra. Computed tomography is the preferred modality to evaluate for malignancies in DSD. Surgical procedures are also utilized in DSD evaluation. An examination under anesthesia combined with a diagnostic vaginoscopy is commonly used to evaluate the lower genital tract in pediatric and adolescent patients.[174] Diagnostic laparoscopy and gonadal biopsy are useful to evaluate gonadal status in complex DSD conditions.

Initial Evaluation of a Newborn with Atypical Genital Appearance

The first step in the DSD evaluation of a newborn is a thorough clinical examination to determine if gonads are palpable (in the labioscrotal folds or inguinal canal), as this will determine the choice and order of the investigation. Infants with atypical genital appearance and nonpalpable gonads must immediately be evaluated for virilizing CAH, due to the possibility of life-threatening complications related to glucocorticoid and/or mineralocorticoid deficiencies. As gonads palpable below the inguinal ligament typically contain testicular tissue, 46,XY DSD may be more likely in this situation. Absence of a uterus (which may be ascertained by ultrasonography) also suggests the presence of testicular tissue that produced AMH. Rapid identification of sex chromosomes via karyotype or other means (e.g., fluorescence *in situ* hybridization for *SRY*) help characterize the DSD (i.e., as 46,XX, 46,XY or sex chromosome DSD), which will provide a framework for the differential diagnosis and subsequent workup (Fig. 101.15). A detailed physical examination may reveal other findings suggestive of the underlying diagnosis. For example, pigmentation of the genitalia and nipples or elevated blood pressure may suggest a form of CAH (see Table 101.4). Other congenital anomalies may suggest an underlying genetic syndrome associated with DSD (see Table 101.2). Family history may be informative to elucidate a genetic mode of transmission; maternal gestational history should also be obtained including prenatal exposures or occurrence of maternal virilization (see the sections entitled "*Maternal-derived Androgens*" and "*Aromatase Deficiency*").

In infants with suspected 46,XX DSD, serum 17-OHP level and electrolytes should be measured to rule out CAH due to 21-OHD. Close monitoring is required, as salt wasting may present at a few days to several weeks of life. Findings of extremely high levels (to 10,000 ng/dL or 300 nmol/L) of 17-OHP are diagnostic of 21-OHD, whereas mildly elevated 17-OHP levels may suggest other forms of virilizing CAH; diagnosis may require measurement of other adrenal metabolites and ACTH stimulation testing (see the section entitled "*Virilizing Congenital Adrenal Hyperplasias*" and Table 101.4). In virilized 46,XX infants, an AMH level above the female reference range is suggestive of testicular tissue; therefore, 46,XX testicular or OT-DSD should be considered.

Evaluation of 46,XY DSD depends on measured testosterone and AMH levels. As discussed above, testosterone should be measured during the "minipuberty" of infancy and with concomitant gonadotropins to ensure activation of the HPG axis. Individuals with low testosterone and AMH are likely to have testicular dysgenesis, as both Leydig and Sertoli cell populations are affected. When testosterone is low but AMH is within or above normal, only Leydig function is affected (see Fig. 101.8). Adrenal function must be assessed to rule out rare forms of CAH that impact both adrenal and testicular steroidogenesis (e.g., LCAH, 3β-HSD, P450c17 deficiency, POR deficiency, see Table 101.4). If adrenal function is normal, disorders that impact testicular biosynthesis or secretion (e.g., Leydig cell hypoplasia, 17β-HSD) should be considered. When both testosterone and AMH are within or above the male range, primary hypogonadism can be ruled out, and androgen end-organ defects should be suspected. In patients with completely female genitalia, CAIS is most likely. Conversely, in patients with ambiguous genitalia, PAIS or 5α-reductase type 2 deficiency should be considered. A high testosterone/DHT ratio is predictive of 5α-reductase type 2 deficiency; however, elevated ratios can sometimes occur in androgen insensitivity syndrome because the expression of 5α-reductase type 2 gene (*SRD5A2*) is upregulated by androgens. Elevated AMH is suggestive of PAIS, in that AMH is normal in 5α-reductase type 2 deficiency, but overlapping can be seen.

Initial Gender Assignment

Gender uncertainty in a newborn is extremely stressful for the family. Speed in achieving a thorough assessment and reaching a decision is required. However, erroneous gender assignment and the need for reassignment could be more harmful for the patient's future than delaying the decision. Although in some cases rapid chromosomal, genetic, and biochemical analyses may lead to an immediate diagnosis and gender assignment, in other situations weeks are needed before a decision can be made. The clinician should tell the parents that recognition of the sex of the child is difficult on simple external examination and can be achieved only after complete investigation rather than saying that the child is neither a boy nor a girl. It is prudent to delay the mention of sex on the birth certificate until biological and molecular results are available to an expert team of specialists and the parents have had time to consider the various options. Factors that influence gender assignment include diagnosis, genital appearance, surgical options, need for lifelong hormone replacement therapy, potential for fertility, views of the family, and, sometimes, circumstances related to cultural practices.

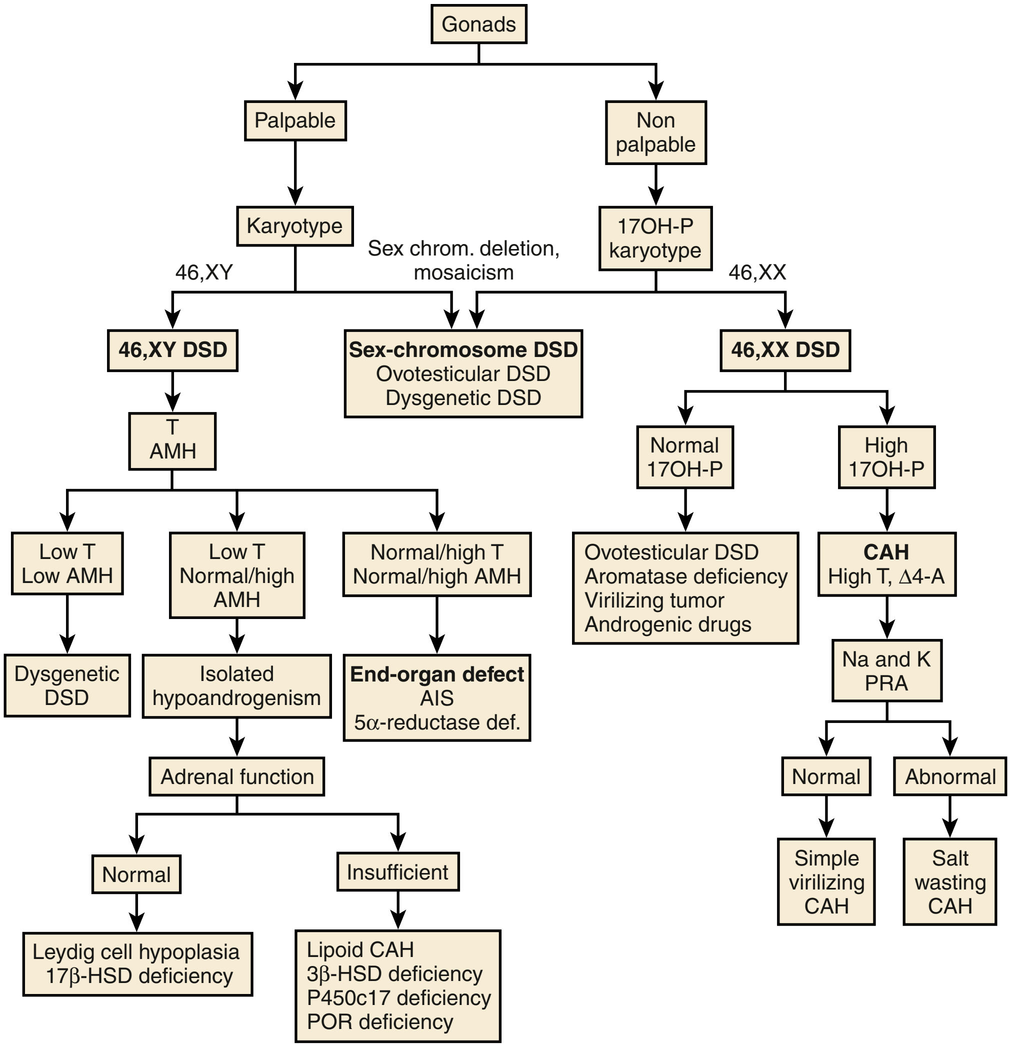

Gonads
Palpable
Non palpable
Karyotype
17OH-P karyotype
46,XY
Sex chrom. deletion, mosaicism
46,XX
46,XY DSD
Sex-chromosome DSD
Ovotesticular DSD
Dysgenetic DSD
46,XX DSD
T
AMH
Normal 17OH-P
High 17OH-P
Low T
Low AMH
Low T
Normal/high AMH
Normal/high T
Normal/high AMH
Ovotesticular DSD
Aromatase deficiency
Virilizing tumor
Androgenic drugs
CAH
High T, Δ4-A
Dysgenetic DSD
Isolated hypoandrogenism
End-organ defect
AIS
5α-reductase def.
Na and K
PRA
Adrenal function
Normal
Abnormal
Normal
Insufficient
Simple virilizing CAH
Salt wasting CAH
Leydig cell hypoplasia
17β-HSD deficiency
Lipoid CAH
3β-HSD deficiency
P450c17 deficiency
POR deficiency

Figure 101.15 Flow chart of the initial steps in the diagnosis of differences of sex development (DSD). (From Rodolfo A. Rey, Nathalie Josso, Chapter 119 - Diagnosis and Treatment of Disorders of Sexual Development⁎, Editor(s): J. Larry Jameson, Leslie J De Groot, David M. de Kretser, Linda C. Giudice, Ashley B. Grossman, Shlomo Melmed, John T. Potts, Gordon C. Weir, Endocrinology: Adult and Pediatric (Seventh Edition), W.B. Saunders, 2016, Pages 2086-2118.e5, ISBN 9780323189071, https://doi.org/10.1016/B978-0-323-18907-1.00119-0. (https://www.sciencedirect.com/science/article/pii/B9780323189071001190))

KEY POINTS

- Initial evaluation of a newborn with atypical genital appearance depends on the location of the gonads; newborns with nonpalpable gonads should be evaluated for congenital adrenal hyperplasia. Identification of sex chromosomes directs subsequent differences of sex development evaluation, which may include genetic testing, hormonal testing, and imaging studies. Gender assignment in newborns should be delayed until expert evaluation has occurred. Open communication with the family is essential.

Management of Differences of Sex Development

Differences of Sex Development Healthcare Teams. A key feature of care recommended for children with a DSD is the presence of a multidisciplinary team. Composition may vary, but this team often includes specialists from endocrinology, genetics, surgery, gynecology, mental health, and ethics. Access to advocacy and/or patient support groups is also extremely important in the context of patient-centered care. A shared decision-making protocol should be in place, emphasizing open communication between the team and the family. Protocols to ensure transition from childhood to adult care should be in place.

Hormonal Management. Sex steroids may be used for induction of puberty and for long-term hormone replacement in DSD patients. Insufficient sex hormone exposure may impact development of secondary sex characteristics, physical growth, bone mineral density, and psychological functioning. Choice of sex hormone replacement is based on the underlying diagnosis, sex of rearing, and gender identity of the patient. Testosterone supplementation, typically with intramuscular depot injections, is used for penile enlargement in infancy and for induction of male puberty in individuals with gonadal dysgenesis and disorders of androgen synthesis (i.e., Leydig cell hypoplasia and testicular steroidogenic defects) who are being reared as males. In males with mild 5α-reductase deficiency, virilization is feasible with androgen therapy (high doses of testosterone or DHT creams) and typically occurs spontaneously at puberty; however, the outcome is most likely satisfactory provided that they show a sufficient degree of virilization at birth. The situation is less clear-cut in PAIS patients, in that the response to androgen treatment is less efficient. Patients with PAIS reared as males may require supraphysiologic doses of testosterone for optimal effect. Topical DHT cream has also been used to treat micropenis in PAIS patients with variable success.[175] Transdermal estradiol is typically used for induction of female puberty in patients being raised as females; various formulations exist for long-term supplementation. Progesterone should also be utilized in patients with a uterus to avoid deleterious effects of unopposed estrogen on the endometrium. Gonadotropin-releasing hormone agonist therapy may be used to suppress endogenous production of pubertal sex hormones. This approach may be useful to prevent exposure to sex hormones that are inconsistent with the gender of rearing or gender identity and may preclude the need for early gonadectomy to prevent pubertal virilization in those being raised as females.

Patients with DSD due to CAH may be treated with glucocorticoid and mineralocorticoid supplementation; management of these patients is discussed in Chapter 94.

Surgery. The traditional approach of surgical care has been to create an anatomy matching the sex of rearing, with the goal of influencing gender identity and allowing sexual intercourse congruent with a heterosexual sexuality in the gender assigned. Needless to say, this view has become more and more controversial. Advocacy groups have challenged the practice of early genital surgery, citing patient autonomy as well as lack of long-term favorable outcome studies, with, in particular, the risk that sexual sensation and function could be impaired because of surgery.

The consensus statement has provided guidelines for a more cautious approach to surgery. The statement proposes that the consequences of surgery be explained in detail, including risks of complications. It also requires that surgeries be performed in centers of excellence by surgeons experienced in DSD care. The statement also suggests that clitoral surgery should only be considered in the most masculinized cases (Prader 3 to 5), and that no vaginal intervention (vaginoplasty) be performed in infancy. The debates about surgery have continued to rage, with a growing movement considering that genital surgery in infancy is equivalent to a human rights violation. In contrast, calls for obtaining outcome data comparing the effects of surgical intervention versus alternative care (nonsurgical) have been made, albeit with limited success.

Risk of Tumor Development. DSD patients with gonads containing Y-chromosome material may be at risk for gonadal malignancy. The risk appears to be greatest in those with dysgenetic gonads and in those with intraabdominal as compared to inguinal or scrotal gonads. The highest risk is among individuals with testis-specific protein Y-encoded (*TSPY*)-positive gonadal dysgenesis (including those with *WT1* mutations such as patients with Frasier or Denys–Drash syndromes) and PAIS who have intraabdominal gonads. Intermediate risk is seen in patients with Y chromosome–positive TS and 17β-HSD, with the lowest risk occurring in CAIS, TS without Y material, and OT-DSD with fully differentiated ovarian and testicular tissue.[29,160,176] The optimal management of DSD patients at risk for gonadal tumors is an area of active study. In the past, prophylactic gonadectomy was routinely performed early on; however, experts now recommend a risk-based approach. Given relatively low risk of gonadal malignancy, gonadectomy in patients with CAIS is now typically delayed until after puberty to avoid the need for hormonal replacement for pubertal induction.[136] Abdominal gonads may also be brought to the scrotum or other fixed anatomic position to better allow for active surveillance. Imaging studies and tumor markers may allow for monitoring of those with retained intraabdominal gonads, however, may be insufficient for early detection of precursor lesions. More research is needed to determine the optimal timing and indications for gonadectomy.

Psychosocial Management. The question of gender assignment at birth remains controversial. Although a majority of patients will not experience gender dysphoria, a significant minority may elect to reassign to a different gender.[177]

The assessment of psychosexual development and psychosocial screening is crucial to the overall management of individuals with a DSD and should be performed by experienced mental health professionals. Diagnosis disclosure, peer and support groups, discussions of gender reassignment, and potential anxiety associated with repeated genital examination/medical photography should be part of the overall team management process.

KEY POINTS

- Best management of differences of sex development (DSD) is patient-centered and involves a multidisciplinary team involved in diagnostic steps, clinical evaluation and management, psychosocial support, and shared decision-making. Hormone replacement therapy depends on the underlying diagnosis, gender of rearing, and gender identity. DSD patients with gonads containing Y-chromosome material may be at risk for gonadal malignancy; optimal timing and indication for gonadectomy in at-risk patients requires further study. Genital surgery in DSD patients is an area of controversy.

ACKNOWLEDGEMENTS

We acknowledge Rodolfo A. Rey, Nathalie Josso, Valerie A. Arboleda, and Charmian A. Quigley for prior text that was utilized in this updated chapter.

REFERENCES

1. Uhlenhaut NH, Jakob S, Anlag K, et al. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell.* 2009;139:1130–1142.
2. Matson CK, Murphy MW, Sarver AL, et al. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature*. 2011;476:101–105.
3. Morohashi KI. The ontogenesis of the steroidogenic tissues. *Genes to Cells*. 1997;2:95–106.
4. Capel B. The battle of the sexes. *Mech Dev.* 2000;92:89–103.
5. Merchant-Larios H, Moreno-Mendoza N. Onset of sex differentiation: dialog between genes and cells. *Arch Med Res.* 2001;32:553–558.
6. Warr N, Greenfield A. The molecular and cellular basis of gonadal sex reversal in mice and humans. *Wiley Interdiscip Rev Dev Biol.* 2012;1:559–577.
7. Meeks JJ, Crawford SE, Russell TA, et al. Dax1 regulates testis cord organization during gonadal differentiation. *Development*. 2003;130:1029–1036.
8. Brennan J, Karl J, Capel B. Divergent vascular mechanisms downstream of sry establish the arterial system in the XY gonad. *Dev Biol.* 2002;244:418–428.
9. McLaren A, Buehr M. Development of mouse germ cells in cultures of fetal gonads. *Cell Differ Dev.* 1990;31:185–195.
10. Tilmann C, Capel B. Mesonephric cell migration induces testis cord formation and Sertoli cell differentiation in the mammalian gonad. *Development*. 1999;126:2883–2890.
11. Capel B, Albrecht KH, Washburn LL, et al. Migration of mesonephric cells into the mammalian gonad depends on Sry. *Mech Dev.* 1999;84:127–131.
12. Coveney D, Cool J, Oliver T, Capel B. Four-dimensional analysis of vascularization during primary development of an organ, the gonad. *Proc Natl Acad Sci USA*. 2008;105:7212–7217.
13. Wang HU, Chen ZF, Anderson DJ. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. *Cell.* 1998;93:741–753.

14. Cotinot C, Pailhoux E, Jaubert F, et al. Molecular genetics of sex determination. *Semin Reprod Med*. 2002;20:157–167.
15. Tilmann C, Capel B. Cellular and molecular pathways regulating mammalian sex determination. *Recent Prog Horm Res.* 2002;57:1–18.
16. Warne GL, Kanumakala S. Molecular endocrinology of sex differentiation. *Semin Reprod Med*. 2002;20:169–179.
17. Köhler B, Lin L, Ferraz-de-Souza B, et al. Five novel mutations in steroidogenic factor 1 (SF1, NR5A1) in 46,XY patients with severe underandrogenization but without adrenal insufficiency. *Hum Mutat*. 2008;29:59–64.
18. Griswold MD, Hogarth CA, Bowles J, et al. Initiating meiosis: the case for retinoic acid. *Biol Reprod*. 2012;86:35.
19. Bleyl SB, Brauer PR, Larsen WJ, et al. Development of the urogenital system. In: *Human Embryology*. 4th ed. Churchill Livingstone; 2008.
20. Sadler TW. *Langman's Medical Embryology*. 9th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2003.
21. Rey R, Picard JY. Embryology and endocrinology of genital development. *Baillieres Clin Endocrinol Metab*. 1998;12:17–33.
22. Zhao F, Franco HL, Rodriguez KF, et al. Elimination of the male reproductive tract in the female embryo is promoted by COUP-TFII in mice. *Science*. 2017;357:717–720.
23. Croft B, Ayers K, Sinclair A, et al. Review disorders of sex development: the evolving role of genomics in diagnosis and gene discovery. *Birth Defects Res Part C - Embryo Today Rev.* 2016;108:337–350.
24. Délot EC, Vilain E. *Disorders of Sex Development*. 8th ed. Philadelphia, PA: Elsevier Inc.; 2018.
25. Ahmed F, Kyriakou A, Tobias E, et al. Disorders of sex development: advances in genetic diagnosis and challenges in management. *Adv Genomics Genet.* 2015;5:165–177.
26. Ohnesorg T, Vilain E, Sinclair AH. The genetics of disorders of sex development in humans. *Sex Dev.* 2014;8:262–272.
27. Ono M, Harley VR. Disorders of sex development: new genes, new concepts. *Nat Rev Endocrinol.* 2013;9:79–91.
28. Hughes IA, Houk C, Ahmed SF, et al. Consensus statement on management of intersex disorders. *Arch Dis Child*. 2006;91:554–563.
29. Lee PA, Houk CP, Ahmed SF, et al. Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. *Pediatrics*. 2006;118:e488–e500.
30. Délot EC, Papp JC, Délot EC, et al. Genetics of disorders of sex development: the DSD-TRN experience. *Endocrinol Metab Clin North Am.* 2017;46:519–537.
31. Hutson JM, Grover SR, O'Connell M, et al. Malformation syndromes associated with disorders of sex development. *Nat Rev Endocrinol*. 2014;10:476–487.
32. Lee PA, Nordenström A, Houk CP, et al. Global disorders of sex development update since 2006: perceptions, approach and care. *Horm Res Paediatr.* 2016;85:158–180.
33. Schiebel K, Winkelmann M, Mertz A, et al. Abnormal XY interchange between a novel isolated protein kinase gene, PRKY and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. *Hum Mol Genet.* 1997;6:1985–1989.
34. Cox JJ, Willatt L, Homfray T, et al. A SOX9 duplication and familial 46,XX developmental testicular disorder. *N Engl J Med*. 2011;364:91–93.
35. Vetro A, Ciccone R, Giorda R, et al. XX males SRY negative: a confirmed cause of infertility. *J Med Genet.* 2011;48:710–712.
36. Huang B, Wang S, Ning Y, et al. Autosomal XX sex reversal caused by duplication of SOX9. *Am J Med Genet.* 1999;87:349–353.
37. Refai O, Friedman A, Terry L, et al. De novo 12;17 translocation upstream of *SOX9* resulting in 46,XX testicular disorder of sex development. *Am J Med Genet Part A*. 2010;152A:422–426.
38. Xiao B, Ji X, Xing Y, et al. A rare case of 46, XX SRY-negative male with a ~74-kb duplication in a region upstream of SOX9. *Eur J Med Genet.* 2013;56:695–698.
39. Bose HS, Sugawara T, Strauss JF, et al. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. *N Engl J Med*. 1996;335:1870–1879.
40. Lee GM, Ko JM, Shin CH, et al. A Korean boy with 46,XX testicular disorder of sex development caused by SOX9 duplication. *Ann Pediatr Endocrinol Metab*. 2014;19:108–112.
41. Kim G-J, Sock E, Buchberger A, et al. Copy number variation of two separate regulatory regions upstream of *SOX9* causes isolated 46,XY or 46,XX disorder of sex development. *J Med Genet.* 2015;52:240–247.
42. Vetro A, Dehghani MR, Kraoua L, et al. Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3. *Eur J Hum Genet.* 2015;23:1025–1032.
43. Seeherunvong T, Perera EM, Bao Y, et al. 46,XX sex reversal with partial duplication of chromosome arm 22q. *Am J Med Genet.* 2004;127:149–151.
44. Rey RA, Belville C, Nihoul-Fékété C, et al. Evaluation of gonadal function in 107 intersex patients by means of serum antiMüllerian hormone measurement. *J Clin Endocrinol Metab*. 1999;84:627–631.
45. Vorona E, Zitzmann M, Gromoll J, et al. Clinical, endocrinological, and epigenetic features of the 46,XX male syndrome, compared with 47,XXY Klinefelter patients. *J Clin Endocrinol Metab*. 2007;92:3458–3465.
46. Aksglaede L, Skakkebaek NE, Juul A. Abnormal sex chromosome constitution and longitudinal growth: Serum levels of insulin-like growth factor (IGF)-I, IGF binding protein-3, luteinizing hormone, and testosterone in 109 males with 47,XXY, 47,XYY, or sex-determining region of the Y chromosome. *J Clin Endocrinol Metab*. 2008;93:169–176.
47. King NW. The reproductive tract. In: *Pathology of Laboratory Animals*. New York, NY: Springer; 1978:509–580.
48. Krob G, Braun A, Kuhnle U. True hermaphroditism: geographical distribution, clinical findings, chromosomes and gonadal histology. *Eur J Pediatr*. 1994;153:2–10.
49. van Niekerk WA, Retief AE. The gonads of human true hermaphrodites. *Hum Genet.* 1981;58:117–122.
50. Wiersma R. The clinical spectrum and treatment of ovotesticular disorder of sexual development. *Adv Exp Med Biol.* 2011;707:101–103.
51. Barseghyan H, Vilain E. Chapter 7 - The Genetics of Ovotesticular Disorders of Sex Development. In: New MI, Lekarev O, Parsa A, Yuen TT, O'Malley BW, Hammer GDBT-GSD, eds. San Diego: Academic Press; 2014:261–263. Available from: https://www.sciencedirect.com/science/article/pii/B978012416006400020X.
52. McElreavey K, Rappaport R, Vilain E, et al. A minority of 46,XX true hermaphrodites are positive for the Y-DNA sequence including SRY. *Hum Genet.* 1992;90:121–125.
53. Wiersma R, Ramdial PK. The gonads of 111 South African patients with ovotesticular disorder of sex differentiation. *J Pediatr Surg*. 2009;44:556–560.
54. Queipo G, Zenteno JC, Peña R, et al. Molecular analysis in true hermaphroditism: demonstration of low-level hidden mosaicism for Y-derived sequences in 46,XX cases. *Hum Genet.* 2002;111:278–283.
55. Hiort O, Gramss B, Klauber GT. True hermaphroditism with 46,XY karyotype and a point mutation in the SRY gene. *J Pediatr*. 1995;126:1022.
56. Abbas N, McElreavey K, Leconiat M, et al. Familial case of 46,XX male and 46,XX true hermaphrodite associated with a paternal-derived SRY-bearing X chromosome. *Comptes Rendus l'Academie des Sci - Ser III*. 1993;316:375–383.
57. Ramos ES, Moreira-Filho CA, Vicente YA, et al. SRY-negative true hermaphrodites and an XX male in two generations of the same family. *Hum Genet.* 1996;97:596–598.
58. Fechner PY, Rosenberg C, Stetten G, et al. Nonrandom inactivation of the Y-bearing X chromosome in a 46,XX individual: evidence for the etiology of 46,XX true hermaphroditism. *Cytogenet Cell Genet.* 1994;66:22–26.
59. Kuhnle U, Krob G, Maier E. True hermaphroditism: Presentation, management, outcome. *Endocrinologist*. 2003;13:214–218.
60. Parma P, Radi O, Vidal V, et al. R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat Genet.* 2006;38:1304–1309.
61. Tomaselli S, Megiorni F, De Bernardo C, et al. Syndromic true hermaphroditism due to an R-spondin1 (RSPO1) homozygous mutation. *Hum Mutat*. 2008;29:220–226.
62. Carvalheira G, Malinverni AM, Moysés-Oliveira M, et al. The natural history of a man with ovotesticular 46,XX DSD caused by a novel 3-Mb 15q26.2 deletion containing NR2F2 gene. *J Endocr Soc.* 2019;3:2107–2113.

63. López-Hernández B, Méndez JP, Coral-Vázquez RM, et al. Duplication of SOX9 associated with 46,XX ovotesticular disorder of sex development. *Reprod Biomed Online*. 2018;37:107–112.
64. Benko S, Gordon CT, Mallet D, et al. Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J Med Genet.* 2011;48:825–830.
65. Mengen E, Kayhan G, Kocaay P, et al. A duplication upstream of SOX9 associated with SRY negative 46,XX ovotesticular disorder of sex development: a case report. *J Clin Res Pediatr Endocrinol*. 2019;12(3):308.
66. Pinti E, Piko H, Lengyel A, et al. Similar cause, different phenotype: SOX9 enhancer duplication in a family. *Horm Res Paediatr.* 2020;92:335–339.
67. Bashamboo A, Donohoue PA, Vilain E, et al. A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development. *Hum Mol Genet.* 2016;25:3446–3453.
68. Eozenou C, Gonen N, Touzon MS, et al. Testis formation in XX individuals resulting from novel pathogenic variants in Wilms' tumor 1 (WT1) gene. *Proc Natl Acad Sci USA*. 2020;117:13680–13688.
69. Fortuño C, Labarta E. Genetics of primary ovarian insufficiency: a review. *J Assist Reprod Genet.* 2014;31:1573–1585.
70. Wittenberger MD, Hagerman RJ, Sherman SL, et al. The FMR1 premutation and reproduction. *Fertil Steril*. 2007;87:456–465.
71. Peng C. The TGF-beta superfamily and its roles in the human ovary and placenta. *JOGC*. 2003;25:834–844.
72. Harris SE, Chand AL, Winship IM, et al. Identification of novel mutations in FOXL2 associated with premature ovarian failure. *Mol Hum Reprod*. 2002;8:729–733.
73. Zangen D, Kaufman Y, Zeligson S, et al. XX ovarian dysgenesis is caused by a PSMC3IP/HOP2 mutation that abolishes coactivation of estrogen-driven transcription. *Am J Hum Genet.* 2011;89:572–579.
74. Bouilly J, Bachelot A, Broutin I, Novel NOBOX loss-of-function mutations account for 6.2% of cases in a large primary ovarian insufficiency cohort, et al. *Hum Mutat*. 2011;32:1108–1113.
75. Caburet S, Arboleda VA, Llano E, et al. Mutant cohesin in premature ovarian failure. *N Engl J Med*. 2014;370:943–949.
76. Fabbri-Scallet H, de Sousa LM, Maciel-Guerra AT, et al. Mutation update for the NR5A1 gene involved in DSD and infertility. *Hum Mutat*. 2020;41:58–68.
77. Aittomäki K, Lucena JL, Pakarinen P, et al. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell.* 1995;82:959–968.
78. Arnhold IJP, Latronico AC, Batista MC, et al. Clinical features of women with resistance to luteinizing hormone. *Clin Endocrinol (Oxf)*. 1999;51:701–707.
79. Quaynor SD, Stradtman Jr EW, Kim H-G, et al. Delayed puberty and estrogen resistance in a woman with estrogen receptor α variant. *N Engl J Med*. 2013;369:164–171.
80. Bulun SE. Clinical review 78: Aromatase deficiency in women and men: would you have predicted the phenotypes? *J Clin Endocrinol Metab*. 1996;81:867–871.
81. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency*. *Endocr Rev*. 2000;21:245–291.
82. Callegari C, Everett S, Ross M, et al. Anogenital ratio: measure of fetal virilization in premature and full-term newborn infants. *J Pediatr*. 1987;111:240–243.
83. van der Straaten S, Springer A, Zecic A, et al. The External Genitalia Score (EGS): A European Multicenter Validation Study. *J Clin Endocrinol Metab*. 2020:105(3):e222–e230.
84. Speiser PW, Arlt W, Auchus RJ, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society Clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:4043–4088.
85. Kai H, Nose O, Iida Y, et al. Female pseudohermaphroditism caused by maternal congenital adrenal hyperplasia. *J Pediatr*. 1979;95:418–420.
86. Auchus RJ, Chang AY. 46,XX DSD: the masculinised female. *Best Pract Res Clin Endocrinol Metab*. 2010;24:219–242.
87. Mendonca BB, Leite MV, De Castro M, et al. Female pseudohermaphroditism caused by a novel homozygous missense mutation of the GR gene. *J Clin Endocrinol Metab*. 2002;87:1805–1809.
88. Hakim C, Padmanabhan V, Vyas AK. Gestational hyperandrogenism in developmental programming. *Endocrinology*. 2017;158(2):199–212.
89. Grumbach MM, Ducharme JR. The effects of androgens on fetal sexual development. *Fertil Steril*. 1960;11:157–180.
90. Brunskill PJ. The effects of fetal exposure to danazol. *BJOG An Int J Obstet Gynaecol*. 1992;99:212–215.
91. Belgorosky A, Guercio G, Pepe C, et al. Genetic and clinical spectrum of aromatase deficiency in infancy, childhood and adolescence. *Horm Res.* 2009;72:321–330.
92. Simpson ER, Clyne C, Rubin G, et al. Aromatase - A brief overview. *Annu Rev Physiol*. 2002;64:93–127.
93. Lin Ercan O, Raza J, et al. Variable phenotypes associated with aromatase (CYP19) insufficiency in humans. *J Clin Endocrinol Metab*. 2007;92:982–990.
94. Zenteno JC, Carranza-Lira S, Kofman-Alfaro S. Molecular analysis of the anti-Müllerian hormone, the anti-Müllerian hormone receptor, and galactose-1-phosphate uridyl transferase genes in patients with the Mayer-Rokitnasky-Küster-Hauser syndrome. *Arch Gynecol Obstet*. 2004;269:270–273.
95. Londra L, Chuong FS, Kolp L. Mayer-Rokitansky-Kuster-Hauser syndrome: a review. *Int J Womens Health*. 2015;7:865.
96. Fontana L, Gentilin B, Fedele L, et al. Genetics of Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome. *Clin Genet.* 2017;91(2):233–246.
97. Jung EJ, Cho MH, Kim DH, et al. Herlyn-Werner-Wunderlich syndrome: an unusual presentation with pyocolpos. *Obstet Gynecol Sci.* 2017;60(4):374–377.
98. Suson KD, Preece J, Di Carlo HN, et al. Complexities of Müllerian anatomy in 46XX cloacal exstrophy patients. *J Pediatr Adolesc Gynecol*. 2016;29(5):424–428.
99. Ogata T, Matsuo N. Comparison of adult height between patients with XX and XY gonadal dysgenesis: support for a y specific growth gene(s). *J Med Genet.* 1992;29:539–541.
100. Josso N, Rey R, Picard JY. Testicular anti-müllerian hormone: clinical applications in DSD. *Semin Reprod Med*. 2012;30:364–373.
101. Looijenga LHJ, Hersmus R, De Leeuw BHCGM, et al. Gonadal tumours and DSD. *Best Pract Res Clin Endocrinol Metab*. 2010;24:291–310.
102. Chemes HE, Muzulin PM, Venara MC, et al. Early manifestations of testicular dysgenesis in children: pathological phenotypes, karyotype correlations and precursor stages of tumour development. *Apmis*. 2003;111:12–24.
103. Morel Y, Rey R, Teinturier C, et al. Aetiological diagnosis of male sex ambiguity: A collaborative study. *Eur J Pediatr*. 2002;161:49–59.
104. Sharpe RM, Skakkebaek NE. Testicular dysgenesis syndrome: mechanistic insights and potential new downstream effects. *Fertil Steril*. 2008;89:e33–e38.
105. McElreavey K, Jorgensen A, Eozenou C, et al. Pathogenic variants in the DEAH-box RNA helicase DHX37 are a frequent cause of 46,XY gonadal dysgenesis and 46,XY testicular regression syndrome. *Genet Med*. 2020;22(1):150–159.
106. Latronico A, Arnhold IP. Inactivating mutations of the human luteinizing hormone receptor in both sexes. *Semin Reprod Med*. 2012;30:382–386.
107. Porter FD. Smith-Lemli-Opitz syndrome: pathogenesis, diagnosis and management. *Eur J Hum Genet.* 2008;16(5):535–541.
108. Bianconi SE, Conley SK, Keil MF, et al. Adrenal function in Smith-Lemli-Opitz syndrome. *Am J Med Genet Part A*. 2011;155(11):2732–2738.
109. Baker BY, Lin L, Kim CJ, et al. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*. 2006;91(12):4781–4785.
110. Goto M, Hanley KP, Marcos J, et al. In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development. *J Clin Invest*. 2006;116:953–960.
111. Dean B, Chrisp GL, Quartararo M, et al. P450 Oxidoreductase deficiency: a systematic review and meta-analysis of genotypes, phenotypes, and their relationships. *J Clin Endocrinol Metab*. 2020;105(3):e42–e52.
112. Auchus RJ. The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol Metab Clin North Am.* 2001;30:101–119.

113. ten Kate-Booij MJ, Cobbaert C, Koper JW, et al. Deficiency of 17,20-lyase causing giant ovarian cysts in a girl and a female phenotype in her 46,XY sister: case report. *Hum Reprod*. 2004;19:456–459.
114. Geller DH, Auchus RJ, Mendonça BB, et al. The genetic and functional basis of isolated 17, 20-lyase deficiency. *Nat Genet*. 1997;17(2):201–205.
115. Miller WL. The syndrome of 17,20 lyase deficiency. *J Clin Endocrinol Metab*. 2012;97:59–67.
116. Lee YS, Kirk JMW, Stanhope RG, et al. Phenotypic variability in 17β-hydroxysteroid dehydrogenase-3 deficiency and diagnostic pitfalls. *Clin Endocrinol (Oxf)*. 2007;67:20–28.
117. Yang Z, Ye L, Wang W, et al. 17β-Hydroxysteroid dehydrogenase 3 deficiency: three case reports and a systematic review. *J Steroid Biochem Mol Biol*. 2017;174:141–145.
118. Thigpen AE, Silver RI, Guileyardo JM, et al. Tissue distribution and ontogeny of steroid 5α-reductase isozyme expression. *J Clin Invest*. 1993;92:903–910.
119. Maimoun L, Philibert P, Cammas B, et al. Phenotypical, biological, and molecular heterogeneity of 5α-reductase deficiency: an extensive international experience of 55 patients. *J Clin Endocrinol Metab*. 2011;96(2):296–307.
120. Mendonca BB, Inacio M, Costa EMF, et al. Male pseudohermaphroditism due to steroid 5α-reductase 2 deficiency: diagnosis, psychological evaluation, and management. *Medicine (Baltimore)*. 1996;75:64–76.
121. Cai LQ, Fratianni CM, Gautier T, et al. Dihydrotestosterone regulation of semen in male pseudohermaphrodites with 5 alpha-reductase-2 deficiency. *J Clin Endocrinol Metab*. 1994;79:409–414.
122. Katz MD, Kligman I, Cai L-Q, et al. Paternity by intrauterine insemination with sperm from a man with 5α-reductase-2 deficiency. *N Engl J Med*. 1997;336:994–998.
123. Walter KN, Kienzle FB, Frankenschmidt A, et al. Difficulties in diagnosis and treatment of 5α-reductase type 2 deficiency in a newborn with 46, XY DSD. *Horm Res Paediatr*. 2010;74(1):67–71.
124. Stuchi-Perez EG, Hackel C, Oliveira LCLEC, et al. Diagnosis of 5α-reductase type 2 deficiency: contribution of anti-Müllerian hormone evaluation. *J Pediatr Endocrinol Metab*. 2005;18:1383–1389.
125. Hughes I, Werner R, Bunch T, et al. Androgen insensitivity syndrome. *Semin Reprod Med*. 2012;30:432–442.
126. Saporita AJ, Zhang Q, Navai N, et al. Identification and characterization of a ligand-regulated nuclear export signal in androgen receptor. *J Biol Chem*. 2003;278:41998–42005.
127. Adachi M. Androgen-insensitivity syndrome as a possible coactivator disease. *Fukuoka Acta Medica*. 2001;92:258–265.
128. van de Wijngaart DJ, Dubbink HJ, van Royen ME, et al. Androgen receptor coregulators: recruitment via the coactivator binding groove. *Mol Cell Endocrinol*. 2012;352:57–69.
129. Matias PM, Donner P, Coelho R, et al. Structural evidence for ligand specificity in the binding domain of the human androgen receptor: implications for pathogenic gene mutations. *J Biol Chem*. 2000;275:26164–26171.
130. Gottlieb B, Beitel LK, Nadarajah A, et al. The androgen receptor gene mutations database: 2012 update. *Hum Mutat*. 2012;33:887–894.
131. Oakes MB, Eyvazzadeh AD, Quint E, et al. Complete androgen insensitivity syndrome-a review. *J Pediatr Adolesc Gynecol*. 2008;21(6):305–310.
132. Danilovic DLS, Correa PHS, Costa EMF, et al. Height and bone mineral density in androgen insensitivity syndrome with mutations in the androgen receptor gene. *Osteoporos Int*. 2007;18(3):369–374.
133. Hannema SE, Scott IS, Hodapp J, et al. Residual activity of mutant androgen receptors explains wolffian duct development in the complete androgen insensitivity syndrome. *J Clin Endocrinol Metab*. 2004;89(11):5815–5822.
134. Corbetta S, Muzza M, Avagliano L, et al. Gonadal structures in a fetus with complete androgen insensitivity syndrome and persistent Müllerian derivatives: Comparison with normal fetal development. *Fertil Steril*. 2011;95. 1119.e9–1119.e14.
135. Hughes IA, Deeb A. Androgen resistance. *Best Pract Res Clin Endocrinol Metab*. 2006;20:577–598.
136. Patel V, Casey RK, Gomez-Lobo V. Timing of gonadectomy in patients with complete androgen insensitivity syndrome-current recommendations and future directions. *J Pediatr Adolesc Gynecol*. 2016;29(4):320–325.
137. Wisniewski AB, Migeon CJ, Meyer-Bahlburg HFL, et al. Complete androgen insensitivity syndrome: long-term medical, surgical, and psychosexual outcome. *J Clin Endocrinol Metab*. 2000;85:2664–2669.
138. Bouvattier C, Carel JC, Lecointre C, et al. Postnatal changes of T, LH, and FSH in 46,XY infants with mutations in the AR gene. *J Clin Endocrinol Metab*. 2002;87:29–32.
139. Batista RL, Costa EMF, Rodrigues A de S, et al. Androgen insensitivity syndrome: a review. *Arch Endocrinol Metab*. 2018;62:227–235.
140. Ahmed SF, Cheng A, Dovey L, et al. Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome. *J Clin Endocrinol Metab*. 2000;85:658–665.
141. Deleted in review.
142. Picard JY, Cate RL, Racine C, et al. The persistent Müllerian duct syndrome: an update based upon a personal experience of 157 cases. *Sex Dev*. 2017;11(3):109–125.
143. Dessens AB, Cohen-Kettenis PT, Mellenbergh GJ, et al. Association of prenatal phenobarbital and phenytoin exposure with genital anomalies and menstrual disorders. *Teratology*. 2001;64(4):181–188.
144. Yiee JH, Baskin LS. Environmental factors in genitourinary development. *J Urol*. 2010;184(1):34–41.
145. Chen M-J, Macias CG, Gunn SK, et al. Intrauterine growth restriction and hypospadias: is there a connection? *Int J Pediatr Endocrinol*. 2014;2014:20.
146. Gravholt CH, Andersen NH, Conway GS, et al. Clinical practice guidelines for the care of girls and women with Turner syndrome: Proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. *Eur J Endocrinol*. 2017;177(3):G1–G70.
147. Bondy CA. Clinical practice guideline: care of girls and women with turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*. 2007;92:10–25.
148. Reynaud K, Cortvrindt R, Verlinde F, et al. Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil Steril*. 2004;81:1112–1119.
149. Ropelato MG, Escobar ME, Gottlieb S, et al. Gonadotropin secretion in prepubertal normal and agonadal children evaluated by ultrasensitive time-resolved immunofluorometric assays. *Horm Res Paediatr*. 1997;48:164–172.
150. Conte FA, Grumbach MM, Kaplan SL. A diphasic pattern of gonadotropin secretion in patients with the syndrome of gonadal dysgenesis. *J Clin Endocrinol Metab*. 1975;40:670–674.
151. Hagen CP, Aksglaede L, Sørensen K, et al. Serum levels of anti-Müllerian hormone as a marker of ovarian function in 926 healthy females from birth to adulthood and in 172 turner syndrome patients. *J Clin Endocrinol Metab*. 2010;95:5003–5010.
152. Klinefelter HF, Reifenstein EC, Albright F. Syndrome characterized by gynecomastia, aspermatogenesis without a-leydigism, and increased excretion of follicle-stimulating hormone. *J Clin Endocrinol Metab*. 1942;2:615–627.
153. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*. 1959;183:302–303.
154. Lanfranco F, Kamischke A, Zitzmann M, et al. Klinefelter's syndrome. *Lancet*. 2004;364:27–283.
155. Tartaglia N, Ayari N, Howell S, et al. 48,XXYY, 48,XXXY and 49,XXXXY syndromes: not just variants of Klinefelter syndrome. *Acta Paediatr Int J Paediatr*. 2011;100:851–860.
156. Aksglæde L, Wikström AM, De Meyts ER, et al. Natural history of seminiferous tubule degeneration in Klinefelter syndrome. *Hum Reprod Update*. 2006;12:39–48.
157. Bastida MG, Rey RA, Bergadá I, et al. Establishment of testicular endocrine function impairment during childhood and puberty in boys with Klinefelter syndrome. *Clin Endocrinol (Oxf)*. 2007;67:863–870.
158. Ratcliffe S. Long term outcome in children of sex chromosome abnormalities. *Arch Dis Child*. 1999;80:192–195.
159. Holland CM. 47,XXX in an adolescent with premature ovarian failure and autoimmune disease. *J Pediatr Adolesc Gynecol*. 2001;14:77–80.
160. Abacl A, Çatll G. Berberoıլlu M. Gonadal malignancy risk and prophylactic gonadectomy in disorders of sexual development. *J Pediatr Endocrinol Metab*. 2015;28:1019–1027.

161. Arboleda VA, Sandberg DE, Vilain E. DSDs: Genetics, underlying pathologies and psychosexual differentiation. *Nat Rev Endocrinol*. 2014;10:603–615.
162. Guercio G, Rey RA. Fertility issues in the management of patients with disorders of sex development. *Endocr Dev.* 2014;27:87–98.
163. Słowikowska-Hilczer J, Hirschberg AL, Claahsen-van der Grinten H, et al. Fertility outcome and information on fertility issues in individuals with different forms of disorders of sex development: findings from the dsd-LIFE study. *Fertil Steril*. 2017;108:822–831.
164. Sandberg DE, Callens N, Wisniewski AB. Disorders of sex development (DSD): networking and standardization considerations. *Horm Metab Res.* 2015;47:387–393.
165. Sandberg DE, Mazur T. A noncategorical approach to the psychosocial care of persons with DSD and their families. In: Kreukels B, Steensma T, de Vries A, eds. *Gender Dysphoria and Disorders of Sex Development. Boston, MA: Springer;* 2014:93–114.
166. Rolston AM, Gardner M, van Leeuwen K, et al. Disorders of sex development (DSD): clinical service delivery in the United States. *Am J Med Genet Part C Semin Med Genet.* 2017;175:268–278.
167. Adam MP, Vilain E. Emerging issues in disorders/differences of sex development (DSD). *Am J Med Genet Part C Semin Med Genet.* 2017;175:249–252.
168. Audí L, Ahmed SF, Krone N, et al. Genetics in endocrinology: approaches to molecular genetic diagnosis in the management of differences/disorders of sex development (DSD): Position paper of EU COST Action BM 1303 "DSDnet.". *Eur J Endocrinol*. 2018;179:R197–R206.
169. Lanciotti L, Cofini M, Leonardi A, et al. Up-to-date review about minipuberty and overview on hypothalamic-pituitary-gonadal axis activation in fetal and neonatal life. *Front Endocrinol (Lausanne)*. 2018;9:410.
170. Steinmetz L, Rocha MN, Longui CA, et al. Inhibin A production after gonadotropin stimulus: a new method to detect ovarian tissue in ovotesticular disorder of sex development. *Horm Res.* 2009;71(2):94–99.
171. Mendez JP, Schiavon R, Diaz-Cueto L, et al. A reliable endocrine test with human menopausal gonadotropins for diagnosis of true hermaphroditism in early infancy. *J Clin Endocrinol Metab*. 1998;83(10):3523–3526.
172. Ahmed SF, Achermann JC, Arlt W, et al. UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development. *Clin Endocrinol (Oxf)*. 2011;75(1):12–26.
173. Moshiri M, Chapman T, Fechner PY, et al. Evaluation and management of disorders of sex development: Multidisciplinary approach to a complex diagnosis. *Radiographics*. 2012;32(6):1599–1618.
174. Nakhal RS, Wood D, Creighton SM. The role of examination under anesthesia (EUA) and vaginoscopy in pediatric and adolescent gynecology: a retrospective review. *J Pediatr Adolesc Gynecol*. 2012;25(1):64–66.
175. Becker D, Wain LM, Chong YH, et al. Topical dihydrotestosterone to treat micropenis secondary to partial androgen insensitivity syndrome (PAIS) before, during, and after puberty - A case series. *J Pediatr Endocrinol Metab*. 2016;29(2):173–177.
176. Cools M, Nordenström A, Robeva R, et al. Caring for individuals with a difference of sex development (DSD): a consensus statement. *Nat Rev Endocrinol*. 2018;14:415–429.
177. Reiner WG. Gender identity and sex-of-rearing in children with disorders of sexual differentiation. *J Pediatr Endocrinol Metab*. 2005;18:549–553.

102

Endocrinology of Sexual Maturation and Puberty

Dennis M. Styne and Lindsey A. Loomba

OUTLINE

Endocrine Physiology of Sexual Maturation, 1717
  Gonadarche, 1717
  Adrenarche, 1719
Control of the Onset of Puberty, 1720
  Genetics of Pubertal Onset, 1720
  Timing of Puberty, 1721
Physical Changes of Puberty, 1722
  Boys, 1722
  Girls, 1724
Pubertal Growth and the Growth Hormone/Insulin Growth Factor Axis, 1725
Additional Changes During Sexual Maturation, 1725
  Changes in Body Composition and Bone Mineral Density, 1725
  Changes in Sleep and Brain Development, 1726
  Other Changes at Puberty, 1727
Laboratory Testing During Puberty, 1727
Summary, 1728

Pubertal development consists of a complex series of interrelated endocrine and physiologic changes that transform a sexually immature individual into a sexually mature one capable of reproduction. Puberty is initiated when gonadotropin-releasing hormone (GnRH) is released in increasing amounts from the hypothalamus, triggering an increase in the secretion of the pituitary hormones luteinizing hormone (LH) and follicle-stimulating hormone (FSH). LH and FSH in turn regulate the release of gonadal steroids, which bring about secondary sexual maturation and growth. The increase in sex steroid production by the gonad is referred to as gonadarche. Adrenarche, the increase in adrenal hormone production by the adrenal cortex, is controlled separately but typically precedes gonadarche (though pubarche often follows). A variety of other hormonal and metabolic changes also occur during the pubertal time period (Fig. 102.1), including changes in body composition, sleep, and brain development.

ENDOCRINE PHYSIOLOGY OF SEXUAL MATURATION

Gonadarche

Overview of the Hypothalamic-Pituitary-Gonadal Axis. Pulses of hypothalamic GnRH regulate the secretion of the pituitary gonadotropins LH and FSH. Gonadotropin pulses occur about every 1 to 3 hours in adults and vary with sleep and in response to sex-steroid feedback. This process does not start *de novo* at puberty but is instead based upon preexisting patterns of hormone secretion (Fig. 102.2). In the fetus, hypothalamic GnRH-secreting neurons develop the ability to release pulses of GnRH. GnRH pulse generation operates at a relatively low level for most of the prepubertal period (with the exception of a brief period of activity during infancy) due to active central nervous system (CNS) inhibition. During childhood, there is a waning of inhibition of the pulse generator and an increase in neurostimulatory factors, and GnRH pulses increase in amplitude. These GnRH pulses cause an amplification of gonadotropin pulses, which ultimately result in a rise in gonadal steroids. Pulses increase further during puberty, and the higher levels of gonadal steroids bring about the characteristic physical changes.

Fetal Development. GnRH-secreting neurons arise in the fetal olfactory placode by 5.5 weeks' gestation and migrate to the hypothalamus by 9 weeks' gestation. By 11 to 12 weeks' gestation, the fetal pituitary gland begins secreting FSH and LH, and the fetal hypothalamic-hypophyseal portal system becomes functional. By this time (the end of the first trimester), the arcuate nucleus of the fetal hypothalamus can act as a *GnRH pulse generator* and functionally transduce neural signals into pulsatile GnRH release. The GnRH pulse generator has been well-characterized by studies of pulsatile LH secretion, but mechanistic questions remain.[1] Until recently, the pattern of episodic GnRH secretion was thought to be an intrinsic characteristic of the GnRH neurons. However, GnRH pulse generation is now understood to be driven by kisspeptin production and action.[2] Kisspeptin is secreted from neurons in the anteroventral periventricular nucleus and others known as "KNDy" neurons in the arcuate nucleus that also secrete neurokinin B and dynorphin.

Gonadotropin levels are high in both sexes by midgestation but wane in late gestation, likely due to further development of negative feedback (including the development of gonadal steroid receptors in the hypothalamus and pituitary). Fetal LH and FSH levels are lower in males than in females, likely due to negative feedback from testosterone secreted by the fetal testes.[3] The prenatal rise in testosterone in males is critical for male sexual differentiation.

KEY POINTS

- Pubertal onset is not a *de novo* event, but is instead the continuation of a process that has origins in fetal life. Gonadotropin-releasing hormone secretion is triggered by kisspeptin production and action.

Infancy and Childhood. Neonatal estrogen levels decline rapidly after birth. Subsequently, the newborn hypothalamic-pituitary axis is released from estrogen inhibition, and gonadotropins increase in both amplitude and frequency. In male neonates, the LH pulses result in increased testosterone levels; in female infants, the FSH pulses lead to increased

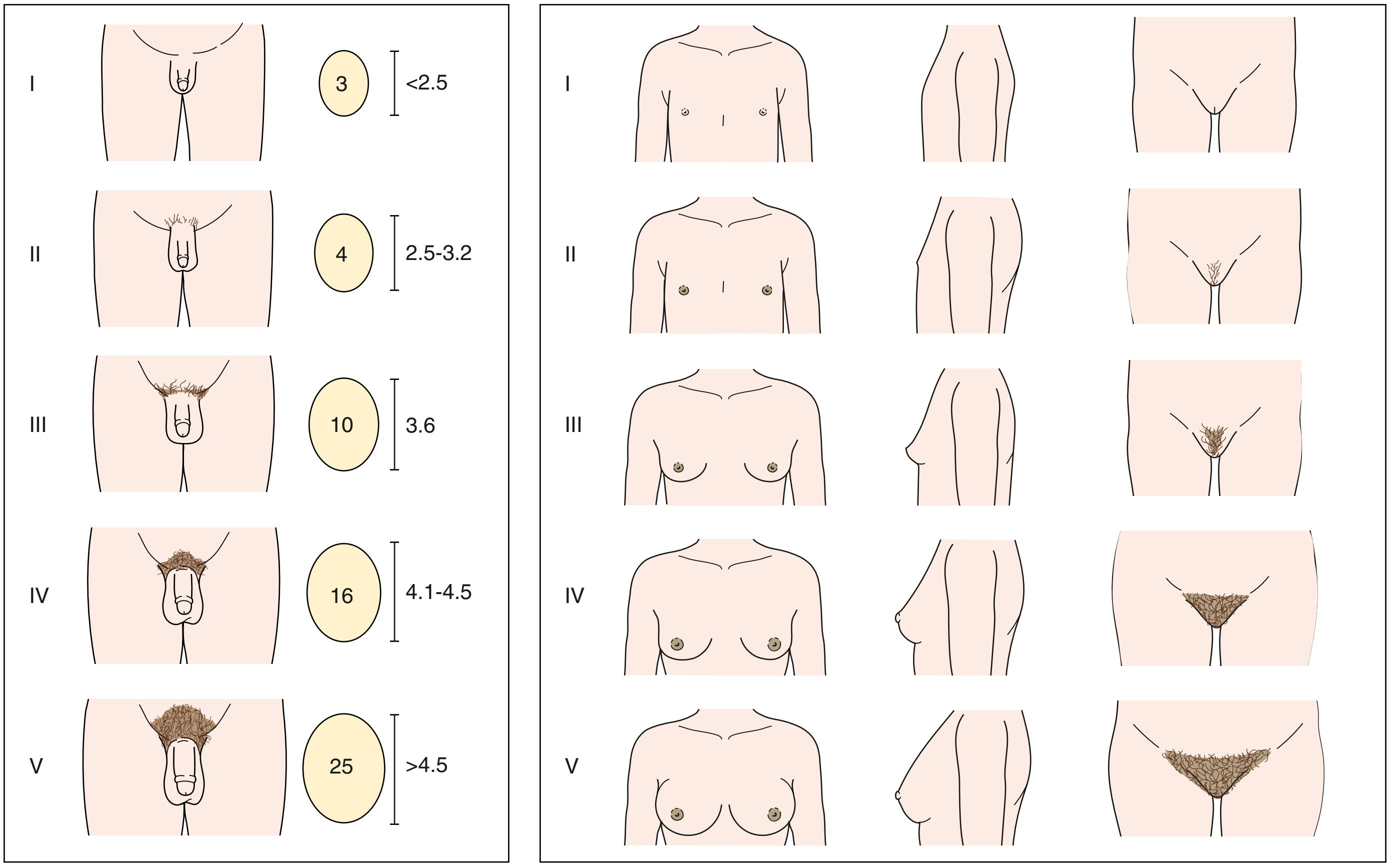


Figure 102.1 **Pubertal stages in boys and girls.** The Tanner staging process describes the pattern of genital and pubic hair development in boys and the pattern of breast and pubic hair development in girls.

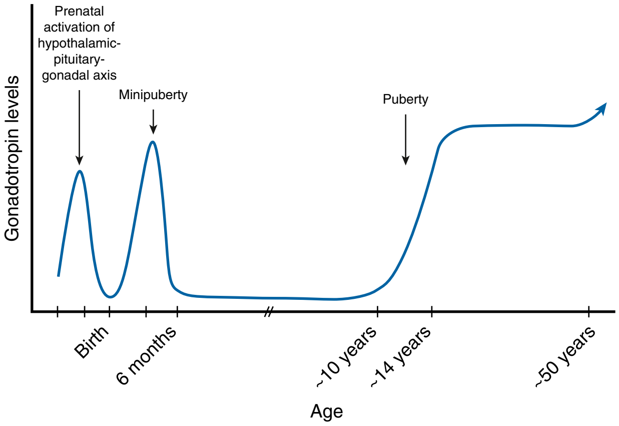


Figure 102.2 **Timeline of changes in the reproductive axis.** Gonadotropin levels are elevated prenatally but decline by late gestation. Levels rise again in infancy ("minipuberty") and then are actively suppressed until the time of pubertal onset.

estradiol. In infant boys, gonadotropins and testosterone levels peak around 1 to 3 months of age and then decline to prepubertal levels by approximately 6 months of age. In girls, gonadotropin and estradiol levels rise and remain high until approximately 6 months, with the exception of FSH, which remains elevated until age 3 to 4 years.[4] This surge of increased gonadotropin and gonadal steroid secretion during infancy is referred to as the postnatal surge, or *minipuberty*.[5]

The high gonadotropin levels of the postnatal surge are associated with rapid expansion of the Sertoli cell population, as well as increases in Leydig and germ cell numbers.[4] This may serve to further prime the LH and FSH response to GnRH later in life. Increased testosterone and dihydrotestosterone (DHT) also appear important for inducing virilization of the male external genitalia (including penile and testicular growth and descent of the testes). In females, the increase in FSH and LH is associated with maturation of the ovarian follicles. The striking sex difference in gonadotropin pulses (with infant females having much higher FSH pulses and infant males having much higher LH pulses) is thought to be due to *in utero* exposure to testicular testosterone in the male fetus.

A lengthy period of quiescence of the hypothalamic-pituitary-gonadal axis follows minipuberty. During this prepubertal juvenile pause, there is active inhibition of the hypothalamic pulse generator. Negative feedback from sex steroids and inhibin contributes to this inhibition, but an intrinsic gonadal steroid-independent CNS mechanism also plays a major role. Notably, gamma-amino butyric acid, the dominant CNS inhibitory neurotransmitter, has been shown to inhibit the hypothalamic pulse generator.[6,7] Increasingly, genetic determinants of this quiescent period have been identified. *MKRN3*, encoding makorin ring finger protein 3, is an imprinted gene that acts to inhibit GnRH secretion during the juvenile pause.[8,9]

The hypothalamic-pituitary-gonadal axis increases further in activity during the peripubertal period, before the physical changes of secondary sexual development are apparent. The mass of hormone secreted during each pulse, and thus the amplitude of each pulse, increases strikingly, suggesting that the sexual-maturation change in LH consists of a specific mass-/amplitude-dependent mechanism. Increased amplitude of LH and FSH pulses (demonstrated by ultrasensitive gonadotropin assays) occurs in most children in midchildhood.[10,11] FSH pulses are less prominent, given the reduced clearance rate of FSH as compared to LH, differences in the factors that modulate the action of GnRH on FSH and LH release by the gonadotrophs, and intrinsic differences in the secretory pattern of the two gonadotropins. LH secretion in early pubertal boys exhibits a diurnal rhythm, with augmentation of pulse amplitude following nocturnal sleep onset.[12-15] Prior to puberty, these low-amplitude and low-frequency pulses are unable to activate gonadal function. As puberty progresses, there is continued augmentation of the LH pulses. The full adult pattern, whereby there is no diurnal rhythm in LH pulse amplitude, does not occur until adulthood.[12] Thus, the increased nocturnal activity of the GnRH pulse generator appears to represent a discrete developmental stage.

A diurnal rhythm of testosterone secretion also exists in peripubertal boys. Testosterone declines gradually in the early part of the night, followed by a postmidnight increase reflecting the nocturnal onset of LH pulsatility.[16] Peak levels of serum testosterone occur in the early morning, with lowest values in the late evening. The delay of approximately 60 minutes between the LH peak and the testosterone peak is presumably due to the time required for synthesis and secretion of testosterone.[17] In prepubertal girls, a diurnal rhythm of testosterone is also seen with augmentation in the early morning. Estradiol also has a diurnal rhythm, with higher values in the early morning in late prepubertal and early pubertal girls.[11]

KEY POINTS

- There is a surge of gonadotropin and gonadal steroid secretion during infancy (referred to as minipuberty), followed by a lengthy period of quiescence of the hypothalamic-pituitary-gonadal axis (the "juvenile pause"). During later childhood, gonadotropin-releasing hormone (GnRH) pulses increase in amplitude. These GnRH pulses cause an amplification of gonadotropin pulses.

Puberty. As puberty approaches, GnRH pulses increase further in amplitude. In response, there is increased amplitude of gonadotropin pulses and increased daytime secretion of LH and FSH. The frequency of these pulses increases slightly as well, which may contribute to making the gonadotrophs more sensitive to pulses of GnRH.[12] Increasing pituitary responsiveness to GnRH can be demonstrated pharmacologically by administration of exogenous GnRH; exogenous GnRH is relatively ineffective in stimulating gonadotropin or gonadal steroid secretion prior to puberty but becomes effective after pubertal onset. If gonadotrophs are exposed to continuous rather than episodic GnRH, the GnRH receptors are downregulated, with decreased responsiveness to GnRH. This phenomenon occurs when GnRH agonists are utilized to treat central precocious puberty (CPP).

During puberty, urinary FSH rises 5-fold in both sexes. In girls, FSH rises in the early stages of puberty, whereas LH rises in the later stages. Urinary LH rises 100-fold in girls and 50-fold in boys; the rise in first morning voided urinary LH levels precedes the physical onset of puberty.[18,19] Qualitative changes in gonadotropins occur in addition to these quantitative ones; glycosylated variants of FSH and LH have distinct signaling properties and provide an additional regulatory mechanism.[20]

In boys, circulating LH activates membrane-bound LH receptors in Leydig cells to stimulate testosterone secretion. The enzyme 5-alpha reductase in certain tissues converts testosterone to DHT, which is a more potent activator of the androgen receptor. The testes additionally secrete other sex steroids, including androstenedione, delta-5-androstenediol, dihydrotestosterone, and relatively small amounts of estradiol. Testosterone rises throughout puberty, with the steepest increase in testosterone levels occurring between pubertal stage 2 and 3 (Fig. 102.1). In boys, FSH supports spermatogenesis by binding to the FSH receptor on the cellular membrane of the seminiferous tubules. Nearly all estrogen in males is converted from testicular and adrenal androgens such as androstenedione by the enzyme aromatase. In all stages of puberty (Fig. 102.1), boys have lower estrogen levels than girls. Estradiol in males is a more potent inhibitor of gonadotropin secretion than testosterone.

In girls, extraglandular conversion of ovarian and adrenal androstenedione gives rise to nearly all the circulating testosterone. FSH binds to cell-surface receptors on ovarian follicular cells to influence follicular development and stimulate estrogen secretion. The ovary produces most of the estradiol in females, although a small fraction arises from extraglandular conversion of testosterone and androstenedione by aromatase. Girls have higher estradiol concentrations than boys prior to puberty, with a rise through puberty until the pubertal growth spurt and a decrease thereafter. This results in more advanced levels of skeletal maturation in girls compared with boys, as it is estrogen that determines the acceleration of bone elongation during puberty, and eventually epiphyseal closure.

In prepubertal girls, small amounts of estrogen will suppress gonadotropin secretion. After midpuberty, estrogen can either stimulate or suppress gonadotropin release, depending on the estrogen level and pattern. Estrogen has an inhibitory effect on FSH and LH secretion early in the menstrual cycle. The onset of positive feedback (whereby higher amounts of estrogen can stimulate GnRH and gonadotropin release) occurs at midcycle once the ovarian follicle is of adequate size to produce enough estrogen and the pituitary has enough readily releasable LH to cause an LH surge (see Chapter 117). The LH surge stimulates prostaglandin, which causes ovulation via an inflammatory response in the dominant follicle. The corpus luteum forms from the remnants of the dominant follicle and produces progesterone, which prepares the endometrium for implantation and pregnancy. The increased estrogenic activity of the midcycle ovary stimulates growth of the endometrium, but the decrease in estrogen and progesterone toward the end of the menstrual cycle leaves the endometrium with no endocrine support. This, in addition to the local action of prostaglandins and constriction of arterioles, leads to endometrial necrosis and menstrual bleeding.

KEY POINTS

- Puberty is initiated when gonadotropin-releasing hormone is released in increasing amounts, triggering an increase in the secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). LH and FSH in turn regulate the release of gonadal steroids (testosterone and estradiol), which bring about secondary sexual maturation and growth.

Adrenarche

The pubertal increase in adrenal androgens (principally dehydroepiandrosterone [DHEA] and its sulfate dehydroepiandrosterone sulfate [DHEAS]) is known as *adrenarche*. The increase in these hormones leads to the physical findings of increased growth of axillary and pubic hair (known as *pubarche*) (Fig. 102.1), as well as body odor, increased sebaceous gland activity, and acne. DHEA and DHE-S are very weak agonists for the androgen receptor, and their androgenic potency is principally derived from peripheral conversion to testosterone and DHT. Clinically, DHEAS is a more useful measurement than DHEA, due to its significantly longer half-life.

During fetal development and early infancy, the adrenal gland consists of separate fetal and adult zones.[21] The fetal zone secretes large amounts of DHEAS due to minimal expression of the enzyme

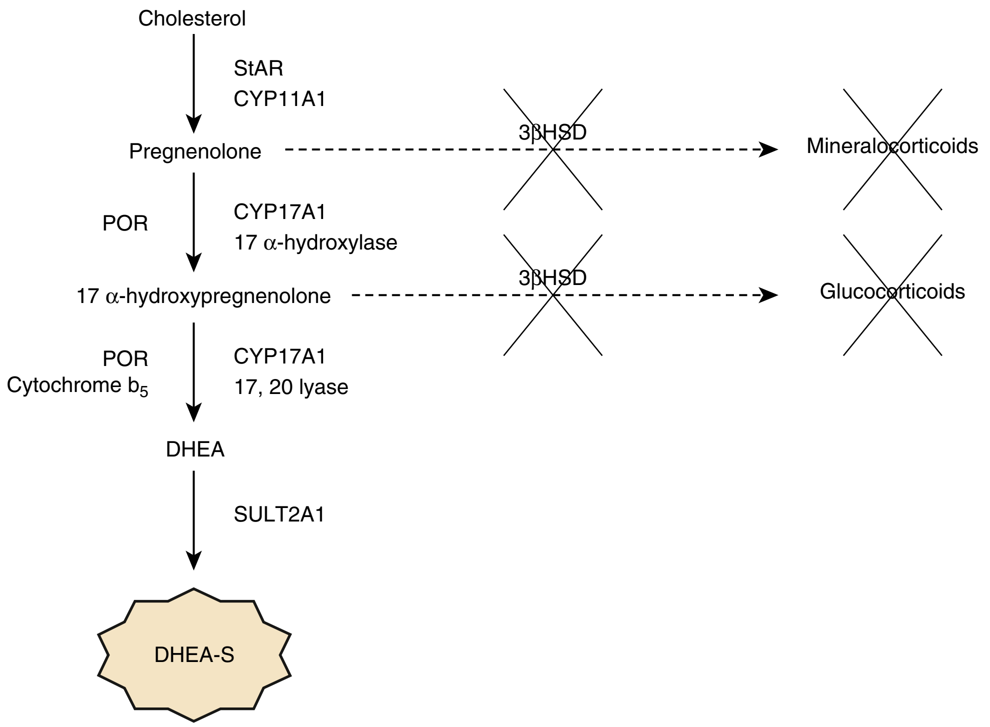


Figure 102.3 **Dehydroepiandrosterone sulfate (DHEAS) production by the zona reticularis of the adrenal glands.** DHEAS production is driven by expression of steroidogenic acute regulatory protein (StAR), cytochrome P450scc (CYP11A1), 17α -hydroxylase/17,20-lyase (CYP17A1), and DHEA sulfotransferase (SULT2A1), along with cofactors cytochrome P450 reductase (POR) and cytochrome b_5. The 17,20-lyase activity of CYP17A1 is enhanced by cytochrome b5. The lack of 3βHSD expression ensures that precursors are not diverted to the mineralocorticoid and glucocorticoid pathways.

3β-hydroxysteroid dehydrogenase, which if expressed at higher levels would direct steroidogenesis toward the mineralocorticoid and glucocorticoid pathways. Regression of the fetal zone in early infancy via apoptosis results in a fall in DHEAS levels. Thereafter, levels stay low until the onset of adrenarche. Differentiation of the adult zone and formation of a continuous zona reticularis results in an increase in DHEAS starting around 6 years of age (roughly 2 years prior to the onset of increased gonadal steroid secretion). The activities of cytochrome P450scc (CYP11A1), 17α-hydroxylase/17,20-lyase (CYP17A1), and DHEA sulfotransferase (SULT2A1), all of which are abundant in the zona reticularis, direct steroidogenesis toward the production of DHEAS (Fig. 102.3).[22] In addition, DHEAS production is increased by expression of cofactor proteins such as cytochrome b5 and cytochrome P450 oxidoreductase, which enhance the activity of CYP17A1. DHEA and DHEAS levels are higher in males than in females and continue to rise in both sexes into the second decade of life (Fig. 102.4).

While certain hormones have been found to help regulate adrenarche, the trigger for its onset remains elusive. Pituitary adrenocorticotropic hormone (ACTH) is necessary, and absence of adrenarche is noted in patients with ACTH receptor mutations and in those with ACTH deficiency.[23,24] However, the relative potency of ACTH to control adrenal androgen versus cortisol secretion is much less, by a factor of approximately 1000. Furthermore, ACTH levels do not rise at the time of adrenarche, indicating that other factors are involved.

KEY POINTS

- The pubertal increase in adrenal androgens (primarily dehydroepiandrosterone and dehydroepiandrosterone sulfate) is known as adrenarche. An increase in these hormones leads to the physical findings of axillary and pubic hair growth, body odor, increased sebaceous gland activity, and acne.

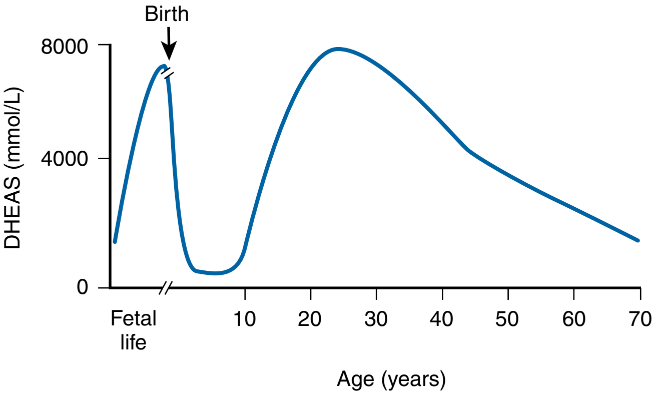


Figure 102.4 **Variation in circulating dehydroepiandrosterone sulfate (DHEAS) concentration through human life.** DHEAS concentrations reach a peak at term; following birth, there is a rapid decline. At approximately 6 years of age, DHEAS concentrations rise again (adrenarche) and reach a peak during early adulthood, declining thereafter in a process often called adrenopause. (From Auchus RJ, Rainey WE. Adrenarche: Physiology, biochemistry and human disease. *Clin Endocrinol (Oxf).* 2004; 60:288–296.)

CONTROL OF THE ONSET OF PUBERTY

Genetics of Pubertal Onset

Pubertal onset involves a complex interplay of inhibitory and stimulatory factors, and numerous genes and regulatory signals are critical to the process. To date, approximately 50 genes have been implicated in patients who fail to progress through puberty normally due to hypogonadotropic hypogonadism,[25] whereas only four monogenic causes of CPP have been identified (including mutations in *KISS1*, *KISS1R*, *MKRN3*, and *DLK1*) (also see Chapter 103).

KISS1 and KISS1R. As previously noted, kisspeptin neurons in the arcuate nucleus function as the hypothalamic pulse generator. Kisspeptin, encoded by *KISS1*, is a critical stimulatory factor for GnRH secretion and functions as a gatekeeper of pubertal onset. Inactivating or loss-of-function mutations in *KISS1* and in the gene encoding G protein–coupled kisspeptin receptor (*KISS1R*, previously *GPR54*) are associated with hypogonadotropic hypogonadism and pubertal delay.[26-28] Alternately, gain-of-function mutations in *KISS1* and *KISS1R* have been identified as causes of CPP. An autosomal dominant *KISS1R* mutation causing prolonged activation of intracellular signaling pathways in response to kisspeptin is associated with CPP.[29] A missense mutation in *KISS1* leading to prolonged kisspeptin effect has also been identified in a 1-year-old male with CPP.[30]

Makorin Ring Finger Protein 3 (MKRN3). *MKRN3* is an imprinted gene located on chromosome 15q11-q13 in the Prader–Willi syndrome region.[9] MKRN3 acts to inhibit GnRH secretion during the juvenile pause and has been described as imposing a neurobiologic break on puberty.[31] The maternal *MKRN3* allele is not expressed, due to imprinting; a paternally inherited loss-of-function mutation of *MKRN3* leads to deficiency of *MKRN3* and autosomal dominant CPP,[8,32] though the precise mechanism by which this occurs is unknown. To date, *MKRN3* mutations are the most commonly identified genetic cause of CPP. Mutations in *MKRN3* are particularly common in boys and in familial cases, though affected girls typically have an earlier age at pubertal onset (median age 6.0 vs. 8.5 years).[33]

Tachykinin Precurson 3 (TAC3) and TAC3R/Neurokinin B. Neurokinin B (NKB), encoded by *TAC3*, also appears to play an important role in puberty initiation via stimulatory effects on kisspeptin release in the arcuate nucleus.[34] Mutations in both *TAC3* and the gene encoding the NKB receptor, *TAC3R*, have been shown to result in failed GnRH release and hypogonadotropic hypogonadism.[35,36] Mutations in the NKB system are now thought to occur in more than 5% of patients with normosmic idiopathic hypogonadotropic hypogonadism.[37]

Delta-Like Noncanonical Notch Ligand 1 (DLK1). Like *MKRN3*, *DLK1* is a paternally expressed imprinted gene that appears to play a role in pubertal onset, though mutations in *DLK1* occur much less frequently than in *MKRN*3. Paternally inherited *DLK1* deletions were found through linkage analysis and whole-genome sequencing in multiple individuals with CPP within the same family.[38] *DLK1* is thought to affect GnRH secretion through the regulation of kisspeptin.

Timing of Puberty

Secular Trends and Ethnic Differences. Pubertal timing describes the pubertal status of a child in relation to same-sex and same-age peers. It is determined largely by genetic factors and is modulated by both intrinsic and environmental factors (including the intrauterine environment, nutritional status, and metabolic stress). Direct genetic effects account for 50% to 80% of the variation of timing of puberty in healthy populations.[39-42] Recently, genome-wide association studies have been used to identify genetic loci that affect pubertal timing, including a 2017 analysis of more than 360,000 women.[43]

Sexual maturation in girls and boys has occurred at younger ages throughout the past 100 years. Between 1850 and 1950, the mean age of menarche decreased by 3 to 4 months each decade in industrialized nations.[44] Most investigators believe that improved nutritional status and improving health conditions worldwide account for most of these changes. An expert panel review of data collected between 1940 and 1994 also reported a decrease in the age of menarche by several months over the past 25 years.[45] These more recent pubertal trends are likely due to the increased prevalence of childhood obesity in this country as well as environmental influences[45] (see later).

Changes in the age of onset of thelarche have been more difficult to characterize, as methods have varied across studies, and some analyses have been limited by small sample sizes. However, a large cross-sectional study of 17,000 girls suggested that thelarche has advanced in the second half of the twentieth century.[46] These data are supported by a relatively recent Danish study, which showed significantly earlier onset of thelarche in a 2006 cohort of girls as compared with a 1991 cohort.[47] Progression of breast development after onset through later stages is not clearly accelerating, however.[45]

There is considerable variability in the age of pubertal events among ethnic groups. Black girls often have pubic hair development prior to the onset of breast development, while the opposite is typically true for White girls.[46] In the United States, the ages of thelarche, menarche, and onset of pubic hair development are all significantly younger in Black girls as compared with non-Hispanic White and Mexican American girls.[48,49] In a relatively recent large longitudinal study, mean ages of breast stage 2 development were 8.8 years for Black girls, 9.2 years for Hispanic girls, 9.6 years for non-Hispanic White girls, and 9.9 years for Asian girls.[50] Data from the longitudinal Bogalusa Heart Study has shown advancement in the age of menarche in White and Black American girls from 1974 to 1994, with the timing advancing more in Black American girls as opposed to White girls.

In general, the timing of thelarche varies more than the timing of menarche. There appears to be a physiologic compensation for puberty that occurs in the early-normal age range, such that girls with thelarche at age 9 years have an average of 2.7 years to menarche, whereas girls with thelarche at age 13 years have an average of 0.7 years to menarche.[51]

In boys, data quality issues make drawing conclusions about secular trends even more difficult. Though some studies have suggested earlier genital development in more recent decades, data are insufficient to conclude this with certainty.[45] As in girls, there are significant ethnic differences with respect to ages of attainment of various pubertal stages.[49,52]

KEY POINTS

- Changes in the age of onset of puberty over recent decades are difficult to characterize, as methods have varied widely across studies. However, evidence points toward younger ages of attainment of both thelarche and menarche in girls. There is variability in the age of pubertal events among ethnic groups.

Nutrition. The pronounced physical changes of puberty require significant energy input, and thus energy status unsurprisingly has a critical regulatory effect on pubertal timing (Fig. 102.5). The metabolic hormone leptin is secreted by adipocytes in concentrations that are highly associated with fat mass, acting as an essential permissive signal that ensures the availability of adequate energy stores prior to entrance into puberty.[53] Appropriate leptin concentrations are required for maturation of the hypothalamic-pituitary-gonadal axis, pubertal progress and gonadal function.[54] Once bound to its receptor, leptin helps promote a feeling of satiety. Leptin deficiency, due to homozygous mutations in the leptin (*LEP*) or leptin receptor (*LEPR*) genes, results in hypogonadotropic hypogonadism and early-onset obesity due to lack of

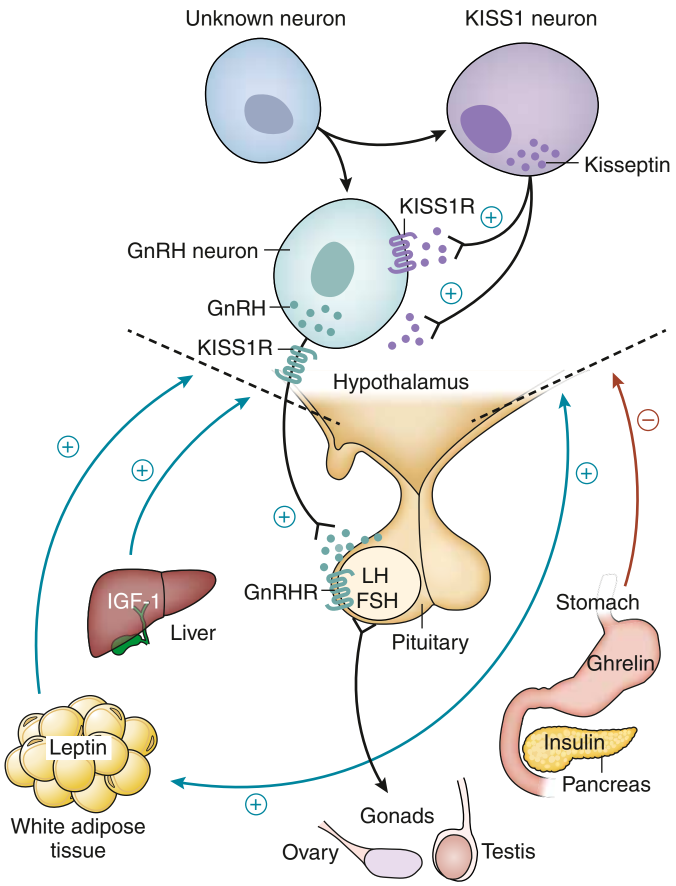


Figure 102.5 **Tentative model for the metabolic control of GnRH secretion by peripheral metabolic hormones.** The peripheral hormones that transmit metabolic signals to GnRH neurons are shown. Leptin, a signal for energy abundance, is a permissive factor for GnRH secretion. Ghrelin, a putative signal of energy insufficiency, is an inhibitory factor for GnRH secretion. Insulin and IGF1 have stimulatory effects on the reproductive axis. Dashed line represents proposed pathway. *FSH*, Follicle-stimulating hormone; *GnRH*, gonadotropin-releasing hormone; *GnRHR*, gonadotropin-releasing hormone receptor; *IGF1*, insulin-like growth factor 1; *KISS1*, kisspeptin; *KISS1R*, kisspeptin receptor; *LH*, luteinizing hormone; *+*, stimulatory effect; *–*, inhibitory effect. (Reprinted with permission from Elsevier. From Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol.* 2016 Mar;4(3):254–264.)

satiety.[55-57] In leptin-deficient patients, there is initiation or resumption of gonadotropin secretion after leptin replacement. Recombinant leptin also restores gonadotropin secretion in women with hypothalamic amenorrhea associated with anorexia nervosa,[58] although is not commonly used, due to limited supply of recombinant leptin.

Ghrelin, the ligand of the growth hormone (GH) secretagogue receptor 1a or ghrelin receptor, is secreted in response to food deprivation and plays a role in signaling negative energy balance.[59] Release of ghrelin from the enteroendocrine cells of the gastrointestinal tract results in the feeling of hunger that results in food intake. Ghrelin decreases GnRH secretion and has an overall inhibitory effect on the gonadotropic axis,[60] consistent with its role as an indicator of insufficient energy. Insulin also appears to modulate GnRH secretion and likely has additional effects on pubertal onset via stimulation of leptin secretion.[61,62]

Recent work has shed light on the process by which metabolic signals (including mammalian target of raptomycin [mTOR], AMP-activated protein kinase [AMPK], and SIRT1) modulate pubertal timing via effects on kisspeptin production in kisspeptin neurons (Fig. 102.6). mTOR is activated in conditions of energy excess and helps mediate the permissive effects of leptin on pubertal onset.[63] AMPK, a major cellular energy sensor, is inhibited by leptin and has a restrictive effect on *KISS1* expression in conditions of negative energy balance.[64] *SIRT1* is a nuclear protein expressed in kisspeptin neurons and restrains female puberty by decreasing *KISS1* promoter activity.[65] As puberty approaches, *SIRT1* is expelled from the KISS1 promoter, which helps permit *KISS1* expression. Overnutrition accelerates these changes, enhancing the expression of *KISS1* and advancing puberty.[65] Undernutrition, in contrast, increases levels of *SIRT1* and delays puberty.

Obese girls have a clear trend toward earlier puberty, and it appears that boys likely do as well.[66,67] Lee et al. followed young girls longitudinally and found that higher body mass index (BMI) z-scores in early childhood and higher rates of BMI change between early and midchildhood are associated with earlier thelarche, suggesting that childhood obesity in the United States may decrease the age of onset of puberty for US girls.[68] Conversely, low body fat, as occurs in anorexia and some avid adolescent female athletes, has long been thought to postpone menarche.[69] It is postulated that local production of estrogen by adipose tissue may account for earlier breast development in obese girls, but this has not been consistently shown.[70]

KEY POINTS

- The pronounced physical changes of puberty require significant energy input, and thus energy status has a critical regulatory effect on pubertal timing. Metabolic signals modulate pubertal timing via effects on kisspeptin production.

Endocrine-Disrupting Substances. Though genetic factors play the greatest role in the determination of pubertal timing, endocrine-disrupting compounds (EDCs) are potential contributors to observed shifts. EDCs are exogenous compounds that disrupt endogenous endocrine activity via effects on hormone synthesis, metabolism, or elimination (see Chapter 137). Although EDC exposure has been typically thought to lead to earlier pubertal development, several studies suggest later pubertal development (or earlier onset but overall longer time to maturation).[71] As one example, phenols, including bisphenol A, and phthalates have gained attention in recent years for their estrogenic and antiandrogenic properties, respectively.[72] In a large study of 1051 girls, higher 2,5-dichlorophenol was associated with earlier menarche by 7 months, and mono-3-carboxypropyl phthalate was associated with later thelarche and menarche.[73] Other smaller studies have also shown possible effects of phthalates and phenols on pubertal timing.[74-76] Additional compounds, including flame retardants/polybrominated diphenyl ethers, polychlorinated biphenyls, and dioxins, have all been associated with alterations in pubertal timing. Effects appear to depend upon the timing, duration, and degree of exposure, as well as on the specific compound and gender of the child.[71]

PHYSICAL CHANGES OF PUBERTY

Physical development of puberty can be denoted by sexual maturation ratings, commonly called Tanner stages. These standards allow for objective descriptions of pubertal progress.

Boys

Table 102.1 describes the sequence of pubertal genital changes in boys. The first sign of central puberty in boys is typically an increase in the size of the testes to over 2.5 cm in the longest diameter (excluding

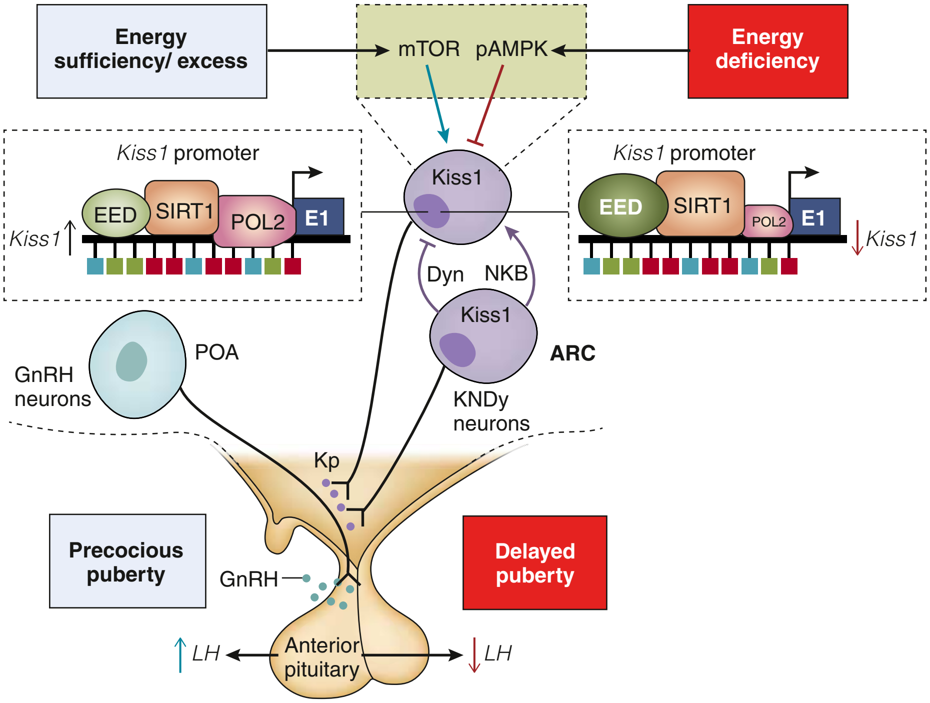


Figure 102.6 **Metabolic regulation of gonadotropin secretion.** The mTOR pathway is activated in conditions of energy excess. This leads to eviction of SIRT1 and EED from the *KISS1* promoter, which enhances Kiss1 expression and stimulates gonadotropin secretion. Alternately, the AMPK pathway is activated in energy-deficient conditions. This is associated with the continued presence of SIRT1 and EED, leading to suppression of Kiss1 expression and reduced gonadotropin secretion. *ARC*, Arcuate nucleus; *NKB*, neurokinin B; *Dyn*, dynorphin; *KNDy*, Kiss1, NKB, and Dyn; *EED*, embryonic ectoderm development; *mTOR*, mammalian target of raptomycin; *AMPK*, AMP-activated protein kinase. (From Vazquez MJ, Velasco I, Tena-Sempere M. Novel mechanisms for the metabolic control of puberty: implications for pubertal alterations in early-onset obesity and malnutrition. *J Endocrinol.* 2019;242:R51–R65.)

TABLE 102.1 **Stages of Genital Development in Boys**

Pubertal Stage	Description	Ethnicity	Median Age at Entry (Years)
1	No change	n/a	
2	No change	Non-Hispanic White	11.98
		Non-Hispanic Black	11.16
		Mexican American	12.30
3	No change	Non-Hispanic White	12.65
		Non-Hispanic Black	12.51
		Mexican American	13.06
4	No change	Non-Hispanic White	13.56
		Non-Hispanic Black	13.73
		Mexican American	14.08
5	No change	Non-Hispanic White	15.67
		Non-Hispanic Black	15.32
		Mexican American	15.75

Source: Sun SS, Schubert CM, Chumlea WC, Roche AF, Kulin HE, Lee PA, Himes JH, Ryan AS. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics.* 2002;110(5):911–919.

the epdidymus), equivalent to a testicular volume of 4 mL or greater. Increasing testicular size early in puberty occurs due to an increase in Sertoli cell number and seminiferous tubule length (stimulated by FSH). Rising testosterone levels result in enlargement of the penis, and scrotal thinning occurs. Pubic hair development occurs as a result of adrenal and testicular androgen secretion and again is classified separately from genital development (Table 102.2). The appearance of spermatozoa in early morning urinary specimens (spermarche) occurs at a mean chronologic age of 13.4 years; this usually occurs at gonadal stages 3 to 4 and pubic hair stages 2 to 4.

Increasing testosterone concentrations (or the metabolic products of testosterone) additionally produce changes in nonreproductive organs. The larynx increases in size and the voice deepens, red blood cell mass increases, high-density lipoprotein levels decline, and body

TABLE 102.2 Stages of Pubic Hair Development in Boys

Pubertal Stage	Description	Median Age (Years)	Age at 2.3 Percentile (Years)
1	No pubic hair.		
2	Sparse growth of long, slightly pigmented, downy hair (straight or slightly curled), appears chiefly at the base of the penis.	11.3	8.3
3	Hair is considerably darker, coarser, and mostly curled. Hair spreads sparsely over the pubic junction.	12.0	10.0
4	Hair is adult in type and distributed as an inverse triangle without spread to the medial surface of the thighs.	13.2	11.1
5	Hair is adult in quantity and type and distributed as an inverse triangle with spread to the medial surface of the thigh.	15.0	12.2

A description of each stage is shown, along with median age at each pubertal stage and age at the 2.3 percentile (approximates –2 standard deviations).
Data from the National Health and Nutrition Examination Survey III.

TABLE 102.3 Stages of Breast Development in Girls

Pubertal Stage	Description	Ethnicity	Median Age at Entry (Years)
1	Preadolescent; only the papilla is elevated.	n/a	
2	Breast bud stage; the breast and papilla are elevated as a small mound. Areolar diameter is enlarged.	Non-Hispanic White	10.38
		Non-Hispanic Black	9.48
		Mexican American	9.80
3	Further enlargement of breast and areola has occurred, with no separation of their contours.	Non-Hispanic White	11.75
		Non-Hispanic Black	10.79
		Mexican American	11.43
4	The areola and papilla project to form a secondary mound above the level of the breast.	Non-Hispanic White	13.29
		Non-Hispanic Black	12.24
		Mexican American	13.07
5	Mature stage; only the papilla projects because of recession of the areola to the general contour of the breast.	Non-Hispanic White	15.47
		Non-Hispanic Black	13.92
		Mexican American	14.70

A description of each stage is shown, along with median age of attainment of each stage.
(Ethnicity data from Sun SS, Schubert CM, Chumlea WC, et al. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics*. 2002;110:911–919.)

hair (including facial hair) increases. Through the ensuing years, hair recedes on the scalp. Testosterone or estradiol produced from testosterone in the nervous system stimulates sexual libido, and behavior changes. Maturation takes many years; for example, although adult testosterone concentrations in blood are reached by age 15 or so, growth of the adult male beard is not usually complete until 20 to 25 years of age.

KEY POINTS

- The first sign of central puberty in a boy is an increase in size of the testes. Increasing testosterone levels result in changes in the genitalia, as well as in nonreproductive organs.

Girls

The first sign of puberty in the female is an increase in linear growth velocity, although breast development is usually detected first. Breast growth is stimulated by ovarian estrogen secretion and occurs in a predictable sequence, with eventual size and shape influenced by genetic and nutritional factors (Table 102.3). Likewise, areolar changes in size, color, and erectility occur in a predictable order. Increasing estrogen levels additionally lead to enlargement of the labia minora and majora. The prepubertal vaginal mucosa is reddish (though not bright red as might be found with an infection) and progresses to a duller pinkish color with clear or slightly whitish discharge due to estrogen stimulation. The ovaries enlarge, the uterus enlarges, and the fundus becomes prominent.[77] Sonographic standards are available; uterine volume greater than 1.8 mL and ovarian volume greater than 1.2 mL occur in CPP or normal puberty. Pubic hair development in girls is determined chiefly by adrenal and ovarian androgen secretion (Table 102.4). The relative timing of growth and pubertal changes in boys and girls is summarized in Fig. 102.7.

Menarche occurs in midpuberty, approximately 2.5 years after the onset of breast development. In a recent longitudinal study of 946 girls, the median age of menarche was 12.25 years.[78] Compared with Black participants (median age of menarche 12.0 years), Hispanic girls were more likely to have menarche earlier (11.83 years), whereas White and Asian girls were more likely to have menarche later (12.67 and 12.75 years, respectively).

KEY POINTS

- The first sign of puberty in the female is an increase in linear growth velocity, although breast development is usually detected first. Menarche occurs at an average age of 12.25 years.

TABLE 102.4 Stages of Pubic Hair Development in Girls.

Pubertal Stage	Description	Ethnicity	Median Age at Entry (Years)
1	No change	n/a	
2	No change	Non-Hispanic White	10.57
		Non-Hispanic Black	9.43
		Mexican American	10.39
3	No change	Non-Hispanic White	11.80
		Non-Hispanic Black	10.57
		Mexican American	11.70
4	No change	Non-Hispanic White	13.00
		Non-Hispanic Black	11.90
		Mexican American	13.19
5	No change	Non-Hispanic White	16.33
		Non-Hispanic Black	14.70
		Mexican American	16.30

Source: Sun SS, Schubert CM, Chumlea WC, Roche AF, Kulin HE, Lee PA, Himes JH, Ryan AS. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics*. 2002;110(5):911–919.

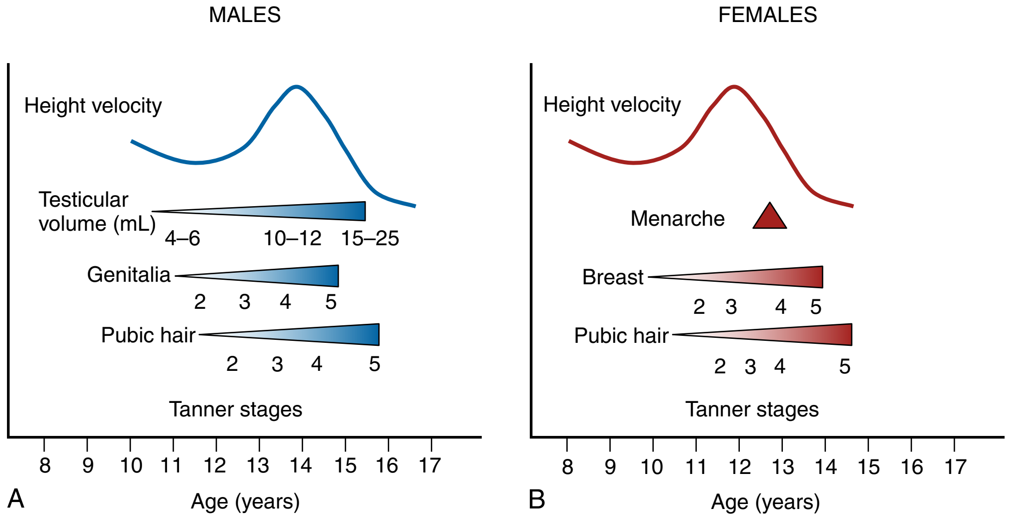


Figure 102.7 **Relative timing of pubertal events in males and females.** (From Dattani MT, Brook CGD. Adolescent medicine. In: Braunwald E, Fauci AS, Kasper DL, et al., eds. *Harrison's Principles of Internal Medicine*. 15th ed. New York: McGraw-Hill, 2001.)

PUBERTAL GROWTH AND THE GROWTH HORMONE/INSULIN GROWTH FACTOR AXIS

There is a close interplay between the gonadal steroids and GH during pubertal development. During puberty, increasing gonadal steroid concentrations stimulate the GH/insulin-like growth factor (IGF) axis and enhanced secretion of GH in both the basal and stimulated state occurs. Likewise, IGF1 acts synergistically with the gonadotropins to stimulate gonadal steroid synthesis. Thus, increasing levels of both GH and the gonadal steroids result in coordinated anabolic effects on growth, and it appears that GH modulates the timing of sexual maturation.[79] As such, children with GH deficiency have delayed pubertal onset and a lower overall rate of pubertal development, though they do not entirely fail to undergo pubertal changes.

Mean growth velocity is approximately 6 cm/year during prepubertal childhood. As a result of increases in GH/IGF1 and the rising gonadal steroid levels, growth velocity increases in girls at 11 years of age, peaking on average at age 12 years (Tanner stage 2–3). In boys, growth velocity increases later, beginning at approximately 13 years and peaking at approximately 14 years of age (Tanner stage 3–4) (Fig. 102.8). The mean 24-hour concentration of GH is approximately 5 to 7 µg/L in prepuberty and early puberty and increases strikingly during middle to late puberty to approximately 13 to 15 µg/L. The rise in the mean GH concentration is caused primarily by increases in pulse amplitude and an increase in the mass of the GH pulse secretory burst rather than by changes in pulse frequency.[80,81]

KEY POINTS

- Increasing levels of growth hormone and the gonadal steroids together result in coordinated anabolic effects on growth during puberty.

ADDITIONAL CHANGES DURING SEXUAL MATURATION

Changes in Body Composition and Bone Mineral Density

Gender differences in body composition before puberty are modest compared with postpubertal differences. Throughout infancy, females have a slightly greater percent body fat than males.[82,83] In early childhood, males and females have comparable absolute fat mass and fat free mass. By later childhood, girls begin to gain fat more readily, and by

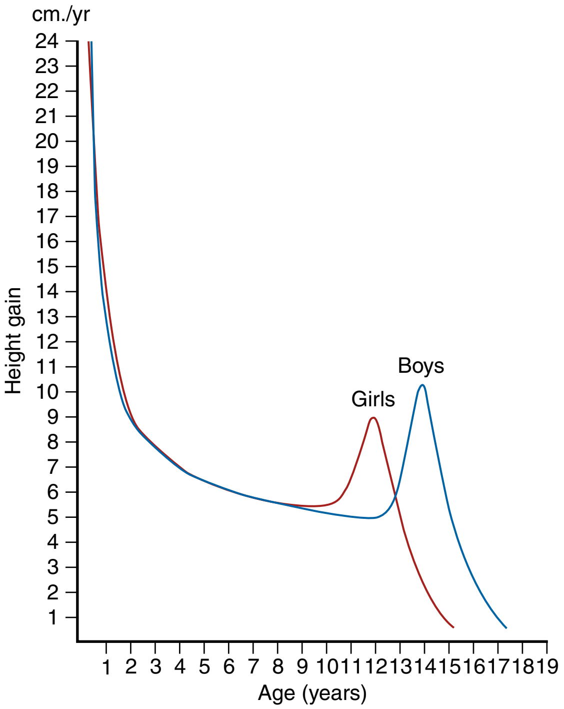


Figure 102.8 **Typical individual curves for height velocity in boys and girls.** Peak height velocity during puberty occurs on average at age 12 years in girls and age 14 years in boys. (From Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. *Arch Dis Child.* 1966; 41:454–471.)

10 years of age, girls have approximately 2 kg more fat mass than boys. Meanwhile, boys gain slightly more fat free mass.[83-85] During puberty, girls gain significantly more fat mass than males and by the end of puberty, have approximately 5 to 6 kg more absolute fat mass. Males do not significantly change their absolute fat mass during puberty. By approximately age 12 years, fat free mass gain in females begins to plateau, and males begin gaining fat free mass at an increased rate.[84] By adulthood, males have 20 kg greater fat free mass than females on average. The net effect of these changes is a decreasing percent body fat throughout puberty in males to an adult value of approximately 13%. In females, the net effect of a plateau in fat free mass and increasing fat mass results in a higher percent body fat of approximately 25%.

In female adolescents, the hips enlarge and the waist-to-hip ratio decreases with advancing pubertal stage.[86] Females generally have gynecoid body fat distribution, with fat centered around the hips and thighs, whereas males have an android body fat distribution, with fat distributed mainly in the abdominal area. In males, the android shape develops primarily during puberty. Females remain gynecoid (though become more so) during puberty, but proportions change toward a more android shape in later life, after menopause, or with increasing parity or with the hyperandrogenic ovulatory disorder polycystic ovarian syndrome (see Chapter 124).

The most important phases of bone accretion occur during infancy and during puberty.[87] Girls reach peak calcium accretion around 12.5 years of age, while boys reach this stage at approximately 14 years. Approximately one-fourth of total skeletal calcium is laid down during the 2 peak pubertal years of bone growth.[88] A significant reduction in bone mass accrual occurs in females after 15 years of age (2–4 years after menarche), while bone mass in males continues to increase between 15 and 18 years of age.[89] In girls, later age at menarche is associated with low bone mineral density and higher fracture risk, likely due to lower peak bone mass.[90-92]

Sex steroids are critical for bone maturation, and hypogonadism in both sexes leads to osteoporosis. Males with aromatase deficiency or estrogen receptor defects have tall stature, unfused epiphyseal cartilages, and reduced bone mineral density/osteoporosis. These men have helped identify estradiol as critical for attainment of normal bone mineral density and epiphyseal fusion.[93] Although estrogen plays the critical role in stimulating new bone formation, augmenting mineralization, promoting epiphyseal maturation, and establishing normal skeletal proportions, androgens also contribute. In addition, GH and IGF1 are required for attainment of peak bone mass.[84,94] Ethnicity and other genetic effects may account for up to 11% of the variation in bone mineral density.[95]

Changes in Sleep and Brain Development

Changes in sleep regulation occur during sexual maturation, and these changes may actually precede the physical changes of puberty.[96] The body's circadian rhythm is driven by a circadian pacemaker in the hypothalamic suprachiasmatic nucleus (SCN), which is genetically regulated but directly influenced by environmental cues such as light. Changes to circadian rhythm during adolescence result in later sleep onset, later wake times (when children are unrestrained by school schedules), and less overall sleep. At nighttime, melatonin production in the SCN promotes sleep onset; this secretion pattern becomes later as children progress though the stages of puberty.[97] Sleep electroencephalogram patterns change during puberty and may account for the marked daytime sleepiness of adolescence, regardless of overall sleep time. Notably, the large, slow ("delta") waves of non-rapid eye movement (NREM) sleep decline significantly during adolescence, though the proportion of NREM to rapid eye movement sleep remains the same.[98] A greater tolerance for sleep deprivation additionally occurs during the maturation process.

There are considerable differences in the cognitive abilities of children versus adults, and much of this change occurs during adolescence. The prefrontal cortex, an area of the brain concerned with forward planning and regulatory control of emotional behavior, continues to develop until the age of 20 to 25 years. Magnetic resonance imaging studies of the adolescent brain demonstrate specific structural changes during puberty. There is a linear increase in white matter due to increased myelination between ages 4 and 22 years, with males showing steeper increases in volume as compared with females.[99-102] Alternately, gray matter volumes change in a region-dependent nonlinear (inverted U-shaped) manner, with peak gray matter volume occurring 1 to 2 years earlier in females than in males.[99,102] A reduction in cortical synaptic density and neuronal density occurs secondary to synaptic pruning, which is more prominent in males compared with females.[101] Gender differences in brain structure likely reflect gonadal steroid influences on brain growth and remodeling.

Brain plasticity decreases during puberty, as evidenced by greater difficulties with foreign language acquisition in adults compared with prepubertal children and improved traumatic brain injury outcomes in young children compared with older teens or adults. Mental illnesses such as mania, depression, obsessive-compulsive disorder, and schizophrenia are all more common after puberty. This increase is postulated to be related to pubertal changes in brain architecture and function.

KEY POINTS

- Fat and fat free mass changes during puberty result in adult males having a much lower percent body fat than adult females (13% vs. 25%). A critical phase of bone accretion occurs during puberty, with peak calcium accretion around 12.5 years of age in girls and 14 years in boys. Changes in sleep regulation also occur, and these changes may precede the physical changes of puberty.

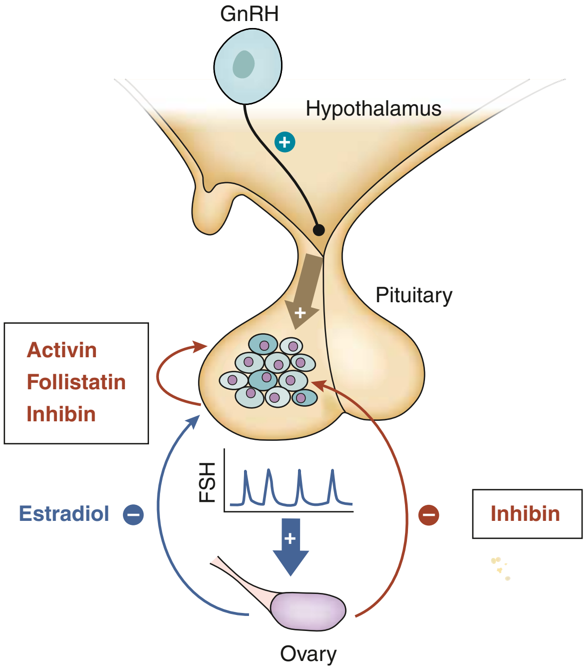


Figure 102.9 Regulation of follicle-stimulating hormone (FSH) secretion by inhibin, activin, and follistatin in females. Gonadotropin-releasing hormone pulses are transported to the anterior pituitary by portal vasculature to stimulate FSH synthesis and secretion by gonadotrope cells. At the anterior pituitary level, a local loop involving activin, inhibin, and follistatin regulates FSH secretion in an autocrine/paracrine fashion. At the ovarian level, estradiol and inhibin are two key negative feedback regulators of FSH secretion. While the ovary also produces activin and follistatin, these hormones are not believed to play an endocrine role in controlling FSH secretion. (From Padmanabhan V, Cardoso RC. Neuroendocrine, autocrine, and paracrine control of follicle-stimulating hormone secretion. *Mol Cell Endocrinol.* 2020;500:110632.)

Other Changes at Puberty

Follicle-Stimulating Hormone Regulatory Proteins: Inhibin, Activin, and Follistatin. Inhibin, produced by multiple tissues, including the ovarian granulosa cells and testicular Sertoli cells, acts to suppress FSH release from the pituitary (Fig. 102.9). Inhibin B rises twice during male development, reflecting the two periods of Sertoli cell proliferation in infancy and in early puberty. In the early stages of male puberty, inhibin B values have a close positive relationship to LH and testosterone levels and increase with pubertal progression. In genital stage 3 (characterized by continued testicular enlargement), inhibin B values peak and instead become inversely related to FSH levels[103,104]; this relationship with FSH persists into adulthood. During childhood, inhibin B is a useful diagnostic marker for the presence and function of Sertoli cells, and thus testicular tissue, in males. In girls, serum levels of inhibin A and B increase in early puberty, and there is a strong positive correlation between inhibin B and FSH and estradiol early in puberty. After pubertal stage 2, these relationships diminish.[105] Inhibin B is predominant in the follicular phase, and inhibin A is predominant during the luteal phase.

Activin is a closely related protein to inhibin, which acts in a biologically opposing manner (Fig. 102.9). Activin stimulates FSH synthesis and secretion from the pituitary gland and increases FSH binding in the ovarian follicle. In males, activin enhances spermatogenesis. Activin is produced in multiple sites, including the gonads, pituitary gland, and placenta. Follistatin, like inhibin, also inhibits FSH expression and therefore is an additional regulator of FSH biosynthesis and secretion (Fig. 102.9).

Antimüllerian Hormone. Antimüllerian hormone (AMH), a glycoprotein produced by the Sertoli cells, acts ipsilaterally via localized effects to cause regression of the Müllerian ducts in boys during early fetal development. AMH values rise from birth to relatively high levels during the first year of life in males. Levels remain stable in early childhood (with a modest late prepubertal peak) and then decline during puberty.[106] Newborn females have low or nondetectable serum levels of AMH, which rise only slightly thereafter and then rise to adult levels. As with inhibin B, AMH can be used as a laboratory marker of Sertoli cell presence and function in young males. AMH is a marker of ovarian reserve in adult women (see Chapter 123).

Prolactin. Prolactin is a peptide hormone released from the anterior pituitary gland that plays a critical role in stimulation of the mammary glands and lactation. The serum concentration of prolactin increases in normal female puberty but is not affected by male puberty. This sex difference is likely a consequence of the higher estradiol levels in females. It is an important component of breast development during and after puberty (see Chapter 127).

Sex Hormone-Binding Globulin. Between 97% and 99% of circulating testosterone and estradiol is reversibly bound to carrier proteins, including sex hormone-binding globulin (SHBG). These carrier proteins regulate the transport and biological activity of the sex steroids. Prepubertal levels of SHBG are approximately equal in boys and girls. At puberty, levels decline in both sexes, but the decline is greater in males,[107] likely due to suppression of SHBG by androgens in males.[108] The rise in adrenal androgen levels at adrenarche may explain the early drop in SHBG levels, which allows more circulating free hormone at a given concentration of testosterone. Estrogen has the opposing effect and leads to higher SHBG levels. DHEA and DHEAS bind primarily to albumin and only minimally to SHBG.

LABORATORY TESTING DURING PUBERTY

Standard assays for gonadotropins and gonadal steroids are calibrated to indicate if LH and FSH levels are very high (as found in postmenopausal women) and whether testosterone is very low (as in men with gonadal failure) or very high (as in a normal adult male). Specialized pediatric assays are needed to accurately determine the low values of sex steroids and gonadotropins found in children during pubertal development.[109] Testosterone in women and children should be measured by high-performance liquid chromatography (HPLC) tandem mass spectroscopy, which is typically performed by national labs due to high equipment cost. Free testosterone (freeT) (not bound to SHBG) may be measured as freeT, with the best results obtained by the dialysis method, and can also be estimated by a bioavailable testosterone assay. Both assays are useful in the evaluation of virilizing disorders in girls and women. Estradiol is best measured in the HPLC tandem mass spectroscopy system.

LH and FSH should be measured via highly sensitive immunometric assay and interpreted with pediatric standards, again typically requiring a national send-out laboratory. Standard LH and FSH assays are far too insensitive to accurately determine childhood values and provide meaningful information for the evaluation of most pubertal disorders. Impressively, these pediatric ultrasensitive gonadotropin and sex steroid assays can demonstrate increasing values before pubertal development is detected on physical examination. Samples are best obtained in the early morning due to the diurnal overnight rise in gonadotropins in early puberty.

KEY POINTS

- High-sensitivity pediatric assays for gonadal steroids and gonadotropins are critical to accurately characterize the low hormone levels found in children.

SUMMARY

Puberty consists of a series of complex hormonal processes that are tightly controlled by an interplay of intrinsic and extrinsic factors. Pubertal onset is not a *de novo* event, but is instead the continuation of a process that has origins in fetal life. Pubertal development includes physical changes in the breasts and genitalia, as well as changes in growth, body composition, sleep, and brain development. Obesity and other environmental factors appear to affect pubertal development, although the extent of these contributions requires further investigation.

REFERENCES

1. Herbison AE. The gonadotropin-releasing hormone pulse generator. *Endocrinology*. 2018;159:3723–3736.
2. Plant TM. The neurobiological mechanism underlying hypothalamic GnRH pulse generation: the role of kisspeptin neurons in the arcuate nucleus. *F1000Res*. 2019;8.
3. Grumbach MM, Roth JC, Kaplan SL, et al. *Hypothalamic-pituitary Regulation of Puberty in Man: Evidence and Concepts Derived from Clinical Research.* New York: John Wiley & Sons.
4. Bizzarri C, Cappa M. Ontogeny of hypothalamus-pituitary gonadal axis and minipuberty: an ongoing debate? *Front Endocrinol*. 2020;11:187.
5. Forest MG, Cathiard AM. Pattern of plasma testosterone and delta4-androstenedione in normal newborns: evidence for testicular activity at birth. *J Clin Endocrinol Metab*. 1975;41:977–980.
6. Terasawa E, Fernandez DL. Neurobiological mechanisms of the onset of puberty in primates. *Endocr Rev*. 2001;22:111–151.
7. Mitsushima D, Hei DL, Terasawa E. gamma-Aminobutyric acid is an inhibitory neurotransmitter restricting the release of luteinizing hormone-releasing hormone before the onset of puberty. *Proc Natl Acad Sci USA*. 1994;91:395–399.
8. Abreu AP, Dauber A, Macedo DB, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med*. 2013;368:2467–2475.
9. Jong MT, Gray TA, Ji Y, et al. A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. *Hum Mol Genet*. 1999;8:783–793.
10. Mitamura R, Yano K, Suzuki N, et al. Diurnal rhythms of luteinizing hormone, follicle-stimulating hormone, and testosterone secretion before the onset of male puberty. *J Clin Endocrinol Metab*. 1999;84:29–37.
11. Mitamura R, Yano K, Suzuki N, et al. Diurnal rhythms of luteinizing hormone, follicle-stimulating hormone, testosterone, and estradiol secretion before the onset of female puberty in short children. *J Clin Endocrinol Metab*. 2000;85:1074–1080.
12. Wu FC, Butler GE, Kelnar CJ, et al. Ontogeny of pulsatile gonadotropin releasing hormone secretion from midchildhood, through puberty, to adulthood in the human male: a study using deconvolution analysis and an ultrasensitive immunofluorometric assay. *J Clin Endocrinol Metab*. 1996;81:1798–1805.
13. Boyar RM, Rosenfeld RS, Kapen S, et al. Human puberty. Simultaneous augmented secretion of luteinizing hormone and testosterone during sleep. *J Clin Invest*. 1974;54:609–618.
14. Wu FC, Borrow SM, Nicol K, et al. Ontogeny of pulsatile gonadotrophin secretion and pituitary responsiveness in male puberty in man: a mixed longitudinal and cross-sectional study. *J Endocrinol*. 1989;123:347–359.
15. Jakacki RI, Kelch RP, Sauder SE, et al. Pulsatile secretion of luteinizing hormone in children. *J Clin Endocrinol Metab*. 1982;55:453–458.
16. Crofton PM, Evans AE, Wallace AM, et al. Nocturnal secretory dynamics of inhibin B and testosterone in pre- and peripubertal boys. *J Clin Endocrinol Metab*. 2004;89:867–874.
17. Veldhuis JD, King JC, Urban RJ, et al. Operating characteristics of the male hypothalamo-pituitary-gonadal axis: pulsatile release of testosterone and follicle-stimulating hormone and their temporal coupling with luteinizing hormone. *J Clin Endocrinol Metab*. 1987;65:929–941.
18. Kolby N, Busch AS, Aksglaede L, et al. Nocturnal urinary excretion of FSH and LH in children and adolescents with normal and early puberty. *J Clin Endocrinol Metab*. 2017;102:3830–3838.
19. Demir A, Voutilainen R, Juul A, et al. Increase in first morning voided urinary luteinizing hormone levels precedes the physical onset of puberty. *J Clin Endocrinol Metab*. 1996;81:2963–2967.
20. Arey BJ, Lopez FJ. Are circulating gonadotropin isoforms naturally occurring biased agonists? Basic and therapeutic implications. *Rev Endocr Metab Disord*. 2011;12:275–288.
21. Ratcliffe J, Nakanishi M, Jaffe RB. Identification of definitive and fetal zone markers in the human fetal adrenal gland reveals putative developmental genes. *J Clin Endocrinol Metab*. 2003;88:3272–3277.
22. Auchus RJ, Rainey WE. Adrenarche - physiology, biochemistry and human disease. *Clin Endocrinol*. 2004;60:288–296.
23. Weber A, Clark AJ, Perry LA, et al. Diminished adrenal androgen secretion in familial glucocorticoid deficiency implicates a significant role for ACTH in the induction of adrenarche. *Clin Endocrinol*. 1997;46:431–437.
24. Boettcher C, Hartmann MF, de Laffolie J, et al. Absent adrenarche in children with hypopituitarism: a study based on urinary steroid metabolomics. *Horm Res Paediatr*. 2013;79:356–360.
25. Topaloglu AK. Update on the genetics of idiopathic hypogonadotropic hypogonadism. *J Clin Res Pediatr Endocrinol*. 2017;9:113–122.
26. de Roux N, Genin E, Carel JC, et al. Hypogonadotropic hypogonadism due to loss of function of the KiSS1-derived peptide receptor GPR54. *Proc Natl Acad Sci U S A*. 2003;100:10972–10976.
27. Seminara SB, Messager S, Chatzidaki EE, et al. The GPR54 gene as a regulator of puberty. *N Engl J Med*. 2003;349:1614–1627.
28. Topaloglu AK, Tello JA, Kotan LD, et al. Inactivating KISS1 mutation and hypogonadotropic hypogonadism. *N Engl J Med*. 2012;366:629–635.
29. Teles MG, Bianco SD, Brito VN, et al. A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*. 2008;358:709–715.
30. Silveira LG, Noel SD, Silveira-Neto AP, et al. Mutations of the KISS1 gene in disorders of puberty. *J Clin Endocrinol Metab*. 2010;95:2276–2280.
31. Hughes IA. Releasing the brake on puberty. *N Engl J Med*. 2013;368:2513–2515.
32. Abreu AP, Macedo DB, Brito VN, et al. A new pathway in the control of the initiation of puberty: the MKRN3 gene. *J Mol Endocrinol*. 2015;54:R131–R139.
33. Valadares LP, Meireles CG, De Toledo IP, et al. MKRN3 mutations in central precocious puberty: a systematic review and meta-analysis. *J Endocr Soc*. 2019;3:979–995.
34. Navarro VM. Interactions between kisspeptins and neurokinin B. *Adv Exp Med Biol*. 2013;784:325–347.
35. Guran T, Tolhurst G, Bereket A, et al. Hypogonadotropic hypogonadism due to a novel missense mutation in the first extracellular loop of the neurokinin B receptor. *J Clin Endocrinol Metab*. 2009;94:3633–3639.
36. Topaloglu AK, Reimann F, Guclu M, et al. TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat Genet*. 2009;41:354–358.
37. Gianetti E, Tusset C, Noel SD, et al. TAC3/TACR3 mutations reveal preferential activation of gonadotropin-releasing hormone release by neurokinin B in neonatal life followed by reversal in adulthood. *J Clin Endocrinol Metab*. 2010;95:2857–2867.
38. Dauber A, Cunha-Silva M, Macedo DB, et al. Paternally inherited DLK1 deletion associated with familial central precocious puberty. *J Clin Endocrinol Metab*. 2017;102:1557–1567.
39. Mustanski BS, Viken RJ, Kaprio J, et al. Genetic and environmental influences on pubertal development: longitudinal data from Finnish twins at ages 11 and 14. *Dev Psychol*. 2004;40:1188–1198.

40. Treloar SA, Martin NG. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. *Am J Hum Genet*. 1990;47:137–148.
41. Kaprio J, Rimpela A, Winter T, et al. Common genetic influences on BMI and age at menarche. *Hum Biol*. 1995;67:739–753.
42. Gajdos ZK, Henderson KD, Hirschhorn JN, et al. Genetic determinants of pubertal timing in the general population. *Mol Cell Endocrinol*. 2010;324:21–29.
43. Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49:834–841.
44. Wyshak G, Frisch RE. Evidence for a secular trend in age of menarche. *N Engl J Med*. 1982;306:1033–1035.
45. Euling SY, Herman-Giddens ME, Lee PA, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*. 2008;121:S172–S191.
46. Herman-Giddens ME, Slora EJ, Wasserman RC, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics*. 1997;99:505–512.
47. Aksglaede L, Sorensen K, Petersen JH, et al. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*. 2009;123:e932–e939.
48. Chumlea WC, Schubert CM, Roche AF, et al. Age at menarche and racial comparisons in US girls. *Pediatrics*. 2003;111:110–113.
49. Sun SS, Schubert CM, Chumlea WC, et al. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics*. 2002;110:911–919.
50. Biro FM, Greenspan LC, Galvez MP, et al. Onset of breast development in a longitudinal cohort. *Pediatrics*. 2013;132:1019–1027.
51. Marti-Henneberg C, Vizmanos B. The duration of puberty in girls is related to the timing of its onset. *J Pediatr*. 1997;131:618–621.
52. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988-1994. *Arch Pediatr Adolesc Med*. 2001;155:1022–1028.
53. O'Rahilly S. Leptin: defining its role in humans by the clinical study of genetic disorders. *Nutr Rev*. 2002;60:S30–S34; discussion S68–S84, 5–7.
54. Farooqi IS, O'Rahilly S. 20 years of leptin: human disorders of leptin action. *J Endocrinol*. 2014;223:T63–T70.
55. Farooqi IS, Wangensteen T, Collins S, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*. 2007;356:237–247.
56. Montague CT, Farooqi IS, Whitehead JP, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*. 1997;387:903–908.
57. Clement K, Vaisse C, Lahlou N, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*. 1998;392:398–401.
58. Sienkiewicz E, Magkos F, Aronis KN, et al. Long-term metreleptin treatment increases bone mineral density and content at the lumbar spine of lean hypoleptinemic women. *Metabolism*. 2011;60:1211–1221.
59. Tena-Sempere M. Interaction between energy homeostasis and reproduction: central effects of leptin and ghrelin on the reproductive axis. *Horm Metab Res*. 2013;45:919–927.
60. Tena-Sempere M. Ghrelin and reproduction: ghrelin as novel regulator of the gonadotropic axis. *Vitam Horm*. 2008;77:285–300.
61. Barr VA, Malide D, Zarnowski MJ, et al. Insulin stimulates both leptin secretion and production by rat white adipose tissue. *Endocrinology*. 1997;138:4463–4472.
62. Moret M, Stettler R, Rodieux F, et al. Insulin modulation of luteinizing hormone secretion in normal female volunteers and lean polycystic ovary syndrome patients. *Neuroendocrinology*. 2009;89:131–139.
63. Roa J, Garcia-Galiano D, Varela L, et al. The mammalian target of rapamycin as novel central regulator of puberty onset via modulation of hypothalamic Kiss1 system. *Endocrinology*. 2009;150:5016–5026.
64. Roa J, Barroso A, Ruiz-Pino F, et al. Metabolic regulation of female puberty via hypothalamic AMPK-kisspeptin signaling. *Proc Natl Acad Sci USA*. 2018;115:E10758–E10767.
65. Vazquez MJ, Toro CA, Castellano JM, et al. SIRT1 mediates obesity- and nutrient-dependent perturbation of pubertal timing by epigenetically controlling Kiss1 expression. *Nat Commun*. 2018;9:4194.
66. De Leonibus C, Marcovecchio ML, Chiavaroli V, et al. Timing of puberty and physical growth in obese children: a longitudinal study in boys and girls. *Pediatr Obes*. 2014;9:292–299.
67. Kaplowitz PB, Slora EJ, Wasserman RC, et al. Earlier onset of puberty in girls: relation to increased body mass index and race. *Pediatrics*. 2001;108:347–353.
68. Lee JM, Appugliese D, Kaciroti N, et al. Weight status in young girls and the onset of puberty. *Pediatrics*. 2007;119:e624–e630.
69. Frisch RE, Revelle R. Height and weight at menarche and a hypothesis of critical body weights and adolescent events. *Science*. 1970;169:397–399.
70. Carlson L, Flores Poccia V, Sun BZ, et al. Early breast development in overweight girls: does estrogen made by adipose tissue play a role? *Int J Obes*. 2019;43:1978–1987.
71. Greenspan LC, Lee MM. Endocrine disrupters and pubertal timing. *Curr Opin Endocrinol Diabetes Obes*. 2018;25:49–54.
72. Leonardi A, Cofini M, Rigante D, et al. The effect of bisphenol a on puberty: a critical review of the medical literature. *Int J Environ Res Publ Health*. 2017;14.
73. Wolff MS, Pajak A, Pinney SM, et al. Associations of urinary phthalate and phenol biomarkers with menarche in a multiethnic cohort of young girls. *Reprod Toxicol*. 2017;67:56–64.
74. Supornsilchai V, Jantarat C, Nosoognoen W, et al. Increased levels of bisphenol A (BPA) in Thai girls with precocious puberty. *J Pediatr Endocrinol Metab*. 2016;29:1233–1239.
75. Miao M, Wang Z, Liu X, et al. Urinary bisphenol A and pubertal development in Chinese school-aged girls: a cross-sectional study. *Environ Health*. 2017;16:80.
76. Wang Z, Li D, Miao M, et al. Urine bisphenol A and pubertal development in boys. *Int J Hyg Environ Health*. 2017;220:43–50.
77. Garel L, Dubois J, Grignon A, et al. US of the pediatric female pelvis: a clinical perspective. *Radiographics*. 2000;21:1393–1407.
78. Biro FM, Pajak A, Wolff MS, et al. Age of menarche in a longitudinal US cohort. *J Pediatr Adolesc Gynecol*. 2018;31:339–345.
79. Bartke A, Sun LY, Longo V. Somatotropic signaling: trade-offs between growth, reproductive development, and longevity. *Physiol Rev*. 2013;93:571–598.
80. Link K, Blizzard RM, Evans WS, et al. The effect of androgens on the pulsatile release and the twenty-four-hour mean concentration of growth hormone in peripubertal males. *J Clin Endocrinol Metab*. 1986;62:159–164.
81. Martha Jr PM, Gorman KM, Blizzard RM, et al. Endogenous growth hormone secretion and clearance rates in normal boys, as determined by deconvolution analysis: relationship to age, pubertal status, and body mass. *J Clin Endocrinol Metab*. 1992;74:336–344.
82. Butte NF, Hopkinson JM, Wong WW, et al. Body composition during the first 2 years of life: an updated reference. *Pediatr Res*. 2000;47:578–585.
83. Fomon SJ, Haschke F, Ziegler EE, et al. Body composition of reference children from birth to age 10 years. *Am J Clin Nutr*. 1982;35:1169–1175.
84. Veldhuis JD, Roemmich JN, Richmond EJ, et al. Endocrine control of body composition in infancy, childhood, and puberty. *Endocr Rev*. 2005;26:114–146.
85. Van Loan MD. *Human Body Composition*. Champaign: Human Kinetics; 1996.
86. Hammer LD, Wilson DM, Litt IF, et al. Impact of pubertal development on body fat distribution among white, Hispanic, and Asian female adolescents. *J Pediatr*. 1991;118:975–980.
87. Leonard MB, Zemel BS. Current concepts in pediatric bone disease. *Pediatr Clin*. 2002;49:143–173.
88. Bailey DA, Martin AD, McKay HA, et al. Calcium accretion in girls and boys during puberty: a longitudinal analysis. *J Bone Miner Res*. 2000;15:2245–2250.

89. Bonjour JP, Theintz G, Buchs B, et al. Critical years and stages of puberty for spinal and femoral bone mass accumulation during adolescence. *J Clin Endocrinol Metab*. 1991;73:555–563.
90. Ribot C, Pouilles JM, Bonneu M, et al. Assessment of the risk of post-menopausal osteoporosis using clinical factors. *Clin Endocrinol*. 1992;36:225–228.
91. Johnell O, Gullberg B, Kanis JA, et al. Risk factors for hip fracture in European women: the MEDOS Study. Mediterranean Osteoporosis Study. *J Bone Miner Res*. 1995;10:1802–1815.
92. Chevalley T, Bonjour JP, Ferrari S, et al. Influence of age at menarche on forearm bone microstructure in healthy young women. *J Clin Endocrinol Metab*. 2008;93:2594–2601.
93. Zirilli L, Rochira V, Diazzi C, et al. Human models of aromatase deficiency. *J Steroid Biochem Mol Biol*. 2008;109:212–218.
94. Soyka LA, Fairfield WP, Klibanski A. Clinical review 117: hormonal determinants and disorders of peak bone mass in children. *J Clin Endocrinol Metab*. 2000;85:3951–3963.
95. Gilsanz V, Skaggs DL, Kovanlikaya A, et al. Differential effect of race on the axial and appendicular skeletons of children. *J Clin Endocrinol Metab*. 1998;83:1420–1427.
96. Sadeh A, Dahl RE, Shahar G, et al. Sleep and the transition to adolescence: a longitudinal study. *Sleep*. 2009;32:1602–1609.
97. Carskadon MA, Acebo C, Jenni OG. Regulation of adolescent sleep: implications for behavior. *Ann NY Acad Sci*. 2004;1021:276–291.
98. Feinberg I, Campbell IG. Sleep EEG changes during adolescence: an index of a fundamental brain reorganization. *Brain Cognit*. 2010;72:56–65.
99. Giedd JN, Blumenthal J, Jeffries NO, et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci*. 1999;2:861–863.
100. Sowell ER, Thompson PM, Toga AW. Mapping changes in the human cortex throughout the span of life. *Neuroscientist*. 2004;10:372–392.
101. De Bellis MD, Keshavan MS, Beers SR, et al. Sex differences in brain maturation during childhood and adolescence. *Cerebr Cortex*. 2001;11:552–557.
102. Lenroot RK, Gogtay N, Greenstein DK, et al. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage*. 2007;36:1065–1073.
103. Chada M, Prusa R, Bronsky J, et al. Inhibin B, follicle stimulating hormone, luteinizing hormone and testosterone during childhood and puberty in males: changes in serum concentrations in relation to age and stage of puberty. *Physiol Res*. 2003;52:45–51.
104. Crofton PM, Evans AE, Groome NP, et al. Inhibin B in boys from birth to adulthood: relationship with age, pubertal stage, FSH and testosterone. *Clin Endocrinol*. 2002;56:215–221.
105. Chada M, Prusa R, Bronsky J, et al. Inhibin B, follicle stimulating hormone, luteinizing hormone, and estradiol and their relationship to the regulation of follicle development in girls during childhood and puberty. *Physiol Res*. 2003;52:341–346.
106. Jeffery A, Streeter AJ, Hosking J, et al. Anti-Mullerian hormone in children: a ten-year prospective longitudinal study (EarlyBird 39). *J Pediatr Endocrinol Metab*. 2015;28:1153–1162.
107. Pinkney J, Streeter A, Hosking J, et al. Adiposity, chronic inflammation, and the prepubertal decline of sex hormone binding globulin in children: evidence for associations with the timing of puberty (Earlybird 58). *J Clin Endocrinol Metab*. 2014;99:3224–3232.
108. Garces C, Oya I, Lasuncion MA, et al. Sex hormone-binding globulin and lipid profile in pubertal children. *Metabolism*. 2010;59:166–171.
109. Wang C, Catlin DH, Demers LM, et al. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. *J Clin Endocrinol Metab*. 2004;89:534–543.

103

Endocrinology of Pubertal Disorders

Nadine G. Haddad and Erica A. Eugster

OUTLINE

Introduction, 1731
Precocious Puberty, 1731
Variants of Precocious Puberty, 1731
  Premature Thelarche, 1732
  Premature Pubarche, 1732
Pathologic Precocious Puberty, 1734
  Central Precocious Puberty, 1734
  Peripheral Precocious Puberty, 1736
  Abnormalities of G Proteins, 1736
  Abnormalities of Gonadotropin Receptors, 1738
Tumors, 1739
  Gonadal Tumors, 1739
  Adrenal Tumors, 1739
  Germ Cell Tumors, 1739
  Exogenous Sex Steroids, 1739
Hypothyroidism, 1739
Delayed Puberty, 1740
  Definition, 1740
Benign Variants of Delayed Puberty, 1741
  Constitutional Delay of Growth and Puberty, 1741
Pathologic Delayed Puberty, 1742
  Hypothalamic-Pituitary Hypogonadism, 1742
  Primary Hypogonadism, 1745
Treatment of Delayed Puberty, 1748
  Girls, 1748
  Boys, 1748
Conclusion, 1748

INTRODUCTION

Disorders of puberty are among the most fascinating and complex areas of pediatric endocrinology. The combination of an extensive differential diagnosis, dramatic clinical manifestations, and intricate pathophysiology provides continual challenges for clinicians and scientists alike. Although many questions remain unanswered, phenomenal breakthroughs have occurred in elucidation of the molecular genetics, lifelong implications, and therapeutic outcomes of different forms of precocious and delayed puberty. Future advances will continue to enhance our understanding of the biological basis of normal and abnormal secondary sexual development.

PRECOCIOUS PUBERTY

Precocious puberty has been defined historically as the onset of any sign of puberty in girls younger than 8 years of age and in boys younger than 9 years of age. However, there has been considerable debate about whether these guidelines are appropriate for girls. This was precipitated by a large, office-based, cross-sectional study in which pubertal development was documented in a substantial proportion of normal girls who were younger than 8 years of age.[1] Subsequent analysis of the original data has demonstrated a correlation between body mass index (BMI) and the onset of puberty, suggesting that heavier girls may be entering puberty at slightly younger ages than lean peers. This was corroborated by a longitudinal study demonstrating that a higher BMI zscore in girls as young as 3 years of age, and a higher rate of change of BMI in early childhood is associated with earlier puberty.[2] Despite observations from international and US studies indicating that the average age of menarche has been stable over several decades, a rapid rise in the incidence of obesity in children may well lead to a secular trend in the onset of puberty in girls.[3] To what degree historical cutoffs should be revised in boys has also been challenged by a report of earlier testicular enlargement in one study, although age at attainment of Tanner V development was indistinguishable from standard norms.[4] In contrast, racial differences in the timing of pubertal development are well documented, with African American children typically starting puberty 6 to 12 months earlier than White children and Hispanic children having an onset of puberty intermediate between these racial groups.[5] Regardless of the ultimate consensus in terms of what should be considered "normal," individual factors such as growth parameters, BMI, skeletal maturation, height prediction, family history, and psychosocial status need to be taken into consideration when one is evaluating a child who presents with early puberty. It is also important to recognize that the tempo of puberty varies among children, with duration of normal puberty often being inversely proportional to age of pubertal onset.[6]

VARIANTS OF PRECOCIOUS PUBERTY

Evidence of pubertal development at a younger than normal age does not necessarily imply a pathologic condition that requires intervention. Two distinct conditions commonly occur in children when there is isolated development of a secondary sexual characteristic. *Premature thelarche* refers to isolated breast enlargement in girls, whereas *premature pubarche* describes early onset of pubic hair in boys or girls. Although different in terms of origin, pathogenesis, and implications for future health, both conditions may be considered variations of normal development. Pubic hair of infancy is a rare condition that has been described in a few case series. Although little is known about its cause and long-term natural history, pubic hair of infancy is also likely a benign entity that may represent a variant of normal pubertal development.

Premature Thelarche

Clinical Features. Premature thelarche occurs most commonly in girls between infancy and 3 years of age, although it can occur any time before the normal onset of puberty. The typical history is that breast tissue has been noted since birth or that gradual enlargement of the breasts occurs over several months without any other evidence of puberty. Growth velocity and skeletal maturation are within the normal range for age.[7] On physical examination, breast tissue usually is easily discernible, with a texture characterized by firm elasticity and a palpable margin at the periphery. Breast enlargement may be unilateral or bilateral and usually does not exceed a Tanner stage III level of development (Fig. 103.1). The nipples typically remain immature, and the vaginal mucosa usually appears unestrogenized, although estrogenization does not exclude a diagnosis of premature thelarche. When an ultrasensitive assay is used, estradiol levels have been found to be higher in girls with premature thelarche compared with controls.[8] Despite this, ovarian and uterine volumes remain in a prepubertal range, in contrast to enlargement of these structures, which is characteristic of central precocious puberty.[9]

Pathophysiology. The precise cause of premature thelarche remains unknown, although several theories about its pathogenesis have been proposed. These include individual differences in peripheral sensitivity to estradiol, as well as an alteration in the ratio of estrogen to androgens from increased sex hormone-binding globulin levels. Gain-of-function mutations in the follicle-stimulating hormone (FSH) receptor or $G_s\alpha$ have been hypothesized to be a cause of premature thelarche, although a search for molecular genetic abnormalities in these patients has been negative thus far.[10] Most cases of premature thelarche are characterized by an FSH-predominant response to gonadotropin-releasing hormone (GnRH) stimulation testing. However, it is still unclear whether this indicates involvement of the hypothalamic-pituitary-gonadal (HPG) axis in causing this condition, because this response is also characteristic of the prepubertal state in young girls.[11] One likely hypothesis is that premature thelarche exists along a continuum of hypothalamic-pituitary-ovarian activity, representing one end of a clinical spectrum that, in some patients, manifests as full-blown central precocious puberty characterized by a luteinizing hormone (LH)-predominant response to GnRH stimulation. What governs whether activation of this axis is mild, partial, transient, or complete remains a mystery. Higher serum levels of kisspeptin (a neuropeptide involved in central regulation of GnRH release) were noted in a study of young girls with premature thelarche, suggesting a central etiology.[12] It is interesting to note that premature thelarche is relatively common in girls, whereas prepubertal gynecomastia is uncommon in boys, and both FSH and estradiol levels are higher in girls than in boys at all ages.[13] Another theory is that premature thelarche and other forms of early puberty occur as a result of exposure to endocrine-disrupting chemicals (EDCs) such as phthalates, bisphenol A, pesticides, or phytoestrogens such as those found in soy formula.[14] Despite concerns about potential deleterious effects of EDCs on the reproductive system, results of existing studies are mixed, and no clear evidence of causality has yet emerged[15] (see Chapter 137).

Natural History. Follow-up studies of patients with premature thelarche have indicated several possible scenarios in terms of clinical course. Complete regression of breast tissue occurs in a substantial percentage of girls with premature thelarche, particularly in girls with onset before 2 years of age. In other patients, breast tissue may persist, or even enlarge, although initial studies failed to identify an increased risk for pubertal progression or additional medical problems.[16] In some series, as many as 14% of girls with classic premature thelarche progress to central precocious puberty, although baseline clinical and biochemical characteristics of these patients are indistinguishable from those whose condition remains stable.[17] It is reassuring that adult stature and the timing of central puberty appear to be normal in girls with nonprogressive premature thelarche.

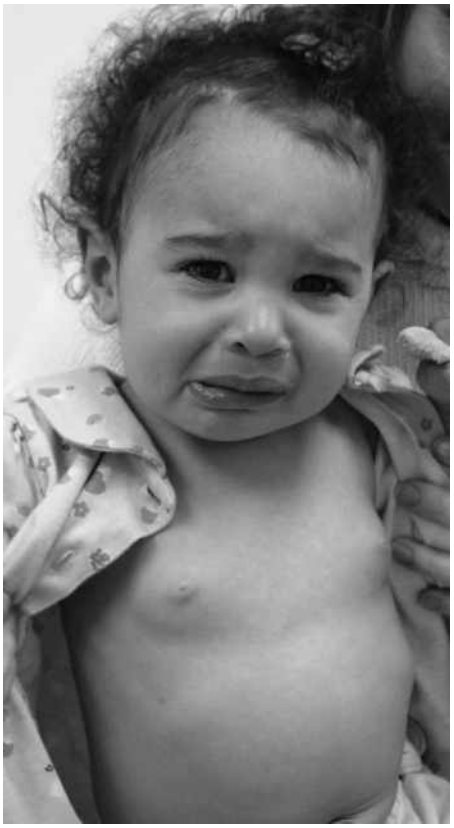


Fig. 103.1 Typical appearance of premature thelarche in a 2.5-year-old girl.

Management. In girls with typical premature thelarche, reassurance and follow-up are all that are indicated. Careful clinical assessment, including determination of growth velocity, should be performed at 3- to 6-month intervals. A bone age radiograph is helpful at the time of diagnosis and is reassuring if it is consistent with the chronologic age. Because of the possibility of progression to central precocious puberty, it is recommended that follow-up by a pediatric endocrinologist or the primary care physician continues.

Premature Pubarche

Clinical Features. Premature pubarche is characterized by the development of pubic hair in prepubertal children. Additional findings often include axillary hair, body odor, and mild acne. Most cases of premature pubarche are secondary to premature adrenarche, which is characterized by mildly elevated serum adrenal androgen levels for age, especially dehydroepiandrosterone sulfate, with values corresponding to the degree of Tanner stage development. Premature adrenarche usually occurs in children between 6 and 8 years of age but may occur earlier. Additional diagnostic criteria include the absence of other signs of puberty, such as breast development. Although growth velocity and skeletal maturation may be slightly advanced, significant virilization, such as clitoromegaly, voice change, or increased muscle mass, is

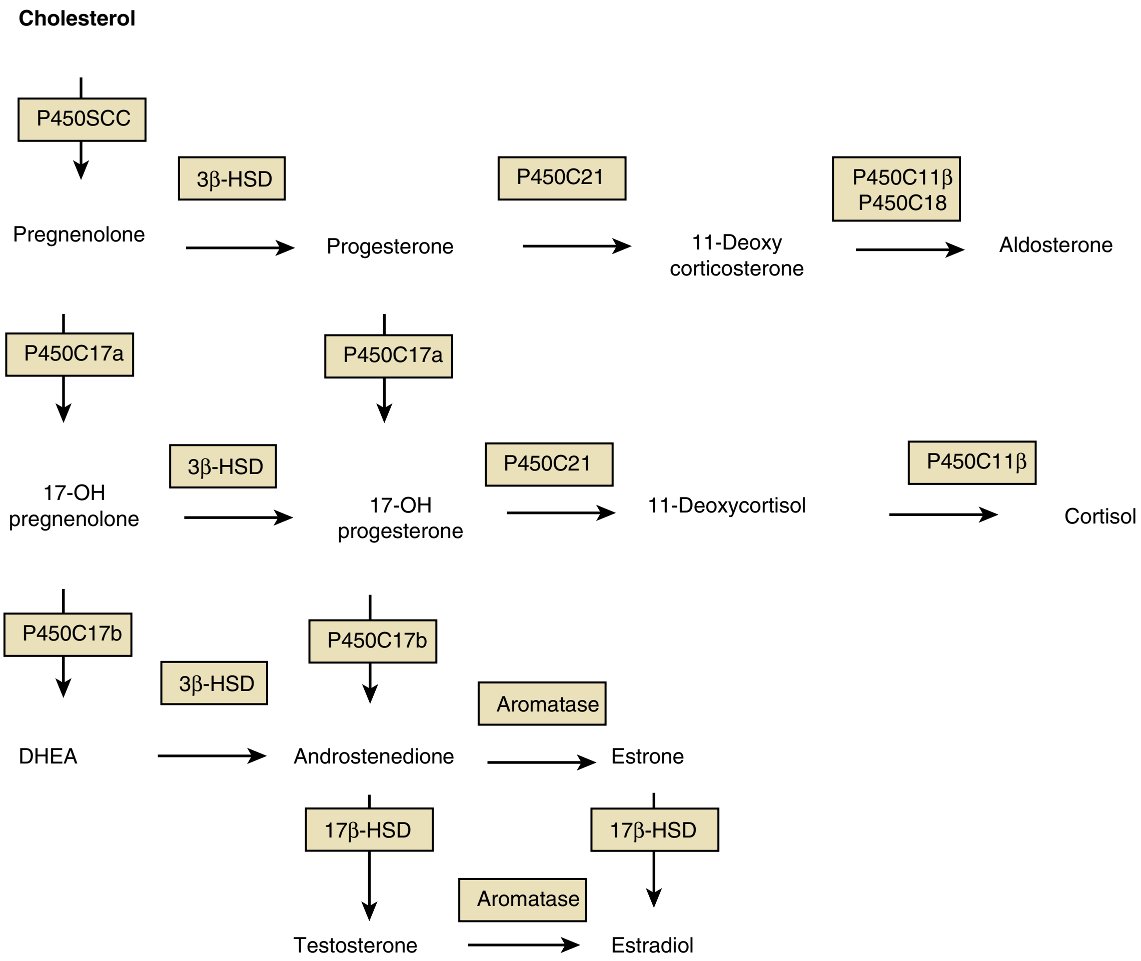


Fig. 103.2 Simplified adrenal steroidogenesis pathway. *P450SCC*, 20-22-desmolase; *P450C11β*, 11β-hydroxylase; *3β-HSD*, 3β-hydroxysteroid dehydrogenase; *P450C17a*, 17 α-hydroxylase; *P450C17b*, 17-20 desmolase; *P450C21,* 21 alpha-hydroxylase; *17β-HSD*, 17β-hydroxysteroid dehydrogenase.

absent.[18] In a study of 266 children with premature adrenarche, a significantly advanced bone age (more than 2 years) was noted in 30%. These patients were significantly taller and more overweight, and the impact of an advanced bone age on their predicted adult height (PAH) was minimal.[19] Because premature pubarche can be the presenting sign in congenital adrenal hyperplasia (CAH) or virilizing adrenal or gonadal tumors, these conditions should be considered and excluded with appropriate laboratory testing.

Pathophysiology. Attempts to define the precise clinical and biochemical profile of premature adrenarche have yielded significant heterogeneity in terms of underlying pathophysiology and concurrent metabolic features. Physiologic adrenarche, which occurs normally during midchildhood, refers to the development of the zona reticularis and an associated increase in adrenal androgen secretion. This is accompanied by alterations in adrenal enzyme activity, characterized by increases in 17,20-desmolase and 17-hydroxylase enzymes and a decrease in 3β-hydroxysteroid dehydrogenase activity[20] (Fig. 103.2). Although much progress has been made in understanding the physiology of adrenarche, the driving force behind this process in both premature and physiologic adrenarche remains unknown. Differences in peripheral metabolism of adrenal androgens have been implicated as a contributing factor in some patients, supported by the finding of elevated 3α-androstenediol glucuronide, a marker of peripheral adrenal androgen activity, in children with premature adrenarche compared with controls.[21] Partial adrenal enzyme deficiencies, including 21-hydroxylase, 3β-hydroxysteroid dehydrogenase, and 11-hydroxylase deficiencies, have been reported at variable frequencies among patients presenting with premature adrenarche,[22] whereas the prevalence of heterozygosity for 21-hydroxylase deficiency has been found to be increased in this population.[23] A search for genetic markers associated with premature adrenarche has revealed sequence variants in a number of genes encoding for steroidogenic enzymes and related receptors, although the functional consequence of these variants is unclear. Last, a subset of patients has been identified with steroid hormone profiles intermediate between those with “typical” premature adrenarche and those with CAH, a condition described as “exaggerated adrenarche.” An increased incidence of premature adrenarche has been noted in children with a history of central nervous system insult such as head trauma or hypoxic-ischemic injury, suggesting that a central process may play a role in its development.

Natural History. Despite the fact that children with premature adrenarche experience a normal onset of central puberty and achieve normal adult height, the historical assumption that this developmental variation was “benign” has proved to be incorrect. For many children, premature adrenarche may now be thought of as one stage in a continuum of metabolic abnormalities that ultimately manifests as metabolic syndrome, which is typified by insulin resistance, hyperandrogenism, and dyslipidemia. Indeed, as many as 45% of girls with premature adrenarche may develop polycystic ovarian syndrome (PCOS) later in life. Even lean girls with premature adrenarche have demonstrable alterations in body composition and insulin and androgen levels compared with controls.[24] Insulin resistance has also been reported in boys with premature adrenarche, suggesting that a propensity for future development of cardiovascular disease and type 2 diabetes pertains to both genders with this condition. Prenatal growth restriction resulting in low birth weight has been established as a common risk factor for premature adrenarche, functional ovarian hyperandrogenism, and hyperinsulinemia, implying that these metabolic derangements may be related via a shared primary abnormality. However, predictive factors in patients who will develop hyperandrogenism in later life thus far have eluded detection. Further biochemical and molecular genetic characterization of patients with premature adrenarche and its postmenarchal corollaries should yield greater understanding of pathogenesis and allow improved diagnostic discrimination and prognostic accuracy.[25]

Management. In patients with isolated pubic hair and normal linear growth, a bone age radiograph may be the only diagnostic study indicated, although the finding of normal or only mildly elevated adrenal androgens provides additional reassurance. Further investigation, including baseline or stimulated serum adrenal androgen and steroid precursor levels, is warranted if there is marked growth acceleration, advanced skeletal maturation, clitoromegaly, or other significant virilization on physical examination. Because the condition is progressive, most patients experience a gradual increase in body hair over time, accompanied by a normal rate of skeletal advancement. Insulin sensitization with metformin early after menarche in girls with premature adrenarche appears to prevent the progression of PCOS in the experimental setting. Metformin has also been shown to improve adipose body composition and may delay or prevent hirsutism, androgen excess, oligomenorrhea, and PCOS in prepubertal girls with premature pubarche and history of low birth weight.[26] However, to date, there is no evidence that pharmacologic intervention in premature adrenarche in the clinical setting is indicated, especially because we are not yet able to predict which children will develop more serious adverse associations.

Pubic Hair of Infancy. Pubic hair of infancy is a unique entity that has been described in both boys and girls. Typically, infants have isolated pubic hair growth, usually in an atypical location such as the scrotum in males and the mons in females. Linear growth, skeletal maturation, and laboratory evaluation are otherwise normal. Pubic hair tends to resolve spontaneously before 1 year of age. Given that long-term follow-up of these patients is not available, it is unclear whether there are any clinical consequences for children with this developmental variant.[27]

KEY POINTS

- Premature thelarche is a benign variant of premature pubertal development characterized by isolated breast development in girls younger than 3 years of age. The breast development is typically nonprogressive, and linear growth and age at central puberty are normal. Clinical observation is indicated because of the possibility of progression to central precocious puberty in a small subset of girls.
- Premature pubarche is characterized by premature development of pubic hair with or without axillary hair, body odor, or mild acne in prepubertal children, and mildly elevated serum adrenal androgen levels for age, especially dehydroepiandrosterone sulfate. The bone age may be mildly advanced, but growth potential is typically normal. Girls with premature adrenarche are at risk to develop polycystic ovarian syndrome later in life.

PATHOLOGIC PRECOCIOUS PUBERTY

Central Precocious Puberty

Also termed *gonadotropin-dependent,* central precocious puberty (CPP) refers to early activation of the HPG, and thus "central," axis. The underlying physiology of this form of precocious puberty therefore is identical to that of normal puberty, with the distinction being that it occurs at an abnormal time and may be the result of a pathologic condition.

Clinical Features. Girls with CPP typically present with breast enlargement and other estrogenic changes accompanied by growth acceleration and skeletal maturation. Unlike puberty that begins at a normal age, in CPP, dissociation is often noted between gonadarche and adrenarche, which may be a result of lower adrenal androgen levels during early childhood. Testicular enlargement, which is the first sign of CPP in boys, often goes unrecognized, with subsequent genital and body hair development prompting medical evaluation.

Pathophysiology. Causes of CPP can be divided into two general categories. Idiopathic CPP is used to designate cases in which no anatomic abnormality is identified, whereas in the remaining cases a central nervous system insult, either congenital or acquired, is believed to be the trigger. CPP is far more common in girls than in boys, with as many as 90% of cases having historically been designated as idiopathic.[28] However, the recognition of four monogenetic causes of CPP has resulted in a decrease in the percentage of cases deemed idiopathic.[29] The first genetic etiology discovered as the basis for CPP was a gain-of-function mutation in the KISS1 receptor (KISS1R) that has thus far been described in a single patient.[30] Subsequently, an activating mutation of the *KISS-1* gene was described in one child with very early-onset CPP.[31] In both instances, a resistance to degradation of the mutated protein compared to the wild-type protein is postulated to be the basis for the physiologic effect. Inactivating mutations in *MKRN3,* the gene encoding makorin RING-finger protein 3 that is a paternally expressed imprinted gene located in the Prader–Willi syndrome region on chromosome 15 and is involved in inhibiting puberty, were initially identified in five individuals from 15 families with familial CPP.[32] *MKRN3* mutations are well established as the most common genetic cause of CPP and have been described in sporadic as well as familial cases, where they account for up to approximately 40% of affected individuals.[33] There is evidence that *MKRN3* mutations may be underreported in boys, who experience the onset of puberty at an age very close to the lower limit of normal compared with the general population. In keeping with evidence suggesting an inhibitory role of *MKRN3* on the HPG axis is the observation that serum levels decrease in healthy children as pubertal onset draws near.[34] Mutations in Delta-like homolog 1 (*DLK1*), another paternally expressed imprinted gene, have been implicated in the pathogenesis of CPP. *DLK1* is expressed in the hypothalamus and kisspeptin-secreting neurons. In a study of a family with multigenerational CPP, a complex genetic defect in *DLK1* was identified.[35] The exact role of *DLK1* in the modulation of pubertal timing is unclear. However, it is known to be a molecular gatekeeper of adipogenesis, and adult women with *DLK1* mutations exhibit a metabolic phenotype marked by insulin resistance, obesity, and dyslipidemia. Thus, *DLK1* appears to be a unique link between metabolism and reproduction.[36] Several forms of intracranial lesions have been implicated as causative of CPP and are seen more frequently in boys. The most common is a hypothalamic hamartoma, a heterotopic mass located typically in the region of the tuber cinereum[37](Fig. 103.3). This congenital malformation consists of GnRH-secreting neurons or transforming growth factor α–producing astroglial cells, both of which

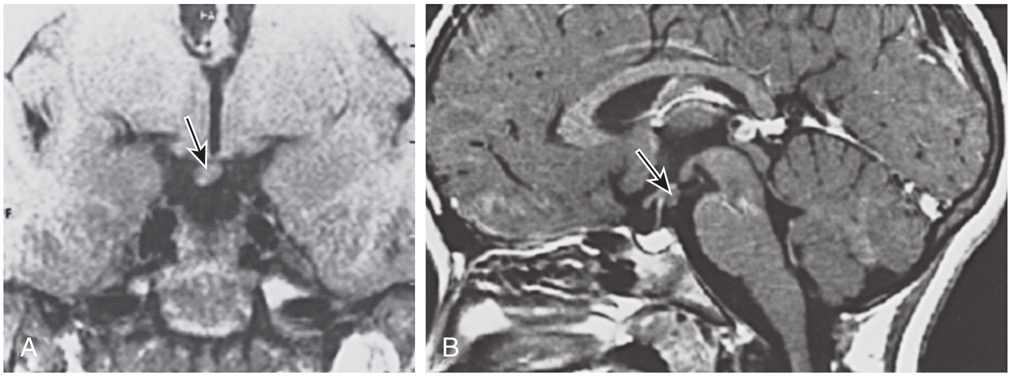


Fig. 103.3 Head magnetic resonance imaging scan revealed a pedunculated hypothalamic hamartoma in the region of the tuber cinereum, as indicated by the *arrows* in frontal (A) and sagittal (B) sections.

are believed to function as ectopic triggers on the hypothalamic pulse generator, resulting in an escape from the normal central nervous system inhibitory constraint on pubertal onset. Rarely, hypothalamic hamartomas are associated with an increased incidence of seizures, particularly gelastic, or laughing, seizures.[38] Other central nervous system abnormalities that are associated with CPP include hypothalamic-pituitary or optic chiasm tumors, pineal or suprasellar cysts, septooptic dysplasia, neurofibromatosis, head trauma, and hydrocephalus. The mechanism of intracranial abnormalities leading to precocious puberty is believed to be disruption of the neural pathways that normally exert tonic inhibition on the brain in terms of activation of puberty. This process results in a seemingly paradoxical situation in which many disorders associated with deficiencies of anterior pituitary hormones may be associated with early gonadotropin secretion.

Diagnosis. The diagnosis of CPP is based on the ability to detect activation of the HPG axis, during which sensitivity to stimulation is increased at each level of the axis. The GnRH stimulation test takes advantage of the fact that a "primed" pituitary gland responds to GnRH with a brisk elevation of gonadotropins, which in turn stimulates gonadal sex steroid secretion. Historically, synthetic GnRH (Factrel) was used for the evaluation of CPP. However, this agent is not available in the United States. Although several different protocols have been developed, the most widely used is the GnRH analog stimulation test with multiple blood sampling, in which nafarelin or leuprolide is used.

A pubertal response to GnRH stimulation is characterized by a pattern of LH predominance indicating an elevation of LH above that of FSH, whereas a prepubertal response is marked by a minimal increase in LH, if any, but may elicit an FSH response above baseline. An intermediate response, typified by an FSH-predominant response, is seen in early central puberty and in premature thelarche. A caveat of the various available tests is their inability to differentiate early central puberty from normal. This stems in part from wide variability, poor reproducibility, and lack of age-based references for the various available gonadotropin assays. However, a random ultrasensitive LH performed via immunochemiluminometric or electrochemiluminescence assay has a high sensitivity for the presence of CPP and is sufficient to confirm the diagnosis without stimulation testing. Additional evaluation of patients with CPP typically includes a magnetic resonance imaging (MRI) scan of the brain to rule out an intracranial lesion, although whether this is necessary in all patients is controversial, and a genetic diagnosis would certainly obviate this test.[39] A bone age radiograph provides important information about the degree of skeletal maturation and remaining growth potential. Pelvic ultrasonography is useful in documenting uterine and ovarian enlargement, ovarian symmetry, or the presence of cysts or other ovarian abnormalities. In patients with a prepubertal or suppressed response to GnRH stimulation, further investigation for causes of peripheral precocious puberty is indicated.

Treatment. Potential negative consequences of precocious puberty include compromise of adult stature because of premature fusion of epiphyseal growth plates and psychosocial issues related to a discrepancy between physical and emotional maturity. Therefore, the primary goal of therapy is to temporarily suppress the HPG axis to allow improved adult height and to simultaneously restore the child to a physiologic state comparable with that of his or her peers. The discovery of safe and effective treatment in the form of GnRH analogs revolutionized the therapy of CPP in the mid-1980s, and this is the treatment of choice for this condition.[40] Although the mechanism of action of GnRH analog therapy is not completely understood, it is believed that downregulation of pituitary gonadotropin-releasing hormone receptor occurs as a result of sustained high levels of GnRH. Several long-acting GnRH analogs have been used in the treatment of CPP, with subcutaneous (SC), intranasal, and intramuscular (IM) routes of administration.[41] Long-acting injectable options for GnRH analog therapy include 3-monthly and 6-monthly IM formulations, as well as a 6-monthly SC injection (Fensolvi).[42] A SC implant containing the potent GnRH analog histrelin is also available and, while marketed for annual use, lasts for at least 2 years.[43] The degree of biochemical suppression of the HPG axis varies by treatment modality, with the lower doses of extended-release injectable preparations generally being less potent than the higher doses and the histrelin implant.[44] Whether these discrepancies will translate into differences in treatment outcomes is unknown. Although untested, it is possible that GnRH antagonists may eventually have a role in the treatment of CPP as well.[45] Therapy for CPP is characterized by a return of gonadal steroids to prepubertal levels, along with a halt in pubertal progression and a deceleration in growth velocity and rate of skeletal maturation. Although ultrasensitive LH values do not predictably revert to a prepubertal range,[46] therapeutic efficacy may be documented by gonadotropin suppression in response to GnRH analog stimulation testing. Pelvic ultrasonography represents a potentially useful adjunctive tool for monitoring the effects of treatment in girls, which are characterized by regression of the internal genitalia from a pubertal to a prepubertal size. In boys, a decrease in testicular volume provides similar information. Although therapy has a profound impact on the outcome of many patients, a subset of patients with precocious puberty has been identified in whom height prognosis is quite favorable, even without treatment. These patients are characterized as having an indolent form of precocious puberty, in which pubertal progression occurs gradually, and bone age advancement is minimal.[47] Distinguishing features at presentation among this subset of patients include an increased height age/bone age ratio and PAH similar to target height when compared with patients who have rapidly progressive CPP. Although psychosocial concerns are often touted as a rationale for treatment, many studies have failed to document psychological distress in girls with early puberty, and no consensus on this aspect of treatment currently exists.[48]

Therapeutic Outcomes. Several different outcome measures have been examined with regards to the safety and efficacy of GnRH analog therapy for CPP, including adult height, bone mineralization, and resumption of puberty after treatment.[49] Each of these is discussed individually.

Adult Height. Evidence suggests that girls with onset of CPP before 6 years of age have the best response to therapy, with a height gain of 2.9 to 12.5 cm with GnRH analogs. In contrast, GnRH analogs appear to be ineffective in increasing height in children treated after age 7.5 to 8 years.[50] Indications for GnRH analog therapy other than CPP are not supported by current data.

Bone Mineralization. The effects of sex steroids on bone density are well described, with peak acquisition of bone mass occurring during puberty. Reports of a rapid decrease in bone mineralization after GnRH analog therapy in adults initially led to concerns about potential similar negative consequences in children with precocious puberty. However, several follow-up studies have now demonstrated normal bone mineral density in adult patients previously treated with a GnRH analog.[51]

Once pubertal suppression is removed via discontinuation of GnRH analog therapy, spontaneous resumption of puberty occurs. Studies have indicated a resumption of pubertal LH pulsatile secretion in most patients within 4 months after therapy, as well as normal menarche, which on average occurs 18 months after treatment is stopped. However, an inverse relationship between age at discontinuation of treatment and time to menarche has been reported.[52] Reproductive function in women treated with a GnRH analog for CPP appears to be equal to that seen in controls.[53] However, an increased incidence of hyperandrogenism has been reported independent of treatment status.

KEY POINTS

- Central precocious puberty (CPP) results from early activation of the hypothalamic-pituitary-gonadal axis, characterized clinically by early breast development in girls or testicular enlargement in boys, accelerated linear growth, and skeletal maturation, with the possibility of compromised adult height.
- CPP is commonly idiopathic, and more so in girls. The recognition of four monogenetic causes of CPP (the *MKRN3*, *KISS1*, *KISS1R*, and *DLK1* genes) has resulted in a decrease in the percentage of cases deemed idiopathic.
- The diagnosis of CPP is based on the demonstration of a luteinizing hormone (LH)-predominant response on a gonadotropin-releasing hormone (GnRH) analog stimulation test or an elevated random ultrasensitive LH performed via immunochemiluminometric or electrochemiluminescence assay.
- Treatment of CPP is with a GnRH analog. Several long-acting GnRH analogs with different routes of administration are available. The primary goal of treatment is height preservation.
- GnRH analogs improve adult height in girls with CPP, with the best response observed in those with onset of puberty before age 6 years and no apparent effect on height seen in those treated after age 7.5 to 8 years.

Peripheral Precocious Puberty

Also termed *gonadotropin-independent,* peripheral precocious puberty (PPP) refers to cases in which sex steroid exposure leads to pubertal development through a source other than activation of the HPG axis. This broad category encompasses an extremely heterogeneous group of abnormalities.[54] Evaluation and differential diagnosis depend on whether the child is a girl or boy, and on whether the clinical presentation reveals changes associated with estrogens, androgens, or both. Individual causes of PPP are discussed here, and the genotype-phenotype relationships seen in these disorders are outlined in Table 103.1.

Abnormalities of Enzymes Involved in Steroid Biosynthesis

Congenital Adrenal Hyperplasia. CAH represents a group of diseases stemming from an autosomal recessive inherited deficiency of an enzyme required for adrenal steroidogenesis (see Chapter 94) (see Fig. 103.2). The resulting disorders represent a wide spectrum of molecular genetic abnormalities and phenotypic characteristics. Excess adrenal androgen secretion as a consequence of CAH is the most common cause of abnormal postnatal virilization in children. The hallmark of these disorders is an enzymatic block in the cortisol biosynthetic pathway leading to lack of negative feedback stimulation at the level of the pituitary gland. The resulting increase in pituitary adrenocorticotropic hormone and buildup of steroid precursors proximal to the enzyme block result in an "overflow" of these precursors to the androgen synthetic pathway with subsequent overproduction of adrenal androgens. Once the diagnosis has been made, glucocorticoid replacement is indicated to result in cessation of adrenal hyperstimulation and a return of adrenal androgens to near-normal levels. The three types of CAH that may result in abnormal adrenal androgen production and subsequent virilization in prepubertal children are 21-hydroxylase, 3β-hydroxysteroid dehydrogenase, and 11-hydroxylase deficiencies.[55] Typical physical findings in girls may include pubic and axillary hair, body odor, growth acceleration, advanced bone age, and clitoromegaly. However, not all these features may be present at diagnosis. Boys present with similar changes, including increased phallic size and increased muscle mass. Testicular volume, however, is in the prepubertal range, in contrast to that observed in CPP. Rare cases of CAH with adrenal rest tissue represent an exception, in which unilateral or even bilateral enlargement of the testes may be found.[56]

Aromatase Excess Syndrome. Aromatase is the enzyme responsible for the conversion of androgens to estrogen in many tissues, including the gonads and adipocytes. Increased aromatization has been reported to be a cause of familial and prepubertal gynecomastia. Kindreds have been described in whom the aromatase excess syndrome appears to be inherited in an autosomal dominant fashion, resulting in gynecomastia in boys and precocious puberty in girls, accompanied by elevations of estradiol and estrone. Aberrant expression of the *P450* aromatase gene is thought to form the basis of the condition in at least one case, whereas an activating mutation in the aromatase gene has also been reported as a cause of familial aromatase excess.[57]

Abnormalities of G Proteins

McCune–Albright Syndrome. McCune–Albright syndrome (MAS), first described in 1937 by Albright and colleagues,[58] is a rare disorder characterized by the clinical triad of precocious puberty, polyostotic fibrous dysplasia of bone, and café-au-lait spots (see Chapter 68). Molecular genetic investigation led to the discovery that MAS is caused by an activating mutation in $G_s\alpha$, the stimulatory subunit of the G protein involved in intracellular signaling[59] encoded for by the *GNAS1* gene. This somatic mutation, which usually involves substitution of arginine for histidine or cysteine at the 201 position, is believed to occur in early embryogenesis, leading to a mosaic distribution of affected tissues. The result is a phenomenon of unregulated intracellular

TABLE 103.1 Genotype-Phenotype Relationships in Peripheral Precocious Puberty

Disorder	Affected Product	Molecular Genetic Abnormality	Functional Consequence	Inheritance	Clinical Presentation
Congenital adrenal hyperplasia	21-hydroxylase, 11-hydroxylase, 3β-hydroxysteroid dehydrogenase	*CYP21, CYP11B1, HSD3B2*	Impaired cortisol biosynthesis Excess androgen production	Autosomal recessive	Classic form: Ambiguous genitalia in newborn girls Adrenal insufficiency in boys and girls Nonclassic form: Premature pubarche in boys and girls
McCune–Albright syndrome	$G_s\alpha$	Missense mutation in *GNAS1* gene resulting in ARG201 → CYS or HIS	Mosaic distribution of constitutive activation of the *GNAS1* gene including the gonads	Sporadic	Girls: recurrent episodes of vaginal bleeding, premature breast development Boys: precocious puberty Café-au-lait macules, fibrous dysplasia of bone, other hyperfunctioning endocrinopathies
Tissue-specific $G_s\alpha$ mutation	$G_s\alpha$	Point mutation in the *GNAS1* gene resulting in ALA366→ SER	Constitutive activation in testes, inactivation in other tissues	Sporadic	Boys: precocious puberty, pseudohypoparathyroidism
Familial male precocious puberty	Luteinizing hormone receptor	Single amino acid substitution in the LH receptor	Constitutive activation of the LH receptor in Leydig cells	Autosomal dominant, sporadic	Precocious puberty in males
Aromatase excess syndrome	Aromatase enzyme	Polymorphism of *P450* aromatase gene; inversions of chromosome 15q21.2-3	Aberrant transcription of aromatase gene; gain-of-function mutation of aromatase gene	Autosomal dominant, X-linked	Girls: precocious puberty Boys: gynecomastia
FSH-activating mutation	FSH receptor	Single amino acid substitutions in the FSH receptor	Constitutive activation of the FSH receptor. Increased affinity of FSH receptor for human chorionic gonadotropin	Autosomal dominant	Females: familial gestational ovarian hyperstimulation syndrome, ovarian cysts*, premature thelarche* Males: preservation of spermatogenesis after hypophysectomy in one male

FSH, Follicle-stimulating hormone; *LH*, luteinizing hormone.
*, (theoretical)

accumulation of cyclic adenosine monophosphate, which in turn stimulates downstream gene transcription. Hyperfunction of endocrine glands is a cardinal feature of MAS, with numerous organs potentially affected, including the adrenal glands, parathyroid glands, pituitary gland, thyroid gland, and gonads.[60] Rare involvement of nonendocrine tissues has also been reported in MAS and has included hepatobiliary disease, cardiac disease, and sudden death.[61] Unregulated sex steroid secretion from the gonads in prepubertal children forms the basis of precocious puberty in MAS, which, although it occurs in both sexes, is more commonly reported in females. In prepubertal girls, precocious puberty is characterized by sporadic ovarian cysts with subsequent elevations in serum estradiol levels. Spontaneous regression of the cysts can lead to withdrawal bleeding, whereas long-term unopposed estrogen secretion can result in uterine breakthrough bleeding.

The clinical presentation of MAS is extremely variable, ranging from patients with extensive skin (Fig. 103.4) and bone (Fig. 103.5) involvement and progressive precocious puberty to patients with only mild manifestations. Patients exhibiting a forme fruste of MAS are also recognized, in whom isolated involvement of bone, adrenal glands, or gonads may occur. Adolescent girls with MAS tend to have a lower ovulatory rate compared with controls (36% vs. 52%).[62] The variable clinical presentation reflects, in part, the fact that the $G_s\alpha$ mutations that cause MAS are postzygotic and mosaic. Macroorchidism is commonly seen in boys with MAS, and Sertoli cell-only *GNAS1* mutations have been reported as a cause.[63] Thus, different individuals and different organs within the same individual may contain variable proportions of cells that harbor the $G_s\alpha$ mutation. Tissue-specific imprinting of the *GNAS1* gene may contribute to the observed phenotypic heterogeneity.[64] In patients with rapidly progressive precocious puberty, historical treatment has involved medroxyprogesterone or an aromatase inhibitor, such as testolactone. Unfortunately, these therapies have been largely disappointing. Although the newer-generation aromatase inhibitor anastrozole was not shown to be beneficial in a study of 27 girls with MAS,[65] therapy with letrozole appears to be more promising. In a pilot study of nine girls with MAS, a significant decrease in frequency of menstrual bleeding and rate of growth was observed.[66] Long-term outcomes of 28 girls, including the nine girls from the initial pilot study, treated with letrozole for up to 10 years showed a significant improvement in PAH, and no adverse events.[67] Another option is tamoxifen, a selective estrogen receptor modulator, which was shown to be safe and efficacious in 25 girls with MAS and precocious puberty.[68] A retrospective study of eight girls treated with tamoxifen for 3 to 8 years and followed for a mean period of 8.3 years showed cessation of vaginal bleeding and increased PAH, with no noted alteration in uterine volumes or endometrial thickness.[69] However, long-term data are lacking.

A 1-year treatment with Fulvestrant, a pure estrogen receptor antagonist, in 30 girls with PPP related to MAS resulted in a decrease in vaginal bleeding (from 12 days to 1 day per year) and a significant decrease in the rate of skeletal maturation.[70]

Regardless of the pharmacologic agent used, the underlying disease process is unchanged. Once the onset of central puberty commences, the HPG axis may override the independent gonadal function, although persistent autonomous ovarian activity has been documented in postmenarchal girls and women with MAS and a history of precocious puberty. Long-term data suggest that these women have successful spontaneous pregnancies, despite a longer than average time to conception.[71]

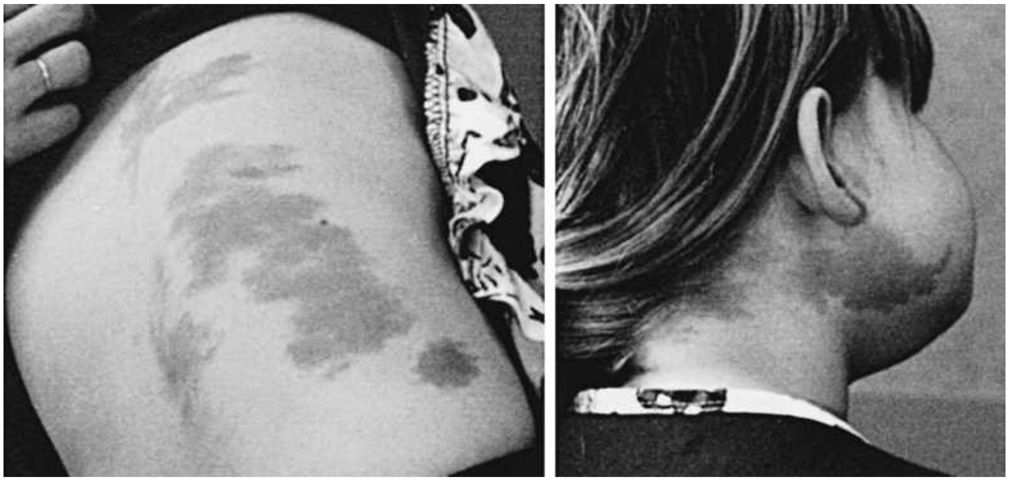


Fig. 103.4 Café-au-lait skin pigmentation in a 4-year-old girl with McCune–Albright syndrome who presented with precocious puberty. The irregular borders of these lesions designate them as "coast of Maine" café-au-lait spots, as opposed to the smooth-bordered "coast of California" variety characteristic of neurofibromatosis. The propensity of these lesions to stop abruptly at the midline is demonstrated.

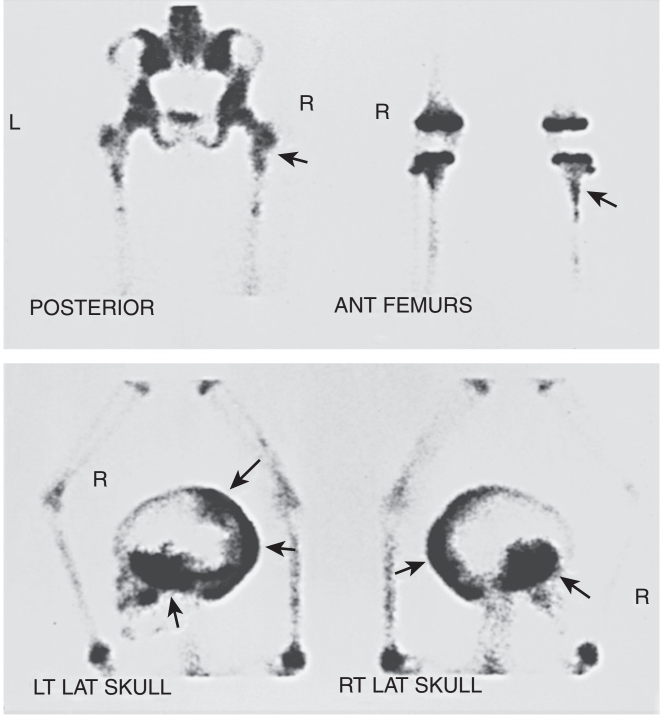


Fig. 103.5 Bone scan of a patient with polyostotic fibrous dysplasia. The *arrows* indicate areas of increased uptake.

Less data are available on the treatment of boys with MAS and PPP, given the rarity of this condition. Similar to the treatment approach in other forms of PPP in boys, the combination of anastrozole and the competitive androgen blocker bicalutamide has been reported and appears to be a promising treatment option.[72]

A rare and fascinating mutation of $G_s\alpha$ has been described in two unrelated males, in which a mutant G protein subunit exhibits temperature dependence resulting in both activating and inactivating properties, depending on the tissue involved. In the testes, which are cooler than the core body temperature, the abnormal $G_s\alpha$ is activated, leading to autonomous sex steroid production and testotoxicosis. In contrast, in skin fibroblasts, which are at 37°C, thermal instability of the abnormal protein is observed, leading to rapid degradation of $G_s\alpha$, along with a subsequent decrease in $G_s\alpha$ activity and pseudohypoparathyroidism.[73]

Abnormalities of Gonadotropin Receptors

Familial Male Precocious Puberty. Familial male precocious puberty (FMPP), also termed *testotoxicosis,* is a form of PPP that occurs only in males. In this disorder, the LH receptor exhibits constitutive activation, leading to autonomous sex steroid secretion and early pubertal development. The underlying abnormality is a heterozygous mutation in the gene that encodes the LH receptor, which results in transcription of an abnormal receptor product with aberrant function. The first described mutation, found in eight affected individuals in different families, involved the substitution of glycine for aspartate at position 578 in the sixth transmembrane helix of the LH receptor. Since then, additional mutations involving the first, second, and third transmembrane domains and the third intracellular loop of the LH receptor have been described.[74] Although the typical inheritance pattern for FMPP is autosomal dominant, patients with sporadic mutations have also been identified. Females with activating mutations of the LH receptor have no phenotypic abnormalities, likely because the actions of both LH and FSH receptors are required for ovarian steroidogenesis.[75]

Patients with FMPP present in early childhood (typically by 4 years of age) with growth acceleration and progressive virilization, including growth of pubic hair and phallic enlargement. The testes are usually small in comparison with the degree of virilization and may be nodular, with histology demonstrating Leydig cell hyperplasia. Serum testosterone levels are in the adult male range, whereas serum gonadotropins are suppressed, as is the response to GnRH stimulation testing. Untreated males with FMPP experience continued skeletal maturation with early epiphyseal closure and final short stature. Fertility is generally normal, although oligospermia and morphologic testicular abnormalities have been described in some affected individuals.[76] Pharmacologic regimens shown to be effective in the treatment of FMPP have included the antisteroidogenic agent ketoconazole, although there are concerns regarding potential liver toxicity. Although effective, combination therapy with spironolactone and the aromatase inhibitor testolactone has been replaced with the combination of the selective androgen receptor blocker bicalutamide and a third-generation aromatase inhibitor because of the need for less-frequent dosing and more selective androgen action blockade. The efficacy of this approach has been reported in several small studies.[77] In one prospective trial involving 14 boys, a decrease in the rate of growth and skeletal maturation was noted after 1 year of treatment with bicalutamide and anastrozole. Gynecomastia and breast tenderness were the most common treatment-related adverse effects.[78] The addition of a long-acting GnRH analog once central puberty begins has provided useful adjunctive therapy and improved adult height in the setting of FMPP.[79] Interestingly, intermittent spontaneous resolution has been reported in a patient with genetically confirmed FMPP.[80]

Ovarian Cysts. Use of high-resolution ultrasonography has revealed that ovarian cysts are common in prepubertal girls of all ages. Although these cysts are typically asymptomatic, they may become transiently functional, causing acute onset of breast development because of elevated ovarian estradiol secretion. Spontaneous resolution of these cysts is followed by a decrease in breast tissue, leading to a history of waxing and waning breast development. The underlying pathophysiology of this cause of precocious puberty is poorly understood and may represent transient activity of the HPG axis, a $G_s\alpha$ mutation (forme fruste of MAS), or an unidentified FSH receptor mutation. Although activating mutations of the FSH receptor have been linked to gestational ovarian hyperstimulation syndrome,[81] no case of precocious puberty resulting from such a mutation has been described thus far.

TUMORS

Sex steroid production by neoplastic tissue in prepubertal children is a rare but important cause of PPP. Presenting symptoms depend on the location and type of tumor and whether androgens, estrogens, or both are being produced.

Gonadal Tumors

Ovarian Tumors. Primary ovarian neoplasms in children may originate from sex cord or stromal tissue, epithelium, or the germ cell line, with granulosa cell tumors being the most frequent ovarian tumor that induces precocious puberty. Although the average age of prepubertal children presenting with ovarian tumors is 10 years, precocious puberty caused by ovarian neoplasms has been reported in infants. Most ovarian tumors produce estrogen, resulting in feminization, although androgen production may also occur, leading to hirsutism and other signs of virilization. Additional clinical findings often include abdominal symptoms such as pain, distention, ascites, or a palpable mass. The frequency of precocious puberty as a feature of ovarian tumors varies, depending on the type of tissue involved, but may be as high as 71%. Primary treatment is surgical with or without adjuvant chemotherapy, and the prognosis is generally good, particularly in patients with the juvenile granulosa cell tumor variety.[82]

Testicular Tumors. Leydig cell tumors, which make up approximately 3% of all testicular neoplasms, are the most common gonadal stromal tumors associated with precocious puberty in boys. Although 10% of such tumors are malignant in adults, they are usually benign in children. Precocious puberty caused by testicular tumors is typically characterized by elevated serum testosterone levels and evidence of virilization. However, sex cord (Sertoli cell) tumors have rarely been described in which gynecomastia, as well as virilization, is a feature of the clinical presentation. In several cases, somatic activating mutations of the LH receptor have been shown to form the basis for Leydig cell tumors associated with precocious puberty.[83]

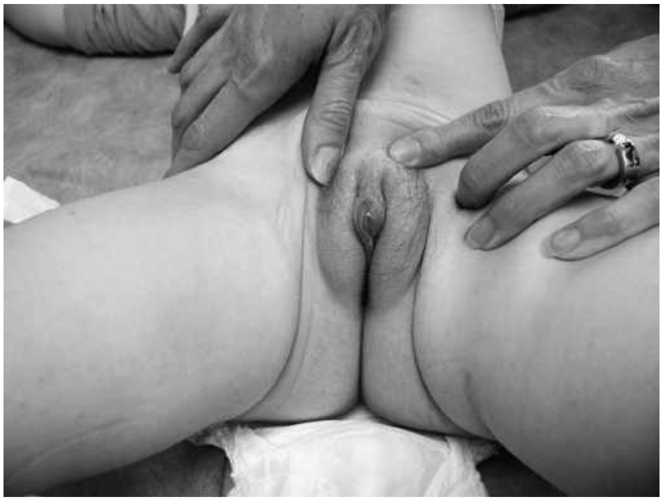


Fig. 103.6 **Virilization of the genitalia in a 2-year-old girl with an androgen-secreting adrenocortical adenoma.** (Courtesy Dr. Emily Walvoord.)

Adrenal Tumors

Adrenocortical adenomas and carcinomas are extremely rare causes of precocious puberty in childhood but have been reported to occur even during infancy. The vast majority of such tumors are virilizing (Fig. 103.6) as the result of elevations in serum adrenal androgen levels, which may or may not be accompanied by hypercortisolism. However, in rare instances, variations in biosynthetic enzyme activity within the tumor may result in increased plasma concentrations of estrogens, with subsequent feminization. Most functioning childhood adrenocortical neoplasms are malignant, and surgical resection is the treatment method of choice. Secondary activation of the HPG axis has been reported after removal of a virilizing adrenal tumor.[84]

Germ Cell Tumors

Nongonadal germ cell tumors as a cause of precocious puberty have been reported to arise in numerous locations, including the liver, lungs, mediastinum, pineal gland, basal ganglia, thalamus, and hypothalamus. Tumor markers useful in the diagnosis and follow-up of germ cell tumors include α-fetoprotein, β-human chorionic gonadotropin (β-hCG), and pregnancy-specific β_1-glycoprotein.[85] Because precocious puberty in this setting occurs as a result of ectopic β-hCG secretion, which mimics the action of LH on testicular Leydig cells, the overwhelming majority of cases are reported in boys. Increased tumor aromatase activity in the setting of simultaneous β-hCG production has been postulated to account for rare cases of precocious feminization in girls presenting with germ cell tumors.

Exogenous Sex Steroids

Exogenous sex steroids—estrogens or androgens—can lead to the development of secondary sexual characteristics in prepubertal children. Examples of widely available exogenous steroids include oral contraceptive pills and estrogen- or androgen-containing topical preparations. Systemic absorption of estrogens found in various cosmetics and hair products has also been linked to precocious puberty.[86]

HYPOTHYROIDISM

More appropriately thought of as a form of pseudoprecocious puberty (PSPP), the rare association between severe long-standing hypothyroidism and early sexual development (Van Wyk–Grumbach syndrome) was described in the 1960s.[87] Although the incidence of this condition has not been systematically evaluated, 24% of 33 children with profound hypothyroidism had evidence of PSPP.[88] Those with PSPP were older and had higher TSH levels. This condition is accompanied by the seemingly paradoxical finding of significantly delayed skeletal maturation and growth arrest. In girls, the clinical presentation consists of estrogenic manifestations, including breast development and changes in the vaginal mucosa. Galactorrhea and vaginal bleeding may be present, and in some patients are severe. Pelvic ultrasonography reveals enlarged ovaries with multiple cysts. Boys present with macroorchidism, often without evidence of virilization (Figs. 103.7 and 103.8), which is consistent with the finding of low serum testosterone levels.

Investigation of the HPG axis of boys with severe primary hypothyroidism and macroorchidism reveals a prepubertal pattern of gonadotropin response to GnRH stimulation, although baseline levels of TSH, prolactin, and pituitary gonadotropins are elevated. The original hypothesis for increased gonadotropins in addition to TSH has been supported by the finding of an elevated FSH response to TSH stimulation in a patient with hypothyroidism-induced macroorchidism,[89] whereas evidence also exists of a component of cross-reactivity of TSH at the level of the FSH receptor.[90] Both of these mechanisms would theoretically result in FSH-mediated gonadal stimulation, causing Sertoli cell proliferation with subsequent testicular enlargement in boys and ovarian hyperstimulation in girls. The lack of LH-induced testosterone secretion in hypothyroid boys with elevated gonadotropins has been attributed to an inhibitory influence of prolactin on LH action at the Leydig cell. Most of the clinical and biochemical abnormalities in these patients gradually resolve with restoration of euthyroid status, although macroorchidism may persist.[89]

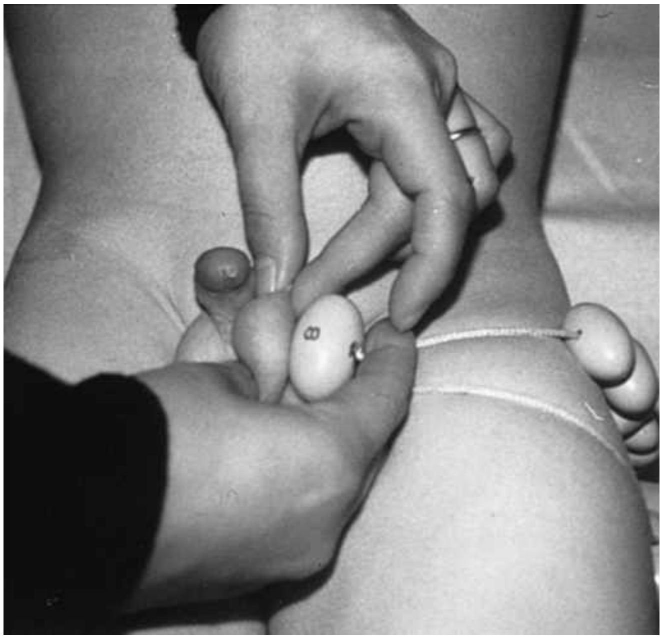


Fig. 103.7 **Macroorchidism in an 8-year-old boy with severe long-standing hypothyroidism secondary to autoimmune thyroiditis.** Testicular volume was 8 mL, in contrast to a normal prepubertal testicular volume of 1–3 mL.

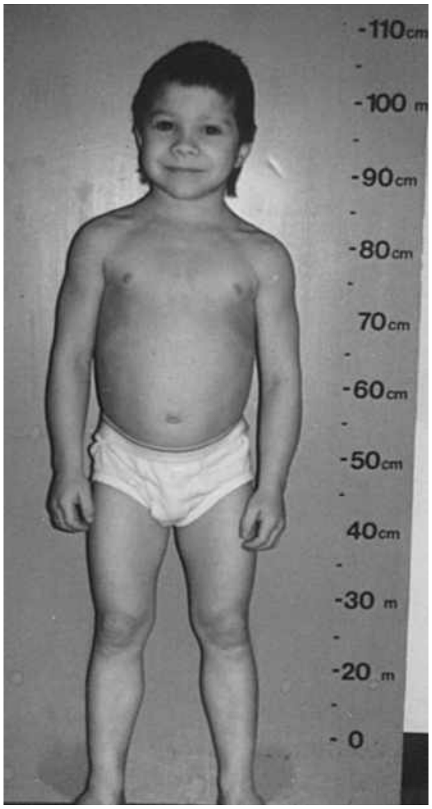


Fig. 103.8 **Additional manifestations of hypothyroidism in this child included severe growth arrest, coarse facial features, dry skin, and sallow complexion.** Bone age radiograph was significantly delayed at 5 years.

KEY POINTS

- There are many etiologies of peripheral precocious puberty (PPP), with diverse manifestations resulting from exposure to androgens, estrogens, or both, independent of activation of the hypothalamic-pituitary-gonadal axis. Most common etiologies include congenital adrenal hyperplasia (CAH), McCune–Albright syndrome (MAS), and familial male precocious puberty (FMPP).
- CAH is the most common cause of abnormal postnatal virilization in children. The excess androgen production results from enzymatic block in the cortisol biosynthetic pathway, most common being 21-hydroxylase deficiency. Glucocorticoid replacement aims to decrease shunting of the enzyme precursors to the androgen pathway.
- The early pubertal development in MAS results from spontaneous ovarian cysts caused by an activating mutation in $G_s\alpha$, a stimulatory subunit involved in intracellular signaling of multiple hormone receptors. Treatment with letrozole, a third-generation aromatase inhibitor, tamoxifen, an estrogen receptor modulator, or Fulvestrant, an estrogen receptor antagonist, appear to decrease the frequency of vaginal bleeding and improve the rate of skeletal maturation.
- FMPP is a male-limited form of PPP that results from activating mutations in the luteinizing hormone receptor leading to autonomous androgen production. Combination therapy with an androgen receptor blocker and an aromatase inhibitor decrease the rate of growth and skeletal maturation and improves growth potential.

DELAYED PUBERTY

Definition

Delayed puberty refers to the absence of pubertal onset by the expected age or, once puberty has commenced, failure of appropriate progression. Although its outward manifestations are subtle, delayed puberty is nonetheless profoundly distressing to the adolescents affected by it, who continually compare themselves to their peers and find themselves lacking. A large array of disorders may manifest as delayed puberty, necessitating a diagnostic process that incorporates patience, astute observation, and consideration of multiple potential etiologies. Discovery of the molecular genetic basis for both benign and pathologic forms of delayed puberty will continue to offer tremendous insight into the biology of human reproduction.

Girls are considered to have delayed puberty if they reach 13 years of age without evidence of pubertal changes; in boys, the analogous age is 14 years. Progression of puberty denotes the development of secondary sexual characteristics according to a predictable and sequential series of stages first described by James Tanner.[91] Whereas a postponement of pubertal onset often represents a benign variation of normal development (discussed in the following section), arrest of puberty once it has begun is almost always abnormal and indicative of an underlying pathologic condition.

BENIGN VARIANTS OF DELAYED PUBERTY

Constitutional Delay of Growth and Puberty

Constitutional delay of growth and puberty (CDGP) refers to a well-described variation of normal growth that is the most common cause of delayed puberty in both boys and girls. Children with this condition are short for their genetic background, often below the third percentile, and have delayed skeletal maturation, as well as delayed puberty. In many cases, there is a family history of "late bloomers," manifested by an older age at menarche in women or significant linear growth occurring in the late teenage years in men. The precise cause of the altered growth pattern observed in CDGP is unknown. Observations of the strong familial nature of the condition have led to investigations of potential inheritance patterns. Evidence thus far suggests the presence of a genetic predisposition with an autosomal dominant effect in addition to environmental modifiers. However, precisely which genes and/or modifiers are responsible for CDGP remains largely unknown. A search for mutations in genes known to cause isolated hypogonadotropic hypogonadism (HH) such as *KISS1R, FGFR1, GNRHR, TAC3,* and *TAC3R* in patients with CDGP, has thus far been negative. Mutations in the *IGSF10* gene have been identified in six unrelated families from a large Finnish cohort. This gene is involved in GnRH neuron migration during embryogenesis.[92] Several inactivating mutations in the growth hormone secretagogue receptor have been described, suggesting a role of ghrelin in the pathogenesis of CDGP.[93] In CDGP, boys and girls appear to be equally affected.[94]

Diagnosis. The diagnosis of CDGP rests on the typical clinical findings of short stature, delayed bone age, and low-normal growth velocity, combined with the absence of any other identifiable cause of short stature and pubertal delay. Baseline gonadotropin levels are low, and the clinical and biochemical features are virtually indistinguishable from those observed in patients with gonadotropin deficiency. This often results in a "wait and see" policy in terms of establishing which patients may ultimately experience spontaneous puberty. However, the ability to differentiate CDGP from pathologic causes is desirable because of anxiety regarding whether or not puberty will ensue. Although no single test is considered foolproof, several different strategies appear useful. One of the earliest employed involves stimulation tests using GnRH analogs such as nafarelin, leuprolide, and triptorelin. Peak LH levels following administration of a GnRH analog are reported to be significantly higher in patients with constitutional delay as compared with patients with HH.[95] High sensitivity and specificity for discrimination of CDGP from other etiologies of delayed puberty in boys has also been reported using human chorionic gonadotropin (hCG) stimulation tests.[96] Another approach relies on measurements of the free α-glycoprotein subunit after GnRH stimulation. Measurement of basal inhibin B also appears to have a good ability to differentiate between CDGP and HH. In a study of 82 boys with pubertal delay, a baseline inhibin B level of 35 pg/mL or less had a 100% sensitivity and specificity to differentiate between HH and CDGP in boys at genital stage 1.[97] The kisspeptin stimulation test shows promise in being able to distinguish youth with CDGP from those with HH, and more study is warranted.[98] Alternatively, failure of spontaneous pubertal onset by a bone age of 12 years for girls and 13 years for boys would clearly be inconsistent with a diagnosis of CDGP, and causes of pathologic hypogonadism should be sought.

Treatment. The practice of using anabolic agents in adolescents with CDGP originated from recognizing the potential adverse psychological consequences of sexual immaturity and from concerns about possible decreased bone mineral density in these patients. Short-term low-dose anabolic therapy has been shown to be both safe and effective in promoting growth and pubertal development without advancing bone age. Specific agents that have been used include fluoxymesterone, oxandrolone, and oral or transdermal testosterone. However, the mainstay of therapy employs low-dose IM testosterone, given monthly on a continuous or intermittent basis for several months and up to 1 year. Another proposed advantage of such therapy is based on the concept that exogenous androgens may advance maturation of the HPG axis, to "jump start" puberty. Despite this hypothesis, there is little evidence that treatment with exogenous testosterone affects the timing and tempo of endogenous pubertal development in boys with constitutional delay in whom considerable heterogeneity in baseline HPG activation has been well documented. Similar therapeutic regimens have been administered in girls with CDGP using conjugated or unconjugated oral estrogens or transdermal 17β-estradiol, with comparable results.[99]

Outcome

Adult Stature. Adult height in patients with CDGP has been examined in relation to predicted height and target height. The majority of studies indicate that individuals with CDGP attain an adult height that is below the target height but within the midparental range,[100] whether or not treatment with short-term anabolic agents is undertaken. Accuracy of predicted heights as determined by the Bayley–Pinneau method is variable. Although most investigators have found that height predictions are fairly reliable,[95] overestimating and underestimating adult heights have also been reported using this method. Disproportionality—when the spine is short in relation to the limbs (*eunuchoid* body habitus)—has been described in CDGP and has been attributed to decreased spinal growth during the characteristically attenuated pubertal growth spurt observed in these patients. A longer period for limb growth is also noted.

Bone Mineralization. The effects of sex steroids on the acquisition of bone mineral density are well known, and hypogonadal states are a cause of osteoporosis in adulthood.[101] Although normal bone mineral density has been reported in boys and men with a history of CDGP, a significant decrease in bone mineral density has also been identified in this population,[102] suggesting a possible negative impact of delayed sex steroid exposure on bone accretion. Treatment with exogenous testosterone during adolescence has been shown to improve bone density and bone turnover, providing additional support for the use of short-term anabolic agents in these patients. The effects of androgens on bone mineralization and epiphyseal fusion are mediated via aromatization to estrogen. This has been demonstrated by rare clinical examples of estrogen receptor mutations and aromatase deficiency in males, conditions that are characterized by profoundly decreased bone mineral density and open epiphyses, despite normal serum testosterone levels.[103,104]

PATHOLOGIC DELAYED PUBERTY

Pathologic causes of delayed puberty have conventionally been grouped into two distinct categories, depending on which portion of the HPG axis is affected. HH refers to conditions in which gonadotropins are low, implying an abnormality at the level of the hypothalamus or pituitary gland, or both, whereas *hypergonadotropic hypogonadism* refers to conditions in which LH and FSH are elevated, indicating a lack of negative feedback at the level of the pituitary gland and localizing the problem to the gonads. This delineation has become less useful with the recognition of genetic mutations involving hypothalamic-pituitary hormones and their receptors, some of which result in hypergonadotropism rather than hypogonadotropism. Therefore, causes of hypogonadism are grouped according to the location within the HPG axis of the affected hormone. The term *hypothalamic-pituitary hypogonadism* is used to designate disorders of hormones (and their receptors) originating from this portion of the axis, whereas *primary hypogonadism* encompasses disorders of the gonads themselves. In both forms of hypogonadism, the end result is a failure of normal ovarian or testicular sex steroid production. Discussed later in this chapter, exogenous sex steroid replacement is therefore the cornerstone of therapy for all types of pathologic delayed puberty, regardless of the cause.

Hypothalamic-Pituitary Hypogonadism

Hypothalamic-pituitary hypogonadism refers to conditions in which hypothalamic GnRH, pituitary gonadotropins, or both, are either deficient or inactive. This includes a diverse group of disorders with varied pathophysiology. Some causes stem from an isolated abnormality in hypothalamic GnRH, and others stem from an isolated abnormality in pituitary gonadotropins (LH and/or FSH), but conditions also exist in which both hypothalamic and pituitary dysfunction are present. Similarly, hypothalamic-pituitary hypogonadism may be the sole manifestation of the disorder, or there may be wide-ranging effects on additional body systems. Partial progression through puberty in a subset of affected individuals is a common feature of many conditions, leading to clinical heterogeneity even within the same kindred. Another shared trait among patients with hypothalamic-pituitary hypogonadism is a characteristic body habitus. An arm span of at least 5 cm more than height and an upper-to-lower segment ratio of less than 0.9 confirm the presence of eunuchoid proportionality. When left untreated, hypothalamic-pituitary hypogonadism is also associated with an increase in adult height owing to the lack of sex steroid-induced skeletal maturation and epiphyseal fusion.[105] Although the exact cause of many cases of hypothalamic-pituitary hypogonadism remains unknown, there has been significant progress in identifying the precise molecular genetic abnormalities responsible for several specific disorders.[106] Single-gene defects are estimated to account for 30% of patients with hypothalamic-pituitary hypogonadism. The list of single-gene defects associated with hypothalamic-pituitary hypogonadism continues to grow and is summarized in Table 103.2, along

TABLE 103.2 Genetic Etiologies of Hypogonadotropic Hypogonadism

Category	Affected Genes	Additional Findings
Disorders of migration of the GnRH neuron (Kallmann syndrome)	*ANOS* (*KAL-1*)	Anosmia/hyposmia, synkinesia, hearing loss, renal agenesis
	FGFR1, FGF8	Dental agenesis, digital bony abnormalities, cleft palate, agenesis of corpus callosum
	PROK2/PROKR2	Fibrous dysplasia, synkinesia, and seizures
	CHD7	Features of CHARGE syndrome
	Other identified genes: *FGF17, WDR11, SEMA3A, IGSF10, SMCHd1, CCDC141, FEZF1*	
Disorders of the GnRH pulse generator	*TAC3/TACR3*	Partial or complete reversal of HH may occur
	KISS1/KISS1R	HH can be reversible
	GNRH1	Cryptorchidism, micropenis
		Phenotype can be reversed by pulsatile GnRH administration
Disorders of pituitary gonadotropes	GnRH receptor	Partial HH
	LH β subunit	M: delayed puberty
		F: homozygous females have oligomenorrhea heterozygous females are unaffected
	FSH β subunit	M: azoospermia, low testicular volume, normal testosterone
		F: ovarian dysgenesis, premature ovarian failure
	DAX-1	Congenital adrenal hypoplasia
	SF-1	XY sex reversal
		Adrenal insufficiency
	Transcription factor defects:	
	PROP1, LHX3	Combined anterior pituitary hormone deficiencies
	HESX1	Septooptic dysplasia, variable anterior pituitary hormone deficiencies
Disorders associated with obesity	Leptin (*LEP*), leptin receptor (*LEPR*)	Hyperphagia, childhood-onset morbid obesity
		Treatment with leptin reverses HH
	Prohormone convertase (PC1)	Childhood morbid obesity, hypocortisolism, elevated proinsulin
Disorders associated with neurodegenerative disease	*POLR3A/POLR3B* (4H syndrome)	Hypodontia, hypomyelination
	RNF216, PNPLA6 (Gordon–Holmes syndrome)	Cerebellar ataxia, early dementia

CHARGE, Coloboma, heart disease, choanal atresia, retarded growth and development, and/or deafness; *F*, females; *FSH*, follicle-stimulating hormone; *GnRH*, gonadotropin-releasing hormone; *LH*, luteinizing hormone; *HH*, hypogonadotropic hypogonadism; *M*, males.

with descriptions of the associated phenotypes. Once a diagnosis of hypothalamic-pituitary hypogonadism has been made, MRI is recommended to rule out an intracranial lesion. Ultrasensitive gonadotropin assays or a GnRH analog stimulation test may be useful in confirming and determining the degree of hypogonadotropism. Hypothalamic GnRH deficiency may be differentiated from pituitary gonadotropin deficiency by administration of pulsatile GnRH, which usually results in subsequent elevations of LH and FSH in individuals with isolated hypothalamic defects. Fertility is therefore potentially possible with the use of exogenous GnRH or gonadotropin therapy, although sex steroid replacement is generally the therapy of choice for induction of puberty and maintenance of secondary sexual characteristics. Interestingly, spontaneous reversal of hypothalamic-pituitary hypogonadism has been noted after discontinuation of sex steroid replacement in approximately 10% of patients with either Kallmann syndrome (KS) or normosmic HH.[107] This suggests that genetic defects in some patients with hypothalamic-pituitary hypogonadism may prolong the time to maturation of the HPG axis. Specific causes of hypothalamic-pituitary hypogonadism are discussed in the following sections.

Kallmann Syndrome. The prototype of GnRH deficiency exists in the form of KS, which refers to the association between HH and anosmia, first described in 1944.[108] Sporadic cases and multiple heritable forms have been described, including X-linked, autosomal recessive, and autosomal dominant. The X-linked form of the disease is caused by mutations in the *KAL* gene on Xp22.3. This gene encodes for the extracellular matrix glycoprotein anosmin-1, which is a cell-adhesion molecule essential for normal neuronal migration during embryogenesis. Highly conserved across species, anosmin-1 appears to have both direct and permissive roles in olfactory bulb development, GnRH migration, and axonal growth.[109] Numerous mutations within the *KAL* gene have been described in patients with KS. Interestingly, significant clinical heterogeneity in the degree of hyposmia/anosmia and hypogonadism exists among patients bearing identical mutations, implying the presence of important modifier genes and/or environmental factors in the expression of the syndrome. The list of specific genes underlying the other forms of KS and normosmic HH is expanding.[110] Loss-of-function mutations in *FGFR1* and mutations in the prokineticin-2 signaling and in nasal embryonic LH-releasing hormone factor genes have been reported in association with the autosomal dominant form of KS. Of these, loss-of-function mutations in *FGFR1* are particularly common and have been detected in 7% of patients with normosmic HH. These mutations likely act in concert with other genetic and environmental factors to explain the variable phenotype in these patients. Autosomal recessive mutations in *TAC3,* the gene encoding neurokinin B, as well as the gene encoding the neurokinin B G protein–bound receptor, have also been commonly reported in patients with idiopathic HH. Although its mechanism of action is unclear, neurokinin B is expressed in the same neurons that express kisspeptin and is thus thought to stimulate GnRH.

Given that HH and anosmia also occur in CHARGE (coloboma, heart disease, choanal atresia, retarded growth and development, and/or deafness) syndrome, mutations in the *CHD7* gene, the putative causative gene in CHARGE, have been investigated in KS and noted in a minority of patients who also tended to have additional features of the syndrome.

It is not surprising that the presence of anosmia in individuals with CHARGE syndrome predicts HH. A wide range of associated congenital anomalies have been reported in patients with KS, including cleft palate, hearing loss, color blindness, abnormal eye movements, unilateral renal agenesis, and synkinesia, although the relationship between these features and specific abnormalities of the *KAL* gene are unclear.

Gonadotropin-Releasing Hormone and Gonadotropin-Releasing Hormone Receptor Mutations. *GnRH*[111] and GnRH receptor gene mutations have been identified as a cause of familial isolated hypothalamic-pituitary hypogonadism.[112] These disorders affect both males and females, and olfaction is normal. Both homozygous and compound heterozygous cases occur. In those with biallelic mutations, the clinical phenotype is determined by the genetic burden defined by the more severe of the two mutations, unlike in those with a single mutation in whom this correlation does not hold, suggesting a role for genetic or environmental modifiers. As in KS, pulsatile GnRH administration may restore normal gonadal function and fertility in some patients with partially inactivating GnRH receptor mutations.

Another etiology for isolated hypothalamic-pituitary hypogonadism has been traced to autosomal recessive mutations in the *KISS1* and *KISS1R* genes. The *KISS1R* gene encodes an orphan G protein-coupled receptor that is a member of the rhodopsin family of receptors.[113] Mutations in the *KISS1R* gene, which can result in both absent and delayed puberty, are a less common cause of hypothalamic-pituitary hypogonadism than GnRH receptor mutations. Several consanguineous families have been identified, as well as one unrelated individual found to be a compound heterozygote for *KISS1R* mutations. As is the case with other causes of HH, fertility can be achieved with exogenous GnRH in men and women with *KISS1R* mutations.

The ligand of *KISS1R* is kisspeptin, a key regulator of hypothalamic GnRH release and pubertal onset. Thus, mutations in the *KISS1* gene also result in normosmic HH.[114]

Mutations in Pituitary Transcription Factors and Orphan Nuclear Receptors. Normal embryologic hypothalamic-pituitary development relies on a tightly regulated, complex cascade of tissue-, cell-, and stage-specific transcription factors. Mutations within the genes encoding for these transcription factors, therefore, may result in a spectrum of combined anterior pituitary hormone deficiencies and abnormal function of the hypothalamic-pituitary axis. While the etiology of many forms of hypopituitarism remains unknown, molecular genetic abnormalities in several pituitary transcription factors have now been elucidated.

Pit-1. The first identified transcription factor, pituitary-specific transcription factor (Pit-1), is essential for normal activation of the genes encoding for growth hormone (GH), prolactin, and TSH, as well as for normal differentiation and proliferation of the pituitary cell types responsible for the expression of these hormones.[115] Following its discovery in 1988, numerous additional transcription factors have been identified, some of which are also involved in normal function of hypothalamic GnRH and pituitary gonadotrophs.[116] Orphan nuclear receptors that regulate development within multiple levels of the HPG axis are also associated with hypothalamic-pituitary hypogonadism. Human genetic mutations that have been identified in these factors are discussed in the following sections.

Prop-1. Prop-1 (prophet of Pit-1) is a paired-like homeodomain transcription factor that is required for normal expression of Pit-1. *PROP1* mutations in humans are associated with GH, prolactin, and TSH deficiency in addition to LH and FSH deficiencies.[117] Pubertal progression has been quite variable in reported cases, with several individuals completing pubertal maturation before the development of permanent hypogonadism by their early twenties.

LHX3. Another etiology of combined pituitary hormone deficiency has been traced to mutations within *LHX3,* a pituitary transcription factor gene located on chromosome 9. Although a limited number of cases have been identified, all affected individuals have had gonadotropin deficiency and unresponsiveness to exogenous GnRH.[118]

LHX4. Heterozygous *LHX4* mutations are a rare cause of combined pituitary hormone deficiency and are associated with variable degrees of LH and FSH deficiencies.[119]

HESX1. Several different mutations within the transcription factor *HESX1* gene have been implicated in septooptic dysplasia associated with variable pituitary function. Although some patients exhibit apparently normal pituitary hormone secretion, a continuum ranging from isolated GH deficiency to panhypopituitarism has been described.[120] Studies from the mouse model have indicated that *HESX1* has an important role in embryologic development of the optic nerves and the hypothalamic-pituitary axis.

SOX2. Several heterozygous mutations have been described within the transcription factor *SOX2* gene. In addition to anophthalmia or severe microphthalmia, affected individuals have gonadotropin deficiency. Other associated findings include developmental delay, defect of the corpus callosum, esophageal atresia, and sensorineural hearing loss.[121]

DAX1. HH is a feature of X-linked adrenal hypoplasia congenita. The gene responsible for this condition, *DAX1,* encodes an orphan nuclear receptor that is expressed in adrenal, pituitary, hypothalamic, and gonadal tissue.[122] More than 100 different mutations within *DAX1* have been identified. Males with the disorder typically present with an adrenal crisis during infancy, whereas the hypothalamic-pituitary hypogonadism emerges in adolescence. However, there appears to be significant clinical variability among individuals with adrenal hypoplasia congenita. Congenital gonadotropin deficiency has been suggested by cryptorchidism in some newborns with *DAX1* mutations, while a normal minipuberty of infancy has also been reported in affected boys.[123] Atypical presentations have included hypothalamic-pituitary hypogonadism in a female with a homozygous *DAX1* mutation, delayed puberty in females with heterozygous mutations, and the delayed onset of adrenal insufficiency and hypogonadism in adolescence and adulthood in a few affected individuals.[124] Studies from the *Ahch* (*Dax1*) knockout mouse suggest that *DAX1* has a critical role in testis development and spermatogenesis, as well as in the function of the HPG axis. This finding, as well as recognition of the combined hypothalamic and pituitary defect in these patients, has been proposed to explain the disappointing results using exogenous GnRH or gonadotropins to induce virilization and fertility.[125]

SF1. Like *DAX1*, steroidogenic factor 1 *(SF1)* encodes an orphan nuclear receptor that regulates gene transcription in multiple tissues including the gonads, the adrenal glands, the hypothalamus, and the pituitary gland. The importance of *SF1* for normal gonadotrope function was first established by the knockout mouse, which exhibits gonadotropin deficiency and hypoplastic gonads. Human *SF1* mutations have been identified as a cause of 46,XY sex reversal and adrenal failure.[126] Some *SF1* mutations cause testicular dysgenesis without affecting adrenal function. Ovarian development appears to be normal thus far in a prepubertal girl with a heterozygous *SF1* mutation and adrenal insufficiency.

Mutations in Gonadotropins and Their Receptors. FSH and LH are composed of a common α subunit and a distinct β subunit and exert their biological effects through G protein-coupled receptors. No mutations of the α subunits have been described to date, but mutations of the FSH-β subunit, LH-β subunit, and gonadotropin receptor have been identified as causes of delayed puberty and abnormal reproductive function in males and females.[127] These mutations continue to provide important insight into the relative roles of FSH and LH in gonadal function and human physiology.

Follicle-Stimulating Hormone–β Subunit Mutations. Mutations in the FSH-β subunit gene have been reported to cause hypogonadism in both females and males.[128] The majority of cases have occurred in females with complete absence of pubertal development, although partial breast development in affected girls has been reported. The phenotype of males with *FSH-β* gene mutations ranges from delayed puberty and azoospermia to azoospermia alone.[129] As expected, serum LH levels are high in these patients, and FSH is undetectable.

Follicle-Stimulating Hormone Receptor Mutations. Homozygous inactivating mutations in the FSH receptor were first identified as the cause of autosomal recessive premature ovarian failure in several Finnish families.[130] The females in these kindreds demonstrated uneven degrees of pubertal development, followed by complete ovarian failure in association with elevated serum gonadotropins and hypoplastic ovaries. Subsequently, a number of novel FSH receptor mutations have been described in association with primary or secondary amenorrhea. In contrast, men with homozygous mutations have normal LH and testosterone levels and low-normal testicular volumes, with elevated FSH levels and unpredictable degrees of impaired spermatogenesis and infertility.[131]

Luteinizing Hormone–β Subunit Mutation. One case of delayed puberty because of an LH gene mutation has been described. The 17-year-old male proband was homozygous for a single base substitution in the LH-β subunit gene.[132] Serum LH was elevated, while FSH and testosterone levels were low. The finding of LH immunoreactivity with complete loss of LH bioactivity in this patient was consistent with the clinical phenotype. Three maternal uncles, all heterozygous for the same mutation, were infertile. Although female carriers appeared unaffected in this family, anovulatory infertility was reported in a woman with a novel mutation of the LH-β subunit.[133] Two novel mutations of the *LH* gene were described in a 31-year-old man with delayed puberty, azoospermia, and absent circulating LH. His sister, who carried the same mutations, had normal pubertal development and oligomenorrhea, indicating variable genetic expression in this condition.[134]

Luteinizing Hormone Receptor Mutations. Homozygous loss-of-function mutations of the LH receptor, inherited via autosomal recessive transmission, have been recognized as a rare cause of hypogonadism in males and females. The first such mutation was reported in two 46,XY individuals with female external genitalia but complete absence of breast development and primary amenorrhea.[135] Serum levels of LH were markedly elevated, while testosterone levels were low, and FSH levels were normal. Histologic examination of the testes revealed Leydig cell hypoplasia. Since then, multiple additional mutations of the LH receptor have been identified. Males with inactivating LH receptor mutations exhibit a clinical spectrum ranging from female external genitalia to hypospadias or micropenis. The testes are usually cryptorchid, and Müllerian structures are absent. In females, LH receptor mutations result in amenorrhea with normal secondary sexual development and elevations of LH and FSH. The finding of ovarian follicular development at all stages suggests that LH may be necessary for ovulation, but that follicular development may occur under the influence of FSH and ovarian paracrine factors alone.

A number of other gene defects associated with HH and neurological manifestations have been described, including defects in the *POLR3A/B* gene associated with 4H (hypomyelination, hypodontia and hypogonadotropic hypogonadism) syndrome, and defects in the *RNF216*, *OTUD4*, and *PNPLA6* genes causing ataxia and HH.[136,137]

Estrogen Receptor Defects. Some 19 years after it was discovered in a man, a homozygous mutation in the estrogen receptor alpha subunit causing estrogen resistance was reported in an 18-year-old woman. This specific mutation impaired estrogen signaling. The patient had no breast development and had significantly elevated estrogen levels and bilateral multicystic ovaries.[138] She had osteopenia and a delayed bone age (13.5 years) and no response to estrogen therapy.

Metabolism-Related Genetic Defects. Naturally occurring genetic mutations associated with obesity have indicated important interactions between metabolism and human reproduction. Hypothalamic-pituitary hypogonadism has been a feature of rare human mutations within the leptin gene and leptin receptor.[139] A similar phenotype was observed in a girl with primary amenorrhea and hypothalamic-pituitary hypogonadism in association with compound heterozygous mutations in prohormone convertase 1, an endopeptidase involved in neuropeptide processing. The precise role of these genes in GnRH release and/or function is unknown.

Chronic Disease. Any significant chronic disease or inflammation can result in delayed puberty, which is usually seen in association with poor growth and retarded skeletal maturation. Malnutrition also has a profound impact on gonadal function through loss of normal hypothalamic-pituitary pulsatile secretion. This is exemplified by the hypothalamic amenorrhea often observed in young women with anorexia nervosa, in whom chronic caloric restriction leads to aberrations in neuroendocrine function. The same phenomenon is observed in individuals engaging in frequent, high-intensity exercise such as runners or ballet dancers, in whom osteopenia is a common correlate. Hypothalamic-pituitary hypogonadism has also been documented in women of normal weight who practice severe restriction of dietary fat intake. Loss of GnRH pulsatility in these settings is likely the result of leptin deficiency, a hypothesis supported by the restoration of ovulatory cycles following human recombinant leptin administration in women with anorexia nervosa. These forms of hypogonadism are transient and characterized by a restoration of normal HPG function once the underlying abnormality is resolved.[140] In contrast, permanent hypothalamic-pituitary hypogonadism may result from acquired central nervous system insults such as infection, cranial irradiation, tumor, or trauma.

Syndromes. Hypothalamic-pituitary hypogonadism is a common feature of a number of different genetic syndromes with otherwise heterogeneous manifestations. For example, individuals with Prader–Willi syndrome often have hypogonadism secondary to hypothalamic-pituitary dysfunction.[141] Other syndromes that share this characteristic include Boucher–Neuhauser, Bardet–Biedl, and Noonan syndromes. Associated abnormalities often include cryptorchidism, micropenis, or both. Hypothalamic-pituitary hypogonadism also occurs in patients with pseudohypoparathyroidism type 1A in association with multiple hormone resistance from a loss-of-function mutation in the *GNAS1* gene.[142]

Primary Hypogonadism

Primary hypogonadism refers to conditions in which there is failure of normal sex steroid production because of an intrinsic abnormality of the gonads. Using traditional assays, individual gonadotropin levels in prepubertal children with primary hypogonadism may be indistinguishable from normal controls, with the development of elevated gonadotropins by the time pubertal age is reached. Using ultrasensitive methods, supranormal LH and FSH concentrations are detectable in prepubertal agonadal children of all ages compared with controls.[143] In contrast, serum inhibin B levels tend to be inversely proportional to gonadotropins in boys with primary gonadal failure.[144] The causes of primary hypogonadism include congenital and acquired gonadal abnormalities, as discussed in the next section.

Congenital Primary Hypogonadism

Chromosomal Abnormalities. It has been recognized for many years that normal ovarian activity depends on the presence and function of specific X chromosomal genetic loci. Therefore, abnormalities involving any significant portion of the X chromosome in females are associated with aberrant ovarian function. Varying degrees of gonadal dysregulation are also observed in males with an abnormal complement of X chromosomes. The most common karyotypic abnormalities leading to primary hypogonadism are discussed in the following sections.

Turner Syndrome. With an incidence of 1 in 2000 to 5000 liveborn females, Turner syndrome is the most common cause of primary ovarian failure. First described in 1938 by Dr. Henry Turner,[145] the syndrome consists of a missing or structurally abnormal X chromosome combined with characteristic phenotypic features. Although a wide range of physical characteristics may manifest, nearly 100% of affected individuals have short stature. Gonadal dysgenesis, as evidenced by "streak ovaries," is present in approximately 94% of patients and appears to occur as a consequence of accelerated oocyte atresia, with subsequent fibrosis of ovarian stroma. The multisystem involvement is demonstrated by the frequent occurrence of cardiac, renal, lymphatic, and musculoskeletal anomalies, while common phenotypic features include low-set, posteriorly rotated ears, high-arched palate, low posterior hairline, webbed neck, the appearance of widely spaced nipples, cubitus valgus, and hypoplastic, hyperconvex fingernails and toenails (Figs. 103.9, 103.10, and 103.11). Although many of these somatic characteristics may be the consequence of intrauterine lymphedema, the Turner syndrome phenotype in general is thought to result from a haploid dosage of genes on the X chromosome that normally escape X inactivation. Girls may come to medical attention in several ways. Lymphedema noted at birth may prompt an early diagnosis. Alternatively, girls may present in midchildhood or adolescence with progressive short stature or delayed puberty. Cytogenetic analysis reveals a heterogeneous group of X chromosomal abnormalities, with approximately 50% of patients exhibiting some form of mosaicism. The diagnosis of ovarian failure is based on the finding of elevated gonadotropins. These follow a diphasic pattern, with plasma LH and FSH levels rising during early and late childhood and exhibiting a nadir during middle years, although increased for age. FSH levels in young girls with Turner syndrome who have mosaic karyotypes are significantly lower than those in girls with monosomy X. Although the vast majority of patients will require exogenous

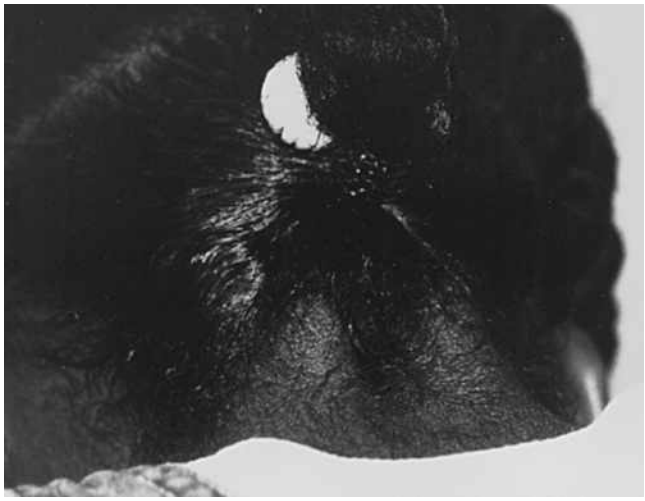


Fig. 103.9 Girls with Turner syndrome often have a low posterior hairline, which typically has a "tripartite" (double arch) configuration.

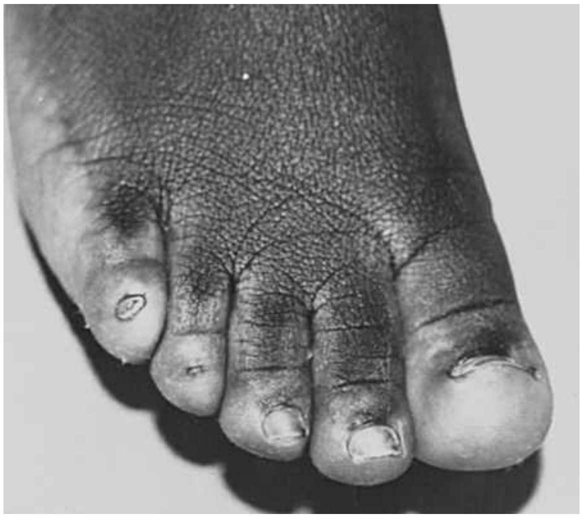


Fig. 103.10 Hypoplasia of the fourth and fifth toenails in a patient with Turner syndrome.

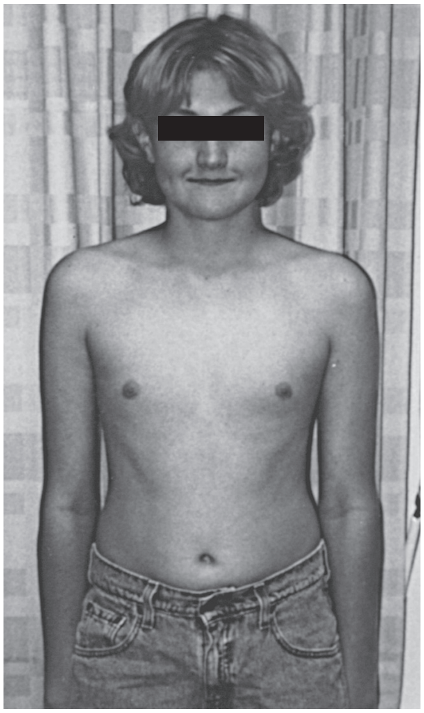
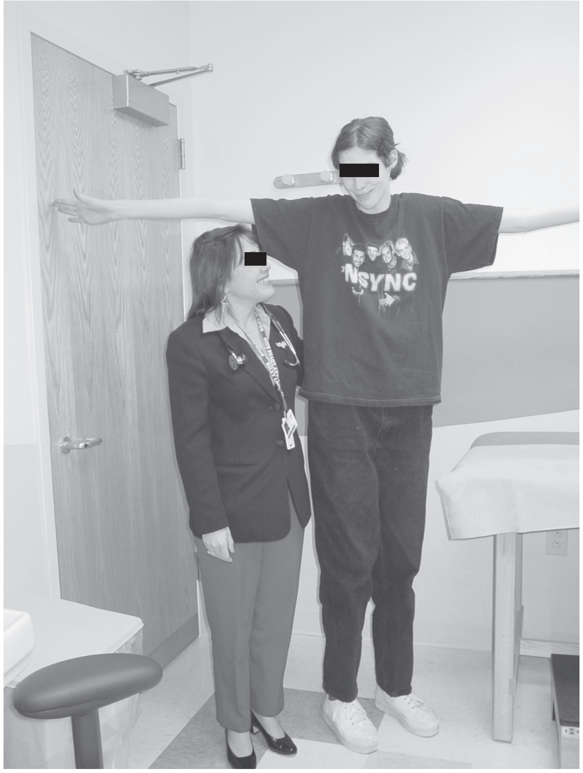

[redacted]

Fig.103.11 Complete absence of breast development in a 15-year-old girl presenting with delayed puberty. Arm span was 8.3 cm greater than height, and upper-to-lower body segment ratio was 0.7, confirming a "eunuchoid habitus." Karyotype was 46,XY.

sex-steroid replacement, approximately 30% of girls will experience spontaneous pubertal onset, and up to 16% percent will achieve menarche unassisted.[146] Rarely, women with Turner syndrome are able to conceive, although pregnancies in this population are complicated by a high incidence of fetal loss and congenital anomalies.[147]

Xq and Other Chromosomal Abnormalities. Patients with deletions or rearrangements of the long arm of the X chromosome typically have premature ovarian failure without phenotypic features of Turner syndrome. Molecular genetic analysis has led to the localization within the distal region of Xq of one or more genes essential for ovarian maintenance.[148] Many cases of sporadic X chromosome:autosomal translocations are also associated with premature ovarian failure, as are mutations within a forkhead transcription gene located on chromosome 3. Triple- and penta-X syndromes represent other examples of X chromosome aneuploidy in which premature ovarian failure is a frequent correlate (Fig. 103.12).



[redacted]
[redacted]

Fig. 103.12 Tall stature with increased arm span and primary ovarian failure in a 16-year-old girl with a 48,XXXX karyotype.

Klinefelter Syndrome. First described in 1942,[149,150] Klinefelter syndrome results from X chromosome polysomy in males, with the majority of patients demonstrating a 47,XXY karyotype. It is the most common disorder of the sex chromosomes, with an estimated incidence ranging from 1 in 400 to 1 in 1000 live-born males. Typical features include tall stature with eunuchoid body proportions, hypogonadism, a variety of learning and behavioral problems, and borderline-to-normal intelligence. The testes, which are usually small and firm and may be cryptorchid, are characterized by seminiferous tubule dysgenesis and a spectrum of Leydig cell dysfunction. Although pubertal onset may occur normally, serum testosterone levels are generally at the low end of the normal range, and inhibin B levels gradually decrease in conjunction with progressive testicular failure. Gynecomastia, because of an abnormal estrogen/androgen ratio, and elevated serum gonadotropins are important clues to the presence of Klinefelter syndrome in a boy at pubertal age. However, many patients remain undiagnosed through adolescence and may present in adulthood with infertility, azoospermia, or hypogonadism.[150] A propensity for the development of certain malignancies has long been recognized in individuals with Klinefelter syndrome. Breast cancer, historically believed to occur at increased frequency, has not been observed in other large series, whereas germ cell tumors have been found to be 50 times more common in this population and have been implicated in rare cases of precocious puberty in boys with Klinefelter syndrome.[151]

Enzymatic Defects in Steroid Biosynthesis. Although the most common forms of CAH (see Chapter 94) are associated with an excess of adrenal androgen secretion, rare defects exist in which there may be a complete inability to synthesize both androgens and estrogens. These

TABLE 103.3 Clinical, Laboratory, and Radiographic Data in a Patient With Pure 46,XY Gonadal Dysgenesis

Clinical Characteristics	Laboratory Data	Radiographic Studies
• Presenting complaint: primary amenorrhea • Chronologic age: 15 years • Height: 162.7 cm (50th percentile on female growth chart, 25th percentile on male growth chart) • Arm span: 171 cm • Upper/lower ratio: 0.7 • Tanner staging: Breasts: I Pubic hair: II • External genitalia: normal female	• LH: 15.8 mIU/mL (1.0–18) • FSH: 64.7 mIU/mL (4–13) • Estradiol: 13 pg/mL (10–178) • Thyroxine: 7.6 μg/dL (4.5–12) • TSH: 3.9 μIU/mL (0.45–6.0) • Prolactin: 10.7 ng/mL (5.3–22) • Karyotype: 46,XY • MIS: 0 ng/mL	• Bone age x-ray: 11.6/12 years • Pelvic ultrasound: small uterus, no visible gonads • Genitogram: normal vagina with cervical impression

LH, Luteinizing hormone; *FSH*, follicle-stimulating hormone; *TSH*, thyroid-stimulating hormone; *MIS*, Müllerian-inhibiting substance.

include deficiencies of 17α-hydroxylase and 17,20-lyase enzymes, as well as a deficiency of the steroidogenic acute regulatory protein, which plays a key role in cholesterol transport, the rate-limiting step of steroid hormone biosynthesis.[152] Lipoid adrenal hyperplasia results from mutations within the steroidogenic acute regulatory gene, with many different mutations identified to date.[153] Clinical manifestations consist of severe electrolyte imbalance usually beginning in infancy, as well as complete deficiency of adrenal and testicular steroidogenesis, resulting in phenotypically female genitalia in both sexes. A fascinating aspect of lipoid adrenal hyperplasia is that some affected 46,XX individuals experience spontaneous puberty and normal menstrual cycles, presumably owing to steroidogenic acute regulatory-independent steroidogenesis within the ovary. Mutations in the *CYP17* gene (responsible for both 17α-hydroxylase and 17,20-lyase activity) is another rare cause of delayed puberty in males and females. Both sexes appear phenotypically female, with affected males demonstrating absence of Müllerian structures and intraabdominal testes. Diagnosis is based on extreme elevations of adrenal steroid precursors, and treatment consists of glucocorticoid and sex steroid replacement as indicated by the phenotypic sex. Genetic males with severe 17α-hydroxylase deficiencies are, therefore, typically raised as females, using estrogen therapy to induce breast development at puberty once the diagnosis is made.

Disorders of Sex Development. Abnormalities of sexual determination and differentiation result in a number of conditions typified by either overvirilization of genetic females or undervirilization of genetic males, and are usually associated with genital ambiguity (see Chapter 101). However, rare disorders exist in which a phenotypically normal appearance is present, which may result in delayed diagnosis until abnormal pubertal development prompts medical evaluation. Discussion of these disorders will be limited to genetic males, who may appear phenotypically female, as discussed in the following sections.

Complete Androgen Insensitivity. A variety of mutations within the androgen receptor gene may result in qualitative or quantitative abnormalities in androgen receptor function.[154] Although partial androgen insensitivity is a cause of undervirilization in a male infant, complete androgen insensitivity results in an apparently normal female phenotype in a 46,XY individual. These patients are typically tall and present in adolescence with normal breast development, primary amenorrhea, and a paucity of pubic and axillary hair. The external genitalia are those of a normal female, although the vagina ends blindly. The testes are intraabdominal, and no female internal reproductive organs are found. As in XY gonadal dysgenesis, women with complete androgen insensitivity are at risk for gonadal neoplasms, although the incidence appears to be low in this setting. Therefore, most clinicians advocate leaving the gonads in place until pubertal development is complete, after which time gonadectomy is recommended, and estrogen replacement is necessary to maintain secondary sexual characteristics.

Pure 46,XY Gonadal Dysgenesis (Swyer Syndrome). First described in 1955,[155,156] pure 46,XY gonadal dysgenesis manifests as a normal female phenotype with complete absence of secondary sexual development (see Fig. 103.11; Table 103.3). Stature is normal to tall, and there are no stigmata of Turner syndrome. A uterus and fallopian tubes are usually present, and the gonads are characterized by fibrotic "streaks." Heritable forms of pure 46,XY gonadal dysgenesis exist, with postulated modes of transmission including X-linked and autosomal inheritance patterns. Although the causes of gonadal dysgenesis are largely unknown, approximately 15% of patients harbor mutations within *SRY*. Additional etiologies implicated in pure XY sex reversal include deletions or mutations in genes involved in male sex determination, including *SOX9, SF1, WT1,* and *DAX1*.[156] Regardless of the precise genetic cause, patients with pure 46,XY gonadal dysgenesis require prophylactic gonadectomy because of an estimated 20% to 30% incidence of gonadoblastoma or dysgerminoma in this population.[157]

Testicular Regression (Vanishing Testes) Syndrome. *Testicular regression syndrome* refers to a rare condition in which the testes become nonfunctional and involute (thus the term *vanishing*) at some point during intrauterine life or the early postnatal period. Depending on the timing of testicular regression, ambiguous or normally formed male genitalia may result. The cause of testicular regression is poorly understood but is hypothesized to be the result of a vascular insult to the testes during intrauterine life. Familial cases have rarely been reported. Males may go undiagnosed until adolescence or even adulthood. An hCG stimulation test is useful in the evaluation of Leydig cell function, while Müllerian-inhibiting substance (antimüllerian hormone) levels, secreted by Sertoli cells, provide additional information regarding the presence and integrity of testicular tissue. There is controversy whether inguinal exploration of these patients is always mandatory, as the incidence of residual gonadal tissue is extremely low.[158]

Acquired Primary Hypogonadism. There are several causes of acquired primary hypogonadism, all stemming from significant insult or injury to the gonads. Although trauma and torsion represent rare causes of acquired primary hypogonadism, improved survival rates of childhood cancer have resulted in a rise in cases of primary hypogonadism occurring as sequelae of irradiation, chemotherapy, and bone marrow transplantation. Autoimmune-related gonadal failure is a feature of type I and type II polyglandular autoimmune syndromes and may have an insidious onset many years after the development of adrenal insufficiency. Conversely, ovarian failure may be the first presenting sign of these disorders, and, as such, girls with unexplained premature ovarian failure should be screened for other autoimmune disorders. Metabolic diseases associated with acquired gonadal failure include galactosemia, where premature ovarian failure may occur in up to 60% of affected females and result in complete or partial lack of pubertal development.[159]

KEY POINTS

- Constitutional delay of growth and puberty is a well-described variation of normal pubertal development that is the most common cause of delayed puberty in both boys and girls. Evidence suggests genetic predisposition with an autosomal dominant effect in addition to environmental modifiers.
- Short-term low-dose anabolic therapy in boys and estrogens in girls is effective in promoting growth and pubertal development without advancing bone age.
- Pathologic causes of delayed puberty encompass two categories: conditions associated with hypothalamic-pituitary hypogonadism and those associated with hypergonadotropic or primary hypogonadism.
- Single-gene defects are estimated to account for 30% of patients with hypothalamic-pituitary hypogonadism.
- Kallmann syndrome is the prototype of gonadotropin-releasing hormone deficiency and refers to the association between hypogonadotropic hypogonadism and anosmia. Sporadic cases and multiple heritable forms have been described, including the X-linked form caused by mutations in the *KAL* gene.
- Turner syndrome and Klinefelter syndrome are the most common karyotypic abnormalities leading to primary hypogonadism.

TREATMENT OF DELAYED PUBERTY

The goal of therapy in children with hypogonadism is to initiate secondary sexual development in a way that simulates normal puberty. This is achieved by using low doses of sex steroids at initiation of therapy and gradually increasing the dosage over time. Once pubertal development is complete, adult replacement doses of gonadal steroids are required for the maintenance of secondary sexual characteristics and normal sexual function.

Girls

Multiple protocols exist for estrogen replacement therapy. These include transdermal estrogen or oral conjugated or unconjugated estrogen preparations, all of which are initiated at very low doses. Concurrently monitoring growth, skeletal maturation, and pubertal progression, the dose of estrogen is then gradually increased at 6- to 12-month increments to an adult replacement dose. Once breakthrough bleeding or Tanner stage IV breast development has been achieved, progesterone is added in the form of medroxyprogesterone acetate or micronized progesterone, which allows for the commencement of monthly bleeding. Once puberty is complete, oral contraceptive pills offer a convenient method of combined estrogen-progesterone therapy. A salutary effect of very-low-dose estrogen on growth has been observed in Turner syndrome, and the optimal estradiol replacement regimen in hypogonadal girls remains under investigation. As has been observed in hypogonadal boys, sex steroid replacement in girls with hypogonadism significantly improves calcium absorption and retention, with important implications in terms of the acquisition of bone mineral density.[160]

Boys

The most commonly employed method of sex steroid replacement therapy in adolescent hypogonadal boys consists of IM depot preparations of testosterone such as testosterone enanthate or cypionate. The typical starting dose is 50 to 100 mg IM or SC every 4 weeks, which is followed by gradual increases of 25 to 50 mg every 6 to 12 months as indicated by rates of growth, skeletal maturation, and pubertal progression. An adult replacement dose of testosterone consists of 200 to 250 mg IM or SC every 2 to 4 weeks, resulting in elevation of serum testosterone levels to within the normal range. Alternate forms of maintenance testosterone replacement include transdermal testosterone patches and testosterone gels. Although experience with these preparations during adolescence is limited, ongoing investigation may reveal that they are safe and effective for pubertal induction.[161]

KEY POINTS

- Treatment of pathologic causes of delayed puberty is with sex steroids, using low doses at initiation of therapy and gradually increasing the dosage over time to simulate normal puberty.

CONCLUSION

In summary, causes of precocious and delayed puberty include both normal variants and abnormalities at any level of the HPG axis. An increasing number of molecular genetic abnormalities have been linked to specific forms of precocious puberty and hypogonadism. Treatment of precocious puberty depends on the specific etiology. Sex steroid replacement leads to successful pubertal induction and the acquisition of adult secondary sexual development in those with hypogonadism.

REFERENCES

1. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988-1994. *Arch Pediatr Adolesc Med*. 2001;155:1022–1028.
2. Lee JM, et al. Weight status in young girls and the onset of puberty. *Pediatrics*. 2007;119:e624–e630.
3. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*. 2009;123:84–88.
4. Herman-Giddens ME, et al. Secondary sexual characteristics in boys: data from the pediatric research in office settings network. *Pediatrics*. 2012;130:e1058–e1068.
5. Chumlea WC, et al. Age at menarche and racial comparisons in US girls. *Pediatrics*. 2003;111:110–113.
6. Marti-Henneberg C, Vizmanos B. The duration of puberty in girls is related to the timing of its onset. *J Pediatr*. 1997;131:618–621.
7. Khokhar A, Mojica A. Premature thelarche. *Pediatr Ann*. 2018;47:e12–e15.
8. Klein KO, et al. Estrogen levels in girls with premature thelarche compared with normal prepubertal girls as determined by an ultrasensitive recombinant cell bioassay. *J Pediatr*. 1999;134:190–192.
9. Badouraki M, et al. Evaluation of pelvic ultrasonography in the diagnosis and differentiation of various forms of sexual precocity in girls. *Ultrasound Obstet Gynecol*. 2008;32:819–827.
10. Hannon TS, et al. Premature thelarche and granulosa cell tumors: a search for FSH receptor and G5alpha activating mutations. *J Pediatr Endocrinol Metab*. 2002;15:891–895.
11. Pescovitz OH, et al. Premature thelarche and central precocious puberty: the relationship between clinical presentation and the gonadotropin response to luteinizing hormone-releasing hormone. *J Clin Endocrinol Metab*. 1988;67:474–479.

12. Akinci A, Cetin D, Ilhan N. Plasma kisspeptin levels in girls with premature thelarche. *J Clin Res Pediatr Endocrinol*. 2012;4:61–65.
13. Klein KO, et al. Estrogen levels in childhood determined by an ultrasensitive recombinant cell bioassay. *J Clin Invest*. 1994;94:2475–2480.
14. Fisher MM, Eugster EA. What is in our environment that effects puberty? *Reprod Toxicol*. 2014;44:7–14.
15. Leonardi A, Cofini M, Rigante D, et al. The effect of bisphenol a on puberty: a critical review of the medical literature. *Int J Environ Res Public Health*. 2017;14:1044.
16. Van Winter JT, et al. Natural history of premature thelarche in Olmsted County, Minnesota, 1940 to 1984. *J Pediatr*. 1990;116:278–280.
17. Pasquino AM, et al. Progression of premature thelarche to central precocious puberty. *J Pediatr*. 1995;126:11–14.
18. Novello L, Speiser PW. Premature adrenarche. *Pediatr Ann*. 2018;47:e7–e11.
19. DeSalvo DJ, et al. In children with premature adrenarche, bone age advancement by 2 or more years is common and generally benign. *J Pediatr Endocrinol Metab*. 2013;26:215–221.
20. Gell JS, et al. Adrenarche results from development of a 3beta-hydroxysteroid dehydrogenase-deficient adrenal reticularis. *J Clin Endocrinol Metab*. 1998;83:3695–3701.
21. Riddick LM, et al. 3 alpha-Androstanediol glucuronide in premature and normal pubarche. *J Clin Endocrinol Metab*. 1991;72:46–50.
22. Balducci R, et al. Isolated precocious pubarche: an approach. *J Clin Endocrinol Metab*. 1994;79:582–589.
23. Dacou-Voutetakis C, Dracopoulou M. High incidence of molecular defects of the CYP21 gene in patients with premature adrenarche. *J Clin Endocrinol Metab*. 1999;84:1570–1574.
24. Ibanez L, et al. Fat distribution in non-obese girls with and without precocious pubarche: central adiposity related to insulinaemia and androgenaemia from prepuberty to postmenarche. *Clin Endocrinol (Oxf)*. 2003;58:372–379.
25. Ibanez L, et al. Premature adrenarche–normal variant or forerunner of adult disease? *Endocr Rev*. 2000;21:671–696.
26. Ibanez L, et al. Early metformin therapy to delay menarche and augment height in girls with precocious pubarche. *Fertil Steril*. 2011;95:727–730.
27. Nebesio TD, Eugster EA. Pubic hair of infancy: endocrinopathy or enigma? *Pediatrics*. 2006;117:951–954.
28. Eugster EA. Update on precocious puberty in girls. *J Pediatr Adolesc Gynecol*. 2019;32:455–459.
29. Aguirre RS, Eugster EA. Central precocious puberty: from genetics to treatment. *Best Pract Res Clin Endocrinol Metab*. 2018;32:343–354.
30. Teles MG, et al. A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*. 2008;358:709–715.
31. Silveira LG, et al. Mutations of the KISS1 gene in disorders of puberty. *J Clin Endocrinol Metab*. 2010;95:2276–2280.
32. Abreu AP, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med*. 2013;368:2467–2475.
33. Canton APM, et al. Pioneering studies on monogenic central precocious puberty. *Arch Endocrinol Metab*. 2019;63:438–444.
34. Busch AS, et al. Circulating MKRN3 levels decline during puberty in healthy boys. *J Clin Endocrinol Metab*. 2016;101:2588–2593.
35. Dauber A, et al. Paternally inherited DLK1 deletion associated with familial central precocious puberty. *J Clin Endocrinol Metab*. 2017;102:1557–1567.
36. Gomes LG, et al. DLK1 is a novel link between reproduction and metabolism. *J Clin Endocrinol Metab*. 2019;104:2112–2120.
37. Robben SG, et al. Idiopathic isosexual central precocious puberty: magnetic resonance findings in 30 patients. *Br J Radiol*. 1995;68:34–38.
38. Jung H, et al. Association of morphological characteristics with precocious puberty and/or gelastic seizures in hypothalamic hamartoma. *J Clin Endocrinol Metab*. 2003;88:4590–4595.
39. Cantas-Orsdemir S, Garb JL, Allen HF. Prevalence of cranial MRI findings in girls with central precocious puberty: a systematic review and meta-analysis. *J Pediatr Endocrinol Metab*. 2018;31:701–710.
40. Bangalore Krishna K, et al. Use of gonadotropin-releasing hormone analogs in children: update by an international consortium. *Horm Res Paediatr*. 2019;91:357–372.
41. Eugster EA. Treatment of central precocious puberty. *J Endocr Soc*. 2019;3:965–972.
42. Klein KO, et al. Phase 3 trial of a small-volume subcutaneous 6-month duration leuprolide acetate treatment for central precocious puberty. *J Clin Endocrinol Metab*. 2020:105:e3660-e3671.
43. Lewis KA, et al. A single histrelin implant is effective for 2 years for treatment of central precocious puberty. *J Pediatr*. 2013;163:1214–1216.
44. Fuqua JS. Treatment and outcomes of precocious puberty: an update. *J Clin Endocrinol Metab*. 2013;98:2198–2207.
45. Roth C. Therapeutic potential of GnRH antagonists in the treatment of precocious puberty. *Expert Opin Investig Drugs*. 2002;11:1253–1259.
46. Lewis KA, Eugster EA. Random luteinizing hormone often remains pubertal in children treated with the histrelin implant for central precocious puberty. *J Pediatr*. 2013;162:562–565.
47. Ghirri P, et al. Final height in girls with slowly progressive untreated central precocious puberty. *Gynecol Endocrinol*. 1997;11:301–305.
48. Schoelwer MJ, et al. One-year follow-up of girls with precocious puberty and their mothers: do psychological assessments change over time or with treatment? *Horm Res Paediatr*. 2017;88:347–353.
49. Cantas-Orsdemir S, Eugster EA. Update on central precocious puberty: from etiologies to outcomes. *Expert Rev Endocrinol Metab*. 2019;14:123–130.
50. Franzini IA, et al. GnRH analog is ineffective in increasing adult height in girls with puberty onset after 7 years of age: a systematic review and meta-analysis. *Eur J Endocrinol*. 2018;179:381–390.
51. Alessandri SB, et al. Bone mineral density and body composition in girls with idiopathic central precocious puberty before and after treatment with a gonadotropin-releasing hormone agonist. *Clinics (Sao Paulo)*. 2012;67:591–596.
52. Thornton P, et al. Review of outcomes after cessation of gonadotropin-releasing hormone agonist treatment of girls with precocious puberty. *Pediatr Endocrinol Rev*. 2014;11:306–317.
53. Lazar L, et al. Treated and untreated women with idiopathic precocious puberty: long-term follow-up and reproductive outcome between the third and fifth decades. *Clin Endocrinol (Oxf)*. 2014;80:570–576.
54. Haddad NG, Eugster EA. Peripheral precocious puberty including congenital adrenal hyperplasia: causes, consequences, management and outcomes. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101273.
55. Witchel SF. Congenital adrenal hyperplasia. *J Pediatr Adolesc Gynecol*. 2017;30:520–534.
56. Speiser PW, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:4043–4088.
57. Fukami M, et al. Aromatase excess syndrome: a rare autosomal dominant disorder leading to pre- or peri-pubertal onset gynecomastia. *Pediatr Endocrinol Rev*. 2014;11:298–305.
58. Albright FB, Hampton AO. Syndrome characterized by osteitis fibrosa disseminata, areas of pigmentation and endocrine dysfunction with precocious puberty in females. *N Engl J Med*. 1937;216:727–746.
59. Weinstein LS, et al. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*. 1991;325:1688–1695.
60. Collins MT, Singer FR, Eugster E. McCune-Albright syndrome and the extraskeletal manifestations of fibrous dysplasia. *Orphanet J Rare Dis*. 2012;7:S4.
61. Shenker A, et al. Severe endocrine and nonendocrine manifestations of the McCune-Albright syndrome associated with activating mutations of stimulatory G protein GS. *J Pediatr*. 1993;123:509–518.
62. Sotomayor K, et al. Ovarian function in adolescents with McCune-Albright syndrome. *J Pediatr Endocrinol Metab*. 2011;24:525–528.
63. Coutant R, et al. Macroorchidism due to autonomous hyperfunction of Sertoli cells and G(s)alpha gene mutation: an unusual expression of McCune-Albright syndrome in a prepubertal boy. *J Clin Endocrinol Metab*. 2001;86:1778–1781.
64. Weinstein LS, Chen M, Liu J. Gs(alpha) mutations and imprinting defects in human disease. *Ann NY Acad Sci*. 2002;968:173–197.
65. Mieszczak J, et al. The aromatase inhibitor anastrozole is ineffective in the treatment of precocious puberty in girls with McCune-Albright syndrome. *J Clin Endocrinol Metab*. 2008;93:2751–2754.

66. Feuillan P, et al. Letrozole treatment of precocious puberty in girls with the McCune-Albright syndrome: a pilot study. *J Clin Endocrinol Metab*. 2007;92:2100–2106.
67. Estrada A, et al. Long-term outcomes of letrozole treatment for precocious puberty in girls with McCune-Albright syndrome. *Eur J Endocrinol*. 2016;175:477–483.
68. Eugster EA, et al. Tamoxifen treatment for precocious puberty in McCune-Albright syndrome: a multicenter trial. *J Pediatr*. 2003;143:60–66.
69. de GBPC, et al. Tamoxifen improves final height prediction in girls with McCune-Albright syndrome: a long follow-up. *Horm Res Paediatr*. 2015;84:184–189.
70. Sims EK, et al. Fulvestrant treatment of precocious puberty in girls with McCune-Albright syndrome. *Int J Pediatr Endocrinol*. 2012;2012:26.
71. Boyce AM, et al. Gynecologic and reproductive outcomes in fibrous dysplasia/McCune-Albright syndrome. *Orphanet J Rare Dis*. 2019;14:90.
72. Neyman A, Eugster EA. Treatment of girls and boys with McCune-Albright syndrome with precocious puberty—update 2017. *Pediatr Endocrinol Rev*. 2017;15:136–141.
73. Nakamoto JM, et al. Concurrent hormone resistance (pseudohypoparathyroidism type Ia) and hormone independence (testotoxicosis) caused by a unique mutation in the G alpha s gene. *Biochem Mol Med*. 1996;58:18–24.
74. Shenker A. Activating mutations of the lutropin choriogonadotropin receptor in precocious puberty. *Recept Channels*. 2002;8:3–18.
75. Macedo DB, et al. Sexual precocity—genetic bases of central precocious puberty and autonomous gonadal activation. *Endocr Dev*. 2016;29:50–71.
76. Egli CA, et al. Pituitary gonadotropin-independent male-limited autosomal dominant sexual precocity in nine generations: familial testotoxicosis. *J Pediatr*. 1985;106:33–40.
77. Lenz AM, et al. Bicalutamide and third-generation aromatase inhibitors in testotoxicosis. *Pediatrics*. 2010;126:e728–e733.
78. Reiter EO, et al. Bicalutamide plus anastrozole for the treatment of gonadotropin-independent precocious puberty in boys with testotoxicosis: a phase II, open-label pilot study (BATT). *J Pediatr Endocrinol Metab*. 2010;23:999–1009.
79. Leschek EW, et al. Effect of antiandrogen, aromatase inhibitor, and gonadotropin-releasing hormone analog on adult height in familial male precocious puberty. *J Pediatr*. 2017;190:229–235.
80. Nabhan ZM, Eugster EA. Testotoxicosis with an episodic course: an unusual case within a series. *AACE Clin Case Rep*. 2019;5:e50–e53.
81. Vasseur C, et al. A chorionic gonadotropin-sensitive mutation in the follicle-stimulating hormone receptor as a cause of familial gestational spontaneous ovarian hyperstimulation syndrome. *N Engl J Med*. 2003;349:753–759.
82. Kelleher CM, Goldstein AM. Adnexal masses in children and adolescents. *Clin Obstet Gynecol*. 2015;58:76–92.
83. Olivier P, et al. Leydig cell tumors in children: contrasting clinical, hormonal, anatomical, and molecular characteristics in boys and girls. *J Pediatr*. 2012;161:1147–1152.
84. Ghazi AA, et al. Functioning adrenocortical tumors in children-secretory behavior. *J Clin Res Pediatr Endocrinol*. 2013;5:27–32.
85. Englund AT, et al. Pediatric germ cell and human chorionic gonadotropin-producing tumors. Clinical and laboratory features. *Am J Dis Child*. 1991;145:1294–1297.
86. Zimmerman PA, Francis GL, Poth M. Hormone-containing cosmetics may cause signs of early sexual development. *Mil Med*. 1995;160:628–630.
87. Van Wyk JJ, Grumbach MM. Syndrome of precocious menstruation and galactorrhea in juvenile hypothyroidism: an example of hormonal overlap in pituitary feedback. *J Pediatr*. 1960;57:P415–P435.
88. Cabrera SM, DiMeglio LA, Eugster EA. Incidence and characteristics of pseudoprecocious puberty because of severe primary hypothyroidism. *J Pediatr*. 2013;162:637–639.
89. Bruder JM, et al. Hypothyroidism-induced macroorchidism: use of a gonadotropin-releasing hormone agonist to understand its mechanism and augment adult stature. *J Clin Endocrinol Metab*. 1995;80:11–16.
90. Anasti JN, et al. A potential novel mechanism for precocious puberty in juvenile hypothyroidism. *J Clin Endocrinol Metab*. 1995;80:276–279.
91. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*. 1969;44:291–303.
92. Howard SR, et al. IGSF10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. *EMBO Mol Med*. 2016;8:626–642.
93. Pugliese-Pires PN, et al. Novel inactivating mutations in the GH secretagogue receptor gene in patients with constitutional delay of growth and puberty. *Eur J Endocrinol*. 2011;165:233–241.
94. Palmert MR, Dunkel L. Clinical practice. Delayed puberty. *N Engl J Med*. 2012;366:443–453.
95. Ehrmann DA, et al. A new test of combined pituitary-testicular function using the gonadotropin-releasing hormone agonist nafarelin in the differentiation of gonadotropin deficiency from delayed puberty: pilot studies. *J Clin Endocrinol Metab*. 1989;69:963–967.
96. Degros V, et al. The human chorionic gonadotropin test is more powerful than the gonadotropin-releasing hormone agonist test to discriminate male isolated hypogonadotropic hypogonadism from constitutional delayed puberty. *Eur J Endocrinol*. 2003;149:23–29.
97. Coutant R, et al. Baseline inhibin B and anti-Mullerian hormone measurements for diagnosis of hypogonadotropic hypogonadism (HH) in boys with delayed puberty. *J Clin Endocrinol Metab*. 2010;95:5225–5232.
98. Chan YM, et al. Divergent responses to kisspeptin in children with delayed puberty. *JCI Insight*. 2018;3:e99109.
99. Howard SR, Dunkel L. Management of hypogonadism from birth to adolescence. *Best Pract Res Clin Endocrinol Metab*. 2018;32:355–372.
100. Arrigo T, et al. Final height outcome in both untreated and testosterone-treated boys with constitutional delay of growth and puberty. *J Pediatr Endocrinol Metab*. 1996;9:511–517.
101. Saggese G, Bertelloni S, Baroncelli GI. Sex steroids and the acquisition of bone mass. *Horm Res*. 1997;48:65–71.
102. Moreira-Andres MN, et al. Bone mineral status in prepubertal children with constitutional delay of growth and puberty. *Eur J Endocrinol*. 1998;139:271–275.
103. Smith EP, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med*. 1994;331:1056–1061.
104. Morishima A, et al. Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens. *J Clin Endocrinol Metab*. 1995;80:3689–3698.
105. Uriarte MM, et al. The effect of pubertal delay on adult height in men with isolated hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 1992;74:436–440.
106. Herbison AE. Genetics of puberty. *Horm Res*. 2007;68:75–79.
107. Raivio T, et al. Reversal of idiopathic hypogonadotropic hypogonadism. *N Engl J Med*. 2007;357:863–873.
108. Kallmann FJ, Schoenfeld WA, Barrera SE. The genetic aspect of primary eunuchoidism. *Am J Ment Retard*. 1944;48:203.
109. Soussi-Yanicostas N, et al. Anosmin-1, defective in the X-linked form of Kallmann syndrome, promotes axonal branch formation from olfactory bulb output neurons. *Cell*. 2002;109:217–228.
110. Topaloglu AK. Update on the genetics of idiopathic hypogonadotropic hypogonadism. *J Clin Res Pediatr Endocrinol*. 2017;9:113–122.
111. Chan YM, et al. GNRH1 mutations in patients with idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci USA*. 2009;106:11703–11708.
112. de Roux N, et al. A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *N Engl J Med*. 1997;337:1597–1602.
113. de Roux N, et al. Hypogonadotropic hypogonadism due to loss of function of the KiSS1-derived peptide receptor GPR54. *Proc Natl Acad Sci USA*. 2003;100:10972–10976.
114. Topaloglu AK, et al. Inactivating KISS1 mutation and hypogonadotropic hypogonadism. *N Engl J Med*. 2012;366:629–635.
115. Radovick S, et al. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science*. 1992;257:1115–1118.

116. Sheng HZ, et al. Specification of pituitary cell lineages by the LIM homeobox gene Lhx3. *Science*. 1996;272:1004–1007.
117. Wu W, et al. Mutations in PROP1 cause familial combined pituitary hormone deficiency. *Nat Genet*. 1998;18:147–149.
118. Netchine I, et al. Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat Genet*. 2000;25:182–186.
119. Pfaeffle RW, et al. Three novel missense mutations within the LHX4 gene are associated with variable pituitary hormone deficiencies. *J Clin Endocrinol Metab*. 2008;93:1062–1071.
120. Thomas PQ, et al. Heterozygous HESX1 mutations associated with isolated congenital pituitary hypoplasia and septo-optic dysplasia. *Hum Mol Genet*. 2001;10:39–45.
121. Kelberman D, et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest*. 2006;116:2442–2455.
122. Zanaria E, et al. An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature*. 1994;372:635–641.
123. Kaiserman KB, et al. Minipuberty of infancy and adolescent pubertal function in adrenal hypoplasia congenita. *J Pediatr*. 1998;133:300–302.
124. Ozisik G, et al. An alternate translation initiation site circumvents an amino-terminal DAX1 nonsense mutation leading to a mild form of X-linked adrenal hypoplasia congenita. *J Clin Endocrinol Metab*. 2003;88:417–423.
125. Tabarin A, et al. A novel mutation in DAX1 causes delayed-onset adrenal insufficiency and incomplete hypogonadotropic hypogonadism. *J Clin Invest*. 2000;105:321–328.
126. Ozisik G, Achermann JC, Jameson JL. The role of SF1 in adrenal and reproductive function: insight from naturally occurring mutations in humans. *Mol Genet Metab*. 2002;76:85–91.
127. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev*. 2000;21:551–583.
128. Matthews CH, et al. Primary amenorrhoea and infertility due to a mutation in the beta-subunit of follicle-stimulating hormone. *Nat Genet*. 1993;5:83–86.
129. Phillip M, et al. Male hypogonadism due to a mutation in the gene for the beta-subunit of follicle-stimulating hormone. *N Engl J Med*. 1998;338:1729–1732.
130. Aittomaki K, et al. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell*. 1995;82:959–968.
131. Tapanainen JS, et al. Men homozygous for an inactivating mutation of the follicle-stimulating hormone (FSH) receptor gene present variable suppression of spermatogenesis and fertility. *Nat Genet*. 1997;15:205–206.
132. Weiss J, et al. Hypogonadism caused by a single amino acid substitution in the beta subunit of luteinizing hormone. *N Engl J Med*. 1992;326:179–183.
133. Lofrano-Porto A, et al. Luteinizing hormone beta mutation and hypogonadism in men and women. *N Engl J Med*. 2007;357:897–904.
134. Basciani S, et al. Hypogonadism in a patient with two novel mutations of the luteinizing hormone beta-subunit gene expressed in a compound heterozygous form. *J Clin Endocrinol Metab*. 2012;97:3031–3038.
135. Kremer H, et al. Male pseudohermaphroditism due to a homozygous missense mutation of the luteinizing hormone receptor gene. *Nat Genet*. 1995;9:160–164.
136. Wolf NI, et al. Clinical spectrum of 4H leukodystrophy caused by POLR3A and POLR3B mutations. *Neurology*. 2014;83:1898–1905.
137. Margolin DH, et al. Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. *N Engl J Med*. 2013;368:1992–2003.
138. Quaynor SD, et al. Delayed puberty and estrogen resistance in a woman with estrogen receptor alpha variant. *N Engl J Med*. 2013;369:164–171.
139. Clement K, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*. 1998;392:398–401.
140. Gordon CM, et al. Functional hypothalamic amenorrhea: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2017;102:1413–1439.
141. Muller J. Hypogonadism and endocrine metabolic disorders in Prader-Willi syndrome. *Acta Paediatr Suppl*. 1997;423:58–59.
142. Germain-Lee EL, et al. Growth hormone deficiency in pseudohypoparathyroidism type 1a: another manifestation of multihormone resistance. *J Clin Endocrinol Metab*. 2003;88:4059–4069.
143. Ropelato MG, et al. Gonadotropin secretion in prepubertal normal and agonadal children evaluated by ultrasensitive time-resolved immunofluorometric assays. *Horm Res*. 1997;48:164–172.
144. Meachem SJ, Nieschlag E, Simoni M. Inhibin B in male reproduction: pathophysiology and clinical relevance. *Eur J Endocrinol*. 2001;145:561–571.
145. Turner H. A syndrome of infantilism, congenital webbed neck and cubitus valgus. *Endocrinology*. 1938;23:566–574.
146. Folsom LJ, et al. Characterization of spontaneous and induced puberty in girls with Turner syndrome. *Endocr Pract*. 2017;23:768–774.
147. Gravholt CH, Backeljauw P. New international Turner syndrome guideline: a multi-society feat. *Eur J Endocrinol*. 2017;177:E1–E2.
148. Tharapel AT, et al. Deletion (X)(q26.1-->q28) in a proband and her mother: molecular characterization and phenotypic-karyotypic deductions. *Am J Hum Genet*. 1993;52:463–471.
149. Klinefelter HF, Reifenstein EC, Albright F. Syndrome characterized by gynecomastia, aspermatogenesis without aleydigism and increased excretion of follicle-stimulating hormone. *J Clin Endocrinol Metab*. 1942;2:615-627.
150. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *J Clin Endocrinol Metab*. 2003;88:622–626.
151. Lizarazo AH, McLoughlin M, Vogiatzi MG. Endocrine aspects of Klinefelter syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2019;26:60–65.
152. Stocco DM. Clinical disorders associated with abnormal cholesterol transport: mutations in the steroidogenic acute regulatory protein. *Mol Cell Endocrinol*. 2002;191:19–25.
153. Bose HS, Pescovitz OH, Miller WL. Spontaneous feminization in a 46,XX female patient with congenital lipoid adrenal hyperplasia due to a homozygous frameshift mutation in the steroidogenic acute regulatory protein. *J Clin Endocrinol Metab*. 1997;82:1511–1515.
154. Quigley CA, et al. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev*. 1995;16:271–321.
155. Swyer GI. Male pseudohermaphroditism: a hitherto undescribed form. *Br Med J*. 1955;2:709–712.
156. Baetens D, et al. Update on the genetics of differences of sex development (DSD). *Best Pract Res Clin Endocrinol Metab*. 2019;33:101271.
157. Huang H, Wang C, Tian Q. Gonadal tumour risk in 292 phenotypic female patients with disorders of sex development containing Y chromosome or Y-derived sequence. *Clin Endocrinol (Oxf)*. 2017;86:621–627.
158. Pirgon O, Dundar BN. Vanishing testes: a literature review. *J Clin Res Pediatr Endocrinol*. 2012;4:116–120.
159. Ladjouze A, Donaldson M. Primary gonadal failure. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101295.
160. Norjavaara E, Ankarberg-Lindgren C, Kristrom B. Sex steroid replacement therapy in female hypogonadism from childhood to young adulthood. *Endocr Dev*. 2016;29:198–213.
161. Bertelloni S, et al. Androgen therapy in hypogonadal adolescent males. *Horm Res Paediatr*. 2010;74:292–296.

Transgender Healthcare

Madeline B. Deutsch

105

Gonadotropin Regulation and Androgen and Estrogen Physiology

Shalender Bhasin, Ravi Jasuja, and Channa N. Jayasena

OUTLINE

Regulation of Gonadotropin, 1760
- The Gonadotropins, 1760
- Gonadotropin-Releasing Hormone, 1760
- Feedback on Gonadotropin Secretion by Sex Steroids, 1762
- Metabolic Regulation of Gonadotropins, 1763
- Regulation of Gonadotropins by Stress, 1765
- Seasonal Regulation of Gonadotropin Secretion, 1765

Androgen Physiology, 1765
- The Role of Aromatization in Mediating the Biologic Effects of Testosterone, 1765
- The Role of the Steroid 5**α**-Reductase System in Mediating the Biologic Effects of Testosterone, 1766
- Activation of Other Signaling Pathways by Metabolites of Testosterone, 1767
- Biosynthesis of Testosterone and Other Androgens, 1767
- Structure Activity Relationship in Steroidal Androgens, 1768
- The Regulation of Testosterone Production by LH, 1768
- Transport and Bioavailability, 1768
- Testosterone Metabolism, 1771
- Mechanisms of Testosterone's Action, 1771

Estrogen Physiology, 1773
- Endogenous Estrogens, 1773
- Regulation of Estradiol Production, 1774
- Estrogen Transport and Bioavailability, 1775
- Estrogen Metabolism, 1775
- Mechanisms of Estrogen Action, 1775

REGULATION OF GONADOTROPIN

The Gonadotropins

The gonadotropin hormones, luteinizing hormone (LH), and follicle-stimulating hormone (FSH), are secreted by specialized cells (gonadotrophs) within the anterior pituitary gland. The secretion of LH and FSH is orchestrated by hypothalamic activity. Gonadotropin-releasing hormone (GnRH) is released from the median eminence of the hypothalamus via the hypophyseal-portal circulation to stimulate gonadotropin secretion. LH and FSH are glycopeptides each consisting of a common α-subunit, and a unique LH-β or FSH-β subunit, respectively. The secretory patterns of LH and FSH have two components, basal and pulsatile. The pituitary secretes tonic (basal) levels of each hormone with superimposed burst of secretion (pulsatile) every 60 to 90 minutes, which is stimulated by the pulsatile GnRH secretion. LH has a more pronounced pulsatile component than FSH, which has a greater component of basal secretion. Push-pull perfusion frequent sampling of the hypophyseal-portal circulation in monkeys has determined that pulses of LH, FSH secretion are entirely dependent on and coincident with hypothalamic GnRH secretion.[1] The frequency of GnRH pulses regulates the ratio of LH to FSH secretion. High GnRH pulse frequencies increase α-subunit and LH-β subunit transcription, thus favoring LH secretion; conversely, low GnRH pulse frequency favors FSH-β gene transcription and FSH secretion.[2]

Gonadotropin-Releasing Hormone

The Nobel Prize-winning isolation of GnRH from hypothalamic porcine extracts and discovery of its potent action to stimulate LH secretion heralded the conceptualization of the hypothalamo-pituitary-gonadal (HPG) or reproductive axis.[3] Several isoforms of GnRH exist and are distributed throughout the mammalian central nervous system (CNS), but only the GnRH I isoform has a known, direct role in regulating gonadotropin secretion. During embryogenesis, GnRH I neurons migrate caudally from the olfactory placode to the basal forebrain, under spatiotemporal regulation by several proteins such as those involved in fibroblast growth factor (FGF) and prokineticin signaling, adhesion molecules such as anosmin-1 (the product of KAL gene), and heparin sulfate glycosaminoglycan compounds (Fig. 105.1).[4] Craniofacial tissues and olfactory ensheathing cells also play important roles in neurogenesis and migration of the GnRH neurons.[4,5] Only 1000 to 2000 GnRH neuronal cell bodies are distributed postnatally within regions of the basal forebrain, particularly the preoptic area (POA). GnRH neuronal axons terminate at the median eminence, which has a fenestrated area of blood-brain barrier, enabling GnRH secretion into the hypophyseal-portal circulation. Failure of appropriate embryonic migration of GnRH neurons leads to congenital hypogonadotropic hypogonadism (CHH, low levels of gonadotropins and sex steroids) associated with anosmia, termed Kallmann syndrome.[4] Failure of GnRH activation (with normal migration), its regulation, or its action on the gonadotroph such as that associated with mutations in GnRHR, GNRH1, KISS1R, leptin, or leptin receptor, Tachykinin 3 (TAC3), TACR3, causes normosmic CHH.[4,5] Some mutations, such as those in PROK2, PROKR2, FGFR1, FGF8, SEMA3A, and CHD7, have been associated with both anosmic and normosmic forms of IHH.[4]

Pubertal Onset of GnRH Secretion. After a brief period of activity during the neonatal period, GnRH secretion from specialized neurons in the medial POA (in rodents and sheep) or mediobasal hypothalamus (in primates) remains quiescent throughout juvenile life.[6] Subsequently, a surge in the pulsatile release of GnRH into the

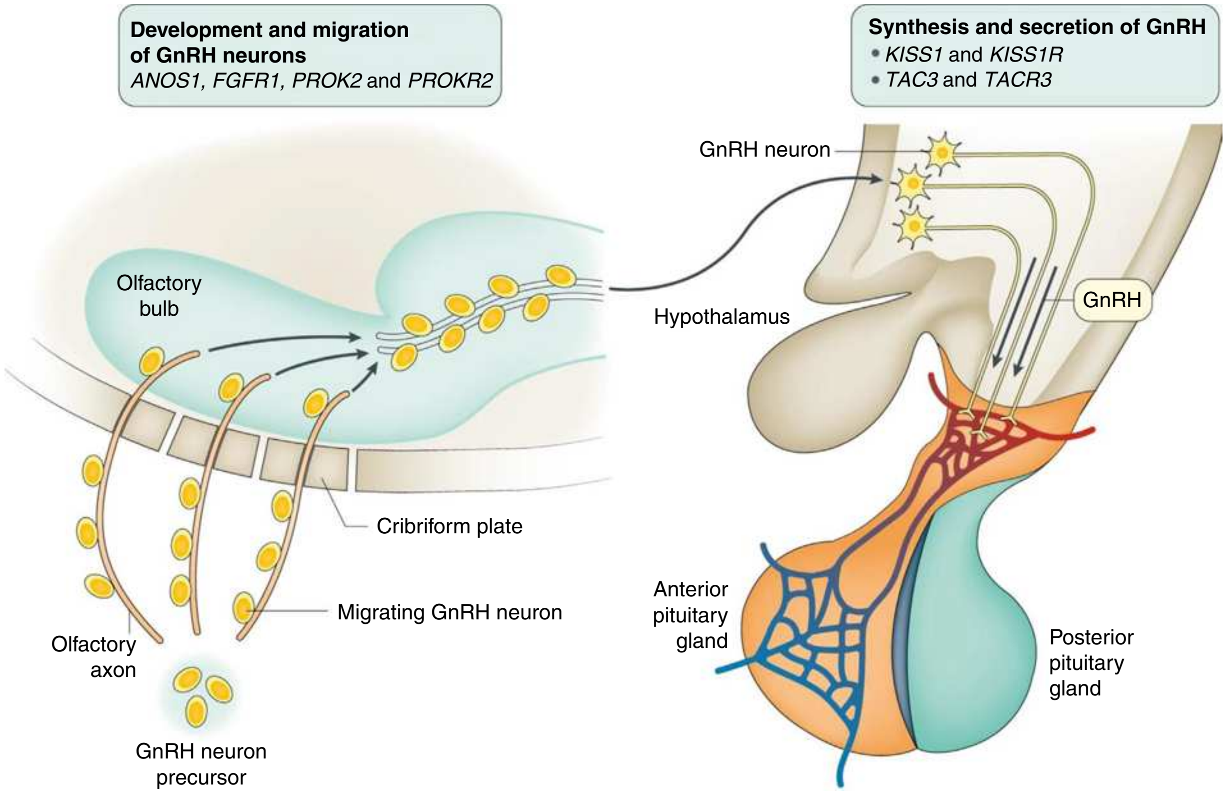


Figure 105.1 Migration of gonadotrophin-releasing hormone (GnRH) neurons during embryogenesis. During embryogenesis, GnRH neuron precursors originating within the olfactory placode migrate caudally to the hypothalamus. Several neurotrophic factors including prokineticin 2 (PROK2), the prokineticin 2 receptor (PROKR2), anosmin-1 (ASOS1), and fibroblast growth factor receptor 1 regulate spaciotemporally the migration of GnRH neurons. Failure of migration of GnRH neurons causes a congenital hypogonadotropic hypogonadism (CHH) with anosmia (Kallmann syndrome). Synthesis and secretion of GnRH is regulated by factors including kisspeptin (KISS1) and its receptor (KISS1R), neurokinin B (TAC3), and the neurokinin 3 receptor (TACR3). Failure of activation of GnRH neurons following normal migration is termed normosmic CHH. GnRH neurons terminate at the median eminence that has an incomplete blood-brain barrier, allowing GnRH to be released into the hypophyseal-portal circulation to reach the anterior pituitary gland. (From Salonia A, Rastrelli G, Hackett G, et al. Paediatric and adult-onset male hypogonadism. *Nat Rev Dis Primers*. 2019;5(1):38.)

hypothalamo-pituitary portal circulation heralds the onset of puberty.[7] GnRH stimulates the pulsatile release of LH and FSH to stimulate gonadal maturation and sex steroid secretion to mediate pubertal maturation and sex-specific physical characteristics.

The GnRH Pulse Generator. Pioneering studies by Knobil unveiled the critical requirement for pulsatile pattern of GnRH secretion to maintain physiological gonadotropin secretion.[8] Constant infusion of synthetic GnRH fails to stimulate gonadotropin release in monkeys with hypogonadotropic hypogonadism due to lesioning of the hypothalamus; however, hourly pulses of GnRH restore pituitary gonadotropin secretion in these animals.[8] Although an intermittent pattern of GnRH release may be observed from the cultured GnRH neurons, their activity is extensively modified by a network of neurohumoral signals. The hypothalamic elements responsible for the physiological pattern of pulsatile GnRH secretion are collectively referred to as the "GnRH pulse generator." Several candidate components of the GnRH pulse generator have been identified[1] and are listed in Table 105.1.[9] The relative balance of these factors determines the pubertal onset and subsequent pulsatile secretion of GnRH and gonadotropin secretion. Disruption of excitatory or modulatory inputs to the GnRH pulse generator such as kisspeptin or neurokinin B may cause hypogonadotropic hypogonadism. Similarly, overactivation of inhibitory components of the GnRH pulse generator such as excessive dynorphin (an endogenous opioid) causes hypogonadotropic hypogonadism.

Kisspeptin as a Regulator of Gonadotropin Secretion. Kisspeptin is a group of RF-amide peptides encoded by the KISS1 gene, which signals via the G-protein-coupled Kisspeptin receptor. Multiple kisspeptin isoforms have been identified; kisspeptin-54 is the major circulating isoform in humans.[10] Expression of kisspeptin and its receptor has been identified in two major neuronal populations within the hypothalamus of rodents: in the arcuate nucleus (ARC) and the anteroventral periventricular nucleus (AVPV).[11] In humans and primates, kisspeptin mRNA is predominantly expressed within the infundibular nucleus (equivalent of the ARC in this order of mammals). Kisspeptin neurons project to GnRH neuronal terminals in the median eminence.

Inactivating mutations in the gene encoding kisspeptin and its receptor cause pubertal failure.[12] Kisspeptin receptor null mice have hypogonadotropic hypogonadism (HH), as suggested by low levels of circulating gonadotropin hormones, with small testes in male mice and a delay in vaginal opening and an absence of follicular maturation in female mice. Administration of exogenous GnRH corrected the HH phenotype observed with inactivation of kisspeptin signaling, which is consistent with the view that kisspeptin acts by stimulating endogenous GnRH. Conversely, activation of kisspeptin signaling

TABLE 105.1 Neurotransmitters Which Modulate Gonadotropin-Releasing Hormone (GnRH) Release

Excitatory	Inhibitory	Excitatory, Inhibitory, or Modulatory
Glutamate	Corticotrophin-releasing hormone	Dopamine
Neuropeptide Y	GABA	Acetylcholine
Substance P	Dynorphin	Oxytocin
Kisspeptin	Serotonin	Vasopressin
Histamine		Neurokinin B
Vasoactive intestinal peptide		
Galanin		
Dopamine		
Noradrenaline		
Cholecystokinin		

Candidate factors have been classified as being excitatory, inhibitory, or both excitatory and inhibitory on GnRH release. *GABA,* gamma-aminobutyric acid.

Adapted from Gore AC. Factors regulating GnRH neurons. In *GnRH: The Master Molecule of Reproduction*. Boston: Kluwer Academic Puiblishers, 2009:93–124.

induces precocious puberty in rodents. Furthermore, activating mutations in the kisspeptin receptor gene have been identified in girls with precocious puberty.[13] In several mammalian species, administration of intracerebroventricular or intravenous kisspeptin has been shown to acutely and potently increase LH and FSH secretion in males and females by stimulating GnRH secretion. Furthermore, intravenous or subcutaneous injection of kisspeptin potently increases LH and FSH in men and women with or without hypogonadism.[14,15] In summary, kisspeptin is a direct and potent stimulator of endogenous GnRH secretion, which gives it potential to treat patients with reproductive disorders in the future.

It has been proposed that kisspeptin may also have direct effects on the anterior pituitary gland to stimulate the release of LH and FSH. Both the KISS1 and KISS1R genes are expressed in pituitary gonadotrophs. The addition of kisspeptin results in dose-dependent LH secretion from cultured pituitary cells and tissue explants.[16] However, kiss1r$^{-/-}$ mice display the phenotype of hypogonadotropic hypogonadism and respond to exogenous GnRH with an appropriate LH response. In addition, intravenous administration of kisspeptin fails to stimulate LH release in GnRH-replaced, ovariectomized ewes in which the hypothalamus and the pituitary have been disconnected. In summary, a direct action of kisspeptin on the pituitary is plausible, but its physiological importance remains unclear.

Neurokinin B and Dynorphin as Regulators of Gonadotropins.

NKB and DYN are coexpressed with kisspeptin within the ARC; these neurons are commonly referred to as kisspeptin/neurokinin B/dynorphin (KNDy) neurons (Fig. 105.2). Like kisspeptin, NKB and DYN play important regulatory roles in mammalian reproduction. Mutations in the genes encoding NKB, TAC3, or its receptor (TACR3) are associated with hypogonadotropic hypogonadism.

The role of NKB in GnRH pulse regulation remains controversial in part due to the variable effects of exogenous NKB administration on LH secretion depending on the animal model. Peripheral administration of NKB in healthy humans has no effect on gonadotropin release. The heterogeneity of observed effects of NK3R agonism in different experimental models may arise due to differences in the steroid hormone milieu during NKB administration.

Exogenous opioids profoundly lower LH and FSH and inhibit LH pulsatility. It is now thought that exogenous opioids act by mimicking the effects of DYN, an endogenous opioid peptide that acts primarily through the κ-opioid receptor (KOR).[17] It has been proposed that DYN may act in an autocrine manner to negatively regulate KNDy neurons that express KOR, rather than directly acting upon GnRH neurons, few of which express KOR.[18]

In summary, KNDy neurons have complex and powerful effects to regulate human reproduction. Kisspeptin activates GnRH secretion, and DYN is autoinhibitory. NKB signaling is obligatory for GnRH activation, but its precise physiological role remains unclear.

KEY POINTS

- Pulsatile secretion of gonadotropin-releasing hormone (GnRH) stimulates luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion, but constant infusion of GnRH paradoxically inhibits LH and FSH secretion.
- Low levels of circulating estradiol suppress LH and FSH secretion, but high levels stimulate an LH surge to trigger ovulation in women.
- KNDy neurons in the preoptic area of the hypothalamus coexpress kisspeptin, neurokinin B, and dynorphin and mediate the negative feedback effects of testosterone, estradiol, and progesterone by inhibition of kisspeptin release.

Feedback on Gonadotropin Secretion by Sex Steroids

Estrogen Feedback. In mammals, estradiol has biphasic effects on GnRH secretion. Low levels of circulating estrogen suppress gonadotropin secretion, but at mid-cycle, high circulating levels of estrogen (from the dominant ovarian follicle) exert a potent positive feedback effect, leading to a surge in the plasma LH and ovulation. ERα is critical for the positive feedback effect of estrogen, which results in the female mid-cycle LH surge. Adult female rodents pre-administered an ERα antagonist fail to ovulate or undergo a preovulatory LH surge and have blunted LH response following administration of kisspeptin-10.[19] Conversely, the preovulatory LH surge is preserved in ERβ knockout mice and rats administered an ERβ antagonist.

GnRH neurons lack estrogen receptor alpha (ERα), which suggests the existence of an intermediary neuronal pathway for mediating estrogen feedback.[20] Kisspeptin is now known to play a critical role in mediating the effects of estrogen on GnRH to mediate the LH surge. Kisspeptin expression in the AVPV of rats is the highest during the evening of proestrus when the LH surge occurs, whereas expression levels in the ARC are at their lowest during this time.[21] In addition, kisspeptin expression is increased in the AVPV at the time of estrogen and progesterone-induced LH surge in ovariectomized rats, but ARC kisspeptin expression is at its lowest during this time (Fig. 105.3).[21] Furthermore, kisspeptin neurons in the AVPV coexpress a surrogate marker of cellular activation, Fos, at the time of the LH surge, whereas minimal Fos expression was observed on diestrus; in contrast, ARC kisspeptin neurons do not express Fos during the LH surge or on diestrus. Finally, most kisspeptin neurons in both the AVPV and ARC express ERα. Blockade of hypothalamic kisspeptin signaling by intracerebroventricular injection by the pharmacological antagonist is sufficient to prevent ovulation in rodents.[22] Taken together, these data provide compelling evidence that a subpopulation of kisspeptin signaling in the AVPV is activated by circulating estrogen to generate the preovulatory LH surge in rats.

The epigenetic regulation of kisspeptin gene expression in mediating the estrogen-positive feedback action has been investigated

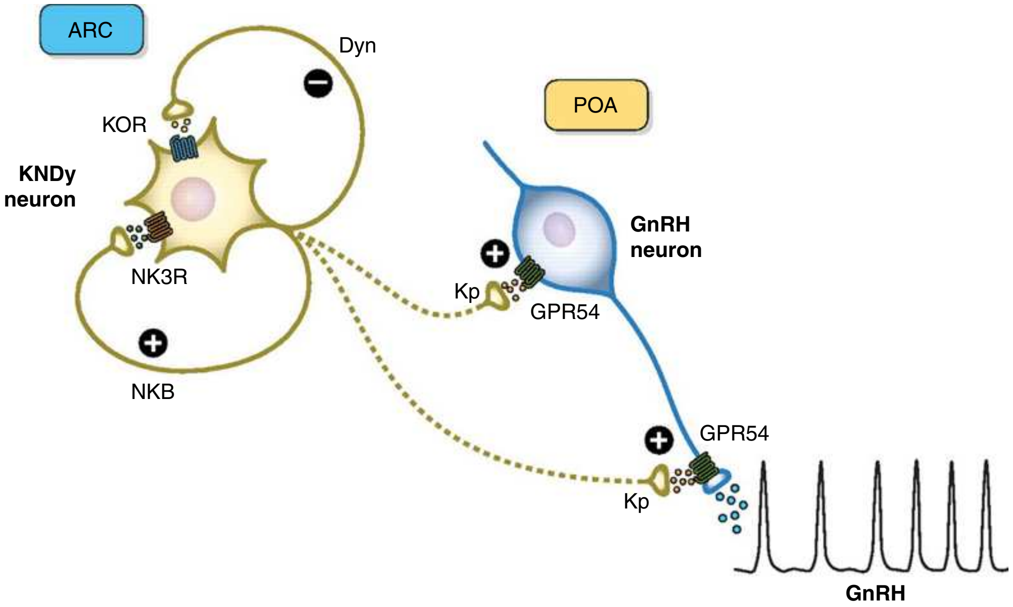


Figure 105.2 The role of kisspeptin-neurokinin B-dynorphin (KNDy) neurons in the control of pulsatile gonadotrophin-releasing hormone (GnRH) secretion. The physiological, pulsatile pattern of GnRH is critical to its ability to stimulate gonadotrophin secretion. KNDy neurons may play an important role in GnRH activation and pulse generation. A population of KNDy neurons within the hypothalamic ARCuate nucleus projects to GnRH neuronal cell bodies and terminals within the preoptic area (POA). KNDy neurons are the major driving signal for GnRH pulsatile release. Kisspeptin released from KDNy neuronal terminals signals via the kisspeptin receptor (GPR54) to potently activate GnRH neuronal activity, leading to GnRH secretion, which stimulates gonadotropin secretion. NKB signals via the neurokinin 3 receptor (NK3R) and is implicated in GnRH pulse initiation, possibly autoactivating KNDy neurons. Conversely, DYNorphin signals via the kappa opioid receptor (KOR), which may mediate GnRH pulse termination by autoinhibition of KNDy neurons. (Adapted from Pinilla L, Aguilar E, Dieguez C, Millar RP, Tena-Sempere M. Kisspeptins and reproduction: physiological roles and regulatory mechanisms. *Physiol Rev*. 2012;92(3):1235–316. doi: 10.1152/physrev.00037.2010.)

in mice. In the mouse AVPV, there is a high degree of acetylation at the histone of kisspeptin gene locus while the same locus in the ARC showed histone deacetylation in response to estrogen.[23] This suggests that epigenetic regulation of kisspeptin contributes to the activation of kisspeptin signaling in the AVPV in response to estrogen underpinning the LH surge.[23]

Testosterone Feedback. Unlike estrogen which has biphasic effects on gonadotropins, testosterone suppresses LH and FSH secretion in most physiologic conditions. GnRH neurons do not express the androgen receptor (AR). Furthermore, reduced levels of testosterone are associated with increased LH and FSH pulse frequency, suggesting that testosterone inhibits GnRH secretion.[24] Kisspeptin neurons within the hypothalamic ARC have been implicated in mediating the negative feedback regulation of GnRH secretion by circulating testosterone. Orchidectomy normally elevates LH and FSH due to reduced negative testosterone feedback; this is prevented by the administration of a selective kisspeptin antagonist. Aromatization to estradiol may partially mediate the negative feedback upon kisspeptin signaling by testosterone as dihydrotestosterone is less effective at suppressing Kiss1 ARC kiss1 mRNA than testosterone or estradiol in rodents.[25]

When the action of GnRH on the gonadotrophs is blocked by a GnRH antagonist, testosterone directly and selectively increases pituitary FSH secretion.[26] Thus, in male rats treated with a GnRH antagonist, graded doses of testosterone are associated with a dose-dependent increase in serum FSH[26] and upregulation of FSH-β mRNA levels,[27] whereas serum LH remains suppressed.

Metabolic Regulation of Gonadotropins

The availability of adequate nutrition is obligatory for fertility. The signals linking body weight, energy stores, and body composition with fertility have been extensively studied. Leptin is a peptide hormone secreted by adipocytes that circulates in proportion to fat mass. Disruption of leptin secretion (e.g., in *ob/ob* mice) or action (*db/db* mice) results in delayed puberty and hypogonadotropic hypogonadism in mice and humans. Furthermore, leptin administration reverses infertility associated with leptin deficiency.[28] Thus, leptin is an important humoral link between nutrition and fertility.

However, GnRH neurons lack receptors for the major metabolic signaling peptides, including insulin and leptin. Kisspeptin has been implicated as an intermediary between leptin signaling and GnRH function. *Ob/Ob* mice have reduced ARC levels of kisspeptin mRNA compared with wild-type controls,[29] and kisspeptin expression is increased following exogenous leptin administration. Fasting reduces hypothalamic kisspeptin mRNA and delays the onset of puberty in rats. In addition, central administration of kisspeptin to chronically undernourished prepubertal rats restores sexual maturation. However, only a small fraction of kisspeptin neurons express the leptin receptor, and genetic disruption of the leptin receptor in kisspeptin neurons does not inhibit reproduction in rodents suggesting the presence of additional circuits of leptin-responsive neurons involved in the metabolic control of GnRH neuronal activity that modulate the kisspeptinergic tone.[30]

To characterize the leptin-responsive neurons that mediate the effects of energy homeostasis on reproductive function in female mice, Martin et al.[31] selectively ablated the leptin receptor in neurons that were predominantly inhibitory (GABAergic) or excitatory (glutamatergic). They found that the ablation of leptin receptors in GABAergic neurons, but not in glutamatergic neurons, delayed sexual maturation and impaired reproductive function and fertility.[31] These data suggest that leptin-responsive GABAergic neurons are involved in conveying signals of energy balance via kisspeptin neurons to regulate reproductive function.

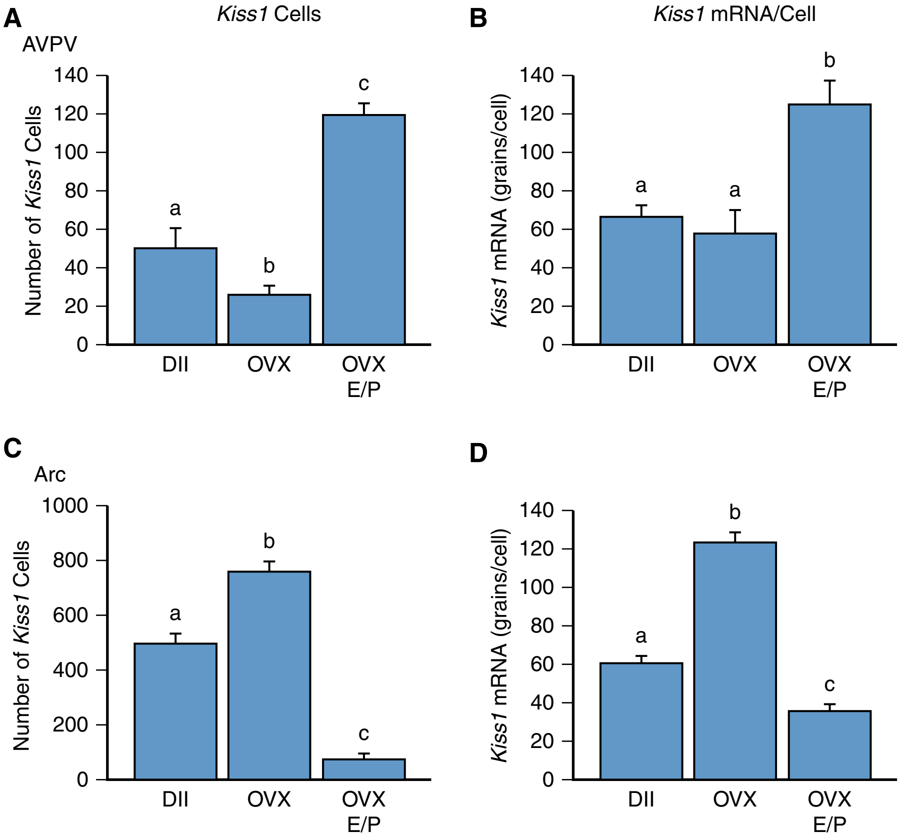


Figure 105.3 Distinct subpopulations of hypothalamic kisspeptin neurons mediate positive and negative feedback of estradiol in rats. **A-B**, In the hypothalamic anteroventral periventricular nucleus (AVPV), ovariectomy (OVX) *reduces* the number of kiss1-positive cells (**A**) and levels of kiss1 expression per cell (**B**). Conversely, experimental induction of the luteinizing hormone (LH) surge (OVX E/P) potently increases the number of kiss1-positive cells and levels of kiss1 expression within the AVPV. Collectively, these data suggest that kisspeptin signaling within the AVPV may mediate the positive feedback by estradiol associated with the LH surge. **C-D**, In the hypothalamic ARCuate nucleus, OVX increases the number of kiss1-positive cells (**C**) and levels of kiss1 expression per cell (**D**). Conversely, experimental induction of the LH surge (OVX E/P) reduces the number of kiss1-positive cells and suppresses kiss1 expression within the ARC. Collectively, these data suggest that kisspeptin signaling within the ARC may mediate negative feedback by estradiol on gonadotropin secretion. Values without common notations (a, b, c) differ significantly ($P < 0.05$). Values are the mean ± SEM. *DII,* Diestrus. (Smith JT, Popa SM, Clifton DK, Hoffman GE, Steiner RA. Kiss1 neurons in the forebrain as central processors for generating the preovulatory luteinizing hormone surge. *J Neurosci.* 2006;26(25):6687–6694. doi: 10.1523/JNEUROSCI.1618-06.2006.)

In contrast to leptin, the regulatory role of insulin in mammalian reproduction is currently unclear due to conflicting experimental data. Multiple studies have demonstrated that streptozotocin-induced insulin deficiency causes hypogonadotropic hypogonadism,[32] and type 1 diabetes is commonly associated with hypogonadotropic hypogonadism. A role for insulin signaling in the central regulation of reproduction is implied by observing that brain-specific deletion of insulin receptors in (NIRKO) mice causes hypogonadotropic hypogonadism.[33] However, specific deletion of the insulin receptor from either GnRH or kisspeptin neurons results in no obvious reproductive phenotype.[34] Taken together, insulin deficiency profoundly impacts gonadotropin secretion but has no obvious, direct effects on GnRH or kisspeptin signaling. The critical locations of insulin-sensitive targets within the brain influencing the reproductive system are therefore an important, unanswered question.

Kisspeptin cells communicate with neuropeptide Y (NPY) and proopiomelanocortin (POMC) cells, and this network of kisspeptin, NPY, and POMC neurons are involved in the coordination of brain control of reproduction and energy homeostasis.[35] NPY is an orexigenic peptide known to increase food intake, while POMC is a precursor of α-melanocyte–stimulating hormone (α-MSH), known for its anorectic effects. Neurons expressing NPY and POMC are in close apposition with kisspeptin neurons in the ARC.[35] Furthermore, central administration of an α-MSH agonist results in increased kisspeptin mRNA in the POA and increased plasma LH levels.

Mammalian target of rapamycin protein (mTOR) is a key player in the regulation of energy homeostasis, acting to reduce cell growth and differentiation in undernutrition. A link between mTOR and kisspeptin was suggested when antagonism of mTOR by rapamycin led to reduced kisspeptin expression in the ARC and reduced plasma LH levels.[36]

Starvation as well as feeding a high-fat diet (HFD) impair testosterone secretion. Fasting for 18 to 24 hours in monkeys is associated with lower and delayed testosterone response to intravenous kisspeptin injection.[37] In a study of the effects of an HFD in sexually immature and sexually mature male rats, the feeding of HFD was associated with increased body weight, impaired glucose tolerance, reduced testosterone levels, decreased hypothalamic kisspeptin receptor expression, and decreased LH responsiveness to kisspeptin.[38]

Regulation of Gonadotropins by Stress

Stress is a profound inhibitor of reproductive function in mammals, causing suppression of hypothalamic GnRH secretion and hypogonadotropic hypogonadism. Increased signaling of the hypothalamic neuropeptide corticotrophin-releasing factor (CRF) inhibits GnRH secretion and downregulates the expression of kisspeptin and its receptor in the ARC and medial POA (mPOA) of mice in response to central injection of CRF.[39] Reduced kisspeptin and kisspeptin receptor expression is also observed in response to other stressors including restraint, insulin-induced hypoglycemia, and injection of lipopolysaccharide (LPS).

The amygdala is a component of the limbic system implicated to have inhibitory effects on GnRH secretion. Manganese-enhanced magnetic resonance imaging (MRI) has shown that peripheral kisspeptin administration markedly reduces signal intensity in the amygdala in rodents. Furthermore, direct intra-medial amygdala (MeA) administration of kisspeptin increases LH secretion, whereas intra-MeA kisspeptin antagonist decreases LH secretion and LH pulse frequency.[40]

The bacterial endotoxin LPS reduces GnRH secretion in several mammalian species. Neonatal exposure to LPS causes delayed puberty and decreased kisspeptin mRNA expression in the mPOA of female rats. Furthermore, intraperitoneal administration of high-dose LPS in both ovariectomized and intact female rats reduces plasma LH levels and hypothalamic mRNA expression levels of kisspeptin and GnRH levels.[41]

Seasonal Regulation of Gonadotropin Secretion

Some animals use seasonal patterns of reproductive activity to maximize the survival of their offspring. During seasons of reproductive nonactivity, GnRH and gonadotropin secretion become quiescent in seasonal breeders. Recent evidence highlights the importance of the hypothalamic neuropeptide kisspeptin in mediating the seasonal activation of GnRH secretion. Increased hypothalamic kisspeptin expression has been reported in Syrian hamsters during long-day conditions, associated with increased sexual activity.[42] Furthermore, administration of kisspeptin-10 to Syrian hamsters under photoinhibitory conditions restored testicular, and therefore reproductive, activity.[43] Sheep are also known to be seasonal breeders, with increased reproductive activity during periods with short photoperiods to allow lambing in Spring. Increased ARC kisspeptin expression is observed in ewes during short photoperiods conditions. Conversely, during the long-day periods, kisspeptin expression in the ARC of ewes is reduced. Furthermore, kisspeptin administration in seasonally acyclic ewes (i.e., during a season of reproductive nonactivity) induces ovulation.[44] Kisspeptin receptor expression also varies seasonally. GnRH (and LH) responses to exogenous kisspeptin are greater in anestrus ewes than in seasonally active ewes during the luteal phase. In addition, kisspeptin receptor expression on GnRH neurons was greater during the nonbreeding season than during the breeding season. A recent study examined the expression of kisspeptin, NKB and DYN (which are coexpressed within hypothalamic KNDy neurons), in Syrian hamsters. They observed that all three neuropeptides were downregulated in the ARC under a short photoperiod.[45] The kisspeptin immunoreactivity is increased in the AVPV in Turkish hamsters exposed to long photoperiod conditions compared to hamsters exposed to short photoperiods as well as pinealectomized hamsters exposed to long photoperiods, suggesting a close relationship of pineal secretion of melatonin in the regulation of seasonal reproduction.[46]

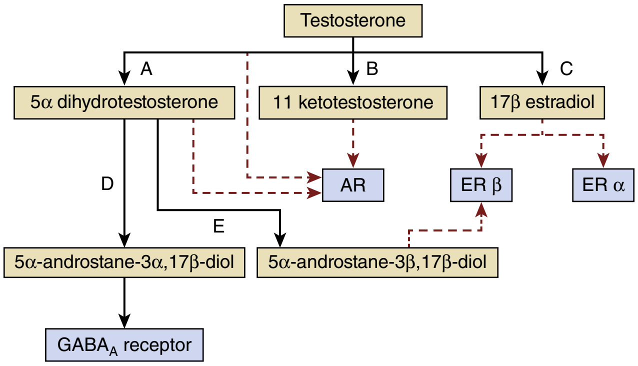


Figure 105.4 Testosterone can activate multiple signaling pathways through its conversion to other active metabolites. Testosterone exerts its biologic effects directly by itself or indirectly through its conversion to a number of metabolites: 5α dihydrotestosterone (DHT) a ligand for the androgen receptor, estradiol 17β a ligand for the estrogen receptor, 5α-androstane-3α,17β-diol, a modulator of gamma-aminobutyric acid (GABA$_A$) receptors[14] and 5α-androstan-3β,17β-diol a ligand for Erβ. The steroidal ligands are shown in *beige boxes* and the cognate receptors in *blue boxes*. A, SRD5A2 and 1; B, CYP11B2; C, CYP19A1; D, 3α-hydroxysteroid dehydrogenase; E, 3β-hydroxysteroid dehydrogenase.

KEY POINTS

- Increased levels of hypothalamic kisspeptin signaling trigger seasonal, reproductive activity in hamsters and sheep.
- Sepsis, hypoglycemia, and other experimental models of physiological stress suppress hypothalamic kisspeptin signaling.
- Starvation induces hypogonadotropic hypogonadism by suppressing hypothalamic kisspeptin signaling.

ANDROGEN PHYSIOLOGY

Testosterone is an important regulator of the masculinization of the structures that develop from the Wolffian ducts and the urogenital sinus; the development and maintenance of secondary sex characteristics; spermatogenesis; muscle and bone mass; hematopoiesis; metabolism; and many reproductive and nonreproductive behaviors. Testosterone exerts its biologic effects either directly through genomic and nongenomic mechanisms or indirectly through its conversion to one of its metabolites: 5α-dihydrotestosterone (DHT) a ligand for the AR, estradiol 17β a ligand for the ER, 5α-androstane-3α,17β-diol, a modulator of GABA$_A$ receptors[14] and 5α-androstan-3β,17β-diol, a ligand for ERβ (Fig. 105.4).[47–52] Thus, testosterone may expose different tissues to several biologically active metabolites that mediate the observed responses through multiple signaling pathways.

The Role of Aromatization in Mediating the Biologic Effects of Testosterone

Important insights into the role of estradiol in mediating testosterone's physiologic effects have emerged from studies of mice with genetic disruption of the CYP19A1 (aromatase enzyme), ERα, and ERβ genes, and human studies in which the conversion of exogenous testosterone was blocked by the administration of an aromatase inhibitor.[49,53–60] The effects of testosterone on bone resorption, epiphyseal closure, high-density lipoprotein (HDL) cholesterol, atherosclerosis progression, sexual desire, and fat mass require its conversion by the product of the CYP19A1 gene (aromatase) to estradiol 17β.[47,49,53,54] Mice with genetic disruption of the P450-linked CYP19 aromatase gene and humans with CYP19A1 mutations have increased fat mass and reduced muscle

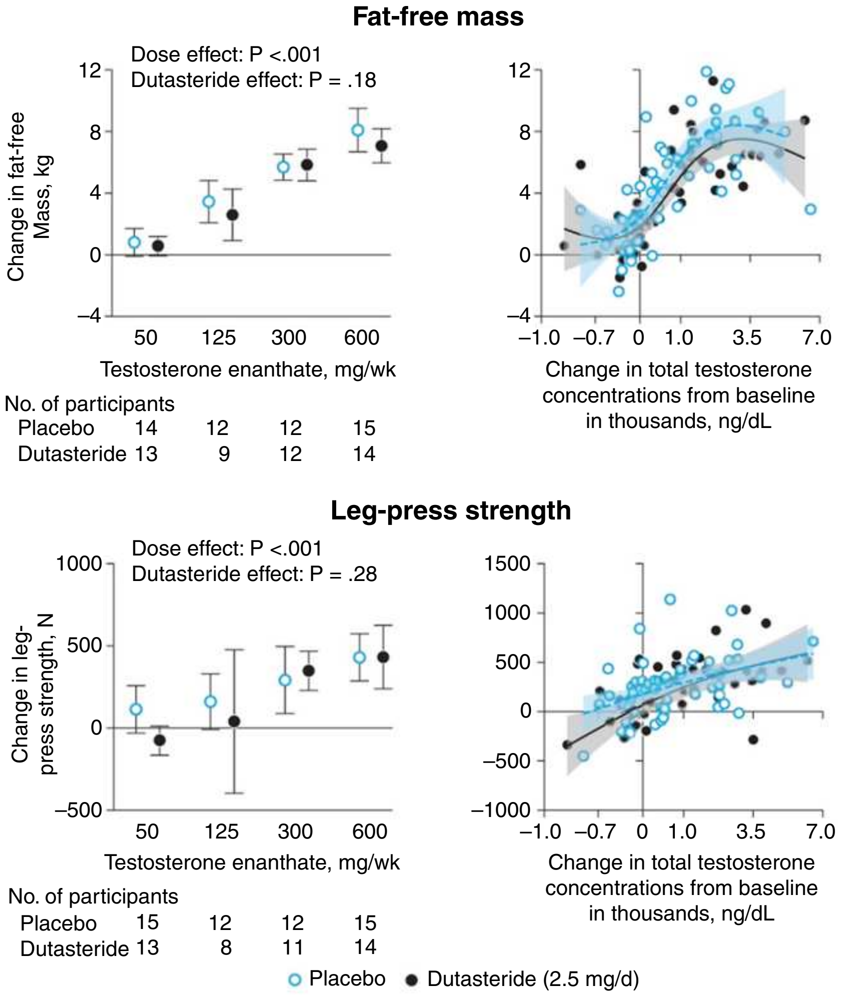


Figure 105.5 The role of 5-alpha-dihydrotestosterone in mediating testosterone's effects on skeletal muscle mass and strength in healthy young men. In this randomized trial, healthy young men were administered a long-acting gonadotrophin-releasing hormone (GnRH) agonist to suppress endogenous testosterone and randomized to either placebo or dutasteride (dual inhibitor of steroid 5α-reductase types 1 and 2) daily, plus 50-, 125-, 300-, or 600-mg testosterone enanthate weekly for 20 weeks. Changes in fat-free mass *(upper panel)* and leg-press strength *(lower panel)* were related to testosterone dose but did not differ between placebo and dutasteride groups. The relationship between change in total testosterone levels and change in fat-free mass and leg-press strength *(right panels)* did not differ between men assigned to placebo or dutasteride arms. (Reproduced with permission from Bhasin S, Travison TG, Storer TW, et al. Effect of testosterone supplementation with and without a dual 5alpha-reductase inhibitor on fat-free mass in men with suppressed testosterone production: a randomized controlled trial. *JAMA*. 2012;307(9):931–939.)

mass and exhibit insulin resistance,[53,61] suggesting that aromatization of testosterone to estradiol is important in mediating androgen effects on body composition. Testosterone's effects on lean mass, muscle size, and strength are not attenuated when its conversion to estradiol is blocked by the administration of an aromatase inhibitor.[49] However, testosterone's effects on fat mass, osteoclastic bone resorption as well as osteoblastic bone formation, and sexual desire are mediated by estradiol.[49] Mendelian randomization studies have revealed a stronger association of polymorphisms in the aromatase gene and genetically determined estradiol levels with bone mineral density and bone fracture risk than with testosterone levels.[62,63]

The Role of the Steroid 5α-Reductase System in Mediating the Biologic Effects of Testosterone

The conversion of testosterone to DHT by a family of steroid 5α-reductase enzymes (SRD5A1 and SRD5A2) is obligatory for mediating its effects on the differentiation of the urogenital sinus and the genital tubercle.[64] Testosterone's conversion to DHT also is believed to mediate its effects on hair growth, scalp hair loss, prostate growth, and acne. Men with benign prostatic hypertrophy who are treated with a steroid 5α-reductase inhibitor do not experience muscle or bone loss.[65] Similarly, individuals with congenital steroid 5α-reductase deficiency have normal muscle development at puberty.[66] In healthy young men in whom endogenous testosterone production was suppressed by the administration of a long-acting GnRH agonist, the anabolic effects of graded doses of testosterone on muscle mass and muscle strength, hemoglobin and hematocrit; bone markers; and sexual function were not blocked by dutasteride, a potent dual inhibitor of steroid 5α-reductase type 1 and 2 enzymes.[48] (Fig. 105.5) Thus, testosterone's conversion to DHT is not required for mediating its effects on muscle mass and strength, or bone mineral density, plasma lipids, and erythropoiesis.

Activation of Other Signaling Pathways by Metabolites of Testosterone

In specific brain regions, testosterone is converted locally to DHT, which is further converted in some cell types to 5α-androstane-3α,17β-diol, a modulator of GABA_A receptors and 5α-androstan-3β,17β-diol, a ligand for ERβ.[50–52,67–69] These metabolites play an important role in mediating testosterone's effects on mood and behavior and Alzheimer's disease neuropathology[50,51] (Fig. 105.4).

KEY POINTS

- Although testosterone is the predominant androgen in most mammalian species, it can also activate other signaling pathways through its conversion to its metabolites, such as 5α-dihydrotestosterone, a ligand for the androgen receptor; 17β estradiol, a ligand for the estrogen receptor; 5α-androstane-3α,17β-diol, a modulator of GABA_A receptors; and 5α-androstan-3β,17β-diol, a ligand for ERβ.
- The effects of testosterone on bone resorption, epiphyseal closure, high-density lipoprotein cholesterol, atherogenesis, sexual desire and some other behaviors, and fat mass, distribution, and metabolism require its conversion to 17β estradiol.
- The conversion of testosterone to 5α-DHT is required for the masculinization of the urogenital sinus and the genital tubercle but is not obligatory for mediating its effects on muscle mass and strength, hematopoiesis, and bone.

Biosynthesis of Testosterone and Other Androgens

More than 90% of circulating testosterone in men is derived from its biosynthesis in the Leydig cells that reside in the interstitial compartment of the testes; the remainder is derived from steroidogenic precursors in the adrenal gland. Cholesterol within the steroidogenic cells is derived mostly from three sources: its *de novo* synthesis in the endoplasmic reticulum; its mobilization from the plasma membrane; uptake of circulating cholesterol esters mediated by receptors found on the plasma membrane; and mobilization of cholesterol stored in lipid droplets.[70] In the mitochondria, the cholesterol available for steroidogenesis is bound to the cholesterol-binding domain of the translocator protein (TSPO)[71] Under the influence of LH in the testes and adrenocorticotrophic hormone in the adrenal gland, cholesterol bound to the cholesterol-binding domain of the TSPO is released and translocates to the inner mitochondrial membrane, where it is converted by the action of cholesterol side-chain cleavage enzyme (CYP11A1) and electron transferring proteins to pregnenolone.[72] The transport of intracellular cholesterol by the steroidogenic acute regulatory (StAR) protein from the outer mitochondrial membrane to the inner mitochondrial membrane is the rate-limiting process in testosterone biosynthesis. Mutations of the StAR protein are associated with congenital lipoid adrenal hyperplasia, a rare form of congenital adrenal hyperplasia (CAH) characterized by very low adrenal and gonadal steroids.[73]

The subsequent steps in testosterone biosynthesis take place in the endoplasmic reticulum. In the human testes, testosterone biosynthesis takes place predominantly via the delta-5 pathway, which involves the conversion of pregnanolone to 17α pregnanolone, by the 17α hydroxylase and the subsequent conversion of 17α pregnanolone to dehydroepiandrosterone (DHEA) by the 17,20-lyase reactions. The 17α hydroxylase and 17,20 lyase reactions are catalyzed by a single protein, CYP17A1; posttranslational modification (phosphorylation) of this enzyme and the presence of specific enzyme cofactors, such as cytochrome B5, confer 17,20-lyase activity selectively in the testis and zona reticularis of the adrenal gland. Abiraterone is a dual inhibitor of 17α hydroxylase and 17,20 lyase activities, which play an important role in local androgen synthesis in castration-resistant prostate cancers.

Although CYP17A1 is able to catalyze the conversion of progesterone to 17α-hydroxyprogesterone, most of δ4-androstenedione in humans is not derived from 17α-hydroxyprogesterone but rather from the conversion of 17α-hydroxy pregnenolone to DHEA in the δ5 pathway and further conversion of DHEA to δ4-androstenedione by the catalytic action of 3β-hydroxysteroid dehydrogenase.[74]

The Classical and the "Backdoor" Pathway for the Synthesis of DHT. Testosterone is typically converted in the testes and in some peripheral tissues to the more potent DHT by a family of steroid 5α-reductase enzymes.[64] At least two isoforms of steroid 5α-reductase, SRD5A1 and SRD5A2, have been described; the well-described kindreds with 5α-reductase deficiency have had mutations in SRD5A2, the predominant form in the prostate and the skin.[64] Finasteride predominantly inhibits SRD5A2, while dutasteride is a dual inhibitor of both SRD5A1 and SRD5A2.

DHT can also be derived through the backdoor pathway in which progesterone is converted by SRD5A1 to 5α-dihydroprogesterone, allopregnanolone, 17-hydroxy allopregnanolone, androsterone, androstanediol, and eventually to DHT. Recent reports of mutations in AKR1C2/4 genes in undervirilized 46, XY individuals suggest that the backdoor pathway for DHT formation, which was originally described in the tammar wallaby, is active in the human fetal testis.[75] The placental progesterone serves as a substrate for the synthesis of androsterone via the backdoor pathway in several nongenital tissues, which is then converted to DHT in the genital tubercle.[76]

The Biosynthesis of 11-Oxygenated Androgens. 11-Oxyandrogens are 19-carbon steroids that contain an oxygen atom on carbon 11 and have androgenic activity. 11β-hydroxysteroid dehydrogenase types 1 (CYP11B1) and 2 (CYP11B2) are expressed in the Leydig cells, the theca cells, and the *zona reticularis* of the adrenal gland[77–79]; testosterone as well as delta 4 androstenedione can serve as substrates for CYP11B1, resulting in the formation of 11β-hydroxytestosterone and 11β-hydroxyandrostenedione, respectively. CYP11B2 then converts 11β-hydroxytestosterone and 11β-hydroxyandrostenedione to 11-ketotestosterone and 11-ketoandrostenedione, respectively, mostly in peripheral tissues. Delta 4 androstenedione is the predominant substrate for CYP11B2 in the adrenal gland due to the high levels of delta 4 androstenedione produced in the zona reticularis although most of the 11-keto androgens are produced in the peripheral tissues.[77,79] 11-ketoandrostenedione can also be converted by AKR1C3 to 11-ketotestosterone. 11-Oxygenated steroids—11-ketotestosterone and 11-ketodihydrotestosterone—are potent androgens, with potencies equivalent to those of testosterone and dihydrotestosterone.[79] In fact, 11-ketotestosterone circulates in the plasma of men and women in appreciable concentrations (~20 to 30 ng/dL) without sexual dimorphism.[78] Growing evidence suggests that 11-oxygenated androgens may play an important role in several disorders of androgen excess, such as premature adrenarche, polycystic ovary syndrome, CAH, as well as in androgen-dependent cancers, such as castration-resistant prostate cancer.[79]

KEY POINTS

- The transport of intracellular cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane by the mediation of the StAR and translocator (TSPO) proteins is the rate-limiting step in testosterone biosynthesis.
- In adults, testosterone is converted in the testes and in some peripheral tissues to 5α-DHT by one of at least two isoforms of steroid 5α-reductase, SRD5A1 and SRD5A2; in the human fetus, 5α-DHT is derived mostly through the "backdoor pathway" in which progesterone is converted by SRD5A1 to 5α-dihydroprogesterone, allopregnanolone, 17-hydroxy allopregnanolone, androsterone, androstanediol, and eventually to DHT.
- 11-Oxygenated steroids—11-ketotestosterone and 11-ketodihydrotestosterone—potent androgens, are produced in appreciable amounts by the adrenal gland, testis, and some peripheral tissues by the action of CYP11B1 and CYP11B2.

TABLE 105.2 Structure Activity Relationship in Androgens

Structure Activity Relationship	Compounds	Chemical Structure
The basic structure of testosterone and the carbon numbering system	Testosterone	<Table Fig 105.2a>
17α-alkyl substitution retards presystemic first-pass metabolism	17α-Methyltestosterone	<Table Fig 105.2b>
Removal of 19 methyl group increases anabolic activity	19-Nortestosterone (nandrolone) series of compounds	<Table Fig 105.2c>
7α-Alkyl substitution increase anabolic activity and prevent its 5α reduction	7α-Methyl-19-nortestosterone	<Table Fig 105.2d>
Esterification of the 17β hydroxyl group renders the molecule more hydrophobic and extends the duration of action when injected into oil	Testosterone enanthate	<Table Fig 105.2e>
C2 carbon atom with an oxygen atom in 17α-alkylated AAS can improve metabolic stability	Oxandrolone	<Table Fig 105.2f>
Introduction of double bonds between the C9 and C10 positions and between the C11 and C12 positions can confer greatly increased potency	Trenbolone	<Table Fig 105.2g>

Structure Activity Relationship in Steroidal Androgens

Testosterone is a 19-carbon androgen with a cyclopentanophenanthrene ring structure that contains a 17-hydroxyl group, a 3-keto group, and a double bond between carbon 4 and 5 (Table 105.2).[80–82] Structural modifications of the testosterone molecule can modify its AR-binding affinity, AR stabilization, coactivator recruitment, nuclear translocation, DNA-binding affinity, presystemic metabolism, specificity as a substrate for the steroid 5α-reductase and the aromatase enzymes and its tissue specificity, and metabolic clearance.[81,82]

The addition of a methyl or an ethyl group to the 17α carbon position inhibits the presystemic metabolism of testosterone, extending its half-life and rendering it orally active; 17α-methyltestosterone is an example (Fig. 105.2).[80–82] However, orally administered 17α-alkylated androgens are potentially hepatotoxic and markedly suppress plasma HDL cholesterol and their use to treat hypogonadism is not recommended.

Removal of the methyl group in C19 position increases the anabolic activity; this modification has formed the basis of the 19-nortestosterone (nandrolone) series of compounds that were approved for the treatment of anemia and osteoporosis and are still widely used as anabolic drugs by recreational bodybuilders and athletes.[80–82]

Alkyl substitutions in the 7 position also increase the anabolic activity and reduce susceptibility for 5α reduction. Among the many 7α alkylated compounds that have been developed, 7α-methyl-19-nortestosterone (MENT) has undergone human trials.

The substitution of C2 with oxygen in oxandrolone increases the stability of the 3-keto group and its anabolic activity and prevents its aromatization. Oxandrolone also has a 17α-methyl group that renders it orally active. Because of its high anabolic activity relative to its androgenic activity, it is often used by female athletes.

In stanozolol, the 3,2-pyrazole group on the first cyclohexane ring enhances AR binding and the C17 methyl group makes it orally active. The 2-hydroxymethylene group in oxymetholone adds stability to the 3-keto group and increases its anabolic activity and the 17α methyl substitution prevents its presystemic metabolism and makes it orally active.

Orally administered testosterone is absorbed rapidly, but it undergoes rapid presystemic metabolism in the intestinal lumen and wall and in the liver. The esterification of the 17β-hydroxyl group with a carboxylic acid group (ester linkage) increases the hydrophobicity of the molecule. Testosterone enanthate, cypionate, and undecanoate are examples of testosterone esters that are currently used in clinical practice. The longer the side chain, the greater the hydrophobicity of the molecule; when injected intramuscularly in oil, these hydrophobic esters are released slowly from the oil depot in the muscle into the aqueous plasma thereby increasing its duration of action.[83] The half-life of testosterone esters in the human plasma is short; their long duration of action is largely due to their slow release from the oil depot in the muscle rather than from slower hydrolysis in plasma.

The Regulation of Testosterone Production by LH

Testosterone production and secretion by the Leydig cells are regulated by LH through a feedback and feedforward system. As discussed earlier, the negative feedback effects of testosterone, estradiol, and progesterone are mediated through KNDy neurons in the POA by inhibition of kisspeptin release[84]; testosterone also exerts direct stimulatory effect on pituitary FSH secretion.[26]

LH binds to a specific rhodopsin-like G-protein-coupled LH/chorionic gonadotropin receptor (LH/CGR), induces a conformational change in the receptor, the dissociation and release of the G-protein stimulatory Gs alpha subunit, activation of the adenylyl cyclase, and a rapid increase in the cyclic adenosine monophosphate (cAMP) and intracellular calcium.[85,86] The second messenger cAMP activates protein kinase A and phosphorylation of the transcription factor cAMP response element-binding (CREB) protein that binds target DNA sequences and regulates the transcription of genes encoding the steroidogenic enzymes in a cell type-specific manner. The activation of pERK1/2 is required for the steroidogenic response to LH in Leydig cells.[86] The signaling cascade is triggered by the Gaq protein activating phospholipase C, with subsequent cleavage of phosphatidylinositol bisphosphate to diacyl glycerol and inositol trisphosphate that binds calcium channels in the endoplasmic reticulum, resulting in calcium release.[85–87] LH receptor mutations cause Leydig cell hypoplasia or agenesis, underscoring the importance of this pathway for Leydig cell development and function.[88]

KEY POINTS

- LH stimulates testosterone production by binding to a specific G-protein–coupled LH/chorionic gonadotropin receptor (LH/CGR) to activate adenylyl cyclase, increase cAMP, and activate protein kinase A.
- LH also activates phospholipase C resulting in calcium release. The activation of pERK1/2 also plays an important role in the steroidogenic response to LH in Leydig cells.

Transport and Bioavailability

The circulating sex hormones—testosterone, dihydrotestosterone, and estradiol—are transported in human plasma bound predominantly to human serum albumin and sex hormone-binding globulin (SHBG),

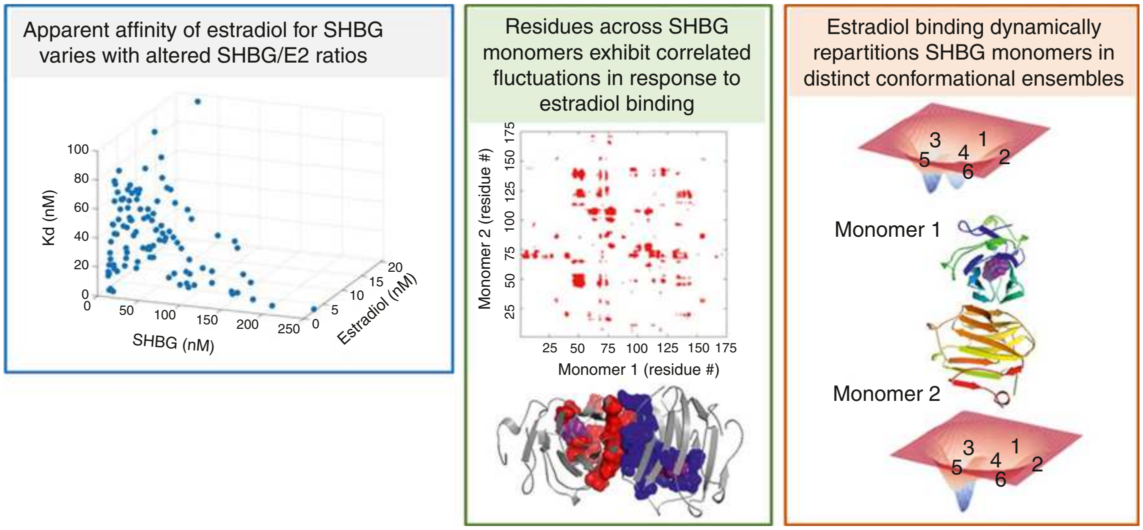


Figure 105.6 *Left,* The apparent binding affinity of estradiol varies as the concentrations of estradiol and sex hormone–binding globulin (SHBG) are varied. *Middle,* Dynamic cross-correlation matrices for SHBG dimer. The quadrants depicting the intermonomer correlation show changes in coordinated movement of the residues in the two monomers. *Red color* intensity stands for strength of residue correlation. Collectively, these data show that allosteric coupling in binding estradiol is manifested through coordinate, dynamic rearrangement in SHBG monomers. *Right,* Allosteric modulation of relative probabilities of the population of conformational clusters in response to estradiol binding (the five major conformational clusters are depicted). Contrary to the prevailing linear model, two unliganded SHBG monomers populate distinct conformational states. The populations of the liganded and unliganded monomers redistribute upon estradiol binding to either of the monomers. (Reproduced with permission from Jasuja R, Spencer D, Jayaraj A, et al. Estradiol induces allosteric coupling and partitioning of sex-hormone-binding globulin monomers among conformational states. *iScience.* 2021;24(6):102414.)

and to a lesser degree to α1-acid glycoprotein (orosomucoid), and cortisol-binding globulin (CBG),[89] and in rodents that lack SHBG, possibly also to lipoproteins. These binding proteins regulate the distribution, metabolism, elimination, and biologic activity of testosterone, dihydrotestosterone, and estradiol.[89]

Sex Hormone-Binding Globulin. SHBG is a homodimeric glycoprotein with a molecular mass of approximately 90 kDa that is produced in the liver of higher mammals but not in rodents and can bind to all three major sex steroid hormones.[90] In men, approximately 44% of testosterone is bound to SHBG, while in women, a greater fraction—approximately 66%—is bound to SHBG.[91] For a number of years, it was believed that testosterone's binding to each of the two SHBG monomers was linear and equivalent and conformed to a 1:1 stoichiometry with a binding affinity of ~1 nM.[92–94] Recent studies using modern biophysical techniques have shown that the binding of testosterone to SHBG is a nonlinear dynamic process such that the binding affinity and conformational states of the two monomers in unliganded as well as liganded states are not equivalent (Fig. 105.6).[95]

Some studies have suggested that extracellular SHBG may bind to a membrane receptor that is coupled to an adenylate cyclase-mediated signaling cascade or may even activate AR signaling in the prostate.[96,97] It has also been suggested that SHBG may transport testosterone intracellularly through a megalin-mediated endocytic process.[98] However, further independent validation of these findings has not emerged.

Does SHBG Exert Biologic Effects Other Than as a Transport Protein?. Whether SHBG exerts biologic effects in addition to serving as a transport protein for sex hormones is unknown. Serum SHBG levels and SHBG polymorphisms that are associated with higher SHBG levels are independent predictors of the risk of type 2 diabetes, metabolic syndrome, and bone fractures, even after adjusting for testosterone concentrations.[99–101]

Human Serum Albumin. Human serum albumin comprises three homologous domains; domains II and III contain binding pockets that are formed mostly of hydrophobic and positively charged residues in which a variety of compounds bind.[102,103] Human serum albumin possesses multiple binding sites for several biomolecules, including nonesterified fatty acids and many commonly used drugs. However, it has been generally believed that testosterone binds to human serum albumin at a single site on domain IIA with low-to-moderate affinity ($K_a \approx 2.0 - 4.1 \times 10^4\ M^{-1}$ at 37°C) and 1:1 stoichiometry.[92–94,104] Studies using two-dimensional nuclear magnetic resonance (2D NMR), fluorescence spectroscopy, bis-ANS partitioning, and molecular modeling have provided evidence of multiple, allosterically coupled binding sites for testosterone on human serum albumin (Fig. 105.7).[105] Testosterone shares these binding sites on human serum albumin with free fatty acids and many commonly used drugs, such as ibuprofen and coumadin, which could displace testosterone from human serum albumin under various physiological states or disease conditions, affecting its bioavailability.[105]

The Significance of the Unbound Fraction: The Free Hormone Hypothesis. The free hormone hypothesis postulates that only the unbound or free testosterone fraction is able to diffuse into the target cell and exert its biologic effect.[106] Consistent with the free hormone hypothesis, the Endocrine Society Guideline suggests measuring free testosterone concentrations in men with conditions in which alterations in binding protein concentrations are suspected.[107] The validity of the free hormone hypothesis has been debated extensively. In an effort to test the free hormone hypothesis, Laurent et al.[108] generated transgenic mice expressing a human SHBG transgene. The SHBG transgenic mice exhibited higher total testosterone and estradiol

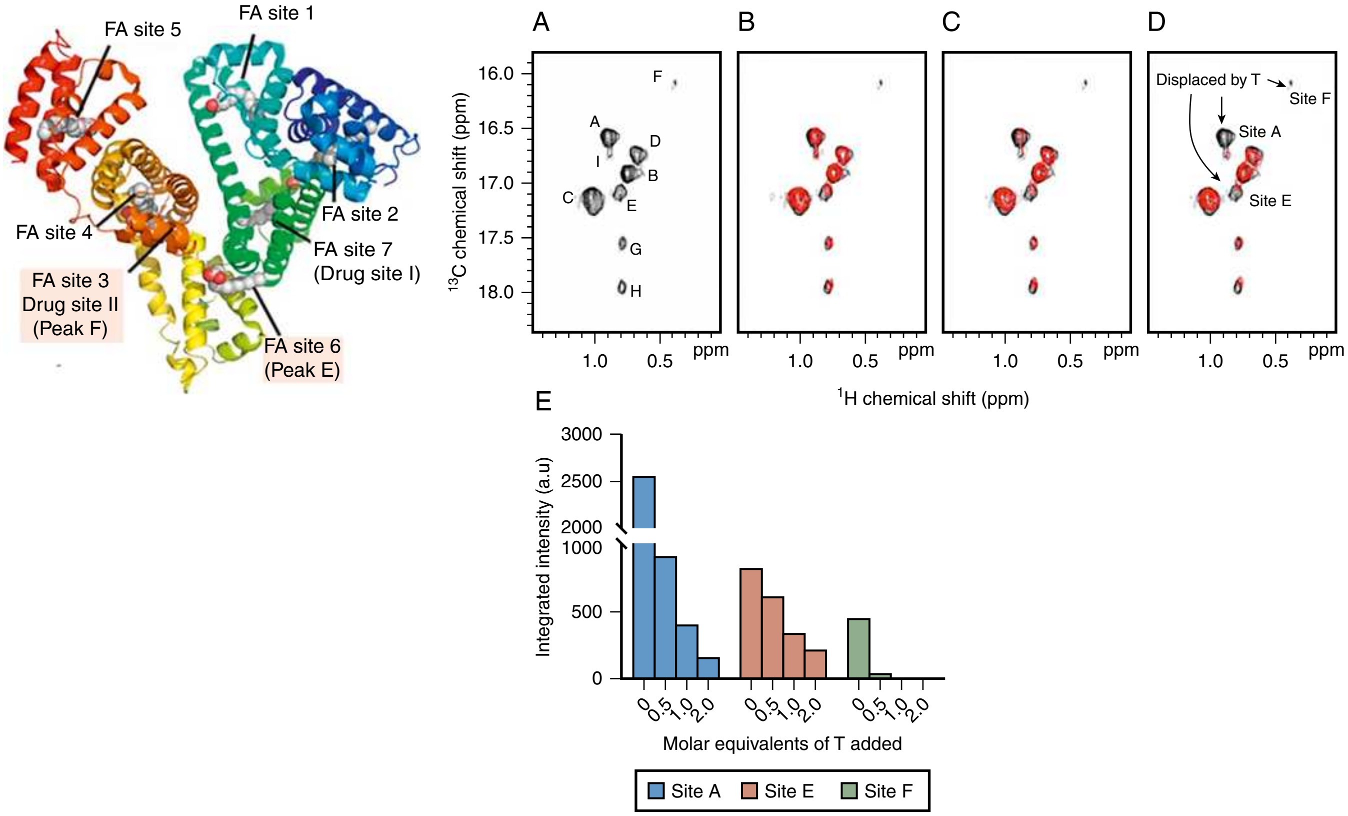


Figure 105.7 The crystal structure of human serum albumin (HSA) and testosterone's binding to multiple fatty acid-binding sites on HAS. *Left,* HSA crystal structure showing the fatty acid–binding sites that have been resolved in the crystal structure. *Right,* **A**, Nine resonances were observed from ^{13}C-OA binding to HSA (2 mM OA, 0.5 mM HSA) bound sites. **B-D**, Decrease in nuclear magnetic resonance signal from binding pockets corresponding to peaks A, E, and F as the OA is displaced by increasing testosterone concentrations. Red resonances in **B**, **C**, and **D** are from sites that are occupied by OA in the presence of testosterone. **A**, The spectra obtained from 4:1 ratio of OA:HSA in absence of testosterone; **B**, in the presence of 0.5 molar equivalents of testosterone; **C**, upon addition of 1 molar equivalent of testosterone; and **D**, in the presence of 2 molar equivalents of testosterone. **E**, The four spectra were analyzed using CCPNMR software to determine the peak intensity at corresponding concentrations of testosterone. (Reproduced with permission from the Endocrine Society (Jayaraj A, Schwanz HA, Spencer DJ, et al. Allosterically coupled multisite binding of testosterone to human serum albumin. *Endocrinology*. 2021;162(2).)

concentrations compared to wild-type controls but had normal free testosterone concentrations.[108] The SHBG transgenic mice had higher LH concentrations and lower prostate and seminal vesicle volumes than wild-type controls. When wild-type mice and SHBG transgenic mice were castrated and given equivalent doses of testosterone implants, free testosterone levels, as well as the wet weights of androgen-sensitive organs (seminal vesicles and levator ani/bulbocavernosus), were significantly lower in SHBG transgenic mice than in wild-type mice.[108] These data support the notion that SHBG regulates the bioavailability of testosterone in androgen-responsive tissues.

Genetic polymorphisms in the human *SHBG* gene have been associated with SHBG levels and binding affinity, as well as with total testosterone levels.[109,110] Patients with polymorphisms associated with undetectable SHBG concentrations offer a unique opportunity to evaluate the free hormone hypothesis. Vos et al.[111] described a brother and sister who were homozygous for a missense mutation within the *SHBG* gene (p.G224R SHBG) and had undetectable SHBG in their plasma. Both affected siblings had normal gonadal development and function suggesting a limited role for SHBG in sexual maturation and physiology. The male subject with the undetectable SHBG had very low total testosterone but normal free testosterone level, normal LH and FSH, and normal pubertal development and spermatogenesis. However, the male subject reported symptoms similar to those reported by men with testosterone deficiency, such as low libido, decreased spontaneous morning erections, fatigue, muscle weakness, depressed mood, and sleep disturbance. The women with undetectable SHBG levels have been reported to exhibit hyperandrogenism, hirsutism, and amenorrhea.[112]

In an analysis of the European male Aging Study,[113] low free testosterone was more strongly associated with clinical features of testosterone deficiency than total testosterone. In another analysis of men who were evaluated for hypogonadism,[114] 62% of patients with low total testosterone had a normal calculated free testosterone concentration. Thus, accurate measurement of serum free testosterone concentration is necessary for most men to confirm the diagnosis of testosterone deficiency.[114]

KEY POINTS

- Binding of circulating sex hormones to SHBG, human serum albumin, CBG, and orosomucoid regulates their distribution, metabolism, elimination, and biologic activity.
- Testosterone's binding to SHBG dimer is a dynamic, nonlinear process that involves allosteric interactions between the two SHBG monomers such that the binding affinities and conformation of the two monomers in the liganded or unbound state are not equivalent.
- Testosterone binds to multiple binding sites of human serum albumin that also bind fatty acids and commonly used drugs.

Testosterone Metabolism

As discussed earlier, testosterone is converted to DHT and estradiol in many peripheral tissues. In addition, testosterone is converted mostly in the liver to various inactive metabolites via enzymatic oxidation, reduction, and hydroxylation of the A, B, C, and D rings and excreted in the urine either as a glucuronide or as a sulfate; only 2% is excreted as unconjugated testosterone.[115,116] The major urinary metabolites of testosterone, androsterone, and etiocholanolone are formed by the catalytic action of the enzyme 17β-hydroxydehydrogenase and are excreted as 17-ketosteroids. These inactive metabolites are conjugated in the liver and kidney with either glucuronic acid through the catalytic actions of UDP-glucuronosyltransferase enzymes or sulfotransferase.[115,116] Some synthetic androgens such as oxandrolone are excreted mostly in the unconjugated form.[115] Detection of androgens or their metabolites in the urine is the basis of current drug testing.

Mechanisms of Testosterone's Action

Testosterone and the Androgen Receptor. Many physiologic actions of testosterone and 5α-DHT, including its effects on sexual differentiation of urogenital structures, spermatogenesis, muscle and bone, immune system, and erythropoiesis, are mediated through the classical AR; the nongenomic actions of testosterone are less well understood. The AR, a member of the nuclear receptor superfamily, is a 919-aa, 110-kDa, ligand-inducible transcription factor that regulates the transcription of androgen-responsive genes and mediates many of the biologic effects of testosterone and other androgens.[117,118]

AR possesses six characteristic structural domains A-B, C, D, E, and F.[117–119] The domain A/B includes activation function 1 (AF-1) (residues 101 to 370) that is required for full ligand-activated transcriptional activity; AF-5 (residues 360 to 485) that is responsible for its constitutive activity without ligand binding; and a dimerization surface involving residues 1 to 36 containing the FxxLF motif and residues 370 to 394 that interact with the ligand-binding domain (LBD).[119–121] The AF-1 transactivation domain, located in the N-terminal region, binds different coregulator proteins that enable liganded AR to subserve diverse physiologic functions.[119–121] Domain C serves as the DNA-binding domain (DBD). The DBD is organized in two zinc finger binding motifs that direct binding as homodimers. The target androgen response element (ARE) typically consists of one or two DNA consensus sequences that are arranged as inverted repeats—the distance between these repeats determines the binding specificity. The DBD (545 to 675 aa) is well conserved and binds with high affinity to the AREs and other transcription factors bound to the AREs.[119–121] Domain D serves as the hinge region that separates the DBD from the LBD, provides flexibility to the AR so it can assume different conformations depending upon the bound ligand, and contains a nuclear localization signal.[122] Domain E contains the LBD, the AF-2, and a nuclear export signal.[118,121,122] The highly conserved LBD in the C terminus (690 to 919 aa) confers ligand specificity for androgens and also contains domains that can mediate the binding of heat shock proteins, receptor dimerization, and nuclear localization.[122] The N-terminal region of domain E contains another activation domain, AF-2, that is responsible for agonist-induced activity (see later).[122] The N-terminal region (1 to 514 aa) is not well conserved, has been difficult to crystallize, and contains several unstructured domains.

AR contains two AF regions located in the N-terminal (AF-1) and the C-terminal halves of the protein (AF-2) that regulate the interactions between N-terminal and C-terminal domains of the AR as well as the recruitment of coactivator and corepressor proteins in a ligand-specific manner.[123–125] The AR AF-2 plays two functional roles: (1) in the unliganded state, it associates with FxxLF (residues 23 to 27) and WxxLF (residues 433 to 437) motifs observed in the N-terminal domain (NTD) to keep the AR in an inactive conformation and (2) upon ligand binding, the liganded AR recruits coregulators that possess LxxLL and FxxLF motifs, including p160 family, ARA55, ARA55, ARA70, NCoRs, and steroid receptor coactivators (SRCs).[125] Selective recruitment of the coregulators elicits ligand-specific, tissue-specific, AR-mediated transcriptional activity.[125] Two transcription activation units, Tau-1 and Tau-5, within the AF-1 are obligatory and sufficient for the activity of the isolated AR-NTD and the full-length AR[126]; mutations in Tau-1/Tau-5 render the AR protein inactive.[126]

Unliganded AR is located primarily in the cytoplasm; in response to ligand binding, AR undergoes distinct conformational changes, leading to dissociation of heat shock proteins, recruitment of coregulators and chaperone proteins, translocation to the nucleus, dimerization, chromatin remodeling, DNA binding, and regulation of transcriptional activity of AR-responsive genes.[127,128] The various domains of AR act in concert to conformationally gate the transcriptional response to agonist or antagonist binding.[128] Collectively, the transcriptional activity of AR is regulated by a spatiotemporal interplay between domains and motifs. Ligand binding and coregulator protein-induced conformational dynamics enable the domains to exhibit multiple functional roles.

The crystal structure of C-terminal LBD of the nuclear hormone superfamily displays highly conserved arrangement of 12 α helices.[129–131] While AR contains 11 helices, they are still numbered from 1 to 12 (helix 2 is skipped) to reflect the shared functional organization with helix 12 acting as a flexible, conformational gate.[121,129–131] Ligand binding is associated with significant rearrangement in the immediate vicinity of the residues in the binding pocket as well as in distant residues in other domains. The DBD binds with high affinity to the AREs to induce transcriptional activity. While AR shares affinity for consensus sequence 5’-TGTTCT-3’ with other nuclear receptors, such as GR and PR, it also has selective ARE, including PB-ARE2 and Slp-HRE2.[121,131] These interactions of the DBD and ARE regulate the recruitment of enhancers/promoters to the macromolecular transcriptional assembly.[131–133] The crystal structures of liganded LBD and DBD have been described, but no crystallographic or solution NMR structures have been solved for the NTD or its fragments. It is conceivable that the conformational flexibility of NTD enables intra- and interdomain steric interactions.

Polymorphisms in Trinucleotide Repeats in the Androgen Receptor. The human AR gene contains two polymorphic trinucleotide repeats of CAG and GGC, which code for polyglutamine and polyglycine tracts in exon 1 of AR.[134] The length of the polyglutamine tracts has been associated negatively with the transcriptional activity of AR. Very long polyglutamine tracts have been associated with a rare inherited X-linked motor neuron disorder, the Spinal and Bulbar Muscular Atrophy, or Kennedy's disease.[135] The AR protein with very long polyglutamine tracts forms aggregates that are toxic to the neuronal cells.[135]

Mechanisms of Testosterone's Effects on the Skeletal Muscle. Testosterone administration increases skeletal muscle mass, maximal voluntary strength, and muscle power. Testosterone-induced increase in skeletal muscle mass is related to the administered dose and circulating testosterone concentration[136,137] and is associated with the hypertrophy of both type 1 and type 2 muscle fibers.[138–140] Testosterone does not increase the numbers or relative proportion of type 1 or type 2 muscle fibers. Testosterone-induced muscle fiber hypertrophy is associated with dose-dependent increases in myonuclear number and satellite cell number.[140] Testosterone promotes the differentiation of mesenchymal multipotent muscle progenitor cells into the myogenic lineage and inhibits the differentiation of these precursor cells into the adipogenic lineage.[141,142] Testosterone promotes the formation of myosin heavy chain II positive myotubes and upregulates markers of myogenic differentiation, such as MyoD and myosin heavy chain.[142] Testosterone and DHT inhibit adipogenic differentiation and downregulate markers of adipogenic differentiation, such as peroxisome proliferator-activated receptor gamma (PPARγ) and C/EBP∝.[142,143]

Testosterone's effects on myogenic differentiation are mediated largely through its binding to the classical AR,[142] which induces a conformational change in the AR, promoting its association with its coactivator, β-catenin, causing the complex to translocate into the nucleus. The AR-β-catenin complex associates with T cell factor 4 (TCF-4) and activates a number of Wnt target genes,[143,144] including follistatin (Fig. 105.8). Follistatin cross-communicates the signal from the AR-β-catenin pathway to the transforming growth factor (TGF)-β signaling pathway, blocking signaling through the TGF-β/Smad 2/3.[144] Follistatin plays an essential role in mediating the effects of testosterone on myogenic differentiation.[143,144]

Testosterone regulates polyamine biosynthesis in multiple androgen-responsive tissues, including the muscle, prostate, and kidney.[145] Testosterone depletion downregulates ornithine decarboxylase 1 (ODC1) enzyme, the rate-limiting step in the polyamine biosynthesis, and testosterone supplementation of orchiectomized mice upregulates ODC1[145,146] in the prostate as well as the muscle. Polyamine availability in the prostate is constrained because polyamines are being continually acetylated and secreted by the prostate. The administration of ODC1 inhibitor, α-difluoromethylornithine, to mice is associated with reduced prostate mass. In orchidectomized mice, testosterone supplementation restores prostate as well as *levator ani* muscle mass but concurrent administration of ODC1 inhibitor blocks testosterone's restorative effects on prostate mass but not on the *levator ani* mass, indicating that ODC1 plays an important role in mediating the effects of testosterone on the prostate.[145] Therefore, combined administration of testosterone plus ODC1 inhibitor provides a novel approach for achieving selectivity of testosterone's anabolic effects on the muscle while sparing the prostate.[145]

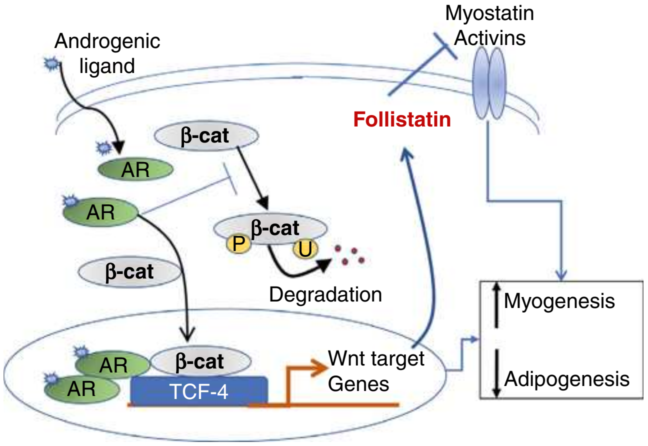


Figure 105.8 The mechanisms by which testosterone promotes myogenic differentiation of mesenchymal progenitor cells. *AR,* Androgen receptor; *β-cat,* β-catenin; *TCF,* T cell factor.

Mechanisms of Testosterone's Effects on Hematopoiesis.

Testosterone treatment is associated with a dose-related increase in hemoglobin and red cell number.[147] Testosterone acts on hematopoietic progenitor cells to increase the numbers of cells in stages that bear the transcriptional bias of multipotent common myeloid progenitors as well as early erythroid progenitors through an AR-mediated mechanism (Fig. 105.9).[148] Additionally, testosterone stimulates erythropoietin production and prevents its compensatory suppression even in the face of elevated red blood cell (RBC) counts, as indicated by higher levels of erythropoietin for any level of hemoglobin in testosterone-treated men compared to their pretreatment baseline.[149] Consistent with this effect on multipotent hematopoietic progenitors in the bone marrow, testosterone also increases the numbers of circulating neutrophils, monocytes, and platelets—not just RBCs—all of which are derived from myeloid series progenitors.[148] Testosterone stimulates iron-dependent erythropoiesis by suppressive hepcidin.[150] Testosterone suppresses hepcidin transcription in the liver by blocking BMP-SMAD signaling, increases iron recycling and export from the spleen by increased expression of ferroportin in the spleen, and increases serum iron and the incorporation of transferrin-bound ^{58}Fe into the RBCs of testosterone-treated mice compared to controls.[150] Additionally, testosterone stimulates erythropoietin production

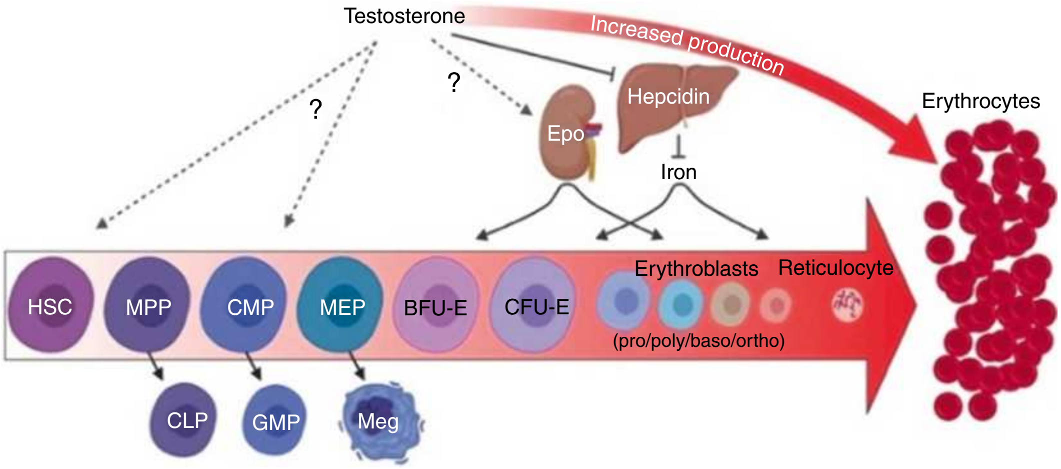
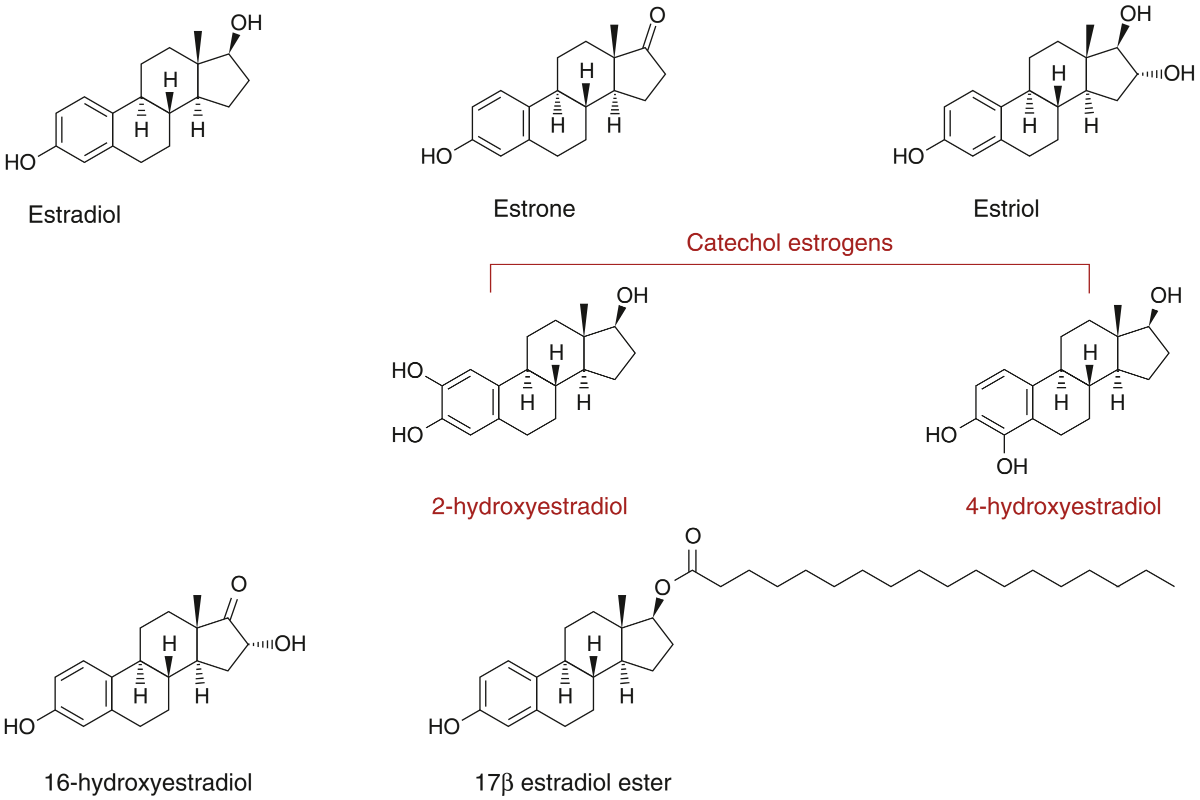


Figure 105.9 An integrated model of the mechanisms by which testosterone increases red cell counts. Testosterone increases the numbers of myeloid series progenitors from which red blood cells, neutrophils, monocytes, and platelets are derived. Testosterone increases iron availability by suppressing hepcidin in an iron-replete state. Iron is essential for mediating testosterone's effects on erythropoiesis. Testosterone also stimulates erythropoietin, which plays a permissive role in enabling the terminal differentiation of erythroid progenitors. *BFU-E,* erythroid burst forming units; *CFU-E,* erythroid colony forming units; *CLP,* common lymphoid progenitors; *CMP,* common myeloid progenitors; *Epo,* erythropoietin; *GMP,* granulocyte-macrophage progenitors; *HSC,* hematopoietic stem cells; *Meg,* megakaryocyte; *MEP,* megakaryocyte-erythrocyte progenitors; *MPP,* multipotential progenitors.



Estradiol

Estrone

Estriol

2-hydroxyestradiol

4-hydroxyestradiol

16-hydroxyestradiol

17β estradiol ester

Figure 105.10 Endogenously produced estrogenic compounds.

and prevents its compensatory suppression even in the face of elevated RBC counts, as indicated by higher levels of erythropoietin for any level of hemoglobin in testosterone-treated men compared to their pretreatment baseline, enabling the terminal differentiation of the increased numbers of erythroid progenitors.[149]

Nongenomic Effects of Testosterone. Some effects of testosterone are mediated through nongenomic mechanisms that are independent of the ligand-dependent transcriptional effects mediated by the classical AR. These nongenomic effects occur rapidly over short periods of time, do not require transcription or protein synthesis, and are not blocked by AR antagonists such as flutamide.[151] Among the most well-characterized nongenomic effects of testosterone is the relaxation of the vascular smooth muscle in the blood vessel walls that results in vasodilatation and increased blood flow.[151,152] The vasorelaxation induced by testosterone is associated with inactivation of the L-type voltage-operated Ca^{2+} channels and activation of K^+ channels especially the voltage-operated K^+ channels and the large conductance Ca+-activated K^+ channel (BK_{Ca}).[151,153] Testosterone also increases endothelium-derived nitric oxide by inducing endothelial nitric oxide synthase (eNOS) phosphorylation (Ser^{1177}) via PI3K, caveolin-1 and c-proto-oncogene tyrosine-protein kinase (Src)[154,155]; these actions are not associated with increased protein synthesis and are not blocked by classical AR antagonists. The nongenomic effects of androgens exhibit structural specificity; thus, 5β-dihydrotestosterone is a substantially more potent vasodilator than 5α-dihydrotestosterone,[151] but in contrast to 5α-dihydrotestosterone, the 5β isoform does not exert any appreciable AR-mediated androgenic effects. Androgens also exert similar nongenomic effects in inducing relaxation of smooth muscles in the bronchi, urinary bladder, gall bladder, and intestines.[151]

KEY POINTS

- Androgens exert their biologic effects through genomic and nongenomic mechanisms.
- In response to ligand binding, AR undergoes conformational changes, leading to dissociation of heat shock proteins, recruitment of coregulators and chaperone proteins, translocation to the nucleus, dimerization, chromatin remodeling, DNA binding, and regulation of transcriptional activity of AR-responsive genes.
- Ligand binding and coregulator protein-induced conformational dynamics enable the domains to exhibit multiple functional roles in a tissue-specific manner.

ESTROGEN PHYSIOLOGY

Endogenous Estrogens

Human ovaries and some peripheral tissues produce several estrogens—estradiol, estrone, estriol, and catechol estrogens (Fig. 105.10)—with varying estrogenic potency and cell type specificity; these estrogenic compounds are 18 carbon steroids that share common structural features, including an aromatic A ring, a phenolic hydroxyl group at C3, and a hydroxyl or a keto group at C17.

Estradiol is the principal estrogen secreted by the ovary in menstruating women although it is also produced by peripheral conversion from testosterone in the adipose tissue, breast, liver, muscle, bone, and in many brain regions.[156,157] Estradiol promotes the development and maintenance of female reproductive tissues such as the breast, vagina, and the uterus during puberty, adulthood, and pregnancy; in women of reproductive age, increasing estradiol concentrations during the menstrual cycle promote the maturation and release of the oocyte and the ripening of the endometrial lining to enable implantation of the fertilized egg. Additionally, estradiol regulates reproductive and nonreproductive behaviors; epiphyseal closure, osteoblastic bone formation as well as osteoclastic bone resorption; adipose tissue distribution and metabolism; and muscle, skin, and neuronal health.

Ovaries also secrete estrone, but a major fraction of circulating estrone is derived from aromatization of delta-4 androstenedione in the peripheral tissues.[156,157] In postmenopausal women, the circulating concentrations of estrone are substantially higher than those of estradiol.[158] Estrone has substantially lower potency than estradiol in activating signaling through ERα; however, it can be converted to estradiol by the catalytic action of 17β-hydroxysteroid dehydrogenase isoenzymes. *In vitro*, the various isoforms of 17β-hydroxysteroid dehydrogenase are typically bidirectional; however, *in vivo*, the reductases mainly use nicotinamide adenine dinucleotide phosphate (NADPH) to preferentially reduce estrone to estradiol. In humans, 17β-hydroxysteroid dehydrogenase type 1 is the principal isoenzyme involved in estradiol production from estrone in the ovaries as well as in peripheral tissues. Estrone is also the most abundant component of Premarin, a commercially available estrogen compound that contains 10 or more estrogenic compounds, of which estrone constitutes 50% to 60%.

Estriol is produced by peripheral metabolism of estradiol and estrone and represents the most abundant estrogen in the urine of menstruating as well as pregnant women. Estriol has very low activity on ERα, but it can be converted to estradiol through the action of 17β-hydroxysteroid dehydrogenase enzymes.

Catechol estrogens are estrogen metabolites that are synthesized in the brain, liver, and other tissues by the addition of a hydroxyl group at C2 or C4.[159,160] The cytochrome P450-dependent monooxygenases, encoded by CYP1A1 and CYP1A2 catalyze 2-hydroxylation of estrogens, and the protein product of CYP1B1 serve as an estrogen 4-hydroxylase.[159,160] Catechol estrogens typically have short half-lives *in vivo* and are postulated to serve as local signaling molecules. However, they can bind ER albeit with lower affinity than estradiol and exert both genomic and nongenomic effects in the target cells. Catechol estrogens are potent inhibitors of the inactivation of catecholamines by catecholamine methyltransferase (COMT). The 2-hydroxyl estradiol exerts antiangiogenic effects[159,161] and inhibits cell proliferation. In contrast, 4-hydroxy estradiol can be further metabolized to reactive catechol estrogen quinones that react with DNA to form mutagenic depurinating adducts that are potentially mutagenic and can promote the initiation of carcinogenesis.[161,162]

16α-Hydroxyestrone, a product of 16α-hydroxylation of estrone, promotes dysregulation of DNA synthesis and anchorage-independent growth in mouse mammary epithelial cells.[163] Circulating concentrations of 16α-hydroxyestrone have been associated with the risk of breast cancer induction.[164,165]

Regulation of Estradiol Production

The follicular cells in the preovulatory follicles produce estradiol prior to ovulation and the corpus luteum produces estradiol and progesterone after ovulation. With the ovulatory LH surge, granulosa cells undergo terminal differentiation into the large luteal cells of the corpus luteum that secrete progesterone. In the ovarian follicle, the production of 17β-estradiol by the granulosa cells is dependent upon LH-stimulated synthesis of androstenedione by the ovarian theca interna cells and the FSH-stimulated aromatization of C19 androgens by CYP19A1 in the granulosa cells.[166–168] The rate-limiting steps in estradiol production by the ovary are the cholesterol transport from the outer to the inner mitochondrial membrane by the StAR protein necessary for the side-chain cleavage of cholesterol to form pregnanolone in the theca-interstitial cells and the aromatization of C19 androgens in the granulosa cells.[166,167]

The LH binds to LH/CG receptor on the theca interna cells of the small follicles to stimulate the expression of steroidogenic enzymes—StAR protein, *CYP11A1* (cytochrome P450 side-chain cleavage enzyme), and *CYP17A1* (cytochrome P450 17α hydrolase)—that are essential for the production of androgens by theca-interstitial cells.[169–171] The stimulatory actions of LH on theca cell steroidogenesis are mediated largely through the activation of the cAMP/protein kinase A signaling pathway.[169–171] Protein kinase A activation also induces the phosphorylation of ERK1/2, inhibition of progesterone synthesis,[172] and stimulation of androgens synthesis by modulating the transcription of genes encoding steroidogenic enzymes. In granulosa cells, LH regulates the expression of multiple genes involved in progesterone synthesis, paracrine signaling, signal transduction, cytoskeletal regulation, and apoptosis.[173]

The rate-limiting step in the acute LH-induced upregulation of steroidogenesis in the theca interna cells is the transport of the cholesterol from the outer to the inner mitochondrial membrane.[174] In steroidogenic cells, the StAR protein interacts with a mitochondrial protein complex to facilitate cholesterol delivery to the inner mitochondrial membrane[174] where it is converted to pregnenolone by the action of cytochrome P450 cholesterol side-chain cleavage enzyme complex that includes cytochrome P450 cholesterol side-chain cleavage enzyme and cofactors, adrenodoxin and flavoprotein.[175] Severe disruption of steroidogenesis observed in preclinical models with genetic disruption of the StAR gene as well as in patients with congenital lipoid adrenal hyperplasia who have naturally occurring mutations of the StAR protein provides strong evidence for the critical role of StAR protein in steroidogenesis.[176] Pregnanolone is converted to androstenedione through the catalytic action of 17-hyroxylase/17,20-lyase enzyme, a product of the CYP17A1, and 3β-hydroxysteroid dehydrogenase enzyme, a product of the HSD3B2 gene.[177] Most of the androstenedione in humans is produced through the delta-5 pathway. Androstenedione, the most important product of the theca cells during the follicular phase of the menstrual cycle, is a weak androgen, but it serves a precursor for both estrone and testosterone in the ovaries and peripheral tissues, which are further converted to estradiol.

The aromatase enzyme encoded by the CYP19A1 gene catalyzes the conversion of theca-derived androstenedione to estradiol.[166,167,178] In granulosa cells from preovulatory follicles, FSH stimulates the synthesis of aromatase (CYP19A1) by inducing cAMP production and activating protein kinase A and certain MAP kinases.[178] The CYP19A1 gene contains a large regulatory region that contains many tissue-specific promoters that are regulated by a distinct set of regulatory sequences and transcription factors.[178,179] The expression of the human aromatase gene in various estrogen-producing tissues is regulated by the tissue-specific promoters upstream of alternative first exons, which encode the tissue-specific 5′-untranslated regions of the aromatase transcripts.[179] In the human ovary, aromatase transcripts contain 5′-untranslated sequences encoded by exon IIa, which lies 110 base pairs upstream of the translation start site.[179] Two orphan members of the Ftz-F1 subfamily of nuclear receptors—the liver receptor homolog-1 (LRH-1; NR5A2) and steroidogenic factor-1 (SF-1; NR5A1)—are expressed in the ovary and play an important role in regulating estradiol biosynthesis in the ovary.[180,181] Steroid biosynthesis in the ovary is enhanced by the transcription factors SF-1 and LRH-1 through the regulation of CYP19A1 and other steroidogenic enzymes StAR, CYP11A1, HSD3B2, and CYP17A1.[181–183] FSH also promotes the binding of SF-1 and LRH-1 to the promoter of the aromatase gene.[182,183]

In addition to the regulation of CYP19A1, FSH stimulates follicular growth and induces LH receptors in the granulosa cells of the preovulatory follicle. FSH receptor is expressed only in the granulosa cells of the ovary, in contrast to LHCG receptor, which is expressed in theca-interstitial cells and by granulosa cells of large preovulatory follicles.[184,185]

Estrogen Transport and Bioavailability

Most of the circulating estradiol is bound to SHBG and human serum albumin. Although sex steroid binding to human serum albumin has been generally described as a low-affinity binding, recent studies using 13C-oleic acid NMR have revealed the presence of multiple binding sites on human serum albumin that are allosterically coupled.[105] Recent studies of estradiol's binding to SHBG show that estradiol binding to SHBG is also a dynamic process in which each estradiol binding to each SHBG monomer influences residues in the ligand-binding pocket of both monomers and differentially alters the conformational and energy landscapes of both monomers[186] (Fig. 105.6). The SHBG monomers are not energetically or conformationally equivalent in an unliganded, singly bound, or fully bound state. Intermonomeric allostery could be nature's mechanism to extend the binding range of SHBG and regulate hormone bioavailability as estradiol concentrations vary widely during life.[186] Estradiol and testosterone are oriented differently in the ligand-binding pocket of SHBG. Naturally occurring SHBG variants differ in their relative affinities for estradiol and testosterone.[187] Ng et al.[188] reported a ligand-dependent interaction between SHBG and two members of the fibulin family of extracellular matrix–associated proteins, fibulin-1D and fibulin-2, in the uterus. The physiologic significance of these observations is unknown, but it is possible that SHBG-fibulin interaction could regulate the local estradiol concentration within the myometrium. SHBG concentrations are increased in pregnancy, hyperthyroidism, liver disease, old age, inflammatory conditions, and by some medications such as estrogens and some antiepileptics.

KEY POINTS

- The rate-limiting steps in estradiol production by the ovary are the cholesterol transport from the outer to the inner mitochondrial membrane by the StAR protein in the theca interna cells and the aromatization of C19 androgens in the granulosa cells.
- Steroid biosynthesis in the ovary is regulated by LH and FSH and enhanced by two orphan members of the Ftz-F1 subfamily of nuclear receptors, the transcription factors SF-1 and LRH-1 through the regulation of CYP19A1, and other steroidogenic enzymes StAR, CYP11A1, HSD3B2, and CYP17A1.
- Estrogen and testosterone can alter each other's bioavailability through dynamic repartitioning on the allosterically coupled binding sites on SHBG and albumin proteins.

Estrogen Metabolism

Estradiol is metabolized through a series of the NADPH-dependent hydroxylation at C2, C4, C15, or C16 or oxidation reactions to form inactive or less active polar (hydroxylated and keto) metabolites as well as to nonpolar metabolites largely in the liver.[189] These compounds are conjugated in the liver by uridine UDP-glucuronosyltransferase 1-3, the product of the UGT1A3 gene that converts estrogenic compounds into water-soluble glucuronides and by sulfotransferase enzyme into sulfates that are excreted in the urine.[189] Estriol glucuronide and sulfate are the major metabolic products excreted in the urine of menstruating and pregnant women. Some of the estrogen conjugates are excreted through the bile duct into the intestine, partially hydrolyzed and reabsorbed, thus constituting an enterohepatic recirculation loop that can extend the half-life of estrogenic products.[190] As discussed earlier, hydroxylation of estradiol and estrone at C2 and C4 can form catechol estrogens that subserve different functional roles.

Mechanisms of Estrogen Action

Estrogens exert their physiological function by genomic as well as nongenomic mechanisms. The genomic actions of estrogenic compounds are mediated by two ER subtypes, which are encoded from distinct genes on different chromosomes: ERα (NR3A1) and ERβ (NR3A2) that regulate the ligand-dependent transcription of specific target genes by binding to DNA regulatory sequences. ERα is expressed in the mammary gland, uterus, theca cells of the ovary, adipose tissue, bone, testes and epididymis, stromal cells of the prostate, vascular endothelium, and hypothalamus and other brain regions. ERβ is found mainly in the granulosa cells of the ovary, adipose tissue, cardiovascular system, immune system, several brain regions, epithelial cells of the prostate, bladder, and colon.

Similar to other members of the nuclear hormone receptors, ERα and ERβ are composed of three distinct functional domains spanning amino- to carboxyl-terminal sequences: the NTD, the DBD, and the LBD[190] (Fig. 105.11). Additionally, two AF domains, AF1 and AF2, located within the NTD and LBD, respectively, regulate the downstream transcriptional activity through ligand-specific conformational gating.[191,192] Liganded ER translocates to nucleus and associates with coactivators/corepressors. These multiprotein complexes associate with specific DNA sequences—estrogen response elements—to modulate transcriptional signaling.

ER Isoforms: Structural Features, Homology, and Functional Dynamics. ERα and ERβ share the same three functional domains (LBD, DBD, and NTD). While NTD and DBD exhibit low sequence homology of 15% and 55%, respectively.[193] DBD possesses ~95% homology between ERα and ERβ that facilitates the formation of heterodimers between the two ER isoforms on the estrogen response elements (EREs). The NTD of the nuclear hormone receptors has substantial conformational heterogeneity as a result of intrinsically disordered (ID) regions. While this dynamic structure allows the AF1 region to fine-tune the transactivation function of the ER, it has prevented the successful crystallization of most of the NTDs. Functionally, the presence of ID regions affords facile recognition surfaces capable of associating with target molecules in diverse cell types. Secondary structural analyses of the ERα and ERβ NTD using network protein sequence analysis show limited structural organization in helical or β-sheet states, while more than 67% of ERα and 80% of ERβ NTDs contains random coil conformation.[194] These large sections of ID regions in NTD may provide a common mechanism for unconstrained conformational sampling of coordinate/energy landscape to identify distinct binding partners in tissues. It has also been proposed that AF domain may populate the folded conformational state in the presence of coregulators such that relative concentrations and/or affinity of the coactivator/repressor could amplify or abrogate the transcriptional activity.

A highly conserved DBD, with 95% homology between the two ER isoforms, adjoins the N-terminal transactivation domain (A/B domain).[19] The minimal consensus ERE, shared by ERα and ERβ, is a 13 base-pair inverted repeat-5′GGTCAnnnTGACC3′. However, two flanking base pairs (resulting in a 15-bp palindrome: 5′AGGTCAnnnTGACCT3′) substantially influence the ER:ERE affinity.[195,196] The ERE sequences modulate the on/off rates for the ER:ERE complexes as well as the formation of higher-order complexes including recruitment of coactivators. The conformational states of ERα DBD:ERE structures have been extensively characterized by several biophysical spectroscopy methods. Solution structures have been solved using NMR techniques to complement the high-resolution,

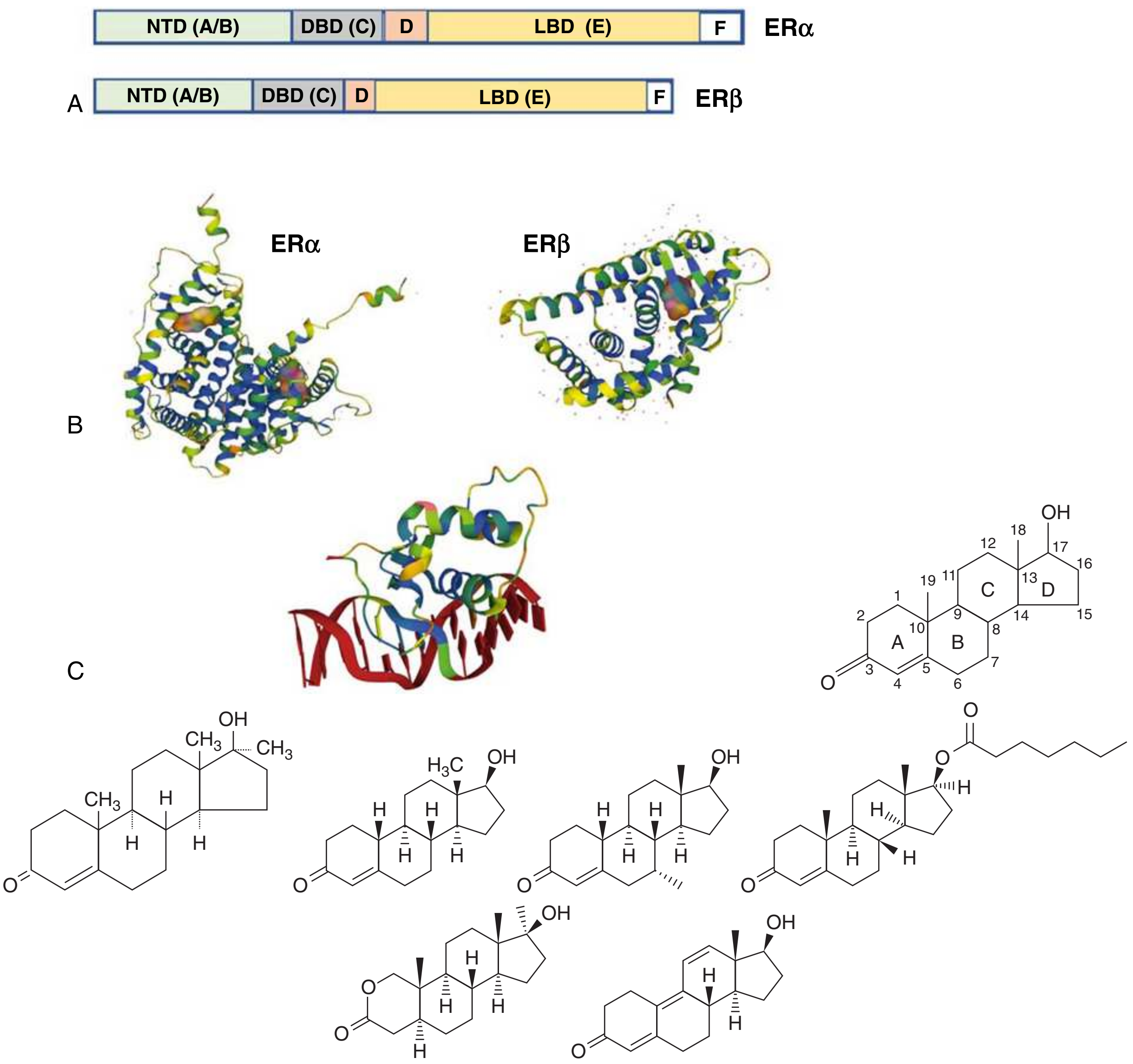


Figure 105.11 Schematic depiction of the various domains of the estrogen receptor (ER) subtypes, their crystal structure, and the three-dimensional structures of estradiol-bound ERα and ERβ. **A**, Schematic representation of the various domains of ERα and ERβ. **B**, Crystal structure of ER-DBD in complex with DNA-response element (obtained from 1LO1 submission in protein data bank). C, Three-dimensional structures of estradiol-bound ERα and ERβ, PDB structures 1A52 and 3OLS, respectively. *D*, hinge region; *DBD*, DNA-binding domain; *F*, C-terminal region; *LBD*, ligand-binding domain; *NTD*, N-terminal domain.

three-dimensional X-ray crystallographic studies.[197,198] These studies have shown a concerted role of P-box and D-box sequences in zinc finger domains. The cysteine-rich Zn finger subdomains coordinate with the two Zn^{2+} ions, P box associates with the ERE nucleotides, the D box has been shown to be present at the dimerization interface. The stoichiometry of ERE:ER complex formation has been actively debated in the literature with conflicting data suggesting that monomeric ER can associate with ERE half site or an ER monomer associating with ERE palindrome.[199]

The Ligand-Binding Domain. Organization of the LBD is a highly conserved structural feature among the members of the nuclear hormone receptor family. Similar to AR, the "E" domain of ER contains LBD spanning 11 α-helices (H1 to H12) with a hormone-binding pocket, a homo- or heterodimerization interface, and coregulator (activator or repressor) recruitment surface. The hormone-binding pocket in LBD mediates the ligand-induced conformational rearrangement and ensuing processes such as coregulator recruitment to AF2, dimerization, and translocation to the nucleus. ERα and ERβ retain overlapping three-dimensional arrangement even with a low overall sequence homology (55%).[193] Amino acid residues 223 to 343 and 404 to 457, which retain a significantly higher homology in ERα and ERβ appear to be buried on the interior, are protected from proteolytic degradation and participate in ligand binding. On the other hand, the sequences with lowest homology to ERβ are solvent accessible.[200] Both isoforms of ER-LBDs form dimers through the interactions spanning helices 10 and 11.

Ligand binding to ERα is accompanied by significant rearrangement in the orientation of helix 12. The hydrophobic ligand-binding pocket is formed by 22 residues. The crystal structure of 17β-estradiol-bound ERα LBD shows an important role of the two hydroxyl groups (3 and 17 positions) in the relative positioning of the ligand in the pocket. In ERα, the hydroxyl groups of the A and D rings form hydrogen bonds with Glu353, Arg394, and His 524 from helices H3, H5, and H11, respectively, and a water molecule.[201] However, there are notable changes in the two amino acid positions in binding pocket of ERα and ERβ: Leu384 and Met421 in ERα are replaced by Met336 and Ile373 in ERβ, respectively. While the changes do not alter the hydrophobicity in the pocket, the crystal structures show that the binding pocket of ERβ is slightly smaller than that of ERα.[202] Collectively, with the conserved spatial positioning of ligand-binding pocket residues and overall electrodensity map, the ERα and ERβ share agonist and antagonist ligands.

Coactivators. The pleiotropic effect of ER signaling on diverse tissues cannot be fully understood without accounting for the dynamic interaction of ER-associated coregulators. Over 200 coregulators have been identified that interact with ER to either enhance (coactivators) or inhibit (corepressors) transcription in a tissue-specific manner. Due to the heterogeneous expression of these coregulators in diverse tissues, their active role in mediating tissue-specific effects of ER is being increasingly recognized. While the coregulator proteins are thought to make higher-order complexes with unliganded or liganded ER states, many of them are epigenetic modifiers that exhibit enzymatic activities such as acetylation, methylation, and ubiquitination, thereby actively regulating transcriptional output. Coactivators belonging to members of SRC/p160 family, the histone acetyltransferase CREB-binding protein (CBP)/p300, ATP-dependent chromatin remodeling complexes like SWI/SNF, E3 ubiquitin-protein ligases, and steroid RNA activator (SRA) have been extensively characterized. Additionally, ERα and ERβ have been shown to differentially interact with the TATA box-binding protein (TBP), which has a central role in the basal transcription machinery. While TBP selectively binds to the ERα-NTD, it fails to associate with the ERβ-NTD and to potentiate ER-activated transcription. Most of the coactivators have been shown to share NR boxes or LxxLL motifs (x, any amino acid; L, Leucine), which stabilize their binding to ER LBD. Interestingly, the protein expression level of many of the coregulators of ER is under transcriptional control by estradiol, thereby providing a sensitive feedback mechanism for transcription regulation. Collectively, coregulators seem to play a critical role in mediating tissue-specific transcriptional outcomes by either physically stabilizing multiprotein complexes or altering post-transcriptional states of the in-tissue protein milieu. The net effect of these multilevel regulatory interactions results in a fine-tuned, ER-subtype–specific modulation of signaling pathways.

KEY POINTS

- Two estrogen receptor subtypes, ERα (NR3A1) and ERβ (NR3A2), modulate diverse physiological outcomes through either genomic (binding to specific DNA sequences to regulate ligand-dependent transcription of specific target genes) or direct nongenomic transcriptional events.
- Ligand-induced transcriptional control through interaction with coregulator families provides a facile mechanism for tissue-specific estrogen receptor-mediated effects.

REFERENCES

1. Plant TM, Barker-Gibb ML. Neurobiological mechanisms of puberty in higher primates. *Hum Reprod Update*. 2004;10(1):67–77.
2. Spratt DI, Finkelstein JS, Butler JP, Badger TM, Crowley Jr WF. Effects of increasing the frequency of low doses of gonadotropin-releasing hormone (GnRH) on gonadotropin secretion in GnRH-deficient men. *J Clin Endocrinol Metab*. 1987;64(6):1179–1186.
3. Kastin AJ, Schally AV, Gual C, et al. Stimulation of LH release in men and women by LH-releasing hormone purified from porcine hypothalami. *J Clin Endocrinol Metab*. 1969;29(8):1046–1050.
4. Stamou MI, Georgopoulos NA. Kallmann syndrome: phenotype and genotype of hypogonadotropic hypogonadism. *Metabolism*. 2018;86:124–134.
5. Salonia A, Rastrelli G, Hackett G, et al. Paediatric and adult-onset male hypogonadism. *Nat Rev Dis Primers*. 2019;5(1):38.
6. Barry J. Immunohistochemistry of luteinizing hormone-releasing hormone-producing neurons of the vertebrates. *Int Rev Cytol*. 1979;60:179–221.
7. Watanabe G, Terasawa E. In vivo release of luteinizing hormone releasing hormone increases with puberty in the female rhesus monkey. *Endocrinology*. 1989;125(1):92–99.
8. Belchetz PE, Plant TM, Nakai Y, Keogh EJ, Knobil E. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. *Science*. 1978;202(4368):631–633.
9. Maffucci JA, Gore AC. Chapter 2: hypothalamic neural systems controlling the female reproductive life cycle gonadotropin-releasing hormone, glutamate, and GABA. *Int Rev Cell Mol Biol*. 2009;274:69–127.
10. Kotani M, Detheux M, Vandenbogaerde A, et al. The metastasis suppressor gene KiSS-1 encodes kisspeptins, the natural ligands of the orphan G protein-coupled receptor GPR54. *J Biol Chem*. 2001;276(37):34631–34636.
11. Kauffman AS, Gottsch ML, Roa J, et al. Sexual differentiation of Kiss1 gene expression in the brain of the rat. *Endocrinology*. 2007;148(4):1774–1783.
12. Seminara SB, Messager S, Chatzidaki EE, et al. The GPR54 gene as a regulator of puberty. *N Engl J Med*. 2003;349(17):1614–1627.
13. Teles MG, Bianco SD, Brito VN, et al. A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*. 2008;358(7):709–715.
14. Shahab M, Mastronardi C, Seminara SB, et al. Increased hypothalamic GPR54 signaling: a potential mechanism for initiation of puberty in primates. *Proc Natl Acad Sci USA*. 2005;102(6):2129–2134.
15. Jayasena CN, Nijher GM, Chaudhri OB, et al. Subcutaneous injection of kisspeptin-54 acutely stimulates gonadotropin secretion in women with hypothalamic amenorrhea, but chronic administration causes tachyphylaxis. *J Clin Endocrinol Metab*. 2009;94(11):4315–4323.
16. Gutierrez-Pascual E, Martinez-Fuentes AJ, Pinilla L, et al. Direct pituitary effects of kisspeptin: activation of gonadotrophs and somatotrophs and stimulation of luteinising hormone and growth hormone secretion. *J Neuroendocrinol*. 2007;19(7):521–530.
17. Wuster M, Schulz R, Herz A. Opiate activity and receptor selectivity of dynorphin1-13 and related peptides. *Neurosci Lett*. 1980;20(1):79–83.
18. Navarro VM, Gottsch ML, Chavkin C, et al. Regulation of gonadotropin-releasing hormone secretion by kisspeptin/dynorphin/neurokinin B neurons in the arcuate nucleus of the mouse. *J Neurosci*. 2009;29(38):11859–11866.
19. Roa J, Vigo E, Castellano JM, et al. Opposite roles of estrogen receptor (ER)-alpha and ERbeta in the modulation of luteinizing hormone responses to kisspeptin in the female rat: implications for the generation of the preovulatory surge. *Endocrinology*. 2008;149(4):1627–1637.
20. Herbison AE, Theodosis DT. Immunocytochemical identification of oestrogen receptors in preoptic neurones containing calcitonin gene-related peptide in the male and female rat. *Neuroendocrinology*. 1992;56(5):761–764.
21. Smith JT, Popa SM, Clifton DK, Hoffman GE, Steiner RA. Kiss1 neurons in the forebrain as central processors for generating the preovulatory luteinizing hormone surge. *J Neurosci*. 2006;26(25):6687–6694.
22. Pineda R, Garcia-Galiano D, Roseweir A, et al. Critical roles of kisspeptins in female puberty and preovulatory gonadotropin surges as revealed by a novel antagonist. *Endocrinology*. 2010;151(2):722–730.
23. Tomikawa J, Uenoyama Y, Ozawa M, et al. Epigenetic regulation of Kiss1 gene expression mediating estrogen-positive feedback action in the mouse brain. *Proc Natl Acad Sci USA*. 2012;109(20):E1294–E1301.
24. Pitteloud N, Dwyer AA, DeCruz S, et al. Inhibition of luteinizing hormone secretion by testosterone in men requires aromatization for

its pituitary but not its hypothalamic effects: evidence from the tandem study of normal and gonadotropin-releasing hormone-deficient men. *J Clin Endocrinol Metab*. 2008;93(3):784–791.

25. Smith JT, Dungan HM, Stoll EA, et al. Differential regulation of KiSS-1 mRNA expression by sex steroids in the brain of the male mouse. *Endocrinology*. 2005;146(7):2976–2984.
26. Bhasin S, Fielder TJ, Swerdloff RS. Testosterone selectively increases serum follicle-stimulating hormonal (FSH) but not luteinizing hormone (LH) in gonadotropin-releasing hormone antagonist-treated male rats: evidence for differential regulation of LH and FSH secretion. *Biol Reprod*. 1987;37(1):55–59.
27. Wierman ME, Wang C. Androgen selectively stimulates follicle-stimulating hormone-beta mRNA levels after gonadotropin-releasing hormone antagonist administration. *Biol Reprod*. 1990;42(3):563–571.
28. Chehab FF, Lim ME, Lu R. Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. *Nat Genet*. 1996;12(3):318–320.
29. Smith JT, Acohido BV, Clifton DK, Steiner RA. KiSS-1 neurones are direct targets for leptin in the ob/ob mouse. *J Neuroendocrinol*. 2006;18(4):298–303.
30. Donato Jr J, Cravo RM, Frazao R, et al. Leptin's effect on puberty in mice is relayed by the ventral premammillary nucleus and does not require signaling in Kiss1 neurons. *J Clin Invest*. 2011;121(1):355–368.
31. Martin C, Navarro VM, Simavli S, et al. Leptin-responsive GABAergic neurons regulate fertility through pathways that result in reduced kisspeptinergic tone. *J Neurosci*. 2014;34(17):6047–6056.
32. Castellano JM, Navarro VM, Fernandez-Fernandez R, et al. Expression of hypothalamic KiSS-1 system and rescue of defective gonadotropic responses by kisspeptin in streptozotocin-induced diabetic male rats. *Diabetes*. 2006;55(9):2602–2610.
33. Bruning JC, Gautam D, Burks DJ, et al. Role of brain insulin receptor in control of body weight and reproduction. *Science*. 2000;289(5487):2122–2125.
34. Evans MC, Rizwan M, Mayer C, Boehm U, Anderson GM. Evidence that insulin signalling in gonadotrophin-releasing hormone and kisspeptin neurones does not play an essential role in metabolic regulation of fertility in mice. *J Neuroendocrinol*. 2014;26(7):468–479.
35. Backholer K, Smith JT, Rao A, et al. Kisspeptin cells in the Ewe brain respond to leptin and communicate with neuropeptide Y and proopiomelanocortin cells. *Endocrinology*. 2010;151(5):2233–2243.
36. Roa J, Garcia-Galiano D, Varela L, et al. The mammalian target of rapamycin as novel central regulator of puberty onset via modulation of hypothalamic Kiss1 system. *Endocrinology*. 2009;150(11):5016–5026.
37. Wahab F, Atika B, Huma T, Shahab M. Primate HPT axis response to the peripheral kisspeptin challenge under different time periods of food restriction in monkeys. *Horm Metab Res*. 2014;46(3):187–192.
38. Sanchez-Garrido MA, Ruiz-Pino F, Manfredi-Lozano M, et al. Obesity-induced hypogonadism in the male: premature reproductive neuroendocrine senescence and contribution of Kiss1-mediated mechanisms. *Endocrinology*. 2014;155(3):1067–1079.
39. Kinsey-Jones JS, Li XF, Knox AM, et al. Down-regulation of hypothalamic kisspeptin and its receptor, Kiss1r, mRNA expression is associated with stress-induced suppression of luteinising hormone secretion in the female rat. *J Neuroendocrinol*. 2009;21(1):20–29.
40. Comninos AN, Anastasovska J, Sahuri-Arisoylu M, et al. Kisspeptin signaling in the amygdala modulates reproductive hormone secretion. *Brain Struct Funct*. 2016;221(4):2035–2047.
41. Iwasa T, Matsuzaki T, Tungalagsuvd A, et al. Hypothalamic Kiss1 and RFRP gene expressions are changed by a high dose of lipopolysaccharide in female rats. *Horm Behav*. 2014;66(2):309–316.
42. Simonneaux V, Ansel L, Revel FG, et al. Kisspeptin and the seasonal control of reproduction in hamsters. *Peptides*. 2009;30(1):146–153.
43. Clarke IJ, Smith JT, Caraty A, Goodman RL, Lehman MN. Kisspeptin and seasonality in sheep. *Peptides*. 2009;30(1):154–163.
44. Caraty A, Smith JT, Lomet D, et al. Kisspeptin synchronizes preovulatory surges in cyclical ewes and causes ovulation in seasonally acyclic ewes. *Endocrinology*. 2007;148(11):5258–5267.
45. Bartzen-Sprauer J, Klosen P, Ciofi P, Mikkelsen JD, Simonneaux V. Photoperiodic co-regulation of kisseptin, neurokinin B and dynorphin in the hypothalamus of a seasonal rodent. *J Neuroendocrinol*. 2014;26(8):510–520.
46. Piekarski DJ, Jarjisian SG, Perez L, et al. Effects of pinealectomy and short day lengths on reproduction and neuronal RFRP-3, kisspeptin, and GnRH in female Turkish hamsters. *J Biol Rhythms*. 2014;29(3):181–191.
47. Santen RJ, Simpson E. History of estrogen: its purification, structure, synthesis, biologic actions, and clinical implications. *Endocrinology*. 2019;160(3):605–625.
48. Bhasin S, Travison TG, Storer TW, et al. Effect of testosterone supplementation with and without a dual 5alpha-reductase inhibitor on fat-free mass in men with suppressed testosterone production: a randomized controlled trial. *JAMA*. 2012;307(9):931–939.
49. Finkelstein JS, Lee H, Burnett-Bowie SA, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*. 2013;369(11):1011–1022.
50. Melcangi RC, Celotti F, Castano P, Martini L. Differential localization of the 5 alpha-reductase and the 3 alpha-hydroxysteroid dehydrogenase in neuronal and glial cultures. *Endocrinology*. 1993;132(3):1252–1259.
51. Reddy DS. Testosterone modulation of seizure susceptibility is mediated by neurosteroids 3alpha-androstanediol and 17beta-estradiol. *Neuroscience*. 2004;129(1):195–207.
52. Handa RJ, Sharma D, Uht R. A role for the androgen metabolite, 5alpha androstane 3beta, 17beta diol (3beta-diol) in the regulation of the hypothalamo-pituitary-adrenal axis. *Front Endocrinol*. 2011;2:65.
53. Fisher CR, Graves KH, Parlow AF, Simpson ER. Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the cyp19 gene. *Proc Natl Acad Sci USA*. 1998;95(12):6965–6970.
54. Falahati-Nini A, Riggs BL, Atkinson EJ, et al. Relative contributions of testosterone and estrogen in regulating bone resorption and formation in normal elderly men. *J Clin Invest*. 2000;106(12):1553–1560.
55. Krege JH, Hodgin JB, Couse JF, et al. Generation and reproductive phenotypes of mice lacking estrogen receptor beta. *Proc Natl Acad Sci USA*. 1998;95(26):15677–15682.
56. Walter P, Green S, Greene G, et al. Cloning of the human estrogen receptor cDNA. *Proc Natl Acad Sci USA*. 1985;82(23):7889–7893.
57. Lubahn DB, Moyer JS, Golding TS, et al. Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. *Proc Natl Acad Sci USA*. 1993;90(23):11162–11166.
58. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med*. 1994;331(16):1056–1061.
59. Lang-Muritano M, Sproll P, Wyss S, et al. Early-onset complete ovarian failure and lack of puberty in a woman with mutated estrogen receptor beta (ESR2). *J Clin Endocrinol Metab*. 2018;103(10):3748–3756.
60. Lin L, Ercan O, Raza J, et al. Variable phenotypes associated with aromatase (CYP19) insufficiency in humans. *J Clin Endocrinol Metab*. 2007;92(3):982–990.
61. Murata Y, Robertson KM, Jones ME, Simpson ER. Effect of estrogen deficiency in the male: the ArKO mouse model. *Mol Cell Endocrinol*. 2002;193(1–2):7–12.
62. Eriksson AL, Perry JRB, Coviello AD, et al. Genetic determinants of circulating estrogen levels and evidence of a causal effect of estradiol on bone density in men. *J Clin Endocrinol Metab*. 2018;103(3):991–1004.
63. Nethander M, Vandenput L, Eriksson AL, et al. Evidence of a causal effect of estradiol on fracture risk in men. *J Clin Endocrinol Metab*. 2019;104(2):433–442.
64. Mendonca BB, Batista RL, Domenice S, et al. Steroid 5alpha-reductase 2 deficiency. *J Steroid Biochem Mol Biol*. 2016;163:206–211.
65. Andriole GL, Kirby R. Safety and tolerability of the dual 5alpha-reductase inhibitor dutasteride in the treatment of benign prostatic hyperplasia. *Eur Urol*. 2003;44(1):82–88.
66. Wilson JD. The role of 5alpha-reduction in steroid hormone physiology. *Reprod Fertil Dev*. 2001;13(7–8):673–678.
67. Flores-Ramos M, Alcauter S, Lopez-Titla M, et al. Testosterone is related to GABA+ levels in the posterior-cingulate in unmedicated depressed women during reproductive life. *J Affect Disord*. 2019;242:143–149.

68. Zuloaga KL, O'Connor DT, Handa RJ, Gonzales RJ. Estrogen receptor beta dependent attenuation of cytokine-induced cyclooxygenase-2 by androgens in human brain vascular smooth muscle cells and rat mesenteric arteries. *Steroids*. 2012;77(8–9):835–844.
69. Chen J, Wang WQ, Lin SX. Interaction of androst-5-ene-3beta,17beta-diol and 5alpha-androstane-3beta,17beta-diol with estrogen and androgen receptors: a combined binding and cell study. *J Steroid Biochem Mol Biol*. 2013;137:316–321.
70. Midzak A, Zirkin B, Papadopoulos V. Translocator protein: pharmacology and steroidogenesis. *Biochem Soc Trans*. 2015;43(4):572–578.
71. Rone MB, Fan J, Papadopoulos V. Cholesterol transport in steroid biosynthesis: role of protein-protein interactions and implications in disease states. *Biochim Biophys Acta*. 2009;1791(7):646–658.
72. Chung JY, Chen H, Papadopoulos V, Zirkin B. Cholesterol accumulation, lipid droplet formation, and steroid production in Leydig cells: role of translocator protein (18-kDa). *Andrology*. 2020;8(3):719–730.
73. Bose HS, Sato S, Aisenberg J, et al. Mutations in the steroidogenic acute regulatory protein (StAR) in six patients with congenital lipoid adrenal hyperplasia. *J Clin Endocrinol Metab*. 2000;85(10):3636–3639.
74. Teerds KJ, Huhtaniemi IT. Morphological and functional maturation of Leydig cells: from rodent models to primates. *Hum Reprod Update*. 2015;21(3):310–328.
75. Fluck CE, Meyer-Boni M, Pandey AV, et al. Why boys will be boys: two pathways of fetal testicular androgen biosynthesis are needed for male sexual differentiation. *Am J Hum Genet*. 2011;89(2):201–218.
76. O'Shaughnessy PJ, Antignac JP, Le Bizec B, et al. Alternative (backdoor) androgen production and masculinization in the human fetus. *PLoS Biol*. 2019;17(2):e3000002.
77. Schiffer L, Barnard L, Baranowski ES, et al. Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: a comprehensive review. *J Steroid Biochem Mol Biol*. 2019;194:105439.
78. Imamichi Y, Yuhki KI, Orisaka M, et al. 11-Ketotestosterone is a major androgen produced in human gonads. *J Clin Endocrinol Metab*. 2016;101(10):3582–3591.
79. Turcu AF, Rege J, Auchus RJ, Rainey WE. 11-Oxygenated androgens in health and disease. *Nat Rev Endocrinol*. 2020;16(5):284–296.
80. Kicman AT. Pharmacology of anabolic steroids. *Br J Pharmacol*. 2008;154(3):502–521.
81. Bhasin S, Calof OM, Storer TW, et al. Drug insight: testosterone and selective androgen receptor modulators as anabolic therapies for chronic illness and aging. *Nat Clin Pract Endocrinol Metab*. 2006;2(3):146–159.
82. Fragkaki AG, Angelis YS, Koupparis M, et al. Structural characteristics of anabolic androgenic steroids contributing to binding to the androgen receptor and to their anabolic and androgenic activities. Applied modifications in the steroidal structure. *Steroids*. 2009;74(2):172–197.
83. Sokol RZ, Palacios A, Campfield LA, Saul C, Swerdloff RS. Comparison of the kinetics of injectable testosterone in eugonadal and hypogonadal men. *Fertil Steril*. 1982;37(3):425–430.
84. Weems PW, Lehman MN, Coolen LM, Goodman RL. The roles of neurokinins and endogenous opioid peptides in control of pulsatile LH secretion. *Vitam Horm*. 2018;107:89–135.
85. Ji TH, Ryu KS, Gilchrist R, Ji I. Interaction, signal generation, signal divergence, and signal transduction of LH/CG and the receptor. *Recent Prog Horm Res*. 1997;52:431–453; discussion 454.
86. Casarini L, Simoni M. Recent advances in understanding gonadotropin signaling. *Fac Rev*. 2021;10:41.
87. Herrlich A, Kuhn B, Grosse R, et al. Involvement of Gs and Gi proteins in dual coupling of the luteinizing hormone receptor to adenylyl cyclase and phospholipase C. *J Biol Chem*. 1996;271(28):16764–16772.
88. Huhtaniemi I. The Parkes lecture. Mutations of gonadotrophin and gonadotrophin receptor genes: what do they teach us about reproductive physiology? *J Reprod Fertil*. 2000;119(2):173–186.
89. Goldman AL, Bhasin S, Wu FCW, et al. A Reappraisal of testosterone's binding in circulation: physiological and clinical implications. *Endocr Rev*. 2017;38(4):302–324.
90. Nakhla AM, Hryb DJ, Rosner W, et al. Human sex hormone-binding globulin gene expression- multiple promoters and complex alternative splicing. *BMC Mol Biol*. 2009;10:37.
91. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. *J Clin Endocrinol Metab*. 1981;53(1):58–68.
92. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*. 1999;84(10):3666–3672.
93. Sodergard R, Backstrom T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. *J Steroid Biochem*. 1982;16(6):801–810.
94. Mazer NA. A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. *Steroids*. 2009;74(6):512–519.
95. Zakharov MN, Bhasin S, Travison TG, et al. A multi-step, dynamic allosteric model of testosterone's binding to sex hormone binding globulin. *Mol Cell Endocrinol*. 2015;399:190–200.
96. Ding VD, Moller DE, Feeney WP, et al. Sex hormone-binding globulin mediates prostate androgen receptor action via a novel signaling pathway. *Endocrinology*. 1998;139(1):213–218.
97. Plymate SR, Loop SM, Hoop RC, et al. Effects of sex hormone binding globulin (SHBG) on human prostatic carcinoma. *J Steroid Biochem Mol Biol*. 1991;40(4–6):833–839.
98. Hammes A, Andreassen TK, Spoelgen R, et al. Role of endocytosis in cellular uptake of sex steroids. *Cell*. 2005;122(5):751–762.
99. Bhasin S, Jasjua GK, Pencina M, et al. Sex hormone-binding globulin, but not testosterone, is associated prospectively and independently with incident metabolic syndrome in men: the Framingham heart study. *Diabetes Care*. 2011;34(11):2464–2470.
100. Lakshman KM, Bhasin S, Araujo AB. Sex hormone-binding globulin as an independent predictor of incident type 2 diabetes mellitus in men. *J Gerontol A Biol Sci Med Sci*. 2010;65(5):503–509.
101. Ding EL, Song Y, Manson JE, et al. Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N Engl J Med*. 2009;361(12):1152–1163.
102. Matsushita S, Isima Y, Chuang VT, et al. Functional analysis of recombinant human serum albumin domains for pharmaceutical applications. *Pharm Res (N Y)*. 2004;21(10):1924–1932.
103. Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 A resolution. *Protein Eng*. 1999;12(6):439–446.
104. Nisula BC, Dunn JF. Measurement of the testosterone binding parameters for both testosterone-estradiol binding globulin and albumin in individual serum samples. *Steroids*. 1979;34(7):771–791.
105. Jayaraj A, Schwanz HA, Spencer DJ, et al. Allosterically coupled multisite binding of testosterone to human serum albumin. *Endocrinology*. 2021;162(2).
106. Mendel CM. The free hormone hypothesis: a physiologically based mathematical model. *Endocr Rev*. 1989;10(3):232–274.
107. Bhasin S, Brito JP, Cunningham GR, et al. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical Practice Guideline. *J Clin Endocrinol Metab*. 2018;103(5):1715–1744.
108. Laurent MR, Hammond GL, Blokland M, et al. Sex hormone-binding globulin regulation of androgen bioactivity in vivo: validation of the free hormone hypothesis. *Sci Rep*. 2016;6:35539.
109. Coviello AD, Haring R, Wellons M, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet*. 2012;8(7):e1002805.
110. Ohlsson C, Wallaschofski H, Lunetta KL, et al. Genetic determinants of serum testosterone concentrations in men. *PLoS Genet*. 2011;7(10):e1002313.
111. Vos MJ, Mijnhout GS, Rondeel JM, Baron W, Groeneveld PH. Sex hormone binding globulin deficiency due to a homozygous missense mutation. *J Clin Endocrinol Metab*. 2014;99(9):E1798–E1802.

112. Hogeveen KN, Cousin P, Pugeat M, et al. Human sex hormone-binding globulin variants associated with hyperandrogenism and ovarian dysfunction. *J Clin Invest*. 2002;109(7):973–981.
113. Antonio L, Wu FC, O'Neill TW, et al. Low free testosterone is associated with hypogonadal signs and symptoms in men with normal total testosterone. *J Clin Endocrinol Metab*. 2016;101(7):2647–2657.
114. Anawalt BD, Hotaling JM, Walsh TJ, Matsumoto AM. Performance of total testosterone measurement to predict free testosterone for the biochemical evaluation of male hypogonadism. *J Urol*. 2012;187(4):1369–1373.
115. Schanzer W. Metabolism of anabolic androgenic steroids. *Clin Chem*. 1996;42(7):1001–1020.
116. Hoffman JR, Kraemer WJ, Bhasin S, et al. Position stand on androgen and human growth hormone use. *J Strength Cond Res*. 2009;23(suppl 5):S1–S59.
117. Tan JA, Joseph DR, Quarmby VE, et al. The rat androgen receptor: primary structure, autoregulation of its messenger ribonucleic acid, and immunocytochemical localization of the receptor protein. *Mol Endocrinol*. 1988;2(12):1276–1285.
118. Jenster G, van der Korput HA, van Vroonhoven C, et al. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Mol Endocrinol*. 1991;5(10):1396–1404.
119. Matsumoto T, Sakari M, Okada M, et al. The androgen receptor in health and disease. *Annu Rev Physiol*. 2013;75:201–224.
120. Pihlajamaa P, Sahu B, Janne OA. Determinants of receptor- and tissue-specific actions in androgen signaling. *Endocr Rev*. 2015;36(4):357–384.
121. Trapman J, Ris-Stalpers C, van der Korput JA, et al. The androgen receptor: functional structure and expression in transplanted human prostate tumors and prostate tumor cell lines. *J Steroid Biochem Mol Biol*. 1990;37(6):837–842.
122. Roy AK, Tyagi RK, Song CS, et al. Androgen receptor: structural domains and functional dynamics after ligand-receptor interaction. *Ann N Y Acad Sci*. 2001;949:44–57.
123. Heinlein CA, Chang C. Androgen receptor (AR) coregulators: an overview. *Endocr Rev*. 2002;23(2):175–200.
124. Edwards DP. The role of coactivators and corepressors in the biology and mechanism of action of steroid hormone receptors. *J Mammary Gland Biol Neoplasia*. 2000;5(3):307–324.
125. Heemers HV, Tindall DJ. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev*. 2007;28(7):778–808.
126. Jenster G, van der Korput HA, Trapman J, Brinkmann AO. Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J Biol Chem*. 1995;270(13):7341–7346.
127. Tyagi RK, Lavrovsky Y, Ahn SC, et al. Dynamics of intracellular movement and nucleocytoplasmic recycling of the ligand-activated androgen receptor in living cells. *Mol Endocrinol*. 2000;14(8):1162–1174.
128. Kemppainen JA, Lane MV, Sar M, Wilson EM. Androgen receptor phosphorylation, turnover, nuclear transport, and transcriptional activation. Specificity for steroids and antihormones. *J Biol Chem*. 1992;267(2):968–974.
129. Sack JS, Kish KF, Wang C, et al. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci USA*. 2001;98(9):4904–4909.
130. Pereira de Jesus-Tran K, Cote PL, Cantin L, et al. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Sci*. 2006;15(5):987–999.
131. Bohl CE, Wu Z, Miller DD, Bell CE, Dalton JT. Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *J Biol Chem*. 2007;282(18):13648–13655.
132. Schoenmakers E, Verrijdt G, Peeters B, et al. Differences in DNA binding characteristics of the androgen and glucocorticoid receptors can determine hormone-specific responses. *J Biol Chem*. 2000;275(16):12290–12297.
133. Cato AC, Henderson D, Ponta H. The hormone response element of the mouse mammary tumour virus DNA mediates the progestin and androgen induction of transcription in the proviral long terminal repeat region. *EMBO J*. 1987;6(2):363–368.
134. Palazzolo I, Gliozzi A, Rusmini P, et al. The role of the polyglutamine tract in androgen receptor. *J Steroid Biochem Mol Biol*. 2008;108(3–5):245–253.
135. Breza M, Koutsis G. Kennedy's disease (spinal and bulbar muscular atrophy): a clinically oriented review of a rare disease. *J Neurol*. 2019;266(3):565–573.
136. Bhasin S, Woodhouse L, Casaburi R, et al. Testosterone dose-response relationships in healthy young men. *Am J Physiol Endocrinol Metab*. 2001;281(6):E1172–E1181.
137. Bhasin S, Woodhouse L, Casaburi R, et al. Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. *J Clin Endocrinol Metab*. 2005;90(2):678–688.
138. Sinha-Hikim I, Artaza J, Woodhouse L, et al. Testosterone-induced increase in muscle size in healthy young men is associated with muscle fiber hypertrophy. *Am J Physiol Endocrinol Metab*. 2002;283(1):E154–E164.
139. Sinha-Hikim I, Roth SM, Lee MI, Bhasin S. Testosterone-induced muscle hypertrophy is associated with an increase in satellite cell number in healthy, young men. *Am J Physiol Endocrinol Metab*. 2003;285(1):E197–E205.
140. Sinha-Hikim I, Cornford M, Gaytan H, Lee ML, Bhasin S. Effects of testosterone supplementation on skeletal muscle fiber hypertrophy and satellite cells in community-dwelling older men. *J Clin Endocrinol Metab*. 2006;91(8):3024–3033.
141. Bhasin S, Taylor WE, Singh R, et al. The mechanisms of androgen effects on body composition: mesenchymal pluripotent cell as the target of androgen action. *J Gerontol A Biol Sci Med Sci*. 2003;58(12):M1103–M1110.
142. Singh R, Artaza JN, Taylor WE, Gonzalez-Cadavid NF, Bhasin S. Androgens stimulate myogenic differentiation and inhibit adipogenesis in C3H 10T1/2 pluripotent cells through an androgen receptor-mediated pathway. *Endocrinology*. 2003;144(11):5081–5088.
143. Singh R, Artaza JN, Taylor WE, et al. Testosterone inhibits adipogenic differentiation in 3T3-L1 cells: nuclear translocation of androgen receptor complex with beta-catenin and T-cell factor 4 may bypass canonical Wnt signaling to down-regulate adipogenic transcription factors. *Endocrinology*. 2006;147(1):141–154.
144. Braga M, Bhasin S, Jasuja R, Pervin S, Singh R. Testosterone inhibits transforming growth factor-beta signaling during myogenic differentiation and proliferation of mouse satellite cells: potential role of follistatin in mediating testosterone action. *Mol Cell Endocrinol*. 2012;350(1):39–52.
145. Jasuja R, Costello JC, Singh R, et al. Combined administration of testosterone plus an ornithine decarboxylase inhibitor as a selective prostate-sparing anabolic therapy. *Aging Cell*. 2014;13(2):303–310.
146. Piik K, Rajamaki P, Guha SK, Janne J. Regulation of L-ornithine decarboxylase and S-adenosyl-L-methionine decarboxylase in rat ventral prostate and seminal vesicle. *Biochem J*. 1977;168(3):379–385.
147. Coviello AD, Kaplan B, Lakshman KM, et al. Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab*. 2008;93(3):914–919.
148. Gagliano-Juca T, Pencina KM, Guo W, et al. Differential effects of testosterone on circulating neutrophils, monocytes, and platelets in men: findings from two trials. *Andrology*. 2020;8(5):1324–1331.
149. Bachman E, Travison TG, Basaria S, et al. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*. 2014;69(6):725–735.
150. Guo W, Bachman E, Li M, et al. Testosterone administration inhibits hepcidin transcription and is associated with increased iron incorporation into red blood cells. *Aging Cell*. 2013;12(2):280–291.
151. Perusquia M, Stallone JN. Do androgens play a beneficial role in the regulation of vascular tone? Nongenomic vascular effects of testosterone metabolites. *Am J Physiol Heart Circ Physiol*. 2010;298(5):H1301–H1307.
152. Lorigo M, Mariana M, Lemos MC, Cairrao E. Vascular mechanisms of testosterone: the non-genomic point of view. *J Steroid Biochem Mol Biol*. 2020;196:105496.

153. Ding AQ, Stallone JN. Testosterone-induced relaxation of rat aorta is androgen structure specific and involves K+ channel activation. *J Appl Physiol (1985)*. 2001;91(6):2742–2750.
154. Yu J, Akishita M, Eto M, et al. Androgen receptor-dependent activation of endothelial nitric oxide synthase in vascular endothelial cells: role of phosphatidylinositol 3-kinase/akt pathway. *Endocrinology*. 2010;151(4):1822–1828.
155. Bulldan A, Dietze R, Shihan M, Scheiner-Bobis G. Non-classical testosterone signaling mediated through ZIP9 stimulates claudin expression and tight junction formation in Sertoli cells. *Cell Signal*. 2016;28(8):1075–1085.
156. Barlow JJ, Emerson Jr K, Saxena BN. Estradiol production after ovariectomy for carcinoma of the breast. *N Engl J Med*. 1969;280(12):633–637.
157. England PC, Skinner LG, Cottrell KM, Sellwood RA. Serum oestradiol-17 beta in normal women. *Br J Cancer*. 1974;29(6):462–469.
158. Chakravarti S, Collins WP, Forecast JD, et al. Hormonal profiles after the menopause. *Br Med J*. 1976;2(6039):784–787.
159. Doherty JA, Weiss NS, Freeman RJ, et al. Genetic factors in catechol estrogen metabolism in relation to the risk of endometrial cancer. *Cancer Epidemiol Biomarkers Prev*. 2005;14(2):357–366.
160. Schendzielorz N, Rysa A, Reenila I, Raasmaja A, Mannisto PT. Complex estrogenic regulation of catechol-O-methyltransferase (COMT) in rats. *J Physiol Pharmacol*. 2011;62(4):483–490.
161. Tripathi K, Mani C, Somasagara RR, et al. Detection and evaluation of estrogen DNA-adducts and their carcinogenic effects in cultured human cells using biotinylated estradiol. *Mol Carcinog*. 2017;56(3):1010–1020.
162. Zhang Y, Gaikwad NW, Olson K, et al. Cytochrome P450 isoforms catalyze formation of catechol estrogen quinones that react with DNA. *Metabolism*. 2007;56(7):887–894.
163. Telang NT, Suto A, Wong GY, Osborne MP, Bradlow HL. Induction by estrogen metabolite 16 alpha-hydroxyestrone of genotoxic damage and aberrant proliferation in mouse mammary epithelial cells. *J Natl Cancer Inst*. 1992;84(8):634–638.
164. Bradlow HL, Hershcopf RE, Fishman JF. Oestradiol 16 alpha-hydroxylase: a risk marker for breast cancer. *Cancer Surv*. 1986;5(3):573–583.
165. Osborne MP, Bradlow HL, Wong GY, Telang NT. Upregulation of estradiol C16 alpha-hydroxylation in human breast tissue: a potential biomarker of breast cancer risk. *J Natl Cancer Inst*. 1993;85(23):1917–1920.
166. Barbieri RL. The endocrinology of the menstrual cycle. *Methods Mol Biol*. 2014;1154:145–169.
167. Bulun SE. Aromatase and estrogen receptor alpha deficiency. *Fertil Steril*. 2014;101(2):323–329.
168. Zachow RJ, Magoffin DA. Granulosa cell modulation of luteinizing hormone-dependent androgen production by ovarian theca-interstitial cells: a temporal switch from suppression to augmentation stimulated by follicle-stimulating hormone in vitro. *Biol Reprod*. 1995;53(4):758–765.
169. Magoffin DA. Evidence that luteinizing hormone-stimulated differentiation of purified ovarian thecal-interstitial cells is mediated by both type I and type II adenosine 3',5'-monophosphate-dependent protein kinases. *Endocrinology*. 1989;125(3):1464–1473.
170. Sasson R, Rimon E, Dantes A, et al. Gonadotrophin-induced gene regulation in human granulosa cells obtained from IVF patients. Modulation of steroidogenic genes, cytoskeletal genes and genes coding for apoptotic signalling and protein kinases. *Mol Hum Reprod*. 2004;10(5):299–311.
171. Ivell R, Tillmann G, Wang H, et al. Acute regulation of the bovine gene for the steroidogenic acute regulatory protein in ovarian theca and adrenocortical cells. *J Mol Endocrinol*. 2000;24(1):109–118.
172. Seger R, Hanoch T, Rosenberg R, et al. The ERK signaling cascade inhibits gonadotropin-stimulated steroidogenesis. *J Biol Chem*. 2017;292(21):8847.
173. Sekar N, Garmey JC, Veldhuis JD. Mechanisms underlying the steroidogenic synergy of insulin and luteinizing hormone in porcine granulosa cells: joint amplification of pivotal sterol-regulatory genes encoding the low-density lipoprotein (LDL) receptor, steroidogenic acute regulatory (stAR) protein and cytochrome P450 side-chain cleavage (P450scc) enzyme. *Mol Cell Endocrinol*. 2000;159(1–2):25–35.
174. Stocco DM, Clark BJ. Regulation of the acute production of steroids in steroidogenic cells. *Endocr Rev*. 1996;17(3):221–244.
175. Miller WL, Strauss 3rd JF. Molecular pathology and mechanism of action of the steroidogenic acute regulatory protein, StAR. *J Steroid Biochem Mol Biol*. 1999;69(1–6):131–141.
176. Miller WL. Congenital lipoid adrenal hyperplasia: the human gene knockout for the steroidogenic acute regulatory protein. *J Mol Endocrinol*. 1997;19(3):227–240.
177. Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. *Cancer Lett*. 2015;356(2 Pt A):231–243.
178. Stocco C. Aromatase expression in the ovary: hormonal and molecular regulation. *Steroids*. 2008;73(5):473–487.
179. Kamat A, Hinshelwood MM, Murry BA, Mendelson CR. Mechanisms in tissue-specific regulation of estrogen biosynthesis in humans. *Trends Endocrinol Metab*. 2002;13(3):122–128.
180. Fayard E, Auwerx J, Schoonjans K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. *Trends Cell Biol*. 2004;14(5):250–260.
181. Hinshelwood MM, Repa JJ, Shelton JM, et al. Expression of LRH-1 and SF-1 in the mouse ovary: localization in different cell types correlates with differing function. *Mol Cell Endocrinol*. 2003;207(1–2):39–45.
182. Leers-Sucheta S, Morohashi K, Mason JI, Melner MH. Synergistic activation of the human type II 3beta-hydroxysteroid dehydrogenase/delta5-delta4 isomerase promoter by the transcription factor steroidogenic factor-1/adrenal 4-binding protein and phorbol ester. *J Biol Chem*. 1997;272(12):7960–7967.
183. Hanley NA, Ikeda Y, Luo X, Parker KL. Steroidogenic factor 1 (SF-1) is essential for ovarian development and function. *Mol Cell Endocrinol*. 2000;163(1–2):27–32.
184. Bortolussi M, Marini G, Dal Lago A. Autoradiographic study of the distribution of LH(HCG) receptors in the ovary of untreated and gonadotrophin-primed immature rats. *Cell Tissue Res*. 1977;183(3):329–342.
185. Nimrod A, Erickson GF, Ryan KJ. A specific FSH receptor in rat granulosa cells: properties of binding in vitro. *Endocrinology*. 1976;98(1):56–64.
186. Jasuja R, Spencer D, Jayaraj A, et al. Estradiol induces allosteric coupling and partitioning of sex-hormone-binding globulin monomers among conformational states. *iScience*. 2021;24(6):102414.
187. Wu TS, Hammond GL. Naturally occurring mutants inform SHBG structure and function. *Mol Endocrinol*. 2014;28(7):1026–1038.
188. Ng KM, Catalano MG, Pinos T, et al. Evidence that fibulin family members contribute to the steroid-dependent extravascular sequestration of sex hormone-binding globulin. *J Biol Chem*. 2006;281(23):15853–15861.
189. Zhu BT, Lee AJ. NADPH-dependent metabolism of 17beta-estradiol and estrone to polar and nonpolar metabolites by human tissues and cytochrome P450 isoforms. *Steroids*. 2005;70(4):225–244.
190. Sandberg AA, Slaunwhite Jr WR. Studies on phenolic steroids in human subjects. II. The metabolic fate and hepato-biliary-enteric circulation of C14-estrone and C14-estradiol in women. *J Clin Invest*. 1957;36(8):1266–1278.
191. Tora L, White J, Brou C, et al. The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell*. 1989;59(3):477–487.
192. Kong EH, Heldring N, Gustafsson JA, et al. Delineation of a unique protein-protein interaction site on the surface of the estrogen receptor. *Proc Natl Acad Sci USA*. 2005;102(10):3593–3598.
193. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A*. 1996;93(12):5925–5930.
194. Kumar R, Thompson EB. Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Mol Endocrinol*. 2003;17(1):1–10.
195. Roche PJ, Hoare SA, Parker MG. A consensus DNA-binding site for the androgen receptor. *Mol Endocrinol*. 1992;6(12):2229–2235.
196. Klein-Hitpass L, Tsai SY, Greene GL, et al. Specific binding of estrogen receptor to the estrogen response element. *Mol Cell Biol*. 1989;9(1):43–49.

197. Schwabe JW, Chapman L, Finch JT, Rhodes D. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*. 1993;75(3):567–578.
198. Schwabe JW, Chapman L, Finch JT, Rhodes D, Neuhaus D. DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure*. 1993;1(3):187–204.
199. Klinge CM. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res*. 2001;29(14):2905–2919.
200. Seielstad DA, Carlson KE, Kushner PJ, Greene GL, Katzenellenbogen JA. Analysis of the structural core of the human estrogen receptor ligand binding domain by selective proteolysis/mass spectrometric analysis. *Biochemistry*. 1995;34(39):12605–12615.
201. Paterni I, Granchi C, Katzenellenbogen JA, Minutolo F. Estrogen receptors alpha (ERalpha) and beta (ERbeta): subtype-selective ligands and clinical potential. *Steroids*. 2014;90:13–29.
202. Shiau AK, Barstad D, Radek JT, et al. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat Struct Biol*. 2002;9(5):359–364.

106

HYPOGONADISM: PATHOGENESIS, DIAGNOSIS, AND TREATMENT

GIULIA RASTRELLI, LEEN ANTONIO, AND FREDERICK C.W. WU

OUTLINE

Introduction, 1783
Pathogenesis, 1784
Primary Hypogonadism, 1784
- Klinefelter Syndrome, 1784
- Other Sex Chromosome Disorders, 1786
- Cryptorchidism and Spermatogenic Failure, 1786
- Congenital Anorchia, 1786
- Acquired Anorchia, 1786
- Orchitis, 1786
- Gonadotoxic Cancer Treatment, 1786
- Noonan Syndrome, 1787
- Myotonic Dystrophy, 1787
- Autoimmune Polyglandular Syndrome, 1787
- Sickle Cell Disease, 1788

Secondary Hypogonadism, 1788
- Congenital Isolated Hypogonadotropic Hypogonadism, 1788
- Congenital Hypogonadotropic Hypogonadism in Complex Genetic Syndromes, 1790
- Congenital Combined Pituitary Hormone Deficiency, 1791
- Pituitary and Nonpituitary Tumors, 1791
- Nontumorous Hypothalamic-Pituitary Diseases/Injuries, 1791
- Drug-induced, 1792

Diagnosis of Hypogonadism, 1792
Clinical Diagnosis, 1792
- History, 1793
- Examination, 1793

Laboratory Diagnosis, 1794
- Testosterone, 1794

Treatment of Hypogonadism, 1797
Testosterone Replacement Therapy, 1797
- Indications, 1797
- Testosterone Preparations, 1797
- Efficacy, 1799
- Contraindications and Adverse Events, 1799
- Monitoring of Treatment, 1800

Gonadotropin Induction of Spermatogenesis and Fertility, 1800
- Indications, 1800
- Gonadotropin Treatment, 1800
- Gonadotropin-Releasing Hormone Treatment, 1801

Future Directions, 1801

INTRODUCTION

Hypogonadism in men is a clinical syndrome arising from failure of the testes to produce physiological amounts of testosterone (T) and/or normal numbers of spermatozoa due to pathologies or dysfunctions in one or more levels of the hypothalamic-pituitary-testicular (HPT) axis (see Chapter 105).

Androgen or T deficiency is synonymous with low circulating and target tissue concentrations of T, which lead to a variable range of androgen-dependent (androgen receptor-mediated gene transcription) physiological deficits, including defective spermatogenesis, manifesting as clinical features of hypogonadism. In contrast, primary/intrinsic disorders of spermatogenesis presenting as male infertility mostly exist in isolation, without overt evidence of T deficiency. This chapter addresses only the diagnosis and treatment of androgen deficiency in hypogonadism in the adult, while the management of spermatogenic disorders and delayed puberty are dealt with elsewhere (see Chapters 112 and 115 and 103).

Hypogonadism can be differentiated by the level of the underlying abnormality in the HPT axis. Testicular abnormalities give rise to primary hypogonadism, common causes of which include Klinefelter syndrome (KS) and cryptorchidism (Table 106.1), where low T concentrations and impaired spermatogenesis are coupled with elevated gonadotropins, reflecting the loss of negative feedback inhibition in the HPT axis. Hypothalamus or pituitary defects such as pituitary tumors and Kallmann syndrome (KallS) or normosmic isolated hypogonadotropic hypogonadism (nIHH) (Table 106.1) cause secondary hypogonadism, where low T concentrations and impaired spermatogenesis are associated with low or inappropriate normal gonadotropin concentrations. Multilevel abnormalities in the HPT axis can also coexist in hypogonadism,[1] usually with a predominant hormone pattern related to either primary (e.g., renal failure, aging) or secondary hypogonadism (e.g., alcohol abuse, systemic diseases) (Table 106.1).

Hypogonadism can also be differentiated by its underlying etiologies (Table 106.1). Organic hypogonadism, also known as classical or pathological hypogonadism, is caused by congenital, structural, vascular, inflammatory, or infiltrative diseases in the HPT axis, which are usually permanent and irreversible (Key Points: Definition and Classification). Functional hypogonadism is caused by suppression of gonadotrophin or T production, usually from conditions unrelated to pathology of the HPT axis, and may be reversible after correction or removal of the underlying disorder[2] (Key Points: Definition and Classification). The topic of T deficiency in aging and chronic illness is dealt with in more detail separately (see Chapter 107).

KEY POINTS: Definition and Classification

- Male hypogonadism is a clinical syndrome arising from the failure of the testes to produce sufficient testosterone to maintain secondary sexual, reproductive, and other androgen-dependent physiological functions.
- Consequences and clinical features of testosterone deficiency vary according to the stage of sexual development:
  - Ambiguous genitalia development characterizes fetal onset.
  - Delayed puberty characterizes childhood onset.
  - Sexual dysfunction, infertility, and alterations in adipose and musculoskeletal mass and function characterize adult onset.
- Organic hypogonadism is caused by structural, vascular, inflammatory, or infiltrative diseases in the hypothalamus-pituitary-testicular axis are recognized. These are usually permanent and irreversible.
- Functional hypogonadism is caused by conditions unrelated to pathology of the hypothalamus-pituitary-testicular axis and may be reversible after correction or removal of the underlying disorder.
- Functional hypogonadism associated with obesity/diabetes, opiate, and anabolic steroid abuse and advanced age is becoming increasingly prevalent.

It should be emphasized that male hypogonadism is a syndromic diagnosis characterized by suggestive clinical features in the presence of low T concentrations; as such, the underlying etiologies should be defined in order to ensure optimal management (Key Points: Diagnosis). Thus, differentiating primary from secondary hypogonadism has significant therapeutic implications. Spermatogenic defects in primary testicular disorders are not amenable to hormonal treatment, and fertility management is limited to options offered by assisted reproduction technologies. In contrast, dormant spermatogenesis in secondary hypogonadism can be stimulated or restored with appropriate exogenous gonadotropin therapy to promote fertility. Secondary hypogonadism arising from pituitary tumors will require specific treatment targeting tumor mass reduction and effects of pituitary hormone excess or deficiencies. Functional secondary hypogonadism associated with obesity, medications (e.g., opioid analgesics, anabolic steroids), or systemic illness is potentially reversible with weight reduction, discontinuation of medications, or successful treatment of or recovery from the underlying illness, respectively.

KEY POINTS: Diagnosis

- The diagnosis of hypogonadism should be made only in men with symptoms and signs consistent with androgen deficiency abetted by consistently and unequivocally low serum testosterone concentrations in fasting morning blood.
- Accurate and reliable measurement of testosterone is essential for the diagnosis of hypogonadism, but technical shortcomings and quality control issues can undermine this mission.
- Free testosterone measurement is indicated in the presence of sex hormone–binding globulin alterations, e.g., in obese and older men, and when testosterone is in the borderline low range.
- Measurement of gonadotrophins can differentiate secondary from primary hypogonadism, and direct further investigations to identify underlying etiologies and optimize specific treatment, e.g., for pituitary tumor or fertility promotion.

PATHOGENESIS

PRIMARY HYPOGONADISM

Both congenital and acquired causes can induce testicular damage and result in primary hypogonadism (Table 106.1). The hormonal profile is marked by elevated luteinizing hormone (LH) and follicle-stimulating hormone (FSH) concentrations, indicating Leydig cell damage with decreased sex steroid production and damage to the seminiferous tubules with impaired spermatogenesis and low inhibin B, respectively.

Klinefelter Syndrome

KS is the most frequent genetic disorder in men, with an estimated prevalence of one to two per 1000 men. Most men with KS have a supernumerary X chromosome (47,XXY), although other karyotypes (such as 48,XXXY, 48,XXYY) and in particular, mosaicisms (46,XY/47,XXY) are found in 10% to 20% of cases.[3] In 50% of patients, the additional X chromosome is from paternal origin and is the result of nondisjunction of sex chromosomes during the first meiotic division. The other 50% of cases have a maternal origin arising from nondisjunction in meiosis I or II, or during mitotic divisions in the early embryo.[3] Advanced parental age has been associated with the occurrence of both maternally and paternally derived KS.

Clinical Features and Diagnosis. The classic phenotype consists of small testes, gynecomastia, tall stature, hypergonadotropic hypogonadism, infertility, and learning disabilities (Fig. 106.1).[4] Other signs and symptoms may include cryptorchidism, decreased penis size, and signs of hypogonadism, such as decreased body hair growth, erectile dysfunction, and low libido (Fig. 106.1). It is important to emphasize that the clinical features in men with KS are highly variable. The classic phenotypic features may not always be extant (with the exception of testicular atrophy and infertility), resulting in diagnostic delay or nondetection (see later).

Diagnosis of KS is based on chromosome karyotype analysis. Few patients (<10%) are diagnosed during childhood or adolescence, when karyotype analysis is performed in boys with cryptorchidism, behavioral problems, or learning difficulties.[4] In adolescents and young adults, gynecomastia and small testis size can lead to the diagnosis. The majority of patients, however, are diagnosed in adulthood when they seek medical care for infertility. Remarkably, it is estimated that 50% to 75% of men with KS are never diagnosed, presumably because of the lack of characteristic clinical features drawing medical attention.[4]

Infertility. Most men (>95%) with KS are azoospermic, with elevated FSH, making it the most frequent genetic cause of azoospermia.[3] Testicular biopsy reveals mostly hyalinized seminiferous tubules with occasional Sertoli cells, as well as Leydig cell hyperplasia. Spermatozoa can be detected in the ejaculate in a limited number of patients (<10%, usually KS mosaic). In 30% to 50% of azoospermic men, focal spermatogenesis can be preserved, offering the opportunity for spermatozoa to be retrieved surgically by testicular sperm extraction (TESE) or microdissection TESE and used for intracytoplasmic sperm injection, with approximately 45% live birth rates in highly specialized centers (see Chapter 115).[4]

Hypogonadism. Almost all men with KS have elevated LH levels, reflecting some degree of Leydig cell dysfunction. T levels, are however more variable and as many as one third of patients can have T levels in the low-normal range. Testosterone replacement therapy (TRT) should be started in younger patients with delayed or incomplete puberty and in adult patients with hypogonadal symptoms and

TABLE 106.1 **Causes and classification of hypogonadism**

Organic or Classical Causes of Hypogonadism	Functional Causes of hypogonadism
Primary Hypogonadism	
Leydig cell dysfunction and impaired spermatogenesis: – *Klinefelter Syndrome* – Other (sex) chromosome disorders – Cryptorchidism (bilateral) – Congenital anorchia – Acquired anorchia (trauma, torsion, orchidectomy) – Orchitis (Mumps, HIV, brucellosis, leprosy, syphilis, tuberculosis, schistosomiasis, and filariasis) – Gonadotoxic cancer treatment (chemo/radiotherapy) – Noonan Syndrome – Myotonic dystrophy – Autoimmune polyglandular syndrome – *Hemoglobinopathies (Sickle cell disease, Thalassemia)* – Advanced age	– Drug-induced: ketoconazole, abiraterone – Renal failure[§]
Impaired spermatogenesis only: – *Y-chromosome microdeletions* – Sertoli-cell only syndrome – Spermatogenic arrest – *Cryptorchidism* (unilateral & bilateral) – *Orchitis* (*Mumps*, *HIV*, brucellosis, leprosy, syphilis, tuberculosis, schistosomiasis, and filariasis) – *Gonadotoxic cancer treatment (chemo/radiotherapy)*	
Secondary Hypogonadism	
– *Congenital IHH* – Congenital IHH in complex genetic syndromes – Congenital Combined Pituitary Hormone Deficiency – *Pituitary and non-pituitary tumors* – Nontumorous hypothalamic-pituitary diseases/injuries ■ Developmental hypothalamic and pituitary alterations (e.g. cysts) ■ Surgery ■ Radiotherapy ■ Traumatic brain injury ■ Apoplexy and other vascular accidents ■ Lymphocytic hypophysitis ■ Langerhans cell histiocytosis ■ Sarcoidosis ■ Infections (i.e. tuberculosis) ■ *Iron overload* ■ Empty sella	– *Hyperprolactinemia* – Cushing syndrome – Drug-induced ■ *Opiates* ■ *anabolic steroids* ■ *glucocorticoids* ■ GnRH analogs – Alcohol abuse[§] – *Severe obesity* – *Diabetes mellitus* – Systemic illnesses – Liver, heart, lung failure[§] – *Comorbid illnesses associated with ageing*[§] – Some sleep disorders – Nutritional deficiency – Excessive exercise

[§] Combined primary and secondary hypogonadism, but classified to usual predominant hormonal pattern.
Italic denotes the most frequent conditions.
Abbreviations: HIV: human immunodeficiency virus, IHH= Isolated Hypogonadotropic Hypogonadism, GnRH= Gonadotropin releasing hormone.

total and/or free T levels below the normal range. Some, but not all, pediatricians may prefer to proactively initiate TRT in puberty once LH becomes elevated to ensure normal progression and maximize virilization and psychosocial development. Normalizing T has a positive impact on erectile function, libido, moods, bone mineral density (BMD), muscle strength, body composition, and school or work performance. However, TRT can suppress residual spermatogenesis and influence to ability to recover occult sperm via testicular sperm extraction (TESE) (see Chapters 112 and 115 xxx). Fertility status should therefore be assessed, if possible, before starting TRT.

Associated Conditions. Men with KS have an increased risk for other medical and psychological conditions. As a consequence of androgen deficiency, Klinefelter patients have a lower BMD and a higher risk of fractures. However, low BMD can also be present in KS men who have T concentrations in the low-normal range. Follow-up of BMD via dual-energy x-ray absorptiometry (DXA) is recommended. Varicose veins/ulcers, venous thromboembolic events, and pulmonary embolism occur more frequently in patients with KS than in men from the general population. Men with KS have a higher risk for obesity with abdominal adiposity, type 2 diabetes, dyslipidemia, and the metabolic syndrome.[4] The prevalence of cardiovascular disease is higher, contributing to the observed increase in mortality in Klinefelter patients. Proactive screening and treatment of cardiovascular risk factors is recommended. The overall risk of cancer in men with KS is not increased. However, certain types of cancers occur more frequently. Breast cancer risk is approximately four

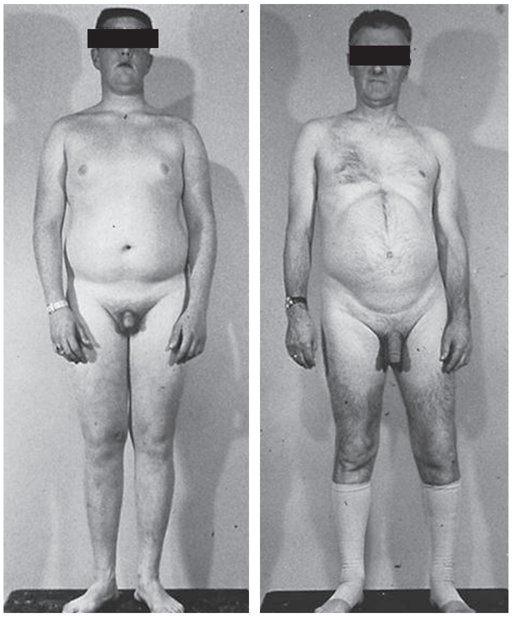


Figure 106.1 Variability in the degree of androgen deficiency manifested in men with Klinefelter syndrome. The man on the left, who has classic 47,XXY Klinefelter syndrome, demonstrates prepubertal androgen deficiency with eunuchoidal body proportions, small penis, sparse chest and pubic hair, poor muscle development, prepubertal fat distribution, and very small testes (2 mL bilaterally). The man on the right, who has mosaic 47,XXY/46,XY Klinefelter syndrome, demonstrates normal body proportions, penis size, and body hair but small testis size (8 mL bilaterally). (From Smyth CM, Bremner WJ. Klinefelter syndrome. *Arch Intern Med.* 1998;158:1309–1314; and Matsumoto AM, Bremner WJ. Testicular disorders. In: Melmed S, Polansky KS, Larsen PR, et al., eds. *Williams Textbook of Endocrinology*. 13th ed. New York: Elsevier; 2016:688–777.)

times higher in Klinefelter patients. The incidence of mediastinal germ cell tumors and hematological cancers is also increased. Autoimmune disorders, such as type 1 diabetes, hypothyroidism, and rheumatic disorders, appear to be more prevalent in men with KS. They can also have dental problems, retinal dysfunction, and tremor. Neurocognitive deficits and psychological morbidities are major problems in many Klinefelter patients. The most characteristic are learning difficulties, impaired speech and language skills (dyslexia), and impaired executive function. Anxiety, depression, autism, and schizophrenia also appear to occur more often in men with KS. Recent studies suggest there is an increase in criminality in KS patients compared with the general population.[4,5]

Other Sex Chromosome Disorders

Other sex chromosome disorders, such as 46,XX with a male phenotype, are discussed in Chapter 101.

Cryptorchidism and Spermatogenic Failure

Cryptorchidism is a prevalent condition affecting 2% to 4% of male full-term infants. Bilateral cryptorchidism is associated with many hypogonadal conditions (primary and secondary) and androgen resistance syndromes. Unilateral or bilateral cryptorchidism not associated with known causes of hypogonadism usually gives rise to isolated impairment of spermatogenesis with low sperm counts, normal T concentrations, and generally a selective elevation in FSH levels. Rarely, however, bilateral cryptorchidism causes Leydig cell failure as well, with low serum T and high LH and FSH. Azoospermia occurs in 50% to 60% and oligozoospermia in 75% to 100% of men with bilateral cryptorchidism. In unilateral cryptorchidism, azoospermia is present in 15% to 20% and oligozoospermia in 20% to 40% (see Chapters 114 and 115).

Cryptorchidism can be associated with hypospadias, infertility, and testicular cancer (testicular dysgenesis syndrome (see Chapter 114)).

Congenital Anorchia

Congenital anorchia is the absence of testicular tissue in 46,XY phenotypic males. In most cases, it is unilateral, and affected men are usually not hypogonadal. Bilateral anorchia in phenotypic males is rare (1:20,000 men). It is hypothesized that testicular torsion occurs when the testes descend into the scrotum during fetal development, resulting in testicular atrophy. As the external male genitalia are normally developed at birth (Fig. 106.2), testicular function had to be normal in first trimester of pregnancy (vanishing testis syndrome or testicular regression syndrome).

In patients with impalpable testes, undetectable antimullerian hormone (AMH) is suggestive of bilateral anorchia, while normal AMH levels predict the presence of intraabdominal testes (bilateral cryptorchidism) and a normal T response to human chorionic gonadotrophin (hCG) stimulation (1000–2000 IU three times weekly for 2 weeks).[6] Imaging studies and laparoscopy or surgical abdominal exploration are necessary for definitive exclusion of cryptorchidism.[7] Patients with bilateral congenital anorchia are infertile and require puberty induction and lifelong TRT.[8]

Acquired Anorchia

Scrotal trauma, damage to vascular structures during testicular surgery, testicular torsion, and orchidectomy (e.g., for testicular cancer) can lead to the loss of one or both testes. The remaining testis will usually hypertrophy to compensate for the loss of one testis in response to increased gonadotrophins.

Orchitis

Inflammation of the testis (orchitis) usually occurs together with inflammation of the epididymis (epididymitis). Mumps orchitis used to be a common cause of orchitis.[9] After introduction of widespread vaccination, the incidence of mumps orchitis decreased, although sporadic outbreaks are reported despite high vaccination coverage. Approximately 15% to 30% of adolescents and adult men who are infected with the mumps virus develop orchitis, which is bilateral in 20% to 40% of cases. Following mumps orchitis, testicular atrophy and decreased fertility can occur, although the development of azoospermia is rare and occurs more frequently in men with bilateral orchitis.

Primary hypogonadism used to be prevalent in men with acquired immunodeficiency syndrome (up to 30%–50%).[10,11] This is caused by orchitis induced by the human immunodeficiency virus (HIV) infection itself or secondary to opportunistic infections in immunocompromised patients, such as toxoplasmosis or cytomegalovirus infection. Since the introduction of highly active antiretroviral therapy, the prevalence of hypogonadism in men with HIV has decreased.[11]

Other infectious diseases, such as brucellosis, leprosy, syphilis, tuberculosis, schistosomiasis, and filariasis occasionally also affect the testis and can induce primary testicular failure.

Gonadotoxic Cancer Treatment

Cancer treatment with chemotherapy or irradiation can severely damage the testis and induce transient or permanent impairment of spermatogenesis and infertility. Often, male cancer patients have impaired spermatogenesis already at the time of cancer diagnosis, before any cytotoxic treatment is started.[12]

The spermatotoxicity of cancer treatments depends on the type and combination of agents and is dose-dependent. Irradiation of greater

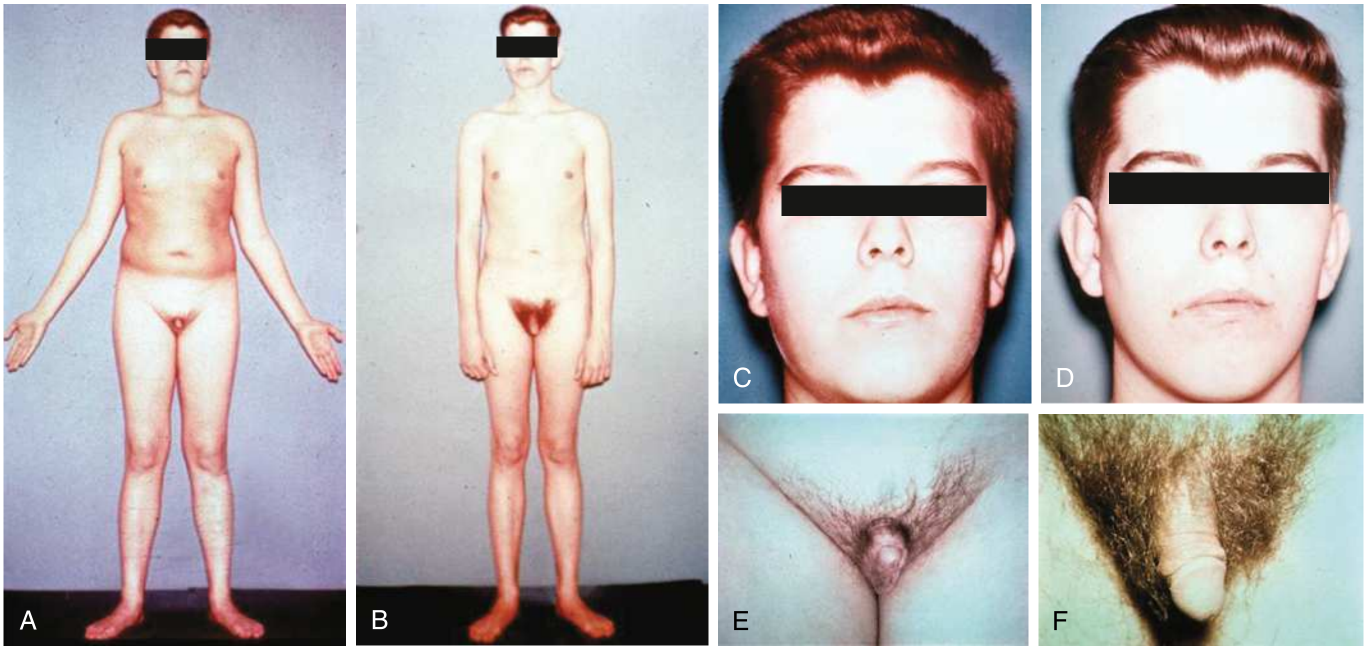


Figure 106.2 A 19-year-old patient with prepubertal androgen deficiency caused by congenital anorchia before (A, C, and E) and after (B, D, and F) 5 years of testosterone treatment. Before testosterone treatment, the patient had features of eunuchoidism, characterized by infantile genitalia (small penis and poorly developed scrotum); lack of chest, pubic, and facial hair; long arms and legs relative to height; and poorly developed muscle mass in the upper body with accumulation of fat in the face, chest, and hips. After testosterone treatment, there was an increase in penis size; an increase in chest, pubic, and facial hair with scalp recession and development of acne; an increase in muscle mass, particularly in the upper body; and loss of fat in the face, chest, and hips. (From Matsumoto AM. The testis. In: Felig P, Frohman LA, eds. *Endocrinology and Metabolism*. 4th ed. New York: McGraw-Hill; 2001:635–705; and Matsumoto AM, Bremner WJ. Testicular disorders. In: Melmed S, Polansky KS, Larsen PR, et al., eds. *Williams Textbook of Endocrinology*. 13th ed. New York: Elsevier; 2016:688–777.)

than 2.5 Gy to the testis, total body irradiation, and the use of alkylating agents (such as cyclophosphamide, melphalan, chlorambucil, busulfan, and procarbazine) are associated with the highest risk of developing prolonged azoospermia in men.[13] Other chemotherapeutic agents (such as cisplatin) can also affect spermatogenesis, resulting in severe oligospermia or azoospermia. Recovery is possible, but it can take months to years, and oligospermia will persist in a substantial number of men.[12] It is now standard of care to refer all patients of reproductive age with a recent cancer diagnosis for fertility counseling and sperm cryopreservation before any gonadotoxic treatment is started.

Chemotherapy and radiotherapy (>600 cGy) can also induce Leydig cell dysfunction, although Leydig cells are more resistant to gonadotoxic agents than the seminiferous epithelium. Long-term cancer survivors can develop compensated or primary hypogonadism later in life. In testicular cancer survivors, the risk for developing T deficiency was highest in men receiving high doses of cisplatin. Men with testicular cancer treated with standard chemotherapy or infradiaphragmatic radiotherapy had a higher risk of developing T deficiency compared with those treated with orchiectomy alone.[14]

Noonan Syndrome

Noonan syndrome is a relatively rare autosomal dominant disease with an estimated incidence of one in 1000 to 2500.[15] A genetic defect can be documented in 80% of cases with a *de novo* (66%) or inherited (33%) mutation in genes encoding for proteins of the RAS-MAPK pathway (*PTPN10* in 50% of cases).[15] Clinical manifestations are heterogeneous and include peculiar facial features (triangular-shaped face, hypertelorism, down-slanting palpebral fissures, micrognathia, and low-set ears), as well as short stature, coagulation disorders, and cardiomyopathy.[15] Cryptorchidism occurs in 70% of boys with Noonan syndrome. Delayed puberty is frequent, although T tends to normalize later in life.[16] Spermatogenesis may be compromised; this may result from cryptorchidism, but it can also be specific to the underlying mutations, as RAS plays a role in spermatogenesis.

Myotonic Dystrophy

Myotonic dystrophy type 1 (Steinert disease) is a rare autosomal dominant disorder affecting 1:8000 people and caused by expansion of *CTG* trinucleotide repeat in the dystrophia myotonica protein kinase gene (*DMPK*). It is characterized by myotonia (involuntary, sustained muscle contraction), muscle weakness/wasting, cardiac conduction abnormalities, frontal balding, and cataract. Testicular atrophy, infertility, and primary hypogonadism are common (80%) in men with myotonic dystrophy type 1. Myotonic dystrophy type 2 is less severe and is caused by *CTG* repeat expansion in the CCHC-type zinc finger, nucleic acid-binding protein gene (*CNBP*).

Autoimmune Polyglandular Syndrome

Mutations in the autoimmune regulator gene (*AIRE*) on chromosome 21 induce autoimmune polyglandular syndrome (APS) type 1. It is an autosomal recessive disorder, characterized by hypoparathyroidism, adrenal insufficiency, and mucocutaneous candidiasis, but other autoimmune diseases can also occur. Primary hypogonadism in adult men with APS is linked to the presence of antibodies targeting side-chain cleavage enzyme in the steroidogenic pathway and occurs in 5% to 7% of patients with APS. Azoospermia is rarely observed.[17]

Sickle Cell Disease

Homozygous sickle cell disease can be associated with primary testicular failure, but secondary hypogonadism can also occur. Vasoocclusion of testicular blood vessels resulting in recurring microinfarctions may be the cause, but other mechanisms have also been postulated. Men with sickle cell disease can also have impaired spermatogenesis.[18]

SECONDARY HYPOGONADISM

Secondary hypogonadism, also known as hypogonadotropic hypogonadism, may result from an intrinsic impairment of pituitary structure/function or a failure of hypothalamic gonadotropin-releasing hormone (GnRH) secretion. Secondary hypogonadism may also be the consequence of a combined dysfunction of hypothalamus and pituitary. Gonadotropin deficiency can be either isolated or combined with other pituitary hormone insufficiencies, and it may result from congenital or acquired diseases. Hypogonadotropic hypogonadism may potentially result from any pathology involving the hypothalamus-pituitary unit. As a consequence of these disorders, hypogonadism may occur alone or together with multiple hormone deficiencies. These conditions are discussed in more detail in chapters 4, 6-11, and 16. In this chapter, the secondary hypogonadism aspects will be addressed.

Congenital Isolated Hypogonadotropic Hypogonadism

Congenital isolated hypogonadotropic hypogonadism (IHH) was formerly known as idiopathic hypogonadotropic hypogonadism in order to denote the absence of structural abnormalities of the hypothalamus/pituitary. IHH has a global prevalence of one per 4000 to 10,000 males.[19] It is now known that mutations in multiple genes involved in GnRH secretion and/or action are the causes of IHH. Nevertheless, half of tested IHH patients still have no identifiable genetic defects.[20] Genes involved in IHH can be classified into (i) genes encoding factors, which play a role in normal differentiation, development, or migration of GnRH neurons from their early embryonal location in the olfactory placode to the hypothalamus and (ii) genes involved in normal GnRH secretion or action. Because the former processes take place in close association with developing axons of olfactory nerves, IHH arising from mutations in these genes is frequently associated with impairment in sense of smell. IHH with an-hyposmia is termed KallS, whereas IHH with normal sense of smell is called nIHH.

At least 30 genes have been implicated in IHH,[21] with an inconsistent genotype–phenotype correlation due to low penetrance and variable expressivity among subjects carrying the same mutation. A familial aggregation is recognized in over 40% of cases. IHH, originally deemed to be a monogenic disorder, is increasingly recognized as an oligogenic (multiple genes) disease, thus amplifying the spectrum of inheritance patterns.

Kallmann Syndrome. KallS represents approximately half of IHH cases. Although IHH is characterized by a limited genotype–phenotype correlation, mutations in some genes tend to be more specifically associated with KallS[21] (Table 106.2). The most frequent amongst these is *ANOS1* (formerly denoted as *KAL1*), encoding an extracellular matrix glycoprotein termed anosmin-1, which participates in the normal embryonal migration of GnRH neurons. Mutations of *ANOS1* are found in up to 60% of familial cases of KallS and in up to 15% of sporadic forms. *ANOS1* is located on chromosome Xp22.3 and, accordingly, familial forms of KallS due to *ANOS1* mutations follow an X-linked inheritance pattern. Other genes are involved in KallS development, but, unlike *ANOS1*, their mutations could also be responsible for nIHH or even congenital combined pituitary hormone deficiency (CPHD) (Table 106.2). These mutations may arise *de novo* or be inherited with autosomal dominant or recessive transmission. Increasingly more cases are attributed to two or more genes (oligogenicity) that do not follow the classical Mendelian inheritance pattern. The list of implicated genes is constantly growing, but among the most frequently involved are *FGFR1* (formerly named *KAL2*), *PROKR2*, *PROK2*, *SEMA3A*, and *CHD7* (Table 106.2). *FGFR1* is the first gene associated with autosomal dominant forms of KallS, although now autosomal recessive or oligogenic forms due to *FGFR1* mutations are also known. FGF8 is one of FGFR1 ligands, and mutations in its codifying gene are rare causes of KallS/nIHH.

CHD7 (Table 106.2) was initially recognized as the gene responsible for the CHARGE syndrome (coloboma, heart defects, atresia choanae, retarded growth, genital hypoplasia, ear abnormalities/deafness). Later, *CHD7* mutations were found in attenuated forms of CHARGE syndrome that had underdevelopment of genitalia and smell impairment as major features. Moreover, missense *CHD7* mutations with partial loss of function were associated with KallS or even nIHH. At present, *CHD7* and *FGFR1* are the most frequently implicated in KallS/nIHH, with autosomal transmission.

PROKR2 and *PROK2* mutations are implicated in KallS/nIHH (Table 106.2). In up to 20% of familial forms of KallS/nIHH attributable to *PROKR2/PROK2,* biallelic mutations (homozygous or compound heterozygous) denote autosomal recessive transmission. However, several subjects are heterozygous, indicating autosomal dominant or oligogenic transmission. Nonetheless, rare *PROKR2/PROK2* mutations are also prevalent in the general population, so finding *PROKR2/PROK2* mutations in KallS/nIHH subjects may not yield a conclusive genetic diagnosis. *PROKR2/PROK2* mutations are rarely associated with nonreproductive phenotypic defects that are observed in KallS or nIHH (see later).

After the first recognition in a family with several KallS cases, *SEMA3A* mutations have been recognized in an increasing number of KallS patients in a heterozygous state (suggesting autosomal dominant inheritance), sometimes in association with other KallS/nIHH genes (suggesting oligogenicity) (Table 106.2).

Besides an/hyposmia, KallS is characterized by clinical features of fetal-onset hypogonadism, which include cryptorchidism, microphallus, and hypospadias, and of delayed puberty. Additionally, nongenital/nonolfactory defects can be variably present (Box 106.1). It is noteworthy that hypogonadism can resolve in up to 10% to 15% of cases of KallS/nIHH.[22] Reversal to eugonadism, usually in adulthood, is denoted by increase in testes size and a sustained normalization of T and gonadotropin levels after withdrawal of T treatment. The mechanism of reversal is still unclear. It has been proposed that the increase in circulating T levels from treatment could upregulate the expression of genes involved in GnRH secretion. Moreover, in animal models, hypothalamic cells have the potential to differentiate into GnRH neurons, thus replacing those missing from fetal development. Reversal has been documented in subjects with different genetic defects, including *FGFR1, PROKR2, ANOS1,* and *CHD7* (Table 106.2). However, reversal may not be permanent, and relapse has been documented in a number of patients who had apparently recovered from hypogonadism.

Normoosmic Congenital Isolated Hypogonadotropic Hypogonadism. The distinction between KallS and nIHH is based on the presence or absence of an/hyposmia. However, the dividing line is not always clear-cut. Indeed, several genes may be implicated in both disorders, and the same mutation may even cause both clinical conditions in different patients (Table 106.2). Similar to KallS, nIHH may also be associated with nongenital disorders. Despite the overlap between KallS and nIHH, there are genes whose mutations have never been associated with smell impairment and, for this reason, are considered specific to nIHH (Table 106.2). These include *GNRH1*, *GNRHR*, *KISS1*, *KISS1R*, *TAC3*, and *TACR3*. Notably, mutations in these genes linked with nIHH are also not associated with any other nongenital conditions (Table 106.2). *GNRHR*, encoding the GnRH receptor, was the first gene causing nIHH that was identified, and it is the most frequent genetic defect in nIHH patients. Mutations in *GNRH1*, encoding GnRH, are rare causes of this disorder. The *KISS1* gene encoding kisspeptin 1 is expressed in the hypothalamus, whereas *KISS1R*, formerly known as *GPR54*, encoding its cognate G protein-coupled receptor, is expressed in the hypothalamus and pituitary. The interaction of kisspeptin 1, secreted by hypothalamic KISS1 neurons, with KISS1R is responsible for the normal processing and release of GnRH and thought to be pivotal for the onset of puberty. Biallelic mutations in *KISS1R* have been found in nIHH-affected families, suggesting autosomal recessive

TABLE 106.2 Genes Most Commonly Associated with Congenital Isolated Hypogonadotropic Hypogonadism (IHH)

Gene *Encoded Protein*	Chr Location	Biological activity	Inheritance	KallS	nIHH	Other Phenotypes	IHH Reversal	Nonreproductive Clinical Features
ANOS1 *Anosmin-1* *Extracellular matrix glycoprotein*	Xp22.31	Normal embryonal migration of GnRH neurons	X-linked	X	-	-	X	Yes
FGFR1 *Fibroblast growth factor receptor 1* *Receptor*	8p11.23	Broad range of functions during embryogenesis, including the development of olfactory bulbs	AD/AR/OG/DN	X	X	CPHD ±SOD	X	Yes
FGF8 *Fibroblast growth factor 8, ligand of FGFR1*	10q24.32	Development of forebrain, olfactory and GnRH system, skeletal structure and heart	AD/OG	X	X	CPHD	-	Yes
FGF17 *Fibroblast growth factor 17* *Ligand of FGFR1*	8p21.3	Development of dorsal frontal cortex	OG	X	X	-	-	?
HS6ST1 Heparan-sulphate 6O-sulpho-transferase-1	2q14.3	Catalyzes the transfer of sulfate to heparan sulfate, in FGFR1 signaling	OG	X	X	-	X	?
IL17RD Interleukin-17 receptor D	3p14.3	Negative regulator of FGFR1	OG	X	X	-	-	?
CHD7 *Chromodomain helicase DNA binding protein 7* *Transcriptional regulator*	8q12.2	Development of neural crest forebrain, craniofacial bones, and heart	AD/AR/OG/DN	X	X	CHARGE	X	Yes
PROK2 *Prokineticin-2* *Ligand of PROKR2*	3p13	Development of olfactory bulbs	AD/AR/OG	X	X	-	-	No
PROKR2 *PROK2 receptor* *G protein–coupled receptor*	20p12.3	Development of olfactory bulbs	AD/AR/OG	X	X	CPHD	X	No
NSMF *Nasal embryonic luteinizing hormone–releasing hormone factor*	9q34.3	Development of olfactory axons required for the migration of GnRH neurons	OG	X	X	-	X	?
WDR11 *WD repeat protein 11*	10q26.12	Expressed in the forebrain and hypothalamus-pituitary-gonadal axis	?	X	X	CPHD	-	?
SEMA3A *Semaphorin-3A*	7q21.11	Guidance molecule for GnRH neuronal migration	AD/OG	X	X	-	-	No
SEMA7A *Semaphorin-7A*	15q24.1	Guidance molecule Favors axon development	OG	X	X	-	-	?
SEMA3E *Semaphorin-3E*	7q21.11	Guidance molecule Involved in GnRH neurons survival	OG	X	X	CHARGE	-	?
GNRH1 *Gonadotropin-releasing hormone*	8p21.2	Ligand of GnRHR-stimulating gonadotrophs	AR/OG	-	X	-	-	No
GNRHR *GnRH receptor* *G protein-coupled receptor binding GnRH*	4q13.2	Expressed in the pituitary gonadotrophs Regulation of gonadotropin secretion	AR/OG	-	X	-	X	No

Continued

TABLE 106.2 Genes Most Commonly Associated with Congenital Isolated Hypogonadotropic Hypogonadism (IHH)—cont'd

Gene *Encoded Protein*	Chr Location	Biological activity	Inheritance	KallS	nIHH	Other Phenotypes	IHH Reversal	Nonreproductive Clinical Features
KISS1 *Kisspeptin-1* *Ligand of KISS1R*	1q32.1	Secreted by the neurons of arcuate and anteroventral periventricular nuclei; regulation of GnRH neurons	AR	-	X	-	-	No
KISS1R *G protein-coupled receptor for kisspeptin-1*	19p13.3	Expressed in hypothalamic GnRH neurons Regulation of GnRH neurons	AR	-	X	-	-	No
TAC3 *Tachykinin-3 Ligand of TACR3*	12q13.3	Secreted by neurons of arcuate nucleus; regulation of the secretion and homeostasis of GnRH neurons	AR	-	X	-	X	No
TACR3 *G protein-coupled receptor for TAC3*	4q24	Expressed in hypothalamic GnRH neurons Regulation of the secretion and homeostasis of GnRH neurons	AR	-	X	-	X	No
LEP *Leptin* *Adipocyte-specific hormone*	7q32.1	Regulation of food intake and energy expenditure	AR	-	X	Morbid obesity	-	No
LEPR *Leptin receptor* *Single-transmembrane domain receptor for leptin*	1p31.1	Expressed in hypothalamus Regulation of food intake and energy expenditure	AR	-	X	Morbid obesity	-	No
FSHB *β-subunit of FSH*	11p14.1	Together with α-subunit makes FSH, which stimulates Sertoli cells	AR	-	-	Isolated FSH deficiency	-	No
LHB *β-subunit of LH*	19q13.33	Together with α-subunit makes LH, which stimulates Leydig cells	AR	-	-	Isolated FSH deficiency	-	No

Chr, Chromosome; *KallS*, Kallmnn syndrome; *nIHH*, normosmic isolated hypogonadotropic hypogonadism; *IHH*, isolated hypogonadotropic hypogonadism; *AD*, autosomal dominant; *AR*, autosomal recessive; *OG*, oligogenic; *DN*, *de novo*; *CPHD*, combined pituitary hormone deficiency; *SOD*, septooptic dysplasia; *CHARGE*, coloboma, heart defects, atresia choanae, retarded growth, genital hypoplasia, ear abnormalities/deafness; *FSH*, follicle-stimulating hormone; *LH*, luteinizing hormone; *GnRH*, gonadotropin-releasing hormone.

BOX 106.1 Extragenital Nonolfactory Congenital Defects Occurring in Congenital Hypogonadotropic Hypogonadism

Neurologic Defects	Nonneurological Defects
Sensorineural deafness	Cleft palate
Nystagmus	Cleft lip
Color blindness	Dental agenesis (hypodontia)
Bimanual synkinesia (mirror movements)	Syndactyly or other hand-foot malformations
Cerebellar ataxia	Renal agenesis or malformations
Spastic paraplegia	Pes cavus
Seizure	

inheritance. Mutations in *KISS1* are much rarer, but they seem to follow the same inheritance pattern. *TAC3* encodes tachykinin-3, precursor of neurokinin-B. It is secreted by hypothalamic neurons in the arcuate nucleus and infundibular region afferent to GnRH neurons. Its cognate receptor TACR3 is expressed in the same area. They interact in normal GnRH secretion, and their mutations have been recognized in nIHH families following an autosomal recessive inheritance pattern. Similar to KallS, nIHH can also spontaneously revert (Table 106.2), most frequently in those with *GNRHR* mutations.

Selective Deficiency of Luteinizing or Follicle-Stimulating Hormone. This is a rare form of nIHH characterized by a selective deficiency of one of the gonadotropins due to mutations in genes encoding for the β chain of LH (*LHB*) or FSH (*FSHB*). Boys with *LHB* mutations are characterized by severe hypogonadism with low T levels and undetectable LH, small testicles, and azoospermia. They fail to initiate puberty and, in some cases, present in adulthood with severe hypogonadism. *FSHB* mutations are even rarer. The few affected males have normal T and LH, but undetectable FSH. Azoospermia is the predominant clinical feature.

Congenital Hypogonadotropic Hypogonadism in Complex Genetic Syndromes

These rare conditions are usually diagnosed in infancy and childhood from the clustering of characteristic developmental and dysmorphic clinical features. These patients are usually introduced into adult practice in transitional clinics for continued endocrine management.

Prader-Willi syndrome. Prader–Willi Syndrome (PWS) is a rare genetic condition with a prevalence of one in 10,000 to 30,000 live births due to deletions (65%–75% of PWS patients), maternal uniparental disomy (23%–30%), or imprinting defects (3%) in a critical region on chromosome 15q11-q13.[23] Several genes important for hypothalamic functioning are located in the 15q11-q13 region, and most PWS clinical features are attributable to hypothalamic dysfunction, including hyperphagia with food obsession leading to pathological obesity, behavioral disorders, increased pain threshold, temperature instability, and sleep apnea.[23] Endocrine dysfunctions in PWS derive from deficient hypothalamic stimulation of the pituitary. PWS subjects therefore frequently have coexisting hypopituitarism, with somatotropic and gonadotropic axes most affected (almost 100% of subjects).[23] Hypogonadism in PWS can also have a primary testicular component in up to 60% of patients, with severe seminiferous tubule hypoplasia and high FSH, even though T and LH are low.

Laurence-Moon-Bardet-Biedl syndrome. (Laurence–Moon) Bardet–Biedl syndrome is an autosomal recessive disease with a prevalence of one in 125,000 to 175,000.[24] More than 20 genes have been associated with this syndrome, all with important roles in cellular cilia development and function.[24] Hypogonadotropic hypogonadism is a frequent finding, together with multiple malformations (eyes and retina, mouth and teeth, hands and feet, kidney), neurologic alterations (cognitive impairment, developmental delay, ataxia), growth retardation, and obesity.[24]

Septo-optic dysplasia. Septooptic dysplasia is a rare condition with a prevalence of around one in 10,000, characterized by midline and forebrain abnormalities, including absent septum pellucidum and optic nerve and pituitary hypoplasia.25 Hypopituitarism is observed in 40% of patients. Multiple endocrine axes may be affected in this condition, and the deficiencies may accumulate throughout life.[25] A genetic defect is found only in a limited proportion of patients. Most cases are sporadic, although familial forms are recognized based on mutations of genes involved in the normal forebrain development, including *HESX1*, *SOX2*, *SOX3*, and *OTX2*.[25]

DAX-1 mutations. DAX-1 mutations are associated with a syndromic condition characterized by the combination of primary adrenal insufficiency (adrenal hypoplasia) and congenital secondary hypogonadism. *DAX-1*, located on chromosome Xp21.2, encodes an orphan receptor expressed in the adrenal, gonads, hypothalamus, and pituitary and participates to the development of these structures. Accordingly, hypogonadism in this syndrome is a combination of primary (testicular) and secondary (hypothalamus-pituitary) dysfunction. Clinical manifestations of hypogonadism are variable, with subjects presenting with normal or delayed puberty and different degrees of altered spermatogenesis. Due to the mixed origin of hypogonadism, patients with *DAX-1* mutations do not have an optimal response to gonadotropin therapy.

Congenital Combined Pituitary Hormone Deficiency

Mutations in transcription factors involved in the development of pituitary may cause congenital CPHD. Further details on CPHD are provided in Chapters 5 and 14.

Pituitary and Nonpituitary Tumors

Pituitary tumors are mainly adenomas that are usually intrasellar masses and represent 10% of all the intracranial neoplasms. Hormone deficiency, including secondary hypogonadism, may result from space-occupying destructive effects, which are more common for macroadenomas,[26] and/or from a specific inhibiting effects of hormone excess. Typically, prolactin-secreting adenomas can impair gonadotropin release through downregulation of GnRH, whereas adrenocorticotropic hormone (ACTH) hypersecretion can induce hypogonadism because of cortisol excess.

Nonpituitary tumors may be either intrasellar or suprasellar and affect gonadotropin secretion due to a space-occupying destructive effect and/or pituitary stalk disconnection. The most common primary tumors in the hypothalamus-pituitary region include craniopharyngioma, meningioma, germ cell tumors, chordomas, lymphoma, and glioma. In addition, this region may be also affected by intracranial metastasis from distant primary malignancies.

Nontumorous Hypothalamic-Pituitary Diseases/Injuries

Developmental. Developmental. Cystic lesions including Rathke's cleft, arachnoid, and dermoid cysts can affect the hypothalamus and pituitary. Secondary hypogonadism occurs as a result of gonadotroph damage or from hyperprolactinemia (pituitary stalk disconnection).

Surgery and Radiotherapy. Surgery and radiotherapy. While pituitary or hypothalamus surgery commonly causes multiple pituitary hormone deficiencies, external beam radiotherapy produces more selective effects. Gonadotrophs are more likely to be injured by cranial radiation doses greater than 40 Gy, but lower doses can impair their function when radiotherapy targets the pituitary directly.

Traumatic Brain Injury. Traumatic brain injury results in deficiency of at least one pituitary hormone in up to 40% of patients[27] depending on the severity of injury. Secondary hypogonadism, together with growth hormone (GH) deficiency, is the most frequent occurrence, with an incidence of approximately 15%.[27,28] The hormone impairments may occur several months after the traumatic event, but they can also revert spontaneously over time.

Apoplexy is a rare complication of pituitary adenomas. It results from hemorrhage or infarction of pituitary, traumatic brain injury, hypotension (including postpartum hemorrhage, so-called Sheehan syndrome), and rarely, some endocrine stimulation tests.[29] Pituitary apoplexy presents acutely with headache, visual disturbances, nausea and vomiting, photophobia, and meningism. Acute endocrine deficiencies may occur; among these, the most frequent is ACTH/cortisol deficiency that may cause life-threatening Addisonian crises. Secondary hypogonadism is evident at presentation in 40% to 75% of apoplectic patients.[29]

Lymphocytic hypophysitis. Lymphocytic hypophysitis is an inflammatory condition of the pituitary thought to be autoimmune-mediated, but the exact pathogenesis is poorly understood. It occurs more frequently during pregnancy and postpartum and is also associated with the newly introduced checkpoint inhibitor immunotherapies such as ipilimumab and nivolumab. Deficiencies in gonadotropins, vasopressin (leading to diabetes insipidus) and ACTH are the earliest and most frequent abnormalities.

Langerhans cell histiocytosis. Langerhans cell histiocytosis is a rare systemic disease characterized by uncontrolled proliferation of Langerhans cells, a specific cell type of the monocyte-macrophage lineage. Secondary hypogonadism and GH deficiency are the most frequent anterior pituitary deficiencies.[30]

Sarcoidosis. Sarcoidosis is a systemic disorder of unknown etiology characterized by the development of noncaseating granulomas in multiple organs and tissues. Hypothalamic or pituitary involvement is rare and leads to diabetes insipidus and anterior pituitary hormone deficiencies.

Iron Overload. Iron overload. Congenital iron metabolism disorders (i.e., hemochromatosis) or congenital or acquired diseases (haemoglobinopathies) requiring frequent blood transfusions may lead to iron overload in several tissues, including liver, heart, pancreas, and pituitary. Iron deposition seems to have a predilection for gonadotrophs, so that iron overload conditions are often complicated by severe secondary hypogonadism without impairment of other hormone axes.

Empty Sella. Empty sella is an anatomic and radiologic finding of herniation of the subarachnoid space within the sella turcica, which causes a variable degree of flattening of the pituitary. Pituitary deficiency is identified in up to 20% of patients, and gonadotropin deficiency is described in up to 30% of these.[31]

Drug-induced

Anabolic androgen steroids. Anabolic androgenic steroids are synthetic derivatives of T with modifications aimed at increasing the anabolic over the androgenic actions of the steroid. These are mainly used as illicit enhancers of physical performance in almost any sport at any level. There have also been attempts to use these drugs to improve quality of life in cachectic cancer patients. Although modified structurally, these androgens will inhibit GnRH-gonadotropin secretion, with consequent suppression of testicular function (both steroidogenesis and spermatogenesis), thus inducing secondary hypogonadism. The topic is described in detail in Chapter 110.

Opiates. Opiates are normally used as analgesic medications, but their abuse as chronic addictive substances has become increasingly prevalent in recent years.[32] Strong opiates (such as morphine, fentanyl, methadone, heroin) cause secondary hypogonadism in up to 90% of men, irrespective of route of administration. Hypogonadism occurs more rarely with buprenorphine, a partial μ-opioid agonist, used as an alternative to methadone for opiate withdrawal treatment.[33] Hypogonadism is due to an inhibitory effect of opiates on GnRH resulting in suppressed LH and T levels without affecting FSH. In heroin addicts, but not in chronic oral opioid users and abusers, sex hormone-binding globulin (SHBG) is elevated. It is generally underappreciated that hypogonadal symptoms are very common with chronic use of opiates.[33] These include low libido, erectile dysfunction, delayed ejaculation, decreased BMD, and depression. The use of opiate antagonists, such as naltrexone, may reverse some features of hypogonadism. Sexual function improves with naloxone treatment even before normalization of T levels, suggesting a direct effect of opiates on male sexuality.[33]

GnRH analogs. GnRH analogs are synthetic long-acting analogs of native GnRH. In men, they are usually administered intramuscularly monthly to once every three months for the treatment of advanced prostate cancer. After an initial and transient increase in gonadotropin and T concentrations (flare effect), GnRH analogs induce the downregulation of GnRH receptors. This leads to a sharp decline in gonadotropin and T levels. This is the desired therapeutic effect aimed at reducing the androgenic stimulation of prostatic malignant cell proliferation. However, the resulting severe hypogonadism is associated with troublesome side effects, including hot flushes, sexual dysfunction, fatigue, depression, osteoporosis and fragility fractures, gain in fat mass and loss of lean mass leading to metabolic consequences (dyslipidemia, impaired glucose tolerance, hypertension), and increased cardiovascular risk.

DIAGNOSIS OF HYPOGONADISM

CLINICAL DIAGNOSIS

The clinical features of androgen deficiency can be distinguished by the stage of sexual development at onset (Table 106.3). Fetal-onset hypogonadism causing incomplete development of male genitalia can present at birth with microphallus, hypospadias, bifid scrotum, and cryptorchidism. Prepubertal onset hypogonadism results in a failure of T-induced secondary sexual development, giving rise to a specific clinical picture of eunuchoidism (Fig. 106.2). These include underdeveloped infantile external genitalia, small, smooth unpigmented scrotum with small testes (<4 mL), lack of secondary sexual hair (pubic, axillary, chest, limbs, and face) with no male escutcheon or temporal hair recession, poor muscle development, female body fat distribution (hips and chest), unbroken high-pitched voice without laryngeal enlargement, and eunuchoidal body habitus of long arms and legs. Gynecomastia, lack of sexual interest, spontaneous erections, and nocturnal emission with an immature and diffident demeanor complete the clinical picture, which can persist into adulthood if diagnosis is delayed. Osteopenia and reduced hematocrit are common.

Postpubertal or adult-onset hypogonadism arises from the inability to maintain previously established androgen-dependent functions such as sexual behavior/activity, spermatogenesis, muscle/bone mass, and facial/body/pubic hair. Clinical features include sexual dysfunction (decreased sexual interest and activity, reduced spontaneous/nocturnal erections, and erectile impotence), infertility (defined as an inability to father children with at least weekly unprotected vaginal intercourse for at least 12 months) with reduced testicular volume, particularly in primary hypogonadism, gynecomastia, hot flushes, fine, dry, wrinkling, smooth facial skin, height loss from vertebral compression fractures, fragility fractures, and a normochromic normocytic anemia. Depending on the severity and duration of androgen deficiency, some loss of secondary sexual hair may be present. However, adult secondary sexual characteristics such as genitalia, scrotum, larynx, and vocal cords generally do not regress to their prepubertal states. Less specific

TABLE 106.3 Symptoms and Signs Suggestive of Androgen Deficiency in Adult Men

Specific	Suggestive	Nonspecific
– Incomplete or delayed sexual development – Loss of body/pubic hair – Very small testes (<6 mL)	– Reduced sexual desire/drive (libido) and activity – Decreased spontaneous erections – Erectile dysfunction – Gynecomastia – Eunuchoidal body proportions – Height loss, low-trauma fracture, low bone mineral density – Hot flushes or sweats	– Decreased energy, motivation, initiative, self-confidence – Feeling sad or low, depressed mood – Poor concentration and memory – Sleep disturbance, increased sleepiness – Mild unexplained anemia (normochromic, normocytic) – Reduced muscle mass and strength – Central obesity

From Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force: Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536–2559.

features may include lack of drive or confidence, low moods or dysthymia, low energy or vitality, poor concentration, sleep disturbance, increased (central) body fat, and reduced muscle mass and strength with or without physical limitations. The nonspecific and variable clinical picture of postpubertal hypogonadism overlaps with that commonly encountered in older men associated with obesity and comorbidities, rendering a clinical diagnosis of androgen deficiency in this population particularly challenging (see Chapter 107) (Table 106.3).

Clinical evaluation of a patient suspected of hypogonadism should include a targeted history and focused physical examination aiming to address the following questions:

- What symptoms and signs are compatible with androgen deficiency? If so, can they be differentiated from any coexisting nonspecific symptoms?
- When was the likely time of onset (fetal, prepubertal, or postpubertal), duration/severity of putative hypogonadism?
- What is the likely etiology of hypogonadism? Is it reversible?
- What may be the likely benefits and risks of hormone treatment if hypogonadism is confirmed?

History

The history should include enquiries into:

- Neonatal or childhood development (hypospadias, cryptorchidism, growth or learning problems), pubertal onset milestones (appearance of pubic hair, growth spurt, nocturnal emission, and shaving), history of testicular insults (toxins, chemotherapy, trauma/torsion, surgery, or infection [e.g., mumps orchitis]).
- Sexual, reproductive, and androgenic functions including reduced sexual interest (libido), decreased spontaneous morning erections, decline in frequency of sexual intercourse or masturbation, erectile dysfunction, marital or other relationship status, previous paternity, hot flushes, loss of facial/body/pubic hair, decreased shaving frequency, breast enlargement or tenderness, height loss, history of low-impact fractures or osteoporosis, and decreased muscle mass and muscle strength affecting physical performance or activity.
- General health including chronic respiratory disease, major organ failure of the heart, liver, or kidney, hemoglobinopathies/blood transfusions, haemochromatosis, diabetes mellitus, ischemic heart disease, thromboembolic disease, headaches, and visual disturbance.
- Past and current medications, (opioid analgesics, glucocorticoids, psychotropics, chemotherapy, radiotherapy, 5α-reductase inhibitors, GnRH analogs, antiandrogens, and anticoagulants).
- Recreational drugs (anabolic steroids, cannabis, opiates).
- Smoking and alcohol intake.
- History of eating disorders.
- Excessive exercise.
- History of sexually transmitted infections, e.g., chlamydia, gonorrhea, syphilis.
- History of excessive snoring, which may suggest obstructive sleep apnea.
- Lower urinary tract symptoms (LUTS).
- Past history of prostatic diseases—benign hyperplasia or cancer.
- Mental state—anxiety, depression, irritability, poor concentration, low vitality/energy, low self-esteem.
- Family history of delayed puberty, anosmia, infertility, haemochromatosis, thromboembolic diseases, androgenic alopecia, and prostate cancer.

A history of obstructive sleep apnea, prostatic diseases, and cardiovascular and thromboembolic diseases are important in assessing the potential risks of TRT.

Examination

Physical examination of a patient suspected of hypogonadism should include assessment of:

- Body habitus (height, arm span, kyphosis), body weight (body mass index) and body fat distribution, muscle development, acne, and sebum production.
- Secondary sexual hair distribution in the face, chest, limbs, axillae, pubis, and male escutcheon.
- External genitalia—penile size/length, position of external urethral meatus.
- Testes location, size, contour, and consistency.
- Prostate by digital rectal examination (in adults over 40 years of age).[34]
- Breast examination.
- Sense of smell.
- Visual fields, iris, and optic fundoscopy (if indicated).
- Mental state and demeanor.

A eunuchoidal body habitus, defined as an arm span exceeding height by greater than 5 cm, is suggestive of prepubertal onset hypogonadism because of delayed long bone epiphyseal closure. Patients with KS may have disproportionally long legs, with lower body segment (symphysis to floor distance) exceeding the upper body segment (crown to symphysis distance). The presence of excessive acne over the shoulders and upper back and excessive sebum production in the forehead should alert the possibility of anabolic steroid abuse, especially in a man with well-developed musculature (presenting with infertility).

Testes position, size, contour, and consistency should be assessed by careful palpation, using a Prader orchidometer or calipers for documenting testicular volume. This should be followed by palpation to locate other scrotal contents: epididymides, vasa deferentia, hydrocoele, varicocoele, and cough impulse.

The size of the testis provides a simple and convenient clinical measurement of the seminiferous tubular mass (which constitutes 98% of the total testis volume). Normal adult testis volume is between 15 and 30 mL (3.5–5.5 cm in length and 2–3 cm in width). Testis volume may be slightly reduced in secondary hypogonadism but is usually significantly reduced in primary hypogonadism and infertility, being the hallmark of KS, where the testes are characteristically small (<3 mL) and firm in consistency. Testis volume measurement is useful in assessing pubertal development, its onset being indicated by testis volumes of greater than or equal to 4 mL. Testis volume remaining at less than 4 mL in an adult patient is therefore suggestive of prepubertal-onset hypogonadotropic hypogonadism. Testis volume of greater than 6 mL in hypogonadotropic patients indicates some degree of seminiferous tubule maturation has taken place and is associated with a favorable fertility outcome in response to gonadotrophin treatment in adulthood.[35]

Examination of the genitalia and scrotum is best performed with the patient initially in a supine position, followed by inspection and palpation for the presence of a varicocele (resting and with Valsalva maneuver) or a cough impulse in the spermatic cord in the upright position.

Digital rectal examination of the prostate may reveal an enlarged smooth gland, induration, or the presence of nodule Detection of undiagnosed prostate pathologies is important in assessing potential risks before starting TRT in middle-aged or older (>40 years) patients.

Breast examination, with the patient supine, should use the thumb and index finger to define the margin of any firm breast tissue from the surrounding soft fatty tissue, allowing the diameter of gynecomastia to be assessed with calipers.

The presence of anosmia or hyposmia (usually tested in the clinic with prepared odorants or more formal olfactometry using the commercially available University of Pennsylvania Smell Identification Test), bimanual synkinesis, cleft lip and/or palate, and hypodontia (agenesis of one or more teeth) point to the diagnosis of KallS. Additional phenotypic features such as severe obesity, short stature, hearing loss, polydactyly, syndactyly, visual impairment, external ear abnormalities, and learning difficulties are encountered in complex syndromes such as PWS, Lawrence–Moon/Bardet–Biedl syndrome, CHARGE syndrome, and septooptic dysplasia.

Examination of visual fields by confrontation is important in the context of expanding pituitary or hypothalamic tumors, and visual acuity and fundoscopy for retinal or optic nerve abnormalities in secondary hypogonadism.

LABORATORY DIAGNOSIS

Testosterone

In men presenting with symptoms and signs suggestive of androgen deficiency, the diagnosis of hypogonadism must be confirmed by documentation of unequivocally and consistently low levels of serum (total and/or free) T (Key Points: Diagnosis).

Blood sampling for measurement of T needs to take into account some physiological variations in serum T concentrations in order to minimize preanalytical noise and maximize the reliability in confirming T deficiency. Serum T concentrations in an individual vary from minute to minute,[36] day to day,[37] and also diurnally,[38] being highest in the early morning (7:00 a.m.–8:00 a.m.) and lowest (up to 25% lower than morning values in young men, but less in older men) in the late afternoon (4:00 p.m.–5:00 p.m.), when serum T can occasionally fall well below the normal range in men with intact reproductive function.[36] It follows that a single low serum T result may not by itself be abnormal and is insufficient to confirm the diagnosis of hypogonadism. It is imperative therefore to confirm an initial low T with at least one, if not two, further measurements. Serum T is transiently suppressed (by 25%–30%) after food intake,[39] so that measurements should be performed after an overnight fast. Acute or critical illnesses, major trauma, and surgery reduce T and gonadotropin production due to temporary suppression of the HPT axis.[40] It is therefore advisable to avoid measuring T in acutely ill patients or in those recovering from acute illnesses in the previous 3 to 6 months. It is also important to be aware of other causes of transient low T, e.g., medications, particularly opiate analgesics and glucocorticoids, malnutrition (eating disorders), and excessive strenuous exercise. In summary, the biochemical confirmation of clinically suspected hypogonadism should be based on measurements of T on fasting morning (before 11:00 a.m.) blood samples obtained on at least two different days preferably four weeks apart.

Testosterone Measurements. The accurate measurement of T (total) with a reliable and well-validated assay is a prerequisite in establishing the diagnosis of hypogonadism. Most clinicians rely on automated nonextraction immunoassays performed in hospital clinical chemistry laboratories using multianalyzer platforms. Though economic, convenient, and hence prevalent, these methods show considerable analytical variabilities between different assays and laboratories[41] due to suboptimal measurement precision, compounded by inappropriate calibration and the use of peer comparison rather than accuracy-based external quality control. Although these automated platform assays can perform reasonably well within the male physiological range,[42] the accuracy of T measurements progressively decreases at lower concentrations, rendering results in the hypogonadal and borderline-low range less reliable.

Mass spectrometry-based methods, which include serum extraction and a chromatography step (gas chromatography–mass spectrometry or, more commonly, liquid chromatography-tandem mass spectrometry [LC-MS/MS]), are considered the state-of-the-art methods for the measurement of T. They offer improved sensitivity and accuracy over a wide range, though incurring higher costs and decreasing sample throughput,[43] but are nevertheless not immune from interlaboratory variation and standardization issues. Currently, the availability of mass spectrometry methods for T measurement is still patchy in most hospital and community practices, where immunoassays continue to be employed. Whether immunoassay or LC-MS/MS is used, it is important for individual laboratories to optimize and standardize T measurements through stringent external quality assurance (EQA). Clinicians should only rely on T results provided by laboratories certified or verified by accuracy-based external quality control.

Total and Free Testosterone. T secreted from the testes is transported in the circulation bound to plasma proteins to prevent rapid diffusion across lipid-rich capillary cell membranes into tissues. Total T in the circulation is the sum of fractions bound mainly to SHBG (40%–60%) with high affinity (1.6×10^{-9} mol/L), to albumin (40%–50%) with lower affinity (1.0×10^{-4} mol/L), and minor fractions to orosomucoid and cortisol-binding protein, with only a small proportion (1%–3%) being unbound as the free hormone.[44] According to the free hormone hypothesis, only this small unbound fraction can freely access target tissues to induce androgenic biological actions.[45] T loosely bound to albumin can dissociate in slower-flowing circulations and potentially access some target tissues, e.g., liver, brain. The combined free and albumin-bound fractions of T are referred to as "bioavailable T."

The synthesis and secretion of SHBG in the liver is regulated by many hormonal, metabolic, and nutritional factors. Consequently, SHBG concentrations are frequently altered in commonly encountered physiologic and pathological conditions (Table 106.4). Serum total T measurements are affected by changes in SHBG levels without altering the free or intracellular (biologically active) T concentrations. Therefore, the biochemical confirmation of hypogonadism by total T alone can be confounded when SHBG levels are decreased or increased, leading respectively to spurious over- or underdiagnosis. Conditions that decrease SHBG (e.g., obesity, type 2 diabetes) can lower total T concentrations in eugonadal men into the subnormal range while free T remains normal, leading to overdiagnosis of hypogonadism. The European Male Aging Study found that obese men with low total T (and low SHBG) in the presence of normal free T did not show evidence of androgen deficiency, but low free T, even in the presence of normal levels of total T in aging men (with high SHBG), was associated with sexual symptoms and other evidence compatible with hypogonadism.[46] Conditions that increase SHBG (e.g., advanced age, HIV infection, thyrotoxicosis, cirrhosis) can maintain or raise total T concentration in the normal or high-normal range even though free T is low, and the diagnosis of hypogonadism is potentially overlooked.[46] It follows that free (or bioavailable) T should be measured in conditions that alter SHBG concentrations where total T is a less reliable indicator of gonadal status (Fig. 106.3). For similar reasons,

TABLE 106.4 Conditions that may Alter SHBG Concentrations

Increase	Decrease
– Ageing	– Obesity
– Thyrotoxicosis	– Diabetes
– Liver disease	– Hypothyroidism
– HIV infection	– Acromegaly, GH treatment
– Anticonvulsants	– Androgens, progestins, glucocorticoids, tyrosine kinase inhibitors
– Estrogens, SERMs, tamoxifen, cyproterone	– Nephrotic syndrome
– Malnutrition	– Polymorphisms in *SHBG* gene
– Polymorphisms in *SHBG* gene	

HIV, Human immunodeficiency virus; *SERMs*, selective estrogen receptor modulators; *GH*, growth hormone; *SHBG*, sex hormone–binding globulin.
(From Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force: Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536–2559.)

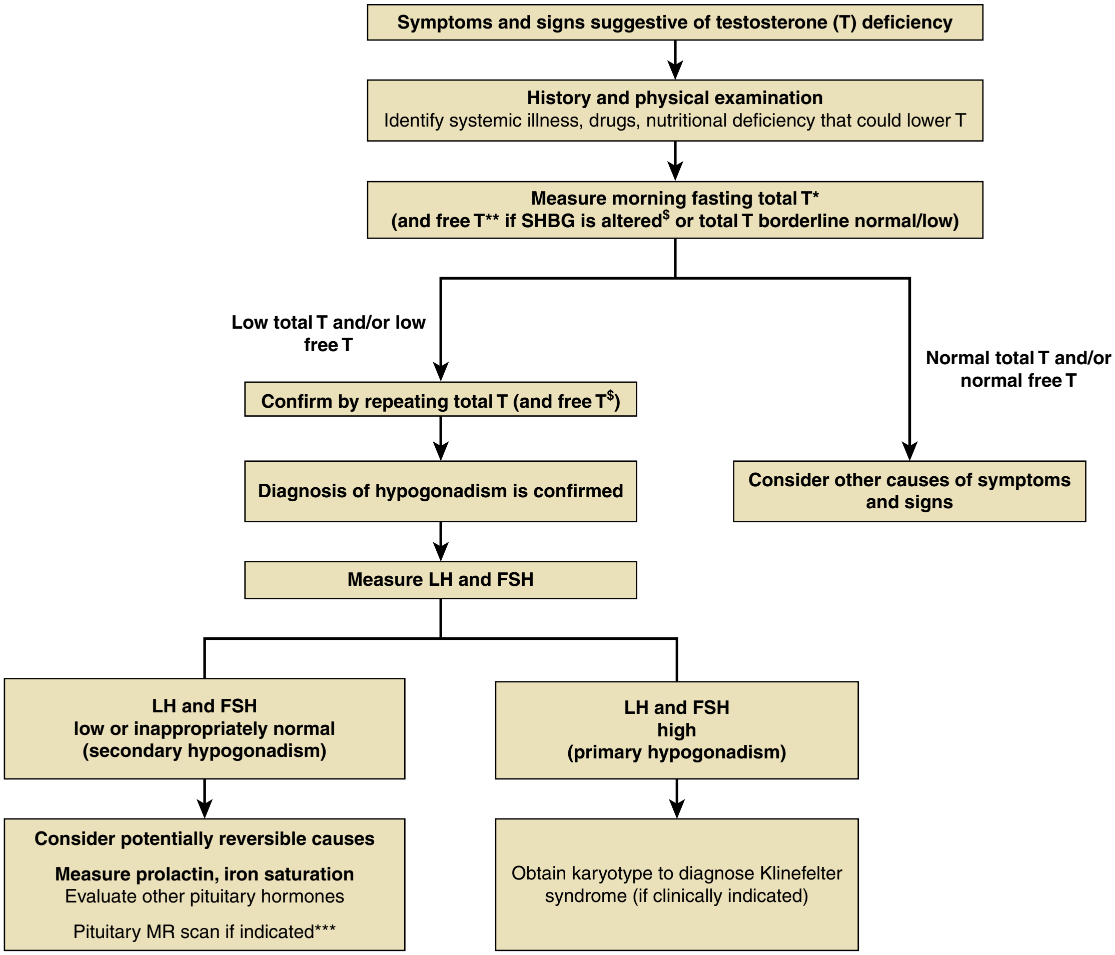

Symptoms and signs suggestive of testosterone (T) deficiency
History and physical examination
Identify systemic illness, drugs, nutritional deficiency that could lower T
Measure morning fasting total T*
(and free T** if SHBG is altered$ or total T borderline normal/low)
Low total T and/or low free T
Normal total T and/or normal free T
Confirm by repeating total T (and free T$)
Diagnosis of hypogonadism is confirmed
Consider other causes of symptoms and signs
Measure LH and FSH
LH and FSH low or inappropriately normal (secondary hypogonadism)
LH and FSH high (primary hypogonadism)
Consider potentially reversible causes
Measure prolactin, iron saturation
Evaluate other pituitary hormones
Pituitary MR scan if indicated***
Obtain karyotype to diagnose Klinefelter syndrome (if clinically indicated)

Figure 106.3 Algorithm for the diagnostic evaluation of adult men suspected of hypogonadism. *The 2.5th percentile of reference range for total T harmonized to CDC standard in healthy nonobese young men is 264 ng/dL (9.2 nmol/L). This limit could be used for total T assays that are CDC-certified. For laboratories that are not CDC-certified and do not participate in an accuracy-based quality control program, the reference range may vary considerably depending on the assay and reference population used. Using the lower limit of the range established in local laboratories may not always accurately identify men with hypogonadism. **A harmonized reference range for free T has not been established, so reference ranges may vary considerably depending on the specific equilibrium dialysis method or the algorithm used to calculate free T. Therefore, until a harmonized reference range is available, the lower limits proferred by local laboratories may be used. $Conditions in which measurement of free T is recommended, where SHBG levels may be altered, are listed in Table 106.5. ***Perform pituitary imaging (MRI) to exclude pituitary and/or hypothalamic tumor or infiltrative disease in severe secondary hypogonadism (serum T <150 ng/dL [5.2 nmol/L]), panhypopituitarism, persistent hyperprolactinemia, or symptoms or signs of tumor mass effect (such as new-onset headache, visual impairment, or visual field defect) are present. *CDC*, Centers for Disease Control; *FSH*, follicle-stimulating hormone; *LH*, luteinizing hormone; *MRI*, magnetic resonance imaging; *SHBG*, sex hormone–binding globulin; *T*, testosterone. (Figure adapted from Bhasin S, Brito JP, Cunningham GR, et al. Testosterone therapy in men with hypogonadism: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2018;103:1715–1744.)

free T measurement is also informative when total T is at the borderline of the lower limit of normal (see "Reference range and thresholds" later and Fig. 106.3)[47] (Key Points: Diagnosis). Nevertheless, it is important to emphasize that total T remains a robust and well-tried measure of testicular function in most situations, especially in classical hypogonadism. Current guidelines unanimously recommend total T as the first line investigation for the diagnosis of suspected hypogonadism[1,2,48] (Fig. 106.3).

Measurement/Calculation of Free Testosterone. Free T concentrations should ideally be measured directly after serum dialysis or ultrafiltration to separate the bound from the free hormone. These state-of-the-art methods are technically demanding and expensive, and availability is limited to only a few reference and commercial laboratories. In routine practice, clinicians usually rely on calculated free T derived from total T, SHBG, and albumin concentrations using several algorithms[44,49] mostly based on affinity constants of T to the binding proteins. The accuracies of different algorithms vary when validated against directly measured free T concentrations. The most widely used is the Vermeulen law of mass action calculations,[50] which correlates well with measured free T and performs well in clinical scenarios across a wide range of SHBG concentrations in men, even though the results have a systematic positive bias.[49]

Bioavailable T is measured by ammonium sulfate precipitation of SHBG-bound T, with the addition of tritium-labeled T tracer to indicate the percentage of non–SHBG-bound T. This is a relatively cumbersome method involving radioactive reagents. Although bioavailable T can also be calculated from total T, SHBG, and albumin,[50] fewer clinical data are available to support its application, compared with free T, in the diagnosis of hypogonadism.

Free T tracer analog immunoassays, though widely available in commercial laboratories, are imprecise, and results are influenced by changes in SHBG; their use is not recommended.[41] Free androgen index (T/SHBG × 100), commonly used in women for the assessment of hirsutism, is not a valid surrogate for free T in men because of the relatively high levels of T compared with SHBG. Furthermore, the feedback regulation of T in the HPT axis in response to changes in SHBG produces a nonlinear relationship between T and SHBG, which is not accurately described by the ratio.

Reference Ranges and Thresholds. Reference ranges for total and free T vary considerably between laboratories and localities due to differences in assay methodologies, calibrators, and the constituents of the reference population. Clinicians have to rely on reference ranges provided by their local laboratories, some of which are of uncertain provenance. To harmonize reference ranges across different geographic regions, local results from four large cohorts of community-dwelling men in the United States and Europe were recently cross-calibrated against the reference mass spectrometry method of the Centers for Diseases Control (CDC). The harmonized reference range for total T in young (aged 19–39 years) healthy nonobese men was 264 to 916 ng/dL (9.2–31.8 nmol/L) representing the 2.5th to 97.5th percentiles and 303–852 ng/dL (10.5–29.5 nmol/L) using the 5th to 95th percentiles. Individual laboratories and EQA schemes can now calibrate their total T results to the CDC reference method so that clinicians may make use of the internationally harmonized reference range.

Serum T concentration declines gradually with age, mostly within the physiological range, but with greater rates of decline in men with obesity and comorbidities.[51] Currently, there is insufficient evidence to support the need for an age-adjusted Z-score approach for the T reference range,[52] and most laboratories continue to use the T-score approach, with reference range based on healthy nonobese men under 40 years of age.[1,2]

T concentrations in the population have a Gaussian continuous distribution, so that a single cutoff value below which hypogonadism pertains is unlikely to exist. Furthermore, there is evidence to suggest that different T threshold levels apply to individual androgen deficiency symptoms or functions.[53] While there is no consensus on what is considered a subnormal T level, many studies have demonstrated that clinical signs of androgen deficiency increase when T levels fall below 260 to 317 ng/dL (9–11 nmol/L).[54-57] This also corresponds to the 2.5th to 5th centile of the CDC-harmonized population reference range (see earlier). Therefore, a threshold "range" or "grey zone" around 260 to 317 ng/dL (9–11 nmol/L), rather than a discreet level of T, would more accurately reflect the available evidence. In clinical practice, there is a prevalent view that men with T concentrations below 230 ng/dL (8 nmol/L) are likely to be clinically hypogonadal, whereas men with T above 346 ng/dL (12 nmol/L) are probably eugonadal, while those with T between 230 and 346 ng/dL (8–12 nmol/L) can be considered to have borderline values when free T measurements can be helpful (see earlier).[2,47]

Given the lack of standardization in free T assays and uncertainties surrounding different algorithms for calculating free T, the reference range for free T is less well established than that for total T. A level around 6.5 ng/dL (225 pmol/L), ranging from 5.0 to 7.0 ng/dL (170–240 pmol/L), has been suggested as the lower limit of normal free T.[41,55,57] Until a harmonized and validated reference range for measured or calculated free T is available, individual laboratories will need to establish their own locally.

Other Hormone Measurements. Circulating T is metabolized to a more potent androgenic metabolite, 5α-dihydrotestosterone (DHT), in target tissues such as prostate, liver, and hair follicles by 5α-reductase 1, 2, and 3 (SRD5A1, SRD5A2, and SRD5A3).[58] DHT in the circulation is almost solely derived from the peripheral conversion of secreted T in target tissues. Measuring blood levels of DHT is therefore not useful in the assessment of testicular function in the diagnosis of hypogonadism and is only performed in rare suspected cases of 5α-reductase deficiency (see Chapters 101 and 102).

Estradiol (E2) plays an important part in male skeletal physiology, fat deposition, and sexual function.[59] Less than 15% of circulating E2 in men is secreted directly by the testes; the remaining 85% is converted from androgen precursors (predominantly T) by the aromatase (CYP19A1) enzyme in tissues such as fat, brain, testis, and bone. Measurement of serum E2 is generally uninformative in the diagnosis of hypogonadism, but it is indicated in the investigation of gynecomastia (see Chapter 108) and in rare cases of aromatase deficiency (see Chapters 101 and 102). Serum E2 levels in men are low (around 10–53 pg/mL [38–196 pmol/L]), and immunoassays usually lack the sensitivity of more reliable LC-MS/MS-based methods.

Further Investigations to Identify Etiologies of Hypogonadism. Gonadotropins should be measured in men with confirmed hypogonadism to distinguish primary from secondary hypogonadism, which then entrains further investigations to identify underlying etiologies (Fig. 106.3) (Key Points: Diagnosis).

Gonadotropins are now routinely measured by modern immunoassays, which have sufficient sensitivity to distinguish between normal and low basal levels of LH and FSH in adult patients without the need for dynamic GnRH stimulation test. The latter test is poorly characterized and cannot consistently distinguish IHH from self-limited delayed puberty (see Chapter 103) or hypothalamic from pituitary pathologies. Despite the pulsatile secretion of gonadotropins, there is generally no difficulty in identifying unequivocally elevated levels encountered in primary hypogonadism; thus, a single measurement for LH and FSH usually suffices.

In men with primary hypogonadism, chromosome karyotyping may be needed, especially in the presence of suggestive clinical features such as small (<6 mL) testes, gynecomastia, and azoospermia, to diagnose possible KS (47,XXY in 90%) (Fig. 106.3). Karyotype analysis is carried out in cultured peripheral blood lymphocytes. To confirm suspected KS mosaicism (47,XXY/46,XY), karyotype analysis is performed on cultured skin fibroblasts or testicular tissue. Selective elevation of FSH (with normal LH and T) in infertile men with oligozoospermia or azoospermia suggests severe primary spermatogenic failure, commonly due to Y chromosome microdeletion (see Chapters 112 and 115), which is not amenable to hormonal/medical management.

In secondary hypogonadism, further investigations to exclude possible hypothalamus or pituitary pathologies such as pituitary tumors, KallS, IHH, hyperprolactinemia, and iron overload are indicated (Fig. 106.3), especially in younger patients. These should include measurement of serum prolactin, iron saturation, and/or serum ferritin, assessment of other anterior pituitary hormone excess (e.g., Cushing syndrome acromegaly) or deficiency (hypopituitarism, diabetes insipidus), and magnetic resonance imaging (MRI) of the pituitary and hypothalamus. The latter is required in the presence of clinical evidence of pituitary mass effects such as visual disturbances, recent-onset headache, hydrocephalus, or significant hyperprolactinemia

(>100 ng/mL or 2000 mIU/L) suggestive of a pituitary prolactinoma (see Chapter 4).

In the evaluation of presumed secondary hypogonadism in middle-aged and older men with low T, where these red-flag findings of pituitary pathologies are usually absent, the cost-effectiveness of MRI of the pituitary has been questioned.[60] The reported prevalence of major hypothalamic-pituitary lesions that require intervention remain very low until T concentrations fall to less than 5.2 nmol/L (<150 ng/dL)—this has been proposed as the action threshold for MRI to improve diagnostic yield in the investigation of functional secondary hypogonadism.

In patients with IHH, the advent of next-generation sequencing has improved our understanding of monogenic, digenic, and oligogenic etiologies of the heterogeneous disorders associated with GnRH deficiency.[61] This has raised the prospects of molecular genetics diagnosis of mutations affecting GnRH regulation, the GnRH receptor, or the *LH* gene, guided by family history (sporadic, X-linked, autosomal dominant/recessive) and the presence of somatic phenotypic syndromic features such as anosmia, hearing loss, bimanual synkinesis, cleft lip and/or palate, polydactyly, hypodontia (agenesis of one or more teeth), and unilateral renal agenesis (Box 106.1). Making a genetic diagnosis of IHH (and differentiating from self-limited delayed puberty) may confer advantages in targeted treatment to minimize virilization deficits/delays, optimize fertility, and promote personalized approaches to genetic counseling (see Chapter 103).

Hypogonadism is associated with osteopenia or osteoporosis and increased risk of low-trauma or fragility fracture.[59] All patients with confirmed hypogonadism should undergo DXA scanning to measure BMD in both spine and hip and calculate fracture risk with fracture risk calculators such as FRAX to inform management and follow-up.[62] However, the cost-effectiveness of BMD measurement for optimizing bone health outcomes, especially in fracture and fall prevention, in hypogonadal men has not been formally assessed.

TREATMENT OF HYPOGONADISM

TESTOSTERONE REPLACEMENT THERAPY

Indications

TRT is used to treat androgen-deficient men with a diagnosis of primary and secondary hypogonadism. The aim of TRT is to improve hypogonadal symptoms, restore circulating T levels to physiological levels, induce or maintain secondary sexual characteristics, and reverse or prevent the consequences of T deficiency (Key Points: Treatment). Current guidelines recommend TRT only in men diagnosed to be hypogonadal on the basis of symptoms and signs of androgen deficiency, as well as repeatedly low T levels.[1,2]

KEY POINTS: Treatment

- Testosterone replacement therapy for confirmed organic hypogonadism is well-established and highly efficacious.
- Patient-important efficacy outcomes and long-term safety of testosterone treatment in men with functional hypogonadism have not yet been clearly demonstrated.
- Appropriate efficacy assessment and safety monitoring should be performed during maintenance of testosterone treatment.
- Gonadotropin induction of spermatogenesis with luteinizing hormone-like and follicle-stimulating hormone preparations allow the majority of isolated hypogonadotropic hypogonadal patients to achieve natural fertility or assisted conception.

Testosterone Preparations

TRT can be delivered by many preparations, with different pharmacokinetics and routes of administration (Table 106.5). Replacing relatively obsolete injectable or implantable T preparations, transdermal gels and long-acting depot injections are now the most frequently used. The choice of the therapeutic regimen depends on its pharmacokinetics, the patient's preference, local availability and cost, and treatment burden.[1] An overview of the recommended dose, interval of administration, advantages, and disadvantages is given in Table 106.5.

Endogenous T secreted by the testis into the circulation is converted to E2 by the aromatase enzyme and to the more potent androgen DHT by 5α-reductase enzymes. All three hormones are required to achieve the full spectrum of male sexual/reproductive and nonreproductive physiological functions. The TRT preparation of choice should therefore also be convertible to E2 and DHT, ideally mimicking endogenous T.

Parenteral Testosterone Preparations

Intramuscular Injectable Preparations. Injected natural T has a short circulatory half-life of approximately 10 minutes. To prolong the half-life and improve solubility in the vehicle oil, T esters, such as T enanthate and T cypionate, have been developed for intramuscular injection. Shortly after injection, serum T increases to supraphysiological levels, followed by a gradual decrease over 1 to 3 weeks. Large excursions in serum T concentrations, which can be associated with symptomatic fluctuations in some patients, are the main disadvantage. The risk of developing erythrocytosis is higher with intramuscular injections due to supraphysiological peak T concentrations.[63,64] There is no advantage in using mixtures that combine short- and longer-acting T esters in one injection.

Testosterone undecanoate (TU) is available as an ultra–long-acting T formulation in a castor oil base injected every 10 to 14 weeks. A rare adverse effect, occurring in 15 to 19/1000 injections, is the development of coughing and dyspnea shortly after injection, possibly due to pulmonary oil microembolization.[63-65] It is advisable to observe patients for 20 to 30 minutes after TU injections.

Subdermal Implantation of Testosterone Pellets. T pellets are implanted subcutaneously in the gluteal region or lower abdominal wall via a minor surgical procedure. Data documenting the efficacy of T pellets in improving hypogonadal symptoms are scarce,[66] but the infrequent administration and relatively stable T concentrations are appreciated by patients.

Subcutaneous Injections. Subcutaneous administration of T enanthate via an autoinjector is under development. Weekly injections of 75 mg yielded stable T concentrations.[67]

Transdermal Testosterone Preparations

Testosterone Patches. One patch contains 2 or 4 mg of T applied on the back, abdomen, upper arms, or thighs. Skin irritation and rash from the excipients are frequently reported, leading to treatment discontinuation.[63]

Testosterone Gels. Transdermal hydroalcoholic T gels are one of the most frequently used formulations to treat hypogonadism. The gel is applied every morning to the upper arms, shoulders, or abdomen. T is then absorbed to form a reservoir in the dermis, from which it is slowly released into the circulation. Transdermal T gels are easy to use and have few side effects. Skin irritation is rare compared with patches. Secondary exposure to women or children via accidental transfer of the gel has been reported, causing virilization. This risk can be minimized by instructing patients to wash their hands and avoid close skin contact with others after

TABLE 106.5 Overview of Available Testosterone Preparations

Route of Administration	Formulation	Starting Dose/Interval	Advantages	Disadvantages
Intramuscular injections	Testosterone enanthate	250 mg Every 2–3 weeks	Relatively inexpensive Flexibility of dosing	Requires IM injections. Excursions in T concentrations may be associated with fluctuating symptoms
	Testosterone cypionate	250 mg Every 2-3 weeks	Relatively inexpensive Flexibility of dosing	Requires IM injections Excursions in T concentrations may be associated with fluctuating symptoms
	Testosterone propionate (not available in United States)	50-100 mg Every 3–5 days		Frequent IM injections, not suited for chronic treatment T concentrations fluctuate
	Mix of testosterone esters (Sustanon)	One vial 250 mg Every 2–3 weeks	Relatively inexpensive Flexibility of dosing	Regular IM injections Excursions in T concentrations may be associated with fluctuating symptoms
	Testosterone undecanoate	1000 mg at week 0 and 6, and then every 12 weeks United States: 750 mg at week 0 and 4, and then every 10 weeks	Infrequent administration	Local discomfort after IM injection of large volume (3 or 4 mL) Coughing/dyspnea within 15–30 minutes of injection in small number of patients (15–19/1000 injections)
Subcutaneous implants	Implantable pellets (only available in United States and Australia)	150–450 mg every 3 months or 600–1200 mg every 6 months	Infrequent administration Relatively stable T concentrations	Requires minor surgical insertion Pellets may extrude Local hematoma or infection rarely
Subcutaneous injections	Testosterone enanthate	75 mg Once weekly	Ease of application	Increased blood pressure, injection site reactions, headache
Transdermal	T patch (only available in some countries)	2 or 4 mg per patch 4 mg every 24 hours	Ease of application	Skin irritation
	T gel	Available as 1%, 1.62%, 2% gel in unit-dose packets or via a metered-dose pump system Dose dependent on the formulation: 50 mg for 1% T gel; 81 mg for T 1.62%; 60 mg for 2% T gel Once daily	Ease of application Flexibility of dosing	Risk of transfer to female partner or child Variable T concentrations from day to day
	T solution (Axiron)	60 mg one to two times/day in the axillae		Risk of transfer to female partner or child Variable T concentrations from day to day
Oral	Testosterone undecanoate capsules (not available in United States)	40-80 mg/capsule Two to three times/day	Convenient administration Avoids first-pass metabolism	Short half-life, requires multiple daily dosing Ingestion with fatty meal Fluctuating T levels due to variable absorption
	Modified testosterone undecanoate	237 mg/capsule Two times/day	Ingestion with regular meal Avoids first-pass metabolism	Twice daily administration
	Testosterone buccal bioadhesive tablet	30 mg/tablet Two times/day	Avoids first-pass metabolism	Local gum irritation Alteration of taste No dose titration
Intranasal	Nasal gel	One pump of 5.5 mg per nostril three times/day	Rapid absorption Avoids first-pass metabolism	Local nasal side effects Multiple applications per day

applying the gel and by covering the gel application area with clothes. A major advantage is that treatment can be stopped quickly if side effects occur.[63,64]

Oral Testosterone Preparations. Orally ingested T is rapidly inactivated by the liver (first-pass metabolism). To avoid this first-pass effect, modified androgen preparations have been developed.

TU in castor oil capsules is absorbed via the gut lymphatic system, thus avoiding hepatic first-pass metabolism.[63] To increase absorption, the capsules should be ingested with a (fatty) meal twice or thrice daily.

A modified form of oral TU using a self-emulsifying drug delivery system has been designed to increase the solubilization of lipophilic testosterone. This twice-a-day oral TU formulation has recently been approved by the US Food and Drug Administration (FDA) for treatment of hypogonadism in the United States. In clinical trials, it produces mean serum T concentrations in the mid-normal range in 87% of patients, with a positive impact on hypogonadal symptoms.[68]

Transbuccal Bioadhesive Testosterone Preparations. A buccal tablet is applied to the gum above the upper incisors, providing a sustained release of T via the buccal mucosa into the circulation. It is effective in increasing serum T levels, but data on improvement of hypogonadal symptoms are lacking.[63,69]

Transnasal Testosterone Preparations. Via a metered-dose pump applicator, nasal T gel can be applied into the nostrils. With the recommended dose of 11 mg, 90% of patients reached T levels in the normal range. In open-labeled trials, erectile function, body composition, and BMD improved, but placebo-controlled data or trials comparing the nasal gel with more commonly used formulations are lacking.[70]

Androgens Other than Testosterone

Dehydroepiandrosterone. Dehydroepiandrosterone (DHEA) is a weak androgen and a precursor of T. A few placebo-controlled trials have shown no convincing evidence that DHEA improves hypogonadal symptoms, and it is not recommended as an alternative treatment in men with androgen deficiency.[71]

Dihydrotestosterone Gel and Mesterolone. A transdermal gel containing DHT (Androstanolone) or an oral preparation containing a synthetic androgen derived from DHT (Mesterolone) are available in some countries. These preparations can only exert androgenic effects, as they cannot be aromatized to E2. For this reason, they are not indicated to treat hypogonadal men.

Efficacy

It is generally accepted that restoration of circulating T concentrations in hypogonadal men to the midphysiological range will invariably ensure an efficacious and sustained clinical response (Key Points: Treatment). Indeed, this has been the longstanding strategy for regulatory approval of new T products, which requires only this pharmacokinetic target to be satisfied, without the need for demonstrating clinical efficacy outcomes. TRT is highly effective in inducing secondary sexual characteristics in younger patients with incomplete sexual development due to organic hypogonadism (Fig. 106.2). In adult men with confirmed hypogonadism, the most consistent clinical benefit of TRT is the improvement in libido, erectile function, sexual activity, and sexual satisfaction.[72] Lean mass (and muscle strength in most studies) increases and fat mass decreases consistently in response to TRT. However, the effects of TRT on mood and well-being are less predictable.

In older men with low T, there is no convincing evidence that TRT improves depressive symptoms or fatigue, nor does it have a positive effect on cognitive function or memory.[73] Despite the increase in lean mass in response to TRT in older men with low T, placebo-controlled trials showed no or only minimal improvements in walking distance or gait speed.[73] There is some evidence that TRT can improve insulin sensitivity in hypogonadal men with type 2 diabetes or the metabolic syndrome, but any benefit on glycemic control is unconvincing.[74] TRT also has a positive effect on vertebral and femoral BMD. However, there are no data indicating that TRT reduces the risk of osteoporotic fractures, and TRT on its own is not approved for treating osteoporosis. In hypogonadal men with a high fracture risk, TRT alone is not sufficient, and a bone-specific osteoporosis treatment should be added.

Contraindications and Adverse Events

Minor adverse events related to TRT include oily skin, acne, and breast tenderness. Formulation-specific adverse effects are described in Table 106.5. Contraindications for TRT are summarized in Box 106.2.

Exogenous T suppresses spermatogenesis, so TRT is contraindicated in men planning fertility in the near future. Alternative treatment in men with secondary hypogonadism seeking fertility is discussed below.

BOX 106.2 Contraindications for Testosterone Replacement Therapy

Desire for fertility in the near future
Untreated and treated breast cancer
Untreated, locally advanced or metastatic prostate cancer
Unevaluated palpable prostate nodule
Unevaluated PSA >4 ng/mL (>3 ng/mL in men with high risk for prostate cancer, e.g., Black Americans or men with affected first-degree relatives)
Severe lower urinary tract symptoms (IPSS >19)
Hematocrit >48%–50%
Thrombophilia
Severe congestive heart failure
Recent myocardial infarction or stroke
Untreated sleep apnea

PSA, Prostate-specific antigen; *IPSS*, international prostate symptom score.

TRT is contraindicated in men with breast cancer and men with untreated, locally advanced or metastatic prostate cancer, both being sex steroid-responsive malignancies. Prostate-specific antigen (PSA) levels and prostate volume can increase within the physiological range shortly after TRT is started in hypogonadal men. There is currently no evidence that TRT increases the risk of prostate cancer, but the data are based on short-term studies with duration up to 36 months only.[75] Because T can promote growth of an occult preexisting prostate tumor, current guidelines recommend screening for undiagnosed prostate cancer prior to starting TRT in men over 40 years of age via digital rectal examination and measurement of PSA. If PSA is greater than 4 ng/mL (>3 ng/mL in men with high risk for prostate cancer, e.g., Black Americans or men with affected first-degree relative) or a palpable prostate nodule is present, further evaluation by a urologist is required. It is also recommended to monitor PSA and prostate volume 3 and 12 months after TRT is started. Afterwards, the local guidelines for prostate cancer screening for the general population should be followed.

In men with a history of low-grade prostate cancer who have been treated with a radical prostatectomy and have undetectable PSA levels for at least 2 years, TRT may be considered in those with low T levels and severe hypogonadal symptoms. However, initiating TRT in men with a history of prostate cancer remains controversial, and there is currently insufficient evidence to support a generally accepted recommendation.[1,76] The risks and benefits of TRT should be discussed with the patient, and PSA levels should be monitored regularly if TRT is started in this group.

Severe LUTS (International Prostate Symptom Score [IPSS] ≥19) associated with benign prostatic hyperplasia is a relative contraindication, but modest LUTS (IPSS <19) is not, because there is no deterioration from TRT.[1,2]

T stimulates erythropoiesis, and TRT is contraindicated in men with a high hematocrit (>48%–50%). Erythrocytosis (hematocrit >54%) is a frequent adverse event in men using TRT. It occurs more frequently in older patients and with intramuscular injections, and is dose-dependent. TRT should be discontinued until hematocrit is normal again, in some cases after venesection.

TRT can increase the risk of venous thromboembolism in men with thrombophilia (often undiagnosed). Personal and family history of thromboembolic events and risk factors should be sought before starting TRT. The frequency of neuroocclusive events in men who developed TRT-induced erythrocytosis is however low, and the number of venous thromboembolic events in randomized controlled trials (RCTs) was too low to draw conclusions about the risk.[1,72]

Untreated sleep apnea, severe heart failure, recent myocardial infarction, and stroke are also considered contraindications for TRT. There has been concern that TRT increases cardiovascular disease risk. This is mainly based on data from retrospective studies, in which there was no information available about hypogonadal symptoms, the underlying condition, and T levels. Available RCTs on TRT with duration of only 6 to 36 months have not been designed or were not sufficiently powered to detect a difference in major adverse cardiovascular events. However, most metaanalyses show no association between TRT and cardiovascular events in men with an established diagnosis of hypogonadism.[75] Following the directive of the FDA, an industry-sponsored multicenter RCT has recently been launched to investigate the cardiovascular disease risk of TRT in older men in the United States.

Monitoring of Treatment

After TRT is started, improvement of hypogonadal signs and symptoms is expected within 3 to 6 months, which should be reviewed clinically. It is generally recommended to measure T 3 months after TRT is started and to aim for serum T concentrations in the mid-normal range for healthy young men.[1,2] The timing to measure serum T levels depends on the formulation used. For intramuscular injections of testosterone cypionate or enanthate, the interval between injections can be optimized by measuring serum T just before the next injection, which should be at the lower limit of the reference range.[77] When using a transdermal T gel, serum T levels should be measured 4 to 12 hours after application. After 2 to 4 weeks of daily usage, the dose can be adjusted to obtain T concentrations in the normal range. Patients maintained on long-term TRT should be clinically evaluated at least annually, including review of therapeutic goals and measuring serum T and hematocrit (Key Points: Treatment). Patients receiving the newest oral formulations of testosterone undecanoate should have T sampled 4 to 7 hours after taking the morning capsule(s).

GONADOTROPIN INDUCTION OF SPERMATOGENESIS AND FERTILITY

Besides being candidates for TRT, gonadotropin-deficient patients with secondary hypogonadism can benefit from treatments that, by activating or reactivating the dormant testis, stimulate endogenous T secretion as well as spermatogenesis. This option is not feasible in primary hypogonadal patients, where the pathological testes will not respond to exogenous gonadotropin stimulation (Key Points: Treatment).

Indications

Gonadotropin treatment may be considered more physiological than TRT in secondary hypogonadism, but it is more burdensome, having to be administered by subcutaneous or intramuscular injections two or three times a week. In addition, it is more expensive than TRT. For these reasons, although potentially suitable for all patient with secondary hypogonadism, its use is generally limited to patients who require fertility.

Gonadotropin Treatment

Preparations with Luteinizing Hormone-Like Action. LH-like preparations can stimulate Leydig cells (Table 106.6). Recombinant LH is expensive, has a very short half-life, and has not been extensively tested or approved for the use in men. hCG binds the LH receptor with the same affinity as native LH but has a much longer half-life. hCG extracted from urine of pregnant women is the most frequently used treatment for Leydig cell stimulation, being administered subcutaneously or intramuscularly two or three times a week.

TABLE 106.6 Gonadotropin and GnRH treatment for Hypogonadotropic Hypogonadism to induce fertility in adulthood

	Dose	Regimen
LH-like preparations		
rhLH	No data available	• Start LH-like preparations alone • Start LH-like and FSH-like preparations together
hCG	1000–2000 IU Two to three times weekly	
rhCG	No data available	
FSH-like preparations		
hMG	75–225 IU Two to three times weekly	• Add FSH-like preparations after 6 months of LH-like therapy with sperm concentration <10 million/mL • Increase FSH-like therapy (in combination with LH-like preparations) after 6 months of lower dosage with sperm concentration <20 million/mL • Start FSH-like preparations together with LH-like preparations and increase FSH after 6 months of lower dosage with sperm concentration <20 million/mL
hpFSH	75–225 IU Two to three times weekly	
rhFSH	75–225 IU Two to three times weekly	
Corifollitropin alfa	150 μg every other week	
GnRH		
Native GnRH	100–400 ng/kg every 90–120 min	GnRH only until outcome is achieved

GnRH, Gonadotropin-releasing hormone; *hMG*, human menotropin; *hpFSH*, highly purified follicle stimulating hormone; *rhFSH*, recombinant human follicle-stimulating hormone; *rhLH*, recombinant human luteinizing hormone; *hCG*, human chorionic gonadotropin; *rhCG*, recombinant human chorionic gonadotropin.

Preparations with Follicle-Stimulating Hormone-Like Action. FSH action can be replicated by human menopausal gonadotropin (hMG) and highly purified FSH (hpFSH), or by synthetic recombinant human FSH (rhFSH) (Table 106.6). hMG is a mixture of gonadotropins extracted from human urine with both FSH- and LH-like action, while predominant or exclusive FSH stimulation is provided by hpFSH and rhFSH, respectively. These selective preparations are more expensive; they are specifically designed for ovulation induction, which requires a precise dose/sequence of FSH and LH stimulation. This is not the case for males, and at present there is no evidence to indicate superiority of hpFSH and rhFSH over hMG in the induction of spermatogenesis.[35] FSH preparations are administered by subcutaneous or intramuscular injections two to three times weekly. To overcome the multiple injection issue, recently, corifollitropin alfa has been evaluated for the use in men (Table 106.6). Corifollitropin alfa is a recombinant gonadotropin, made of the α-subunit of human FSH and a hybrid subunit composed of the β-subunit of human FSH and the C-terminal of the β-subunit of hCG, with similar bioactivity to rhFSH. Corifollitropin alfa has longer circulating half-life and allows administration every other week, a characteristic that may improve compliance in prolonged treatment required in the induction of spermatogenesis.

Treatment Regimens. Adult patients with secondary hypogonadism can start gonadotropins when they are seeking fertility. There is no consensus on the best dose and regimen for stimulation of spermatogenesis in adult men (Table 106.6). With a view to cost-effectiveness, it is generally advised to start with hCG 1000 to 2000 IU two or three times weekly. If sperm concentration remains below 10 million/mL and conception is not achieved within 6 months, FSH/hMG is added with a starting dose of 75 IU alternate days or three times weekly.[78] After a further 6 months of treatment, if conception does not occur and sperm concentration is below 20 million/mL, FSH/hMG may be titrated up to 150 IU.[78] It should be noted that, if costs are not a concern, combined hCG and FSH/hMG treatment may be started from the beginning.[78]

It should be emphasized that promotion of spermatogenesis requires relatively prolonged treatment, because one complete cycle of germ cell differentiation takes 3 months (see Chapter 111). Usually, at least 6 months (two cycles of spermatogenesis) of therapy are required to produce clinically significant results.

Outcomes

Semen Parameters. A meta-analysis of studies considering semen outcomes in adult azoospermic secondary hypogonadal men treated with gonadotropins has shown that sperm in the ejaculate is obtained on average in 75% of men, with a sperm concentration of approximately 6 million/mL.[35] Larger baseline testis volume has been consistently found to be a predictor of earlier achievement of sperm output with gonadotropin therapy.[79] Moreover, better results were obtained in men with postpubertal-onset secondary hypogonadism and in those with lower endogenous pretreatment LH and FSH levels.[35] hCG alone could induce sperm output in almost 50% of cases, but sperm concentration remained below 1 million/mL, whereas combined FSH/hMG and hCG treatment is associated with sperm appearance in the ejaculate in 80% of previously azoospermic men, achieving an average sperm concentration of 12 million/mL.[35] Different FSH-like preparations (hMG, hpFSH, rhFSH) produced similar outcomes.[35] A potential detrimental role of previous TRT (usually for induction of secondary sexual development) on semen quality in response to gonadotropin treatment in adulthood was not confirmed.[35] However, it should be noted that this conclusion was based on cumulative data of final outcomes irrespective of different treatment durations. When considering the time needed to achieve sperm output in response to gonadotropin therapy, previous TRT may be a factor in delaying sperm output (40%–60% longer time than TRT-naïve subjects) and attaining peak sperm count (16 months later than TRT-naïve subjects).[79,80] This possible delay in achieving fertility can be a significant issue in light of the trend of increasing age at which couples seek fertility.

Fertility (Also See Chapters 112 and 115). Assessing fertility outcomes in secondary hypogonadal men receiving gonadotropin therapy is challenging, because the fecundity of the female partner plays a critical role. It should be noted that, in secondary hypogonadism without adverse testicular factors (e.g., cryptorchidism), semen quality is usually good despite lower sperm concentrations, so that most patients are nevertheless able to achieve spontaneous pregnancy.[35,78] In the aforementioned meta-analysis, spontaneous pregnancy was obtained on average by 50% of subjects, but it should be noted that not all men included in the meta-analyzed studies were seeking paternity, so fertility rate may be even higher.[35] Fertility from gonadotropin treatment is less likely in men with cryptorchidism or in those having a partner with adverse fertility factors. Unlike sperm output, pretreatment testis volume does not affect pregnancy outcome.[80] In line with sperm output, previous TRT does not affect fertility rate, although it has been associated with a delay in achieving pregnancy.[80]

Adverse Effects. Gonadotropin treatment is generally well tolerated and safe. The side effect most frequently reported is gynecomastia (9% of treated adult men), due to hCG/LH-mediated stimulation of E2 production.[75] The highest incidence was reported with the use of hMG, which has both FSH- and LH-like action.[75] Acne is the second most reported side effect (7%), followed by local injection site reaction and weight gain (both 5%).[75]

Gonadotropin-Releasing Hormone Treatment

Pulsatile GnRH treatment is an alternative to gonadotropin therapy in secondary hypogonadal patients with intact pituitary function. It is administered by subcutaneous pumps that release pulses of 100 to 400 ng/kg of GnRH every 90 to 120 minutes, thus mimicking the physiological hypothalamic pulsatile release of GnRH (Table 106.6) and avoiding downregulation of the GnRH receptors. The need to wear a pump makes this treatment expensive and burdensome, limiting its routine clinical application. In young adults with secondary hypogonadism treated with GnRH, spermatozoa in the ejaculate was obtained on average in 75% of azoospermic subjects, with a mean sperm concentration of 4.3 million/mL. However, direct head-to-head comparison between GnRH and gonadotropin treatment in a limited number of studies suggested a similar efficacy in improving sperm count, although shorter time was required with GnRH, which also produced greater testicular size.[81-83]

FUTURE DIRECTIONS

Functional hypogonadism:
- increase diagnostic awareness and promote its reversibility if possible
- conduct large-scale RCTs to determine long-term efficacy and safety of TRT

Klinefelter syndrome:
- improve diagnosis in “subclinical” cases
- establish postnatal screening programs
- establish multidisciplinary care to:
  - preserve/improve fertility potential
  - early intervention for neurocognitive and behavioral problems
  - promote general health
- Improve the accuracy and standardization of T and free T measurements and harmonize international reference ranges.
- promote general health

REFERENCES

1. Bhasin S, Brito JP, Cunningham GR, et al. Testosterone Therapy in Men With Hypogonadism: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2018;103:1715-1744.
2. Corona G, Goulis DGDG, Huhtaniemi I, et al. European Academy of Andrology (EAA) guidelines on investigation, treatment and monitoring of functional hypogonadism in males. *Andrology*. 2020;8:970–987.
3. Tüttelmann F, Gromoll J. Novel genetic aspects of Klinefelter's syndrome. *Mol Hum Reprod*. 2010;16:386–395.
4. Gravholt CH, Chang S, Wallentin M, et al. Klinefelter syndrome: Integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*. 2018;39:389–423.
5. Stochholm K, Bojesen A, Jensen AS, et al. Criminality in men with Klinefelter's syndrome and XYY syndrome: a cohort study. BMJ Open. 2012;2:e000650.
6. Edelsztein NY, Grinspon RP, Schteingart HF, et al. Anti-Müllerian hormone as a marker of steroid and gonadotropin action in the testis of children and adolescents with disorders of the gonadal axis. *Int J Pediatr Endocrinol*. 2016;2016:1–10.
7. Lucas-Herald AK, Kyriakou A, Alimussina M, et al. Serum anti-Müllerian hormone in the prediction of response to HCG stimulation in children with DSD. *J Clin Endocrinol Metab*. 2020;105:1608–1616.
8. Brauner R, Neve M, Allali S, et al. Clinical, biological and genetic analysis of anorchia in 26 boys. *PLoS One*. 2011;6:4–8.
9. Hviid A, Rubin S, Mühlemann K. Mumps. *Lancet*. 2008;371:932–944.

10. Dobs AS, Dempsey MA, Ladenson PW, et al. Endocrine disorders in men infected with human immunodeficiency virus. *Am J Med*. 1988;84:611–616.
11. Rochira V, Zirilli L, Orlando G, et al. Premature decline of serum total testosterone in HIV-infected men in the HAART-Era. PLoS One. 2011;6:e28512.
12. Halpern JA, Hill R, Brannigan RE. Guideline based approach to male fertility preservation. *Urol Oncol Semin Orig Investig*. 2020;38:31–35.
13. Lambertini M, Del Mastro L, Pescio MC, et al. Cancer and fertility preservation: international recommendations from an expert meeting. *BMC Med*. 2016;14:1.
14. Bandak M, Jørgensen N, Juul A, et al. Testosterone deficiency in testicular cancer survivors - a systematic review and meta-analysis. *Andrology*. 2016;4:382–388.
15. Turner AM. Noonan syndrome. *J Paediatr Child Health*. 2014;50:E14–E20.
16. Moniez S, Pienkowski C, Lepage B, et al. Noonan syndrome males display Sertoli cell-specific primary testicular insufficiency. *Eur J Endocrinol*. 2018;179:409–418.
17. Khan GQ, Hassan G, Qureshi W, et al. Polyglandular autoimmune syndrome type - I. *J, Indian Acad Clin Med*. 2005;6:155–157.
18. Smith-Whitley K. Reproductive issues in sickle cell disease. *Blood*. 2014;124:3538–3543.
19. Bonomi M, Vezzoli V, Krausz C, et al. Characteristics of a nationwide cohort of patients presenting with isolated hypogonadotropic hypogonadism (IHH). *Eur J Endocrinol*. 2018;178:23–32.
20. Cassatella D, Howard SR, Acierno JS, et al. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*. 2018;178:377–388.
21. Boehm U, Bouloux PM, Dattani MT, et al. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism-pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*. 2015;11:547–564.
22. Dwyer AA, Raivio T, Pitteloud N. Reversible hypogonadotropic hypogonadism. *Eur J Endocrinol*. 2016;174:R267–R274.
23. Costa RA, Ferreira IR, Cintra HA, et al. Genotype-phenotype relationships and endocrine findings in Prader-Willi syndrome. *Front Endocrinol (Lausanne)*. 2019;10:864.
24. Suspitsin EN, Imyanitov EN. Bardet-biedl syndrome. *Mol Syndromol*. 2016;7:62–71.
25. McCabe MJ, Alatzoglou KS, Dattani MT. Septo-optic dysplasia and other midline defects: the role of transcription factors: HESX1 and beyond. *Best Pract Res Clin Endocrinol Metab*. 2011;25:115–124.
26. Tresoldi AS, Carosi G, Betella N, et al. Clinically nonfunctioning pituitary incidentalomas: characteristics and natural history. *Neuroendocrinology*. 2020;110:595–603.
27. Sav A, Rotondo F, Syro LV, et al. Pituitary pathology in traumatic brain injury: a review. *Pituitary*. 2019;22:201–211.
28. Alavi SA, Tan CL, Menon DK, et al. Incidence of pituitary dysfunction following traumatic brain injury: a prospective study from a regional neurosurgical centre. *Br J Neurosurg*. 2016;30:302–306.
29. Briet C, Salenave S, Bonneville J-F, et al. Pituitary apoplexy. *Endocr Rev*. 2015;36:622–645.
30. Kurtulmus N, Mert M, Tanakol R, et al. The pituitary gland in patients with Langerhans cell histiocytosis: a clinical and radiological evaluation. *Endocrine*. 2015;48:949–956.
31. Chiloiro S, Giampietro A, Bianchi A, et al. Diagnosis of endocrine disease: primary empty sella: a comprehensive review. *Eur J Endocrinol*. 2017;177:R275–R285.
32. Pan Z, Zhang J, Cheng H, et al. Trends of the incidence of drug use disorders from 1990 to 2017: an analysis based on the Global Burden of Disease 2017 data. *Epidemiol Psychiatr Sci*. 2020;29:e148.
33. Vuong C, Van Uum SHM, O'Dell LE, et al. The effects of opioids and opioid analogs on animal and human endocrine systems. *Endocr Rev*. 2010;31:98–132.
34. Pinsky PF, Prorok PC, Kramer BS. Prostate cancer screening — a perspective on the current state of the evidence. *N Engl J Med*. 2017;376:1285–1289.
35. Rastrelli G, Corona G, Mannucci E, et al. Factors affecting spermatogenesis upon gonadotropin-replacement therapy: a meta-analytic study. *Andrology*. 2014;2:794–808.
36. Spratt DI, O'Dea LS, Schoenfeld D, et al. Neuroendocrine-gonadal axis in men: frequent sampling of LH, FSH, and testosterone. *Am J Physiol Metab*. 1988;254:E658–E666.
37. Brambilla DJ, O'Donnell AB, Matsumoto AM, et al. Intraindividual variation in levels of serum testosterone and other reproductive and adrenal hormones in men. *Clin Endocrinol (Oxf)*. 2007;67:853–862.
38. Diver MJ, Imtiaz KE, Ahmad AM, et al. Diurnal rhythms of serum total, free and bioavailable testosterone and of SHBG in middle-aged men compared with those in young men. *Clin Endocrinol (Oxf)*. 2003;58:710–717.
39. Caronia LM, Dwyer AA, Hayden D, et al. Abrupt decrease in serum testosterone levels after an oral glucose load in men: implications for screening for hypogonadism. *Clin Endocrinol (Oxf)*. 2013;78:291–296.
40. Spratt DI, Kramer RS, Morton JR, et al. Characterization of a prospective human model for study of the reproductive hormone responses to major illness. *Am J Physiol Metab*. 2008;295:E63–E69.
41. Rosner W, Auchus RJ, Azziz R, et al. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2007;92:405–413.
42. Huhtaniemi IT, Tajar A, Lee DM, et al. Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men. *Eur J Endocrinol*. 2012;166:983–991.
43. Handelsman DJ. Mass spectrometry, immunoassay and valid steroid measurements in reproductive medicine and science. *Hum Reprod*. 2017;32:1147–1150.
44. Goldman AL, Bhasin S, Wu FCW, et al. A reappraisal of testosterone's binding in circulation: physiological and clinical implications. *Endocr Rev*. 2017;38:302–324.
45. Hammond GL. Plasma steroid-binding proteins: primary gatekeepers of steroid hormone action. *J Endocrinol*. 2016;230:R13–R25.
46. Antonio L, Wu FCW, O'Neill TW, et al. Low free testosterone is associated with hypogonadal signs and symptoms in men with normal total testosterone. *J Clin Endocrinol Metab*. 2016;101:2647–2657.
47. Anawalt BD, Hotaling JM, Walsh TJ, et al. Performance of total testosterone measurement to predict free testosterone for the biochemical evaluation of male hypogonadism. *J Urol*. 2012;187:1369–1373.
48. Yeap BB, Grossmann M, McLachlan RI, et al. Endocrine Society of Australia position statement on male hypogonadism (part 1): assessment and indications for testosterone therapy. *Med J Aust*. 2016;205:173–178.
49. Fiers T, Wu F, Moghetti P, et al. Reassessing free-testosterone calculation by liquid chromatography–tandem mass spectrometry direct equilibrium dialysis. *J Clin Endocrinol Metab*. 2018;103:2167–2174.
50. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*. 1999;84:3666–3672.
51. Wu FCW, Tajar A, Pye SR, et al. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metab*. 2008;93:2737–2745.
52. Handelsman DJ, Yeap BB, Flicker L, et al. Age-specific population centiles for androgen status in men. *Eur J Endocrinol*. 2015;173:809–817.
53. Finkelstein JS, Lee H, Burnett-Bowie S-AAM, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*. 2013;369:1011–1022.
54. Bhasin S, Woodhouse L, Casaburi R, et al. Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. *J Clin Endocrinol Metab*. 2005;90:678–688.
55. Wu FCW, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*. 2010;363:123–135.
56. Finkelstein JS, Lee H, Burnett-Bowie S-AM, et al. Dose-response relationships between gonadal steroids and bone, body composition, and sexual function in aging men. *J Clin Endocrinol Metab*. 2020;105:2779–2788.
57. Bhasin S, Pencina M, Jasuja GK, et al. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham heart study and applied to three geographically distinct cohorts. *J Clin Endocrinol Metab*. 2011;96:2430–2439.

58. Swerdloff RS, Dudley RE, Page ST, et al. Dihydrotestosterone: Biochemistry, physiology, and clinical implications of elevated blood levels. *Endocr Rev*. 2017;38:220–254.
59. Almeida M, Laurent MR, Dubois V, et al. Estrogens and androgens in skeletal physiology and pathophysiology. *Physiol Rev*. 2017;97:135–187.
60. Dalvi M, Walker BR, Strachan MWJ, et al. The prevalence of structural pituitary abnormalities by MRI scanning in men presenting with isolated hypogonadotrophic hypogonadism. *Clin Endocrinol (Oxf)*. 2016;84:858–861.
61. Howard SR, Dunkel L. Delayed puberty - phenotypic diversity, molecular genetic mechanisms, and recent discoveries. *Endocr Rev*. 2019;40:1285–1317.
62. Rochira V, Antonio L, Vanderschueren D. EAA clinical guideline on management of bone health in the andrological outpatient clinic. *Andrology*. 2018;6:272–285.
63. Barbonetti A, D'Andrea S, Francavilla S. Testosterone replacement therapy. *Andrology*. 2020. February:andr.12774.
64. Surampudi P, Swerdloff RS, Wang C. An update on male hypogonadism therapy. *Expert Opin Pharmacother*. 2014;15:1247–1264.
65. Middleton T, Turner L, Fennell C, et al. Complications of injectable testosterone undecanoate in routine clinical practice. *Eur J Endocrinol*. 2015;172:511–517.
66. McCullough A. A review of testosterone pellets in the treatment of hypogonadism. *Curr Sex Heal Reports*. 2014;6:265–269.
67. Kaminetsky JC, Mccullough A, Hwang K, et al. A 52-week study of dose adjusted subcutaneous testosterone enanthate in oil self-administered via disposable auto-injector. *J Urol*. 2019;201:587–594.
68. Swerdloff RS, Wang C, White WB, et al. A new oral testosterone undecanoate formulation restores testosterone to normal concentrations in hypogonadal men. *J Clin Endocrinol Metab*. 2020;105:1–17.
69. Dinsmore WW, Wyllie MG. The long-term efficacy and safety of a testosterone mucoadhesive buccal tablet in testosterone-deficient men. *BJU Int*. 2012;110:162–169.
70. Rogol AD, Tkachenko N, Bryson N. Natesto^{TM}, a novel testosterone nasal gel, normalizes androgen levels in hypogonadal men. *Andrology*. 2016;4:46–54.
71. Buvat J. Androgen therapy with dehydroepiandrosterone. *World J Urol*. 2003;21:346–355.
72. Ponce OJ, Spencer-Bonilla G, Alvarez-Villalobos N, et al. The efficacy and adverse events of testosterone replacement therapy in hypogonadal men: a systematic review and meta-analysis of randomized, placebo-controlled trials. *J Clin Endocrinol Metab*. 2018;103:1745–1754.
73. Snyder PJ, Bhasin S, Cunningham GR, et al. Lessons from the testosterone trials. *Endocr Rev*. 2018;39:369–386.
74. Corona G, Monami M, Rastrelli G, et al. Type 2 diabetes mellitus and testosterone: a meta-analysis study. *Int J Androl*. 2011;34:528–540.
75. Corona G, Rastrelli G, Reisman Y, et al. The safety of available treatments of male hypogonadism in organic and functional hypogonadism. *Expert Opin Drug Saf*. 2018;17:277–292.
76. Salonia A, Bettocchi C, Boeri L, et al.; EAU Working Group on Male Sexual and Reproductive Health. European Association of Urology Guidelines on Sexual and Reproductive Health-2021 Update: Male Sexual Dysfunction. Eur Urol. 2021;80:333–357.
77. Snyder PJ, Lawrence DA. Treatment of male hypogonadism with testosterone enanthate. *J Clin Endocrinol Metab*. 1980;51:1335–1339.
78. Anawalt BD. Approach to male infertility and induction of spermatogenesis. *J Clin Endocrinol Metab*. 2013;98:3532–3542.
79. Liu PY, Gebski VJ, Turner L, et al. Predicting pregnancy and spermatogenesis by survival analysis during gonadotrophin treatment of gonadotrophin-deficient infertile men. *Hum Reprod*. 2002;17:625–633.
80. Liu PY, Baker HWG, Jayadev V, et al. Induction of spermatogenesis and fertility during gonadotropin treatment of gonadotropin-deficient infertile men: predictors of fertility outcome. *J Clin Endocrinol Metab*. 2009;94:801–808.
81. Gong C, Liu Y, Qin M, et al. Pulsatile GnRH is superior to hCG in therapeutic efficacy in adolescent boys with hypogonadotropic hypogonadodism. *J Clin Endocrinol Metab*. 2015;100:2793–2799.
82. Mao J-F, Liu Z-X, Nie M, et al. Pulsatile gonadotropin-releasing hormone therapy is associated with earlier spermatogenesis compared to combined gonadotropin therapy in patients with congenital hypogonadotropic hypogonadism. *Asian J Androl*. 2017;19:680.
83. Lin J, Mao J, Wang X, et al. Optimal treatment for spermatogenesis in male patients with hypogonadotropic hypogonadism. *Medicine (Baltim)*. 2019;98:e16616.

107

Testosterone Deficiency: Aging and Chronic Disease

Peter J. Snyder

OUTLINE

Aging, 1804
Acute and Chronic Disease, 1807
Medications, 1808

Serum testosterone concentrations decrease not only in diseases of the hypothalamus, pituitary, and testes, but also in association with aging, systemic diseases, and certain medications.

AGING

KEY POINTS

- With aging, a man's serum testosterone concentration decreases slightly, but because the sex hormone–binding globulin concentration increases, the free testosterone decreases more.

Decrease in Testosterone as Men Age. Serum testosterone concentrations decrease with increasing age. In contrast to the decrease in estradiol in women at menopause, which usually occurs over 2 years or less and to very low levels, the decrease in testosterone in men is gradual, over many decades, and to a modest degree. Also in contrast to menopause, which is the result of primary hypogonadism, the fall in testosterone appears to be caused by both primary and secondary hypogonadism.

Total Testosterone. Several cross-sectional studies demonstrated small decreases in serum concentrations of total testosterone from age 20 to 80 years. The European Male Aging Study evaluated 3220 men aged 40 to 79 years old and found that the serum total testosterone concentration fell only 0.04% per year (Fig. 107.1).[1] A study in Australia of 10,904 serum samples from community-dwelling men aged 35 to 99 years showed a gradual decline in testosterone concentration until age 80 years and a steeper decrease thereafter.[2] A study that combined samples from four other studies of 9054 community-dwelling men aged 19 to 99 years showed that the median serum testosterone concentration in men decreased only from 507 ng/dL in men 19 to 39 years of age to 446 ng/dL in men 80 to 99 years of age, but the fifth percentile decreased from 273 to 203 ng/dL, and the 2.5th percentile decreased from 229 to 119 ng/dL.[3]

Several longitudinal studies also demonstrated decreases in serum total testosterone concentrations, and to a somewhat greater degree than the cross-sectional studies. In the Baltimore Longitudinal Study on Aging, 890 men demonstrated a decrease in serum total testosterone concentration of approximately 519 ng/dL at age 30 years to 346 ng/dL at age 80 years.[4] In the Massachusetts Male Aging Study, of 1709 men evaluated initially, 1156 were evaluated a second time, an average of 8 years later, and 853 a third time, an average of 7 years after the second time.[5] Regression modeling estimated a decrease in serum total testosterone concentration of 14.5% per decade of aging.

The decrease in serum testosterone concentration with aging occurs particularly in the early morning. Men 20 to 40 years old exhibit a distinct diurnal pattern of serum testosterone, with prominent peak values occurring at approximately 8:00 a.m. and nadir values at 8:00 p.m., whereas men 60 to 80 years old exhibit a dampened early morning peak.[6]

Sex Hormone–Binding Globulin and Free Testosterone. Sex hormone–binding globulin (SHBG) increased with increasing age in the European Male Aging Study; the increase was approximately 50% from the early 40s to late 70s.[1] As a consequence of the increase in SHBG with aging, the free testosterone decreases to a greater degree than the total.[1] The concentrations in men 75 to 79 years old were about 30% lower than those in men 40 to 44 years old.

Luteinizing Hormone. The serum concentration of luteinizing hormone (LH) increases with increasing age,[1] but not as much as one would expect from the decrease in testosterone if the cause of the hypogonadism was exclusively as a result of Leydig cell damage, suggesting that the fall in testosterone is caused by a combination of primary and secondary hypogonadism.

Summary of the Changes in Testosterone with Aging. As men age, their serum total testosterone concentrations fall modestly, but because the concentrations of SHBG increase with increasing age, the fall in free testosterone is greater.

Effects of Testosterone Treatment of Older Men With Low Testosterone

KEY POINTS

- Testosterone replacement in older men with low testosterone improves all aspects of sexual function and increases walking distance, erythropoiesis, and bone mineral density, but does not improve vitality or cognitive function. This treatment also increases coronary artery plaque volume.

The parallels between many of the consequences of severely low testosterone and many of the consequences of normal aging in men—decreased sexual function, decreased muscle mass and strength, decreased bone mineral density, and decreased hemoglobin—suggest that decreasing testosterone might be one cause of these consequences

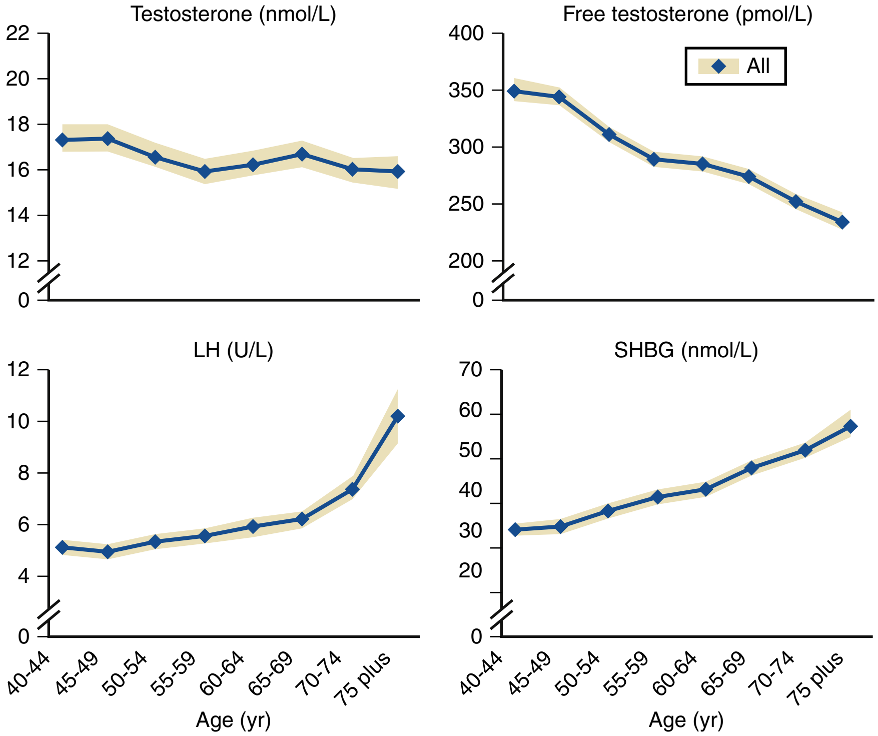


Fig. 107.1 Mean (95% confidence interval) serum concentrations of testosterone, free testosterone, luteinizing hormone (*LH*), and sex hormone–binding globulin (*SHBG*) with aging in 3220 men in the European Male Aging Study. (From Wu FC, Tajar A, Pye SR, et al. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metabol.* 2008;93:2737–2745.)

in older men. Small clinical trials of testosterone treatment of older men in the 1990s suggested possible improvement in some of these parameters, but a committee of the National Academy of Medicine was not convinced. In 2004, the committee recommended that a coordinated set of clinical trials be conducted to determine if testosterone treatment of older men with low testosterone would have any beneficial effect.[7] The Testosterone Trials were designed to implement this recommendation, and the results were reported from 2010 to 2020. A few other trials were conducted during the same period.[8-10]

The Testosterone Trials. The Testosterone Trials were a coordinated group of seven trials (Sexual Function, Physical Function, Vitality, Cognitive Function, Anemia, Bone, and Cardiovascular) conducted to determine if raising the serum testosterone levels in older men with low testosterone would have any benefit.[11] More than 50,000 men were screened to find 790 men who met all of the entry criteria and enrolled.[12] Only 15% of men screened met the enrollment criterion of a serum testosterone concentration of less than 275 ng/dL on two occasions, indicating that, even in elderly men, a low testosterone level is not common. Their mean age was 72 years, and their mean body mass index was 31 kg/m^2; 89% were Caucasian.[13] Enrollees were allocated to treatment with a testosterone gel or placebo; the dose was adjusted to maintain the serum testosterone concentration within the normal range for 1 year. Participants were evaluated for efficacy periodically during the year of treatment.

A total of 705 men (89%) completed the 1-year course of treatment and evaluations.[13] During the year of treatment, the serum testosterone concentrations increased from a median of 234 ng/dL at baseline to the mid-500s during treatment. Serum free testosterone and estradiol also increased to midnormal during treatment.

Beneficial Effects of Testosterone Treatment. Testosterone treatment of men in the Testosterone Trials improved sexual function, walking, mood, hemoglobin, bone density, and strength, each to variable degrees.[14]

Sexual Function. Testosterone treatment, compared with placebo, improved several aspects of sexual function, especially sexual activity—from flirting to intercourse—and libido and, to a lesser extent, erectile function (Fig. 107.2).[13,15] The increases in sexual function and libido, but not erectile function, were significantly associated with the increases in total and free testosterone and estradiol. Testosterone treatment had similar effects in 751 middle-aged men who had low testosterone concentrations.[10]

Physical Function. Testosterone treatment was associated with a small improvement in walking—the distance walked in 6 minutes—in all men in the Testosterone Trials (Fig. 107.3) and also in their perception that their walking had improved.[13,16] These increases were associated with increases in free and total testosterone and hemoglobin. In two smaller trials, testosterone did not improve walking, although it did improve strength.[8,9]

Mood. Testosterone treatment was associated with small but significant improvements in mood, both an increase in positive mood and a decrease in negative mood, as measured by the Positive and Negative Assessment Scales.[13] Treatment was also associated with a small but significant decrease in depressive symptoms, as measured by the Patient Health Questionnaire-9.

Erythropoiesis. In the 126 men in the Testosterone Trials who were anemic at baseline, testosterone increased hemoglobin more than placebo, both in men who were classified as having unexplained anemia (Fig. 107.4) and in men who had anemia of known causes, such as iron deficiency.[17] The average increase in hemoglobin was approximately 1.0 g/dL in both groups.

Bone Mineral Density. A total of 211 men in the Testosterone Trials had measurements of volumetric bone mineral density (vBMD) by

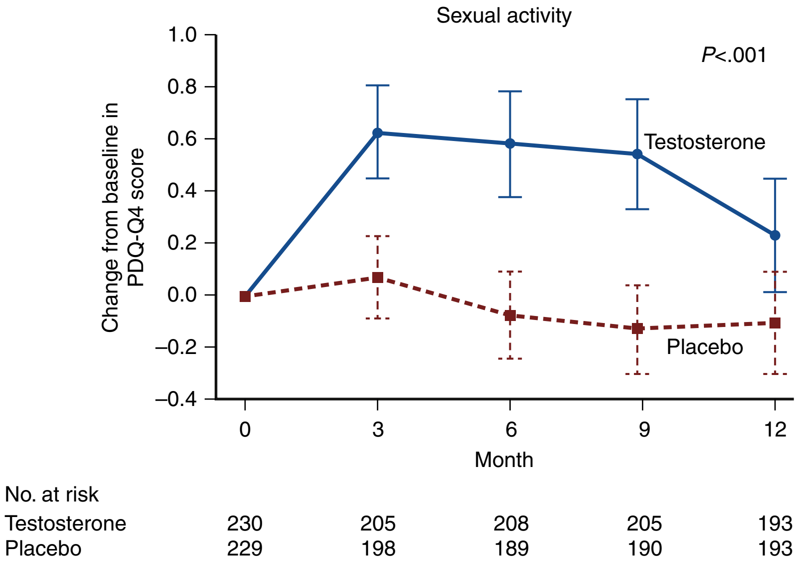


Fig. 107.2 **Mean (95% confidence interval) change in sexual activity when older men with low testosterone were treated with testosterone or placebo for 1 year. Sexual activity was assessed by question 4 of the Patient Daily Questionnaire (PDQ-Q4).** (From Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *NEJM.* 2016;374:611–624.)

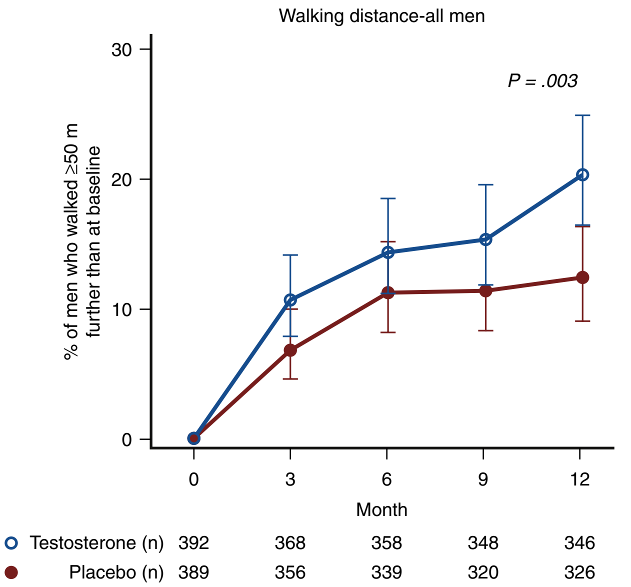


Fig. 107.3 **The mean (95% confidence interval) change in walking distance when older men with low testosterone were treated with testosterone or placebo for 1 year.** (From Snyder PJ, Bhasin S, Cunningham GR, et al. Lessons from the testosterone trials. *Endocr Rev.* 2018;39:369–386.)

quantitative computed tomography (QCT) at baseline and again after 1 year of testosterone treatment.[18] Bone strength was estimated from QCT data by finite element analysis. Testosterone treatment increased vBMD of trabecular bone in the lumbar spine 6.8% more than placebo (Fig. 107.5); the effects on whole bone in the spine and trabecular and whole bone in the hip were less but still significant. Estimated bone strength increased similarly.

Lack of Beneficial Effects. Prior studies of testosterone had suggested that testosterone treatment of older men with low testosterone might have several other beneficial effects, including on vitality, cognition, lipids, and glucose metabolism, but testosterone treatment of the 790 men in the Testosterone Trials had no effect on any of these parameters.

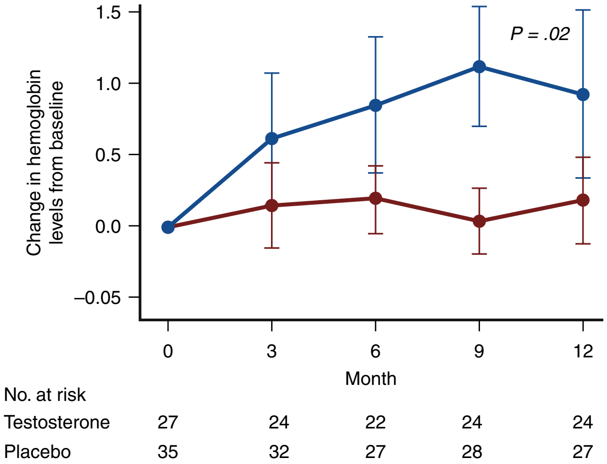


Fig. 107.4 **The mean (95% confidence interval) change in hemoglobin (g/dL) when older men with low testosterone and unexplained anemia were treated with testosterone *(dark green)* or placebo *(orange)* for 1 year.** (From Snyder PJ, Bhasin S, Cunningham GR, et al. Lessons from the testosterone trials. *Endocr Rev.* 2018;39:369–386.)

Vitality. The effect of testosterone on vitality, as measured by the Functional Assessment of Chronic Illness Therapy Fatigue scale, was very small and of borderline significance for all men in the Testosterone Trials.[13]

Cognition. Testosterone treatment had no effect on any aspect of cognition, including delayed or immediate paragraph recall, visual memory, spatial ability, and executive function, in the 493 men in the Testosterone Trials who had age-associated memory impairment.[19]

Lipids and Glucose. Testosterone treatment of all 788 men in the Testosterone Trials was associated with very small changes in total cholesterol, high density lipoprotein cholesterol, and low density lipoprotein cholesterol, none of sufficient magnitude to suggest clinical significance.[20] Treatment was also associated with small decreases in insulin and homeostatic model assessment–insulin resistance but no changes in fasting glucose or hemoglobin A1c concentrations.

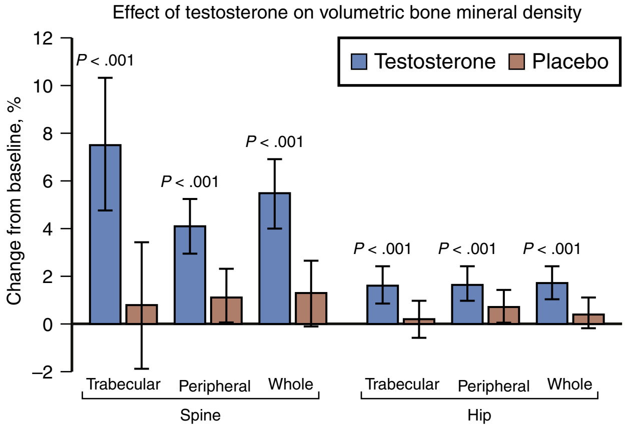


Fig. 107.5 The mean (95% confidence interval) percent change in volumetric bone mineral density of trabecular, peripheral, and whole bone of the spine and hip as determined by quantitative computed tomography when older men with low testosterone were treated with testosterone or placebo for 1 year. (From Snyder PJ, Kopperdahl DL, Stephens-Shields AJ, et al. Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone: a controlled clinical trial. *JAMA Intern Med.* 2017;177:471–479.)

Risks of Testosterone Treatment. Several potential risks of treating older men with testosterone have been proposed. These include stimulating the growth of dormant prostate cancer, worsening benign prostatic hyperplasia, and causing adverse cardiovascular events.

Prostate Cancer. The possibility that testosterone treatment of older men may stimulate the growth of previously dormant prostate cancer has been raised, because the majority of older men harbor silent prostate cancer, and because metastatic prostate cancer can be ameliorated for a period by drastic reduction of the serum testosterone concentration. In the Testosterone Trials, only one man, who was in the testosterone arm, was diagnosed with prostate cancer during the year of treatment.[13] In a metaanalysis of 51 randomized trials of testosterone treatment designed to evaluate a variety of primary outcomes, testosterone treatment was not associated with an increased incidence of prostate cancer.[21] The numbers of men were too small, and the duration of the observations in these studies were too short, however, to draw conclusions about this risk.

Benign Prostatic Hyperplasia. In the Testosterone Trials, all 788 participants were evaluated for lower urinary tract symptoms using the International Prostate Symptom Score. The number of men whose score increased to a level indicating moderately severe symptoms was similar in the men treated with testosterone and the men treated with placebo.[13] In a metaanalysis of 1779 men in four clinical trials, testosterone had no significant effect on lower urinary tract symptoms.[22] The numbers of men and the duration of treatments were too short to draw conclusions about this risk.

Adverse Cardiovascular Events. Several studies suggest the possibility that testosterone treatment of older men with low testosterone may increase the risk of cardiovascular adverse events. Two retrospective cohort studies reported that slightly more men who received testosterone had major adverse cardiovascular events than men who did not.[23,24] A clinical trial of testosterone in men with mobility limitations was stopped early because more men treated with testosterone reported cardiovascular events than men treated with placebo.[25] In a subgroup of 170 men in the Testosterone Trials, testosterone treatment increased coronary artery plaque volume, as determined by computed tomography angiography, by 40 mm^3 in 1 year, compared with 4 mm^3 in the placebo treatment group (Fig. 107.6).[26]

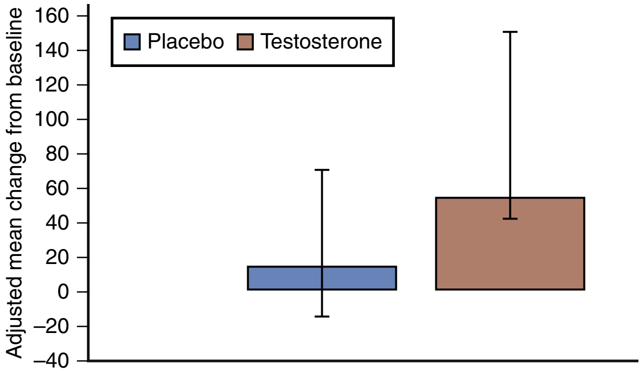


Fig. 107.6 The least mean square (95% confidence interval) percent change in coronary artery plaque volume as determined by computed tomographic angiography when older men with low testosterone were treated with testosterone or placebo for 1 year. (Adapted from Roy CN, Snyder PJ, Stephens-Shields AJ, et al. Association of Testosterone Levels With Anemia in Older Men: A Controlled Clinical Trial. *JAMA internal medicine.* 2017; 177: 480–490. 2017/02/28. DOI: 10.1001/jamainternmed.2016.9540.)

Other studies, however, do not confirm this risk. Another retrospective cohort study showed no increased cardiovascular risk of testosterone.[27] A second clinical trial of testosterone in older men with mobility limitations reported few adverse cardiovascular events.[8] In a clinical trial of 308 older men, testosterone treatment for 3 years did not increase carotid intima media thickness more than placebo.[28] In the Testosterone Trials, the number of major adverse cardiovascular events was similar in the testosterone-treated and placebo-treated men.[13] Although the Testosterone Trials was the largest study of testosterone treatment, it was not nearly large enough and long enough to draw conclusions about cardiovascular risk.

Summary of the Effects of Testosterone Treatment of Older Men with Low Testosterone. Testosterone treatment of older men who have low testosterone has clear benefits—on sexual function, walking, mood, erythropoiesis, and bone density—but the risks are not yet known.

ACUTE AND CHRONIC DISEASE

KEY POINTS

- Several diseases are associated with low testosterone. Improvement of those conditions usually leads to increases in testosterone.

Many acute and chronic diseases cause hypogonadism. The specific mechanisms by which these conditions cause hypogonadism are not known, but most cause secondary hypogonadism, and a few cause a combination of primary and secondary hypogonadism. In many cases, the hypogonadism is reversible; if the disease remits or can be successfully treated, the hypogonadism is reversed.

Critical Illness. Two prospective studies document that critical illness of any kind causes secondary hypogonadism. One study found that 35 men who had trauma, surgery, or myocardial infarction experienced declines in serum testosterone concentrations of approximately 200 ng/dL, accompanied by decreases in LH concentrations, within the first few days after the event.[29] Another study of 59 men found similar decreases in testosterone and LH in response to critical illness and also found that the fall in testosterone was greater the more severe the illness.[30] Serum testosterone increased to normal as the patients recovered from the illness.

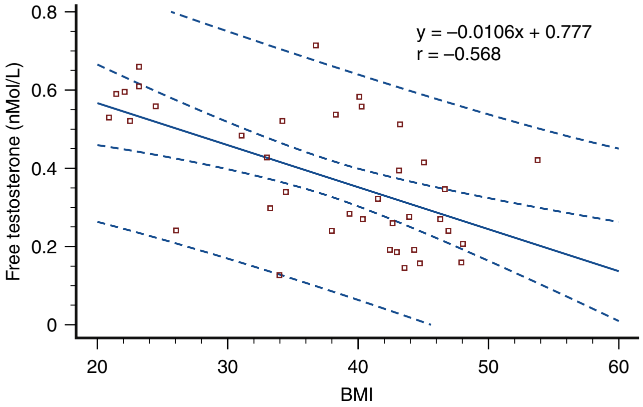


Fig. 107.7 Correlation of free testosterone with body mass index *(BMI)*. (From Vermeulen A, Kaufman JM, Deslypere JP, et al. Attenuated luteinizing hormone (LH) pulse amplitude but normal LH pulse frequency, and its relation to plasma androgens in hypogonadism of obese men. *J Clin Endocrinol Metabol*. 1993;76:1140–1146.)

Anorexia Nervosa. Males who have anorexia nervosa have hypogonadotropic hypogonadism, as do females. In a study of 15 adolescent boys, the mean testosterone concentration was 187 ng/dL, compared with 414 ng/dL in 15 age-matched controls.[31]

Diabetes Mellitus. Many studies have explored testosterone levels in men with diabetes mellitus. Overall, testosterone appears to be lower in diabetics than nondiabetics, but only to a modest degree, and LH is not elevated, indicating that the hypogonadism is secondary. In one metaanalysis, the serum testosterone concentration was 77 ng/dL lower in the diabetic men.[32] In another metaanalysis, testosterone was 86 ng/dL lower in the diabetic than the nondiabetic men.[33] Many studies have also explored the possibility that testosterone treatment would improve glucose control in men who have diabetes. A metaanalysis showed that testosterone treatment of diabetic men with low testosterone resulted in no overall improvement in hemoglobin A1c levels.[33]

Obesity. Obesity causes low serum testosterone by two mechanisms. Men who are obese have lower serum levels of SHBG than men of normal weight, which results in a low total testosterone.[34,35] Men who are obese also have secondary hypogonadism, as determined by low levels of free testosterone; the degree of hypogonadism is proportional to the degree of obesity (Fig. 107.7).[35-37] The hypogonadism is secondary because serum LH levels are not elevated because of decreased LH pulse frequency.[37]

Sleep Apnea. Although some studies report low serum testosterone concentrations in men who have sleep apnea,[38,39] a metaanalysis of five observational studies and two clinical trials comprising a total of 232 men showed no effect of continuous positive airway pressure on serum testosterone concentrations.[40] This finding suggests that the low testosterone in these men may be because of obesity.

Human Immunodeficiency Virus Infection. Before the availability of antiretroviral drugs, human immunodeficiency virus (HIV) infection was sometimes associated with hypogonadism,[41] mostly secondary, especially when accompanied by muscle wasting. Even then, it was not common, as shown by a study of 213 HIV-infected men.[42] Before antiretroviral treatment, the mean serum free testosterone concentration was 92 pg/mL, and only 6% of patients had subnormal values (<50 pg/mL). Antiretroviral treatment of these men for 64 weeks increased the serum free testosterone concentration by 15 pg/mL. In a study of men participating in the Multicenter AIDS Cohort Study Cardiovascular Substudy, the mean total serum testosterone concentration in 364 HIV-infected men was similar to that in 166 uninfected men (629 ng/dL vs. 650 ng/dL, $P = .43$), but the SHBG concentration was slightly higher (52.6 nmol/L vs. 47.3 nmol/L, $P = .02$) and the calculated free testosterone concentration slightly lower (100.2 pg/mL vs. 107.3 pg/mL, $P = .005$) in the HIV-infected men (Fig. 107.8).[43]

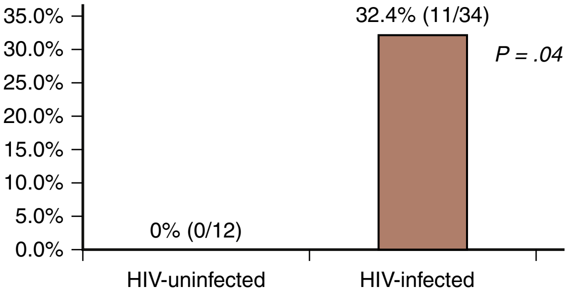


Fig. 107.8 Percentage of men by human immunodeficiency virus *(HIV)* status who had normal testosterone but subnormal free testosterone. (From Monroe AK, Dobs AS, Palella FJ, et al. Morning free and total testosterone in HIV-infected men: implications for the assessment of hypogonadism. *AIDS Res Ther*. 2014;11:6.)

Cirrhosis. Cirrhosis of the liver may be associated with hypogonadism, which may be primary, indicated by elevated serum concentrations of LH and follicle-stimulating hormone (FSH) and by subnormal testosterone responses to administration of human chorionic gonadotropin.[44] Other patients with hypogonadism associated with cirrhosis do not have elevated gonadotropin concentrations. An additional feature of this hypogonadism is elevated serum estradiol concentrations, which cause gynecomastia and elevated serum SHBG concentrations. Hepatic transplantation corrects the hypogonadism.[45]

Chronic Kidney Disease. Severe chronic kidney disease causes hypogonadism primarily by testicular damage. In a study of 35 men whose creatinine clearance rates were less than 4 mL/min/1.7 m², low testosterone was accompanied by elevations of FSH and LH and severe spermatogenic damage as determined by biopsy.[46] The hypogonadism is often corrected following renal transplantation, as illustrated by a study of 97 men with end-stage renal disease.[47] Before renal transplantation, 40% were hypogonadal, but 1 year afterwards, only 18% were. Recovery from hypogonadism was greater in men less than 50 years of age than in those over 50 years of age.

MEDICATIONS

KEY POINTS

- Several medications suppress gonadotropin secretion and thereby cause hypogonadism. Discontinuation of the medication allows recovery of the pituitary-testicular axis, but the time needed for recovery may be prolonged.

Several types of medications cause hypogonadism, some by suppressing gonadotropin secretion and some by damaging the testes.

Gonadotropin-Releasing Hormone Analogs. Repeated administration of superactive analogs of gonadotropin-releasing hormone downregulate their receptors on gonadotroph cells, causing diminished secretion of LH and profound hypogonadism. This property has been employed in the treatment of metastatic prostate cancer since 1982.[48] The resulting chronic hypogonadism increases risk of fracture.[49,50]

Gonadal Steroids. Administration of all gonadal steroids—androgens, estrogens, and progestins—inhibits LH secretion and thereby lowers endogenous testosterone secretion. Although this suppression usually remits after discontinuation of the exogenous steroid, the effect may be prolonged if the dose is high and the duration long. The high doses of androgens that some athletes take to attempt to improve performance

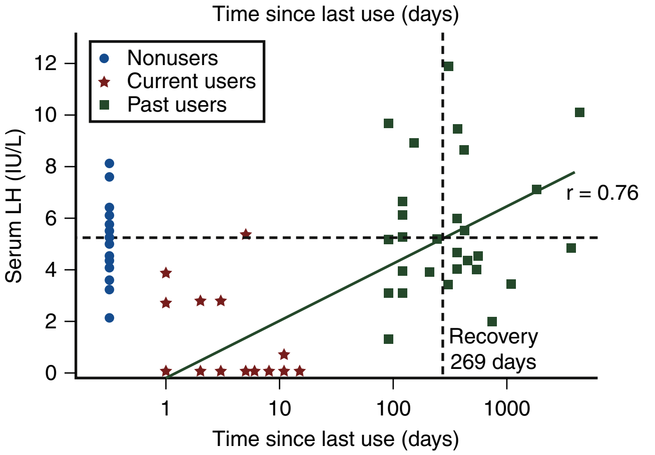


Fig. 107.9 Serum luteinizing hormone (*LH*) concentration as a function of time since discontinuing androgen abuse. (From Shankara-Narayana N, Yu C, Savkovic S, et al. Rate and extent of recovery from reproductive and cardiac dysfunction due to androgen abuse in men. *J Clin Endocrinol Metabol.* 2020;105:dgz324.)

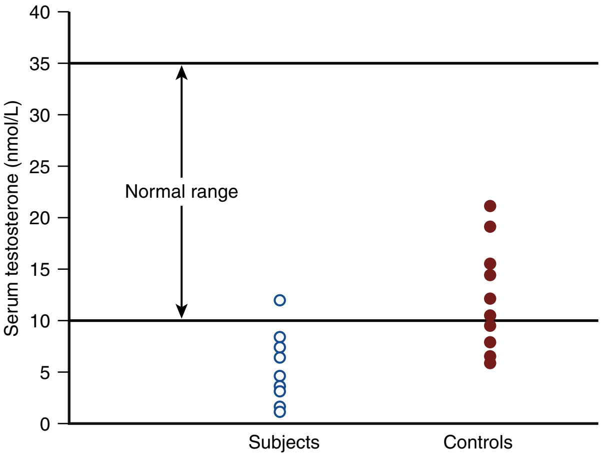


Fig. 107.10 Serum testosterone concentrations in ten men treated with intrathecal morphine for chronic pain and in ten men of similar age and with chronic pain not treated with an opioid. (From Finch PM, Roberts LJ, Price L, et al. Hypogonadism in patients treated with intrathecal morphine. *Clin J Pain.* 2000;16:251–254.)

may take months or even years to remit. In a metaanalysis of four studies involving 22 androgen users evaluated before, during, and after androgen administration, serum testosterone concentrations 13 to 24 weeks after discontinuation were an average of 271 ng/dL lower than before.[51] This phenomenon may be common in young men who seek testosterone treatment, as illustrated by the finding that, of 382 men being treated for hypogonadism at a urology clinic, 80 men, mean age 40 years, said they had previously taken androgenic steroids.[52] An observational study compared reproductive and cardiac parameters in 41 current and 31 former androgen users with those of 21 nonusers.[53] Current users had lower testis volume, sperm output, and serum inhibin B concentrations, as well as impaired cardiac systolic function, compared with nonusers. Variable times since previous use among former users, from 3 months to several years, allowed estimation of the time to recovery of the parameters. The mean time for recovery was 10.7 months for LH (Fig. 107.9), 19.6 months for FSH, and 14.1 months for sperm output. Cardiac function also improved. A longer duration of androgen abuse was associated with a longer duration of recovery.

Opioids. Administration of opioids, especially when chronically and in high doses, causes secondary hypogonadism, often to a marked degree. This effect was noted first in patients who were receiving the drugs intrathecally for chronic pain (Fig. 107.10),[54,55] but later was observed in patients who were taking preparations orally, especially sustained-acting preparations. In a metaanalysis of 15 studies involving 3250 patients, 63% were hypogonadal, more so if the drug was long-acting and if the dose was higher.[56]

Glucocorticoids. High-dose treatment with glucocorticoids can lower serum testosterone concentrations. One study reported that 16 men with chronic pulmonary disease who were treated with high-dose glucocorticoids for 1 month had a mean serum testosterone concentration of 211 ng/dL, compared with 449 ng/dL in 16 men matched for age and disease.[57]

Ketoconazole. Oral administration of ketoconazole, used to treat fungal infections or excessive cortisol secretion, inhibits steroid biosynthesis in both the adrenals and gonads by blocking cholesterol side chain cleavage, as well as blocking 17,20-desmolase, thereby blocking androgen production. Decreased testosterone production has been demonstrated *in vitro* and in response to acute administration in men, but sustained hypogonadism in response to chronic administration has not been demonstrated.

Cytotoxins. Alkylating agents can damage the seminiferous tubules and thereby decrease sperm production but have less of an effect on the Leydig cells and testosterone production.[58,59] The effects are greater when the drugs are given in childhood or adolescence and with longer duration of treatment. The effects may be reversible.

REFERENCES

1. Wu FC, Tajar A, Pye SR, et al. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metabol.* 2008;93:2737–2745.
2. Handelsman DJ, Yeap B, Flicker L, et al. Age-specific population centiles for androgen status in men. *Eur J Endocrinol.* 2015;173:809–817.
3. Travison TG, Vesper HW, Orwoll E, et al. Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *J Clin Endocrinol Metabol.* 2017;102:1161–1173.
4. Harman S, Metter E, Tobin J, et al. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. *J Clin Endocrinol Metabol.* 2001;86:724–731.
5. Travison TG, Araujo AB, Kupelian V, et al. The relative contributions of aging, health, and lifestyle factors to serum testosterone decline in men. *J Clin Endocrinol Metabol.* 2007;92:549–555.
6. Bremner WJ, Vitiello MV, Prinz PN. Loss of circadian rhythmicity in blood testosterone levels with aging in normal men. *J Clin Endocrinol Metabol.* 1983;56:1278–1281.
7. Liverman CTBD. *Testosterone and Aging*. Institute of Medicine of the National Academies; 2003:219.
8. Srinivas-Shankar U, Roberts SA, Connolly MJ, et al. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metabol.* 2010;95:639–650.
9. Travison TG, Basaria S, Storer TW, et al. Clinical meaningfulness of the changes in muscle performance and physical function associated with testosterone administration in older men with mobility limitation. *J Gerontol A Biol Sci Med Sci.* 2011;66:1090–1099.
10. Brock G, Heiselman D, Maggi M, et al. Effect of testosterone solution 2% on testosterone concentration, sex drive and energy in hypogonadal men: results of a placebo controlled study. *J Urol.* 2016;195:699–705.
11. Snyder PJ, Ellenberg SS, Cunningham GR, et al. The Testosterone Trials: Seven coordinated trials of testosterone treatment in elderly men. *Clin Trials.* 2014;11:362–375.

12. Cauley JA, Fluharty L, Ellenberg SS, et al. Recruitment and screening for the testosterone trials. *J Gerontol A Biol Sci Med Sci*. 2015;70:1105–1111.
13. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *NEJM*. 2016;374:611–624.
14. Snyder PJ, Bhasin S, Cunningham GR, et al. Lessons from the testosterone trials. *Endocr Rev*. 2018;39:369–386.
15. Cunningham GR, Stephens-Shields AJ, Rosen RC, et al. Testosterone treatment and sexual function in older men with low testosterone levels. *J Clin Endocrinol Metabol*. 2016;101:3096–3104.
16. Bhasin S, Ellenberg SS, Storer TW, et al. Effect of testosterone replacement on measures of mobility in older men with mobility limitation and low testosterone concentrations: secondary analyses of the Testosterone Trials. *Lancet Diabetes Endocrinol*. 2018;6:879–890.
17. Roy CN, Snyder PJ, Stephens-Shields AJ, et al. Association of testosterone levels with anemia in older men: a controlled clinical trial. *JAMA Intern Med*. 2017;177:480–490.
18. Snyder PJ, Kopperdahl DL, Stephens-Shields AJ, et al. Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone: a controlled clinical trial. *JAMA Intern Med*. 2017;177:471–479.
19. Resnick SM, Matsumoto AM, Stephens-Shields AJ, et al. Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. *JAMA*. 2017;317:717–727.
20. Mohler 3rd ER, Ellenberg SS, Lewis CE, et al. The effect of testosterone on cardiovascular biomarkers in the Testosterone Trials. *J Clin Endocrinol Metabol*. 2018;103:681–688.
21. Fernandez-Balsells MM, Murad MH, Lane M, et al. Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metabol*. 2010;95:2560–2575.
22. Ponce OJ, Spencer-Bonilla G, Alvarez-Villalobos N, et al. The efficacy and adverse events of testosterone replacement therapy in hypogonadal men: a systematic review and meta-analysis of randomized, placebo-controlled trials. *J Clin Endocrinol Metabol*. 2018. [Epub ahead of print].
23. Vigen R, O'Donnell CI, Baron AE, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *JAMA*. 2013;310:1829–1836.
24. Finkle WD, Greenland S, Ridgeway GK, et al. Increased risk of non-fatal myocardial infarction following testosterone therapy prescription in men. *PLoS One*. 2014;9:e85805.
25. Basaria S, Coviello AD, Travison TG, et al. Adverse events associated with testosterone administration. *NEJM*. 2010;363:109–122.
26. Budoff MJ, Ellenberg SS, Lewis CE, et al. Testosterone treatment and coronary artery plaque volume in older men with low testosterone. *JAMA*. 2017;317:708–716.
27. Shores MM, Smith NL, Forsberg CW, et al. Testosterone treatment and mortality in men with low testosterone levels. *J Clin Endocrinol Metabol*. 2012;97:2050–2058.
28. Basaria S, Harman SM, Travison TG, et al. Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: a randomized clinical trial. *JAMA*. 2015;314:570–581.
29. Woolf PD, Hamill RW, McDonald JV, et al. Transient hypogonadotropic hypogonadism caused by critical illness. *J Clin Endocrinol Metabol*. 1985;60:444–450.
30. Spratt DI, Cox P, Orav J, et al. Reproductive axis suppression in acute illness is related to disease severity. *J Clin Endocrinol Metabol*. 1993;76:1548–1554.
31. Misra M, Katzman DK, Cord J, et al. Percentage extremity fat, but not percentage trunk fat, is lower in adolescent boys with anorexia nervosa than in healthy adolescents. *Am J Clin Nutr*. 2008;88:1478–1484.
32. Ding EL, Song Y, Malik VS, et al. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA*. 2006;295:1288–1299.
33. Corona G, Monami M, Rastrelli G, et al. Type 2 diabetes mellitus and testosterone: a meta-analysis study. *Int J Androl*. 2011;34:528–540.
34. Glass AR, Swerdloff RS, Bray GA, et al. Low serum testosterone and sex-hormone-binding-globulin in massively obese men. *J Clin Endocrinol Metabol*. 1977;45:1211–1219.
35. Amatruda JM, Harman SM, Pourmotabbed G, et al. Depressed plasma testosterone and fractional binding of testosterone in obese males. *J Clin Endocrinol Metabol*. 1978;47:268–271.
36. Zumoff B, Strain GW, Miller LK, et al. Plasma free and non-sex-hormone-binding-globulin-bound testosterone are decreased in obese men in proportion to their degree of obesity. *J Clin Endocrinol Metabol*. 1990;71:929–931.
37. Vermeulen A, Kaufman JM, Deslypere JP, et al. Attenuated luteinizing hormone (LH) pulse amplitude but normal LH pulse frequency, and its relation to plasma androgens in hypogonadism of obese men. *J Clin Endocrinol Metabol*. 1993;76:1140–1146.
38. Luboshitzky R, Aviv A, Hefetz A, et al. Decreased pituitary-gonadal secretion in men with obstructive sleep apnea. *J Clin Endocrinol Metabol*. 2002;87:3394–3398.
39. Gambineri A, Pelusi C, Pasquali R. Testosterone levels in obese male patients with obstructive sleep apnea syndrome: relation to oxygen desaturation, body weight, fat distribution and the metabolic parameters. *J Endocrinol Invest*. 2003;26:493–498.
40. Zhang XB, Jiang XT, Du YP, et al. Efficacy of continuous positive airway pressure on testosterone in men with obstructive sleep apnea: a meta-analysis. *PLoS One*. 2014;9:e115033.
41. Dobs AS, Dempsey MA, Ladenson PW, et al. Endocrine disorders in men infected with human immunodeficiency virus. *Am J Med*. 1988;84:611–616.
42. Dube MP, Parker RA, Mulligan K, et al. Effects of potent antiretroviral therapy on free testosterone levels and fat-free mass in men in a prospective, randomized trial: A5005s, a substudy of AIDS Clinical Trials Group Study 384. *Clin Infect Dis*. 2007;45:120–126.
43. Monroe AK, Dobs AS, Palella FJ, et al. Morning free and total testosterone in HIV-infected men: implications for the assessment of hypogonadism. *AIDS Res Ther*. 2014;11:6.
44. Baker HW, Burger HG, de Kretser DM, et al. A study of the endocrine manifestations of hepatic cirrhosis. *Q J Med*. 1976;45:145–178.
45. Guechot J, Chazouilleres O, Loria A, et al. Effect of liver transplantation on sex-hormone disorders in male patients with alcohol-induced or post-viral hepatitis advanced liver disease. *J Hepatol*. 1994;20:426–430.
46. Holdsworth S, Atkins RC, de Kretser DM. The pituitary-testicular axis in men with chronic renal failure. *NEJM*. 1977;296:1245–1249.
47. Reinhardt W, Kubber H, Dolff S, et al. Rapid recovery of hypogonadism in male patients with end stage renal disease after renal transplantation. *Endocrine*. 2018;60:159–166.
48. Tolis G, Ackman D, Stellos A, et al. Tumor growth inhibition in patients with prostatic carcinoma treated with luteinizing hormone-releasing hormone agonists. *Proc Natl Acad Sci U S A*. 1982;79:1658–1662.
49. Shahinian VB, Kuo YF, Freeman JL, et al. Risk of fracture after androgen deprivation for prostate cancer. *NEJM*. 2005;352:154–164.
50. Smith MR, Boyce SP, Moyneur E, et al. Risk of clinical fractures after gonadotropin-releasing hormone agonist therapy for prostate cancer. *J Urol*. 2006;175:136–139;discussion 139.
51. Christou MA, Christou PA, Markozannes G, et al. Effects of anabolic androgenic steroids on the reproductive system of athletes and recreational users: a systematic review and meta-analysis. *Sports Med*. 2017;47:1869–1883.
52. Coward RM, Rajanahally S, Kovac JR, et al. Anabolic steroid induced hypogonadism in young men. *J Urol*. 2013;190:2200–2205.
53. Shankara-Narayana N, Yu C, Savkovic S, et al. Rate and Extent of recovery from reproductive and cardiac dysfunction due to androgen abuse in men. *J Clin Endocrinol Metabol*. 2020;105:dgz324.
54. Finch PM, Roberts LJ, Price L, et al. Hypogonadism in patients treated with intrathecal morphine. *Clin J Pain*. 2000;16:251–254.
55. Abs R, Verhelst J, Maeyaert J, et al. Endocrine consequences of long-term intrathecal administration of opioids. *J Clin Endocrinol Metabol*. 2000;85:2215–2222.
56. de Vries F, Bruin M, Lobatto DJ, et al. Opioids and their endocrine effects: a systematic review and meta-analysis. *J Clin Endocrinol Metabol*. 2020;105:1020–1029.
57. MacAdams MR, White RH, Chipps BE. Reduction of serum testosterone levels during chronic glucocorticoid therapy. *Ann Intern Med*. 1986;104:648–651.
58. Watson AR, Rance CP, Bain J. Long term effects of cyclophosphamide on testicular function. *Br Med J (Clin Res Ed)*. 1985;291:1457–1460.
59. Friedman NM, Plymate SR. Leydig cell dysfunction and gynaecomastia in adult males treated with alkylating agents. *Clin Endocrinol*. 1980;12:553–556.

108

Gynecomastia

Bradley D. Anawalt

OUTLINE

Physiologic Forms of Gynecomastia, 1811
  Newborn Gynecomastia, 1811
  Pubertal Gynecomastia, 1811
  Adult and Senescent Gynecomastia, 1812
Pathologic Forms of Gynecomastia, 1812
  Related to Elevated Estrogen, 1813
  Related to Testosterone Deficiency, 1814
  Related to Estradiol/Testosterone Imbalance, 1814
  Related to Modulatory Hormones, 1815
  Other Causes, 1815
  Uncertain Causes, 1816
Treatment and Prevention, 1817

KEY POINTS

- Growth of male breast tissue (gynecomastia) is caused by high ratio of estrogen to androgen effect in the breast; this high ratio may be caused by low androgen or high estrogen serum concentrations, increased local aromatization of androgens to estrogens in the breast, or altered activation of the androgen or estrogen breast receptors.
- Other hormones, such as prolactin, growth hormone, insulin-like growth factor-1, and thyroid hormone, facilitate breast tissue growth when there is a relatively high ratio of breast tissue estrogen to androgen effect.
- Physiologic gynecomastia occurs commonly in neonatal boys, peripubertal boys, and older adult men.
- Most neonatal boys, peripubertal boys, and men with gynecomastia do not need an initial evaluation beyond a history and physical examination.
- Males with new onset of gynecomastia that is tender, >4 cm, or increasing size should be evaluated. The minimum evaluation is serum testosterone, follicle-stimulating hormone, and luteinizing hormone. Additional evaluation is based on the history, examination, and the results from the initial tests.
- Most males with gynecomastia do not need medical treatment. A 3-month trial of oral daily tamoxifen appears to be safe and effective in peripubertal boys and adult men with clinically significant gynecomastia.
- It is seldom necessary to recommend surgery as a treatment for gynecomastia. With rare exception (e.g., men whose ability to work might be affected by gynecomastia), surgery should only be recommended after treatment of the underlying cause of gynecomastia and only to males with significant, large, chronic gynecomastia (>5 cm).

PHYSIOLOGIC FORMS OF GYNECOMASTIA

Estrogens stimulate and androgens inhibit the growth and differentiation of mammary tissue.[1-3] Estradiol binds to estrogen receptors and stimulates ductal and glandular cells in the male breast. Testosterone exerts a generalized inhibitory action on growth and differentiation, perhaps through a specific antiestrogenic action.[2] The ratio between estrogens and androgens is the principal regulator of the development and maintenance of breast tissue in men. Prolactin, growth hormone, insulin, insulin-like growth factor 1 (IGF-1), and cortisol are permissive trophic factors that require an imbalance in estrogen and androgens to cause gynecomastia.[1] Several hormones alter the estrogen-to-androgen ratio and therefore may cause gynecomastia. Thyroxine increases the level of sex hormone–binding globulin (SHBG), a protein that binds testosterone more avidly than estradiol (thereby increasing the free estrogen-to-androgen ratio). Corticosteroids and prolactin lower circulating testosterone concentrations indirectly by suppressing hypothalamic secretion of gonadotropin-releasing hormone (GnRH).[4-6] Corticosteroids also directly inhibit testicular steroidogenesis.[7,8] The application of these physiologic principles provides a framework for evaluation of gynecomastia on a physiologic basis.

There are two phases of life when gynecomastia is due to normal physiology: just after birth and at the onset of puberty. In adulthood, nontender gynecomastia less than 4 cm is common, and there is often not an identifiable pathology. Nontender gynecomastia is particularly common in older men due to physiological changes that are associated with aging; some experts regard gynecomastia in older men as an effect of normal physiology.

Newborn Gynecomastia

The high concentrations of estradiol and progesterone produced by the mother during fetal life stimulate newborn breast tissue that persists for several weeks after birth in boys. Minimal galactorrhea called *witch's milk* occurs in up to 5% of newborn boys.[9] Galactorrhea is more common in infant boys with larger breast nodules at birth and may persist for 2 months or longer.

Pubertal Gynecomastia

In early puberty (average 11–12 years of age), approximately 30% of boys develop detectable gynecomastia (i.e., glandular tissue >0.5 cm in diameter). By 14 years of age, gynecomastia is detectable in approximately 65%.[10] While it is usually bilateral, gynecomastia is unilateral in 20% of pubertal boys. Spontaneous regression occurs in the majority of boys after 1 to 2 years; long-term gynecomastia persists in less than 5%.[10] There are variable hormonal alterations in pubertal boys with gynecomastia: low free testosterone, increased estradiol, elevated estradiol-to-testosterone ratio, and increased SHBG

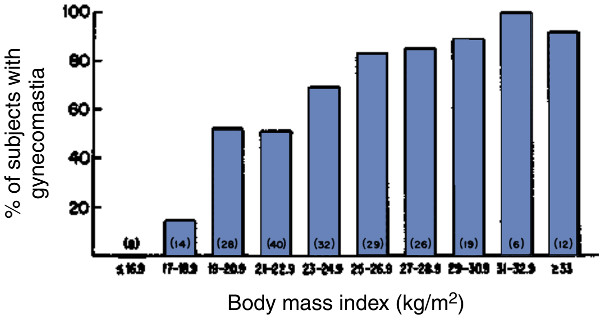


Figure 108.1 Correlation of percentage of subjects with gynecomastia and body mass index. (From Niewoehner CB, Nutall FQ. Gynecomastia in a hospitalized male population. *Am J Med.* 1984;77:633–638.)

concentrations.[11-16] The likely explanation for peripubertal gynecomastia is relative excess of circulating estrogen concentrations compared with circulating androgen concentrations due in part to highly variable aromatization of testosterone during early puberty.[15] In addition, recent data indicate that there are relatively higher concentrations of ERβ than ERα in stromal cells of breast tissue obtained from patients with pubertal gynecomastia. These data suggest that stromal ERβ receptors and associated downstream pathways play an important role in the development of pubertal gynecomastia.[17] In addition, polymorphisms of an estrogen receptor (leading to increased sensitivity to estrogen) may increase the risk of pubertal gynecomastia.[18]

Adult and Senescent Gynecomastia

The best definition of adult gynecomastia is benign glandular tissue that is palpable, greater than 2 cm in diameter, and generally extends outside the area under the nipple. A retrospective study of more than 500 adolescent boys and men who underwent chest computerized tomography for the evaluation of trauma demonstrated that the 90th, 95th, and 97.5th percentile of the diameter of normal breast tissue is 2.2, 2.6, and 3.6 cm, respectively.[19] The correlation of breast tissue volume or diameter between assessment by physical examination and radiographic imaging has not been systematically assessed, but radiographic tissue volume and diameter are highly correlated. Gynecomastia greater than 2 cm of palpable breast tissue is very common in normal ambulatory adults and in hospitalized patients. In a study of 306 normal military recruits (17–59 years of age), 36% had 2 to 4 cm of gynecomastia.[20] In this study, 4% had palpable breast tissue greater than 4 cm, but less than 1% had breast tissue greater than 5 cm. In another study of 954 young men (ages 18-26 years), the prevalence of gynecomastia (> 2 cm) was 40.5%.[21] Hospitalized patients have a higher prevalence of gynecomastia (up to 60%–70% in men over 45 years of age), and approximately 20% have gynecomastia greater than or equal to 5 cm.[22,23] The prevalence and diameter of gynecomastia increase with age and body mass index (Fig. 108.1).[21,23] Autopsy studies confirm the high prevalence of gynecomastia. In large autopsy series, the prevalence of breast tissue is approximately 40%.[24,25]

Most adult men with asymptomatic gynecomastia (i.e., found incidentally on physical examination) will not have an identifiable cause. Because asymptomatic gynecomastia less than 4 cm is so common in adult men, it is generally considered a normal variant or physiologic when there is no clearly identifiable cause. Note that there is no agreement that gynecomastia in adult men may represent a normal (physiologic) finding.[3,20,25]

The cause of physiologic adult gynecomastia in older men is not known, and it is likely to be multifactorial. Asymptomatic gynecomastia that is detected incidentally on physical examination is likely to have developed months or years earlier; the initial cause(s) might not be present. It is likely that age-related declines in serum free and total testosterone, age-related increases in obesity, environmental exposures, and systemic illness, contribute to the high prevalence of adult gynecomastia. In addition, older men often have used or been exposed to drugs and topical agents that decrease circulating or breast tissue androgen concentrations, block the effects of androgens, increase circulating and breast tissue estrogens, or have estrogen-like effects.

KEY POINTS

- Physiological gynecomastia occurs in the neonatal, pubertal, and late stages of adulthood. It is generally bilateral but may be asymmetric or even unilateral.
- Neonatal and pubertal boys commonly have small amounts (0.5–2 cm) of palpable breast tissue that may be tender. This breast tissue growth is due to changes in sex hormone concentrations that normally occur during these stages.
- In older adults, nontender breast tissue that is 2–4 cm in diameter and stable in size is considered physiologic gynecomastia. Nontender breast tissue <2 cm in adults is a normal variant.

PATHOLOGIC FORMS OF GYNECOMASTIA

There are two histologic subtypes of gynecomastia.[26,27] Histologic examination of breast tissue from men with recent onset of gynecomastia demonstrates glandular predominance: hyperplasia of the ductal epithelium, infiltration of periductal tissue with inflammatory cells, and increased subareolar fat. Histologic examination of breast tissue from men with asymptomatic, long-standing gynecomastia demonstrates stromal tissue predominance: dilated ducts with periductal fibrosis, stromal hyalinization, and increased subareolar fat.[28,29] The clinical correlate of these pathologic findings is that longstanding gynecomastia (with fibrosis) is unlikely to spontaneously remit even after resolution of the underlying cause of gynecomastia, but gynecomastia of recent onset may remit.

The clinical definition of pathologic and physiologic gynecomastia is *not* determined by histology. The important clinical distinction between pathologic and physiologic gynecomastia is to define the clinician's extent of evaluation and treatment. Many experts define pathologic gynecomastia pragmatically based on the epidemiology of gynecomastia. Because symptomatic (tenderness) and large (>4 cm in diameter) gynecomastia is uncommon in healthy young men, pathologic gynecomastia is defined as "tender, palpable breast tissue" or "nontender breast tissue >4 cm." Documented growth of breast tissue over time is also considered to be pathologic gynecomastia (even when the diameter is <4 cm). These definitions serve to limit extensive evaluation of gynecomastia, a common condition that is usually not indicative of significant underlying disease. Although these definitions of physiologic and pathologic gynecomastia are practical and clinically useful, they are empirical and not based on published longitudinal studies of men with gynecomastia.

Pathologic gynecomastia is caused by a deficiency of circulating and breast tissue androgen (usually testosterone) concentrations, inhibition of androgen effect on breast tissue, or an excess of circulating or breast tissue estrogen (usually estradiol) concentrations. Many causes of gynecomastia do not fit neatly into one category, but all are associated with an imbalance in the ratio of estrogen to androgen concentrations and/or their effects on the breast. However, it is clinically useful to determine whether the cause of the gynecomastia is due to a loss of the antiproliferative effects of androgen on the breast, an increase in the trophic effects of increased estrogen, or a disturbance of both.

TABLE 108.1 Common Causes of Gynecomastia in Adult (% Prevalence)

Persistent pubertal gynecomastia (25%)
Idiopathic with no detectable abnormality (25%)
Drugs 10%–25%
Hypogonadism 10%
Cirrhosis or malnutrition 8%–10%
Testicular tumors 3%
Hyperthyroidism 1%–2%
Chronic kidney injury 1%

There are many drugs, endocrinopathies, and systemic illnesses that may cause gynecomastia through these mechanisms. Of men with gynecomastia, the two most common etiologies are postpubertal (~25%) and idiopathic (~25%) gynecomastia.[30] Of the men who present with new onset of gynecomastia as adults, a majority (~60%) will have no detectable abnormalities. Drugs or medications and hypogonadism are the next most common causes (Table 108.1).

Related to Elevated Estrogen

Some causes of gynecomastia are driven more by estrogen excess than by androgen deficiency. Because circulating concentrations of estradiol (or other estrogens) inhibit pituitary gonadotropin secretion[31] and therefore reduce testicular production of testosterone, causes of gynecomastia due to estrogen excess are also associated with lower serum concentrations of testosterone.

Tumors. Testicular tumors of germ, sex cord, Leydig, or Sertoli cell origin can secrete an excess of estrogen or estrogen precursors.[32-42] When testicular tumors are the cause of gynecomastia, they are often large enough to be palpated patients,[41,42] but many, including Leydig cell tumors (the most common testicular hormone-secreting tumors), require detection by testicular ultrasound. Choriocarcinomas may produce human chorionic gonadotropin (hCG), a hormone that acts on luteinizing hormone (LH) receptors in the testes to stimulate production of estradiol and testosterone. Ectopic production of hCG by lung, kidney, liver, and gastric carcinomas also stimulates testosterone and estradiol production and causes gynecomastia; hCG and LH increase the activity of aromatase and therefore raise estrogen concentrations more than testosterone concentrations.[43,44]

Rarely, adrenal adenomas or carcinomas produce large amounts of estradiol.[45-47] Alternatively, the adrenal tumor secretes androstenedione, dehydroepiandrosterone (DHEA), and dehydroepiandrosterone sulfate (DHEAS), which are aromatized to estradiol. Patients with adrenal tumors as the cause of gynecomastia usually present with rapid onset of hypertension, elevated estradiol, DHEA, and DHEAS concentrations, and a large adrenal mass.[48]

Aromatase Excess. Overexpression of the aromatase is a unifying feature of several clinical causes of gynecomastia (Table 108.2). Overexpression of aromatase results in increased conversion of androstenedione to estrone and testosterone to estrogen. Obesity is associated with an increased total body production of aromatase, and in one study of 214 carefully assessed men, body mass index correlated significantly ($r = 0.52$) with breast tissue diameter.[49,50] Hyperthyroidism is also associated with increased total body aromatase.[51] Total body aromatase activity also increases with aging, perhaps because of the increased percentage of total body fat or age-related increases in circulating LH concentrations.[52,53]

TABLE 108.2 Aromatase-Associated Causes of Gynecomastia

Increased Amount of Aromatase Enzyme
Increased Activity in Normal Tissue
  1. Obesity
  2. Aging
Aromatase Dysregulation
  1. Aromatase excess syndrome
    a. Familial
    b. Sporadic
Neoplasms
  1. Gonadal tumors
    a. Sertoli cell tumors
      i. Isolated
      ii. Peutz–Jeghers syndrome
      iii. Carney complex
    b. Trophoblastic tumors
  2. Extragonadal tumors
    a. Feminizing adrenocortical neoplasms
    b. Hepatocellular carcinoma
    c. Melanoma
Hormonal Stimulation of Aromatase
Thyrotoxicosis
Idiopathic

Adapted from Braunstein GD. Aromatase and gynecomastia. *Endocr Relat Cancer*. 1999;6:315–324.

Familial gynecomastia, a rare cause of gynecomastia, is due to congenital overexpression of aromatase.[54] Early onset of gynecomastia is the most prominent initial clinical feature of these patients, but it is also associated with accelerated growth before puberty, decreased final adult height, and possibly reduced fertility.[54,55] Other familial disorders are associated with gynecomastia and testicular tumors. Peutz–Jeghers syndrome is associated with pigmented lesions around the mouth, colonic carcinoma, and Sertoli cell tumors.[56-60] These Sertoli cell testicular tumors may contain high amounts of aromatase. The Carney complex of pigmented, nodular adrenal hyperplasia, cardiac myxomas, and lentigenes is also associated with Sertoli cell tumors that often contain high amounts of aromatase.[60,61]

Several sporadic neoplasms cause gynecomastia through overexpression of aromatase. Isolated Sertoli cell tumors, choriocarcinomas, hepatocellular carcinomas, and adrenocortical adenomas, may overexpress aromatase that may cause gynecomastia.[34,38,45,62-64] Choriocarcinomas may cause gynecomastia by both overexpression of aromatase in the tumor and by neoplastic production of hCG that has LH-like effects on aromatase.[34,43,44,62]

There are hints in the literature that a mild increase in peripheral aromatase activity may be responsible for idiopathic gynecomastia in subsets of patients.[65] In one small study, 11 in 30 cases of idiopathic gynecomastia had aromatase excess evidenced by immunochemical assessment of biopsy tissue using a specific antiaromatase antibody.[66] Another study examined aromatase polymorphisms in DNA from whole blood in 100 men with gynecomastia compared with 99 healthy controls.[67] There was a marginally significant increase in T and TT single-nucleotide polymorphisms (SNPs) at *rs10046* in the 3′ untranslated region of exon 10 in the men with gynecomastia (40% vs. 26.3%, $P = 0.04$), but there was no direct evidence that these SNPs cause increased aromatase activity. In a case report, two patients with new onset of idiopathic gynecomastia had increased breast tissue aromatase mRNA compared with two patients with longstanding fibrotic gynecomastia.[68]

Excess Secretion of Estrogen Precursors. Congenital adrenal hyperplasia may cause gynecomastia through excess production of androstenedione, DHEA, and DHEAS, which are converted to estrone and then estradiol.[69-71]

Exogenous Estrogen and its Precursors

Estrogens. Several medications or topical creams provide estrogen or estrogenic substances for absorption and systemic action. Diethylstilbestrol (DES), which was used for treatment of prostate cancer, resulted in gynecomastia with a very high frequency, but it is no longer produced for human use, although it still can be obtained at pharmacies that compound medications.[72] Epidemics of gynecomastia have occurred in which an excess of estrogens present in meats was ingested. Regulatory agencies inspect for illegal feeding of DES or other estrogens to animals prior to slaughter.[73] The practice of estrogen use as an anabolic steroid for animals continues in some areas of the world and can cause gynecomastia in men who consume enough meat with residual contamination. Inadvertent ingestion of oral contraceptives by boys from their mothers' prescriptions or industrial exposure to oral contraceptives during the manufacturing process also can cause gynecomastia.[74] Other drugs such as digitoxin or digoxin have intrinsic estrogenic properties and stimulate breast enlargement.[75] Because this class of drug is administered to patients with severe systemic illnesses, other factors may also be involved in the development of gynecomastia.

Occupational and inadvertent secondary exposures may cause gynecomastia. Embalming fluid has estrogenic effects and may produce gynecomastia.[76] Unintended exposure to estrogen can occur from coital exposure to women using vaginal estrogen cream or even from women using estrogen-containing cosmetics.[77] Exposure of prepubertal boys to estradiol creams utilized as a skin cream by their mothers can cause gynecomastia.[78] In France, estrogen cream is distributed over large portions of the body as a means of transdermal estrogen delivery. Men may develop gynecomastia from secondary transfer with prolonged skin-to-skin contact with women who are applying these estrogen creams.

Estrogen Precursors. Exogenous androgens that can be converted to estrogen through aromatization are a common cause of gynecomastia. Intramuscular testosterone enanthate and cypionate increase estradiol concentrations substantially[79] and commonly result in gynecomastia for a few weeks after initiation of intramuscular therapy. This adverse effect is most common in hypogonadal boys during testosterone treatment, and it generally remits spontaneously (without treatment) after a few weeks to months after initiation of a physiologic dosage of intramuscular testosterone. Because transcutaneous formulations of testosterone are associated with smaller peaks in testosterone concentrations, physiologic dosages of these formulations may cause less gynecomastia. Adult athletes abusing anabolic androgenic steroids that are aromatizable and men being treated with high-dosage intramuscular testosterone for hypogonadism or contraception also may experience gynecomastia.[80,81] When hCG is given exogenously to men with secondary hypogonadism (due to hypothalamic or pituitary disease), it raises testosterone and estradiol concentrations, but because of its LH-like activity, hCG increases aromatization, resulting in a relatively high estrogen-to-androgen ratio, and therefore may cause gynecomastia. At physiologic dosages of hCG, the gynecomastia usually remits after several weeks to months.

Related to Testosterone Deficiency

Some causes of gynecomastia are driven more by androgen deficiency than by estrogen excess. However, most causes of androgen deficiency are associated with relatively greater absolute decreases in serum androgens compared to estrogens. Thus, even in states of androgen deficiency, there is a relative excess of estrogen.

Hypogonadotropic Hypogonadism. Patients with secondary hypogonadism due to hypothalamic or pituitary disease often have gynecomastia due to greater relative declines in serum testosterone than estradiol.[81] Cushing syndrome causes gynecomastia by at least two mechanisms. Glucocorticoids suppress hypothalamic secretion of GnRH and directly inhibit testosterone synthesis, resulting in lower circulating testosterone concentrations and a decreased testosterone-to-estrogen ratio.[4-6]

Anorchia. Fifty percent of anorchic patients also have gynecomastia, due to the absence of testicular androgens and adrenal secretion of aromatizable precursors of estrogen.[25]

Related to Estradiol/Testosterone Imbalance

Some causes of gynecomastia are driven equally by estrogen excess and androgen deficiency.

Hypergonadotropic Hypogonadism. Patients with primary hypogonadism tend to have more gynecomastia than men with secondary hypogonadism because of the dual effects of decreased circulating androgen and markedly increased circulating estrogen concentrations. The markedly increased circulating estrogen concentrations are due to the effects of LH to increase aromatization of testosterone. Klinefelter syndrome is the classic example of this hormonal profile.[81] Men with Klinefelter syndrome have reduced total serum testosterone, a greater reduction in free testosterone, and increases in circulating SHBG, estradiol, follicle-stimulating hormone (FSH), and LH concentrations, and approximately 85% of men with Klinefelter syndrome have gynecomastia. Men with Klinefelter syndrome often have gynecomastia before serum testosterone concentrations have declined below the lower limit of normal. All other forms of primary testicular disease, are associated with an increased risk for gynecomastia based on the same mechanism.[81,82]

Partially Compensated Testicular Dysfunction Associated with Aging. Androgen concentrations gradually decline with aging in men, with free testosterone concentrations declining to a greater extent than total testosterone concentrations.[83] As men age, circulating LH and FSH concentrations also tend to increase; as a result, estrogen concentrations decline to a lesser extent than androgen concentrations. These effects of aging increase the estrogen-to-androgen ratio, even when both circulating steroid concentrations remain in the normal range,[83] a finding that partially explains the high prevalence of gynecomastia in men older than 50 years of age.[22]

Androgen Resistance. A number of disorders and drugs cause decreased tissue responsiveness to androgen.[84-86] Because androgens inhibit breast development, these conditions are associated with gynecomastia. Male patients with complete androgen resistance have female-type breast development despite being genetically male (XY). Because the hypothalamus and pituitary gland of these genotypic males are also resistant to androgens, LH and testosterone are elevated in patients with androgen resistance, resulting in significantly increased circulating estradiol concentrations.

In syndromes of partial androgen resistance, a lesser degree of gynecomastia occurs due to the decreased antitrophic effects of androgen on the breast.[87-89] Kennedy syndrome is another form of androgen resistance due to the androgen receptor, a syndrome associated with variably expanded numbers of CAG repeats in exon 1 of the androgen receptor gene.[90] The diagnosis of Kennedy syndrome is usually delayed until the third to fifth decade.[90,91] Features of Kennedy syndrome include gynecomastia and neurodegenerative findings of symmetric muscular atrophy, weakness, and fasciculations of the bulbar, facial, and proximal muscles of the extremities.[90,91]

Systemic Illness. Any severe systemic illness may cause gynecomastia by causing transient or prolonged hypogonadotropic hypogonadism. Chronic liver disease and human immunodeficiency virus (HIV) infection may cause gynecomastia by additional mechanisms.

Chronic liver disease may produce gynecomastia through a number of mechanisms.[92] In addition to hypogonadotropic hypogonadism due to chronic liver disease, hepatopathy is associated with increased hepatic SHBG production and higher serum SHBG concentrations. As a result, serum free testosterone concentration declines disproportionately to the decline in serum total testosterone concentration. Acute alterations in nutritional state and ingestion of drugs such as spironolactone can further contribute. Of note, although liver disease is commonly believed to cause gynecomastia, one study found an equal prevalence of gynecomastia in hospitalized patients without and with liver disease.[47]

HIV infection may be associated with higher prevalence of gynecomastia, and multiple causes have been identified. The increased prevalence of gynecomastia in patients with HIV may be due to hypogonadism, increased use of illicit and prescription drugs known to cause gynecomastia, and severe chronic kidney or liver disease.[93,94]

Men with HIV infection are often treated with multiple antiretroviral drugs (highly active antiretroviral therapy [HAART]).[93-96] Any form of a HAART regimen may be associated with increased risk for prevalence of gynecomastia, but regimens that include efavirenz appear to be the most likely to cause gynecomastia.[93] The mechanism of HAART-induced gynecomastia is unknown, but one hypothesis is that HAART increases local production of interleukins that increase breast aromatase activity and/or directly stimulate breast growth.[93]

Related to Modulatory Hormones

Prolactin, growth hormone, IGF-1, and thyroxine modulate the actions of androgens and estrogens on breast tissue. Disorders of these hormones may cause or facilitate the development of gynecomastia.

Prolactin is a permissive lactotrophic hormone that stimulates breast growth when estrogen effects are relatively high compared with testosterone effects, but it is unlikely that prolactin directly causes gynecomastia, because many men with significant hyperprolactinemia lack this finding. Hyperprolactinemia, however, reduces GnRH (and therefore LH and testosterone) secretion in many men, lowering the ratio of circulating and tissue testosterone to estradiol concentrations. Therefore, prolactin commonly causes gynecomastia due to the combination of its permissive lactotrophic effects and suppression of circulating and tissue testosterone-to-estradiol concentrations. Prolactinomas and antidopaminergic drugs may cause gynecomastia; antidopaminergic drugs increase prolactin by blocking the tonic inhibitory effects of central dopaminergic pathways. A wide range of drugs with antidopaminergic actions is associated with increased prolactin concentrations.

Hyperthyroidism is commonly associated with breast tenderness and gynecomastia.[98-100] Thyroxine stimulates the production of SHBG, which binds circulating testosterone and estradiol: SHBG binds testosterone more avidly than estrogen.[101] Hyperthyroidism is also associated with increased aromatization of androgens to estrogens.[49] These effects of hyperthyroidism result in lower concentrations of unbound testosterone, higher concentrations of estradiol, and a lower ratio of circulating testosterone to estradiol.[102] Primary hypothyroidism is associated with hyperprolactinemia in women, but rarely in men.[103]

Growth hormone also has permissive trophic effects on breast tissue growth. Patients with acromegaly due to large pituitary tumors may have hypogonadotropic hypogonadism due to compression of the gonadotropes. In this setting of low androgen concentrations, growth hormone excess contributes to the development of gynecomastia. Because there are growth hormone receptors in breast tissue of men with gynecomastia, growth hormone induces breast tissue development directly.[104] It is also likely that growth hormone contributes to breast growth (at least partially) via increased circulating and tissue IGF-1 concentrations. In a recent small study, boys with significant peripubertal gynecomastia had higher serum IGF-1 concentrations than normal controls without significant peripubertal gynecomastia, and there were no significant differences in testosterone, estradiol, or in the ratio of testosterone-to-estradiol concentrations.[105] This study suggests that IGF-1 has a permissive and perhaps direct role in breast tissue development in boys. There have been a few case reports of gynecomastia attributed to exogenous growth hormone therapy in prepubertal boys.[106,107] Gynecomastia that occurs during growth hormone therapy usually spontaneously remits, but occasionally the growth hormone dosage must be decreased or discontinued.[107]

Other Causes

Drug-Induced. Drugs inhibiting testosterone biosynthesis, such as ketoconazole, may cause gynecomastia.[108] Several common drugs, such as cimetidine and spironolactone (but not eplerenone), inhibit the effects of androgens by binding to the androgen receptor.[84-86,] These drugs cause gynecomastia by two mechanisms: 1) reduction of the inhibitory effects of androgen on breast tissue and 2) interruption of negative feedback of testosterone on the pituitary gland, resulting in increased circulating LH concentrations, increased peripheral aromatization of testosterone to estradiol, and a lower testosterone-to-estradiol ratio. 5α-reductase inhibitors used to treat benign prostatic hyperplasia or androgenic alopecia cause gynecomastia in approximately 1% of men by reduced conversion of testosterone to the more potent androgen, dihydrotestosterone (and reduced overall androgen effect in breast tissue).[109] Cyproterone acetate, a drug available outside of the United States, has antiandrogenic effects and may cause gynecomastia.[110] Marijuana, a weak antiandrogen, has been associated with gynecomastia, although a small case-control series did not confirm this association.[111-113]

Treatment of prostate cancer with estrogens or with antiandrogens such as bicalutamide and flutamide frequently causes gynecomastia.[114,116] Approximately 50% of men with prostate cancer have gynecomastia, but multifactorial causes may be involved.[117] High-dosage cytotoxic chemotherapeutic drugs or irradiation can occasionally cause gynecomastia by producing transient or permanent decreases in testosterone production.[118-121] In one series of 190 men with metastatic testicular cancer treated with high-dosage cytotoxic chemotherapy, 4% developed gynecomastia 2 to 5 months after cessation of chemotherapy.[118] Gynecomastia may also occur after chemotherapy or bone marrow transplantation for lymphoma or other cancers.[122]

A large number of other drugs and environmental exposures have been reported to cause gynecomastia (Table 108.3).[84-86,106-160] For many of these drugs, it is not clear whether there is a causal relationship to gynecomastia.

Local Trauma. Trauma to the chest wall may cause gynecomastia, but causality remains unproven. In one small case series, 21 men developed gynecomastia (19 unilateral, two bilateral) after chest trauma.[142] Traumatic gynecomastia has been reported following thoracotomy, in military recruits wearing military packs, in patients with hip spica casts irritating the chest wall, following herpes zoster infection of the chest wall, and following local chest wall irradiation.[81,143] Spinal cord injuries also may rarely be associated with gynecomastia.[144,145]

TABLE 108.3 Drugs Associated with Gynecomastia

Exogenous Estrogens or Estrogenic Substances
Drug therapy with estrogens
Estrogen creams and lotions
Embalming fluid exposure
Delousing powder
Hair oil
Marijuana
Lavender and tea tree oils
Estrogen analogues: digitoxin
Nettle *(Urtica dioica)*
Exogenous Androgens or Luteinizing Hormone Analogs
Testosterone esters
Anabolic androgenic steroids
Human chorionic gonadotropin
Drugs That Inhibit Testosterone Synthesis
Opioids, corticosteroids, alcohol
Androgen-Antagonistic Drugs
Bicalutamide, cimetidine, flutamide, cyproterone, spironolactone
Blockers of 5α-Reductase Inhibitors
Finasteride, dutasteride
Uncertain Mechanisms
Drugs potentially associated with gynecomastia with uncertain mechanisms (incomplete list):
Angiotensin-converting enzyme inhibitors, amiodarone, amphetamines, atorvastatin (and other statins), auronofin, beta blockers, calcium channel blockers, cyclosporine, diazepam, etritinate, efavirenz (and other antiretrovirals), fenofibrate, fluoxetine, griseofulvin, heparin, isoniazid, metronidazole, omeprazole, phenytoin, pregabalin, quinidine, tacrolimus, and theophylline

Macrophages contain large amounts of aromatase, and it is possible that macrophage activity in areas of trauma could increase local breast estrogen production in trauma-induced gynecomastia.[146]

Uncertain Causes

Following World War II, starved prisoners developed gynecomastia upon refeeding. Starvation suppresses gonadotropin production, resulting in a biochemical pattern similar to hypogonadotropic hypogonadism.[147] The recovery phase with refeeding produces changes in testosterone and estrogen concentrations similar to early puberty. The resultant low testosterone-to-estrogen ratio is the likely cause of the gynecomastia that occurs with refeeding after starvation. "Refeeding gynecomastia" may occur in any man recovering from any chronic illness associated with undernutrition due to decreased caloric and nutrient intake or due to gut malabsorption. This mechanism is probably responsible for the gynecomastia observed in some patients with very poorly controlled diabetes mellitus, tuberculosis, and Hansen disease (leprosy).[148,149] Some men with a wasting syndrome from HIV infection may develop gynecomastia by a similar mechanism upon resolution of the wasting syndrome after an effective response to HAART.[150,151]

Pubertal Macromastia. *Pubertal macromastia* is the term used to describe unusually severe pubertal gynecomastia (i.e., female-appearing—Tanner stage III, IV, and V).[152] Unlike mild cases of physiologic pubertal gynecomastia, macromastia is much less likely to spontaneously remit. As with mild pubertal gynecomastia, there is generally no hormonal abnormality seen with macromastia. In one case series of 60 boys with macromastia, only seven had a primary endocrine abnormality or a systemic disease associated with increased aromatase activity.[153]

Idiopathic Gynecomastia. In the outpatient setting, the majority of patients with gynecomastia will have no detectable cause of the gynecomastia after careful endocrinological evaluation.[154]

Breast Enlargement Due to Malignancy or Infection. A carcinoma of the breast is very uncommon in men. Men with Klinefelter syndrome have an increased risk for breast cancer.[155] A firm, irregular, and unilateral mass is characteristic and suggests the diagnosis. Certain infectious diseases (e.g., Hansen disease, filariasis) may also cause breast enlargement due to infection of breast tissue.[156,157]

KEY POINTS

- Pathologic forms of gynecomastia are defined as gynecomastia >4 cm or of recent onset (<6 months) that may be tender (indicating active growth) or measurably growing over time. The causes of pathological gynecomastia are generally increased estrogen effects or decreased testosterone effects on the male breast tissue. There are a number of causes of gynecomastia with unknown mechanism, including refeeding gynecomastia, pubertal macromastia, and idiopathic gynecomastia. The most common cause of pathologic gynecomastia is idiopathic with no identifiable cause.

Evaluation. The first important step in the evaluation is to exclude pseudogynecomastia—fat tissue underneath the nipple. The appropriate technique for physical examination to detect gynecomastia is to pinch the tissue between thumb and forefinger lateral to the nipple or between fingers of both hands (Fig. 108.2). Breast tissue feels firm and rubbery with a firm edge. Ability to feel an edge of tissue at the interface of normal and glandular tissue confirms the presence of gynecomastia. Comparison of consistency with that of the fat tissue over the abdomen or along the axillary line of the breast is useful. Pressing down on the breast toward the chest wall is insensitive in detecting gynecomastia and is not useful. It is rarely necessary to use mammography to distinguish breast tissue from adipose tissue.[158-161] Ultrasonography can also be used when a patient refuses mammography or when mammography is infeasible, but it is generally a second-line imaging technique.[162-163]

A careful history and physical examination is generally sufficient to evaluate neonatal or peripubertal gynecomastia. It is also unnecessary to perform biochemical or radiographic evaluation in adult men with chronic, stable gynecomastia less than 5 cm in diameter. It is prudent to perform a more extensive evaluation in peripubertal boys with macromastia and adult men with new onset of gynecomastia, rapidly enlarging, eccentric, or hard, irregular masses, or gynecomastia greater than 4 cm in diameter. Firm, irregular lesions raise concern about neoplasm, and they should be evaluated by excisional biopsy.

History. The key questions are the following: (1) the time of onset, rate of progression, and degree of pain associated with the gynecomastia; (2) symptoms of androgen deficiency; (3) use of medications, prescription and nonprescription drugs, "nutritional" supplements, and skin products; (4) the presence of severe systemic disease such as poorly controlled diabetes mellitus or severe renal, hepatic, cardiac, or pulmonary disease that could suppress the hypothalamic-pituitary-gonadal axis or increase SHBG levels (e.g., hepatic disease); (5) recent nutritional deficiencies due to diet or gut malabsorption; (6) symptoms of underlying malignancy, especially testicular; (7) symptoms associated with estradiol, prolactin, growth hormone, cortisol, or thyroxine

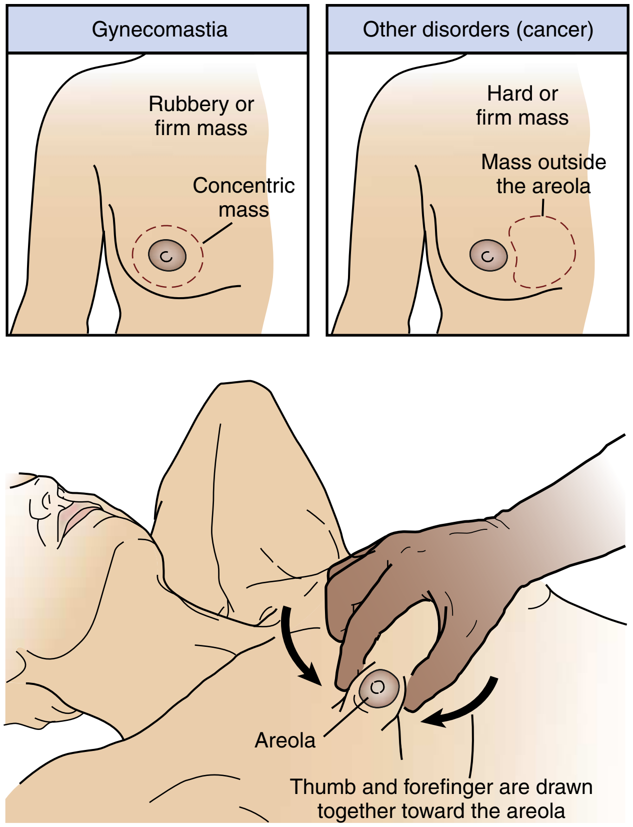


Figure 108.2 Method of examination to detect gynecomastia. **A,** Palpation by pressing down on tissue beneath the nipple in a supine patient is an insensitive method for detecting gynecomastia and is considered incorrect. **B,** Squeezing tissue between the thumb and forefinger in a sitting patient is a more sensitive technique and is considered correct. The examiner should try to palpate and "flip an edge" between the normal fat tissue and the glandular tissue to distinguish the outer limits of the gynecomastia. (Adapted from Santen RJ, Cerilli LA, Harvey JA, et al. In: Besser G, Thorner M, eds. *Comprehensive Clinical Endocrinology*. 3rd ed. Amsterdam: Mosby; 2002.)

excess; (8) a recent history of chest trauma; and (9) family history of gynecomastia.

It is important to specifically ask about the use of new skin products or medications by the patient or intimate contacts. There have been case reports of gynecomastia related to many skin products, including case reports of gynecomastia due to lavender oil and secondary exposure to estradiol cream used by a female partner or family member.[166,167]

Physical Examination. After using the "pinch test" to exclude pseudogynecomastia, the key elements of the physical examination are the following: 1) assessment for obesity with body mass index and examination for subcutaneous and abdominal adiposity; 2) assessment for evidence of malnutrition and severe systemic disease; 3) skin examination for signs of Cushing syndrome (striae, ecchymoses), acromegaly, and hyperthyroidism; 4) thyroid examination for goiter; 5) careful breast examination for galactorrhea or a hard, firm mass; 6) lymph node examination for large, hard, or fixed axillary lymphadenopathy; 7) deep palpation of the abdomen to detect large masses; 8) genital examination for hypospadias and assessment of testicular volume and testicular mass; and 9) assessment of evidence of other findings of systemic disease or endocrinopathy as indicated.

In peripubertal boys with macromastia and adult men with new onset of gynecomastia, rapid enlargement, and eccentric or hard, irregular masses or gynecomastia greater 4 cm in diameter, measurement of circulating testosterone, FSH, and LH concentrations is a reasonable initial laboratory assessment. Many experts recommend measurement of thyroid-stimulating hormone to detect subclinical hyperthyroidism. This recommendation is reasonable, and the test is relatively inexpensive, but, in the author's experience, it is not useful. Men with new onset of gynecomastia due to hyperthyroidism have clinical manifestations of hyperthyroidism, and gynecomastia is rarely (if ever) due to subclinical hyperthyroidism.

Many experts also recommend measurement of adrenal androgen and estrogen precursors such as DHEA or DHEAS, but these tests are not necessary in the routine initial evaluation. Adrenal sex-steroid–producing tumors are very rare causes of new onset gynecomastia, and are virtually always associated with large, bulky abdominal masses, abdominal pain, constitutional signs of malignancy, and suppressed gonadotropin levels. Measurement of estradiol and hCG concentrations is likewise not indicated on a routine basis. Tumors that produce estradiol or hCG will be associated with suppressed circulating FSH and LH concentrations.[168,169]

Finally, there is some debate about whether to measure estradiol concentrations on a routine basis. Commonly available commercial immunoassays for estradiol are notoriously inaccurate at the low concentrations typically found in men. Estradiol assays using tandem liquid chromatography followed by mass spectrometry are accurate in men, but these assays are more expensive and not widely available. The vast majority of men with gynecomastia have normal or only slightly elevated estradiol concentrations, and determination of estradiol concentrations does not affect the management of gynecomastia in these men. Because estradiol is a potent inhibitor of endogenous gonadotropin (and therefore endogenous testosterone) secretion, this author reserves measurement of estradiol and hCG for the rare patient with gynecomastia and suppressed circulating LH and FSH concentrations. When assessing for an hCG-producing tumor, it is best to use an assay with monoclonal antibodies that detects all forms of intact hCG and the free beta subunit of hCG.[172] Fig. 108.3 outlines a reasonable algorithmic evaluation of gynecomastia.

Imaging. As noted earlier, imaging is generally not required to make the diagnosis of gynecomastia. If a breast tumor is suspected, then imaging with mammography or ultrasound and obtaining tissue by fine-needle aspiration or core biopsy should be performed for pathologic examination.[158-165,170-175] In general, ultrasound does not add significant diagnostic information if a mammogram has been performed.[176]

KEY POINTS

- The diagnosis of pathologic gynecomastia is based on the history and examination.
- The evaluation for pathologic gynecomastia includes review of medications and topical agents associated with gynecomastia and history and physical examination for evidence of hypogonadism, hyperthyroidism, and chronic diseases, including kidney and liver disease. Measurement of serum testosterone, follicle-stimulating hormone, and luteinizing hormone in an early morning blood sample is generally the sole initial laboratory assessment. Further testing is based on the results of these tests and clinical suspicion of other systemic disease. Imaging is virtually never necessary unless breast malignancy is suspected based on palpation of a hard, immobile mass.

TREATMENT AND PREVENTION

Management of gynecomastia depends upon the clinical circumstances, severity, and presence of pain. The first principle is to stop

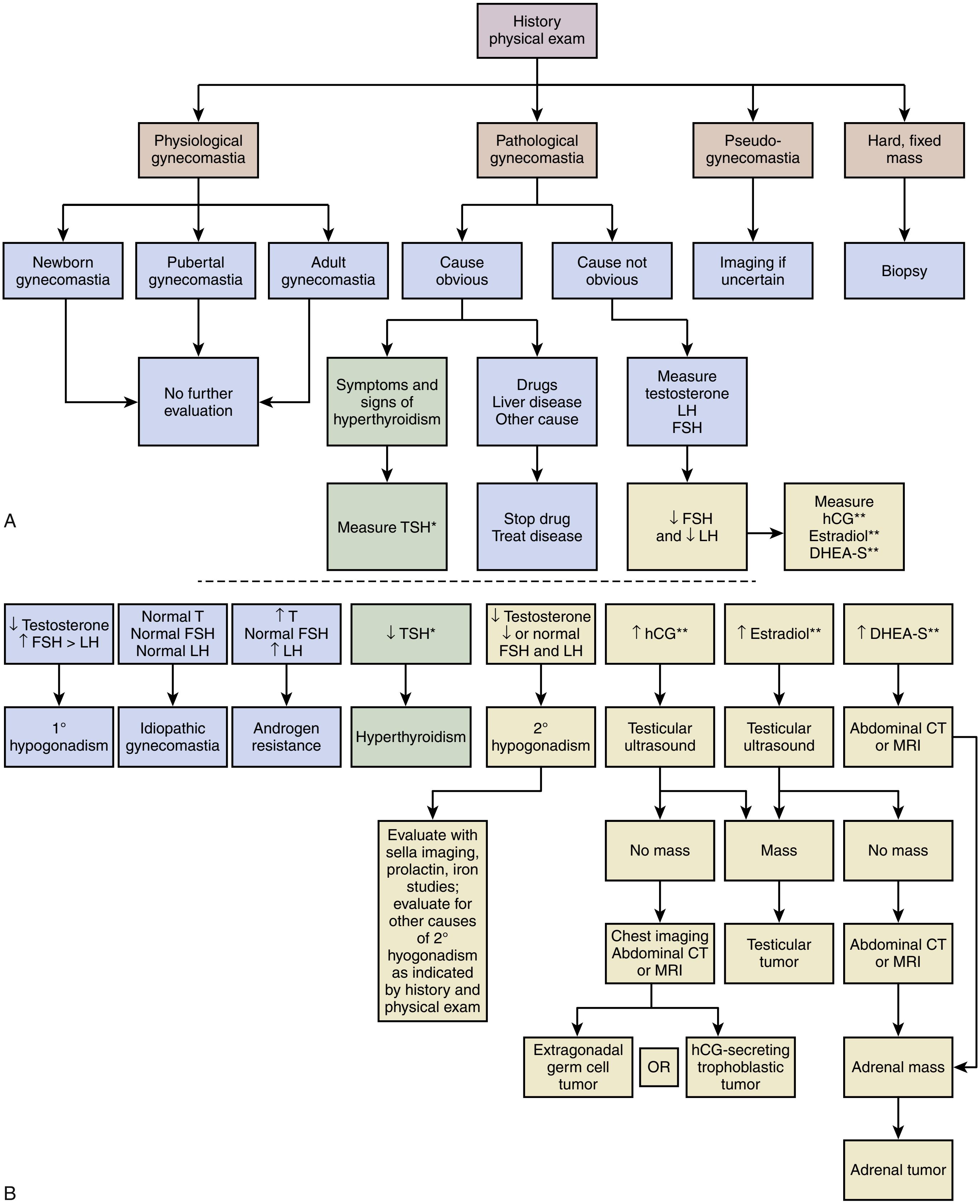

History physical exam
Physiological gynecomastia
Pathological gynecomastia
Pseudo-gynecomastia
Hard, fixed mass
Newborn gynecomastia
Pubertal gynecomastia
Adult gynecomastia
Cause obvious
Cause not obvious
Imaging if uncertain
Biopsy
No further evaluation
Symptoms and signs of hyperthyroidism
Drugs Liver disease Other cause
Measure testosterone LH FSH
A
Measure TSH*
Stop drug Treat disease
↓ FSH and ↓ LH
Measure hCG** Estradiol** DHEA-S**
↓ Testosterone ↑ FSH > LH
Normal T Normal FSH Normal LH
↑ T Normal FSH ↑ LH
↓ TSH*
↓ Testosterone ↓ or normal FSH and LH
↑ hCG**
↑ Estradiol**
↑ DHEA-S**
1° hypogonadism
Idiopathic gynecomastia
Androgen resistance
Hyperthyroidism
2° hypogonadism
Testicular ultrasound
Testicular ultrasound
Abdominal CT or MRI
Evaluate with sella imaging, prolactin, iron studies; evaluate for other causes of 2° hyogonadism as indicated by history and physical exam
No mass
Mass
No mass
Chest imaging Abdominal CT or MRI
Testicular tumor
Abdominal CT or MRI
Extragonadal germ cell tumor
OR
hCG-secreting trophoblastic tumor
Adrenal mass
Adrenal tumor
B

Figure 108.3 A, Algorithm depicting the initial evaluation of patients with gynecomastia. History and physical examination generally suffices to distinguish between the four most common categories of male breast enlargement *(pale blue boxes).* For men with recent onset of gynecomastia, increasing size, or breast diameter >4 cm, measurement of testosterone, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) is indicated. A serum thyroid-stimulating hormone (TSH) level should be performed if the patient has symptoms and signs of thyrotoxicosis *(yellow boxes).* Evaluation for causes of hypogonadotropic (secondary) and measurement of serum human chorionic gonadotropin (hCG), estradiol, and dehydroepiandrosterone sulfate (DHEAS) concentrations should be performed in men with pathologic gynecomastia and low serum testosterone and gonadotropin concentrations *(lavender boxes).* B, Interpretation and further evaluation after initial diagnostic evaluation of gynecomastia. *Some experts recommend measurement of serum TSH in all men with pathologic gynecomastia. **Some experts recommend measurement of serum hCG, estradiol, and DHEAS in all men with pathologic gynecomastia.

using offending drugs and treat any specific abnormality amenable to therapy. Boys with pubertal gynecomastia can be reassured that regression of breast tissue usually occurs after 1 year and at most after 3 years, and that no treatment is usually necessary.[10,11]

Studies of pubertal and adult gynecomastia require cautious interpretation, because gynecomastia resolves spontaneously, and most of the therapeutic trials do not use placebo controls for comparison. For example, in one randomized trial, investigators compared the aromatase inhibitor anastrozole with placebo in 80 boys with pubertal gynecomastia.[177] Although the aromatase inhibitor decreased estradiol concentrations in the active treatment arm, the percentage of patients with decreased breast tissue diameter was 38.5% with anastrozole and 31.4% with placebo (P = not significant). This study used state-of-the-art clinical trial methods and highlights the need for a placebo control in all clinical trials of medical therapy for gynecomastia.

Based on the pathophysiology of gynecomastia, responses to aromatase inhibitors would be expected both in pubertal and adult subjects. Although anecdotal reports suggested that aromatase inhibitors are effective, the clinical trials have yielded conflicting results.[177-181] The largest aromatase inhibitor study demonstrated no benefit for anastrozole over placebo in the treatment of significant pubertal gynecomastia (≥3 cm).[177] Because aromatase inhibitors cause adverse effects on bone (decreased bone density) and fat (increased total body fat), these drugs are not recommended for the treatment or prevention of gynecomastia.

Selective estrogen receptor modulators (tamoxifen, clomiphene, and raloxifene) have been shown to significantly reduce the amount of breast tissue in males with gynecomastia of recent onset.[182-195] Although these studies had methodologic flaws, it appears that this class of drugs can be used safely and effectively in the short term for some pubertal boys and adult men with recent onset of gynecomastia.

Treatment of Pubertal Gynecomastia. A prudent approach to management of pubertal gynecomastia is to observe for spontaneous regression and reserve medical therapy for boys with macromastia or significant persistent pain. Breast tenderness resolves within a few months in virtually all cases of pubertal gynecomastia, and breast tissue remits in approximately 80% of cases. Medical therapy for macromastia or very painful pubertal gynecomastia should wait until *after* any reversible cause of gynecomastia has been treated. Although there have been no trials with head-to-head comparisons, the data suggest that tamoxifen may be the most effective therapy. However, even for tamoxifen, the best-studied of the available medical therapies, the data are weak. A 2013 systematic review determined that only six of 164 publications on the use of tamoxifen in pubertal gynecomastia were worth including in the study; there were no randomized, controlled trials.[195] The authors concluded that tamoxifen appears to be safe and may be effective in the treatment of pubertal gynecomastia.

For macromastia or pubertal gynecomastia with significant pain and tenderness, it is reasonable to treat with tamoxifen (10 or 20 mg daily) for 3 months and reassess. For persistent (>2 years) pubertal macromastia, surgical excision may be considered. It is important to consult a surgeon who is experienced in mammoplasty to ensure optimal cosmetic results.

Treatment of Adult Gynecomastia. In adult men with gynecomastia, specific therapy for causative problems should be used when feasible, and offending drugs should be discontinued when possible. For example, eplerenone, a mineralocorticoid inhibitor without antiandrogenic effects, may ameliorate gynecomastia due to aldactone, a commonly used drug for cirrhosis and heart failure.[196,197]

In men with idiopathic gynecomastia, carefully controlled randomized studies of medical therapy are limited. Two studies compared tamoxifen with placebo and detected a reduction in size in 10 of 16 (62%) men receiving the drug and one of 16 men receiving placebo. Pain relief occurred in nine of 10 on tamoxifen and one of 10 with placebo.[182,183] Three noncontrolled trials reported six of 16 (37%) with size reduction and 13 of 15 (87%) with pain relief.[184-186] Danazol, a weak androgen, and dihydrotestosterone, a nonaromatizable androgen, have also been shown in small trials to be modestly effective in the treatment of gynecomastia, but neither androgen is used commonly in the treatment of gynecomastia.[198]

Based upon these data, a trial of tamoxifen 20 mg daily for 3 months may be useful in some patients with recent-onset, very symptomatic, idiopathic gynecomastia. If gynecomastia persists for more than 1 year and is cosmetically very bothersome, reduction mammoplasty may be necessary. A surgeon who has experience with this procedure should be selected in order to achieve optimal cosmetic effects with a minimally invasive procedure.[199-202]

Prevention of Adult Gynecomastia. Prevention of gynecomastia may be useful in men undergoing antiandrogen therapy for treatment of prostate cancer. Prophylactic tamoxifen is the best-studied effective preventive therapy. Prophylactic tamoxifen reduces the frequency of gynecomastia due to antiandrogen therapy for prostate cancer by 60% to 80%.[203-207] Prophylaxis with low-dosage irradiation of the breast also reduces the incidence of gynecomastia in men treated with antiandrogen therapy, but this preventive treatment is rarely used now, in part because tamoxifen appears to be more effective.[208-216]

KEY POINTS

- Gynecomastia is generally treated by managing the underlying cause. Medical therapy is seldom indicated. For patients with severe persistent breast tenderness due to pubertal idiopathic gynecomastia, a short course of tamoxifen (<6 months) may be beneficial. Breast reduction surgery is reserved for those men who are very bothered by the amount of breast tissue, and this surgery should be delayed until the underlying cause has been treated and breast tissue growth and tenderness has resolved for at least 6 months.

REFERENCES

1. McNally S, Martin F. Molecular regulators of pubertal mammary gland development. *Ann Med*. 2011;43:212–234.
2. Casey RW, Wilson JD. Antiestrogenic action of dihydrotestosterone in mouse breast. Competition with estradiol for binding to the estrogen receptor. *J Clin Invest*. 1984;74:2272–2278.
3. Narula HS, Carlson HE. Gynecomastia. *Endocrinol Metab Clin North Am*. 2007;36:497–519.
4. MacAdams MR, White RH, Chipps BE. Reduction of serum testosterone levels during chronic glucocorticoid therapy. *Ann Intern Med*. 1986;104:648–651.
5. Contreras LN, Masini AM, Danna MM, et al. Glucocorticoids: their role on gonadal function and LH secretion. *Minerva Endocrinol*. 1996;21:43–46.
6. Bouchard P, Lagoguey M, Brailly S, et al. Gonadotropin-releasing hormone pulsatile administration restores luteinizing hormone pulsatility and normal testosterone concentrations in males with hyperprolactinemia. *J Clin Endocrinol Metab*. 1985;60:258–262.
7. Welsh Jr TH, Bambino TH, Hsueh AJ. Mechanism of glucocorticoid-induced suppression of testicular androgen biosynthesis in vitro. *Biol Reprod*. 1982;27:1138–1146.

8. Bambino TH, Hsueh AJ. Direct inhibitory effect of glucocorticoids upon testicular luteinizing hormone receptor and steroidogenesis in vivo and in vitro. *Endocrinology*. 1981;108:2142–2148.
9. Madlon-Kay DJ. “Witch’s milk”. Galactorrhea in the newborn. *Am J Dis Child*. 1986;140:252–253.
10. Nydick M, Bustos J, Dale Jr JH, et al. Gynecomastia in adolescent boys. *J Am Med Assoc*. 1961;178:449–454.
11. Biro FM, Lucky AW, Huster GA, et al. Hormonal studies and physical maturation in adolescent gynecomastia. *J Pediatr*. 1990;449:450–455.
12. Lee PA. The relationship of concentrations of serum hormones to pubertal gynecomastia. *J Pediatr*. 1975;86:212–215.
13. LaFranchi SH, Parlow AF, Lippe BM, et al. Pubertal gynecomastia and transient elevation of serum estradiol level. *Am J Dis Child*. 1975;129:927–931.
14. Knorr D, Bidlingmaier F. Gynaecomastia in male adolescents. *Clin Endocrinol Metab*. 1975;4:157.
15. Large DM, Anderson DC. Twenty-four hour profiles of circulating androgens and oestrogens in male puberty with and without gynaecomastia. *Clin Endocrinol (Oxf)*. 1979;11:505–521.
16. Moore DC, Schlaepfer LV, Paunier L, et al. Hormonal changes during puberty: V: transient pubertal gynecomastia: abnormal androgen-estrogen ratios. *J Clin Endocrinol Metab*. 1984;58:492–499.
17. Nicoletti GF, D’Andrea F, Ferraro G, et al. Differential expression of estrogen receptor α and β transcripts in tissues and in primary culture cells from pubertal gynecomastia. *J Endocrinol Invest*. 2012;35:150–153.
18. Korkmaz HA, Edgünlü T, Eren E, et al. GPR30 gene polymorphisms are associated with gynecomastia risk in adolescents. *Horm Res Paediatr*. 2015;83:177–182.
19. Klang E, Kanana N, Grossman A, et al. Quantitative CT assessment of gynecomastia in the general population and in dialysis, cirrhotic and obese patients. *Acad Radiol*. 2018;25:626–635.
20. Nuttall FQ. Gynecomastia as a physical finding in normal men. *J Clin Endocrinol Metab*. 1979;48:338–340.
21. Georgiadis E, Papandreou L, Evangelopoulou C, et al. Incidence of gynaecomastia in 954 young males and its relationship to somatometric parameters. *Ann Hum Biol*. 1994;21:579–587.
22. Carlson HE. Gynecomastia. *N Engl J Med*. 1980;303:795–799.
23. Niewoehner CB, Nuttall FQ. Gynecomastia in a hospitalized male population. *Am J Med*. 1984;77:633–638.
24. Williams MJ, Gynecomastia. Its incidence, recognition and host characterization in 447 autopsy cases. *Am J Med*. 1963;34:103–112.
25. Braunstein GD. Clinical practice. Gynecomastia. *N Engl J Med*. 2007;357:1229–1237.
26. Mandava U, Kirma N, Tekmal RR. Aromatase overexpression transgenic mice model: cell type specific expression and use of letrozole to abrogate mammary hyperplasia without affecting normal physiology. *J Steroid Biochem Mol Biol*. 2001;79:27–34.
27. Li X, Warri A, Makela S, et al. Mammary gland development in transgenic male mice expressing human P450 aromatase. *Endocrinology*. 2002;143:4074–4083.
28. Andersen JA, Gram JB. Male breast at autopsy. *Acta Pathol Microbiol Immunol Scand A*. 1982;90:191–197.
29. Sandison AT. An autopsy study of the adult human breast: with special reference to proliferative epithelial changes of importance in the pathology of the breast. *J Natl Cancer Inst Monogr*. 1962;4:1–145.
30. Bannayan GA, Hajdu SI. Gynecomastia: clinicopathologic study of 351 cases. *Am J Clin Pathol*. 1972;57:431–437.
31. Nicolis GL, Modlinger RS, Gabrilove JL. A study of the histopathology of human gynecomastia. *J Clin Endocrinol Metab*. 1971;32:173–178.
32. Veldhuis JD, Dufau ML. Steroidal regulation of biologically active luteinizing hormone secretion in men and women. *Hum Reprod*. 1993;8:84–96.
33. Anderson MS, Brogi E, Biller BM. Occult Leydig cell tumor in a patient with gynecomastia. *Endocr Pract*. 2001;7:267–271.
34. Daniels IR, Layer GT. Testicular tumours presenting as gynaecomastia. *Eur J Surg Oncol*. 2003;29:437–439.
35. Klang E, Rozendorn N, Raskin S, et al. CT measurement of breast glandular tissue and its association with testicular cancer. *Eur Radiol*. 2017;27:536–542.
36. Polat AV, Ozturk M, Camlidag I, Akyuz B. Is gynecomastia related to the disease characteristics and prognosis in testicular germ cell tumor patients? *Diagn Interv Radiol*. 2019;25:189–194.
37. Kuhn JM, Mahoudeau JA, Billaud L, et al. Evaluation of diagnostic criteria for Leydig cell tumours in adult men revealed by gynaecomastia. *Clin Endocrinol (Oxf)*. 1987;26:407–416.
38. Whitcomb RW, Schimke RN, Kyner JL, et al. Endocrine studies in a male patient with choriocarcinoma and gynecomastia. *Am J Med*. 1986;81:917–920.
39. Foppiani L, Bernasconi D, Del MP, et al. Leydig cell tumour-induced bilateral gynaecomastia in a young man: endocrine abnormalities. *Andrologia*. 2005;37:36–39.
40. Sugimoto K, Matsumoto S, Nose K, et al. A malignant Leydig cell tumor of the testis. *Int Urol Nephrol*. 2006;38:291–292.
41. Gabrilove JL, Nicolis GL, Mitty HA, et al. Feminizing interstitial cell tumor of the testis: personal observations and a review of the literature. *Cancer*. 1975;35:1184–1202.
42. Hendry WS, Garvie WH, Ah-See AK, et al. Ultrasonic detection of occult testicular neoplasms in patients with gynaecomastia. *Br J Radiol*. 1984;57:571–572.
43. Valladares LE, Payne AH. Acute stimulation of aromatization in Leydig cells by human chorionic gonadotropin in vitro. *Proc Natl Acad Sci USA*. 1979;76:4460–4463.
44. Cuttler L, Rosenfield RL, Ehrmann DA, et al. Maturation of gonadotropin and sex steroid responses to gonadotropin-releasing hormone agonist in males. *J Clin Endocrinol Metab*. 1993;76:362–366.
45. Kidd MT, Karlin NJ, Cook CB. Feminizing adrenal neoplasms: case presentations and review of the literature. *J Clin Oncol*. 2011;29:e127–e130.
46. Kuhn JM, Lefebvre H, Duparc C, et al. Cosecretion of estrogen and inhibin B by a feminizing adrenocortical adenoma: impact on gonadotropin secretion. *J Clin Endocrinol Metab*. 2002;87:2367–2375.
47. Bouraima H, Lireux B, Mittre H, et al. Major hyperestrogenism in a feminizing adrenocortical adenoma despite a moderate overexpression of the aromatase enzyme. *Eur J Endocrinol*. 2003;148:457–461.
48. Morio H, Terano T, Yamamoto K, et al. Serum concentrations of dehydroepiandrosterone sulfate in patients with asymptomatic cortisol producing adrenal adenoma: comparison with adrenal Cushing’s syndrome and nonfunctional adrenal tumor. *Endocr J*. 1996;43:387–396.
49. Schneider G, Kirschner MA, Berkowitz R, et al. Increased estrogen production in obese men. *J Clin Endocrinol Metab*. 1979;48:633–638.
50. Schulte DM, Hahn M, Oberhäuser F, et al. Caloric restriction increases serum testosterone concentrations in obese male subjects by two distinct mechanisms. *Horm Metab Res*. 2014;46:283–286.
51. Southren AL, Olivo J, Gordon GG, et al. The conversion of androgens to estrogens in hyperthyroidism. *J Clin Endocrinol Metab*. 1974;38:207–214.
52. Cleland WH, Mendelson CR, Simpson ER. Effects of aging and obesity on aromatase activity of human adipose cells. *J Clin Endocrinol Metab*. 1985;60:174–177.
53. Feldman HA, Longcope C, Derby CA, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *J Clin Endocrinol Metab*. 2002;87:589–598.
54. Stratakis CA. An aroma of complexity: how the unique genetics of aromatase (CYP19A1) explain diverse phenotypes from hens and hyenas to human gynecomastia, and testicular and other tumors. *J Clin Endocrinol Metab*. 2013;98:4676–4681.
55. Shozu M, Fukami M, Ogata T. Understanding the pathological manifestations of aromatase excess syndrome: lessons for clinical diagnosis. *Expert Rev Endocrinol Metab*. 2014;9:397–409.
56. Coen P, Kulin H, Ballantine T, et al. An aromatase-producing sex-cord tumor resulting in prepubertal gynecomastia. *N Engl J Med*. 1991;324:317–322.
57. Hertl MC, Wiebel J, Schafer H, et al. Feminizing Sertoli cell tumors associated with Peutz-Jeghers syndrome: an increasingly recognized cause of prepubertal gynecomastia. *Plast Reconstr Surg*. 1998;102:1151–1157.

58. Lefevre H, Bouvattier C, Lahlou N, et al. Prepubertal gynecomastia in Peutz-Jeghers syndrome: incomplete penetrance in a familial case and management with an aromatase inhibitor. *Eur J Endocrinol*. 2006;154:221–227.
59. Ulbright TM, Amin MB, Young RH. Intratubular large cell hyalinizing Sertoli cell neoplasia of the testis: a report of 8 cases of a distinctive lesion of the Peutz-Jeghers syndrome. *Am J Surg Pathol*. 2007;31:827–835.
60. Gourgari E, Saloustros E, Stratakis CA. Large-cell calcifying Sertoli cell tumors of the testes in pediatrics. *Curr Opin Pediatr*. 2012;24:518–522.
61. Brown B, Ram A, Clayton P, et al. Conservative management of bilateral Sertoli cell tumors of the testicle in association with the Carney complex: a case report. *J Pediatr Surg*. 2007;42:E13–E15.
62. Templ H, Kuzmits R, Kratzik C. Microsomal aromatase in testicular tumour tissue: increased activity in choriocarcinoma compared to embryonal carcinoma. *Horm Metab Res*. 1988;20:522–524.
63. Agarwal VR, Takayama K, Van Wyk JJ, et al. Molecular basis of severe gynecomastia associated with aromatase expression in a fibrolamellar hepatocellular carcinoma. *J Clin Endocrinol Metab*. 1998;83:1797–1800.
64. Hsiao HP, Chao MC, Lin CY, et al. Feminizing adrenocortical adenoma presenting as heterosexual precocious puberty: report of one case. *Acta Paediatr Taiwan*. 2005;46:97–100. [Erratum appears in *Acta Paediatr Taiwan*. 2005;46:392.
65. Bulard J, Mowszowicz I, Schaison G. Increased aromatase activity in pubic skin fibroblasts from patients with isolated gynecomastia. *J Clin Endocrinol Metab*. 1987;64:618–623.
66. Sasano H, Kimura M, Shizawa S, et al. Aromatase and steroid receptors in gynecomastia and male breast carcinoma: an immunohistochemical study. *J Clin Endocrinol Metab*. 1996;81:3063–3067.
67. Czajka-Oraniec I, Zgliczynski W, Kurylowicz A, et al. Association between gynecomastia and aromatase (CYP19) polymorphisms. *Eur J Endocrinol*. 2008;158:721–727.
68. Irahara N, Miyoshi Y, Taguchi T, et al. Possible involvement of aromatase overexpression induced by cyclo-oxygenase-2 in the pathogenesis of idiopathic gynecomastia. *Endocr Res*. 2005;31:219–227.
69. Maclaren NK, Migeon CJ, Raiti S. Gynecomastia with congenital virilizing adrenal hyperplasia (11-betahydroxylase deficiency). *J Pediatr*. 1975;86:579–581.
70. Kadair RG, Block MB, Katz FH, et al. "Masked" 21-hydroxylase deficiency of the adrenal presenting with gynecomastia and bilateral testicular masses. *Am J Med*. 1977;62:278–282.
71. Frank-Raue K, Raue F, Korth-Schutz S, et al. Clinical features and diagnosis of mild 3-beta-hydroxysteroid dehydrogenase deficiency in men. *Dtsch Med Wochenschr*. 1989;114:331–334.
72. Hendrickson DA, Anderson WR. Diethylstilbestrol therapy gynecomastia. *J Am Med Assoc*. 1970;213:468.
73. Henricks DM, Gray SL, Hoover JL. Residue concentrations of endogenous estrogens in beef tissues. *J Anim Sci*. 1983;57:247–255.
74. Harrington JM, Stein GF, Rivera RO, et al. The occupational hazards of formulating oral contraceptives—a survey of plant employees. *Arch Environ Health*. 1978;33:12–15.
75. Rifka SM, Pita JC, Vigersky RA, et al. Interaction of digitalis and spironolactone with human sex steroid receptors. *J Clin Endocrinol Metab*. 1978;46:338–344.
76. Finkelstein JS, McCully WF, MacLaughlin DT, et al. The mortician's mystery. Gynecomastia and reversible hypogonadotropic hypogonadism in an embalmer. *N Engl J Med*. 1988;318:961–965.
77. DiRaimondo CV, Roach AC, Meador CK. Gynecomastia from exposure to vaginal estrogen cream. *N Engl J Med*. 1980;302:1089–1090.
78. Felner EI, White PC. Prepubertal gynecomastia: indirect exposure to estrogen cream. *Pediatrics*. 2000;105:E55.
79. Wu FC, Farley TM, Peregoudov A, et al. Effects of testosterone enanthate in normal men: experience from a multicenter contraceptive efficacy study. World Health Organization Task Force on Methods for the Regulation of Male Fertility. *Fertil Steril*. 1996;65:626–636.
80. Babigian A, Silverman RT. Management of gynecomastia due to use of anabolic steroids in bodybuilders. *Plast Reconstr Surg*. 2001;107:240–242.
81. Kanakis GA, Nieschlag E. Klinefelter syndrome: more than hypogonadism. *Metabolism.* 2018;86:135–144.
82. Basaria S. Male hypogonadism. *Lancet*. 2014;383:1250–1263.
83. Bhasin S, Gagliano-Jucá T, Huang G, Basaria S. Age-Related Changes in the Male Reproductive System. 2022 February 5. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, et al., editors. *Endotext* [Internet]. South Dartmouth, MA: MDText.com, Inc.; 2000.
84. Jensen RT, Collen MJ, Pandol SJ, et al. Cimetidine-induced impotence and breast changes in patients with gastric hypersecretory states. *N Engl J Med*. 1983;308:883–887.
85. Caminos-Torres R, Ma L, Snyder PJ. Gynecomastia and semen abnormalities induced by spironolactone in normal men. *J Clin Endocrinol Metab*. 1977;45:255–260.
86. Deepinder F, Braunstein GD. Drug-induced gynecomastia: an evidence-based review. *Expert Opin Drug Saf*. 2012;11:779–795.
87. Brinkmann AO. Molecular basis of androgen insensitivity. *Mol Cell Endocrinol*. 2001;179:105–109.
88. Grino PB, Griffin JE, Cushard Jr WG, et al. A mutation of the androgen receptor associated with partial androgen resistance, familial gynecomastia, and fertility. *J Clin Endocrinol Metab*. 1988;66:754–761.
89. Hellmann P, Christiansen P, Johannsen TH, et al. Male patients with partial androgen insensitivity syndrome: a longitudinal follow-up of growth, reproductive hormones and the development of gynaecomastia. *Arch Dis Child*. 2012;97:403.
90. Abel A, Danek A, Borasio GD, et al. X-chromosomal bulbospinal neuropathy (X-BSN, Kennedy syndrome): an illness with repetitive triplet sequences. Case report, differential diagnosis and molecular genetics aspects. *Nervenarzt*. 1996;67:1011–1019.
91. Berkhoff M, Sturzenegger M, Spiegel R, et al. X-chromosomal bulbospinal muscular atrophy (Kennedy syndrome). *Schweiz Med Wochenschr*. 1998;128:817–823.
92. Baker HW, Burger HG, de Kretser DM, et al. A study of the endocrine manifestations of hepatic cirrhosis. *Q J Med*. 1976;45:145–178.
93. Braunstein GD. Editorial comment: unraveling the cause of HIV-related gynecomastia. *AIDS Read*. 2004;14:38–39.
94. Biglia A, Blanco JL, Martinez E, et al. Gynecomastia among HIV-infected patients is associated with hypogonadism: a case-control study. *Clin Infect Dis*. 2004;39:1514–1519.
95. Deonarain J, Ramdial PK, Singh B. Bilateral lipomastia in men: a side effect of highly active antiretroviral therapy. *Int J Surg Pathol*. 2008;16:171–175.
96. Gewurz BE, Dezube BJ, Pantanowitz L. HIV and the breast. *AIDS Read*. 2005;15:392–396.
97. Mira JA, Lozano F, Santos J, et al. Gynaecomastia in HIV-infected men on highly active antiretroviral therapy: association with efavirenz and didanosine treatment. *Antivir Ther*. 2004;9:511–517.
98. Abalovich M, Levalle O, Hermes R, et al. Hypothalamic-pituitary-testicular axis and seminal parameters in hyperthyroid males. *Thyroid*. 1999;9:857–863.
99. Becker KL, Winnacker JL, Matthews MJ, et al. Gynecomastia and hyperthyroidism. An endocrine and histological investigation. *J Clin Endocrinol Metab*. 1968;28:277–285.
100. Tan YK, Birch CR, Valerio D. Bilateral gynaecomastia as the primary complaint in hyperthyroidism. *J R Coll Surg Edinb*. 2001;46:176–177.
101. Mean F, Pellaton M, Magrini G. Study on the binding of dihydrotestosterone, testosterone and oestradiol with sex hormone binding globulin. *Clin Chim Acta*. 1977;80:171–180.
102. Chan WB, Yeung VT, Chow CC, et al. Gynaecomastia as a presenting feature of thyrotoxicosis. *Postgrad Med J*. 1999;75:229–231.
103. Iranmanesh A, Lizarralde G, Veldhuis JD. Robustness of the male lactotropic axis to the hyperprolactinemic stimulus of primary thyroidal failure. *J Clin Endocrinol Metab*. 1992;74:559–564.
104. Mertani HC, Garcia-Caballero T, Lambert A, et al. Cellular expression of growth hormone and prolactin receptors in human breast disorders. *Int J Cancer*. 1998;79:202–211.
105. Mieritz M, Sorensen K, Aksglaede L, et al. Elevated serum IgF-1, but normal sex steroid levels in healthy boys with pubertal gynaecomastia. *Clin Endocrinol*. 2014;80:691–698.
106. Acharya SV, Gopal RA, Bandgar TR, et al. Prepubertal gynecomastia a rare complication of growth hormone therapy. *Indian J Pediatr*. 2010;77:443–444.
107. Malozowski S, Stadel BV. Prepubertal gynecomastia during growth hormone therapy. *J Pediatr*. 1995;126:659–661.

108. Feldman D. Ketoconazole and other imidazole derivatives as inhibitors of steroidogenesis. *Endocr Rev*. 1986;7:409–420.
109. De Vecchis R, Cantatrione C, Mazzei D, et al. The impact exerted on clinical outcomes of patients with chronic heart failure by aldosterone receptor antagonists: a meta-analysis of randomized controlled trials. *J Clin Med Res*. 2017;9:130.
110. Hagberg KW, Divan HA, Fang SC, et al. Risk of gynecomastia and breast cancer associated with the use of 5-alpha reductase inhibitors for benign prostatic hyperplasia. *Clin Epidemiol*. 2017;9:83–91.
111. Geller J, Vazakas G, Fruchtman B, et al. The effect of cyproterone acetate on advanced carcinoma of the prostate. *Surg Gynecol Obstet*. 1968;127:748–758.
112. Harmon JW, Aliapoulios MA. Marijuana-induced gynecomastia: clinical and laboratory experience. *Surg Forum*. 1974;25:423–425.
113. Purohit V, Ahluwahlia BS, Vigersky RA. Marihuana inhibits dihydrotestosterone binding to the androgen receptor. *Endocrinology*. 1980;107:848–850.
114. Cates Jr W, Pope JN. Gynecomastia and cannabis smoking. A nonassociation among US Army soldiers. *Am J Surg*. 1977;134:613–615.
115. Boccardo F, Rubagotti A, Conti G, et al. Exploratory study of drug plasma concentrations during bicalutamide 150 mg therapy co-administered with tamoxifen or anastrozole for prophylaxis of gynecomastia and breast pain in men with prostate cancer. *Cancer Chemother Pharmacol*. 2005;56:415–420.
116. Boccardo F, Rubagotti A, Battaglia M, et al. Evaluation of tamoxifen and anastrozole in the prevention of gynecomastia and breast pain induced by bicalutamide monotherapy of prostate cancer. *J Clin Oncol*. 2005;23:808–815.
117. Dobs A, Darkes MJ. Incidence and management of gynecomastia in men treated for prostate cancer. *J Urol*. 2005;174:1737–1742.
118. Aki FT, Tekin MI, Ozen H. Gynecomastia as a complication of chemotherapy for testicular germ cell tumors. *Urology*. 1996;48:944–946.
119. Uygur MC, Ozen H. Gynecomastia following chemotherapy for testicular cancer. *Urol Int*. 2003;70:253–254.
120. Saeter G, Fosså SD, Norman N. Gynaecomastia following cytotoxic therapy for testicular cancer. *Br J Urol*. 1987;59:348–352.
121. Trump DL, Pavy MD, Staal S. Gynecomastia in men following antineoplastic therapy. *Arch Intern Med*. 1982;142:511–513.
122. Chatterjee R, Kottaridis PD, McGarrigle HH, et al. Patterns of Leydig cell insufficiency in adult males following bone marrow transplantation for hematological malignancies. *Bone Marrow Transplant*. 2001;28:497–502.
123. Allen RC, Wallace AM, Royce M. Marinol-induced gynecomastia: a case report. *Am J Med*. 2007;120:e1.
124. Thompson DF, Carter JR. Drug-induced gynecomastia. *Pharmacotherapy*. 1993;13:37–45.
125. Colin JF. Gynaecomastia: drugs and surgical concerns. *BMJ*. 2008;336:790.
126. Desgrandchamps F, Droupy S, Irani J, et al. Effect of dutasteride on the symptoms of benign prostatic hyperplasia, and patient quality of life and discomfort, in clinical practice. *BJU Int*. 2006;98:83–88.
127. Gardette V, Vezzosi D, Maiza JC, et al. Gynecomastia associated with fenofibrate. *Ann Pharmacother*. 2007;41:508–510.
128. Hammons KB, Edwards RF, Rice WY. Golf-inhibiting gynecomastia associated with atorvastatin therapy. *Pharmacotherapy*. 2006;26:1165–1168.
129. Henley DV, Lipson N, Korach KS, et al. Prepubertal gynecomastia linked to lavender and tea tree oils. *N Engl J Med*. 2007;356:479–485.
130. Iaria G, Urbani L, Catalano G, et al. Switch to tacrolimus for cyclosporine-induced gynecomastia in liver transplant recipients. *Transplant Proc*. 2005;37:2632–2633.
131. Jelenkovic AV, Ukanovic-Golubovic LD. Diazepam-associated gynecomastia. *Ann Pharmacother*. 2005;39:201.
132. Malaga I, Sanmarti FX. Two cases of painful gynecomastia and lower extremity pain in association with pregabalin therapy. *Epilepsia*. 2006;47:1576–1579.
133. Mendhekar DN. Clonidine-induced gynecomastia and hyperprolactinemia in a 6-year-old child. *J Clin Psychiatry*. 2005;66:1616–1617.
134. Mendhekar DN, Andrade CR. Unilateral gynecomastia induced by risperidone in a geriatric male patient. *Indian J Med Sci*. 2005;59:361–362.
135. Mendhekar DN, Srilakshmi P. Olanzapine-induced gynecomastia. *Aust N Z J Psychiatry*. 2005;39:736.
136. Oteri A, Catania MA, Travaglini R, et al. Gynecomastia possibly induced by rosuvastatin. *Pharmacotherapy*. 2008;28:549–551.
137. Pandhi D, Gupta R, Singal A. Gynaecomastia with oligospermia: an unusual complication of low-dose methotrexate for pustular psoriasis. *Clin Exp Dermatol*. 2006;31:138–140.
138. Sahin M, Yilmaz H, Guvener ND. A possible case of gynecomastia with fluoxetine. *Ann Pharmacother*. 2005;39:1369.
139. Satoh T, Tomikawa Y, Takanashi K, et al. Studies on the interactions between drugs and estrogen. III. Inhibitory effects of 29 drugs reported to induce gynecomastia on the glucuronidation of estradiol. *Biol Pharm Bull*. 2004;27:1844–1849.
140. Kakisaka Y, Ohara T, Tozawa H, et al. Panax ginseng: a newly identified cause of gynecomastia. *Tohoku J Exp Med*. 2012;228:143–145.
141. Braunstein GD. Environmental gynecomastia. *Endocr Pract*. 2008;14:409–411.
142. Greene WW, Howard NJ. Relation of trauma to lesions of the male breast. *Am J Surg*. 1953;85:431–437.
143. Field JB, Solis RT, Dear WE. Unilateral gynecomastia associated with thoracotomy following resection of carcinoma of the lung. *Am J Med Sci*. 1989;298:402–406.
144. Heruti RJ, Dankner R, Berezin M, et al. Gynecomastia following spinal cord disorder. *Arch Phys Med Rehabil*. 1997;78:534–537.
145. Biju G, Taly AB, Sinha S, et al. Painful gynecomastia following traumatic spinal cord injury. *Clin Neurol Neurosurg*. 2005;107:442–443.
146. Mor G, Yue W, Santen RJ, et al. Macrophages, estrogen and the microenvironment of breast cancer. *J Steroid Biochem Mol Biol*. 1998;67:403–411.
147. Morley JE, Distiller LA, Sagel J, et al. Hormonal changes associated with testicular atrophy and gynaecomastia in patients with leprosy. *Clin Endocrinol (Oxf)*. 1977;6:299–303.
148. Zurbiran S, Gomez-Mont F. Endocrine disturbances in chronic human malnutrition. *Vitam Horm*. 1953;11:97–132.
149. Hunfeld KP, Bassler R, Kronsbein H. "Diabetic mastopathy" in the male breast—a special type of gynecomastia. A comparative study of lymphocytic mastitis and gynecomastia. *Pathol Res Pract*. 1997;193:197–205.
150. Toma E, Therrien R. Gynecomastia during indinavir antiretroviral therapy in HIV infection. *AIDS*. 1998;12:681–682.
151. Benveniste O, Simon A, Herson S. Successful percutaneous dihydrotestosterone treatment of gynecomastia occurring during highly active antiretroviral therapy: four cases and a review of the literature. *Clin Infect Dis*. 2001;33:891–893.
152. Lazala C, Saenger P. Pubertal gynecomastia. *J Pediatr Endocrinol Metab*. 2002;15:553–560.
153. Sher ES, Migeon CJ, Berkovitz GD. Evaluation of boys with marked breast development at puberty. *Clin Paediatr (Phila)*. 1998;37:367–371.
154. Ersoz H, Onde ME, Terekeci H, et al. Causes of gynaecomastia in young adult males and factors associated with idiopathic gynaecomastia. *Int J Androl*. 2002;25:312–316.
155. Brinton LA. Breast cancer risk among patients with Klinefelter syndrome. *Acta Paediatr*. 2011;100:814–818.
156. Jaiswal AK, Subbarao NT. Gynaecothelia—a common yet ignored sign of multibacillary leprosy in males: a case series with review of literature. *Lepr Rev*. 2012;83:384–388.
157. Kohli K, Jain S. Filariasis presenting as gynecomastia. *Breast J*. 2012;18:83–84.
158. Hines SL, Tan WW, Yasrebi M, et al. The role of mammography in male patients with breast symptoms. *Mayo Clin Proc*. 2007;82:297–300.
159. Dershaw DD. Male mammography. *Am J Roentgenol*. 1986;146:127–131.
160. Rissanen TJ, Makarainen HP, Kallioinen MJ, et al. Radiography of the male breast in gynecomastia. *Acta Radiol*. 1992;33:110–114.
161. Appelbaum AH, Evans GF, Levy KR, et al. Mammographic appearances of male breast disease. *Radiographics*. 1999;19:559–568.
162. Dialani V, Baum J, Mehta TS. Sonographic features of gynecomastia. *J Ultrasound Med*. 2010;29:539–547.
163. Gunhan-Bilgen I, Bozkaya H, Ustun EE, et al. Male breast disease: clinical, mammographic, and ultrasonographic features. *Eur J Radiol*. 2002;43:246–255.

164. Weinstein SP, Conant EF, Orel SG, et al. Spectrum of US findings in pediatric and adolescent patients with palpable breast masses. *Radiographics*. 2000;20:1613–1621.
165. Evans GF, Anthony T, Turnage RH, et al. The diagnostic accuracy of mammography in the evaluation of male breast disease. *Am J Surg*. 2001;181:96–100. [Erratum appears in *Am J Surg* 2001;181:579].
166. Skeldon SC, Carleton B, Brophy JM, et al. Statin medications and the risk of gynecomastia. *Clin Endocrinol (Oxf)*. 2018;89:470.
167. Ramsey JT, Li Y, Arao Y, et al. Lavender products associated with premature thelarche and prepubertal gynecomastia: case reports and endocrine-disrupting chemical activities. *J Clin Endocrinol Metab*. 2019;104:5393–5405.
168. Coviello AD, Matsumoto AM, Bremner WJ, et al. Low-dose human chorionic gonadotropin maintains intratesticular testosterone in normal men with testosterone-induced gonadotropin suppression. *J Clin Endocrinol Metab*. 2005;90:2595–2602.
169. McLachlan RI, de Kretser DM. Hypogonadotropism with elevated serum testosterone: reversible causes of secondary infertility. *Nat Clin Pract Urol*. 2006;3:560–565.
170. Jackson VP, Gilmor RL. Male breast carcinoma and gynecomastia: comparison of mammography with sonography. *Radiology*. 1983;149:533–536.
171. Westenend PJ, Jobse C. Evaluation of fine-needle aspiration cytology of breast masses in males. *Cancer*. 2002;96:101–104.
172. Olsson H, Bladstrom A, Alm P. Male gynecomastia and risk for malignant tumours–a cohort study. *BMC Cancer*. 2002;2:26.
173. Kapila K, Verma K. Cytomorphological spectrum in gynaecomastia: a study of 389 cases. *Cytopathology*. 2002;13:300–308.
174. Volpe CM, Raffetto JD, Collure DW, et al. Unilateral male breast masses: cancer risk and their evaluation and management. *Am Surg*. 1999;65:250–253.
175. van Geel AN, van Slooten EA, Mavrunac M, et al. A retrospective study of male breast cancer in Holland. *Br J Surg*. 1985;72:724–727.
176. Chen PH, Slanetz PJ. Incremental clinical value of ultrasound in men with mammographically confirmed gynecomastia. *Eur J Radiol*. 2014;83:123–129.
177. Plourde PV, Reiter EO, Jou HC, et al. Safety and efficacy of anastrozole for the treatment of pubertal gynecomastia: a randomized, double-blind, placebo-controlled trial. *J Clin Endocrinol Metab*. 2004;89:4428–4433.
178. Mauras N, Bishop K, Merinbaum D, et al. Pharmacokinetics and pharmacodynamics of anastrozole in pubertal boys with recent onset gynecomastia. *J Clin Endocrinol Metab*. 2009;94:2975–2978.
179. Zachmann M, Eiholzer U, Muritano M, et al. Treatment of pubertal gynaecomastia with testolactone. *Acta Endocrinol (Copenh)*. 1986;279:218–226.
180. Rhoden EL, Morgentaler A. Treatment of testosterone-induced gynecomastia with the aromatase inhibitor, anastrozole. *Int J Impot Res*. 2004;16:95–97.
181. Riepe FG, Baus I, Wiest S, et al. Treatment of pubertal gynecomastia with the specific aromatase inhibitor anastrozole. *Horm Res*. 2004;62:113–118.
182. McDermott MT, Hofeldt FD, Kidd GS. Tamoxifen therapy for painful idiopathic gynecomastia. *South Med J*. 1990;83:1283–1285.
183. Parker LN, Gray DR, Lai MK, et al. Treatment of gynecomastia with tamoxifen: a double-blind crossover study. *Metabolism*. 1986;35:705–708.
184. Fairlamb D, Boesen E. Gynaecomastia associated with gonadotrophin-secreting carcinoma of the lung. *Postgrad Med J*. 1977;53:269–271.
185. Jefferys DB. Painful gynaecomastia treated with tamoxifen. *Br Med J*. 1979;1:1119–1120.
186. Eversmann T, Moito J, VonWerder K. Testosterone and estradiol concentrations in male gynecomastia. Clinical and endocrine findings during treatment with tamoxifen. *Dtsch Med Wochenschr*. 1984;109:1678–1682.
187. Stepanas AV, Burnet RB, Harding PE, et al. Clomiphene in the treatment of pubertal-adolescent gynecomastia: a preliminary report. *J Pediatr*. 1977;90:651–653.
188. Plourde PV, Kulin HE, Santner SJ. Clomiphene in the treatment of adolescent gynecomastia. Clinical and endocrine studies. *Am J Dis Child*. 1983;137:1080–1082.
189. LeRoith D, Sobel R, Glick SM. The effect of clomiphene citrate on pubertal gynaecomastia. *Acta Endocrinol (Copenh)*. 1980;95:177–180.
190. Derman O, Kanbur NO, Kutluk T. Tamoxifen treatment for pubertal gynecomastia. *Int J Adolesc Med Health*. 2003;15:359–363.
191. Hanavadi S, Banerjee D, Monypenny IJ, et al. The role of tamoxifen in the management of gynaecomastia. *Breast*. 2006;15:276–280.
192. Konig R, Schonberger W, Neumann P, et al. Treatment of marked gynecomastia in puberty with tamoxifen. *Klin Pädiatr*. 1987;199:389–391.
193. Lawrence SE, Faught KA, Vethamuthu J, et al. Beneficial effects of raloxifene and tamoxifen in the treatment of pubertal gynecomastia. *J Pediatr*. 2004;145:71–76.
194. Khan HN, Rampaul R, Blamey RW. Management of physiological gynaecomastia with tamoxifen. *Breast*. 2004;13:61–65.
195. Lapid O, van Wingerden J, Perlemuter L. Tamoxifen therapy for the management of pubertal gynaecomastia: a systematic review. *J Pediatr Endocrinol Metab*. 2013;26:803–807.
196. Karagiannis A, Tziomalos K, Kakafika A, et al. Eplerenone relieves spironolactone-induced painful gynaecomastia in a patient with primary aldosteronism. *Nephrol Dial Transplant*. 2007;22:293.
197. Mimidis K, Papadopoulos V, Kartalis G. Eplerenone relieves spironolactone-induced painful gynaecomastia in patients with decompensated hepatitis B-related cirrhosis. *Scand J Gastroenterol*. 2007;42:1516–1517.
198. Jones DJ, Holt SD, Surtees P, et al. A comparison of danazol and placebo in the treatment of adult idiopathic gynaecomastia: results of a prospective study in 55 patients. *Ann R Coll Surg Engl*. 1990;72:296–298.
199. Handschin AE, Bietry D, Husler R, et al. Surgical management of gynecomastia—a 10-year analysis. *World J Surg*. 2008;32:38–44.
200. Lista F, Ahmad J. Power-assisted liposuction and the pull-through technique for the treatment of gynecomastia. *Plast Reconstr Surg*. 2008;121:740–747.
201. Mentz HA, Ruiz-Razura A, Newall G, et al. Correction of gynecomastia through a single puncture incision. *Aesthetic Plast Surg*. 2007;31:244–249.
202. Tashkandi M, Al-Qattan MM, Hassanain JM, et al. The surgical management of high-grade gynecomastia. *Ann Plast Surg*. 2004;53:17–20.
203. Serels S, Melman A. Tamoxifen as treatment for gynecomastia and mastodynia resulting from hormonal deprivation. *J Urol*. 1998;159:1309.
204. Saltzstein D, Sieber P, Morris T, et al. Prevention and management of bicalutamide-induced gynecomastia and breast pain: randomized endocrinologic and clinical studies with tamoxifen and anastrozole. *Prostate Cancer Prostatic Dis*. 2005;8:75–83.
205. Sieber PR. Treatment of bicalutamide-induced breast events. *Expert Rev Anticancer Ther*. 2007;7:1773–1779.
206. Fradet Y, Egerdie B, Andersen M, et al. Tamoxifen as prophylaxis for prevention of gynaecomastia and breast pain associated with bicalutamide 150 mg monotherapy in patients with prostate cancer: a randomised, placebo-controlled, dose-response study. *Eur Urol*. 2007;52:106–114.
207. Nuttall MC, Harris JP, Dawkins GP. The role of tamoxifen in reducing bicalutamide-induced gynaecomastia and breast pain. *BJU Int*. 2007;99:243–244.
208. Widmark A, Fossa SD, Lundmo P, et al. Does prophylactic breast irradiation prevent antiandrogen-induced gynecomastia? Evaluation of 253 patients in the randomized Scandinavian trial SPCG-7/SFUO-3. *Urology*. 2003;61:145–151.
209. Dicker AP. The safety and tolerability of low-dose irradiation for the management of gynaecomastia caused by antiandrogen monotherapy. *Lancet Oncol*. 2003;4:30–36.
210. Tyrrell CJ, Payne H, Tammela TL, et al. Prophylactic breast irradiation with a single dose of electron beam radiotherapy (10 Gy) significantly reduces the incidence of bicalutamide-induced gynecomastia. *Int J Radiat Oncol Biol Phys*. 2004;60:476–483.
211. Di Lorenzo G, Perdonà S, De Placido S, et al. Gynecomastia and breast pain induced by adjuvant therapy with bicalutamide after radical prostatectomy in patients with prostate cancer: the role of tamoxifen and radiotherapy. *J Urol*. 2005;174:2197–2203.

212. Viani GA, Bernardes da Silva LGS. Prevention of gynecomastia and breast pain caused by androgen deprivation therapy in prostate cancer: tamoxifen or radiotherapy? *Int J Radiat Oncol Biol Phys*. 2012;83:e519–e524.
213. Perdona S, Autorino R, DePlacido S, et al. Efficacy of tamoxifen and radiotherapy for prevention and treatment of gynaecomastia and breast pain caused by bicalutamide in prostate cancer: a randomised controlled trial. *Lancet Oncol*. 2005;6:295–300.
214. Serretta V, Altieri V, Morgia G, et al. A randomized trial comparing tamoxifen therapy vs. tamoxifen prophylaxis in bicalutamide-induced gynecomastia. *Clin Genitourin Cancer*. 2012;10:174–179.
215. Tunio MA, Al-Asiri M, Al-Amro S, et al. Optimal prophylactic and definitive therapy for bicalutamide-induced gynecomastia: results of a meta-analysis. *Curr Oncol*. 2012;19:280–288.
216. Fagerlund A, Cormio L, Palangi L, et al. Gynecomastia in patients with prostate cancer: a systematic review. *PLoS One*. 2015;10:e0136094.

109

Erectile Dysfunction

Amanda B. Reed-Maldonado, Alan W. Shindel, and Tom F. Lue

OUTLINE

Introduction, 1825
Guidelines, 1825
Epidemiology, 1826
- Classification of Erectile Dysfunction Based on Pathogenesis, 1826
- Psychogenic Erectile Dysfunction, 1826
- Organic Erectile Dysfunction, 1826

Penile Physiology, 1828
- Anatomy of the Penis, 1828
- Hemodynamics of Penile Erection, 1828
- Innervation of the Penis, 1829
- Neurotransmitters and Pharmacology of Erection, 1829
- Testosterone and Male Sexual Function, 1829

Clinical Evaluation of the Man with Erectile Dysfunction, 1830
- Discussing Sexual Health with Patients, 1830
- History, 1831
- Physical Examination, 1831
- Laboratory Testing, 1831
- Specialized Testing for Erectile Dysfunction, 1831

Treatment of Erectile Dysfunction, 1833
- Adaptive Sexuality, 1833
- Lifestyle Modification, 1833
- Mental Health, 1833
- Vacuum Erection Device, 1834

Medical Therapy for Erectile Dysfunction, 1834
- Herbal Supplements and Complementary Medicine, 1834
- Phosphodiesterase Type 5 Inhibitors, 1835
- Intracavernosal Injection Therapy, 1836
- Urethral Prostaglandin Suppositories, 1837

Surgical Therapies for Erectile Dysfunction, 1837
- Malleable Penile Prostheses, 1837
- Inflatable Penile Prostheses, 1837
- Penile Vascular Surgery, 1838

Additional Conditions Contributing to Male Sexual Dysfunction, 1839
- Disorders of Ejaculation, Emission, and Orgasm, 1839
- Peyronie Disease, 1839
- Priapism, 1840

Future Directions for the Treatment of Erectile Dysfunction, 1841
- New Pharmacotherapeutics, 1841
- Stem Cell Therapy, 1841
- Low-Intensity Extracorporeal Shock Wave Therapy, 1841
- Platelet-Rich Plasma, 1842

Conclusions, 1842

INTRODUCTION

Erectile dysfunction (ED) is a pervasive disorder defined as the persistent inability to achieve and/or sustain penile erection adequate for sexual satisfaction.[1] This definition from the International Consultation on Sexual Medicine incorporates not only the mechanics of obtaining and maintaining an erection but also critical subjective factors, including overall sexual satisfaction, bother, and interpersonal dimensions, which are critical determinants of clinically relevant ED. Although dysfunction of penile erection is the most widely recognized male sexual condition, other conditions, such as decreased libido, premature ejaculation (PE), delayed ejaculation (DE), and Peyronie disease (PD), may sometimes be comorbid with or confused for ED.

In this chapter, we will focus on the evaluation and management of the man presenting with difficulty attaining or maintaining penile erection. We will review penile anatomy, mechanisms of penile erection, and the pathogenesis of ED, in addition to reviewing diagnostic modalities and medical and surgical therapeutic options for the treatment of ED. Next, we will briefly review related male sexual health issues, including disorders of ejaculation and orgasm, PD, and priapism, which may be associated with ED but are less commonly addressed due to persistent social stigma. Finally, we will review future directions in the development of therapeutics for ED.

We hope to provide a reference for all clinicians regarding optimal management of ED and to improve the reader's understanding of the physiological basis of penile erection and how vascular, neuronal, hormonal, iatrogenic, and psychosocial factors operate in concert together to produce or inhibit penile erection.

GUIDELINES

When treating patients with ED, it is valuable to recognize that many professional medical societies have published recent guidelines regarding best practices related to the condition. Written by well-respected experts in the field of sexual medicine, these are living documents that continually evolve as medical knowledge expands and technology advances. The European Association of Urology has published a thorough set of male sexual health guidelines including disorders such as ED, ejaculatory disorders, decreased libido, and penile curvature.[2] The American Urological Association (AUA) has published the recent guidelines on ED, which is a current, evidenced-based document that describes algorithms for both the medical and surgical management of ED.[3] The AUA has also recently updated its guidelines on testosterone (T) therapy, another evidence-based document that provides direction on the related topic of androgen insufficiency.[4] The Endocrine Society also has recently published guidelines on the evaluation of hypogonadism and treatment with T therapy.[5] PE is another common male sexual concern,[6] and recent guidance from the International Society of Sexual Medicine provides an algorithm for management of this poorly understood yet disturbing condition.[6,7] Becoming acquainted with these and other evidence-based guidelines documents, which are available online to all clinicians, will not only facilitate patient encounters but also improve quality of care and reduce risk for male sexual dysfunction patients.

EPIDEMIOLOGY

In the Massachusetts Male Aging Study, a community-based survey of men between 40 and 70 years of age, 52% of respondents reported some degree of ED: 17% mild, 25% moderate, and 10% complete.[8] Although the prevalence of mild ED remained constant (17%) between the age of 40 and 70 years, there was a doubling in the number of men reporting moderate ED (from 17% to 34%) and a tripling in the number of men reporting complete ED (from 5% to 15%).[9] More than 70% of men over 65 years reported that they are sexually active. However, 40% were dissatisfied with their sexual function.

Among the major predictors of ED are hypertension, hyperlipidemia, diabetes mellitus, and heart disease. Risk of ED increases with smoking and appears to do so in a dose-dependent manner.[10] There is also a higher prevalence of ED in men who have undergone radiation or surgery for prostate cancer or other pelvic conditions such as malignancy or large bowel disease. The psychological correlates of ED include decreased self-esteem, depression, anxiety, anger, and relationship dissatisfaction.[11] Other male sexual dysfunctions, such as PE, DE, and decreased libido, have also been found to be highly prevalent and comorbid in men with ED.

Classification of Erectile Dysfunction Based on Pathogenesis

Penile erection is a multidimensional process that involves the complex interplay among psychological, organic, and relational components. When any of these components is altered, ED may result. Often, clinicians attempt to over simplify the condition by seeking to classify its etiologies into organic and psychogenic groups (Table 109.1). This unquestionably helps to identify a good starting point for treatment. Most commonly, however, "organic" ED involves a significant psychological component, as ED universally inflicts a negative effect on mood, interpersonal relations, and quality of life. In the 1950s, 90% of cases of ED were believed to be exclusively psychogenic in origin. Currently, most clinicians believe that mixed ED is most commonly encountered. The classification system of ED most frequently utilized for research encompasses organic, psychogenic, and mixed etiologies of ED and is endorsed by the International Society of Impotence Research (Table 109.2). This classification provides an excellent framework for beginning the clinical evaluation of a male with complaints of sexual dysfunction.

Psychogenic Erectile Dysfunction

Many psychological conditions can either trigger or aggravate ED. Sexual behavior and penile erections are controlled by the hypothalamus, cerebral cortex, and limbic systems. Given the number and complexity of known and as-yet-unidentified factors involved, it is not surprising that the pathogenesis of psychogenic ED is still under investigation. Possible mechanisms proposed include an imbalance of central neurotransmitters, overinhibition of spinal erection center by the brain, inadequate nitric oxide (NO) release, and sympathetic overactivity.[12]

TABLE 109.1 **Information Gained from Patient History that Helps to Distinguish Psychogenic from Organic Erectile Dysfunction, Which May Be Helpful in Determining the Most Useful Therapeutic Options for the Patient**

Psychogenic	• Sudden onset • Intermittent • Situational (erection may be intact in certain situations, such as with self-stimulation, but not in others, such as with a partner or specific partner) • Loss of ability to maintain erection • Nocturnal erections intact • Likely to respond well to PDE5 inhibitors if libido is intact
Organic	• Gradual onset • Progressive • Consistently poor erections despite situation • Difficulty with obtaining and maintaining erection

Organic Erectile Dysfunction

Sexual function gradually declines as men age, though this decline should not be considered any more acceptable than the development of diabetes, hypercholesterolemia, or hypertension, which also increase in incidence with age. It is physiologically normal for the aging male to anticipate the latency period between sexual stimulation and erection to increase, the refractory period between erections to increase, the erection to be less rigid, ejaculation to be less forceful, and ejaculate volume to decline.[8] Aging is also associated with a decrease in penile sensitivity to tactile stimulation, a gradual decrease in serum T concentration, and an increase in cavernous muscle tone.

ED is more common in patients with neurologic disorders, such as Parkinson or Alzheimer diseases, stroke, or traumatic brain injury, and it is estimated that up to 20% of all ED is neurogenic in origin. This may be due to both a decrease in libido and/or inability to initiate the erectile process. Spinal cord injury patients have varying degrees of ED chiefly determined by the location and extent of the spinal cord lesion. Men with sacral spine lesions may have disruption of the sacral reflex arc and may not respond to genital stimulation. Men with thoracic spinal lesions may retain reflexogenic erections but lose psychogenic erections, which are generated by suppression of sympathetic penile tone in response to erotic stimuli. Although erections of some form are often possible in men with spinal cord injury, they are typically less rigid and of shorter duration compared with those in men without spinal cord injury. Even in men without spinal cord injury, sensory input from the genitalia remains essential to achieving and maintaining reflexogenic erection. This input may be compromised due to peripheral neuropathy in men with diabetes and other chronic medical conditions such as renal insufficiency.[13]

Multiple chronic medical conditions contribute to the development and progression of ED. Chronic diabetes mellitus is highly associated with ED, in that it causes neurogenic, hormonal, and vascular consequences. In addition to its effect on small vessels, it may also affect the cavernous nerve terminals and endothelial cells, resulting in deficiency of neurotransmitters. Additionally, corporal smooth muscle relaxation in response to NO may also be impaired, possibly due to the accumulation of glycosylation products. Many diabetic males also have low serum T as a consequence of the disease. Chronic renal insufficiency is commonly associated with ED, impaired libido, and infertility. The mechanism is likely multifactorial and includes low T and decreased libido, vascular insufficiency, polypharmacy, and neuropathy.

TABLE 109.2 **The Erectile Dysfunction Classification System Recommended by the International Society of Impotence Research that was Created Due to the Need for Standardization of Definitions and Terms Used in the Diagnosis and Treatment of the Disorder and Later Modified to Include Specific Conditions That May be Associated with Each Classification**

Psychogenic

1. Generalized type
    - A. Generalized unresponsiveness
        - a. Primary lack of sexual arousability
        - b. Aging-related decline in sexual arousability
    - B. Generalized inhibition—chronic disorder of sexual intimacy
2. Situational type
    - A. Partner-related
        - a. Lack of arousability in specific relationship
        - b. Lack of arousability due to sexual object preference
        - c. High central inhibition due to partner conflict or threat
    - B. Performance-related
        - a. Associated with other sexual dysfunction/s (e.g., rapid ejaculation)
        - b. Situational performance anxiety (e.g., fear of failure)
    - C. Psychological distress or adjustment-related
        - a. Associated with negative mood state (anxiety, depression) or major life stress (death of partner, divorce, loss of employment, illness)

Organic

1. Neurogenic
    - Central nervous system: spinal cord injury, multiple sclerosis, cerebrovascular accident
    - Peripheral nervous system: neuropathy
    - Iatrogenic from surgery or radiation
2. Hormonal
    - Hypogonadism
    - Hyperprolactinemia
    - Diabetes mellitus
    - Thyroid disorders
3. Arterial
    - Hypertension
    - Dyslipidemia
    - Obesity and sedentary lifestyle
    - Renal insufficiency
    - Tobacco smoking
    - Chronic obstructive pulmonary disease
    - Sleep apnea
    - Coronary artery or peripheral arterial disease
    - Focal arterial occlusion after trauma
4. Cavernosal
    - Venoocclusive disease
    - Age-related muscle atrophy
    - Cavernosal fibrosis of the penis
    - Peyronie disease
    - Penile fracture
5. Drug-induced
    - Antihypertensives
    - Antidepressants
    - Antiandrogens
    - Marijuana
    - Opiates

Mixed Organic/Psychogenic = most frequently encountered type of erectile dysfunction

Endocrinologic conditions may also contribute to ED. Androgen deficiency and androgen deprivation therapy for prostate cancer are associated with decreased libido and less frequent nocturnal erections. However, erectile response is preserved in many men with low serum T, suggesting that androgens are beneficial but not essential for erection. Hyperprolactinemia has been linked to sexual dysfunction, as prolactin has an inhibitory effect on both central dopaminergic activity and gonadotropin-releasing hormone secretion.

Vascular disease is a common and important cause of ED. Common vascular risk factors associated with generalized penile arterial insufficiency include hypertension, hyperlipidemia, cigarette smoking, and diabetes mellitus. Focal stenosis of the common penile artery is uncommon but can be a cause of ED in men who have sustained repetitive or severe pelvic or perineal trauma, such as with biking accidents or pelvic fractures. Venoocclusive dysfunction results from disruption of the blood-trapping capacity of the tunica albuginea: insufficient compression of the subtunical venules allows for rapid drainage of blood soon after filling, leading to loss of erection. Venous-leak ED may also be the result of degenerative conditions that affect the penis, including PD, penile scarring, trauma, and diabetes mellitus. Commonly found in young men, venous leak can also be seen in anxious men with excessive adrenergic tone causing structural alterations of the cavernous smooth muscle and endothelium in addition to insufficient trabecular smooth muscle relaxation.[14]

Many drugs are associated with ED, with some classes more commonly associated with sexual side effects than others (Table 109.3). Antipsychotics and antidepressants may disrupt central neurotransmitter pathways involved in sexual function. Antihypertensives are hypothesized to contribute to ED by decreasing the availability of blood to the corporal arteries. However, angiotensin-converting enzyme inhibitors and angiotensin receptor blockers are less likely to cause ED than beta-adrenergic blocking drugs, which potentiate a-1 adrenergic activity. Thiazide diuretics cause ED via an unknown mechanism.

Cigarette smoking acutely induces vasoconstriction and penile venous leakage secondary to the contractile effect on the cavernous smooth muscle, and chronic use likely accelerates atherosclerotic changes in the penile microvasculature over time. Alcohol in small amounts may improve erections and increase libido because of its vasodilatory effect and the suppression of performance anxiety, but increased use may cause sedation, decreased libido, and transient ED. Chronic alcoholism can cause hypogonadism and polyneuropathy, among other harmful conditions.

KEY POINTS

- Erectile dysfunction (ED) is defined as the persistent inability to achieve and/or sustain penile erection adequate for sexual satisfaction. Penile erection is a complex process that involves the interplay among psychological, organic, and relational components. When diagnosing the etiologies of ED, it is helpful to consider both organic and psychogenic causes with the understanding that the majority of patients have "mixed" ED with multiple etiologies. Common factors contributing to ED are drugs (prescription and nonprescription); tobacco and alcohol misuse; psychological, neurological, endocrinologic, and vascular disease states; and age-related changes.

TABLE 109.3 Medical Conditions and Medications Frequently Associated with Erectile Dysfunction

Condition	Medication Associated with Erectile Dysfunction
Antihypertensives	Thiazide diuretics Beta-blockers Spironolactone
Atrial fibrillation	Digoxin
Benign prostatic hyperplasia	Alpha-blockers 5 alpha reductase inhibitors
Chronic pain	Opiates
Gastroesophogeal reflux disease	H2-blockers
Prostate cancer	Antiandrogens Luteinizing hormone–releasing hormone agonists Luteinizing hormone–releasing hormone antagonists
Psychiatric conditions	Antipsychotics Benzodiazepines Phenytoin Selective serotonin reuptake inhibitors Tricyclic antidepressants

PENILE PHYSIOLOGY

Anatomy of the Penis

The penis is a highly dynamic organ with complex anatomy that facilitates its functions for sexual reproduction. The penis is grossly made up of three anatomic structures: the paired corpora cavernosa, and the corpus spongiosum, which surrounds the urethra and terminates at the glans penis. These structures are contained in the penile shaft by the tunica albuginea, which is composed of a dense network of elastin fibers on which collagen fibers rest (Fig. 109.1). The tunica of the corpora cavernosa is a bilayered structure.[15] The inner circular layer supports and contains the cavernous tissue and projects intracavernosal pillars that connect to the midline septum of the tunica. These pillars, or struts, and the midline septum provide structural support to the corpus cavernosum during erection. The outer layer of the tunica is oriented longitudinally and extends from the crura of the penis proximally to the glans penis distally. These fibers insert into the inferior pubic ramus and are absent ventrally between the 5- and 7-o'clock positions. The corpus spongiosum, which surrounds the urethra and ends in the bulbous glans penis, lacks an outer layer and has no intracorporal pillars. This ensures a low-pressure, compliant structure during erection/intercourse and expulsion of the ejaculate fluid. Emissary veins pierce the outer bundles obliquely and run between the inner and outer layers of the tunica of the corpora cavernosa for a short distance. Branches of the dorsal artery take a more direct perpendicular route through the tunica and are surrounded by a fibrous sheath.[16]

The arterial blood supply to the penis comes from the paired internal pudendal arteries. The pudendal divides into the dorsal artery (to skin, subcutaneous tissue, glans penis), the bulbourethral artery (to the corpus spongiosum), and the cavernous arteries (to the corpora cavernosa). Accessory pudendal arteries may also supply a major portion of the penis, and these arteries tend to originate from the external iliac or obturator arteries. Collaterals among the three branches are frequently encountered. The venous drainage of the penis is also multiple. The venous drainage of all three corporal bodies originates in the corporal subtunical venules, which unite to form the emissary veins. The mid and distal corpora cavernosa drain via the deep dorsal and periarterial veins to the preprostatic plexus, while the proximal portion drains via the cavernous and crural veins to the preprostatic plexus and internal pudendal vein. The glans penis drains via the deep dorsal and dorsal veins, and the corpus spongiosum drains via the circumflex, urethral, and bulbar veins. The penile skin and subcutaneous tissue drain via the superficial dorsal veins, which then empty into the saphenous veins.

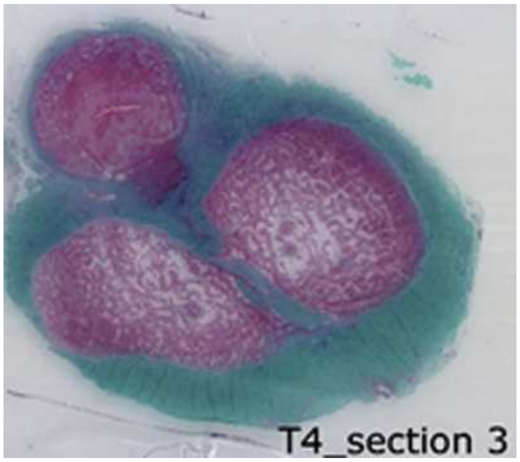


Figure 109.1 **Trichrome histology slide showing a cross section of the penis.** The paired corpora cavernosa and the singular corpus spongiosum are stained pink, and the tunica is stained blue. (Image courtesy of Lia Banie.)

Hemodynamics of Penile Erection

The penile erectile tissue, composed of cavernous, arteriolar, and arterial wall smooth muscle, is fundamental to the erection process. The erection process can be divided into six phases (Table 109.4). In the flaccid state, the vascular smooth muscles are tonically contracted secondary to baseline adrenergic discharge. This allows only a trickle of arterial flow, which is all that is needed to adequate to maintain cellular respiration. The blood partial pressure of oxygen (PO_2) in the flaccid state is approximately 35 mm Hg.

Activation of the cavernous autonomic nerves produces a full erection secondary to rapid filling and trapping of blood in the paired corpora cavernosa.[17] When the penile smooth muscle relaxes secondary to neurotransmitter release, resistance to incoming arterial flow drops to a minimum. This leads to arterial and arteriolar vasodilatation and expansion of the sinusoids to receive a large volume of arterial blood. The rapid delivery and trapping of blood in the penile sinusoids cause the corpora cavernosa to lengthen and widen until the capacity of the tunica albuginea is reached. The sinusoidal walls expand and push outward against the tunica albuginea, resulting in compression of the subtunical venous plexus. Stretching of the layers of the tunica albuginea compresses the emissary veins and effectively reduces the venous flow to a minimum.[17] Intracavernous pressure (ICP) and PO_2 increase to approximately 100 and 90 mm Hg, respectively, raising the penis from a dependent position to the erect state. Contraction of the ischiocavernosus muscles (somatic nerves) causes ICP to increase further, to several hundred millimeters of mercury and well above systolic blood pressure, resulting in rigid erection.[18,19]

During erection, the hemodynamics of the corpus spongiosum and penile glans are distinctly different from those of the corpora cavernosa because of the lack of the outer layer of the tunica albuginea. Arterial flow increases in a manner similar to that in the corpora cavernosa, but the glans functions as an arteriovenous fistula during the full erection

TABLE 109.4 The Six Phases of Penile Erection

Phases of Penile Erection

1. **Flaccid**
    - Minimal arterial or venous blood flow, with blood gas values equal to venous blood.
2. **Filling (Latent)**
    - Increased flow in the cavernous artery during systole and diastole.
    - Unchanged intracavernous pressure.
3. **Tumescence**
    - Climbing intracavernous pressure until full erection is achieved.
    - Expansion and elongation of the penis with visible pulsation.
    - Arterial flow decreases as the pressure rises.
    - Intracavernous pressure rises above diastolic pressure so flow occurs only during systole.
4. **Full Erection**
    - Intracavernous pressure can rise to as much as 80%–90% of the systolic pressure, with blood gas values similar to arterial blood.
5. **Rigid Erection**
    - Contraction of the ischiocavernous leads to intracavernous pressure well above systolic pressure, leading to complete rigidity.
    - Almost no blood flow through the cavernous artery for the short duration of rigid erection.
6. **Detumescence**
    - After ejaculation, tonic sympathetic discharge resumes.
    - Smooth muscle contracts around the sinusoids and arterioles.
    - Arterial flow returns to flaccid levels, and venous channels reopen.
    - The penis returns to its flaccid length and girth.

Adapted from Smith & Tanagho's General Urology, 19th Ed.[5]

phase. Perhaps this is to serve as a cushion for the corpora during intercourse or to serve as a conduit for the ejaculate after orgasm. Nevertheless, during rigid erection, most of the venous channels are temporarily compressed, and further engorgement of the glans can be observed.[17]

Innervation of the Penis

Humans are known to have three types of erections based on the source of the erection. *Genital-stimulated erections* are contact, or reflexogenic, erections that are provoked by physical contact with or stimulation of the genitals. This type of erection is preserved in patients with upper spinal cord lesions, though erection is usually limited in duration and poorly controlled by the patient. *Central-stimulated erections* are more complex and are noncontact, or psychogenic, erections provoked by memory, fantasy, or audiovisual stimuli. *Central-originated* erections are erections that occur during sleep or spontaneously without any type of intentional stimulation.[15] Most sleep-related erections occur during rapid eye movement sleep, with the activation of cholinergic neurons in the lateral pontine tegmentum but suppression of adrenergic neurons in the locus coeruleus and serotonergic neurons in the midbrain raphe.[15,20] Of note, central-originated, or sleep-related, erections are significantly reduced in men with hypogonadism.[20]

Penile erection is a complex neurologic process that involves the interplay amongst the sympathetic, parasympathetic, and somatic nervous systems. The autonomic spinal erection center is located in the intermediolateral nuclei of the spinal cord at levels S2 to S4 and T12 to L2. Nerve fibers from the thoracolumbar (sympathetic) and sacral (parasympathetic) spinal segments unite to become the inferior hypogastric and pelvic plexuses, which send branches to the pelvic organs. The cavernous nerves are the fibers that innervate the penis; these nerves course posterolaterally along the seminal vesicles to the prostate and continue through the genitourinary diaphragm along the membranous urethra. Some of these fibers travel with the cavernous and bulbourethral arteries to enter the corpora cavernosa and corpus spongiosum. Others travel distally with the dorsal nerve of the penis to enter the corporal bodies in various locations to supply the mid and distal portions of the penis and urethra. Terminal branches of the cavernous nerves are responsible for vascular events during erection and detumescence via innervation of the helicine arteries and trabecular smooth muscle of the penis.[15]

Somatic sensory nerves originate at receptors in the penis and convey sensation of touch, temperature, vibration, and pain. The center for somatic motor nerves is located at Onuf's nucleus in the ventral horn of the S2 to S4 segments. Motor nerve fibers travel along the pudendal nerve to innervate the bulbocavernosus and ischiocavernosus muscles of the penis. The brain has a modulatory effect on the spinal pathways of erection. The regions of the brain specifically involved in penile erection are the medial preoptic area and paraventricular nucleus of the hypothalamus, the periaqueductal gray of the midbrain, and the nucleus paragigantocellularis of the medulla. Likely there are many more neurologic pathways involved in penile erection that have yet to be discovered. In the future, advancements in imaging modalities, such as positron emission tomography and functional magnetic resonance imaging, may allow for improved understanding of central activation during sexual arousal, orgasm, and ejaculation and may help to clarify the pathophysiology triggering sexual disorders, such as psychogenic ED, orgasmic dysfunction, and PE.[21]

Neurotransmitters and Pharmacology of Erection

Neural control of penile erection involves adrenergic, cholinergic, and nonadrenergic-noncholinergic (NANC) neuroeffector systems.[15] In the flaccid state, adrenergic nerves mediate tonic intracavernous smooth muscle contraction, which is the resting state of the penis. This sympathetic contraction is mediated by activation of postsynaptic alpha-1a- and alpha-1d-adrenergic receptors and modulated by presynaptic alpha-2-adrenergic receptors.[22] For penile erection, cholinergic nerves contribute to smooth muscle relaxation through inhibition of adrenergic nerves via inhibitory interneurons and the release of NO from the endothelium by acetylcholine.[23]

The fundamental neurotransmitter that has been identified in penile erection is NO from parasympathetic NANC nerve terminals. NO is released from the cavernous nerves and endothelium of the arterial vasculature. NO activates guanylate cyclase to produce produces cyclic guanosine monophosphate (cGMP) from guanosine triphosphate.[15] cGMP activates G protein–linked cellular processes, leading to decreased intracellular calcium and relaxation of actin–myosin crossbridges in the corporal and arterial smooth musculature. The net effect is arterial vasodilation and increased penile blood flow.[24] Once blood rushes into the sinusoids, shear stress also promotes release of NO from the endothelium, augmenting smooth muscle relaxation and erection. cGMP is broken down by the enzyme phosphodiesterase type 5 (PDE5)

Oxygen tension and substances secreted by endothelial cells lining the sinusoidal spaces, prostaglandins, endothelins, and angiotensin may be involved in penile detumescence.[25] Although mechanisms of action vary, erection-inducing substances cause smooth muscle relaxation, and detumescing agents cause smooth muscle contraction.

Testosterone and Male Sexual Function

A thorough discussion of the relationship between T and erectile function is beyond the scope of this chapter. We will briefly discuss the topic here to be thorough but refer the reader to other chapters in this text for additional information. T levels do not correspond to severity of ED, and T supplementation alone would rarely be utilized as a singular treatment for ED. Lower T levels are associated with low libido and a reduction in the frequency, magnitude, and latency of nocturnal penile erections. However, erectile response to visual sexual stimulation is preserved in men with hypogonadism. This suggests that androgens are not essential for erection.[26]

The importance of normal serum T for the maintenance of male secondary sex characteristics and libido is well established.[4] Low serum T has been associated with a multitude of nonspecific somatic concerns, including fatigue, depressed mood, loss of muscle mass, weight gain, decreased exercise tolerance, and ED.[4] These symptoms are also highly prevalent in middle-aged and older men. The indications and considerations for T therapy are discussed in Chapters 106 and 107. Data are conflicting on whether T supplementation improves response to ED therapy.[26,27]

As men age, there is a progressive decline of such hormones as T, dehydroepiandrosterone, thyroxine, melatonin, and growth hormone, and there is an increase in sex hormone–binding globulin, pituitary gonadotropins, and prolactin.[28] These hormonal changes, in addition to age-related anatomic changes, such as atherosclerosis, penile smooth muscle atrophy, and loss of tissue elasticity, likely both contribute to the increasing incidence of ED as men age.[29]

KEY POINTS

- The penis is composed of the paired corpora cavernosa and the corpus spongiosum, which surrounds the urethra and terminates at the glans penis. Erection is controlled by the sympathetic, parasympathetic, and somatic nervous systems, and neuronal integrity at the brain, spinal cord, and peripheral level is required for modulation of erection. Erection requires a rapid increase in arterial blood flow with resultant compression of venous structures, which reduces venous outflow and maintains erection. The major neurotransmitter involved in penile erection is nitric oxide. Multiple other neurotransmitters, including dopamine, serotonin, oxytocin, and norepinephrine, are also involved with regulation of sexual function.
- Changes associated with aging, such as hormonal changes, atherosclerosis, penile smooth muscle atrophy, and loss of tissue elasticity, contribute to the increased incidence of ED in older men.
- Testosterone levels do not correspond to severity of ED, and studies suggest that androgens are not essential for erection. Lower T levels are associated with low libido and a reduction in the frequency, magnitude, and latency of nocturnal penile erections, but erectile response to visual sexual stimulation is preserved in hypogonadal men.

CLINICAL EVALUATION OF THE MAN WITH ERECTILE DYSFUNCTION

A focused history and physical examination are fundamental for the evaluation of men with complaints of ED. As sexual health can be a sensitive and potentially discomfiting subject, many patients are hesitant to broach the topic with their medical providers.

Discussing Sexual Health with Patients

Patient-reported validated questionnaires, such as the International Index of Erectile Function: Erectile Function Domain (IIEF-EF) or the Sexual Health Inventory for Men (SHIM; Table 109.5), are valuable tools that allow patients to bring sexual concerns to their provider's attention.[30,31] Because these questionnaires were developed with heterosexual intercourse in mind, it is important to understand the context of the patient's sexual relationship when interpreting the results of these tools and ask additional questions as needed. The SHIM is comprised of five questions based on a five-point Likert scale and was derived from the 15-question IIEF, which is a more comprehensive and lengthier instrument developed for use in clinical trials. The SHIM is specifically intended as an office screening tool for ED. Lower scores indicate more severe ED, with classification into five severity grades between no ED (SHIM score 22–25) and severe ED (SHIM score 1–7).

TABLE 109.5 Sexual Health Inventory for Men Questionnaire

Sexual Health Inventory for Men Questionnaire

Over the past 6 months:

1. How do you rate your confidence that you could get and keep an erection?
    1. Very low
    2. Low
    3. Moderate
    4. High
    5. Very high
2. When you had erections with sexual stimulation, how often were your erections hard enough for penetration (entering your partner)?
    0. No sexual activity
    1. Almost never or never
    2. A few times (much less than half the time)
    3. Sometimes (about half the time)
    4. Most times (much more than half the time)
    5. Almost always or always
3. During sexual intercourse, how often were you able to maintain your erection after you had penetrated (entered) your partner?
    0. Did not attempt intercourse
    1. Almost never or never
    2. A few times (much less than half the time)
    3. Sometimes (about half the time)
    4. Most times (much more than half the time)
    5. Almost always or always
4. During sexual intercourse, how difficult was it to maintain your erection to completion of intercourse?
    0. Did not attempt intercourse
    1. Extremely difficult
    2. Very difficult
    3. Difficult
    4. Slightly difficult
    5. Not difficult
5. When you attempted sexual intercourse, how often was it satisfactory for you?
    0. Did not attempt intercourse
    1. Almost never or never
    2. A few times (much less than half the time)
    3. Sometimes (about half the time)
    4. Most times (much more than half the time)
    5. Almost always or always

Score

- 1–7 = Severe ED
- 8–11 = Moderate ED
- 12–16 = Mild to Moderate ED
- 17-21 = Mild ED
- 22–25 = No signs of ED

ED, Erectile dysfunction.

© 2013 Pfizer Inc. https://www.pfizerpro.com/sites/default/files/shim_vgu610709-01.pdf.

Simple open-ended questions about sexual health may be even more helpful in prompting a conversation about sexual wellness.[31] Facilitating the conversation and letting patients know that it is safe and appropriate to discuss sexual wellness with their providers can alleviate anxiety and give the patient (and possibly the partner) an opportunity to open up about sexual concerns.[32] It is important for healthcare providers in all medical and surgical fields to be comfortable discussing sexual wellness with patients. This is particularity important when a condition or an intervention is being considered that may have sexual health ramifications.[32] Unfortunately, many providers are also hesitant to broach the topic of sexuality with patients.[33] We understand that sexual health need not figure prominently in every patient visit, but we hope to encourage providers to become comfortable with the topic and attentive to the sexual concerns that their patients may have.

History

If the patient expresses an issue with sexual function, it is important to define exactly what his concerns are. Many men will complain of ED but actually have other sexual concerns, such as decreased libido, penile deformity, or PE. ED may be either the cause or the effect of other these sexual problems, such as PE or decreased sexual desire. For instance, a man who lacks desire for sex will typically have difficulty obtaining an erection, and a man who lacks confidence in his erection may engage in vigorous thrusting that leads to rapid ejaculation due to fear of losing his erection.[34] A man with ED may also have reduced sexual interest secondary to performance anxiety.

For men with sexual complaints, a thorough history should include multiple components. First, a medical and surgical history is essential to evaluate the potential roles for medical or surgical conditions as the etiology for ED and to investigate both the organic and psychological causes of ED. A detailed sexual history aimed at determining onset, severity, and duration of ED should also be obtained. Finally, a psychosocial history is also critical, as ED is associated with confidence, self-esteem, relationship quality, and overall quality of life.[35] It is most ideal to define the patient's primary sexual complaint and to focus initial treatment on that specific issue. No matter the etiology or the sexual complaint, erectogenic therapy figures prominently in the management of most male sexual concerns, as the available therapies are efficacious. There is a lack of commonly accessible and approved therapies for many other male sexual concerns. Improved ability to maintain rigid erection typically increases sexual confidence, while other sexual issues, such as PE, Peyronie's disease, and orgasmic dysfunction, are addressed by other means.

ED may be generalized or situational. The context of the sexual relationship is important to determining the most appropriate treatment. For men who do not have a regular partner or those who have multiple partners, inquiry should be directed towards partner-specific differences in erectile response versus erection with masturbation alone. Impaired penile rigidity may be of differing consequence dependent on types of sexual activity. For instance, penile rigidity is less important for masturbation or oral intercourse, more important for vaginal penetration, and even more crucial for anal penetration. Onset of sexual dysfunction should be discussed in addition to any precipitating events, treatments, or conditions such as relationship stress, medical diagnoses, trauma, surgery, or addition of medications.[3,32] In some cases, if the patient has a sexual partner, it is helpful to discuss sexual issues with both partners in order to obtain additional context and to facilitate the candid communication that can be useful in sexual therapy.[36] In other cases, partner involvement may cause patient embarrassment or humiliation. It is notable that improvements in the male partner's sexual function have been associated with improvements in the female partner's sexual function.[37] It is logical that similar parallels would be noted in the sexual function of male partners of men with effectively managed ED.

Physical Examination

A targeted physical examination is important for the evaluation of ED but typically cannot precisely diagnose the etiology of ED or the severity of the condition. A thorough examination may be reassuring to anxious patients, particularly if the provider affords context and explains of the purpose of the various assessments.[3] Vital signs should be obtained, as hypertension is a major risk factor for ED. The patient's general state of being should be assessed for indications of anxiety, depression, and general wellness. Assessment of facial hair and virilization may provide evidence regarding androgen status. Examination of the chest and abdomen may reveal gynecomastia, obesity, ascites, aortic aneurysm, or possibly other findings that point toward an endocrine or metabolic etiology that may contribute to ED or complicate its management.[3] Examination of the penis includes notation of the position of the urethral meatus, circumcision status, presence of scars, and any distinct lesions that may warrant either biopsy or screening for sexually transmitted infection. Examination of the scrotum and testicles can provide information regarding androgen status. Digital rectal examination to assess anal sphincter tone is informative if there is concern for a neurological lesion as a potential cause for ED. Assessment of the bulbocavernosus reflex, which is contraction of the anal sphincter in response to pressure on the glans penis, may also be useful as an assessment of neuropathy. Examination of peripheral pulses and lower extremities may also reveal arterial or venous vascular disease relevant to the ED diagnosis.[3]

Laboratory Testing

Laboratory evaluation of the man with ED is focused on detection of metabolic conditions that may contribute to ED. Fasting lipids, glucose, and glycated hemoglobin should all be evaluated. A basic electrolyte and/or hepatic function panel may be of use for assessment of general health, but does not typically change management of ED.[3] The 2018 AUA guidelines on management of ED recommend assessment of morning T levels in all men presenting with ED.[4] Men with bothersome sexual symptoms in addition to low T as determined by the established reference ranges[38] should undergo repeat testing, including early morning serum T, luteinizing hormone, and prolactin.[4] The decision of whether to supplement T should be made after careful consideration of the risks and benefits of T treatment for each individual patient.[4] T treatment alone is rarely adequate to treat ED, but normalizing serum T may improve patient response to other pharmacologic treatments specific for ED.[39]

Specialized Testing for Erectile Dysfunction (Table 109.6)

The **combined injection and stimulation test (CIS)** is a valuable augment to the physical examination in many men who have complaints of ED. Typically, this test is reserved for patients who are not responding to oral agents for ED or those who have concerns about penile deformity. It is valuable to describe the process for the patient prior to proceeding, as anxiety about the process may inhibit the erectile response through mechanisms discussed previously. For this test, an erectogenic medication, typically a combination of papaverine and phentolamine and/or prostaglandin E1, is injected into the penis in the clinical setting. The type of medication and the dosage is selected contingent on the suspected etiology of the disorder. The patient's erectile response allows for an assessment of the arteriogenic, venogenic, and possibly psychogenic etiologies of ED, in addition to

TABLE 109.6 Specialized Tests That May be Helpful in the Evaluation of Erectile Dysfunction Patients

Test	Indications
Combined injection and stimulation test (CIS)	• Assesses penile vascular function and presence of penile deformities with erection • Allows for correct dosing and concentration of medication for men who choose ICI therapy for ED
Penile duplex ultrasonography (PDU)	• Allows for same as CIS, which is part of the PDU process • Assesses arterial inflow and venoocclusive function of the penis • Assesses internal anatomy of the penis to evaluate vessels for calcification, corpora and septum for fibrosis, and tunica for plaque formation or calcification
Cavernosometry and cavernosography*	• Assessment for congenital or traumatic venous leakage in men with venoocculsive ED
Selective internal pudendal pelvic arteriography*	• Assess for the presence of focal arterial lesions in men with arterial ED on PDU • Typically reserved for very young men with a history or traumatic ED
Ambulatory nocturnal penile tumescence and rigidity*	• Assess nocturnal erectile responses • Historically used to differentiate psychogenic from organic ED • Unreliable

*Denotes that this test is rarely used in contemporary practice.
ICI. Intracavernosal injection; *ED,* erectile dysfunction.

allowing for physical examination of the rigid penis, which may reveal penile deformities. CIS facilitates a more thorough discussion of the etiology/etiologies of ED and may help to guide additional medical therapy or suggest instances when surgical therapy would be most effective.

Penile color duplex ultrasound (PDU) is considered by many to be the gold-standard assessment for ED due to its ability to provide functional hemodynamic data, but in most cases it does not change the management of ED. In addition, the quality of PDU is exceedingly operator-dependent, and studies show that many patients may need repeated testing (i.e., repeated penile injections or even additional clinic visits) in order to fully evaluate their complaints.[40] With these issues in mind, PDU is best conducted in high-volume centers by examiners with extensive training and experience in conducting and interpreting the study.

PDU does have utility in specific clinical situations. It is best reserved for patients in whom the results may change the management strategy, such as in young men with a history of severe pelvic trauma or lifelong, severe ED. In some cases, PDU may suggest cases in which medical therapy is unlikely to succeed. Patients with very poor response to penile injection or with poor hemodynamic parameters on PDU may be counseled that they are unlikely to succeed with medical ED therapy; such men may consider more rapid selection of invasive ED therapy. Most patients and urologists are likely to agree, however, that a trial of medical therapy is warranted prior to launching into irreversible surgical interventions, regardless of what the PDU reveals. In addition to being a valuable diagnostic tool for some men, PDU may also have therapeutic value in a certain subset of men with ED who value knowing exactly what the anatomic or psychological basis for their condition may be.

Standard operating procedures for the performance and interpretation of PDU are available from knowledgeable clinicians.[41] PDU is most commonly performed in several steps—on the flaccid penis, after injection of vasodilators, and finally after self-stimulation to erection—to assess blood flow in the cavernous arteries. The most important hemodynamic parameters are peak systolic velocity (PSV) and end diastolic velocity (EDV) of the rigid erection. A PSV of greater than 30 cm/s is generally considered normal. Arterial flow of less than 25 cm/s is considered evidence of arterial insufficiency.[41] Resistive index (RI) is a derived variable ([PSV– EDV]/PSV) that is used to assess the presence of venous leak or venooclussive disease; RI of 0.6 or less is consistent with corporal venous leakage.[41] In addition, PDU assesses the echogenicity of the spongy tissue to evaluate for corporal fibrosis and for plaques of the tunica albuginea that might suggest of Peyronie disease.[41,42]

Biothesiometry is sensory testing that may be utilized in patients who have concerns about decreased penile sensation. Most often, a handheld, electromagnetic, adjustable, vibrating device known as a biothesiometer is utilized, but other modalities such as heat/cold or sharp/dull or soft/hard may also be employed. The stimulus is applied to the glans penis and both sides of the penile shaft and with varying degrees of intensity. For vibratory stimulus, the minimum vibratory intensity that is detectable to the patient is assessed. These values are compared with the sensitivity of other skin regions, most commonly the pulp of the fingertips and big toes bilaterally.[43] Penile thermal sensory testing may be performed to quantify conductance of small sensory nerves, which may reflect autonomic disturbances in the context of diffuse neuropathy. Although biothesiometry may be informative, it seldom leads to a change in management. It may, however, prompt further evaluation and possibly the need for neurology consultation.

Cavernosometry with or without **cavernosography** is rarely utilized, due to invasiveness and lack of clinical applicability in most cases of ED. These tests require injection of an erectogenic agent followed by assessment of real-time ICP during saline infusion into the penis. Important metrics include ICP and flow rate needed to maintain erection. Lower pressure and higher flow rate are indicative of venous leak as the etiology of ED. If contrast infusion is performed, fluoroscopic images or even computed tomography scan can be obtained to evaluate for venous leakage. It is essential that maximal penile erection be obtained for accurate assessment of leakage.[44] Diffuse leakage of contrast is the most common finding in older men with diabetes or other conditions leading to tissue senescence, and this is not amenable to surgical repair. Rarely, a young man with ED may have a discrete area of venous leakage from the corpora cavernosa and may be considered a candidate for a venous ligation procedure. However, this intervention that is not supported by the current AUA guidelines on ED.[3]

Selective internal pudendal angiography (SIPA) may be indicated in the case of a young man with a history of ED after pelvic trauma with a unilateral or bilateral cavernosal PSV on PDU that is below expected thresholds. In select cases, SIPA may reveal a focal occlusion of the pudendal or more distal arteries that is potentially amenable to bypass for penile revascularization.[45] This procedure should only be performed in centers of excellence with highly experienced surgeons after meticulous patient selection.

Various other tests, including a variety of reflex and autonomic nervous system tests, have been promoted for assessment of penile erection response; examples include bulbocavernosus reflex time, sympathetic skin response, pudendal somatosensory evoked potentials, and cavernous electromyography. At this time the clinical utility of these examinations is unclear.[46]

KEY POINTS

- History and physical examination should focus on the various etiologies (psychogenic and organic) of ED. Validated questionnaires may be useful clinical tools to initiate a discussion about sexual dysfunction, especially for patients who are embarrassed or uncomfortable with the topic. In addition to obtaining a thorough history, assessing vital signs, and performing a thorough physical exam, consider ordering labs, such as fasting lipids, glucose, HbA1c, basic renal and/or hepatic function panel, and early morning testosterone levels, to evaluate for medical conditions associated with ED.
- If clinically indicated, there are many specialized tests, including combined injection and stimulation test, duplex ultrasonography, biothesiometry, cavernosometry/cavernosography, and selective internal pudendal arteriography, that a specialist may offer to further evaluate the severity and etiologies of ED.

TREATMENT OF ERECTILE DYSFUNCTION

The patient's baseline state of health, relationship status, objectives for treatment, and threshold for risk must all be considered when discussing treatment options for ED. Each man will have his own preferences regarding how to treat ED, and it is best to outline the treatment options to allow for an informed, shared decision.[3] As part of the ED treatment discussion, men should be counseled on how other medical conditions may contribute to the condition. Comorbid conditions that are detected during ED evaluation, such as obesity, diabetes, hypertension, hyperlipidemia, and arterial vascular disease, should be appropriately managed by the diagnosing physician or referrals placed to consultants, such as nutrition, cardiology, endocrinology, neurology, bariatric surgery, or vascular surgery.[3]

The Princeton III guidelines provide recommendations regarding management of ED in light of cardiovascular risk.[47] Patients who are deemed high-risk for cardiac events, such as those with unstable angina, high-risk arrhythmia, severe congestive heart failure, myocardial infarction within the past 2 weeks, and severe cardiac valve disease, should have stabilization of their cardiac status before engaging in sexual activity. Intermediate-risk patients, such as men with noncardiac vascular disease, including peripheral artery disease or history of transient ischemic attack, mild to moderate angina, moderate congestive heart failure, or other conditions that portent moderate cardiac risk, may consider cardiac evaluation before engaging in sexual activity or receiving treatment for ED.[47]

Adaptive Sexuality

Some patients may not be candidates for medical or surgical therapy for ED due to medical or surgical comorbidities, or simply personal preference. Many of these patients can maintain a satisfying sexual life by adapting their sexual practices to accommodate the absence of rigid penile erections.[3] Alternative sexual activities, such as mutual masturbation, oral sex, or anal stimulation, may be satisfying options for some couples if penile rigidity is insufficient for coital intercourse. Use of assistive devices, such as vibratory stimulators or lubricants and receptive devices, and may also be valuable adjuncts for some couples.[48] For patients with physical limitations that make the act of intercourse challenging, there are special positioning devices that may augment sexual encounters to improve mechanics and sexual satisfaction. For patients who are interested in exploring adaptive sexuality, referral to a sex therapist is recommended. Sex therapists are behavioral health specialists trained to educate patients and their partners and to facilitate development of a satisfying sexual relationship tailored to patient goals and limitations.

Lifestyle Modification

Men with ED should be counseled on the importance of healthy lifestyle in terms of diet, nutrition, activity, body weight, and sleep. The Massachusetts Male Aging Study revealed that men who are sedentary are at greater risk of ED than men who are active, as well as men who are sedentary but become physically active.[49] This study also identified that the risk factors for ED are very similar to those of cardiovascular disease (CVD), including diabetes, smoking, and dyslipidemia. With this in mind, ED should be considered at the very least a marker for possible further vascular disease in all men. Men with new-onset ED and no signs or symptoms of CVD are at increased risk of a cardiovascular event in the subsequent 3 to 5 years.[50] The presentation of a *young* man with ED gives the clinician an opportunity to assess for CVD and to discuss prevention as well. Men in their 40s with ED have a 50-fold increased risk of CVD, and men in their 20s and 30s have a 7-fold increased risk.[51] Atherosclerosis affects all vascular beds equally, but smaller arteries are more likely to become occluded than larger vessels. As penile arteries are only 1 to 2 mm in size, and coronaries are larger (3–4 mm), penile arteries may serve as a sensitive indicator for consequent development of CVD.[52]

Although positive lifestyle changes are challenging to implement, and even more challenging to maintain in many cases,[3] the possibility of improving sexual response may be a powerful motivator for some men and their partners. Robust evidence demonstrates that men who are sedentary and/or obese are at greater risk for ED compared with men who are active and of normal body weight.[53] A randomized trial of men with metabolic syndrome revealed that intensive lifestyle change, including exercise and reduced intake of processed, high-fat foods, improved metabolic parameters and produced a clinically significant improvement in erectile function as measured by the SHIM.[54] Furthermore, a systematic review of studies on exercise for ED intervention revealed that a supervised program of at least 40 minutes of moderate to vigorous physical activity four times a week was associated with meaningful improvements in erectile function.[55]

Mental Health

Mental health plays a central role in male sexual wellness. Mental distraction inhibits the ability of the man to fully engage in sexual relations. Interpersonal issues with partners may be the primary precipitating factor in some cases of ED, but other factors such as depression, anxiety, adjustment disorders, or medications taken for these conditions may also contribute to male sexual dysfunction.[56] Psychological stress frequently leads to a physiologic stress response that involves activation of the sympathetic nervous system, peripheral vasoconstriction, and inhibition of penile erection. Explaining the biologic basis for ED associated with psychological conditions may be therapeutic for some men and for their partners.[14] Even in men for whom psychological or interpersonal issues are not the primary cause of ED, anxiety over sexual performance may perpetuate ED via the same biologic mechanism.[56] If sexual partners interpret ED as a lack of interest in the relationship or as a sign of infidelity, this may further exacerbate the existing heightened adrenergic tone and worsen the overall condition.

Human sexuality is complex, and it is common for human beings to have deep-seated psychological issues relating to sexual expression. Considering that a sexual relationship typically involves two, or more, individuals who may have distinct sexual concerns, sex therapy or counseling can be an integral element in ED management. Mental health interventions can be valuable as primary therapy[57] in some cases or as an adjunct to medical or surgical therapy for ED.[58] Referral to behavioral health consultants, including sex educators, counselors, and therapists, with expertise in mental health and sexual wellness is

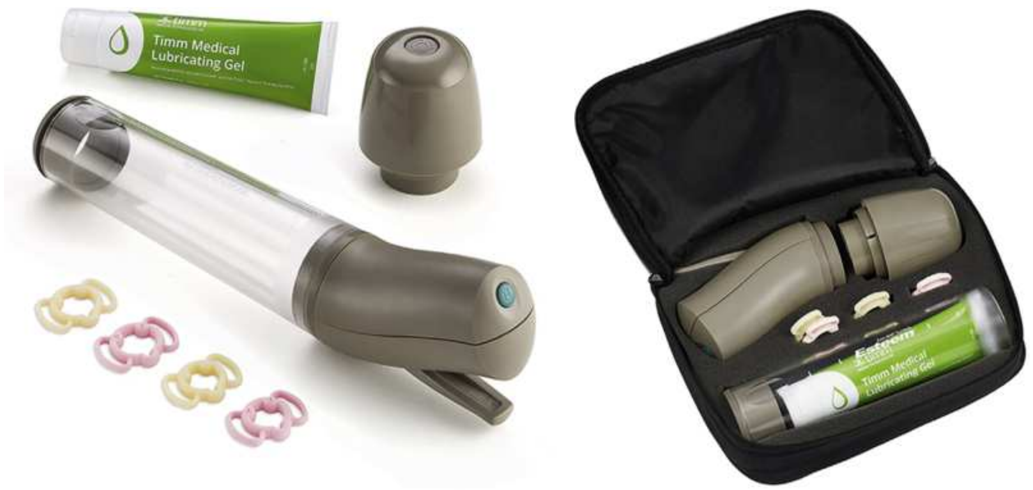


Figure 109.2 Osbon Esteem Manual Vacuum Therapy System manufactured by Pos-T-Vac Medical.

a low-risk intervention with the potential to educate about sexuality and sexual response, which is extremely valuable for most patients and their partners.

Vacuum Erection Device

Vacuum erection device (VED) therapy is a low-risk, nonpharmaceutical, nonsurgical option for management of ED. VED can be utilized in nearly all etiologies of ED and may be used as monotherapy or in combination with PDE5 inhibitors or intracavernosal injection to augment the erectogenic effects of these medications. These devices are cylindrical in shape to fit over the phallus and are designed to create a vacuum around the penis using either a hand-driven or electronic pump. This negative pressure draws venous blood into the corpora cavernosa, leading to engorgement, and in some cases turgidity.[59,60] To maintain the erection, a constrictive ring can then be placed at the base of the penis to prevent outflow from the penis (Fig. 109.2). The combination of VED and constrictive ring restores adequate rigidity for some men,[59] but others may find that it is cumbersome, uncomfortable, or embarrassing, or that the ring does not fit well due to body habitus. Some patients may find the VED-induced erection unstable, uncomfortable, cold, or of insufficient rigidity to facilitate penile intromission.[61,62] Minor risks include petechiae, bruising (particularly in anticoagulation patients), skin edema, or penile discomfort. Constrictive rings are generally safe for short periods (30–60 minutes), but severe penile tissue damage from ischemia can result if left in place for prolonged periods of time. VEDs have efficacy rates approaching those of PDE5Is and intracavernosal injection but also have low overall acceptance and high long-term discontinuation rates.

VEDs may also be recommended as a component of erectile preservation programs for patients who have had radical pelvic surgery or surgery for PD. Clinicians theorize that VED use may prevent penile shrinkage after pelvic surgery and potentially prevent PD occurrence after pelvic surgery and recurrence after penile surgery for Peyronie disease. For postsurgical patients, VEDs are frequently used in concert with oral agents, intracavernosal injection therapy, or transurethral therapies for ED.

MEDICAL THERAPY FOR ERECTILE DYSFUNCTION

In 1998, the management of ED was transformed by the introduction the first highly effective, safe oral pharmacotherapy (sildenafil) for ED.[63] This revolution not only improved access to care for ED management by allowing primary care providers to treat ED, it also changed public perception about the medical condition itself and allowed the topic to be discussed openly in various forums, particularly in mainstream media. The resulting cultural impact of oral pharmacotherapy for ED has been virtually unparalleled by any other medication in recent years. This phenomenon has encouraged additional research and development of other options for the treatment of ED.

Herbal Supplements and Complementary Medicine

Over the past two decades or so, there has been a boom in both the production and mass marketing of products and procedures targeted at men with ED. For various reasons, including skepticism about pharmacotherapy, cost, and embarrassment, many patients are enthusiastic about “natural remedies,” such as acupuncture, herbal supplements, nutraceuticals, and other nonprescription treatments, for ED. As the population of men with ED tends to be vulnerable, and in many cases desperate for an effective treatment, it is important that providers understand what products are being marketed in order to protect patients from potential harm.

Unfortunately, manufacturers of supplements and nutraceuticals are able to make expansive claims about efficacy in the absence of evidence by providing a general disclaimer statement on the product label.[64] In addition, the production of supplements is not closely regulated, and the safety, content, and quality vary substantially among herbal supplements sold by different manufacturers.[65] Most troubling, some “natural” therapeutics may be contaminated with pharmaceuticals, including prescription ED medications or similar compounds, posing serious risk to individuals with certain medical conditions.[66] Most of the trials investigating the utility of herbal formulations have been small, poorly designed studies.[67]

Korean red ginseng is the supplement with the most robust data on efficacy in ED. Even so, this is based on a few small, randomized controlled trials with inadequate sample sizes and methodological quality too low to draw definitive conclusions.[68] Yohimbine, derived from the bark of the yohimbe tree, has been used for many years to treat ED, specifically in the era prior to widespread commercial availability of the PDE5 inhibitors. Yohimbine functions as an alpha-2-adrenergic receptor antagonist. Some small studies have suggested that yohimbine is superior to placebo for the treatment of ED, including a 1998 metanalysis,[69] but the AUA guidelines panel concluded that there were insufficient quality data to recommend yohimbine as a standard treatment for ED.[3]

L-arginine, a precursor of NO, has been studied for use in ED. A recent double-blind, placebo-controlled study of 124 men aged 30 to 50 years of a commercial L-arginine product revealed a significant improvement in SHIM score.[70] Because the improvement was so dramatically different than findings from previous similar trials, however, it was determined that additional larger trials with large sample size and rigorous methodological quality are needed.

Acupuncture has been used for centuries for a variety of medical conditions, and more recently it has been employed for ED. A recent metanalysis of 22 randomized clinical trials, with 14 focused on psychogenic ED, reported that there is low-quality evidence showing a beneficial effect of acupuncture as an adjunct to other treatments for ED, including oral medications and oral herbal supplements, for patients with mainly psychogenic ED.[71] Most of the trials were identified to have small sample sizes and high or unclear bias.

Phosphodiesterase Type 5 Inhibitors

Oral therapy for ED focuses on the NO pathway. Currently, PDE5 inhibitors are the only class of orally bioavailable medications commercially available and approved by the US Food and Drug Administration (FDA) for the management of ED. These drugs inhibit the breakdown of cGMP by PDE5 to perpetuate the NO cascade.[72] All four PDE5 inhibitors that are commercially available (avanafil, sildenafil, tadalafil, and vardenafil) in the United States are highly efficacious for management of ED. There are additional oral agents available in other countries.

It is important that patients understand that PDE5 inhibitors alone do not produce penile erection, as they do not alone increase production of cGMP. Sexual stimulation to arouse the erectile response must be present in order for PDE5 inhibitors to have an effect. Men who take a PDE5 inhibitor and do not engage in erotic stimulation are not likely to have much response.[73] Likewise, men whose primary concern is decreased libido may find their response to PDE5 inhibitors suboptimal. Finally, men who have impaired NO responses secondary to cavernous nerve damage or advanced endothelial and neuronal injury from medical conditions tend to respond poorly to PDE5 inhibitors.[74]

Contraindications to the use of PDE5 inhibitors are relatively few and mostly relate to CVD. Most importantly, PDE5 inhibitor use in combination with nitrates has the potential to induce severe hypotension. PDE5 inhibitors are contraindicated for patients taking stable nitrate therapy.[75] For patients who utilize nitroglycerin as needed for angina pectoris, PDE5 inhibitors may be prescribed at the discretion of the provider after a thorough discussion about avoiding the medication within 24 to 36 hours of a PDE5 inhibitor. For these patients, topical nitroglycerin may utilized in the emergency setting as needed, as it is easily reversed if severe hypotension develops. Hypotension has also been reported in men using PDE5 inhibitors concomitant with alpha-adrenergic blockers, and dosing should be separated by at least 4 hours.[75] PDE5 inhibitors rarely cause priapism in otherwise healthy patients. The risk of priapism may be higher in patients who use these medications in conjunction with recreational or certain prescription drugs.[76]

All currently available PDE5 inhibitors are metabolized via the cytochrome CYP3A4 system. For this reason, CYP3A4 inhibitors, such as cimetidine, erythromycin, ketoconazole, itraconazole, and protease inhibitors (such as indinavir, ritonavir, and saquinavir), may lead to elevated serum levels of PDE5 inhibitors.[77,78] Lower dosages of PDE5 inhibitors in men using these CYP3A4 inhibitor medications may be considered.[79]

PDE5 inhibitor treatment has been linked to of various conditions, including vision loss from nonarteritic ischemic optic neuropathy (NAION),[80] hearing loss, melanoma, and prostate cancer recurrence. There is epidemiological evidence to support an association of PDE5 inhibitor use and some of these conditions, particularly NAION.[81] Causality remains unclear in many of these cases, and the magnitude of increased absolute risk for any of these conditions appears to be very low. Men who have or are at particular risk of these conditions should be counseled about the relative risks and benefits of PDE5 inhibitors prior to initiating therapy.[81]

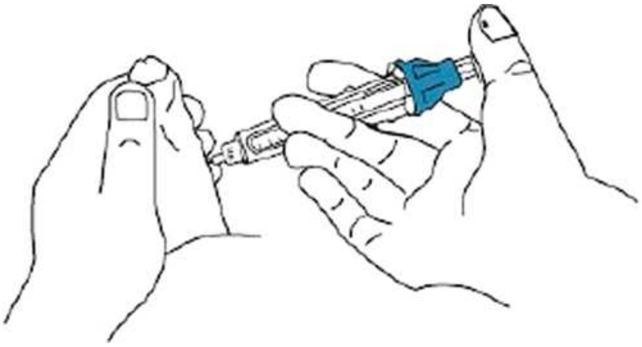


Figure 109.3 Intracavernosal injection into the lateral aspect of the penile shaft for erectile dysfunction. (From the Edex package insert.)

The most common side effects of PDE5 inhibitors are related to vasodilation and include headache, dizziness, flushing, nasal congestion, and dyspepsia. Other side effects are blue-tinted vision, back pain, myalgia, and limb pain. Depending on the drug and the dosage, the incidence of side effects from PDE5Is ranges from 8.5% to 25.1%.[82]

Sildenafil. Sildenafil citrate (Viagra, Pfizer), the first commercially available PDE5 inhibitor, has a half-life of 4 hours and reaches peak serum concentration 60 minutes after dosing. Absorption in the bowel is slowed and the peak serum concentration reduced when taken with high-lipid foods. Sildenafil should ideally be taken on an empty stomach 30 to 60 minutes prior to sexual stimulation. It is available in 25-, 50-, and 100-mg doses for the management of ED.[78]

Vardenafil. Vardenafil hydrochloride (Levitra, Bayer/GlaxoSmithKline) has a half-life of 4 to 5 hours and reaches peak serum concentration 60 minutes after dosing. Like sildenafil, vardenafil should ideally be taken on an empty stomach.[77] Vardenafil is available in 5-, 10-, and 20-mg doses. A 10-mg sublingual dissolving formulation of vardenafil (Staxyn, Bayer/GlaxoSmithKline) is available and is less affected by food intake compared with the oral tablet formulation. Half-life of the sublingual vardenafil is 4 to 5 hours, but time to peak serum concentration is 90 minutes. Absolute contraindications to vardenafil are congenital QT syndrome and class IA or III antiarrhythmics.[77]

Tadalafil. Tadalafil (Cialis, Lilly/ICOS), with a chemical structure distinct from the other PDE5 inhibitors,[83] has a half-life of 18 hours and reaches peak serum concentration 2 hours after dosing. Absorption is not influenced by food intake. Tadalafil is available as 5-, 10-, and 20-mg on-demand tablets. In addition, a once-daily dose of tadalafil 2.5- and 5-mg is available for management of lower urinary tract symptoms and may be useful for ED patients who have side effects with higher dosages of other PDE5 inhibitors. After 5 days of once-daily dosing, steady-state serum drug levels are obtained and are roughly proportional to 60% higher than the daily dose.

Avanafil. Avanafil (Stendra, Mist Pharmaceuticals/TIMM Medical), a pyrimidine carboxamide PDE5 inhibitor,[84] has a half-life of 5 hours and reaches peak serum concentration 30 to 45 minutes after dosing. Avanafil should be taken on an empty stomach, particularly with respect to fatty foods, to avoid slowed absorption and reduced peak concentrations.[84] It is available as 50-, 100-, and 200-mg on-demand doses.

Other Phosphodiesterase Type 5 Inhibitors. Multiple other PDE5 inhibitors have been approved for the management of ED outside of the United States. Examples include lodenafil, mirodenafil, and udenafil.[85-87] The efficacy and side effect profile of these drugs appear generally similar to the other PDE5 inhibitors previously discussed.

Intracavernosal Injection Therapy

In the late 1970s, scientists discovered that penile erection could be induced by direct injection of vasodilators into the corpora cavernosa.[88] Subsequently, in 1983, Dr. Giles Brindley administered papaverine to himself and subsequently displayed his erect phallus to all attendees at a national urological conference.[89,90] Ever since, injectable agents have been a mainstay of ED evaluation and management. Penile injections are a particularly important option for men who PDE5 inhibitor nonresponders and for those who are not candidates for PDE5 inhibitors.[91]

Prostaglandin E1 (alprostadil, marketed under the trade names CaverJect[92] [Pfizer] and Edex[93] [Endo Pharmaceuticals, Inc]) is the only FDA-approved injection agent for intracavernosal injection (Fig. 109.3). It is typically dosed at between 10 and 40 mcg and is manufactured in a vial along with a prepackaged delivery system.[94] Papaverine, a nonselective PDE5 inhibitor, and phentolamine, an alpha-blocker, are also frequently utilized as injectable erectogenic agents. Atropine, an anticholinergic, has been employed for intracavernosal therapy, but less often. Most commonly, these medications are compounded into mixtures of several drugs, named generically as “bimix,” “trimix,” or “quadmix,” depending on the number of component drugs. Most commonly, “bimix” refers to mixtures containing papaverine and phentolamine, whereas “trimix” refers to mixtures containing papaverine, phentolamine, and prostaglandin E1.[95] “Quadmix” would include atropine as the additional agent. The compounded medications generally cost less and can be concentrated to produce greater efficacy at lower doses of each individual component, reducing the risk of side effects and decreasing the volume of medication required. A test dose should be performed in the clinic along with self-injection teaching and demonstration of proficiency prior to at-home use.

Patients may express apprehension about the penile self-injection, but discomfort from the actual injection is rare, as the needle used is most often 30 gauge or smaller. On the other hand, men may experience discomfort from the injected agent itself, with approximately 15% to 30% of men experiencing penile pain with prostaglandin E1 injection. Petechiae and bruising may occur with self-injection of any agent, but compression at the injection site for a brief period of time (5 minutes) after injection reduces the risk of bleeding. A longer period of compression (7–10 minutes) is necessary for patients who have clotting disorders or take anticoagulant therapy. The risk of priapism with injectable agents is significant, and patients should be counseled to start at a low dose and titrate in small increments to obtain an erection satisfactory for their sexual goals. If satisfactory results cannot be obtained with a given prescription, a more concentrated mixture or one that includes additional vasoactive drugs (e.g., escalation from bimix to trimix to quadmix) may be considered. Typical volumes of medication utilized for intracavernosal injection are small, between 0.1 and 1 cc. If additional volume is required, frequently the provider will order a more concentrated medication.

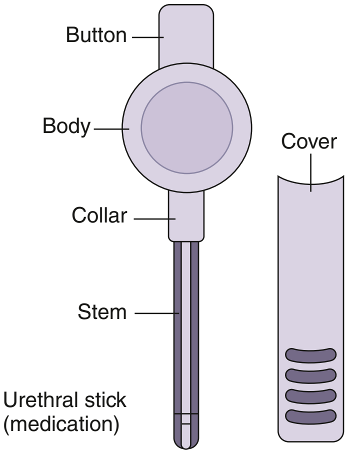


Figure 109.4 Diagram of the MUSE Transurethral System for administration of alprostadil for erectile dysfunction. (© *Viatris Inc.*)

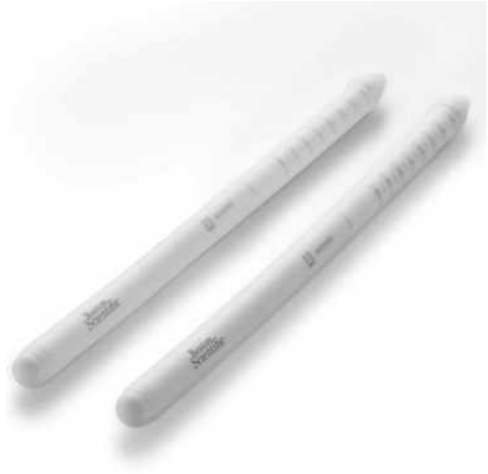


Figure 109.5 Boston Scientific Tactra· malleable penile prosthesis. (© *Boston Scientific Corporation*)

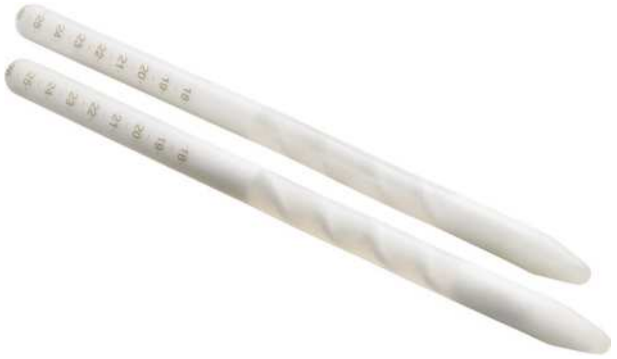


Figure 109.6 Coloplast Genesis malleable penile prosthesis. (*Used with the permission of Coloplast A/S.*)

Besides priapism, the major risks with repeated intracavernosal injections and/or the injectable agents themselves are corporal scarring, progressive fibrosis, and possibly penile deformities such as Peyronie disease. Whether these conditions occur in response to the therapy or are due to natural progression of ED is unclear. Men should alternate the side of injection and the location along the shaft at which the injection is administered and firmly compress the injection site after each injection (5–10 minutes) to minimize these potentially devastating conditions. Physical examination after 3 to 6 months of therapy should be performed to identify these conditions early and to potentially modify technique or withdraw therapy as needed.

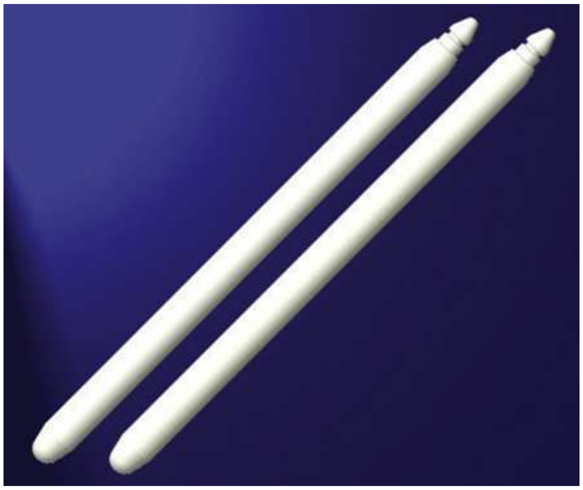


Figure 109.7 Spectra from Boston Scientific. (© *Boston Scientific Corporation*)

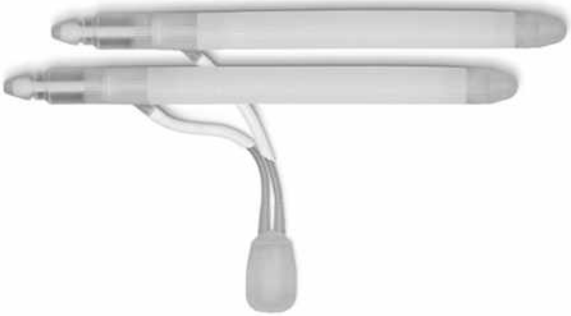


Figure 109.8 Ambicor two-piece inflatable penile prosthesis from Boston Scientific. (© *Boston Scientific Corporation*)

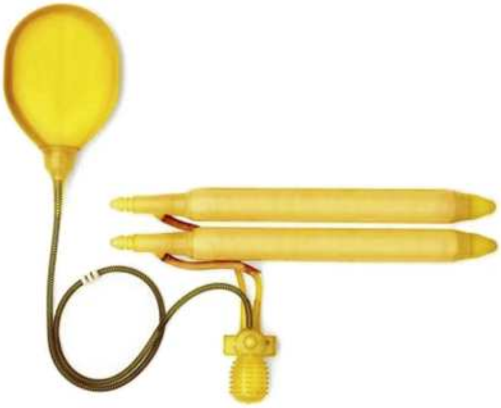


Figure 109.9 AMS700 LGX three-piece inflatable penile prosthesis with Inhibizone from Boston Scientific. (© *Boston Scientific Corporation*)

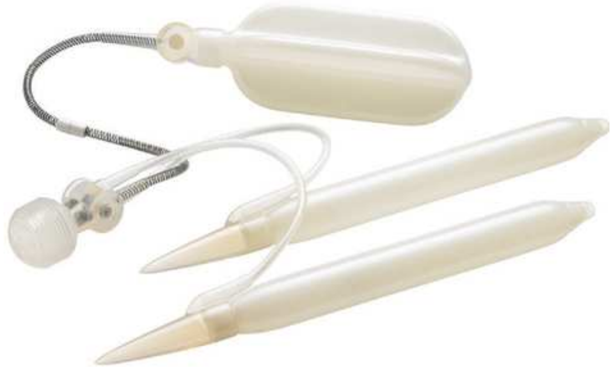


Figure 109.10 Coloplast Titan Touch three-piece inflatable penile prosthesis. (*Used with the permission of Coloplast A/S.*)

Urethral Prostaglandin Suppositories

Medicated Urethral Suppository for Erections (MUSE, Meda Pharmaceuticals[96]) is a urethral suppository containing PGE-1 (Fig. 109.4). MUSE, available in doses of 125, 250, 500, and 1000 mcg, causes corporal arterial vasodilation after absorption through the urethra into the corporal bodies.[97] Side effects may include urethral discomfort, penile pain, and mucous membrane irritation in sexual partners. Due to the potential risk of syncope, it is recommended that a test dose be applied in the clinic prior to at-home use. The most common adverse event with MUSE is genital pain, with pain in the penis, urethra, or testes reported in some men.[96] Typically, the pain is transient and mild, but it may be more severe in patients after pelvic nerve injury, such as after prostatectomy or cystectomy.

SURGICAL THERAPIES FOR ERECTILE DYSFUNCTION

The 2018 AUA guidelines on ED affirm that properly informed patients may elect to proceed immediately to surgical therapy for ED if desired.[3] Surgery may certainly be considered first-line intervention in men with severe vascular disease or ED related to radical pelvic surgery. Most often, however, clinicians recommend gradual escalation in therapy from least to most invasive, choosing to initiate therapy with an oral PDE5 inhibitor then progress to second- and third-line medical options, and finally to surgery if all other options fail. Prosthesis surgery is irreversible—the corporal tissue is permanently altered, such that physiologic erection is no longer possible without the device. The penile prosthesis is the mainstay of surgical treatment for ED, and there are two classes of penile prostheses: malleable and inflatable.

Malleable Penile Prostheses

Malleable, or semirigid, penile implants are most commonly composed of two symmetric, malleable (noninflatable) cylinders made of silicone with a central intertwined metallic bendable core. One cylinder is placed within each corpus cavernosum (Figs. 109.5 and 109.6). Another of these devices is made of a silicone rubber sheath containing (*Used with the permission of Coloplast A/S.*)rod column, which provides rigidity when the rings are lined up in a straight line and flaccidity when the penis is bent (Fig. 109.7). The advantages of all types of malleable implants are that they are easy to implant, easy to operate, and durable, as they lack mechanical parts that may break. These are also considerably cheaper than the inflatable devices, making them a popular option for patients with limited insurance coverage or resources. Besides the usual risks of penile implant (bleeding, infection, and urethral or crural perforation), the greatest risk of malleable implants is the risk of distal tip erosion into the glans penis. This may be a consequence of oversizing the device or of patient vascular disease. Additional disadvantages are that the device may interfere with urination in some cases and may be difficult to conceal under clothing. Possible candidates for malleable devices may be patients without the manual dexterity to manipulate the pump of an inflatable device, patients who desire penile rigidity for fit of a condom catheter, or patients who desire minimal complexity.

Inflatable Penile Prostheses

Inflatable penile prostheses are hydraulic devices that account for approximately 85% of the US market. These devices may have two components and or three components, with the distinction being the presence or absence of a reservoir for storage of fluid that is shifted into two inflatable cylinders placed inside the corpora cavernosa.[98]

A **two-piece inflatable prosthesis** consists of a pair of inflatable cylinders attached to a scrotal pump (Fig. 109.8). The reservoir for the fluid is in the proximal portion of the cylinders. The prosthesis is activated with the pump in the scrotum and deflated by manually bending the penis at the midshaft. These devices do permit the appearance of improved flaccidity versus the malleable implants, but there is limited detumescence that can be attained, as there is no separate reservoir to contain the fluid. The two-piece device may be considered a compromise between the ease of use of the malleable device and the potential for expansion of the three-piece device.

The **three-piece inflatable prosthesis** can be considered the gold standard for surgical treatment of ED (Figs. 109.9 and 109.10). These devices again have paired penile cylinders and a scrotal pump, but they also have a reservoir for saline that is placed in the retropubic space or posterior to the abdominal wall fascia. The advantages of three-piece prostheses are excellent rigidity when erect and a natural appearance when flaccid. These devices have high patient and partner satisfaction rates, in excess of 90%, and tend to be durable, with mean lifespans of 10 to 15 years.[99,100] The most feared, and most devastating, complication of penile implant surgery is infection. Modern prostheses incorporate antibiotic impregnation and elution, and in the hands of high-volume surgeon, infection rates are less than 2% for virgin implants.[101] Poor short-term blood glucose control has been associated with a higher risk of infection.

Careful patient selection, expectation management, and meticulous preoperative counseling are essential for optimal outcomes with all types of penile prostheses. Risks include mechanical failure, cylinder or tubing leaks, reservoir herniation, palpable tubing, perforation, persistent penile or pelvic pain, and autoinflation. Patients may also be disappointed with their erection length or girth after prosthesis placement. This can be estimated preoperatively for the patient in the clinic based on the stretched penile length in the flaccid state. The size of the device implanted is specific to the measured length of the corpora cavernosa at the time of implantation. Additionally, the glans penis does not erect with the penile prosthesis, and some patients may find this disappointing. Finally, postoperatively it may take a significant amount of time for the patient and his partner to become comfortable operating the implant. Close follow-up will help to mitigate these postoperative concerns.

Penile Vascular Surgery

Penile vascular surgery is rarely indicated for management of ED, and it is rarely effective, except in very select patients. With this in mind, we recommended that patients who may be candidates for penile vascular surgery be referred to high-volume centers. In 1902, dorsal vein ligation was first introduced as the earliest contemporary surgical approach to ED. In the 1970s, penile arterial bypass surgery was reported as a treatment for ED. In the 1980s, venous surgery was reintroduced with modifications for the treatment of venoocclusive dysfunction. The current AUA guidelines for ED management suggest that the only patients who may be candidates for arterial vascular surgery are young healthy men, without generalized vascular disease or venoocclusive dysfunction, who have focal pelvic or penile arterial occlusion. Even in these patients, penile prosthesis placement may be the surgical treatment of choice. The AUA guidelines recommend against venous surgery.[3]

Arterial Surgery for Erectile Dysfunction. Though rarely discovered, isolated stenosis or focal occlusion of pelvic or penile arteries may be amenable to surgical repair. Arterial revascularization surgery for these patients could allow for the restoration of erectile function without the need for medications, injections, or devices. Arterial reconstructive surgery is a treatment option reserved for healthy young men (usually aged 55 years or younger) with acquired ED secondary to focal arterial occlusion, typically secondary to trauma, and in the absence of severe venoocclusive disease or generalized vascular disease.[102] Most frequently, however, penile arterial insufficiency is the result of generalized arteriosclerosis associated with medical conditions such as cigarette smoking, diabetes mellitus, hypercholesterolemia, and hypertension. For patients with these conditions, arterial revascularization is not a surgical option, because diffuse small vessel disease is present.

Penile arterial insufficiency is best diagnosed by performing a CIS test followed by self-stimulation. Physical exam and PDU can then be performed to evaluate for rigidity and duration of erection in addition to arterial flow velocity and venoocclusive function. If ED secondary to arterial insufficiency is confirmed, selective pudendal/penile arteriogram can be performed to evaluate penile vascular anatomy and to identify the location of any specific arterial lesion. The most commonly used technique for penile arterial vascular surgery is to bypass the occlusion by bringing the inferior epigastric artery to either the dorsal penile artery (revascularization) or to the deep dorsal vein (arterialization). The most frequent side effect with penile arterialization is glans hyperemia, with other complications being infection, hematoma, and thrombosis of the anastomosis. Long-term results in these patients are highly variable.

TABLE 109.7 Terms Frequently Encountered in Sexual Medicine

Anorgasmia is the inability to achieve an orgasm during conscious sexual activity, although nocturnal emission may occur.

Premature (rapid) ejaculation refers to persistent or recurrent occurrence of ejaculation with minimal sexual stimulation before, on, or shortly after penetration and before the person wishes it.

Retarded ejaculation is undue delay in reaching a climax during sexual activity.

Retrograde ejaculation denotes backflow of semen into the bladder during ejaculation owing to an incompetent bladder neck mechanism.

Priapism is a prolonged erection that continues beyond or in the absence of sexual stimulation. Priapism is often painful and is considered a urologic emergency. Any erection that lasts longer than 4 hours requires urgent medical evaluation.

Peyronie disease is a fibrotic condition of the tunica of the corporal bodies of the penis that causes plaque (scar) development with subsequent penile curvature or other deformity, such as indents or hourglass deformities, of the erect penis.

Venous Ligation Surgery. Penile venous surgery is not recommended by the AUA guideline for the management of ED.[3] However, there are instances in which venous surgery may be considered in young patients with clearly defined and visualized venous leakage. Venogenic ED may be suspected in a young man who has congenital ED or posttraumatic ED. These men will typically have robust arterial inflow (>25 cm/sec) as measured by PDU, but physical exam reveals rapid resolution of erection, and PDU reveals an end-diastolic flow of greater than 5 cm/sec, signifying venous leak. The confirmatory test for the diagnosis of venous leak is dynamic infusion cavernosometry (DICC), as previously discussed. DICC requires injection of an erectogenic agent and infusion of saline to maintain an erection. An abnormal results requires a flow of more than 10 cc/minute to maintain the erection or a drop of ICP greater than 50 mm Hg within 30 seconds of infusion termination. Cavernosogram can then be used to visualize the location of the venous leakage of radioopaque contrast.

Results of penile venous surgery are mixed, and there is unanimous agreement that only young men with congenital or traumatic venous leakage should be offered venous surgery for ED. In congenital venous leakage, the venous insufficiency is typically through abnormal crural veins or the superficial dorsal vein and is amenable to surgical cure. Traumatic venous leak is usually due to localized damage to the tunica albuginea or formation of a fistula between the corpus cavernosum and the corpus spongiosum. Repair of the tunic or closure of the fistula may result in significant improvement of erectile function. In older men with chronic systemic diseases, venous leakage is usually caused by age-related cavernous smooth muscle atrophy, endothelial dysfunction, and degeneration of tunica albuginea with loss of function. In these men, venous ligation surgery is not recommended, and penile prosthesis is preferred.

KEY POINTS

- For all men with ED, it is important to address co-morbid medical and behavioral health conditions and to encourage a healthy lifestyle (diet, nutrition, activity, body weight, stress reduction, and sleep). Mental health plays a critical role in sexual wellness, and a psychological component, either as a consequent development or as principal etiology, is virtually always present in ED patients.
- Young men with ED are at increased risk of CV events, and ED may be the first sign of significant vascular disease.
- ED treatment options include nonmedical/nonsurgical (VED, adaptive sexuality), medical (PED5Is, ICI, MUSE), and surgical options (penile prosthesis). PDE5 inhibitors perpetuate the effect of nitric oxide; they are contraindicated in patients who are on stable nitrate therapy due to the risk of profound hypotension. After thorough counseling, all ED treatment options can be considered first-line therapy, based on patient preference.

  Most clinicians recommend starting with the least invasive treatments and progressing to the more invasive or irrevocable options when other options fail.
- Vascular surgery for ED is very rarely indicated but for very young men with focal arterial occlusion, and venous surgery for ED is not recommended.

ADDITIONAL CONDITIONS CONTRIBUTING TO MALE SEXUAL DYSFUNCTION

Disorders of Ejaculation, Emission, and Orgasm

Distinct physiologic mechanisms are involved in erection, emission, ejaculation, and orgasm (Table 109.7). Emission and ejaculation require stimulation of the external genitalia, except for nocturnal emissions. Signals from the pudendal nerves travel to the upper lumbar spinal sympathetic nuclei. Efferent signals travel in the hypogastric nerve to activate secretions and coordinate the passage of sperm and fluid from the epididymis, vasa deferentia, seminal vesicles, and prostate to the prostatic urethra. The internal urethral sphincter then closes, and the external sphincters relaxes in a synchronized fashion to force the semen into the bulbous urethra, a process called emission. The bulbocavernosus muscles then contract rhythmically and force the semen through the urethra to expel the ejaculate. The external ejaculation process involves the somatomotor efferent of the pudendal nerve to contract the bulbocavernous muscle, but this is involuntary, requiring integrated autonomic and somatic activation.[15] The mechanisms of orgasm, which involves cerebral interpretation of and response to sexual stimuli, is the least understood of sexual processes. Along with emission and ejaculation, several nongenital effects occur, including tachycardia, tachypnea, hypertension, and pulsatile contraction of the anal sphincter.

Retrograde ejaculation (RE) and **failure of emission** present in a similar way, with no visible antegrade ejaculate after orgasm. A thorough medical and surgical history, along with a post-orgasmic urinalysis (POU), will likely differentiate the two conditions. RE is the result of dysfunction of the internal sphincter or the bladder neck, commonly seen with alpha-blocker therapy for benign prostatic hyperplasia, in autonomic neuropathy from diabetes, and after transurethral resection of the prostate. Disruption of the retroperitoneal sympathetic nerves, such as during retroperitoneal lymphadenectomy for testis malignancy or from spinal injury at the level of L2, may result in failure of emission. To differentiate these conditions, a postorgasmic microscopic evaluation of the urine, or POU, should be performed. If there are no sperm in the POU, failure of emission can be diagnosed. If there are sperm in the POU, RE can be diagnosed.

Cessation of alpha-adrenergic blocker therapy may reverse RE in some patients. Alpha-sympathomimetics, such as ephedrine and pseudoephedrine, have been used successfully in patients with RE.[103] For spinal cord injury patients with failure of emission, electroejaculation, via a rectal probe that directly stimulates nerves via electrical impulses, has been applied with some success.[104] For patients who have normal nocturnal emissions but cannot achieve orgasm and ejaculation, psychosexual therapy is recommended.

Delayed orgasm, **DE**, and **decreased penile sensation** may also be complaints encountered during evaluation of sexual dysfunction**.** For these patients, finding an effective and satisfactory treatment can be challenging**.** Complaints of DE may be associated with certain medications, such as selective serotonin reuptake inhibitors (SSRIs) for depression. If this is the case, the patient should be referred to the prescriber for a discussion of the medication and sexual side effects to see if an alternative agent might be equally beneficial. For other men with these conditions, a handheld medical device called Viberect (Reflexonic, LLC, Chambersburg, PA, USA) may be considered as primary therapy. This device provides vibratory stimulation to the penile genital afferent nerves, provoking erection via neurostimulation. In some patients, it may also provoke emission and ejaculation.

PE, or **rapid ejaculation**, a persistent or recurrent occurrence of ejaculation with minimal sexual stimulation before, on, or shortly after penetration and before the person wishes it, is reported by up to 20% to 30% of men.[7] There are multiple recent clinical guidelines from various urological societies that may serve as a clinical reference for evaluation and treatment of PE.[6,34,39] To reduce penile sensitivity, a condom or a topical anesthetizing cream (lidocaine-prilocaine) can be applied to the glans penis. Although time to ejaculation may be prolonged, many men report penile numbness or loss of erection with this treatment. In addition to psychotherapy or behavioral therapy, an SSRI, such as paroxetine (10–40 mg daily or 20 mg 3–4 hours before intercourse) may be used as pharmacotherapy for PE, although this indication is not FDA-approved.[6] A short-acting SSRI, dapoxetine hydrochloride, has been developed for on-demand use in PE but has not yet been approved by FDA, despite several large phase III studies demonstrating efficacy. In the trials, the most common side effects were nausea, dizziness, diarrhea, and headache, and there was no SSRI discontinuation syndrome after abrupt withdrawal.[105]

Peyronie Disease

PD, or plastic induration of the penis, is an acquired, progressive connective tissue disorder of the tunica albuginea of the penis that effects between 3% and 9% of men.[106,107] Some experts suspect that the true prevalence of PD may be even higher. PD occurs most commonly in Caucasian men between the ages of 45 and 60 years, but can occur in men regardless of race or age. The underlying pathogenesis of PD is unknown, but the process is likely related to aberrant would healing, genetic predisposition, repetitive microtrauma to the erect penis during intercourse, and subsequent tissue ischemia.[108] Symptoms of PD include penile pain at rest or with erection, penile curvature or other deformity, and difficulty with erection and penetrative intercourse. Physical examination may reveal palpable penile plaque or dense scar of the tunica. Frequently, PD has significant adverse effects on quality of life, self-esteem, and relationships, resulting in considerable distress in both the patients and their sexual partners.[109] PD is described as a two-phase disease for clinical purposes and to help in staging definitive surgical therapy when indicated and desired. The acute (inflammatory or active) phase is characterized by penile inflammation, penile pain in the flaccid or erect state, growth of penile plaque, and progressive deformity. Pain usually resolves in 90% of patients during the first 6 months.[110] The chronic (fibrotic or stable) phase usually occurs 6 to 12 months later, when the inflammation resolves and the pain, plaque, and penile deformity stabilize. At this stage, improvement in penile curvature and deformity are rare.[111] Studies show that, without treatment, penile deformity worsens in 30% to 50% of patients, stabilizes in 47% to 67% of patients, and improves in 3% to 13% of patients.[111-113]

Determining the appropriate treatment for PD depends on a variety of factors: patient bother, stage of the disease, the presence of pain, severity and direction of the curvature, penile length, and erectile function.[114] There are published guidelines for the treatment of PD,[110,115] but the authors of this chapter diverge from the guidelines regarding treatment strategies for PD. Nonsurgical treatments, including pharmacotherapies, topical treatments, and intralesional injection therapy, endeavor to alleviate pain, halt disease progression, and stabilize deformity.[116] Oral therapy is indicated in men with acute disease and penile deformity associated with bother. We utilize pentoxifylline as first-line oral therapy as a key component of multimodal therapy that also includes intralesional injection[117] and traction[118] at the appropriate time during disease progression. Pentoxifylline is a nonspecific phosphodiesterase inhibitor that inhibits TGF-β1-mediated inflammation, prevents deposition of collagen type 1, and increases NO. Various other options for oral therapy, which are not recommended, have been suggested, including colchicine, PDE5 inhibitors, vitamin E, potaba, tamoxifen, carnitine, coenzyme Q10, and omega 3 fatty acids. Well-designed randomized, placebo-controlled trials have been disappointing and failed to demonstrate much benefit of any of these therapies.[110]

Intralesional therapy involves direct injection of a pharmacological agent into the penile plaque. Three intralesional drugs have shown efficacy in randomized trials: collagenase, verapamil, and interferon alpha-2b. Collagenase clostridium histolyticum is the only FDA-approved intralesional treatment for PD and is the intralesional treatment of choice. However, the treatment course is lengthy, consisting of two separate injections (separated by 1–3 days) every 6 weeks, with up to four cycles per plaque. Manual modeling procedures should begin 48 hours after the second injection of the cycle or sooner, if tolerated.[119] Mild adverse effects include penile pain, ecchymosis, erythema, and edema around the injection site. Major adverse events include penile rupture. Other medical modalities such as iontophoresis, topical therapies, extracorporeal shock wave therapy, and radiation therapy, are not recommended.[106]

When nonsurgical treatment for PD fails, surgical treatment can be considered. When a patient has a deformity that impairs sexual function, stable disease (unchanged deformity for 3–12 months), extensive plaque calcification, and a desire for rapid and reliable results, surgery may be indicated.[106,110,115] The goals of surgery are to correct the deformity, to maintain satisfactory penile length and girth, and to preserve rigidity to allow for restoration of penetrative intercourse.[114] Surgical risks include *de novo* or worsened ED, penile shortening, temporary or permanent penile hypoesthesia, chronic pain, the potential for palpation of permanent suture below the skin, or poor cosmesis of the surgical scar.[120] Recurrent curvature may also develop due to performing surgery prematurely prior to stabilization, reactivation of the condition, or the use of absorbable sutures. Procedures for correction of deformities caused by PD are classified as either penile shortening procedures (penile plication) performed on the convex side of the penis or penile lengthening procedures performed on the concave side of the penis. Lengthening procedures are used to avoid penile shortening and require the use of a graft, which carries a higher risk of ED and penile hypoesthesia. Penile prosthesis should be considered in PD associated with ED refractory to medical treatments.[115]

Priapism

Priapism is defined as a prolonged and persistent penile erection that is unrelated to sexual interest or stimulation and lasts longer than 4 hours.[121] Three types of priapism have been described: ischemic (low-flow), nonischemic (high-flow), and stuttering (recurrent).

Ischemic priapism, considered a urologic emergency, is the most common presentation of priapism and is most often idiopathic. Recognized etiologies include malignancy, medications, spinal cord injury, sickle cell disease, and glucose-6-phosphate deficiency.[122,123] The underlying mechanisms of ischemic priapism are poorly understood, but this condition results when occlusion of the venous outflow of the corpora cavernosa prevents arterial inflow, leading to a cascade of tissue ischemia, endothelial and smooth muscle damage, and eventual fibrosis.[121,124] Prompt treatment is essential to ensure the least amount of long-term disability. Treatment can be initiated with corporal aspiration of blood with injection of a diluted alpha-adrenergic agent, such as phenylephrine (500 mcg delivered every 5 minutes), the preferred agent.[121] Other vasoactive agents, such as oral terbutaline, have been deemed less effective.[125] If this is ineffective, or if the duration has been prolonged (>24–48 hours), a shunt procedure may be performed. Typically, a distal shunt between the corpora and the glans is created using a percutaneous or open technique,[126-128] and tunneling procedures may also be indicated in refractory cases.[127-130]

For patients with ischemia lasting longer than 2 days, early implantation of a penile prosthesis may be an option if there is zero response to injectable phenylephrine. Early prosthesis may result in fewer intraoperative and postoperative complications than implantation delayed for several months, with patients reporting higher satisfaction and less penile shortening with early implantation.[131] Early implantation of a penile prosthesis is not obligatory.

Nonischemic priapism is less common and is almost always the consequence or perineal or penile trauma.[132] This condition is not a medical emergency. After trauma, vascular injury to the cavernosal artery or helicine arteries is exacerbated by cyclical penile erection.[133,134] When a fistula forms between an artery and the sinusoidal spaces of the corpora, unregulated arterial inflow results in persistent, painless penile tumescence. Because venous drainage is preserved, the penile tissue does not become ischemic, and the condition is generally painless.[135]

Treatment progresses from least to most invasive and includes observation with penile compression, medical management with androgen ablation, selective arterial embolization, and surgical intervention as the last resort.[121] Androgen ablation, which reduces sleep-related erection to facilitate closure of the vascular injury, for 1 to 3 months can provide an effective option if there is damage to the cavernous artery.[133,134] If androgen ablation fails, more invasive therapies are then recommended.[136]

Recurrent ischemic, or stuttering, priapism is defined as periodic episodic prolonged erections that may evolve into ischemia priapism. Because this condition affects patients with sickle cell disease more commonly than others, microvascular occlusion secondary to hemolysis with stasis of sickled erythrocytes is thought to be a significant contributing factor.[137] Animal models and clinical studies have revealed a role for improper PDE5 regulation and abnormal NO and cGMP signaling and regulation in the pathogenesis of stuttering priapism.[138-140] Therapeutic options for this condition are limited, and the primary goal of contemporary management of stuttering priapism is to prevent additional episodes to reduce the risk of progression to a major ischemic event.[141]

Androgen ablation therapy is an effective, proven therapy to eliminate the erection-promoting effects of androgens.[142] In men with sickle cell disease, treatment with continuous, low-dose daily PDE5 inhibitors may normalize dysregulated PDE5 levels and decrease the frequency of priapism episodes.[143,144]

KEY POINTS

- Patients may refer to all sexual complaints, including problems with penile sensation, premature/delayed ejaculation, emission, orgasm, or penile deformity, as ED. Recording a good sexual history is essential to defining the actual sexual concerns as each issue has specific treatment options. Peyronie's disease is an acquired, progressive, connective tissue disease of the tunica albuginea that causes penile deformities, such as curvature, hourglass, indent, and shortening. Priapism is defined as a prolonged penile erection that is independent of sexual stimulation. There are two main categories of priapism: ischemic (emergency) and non-ischemic (nonurgent). Ejaculatory disorders and orgasmic disorders may also present as ED.

FUTURE DIRECTIONS FOR THE TREATMENT OF ERECTILE DYSFUNCTION

With the exceptions of lifestyle modifications and revascularization procedures, currently available interventions for ED merely treat the problem, offering symptomatic relief rather than a cure for the underlying disease process. The collective desire to develop a cure has stimulated interest in development of novel therapeutics for ED. Multiple preliminary trials have investigated the restorative potential of promising regenerative therapies, including stem cell therapy (SCT), low-intensity extracorporeal shock wave therapy (LiESWT), and platelet-rich plasma (PRP) injection therapy. The underlying mechanisms and the short- and long-term biological effects of these therapies remain to be fully elucidated. Well-designed, rigorous studies are needed to further define the mechanisms of action, safety profiles, efficacy, and protocols for treatment in humans. At this time, these treatments should be reserved for use only in the context of institutional review board–approved clinical trials until such time as a sufficient body of evidence has demonstrated their efficacy and safety.[3]

New Pharmacotherapeutics

Ongoing basic science research is attempting to identify effective therapeutics that target alternative erectogenic molecular pathways for PDE5 inhibitor–refractory patients. There is a specific need for pharmacotherapeutics that induce smooth muscle relaxation in the absence of NO. ROCK inhibitors and soluble guanylyl cyclase activators are of particular interest for PDE5 inhibitor nonresponders, as they are independent of the NO pathway. The ROCK pathway plays an important role in maintaining the flaccid state of the penis, in that ROCK phosphorylates and inactivates myosin light chain phosphatase, allowing myosin light chain to stay bound to smooth muscle actin.[145] Inhibition of ROCK produces a mechanism for smooth muscle relaxation. Soluble guanylyl cyclase is an enzyme in the NO pathway that can be directly activated to result in smooth muscle relaxation in the absence of NO production. Further research is warranted regarding these and other pathways and to identify additional promising treatments for ED patients.

Stem Cell Therapy

Over the past decades, there has been considerable excitement about the potential use of stem cell (SC)-based therapies in medicine in general and in the field of urology in specific.[146] SCT has been in use in other fields, such as for graft-versus-host-disease to enhance engraftment in patients receiving allogeneic hematopoietic SC transplantation, since the 1990s.[147] In addition, the antiinflammatory, restorative, and immunomodulatory qualities of SCs are currently being exploited clinically to treat hematologic diseases,[148-150] CVD,[151] neurologic diseases such as Parkinson disease[152] and stroke,[153] and autoimmune disease such as multiple sclerosis[153,154] and lupus,[155] as well as refractory wounds,[156] spinal cord injury,[157] and cartilage defects.[158] It makes sense that the field of urology would also be interested in utilization of this type of approach for acute and chronic urologic illnesses.

SCs have been investigated for ED based on the premise that SCs can differentiate into various cell types, including endothelial cells, smooth muscle cells, Schwann cells, and neurons.[159] One hypothesis suggests that SC transplantation into the penis may replenish these cell concentrations in the corporal tissue in order to restore erectile function. An alternative hypothesis is that transplanted SCs may encourage regeneration of the host's own endothelial and smooth muscle cells or may restore proper interactions between these cells via a paracrine effect.[160] The exact mechanisms through which SCs may augment erectile function are poorly understood at this time.

There have been many preclinical studies utilizing SCs in animal models of ED. For animal trials, most authors utilize an animal model that simulates either an acute iatrogenic trauma to the neurovascular bundle or a chronic disease state, such as aging, or hyperlipidemia. In acute ED models, the mechanism of action of SCs is presumed to be by paracrine action.[161-163] In contrast, in chronic ED the theoretical method of SC action is proposed to be both engraftment and cellular differentiation.[164] However, the exact mechanism of action of SCs in the treatment of ED remains uncertain.[165]

There have been a limited number of clinical trials investigating SC therapy for male sexual dysfunction. Small studies have utilized SCs in human patients with no adverse outcomes, but little can be concluded from these studies regarding the potential restorative effects of SCT. Many questions remain unanswered. Future studies should focus on robust, placebo-controlled, double-blind, randomized design in order to assess the true efficacy of SCT for male sexual dysfunction.

Low-Intensity Extracorporeal Shock Wave Therapy

On the horizon for the treatment of ED is LiESWT, modified as low-intensity pulsed ultrasound or low-energy shock wave therapy (LESWT) in different studies. Provoking widespread enthusiasm from urologists and patients alike for novel noninvasive therapies, LiESWT has already been discussed in the first-line therapy section of the European Association of Urology Sexual Dysfunction Guidelines.[166] The FDA has not yet approved LiESWT for mainstream treatment of ED in the United States, as it is important first to establish the mechanisms of action of the treatment in addition to performing rigorous clinical trials with large numbers of patients, using validated and standardized protocols, and ensuring long-term follow-up to ensure the safety and efficacy of treatment.

Many studies have investigated the mechanisms of action of LiESWT for ED, and it is believed that this therapy has the ability to activate SCs *in situ* to produce such downstream effects as endogenous

SC differentiation with subsequent angiogenesis and tissue regeneration. Animal studies reveal that LiESWT stimulates cell proliferation through activation of integrin receptors and the Rho/ROCK/Src/ERK signaling pathway[167] and by promoting multilineage differentiation of mesenchymal SC lines through the ROCK-Cot/Tpl2-MEK-ERK signaling pathway.[168] Additionally, endogenous SC recruitment and Schwann cell activation coincided with angiogenesis, tissue regeneration, and nerve generation in a rat model of pelvic neurovascular injury after LiESWT treatment.[169] The angiogenesis-promoting effect of LiESWT is believed to involve upregulation of vascular endothelial growth factor. Additionally, LESWT has been shown to markedly improve erectile function[170] through downregulation of receptors for advanced glycation end products[171] and through recruitment of endogenous mesenchymal SCs.[172]

There have been many small, short-term clinical trials and subsequent systematic reviews and metaanalyses to evaluate the efficacy of LiESWT for the treatment of ED. Though there is heterogeneity regarding study design, treatment protocol (dosing, frequency, number of shocks, linear versus focuses device), and follow-up interval, the overall conclusions suggest that there is clinical improvement with LiESWT, with no significant adverse events reported.

Proponents of LiESWT suggest that this regenerative treatment modality will be able to potentially mitigate or eliminate the safety and methodological concerns about SC acquisition, preparation, and delivery and also reduce or eliminate the need for expensive and invasive medical and surgical therapies for ED in many patients. As with all regenerative therapies, the underlying mechanisms of therapeutic ultrasound and the biological effects it has on the human body remain to be thoroughly investigated. Additional well-designed, rigorous studies are needed to further define the mechanisms of action, safety profile, efficacy, and optimal treatment protocols for this promising therapy.

Platelet-Rich Plasma

In recent years, PRP injection therapies have been utilized in many fields of medicine in hopes of restoring structure and function to diseased tissues. This therapy is based on the hypothesis that biologically active molecules including growth factors can upregulate downstream cell regeneration. For the PRP procedure, whole blood is withdrawn from the patient and then centrifuged to produce autologous plasma containing four times the normal concentration of platelets and growth factors. This concentrated plasma is then injected into the structure of concern. Beyond promising initial studies, which were limited by a low number of animals and unconventional ICP monitoring techniques, there has been a lack of significant basic science exploring the mechanisms and efficacy of PRP for ED. Additional animal studies are necessary for elucidation of the mechanisms of RPR in erectile recovery.

Across the globe, however, there are already human trials underway to investigate the use and safety of PRP in human subjects. Thus far, there have been no major adverse outcomes, and no patients have reported worsening of their ED. These studies, however, have had low patient numbers, have suffered from the lack of a control group, and have shown uncertain efficacy. Overall, current data suggest that PRP for ED is safe in humans, but no studies have determined whether there is any long-term functional benefit in the context of sexual medicine.[173] Good-quality randomized, prospective, placebo-controlled trials are needed before PRP can be recommended as a therapy for ED.

KEY POINTS

- Multiple regenerative therapies, including stem cell therapy, low intensity extracorporeal shock wave therapy, and platelet rich plasma, are under investigation in hopes of providing either a durable treatment or curative therapy for ED. At the time of this writing, these potentially restorative and regenerative therapies remain under investigation and are not recommended by professional guidelines for the treatment of ED outside of clinical trials.

CONCLUSIONS

Over the past three decades, the treatment algorithm for ED has been transformed by discovery. Whereas psychosexual therapy was once the only therapeutic option, patients can now choose from multiple efficacious pharmacologic agents and progress, if necessary, to highly effective and safe surgical interventions. New and improved therapies for ED will continue to emerge as our understanding of the basic mechanisms of penile erection continues to evolve. Penile erection is a complex physiological phenomenon that is reliant not only on neurovascular components, but also on psychological wellness. A thorough understanding of the mechanisms of penile erection and the medical, surgical, and psychosocial conditions that contribute to ED facilitates the evaluation and management of patients with this distressing condition.

REFERENCES

1. McCabe MP, Sharlip ID, Atalla E, et al. Definitions of sexual dysfunctions in women and men: a consensus statement from the Fourth International Consultation on Sexual Medicine 2015. *J Sex Med*. 2016;13:135.
2. Salonia A, Bettocchi C, Carvalho J, et al. EAU Guidelines. Edn. presented at the EAU Annual Congress Amsterdam 2022. EAU Guidelines Office, Arnhem, The Netherlands. http://uroweb.org/guidelines/compilations-of-all-guidelines/.
3. Burnett AL, Nehra A, Breau RH, et al. Erectile dysfunction: AUA guideline. *J Urol*. 2018;200:633.
4. Mulhall JP, Trost LW, Brannigan RE, et al. Evaluation and management of testosterone deficiency: AUA Guideline. *J Urol*. 2018;200:423.
5. Bhasin S, Brito JP, Cunningham GR, et al. Testosterone therapy in men with hypogonadism: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2018;103:1715.
6. Montague DK, Jarow J, Broderick GA, et al. AUA guideline on the pharmacologic management of premature ejaculation. *J Urol*. 2004;172:290.
7. Althof SE. Prevalence, characteristics and implications of premature ejaculation/rapid ejaculation. *J Urol*. 2006;175:842.
8. Araujo AB, Mohr BA, McKinlay JB. Changes in sexual function in middle-aged and older men: longitudinal data from the Massachusetts Male Aging Study. *J Am Geriatr Soc*. 2004;52:1502.
9. Feldman HA, Goldstein I, Hatzichristou DG, et al. Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *J Urol*. 1994;151:54.
10. Polsky JY, Aronson KJ, Heaton JP, et al. Smoking and other lifestyle factors in relation to erectile dysfunction. *BJU Int*. 2005;96:1355.
11. Althof SE, O'Leary MP, Cappelleri JC, et al. Sildenafil citrate improves self-esteem, confidence, and relationships in men with erectile dysfunction: results from an international, multi-center, double-blind, placebo-controlled trial. *J Sex Med*. 2006;3:521.
12. Bodie JA, Beeman WW, Monga M. Psychogenic erectile dysfunction. *Int J Psychiatry Med*. 2003;33:273.
13. Shamsa A, Motavalli SM, Aghdam B. Erectile function in end-stage renal disease before and after renal transplantation. *Transplant Proc*. 2005;37:3087.
14. Reed-Maldonado AB, Lue TF. A syndrome of erectile dysfunction in young men? *Transl Androl Urol*. 2016;5:228.
15. Reed-Maldonado A, Lue TF. Chapter 39: male sexual dysfunction. In: McAninch JW, Lue TF, eds. *Smith & Tanagho's General Urology*. 19th ed. New York: McGraw-Hill; 2020.
16. Hsu GL, Hsieh CH, Wen HS, et al. Anatomy of the human penis: the relationship of the architecture between skeletal and smooth muscles. *J Androl*. 2004;25:426.

17. Lue TF. Erectile dysfunction. *N Engl J Med*. 2000;342:1802.
18. Dean RC, Lue TF. Physiology of penile erection and pathophysiology of erectile dysfunction. *Urol Clin North Am.* 2005;32:379.
19. Gratzke C, Angulo J, Chitaley K, et al. Anatomy, physiology, and pathophysiology of erectile dysfunction. *J Sex Med*. 2010;7:445.
20. Granata AR, Rochira V, Lerchl A, et al. Relationship between sleep-related erections and testosterone levels in men. *J Androl*. 1997;18:522.
21. Georgiadis JR, Holstege G. Human brain activation during sexual stimulation of the penis. *J Comp Neurol*. 2005;493:33.
22. Giuliano F. Impact of medical treatments for benign prostatic hyperplasia on sexual function. *BJU Int*. 2006;97:34.
23. Saenz de Tejada I, Angulo J, Gonzalez-Cadavid N, et al. Physiology of erection and pathophysiology of erectile dysfunction. In: Lue T, Basson R, Rosen R, et al., eds. *Sexual Medicine: Sexual Dysfunctions in Men and Women.* Paris: *Health Publications.* 2004:287.
24. Andersson KE. Mechanisms of penile erection and basis for pharmacological treatment of erectile dysfunction. *Pharmacol Rev*. 2011;63:811.
25. Musicki B, Burnett AL. eNOS function and dysfunction in the penis. *Exp Biol Med (Maywood)*. 2006;231:154.
26. Corona G, Isidori AM, Buvat J, et al. Testosterone supplementation and sexual function: a meta-analysis study. *J Sex Med*. 2014;11:1577.
27. Spitzer M, Basaria S, Travison TG, et al. Effect of testosterone replacement on response to sildenafil citrate in men with erectile dysfunction: a parallel, randomized trial. *Ann Intern Med*. 2012;157:681.
28. Morales A. Men's aging and sexual disorders: an update on diagnosis and treatment. *Rev Endocr Metab Disord*. 2005;6:85.
29. Feldman HA, Longcope C, Derby CA, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *J Clin Endocrinol Metab*. 2002;87:589.
30. Rosen RC, Allen KR, Ni X, et al. Minimal clinically important differences in the erectile function domain of the International Index of Erectile Function scale. *Eur Urol*. 2011;60:1010.
31. Bayer CR, Eckstrand KL, Knudson G, et al. Sexual health competencies for undergraduate medical education in North America. *J Sex Med*. 2017;14:535.
32. Althof SE, Rosen RC, Perelman MA, et al. Standard operating procedures for taking a sexual history. *J Sex Med*. 2013;10:26.
33. Nusbaum MR, Lenahan P, Sadovsky R. Sexual health in aging men and women: addressing the physiologic and psychological sexual changes that occur with age. *Geriatrics*. 2005;60:18.
34. Althof SE, McMahon CG, Waldinger MD, et al. An update of the International Society of Sexual Medicine's guidelines for the diagnosis and treatment of premature ejaculation (PE). *Sex Med*. 2014;2:60.
35. Pastuszak AW. Current diagnosis and management of erectile dysfunction. *Curr Sex Health Rep*. 2014;6:164.
36. Perelman MA. Sex coaching for physicians: combination treatment for patient and partner. *Int J Impot Res*. 2003;15:S67.
37. McCabe MP, O'Connor EJ, Conaglen JV, et al. The impact of oral ED medication on female partners' relationship satisfaction. *J Sex Med*. 2011;8:479.
38. Travison TG, Vesper HW, Orwoll E, et al. Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *J Clin Endocrinol Metab*. 2017;102:1161.
39. Krader CG. EAU updates evidence-based guidelines on male sexual dysfunction. *Med Econ*. 2014;91:10.
40. Nascimento B, Miranda EP, Terrier JE, et al. A critical analysis of methodology pitfalls in duplex Doppler ultrasound in the evaluation of patients with erectile dysfunction: technical and interpretation deficiencies. *J Sex Med*. 2020;17:1416.
41. Sikka SC, Hellstrom WJ, Brock G, et al. Standardization of vascular assessment of erectile dysfunction: standard operating procedures for duplex ultrasound. *J Sex Med*. 2013;10:120.
42. Goldstein I, McLane MP, Xiang Q, et al. Long-term curvature deformity characterization in men previously treated with collagenase clostridium histolyticum for Peyronie's disease, subgrouped by penile plaque calcification. *Urology*. 2020;146:145–151.
43. Bemelmans BL, Hendrikx LB, Koldewijn EL, et al. Comparison of biothesiometry and neuro-urophysiological investigations for the clinical evaluation of patients with erectile dysfunction. *J Urol*. 1995;153:1483.
44. Glina S, Ghanem HSOP. Corpus cavernosum assessment (cavernosography/cavernosometry). *J Sex Med*. 2013;10:111.
45. Spiliopoulos S, Shaida N, Katsanos K, et al. The role of interventional radiology in the diagnosis and management of male impotence. *Cardiovasc Intervent Radiol*. 2013;36:1204.
46. Giuliano F, Rowland DL. Standard operating procedures for neurophysiologic assessment of male sexual dysfunction. *J Sex Med*. 2013;10:1205.
47. Nehra A, Jackson G, Miner M, et al. The Princeton III Consensus recommendations for the management of erectile dysfunction and cardiovascular disease. *Mayo Clin Proc*. 2012;87:766.
48. Miranda EP, Taniguchi H, Cao DL, et al. Application of sex aids in men with sexual dysfunction: a review. *J Sex Med*. 2019;16:767.
49. Derby CA, Mohr BA, Goldstein I, et al. Modifiable risk factors and erectile dysfunction: can lifestyle changes modify risk? *Urology*. 2000;56:302.
50. Hodges LD, Kirby M, Solanki J, et al. The temporal relationship between erectile dysfunction and cardiovascular disease. *Int J Clin Pract*. 2007;61:2019.
51. Inman BA, Sauver JL, Jacobson DJ, et al. A population-based, longitudinal study of erectile dysfunction and future coronary artery disease. *Mayo Clin Proc*. 2009;84:108.
52. Mobley DF, Khera M, Baum N. Recent advances in the treatment of erectile dysfunction. *Postgrad Med J*. 2017;93:679.
53. Simon RM, Howard L, Zapata D, et al. The association of exercise with both erectile and sexual function in black and white men. *J Sex Med*. 2015;12:1202.
54. Esposito K, Ciotola M, Giugliano F, et al. Effects of intensive lifestyle changes on erectile dysfunction in men. *J Sex Med*. 2009;6:243.
55. Gerbild H, Larsen CM, Graugaard C, et al. Physical activity to improve erectile function: a systematic review of intervention studies. *Sex Med*. 2018;6:75.
56. McCabe M, Althof SE, Assalian P, et al. Psychological and interpersonal dimensions of sexual function and dysfunction. *J Sex Med*. 2010;7:327.
57. Andersson E, Walen C, Hallberg J, et al. A randomized controlled trial of guided Internet-delivered cognitive behavioral therapy for erectile dysfunction. *J Sex Med*. 2011;8:2800.
58. Aubin S, Heiman JR, Berger RE, et al. Comparing Sildenafil alone vs. Sildenafil plus brief couple sex therapy on erectile dysfunction and couples' sexual and marital quality of life: a pilot study. *J Sex Marital Ther*. 2009;35:122.
59. Brison D, Seftel A, Sadeghi-Nejad H. The resurgence of the vacuum erection device (VED) for treatment of erectile dysfunction. *J Sex Med*. 2013;10:1124.
60. Kolettis PN, Lakin MM, Montague DK, et al. Efficacy of the vacuum constriction device in patients with corporeal venous occlusive dysfunction. *Urology*. 1995;46:856.
61. Derouet H, Caspari D, Rohde V, et al. Treatment of erectile dysfunction with external vacuum devices. *Andrologia*. 1999;31:89.
62. Vrijhof HJ, Delaere KP. Vacuum constriction devices in erectile dysfunction: acceptance and effectiveness in patients with impotence of organic or mixed aetiology. *Br J Urol*. 1994;74:102.
63. Goldstein I, Lue TF, Padma-Nathan H, et al. Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. *N Engl J Med*. 1998;338:1397.
64. Shaw D, Graeme L, Pierre D, et al. Pharmacovigilance of herbal medicine. *J Ethnopharmacol*. 2012;140:513.
65. Gilard V, Balayssac S, Tinaugus A, et al. Detection, identification and quantification by 1H NMR of adulterants in 150 herbal dietary supplements marketed for improving sexual performance. *J Pharm Biomed Anal*. 2015;102:476.
66. Campbell N, Clark JP, Stecher VJ, et al. Adulteration of purported herbal and natural sexual performance enhancement dietary supplements with synthetic phosphodiesterase type 5 inhibitors. *J Sex Med*. 2013;10:1842.
67. Bella AJ, Shamloul R. Traditional plant aphrodisiacs and male sexual dysfunction. *Phytother Res*. 2014;28:831.

68. Jang DJ, Lee MS, Shin BC, et al. Red ginseng for treating erectile dysfunction: a systematic review. *Br J Clin Pharmacol*. 2008;66:444.
69. Ernst E, Pittler MH. Yohimbine for erectile dysfunction: a systematic review and meta-analysis of randomized clinical trials. *J Urol*. 1998;159:433.
70. Ledda A, Belcaro G, Cesarone MR, et al. Investigation of a complex plant extract for mild to moderate erectile dysfunction in a randomized, double-blind, placebo-controlled, parallel-arm study. *BJU Int*. 2010;106:1030.
71. Lai BY, Cao HJ, Yang GY, et al. Acupuncture for treatment of erectile dysfunction: a systematic review and meta-analysis. *World J Men's Health*. 2019;37:322.
72. Eardley I. The role of phosphodiesterase inhibitors in impotence. *Expert Opin Investig Drugs*. 1997;6:1803.
73. Atiemo HO, Szostak MJ, Sklar GN. Salvage of sildenafil failures referred from primary care physicians. *J Urol*. 2003;170:2356.
74. Fode M, Ostergren PB, Jensen CFS, et al. Treatment effects of phosphodiesterase-5 inhibitors may improve with time following nerve-sparing radical prostatectomy. *Scand J Urol*. 2018;52:108.
75. Doumas M, Lazaridis A, Katsiki N, et al. PDE-5 inhibitors: clinical points. *Curr Drug Targets*. 2015;16:420.
76. Giuliano F, Jackson G, Montorsi F, Martin-Morales A, Raillard P. Safety of sildenafil citrate: review of 67 double-blind placebo-controlled trials and the postmarketing safety database. *Int J Clin Pract*. 2010;64(2):240–255.
77. FDA: Levitra Package Insert.
78. FDA: Sildenafil Package Insert.
79. Giuliano F, Jackson G, Montorsi F, et al. Safety of sildenafil citrate: review of 67 double-blind placebo-controlled trials and the postmarketing safety database. *Int J Clin Pract*. 2010;64:240.
80. Laties A, Sharlip I. Ocular safety in patients using sildenafil citrate therapy for erectile dysfunction. *J Sex Med*. 2006;3:12.
81. Yafi FA, Sharlip ID, Becher EF. Update on the safety of phosphodiesterase type 5 inhibitors for the treatment of erectile dysfunction. *Sex Med Rev*. 2018;6:242.
82. Chen L, Staubli SE, Schneider MP, et al. Phosphodiesterase 5 inhibitors for the treatment of erectile dysfunction: a trade-off network meta-analysis. *Eur Urol*. 2015;68:674.
83. FDA: Cialis Package Insert.
84. FDA: Avanafil Package Insert.
85. Moon KH, Ko YH, Kim SW, et al. Efficacy of once-daily administration of udenafil for 24 weeks on erectile dysfunction: results from a randomized multicenter placebo-controlled clinical trial. *J Sex Med*. 2015;12:1194.
86. Glina S, Fonseca GN, Bertero EB, et al. Efficacy and tolerability of lodenafil carbonate for oral therapy of erectile dysfunction: a phase III clinical trial. *J Sex Med*. 2010;7:1928.
87. Du W, Li J, Fan N, et al. Efficacy and safety of mirodenafil for patients with erectile dysfunction: a meta-analysis of three multicenter, randomized, double-blind, placebo-controlled clinical trials. *Aging Male*. 2014;17:107.
88. Virag R. Intracavernous injection of papaverine for erectile failure. *Lancet*. 1982;2:938.
89. Klotz L. How (not) to communicate new scientific information: a memoir of the famous Brindley lecture. *BJU Int*. 2005;96:956.
90. Brindley GS. Cavernosal alpha-blockade: a new technique for investigating and treating erectile impotence. *Br J Psychiatry*. 1983;143:332.
91. Coombs PG, Heck M, Guhring P, et al. A review of outcomes of an intracavernosal injection therapy programme. *BJU Int*. 2012;110:1787.
92. FDA: Caverject Package Insert.
93. FDA: Edex Package Insert.
94. Tsai YS, Lin JS, Lin YM. Safety and efficacy of alprostadil sterile powder (S. Po., CAVERJECT) in diabetic patients with erectile dysfunction. *Eur Urol*. 2000;38:177.
95. Porst H, Burnett A, Brock G, et al. SOP conservative (medical and mechanical) treatment of erectile dysfunction. *J Sex Med*. 2013;10:130.
96. NIH: MUSE Package insert.
97. Costa P, Potempa AJ. Intraurethral alprostadil for erectile dysfunction: a review of the literature. *Drugs*. 2012;72:2243.
98. Glina S, Montague DK, Torres LO, et al. Management of erectile impotence: use of inflatable prosthesis. *J Sex Med*. 2008;5:2494.
99. Knoll LD, Henry G, Culkin D, et al. Physician and patient satisfaction with the new AMS 700 momentary squeeze inflatable penile prosthesis. *J Sex Med*. 2009;6:1773.
100. Pryor MB, Carrion R, Wang R, et al. Patient satisfaction and penile morphology changes with postoperative penile rehabilitation 2 years after Coloplast Titan prosthesis. *Asian J Androl*. 2016;18:754.
101. Serefoglu EC, Mandava SH, Gokce A, et al. Long-term revision rate due to infection in hydrophilic-coated inflatable penile prostheses: 11-year follow-up. *J Sex Med*. 2012;9:2182.
102. Montague DK, Jarow JP, Broderick GA, et al. Chapter 1: the management of erectile dysfunction: an AUA update. *J Urol*. 2005;174:230.
103. McMahon CG, Abdo C, Incrocci L, et al. Disorders of orgasm and ejaculation in men. *J Sex Med*. 2004;1:58.
104. Soeterik TF, Veenboer PW, Oude-Ophuis RJ, et al. Electroejaculation in patients with spinal cord injuries: a 21-year, single-center experience. *Int J Urol*. 2016;24(2):157–161.
105. Owen RT. Dapoxetine: a novel treatment for premature ejaculation. *Drugs Today (Barc)*. 2009;45:669.
106. Chung E, Ralph D, Kagioglu A, et al. Evidence-based management guidelines on Peyronie's disease. *J Sex Med*. 2016;13:905.
107. Mulhall JP, Creech SD, Boorjian SA, et al. Subjective and objective analysis of the prevalence of Peyronie's disease in a population of men presenting for prostate cancer screening. *J Urol*. 2004;171:2350.
108. Bella AJ, Perelman MA, Brant WO, et al. Peyronie's disease (CME). *J Sex Med*. 2007;4:1527.
109. Smith JF, Walsh TJ, Conti SL, et al. Risk factors for emotional and relationship problems in Peyronie's disease. *J Sex Med*. 2008;5:2179.
110. Hatzimouratidis K, Eardley I, Giuliano F, et al. EAU guidelines on penile curvature. *Eur Urol*. 2012;62:543.
111. Mulhall JP, Schiff J, Guhring P. An analysis of the natural history of Peyronie's disease. *J Urol*. 2006;175:2115.
112. Kadioglu A, Tefekli A, Erol B, et al. A retrospective review of 307 men with Peyronie's disease. *J Urol*. 2002;168:1075.
113. Gelbard MK, Dorey F, James K. The natural history of Peyronie's disease. *J Urol*. 1990;144:1376.
114. Garaffa G, Trost LW, Serefoglu EC, et al. Understanding the course of Peyronie's disease. *Int J Clin Pract*. 2013;67:781.
115. Nehra A, Alterowitz R, Culkin DJ, et al. Peyronie's disease: AUA guideline. *J Urol*. 2015;194:745.
116. Gur S, Limin M, Hellstrom WJ. Current status and new developments in Peyronie's disease: medical, minimally invasive and surgical treatment options. *Expert Opin Pharmacother*. 2011;12:931.
117. Alwaal A, Hussein AA, Zaid UB, et al. Management of Peyronie's disease after collagenase (Xiaflex:(R)). *Curr Drug Targets*. 2015;16:484.
118. Usta MF, Ipekci T. Penile traction therapy for Peyronie's disease-what's the evidence? *Transl Androl Urol*. 2016;5:303.
119. Tan RB, Sangkum P, Mitchell GC, et al. Update on medical management of Peyronie's disease. *Curr Urol Rep*. 2014;15:415.
120. Montorsi F, Adaikan G, Becher E, et al. Summary of the recommendations on sexual dysfunctions in men. *J Sex Med*. 2010;7:3572.
121. Montague DK, Jarow J, Broderick GA, et al. American Urological Association guideline on the management of priapism. *J Urol*. 2003;170:1318.
122. Broderick GA, Kadioglu A, Bivalacqua TJ, et al. Priapism: pathogenesis, epidemiology, and management. *J Sex Med*. 2010;7:476.
123. Roghmann F, Becker A, Sammon JD, et al. Incidence of priapism in emergency departments in the United States. *J Urol*. 2013;190:1275.
124. Spycher MA, Hauri D. The ultrastructure of the erectile tissue in priapism. *J Urol*. 1986;135:142.
125. Martin C, Cocchio C. Effect of phenylephrine and terbutaline on ischemic priapism: a retrospective review. *Am J Emerg Med*. 2016;34:222.
126. Kovac JR, Mak SK, Garcia MM, et al. A pathophysiology-based approach to the management of early priapism. *Asian J Androl*. 2013;15:20.
127. Segal RL, Readal N, Pierorazio PM, et al. Corporal Burnett "Snake" surgical maneuver for the treatment of ischemic priapism: long-term followup. *J Urol*. 2013;189:1025.
128. Garcia MM, Shindel AW, Lue TF. T-shunt with or without tunnelling for prolonged ischaemic priapism. *BJU Int*. 2008;102:1754.

129. Brant WO, Garcia MM, Bella AJ, et al. T-shaped shunt and intracavernous tunneling for prolonged ischemic priapism. *J Urol*. 2009;181:1699.
130. Burnett AL, Pierorazio PM. Corporal "snake" maneuver: corporoglanular shunt surgical modification for ischemic priapism. *J Sex Med*. 2009;6:1171.
131. Zacharakis E, Garaffa G, Raheem AA, et al. Penile prosthesis insertion in patients with refractory ischaemic priapism: early vs delayed implantation. *BJU Int*. 2014;114:576.
132. Shigehara K, Namiki M. Clinical management of priapism: a review. *World J Men's Health*. 2016;34:1.
133. Mwamukonda KB, Chi T, Shindel AW, et al. Androgen blockade for the treatment of high-flow priapism. *J Sex Med*. 2010;7:2532.
134. Wu AK, Lue TF. Commentary on high flow, non-ischemic, priapism. *Transl Androl Urol*. 2012;1:109.
135. Brock G, Breza J, Lue TF, et al. High flow priapism: a spectrum of disease. *J Urol*. 1993;150:968.
136. Hakim LS, Kulaksizoglu H, Mulligan R, et al. Evolving concepts in the diagnosis and treatment of arterial high flow priapism. *J Urol*. 1996;155:541.
137. Adeyoju AB, Olujohungbe AB, Morris J, et al. Priapism in sickle-cell disease; incidence, risk factors and complications - an international multicentre study. *BJU Int*. 2002;90:898.
138. Champion HC, Bivalacqua TJ, Takimoto E, et al. Phosphodiesterase-5A dysregulation in penile erectile tissue is a mechanism of priapism. *Proc Natl Acad Sci U S A*. 2005;102:1661.
139. Lagoda G, Sezen SF, Hurt KJ, et al. Sustained nitric oxide (NO)-releasing compound reverses dysregulated NO signal transduction in priapism. *FASEB J*. 2014;28:76.
140. Anele UA, Burnett AL. Nitrergic mechanisms for management of recurrent priapism. *Sex Med Rev*. 2015;3:160.
141. Anele UA, Le BV, Resar LM, et al. How I treat priapism. *Blood*. 2015;125:3551.
142. Salonia A, Eardley I, Giuliano F, et al. European Association of Urology guidelines on priapism. *Eur Urol*. 2014;65:480.
143. Burnett AL, Bivalacqua TJ, Champion HC, et al. Feasibility of the use of phosphodiesterase type 5 inhibitors in a pharmacologic prevention program for recurrent priapism. *J Sex Med*. 2006;3:1077.
144. Burnett AL, Anele UA, Trueheart IN, et al. Randomized controlled trial of sildenafil for preventing recurrent ischemic priapism in sickle cell disease. *Am J Med*. 2014;127:664.
145. Yafi FA, Jenkins L, Albersen M, et al. Erectile dysfunction. *Nat Rev Dis Primers*. 2016;2:16003.
146. Khera M, Albersen M, Mulhall JP. Mesenchymal stem cell therapy for the treatment of erectile dysfunction. *J Sex Med*. 2015;12:1105.
147. Casiraghi F, Remuzzi G, Abbate M, et al. Multipotent mesenchymal stromal cell therapy and risk of malignancies. *Stem Cell Rev*. 2013;9:65.
148. Arima N, Nakamura F, Fukunaga A, et al. Single intra-arterial injection of mesenchymal stromal cells for treatment of steroid-refractory acute graft-versus-host disease: a pilot study. *Cytotherapy*. 2010;12:265.
149. Le Blanc K, Frassoni F, Ball L, et al. Mesenchymal stem cells for treatment of steroid-resistant, severe, acute graft-versus-host disease: a phase II study. *Lancet*. 2008;371:1579.
150. Ball LM, Bernardo ME, Roelofs H, et al. Multiple infusions of mesenchymal stromal cells induce sustained remission in children with steroid-refractory, grade III-IV acute graft-versus-host disease. *Br J Haematol*. 2013;163:501.
151. Chen S, Liu Z, Tian N, et al. Intracoronary transplantation of autologous bone marrow mesenchymal stem cells for ischemic cardiomyopathy due to isolated chronic occluded left anterior descending artery. *J Invasive Cardiol*. 2006;18:552.
152. Venkataramana NK, Kumar SK, Balaraju S, et al. Open-labeled study of unilateral autologous bone-marrow-derived mesenchymal stem cell transplantation in Parkinson's disease. *Transl Res*. 2010;155:62.
153. Bang OY, Lee JS, Lee PH, et al. Autologous mesenchymal stem cell transplantation in stroke patients. *Ann Neurol*. 2005;57:874.
154. Lee JS, Hong JM, Moon GJ, et al. A long-term follow-up study of intravenous autologous mesenchymal stem cell transplantation in patients with ischemic stroke. *Stem Cell*. 2010;28:1099.
155. Sun L, Akiyama K, Zhang H, et al. Mesenchymal stem cell transplantation reverses multiorgan dysfunction in systemic lupus erythematosus mice and humans. *Stem Cell*. 2009;27:1421.
156. Dash NR, Dash SN, Routray P, et al. Targeting nonhealing ulcers of lower extremity in human through autologous bone marrow-derived mesenchymal stem cells. *Rejuvenation Res*. 2009;12:359.
157. Pal R, Venkataramana NK, Bansal A, et al. Ex vivo-expanded autologous bone marrow-derived mesenchymal stromal cells in human spinal cord injury/paraplegia: a pilot clinical study. *Cytotherapy*. 2009;11:897.
158. Wakitani S, Okabe T, Horibe S, et al. Safety of autologous bone marrow-derived mesenchymal stem cell transplantation for cartilage repair in 41 patients with 45 joints followed for up to 11 years and 5 months. *J Tissue Eng Regen Med*. 2011;5:146.
159. Lin CS, Xin ZC, Deng CH, et al. Recent advances in andrology-related stem cell research. *Asian J Androl*. 2008;10:171.
160. Lin CS. Advances in stem cell therapy for the lower urinary tract. *World J Stem Cells*. 2010;2:1.
161. Bochinski D, Lin GT, Nunes L, et al. The effect of neural embryonic stem cell therapy in a rat model of cavernosal nerve injury. *BJU Int*. 2004;94:904.
162. Kendirci M, Trost L, Bakondi B, et al. Transplantation of nonhematopoietic adult bone marrow stem/progenitor cells isolated by p75 nerve growth factor receptor into the penis rescues erectile function in a rat model of cavernous nerve injury. *J Urol*. 2010;184:1560.
163. Zhang H, Yang R, Wang Z, et al. Adipose tissue-derived stem cells secrete CXCL5 cytokine with neurotrophic effects on cavernous nerve regeneration. *J Sex Med*. 2011;8:437.
164. Soebadi MA, Moris L, Castiglione F, et al. Advances in stem cell research for the treatment of male sexual dysfunctions. *Curr Opin Urol*. 2016;26:129.
165. Albersen M, Lin CS, Lue T. Stem-cell therapy for erectile dysfunction. *Arab J Urol*. 2013;11:237.
166. Hatzimouratidis K, Eardley I, Giuliano F, et al. EAU guidelines on penile curvature. *Eur Urol*. 2012;62:543.
167. Zhou S, Schmelz A, Seufferlein T, et al. Molecular mechanisms of low intensity pulsed ultrasound in human skin fibroblasts. *J Biol Chem*. 2004;279:54463.
168. Kusuyama J, Bandow K, Shamoto M, et al. Low intensity pulsed ultrasound (LIPUS) influences the multilineage differentiation of mesenchymal stem and progenitor cell lines through ROCK-Cot/Tpl2-MEK-ERK signaling pathway. *J Biol Chem*. 2014;289:10330.
169. Li H, Matheu MP, Sun F, et al. Low-energy shock wave therapy ameliorates erectile dysfunction in a pelvic neurovascular injuries rat model. *J Sex Med*. 2016;13:22.
170. Vardi Y, Appel B, Jacob G, et al. Can low-intensity extracorporeal shockwave therapy improve erectile function? A 6-month follow-up pilot study in patients with organic erectile dysfunction. *Eur Urol*. 2010;58:243.
171. Liu J, Zhou F, Li GY, et al. Evaluation of the effect of different doses of low energy shock wave therapy on the erectile function of streptozotocin (STZ)-induced diabetic rats. *Int J Mol Sci*. 2013;14:10661.
172. Qiu X, Lin G, Xin Z, et al. Effects of low-energy shockwave therapy on the erectile function and tissue of a diabetic rat model. *J Sex Med*. 2013;10:738.
173. Campbell JD, Milenkovic U, Usta MF, et al. The good, bad, and the ugly of regenerative therapies for erectile dysfunction. *Transl Androl Urol*. 2020;9:S252.

110

Misuse and Abuse of Anabolic Hormones

Thiago Gagliano-Jucá and Shehzad Basaria

OUTLINE

AAS: Use, Misuse, and Abuse, 1846
  Use, 1846
  Misuse, 1846
  Abuse, 1846
History and Epidemiology of Anabolic Steroid Abuse, 1847
  History of Anabolic Steroid Abuse, 1847
  Epidemiology of Anabolic Steroid Abuse, 1848
Ergogenic Potential of Anabolic Steroids, 1848
Types and Patterns of Androgen Doping, 1848
  Direct Doping, 1849
  Indirect Doping, 1850
Detection of Androgen Doping, 1850
  Detection of Compounds Used in Direct Androgen Doping, 1850
  Detection of Compounds Used in Indirect Androgen Doping, 1852
  Evaluation Process of AAS Abuse in Elite Athletes, 1852
  Assessment of Men Abusing AAS in the Clinics, 1852
Strategies Used by Athletes to Avoid Detection of Androgen Abuse, 1853
  Administration of Epitestosterone, 1853
  Masking Agents, 1853
Health Consequences of Anabolic Steroid Abuse, 1853
  Cardiovascular, 1853
  Reproductive, 1855
  Gynecomastia, 1855
  Hepatic, 1855
  Musculoskeletal, 1855
  Renal, 1855
  Dermatologic, 1856
  Neuropsychiatric and Behavioral, 1856
  Dependence, 1856
Management Options of Anabolic Steroid Abuse, 1856
  Elite Athletes, 1856
  General Public, 1856
Conclusion, 1857

Anabolic androgenic steroids (AAS) are a group of steroid hormones, including testosterone and its synthetic derivatives, that have been in use for the past eight decades. In addition to their anabolic effects on the skeletal muscles and bone, androgens promote and maintain male secondary sexual characteristics, sexual function, and sexual behavior, and are used in the treatment of male androgen deficiency. However, there is an important distinction between treatment with testosterone at physiologic doses in organic male hypogonadism and abuse of AAS as pharmacotherapy for the purposes of achieving competitive advantage in sports or for “image enhancement” for cosmetic reasons. Indeed, few topics in medicine have generated more interest and controversy, both among the clinicians and lay public, than abuse of AAS for nonmedical reasons. This controversy is further compounded by the fact that the long-term safety of these steroids remains unclear. Additionally, there are no effective strategies to treat men who have discontinued or plan to discontinue AAS and are either experiencing or are concerned about AAS-withdrawal syndrome. The paucity of high-quality data in these areas foments the disregard toward medical advice related to adverse effects associated with AAS abuse,[1,2] with users often assuming that the physicians are misinformed regarding the consequences of AAS abuse.[3]

This chapter focuses on the pharmacological use of AAS in men, including (1) the history and epidemiology of AAS abuse, (2) types of doping, (3) methods used in the detection of AAS, (4) adverse health consequences of AAS abuse, and (5) management options in men who are rendered hypogonadal as a result of prior AAS abuse.

AAS: USE, MISUSE, AND ABUSE

At the beginning of this chapter, we highlight various terms associated with the use of AAS.

Use

The *use* of AAS implies taking prescription androgens (mainly testosterone) for medical conditions. An example is the treatment of hypogonadal men with organic diseases of the hypothalamus, pituitary, or testes. Testosterone therapy, at physiologic dose, is indicated to induce or maintain secondary sexual characteristics.

Misuse

The *misuse* constitutes systematic prescribing of AAS for unproven medical indications in men who do not have organic androgen deficiency. This includes prescribing androgens to treat obesity, diabetes, impaired cognition, or reverse aging.

Abuse

The *abuse* of AAS constitutes their use for nonmedical purposes where they are often used in supraphysiologic doses for either the purposes of performance enhancement (competitive sports) or image

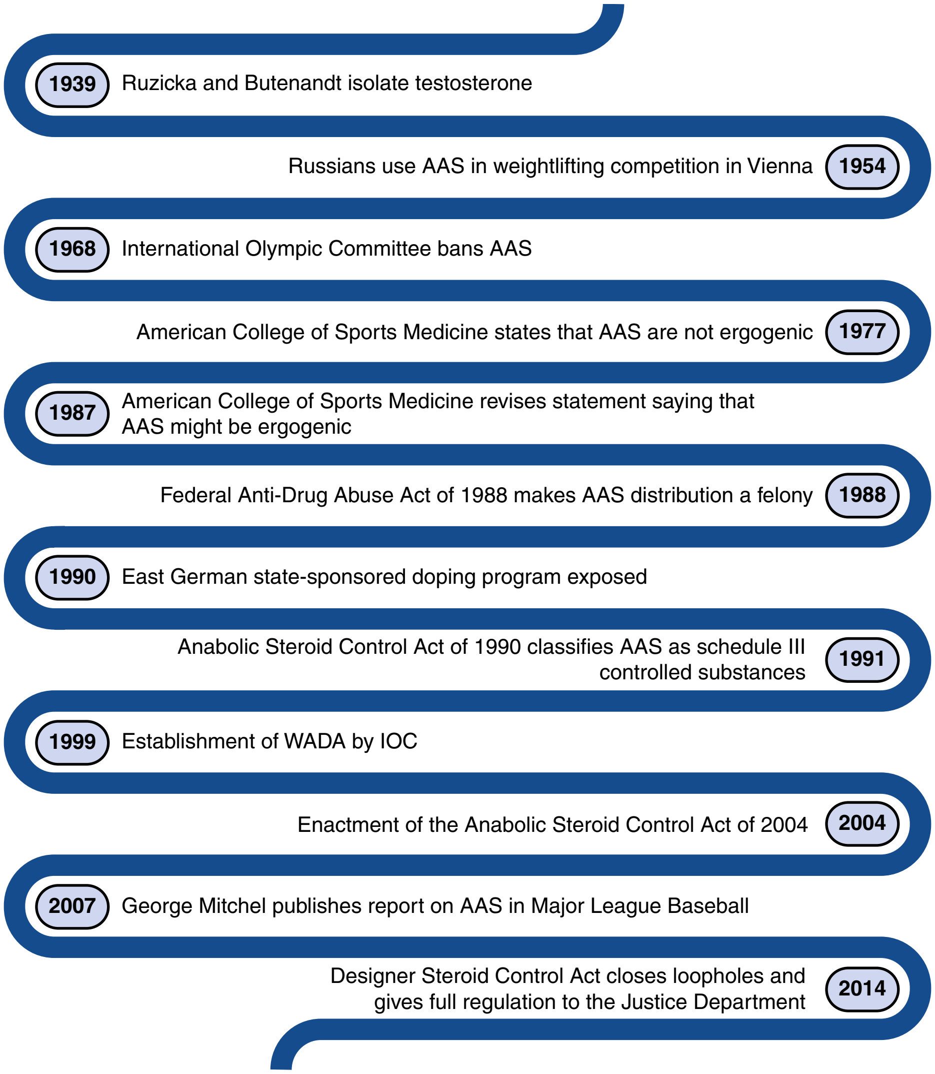

1939 Ruzicka and Butenandt isolate testosterone
1954 Russians use AAS in weightlifting competition in Vienna
1968 International Olympic Committee bans AAS
1977 American College of Sports Medicine states that AAS are not ergogenic
1987 American College of Sports Medicine revises statement saying that AAS might be ergogenic
1988 Federal Anti-Drug Abuse Act of 1988 makes AAS distribution a felony
1990 East German state-sponsored doping program exposed
1991 Anabolic Steroid Control Act of 1990 classifies AAS as schedule III controlled substances
1999 Establishment of WADA by IOC
2004 Enactment of the Anabolic Steroid Control Act of 2004
2007 George Mitchel publishes report on AAS in Major League Baseball
2014 Designer Steroid Control Act closes loopholes and gives full regulation to the Justice Department

Figure 110.1 Timeline of important events in the discovery and abuse of anabolic androgenic steroids (AAS). *IOC,* International Olympic Committee; *WADA,* World Anti-Doping Agency.

enhancement. In these situations, AAS are usually procured illicitly from compounding pharmacies, fitness centers, or veterinary clinics. This chapter will focus on this patient population.

HISTORY AND EPIDEMIOLOGY OF ANABOLIC STEROID ABUSE

History of Anabolic Steroid Abuse

The culture of using ergogenic (performance-enhancing) compounds among athletes dates to ancient Greece, where herbal remedies and animal extracts were used by athletes to gain competitive edge. Thus, the history of “doping” began long before the discovery of testosterone by Butenandt in the 1930s. However, the attribution of athletic prowess to testicular secretion dates back to the 18th century when experiments by John Hunter showed that testicular transplantation in any anatomical location maintained cock combs.[4] In the 1870s, Charles Edouard Brown-Séquard, a 72-year-old physician-scientist, hypothesized that injections of testicular extracts from young animals possessed rejuvenating effects and decided to use himself as the first human subject. He self-administered 10 doses of this extract subcutaneously over 2 weeks and reported improvement in strength and endurance.[5] The news of Brown-Séquard’s “Elixir of Life” quickly spread throughout Europe and North America, where it was commercially sold for decades. However, replication of Brown-Séquard’s published methods demonstrated that his elixir had insufficient concentrations of testosterone to produce a biological response[6]; thus, the birth of Andrology began with demonstration of the *placebo effect*. Six decades later, Butenandt and Ruzicka won the Nobel Prize for the synthesis of testosterone. Subsequently, chemists and compounding pharmacies made numerous derivatives of testosterone in the 1950s. Although powerlifters were the first group of athletes to abuse AAS for competitions (Fig. 110.1),[7] AAS abuse soon spread to other athletes, including those participating in the Olympic Games.[8] The most well-known phase of AAS abuse in the history of Olympics was the systematic administration of AAS to the Soviet weightlifting team in the 1952 and 1956 Olympic Games.[8] In the United States, the credit for introducing AAS to American athletes goes to Dr. John Ziegler, a physician member of the US Weightlifting Team, who learned about the use of AAS by the Russian team during his trip to weightlifting championships in Vienna in 1954.[9] Upon return, he administered testosterone to weightlifters in the York Babel Club in Pennsylvania. In the 1960s, the German Democratic Republic instituted a program of systematic doping of their Olympic athletes with AAS, which led to winning 153 Olympic gold medals from 1968 to 1988, despite having a population of only 18 million.[7]

While athletes were well aware of the ergogenic effects of AAS, the scientific community and the general public remained skeptical. Although the International Olympic Committee (IOC) prohibited the use of AAS in 1974, a position paper from the American College of Sports Medicine in 1977 still raised doubts regarding the ergogenic potential of AAS.[10] It was not until 1996 that the effects of AAS on strength were demonstrated in a well-controlled study.[11] Currently, the scrutiny of doping in sports is conducted by the World Anti-Doping Agency (WADA), a subsidiary of IOC. Every year WADA publishes an ever-growing list of prohibited drugs; however, androgens remain the most commonly used ergogenic drugs in sports.[7]

Epidemiology of Anabolic Steroid Abuse

The epidemiology of AAS abuse in athletes and in the general public has been difficult to ascertain.[12,13] Questionnaire-based studies of elite college athletes in the United States have demonstrated a lifetime prevalence of 20%. In one frequently cited study, Goldman et al. presented a Faustian bargain to world-class athletes biannually between 1982 and 1995, asking if they would agree to take a drug that would guarantee an Olympic gold medal without being caught, but which would also result in their death within 5 years (The Goldman Dilemma).[14] Interestingly, 52% of the athletes said they would accept the "death-for-gold" deal. In 2013, another study revisited the Goldman Dilemma in 212 elite track athletes and found that 12% would still take a hypothetical illegal performance-enhancing drug if it was undetectable and guaranteed an Olympic gold medal, but only 2% would accept the *death-for-gold* deal.[15]

In the general American public, questionnaire-based studies have suggested lifetime prevalence of up to 15%,[13] with 98% of users being men.[16] It is estimated that 3 to 4 million males between age 13 and 50 years in the United States have used AAS. The lifetime prevalence of AAS use is higher among gay and bisexual men (21%) as well as among ethnic minorities (25%).[17] Over the decades, the abuse of AAS has permeated beyond competitive sports, as elite athletes account for only a small fraction of AAS users in the United States.[16] Most of the users are noncompetitive weightlifters and men who are keen to acquire muscular appearance for image-enhancing purposes. The desire to become muscular among some men is influenced by exposure to muscular male images in mass media while others may have muscle dysmorphia, a body image disorder that manifests as an obsessive preoccupation with a muscular appearance. It is estimated that approximately 45% of men with muscle dysmorphia may have used AAS.[18]

Worldwide, the lifetime prevalence of AAS abuse among men is estimated to be 1% to 5%. Although most Western cultures currently have an endemic problem with AAS abuse, East Asian cultures, such as Chinese and Japanese cultures, appear to be largely shielded from this public health issue; this is likely due to the fact that Confucian tradition associates masculinity with intellect and integrity rather than physical strength.[19] In contrast, in the Nordic culture, where physical strength has been valued since the origins of Norse mythology, the lifetime prevalence of AAS abuse is high at 2.4% in Norwegians and 4.4% among the Swedes.[20]

KEY POINTS

- AAS abuse, initially limited to elite athletes, has now permeated the general population where it is mainly used for the purposes of image enhancement.
- The lifetime prevalence of AAS abuse in the general population is estimated to be 5%, 98% of these being men.

ERGOGENIC POTENTIAL OF ANABOLIC STEROIDS

While elite and amateur athletes had been aware of the anabolic effects of AAS on skeletal muscles for decades, the scientific community continued to attribute ergogenic effects of AAS solely to their influence on motivation, arguing that only men who trained harder and longer while taking AAS experienced improved performance. This skepticism started to abate in 1987 when the American College of Sports Medicine revised its original conclusion that AAS did not increase muscle strength.[21] However, it was only in 1996 that the medical community fully accepted the anabolic effects of AAS after a well-controlled study conclusively demonstrated that administration of supraphysiologic doses of testosterone (600 mg/week) increased muscle strength in young healthy men; this study also showed that the effects of testosterone and exercise were additive.[11] Men who only received testosterone experienced an increase in fat-free mass of 3.2 kg compared to the placebo plus exercise group who had an increase of 1.9 kg. On the other hand, the increase in fat-free mass in the testosterone plus exercise group was 6.1 kg.[11] The synergistic effects of testosterone and resistance training on muscle strength have since been verified in other studies.[22] Similarly, studies have shown that suppression of endogenous testosterone in men reduces muscle mass and strength.[23]

The anabolic actions of AAS on skeletal muscle involve several mechanisms that include: (1) muscle fiber hyperplasia and hypertrophy, (2) differentiation of mesenchymal pluripotent stem cells toward myogenic rather than adipogenic lineage,[24] (3) stimulation of skeletal muscle protein synthesis,[25,26] and (4) reduction of muscle catabolism via inhibition of the ubiquitin proteasome and myostatin pathways.[25,27] Testosterone administration also leads to the activation of muscle satellite cells.[28,29]

In addition to their direct effects via androgen receptor (AR), AAS also exert their anabolic effects via stimulation of growth hormone (GH) secretion. Indeed, testosterone replacement in hypogonadal men has been shown to increase serum concentrations of GH[30] and an increase in intramuscular insulin-like growth factor-1 (IGF-1) mRNA expression. Additionally, AAS may also improve physical performance via stimulation of erythropoiesis, which results in an increase in oxygen-carrying capacity and endurance.

KEY POINTS

- The ergogenic effects of AAS are due to their direct anabolic action on skeletal muscles where they stimulate muscle fiber hyperplasia and hypertrophy, promote differentiation of mesenchymal pluripotent stem cells toward myogenic lineage, and increase protein incorporation.
- The end result is an increase in muscle mass and strength. Additionally, stimulation of erythropoiesis by androgens increases endurance.

TYPES AND PATTERNS OF ANDROGEN DOPING

Men who abuse AAS rely on two main doping strategies (Table 110.1).

Indirect Doping: This involves use of drugs that increase endogenous testosterone production.

Direct Doping: This includes direct use of both natural and synthetic steroids (including testosterone).

In addition to traditional AAS, there has been a remarkable proliferation of new nonsteroidal anabolic agents that are increasingly being used in conjunction with traditional AAS, such as selective androgen receptor modulators (SARMs).[31–33] The practice of abusing multiple anabolic agents simultaneously is known as "stacking," which involves

TABLE 110.1 Drugs Used in Direct and Indirect Doping

Direct Doping

Naturally Occurring Anabolic Steroids
Testosterone
Epitestosterone
Dihydrotestosterone
Nandrolone
Boldenone

Synthetic Anabolic Steroids
Clostebol
Danazol
Drostanolone
Ethylestrenol
Fluoxymesterone
Gestrinone
Methandienone
Methenolone
Methyltestosterone
Norethandrolone
Oxandrolone
Oxymetholone
Stanozolol
Tibolone
Trenbolone

Designer Steroids
Boldione (precursor to boldenone)
Dimethazine
Mentabolan
Methasteron
Methoxygonadiene
Methylclostebol
Methylepistiostanol
Methylstenbone
Methylstenbolone
Prostanozolol (precursor to stanozolol)

Selective Androgen Receptor Modulators
GTx-024
LGD-4033
RAD140
S-4
S-23
YK11

Indirect Doping

Testosterone Precursors
DHEA
Androstenedione

Selective Estrogen Receptor Modulators
Clomiphene
Droloxifene
Raloxifene
Tamoxifen
Toremifene

Gonadotropins
Human chorionic gonadotropin
Recombinant human LH

Aromatase inhibitors
Anastrozole
Letrozole
Vorozole
Atamestane
Exemestane
Testolactone

Selective Estrogen Receptor Degraders
Fulvestrant

mixing a variety of oral and injectable drugs that are considered to have synergistic effects on muscle mass and strength. Some men follow the practice of "cycling," with planned periods of AAS use ("on") followed by abstinence ("off"). It is believed that the "off" period prevents desensitization of AR (an unproven concept). Some users apply the "blast and cruise" technique, in which continuous "blasts" of high-dose AAS use is interspaced with lower-dose "cruise."

Direct Doping

This type of doping involves the direct use of testosterone and its steroid derivatives. Additionally, other compounds, such as SARMs (designed to have selective effects on some but not all tissues), are also increasingly being used. The following are the classes of drugs used in direct doping.

Naturally Occurring AAS. The naturally occurring steroids commonly used in direct doping include testosterone, nandrolone, and boldenone. Testosterone esters are by far the most commonly abused AAS.[34–36] The physiological dose of testosterone replacement in hypogonadal men is 100 mg every week intramuscularly; men who abuse testosterone may take as much as 6000 mg weekly. Testosterone is aromatized to estradiol by the aromatase enzyme complex via two successive hydroxylations on the C19 methyl group of testosterone; *Nandrolone* (19-nortestosterone) is a penultimate step in the aromatization process,[37] thus making it a naturally occurring AAS. Exogenous nandrolone is poorly aromatized, but it upregulates aromatase expression.[38] Therefore, men "stacking" nandrolone with testosterone are more likely to develop gynecomastia. Nandrolone is also 5α-reduced, a process that increases its androgenic effects. Nandrolone was previously used medically in the treatment of anemia of chronic kidney disease[39] but is no longer manufactured in the United States. However, it still remains a popular AAS, both for the purposes of performance and image enhancement.[34,36] *Boldenone*, also known as equipoise, is a naturally occurring androgen in animals that can rarely be detected in urine samples of nondoping athletes.[40] Boldenone is currently marketed for veterinary purposes and is among the most abused AAS.

Synthetic Androgenic Steroids. Shortly after the isolation and synthesis of testosterone, several synthetic AAS were developed for clinical use (Table 110.1). Among the most commonly abused synthetic AAS are stanozolol, oxandrolone, trenbolone, and methandienone.[34,36] Some of these steroids are still used for select medical conditions. For example, oxandrolone is often used as an anabolic agent in patients with severe burn injuries to counter catabolism[41] and in patients with Turner syndrome receiving GH therapy. However, the remaining synthetic AAS developed are no longer used for medical purposes.

Designer Steroids. A "designer steroid" is an AAS that is synthesized from a known parent compound and is chemically modified with the intent to circumvent substance control laws.[42] These agents were initially sold legally as supplements until 2015 but were labeled as controlled substances after passing of the Designer Anabolic Steroid Control Act of 2014.[43] The modifications in these compounds are also thought to alter the anabolic and androgenic effects of these steroids. Norbolethone has been credited as the first designer androgen that was originally discovered in the 1960s and its chemical signature was identified in 2002.[44] A year later, tetrahydrogestrinone, a derivative of gestrinone (a progestin), was identified.[45] In 2005, a third designer androgen, desoxymethyltestosterone, was identified. These compounds are banned by the World Anti-Doping Agency (WADA), though their ergogenic potential remains unclear.

SARMs. For decades, scientists have been on a quest to dissect out the anabolic properties of androgens from their androgenic potential. With the discovery of SARMs, this endeavor came to fruition, as these compounds achieved tissue selectivity by exerting anabolic effects on the skeletal muscle and bone while sparing the prostate. This tissue selectivity is considered important in the elderly, who are at a higher risk of prostate abnormalities.[46] The first SARM was synthesized in 1998 and derived from AR antagonists, flutamide and bicalutamide. Over the years, four groups of SARMs have been synthesized: (1) arylpropionamides, (2) bicyclic hydantoins, (3) quinolines, and (4) tetrahydroquinolines.

Indirect Doping

Indirect doping denotes use of drugs that increase endogenous testosterone secretion. This can be achieved with compounds that (1) stimulate the hypothalamic-pituitary-gonadal (HPG) axis such as estrogen receptor (ER) antagonists and aromatase inhibitors; (2) directly stimulate the testes, including luteinizing hormone (LH) and its analogs like human chorionic gonadotropin (hCG); and (3) testosterone precursors such as dehydroepiandrosterone (DHEA) and androstenedione (Table 110.1).

Estrogen Blockers. The rationale for using ER antagonists and aromatase inhibitors stems from the fact that estradiol is a more potent negative regulator of the HPG axis than testosterone.[47] This is evident in men with congenital aromatase deficiency in whom serum concentrations of estradiol are undetectable but both serum testosterone and gonadotropins are elevated; estrogen administration suppresses the HPG axis in these patients.[48] Aromatase is a product of the *CYP19* gene and converts testosterone to estradiol. Aromatase inhibitors inhibit aromatase activity, thereby decreasing estradiol synthesis; as a result, serum testosterone concentrations increase by ~50% to 65%,[49] which can still induce myotrophic effects. Among steroidal aromatase inhibitors, aminoglutethimide was the first drug discovered in this class, followed by testolactone, atamestane, and exemestane. Nonsteroidal aromatase inhibitors include anastrozole, letrozole, and vorozole. The ER antagonists increase gonadotropins by blocking the negative feedback of estradiol on the hypothalamus and the pituitary, which in turn increases serum testosterone concentrations. The original antiestrogens were nonsteroidal drugs such as clomiphene and tamoxifen that bind to both ERα and ERβ. Newer nonsteroidal agents have since become available such as raloxifene, toremifene, and droloxifene. In a trial of clomiphene administered for 30 days to healthy eugonadal young men, there was a progressive increase in serum testosterone concentrations by ~146% by day 30, well into the supraphysiological range.[50] Notably, testosterone concentrations were still increasing at the end of the planned intervention period of 30 days.[50] Thus, it remains unclear if a longer treatment duration could result in even higher testosterone levels. Both ER antagonists and aromatase inhibitors are on the WADA's list of banned drugs.

Men who abuse aromatizable AAS commonly "stack" ER antagonists or aromatase inhibitors to avoid development of gynecomastia.[34,35,51] Some men use these drugs after completing a cycle of AAS in an attempt to "hasten" the recovery of the HPG axis, known as "post-cycle therapy."[35]

Gonadotropins. Another form of indirect doping is the use of gonadotropins, such as recombinant LH or hCG, to stimulate the testes directly, bypassing the hypothalamus and the pituitary. Both drugs are currently on the WADA's list of prohibited drugs for male athletes. hCG is a dimeric glycoprotein containing an α and a β subunit that is normally produced by the human placenta[52]; its α-subunit is similar to other glycoprotein hormones, namely, follicle-stimulating hormone (FSH), LH, and thyroid-stimulating hormone (TSH). It undergoes glycosylation with sialic acid residues that prolongs its half-life and makes it a long-acting analog of LH. Clinically, recombinant LH or hCG are used to stimulate spermatogenesis in men who have central hypogonadism and desire fertility.[53] These therapies are expensive and require injections several times per week, but the evidence regarding their ergogenic potential is sparse. Small trials in older men have shown that hCG therapy increases serum testosterone levels and lean body mass but does not improve muscle strength.[54] Men abusing AAS also use hCG to prevent testicular atrophy.

Androgen Precursors. This class of agents include drugs like androstenedione and DHEA. Androstenedione is a prohormone that is produced in both the gonads and the adrenals in both sexes. It is synthesized from DHEA and is converted to testosterone by 17β-hydroxysteroid dehydrogenase. On the other hand, DHEA is predominantly secreted by the adrenal glands. Although the IOC had banned these compounds in the late 1990s, they were sold in an uncontrolled fashion as dietary supplements for the subsequent two decades under the Dietary Supplement Health and Education Act until 2004, when the US Congress added androstenedione to the list of banned drugs.[55] Data suggest that androstenedione at a dose of 300 mg increases serum testosterone by 30%,[56] whereas a dose of 1500 mg improves lean body mass and muscle strength. Similarly, DHEA at high doses (100 mg) improves muscle strength.[57] Although banning of these prohormones by the IOC is justified because of their ergogenic potential, there is still no conclusive evidence that testosterone precursors significantly impact the performance of young athletes.

KEY POINTS

- Direct doping involves the use of testosterone, its synthetic derivatives, and designer androgens (such as SARMs).
- Indirect doping involves the use of drugs that increase endogenous testosterone secretion by stimulating the HPG axis (ER antagonists and aromatase inhibitors) or via direct stimulation of the testes (LH and hCG).

DETECTION OF ANDROGEN DOPING

Detection of Compounds Used in Direct Androgen Doping

Although some athletes use both natural and synthetic AAS to circumvent the detection of synthetic androgens, athletes have resorted to doping mainly with *testosterone*. Hence, the detection of illegal use depends upon distinguishing between endogenous and exogenous testosterone. Detection of exogenous testosterone abuse can be accomplished by two methods: measurement of testosterone-to-epitestosterone (T/E) ratio and isotope ratio mass spectrometry (IRMS), the latter is only performed by specialized laboratories working with antidoping agencies to test elite athletes (not available to clinicians).

Testosterone-to-Epitestosterone Ratio. Epitestosterone (17α-hydroxy-4-androsten-3-one), a 17-epimer of testosterone, is also secreted by the Leydig cells. Epitestosterone is biologically inactive, and there is no interconversion between testosterone and epitestosterone. It is mainly excreted in the urine as a glucuronide.[58] Its production rate is <5% of testosterone, but its urinary excretion is 33% that of testosterone.[59] Because of this rapid clearance, excretion rates of testosterone and epitestosterone are similar, resulting in urinary T/E ratio of ~1. Because there is no interconversion between the two compounds, administration of exogenous testosterone results in an increase in the T/E ratio. Assessment of T/E ratio in a large number of athletes has shown that generally the ratio remains below 6.[60] This initially led the IOC to adopt the cutoff threshold of 6 to consider a test as "positive." However, it was later recognized that genetic differences in various populations influences T/E ratio as some athletes have lower rates of endogenous epitestosterone production; thus, their T/E ratio naturally exceeds 6. On the other side of the spectrum, a deletion polymorphism in uridine diphosphate glucuronyl transferase 2B17, an enzyme that facilitates epitestosterone excretion, lowers T/E ratios.[61] Considering these genetic variants, the IOC recently lowered the T/E cutoff ratio to 4. Hence, any value above 4 is considered as "suspicious." To complicate

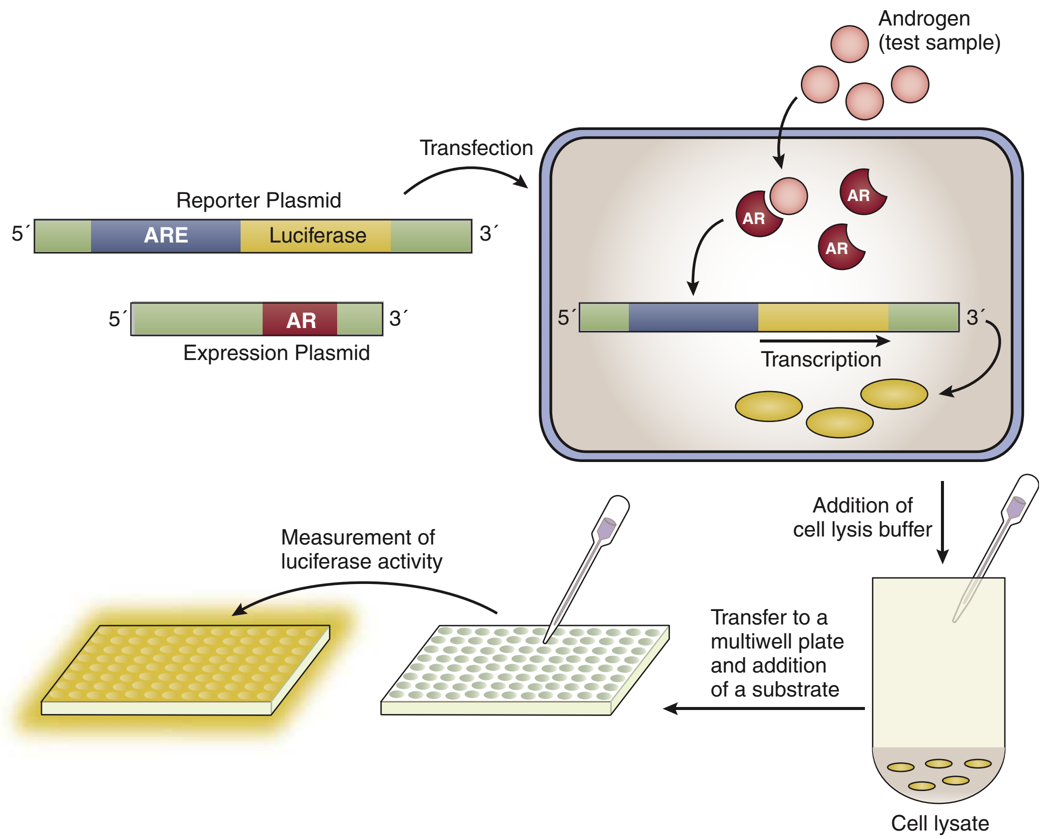


Figure 110.2 Detection of designer androgens by the luciferase reporter gene bioassay. *AR,* androgen receptor; *ARE,* androgen response element.

matters, some athletes may mask the use of exogenous testosterone by coadministration of epitestosterone.[62] In such cases, the exogenous use of epitestosterone can be detected by elevated concentrations of its urinary metabolites. Generally, the T/E ratio remains largely unchanged in athletes not using AAS. Therefore, serial T/E ratio measurement in the same individual is an important tool that aids in the detection of exogenous administration of testosterone.

Isotope Ratio Mass Spectrometry (IRMS). If the T/E ratio is abnormal among elite athletes, the WADA requires additional confirmation by gas chromatography-combustion-IRMS, which involves measurement of $^{13}C/^{12}C$ isotope ratio in testosterone. This method is based on the premise that the percentage of ^{13}C (a naturally occurring isotope of carbon) in endogenous testosterone (synthesized from carbon sources derived from animals and plants) is higher than in synthetic testosterone (synthetic testosterone has a lower $^{13}C/^{12}C$ ratio). The IRMS involves separation of steroids by gas chromatography and oxidized to carbon dioxide in a combustion chamber. The ratio of $^{13}CO_2/^{12}CO_2$ is measured in a spectrometer.[63] A lower $^{13}C/^{12}C$ ratio suggests exogenous testosterone use. This technology can also distinguish endogenous nandrolone, dihydrotestosterone (DHT), and DHEA from exogenous sources.

Synthetic Androgen Bioassays. Bioassays are increasingly being used in the detection of AAS. Bioassays are functional assays that are used to determine the bioactivity (and potency) of a compound, with the unit of bioactivity defined as the lowest concentration that generates a functional response. This unit is then compared with the known bioactivity of a standard compound (testosterone or DHT). The main utility of these bioassays is to detect novel designer androgens, compounds whose chemical structure is unknown (radioimmunoassay and mass spectrometry only detect AAS whose chemical signature is known). These bioassays are available in laboratories associated with doping agencies that monitor elite athletes and are not available to clinicians. Bioassays can be divided into *in vivo* and *in vitro* assays. The *in vivo* assays assess the response of androgenic stimulation on androgen-dependent tissues such as capon comb, skeletal muscles, or seminal vesicles. As these assays are time consuming, the *in vitro* assays are used more commonly as they are faster, relatively inexpensive, and have high throughput. There are three types of *in vitro* assays:

1. **Receptor Binding Assays:** These assays evaluate the ability of an androgen (test sample) to bind to the AR in the presence of a known radioactive ligand for the receptor. The displacement of the ligand from the AR by the test sample correlates with its bioavailability.
2. **Cell Proliferation Assays:** These assays assess the bioactivity of an unknown AAS by its ability to stimulate proliferation of androgen-dependent cell lines, e.g., human prostate cancer cells. These assays are sensitive but not specific for androgens.
3. **Reporter Gene Assays:** In these assays, the bioactivity of AAS is measured by the degree of expression of the reporter gene (Fig. 110.2).[64] These assays are performed in yeast or mammalian cells. The process involves selecting a host cell line that does not express an endogenous AR. The next step involves introduction of two plasmids into the cell line: (1) an expression plasmid for constitutive expression of AR in the cells and (2) a reporter plasmid in which the androgen response element sequences drive the expression of firefly luciferase that serves as the reporter gene.[64] The amount of luciferase expression is proportional to the bioactivity of the AAS being tested and is measured by luminometry. These bioassays have been optimized for high-throughput screening and can detect a variety

of designer steroids but are mainly available in specialized laboratories working with antidoping agencies.

Ketoconazole. Ketoconazole, an antifungal agent, is an inhibitor of endogenous steroid synthesis, including testosterone and epitestosterone.[75] Because there is no interconversion between testosterone and epitestosterone and because epitestosterone is more readily excreted in urine than testosterone, ketoconazole increases T/E ratio in men using exogenous testosterone (as it inhibits testicular epitestosterone secretion). In contrast, in men with naturally occurring high urinary T/E ratios, ketoconazole does not affect the T/E ratio as it inhibits secretion of both testosterone and epitestosterone.[76] This forms the basis for the ketoconazole test, which is used to differentiate exogenous from endogenous testosterone in men who present with an elevated T/E ratio.

Detection of Compounds Used in Indirect Androgen Doping

High-performance immunoassays are used to detect LH in men suspected of using it for indirect doping. The detection of hCG is initially performed with immunoassays, and if positive, confirmation is carried out with immune extraction and mass spectrometry.[65] As hCG does not alter T/E ratio, this methodology is not useful in the detection of indirect doping with hCG. Detection of doping with aromatase inhibitors and ER antagonists is performed by liquid or gas chromatography mass spectrometry (GC-MS).[66,67] Use of DHEA and androstenedione result in an increase in T/E ratio,[68,69] which is then confirmed by measurement of $^{13}C/^{12}C$ ratio with IRMS. Confirmation of doping with androstenedione can also be accomplished by measurement of its metabolite, 4-hydroxyandrostenedione, by either mass spectrometry or IRMS.[70] Use of high doses of testosterone precursors also alters the ratios of metabolites of endogenous androgens. For example, administration of DHEA increases urinary androsterone and results in a high androsterone/testosterone ratio.[68]

KEY POINTS

- Detection of doping with exogenous testosterone can be accomplished by measuring the testosterone-to-epitestosterone (T/E) ratio or via IRMS.
- Bioassays can aid in the detection of AAS whose molecular structure is not known.
- Indirect doping with LH or hCG can be detected by immunoassays.

Evaluation Process of AAS Abuse in Elite Athletes

In this section, we elaborate on the procedures that involve the collection and transportation of urine samples for doping tests in elite athletes. The WADA requires all urine samples to be collected under direct visual oversight of an accredited supervisor. After collection, each sample is divided into two parts (samples A and B) before transporting them to a WADA-accredited laboratory. Confirmatory testing consists of remeasurement of any suspicious analytes in an additional aliquot of the athlete's original "A" sample; the "B" sample is used as necessary for additional confirmatory testing if the "A" sample tests positive for a suspicious substance. If the second sample is also positive, then doping is confirmed. This information is then conveyed to the appropriate sports organization. In 2006, approximately 50% of the positive doping results were due to androgens, which were three times higher than the next category of banned drugs. As a result, WADA conducts screening for AAS using mass spectrometry to detect even trace quantities of natural or synthetic androgens.

An important tool in the detection of AAS abuse (and other performance-enhancing drugs) has been the "biological passport." The concept of a biological passport assumes that an individual's serum and urine hormone and prohormone concentrations do not change significantly over time.[71] An athlete's urine and serum testosterone, DHT, epitestosterone, and their precursors are measured before the first competition and are followed longitudinally. Additional testing follows any deviation from baseline that is considered significant. The biological passport is intended to detect attempts by athletes to use exogenous AAS or attempts to increase endogenous AAS by drugs that stimulate the HPG axis and increase endogenous production of AAS. The shortcomings of this process is that it assumes that the athletes have not used AAS prior to the first collection of samples at the time of enrollment. Therefore, this initial collection of samples may not reflect the "true" baseline status in such athletes.[72] The laboratory will also assess for the presence of compounds that athletes might have used to mask measurement of AAS. Based on the initial result, the laboratory might perform gas chromatography-combustion-IRMS. The final steps of the process of determination of whether an athlete has used an AAS involves expert interpretation of the laboratory results and comparison to previous results in the athlete's biological passport. Determination of an AAS use violation requires a unanimous conclusion of "doping likely" by three independent experts.

TABLE 110.2 Effect of Different Doping Strategies on Serum Hormone Concentrations

	LH and FSH	Testosterone	Estradiol
Indirect Doping			
SERMs	↑	↑	↑
Aromatase inhibitors	↑	↑	↓
hCG	↓	↑	↑
Testosterone precursors	↓	↑	↑
Direct Doping			
Testosterone	↓	↑	↑
Nontestosterone AAS	↓	↓	↓

Assessment of Men Abusing AAS in the Clinics

Unlike the laboratory services available to antidoping agencies, clinicians do not have access to these tools. Thus, they must rely on their clinical impression and limited assays provided by the commercial laboratories. Abuse of AAS should be considered when a muscular man presents either with low testosterone, infertility, or gynecomastia. Many men who abuse testosterone will give themselves a "drug holiday" from AAS and present to the clinic to seek testosterone treatment for hypogonadal symptoms at the time when their endogenous testosterone is still suppressed (men abusing SARMs or AAS other than testosterone will present with this biochemical profile while still taking these drugs). This incongruence between their muscular phenotype and suppressed serum testosterone levels is often the first indication of AAS abuse. If AAS abuse is suspected, the clinician should directly ask the patient about AAS abuse in a nonjudgmental manner. Clinicians should consider measurement of serum testosterone, FSH, and LH concentrations. As exogenous testosterone, nontestosterone AAS and hCG suppress circulating gonadotropins (Table 110.2), measurement of these hormones is useful in determining the likelihood of AAS abuse. Additionally, elevated hematocrit, low serum high-density lipoprotein (HDL) cholesterol and sex hormone–binding globulin (SHBG) concentrations also offer a clue to AAS abuse.

STRATEGIES USED BY ATHLETES TO AVOID DETECTION OF ANDROGEN ABUSE

Athletes use several methods to avoid detection of AAS, from urine replacement or adulteration to administration of masking agents. The following includes specific examples of strategies that are used to avoid detection:

Administration of Epitestosterone

The urinary T/E ratio is used to detect doping with exogenous testosterone and the test is considered positive when the T/E ratio is >4. Athletes doping with testosterone may concurrently administer epitestosterone, which decreases T/E ratio and results in a false-negative result.[62] As urinary excretion of epitestosterone is about one-third that of testosterone, coadministration of testosterone and epitestosterone in 30:1 ratio will generally result in a normal T/E ratio.[73] Deviations in the T/E ratio from an athlete's biological passport trigger further investigation but in itself do not constitute a doping violation. These urine samples can be submitted for IRMS confirmation.

Masking Agents

The WADA's list of prohibited substances also includes drugs that are used to mask detection of prohibited agents. Masking agents belong to different classes of medications and have different mechanisms of action. In this section, we will discuss such agents and how they mask detection of AAS.

Probenecid. Probenecid is a uricosuric agent that is used in the treatment of gout. Probenecid is known to interfere with renal excretion of steroid glucuronide conjugates and has been shown to reduce urinary excretion of endogenous (testosterone, epitestosterone, androsterone) and exogenous (norandrosterone, noretiocholanolone) steroid conjugates.[73] Probenecid is detected by GC-MS, and its identification alone is considered sufficient evidence to establish a doping violation.

Steroid 5α-Reductase Inhibitors. These drugs inhibit 5α-reductase, the enzyme that inhibits conversion of testosterone to DHT, and are used in the treatment of benign prostatic hyperplasia and male-pattern baldness. Endogenously produced 5α-reduced steroids (DHT and androsterone) are part of the biological passport in athletes. Concomitant use of finasteride with testosterone alters the urinary steroid profile by reducing the excretion of 5α-reduced steroids without changing the excretion of corresponding 5β-reduced steroids, resulting in a decreased ratio between epimeric 5α- and 5β-reduced steroids. If a 5α-reduced metabolite is targeted for testing, the use of finasteride can mask its detection. Similarly, finasteride inhibits the formation of 19-norandrosterone in nandrolone abuse. Finasteride can be measured by GC-MS and high-performance liquid chromatography (HPLC).

Diuretics. There are several classes of diuretics used in clinical practice to treat hypertension and hypervolemic states. Athletes use diuretics to achieve rapid weight loss in sports with weight restrictions such as boxing and wrestling. Diuretics are also effective masking agents as they increase urine volume and dilute urinary concentrations of AAS and their metabolites to either below the limits of detection or below allowed limits.[74]

KEY POINTS

- Athletes use several methods to avoid detection of AAS, including administration of epitestosterone and use of masking agents such as probenecid, 5α-reductase inhibitors, and diuretics.

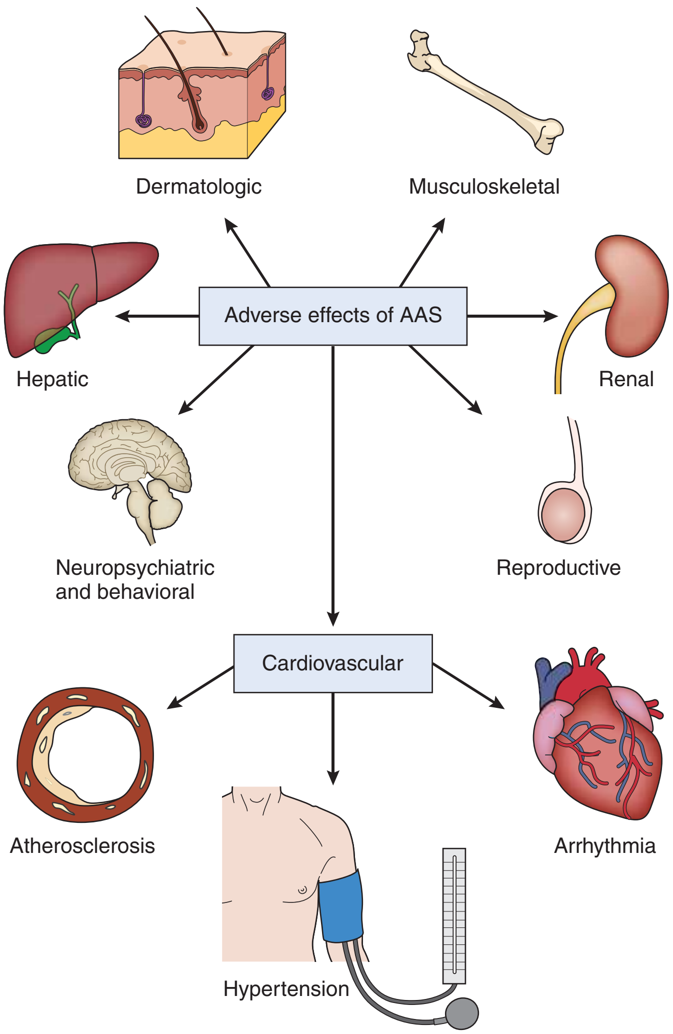


Figure 110.3 Adverse effects of anabolic androgenic steroids (AAS) by physiological systems.

HEALTH CONSEQUENCES OF ANABOLIC STEROID ABUSE

High-quality data on adverse health effects of AAS are scarce as there are no systematic prospective studies in individuals abusing AAS. Additionally, the current secrecy surrounding AAS abuse further confounds information from retrospective studies. Thus, the majority of the data on the adverse effects of AAS are derived from case reports, case series and cross-sectional case-control studies. Furthermore, there is evidence that many men who abuse AAS also indulge in other illicit drugs,[77] which could potentially contribute to the observed adverse effects. Additionally, these men often take high doses of dietary supplements and nutraceuticals that have not been carefully studied for safety or content.[77] As the widespread use of AAS in the general population only began 30 years ago, most abusers are still relatively young and may not have yet experienced adverse effects from AAS abuse.[78] Thus, most healthcare providers have not yet seen a large number of patients with complications of AAS abuse but will likely see them in their clinics as this cohort gets older. Below is the summary of the potential adverse effects of AAS across various organ systems (Fig. 110.3).

Cardiovascular

The AR is expressed in the human myocardium, vascular smooth muscle, and endothelium[79]; therefore, the abuse of AAS can have an impact on several aspects of human cardiovascular physiology. Indeed, a growing body of evidence supports that AAS abuse directly or indirectly impacts cardiovascular morbidity.

Blood Pressure: Supraphysiological doses of AAS have been associated with the incident or exacerbation of hypertension via multiple mechanisms. These include (1) inhibition of extraneuronal uptake of catecholamines,[80] (2) salt and water retention by inhibition of 11β-hydroxylase activity (which results in cortisol-induced activation of the mineralocorticoid receptor),[81] and (3) reduced aortic elasticity. Studies have shown that normalization of blood pressure after discontinuation of AAS may take up to a year.[82,83] To the contrary, some studies have not found such an association.[84] This lack of consensus is likely a result of the observational design and small sample size of these studies.

Dyslipidemia: AAS abuse is known to result in dyslipidemia. The most common observation is a 20% to 70% reduction in HDL cholesterol[85]; this reduction is more profound with the use of oral non-aromatizable AAS. In addition, HDL-mediated cholesterol efflux (a measure of HDL functionality) is also reduced in men who abuse AAS. This is important as HDL cholesterol efflux capacity is inversely associated with subclinical cardiovascular risk in young adults. Reductions in apolipoprotein A1[86] and increases in low-density lipoprotein (LDL) cholesterol and triglycerides have also been observed.[85] This atherogenic lipid profile may develop within a few weeks of AAS abuse and may take several months to resolve after their discontinuation.

Erythrocytosis: Erythrocytosis is the most common adverse effect of androgen use[87] and can result in increased blood viscosity and blood-flow resistance.[88] *In vitro* studies have shown a direct correlation between hematocrit and platelet aggregation. Mechanisms by which androgens stimulate erythrocytosis include stimulation of erythropoietin secretion, increasing iron availability by suppression of hepcidin, and direct stimulation of erythroid elements in the bone marrow.[89] Elevation of hematocrit above 52% may result in thromboembolic events.[85,90]

Myocardial Function: Long-term AAS abuse influences cardiac remodeling. Autopsy studies in young men who have abused AAS have shown left ventricle hypertrophy and increased overall cardiac mass.[91] These *post mortem* findings are corroborated by echocardiographic studies that show impairment of cardiac function in individuals with long-term abuse of AAS.[92,93] In a recent report, young men with at least 2 years of exposure to supraphysiologic doses of AAS had impaired systolic and diastolic function on echocardiography compared with experienced weightlifters reporting no exposure to AAS.[93] Cardiac dysfunction was worse in men who were still taking AAS at the time of evaluation compared with previous users, suggesting that discontinuation of AAS at least partially restores cardiac function (Fig. 110.4).[93] Systolic dysfunction with AAS abuse has also been observed by other authors.[92] However, as left ventricular hypertrophy is considered a physiologic adaptation to endurance and resistance training,[94–96] it is difficult to differentiate between the impact of exercise training versus the direct effect of AAS on this remodeling. Long-term exposure to high doses of AAS results in a decrease in myocardial contractile reserve to β-adrenoceptor stimulation,[97] which likely contributes to cardiac dysfunction. Furthermore, subtle histological changes in the myocardium as a result of AAS abuse are seen even before the onset of overt ventricular hypertrophy.[98]

Arrhythmias: Long-term abuse of AAS may lead to myocardial scarring and fibrosis even in individuals without coronary artery disease,[99] possibly by inducing cardiomyocyte apoptosis. Myocardial fibrosis alters cardiac electrophysiology predisposing to lethal arrhythmias.[100] Indeed, animal studies have suggested that nandrolone can potentiate the arrhythmogenic effects of cardiac ischemia by altering cardiomyocyte calcium handling. Arrhythmias in AAS abusers often occur during or shortly after physical activity[101] and include atrial fibrillation, supraventricular or ventricular tachycardia, and ventricular fibrillation.[102] Arrhythmias appear to be a common cause of sudden death in young AAS abusers, as some autopsy studies did not find evidence of either atherosclerosis or thrombosis in coronary arteries of these patients.[103,104]

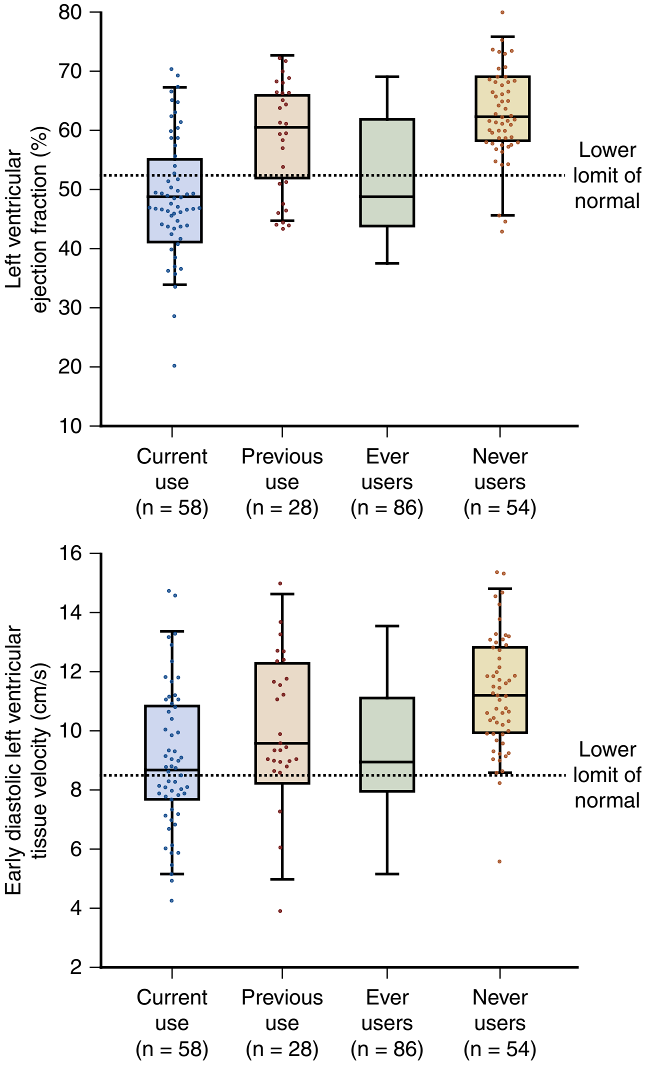


Figure 110.4 Left ventricular systolic and diastolic function in anabolic-androgenic steroid users and nonusers. (Adapted from Baggish AL, Weiner RB, Kanayama G, et al. Cardiovascular toxicity of illicit anabolic-androgenic steroid use. *Circulation* 2017;135:1991–2002.)

Atherosclerosis: Abuse of AAS has been associated with accelerated atherosclerosis. Men who abuse AAS have greater coronary artery plaque volume on computed tomography coronary angiography compared with nonusers, with greater plaque burden associated with cumulative AAS exposure.[93,105] These findings are corroborated by the numerous case reports of myocardial infarction in young men abusing AAS, which is likely a result of plaque rupture and thrombus formation.[91,106] Many cases of stroke in young AAS abusers have also been reported; however, most are thought to be due to embolic events.[107,108] Indeed, testosterone treatment has been shown to augment *ex vivo* platelet aggregability by increasing thromboxane A2 receptor density on human platelets.[109] Androgens also decrease production of prostacyclin, an inhibitor of platelet aggregation. Additionally, current AAS abusers have reduced fibrinolysis compared to nonusers and former users.[110]

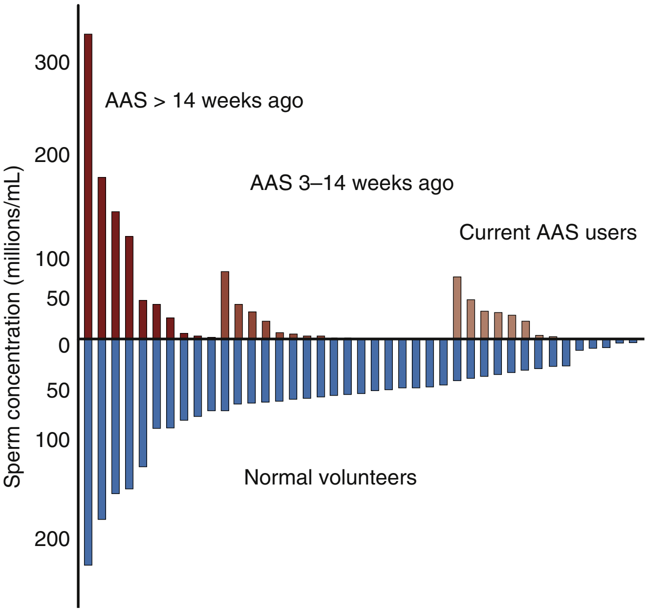


Figure 110.5 Sperm concentrations in 41 bodybuilders currently using anabolic androgenic steroids (AAS), those who discontinued AAS 3 to 14 weeks ago or discontinued >14 weeks ago *(red bars)* and in 41 drug-free volunteers *(blue bars)*. (Adapted from Nieschlag E, Vorona E. Mechanisms in endocrinology: medical consequences of doping with anabolic androgenic steroids: effects on reproductive functions. *Eur J Endocrinol* 2015;173:R47–R58.)

Reproductive

Considering that exogenous androgens suppress the HPG axis, it is not surprising that individuals involved in direct doping have suppression of testicular testosterone secretion and spermatogenesis. Although systematic evaluation of reproductive function in men who abuse AAS is lacking, data from studies of male hormonal contraception that used supraphysiologic doses of testosterone, usually in combination with a progestin, show profound oligozoospermia during the intervention phase.[111] Reassuringly, sperm concentrations returned to normal in 90% of participants within 12 months of cessation of intervention and in 100% within 24 months. The results of semen analysis in men abusing AAS show a spectrum that ranges from normal sperm count to azoospermia; the recovery of spermatogenesis is associated with time since last AAS (Fig. 110.5).[112,113] As the seminiferous tubules account for ~95% of the testicular volume, abuse of AAS often results in testicular atrophy. This explains why some AAS users choose to coadminister hCG to maintain testicular volume and spermatogenesis.

Sexual dysfunction is also a common adverse effect of AAS abuse.[78] As estrogens play an important role in male libido,[114] men who abuse nonaromatizable AAS experience low libido and erectile dysfunction as endogenous testosterone and estradiol are both suppressed. Additionally, even aromatizable synthetic AAS may not generate metabolites with similar estrogenic activity as estradiol, resulting in a detrimental effect on libido similar to nonaromatizable AAS. For these reasons, testosterone remains the most abused AAS, both among the athletes and general public. However, hypogonadism and sexual dysfunction are an integral part of the “AAS-withdrawal syndrome” that occurs upon discontinuation of all AAS,[115] and long-term abuse may result in prolonged androgen deficiency that may persist for months to years after discontinuation of AAS.[116]

Gynecomastia

Abuse of AAS may result in gynecomastia in some men, particularly those using aromatizable AAS in supraphysiologic doses. To prevent this, many men take ER antagonists or aromatase inhibitors. Gynecomastia is also seen in men who have persistent androgen deficiency after discontinuation of both aromatizable and nonaromatizable AAS.[117]

Hepatic

Hepatic adverse effects of AAS are generally limited to oral 17α-alkylated agents, such as methyltestosterone, methandrostenolone, oxandrolone, oxymetholone, and stanozolol. However, case reports of hepatic injury due to abuse of SARMs are emerging. The clinical assessment of AAS-induced hepatotoxicity poses some challenges. Rhabdomyolysis from high-intensity exercise can increase transaminases that may mislead clinicians into considering these elevations as a sign of hepatic injury.[78] To the contrary, AAS-induced hepatic injury sometimes does not result in an increase in transaminase levels.[118] Long-term use of 17α-alkylated agents reduces the activities of the microsomal drug-metabolizing enzyme system and the hepatic mitochondrial respiratory chain complex, and increases hepatic lysosomal hydrolases, without any increase in liver function tests. Macroscopic signs of AAS-induced hepatotoxicity are related to the cumulative dose and duration of abuse of 17α-alkylated steroids, and include cholestatic jaundice and hepatic peliosis (proliferation of sinusoidal capillaries that result in blood-filled cysts).[118] Data suggest that peliosis and hepatic adenomas are rare and not clinically apparent, with most cases found incidentally or on autopsy, although there have been rare reports of spontaneous hepatic rupture in men abusing AAS.[119]

Musculoskeletal

Compared with bodybuilders who do not abuse AAS, men abusing androgens have a much higher risk of tendon ruptures, mainly of the upper body.[120] This might be due to a rapid increase in skeletal muscle mass that outpaces adaptation by the connective tissues. This discrepancy in the speed of adaptation between the muscles and the tendons may also contribute to the higher patellar tendon stiffness observed in men with long-term AAS abuse. Some data also suggest altered metabolism of type-1 and type-3 collagen fibers in these men.[121]. Another important musculoskeletal adverse effect of AAS abuse is premature closure of the epiphyses in teenagers, which mainly occurs with the use of aromatizable AAS, resulting in reduced final height. Other musculoskeletal complications include myositis ossificans, a form of heterotopic ossification, that occurs at the site of intramuscular injections,[122] and generalized necrotizing myopathy, which rarely leads to severe rhabdomyolysis resulting in renal injury or multiorgan failure.

Renal

In addition to renal injury related to rhabdomyolysis, some cases of acute bile nephropathy secondary to AAS have been reported.[123] As previously discussed, oral AAS can cause cholestasis, increasing circulating concentrations of bile acids that lead to formation of bile casts in the renal tubules resulting in acute tubular injury; chronic exposure leads to progressive interstitial injury and tubulointerstitial fibrosis. Additionally, there have been reports of AAS-induced focal segmental glomerulosclerosis.[124] Though these serious forms of renal injury have been reported, the use of AAS generally presents as progressive subclinical deterioration of renal function. Indeed, a recent report that assessed creatinine and cystatin-C–derived estimates of glomerular filtration rate (eGFR) in weightlifters (57 current AAS users, 28 past users, and 52 never users) showed that eGFR was lowest in the current users, intermediate among the past users, and highest in nonusers.[125] Animal studies suggest that increases in oxidative stress and proinflammatory cytokines lead to renal cell apoptosis. Similarly, an increase in glomerular expression of CD34, a marker of endothelial dysfunction, has also been reported. Clinicians should be aware that higher muscle mass in these men is reflected by an increase in serum creatinine, which does not necessarily suggest impaired renal function. Thus, they should consider measurement of cystatin-C and its clearance to assess renal function.

Dermatologic

AR is expressed in the skin, and testosterone has been shown to dose-dependently increase sebum production.[126] Use of AAS affects the pilosebaceous unit and can lead to or exacerbate acne vulgaris and folliculitis; these conditions generally resolve upon discontinuation of AAS. Men abusing AAS may also develop *striae distensae* (stretch marks) due to the rapid increase in muscle size that outpaces expansion of the skin, likely due to reduced cutaneous elasticity as a result of AAS-induced reduced collagen production.[127] Keloid formation is also seen in some men.

Neuropsychiatric and Behavioral

Although there is a widespread notion that AAS abuse increases aggressive behavior, this is observed in a small number of individuals. In addition to experiencing neuropsychiatric symptoms during AAS abuse, androgen withdrawal is also associated with major depression, including suicidal ideation.[116] These mood changes might be related to a decrease in functional connectivity between the amygdala and the default mode network, as well as with other cerebral centers involved in control of emotions.[128] Moreover, reduced volumes of total gray matter, cerebral cortex, and putamen have been demonstrated in weightlifters with cumulative exposure to AAS of at least 1 year compared with weightlifters who never used androgens.[129] These data are also supported by preclinical studies showing that AAS induce neuronal apoptosis and reduce expression of nerve growth factor.

Dependence

Since the late 1980s, reports of dependence on AAS have been emerging, but only in recent years, these dependence syndromes have been widely recognized. Pooled data from 10 international studies suggest that 30% of AAS abusers may develop dependence—a prevalence much greater than seen for other recreational drugs.[16] Multiple factors likely contribute to the development of AAS dependence.[130] First, men with muscle dysmorphia may become "addicted" to AAS to maintain their physique as they fear losing muscle mass upon discontinuation of androgens. Second, AAS suppresses the HPG axis[115,116]; thus, upon discontinuation of AAS (especially after prolonged use), many men experience androgen deficiency. Recent data suggest that this androgen deficiency may persist for months or even years following cessation of AAS[116,131] and may result in sexual dysfunction, fatigue, and mood changes. Thus, many men resume AAS to avoid these adverse effects. Third, AAS may cause dependence via their hedonic properties. Although AAS use is not associated with intoxication that is associated with alcohol, opiates, or cocaine, their use is associated with feelings of self-confidence and "invincibility."[78] Animal data show that when male hamsters are given the opportunity to inject themselves with AAS via a nose poke device, they will self-administer AAS to the point of death.[132] This phenomenon has been demonstrated even in those animal studies where AAS were administered directly into the central nervous system, suggesting that the effects of AAS are central. All these factors might be responsible in predisposing men to dependence on AAS.

KEY POINTS

- Although high-quality data on adverse health consequences of AAS abuse are lacking, evidence from case reports, case series, and cohort studies suggest that AAS abuse may result in cardiovascular dysfunction, erythrocytosis, infertility, suppression of gonadal axis, gynecomastia, acne, and neuropsychological disturbances.

MANAGEMENT OPTIONS OF ANABOLIC STEROID ABUSE

Although the general principles of the medical management of AAS abuse are the same for both the elite athletes and the general public, there are some important differences that are highlighted in the section below.

Elite Athletes

Therapeutic Use Exemption: As in the general public, athletes may also have hormonal deficiencies for which they might be receiving appropriate hormone replacement. These include athletes with androgen deficiency due to organic diseases of the hypothalamus, pituitary, or the testes, such as Klinefelter's syndrome, chemotherapy-induced testicular injury, or central hypogonadism after hypophysectomy. For elite athletes participating in competitive sports in which AAS are prohibited, the process for obtaining permission to continue treatment from the relevant antidoping agency is known as Therapeutic Use Exemption (TUE). Clinicians, particularly endocrinologists, may be asked to verify that testosterone therapy is truly indicated for a particular athlete. The granting of a TUE for testosterone follows rigid guidelines, including documentation of specific signs and symptoms of androgen deficiency and unequivocally low morning testosterone concentrations on more than one occasion. For documented organic male androgen deficiency, the TUE after vigorous review can be granted for testosterone therapy and in some circumstances for gonadotropins.

If an elite athlete is found to be abusing AAS to purely achieve competitive advantage (i.e., not under TUE), the athlete is instructed to discontinue the use of the banned drug and may be blocked from competition for a proscribed period of time. These athletes are not given testosterone to mitigate the risk of AAS-withdrawal syndrome if they are still actively competing. On the other hand, a clinician may encounter retired elite male athletes who previously abused AAS while competing and now present with persistently low serum testosterone concentrations. Many of these men have low or inappropriately normal serum gonadotropins suggesting continued suppression of their HPG axis. The medical management of these retired elite athletes with hypogonadal symptoms is similar to members of the general public who have previously abused AAS (see below).

General Public

As there are no clinical trials on the management of long-term AAS abuse in the general public, care is based on the likelihood of severe withdrawal symptoms, the goals of the patient, and a risk-benefit analysis of the available treatment options.

Men Not Willing to Discontinue AAS: Some men who abuse AAS, either for competing in amateur sports or for image enhancement, are not willing to discontinue AAS. In this case, the main role of the endocrinologist is to ensure that the patient is fully informed about the known adverse effects of all types of AAS such as erythrocytosis, reduced fertility, dyslipidemia, and potential cardiovascular risks. Additionally, clinicians should counsel on the formulation-specific adverse effects of AAS (e.g., hepatotoxicity associated with 17α-alkylated steroids). In men who are already using supraphysiologic doses of testosterone, a prudent approach would be to gradually reduce the dose over the course of weeks to months to a physiologic replacement dose. This approach may help avoid severe symptoms of AAS withdrawal. On the other hand, men who are taking AAS other than testosterone could be switched to prescription testosterone. Based on the current dose of their AAS, replacement with testosterone can be initiated up to twice the usual replacement dose with a gradual taper to physiological dose over subsequent weeks. Anecdotal experience suggests that this approach (1) foments trust with the patient, (2) discourages purchase of AAS from unregulated sources, and (3) allows continuity of care with the physician. The clinician should continue to inform the patient that continued use of testosterone will result in continued suppression of the HPG axis.

Men Willing to Discontinue AAS: There are no randomized controlled trials comparing medical therapy versus no treatment on the degree of withdrawal symptoms related to discontinuation of AAS. Furthermore, available medical therapies are not Food and Drug Administration approved and are used in the clinics in an off-label fashion. For men who are willing to discontinue AAS, the management options include (1) discontinuation without additional medical therapy, (2) initiation of physiological testosterone replacement to avoid withdrawal symptoms (as mentioned earlier), and (3) short-term treatment with clomiphene or hCG. The selection of these options is dictated by the predominant reason that has led these men to discontinue AAS; in most cases, it is infertility. Clinicians should perform the usual workup indicated in the evaluation of infertility including measurement of serum testosterone and gonadotropins and perform semen analyses. If the working diagnosis of infertility is prior AAS abuse, then the choice of therapy is based on the couples' timeline for starting a family. If this is not an immediate goal, then the clinician could consider discontinuation of AAS with monitoring of serum testosterone and gonadotropins periodically. Once testosterone has normalized, serial semen analyses can be performed to assess resumption of spermatogenesis. This approach is based on the observation of spontaneous recovery of the HPG axis in trials of male contraception, where recovery of endogenous testosterone and spermatogenesis occurs within 3 and 12 months, respectively, after discontinuation of exogenous testosterone. Similarly, in men who have abused AAS for ≤1 year, recovery of the HPG axis occurs within a few months. To the contrary, among men who have abused high doses of AAS for >1 year, the recovery time is longer; such men might benefit from a short course of clomiphene or gonadotropin therapy after discontinuation of AAS.

Clomiphene Therapy: Clomiphene is a selective estrogen receptor modulator (SERM) that increases serum gonadotropins and testosterone concentrations within 2 to 4 weeks in men with an intact HPG axis.[133,134] Thus, the integrity of the HPG axis is a prerequisite for its efficacy. Limited data suggest that some men with a history of long-term AAS abuse experienced recovery of their HPG axis with clomiphene therapy.[135] The benefits of hastening the recovery of the HPG axis are of interest to men who are eager to have a child soon. Initiation of clomiphene at 25 to 50 mg every other day is a reasonable approach as it has been efficacious in normalizing serum testosterone concentrations. However, this benefit should be weighed against the potential side effects such as venous thrombosis and other cardiovascular events. The clinician should inform the patient that the safety and efficacy of this approach has not been shown in clinical trials.

hCG Therapy: Treatment with hCG is another option in patients with similar presentation. Subcutaneous hCG therapy stimulates testosterone production and spermatogenesis in almost all men with postpubertal hypogonadotropic hypogonadism. The increase in serum testosterone that results from administration of hCG suppresses gonadotropins; therefore, the HPG axis will remain suppressed after discontinuation of hCG. The usual dose of hCG is 1000 to 2000 IU subcutaneously 2 to 3 times weekly. The dose is adjusted until serum total testosterone is in the normal range and continued at least until conception. Data on the comparative effectiveness of clomiphene and hCG are lacking.

KEY POINTS

- Elite athletes abusing AAS should be advised to discontinue them, except in whom TUE has been granted.
- In the general population, men who abuse AAS and are unwilling to discontinue could be transitioned to physiologic testosterone replacement.
- Men who are willing to discontinue AAS could either be offered a drug holiday from testosterone, treatment with hCG, or a short-term trial with clomiphene (in an attempt to hasten the recovery of the gonadal axis).

CONCLUSION

The abuse of AAS continues to increase. Although legislative efforts have attempted to curtail access to these agents,[136] they continue to be easily procured through the Internet and other unregulated sources.[137,138] Although antidoping agencies tightly regulate doping among elite athletes, most men abusing AAS are not elite athletes but young men who want to enhance their image by improving their muscularity. Furthermore, the burden of adverse effects of these compounds will only become apparent when these men enter their middle age. Thus, tremendous amount of work needs to be done, including (1) organization of an education campaign to raise public awareness regarding the risks of AAS use, (2) collaboration between doping control laboratories and commercial laboratories to make doping tests available to clinicians, (3) creation of a steroid hormone user registry so that men abusing AAS can be followed systematically to assess doping practices and potential risks, and (4) clinical trials with ER antagonists and aromatase inhibitors to determine their efficacy in hastening the recovery of the HPG axis. This can be accomplished when patients, clinicians, researchers, and policy makers join hands and make a concerted effort toward achieving these objectives.

REFERENCES

1. Dawson RT. Drugs in sport - the role of the physician. *J Endocrinol*. 2001;170:55–61.
2. Brennan BP, Kanayama G, Pope Jr HG. Performance-enhancing drugs on the web: a growing public-health issue. *Am J Addict*. 2013;22:158–161.
3. Pope HG, Kanayama G, Ionescu-Pioggia M, et al. Anabolic steroid users' attitudes towards physicians. *Addiction*. 2004;99:1189–1194.
4. Sengoopta C. 'Dr Steinach coming to make old young!': sex glands, vasectomy and the quest for rejuvenation in the roaring twenties. *Endeavour*. 2003;27:122–126.
5. Brown-Séquard C. Note on the effects produced on man by subcutaneous injections of a liquid obtained from the testicles of animals. *Lancet*. 1889;134:105–107.
6. Cussons AJ, Bhagat CI, Fletcher SJ, et al. Brown-Sequard revisited: a lesson from history on the placebo effect of androgen treatment. *Med J Aust*. 2002;177:678–679.
7. Kanayama G, Pope Jr HG. History and epidemiology of anabolic androgens in athletes and non-athletes. *Mol Cell Endocrinol*. 2018;464:4–13.
8. Wade N. Anabolic steroids: doctors denounce them, but athletes aren't listening. *Science*. 1972;176:1399–1403.
9. Bahrke MS, Yesalis CE. Abuse of anabolic androgenic steroids and related substances in sport and exercise. *Curr Opin Pharmacol*. 2004;4:614–620.
10. American College of Sports Medicine. Position statement on the use and abuse of anabolic-androgenic steroids in sports. *Med Sci Sports*. 1977;9:xi–xii.
11. Bhasin S, Storer TW, Berman N, et al. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*. 1996;335:1–7.
12. de Hon O, Kuipers H, van Bottenburg M. Prevalence of doping use in elite sports: a review of numbers and methods. *Sports Med*. 2015;45:57–69.

13. Sagoe D, Molde H, Andreassen CS, et al. The global epidemiology of anabolic-androgenic steroid use: a meta-analysis and meta-regression analysis. *Ann Epidemiol*. 2014;24:383–398.
14. Goldman B, Bush PJ, Klatz R. *Death in the Locker Room*. London: Century; 1984.
15. Connor J, Woolf J, Mazanov J. Would they dope? Revisiting the Goldman dilemma. *Br J Sports Med*. 2013;47:697–700.
16. Pope Jr HG, Kanayama G, Athey A, et al. The lifetime prevalence of anabolic-androgenic steroid use and dependence in Americans: current best estimates. *Am J Addict*. 2014;23:371–377.
17. Blashill AJ, Safren SA. Sexual orientation and anabolic-androgenic steroids in U.S. adolescent boys. *Pediatrics*. 2014;133:469–475.
18. Pope Jr HG, Khalsa JH, Bhasin S. Body image disorders and abuse of anabolic-androgenic steroids among men. *JAMA*. 2017;317:23–24.
19. Yang CF, Gray P, Pope Jr HG. Male body image in Taiwan versus the west: Yanggang Zhiqi meets the Adonis complex. *Am J Psychiatry*. 2005;162:263–269.
20. Sagoe D, Tr T, Molde H, et al. Anabolic-androgenic steroid use in the Nordic countries: a meta-analysis and meta-regression analysis. *Nord Stud Alcohol Dr*. 2017;32:7–20.
21. American College of Sports Medicine. Position stand on the use of anabolic-androgenic steroids in sports. *Med Sci Sports Exerc*. 1987;19:534–539.
22. Storer TW, Magliano L, Woodhouse L, et al. Testosterone dose-dependently increases maximal voluntary strength and leg power, but does not affect fatigability or specific tension. *J Clin Endocrinol Metab*. 2003;88:1478–1485.
23. Basaria S, Lieb 2nd J, Tang AM, et al. Long-term effects of androgen deprivation therapy in prostate cancer patients. *Clin Endocrinol*. 2002;56:779–786.
24. Singh R, Artaza JN, Taylor WE, et al. Androgens stimulate myogenic differentiation and inhibit adipogenesis in C3H 10T1/2 pluripotent cells through an androgen receptor-mediated pathway. *Endocrinology*. 2003;144:5081–5088.
25. Rossetti ML, Steiner JL, Gordon BS. Androgen-mediated regulation of skeletal muscle protein balance. *Mol Cell Endocrinol*. 2017;447:35–44.
26. Ferrando AA, Sheffield-Moore M, Paddon-Jones D, et al. Differential anabolic effects of testosterone and amino acid feeding in older men. *J Clin Endocrinol Metab*. 2003;88:358–362.
27. Mendler L, Baka Z, Kovacs-Simon A, et al. Androgens negatively regulate myostatin expression in an androgen-dependent skeletal muscle. *Biochem Biophys Res Commun*. 2007;361:237–242.
28. Kadi F, Eriksson A, Holmner S, et al. Effects of anabolic steroids on the muscle cells of strength-trained athletes. *Med Sci Sports Exerc*. 1999;31:1528–1534.
29. Sinha-Hikim I, Roth SM, Lee MI, et al. Testosterone-induced muscle hypertrophy is associated with an increase in satellite cell number in healthy, young men. *Am J Physiol Endocrinol Metab*. 2003;285:E197–E205.
30. Illig R, Prader A. Effect of testosterone on growth hormone secretion in patients with anorchia and delayed puberty. *J Clin Endocrinol Metab*. 1970;30:615–618.
31. Krug O, Thomas A, Walpurgis K, et al. Identification of black market products and potential doping agents in Germany 2010-2013. *Eur J Clin Pharmacol*. 2014;70:1303–1311.
32. Thevis M, Lagojda A, Kuehne D, et al. Characterization of a non-approved selective androgen receptor modulator drug candidate sold via the Internet and identification of in vitro generated phase-I metabolites for human sports drug testing. *Rapid Commun Mass Spectrom*. 2015;29:991–999.
33. Van Wagoner RM, Eichner A, Bhasin S, et al. Chemical composition and labeling of substances marketed as selective androgen receptor modulators and sold via the Internet. *JAMA*. 2017;318:2004–2010.
34. Ip EJ, Barnett MJ, Tenerowicz MJ, et al. The Anabolic 500 survey: characteristics of male users versus nonusers of anabolic-androgenic steroids for strength training. *Pharmacotherapy*. 2011;31:757–766.
35. de Ronde W, Smit DL. Anabolic androgenic steroid abuse in young males. *Endocr Connect*. 2020;9(4):R102–R111.
36. Mullen C, Whalley BJ, Schifano F, et al. Anabolic androgenic steroid abuse in the United Kingdom: an update. *Br J Pharmacol*. 2020;177:2180–2198.
37. Simpson ER, Mahendroo MS, Means GD, et al. Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocr Rev*. 1994;15:342–355.
38. Sirianni R, Capparelli C, Chimento A, et al. Nandrolone and stanozolol upregulate aromatase expression and further increase IGF-I-dependent effects on MCF-7 breast cancer cell proliferation. *Mol Cell Endocrinol*. 2012;363:100–110.
39. Navarro JF. In the erythropoietin era, can we forget alternative or adjunctive therapies for renal anaemia management? The androgen example. *Nephrol Dial Transplant*. 2003;18:2222–2226.
40. Schanzer W. Metabolism of anabolic androgenic steroids. *Clin Chem*. 1996;42:1001–1020.
41. Ring J, Heinelt M, Sharma S, et al. Oxandrolone in the treatment of burn injuries: a systematic review and meta-analysis. *J Burn Care Res*. 2020;41:190–199.
42. Rahnema CD, Crosnoe LE, Kim ED. Designer steroids - over-the-counter supplements and their androgenic component: review of an increasing problem. *Andrology*. 2015;3:150–155.
43. Congress US. *Designer Anabolic Steroid Control Act of 2014*; 2014.
44. Catlin DH, Ahrens BD, Kucherova Y. Detection of norbolethone, an anabolic steroid never marketed, in athletes' urine. *Rapid Commun Mass Spectrom*. 2002;16:1273–1275.
45. Jasuja R, Catlin DH, Miller A, et al. Tetrahydrogestrinone is an androgenic steroid that stimulates androgen receptor-mediated, myogenic differentiation in C3H10T1/2 multipotent mesenchymal cells and promotes muscle accretion in orchidectomized male rats. *Endocrinology*. 2005;146:4472–4478.
46. Basaria S, Dobs AS. Risks versus benefits of testosterone therapy in elderly men. *Drugs Aging*. 1999;15:131–142.
47. Winters SJ, Janick JJ, Loriaux DL, et al. Studies on the role of sex steroids in the feedback control of gonadotropin concentrations in men. II. Use of the estrogen antagonist, clomiphene citrate. *J Clin Endocrinol Metab*. 1979;48:222–227.
48. Bilezikian JP, Morishima A, Bell J, et al. Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. *N Engl J Med*. 1998;339:599–603.
49. Leder BZ, Rohrer JL, Rubin SD, et al. Effects of aromatase inhibition in elderly men with low or borderline-low serum testosterone levels. *J Clin Endocrinol Metab*. 2004;89:1174–1180.
50. Miller GD, Moore C, Nair V, et al. Hypothalamic-pituitary-testicular axis effects and urinary detection following clomiphene administration in males. *J Clin Endocrinol Metab*. 2019;104:906–914.
51. Parkinson AB, Evans NA. Anabolic androgenic steroids: a survey of 500 users. *Med Sci Sports Exerc*. 2006;38:644–651.
52. Casarini L, Santi D, Brigante G, et al. Two hormones for one receptor: evolution, biochemistry, actions, and pathophysiology of LH and hCG. *Endocr Rev*. 2018;39:549–592.
53. Finkel DM, Phillips JL, Snyder PJ. Stimulation of spermatogenesis by gonadotropins in men with hypogonadotropic hypogonadism. *N Engl J Med*. 1985;313:651–655.
54. Liu PY, Wishart SM, Handelsman DJ. A double-blind, placebo-controlled, randomized clinical trial of recombinant human chorionic gonadotropin on muscle strength and physical function and activity in older men with partial age-related androgen deficiency. *J Clin Endocrinol Metab*. 2002;87:3125–3135.
55. Congress US. *Anabolic Steroid Control Act of 2004*; 2004.
56. Leder BZ, Longcope C, Catlin DH, et al. Oral androstenedione administration and serum testosterone concentrations in young men. *JAMA*. 2000;283:779–782.
57. Morales AJ, Haubrich RH, Hwang JY, et al. The effect of six months treatment with a 100 mg daily dose of dehydroepiandrosterone (DHEA) on circulating sex steroids, body composition and muscle strength in age-advanced men and women. *Clin Endocrinol*. 1998;49:421–432.
58. Starka L. Epitestosterone. *J Steroid Biochem Mol Biol*. 2003;87:27–34.

59. Wilson H, Lipsett MB. Metabolism of epitestosterone in man. *J Clin Endocrinol Metab*. 1966;26:902–914.
60. Basaria S. Androgen abuse in athletes: detection and consequences. *J Clin Endocrinol Metab*. 2010;95:1533–1543.
61. Schulze JJ, Lundmark J, Garle M, et al. Doping test results dependent on genotype of uridine diphospho-glucuronosyl transferase 2B17, the major enzyme for testosterone glucuronidation. *J Clin Endocrinol Metab*. 2008;93:2500–2506.
62. Dehennin L. Detection of simultaneous self-administration of testosterone and epitestosterone in healthy men. *Clin Chem*. 1994;40:106–109.
63. Shackleton CH, Phillips A, Chang T, et al. Confirming testosterone administration by isotope ratio mass spectrometric analysis of urinary androstanediols. *Steroids*. 1997;62:379–387.
64. Michelini E, Cevenini L, Mezzanotte L, et al. A sensitive recombinant cell-based bioluminescent assay for detection of androgen-like compounds. *Nat Protoc*. 2008;3:1895–1902.
65. Kicman AT, Parkin MC, Iles RK. An introduction to mass spectrometry based proteomics-detection and characterization of gonadotropins and related molecules. *Mol Cell Endocrinol*. 2007;260–262:212–227.
66. Mareck U, Sigmund G, Opfermann G, et al. Identification of the aromatase inhibitor letrozole in urine by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom*. 2005;19:3689–3693.
67. Mareck U, Geyer H, Guddat S, et al. Identification of the aromatase inhibitors anastrozole and exemestane in human urine using liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2006;20:1954–1962.
68. Dehennin L, Ferry M, Lafarge P, et al. Oral administration of dehydroepiandrosterone to healthy men: alteration of the urinary androgen profile and consequences for the detection of abuse in sport by gas chromatography-mass spectrometry. *Steroids*. 1998;63:80–87.
69. Bowers LD. Oral dehydroepiandrosterone supplementation can increase the testosterone/epitestosterone ratio. *Clin Chem*. 1999;45:295–297.
70. Cawley AT, Trout GJ, Kazlauskas R, et al. The detection of androstenedione abuse in sport: a mass spectrometry strategy to identify the 4-hydroxyandrostenedione metabolite. *Rapid Commun Mass Spectrom*. 2008;22:4147–4157.
71. Catlin DH, Hatton CK, Starcevic SH. Issues in detecting abuse of xenobiotic anabolic steroids and testosterone by analysis of athletes' urine. *Clin Chem*. 1997;43:1280–1288.
72. Evans NA. Gym and tonic: a profile of 100 male steroid users. *Br J Sports Med*. 1997;31:54–58.
73. van de Kerkhof DH, de Boer D, Thijssen JH, et al. Evaluation of testosterone/epitestosterone ratio influential factors as determined in doping analysis. *J Anal Toxicol*. 2000;24:102–115.
74. Cadwallader AB, de la Torre X, Tieri A, et al. The abuse of diuretics as performance-enhancing drugs and masking agents in sport doping: pharmacology, toxicology and analysis. *Br J Pharmacol*. 2010;161:1–16.
75. Sonino N. The use of ketoconazole as an inhibitor of steroid production. *N Engl J Med*. 1987;317:812–818.
76. Oftebro H, Jensen J, Mowinckel P, et al. Establishing a ketoconazole suppression test for verifying testosterone administration in the doping control of athletes. *J Clin Endocrinol Metab*. 1994;78:973–977.
77. Hakansson A, Mickelsson K, Wallin C, et al. Anabolic androgenic steroids in the general population: user characteristics and associations with substance use. *Eur Addict Res*. 2012;18:83–90.
78. Pope Jr HG, Wood RI, Rogol A, et al. Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. *Endocr Rev*. 2014;35:341–375.
79. Cai J, Hong Y, Weng C, et al. Androgen stimulates endothelial cell proliferation via an androgen receptor/VEGF/cyclin A-mediated mechanism. *Am J Physiol Heart Circ Physiol*. 2011;300:H1210–H1221.
80. Baker PJ, Ramey ER, Ramwell PW. Androgen-mediated sex differences of cardiovascular responses in rats. *Am J Physiol*. 1978;235:H242–H246.
81. Johannsson G, Gibney J, Wolthers T, et al. Independent and combined effects of testosterone and growth hormone on extracellular water in hypopituitary men. *J Clin Endocrinol Metab*. 2005;90:3989–3994.
82. Urhausen A, Albers T, Kindermann W. Are the cardiac effects of anabolic steroid abuse in strength athletes reversible? *Heart*. 2004;90:496–501.
83. Kuipers H, Wijnen JA, Hartgens F, et al. Influence of anabolic steroids on body composition, blood pressure, lipid profile and liver functions in body builders. *Int J Sports Med*. 1991;12:413–418.
84. Sader MA, Griffiths KA, McCredie RJ, et al. Androgenic anabolic steroids and arterial structure and function in male bodybuilders. *J Am Coll Cardiol*. 2001;37:224–230.
85. Achar S, Rostamian A, Narayan SM. Cardiac and metabolic effects of anabolic-androgenic steroid abuse on lipids, blood pressure, left ventricular dimensions, and rhythm. *Am J Cardiol*. 2010;106:893–901.
86. Singh AB, Hsia S, Alaupovic P, et al. The effects of varying doses of T on insulin sensitivity, plasma lipids, apolipoproteins, and C-reactive protein in healthy young men. *J Clin Endocrinol Metab*. 2002;87:136–143.
87. Bhasin S, Brito JP, Cunningham GR, et al. Testosterone therapy in men with hypogonadism: an endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103:1715–1744.
88. Guo W, Bachman E, Vogel J, et al. The effects of short-term and long-term testosterone supplementation on blood viscosity and erythrocyte deformability in healthy adult mice. *Endocrinology*. 2015;156:1623–1629.
89. Gagliano-Juca T, Pencina KM, Ganz T, et al. Mechanisms responsible for reduced erythropoiesis during androgen deprivation therapy in men with prostate cancer. *Am J Physiol Endocrinol Metab*. 2018;315:E1185–E1193.
90. Lippi G, Banfi G. Doping and thrombosis in sports. *Semin Thromb Hemost*. 2011;37:918–928.
91. Frati P, Busardo FP, Cipolloni L, et al. Anabolic Androgenic Steroid (AAS) related deaths: autoptic, histopathological and toxicological findings. *Curr Neuropharmacol*. 2015;13:146–159.
92. Rasmussen JJ, Schou M, Madsen PL, et al. Cardiac systolic dysfunction in past illicit users of anabolic androgenic steroids. *Am Heart J*. 2018;203:49–56.
93. Baggish AL, Weiner RB, Kanayama G, et al. Cardiovascular toxicity of illicit anabolic-androgenic steroid use. *Circulation*. 2017;135:1991–2002.
94. Scharf M, Oezdemir D, Schmid A, et al. Myocardial adaption to HI(R)T in previously untrained men with a randomized, longitudinal cardiac MR imaging study (Physical adaptions in Untrained on Strength and Heart trial, PUSH-trial). *PLoS One*. 2017;12:e0189204.
95. Pluim BM, Zwinderman AH, van der Laarse A, et al. The athlete's heart. A meta-analysis of cardiac structure and function. *Circulation*. 2000;101:336–344.
96. Dickerman RD, Schaller F, McConathy WJ. Left ventricular wall thickening does occur in elite power athletes with or without anabolic steroid Use. *Cardiology*. 1998;90:145–148.
97. Norton GR, Trifunovic B, Woodiwiss AJ. Attenuated beta-adrenoceptor-mediated cardiac contractile responses following androgenic steroid administration to sedentary rats. *Eur J Appl Physiol*. 2000;81:310–316.
98. Hassan NA, Salem MF, Sayed MA. Doping and effects of anabolic androgenic steroids on the heart: histological, ultrastructural, and echocardiographic assessment in strength athletes. *Hum Exp Toxicol*. 2009;28:273–283.
99. Baumann S, Jabbour C, Huseynov A, et al. Myocardial scar detected by cardiovascular magnetic resonance in a competitive bodybuilder with longstanding abuse of anabolic steroids. *Asian J Sports Med*. 2014;5:e24058.
100. Lusetti M, Licata M, Silingardi E, et al. Pathological changes in anabolic androgenic steroid users. *J Forensic Leg Med*. 2015;33:101–104.
101. Sculthorpe N, Grace F, Jones P, et al. Evidence of altered cardiac electrophysiology following prolonged androgenic anabolic steroid use. *Cardiovasc Toxicol*. 2010;10:239–243.
102. Nieminen MS, Ramo MP, Viitasalo M, et al. Serious cardiovascular side effects of large doses of anabolic steroids in weight lifters. *Eur Heart J*. 1996;17:1576–1583.
103. Kennedy MC, Lawrence C. Anabolic steroid abuse and cardiac death. *Med J Aust*. 1993;158:346–348.
104. Fineschi V, Riezzo I, Centini F, et al. Sudden cardiac death during anabolic steroid abuse: morphologic and toxicologic findings in two fatal cases of bodybuilders. *Int J Legal Med*. 2007;121:48–53.
105. Souza FR, Dos Santos MR, Porello RA, et al. Diminished cholesterol efflux mediated by HDL and coronary artery disease in young male anabolic androgenic steroid users. *Atherosclerosis*. 2019;283:100–105.

106. Ferenchick GS, Adelman S. Myocardial infarction associated with anabolic steroid use in a previously healthy 37-year-old weight lifter. *Am Heart J*. 1992;124:507–508.
107. Shimada Y, Yoritaka A, Tanaka Y, et al. Cerebral infarction in a young man using high-dose anabolic steroids. *J Stroke Cerebrovasc Dis*. 2012;21:906–911. e909.
108. Nieschlag E, Vorona E. Doping with anabolic androgenic steroids (AAS): adverse effects on non-reproductive organs and functions. *Rev Endocr Metab Disord*. 2015;16:199–211.
109. Ajayi AA, Mathur R, Halushka PV. Testosterone increases human platelet thromboxane A2 receptor density and aggregation responses. *Circulation*. 1995;91:2742–2747.
110. Sidelmann JJ, Gram JB, Rasmussen JJ, et al. Anabolic-androgenic steroid abuse impairs fibrin clot lysis. *Semin Thromb Hemost*. 2021;47(1):11–17.
111. Nieschlag E. Clinical trials in male hormonal contraception. *Contraception*. 2010;82:457–470.
112. Basaria S, Collins L, Dillon EL, et al. The safety, pharmacokinetics, and effects of LGD-4033, a novel nonsteroidal oral, selective androgen receptor modulator, in healthy young men. *J Gerontol A Biol Sci Med Sci*. 2013;68:87–95.
113. Christou MA, Christou PA, Markozannes G, et al. Effects of anabolic androgenic steroids on the reproductive system of athletes and recreational users: a systematic review and meta-analysis. *Sports Med*. 2017;47:1869–1883.
114. Finkelstein JS, Lee H, Burnett-Bowie SA, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*. 2013;369:1011–1022.
115. Coward RM, Rajanahally S, Kovac JR, et al. Anabolic steroid induced hypogonadism in young men. *J Urol*. 2013;190:2200–2205.
116. Kanayama G, Hudson JI, DeLuca J, et al. Prolonged hypogonadism in males following withdrawal from anabolic-androgenic steroids: an under-recognized problem. *Addiction*. 2015;110:823–831.
117. Narula HS, Carlson HE. Gynaecomastia--pathophysiology, diagnosis and treatment. *Nat Rev Endocrinol*. 2014;10:684–698.
118. Neri M, Bello S, Bonsignore A, et al. Anabolic androgenic steroids abuse and liver toxicity. *Mini Rev Med Chem*. 2011;11:430–437.
119. Patil JJ, O'Donohoe B, Loyden CF, et al. Near-fatal spontaneous hepatic rupture associated with anabolic androgenic steroid use: a case report. *Br J Sports Med*. 2007;41:462–463.
120. Kanayama G, DeLuca J, Meehan 3rd WP, et al. Ruptured tendons in anabolic-androgenic steroid users: a cross-sectional cohort study. *Am J Sports Med*. 2015;43:2638–2644.
121. Parssinen M, Karila T, Kovanen V, et al. The effect of supraphysiological doses of anabolic androgenic steroids on collagen metabolism. *Int J Sports Med*. 2000;21:406–411.
122. Schultzel MM, Johnson MH, Rosenthal HG. Bilateral deltoid myositis ossificans in a weightlifter using anabolic steroids. *Orthopedics*. 2014;37:e844–e847.
123. Luciano RL, Castano E, Moeckel G, et al. Bile acid nephropathy in a bodybuilder abusing an anabolic androgenic steroid. *Am J Kidney Dis*. 2014;64:473–476.
124. Herlitz LC, Markowitz GS, Farris AB, et al. Development of focal segmental glomerulosclerosis after anabolic steroid abuse. *J Am Soc Nephrol*. 2010;21:163–172.
125. Hudson JI, Kanayama G, Pope Jr HG, et al. Glomerular filtration rate and supraphysiologic-dose anabolic-androgenic steroid use: a cross-sectional cohort study. *Am J Kidney Dis*. 2020;76:152–155.
126. Bhasin S, Travison TG, Storer TW, et al. Effect of testosterone supplementation with and without a dual 5alpha-reductase inhibitor on fat-free mass in men with suppressed testosterone production: a randomized controlled trial. *J Am Med Assoc*. 2012;307:931–939.
127. Walker J, Adams B. Cutaneous manifestations of anabolic-androgenic steroid use in athletes. *Int J Dermatol*. 2009;48:1044–1048. quiz 1048.
128. Westlye LT, Kaufmann T, Alnaes D, et al. Brain connectivity aberrations in anabolic-androgenic steroid users. *Neuroimage Clin*. 2017;13:62–69.
129. Bjornebekk A, Walhovd KB, Jorstad ML, et al. Structural brain imaging of long-term anabolic-androgenic steroid users and nonusing weightlifters. *Biol Psychiatry*. 2017;82:294–302.
130. Kanayama G, Brower KJ, Wood RI, et al. Treatment of anabolic-androgenic steroid dependence: emerging evidence and its implications. *Drug Alcohol Depend*. 2010;109:6–13.
131. Rasmussen JJ, Selmer C, Ostergren PB, et al. Former abusers of anabolic androgenic steroids exhibit decreased testosterone levels and hypogonadal symptoms years after cessation: a case-control study. *PLoS One*. 2016;11:e0161208.
132. Wood RI. Anabolic-androgenic steroid dependence? Insights from animals and humans. *Front Neuroendocrinol*. 2008;29:490–506.
133. Handelsman DJ. Clinical review: the rationale for banning human chorionic gonadotropin and estrogen blockers in sport. *J Clin Endocrinol Metab*. 2006;91:1646–1653.
134. Habous M, Giona S, Tealab A, et al. Clomiphene citrate and human chorionic gonadotropin are both effective in restoring testosterone in hypogonadism: a short-course randomized study. *BJU Int*. 2018;122:889–897.
135. Guay AT, Jacobson J, Perez JB, et al. Clomiphene increases free testosterone levels in men with both secondary hypogonadism and erectile dysfunction: who does and does not benefit? *Int J Impot Res*. 2003;15:156–165.
136. Goldman AL, Pope HG, Bhasin S. The health threat posed by the hidden epidemic of anabolic steroid use and body image disorders among young men. *J Clin Endocrinol Metab*. 2019;104:1069–1074.
137. McBride JA, Carson 3rd CC, Coward RM. The availability and acquisition of illicit anabolic androgenic steroids and testosterone preparations on the Internet. *Am J Men's Health*. 2018;12:1352–1357.
138. Westerman ME, Charchenko CM, Ziegelmann MJ, et al. Heavy testosterone use among bodybuilders: an uncommon cohort of illicit substance users. *Mayo Clin Proc*. 2016;91:175–182.
139. Nieschlag E, Vorona E. Mechanisms in endocrinology: medical consequences of doping with anabolic androgenic steroids: effects on reproductive functions. *Eur J Endocrinol*. 2015;173:R47–R58.

111

Regulation of Spermatogenesis

Kate L. Loveland, Liza O'Donnell, David de Kretser, Robin M. Hobbs, and Mark P. Hedger

OUTLINE

Macroscopic Organization of the Adult Human Testis, 1861
- The Seminiferous Tubules, 1861
- The Intertubular Tissue, 1861
- Postnatal Growth, 1862

Cellular Development and Organization for Functional Control, 1862
- The Spermatogenic Epithelium, 1862
- Sertoli Cells, 1862
- Spermatogonia and Spermatogonial Stem Cells, 1864
- The Spermatogonial Stem Cell Niche, 1865
- Morphological Features of Spermatogonia in Primate and Human, 1866
- Spermatogonial Differentiation and Its Control, 1866
- Meiosis, 1866
- Spermiogenesis: Emergence of Spermatozoa, 1867
- Germ Cell Associations and the Spermatogenic Cycle, 1869
- Spermatogenic Efficiency and Germ Cell Loss, 1869
- Genetic Defects and Spermatogenesis, 1870
- Epigenetics in the Male Germline, 1870

The Intertubular Compartment, 1870
- Leydig Cells, 1871
- Other Interstitial Cells, 1871

The Immunologic Environment of the Testis, 1871

Endocrine and Local Control of Spermatogenesis, 1871
- Interactions Between the Testis, Hypothalamus, and Pituitary Gland, 1871

Control of Sertoli Cell Function, 1871
- Sertoli Cells During Fetal and Postnatal Development, 1872
- Control of Spermatogenesis: Challenges Ahead, 1874

MACROSCOPIC ORGANIZATION OF THE ADULT HUMAN TESTIS

The adult human testis is 4- to 5-cm long, 15 to 35 mL in weight, enclosed on anterolateral and medial surfaces by serous membranes of the tunica vaginalis. Posteriorly, the associated epididymis and spermatic cord include the vas deferens and testicular neurovascular pedicle. Vessels and autonomic nerves from renal and abdominal aortic plexuses traverse the tunica albuginea, a 1-mm-thick, collagen-rich capsule deep within the tunica vaginalis.

The testicular artery enters posteriorly, descending to the inferior pole under the tunica albuginea to form the tunica vasculosa. In human testes, the main artery stem ascends beneath the anterior surface before smaller branches penetrate the parenchyma. Posteriorly, the thickened tunica albuginea projects into the testis parenchyma, forming the honeycomb-like mediastinum through which the rete testis links the testis seminiferous tubules via efferent ductules to the epididymis. Numerous imperfect, thin connective tissue septa extend from the tunica albuginea toward the mediastinum, creating several hundred incomplete pyramidal lobules containing the seminiferous tubules surrounded by loose connective tissue.

The Seminiferous Tubules

Spermatogenesis takes place within the seminiferous tubules; these are ~200 μm in diameter, ~600-m long, and occupy ~50% of the human testis volume. The numerous individual seminiferous tubules are highly convoluted linear segments. Each tubule ends in the mediastinum and empties via straight extensions, termed tubuli recti.[1]

The seminiferous epithelium is formed by contiguous Sertoli cells extending from the tubule base to its lumen. Fine and elaborate Sertoli cell cytoplasmic extensions envelope spermatogenic cells at all development phases. The most immature germ cells are mitotic spermatogonia that contact the tubule basement membrane. Spermatogonia undergo both self-renewing and multiplicative divisions, the latter with incomplete cytokinesis; clonal siblings are connected by cytoplasmic bridges through all subsequent differentiation steps. Spermatogonia differentiate and move away from the tubule base into the epithelium, forming the primary spermatocytes that commence meiosis. After the first meiotic division, these become secondary spermatocytes that undergo a second nuclear division to complete meiosis and become haploid round spermatids. Spermatids do not divide but undergo the complex metamorphosis called spermiogenesis. They are released into the tubule lumen as individual spermatozoa by the processes of spermiation.

Direct movement of substances between adjacent Sertoli cells is prevented by specialized tight cell junctions that also create basal and adluminal compartments in the seminiferous tubules. These junctions form the blood-testis barrier, which transiently opens as spermatogonia lose contact with the basement membrane to enter the adluminal compartment when they mature. Peritubular myoid cells completely cover the external tubule surfaces; together with the Sertoli cells, they synthesize the basement membrane that surrounds the tubules and upon which spermatogonia reside. These myoid cells contract irregularly to move sperm and Sertoli cell-secreted fluid in the tubule lumen through the mediastinum into the rete testis, which drains into the epididymis. Three successive rete testis zones are (1) straight tubules of the septal rete, (2) anastomosing mediastinal rete channels, and (3) wider extratesticular rete spaces that connect to the 6 to 12 fine efferent ductules that lead into the epididymal head.

The Intertubular Tissue

Adult human seminiferous tubules are surrounded by multiple layers of myofibroblasts, macrophages, and fibroblasts interspersed with collagen and extracellular matrix[1]; The peristaltic contraction and relaxation of peritubular myoid cells is governed by local vasoconstrictive and vasorelaxant products such as endothelin, vasopressin, oxytocin, and nitric oxide. Peritubular cells cosynthesize the tubule basement membrane with Sertoli cells, secreting proteins that modify the intratubular environment and directly affect spermatogenesis.[2]

The loose connective tissue of the interstitial area contains a species-specific cell composition. For example, the frequency and proportion of the androgen-secreting Leydig cells varies between species, typically occupying 10% to 20% of this area. Blood vessels occur randomly throughout the interstitium, and capillaries are nonfenestrated. Although lymphatic vessels and large lymphatic sinusoids exist in testes of several rodent species, in humans these reside only in the main connective tissue septa extending inward from the tunica albuginea. Because blood vessels do not enter the seminiferous tubules, germ cells embedded within the epithelium are dependent on Sertoli cells for direct support.[1]

Interstitial areas also contain fibroblasts and macrophages, both opposed to the tubules (peritubular macrophages) and also within larger spaces, usually adjacent to Leydig cells and blood vessels (interstitial macrophages). Lymphocytes and mast cells are less frequent and are typically more abundant near the tunica vasculosa. The macrophage and lymphocyte frequency and function relate to their roles in maintaining the testis as an immune-privileged organ.[3]

Postnatal Growth

Testis volume increases significantly from 1.1 mL (median paired testes volume) at 0 to 1 year of age, to 3 mL at 5 to 10 years, to 23 mL at 14 to 18 years, and reaching 40 mL in an adult group of 18 to 50 years.[4] Seminiferous tubule length per testis increases in prepubertal and postpubertal age ranges in a biphasic pattern, reaching a median length of 600 m in adults, with a wide variation that exceeds 1000 m in some individuals. The tubule diameter does not appreciably increase from birth to 14 years (50 to 60 μm) but expands to ~130 μm at 14 to 18 years and is ~200 μm in adults. Growth in seminiferous tubule length, but not diameter, in the prepubertal period reflects the significant increase in total germ cell numbers, from 13 to 83 million from 1 to 10 years of age, as spermatogonia populate the growing seminiferous tubules by migrating along the tubule length and forming more expansive colonies. This is followed by an exponential numerical expansion in the germ cell population.[5–7]

CELLULAR DEVELOPMENT AND ORGANIZATION FOR FUNCTIONAL CONTROL

The Spermatogenic Epithelium

Adult male fertility requires sustained proliferation and maturation of self-renewing germline stem cells and spermatogonial stem cells (SSCs) capable of differentiating into fully formed spermatozoa. The duration of one adult spermatogenic cycle (from stem cell to sperm) varies between species, taking 35 days in mice and hamsters, 50 days in rats, 45 to 65 days in several nonhuman primates, and 70 days in humans. The first release of spermatozoa (spermarche) in human puberty is at ~13.5 years of age, or earlier. Adult men produce ~1000 sperm per second, with 1.5 to 2.5 × 10^8 sperm typically present in an ejaculate.

KEY POINTS

- Spermatogenesis occurs with all germ cells embedded in the seminiferous epithelium created by Sertoli cells.

Germ cells occupy up to 75% of the seminiferous epithelium, arranged in multiple layers with least mature germ cells adjacent to the base and more mature cells closer to the tubule lumen. Maturing cells move apically through the epithelium connected as clonal siblings by cytoplasmic bridges until the most mature testicular germ cells, spermatozoa, are released into the tubule lumen (Figs. 111.1 and 111.2). This basal to luminal arrangement of maturing germ cells within the seminiferous epithelium is fundamentally similar in all mammals, although the associations between differentiating germ cell types along its length, termed *stages*, differ between species. Broadly speaking, regulation of adult spermatogenesis involves (1) endocrine stimulation of the testis by gonadotropins and steroid hormones (notably androgens), (2) local paracrine and autocrine factors, (3) tightly regulated gene expression within each germ and somatic cell type, (4) dynamic interactions between individual Sertoli cells and the changing variety of developing germ cells they support, and (5) Sertoli cell maintenance of the unique seminiferous tubule environment that restricts entry of cells, macromolecules, and immunogenic agents originating from the intertubular tissues and vasculature. Disturbance or impairment of any one of these parameters may disrupt or suppress spermatogenesis, contribute to defects in spermatozoa, and/or decrease sperm count.

Sertoli Cells

In the adult testis, Sertoli cells are the only nondividing cell type within the seminiferous tubules, and their function is central to sperm production. A single Sertoli cell partially or completely surrounds and supports several dozen or more germ cells.[8] The columnar Sertoli cell base rests on the basement membrane and overarches spermatogonia, extending apically to the lumen with elaborate lateral branches that interdigitate and surround all other germ cell types (Fig. 111.1). Each Sertoli cell contacts the base of five or six other Sertoli cells, simultaneously supporting up to 50 germ cells at different developmental stages.[9] Sertoli cells change shape as germ cells multiply, move, and undergo their own shape and size transformations. Sertoli cells divide only during fetal and postnatal life, stimulated by follicle-stimulating hormone (FSH), activin, and other factors; their number in an adult testis determines its spermatogenic capacity.[10] Proliferation during puberty occurs in species with a lengthy interval between birth and puberty, ending by approximately 15 years of age in humans.[11]

Blood-Testis Barrier. The blood-testis barrier exists in all animals. It is a structural and physiological compartment created by inter–Sertoli cell tight junctions on the luminal side of spermatogonia and their immediate descendants, the preleptotene/leptotene primary spermatocytes (Figs. 111.1 and 111.2). It includes several junction types, including tight junctions between adjacent Sertoli cell membranes that overlay a region of smooth endoplasmic reticulum with microfilaments projecting into the cytoplasm. The blood-testis barrier is stabilized by many intracellular components just beneath the cell membrane, including actin filaments and actin-binding proteins such as espin.[9,12] The junction positioning at the seminiferous epithelium base subdivides the anatomically and functionally distinct basal and adluminal compartments; spermatogonia develop within the basal compartment, while the meiotic and postmeiotic germ cells mature within the restricted adluminal compartment. Barrier integrity is essential for (1) excluding meiotic and postmeiotic germ cells from immunologic challenge arising from outside the seminiferous epithelium, (2) maintaining the stem and spermatogonial cell niche, (3) preventing intercellular fluid and macromolecular flow into the adluminal compartment and thus allowing its milieu to be determined by the Sertoli cells, and (4) enabling selective exposure of basal germ cells to regulatory factors originating from interstitial cells, such as Leydig and peritubular cells and macrophages.

The blood-testis barrier forms first in the peripubertal period when spermatogenesis is initiated by gonadotropic stimulation and coincides with the appearance of zygotene-pachytene primary spermatocytes.[13] Developing spermatocytes move through this tight junction barrier, which opens during their transit into the adluminal compartment and then reforms basally.[14]

In humans, junctional specializations between Sertoli cells are observed after 8 years of age and assembled in early puberty (11 to 13 years)[15] accompanied by increasing spermatocyte numbers. Coordinated signals by hormones (testosterone and FSH), retinoic acid (RA), and other factors differentially regulate tight junction dynamics to maintain blood-testis barrier integrity, yet allow germ cells to cross the barrier at the appropriate stage for meiotic maturation (see "Adult Sertoli Cell Function").

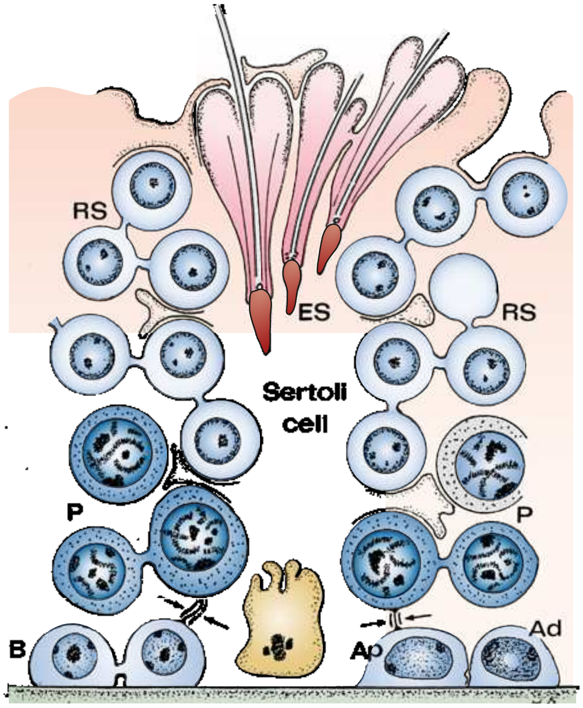


Figure 111.1 Diagram of the human seminiferous epithelium showing the general relationship between Sertoli cells and germ cells. Type A pale *(Ap)*, A dark *(Ad)*, and B *(B)* spermatogonia are shown in the basal compartment, together with primary spermatocytes *(P)*, early round spermatids *(RS)*, and elongating spermatids *(ES)* in the adluminal compartment. Inter–Sertoli cell tight junctions are indicated by *arrows*.

KEY POINTS

- The haploid spermatids undergo complex differentiation, termed spermiogenesis, involving sperm head nuclear compaction and acrosome development, sperm tail formation, cell organelle reorganization, and shedding of excess cytoplasm.

Plasticity and Links With Germ Cells. Concordant with elongated spermatid release at the end of spermiogenesis, Sertoli cells form specialized junctions with a new cohort of elongating spermatids and restructure inter–Sertoli cell junctions to permit spermatocyte translocation into the adluminal compartment (Fig. 111.3). The elaborate and dynamic cytoskeleton of adult Sertoli cells maintains their columnar shape, determines the distribution of its organelles, and mediates adhesion with germ cells. The cytoplasm around its nucleus contains microtubules and abundant motor proteins, dynein and kinesin, that are especially abundant in crypts surrounding elongating spermatid heads.[16] These cytoskeletal elements, together with intermediate filaments, are positioned to influence the dynamic shape and position of maturing germ cells as proceed through the spermatogenic cycle. Mouse models with defects in microtubule dynamics often exhibit phenotypes of premature germ cell release and sterility.

The actin-containing ectoplasmic specialization (ES) is another Sertoli cell cytoskeleton component that connects Sertoli and germ cells and facilitates spermatid orientation within the seminiferous epithelium. This structure in the Sertoli cells is positioned opposite the nucleus of newly elongating spermatids and facilitates their translocation in the epithelium as they elongate and develop an extensive flagellum.[17] The abundant Sertoli cell microtubular system moves spermatids by coupling its associated motor proteins to the ES. The microtubule-associated protein kinesin directs the

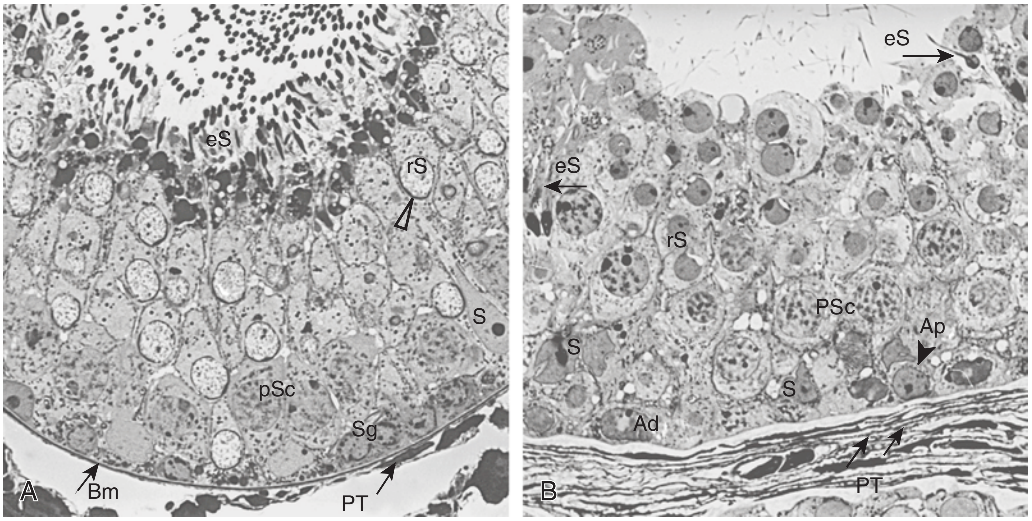


Figure 111.2 Histological sections showing seminiferous epithelium cross sections of (A) rat and (B) human testes. *Ad,* spermatogonia types A dark; *Ap,* A pale; *Bm,* basement membrane; *eS,* step 19 elongating spermatids; *P,* pachytene spermatocytes; *pS and eS,* round and elongated spermatids; *pSc,* primary spermatocytes; *PT,* peritubular tissue; *rS,* step 8 round spermatids; *S,* Sertoli cell nucleus; *Sg,* spermatogonia.

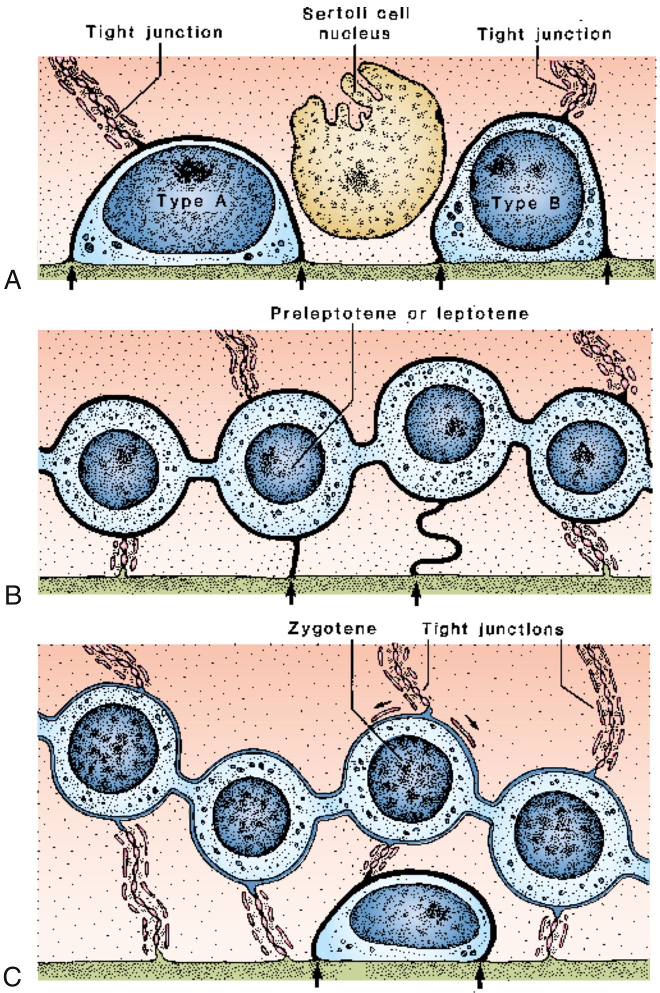


Figure 111.3 Schema showing tight occluding junctions between adjacent Sertoli cells that constitute the blood-testis barrier. (A) Tight occluding junctions are formed above the Type A and Type B spermatogonia positioned on the basement membrane, and (B) above and below meiotic spermatocytes (labeled), as developing germ cells move (C) from the basal to the adluminal epithelial compartment. Positions where tracers, representing macromolecules required for germline support, can penetrate are indicated with *arrows.*

initial basal movement of elongating spermatids into deep Sertoli cell crypts, then the microtubule motor protein dynein reverses spermatid translocation, toward the lumen.[17] This transient penetration of elongating spermatids deep within the epithelium likely facilitates spermatid-Sertoli cell communication during spermiogenesis.

Phagocytosis, Endocytosis, and Secretion. Many germ cells are eliminated via apoptosis during normal spermatogenesis, and the spermatid residual cytoplasm remains within the epithelium as a residual body after spermiation. Sertoli cells process these apoptotic germ cells and residual bodies via phagocytic and endocytic pathways. Residual bodies move basally, while excess plasma membrane removal and luminal fluid absorption occur at the Sertoli cell apical cytoplasm.[18] Sertoli cells must also transport or generate metabolites required for germ cells in the adluminal compartment, as the blood-testis barrier prevents their direct delivery. Receptor-mediated endocytosis of macromolecules occurs at both apex and base of the epithelium, including to deliver iron from serum to germ cells.[19]

Spermatogonia and Spermatogonial Stem Cells

Spermatogonial proliferation and differentiation are regulated by hormones, paracrine/autocrine factors, and systemic nutrients.[20] The proliferative, sexually indifferent primordial germ cells formed in early embryogenesis adopt the male fate in response to somatic cell cues following *SRY* expression in Sertoli cells. Termed gonocytes, or prospermatogonia, in the fetal testis, they enter a period of mitotic quiescence then transform immediately after birth into mitotic spermatogonia in rodents or asynchronously in humans from fetal life to about 2 years. Fibroblast growth factor (FGF) signaling mediated by FGF8 helps initiate this transition by germ cells in late fetal life in mice[21]; transcripts encoding FGF receptors are present in human spermatogonia but their functionality is undetermined.[22] All spermatogonia undergo mitosis. Continuous production and release of spermatozoa requires a stable population of SSCs that either self-renew to generate more stem cells or produce progeny, which commence spermatogenesis and transform into sperm.

KEY POINTS

- The entire spermatogenic process involves highly regulated developmental steps, mediated by local and endocrine somatic cell products and requiring ongoing spermatogonial stem cell replication for continuous sperm production over the life span.

All spermatogonia are partly surrounded by Sertoli cells and also directly contact the basal lamina, so secretory products of peritubular tissues, Leydig cells, immune cells, and circulating factors including hormones can influence their maturation (Fig. 111.4). In adult mammalian testes, spermatogonia span a range of differentiation states that include the least differentiated SSCs, progenitor spermatogonia primed to differentiate but that can still form a stem cell pool, and the differentiated spermatogonia that cannot function as stem cells. Spermatogonia in mammals were originally classified into types A and B according to their nuclear chromatin patterns in histologic preparations.[23] Type A spermatogonia were classified into undifferentiated and differentiated subtypes, based on morphological criteria and their potential to function as stem cells,[24] and type B spermatogonia are those that divide to produce preleptotene spermatocytes. These categories and species-specific differences are being revealed by single cell transcriptome data (discussed later).

Transplantation analyses functionally identify SSCs as those germ cells that can reconstitute full spermatogenesis when transferred into a recipient testis. This can assess how specific signaling pathways contribute to SSC maintenance and differentiation, with one goal being the preservation or restoration of fertility for individuals undergoing stem cell–damaging chemotherapy.[25] Both organ and cell cryopreservation and culture are used for SSC preservation and recovery, and genetic modifications prior to transplantation can correct infertility-causing mutations in animals.[25] Transplantation of heterologous or xenologous germ cells has also revealed fundamental or conserved features of testis development. For example, rat or hamster germ cells undergo spermatogenic development in recipient mouse testes[26,27] with the rate of germ cell development intrinsic to the transplanted donor germ cells, rather than determined by the host niche. Mouse and other host testes will maintain human SSCs for months, but they do not differentiate[25]; however, allogeneic transplantation of Rhesus macaque spermatogonia into either juvenile or adult recipient testes can undergo complete spermatogenesis.[28] Thus autologous transplantation of preserved self-germline stem cells may enable men requiring chemotherapy to produce healthy sperm.

The prevailing model of the rodent spermatogonial hierarchy is based on markers for undifferentiated cell subsets and development of

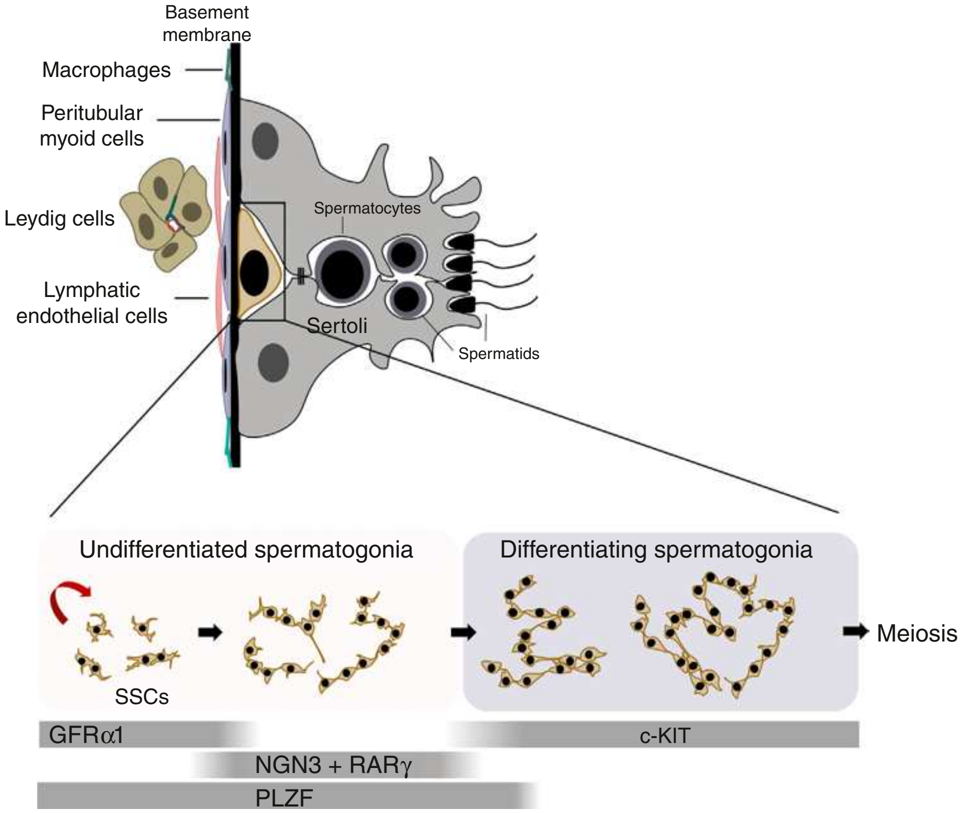


Figure 111.4 **Illustration of rodent spermatogonial hierarchy and components of the stem cell niche.** Spermatogonia are attached to the seminiferous tubule basement membrane and surrounded by multiple supporting somatic cell types *(top panel)*. A segment of the seminiferous tubule is shown. The spermatogonial population is divided into undifferentiated and differentiating fractions *(bottom panel)*. The undifferentiated population contains spermatogonial stem cells (SSCs) and differentiation-destined transit-amplifying cells. Differentiating spermatogonia undergo a series of coordinated mitotic divisions before generating meiotic spermatocytes that produce haploid spermatids. Molecular markers of the distinct spermatogonial fractions are shown.

functional SSC assays (lineage-tracing and transplantation).[20] It proposes that SSCs of the homeostatic testis are spermatogonia expressing the coreceptor for growth factor GDNF (GFRα1), which includes most single (A_s) and a substantial fraction of paired (A_{pr}) cells (Fig. 111.4). Transcriptional regulators ID4, EOMES, and PDX1 are linked to SSCs and are expressed within the GFRα1+ population.[29] GFRα1+ cells continually interconvert between syncytial states (A_{pr} plus short chains of aligned cells [A_{al}] and A_s through chain fragmentation).[30]

In contrast, undifferentiated cells positive for *Neurog3* (*Ngn3*), primarily longer chains (A_{al}), are destined to differentiate under the influence of RA acting via RA receptor γ (RARγ).[31] Spermatogonial differentiation is marked by KIT receptor induction and epigenetic changes, accompanied by loss of stem cell potential.[32] Importantly, Ngn3+ A_{al} can fragment to A_s and A_{pr} and contribute to the GFRα1+ stem cell pool, particularly under regenerative conditions, for example, following germline depletion or transplantation.[24] The role of transit-amplifying populations in germline regeneration was confirmed by studies of cells expressing *Piwil4* (*Miwi2*), a subset of the Ngn3+ population. Further, differentiation-primed progenitors can revert to an SSC state under the influence of niche factors such as GDNF, as demonstrated through culture and transplant-based studies.[29] Markers such as GFRα1, RARγ, KIT, and others are routinely used to functionally categorize spermatogonia (Fig. 111.4).[29]

Rodents have a distinct first round of spermatogenesis in which the first sperm arise soon after birth directly from gonocytes that transition directly into differentiating spermatogonia; all subsequent rounds of spermatogenesis derive from gonocytes that transition first into an undifferentiated spermatogonial state. Recent single cell-RNA sequencing (scRNAseq) analyses indicate that this is different in humans.[22,33] Human spermatogonia are maintained in one of two undifferentiated states in low numbers (States 0 and 1; 3% to 4% of all testis cells) from infancy through adulthood. This “reserve” population is amplified by multiplication at the onset of puberty, from 11 years; differentiating spermatogonia emerge (stages 2 to 4) with transcription of meiotic genes evident from 13 years.[22] Thus the mechanisms that control the initiation of spermatogenesis in juveniles are likely to be both shared and distinct between rodents and humans, and they may also be distinct from those active in adults.

The Spermatogonial Stem Cell Niche

The specialized stem cell niche microenvironment conveys extrinsic signals to modulate the intrinsic stem cell regulators that determine fate choice: between self-renewal, proliferation, differentiation, or cell death. Adult fertility requires an uninterrupted germ cell supply along the complete length of each tubule, and hence stem cells must be appropriately distributed. In the postnatal prepubertal rodent testis, spermatogenesis is initiated nonuniformly throughout the testis, with RA a fundamental driver for this in both neonatal and adult rodents.[34,35] Over time, complete cycles of the seminiferous epithelium are established with fairly precise spatial distribution along the tubules. Histological and real-time video evidence from rodent testes indicates that SSCs are nonrandomly positioned close to the vascular-rich intertubular regions rather than in regions where neighboring tubules are closely opposed; this distribution was reestablished if tagged tubule segments were freed of vessels and interstitial tissues subsequently transplanted into a recipient testis (Fig. 111.4). SSC density and localization are regulated by competition for limiting amounts of FGFs produced by lymphatic endothelial cells and other interstitial cells surrounding the tubules.[36] The Sertoli cell–produced chemokine, CXCL12, provides a critical signal via the CXCR4 receptor on SSCs, maintaining them in an undifferentiated state and controlling their homing to the niche.[37] CXCL12 may also influence the behavior of seminoma cells that arise in young men from fetal germ cells that fail to differentiate normally into spermatogonia.[37,38]

Glial cell–derived neurotrophic factor (GDNF) synthesis by Sertoli cells is essential for the maintenance and proliferation of mouse, rat, hamster, bull, and human SSCs. Its synthesis is stimulated by pituitary-derived hormone FSH and locally produced factors such RA, basic fibroblast growth factor (bFGF), tumor necrosis factor-α (TNFα), and interleukin 1b (IL1β).[39,40] Other cytokines including FGFs and colony-stimulating factor-1 (CSF-1) complement the role of GDNF in SSC renewal and proliferation[20] but how circulating factors affect their synthesis is unknown. Balanced signaling by WNT ligands and inhibitors (via both canonical β-catenin–mediated and noncanonical pathways) also contributes to SSC maintenance and differentiation.[20]

Morphological Features of Spermatogonia in Primate and Human

Type A spermatogonia are precursors of all further differentiated germ cells. Primate and human testes exhibit dark and pale type A spermatogonia. The A dark (Ad) spermatogonia stain deeply with hematoxylin and have homogeneous, fine chromatin with an obvious central pale-stained vacuole-type area. Pale type A (Ap) spermatogonia have uniformly pale-staining nuclei and one or two nucleoli near the nuclear membrane. Type B spermatogonia nuclei feature coarse, deeply stained heterochromatin clumps next to the nuclear membrane, with a central nucleolus.

Ad and Ap spermatogonia are in pairs and groups along the seminiferous tubule (see Fig. 111.2). Type Ad spermatogonia rarely divide in the human testis and are considered as reserve stem cells that only divide and transform into Ap cells after irradiation or cytotoxic treatment.[41] Type Ap spermatogonia undergo both self-renewing mitotic divisions and mitotic amplification as they transition into type B spermatogonia.[42] Type B spermatogonia divide by mitosis then become primary and secondary spermatocytes while undergoing the two nuclear cell divisions of meiosis to yield the round haploid spermatids that then transform into spermatozoa. In humans, the ratio of Ad spermatogonia:Ap spermatogonia:type B spermatogonia:spermatocytes is 1:1:2:4. The relationship of these morphologically defined populations to those identified by scRNAseq is yet to be revealed.

Spermatogonial Differentiation and Its Control

Application of scRNAseq to testes of different ages and in different species has identified the progressive expression of transcripts through their maturation, including those that encode signaling receptors and thereby reveal regulatory components that act successively. Common to both mouse and human spermatogonia is the delineation of sequential maturation states, transitioning generally from a more metabolically quiescent to active, replicative state, and acquiring machinery for meiosis.[43,44] The first two spermatogonial states in adult human testes (referred to as States 0 and 1) express stem cell–associated factors including *ID4* and *PLZF*, suggesting that they represent SSCs, while proliferation and differentiation-associated markers are upregulated in subsequent states (e.g., *KI67* and *KIT*, respectively).[33] State 0 cells most closely resemble germ cells of the immature infant testis, indicating they are more primitive than State 1 cells. Surprisingly, expression of *GFRA1*, an SSC marker commonly used in rodents, is predominately observed in State 1 spermatogonia, while State 0 cells, a potential quiescent reserve SSC population, are uniquely marked by transcripts such as *PIWIL4* plus *TSPAN33* and lacking in rodents.[33] However, in common with rodents, RNA-velocity analyses indicate that human spermatogonia can dynamically interconvert between functional states, and their transplantation potential is not restricted to the most primitive state.[33,45] This cellular plasticity is associated with limited changes in the epigenetic landscape between spermatogonial states.[33] Notably, markers of State 0 and 1 cells are expressed by both Ad and Ap spermatogonia, indicating the disconnect between transcriptional and histological cell features historically used to assess human spermatogonial functions[22,45]

Multiple signaling molecules produced by testis somatic cells that are known, and new ones predicted from scRNAseq analysis, support germ cells transitioning through spermatogenesis.[45] RA signaling is essential for spermatogonial differentiation in postnatal and adult testes. Induced RA deficiency in rats and mice through either dietary or pharmacological means reversibly arrests type A spermatogonial proliferation and differentiation.[34] Produced by Sertoli cells, RA represses *GDNF* transcription[40] and elevates transcripts associated with spermatogonial differentiation in both Sertoli (*Kitl* and *Bmp4*) and germ cells (*Stra8* and *Kit*). Activation of the KIT tyrosine kinase receptor via KITL (termed KIT ligand or stem cell factor [SCF]) from Sertoli cells is essential for survival and proliferation of differentiating spermatogonia.[46,47] This pathway is also central to human fertility; abnormal pathway component expression is documented in testes with defective spermatogenesis, and polymorphisms and mutations in *KIT* and *KITL* are frequently associated with infertility and testicular germ cell tumors.[48] The transforming growth factor-β (TGFβ) superfamily ligands BMP4 and activin A also influence spermatogonial differentiation, including by modulating levels of KIT, the hallmark of differentiating spermatogonia.[49] Additional factors affecting spermatogonial differentiation and proliferation include nerve growth factor, epidermal growth factor, and platelet-derived growth factor.

Also key for postnatal spermatogenesis is the doublesex-related transcription factor (DMRT1) involved with both Sertoli and germ cell development; *Dmrt1* deletion in mouse embryos resulted in azoospermic adults. Cell-selective deletion of *Dmrt1* in Sertoli cells prevented their postnatal differentiation and led to the loss of the male phenotype; deletion in germ cells impaired the initial establishment and subsequent maintenance of spermatogonia,[50,51] premature entry into meiosis, and eventual stem cell depletion. This outcome was dependent on vitamin A actions, including through direct interaction of Dmrt1 with genes encoding proteins required at the onset of meiosis such as Stra8 and the mitotic inhibitor p21Cip1.52 Mutations and single-nucleotide polymorphisms (SNPs) in *DMRT1* are implicated in human infertility and increased testicular germ cell tumor risk, with the latter implicating a germ cell–intrinsic defect as critical in the establishment of this condition in some men. As mentioned in "Genetic Defects", mutations in sex chromosomes may also affect spermatogonial proliferation and/or maturation.

Meiosis

Type B spermatogonia enter prophase of the first meiotic division to form primary spermatocytes. Sibling spermatocytes remain connected via the cytoplasmic bridges through which mRNAs and other macromolecules can transit.[53] Although these bridges first form as SSCs commence differentiation, their loss does not visibly impact progression until meiosis.[54] Meiosis involves a long prophase whereby DNA is replicated and there is homologous recombination between paired chromosomes to facilitate genetic diversity of the games. Preleptotene spermatocytes commence DNA synthesis, replicating each chromosome to form twin copies called *sister chromatids*, marking the commencement of prophase. As a consequence, the preleptotene primary spermatocyte chromosome number remains diploid, but the DNA content is 4C, twice that found in spermatogonia.

Prophase is subdivided into five phases based on gross chromosomal activities and appearance: (1) leptotene, thread-like chromosomes, (2) zygotene, homologous chromosome pairing, (3) pachytene, paired chromosomes thicken and undertake crossing over, (4) diplotene, paired chromosomes decondense and partially separate, and (5) diakinesis, chromosomes recondense. During prophase, homologous chromosomes pair and the *synaptonemal complex* facilitate the close apposition of each homologue pair, and the chromosomes form a synapse that permits crossover and exchange between them at *chiasmata*. Genetic exchange between the paired chromosomes requires error-free double-strand DNA breakage and repair; key to this is the strong upregulation in mid-phase primary spermatocytes of highly conserved topoisomerases.[55]

Metaphase follows prophase, as the highly condensed chromosomes align on a metaphase spindle. During anaphase, the chromosome pairs move to opposite poles of the cell. During telophase, complete nuclear separation forms daughter cell nuclei, producing *secondary spermatocytes*. The second meiotic division proceeds rapidly (<24 hours) without DNA replication, yielding two haploid spermatids, each with 1C DNA content. Thus four spermatids form from each primary spermatocyte. Errors during meiosis have severe consequences for fertility, ranging from azoospermia to oligospermia, to aneuploidy and elevated fetal loss.[56]

KEY POINTS

- Spermatogenesis features incomplete cytokinesis between clonal siblings, with mitosis in spermatogonia followed by meiosis in primary and secondary spermatocytes to achieve the haploid state.

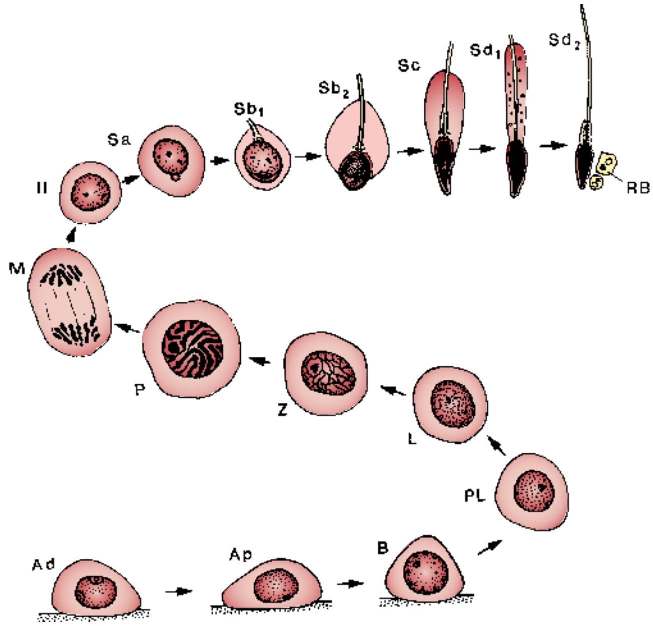


Figure 111.5 Sequence of germ cell types in human spermatogenesis commencing with a spermatogonium, progressing through an orderly sequence of cell proliferation and maturation, and terminating in a mature spermatozoon. Spermatogonia (shown on epithelial base): Type A dark *(Ad)*, A pale *(Ap)*, and B *(B)*; primary spermatocytes: preleptotene *(PL)*, leptotene *(L)*, zygotene *(Z)*, pachytene *(P)*, and meiotic division *(M)*; secondary spermatocyte *(II)*; spermatids *(Sa, Sb1, Sb2, Sc, Sd1, Sd2)*; residual body *(RB)*.

Spermiogenesis: Emergence of Spermatozoa

Spermiogenesis is the elaborate transformation of round haploid spermatids into elongated spermatozoa; it takes 24 days in humans (Fig. 111.5). Sibling spermatids remain connected by cytoplasmic bridges until their release into the tubule lumen at spermiation. The complex morphological features of spermatid maturation transform a spherical cell into the highly polarized spermatozoon, featuring a streamlined head containing a tightly compacted nucleus overlayed by the acrosomal vesicle, a tail for motility, and minimal cytoplasm (Fig. 111.6). The human spermatozoon head is spatulate in form and flattened, while the rodent sperm head is hooked or falciform. A thin membrane-bound cap, the acrosome, extends from the tip, or leading edge, of the head and covers approximately 60% of the sperm head. The nucleus is enveloped by a perinuclear theca, and the acrosome is attached to the apical pole via the acroplaxome.[57] The entire cell is covered by a plasma membrane. The complex assembly of the various components in each domain is a fascinating process that is heavily dependent on both actin- and tubulin-based cytoskeletal elements and their associated motor proteins (reviewed in[58]).

The Sperm Head. The sperm head contains the highly condensed haploid nucleus surrounded by a perinuclear theca enveloped at the apical end by the membrane-bound acrosomal vesicle containing enzymes and surrounded by the plasma membrane. The genomic DNA in sperm becomes compacted to occupy ~5% of the typical somatic cell volume in a highly regulated manner. Most histones are removed in developing spermatids and replaced first with transition proteins (TP1, TP2) and finally the protamines (1 and 2), retained in mature sperm. Acetylation of histones mediates their disassociation from DNA and subsequent ubiquitination, allowing histone replacement with the transition proteins that facilitate DNA compaction and the repair of DNA breaks arising during chromatin remodeling. Mouse data demonstrate the sequential incorporation of transition proteins, and protamines are absolutely required for normal fertility,[59] and altered protamine 1 and 2 ratios in sperm are associated with human infertility.[60] Protamines organize the spermatid genome into distinctive doughnut-shaped chromatin components. This exchange is incomplete, and approximately 10% to 15% of genes remain associated with histones in humans. Sites retaining nucleosomes are preferentially linked to early embryonic development and biased toward genes encoding transcription factors,[61] indicating that histone retention contributes to reproductive fitness.[62] Defects in histone retention or positioning may contribute to human infertility[61] and changes to ribosomal and genomic DNA methylation profiles in sperm correlate directly with aging.[63,64]

The onset of spermiogenesis requires a massive wave of transcriptional activity in haploid cells.[65] However, because nuclear compaction in elongating spermatids renders much of the genome inaccessible, RNA translation is regulated to ensure that proteins are synthesized at the appropriate time in late spermiogenesis. Translational regulation includes transcript storage in messenger ribonucleoprotein particles (mRNPs)[66] relating to particular sequences present in untranslated regions (UTRs), functions of small RNAs and a newly discovered class of abundance circular RNAs.[67]

Concordant with histone to protamine exchange, the sperm head species-specific shape is sculpted by the microtubule-based manchette. It forms as a girdle-like perinuclear ring in early elongating spermatids, with multiple microtubules extending toward the growing tail[58] and moving distally, facilitated by the kinesin-related molecular motor, KIFC1.[57] Nuclear shaping is aided by progressive manchette ring constriction involving cytoplasmic dynein.[68] Failure of manchette assembly or function results in teratozoospermia. Manchette microtubules also deliver essential components to the developing tail.[58]

While the sperm head is being shaped, the acrosome forms at the germ cell apical pole in four phases: Golgi, cap, acrosome, and maturation. Its assembly uses cytoskeletal machinery to incorporate membranes both

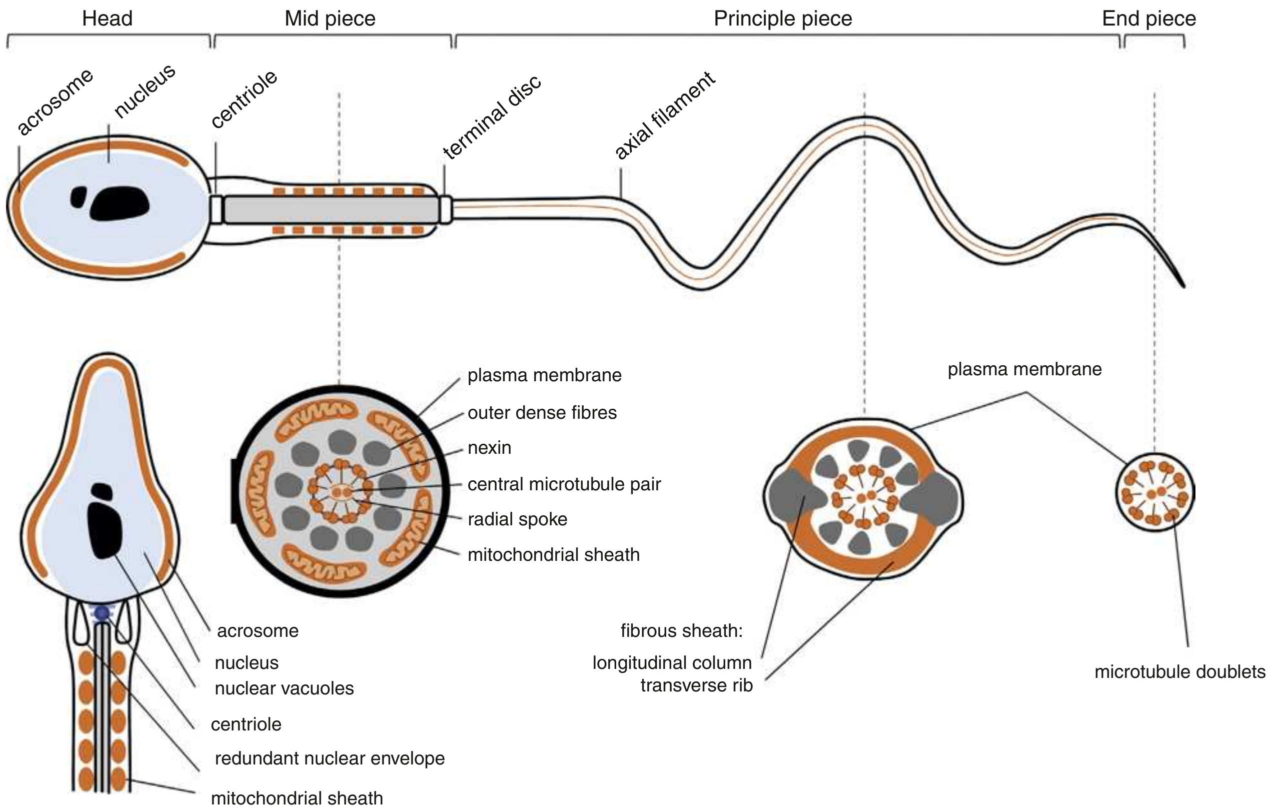


Figure 111.6 Diagram of a typical mammalian spermatozoan. Cross sections show distinct regions in the head and along the tail length, with 9 + 2 microtubule doublets forming the central axoneme.

from the Golgi and the plasma membrane.[58] The acrosome membranes and contents are regionally distinct, with the acrosome containing soluble proteins, including hyaluronidase and dipeptidyl peptidase, and an insoluble matrix containing proacrosin and proacrosin-binding protein. During fertilization, exocytosis of the acrosomal granule involves selective fusion of the apical, outer acrosomal membrane with the overlying plasma membrane, releasing soluble hydrolases required for sperm penetration through the oocyte vestments. Men in whom acrosome formation fails have globular-shaped sperm, termed globozoospermia, or sperm with misshaped heads (acrosomal hypoplasia), and they are infertile. In humans, mutations affecting several elements disrupt the connection between sperm head and tail and thereby cause infertility.

Another further key step in sperm head development is the assembly of sperm-oocyte receptor components, including by molecular chaperones.[69] Proteins required for or implicated in sperm fusion with the oocyte plasma membrane are mainly localized to the equatorial segment in acrosome-reacted sperm.[70,71] Post-testicular sperm modifications require dynamin proteins that mediate interaction with epididymosomes, vesicles secreted by cells lining the epididymis that preferentially localize to the post-acrosomal domain during epididymal transit.[72] This results in the gain of proteins and small nonprotein-coding RNAs that contribute significantly to fertility and offspring health.[73–75] Thus the appropriate formation and maintenance of complex sperm head domains and their dynamic modifications during post-testicular maturation are key for male fertility.

The chromatoid body, nuage, and intermitochondrial cement are important components of the maturing spermatid cytoplasm, where small RNAs (miRNAs and piRNA) and the machinery that produces them are transcribed and stored.[76] Circular RNAs (circRNAs) become predominant in elongating spermatids as templates for translation.[67]

The Sperm Tail. The complex sperm tail contains four structurally distinct regions: the neck or connecting piece, and the middle, principal, and end pieces (Fig. 111.6). The neck and head articulate at the nuclear caudal pole, with nine cross-striated columns continuous with the nine outer dense fibers (ODFs) extending the entire tail length. Within the neck core, the distal centriole remnant connecting the tail and head is the origin of the axoneme, the microtubule-based motility apparatus of sperm. The midpiece (~5-μm long in human sperm) comprises the axoneme encircled by nine ODFs, which is in turn surrounded by a mitochondrial sheath. More distally, the principal piece (~45 μm in human) contains modified ODFs that are partially replaced by the two fibrous sheath (FS) columns, connected by transverse ribs. The ODFs confer elasticity and have tensile strength integral to flagellar motility. Both ODF and FS taper along the sperm tail length, terminating at the end piece.

The sperm axoneme consists of nine outer microtubule doublets encircling two central microtubules, extending the entire tail length. Dynein arm projections from the outer microtubule doublets drive axoneme movement to facilitate sperm motility through coordinated ATP hydrolysis, based on anchoring between the axoneme and sperm head and between microtubule pairs 3 and 4.[58] The circular axoneme arrangement of nine microtubules, but lacking a central pair, is also present in motile cilia and flagella and in nonmotile nodal (in embryos) and primary cilia (on most nondividing somatic cells). Because of the fundamental importance of this conserved structure, male infertility and respiratory dysfunction are common features in ciliopathies arising from primary ciliary dyskinesia.[77]

Axoneme Growth and Composition. Axoneme growth commences in round spermatids from one of the pair of centrioles that initially lies close to the Golgi. The axoneme emerges from the distal, or mother, centriole, and the intact companion proximal (daughter) centriole is lodged within the neck cone at 90 degrees to the axoneme. The nine cross-striated columns of the neck are continuous with the nine ODFs (Fig. 111.6).

Axoneme development is concordant with sperm head elongation and condensation. Transcription of genes encoding sperm neck, ODFs, and FS components commences in early round spermatids and peaks during the acrosome phase.[78] Axoneme assembly requires conserved intraflagellar transport (IFT) processes for protein delivery and placement. The ODF forms from proteins in granular bodies in the periaxonemal cytoplasm in a proximal-to-distal direction along the full axoneme length. In contrast, FS proteins distributed in the caudal spermatid cytoplasm assemble in distal-to-proximal orientation along the axoneme, eventually overlapping with the ODF. In addition to its role in sperm head shaping, the manchette is implicated as a transport highway for proteins involved in sperm tail development, *intramanchette transport.*[58] Many of the more than 20 FS proteins are essential for fertility. Among these are specific A-kinase anchoring proteins (AKAPs), which position regulatory protein kinase A (PKA) subunits near appropriate substrates for tyrosine phosphorylation.[79]

The sperm midpiece terminates at the annulus, a septin-containing ringlike structure that restricts protein diffusion. Defects in annulus formation or positioning can contribute to infertility. The principal piece contains modified ODFs, and ODF-3 and ODF-8 are replaced by the longitudinal FS columns that are connected by transverse ribs. The ODF and FS taper along the length of the sperm tail and terminate at the end piece junction, so that the end piece solely consists of the axoneme surrounded by the plasma membrane.

The ODF and FS columns provide structural rigidity to sperm tail movement and protection against shearing forces. Species-specific oxidative phosphorylation by the mitochondria and ATP production in the FS via glycolysis are also required for axoneme function.

Release of Spermatozoa. At the end of spermiogenesis, spermatids undergo spermiation over several days[18] to separate from the Sertoli cell and become individual haploid spermatozoa. This begins at the start of the seminiferous cycle stage II in humans, with elongated spermatids aligned at the seminiferous tubule lumen. Key steps are (1) removal of the ES that anchors the sperm head to the Sertoli cell, (2) development and then dissolution of tubulobulbar complexes that both anchor germ cells and remove their cytoplasm, (3) residual body formation from excess germ cell organelles and cytoplasm, and (4) complete disengagement of spermatozoa from the seminiferous epithelium. See "Sertoli Cell–Mediated Regulation of Spermatogenesis" for an overview of sperm release regulation.

Germ Cell Associations and the Spermatogenic Cycle

The tight coordination of germ cell development within the seminiferous tubule is evident in both cross sections and whole-mount preparations. Tubule cross sections exhibit species-specific cell associations as consistent, co-located combinations of particular mitotic, meiotic, and haploid germ cell subtypes within the Sertoli cell epithelium. These repeat along the length of each tubule, in associations designated by Roman numerals as stages I to VI for humans. Over a period of days, a given point in the epithelium corresponding to an individual Sertoli cell harbors every stage in a repeating sequence. Individual stages may be several millimeters in length, reflecting simultaneous development of germ cell cohorts linked by cytoplasmic bridges.

The wave of the seminiferous epithelium cycle is defined as the time or distance along the tubule taken to progress through all germ cell stages. A complete sequence of 16 stages occupies a species-specific interval of days, referred to as one cycle of the seminiferous epithelium. Completion of spermatogenesis requires a spermatogonium to commence proliferation, pass through four cycles of the seminiferous epithelium, and transform into a spermatozoon. Xenotransplantation between rodent species shows this pace is endogenous to the germ cells (see "Spermatogonia and Spermatogonial Stem Cells"), while tissue grafts into mice demonstrate the host hypothalamic-pituitary axis can speed the initial development of monkey or pig germ cells.[80,81] In humans, one spermatogonium can yield a theoretical maximum of 32 spermatozoa,[82] while rodents can produce over 1000, due to undergoing 9 to 10 mitotic "amplification divisions" before meiosis commences.

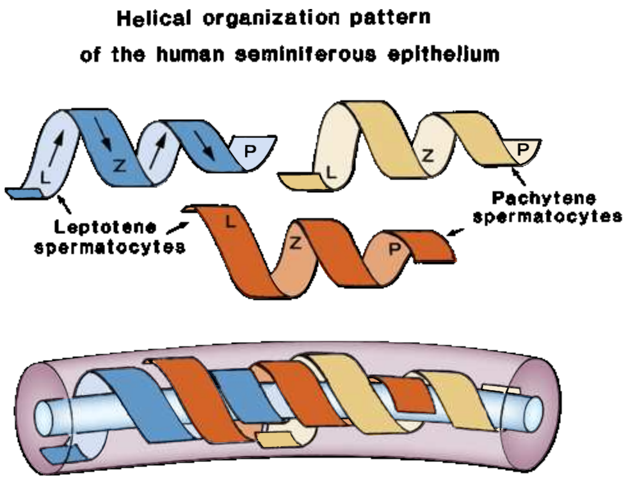

Helical organization pattern
of the human seminiferous epithelium
L
Z
P
L
Z
P
L
Z
P
Leptotene spermatocytes
Pachytene spermatocytes

Figure 111.7 Helical arrangement of germ cell development in human testis. Primary spermatocyte (leptone *[L]*, zygotene *[Z]*, pachytene *[P]*) positions drawn as strips that overlap and gyrate toward the tubule lumen. This pattern occupies only small seminiferous tubule segments, and elsewhere individual spermatogenic stages (I to VI) occur randomly along tubules.

While many species exhibit a single seminiferous cycle stage occupying each individual tubule cross section (Fig. 111.3), humans and some primates (baboon, crab-eating macaque, orangutan, chimpanzee) display several within the same sectioned tubules (Fig. 111.3). In some, but not all human tubules, the developing germ cell arrangement follows a series of overlapping helical strips (Fig. 111.7).[83] One relatively new set of criteria distinguishes 12 stages in the human spermatogenic cycle based on acrosomal morphology using immunohistochemical detection of proacrosin.[84] Findings from rodent models indicate that Sertoli cells determine the spatiotemporal patterning of the seminiferous epithelium cycle.

Spermatogenic Efficiency and Germ Cell Loss

Spermatogenic cell degeneration is a normal feature of testicular development and required for ongoing spermatogenesis; genetic ablation of cell death mediators in mice can cause infertility. Spermatogenic efficiency is defined as the theoretical versus actual yield of mature spermatozoa from a single SSC.[85] This species-specific number can be affected by external factors, such as seasonality, nutrition, infection, exposure to medication or environmental agents, or it may arise from genetically determined dysfunction. Up to 75% of the maximum predicted sperm yield is normally lost in some species. Apoptotic cell death in the testis, as in other tissues, balances cell proliferation and elimination for normal cell growth and population maintenance.[85] Germ cell apoptosis occurs in all phases of spermatogenesis in response to hormone deprivation, heat stress, radiation, antimetabolic agents, cytoskeletal disrupters, cytotoxic agents, and tumor-inducing drugs. This mediates the removal of mitotic germ cells that harbor unrepaired DNA, spermatocytes with chromosomal breaks or unequal chromosomal exchange, and germ cells too numerous or inadequately supported by Sertoli cells.

Distinct from necrosis, apoptotic cells fragment and form condensed bodies that include the pyknotic nucleus. Importantly, these bodies do not lyse and thus do not provoke an inflammatory response; instead, they are phagocytosed and their contents "recycled," typically by the Sertoli cells. Homodimers or heterodimers of Bcl-2 family proteins either promote (Bax, Bak, Bcl-x_s, Bad) or prevent (Bcl-2, Bcl-x_L, Bcl-w, Mcl, Al) apoptosis; their relative proportions determine whether apoptosis proceeds. Bcl-2 family expression is regulated during spermatogenesis, with spermatogonia requiring Bcl-w for survival.[86] Apoptosis is triggered by cysteine protease caspase activation, causing sequence-specific protein cleavage. Caspase 2 synthesis in germ cells establishes a sustainable balance of germ cells relative to niche size in the first spermatogenic wave, while caspase 9 and BAX mediate germ cell degradation following torsion-mediated ischemia in adults.[87] Sertoli cell engulfment of apoptotic mouse germ cells is required for ongoing spermatogenesis.[88]

Genetic Defects and Spermatogenesis

Knowledge of genetic contributions to idiopathic spermatogenic disruption in men is being expanded by analyses of large population datasets and new technologies[89]; this topic is addressed in Chapter 115. Mutations in over 450 genes can cause male infertility in mice, and many correspond to known genetic changes that impair human fertility.[90] Genetic causes include aneuploidies, with Klinefelter's syndrome (XXY) most frequent, chromosomal translocations, large chromosome deletions, particularly of the Y chromosome, copy number variations, and point mutations. In addition to human sex chromosome mutations that underpin 10% to 20% of male infertility cases,[91] extrinsic factors affecting chromosomal integrity, including exposure to oxidative stress, are implicated in up to 40% of male infertility cases.[92]

Epigenetics in the Male Germline

Chemical modifications to gDNA and histones determine which genes are open for transcription factor access and which are closed and silent. Transmission of this epigenetic information from parent to offspring occurs in gametes and can affect multiple generations.[93,94] Essential reprogramming of primordial germ cells involves global gDNA epigenetic erasure; this provides a "blank slate" for regulating transcription of genes needed in development that are later switched off in differentiated cells.

Methyl group addition to cysteine residues in gene promoter CpG islands (5-methylcytosine [5mC]) can silence their transcription, mediated by DNA methyl-transferase enzymes. During cell replication, DNMT1 transfers 5mC to maintain the parental cell transcription profile. However, *de novo* methyltransferases DNMT3A, DNMT3B, and DNMT3L act in fetal life to remethylate of germline precursor cell gDNA following their unique genome-wide demethylation. Histone modifications include acetylation, methylation (at lysine or arginine), or phosphorylation of histone tails.[93]

Epigenetic inheritance can depart from traditional Mendelian inheritance of one allele from each parent for every gene. Instead, a critical, parental-specific imprint can silence either the maternal or the paternal allele in certain genes. These imprints must be erased in the primordial germ cells for correct inheritance of parental epigenetic information; maternally imprinted genes must be demethylated in the male offspring's germline in which only paternal imprinted genes will be remethylated, and vice versa for the female germline.[94]

Retrotransposons are mobile DNA elements that have the potential to excise themselves from the genome, copy and paste themselves back into host cells, and thus can create new mutations. Occupying up to 40% of the human genome, their movement within a genome is usually suppressed by the same type of methylation marks that create closed chromatin structures. However, during the period of global demethylation in primordial germ cells, their gDNA is vulnerable to retrotransposon activity. Specific PIWI/piRNA pathway components are crucial for repression of retrotransposons during *de novo* methylation in fetal life. The P-element–induced wimpy testis (PIWI) proteins in the Argonaute family of RNA-binding proteins are highly expressed in germ cells.[95] PIWI proteins cleave and destroy mRNAs transcribed from transposons sequences; this cleavage creates the PIWI-interacting RNAs (piRNAs) that bind PIWI proteins then cooperatively silence transposons. In postnatal life, the PIWI/piRNA pathway regulates genes required for male fertility and many other cellular processes essential for life. Key components of this pathway essential for spermatogenesis are PIWIs, Tudor-domain containing (TDRD) proteins, DNMTs, and piRNAs.[96] The transcription factor A-MYB (MYBL1) drives transcription at loci encoding these "pachytene piRNAs," facilitating their biogenesis in early pachytene spermatocyte development,[97] though the trigger for this is unknown.

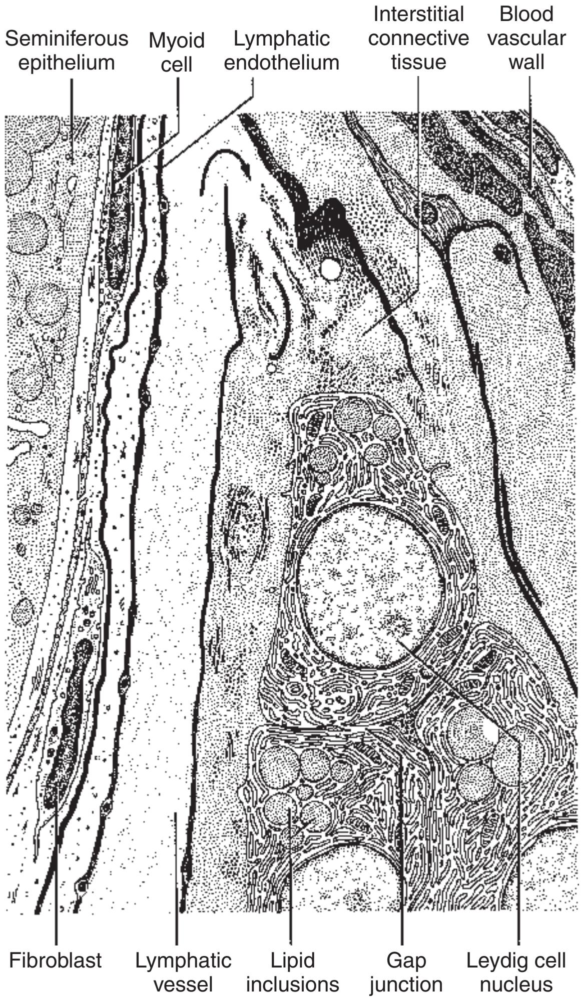


Figure 111.8 Cellular arrangement with the testis intertubular tissue, observed by electron microscopy to illustrate organelles and lipid droplets. (From de Kretser D, Kerr J. The cytology of the testis. In: Knobil E, Neill JD, eds. *The Physiology of Reproduction*. 2nd ed. New York: Raven Press; 1994:1177–1290.)

THE INTERTUBULAR COMPARTMENT

The space between seminiferous tubules is occupied by a loose connective tissue containing Leydig cells, macrophages, blood and lymphatic vessels, and the testicular innervation (Fig. 111.8).

KEY POINTS

- Spermatogonia are the only germ cells directly contacting the extratubular environment consisting of blood vessels, immune cells and fibroblasts, testosterone-producing Leydig cells, and the circulating gonadotropic hormones, follicle-stimulating hormone and luteinizing hormone.

Leydig Cells

Fetal, immature, and adult-type Leydig cells are epithelioid in shape. Their basophilic staining property arises from abundant smooth endoplasmic reticulum, where enzymes catalyze steroidogenic conversions, most notably conversion of acetate or cholesterol into testosterone. Leydig cells also contain Golgi membranes, microtubules, mitochondria with tubular cristae (unique to steroidogenic cells), peroxisomes, and scattered rough endoplasmic reticulum. Lipofuscin pigment granules are prevalent in human Leydig cells.

Other Interstitial Cells

The intertubular tissue contains abundant resident macrophages, characterized in several species using histochemical, immunologic, and ultrastructural techniques. Testicular resident macrophages have anti-inflammatory and tolerogenic properties, allowing tissue allografts or xenografts survive for long periods in recipient rodent testes[98]; consequently, the testis is designated as one of the rare immunologically privileged organs.

The mesenchymal or fibroblast-type cell populations in intertubular tissue are a source of new Leydig cells. These Leydig precursor or stem cells reside in peritubular and perivascular locations; they can produce testosterone following stimulation by luteinizing hormone (LH) and ultimately differentiate into Leydig cells.[99]

THE IMMUNOLOGIC ENVIRONMENT OF THE TESTIS

Germ cells excluding spermatogonia and very early spermatocytes develop within the adluminal testis compartment, isolated from the intertubular environment by the tight inter–Sertoli cell blood-testis barrier junctions. This sequesters more mature germ cells from the systemic immune system. Immune cells are normally confined to the testis intertubular tissue and include resident macrophages, antigen-presenting dendritic cells, circulating lymphocytes and significant numbers of CD8+ cytotoxic T cells and natural killer (NK) cells, and immunoregulatory T-cell subsets.[100] Interactions between resident and circulating immune cells are considered to regulate testicular immune responses and facilitate immune privilege.

KEY POINTS

- Spermatocytes and spermatids develop within the Sertoli cell intratubular epithelium that provides their structural and nutritional support and forms the blood-testis barrier that creates an immunologically privileged environment essential for fertility.

In early observations of the rat, functional interactions between Leydig cells and macrophages were deduced from their intimate association via specialized cytoplasmic interdigitations.[101] Testicular macrophage secretions including cytokines and steroidogenic precursors have the potential to regulate steroidogenesis by Leydig cells. Moreover, inflammation-like responses induced by Leydig cell hyperstimulation with the LH agonist, human chorionic gonadotropin (hCG),[102] implicates the close association between Leydig cells with macrophages in intratesticular inflammatory responses involving mast cell secretion, polymorphonuclear leukocyte infiltration, and in increased blood flow, capillary permeability, and interstitial fluid accumulation.

The heterogeneous resident macrophage population comprises two main subsets defined by shape and distribution: ovoid intertubular macrophages close to Leydig cell clusters and blood vessels, and flattened peritubular macrophages with dendritic-shaped projections, directly adjacent to the tubules.[103,104] Most resident macrophages arise from the fetal yolk sac and liver, although many details of their origins and maintenance remain to be fully elaborated.[105] Testicular macrophages are crucial for normal testis somatic and spermatogenic development.[103]

ENDOCRINE AND LOCAL CONTROL OF SPERMATOGENESIS

Interactions Between the Testis, Hypothalamus, and Pituitary Gland

After puberty and in adults, control of testicular function is exerted by the hypothalamic-pituitary axis through episodic secretion of the gonadotropic hormones FSH and LH from the pituitary. These are in turn regulated by the pulsatile secretion of gonadotropin-releasing hormone (GnRH) from the hypothalamus into the pituitary portal vessels. Testosterone production by Leydig cells is regulated by LH acting via specific receptors. The secretion of GnRH, FSH, and LH is modulated by negative feedback from the testis via testosterone and by the Sertoli cell product, inhibin.[106] Testosterone, estradiol, and dihydrotestosterone (DHT) can suppress pituitary LH secretion.[107] Testosterone primarily acts at the hypothalamus by decreasing GnRH pulse frequency without changing pulse amplitude.[107] Thus although the seminiferous tubule and intertubular testis compartments may be considered separately, their functions are intimately related through paracrine and endocrine control mechanisms. The stimulatory actions of FSH and testosterone on seminiferous tubule function are modulated by local factors.

Inhibin is a dimer of an α and a β subunit, synthesized by males almost exclusively in Sertoli cells,[106] with inhibin B, an α:βB dimer, the predominant form in adult human testes. Inhibin secretion by Sertoli cells is directly related to FSH stimulation,[108] and serum inhibin B levels are used as a noninvasive marker of Sertoli cell number and function.

The activin proteins are produced throughout the body as dimers of inhibin β subunits that affect many cellular processes. Activins stimulate pituitary FSH production and are produced in the testis by spermatogenic, Sertoli, and Leydig cells and by macrophages where it acts on spermatogonia, spermatocytes, round spermatids, and Sertoli cells.[109] Many proteins that regulate activin signaling are produced in a highly regulated manner during spermatogenesis, including its potent inhibitor, follistatin.[110] Activin A drives Sertoli cell proliferation, and thereby contributes to determining the spermatogenic capacity of the testis; its overproduction can lead to tumor formation.[111]

In the fetal mouse testis, HSD17B1 and HSD17B3, the enzymes that convert androstenedione into testosterone, are synthesized in Sertoli cells in response to activin A.[112] Thus mouse fetal Leydig cells can only produce androstenedione, and circulating activin A levels may profoundly affect fetal hormone production, a relationship currently unexplored in the human fetal testis. LH stimulation of Leydig cells is central to normal pubertal onset.[113] It rapidly stimulates cholesterol mobilization and increases testosterone levels within 30 to 60 minutes of LH stimulation. Particularly in human, a repeated LH stimulus or a single hCG injection, which has a longer half-life, generates a prolonged testosterone response.[114]

Testosterone, the major adult Leydig cell androgenic product, enters the circulation through nearby capillaries. Testicular cells produce smaller amounts of weaker androgens, including androstenedione and dehydroepiandrosterone, and small amounts of DHT, which binds the androgen receptor with high affinity and generates ~25% of circulating 17β-estradiol.[115]

CONTROL OF SERTOLI CELL FUNCTION

The potential adult spermatogenic output of a testis is determined by its number of Sertoli cells.[116] As the nurse cells to developing sperm, Sertoli cells have receptors for the main hormonal regulators of spermatogenesis, functioning to “transduce” FSH and androgens signals to support germ cell development.

Sertoli Cells During Fetal and Postnatal Development

Sertoli cells are specified in early fetal life following expression of male sex determining genes, *Sry* and *Sox9* in mice.[116] They migrate from the coelomic epithelium into the nascent gonad and form the seminiferous cords as they envelope and direct masculinization of the primordial germ cells, which become gonocytes or prospermatogonia. Together with the fetal peritubular myoid cells, they secrete the extracellular matrix components that surround spermatogenic cords and form the basement membrane upon which spermatogonia ultimately reside. Because the cords later transform into adult testis seminiferous tubules, Sertoli cell proliferation during fetal life is a key determinant of future testis size and fertility, and this is stimulated by activin A from fetal Leydig cells. Activin, FSH, and androgens affect other Sertoli functions,[117] including synthesis of AMH required for Müllerian duct regression[116] and testosterone from the Leydig cell–derived androstenedione.[112,118] Endocrine disruptors, including environmental chemicals and pharmaceuticals, and lifestyle factors, such as diet and smoking, can restrict fetal testis development, including via effects on Sertoli cells.[119]

Many Sertoli cell products support gonocytes.[120] The enzyme CYP26B1 facilitates primordial germ cell masculinization by removing RA, degrading this trigger for meiosis. Platelet-derived growth factor (PDGF) and Notch signaling enhance gonocyte proliferation, while TGFβ, activin A, and testosterone each promote their entry into quiescence.[121] Intriguing evidence of cyclic functionality in fetal Sertoli cells suggests that these are a heterogeneous population.[122]

Sertoli cells proliferate through early postnatal life. In humans, fetal-neonatal proliferation increases their number ~5-fold, and a second prepubertal wave more than doubles it.[123] FSH and thyroid-stimulating hormone (TSH), and local products activin A and insulin-like growth factor, each stimulate postnatal Sertoli cell proliferation.[10]

Sertoli cells cease proliferation when puberty sets the upper limit for adult sperm production. As Sertoli cells “switch” to a mature phenotype, they acquire the capacity to support spermatogenesis, changing from PI3K/AKT/mTORC1 and MEK/ERK1/2 signaling pathways that promote proliferation to cyclic adenosine monophosphate–stimulated pathways for differentiation.[124] These are influenced by insulin-like growth factors, including relaxin,[125] and the interplay between RA and activin, collectively allowing formation of tight junctions between Sertoli cells.[10] Other endocrine and local products, including androgen, estrogens, and TSH, and exposure to exogenous agents such as inflammation and endocrine disrupting chemicals are implicated as affecting the timing of this switch and tight junction integrity.[10] Testis function during the pubertal transition in primates is discussed by Bhattacharya and colleagues.[126.]

Roles and Clinical Significance of Androgen and FSH in Sertoli Cell Development and Establishment of Spermatogenesis

Spermatogenesis is first established in conditions different from those present in adults. Both FSH and androgens are required for quantitatively normal spermatogenesis in rodents and primates, with roles that differ between species.[127,128] Androgen and FSH act independently, cooperatively, and synergistically to support establishment of spermatogenesis,[127] while RA produced in the seminiferous epithelium governs spermatogenic cycle initiation and drives germ cell meiotic entry.[129]

Cooperative actions of androgens and FSH on testicular somatic cells create an optimal germ:Sertoli cell ratio during the first spermatogenic wave in rodents, with proapoptotic factors balanced by FSH support of spermatogenesis.[117]

The rodent and primate studies establishing that both androgen and FSH are needed for optimal postnatal spermatogenic development have informed clinical therapies. Men with congenitally acquired or prepubertal onset hypogonadotrophic hypogonadism require coadministration of hCG or LH (stimulates Leydig cell androgen production) and FSH (stimulates Sertoli cell proliferation and maturation). In some men, treatments of 1 to 2 years may be required to allow for testis maturation and 64 days for sperm production from spermatogonia.[130]

Adult Sertoli Cell Function. The adult Sertoli cell is extraordinarily complex in terms of its cytoskeletal organization, polarity and bidirectional secretion, ability to rapidly and cyclically respond to the changing needs of the germ cells as they progress through their development, and its capacity to respond to endocrine and paracrine “cues” from the vasculature and adjacent somatic cells (see “Sertoli Cell–Mediated Regulation of Spermatogenesis”). For a comprehensive review on this extraordinary cell, refer to *The Sertoli Cell* book[131]; however, a few aspects of adult Sertoli cell function are discussed briefly here.

RA is a particularly important regulator of adult Sertoli cell function. It is cooperatively generated from dietary-derived retinol in the circulation via enzymes present in Sertoli and germ cells in the postnatal seminiferous epithelium.[129] RA actions in the testis are regulated via a complex interplay between enzymes that produce RA, storage and binding proteins that direct RA toward either classic RA-mediated or alternative PPAR-related signaling pathways, and enzymes that degrade RA. Specific roles of RA in germ cell development are detailed below (see “Sertoli Cell–Mediated Regulation of Spermatogenesis” and Fig. 111.9). Importantly, RA drives Sertoli cell cyclic function,[129] with vitamin A deficiency reversibly blocking spermatogonial differentiation and Sertoli cell cyclic behaviors in rodents[129]; RA produced by advanced germ cells contributes to regulation of the Sertoli cell, potentially explaining how transplanted germ cells control the cycle in the host niche.[27] This may involve enzymes other than the well-known ALDH1A enzymes.[132] Thus the Sertoli cell cycle, or “clock”, that arises during testis development is “set” by RA signaling and partially maintained by germ cells in the adult testis.[129]

Each adult Sertoli cell produces many components required by the stage-specific cohorts of spermatogenic cells it supports,[65] including lactate, the preferred substrate for germ cell glycolysis, metabolized from glucose. The dynamic junctions between Sertoli and germ cells required for adhesion and intercellular communication include desmosome-like, gap and adherens junctions with spermatogonia and spermatocytes, and the ESs with elongating spermatids.[133] The blood-testis barrier comprises inter–Sertoli cell junctions (Fig. 111.3) that are constantly remodeled in mid-spermatogenic stages to permit translocation of leptotene spermatocytes into the specialized milieu of the apical epithelium. Formation, maintenance, and remodeling of this barrier is regulated by androgens and FSH (see below), and it is locally regulated by paracrine factors that affect endocytosis and many signaling pathways, for example, by RA, activin, and TGFβ ligands.[13] In particular, RA and activin A appear to coordinate adult inter–Sertoli cell junction remodeling, with activin-promoting junction disassembly and RA-promoting junction assembly.[13]

Roles and Clinical Significance of Androgens and FSH in Adult Sertoli Cell Function. In adult mammals, androgens and FSH are both required to maintain quantitatively normal spermatogenesis. The functions of androgen and FSH on spermatogenesis are executed via receptors in Sertoli cells, although androgen also acts via AR in other testicular somatic cells.[127] The hormonal regulation of spermatogenesis has been addressed in Chapters 105–107, the requirement of gonadotropins for adult sperm production is the basis for hormonal male contraceptive development. A brief summary of the major actions of androgens and FSH to support adult spermatogenesis and their clinical significance is provided below.

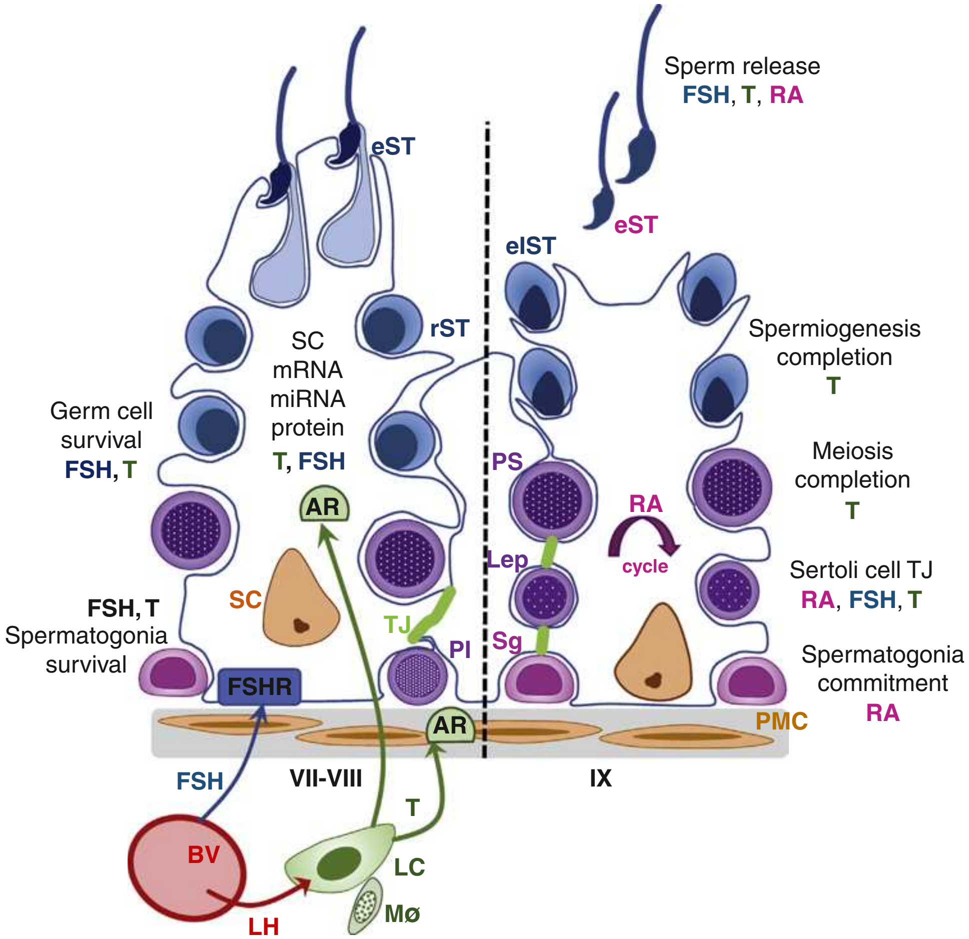


Figure 111.9 Overview of the regulation of spermatogenesis by follicle-stimulating hormone *(FSH)*, luteinizing hormone *(LH)*/testosterone and retinoic acid *(RA)*. *VII, VIII, and IX* denote the stages of the spermatogenic cycle represented in the diagram; *AR,* androgen receptor; *BV,* blood vessel; *elST,* elongating spermatid; *eST,* elongated spermatid; *FSHR,* FSH receptor; *LC,* Leydig cell; *Lep,* leptotene spermatocyte; *MØ,* macrophage; *Pl,* preleptotene spermatocyte; *PMC,* peritubular myoid cell; *PS,* pachytene spermatocyte; *rST,* round spermatid; *SC,* Sertoli cell; *Sg,* spermatogonia; *T,* testosterone; *TJ,* Sertoli cell tight junctions.

Withdrawal of gonadotrophic support, and the resulting suppression of testicular androgen production, suppresses sperm output and results in spermatogenic defects in adult rodents or primates, with variation between species[127] and dependence on the extent of gonadotropin suppression. In general, acute or long-term gonadotropin suppression in adults results in reduced spermatogonial entry into spermatogenesis, deficient blood-testis barrier function, reduced germ cell viability, particularly of pachytene spermatocytes and round spermatids in the mid-spermatogenic stages, and reduced ability of germ cells to enter meiosis or complete spermiogenesis or spermiation[127,128] (Fig. 111.9).

These processes have different thresholds for gonadotropin and androgen support. For example, in rodents, suppression without complete elimination of intratesticular testosterone prevents round spermatid maturation beyond mid-spermiogenesis, whereas the absence of androgen action on Sertoli cells due to genetic ablation of Sertoli cell AR blocks progression of meiosis by halting Sertoli cell functions that support the germline meiotic initiation transcriptional program.[134] Different spermatogenic defects arise from acute and long-term gonadotropin withdrawal. In primates, which are particularly sensitive to gonadotropin levels, acute suppression causes spermiation failure and alters spermatogonial kinetics, whereas long-term suppression is associated with a failure of spermatogonia to progress through spermatogenesis.[135]

Androgens and FSH synergize to support adult spermatogenesis, probably due to their overlapping influence on Sertoli cell signaling and gene transcription.[136] However, whether androgens and FSH have independent effects is a key, long-standing question. Spermatogonial development has a particular requirement for FSH, whereas meiosis is thought to absolutely require androgens.[127,128] Mice with a constitutively and strongly activated FSH receptor that lack an LH receptor do maintain sperm production, albeit reduced, even when residual testicular androgen action is blocked.[137] Thus very high FSH levels/receptor activation can compensate for a lack of LH/androgen in terms of stimulating fertility, suggesting that elevated FSH activity could stimulate androgen-dependent signaling pathways in Sertoli cells, despite their use of entirely different receptors.[137]

An acute and consistent suppression of both LH and FSH is the basis for hormonal contraception in men. Many different formulations have been trialed to promote maximum gonadotropin suppression while maintaining peripheral androgen actions, with some formulations showing particular promise.[138] The variability in individual responses is a key challenge: not every man shows adequate and consistent sperm count suppression, perhaps due to innate differences in the testicular response to gonadotropin withdrawal.[138] Additionally, the speed of sperm count suppression can vary, perhaps depending on whether or not a man shows an acute arrest of spermiation.[139]

Aspects relating to treatment of infertility are discussed in Chapter 112. Briefly, in terms of spermatogenic stimulation in men with infertility due to insufficient gonadotropin production (hypogonadotrophic hypogonadism) post-puberty, men are often treated with hCG (as an LH substitute), frequently in conjunction with FSH, for several months until sperm counts increase.[140] While hCG alone can stimulate spermatogenesis in some men, a combination of hCG and FSH administration is often administered to achieve optimal sperm production.[140]

Sertoli Cell–Mediated Regulation of Spermatogenesis. Sertoli cells coordinate the main endocrine signals from the HPG axis to stimulate and maintain spermatogenesis, and androgen and FSH receptors within Sertoli cells are necessary for optimal germ cell development.[128] The absolute dependency of germ cells on Sertoli cells was shown by diphtheria toxin (DTX) administration to transgenic mice expressing DTX receptors in Sertoli cells that caused their ablation.[141] This approach demonstrated that Sertoli cells (1) actively maintain peritubular myoid cell fate during development, (2) support adult Leydig cell and progenitor cell development and maintenance required for testosterone production, and (3) maintain the testicular vasculature structure and function central to Leydig cell steroidogenesis[142] (reviewed in 141). By showing that selective adult Sertoli cell ablation determines germ and Leydig cell numbers,[143] the DTX model unequivocally demonstrated that adult Sertoli cells determine both sperm output and androgen production.

RA signaling in Sertoli cells operates at a specific point in the spermatogenic cycle (also see "Adult Sertoli Cell Function") to coordinate multiple events and stimulate the spermatogenic "wave" (Fig. 111.9). RA signaling regulates spermatogonial commitment to germ cell entry into meiosis, blood-testis barrier remodeling, and elongated spermatid release at spermiation, events that occur in mid-spermatogenic stages corresponding to stages II-III in humans and VII-IX in mice. Their coordination may be achieved by local regulation of RA production and degradation between stages VII and IX, resulting in a pulse of RA signaling at VIII, which triggers RA-responsive processes (reviewed in 129).

The Sertoli cell creation of a stem cell niche and support for spermatogonial development are described earlier in this chapter (*Spermatogonia and Spermatogonial Stem Cells*).

Support of Meiosis and Spermiogenesis. Germ cells physically begin meiosis when type B spermatogonia divide to produce preleptotene spermatocytes.[127] The commitment to the meiotic program is made in rodents with differentiation of A to A1 spermatogonia, regulated by a pulse of RA in mid-spermatogenic stages and dependent on the RA-induced gene *Stra8* in mice.[129]

Preleptotene spermatocytes transition into leptotene spermatocytes as they traverse the inter–Sertoli cell junctions into the adluminal compartment where meiosis is completed[14] (Fig. 111.9). Junction remodeling facilitates meiotic progression, involves a multitude of adhesion and related signaling molecules (reviewed in 13) and is responsive to FSH, testosterone, RA, and activin signaling in Sertoli cells (see "Adult Sertoli Cell Function" and Fig. 111.9).

The completion of meiosis takes several weeks and depends on a complex transcriptional program that involves post-transcriptional regulation of mRNA stability via the "Meiosis Specific With Coiled-Coil Domain" (MEIOC) protein.[144] RNA sequencing has revealed the complex meiotic transcriptional program,[33] which is sensitive to androgen deficiency. When androgen levels or action are suppressed, meiosis can occur and spermatids are reduced; however, complete ablation of androgen action in Sertoli cells arrests spermatogenesis prior to meiotic division,[127] showing that round spermatid production absolutely requires androgen signaling (Fig. 111.9).

Pachytene spermatocyte and round spermatid survival requires hormonal support.[127] Some apoptosis of these cells occurs during normal spermatogenesis, and they are removed by Sertoli cells via phagocytosis and autophagy-related processes regulated by an androgen-dependent miRNA essential for normal fertility.[145]

Importantly, defects in meiosis and spermiogenesis observed in rodent models of hormone manipulation are not recapitulated in gonadotropin-suppressed monkeys and men. This is because the spermatogonial entry into meiosis in primates is exquisitely sensitive to gonadotropin suppression,[127] and thus monkeys and men undergoing gonadotropin suppression predominantly exhibit arrested spermatogonial development, likely due to reduced germ cell survival.

Sperm Release. The final step in testicular germ cell development is spermiation, whereby elongated spermatids are released into the tubule lumen for transport and further maturation in the epididymis (reviewed in 18). The involvement of a multitude of cytoskeletal and adhesion proteins and signaling cascades in Sertoli cells renders it highly vulnerable to disruption.[18,146] It is one of the first spermatogenic processes to be altered by FSH and androgen suppression and RA inhibition (Fig. 111.9), and it is also responsive to changes in estrogen production and signaling.[18,146] The initial phases involving spermatid translocation and adhesion structure removal are particularly vulnerable to altered RA signaling.[146] Changes in estrogen activity impair tubulobulbar complex formation required to remove the ES junctions between spermatids and Sertoli cells.[146] FSH and androgen actions on Sertoli cells are involved in mediating the final spermatid release via a FAK-integrin–based adhesion complex. Defects within spermatids can also prevent their release from Sertoli cells.[18] The consequences of acute defects in spermiation can significantly reduce sperm counts, an important response to hormone-based contraceptives.[18]

Control of Spermatogenesis: Challenges Ahead

Increased fundamental knowledge of the processes that underpin development of the healthy testis and fertile sperm are vital to address clinical and agricultural needs. The transgenerational effects of sperm on offspring health demand that information about spermatozoa formation and function is used to inform best practice in industrial, farming, and clinical settings. Delineation of epigenetic and genetic effects on germline formation at each life stage is critical, with species-specific and conserved impacts requiring evaluation. Current extensive reliance on artificial reproductive technologies and development of fertility preservation techniques, coupled with an ongoing demand for new male contraceptive approaches continues to underscore the need for research-based evidence of how spermatogenesis is controlled.

REFERENCES

1. de Kretser D, Kerr J. The cytology of the testis. In: Knobil E, Neill JD, eds. *The Physiology of Reproduction*. New York: Raven Press; 1994:1177–1290.
2. Dym M. Basement membrane regulation of Sertoli cells. *Endocr Rev*. 1994;15:102–115.
3. Hedger MP. Immunophysiology and pathology of inflammation in the testis and epididymis. *J Androl*. 2011;32:625–640.
4. Muller J, Skakkebaek NE. Fluctuations in the number of germ cells during the late foetal and early postnatal periods in boys. *Acta Endocrinol (Copenh)*. 1984;105:271–274.
5. Muller J, Skakkebaek NE. Quantification of germ cells and seminiferous tubules by stereological examination of testicles from 50 boys who suffered from sudden death. *Int J Androl*. 1983;6:143–156.
6. Paniagua R, Nistal M. Morphological and histometric study of human spermatogonia from birth to the onset of puberty. *J Anat*. 1984;139(Pt 3):535–552.
7. de Rooij DG. Stem cells in the testis. *Int J Exp Pathol*. 1998;79:67–80.
8. Russell LD, Griswold MD. *Sertoli Cell Biology*. Boston: Elsevier; 2004.
9. Vogl AW, Soucy LJ. Arrangement and possible function of actin filament bundles in ectoplasmic specializations of ground squirrel Sertoli cells. *J Cell Biol*. 1985;100:814–825.

10. Meroni SB, Galardo MN, Rindone G, et al. Molecular mechanisms and signaling pathways involved in Sertoli cell proliferation. *Front Endocrinol (Lausanne)*. 2019;10:224.
11. Cortes D, Muller J, Skakkebaek NE. Proliferation of Sertoli cells during development of the human testis assessed by stereological methods. *Int J Androl*. 1987;10:589–596.
12. Mruk DD, Cheng CY. The mammalian blood-testis barrier: its biology and regulation. *Endocr Rev*. 2015;36:564–591.
13. Stanton PG. Regulation of the blood-testis barrier. *Semin Cell Dev Biol*. 2016;59:166–173.
14. Russell L. Movement of spermatocytes from the basal to the adluminal compartment of the rat testis. *Am J Anat*. 1977;148:313–328.
15. Furuya S, Kumamoto Y, Sugiyama S. Fine structure and development of Sertoli junctions in human testis. *Arch Androl*. 1978;1:211–219.
16. Vogl AW, Pfeiffer DC, Mulholland D, et al. Unique and multifunctional adhesion junctions in the testis: ectoplasmic specializations. *Arch Histol Cytol*. 2000;63:1–15.
17. Vogl AW, Vaid KS, Guttman JA. The Sertoli cell cytoskeleton. *Adv Exp Med Biol*. 2008;636:186–211.
18. O'Donnell L, Nicholls PK, O'Bryan MK, et al. Spermiation: the process of sperm release. *Spermatogenesis*. 2011;1:14–35.
19. Petrie Jr RG, Morales CR. Receptor-mediated endocytosis of testicular transferrin by germinal cells of the rat testis. *Cell Tissue Res*. 1992;267:45–55.
20. La HM, Hobbs RM. Mechanisms regulating mammalian spermatogenesis and fertility recovery following germ cell depletion. *Cell Mol Life Sci*. 2019;76:4071–4102.
21. Pui HP, Saga Y. Gonocytes-to-spermatogonia transition initiates prior to birth in murine testes and it requires FGF signaling. *Mech Dev*. 2017;144:125–139.
22. Guo J, Nie X, Giebler M, et al. The dynamic transcriptional cell atlas of testis development during human puberty. *Cell Stem Cell*. 2020;26:262–276.e4.
23. Clermont Y. Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev*. 1972;52:198–236.
24. Nakagawa T, Sharma M, Nabeshima Y, et al. Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science*. 2010;328:62–67.
25. David S, Orwig KE. Spermatogonial stem cell culture in oncofertility. *Urol Clin North Am*. 2020;47:227–244.
26. Clouthier DE, Avarbock MR, Maika SD, et al. Rat spermatogenesis in mouse testis. *Nature*. 1996;381:418–421.
27. Ogawa T, Dobrinski I, Avarbock MR, et al. Xenogeneic spermatogenesis following transplantation of hamster germ cells to mouse testes. *Biol Reprod*. 1999;60:515–521.
28. Hermann BP, Sukhwani M, Winkler F, et al. Spermatogonial stem cell transplantation into rhesus testes regenerates spermatogenesis producing functional sperm. *Cell Stem Cell*. 2012;11:715–726.
29. La HM, Makela JA, Chan AL, et al. Identification of dynamic undifferentiated cell states within the male germline. *Nat Commun*. 2018;9:2819.
30. Hara K, Nakagawa T, Enomoto H, et al. Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell Stem Cell*. 2014;14:658–672.
31. Ikami K, Tokue M, Sugimoto R, et al. Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis. *Development*. 2015;142:1582–1592.
32. Shirakawa T, Yaman-Deveci R, Tomizawa S, et al. An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a Kit-positive identity. *Development*. 2013;140:3565–3576.
33. Guo J, Grow EJ, Mlcochova H, et al. The adult human testis transcriptional cell atlas. *Cell Res*. 2018;28:1141–1157.
34. Agrimson KS, Onken J, Mitchell D, et al. Characterizing the spermatogonial response to retinoic acid during the onset of spermatogenesis and following synchronization in the neonatal mouse testis. *Biol Reprod*. 2016;95:81.
35. Hogarth CA, Arnold S, Kent T, et al. Processive pulses of retinoic acid propel asynchronous and continuous murine sperm production. *Biol Reprod*. 2015;92:37.
36. Kitadate Y, Jorg DJ, Tokue M, et al. Competition for mitogens regulates spermatogenic stem cell homeostasis in an open niche. *Cell Stem Cell*. 2019;24:79–92.e6.
37. Yang QE, Kim D, Kaucher A, et al. CXCL12-CXCR4 signaling is required for the maintenance of mouse spermatogonial stem cells. *J Cell Sci*. 2013;126:1009–1020.
38. McIver SC, Loveland KL, Roman SD, et al. The chemokine CXCL12 and its receptor CXCR4 are implicated in human seminoma metastasis. *Andrology*. 2013;1:517–529.
39. Hofmann MC. Gdnf signaling pathways within the mammalian spermatogonial stem cell niche. *Mol Cell Endocrinol*. 2008;288:95–103.
40. Saracino R, Capponi C, Di Persio S, et al. Regulation of Gdnf expression by retinoic acid in Sertoli cells. *Mol Reprod Dev*. 2020;87:419–429.
41. Meistrich ML, an Beek MEAB. Spermatogonial stem cells. In: Desjardins C, Ewing LL, eds. *Cell and Molecular Biology of the Testis*. New York: Oxford University Press; 1993:266–295.
42. Sharma S, Wistuba J, Pock T, et al. Spermatogonial stem cells: updates from specification to clinical relevance. *Hum Reprod Update*. 2019;25:275–297.
43. Guo J, Grow EJ, Yi C, et al. Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell*. 2017;21:533–546.e6.
44. Hermann BP, Cheng K, Singh A, et al. The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep*. 2018;25:1650–1667.e8.
45. Shami AN, Zheng X, Munyoki SK, et al. Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev Cell*. 2020;54:529–547.e12.
46. Loveland KL, Schlatt S. Stem cell factor and c-kit in the mammalian testis: lessons originating from Mother Nature's gene knockouts. *J Endocrinol*. 1997;153:337–344.
47. Yoshinaga K, Nishikawa S, Ogawa M, et al. Role of c-kit in mouse spermatogenesis: identification of spermatogonia as a specific site of c-kit expression and function. *Development*. 1991;113:689–699.
48. Shen H, Shih J, Hollern DP, et al. x Integrated molecular characterization of testicular germ cell tumors. *Cell Rep*. 1991;23:3392–3406.
49. Tan K, Song HW, Wilkinson MF. Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. *Development*. 2020:147.
50. Kim S, Bardwell VJ, Zarkower D. Cell type-autonomous and non-autonomous requirements for Dmrt1 in postnatal testis differentiation. *Dev Biol*. 2007;307:314–327.
51. Matson CK, Murphy MW, Sarver AL, et al. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature*. 2011;476:101–104.
52. Matson CK, Murphy MW, Griswold MD, et al. The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev Cell*. 2010;19:612–624.
53. Caldwell KA, Handel MA. Protamine transcript sharing among postmeiotic spermatids. *Proc Natl Acad Sci U S A*. 1991;88:2407–2411.
54. Rezende-Melo CA, Caldeira-Brant AL, Drumond-Bock AL, et al. Spermatogonial asynchrony in Tex14 mutant mice lacking intercellular bridges. *Reproduction*. 2020;160:205–215.
55. Baarends WM, van der Laan R, Grootegoed JA. DNA repair mechanisms and gametogenesis. *Reproduction*. 2001;121:31–39.
56. Pellestor F, Anahory T, Lefort G, et al. Complex chromosomal rearrangements: origin and meiotic behavior. *Hum Reprod Update*. 2001;17:476–494.
57. Kierszenbaum AL, Tres LL. The acrosome-acroplaxome-manchette complex and the shaping of the spermatid head. *Arch Histol Cytol*. 2004;67:271–284.
58. Pleuger C, Lehti MS, Dunleavy JE, et al. Haploid male germ cells-the Grand Central Station of protein transport. *Hum Reprod Update*. 2020;26:474–500.
59. Cho C, Willis WD, Goulding EH, et al. Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet*. 2001;28:82–86.
60. Nanassy L, Liu L, Griffin J, et al. The clinical utility of the protamine 1/protamine 2 ratio in sperm. *Protein Pept Lett*. 2011;18:772–777.

61. Hammoud SS, Nix DA, Zhang H, et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature*. 2009;460:473–478.
62. Oikawa M, Simeone A, Hormanseder E, et al. Epigenetic homogeneity in histone methylation underlies sperm programming for embryonic transcription. *Nat Commun*. 2020;11:3491.
63. Denomme MM, Haywood ME, Parks JC, et al. The inherited methylome landscape is directly altered with paternal aging and associated with offspring neurodevelopmental disorders. *Aging Cell*. 2020;19:e13178.
64. Potabattula R, Zacchini F, Ptak GE, et al. Increasing methylation of sperm rDNA and other repetitive elements in the aging male mammalian germline. *Aging Cell*. 2020;19:e13181.
65. Soumillon M, Necsulea A, Weier M, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*. 2013;3:2179–2190.
66. Sassone-Corsi P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science*. 2002;296:2176–2178.
67. Tang C, Xie Y, Yu T, et al. m(6)A-dependent biogenesis of circular RNAs in male germ cells. *Cell Res*. 2020;30:211–228.
68. Lehti MS, Sironen A. Formation and function of the manchette and flagellum during spermatogenesis. *Reproduction*. 2016;151:R43–R54.
69. Dun MD, Aitken RJ, Nixon B. The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. *Hum Reprod Update*. 2012;18:420–435.
70. Fujihara Y, Lu Y, Noda T, et al. Spermatozoa lacking fertilization influencing membrane protein (FIMP) fail to fuse with oocytes in mice. *Proc Natl Acad Sci U S A*. 2020;117:9393–9400.
71. Noda T, Lu Y, Fujihara Y, et al. Sperm proteins SOF1, TMEM95, and SPACA6 are required for sperm-oocyte fusion in mice. *Proc Natl Acad Sci U S A*. 2020;117:11493–11502.
72. Zhou W, Stanger SJ, Anderson AL, et al. Mechanisms of tethering and cargo transfer during epididymosome-sperm interactions. *BMC Biol*. 2019;17:35.
73. Conine CC, Sun F, Song L, et al. Small RNAs gained during epididymal transit of sperm are essential for embryonic development in mice. *Dev Cell*. 2018;46:470–480.e3.
74. Nixon B, De Iuliis GN, Dun MD, et al. Profiling of epididymal small non-protein-coding RNAs. *Andrology*. 2019;7:669–680.
75. Nixon B, De Iuliis GN, Hart HM, et al. Proteomic profiling of mouse epididymosomes reveals their contributions to post-testicular sperm maturation. *Mol Cell Proteomics*. 2019;18:S91–s108.
76. Ding D, Liu J, Dong K, et al. Mitochondrial membrane-based initial separation of MIWI and MILI functions during pachytene piRNA biogenesis. *Nucleic Acids Res*. 2019;47:2594–2608.
77. Hildebrandt F, Benzing T, Katsanis N. Ciliopathies. *N Engl J Med*. 2011;364:1533–1543.
78. Lehti MS, Sironen A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol Reprod*. 2017;97:522–536.
79. Xu K, Yang L, Zhang L, et al. Lack of AKAP3 disrupts integrity of the subcellular structure and proteome of mouse sperm and causes male sterility. *Development*. 2020:147.
80. Honaramooz A, Li MW, Penedo MC, et al. Accelerated maturation of primate testis by xenografting into mice. *Biol Reprod*. 2004;70:1500–1503.
81. Zeng W, Avelar GF, Rathi R, et al. The length of the spermatogenic cycle is conserved in porcine and ovine testis xenografts. *J Androl*. 2006;27:527–533.
82. Schlatt S, Ehmcke J. Regulation of spermatogenesis: an evolutionary biologist's perspective. *Semin Cell Dev Biol*. 2014;29:2–16.
83. Johnson L, McKenzie KS, Snell JR. Partial wave in human seminiferous tubules appears to be a random occurrence. *Tissue Cell*. 1996;28:127–136.
84. Muciaccia B, Boitani C, Berloco BP, et al. Novel stage classification of human spermatogenesis based on acrosome development. *Biol Reprod*. 2013;89:60.
85. Sinha Hikim AP, Swerdloff RS. Hormonal and genetic control of germ cell apoptosis in the testis. *Rev Reprod*. 1999;4:38–47.
86. Meehan T, Loveland KL, de Kretser D, et al. Developmental regulation of the bcl-2 family during spermatogenesis: insights into the sterility of bcl-w-/- male mice. *Cell Death Differ*. 2001;8:225–233.
87. Lysiak JJ, Zheng S, Woodson R, et al. Caspase-9-dependent pathway to murine germ cell apoptosis: mediation by oxidative stress, BAX, and caspase 2. *Cell Tissue Res*. 2007;328:411–419.
88. Elliott MR, Zheng S, Park D, et al. Unexpected requirement for ELMO1 in clearance of apoptotic germ cells in vivo. *Nature*. 2010;467:333–337.
89. Krausz C, Riera-Escamilla A, Moreno-Mendoza D, et al. Genetic dissection of spermatogenic arrest through exome analysis: clinical implications for the management of azoospermic men. *Genet Med*. 2020.
90. Jiao SY, Yang YH, Chen SR. Molecular genetics of infertility: loss-of-function mutations in humans and corresponding knockout/mutated mice. *Hum Reprod Update*. 2020.
91. Kuroda S, Usui K, Sanjo H, et al. Genetic disorders and male infertility. *Reprod Med Biol*. 2020;19:314–322.
92. Aitken RJ, Baker MA. The role of genetics and oxidative stress in the etiology of male infertility-a unifying hypothesis? *Front Endocrinol (Lausanne)*. 2020;11:581838.
93. Bunkar N, Pathak N, Lohiya NK, et al. Epigenetics: a key paradigm in reproductive health. *Clin Exp Reprod Med*. 2016;43:59–81.
94. Hogg K, Western PS. Refurbishing the germline epigenome: out with the old, in with the new. *Semin Cell Dev Biol*. 2015;45:104–113.
95. Zhai L, Wang L, Teng F, et al. Argonaute and argonaute-bound small RNAs in stem cells. *Int J Mol Sci*. 2016;17:208.
96. Wu PH, Fu Y, Cecchini K, et al. The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet*. 2020;52:728–739.
97. Özata DM, Yu T, Mou H, et al. Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol*. 2020;4:156–168.
98. Winnall WR, Hedger MP. Phenotypic and functional heterogeneity of the testicular macrophage population: a new regulatory model. *J Reprod Immunol*. 2013;97:147–158.
99. Chemes H, Cigorraga S, Bergada C, et al. Isolation of human Leydig cell mesenchymal precursors from patients with the androgen insensitivity syndrome: testosterone production and response to human chorionic gonadotropin stimulation in culture. *Biol Reprod*. 1992;46:793–801.
100. Bhushan S, Theas MS, Guazzone VA, et al. Immune cell subtypes and their function in the testis. *Front Immunol*. 2020;11:583304.
101. Meinhardt A, Hedger MP. Immunological, paracrine and endocrine aspects of testicular immune privilege. *Mol Cell Endocrinol*. 2011;335:60–68.
102. Setchell BP, Rommerts FF. The importance of the Leydig cells in the vascular response to hCG in the rat testis. *Int J Androl*. 1985;8:436–440.
103. DeFalco T, Potter SJ, Williams AV, et al. Macrophages contribute to the spermatogonial niche in the adult testis. *Cell Rep*. 2015;12:1107–1119.
104. Hume DA, Perry VH, Gordon S. The mononuclear phagocyte system of the mouse defined by immunohistochemical localisation of antigen F4/80: macrophages associated with epithelia. *Anat Rec*. 1984;210:503–512.
105. DeFalco T, Bhattacharya I, Williams AV, et al. Yolk-sac-derived macrophages regulate fetal testis vascularization and morphogenesis. *Proc Natl Acad Sci U S A*. 2014;111:E2384–E2393.
106. Wijayarathna R, de Kretser DM. Activins in reproductive biology and beyond. *Hum Reprod Update*. 2016;22:342–357.
107. Santen RJ. Is aromatization of testosterone to estradiol required for inhibition of luteinizing hormone secretion in men? *J Clin Invest*. 1975;56:1555–1563.
108. Anderson RA, Wallace EM, Groome NP, et al. Physiological relationships between inhibin B, follicle stimulating hormone secretion and spermatogenesis in normal men and response to gonadotrophin suppression by exogenous testosterone. *Hum Reprod*. 1997;12:746–751.
109. Loveland KL. Activins in male reproduction. In: Jegou B, Skinner M, eds. *Encyclopedia of Reproduction.* Vol I. Male Reproduction. Boston: Academic Press; 2018:208–214.
110. Itman C, Wong C, Whiley PA, et al. TGFbeta superfamily signaling regulators are differentially expressed in the developing and adult mouse testis. *Spermatogenesis*. 2011;1:63–72.

111. Matzuk MM, Finegold MJ, Mather JP, et al. Development of cancer cachexia-like syndrome and adrenal tumors in inhibin-deficient mice. *Proc Natl Acad Sci U S A*. 1994;91:8817–8821.
112. Whiley PAF, O'Donnell L, Moody SC, et al. Activin A determines steroid levels and composition in the fetal testis. *Endocrinology*. 2020:161.
113. Simoni M, Gromoll J, Nieschlag E. Molecular pathophysiology and clinical manifestations of gonadotropin receptor defects. *Steroids*. 1998;63:288–293.
114. Padron RS, Wischusen J, Hudson B, et al. Prolonged biphasic response of plasma testosterone to single intramuscular injections of human chorionic gonadotropin. *J Clin Endocrinol Metab*. 1980;50:1100–1104.
115. Chen YD, Kraemer FB, Reaven GM. Identification of specific high density lipoprotein-binding sites in rat testis and regulation of binding by human chorionic gonadotropin. *J Biol Chem*. 1980;255:9162–9167.
116. Makela JA, Koskenniemi JJ, Virtanen HE, et al. Testis development. *Endocr Rev*. 2019;40:857–905.
117. O'Shaughnessy PJ, Monteiro A, Abel M. Testicular development in mice lacking receptors for follicle stimulating hormone and androgen. *PLoS One*. 2012;7:e35136.
118. Archambeault DR, Yao HH. Activin A, a product of fetal Leydig cells, is a unique paracrine regulator of Sertoli cell proliferation and fetal testis cord expansion. *Proc Natl Acad Sci U S A*. 2010;107:10526–10531.
119. Juul A, Almstrup K, Andersson AM, et al. Possible fetal determinants of male infertility. *Nat Rev Endocrinol*. 2014;10:553–562.
120. Manku G, Culty M. Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. *Reproduction*. 2015;149:R139–R157.
121. Young JC, Wakitani S, Loveland KL. TGF-beta superfamily signaling in testis formation and early male germline development. *Semin Cell Dev Biol*. 2015;45:94–103.
122. Timmons PM, Rigby PW, Poirier F. The murine seminiferous epithelial cycle is pre-figured in the Sertoli cells of the embryonic testis. *Development*. 2002;129:635–647.
123. Yang Q-E, Oatley JM. Early postnatal interactions between Sertoli and germ cells. In: Griswold MD, ed. *Sertoli Cell Biology*. 2nd ed. Waltham, MA: Elsevier; 2015:81–98.
124. Bhattacharya I, Pradhan BS, Sarda K, et al. A switch in Sertoli cell responsiveness to FSH may be responsible for robust onset of germ cell differentiation during prepubartal testicular maturation in rats. *Am J Physiol Endocrinol Metab*. 2012;303:E886–E898.
125. Nascimento AR, Macheroni C, Lucas TF, et al. Crosstalk between FSH and relaxin at the end of the proliferative stage of rat Sertoli cells. *Reproduction*. 2016;152:613–628.
126. Bhattacharya I, Sen Sharma S, Majumdar SS. Pubertal orchestration of hormones and testis in primates. *Mol Reprod Dev*. 2019;86:1505–1530.
127. O'Donnell L, Stanton P, de Kretser DM. Endocrinology of the male reproductive system and spermatogenesis. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext*. South Dartmouth, MA: MDText.com; 2016.
128. O'Shaughnessy PJ. Hormonal control of germ cell development and spermatogenesis. *Semin Cell Dev Biol*. 2014;29:55–65.
129. Hogarth C. Retinoic acid metabolism, signaling, and function in the adult testis. In: Griswold MD, ed. *Sertoli Cell Biology*. Waltham, MA: Elsevier; 2015:247–272.
130. McLachlan RI. Approach to the patient with oligozoospermia. *J Clin Endocrinol Metab*. 2013;98:873–880.
131. Griswold MD, ed. *Sertoli Cell Biology*. 2nd ed. Waltham, MA: Elsevier; 2015.
132. Beedle MT, Stevison F, Zhong G, et al. Sources of all-trans retinal oxidation independent of the aldehyde dehydrogenase 1A isozymes exist in the postnatal testis. *Biol Reprod*. 2019;100:547–560.
133. Cheng CY, Mruk DD. Biochemistry of Sertoli cell/germ cell junctions, germ cell transport, and spermiation in the seminiferous epithelium. In: Griswold MD, ed. *The Sertoli Cell Biology*. Waltham, MA: Elsevier; 2015:333–383.
134. Larose H, Kent T, Ma Q, et al. Regulation of meiotic progression by Sertoli-cell androgen signaling. *Mol Biol Cell*. 2020;mbcE20050334.
135. McLachlan RI, O'Donnell L, Meachem SJ, et al. Hormonal regulation of spermatogenesis in primates and man: insights for development of the male hormonal contraceptive. *J Androl*. 2002;23:149–162.
136. Santi D, Crepieux P, Reiter E, et al. Follicle-stimulating Hormone (FSH) Action on spermatogenesis: a focus on physiological and therapeutic roles. *J Clin Med*. 2020;9:10–14.
137. Oduwole OO, Peltoketo H, Poliandri A, et al. Constitutively active follicle-stimulating hormone receptor enables androgen-independent spermatogenesis. *J Clin Invest*. 2018;128:1787–1792.
138. Thirumalai A, Page ST. Male hormonal contraception. *Annu Rev Med*. 2020;71:17–31.
139. McLachlan RI, O'Donnell L, Stanton PG, et al. Effects of testosterone plus medroxyprogesterone acetate on semen quality, reproductive hormones, and germ cell populations in normal young men. *J Clin Endocrinol Metab*. 2002;87:546–556.
140. Hayes F, Dwyer A, Pitteloud N. Hypogonadotropic hypogondism (HH) and gonadotropin therapy. In: McLachlan RI, Section, eds. *www.ENDOTEXT.org, Endocrinology of Male Reproduction*. South Dartmouth, MA: MDTEXTCOM Inc.; 2013. [chapter 5].
141. Smith LB, O'Shaughnessy PJ, Rebourcet D. Cell-specific ablation in the testis: what have we learned? *Andrology*. 2015;3:1035–1049.
142. Rebourcet D, Wu J, Cruickshanks L, et al. Sertoli cells modulate testicular vascular network development, structure, and function to influence circulating testosterone concentrations in adult male mice. *Endocrinology*. 2016;157:2479–2488.
143. Rebourcet D, Darbey A, Monteiro A, et al. Sertoli cell number defines and predicts germ and Leydig cell population sizes in the adult mouse testis. *Endocrinology*. 2017;158:2955–2969.
144. Abby E, Tourpin S, Ribeiro J, et al. Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. *Nat Commun*. 2016;7:10324.
145. Panneerdoss S, Viswanadhapalli S, Abdelfattah N, et al. Cross-talk between miR-471-5p and autophagy component proteins regulates LC3-associated phagocytosis (LAP) of apoptotic germ cells. *Nat Commun*. 2017;8:598.
146. Kumar A, Raut S, Balasinor NH. Endocrine regulation of sperm release. *Reprod Fertil Dev*. 2018;30:1595–1603.

112

Prevalence and Causes of Male Infertility

Christina Wang, Ronald S. Swerdloff, and Csilla Krausz

OUTLINE

Introduction, 1878
Definition and Prevalence of Male Infertility, 1878
Common Causes of Male Infertility, 1878
  Hypothalamic-Pituitary Disorders, 1879
  Testicular Disorders With Spermatogenesis Defects, 1881
  Sperm Motility, Morphology, Function, or Genetic Incompetence—Spermatogenesis Qualitative Defects, 1884
  Ductal Obstruction or Dysfunction, 1885
  Idiopathic Azoospermia, Oligozoospermia, Asthenozoospermia, and Teratozoospermia, 1886
Other Possible Contributing Factors to Male Infertiliy, 1886
  Obesity, 1886
  Environmental Factors: Endocrine Disruptors, 1886
  Cigarette Smoking, 1887
  Cell Phone, 1887
  Scrotal Hyperthermia, 1887
Epigenetics and Male Infertility, 1887
Summary and Future Directions, 1888

INTRODUCTION

In this chapter, we will consider the prevalence and common causes of male infertility. We will briefly describe environmental factors that may impact sperm production and development. Lastly, though not in clinical use, we will discuss the exploratory research on epigenetic regulation of spermatogenesis, as these epigenetic modifications may not only influence the reproductive health of the man but also his progeny. This chapter should be read in conjunction with chapters on Regulation of Spermatogenesis (see Chapters 111 and 113).

DEFINITION AND PREVALENCE OF MALE INFERTILITY

The World Health Organization (WHO) defines infertility as "a disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse."[1] However, it should be noted that many men with male infertility due to defects in spermatogenesis are otherwise healthy. Thus whether male infertility is a disease is debatable but conditions with a pathophysiological basis should qualify and be treated as a disease.[2] The main factors influencing spontaneous conception are the duration of unwanted infertility, the age of the female partner, and disease-related infertility (e.g., known chromosomal anomalies, testicular diseases, and hypogonadotropic hypogonadism).[3] Infertility affects about 15% of couples. Males are solely responsible for about 20% to 30% of the infertility and contribute to about 50% (ranges between 20% in sub-Saharan Africa and 70% in Middle East) of all cases of infertility globally. A meta-analysis of the global prevalence indicates that 2.5% to 12% of men in the general population are infertile. The prevalence is estimated to be 8% to 12% in Central/Eastern Europe, 9% in Australia, 7.5% in Europe, 4.5% to 6% in North America, and 2.8% to 5.7% in sub-Saharan Africa and 8.5% to 11.7% in North/West Africa.[4] Causes of this regional variation in male infertility rates in the general population are not known and may be subject to reporting errors.

KEY POINTS

- Male factors are solely responsible for about 20% to 30% of infertility and may be a factor in 50% of all cases.
- Male infertility rates differ in different geographical regions of the world.

COMMON CAUSES OF MALE INFERTILITY

While many men with male infertility have oligozoospermia (decrease in number of sperm cells in the ejaculate compared with reference range) or azoospermia (no spermatozoa in the ejaculate), some infertile men have normal sperm concentration. The generally accepted reference ranges for fertile men in the WHO Manual for the Examination of and Processing of Human Semen[5] are shown in Chapter 12 (Table 12.3). The data were based on semen samples of over 1941 fertile men whose partners had a time to pregnancy within 12 months.[6] Azoospermia occurs in about 16% of infertile men presenting at infertility and andrology clinics.[7,8] Over 80% of men with infertility have low sperm concentration (oligozoospermia) that may be associated with a decrease in sperm motility (asthenozoospermia) and a decreased percentage of spermatozoa with normal morphology (teratozoospermia). Others do not have low sperm concentration but may have qualitative defects in sperm motility, abnormal sperm morphology, and/or function. The common causes of male infertility are categorized into: (1) hypothalamic-pituitary disorders; (2) spermatogenesis defects with low sperm concentration or absent spermatozoa; (3) abnormal sperm motility, morphology, functional defects, and/or genetic incompetency; (4) posttesticular duct obstruction; and (5) idiopathic (unknown origin). In each category, the cause can be congenital (usually due or ascribed to possible genetic defects) or acquired (Table 112.1).[9] There is currently a large number of infertile men categorized as unknown or idiopathic infertility. With modern genetic analyses tools, many of these men may be shown to have a genetic cause of infertility although environmental factors may also contribute.[10] The frequency of the causes may vary between regions and Table 112.1 shows the common causes of infertility based on data from andrology and reproductive medicine clinics.[7–9,11]

Hypothalamic-Pituitary Disorders

Any hypothalamic or pituitary disease/dysfunction may cause gonadotropin-releasing hormone (GnRH) or gonadotropin deficiency (hypogonadotropic hypogonadism) presenting with secondary male hypogonadism (testosterone deficiency) associated with infertility. These conditions are characterized by low serum concentrations of luteinizing hormone (LH), follicle-stimulating hormone (FSH), and testosterone. In a large andrology clinic, hypothalamic-pituitary disorders occur in 1% to 2% of infertile men.[8,11] It is important to diagnose hypogonadotropic hypogonadism, as there are specific therapies available (pulsatile GnRH or gonadotropins). These conditions can be subdivided into congenital or acquired disorders (see Chapters 105 and 106).

KEY POINTS: Hypothalamic-Pituitary Disorders

- Hypothalamic-pituitary endocrine disorders (with low serum testosterone) represent a small percentage of overall cases of male infertility.
- Congenital hypogonadotropic hypogonadism can be due to defects in the development and migration of the GnRH neurons (Kallmann syndrome associated with anosmia and other developmental defects) or in impaired GnRH secretion or action (GnRH receptor defects) not associated with olfactory anomalies. Congenital hypogonadotropic hypogonadism undergoes spontaneous remission in 10 to 20% of patients (both adolescents and adults).
- Other genetic defects such as gonadotropins and gonadotropin receptors are very rare.
- More common acquired hypothalamic-pituitary disorders are sellar and suprasellar lesions: pituitary adenomas, craniopharyngiomas, meningiomas, lymphocytic hypophysitis, Rathke's cleft cysts, germinomas, and drugs (opioids and androgenic steroids) and severe systemic illness.
- Since specific therapies are available, it is important to diagnose these disorders

Congenital Hypogonadotropic Hypogonadism (Table 112.2)

GnRH Neurons Development or Migration Defects (Associated With Anosmia/Hyposmia). Patients with isolated congenital hypogonadotropic hypogonadism (cHH) have defective GnRH production or secretion, low LH and FSH levels, and present with delayed puberty (see Chapters 102 and 103). Genetic defects can be identified in over 40% of cHH using current gene panel screening methodologies.[12,13] The genes associated with cHH can be subdivided to those that are associated with GnRH neuron development and/or migration (Kallmann syndrome) with anosmia or hyposmia or those associated with defective GnRH secretion (normosmia) or action. Many heterogeneous genetic mutations cause cHH (Table 112.2) and the mutations can be digenic or oligogenic.[14] Genetic analyses may help to distinguish between cHH and constitutional delayed puberty.[15,16] About 10% to 20% of patients with cHH undergo spontaneous remissions after testosterone treatment suggesting a functional defect or misclassification of patients with delayed puberty. For this reason, endocrinologists may elect to withdraw hormone replacement therapy periodically in adult and adolescent patients diagnosed with cHH to assess for recovery.[17] Kallmann syndrome, the X-linked form, is due to mutations of *ANOS1* (*KAL1*) and hypogonadism is associated with anosmia or hyposmia and midline facial defects such as harelip and cleft palate and unilateral kidney agenesis. Other genetic defects that cause cHH include fibroblast growth factor 1 receptor *(FGFR1)* or *FGF 8*, and prokineticin 2 (*PROK2*) and its receptor (*PROKR2*) that may present clinically with a spectrum of findings including anosmia to normosmia, midline defects, and skeletal and dental abnormalities. Gene mutations of chromodomain helicase-DNA-binding protein 7 (*CHD7*) are part of the CHARGE syndrome (coloboma, heart disease, choanal atresia, retarded growth and development/central nervous system anomalies, genital anomalies/hypogonadism, and ear anomalies/deafness) and patients may have anosmia to normosmia.

GnRH Secretion Defects (Normosmia). There are many genes involved in GnRH secretion, mutations of which result in cHH in males. *KISS1* and *KISS1R* genes encoding kisspeptin 1 and kisspeptin 1 receptor regulate GnRH secretion and onset of puberty. Activating mutations of *KISS1* are associated with precocious puberty, whereas disrupting mutations of *KISS1R* lead to cHH.[18,19] Mutations of the leptin (*LEP*) and leptin receptor (*LEPR*) genes result in extreme obesity and cHH. Another important regulator of GnRH secretion is neurokinin B (NK3R) that encoded by the gene *TAC* and mutations of *TAC3* and its receptor *TACR3* results in cHH and delayed puberty. *PCSK1* gene product protein convertase subtilisin/kexin-type 1 processes several precursor hormones (proinsulin, proopiomelanocortin) to release bioactive hormones. Mutations of *PCSK1* are associated with obesity and cHH. DAX-1 encoded by *NROB1* is important for the development of the adrenals and hypothalamic-pituitary-gonadal axis. NROB1 mutations result in adrenal insufficiency and cHH.[10]

GnRH and Its Receptor. Mutations of *GnRH* gene are very rare but mutation of its receptor gene (*GnRHR)* results in cHH with normosmia and accounts for 3.5% to 16% of sporadic cHH and over 40% of familial normosmic cHH.[13,20]

LH and LH Receptor; FSH and FSH Receptor Defects. LH gene mutations are very rare and the patient will present with delayed puberty, cryptorchidism, and hypogonadism. LH receptor gene inactivating mutations have variable presentation from female phenotype to only male infertility similar to androgen receptor defects. Mutations of FSH or FSH receptor result in variable phenotypes from azoospermia to normal spermatogenesis.[21]

Acquired Hypogonadotropic Hypogonadism. Acquired hypogonadotropic hypogonadism may be associated with other pituitary hormone deficiencies (see Chapter 106).

Tumors. The commonest tumors associated with hypogonadotropic hypogonadism are pituitary macroadenomas, microprolactinomas, craniopharyngiomas, or meningiomas. Hypogonadotropic hypogonadism often persists in men with macroprolactinoma even after treatment with dopamine agonists and serum prolactin concentration normalized. After surgical or radiation therapy of either macroadenomas or craniopharyngiomas, hypogonadotropic hypogonadism often ensues.

Infiltrative Diseases. Infiltrative diseases include sarcoidosis, histiocytosis, tuberculosis, fungal infections, transfusion-related hemosiderosis,[22] and hemochromatosis (a genetic disorder of mucosal iron transport that results in increased deposition of iron in many tissues including the pituitary gland). Hemochromatosis-associated hypogonadotropic hypogonadism is nearly always of postpubertal onset.[23,24]

Lymphocytic Hypophysitis or Infundibulitis. This autoimmune condition affecting the pituitary and/or the infundibulum results in enlargement of the pituitary gland. Lymphocytic hypophysitis or

TABLE 112.1 Common Causes of Male Infertility

Congenital (Genetic)	Acquired (Nongenetic)
Hypothalamic-Pituitary Disorders	
• **Congenital hypogonadotropic hypogonadism with anosmia or hyposmia (Kallmann syndrome)** • **Congenital hypogonadotropic hypogonadism with normosmia (see Table 112.2)** • **LH or LH receptor, FSH or FSH receptor gene defects**	• Pituitary or CNV tumors • Postsurgery or irradiation, trauma, vascular • Infiltrative diseases (hemochromatosis, sarcoidosis) • Lymphocytic hypophysitis • Traumatic brain injury • Drugs • Anabolic steroids • Systemic illness
Spermatogenesis Defects With Decrease in Sperm Concentration	
• **Chromosomal aneuploidy (XXY, XX males)** • **Chromosomal structural anomalies (translocation, inversion)** • **Y chromosome microdeletion (AZFa, AZFb, and AZFc; gr/gr deletion in AZFc region)** • **Chromosomal copy number variations** • **Mild partial androgen insensitivity (*AR*)** • **Monogenic X chromosome and autosome gene mutations (*TEX11, 14, 15; STAG3; SYCE1; DMRT1; MEIOB; FANCA; FAMCM; SYCP3; MEI1; MAGEA9; and many others)*** • **Cryptorchidism/testicular dysgenesis syndrome (probably genetic or epigenetic)**	• Orchitis • Drugs, toxins, irradiation • Postorchiectomy, complicated inguinal surgery (iatrogenic), trauma, testicular torsion, varicocele (grade 3)
Sperm Motility, Morphology, Function, or Genetic Incompetence	
• **Sperm flagella: primary cilia dyskinesia, multiple morphological abnormality of sperm flagella** • **Sperm head: globozoospermia, macrocephalic sperm** • **Advanced paternal age** • **Autoimmunity (probably genetic)**	• Oxidative stress • DNA damage
Ductal Obstruction or Dysfunction	
• **Congenital bilateral absence of vas deferens (mutation in *CFTR, ADGRG2*)** • **Young syndrome (probably genetic)**	• Epididymal obstruction • Vasectomy/vas occlusion • Seminal vesicle and prostate disorders • Ejaculatory duct dysfunction • Posthernia repair, pelvic surgery, retroperitoneal lymph node dissection • Severe sexual erectile or ejaculatory dysfunction (spinal cord injury, multiple sclerosis, neurotube defects)
Idiopathic	
• **Idiopathic azoospermia or oligozoospermia, athenozoospermia, and teratozoospermia (Likely genetic)**	

CFTR, Cystic fibrosis transmembrane conductance regulator; *FSH,* follicle-stimulating hormone; *LH,* luteinizing hormone.
Adapted from Çayan S, Shavakhabov S, Kadioglu A. Treatment of palpable varicocele review in infertile men: A meta-analysis to define the best technique. J Androl. 2009;30(1):33-40. Accessed 29 September 2015. https://doi.org/10.2164/jandrol.108.005967.

infundibulitis may spontaneously involute or require pharmacological treatment of the inflammation and may cause hypogonadotropic hypogonadism.[25] Recent studies showed that cancer immunotherapy with checkpoint inhibitors may induce hypophysitis resulting in adrenal, thyroid, and gonadal insufficiency. Recovery from central hypogonadism from checkpoint inhibitors is common when compared to adrenal or thyroid insufficiency.[26]

Vascular Lesions. Vascular lesions include pituitary infarction and carotid aneurysm. Sheehan syndrome is an example of the former.

Traumatic Brain Injury. Anterior pituitary dysfunction occurs in the early posttraumatic brain injury period and may persist after the acute period. The frequency of pituitary dysfunction depends on the severity and type of trauma. Though hypogonadism may be transient, surveillance is required to ensure the patient does not develop persistent hypogonadotropic hypogonadism.[27]

Endocrinopathies. Suppression of the hypothalamus-pituitary axis may be caused by other endocrinopathies. While any cause of hyperprolactinemia (drugs, hypothyroidism) may be incriminated, prolactinoma is the most common cause of hypogonadotropic hypogonadism in men. Chronic supraphysiological glucocorticoid secretion and other causes of Cushing syndrome in men may result in low serum testosterone concentrations and inappropriately normal serum gonadotropins, due to suppression of hypothalamic GnRH secretion as well as suppressed testicular androgen synthesis.

Androgens and Other Sex Steroids. Administration of exogenous testosterone, other anabolic steroids, as well as estrogens and progestins suppress both LH and FSH secretion and serum concentrations and result in oligo- or azoospermia. Anabolic steroid use should be suspected in men with low sperm concentration, low serum LH concentrations,

TABLE 112.2 **Most Frequently Reported Genes Associated With Congenital Hypogonadotropic Hypogonadism**

Gene	Gene Product/Function	Clinical Features
Genes Associated With GnRH Neurons Development or Migration (Kallmann Syndrome)		
***ANOS1* (formerly *KAL1*)**	Anosmin 1 (X-linked)	Anosmia or hyposmia, unilateral renal agenesis, midline facial defects, synkinesia (cerebellar abnormality)
***FGFR1* (formerly *KAL2*)**	Fibroblast growth factor receptor 1 (autosomal dominant, digenic)	Anosmia to normosmia, cleft palate/lip, synkinesia, syndactyly and other skeletal abnormalities, dental agenesis
FGF8	Fibroblast growth factor 8 (autosomal dominant or recessive)	Anosmia to normosmia, cleft palate/lip
PROK2	Prokineticin 2	Anosmia to normosmia
PROKR2	Prokineticin receptor 2 (autosomal dominant or recessive)	
CHD7	Chromodomain helicase DNA-binding protein 7 (autosomal dominant)	Mild phenotype of CHARGE syndrome (eye coloboma, heart defects, choanal atresia, retarded growth, genital hypoplasia, ear anomalies, and deafness), anosmia to normosmia
Others: *NELF* (nasal embryonic LHRH factor), *WDR11* (WD repeat containing protein), and many others		
Genes Associated With GnRH Secretion and Action		
KISS	Kisspeptin	Central precocious puberty, no infertility
KISSR	KISS receptor, autosomal recessive	Normosmia
LEP	Autosomal recessive	Severe obesity
LEPR	Leptin	Severe obesity
	Leptin receptor	
TAC3	Neurokinin B	Cryptorchidism, microphallus
TAC3R	TAC3 receptor (autosomal dominant of recessive)	
GnRH1	GnRH (autosomal recessive)	No smell or midline defect
GnRHR	GnRH receptor (autosomal recessive or digenic) inheritance	
Others: *PCSK1* (protein convertase subtilisin/kexin-type 1); *NROB1* (dosage sensitive sex reversal adrenal hypoplasia congenital *DAX1*), and many others		

Adapted from Krausz C, Rosta V, Swerdloff RS, Wang C. Genetics of male infertility. In: Pyertitz R, Korf B, Grody W, eds. *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics.* 7th ed. Elsevier; 2021:in press.[10]

but with a well-androgenized phenotype.[28,29] A recent observational study of current and past androgen abuser showed that recovery of LH occurs faster than FSH, spermatogenesis, and sperm output and the slower recovery is related longer duration of drug abuse.[30,31] Healthy men who were administered androgen together with progestins in male contraceptive studies also demonstrate marked suppression of sperm output, which is reversible after discontinuation of treatment; the time to recovery is influenced by the type and duration of androgen administration[32] (see Chapter 115). In a review of 1540 men evaluated for infertility, 110 (7.1%) were on supplemental testosterone. Median sperm concentration 4.5 months after stopping testosterone increased from 0 to 6.3 million/mL, but some men remained azoospermic for more protracted periods indicating a longer follow-up period may be required.[33]

Other Drugs. Many drugs can affect male fertility including those acting on the hypothalamic-pituitary-gonadal axis (see Chapter 113; Table 113.2). Thus assessment of a man with infertility must include detailed drug history.[34] Chronic use of pharmacologic doses of glucocorticoids may lead to decreased intratesticular testosterone concentration, inhibition of spermatogenesis, and male infertility. Opioid-like or other central nervous system–activating drugs, including many psychotropic drugs, can inhibit GnRH or gonadotropin secretion, resulting in secondary hypogonadism and infertility.[35] Administration of GnRH analogs (agonists and antagonists) to suppress gonadotropin secretion in men with prostatic carcinoma results in hypogonadism and infertility.[36]

Systemic Illness. Any serious systemic illness or chronic nutritional deficiency can cause hypogonadotropic hypogonadism, primary hypogonadism, and infertility.

KEY POINTS

- Hypothalamic-pituitary endocrine disorders (with low serum testosterone) represent a small percentage of overall cases of male infertility.
- Congenital hypogonadotropic hypogonadism can be due to defects in the development and migration of the GnRH neurons (Kallmann syndrome associated with anosmia and other developmental defects) or in impaired GnRH secretion or action (GnRH receptor defects) not associated with olfactory anomalies. Congenital hypogonadotropic hypogonadism undergoes spontaneous remission in 10% to 20% of patients (both adolescents and adults).
- Other genetic defects such as gonadotropins and gonadotropin receptors are very rare.
- More commonly acquired hypothalamic-pituitary disorders are sellar and suprasellar lesions: pituitary adenomas, craniopharyngiomas, meningiomas, lymphocytic hypophysitis, Rathke's cleft cysts, germinomas, and drugs (opioids and androgenic steroids) and severe systemic illness.
- Since specific therapies are available, it is important to diagnose these disorders.

Testicular Disorders With Spermatogenesis Defects

About 75% of infertile men have primary testicular impairment; including 20% due to known genetic causes and 30% to 40% acquired disorders and 40% to 50% "idiopathic."[8,10] This group of disorders affect spermatogenesis leading to decreased or absent production of spermatozoa and oligozoo- or azoospermia.[9] Current recommendations suggest that men with severe oligozoospermia (sperm concentration <5 million/mL ejaculate) and nonobstructive azoospermia should have genetic screening studies performed.[37] In selected cases based on the phenotype, monogenic defects can be screened.[10,38]

Congenital Spermatogenesis Defects

Chromosomal Defects. About 13.7% of azoospermic and 4.6% of oligozoospermic males have sex chromosomal disorders. Azoospermia or severe oligozoospermia (<5 million spermatozoa/mL ejaculate) is associated with sex chromosome abnormality and Y chromosome microdeletions,[39] whereas less severe oligozoospermia (>5 million spermatozoa/mL ejaculate) with autosome anomalies such as Robertsonian and reciprocal translocations or inversions.[40]

Klinefelter Syndrome, XYY and XX Males. The most common chromosomal disorder associated with male infertility is Klinefelter syndrome (KS) occurring in 1:600 males.[41] The karyotype is usually 47, XXY but mosaicism (46,XY/47,XXY) does occur. Addition of another X chromosome to XXY causes a more severely affected phenotype. Infertility, bilateral small, firm testes, and azoospermia occurred in 99% of men with 47,XXY karyotype. They are tall and usually hypogonadal with decreased body hair, gynecomastia, eunuchoid proportions, and decreased muscle mass and strength and abdominal obesity (in 60% to 80% of KS men).[41,42] In addition, they have higher incidence of learning disability, delayed speech development in childhood and metabolic syndrome, type 2 diabetes mellitus, osteopenia, and autoimmune diseases in adulthood.[43] Men with untreated KS usually have elevated FSH, LH, and low or borderline low testosterone concentrations. The testis in adult KS men showed hyalinization of the seminiferous tubules and hyperplasia of the Leydig cells. However, about 50% of patients with KS have pocket of spermatozoa in the seminiferous tubules that is even higher (70%) in adolescents 12 to 20 years old (see Chapter 113). The current treatment for these patients with azoospermia who desire fertility is to harvest spermatozoa using microsurgical testicular sperm extraction (TESE) followed by intracytoplasmic sperm injection (ICSI) (see Chapter 113). Though not currently the standard of care, early fertility preservation with sperm and testicular tissue cryopreservation in adolescent boys with KS are under investigation.[44]

Other sex chromosome anomalies include XYY males and XX males. XYY men have normal-to-severe spermatogenesis defect. In most of the XX males, there is a translocation of the *SRY* gene (sex-determining gene, located right below the pseudoautosomal region of the Y chromosome) to the X chromosome. Though the genotype is XX, due to the presence of the SRY gene, the phenotype will result in a male but with spermatogenic failure. XX males are not tall, does not have eunuchoid proportions, and have no spermatozoa in the testis because of the absence of spermatogenesis genes mapping to the long arm of the Y chromosome.[10]

Structural Chromosomal Aberrations. Robertsonian translocation, reciprocal translocations, and inversions are most frequently found in oligozoospermic men. Reciprocal translocations occur with breaks in each of two nonhomologous chromosomes that exchange the broken segments. Robertsonian translocation involving chromosomes 13, 14, 15, 21, and 22 occurs in 1 in 1000 live births with the fusion of two acrocentric chromosomes, where the long arms are fused and the short arms are lost resulting in 45 chromosomes. The balanced form usually has no health problems. In unbalanced translocations, monosomy or trisomy are associated with phenotypes including trisomy 21.[45] In a large cohort, 95% of adults with balanced Robertsonian translocations have infertility, miscarriage, or offspring with chromosomal abnormalities.[46] In another study, balanced reciprocal translocations are associated also with infertility.[47] It is important to detect these structural chromosomal anomalies because of increased risk of aneuploidy and unbalanced translocations in the offspring that may lead to Angelman or Prader-Willi syndromes. New molecular and sequencing techniques can now correctly diagnose chromosomal rearrangements in the parents and in the embryo.[48,49]

Y Chromosome Microdeletions. The Y chromosome is an acrocentric chromosome with pseudoautosomal regions that are homologous with the X chromosome sequences and the male-specific Y region that is critical for spermatogenesis. This chromosome has many segmental duplications, direct or inverted repeats, including palindromes allowing for recurrent deletions/duplications controlling the dosage of different Y genes. The AZF (azoospermia factor) region has recurrent microdeletions in three subregions in the proximal, middle, and distal Yq11 named AZFa, AZFb, and AZFc. AZFa region contains *USP9Y* and *DDX3Y genes* and deletions of AZFa result in Sertoli cell–only syndrome. Males with AZFa deletions are infertile without germ cells in the testis. Complete AZFb deletions result in spermatogenic arrest, absence of mature germ cells, and male infertility. In contrast, deletions of the AZFc region may present in men with oligozoospermia to azoospermia. Azoospermia may be due to progressive loss of germinal epithelium with time or the consequence of 46,XY/45,XO mosaicism or environmental influences. Y chromosome deletions occur in about 5% to 7.5% of infertile men and the highest incidence is found in azoospermic men. AZF microdeletion testing is standardized and generally requested for men with severe oligozoospermia <5 million/mL, not only for diagnosis but in case of azoospermia it is also relevant for predicting whether TESE is a viable option.[50,51] For men who will undergo TESE and ICSI, genetic counseling is necessary as the son will carry the Y microdeletion and will be affected by impaired sperm production.[10]

Partial deletion and duplication of the AZFc region, gr/gr deletion, is a significant risk factor associated with male subfertility in some populations such as East Asian and White American men but not in Black and some European men.[51–54] The risk of spermatogenesis defects is increased by 2- to 2.5-fold in men with gr/gr deletions and is highest in men of Mediterranean origin. Testing for this deletion in infertile men should be offered for the couple because this deletion is transmitted to the male progeny and is associated with a risk of developing testicular germ cell tumors (TGCTs). In multicenter studies, gr/gr deletions increased the risk of TGCTs especially in normospermic men by 3.7-fold.[55]

Copy Number Variations. These chromosome variations involve gains and losses of homologous sequences in chromosomal segments. Men with spermatogenesis defects have higher load of copy number variations.[56–58] If this occurs in the single X chromosome in the male that may lead directly to a deleterious effect in the absence of a compensatory normal allele. X-linked deletion, CNV67, leads to deletion of the proximal copy of the testis-specific *MAGEA9* (*melanoma antigen family A,9B*) gene specifically expressed in spermatogonia.[59] The associated semen phenotype ranges from azoospermia to oligozoospermia. The intragenic deletion of the X-linked *TEX11* (*testis-expressed 11*), a gene required for male germ cell meiosis, leads to azoospermia due to meiotic arrest.[60]

Mild Partial Androgen Receptor Insensitivity. Androgens are essential for postmeiotic development of spermatocyte, spermatids to spermatozoa, and for the release of spermatozoa from the germinal epithelium to the lumen. Mutations of the androgen receptor gene are uncommon and result in variable phenotype from a phenotypic female in complete androgen insensitivity syndrome, undervirilized male with ambiguous genitalia in partial androgen insensitivity to a normal male with infertility, and defective spermatogenesis (see Chapter 105).[61] Patients with complete or partial androgen insensitivity syndrome are infertile. Mutations in the transactivation or the ligand-binding domain of the androgen receptor may result in defective spermatogenesis.[62]

The transactivation domain of the androgen receptor contains two segments of glutamine (CAG) and glycine (GCN) repeats. Expansion of the CAG repeats (>39 CAG) in exon 1 of the transactivation domain of the androgen receptor leads to Kennedy syndrome, spinal bulbar atrophy, that is associated with progressive hypogonadism and testicular atrophy. The relationship between CAG or GCN repeats with male infertility is controversial. Meta-analyses show that such association may be present in Asians but not in White Europeans and thus expansion of CAG repeats (in the polymorphic range, i.e., <39 CAG) may only play a minor role in male infertility in some populations but not others.[62–64]

Monogenic and Autosomal Mutation. The advent of next-generation sequencing and whole-exome sequencing allows the identification of many novel spermatogenesis-related gene defects in men with nonobstructive azoospermia or oligozoospermia (Table 112.1).[65–68] The large majority of gene defects are recessive and often are found in consanguineous families. The most frequently reported monogenic defect of meiotic arrest is due to *TEX11* mutations. The clinical importance of diagnosing these genetic factors is not only for identifying the monogenic defects but also because some of these defects are consistently associated with Sertoli cell–only syndrome or maturation arrest with prognostic and treatment implications. A future nonobstructive azoospermia gene panel based on genes with strong/definitive clinical evidence would avoid unnecessary surgical sperm extraction.[69]

Recessive mutations in the Fanconi's anemia gene pathway (*FANCA*) have been identified in infertile men with Sertoli cell–only syndrome and mild hematological abnormalities. This finding implies that the two different compartments, germinal epithelium and bone marrow, may share common genetic defects involved in DNA damage repair. Late-onset Fanconi's anemia can be diagnosed during genetic investigations for male infertility.[70] Mutations in other genes involved in DNA damage repair (FANCM) have also been reported in infertile men implying that carriers of these genetic defects may be predisposed to cancer.[67]

Cryptorchidism. Cryptorchidism affects 2% to 9% of full-term newborn boys. Testicular descent is regulated by testosterone and insulin-like 3 (INSL3), both are products of Leydig cells. Cryptorchidism associated with hypospadias and testis cancer may originate *in utero* and are components of the "testicular dysgenesis syndrome." *In utero* environmental exposure to antiandrogens such as phthalates and estrogens like dioxin have been suggested to be the cause of the increasing prevalence of cryptorchidism (and other reproductive defects) in some but not all regions (discussed later in the chapter).[71] Men with a history of undescended testes have lower sperm counts, poor sperm quality, and lower fertility rates than men with normally descended testes. Impaired spermatogenesis in the undescended testis is probably related to underlying genetic, hormonal, and developmental abnormalities, some of which may be partially reversible through early surgical intervention. The degree of germ cell dysfunction with cryptorchidism is correlated with the duration of suprascrotal location of the testes. Early surgical correction of cryptorchidism may improve fertility but not the risk of testicular cancer.[72] TGCTs are associated with cryptorchidism and impaired spermatogenesis.[73] TGCT occurs mainly in adolescent and young adults due to impaired gonadal development, arrest of prespermatogonia differentiation, and retention of embryonic pluripotent features associated with genomic instability. Cryptorchidism in Klinefelter and Kallmann syndrome is related to decreased production of testosterone and INSL3 by the Leydig cells. Though INSL3 and its receptor are important in the descent of the testis, no mutations of these genes have been reported. Apart from the above-mentioned syndromic forms, in the majority of cases the pathophysiology of cryptorchidism is unknown. Genome-wide association studies identified gene polymorphisms associated with cryptorchidism, yet no single-gene defect has been identified.[74]

KEY POINTS

- Genetic studies are recommended for infertile men with nonobstructive azoospermia and severe oligozoospermia (<5 million spermatozoa/mL ejaculate).
- Klinefelter syndrome is the commonest chromosomal abnormality in males and is associated with azoospermia and hypogonadism in the majority of patients.
- Robertsonian and reciprocal translocations are more frequent in infertile men with oligozoospermia and the diagnosis is important as these rearrangements may become unbalanced in the offspring.
- Y chromosome microdeletions are the most common molecular genetic causes of spermatogenic impairment. Genetic counseling is important as this genetic defect is transmitted to the male progeny.
- Monogenic spermatogenesis gene defects are increasingly being identified using next-generation sequencing and are mostly recessive.
- Genetic defects can be transmitted through assisted reproductive techniques to the future offspring. Preintervention/conception counseling and genetic testing should be considered.
- Cryptorchidism is the commonest male reproductive developmental defect, but the exact pathophysiology is yet to be discovered, probably involving genetic susceptibility, hormonal, and other factors during development.

Syndromic Monogenic Defects. The genetic defects associated with these syndromes may disrupt single-gene function, may lead to the inactivation of imprinted genes, or are related to expansion of trinucleotide repeats.

Bardet-Biedl Syndrome. Patients with Bardet-Biedl syndrome present with obesity, polydactyly, mental retardation, renal anomalies, hypogonadism with cryptorchidism, hypospadias, and micropenis. This syndrome is associated with mutations of *BBS*.

Prader-Willi Syndrome. This syndrome is associated with hypotonia, developmental delay, mental retardation, and obesity. Prader-Willi syndrome is associated with the absence of genes in the imprinted region of paternal chromosome 15q11-q13 in 70% of cases; the remainder is associated with maternal gene disomy or defective imprinting.

Noonan Syndrome. Patients with Noonan syndrome have short stature, heart failure due to cardiomyopathy or pulmonary valve lesions, dysmorphic facial features, and cryptorchidism in boys. The most common gene mutation is in *PTPN11* that encodes the nonreceptor protein phosphate SHP-2.

Myotonic Dystrophy. Myotonic dystrophy is caused by an expansion of the trinucleotide CTG repeats at chromosome 10p13.3. Genes with 50 to 80 repeats are unstable and may expand more in transmission to the offspring. The clinical severity of the disease is related to the number of repeats. Muscle dystrophy is associated with cardiac conductance problems, cataracts, progressive testicular atrophy, infertility, and hypogonadism.

Acquired Spermatogenesis Defects. This section will discuss infections, drugs, toxins, irradiation, and trauma to the testis and presence of varicocele on male infertility.

Infection. Viral orchitis, especially mumps, is a well-recognized cause of infertility. Among those with mumps, unilateral clinical orchitis is rare in prepubertal males but occurs in about 10% to 20% of adult men. The virus can directly affect Leydig and germ cells or through an indirect immune mechanism. Orchitis is frequently accompanied by epididymitis. Prior vaccination against mumps virus decreases the severity of the disease. Outbreaks of mumps orchitis occur that may be due to incomplete vaccination or waning immunity.[75] The affected testis may become atrophic in about half of the cases and is associated with oligozoospermia and infertility. In mumps and other viral causes of orchitis (echovirus and arbovirus), germ cell failure is much more common than androgen deficiency.[76]

Other infectious causes of orchitis and infertility include tuberculosis and leprosy; the former may also cause epididymal obstruction. Sexually transmitted infections (STIs) such as gonorrhea, chlamydia, ureaplasma, herpesviruses (HSV), human papillomaviruses (HPV), hepatitis B and C viruses, human cytomegalic virus (HCMV), and human immunodeficiency virus (HIV) are detected in symptomatic as well as asymptomatic men with testis, accessory gland, and urethral infections. These bacteria, viruses, and protozoa are associated with decreased semen quality, including lower concentration and motility, and may cause male infertility. White blood cells may be present in the semen, especially if HIV was associated with other STIs such as gonorrhea. Very sensitive molecular methods are now available to allow investigation of the relationship between STD infections and male infertility. Chronic inadequate STI treatment may result in more damage and infertility than acute infections.[77–80] All the pathogens present in the semen can be transmitted to the partners and newborns, and screening for HIV, HPV, HSV, and HCMV in male patients is advised to prevent viral transmission and adverse outcome of fertility.[80,81]

Drugs and Radiation. Chemotherapy: Many drugs are associated with impaired spermatogenesis and/or Leydig cell dysfunction. Among them, the most important are alkylating drugs used in chemotherapy of hematological malignancies in children and adults. Childhood cancer has a survival rate of over 85%. The Childhood Cancer Survivor study showed that males have reduced likelihood of siring a pregnancy. This reduction in fertility is related to the dose of alkylating agents (cyclophosphamide, ifosfamide) or DNA interstrand crosslinking drugs (procarbazine and cisplatin).[82] Fertility preservation is recommended for boys and young adults undergoing chemotherapy.[83–85] Semen cryopreservation should be performed before chemotherapy for those that can produce a sample. For those where semen samples cannot be obtained, testicular biopsy and cryopreservation of testicular tissue for future *in vitro/in vivo* germ cell development are available in some centers.[86,87]

Ionizing radiation: Irradiation of the scrotal area impairs spermatogenesis but not Leydig cell function until very high dose >30 Gy is administered. Doses as low as 0.5 Gy may transiently suppress spermatogenesis, whereas doses above 3 Gy (600 rad) usually cause prolonged azoospermia and infertility. Doses over 16 to 18 Gy result in permanent azoospermia.[88]

Antiandrogens: Flutamide, cyproterone, bicalutamide, spironolactone, ketoconazole, and cimetidine impair spermatogenesis by disrupting the hypothalamic-pituitary-testis (HPT) axis altering androgen action (see Chapter 113, Table 113.2).

Testicular Torsion and Trauma. Testicular torsion and injury lead to ischemia and rapid loss of germ cells. In addition, ischemia followed by reperfusion triggers the generation of reactive oxygen species, proinflammatory cytokines, anoxia, and apoptosis with a high risk of subsequent infertility.[89] Emergent surgical intervention in emergency with unilateral orchiopexy and orchiectomy may reduce future subfertility.[90]

Varicocele. Varicoceles are dilatations of the pampiniform plexus of the spermatic veins in the scrotum. Left-sided varicoceles are 10 times more common than right-sided ones, perhaps because of the almost 90-degree angle with which the left spermatic vein enters into the renal vein causing a lower blood efflux. Varicoceles are found in about 10% to 15% of normal men while this percentage is higher in infertile men. In a large study of over 9000 infertile couples, the presence of a varicocele was accompanied by decreased testis volume and impaired semen parameters that may contribute to spermatogenesis defects.[91] The mechanisms by which varicocele causes infertility and the reversibility of infertility after varicocele surgery are controversial. Increased oxidative stress may be one of the mechanisms. Current evidence suggests that varicocelectomy may have a beneficial effect on semen quality and pregnancy outcomes only if the male has a clinically palpable varicocele (grade 3) and abnormal semen parameters.[92,93]

Sperm Motility, Morphology, Function, or Genetic Incompetence—Spermatogenesis Qualitative Defects

Congenital Qualitative Sperm Defects

Sperm Flagellum Defects. Primary cilia dyskinesia (immotile cilia syndrome) is characterized by abnormal ciliary motion and mucociliary clearance in the bronchus, sinuses, and nose. Males are infertile because of poor sperm motility due to defective function of the flagella. This defect is associated with chronic respiratory disease and situs inversus in about half of the patients (Kartagener syndrome). There are over 40 genes related to ciliary dyskinesia. Recessive mutations of the genes *DNAH5* and *DNAH11* result in the absence of the dynein arms in the cilia and the flagellum of spermatozoa resulting in impaired movement and function and male infertility.[94]

Multiple Morphological Abnormalities of the Sperm Flagella. Patients with multiple morphological abnormalities of the sperm flagella (MMAF) have multiple abnormalities of the sperm flagella (absent, coiled, bent, angulated, irregular). MMAF is one of the most severe forms of sperm flagellum defects that are not associated with recurrent respiratory infections. Because in the human sperm flagella several hundreds of proteins are involved, many gene defects have been identified in MMAF. These include *DNAH1, CFAP43, CFAP44, CFAP69, SPEF2*, and many others that affect axonemal or periaxonemal protein complexes. However, patients with *DNAH1* have good outcomes following ICSI.[95–97]

Sperm Head Defects. Sperm head defects that result in severe teratozoospermia have genetic bases. An example is the identification of new genes using whole-exome sequencing for human globozoospermia (sperm with round heads and acrosome defects). Mutations of *DPY19L2* are responsible for 70% of all cases of globozoospermia; using exome sequencing six other new genes have been identified.[98] Another sperm head morphological abnormality is macrozoospermia (macrocephalic sperm heads) that is associated with mutations of *AURKC* in up to 80% of patients. These sperm head defects are usually not compatible with normal embryo development due to tetraploidy of the spermatozoa.[99]

Advanced Paternal Age. As men age, sperm concentration does not usually decrease but semen volume, sperm motility, and spermatozoa with normal morphology decrease[100] and sperm DNA fragmentation[101] increases. DNA damage in the spermatozoa with advanced paternal age may affect the pregnancy outcome, whereas epigenetic changes are predicted to affect health of the offspring. Some dominant diseases, such as Apert, Noonan, Costello, Crouzon, Pfeiffer syndromes, achondroplasia, and multiple endocrine neoplasia (types 2A and 2B) present a higher frequency in children born to older fathers. These genetic diseases are considered as paternal age effect diseases.[102,103] In addition, mental disorders such as autism, schizophrenia, bipolar disorder, low IQ, as well as attention deficit hyperactivity disorder occur more frequently in offspring from fathers of advanced age.[104]

Sperm Autoimmunity. The presence of sperm agglutination in the semen should trigger the laboratory to test for antisperm antibodies. Whether antibodies occur spontaneously or only after testicular injury or infection is not known. About 2% to 4% of infertile men have antisperm antibodies in serum or semen and both presumably could impair spermatogenesis.[105] Men with antisperm antibodies have lower sperm concentration (but not oligozoospermia) and sperm progressive motility.[106] A multicenter case-control study found that infertile men with normozoospermia and positive sperm antibodies have more acrosome reaction and hyperactivation disorders, oxidative stress, and DNA fragmentation than fertile men suggesting that sperm autoimmunity may impair sperm function.[107] However, the exact role of sperm antisperm antibodies in male infertility remains controversial. Experimental models have demonstrated that infection and inflammation of the testis may result in autoimmune orchitis. Patients with autoimmune polyendocrinopathy and those with systemic autoimmune disease may have autoimmune testicular disease but there are many other unknown factors.[108]

Acquired Sperm Qualitative Defects. Infection and inflammation of the genital system may increase the production of reactive oxidative species by leucocytes impairing the functional competence of spermatozoa. Oxidative stress results in DNA damage that may affect *in vitro* fertilization success rate but the effects on the health of the offspring are not well established.[109]

KEY POINTS: Sperm Qualitative Defects (sperm motility, morphology, function or genetic incompetence)

- Genetic analysis of large consanguineous families and the recent use of whole exome sequencing allowed the identification of genetic mutations causing flagellum defects in ciliary dyskinesia and multiple morphological abnormality of sperm flagella and gross sperm head morphological defects including globozoospermia and macrozoospermia.
- Advanced paternal age affects sperm motility and DNA integrity. Moreover, spermatozoa of aged men carry dominant mutations more frequently causing the so-called Paternal Age Effect diseases such as Apert syndrome, achondroplasia and multiple endocrine neoplasia (Type 2 A and B), as well as mental disorders.
- Sperm autoantibodies are detected in about 2 to 4% of infertile men and their effect on male fertility is not clear.
- Acquired sperm function defects with oxidative stress and DNA damage may be associated with inflammation and infection of the genital tract.

Ductal Obstruction or Dysfunction

Infertile men with ductal obstruction have normal reproductive and endocrine function. Physical examination with imaging is useful in the diagnosis of the site of the obstruction in the urogenital tract.

KEY POINTS: Ductal Obstruction

- Congenital Bilateral Absence of Vas Deferens is the commonest genetic cause of ductal obstruction mainly due to mutations of the *CFTR* gene. Small proportion of these patients may have mutations of the X-linked *ADGRG2* gene or associated with unilateral renal agenesis where the underlying mechanisms have yet to be identified.
- Obstructive azoospermia with chronic sinusitis-pulmonary infections called Young's syndrome is a rare disease and the genetic cause has not been defined.
- Epididymal obstruction usually due to prior infection is the most common cause of acquired proximal obstruction followed by vasectomy or vas occlusion procedures. Midline prostatic cyst and ejaculatory duct obstruction are correctable by minimally invasive aspiration of the cyst or surgery.
- Severe erectile and ejaculatory dysfunction related to neurological diseases, spinal cord injury or systemic disorders such as diabetes mellitus must be recognized and treated.
- Men with congenital bilateral obstruction of the vas deferens must be counseled and tested for the CFTR gene before epididymis sperm extraction and ICSI.

:Genetic Causes of Ductal Obstruction

Congenital Bilateral Obstruction of the Vas Deferens. Congenital bilateral absence of the vas deferens (CBAVD) is the cause of about 1% to 2% of male infertility and up to 6% of obstructive azoospermia.[8] Men affected by cystic fibrosis have CBAVD. Mutations of the gene encoding cystic fibrosis transmembrane conductance regulator (*CFTR*) can lead to CBAVD as a monosymptomatic, mild form (male infertility) of cystic fibrosis. Some *CFTR* variants are more frequently found in CBAVD. CBAVD without other symptoms of cystic fibrosis are due to the combination of either two mild mutations or one mild mutation and one severe mutation. The polymorphic poly-T tract in intron 8 may influence the severity of the phenotype and IVS8-5T is considered a mild mutation. CBAVD men can sire a progeny by either epididymal or TESE followed by ICSI. Genetic counseling and testing are essential to prevent the couple having a child with cystic fibrosis. The female partner of a patient affected by CF or CBAVD must be screened for *CFTR* mutations since in some populations the carrier frequency of *CFTR* mutations can be as high as 1:25. Depending on the genotype of each partner, the option for preimplantation genetic diagnosis should be discussed with the couple before assisted reproductive techniques are used.[110] About 20% of CBAVD men who are negative for *CFTR* mutations are hemizygous for a mutation of the X-linked *ADGRG2* gene. Some men with this obstructive disorder present with unilateral kidney agenesis suggesting early organogenesis disorder that is independent from CFTR.[111] The underlying mechanisms of this pathological condition are still unknown.

Young Syndrome. Men with Young syndrome present with obstructive azoospermia and chronic sinopulmonary infections.[112] The obstruction in the bronchi and epididymides is due to inspissated secretions. This disease is very rare and has been related to mercury exposure in childhood.[113] Presumably genetic in origin, but no specific genes have been identified.

Acquired Ductal Obstruction

Abnormalities of the Epididymis. Absence, dysfunction, infection,[108] or obstruction of the epididymis leads to infertility even though testicular sperm production may be normal. Intrauterine exposure to estrogens may cause epididymal dysfunction. While poorly documented, it is presumed that some men with isolated asthenozoospermia (impaired motility) have defects of epididymal function.[114]

Abnormalities of the Vas Deferens. Bilateral obstruction, ligation, or altered peristalsis of the vas deferens results in infertility. Obstruction may result from infection (gonorrhea, chlamydia, tuberculosis), whereas ligation of the vas deferens (vasectomy) is an intentional, medically induced cause of infertility (male contraception). Though the patency of the vas can be restored by vasovasostomy of vasoepididymostomy, the pregnancy rate in the female partner is much lower.[115]

Seminal Vesicles and Prostate. Congenital absence of the seminal vesicles is associated with CBAVD. The prostate gland secretes zinc, citrate, and kallikreins that may be important for sperm function. Inflammation of the prostate or seminal vesicles may be associated with abnormal sperm parameters and function.[116] Midline prostatic cysts were present in 10.2% male partners of infertile couples compared to 5.8% fertile men. Men with larger midline prostatic cyst had low sperm concentration. Transrectal ultrasound-guided cyst aspiration may improve sperm concentration.[117]

Ejaculatory Duct Obstruction. Patients with ejaculatory duct obstruction present with a low ejaculate volume and no sperm and/or very low sperm motility. Ejaculatory duct obstruction is found in up to 5% of infertile men and can be diagnosed and treated surgically with minimally invasive techniques (transurethral resection of the ejaculatory ducts).[118]

Postsurgery. Prior inguinal hernia repair may inadvertently injure leading to scarring of the vas deferens affecting sperm transport. Retroperitoneal lymph node dissection and posterior pelvic surgery may damage the sympathetic supply to the vasa, seminal vesicles, and bladder neck resulting in interruption of sperm transport.

Severe Sexual and Ejaculatory Dysfunction. Erectile dysfunction, mechanical obstruction (condoms and diaphragm use), premature ejaculation, and ejaculatory failure may cause male infertility in about 4% of infertile men.[8] Neurological disease including multiple sclerosis, transverse myelitis, and defects of the neural tube interferes with normal ejaculation. Long-standing diabetes mellitus with autonomic neuropathy may result in retrograde ejaculation and severe erectile and ejaculatory dysfunction.[8,9] These causes should be determined by careful history, examination, and diagnostic workup.

KEY POINTS

- Congenital bilateral absence of vas deferens is the commonest genetic cause of ductal obstruction mainly due to mutations of the *CFTR* gene. Small proportion of these patients may have mutations of the X-linked *ADGRG2* gene or associated with unilateral renal agenesis where the underlying mechanisms have yet to be identified.
- Obstructive azoospermia with chronic sinusitis-pulmonary infections called Young syndrome is a rare disease and the genetic cause has not been defined.
- Epididymal obstruction usually due to prior infection is the most common cause of acquired proximal obstruction followed by vasectomy or vas occlusion procedures. Midline prostatic cyst and ejaculatory duct obstruction are correctable by minimally invasive aspiration of the cyst or surgery.
- Severe erectile and ejaculatory dysfunction related to neurological diseases, spinal cord injury, or systemic disorders such as diabetes mellitus must be recognized and treated.
- Men with congenital bilateral obstruction of the vas deferens must be counseled and tested for the CFTR gene before epididymis sperm extraction and ICSI.

Idiopathic Azoospermia, Oligozoospermia, Asthenozoospermia, and Teratozoospermia

Despite the advances in the management of male infertility using assisted reproductive technologies, the cause of male infertility remains not yet identifiable in about 40% of infertile men. As there are over 1500 genes that are involved in human spermatogenesis, only a small proportion have been identified causing quantitative and qualitative spermatogenesis defects. In the most severe forms of infertility, we expect to identify new single-gene mutations affecting spermatogenesis that are sporadic. Environmental factors may cause infertility through interaction with individual genetic background or by inducing epigenetic modifications. The challenge in male infertility is to identify the causes of idiopathic infertility, find specific treatment to correct the problem, and prevent the recurrence of the problems in susceptible individuals.[9,10,119]

OTHER POSSIBLE CONTRIBUTING FACTORS TO MALE INFERTILIY

Obesity

Obesity is a pandemic that is affecting all regions of the world. It is associated with increased risk of cardiovascular disease, type 2 diabetes mellitus, and reproductive disorders including hypogonadism and male infertility. Obesity in men may result in suppression of the HPT axis with low total testosterone, free testosterone, and low or inappropriately normal gonadotropin concentrations. The obesity-associated decrease in serum sex hormone–binding globulin (SHBG) contributes to the low serum total testosterone concentrations.[120] In addition, free testosterone concentrations in the blood appear to be inversely related to body weight and body mass index (BMI), independent of changes in SHBG. Other factors contributing to the hypogonadotropic hypogonadism associated with obesity include an increase in estrogens through aromatization in adipose tissue, increased inflammatory cytokines[121] resulting in insulin resistance, metabolic syndrome, diabetes mellitus, and sleep apnea.

More recent studies suggest paternal obesity may affect sperm epigenomics, which can be transmitted and affect offspring reproductive function by epigenetic reprogramming of spermatogonial stem cells.[122,123] A meta-analysis of 21 studies showed that underweight, overweight, and obese men have higher risk of oligozoospermia and azoospermia.[124,125] The relationship between obesity and semen parameters is still unclear, with some studies reporting lower sperm concentrations/motility and semen volume with increasing BMI[126,127] and others finding no association or only with morbid obesity.[120,128] Both lifestyle changes, medical weight loss and bariatric surgery, demonstrate increase in testosterone but the effect on sperm concentration and infertility needs further controlled studies.[129,130] Moreover, metabolically fit versus unfit obese individual does not differ in semen parameters but evidence suggests that metabolically unfit obese men have higher cardiovascular disease risks.[131] Although these data are conflicting, advice on lifestyle changes and weight loss should be provided to obese men seeking infertility treatment, given the known negative effects of obesity on serum SHBG and testosterone concentrations. Weight loss and bariatric surgery normalize the hypothalamus-pituitary axis and increase testosterone levels and may increase sperm output, but the effect on improving fertility is controversial.[129,132,133] Thus obesity is generally considered a contributing rather than a causal factor for male infertility.[8]

Environmental Factors: Endocrine Disruptors

Testicular Dysgenesis Syndrome and Semen Quality. Although controversial, environmental factors including exposure to endocrine disruptors (antiandrogens or estrogenic compounds) *in utero* have been linked to cryptorchidism, hypospadias, poor semen quality, male infertility, and testicular cancer, the so-called "testicular dysgenesis syndrome"[71,134] (see Chapter 115). These environmental factors including pesticides (organochlorides), fungicides (vinclozolin), byproducts of solvents (dioxin and polychlorinated biphenyls), and plasticizers (phthalates, bisphenol A) affect the fetal testis, but some may also affect the adult testis. Environmental factors most probably act directly on spermatogenesis or via epigenetic modification of spermatogenesis. The transgeneration transmission of these epimutations may result in reproductive and health problems in the progeny.[71,135–138] Environmental exposure may also be the cause of variation of semen parameters across regions where men from agriculture intensive area have higher concentrations of pesticides and lower sperm concentration.[139,140] The causal effects of environmental exposure on male reproductive processes can be demonstrated in animal models. The direct testing of the effects of environmental exposure on male reproductive function in men is not possible. Data derived from epidemiology studies show inconsistent association of exposure to endocrine disruptors with semen quality or reproductive developmental defects in men.[141]

Male Infertility and Testicular Cancer. Testis cancer has been linked to environmental exposure to endocrine disruptors.[71,141] There is an increased incidence of testicular cancer in men presenting with infertility (even in the absence of a history of cryptorchidism).[142] In a case-control study, a prior diagnosis of infertility was associated with an increased risk of testicular cancer (odds ratio [OR] 2.4).[143] In another infertility registry cohort case-control study, men with known male factor infertility had 2.8-fold higher chance of developing a testicular cancer.[144] Testicular cancers are mainly derived from germ cells and occur in young adults. Men with poor semen quality have a much higher risk of testicular cancer (up to 20-fold). Endocrine disruptors with estrogenic or antiandrogenic activity may affect development of both the somatic and germ cells and predispose to infertility and testicular cancer. Cryptorchidism is associated with azoospermia and oligozoospermia and significantly increased the risk of testicular cancer. While no common genetic factors have been identified between impaired sperm production and testis cancer, a recent study based on bioinformatics analyses has suggested the existence of shared genetic factors between oligo-/azoospermia and general cancer predisposition, including testis cancer.[144]

Cigarette Smoking

Data on cigarette smoking and its possible effect on sperm counts are inconsistent. In a meta-analysis of 20 observational studies, men who smoked cigarettes were more likely to have lower sperm counts, motility, and normal morphology.[145] The possibility that *in utero* exposure to smoking may have a detrimental effect on sperm count in adulthood was studied in 1770 young, healthy potential military recruits; the results showed the possibility of a small detrimental effect.[146] In a second related study, there were no significant differences in mean sperm concentrations in men whose mothers smoked or not during pregnancy.[147] Smoking has been shown to change DNA methylation patterns[148] and micro-RNA content in spermatozoa.[149] The significance of this finding, if any, on infertility and progeny is not known.[150] It is advisable that men should stop smoking for fecundity and smoking-related health reasons.[151]

Cell Phone

Because of the rapid increase in cell phone use around the world, studies investigating whether cell phone usage and associated electromagnetic field has a detrimental effect on sperm parameters, are ongoing. This issue is controversial with definitive data not yet available.[152]

Scrotal Hyperthermia

Prolonged high testicular temperature may explain the infertility associated with spinal cord injuries, varicocele, and chronic sauna and Jacuzzi exposure.[153] Studies in men have shown that small increases in testicular temperature accelerate germ cell loss through apoptosis.[154] Similarly, febrile illness, prolonged sitting during work or truck driving, welding, baking, tight-fitting underwear, and laptop use with increased heat to the testes have been proposed to adversely affect male fertility. The data to support these associations are inconsistent and may be a very weak risk factor for infertility.[155]

KEY POINTS

- Though it is debatable whether obese men have higher risk of having abnormal semen analyses and male infertility, for health reasons lifestyle modifications are recommended for infertile couples seeking fertility treatment.
- Environmental exposure to endocrine disruptor *in utero* may result in male reproductive developmental defects such as cryptorchidism, hypospadias, and poor semen quality and testicular cancer. Data are from experimental studies in rodent models and epidemiological association studies in men.
- Men with poor semen quality have higher risk of testis cancer suggesting common factors (environmental and/or genetic) involved in both conditions.
- Cigarette smoking and scrotal hyperthermia should be avoided for men with infertility problems for general health reasons and possible risk for poor semen quality.

EPIGENETICS AND MALE INFERTILITY

With the advent of whole-exome sequencing, many new genetic factors have been identified in the past few years as monogenetic defects causing spermatogenesis dysfunction and male infertility.[156] Much less is known about the contribution of epigenetics to male infertility. Data are sparse and inconsistent, and epigenetic analyses are not used clinically for the clinical diagnosis of male infertility. Nevertheless, epigenetic modifications may alter the expression of genes involved in spermatogenesis theoretically leading to its impairment. Sperm epigenetic modifications such as changes in DNA methylation, histone modification (protamine in sperm), chromatin structure, and noncoding RNA expression may not only affect reproductive function of men but also their progeny.[71,157–159] Modifications of DNA methylation and noncoding RNAs have been implicated in environmental effects on sperm quality (e.g., obesity, smoking)[148,160–162] and in transgenerational effects of pesticides and plasticizers on offspring.[71] Large-scale studies have identified sperm micro-RNAs species and DNA methylation signatures in men with infertility.[163,164,165] Deploying epigenetic biomarkers of male infertility to discover new causes of infertility will require specialized laboratory with technical and bioinformatics capabilities to decipher the epigenetics data and to relate to clinical phenotype and genetics profiling that is not currently used in diagnosis and management of male infertility.[166]

KEY POINTS

- Sperm epigenetics may be associated with male infertility and health of the progeny.
- The importance of epigenetic sperm biomarkers is under investigation but not currently used in clinical practice.

SUMMARY AND FUTURE DIRECTIONS

Male infertility is a common problem that affects 2.5% to 12% of all men globally. Identification of HPT axis abnormalities is important because these conditions may be treatable with medical and surgical approaches. Many testicular causes of infertility do not have specific treatment but recent advances in molecular genetics allow the identification of genes responsible for specific defects of spermatogenesis that may be treatable in the future. The genetics of male infertility are complex but important as they account for about 20% to 30% of the causes of male infertility.[156] The common causes include chromosomal aneuploidy and translocations, Y chromosome microdeletions, monogenetic mutations responsible for quantitative and qualitative spermatogenesis defects, and other causes of endocrine male infertility. Epigenetic modification of gene expression in spermatozoa may provide explanation in some men with idiopathic infertility and may be the conduit for the environmental effects on spermatogenesis.

The challenge for clinicians and investigators in the field of male infertility is the ability to identify the cause of male infertility in each infertile man. Starting with definition of the clinical phenotype, accurate and reproducible sperm quality assessment, hormone measurements, ultrasound imaging, coupled with comprehensive modern genetics, and causative diagnosis are made in about 50% to 60% of infertile men. Coupling the current genetic tools with the impact of epigenetic modifications, proteomics and metabolomics defects may be assessed using bioinformatics toolbox to integrate the information to resolve the etiology of male infertility. However, these assessments at the present required technical expertise concentrated in a few centers that would be restrictive and not available to all patients. Based on each etiology identified, the future is to develop panels of biomarkers that can be used to specifically identify spermatogenesis defects and specific treatment for each infertile man.[167,168] In the meantime, environmental factors such as obesity, cigarette smoking, endocrine disruptors, and others should be eliminated to allow optimization of spermatogenesis and the health of the future progeny.

REFERENCES

1. World Health Organization. www.who.int/teams/sexual-and-reproductive-health-and-research/key-areas-of-work/fertility-care/infertility-definitions-and-terminology#:~:text=Clinical%20definitions; Accessed December 13, 2020.
2. Maung HH. Is infertility a disease and does it matter? *Bioethics*. 2019;33(1):43–53.
3. Vander Borght M, Wyns C. Fertility and infertility: definition and epidemiology. *Clin Biochem*. 2018;62:2–10.
4. Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol*. 2015;13(1):37.
5. World Health Organization. *World Health Organization Laboratory Manual for the Examination and Processing of Human Semen*. 5th ed. Geneva, Swtizerland: World Health Organization; 2010.
6. Cooper TG, Noonan E, Von ES, et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*. 2010;16(3):231–245.
7. Ventimiglia E, Capogrosso P, Boeri L, et al. Infertility as a proxy of general male health: results of a cross-sectional survey. *Fertil Steril*. 2015;104(1):48–55.
8. Punab M, Poolamets O, Paju P, et al. Causes of male infertility: a 9-year prospective monocentre study on 1737 patients with reduced total sperm counts. *Hum Reprod*. 2016;32(1):18–31.
9. Tournaye H, Krausz C, Oates RD. Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabet Endocrinol*. 2017;5(7):554–564.
10. Krausz C, Rosta V, Swerdloff RS, Wang C. Genetics of male infertility. In: Pyertitz R, Korf B, Grody W, eds. *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics*. 7th ed. Elsevier; 2021(in press).
11. Krausz C. Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab*. 2011;25(2):271–285.
12. Stamou MI, Cox KH, Crowley Jr WF. Discovering genes essential to the hypothalamic regulation of human reproduction using a human disease model: adjusting to life in the "-omics" era. *Endocr Rev*. 2015;36(6):603–621.
13. Amato LGL, Montenegro LR, Lerario AM, et al. New genetic findings in a large cohort of congenital hypogonadotropic hypogonadism. *Eur J Endocrinol*. 2019;181(2):103–119.
14. Boehm U, Bouloux PM, Dattani MT, et al. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism: pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*. 2015;11(9):547–564.
15. Cassatella D, Howard SR, Acierno JS, et al. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*. 2018;178(4):377–388.
16. Raivio T, Miettinen PJ. Constitutional delay of puberty versus congenital hypogonadotropic hypogonadism: genetics, management and updates. *Best Pract Res Clin Endocrinol Metab*. 2019;33(3):101316.
17. Dwyer AA, Raivio T, Pitteloud N. Management of endocrine disease: reversible hypogonadotropic hypogonadism. *Eur J Endocrinol*. 2016;174(6):R267–R274.
18. Seminara SB. We all remember our first kiss: kisspeptin and the male gonadal axis. *J Clin Endocrinol Metab*. 2005;90(12):6738–6740.
19. Seminara SB, Messager S, Chatzidaki EE, et al. The GPR54 gene as a regulator of puberty. *N Engl J Med*. 2003;349(17):1614–1627.
20. Hietamäki J, Hero M, Holopainen E, et al. GnRH receptor gene mutations in adolescents and young adults presenting with signs of partial gonadotropin deficiency. *PLoS One*. 2017;12(11):e0188750.
21. Huhtaniemi IT, Themmen AP. Mutations in human gonadotropin and gonadotropin-receptor genes. *Endocrine*. 2005;26(3):207–218.
22. De Sanctis V, Elsedfy H, Soliman AT, et al. Acquired hypogonadotropic hypogonadism (AHH) in thalassaemia major patients: an underdiagnosed condition*? Mediterr J Hematol Infect Dis*. 2016;8(1):e2016001.
23. Gabrielsen JS, Lamb DJ, Lipshultz LI. Iron and a man's reproductive health: the good, the bad, and the ugly. *Curr Urol Rep*. 2018;19(8):60.
24. El Osta R, Grandpre N, Monnin N, et al. Hypogonadotropic hypogonadism in men with hereditary hemochromatosis. *Basic Clin Androl*. 2017;27:13.
25. Gubbi S, Hannah-Shmouni F, Verbalis JG, Koch CA. Hypophysitis: an update on the novel forms, diagnosis and management of disorders of pituitary inflammation. *Best Pract Res Clin Endocrinol Metab*. 2019;33(6):101371.
26. Chang LS, Barroso-Sousa R, Tolaney SM, et al. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. 2019;40(1):17–65.
27. Ntali G, Tsagarakis S. Traumatic brain injury induced neuroendocrine changes: acute hormonal changes of anterior pituitary function. *Pituitary*. 2019;22(3):283–295.
28. Anawalt BD. Diagnosis and management of anabolic androgenic steroid use. *J Clin Endocrinol Metab*. 2019;104(7):2490–2500.
29. Windfeld-Mathiasen J, Dalhoff KP, Andersen JT, et al. Male fertility before and after androgen abuse. *J Clin Endocrinol Metab*. 2020;106(2):442–449.
30. Shankara-Narayana N, Yu C, Savkovic S, et al. Rate and extent of recovery from reproductive and cardiac dysfunction due to androgen abuse in men. *J Clin Endocrinol Metab*. 2020;105(6):dgz324.
31. McBride JA, Coward RM. Recovery of spermatogenesis following testosterone replacement therapy or anabolic-androgenic steroid use. *Asian J Androl*. 2016;18(3):373–380.
32. Liu PY, Swerdloff RS, Christenson PD, et al. Rate, extent, and modifiers of spermatogenic recovery after hormonal male contraception: an integrated analysis. *Lancet*. 2006;367(9520):1412–1420.

33. Kolettis PN, Purcell ML, Parker W, et al. Medical testosterone: an iatrogenic cause of male infertility and a growing problem. *Urology*. 2015;85(5):1068–1073.
34. Semet M, Paci M, Saïas-Magnan J, et al. The impact of drugs on male fertility: a review. *Andrology*. 2017;5(4):640–663.
35. Subirán N, Casis L, Irazusta J. Regulation of male fertility by the opioid system. *Mol Med*. 2011;17(7–8):846–853.
36. Anderson RC, Newton CL, Anderson RA, Millar RP. Gonadotropins and their analogs: current and potential clinical applications. *Endocr Rev*. 2018;39(6):911–937.
37. Schlegel PN, Sigman M, Collura B, et al. Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. *Fertil Steril*. 2021;115(1):54–61.
38. Oud MS, Volozonoka L, Smits RM, et al. A systematic review and standardized clinical validity assessment of male infertility genes. *Hum Reprod*. 2019;34(5):932–941.
39. Vegetti W, Van Assche E, Frias A, et al. Correlation between semen parameters and sperm aneuploidy rates investigated by fluorescence in-situ hybridization in infertile men. *Hum Reprod*. 2000;15(2):351–365.
40. Van Assche E, Bonduelle M, Tournaye H, et al. Cytogenetics of infertile men. *Hum Reprod*. 1996;11(suppl 4):1–24; discussion 25-26.
41. Gravholt CH, Chang S, Wallentin M, et al. Klinefelter syndrome: integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*. 2018;39(4):389–423.
42. Bonomi M, Rochira V, Pasquali D, et al. Klinefelter syndrome (KS): genetics, clinical phenotype and hypogonadism. *J Endocrinol Invest*. 2017;40(2):123–134.
43. Kanakis GA, Nieschlag E. Klinefelter syndrome: more than hypogonadism. *Metabolism*. 2018;86:135–144.
44. Hawksworth DJ, Szafran AA, Jordan PW, et al. Infertility in patients with Klinefelter syndrome: optimal timing for sperm and testicular tissue cryopreservation. *Rev Urol*. 2018;20(2):56–62.
45. Wilch ES, Morton CC. Historical and clinical perspectives on chromosomal translocations. *Adv Exp Med Biol*. 2018;1044:1–14.
46. Zhao WW, Wu M, Chen F, et al. Robertsonian translocations: an overview of 872 Robertsonian translocations identified in a diagnostic laboratory in China. *PLoS One*. 2015;10(5):e0122647.
47. Zhang H, Wang R, Yu Y, et al. Non-Robertsonian translocations involving chromosomes 13, 14, or 15 in male infertility: 28 cases and a review of the literature. *Medicine (Baltim)*. 2019;98(9):e14730.
48. Wiland E, Olszewska M, Woźniak T, Kurpisz M. How much, if anything, do we know about sperm chromosomes of Robertsonian translocation carriers? *Cell Mol Life Sci*. 2020;77(23):4765–4785.
49. Morin SJ, Eccles J, Iturriaga A, Zimmerman RS. Translocations, inversions and other chromosome rearrangements. *Fertil Steril*. 2017;107(1):19–26.
50. Krausz C, Hoefsloot L, Simoni M, Tuttelmann F. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology*. 2014;2(1):5–19.
51. Colaco S, Modi D. Genetics of the human Y chromosome and its association with male infertility. *Reprod Biol Endocrinol*. 2018;16(1):14.
52. Iijima M, Shigehara K, Igarashi H, et al. Y chromosome microdeletion screening using a new molecular diagnostic method in 1030 Japanese males with infertility. *Asian J Androl*. 2020;22(4):368–371.
53. Liu C, Zhao X, Mu C, et al. The association of partial azoospermia factor C deletions and male infertility in Northwestern China. *Hum Hered*. 2019;84(3):144–150.
54. Krausz C, Casamonti E. Spermatogenic failure and the Y chromosome. *Hum Genet*. 2017;136(5):637–655.
55. Moreno-Mendoza D, Casamonti E, Paoli D, et al. gr/gr deletion predisposes to testicular germ cell tumour independently from altered spermatogenesis: results from the largest European study. *Eur J Hum Genet*. 2019;27(10):1578–1588.
56. Lo Giacco D, Chianese C, Ars E, et al. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *J Med Genet*. 2014;51(5):340–344.
57. Krausz C, Giachini C, Lo Giacco D, et al. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS One*. 2012;7(10):e44887.
58. Lopes AM, Aston KI, Thompson E, et al. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet*. 2013;9(3):e1003349.
59. Shen Y, Xu J, Yang X, et al. Evidence for the involvement of the proximal copy of the MAGEA9 gene in Xq28-linked CNV67 specific to spermatogenic failure. *Biol Reprod*. 2017;96(3):610–616.
60. Yatsenko AN, Georgiadis AP, Ropke A, et al. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N Engl J Med*. 2015;372(22):2097–2107.
61. McPhaul MJ, Marcelli M, Zoppi S, et al. Genetic basis of endocrine disease. 4. The spectrum of mutations in the androgen receptor gene that causes androgen resistance. *J Clin Endocrinol Metab*. 1993;76(1):17–23.
62. Yong EL, Loy CJ, Sim KS. Androgen receptor gene and male infertility. *Hum Reprod*. 2003;9(1):1–7.
63. Pan B, Li R, Chen Y, et al. Genetic association between androgen receptor gene CAG repeat length polymorphism and male infertility: a meta-analysis. *Medicine (Baltim)*. 2016;95(10):e2878.
64. Ferlin A, Bartoloni L, Rizzo G, et al. Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility. *Mol Hum Reprod*. 2004;10(6):417–421.
65. Thirumavalavan N, Gabrielsen JS, Lamb DJ. Where are we going with gene screening for male infertility? *Fertil Steril*. 2019;111(5):842–850.
66. Hodžić A, Maver A, Plaseska-Karanfilska D, et al. De novo mutations in idiopathic male infertility-a pilot study. *Andrology*. 2021;9(1):212–220.
67. Capalbo A, Poli M, Riera-Escamilla A, et al. Preconception genome medicine: current state and future perspectives to improve infertility diagnosis and reproductive and health outcomes based on individual genomic data. *Hum Reprod Update*. 2021;27(2):254–279.
68. Tuttelmann F, Ruckert C, Ropke A. Disorders of spermatogenesis: perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Med Genet: Mitteilungsblatt des Berufsverbandes Medizinische Genetik eV*. 2018;30(1):12–20.
69. Krausz C, Riera-Escamilla A, Moreno-Mendoza D, et al. Genetic dissection of spermatogenic arrest through exome analysis: clinical implications for the management of azoospermic men. *Genet Med*. 2020;22(12):1956–1966.
70. Krausz C, Riera-Escamilla A, Chianese C, et al. From exome analysis in idiopathic azoospermia to the identification of a high-risk subgroup for occult Fanconi anemia. *Genet Med*. 2019;21(1):189–194.
71. Skakkebaek NE, Rajpert-De Meyts E, Buck Louis GM, et al. Male reproductive disorders and fertility trends: influences of environment and genetic susceptibility. *Physiol Rev*. 2016;96(1):55–97.
72. Virtanen HE, Toppari J. Cryptorchidism and fertility. *Endocrinol Metab Clin North Am*. 2015;44(4):751–760.
73. Rajpert-De Meyts E. Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects. *Hum Reprod Update*. 2006;12(3):303–323.
74. Kalfa N, Gaspari L, Ollivier M, et al. Molecular genetics of hypospadias and cryptorchidism recent developments. *Clin Genet*. 2019;95(1):122–131.
75. Ramanathan R, Voigt EA, Kennedy RB, Poland GA. Knowledge gaps persist and hinder progress in eliminating mumps. *Vaccine*. 2018;36(26):3721–3726.
76. Rubin S, Eckhaus M, Rennick LJ, et al. Molecular biology, pathogenesis and pathology of mumps virus. *J Pathol*. 2015;235(2):242–252.
77. Fode M, Fusco F, Lipshultz L, Weidner W. Sexually transmitted disease and male infertility: a systematic review. *Eur Urol Focus*. 2016;2(4):383–393.
78. Moazenchi M, Totonchi M, Salman Yazdi R, et al. The impact of Chlamydia trachomatis infection on sperm parameters and male fertility: a comprehensive study. *Int J STD AIDS*. 2018;29(5):466–473.
79. Gimenes F, Souza RP, Bento JC, et al. Male infertility: a public health issue caused by sexually transmitted pathogens. *Nat Rev Urol*. 2014;11(12):672–687.
80. Moreno-Sepulveda J, Rajmil O. Seminal human papillomavirus infection and reproduction: a systematic review and meta-analysis. *Andrology*. 2021;9(2):478–502.
81. Garolla A, Pizzol D, Bertoldo A, et al. Sperm viral infection and male infertility: focus on HBV, HCV, HIV, HPV, HSV, HCMV, and AAV. *J Reprod Immunol*. 2013;100(1):20–29.

82. Chow EJ, Stratton KL, Leisenring WM, et al. Pregnancy after chemotherapy in male and female survivors of childhood cancer treated between 1970 and 1999: a report from the Childhood Cancer Survivor Study cohort. *Lancet Oncol*. 2016;17(5):567–576.
83. Oktay K, Harvey BE, Partridge AH, et al. Fertility preservation in patients with cancer: ASCO Clinical Practice Guideline update. *J Clin Oncol*. 2018;36(19):1994–2001.
84. Martinez F. Update on fertility preservation from the Barcelona International Society for Fertility Preservation-ESHRE-ASRM 2015 expert meeting: indications, results and future perspectives. *Fertil Steril*. 2017;108(3):407–415. e11.
85. Skinner R, Mulder RL, Kremer LC, et al. Recommendations for gonadotoxicity surveillance in male childhood, adolescent, and young adult cancer survivors: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium. *Lancet Oncol*. 2017;18(2):e75–e90.
86. Gassei K, Orwig KE. Experimental methods to preserve male fertility and treat male factor infertility. *Fertil Steril*. 2016;105(2):256–266.
87. Nassau DE, Chu KY, Blachman-Braun R, et al. The pediatric patient and future fertility: optimizing long-term male reproductive health outcomes. *Fertil Steril*. 2020;113(3):489–499.
88. De Felice F, Marchetti C, Marampon F, et al. Radiation effects on male fertility. *Andrology*. 2019;7(1):2–7.
89. Shimizu S, Tsounapi P, Dimitriadis F, et al. Testicular torsion-detorsion and potential therapeutic treatments: a possible role for ischemic postconditioning. *Int J Urol*. 2016;23(6):454–463.
90. Gielchinsky I, Suraqui E, Hidas G, et al. Pregnancy rates after testicular torsion. *J Urol*. 2016;196(3):852–855.
91. The influence of varicocele on parameters of fertility in a large group of men presenting to infertility clinics. World Health Organization. *Fertil Steril*. 1992;57(6):1289–1293.
92. Jensen CFS, Østergren P, Dupree JM, et al. Varicocele and male infertility. *Nat Rev Urol*. 2017;14(9):523–533.
93. Johnson D, Sandlow J. Treatment of varicoceles: techniques and outcomes. *Fertil Steril*. 2017;108(3):378–384.
94. Sironen A, Shoemark A, Patel M, et al. Sperm defects in primary ciliary dyskinesia and related causes of male infertility. *Cell Mol Life Sci*. 2020;77(11):2029–2048.
95. Liu C, Lv M, He X, et al. Homozygous mutations in SPEF2 induce multiple morphological abnormalities of the sperm flagella and male infertility. *J Med Genet*. 2020;57(1):31–37.
96. Tang S, Wang X, Li W, et al. Biallelic mutations in CFAP43 and CFAP44 cause male infertility with multiple morphological abnormalities of the sperm flagella. *Am J Hum Genet*. 2017;100(6):854–864.
97. Touré A, Martinez G, Kherraf ZE, et al. The genetic architecture of morphological abnormalities of the sperm tail. *Hum Genet*. 2021;140(1):21–42.
98. Oud MS, Okutman Ö, Hendricks LAJ et al. Exome sequencing reveals novel causes as well as new candidate genes for human globozoospermia. *Hum Reprod*. 2020;35(1):240–252.
99. Chianese C, Fino MG, Riera Escamilla A, et al. Comprehensive investigation in patients affected by sperm macrocephaly and globozoospermia. *Andrology*. 2015;3(2):203–212.
100. Johnson SL, Dunleavy J, Gemmell NJ, Nakagawa S. Consistent age-dependent declines in human semen quality: a systematic review and meta-analysis. *Ageing Res Rev*. 2015;19:22–33.
101. Vaughan DA, Tirado E, Garcia D, et al. DNA fragmentation of sperm: a radical examination of the contribution of oxidative stress and age in 16 945semen samples. *Hum Reprod*. 2020;35(10):2188–2196.
102. Cioppi F, Casamonti E, Krausz C. Age-dependent de novo mutations during spermatogenesis and their consequences. *Adv Exp Med Biol*. 2019;1166:29–46.
103. Goriely A, Wilkie AO. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet*. 2012;90(2):175–200.
104. Janeczko D, Hołowczuk M, Orzeł A, et al. Paternal age is affected by genetic abnormalities, perinatal complications and mental health of the offspring. *Biomed Rep*. 2020;12(3):83–88.
105. Barbonetti A, Castellini C, D’Andrea S, et al. Prevalence of anti-sperm antibodies and relationship of degree of sperm auto-immunization to semen parameters and post-coital test outcome: a retrospective analysis of over 10 000 men. *Hum Reprod*. 2019;34(5):834–841.
106. Cui D, Han G, Shang Y, et al. Antisperm antibodies in infertile men and their effect on semen parameters: a systematic review and meta-analysis. *Clin Chim Acta*. 2015;444:29–36.
107. Bozhedomov VA, Nikolaeva MA, Ushakova IV, et al. Functional deficit of sperm and fertility impairment in men with antisperm antibodies. *J Reprod Immunol*. 2015;112:95–101.
108. Fijak M, Pilatz A, Hedger MP, et al. Infectious, inflammatory and “autoimmune” male factor infertility: how do rodent models inform clinical practice? *Hum Reprod Update*. 2018;24(4):416–441.
109. Simon L, Emery BR, Carrell DT. Review: diagnosis and impact of sperm DNA alterations in assisted reproduction. *Best Pract Res Clin Obstet Gynaecol*. 2017;44:38–56.
110. de Souza DAS, Faucz FR, Pereira-Ferrari L, et al. Congenital bilateral absence of the vas deferens as an atypical form of cystic fibrosis: reproductive implications and genetic counseling. *Andrology*. 2018;6(1):127–135.
111. Bieth E, Hamdi SM, Mieusset R. Genetics of the congenital absence of the vas deferens. *Hum Genet*. 2021;140(1):59–76.
112. Handelsman DJ, Conway AJ, Boylan LM, Turtle JR. Young’s syndrome. Obstructive azoospermia and chronic sinopulmonary infections. *N Engl J Med*. 1984;310(1):3–9.
113. Arya AK, Beer HL, Benton J, et al. Does Young’s syndrome exist? *J Laryngol Otol*. 2009;123(5):477–481.
114. Elbashir S, Magdi Y, Rashed A, et al. Epididymal contribution to male infertility: an overlooked problem. *Andrologia*. 2021;53(1):e13721.
115. Namekawa T, Imamoto T, Kato M, et al. Vasovasostomy and vasoepididymostomy: review of the procedures, outcomes, and predictors of patency and pregnancy over the last decade. *Reprod Med Biol*. 2018;17(4):343–355.
116. Verze P, Cai T, Lorenzetti S. The role of the prostate in male fertility, health and disease. *Nat Rev Urol*. 2016;13(7):379–386.
117. Lotti F, Corona G, Cocci A, et al. The prevalence of midline prostatic cysts and the relationship between cyst size and semen parameters among infertile and fertile men. *Hum Reprod*. 2018;33(11):2023–2034.
118. Chan PTK, Aubé M. Infertility due to ejaculatory duct obstruction can be surgically corrected-a forgotten lesson in male reproductive urology. *Fertil Steril*. 2018;110(7):1275–1276.
119. Krausz C, Escamilla AR, Chianese C. Genetics of male infertility: from research to clinic. *Reproduction*. 2015;150(5):R159–R174.
120. MacDonald AA, Herbison GP, Showell M, Farquhar CM. The impact of body mass index on semen parameters and reproductive hormones in human males: a systematic review with meta-analysis. *Hum Reprod Update*. 2010;16(3):293–311.
121. Lainez NM, Coss D. Obesity, neuroinflammation, and reproductive function. *Endocrinology*. 2019;160(11):2719–2736.
122. Craig JR, Jenkins TG, Carrell DT, Hotaling JM. Obesity, male infertility, and the sperm epigenome. *Fertil Steril*. 2017;107(4):848–859.
123. Donkin I, Barrès R. Sperm epigenetics and influence of environmental factors. *Mole Metab*. 2018;14:1–11.
124. Kahn BE, Brannigan RE. Obesity and male infertility. *Curr Opin Urol*. 2017;27(5):441–445.
125. Sermondade N, Faure C, Fezeu L, et al. BMI in relation to sperm count: an updated systematic review and collaborative meta-analysis. *Hum Reprod Update*. 2013;19(3):221–231.
126. Sermondade N, Faure C, Fezeu L, et al. Obesity and increased risk for oligozoospermia and azoospermia. *Arch Intern Med*. 2012;172(5):440–442.
127. Martini AC, Tissera A, Estofan D, et al. Overweight and seminal quality: a study of 794 patients. *Fertil Steril*. 2010;94(5):1739–1743.
128. Chavarro JE, Toth TL, Wright DL, et al. Body mass index in relation to semen quality, sperm DNA integrity, and serum reproductive hormone levels among men attending an infertility clinic. *Fertil Steril*. 2010;93(7):2222–2231.
129. Di Vincenzo A, Busetto L, Vettor R, Rossato M. Obesity, male reproductive function and bariatric surgery. *Front Endocrinol*. 2018;9:769.

130. Hayden RP, Flannigan R, Schlegel PN. The role of lifestyle in male infertility: diet, physical activity, and body habitus. *Curr Urol Rep*. 2018;19(7):56.
131. Rastrelli G, Lotti F, Reisman Y, et al. Metabolically healthy and unhealthy obesity in erectile dysfunction and male infertility. *Exp Rev Endocrinol Metab*. 2019;14(5):321–334.
132. Reis LO, Dias FG. Male fertility, obesity, and bariatric surgery. *Reprod Sci*. 2012;19(8):778–785.
133. Hakonsen LB, Thulstrup AM, Aggerholm AS, et al. Does weight loss improve semen quality and reproductive hormones? Results from a cohort of severely obese men. *Reprod Health*. 2011;8:24.
134. Gore AC, Chappell VA, Fenton SE, et al. EDC-2: the Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36(6):E1–E150.
135. Lombó M, Fernández-Díez C, González-Rojo S, Herráez MP. Genetic and epigenetic alterations induced by bisphenol A exposure during different periods of spermatogenesis: from spermatozoa to the progeny. *Sci Rep*. 2019;9(1):18029.
136. Gillette R, Son MJ, Ton L, et al. Passing experiences on to future generations: endocrine disruptors and transgenerational inheritance of epimutations in brain and sperm. *Epigenetics*. 2018;13(10–11):1106–1126.
137. Ho SM, Cheong A, Adgent MA, et al. Environmental factors, epigenetics, and developmental origin of reproductive disorders. *Reprod Toxicol*. 2017;68:85–104.
138. Stenz L, Escoffier J, Rahban R, et al. Testicular dysgenesis syndrome and long-lasting epigenetic silencing of mouse sperm genes involved in the reproductive system after prenatal exposure to DEHP. *PLoS One*. 2017;12(1):e0170441.
139. Swan SH, Kruse RL, Liu F, et al. Semen quality in relation to biomarkers of pesticide exposure. *Environ Health Perspect*. 2003;111(12):1478–1484.
140. Swan SH, Brazil C, Drobnis EZ, et al. Geographic differences in semen quality of fertile U.S. males. *Environ Health Perspect*. 2003;111(4): 414–420.
141. Bonde JP, Flachs EM, Rimborg S, et al. The epidemiologic evidence linking prenatal and postnatal exposure to endocrine disrupting chemicals with male reproductive disorders: a systematic review and meta-analysis. *Hum Reprod Update*. 2016;23(1):104–125.
142. Hanson BM, Eisenberg ML, Hotaling JM. Male infertility: a biomarker of individual and familial cancer risk. *Fertil Steril*. 2018;109(1):6–19.
143. Doria-Rose VP, Biggs ML, Weiss NS. Subfertility and the risk of testicular germ cell tumors (United States). *Cancer Causes Control*. 2005;16(6):651–656.
144. Walsh TJ, Croughan MS, Schembri M, et al. Increased risk of testicular germ cell cancer among infertile men. *Arch Intern Med*. 2009;169(4):351–356.
145. Sharma R, Harlev A, Agarwal A, Esteves SC. Cigarette smoking and semen quality: a new meta-analysis examining the effect of the 2010 World Health Organization laboratory methods for the examination of human semen. *Eur Urol*. 2016;70(4):635–645.
146. Jensen TK, Jorgensen N, Punab M, et al. Association of in utero exposure to maternal smoking with reduced semen quality and testis size in adulthood: a cross-sectional study of 1,770 young men from the general population in five European countries. *Am J Epidemiol*. 2004;159(1):49–58.
147. Ramlau-Hansen CH, Thulstrup AM, Storgaard L, et al. Is prenatal exposure to tobacco smoking a cause of poor semen quality? A follow-up study. *Am J Epidemiol*. 2007;165(12):1372–1379.
148. Jenkins TG, James ER, Alonso DF, et al. Cigarette smoking significantly alters sperm DNA methylation patterns. *Andrology*. 2017;5(6):1089–1099.
149. Marczylo EL, Amoako AA, Konje JC, et al. Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? *Epigenetics*. 2012;7(5):432–439.
150. Practice Committee of the American Society for Reproductive Medicine. Smoking and infertility: a committee opinion. *Fertil Steril*. 2018;110(4):611–618.
151. Medicine. PCotASfR. Smoking and infertility: a committee opinion. *Fertil Steril*. 2018;110(4):611–618.
152. Altun G, Deniz Ö G, Yurt KK, et al. Effects of mobile phone exposure on metabolomics in the male and female reproductive systems. *Environ Res*. 2018;167:700–707.
153. Kandeel FR, Swerdloff RS. Role of temperature in regulation of spermatogenesis and the use of heating as a method for contraception. *Ferti Steril*. 1988;49(1):1–23.
154. Wang C, Cui YG, Wang XH, et al. Transient scrotal hyperthermia and levonorgestrel enhance testosterone-induced spermatogenesis suppression in men through increased germ cell apoptosis. *J Clin Endocrinol Metab*. 2007;92(8):3292–3304.
155. Thonneau P, Ducot B, Bujan L, et al. Effect of male occupational heat exposure on time to pregnancy. *Int J Androl*. 1997;20(5):274–278.
156. Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol*. 2018;15(6):369–384.
157. McSwiggin HM, O'Doherty AM. Epigenetic reprogramming during spermatogenesis and male factor infertility. *Reproduction*. 2018;156(2): R9-R21.
158. Marcho C, Oluwayiose OA, Pilsner JR. The preconception environment and sperm epigenetics. *Andrology*. 2020;8(4):924–942.
159. Jenkins TG, Aston KI, James ER, Carrell DT. Sperm epigenetics in the study of male fertility, offspring health, and potential clinical applications. *Syst Biol Reprod Med*. 2017;63(2):69–76.
160. Donkin I, Barres R. Sperm epigenetics and influence of environmental factors. *Mole Metab*. 2018;14:1–11.
161. Donkin I, Versteyhe S, Ingerslev LR, et al. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab*. 2016;23(2):369–378.
162. Morgan HL, Paganopoulou P, Akhtar S, et al. Paternal diet impairs F1 and F2 offspring vascular function through sperm and seminal plasma specific mechanisms in mice. *J Physiol*. 2020;598(4):699–715.
163. Burl RB, Clough S, Sendler E, et al. Sperm RNA elements as markers of health. *Syst Biol Reprod Med*. 2018;64(1):25–38.
164. Gòdia M, Swanson G, Krawetz SA. A history of why fathers' RNA matters. *Biol Reprod*. 2018;99(1):147–159.
165. Jenkins TG, Aston KI, Hotaling JM, et al. Teratozoospermia and asthenozoospermia are associated with specific epigenetic signatures. *Andrology*. 2016;4(5):843–849.
166. Abbasi M, Smith AD, Swaminathan H, et al. Establishing a stable, repeatable platform for measuring changes in sperm DNA methylation. *Clin Epigenet*. 2018;10(1):119.
167. Krzastek SC, Smith RP, Kovac JR. Future diagnostics in male infertility: genomics, epigenetics, metabolomics and proteomics. *Translat Androl Urol*. 2020;9(suppl 2):S195–S205.
168. Patel DP, Jenkins TG, Aston KI, et al. Harnessing the full potential of reproductive genetics and epigenetics for male infertility in the era of "big data". *Fertil Steril*. 2020;113(3):478–488.

113

Clinical Management of Male Infertility

Peter N. Schlegel

OUTLINE

Introduction, 1892
Health Considerations For Infertile Men, 1892
  Evaluation of Male Infertility, 1892
  History and Physical Examination, 1893
  Semen Analysis, 1894
  Further Investigations and Complete Evaluation, 1895
  Endocrine Evaluation, 1895
  Low Semen Volume, 1895
  Postejaculatory Urinalysis, 1895
  Transrectal Ultrasound, 1896
  Oligospermia/Asthenospermia/Teratospermia, 1896
  Semen Leukocytes and Screening for Genital Infection, 1896
  Antisperm Antibodies (ASA), 1896
  Genetic Testing, 1897
  Sperm DNA Fragmentation, 1897
  Other Genetic Anomalies, 1898
  Azoospermia, 1898
Management of Male Infertility, 1898
  Lifestyle Modifications, 1898
  Hormone Optimization, 1899
  Ejaculatory Dysfunction, 1899
  Ejaculatory Duct Obstruction, 1899
  Pyospermia, 1899
  Immunological Infertility, 1900
  Varicocele Repair, 1900
  Vasal Reconstruction, 1901
  Assisted Reproductive Techniques (ART), 1902
  Sperm Retrieval, 1902
Conclusion, 1902

INTRODUCTION

Infertility, defined as the inability to conceive after 12 months of regular, unprotected intercourse, affects 8 million couples in the United States, of which up to 50% have a male contributory factor.[1,2] Overall, up to 12% of men of reproductive age worldwide suffer from male infertility.[3] Fortunately, a tremendous increase in our understanding of male infertility and our ability to provide novel diagnostic and therapeutic approaches has occurred over the past 50 years. Of note, advanced reproductive technologies such as *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI) have both enabled men for whom fatherhood could be achieved previously only through donor sperm or adoption to now father biological offspring as well as to provide greater insight into the role of sperm in the outcomes of fertility treatment.

HEALTH CONSIDERATIONS FOR INFERTILE MEN

Progressive accumulation of data has indicated that men who are diagnosed as being infertile or identified as having impaired semen parameters (oligospermia) may be at increased risk of medical conditions later in life, including testicular cancer.[4] For example, men who require treatment with assisted reproduction for male factor infertility are at increased risk of subsequent development of prostate cancer.[5] Men with azoospermia are at increased risk of several different types of cancer later in life, so routine screening or follow-up for cancer should be considered after fertility treatment.[6] Men with infertility are also at increased risk of comedical conditions including obesity, diabetes, and hyperlipidemia, so men may reach medical care because of an infertility diagnosis and evaluation.[7] Obviously, follow-up management of these other comorbidities is needed for long-term health optimization. In addition, a specific diagnosis (e.g., diagnosis of Klinefelter syndrome) may result in certain risks of disease, including extragonadal germ cell tumor, breast cancer, and metabolic syndrome. Conditions associated with infertility such as low testosterone are also associated with other medical conditions, including diabetes, metabolic syndrome, cardiovascular disease, hypertension, all-cause mortality, and Alzheimer's disease.[8]

KEY POINTS

- Infertile men, especially those with azoospermia, are at increased risk of cancer and other medical comorbidities.
- Counseling regarding these risks is indicated after treatment of infertility. Specific risks associated with conditions causing infertility, such as Klinefelter syndrome, are also needed (see Chapter 107).

Recommendations from the American Urological Association and the American Society for Reproductive Medicine advocate that initial evaluation of the infertile couple includes investigation of both partners in parallel.[9] In this chapter, we provide an updated outline of evaluation and treatment algorithms that guide the management of the infertile male.

Evaluation of Male Infertility

When and on Whom to Perform an Evaluation of Male Infertility. An initial screening evaluation on both partners is recommended if a couple has not conceived a pregnancy after 12 months of regular,

BOX 113.1 Significant Medical Conditions That May Present as Male Infertility[84,85]

Adult polycystic kidney
Brain tumors
Cystic fibrosis
Diabetes
Hemochromatosis
Hypopituitarism
Klinefelter syndrome
Multiple sclerosis
Pituitary adenoma
Prostate cancer
Retroperitoneal tumors
Spinal cord tumors
Testis cancer
Thyroid disease
Urinary tract infection

unprotected intercourse. Alternatively, an earlier evaluation is justified if the female partner is older than 35 years old, if either partner has a medical history or physical findings suggestive of decreased reproductive potential, or if either partner is concerned about the male's reproductive potential. Men who do not have a partner but who are concerned about their reproductive potential should also undergo an initial screening evaluation.

The goals of the evaluation of male infertility are to identify and treat correctable causes of male infertility that could then result in natural pregnancy, to guide treatment options in those for whom natural pregnancy is unlikely to occur, and to identify significant medical problems that are associated with infertility (Box 113.1).

Initial Screening Evaluation

KEY POINTS

- Initial screening evaluation for male infertility should involve a reproductive history and at least one semen analysis.
- For men with abnormalities on initial evaluation, a complete history, physical examination, and at least two semen analyses are needed (see Chapter 116).
- While abnormalities on semen analysis are not diagnostic of male infertility, semen analysis is a cornerstone in the evaluation of male infertility.
- Abnormalities on semen analysis are more predictive of infertility if there are multiple abnormal semen parameters.

History and Physical Examination

A thorough reproductive history (Table 113.1) is a critical component of the initial screening evaluation of the infertile male and should cover coital frequency and timing, duration of infertility, prior fertility, developmental history and childhood illnesses (including hypospadias, congenital anomalies, and onset of puberty), current systemic illnesses (especially diabetes mellitus, cystic fibrosis, cancer, and infections), surgical history (with a focus on scrotal, pelvic, retroperitoneal, and inguinal surgeries), sexual history (with a history of sexually transmitted diseases), medications and allergies, and exposure to gonadotoxins (both chemical and environmental toxins such as heat; see Table 113.2). Additionally, a family history of cryptorchidism, infertility, disordered sexual differentiation, or hypogonadism and/or a social history with frequent alcohol, tobacco, or recreational drug use may suggest potential etiologies for male infertility and lead to further evaluation.

A general physical examination with an emphasis on the genitalia can provide additional important information on conditions that can affect male fertility. Upon initial inspection,

note should be taken of obesity, secondary sex characteristics, hair distribution, and gynecomastia. Specific examination of the genitalia should include the phallus, which should be examined for hypospadias, penile curvature, and plaques that may interfere with proper semen deposition within the vagina. Both testes should be examined for size, consistency, and the presence of masses. Over 70% of human testicular volume is made up of seminiferous tubules in which spermatogenesis occurs and testicles smaller than the normal adult size of 20 mL (or 4.5 cm × 3 cm) may be associated with impaired spermatogenesis.[10]

Careful attention should be paid to the presence, consistency, and nodularity of the epididymides that may help to identify obstruction and/or infection while palpation of both vasa deferens can diagnose vasal agenesis, atresia, or injury. Congenital bilateral absence of the vas deferens (CBAVD) is a clinical diagnosis that should be made on physical examination. Palpation of both spermatic cords in the supine and standing positions with and without Valsalva enables the diagnosis of varicocele, which is also a clinical diagnosis. Scrotal ultrasound may be used to confirm diagnosis of varicocele, but those subclinical varicoceles found only on ultrasound (and not on physical examination)

TABLE 113.1 Important Parts of the Male Infertility History

Infertility History
- Duration
- Prior pregnancies
    - Current vs previous partner
    - Outcome
- Prior fertility evaluation

Sexual History
- Coital frequency and timing
- Erectile function
- Use of lubricants

Past Medical History
Developmental History
- Cryptorchidism
- Puberty
- Testicular pathology (e.g., trauma or torsion)
- Hypospadias

Medical History
- Diabetes mellitus
- Neurologic disease (spinal cord injury, multiple sclerosis)
- Fevers or infections (urinary, sexually transmitted, tuberculosis, prostatitis)
- Cancer and treatment (chemotherapy or radiation)
- Genetic disease (Klinefelter, cystic fibrosis)

Surgical History
- Genital (orchidopexy, testicular trauma, vasectomy)
- Pelvic (herniorrhaphy, prostatic, bladder neck)
- Abdominal/retroperitoneal

Family History
- Infertility
- Cryptorchidism
- Disorders of sexual differentiation
- Midline defects
- Cystic fibrosis

Medications
Social History
- Recreational drugs (smoking, alcohol, marijuana, cocaine)
- Environmental/occupational exposure (pesticides, radiation, heavy metals, heat)

TABLE 113.2 List of Agents and Factors With Potential Adverse Impact on Male Fertility

Medications

- Gonadotoxic agents
  - Antibiotics
  - Nitrofurantoin
  - Erythromycin
  - Tetracycline
  - Gentamycin
  - Sulfasalazine
  - Chemotherapeutic agents
  - Alkylating agents (cyclophosphamide, ifosfamide)
  - Cisplatin
  - Actinomycin D
  - Marijuana
  - Alcohol
- Androgen production inhibitors
  - Spironolactone
  - Ketoconazole
  - Cimetidine
  - Flutamide
  - 5α-reductase inhibitors
- Ejaculatory dysfunction
  - Alpha-blockers
  - Lithium
  - Antipsychotics
  - Tricyclic antidepressants
  - Valproic acid
  - Phenytoin
  - SSRIs
- Asymptomatic sperm DNA damage
  - SSRIs
- Hypothalamus–pituitary–gonadal axis inhibitors
  - Testosterone
  - Anabolic steroids
  - Monoamine oxidase inhibitors
  - Phenothiazines
  - DES
  - Digoxin
  - Alcohol
  - Opioids
- Sperm toxic agents
  - Nicotine
  - Calcium channel blockers
  - Sulfasalazine

Environmental

- Heavy metals (lead)
- Pesticides (dibromochloropropane)
- Heat
- Bisphenol A
- Organic solvents
- Ionizing radiation

DES, Diethylstilbestrol; *SSRI,* selective serotonin reuptake inhibitor.

are unlikely to be clinically relevant.[9] Finally, digital rectal examination may help identify dilated seminal vesicles, ejaculatory duct cysts, or utricular cysts that may contribute to obstructive etiologies of male infertility, such examination is most relevant for men with abnormal semen volume.

TABLE 113.3 WHO (5th Edition) Reference Values for Semen Analysis[29]

Semen Analysis Parameter	Reference Value
Ejaculate volume	1.5 mL
Sperm concentration	15 million sperm/mL
Sperm count	39 million sperm/ejaculate
Total motility	40%
Total progressive motility	32%
Sperm morphology	4% normal (strict criteria)
Sperm viability	58%
Sperm agglutination	Absent
White blood cells	<1 million leukocytes/mL

Semen Analysis

Along with a thorough history and physical examination, semen analysis has long been the pillar upon which male infertility is evaluated and managed. While an abnormal semen analysis (with the exception of azoospermia) cannot delineate between fertile and sterile, abnormal semen quality is associated with decreased chances of natural conception.[11]

To account for the large degree of biological variability between semen samples, at least two semen analyses should be performed. Because environmental exposures can change over time and spermatogenesis takes more than 2 months, repeat semen samples after a month or more may be needed to validate abnormal semen parameters. Patients should be given specific instructions regarding the need for 2 to 5 days of abstinence prior to undergoing semen analysis with the two semen analyses.[14] Although the standard abstinence period is important for comparison of samples to "normal" standards, it is possible that men with low sperm production, such as those with severe oligospermia, may actually have better semen parameters with very short abstinence, even abstinence of a few hours. Indeed, men with severe oligospermia or cryptospermia may benefit from providing multiple semen specimens within hours, especially for assisted reproductive treatments such as IVF/ICSI.

A semen sample can be collected either by masturbation into a specimen cup or by intercourse with the use of a special semen collection condom that does not contain any spermatotoxic agents. In particular, lubricants should be avoided. If a specimen cannot be produced at the laboratory for analysis, it should be kept at room or body temperature during transport and delivered to the laboratory for analysis, preferably within 1 hour. Further delays can affect semen and sperm parameters such as sperm motility, which may decrease dramatically after 2 hours.

Parameters examined in routine semen analysis can be broken up into macroscopic and microscopic parameters. Key macroscopic parameters include semen volume, viscosity, color, pH, and coagulation/liquefaction, while key microscopic parameters include sperm count/concentration, sperm motility, and sperm morphology. After analyzing the semen analyses of 1859 new fathers from eight countries across three continents, the World Health Organization (WHO) established the normal reference ranges for most of the key parameters of semen analysis (Table 113.3).[15] To establish the reference ranges, the WHO applied a one-sided lower reference limit of the 5th percentile to the semen parameters of the 1859 new fathers. While this methodology provides lower thresholds for semen parameters in fertile men, it fails to answer the more relevant clinical question of which specific semen parameters and cutoffs are able to delineate between male fertility and subfertility. Efforts to identify specific semen parameters that can distinguish between male fertility and subfertility have instead found that while semen parameters are associated with fecundity, neither sperm concentration, morphology, nor motility could be considered diagnostic of infertility either alone or in combination.[13,16] Rather than a direct diagnostic test of male infertility, the semen analysis should instead be considered an important contributor to the overall picture of male fertility. The presence of multiple semen parameter abnormalities, especially low sperm concentration, motility, and morphology, may have the greatest effects on fertility potential.

Further Investigations and Complete Evaluation

KEY POINTS

- Further investigations of male infertility are directed by the initial evaluation. Critical abnormal semen parameters may include detection of low semen volume, oligoasthenoteratospermia, or azoospermia.
- Endocrine evaluation of male infertility should include, at a minimum, serum testosterone and follicle-stimulating hormone (FSH). Low serum testosterone should be further evaluated with LH determination; men with abnormal libido or hypogonadotropic hypogonadism should be considered for testing with serum prolactin evaluation.

Men with an unremarkable reproductive history and normal semen analysis(es) are highly unlikely to harbor male infertility and a greater focus on potential female factors should be undertaken. However, if any abnormalities are detected on the initial screening evaluation of male infertility, further investigation, as discussed previously, is indicated.

Endocrine Evaluation

While a specific endocrine cause of impaired sperm production is identified in only 2% to 5% of men with male factor infertility, endocrine evaluation is indicated for men with abnormal semen analyses (especially in those with sperm concentrations <10 million/mL), abnormal sexual function, or other findings suggestive of endocrinopathy.[9,17] At minimum, an endocrine evaluation for male infertility should include serum FSH and morning testosterone levels. A low serum testosterone level (<300 ng/dL) should prompt the testing of a second morning serum testosterone level alongside serum luteinizing hormone (LH) and prolactin levels.[18] The importance of obtaining a second morning serum testosterone level should not be understated as gonadotropins and testosterone are typically released in a pulsatile manner and testosterone levels commonly demonstrate a physiologic variation over the course of the day with the highest levels in the early morning. The characteristic hormone profiles for various etiologies of male infertility are described in Table 113.4. In particular, while normal FSH levels (≤7.6 IU/L) do not guarantee normal spermatogenesis, elevated FSH levels often indicate an abnormality with spermatogenesis.[18]

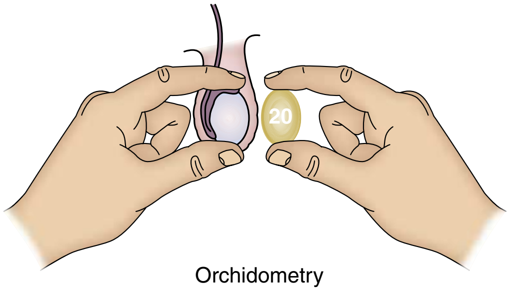
Low Semen Volume

KEY POINTS

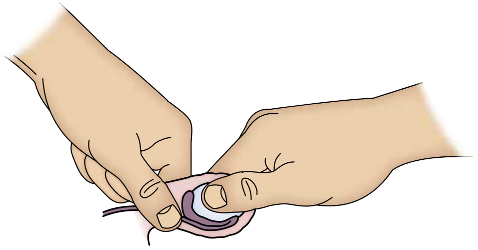
- Low semen volume may be attributable to ejaculatory dysfunction, hypogonadism, ejaculatory duct obstruction, or CBAVD, as well as incomplete semen collection.
- An endocrine evaluation, postejaculatory urinalysis, and transrectal ultrasound can help identify the cause of low semen volume.

TABLE 113.4 Characteristic Hormone Profiles for Various Etiologies of Male Infertility

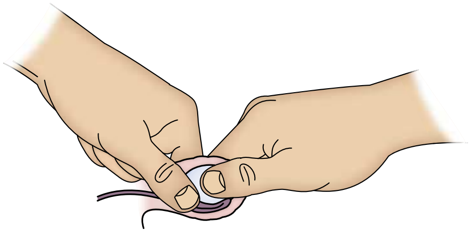
Condition	Testosterone	FSH	LH	Prolactin
Normal	Normal	Normal	Normal	Normal
Primary Testicular Failure				
• Hypospermatogenesis	Normal/low	High	High/normal	Normal
• Maturation arrest	Normal	Normal	Normal	Normal
• Sertoli cell only	Low	Very high	High	Normal
Hypogonadotropic hypogonadism	Low	Low	Low	Normal
Hyperprolactinemia	Low	Normal/low	Low	High
Klinefelter syndrome	Low	Very high	High	Normal



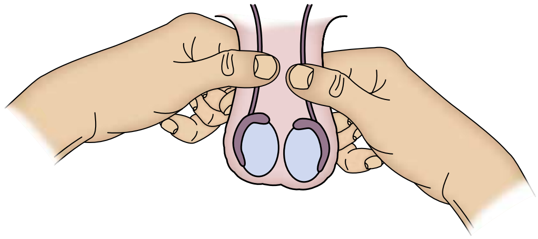
Orchidometry



Palpatation of testis



Palpatation of epididymis



Checking for varicocele

Figure 113.1 Clinical Examination of the Scrotum.

Postejaculatory Urinalysis

The presence of low semen volume (<1.0 mL) should prompt a systematic algorithm to delineate the cause (Fig. 113.1). Potential etiologies for low semen volume include incomplete semen specimen collection, CBAVD, hypogonadism, ejaculatory dysfunction, and ejaculatory duct obstruction. In men with normal serum testosterone levels and palpable vasa deferens, a postejaculatory urinalysis should be performed to rule out retrograde ejaculation. The presence of more sperm than in the urine specimen pellet postejaculation is suggestive of retrograde ejaculation.[15]

Transrectal Ultrasound

Transrectal ultrasound (TRUS) has limited indications for the evaluation of male infertility. Only in the case of significantly low semen volume (<1.0 mL) that is not explained by vasal agenesis, low serum testosterone, retrograde ejaculation, or incomplete collection, should TRUS be considered. The absence of seminal fructose or a seminal pH <7 may also suggest the absence/atrophy of the seminal vesicles or ejaculatory duct obstruction. The diagnostic tool of choice for ejaculatory duct obstruction is TRUS. The presence of midline cysts, dilated seminal vesicles (anterior-posterior diameter >1.5 cm), or dilated ejaculatory ducts (>2.3 mm) on TRUS is suggestive, but not diagnostic, of complete or partial ejaculatory duct obstruction.[16,17] If dilated seminal vesicles are seen on TRUS, TRUS-guided seminal vesicle aspiration may be performed to confirm the diagnosis (with simultaneous cryopreservation of any sperm found).[18] Patients with unilateral absence of the vas deferens and low volume azoospermia may have a variant of CBAVD and should undergo genetic screening for cystic fibrosis (CFTR) before TRUS. Forty percent of men with CBAVD and a negative CFTR screen will demonstrate renal anomalies and should be additionally screened with abdominal ultrasound. On the other hand, those with CBAVD and a positive CFTR screen are at lower risk of renal.[19,20] Men with detectable CFTR mutations should undergo thorough genetic counseling, and their female partner must be tested for genetic abnormalities prior to any fertility treatment, as recommended by the American College of Obstetricians and Gynecologists (ACOG) guidelines for preconceptual counseling.

Oligospermia/Asthenospermia/Teratospermia

KEY POINTS

- Oligoasthenoteratospermia may be caused by varicocele, genetic abnormalities, immunologic infertility, or gonadotoxic exposures.
- Genetic testing with karyotype and Y-microdeletion testing is indicated in men with severe oligospermia (<1 million sperm/mL) or nonobstructive azoospermia (see Chapter 112).
- Genetic testing with CFTR gene mutation analysis is indicated for men with Wolffian abnormalities, including CBAVD.

Infertile men may demonstrate abnormalities in sperm concentration (oligospermia), motility (asthenospermia), and/or morphology (teratospermia), which are often seen together. Potential causes for these abnormal semen parameters include endocrinopathies, genital tract infections, immunologic antisperm antibodies (ASA), varicoceles, and genetic abnormalities (Fig. 113.2).

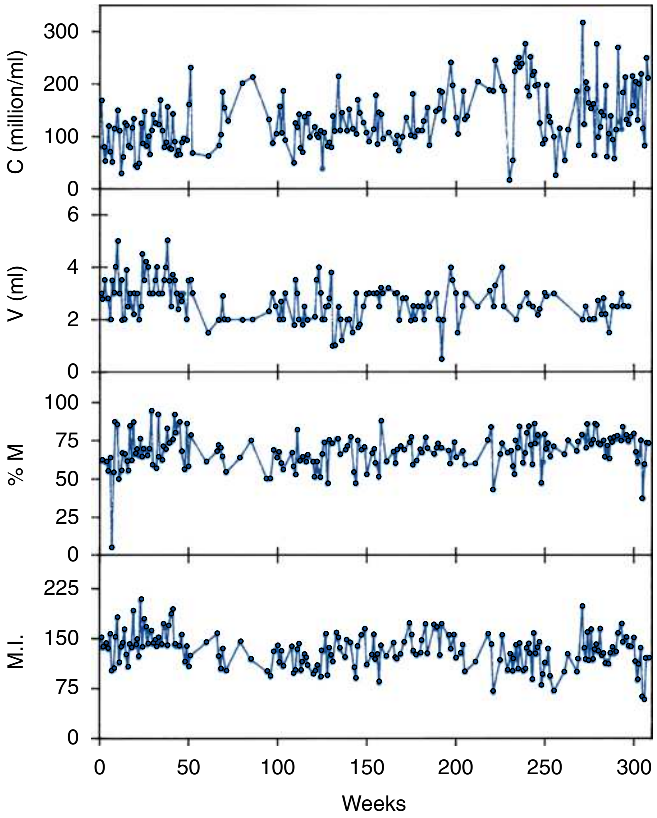


Figure 113.2 Variability of Semen Analysis Results in a Fertile Sperm Donor. C, Sperm concentration; M, total motility; M.I., motility index—product of grade and percentage of sperm with progressive motility graded 0 to 3; V, semen volume. (Data from Mallidis C, Howard EJ, Baker HW: Variation of semen quality in normal men, Int J Androl 14:99–107, 1991. Used by permission, Blackwell Scientific Publications.)

KEY POINTS

- Antisperm antibody testing and routine semen cultures are not indicated for the initial evaluation of male infertility.

Semen Leukocytes and Screening for Genital Infection

Routine semen cultures do not have value in the evaluation of male infertility. However, genital tract infection or inflammation may result in the presence of leukocytes in the semen that can negatively affect sperm motility and function via the overproduction of reactive oxygen species as well as direct effects on spermatozoa. Because both leukocytes and immature germ cells appear as “round cells” on microscopy, differentiating between the two types of cells can be tricky but can be accomplished using cytologic staining or immunohistochemistry.[21] Those with pyospermia (>1 million leukocytes/mL) should be evaluated further to rule out or treat genital tract infection or inflammation. Because 83% of semen cultures in asymptomatic infertile men with pyospermia are positive with multiple organisms, the evaluation may include an examination of expressed prostatic secretion for leukocytes as well as a urethral culture (or urine ligase chain reaction [LCR] test) for Chlamydia and Mycoplasma.[22]

Antisperm Antibodies (ASA)

Patients with isolated asthenospermia or sperm agglutination may be bound by ASA, which is a rare cause of male infertility that can form following a breach of the blood–testis barrier after vasectomy, trauma, torsion, biopsy, or testicular cancer. Routine testing for ASA is not indicated in the evaluation of male infertility as the condition is rare, and the presence of ASA may not affect management of infertility. ASA can be a diagnostic test for obstruction in azoospermic men. Approximately 50% to 80% of men with obstructive azoospermia (OA) have ASA detectable in serum. ASA can be found in the serum or seminal plasma via indirect antibody agglutination assays or directly bound to the sperm via a direct immunobead test. ASA are considered significant if over 50% of sperm are bound by ASA.[23] Because the presence of ASA correlates with spermatogenesis, the detection of serum ASA may also be clinically useful in differentiating azoospermic men with obstructive etiologies versus those with nonobstructive etiologies without the need for a diagnostic testicular biopsy.[24] Azoospermic men with positive ASA are more likely to have an obstructive etiology, with the presence of ASA having an 85% specificity for obstruction.[25]

Genetic Testing (see Chapter 112)

Genetic abnormalities can cause problems with sperm production or transport and men with severe oligospermia (sperm concentration <1 million/mL) or non-OA (NOA) should be worked up for genetic abnormalities (Table 113.5).[4] In addition to the previously discussed CFTR mutations, karyotypic chromosomal abnormalities and Y-microdeletions are of concern. In total, genetic abnormalities are estimated to account for 15% to 30% of male infertility, though the overwhelming majority of these abnormalities have yet to be discovered or understood.[26]

Chromosomal abnormalities are more common in infertile men, with the prevalence of chromosomal abnormalities inversely proportional to sperm count. Ten to fifteen percent of azoospermic men harbor chromosomal abnormalities compared with 5% of men with severe oligospermia and <1% of men with normal sperm concentrations.[27,28] The most common chromosomal abnormality in infertile men is 47,XXY (Klinefelter syndrome), though inversions and balanced translocations are also more prevalent in infertile men when compared with fertile men.[28]

Deletions of clinically relevant sections from the azoospermia factor regions (AZF) on the long arm of the Y chromosome (Yq11) are found in up to 17% of men with non-OA or severe oligospermia.[25] These Y-microdeletions can be categorized by their specific locations, which are prognostic of their ultimate impact on spermatogenesis and infertility. Men with deletions in the AZFa or AZFb regions are azoospermic with a near 0% chance for successful sperm retrieval.[30] In contrast, men with deletions in the AZFc region may have severe oligospermia while 65% to 75% of those with azoospermia produce sufficient sperm for testicular sperm extraction and use in ICSI.[30]

KEY POINTS

- Sperm DNA fragmentation is not required in the initial evaluation of male infertility.
- Couples with recurrent miscarriage or failed IVF may benefit from sperm DNA fragmentation testing, as it could identify the cause of their infertility problem.

Sperm DNA Fragmentation

The standard semen analysis has a limited ability to predict male fertility, except in azoospermia or specific, uniform abnormalities

TABLE 113.5 **Common and Known Genetic Abnormalities in Male Infertility[86–88] (see Chapter112)**

Genetic Abnormality	Phenotype	Frequency
Chromosomal Abnormality		
• Klinefelter syndrome (47,XXY)	• NOA, sperm retrieval rate up to 65% at micro-TESE • Hypogonadism	1:500 of newborn males
• 47,XYY	• Often phenotypically normal with variable sperm counts (normozoospermia to azoospermia)	1:1000 of newborn males
• 46,XX male	• Azoospermia with Sertoli–cell-only testicular histopathology	1:20,000 of newborn males
• Chromosomal translocations/ inversions	• Autosomal inversions can result in phenotypically normal males, but with oligospermia or azoospermia • Robertsonian translocations can cause recurrent miscarriage and/or offspring with genetic abnormality	1:600 to 1:1000 of newborn males
YCMD		10% to 15% of men with severe oligospermia or NOA
AZFa	• Azoospermia and Sertoli–cell-only pattern on testicular histopathology • Sperm retrieval rate 0% with micro-TESE	2% to 5% of YCMD
AZFb	• Azoospermia and uniform maturation arrest pattern on testicular histopathology • Sperm retrieval rate 0% with micro-TESE	13% to 16% of YCMD
AZFc	• Severe oligospermia to azoospermia • Sperm retrieval rate of up to 70% with micro-TESE	60% to 75% of YCMD
Combination deletions	• Azoospermia with very poor sperm retrieval prognosis if AZFa or AZFb regions involved	10% to 14% of YCMD
Single Gene Mutations		
• CFTR gene mutations	• OA due to congenital bilateral absence of vas deferens • Clinical cystic fibrosis may also have chronic pulmonary obstruction/infection and pancreatic insufficiency	1:25 carriers, 1:2500 of live births
• Kallman syndrome (KAL1, FGFR1, and others)	• Hypogonadotropic hypogonadism • Range of testicular histopathology with the recovery of spermatogenesis following hormonal treatment	1:30,000 of live births
• AR gene mutations	• Varies depending on the severity of AR deficiency • Female phenotype and Sertoli–cell-only in AR-negative or low AR patients • Male phenotype with infertility for AR-positive patients • Intact spermatogenesis for abnormal AR patients	1:60,000 of live births
• INSL3-LGR8	• Cryptorchidism	4% to 5% of newborn males

AR, Androgen receptor; *NOA,* nonobstructive azoospermia; *OA,* obstructive azoospermia; *TESE,* testicular sperm extraction; *YCMD,* Y-chromosome microdeletions.

such as globozoospermia. Efforts to identify sperm characteristics that may provide unique insights into the functional contributions of male gametes led to the application of sperm DNA fragmentation testing. Sperm DNA damage is rarely seen in fertile males but can be found in 8% of infertile men with normal semen analyses and 17% of infertile men with abnormal semen analyses.[31] Causes of sperm DNA damage may include intrinsic causes such as abnormalities in DNA compaction and protamine deficiency or extrinsic causes such as heat, gonadotoxins, varicocele, smoking, and radiation. Clinical observations have elucidated that sperm DNA damage occurs during sperm transport after sperm have left the testis. These observations include the increased sperm DNA damage in men with defects in sperm transport, including men with ejaculatory dysfunction associated with spinal cord injury or other abnormalities, increased sperm DNA damage after ineffective microsurgical reconstruction of the male reproductive tract, and for men who have received selective serotonin reuptake inhibitors (SSRIs). Several assays have been used to measure sperm DNA fragmentation, including direct assays like Comet gel electrophoresis and terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate-nick end labeling (TUNEL) and indirect assays such as sperm chromatin structure assay (SCSA) and sperm chromatin dispersion (SCD). Although studies on male contributions to fertility are subject to substantial heterogeneity and bias because of the confounding effects of important female factors, existing evidence supports a substantial role for sperm DNA fragmentation in predicting IVF and ICSI outcomes, with elevated sperm DNA fragmentation associated with decreased assisted reproductive technique (ART) success rates for clinical pregnancy. Even more dramatic are the effects of abnormal sperm DNA fragmentation of miscarriage rates. Adverse effects of sperm DNA fragmentation are detectable for both IVF and ICSI.[32–34]

Other Genetic Anomalies

While thousands of genes have been implicated in affecting spermatogenesis, our understanding of their causative roles in male infertility is still in nascent stages.[35] However, advances in genetic evaluations including next-generation whole-exome or whole-genome sequencing, single nucleotide polymorphism (SNP) arrays, and array comparative genomic hybridization analysis improve our understanding of specific gene defects on male infertility by enabling the study of multiple genes in parallel. For example, investigation of asthenozoospermic men with primary ciliary dyskinesia has enabled the identification of new genetic abnormalities that impair the axoneme's structure and function (DNAI1, DNAI2, DNAAF1/LRRC50, DNAAF2/Ktu, LRRC6, ZMYND10, CCDC40).[36]

Other genes known to cause male infertility include the X-linked KAL1 and the autosomal FGFR1, which both result in idiopathic hypogonadotropic hypogonadism (HH), or Kallman syndrome when also associated with anosmia.[37] Genes causing teratozoospermia include Aurora kinase c, which has been found to cause macrozoospermia particularly in North African men, and SPATA16, PICK1, and DPY19L2, which have been shown to cause globoozospermia.[38,39]

In addition to the aforementioned single gene mutations, promising work is occurring in the field of epigenetics or the regulation of gene expression without modification of the gene sequence. Investigation of various epigenetic processes such as DNA methylation, post-translational histone modification, chromatin remodeling, and micro-RNAs may help unlock the etiology behind many cases of male infertility currently classified as idiopathic and identify new diagnostic and therapeutic strategies for the management of male infertility.

Azoospermia

KEY POINTS

- A critical distinction to make in the evaluation of azoospermia is obstructive versus nonobstructive azoospermia.
- The cause of azoospermia directs genetic evaluation and the prognosis for sperm retrieval and successful fertility treatment.

A complete lack of sperm in the ejaculate can be caused by either obstructive etiologies (i.e., ductal obstruction) or nonobstructive etiologies (i.e., testicular failure), which represents the critical distinction to make in the evaluation of men with azoospermia. Using many of the investigations outlined in Fig. 113.3 provides the information necessary to make the diagnosis that will drive the ultimate management of the azoospermic male. In rare equivocal cases where a testicular biopsy is needed to differentiate obstructive from non-OA, it should be performed in the operating room with the surgeon ready to proceed immediately either to reconstruction of the male reproductive tract or to sperm retrieval.

MANAGEMENT OF MALE INFERTILITY

Lifestyle Modifications

KEY POINTS

- Men should discontinue their use of anabolic steroids or exogenous testosterone.
- Potentially effective lifestyle modifications for male infertility include counseling on coital practices, avoidance of spermatotoxic lubricants, smoking cessation, obesity reduction, and decreased exposure to wet heat.
- Evidence to support other lifestyle interventions is very limited, so effective counseling on other lifestyle changes cannot be directed by current data.

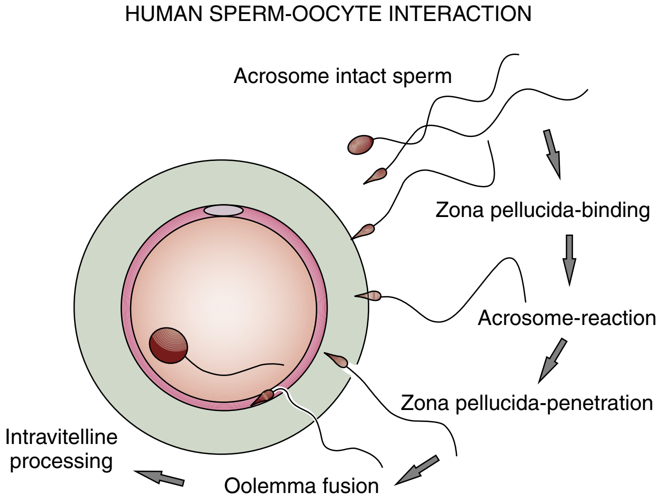

HUMAN SPERM-OOCYTE INTERACTION
Acrosome intact sperm
Zona pellucida-binding
Acrosome-reaction
Zona pellucida-penetration
Intravitelline processing
Oolemma fusion

Figure 113.3 Stages of Human Fertilization. Spermatozoa swim through the surrounding medium and cumulus mass *(not shown)* and bind to the surface of the zona pellucida. The acrosome reaction is stimulated by zona proteins, and the acrosome-reacted sperm penetrates the zona, enters the perivitelline space, and binds to the oolemma via the equatorial segment. Oocyte processes surround the sperm head, and it enters the ooplasm and decondenses. Infertility could result from defects in any of these processes. For example, abnormal sperm, particularly with defective head morphology, bind poorly to the zona.

Counseling on coital practices, including the appropriate frequency and timing of intercourse and the avoidance of potentially spermatotoxic lubricants, may be an important and productive intervention for many patients. If required, rather than using many of the commercially available lubricants, Pre-Seed, canola oil, and raw egg white appear to be sperm-friendly coital lubricants.[40,41]

Men using anabolic steroids or exogenous testosterone should be informed of the deleterious impact these medications have on spermatogenesis via inhibition of the pituitary-gonadal hormone axis and counseled to discontinue their use. Similar warnings should be given to men using other medications or recreational drugs known to hinder spermatogenesis (Table 113.2). Smoking, which is linked with decreased semen quality that results in decreased fecundity and lower success rates even with assisted reproduction, should be strongly discouraged.[42–44] Since obesity is associated with decreased semen parameters, weight loss should be encouraged.[45–47] Increased physical activity, which has been shown to improve semen parameters, but not necessarily clinical pregnancy or live birth rates following infertility treatment, should be encouraged.[48]

Routine exposure to wet heat via saunas, hot tubs, or hot baths should be discouraged as it can increase intratesticular temperature and transiently hinder spermatogenesis and sperm quality.[49,50] While an excellent form of aerobic exercise, bicycling, may negatively affect semen parameters due to the increased scrotal temperature associated with scrotal compression against the saddle.[51] Men who biked more than 1.5 hours/week had 34% lower sperm concentrations than those who did not bike.[48] Of note, given that the spermatogenic cycle is typically 60 to 76 days long, recovery from a spermatogenic insult may take up to 3 months to resolve.[52]

Hormone Optimization

KEY POINTS

- Men with low serum testosterone should be treated with medications that will not inhibit the hypothalamic–pituitary–gonadal axis such as selective estrogen receptor modulators or gonadotropins.
- Men with excess estrogen levels may benefit from treatment with aromatase inhibitors .

Treatment of diagnosed hormonal abnormalities to optimize sperm production should occur before more invasive strategies are undertaken. Treatable endocrinopathies include hyperprolactinemia, hypothyroidism, and congenital adrenal hyperplasia. Men with low serum testosterone (<300 ng/mL) should be managed with agents that will increase serum testosterone without suppressing the hypothalamic–pituitary–gonadal axis such as a selective estrogen receptor modulator (SERM) or human chorionic gonadotropin (hCG). A recent meta-analysis reviewing the use of SERM (clomiphene or tamoxifen) in infertile men with oligo and/or asthenoteratozoospermia showed an association between SERM use and increased pregnancy rates (odds ratio [OR] 2.42, $P = 0.0004$) with no significant differences in complication rates between the treated and control groups.[58] Men with HH may benefit from treatment with both hCG to induce intratesticular testosterone production and FSH to stimulate spermatogenesis.[59–62] Another treatment option for men with HH and a functioning pituitary gland is with pulsatile gonadotropin-releasing hormone (GnRH). Patients with excess estrogen levels manifested by low serum testosterone (<300 ng/mL) and decreased testosterone/estradiol ratios (<10) should be treated with an aromatase inhibitor (such as anastrozole, 1 mg P.O. daily), which increases circulating testosterone and decreases estradiol to enhance intratesticular testosterone levels and improve spermatogenesis.[62–64]

Ejaculatory Dysfunction

KEY POINTS

- Retrograde ejaculation may be treated with sympathomimetic medications or with bladder harvest of sperm for use in ART.
- Men with anejaculation may be treated with penile vibratory stimulation or rectal probe electroejaculation.

Retrograde ejaculation diagnosed on postejaculatory urinalysis can be treated with sympathomimetic medications to promote closure of the bladder neck at ejaculation or with bladder harvest of sperm for use in ART. Anejaculation, or lack of ejaculation, can occur in men with pelvic nerve damage from diabetes mellitus, retroperitoneal surgery, multiple sclerosis, spinal cord injury, or psychosocial causes. The use of penile vibratory stimulation (PVS) or rectal probe electroejaculation (EEJ) can help to procure a semen sample containing sperm for use with ART. In particular, PVS, which often does not require general anesthesia, works best in men with complete spinal cord injuries above T10, with up to 86% of patients able to achieve ejaculation with PVS.[65,68] On the other hand, only 0% to 15% of men with spinal cord injuries below T10 were able to achieve ejaculation with PVS.[60,62] In men with anejaculation from other etiologies or after failure with PVS, EEJ is preferred and can produce ejaculation in 92% of men with anejaculation from spinal cord injury.[60] Of note, because both PVS and EEJ can cause autonomic dysreflexia in men with spinal cord injuries at or above T6, blood pressure monitoring and premedication with nifedipine should be considered.[63] If these techniques fail, then surgical sperm retrieval with ART can be employed to circumvent the inability to ejaculate.

Ejaculatory Duct Obstruction

KEY POINTS

- Azoospermic men with ejaculatory duct obstruction may be effectively managed with transurethral resection of the ejaculatory ducts or transurethral ejaculatory duct dilation.
- Another treatment for this condition is sperm retrieval and ART.

Ejaculatory duct obstruction is suspected in men with azoospermia, severe oligospermia, and/or asthenospermia with low semen volume, at least one palpable vas deferens, low semen pH, and negative semen fructose. Once confirmed, ejaculatory duct obstruction can be managed via transurethral resection of the ejaculatory ducts (TURED) or transurethral dilation of the ejaculatory ducts, with 65% of men having significant and durable improvements in semen quality (Fig. 113.4).[64] Potential complications of transurethral management of ejaculatory duct obstruction can occur in 20% to 26% of patients and may include epididymitis, urine reflux into the excurrent ducts, and retrograde ejaculation.[64,65] Given the high complication rates associated with TURED for the management of ejaculatory duct obstruction, it should be reserved for patients who are counseled of its risks and who are unwilling to undergo IVF.

Pyospermia

KEY POINTS

- Increased round cells on semen analysis should be evaluated further to determine whether germ cells or white blood cells (pyospermia) are present in semen.
- Pyospermia should be treated with antibiotic therapy only after culture confirmation of bacterial infection.

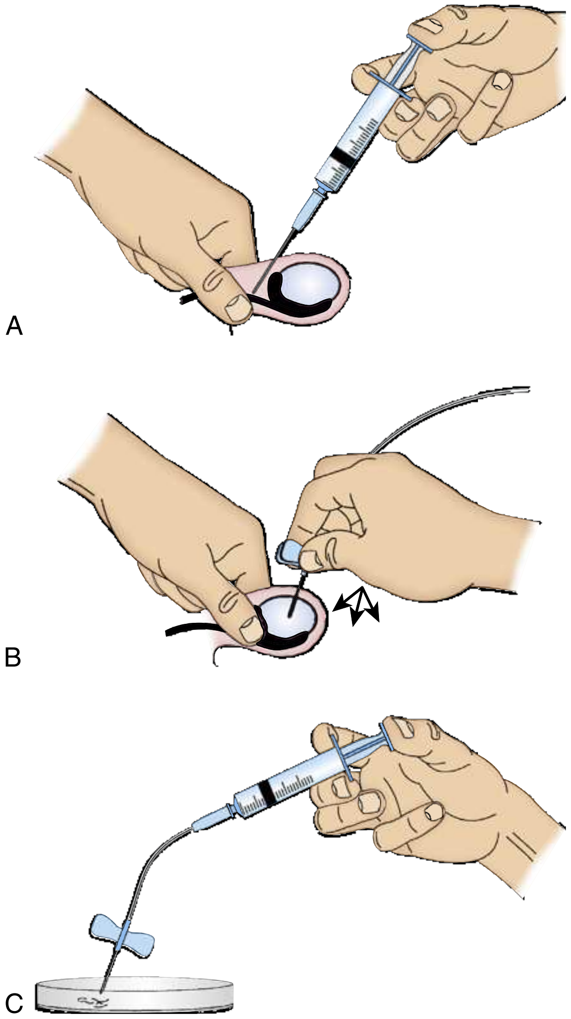


Figure 113.4 Fine-Needle Tissue Aspiration Biopsy of the Testis. A, Local anesthetic is injected around the vas to block testicular sensation. B, A 21-gauge butterfly needle is inserted into the testis. An assistant applies suction to the needle tubing via a syringe, and the operator makes thrusting movements of the needle into the substance of the testis. The appearance of some tissue fluid in the tubing (1 to 2 mm) usually indicates that adequate tissue has entered the needle. While maintaining the suction, the needle is removed carefully, and any seminiferous tubules protruding from the needle are grasped with fine forceps to avoid their falling back into the puncture hole. C, Seminiferous tubule sections sucked into the needle are expelled into culture medium. Portions can be sent for histology or used for extraction of sperm for intracytoplasmic sperm injection (ICSI) by stripping the seminiferous tissue out of the connective tissue membrane of the seminiferous tubule.

Pyospermia associated with a bacteriological infection should be managed with an appropriate course of culture-specific antibiotics. However, without concrete evidence of bacteriological infection, the management of pyospermia is controversial, though a limited course of a broad-spectrum antibiotic such as doxycycline or trimethoprim-sulfamethoxazole along with frequent ejaculation has been shown to decrease semen leukocyte concentration.[66,67]

Immunological Infertility

KEY POINTS

- The effects of antisperm antibodies are effectively managed using ICSI, but evidence suggests that they can also be managed with long-term steroid therapy.

Because the presence of ASA is often associated with obstruction of the genital tract, any lesions should be sought out and corrected. In men with persistent ASA, the management focuses on two broad strategies—immunosuppression or ART. Immunosuppression can be achieved using 6 to 9 months of corticosteroid therapy, which results in consistent reductions in antibody titers and in pregnancy rates ranging from 6% to 50%.[68] The potential complications of long-term steroid therapy are severe and must be accounted for prior to initiation of immunosuppression.[69] Semen can also either be processed in the laboratory in an attempt to remove ASA prior to intrauterine insemination (IUI) or injected directly into an oocyte with ICSI. Due to the potentially significant complications of long-term steroid therapy and the advent of ART, only patients with an objection to IVF should be offered steroid therapy.

Varicocele Repair

KEY POINTS

- Couples with varicocele-associated infertility should be offered varicocele repair when the varicocele is palpable; abnormal semen parameters are documented and no female factors or treatable female factors are present.
- There is no benefit to repairing subclinical varicoceles.
- Microsurgical inguinal or subinguinal repair is the technique of choice for varicocele repair.
- Right varicoceles that are longstanding or decompress in the supine position should not be evaluated further with abdominal imaging.

The decision on whether or not to repair clinical varicocele in the subfertile male remains controversial, though growing evidence suggests a beneficial role for varicocele repair in this population.[12] While meta-analyses trying to answer the question of the impact of varicocele repair on male fertility have been fraught with heterogeneity that make it difficult to draw definitive conclusions, a recent Cochrane subgroup metaanalysis that excluded studies of men with normal semen analyses and/or subclinical varicoceles suggests an association between varicocele repair and improved pregnancy rates.[70] In particular, the evidence suggests that 30% to 50% of infertile couples who are candidates for only ART due to low semen quality may improve to the point of being able to downstage the level of ART needed to conceive or to even enable natural conception following varicocele repair.[71] Currently, the American Society for Reproductive Medicine guidelines suggest varicocele repair for two patient populations: adolescents with reduced ipsilateral testicular size and the infertile couple in which the man has a clinical varicocele, an abnormal semen analysis, and a partner with normal or correctable fertility.[72] On the other hand, the question of whether to repair subclinical varicoceles is less controversial, with the evidence suggesting a minimal benefit to the repair of these varicoceles.[73]

There exist many techniques for varicocele repair, including microsurgical subinguinal or inguinal, open subinguinal or inguinal, retroperitoneal, laparoscopic, percutaneous radiographic embolization, and sclerotherapy. While the success rates between the various techniques are similar, there are differences in the rates of complications such as hydrocele and recurrence (Table 113.6). A metaanalysis from

TABLE 113.6 Summary of Varicocele Repair Techniques[89,90]

Technique	Hydrocele Formation (%)	Recurrence Rate (%)
Open retroperitoneal	6.4 to 10	7 to 35
Open conventional inguinal	7.3	0 to 37
Laparoscopic retroperitoneal	0 to 9.4	2.2 to 6.1
Microsurgical inguinal/subinguinal	0.2 to 1.6	0 to 3.0
Radiographic	0	2 to 24

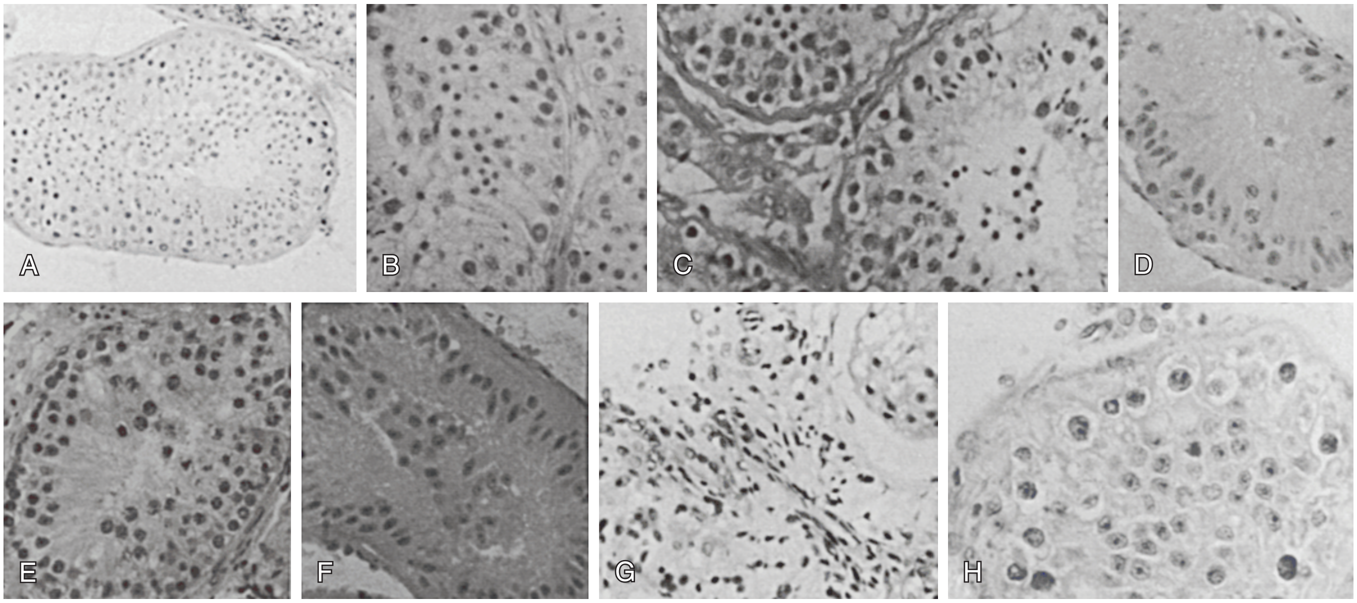


Figure 113.5 Testicular Histology From Fine-Needle Aspiration Samples. **A,** Normal (×200); **B,** mild hypospermatogenesis (×500); **C,** moderate hypospermatogenesis; **D,** severe hypospermatogenesis; **E,** germ cell arrest at the primary spermatocyte stage; **F,** Sertoli-cell–only syndrome; **G,** Leydig-cell tumor in an undescended testis with hyalinized Sertoli-cell–only tubules; H, carcinoma in situ, only transformed spermatogonia and Sertoli cells present.

2009 suggested that microsurgical varicocele repair (Fig. 113.5) was the gold standard for varicocele repair with the lowest rate of both hydrocele (0.4%) and recurrence (1%).[74]

Vasal Reconstruction

KEY POINTS

- Obstructive azoospermia may be treated with vasal reconstruction with either vasovasostomy or vasoepididymostomy as well as sperm retrieval with assisted reproductive technologies.

The most common etiology for OA is vasectomy, though it can also result from iatrogenic injuries (especially after inguinal hernia repair) and idiopathic causes. Of the 7% of American men who report having had a vasectomy, 20% also report a desire for future children.[75]

At the time of vasal reconstruction, if distal and proximal patency is confirmed, then a vasovasostomy is performed, in which the two cut ends of the vas deferens are anastomosed (Fig. 113.6). Seventy to ninety-seven percent of men are likely to have a return of sperm to the ejaculate following vasovasostomy with pregnancy rates ranging from 38% to 93%.[76,77] However, if there is suspicion of secondary epididymal obstruction, then a vasoepididymostomy is performed instead, with the abdominal vas deferens anastomosed to the epididymis (Fig. 113.7). Thirty to ninety percent of men will have a return of sperm to the ejaculate following vasoepididymostomy with subsequent pregnancy rates ranging from 20% to 50%.[78] Due to the delicate nature of both surgeries, almost all vasal reconstructions are performed microsurgically.

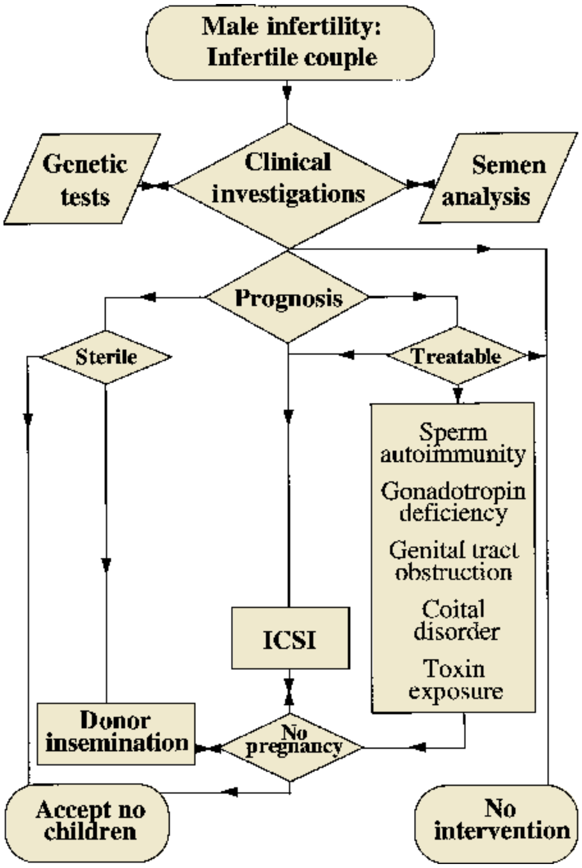

Male infertility: Infertile couple
Genetic tests
Clinical investigations
Semen analysis
Prognosis
Sterile
Treatable
Sperm autoimmunity
Gonadotropin deficiency
Genital tract obstruction
Coital disorder
Toxin exposure
ICSI
Donor insemination
No pregnancy
Accept no children
No intervention

Figure 113.6 Flowchart for the Management of Male Infertility.

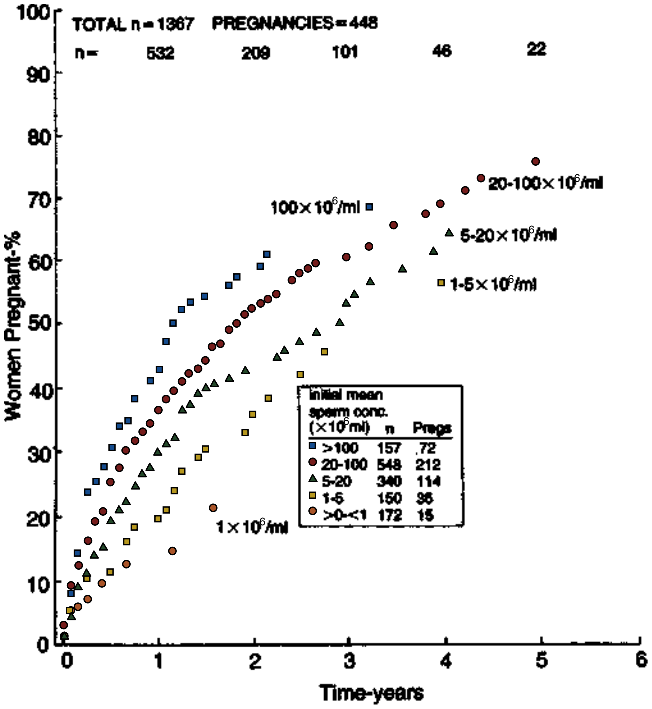


Figure 113.7 **Pregnancy Rate Curves Grouped According to Average Pretreatment Sperm Concentration.** The number of patients (n) followed each year is shown. The numbers of men and pregnancies in each sperm concentration group are shown in the inset.

Assisted Reproductive Techniques (ART)

KEY POINTS

- ART are effective treatments in the management of the infertile male that should be applied when no specific treatments to correct the male factor are available.
- Different types of ART are recommended for men with differing severities of infertility: IUI for those with moderate infertility, IVF for severe infertility, and ICSI for extreme infertility.
- Sperm can be retrieved from men with non–obstructive azoospermia using microdissection testicular sperm extraction.

For men in whom the aforementioned lifestyle, medical, and surgical strategies fail, ART is the final step. ART can be subdivided into IUI, IVF, and ICSI. From the male perspective, the decision between which ART strategy to use depends primarily on the level of impairment as defined by the number of sperm that can be retrieved (Fig. 113.8). Men with total motile sperm concentrations above 5 million/mL in the ejaculate, those with immunological infertility, and those with anatomic problems precluding sperm delivery (such as hypospadias) are candidates for IUI. Men with a total motile sperm concentration of 500,000 to 5 million/mL either surgically retrieved or in the ejaculate are suitable candidates for IVF while those with total motile sperm concentrations below 500,000/mL should consider ICSI, which theoretically requires only one single viable sperm to produce a pregnancy.

For men with iatrogenic OA, vasal reconstruction has been found to be a more cost-effective management strategy than ART, with the value proposition more dependent on female age and fertility potential than on obstructive interval.[79,80] Ultimately, the decision on whether to pursue vasal reconstruction or ART in men postvasectomy will vary with each couple and rely on factors such as female age/fertility potential, desired number of children, and willingness to undergo IVF.

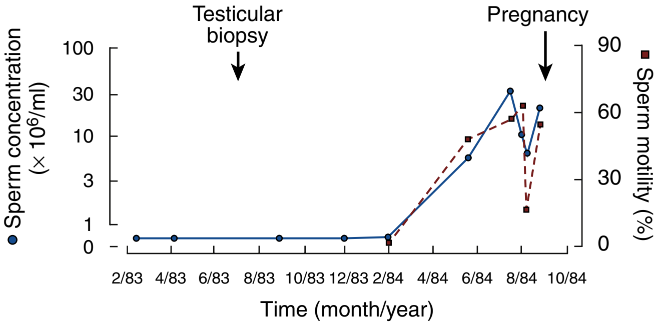


Figure 113.8 **Sperm Concentration and Motility in a Man With Severe Oligospermia and Severe Hypospermatogenesis Included in a Therapeutic Trial of Clomiphene.** Semen quality improved, and his wife conceived. He was given the placebo! (Data from Baker HW: Requirements for controlled therapeutic trials in male infertility, Clin Reprod Fertil 4:13-25, 1986.)

In short, couples in whom the maternal fertility potential is less than 1 year should favor ART while those in whom the maternal fertility potential is greater than 1 year and/or who desire multiple children should favor vasal reconstruction.

Sperm Retrieval

The goal of sperm retrieval is to obtain a maximal amount of viable sperm for use in ART while minimizing the amount of damage to the reproductive tract. Different techniques are available for use in men with OA and for those with NOA (Table 113.7). In men with OA, sperm can be successfully retrieved in almost all patients from the vas deferens via vasal aspirate,[96] from the epididymis via either percutaneous epididymal sperm aspiration (PESA)[97] or microsurgical epididymal sperm aspiration (MESA),[98] and/or from the testicle via either testicular sperm aspiration (TESA) or testicular sperm extraction (TESE).

On the other hand, men with NOA present a much greater challenge, with traditional sperm retrieval rates of only 25% to 50%. Because of the poor sperm production in men with NOA, sperm can often be retrieved only from the testicle. Variations of TESE that have been developed to improve the success rate of sperm retrieval for use in ICSI in men with NOA include microdissection TESE (micro-TESE), in which an operating microscope is used to identify and remove fuller seminiferous tubules within the testicle that are more likely to harbor spermatogenesis (Fig. 113.9).[81] Micro-TESE, which results in the removal of much less testicular tissue than TESE, has resulted in both an improvement in sperm retrieval rates (Table 113.8) and a decrease in complication rates when compared with TESE.[82,99] Other strategies that have been developed include fine-needle aspiration (FNA) mapping, in which multiple diagnostic fine needle aspirations are systematically taken in a bid to create a map of spermatogenesis locations within the testis that can be used to direct future TESA or TESE at the time of ICSI.[83]

CONCLUSION

With male factor contributing to 50% of infertility, the appropriate and concurrent management of the male is critical. While the evaluation of the infertile male aims to identify correctable causes of infertility that can guide therapeutic decision making for the infertile couple, it can also detect significant medical problems related to infertility. The management of male infertility involves a combination of lifestyle, medical, and surgical strategies, but with the advent of ICSI and surgical sperm retrieval techniques, even the most severe male factor infertility can often be successfully overcome.

TABLE 113.7 Summary of Sperm Retrieval Techniques for Men With OA[91–93]

Procedure	Anesthesia	Sperm Retrieval Rate	Advantages	Disadvantages
Vasal aspiration	General/regional	100%	Most mature sperm	Requires delicate microsurgical reconstruction of vasotomy
PESA	Local	80% to 100%	Office procedure Minimal recovery	Hematoma May not retrieve adequate sperm Can induce epididymal obstruction
MESA	General/regional	95% to 100%	High likelihood of obtaining adequate sperm May perform simultaneous reconstruction	Requires operating room and microsurgical skills
TESA	Local	52% to 100%	Office procedure Minimal recovery	Hematoma May not retrieve adequate sperm
TESE	General/regional	100%	Fast procedure No microsurgical skills required	Risk of testicular atrophy and hypofunction Requires operating room More invasive and longer recovery

MESA, Microsurgical epididymal sperm aspiration; *OA,* obstructive azoospermia; *PESA,* percutaneous epididymal sperm aspiration; *TESA,* testicular sperm aspiration; *TESE,* testicular sperm extraction.

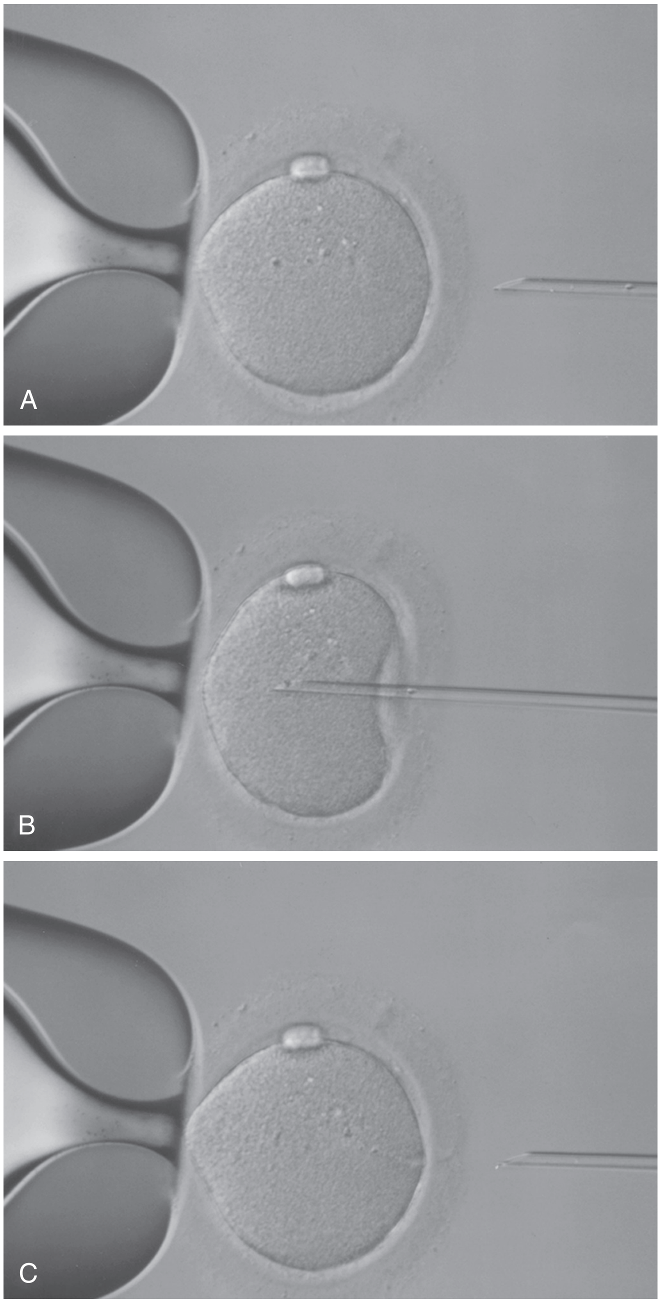


Figure 113.9 Intracytoplasmic Sperm Injection.

TABLE 113.8 Micro-TESE Sperm Retrieval and Pregnancy Rates for Various Azoospermia Patient Cohorts[94]

Patient Condition	Sperm Retrieval Rate	Pregnancy Rate (%)
AZFc deletion	72% (39/54)	46
Klinefelter syndrome	65% (100/155)	40
Cryptorchidism	64% (116/181)	50
Postchemotherapy	48% (55/114)	40
All patients	56% (794/1414)	48

TESE, Testicular sperm extraction.

REFERENCES

1. Martinez G, Daniels K, Chandra A. Fertility of men and women aged 15-44 years in the United States: national survey of family growth, 2006-2010. *Natl Health Stat Report*. 2012;(51):1–28.
2. Thonneau P, Marchand S, Tallec A, et al. Incidence and main causes of infertility in a resident population (1,850,000) of three French regions (1988-1989). *Hum Reprod*. 1991;6(6):811–816.
3. Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol*. 2015;13(1):37–45.
4. Hanson HA, Anderson RE, Aston KI, et al. Subfertility increases risk of testicular cancer: evidence from population-based semen samples. *Fertil Steril*. 2016;105(2):322–328. e1.
5. Al-Jebari Y, Elenkov A, Wirestrand E, et al. Risk of prostate cancer for men fathering through assisted reproduction: nationwide population based register study. *BMJ*. 2019;366:l5214.
6. Eisenberg ML, Betts P, Herder D, Lamb DJ, Lipshultz LI. Increased risk of cancer among azoospermic men. *Fertil Steril*. 2013;100(3):681–685.
7. Kasman AM, Li S, Luke B, et al. Male infertility and future cardiometabolic health: does the association vary by sociodemographic factors? *Urology*. 2019;133:121–128.
8. Glazer CH, Bonde JP, Eisenberg ML, et al. Male infertility and risk of nonmalignant chronic diseases: a systematic review of the epidemiological evidence. *Semin Reprod Med*. 2017;35(3):282–290.
9. Schlegel PN, Sigman M, Collura B, et al. Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. *J Urol*. 2021;205(1):36–43.
10. Lipshultz LI, Corriere Jr JN. Progressive testicular atrophy in the varicocele patient. *J Urol*. 1977;117(2):175–176.
11. Garcia-Peiro A, Ribas-Maynou J, Oliver-Bonet M, et al. Multiple determinations of sperm DNA fragmentation show that varicocelectomy is not indicated for infertile patients with subclinical varicocele. *BioMed Res Int*. 2014;2014:181396.
12. Will MA, Swain J, Fode M, Sonksen J, Christman GM, Ohl D. The great debate: varicocele treatment and impact on fertility. *Fertil Steril*. 2011;95(3):841–852.

13. Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med*. 2001;345(19):1388–1393.
14. Agarwal A, Gupta S, Du Plessis S, et al. Abstinence time and its impact on basic and advanced semen parameters. *Urology*. 2016;94:102–110.
15. Cooper TG, Noonan E, von Eckardstein S, et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*. 2010;16(3):231–245.
16. Buck Louis GM, Sundaram R, Schisterman EF, et al. Semen quality and time to pregnancy: the longitudinal investigation of fertility and the environment study. *Fertil Steril*. 2014;101(2):453–462.
17. Sigman M, Jarow JP. Endocrine evaluation of infertile men. *Urology*. 1997;50(5):659–664.
18. Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab*. 2010;95(6):2536–2559.
19. Yang Q, Huang YP, Wang HX, et al. Follicle-stimulating hormone as a predictor for sperm retrieval rate in patients with nonobstructive azoospermia: a systematic review and meta-analysis. *Asian J Androl*. 2015;17(2):281–284.
20. Mehta A, Jarow JP, Maples P, Sigman M. Defining the "normal" postejaculate urinalysis. *J Androl*. 2012;33(5):917–920.
21. Purohit RS, Wu DS, Shinohara K, Turek PJ. A prospective comparison of 3 diagnostic methods to evaluate ejaculatory duct obstruction. *J Urol*. 2004;171(1):232–235; discussion 235-6.
22. Jarow JP. Transrectal ultrasonography of infertile men. *Fertil Steril*. 1993;60(6):1035–1039.
23. Engin G. Transrectal US-guided seminal vesicle aspiration in the diagnosis of partial ejaculatory duct obstruction. *Diagn Interv Radiol*. 2012;18(5):488–495.
24. Schlegel PN, Shin D, Goldstein M. Urogenital anomalies in men with congenital absence of the vas deferens. *J Urol*. 1996;155(5):1644–1648.
25. Lee R, Goldstein M, Ullery BW, et al. Value of serum antisperm antibodies in diagnosing obstructive azoospermia. *J Urol*. 2009;181(1):264–269.
26. JU S, Schwarz M. Significance of CFTR gene mutations in patients with congenital aplasia of vas deferens with special regard to renal aplasia. *Andrologia*. 2012;44(5):305–307.
27. Wolff H, Anderson DJ. Immunohistologic characterization and quantitation of leukocyte subpopulations in human semen. *Fertil Steril*. 1988;49(3):497–504.
28. Eggert-Kruse W, Pohl S, Naher H, Tilgen W, Runnebaum B. Microbial colonization and sperm–mucus interaction: results in 1000 infertile couples. *Hum Reprod*. 1992;7(5):612–620.
29. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*. 5th ed. Geneva, Switzerland: WHO Press; 2010.
30. Tahmasbpour E, Balasubramanian D, Agarwal A. A multi-faceted approach to understanding male infertility: gene mutations, molecular defects and assisted reproductive techniques (ART). *J Assist Reprod Genet*. 2014;31(9):1115–1137.
31. Samli H, Samli MM, Solak M, Imirzalioglu N. Genetic anomalies detected in patients with non-obstructive azoospermia and oligozoospermia. *Arch Androl*. 2006;52(4):263–267.
32. Ravel C, Berthaut I, Bresson JL, Siffroi JP. Genetics Commission of the French Federation of CECOS. Prevalence of chromosomal abnormalities in phenotypically normal and fertile adult males: large-scale survey of over 10,000 sperm donor karyotypes. *Hum Reprod*. 2006;21(6):1484–1489.
33. Yatsenko AN, Yatsenko SA, Weedin JW, et al. Comprehensive 5-year study of cytogenetic aberrations in 668 infertile men. *J Urol*. 2010;183(4):1636–1642.
34. Liu XG, Hu HY, Guo YH, Sun YP. Correlation between Y chromosome microdeletion and male infertility. *Genet Mol Res*. 2016;15(2).
35. Hopps CV, Mielnik A, Goldstein M, et al. Detection of sperm in men with Y chromosome microdeletions of the AZFa, AZFb and AZFc regions. *Hum Reprod*. 2003;18(8):1660–1665.
36. Zini A, Fischer MA, Sharir S, et al. Prevalence of abnormal sperm DNA denaturation in fertile and infertile men. *Urology*. 2002;60(6):1069–1072.
37. Simon L, Zini A, Dyachenko A, Ciampi A, Carrell DT. A systematic review and meta-analysis to determine the effect of sperm DNA damage on in vitro fertilization and intracytoplasmic sperm injection outcome. *Asian J Androl*. 2017;19(1):80–90.
38. Osman A, Alsomait H, Seshadri S, El-Toukhy T, Khalaf Y. The effect of sperm DNA fragmentation on live birth rate after IVF or ICSI: a systematic review and meta-analysis. *Reprod Biomed Online*. 2015;30(2):120–127.
39. Zhao J, Zhang Q, Wang Y, Li Y. Whether sperm deoxyribonucleic acid fragmentation has an effect on pregnancy and miscarriage after in vitro fertilization/intracytoplasmic sperm injection: a systematic review and meta-analysis. *Fertil Steril*. 2014;102(4):998–1005. e8.
40. Aston KI. Genetic susceptibility to male infertility: news from genome-wide association studies. *Andrology*. 2014;2(3):315–321.
41. Praveen K, Davis EE, Katsanis N. Unique among ciliopathies: primary ciliary dyskinesia, a motile cilia disorder. *F1000Prime Rep*. 2015;7. https://doi.org/10.12703/P7-36.
42. Bhagavath B, Layman LC. The genetics of hypogonadotropic hypogonadism. *Semin Reprod Med*. 2007;25(4):272–286.
43. Eloualid A, Rouba H, Rhaissi H, et al. Prevalence of the aurora kinase C c.144delC mutation in infertile Moroccan men. *Fertil Steril*. 2014;101(4):1086–1090.
44. De Braekeleer M, Nguyen MH, Morel F, Perrin A. Genetic aspects of monomorphic teratozoospermia: a review. *J Assist Reprod Genet*. 2015;32(4):615–623.
45. Agarwal A, Deepinder F, Cocuzza M, Short RA, Evenson DP. Effect of vaginal lubricants on sperm motility and chromatin integrity: a prospective comparative study. *Fertil Steril*. 2008;89(2):375–379.
46. Sandhu RS, Wong TH, Kling CA, Chohan KR. In vitro effects of coital lubricants and synthetic and natural oils on sperm motility. *Fertil Steril*. 2014;101(4):941–944.
47. Sharma R, Harlev A, Agarwal A, Esteves SC. Cigarette smoking and semen quality: a new meta-analysis examining the effect of the 2010 World Health Organization laboratory methods for the examination of human semen. *Eur Urol*. 2017;70(4):635–645.
48. Yang F, Li L, Chen JP, et al. Couple's infertility in relation to male smoking in a Chinese rural area. *Asian J Androl*. 2017;19(3):311–315.
49. Firns S, Cruzat VF, Keane KN, et al. The effect of cigarette smoking, alcohol consumption and fruit and vegetable consumption on IVF outcomes: a review and presentation of original data. *Reprod Biol Endocrinol*. 2015;13:134.
50. Sermondade N, Faure C, Fezeu L, et al. BMI in relation to sperm count: an updated systematic review and collaborative meta-analysis. *Hum Reprod Update*. 2013;19(3):221–231.
51. Andersen JM, Herning H, Aschim EL, et al. Body mass index is associated with impaired semen characteristics and reduced levels of anti-Müllerian hormone across a wide weight range. *PLoS One*. 2015;10(6):e0130210.
52. Eisenberg ML, Kim S, Chen Z, et al. The relationship between male BMI and waist circumference on semen quality: data from the LIFE study. *Hum Reprod*. 2014;29(2):193–200.
53. Gaskins AJ, Afeiche MC, Hauser R, et al. Paternal physical and sedentary activities in relation to semen quality and reproductive outcomes among couples from a fertility center. *Hum Reprod*. 2014;29(11):2575–2582.
54. Rao M, Zhao XL, Yang J, et al. Effect of transient scrotal hyperthermia on sperm parameters, seminal plasma biochemical markers, and oxidative stress in men. *Asian J Androl*. 2015;17(4):668–675.
55. Rao M, Xia W, Yang J, et al. Transient scrotal hyperthermia affects human sperm DNA integrity, sperm apoptosis, and sperm protein expression. *Andrology*. 2016;4(6):1054–1063.
56. Hajizadeh Maleki B, Tartibian B. Long-term low-to-intensive cycling training: impact on semen parameters and seminal cytokines. *Clin J Sport Med*. 2015;25(6):535–540.
57. Misell LM, Holochwost D, Boban D, et al. A stable isotope-mass spectrometric method for measuring human spermatogenesis kinetics in vivo. *J Urol*. 2006;175(1):242–246. discussion 246.
58. Chua ME, Escusa KG, Luna S, et al. Revisiting oestrogen antagonists (clomiphene or tamoxifen) as medical empiric therapy for idiopathic male infertility: a meta-analysis. *Andrology*. 2013;1(5):749–757.

59. Zacharin M, Sabin MA, Nair VV, Dabadghao P. Addition of recombinant follicle-stimulating hormone to human chorionic gonadotropin treatment in adolescents and young adults with hypogonadotropic hypogonadism promotes normal testicular growth and may promote early spermatogenesis. *Fertil Steril*. 2012;98(4):836–842.
60. Sato N, Hasegawa T, Hasegawa Y, et al. Treatment situation of male hypogonadotropic hypogonadism in pediatrics and proposal of testosterone and gonadotropins replacement therapy protocols. *Clin Pediatr Endocrinol*. 2015;24(2):37–49.
61. Boehm U, Bouloux PM, Dattani MT, et al. Expert consensus document: European consensus statement on congenital hypogonadotropic hypogonadism--pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*. 2015;11(9):547–564.
62. Pavlovich CP, King P, Goldstein M, Schlegel PN. Evidence of a treatable endocrinopathy in infertile men. *J Urol*. 2001;165(3):837–841.
63. Raman JD, Schlegel PN. Aromatase inhibitors for male infertility. *J Urol*. 2002;167(2, Part 1):624–629.
64. Schlegel PN. Aromatase inhibitors for male infertility. *Fertil Steril*. 2012;98(6):1359–1362.
65. Brackett NL, Ibrahim E, Iremashvili V, Aballa TC, Lynne CM. Treatment for ejaculatory dysfunction in men with spinal cord injury: an 18-year single center experience. *J Urol*. 2010;183(6):2304–2308.
66. Castle SM, Jenkins LC, Ibrahim E, Aballa TC, Lynne CM, Brackett NL. Safety and efficacy of a new device for inducing ejaculation in men with spinal cord injuries. *Spinal Cord*. 2014;52(suppl 2):S27–S29.
67. Chehensse C, Bahrami S, Denys P, Clement P, Bernabe J, Giuliano F. The spinal control of ejaculation revisited: a systematic review and meta-analysis of anejaculation in spinal cord injured patients. *Hum Reprod Update*. 2013;19(5):507–526.
68. Trofimenko V, Hotaling JM. Fertility treatment in spinal cord injury and other neurologic disease. *Transl Androl Urol*. 2016;5(1):102–116.
69. Turek PJ, Magana JO, Lipshultz LI. Semen parameters before and after transurethral surgery for ejaculatory duct obstruction. *J Urol*. 1996;155(4):1291–1293.
70. El-Assmy A, El-Tholoth H, Abouelkheir RT, Abou-El-Ghar ME. Transurethral resection of ejaculatory duct in infertile men: outcome and predictors of success. *Int Urol Nephrol*. 2012;44(6):1623–1630.
71. Yanushpolsky EH, Politch JA, Hill JA, Anderson DJ. Antibiotic therapy and leukocytospermia: a prospective, randomized, controlled study. *Fertil Steril*. 1995;63(1):142–147.
72. Branigan EF, Muller CH. Efficacy of treatment and recurrence rate of leukocytospermia in infertile men with prostatitis. *Fertil Steril*. 1994;62(3):580–584.
73. Turek PJ, Lipshultz LI. Immunologic infertility. *Urol Clin North Am*. 1994;21(3):447–468.
74. McLachlan RI. Basis, diagnosis and treatment of immunological infertility in men. *J Reprod Immunol*. 2002;57(1–2):35–45.
75. Kroese AC, de Lange NM, Collins J, Evers JL. Surgery or embolization for varicoceles in subfertile men. *Cochrane Database Syst Rev*. 2012;10:CD000479.
76. Cayan S, Erdemir F, Ozbey I, et al. Can varicocelectomy significantly change the way couples use assisted reproductive technologies? *J Urol*. 2002;167(4):1749–1752.
77. Practice Committee of the American Society for Reproductive Medicine. Society for Male Reproduction and Urology. Report on varicoceleand infertility: a committee opinion. *Fertil Steril*. 2014;102(6):1556–1560.
78. Pastuszak AW, Wang R. Varicocele and testicular function. *Asian J Androl*. 2015;17(4):659–667.
79. Wang J, Xia S-J, Liu Z-H, et al. Inguinal and subinguinal micro-varicocelectomy, the optimal surgical management of varicocele: a meta-analysis. *Asian J Androl*. 2015;17(1):74–80.
80. Sharma V, Le BV, Sheth KR, et al. Vasectomy demographics and postvasectomy desire for future children: results from a contemporary national survey. *Fertil Steril*. 2013;99(7):1880–1885.
81. Belker AM, Thomas Jr AJ, Fuchs EF, Konnak JW, Sharlip ID. Results of 1,469 microsurgical vasectomy reversals by the vasovasostomy study group. *J Urol*. 1991;145(3):505–511.
82. Herrel LA, Goodman M, Goldstein M, Hsiao W. Outcomes of microsurgical vasovasostomy for vasectomy reversal: a meta-analysis and systematic review. *Urology*. 2015;85(4):819–825.
83. Wosnitzer M, Goldstein M, Hardy MP. Review of azoospermia. *Spermatogenesis*. 2014;4:e28218.
84. Pavlovich CP, Schlegel PN. Fertility options after vasectomy: a cost-effectiveness analysis. *Fertil Steril*. 1997;67(1):133–141.
85. Hsieh MH, Meng MV, Turek PJ. Markov modeling of vasectomy reversal and ART for infertility: how do obstructive interval and female partner age influence cost effectiveness? *Fertil Steril*. 2007;88(4):840–846.
86. Schlegel PN. Testicular sperm extraction: microdissection improves sperm yield with minimal tissue excision. *Hum Reprod*. 1999;14(1):131–135.
87. Donoso P, Tournaye H, Devroey P. Which is the best sperm retrieval technique for non-obstructive azoospermia? A systematic review. *Hum Reprod Update*. 2007;13(6):539–549.
88. Beliveau ME, Turek PJ. The value of testicular 'mapping' in men with non-obstructive azoospermia. *Asian J Androl*. 2011;13(2):225–230.
89. Honig SC, Lipshultz LI, Jarow J. Significant medical pathology uncovered by a comprehensive male infertility evaluation. *Fertil Steril*. 1994;62(5):1028–1034.
90. Walsh TJ, Schembri M, Turek PJ, et al. Increased risk of high-grade prostate cancer among infertile men. *Cancer*. 2010;116(9):2140–2147.
91. Song SH, Chiba K, Ramasamy R, Lamb DJ. Recent advances in the genetics of testicular failure. *Asian J Androl*. 2016;18(3):350–355.
92. Walsh TJ, Pera RR, Turek PJ. The genetics of male infertility. *Semin Reprod Med*. 2009;27(2):124–136.
93. Tahmasbpour E, Balasubramanian D, Agarwal A. A multi-faceted approach to understanding male infertility: gene mutations, molecular defects and assisted reproductive techniques (ART). *J Assist Reprod Genet*. 2014;31(9):1115–1137.
94. Çayan S, Shavakhabov S, Kadioglu A. Treatment of palpable varicocele review in infertile men: a meta-analysis to define the best technique. *J Androl*. 2009;30(1):33–40.
95. Levine LA, Fakouri BJ. Experience with vasal sperm aspiration. *J Urol*. 1998;159(5):1551–1553.
96. Qiu Y, Wang S, Yang D, Wang L. Percutaneous vasal sperm aspiration and intrauterine insemination for infertile males with anejaculation. *Fertil Steril*. 2003;79(3):618–620.
97. Bernie AM, Ramasamy R, Stember DS, Stahl PJ. Microsurgical epididymal sperm aspiration: indications, techniques and outcomes. *Asian J Androl*. 2013;15(1):40–43.
98. Dabaja AA, Schlegel PN. Microdissection testicular sperm extraction: an update. *Asian J Androl*. 2013;15(1):35–39.

BIBLIOGRAPHY

Practice Committee of the American Society for Reproductive Medicine. Diagnostic evaluation of the infertile male: a committee opinion. *Fertil Steril*. 2015;103(3):e18–e25.

Cooper TG, Noonan E, von Eckardstein S, et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*. 2010;16(3):231–245.

Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med*. 2001;345(19):1388–1393.

Hopps CV, Mielnik A, Goldstein M, Palermo GD, Rosenwaks Z, Schlegel PN. Detection of sperm in men with Y chromosome microdeletions of the AZFa, AZFb and AZFc regions. *Hum Reprod*. 2003;18(8):1660–1665.

Simon L, Zini A, Dyachenko A, Ciampi A, Carrell DT. A systematic review and meta-analysis to determine the effect of sperm DNA damage on in vitro fertilization and intracytoplasmic sperm injection outcome. *Asian J Androl*. 2017;19(1):80–90.

Firns S, Cruzat VF, Keane KN, et al. The effect of cigarette smoking, alcohol consumption and fruit and vegetable consumption on IVF outcomes: a review and presentation of original data. *Reprod Biol Endocrinol*. 2015;13:134.

Chua ME, Escusa KG, Luna S, Tapia LC, Dofitas B, Morales M. Revisiting oestrogen antagonists (clomiphene or tamoxifen) as medical empiric therapy for idiopathic male infertility: a meta-analysis. *Andrology*. 2013;1(5):749–757.

Schlegel PN. Aromatase inhibitors for male infertility. *Fertil Steril*. 2012;98(6):1359–1362.

Brackett NL, Ibrahim E, Iremashvili V, Aballa TC, Lynne CM. Treatment for ejaculatory dysfunction in men with spinal cord injury: an 18-year single center experience. *J Urol*. 2010;183(6):2304–2308.

Herrel LA, Goodman M, Goldstein M, Hsiao W. Outcomes of microsurgical vasovasostomy for vasectomy reversal: a meta-analysis and systematic review. *Urology*. 2015;85(4):819–825.

Donoso P, Tournaye H, Devroey P. Which is the best sperm retrieval technique for non-obstructive azoospermia? A systematic review. *Hum Reprod Update*. 2007;13(6):539–549.

Kroese AC, de Lange NM, Collins J, Evers JL. Surgery or embolization for varicoceles in subfertile men. *Cochrane Database Syst Rev*. 2012;10:CD000479.

Wang J, Xia S-, Liu Z-, et al. Inguinal and subinguinal micro-varicocelectomy, the optimal surgical management of varicocele: a meta-analysis. *Asian J Androl*. 2015;17(1):74–80.

114

Male Contraception

Christina Wang, Ronald S. Swerdloff, and Stephanie T. Page

OUTLINE

The Need for Male Contraception, 1907
Current Methods, 1907
  Withdrawal, 1907
  Male Condom, 1907
  Vas Occlusion, 1908
Male-Directed Contraceptive Methods in Development, 1909
  Hormonal Male Contraception, 1909
  Other New Hormonal Agents for Male Contraception, 1912
  Nonhormonal Targets for Male Contraception, 1912
  Men's Attitude and Acceptability of Male Contraception, 1914
Summary and Future Directions, 1914

THE NEED FOR MALE CONTRACEPTION

Globally, the estimated unintended pregnancy rate between the years 1990 and 2014 was 44%. During this same period, a decline in unintended pregnancies of 30% in developed countries and 16% in developing countries was observed.[1,2] Over 50% of unintended pregnancies ended in abortions.[3] The decline in unintended pregnancy is attributable to improvements in reproductive and sexual health services and family planning programs. In addition to the mental, physical, and social impact of unintended pregnancies, the medical costs in the United States are estimated to be $4.5 to $11.1 billion; this cost could be substantially reduced by adherence to modern contraceptive methods.[4,5]

The female partner has traditionally accepted family planning responsibilities, though surveys indicate that both partners make contraceptive choices. Despite the many methods available to women, the failure rate of many types of reversible contraception remains high (Fig. 114.1).[6–9] Effective female hormonal methods include implants and intrauterine devices with typical use failure of less than 1%, whereas higher but acceptable typical use failure rates are observed with injections (6%), daily oral pill, patch, or vaginal rings (9%). Timing-dependent methods such as the female condom, spermicides, sponge, and fertility awareness methods have high failure rates of over 20%. More importantly the adherence rate of female hormonal methods after 12 months is only 50% to 70% with lower adherence in young women and higher adherence with long-acting reversible contraception compared with short-acting methods.[10,11] The adherence rate does not appear to improve with intensive counseling.[12]

In contrast to female methods of contraception, male-controlled methods include coital dependent, reversible methods of withdrawal and condom, and the irreversible method of vas occlusion. Thus there is a need for increasing the options for male contraception to meet the needs of men at different stages of their sexual and reproductive life and to provide them with choices that are not timing dependent yet reversible. The attitudes of men are changing with more men willing to share responsibilities in family planning.[13] According to a pharmacoeconomics study, about 50% of men between the ages of 15 and 64 years may wish to use a male contraceptive, which is equivalent to a potential market of about 44 million men.[14] Thus this is the time to develop and introduce new male contraceptives.[15]

KEY POINTS

- Availability and use of modern family planning methods is a key determinant of unintended pregnancies.
- Many men are willing to share responsibility in family planning; however, few modern methods are available to them.

CURRENT METHODS

In addition to sexual abstinence, the currently available contraceptive methods for men include withdrawal and condom with high typical use contraceptive failure rates of 22% and 18%, respectively (Fig. 114.1). Though these methods are reversible, they are dependent on the male to correctly and predictably use the method. The alternative available method is vas occlusion, which is considered irreversible because pregnancy rates after vasectomy reversal are low despite patency of the reanastomosis.

Withdrawal

Coitus interruptus or withdrawal requires the user's self-control as the ejaculate with spermatozoa will enter the vagina if withdrawal is not timed before ejaculation. The female partner also must trust that the male can and will withdraw his penis from the vagina before ejaculation. Few studies have been performed on withdrawal.[16] The failure rate is high (22%),[9] but a recent survey study in 43 countries estimated a lower failure rate of 13.4%.[17] Because of the high failure rate, family planning providers should provide emergency contraception for the female partner for couples relying on withdrawal.

Male Condom

The male condom is the most frequently used male method globally at 7.7%. Percentage varies among regions with the highest rate in Europe and lowest in Africa.[13] In the United States, the prevalence of condom use in men aged 15 to 44 years is 33.7%, but only about one-third use condoms all the time.[18] Condoms when used appropriately provide a barrier between the penis, the seminal fluid containing spermatozoa, and the vagina. The added value of condoms is prevention of sexually transmitted infections including the human immunodeficiency virus,

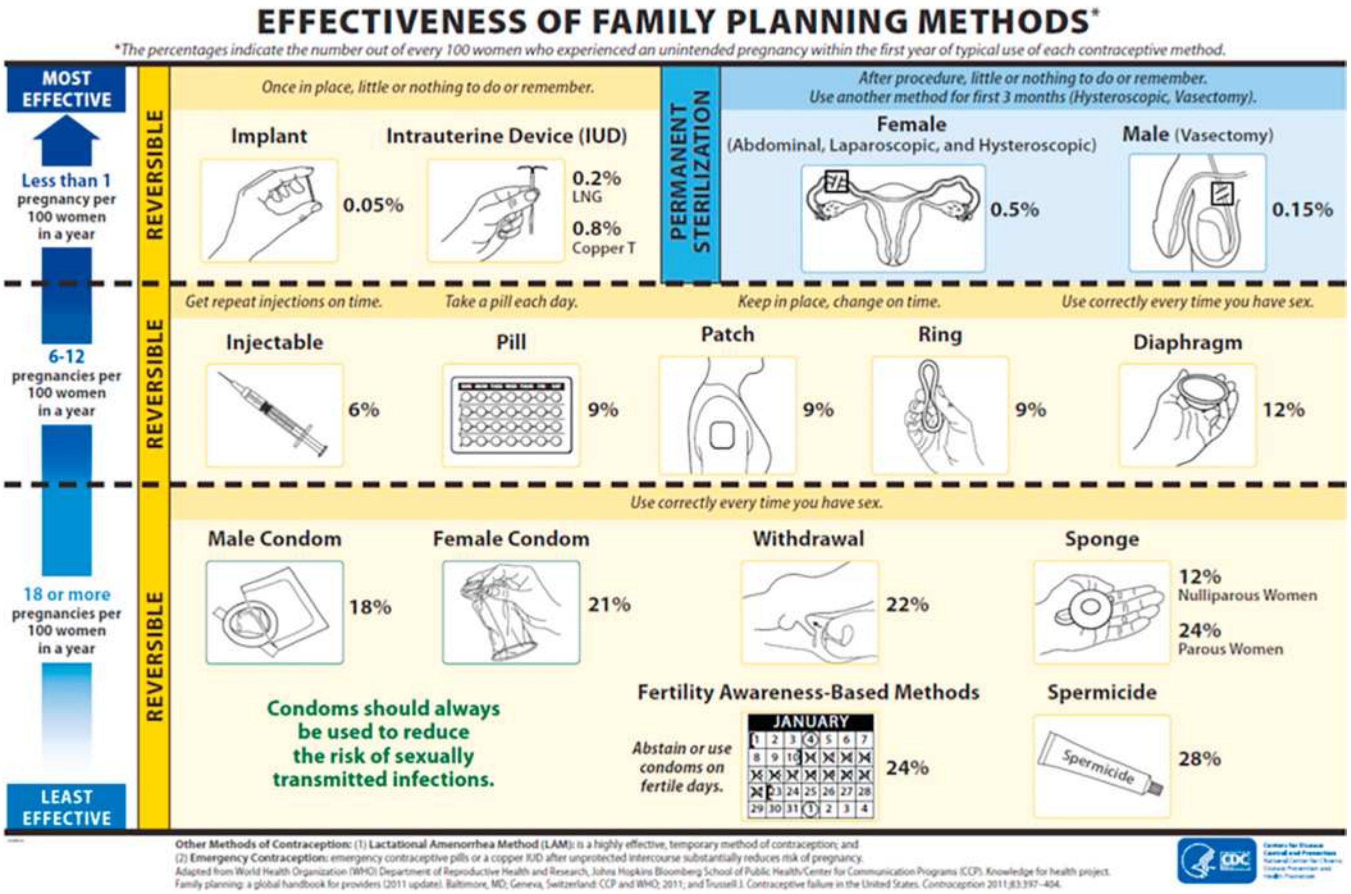

EFFECTIVENESS OF FAMILY PLANNING METHODS*
*The percentages indicate the number out of every 100 women who experienced an unintended pregnancy within the first year of typical use of each contraceptive method.
MOST EFFECTIVE
Less than 1 pregnancy per 100 women in a year
6-12 pregnancies per 100 women in a year
18 or more pregnancies per 100 women in a year
LEAST EFFECTIVE
REVERSIBLE
Once in place, little or nothing to do or remember.
Implant
0.05%
Intrauterine Device (IUD)
0.2% LNG
0.8% Copper T
PERMANENT STERILIZATION
After procedure, little or nothing to do or remember.
Use another method for first 3 months (Hysteroscopic, Vasectomy).
Female (Abdominal, Laparoscopic, and Hysteroscopic)
0.5%
Male (Vasectomy)
0.15%
Get repeat injections on time.
Take a pill each day.
Keep in place, change on time.
Use correctly every time you have sex.
Injectable
6%
Pill
9%
Patch
9%
Ring
9%
Diaphragm
12%
Use correctly every time you have sex.
Male Condom
18%
Female Condom
21%
Withdrawal
22%
Sponge
12% Nulliparous Women
24% Parous Women
Condoms should always be used to reduce the risk of sexually transmitted infections.
Fertility Awareness-Based Methods
Abstain or use condoms on fertile days.
JANUARY
24%
Spermicide
28%
Other Methods of Contraception: (1) Lactational Amenorrhea Method (LAM): is a highly effective, temporary method of contraception; and
(2) Emergency Contraception: emergency contraceptive pills or a copper IUD after unprotected intercourse substantially reduces risk of pregnancy.
Adapted from World Health Organization (WHO) Department of Reproductive Health and Research, Johns Hopkins Bloomberg School of Public Health/Center for Communication Programs (CCP). Knowledge for health project. Family planning: a global handbook for providers (2011 update). Baltimore, MD; Geneva, Switzerland: CCP and WHO; 2011; and Trussell J. Contraceptive failure in the United States. Contraception 2011;83:397–404.
CDC

Figure 114.1 Effectiveness of family planning methods. This figure shows the most effective currently available method from the top to the least effective at the bottom. The percentage is the Pearl index that is the number of unintended pregnancies per 100 women within the first year of typical use of the method. Male method includes vasectomy with a failure rate of 0.15% to condom of 18% and withdrawal of 22%. (From *Centers for Disease Control and Prevention,* How effective are birth control methods? https://www.cdc.gov/reproductivehealth/unintendedpregnancy/pdf/family-planning-methods-2014.pdf.)

herpes simplex virus, gonorrhea, chlamydia, and syphilis.[19] Thus condoms are frequently used by single men during casual sexual encounters.

The success of condom use depends on the correct application of the condom on the erect penis and removal of the condom after intercourse.[20] It is important that family planning providers counsel the user on the correct application and disposal of the condom.[21] The typical use failure rate of condom is estimated to be 18%,[9] but recent evidence suggests that the failure rate is lower at 12.6%.[22] The common causes of contraceptive failure of condoms reported by women seeking abortions are inconsistent use and breakage or slippage of the condom.[23] Condoms are usually made of latex and nonlatex condoms are made of polyurethane for individuals who are sensitive to latex. Compared with latex condoms, polyurethane condoms have higher breakage rates.[24] Natural membrane condoms are not recommended because of the high breakage rates.

KEY POINTS

- Current reversible male contraceptive methods are timing dependent and have a high typical use failure rate.
- Withdrawal during coitus requires the male user's self-control and understanding.
- Condom success in pregnancy prevention depends on the correct application at the right time.

Vas Occlusion

Vas occlusion shows a large difference in acceptability by couples from different geographical areas; higher prevalence (12% to 19%) has been reported from Australia, Bhutan, Canada, The Netherlands, New Zealand, Korea, United Kingdom, and United States, compared with very low use in Africa (0.1%), Central and South America (2%).[25] The number of couples relying on vasectomy as a contraceptive method has remained relatively constant since the 1990s. Vasectomy prevalence is about 6.6% in men aged between 18 and 45 years in the United States and is higher in Whites, older age, nonimmigrants, and those with higher income and education.[26] There has been a suggested link between vasectomy and an increased risk of prostate cancer but this has been refuted by epidemiological studies.[27,28]

The current recommended method of vas occlusion is performed under local anesthesia, using a "no-scalpel vasectomy" or other minimally invasive techniques.[29] The ends of the vas should be closed with cautery with fascial interposition or by ligature or clips. The success of vas occlusion is determined by postvasectomy semen analyses. The couple should use another method of contraception until semen samples showed azoospermia or rare nonmotile sperm concentration (<100,000 nonmotile sperm). It takes from 8 to 16 weeks to reach this very low nonmotile sperm concentration in the ejaculate. The failure rate of vasectomy is rare (<1%).[30] If after 6 months, azoospermia or near azoospermia is not reached, repeat vasectomy should be considered. With minimally invasive surgery, postoperative complications are very few, about 1% to 2% including hematoma and infection. Few men (1% to 2%) developed chronic scrotal pain after vasectomy that rarely impairs quality of life.[30]

Men are generally satisfied with vas occlusion and accept this as a permanent method of contraception. Counseling men and the couple before vasectomy increases postvasectomy satisfaction.[31] Reversibility of vasectomy can be achieved by vasovasostomy or vasoepididymostomy. Although the patency after surgery is about 87%, the pregnancy rate in the partner is only about 49%. Factors that adversely affect the success of reversal and pregnancy include a longer interval between vasectomy and reversal, older age of female partner, and small testicular volume. Presence of sperm and a sperm granuloma is indicative of higher success in reversal and pregnancy.[32]

There are new developments in vas occlusion methods to gear toward enabling reversibility. RISUG,[33] Adam, and Vasalgel are in development. These methods inject vas occlusion material into the vas deferens to block sperm transport with the notion that the material can be flushed or removed when men desire reversal. The intravas injection of Vasalgel proved to be an effective contraception in rhesus monkeys, though one out of nine monkeys had incorrect placement of gel and development of a sperm granuloma in another.[34] Clinical trials demonstrating reversal of these novel vas occlusive methods have not yet been reported.[35,36]

KEY POINTS

- Vas occlusion is an effective permanent method of contraception with few complications and acceptability by users.
- Reversal leads to patency in a high proportion of men but pregnancy rate is much lower in the partner.
- New methods of vas occlusion are in development to make vas occlusion more reversible.

MALE-DIRECTED CONTRACEPTIVE METHODS IN DEVELOPMENT

Hormonal Male Contraception

Since the 1970s, investigation of hormonal male contraception has been spearheaded by the Eunice Shriver National Institute of Child Health and Human Development (NICHD).[37] The NICHD initiated studies using the testosterone ester (testosterone enanthate) as the sole agent for hormonal male contraception given its ability to suppress gonadotropin secretion and to maintain androgenicity in men.[38–40] There is a quest for a reversible method of male contraception based on sex hormones used for the treatment of male hypogonadism and on hormonal methods for the female.

Principles of Hormonal Contraception. Hormonal male contraception depends on negative feedback mechanisms in the endocrine system wherein the target organ hormone, testosterone, restrains the secretions of regulatory hormones from the hypothalamus and pituitary. In the male, the hypothalamic hormone gonadotropin-releasing hormone (GnRH) stimulates the pituitary to produce two hormones: luteinizing hormone (LH) that stimulates the Leydig cell in the testis to produce testosterone and follicle-stimulating hormone (FSH) that acts on the Sertoli cell to support spermatogenesis (Fig. 114.2). The high intratesticular concentration of testosterone is essential for germ cell development into mature spermatozoa. Knockout of the androgen receptors in the Sertoli cells of the testis in mice results in arrest of spermatogenesis at the primary spermatocytes, whereas knockout of androgen receptor in Leydig cells results in disruption of steroidogenesis, testosterone deficiency, and arrest of spermatogenesis at the round spermatid stage.[41] These studies show the critical role of testosterone within the testis for progression of spermatogenesis.

Hormonal male contraception utilizes exogenous administration of an androgen or an androgen plus a progestin or a GnRH agonist/antagonist with the goal of maximally suppressing both LH and FSH. This results in marked decrease in intratesticular testosterone production, but the exogenously administered androgen maintains sexual function and the important effects of androgens on the muscle, bone, and erythropoiesis. The serum concentrations of testosterone attained are within the physiological range, maintaining androgen action in the body, but are insufficient to maintain spermatogenesis.[42] The low intratesticular testosterone results in arrest of the development of spermatogenesis mainly at the differentiated spermatogonia and primary spermatocytes.[43] Upon discontinuation of the exogenous hormones, the hypothalamic-pituitary suppression rebounds, leading to increased intratesticular testosterone production and gradual recovery of spermatogenesis from undifferentiated spermatogonia (Fig. 114.2).[44,45] All hormonal male contraceptives will need testosterone or an androgen to maintain androgenic effects in men.

It takes about 74 days to complete the development of spermatogonia to mature spermatozoa.[46] Therefore hormonal male contraception that relies on the suppression of spermatogenesis takes an average of 60 to 90 days to produce very low sperm output and about the same time for recovery to full spermatogenesis. An integrated analysis of 30 male hormonal contraceptive clinical trials in over 1500 men showed that the probability of recovery to a sperm concentration within the reference range of fertile men was 67% in 6 months, 90% in 12 months, and 100% in 24 months. Recovery rate was higher in older men, men of Asian origin, those with shorter treatment duration, treated with short-acting testosterone preparations, higher baseline sperm concentration, faster spermatogenesis suppression, and lower serum LH at baseline.[47] Thus hormonal male methods are reversible. Because hormones act on many tissues of the body, male hormonal contraceptives may have beneficial off-target effects (e.g., increase in muscle, bone mass, and maintenance of erythropoiesis). The full implications of long-term effects on lipids, cardiovascular disease risks, and prostate diseases will not be known until hormonal male contraception is available for use and long-term large-scale epidemiological studies are performed as has been done for the female hormonal methods.[44] The androgens and progestins that have been tested in clinical trials in men are shown in Table 114.1.

Androgens. Testosterone enanthate and undecanoate intramuscular injections, testosterone implants, oral testosterone undecanoate, testosterone patch, and gel have been used in hormonal male contraceptive trials where suppression of spermatogenesis was the primary outcome.[44,45] Oral testosterone undecanoate had inadequate suppression of gonadotropins and spermatogenesis.[48] Based on the promising studies showing that testosterone enanthate administered as intramuscular injections resulted in suppression of spermatogenesis to azoospermia (no sperm detected in the ejaculate) or severe oligozoospermia (arbitrarily defined as <5 million/mL),[38–40] the World Health Organization (WHO) designed two consecutive studies that administered testosterone enanthate intramuscular injections (200 mg once per week) to healthy male volunteers for 1 year. The primary outcome was the prevention of pregnancy in the female partner when the sperm concentration is suppressed to azoospermia in the first study,[49] and to severe oligozoospermia in the second study[50] (Table 114.2). As in all hormonal male contraceptive studies, the efficacy studies recruit couples after screening and require male participants to reach a sperm

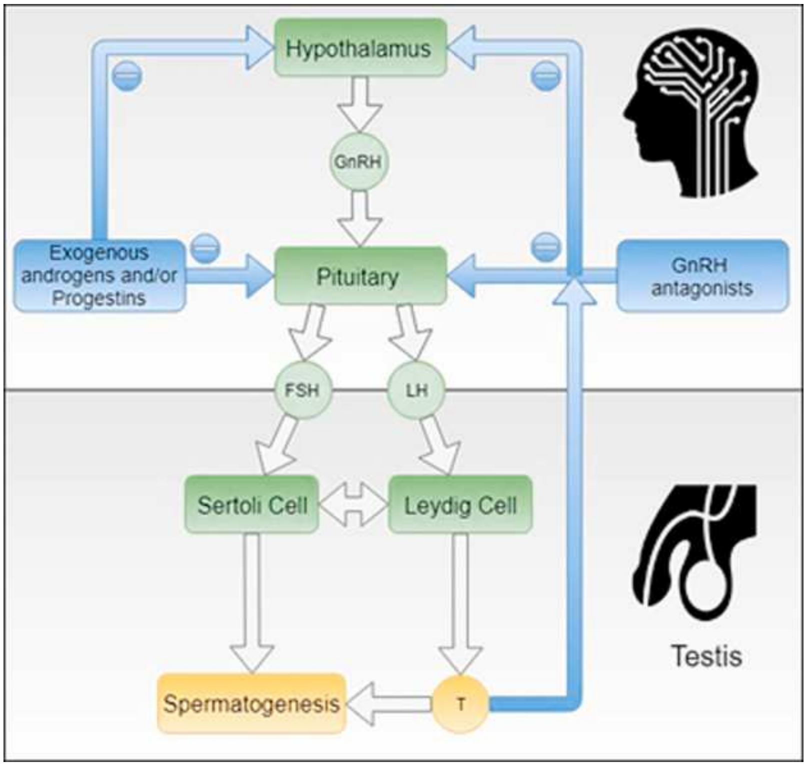


Figure 114.2 Principles of hormonal male contraception. Exogenously administered androgens and/or progestins exert negative feedback on the hypothalamic-pituitary reducing the secretion of gonadotropin-releasing hormone *(GnRH)* from the hypothalamus and both luteinizing hormone *(LH)* and follicle-stimulating hormone *(FSH)* from the pituitary. GnRH antagonists compete with the action of endogenous GnRH on the pituitary and inhibit the section of both gonadotropins. LH regulates the secretion of testosterone *(T)* by the Leydig cells and FSH regulates the function of the Sertoli cells. Inhibition of LH secretion results in very low concentration of intratesticular T, which is critical for initiation, and maintenance of spermatogenesis. Concomitant reduction of FSH further inhibits progression of spermatogenesis. Despite the marked suppression of endogenous T production by Leydig cells, prevention of testosterone deficiency and maintenance of androgen effects are provided by the exogenously administered testosterone or another androgen that is an essential component of hormonal male contraception. (From Yuen F, Nguyen BT, Swerdloff RS, Wang C. Continuing the search for a hormonal male contraceptive. *Best Pract Res Clin Obstet Gynaecol*. 2020;66:83-94.)

TABLE 114.1 **Androgens and Progestins Tested in Clinical Trials**

Route	Androgens	Progestins
Oral	Testosterone undecanoate Dimethandrolone undecanoate (DMAU)* 11β-methyl-19-nortestosterone-17β-dodecylcarbonate (11β-MNTDC)*	Levonorgestrel Desogestrel Cyproterone acetate
Transdermal	Testosterone gel,* patch, cream	Nestorone gel*
Injection	Testosterone enanthate, undecanoate, and decanoate 19-nortestosterone-hexoxyphenylpropionate Dimethandrolone undecanoate (DMAU)*	Depot-medroxyprogesterone acetate (DMPA) Norethisterone enanthate
Implant	Testosterone, 7α-methyl-19-nortestosterone (MENT)	Levonorgestrel Etonogestrel

*Currently in clinical trials.

From Yuen F, Nguyen BT, Swerdloff RS, Wang C. Continuing the search for a hormonal male contraceptive. *Best Pract Res Clin Obstet Gynaecol*. 2020;66:83-94.

concentration threshold in two or three samples in the suppression phase, initially defined as ≤5, then reduced to ≤3, and finally to ≤1 million/mL ejaculate based on expert opinion and available data.[51] Then the couple would enter the efficacy phase where they will use the method under investigation as the sole method of contraception. Recovery is defined when the sperm concentration returns to values compatible with fertility in a male. The WHO studies showed that testosterone enanthate weekly injections were efficient at inducing azoospermia in about 70% of men after 6 months and severe oligozoospermia or azoospermia in almost all men (97.8%). These two studies indicated that hormonal male contraception using androgen alone was very effective in preventing pregnancy. Because the dosage of testosterone injections in these two studies was high (approximately twice the normal physiologic production of testosterone in healthy men) and required weekly injections, subsequent studies sought to administer testosterone implants[52] or testosterone undecanoate injections[53,54] that had a longer duration of action. The contraceptive efficacy of these studies (Table 114.2) confirmed that testosterone injections and implants were as efficacious as injectable used for hormonal female contraception and much better than condom or withdrawal (Fig. 114.1). The adverse effects reported by men participating in these clinical trials were anticipated, including increased acne, oiliness of skin, mild weight and hematocrit increases, and small decreases in high-density

TABLE 114.2 Contraceptive Efficacy Clinical Trials of Testosterone Injections and Implants

Drugs	Sperm Concentration Threshold (million/mL)	Number Enrolled	Number Completing Suppression Phase	Number Reaching Threshold	Number Entering Efficacy	Pregnancies/ Failure Rate (Pearl Index)	Study Reference
TE 200 mg/week	Azoospermia	271	225 (83%)	157 (70%)	157	1 0.8 (0.0–4.5)	WHO (1990)[45]
TE 200 mg/week	≤3 (reduced from ≤5)	399	357 (89%)	349 (97.8%)	268	4 1.4 (0.4–3.7)	WHO (1996)[46]
T implants 800 or 1200 mg/3 months	≤1	36	29	21 (72%)	13	0	McLachlan et al. (2000)[48]
TU 1000 mg loading 500 mg/month	≤3	308	308	299 (97.1%)	296	1 2.3 (0.5–4.2)	Gu et al. (2003)[49]
TU 1000 mg loading 500 mg/month	≤1	1045	898	855 (95.2%)	855	9 1.1 (0.4–1.8)	Gu et al. (2009)[50]

T, Testosterone; *TE*, testosterone enanthate; *TU*, testosterone undecanoate; Norethisterone Enanthate. Pearl Index (95% confidence interval) per 100 couple-year

TABLE 114.3 Contraceptive Efficacy of Clinical Trials of Testosterone With Progestins

Drugs	Sperm Concentration Threshold (million/mL)	Number Enrolled	Number Completing Suppression Phase	Number Reaching Threshold	Number Entering Efficacy	Pregnancies/ Failure Rate (Pearl Index)	Study Reference
T implants 800/1200 mg/4–6 months + DMPA 300 mg/3 months	≤1	55	55	53 (94%)	51	0 0 (0–8)	Turner et al. (2003)[55]
TU 1000 mg IM and norethisterone enanthate IM 200 mg/8 weeks	≤1	320	283	274 (95.9%)	266	4 1.57 (0.59–4.14)	Behre et al. (2016)[56]
T 62 to 75 mg + nestorone 8 mg gel transdermal daily	≤1	420 ongoing					

T, Testosterone; *TU*, testosterone undecanoate.

lipoprotein-cholesterol levels. These side effects did not cause early discontinuations.[55] The dosages of androgens used in these studies may be considered higher than dosages used for testosterone replacement in hypogonadal men. Furthermore, the suppression of sperm output showed an ethnic/racial difference in that the suppression to severe oligozoospermia or azoospermia was higher in Asian (>90%) versus non-Asian men (~60% to 70%).[49,50,56] For these reasons, clinical trials began utilizing an additional gonadotropin-suppressing agent, a progestin, or a GnRH agonist or antagonist, together with a male hormone to suppress spermatogenesis.

Androgens and Progestins. The addition of a progestin to testosterone in hormonal male contraceptive clinical trials more effectively suppresses the hypothalamic–pituitary–testis axis. It is believed that this combined hormonal approach will allow use of reduced dosages of testosterone to physiological concentrations with retained efficacy by the combined treatment. This remains under study, as the use of two agents instead of testosterone alone adds complexity in regulatory development and requires more data on the safety of the progestin in men. Despite these caveats, many progestins have been tested in male contraceptive clinical trials. Levonorgestrel, desogestrel, and cyproterone acetate oral pills, intramuscular depot-medroxyprogesterone acetate (DMPA) and norethisterone enanthate injections, and levonorgestrel and etonogestrel implants have been studied and transdermal nestorone (NES-T) gel is currently being studied. These progestins have been used in combination with oral testosterone undecanoate, testosterone enanthate, decanoate, and undecanoate injections, testosterone implants, and transdermal testosterone gel (Table 114.1).[44,57] In spermatogenesis suppression studies with small numbers of men, these combinations of testosterone injections or implants plus progestin oral preparations, injections, implants, and gel have effectively suppressed sperm output to azoospermia and/or very severe oligozoospermia in over 95% of men. In addition to improved suppression of sperm output, the combined androgen-progestin regimens appear to suppress spermatogenesis faster than androgens alone.[58]

Two contraceptive efficacy studies with testosterone plus a progestin have been reported (Table 114.3). Testosterone 800 mg implants every 4 to 6 months together with intramuscular DMPA 300 mg every 12 weeks suppressed sperm concentration to ≤1 million/mL in over 94% of men and prevented pregnancy in their female partners.[59] In a larger study, 1000 mg of testosterone undecanoate was administered with 200 mg norethisterone enanthate as intramuscular injections every 8 weeks to 320 couples. Results showed 95.9% of men suppressed to a sperm concentration ≤1 million/mL (the threshold for contraceptive efficacy) and only four pregnancies occurred in 266 couples giving a failure rate of 1.57 per 100 continuing users (Table 114.3).[60] The common adverse events were pain at the injection site, acne, increased libido, and mood changes. Most of the adverse events were mild and concentrated at one clinical study site, but an external safety review committee prematurely stopped the study because of mood disorders including depression. Some of the adverse events may be related to low serum testosterone levels at the end of the injection interval and the different pharmacokinetics of testosterone undecanoate and norethisterone enanthate. Moreover, the participants were not screened for depressive symptoms and prospective monitoring of mood and sexual function was not included in the study. Despite these adverse events, the male participants expressed a high level of satisfaction and over 87% of both male and female participants affirmed they would use this as a method of contraception if available.[60] While other long-acting reversible male contraception regimens are being developed, user-friendly, provider-independent methods of drug delivery are under investigation. Daily transdermal application of testosterone gel together with a potent novel progestin without androgenic and estrogenic activity, Nestorone (segesterone acetate), was shown to be more effective in suppressing gonadotropins and spermatogenesis than testosterone gel alone in men. Minimal adverse events were reported.[61,62] Currently, a contraceptive efficacy study of transdermal testosterone and Nestorone gel with a plan of recruiting 420 couples with over 50% of them completing 12 months of the efficacy phase is ongoing in 13 centers in four continents.

KEY POINTS

- Addition of a progestin to testosterone suppresses spermatogenesis faster than testosterone alone.
- Oral pills, transdermal gel, injections, and implants of various progestins have been combined with testosterone injection, implants, and transdermal gels in male contraceptive clinical trials.
- Contraceptive efficacy studies with progestins combined with injections or implants of testosterone result in severe oligozoospermia in most men and prevention of pregnancy rates in couples approximating efficacy as for hormonal female methods.

Gonadotropin-Releasing Hormone Agonist and Antagonists. Both GnRH agonists and antagonists in combination with testosterone injections have undergone clinical trials in men. GnRH agonists initially induce increases in both LH and FSH but then owing to downregulation of the gonadotropin receptors result in marked suppression of both gonadotropins and testosterone production. These agents have been widely used in androgen depletion therapy for prostate cancer and for precocious puberty. However, GnRH agonists' injection or infusion together with testosterone injections failed to suppress sperm output to severe oligozoospermia in most men.[63,64] In contrast, GnRH antagonists administered as a subcutaneous daily injection together with testosterone supplementation were efficacious in suppressing sperm output to azoospermia or severe oligozoospermia in most men.[65–68] It was also demonstrated that once sperm suppression was induced by the GnRH antagonist plus testosterone enanthate injections, azoospermia could be maintained for 6 months with replacement dosages of testosterone enanthate alone.[69] Despite the success with GnRH antagonists, the cost of production and the requirement of daily injections impeded the further development of these agents as viable hormonal male contraceptives. Oral bioavailable GnRH antagonists are being developed for the treatment of endometriosis and uterine fibroids.[70,71] Such agents could be tested in clinical trials in men combined with new orally bioavailable testosterone undecanoate preparations that sufficiently raise testosterone levels to the adult range in hypogonadal men.[72]

Other New Hormonal Agents for Male Contraception

A modified androgen 7-α methyl 19-nortestosterone (MENT) formulated as an implant showed promise as a possible male contraceptive agent.[73] However, the release of testosterone was not maintained in all batches of MENT implants[74] and the redevelopment of the implant is in progress. Novel-modified androgens related to MENT that bind to and activate both the androgen and progesterone receptors have the potential of one compound possessing the effects of two steroids and may mimic the action of combined androgen and progestins regimens. These include dimethandrolone (DMA, 7α,11β-dimethyl-19-nortestosterone) and 11β-MNT (11β-methyl-19-nortestosterone) administered as prodrugs DMA undecanoate (DMAU)[75,76] and 11β-MNT dodecylcarbonate (11β-MNTDC).[77] In contrast to 19-nortestosterone, the backbone of these modified androgens, these compounds are not aromatized to estrogenic compounds and are not subjected to 5α-reductase activity, resulting in less prostate-stimulating activity. Both compounds have been administered to healthy men as a daily oral pill with marked suppression of LH, FSH, and testosterone while maintaining sexual and other androgenic functions.[44,77] Both DMAU and 11β-MNTDC are being developed not only as oral pills but also as intramuscular injections as potential male contraceptives.

KEY POINTS

- GnRH antagonists are very effective suppressors of the hypothalamic–pituitary–testis axis and when administered with testosterone suppresses sperm concentration to very low levels compatible with contraceptive efficacy. New orally bioavailable agents are being developed.
- New orally bioavailable androgens with progestational activity are being tested in clinical trials to assess whether these agents can suppress spermatogenesis.

Nonhormonal Targets for Male Contraception

Since hormones act on many other biological systems other than the hypothalamic–pituitary–testis axis, off-target effects can occur. Nonhormonal agents acting on a specific cellular protein or biological process in the male reproductive tract mainly exert localized effects for inhibiting sperm production, motility, or fertilizing capacity. The common approach to nonhormonal male contraceptive development is to identify a target (usually a protein with expression mostly restricted to the male reproductive tract); to demonstrate regulation/disruption of the target gene leads to infertility; and demonstration that the target is druggable (i.e., proteins with binding niches where small-molecule inhibitors can bind). Once the target is identified, drug development usually requires screening for small molecules that bind to the target scaffold using high-throughput screening with measurable readouts. Alternatively, new targets can be discovered by integrating human and mouse RNA-seq datasets combined with additional analyses of human and mouse testis and epididymis.[78] Targets include spermatogenesis, spermiogenesis, sperm maturation in the epididymis, sperm motility, hyperactivated sperm motility, capacitation, acrosome reaction, binding to the zona pellucida, and sperm–egg fusion (Fig. 114.3).[79–81] Examples of a few nonhormonal targets are discussed.

Spermatogenesis as a Target. Hormonal methods act on spermatogenesis with arrest of spermatogenesis at the differentiating spermatogonia. Retinoic acid receptor α (RARα) is required for spermatogenesis.[82] Antagonists to RARα induce reversible inhibition of spermatogenesis and infertility in mice.[83] Structure-activity-based drug design may yield small molecules that will prevent binding of RA to RARα, specifically inhibiting spermatogenesis. Another approach is to disrupt the synthesis of RA using bisdichloroacetyldiamine (BDAD), which leads to spermatogenesis suppression in men.[84] Unfortunately, BDAD causes a severe disulfiram reaction after alcohol consumption

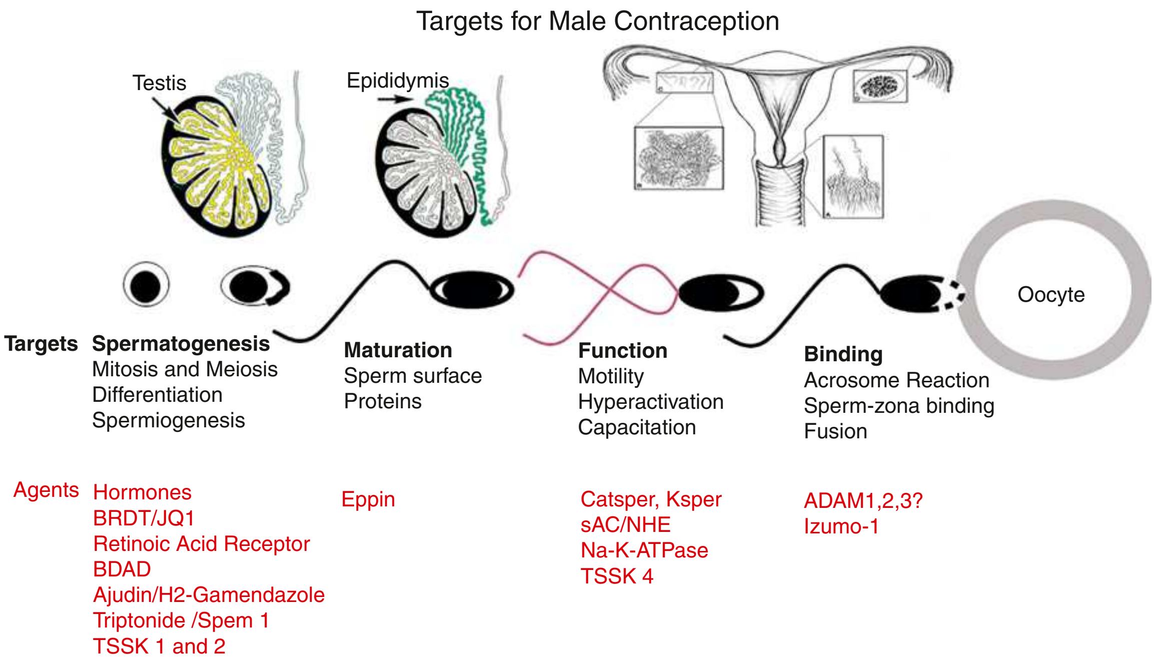


Figure 114.3 Targets for male contraception. The targets for male contraception include processes in sperm development (spermatogenesis and spermiogenesis in the testis), maturation, acquisition of motility (transit through the epididymis), and the ability to hyperactivate, undergo acrosome reaction, and bind and fuse with oocyte. Possible contraceptive agents include those that act in the *testis* and affect *spermatogenesis*: Retinoic acid receptor (RARα) antagonist, bisdichloroacetyldiamine (BDAD), bromodomain testis-specific protein (BRDT); *tight junctions* between Sertoli and germ cells: Adjudin and H2-gamendazole; *spermiogenesis*: Spermatid maturation 1 (Spem1) and testis-specific serine kinase protein (TSSK 1 and 2). In the *epididymis*, where *sperm maturation and acquisition of motility* occurs, epididymal protease inhibitor (EPPIN) antibodies and antagonist reduce sperm motility and cause infertility. In the *female reproductive trac*t, agents that affect *sperm motility and hyperactivated motility*: ion channels CatSper and KSper; soluble adenylyl cyclase sAC and sodium-hydrogen exchangers (NHE), sodium-potassium ATPase (Na-K-ATPase) and testis-specific serine kinase 4 (TSSK4); *sperm transport and fertilization*: a disintegrin and metalloproteases (ADAMs) family proteins ADAM 1a, 2, and 3; IZUMO1 and JUNO. (Adapted from Gottwald U, Davies B, Fritsch M, Habenicht UF. New approaches for male fertility control: HE6 as an example of a putative target. *Mol Cell Endocrinol*. 2006;250(1-2):49-57.)

due to the inhibition of aldehyde dehydrogenase 2 resulting in aldehyde intoxication.[85] Studies are currently underway to custom design and synthesize an inhibitor of the testis-specific aldehyde dehydrogenase 1A2 enzyme to avoid the disulfiram reaction while retaining the intratesticular effect of BDAD on spermatogenesis.

Bromodomain testis-specific protein (BRDT) knockout mice are infertile. A small-molecule inhibitor (JQ1) of BRDT causes reversible suppression of spermatogenesis in mice. BRDT is essential for chromatin remodeling during spermatogenesis. Current efforts are directed to screen molecules based on structure-activity relationships for new testis-specific compounds.[86] A number of indazole carboxylic acid compounds disrupt junctions between Sertoli and male germ cells causing immature germ cell loss from the seminiferous epithelium.[87] These compounds, including adjudin[88,89] and H2-gamendazole,[90] induce reversible infertility in rodents. Unfortunately, these compounds have little penetrance across the blood–testis barrier, raising a concern for irreversible effects with high cumulative dosages.

Spermiogenesis as a Target. Spermiogenesis is the process whereby postmeiotic haploid round cells develop into elongated spermatids and spermatozoa with the sperm head containing the nucleus and the acrosome, the mid-piece with mitochondria and the flagellum. Spermiogenesis genes are testis specific and are conserved.[91] Gene knockouts of Spermatid maturation 1 (*Spem1*)[92] and Spermatogenesis-associated Protein 6 (*Spata6*)[93] affect late spermiogenesis and result in mice with normal testicular histology but with malformed spermatozoa that are not capable of demonstrating forward progressive motility. A recent study reported that triptonide, purified from *Tripterygium wilfordii*, when administered to mice and monkeys caused deformed sperm and infertility. Triptonide appeared to disrupt the interaction of the junction plakoglobin with SPEM1.[94] Testis-specific serine kinase (TSSK) 1 and 2 are critical for spermiogenesis and double-knockout mice show sperm head abnormalities and are infertile. TSK6 knockout mice showed altered head morphology and impairment attachment of the sperm head to the mid-piece due to defects in sperm DNA condensation, chromatin remodeling, and histone to protamine transition.[95] The hope is to use high-throughput small-molecule screening to identify specific molecules that target these crucial gene products and thus reversibly impair spermiogenesis.

Sperm Maturation and Sperm Motility as Targets. While transiting through the epididymis, spermatozoa acquire many sperm surface proteins essential for sperm maturation, acquisition of motility, and development of fertilizing capacity. Thus epididymal[96] and sperm motility signaling pathway proteins[97] are potential targets of male contraception. The sperm surface protein, epididymal protease inhibitor (EPPIN), is important for forward progressive motility of spermatozoa. Antibodies to EPPIN and a specific inhibitor of EPPIN (EP055) reduced sperm motility and caused infertility in rodents and monkeys.[98,99]

Ion channel proteins are produced in the tail end of the flagellum and control the calcium influx. Hyperactivated motility is activated through voltage-gated ion channels CatSper (calcium ion channel) and KSper (potassium ion channel). Hyperactivated motility is necessary for spermatozoa transit through the female reproductive tract and penetration of the zona pellucida. The structure and dynamics of these ion channels are very complex. CatSper is regulated by progesterone, intracellular pH, and endocannabinoids. CatSper deficient mice and men with certain CatSper gene mutations are infertile. Spermatozoa deficient in CatSper fail to fertilize an oocyte and those deficient in KSper can fertilize but at a markedly reduced rate.[100] Because CatSper is sperm specific and affects hyperactivated motility in the female reproductive tract, small molecule targeting the channel subunits may be developed as both male and female nonhormonal contraceptives.[101]

Other potential target is the soluble adenylyl cyclase. Disruption of this enzyme results in loss of sperm motility and infertility in mice.[102] The Na-H exchanger (NHE) is critical for the regulation of adenylyl cyclase and sperm motility.[103] In addition, loss of NHE 8 disrupts acrosome formation and causes infertility.[104] The Na-K ATPase (NKA) catalytic subunit NKAα4 (ATP1A4) is testis specific, present in spermatozoa, and required for sperm motility. This catalytic subunit is being used as a target to block ion transport and disrupt male fertility.[105] TSSK4 knockout mice have decreased sperm motility because of abnormality of the annulus at the principle and mid-piece junction resulting in bent flagellum.[95] Current research focuses on the development of cell membrane permeable inhibitors that are druggable.

Sperm Transport and Sperm–Egg Binding and Interaction. Based on gene knockout studies in mice, many gene products have been reported to impair sperm function and sperm–egg interactions. A family of disintegrins and metalloproteases (ADAMs), ADAM 1a, 2, and 3 are important in mouse sperm migration through the uterotubal junction. ADAM 1a/2 and ADAM 3 are important in inducing the ability of mouse spermatozoa to bind to the zona pellucida. However, the role of ADAM 3 in human fertilization is not clear. Many other gene products may also be required for sperm–egg interaction.[106] More recently, the sperm membrane protein IZUMO1 binding to its receptor JUNO on the oocyte has been identified to be critical for fertilization.[107,108] IZUMO-I-JUNO complex is a target of contraception. Because sperm transport, the acrosome reaction, and sperm–egg binding occur mainly in the female reproductive tract, these are also targets for nonhormonal female methods of contraception as drug delivery to the female reproductive tract may be more streamline for such a method.

KEY POINTS

- Nonhormonal methods target specific gene products essential for spermatogenesis, spermiogenesis, sperm maturation, sperm motility, sperm hyperactivated motility, and sperm–egg binding.
- Drug development involves screening of small molecules or structure-activity-based synthesis of compounds that specifically target a gene product that is critical for fertility.
- Some of the inhibitors are available for testing in rodents and nonhuman primates but none has reached clinical testing in men.
- Unanticipated off-target effects are unknown and require extensive preclinical testing.

Men’s Attitude and Acceptability of Male Contraception

The attitude of men toward family planning responsibility seems to have changed in the last decade. Advancements in the understanding of male reproduction stimulate new developments in male contraception. Gender equity and social justice concepts propel men to acknowledge and accept that family planning and contraceptive use should be shared between the partners.[109,110] The question of whether men will use contraception was examined by multicenter surveys conducted over a decade ago. In a study where over 9000 men in nine countries in four continents were surveyed by phone, 55% to 81.3% of the couples responded that the method of contraception was decided together by the couple. The overall acceptability of a hypothetical, reversible male contraceptive method was 55%, varying between countries.[111] In another multicenter study of over 1800 couples in four countries in three continents, 44% to 83% of the men would use a new contraceptive especially if this was a pill.[112] In this survey, when 1894 female partners were asked, 83% of the women indicated that a new male contraceptive method was a good idea and only 2% of the women would not trust their partner to use the male method.[113] Surveys on men who were participating in clinical trials consistently show that over half of the trial participants will use the method when available, methods that have included injections, transdermal gels, and daily oral pills.[114–117] In an ongoing contraceptive clinical trial of daily NES-T transdermal gel, comprehensive acceptability questionnaires are being collected while in-depth interviews of the couples will be conducted to determine the attitudes and behavior of men and women who are interested and elected to participate in the contraceptive efficacy study that lasts over 18 months.

With social media penetrating the everyday life of younger adults, contraceptive dialogues with online audiences suggest that male contraception is a topic of interest in couples of the reproductive age.[44] Before launching a new male contraceptive method, landscape studies in different countries on different racial/ethnic, socioeconomic groups must be planned to allow the community to inform the researchers what are the best methods to disseminate information and launch new male contraceptive methods.

KEY POINTS

- Multicountry surveys performed a decade ago indicate that over 50% of men would accept a new male-controlled method of contraception.
- Over 95% of women in stable relationship will trust their male partner to use the new male method when available.
- With gender equity, men are willing to accept family planning responsibility.

SUMMARY AND FUTURE DIRECTIONS

The demand for new, modern, reversible methods of male contraception independent of timing of sexual activity will drive the development and approval of a new method of male contraception. There are many possible candidate approaches in the pipeline. As an example, this could be a reversible vas occlusion or a hormonal method. Large-scale studies will need to be performed to demonstrate safety and reversibility of these methods. Launch of new methods must be preceded by landscape studies to optimize dissemination of information surrounding these methods and engage men and couples in male contraceptive utilization. With progress in our understanding of the molecular and cellular processes regulating spermatogenesis, sperm maturation, and acquisition of motility and fertilizing potential of spermatozoa, new testis-/spermatozoa-specific targets are being identified. Using new molecular technologies and high-throughput screening of small molecules that inhibit the critical targets affecting development, maturation, and acquisition of essential functions of spermatozoa, new agents with specific actions on the testis, epididymis, or reproductive tract will be identified. Novel nonhormonal agents will need to pass preclinical toxicology to insure there are no significant off-target adverse effects. Novel male contraceptive agents will preferably induce reversible contraception to expand the contraceptive options for men and couples. Lastly, minimizing the cost of novel male contraceptive methods is of great importance in ensuring these agents have an impact on unplanned pregnancy rates. Only if they are affordable methods for all men, irrespective of their income, will novel male contraceptives realize their potential to have a major impact on global public health.

REFERENCES

1. Bearak J, Popinchalk A, Alkema L, Sedgh G. Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a Bayesian hierarchical model. *Lancet Global Health*. 2018;6(4):e380–e389.
2. Sedgh G, Singh S, Hussain R. Intended and unintended pregnancies worldwide in 2012 and recent trends. *Stud Fam Plann*. 2014;45(3):301–314.
3. Sedgh G, Bearak J, Singh S, et al. Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends. *Lancet*. 2016;388(10041):258–267.
4. Trussell J, Henry N, Hassan F, et al. Burden of unintended pregnancy in the United States: potential savings with increased use of long-acting reversible contraception. *Contraception*. 2013;87(2):154–161.
5. Sonfield A, Kost K, Gold RB, Finer LB. The public costs of births resulting from unintended pregnancies: national and state-level estimates. *Perspect Sex Reprod Health*. 2011;43(2):94–102.
6. Gavin L, Moskosky S, Carter M, et al. Providing quality family planning services: recommendations of CDC and the US Office of Population Affairs. *Center for Disease Control and Prevention: Morbidity and Mortality Weekly Report*. 2014;63(4):1–50.
7. World Health Organization, Johns Hopkins School of Public Health. *Family Planning: A Global Handbook for Providers*; 2018. https://www.fphandbook.org/.
8. Centers for Disease Control and Prevention. *How Effective Are Birth Control Methods?* 2014. https://www.cdc.gov/reproductivehealth/unintendedpregnancy/pdf/family-planning-methods-2014.pdf.
9. Trussell J. Contraceptive failure in the United States. *Contraception*. 2011;83(5):397–404.
10. Caetano C, Peers T, Papadopoulos L, et al. Millennials and contraception: why do they forget? An international survey exploring the impact of lifestyles and stress levels on adherence to a daily contraceptive regimen. *Eur J Contracept Reprod Health Care*. 2019;24(1):30–38.
11. Hubacher D, Spector H, Monteith C, et al. Long-acting reversible contraceptive acceptability and unintended pregnancy among women presenting for short-acting methods: a randomized patient preference trial. *Am J Obstet Gynecol*. 2017;216(2):101–109.
12. Mack N, Crawford TJ, Guise JM, et al. Strategies to improve adherence and continuation of shorter-term hormonal methods of contraception. *Cochrane Database Syst Rev*. 2019;4(4):Cd004317.
13. Ross J, Hardee K. Use of male methods of contraception worldwide. *J Biosoc Sci*. 2017;49(5):648–663.
14. Dorman E, Bishai D. Demand for male contraception. *Expert Rev Pharmacoecon Outcomes Res*. 2012;12(5):605–613.
15. Wang C, Sitruk-Ware R, Serfaty D. It is time for new male contraceptives. *Andrology*. 2016;4(5):773–775.
16. Rogow D, Horowitz S. Withdrawal: a review of the literature and an agenda for research. *Stud Fam Plann*. 1995;26(3):140–153.
17. Polis CB, Bradley SE, Bankole A, et al. Typical-use contraceptive failure rates in 43 countries with Demographic and Health Survey data: summary of a detailed report. *Contraception*. 2016;94(1):11–17.
18. Copen CE. Condom use during sexual intercourse among women and men aged 15-44 in the United States: 2011-2015 National Survey of Family Growth. *Natl Health Stat Report*. 2017;105(105):1–18.
19. Giannou FK, Tsiara CG, Nikolopoulos GK, et al. Condom effectiveness in reducing heterosexual HIV transmission: a systematic review and meta-analysis of studies on HIV serodiscordant couples. *Expert Rev Pharmacoecon Outcomes Res*. 2016;16(4):489–499.
20. Sanders SA, Yarber WL, Kaufman EL, et al. Condom use errors and problems: a global view. *Sex Health*. 2012;9(1):81–95.
21. Centers for Disease Control and Prevention. *The Right Way to Use a Male Condom*; 2016. https://www.cdc.gov/condomeffectiveness/docs/male-condom-use-508.pdf.
22. Sundaram A, Vaughan B, Kost K, et al. Contraceptive failure in the United States: estimates from the 2006-2010 national survey of family Growth. *Perspect Sex Reprod Health*. 2017;49(1):7–16.
23. Jones RK, Darroch JE, Henshaw SK. Contraceptive use among U.S. women having abortions in 2000-2001. *Perspect Sex Reprod Health*. 2002;34(6):294–303.
24. Gallo MF, Grimes DA, Lopez LM, Schulz KF. Non-latex versus latex male condoms for contraception. *Cochrane Database Syst Rev*. 2006;1(1):Cd003550.
25. United Nations. *World Contraceptive Use*; 2020. https://www.un.org/en/development/desa/population/publications/dataset/contraception/wcu2020asp.
26. Sharma V, Le BV, Sheth KR, et al. Vasectomy demographics and postvasectomy desire for future children: results from a contemporary national survey. *Fertil Steril*. 2013;99(7):1880–1885.
27. Byrne KS, Castano JM, Chirlaque MD, et al. Vasectomy and prostate cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Clin Oncol*. 2017;35(12):1297–1303.
28. Shoag J, Savenkov O, Christos PJ, et al. Vasectomy and risk of prostate cancer in a screening trial. *Cancer Epidemiol Biomarkers Prev*. 2017;26(11):1653–1659.
29. Cook LA, Pun A, Gallo MF, et al. Scalpel versus no-scalpel incision for vasectomy. *Cochrane Database Syst Rev*. 2014;3:CD004112.
30. Sharlip IM, Belker AM, Honig S, et al. *AUA Clinical Guideline: Vasectomy*; 2012. https://www.auanet.org/guidelines/vasectomy-(2012-reviewed-for-currency-2015.
31. Sandlow JI, Westefeld JS, Maples MR, Scheel KR. Psychological correlates of vasectomy. *Fertil Steril*. 2001;75(3):544–548.
32. Namekawa T, Imamoto T, Kato M, et al. Vasovasostomy and vasoepididymostomy: review of the procedures, outcomes, and predictors of patency and pregnancy over the last decade. *Reprod Med Biol*. 2018;17(4):343–355.
33. Guha SK, Singh G, Ansari S, et al. Phase II clinical trial of a vas deferens injectable contraceptive for the male. *Contraception*. 1997;56(4):245–250.
34. Colagross-Schouten A, Lemoy M-J, Keesler RI, et al. The contraceptive efficacy of intravas injection of Vasalgel™ for adult male rhesus monkeys. *Basic Clin Androl*. 2017;27(1):4.
35. Guha SK, Singh G, Anand S, et al. Phase I clinical trial of an injectable contraceptive for the male. *Contraception*. 1993;48(4):367–375.
36. Sharma RS, Mathur AK, Singh R, et al. Safety & efficacy of an intravasal, one-time injectable & non-hormonal male contraceptive (RISUG): a clinical experience. *Indian J Med Res*. 2019;150(1):81–86.
37. Patanelli D. *Hormonal Control of Male Fertility*. Bethesda, Maryland: Department of Health, Education, and Welfare; 1978:145–172. (DHEW Publication No. NIH 78-1097), 1978.
38. Cunningham GR, Silverman VE, Thornby J, Kohler PO. The potential for an androgen male contraceptive. *J Clin Endocrinol Metab*. 1979;49(4):520–526.
39. Steinberger E, Smith KD, Rodriguez-Rigau LJ. Suppression and recovery of sperm production in men treated with testosterone enanthate for one year. A study of a possible reversible male contraceptive. In: *Endocrine Approach to Male Contraception*. Copenhagen: Scriptor; 1978:748–760.
40. Swerdloff RS, Campfield LA, Palacios A, McClure RD. Suppression of human spermatogenesis by depot androgen: potential for male contraception. *J Steroid Biochem*. 1979;11(1B):663–670.
41. Wang RS, Yeh S, Tzeng CR, Chang C. Androgen receptor roles in spermatogenesis and fertility: lessons from testicular cell-specific androgen receptor knockout mice. *Endocr Rev*. 2009;30(2):119–132.
42. Coviello AD, Bremner WJ, Matsumoto AM, et al. Intratesticular testosterone concentrations comparable with serum levels are not sufficient to maintain normal sperm production in men receiving a hormonal contraceptive regimen. *J Androl*. 2004;25(6):931–938.
43. McLachlan RI, O'Donnell L, Meachem SJ, et al. Hormonal regulation of spermatogenesis in primates and man: insights for development of the male hormonal contraceptive. *J Androl*. 2002;23(2):149–162.

44. Yuen F, Nguyen BT, Swerdloff RS, Wang C. Continuing the search for a hormonal male contraceptive. *Best Pract Res Clin Obstet Gynaecol*. 2020;66:83–94.
45. Thirumalai A, Page ST. Recent developments in male contraception. *Drugs*. 2019;79(1):11–20.
46. Clermont Y. Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev*. 1972;52(1):198–236.
47. Liu PY, Swerdloff RS, Christenson PD, et al. Rate, extent, and modifiers of spermatogenic recovery after hormonal male contraception: an integrated analysis. *Lancet*. 2006;367(9520):1412–1420.
48. Nieschlag E, Hoogen H, Bolk M, et al. Clinical trial with testosterone undecanoate for male fertility control. *Contraception*. 1978;18(6):607–614.
49. World Health Organization Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia in normal men. World Health Organization Task Force on methods for the regulation of male fertility. *Lancet*. 1990;336(8721):955–959.
50. World Health Organization Task Force on the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil Steril*. 1996;65(4):821–829.
51. Aaltonen P, Amory JK, Anderson RA, et al. 10th Summit Meeting consensus: recommendations for regulatory approval for hormonal male contraception. *J Androl*. 2007;28(3):362–363.
52. McLachlan RI, McDonald J, Rushford D, et al. Efficacy and acceptability of testosterone implants, alone or in combination with a 5alpha-reductase inhibitor, for male hormonal contraception. *Contraception*. 2000;62(2):73–78.
53. Gu YQ, Wang XH, Xu D, et al. A multicenter contraceptive efficacy study of injectable testosterone undecanoate in healthy Chinese men. *J Clin Endocrinol Metab*. 2003;88(2):562–568.
54. Gu Y, Liang X, Wu W, et al. Multicenter contraceptive efficacy trial of injectable testosterone undecanoate in Chinese men. *J Clin Endocrinol Metab*. 2009;94(6):1910–1915.
55. Ilani N, Swerdloff RS, Wang C. Male hormonal contraception: potential risks and benefits. *Rev Endocr Metab Disord*. 2011;12(2):107–117.
56. Ilani N, Liu PY, Swerdloff RS, Wang C. Does ethnicity matter in male hormonal contraceptive efficacy? *Asian J Androl*. 2011;13(4):579–584.
57. Meriggiola MC, Farley TM, Mbizvo MT. A review of androgen-progestin regimens for male contraception. *J Androl*. 2003;24(4):466–483.
58. Liu PY, Swerdloff RS, Anawalt BD, et al. Determinants of the rate and extent of spermatogenic suppression during hormonal male contraception: an integrated analysis. *J Clin Endocrinol Metab*. 2008;93(5):1774–1783.
59. Turner L, Conway AJ, Jimenez M, et al. Contraceptive efficacy of a depot progestin and androgen combination in men. *J Clin Endocrinol Metab*. 2003;88(10):4659–4667.
60. Behre HM, Zitzmann M, Anderson RA, et al. Efficacy and safety of an injectable combination hormonal contraceptive for men. *J Clin Endocrinol Metab*. 2016;101(12):4779–4788.
61. Mahabadi V, Amory JK, Swerdloff RS, et al. Combined transdermal testosterone gel and the progestin nestorone suppresses serum gonadotropins in men. *J Clin Endocrinol Metab*. 2009;94(7):2313–2320.
62. Ilani N, Roth MY, Amory JK, et al. A new combination of testosterone and nestorone transdermal gels for male hormonal contraception. *J Clin Endocrinol Metab*. 2012;97(10):3476–3486.
63. Bhasin S, Yuan QX, Steiner BS, Swerdloff RS. Hormonal effects of gonadotropin-releasing hormone (GnRH) agonist in men: effects of long term treatment with GnRH agonist infusion and androgen. *J Clin Endocrinol Metab*. 1987;65(3):568–574.
64. Bhasin S, Steiner B, Swerdloff R. Does constant infusion of gonadotropin-releasing hormone agonist lead to greater suppression of gonadal function in man than its intermittent administration? *Ferti Steril*. 1985;44(1):96–101.
65. Bagatell CJ, Matsumoto AM, Christensen RB, et al. Comparison of a gonadotropin releasing-hormone antagonist plus testosterone (T) versus T alone as potential male contraceptive regimens. *J Clin Endocrinol Metab*. 1993;77(2):427–432.
66. Bagatell CJ, Rivier JE, Bremner WJ. Dose effects of the gonadotropin-releasing hormone antagonist, Nal-Glu, combined with testosterone enanthate on gonadotropin levels in normal men. *Ferti Steril*. 1995;64(1):139–145.
67. Salameh W, Bhasin S, Steiner BS, et al. Comparative effects of two different delivery systems on gonadotropin-releasing hormone (GnRH) antagonist-induced suppression of gonadotropins and testosterone in man. *J Androl*. 1994;15(1):22–28.
68. Tom L, Bhasin S, Salameh W, et al. Induction of azoospermia in normal men with combined Nal-Glu gonadotropin-releasing hormone antagonist and testosterone enanthate. *J Clin Endocrinol Metab*. 1992;75(2):476–483.
69. Swerdloff RS, Bagatell CJ, Wang C, et al. Suppression of spermatogenesis in man induced by Nal-Glu gonadotropin releasing hormone antagonist and testosterone enanthate (TE) is maintained by TE alone. *J Clin Endocrinol Metab*. 1998;83(10):3527–3533.
70. Osuga Y, Enya K, Kudou K, Hoshiai H. Relugolix, a novel oral gonadotropin-releasing hormone antagonist, in the treatment of pain symptoms associated with uterine fibroids: a randomized, placebo-controlled, phase 3 study in Japanese women. *Fertil Steril*. 2019;112(5):922–929.e2.
71. Taylor HS, Giudice LC, Lessey BA, et al. Treatment of endometriosis-associated pain with Elagolix, an oral GnRH antagonist. *N Engl J Med*. 2017;377(1):28–40.
72. Swerdloff RS, Wang C, White WB, et al. A new oral testosterone undecanoate formulation restores testosterone to normal concentrations in hypogonadal men. *J Clin Endocrinol Metab*. 2020;105(8):2515–2531.
73. von Eckardstein S, Noe G, Brache V, et al. A clinical trial of 7 alpha-methyl-19-nortestosterone implants for possible use as a long-acting contraceptive for men. *J Clin Endocrinol Metab*. 2003;88(11):5232–5239.
74. Anderson RA, Martin CW, Kung AW, et al. 7 Alpha-methyl-19-nortestosterone maintains sexual behavior and mood in hypogonadal men. *J Clin Endocrinol Metab*. 1999;84(10):3556–3562.
75. Surampudi P, Page ST, Swerdloff RS, et al. Single, escalating dose pharmacokinetics, safety and food effects of a new oral androgen dimethandrolone undecanoate in man: a prototype oral male hormonal contraceptive. *Andrology*. 2014;2(4):579–587.
76. Ayoub R, Page ST, Swerdloff RS, et al. Comparison of the single dose pharmacokinetics, pharmacodynamics, and safety of two novel oral formulations of dimethandrolone undecanoate (DMAU): a potential oral, male contraceptive. *Andrology*. 2017;5(2):278–285.
77. Wu S, Yuen F, Swerdloff RS, et al. Safety and pharmacokinetics of single-dose novel oral androgen 11beta-methyl-19-nortestosterone-17beta-dodecylcarbonate in men. *J Clin Endocrinol Metab*. 2019;104(3):629–638.
78. Robertson MJ, Kent K, Tharp N, et al. Large-scale discovery of male reproductive tract-specific genes through analysis of RNA-seq datasets. *BMC Biol*. 2020;18(1):103.
79. Gottwald U, Davies B, Fritsch M, Habenicht UF. New approaches for male fertility control: HE6 as an example of a putative target. *Mol Cell Endocrinol*. 2006;250(1–2):49–57.
80. Long JE, Lee MS, Blithe DL. Male contraceptive development: update on novel hormonal and nonhormonal methods. *Clin Chem*. 2019;65(1):153–160.
81. Reynolds-Wright JJ, Anderson R. Male contraception: where are we going and where have we been? *BMJ Sex Reprod Health*. 2019;45(4):236–242.
82. Griswold MD. Spermatogenesis: the commitment to meiosis. *Physiol Rev*. 2016;96(1):1–17.
83. Chung SS, Wang X, Roberts SS, et al. Oral administration of a retinoic acid receptor antagonist reversibly inhibits spermatogenesis in mice. *Endocrinology*. 2011;152(6):2492–2502.
84. Amory JK, Muller CH, Shimshoni JA, et al. Suppression of spermatogenesis by bisdichloroacetyldiamines is mediated by inhibition of testicular retinoic acid biosynthesis. *J Androl*. 2011;32(1):111–119.
85. Paik J, Haenisch M, Muller CH, et al. Inhibition of retinoic acid biosynthesis by the bisdichloroacetyldiamine WIN 18,446 markedly suppresses spermatogenesis and alters retinoid metabolism in mice. *J Biol Chem*. 2014;289(21):15104–15117.
86. Matzuk MM, McKeown MR, Filippakopoulos P, et al. Small-molecule inhibition of BRDT for male contraception. *Cell*. 2012;150(4):673–684.

87. Cheng CY, Silvestrini B, Grima J, et al. Two new male contraceptives exert their effects by depleting germ cells prematurely from the testis. *Biol Reprod*. 2001;65(2):449–461.
88. Chen H, Mruk DD, Xia W, et al. Effective delivery of male contraceptives behind the blood-testis barrier (BTB) - lesson from Adjudin. *Curr Med Chem*. 2016;23(7):701–713.
89. Li L, Tang EI, Chen H, et al. Sperm release at spermiation is regulated by changes in the organization of Actin- and microtubule-based cytoskeletons at the apical ectoplasmic specialization-a study using the Adjudin model. *Endocrinology*. 2017;158(12):4300–4316.
90. Tash JS, Attardi B, Hild SA, et al. A novel potent indazole carboxylic acid derivative blocks spermatogenesis and is contraceptive in rats after a single oral dose. *Biol Reprod*. 2008;78(6):1127–1138.
91. Yan W, Ma L, Burns KH, Matzuk MM. Haploinsufficiency of kelch-like protein homolog 10 causes infertility in male mice. *Proc Natl Acad Sci U S A*. 2004;101(20):7793–7798.
92. Zheng H, Stratton CJ, Morozumi K, et al. Lack of Spem1 causes aberrant cytoplasm removal, sperm deformation, and male infertility. *Proc Natl Acad Sci USA*. 2007;104(16):6852–6857.
93. Yuan S, Stratton CJ, Bao J, et al. Spata6 is required for normal assembly of the sperm connecting piece and tight head-tail conjunction. *Proc Natl Acad Sci USA*. 2015;112(5):E430–E439.
94. Chang Z, Qin W, Zheng H, et al. Triptonide is a reversible non-hormonal male contraceptive agent in mice and non-human primates. *Nat Commun*. 2021;12(1):1253.
95. Salicioni AM, Gervasi MG, Sosnik J, et al. Testis-specific serine kinase protein family in male fertility and as targets for non-hormonal male contraception. *Biol Reprod*. 2020;103(2):264–274.
96. Drevet JR. Epididymal approaches to male contraception. *Basic Clin Androl*. 2018;28:12.
97. Freitas MJ, Vijayaraghavan S, Fardilha M. Signaling mechanisms in mammalian sperm motility. *Biol Reprod*. 2017;96(1):2–12.
98. O'Rand MG, Hamil KG, Adevai T, Zelinski M. Inhibition of sperm motility in male macaques with EP055, a potential non-hormonal male contraceptive. *PLoS One*. 2018;13(4):e0195953.
99. O'Rand MG, Silva EJ, Hamil KG. Non-hormonal male contraception: a review and development of an Eppin based contraceptive. *Pharmacol Ther*. 2016;157:105–111.
100. Brown SG, Publicover SJ, Barratt CLR, Martins da Silva SJ. Human sperm ion channel (dys)function: implications for fertilization. *Hum Reprod Update*. 2019;25(6):758–776.
101. Lishko PV, Mannowetz N. CatSper: a unique calcium channel of the sperm flagellum. *Curr Opin Physiol*. 2018;2:109–113.
102. Esposito G, Jaiswal BS, Xie F, et al. Mice deficient for soluble adenylyl cyclase are infertile because of a severe sperm-motility defect. *Proc Natl Acad Sci USA*. 2004;101(9):2993–2998.
103. Wang D, Hu J, Bobulescu IA, et al. A sperm-specific Na+/H+ exchanger (sNHE) is critical for expression and in vivo bicarbonate regulation of the soluble adenylyl cyclase (sAC). *Proc Natl Acad Sci USA*. 2007;104(22):9325–9330.
104. Oberheide K, Puchkov D, Jentsch TJ. Loss of the Na(+)/H(+) exchanger NHE8 causes male infertility in mice by disrupting acrosome formation. *J Biol Chem*. 2017;292(26):10845–10854.
105. Syeda SS, Sánchez G, McDermott JP, et al. The Na+ and K+ transport system of sperm (ATP1A4) is essential for male fertility and an attractive target for male contraception. *Biol Reprod*. 2020;103(2):343–356.
106. Ikawa M, Inoue N, Benham AM, Okabe M. Fertilization: a sperm's journey to and interaction with the oocyte. *J Clin Invest*. 2010;120(4):984–994.
107. Okabe M. Sperm-egg interaction and fertilization: past, present, and future. *Biol Reprod*. 2018;99(1):134–146.
108. Ohto U, Ishida H, Krayukhina E, et al. Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. *Nature*. 2016;534(7608):566–569.
109. Campo-Engelstein L. Contraceptive justice: why we need a male pill. *Virtual Mentor*. 2012;14(2):146–151.
110. Nguyen BT, Shih G, Turok DK. Putting the man in contraceptive mandate. *Contraception*. 2014;89(1):3–5.
111. Heinemann K, Saad F, Wiesemes M, et al. Attitudes toward male fertility control: results of a multinational survey on four continents. *Hum Reprod*. 2005;20(2):549–556.
112. Martin CW, Anderson RA, Cheng L, et al. Potential impact of hormonal male contraception: cross-cultural implications for development of novel preparations. *Hum Reprod*. 2000;15(3):637–645.
113. Glasier AF, Anakwe R, Everington D, et al. Would women trust their partners to use a male pill? *Hum Reprod*. 2000;15(3):646–649.
114. Amory JK, Page ST, Anawalt BD, et al. Acceptability of a combination testosterone gel and depot medroxyprogesterone acetate male contraceptive regimen. *Contraception*. 2007;75(3):218–223.
115. Meriggiola MC, Cerpolini S, Bremner WJ, et al. Acceptability of an injectable male contraceptive regimen of norethisterone enanthate and testosterone undecanoate for men. *Hum Reprod*. 2006;21(8):2033–2040.
116. Roth MY, Shih G, Ilani N, et al. Acceptability of a transdermal gel-based male hormonal contraceptive in a randomized controlled trial. *Contraception*. 2014;90(4):407–412.
117. Nguyen BT, Farrant MT, Anawalt BD, et al. Acceptability of oral dimethandrolone undecanoate in a 28-day placebo-controlled trial of a hormonal male contraceptive prototype. *Contraception*. 2020;102(1):52–57.

115

Testicular Dysgenesis Syndrome and Testicular Tumors

Ewa Rajpert-De Meyts, Jorma Toppari, Katharina M. Main, Anders Juul, and Niels Jørgensen

OUTLINE

Testicular Dysgenesis Syndrome, 1918
  Epidemiology, 1918
  Biological Evidence, 1919
Cryptorchidism, 1919
  Classification, 1919
  Embryology, 1920
  Etiology, 1920
  Prevalence, 1921
  Treatment, 1921
  Long-Term Effects, 1921
Hypospadias, 1921
  Embryology, 1922
  Etiology, 1922
  Prevalence, 1922
  Treatment, 1922
  Long-term Effects, 1922
Minipuberty, 1923
  Minipuberty in Healthy Infants, 1923
  Minipuberty in Various DSD and TDS Conditions, 1923
Testicular Tumors, 1923
  Classification, 1923
  Testicular Germ Cell Tumors, 1924
  Fertility Preservation and Long-Term Effects of Testicular Cancer, 1929
  Sex Cord-Stromal Tumors of the Testis, 1930
  Other Testicular Tumors, 1931
Conclusion, 1932

TESTICULAR DYSGENESIS SYNDROME

There is growing evidence that a range of male reproductive disorders diagnosed in children and adults, including cryptorchidism, some forms of hypospadias, testicular germ cell tumors (TGCT) of young adults, and some forms of decreased spermatogenesis—with smaller testis volume and poor semen quality—are linked with disturbed fetal testicular development. Clinical and epidemiological observations in patients with these disorders, combined with biological evidence—especially from studies of preinvasive TGCT—led to the concept of a testicular dysgenesis syndrome (TDS).[1,2] The initial TDS concept has been subsequently enlarged to include additional discrete manifestations of testicular dysgenesis, such as shortened anogenital distance (AGD) and low testosterone levels (adult hypogonadism), as presented in Fig. 115.1. It is important to underline that the main concept of the TDS hypothesis is the predominantly environmental-/lifestyle-related etiology of the syndrome, demonstrated by the changing incidences of its components. On the other hand, TDS prevalence and severity are undoubtedly modulated by the genetic background. Hence, more severe TDS manifestations, especially if present together in one patient (e.g., severe hypospadias and cryptorchidism), overlap with disorders of sex development (DSD).[3] The complexity of the syndrome and its multifactorial pathogenesis led to discussion in the scientific literature, with some authors casting doubt on the existence of TDS but also acknowledging that the TDS hypothesis contributed to stimulating research on the apparently increasing problems in male reproductive health.[4] Even if the existence of the syndrome as one entity can be discussed, a general link between testicular developmental disorders and postpubertal reproductive problems has been proven beyond doubt (reviewed in[2,5]). The epidemiological and biological evidence supporting the concept of TDS is briefly reviewed later.

Further in this chapter, the individual TDS-related disorders are reviewed in detail, including the recently added mild manifestations and possible influence on hormonal changes in male minipuberty. Regarding testicular cancer, only TGCT derived from germ cell neoplasia *in situ* (GCNIS) are considered a part of TDS. However, other testicular tumors, including childhood germ cell tumors and tumors derived from somatic cells (sex cord stromal tumors), are also described in this chapter because of their frequent endocrine manifestations.

Epidemiology

Epidemiological studies of male reproductive disorders have intensified in recent decades, partly in response to a dramatic rise in the incidence of testicular cancer in many countries and partly because of the TDS hypothesis and an increased attention given to reports of the declining semen quality. The adverse trends in testis cancer incidence coinciding with declining semen quality were first reported in Nordic countries, and the suggestion was made that increasing testis cancer rates in a population can be a warning sign of other reproductive health problems in the same area.[2] Positive correlations between the incidence of testicular cancer and hypospadias and between hypospadias and cryptorchidism and a negative correlation between testicular cancer and sperm concentration were subsequently found in other registries at the population level.[6] In concert with these observations, childhood TDS conditions, for example, cryptorchidism, are risk factors for TGCT and decreased spermatogenesis in adults. Isolated hypospadias but not

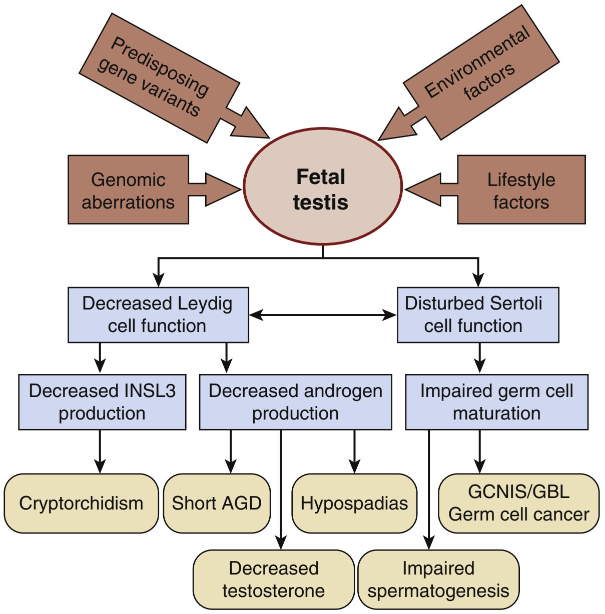

Predisposing gene variants
Environmental factors
Genomic aberrations
Fetal testis
Lifestyle factors
Decreased Leydig cell function
Disturbed Sertoli cell function
Decreased INSL3 production
Decreased androgen production
Impaired germ cell maturation
Cryptorchidism
Short AGD
Hypospadias
GCNIS/GBL Germ cell cancer
Decreased testosterone
Impaired spermatogenesis

Figure 115.1 Schematic representation of the concept of testicular dysgenesis syndrome (TDS). The TDS hypothesis stipulates that various male reproductive disorders are linked to impaired fetal testis development. The etiology is multifactorial and includes adverse environmental/lifestyle factors acting on genetically susceptible individuals. The cellular targets are mainly fetal Leydig and Sertoli cells, and the mechanisms involve testicular hormone insufficiency and failure of germ cell maturation, leading to different postnatal outcomes. *AGD,* Anogenital distance; *GCNIS,* germ cell neoplasia *in situ*; *GBL,* gonadoblastoma. (Modified and updated from Skakkebæk NE, Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects, *Hum Reprod.* 2001;6:972–978.)

cryptorchidism has also been more frequent in boys born after fertility treatment and was associated with delayed pubertal onset in the Danish population.[7] A dissociation between cryptorchidism and hypospadias was also noted in New Zealand, where indigenous populations have a greater risk of cryptorchidism and TGCT than the white population, but not hypospadias, suggesting that a combination of environmental exposures and genetic background can modulate TDS manifestations in different populations.[8]

Biological Evidence

Growing biological evidence suggests that the TDS components, despite having different age of manifestation, are characterized by impaired development of the Leydig and Sertoli cells, leading to the changed hormonal microenvironment of the testis and consequently disturbed maturation of germ cells. The strongest evidence comes from studies of the early stages of TGCT, which closely resemble fetal gonocytes.[1,9] A number of studies examining the testis tissue samples from patients with TDS demonstrated focal signs of dysgenesis, including clusters of poorly developed tubules, incompletely differentiated Sertoli cells, Leydig cells nodules, and microcalcifications.[1] Similar morphological changes have been observed in experimental animals exposed to various endocrine disrupters, especially those that affect the development of the reproductive tract during the masculinization programming window.[10] Disturbed masculinization window in rodents is reflected by a shorter AGD, which is a sex-dimorphic readout of the prenatal androgen activity.[10,11] Similarly, in human patients, shorter AGD was associated at the cohort level with hypospadias, cryptorchidism, testicular cancer and poorer semen quality,[11,12,12a] supporting the syndromic nature of these conditions and their relation to the intrauterine development.[5]

The etiology of TDS is complex and likely multifactorial, but a growing number of studies identified components of TDS in subjects exposed *in utero* to environmental endocrine disrupters, for example, maternal mid-gestation serum levels of paracetamol, bisphenol A (BPA), or n-propyl paraben were associated with cryptorchidism or a shorter AGD in infants.[5,13,14] Similar studies of human prenatal exposure to phthalates were somewhere contradictory and remain inconclusive.[15,16,17]

KEY POINTS

- Testicular dysgenesis syndrome (TDS) comprises disorders linked to disturbed testis development, including cryptorchidism, testicular germ cell tumors derived from germ cell neoplasia in situ (GCNIS), some forms of hypospadias, short anogenital distance, reduced spermatogenesis, and low testosterone levels in adulthood.
- Pathogenesis of TDS is predominantly environmental, but genetic background modulates the risk.

CRYPTORCHIDISM

Cryptorchidism, or undescended testis, is one of the most common urogenital abnormalities in boys. It is mostly found isolated, but it can also occur in combination with other malformations and syndromes. An undescended testis is often indicative of some degree of testicular dysfunction, which may have long-lasting effects on especially semen quality and the risk of testicular cancer in adulthood.

Classification

Cryptorchidism is divided into a congenital and an acquired form. Synonymous terms for these two entities are *primary* and *secondary cryptorchidism.* Traditionally, cryptorchid testes are divided into nonpalpable and palpable testes, and synonyms are *undescended testis, maldescensus testis*, and *retentio testis.* During laparoscopy or surgery, a nonpalpable testis may be completely absent (vanished) or "nubbin," which refers to an atrophic remnant of tissue at the end of vessels.[18] Ideally, the precise position of the testis along its normal pathway of descent should be described in more detail: high/low abdominal testes, inguinal (canalicular), suprascrotal (superficial inguinal pouch), high scrotal (prescrotal), and scrotal. Such a distinction would ensure a better basis for future studies about the long-term impact of treatment regimens and the disease *per se* on testis function. In rare cases, the testis can be found in an ectopic position, usually in perineum, contralateral scrotum, or femoral. Unilateral cryptorchidism occurs more frequently than bilateral cryptorchidism, and in only approximately 10% of orchiopexies, the testis is nonpalpable before surgery (Fig. 115.2). Retractile testes were considered a normal variant, but this has been challenged.[19] In severely retractile cases or in children who are distressed during the physical examination, these testes may be difficult to distinguish from undescended testes. In addition, retractile testes have an increased risk to become truly cryptorchid over time, and their position needs to be monitored.[20] Testes normally descend to the lowest position during the first 3 months after birth (minipuberty) and then ascend a little again in association with changes in serum levels of reproductive hormones and IGF1.[21]

Registry data, either of congenital malformations or hospital diagnoses after orchiopexy, may differ considerably in their selection criteria from prospective standardized studies and be biased toward the more severe forms of testicular maldescent. Cumulative orchiopexy rates may also include cases of ascended testes and encompass historical changes in referral and treatment regimens. Thus reported prevalence rates may differ even within the same population.

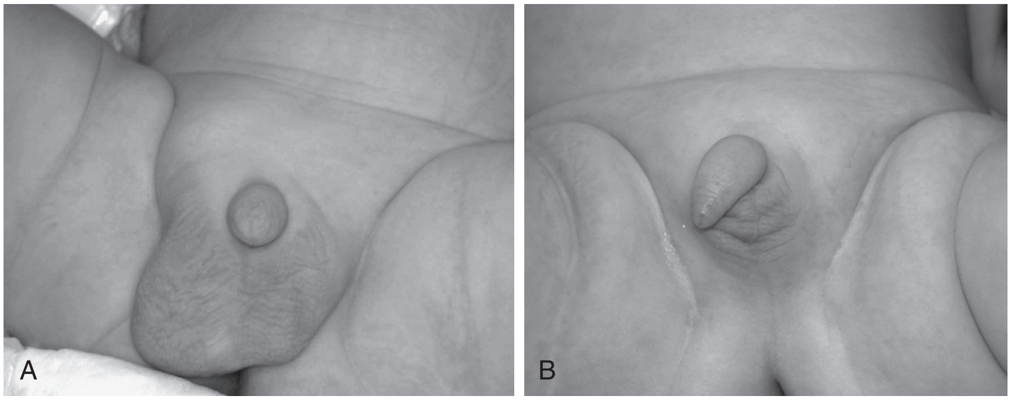


Figure 115.2 Infants with unilateral (**A**) and bilateral (**B**) cryptorchidism. It is important that the diagnosis is not based on the visual inspection but by a careful palpation of scrotum in warm and comfortable conditions.

The clinical examination can on occasion be technically difficult. It requires an experienced observer, warm and relaxed conditions, and may need to be repeated if the child is upset. Examination should be done not only in the supine position but also cross-legged or upright before a definite diagnosis is made. An experienced examiner may have a higher sensitivity for locating an undescended testis than ultrasound or other imaging techniques. The American Urological Association guidelines recommend withholding of ultrasound or other imaging before referral to a pediatric surgeon.[20] The testis is brought into the lowest possible position along its normal pathway of descent by gentle, not forced, traction and should remain there for a while after exhaustion of the cremaster reflex. An asymmetric or hypoplastic scrotum may further suggest unilateral or bilateral cryptorchidism, respectively.

The position of the testis during infancy and childhood is not necessarily stationary, and there is a consensus that boys should be examined repetitively as part of their standard health care program. A significant proportion of the testes in congenital cryptorchidism descend spontaneously during the first months of life.[22] Testes, scrotally placed at birth or retractile, can ascend during infancy and childhood. The acquired undescended testis is more frequent than previously thought, but the etiology and consequences are yet unclear.[23]

Embryology

The development of the testis and its descent from the abdomen into the scrotum is a highly complex process, with numerous genetic and hormonal factors involved.[24] Testicular descent in humans occurs in two phases, the transabdominal phase and the inguinoscrotal phase, which are usually completed around gestational weeks 15 and 35, respectively. During the first phase, the testis is anchored at the internal inguinal ring by enlargement of the gubernaculum. In rodents, the masculinization of the gubernaculum is highly dependent on insulin-like hormone 3 (INSL3), secreted by Leydig cells, and INSL3 receptor, leucine-rich repeat-containing G-protein–coupled receptor (also known as LGR8, RXFP2), whereas mutations of these genes in humans have only rarely been reported in large series of patients with cryptorchidism. The regression of the cranial suspensory ligament and the second phase of descent are highly dependent on androgens. The inguinoscrotal descent is often compromised in hormonal disorders of the pituitary–gonadal axis or steroidogenesis and androgen receptor mutations. The role of the genitofemoral nerve (GFN) and its neurotransmitter, calcitonin gene-related peptide (CGRP), in the inguinoscrotal phase of descent is less well established in humans than in rodents. However, the high prevalence of cryptorchidism in spina bifida and caudal regression syndromes indicates that GFN also plays a role in testis descent in humans.

Etiology

Cryptorchidism can be caused by a multitude of genetic and hormonal disorders, as well as complex syndromes with midline and caudal defects. In addition, lifestyle factors and the environment may have adverse effects on testicular descent in humans.[25] Disorders of the hypothalamic-pituitary-gonadal (HPG) axis and testosterone biosynthesis have more frequently been reported in association with isolated cryptorchidism than other genetic defects, such as androgen receptor mutations.[26] However, there are familial cases of cryptorchidism, and one study estimated the relative risk of cryptorchidism to be increased to 6.9 if a brother and 4.6 if the father also had cryptorchidism.[27] However, the concordance of cryptorchidism between monozygotic and dizygotic twins is rather similar (27.3% vs. 24.1%, respectively), emphasizing the role of shared pregnancy rather than heritability.[28]

Epidemiologic studies have revealed a multitude of risk factors for congenital cryptorchidism,[25] the most established being low birth weight, being small for gestational age, and prematurity. Pregnancy complications such as placental insufficiency, preeclampsia, and maternal diabetes have also been associated with cryptorchidism. Even mild gestational diabetes appears to increase the risk of cryptorchidism in the offspring, potentially via reduced levels of maternal sex–hormone-binding globulin (SHBG) and fetal hyperinsulinemia leading to an imbalance in the fetal estrogen-androgen action.[29]

Risk factors in maternal behavior, such as continuous smoking, drinking, caffeine intake, and use of mild analgesics during pregnancy, seem to have adverse effects on the development and descent of the testis in their offspring, although these findings are still controversial.[25] Some discrepancies in the results between studies are caused by the differences in the inclusion criteria related to mild and transient forms of cryptorchidism in some and only persistent or severe forms in others. Different types of cryptorchidism may not share the same etiology.

Several epidemiologic studies suggest an increased prevalence of cryptorchidism and related urogenital malformations in specific professions such as farmers and males from geographic regions with high agricultural activity.[25] In animal studies, prenatal exposure to many man-made chemicals with endocrine-disrupting properties can cause cryptorchidism. Data in humans are still sparse, although they are exposed to a multitude of chemicals over their entire lifespan. There is some evidence that halogenated compounds such as polychlorinated compounds and polybrominated flame retardants may be associated with an increased risk of congenital cryptorchidism and hypospadias.[30-33] In addition, phthalate exposure of the fetus and infant was associated with testicular hormonal function[34] and reduced AGD as a sign of reduced virilization.[15] Short AGD was also linked to cryptorchidism and hypospadias[35] and to testicular cancer and semen quality in adults.[36,12,12a] In most cases with cryptorchidism, a distinct etiology cannot be ascertained, suggesting that genetic predisposition together with environmental and lifestyle factors may exert adverse effects.

Prevalence

The reported prevalence of cryptorchidism varies considerably, depending on the type of study, geographic region, the selection, age, and ethnicity of participants, and the clinical classification. At birth, reported rates from prospective studies vary between 1.6% and 9% if premature boys are also included and between 1.8% and 8.4% for boys ≥2500-g birth weight.[22,37] Some countries, such as Denmark and Great Britain, have noted an increase of cryptorchidism over the past decades. A doubling of congenital cryptorchidism was also reported in the United States, Canada, and South America in the 1970s and 1980s. In contrast, orchiopexy data from registries suggest a decrease since the 1990s. In general, rates of cryptorchidism reported in registries tend to be lower than from prospective studies because ascertainment and registration traditions differ from systematic studies. For orchiopexies, registries may also reflect changes in treatment traditions and surgical referrals.

Owing to the spontaneous postnatal descent of retained testes during the first months of life, prevalence rates drop to between 0.9% and 1.8% at 3 months of age. During childhood, retractile testes and ascensus testis become more common, and reported prevalence data from prospective studies increase up to 7%. Most ascended testes seem to descend spontaneously during puberty, thus the age and pubertal stage at examination has a major impact on reported rates.

Treatment

Cryptorchidism should be treated early in life, preferably between 6 and 12 months of age or as soon as the diagnosis is established, in order to prevent further damage to spermatogenesis. There have been two treatment modalities available for cryptorchidism: hormonal or surgical treatment. Recent recommendations strongly favor surgery,[20,38] which offers the best balance between efficacy and side effects and improves testis growth.[39] Routine biopsy at surgery is not recommended except in cases of chromosomal abnormality or ambiguous genitalia, where there is an increased risk of gonadoblastoma or GCNIS.

The success rate (defined as a scrotal testis without atrophy) after surgical treatment in expert hands has improved over the years and is now approaching >95% for inguinal testes and 85% to 90% for abdominal testes, with slightly varying results depending on techniques and patient age. Side effects include pain, hematoma, infection, and anesthetic complications. Complications such as testicular atrophy and damage of the vas deferens are rare but become more frequent if surgery needs to be repeated due to relapse.

Hormonal treatments with human chorionic gonadotropin (hCG), gonadotropin-releasing hormone (GnRH), or luteinizing hormone–releasing hormone (LHRH) are now rarely used; their overall efficacy is only 20%, and reascent appears in approximately 25%.[40] The lower the initial testis position, the higher the rate of success. Acute side effects include pain at injection site, penile growth, pubic hair, erection pain, behavioral problems, and pain in the groin. More severely, germ cell apoptosis and inflammatory changes in the testes with the risk of long-term reduction in germ cell number and testis size in adulthood have been reported, especially for the youngest children.[41] Hormonal treatment after orchiopexy has been proposed to have beneficial effects on sperm count,[42] but this treatment regimen remains controversial and needs further scrutiny.

Treatment of retractile testes and high scrotal testes is currently not recommended because systematic studies are not yet available to assess whether treatment would be of long-term benefit. However, some retractile testes may show histologic deterioration and a reduction in volume over time.

Long-Term Effects

Impairment of spermatogenesis is a major problem of cryptorchidism. This impairment is partly related to a dysgenetic development *in utero*, but failure or delay to treat will additionally damage future semen quality and fertility. The precise adverse mechanisms that affect testicular function in untreated cryptorchid testes are not yet clear. Testes with a history of cryptorchidism grow slower and end up smaller in puberty than normal testes,[43] but there is no difference at the age of the onset of puberty. Although semen quality is reduced in unilateral cryptorchidism, reported fertility rates are usually not compromised. In contrast, bilateral cryptorchidism will lead to azoospermia if not treated. Long-term follow-up studies show that treatment before 4 years of age improves semen quality outcome. Whether the current recommendation for treatment at an even younger age will lead to further improvement of the long-term outcome remains to be seen. In a prospective study of unilateral cryptorchidism, the affected testis grew better over the first 2 years of life if orchidopexy was performed early.[39] Surgical techniques vary and may affect fertility outcome.[44,45]

Hormonal effects related to cryptorchidism are very subtle in childhood, and controversial results have been reported. At 3 months of age, increased follicle-stimulating hormone (FSH) and decreased inhibin B levels were found in patients with persistently cryptorchid testes.[46] Low serum inhibin B levels at orchiopexy may reflect a low number of germ cells at biopsy. In adults, in addition to reduced semen quality, the predominant effect of cryptorchidism is impaired Sertoli cell function, reflected in higher FSH and lower inhibin B levels. The effect on Leydig cell function, reflected by increased luteinizing hormone (LH) and low or normal testosterone level, is quite common but more pronounced if the orchiopexy was performed at a relatively late age.

Cryptorchidism increases the risk of testicular cancer, and both conditions share several prenatal and perinatal risk factors.[2,47] The relative risk of neoplasia is greatest in the undescended testis (RR = 6.3) but is also slightly increased in the contralateral, normally descended testis in unilateral cryptorchidism. Historical data suggest that orchidopexy before the age of 11 years lowered the risk of testis cancer, but the outcomes of surgery at even earlier ages require further long-term follow-up to determine if testicular cancer rates decline further.

KEY POINTS

- Cryptorchidism can be congenital or acquired, with one or both testicles positioned from high scrotal to abdominal.
- Prematurity and low birth weight are the greatest risk factors, and the prevalence varies from 1.8% to 8.4% at birth according to the geographical region to 0.9% to 1.8% at 3 months of age.
- Treatment of choice is by surgery, preferably between 6 and 12 months of age to prevent further damage to spermatogenesis.
- Cryptorchidism increases the risk of germ cell cancer (RR = 6.3).

HYPOSPADIAS

Hypospadias is a penile malformation in newborn boys and traditionally categorized into mild or distal (glanular, coronal) and severe or proximal (penile, penoscrotal, perineal) forms. Recently, another classification was proposed into four categories: *forme fruste* (incomplete), standard, severe hypospadias, and hypospadias variations.[48] The lack of a standardized nomenclature is a considerable limitation in the hypospadias literature.

Embryology

After formation of the testis from the undifferentiated gonad, the primitive phallus begins to grow from week 8 in gestation under the influence of fetal androgens, especially dihydrotestosterone. In parallel, the inner genital folds fuse in midline, and the penile urethra forms from the endodermal urethral plate in a proximal-distal direction. Thereafter, the glandular urethra is formed by distal proliferation and canalization of the urethral plate into the glans penis. The preputial skin enlarges from a dorsal fold to eventually cover the ventral surface of the glans and fuse. Hypospadias usually arises in the first trimester (gestational weeks 8 to 14), leading to a ventral positioning of the meatus and agenesis of the surrounding tissues.[49]

Very mild types of hypospadias may present with a normally shaped prepuce or with isolated chordee and normal placement of the urethral meatus. In rare cases, only a hooded foreskin is found, with a normal urethral opening. In contrast, severe hypospadias with early disturbances of urethral fusion may compromise the growth of the periurethral tissue and thereby lead to chordae and ventral curvature of the penis. In very severe cases, which are often combined with bifid scrotum and cryptorchidism, the clinical appearance of the external genitalia at birth may cause difficulties in gender assignment. Such cases need thorough assessment according to the guidelines for children with differences of sexual development, DSD.[50] A large number of complex syndromes are associated with hypospadias, that is, the Smith-Lemli-Opitz syndrome or the hand-foot-genital syndrome. Such entities should be kept in mind, especially in cases with other concomitant malformations and/or psychomotor developmental delay.

Etiology

The etiology of hypospadias remains obscure in many cases, and evidence suggests that it is a multifactorial disease with genetic, hormonal, and environmental components. Familial cases account for approximately 10% of all hypospadias, but truly monogenic disorders, such as 5α-reductase gene (*SRD5A2*) mutations and polymorphisms, MAMLD1 (CXorf6) mutations, Wilms tumor gene 1 (*WT1*) mutation or chromosomal aberrations are seldom found.[51,52] Monozygotic twinning appears to increase the risk of hypospadias by yet unknown mechanisms.[53] Fathers of sons with hypospadias present with reduced semen quality, increased need for fertility treatment, and a higher rate of genital disorders.[54] Assisted reproduction may contribute to a higher prevalence of hypospadias, potentially through the higher rate of twinning, compromised *in utero* growth, and paternal subfertility.[55]

Hypospadias shares some risk factors with other diseases of the male reproductive tract, that is, low birth weight, being a twin/triplet, being born small for gestational age and low parity.[56,57] Maternal obesity, hormonal contraception, preeclampsia, placental insufficiency and retention, weak contractions during delivery, increased rate of cesarean sections, and increased need for assisted reproductive techniques due to parental infertility have also been reported in hypospadias.

Environmental and lifestyle factors play a role in the etiology of hypospadias, that is, exposure to pesticides, residency close to hazardous waste sites, and vegetarian diets.[58,59] The use of some medications such as diethylstilbestrol and mild analgesics in pregnancy has also been linked with hypospadias.[60,61]

AGD measurement in rodents is a biomarker of prenatal androgen effect and used in toxicologic research. One study reported shorter AGD in boys born after high phthalate exposure in the second trimester of pregnancy[15] and another showed that boys with hypospadias and cryptorchidism had a significantly shorter AGD than controls.[62]

Prevalence

The reported prevalence of hypospadias ranges from 2 to 90 per 10,000 male births, depending on the geographical region, ethnicity, study design, and quality of public registries.[63] In registries, mild forms are often underreported and mild hypospadias with an intact prepuce may not be recognized at birth because of the high rate of physiologic phimosis.[56] The reported relative contribution of mild (glanular) forms to the total prevalence of hypospadias ranges from 29% to 72% and is highly dependent on the age and selection criteria for participation in studies and the completeness of recruitment. This causes considerable difficulties in comparing data across countries and over time. Nevertheless, a large metaanalysis of datasets from 27 surveillance birth defect programs around the world reported a continuous increase in the prevalence of hypospadias in the period 1980 to 2010.[64]

Treatment

Very mild types of hypospadias may not require any treatment, but severe cases will require surgical correction. The aims of surgical treatment are to extend the urethra to the glans penis to allow normal micturition, straightening of the penile shaft to allow normal intercourse, and improvement of cosmetic appearance. There is consensus that surgery should be done by experts in the field and current guidelines favor an early operation timing during infancy. However, parents may find decision-making difficult.[65] There is an ongoing debate as to whether surgery should be postponed obtaining informed consent from the patient himself.

Some surgical centers use preoperative hormone treatment with testosterone, dihydrotestosterone, hCG, or promestriene,[66] but very few clinical trials exist. While androgens induce penile growth, it remains to be seen if such treatment affects postoperative complication rates.[67] Cosmetic results with surgery are overall good, but complications such as wound infections, hematoma, urethral fistula, and strictures can cause severe problems.[68]

Long-term Effects

An increasing number of studies now focus on the consequences of hypospadias and hypospadias repair on semen quality, fertility, sexual function, cosmetic satisfaction, and quality of life. Men with hypospadias have fewer partners, fewer children, a higher need for fertility treatment, and sexual problems such as genital pain, mechanical difficulties, and ejaculation difficulties.[69] Men with severe forms of hypospadias combined with cryptorchidism may have poor semen quality, whereas most men with mild hypospadias have normal semen quality, indicating that isolated hypospadias is not necessarily associated with testicular dysgenesis.[70] There is still a considerable lack of knowledge on long-term outcomes including patient-reported outcome measures, and several countries have established international reference groups, that is, eUROGEN[71] to improve future care for patients.

KEY POINTS

- Hypospadias is the most common congenital abnormality of penis, with a prevalence ranging from 2 to 90 per 10,000 male births.
- It can be mild (glanular, coronal) or severe (penile, penoscrotal, perineal).
- Severe forms can be part of disorders/differences of sex development (DSD) or other complex syndromes.
- The mild forms are likely caused by environmental or lifestyle factors that adversely affect fetal androgen synthesis or action.
- Severe cases require surgical correction.

MINIPUBERTY

Activation of the HPG axis happens in three phases: in fetal life, in early postnatal life (minipuberty), and at puberty. Minipuberty consists of the neonatal activation of the HPG axis, mainly in the first 1 to 6 months of postnatal life.[72-74] Minipuberty has been suggested as a critical window of programming with lifelong implications. It has been proposed that minipuberty is linked to the timing of puberty and that it may be a marker of reproductive capacity in adult life, although the physiological role and clinical importance of the elevated sex steroids levels during minipuberty still remain to be elucidated. We suggest that minipuberty may be used as an early "window of opportunity" to detect and even treat reproductive disorders.

Minipuberty in Healthy Infants

At birth, serum concentrations of LH and FSH are low due to the suppressive effects of the high concentrations of placental hormones,[72,73] which are gradually cleared after approximately 1 week. This removes the restraint, leading to elevated gonadotropin levels until 6 to 9 months of age, with the exception of FSH, which remains elevated in girls for up to 3 to 4 years.[72,73]

In boys, the postnatal rise in LH and FSH is associated with increases in INSL-3 and testosterone from the Leydig cells and inhibin B and anti-Müllerian hormone (AMH) from the Sertoli cells.[75,76] Likewise, testosterone levels are low in cord blood but increase soon after birth and reach peak levels between 1 and 3 months of age, with levels reported just below adult concentrations, before declining and stabilizing after approximately 4 to 6 months at undetectable levels.[72,73] The number of Leydig cells increases concomitantly until 3 months of age, when the fetal Leydig cells go into apoptosis.[77] Similarly, the Sertoli cells develop and the number of germ cells increases in the first months of life[78] but do not express androgen receptors (ARs) during infancy. Thus the Sertoli cells produce AMH at very high concentrations during infancy, peaking at around 3 months of age, even in the presence of high levels of testosterone.[79]

Minipuberty and the associated sex steroid levels are important for genital development in boys. Thus penile growth is positively associated with testosterone levels from birth until 3 years of age. Likewise, minipuberty is associated with testicular growth due to FSH stimulation of Sertoli cells and seminiferous tubule volume. A positive correlation between FSH levels and testicular growth has been found, with testicular volume increasing from 0.27 to 0.44 cm from birth until 5 to 6 months and then decreasing to 0.31 cm at 9 months of age.[80]

Minipuberty in Various DSD and TDS Conditions

The HPG axis can readily be evaluated in newborns suspected of DSD conditions during minipuberty. A few studies on reproductive hormones in infants with DSD conditions exist. Infants with congenital hypogonadotropic hypogonadism (CHH) have extremely low FSH, LH, and inhibin B levels in minipuberty,[81] and the postnatal surge is absent in boys with CHH, who often present with cryptorchidism with or without micropenis. By contrast, hypospadias rarely occurs in CHH because first-trimester placental hCG secretion ensure sufficient androgen production needed for early urethral closure, whereas testicular descent and penile growth are primarily dependent on sufficient testosterone and INSL3 secretion in the second half of pregnancy.[76,81] In newborns with cryptorchidism, reproductive hormones during minipuberty may be altered. Thus higher FSH and LH levels and reduced inhibin B and INSL3 have been demonstrated in cryptorchid boys.[46,76] Newborn boys with nonpalpable testes remain a diagnostic challenge, and it is difficult to distinguish boys with anorchia who do not have residual testicular tissue from boys with bilateral cryptorchidism with intraabdominal testes. However, serum concentrations of LH, FSH, AMH, inhibin B, and testosterone hormone levels in minipuberty can be used as a clinical tool to determine the presence of functional testicular tissue in infant boys.[82]

Klinefelter syndrome (47,XXY) is characterized by primary testicular failure and infertility in adulthood, whereas data on the HPG axis during minipuberty are sparse and contradicting. However, a pivotal study showed that testosterone concentrations (measured by liquid chromatography/tandem mass spectrometry [LC-MS/MS]) in infants with Klinefelter syndrome were lower than age-matched controls, while INSL3 levels were normal.[83] Infants with AR mutations resulting in complete androgen-insensitivity syndrome (AIS) have lowered LH and testosterone levels in minipuberty, whereas infants with partial AIS have normal or high LH and testosterone levels, suggesting some physiological role of androgen signaling in in the HPG axis activation during minipuberty.

In conclusion, the regulation of the HPG axis is quite complex with many biological and environmental factors playing a part. The mechanisms behind the quiescence of the HPG axis between minipuberty and puberty remain to be elucidated. Nevertheless, minipuberty allows for the maturation of sexual organs and forms a platform for future fertility. It provides a window of opportunity in the early detection of disorders of sexual development, offering the possibility of initiating treatment at an early age in some cases.

KEY POINTS

- Minipuberty is a term for the transient activation of the HPG hormone axis during the first months of postnatal life (mainly between 1 and 6 months).
- This period is positively associated with penile and testicular growth, and maturation of germ cells, and may be a marker of reproductive capacity in adult life.
- Abnormal hormone levels at minipuberty can potentially alert about TDS.

TESTICULAR TUMORS

It is important that endocrinologists keep in mind a possibility of neoplasia in patients with fertility problems or other reproductive disorders, especially in view that testicular tumors have become more common in recent decades. In the vast majority of cases, the neoplasm is derived from germ cells, and most of these tumors occur in young men. The rapidly changing incidence trends suggestive of predominantly environmental causative factors and links to development have led to TGCT being considered a part of TDS. Tumors derived from testicular somatic cells, termed sex cord stromal neoplasms, which include Leydig cells, Sertoli cells, and other rare types, are not considered to be a part of TDS.[1,2] Regardless of the pathogenesis, all types of overt testicular neoplasms may cause endocrine effects, and the management nearly always causes some degree of late endocrine complications.

Classification

The latest classification proposed by the World Health Organization in 2016 is presented in a slightly simplified form in Table 115.1.[84] TGCT are divided into two major groups according to their association (or lack thereof) with GCNIS. The most common TGCT that typically occur in young adult men, seminoma and nonseminoma, are derived from GCNIS—and these tumors are considered a part of TDS. Infantile/childhood TGCT and spermatocytic tumor of older men are not associated with GCNIS.

Testicular tumors derived from somatic cells are jointly termed sex cord-stromal tumors and include quite common Leydig cell tumors

TABLE 115.1 Classification of Testicular Tumors, Based on the WHO Consensus (2016)

Testicular Germ Cell Tumors (TGCT)

- **TGCT derived from germ cell neoplasia *in situ* (GCNIS)**
  - Preinvasive lesions
    - GCNIS
    - Specific forms of intratubular germ cell neoplasia
  - Seminoma, pure
  - Nonseminomatous TGCT, pure
    - Embryonal carcinoma
    - Yolk sac tumor, postpubertal type
    - Trophoblastic tumors, including choriocarcinoma
    - Teratoma, postpubertal type
  - Nonseminomatous TGCT of more than one type (mixed)
- **TGCT unrelated to GCNIS**
  - Spermatocytic tumor
  - Childhood (infantile/pediatric) tumors
    - Teratoma, prepubertal type, including (epi)dermoid cysts
    - Yolk sac tumor, prepubertal type
    - Mixed prepubertal-type tumors

Sex Cord-Stromal Tumors (SCT) of the Testis

- **Leydig cell tumors (malignant or benign)**
- Sertoli cell tumors
  - Sertoli cell tumor, NOS (not otherwise specified)
  - Malignant SCT
  - Large cell calcifying SCT
  - Intratubular large cell hyalinizing SCT
- **Granulosa cell tumors**
  - Juvenile-type granulosa cell tumors
  - Adult-type granulosa cell tumors
- **The fibroma-thecoma tumors**
- **Mixed and unclassified sex cord-stromal tumors**
- **Unclassified sex cord-stromal tumors**

Tumors Containing Both Germ Cells and Sex Cord Stromal Elements

- **Gonadoblastoma**

Other tumors and malignancies

- **Miscellaneous** (including xanthogranuloma, hemangioma)
- **Hematolymphoid tumors**
  - Large B-cell, NK/T-cell or follicular lymphoma
  - Plasmacytoma
  - Myeloid sarcoma
- **Tumors of collecting ducts and rete testis**
  - Adenoma or adenocarcinoma

and rare tumors derived from Sertoli cells, granulosa cells, and other cell types. A similar to GCNIS precursor lesion, gonadoblastoma, which occurs in individuals with DSD and the presence of the Y-chromosome material, is currently classified as a separate entity grouping both germ cells and somatic cells, but this has been discussed because of a strong overlap with GCNIS. Finally, it is also important to remember that the testis can be a site of hematolymphoid malignancies.

Testicular Germ Cell Tumors

Tumors derived from germ cells account for an estimated 90% to 95% of cases of testicular cancer and are by far the most frequent neoplasms of the testis. TGCT can occur in any age: they are rare in infants and toddlers but common in postpubertal adolescents and young men, with subsequently decreasing prevalence with age, which is unusual for solid tumors. TGCT of each age group are derived from different stages of germ cell maturation and have very distinct pathogenesis (Fig. 115.3). TGCT are characterized by a striking morphologic variety, with teratomas mimicking any cell type of the body, including undifferentiated and even extraembryonic elements.[85,86] The majority of TGCTs are extremely radio- and chemosensitive, except the highly differentiated mature teratomas, hence, survival prognosis is excellent and mortality is relatively low.

TGCT Derived From GCNIS. This major group of TGCT, comprising seminomas and nonseminomas, is characterized by the origin from a common precursor lesion GCNIS, previously known as *carcinoma in situ testis*.[87] GCNIS is a prototypic sign of TDS, because of its biological features of an undifferentiated germ cell.

Epidemiology and Risk Groups. In accordance with the origin from the common precursor, the relative distribution of seminomas and nonseminomas is about half and half, but the two main types differ somewhat regarding their age-specific peak incidence. Seminoma peaks around 30 to 35 years of age, while nonseminoma occurs in relatively younger men and peaks around 20 to 25 years. Even if the incidence of testicular cancer is much lower than other cancers (e.g., prostate or lung cancers), TGCT is the most common cancer in young White American men, while among young Black men, it is close second after colorectal cancer.[88] The incidence rates are highest in some areas of Europe (>10/100,000) and lowest (<0.5/100,000) in African and East-Asian populations.[89] The incidence has been rising in recent decades but with marked geographic differences. The highest rates have been seen in Northern and Western Europe, but the incidence has been stabilizing in this region.[89,90] Conversely, steep increases have been seen among men in Southern-Eastern Europe, which previously has been a low incidence area, and among the Latin/Hispanic populations in the Americas.[89]

Cryptorchidism is a strong risk factor for testicular cancer, and approximately 5% of patients with a history of the undescended testis would develop a TGCT. The two conditions have an almost identical pattern of risk factors, consistent with a pathogenetic link between them.[47] Most of the risk factors are operating pre- or perinatally: premature birth, birth weight, birth order, high maternal age or body weight of the mother, neonatal jaundice, inguinal hernia.[91] There is a relative paucity of occupational and adult lifestyle-related risk factors, and no strong associations have been found, except that heavy smokers of marijuana have a slightly increased risk of nonseminoma (but not seminoma).[92] Tall stature during childhood and in adulthood have also been associated with TGCT risk.[93]

Individuals with decreased AGD[12b], developmental abnormalities of the gonads and DSD are at risk of germ cell neoplasia, but the risk is variable depending on the diagnosis, with the mildest forms of DSD overlapping with TDS.[3] Sex chromosome aneuploidy and structural abnormalities of chromosome Y carry the greatest risk of GCT.[50,94] Individuals with AIS are also at risk of TGCT, but those with the complete AIS (usually phenotypic females) rarely develop overt tumors.[95]

Patients with a unilateral TGCT are at increased risk for a tumor in the contralateral testis, especially if the remaining testis is small.[96] Among other phenotypic associations or comorbidities of importance for endocrinologists and andrologists, patients with testicular atrophy and impaired spermatogenesis have increased TGCT risk and should be carefully screened for malignancy.[97,98]

Genetics. TGCT have a relatively high heritability but are characterized by complex polygenic background. The familial risk of testicular GCT is quite strong: brothers and sons of cases have, respectively, 8- to 10-fold and 4- to 6-fold increased risk. However, a

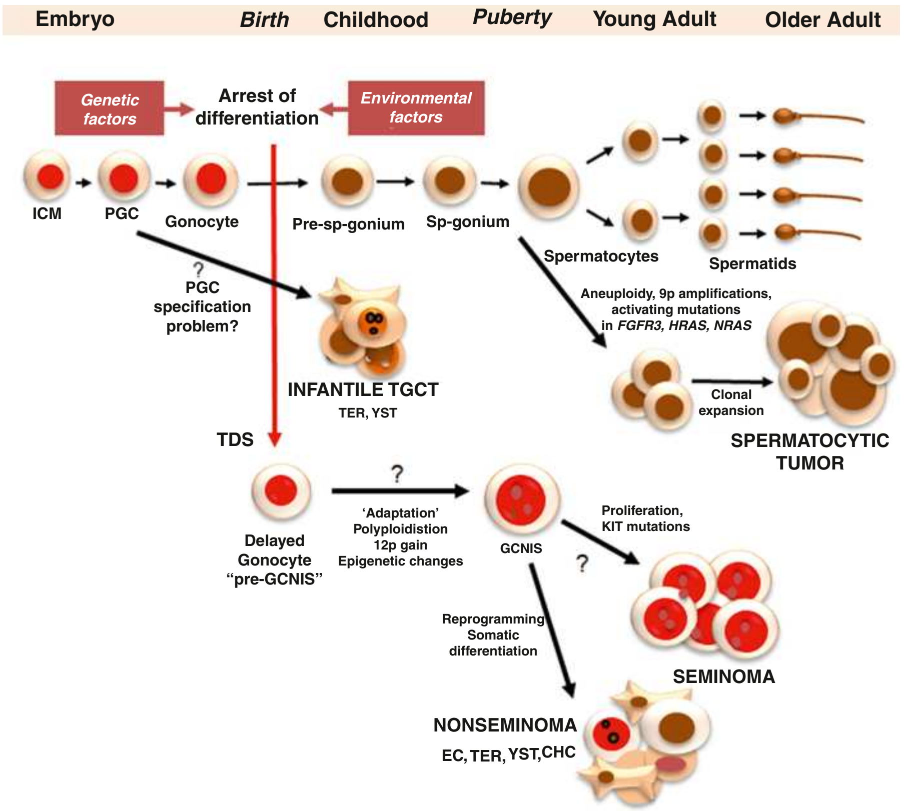

Embryo
Birth
Childhood
Puberty
Young Adult
Older Adult
Genetic factors
Arrest of differentiation
Environmental factors
ICM
PGC
Gonocyte
Pre-sp-gonium
Sp-gonium
Spermatocytes
Spermatids
? PGC specification problem?
INFANTILE TGCT
TER, YST
Aneuploidy, 9p amplifications, activating mutations in *FGFR3, HRAS, NRAS*
Clonal expansion
SPERMATOCYTIC TUMOR
TDS
Delayed Gonocyte "pre-GCNIS"
?
'Adaptation' Polyploidistion 12p gain Epigenetic changes
GCNIS
Proliferation, KIT mutations
?
SEMINOMA
Reprogramming Somatic differentiation
NONSEMINOMA
EC, TER, YST, CHC

Figure 115.3 Histogenesis of testicular germ cell tumors *(TGCT)* in relation to the normal development of germ cells over the lifetime (shown at the *top* part of the picture). Cell types expressing embryonic pluripotency factors are in *red*. The most common TGCT of young adults are considered a part of testicular dysgenesis syndrome (TDS), derived from fetal germ cells (delayed gonocytes) via the stage of germ cell neoplasia *in situ (GCNIS)*, which progresses to invasive tumors (seminoma and nonseminoma) in young adulthood. Rare TGCTs that are unrelated to GCNIS include infantile germ cell tumors derived from primordial germ cells (PGC) and spermatocytic tumor of the older men, which originates from mature spermatogonia /early primary spermatocytes. *CHC,* choriocarcinoma; *ICM,* inner cell mass; *EC,* embryonal carcinoma; *TER,* teratoma; *YST,* yolk sac tumor. (Updated from Rajpert-De Meyts E, Developmental model for the pathogenesis of testicular carcinoma in situ: genetic and environmental aspects. *Hum Reprod Update.* 2006;12:303–323.)

specific oncogenic driver mutation, similar to *BRCA1* for breast cancer, has never been identified, and this was confirmed by large whole-exome sequencing studies.[99] Invasive TGCT are characterized by polyploidization and a pathognomonic amplification of chromosome 12p.[86] These tumors also develop secondary adapting mutations; gain-of-function *KIT* mutations occur mainly in seminomas, while *KRAS* mutations have been mainly detected in nonseminomas.[100,101]

Genome-wide association studies (GWAS) identified a large number of genetic variants significantly associated with testicular cancer. Among the significant loci and candidate genes, several are associated with germ cell development, *KITLG, SPRY4, DMRT1*, *DAZL*, *PRDM14*, but also several informative loci have implicated the centrosome cycle, telomerase, and DNA-repair pathways.[102,103] It is interesting from the biological point of view that a large proportion of the predisposing genetic markers implicate pathways involved in sex differentiation and germ cell development. The strongest genetic marker, *KITLG*, encodes a tyrosine kinase receptor KIT, which has been long known as essential survival factor for germ cells and for their interaction with Sertoli cells (which express KITLG).

Etiology and Environmental Factors. The etiology of TGCT remains unclear. The changing incidence in well-developed countries, with striking geographic differences, and the epidemiologic association with other components of TDS jointly suggest a predominant influence of environmental factors acting prenatally or in early infancy. This hypothesis is supported by studies of the aforementioned risk factors, a greater risk for brothers than for sons of the cases and by a consistent association of the prevalence rates with the birth calendar year (birth cohort effect) and with the geographic location at birth.[2,90]

The link to early development is consistent with the biological features of the GCNIS, which is similar to fetal gonocytes.[1,9] Clinical and experimental evidence suggest that an imbalance in the hormonal and paracrine microenvironment of the developing gonad, in particular a disturbance of the differentiation and cross talk between germ cells and somatic cells (fetal/perinatal Sertoli cells and fetal Leydig cells), can lead to testicular dysgenesis and arrested development of gonocytes.[1,10,104] Possible mechanisms can be deduced from studies of DSD with specific genetic defects, for example, *AR*, *SRY* mutations, or 45,X/46,XY aneuploidy. But the causative lifestyle or environmental factors that can mimic these effects remain largely obscure. Since a couple of decades, attention has been on environmental endocrine disrupters, but few conclusive results were obtained, except weak associations with maternal exposure to organochlorides and organohalogens.[104a] Further studies in large and well-characterized cohorts are needed to identify the causative factors, and these studies need to take into account the constellation of genetic predisposing variants of the studied individuals, which undoubtedly modulate the risk.

Histopathology and Tumor Markers. The preinvasive GCNIS cells are located inside seminiferous tubules, with usually suppressed

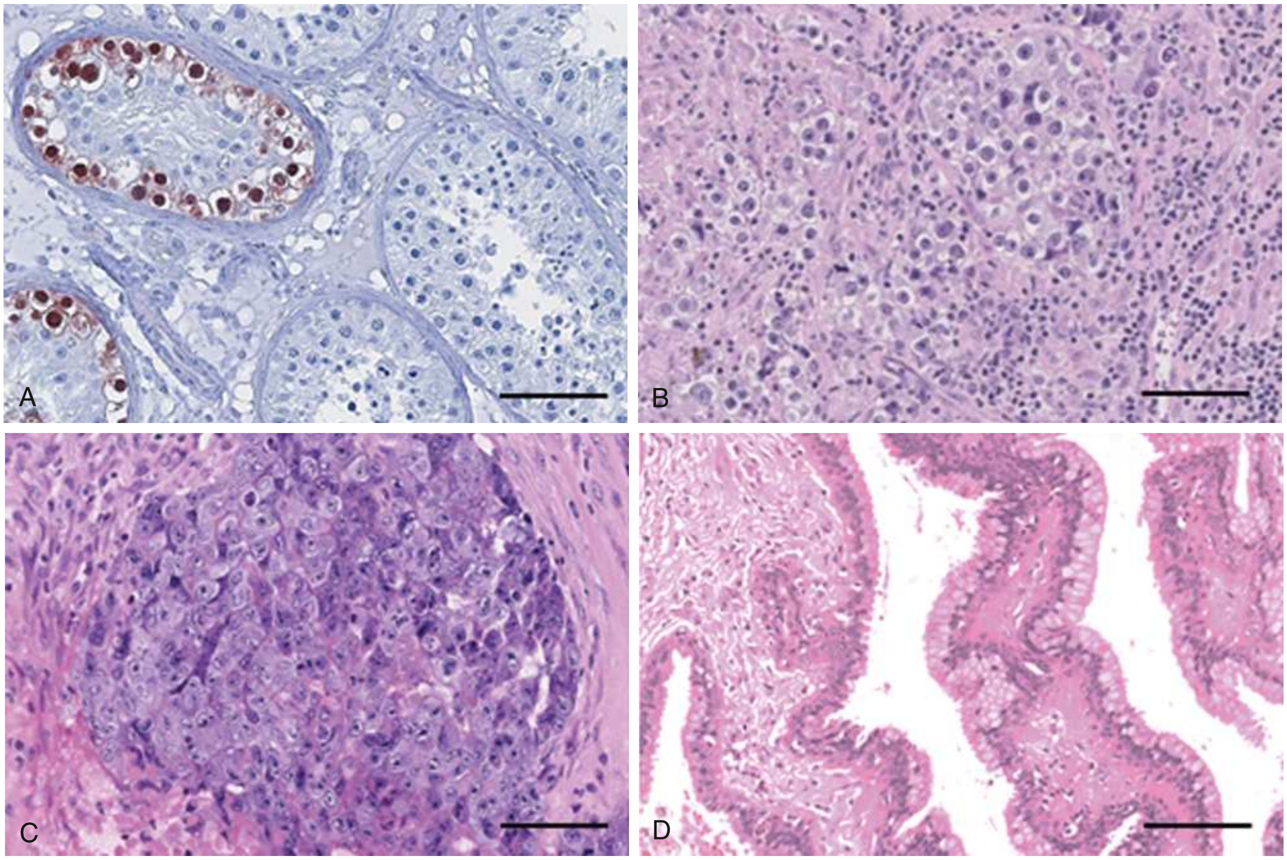


Figure 115.4 Histology of germ cell neoplasia *in situ* (GCNIS) and typical germ cell tumors. **A**, Tubules with GCNIS shown side by side with healthy tubules containing ongoing spermatogenesis. GCNIS cells are stained immunohistochemically for OCT4 (nuclear pluripotency factor). **B**, Seminoma with tumor cells closely resembling GCNIS cells and typical prominent infiltration with immune cells. **C**, Embryonal carcinoma, the undifferentiated component of nonseminomas. **D**, Teratoma, the most differentiated component of nonseminomas, mimicking morphology of somatic tissues. All bars 50 μm.

spermatogenesis, in the place normally occupied by spermatogonia and can be distinguished by large and irregular nuclei (Fig. 115.4). In testes with GCNIS, the tubular structure is usually preserved, but some dysgenetic features are frequently seen, such as undifferentiated Sertoli cells or intratubular hyaline bodies (also visible as microlithiasis in the scrotal ultrasound) (Fig. 115.5). Gonadoblastoma, found predominantly in individuals with DSD, is composed of tumor germ cells (similar to GCNIS cells) intertwined with immature somatic cells resembling Sertoli or granulosa cells, and the tubular structure is usually distorted. The presence of gonadoblastoma or GCNIS depends on the relative degree of virilization of Sertoli cells surrounding the germ cells; the less virilized the cells are, the more likely is the presence of gonadoblastoma-like structures but the two entities can overlap in some patients.[3,87]

GCNIS cells (and to a large extent seminoma tumor cells) share the gene expression profile with a subset of fetal and early infantile gonocytes, which retain embryonic stem cell-related pluripotency markers, such as OCT4/*POU5F4*, NANOG, AP2 gamma/TFAP2C, LIN28, but also express immature germ–cell-specific genes, for example, placental-like alkaline phosphatase (PLAP).[9] This expression profile, normally not present in normal adult germ cells, can be exploited for diagnosis by immuno- or cytochemistry in the adult testis or semen smears.[105]

Seminoma tumor cells resemble GCNIS cells and grow as a relatively homogeneous tumor, usually with a pronounced infiltration by immune cells[106], as shown in Fig. 115.4. Seminoma can secret lactate dehydrogenase (LDH, LD-1), a serum tumor marker used in the clinics. Nonseminomas display a variety of histologic forms, ranging from undifferentiated embryonal carcinoma to mature teratoma (see Fig. 115.4), and can contain clusters of seminoma.[85] Because of this heterogeneity, there is no uniform expression profile, and each component has to be investigated separately.[105] Among nonseminomatous components, two extraembryonic types deserve special note because of their endocrine activity. These are choriocarcinoma, which resembles gestational trophoblast and produces large quantities of hCG, preferentially in a hyperglycosylated form, and yolk sac tumors (YSTs), which are similar in morphology to the embryonic yolk sac and secrete α-fetoprotein (AFP). Both hormones, and LDH are commonly used in the clinics as *biochemical serum tumor markers*, currently indispensable for TGCT diagnosis and staging. However, TGCT rarely occur in pure histologic forms, and the presence of just a few giant cells in a seminoma may be sufficient to detect hCG in serum, hampering the specificity of this marker for nonseminomas.[107] A relatively recently discovered *specific micro-RNA* (miRNA) secreted by malignant TGCT (also in children) are very promising, and commercial assays are under development in several countries.[107,108] Clinical studies suggest that miR371a-3p is currently the most robust marker, but it has a drawback of not being secreted by teratomas.[109,110]

Diagnosis, Staging, and Management. **Preinvasive Lesions.** GCNIS is usually asymptomatic, so diagnosis at this early stage

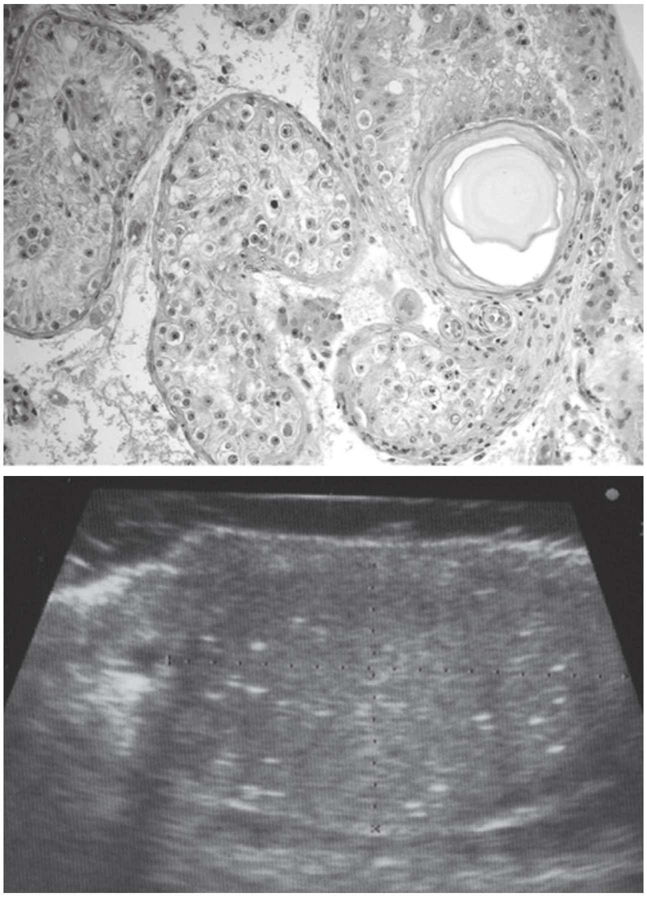


Figure 115.5 Signs of testicular dysgenesis revealed by testicular biopsy or scrotal ultrasound examination. *Top,* Germ cell neoplasia *in situ* and microcalcifications (microliths) in a biopsy of a young adult man. *Bottom,* Ultrasound image of testicular microlithiasis.

is sporadic and happens mainly in men with fertility problems or men with a TGCT who are screened by a contralateral biopsy. Ultrasonographic examination is helpful in many cases. Testicular dysgenesis and GCNIS or incipient TGCT are frequently associated with an irregular echo pattern or microcalcifications (see Fig. 115.5). It is important to remember that ultrasonic microlithiasis can be present in mild disorders or in normal men. A testicular biopsy is indicated only if other risk factors are present, for example, history of cryptorchidism, atrophic testis, and poor semen quality.[111] If a biopsy is performed, technical aspects and immunohistochemical staining for specific markers are important. The presence of genital malformations and abnormal karyotype in a patient with focal microlithiasis is suggestive of a DSD and gonadoblastoma, and these patients require additional genetic investigation.

GCNIS is present in the testis contralateral to a testicular tumor in approximately 4% to 8% of cases. Therefore, many European centers advocate performing a contralateral surgical biopsy simultaneously with an orchidectomy procedure, especially in men with small testes (<12 mL) and history of cryptorchidism, and only in young men below 40 years of age.[96,112] To increase the chance of GCNIS detection, some centers suggest taking two biopsies.[112a] In the United States, contralateral biopsies are usually not performed.

The management of GCNIS depends on a specific situation of the patient. If GCNIS is found in one testicle, without signs of malignancy in the other testis, the patient is treated by orchiectomy and surveillance. If GCNIS is detected in the testis contralateral to a TGCT or is bilateral, low-dose radiotherapy is currently preferred treatment, because the patient can avoid bilateral castration and retain hormone production. Radiation will destroy both neoplastic and normal germ cells, so radiotherapy can be delayed in patients who wish to secure biological fatherhood, and surveillance is practiced in some centers. The patient must be informed about the risk of progression to invasive cancer and should undergo careful monitoring with ultrasound examinations. Chemotherapy is not recommended to treat GCNIS, because it is not successful in a high proportion of patients. However, chemotherapy for the primary disseminated TGCT will decrease the prevalence of bilateral TGCT, apparently by eradicating GCNIS is some cases.[113] All patients with GCNIS should be offered semen analysis and cryopreservation of sperm and they should be monitored after treatment for androgen insufficiency and hypogonadism.

Overt Tumors (Seminoma and Nonseminoma). The most common manifestation of an overt testicular cancer is a scrotal mass, with pain reported by only a few patients. In the differential diagnosis, it is important to distinguish between the TGCTs and other benign malignancies. Other conditions ought to be considered, such as orchitis, epididymitis, hydrocele, varicocele, and inguinal hernia. In a small percentage of cases of TGCT, the initial symptoms are those of metastatic disease. These symptoms are usually uncharacteristic and may include lumbar pain, a palpable abdominal mass, supraclavicular lymph node enlargement, and rarely, pulmonary symptoms.

Ultrasound examination should be performed in each patient with a suspicion of TGCT. Seminomas are more homogeneous, whereas nonseminomas are heterogeneous with frequently occurring calcified hyperechoic areas. Cystic intratesticular lesions are usually benign, while solid and mixed echogenicity lesions are usually malignant.

Prior to treatment, all patients with TGCT should be offered andrological consultation, semen analysis, and cryopreservation of sperm. Endocrinological evaluation of the patient is recommended, including serum testosterone, gonadotropins, and inhibin B, if possible. Differential diagnosis of TGCT requires measurements of circulating tumor markers, which should be performed prior to the histological analysis of the tumor, which though is the decisive diagnostic procedure. Careful staging is necessary in each case to determine the most appropriate treatment strategy.[114,115] Possible spread of disease has to be evaluated by various imaging procedures such as retroperitoneal computed tomography (CT) or magnetic resonance imaging (MRI) scans, bipedal lymphangiography, thoracic radiography, or ultrasonography.[115,116] Histopathological evaluation of the orchidectomy specimen is indispensable for the staging. The important risk factors include tumor size, vascular/lymphatic or rete testis invasion, and the presence of pure embryonal carcinoma.[117] These procedures help classify patients into three stages according to the TNM (tumor, node, metastases) System of the International Germ Cell Cancer Collaborative Group and the American Joint Committee on Cancer[118] presented in Table 115.2.

As far as the treatment of TGCT is concerned, radical orchiectomy remains the primary method. Primary dissection of retroperitoneal lymph nodes (RPLND), previously commonly practiced, is currently performed only in selected patients by nerve-sparing techniques, preferably in high-volume experienced centers.[119] Postsurgical management is variable, depending on the histologic type of tumor, levels of serum markers, stage of disease, and the presence of residual retroperitoneal masses.[112,115,116] For details concerning the treatment of testicular tumors, the reader should consult specialist oncologic literature.

A surveillance strategy is currently preferred for patients in clinical stage I seminoma, which has a good prognosis. Nonseminomas have a relapse rate of around 30% and a poorer than seminoma prognosis, hence, even though surveillance is currently advised,[116,120] some

TABLE 115.2 Stage Grouping of Testicular Germ Cell Tumors According to TNM Classification

Stage 0 (pTis): Germ cell neoplasia *in situ*
Stage IA (pT1, N0, M0, S0): Tumor limited to testis and epididymis
Stage IB (pT2-4): as IA but with vascular/lymphatic, tunica or scrotal invasion
Stage II (any pT, N1-3, M0, S0-1): Metastasis in lymph nodes, serum markers normal or moderately increased
  IIA (N1): lymph nodes <2 cm
  IIB (N2, S1): lymph nodes >2 cm but <5 cm
  IIC (N3, S1): lymph nodes >5 cm
Stage III (any pT, any N, M1, S0-3): Distant metastasis (spread beyond regional nodes)
  IIIA (M1a, S0-1): Spread to non-regional nodes or lung
  IIIB (M1a, S2): as IIIA but high serum markers
  IIIC (M1a-b, S3): distant metastasis to sites other than IIIA, or very high serum markers

M, Distant metastasis; *N,* regional lymph nodes, *pT,* primary tumor; *S,* serum tumor markers (X = unknown, not available).
Adapted from Sobin LH, Gospodarowicz M, Wittekind C, eds. *TNM Classification of Malignant Tumours, UICC International Union Against Cancer*. 7th ed. Hoboken, NJ: Wiley-Blackwell; 2009: 249–254; Moch H, Humphrey PA, Ulbright TM, Reuter VE, eds. *WHO Classification of Tumours of the Urinary System and Male Genital Organs*. 4th ed. Lyon: IARC Press; 2016.

centers opt for more aggressive treatment protocol, with one course of adjuvant chemotherapy.[121] Seminomas with increased serum hCG > 200 IU should be treated as nonseminomas.

KEY POINTS

- Seminoma and nonseminoma are the most frequent types of testicular germ cell tumors (TGCT) and occur in young adults.
- These TGCT are derived from germ cell neoplasia *in situ* (GCNIS), which has characteristics of fetal gonocytes (oncoantigens, fetal miRNAs), and this expression profile helps in diagnosis.
- The incidence of GCNIS-derived TGCT depends on ethnicity and geographic location and has been increasing worldwide, consistent with the action of environmental factors acting on genetically susceptible individuals.
- GCNIS is often found in testis biopsies from men with infertility or history of cryptorchidism.
- TGCT are highly curable by radical orchiectomy (most cases) or cisplatin-based chemotherapy regimens in disseminated cases.

The most common therapy of disseminated disease after orchiectomy is a systemic combination chemotherapy with cytotoxic drugs, and the standard first-line regimen is BEP (bleomycin, etoposide, and cisplatin), administered for 3 or 4 cycles, depending on the patient's prognosis.[115,116] Dynamic monitoring of serum markers is obligatory for accurate prognosis and assessment of treatment, especially in metastatic disease. Endocrinological/andrological follow-up should be coordinated with the postsurgery oncological management. TGCTs are in general sensitive to both chemotherapy and radiotherapy and have become a model of a highly curable malignant disease, with more than 80% of patients reaching a sustained complete remission. The majority of relapses occur within 2 years after primary treatment, but some relapse can occur much later, therefore lifetime follow-up is advocated by some centers. In patients with resistance to chemotherapy and poor prognosis, salvage regimens and complex surgery for residual tumors need to be performed in specialized oncology tertiary centers.

Pediatric (Infantile) Germ Cell Tumors

Biological Features. This group includes germ cell tumors of early childhood, also called infantile TGCT. These tumors are diagnosed in boys from birth to 4 years, with occasional cases until 6 years of age, and rare cases even later. No increase in the incidence of childhood TGCT has been observed.[122] The childhood TGCT are not associated with GCNIS and present two main histological types (see Table 115.1): mature teratoma (including dermoid and epidermoid cysts) and pure YST, termed prepubertal type to distinguish them from similar TGCT that occurs in adults.[85] Prepubertal-type teratomas have also been described in very rare cases of young adults. In contrast to GCNIS-associated tumors, there is no isochromosome 12p in the tumor karyotype, which is pathognomonic for TGCTs of young adults. These distinct biological features are consistent with a different pathogenesis of childhood TGCTs, without association with testicular dysgenesis. The childhood TGCT are apparently caused by yet unknown gene aberrations/mutations that affect differentiation of primordial germ cells.[86,123]

Diagnosis and Management. Pediatric teratomas are well differentiated and benign in clinical course. There are no specific markers, so the tumor is diagnosed by its characteristic morphology. YST is malignant and in about 20% of cases is metastatic, so an early diagnosis is important.[124] The most helpful in differential diagnosis is serum tumor marker, AFP. Importantly, neonates have high physiological levels of AFP, which persist until approximately 2 years of age, when they fall to normal "adult" levels (<10 kU/L or around 100 ng/mL), so dynamic assessment of measurements is needed. AFP can also be detected in tumor tissue, together with SALL4 as an additional useful immunohistochemical marker for YST.[125] A novel blood test, based on the detection of GCT-specific miR-371-3 cluster, is promising for diagnosis of YST, but not mature teratomas.[107,123]

Orchiectomy is a treatment of choice for malignant YST, whereas clinically benign tumors, mature teratoma and epidermoid/dermoid cysts, which are never associated with GCNIS, should be treated by testis-sparing surgery. Postsurgical histopathological description of the tumor, possible extragonadal spread, and AFP levels are used to classify each child into a clinical stage, using the staging system recommended by the Children Oncology Group (USA) and World Health Organization (WHO).[84,126] For chemotherapy regimens used to treat children with metastatic YST, the readers should consult specialized pediatric oncology literature.

Spermatocytic Tumor

This very rare tumor (<1% of TGCT in adults), previously called spermatocytic seminoma, is the only TGCT without an extragonadal or ovarian counterpart.[85,86] Spermatocytic tumor occurs in relatively older men, with median age at diagnosis of around 52 to 56 years.[115,127] This tumor is not associated with GCNIS and has a different pathogenesis than tumors of young adults, which are a part of TDS. Spermatocytic tumors grow from clonally expanding mature spermatogonia or early primary spermatocytes through a stage of intratubular clusters of increasingly aneuploid germ cells. Hence, immunohistochemical markers of spermatogonia, for example, MAGE-A4 or SSX, are helpful in differential diagnosis.[105] Spermatogonial proliferation is caused by *de novo* genetic aberrations, including rare gain-of-function mutations in genes promoting survival of spermatogonia, especially in the RAS-MAKP signaling pathway (*FGFR2, FGFR3, PTPN11, HRAS, NRAS*) or large chromosomal gains of chr9 and chr20 combined with a loss of chr7.[128,129] These mutations are associated with aging of germ cells (the so-called "selfish mutations") and if transmitted to next generation by older fathers, can cause severe diseases in children, such as thanatophoric dysplasia, achondroplasia, Noonan syndrome, or Costello syndrome.[128]

The clinical course of spermatocytic tumors is benign and asymptomatic in most cases, and the tumors can grow quite large. Only up to 7% of the patients can either develop metastatic disease or their tumor undergoes malignant transformation to rhabdomyoid sarcoma, and these complications are associated with a poor prognosis.[127] Treatment of choice is radical orchiectomy.

KEY POINTS

- Two rare types of testicular germ cell tumors (TGCT) are not derived from GCNIS: childhood TGCT (malignant yolk sac tumor and benign teratomas) and spermatocytic tumor of older men (median age of diagnosis around 50 years).
- These tumors are treated by orchiectomy, except some benign tumors in children.

Fertility Preservation and Long-Term Effects of Testicular Cancer

Fertility Preservation. Testicular dysfunction and imbalance in reproductive hormones are an important clinical problem, especially in view of the fact that the vast majority of patients with testicular cancer are of reproductive age. Adult patients with TGCT associated with GCNIS and TDS have in general decreased fertility and poor spermatogenesis even before the overt tumor has developed.[130] Abnormalities include smaller testicular volume, elevated serum FSH levels, frequent oligozoospermia, and in some cases microlithiasis in the ultrasonic examination. A variable degree of testicular dysgenesis, undifferentiated Sertoli cells, and hyaline bodies (microcalcifications) are often revealed in the orchiectomy tissue or in a testicular biopsy of the contralateral testicle. The contralateral biopsy will detect GCNIS in 4% to 8% of patients, with the greatest risk in patients with low sperm count and small testis volume.[96,111]

All patients treated for testicular cancer require careful evaluation of their reproductive hormones, spermatogenic capacity, and sexual function, in consideration of their wishes of fatherhood in near or distant future. In many patients, fertility is impaired and signs of Leydig cell dysfunction and testosterone deficiency are common.[130-132] Testicular function in patients with testicular tumors may be further disturbed by treatment, especially if cytotoxic therapy is required in addition to orchiectomy.[115,116,133] Radiotherapy and chemotherapy induce severe and dose-dependent impairment in spermatogenesis and sperm DNA damage. In many patients, treatment-induced azoospermia recovers within 1 to 2 years.[132] However, in many cases, the therapy induces irreversible infertility, especially if the patient undergoes scrotal irradiation for example to eradicate GCNIS.

Even though established chemotherapy protocols should be followed in general, it is important to refine the dosage individually in each patient in order to eradicate the malignancy with least possible damage to the endocrine function. Men treated for testicular cancer require careful counseling about their reproductive function with respect to the chance of recovery of spermatogenesis and fertility, and they should all be offered sperm banking before treatment. Cryopreservation of sperm will allow assisted reproductive treatment if spermatogenesis is affected by the cancer therapy. If sperm banking is not possible due to azoospermia or severe oligoasthenozoospermia, testicular sperm extraction (TESE) may be attempted at the time of orchidectomy ("onco-TESE" or "onco-micro-TESE") with subsequent intracytoplasmic sperm injection (ICSI).[132,134] The success of TESE depends on the size of the tumor (worse in large TGCT) and is greater if the targeted seminiferous tubules are as far from the tumor as possible.[135] Adolescent boys who are unable to produce a semen sample by masturbation can be offered penile vibratory stimulation or electroejaculation, and in experienced centers, TESE can be considered.[136,137]

Fertility preservation in prepubertal children who are subjected to cytotoxic treatment is challenging. In young boys, the attempts to preserve testicular tissue are still experimental but the developing methodology of cryopreservation and inducing spermatogenesis is improving and is increasingly promising.[138]

Long-Term Effects and Quality of Life Issues. The most common types of TGCT affect young adults, adolescents, and children, so long-term effects are a concern for the survivors who have life expectancy of several decades. After cancer treatment, most of the patients should have essentially a lifelong follow-up. In addition to the usual oncology follow-up, all patients should be carefully monitored by an endocrinologist or andrologist for signs of metabolic syndrome and hypogonadism with respect to the possible need for androgen-replacement therapy.[131,133] Metabolic syndrome and a risk of cardiovascular disease are a concern especially in patients treated with chemotherapy or with bilateral testicular cancer. However, the risk of cardiovascular disease is most pronounced in the first period of follow-up and decreases later in life.[139] Survivors of disseminated/advanced TGCT treated with cisplatin-based complex chemotherapy regimens face also increased risk of peripheral neuropathy, hepatotoxicity, and ototoxicity.[140]

Survivors of TGCT have excess mortality compared to the general male population. In a study from Norway, a great majority of excess deaths within 5 years from diagnosis were caused by the testicular cancer itself, but during longer follow-up, 65% of excess deaths were caused by second cancers, primarily gastric, pancreatic, or bladder cancer.[141] A large study reported significant excesses of second non-TGCT solid tumors 30 years after radiotherapy and a 2.7-fold increased risk of leukemia after chemotherapy.[142] The increased risk of secondary malignancies has been noted in numerous previous studies in several countries, causing increasing calls to reduce as much as possible radiotherapy and radiologic imaging diagnostic procedures during the follow-up.[143] In addition to inducing DNA damage and mutations in patients' somatic cells and spermatozoa, the latter potentially transmissible to the next generation;[144] there is some evidence that cytotoxic treatment can cause accelerated development of cellular senescence, resulting in secretion of inflammatory factors contributing to early ageing phenotype in survivors of testicular cancer.[145]

Psychological health issues are also of concern in the survivors of testicular cancer. Reported problems are mainly related to prolonged stress and anxiety, often linked to the reproductive problems and sexual dysfunction, but also unemployment or loss in socioeconomic status. Hence, depending on the individual patient's need, a psychologist or social advisor should participate in the long-term follow-up and care for testicular cancer survivors.

KEY POINTS

- Patients with TGCTs are at an increased risk of infertility and androgen deficiency.
- Cryopreservation of semen should be offered to all patients prior to treatment, and onco-TESE can be considered in some patients.
- The patients require lifelong andrological/endocrine follow-up.

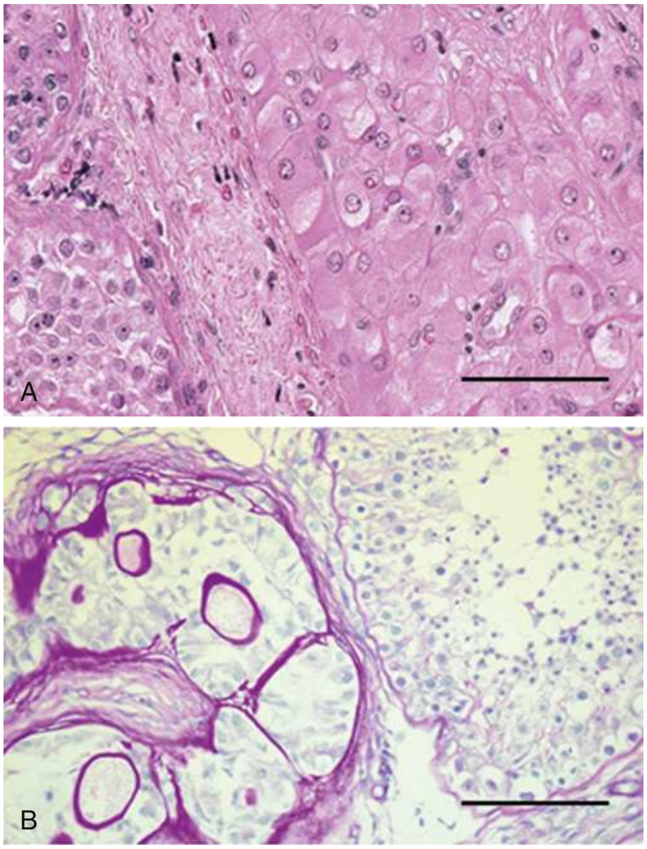


Figure 115.6 Histology of the most common testicular sex cord-stromal tumors. A, Leydig cell tumor in an 8-year-old boy, shown adjacent to preserved seminiferous tubules *(on the left)*. B, Large cell calcifying Sertoli cell tumor isolated from a 12-year-old boy. The tubule with the tumor contains only large pale Sertoli cells and visible calcifications in the lumen. The adjacent normal tubule shows advanced spermatogenesis.

Sex Cord-Stromal Tumors of the Testis

These tumors originate from testicular somatic cells. The classification of sex cord-stromal tumors of the testis according to the WHO[84,146] is presented together with TGCT in Table 115.1. The tumors can occur as pure or mixed types. In contrast to adults—in whom sex cord-stromal tumors are rare and account for 2% to 5% of testicular neoplasms—children develop these tumors in approximately 25% of cases.[112,146] The majority of these tumors (95%) are benign, and some types are a part of specific syndromes.

Leydig Cell Tumors

Histopathology and Genetics. Leydig cell tumors histologically resemble a mass of normal Leydig cells (Fig. 115.6), although malignant tumors can show some cellular atypia often with DNA aneuploidy, mitotic figures, and angiolymphatic invasion.[146] Malignant Leydig cell tumors are relatively rare and have only been reported in adults. Immunohistochemical markers include inhibin A, calretinin, SF-1, melan A, INSL3, vimentin, AR and WT1, AR, chromogranin, synaptophysin, and CD99.[146,147]

Tumors derived from Leydig cells are the most common in the testis besides TGCTs. Leydig cell tumors occur in any age but approximately 20% are found in children between 5 and 10 years of age, and the second peak is in adulthood.[146] There is a perception that the incidence of Leydig cell tumors in adults might have increased in recent decades, but the increase may reflect improved diagnostics of nonpalpable nodules with modern imaging techniques. There is an overlap between Leydig cell tumors, Leydig cell adenoma, micronodules, and florid hyperplasia, most of these belonging to nonpalpable testicular lesions. An extreme example of large Leydig cell nodules is Klinefelter syndrome, although these usually do not progress to malignant tumors. In various disorders of testicular function, including TDS-related cryptorchidism and infertility with testicular atrophy, Leydig cell nodules of variable size are common and their functional insufficiency is often reflected by decreased testosterone/LH ratio,[147,148] whereas patients with overt Leydig cell tumors usually have an increased testosterone/LH ratio and high estradiol.

The etiology of the majority of Leydig cell tumors remains to be elucidated. Genetic defects have been identified only in a small subset of patients. In children with precocious puberty, activating mutations of the LH receptor, *LHCGR* or G proteins can be the cause of Leydig cell tumors.[149,150] In adults with Leydig cell tumors, somatic-activating *GNAS* mutations were identified in a few cases[151] and mutations of germline fumarate hydratase (*FH*) in association with hereditary leiomyomatosis and renal cancer.[152]

Diagnosis and Treatment. The most pronounced symptoms of Leydig cell tumors are in children and young adolescents because of excess of steroid hormones, with precocious puberty often the presenting manifestation (sometimes called testotoxicosis). The typical signs are enlarged penis, pubarche, accelerated growth, skin changes (e.g., acne), and an adult-type odor of sweat. Required hormone profiling in boys with precocious puberty must include serum testosterone, estradiol, LH, FSH, 17-OHP, and androstenedione. Despite signs of puberty, LH and FSH usually remain low because the androgen secretion is gonadotropin-independent.[153] A subset of boys can also have gynecomastia as a result of high aromatase activity in some of the tumors or peripheral aromatization of testosterone. The attending physician should bear in mind that young infants often have physiological transient gynecomastia, obese boys can have pseudo-gynecomastia, and transient gynecomastia can be caused by a quick rise in estradiol at puberty.[154,155]

In adults, adverse hormone effects of a Leydig cell tumor are found in less than half of the patients, most often manifested as gynecomastia, impairment of spermatogenesis, and occasionally sexual dysfunction with loss of libido.[156] Malignant Leydig cell tumors are usually not hormonally active and present as a palpable testicular mass or in about 10% of cases as metastatic disease. Risk factors for metastatic disease include the age >42 years and a large tumor.[148] The ultrasound examination helps in diagnosis in most cases (children and adults), revealing solid well-demarcated hypoechoic lesions, with internal vascularization and occasionally necrosis or cystic areas.[156]

Treatment of Leydig cell tumors is surgical, with testis-sparing surgery recommended for benign tumors. In children, Leydig cell tumors are always benign and should be treated by surgical excision, especially if the tumor is encapsulated. Likewise, in adults, testis-sparing surgery should be attempted if the examination of the frozen preparation does not show signs of malignancy. Malignant tumors have been reported in 10% to 15% of adult patients, and these are usually treated by radical orchiectomy. Testis-sparing surgery can be attempted in isolated and well-demarcated malignant tumors, especially that local recurrence is low (<7%).[148] Adjuvant retroperitoneal lymphadenectomy is no longer recommended as treatment of choice, but some centers practice it in cases with risk factors or when micrometastatic disease is suspected.[148,157] Metastatic tumors, which can appear several years after the initial diagnosis, require salvage surgery and a combination chemotherapy and palliative radiotherapy, but the prognosis is poor.[148,157] Therefore lifelong follow-up of patients with malignant Leydig cell tumors is recommended.

Sertoli Cell Tumors

Histopathology and Genetics. Sertoli cells are the somatic cells in the seminiferous epithelium (sex cord cells) that support differentiating spermatogenic cells. The classification of Sertoli cell tumors according to the WHO[84] is shown in Table 115.1. These tumors are rare (around 1% of all testicular tumors) and are reported in a wide range of age, from 18 to 80 years, but most of them occur in young adults. Sertoli cell tumors are frequently found in association with two distinct multiple neoplasia syndromes, the Carney complex and Peutz-Jeghers syndrome.[146,158]

Isolated Sertoli cell tumors are referred to as *not otherwise specified* (NOS). These tumors occur primarily in adults, are usually unilateral, and occur in descended testes. Juvenile Sertoli cell tumors often showed infiltrative growth into adjacent tissue and considerable proliferative activity, but after surgical excision, no local recurrences or metastases occurred. Histologically, the tumors are composed of sex cord cells with tubular differentiation, frequent lipid vacuoles, and lymphocytic infiltrates, with a subset of tumors hyalinized.[146] Immunohistochemical markers include inhibin, β-catenin, SOX9, SF1, melan A, and WT1.[146] The molecular origin is known only in a subset of tumors, mainly with sclerosing appearance, in which *CTNNB1* mutations causing nuclear accumulation of β-catenin were identified.[159]

In the *Carney complex*, patients have skin myxomas, heart myxomas, typical skin pigmentations, and pituitary, adrenal, and testicular tumors, other tumors can also occur.[158] The testicular tumors typical for this syndrome are *large cell calcifying Sertoli cell tumors* (LCCSTs) that are usually multifocal and bilateral (see Fig. 115.6). The tumors often appear in adolescents and are in most cases benign. Only one malignant case has been reported in association with the Carney complex, whereas malignant course is more frequent in young adult men and in sporadic cases (about 60% of LCCSTs occur sporadically).[84,146] The tumors have characteristic histology, with numerous calcifications, and are typically positive for inhibin and SF1.[146] LCCSTs are usually not hormonally active, although elevated levels of serum inhibin have been reported,[160] but other tumors of Carney complex, including Leydig cell tumors, can cause endocrine manifestations. Genetic loci for Carney complex have been identified on chromosome 2p16 and 17q23-24.[158] The identified germline mutations most often occur in type I-alpha regulatory subunit of protein kinase A, *PRKAR1A*. Inactivating mutations of phosphodiesterase 11A212 *(PDE11A)* and phosphodiesterase 8B213 are associated with bilateral adrenocortical hyperplasia in Carney complex patients. Genetic variation of the phosphodiesterase 11A gene can modify the development of the testicular tumors.[161]

Sertoli cell tumors associated with *Peutz-Jeghers syndrome* are practically exclusive for this syndrome and exhibit a benign behavior. These tumors can be distinguished from those associated with the Carney complex by typical intratubular proliferation, hence are called *intratubular large cell hyalinizing Sertoli cell neoplasia* (LCHSCN). Histological features include prominent basement membrane deposits and annular tubules, and the tumors are usually positive for inhibin and aromatase.[146] Peutz-Jeghers syndrome and the associated tumors are caused by germline loss-of-function mutations in the *STK11/LKB1* gene, encoding a serine-threonine kinase.[162]

Diagnosis and Treatment. Differential diagnosis of Sertoli cell tumors should focus on identification of a possible underlying multicancer syndrome. Clinicians should bear in mind the association with other neoplasms, particularly heart myxomas and adrenal or pituitary tumors in the Carney complex and gastrointestinal tumors in Peutz-Jeghers syndrome, to reach an early diagnosis of these diseases. The patients with Peutz-Jeghers syndrome have a highly increased risk of other neoplasms, especially colorectal, breast, and pancreatic cancers. Patients with the Carney syndrome can usually have spotty skin hyperpigmentation and adrenocortical disease. In addition, the patients with Peutz-Jeghers syndrome often have bilateral gynecomastia, because the associated Sertoli cell tumors can exhibit strong aromatase activity.[163]

Ultrasound examination is helpful, but Sertoli cell tumors have no specific features and can appear as single or multiple masses, either hypo- or hyperechoic, with the presence of large calcifications suggestive of the Carney complex. There are no specific serum tumor markers, but a reproductive hormone profiling should be done, especially in patients with gynecomastia. Final diagnosis is based on the intra- and post-surgery histopathological analysis.

Conservative watchful surveillance management of Sertoli cell tumors is recommended in children, at least until adulthood is reached and sperm banking can be secured, because the tumors are multifocal and bilateral. The follow-up protocols should include ultrasound examinations every 3 months to assess possible changes in size of the lesions and endocrine assessment of testicular function. Growth of the tumors can necessitate orchiectomy. In patients with precocious puberty and/or gynecomastia, aromatase inhibitors and antiandrogens can be used to block estrogen formation and androgen action.

Other Sertoli cell tumors (NOS), which are most often small and well demarcated, usually benign and without adverse endocrine activity, should preferably be treated by orchiectomy or testis-sparing surgery. The prognosis is good, except in exceedingly rare malignant metastatic cases.

Juvenile- and Adult-Type Granulosa Cell Tumors. *Juvenile-type granulosa cell tumors* are rare but are the most common testicular tumors in infants; the diagnosis is usually made before the age of 6 months, sometimes soon after birth.[164] Juvenile granulosa cell tumors have been found in children with ambiguous genitalia and undescended testes, usually with sex chromosomal DSD, primarily XY aneuploidy. The tumors are characterized by the presence of follicles of variable sizes mixed with cord-like structures and are typically positive for inhibin, CD99, SOX9, and FOXL2.[146] In contrast to their ovarian counterparts, testicular juvenile granulosa cell tumors do not show endocrine activity. Juvenile granulosa cell tumors have a benign clinical course and a good prognosis. The treatment should be a part of a multidisciplinary DSD management and may include orchiectomy in selected cases.

Adult-type testicular granulosa cell tumors are extremely rare and occur in adolescents and adults, with mean age of 42 years.[146] These tumors are similar to the ovarian granulosa tumors, which are common in women. Only some of the patients have manifestations of feminization, primarily gynecomastia, resulting from the hormonal activity of the tumor. Most of the tumors are benign, but malignant cases have also been reported. Mutations in *FOXL2* have been described in a subset of male granulosa cell tumors.[165] Treatment is only surgical, with orchiectomy as the primary modality.

Mixed Sex Cord-Stromal Tumors. These rare tumors, which can occur at any age, can contain combinations of Leydig, Sertoli, granulosa, and theca cells and are therefore called *mixed or unclassified sex cord-stromal tumors.*[84,146] These tumors are always benign in children but can be malignant in adults. Most of the patients can be treated by orchiectomy.

Other Testicular Tumors

Testicular Adrenal Rest Tumors

Excessive secretion of adrenocorticotropic hormone (ACTH) in patients with Congenital Adrenal Hyperplasia or Nelson's syndrome (postadrenalectomy status) may lead to the development of testicular adrenal rest tumors (TART).[166] These cells may resemble Leydig cell hyperplasia or tumor and have been hypothesized to originate from persistent fetal Leydig cells or precursors of adult Leydig cells.[167] TART has to be distinguished from Leydig cell hyperplasia and malignant Leydig cell tumors. Histopathological and endocrine evaluation covering both testicular and adrenal steroids and pituitary gonadotropins and ACTH are important to make differential diagnosis.[166,168] Careful endocrine evaluation is necessary to distinguish from TGCT, because TART should not be treated surgically. Some but not all newly appearing TART may regress if appropriate glucocorticoid replacement is initiated, whereas longer-lasting TART does not regress, most likely due to fibrosis and hyalinization. In those rare cases, testis-sparing surgery (nodule enucleation) can be considered.

KEY POINTS

- Testicular tumors derived from somatic cells (sex cord-stromal testis tumors) are rare, except Leydig cell tumors that are considered underdiagnosed and more common than previously thought.
- These tumors can occur at any age.
- Histopathological diagnosis is essential for management planning.
- These tumors are usually benign, and testis-sparing surgery should always be considered.

Secondary Malignancies

The attending physician must be aware that the testis can be a sight of malignancies not originating from testis-specific cell types. In children and adolescents, the most common secondary malignancies involving the testis are acute lymphoblastic leukemia, neuroblastoma, nephroblastoma (Wilms tumor), and rhabdomyosarcoma.[84] In the adults, hematolymphoid malignancies are common and malignant lymphoma (predominantly diffuse large B-cell lymphoma) is the most common testicular malignancy of men older than 50 years.[84] Other malignancies include plasmacytoma, myeloid sarcoma, Rosai-Dorfman disease (histiocytosis), leukemia, myeloid and monocytic sarcoma, and adenocarcinomas of the rete testis. In addition, secondary tumors can metastasize into the testis from other organs, most commonly from the prostate gland, gastrointestinal tract, kidney, lung, pancreas, bladder and urinary tract, and skin (malignant melanoma).

CONCLUSION

The concept of TDS has unified pathogenesis of several male reproductive disorders into one entity, with the underlying cause being an impaired development and paracrine cross talk between multiple cell lineages in the developing testis, resulting in postnatal genital malformations, reduced spermatogenesis, and germ cell cancer. The etiology is multifactorial and polygenetic, with the predominant role of adverse environmental factors acting on genetically susceptible populations and/or individuals. More research is needed to identify the adverse factors in the environment or related to the lifestyle and elucidate the mechanisms of gene-environment interactions, including epigenetic modulation. It is also imperative to increase research efforts to understand the pathogenesis of non-TDS types of testicular tumors and improve their diagnosis and treatment.

Diagnostic tools and therapy of patients with testicular disorders have improved considerably in recent years, but there is a need for more individualized therapy for each patient. Furthermore, most of the patients are young and face long posttreatment life. Endocrinological/andrological participation in a multidisciplinary follow-up team is essential to preserve fertility, prevent late effects of testicular dysfunction, and ensure quality of life of the survivors.

Acknowledgment

The authors wish to thank Professor Niels E. Skakkebæk for his contributions to the previous editions of this chapter.

REFERENCES

1. Skakkebæk NE, Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod*. 2001;6:972–978.
2. Skakkebæk NE, Rajpert-De Meyts E, Buck Louis GM, et al. Male reproductive disorders and fertility trends: influences of environment and genetic susceptibility. *Physiol Rev*. 2016;96:55–97.
3. Jørgensen A, Lindhardt Johansen M, et al. Pathogenesis of germ cell neoplasia in testicular dysgenesis and disorders of sex development. *Sem Cell Dev Biol*. 2015;45:124–137.
4. Xing JS, Bai ZM. Is testicular dysgenesis syndrome a genetic, endocrine, or environmental disease, or an unexplained reproductive disorder? *Life Sci*. 2018;194:120–129.
5. Juul A, Almstrup K, Andersson AM, et al. Possible fetal determinants of male infertility. *Nat Rev Endocrinol*. 2014;10:553–562.
6. Serrano T, Chevrier C, Multigner L, et al. International geographic correlation study of the prevalence of disorders of male reproductive health. *Hum Reprod*. 2013;28:1974–1986.
7. Arendt LH, Ernst A, Braskhøj Lauridsen LL, et al. Timing of pubertal development in boys born with cryptorchidism and hypospadias: a nationwide cohort study. *Asian J Androl*. 2019;21:551–556.
8. Gurney JK, Stanley J, Shaw C, et al. Ethnic patterns of hypospadias in New Zealand do not resemble those observed for cryptorchidism and testicular cancer: evidence of differential aetiology? *Andrology*. 2016;4(1):82–86.
9. Sonne SB, Almstrup K, Dalgaard M, et al. Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma *in situ* is an arrested gonocyte. *Cancer Res*. 2009;69:5241–5250.
10. Van den Driesche S, Kilcoyne KR, Wagner I, et al. Experimentally induced testicular dysgenesis syndrome originates in the masculinization programming window. *JCI Insight*. 2017;2(6):e91204.
11. Thankamony A, Pasterski V, Ong KK, et al. Anogenital distance as a marker of androgen exposure in humans. *Andrology*. 2016;4:616–625.
12. Priskorn L, Bang AK, Nordkap L, et al. Anogenital distance is associated with semen quality but not reproductive hormones in 1106 young men from the general population. *Hum Reprod*. 2019;34:12–24.

12a. Moreno-Mendoza D, Casamonti E, Riera-Escamilla A, et al. Short anogenital distance is associated with testicular germ cell tumour development. *Andrology*. 2020;8:1770–1778. PubMed PMID: 32683775.

13. Fisher BG, Thankamony A, Hughes IA, et al. Prenatal paracetamol exposure is associated with shorter anogenital distance in male infants. *Hum Reprod*. 2016;31:2642–2650.
14. Fisher BG, Thankamony A, Mendiola J, et al. Maternal serum concentrations of bisphenol A and propyl paraben in early pregnancy are associated with male infant genital development. *Hum Reprod*. 2020;35:913–928.
15. Swan SH, Main KM, Liu F, et al. Study for Future Families Research Team: anogenital distance—a marker of fetal androgen action—is decreased in male infants following phthalate exposure during pregnancy. *Environ Health Perspect*. 2005;113:1056–1061.
16. Bornehag CG, Carlstedt F, Jönsson BA, et al. Prenatal phthalate exposures and anogenital distance in Swedish boys. *Environ Health Perspect*. 2015;123:101–107.
17. Arbuckle TE, Agarwal A, MacPherson SH, et al. Prenatal exposure to phthalates and phenols and infant endocrine-sensitive outcomes: the MIREC study. *Environ Int*. 2018;120:572–583.
18. Sturm R, Kurzrock E, Amend G, et al. Blind ending vessels on diagnostic laparoscopy for nonpalpable testis: is a nubbin present? *J Pediatr Urol*. 2017;13:392.e1–392.e6.
19. Favorito LA, Anderson KM, Costa SF, et al. Structural study of the cremaster muscle in patients with retractile testis. *J Pediatr Surg*. 2018;53:780–783.
20. Kolon TF, Herndon CD, Baker LA, American Urological Association, et al. Evaluation and treatment of cryptorchidism: AUA guideline. *J Urol*. 2014;192:337–345.

21. Koskenniemi JJ, Virtanen HE, Wohlfahrt-Veje C, et al. Postnatal changes in testicular position are associated with IGF-I and function of Sertoli and Leydig cells. *J Clin Endocrinol Metab*. 2018;103:1429–1437.
22. Boisen KA, Kaleva M, Main KM, et al. Difference in prevalence of congenital cryptorchidism in infants of two Nordic countries. *Lancet*. 2004;363:1264–1269.
23. Boehme P, Degener S, Wirth S, et al. Multicenter analysis of acquired undescended testis and its impact on the timing of orchidopexy. *J Pediatr*. 2020;S0022–3476(20):30492–30493.
24. Mäkelä JA, Koskenniemi JJ, Virtanen HE, et al. Testis development. *Endocr Rev*. 2019;40:857–905.
25. Barthold JS, Reinhardt S, Thorup J. Genetic, maternal, and environmental risk factors for cryptorchidism: an update. *Eur J Pediatr Surg*. 2016;26:399–408.
26. Rodprasert W, Virtanen HE, Mäkelä JA, et al. Hypogonadism and cryptorchidism. *Front Endocrinol*. 2020;10:906.
27. Elert A, Jahn K, Heidenreich A, et al. The familial undescended testis. *Klin Pädiatr*. 2003;215:40–45.
28. Jensen MS, Toft G, Thulstrup AM, et al. Cryptorchidism concordance in monozygotic and dizygotic twin brothers, full brothers, and half-brothers. *Fertil Steril*. 2010;93:124–129.
29. Virtanen HE, Tapanainen AE, Kaleva MM, et al. Mild gestational diabetes as a risk factor for congenital cryptorchidism. *J Clin Endocrinol Metab*. 2006;91:4862–4865.
30. Damgaard IN, Skakkebæk NE, Toppari J, et al. Persistent pesticides in human breast milk and cryptorchidism. *Environ Health Perspect*. 2006;114:1133–1138.
31. Main KM, Kiviranta H, Virtanen HE, et al. Flame retardants in placenta and breast milk and cryptorchidism in newborn boys. *Environ Health Perspect*. 2007;115:1519–1526.
32. Fernandez MF, Olmos B, Granada A, et al. Human exposure to endocrine-disrupting chemicals and prenatal risk factors for cryptorchidism and hypospadias: a nested case-control study. *Environ Health Perspect*. 2007;115(suppl 1):8–14.
33. Brucker-Davis F, Wagner-Mahler K, Delattre I, et al. Cryptorchidism at birth in Nice area (France) is associated with higher prenatal exposure to PCBs and DDE, as assessed by colostrum concentrations. *Hum Reprod*. 2008;23:1708–1718.
34. Main KM, Mortensen GK, Kaleva MM, et al. Human breast milk contamination with phthalates and alterations in endogenous reproductive hormones in three months old children. *Environ Health Perspect*. 2006;114:270–276.
35. Hua XG, Hu R, Hu CY, et al. Associations between hypospadias, cryptorchidism and anogenital distance: systematic review and meta-analysis. *Andrologia*. 2018;50:e13152.
36. Eisenberg ML, Hsieh MH, Walters RC, et al. The relationship between anogenital distance, fatherhood, and fertility in adult men. *PLoS One*. 2011;6:e18973.
37. Bergbrant S, Omling E, Björk J, et al. Cryptorchidism in Sweden: a nationwide study of prevalence, operative management, and complications. *J Pediatr*. 2018;194:197–203.
38. Radmayr C, Dogan HS, Hoebeke P, et al. Management of undescended testes: European association of Urology/European Society for paediatric Urology guidelines. *J Pediatr Urol*. 2016;12:335–343.
39. Kollin C, Granholm T, Nordenskjöld A, et al. Growth of spontaneously descended and surgically treated testes during early childhood. *Pediatrics*. 2013;131:1174–1180.
40. Ong C, Hasthorpe S, Hurson JM. Germ cell development in the descended and cryptorchid testis and the effects of hormonal manipulation. *Pediatr Surg Int*. 2005;21:240–254.
41. Dunkel L, Taskinen S, Hovotta O, et al. Germ cell apoptosis after treatment of cryptorchidism with hCG is associated with impaired reproductive function in the adult. *J Clin Invest*. 1997;100:2341–2346.
42. Huff DS, Snyder 3rd HM, Rusnack SL, et al. Hormonal therapy for the subfertility of cryptorchidism. *Horm Res*. 2001;55:38–40.
43. Sadov S, Koskenniemi JJ, Virtanen HE, et al. Testicular growth during puberty in boys with and without a history of congenital cryptorchidism. *J Clin Endocrinol Metab*. 2016;101:2570–2577.
44. Shepard CL, Kraft KH. The nonpalpable testis: a narrative review. *J Urol*. 2017;198:1410–1417.
45. Uijldert M, Meißner A, Kuijper CF, et al. Orchidopexy for bilateral undescended testes: a multicentre study on its effects on fertility and comparison of two fixation techniques. *Andrologia*. 2019;51:e13194.
46. Suomi AM, Main KM, Kaleva M, et al. Hormonal changes in 3-month-old cryptorchid boys. *J Clin Endocrinol Metab*. 2006;91:953–958.
47. Trabert B, Zugna D, Richiardi L, et al. Congenital malformations and testicular germ cell tumors. *Int J Cancer*. 2013;133:1900–1904.
48. Baskin L. What is hypospadias? *Clin Pediatr*. 2017;56:409–418.
49. Rey R, Josso N, Racine C. Sexual differentiation. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext. South Dartmouth*. MA: MDText.com, Inc.; 2000–2020.
50. Cools M, Nordenström A, Robeva R, et al. COST Action BM1303 working group 1. Caring for individuals with a difference of sex development (DSD): a consensus statement. *Nat Rev Endocrinol*. 2018;14:415–429.
51. Joodi M, Amerizadeh F, Hassanian SM, et al. The genetic factors contributing to hypospadias and their clinical utility in its diagnosis. *J Cell Physiol*. 2019;234:5519–5523.
52. Kalfa N, Gaspari L, Ollivier M, et al. Molecular genetics of hypospadias and cryptorchidism recent developments. *Clin Genet*. 2019;95:122–131.
53. Fredell L, Lichtenstein P, Pedersen NL, et al. Hypospadias is related to birth weight in discordant monozygotic twins. *J Urol*. 1998;160:2197–2199.
54. Asklund C, Jørgensen N, Skakkebæk NE, et al. Increased frequency of reproductive health problems among fathers of boys with hypospadias. *Hum Reprod*. 2007;10:2639–2646.
55. Fedder J, Loft A, Parner ET, et al. Neonatal outcome and congenital malformations in children born after ICSI with testicular or epididymal sperm: a controlled national cohort study. *Hum Reprod*. 2013;28:230–240.
56. Boisen KA, Chellakooty M, Schmidt IM, et al. Hypospadias in a cohort of 1072 Danish newborn boys: prevalence and relationship to placental weight, anthropometrical measurements at birth and reproductive hormone levels at 3 months of age. *J Clin Endocrinol Metab*. 2005;90:4041–4046.
57. van Rooij IA, van der Zanden LF, Brouwers MM, et al. Risk factors for different phenotypes of hypospadias: results from a Dutch case-control study. *BJU*. 2013;112:121–128.
58. Raghavan R, Romano ME, Karagas MR, et al. Pharmacologic and environmental endocrine disruptors in the pathogenesis of hypospadias: a review. *Curr Environ Health Rep*. 2018;5:499–511.
59. White JT, Kovar E, Chambers TM, et al. Hypospadias risk from maternal residential exposure to heavy metal hazardous air pollutants. *Int J Environ Res Public Health*. 2019;16(6):930.
60. Henderson BE, Benton B, Cosgrove M, et al. Urogenital tract abnormalities in sons of women treated with diethylstilbestrol. *Pediatrics*. 1976;58:505–507.
61. Snijder CA, Kortenkamp A, Steegers EA, et al. Intrauterine exposure to mild analgesics during pregnancy and the occurrence of cryptorchidism and hypospadias in the offspring: the Generation R study. *Hum Reprod*. 2012;27:1191–1201.
62. Thankamony A, Lek N, Carroll D, et al. Anogenital distance and penile length in infants with hypospadias or cryptorchidism: comparison with normative data. *Environ Health Perspect*. 2014;122:207–211.
63. Springer A, van den Heijkant M, Baumann S. Worldwide prevalence of hypospadias. *J Pediatr Urol*. 2016;12. 152.e1-e152.e7.
64. Yu X, Nassar N, Mastroiacovo P, et al. Hypospadias prevalence and trends in international birth defect surveillance systems, 1980-2010. *Eur Urol*. 2019;76:482–490.
65. Bethell GS, Chhabra S, Shalaby MS, et al. Parental decisional satisfaction after hypospadias repair in the United Kingdom. *J Pediatr Urol*. 2020;16:164.e1–164.e7.
66. Gorduza D, Plotton I, Remontet L, et al. Preoperative topical estrogen treatment vs placebo in 244 children with midshaft and posterior hypospadias. *J Clin Endocrinol Metab*. 2020;105:dgaa231.
67. Wright I, Cole E, Farrokhyar F, et al. Effect of preoperative hormonal stimulation on postoperative complication rates after proximal hypospadias repair: a systematic review. *J Urol*. 2013;190:652–659.

68. Gong EM, Cheng EY. Current challenges with proximal hypospadias: we have a long way to go. *J Pediatr Urol*. 2017;13:457–467.
69. Dodds PR, Batter SJ, Shiled DE, et al. Adaptation of adults to uncorrected hypospadias. *Urology*. 2008;71:682–685.
70. Asklund C, Jensen TK, Main KM, et al. Semen quality, reproductive hormones and fertility of men operated for hypospadias. *Int J Androl*. 2009;33:80–87.
71. European Reference Network, Urogenital Diseases, eUROGEN: https://eurogen-ern.eu/.
72. Andersson AM, Toppari J, Haavisto AM, et al. Longitudinal reproductive hormone profiles in infants: peak of inhibin B levels in infant boys exceeds levels in adult men. *J Clin Endocrinol Metab*. 1998;83:675–681.
73. Winter JSD, Hughes IA, Reyes FI, et al. Pituitary-gonadal relations in infancy: I. Patterns of serum gonadotropin concentrations from birth to four years of age in man and chimpanzee. *J Clin Endocrinol Metab*. 1975;40:545–551.
74. Winter JSD, Hughes IA, Reyes FI, et al. Pituitary-gonadal relations in infancy: II. Patterns of serum gonadal steroid concentrations in man from birth to two years of age. *J Clin Endocrinol Metab*. 1976;42:679–686.
75. Aksglaede L, Sørensen K, Boas M, et al. Changes in anti-Müllerian hormone (AMH) throughout the life span: a population-based study of 1027 healthy males from birth (cord blood) to the age of 69 years. *J Clin Endocrinol Metab*. 2010;95:5357–5364.
76. Bay K, Virtanen HE, Hartung S, et al. Insulin-like factor 3 levels in cord blood and serum from children: effects of age, postnatal hypothalamic-pituitary-gonadal axis activation, and cryptorchidism. *J Clin Endocrinol Metab*. 2007;92:4020–4027.
77. Codesal J, Regadera J, Nistal M, et al. Involution of human fetal Leydig cells. An immunohistochemical, ultrastructural and quantitative study. *J Anatomy*. 1990;172:103–114.
78. Cortes D, Müller J, Skakkebæk NE. Proliferation of Sertoli cells during development of the human testis assessed by stereological methods. *Int J Androl*. 1987;10:589–596.
79. Chemes HE, Rey RA, Nistal M, et al. Physiological androgen insensitivity of the fetal, neonatal, and early infantile testis is explained by the ontogeny of the androgen receptor expression in Sertoli cells. *J Clin Endocrinol Metab*. 2008;93:4408–4412.
80. Kuiri-Hänninen T, Seuri R, Tyrväinen E, et al. Increased activity of the hypothalamic-pituitary-testicular axis in infancy results in increased androgen action in premature boys. *J Clin Endocrinol Metab*. 2011;96(1):98–105. https://doi.org/10.1210/jc.2010-1359.
81. Main KM, Schmidt IM, Skakkebæk NE. A possible role for reproductive hormones in newborn boys: progressive hypogonadism without the postnatal testosterone peak. *J Clin Endocrinol Metab*. 2000;85:4905–4907.
82. Jespersen K, Ljubicic ML, Johannsen TH, et al. Distinguishing between hidden testes and anorchia: the role of endocrine evaluation in infancy and childhood. *Eur J Endocrinol*. 2020;183:107–117.
83. Cabrol S, Ross JL, Fennoy I, et al. Assessment of Leydig and Sertoli cell functions in infants with nonmosaic Klinefelter syndrome: insulin-like peptide 3 levels are normal and positively correlated with LH levels. *J Clin Endocrinol Metab*. 2011;96:E746–E753.
84. Moch H, Humphrey PA, Ulbright TM, Reuter VE, eds. *WHO Classification of Tumours of the Urinary System and Male Genital Organs*. 4th ed. Lyon: IARC Press; 2016.
85. Ulbright TM, Amin MB, Balzer B, et al. Germ cell tumours. In: Moch H, Humphrey PA, Ulbright TM, Reuter VE, eds. *WHO Classification of Tumours of the Urinary System and Male Genital Organs*. 4th ed. Lyon: IARC Press; 2016:189–226.
86. Oosterhuis JW, Looijenga LHJ. Human germ cell tumours from a developmental perspective. *Nat Rev Cancer*. 2019;19:522–537.
87. Skakkebaek NE. Possible carcinoma-in-situ of the testis. *Lancet*. 1972;2(7776):516–517.
88. Ward EM, Sherman RL, Henley SJ, et al. Annual report to the nation on the status of cancer, featuring cancer in men and women age 20-49 years. *J Natl Cancer Inst*. 2019;111:1279–1297.
89. Gurney JK, Florio AA, Znaor A, et al. International trends in the incidence of testicular cancer: lessons from 35 years and 41 countries. *Eur Urol*. 2019;76:615–623.
90. Ylönen O, Jyrkkiö S, Pukkala E, et al. Time trends and occupational variation in the incidence of testicular cancer in the Nordic countries. *BJU Int*. 2018;122:384–393.
91. Cook MB, Akre O, Forman D, et al. A systematic review and meta-analysis of perinatal variables in relation to the risk of testicular cancer--experiences of the son. *Int J Epidemiol*. 2010;39:1605–1618.
92. Gurney J, Shaw C, Stanley J, et al. Cannabis exposure and risk of testicular cancer: a systematic review and meta-analysis. *BMC Cancer*. 2015;15:897.
93. McGlynn KA, Petrick JL, Gamborg M, et al. Childhood height and risk of testicular germ cell tumors in adulthood. *Int J Cancer*. 2018;143(4):767–772.
94. Ljubicic ML, Jørgensen A, Acerini C, et al. Clinical but not histological outcomes in males with 45,X/46,XY mosaicism vary depending on reason for diagnosis. *J Clin Endocrinol Metab*. 2019;104(10):4366–4381.
95. Cools M, Wolffenbuttel KP, Hersmus R, et al. Malignant testicular germ cell tumors in postpubertal individuals with androgen insensitivity: prevalence, pathology and relevance of single nucleotide polymorphism-based susceptibility profiling. *Hum Reprod*. 2017;32:2561–2573.
96. Rud CN, Daugaard G, Rajpert-De Meyts E, et al. Sperm concentration, testicular volume and age predict risk of carcinoma-in-situ in contralateral testis of men with testicular germ-cell cancer. *J Urol*. 2013;190:2074–2080.
97. Olesen IA, Andersson AM, Aksglaede L, et al. Clinical, genetic, biochemical, and testicular biopsy findings among 1,213 men evaluated for infertility. *Fertil Steril*. 2017;107:74–82. e7.
98. Xu R, Centola GM, Tanrikut C. Genitourinary cancer patients have worse baseline semen parameters than healthy sperm bankers. *Andrology*. 2019;7:449–453.
99. Litchfield K, Loveday C, Levy M, et al. Large-scale sequencing of testicular germ cell tumour (TGCT) cases excludes major TGCT predisposition gene. *Eur Urol*. 2018;73:828–831.
100. Taylor-Weiner A, Zack T, O'Donnell E, et al. Genomic evolution and chemoresistance in germ-cell tumours. *Nature*. 2016;540:114–118.
101. Shen H, Shih J, Hollern DP, et al. Integrated molecular characterization of testicular germ cell tumors. *Cell Rep*. 2018;23:3392–3406.
102. Litchfield K, Levy M, Orlando G, et al. Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. *Nat Genet*. 2017;49:1133–1140.
103. Wang Z, McGlynn KA, Rajpert-De Meyts E, et al. Testicular Cancer Consortium. Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat Genet*. 2017;49:1141–1147.
104. Jørgensen A, Macdonald J, Nielsen JE, et al. Nodal signalling regulates germ cell development and the establishment of seminiferous cords in human fetal testes. *Cell Rep*. 2018;25:1924–1937.
104a. Bräuner EV, Lim YH, Koch T, et al. Endocrine disrupting chemicals and risk of testicular cancer: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2021;106:e4834–e4860. PubMed PMID: 34270734.
105. Rajpert-De Meyts E, Nielsen JE, Skakkebæk NE, et al. Diagnostic markers for germ cell neoplasms: from placental-like alkaline phosphatase to micro-RNAs. *Folia Histochem Cytobiol*. 2015;53:177–188.
106. Klein B, Haggeney T, Fietz D, et al. Specific immune cell and cytokine characteristics of human testicular germ cell neoplasia. *Hum Reprod*. 2016;31:2192–2202.
107. Murray MJ, Huddart RA, Coleman N. The present and future of serum diagnostic tests for testicular germ cell tumours. *Nat Rev Urol*. 2016;13:715–725.
108. Almstrup K, Lobo J, Mørup N, et al. Application of miRNAs in the diagnosis and monitoring of testicular germ cell tumours. *Nat Rev Urol*. 2020;17:201–213.
109. Dieckmann KP, Radtke A, Geczi L, et al. Serum levels of microRNA-371a-3p (M371 Test) as a new biomarker of testicular germ cell tumors: results of a prospective multicentric study. *J Clin Oncol*. 2019;37:1412–1423.
110. Nappi L, Thi M, Lum A, et al. Developing a highly specific biomarker for germ cell malignancies: plasma miR371 expression across the germ cell malignancy spectrum. *J Clin Oncol*. 2019;37:3090–3098.

111. Dohle GR, Elzanaty S, van Casteren NJ. Testicular biopsy: clinical practice and interpretation. *Asian J Androl*. 2012;14(1):88–93.
112. Albers P, Albrecht W, Algaba F, et al. European Association of Urology. Guidelines on testicular cancer: 2015 update. *Eur Urol*. 2015;68:1054–1068.
112a. Dieckmann KP, Kulejewski M, Pichlmeier U, Loy V. Diagnosis of contralateral testicular intraepithelial neoplasia (TIN) in patients with testicular germ cell cancer: systematic two-site biopsies are more sensitive than a single random biopsy. *Eur Urol*. 2007;51:175–183.
113. Hellesnes R, Myklebust TÅ, Bremnes RM, et al. Metachronous contralateral testicular cancer in the cisplatin era: a population-based cohort study. *J Clin Oncol*. 2021;39:308–318.
114. Gilligan TD, Seidenfeld J, Basch EM, American Society of Clinical Oncology, et al. American Society of Clinical Oncology Clinical Practice Guideline on uses of serum tumor markers in adult males with germ cell tumors. *J Clin Oncol*. 2010;28:3388–3404.
115. Rajpert-De Meyts E, McGlynn KA, Okamoto K, et al. Testicular germ cell tumours. *Lancet*. 2016;387:1762–1774.
116. Honecker F, Aparicio J, Berney D, et al. ESMO Consensus Conference on testicular germ cell cancer: diagnosis, treatment and follow-up. *Ann Oncol*. 2018;29:1658–1686.
117. Verrill C, Yilmaz A, Srigley JR, et al. Members of the international Society of Urological pathology testicular tumor Panel. Reporting and staging of testicular germ cell tumors: the international Society of Urological pathology (ISUP) testicular cancer consultation Conference recommendations. *Am J Surg Pathol*. 2017;41:e22–e32.
118. UICC International Union Against Cancer. In: Sobin LH, Gospodarowicz M, Wittekind C, eds. *TNM Classification of Malignant Tumours*. 7th ed. Hoboken, NJ: Wiley-Blackwell; 2009:249–254.
119. Heidenreich A, Paffenholz P, Nestler T, et al. Primary and postchemotherapy retroperitoneal lymphadenectomy for testicular cancer. *Oncol Res Treat*. 2018;41:370–378.
120. Daugaard G, Gundgaard MG, Mortensen MS, et al. Surveillance for stage I nonseminoma testicular cancer: outcomes and long-term follow-up in a population-based cohort. *J Clin Oncol*. 2014;32:3817–3823.
121. Tandstad T, Ståhl O, Håkansson U, et al. One course of adjuvant BEP in clinical stage I nonseminoma mature and expanded results from the SWENOTECA group. *Ann Oncol*. 2014;25:2167–2172.
122. Poynter JN, Amatruda JF, Ross JA. Trends in incidence and survival of pediatric and adolescent patients with germ cell tumors in the United States, 1975 to 2006. *Cancer*. 2010;116:4882–4891.
123. Murray MJ, Nicholson JC, Coleman N. Biology of childhood germ cell tumours, focussing on the significance of microRNAs. *Andrology*. 2015;3:129–139. https://doi.org/10.1111/andr.277.
124. Cornejo KM, Frazier L, Lee RS, et al. Yolk sac tumor of the testis in infants and children: a clinicopathologic analysis of 33 cases. *Am J Surg Pathol*. 2015;39:1121–1131.
125. Mosbech CH, Svingen T, Nielsen JE, et al. Expression pattern of clinically relevant markers in paediatric germ cell- and sex-cord stromal tumours is similar to adult testicular tumours. *Virchows Arch*. 2014;465:567–577.
126. Frazier AL, Hale JP, Rodriguez-Galindo C, et al. Revised risk classification for pediatric extracranial germ cell tumors based on 25 years of clinical trial data from the United Kingdom and United States. *J Clin Oncol*. 2015;33:195–201.
127. Grogg JB, Schneider K, Bode PK, et al. A systematic review of treatment outcomes in localised and metastatic spermatocytic tumors of the testis. *J Cancer Res Clin Oncol*. 2019;145:3037–3045.
128. Goriely A, Hansen RM, Taylor IB, et al. Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. *Nat Genet*. 2009;41:1247–1252.
129. Giannoulatou E, Maher GJ, Ding Z, et al. Whole-genome sequencing of spermatocytic tumors provides insights into the mutational processes operating in the male germline. *PLoS One*. 2017;12:e0178169. https://doi.org/10.1371/journal.pone.0178169.
130. Petersen PM, Skakkebaek NE, Vistisen K, et al. Semen quality and reproductive hormones before orchiectomy in men with testicular cancer. *J Clin Oncol*. 1999;17:941–947.
131. Bandak M, Jørgensen N, Juul A, et al. Leydig cell dysfunction, systemic inflammation and metabolic syndrome in long-term testicular cancer survivors. *Eur J Cancer*. 2017;84:9–17.
132. Moody JA, Ahmed K, Yap T, et al. Fertility management in testicular cancer: the need to establish a standardized and evidence-based patient-centric pathway. *BJU Int*. 2019;123:160–172.
133. Haugnes HS, Bosl GJ, Boer H, et al. Long-term and late effects of germ cell testicular cancer treatment and implications for follow-up. *J Clin Oncol*. 2012;30:3752–3763.
134. Hsiao W, Stahl PJ, Osterberg EC, et al. Successful treatment of postchemotherapy azoospermia with microsurgical testicular sperm extraction: the Weill Cornell experience. *J Clin Oncol*. 2011;29:1607–1611.
135. Suzuki K, Shin T, Shimomura Y, et al. Spermatogenesis in tumor-bearing testes in germ cell testicular cancer patients. *Hum Reprod*. 2015;30:2853–2858.
136. Gat I, Toren A, Hourvitz A, et al. Sperm preservation by electroejaculation in adolescent cancer patients. *Pediatr Blood Cancer*. 2014;61:286–290.
137. Berookhim BM, Mulhall JP. Outcomes of operative sperm retrieval strategies for fertility preservation among males scheduled to undergo cancer treatment. *Fertil Steril*. 2014;101:805–811.
138. Goossens E, Jahnukainen K, Mitchell RT, et al. Fertility preservation in boys: recent developments and new insights. *Hum Reprod Open*. 2020;2020(3):hoaa016.
139. Lauritsen J, Hansen MK, Bandak M, et al. Cardiovascular risk factors and disease after male germ cell cancer. *J Clin Oncol*. 2020;38:584–592.
140. Chovanec M, Abu Zaid M, Hanna N, et al. Long-term toxicity of cisplatin in germ-cell tumor survivors. *Ann Oncol*. 2017;28:2670–2679.
141. Kvammen Ø, Myklebust TÅ, Solberg A, et al. Causes of inferior relative survival after testicular germ cell tumor diagnosed 1953-2015: a population-based prospective cohort study. *PLoS One*. 2019;14:e0225942.
142. Milano MT, Dinh PC, Yang H, et al. Solid and hematologic neoplasms after testicular cancer: a US population-based study of 24900 survivors. *JNCI Cancer Spectr*. 2020;4(3):pkaa017.
143. Nichols CR, Nappi L, Kollmannsberger C, et al. Back to the future-moving forward for testicular cancer survivors. *JNCI Cancer Spectr*. 2019;4(2):pkz082.
144. Meistrich ML. Risks of genetic damage in offspring conceived using spermatozoa produced during chemotherapy or radiotherapy. *Andrology*. 2020;8:545–558.
145. Lubberts S, Meijer C, Demaria M, et al. Early ageing after cytotoxic treatment for testicular cancer and cellular senescence: time to act. *Crit Rev Oncol Hematol*. 2020;151:102963.
146. Idrees MT, Ulbright TM, Oliva E, et al. Members of the international Society of Urological pathology testicular tumour Panel. The world health Organization 2016 classification of testicular non-germ cell tumours: a review and update from the international Society of Urological pathology testis consultation Panel. *Histopathology*. 2017;70:513–521.
147. Lottrup G, Nielsen JE, Maroun LL, et al. Expression patterns of DLK1 and INSL3 identify stages of Leydig cell differentiation during normal development and in testicular pathologies, including testicular cancer and Klinefelter syndrome. *Hum Reprod*. 2014;29:1637–1650.
148. Fankhauser CD, Grogg JB, Hayoz S, et al. Risk factors and treatment outcomes of 1,375 patients with testicular Leydig cell tumors: analysis of published case series data. *J Urol*. 2020;203:949–956.
149. Liu G, Duranteau L, Carel JC, et al. Leydig-cell tumors caused by an activating mutation of the gene encoding the luteinizing hormone receptor. *N Engl J Med*. 1999;341:173–176.
150. Fragoso MC, Latronico AC, Carvalho FM, et al. Activating mutation of the stimulatory G protein (gsp) as a putative cause of ovarian and testicular human stromal Leydig cell tumors. *J Clin Endocrinol Metab*. 1998;83:2074–2081.
151. Libe R, Fratticci A, Lahlou N, et al. A rare cause of hypertestosteronemia in a 68-year-old patient: a Leydig cell tumor due to a somatic GNAS (guanine nucleotide-binding protein, alpha-stimulating activity polypeptide 1)-activating mutation. *J Androl*. 2012;33:578–584.

152. Carvajal-Carmona LG, Alam NA, Pollard PJ, et al. Adult Leydig cell tumors of the testis caused by germline fumarate hydratase mutations. *J Clin Endocrinol Metab*. 2006;91:3071–3075.
153. Mortensen LJ, Blomberg Jensen M, Christiansen P, et al. Germ cell neoplasia in situ and preserved fertility despite suppressed gonadotropins in a patient with testotoxicosis. *J Clin Endocrinol Metab*. 2017;102:4411–4416.
154. Mieritz MG, Raket LL, Hagen CP, et al. A longitudinal study of growth, sex steroids, and IGF-1 in boys with physiological gynecomastia. *J Clin Endocrinol Metab*. 2015;100:3752–3759.
155. Kanakis GA, Nordkap L, Bang AK, et al. EAA clinical practice guidelines - gynecomastia evaluation and management. *Andrology*. 2019;7:778–793.
156. Pozza C, Pofi R, Tenuta M, et al. Clinical presentation, management and follow-up of 83 patients with Leydig cell tumors of the testis: a prospective case-cohort study. *Hum Reprod*. 2019;34:1389–1403.
157. Calaway AC, Tachibana I, Masterson TA, et al. Oncologic outcomes following surgical management of clinical stage II sex cord stromal tumors. *Urology*. 2019;127:74.
158. Stratakis CA, Kirschner LS, Taymans SE, et al. Carney complex, Peutz-Jegher syndrome, Cowden disease, and Bannayan-Zonana syndrome share cutaneous and endocrine manifestations, but not genetic loci. *J Clin Endocrinol Metab*. 1998;83:2972–2976.
159. Perrone F, Bertolotti A, Montemurro G, et al. Frequent mutation and nuclear localization of β-catenin in Sertoli cell tumors of the testis. *Am J Surg Pathol*. 2014;38:66–71.
160. Toppari J, Kaipia A, Kaleva M, et al. Inhibin gene expression in a large cell calcifying Sertoli cell tumour and serum inhibin and activin levels. *APMIS*. 1998;106(1):101–113.
161. Libé R, Horvath A, Vezzosi D, et al. Frequent phosphodiesterase 11A gene (PDE11A) defects in patients with Carney complex (CNC) caused by PRKAR1A mutations: PDE11A may contribute to adrenal and testicular tumors in CNC as a modifier of the phenotype. *J Clin Endocrinol Metab*. 2011;96:E208–E214.
162. Hemminki A, Markie D, Tomlison I, et al. A serine/threonine kinase gene defective in Peutz-Jegher syndrome. *Nature*. 1998;391:184–187.
163. Ham S, Meachem SJ, Choong CS, et al. Overexpression of aromatase associated with loss of heterozygosity of the STK11 gene accounts for prepubertal gynecomastia in boys with Peutz-Jeghers syndrome. *J Clin Endocrinol Metab*. 2013;98:E1979–E1987.
164. Hofmann M, Schlegel PG, Hippert F, et al. MAKEI study group. Testicular sex cord stromal tumors: analysis of patients from the MAKEI study. *Pediatr Blood Cancer*. 2013;60:1651–1655.
165. Lima JF, Jin L, de Araujo AR, et al. FOXL2 mutations in granulosa cell tumors occurring in males. *Arch Pathol Lab Med*. 2012;136:825–828.
166. Engels M, Span PN, van Herwaarden AE, et al. Testicular adrenal rest tumors: current insights on prevalence, characteristics, origin, and treatment. *Endocr Rev*. 2019;40:973–987.
167. Shima Y, Morohashi KI. Leydig progenitor cells in fetal testis. *Mol Cell Endocrinol*. 2017;445:55–64.
168. Lottrup G, Nielsen JE, Skakkebæk NE, et al. Abundance of DLK1, differential expression of CYP11B1, CYP21A2, MC2R and lack of INSL3 distinguish testicular adrenal rest tumours from Leydig cell tumours. *Eur J Endocrinol*. 2015;172:491–499.

116

Benign Prostatic Hyperplasia and Prostate Cancer

Gail P. Risbridger, Mark Frydenberg, David W. Pook, and Paul Manohar

OUTLINE

Introduction, 1937
Benign Prostatic Hyperplasia, 1937
  Diagnosis, 1937
  Management and Treatment of BPH, 1938
  Summary and Future Research, 1940
Prostate Cancer, 1940
  Diagnosis and Detection of Localized Prostate Cancer, 1940
  Management and Treatment of Localized Prostate Cancer, 1942
  Diagnosis and Detection of Advanced Prostate Cancer, 1944
  Management and Treatment of Advanced Prostate Cancer, 1945
Summary and Future Challenges, 1946

INTRODUCTION

The prostate gland is a hormonally regulated organ that becomes bothersome in some men as they become older. Emerging as part of the normal aging process, benign disease or benign prostatic hyperplasia (BPH) is due to enlargement of the prostate gland and changes in smooth muscle tone. Malignancy or prostate cancer has a different etiology, but also becomes more prevalent as men reach late middle and older age.

Both BPH and prostate cancer have an inestimable impact on the quality of life of patients and their families. The worldwide prevalence of BPH is estimated as 26.2%, that is, nearly 1 in 4 men will suffer from BPH over their lifetime.[1] Worldwide, more than 1.2 million men were diagnosed with prostate cancer in 2018 and there were >350,000 deaths.[2] Australia has the highest age standardized (world) incidence and mortality rates at 86.2 and 10.2 per 100,000 men, respectively. Nearly half of all Australian/American men will develop prostate cancer, and one in four will die from the disease. Although BPH and prostate cancer can occur simultaneously in the aging male prostate, there is no direct relationship between the two pathologies and they are regarded as separate diseases that mainly occur in different regions of the prostate gland (see Fig. 116.1).

The prostate gland is a lobular structure and in 1968, McNeal described the anatomy in terms of four anatomic zones using the urethra as the primary anatomic reference point, that is, peripheral, central, and two transition zones (TZs) (see Ref. 3). The TZ is the site of development of BPH, whereas prostate cancer more commonly occurs in the other anatomic zones.

BENIGN PROSTATIC HYPERPLASIA

Diagnosis

Many men will have BPH; however, if it remains asymptomatic and never causes complications, it is an unnecessary diagnosis. A histological diagnosis of BPH is defined as a hyperplasia of the epithelium, stroma, or both in the periurethral TZ of the prostate. In men aged >80 years old, 80% had BPH at autopsy, whereas no men aged under 30 had histological BPH.[5] A clinical diagnosis of BPH is more complex due to the variation in case definition used by different studies. BPH plays a key role in male lower urinary tract symptoms (LUTS), but it is not the only cause. It is important to distinguish symptoms related to bladder outlet obstruction (BOO) from symptoms related to storage disorders as their cause and treatment are significantly different. The interrelationship between BPH, LUTS, and BOO is represented as a 3-ring diagram, known as Hald's rings, to portray BPH as a syndrome of 3 overlapping, yet distinct entities (see Fig. 116.2). BPH histology, enlargement, obstruction, and symptoms while linked are not always the cause and effect. There are certainly people who will get obstructive symptoms from a modest-sized gland and others who will remain symptom-free with a large gland. As a result, bothersome LUTS due to BPH causing symptomatic BOO are of more concern.[6]

A diagnosis of BPH includes the patient history, examination, and investigations. The **patient history** is a guide to both symptom severity and the likely chance of progression of risks of the sequelae outlined above. It is also critical to identify other potential causes of LUTS unrelated to BPH. These include medical diseases and neurological risk factors. The European Association of Urology (EAU) guidelines state that current medications, lifestyle habits, and emotional and psychological habits must be a part of the history. Reassuring the patient of no definite link between LUTS and prostate cancer is also critical. Common tools used to assess the severity of the patient's symptoms include the **International Prostate Symptom Score (IPSS) and a Bladder diary**. The IPSS is an eight-question screening tool written by the patient and used to screen, diagnose, and track the symptoms of BPH. Seven questions are about symptoms related to BPH and one is related to the patient's perceived quality of life. The **mnemonic** "FUN WISE" can be used to remember the questions where the first 3 relate to bladder function and the last 4 to prostate function (see Box 116.1). A score out of 35 correlates with mild, moderate, or severe symptoms (see Box 116.2).

In addition to diagnosis, the IPSS can be repeated and chart disease progression. For patients whose main complaint is nocturia, a **bladder diary** is a cheap, simple, reliable assessment of the patient's symptoms[7] (see Box 116.3). It can outline other causes for LUTS, track the progress of treatment, and understand the lifestyle causes and impact of the symptoms on lifestyle in a more formal process.

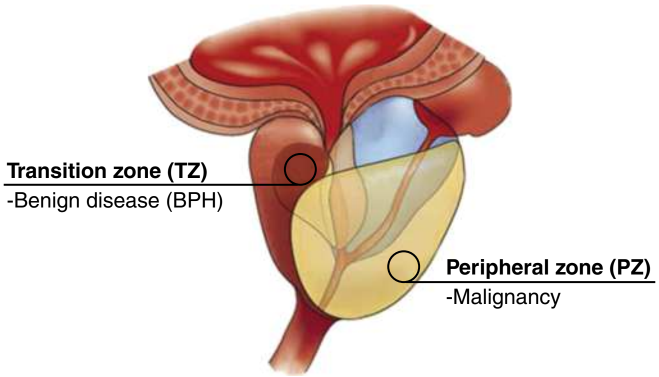


Figure 116.1 **Diagram showing the zonal anatomy of the human prostate gland.** Benign disease most commonly occurs in the TZ, whereas malignant nodules are most prevalent in the peripheral zone (PZ). (Reproduced with permission from Ellem SJ, Risbridger GP. Treating prostate cancer: a rationale for targeting local oestrogens. *Nat Rev Cancer*. 2007;7(8):621-627.)

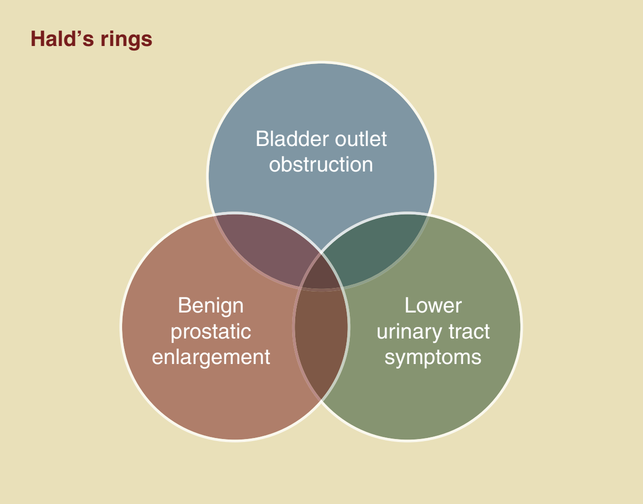

Hald's rings
Bladder outlet obstruction
Benign prostatic enlargement
Lower urinary tract symptoms

Figure 116.2 Hald's rings demonstrate that while linked, BPH, LUTS, and BOO can also occur independently of each other.

BOX 116.1 The IPSS Mnemonic "FUN WISE" Is a Reminder of the Seven Questions to Ask the Patient About Bladder and Prostate Function

F	Frequency
U	Urgency
N	Nocturia
W	Weak stream
I	Intermittency
S	Straining
E	Sensation of incomplete emptying

BOX 116.2 IPSS (Out of 35)

Score	Correlation
0 to 7	Mildly symptomatic
8 to 19	Moderately symptomatic
20 to 35	Severely symptomatic

Examination of the patient should include a lower abdominal examination, an inspection of the genitalia, and a digital rectal examination. Key findings from the abdominal examination would include a distended, palpable bladder and scars indicative of previous surgery. On digital rectal examination, anal tone, prostate size, and texture (e.g. smooth, irregular) should all be assessed. The genital examination is essential to assess for signs of phimosis, lichen sclerosis, meatal stenosis, previous genital surgery, hypospadias, and penile cancer, which can all give obstructive LUTS.

Commonly used **investigations** include flow rate and post-void residual (PVR). This is a simple and cheap test that can give guidance as to the severity of the disease. A maximum flow rate of >15 mL/s is normal and a maximum flow rate of <10 mL/s usually indicates obstruction. A PVR is a highly variable number even on the same day, but a rising residual over time usually indicated progressive disease.[8] To investigate prostate size, ultrasound scan (either transabdominal or transrectal) or MRI can be performed and is a helpful guide to treatment, although size does not always correlate with worsening disease. A test for PSA (prostate-specific antigen) can be conducted and this may be elevated due to the size of the prostate. PSA can also be used to help guide treatment options with some treatments more effective at higher PSAs. It is also useful to detect cancer.[9]

Management and Treatment of BPH

The absence of testicular function from a young age, from either castration or hypopituitarism, prevents the occurrence of BPH in elderly men (<55 years). The link between testicular androgens and BPH led to early attempts to remove androgenic stimulation in BPH patients by castration, a cheap and simple, but understandably unpopular treatment. Antiandrogens and luteinizing hormone-releasing hormone (LHRH) agonists/antagonists block the effects and production, respectively, of androgens. While producing measurable increases in urine flow in ~50% of patients, this form of treatment does have significant associated morbidities, many of which become progressively worse with the longer duration of therapy. These include gynecomastia, osteoporosis, loss of libido, erectile dysfunction, altered serum lipid profiles, decreased insulin sensitivity, and perhaps cognitive decline.

Due to the interrelationship between testosterone and BPH, there has been considerable debate and concern about worsening LUTS and BPH by replacing testosterone. If the testosterone is replaced adequately and supratherapeutic levels are not achieved, testosterone replacement will not make BPH-related LUTS worse.[10] It may even improve the LUTS.[11]

The success of inexpensive and effective medical treatments for BPH has dramatically changed the paradigm for the clinical management of BPH such that this is now first-line treatment for BPH (see Box 116.4).

5-Alpha-reductase inhibitors (such as finasteride and dutasteride) block the conversion of testosterone to the more potent dihydrotestosterone (DHT) and are used therapeutically for the management of BPH.[12–15] This medication induces apoptosis of prostate epithelial cells resulting in reduction in prostate size (by 18 to 28%), PSA level (50%), and symptoms (by 15 to 30%) within 6 to 12 months. The advantages of reductase inhibitors are the reduction in risk of urinary retention by 68% and the risk of requiring surgery by 64%; the amount of bleeding caused during surgery in patients who have been on 5-ARIs is also reduced. The main adverse events caused are loss of libido, erectile dysfunction, decreased ejaculatory volume, and also gynecomastia. One study suggested that they may increase the likelihood of significant prostate cancer, but this was not confirmed in other studies.[9]

Phosphodiesterase-5 inhibitors (PDE-5i) are most widely used for erectile dysfunction and have also been used to treat **BPH.**[16] These medications are thought to cause smooth muscle relaxation in the bladder, prostate, and surrounding arterioles. They do appear to be better

BOX 116.3 A Typical Bladder Diary Used to Chart Patient's Symptoms

Bladder Diary

Please complete your bladder diary each day for three continuous days.

Name: ____________________

URINE					DRINKS		
Date/time AM/PM	Amount in mL	How strong was the urge to go? 0, +. ++	Did you experience accidental leakage?	Comments What were you doing?	Time	Amount in mL or cups	Type what kind?
This "sample" line shows you how to use the diary							
Tuesday 6.30 AM	150 mL	0	No	Got out of bed	7 AM	500 mL 2 cups	Tea

BOX 116.4 Medical Management of BPH

5-Alpha-Reductase Inhibitors

Block conversion of testosterone to the more potent dihydrotestosterone (DHT)

Reduce prostate size, PSA level, and symptoms in 6 to 12 months

Finasteride and Dutasteride

Alpha-Adrenergic Receptor Antagonists

Relax prostate smooth muscle

Improve symptoms but do not reduce risk of urinary retention

Tamsulosin and Silodosin—Selective alpha1 blockers

Combination

5-alpha reductase inhibitor combined with an alpha blocker

Dutasteride/tamsulosin available as a single tablet

than placebo in reducing IPSS—(mean difference [MD] 1.89 lower, 95% confidence interval [CI] 2.27 lower to 1.50 lower; $n = 4293$), and also aid in combination with an alpha blocker (MD 2.56 lower, 95% CI 3.92 lower to 1.19 lower; $n = 193$), but these changes are small.

Alpha-adrenergic receptor antagonists relax the muscle of the prostate and thus improve clinical symptoms. *Alpha blockers* can be selective, particularly targeting the alpha 1a receptor, or they can be nonselective, targeting all alpha receptors. However, the view that alpha blockers work by relaxing the bladder neck appears now to be an oversimplification of their action, since they also appear to work on patients without obvious BOO, and also work without changing the BOO in other men. Tamsulosin and silodosin are the most widely used selective alpha blockers. The nonspecific alpha blocker choice varies widely globally depending on access and funding issues. Silodosin and tamsulosin appear to have very similar efficacy, but the rates of sexual side effects may be higher with silodosin. Alpha blockers do improve symptoms but do not reduce the risk of urinary retention. They can improve symptom scores (IPSS) by 30% to 40%.[9]

Many **herbal medicinal products** are sold for treatment of BPH. These include *Serenoa repens* (saw palmetto), *Cucurbita pepo* L. (pumpkin seed), *Prunus africana* cortex (Pygeum africanum bark), and *Epilobium angustifolium* (Willow herb). The quality of these trials is poor, so it is difficult to make definitive statements based on evidence, about their use.

Combinations of therapies can be even more effective for BPH than single agents. The combination of an alpha blocker combined with a 5-ARI is used widely and dutasteride/tamsulosin is available on single tablet combination. This allows for short-term symptom relief by the alpha blocker and the long-term effects of the 5-ARI. It reduces rates of retention by 68%, BPH-related surgery by 71%. COMBAT TRIAL. It works best in men with PSA >1.4 and prostate size >40 cc. The alpha blocker component can be discontinued after 6 months. Fig. 116.3 demonstrates the effect of placebo versus alpha blocker versus 5 alpha reductase inhibitor vs combination of both.[17]

Surgical intervention for BPH is now considered a second line therapy (see Box 116.5) and the numbers of surgical procedures for BPH have significantly reduced compared to those in the early to mid-1990s. Classical **transurethral resection of the prostate (TURP)** is still the standard treatment to which all others are compared. It can be done with a monopolar system or bipolar system. They are of equivalent efficacy, but bipolar TURP eliminates the risk of water or glycine toxicity. Noted changes include improvement (+162%), a significant reduction in IPSS (−70%), QoL score (−69%), and PVR (−77%). **Laser ablation** is used to vaporize prostatic tissue. It results in the immediate removal of tissue. There are many lasers that perform this task. The most widely used is the KTP 532Hz "Green light laser." These are considered to be safe in patients on anticoagulant therapy although the evidence is weak.[9] This technique is not inferior to TURP in terms of symptom score (IPSS), maximum flow rate, prostate volume reduction, PSA reduction, and quality of life scores. Postoperative storage symptoms may last up to 1 year after photoselective vaporization laser treatment; however, nonsteroidal anti-inflammatory drug (NSAIDs) may address this issue as they cause a cytokine-induced T-cell response.[18]

Laser enucleation is a technique performed endoscopically with the use of a high-power holmium laser. Using the principles of an open prostatectomy, the entire TZ of the prostate is removed by dissecting the pseudocapsule. The holmium laser enucleation of the prostate (HoLEP) is at least as effective as TURP and possibly better in terms of flow rates, symptom score (IPSS), and need for reoperation. This procedure is technically more difficult to perform than other options, which has hindered its take up.

Open prostatectomy is the most invasive option requiring the longest recovery time in hospital, but efficacy is maintained for up to 6 years. It will improve symptoms by 63% to 86% as demonstrated by an improvement in IPSS of 12.5 to 23.3 points. The quality of life scores improves by 60% to 87%. The urodynamics, both PVR and maximum flow, are also improved; the flow rate increases the mean Qmax by 375% (+16.5 to 20.2 mL/s) and PVR was reduced by 86% to 98%.

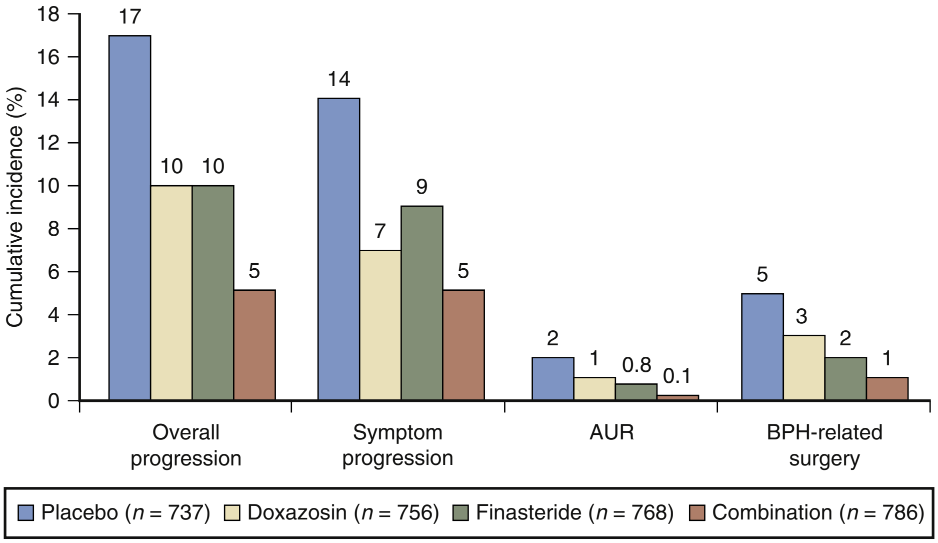


Figure 116.3 MTOPS Clinical Trial results demonstrated that while both alpha blockers and 5ARIs work well, the combination is more effective than individual treatment. *AUR,* acute urinary retention.

BOX 116.5 Surgical Management of BPH

Considered Second Line Therapy

Transurethral Resection of the Prostate (TURP)

The standard treatment—significant improvement in symptoms and post-void residual

Risk of water/glycine toxicity—which causes fluid overload and hyponatremia as a result of the large volumes of irrigation fluid being absorbed

Laser Enucleation

High-power holmium laser used endoscopically to remove the entire transition zone of the prostate by dissecting the pseudocapsule.

Similar efficacy to TURP but it is a technically challenging procedure

Open Prostatectomy

Most invasive option requiring the longest recovery time in hospital, but efficacy is maintained for up to 6 years

Laser Ablation

Laser used to vaporize prostatic tissue

Similar efficacy to TURP, although postoperative storage symptoms may last up to 1 year after.

Urolift, that is, prostatic urethral lift, is a simple procedure where clips are placed in the prostatic urethra, and opening it up has become a widely used treatment.[19] It is less effective in the treatment of LUTS in both the long and short term, but comes with significantly less side effects, particularly less retrograde ejaculation.

Rezum is a relatively new procedure using convective radiofrequency water vapor thermal therapy for LUTS in men with BPH.[20] It appears to be better than a sham procedure in treating LUTS also without the side effect of retrograde ejaculation. However, evidence of its efficacy is limited and there are no data on its role in the prevention of acute urinary retention.

Summary and Future Research

The accuracy of a diagnosis is critical to guiding appropriate intervention for optimal patient outcomes and tools to establish the patient history are essential, in combination with careful physical examination. Sometimes overlooked, the examination of the genital area can be useful in eliminating other causes of bladder symptoms such as the presence of a very tight phimosis.

KEY POINTS

- Important to include physical examination in a thorough work up of patients to diagnose BPH
- There are multiple non-surgical treatments to manage BPH before considering surgery

Future challenges are to improve treatment for patients with BPH. There is a need for surgical procedures that are effective but do not cause retrograde ejaculation, as well as interventions that can be performed under local anesthetic for patients with significant comorbidities. For medical management, the 5-ARIs are ideal for older patients but less so for men for whom libido/sexual issues may be of a concern.

PROSTATE CANCER

Diagnosis and Detection of Localized Prostate Cancer

The increased incidence of the prostate cancer correlates with the introduction of **PSA** testing in the late 1990s. With PSA testing, the lifetime risk of diagnosing prostate cancer in the United States increased from 7.3% to 17%, with a reduction in diagnosis of metastatic disease from 28 to 11 per 100,000 men, and with a concomitant increase in the diagnosis of localized disease.[21]

Two major trials were conducted to determine if PSA-based screening would reduce prostate cancer mortality. The first was the US-based Prostate, Lung, Colorectal and Ovarian (PLCO) trial where 76,693 men underwent annual PSA-based screening for 6 years. The 15-year results indicated that 255 deaths from prostate cancer occurred in the screening arm, and 244 deaths in the control arm, and the differences were not significant.[22] However, the study has limitations as 86% of men in the control arm underwent PSA testing at least once and 46% underwent PSA testing annually. Also, the decision to refer for biopsy and treatment was left up to the individual clinician, and many men with raised PSA were not referred for biopsy or treatment, and hence, no survival advantage could have occurred under those circumstances.

BOX 116.6 Main Benefits and Harms of PSA Testing

Benefits	Limitations
Increased detection of prostate cancer earlier in disease course.	False-positive PSA tests lead to unnecessary biopsies which have potential complications including urinary incontinence, erectile dysfunction, and bowel disturbance.
Reduction in prostate cancer mortality.	The advent of new risk stratification scores may help to avoid unnecessary biopsies following positive PSAs.
	May identify cancers that would not have become clinically relevant in a man's lifetime.

The second trial was the European Randomized Screening trial in Prostate Cancer (ERSPC) which screened 162,387 men, aged 55 to 69 years, from multiple European countries using predominantly 4-year PSA intervals between screens, apart from Sweden that was 2 yearly. The PSA cutoff level was lower at 3 ng/mL, and all participants were mandated to have a biopsy if screening PSA was abnormal. A 16-year follow-up demonstrated a 20% relative risk reduction, with a rate ratio of prostate cancer mortality of 0.80 (95% CI 0.72 to 0.89, $P < 0.001$). The number of men needed to be invited for screening to prevent a prostate cancer death was 570, and the number needed to be diagnosed was 18. The difference in absolute prostate cancer mortality was 0.18% at 16 years.[23] Interestingly, men attending one screening round had a mortality reduction of 25%, but those who attended at least 2 screening rounds had a decrease of 48%, highlighting the need for regular testing. In the Rotterdam cohort the relative risk of metastatic disease and prostate cancer death was 0.46 and 0.48, respectively, in favor of screening, suggesting that in some cohorts the benefits of prostate cancer screening in reducing metastasis and death may be greater.[24]

Nonetheless, **harms of screening** persisted (see Box 116.6), namely, false-positive PSA tests leading to unnecessary biopsies, transrectal biopsies potentially leading to complications such as urosepsis, and treatments leading to morbidity such as urinary incontinence, erectile dysfunction, and bowel disturbances. Risk mitigation strategies were researched to reduce the harms associated with screening, without compromising the benefits. These included improved diagnostics utilizing multiparametric prostate magnetic resonance imaging (mpMRI), altered biopsy methods (transperineal biopsy), and increased reliance of monitoring low-risk prostate cancer (active surveillance [AS]) rather than subjecting patients to definitive therapies of surgery and radiotherapy.

The use of **multiparametric prostate magnetic resonance imaging (mpMRI)** was included in recent updates of the EAU and American Urological Association guidelines, with validated risk stratification using PIRADS (Prostate Imaging-Reporting And Data System) scoring, prior to a patient's first biopsy.[25,26] The mpMRI of the prostate gland allows identification of a focal lesion that could more accurately be targeted at biopsy, as well as the potential of avoiding biopsy in those deemed unlikely to harbor clinically significant disease, defined as Gleason ≥7 or International Society of Urologic Pathologists (ISUP) Grade Group ≥2. A systematic review reported that prebiopsy prostate mpMRI reported as PIRADS 1 or 2, and hence defined as normal, has a negative predictive value (NPV) of 90.8% for ruling out prostate cancer grade group ≥ 2. The NPV for ruling out Gleason Grade ≥ 3 was 97%. This avoids biopsy in 30% of cases and reduces the detection of low-risk prostate cancer that may not require immediate therapy in 18% of men.[27] The PROstate Magnetic resonance Imaging Study (PROMIS) study was a multicenter, paired-cohort, confirmatory study to test the diagnostic accuracy of mpMRI prostate against a standard transrectal ultrasound guided biopsy of the prostate using a reference standard of a transperineal mapping biopsy. Seven hundred forty men were enrolled, with cancer identified in 71% of cases, and clinically significant cancer in 40%. mpMRI prostate was more sensitive (93% vs 48%) but less specific (41% vs 96%) compared to transrectal ultrasound (TRUS). However, the use of mpMRI to triage patients would have allowed 27% of men to avoid a biopsy, while missing only 5% of clinically significant cancers.[28] The other major study was the PRECISION study, where men were assigned to mpMRI prior to biopsy, and compared to the previous paradigm of a systematic TRUS biopsy. In this study, 28% of men avoided a biopsy, and more clinically significant cancer was detected in the mpMRI arm (38% vs 26% $P = 0.005$), with a reduced detection of clinically insignificant disease.[29]

Several authors have considered whether **systematic biopsies** should still be performed in the presence of an mpMRI visible lesion that could be specifically targeted. Miah et al. demonstrated that 43.1% men were diagnosed with clinically significant prostate cancer, with 41.1% able to be diagnosed solely on the target cores alone and only 0.8% on the nontargeted systematic cores exclusively. Like PROMIS the definition of clinically significant disease was Gleason score $\geq 4 + 3 = 7$, or any core ≥6 mm in length.[30] Another study used a definition of clinically significant prostate cancer as Gleason $\geq 3 + 4 = 7$ to detect patients with clinically significant disease and a visible PIRADS 4/5 lesion at mpMRI. They showed 9% of patients would have had been undiagnosed if a systematic biopsy had not been performed, but in patients with Gleason $\geq 4 = 3 = 7$ this was only 1%.[31] Similarly, Hanna et al. demonstrated that 16% of clinically significant prostate cancers would be missed by avoiding the systematic biopsy.[32] The likelihood of missing the diagnosis of a clinically significant cancer decreased with increasing Likert/PIRADS score. The more visible and larger lesions were reliably sampled using target biopsies alone, but systematic biopsies were required to ensure cancers were not missed in the smaller PIRADS/Likert 3 and 4 lesions. To improve the accuracy of biopsies, computer software was developed to fuse the mpMRI images to the ultrasound images seen at the time of biopsy. However, trials failed to show superiority of these computer software platforms over visual co-registration, otherwise known as cognitive fusion. Hamid et al. reported that both techniques diagnosed 90% of clinically significant cancers and 10% were missed by cognitive fusion detected using software assistance and vice versa.[33] Yaxley et al. did not demonstrate superiority of an in-gantry mpMRI guided biopsy to cognitive fusion.[34]

There is emerging interest in performing **transperineal biopsies (TPB)** rather than transrectal biopsies. PROMIS compared TRUS biopsies to mapping transperineal biopsies, and showed that of the men diagnosed with Gleason 6 disease at TRUS, 81% harbored Gleason $3 + 4 = 7$ cancer at TPB and 11% Gleason $4 + 3 = 7$. Reliance on TRUS biopsies, especially in those without a clear mpMRI target, risks misclassifying the patient, leading to undertreatment.[35] Comparison of the results of TRUS to TPB vs prostatectomy specimens also demonstrated the superiority of TPB in predicting final surgical histopathology outcomes.[36] Together with a reduced sepsis rate by avoiding the rectal flora during a transrectal biopsy, TPB is emerging as the preferred method of prostate biopsies in many countries, albeit with a slightly higher rate of temporary urinary retention and erectile dysfunction (24% to 34%), hematuria (20% to 70%), and hematospermia (50% to 90%), which may occur in patients undergoing biopsy regardless of the technique used.[37]

KEY POINTS

- PSA is useful for screening in order to assist in the early detection of prostate cancer but mpMRI should be utilized before consideration is given to a biopsy

BOX 116.7 Localized Prostate Cancer Management

Active Surveillance

Includes

- Regular PSA assessment (3 to 6 monthly)
- Multiparametric MRI (1 to 3 yearly)
- Repeat biopsies (1 to 3 yearly) with treatment reserved for those patients who demonstrate disease progression

For low- or intermediate-risk localized prostate cancer

Radical Prostatectomy (RP)

Removal of prostate with adjoining seminal vesicles, with a vesicourethral anastomosis +/– pelvic lymph node dissection

For low-, intermediate-, or high-risk localized prostate cancer

Radiotherapy

Radical external beam radiotherapy

Low-dose rate or high-dose brachytherapy

For low-, intermediate-, or high-risk localized prostate cancer

Can be administered with or without neoadjuvant/adjuvant androgen deprivation therapy

Androgen Receptor–Signaling Inhibitors (ARSIs) or Chemotherapy with Docetaxel

Can be considered in those with high-risk, localized prostate cancer

Management and Treatment of Localized Prostate Cancer

Based on the risk assessment of the patient following biopsy, treatment recommendations are made (see Box 116.7). Those with **low-volume, low-risk disease** (Gleason score 6, ISUP grade 1) remain at low risk of disease progression, and **AS** is a popular method for managing patients, avoiding overtreatment, and minimizing treatment-related morbidities, first popularized by Klotz et al.[38] The Prostate Testing for Cancer and Treatment (ProtecT) study compared active monitoring, radical prostatectomy (RP), and external beam radiotherapy for the treatment of localized prostate cancer, in 1643 men. There were only 17 prostate cancer–specific deaths in this relatively low-risk cohort, 8 in the active monitoring arm, 5 in the surgery group, and 4 in the radiotherapy group ($P = 0.48$). Metastases did develop more frequently in the active monitoring arm (33, 13, and 16 men, respectively, $P = 0.004$).[39] However, the group demonstrated reduced morbidity from the monitoring approach across the important domains of urinary, bowel, and sexual function.[40] A recent meta-analysis of 26 AS cohorts with relatively short follow-up (3.5 years) demonstrated 0.1% risk of prostate cancer death and metastases, and with 8.8% of men undergoing active treatment per year based on progression on biopsy, PSA levels are based on patient anxiety. The Prostate Cancer Intervention versus Observation Trial (PIVOT), comparing surgery to observation, demonstrated no difference in all-cause mortality after a median of 12.7 years, in men diagnosed with low-risk disease but with greater morbidity in the treatment arm.[41] As such, all studies demonstrated a low risk of prostate cancer death, with improved quality of life for patients with AS compared to definitive therapy. Surgery has higher risks of urinary incontinence and erectile dysfunction, while radiation has higher incidence of urinary and bowel bother, and urethral stricture disease with the incidence of Grade 3 toxicities on the order of 6% to 7% with a rare incidence of serious complications, such as fistulae requiring invasive reconstructive surgery to manage.[42–44]

KEY POINTS

- Prostate cancer can run an indolent course and not all tumors require active therapy. In well selected low risk patients managed with active surveillance, cause specific survivals are 99% at 10 years.

Although **AS protocols vary,** most would include a regular assessment of serum PSA (3 to 6 monthly), multiparametric MRI (1 to 3 yearly), and repeat biopsies (1 to 3 yearly) with treatment reserved for those patients who demonstrate disease progression, which is ~50% at 10 years. Biopsies are currently still required, as mpMRI prostate can miss the occasional high-risk prostate cancer, although whether or not missing a small cancer affects long survival remains unclear. Chu et al. identified that the NPV of mpMRI was 79.5% in patients managed with AS, and as low as 65.5% in those with a normal mpMRI prostate but with a PSA density (serum PSA/prostate volume at mpMRI) > 0.15 ng/mL/cm^3. Therefore, biopsy remains essential in the management of men with AS.[45] However, using a protocol omitting biopsy unless mpMRI identified abnormalities, 71.8% of men remained on AS at 5 years. This rate was lower in men with PIRADS 4 or greater lesions noted on mpMRI or those with secondary Gleason pattern 4 (ISUP ≥2) in their biopsies. There were no deaths from prostate cancer, and only 1.2% of men developed metastatic disease, highlighting the feasibility of this approach.[46]

In men with **intermediate- or high-risk prostate cancer**, treatment is indicated, due to the increased risk of prostate cancer metastases and mortality with observation. The standard whole-gland treatments to consider are **RP as well as radiotherapy (RT)** options such as low-dose rate brachytherapy, high-dose brachytherapy, and external beam radiotherapy with or without the neoadjuvant and adjuvant use of androgen deprivation therapy (ADT). Unfortunately, no appropriately conducted randomized controlled trials (RCTs) have been conducted to establish superiority of one technique over another. A recent review comparing surgery to RT demonstrated the risk of overall survival (hazard ratio [HR] 1.63) and prostate cancer–specific survival (HR 2.08) was higher with surgery. It was supported by a Swedish study demonstrating similar outcomes, but whether these differences reflect selection bias remains unclear and possible.[47] A recent multidisciplinary systematic review of 30 studies comparing RP +/− salvage RT to RT +/− ADT reported 27 studies showing an overall and prostate-specific survival advantage for surgery.[48] The improvement in overall survival was estimated as 10% to 28% and cause-specific survival ranged from 4% to 8%, with greater sexual dysfunction and urinary incontinence in the surgical arm, compared to worsening bowel toxicity after RT.[49] It appears that the contemporary management of higher-risk patients with radiotherapy must include the use of ADT and not just RT alone.[50] Further analysis of 4 RCTs noted an advantage of long-term ADT over short-term ADT with improvements in overall and cause-specific survivals in those patients with higher-risk disease.

In addition, studies demonstrated that the addition of a low-dose **brachytherapy** boost to standard RT + ADT improved the biochemical-free survival from 73% to 83%, with improved overall and cause-specific survivals in nonrandomized studies. However, this combination was associated with worsening morbidity in the Androgen Suppression Combined with Elective Nodal and Dose Escalated Radiation Therapy (ASCENDE-RT) trial with grade 3 genitourinary toxicities of 18.3% compared to 5.2%.[51] Whether this provides equivalent care compared to surgery remains unclear, but a recent study demonstrated a superior overall survival outcome of RP +/− salvage RT compared to EBRT+ADT+BT with HR 1.22.[52] Further advances in radiotherapy techniques include strategies to dose escalate, hypofractionation, and the prophylactic irradiation of pelvic lymph nodes in higher-risk patients

BOX 116.8 Robotic Surgery vs Open Radical Prostatectomy

Small but significant advantage of robotic over open surgery, with regard to erectile dysfunction (68 vs 74%).

No difference identified in:

- Incontinence
- Sexual dysfunction
- Positive surgical margins
- Oncological outcomes

and ongoing RCTs should assist in determining the value and safety of these newer radiation techniques. Local control of the cancer remains crucial in the delivery of radiotherapy for localized prostate cancer because local failure was associated with reduced overall survival (HR 1.7), prostate cancer–specific survival (HR 3.1), and metastasis-free survival (HR 1.92).[53]

RP is surgery that removes the prostate with adjoining seminal vesicles, with a vesicourethral anastomosis +/− pelvic lymph node dissection (PLND). The EAU guidelines recommend an extended PLND in **high- and intermediate-risk prostate cancer**, but the therapeutic benefit of this strategy remains unclear. A recent meta-analysis has demonstrated that while pelvic node dissection improves local staging, there was no difference in cancer-specific or overall mortality comparing PLND vs no PLND. In addition, comparing a standard PLND to an extended PLND demonstrated an improved biochemical relapse free rate in some but not all studies. Furthermore, this did not translate into an overall or cause-specific survival advantage and there was a rise in complications by increasing the rate of pelvic lymphocele formation.[54] Of those who demonstrated pathologically involved pelvic nodes, 10-year biochemical recurrence–free survival was 28% to 56%, and cancer-specific survival was 72% to 98%. As such, pelvic node involvement at surgery does not always translate to a poor outcome for patients and can be managed by observation, salvage radiotherapy, or ADT with some evidence to suggest an improved outcome for salvage RT together with ADT compared to ADT alone or observation.[55,56]

Another area of controversy within Urology is the role of **open versus robotic RP** with the latter now the dominant technique without significant evidence confirming benefit (see Box 116.8). The A prospective, nonrandomized comparative trial, LAParoscopic Prostatectomy Robot Open (LAPPRO) study compared the functional and oncological outcomes between open and robotic RP after 2-year follow-up. At 24 months, there was a small but significant advantage of robotic surgery over open RP, with regard to erectile dysfunction (68 vs 74%), and a small but not statistically significant advantage of open surgery over robotic surgery with regard to incontinence (16 vs 19%) and no oncological differences were identified.[57] In the only RCT comparing the techniques, there was no difference in incontinence rates, sexual dysfunction, or positive surgical margins. The conclusions reached by the authors were encouragement for patients to choose an experienced surgeon they trust, with whom they have a rapport and who has documented satisfactory outcomes, rather than choosing a specific technique.[58]

Biochemical recurrence can occur after either definitive treatment for prostate cancer. **Salvage therapies** would include RT +/− ADT after surgery and RP +/− ADT after radiotherapy. Adding salvage therapies increases patient morbidity, with patients potentially experiencing the side effects of androgen deprivation, but both salvage surgery and radiotherapy increase the risk of urethral stricture disease and urinary incontinence. The effect or size of the biochemical relapse as a risk factor for prostate cancer mortality was highly variable. A recent meta-analysis demonstrated that the effects were more pronounced if patients had a PSA doubling time of < 12 months (especially if < 3 months), a shorter interval to biochemical failure, and increasing Gleason score pathologically. It has been proposed that not all men with biochemical recurrence should be treated equally, as a low-risk patient may not benefit from salvage therapies and those with high-risk disease may fail with distant disease and may need ADT as well as possibly salvage local treatments. Those who have a persistently detectable PSA after surgery (as opposed to an undetectable PSA that subsequently rises) are at a higher risk of metastasis (HR 3.59), death (HR 1.86), and cancer-specific death (HR 3.15).[59] These patients at higher risk appear to benefit from the addition of ADT to salvage radiotherapy, especially at presalvage radiotherapy PSA of ≥0.7 ng/mL, high-grade disease, and positive surgical margins.[60] This was proven by the RTOG 9601 and GETUG 16 studies, while a recent Japanese study confirmed that salvage RT+ ADT prolongs time to treatment failure compared to ADT alone in this setting (HR 0.56).[61] The duration of ADT was examined and is reflective of risk factors, as those at high risk should be on life long or prolonged (3 years) ADT, those with 1 to 2 risk factors required < 12 months of ADT, while those without risk factors did not require ADT.[62]

The **standard of care of managing biochemical recurrence** after definitive treatment for prostate cancer is based on data from trials utilizing standard imaging techniques of nuclear bone scan and computerized tomography (CT) chest abdomen and pelvis that has extremely poor sensitivity especially at low PSA values. Improved imaging may provide the opportunity of personalized care and metastasis directed therapy, potentially avoiding the use of salvage treatments to the bladder neck region or the use of ADT, both of which carry additional toxicities. **Prostate-Specific Membrane Antigen (PSMA)** is a transmembrane protein on the cell surface of prostate cells that is overexpressed in prostate cancer. Both antibodies and small molecules that bind to PSMA were developed and radiolabeled with 68 Ga to allow nuclear **positron emission tomographic imaging (PET).** A recent meta-analysis demonstrated that the likelihood of a positive **PSMA PET scan** was 33% at PSA < 0.2 ng/mL, 45% with PSA 0.2 to 0.49 ng/mL, 59% for PSA 0.5 to 0.99 ng/mL, 75% for PSA 1.0 to 1.99 ng/mL, and 95% for PSA > 2 ng/mL.[63] Interestingly, the distribution of the PET avid lesions was highly variable. Only 28% of lesions were in the prostate bed, but 38% in pelvic lymph nodes, 13% in extra pelvic lymph nodes, 22% in bones, and 5% in distant organs. As such, only the 28% destined to fail in the prostate bed may benefit from salvage radiotherapy to that location, and if treatment were confined to those patients, morbidity could potentially be reduced. Several studies reported on the PSMA PET imaging results compared to histological verification of lesions and on a per-lesion basis; the sensitivity was 75% sensitivity 99%, respectively. However, 68 Ga PSMA PET can miss local recurrences due to the activity of the tracer in the bladder urine, and PSA responses do occur with salvage radiation in patients with a negative PSMA PET in the prostate fossa.[64] Additional reasons could also include a low-grade cancer recurrence and the fact that the spatial resolution of the PET tracer can miss microscopic lesions. However, detecting disease recurrence outside the prostate fossa changed salvage radiotherapy planning in 41% of men with PSA levels 0.2 to 0.59 ng/mL.[65] There are now several published series on metastasis directed therapy, including **stereotactic ablative radiotherapy (SABR)** to involved nodal or bony disease, or less commonly salvage lymph node dissection. While early disease control is promising, the likelihood of PSMA PET of missing microscopic nodal disease would suggest that this would delay rather than avoiding therapy in most patients. However, a meta-analysis of major international trials (Observation vs Stereotactic Ablative Radiation for Oligometastatic Prostate Cancer [ORIOLE]; Pilot Study of patients with Oligometastases from Prostate cancer treated with STereotactic Ablative Radiotherapy [POPSTAR]; Surveillance or metastasis-directed Therapy for OligoMetastatic Prostate cancer recurrence [STOMP]) has shown that local control of disease is reported at 93% to 100%, with low toxicities from SABR with 8% grade 2 toxicities, and only 1% grade 3 toxicity.[66] Studies have also compared **extended nodal radiation (eNRT) therapy** to SABR with fewer recurrences with eNRT, but higher toxicities.[67] Studies have also looked at **salvage lymph node dissection** with 5-year biochemical recurrence–free survivals ranging from 6% to 31%, highlighting that if a node dissection was being considered, a wide dissection would be required due to the high likelihood of more extensive nodal disease than detected on PSMA PET.[68] Unfortunately, the rates of grade 3 complications range from 10% to 20% and can include lymphocele formation, ureteric injury, pelvic abscesses, and venous thromboembolism. Given the lower complication rates of SABR, and the fact that in most cases sLND does not lead to long-term cure, the use of salvage nodal surgery is used with extreme caution and ideally as part of a clinical trial.

The management of localized prostate cancer has always relied on whole-gland therapy, with the understanding that this may represent significant overtreatment in some men, also leading to morbidity and loss of quality of life. An emerging treatment is the use of **focal therapy**, the targeted destruction of prostate cancer within a specific region of the prostate, identifiable on mpMRI prostate and PSMA PET scanning. This would spare the rest of the prostate gland, potentially reducing side effects and morbidity while effectively treating the index tumor, namely, the tumor of the greatest grade and size. In order for focal therapy to be effective, the index tumor needs to be identifiable ideally on both mpMRI prostate and PSMA PET scan, and targeted biopsy of this region confirms cancer with the rest of the prostate demonstrating benign changes, or at worse low-volume, low-grade cancer that could be managed by AS. The treatment must be equally effective as whole-gland therapy, be proven to have less side effects, and potential salvage therapies, for example, surgery or radiotherapy should not be impaired by the focal treatment. At this stage robust long-term data are not available and these treatments should be considered experimental and ideally conducted within a clinical trial. Several treatment techniques have been tried including hemigland or quadrant ablation, or ablation targeted to the lesion with a 5 to 10 mm safety margin. Different energy sources have also been tried including **high-intensity focused ultrasound (HIFU), irreversible electroporation (IRE),** interstitial laser, low-dose rate brachytherapy, photodynamic therapy, and cryotherapy with no energy source being proven to be superior to another.[69] The best studied of these techniques at this time are HIFU and IRE.

HIFU is delivered noninvasively, transrectally, under ultrasound guidance in order to ablate the prostate tissue by both thermal and mechanical effects. A recent systematic review demonstrated that only 8% of patients having repeat biopsies harbored clinically significant disease, subsequently needing additional therapies. Functional outcomes were satisfactory but varied considerably due to differing definitions of erectile dysfunction and incontinence, highlighting the need for international standards for follow-up post treatment and assessment of functional outcomes. A further study reported that 83% of patients had infield negative biopsies with 13% having *de novo* out of field clinically significant disease at 12 months follow-up. Urinary symptoms returned to baseline by 3 to 6 months, and sexual function returned to baseline by 12 months.[70] Several studies showed that salvage RP is safe and feasible after treatment, but oncological follow-up is short and difficult to compare with surgery in treatment naïve patients. In one study there was a 38% positive margin rate, 17.6% required salvage radiotherapy, 20.6% developed biochemical recurrence. 91.2% of patients were continent but there was still a significant decline in the International Index of Erectile function scores post treatment.[71] Another study reported an 83% continence rate and only a 48% progression free survival at 48 months post salvage surgery, identifying that suboptimal outcomes may occur after failed focal HIFU requiring salvage surgery.[72]

Irreversible electroporation (IRE) involves using pulsed electric fields to create destabilization of the cell membrane, causing excessive cell membrane permeability, leading to disruption of osmotic balance within the cell and subsequent cell death, but avoiding harmful thermal effects. Theoretically, damage to surrounding structures such as the external sphincter, rectum, and neurovascular bundles is minimized. A review of 123 patients demonstrated infield recurrences in only 2.7% to 9.8% of patients, failure-free survival of 96.75%, metastasis-free survival of 99%, and overall survival at 100%. 98.8% remained pad free and 76% had no change in erectile function, with only 7% having complete erectile dysfunction. Given these promising short-term oncological and functional outcomes, together with improved imaging techniques, further evaluation of these techniques and developing an RCT comparing focal therapy to whole-gland therapy would be important in the coming years.[73]

Diagnosis and Detection of Advanced Prostate Cancer

Some men develop metastases after failing local treatment with surgery or radiation therapy and disease progression is detected biochemically or radiologically. **Biochemical recurrence** is defined as an elevation in serum PSA. The sensitivity of PSA in this setting is improved because removal of the prostate or radiotherapy to the prostate reduces the amount of noncancer PSA present. If there is no evidence of disease on conventional imaging with CT scan and Tc-99 whole body bone scan (WBBS), this is termed **M0 prostate cancer**, and treatment decisions are based on the PSA level. If disease is present on conventional imaging, it is termed **metastatic prostate cancer.**

Men with a low PSA velocity and long PSA doubling time on average have a delayed onset of symptomatic prostate cancer. In this case the toxicity of treatment must be balanced against the risk posed by prostate cancer. In older men, there is a proportion of men who will never develop symptomatic disease and can avoid any treatment at all. Regular follow-up with PSA testing is required in this setting to ensure that PSA velocity and doubling time do not change and switch the balance in favor of treatment.

Some men will not undergo treatment of localized prostate cancer because of comorbidities or age. These men will be followed with

BOX 116.9 Advanced Prostate Cancer Management

Androgen Deprivation Therapy

Gonadotropin-releasing hormone (GnRH) agonists (goserelin, leuprorelin, diphereline)

GnRH antagonists are also available (degarelix, relugolix)

Most men initially respond to ADT, but inevitably, patients develop castration-resistant prostate cancer (CRPC).

Androgen receptor–signaling inhibitors (ARSIs)

Abiraterone, enzalutamide, apalutamide, and darolutamide

Improve survival in men with CRPC and recently shown to improve survival in high-risk, castrate sensitive prostate cancer (CSPC)

Chemotherapy

Taxane Chemotherapy

Doctetaxel—*Improves overall survival* in CRPC and CSPC

Cabazitaxel—*Used subsequent to docetaxel and ARSIs*

Personalized therapy

Treatment based upon genomic abnormalities within a particular tumor

Poly-ADP ribose polymerase (PARP) inhibitors—*Olaparib*

AKT-signaling inhibitors—*Ipatasertib*

PSA testing and commenced on ADT, based on PSA level, velocity, and doubling time. This group of men will also be designated as having M0 prostate cancer. Metastatic disease at diagnosis ***(de novo metastatic disease)*** occurs in some men and in this situation, it is necessary to treat all sites of disease.

Management and Treatment of Advanced Prostate Cancer

The most important target for treatment of prostate cancer remains testosterone-driven growth of prostate cancer cells. **ADT** has been the cornerstone of treatment of advanced prostate cancer for nearly 80 years.[74] This was initially achieved by a surgical orchidectomy but the advent of 3 or 6-monthly subcutaneous or intramuscular administered gonadotropin-releasing hormone (GnRH) agonists (goserelin, leuprorelin, diphereline) has removed the need for surgery and allowed intermittent therapy for patients who respond well to treatment but are troubled by side effects.

GnRH antagonists are also available (degarelix, relugolix) and have several advantages over **GnRH agonists.** They cause a rapid fall in testosterone rather than the spike in testosterone responsible for painful disease flares seen in patients with bone metastases treated with GnRH agonists. Degarelix is given by monthly injection and has limited long-term use, but relugolix is a newer oral preparation that has a lower rate of major cardiovascular events than leuprorelin (2.9% vs 6.2%) in men with advanced prostate cancer.[75]

Although most men initially respond to ADT, prostate cancer inevitably becomes castration-resistant, requiring additional therapeutic approaches (see Box 116.9). **Chemotherapy** is effective treatment for metastatic castration-resistant prostate cancer (CRPC) and the taxane drug docetaxel provides a significant survival advantage.[76] Treatments are given via intravenous infusion every 3 weeks in combination with 10 mg prednisolone daily and standard side effects include alopecia, marrow toxicity, neuropathy, and paronychia. In addition, high doses of dexamethasone are required prior to treatment to prevent infusion reactions and fluid retention. This can be challenging in patients with diabetes and impaired glucose tolerance. Cabazitaxel is also active when used subsequent to docetaxel and androgen receptor–signaling inhibitors, resulting in disease response in 37% of patients and 4.4 months of disease control. Standard side effects of marrow suppression and diarrhea are commonly observed.[77]

Oral **androgen receptor–signaling inhibitors** are used in the treatment of prostate cancer for decades. First generation agents such as bicalutamide, flutamide, and cyproterone are useful in preventing disease flare which can occur on commencing GnRH agonists. They are used also in castrate-resistant disease where PSA responses are observed, but improvements in survival have not been demonstrated.

More recently, **potent newer generation androgen receptor–signaling inhibitors** have shown survival benefits as well as high rates of PSA response in CRPC. The enhanced potency is believed to block local intratumoral androgen levels that sustain tumor growth in the presence of low systemic levels of androgens. Abiraterone acetate inhibits any androgen biosynthesis when it is metabolized to abiraterone. Abiraterone acts to block CYP17A, an enzyme required for the conversion of pregnenolone and progesterone to testosterone. However, as this drug also blocks cortisol synthesis, it is given in conjunction with corticosteroids, usually in the form of prednisolone at 10 mg/day. If corticosteroids are not given concurrently, some patients will develop hyperaldosteronism. Abiraterone is well tolerated but has been associated with hepatotoxicity, requiring liver function to be monitored in the first couple of months of therapy.

In patients with metastatic CRPC who have received docetaxel chemotherapy, abiraterone with prednisolone improved survival to 14.8 months, compared with 10.9 months in patients who received prednisolone alone.[78]

More potent androgen receptor inhibitors have been developed in the last decade to block androgen receptor–signaling including enzalutamide, apalutamide, and darolutamide. These agents also prevent receptor dimerization, transport to the cell nucleus, and recycling back to the cell surface to allow further signaling. Enzalutamide improved survival of men with CRPC following docetaxel from 13.6 months to 18.4 months compared with placebo.[79] Although the quality of life in the treated men is better, enzalutamide causes fatigue, diarrhea, and hot flushes. There was a small risk of seizures (<1%) that appear to be unrelated to mechanism of action. Hypertension has been reported with longer duration of treatment. Apalutamide is a similar acting agent with a similar side effect profile. Darolutamide differs as it does not cross the blood-brain barrier, resulting in minimal fatigue compared to ADT alone and has not been associated with seizures.

Enzalutamide, apalutamide, and darolutamide have all been evaluated in M0 CRPC. Enzalutamide improved overall survival from 56.3 to 67 months, suggesting a greater benefit in using the drug earlier in disease progression.[80] However, this is offset by increased fatigue, hypertension, and falls. Apalutamide also improved OS with an increase from 73.5% to 82.4% at 24 months and with some increase in the rate of fractures compared with placebo alone.[81] Darolutamide improved OS at 3 years from 77% to 83% but did not cause any significant increase in toxicity compared with ADT alone and actually delayed symptomatic skeletal events with a median follow-up time of 29 months.[82]

Several treatments showed benefit in metastatic disease when commenced within 12 weeks of initiation of ADT. Docetaxel given every 3 weeks for 6 cycles improved median survival from 47.2 to 57.6 months. Most of this benefit was driven by patients with high-volume disease, defined by the investigators as either the presence of visceral metastases or four or more bone metastases with at least one outside the vertebral column and pelvis.[83] Similar benefits were seen with androgen receptor–targeting agents. Enzalutamide when used upfront, significantly improves survival but median survival has not yet been reached (3-year overall survival improved from 72% to 80%). In patients who received upfront docetaxel, no additional improvement is seen with the addition of enzalutamide.[84] Similarly, improved survival is seen with upfront abiraterone with an increase from 36.5 to 53.3 months in patients with high-risk disease defined as two of: High-grade tumor (Gleason score >7); >2 bone lesions; visceral metastases.[85] Upfront docetaxel and abiraterone were tested as part of the large UK-based STAMPEDE trial and showed improvement in survival with both agents in all patient groups, not just high-risk or high-volume patients.[86,87] More recently, apalutamide showed an improvement in survival at 24 months of 82.4% compared with 73.5% with placebo alone with benefit for high- and low-volume disease groups of patients.[81]

Falls and fractures occur more frequently in patients treated with inhibitors of androgen receptor signaling. As it appears that earlier use of these agents improves survival in advanced prostate cancer, it is likely that the duration of treatment will increase and so may the fracture rate. This occurred with enzalutamide, abiraterone, and apalutamide, but the absolute fracture rate was low. Longer follow-up is required to see if the rate of fracture increases and whether this effect will also be seen with darolutamide. The use of bisphosphonates and receptor activator of nuclear factor kappa-B ligand (RANKL) inhibitors reduces the fracture risk when used with abiraterone. Structured exercise programs are also recommended to reduce muscle weakness and prevent osteoporosis.

KEY POINTS

- Before commencing ADT in men with advanced prostate cancer, the risk vs benefit must be weighed in each case
- Androgen signaling plays a critical role, and remains a target for therapies for men with advanced prostate cancer, but eventually androgen blockade is insufficient to eliminate disease

Novel imaging modalities have made the classification of metastatic vs nonmetastatic prostate cancer more difficult. **PSMA** is a highly expressed protein on prostate cancer cells which can be bound by the peptide PSMA-617. PET imaging utilizes the positron emitter Gallium-68 bound to PSMA-617 and is much more sensitive than CT scan and WBBS in detecting metastases in M0 CRPC, detecting disease in 98% of patients and distant disease in 55%.[88] PSMA is also used as a target for radioisotope therapy by binding the ß-emitter Lutetium-177 to PSMA-617. Early results demonstrated a better PSA response to **177-Lu-PSMA** compared with cabazitaxel chemotherapy and with less side effects.[89]

The use of personalized therapy, that is, treatment based upon genomic abnormalities within a particular tumor, shows some promise in prostate cancer. A proportion of men with prostate cancer will carry DNA repair defects such as BRCA1, BRCA2, and ATM mutations. In patients randomized between the PARP inhibitors, olaparib, and enzlalumide or abiraterone, survival was improved from 14.7 months to 19.1 months. Similarly, PTEN loss occurs in about half of prostate cancers and the addition of the AKT-signaling inhibitor ipatasertib to abiraterone delayed disease progression by 2 months. Although the improvements at present are not clinically significant, it is hoped that combination therapy and better identification of target groups may improve outcomes.

SUMMARY AND FUTURE CHALLENGES

Despite advances in the treatment of men with advanced prostate that extend their lifetime, many thousands of patients still die of this disease because it remains incurable. There is increasing agreement that blockade of the androgen receptor and/or androgen biosynthesis is required but is insufficient to eliminate disease. Instead, combination drug therapy is necessary, similar to that proven to be effective in other malignancies such as breast cancer. However, there are many challenges to evaluate combination therapies in men with prostate cancer. Although there is a rich pipeline of therapeutics, it is not feasible to systematically evaluate large numbers of promising drug combinations in patients themselves, given the sheer number of combinations, sequences and timing of drugs. The numbers of patient available to enroll for clinical trial is not unlimited, yet there is redundancy in active trials, with many drugs directed at the same targets. A paucity of preclinical models is another limitation to the process of drug development and a reliance on decades-old cell lines that do not accurately reflect the clinical responses in patients seen in the clinic today. Nevertheless, these barriers need to be overcome, and trials could be designed using precision medicine paradigms where the patients are selected, based on molecular biomarkers that predict response.

Acknowledgment

The authors acknowledge Dr. Lara Pemberton for her helpful comments on the manuscript.

REFERENCES

1. Lee SWH, Chan EMC, Lai YK. The global burden of lower urinary tract symptoms suggestive of benign prostatic hyperplasia: a systematic review and meta-analysis. *Sci Rep*. 2017;7(1):7984.
2. Lancet T. Globocan 2018: counting the toll of cancer. *Lancet*. 2018;392(10152):985.
3. Selman SH. The McNeal prostate: a review. *Urology*. 2011;78(6):1224–1228.
4. Ellem SJ, Risbridger GP. Treating prostate cancer: a rationale for targeting local oestrogens. *Nat Rev Cancer*. 2007;7(8):621–627.
5. Berry SJ, Coffey DS, Walsh PC, et al. The development of human benign prostatic hyperplasia with age. *J Urol*. 1984;132(3):474–479.
6. Slawin KM. The medical therapy of prostatic symptoms study: what will we learn? *Rev Urol*. 2003;5(suppl 4):S42–S47.
7. Kaplan SA. Update on the American Urological Association guidelines for the treatment of benign prostatic hyperplasia. *Rev Urol*. 2006;8(suppl 4):S10–S17.
8. Roehrborn CG. Benign prostatic hyperplasia: an overview. *Rev Urol*. 2005;7(suppl 9):S3–S14.
9. Gratzke C, Bachmann A, Descazeaud A, et al. EAU guidelines on the assessment of non-neurogenic male lower urinary tract symptoms including benign prostatic obstruction. *Eur Urol*. 2015;67(6):1099–1109.
10. Baas W, Kohler TS. Testosterone replacement therapy and BPH/LUTS. What is the evidence? *Curr Urol Rep*. 2016;17(6):46.
11. Rastrelli G, Vignozzi L, Corona G, et al. Testosterone and benign prostatic hyperplasia. *Sex Med Rev*. 2019;7(2):259–271.
12. Geller J, Sionit L. Castration-like effects on the human prostate of a 5 alpha-reductase inhibitor, finasteride. *J Cell Biochem Suppl*. 1992;50:109–112.
13. Grino P, Stoner E. Finasteride for the treatment and control of benign prostatic hyperplasia: summary of phase III controlled studies. The Finasteride Study Group. *Eur Urol*. 1994;25(suppl 1):24–28.
14. Yoshida O, Oishi K, Okada Y, et al. Effect of long-term administration of finasteride (MK-906), an inhibitor of 5 alpha-reductase, in patients with benign prostatic hyperplasia. *Acta Urol Jpn*. 1996;42(4):323–331.
15. Lowe FC, McConnell JD, Hudson PB, et al. Long-term 6-year experience with finasteride in patients with benign prostatic hyperplasia. *Urology*. 2003;61(4):791–796.
16. Pattanaik S, Mavuduru RS, Panda A, et al. Phosphodiesterase inhibitors for lower urinary tract symptoms consistent with benign prostatic hyperplasia. *Cochrane Database Syst Rev*. 2018;11:CD010060.
17. McConnell JD, Roehrborn CG, Bautista OM, et al. The long-term effect of doxazosin, finasteride, and combination therapy on the clinical progression of benign prostatic hyperplasia. *N Engl J Med*. 2003;349(25):2387–2398.
18. Pascoe C, Ow D, Perera M, et al. Optimising patient outcomes with photoselective vaporization of the prostate (PVP): a review. *Transl Androl Urol*. 2017;6(suppl 2):S133–S141.
19. Jung JH, Reddy B, McCutcheon KA, et al. Prostatic urethral lift for the treatment of lower urinary tract symptoms in men with benign prostatic hyperplasia. *Cochrane Database Syst Rev*. 2019;5:CD012832.
20. Kang TW, Jung JH, Hwang EC, et al. Convective radiofrequency water vapour thermal therapy for lower urinary tract symptoms in men with benign prostatic hyperplasia. *Cochrane Database Syst Rev*. 2020;3:CD013251.
21. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
22. Pinsky PF, Prorok PC, Yu K, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. *Cancer*. 2017;123(4):592–599.
23. Hugosson J, Roobol MJ, Mansson M, et al. A 16-yr follow-up of the European randomized study of screening for prostate cancer. *Eur Urol*. 2019;76(1):43–51.
24. Osses DF, Remmers S, Schroder FH, et al. Results of prostate cancer screening in a unique cohort at 19 yr of follow-up. *Eur Urol*. 2019;75(3):374–377.
25. EAU Guidelines. *EDN Presented at the EAU Annual Congress Amsterdam*; 2020. http://uroweb.org/guidelines/compilations-of-all-guidelines/.
26. Bjurlin MA, Carroll PR, Eggener S, et al. Update of the standard operating procedure on the use of multiparametric magnetic resonance imaging for the diagnosis, staging and management of prostate cancer. *J Urol*. 2020;203(4):706–712.
27. Sathianathen NJ, Omer A, Harriss E, et al. Negative predictive value of multiparametric magnetic resonance imaging in the detection of clinically significant prostate cancer in the prostate imaging reporting and data system era: a systematic review and meta-analysis. *Eur Urol*. 2020;78(3):402–414.

28. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017;389(10071):815–822.
29. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med*. 2018;378(19):1767–1777.
30. Miah S, Hosking-Jervis F, Connor MJ, et al. A multicentre analysis of the detection of clinically significant prostate cancer following transperineal image-fusion targeted and nontargeted systematic prostate biopsy in men at risk. *Eur Urol Oncol*. 2020;3(3):262–269.
31. Hansen NL, Kesch C, Barrett T, et al. Multicentre evaluation of targeted and systematic biopsies using magnetic resonance and ultrasound image-fusion guided transperineal prostate biopsy in patients with a previous negative biopsy. *BJU Int*. 2017;120(5):631–638.
32. Hanna N, Wszolek MF, Mojtahed A, et al. Multiparametric magnetic resonance imaging-ultrasound fusion biopsy improves but does not replace standard template biopsy for the detection of prostate cancer. *J Urol*. 2019;202(5):944–951.
33. Hamid S, Donaldson IA, Hu Y, et al. The SmartTarget biopsy trial: a prospective, within-person randomised, blinded trial comparing the accuracy of visual-registration and magnetic resonance imaging/ultrasound image-fusion targeted biopsies for prostate cancer risk stratification. *Eur Urol*. 2019;75(5):733–740.
34. Yaxley AJ, Yaxley JW, Thangasamy IA, et al. Comparison between target magnetic resonance imaging (MRI) in-gantry and cognitively directed transperineal or transrectal-guided prostate biopsies for Prostate Imaging-Reporting and Data System (PI-RADS) 3-5 MRI lesions. *BJU Int*. 2017;120(suppl 3):43–50.
35. Lovegrove CE, Miah S, El-Shater Bosaily A, et al. Comparison of transrectal ultrasound biopsy to transperineal template mapping biopsies stratified by multiparametric magnetic resonance imaging score in the PROMIS trial. *J Urol*. 2020;203(1):100–107.
36. Scott S, Samaratunga H, Chabert C, et al. Is transperineal prostate biopsy more accurate than transrectal biopsy in determining final Gleason score and clinical risk category? A comparative analysis. *BJU Int*. 2015;116(suppl 3):26–30.
37. Borghesi M, Ahmed H, Nam R, et al. Complications after systematic, random, and image-guided prostate biopsy. *Eur Urol*. 2017;71(3):353–365.
38. Klotz L, Zhang L, Lam A, et al. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol*. 2010;28(1):126–131.
39. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med*. 2016;375(15):1415–1424.
40. Donovan JL, Hamdy FC, Lane JA, et al. Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med*. 2016;375(15):1425–1437.
41. Wilt TJ, Jones KM, Barry MJ, et al. Follow-up of prostatectomy versus observation for early prostate cancer. *N Engl J Med*. 2017;377(2):132–142.
42. Lardas M, Liew M, van den Bergh RC, et al. Quality of life outcomes after primary treatment for clinically localised prostate cancer: a systematic review. *Eur Urol*. 2017;72(6):869–885.
43. Matta R, Chapple CR, Fisch M, et al. Pelvic complications after prostate cancer radiation therapy and their management: an international collaborative narrative review. *Eur Urol*. 2019;75(3):464–476.
44. Simpkin AJ, Tilling K, Martin RM, et al. Systematic review and meta-analysis of factors determining change to radical treatment in active surveillance for localized prostate cancer. *Eur Urol*. 2015;67(6):993–1005.
45. Chu CE, Lonergan PE, Washington SL, et al. Multiparametric magnetic resonance imaging alone is insufficient to detect grade reclassification in active surveillance for prostate cancer. *Eur Urol*. 2020;78(4):515–517.
46. Stavrinides V, Giganti F, Trock B, et al. Five-year outcomes of magnetic resonance imaging-based active surveillance for prostate cancer: a large cohort study. *Eur Urol*. 2020;78(3):443–451.
47. Robinson D, Garmo H, Lissbrant IF, et al. Prostate cancer death after radiotherapy or radical prostatectomy: a nationwide population-based observational study. *Eur Urol*. 2018;73(4):502–511.
48. Wallis CJD, Saskin R, Choo R, et al. Surgery versus radiotherapy for clinically-localized prostate cancer: a systematic review and meta-analysis. *Eur Urol*. 2016;70(1):21–30.
49. Moris L, Cumberbatch MG, Van den Broeck T, et al. Benefits and risks of primary treatments for high-risk localized and locally advanced prostate cancer: an international multidisciplinary systematic review. *Eur Urol*. 2020;77(5):614–627.
50. Reichard CA, Hoffman KE, Tang C, et al. Radical prostatectomy or radiotherapy for high- and very high-risk prostate cancer: a multidisciplinary prostate cancer clinic experience of patients eligible for either treatment. *BJU Int*. 2019;124(5):811–819.
51. Rodda S, Tyldesley S, Morris WJ, et al. ASCENDE-RT: an analysis of treatment-related morbidity for a randomized trial comparing a low-dose-rate brachytherapy boost with a dose-escalated external beam boost for high- and intermediate-risk prostate cancer. *Int J Radiat Oncol Biol Phys*. 2017;98(2):286–295.
52. Berg S, Cole AP, Krimphove MJ, et al. Comparative effectiveness of radical prostatectomy versus external beam radiation therapy plus brachytherapy in patients with high-risk localized prostate cancer. *Eur Urol*. 2019;75(4):552–555.
53. Kishan AU, Chu FI, King CR, et al. Local failure and survival after definitive radiotherapy for aggressive prostate cancer: an individual patient-level meta-analysis of six randomized trials. *Eur Urol*. 2020;77(2):201–208.
54. Fossati N, Willemse PM, Van den Broeck T, et al. The benefits and harms of different extents of lymph node dissection during radical prostatectomy for prostate cancer: a systematic review. *Eur Urol*. 2017;72(1):84–109.
55. Marra G, Valerio M, Heidegger I, et al. Management of patients with node-positive prostate cancer at radical prostatectomy and pelvic lymph node dissection: a systematic review. *Eur Urol Oncol*. 2020;3(5):565–581.
56. Touijer KA, Karnes RJ, Passoni N, et al. Survival outcomes of men with lymph node-positive prostate cancer after radical prostatectomy: a comparative analysis of different postoperative management strategies. *Eur Urol*. 2018;73(6):890–896.
57. Nyberg M, Hugosson J, Wiklund P, et al. Functional and oncologic outcomes between open and robotic radical prostatectomy at 24-month follow-up in the Swedish LAPPRO trial. *Eur Urol Oncol*. 2018;1(5):353–360.
58. Yaxley JW, Coughlin GD, Chambers SK, et al. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. *Lancet*. 2016;388(10049):1057–1066.
59. Preisser F, Chun FKH, Pompe RS, et al. Persistent prostate-specific antigen after radical prostatectomy and its impact on oncologic outcomes. *Eur Urol*. 2019;76(1):106–114.
60. Spratt DE, Dess RT, Zumsteg ZS, et al. A systematic review and framework for the use of hormone therapy with salvage radiation therapy for recurrent prostate cancer. *Eur Urol*. 2018;73(2):156–165.
61. Yokomizo A, Wakabayashi M, Satoh T, et al. Salvage radiotherapy versus hormone therapy for prostate-specific antigen failure after radical prostatectomy: a randomised, multicentre, open-label, phase 3 trial (JCOG0401) (dagger). *Eur Urol*. 2020;77(6):689–698.
62. Fossati N, Robesti D, Karnes RJ, et al. Assessing the role and optimal duration of hormonal treatment in association with salvage radiation therapy after radical prostatectomy: results from a multi-institutional study. *Eur Urol*. 2019;76(4):443–449.
63. Perera M, Papa N, Roberts M, et al. Gallium-68 prostate-specific membrane antigen positron emission tomography in advanced prostate cancer-updated diagnostic utility, sensitivity, specificity, and distribution of prostate-specific membrane antigen-avid lesions: a systematic review and meta-analysis. *Eur Urol*. 2020;77(4):403–417.
64. Emmett L, van Leeuwen PJ, Nandurkar R, et al. Treatment outcomes from (68)Ga-PSMA PET/CT-informed salvage radiation treatment in men with rising PSA after radical prostatectomy: prognostic value of a negative PSMA PET. *J Nucl Med*. 2017;58(12):1972–1976.
65. Boreta L, Gadzinski AJ, Wu SY, et al. Location of recurrence by gallium-68 PSMA-11 PET scan in prostate cancer patients eligible for salvage radiotherapy. *Urology*. 2019;129:165–171.

66. Connor MJ, Smith A, Miah S, et al. Targeting oligometastasis with stereotactic ablative radiation therapy or surgery in metastatic hormone-sensitive prostate cancer: a systematic review of prospective clinical trials. *Eur Urol Oncol*. 2020;3(5):582–593.
67. De Bleser E, Jereczek-Fossa BA, Pasquier D, et al. Metastasis-directed therapy in treating nodal oligorecurrent prostate cancer: a multi-institutional analysis comparing the outcome and toxicity of stereotactic body radiotherapy and elective nodal radiotherapy. *Eur Urol*. 2019;76(6):732–739.
68. Ploussard G, Gandaglia G, Borgmann H, et al. Salvage lymph node dissection for nodal recurrent prostate cancer: a systematic review. *Eur Urol*. 2019;76(4):493–504.
69. van der Poel HG, van den Bergh RCN, Briers E, et al. Focal therapy in primary localised prostate cancer: the European Association of Urology Position in 2018. *Eur Urol*. 2018;74(1):84–91.
70. Nahar B, Bhat A, Reis IM, et al. Prospective evaluation of focal high intensity focused ultrasound for localized prostate cancer. *J Urol*. 2020;204(3):483–489.
71. Herrera-Caceres JO, Nason GJ, Salgado-Sanmamed N, et al. Salvage radical prostatectomy following focal therapy: functional and oncological outcomes. *BJU Int*. 2020;125(4):525–530.
72. Marconi L, Stonier T, Tourinho-Barbosa R, et al. Robot-assisted radical prostatectomy after focal therapy: oncological, functional outcomes and predictors of recurrence. *Eur Urol*. 2019;76(1):27–30.
73. Blazevski A, Scheltema MJ, Yuen B, et al. Oncological and quality-of-life outcomes following focal irreversible electroporation as primary treatment for localised prostate cancer: a biopsy-monitored prospective cohort. *Eur Urol Oncol*. 2020;3(3):283–290.
74. Huggins C, Hodges CV. Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *J Urol*. 2002;168(1):9–12.
75. Shore ND, Saad F, Cookson MS, et al. Oral relugolix for androgen-deprivation therapy in advanced prostate cancer. *N Engl J Med*. 2020;382(23):2187–2196.
76. Berthold DR, Pond GR, Soban F, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer: updated survival in the TAX 327 study. *J Clin Oncol*. 2008;26(2):242–245.
77. de Wit R, de Bono J, Sternberg CN, et al. Cabazitaxel versus abiraterone or enzalutamide in metastatic prostate cancer. *N Engl J Med*. 2019;381(26):2506–2518.
78. Rosmaninho A, Lobo I, Selores M. Sweet's syndrome associated with the intake of a selective cyclooxygenase-2 (COX-2) inhibitor. *Cutan Ocul Toxicol*. 2011;30(4):298–301.
79. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med*. 2012;367(13):1187–1197.
80. Sternberg CN, Fizazi K, Saad F, et al. Enzalutamide and survival in nonmetastatic, castration-resistant prostate cancer. *N Engl J Med*. 2020;382(23):2197–2206.
81. Chi KN, Agarwal N, Bjartell A, et al. Apalutamide for metastatic, castration-sensitive prostate cancer. *N Engl J Med*. 2019;381(1):13–24.
82. Fizazi K, Shore N, Tammela TL, et al. Nonmetastatic, castration-resistant prostate cancer and survival with darolutamide. *N Engl J Med*. 2020;383(11):1040–1049.
83. Kyriakopoulos CE, Chen YH, Carducci MA, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer: long-term survival analysis of the Randomized Phase III E3805 CHAARTED Trial. *J Clin Oncol*. 2018;36(11):1080–1087.
84. Davis ID, Martin AJ, Stockler MR, et al. Enzalutamide with standard first-line therapy in metastatic prostate cancer. *N Engl J Med*. 2019;381(2):121–131.
85. Fizazi K, Tran N, Fein L, et al. Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2019;20(5):686–700.
86. Clarke NW, Ali A, Ingleby FC, et al. Addition of docetaxel to hormonal therapy in low- and high-burden metastatic hormone sensitive prostate cancer: long-term survival results from the STAMPEDE trial. *Ann Oncol*. 2019;30(12):1992–2003.
87. Hoyle AP, Ali A, James ND, et al. Abiraterone in "high-" and "low-risk" metastatic hormone-sensitive prostate cancer. *Eur Urol*. 2019;76(6):719–728.
88. Fendler WP, Weber M, Iravani A, et al. Prostate-specific membrane antigen ligand positron emission tomography in men with nonmetastatic castration-resistant prostate cancer. *Clin Cancer Res*. 2019;25(24):7448–7454.
89. Hofman MS, Emmett L, Violet J, et al. TheraP: a randomized phase 2 trial of (177) Lu-PSMA-617 theranostic treatment vs cabazitaxel in progressive metastatic castration-resistant prostate cancer (clinical trial protocol ANZUP 1603). *BJU Int*. 2019;124(suppl 1):5–13.

117

The Menstrual Cycle and Disorders of Ovulation

Christopher R. McCartney

OUTLINE

The Menstrual Cycle: An Introduction, 1950
Overview of the Hypothalamic-Pituitary-Ovarian Axis, 1951
- The GnRH Pulse Generator, 1951
- Anterior Pituitary Gonadotropes and Regulation of Gonadotropin Secretion, 1954
- Ovarian Folliculogenesis and Steroidogenesis, 1956
- Mechanisms of Ovarian Feedback Control of GnRH and Gonadotropin Secretion, 1957
- Uterine Events During the Menstrual Cycle, 1959

Integrated Physiology of the Menstrual Cycle, 1960
- Luteal-Follicular Transition, 1960
- Follicular Phase, 1960
- Midcycle Gonadotropin Surge and Ovulation, 1962
- Luteal Phase, 1962

Neuroendocrine Mechanisms of Anovulation, 1963
- Physiological Anovulation, 1963
- Selected Causes of Pathological Anovulation, 1964

Conclusion, 1967

THE MENSTRUAL CYCLE: AN INTRODUCTION

The menstrual cycle represents repetitive (cyclic) activity of the hypothalamic-pituitary-ovarian axis that achieves episodic—approximately once-monthly—presentation of a mature oocyte for fertilization. Additionally, coordinated changes in ovarian hormone production prepare the endometrium for possible pregnancy. In the absence of pregnancy, the major clinical manifestation of cyclic hypothalamic-pituitary-ovarian activity is episodic menstrual (vaginal) bleeding related to endometrial shedding.

Menstrual cycles are often anovulatory, and intermenstrual length is often irregular, during the initial postmenarcheal years. Once regular cyclic ovulation is established, however, median (interquartile range) intermenstrual length approximates 28 (26–31), 27 (25–30), and 26 (24–28) days at ages 20, 30, and 40 years, respectively.[1] Intermenstrual length tends to be stereotyped—and thus rather predictable—for individual women with consistent ovulatory cycles. Anovulation becomes more common and intermenstrual length variability increases in the menopausal transition, and menopause represents the permanent cessation of menses related to depletion of ovarian follicles (see Chapter 122)

The menstrual cycle is governed by complex interactions among the hypothalamus, pituitary, and ovary. Simplistically, hypothalamic gonadotropin-releasing hormone (GnRH) stimulates luteinizing hormone (LH) and follicle-stimulating hormone (FSH) production from pituitary gonadotropes, and these gonadotropins direct the major functions of the ovaries: follicular maturation, ovulation, and sex steroid production. In turn, ovarian products—estradiol, progesterone, and inhibins in particular—modulate GnRH pulse generator activity and gonadotropin responses to GnRH. These bidirectional interactions between the hypothalamic-pituitary unit and the ovaries are highly coordinated and sequenced.

By convention, menstrual cycle day numbering begins with the first day of menses (i.e., cycle day 1 is the first day of menses), and the cycle is conceptually divided into two phases. The time period from the onset of menses to the onset of the midcycle gonadotropin surge is called the follicular phase, so named because it is characterized by the growth and maturation of ovarian follicles, a process that culminates in the emergence of a mature follicle prepared for ovulation. The early follicular phase coincides with endometrial shedding (menses), but follicular development is associated with a progressive increase in estradiol production, which thereafter stimulates endometrial proliferation. Follicular phase length is typically 14–21 days, but it can be of variable length, especially during the years following menarche and the years leading up to menopause.

The gonadotropin surge marks the follicular-to-luteal transition, which is also known as “midcycle.” The time period from the gonadotropin surge to the onset of menstrual bleeding is called the luteal phase, so named because the corpus luteum is a prominent ovarian structure during this phase. The ovarian luteal phase coincides with the uterine (endometrial) secretory phase: corpus luteum-derived progesterone directs endometrial changes that optimize endometrial receptivity in the case of fertilization. In the absence of fertilization, however, the corpus luteum regresses 9–11 days after ovulation, and the associated decrease in progesterone production results in endometrial shedding (menses). Typical luteal phase length is 10–16 days and tends to be stereotyped for individuals.

OVERVIEW OF THE HYPOTHALAMIC-PITUITARY-OVARIAN AXIS

KEY POINTS

- Pulsatile GnRH release from the hypothalamus is the final common pathway for the central neuronal control of reproduction.
- GnRH neuron function is regulated by afferent neuron populations including those that secrete kisspeptin, neurokinin B (NKB), and dynorphin; these and other afferents mediate the negative feedback effects of ovarian estradiol and progesterone.
- GnRH signal transduction within pituitary gonadotropes involves complex intracellular pathways that culminate in gonadotropin subunit gene expression, gonadotropin assembly, and gonadotropin release into the systemic circulation.
- Within the physiological range, higher and lower GnRH pulse frequencies favor LH and FSH secretion, respectively.
- At low physiological concentrations, estradiol reduces GnRH pulse amplitude and inhibits pituitary FSH release, while progesterone is the principal negative feedback regulator of GnRH pulse frequency.
- Ovarian-derived inhibins selectively restrain FSH secretion from gonadotropes, while pituitary activins and follistatin selectively augment and reduce, respectively, FSH secretion in an autocrine/paracrine manner.
- At high physiological concentrations, estradiol markedly augments gonadotrope responsiveness to GnRH—a unique long-loop positive feedback mechanism that underlies the midcycle gonadotropin surge.

The GnRH Pulse Generator

GnRH is a decapeptide hormone synthesized in some 7000 specialized hypothalamic neurons whose cell bodies primarily reside in the infundibular nucleus of the mediobasal hypothalamus—human homologue of the arcuate nucleus in lower mammalian species—and to a lesser degree in the preoptic area of the anterior hypothalamus. The principal form of GnRH (pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH2) is a product of the GnRH type I gene (*GNRH1*) located on chromosome 8.

Nascent GnRH neurons originate in the nasal (olfactory) placode and migrate to the hypothalamus during embryological development. This migration is guided by olfactory and vomeronasal neurons: GnRH cells adhere to such axons and are thus guided through the nasal epithelium, across the cribriform plate, and through the forebrain toward the hypothalamus. Upon reaching their hypothalamic destinations, GnRH neurons extend projections to the median eminence, where GnRH is released into the hypophyseal portal vessels for transport to the anterior pituitary (Fig. 117.1). These neuronal processes have recently been called "dendrons" because they exhibit characteristics of both axons and dendrites.

GnRH stimulates synthesis and secretion of both LH and FSH from gonadotrope cells in the anterior pituitary. Appropriate ovarian function—gametogenesis, ovulation, sex steroid production—is absolutely dependent on appropriate gonadotropin signaling, and appropriate gonadotropin secretion is absolutely dependent on pulsatile GnRH release.

Pulsatile GnRH Release. Although GnRH cell bodies are loosely associated anatomically, GnRH neurons are functionally integrated and secrete GnRH in a coordinated and pulsatile (episodic) fashion. GnRH is released at the median eminence into the hypophyseal portal system en route to the anterior pituitary gland. However, GnRH is rapidly degraded and has an approximately 2- to 4-minute half-life in hypophyseal portal blood. Episodic GnRH release into the portal system, along with the very short half-life of GnRH, renders discrete episodes of GnRH stimulation of gonadotropes.

Pulsatile GnRH release is essential for normal LH and FSH secretion. GnRH upregulates its own receptors (self-priming or autopriming) on pituitary gonadotropes, but this effect is most prominent when GnRH pulses occur every 60–90 minutes. Continuous GnRH receptor (*GnRHR*) stimulation leads to markedly reduced gonadotropin secretion (desensitization), as demonstrated in GnRH-deficient rhesus monkeys (Fig. 117.2).[2] Desensitization involves a reduction in *GnRHR* expression

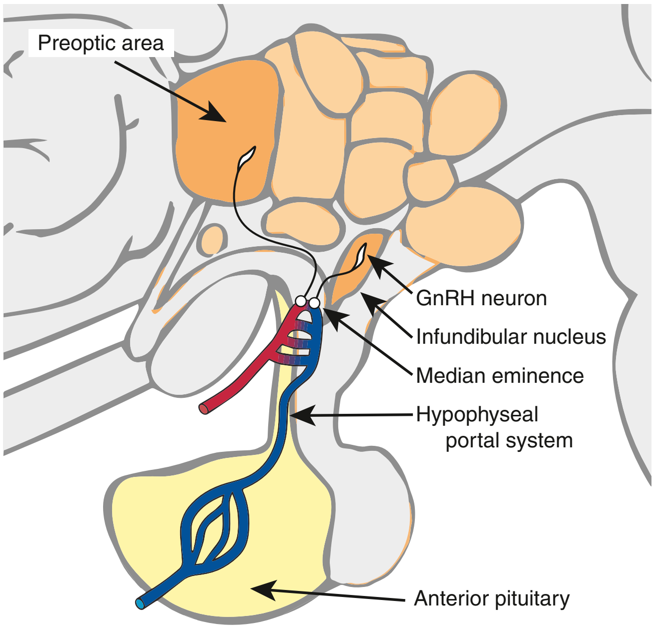


Figure 117.1 **Anatomy of the hypothalamic-pituitary unit (simple schematic).** GnRH neuron cell bodies primarily reside in the infundibular nucleus (mediobasal hypothalamus) and preoptic area. GnRH neuron projections (dendrons) extend to the median eminence, where pulses of GnRH are released into the hypophyseal portal system, through which GnRH is transported to gonadotropes in the anterior pituitary. *GnRH,* Gonadotropin-releasing hormone.

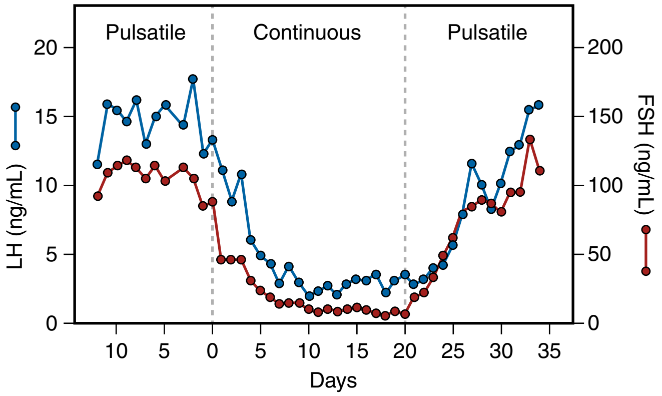


Figure 117.2 **Effects of pulsatile versus continuous GnRH stimulation on gonadotropin secretion.** These data were obtained in gonadectomized but sex steroid-replaced monkeys after arcuate nucleus ablation, which produced GnRH deficiency. Exogenous GnRH administered in hourly pulses (1 μg/minute for 6 minutes per hour) reconstituted gonadotropin secretion. However, continuous GnRH infusion (1 μg/minute) markedly reduced LH and FSH release. Gonadotropin secretion was again restored with hourly GnRH pulses. *FSH,* Follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Belchetz PE, Plant TM, Nakai Y, et al. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. *Science*. 1978;202:631–633.)

on gonadotropes in addition to the uncoupling of *GnRHR* agonism and gonadotropin synthesis. Notably, this phenomenon can be used to therapeutic advantage with the use of long-acting synthetic GnRH agonists—peptides similar to GnRH but with amino acid substitutions that enhance *GnRHR* binding affinity, reduce proteolytic degradation, or both.

Gonadotropin secretion is sensitive to GnRH pulse frequency and amplitude. In GnRH-deficient rhesus monkeys treated with once-hourly exogenous GnRH pulses, plasma LH and FSH concentrations decreased by approximately 50%–60% over 1–2 weeks when frequency was increased to 2 pulses per hour; and profound gonadotropin suppression occurred with GnRH pulse frequencies of 3–5 pulses per hour (Fig. 117.3).[3] In this same study, a 90% decrease in GnRH dose per pulse yielded unmeasurable gonadotropins within 2–7 days, whereas a 10-fold increase in GnRH dose per pulse resulted in a 34% decrease in FSH but no change in LH concentrations. These results highlight that gonadotropin secretion is dependent on both GnRH pulse frequency and GnRH pulse amplitude.

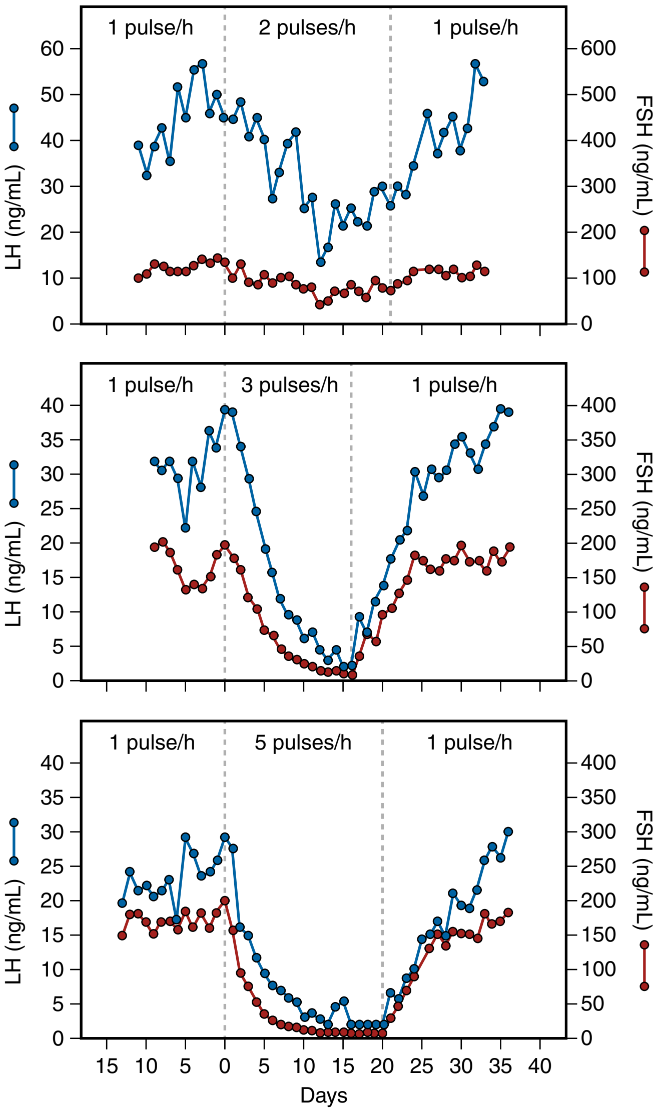


Figure 117.3 Effects of supraphysiological GnRH pulse frequencies on gonadotropin secretion. These data were obtained in gonadectomized but sex steroid-replaced monkeys after arcuate nucleus ablation, which produced GnRH deficiency. Exogenous GnRH administered in hourly pulses (1 μg/minute for 6 minutes per hour) reconstituted gonadotropin secretion. However, circulating LH and FSH release concentrations decreased when exogenous GnRH pulse frequency was increased to 2, 3, or 5 pulses per hour. *FSH,* Follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Wildt L, Hausler A, Marshall G, et al. Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey. *Endocrinology.* 1981;109:376–385.)

Cyclic Changes in GnRH Pulse Secretion. Animal studies confirm that each GnRH pulse elicits a secretory burst of LH (Fig. 117.4).[4] Thus, in human studies, patterns of pulsatile GnRH secretion may be inferred from LH pulse patterns. The circulating half-life of LH approximates 30–60 minutes, so LH pulses are readily discernable in the peripheral circulation in most physiological situations. GnRH pulses also elicit secretory bursts of free α-subunit, which has a shorter half-life than LH; thus, free α-subunit pulse analysis may be advantageous when GnRH pulse frequency is at its highest (e.g., after menopause). Although GnRH pulses elicit secretory bursts of FSH, such pulses are often obscured by the longer circulating half-life of FSH.

LH (GnRH) pulse frequency varies substantially across the menstrual cycle (Fig. 117.5).[5,6] Available data are consistent with the notion that the GnRH pulse generator has an intrinsic pulse frequency of approximately 1 pulse per hour and that cyclic changes in GnRH pulse frequency primarily reflect the imposition and removal of sex steroid negative feedback. For example, isolated human mediobasal hypothalami secrete GnRH pulses every 60–100 minutes,[7] and GnRH pulse frequency is around once hourly in the near absence of negative feedback (e.g., after menopause).[8,9] In contrast, GnRH pulse frequency is low (1 pulse every 3–5 hours) during the luteal phase, reflecting negative feedback restraint from corpus luteum-derived progesterone.

With the demise of the corpus luteum, falling progesterone concentrations allow 24-hour GnRH pulse frequency to increase to 1 pulse every 90–100 minutes in the early follicular phase. The 24-hour GnRH pulse frequency gradually increases to about 1 pulse every hour by the mid- to late follicular phase. This increase appears to reflect a gradual escape from (luteal) progesterone negative feedback.[10] It also reflects changes in sleep-related LH pulse frequency across the follicular phase. In particular, sleep-related decreases in LH pulse frequency are most prominent in the early follicular phase and less prominent in the later follicular phase.[11,12] A recent study suggested that the degree of sleep-associated frequency slowing in the early follicular phase correlates with the degree of previous progesterone exposure.[13]

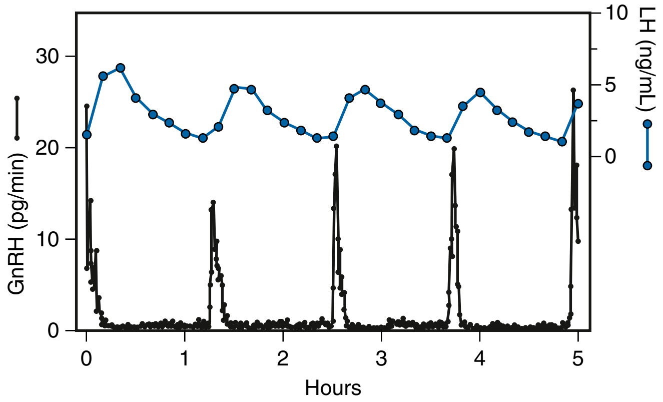


Figure 117.4 GnRH and LH concentration time series in the ewe, obtained via frequent blood sampling from the hypophyseal portal system and jugular vein, respectively. Note the close temporal relationship between GnRH pulses and subsequent LH pulses. *GnRH,* Gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Moenter SM, Brand RM, Midgley AR, et al. Dynamics of gonadotropin-releasing hormone release during a pulse. *Endocrinology.* 1992;130:503–510.)

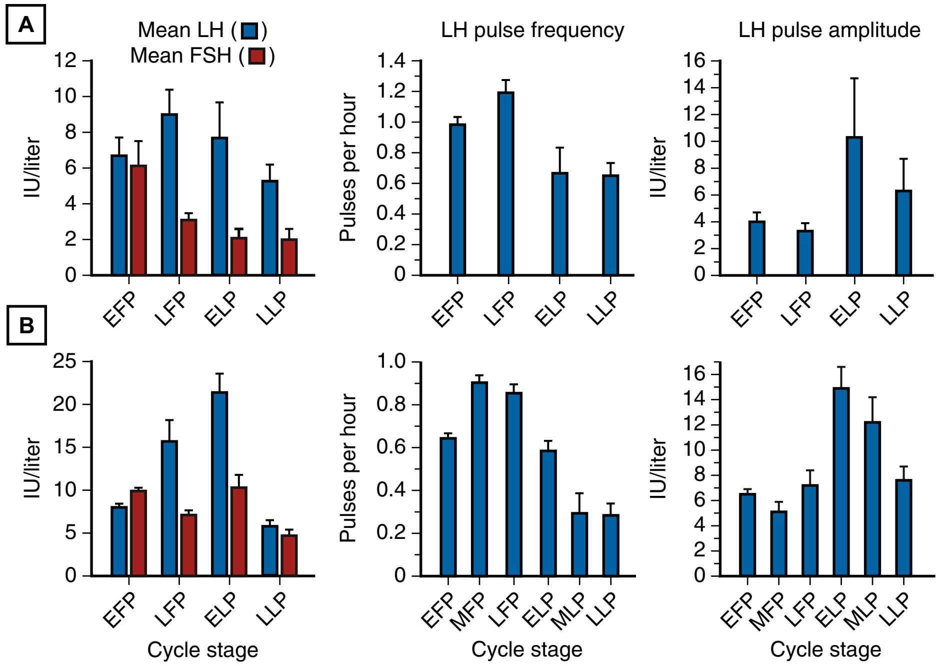


Figure 117.5 **Mean gonadotropin concentrations, LH pulse frequency, and LH pulse amplitude across the menstrual cycle, derived from two early studies.** Data are depicted as mean ± standard error. *EFP,* Early follicular phase; *ELP,* early luteal phase; *FSH,* follicle-stimulating hormone; *LFP,* late follicular phase; *LH,* luteinizing hormone; *LLP,* late luteal phase; *MFP,* midfollicular phase; *MLP,* midluteal phase. (Study A: Reame N, Sauder SE, Kelch RP, et al. Pulsatile gonadotropin secretion during the human menstrual cycle: evidence for altered frequency of gonadotropin-releasing hormone secretion. *J Clin Endocrinol Metab.* 1984;59:328–337. Study B: Filicori M, Santoro N, Merriam GR, et al. Characterization of the physiological pattern of episodic gonadotropin secretion throughout the human menstrual cycle. *J Clin Endocrinol Metab.* 1986;62:1136–1144.)

LH pulse amplitude also changes throughout the menstrual cycle (Fig. 117.5).[5,6] Importantly, LH pulse amplitude reflects both the amount of GnRH released per pulse and gonadotrope responsiveness to GnRH stimulation. Studies in GnRH-deficient subjects treated with exogenous pulsatile GnRH also disclose that GnRH-stimulated LH release varies inversely with the preceding GnRH inter-pulse interval, at least within physiological ranges.[14]

Pulsatile exogenous GnRH restores cycles and fertility in women with isolated GnRH deficiency, and such treatment may employ variable GnRH doses and frequencies to mimic normal physiology.[15] Maintenance of high-frequency exogenous GnRH pulses throughout the luteal phase may impair subsequent follicular development and corpus luteum function,[16,17] suggesting that cyclic changes in GnRH pulse frequency—high physiological pulse frequency in the follicular phase with slowing during the luteal phase—may be an important factor for long-term cyclic ovulation. However, in women with isolated GnRH deficiency, ovulation also occurs with exogenous GnRH administered at a fixed dose and frequency; in such cases, supraphysiological GnRH dosing might potentially compensate for the absence of periodic pulse frequency slowing.

Higher-Order Control of GnRH Secretion. GnRH neuron activity is characterized by episodic increases in action potential firing rates; these increases presumably provoke pulses of GnRH release. However, the fundamental nature of the GnRH pulse generator remains unclear. GnRH neurons intrinsically exhibit episodic bursts of action potential firing, and both immortalized GnRH-secreting neurons and cultured GnRH neurons secrete GnRH in a pulsatile fashion. Coordination of episodic action potential firing and episodic GnRH release could reflect extensive cell-to-cell interconnections. Nonetheless, currently available data suggest that neuronal populations afferent to GnRH neurons—including those that secrete kisspeptin, NKB, and dynorphin—represent critical components of the GnRH pulse generator.

Kisspeptin is the neuropeptide product of the *KISS1* gene and the natural ligand of the seven transmembrane domain, G protein-coupled kisspeptin receptor (KISS1R). Patients and animal models with inactivating *KISS1* or *KISS1R* mutations have pubertal failure and hypogonadotropic hypogonadism. Exogenous kisspeptin administration markedly stimulates LH release, and this is mediated through GnRH neurons. Kisspeptin fibers synapse with GnRH cell bodies and GnRH neuron terminals at the median eminence—both of which express KISS1R—and kisspeptin stimulates GnRH neuron activity and GnRH secretion. Moreover, kisspeptin-stimulated LH release is blocked by GnRH antagonists, and pulsatile GnRH restores reproductive function in patients with loss-of-function inactivating KISS1R mutations.

NKB is a neuropeptide encoded by the tachykinin 3 gene (*TAC3*). NKB exerts its actions primarily through its cognate receptor NK3R (encoded by the *TACR3* gene). Patients with homozygous inactivating mutations of *TAC3* or *TACR3* may demonstrate pubertal failure and hypogonadotropic hypogonadism. Senktide, a selective NK3R agonist, stimulates LH secretion in animals, and this action is blocked by *GnRHR* antagonism. Data suggest that the effects of NKB on GnRH secretion are primarily mediated through kisspeptin release: kisspeptin neurons express NK3R; senktide increases kisspeptin neuronal activity; LH responses to senktide are absent or markedly impaired in the absence of kisspeptin neuron activation; and kisspeptin infusion can restore pulsatile LH secretion in patients with loss-of-function *TAC3* or *TACR3* mutations.

Endogenous opioid peptides also direct GnRH neuronal function, and they play a major role in mediating progesterone negative feedback on GnRH release. For example, hypothalamic portal blood concentrations of β-endorphin are elevated during the luteal phase in monkeys, and opiate receptor antagonists increase LH pulse frequency in luteal phase women and progestin-treated postmenopausal women. Numerous animal studies indicate that dynorphin is a key mediator of progesterone negative feedback on GnRH pulse frequency, reducing GnRH secretion via reduced kisspeptin release.

Interestingly, kisspeptin, NKB, and dynorphin are frequently coexpressed in the same neuron—KNDy neurons (Kisspeptin, NKB, Dynorphin). KNDy neurons form an extensively interconnected network and interact with GnRH neurons in the arcuate (infundibular) nucleus in addition to GnRH fibers in the internal zone of the median eminence. Extensive animal data suggest that KNDy neurons mediate sex steroid negative feedback on GnRH secretion, and many consider the kisspeptin/KNDy neuronal network to be a fundamental component of the GnRH pulse generator (Fig. 117.6).

Many other neuronal populations (e.g., those secreting gamma aminobutyric acid [GABA], glutamate, neuropeptide Y [NPY], corticotropin-releasing hormone [CRH], dopamine) and endocrine signals (e.g., leptin, insulin) appear to influence GnRH secretion. A growing body of evidence also suggests that RFamide-related peptides (mammalian orthologues of gonadotropin-inhibitory hormone) are important regulators of GnRH and gonadotropin secretion.[18]

Anterior Pituitary Gonadotropes and Regulation of Gonadotropin Secretion

Gonadotropes constitute roughly 10% of the anterior pituitary cell population, and a majority of gonadotropes produce both LH and FSH. GnRH binding to the membrane-associated *GnRHR* on gonadotropes initiates a cascade of intracellular events leading to LH and FSH production and release (Fig. 117.7). In turn, LH and FSH act in concert to regulate ovarian function and thus are critical regulators of sexual maturation

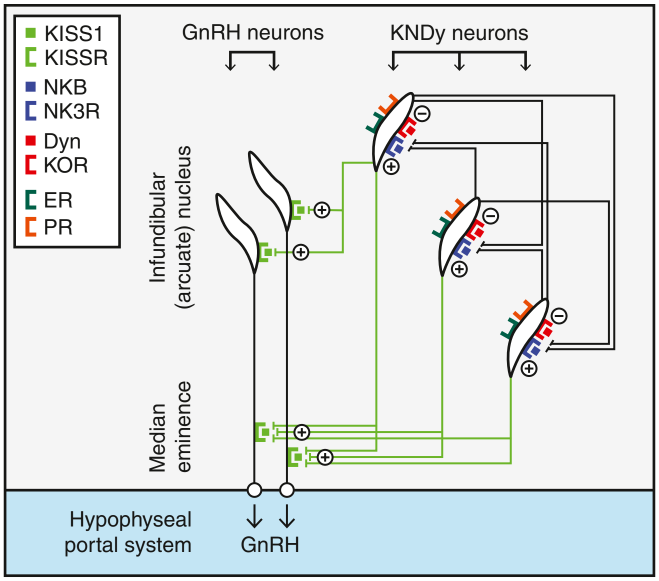


Figure 117.6 **Simplistic model of the infundibular (arcuate) KNDy neuronal network.** KNDy neurons form an interconnected network. NKB release activates KNDy neurons and stimulates kisspeptin release at synapses with GnRH neuron cell bodies and at GnRH dendrons at the median eminence. Conversely, Dyn release inhibits KNDy neurons and suppresses kisspeptin release at synapses with GnRH neuron cell bodies and GnRH dendrons at the median eminence. This schematic does not incorporate the possible roles of NKB and Dyn release directly onto GnRH neurons. *Dyn,* Dynorphin; *ER,* estradiol receptor; *GnRH,* gonadotropin-releasing hormone; *KISS1,* kisspeptin; *KISSR,* kisspeptin receptor; *KNDy,* kisspeptin, neurokinin B, dynorphin; *KOR,* kappa-opioid receptor; *NKB,* neurokinin B; *NK3R,* neurokinin 3 receptor; *PR,* progesterone receptor.

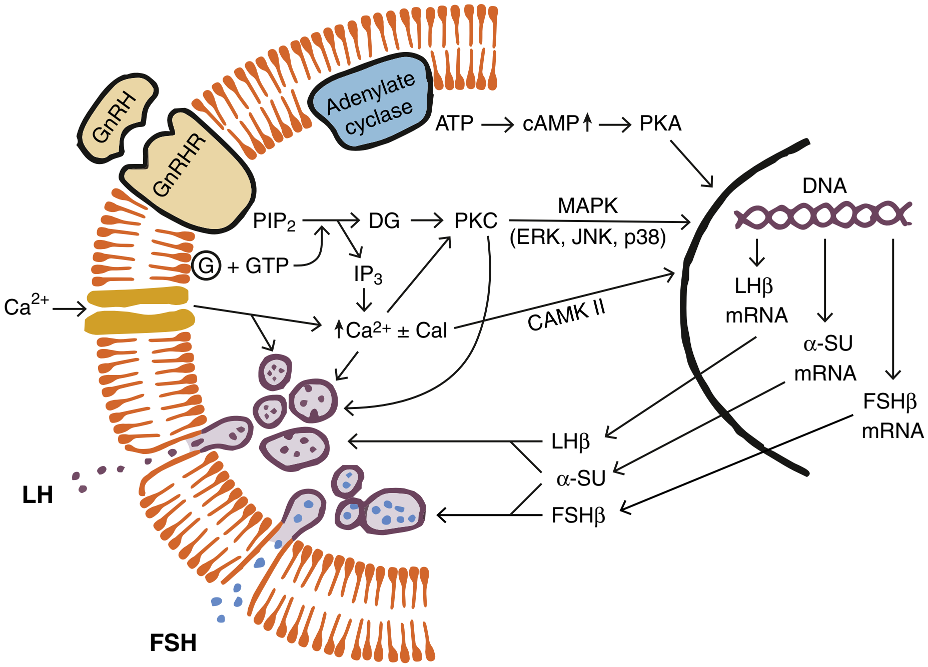


Figure 117.7 **Mechanisms involved in GnRH action on the gonadotrope.** *α-SU,* α-Subunit; *ATP,* adenosine triphosphate; *Ca^{2+},* calcium; *Cal,* calmodulin; *CAMK II,* calcium/calmodulin dependent kinase II; *cAMP,* cyclic adenosine monophosphate; *DG,* diacylglycerol; *ERK,* extracellular signal-regulated kinase; *FSH,* follicle-stimulating hormone; *G,* G protein; *GnRH,* gonadotropin-releasing hormone; *GnRHR,* GnRH receptor; *GTP,* guanosine triphosphate; *IP3,* inositol triphosphate; *JNK,* c-jun N-terminal kinase; *LH,* luteinizing hormone; *MAPK,* mitogen-activated protein kinase; *PIP2,* phosphatidylinositol 4,5-diphosphate; *PKA,* protein kinase A; *PKC,* protein kinase C.

and fertility. Primary roles of LH relate to theca cell steroidogenesis, ovulation, follicular luteinization, and corpus luteum steroidogenesis, while primary functions of FSH relate to follicular development, including dominant follicle recruitment and growth, and induction of granulosa cell aromatase activity (see Chapters 118 and 119).

Mechanisms of GnRH Signal Transduction. The type I *GnRHR* is encoded by the *GNRHR* gene on chromosome 4. Part of the rhodopsin-like G protein-coupled receptor superfamily, the *GnRHR* has a seven-transmembrane-domain structure with both extracellular and intracellular loops. Some of the highly complex intra-gonadotrope mechanisms involved in GnRH action are summarized in Fig. 117.7.[19–21] Briefly, GnRH binding to the *GnRHR* activates guanosine triphosphate (GTP)-associated proteins (e.g., $G\alpha_{q/11}$, $G\alpha_s$), with subsequent phospholipase Cβ hydrolysis of phosphatidylinositol 4,5-diphosphate (PIP_2) to inositol triphosphate (IP_3) and diacylglycerol. IP_3 provokes rapid intracellular calcium release from endoplasmic reticulum, which initiates the first 1–2 minutes of LH pulse release. This is followed by the opening of voltage-gated (L-type) calcium channels within the plasma membrane leading to a more sustained increase in intracellular calcium, supporting the remainder of LH pulse release.

Elevated intracellular calcium and diacylglycerol activate various protein kinase C (PKC) isoforms, which in turn activate a number of mitogen-activated protein kinases (MAPKs) such as extracellular signal-regulated kinase (ERK), c-jun N-terminal kinase (JNK), and p38-α. These MAPKs are involved in stimulating gonadotropin subunit transcription. Activated PKC also contributes to the sustained phase of gonadotropin release. Interestingly, GnRH-related calcium oscillations within the gonadotrope cells also increase gonadotropin subunit mRNA expression, which is at least partly mediated by calcium-/calmodulin-dependent kinase II (CAMK II) and calcineurin. GnRH action also appears to be partly mediated through a pathway involving cyclic adenosine monophosphate (cAMP), protein kinase A (PKA), and cAMP response element binding proteins.

LH is stored in granules, and LH release is closely tied to GnRH pulse stimulation. Although GnRH pulses acutely provoke FSH release as well, FSH exhibits some degree of constitutive secretion. Also of interest, GnRH stimulation increases gonadotrope responsiveness to GnRH stimulation (self-priming or autopriming), which appears to reflect changes in *GnRHR* density on the gonadotrope cell membrane. In rodents, maximum *GnRHR* density is achieved with the high GnRH pulse frequency of 1 pulse every 30 minutes; and *GnRHR* density correlates positively with GnRH secretion, being highest during gonadotropin surges or after castration.

Gonadotropin Structure. Each gonadotropin is a heterodimeric glycoprotein consisting of two different protein subunits, α and β. The 92-amino acid α-subunit, encoded by a single gene (*CGA*) on chromosome 6, is common to LH, FSH, chorionic gonadotropin (CG), and thyroid-stimulating hormone. The β-subunit for LH (LHβ) is a 121-amino acid protein derived from the *LHB* gene on chromosome 19, whereas the β-subunit for FSH (FSHβ) is a 111-amino acid protein derived from the *FSHB* gene on chromosome 11. The unique β-subunits account for the biological specificity of each gonadotropin.

GnRH stimulates LHβ, FSHβ, and α-subunit gene transcription and translation. Disulfide bonds between cysteine residues play important roles in the tertiary structures of each subunit. The α-subunit dimerizes non-covalently with an LHβ or FSHβ subunit to form LH or FSH, respectively. In contrast to the LHβ or FSHβ subunits, α-subunit is synthesized in excess and is independently secreted into the circulation. However, the α- and β-subunits exhibit no known independent biological actions.

Subunits also undergo variable post-translational glycosylation, namely the addition of N-linked oligosaccharide chains to specific asparagine residues. This occurs at positions 52 and 78 of α-subunit, positions 7 and 24 of FSHβ, and position 30 of LHβ. The quantity and types of oligosaccharide additions appear to be involved with gonadotropin assembly, and they influence both gonadotropin bioactivity and circulating half-life. N-linked oligosaccharide chains on FSH and CG tend to end with sialic acid moieties, while N-linked carbohydrate chains on LH tend to end with sulfate moieties. Higher numbers of sialic acid moieties are associated with slower metabolic clearance, while sulfonated N-acetylgalactosamine targets LH for hepatic removal. These factors contribute to the different circulating half-lives for LH and FSH in cycling women: approximately 30–60 minutes for LH and 120–240 minutes for FSH. Interestingly, the ratio of sialic acid to sulfonated N-acetylgalactosamine can vary across the cycle. For example, at midcycle, LH sulfonation is decreased and sialylation slightly increased, which is expected to increase LH half-life, contributing to the preovulatory LH surge.[22,23]

CG is produced by syncytiotrophoblast cells of the developing placenta. The CGβ subunit is a 145-amino acid protein that is 82% homologous with LHβ and includes a unique final C-terminal peptide (amino acid positions 121–145). Both LH and CG activate the LH receptor (LHCG receptor), and CG is critical for the maintenance of corpus luteum function and, thus, maintenance of favorable endometrial conditions during early pregnancy. CG exhibits glycosylation on two asparagine residues (positions 13 and 30) in addition to four mucin-type O-linked glycans at four serine residues (positions 121, 127, 132, and 138). Although CG shares biological activity with LH, higher sialic acid content of CGβ contributes to a much longer circulating half-life.

Feedforward Control of LH and FSH Secretion. A single hormone (GnRH) stimulates the synthesis and secretion of two hormones (LH and FSH) from a single cell type (gonadotrope). Yet circulating LH and FSH concentrations vary differentially across the cycle. For example, the luteal-follicular transition is characterized by a 3-fold increase in FSH, but less than a 2-fold increase in LH.[24] In contrast, FSH decreases while LH tends to increase in the latter half of the follicular phase.

Differential regulation of LH and FSH secretion is partly dictated by GnRH pulse frequency. Specifically, higher physiological GnRH pulse frequencies favor LH synthesis and secretion, whereas low physiological GnRH pulse frequencies favor FSH synthesis and secretion. Studies in GnRH-deficient monkeys disclosed that decreasing the frequency of exogenously administered GnRH pulses from 1 pulse every hour to 1 pulse every 3 hours led to a 50% decrease in plasma LH concentration but a 65% *increase* in plasma FSH concentration (Fig. 117.8).[3] In addition, FSH responses to each GnRH pulse increased 4-fold at this lower frequency, while LH responses increased by approximately 67%.[3] Similar findings have been described in humans. In one study of men with GnRH deficiency, pulsatile GnRH administration restored normal gonadotropin levels, yet FSH progressively increased when pulse frequency was reduced from 1 pulse every hour to 1 pulse every 2 hours to 1 pulse every 3 hours, whereas LH concentrations remained similar across these frequencies.[25] In other studies of men and women with GnRH deficiency, increasing GnRH pulse frequency was associated with increased LH concentrations, but no clear changes in FSH.[26,27]

Animal studies demonstrate that different GnRH pulse frequencies differentially impact LHβ and FSHβ subunit production.[19–21] In rodents, high-frequency GnRH pulse stimulation favors α-subunit and LHβ mRNA expression, whereas low-frequency stimulation favors FSHβ mRNA expression. Different signaling pathways appear to be differentially regulated by high and low GnRH pulse frequencies.

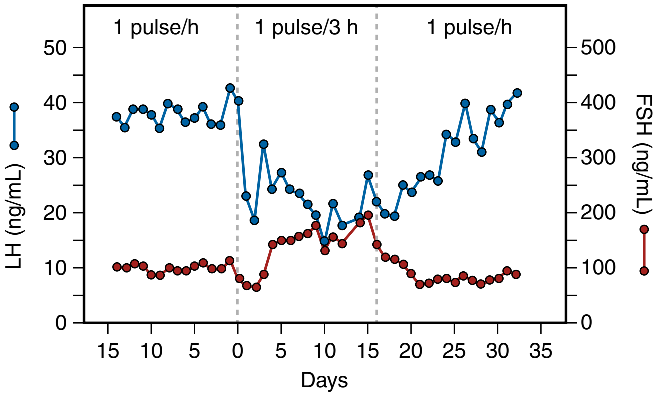


Figure 117.8 Differential effects of GnRH pulse frequency on LH and FSH secretion. These data were obtained in gonadectomized but sex steroid-replaced monkeys after arcuate nucleus ablation, which produced GnRH deficiency. Exogenous GnRH administered in hourly pulses (1 µg/minute for 6 minutes per hour) reconstituted gonadotropin secretion. Changing the frequency of exogenous GnRH pulse administration from a high physiological frequency (hourly) to a relatively low frequency (every 3 hours) resulted in decreased LH but increased FSH concentrations. *FSH,* Follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Wildt L, Hausler A, Marshall G, et al. Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey. *Endocrinology.* 1981;109:376–385.)

For example, high GnRH pulse frequency increases *GnRHR* density at the cell surface, and higher *GnRHR* density at the gonadotrope cell surface favors LHβ gene expression, whereas lower *GnRHR* density favors FSHβ gene expression. Additional post-receptor mechanisms contribute to this phenomenon as well, including differential GnRH pulse frequency-specific effects on MAPK pathways, the cAMP/PKA pathway, etc. Indeed, different frequencies of experimentally induced intracellular calcium oscillations differentially impact LHβ and FSHβ mRNA expression. In addition, higher GnRH pulse frequencies enhance gonadotrope follistatin production, which leads to a selective reduction in FSHβ. Although different GnRH pulse amplitudes differentially impact LHβ and FSHβ translation, the impact of this phenomenon is modest compared with frequency effects.

Of critical importance, LH and FSH secretion is differentially influenced by ovarian hormone feedback. For example, high physiological concentrations of estradiol markedly increase gonadotropin responses to GnRH, and this effect is more prominent for LH than FSH. In addition, ovarian inhibins and low physiological estradiol concentrations selectively restrain FSH secretion in the later follicular and luteal phases. These feedback effects on gonadotropin secretion are discussed further as follows.

Ovarian Folliculogenesis and Steroidogenesis

The ovarian follicle—an oocyte surrounded by granulosa and theca cells—is the functional unit of the ovary, and circulating gonadotropins direct ovarian function. LH and FSH receptors are seven-transmembrane domain, G protein-coupled receptors with large extracellular domains. FSH receptors localize to granulosa cells: FSH regulates follicular growth and maturation, and it stimulates estradiol production from granulosa cells by enhancing aromatase expression. LH receptors are expressed by theca cells; granulosa cells in large, preovulatory follicles; and steroidogenic cells of the corpus luteum. LH stimulates sex steroid (e.g., androstenedione) production from theca cells during the follicular phase; provokes ovulation at midcycle; and controls sex steroid (e.g., progesterone) production from the corpus luteum during the luteal phase.

Folliculogenesis, ovulation, and luteogenesis are addressed in Chapter 118, and ovarian hormone synthesis is addressed in Chapter 119. However, the following brief overviews are provided for context.

Ovarian Follicular Development. Oogonia begin, but do not complete, the first meiotic division between 8 and 13 weeks of gestation, thus becoming primary oocytes. At approximately 20 weeks of gestation, primary oocytes will be surrounded by pre-granulosa cells to form primordial follicles. The number of oogonia reaches a peak of approximately 6–7 million at 20 weeks of gestation, but ongoing atresia will reduce the number of primordial follicles to roughly 700,000 at birth and 300,000 at puberty.

Primordial follicles are dormant, but can be engaged for further development—a process called follicular activation or initial follicular recruitment. This process, which occurs continuously from the fifth or sixth gestational month until menopause, is primarily governed by local growth factors through complex autocrine and paracrine interactions among the oocyte, follicular cells, and surrounding cells. The endgame of follicular development is ovulation, and the progression from recruited primordial follicle to ovulatory follicle usually takes about a year. However, since only 400–600 follicles will proceed to ovulation in a woman's lifetime, the vast majority of follicles will undergo atresia instead.

When a follicle is initially recruited for further development, the flat pre-follicular cells begin to proliferate and become cuboidal granulosa cells; the oocyte increases in size; and a zona pellucida surrounds the egg. This structure is now called a primary (unilaminar) follicle. A developing follicle becomes a secondary (multilaminar) follicle when granulosa cells replicate, accumulate, and form layers around the oocyte. Around this time, granulosa cells begin to express FSH and estrogen receptors; and some of the ovarian stromal cells surrounding the granulosa cells become theca cells, which develop LH receptors and begin to express the steroidogenic machinery to produce androgens. The transition from primordial to full-grown secondary follicle occurs over about 10 menstrual cycles. While gonadotropins are not absolutely required for preantral follicular development, FSH appears to play a facilitatory role in this regard.

Secondary follicles can respond to the increase in FSH during the luteal-follicular transition, thus beginning the "gonadotropin-dependent phase" of follicular development, although local growth factors still participate in this process. FSH stimulation promotes granulosa cell division, which increases inhibin B production. Granulosa cell FSH receptor and aromatase expression increases, and estrogen also stimulates granulosa cells to divide and to express more FSH receptors. With continued stimulation, fluid appears between granulosa cells and coalesces to form a fluid-filled cavity—the antrum. As follicle size increases further, there is a progressive, gonadotropin-stimulated increase of androgen production (especially androstenedione) from theca cells, with subsequent aromatization to estradiol in granulosa cells.

During the luteal-follicular transition, the increase in FSH secretion elicits growth of a cohort of antral follicles measuring approximately 4–5 mm in diameter—a process that is sometimes called cyclic recruitment—in addition to lead follicle selection and growth. Notably, the lead follicle destined to ovulate can be distinguished in the early follicular phase (e.g., its granulosa cells have higher mitotic rates and produce more estradiol). Through a marked increase in granulosa cell mitosis and antral fluid accumulation, the diameter of the lead follicle reaches approximately 10 mm during the midfollicular phase and thereafter grows approximately 2 mm per day until ovulation. In contrast, granulosa cell proliferation tapers off in the non-selected follicles. Mechanisms underlying monofollicular dominance remain unclear, but this phenomenon partly reflects the ability of estradiol and inhibin

B to restrain FSH secretion, thus reducing a stimulatory growth factor for non-selected follicles. Indeed, exogenous FSH administration to maintain an FSH elevation for 5–6 additional days encourages multifollicular development,[28] while suppression of FSH with exogenous estradiol delays dominant follicle emergence.[29]

Five to six days prior to ovulation, the dominant follicle grows rapidly and moves toward the ovarian surface. Toward the end of this maturation process, estradiol and FSH stimulate the appearance of LH receptors on the granulosa cell layers closest to the basal lamina. LH will act on granulosa cells to begin luteinization and stimulate production of progesterone and prostaglandins, which appear to be important in the ovulatory process. High estradiol concentrations derived from the mature, preovulatory follicle will then elicit a preovulatory gonadotropin surge.

This midcycle gonadotropin surge provokes ovulation—expulsion of the cumulus-oocyte complex—and initiates formation of the corpus luteum. Notably, the mature preovulatory (Graafian) follicle typically measures 20–25 mm in diameter at ovulation, which occurs approximately 24–36 hours after the start of the midcycle gonadotropin surge. Follicle rupture and oocyte release involve factors such as plasminogen activator, inflammatory cells (leukocytes, macrophages), prostaglandins, chemokines, and cytokines. Oocyte meiosis I is resumed in response to the preovulatory gonadotropin surge, and the second meiotic division is arrested at metaphase until fertilization.

After ovulation, the remnant of the mature follicle becomes the corpus luteum. Granulosa cells are transformed into granulosa-lutein cells, which acquire the capacity to synthesize progesterone from cholesterol, but also continue to express aromatase. High luteal progesterone concentrations are important for preparing the endometrium for embryonic implantation and suppressing the GnRH pulse generator. Theca cells from the ovulated follicle are also integrated into the corpus luteum; these theca-lutein cells continue to express the steroidogenic enzymes necessary to produce androgen precursors that are aromatized to estrogens by granulosa-lutein cells. Importantly, LH is required for normal luteal function in primates: corpus luteum life span is shortened in the absence of LH, and corpus luteum life span can be prolonged by exogenous LH or placental chorionic gonadotropin (CG).

In a fertile cycle, trophoblast-derived CG prevents luteolysis and maintains progesterone production, which is critical for the maintenance of pregnancy—especially in the first 7–10 weeks of gestation. In the absence of fertilization, the functional lifespan of the corpus luteum approximates 14 (12–16) days, after which it will undergo luteolysis—a complex and incompletely understood process of regression. Reduced progesterone production during luteolysis partly reflects disruptions in post-receptor LH signaling. In addition, luteal cell apoptosis and autophagy, along with reduced perfusion, lead to tissue involution and formation of a fibrotic corpus albicans. Mechanisms underlying the preprogrammed lifespan of the corpus luteum remain uncertain.

Ovarian Steroidogenesis. The primary theca cell product is the weak androgen androstenedione. The initial step of theca cell steroidogenesis is the shuttling of cholesterol—either imported from circulating low-density lipoprotein-cholesterol particles or synthesized de novo—from outer to inner mitochondrial membranes by steroidogenic acute regulatory protein (StAR). Cholesterol is then converted to pregnenolone by the mitochondrial side-chain cleavage enzyme complex (P450scc). Notably, StAR expression in secondary and antral follicles is largely specific to theca cells, and StAR and P450scc activities are stimulated by LH. Within the smooth endoplasmic reticulum, the P450c17 enzyme converts pregnenolone to 17-hydroxypregnenolone via its 17α-hydroxylase activity, then to dehydroepiandrosterone via its 17,20-lyase activity. (Synthesis of androstenedione appears to proceed primarily along this Δ5 pathway; it is unclear to what degree procession along the Δ4 pathway—conversion of progesterone to 17-hydroxyprogesterone to androstenedione—is operative *in vivo*.) The enzyme 3β-hydroxysteroid dehydrogenase then converts dehydroepiandrosterone to the weak androgen androstenedione, which may be further transformed to the more potent testosterone by 17β-hydroxysteroid dehydrogenase.

The primary secretory products of the different sex steroid-producing cells depend on the particular complement of steroidogenic enzymes expressed in the cells. Theca cells produce androstenedione in response to LH stimulation, but theca cells do not express the aromatase enzyme. While granulosa cells are incapable of making androstenedione from cholesterol, they express aromatase in response to FSH stimulation. Thus, some theca cell-derived androstenedione diffuses into granulosa cells where it is aromatized into estrone, a weak estrogen. Estrone is then transformed to the potent estradiol by 17β-hydroxysteroid dehydrogenase. Therefore, ovarian estradiol synthesis depends on both LH stimulation of theca cells and FSH stimulation of granulosa cells—the "two-cell, two-gonadotropin concept" of estradiol synthesis.[30]

Granulosa-lutein cells of the corpus luteum gain the capacity to synthesize progesterone from cholesterol, and they also continue to express aromatase. Theca-lutein cells incorporated into the corpus luteum continue to express the steroidogenic enzymes necessary to produce androgen precursors, which are aromatized to estrogens by granulosa-lutein cells.

Mechanisms of Ovarian Feedback Control of GnRH and Gonadotropin Secretion

Estrogen Negative Feedback. Estrogens play a critical role in the feedback control of gonadotropin secretion. At low physiological concentrations, estrogens inhibit gonadotropin secretion. This phenomenon largely accounts for the observation that women with natural or surgical menopause, in addition to patients with aromatase deficiency, exhibit marked elevations in LH and FSH.

Estradiol negative feedback primarily reflects hypothalamic actions, although direct effects at the gonadotrope may also pertain. Studies in sheep, monkeys, and postmenopausal women suggest that estrogens selectively restrain GnRH pulse amplitude (Fig. 117.9). For example, in postmenopausal women, estradiol replacement enhances the ability of incomplete *GnRHR* blockade to reduce LH concentrations, implying an estradiol-mediated reduction in GnRH secretion.[31] But while estradiol replacement reduces free α-subunit pulse amplitude, it does not reduce free α-subunit pulse frequency in postmenopausal women.[9] Moreover, functional neuroimaging in postmenopausal women suggests that low-dose estradiol replacement reduces metabolic activity at the hypothalamus, but not at the pituitary.[32] Some animal studies also suggest that estrogens decrease pituitary LH responses to GnRH, at least transiently, but data in women are less clear.[33,34]

Estradiol also inhibits pituitary FSH secretion from gonadotropes. Maintenance of midluteal plasma estradiol concentrations across the luteal-follicular transition delays the increase in circulating FSH.[35,36] In women with GnRH deficiency, exogenous estradiol reduces FSH responses to constant-dose GnRH pulse administration.[33] These studies suggest a direct pituitary effect to reduce FSH release. While estrogen-receptor blockade increases FSH in the luteal and early follicular phases in normally cycling women, it does not increase early follicular phase FSH in GnRH-deficient women receiving fixed-dose GnRH pulses, suggesting that estradiol may also restrain FSH release via a hypothalamic mechanism.[37]

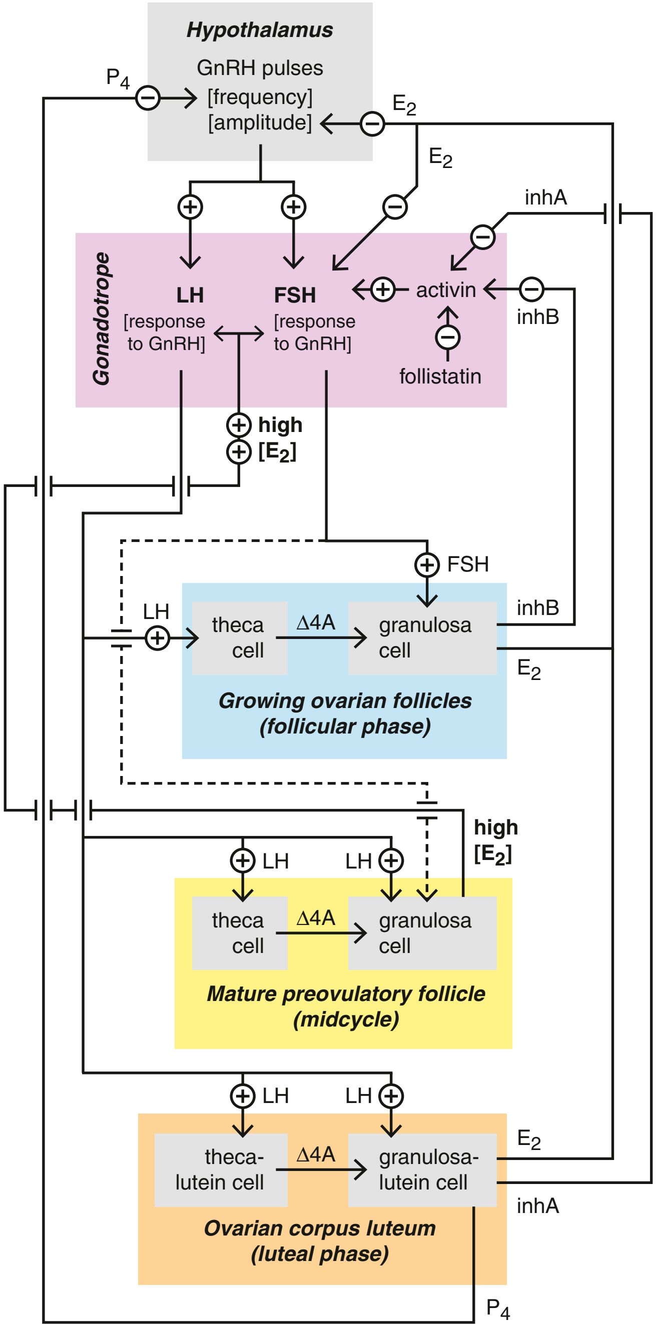


Figure 117.9 Schematic representation of feedforward and feedback relationships governing hypothalamic-pituitary-ovarian function during the menstrual cycle. *Δ4A,* Androstenedione; *E2,* estradiol; *FSH,* follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *inhA,* inhibin A; *inhB,* inhibin B; *LH,* luteinizing hormone; *P4,* progesterone.

Progesterone Negative Feedback. Progesterone appears to be the principal negative feedback regulator of day-to-day GnRH pulse frequency in women. Increased progesterone production in the early luteal phase is associated with a reduction in LH pulse frequency, and decreased progesterone production at the luteal-follicular transition is associated with an increase in LH pulse frequency. GnRH pulse frequency remains low if luteal progesterone and estradiol concentrations are maintained across the luteal-follicular transition,[38] and progesterone administration reduces LH pulse frequency during the follicular phase.[39] In postmenopausal women, progesterone plus low-dose estradiol reduces LH pulse frequency, but estradiol alone does not.[9]

The ability of progesterone to restrain GnRH pulse frequency requires the permissive presence of estradiol, most likely reflecting estrogen-related increases in hypothalamic progesterone receptor expression. In contrast, androgens antagonize the ability of progesterone to suppress GnRH pulse frequency, likely relating to androgen-induced reductions in hypothalamic progesterone receptor expression.

Inhibins, Activins, and Follistatin. FSH secretion, but not LH secretion, is substantially influenced by activins, follistatin, and inhibins. Activins are homodimers of two β-subunits, either βA or βB. Activin B is produced by pituitary cells, and activin receptors are present on gonadotropes. Activin increases FSHβ mRNA expression and stabilizes FSHβ mRNA, thus increasing FSH production. Follistatin, which is produced by gonadotropes and pituitary folliculostellate cells, binds to extracellular activin and reduces its action at the activin receptor; accordingly, follistatin reduces FSH production. Available data suggest that activins and follistatin primarily act in an autocrine/paracrine fashion within the pituitary.

Inhibin is a heterodimer peptide with two isoforms, inhibin A and inhibin B, which contain a common α-subunit disulfide-linked to one of two β-subunits—βA for inhibin A and βB for inhibin B. (Note that these β-subunits are identical to those in activins.) Inhibin B is mainly secreted by ovarian granulosa cells in small antral follicles during the early- and midfollicular phases. Circulating inhibin B concentrations increase in the early follicular phase in response to FSH-driven increases in granulosa cell numbers, but decrease in the latter follicular phase and are low during the luteal phase (Fig. 117.10).[40] Inhibin A is primarily produced by granulosa cells of the dominant follicle and granulosa-lutein cells of the corpus luteum; it therefore increases in the latter follicular phase and peaks in the luteal phase. Both inhibins bind to and interfere with the function of the activin receptor complex, thereby reducing FSH release. Thus, inhibins A and B act in an endocrine fashion to selectively restrain FSH secretion during the mid- to late follicular and luteal phases.

Estrogen and Progesterone Positive Feedback. The influence of estradiol on LH and FSH release is concentration- and time-dependent. High-dose exogenous estradiol has biphasic effects on circulating LH concentrations, with initial suppression over 2–3 days followed by marked augmentation (Fig. 117.11).[41–43] Similarly, high-dose exogenous estradiol initially reduces but later increases LH responses to exogenous GnRH. This latter phenomenon largely accounts for the midcycle (preovulatory) gonadotropin surge.

This positive feedback phenomenon—the only known example of long-loop positive feedback in endocrinology—primarily reflects the effects of high estradiol concentrations, as high-dose exogenous estradiol triggers a gonadotropin surge in the early- and midfollicular phases and in postmenopausal women.[41–43] Positive feedback is at least partly mediated at the pituitary gonadotrope, as gonadotropin responses to exogenous GnRH are increased at midcycle. Numerous mechanisms are involved, including increased *GnRHR* expression, increased gonadotropin subunit expression, and an increase in the number of LH-secreting cells.

In rodents, sheep, and monkeys, the gonadotropin surge is associated with increased GnRH pulse frequency and amplitude—a "GnRH surge." In women, the midcycle gonadotropin surge can be blocked by *GnRHR* antagonists, indicating its GnRH dependence. However, while LH and free α-subunit pulse amplitudes increase at midcycle in women, LH and free α-subunit pulse frequency does not.[44] In addition, a study employing incomplete *GnRHR* antagonism suggested that

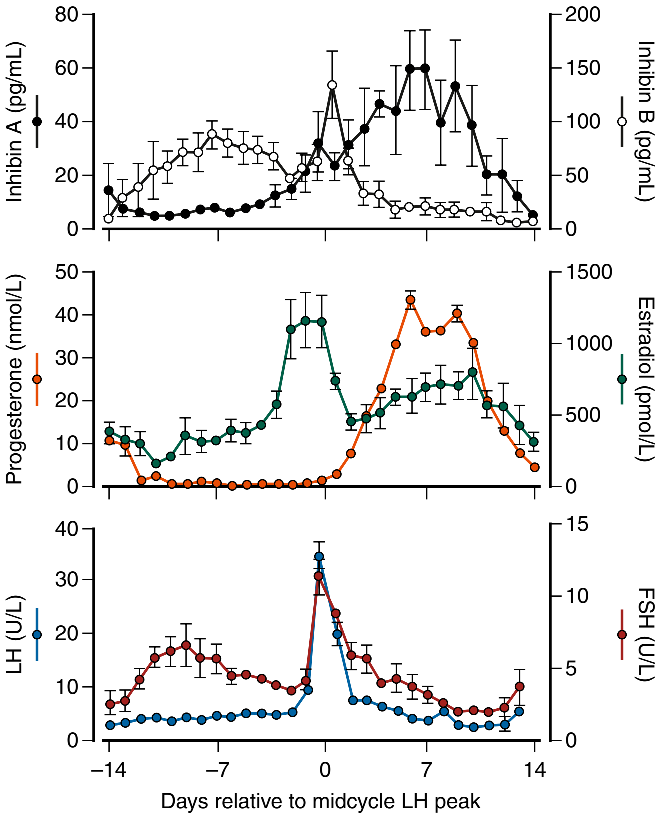


Figure 117.10 **Circulating concentrations of inhibins A and B, progesterone, estradiol, LH, and FSH across the ovulatory cycle.** Data are aligned to the day of the midcycle LH peak (day 0) and represented as mean ± standard error. Conversion of International System (SI) units to conventional units: progesterone (nmol/L) divided by 3.18 (ng/mL); estradiol (pmol/L) divided by 3.67 (pg/mL). *FSH,* Follicle-stimulating hormone; *LH,* luteinizing hormone. (From Groome NP, Illingworth PJ, O'Brien M, et al. Measurement of dimeric inhibin B throughout the human menstrual cycle. *J Clin Endocrinol Metab.* 1996;81:1401–1405.)

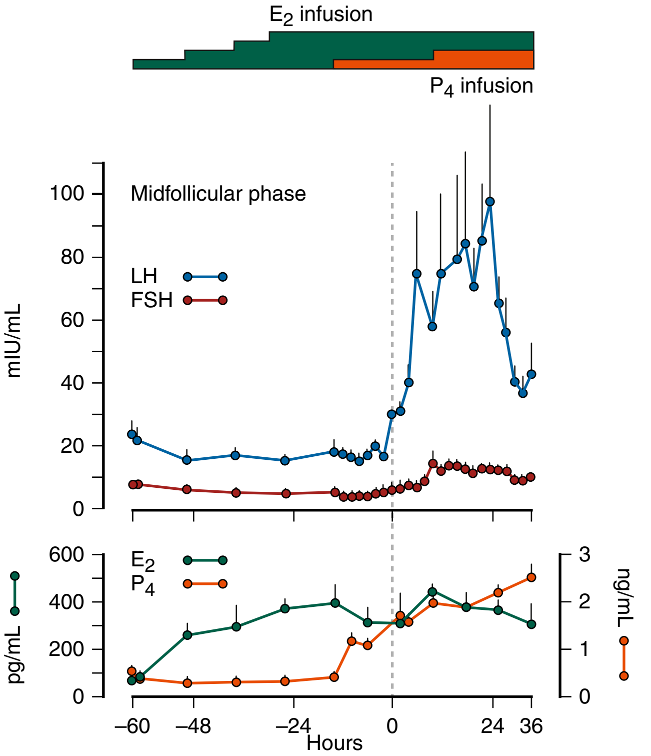


Figure 117.11 **Circulating luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), and progesterone (P4) concentrations in response to graded doses of estradiol and progesterone administered to normally-cycling women during the midfollicular phase (cycle days 7–9).** Data are represented as mean ± standard error. Conversion of conventional to International System (SI) units: estradiol (pg/mL) multiplied by 3.67 (pmol/L); progesterone (ng/mL) multiplied by 3.18 (nmol/L). (From Liu JH, Yen SS. Induction of midcycle gonadotropin surge by ovarian steroids in women: a critical evaluation. *J Clin Endocrinol Metab.* 1983;57:797–802.)

GnRH release decreases at midcycle in women[45]; and functional neuroimaging suggests that pituitary metabolic activity, but not hypothalamic metabolic activity, increases at midcycle in women.[32] Moreover, in GnRH-deficient monkeys and women, gonadotropin surges occur with fixed-dose pulsatile GnRH therapy. Indeed, apparently-normal gonadotropin surges occurred in women with GnRH deficiency even when GnRH doses were reduced from 75 ng/kg to 25 ng/kg at midcycle.[46] Taken together, these data suggest that the midcycle gonadotropin surge in women primarily—if not exclusively—reflects an increase in gonadotrope responsiveness to GnRH.

The addition of progesterone to estradiol appears to augment GnRH-stimulated gonadotropin release. For example, in GnRH-deficient women treated with fixed-dose GnRH pulses, progesterone enhances pituitary gonadotropin responses to GnRH.[47] Similarly, progesterone-receptor antagonism decreases mean LH, LH amplitude, and LH responses to exogenous GnRH administered during the luteal phase.[48,49] Notably, progesterone also has an indirect effect to increase LH pulse amplitude, related to the inverse relationship between GnRH pulse frequency and LH pulse amplitude.

Uterine Events During the Menstrual Cycle

Cyclic changes of the endometrium are geared toward preparation for embryo implantation in the event of fertilization.[50] The follicular phase corresponds to menses and the proliferative phase of the uterine cycle: after menses, estradiol from developing ovarian follicles stimulates growth of the endometrial functional layer, prompting the growth of endometrial glands and an increase in the stromal extracellular matrix. The luteal phase corresponds to the secretory phase of the uterine cycle. Corpus luteum-derived progesterone inhibits estradiol-induced mitosis and induces differentiation into the secretory endometrium: endometrial glands become coiled, they secrete glycogen and other substances needed for implantation, and spiral arteries develop.

In a fertile cycle, implantation begins approximately 6–7 days after fertilization (7–10 days after ovulation). An approximately 3-day window surrounding the peak of luteal progesterone and estradiol production represents the "implantation window"—the time when the secretory endometrium is optimal for implantation. Trophoblast-derived CG maintains function of the corpus luteum and thus endometrial decidualization, which is critical for maintenance of pregnancy. If implantation does not occur, however, the corpus luteum regresses, and the withdrawal of progesterone support initiates a cascade of events that culminate in menstruation, or shedding of the functional portion of the uterine lining. The beginning of this process is characterized by contraction of uterine spiral arteries and arterioles, leukocyte infiltration, and degradation of the endometrial stroma. By the vasoconstriction model of menstruation, menses result from anoxia and reperfusion injury after uterine spiral artery and arteriole constriction and relaxation, respectively, whereas the inflammation hypothesis of menstruation holds that menses primarily reflect the inflammatory response induced by progesterone withdrawal. It is likely that a combination of vasoconstriction and inflammation contribute to menstruation. Menstrual flow, which typically lasts 4–6 days, primarily consists of endometrial debris and blood. A basal layer of endometrium remains intact for regeneration in response to follicular estrogen production.

INTEGRATED PHYSIOLOGY OF THE MENSTRUAL CYCLE

KEY POINTS

- Cyclical ovulation involves complex, coordinated, sequenced, and bidirectional interactions between the hypothalamic-pituitary unit and the ovary.
- During the follicular phase, FSH and LH act in concert to stimulate ovarian follicular development, selection of a dominant follicle, and an increase in estradiol production.
- In the midfollicular phase, ovarian estradiol stimulates endometrial proliferation, while both estradiol and inhibin B selectively restrain FSH release, discouraging the development of multiple dominant follicles.
- At midcycle, high concentrations of estradiol and a modest rise in progesterone—both derived from the mature preovulatory follicle—markedly augment pituitary LH and FSH responses to GnRH, inducing the midcycle gonadotropin surge, which in turn provokes ovulation.
- The remnant of the dominant follicle becomes a corpus luteum and secretes progesterone, estradiol, and inhibin A; these hormones suppress GnRH pulse frequency, inhibit FSH release, and prepare the endometrium for possible pregnancy.
- In a non-fertile cycle, the corpus luteum regresses, and circulating progesterone, estradiol, and inhibin A concentration levels fall; the loss of progesterone support leads to endometrial shedding (menstruation), and an increase in GnRH pulse frequency drives a monotropic rise in plasma FSH levels, which stimulates the next wave of follicular recruitment and selection of a dominant follicle.

Luteal-Follicular Transition

In a non-fertile cycle, the corpus luteum regresses toward the end of the luteal phase. Plasma concentrations of progesterone, estradiol, and inhibin A—products of the corpus luteum—decrease during the last 4–5 days of the luteal phase (Fig. 117.12). In addition to provoking endometrial shedding (menstruation), the decrease in progesterone and estradiol allows GnRH pulse frequency to increase from about 1 pulse every 3–5 hours to 1 pulse every 90–100 minutes in the early follicular phase (Fig. 117.13).[51] In one careful study, GnRH pulse frequency increased 4.5-fold (from 3 to 13.5 pulses per 24 hours), mean LH concentrations increased 67%, and mean FSH concentrations increased 3.5-fold across the luteal-follicular transition (Fig. 117.14).[24]

The marked increase in FSH across the luteal-follicular transition appears to reflect increased GnRH pulse frequency in the setting of decreasing estradiol and inhibin A concentrations—hormones that inhibited pituitary FSH release during much of the luteal phase. Studies in GnRH-deficient women receiving pulsatile exogenous GnRH indicate that the transition from a low physiological GnRH pulse frequency (every 240 minutes) to a higher physiological GnRH pulse frequency (every 90 minutes) is required for the normal intercycle rise in FSH.[52] The preferential increase in circulating FSH concentrations compared with LH likely reflects the longer circulating half-life of FSH as well. Importantly, the marked increase in FSH initiates dominant follicle selection and the next wave of cyclic follicular recruitment.

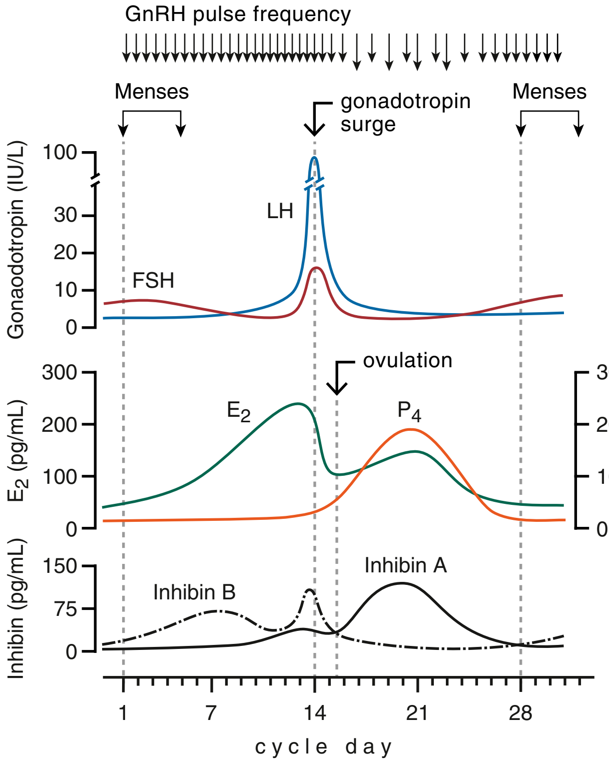


Figure 117.12 **Schematic representation of circulating concentrations of luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), progesterone (P4), and inhibins A and B across an ovulatory menstrual cycle.** Arrows at the top of the diagram indicate relative changes in pulsatile gonadotropin-releasing hormone (GnRH) secretion. Conversion of conventional to International System (SI) units: estradiol (pg/mL) multiplied by 3.67 (pmol/L); progesterone (ng/mL) multiplied by 3.18 (nmol/L). (After Marshall JC, Dalkin AC, Haisenleder DJ, et al. Gonadotropin-releasing hormone pulses: regulators of gonadotropin synthesis and ovulatory cycles. *Recent Prog Horm Res*. 1991; 47:155–187.)

Follicular Phase

Early Follicular Phase. The early follicular phase is characterized by relatively high plasma FSH concentrations and follicular growth. FSH stimulates growth of a cohort of antral follicles by activating granulosa cell proliferation and estradiol secretion, and FSH animates lead follicle growth and the establishment of dominance. Sleep-related slowing of GnRH pulse frequency is very prominent in the early follicular phase, which may support the robust FSH secretion characteristic of this portion of the cycle.[11] FSH-stimulated granulosa cell proliferation also leads to an increase in inhibin B production.

By definition, the initial days of the menstrual cycle are marked by menstrual bleeding. After completion of menstruation, a basal layer of endometrium remains. Thereafter, estradiol production from developing ovarian follicles stimulates endometrial proliferation.

Midfollicular Phase. FSH continues to stimulate the growth of recruited follicles in the first part of the follicular phase. However, only one follicle typically establishes dominance—at approximately 10 mm diameter—and progresses further towards full maturity and ovulation. Mechanisms accounting for the emergence of a single dominant follicle across the follicular phase are unclear, but estradiol and inhibin B appear to play important roles by reducing FSH secretion across the follicular phase. As described previously, inhibin B and estradiol from growing follicles suppress FSH release via a direct and selective action on the gonadotropes: this midfollicular reduction in circulating

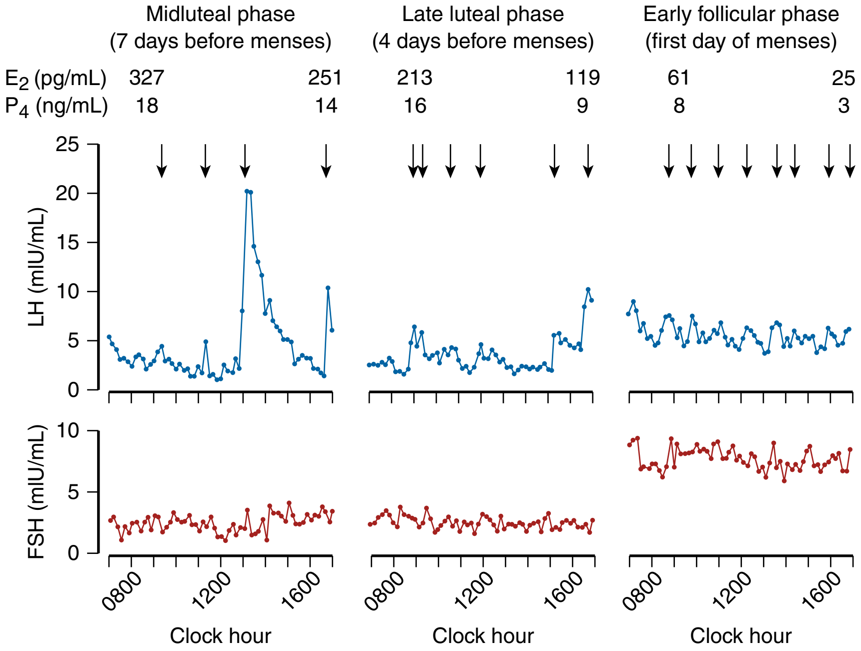


Figure 117.13 **Luteinizing hormone (LH) and follicle-stimulating hormone (FSH) concentration time series in the midluteal, late luteal, and early follicular phases.** Arrows indicate significant LH pulses. *E2,* Estradiol; *P4,* progesterone. Conversion of conventional to International System (SI) units: estradiol (pg/mL) multiplied by 3.67 (pmol/L); progesterone (ng/mL) multiplied by 3.18 (nmol/L). (From Marshall JC, Dalkin AC, Haisenleder DJ, et al. Gonadotropin-releasing hormone pulses: regulators of gonadotropin synthesis and ovulatory cycles. *Recent Prog Horm Res.* 1991;47:155–187; discussion 188–189.)

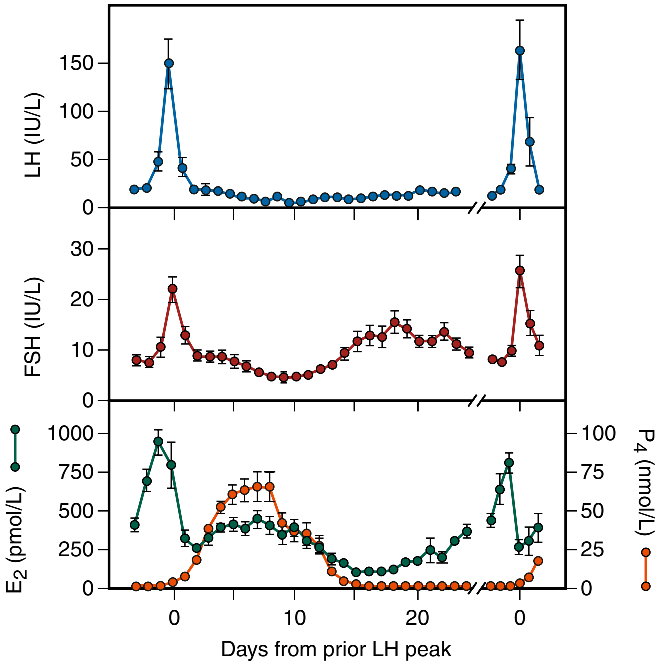


Figure 117.14 **Luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), and progesterone (P4) concentrations across the luteal-follicular transition.** Data are aligned to the day of the LH surge and represented as mean ± standard error. Conversion of International System (SI) units to conventional units: progesterone (nmol/L) divided by 3.18 (ng/mL); estradiol (pmol/L) divided by 3.67 (pg/mL). (From Hall JE, Schoenfeld DA, Martin KA, et al. Hypothalamic gonadotropin-releasing hormone secretion and follicle-stimulating hormone dynamics during the luteal-follicular transition. *J Clin Endocrinol Metab.* 1992;74:600–607.)

FSH concentrations reduces the growth stimulus for non-dominant follicles. Additional mechanisms likely pertain to regression of non-selected follicles.

LH pulse frequency increases to approximately 1 pulse every 60 minutes by the midfollicular phase—presumably reflecting a gradual release from progesterone negative feedback—while increasing estradiol from the dominant follicle restrains LH (GnRH) pulse amplitude. However, in contrast to the decrease in FSH concentrations, average LH concentrations increase across the follicular phase, perhaps in part reflecting the increase in GnRH pulse frequency.

Increasing estradiol concentrations in the midfollicular phase stimulate endometrial proliferation and development of endometrial glands. This process will continue into the late follicular phase.

Late Follicular Phase. In the latter part of the follicular phase, the dominant follicle grows at a rate of approximately 2 mm daily—related to continued granulosa cell proliferation and antral fluid accumulation. The dominant follicle typically reaches a diameter of 20–25 mm (range, 18–30 mm) at the time of ovulation. Although inhibin B levels decline in the later follicular phase, the growing dominant follicle produces increasing amounts of estradiol and inhibin A, both of which selectively restrain FSH release.

Toward the end of the follicular phase, estradiol and FSH stimulate the appearance of LH receptors on the granulosa cell layers closest to the basal lamina. Granulosa cells in the dominant follicle begin to luteinize and produce progesterone, plasma concentrations of which increase to approximately 1.0–1.5 ng/mL (~3.2–4.8 nmol/L) on the day of gonadotropin surge generation. Plasma estradiol concentration peaks at approximately 200–400 pg/mL (~730–1470 pmol/L) on the day before gonadotropin surge initiation.[42,53] The sustained increase in circulating estradiol concentrations provokes a marked increase in gonadotrope responsivity to GnRH, and the gonadotropin surge is initiated: in essence, the mature follicle signals readiness for ovulation by entraining the gonadotropin surge, thus coordinating the ovulatory stimulus with full dominant follicle maturation.

Midcycle Gonadotropin Surge and Ovulation

LH and FSH concentrations increase approximately 10- and 4-fold, respectively, at midcycle (Fig. 117.15).[53] This preovulatory gonadotropin surge relates to a marked increase in gonadotrope responsiveness to GnRH, which primarily reflects the estradiol positive feedback phenomenon.[41,54–56] Other factors may also be required for normal gonadotropin surges at midcycle, and candidate contributors include progesterone, inhibin A, neuropeptide, galanin, and kisspeptin, among others. For example, circulating progesterone concentrations begin to increase approximately 12 hours before gonadotropin surge generation (Fig. 117.15)[41,53]; progesterone appears to augment estradiol positive feedback at midcycle[41,55]; progesterone-receptor antagonism delays gonadotropin surge generation[57]; and some studies suggest that progesterone is important for normal surge timing, surge duration, and/or the FSH surge.[41,42,55,56]

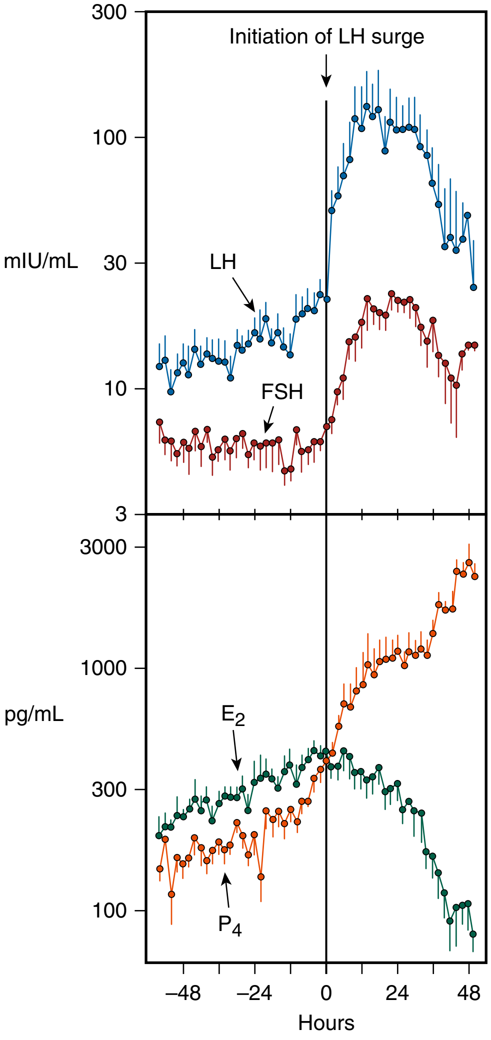


Figure 117.15 **Luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), and progesterone (P4) concentrations at midcycle, aligned to gonadotropin surge initiation.** Data are shown as mean ± standard error and displayed on a logarithmic scale. Conversion of conventional to International System (SI) units: estradiol (pg/mL) multiplied by 3.67 (pmol/L); progesterone (pg/mL) multiplied by 0.00318 (nmol/L). (From Hoff JD, Quigley ME, Yen SS. Hormonal dynamics at midcycle: a reevaluation. *J Clin Endocrinol Metab.* 1983;57:792–796.)

By acting on newly acquired LH receptors expressed by the outer granulosa cells, the LH surge triggers ovulation—rupture of the mature follicle with release of the ovum and surrounding granulosa cells (cumulus-oocyte complex) into the peritoneal space en route to the fallopian tubes. Ovulation occurs approximately 24–36 hours after the start of the surge. The LH surge also triggers resumption of meiosis I and luteinization of the follicular remnant (i.e., formation of the corpus luteum). The preovulatory FSH surge may play a role in the recruitment of follicles that will develop in subsequent cycles. The midcycle gonadotropin surge lasts approximately 48 hours,[53] but mechanisms underlying surge termination remain unclear.

Notably, there appear to be substantial species differences in the mechanisms underlying gonadotropin surge generation.[58] In rodents, GnRH and gonadotropin surges depend on estradiol-induced activation of kisspeptin neurons in the anteroventral periventricular nucleus (AVPV) of the preoptic area. In primates, however, kisspeptin neurons primarily reside in the infundibular (arcuate) nucleus, and gonadotropin surges occur in female monkeys after surgical separation of the mediobasal hypothalamus from the remainder of the brain. While kisspeptin neurons have been identified in the preoptic area in primates, functional analogues to rodent AVPV kisspeptin neurons have not been identified. Also, while LH surges in female rodents are temporally confined to the late afternoon—believed to reflect a circadian signal that is relayed only in the presence of preovulatory estradiol concentrations—gonadotropin surges do not appear to be constrained to a specific time of day in primates.

Luteal Phase

Serum estradiol concentration falls precipitously around the time of the gonadotropin surge, reflecting functional changes in the newly luteinized follicle. In contrast, progesterone concentrations continue to rise, reflecting LH-induced luteinization of granulosa cells shifting steroidogenesis toward progesterone secretion. During the luteal phase, the corpus luteum is the dominant structure in the ovary, and progesterone is the dominant steroid produced by the corpus luteum.

FSH and LH levels decrease rapidly and remain relatively low during much of the luteal phase. LH pulse frequency progressively decreases across the luteal phase, achieving a nadir frequency of approximately 1 pulse every 3–5 hours (or slower) at the mid- to late luteal stage. Low luteal GnRH pulse frequency reflects negative feedback by high levels of progesterone and estradiol, products of the corpus luteum. While high-amplitude LH pulses occur at regular intervals in the early luteal phase, LH pulse frequency and amplitude exhibit more variability in the mid- and late luteal phases.[59]

While progesterone secretion does not initially exhibit periodicity, prominent LH-induced pulses of progesterone secretion—coinciding with LH pulses—are evident in the mid- and late luteal phases.[59] The corpus luteum also produces estradiol, a number of other steroids (e.g., 17-hydroxyprogesterone, estrone), and inhibin A. Inhibin A and estradiol suppress FSH release from gonadotropes, helping to ensure that the antral follicle growth does not proceed prematurely.

Ovarian secretion of the above products increases in the 7–8 days after ovulation, and both progesterone and estradiol levels peak in the midluteal phase; this is when the secretory endometrium is most hospitable to implantation of the fertilized ovum—the “implantation window.” Continued progesterone secretion is a critical requirement for maintenance of pregnancy. If oocyte fertilization occurs, the resulting trophoblast produces CG, which activates LH receptors and maintains corpus luteum function. If fertilization does not occur, however, the corpus luteum eventually regresses. As corpus luteum function wanes, circulating progesterone, estradiol, and inhibin A concentrations fall. The release from negative feedback permits an increase in GnRH pulse frequency and FSH release, which initiates the next wave of follicular recruitment and dominant follicle selection. The reduction in circulating progesterone concentrations also initiates a cascade of events leading to endometrial shedding (menses).

NEUROENDOCRINE MECHANISMS OF ANOVULATION

KEY POINTS

- Anovulation is relatively common in recently-menarcheal girls and in the years leading up to menopause.
- Physiological anovulation occurs during pregnancy, lactation, and in the setting of inadequate energy availability.
- Numerous genetic defects can cause isolated hypogonadotropic hypogonadism (IHH) with accompanying pubertal failure and infertility.
- In functional hypothalamic amenorrhea (FHA), decreased energy availability, stress, and other factors lead to central inhibition of GnRH pulse frequency and hypogonadotropic hypogonadism.
- Hyperprolactinemia acts in a dose-dependent fashion to suppress GnRH pulse frequency, which can lead to gonadotropin deficiency.
- Polycystic ovary syndrome (PCOS) is associated with a number of neuroendocrine abnormalities that contribute to ovulatory dysfunction: relative resistance of the GnRH pulse generator to progesterone and estradiol negative feedback, persistently high GnRH pulse frequency, relatively high LH concentrations, and relatively low FSH concentrations.

Anovulation most commonly manifests as a disruption in expected menstrual patterns. Failure to achieve menarche by age 15 years, or by 3 years after thelarche, represents primary amenorrhea. In the absence of pregnancy, secondary (postmenarcheal) amenorrhea is defined by the absence of menstrual bleeding for at least 3 months. Although normal intermenstrual length is considered to be 21–35 days, oligomenorrhea is often defined as an intermenstrual length greater than 45 days or, alternatively, fewer than 9 menses per year.

Anovulation may relate to defects at any of the three levels of the hypothalamic-pituitary-ovarian axis. Anovulation may also relate to normal physiological processes (e.g., pregnancy, lactation) or reflect mechanisms that maintain whole-body homeostasis (e.g., in the setting of inadequate energy availability). Anovulation is also common in recently-menarcheal girls (see Chapter 102) and in the years leading up to menopause (see Chapter 122).

Physiological Anovulation

Peripubertal (Maturational) Anovulation. During childhood, LH (GnRH) pulse frequency and amplitude are low, and gonadotropin concentrations are inadequate to stimulate ovarian follicular development or substantial ovarian sex steroid production. Puberty is heralded by sleep-associated increases in both LH (GnRH) pulse frequency and amplitude. The resulting increases in circulating gonadotropin concentrations begin to stimulate ovarian follicular development and sex steroid production, especially estradiol, which largely accounts for the physical changes of early puberty (e.g., thelarche).

Adult patterns of gonadotropin secretion are established over time, and menstrual irregularity and anovulation are relatively common in recently-postmenarcheal adolescent girls. Available data suggests that the establishment of the complex cyclic relationships among the hypothalamus, pituitary, and ovary is contingent on maturation of both the hypothalamic-pituitary unit and the ovaries. For example, cyclic hormone activity with small amplitude LH surges have been demonstrated prior to menarche, with more prominent cyclic changes in LH and progesterone being observed across pubertal maturation.[60] In keeping with these findings, transdermal estradiol administration can provoke an approximately 5-fold increase in urinary LH in premenarcheal, early to mid-pubertal girls (Tanner II-III, ages 9–12 years).[61] While this proportional increase was similar to that in adult women, the absolute urinary LH peak was approximately one-quarter of that in adult women. Nonetheless, these data suggest early functioning of the estrogen positive feedback mechanism underlying gonadotropin surge generation, and these investigators hypothesized that ovarian factors play an important role in physiological peripubertal anovulation.

A recent detailed study compared cycles in healthy Tanner V girls (0.4–3.5 years postmenarche) with ovulatory cycles in adults.[62] Ovulatory adolescent cycles were associated with lower LH and FSH levels in the early follicular and luteal phases; lower estradiol concentrations in the early follicular, late follicular, preovulatory, and luteal phases; and a suggestion of lower luteal progesterone concentrations. After adjusting for lower preovulatory estradiol levels, the magnitude of the midcycle LH peak was similar to that in adult cycles, which may imply maturity of the estrogen positive feedback mechanism in such cycles. Ovulatory adolescent cycles were associated with lower early follicular phase inhibin B concentrations, even after statistically adjusting for lower FSH concentrations, in addition to lower mid- to late follicular estradiol concentrations, even after statistically adjusting for lower FSH and smaller follicle sizes: these findings may suggest ovarian immaturity. In adolescent ovulatory cycles with short luteal phases, LH, FSH, estradiol, and progesterone levels were low during the luteal phase. The magnitude of the midcycle LH peak was also reduced, even after adjusting for preovulatory estradiol concentrations, suggesting immature positive feedback as a possible mechanism for luteal insufficiency. Girls with anovulatory cycles demonstrated variable patterns: some demonstrated minimal follicle growth, whereas others demonstrated preovulatory follicles without ovulation (at times despite a robust midcycle LH peak), either with or without follicle luteinization. Anovulatory adolescent cycles were marked by lower LH and FSH in the early follicular, midcycle, and luteal phases; and the midcycle LH peak was attenuated even after statistically adjusting for estradiol levels. Anovulatory cycles also demonstrated reduced inhibin B concentrations in the early follicular phase, even after statistically adjusting for lower FSH, suggesting ovarian immaturity.

In summary, adult patterns of cyclic hypothalamic-pituitary-ovarian activity are established over time. Early cycles may be characterized by longer follicular phases, relatively slow follicle growth, delayed dominant follicle emergence, failure to achieve a preovulatory follicle, failure to achieve ovulation despite a midcycle LH surge, and/or ovulation with short luteal phases. It is tempting to speculate that the likelihood of more-mature cycle patterns increases as peripubertal girls mature—anovulatory cycles without follicle luteinization, anovulatory cycles with follicle luteinization, ovulatory cycles with luteal insufficiency, and normal ovulatory cycles[63]—but this notion requires further study. The endocrinology of sexual maturation and puberty is addressed in more detail in Chapter 102.

Pregnancy and Lactation. Pregnancy is the most common cause of amenorrhea. Markedly elevated placental sex steroid concentrations and significant hyperprolactinemia profoundly suppress GnRH and gonadotropin secretion, thus halting follicular development during pregnancy. The endocrinology of pregnancy and parturition is addressed in more detail in Chapter 129.

In mothers who do not breastfeed, ovulation typically resumes over the 8 weeks following delivery. In lactating mothers, however, GnRH secretion remains suppressed, and pregnancy is very uncommon in fully breastfeeding, amenorrheic women.[64,65] This presumably represents an advantageous physiological adaptation since short intervals between births can negatively impact infant well-being. Pulsatile exogenous GnRH can restore ovarian function in amenorrheic lactating women, suggesting that suppressed GnRH secretion is the primary cause of lactational amenorrhea.[66]

Hyperprolactinemia suppresses GnRH pulsatility, yet circulating prolactin concentrations gradually decrease towards normal despite continued breastfeeding. Animal studies suggest that lactation is associated with increased orexigenic neural system activity (e.g., NPY) and reduced anorexigenic neural system activity, and these alterations may suppress GnRH neuronal activity.[67] Indeed, the very high energy requirements of lactation likely impair GnRH secretion in many of the same ways described below in the context of reduced energy availability. In addition, the suckling stimulus likely inhibits the hypothalamic-pituitary-ovarian axis through additional neural mechanisms, and the intensity of suckling appears to be an important determinant of contraceptive effectiveness in women.

Selected Causes of Pathological Anovulation

The World Health Organization (WHO) has categorized anovulation into three primary classes. Although this classification system is not often used in the clinical setting, it represents a useful conceptual framework. Hypogonadotropic hypogonadal anovulation (WHO class 1) is characterized by estrogen deficiency and low or inappropriately normal serum LH and FSH concentrations. At its foundation, this relates to inadequate gonadotropin secretion for maintenance of cyclic ovulation, as may occur with isolated GnRH deficiency, FHA, and structural hypothalamic-pituitary damage (e.g., related to tumors, surgery, radiation, or pituitary infarction [Sheehan's syndrome]). Although hyperprolactinemic anovulation constitutes a separate category according to the WHO schema, its pathophysiology primarily relates to hypogonadotropic hypogonadism. The most common cause of normogonadotropic normoestrogenic anovulation (WHO class 2) is PCOS, although congenital adrenal hyperplasia is another important example. Women with PCOS are estrogen sufficient and typically have serum gonadotropin concentrations within the reference range, although as a group these patients demonstrate relative LH excess and relative FSH deficiency. Hypergonadotropic hypoestrogenic anovulation (WHO class 3) most commonly relates to a marked reduction in ovarian follicles (primary ovarian insufficiency). Such women are estrogen deficient and have elevated plasma LH and FSH concentrations due to the loss of central negative feedback by estrogens and inhibins.

Isolated Hypogonadotropic Hypogonadism. IHH typically refers to inadequate GnRH secretion or action related to one or more genetic defects. Patients with genetic defects that severely impair GnRH secretion or action demonstrate pubertal failure, hypogonadotropic hypogonadism, and infertility.[68]

The first discovered cause of Kallmann syndrome—a congenital disorder characterized by GnRH deficiency and abnormal sense of smell (anosmia or hyposmia)—was deletion of *ANOS1* (formerly *KAL1*), which encodes a secreted matrix glycoprotein (anosmin-1) believed to be critical for olfactory system development. Loss-of-function *ANOS1* mutations lead to abnormal olfactory system development, which interrupts the guidance infrastructure for migrating GnRH neurons. Accordingly, GnRH neurons fail to reach the hypothalamus appropriately. This syndrome exemplifies the importance of the olfactory guidance system as nascent GnRH neurons migrate from the nasal (olfactory) placode to the hypothalamus during embryological development.

Additional genes associated with Kallmann syndrome include those for *SEMA3A* (semaphorin 3A) and *SOX10* (SRY-box 10), among others. Numerous genes have been associated with both Kallmann syndrome and normosmic IHH, including *PROK2* (prokineticin 2), *PROKR2* (prokineticin 2 receptor), *FGF8* (fibroblast growth factor-8), *FGFR1* (fibroblast growth factor receptor 1), *CHD7* (chromodomain helicase DNA binding protein 7), and *NSMF* (N-methyl-D-aspartate [NMDA] receptor synaptonuclear signaling and neuronal migration factor), among others. Normosmic IHH may also be related to mutations of *GNRH1*, *GNRHR*, *KISS1*, *KISS1R*, *TAC3*, and *TAC3R*. Inheritance patterns are variable depending on the genetic defect(s), although some cases appear to be sporadic. Moreover, penetrance and manifestations can be variable among affected members of the same family, in part related to oligogenic IHH gene mutations.

In most cases of IHH—with notable exceptions including *GNRHR* mutations—pubertal changes can be induced by pulsatile exogenous GnRH therapy.[15,69] However, treatment more commonly involves sex steroid (estradiol) administration to induce and maintain secondary sex characteristics, with progesterone eventually added to prevent endometrial hyperplasia. In most patients, ovulation and fertility can be achieved with exogenous gonadotropin therapy or, much less commonly, pulsatile GnRH therapy.

Functional Hypothalamic Amenorrhea. Since pregnancy and lactation are metabolically demanding, reproductive function may be suppressed when energy availability is reduced (nutritional infertility). This may occur with absolute reductions in calorie intake, insufficient calorie intake vis-à-vis energy demands (e.g., exercise), or both. Stress may also impair reproductive function, presumably because the presence of stressors can herald an inauspicious environment for childbearing.

FHA represents the reversible suppression of hypothalamic-pituitary function that may accompany reduced body weight, disordered eating (e.g., restrictive eating patterns), excessive exercise, psychological stress, or a combination of these factors. These relationships are mediated by neural systems located in the hypothalamus that impair GnRH and gonadotropin secretion. The primacy of dysfunctional GnRH secretion is supported by studies demonstrating that pulsatile exogenous GnRH restores cyclic ovulation and fertility in women with FHA.[15] Women with FHA demonstrate variable patterns of LH pulse secretion, including absent pulses, low frequency and amplitude, low frequency only, low amplitude only, and (apparently) normal frequency and amplitude.[15,70,71] Some women with FHA demonstrate variable patterns over time (Fig. 117.16).[15,70,71] Many such patients exhibit a GnRH pulse frequency approximating 1 pulse every 3–4 hours, similar to that observed in the luteal phase. Regardless of the particular secretory patterns in women with FHA, GnRH and gonadotropin secretion is inadequate for normal follicular development, estrogen production, and ovulation. For example, while hourly GnRH pulses restore ovulatory function in GnRH-deficient monkeys, a long-term GnRH pulse frequency of 1 pulse every 2 hours was associated with reduced LH concentrations, limited follicular development, and anovulatory cycles; and a long-term pulse frequency of 1 pulse every 3 hours largely abolished follicular development.[72]

The observation that FHA can follow significant weight loss contributed to the "critical fatness hypothesis," which holds that FHA largely reflects insufficient body fat stores. However, subsequent studies suggested that reduced energy availability—rather than reduced fat stores per se—is a primary cause of impaired reproductive neuroendocrine function in these women. For example, calorie restriction can impair menstrual cyclicity before weight loss is evident, and amenorrhea can persist after weight restoration in those with a history

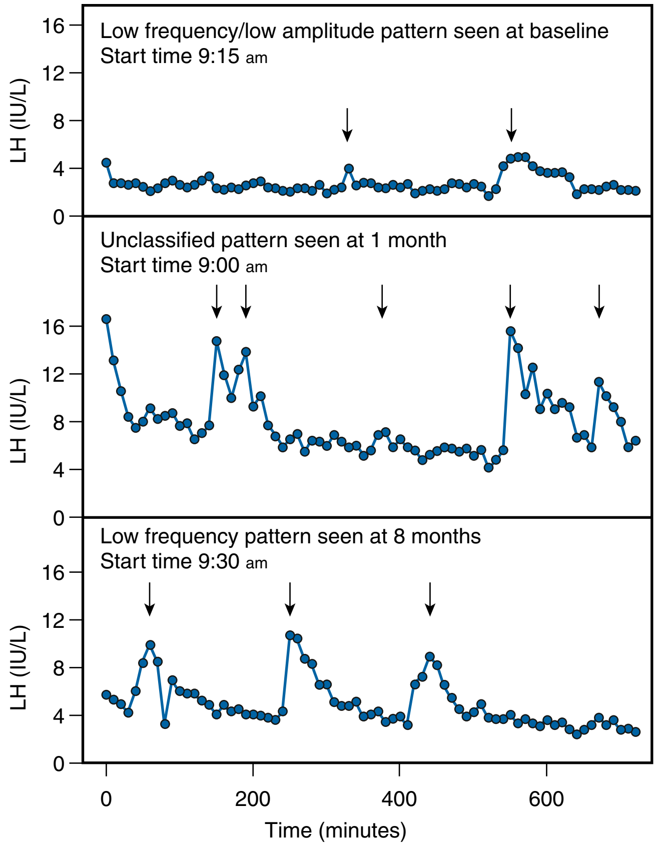


Figure 117.16 Three luteinizing hormone (LH) concentration time series in a single woman with functional hypothalamic amenorrhea studied longitudinally. This patient demonstrated three different LH pulse patterns at three different time points. Arrows indicate significant LH pulses. (From Perkins RB, Hall JE, Martin KA. Neuroendocrine abnormalities in hypothalamic amenorrhea: spectrum, stability, and response to neurotransmitter modulation. *J Clin Endocrinol Metab.* 1999;84:1905–1911.)

of eating disorders.[73] Studies in women and monkeys suggest that reproductive neuroendocrine dysfunction may not occur until energy availability is reduced by at least 30%, and altered LH pulsatility occurs within 5 days of controlled reductions of energy availability.[74,75] While some have suggested that excessive exercise per se may impair GnRH pulsatility, enhancing calorie intake to maintain adequate energy availability can prevent LH secretory abnormalities despite significant exercise loads.[76,77]

The neurobiological mechanisms underlying FHA are complex and incompletely understood. Decreased energy availability provokes numerous hormonal changes including reduced leptin, insulin, insulin-like growth hormone-1, and thyroid hormone concentrations; increased growth hormone and ghrelin concentrations; and activation of the hypothalamic-pituitary-adrenal axis.[78] Of interest, leptin plays a critical permissive role in reproductive function, and recombinant human leptin administration may rapidly (within 2 weeks) increase LH pulsatility in women with FHA, followed by increases in dominant follicle number, ovarian volume, and estradiol concentrations, with ovulation in a subset.[79]

A number of neuropeptides appear to mediate reproductive impairments in the setting of reduced energy availability, including kisspeptin, NPY, galanin-like peptide, beta-endorphin, CRH, ghrelin, and polypeptide YY. Such effects are at least partly mediated by kisspeptin neurons.[80,81] For example, leptin-deficient (ob/ob) mice demonstrate reduced hypothalamic *Kiss1* expression, which is partially reversed with leptin administration. Increased opioid tone also appears to contribute to low GnRH pulse frequency in FHA, as it can be rapidly reversed with opiate receptor blockade (naloxone).[82]

Stress can also contribute to the pathophysiology of FHA. Women with FHA may be more likely to exhibit perfectionism and difficulty coping with stressors; and amenorrheic athletes and women with anorexia nervosa have evidence of hypothalamic-pituitary-adrenal axis activation including elevated serum cortisol concentrations and blunted cortisol responses to exogenous CRH.[78] Mechanisms underlying the suppression of reproductive function during stress are highly complex; and while suppression of GnRH secretion is a major component, direct pituitary and gonadal effects may occur as well. A number of mediators have been implicated in stress-related inhibition of GnRH secretion, including CRH, urocortins (CRH-like peptides), adrenocorticotropic hormone (ACTH), arginine vasopressin (AVP), endogenous opioid peptides, and cortisol, in addition to noradrenergic, GABAergic, and serotoninergic neural pathways.

Women exhibit variable reproductive sensitivity to reduced energy availability and other forms of stress: some are more “stress sensitive,” whereas others are more “stress resilient.” Mechanisms for such variability remain incompletely defined, but one study suggested that mutations in some IHH-associated genes (e.g., *ANOS1*, *FGFR1*, *PROKR2*, and *GNRHR*) are more likely in women with FHA, suggesting that underlying genetic (and perhaps epigenetic) architecture plays an important role.[83]

Mainstays of FHA management include behavioral approaches to restoring sufficient energy availability (e.g., increased calorie intake, decreased physical activity) along with psychological support (e.g., cognitive behavioral therapy).[84] Ovarian hormone replacement—particularly transdermal estradiol with cyclic progestin—can also be used as needed. When pursued, pharmacological ovulation induction will generally require exogenous gonadotropins or pulsatile GnRH therapy.

Hyperprolactinemic Anovulation. Elevated circulating prolactin concentrations are commonly associated with ovulatory dysfunction. The causes of hyperprolactinemia are numerous and include pregnancy, lactation, and prolactin-secreting pituitary adenomas (prolactinoma). Moreover, since prolactin production and release from pituitary lactotrophs is tonically suppressed by hypophysiotropic dopamine, hyperprolactinemia may result from impaired hypothalamic dopamine secretion, interruption of hypothalamic dopamine transport to lactotrophs, or impaired action at lactotroph dopamine D2 receptors (e.g., some medications).

Hyperprolactinemia is associated with low-frequency, irregular LH pulse patterns, which are insufficient to maintain normal gonadotropin secretion.[85,86] Significant hyperprolactinemia can cause secondary hypogonadism with low or inappropriately normal serum gonadotropin concentrations, impaired follicular maturation, and failure to ovulate. The degree of menstrual cycle dysfunction generally correlates with the degree of hyperprolactinemia. As a general rule, mild hyperprolactinemia (2- to 3-fold elevated) can impair luteal progesterone secretion and reduce fertility, but may not disrupt menstrual cyclicity, while higher prolactin concentrations (2- to 5-fold elevated) tend to be associated with ovulatory dysfunction and either oligomenorrhea or amenorrhea. Women with severe hyperprolactinemia (more than 4- to 5-fold elevated) typically exhibit conspicuous secondary hypogonadism marked by estrogen deficiency and amenorrhea. It seems highly likely that different individuals exhibit hypothalamic-pituitary-ovarian dysfunction at different prolactin thresholds.

Prolactin-mediated GnRH suppression appears to involve hypothalamic opioids, as pulsatile LH (GnRH) secretion is improved with

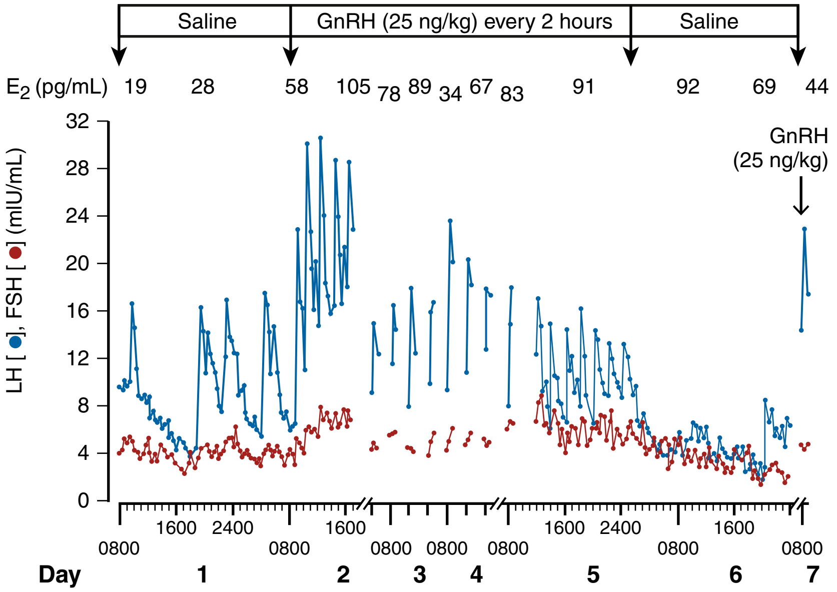


Figure 117.17 LH and FSH concentration time series in a woman with hyperprolactinemia (average serum prolactin 134 ng/mL [5.8 nmol/L]). LH pulse frequency and estradiol concentrations were low during saline infusion. Exogenous low-dose GnRH pulses administered every 2 hours for the following 4 days (days 2–5) led to an approximately 3-fold increase in serum estradiol (E_2) concentrations (circa 30 to 90 pg/mL). Conversion of conventional to International System (SI) units: estradiol (pg/mL) multiplied by 3.67 (pmol/L). *FSH,* Follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Sauder SE, Frager M, Case GD, et al. Abnormal patterns of pulsatile luteinizing hormone secretion in women with hyperprolactinemia and amenorrhea: responses to bromocriptine. *J Clin Endocrinol Metab.* 1984;59:941–948.)

naloxone (opiate receptor antagonist) despite continued hyperprolactinemia.[87] Impaired GnRH secretion also appears to reflect suppression of kisspeptin neuron function. For example, exogenous prolactin decreases kisspeptin expression in the arcuate nucleus in mice, but prolactin does not suppress pulsatile LH secretion in mice lacking prolactin receptors in arcuate nucleus kisspeptin neurons.[88] Moreover, exogenous kisspeptin restores gonadotropin secretion, ovulation rates, and estrus cyclicity in mice subjected to experimental hyperprolactinemia[89]; and exogenous kisspeptin may at least partially restore gonadotropin and estradiol secretion in women with hyperprolactinemia.[90]

Pulsatile GnRH treatment reverses hyperprolactinemia-related hypogonadotropic hypogonadism (Fig. 117.17)[85] and infertility, suggesting that impaired GnRH secretion is the primary cause of hyperprolactinemia-associated anovulation. However, the first-line medical approach to hyperprolactinemia-related ovulatory dysfunction is dopamine agonist therapy (e.g., cabergoline, bromocriptine), which suppresses prolactin secretion and thus restores cyclic ovulation.

Polycystic Ovary Syndrome. PCOS is a complex and heterogeneous disorder characterized by androgen excess, anovulation, and polycystic ovarian morphology (see Chapter 124). Although the precise causes of PCOS remain enigmatic, important factors contributing to PCOS pathophysiology include (potentially inherent) abnormalities of ovarian steroidogenesis, the effects of insulin resistance and compensatory hyperinsulinemia, and neuroendocrine dysfunction. Disordered gonadotropin secretion in PCOS contributes to both dysregulated steroidogenesis (hyperandrogenemia) and abnormal follicular dynamics (e.g., enhanced androgen-stimulated follicular activation but later follicular arrest).

Ovarian hyperandrogenemia in PCOS is LH dependent, as long-acting GnRH agonists markedly reduce androgen production in women with PCOS. Moreover, women and adolescents with PCOS exhibit persistently high GnRH pulse frequency, increased LH pulse amplitude, exaggerated LH responses to exogenous GnRH, relative LH excess, and relative FSH deficiency. In one study, after excluding women with PCOS who had recently ovulated, 75% exhibited elevated serum LH and 94% had elevated LH-to-FSH ratio.[91] LH concentrations tend to be lower in those with concomitant obesity, but LH (GnRH) pulse frequency is high in PCOS regardless of obesity status (Fig. 117.18).[92] Also of interest, a number of gonadotropin-related genes including *FSHB*, *FSHR* (FSH receptor), *LHB*, and *LHCGR* (LHCG receptor) have been implicated in the etiology of PCOS.[93,94]

High- and low-frequency GnRH pulses favor LH or FSH production, respectively; accordingly, elevated GnRH pulse frequencies contribute to relative excess and deficiency of LH and FSH, respectively, observed in PCOS. High GnRH pulse frequency in PCOS partly reflects relative resistance to progesterone and estradiol negative feedback, and this feedback defect can be reversed with androgen receptor blockade (Fig. 117.19).[95] Additional basic science observations suggest that hyperandrogenemia plays a central role in the neuroendocrine dysfunction of PCOS.[96] For example, dihydrotestosterone (DHT) increases GnRH neuron firing rates in murine brain slices. Prenatally androgenized (PNA) animals, which exhibit both endogenous ovarian hyperandrogenism and ovulatory dysfunction, exhibit high GnRH neuron firing frequency, increased LH secretion, and central resistance to the feedback effects of sex steroids. Similarly, in female rhesus monkeys, mild experimental hyperandrogenemia beginning at 1 year of age (prepubertal) increased LH pulse frequency 2.6-fold at 5 years of age.

Animal studies suggest that elevated GnRH pulsatility in PCOS is at least partly mediated by afferent neurons.[96] For example, some studies suggest that PNA rodents exhibit increased kisspeptin and NKB expression in the arcuate nucleus. The roles of kisspeptin and KNDy neurons in women with PCOS remain unclear. Of interest, however, neurokinin-3 receptor antagonism appears to reduce LH pulse frequency and serum LH concentrations in women with PCOS.[97]

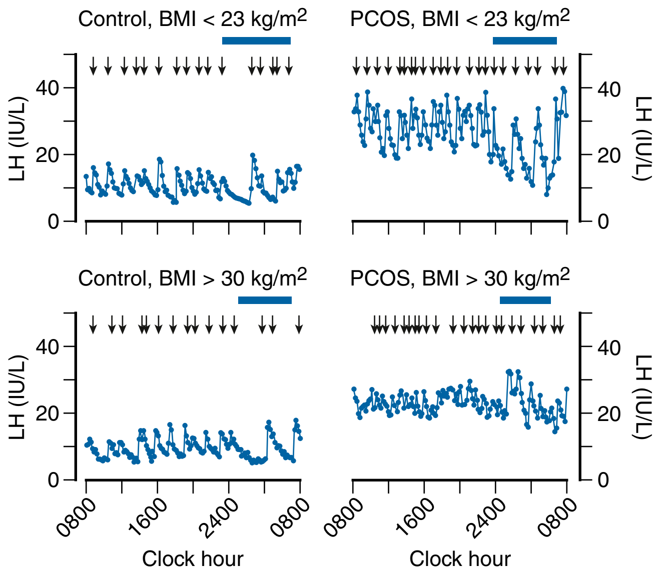


Figure 117.18 **LH concentration time series in women with and without PCOS and with or without obesity.** Arrows indicate significant LH pulses, and horizontal bars at the top of the panels indicate sleep periods. In this study, lean women with PCOS demonstrated 38% higher LH pulse frequency and 2.7-fold higher LH pulse amplitude compared to lean controls, rendering 3-fold elevated mean LH levels. As a group, women with PCOS and obesity demonstrated 50% higher LH pulse frequency, but no significant difference in LH pulse amplitude, rendering 1.9-fold elevated mean LH levels compared to controls with obesity. *BMI,* Body mass index. *LH,* luteinizing hormone; *PCOS,* polycystic ovary syndrome. (From Morales AJ, Laughlin GA, Butzow T, et al. Insulin, somatotropic, and luteinizing hormone axes in lean and obese women with polycystic ovary syndrome: common and distinct features. *J Clin Endocrinol Metab.* 1996;81:2854–2864.)

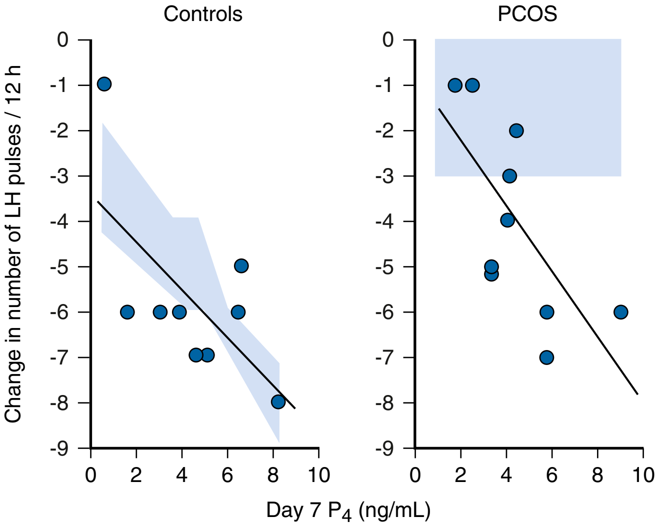


Figure 117.19 **Impact of androgen-receptor blockade (flutamide) on progesterone dose-LH pulse frequency response relationships in normally cycling women and women with PCOS.** In this study, GnRH pulse generator sensitivity to progesterone (P_4) negative feedback is assessed as the change in LH pulse frequency occurring with 7 days of exogenous estradiol and progesterone administration. The shaded areas indicate the range of responses in the absence of flutamide pretreatment, while data points represent subjects who were studied after 4 weeks of flutamide pretreatment. These data indicate that, in women with PCOS, flutamide pretreatment normalized the ability of progesterone to suppress LH (GnRH) pulse frequency. Conversion of conventional to International System (SI) units: progesterone (ng/mL) multiplied by 3.18 (nmol/L). *LH,* Luteinizing hormone; *PCOS,* polycystic ovary syndrome (From Eagleson CA, Gingrich MB, Pastor CL, et al. Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab.* 2000;85:4047–4052.)

GABAergic neurons also substantially impact GnRH neuron activity and appear to partly mediate effects of progesterone and DHT on GnRH neuronal function.[96] For example, PNA mice exhibit increased excitatory GABAergic drive onto GnRH neurons—a phenomenon that can be reversed with post-pubertal androgen receptor blockade.[98,99] Moreover, long-term, selective activation of arcuate nucleus GABAergic neuron terminals—which densely contact GnRH neurons—leads to PCOS-like features such as abnormal estrous cyclicity, hyperandrogenemia, and a trend toward increased LH pulse frequency.[100] Although the role of GABAergic neurons in the neuroendocrine dysfunction of PCOS remains unclear, cerebrospinal fluid GABA concentrations may be elevated in women with PCOS; and the use of valproate, which increases GABAergic tone, has been associated with PCOS.[96]

Recent interest has also focused on the potential role of anti-Müllerian hormone (AMH) in the neuroendocrine dysfunction of PCOS. Serum concentrations of AMH—derived from granulosa cells in pre- and small antral ovarian follicles (2–9 mm)—are 2- to 4-fold elevated in PCOS, and AMH directly stimulates GnRH neuron activity in mice.[101] In addition, when pregnant mice are exposed to experimental AMH excess, their female progeny demonstrate increased GABAergic appositions onto GnRH neurons, increased GnRH neuron firing rate, high LH pulse frequency, and circulating LH excess.[102] These findings were prevented by maternal co-treatment with a GnRH antagonist, suggesting that the fetal effects were mediated by abnormal maternal GnRH secretion.

No current treatments directly address the underlying pathophysiology in PCOS, and treatments are largely symptomatic. Combined hormonal (estrogen-progestin) oral contraceptives are generally considered the first-line pharmacological approach to hyperandrogenism and oligo-/amenorrhea in PCOS. The synthetic estrogen and progestin suppress GnRH and gonadotropin secretion, thus reducing ovarian androgen production; the synthetic estrogen also increases plasma sex hormone-binding globulin concentrations, which reduces androgen bioavailability. Anovulation in PCOS is associated with unopposed estrogen stimulation of the endometrium, which can lead to endometrial hyperplasia and increase the risk for endometrial carcinoma. These risks can be mitigated with exogenous progestin therapy. Lifestyle and weight optimization are foundational components of the overall treatment approach to PCOS, including for those with infertility. When needed, ovulation induction with an aromatase inhibitor is typically the first-line pharmacological approach to anovulatory infertility in PCOS, although more advanced treatments (e.g., exogenous gonadotropin therapy) are sometimes required.

CONCLUSION

The menstrual cycle is characterized by highly complex cyclic interactions among the hypothalamus, pituitary, ovaries, and endometrium. The fundamental teleology of the menstrual cycle relates to episodic presentation of a mature oocyte for fertilization in addition to preparation for early pregnancy. In short, the incompletely understood cyclic interactions described in this chapter undergird fertility in women and, thus, the very survival of the human species. Although it is tempting to speculate that most of the endocrine physiology of the menstrual cycle has been reasonably well described, it seems presumptuous to assume too much in this regard. With regard to neuroendocrine control of the cycle, much recent research attention has focused on the roles of neuronal populations afferent to GnRH neurons (e.g., kisspeptin neurons). Ovulatory dysfunction is a major cause of female-factor infertility, and accompanying physiological states (e.g., estrogen deficiency) may lead to negative long-term health consequences for women. As such, much research attention has rightly focused on the mechanisms underlying ovulatory disorders such as FHA and PCOS.

REFERENCES

1. Treloar AE, Boynton RE, Behn BG, et al. Variation of the human menstrual cycle through reproductive life. *Int J Fertil*. 1967;12:77–126.
2. Belchetz PE, Plant TM, Nakai Y, et al. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. *Science*. 1978;202:631–633.
3. Wildt L, Hausler A, Marshall G, et al. Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey. *Endocrinology*. 1981;109:376–385.
4. Moenter SM, Brand RM, Midgley AR, et al. Dynamics of gonadotropin-releasing hormone release during a pulse. *Endocrinology*. 1992;130:503–510.
5. Reame N, Sauder SE, Kelch RP, et al. Pulsatile gonadotropin secretion during the human menstrual cycle: evidence for altered frequency of gonadotropin-releasing hormone secretion. *J Clin Endocrinol Metab*. 1984;59:328–337.
6. Filicori M, Santoro N, Merriam GR, et al. Characterization of the physiological pattern of episodic gonadotropin secretion throughout the human menstrual cycle. *J Clin Endocrinol Metab*. 1986;62:1136–1144.
7. Rasmussen DD, Gambacciani M, Swartz W, et al. Pulsatile gonadotropin-releasing hormone release from the human mediobasal hypothalamus in vitro: opiate receptor-mediated suppression. *Neuroendocrinology*. 1989;49:150–156.
8. Rossmanith WG, Liu CH, Laughlin GA, et al. Relative changes in LH pulsatility during the menstrual cycle: using data from hypogonadal women as a reference point. *Clin Endocrinol*. 1990;32:647–660.
9. Gill S, Lavoie HB, Bo-Abbas Y, et al. Negative feedback effects of gonadal steroids are preserved with aging in postmenopausal women. *J Clin Endocrinol Metab*. 2002;87:2297–2302.
10. McCartney CR, Gingrich MB, Hu Y, et al. Hypothalamic regulation of cyclic ovulation: evidence that the increase in gonadotropin-releasing hormone pulse frequency during the follicular phase reflects the gradual loss of the restraining effects of progesterone. *J Clin Endocrinol Metab*. 2002;87:2194–2200.
11. Hall JE, Sullivan JP, Richardson GS. Brief wake episodes modulate sleep-inhibited luteinizing hormone secretion in the early follicular phase. *J Clin Endocrinol Metab*. 2005;90:2050–2055.
12. Lu C, Hutchens EG, Farhy LS, et al. Influence of sleep stage on LH pulse initiation in the normal late follicular phase and in polycystic ovary syndrome. *Neuroendocrinology*. 2018;107:60–72.
13. Sun BZ, Kangarloo T, Adams JM, et al. The relationship between progesterone, sleep, and LH and FSH secretory dynamics in early postmenarchal girls. *J Clin Endocrinol Metab*. 2019;104:2184–2194.
14. O'Dea LS, Finkelstein JS, Schoenfeld DA, et al. Interpulse interval of GnRH stimulation independently modulates LH secretion. *Am J Physiol*. 1989;256:E510–E515.
15. Santoro N, Filicori M, Crowley Jr WF. Hypogonadotropic disorders in men and women: diagnosis and therapy with pulsatile gonadotropin-releasing hormone. *Endocr Rev*. 1986;7:11–23.
16. Lam NY, Ferin M. Is the decrease in the hypophysiotropic signal frequency normally observed during the luteal phase important for menstrual cyclicity in the primate? *Endocrinology*. 1987;120:2044–2049.
17. Soules MR, Clifton DK, Bremner WJ, et al. Corpus luteum insufficiency induced by a rapid gonadotropin-releasing hormone-induced gonadotropin secretion pattern in the follicular phase. *J Clin Endocrinol Metab*. 1987;65:457–464.
18. Ubuka T, Tsutsui K. Reproductive neuroendocrinology of mammalian gonadotropin-inhibitory hormone. *Reprod Med Biol*. 2019;18:225–233.
19. Burger LL, Haisenleder DJ, Dalkin AC, et al. Regulation of gonadotropin subunit gene transcription. *J Mol Endocrinol*. 2004;33:559–584.
20. Thackray VG, Mellon PL, Coss D. Hormones in synergy: regulation of the pituitary gonadotropin genes. *Mol Cell Endocrinol*. 2010;314:192–203.
21. Stamatiades GA, Kaiser UB. Gonadotropin regulation by pulsatile GnRH: signaling and gene expression. *Mol Cell Endocrinol*. 2018;463:131–141.
22. Wide L, Naessen T, Sundstrom-Poromaa I, et al. Sulfonation and sialylation of gonadotropins in women during the menstrual cycle, after menopause, and with polycystic ovarian syndrome and in men. *J Clin Endocrinol Metab*. 2007;92:4410–4417.
23. Wide L, Eriksson K, Sluss PM, et al. Serum half-life of pituitary gonadotropins is decreased by sulfonation and increased by sialylation in women. *J Clin Endocrinol Metab*. 2009;94:958–964.
24. Hall JE, Schoenfeld DA, Martin KA, et al. Hypothalamic gonadotropin-releasing hormone secretion and follicle-stimulating hormone dynamics during the luteal-follicular transition. *J Clin Endocrinol Metabol*. 1992;74:600–607.
25. Gross KM, Matsumoto AM, Bremner WJ. Differential control of luteinizing hormone and follicle-stimulating hormone secretion by luteinizing hormone-releasing hormone pulse frequency in man. *J Clin Endocrinol Metab*. 1987;64:675–680.
26. Spratt DI, Finkelstein JS, Butler JP, et al. Effects of increasing the frequency of low doses of gonadotropin-releasing hormone (GnRH) on gonadotropin secretion in GnRH-deficient men. *J Clin Endocrinol Metab*. 1987;64:1179–1186.
27. Hall JE, Taylor AE, Hayes FJ, et al. Insights into hypothalamic-pituitary dysfunction in polycystic ovary syndrome. *J Endocrinol Invest*. 1998;21:602–611.
28. Schipper I, Hop WC, Fauser BC. The follicle-stimulating hormone (FSH) threshold/window concept examined by different interventions with exogenous FSH during the follicular phase of the normal menstrual cycle: duration, rather than magnitude, of FSH increase affects follicle development. *J Clin Endocrinol Metab*. 1998;83:1292–1298.
29. Zeleznik AJ. Premature elevation of systemic estradiol reduces serum levels of follicle-stimulating hormone and lengthens the follicular phase of the menstrual cycle in rhesus monkeys. *Endocrinology*. 1981;109:352–355.
30. Armstrong DT, Goff AK, Dorrington JH. Regulation of follicular estrogen biosynthesis. In: Midgley AR, Sadler WA, eds. *Ovarian Follicular Development and Function*. New York: Raven Press; 1979:169–182.
31. Gill S, Sharpless JL, Rado K, et al. Evidence that GnRH decreases with gonadal steroid feedback but increases with age in postmenopausal women. *J Clin Endocrinol Metab*. 2002;87:2290–2296.
32. Ottowitz WE, Dougherty DD, Fischman AJ, et al. [18F]2-fluoro-2-deoxy-D-glucose positron emission tomography demonstration of estrogen negative and positive feedback on luteinizing hormone secretion in women. *J Clin Endocrinol Metab*. 2008;93:3208–3214.
33. Marshall JC, Case GD, Valk TW, et al. Selective inhibition of follicle-stimulating hormone secretion by estradiol. Mechanism for modulation of gonadotropin responses to low dose pulses of gonadotropin-releasing hormone. *J Clin Invest*. 1983;71:248–257.
34. Shaw ND, Histed SN, Srouji SS, et al. Estrogen negative feedback on gonadotropin secretion: evidence for a direct pituitary effect in women. *J Clin Endocrinol Metab*. 2010;95:1955–1961.
35. le Nestour E, Marraoui J, Lahlou N, et al. Role of estradiol in the rise in follicle-stimulating hormone levels during the luteal-follicular transition. *J Clin Endocrinol Metab*. 1993;77:439–442.
36. Lahlou N, Chabbert-Buffet N, Christin-Maitre S, et al. Main inhibitor of follicle stimulating hormone in the luteal-follicular transition: inhibin A, oestradiol, or inhibin B? *Hum Reprod*. 1999;14:1190–1193.
37. Welt CK, Pagan YL, Smith PC, et al. Control of follicle-stimulating hormone by estradiol and the inhibins: critical role of estradiol at the hypothalamus during the luteal-follicular transition. *J Clin Endocrinol Metab*. 2003;88:1766–1771.
38. Nippoldt TB, Reame NE, Kelch RP, et al. The roles of estradiol and progesterone in decreasing luteinizing hormone pulse frequency in the luteal phase of the menstrual cycle. *J Clin Endocrinol Metab*. 1989;69:67–76.

39. Soules MR, Steiner RA, Clifton DK, et al. Progesterone modulation of pulsatile luteinizing hormone secretion in normal women. *J Clin Endocrinol Metab*. 1984;58:378–383.
40. Groome NP, Illingworth PJ, O'Brien M, et al. Measurement of dimeric inhibin B throughout the human menstrual cycle. *J Clin Endocrinol Metab*. 1996;81:1401–1405.
41. Liu JH, Yen SS. Induction of midcycle gonadotropin surge by ovarian steroids in women: a critical evaluation. *J Clin Endocrinol Metab*. 1983;57:797–802.
42. Taylor AE, Whitney H, Hall JE, et al. Midcycle levels of sex steroids are sufficient to recreate the follicle-stimulating hormone but not the luteinizing hormone midcycle surge: evidence for the contribution of other ovarian factors to the surge in normal women. *J Clin Endocrinol Metab*. 1995;80:1541–1547.
43. Shaw ND, Srouji SS, Histed SN, et al. Differential effects of aging on estrogen negative and positive feedback. *Am J Physiol Endocrinol Metab*. 2011;301:E351–E355.
44. Adams JM, Taylor AE, Schoenfeld DA, et al. The midcycle gonadotropin surge in normal women occurs in the face of an unchanging gonadotropin-releasing hormone pulse frequency. *J Clin Endocrinol Metab*. 1994;79:858–864.
45. Hall JE, Taylor AE, Martin KA, et al. Decreased release of gonadotropin-releasing hormone during the preovulatory midcycle luteinizing hormone surge in normal women. *Proc Natl Acad Sci U S A*. 1994;91:6894–6898.
46. Martin KA, Welt CK, Taylor AE, et al. Is GnRH reduced at the midcycle surge in the human? Evidence from a GnRH-deficient model. *Neuroendocrinology*. 1998;67:363–369.
47. Couzinet B, Brailly S, Bouchard P, et al. Progesterone stimulates luteinizing hormone secretion by acting directly on the pituitary. *J Clin Endocrinol Metab*. 1992;74:374–378.
48. Schaison G, George M, Lestrat N, et al. Effects of the antiprogesterone steroid RU 486 during midluteal phase in normal women. *J Clin Endocrinol Metab*. 1985;61:484–489.
49. Garzo VG, Liu J, Ulmann A, et al. Effects of an antiprogesterone (RU486) on the hypothalamic-hypophyseal-ovarian-endometrial axis during the luteal phase of the menstrual cycle. *J Clin Endocrinol Metab*. 1988;66:508–517.
50. Critchley HOD, Maybin JA, Armstrong GM, et al. Physiology of the endometrium and regulation of menstruation. *Physiol Rev*. 2020;100:1149–1179.
51. Marshall JC, Dalkin AC, Haisenleder DJ, et al. Gonadotropin-releasing hormone pulses: regulators of gonadotropin synthesis and ovulatory cycles. *Recent Prog Horm Res*. 1991;47:155–187. discussion 88–89.
52. Welt CK, Martin KA, Taylor AE, et al. Frequency modulation of follicle-stimulating hormone (FSH) during the luteal-follicular transition: evidence for FSH control of inhibin B in normal women. *J Clin Endocrinol Metab*. 1997;82:2645–2652.
53. Hoff JD, Quigley ME, Yen SS. Hormonal dynamics at midcycle: a reevaluation. *J Clin Endocrinol Metab*. 1983;57:792–796.
54. Yen SS, Tsai CC. Acute gonadotropin release induced by exogenous estradiol during the mid-follicular phase of the menstrual cycle. *J Clin Endocrinol Metab*. 1972;34:298–305.
55. Chang RJ, Jaffe RB. Progesterone effects on gonadotropin release in women pretreated with estradiol. *J Clin Endocrinol Metab*. 1978;47:119–125.
56. March CM, Goebelsmann U, Nakamura RM, et al. Roles of estradiol and progesterone in eiliciting the midcycle luteinizing hormone and follicle-stimulating hormone surges. *J Clin Endocrinol Metab*. 1979;49:507–513.
57. Batista MC, Cartledge TP, Zellmer AW, et al. The antiprogestin RU486 delays the midcycle gonadotropin surge and ovulation in gonadotropin-releasing hormone-induced cycles. *Fertil Steril*. 1994;62:28–34.
58. Plant TM. A comparison of the neuroendocrine mechanisms underlying the initiation of the preovulatory LH surge in the human, Old World monkey and rodent. *Front Neuroendocrinol*. 2012;33:160–168.
59. Filicori M, Butler JP, Crowley Jr WF. Neuroendocrine regulation of the corpus luteum in the human. Evidence for pulsatile progesterone secretion. *J Clin Invest*. 1984;73:1638–1647.
60. Zhang K, Pollack S, Ghods A, et al. Onset of ovulation after menarche in girls: a longitudinal study. *J Clin Endocrinol Metab*. 2008;93:1186–1194.
61. Rovner P, Keltz J, Allshouse A, et al. Induction of the LH surge in premenarchal girls confirms early maturation of the hypothalamic-pituitary-ovarian axis. *Reprod Sci*. 2018;25:33–38.
62. Sun BZ, Kangarloo T, Adams JM, et al. Healthy post-menarchal adolescent girls demonstrate multi-level reproductive axis immaturity. *J Clin Endocrinol Metab*. 2019;104:613–623.
63. Carlson LJ, Shaw ND. Development of ovulatory menstrual cycles in adolescent girls. *J Pediatr Adolesc Gynecol*. 2019;32:249–253.
64. Tay CC, Glasier AF, McNeilly AS. The 24 h pattern of pulsatile luteinizing hormone, follicle stimulating hormone and prolactin release during the first 8 weeks of lactational amenorrhoea in breastfeeding women. *Hum Reprod*. 1992;7:951–958.
65. Van der Wijden C, Kleijnen J, Van den Berk T. Lactational amenorrhea for family planning. *Cochrane Database Syst Rev*. 2003:CD001329.
66. Zinaman MJ, Cartledge T, Tomai T, et al. Pulsatile GnRH stimulates normal cyclic ovarian function in amenorrheic lactating postpartum women. *J Clin Endocrinol Metab*. 1995;80:2088–2093.
67. Smith MS, True C, Grove KL. The neuroendocrine basis of lactation-induced suppression of GnRH: role of kisspeptin and leptin. *Brain Res*. 2010;1364:139–152.
68. Boehm U, Bouloux PM, Dattani MT, et al. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*. 2015;11:547–564.
69. Young J, Xu C, Papadakis GE, et al. Clinical management of congenital hypogonadotropic hypogonadism. *Endocr Rev*. 2019;40:669–710.
70. Reame NE, Sauder SE, Case GD, et al. Pulsatile gonadotropin secretion in women with hypothalamic amenorrhea: evidence that reduced frequency of gonadotropin-releasing hormone secretion is the mechanism of persistent anovulation. *J Clin Endocrinol Metab*. 1985;61:851–858.
71. Perkins RB, Hall JE, Martin KA. Neuroendocrine abnormalities in hypothalamic amenorrhea: spectrum, stability, and response to neurotransmitter modulation. *J Clin Endocrinol Metab*. 1999;84:1905–1911.
72. Pohl CR, Richardson DW, Hutchison JS, et al. Hypophysiotropic signal frequency and the functioning of the pituitary-ovarian system in the rhesus monkey. *Endocrinology*. 1983;112:2076–2080.
73. Sterling WM, Golden NH, Jacobson MS, et al. Metabolic assessment of menstruating and nonmenstruating normal weight adolescents. *Int J Eat Disord*. 2009;42:658–663.
74. Loucks AB, Thuma JR. Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. *J Clin Endocrinol Metab*. 2003;88:297–311.
75. Lane MA, Black A, Handy AM, et al. Energy restriction does not alter bone mineral metabolism or reproductive cycling and hormones in female rhesus monkeys. *J Nutr*. 2001;131:820–827.
76. Loucks AB, Verdun M, Heath EM. Low energy availability, not stress of exercise, alters LH pulsatility in exercising women. *J Appl Physiol*. 1998;84:37–46.
77. Williams NI, Helmreich DL, Parfitt DB, et al. Evidence for a causal role of low energy availability in the induction of menstrual cycle disturbances during strenuous exercise training. *J Clin Endocrinol Metab*. 2001;86:5184–5193.
78. Misra M, Klibanski A. Neuroendocrine consequences of anorexia nervosa in adolescents. *Endocr Dev*. 2010;17:197–214.
79. Welt CK, Chan JL, Bullen J, et al. Recombinant human leptin in women with hypothalamic amenorrhea. *N Engl J Med*. 2004;351:987–997.
80. De Bond JA, Smith JT. Kisspeptin and energy balance in reproduction. *Reproduction*. 2014;147:R53–R63.
81. Navarro VM, Kaiser UB. Metabolic influences on neuroendocrine regulation of reproduction. *Curr Opin Endocrinol Diabetes Obes*. 2013;20:335–341.
82. Khoury SA, Reame NE, Kelch RP, et al. Diurnal patterns of pulsatile luteinizing hormone secretion in hypothalamic amenorrhea: reproducibility and responses to opiate blockade and an alpha 2-adrenergic agonist. *J Clin Endocrinol Metab*. 1987;64:755–762.

83. Caronia LM, Martin C, Welt CK, et al. A genetic basis for functional hypothalamic amenorrhea. *N Engl J Med*. 2011;364:215–225.
84. Gordon CM, Ackerman KE, Berga SL, et al. Functional hypothalamic amenorrhea: an endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2017;102:1413–1439.
85. Sauder SE, Frager M, Case GD, et al. Abnormal patterns of pulsatile luteinizing hormone secretion in women with hyperprolactinemia and amenorrhea: responses to bromocriptine. *J Clin Endocrinol Metab*. 1984;59:941–948.
86. Klibanski A, Beitins IZ, Merriam GR, et al. Gonadotropin and prolactin pulsations in hyperprolactinemic women before and during bromocriptine therapy. *J Clin Endocrinol Metab*. 1984;58:1141–1147.
87. Cook CB, Nippoldt TB, Kletter GB, et al. Naloxone increases the frequency of pulsatile luteinizing hormone secretion in women with hyperprolactinemia. *J Clin Endocrinol Metab*. 1991;73:1099–1105.
88. Brown RSE, Khant Aung Z, Phillipps HR, et al. Acute Suppression of LH secretion by prolactin in female mice is mediated by kisspeptin neurons in the arcuate nucleus. *Endocrinology*. 2019;160:1323–1332.
89. Sonigo C, Bouilly J, Carre N, et al. Hyperprolactinemia-induced ovarian acyclicity is reversed by kisspeptin administration. *J Clin Invest*. 2012;122:3791–3795.
90. Millar RP, Sonigo C, Anderson RA, et al. Hypothalamic-pituitary-ovarian axis reactivation by kisspeptin-10 in hyperprolactinemic women with chronic amenorrhea. *J Endocr Soc*. 2017;1:1362–1371.
91. Taylor AE, McCourt B, Martin KA, et al. Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 1997;82:2248–2256.
92. Morales AJ, Laughlin GA, Butzow T, et al. Insulin, somatotropic, and luteinizing hormone axes in lean and obese women with polycystic ovary syndrome: common and distinct features. *J Clin Endocrinol Metab*. 1996;81:2854–2864.
93. Hiam D, Moreno-Asso A, Teede HJ, et al. The genetics of polycystic ovary syndrome: an overview of candidate gene systematic reviews and genome-wide association studies. *J Clin Med*. 2019;8.
94. Deswal R, Nanda S, Dang AS. Association of luteinizing hormone and LH receptor gene polymorphism with susceptibility of polycystic ovary syndrome. *Syst Biol Reprod Med*. 2019;65:400–408.
95. Eagleson CA, Gingrich MB, Pastor CL, et al. Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab*. 2000;85:4047–4052.
96. McCartney CR, Campbell RE. Abnormal GnRH pulsatility in polycystic ovary syndrome: recent insights. *Curr Opin Endocr Metab Res*. 2020;12:78–84.
97. George JT, Kakkar R, Marshall J, et al. Neurokinin B receptor antagonism in women with polycystic ovary syndrome: a randomized, placebo-controlled trial. *J Clin Endocrinol Metab*. 2016;101:4313–4321.
98. Sullivan SD, Moenter SM. Prenatal androgens alter GABAergic drive to gonadotropin-releasing hormone neurons: implications for a common fertility disorder. *Proc Natl Acad Sci U S A*. 2004;101:7129–7134.
99. Silva MS, Prescott M, Campbell RE. Ontogeny and reversal of brain circuit abnormalities in a preclinical model of PCOS. *JCI Insight*. 2018;3.
100. Silva MSB, Desroziers E, Hessler S, et al. Activation of arcuate nucleus GABA neurons promotes luteinizing hormone secretion and reproductive dysfunction: implications for polycystic ovary syndrome. *EBioMedicine*. 2019;44:582–596.
101. Cimino I, Casoni F, Liu X, et al. Novel role for anti-Mullerian hormone in the regulation of GnRH neuron excitability and hormone secretion. *Nat Commun*. 2016;7:10055.
102. Tata B, Mimouni NEH, Barbotin AL, et al. Elevated prenatal anti-Mullerian hormone reprograms the fetus and induces polycystic ovary syndrome in adulthood. *Nat Med*. 2018;24:834–846.

118

Folliculogenesis, Ovulation, and Luteogenesis

Marco Conti and Paolo Rinaudo

OUTLINE

Ovary Anatomy and Histology, 1971
Folliculogenesis, 1971
  Preantral Follicles, 1972
  Antral or Graafian Follicles, 1976
Ovulation, 1981
  Oocyte Maturation Following the LH Surge, 1981
Luteogenesis, 1982
  Luteinization, 1983
  Luteolysis, 1984
Conclusion, 1984

The human ovary has two primary functions during the reproductive years: (1) to produce every 28 days a single Graafian follicle that secretes estradiol (E_2) and ovulates a mature oocyte at about midpoint in the menstrual cycle; and (2) to produce an endocrine structure called the *corpus luteum* that secretes large amounts of E_2 and progesterone (P_4), which act on the uterine endometrium to prepare it for implantation of an embryo and sustain a conceptus in early gestation. A major feature of this ovarian activity is its cyclic nature, which is reflected in the growth and development of the dominant follicle and corpus luteum, respectively. These cyclic changes are regulated predominantly by changing concentrations of the anterior pituitary hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH). The objective of this chapter is to review the structural and functional changes that occur in the human ovary to evoke a menstrual cycle, namely, folliculogenesis, ovulation, and luteogenesis. Although outside the scope of this chapter, it should be noted that ovarian endocrine secretions have major systemic effects on body homeostasis.

KEY POINTS

- Less than 0.1% of follicles ovulate over a woman's reproductive lifespan.
- Follicle development is a highly coordinated process involving intraovarian factors and hormonal stimulation.
- The completion of oocyte maturation requires epidermal growth factors (EGFs) as well as other local factors.
- Follicle rupture and ovulation are induced by local growth factor, P_4, and prostaglandin production.

OVARY ANATOMY AND HISTOLOGY

In normal adult women, the ovaries are oval-shaped organs, each measuring 2.5 to 5.0 cm in length, 1.5 to 3.0 cm in width, and 0.6 to 1.5 cm in thickness. The ovaries are covered by a sheet of squamous or cuboidal epithelium, the ovarian surface epithelium (also called the *germinal epithelium)*, which rests on a basement membrane. The ovarian surface epithelium is of mesodermal origin and may give rise to surface epithelial-stromal tumors. Beneath the surface epithelium is a layer of dense connective tissue termed the *tunica albuginea*. The proximal edge of the ovary, the hilum, is attached to the broad ligament by the mesovarium and to the vasculature.

The ovary is organized into two principal regions: a central region called the medulla and a peripheral region called the cortex (Fig. 118.1). Embedded in the stroma of the medulla is a mass of blood vessels and associated nerves, which radiate toward the cortex. In the hilum are nests of Leydig-like cells, called *hilus cells,* which secrete testosterone in response to LH stimulation.[1]

The cortex is a dynamic structure with follicles and corpora lutea (see Fig. 118.1). These histologic units exhibit a specialized architecture that reflects their stage of growth and development. The pool of follicles is divided into two major classes: growing and non-growing. Most (90% to 95%) follicles are non-growing or "primordial." Once a primordial follicle is recruited to grow, its size, structure, and position in the cortex change dramatically. During this process, the follicle transitions through three developmental stages: primary, secondary, and tertiary or Graafian stages (see Fig. 118.1). Selection of the single dominant follicle destined to ovulate occurs during the Graafian stage, and those follicles that are not selected degenerate by a process called *atresia*. After ovulation of the large dominant follicle, the follicle wall transforms into a corpus luteum through a process called *luteinization*. The corpus luteum, in the absence of pregnancy, ultimately degenerates by a process called *luteolysis*.

KEY POINTS

- While the exact length of time needed for a primordial follicle to reach the ovulatory stage is unknown, it is accepted that it is at least several months long.
- Preantral follicles are gonadotropin-responsive, whereas antral follicles require gonadotropins for their growth.
- The phosphatidylinositol-3-kinase /AKT/mTOR and Hippo pathways are critical regulators of follicular growth.

FOLLICULOGENESIS

The ability of a primordial follicle to develop into an ovulatory follicle is at the very foundation of menstrual cyclicity and ovarian hormone

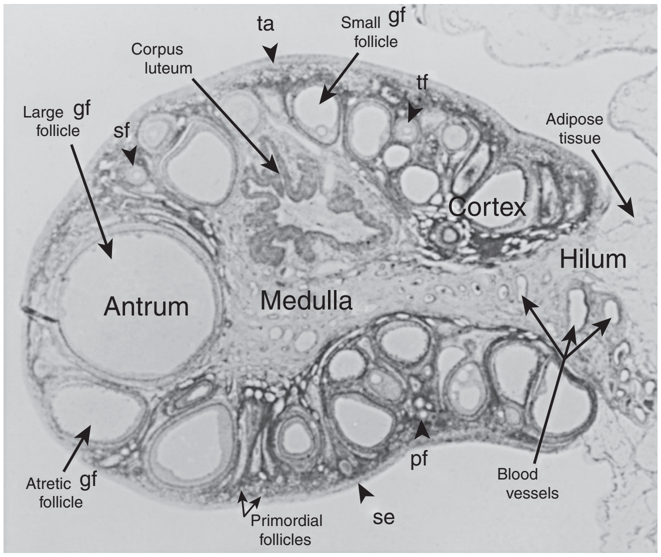


Figure 118.1 Photomicrograph of an adult primate ovary showing the follicular and luteal units in the cortex and large blood vessels and nerves in the medulla. *Gf,* Graafian follicle; *pf,* primary follicle; *se,* serous or surface epithelium; *sf,* secondary follicle; *ta,* tunica albugínea; *tf,* tertiary follicle. (Modified from Bloom W, Fawcett DW, eds. *A textbook of histology.* Philadelphia: WB Saunders; 1975:860.)

production. As a follicle grows and develops, important changes occur within the three major cell populations: oocytes, granulosa cells, and theca cells (TCs). These cellular changes include proliferation, differentiation, growth, and apoptosis. Two major classes of regulatory molecules control these cellular activities, namely, hormones and growth factors. Folliculogenesis can be divided into two phases: preantral growth and antral growth. The interval from recruitment of a primordial follicle to the completion of its growth into a secondary follicle is referred to as the *preantral* phase. While primordial follicles are not sensitive to gonadotropins, primary and secondary follicles show progressively increased responsiveness to gonadotropins.

In the antral phase, a fluid-filled cavity or "antrum" develops within the now *tertiary* or *Graafian follicle*, which is dependent on gonadotropins for growth. Graafian follicles grow progressively larger, mainly by accumulating increasing amounts of fluid in the antrum; usually only one follicle (the *dominant follicle*) will be larger than the remaining ones. The fully developed Graafian follicle is termed the *preovulatory follicle.*

Preantral Follicles

The development of a preantral follicle is characterized by three major events: (1) recruitment of a primordial follicle into the growing pool; (2) growth and differentiation of the oocyte; and (3) acquisition of FSH and LH receptors in the granulosa cells and TCs, respectively.

Primordial Follicle. The primordial follicles comprise a pool of nongrowing follicles from which all preovulatory follicles are ultimately derived. A primordial follicle consists of a small (≈15 μm diameter) oocyte arrested in the dictyotene stage of meiosis, a single layer of flat, squamous pregranulosa cells, and a thin basal lamina that encircles both cell types (Fig. 118.2). Primordial follicles are localized in the ovarian cortex. No blood vascular system is directly associated with the primordial follicles[2]; however, evidence in the rhesus monkey suggests that they may be innervated, most notably by vasoactive intestinal peptide nerves.[3] Although dormant, primordial follicles express genes required to maintain quiescence, until they are recruited into the growing pool.[4]

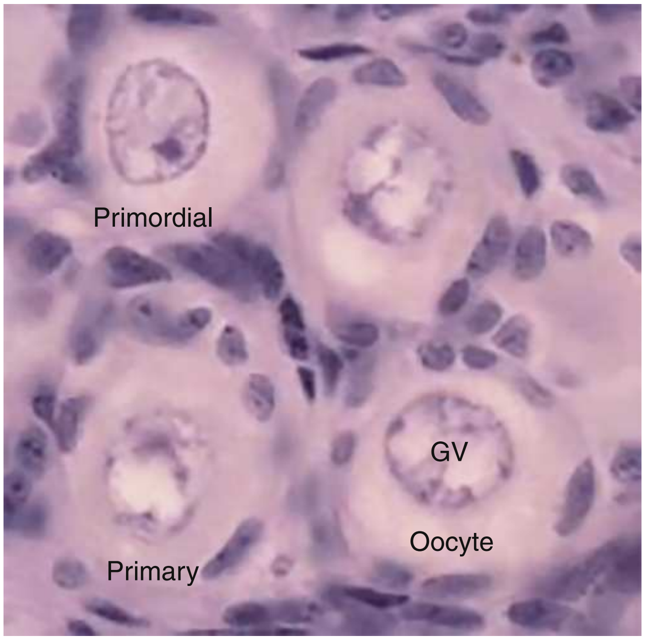


Figure 118.2 Photomicrograph of a portion of the cortex of the ovary. Primordial (non-growing) follicles consist of a small oocyte arrested in diplotene of meiosis I surrounded by a single layer of squamous granulosa cells. A primary follicle contains a growing oocyte and a single layer of cuboidal granulosa cells. Secondary follicles contain a nearly full-grown diplotene oocyte with an intact germinal vesicle or nucleus, a thin zona pellucida, and two or more layers of cuboidal granulosa cells. *GV,* germinal vesicle.

In humans, primordial follicles are formed in the fetal ovaries between the sixth and ninth month of gestation (Fig. 118.3).[5] At this time, all viable female germ cells have committed to meiosis. As a result, by the end of gestation there is no further production of new oocytes. This dogma has been challenged with the isolation of putative "oocyte stem cells" from mouse and human ovary.[6,7] However, this possibility is contradicted by several recent publications and remains highly controversial.[8,9] Consequently, it is most likely that all oocytes in a woman's ovaries are present at birth. Once the primordial follicles are formed, some are recruited to grow. As a woman ages, the process of recruitment continues on a regular basis until the total primordial follicle pool is exhausted. This event, the menopause, occurs in most women at approximately 51 years of age. An important concept is that the time course of loss of primordial follicles or ovarian reserve (OR) follows a simple power curve and can be monitored by measuring anti-Müllerian hormone (AMH) levels or by counting the number of antral follicles of size 2–10 mm visualized by transvaginal ultrasonography (this count is called the "antral follicular count") (Fig. 118.4). Although the mechanisms regulating this accelerated loss are unknown, this phenomenon is of clinical importance because it is associated with a significant decrease in fecundity.[10]

Certainly, one of the foremost questions in reproductive biology concerns the basis of initial primordial follicle growth and maturation. Structurally, the process of primordial-to-primary follicle transition begins with a change in shape in the granulosa cells from a squamous to a cuboidal shape, which, in turn, is accompanied by or associated with a commitment of the granulosa to divide, albeit very slowly.[11] That the first visible signs of primordial follicle growth occur in the granulosa cells fits the prediction that granulosa cells may play a key role in the process.

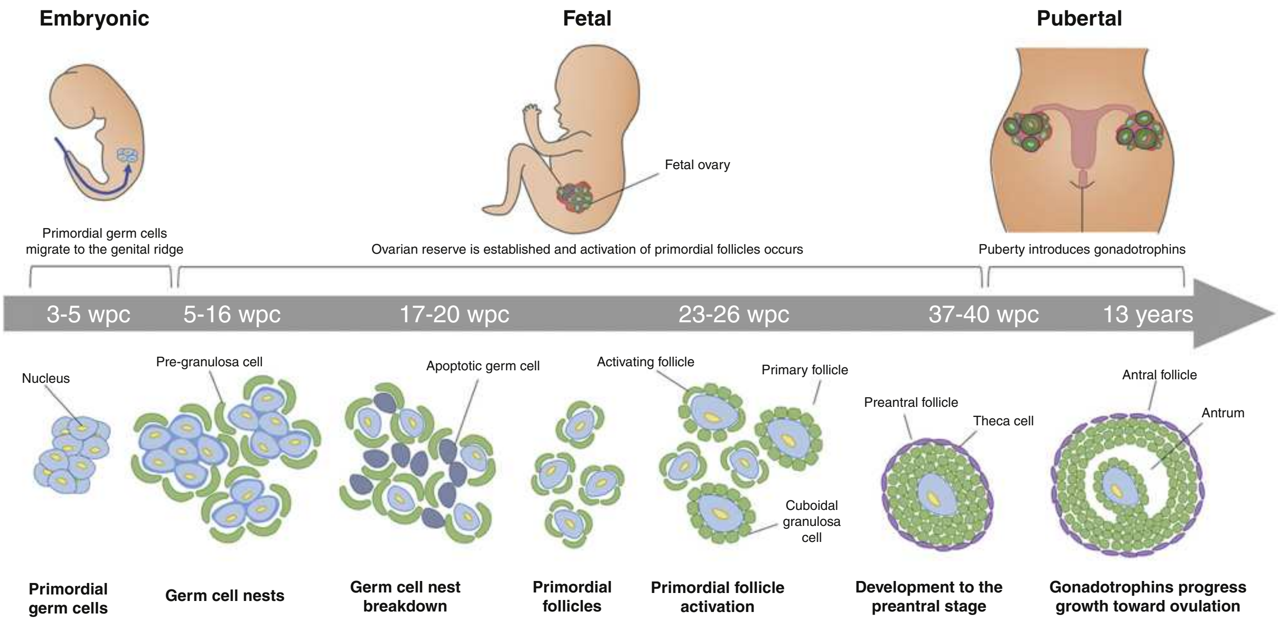

Embryonic
Fetal
Pubertal
Fetal ovary
Primordial germ cells migrate to the genital ridge
Ovarian reserve is established and activation of primordial follicles occurs
Puberty introduces gonadotrophins
3-5 wpc
5-16 wpc
17-20 wpc
23-26 wpc
37-40 wpc
13 years
Nucleus
Pre-granulosa cell
Apoptotic germ cell
Activating follicle
Primary follicle
Preantral follicle
Theca cell
Antral follicle
Antrum
Cuboidal granulosa cell
Primordial germ cells
Germ cell nests
Germ cell nest breakdown
Primordial follicles
Primordial follicle activation
Development to the preantral stage
Gonadotrophins progress growth toward ovulation

Figure 118.3 The development of ovarian reserve. Primordial germ cells migrate to the genital ridge 3–5 weeks post conception (wpc). By 5–16 wpc and up to 17–20 wpc, following rapid proliferation, germ cell nests are visible. At this point, germ cells nest break down, some germs cells undergo apoptosis, whereas others are surrounded by pre-granulosa cells, forming primordial follicles. Oocytes in the primordial follicles start meiosis and remain arrested at prophase I. Starting at 23–26 wpc, the first wave of primordial follicle activation commences and at birth (37–40 wpc), a small proportion of preantral and antral follicles are present. Before puberty, activated follicles can only grow until the early antral stage. At puberty, the cyclical changes in gonadotrophin levels allow progression beyond the early antral stage and ovulation may occur. (From Ford EA, Beckett EL, Roman SD, et al. Advances in human primordial follicle activation and premature ovarian insufficiency. *Reproduction*. 2020;159:R15–R29.)

Primary Follicle. A primary follicle consists of a growing oocyte, a single layer of cuboidal granulosa cells, and a basal lamina (see Figs. 118.2 and 118.5). In the primary follicle, several important events occur, including expression of FSH receptors by the granulosa cells, formation of gap junctions between the oocyte and granulosa cells, dramatic oocyte growth, and zona pellucida (ZP) deposition by the oocyte.

The expression of FSH receptor in the granulosa cells[12] opens the possibility that the follicle may become responsive to gonadotropin, even at this early stage. Thus, although preantral follicle growth can occur in the absence of FSH, granulosa cells of this immature structure may be gonadotropin-responsive.

A second key event that occurs in the primary follicle (Fig. 118.6) is the development of gap junctions between the granulosa cells and the oocyte.[13] Gap junctions comprise a family of proteins termed *connexins (Cx)*.[14] Functionally, they allow for transfer between cells of low-molecular-weight molecules, including the second messengers cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), calcium, nutrients, and ATP.[13–15] Gap junctions between granulosa cells are expressed at the earliest stages of primary follicle development and are mainly composed of Cx43.[13] Gap junctions between the oocyte and surrounding granulosa cells comprise Cx37.[16] It is important to note that Cx37-deficient female mice (Cx37 knockouts) lack Graafian follicles, fail to ovulate, and develop numerous abnormal corpora lutea.[16] Thus, physiologically important interactions between oocytes and granulosa cells via Cx37 channels are obligatory for normal folliculogenesis and fertility in mice.

A third event is the growth and differentiation of the oocyte in the primary follicle (see Figs. 118.2 and 118.5). Studies in rodents have demonstrated that the granulosa cells are essential to oocyte growth and development.[17,18] It seems likely that Cx37 connections between the granulosa and the oocyte allow for the exchange of nutrient and regulatory molecules that are required for proper oocyte function. Specifically, they evoke changes in oocyte metabolism that influence gene expression. One example is a marked increase in ZP gene expression. As the oocyte grows, it synthesizes and secretes an extracellular matrix, ZP, which in time encapsulates the egg (see Figs. 118.2, 118.5, and 118.6). The importance of ZP is demonstrated by the fact that it contains the species-specific receptor for capacitated sperm and provides a block to polyspermy.[19] The ZP in humans consists of four glycoproteins, ZP-1, ZP-2, ZP-3, and ZP4, whereas three ZP proteins are present in the mouse.[20] In mice, ZP-3 was thought to be particularly important in sperm binding through its carbohydrate moiety.[21] However, more recent data indicate that ZP-2 may be the primary zona ligand for spermatozoa.[22] Clinically, considerable interest has been devoted in the development of immunocontraceptive vaccines based on the ZP-3 antigen.[23,24]

Secondary Follicle. The secondary follicle consists of a larger oocyte (80 μm) surrounded by a complete ZP, two to eight layers of cuboidal or columnar granulosa cells, and a presumptive theca layer immediately peripheral to the basal lamina (see Figs. 118.2 and 118.5). The acquisition of a theca layer is a major new feature of developing secondary follicles. The presumptive theca consists of several layers of elongated fibroblast-like cells that run radially around the entire follicle (see Fig. 118.2). Theca development is accompanied by angiogenesis. Consequently, the secondary follicle is exposed to important hormones such as FSH, LH, and insulin present in the circulation. Mechanisms underlying theca formation and angiogenesis during secondary follicle development are not clear.[1] However, because theca

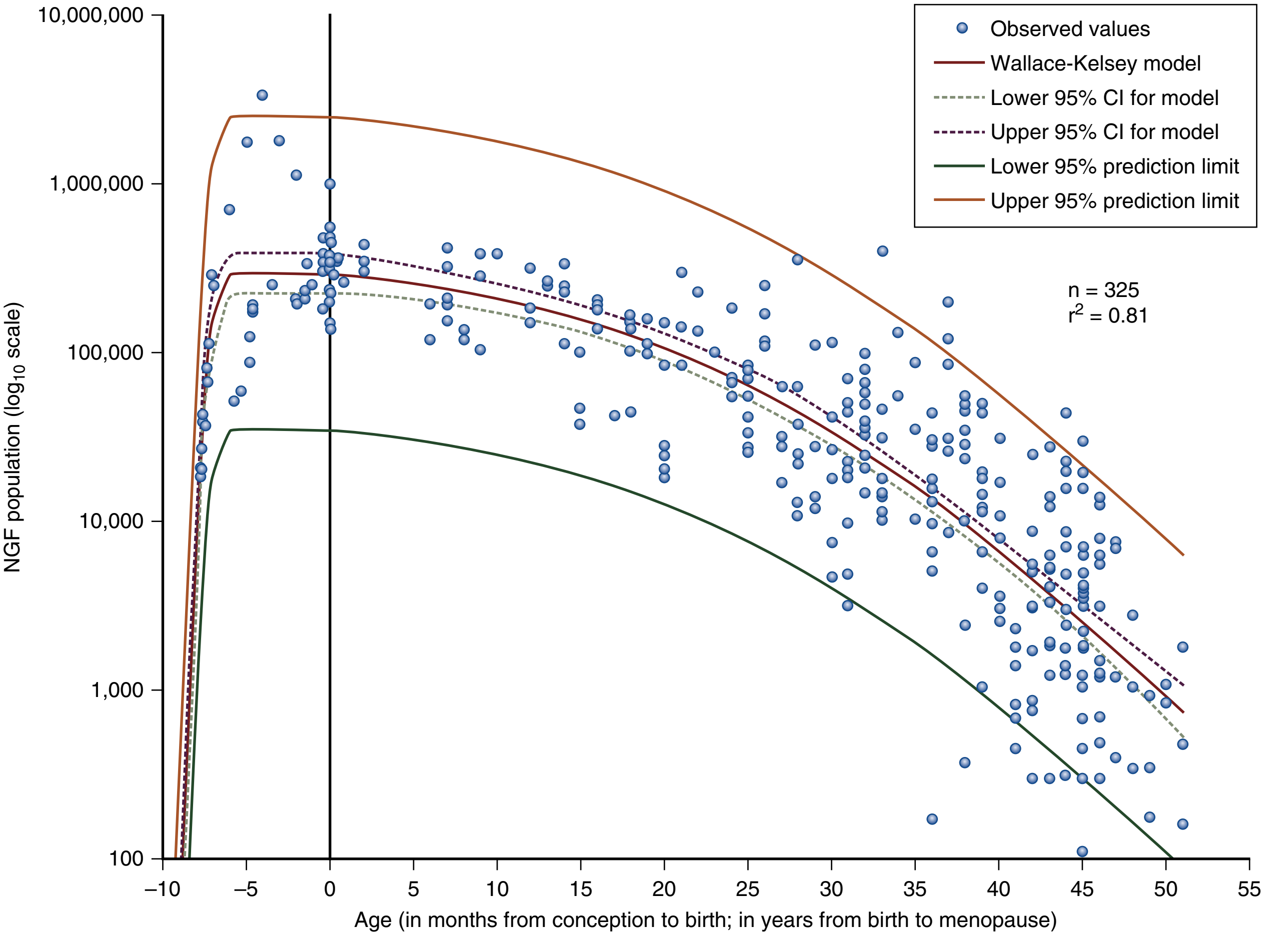


Figure 118.4 **Morphometric analysis of normal human ovaries showing the age-related decrease in the total number of primordial follicles (NGF) within both ovaries from birth to the menopause.** As a consequence of recruitment, the number of NGFs decreases progressively from approximately 1 million at birth to about 25,000 at 37 years of age. Note that the rate of loss of primordial follicles accelerates approximately 2-fold at 37.5 ± 1.2 years, with the number being reduced to about 1000 at approximately 51 years of age. (From Wallace WH, Kelsey TW. Human ovarian reserve from conception to the menopause. *Plos One.* 2010;5:e8772.)

formation requires the presence of follicles, it is believed that granulosa cells, or possibly the oocyte itself, produce factors that direct this process. Among those peptides under consideration are Kit ligand (KL), growth differentiation factor-9 (GDF-9), insulin, insulin-like growth factor-1 (IGF-1) and/or IGF-2, and activin.[25–27] In addition to these growth factors, ample evidence indicates that LH is responsible for early TC differentiation.[1,28,29] The degree to which LH influences TC growth is uncertain.

Role of Growth Factors in Folliculogenesis. Although it is clear that the ovarian cycle is governed by the combinational and sequential actions of key circulating hormones, including FSH, LH, P_4, androgens, E_2, and insulin, large-scale cellular and mutational analyses have led to the identification of local signaling pathways that can modulate the effects of circulating hormones.[30] These local ligands, termed *growth factors*, are capable of interacting with receptors to activate signal transduction pathways that affect multifunctional responses, including cell proliferation, differentiation, and apoptosis. The most intensely and best studied ovarian growth factors include IGF,[31,32] transforming growth factor-β (TGF-β),[30,33] transforming growth factor-α (TGF-α), EGF and related growth factors,[34] fibroblast growth factor (FGF), and cytokines. Thus, the processes of folliculogenesis, ovulation, and luteogenesis require a precise spatial and temporal pattern of expression of different growth factors. In particular, the expression patterns of these growth factors form a controlling network that ensures the proper sequence and timing of developmental events triggered by key hormones such as FSH and LH and play a critical role in the bidirectional interactions between oocytes and somatic cells.

Accordingly, it is recognized that the function of bone morphogenetic proteins (BMPs) is integral to follicular development.[35] For example, BMP-15 and GDF-9 are oocyte-derived growth factors that act to control the ovulation quota[36] by regulating mitosis and the supporting activities of FSH and LH. Therefore, BMPs appear to be obligatory factors that control the hormone-dependent ovarian cycle. In the same vein, EGF-like growth factors are essential for oocyte maturation and ovulation. It is almost certain that growth factors play critical roles in directing the proliferation, differentiation, and death of cells in the human ovary; indeed, evidence suggests aberrant GDF-9 expression in oocytes of women with polycystic ovary syndrome (PCOS).[37] Understanding how specific growth factor signaling pathways lead to activation of particular effector genes could have far-reaching implications for understanding ovarian physiology, infertility, and pathophysiology, including cancer.

Regulation of Primordial Follicular Development. Despite considerable advances, the exact mechanisms regulating primordial

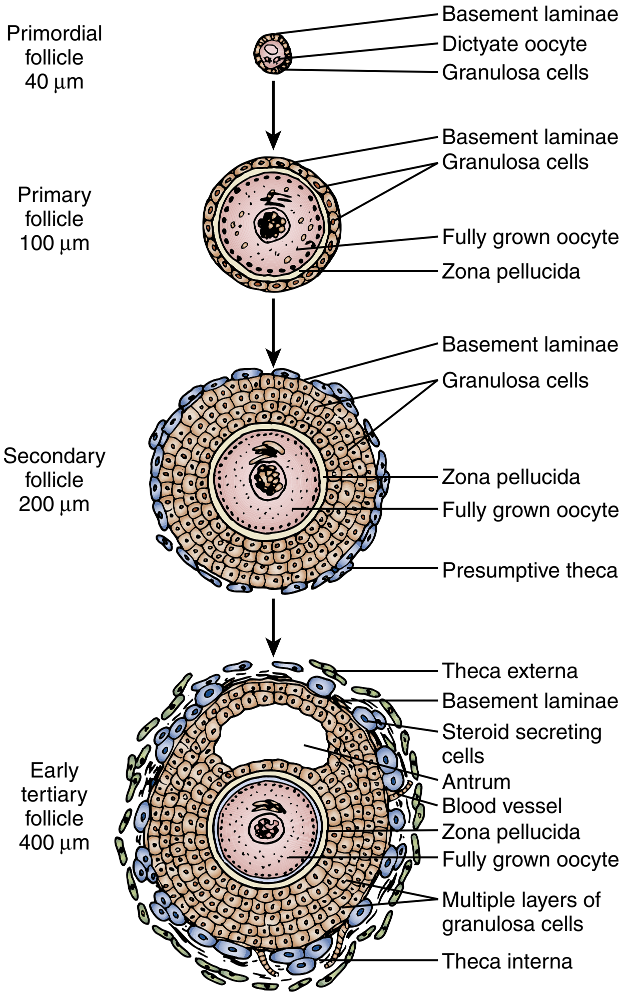

Primordial follicle 40 μm
Basement laminae
Dictyate oocyte
Granulosa cells
Primary follicle 100 μm
Basement laminae
Granulosa cells
Fully grown oocyte
Zona pellucida
Secondary follicle 200 μm
Basement laminae
Granulosa cells
Zona pellucida
Fully grown oocyte
Presumptive theca
Early tertiary follicle 400 μm
Theca externa
Basement laminae
Steroid secreting cells
Antrum
Blood vessel
Zona pellucida
Fully grown oocyte
Multiple layers of granulosa cells
Theca interna

Figure 118.5 Diagram of developing preantral follicles showing steps in the gonadotropin-independent stages of folliculogenesis from recruitment of a primordial follicle into the growing pool through its development to the early antrum or tertiary (cavitation) stage. (From Erickson GF. The ovary: basic principles and concepts. In Felig P, Baxter JD, Frohman LA, eds. *Endocrinology and metabolism,* 3rd ed. New York: McGraw-Hill; 1995:973–1015.)

follicle recruitment remain unclear.[38] While initiation of follicle growth can be regulated by negative and positive growth factors,[39,40] primordial follicle development is likely under constitutive repressive control. In fact, when the ovarian cortex is placed into culture, precocious activation of primordial follicles occurs.[41] Genetic evidence in rodents indicates that Notch and Wnt signaling are critical for the formation and maintenance of the primordial follicle.[42,43] Similarly, mice bearing mutations in the transcription factors Nobox, Sohlh1, or Sohlh2 or the Cdk inhibitor CDKN1B show defects in the primordial to primary follicle transition.[44,45] Importantly, genetic models provided evidence that phosphoinositide 3-kinase (PI3K)/AKT/mTOR/FOXO pathway is involved in the primordial to primary follicle transition (Fig. 118.7). Mice with loss of function of FOXO3,[46] PTEN,[47] or Tsc1 and Tsc2[48] genes in primordial oocytes display massive activation of all dormant follicles and, eventually, depletion of the follicular reserve. Indeed, pharmacological manipulation of this pathway in the "*in vitro* activation" technology may be a strategy to promote primordial follicle recruitment and may be used clinically to induce *in vitro* maturation of primordial follicles in patients with premature ovarian insufficiency.[38]

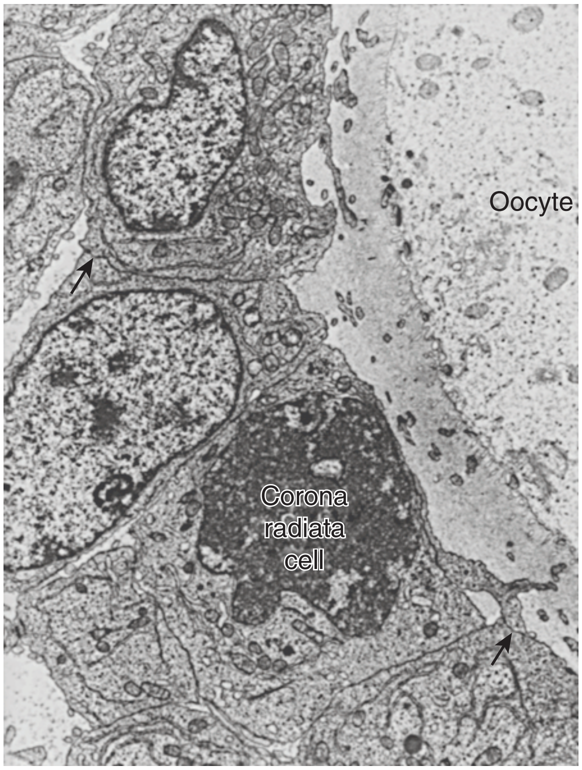

Oocyte
Corona radiata cell

Figure 118.6 **Electron micrograph of a portion of an oocyte/cumulus complex.** Gap junctions (comprising connexin 37) between the processes of corona radiata cells traversing the zona pellucida (ZP) and the oolemma are illustrated by *arrowheads.* Gap junctions between the corona granulosa cells are illustrated by *black arrows.* (From Gilula NB, Epstein ML, Beers WH. Cell-to-cell communication and ovulation: a study of the cumulus-oocyte complex. *J Cell Biol.* 1978;78:58–75.)

KL is one of the key factors involved in initiating follicle activation.[27] In addition, insulin,[49,22] testosterone,[50] BMP-7 and BMP-15,[51,52] GDF-9,[53] and restricted food intake have been shown to stimulate follicle development.[54] For example, in the rat ovary, BMP-7 activates advancement of primordial follicle growth.[52] In sheep, but not mice, BMP-15 has an essential role in promoting the primary to secondary follicle transition during early folliculogenesis, while restraining the transition of follicles to the dominant preovulatory stage.[51,55] Absence of this oocyte-derived peptide as in homozygous carriers of the ovine Inverdale and Hanna *BMP-15* mutations results in infertility with streak ovaries and a block at the primary stage of folliculogenesis.[51] Similarly, targeted deletion of GDF-9 in the oocyte precludes the development of ovarian follicles past the primary stage in mice.[53] Preliminary evidence in women with premature ovarian failure showing mutations in *BMP-15* and *GDF-9* further support these findings.[56–58]

Conversely, evidence in laboratory animals suggests that follicle maturation is inhibited by EGF[54] and AMH (Müllerian-inhibiting substance[59]). In AMH-null mice, the numbers of preantral and small antral follicles were greater in ovaries of 25-day-old and 4-month-old females compared with wild-type females. Moreover, in 4-month-old AMH-null females, smaller numbers of primordial follicles were present, and in 13-month-old animals, essentially no primordial follicles were observed.[60] These findings reflect a generalized progression of follicle growth in the absence of AMH. Importantly, these early stages of follicular development appear to be exquisitely sensitive to environmental endocrine disruptors.[61]

Regulation of Further Follicular Development. Other factors play important roles in follicular development, including FSH, the Hippo pathways, and possibly androgens. As mentioned above, during

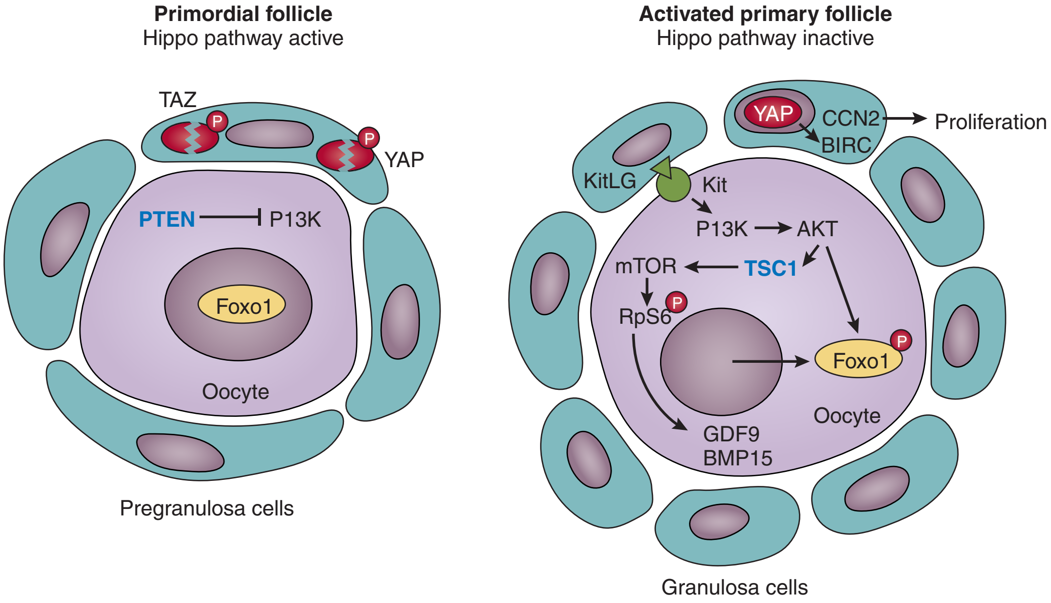


Figure 118.7 Regulation of primordial follicle growth. Genetic and pharmacological models indicate that activation of primordial follicles growth require activation of the Pi3k AKT mTOR FOXO pathway in the oocyte. While the exact nature of ligands involved in the activation is unknown, KIT ligand plays an important role. In addition, disruption of the Hippo pathway in granulosa cells, for example, following ovarian cortex fragmentation, results in proliferation and growth of granulosa cells. PTEN, TSC1, and FOXO are in blue color to indicate that genetic knockout models of these genes show activation of primordial follicle growth.

differentiation of the primary follicle, the gene that encodes the FSH receptor is activated in granulosa cells[12]; this, in turn, is followed by acquisition of cell-surface FSH receptors that can be visualized using ^{125}I-labeled FSH autoradiography.[62] This establishes the follicle potential to be stimulated by FSH-signaling pathways. Although the process of FSH receptor expression is poorly understood in women, evidence in rodents indicates that activin produced by the granulosa cells increases FSH receptor expression by autocrine mechanisms.[63] While follicles at this stage are not dependent on FSH for growth, they may be responsive to it.

The Hippo pathway is a key player in regulating preantral follicular development. Specifically, preantral and antral follicle development is restrained by the inhibitory Hippo signaling pathway, and disruption of this pathway by ovarian fragmentation has resulted in follicular development[64] (Fig 118.7).

Pharmacological and genetic models suggest that androgens may affect preantral follicular development. Administration of androgen[50] or knockout of granulosa-specific androgen receptor[65] results in changes in antral follicle number. Although the weak androgen dehydroepiandrosterone (DHEA) is often prescribed to women with low OR, evidence suggesting benefit is promising but still is inconclusive.[66]

Antral or Graafian Follicles

KEY POINTS

- Coordinated gonadotropin regulation of TCs and granulosa cells defines antral follicle development.
- As in other developmental processes, follicle development and final stages of maturation preceding follicle ruptures are under the control of circulating gonadotropins but also locally released bioactive molecules including growth factors, steroid hormones, cytokines, and other mediators.
- These regulators activate or repress the function of different layers of follicular cells. Moreover, they control the physical coupling of somatic cells with the oocytes and the flow of metabolic and signaling molecules.

During the late stages of growth of the secondary follicle, a clear fluid begins to accumulate between some granulosa cells. The fluid is a transudate of serum components and the secretion of TCs and granulosa cells. When the follicle reaches approximately 400 µm in diameter, the fluid coalesces into a crescent-shaped space called the *antrum*.[67] By definition, the follicle is now a *tertiary* or *Graafian follicle* (see Figs. 118.5 and 118.8). This process, termed *cavitation* or *beginning of antrum formation*, results in remodeling granulosa cells to form a miniature Graafian follicle (see Fig. 118.8). Cavitation is gonadotropin-independent. Precisely how the antrum forms is unknown, but direct evidence of a role for activin has been provided.[68]

TC differentiation is an additional important step that, while started at the secondary follicle stage, continues during cavitation. During cavitation, a subpopulation of cells that surround the growing follicle becomes epithelial-like and acquires ultrastructural and functional characteristics of active steroid-secreting cells.[1] As these changes occur, the cells, now called *theca interstitial cells*, express LH receptors.[1] At the periphery of the *theca interna* is a layer of concentrically arranged fusiform cells that have the ultrastructural and functional characteristics of smooth muscle cells.[69] They constitute the theca externa of the tertiary follicle (see Fig. 118.8).

By the time of cavitation, the oocyte has attained its full size (≈120 µm in diameter) and appears fully developed structurally (see Fig. 118.8). Therefore, the oocyte does not grow further despite the fact that the Graafian follicle as a whole might continue to enlarge to a diameter of 20 mm or more.[70]

Classification of Graafian Follicles. Graafian follicles can be defined structurally as a heterogeneous family of relatively large follicles (400 µm to >20 mm at ovulation or 50-fold increase in diameter) that display an antrum containing follicular fluid, or *liquor folliculi*. The antrum is a characteristic structural feature of all Graafian follicles. As such, the term *antral follicle* is synonymous with the Graafian follicle.

In a broad sense, Graafian follicles can be divided into two major groups: healthy and atretic. The main difference between these is whether or not apoptosis, or programmed cell death, is occurring in the granulosa cells. All Graafian follicles (healthy or atretic) follow a progressive and divergent trajectory during their development. The healthy Graafian follicle becomes progressively larger and more differentiated with time until it reaches the preovulatory stage. The mechanism by which a healthy Graafian follicle is formed depends on pituitary gonadotropins and involves the differential and temporal pattern of expression of specific genes that lead to follicle cell cytodifferentiation and proliferation.[71] In a similar manner, atretic or non-dominant Graafian follicles undergo a temporal pattern of gene expression, which, in turn, results in cessation of growth and activation of apoptosis.

The size of a Graafian follicle is determined largely by the size of the antrum, which, in turn, is determined by the volume of follicular fluid. An atretic follicle rarely reaches a size greater than 10 mm.[69]

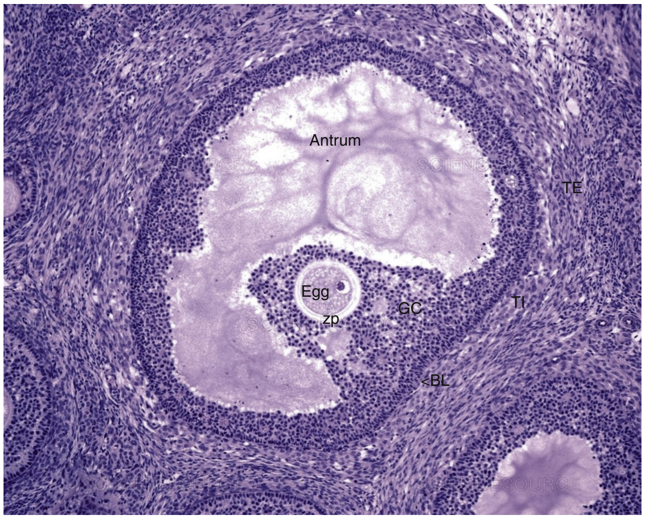


Figure 118.8 **Photomicrograph of a tertiary follicle approximately 400 μm in diameter at the time of cavitation or early antrum formation.** It comprises a fully growing dictyate oocyte (egg) surrounded by a thick zona pellucida; an antrum containing follicular fluid; multiple layers of granulosa cells organized morphologically into recognizable zones; a basal lamina; a theca interna; and a theca externa. *BL,* Basal lamina; *GC,* granulosa cells; *TE,* theca externa; *TI,* theca interna; *zp,* zona pellucida. (From Bloom W, Fawcett DW, eds. *A textbook of histology.* Philadelphia: Saunders; 1975:869.)

Structure of the Graafian Follicle. A healthy Graafian follicle is a multilayered spherical unit that comprises sheets of precisely positioned cells (Figs. 118.8–118.10). The theca externa is innervated by autonomic nerves that play an active role in their contraction.[69] The full functional significance of the theca externa is not known, but regulated contractions of these cells have been implicated in both ovulation and atresia.[69] For example, a role for local release of endothelin-2 by the theca externa in vasoconstriction and ovulation has been reported.[72,73]

The number of differentiated theca interstitial cells that constitute the *theca interna* increases progressively, and their endocrine function is reflected in the large capillary plexus. It is not uncommon to have five to eight layers of theca interstitial cells in a Graafian follicle.[1]

Internal to the theca is a basal lamina that functions as a barrier to vascular tissue. Encircled by the basal lamina are multiple layers of granulosa cells that appear heterogeneous by virtue of their position within the follicle wall (see Figs. 118.9 and 118.10). The position and morphology of the granulosa cells identify at least three different domains: (1) an outer or mural domain (also called membrane) that comprises a pseudostratified epithelium, with cells that make contact with the basal lamina; (2) an inner, periantral or luminal domain lining the antrum; (3) the cumulus domain, which surrounds the oocyte and makes contact with the mural and periantral cells; (see Figs. 118.9 and Fig. 118.10). Electron microscopic reconstruction of granulosa cells has confirmed that granulosa cells have complex elongated shapes with long protrusions that allow inner cells to make contact with the basal lamina.[74] By virtue of their position, the granulosa cells assume different phenotypes, express different genes, and have different functions in response to FSH stimulation.[35,75] For example, the cumulus, and periantral granulosa cells (but not the mural cells) continue to divide throughout Graafian follicle development. By contrast, the cells in the mural domain stop dividing and become highly responsive to FSH and LH (see Fig. 118.9). Although all granulosa cells layers contain similar amount of FSH receptors, cumulus cells express low levels of LH receptors and steroidogenic enzymes. These divergent phenotypes and functions of granulosa cells are generated by two opposing gradients of diffusible signals, one emanating from the theca layers of the follicle and/or the vasculature and the opposing one by the oocyte through secretion of diffusible bioactive molecules.[35,75] In particular, it is well established that oocyte-secreted GDF-9 and BMP-15 play an important role in inducing the unique phenotypic properties of the surrounding cumulus cells.

Chronology of Follicle Development. Folliculogenesis in women is a very lengthy process whose exact duration is uncertain.[11,76,77] In normal cycling women, the dominant follicle that will ovulate originates from a primordial follicle that was recruited several months earlier

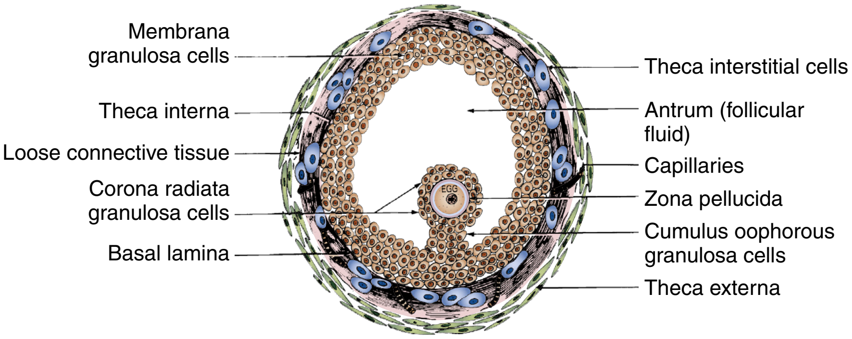


Figure 118.9 **Diagram of the cross-sectional appearance of a typical healthy Graafian follicle showing the organization of its various cell types.** (From Erickson GF. Primary cultures of ovarian cells in serum-free medium as models of hormone-dependent differentiation. *Mol Cell Endocrinol.* 1983;29:21–49.)

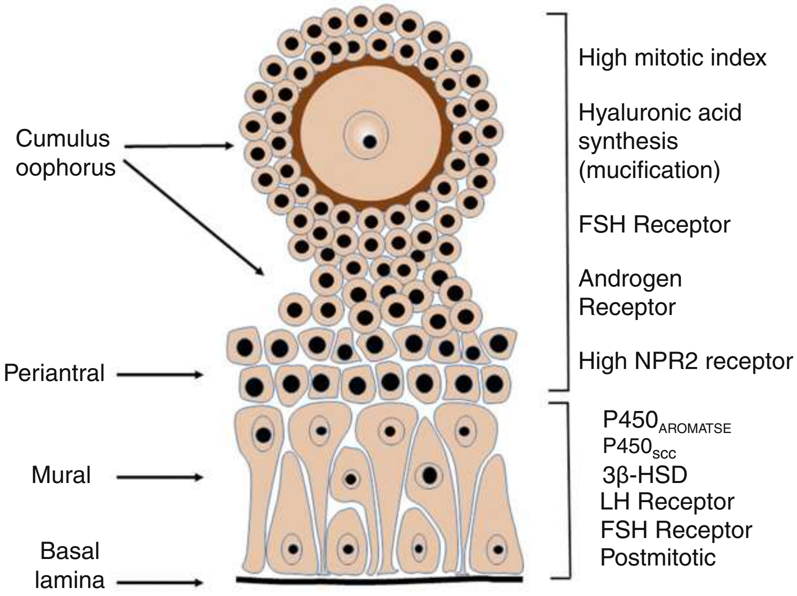


Figure 118.10 **Diagram of the structure-function heterogeneity of the granulosa cells in a healthy Graafian follicle.** By virtue of their position in the follicle wall, the granulosa cells become different from one another, and this is reflected in different patterns of proliferation and differentiation. *FSH,* Follicle-stimulating hormone; *LH,* luteinizing hormone. (Modified from Erickson GF. The graafian follicle: a functional definition. In: Adashi EY, ed. *Ovulation: evolving scientific and clinical concepts.* New York: Springer-Verlag; 2000.)

(Fig. 118.11). While the primordial to primary and secondary follicle transition proceeds very slowly—the growth of a secondary follicle to ovulation is well described and takes approximately 85 days (see Fig. 118.11). The final phase of growth (*selection* of dominant follicle) takes approximately 10 days, The slow nature of the development of preantral follicles results from the long doubling time (≈250 hours) of granulosa cell replication. The rate of follicle growth increases dramatically when follicular fluid begins to accumulate and the healthy Graafian follicle passes through the gonadotropin-dependent or antral phase.

Recruitment and Selection of the Dominant Follicle. Selection of the dominant follicle is a critical event in the ovarian cycle. In normal women, it appears that the dominant follicle is selected from a cohort of small Graafian follicles (4.7 ± 0.7 mm) that are recruited during the late luteal phase of the antecedent cycle (see Fig. 118.11).[11,76,77] On the basis of ultrasound observation, it has been proposed that two or three major waves of follicular recruitment emerge during every menstrual cycle.[78] In addition, minor waves of antral follicle may also be present. The dominant follicle will emerge from one of these waves. With the emergence of the dominant follicle, the remaining cohort of recruited follicles begins to degenerate and undergo atresia. While atresia can occur at any stage, it is most frequent in small and medium Graafian follicles (see Fig. 118.11).

Function of Granulosa Cells of the Dominant Follicle. As described above, the granulosa cells of a dominant follicle are characterized by a sharp increase (~2-fold) in the rate of mitosis.[11,76,77] Indeed, the first visible sign of follicular selection is that granulosa cell division continues at a rapid rate in the chosen follicle, whereas the mitotic rate becomes low in subordinate, non-dominant follicles.[11,76] Because this distinction becomes apparent in the late luteal phase of the cycle antecedent to ovulation, it is argued that selection takes place at this stage (Fig. 118.12). The dominant follicle grows rapidly, reaching an average size of 6.9 ± 0.5 mm at days 1 to 5, 13.7 ± 1.2 mm at days 6 to 10, and 18.8 ± 0.5 mm at days 11 to 14 of the cycle.

A fundamental question is how selection of the dominant follicle occurs. Responsiveness to FSH and the estrogen-FSH feedback in the pituitary are likely critical to establish dominance. According to the FSH threshold/window/gate hypothesis, the increase in plasma FSH occurring at the end of the luteal phase of the preceding cycle is obligatory for, and serves as the basis of, dominant follicle selection.[79] In women, plasma FSH begins to rise at about the same time that plasma P_4 and E_2 levels fall to baseline levels at the end of the luteal phase (Fig. 118.13), concomitantly with decreases in inhibin A emanating from the corpus luteum (see Fig. 118.13). The inverse relationship between plasma FSH and inhibin A (a known inhibitor of pituitary FSH secretion) suggests an endocrine role for inhibin A in regulation of the late luteal FSH rise. By contrast, inhibin B increases before menses coincident with the FSH rise. It is possible that elevated inhibin B levels emanating from the growing follicles play a role in suppressing the FSH rise during the follicular phase (see Fig. 118.13). Given the estrogen (and/or inhibin) feedback in the pituitary, the ensuing decrease in FSH below a threshold at the appropriate time prevents development of subordinate follicles. Continuous growth of a dominant follicle on a background of diminishing FSH may be due to the increased sensitivity to LH.

One effect of the late-luteal FSH rise is a progressive increase in the concentration of FSH in follicular fluid of the dominant follicle. It is noteworthy that FSH levels become low or non-detectable in follicular fluid of non-dominant follicles.[80,81] How a dominant follicle sequesters FSH is not known, even though high concentrations of FSH receptor may play a role. Nonetheless, when the concentration of FSH in the microenvironment reaches threshold levels, dominance is established, and the chosen follicle becomes committed to developing along the ovulatory pathway. In the absence of threshold levels of FSH, the mitotic rate falls to low levels, and non-dominant cohort of follicles proceed along the atretic pathway. Treating women with exogenous FSH stimulates mitosis of granulosa cells in non-dominant follicles,[76] rescuing non-dominant follicles from atresia. This explains how exogenous FSH provokes multiple ovulatory follicles in women undergoing ovulation induction without depleting OR.

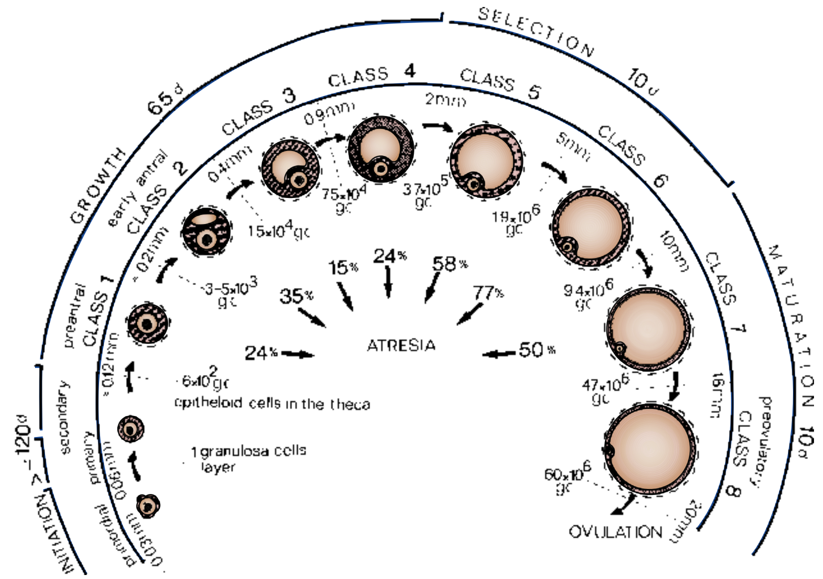

INITIATION >120d
GROWTH 65d
SELECTION 10d
MATURATION 10d
primordial 0.03mm
primary 0.06mm
secondary 0.12mm
preantral CLASS 1 0.2mm
early antral CLASS 2 0.4mm
CLASS 3 0.9mm
CLASS 4 2mm
CLASS 5 5mm
CLASS 6 10mm
CLASS 7 16mm
preovulatory CLASS 8 20mm
1 granulosa cells layer
epitheloid cells in the theca
6×10² gc
3-5×10³ gc
15×10⁴ gc
75×10⁴ gc
37×10⁵ gc
19×10⁶ gc
9.4×10⁶ gc
47×10⁶ gc
60×10⁶ gc
ATRESIA
24%
35%
15%
24%
58%
77%
50%
OVULATION

Figure 118.11 **The chronology of folliculogenesis in normal human ovaries.** The steps from primordial follicle recruitment, dominant follicle selection, and ovulation can be seen flanked by a timeline. The follicles are labeled to indicate developmental stage, size, and the percentage in atresia. (From Gougeon A. Dynamics of follicular growth in the human: a model from preliminary results. *Hum Reprod.* 1986;1:81–87.)

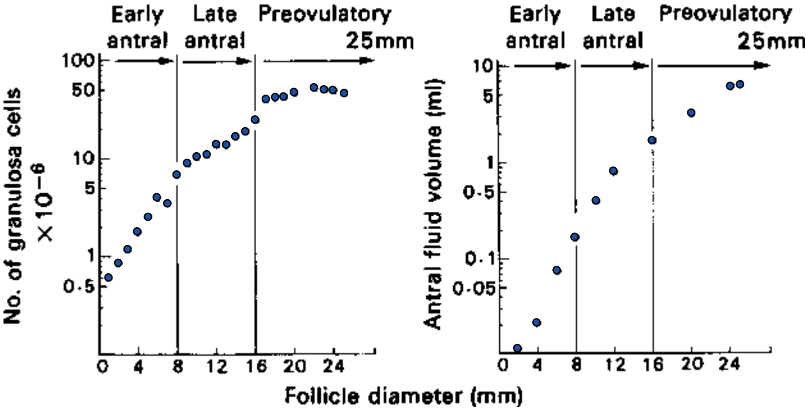


Figure 118.12 **Changes in the number of granulosa cells *(left panel)* and the volume of follicular fluid (right panel) during the growth (diameter) of a healthy (dominant) human Graafian follicle.** (From McNatty KP. Hormonal correlates of follicular development in the human ovary. *Aust J Biol Sci.* 1981;34:249–268.)

As development of the dominant follicle proceeds, granulosa and TCs show a high continual proliferative capacity (see Fig. 118.12).[11,76] In addition to FSH, FGF and EGF are potent stimulators of mitosis by mechanisms independent of FSH.[82] While the numbers of both theca interna and theca externa cells increase concurrently with granulosa cells,[83] mechanisms that regulate their proliferation are virtually unknown.

An important event in the selection of the dominant follicle is the ability of FSH to elicit massive increases in P450arom gene expression in the granulosa cells. Expression of this steroidogenic enzyme results in progressive increases in production of E_2, which, in turn, promotes endometrial growth and the preovulatory surge of LH at midcycle.

The production of E_2 in the follicle requires the coordinated function of theca and granulosa cells, as well as LH and FSH. These integrated functions are collectively summarized in the two-cell, two-gonadotropin model.[69] The theca interstitial cells produce androstenedione (aromatase substrate) in response to LH stimulation.[69] When LH interacts with its transmembrane receptor, the binding event is transduced into an intracellular signal by means of G proteins. The LH-activated receptor is coupled to the trimeric (α, β, γ subunits) stimulatory guanylyl nucleotide binding protein (Gs) adenylate cyclase, cAMP, protein kinase A (PKA) signal transduction pathway, which, in turn, recoupled to the differential activation of genes encoding steroidogenic enzymes. These regulatory proteins include steroidogenic acute regulatory protein (StAR), cytochrome P450 side-chain cleavage (P450c22), 3β-hydroxysteroid dehydrogenase (3β-HSD), and P450 17α-hydroxylase and 17,20-lyase (P450c17). The end result is the production of androstenedione, which accumulates at very high concentrations in the follicular fluid. It is noteworthy that the androstenedione synthetic activity of the theca interstitial cells can be markedly stimulated and enhanced by other regulatory agents, including insulin, IGF-1, low-density lipoprotein (LDL), and high-density lipoprotein (HDL).[1,69] Although the functional significance of these regulatory molecules is poorly understood, it is notable that hyperinsulinemia in women with PCOS is marked by hyperandrogenism and infertility.[84] Thus, although LH is clearly the most important stimulator of follicular androstenedione production, other proteins can affect changes in theca androgen production.

Although detected in some cells outside of the ovary, the granulosa cells are the major site of expression of FSH receptors. Similar to the LH mode

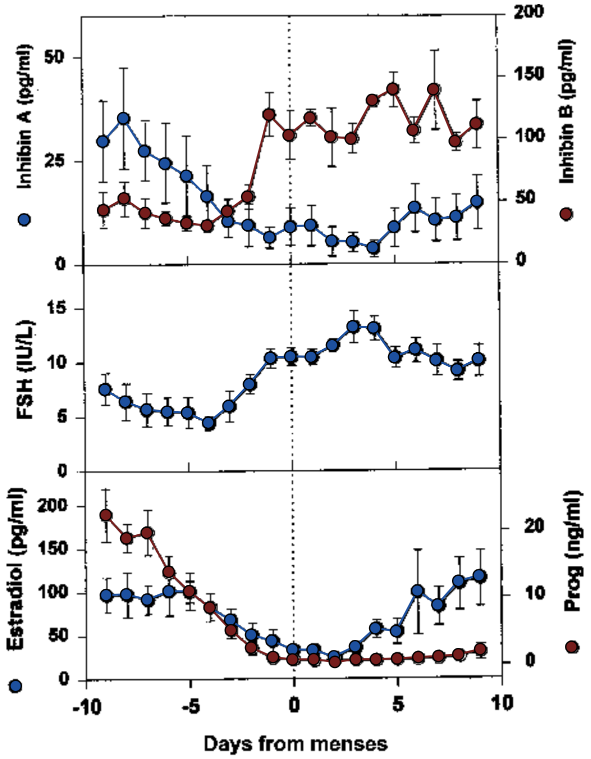


Figure 118.13 **The rise in plasma follicle-stimulating hormone (FSH) as it occurs in normally cycling women during the luteal-follicular transition.** Data are mean ± standard error of the mean (SEM) of daily FSH, estradiol, progesterone, inhibin A, and inhibin B levels in plasma of five normal cycling women. Data are centered to the day of menses. (From Welt CK, Martin KA, Taylor AE, et al. Frequency modulation of follicle-stimulating hormone [FSH] during the luteal-follicular transition: evidence for FSH control of inhibin B in normal women. *J Clin Endocrinol Metab.* 1997;82:2645–2652; © 1997, The Endocrine Society.)

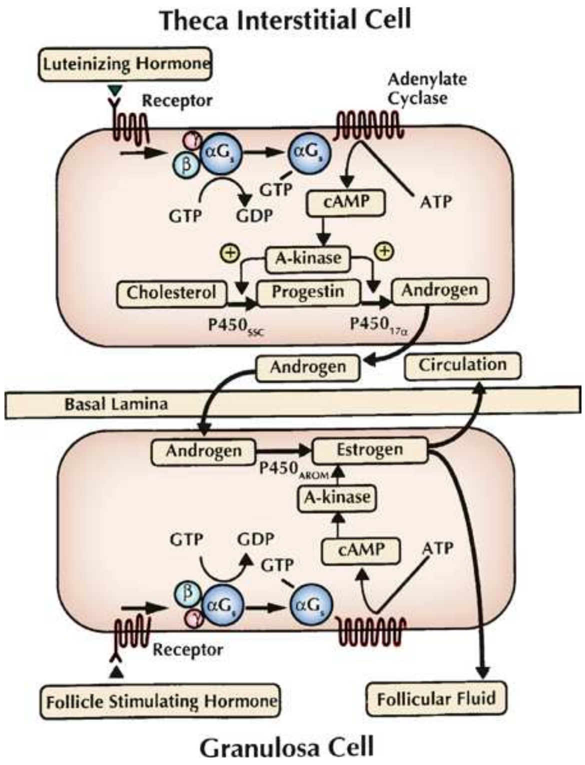


Figure 118.14 **Diagram of the two-cell, two-gonadotropin concept for follicular estradiol biosynthesis.** G proteins include GSα, β, γ; A kinase or cyclic adenosine monophosphate (cAMP)-dependent protein kinase A. *ATP,* Adenosine triphosphate; *GDP,* guanosine diphosphate; *GTP,* guanosine triphosphate. (From Kettel LM, Erickson GF. Basic and clinical concepts of ovulation induction. In: Rock J, Alverez-Murphy A, eds. *Advances in obstetrics and gynecology.* St Louis: Mosby; 1994.)

of action, FSH binds to a cell-surface transmembrane receptor, which activates a stimulatory G protein, the adenylate cyclase, cAMP production, and the PKA signaling pathway.[85] Aside from cAMP, recent data show that several additional signaling pathways mediate the FSH effects on granulosa cells via activation of other G proteins.[86] Activation of these multiple FSH-dependent signaling mechanisms (Fig. 118.14) leads to the expression of a large array of genes including those that control the potential for E_2 production, namely, expression of the P450 aromatase *(P450arom87)* and the type 1 17α-hydroxysteroid dehydrogenase *(17α-HSD)* genes.[88,89] By virtue of the expression of these key steroidogenic enzymes, the granulosa cells have the capacity to convert androstenedione to testosterone, which then is aromatized to E_2 by *P450arom.* The temporal pattern and level of expression of *P450arom* enzyme determine the time course and the amount of E_2 produced by the dominant follicle during the follicular phase.

FSH induces additional receptors that are critical for follicle development and ovulation. In particular, FSH induces expression of LH receptors in granulosa cells. As with P450arom, only mural granulosa cells in the dominant follicle express the LH receptor gene; however, in contrast to the relatively early expression of P450arom in granulosa, LH receptors are not expressed until late in the follicular phase of the cycle.[62,90] LH receptors provide the granulosa cells in the dominant follicle with the capacity to respond to the midcycle surge of LH and for the follicle to undergo ovulation. Furthermore, LH signaling plays a fundamental role in granulosa cells by inducing expression of P_4 receptors.[87,91] Importantly, P_4 receptors play an obligatory role in the physiological process of ovulation.[91] Finally, FSH induces the expression of the natriuretic peptide receptor 2 (NPR2) and of the cognate ligand natriuretic peptide precursor type C (NPPC, also known as C-type natriuretic peptide [CNP]).[92] This ligand/receptor pair is responsible for the accumulation of cGMP in the follicle required to maintain oocyte meiotic arrest therein (see Fig. 118.15).[93]

Oocyte Meiotic Arrest in the Antral Follicle. In a fully developed Graafian follicle, the oocyte has acquired competence to undergo meiotic maturation. However, it is held in a meiotic arrest at the dictyate stage by signals originating from the granulosa cells. Seminal work in rodents and humans has demonstrated that removal of the oocyte from the follicular environment causes meiotic resumption, suggesting the presence of meiotic inhibitory factors in the follicular environment.[94,95] Although a low-molecular-weight oocyte maturation inhibitor had been isolated from the follicular fluid and granulosa cell extracts,[96] the nature of these factors has remained controversial for several decades. However, recent observations in rodents indicate that an interplay of two cyclic nucleotides, cAMP, and cGMP, as well as the cyclic nucleotide phosphodiesterase (PDE3A) expressed in the mature oocyte, comprise the primary mechanism to maintain meiotic arrest (see Fig. 118.15).[97,98] PDE3A degrades both cyclic nucleotides, with cGMP functioning as an inhibitor of the enzyme activity.[99] The current model posits that FSH and E_2 induce synthesis of CNP in granulosa cells. CNP interacts and activates the cognate NPR2 receptor on the same granulosa cells, resulting in accumulation of cGMP.[93] In turn, cGMP diffuses via gap junctions into oocytes where it inhibits oocyte PDE3A activity. The net result is cAMP levels prevent reentry into the meiotic cell cycle[97,98] (see Fig. 118.15). This regulatory loop is functional in all species studied, including humans.

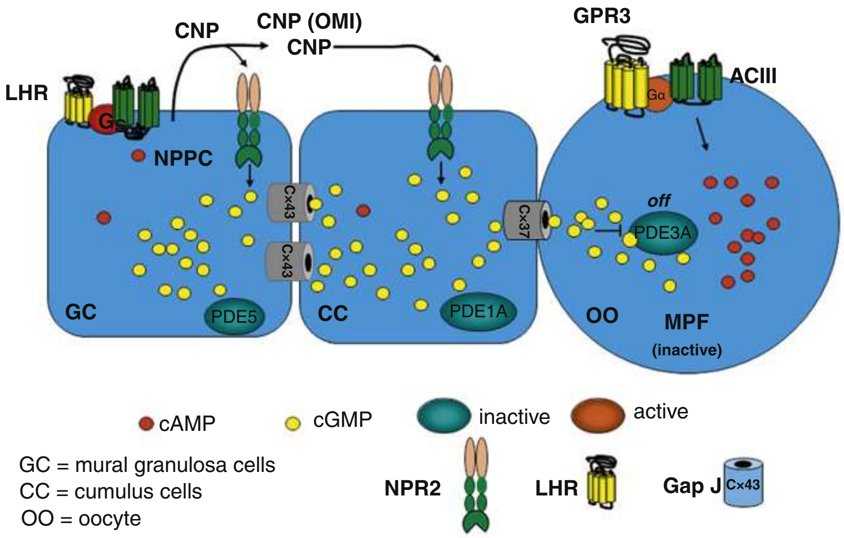


Figure 118.15 **Mechanisms maintaining oocyte meiotic arrest in the antral follicle. Schematic representation of a granulosa cell (GC), a cumulus cell (CC), and an oocyte (OO).** Communication among cells is mediated by connexins (Cx43 and Cx37). Mural granulosa cells produce the NPPC precursor which is then proteolitically processed and secreted as active C natriuretic peptide (CNP). CNP activates the cognate receptor NPR2 activating production of cGMP (yellow symbols). Cyclic GMP diffuses through gap junctions to the oocyte where it inhibits the activity of the oocyte phosphodiesterase PDE3A. In the oocyte, cAMP produced by the GPR3 receptor coupled to adenylyl cyclase ACIII prevents MPF activation and meiotic progression

Atresia. Of the nearly 1 million (400,000 to 2 million) or so primordial follicles present in the ovaries at birth (see Fig. 118.3), about 400 will be ovulated during the reproductive life span. Thus, 99.9% of oocytes are eliminated by apoptosis, which is governed by activation of a genetically controlled cell suicide program.[100] Studies over the past decade have highlighted pathways for directing the expression of proteins involved in promoting and suppressing apoptosis in the ovary.[101,102] The current challenges are to identify how specific atretogenic ligands promote apoptosis and to explain how these interactions are integrated into the overall pattern of folliculogenesis during the cycle. Notably, the ability of high concentrations of FSH to inhibit apoptosis in rodent granulosa cells[103] supports the concept that apoptosis in non-dominant follicles may be causally connected to subthreshold levels of FSH in the microenvironment. Implicit therein is the concept that FSH not only exerts an inductive effect on granulosa cell proliferation and differentiation but also acts as a survival factor through its ability to inhibit the activation of apoptosis. Understanding the nature of FSH signaling pathways that control granulosa apoptosis and follicle atresia is a major goal of reproductive research.

OVULATION

KEY POINTS

- Events associated with ovulation are similar to an inflammatory response.
- Even though LH and FSH share similar mechanisms of action, ovulation is mainly triggered by the gonadotropin LH.
- Intracellular regulation of cyclic nucleotide signaling is only one of the many pathways regulated by gonadotropins.
- This initial signal branches into regulation of a myriad of additional pathways with different biological outcomes.
- Temporal and spatial propagation of these pathways are additional dimensions required to maintain the phenotypic properties of granulosa cells within the follicle.

Thirty-six hours following the LH surge (on or about the 15th day of a 28-day cycle), the preovulatory follicle breaches the surface of the ovary and releases the egg/cumulus complex through a process termed *ovulation* (Fig. 118.16). Normal ovulation requires the coordinated actions of LH, FSH, and P_4, as well coordinates changes in the somatic and germ cell compartments, driven by release of local factors. Given the properties of the signals exchanged during ovulation, this process shares many similarities with an inflammatory response.

Oocyte Maturation Following the LH Surge

At the time of ovulation, oocytes undergo resumption of meiosis. At meiotic reentry, the oocyte dissolves the nuclear membrane (germinal vesicle break down) and extrudes the first polar body; it then arrests in metaphase II. The meiotic process proceeds no farther. The second meiotic division will be completed only when the ovulated egg is fertilized. This critically important transition into the cycle is triggered by the preovulatory surge of LH (Fig. 118.17). The completion of meiosis is termed nuclear maturation and is required to generate a haploid oocyte or "mature egg." A decrease in cGMP diffusing from the somatic compartment is regarded as the primary signal for meiotic reentry: the decrease in cGMP causes an activation of PDE3A and a decrease in cAMP in the oocyte.[99] This in turn causes activation of the meiotic cell cycle machinery.

In addition to nuclear maturation, a second set of less-well characterized changes occurs in the cytoplasm of the oocyte. These complex changes are called "cytoplasmic maturation," which endows the oocyte with the competence to support embryonic development. It is associated with changes in organelle distribution within the oocyte, alteration in membrane and cytoskeleton, and by a complex regulation of gene expression without *de novo* mRNA synthesis. However, selective translation and repression/destabilization of subsets of maternal mRNA will result in oocyte acquisition of developmental competence.[104]

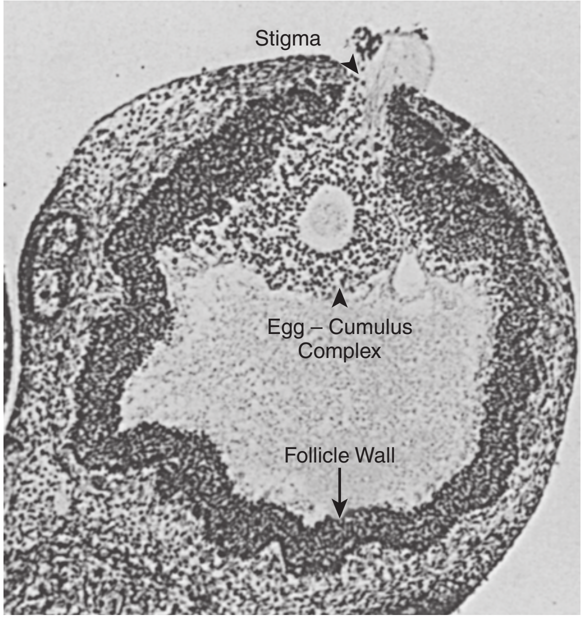


Figure 118.16 Photomicrograph of ovulation *in situ*. (From Hartman CG, Leathem JH. Oogenesis and ovulation. In: *Conference on physiological mechanisms concerned with conception*. New York: Pergamon; 1959.)

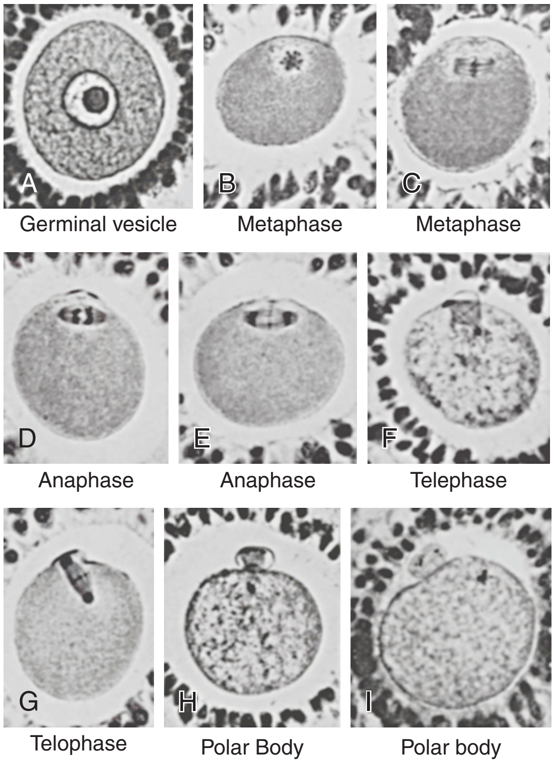


Figure 118.17 Photomicrograph showing the stages of meiotic maturation of the oocyte following resumption of meiosis in the dominant follicle. (From Witschi E. *Development of vertebrates*. Philadelphia: Saunders; 1956.)

Events Triggered by LH in the Somatic Compartment of the Follicle.

The LH surge induces major changes in all granulosa cell compartments of the follicle via modulation of gene expression, steroidogenesis, secretion, and cell-cell communication. Recent observations provide an explanation of how the LH signal propagates from the mural granulosa cells (rich in LH receptors) to the cumulus cells (with LH receptor almost undetectable). In all species studied, the LH surge causes transient increases in the expression of EGF-like growth factors (amphiregulin, epiregulin, betacellulin, and neuregulin) in mural granulosa cells[105] (Fig. 118.18). These growth factors interact with their cognate receptors (EGF receptors) present in all granulosa cells of the follicle, although highly enriched in cumulus cells. In mural granulosa cells, EGF-like growth factors function as autocrine factors amplifying signals originating from the LH receptors. On the other hand, they serve as paracrine factors for the cells that do not express LH receptors. In this case, as in cumulus cells, they function to propagate the LH signal toward the center of the follicle. As the net result of the combined LH/growth factor action, decreased cGMP in the follicle and decreased GAP junction permeability cause a decrease in cGMP in the oocyte and meiotic cell cycle reentry.

Concomitantly with oocyte reentry into the cell cycle, the cumulus granulosa cells undergo a series of structural and functional changes called *mucification* (see Fig. 118.10). In response to the preovulatory LH surge, the cumulus cells secrete large quantities of a glycoproteic mucous substance (primarily hyaluronic acid) into the extracellular space. This accumulation results in the dispersal of cumulus cells and causes the egg/cumulus complex to expand 3- to 5-fold.[86,106] The process of mucification is critical for the transport and fertilization of the egg within the fallopian tube. Mucification is stimulated by the release of multiple factors, including the EGF-like growth factors and prostaglandins. The actual release of the egg/cumulus complex is intimately coupled to the production of proteases[107] that degrade the ovarian tissues in a small, highly localized area called the macula pellucida or stigma. Animal studies have demonstrated that LH-stimulated P_4 and prostaglandin production by the follicle wall play an obligatory role in the expression of the proteolytic activities that occur during ovulation. Indeed, female mice lacking either the P_4 receptor[108] or the prostaglandin synthase gene (cyclooxygenase [COX-2],[109] the first rate-limiting enzyme in the biosynthesis of prostaglandins),[110] fail to ovulate and are infertile. Similarly, histological studies of rodents treated with indomethacin (an inhibitor of prostaglandin production) show that LH-induced prostaglandins synthesis is critical for stigma formation.

The LH surge also induces major changes in gene expression in TCs in preparation for ovulation. These will direct changes in the vasculature and *de novo* angiogenesis in preparation for vasculature invasion of the granulosa cell layer. Recruitment of immune cells at this stage is also essential for ovulation and vasculogenesis.

On the basis of the above changes, the following model for follicle rupture has been proposed (Fig. 118.19): the midcycle LH surge stimulates production of P_4; the P_4 ligand interacts with its receptor in the follicle cells, which, in turn, induce prostaglandin synthase and prostaglandin production. Prostaglandins interact with specific receptors in the surface epithelial cells of the presumptive stigma; prostaglandin receptor signaling causes release of lysosomal (proteolytic) enzymes that degrade the underlying tissue, ultimately resulting in stigma formation and follicle rupture.

LUTEOGENESIS

KEY POINTS

- The development of a transient endocrine structure, the corpus luteum, follows ovulation.
- Coordinated production of estrogen and P_4 by the corpus luteum controls the cyclic changes of the reproductive tract and the pituitary.

After ovulation, the follicle transforms into a novel endocrine organ, the corpus luteum (see Figs. 118.16 and 118.20). The theca interna cells and granulosa cells become the theca lutein and granulosa lutein cells, respectively. The corpus luteum becomes richly vascularized and serves as the site of P_4 and E_2 production during the luteal phase of the cycle. Major changes that occur during the life and death of the

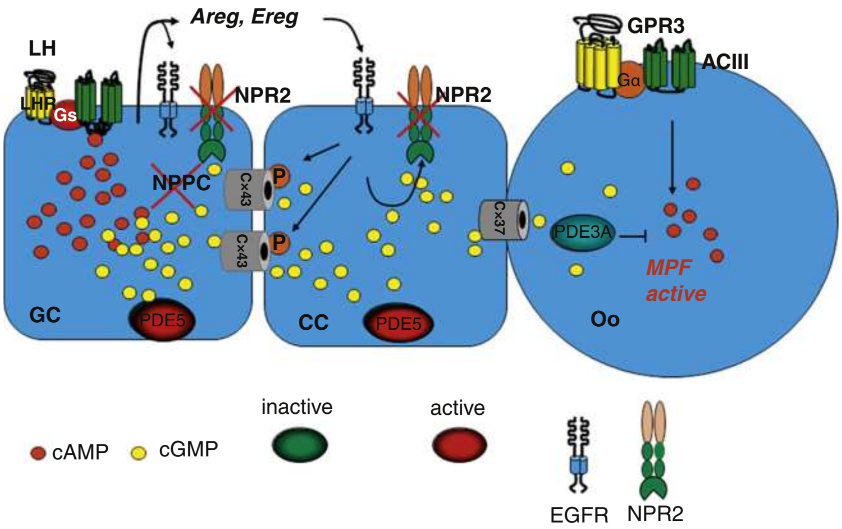

Areg, Ereg
LH
LHR
Gs
NPR2
NPR2
GPR3
Gα
ACIII
NPPC
Cx43
P
Cx43
P
Cx37
PDE3A
MPF active
GC
PDE5
CC
PDE5
Oo
inactive
active
cAMP
cGMP
EGFR
NPR2

Figure 118.18 **Mechanisms of LH-induced oocyte maturation.** In this scheme, symbols are as in Fig. 118.15. LH interaction with the cognate receptor on mural granulosa cells activates adenylyl cyclase to produce cAMP. cAMP accumulation induces the synthesis of EGF-like growth factors (Areg, Ereg). These growth factors act in an autocrine fashion by activating the EGF receptor. As a consequence of the LH stimulation of cAMP activation of EGF receptors causes the inactivation NPR2 and phosphorylation and closure of GAP junctions. As a consequence, cGMP decrease throughout the follicle. In the oocyte, inhibition of PDE3A is relieved and cAMP is degraded leading to activation of MPF and induction of meiotic maturation. *Areg,* amphiregulin; *CC,* cumulus cells; *Cx43,* connexin 43; *Ereg,* epiregulin; *GC,* granulosa cells; *GPR3,* G protein–coupled receptor 3; *MPF,* metaphase-promoting factor; *NPPC,* natriuretic peptide precursor C; *NPR2,* natriuretic peptide receptor 2; *Oo,* oocyte; *PDE3A,* phosphodiesterase type 3A; *PDE5,* phosphodiesterase type 5.

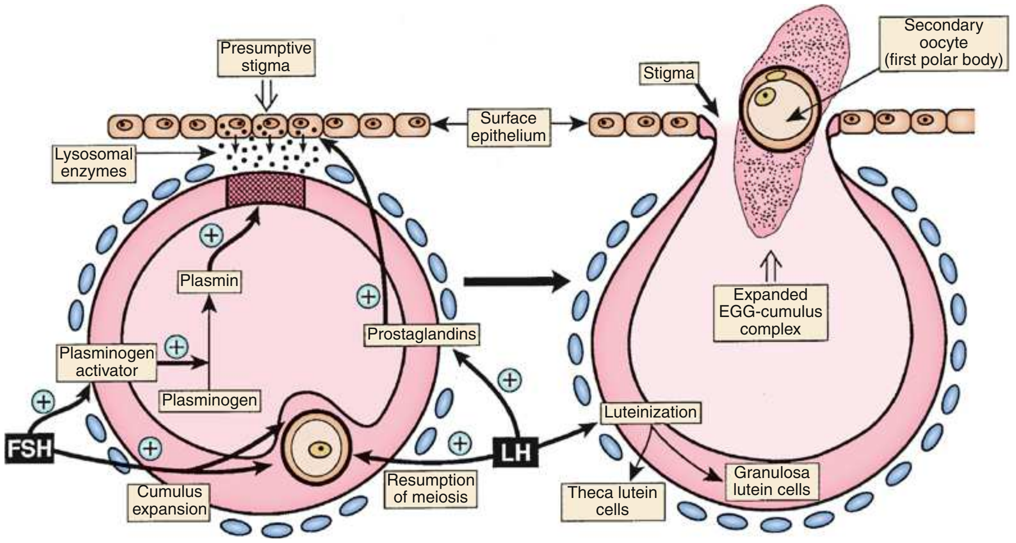

Presumptive stigma
Stigma
Secondary oocyte (first polar body)
Lysosomal enzymes
Surface epithelium
Plasmin
Expanded EGG-cumulus complex
Prostaglandins
Plasminogen activator
Plasminogen
Luteinization
FSH
LH
Cumulus expansion
Resumption of meiosis
Theca lutein cells
Granulosa lutein cells
+

Figure 118.19 **A model illustrating the cellular activities involved in gonadotropin (follicle-stimulating hormone [FSH] and luteinizing hormone [LH])-induced ovulation.** (From Erickson GF. The ovary: basic principles and concepts. In: Felig P, Baxter JD, Frohman LA, eds. *Endocrinology and metabolism,* 3rd ed. New York: McGraw Hill; 1995:973–1015.)

corpus luteum constitute the processes called *luteogenesis* and *luteolysis*, respectively. Luteinization begins shortly after the preovulatory LH surge, and full expression of the differentiated state of the lutein cells begins by the end of 1 week (day 21 or 22 of a 28-day menstrual cycle). If the egg is not fertilized, the corpus luteum undergoes apoptosis,[111] leading to luteolysis. The process of luteogenesis is reflected in the production of P_4, E_2, and inhibin A during the luteal phase of the cycle.

Luteinization

LH is a central regulator of luteinization. In response to the preovulatory LH surge, genes that encode StAR, P450c22, 3β-HSD, and P450arom are expressed at high levels in the granulosa lutein cells, leading to the production of large amounts of P_4 and E_2. Although mean plasma LH levels are low during the luteal phase, small amounts of LH are critical for proper corpus luteum function.[112] It should be mentioned that luteinization is almost certainly more complex and most certainly involves other regulatory molecules, including LDL,[113] blood cells, cytokines, and angiogenic factors.[114] If fertilization and embryo implantation occur, the corpus luteum enlarges further to become the corpus luteum of pregnancy, in response to hCG produced by the blastocyst. The corpus luteum of pregnancy persists for about 6 weeks, after which it declines in size and function, as the placenta replaces the corpus luteum and produces steroid hormones for pregnancy maintenance.

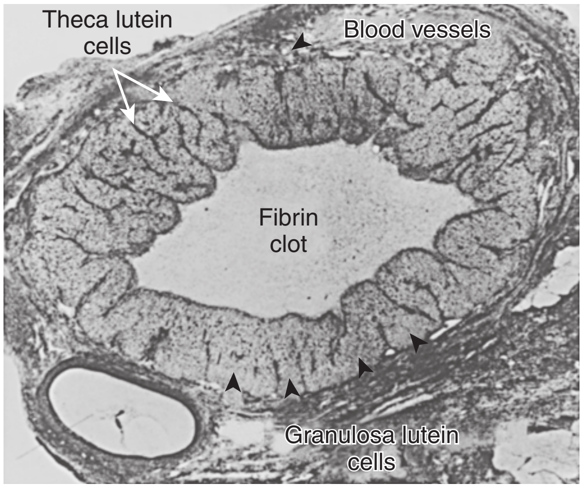


Figure 118.20 Photomicrograph of section of a corpus luteum from the human ovary. (From Bloom W, Fawcett DW, eds. *A textbook of histology*. Philadelphia: Saunders; 1975:875.)

Luteolysis

Luteolysis, or corpus luteum demise, is accompanied by progressive expression of apoptosis in the lutein and vascular cells. During this process, P_4 and E_2 production by the corpus luteum decreases markedly. Despite its great significance to female fertility, the physiological mechanism that governs luteolysis in humans remains unknown. It is possible that prostaglandin $F_2\alpha$ is involved, but its importance in women has yet to be proved.

CONCLUSION

The developmental programs of folliculogenesis, ovulation, and luteogenesis form the basis of the ovarian cycle, which, in turn, generates menstrual cyclicity and female fertility. These programs depend on coordinated regulation of a large array of cellular functions, including mitosis, meiosis, cellular differentiation, growth, and apoptosis. In addition to gonadotropins, recent studies have established the exciting new concept that growth factor-signaling pathways play an obligatory role in specifying ovarian cell fates through local autocrine/paracrine mechanisms. The emerging model is that ovarian growth factors are the controlling modulators that ensure the proper level, timing, and spatial pattern of the developmental programs evoked by hormones, offering a new model for understanding and treating fertility and infertility in women.

REFERENCES

1. Erickson GF, Magoffin DA, Dyer CA, et al. The ovarian androgen producing cells: a review of structure/function relationships. *Endocr Rev. Summer*. 1985;6:371–399.
2. Reynolds SR. The vasculature of ovary and ovarian function. *Omnia Med*. 1950;28:456–457.
3. Schultea TD, Dees WL, Ojeda SR. Postnatal development of sympathetic and sensory innervation of the rhesus monkey ovary. *Biol Reprod*. 1992;47:760–767.
4. Serafica MD, Goto T, Trounson AO. Transcripts from a human primordial follicle cDNA library. *Hum Reprod*. 2005;20:2074–2091.
5. Ohno S, Klinger HP, Atkin NB. Human oogenesis. *Cytogenetics*. 1962;1:42–51.
6. Johnson J, Canning J, Kaneko T, et al. Germline stem cells and follicular renewal in the postnatal mammalian ovary. *Nature*. 2004;428:145–150.
7. White YA, Woods DC, Takai Y, et al. Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. *Nat Med*. 2012;18:413–421.
8. Lei L, Spradling AC. Female mice lack adult germ-line stem cells but sustain oogenesis using stable primordial follicles. *Proc Natl Acad Sci USA*. 2013;110:8585–8590.
9. Zhang ZP, Liang GJ, Zhang XF, et al. Growth of mouse oocytes to maturity from premeiotic germ cells in vitro. *PLoS One*. 2012;7:e41771.
10. Hansen KR, Knowlton NS, Thyer AC, et al. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod*. 2008;23:699–708.
11. Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. *Endocr Rev*. 1996;17:121–155.
12. Oktay K, Briggs D, Gosden RG. Ontogeny of follicle-stimulating hormone receptor gene expression in isolated human ovarian follicles. *J Clin Endocrinol Metab*. 1997;82:3748–3751.
13. Grazul-Bilska AT, Reynolds LP, Redmer DA. Gap junctions in the ovaries. *Biol Reprod*. 1997;57:947–957.
14. Bruzzone R, White TW, Paul DL. Connections with connexins: the molecular basis of direct intercellular signaling. *Eur J Biochem*. 1996;238:1–27.
15. Kumar NM, Gilula NB. The gap junction communication channel. *Cell*. 1996;84:381–388.
16. Simon AM, Goodenough DA, Li E, et al. Female infertility in mice lacking connexin 37. *Nature*. 1997;385:525–529.
17. Brower PT, Schultz RM. Intercellular communication between granulosa cells and mouse oocytes: existence and possible nutritional role during oocyte growth. *Dev Biol*. 1982;90:144–153.
18. Heller DT, Cahill DM, Schultz RM. Biochemical studies of mammalian oogenesis: metabolic cooperativity between granulosa cells and growing mouse oocytes. *Dev Biol*. 1981;84:455–464.
19. Wassarman PM. The biology and chemistry of fertilization. *Science*. 1987;235:553–560.
20. Moos J, Faundes D, Kopf GS, et al. Composition of the human zona pellucida and modifications following fertilization. *Hum Reprod*. 1995;10:2467–2471.
21. Gong X, Dubois DH, Miller DJ. Activation of a G protein complex by aggregation of beta-1,4-galactosyltransferase on the surface of sperm. *Science*. 1995;269:1718–1721.
22. Avella MA, Xiong B, Dean J. The molecular basis of gamete recognition in mice and humans. *Mol Hum Reprod*. 2013;19:279–289.
23. Skinner SM, Prasad SV, Ndolo TM, et al. Zona pellucida antigens: targets for contraceptive vaccines. *Am J Reprod Immunol*. 1996;35:163–174.
24. Tesarik J. Targeting the zona pellucida for immunocontraception: a minireview. *Hum Reprod*. 1995;10(2):132–139.
25. Duleba AJ, Pehlivan T, Carbone R, et al. Activin stimulates proliferation of rat ovarian thecal-interstitial cells. *Biol Reprod*. 2001;65:704–709.
26. Magoffin DA. The ovarian androgen-producing cells: a 2001 perspective. *Rev Endocr Metab Dis*. 2002;3:47–53.
27. Parrott JA, Skinner MK. Kit-ligand/stem cell factor induces primordial follicle development and initiates folliculogenesis. *Endocrinology*. 1999;140:4262–4271.
28. Gelety TJ, Magoffin DA. Ontogeny of steroidogenic enzyme gene expression in ovarian theca-interstitial cells in the rat: regulation by a paracrine theca-differentiating factor prior to achieving luteinizing hormone responsiveness. *Biol Reprod*. 1997;56:938–945.
29. Magoffin DA, Magarelli PC. Preantral follicles stimulate luteinizing hormone independent differentiation of ovarian theca-interstitial cells by an intrafollicular paracrine mechanism. *Endocrine*. 1995;3:107–112.
30. Chang H, Brown CW, Matzuk MM. Genetic analysis of the mammalian transforming growth factor-beta superfamily. *Endocr Rev*. 2002;23:787–823.
31. Monget P, Fabre S, Mulsant P, et al. Regulation of ovarian folliculogenesis by IGF and BMP system in domestic animals. *Domest Anim Endocrinol*. 2002;23:139–154.
32. Poretsky L, Cataldo NA, Rosenwaks Z, et al. The insulin-related ovarian regulatory system in health and disease. *Endocr Rev*. 1999;20:535–582.
33. Matzuk MM, Kumar TR, Shou W, et al. Transgenic models to study the roles of inhibins and activins in reproduction, oncogenesis, and development. *Recent Prog Horm Res*. 1996;51:123–154; discussion 155–127.

34. Conti M, Hsieh M, Park JY, et al. Role of the epidermal growth factor network in ovarian follicles. *Mol Endocrinol*. 2006;20:715–723.
35. Shimasaki S, Moore RK, Otsuka F, et al. The bone morphogenetic protein system in mammalian reproduction. *Endocr Rev*. 2004;25:72–101.
36. Juengel JL, Hudson NL, Heath DA, et al. Growth differentiation factor 9 and bone morphogenetic protein 15 are essential for ovarian follicular development in sheep. *Biol Reprod*. 2002;67:1777–1789.
37. Teixeira Filho FL, Baracat EC, Lee TH, et al. Aberrant expression of growth differentiation factor-9 in oocytes of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2002;87:1337–1344.
38. Hsueh AJ, Kawamura K, Cheng Y, et al. Intraovarian control of early folliculogenesis. *Endocr Rev*. 2015;36:1–24.
39. Fortune JE, Cushman RA, Wahl CM, et al. The primordial to primary follicle transition. *Mol Cell Endocrinol*. 2000;163:53–60.
40. McGee EA, Hsueh AJ. Initial and cyclic recruitment of ovarian follicles. *Endocr Rev*. 2000;21:200–214.
41. Smitz JE, Cortvrindt RG. The earliest stages of folliculogenesis in vitro. *Reproduction*. 2002;123:185–202.
42. Vainio S, Heikkila M, Kispert A, et al. Female development in mammals is regulated by Wnt-4 signalling. *Nature*. 1999;397:405–409.
43. Vanorny DA, Prasasya RD, Chalpe AJ, et al. Notch signaling regulates ovarian follicle formation and coordinates follicular growth. *Mol Endocrinol*. 2014;28:499–511.
44. Edson MA, Nagaraja AK, Matzuk MM. The mammalian ovary from genesis to revelation. *Endocr Rev*. 2009;30:624–712.
45. Rajareddy S, Reddy P, Du C, et al. p27kip1 (cyclin-dependent kinase inhibitor 1B) controls ovarian development by suppressing follicle endowment and activation and promoting follicle atresia in mice. *Mol Endocrinol*. 2007;21:2189–2202.
46. Castrillon DH, Miao L, Kollipara R, et al. Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. *Science*. 2003;301:215–218.
47. Reddy P, Liu L, Adhikari D, et al. Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. *Science*. 2008;319:611–613.
48. Adhikari D, Zheng W, Shen Y, et al. Tsc/mTORC1 signaling in oocytes governs the quiescence and activation of primordial follicles. *Hum Mol Genet*. 2010;19:397–410.
49. Kezele PR, Nilsson EE, Skinner MK. Insulin but not insulin-like growth factor-1 promotes the primordial to primary follicle transition. *Mol Cell Endocrinol*. 2002;192:37–43.
50. Vendola K, Zhou J, Wang J, et al. Androgens promote oocyte insulin-like growth factor I expression and initiation of follicle development in the primate ovary. *Biol Reprod*. 1999;61:353–357.
51. Galloway SM, McNatty KP, Cambridge LM, et al. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat Genet*. 2000;25:279–283.
52. Lee WS, Otsuka F, Moore RK, et al. Effect of bone morphogenetic protein-7 on folliculogenesis and ovulation in the rat. *Biol Reprod*. 2001;65:994–999.
53. Dong J, Albertini DF, Nishimori K, et al. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature*. 1996;383:531–535.
54. Lintern-Moore S, Everitt AV. The effect of restricted food intake on the size and composition of the ovarian follicle population in the Wistar rat. *Biol Reprod*. 1978;19:688–691.
55. Yan C, Wang P, DeMayo J, et al. Synergistic roles of bone morphogenetic protein 15 and growth differentiation factor 9 in ovarian function. *Mol Endocrinol*. 2001;15:854–866.
56. Di Pasquale E, Rossetti R, Marozzi A, et al. Identification of new variants of human BMP15 gene in a large cohort of women with premature ovarian failure. *J Clin Endocrinol Metab*. 2006;91:1976–1979.
57. Dixit H, Rao LK, Padmalatha VV, et al. Missense mutations in the BMP15 gene are associated with ovarian failure. *Hum Genet*. 2006;119:408–415.
58. Ledig S, Ropke A, Haeusler G, et al. BMP15 mutations in XX gonadal dysgenesis and premature ovarian failure. *Am J Obstet Gynecol*. 2008;198:e81–e85. 84.
59. Durlinger AL, Gruijters MJ, Kramer P, et al. Anti-Mullerian hormone inhibits initiation of primordial follicle growth in the mouse ovary. *Endocrinology*. 2002;143:1076–1084.
60. Durlinger AL, Kramer P, Karels B, et al. Control of primordial follicle recruitment by anti-Mullerian hormone in the mouse ovary. *Endocrinology*. 1999;140:5789–5796.
61. Mark-Kappeler CJ, Hoyer PB, et al. Xenobiotic effects on ovarian preantral follicles. *Biol Reprod*. 2011;85:871–883.
62. Yamoto M, Shima K, Nakano R. Gonadotropin receptors in human ovarian follicles and corpora lutea throughout the menstrual cycle. *Horm Res*. 1992;37(1):5–11.
63. Nakamura M, Minegishi T, Hasegawa Y, et al. Effect of an activin A on follicle-stimulating hormone (FSH) receptor messenger ribonucleic acid levels and FSH receptor expressions in cultured rat granulosa cells. *Endocrinology*. 1993;133:538–544.
64. Kawamura K, Cheng Y, Suzuki N, et al. Hippo signaling disruption and Akt stimulation of ovarian follicles for infertility treatment. *Proc Natl Acad Sci USA*. 2013;110:17474–17479.
65. Sen A, Hammes SR. Granulosa cell-specific androgen receptors are critical regulators of ovarian development and function. *Mol Endocrinol*. 2010;24:1393–1403.
66. Nagels HE, Rishworth JR, Siristatidis CS. Androgens (dehydroepiandrosterone or testosterone) for women undergoing assisted reproduction. *Cochrane Database Syst Rev*. 2015:CD009749.
67. Erickson GF, Magoffin DA. 3. Ovarian function: follicle. The endocrine control of follicle androgen biosynthesis. *J Steroid Biochem*. 1983;19:113–117.
68. Li R, Phillips DM, Mather JP. Activin promotes ovarian follicle development in vitro. *Endocrinology*. 1995;136:849–856.
69. G.F. E. The graafian follicle: a functional definition. In: Adashi EY, ed. *Ovulation: Evolving Scientific and Clinical Concepts*. New York: Springer-Verlaag; 2000:31–48.
70. Green SH, Zuckerman S. Quantitative aspects of the growth of the human ovum and follicle. *J Anat*. 1951;85:373–375.
71. Richards JS. Hormonal control of gene expression in the ovary. *Endocr Rev*. 1994;15:725–751.
72. Cacioppo JA, Oh SW, Kim HY, et al. Loss of function of endothelin-2 leads to reduced ovulation and CL formation. *PLoS One*. 2014;9:e96115.
73. Migone FF, Cowan RG, Williams RM, et al. In vivo imaging reveals an essential role of vasoconstriction in rupture of the ovarian follicle at ovulation. *Proc Natl Acad Sci USA*. 2016;113:2294–2299.
74. Baena V, Terasaki M. Three-dimensional organization of transzonal projections and other cytoplasmic extensions in the mouse ovarian follicle. *Sci Rep*. 2019;9:1262.
75. Erickson GF, Shimasaki S. The role of the oocyte in folliculogenesis. *Trends Endocrinol Metab*. 2000;11:193–198.
76. Gougeon A. Dynamics of follicular growth in the human: a model from preliminary results. *Hum Reprod*. 1986;1:81–87.
77. Gougeon A, Lefevre B. Evolution of the diameters of the largest healthy and atretic follicles during the human menstrual cycle. *J Reprod Fertil*. 1983;69:497–502.
78. Baerwald A, Pierson R. Ovarian follicular waves during the menstrual cycle: physiologic insights into novel approaches for ovarian stimulation. *Fertil Steril*. 2020;114:443–457.
79. Zeleznik AJ. Dynamics of primate follicular growth. In: Adashi E, ed. *The Ovary*. New York: Raven Press; 1993.
80. McNatty KP, Hunter WM, MacNeilly AS. Changes in the concentration of pituitary and steroid hormones in the follicular fluid of human graafian follicles throughout the menstrual cycle. *J Endocrinol*. 1975;64:555–571.
81. McNatty KP, Sawers RS. Relationship between the endocrine environment within the Graafian follicle and the subsequent rate of progesterone secretion by human granulosa cells in vitro. *J Endocrinol*. 1975;66:391–400.
82. Gospodarowicz D, Bialecki H. Fibroblast and epidermal growth factors are mitogenic agents for cultured granulosa cells of rodent, porcine, and human origin. *Endocrinology*. 1979;104:757–764.
83. McNatty KP. Hormonal correlates of follicular development in the human ovary. *Aust J Biol Sci*. 1981;34:249–268.
84. Franks S. Polycystic ovary syndrome. *N Engl J Med*. 1995;333:853–861.

85. Simoni M, Gromoll J, Nieschlag E. The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. *Endocr Rev*. 1997;18:739–773.
86. Richards JS, Pangas SA. The ovary: basic biology and clinical implications. *J Clin Investig*. 2010;120:963–972.
87. Suzuki T, Sasano H, Kimura N, et al. Immunohistochemical distribution of progesterone, androgen and oestrogen receptors in the human ovary during the menstrual cycle: relationship to expression of steroidogenic enzymes. *Hum Reprod*. 1994;9:1589–1595.
88. Ghersevich SA, Poutanen MH, Martikainen HK, et al. Expression of 17 beta-hydroxysteroid dehydrogenase in human granulosa cells: correlation with follicular size, cytochrome P450 aromatase activity and oestradiol production. *J Endocrinol*. 1994;143:139–150.
89. Sawetawan C, Milewich L, Word RA, et al. Compartmentalization of type I 17 beta-hydroxysteroid oxidoreductase in the human ovary. *Mol Cell Endocrinol*. 1994;99:161–168.
90. Minegishi T, Tano M, Abe Y, et al. Expression of luteinizing hormone/human chorionic gonadotrophin (LH/HCG) receptor mRNA in the human ovary. *Mol Hum Reprod*. 1997;3:101–107.
91. Graham JD, Clarke CL. Physiological action of progesterone in target tissues. *Endocr Rev*. 1997;18:502–519.
92. Zhang M, Su YQ, Sugiura K, et al. Estradiol promotes and maintains cumulus cell expression of natriuretic peptide receptor 2 (NPR2) and meiotic arrest in mouse oocytes in vitro. *Endocrinology*. 2011;152:4377–4385.
93. Zhang M, Su YQ, Sugiura K, et al. Granulosa cell ligand NPPC and its receptor NPR2 maintain meiotic arrest in mouse oocytes. *Science*. 2010;330:366–369.
94. Pincus G, Enzmann EV. The comparative behavior of mammalian eggs in vivo and in vitro : I. the activation of ovarian eggs. *J Exp Med*. 1935;62:665–675.
95. Edwards RG. Meiosis in ovarian oocytes of adult mammals. *Nature*. 1962;196:446–450.
96. Tsafriri A, Pomerantz SH. Oocyte maturation inhibitor. *Clin Endocrinol Metab*. 1986;15:157–170.
97. Vaccari S, Weeks 2nd JL, Hsieh M, et al. Cyclic GMP signaling is involved in the luteinizing hormone-dependent meiotic maturation of mouse oocytes. *Biol Reprod*. 2009;81:595–604.
98. Norris RP, Ratzan WJ, Freudzon M, et al. Cyclic GMP from the surrounding somatic cells regulates cyclic AMP and meiosis in the mouse oocyte. *Development*. 2009;136:1869–1878.
99. Conti M, Hsieh M, Zamah AM, et al. Novel signaling mechanisms in the ovary during oocyte maturation and ovulation. *Mol Cell Endocrinol*. 2012;356:65–73.
100. Erickson GF. Defining apoptosis: players and systems. *J Soc Gynecol Investig*. 1997;4:219–228.
101. Hsueh AJ, Billig H, Tsafriri A. Ovarian follicle atresia: a hormonally controlled apoptotic process. *Endocr Rev*. 1994;15:707–724.
102. Tilly JL. Apoptosis and ovarian function. *Rev Reprod*. 1996;1:162–172.
103. Chun SY, Billig H, Tilly JL, et al. Gonadotropin suppression of apoptosis in cultured preovulatory follicles: mediatory role of endogenous insulin-like growth factor I. *Endocrinology*. 1994;135:1845–1853.
104. Conti M, Franciosi F. Acquisition of oocyte competence to develop as an embryo: integrated nuclear and cytoplasmic events. *Hum Reprod Update*. 2018;24:245–266.
105. Park JY, Su YQ, Ariga M, et al. EGF-like growth factors as mediators of LH action in the ovulatory follicle. *Science*. 2004;303:682–684.
106. Eppig J. Oocyte-somatic cell communication in the ovarian follicles of mammals. *Semin Dev Biol*. 1994;5:51–59.
107. Tsafriri A, Abisogun AO, Reich R. Steroids and follicular rupture at ovulation. *J Steroid Biochem*. 1987;27:359–363.
108. Lydon JP, DeMayo FJ, Funk CR, et al. Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev*. 1995;9:2266–2278.
109. Lim H, Paria BC, Das SK, et al. Multiple female reproductive failures in cyclooxygenase 2-deficient mice. *Cell*. 1997;91:197–208.
110. Herschman HR. Prostaglandin synthase 2. *Biochim Biophys Acta*. 1996;1299:125–140.
111. Shikone T, Yamoto M, Kokawa K, et al. Apoptosis of human corpora lutea during cyclic luteal regression and early pregnancy. *J Clin Endocrinol Metab*. 1996;81:2376–2380.
112. Vande Wiele RL, Bogumil J, Dyrenfurth I, et al. Mechanisms regulating the menstrual cycle in women. *Recent Prog Horm Res*. 1970;26:63–103.
113. Carr BR, MacDonald PC, Simpson ER. The role of lipoproteins in the regulation of progesterone secretion by the human corpus luteum. *Fertil Steril*. 1982;38:303–311.
114. Brannstrom M, Norman RJ. Involvement of leukocytes and cytokines in the ovulatory process and corpus luteum function. *Hum Reprod*. 1993;8:1762–1775.

119

Ovarian Hormone Synthesis

Elizabeth A. McGee, Jerome F. Strauss III, and Callum A. Potts

OUTLINE

Ovarian Steroid Hormone Biosynthesis, 1987
- Acquisition of Cholesterol, 1988
- The Cholesterol Side-Chain Cleavage Reaction, 1989
- Synthesis of Androgens, 1990
- Synthesis of Estrogens, 1990
- Steroid Hormone Metabolites, 1991
- MicroRNA Regulation of Steroidogenic Enzymes, 1991

Peptide Hormones and Growth Factors Produced by the Ovary, 1992
- Transforming Growth Factor-β Superfamily, 1992
- Relaxin, 1992

Gonadotropin Control of Ovarian Steroid Production, 1992
- Follicle-Stimulating Hormone and Luteinizing Hormone, 1992
- Mechanisms of Gonadotropin Action, 1993
- Other Hormones with Possible Actions in the Ovary, 1993

Intraovarian Regulators, 1994
- Cholesterol Precursors, 1994
- Role of Estrogens, 1995
- Role of Androgens, 1995
- Role of Progesterone, 1996
- Transforming Growth Factor-β Superfamily, 1996
- Eicosanoids, 1997
- Insulin-Like Growth Factors and Their Binding Proteins, 1997
- Other Intraovarian Regulators, 1997

Hormone Production During the Ovarian Life Cycle, 1998
- The Fetal and Prepubertal Ovary, 1998
- The Ovary During Reproductive Life, 1998
- Endocrine Activity of the Postmenopausal Ovary, 1999

Conclusion, 1999

Ovarian hormones can be divided into two categories according to their chemical nature: (1) lipids, including steroids and eicosanoids, and (2) peptide hormones. These hormones act locally within the ovary and adjacent structures and systemically. In the ovary, germ cell growth and maturation and most hormone synthesis take place in the follicle or its derivative, the corpus luteum. The follicle consists of an outer layer of theca cells encircling the inner layers of granulosa cells, which, in turn, surround the oocyte. The androgen-producing theca cells and the estrogen-producing granulosa cells respond to the pituitary gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH), respectively. During the final stages of follicle maturation, LH receptors are also expressed on granulosa cells, and the preovulatory LH surge acts on both granulosa and theca cells of the preovulatory (Graafian) follicle, inducing ovulation and transformation of the follicle into a corpus luteum. Circulating and locally produced peptide growth factors, cytokines, and growth factor–binding proteins, as well as the sex steroids produced by the ovarian cells themselves, amplify or attenuate the actions of gonadotropins. Together, they participate in the intricate dialog between the oocyte, granulosa cells, theca cells, vasculature, immune cells, and adjacent stroma to coordinate growth and maturation of the follicle and oocyte, as well as the function of the corpus luteum (see Chapter 120).

Prior to puberty, the human ovary is relatively quiescent in terms of steroidogenesis despite the presence of significant levels of circulating gonadotropins during fetal life and early infancy. The factors that restrain ovarian sex steroid secretion during these times include the relative absence of expression of LH receptors and the low level of expression of key steroidogenic enzyme genes. Gonadarche, the first evidence of substantial ovarian function in response to rising pituitary gonadotropin secretion at puberty, heralds the transition into the reproductive years. During the normal reproductive stage of life, some 400 to 500 follicles will complete the full developmental cycle. Generally, one follicle in a cohort of growing follicles will be selected to ovulate in each cycle. The remainder of the cohort is lost to atresia via autophagy and apoptosis. Under the influence of gonadotropins, the follicle selected to ovulate produces increasing amounts of estradiol through the aromatization of thecal androgen precursors, which ultimately triggers the preovulatory LH surge, ovulation and the transformation of the follicle into a corpus luteum. The corpus luteum in turn produces large quantities of progesterone, and also secretes estrogens. After the depletion of follicles with menopause, the ovarian capacity to cyclically synthesize estradiol and progesterone is lost. However, residual hilar and subsets of stromal cells maintain some capacity for limited steroidogenesis.

OVARIAN STEROID HORMONE BIOSYNTHESIS

The human ovary produces three main classes of steroid hormones: estrogens, androgens, and progestins. Enzymes involved in steroid biosynthesis and metabolism can be classified into two broad groups, the cytochrome P450s and the oxidoreductases. Cytochrome P450s are heme-containing proteins, named for their characteristic light absorption maximum at 450 nm in the presence of carbon monoxide. The second main group of steroid-metabolizing enzymes, oxidoreductases, can be further divided into two families based on their structure and function, namely, short-chain dehydrogenases/reductases (SDR) and aldo-keto reductase (AKR) families.[1] SDR include 3β-hydroxysteroid dehydrogenase (3β-HSD) types I and II and most of the known 17β-HSDs. Of the AKR family, 17β-HSD type V *(AKR1C3)*, is expressed in the ovary, and is thought to be essential for androgen production.

The rate of production of steroid hormones is determined largely by the delivery of cholesterol to the first enzyme in the steroidogenic

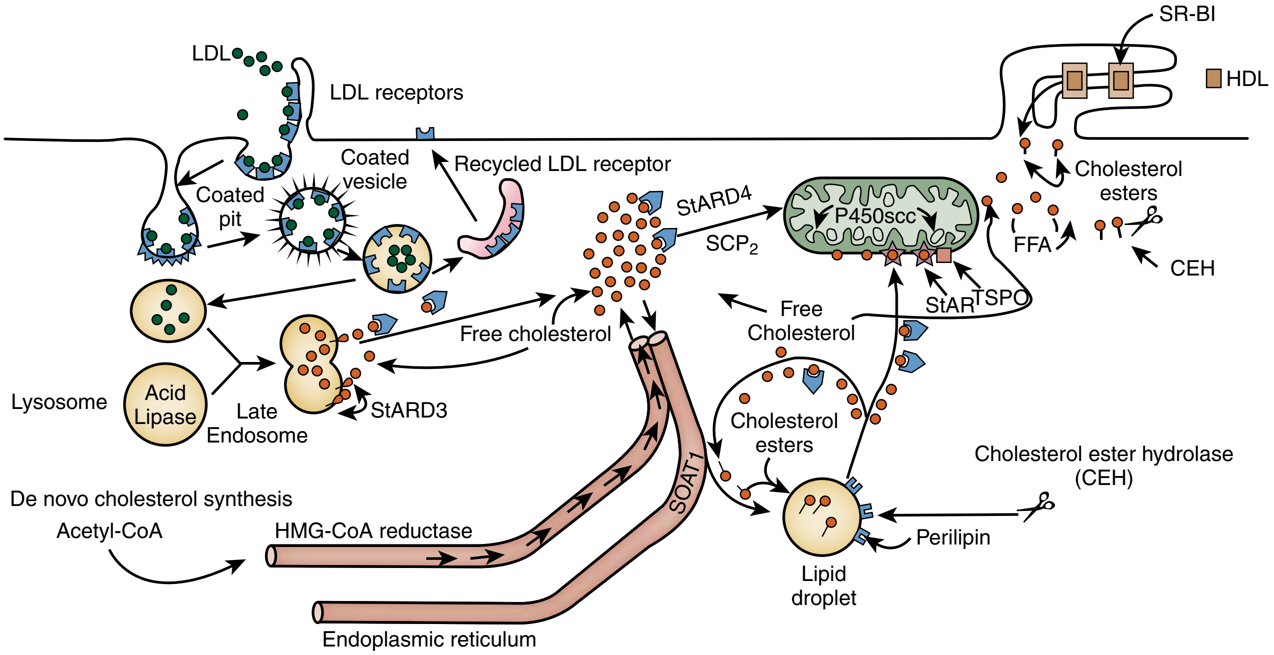


Figure 119.1 The Acquisition, Storage, and Trafficking of Cholesterol in Steroidogenic Cells. *FFA,* free fatty acid; *HMG-CoA,* 3-hydroxy-3-methylglutaryl-coenzyme A; *HDL,* high-density lipoprotein; *LDL,* low-density lipoprotein; *StARD3,* steroidogenic acute regulatory protein–related lipid transfer domain 3; *P450scc,* cholesterol side chain-cleavage enzyme; *SCP2*, sterol carrier protein 2; *StAR,* steroidogenic acute regulatory protein; *StARD4,* START domain protein 4; *SOAT1,* sterol-O-acyltransferase-1; *TSPO,* translocator protein/peripheral benzodiazepine receptor.

pathway, cytochrome P450 side chain-cleavage enzyme (P450scc encoded by *CYP11A1*), and also by the levels and activities of this and other enzymes that catalyze subsequent biosynthetic steps.[2] Gonadotropins acutely increase ovarian steroid synthesis by stimulating cholesterol transport to P450scc and, in the longer term, control the content of the steroidogenic machinery, mainly by stimulating transcription of the genes encoding steroidogenic enzymes. These acute and long-term changes are primarily, but not exclusively, initiated by a cyclic adenosine monophosphate (cAMP)/protein kinase A (PKA)–mediated signaling system in thecal and luteal cells. FSH regulates the expression of steroidogenic enzymes in granulosa cells, including aromatase, mainly via a cAMP/PKA signaling system, although other signaling pathways may contribute (e.g., Akt/protein kinase B [PKB]).

Acquisition of Cholesterol

Ovarian cells, like other steroidogenic cells, acquire cholesterol either by *de novo* synthesis or by the uptake of lipoprotein-carried cholesterol. Trophic hormones stimulating steroidogenesis generally increase both cholesterol synthesis and lipoprotein uptake. Lipoprotein-associated cholesterol and cholesterol esters enter steroidogenic cells by two receptor-mediated pathways: the low-density lipoprotein (LDL) and the high-density lipoprotein (HDL) pathways (Fig. 119.1). Human steroidogenic cells are enriched with LDL receptors that mediate endocytotic lipoprotein uptake. The endocytosed lipoproteins are delivered to the lysosomal compartment where the apolipoproteins are degraded and the cholesterol esters are hydrolyzed by lysosomal acid lipase. The free cholesterol is then released from the lysosomes through the action of Niemann-Pick C proteins, NPC2 and NPC1. Stimulation of steroidogenic cells by trophic hormones increases the number of LDL receptors, and also accelerates the rate of LDL internalization and degradation. The uptake of HDL cholesterol involves a different process. HDL particles bind to scavenger receptors (e.g., SR-BI), with broad ligand specificity, recognizing HDL apolipoproteins as well as lipids. HDL cholesterol esters are selectively internalized, leaving the apolipoproteins on the cell surface. Evidence indicates that the internalized HDL cholesterol esters are not cleaved by lysosomal acid lipase, but via the action of cytosolic cholesterol ester hydrolase (CEH) or carboxyl ester lipase.[2,3]

Steroidogenic cells producing large quantities of hormone (e.g., luteal cells) accumulate the majority of their cholesterol from circulating lipoproteins. The access of lipoproteins to certain ovarian cells may therefore affect their ability to synthesize hormones. A partial blood-follicle barrier excludes high-molecular-weight substances from the follicular antrum, preventing LDL from reaching granulosa cells until after breakdown of the follicular basement membrane in the ovulatory process, which subsequently allows luteal cells access to LDL.

A characteristic feature of steroidogenic cells is the presence of numerous cytoplasmic lipid droplets containing cholesterol esters. Sterol esters in these droplets are synthesized by sterol-O-acyltransferase type 1 (SOAT1), also known as *acyl coenzyme A: cholesterol acyltransferase* (ACAT1), an endoplasmic reticulum enzyme.[4] The esters generated by SOAT1 accumulate within the endoplasmic reticulum membranes and subsequently bud off as lipid droplets. The sterol esters in lipid droplets are hydrolyzed by the soluble CEHs (see Fig. 119.1). Gonadotropin stimulation of cAMP-dependent protein kinases activates certain CEHs by phosphorylation of serine residues, which promotes the association of the enzyme with perilipins, proteins that coat lipid droplets and normally restrain mobilization of sterol esters. But when phosphorylated by cAMP-activated protein kinases, perilipins facilitate the action of hormone-activated esterases, resulting in hydrolysis of the lipid droplet sterol esters.[5] The size and the number of lipid droplets change as the ester pool expands or contracts. Cellular cholesterol ester storage is dependent on the *de novo* synthesis or accumulation of cholesterol on the one hand and cholesterol depletion through steroidogenic activity of the cell on the other. Trophic stimulation promotes cholesterol ester hydrolysis and diverts

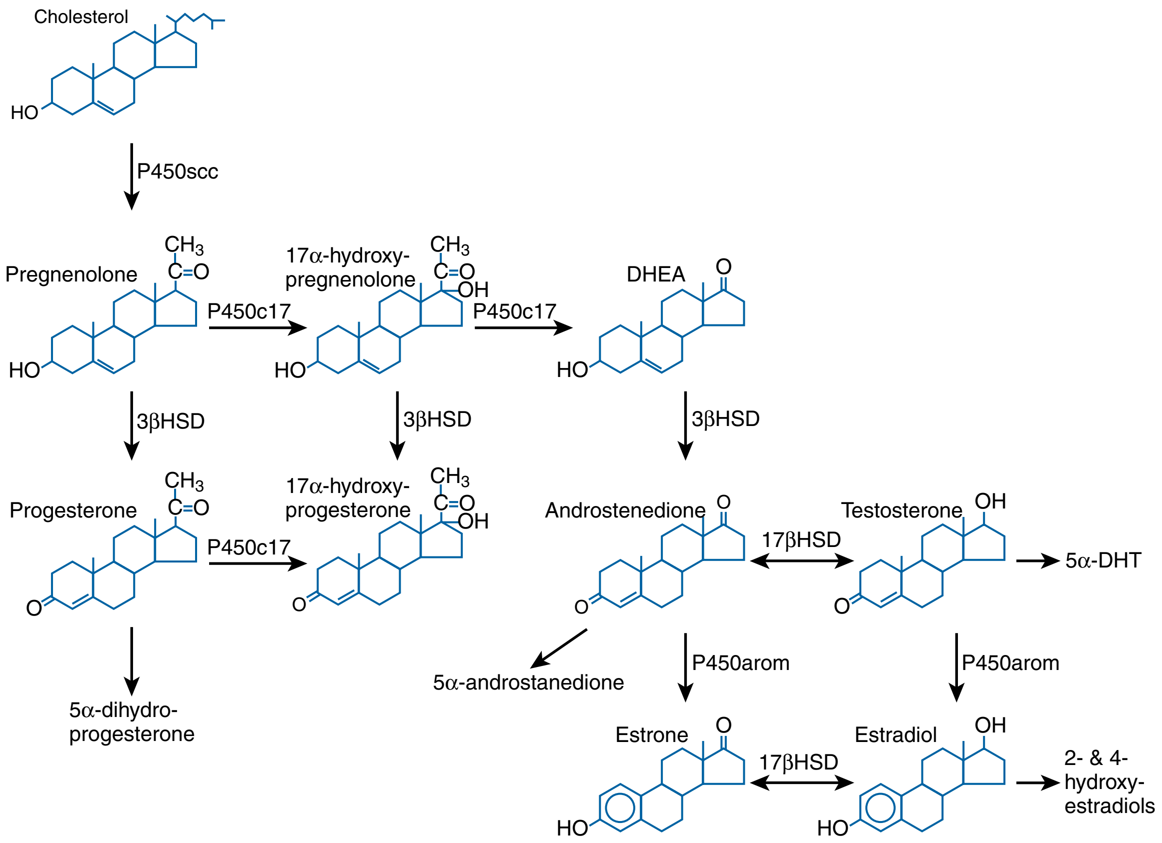


Figure 119.2 Major Pathways of Steroid Biosynthesis and Metabolism in the Human Ovary. Steroid metabolites in italics have unclear roles in the ovary, and their speculative functions are described in the text. *17βHSD,* 17β-Hydroxysteroid dehydrogenase; *3βHSD,* 3β-hydroxysteroid dehydrogenase; *DHEA,* dehydroepiandrosterone; *DHT,* 5α-dihydrotestosterone.

cholesterol into the steroidogenic pool away from SOAT1, preventing reesterification and resulting in a net depletion of cholesterol from the lipid droplets.

The Cholesterol Side-Chain Cleavage Reaction

The cholesterol side-chain cleavage reaction is the first committed step, as well as the rate-limiting process, in steroid hormone synthesis. This reaction takes place on the inner mitochondrial membranes, catalyzed by P450scc (Fig. 119.2) and its associated electron transport system, consisting of a flavoprotein reductase (ferredoxin or adrenodoxin reductase) and an iron sulfoprotein (ferredoxin or adrenodoxin). The dissociation constant (K_d) for binding of cholesterol is about 5000 nM, whereas P450scc's affinities for the intermediate substrates (22R) 22-hydroxycholesterol and (20R, 22R) 20,22-dihydroxycholesterol are more than 60-fold higher ($K_d = 4.9$ and 81 nM, respectively). Thus, once cholesterol is bound to the enzyme, it is committed to complete the reaction sequence. The estimated K_d for pregnenolone is 2900 nM, which permits the dissociation of the final reaction product from the enzyme.

The rate of formation of pregnenolone is determined by multiple factors: (1) delivery of cholesterol to the mitochondria; (2) access of cholesterol to the inner mitochondrial membranes; (3) amount of cholesterol side-chain cleavage enzyme, and secondarily its flavoprotein and iron-sulfur protein electron transport chain; and (4) catalytic activity of P450scc. Acute alterations in steroidogenesis generally result from changes in the delivery of cholesterol to P450scc or availability of electrons from the adrenodoxin/adrenodoxin reductase, whereas long-term alterations involve changes in the quantities of enzyme proteins as well as cholesterol delivery.

The transport of cholesterol from various substrate pools (including lipid droplets, the plasma membrane, and endoplasmic reticulum) to the mitochondria involves the action of sterol transport proteins, including members of the steroidogenic acute regulatory protein–related lipid transfer (StART) domain superfamily (StARD1, StARD3, StARD4, StARD5), Aster-B (also known as GRAMD1B) and sterol carrier protein-2. These transport proteins may interact with other cytosolic proteins, including CEH, to facilitate cholesterol movement to the mitochondria.[6] However, it is the translocation of cholesterol deposited in the outer mitochondrial membranes to the relatively sterol-poor inner mitochondrial membranes through a multiprotein transducer complex that is the critical step in the cholesterol side-chain cleavage process.[7]

The steroidogenic acute regulatory protein (StAR or StARD1), the prototypical member of the StART domain family, plays a pivotal role in the translocation process.[8] Expression of StARD1 mRNA and protein is correlated with steroidogenesis, and transfection of nonsteroidogenic cells with StARD1 and cholesterol side-chain cleavage enzyme system expression plasmids endows them with the capacity for pregnenolone synthesis. The discovery that inactivating mutations in the *StARD1* gene cause congenital lipoid adrenal hyperplasia (lipoid CAH) conclusively established an essential role for StARD1 in adrenal and gonadal steroidogenesis.[8] Patients suffering from this relatively rare autosomal recessive disorder are characterized by impaired biosynthesis of all classes of adrenal and gonadal steroid hormones due to a defect at the cholesterol side-chain cleavage step. However, 46,XX females with lipoid CAH develop female sexual characteristics at puberty and menstrual bleeding despite nonfunctional StARD1 proteins and adrenal insufficiency.[9] These findings demonstrate that production of small amounts of hormone can occur by StARD1-independent processes, including estrogen production by the ovary, which can be sufficient to produce secondary sexual development and endometrial growth.

In response to gonadotropin stimulation, StARD1 is synthesized as a 37-kDa preprotein containing an amino-terminal targeting sequence that directs it to the mitochondria. On the mitochondrial outer membrane, StARD1 interacts with a multiprotein complex (transducesome) to promote cholesterol translocation to the inner mitochondrial membrane through cycles of unfolding and refolding of the StART domain. The transduceome is thought to contain translocator protein (TSPO), voltage-dependent anion channel 1 (VDAC), acyl-coenzyme A binding domain-containing protein 3 (also known as PAP7), PKA regulatory subunit (RI alpha), and proteins of the 14-3-3 family. The specific roles of each component in StARD1 action are not yet known. For example, TSPO (also known as the *peripheral benzodiazepine receptor*) was once thought to play a critical role in StARD1 action, but steroidogenesis occurs normally in mice with a targeted mutation in the *Tspo* gene. TSPO has more recently been linked to the regulation of fatty acid oxidation and reactive oxygen species formation in steroidogenic cells.[10,11]

StARD1 contains several consensus sequences for cAMP-dependent protein kinase phosphorylation. Phosphorylation of Ser57 mediates StARD1 binding to the transduceosome, with the assistance of 14-3-3 protein family members. Phosphorylation of Ser195, which is located in the StART domain, is thought to affect the transition between folded and unfolded state and endow StARD1 with enhanced activity.[12] Although the StARD1 StART domain can bind cholesterol, sterol binding does not appear to be a requirement for its activity.

Residence time on the outer membrane is a determinant of the steroidogenic response. The action of StARD1 is terminated by import into the mitochondria and proteolytic processing of the targeting sequence to yield the 30-kDa "mature" form of the protein.[13] With the removal of StARD1 from the outer mitochondrial membrane, the translocation of cholesterol to the inner membranes ceases. Turning steroidogenesis back on requires the synthesis of new StARD1 molecules.

Pregnenolone generated by the action of the cholesterol side-chain cleavage enzyme can undergo further metabolism by enzymes closely associated with the mitochondria. Tom22, spanning the outer mitochondrial membrane to the inner mitochondrial space, interacts with the 3β-hydroxysteroid/Δ^5-Δ^4 isomerases, which are required for the conversion of pregnenolone to progesterone, 17α-hydroxypregnenolone to 17α-hydroxyprogesterone, and dehydroepiandrosterone to androstenedione (see Fig. 119.2).[14] Using nicotinamide adenine dinucleotide (NAD^+) as a cofactor, these enzymes catalyze the dehydrogenation of the 3β-hydroxyl group and the subsequent isomerization of the Δ^5 olefinic bond to yield the Δ^4,3-ketone structure.[14]

The importance of 3β-HSD to steroidogenesis is highlighted by the observation that inhibitors of 3β-HSD, such as epostane, effectively block progesterone synthesis and interrupt early human pregnancy. Two different but highly homologous human *3β-HSD* genes have been identified and localized to chromosome 1p13.1. The type I gene *(HSD3B1)* is expressed primarily in the human placenta, skin, and adipose tissue. The type II gene *(HSD3B2)* encodes the primary 3β-HSD enzyme expressed in the gonads and adrenal cortex. The type II 3β-HSD in human corpus luteum has been localized to the peri-mitochondrial endoplasmic reticulum by electron microscope cytochemistry. The enzyme is, therefore, positioned to convert pregnenolone produced by the mitochondrial cholesterol side chain-cleavage system to progesterone. Because steroidogenic cells have a large capacity to generate progesterone when presented with exogenous pregnenolone, 3β-HSD is not thought to be a rate-determining enzyme. However, the level of 3β-HSD increases following trophic hormone stimulation.[15]

Synthesis of Androgens

Theca cells of the follicle and theca lutein cells of the corpus luteum express P450c17 (CYP17A1), the key enzyme in androgen biosynthesis. This endoplasmic reticulum enzyme catalyzes two reactions in the ovary: (1) hydroxylation of pregnenolone and progesterone at C17 and (2) conversion of 17α-hydroxypregnenolone into dehydroepiandrosterone (lyase reaction). The 17α-hydroxylation reaction requires a pair of electrons and one molecule of O_2; a second electron pair and a molecule of O_2 are needed for the lyase reaction. The electrons are transferred from NADPH to the P450c17 heme iron by NADPH-cytochrome P450 reductase (POR), the importance of which is demonstrated by disordered steroidogenesis in patients harboring mutations of the *POR* gene, encoding NADPH-cytochrome POR.[16] The importance of POR is highlighted by the spectrum of conditions uncovered by its deficiency, chiefly menstrual cycle disorders, adrenal insufficiency, and disorders of sexual development.[17,18]

Several factors determine whether substrates undergo only 17α-hydroxylation or the subsequent scission of the 17,20 bond. These include the nature of the substrate, the flux of reducing equivalents, post-translational modification of P450c17, and allosteric effectors. The human P450c17 only utilizes Δ^5 substrates for 17,20 bond cleavage (see Fig. 119.2). Thus, 17α-hydroxyprogesterone is not metabolized to androstenedione at significant rates in humans. Phosphorylation of P450c17 by the kinase, p38α, appears to be necessary for maximal 17,20-lyase activity.[19] Cytochrome *b5* increases 17,20-lyase activity by acting as an allosteric modulator between POR and P450c17 rather than an electron donor. Females with mutations in the *CYB5A* gene, which encodes cytochrome *b5* can undergo normal sexual development and are fertile.[20]

The 11-oxyandrogens (e.g., 11-ketotestosterone, 11-ketoandrostenedione, 11β-hydroxy testosterone, and 11β-hydroxyandrostenedione) are produced by the adrenal cortex, which expresses 11β-hydroxylase (*CYP11B1*) and 11β-HSDs.[21] The 11-oxyandrogens bind to and activate androgen receptors (ARs), but evidently cannot be aromatized *in vivo* since 11-oxoestrogens cannot be detected in blood. There is evidence that ovarian theca cells also express *CYP11B1* and *HSD11B2*, albeit at lower levels than in the adrenal cortex, and can be a source of these bioactive steroids.[22]

Synthesis of Estrogens

Estrogens are synthesized in granulosa cells and granulosa lutein cells by the action of P450arom. In most cases, P450arom-expressing cells utilize androgen precursors derived from another cell type, placing ovarian estrogen synthesis under complex regulation epitomized by the "two-cell, two-gonadotropin" model of follicular steroid biosynthesis. According to this model, LH drives thecal synthesis of androgens, which are subsequently aromatized to estrogens in adjacent granulosa cells (Fig. 119.3). FSH is the primary stimulator of *P450arom* gene expression in granulosa cells, which contain more than 99% of the aromatization capacity of the follicle.[23] However, all granulosa cells are not equal in their capacity to form estrogens. There is a gradient of activity, with the highest aromatase expression located in the mural granulosa cells at the basement membrane, and the lowest levels in cells adjacent to the antrum.

Aromatase catalyzes three sequential hydroxylations of a C19 substrate using three molecules of NADPH and three molecules of oxygen to produce a C18 estrogen characterized by a phenolic A ring, with the elimination of the C19 methyl group as formic acid. The

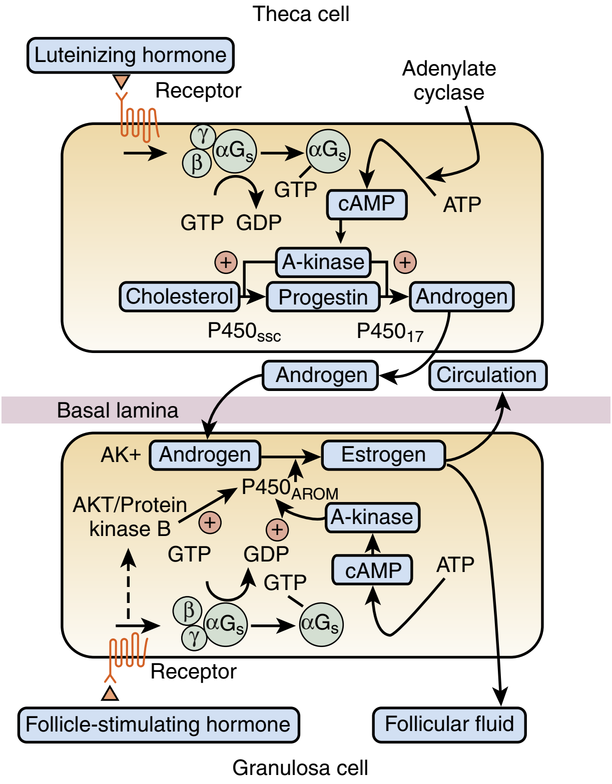


Figure 119.3 Gonadotropin Regulation of Follicular Androgen and Estrogen Biosynthesis and the Two-Cell, Two-Gonadotropin Model. *P450arom,* Aromatase; *P450c17,* l7a-hydroxylase; *P450scc,* cholesterol side chain–cleavage enzyme.

first hydroxylation yields a C19 hydroxyl derivative; this is followed by a second hydroxylation step to produce an unstable gem-diol, which collapses to yield a C19 aldehyde. The final hydroxylation event involves the formation of a 19-hydroxy-19-hydroperoxide intermediate, resulting in the subsequent elimination of a formic acid molecule and aromatization of the steroid A ring. This sequence of reactions takes place at a single active site on the enzyme, with reducing equivalents transferred to P450arom by NADPH-cytochrome POR.[24] The aromatase protein is encoded by a single gene *(CYP19A1)* on chromosome 15q21.1, which gives rise to mRNAs of various sizes as a result of the use of tissue-specific alternative promoters.[25] The promoter (promoter II) that drives ovarian aromatase expression lies adjacent to the exon encoding the translation start site. In granulosa cells, FSH stimulates the transcription of both *CYP19A1* and the gene encoding NADPH-cytochrome POR. Mutations in the gene encoding aromatase may be associated with ambiguous genitalia at birth, abnormal pubertal development and polycystic ovaries, and aromatase deficiency may lead to premature ovarian insufficiency compared with other steroid biosynthesis disorders like lipoid CAH.[26]

Gonadal tissues reduce 17-ketosteroids to 17β-hydroxysteroids, which have greater biological potency. The type I enzyme is also referred to as the *estrogenic* 17β-HSD since it catalyzes the final step in estrogen biosynthesis by preferentially reducing the weak estrogen, estrone, to yield 17β-estradiol (see Fig. 119.2), a potent estrogen, using either NADH or NADPH as a cofactor. Having a 100-fold higher affinity for C18 steroids over C19 steroids, the type I enzyme is located in the cytosol and is expressed in granulosa cells under the influence of FSH.[27] Although it is not normally expressed in the ovary, the type III enzyme is referred to as the *androgenic* 17β-HSD because it catalyzes the final step in androgen biosynthesis, the reduction of the weak androgen androstenedione to the more potent androgen testosterone. The primary androgen produced by the theca is androstenedione.[27] However, a small amount of testosterone is also made by the ovary. The absence of type III enzyme mRNA from the human ovary indicates that another 17β-HSD must be responsible for its production. Type V and Type VII 17β-HSDs are both candidates, but the type V enzyme, also known as AKR1C3, most likely is the enzyme because it is expressed in theca and luteal cells.

Steroid Hormone Metabolites

Metabolites of progesterone, testosterone, and estradiol are produced in the ovary. This is particularly true for the corpus luteum, which produces substantial amounts of 17α-hydroxyprogesterone and 5α-dihydroprogesterone. The secretion of both of these progesterone metabolites parallels the production of progesterone during the luteal phase. Since pregnancy can be established and maintained in the absence of these molecules, they do not appear to have obligatory roles in reproductive physiology.

Two distinct isozymes of 5α-reductase, namely type I *(SRD5A1)* and type II *(SRD5A2)*, have been identified. Both isozymes reduce the double bond between C4 and C5 and can convert progesterone to 5α-dihydroprogesterone, testosterone to dihydrotestosterone (DHT), and androstenedione to 5α-androstenedione. Whereas *SRD5A1* mRNA expression has been clearly shown in the ovary, mRNA expression of *SRD5A2* is debated.[28,29] DHT is present in follicular fluid and ARs are present in granulosa cells, but the precise role of DHT in normal human ovarian physiology remains to be elucidated (see the section entitled "Role of Androgens" later in the chapter). The primary estrogens, estradiol and estrone, can undergo hydroxylation at C2 and C4 positions in the ovary, resulting in the formation of catechol estrogens, which contain a dihydroxybenzene A ring. Metabolism of estradiol to the catechol estrogens, 2-hydroxyestradiol and 4-hydroxyestradiol, is catalyzed by estrogen 2-hydroxylase *(CYP1A1)* and estrogen 4-hydroxylase *(CYP1B1)*, respectively. Both of these metabolites bind to ERα and ERβ with 7% to 13% relative binding affinity compared with estradiol.[30] Catechol estrogens can be further *O*-methylated by catechol-*O*-methyl transferase (COMT) to form methoxy estrogens. These metabolites appear to have important roles in controlling ovarian angiogenesis, acting as inhibitors of hypoxia factor-1a.

MicroRNA Regulation of Steroidogenic Enzymes

MicroRNA (miRNA), small interfering RNA and piwiRNA, are subtypes of small non-coding RNAs. These RNAs are numerous and are present in cells throughout the body, playing a key role in the post-transcriptional regulation of a number of biological processes. Notably, a single miRNA may regulate up to hundreds if not thousands of different mRNAs, and a single mRNA may be affected by multiple miRNAs. The effects of miRNAs can be positive or negative, upregulating gene function or silencing translation of mRNA depending on the conditions present within the cell. The multitude of possible effects has led to difficulty in interpreting the physiological role of specific miRNAs, although there is growing evidence that they play a significant role in modulating the coordination required for follicle recruitment, maturation, and ovulation. In the ovary, they appear to have roles in regulating cell growth, differentiation, apoptosis, and other cell functions,[31] including gonadotropin receptor function, growth factor signaling, and steroidogenesis. For example, miRNA-33 has been shown to play a key role in controlling cholesterol metabolism[32]; miRNA-136-3p has been reported to affect LH receptor expression[33]; has-miR-548ba resides within the FSH receptor gene and regulates follicle growth and activation; and has-miR-7973 resides within the aromatase gene and modulates extracellular matrix and cell-cell interactions.[34]

KEY POINTS

- Cholesterol acquisition and metabolism to produce a variety of steroid hormones is a major function of the ovary.

PEPTIDE HORMONES AND GROWTH FACTORS PRODUCED BY THE OVARY

Transforming Growth Factor-β Superfamily

The transforming growth factor-β (TGF-β) superfamily, which includes TGF-β molecules, activins, inhibins, anti-Müllerian hormone (AMH), and the oocyte-derived proteins growth differentiation factor-9 (GDF-9) and bone morphogenetic protein-15 (BMP-15), play critical roles in ovarian function. These peptides mainly act locally to control follicular function, but several—including inhibins, activins, and AMH—are released into the systemic circulation. Activins and inhibins are disulfide-linked dimers of subunits derived from larger preproteins. Activins are dimers consisting of two β-subunits, βA and βB, assembled into three different combinations, namely activin A (βA-βA), activin B (βB-βB), and activin AB (βA-βB). One unique α-subunit is complexed to either of the two β-subunits to give two isoforms of inhibin, inhibin A (α-βA) and inhibin B (α-βB). In addition to their role as intraovarian regulators (described later in this chapter), inhibins appear to have important functions in the control of FSH secretion (see Chapters 118 and 120). Inhibins and estrogens inhibit, whereas gonadotropin-releasing hormone (GnRH) stimulates FSH secretion by the anterior pituitary gland. Although infusion of activin A has been shown to stimulate FSH secretion in monkeys, it is believed that circulating activins are rendered inactive by binding to follistatin, a single-chain glycoprotein that does not share significant homology with the actin/inhibin α- or β-subunits. Less is known about the regulation of inhibins, although an alternative cleavage site within the α-subunit releases a pro-peptide which appears to limit the synthesis and bioactivity of inhibins.[35,36]

The two different forms of inhibin display a divergent secretory pattern. Circulating inhibin B levels closely parallel serum FSH levels, rising in the early follicular phase and then declining, followed by a midcycle peak and a subsequent decline during the luteal phase. In contrast, serum inhibin A levels are low during the early follicular phase and rise with the growth of the dominant follicle. In the luteal phase, the granulosa and theca-luteal cells of the corpus luteum are a source of inhibin α-subunits, with the corpus luteum primarily secreting inhibin A.[37] Because inhibin B is produced by growing small antral follicles, low circulating levels of inhibin B are correlated with diminished ovarian reserve. Both inhibin A and inhibin B are used as granulosa cell tumor markers.

AMH is produced as a preprohormone and must be processed to become the active.[38] AMH is cleaved by furin, or possibly related convertases, at a conserved N-terminal site. Mutations that eliminate this site result in an inactive hormone and persistent Müllerian duct syndrome in males.[38] Following cleavage, the pro-region remains associated with the mature form of AMH until it binds the AMH receptor, and may even play a role in receptor binding.[39] Both AMH cleavage, and the subsequent competence of cleaved AMH, appear to vary in different body fluids. For example, cleaved AMH is significantly more capable of binding its receptor in follicular fluid than in serum, resulting in lower proportions of cleaved AMH required to exert an effect. Conversely, cleaved AMH is much more prevalent in serum and much less capable of binding its receptor. This raises the possibility of a mechanism by which cleaved AMH is rendered inactive in serum, and suggests that the level of cleaved AMH in biological fluids does not necessarily reflect AMH activity.[40]

Relaxin

Relaxin is an important hormone for maternal adaptation to pregnancy. In humans, there are seven relaxin genes; relaxins 1 to 3 and insulin-like peptides (INSLs) 3 to 6. Relaxin 2 is produced by the corpus luteum, mainly during pregnancy. All the relaxin family members are two-chain peptide hormones processed by cleaving a prohormone. Immunohistochemical studies have localized relaxin 2 to large luteal cells but not small luteal cells. During a non-conception cycle, luteal relaxin immunostaining was low in the early luteal phase; it increased progressively, reaching the highest level in late luteal phase, and then decreased greatly in corpora albicantia.[41] The pattern of relaxin secretion by the corpus luteum of pregnancy mirrors its steroidogenic function, with levels reaching peak concentrations between 8 and 12 weeks of gestation, declining subsequently to levels that are maintained until term. Human chorionic gonadotropin (hCG) appears to be the primary stimulus for luteal relaxin secretion, which is independent of progesterone secretion. INSL3 is the most studied of the other family members; it is involved in both follicle growth and thecal androgen production.[42] It acts in an autocrine/paracrine manner within the follicles to drive the production of androstenedione and its conversion to estrogens by granulosa cells, which in turn promotes expression of more INSL3 in a positive feedback loop, a system countered by BMPs and the LH surge. This system governs estradiol production within an antral follicle through the follicular phase. Higher levels are seen in women with polycystic ovary syndrome (PCOS), and lower levels in women with diminished ovarian reserve.[42] Although other INSLs are expressed in the ovary, their functions are not well established.

KEY POINTS

- The ovary is a dynamic enocrine organ that produces both lipid and peptide hormones.

GONADOTROPIN CONTROL OF OVARIAN STEROID PRODUCTION

Follicle-Stimulating Hormone and Luteinizing Hormone

Gonadotropins control the growth and differentiation of the steroid hormone–secreting cells of the ovary, intrinsically linking form and function. A defined sequence of gonadotropin action propels the growth of follicles and the production of steroid hormones. Positive feedback on the pituitary gland by high concentrations of estrogen leads to the ovulatory surge of LH, which in turn triggers a dramatic differentiation event, resulting in structural reorganization of the preovulatory follicle, release of the ovum, and striking changes in the steroidogenic capacity of the luteinizing cells (see Chapter 120).

Follicular growth, which culminates in ovulation and corpus luteum formation, requires both FSH and LH. Steroidogenic competence of the ovarian follicle is not achieved in the absence of FSH, even if LH is present in abundance. FSH promotes proliferation of the granulosa cells and increase in follicle size, and it induces the expression of genes involved in estradiol biosynthesis.[43] Mutations resulting in inactive FSH β-subunit or inactivation of both alleles of the FSH receptor gene produce symptoms of primary hypogonadism with deficient ovarian estrogen production in the face of elevated plasma LH levels. Likewise, FSH cannot achieve complete regulation of follicular development by itself. In gonadotropin-deficient women or LH-deprived monkeys treated with FSH alone, follicles grow to a preovulatory size, but because of a relative deficiency in androgen precursors, there is no parallel increase in estradiol production.[44,45] The small amount of estradiol that is made in this situation may be derived from androgens secreted by the adrenal cortex or thecal androgens produced in response to paracrine factors released by the granulosa cells. During the last phase of follicular maturation, when granulosa cells acquire LH receptors, LH alone can sustain follicular estradiol synthesis.[45] This LH substitution is thought to compensate for the diminished levels of FSH of the late follicular phase and allow LH action on the granulosa to rescue the dominant follicle from the fate of atresia of the remainder of the cohort.[45,46]

LH stimulation is also indispensable for normal luteal hormone production. Suppression of LH release leads to a prompt decline in progesterone levels that precede changes in the abundance of mRNAs encoding steroidogenic enzymes and structural luteolysis. This acute regulation of ovarian progesterone secretion is controlled by LH via the expression of StARD1. StARD1 mRNA and protein, present in both theca-luteal and granulosa-luteal cells throughout the luteal phase, are highly expressed in the early and midluteal phases, whereas declining StARD1 mRNA and protein levels are characteristic of the late luteal phase. Moreover, StARD1 protein levels in the corpus luteum are highly correlated with plasma progesterone levels. Suppression of LH levels during the midluteal phase markedly decreases plasma progesterone levels and abundance of StARD1 mRNA transcripts in the corpus luteum.[47] In addition, primate luteal cells activate cholesterol efflux mechanisms during luteolysis.[48]

Mechanisms of Gonadotropin Action

Activation of the LH receptor results in a rapid increase in steroid production from the ovary. This is due to post-translational changes to existing proteins such as the phosphorylation and activation of CEH and the phosphorylation of StARD1, actions mediated mainly by proteins phosphorylated by the cAMP-dependent protein kinase (PKA) and other kinases, including extracellular signal-regulated kinase (ERK) 1/2.[2] This results in a rapid enhancement of cholesterol delivery to the inner mitochondrial membranes.

Longer-term regulation of steroid hormone production over hours or days is exerted at the level of gene transcription. Increases in cAMP concentrations stimulate transcription of virtually all of the genes involved in steroid hormone synthesis. Although these genes do not all contain consensus cAMP response elements in their promoters, steroidogenic factor-1 (SF-1), encoded by the *NR5A1* gene, is involved in the control of most of the steroid synthetic genes. SF-1–binding elements are present in the promoters of the *StARD1, CYP11A1, HSD3B2, CYP17A1,* and *CYP19A1* genes.[49] The interaction of SF-1 with other transcription factors, including cAMP response element–binding protein (CREB)–binding protein (CBP), its coactivators, and steroid receptor coactivator-1 (SRC-1), may coordinate the cAMP response.

SF-1 expression decreases after ovulation, but mRNA levels of another orphan nuclear receptor controlling steroidogenic genes, liver receptor homologue - 1 (LRH-1) (encoded by *NR5A2*), are higher in corpus luteum than in mature follicles. Unlike SF-1, the expression of LRH-1 is induced significantly in the corpus luteum during pregnancy.[50] The stage-specific expression of LRH-1 in the ovary implies its specific role in regulating follicular and luteal functions. Transfection of granulosa cells with LRH-1 induces dose-dependent *HSD3B2* promoter activity.[51] Mice lacking LRH-1 expression in granulosa cells fail to ovulate, and expression of the *Stard1* and *Cyplla1* genes is impaired.[52] Collectively, these data demonstrate the importance of LHR-1 in luteal steroidogenesis.

The GATA family of transcription factors has recently emerged as an important group of transcription factors regulating expression of multiple steroidogenic enzymes. This basic helix-loop-helix (bHLH) family includes GATA-4 and GATA-6, which have roles in regulating the expression of *StARD1*, *CYP11A1*, and *CYP17A1*, among other steroidogenic enzyme genes.[53] The GATA family of transcription factors also interacts with other transcription factors, including SF-1 and Sp1, as part of a combinatorial code that regulates steroidogenic enzyme gene expression.

FSH and LH mediate granulosa cell gene expression and differentiation via their respective cell surface receptors through three major signaling pathways, PKA, PKB and protein kinase C (PKC). FSH activates multiple pathways, including phosphatidylinositol 3-kinase (PI3K)/PKB, mitogen-activated protein kinase (MAPK)/ERK, and p38 MAPK.[54] Studies of human granulosa-like tumor cells demonstrate that PKA is the main protein kinase activated by FSH, and that cAMP pathways are mainly mediated by PKA. PKA in turn regulates many factors which promote early follicular differentiation and inhibit cell proliferation.[55] In fact, even in the absence of FSH, expression of a constitutively active PKA subunit qualitatively mimics most of the transcriptional regulation exerted by FSH, with even more robust upregulation of transcripts involved in ovulation.[56] PKB is activated via PKA-dependent and PKA-independent signaling, and appears to play a role in regulating up to 60% of the FSH-dependent genes in granulosa cells through one of its major targets, forkhead box protein O1 (FOXO1). Stimulation of granulosa cells by FSH in the presence of PKB led to an amplification of FSH-induced aromatase and LH receptor mRNA levels, and its role in the synthesis of estradiol may have clinical relevance in controlled ovarian hyperstimulation treatments for fertility.[55]

The transcription of genes encoding steroidogenic enzymes is also suppressed at specific stages of follicular and luteal function. For example, after the LH surge, the expression of thecal P450c17 and granulosa aromatase and 17β-HSD decline transiently. Activation of the LH receptor also activates the inositol phosphate cascade, which is able to reduce aromatase expression and enhance progesterone production. Aromatase is similarly suppressed with the activation of PKC or expression of the constitutively active form of $G\alpha_q$, but mutant forms of the LH receptor that do not activate the phosphoinositol cascade do not inhibit aromatase.[57] PKC appears to function in parallel with the PKA pathway in later stages of follicular maturation, regulating several inflammatory factors which are suspected to facilitate ovulation in response to the LH surge, and others involved in granulosa cell luteinization.[55]

Factors that block expression of steroidogenic proteins include the orphan nuclear receptor, DAX-1, encoded by the *NR0B1* gene, which interacts directly with SF-1 to prevent activation of the transcriptional machinery. Stimulation of P450arom promoter activity by SF-1 in ovarian granulosa cells is inhibited by DAX-1 and WT1; all three transcription factors are expressed in human granulosa cells, indicating finely coordinated regulation of P450arom expression.[58,59] Post-transcriptional control (e.g., changes in mRNA stability) is also important in certain circumstances, and may account in part for the actions of insulin-like growth factors (IGFs) and small RNAs on the expression of components of the steroidogenic machinery.[59]

Other Hormones with Possible Actions in the Ovary

Besides FSH and LH, other hormones have been reported to exhibit gonadotropic activity. These include prolactin, growth hormone (GH), insulin, IGFs, GnRH, and kisspeptin. Although prolactin plays an important role in regulating luteal function in rodents, the physiological role of prolactin in human ovarian function is less certain. The human ovary expresses prolactin, which can be measured in follicular fluid in the long and short prolactin receptor isoforms. Although high levels of prolactin can suppress granulosa cell estrogen production, the menstrual disturbances observed in hyperprolactinemic women are primarily due to suppression of pituitary gonadotropin secretion.

GH has been reported to have direct effects on animal and human granulosa cells. Its cognate receptor is expressed on granulosa cells of human antral follicles and luteal cells. In addition, GH administration increases the ovarian response to gonadotropins in hypopituitary- or growth hormone–deficient individuals. Yet, supplemental GH does not appear to have dramatic effects on gonadotropin action in individuals with normal pituitary function.

Ovarian theca and granulosa cells express the insulin receptor through which insulin augments steroid production. The mechanisms underlying the insulin effect on steroidogenesis remain to be clarified. Insulin can increase steady-state levels of steroidogenic enzyme mRNAs, but it is not yet known whether this is a transcriptional or post-transcriptional response. To date, insulin response elements have not been identified in steroidogenic enzyme genes. There is, however, evidence for insulin effects on mRNA stability. Additionally, insulin is known to rapidly activate translation, but it is yet to be determined if such a translational action takes place in ovarian cells.

Kisspeptin (KISS1), acting through its G-coupled protein receptor (KISS1R), plays a critical role in female reproduction, controlling gonadotropin secretion, onset of puberty, ovulation, and the metabolic regulation of fertility. More recently, the action of kisspeptin/KISS1R in the ovary has been explored, and there is convincing evidence that it exerts a wide variety of effects in a number of mammalian species, including humans. KISS1 can both attenuate initial follicle recruitment and promote development of preovulatory follicles by downregulating FSH receptor and increasing circulating AMH, suggesting a role as a downstream target of LH signaling. In porcine and ovine models, treatment with kisspeptin-10 temporally increases GDF-9 and BMP-15 expression to mediate oocyte maturation. In murine and ovine models, kisspeptins have been shown to induce ovulation and potentially functions as a downstream target of prostaglandins (PGs), although KISS1R knockout models can still ovulate with extended GnRH and gonadotropin stimulation. Finally, KISS1 directly stimulates progesterone synthesis and secretion from granulosa cells by increasing mRNA expression of StARD1, P450scc, and 3β-HSD, and may play a role as a downstream mediator of neurokinin B (NKB) stimulation of estrogen production in granulosa cells, although this relationship requires further investigation.[60]

KEY POINTS

- Gonadotropins are essential regulators of steroidogenesis.

INTRAOVARIAN REGULATORS

While gonadotropins are the primary regulators of ovarian function, intraovarian factors modulate gonadotropin action. Their stage- and cell-specific actions orchestrate the complex dynamics of folliculogenesis and ovarian function over the life span. These intraovarian factors modulate gonadotropin receptor expression and function, as well as downstream signaling pathways. They have roles in follicle formation, granulosa and theca cell differentiation, basement membrane development, and other critical aspects of ovarian morphogenesis and function. They also likely account for differences in the rate of individual follicle growth, rescue of follicles from atresia, arrest and resumption of meiosis, dominant follicle selection, and the timing of key events of ovulation and luteolysis. The list of potential paracrine factors that can influence steroid production by theca and granulosa cells is long and diverse. It includes growth factors belonging to the TGF-β superfamily (e.g., AMH, GDF-9, BMP-15, activins, inhibins), cytokines, epidermal growth factor-related peptides (e.g., epiregulin, amphiregulin, and betacellulin), platelet-derived growth factor, fibroblast growth factors, TGF-α, IGFs, INSL3, hypothalamic peptides (e.g., kisspeptin and GnRH) estradiol, and progesterone (Table 119.1).[61–63]

TABLE 119.1 Intraovarian Regulators That Affect Ovarian Hormone Production

Growth Factor/ Cytokine	Ovarian Actions
Activin	Stimulates follicular growth and maturation and FSH receptor expression Inhibits thecal steroidogenesis
Adiponectin	Regulates granulosa steroidogenesis
Chemerin	Inhibits granulosa steroidogenesis
EGF	Stimulates mitosis of granulosa cells Inhibits expression of aromatase
FGF	Stimulates mitosis of granulosa cells Inhibits steroidogenesis
GDF-9	Increases follicular growth Regulates granulosa cell steroidogenesis Enhances thecal development
GnRH	Inhibits steroidogenesis Activates oocyte maturation
IFN-γ	Decreases granulosa cell steroidogenesis
IGF-1, IGF-2	Stimulates granulosa cell and theca cell mitosis and steroidogenesis, augments gonadotropic-stimulated steroidogenesis
IL-1	Inhibits gonadotropin-stimulated steroidogenesis
IL-2	Augments progesterone production
IL-6/IL-11/OSM/LIF	Suppresses FSH-stimulated steroidogenesis
Inhibin A, inhibin B	Increase thecal androgen synthesis
INSL3	Stimulates thecal androstedione production
KGF	Enhances luteal progesterone production Stimulates granulosa cell division Increases inhibin-α expression
Kisspeptin	Follicular development, oocyte maturation and ovulation, granulosa cell progesterone production
Leptin	Low dose enhances steroidogenesis
PDGF	Higher doses suppress estrogen and progesterone production Augments FSH-induced LH receptor expression and steroidogenesis, enhances theca proliferation
Prohibitin	Knockdown in granulosa cells is associated with suppressed StARD1, P450scc, 3β-HSD, and aromatase activity, decreased FSH-mediated steroidogenesis
Resistin	Stimulates some steroidogenic enzymes Stage-specific IGF modulation
TGF-β	Augments FSH-mediated steroidogenesis Augments FSH receptor expression Inhibits thecal steroidogenesis
TNF-α	Inhibits gonadotropin-stimulated steroidogenesis

EGF, Epidermal growth factor; *FGF,* fibroblast growth factor; *FSH,* follicle-stimulating hormone; *GDF-9,* growth and differentiation factor 9; *GnRH,* gonadotropin-releasing hormone; *IFN,* interferon; *IGF,* insulin-like growth factor; *IL,* interleukin; *KGF,* keratinocyte growth factor; *LH,* luteinizing hormone; PDGF, platelet-derived growth factor; *TGF,* transforming growth factor; *TNF,* tumor necrosis factor.

Cholesterol Precursors

Intermediates in the cholesterol biosynthetic pathway from lanosterol have been found to induce oocytes to resume meiosis. The compound 4,4-dimethyl-5α-cholesta-8,14,24-triene-3β-ol was extracted from human follicular fluid and named *follicular fluid meiosis-activating substance (FF-MAS)*. A related compound, 4,4-dimethyl-5α-cholest-8,24-diene-3β-ol, was isolated from bull testis and called *T-MAS*. Present in micromolar concentrations in preovulatory follicular fluid, both compounds are synthesized from lanosterol by P450 14α-demethylase encoded by the *CYP51* gene. The accumulation of FF-MAS and T-MAS in mature follicles may be the result of increased synthesis or inhibition of subsequent steps in cholesterol synthesis. Gonadotropins have been reported to cause a several-fold increase in *Cyp51* gene expression in rodent ovaries, which could contribute to enhanced MAS formation, and blocking CYP51 function reduced the ability of FSH to induce meiosis in cultured rat follicles.[64] However, the physiologic roles of FF-MAS and T-MAS in human oocyte maturation *in vivo*, if any, remain uncertain at this time.

Role of Estrogens

In addition to their systemic effects on the reproductive tract, hypothalamus, and pituitary gland, estrogens act on granulosa, theca, and luteal cells in the ovaries of laboratory animals and domestic species. Not only are estrogen receptors (ERs) expressed in the ovary, but high levels of estrogens are also characteristically found in the antral fluid of healthy mature follicles. There are conflicting reports in the literature regarding the expression of the two ER subtypes, ERα and ERβ, in the primate ovary.[65] The most convincing of these studies indicates that both ER subtypes are expressed in the germinal epithelium, and granulosa cells, with ERβ predominating over ERα in medium-sized and preovulatory follicles, theca cells, and luteinized granulosa cells. ERα transcripts have also been detected by polymerase chain reaction in human oocytes by some researchers, but these findings have not been confirmed by others. The discrepancies among various reports likely reflect sensitivities of the detection methods used. Existing data support the notion that the ovary is a site of estrogen action via the classical receptor-mediated signaling pathways.

Given the high intraovarian estrogen concentrations achieved during follicular maturation and corpus luteum function, the physiologic roles of estrogen within the primate ovary and the mechanisms by which they might influence cellular function (i.e., genomic versus nongenomic actions) are matters of debate.[65] Indeed, the extremely high levels of estradiol reached in the antrum of the preovulatory follicle (~1 μg/mL) raise serious questions as to the roles of the classical ERs, which would be fully saturated by ligand during the later stages of follicular maturation. Estrogens have pleiotropic effects on animal granulosa cells. They promote proliferation and exert anti-atretogenic effects. Estrogens also augment intercellular gap junction and antrum formation, and increase the ER content of granulosa cells. Estrogens synergize with gonadotropins at several levels (e.g., promotion of ovarian growth, LH and FSH receptor expression) and augment aromatase activity. However, high levels of estrogens are not essential for the growth of follicles to the size equivalent to the preovulatory stage.[66] There are pharmacologic data suggesting that estrogens are important for oocyte function.[67] *In vitro* studies on primate granulosa cells have yielded inconsistent findings with respect to the actions of estrogens. Estradiol inhibits progesterone secretion by rhesus monkey granulosa cells, whereas in marmoset granulosa cells it has no effect on progesterone production, but stimulates aromatase in the presence of IGF-1. Exogenous estrogen exerts a luteolytic effect on the primate corpus luteum, probably through actions on the central nervous system.

In summary, although the primate ovary expresses the receptors that allow a variety of cells to respond to estradiol, the physiologic significance of estrogens in follicular maturation and luteal function in the primate ovary remain to be elucidated. It is evident that follicle growth per se does not require high levels of estradiol, but the orchestration of events that result in a mature oocyte capable of developing into a viable embryo following fertilization may require estrogen action.

Estrogen metabolites (discussed earlier in this chapter) have been proposed to be paracrine/autocrine regulators of steroidogenesis and follicular development. A balance of pro-angiogenic and anti-angiogenic estrogen metabolites appears to be necessary for follicular development. Furthermore, 4-hydroxyestrone, 16-ketoestradiol, and 2-hydroxyestradiol have proliferative and pro-angiogenic actions,[68] whereas 2-methoxyestradiol and 2-methoxyestrone have anti-angiogenic activities. The anti-angiogenic activity of 2-methoxyestradiol can be ascribed to inhibition of the transcription factor, hypoxia-induced factor-1α,[69] which may have a role in controlling neovascularization of the developing corpus luteum and changes in luteal vascular during luteolysis.[70] Genetic variants of the *COMT* gene associated with decreased 2-methoxyestradiol have been associated with premature ovarian insufficiency.[71] Alterations in the balance of pro-/anti-angiogenic estrogen metabolites are found in the follicular fluid of anovulatory women with PCOS.[68]

Role of Androgens

In the human, infra-human primate, and rat ovary, ARs are expressed mainly in granulosa cells and to a lesser extent in theca cells and stroma, suggesting that in addition to serving as substrates for estrogen production, thecal androgens may have paracrine and autocrine functions. Androgens have a number of effects on the primate ovary. It has been proposed that androgens have complex actions on granulosa cells, amplifying the effects of FSH during intermediate stages of follicular development, but potentially inducing atresia during the late preovulatory period. Administration of testosterone or DHT to rhesus monkeys promotes activation of primordial follicles, as well as increased granulosa cell division in small follicles, suggesting a folliculotropic action. Moreover, ARs were positively correlated with a marker of cell proliferation (Ki-67) and negatively correlated with apoptosis. These observations are in sharp contrast with the notion that androgens are atretogenic, a concept that emerged primarily from studies on the rodent ovary in which androgens block granulosa cell proliferation *in vitro* and promote follicular atresia. Taken together, these discrepancies suggest species-specific androgen actions in the ovary.

Stage-dependent effects of androgens on granulosa-cell function *in vitro* have been reported in the marmoset.[72] Androgens enhanced FSH-stimulated aromatase expression and progesterone production, while inhibiting hCG-stimulated aromatase activity and progestin synthesis in cells from large preovulatory follicles. In the rhesus monkey, high levels of DHT given after pituitary downregulation, extending into to a period of exogenous gonadotropin stimulation, resulted in a marked decrease in estrogen production and lower ovarian weights.[73] In mice, there appear to be dose-dependent actions, with low levels of androgen inducing rapid follicular maturation and high levels of androgen inhibiting follicular growth and oocyte maturation.[74] Evidence that androgens have a detrimental effect on human follicular function includes the observation that follicular fluid enriched in DHT and poor in estradiol is characteristic of atresia. However, this steroid profile may be a consequence rather than a cause of atresia. Favoring a causal relationship are reports that high follicular concentrations of 5α-reduced androgens such as DHT act as competitive inhibitors of granulosa cell aromatase activity.[75] Thus, androgens may exert both positive and negative effects on follicular growth and function in a stage- and dose-dependent manner through ARs, as well as by non–receptor-mediated mechanisms.

There has been much discussion in recent years about the possibility of using androgen priming to increase oocyte yield and quality in IVF cycles by increasing the antral follicle pool prior to hyperstimulation in women with diminished ovarian reserve. The logic and caveats to this approach are reviewed by a recent article from the Andersen group.[76] Though systemic androgen can influence smaller follicles, it does not appear to alter follicular fluid concentrations of androgen in larger follicles. Clinical trials employing treatments to influence intrafollicular androgen levels are ongoing. As data emerges from these trials and analysis of systemic versus local effects of androgen on intact human follicles is completed, we may gain a better understanding of the specific roles of androgens in the ovary.

Role of Progesterone

Progesterone plays a key role in the periovulatory period and appears to be essential for the normal luteinization process and maintenance of luteal function. Ovulation does not occur in mice with targeted deletions of the progesterone receptor (PR) gene (the A form of the receptor), or in animals treated with the PR antagonist, RU486. A number of genes encoding enzymes implicated in ovulation that appear to be dependent on periovulatory progesterone have been identified, including those encoding metalloproteinases and cathepsin L. Pharmacological blockade of ovarian progesterone production with a 3β-HSD inhibitor indicated that progesterone exerts antiapoptotic and pro-differentiation effects on luteinizing cells and maintains luteal function.

In women, granulosa cells from preovulatory follicles do not contain significant amounts of PRs, although progesterone may play a role in inhibiting granulosa cell differentiation and steroidogenic activity through PR membrane component 1 (PGRMC1) in rats.[77] Granulosa cells of the dominant follicle, however, highly express PR at the time of the LH surge.[78] After follicle rupture, PR expression persists in the luteinized granulosa cells of the corpus luteum,[79] with mRNA concentrations of both PR isoforms (PR-A and PR-B) at the early and midluteal phases being higher than those at the late luteal phase.[79] These data are concordant with the hypothesis that progesterone is a “local luteotropin” that promotes luteal development and sustains luteal structure and function.[80]

Progesterone also appears to play a role in oocyte maturation. In rhesus monkeys treated with the progestin R5020, a significant percentage of oocytes resumed meiosis and entered metaphase II despite the absence of a gonadotropin surge or follicle rupture.[81] Collectively, the cited studies indicate that progesterone produced in response to the midcycle LH surge is important for luteinization of the dominant follicle, oocyte maturation, ovulation, and subsequent maintenance of luteal function.

Transforming Growth Factor-β Superfamily

Inhibins are produced mainly by granulosa cells during the follicular phase and augment LH-stimulated thecal androgen synthesis, but they have no direct effect on granulosa cell steroidogenesis. However, FSH-stimulated inhibin production by granulosa cells indirectly augments estrogen production by stimulating thecal androgen production for use as estrogen precursors in the granulosa cells. Since FSH also increases the expression of P450arom and NADPH cytochrome POR, the increased substrate is readily metabolized to estrogen by the dominant follicle.[82]

Activins, primarily produced by granulosa cells, have more complex actions in the follicle. Both activin and TGF-β exert their effect, in part, through a canonical Smad pathway, activating Smad2 and Smad3 which then interact with Smad4 and translocate to the nucleus to exert a variety of stage-specific effects.[83] They potentially play a role early in follicular growth by stimulating the expression of FSH receptors, aromatase and inhibin by granulosa cells. TGF-β also increases the expression of the inhibitory Smad7, which provides a negative feedback loop and promotes granulosa cell apoptosis.[83] Activins suppress thecal androgen production and block the stimulatory effects of inhibin on thecal steroidogenesis. Consequently, activins can keep thecal androgen synthesis in check during early follicular development. Activins also increase the production of IGF-binding proteins (IGFBPs) IGFBP-2 and IGFBP-4 by granulosa cells, which, in turn, modulate IGF effects. In primates, inhibin A is not expressed in small follicles, but gradually increases in enlarging follicles. This results in a shift from an activin-dominated environment to one that is progressively more dominated by inhibin.[83,84] Follistatin is a monomeric protein first identified in and isolated from follicular fluid. An activin-binding protein, it is produced by granulosa cells under FSH stimulation and blocks the actions of activin. Serum levels of follistatin do not fluctuate across normal menstrual cycles. Likewise, intrafollicular concentrations of free follistatin do not vary with size or maturity of the follicle. However, follistatin levels are increased in women with PCOS,[85] and female transgenic mice overexpressing follistatin exhibit small ovaries due to a block in folliculogenesis at various stages, confirming its role in follicular development.[86]

AMH is expressed in granulosa cells of preantral and small antral follicles of the ovary. Expression of AMH receptor has been localized to granulosa and theca cells of growing follicles of the rat ovary.[87] Compared with wild-type controls, ovaries of prepubertal AMH-null mice contained more growing follicles, whereas ovaries of adult AMH-null mice had significantly fewer primordial follicles.[88] By 13 months of age, AMH-null ovaries were almost depleted of primordial follicles. Collectively, these observations suggest an indirect role for AMH in inhibiting the recruitment of primordial follicles to the growing pool. AMH appears to prevent entry into the growth phase by coordinated regulation of growth and inhibitory genes, and normal follicular recruitment and ovulation are dependent on a thecal androgen-induced reduction in production of AMH from granulosa cells. It also inhibits aromatase induction by FSH in granulosa cells, and on the basis of *in vitro* studies conducted on a murine Leydig cell line, it is likely that AMH inhibits theca cell androgen production.[89]

In women, serum AMH levels are positively correlated with the number of small antral follicles detectable by ultrasonography, but are negatively correlated with age. As antral follicle count declines with age, reflecting diminishing ovarian reserve, decreased follicular production of inhibin B leads to reduced pituitary suppression of FSH. The rapid decline in AMH with age is purported to be related to this rise in FSH, either by direct inhibition or indirectly by inhibiting oocyte-specific growth factors GDF-9 and BMP-15.[90] Women with PCOS, who have increased numbers of small growing follicles, have elevated AMH levels. The persistently elevated levels of androgens in these women may disrupt the inhibitory role androgens play in decreasing AMH to promote follicular recruitment.[91] Women on hormonal contraceptives also can have lower AMH levels,[92] although women with ovarian insufficiency due to *FSHR* mutations have near normal AMH levels.[93] Clinical studies suggest that AMH can be a good serum marker of ovarian reserve, although best utilized as a predictor of ovarian response to exogenous gonadotropins in fertility therapies. Note that standardization of existing assays remains to be accomplished, and normative values need to be further validated across age ranges and other conditions such as obesity, that may influence results.[94] AMH secretion is thought to be independent of the ovarian cycle. However, some reports suggest that AMH levels can vary across the follicular and luteal phases.[95,96]

The oocyte also participates in the paracrine dialogue involving the somatic cells in the follicle. GDF-9 is highly expressed by growing oocytes, and to a lesser extent by primate granulosa cells. Disruption of the *Gdf9* gene in mice prevents follicle development beyond the primary/early secondary stage. The follicles of GDF9-deficient mice arrest in growth at the early secondary stage, yet their oocytes continue to grow at a faster rate than wild-type oocytes, and progress to advanced stages of differentiation seen in the antral follicles of normal mice. However, there are ultrastructural abnormalities in the interconnections between granulosa cells and oocytes. The theca also fails to form around the follicles, implicating GDF-9 in the organization and/or proliferation of this follicular component.

In vitro, GDF-9 has a variety of species-specific effects on granulosa and theca cells.[97] In rodents, GDF-9 stimulates granulosa cell differentiation, including induction of LH receptors and steroidogenesis. In early preantral follicles, GDF-9 induces follicle growth and thecal differentiation.[98] In cumulus cells, GDF-9 promotes expression of proteins that are incorporated into the proteoglycan extracellular matrix of the cumulus oophorus and follicular fluid. It also stimulates COX-2 and PG synthesis and progesterone formation.[99] LH receptor expression is suppressed, which would discourage luteinization of the cumulus cells. These combined actions of GDF9 confer a unique phenotype to those granulosa cells immediately adjacent to the oocyte that become the cumulus cells with further follicle development. GDF-9 inhibits human granulosa-luteal and antral theca cell steroidogenesis *in vitro*, with the inhibitory effect on theca cells being more pronounced.[100] It also stimulates theca cell proliferation, a finding consonant with the apparent role of GDF-9 in controlling thecal development in the murine ovary.

BMP-15 is another member of the TGF-β superfamily produced by oocytes. It is structurally related to and shares a similar expression pattern with GDF9, with which it can heterodimerize. Although BMP-15 has no effect on FSH-stimulated estradiol production, it decreases FSH-induced progesterone production, indicating that BMP-15 is a selective modulator of FSH action. *In vitro*, BMP-15 stimulates granulosa cell mitosis. Mutations in the human *BMP15* gene that reduce mature protein production, activity, or synergy with GDF-9 cause primary ovarian insufficiency, as do missense variants in the BMP receptor genes, *BMPR1A* and *BMPR1B*.[101,102]

Eicosanoids

PGs are members of the eicosanoid family of biologically active molecules derived from arachidonic acid, which is in turn mobilized from cell membrane phospholipids mainly via the action of phospholipase A2 (PLA2). Arachidonic acid is converted to PGs by the two isozymes of PG endoperoxide synthase-1 and -2 (PTGS-1/-2, also known as cyclooxygenase-1/-2, COX-1 and COX-2). COX-1 is constitutively expressed at low levels, but induction of COX-2 in granulosa cells before ovulation is highly conserved across many mammalian species. COX-2 knockout mice are infertile due to ovulation failure.[103] The preovulatory gonadotropin surge markedly increases COX-2 expression and concentrations of both PGE_2 and $PGF_2\alpha$ in follicular fluid. Transgenic mice lacking the PGE_2 receptor subtype EP2 show decreased ovulation and exhibit abnormal cumulus expansion.[103] A role for $PGF_2\alpha$ in ovulation is unclear, but it has a well-established role as a physiologic luteolysin in animals. For example, administration of $PGF_2\alpha$ to ewes decreases serum progesterone levels and mRNA abundance of both StARD1 and 3β-HSD type II in the corpora lutea.[104]

Insulin-Like Growth Factors and Their Binding Proteins

In response to FSH, granulosa cells produce IGFs as well as IGFBPs, which have higher affinities for the IGFs than the receptors that mediate IGF action.[61–63] IGF-1 is the primary IGF produced in animal follicles, whereas IGF-2 is the major IGF in the human ovary. Both IGFs bind the type I IGF receptor (IGF1R), which is a tyrosine kinase receptor similar to the insulin receptor and is responsible for IGF downstream signaling in the cell. An *in vivo* IGF1R knockdown model in mice demonstrates its central importance in the regulation of follicle growth, and the survival and differentiation of granulosa cells. Mice without functional IGF1R in granulosa cells showed increased apoptosis at all levels of follicular development, lack of responsiveness to FSH (without alteration in FSH receptor expression), and lower levels of preovulatory aromatase, LH receptor, and inhibin α.[105] The type II receptor has a very different structure with only a single transmembrane domain and a small intracytoplasmic component. It does not transmit signals but binds IGF2 and controls its bioavailability. In general, IGFs amplify the actions of FSH on granulosa cells, including effects on cell proliferation, cAMP accumulation, and transcriptional activation of some genes, including the *StARD1* gene. Similarly, IGFs increase theca cell androgen and progesterone production and amplify the stimulatory effects of LH.[106]

The pattern of expression of IGFBPs is determined by the stage of follicular maturation and the status of the follicle.[61–63] IGFBP-2 is expressed by theca and granulosa cells of small antral atretic follicles, whereas IGFBP-3 is expressed by theca and granulosa cells of the dominant follicle. IGFBP-4, expressed by ovarian stroma, theca cells, and the granulosa cells of small antral follicles and the dominant follicle, is a potent inhibitor of IGF action *in vitro*. Inactivation by cleavage by the IGFBP-4 protease, also known as *pregnancy-associated plasma protein-A* (PAPP-A), acts to increase IGF bioavailability. PAPP-A is secreted by granulosa cells of large follicles but not small follicles. Its proteolytic activity is modulated by stanniocalcins (stanniocalcin-1 and -2), which are potent PAPP-A inhibitors.

Healthy follicles are characterized by a microenvironment in which IGFs are present in the unbound form.[107] The high levels of free IGF are achieved by both reduced expression of IGFBP-2 and increased inactivation of IGFBP-4. In contrast, atretic follicles have higher levels of both IGFBP-2 and IGFBP-4, which restrict bioavailability of IGF. The IGF system represents a prototype for amplifying the effects of FSH during the follicular phase, permitting the dominant follicle to survive in the face of declining FSH levels.

Other Intraovarian Regulators

Oxytocin is implicated in the regulation of gap junction formation between luteal cells, stimulation of estradiol synthesis, and the process of luteolysis. Primate corpora lutea have the highest contents of oxytocin during the midluteal phase, and oxytocin receptor transcripts are detectable in the corpus luteum at this time. However, the observation that oxytocin knockout mice do not show abnormalities in corpus luteum function suggests that oxytocin is not a major factor in controlling the rodent corpus luteum.

The proinflammatory cytokines IL-1α and IL-1β inhibit steroid production and suppress the steady-state levels of StARD1 and steroidogenic enzymes in ovarian cells.[108] These cytokines, together with their cognate receptor IL-1R1, are produced by resident ovarian macrophages and both granulosa and theca cells. Studies have shown multiple functions of the IL-1 system in the developing follicle. *In vitro*, IL-1 inhibited FSH-induced estradiol and progesterone production by granulosa cells and also gonadotropin-induced androstenedione production by theca cells. IL-1 treatment can increase PG accumulation in the rat ovary by stimulating the expression of two of the key enzymes in PG formation, PLA2 and COX-2. Although the physiological roles of these cytokines in the ovary remain to be fully characterized, evidence suggests that the IL-1 system participates in the control of folliculogenesis, ovulation, and luteal function.

In addition to playing a pivotal role in the hypothalamus-pituitary-gonadal axis, GnRH also regulates steroidogenesis in the ovary as a local autocrine/paracrine factor. Two isoforms of this decapeptide hormone have been identified, GnRH-1 and GnRH-2. mRNA transcripts of both GnRH isoforms and GnRH-1 receptor are expressed in granulosalutein and ovarian surface epithelial cells.[109] Progesterone secretion from granulosa cells in culture was decreased by the addition of a GnRH agonist, an effect reversed by the coadministration of a GnRH antagonist. At low concentrations, the GnRH agonist, buserelin, stimulates estrogen production in granulosa-lutein cells, but suppresses steroidogenesis at higher concentrations.[110] GnRH-1, GnRH-2, and GnRH-1 receptor mRNA expression levels in the ovary are regulated by multiple factors, including FSH, hCG, estradiol, and GnRH-1 and GnRH-2 themselves.

Many substances originally identified as adipokines because they are secreted by adipose tissue are also produced by ovarian cells or have receptors that are expressed in the follicles or the corpus luteum. Interaction between these metabolic regulators and regulation of reproductive functions has long been suspected, but ongoing characterization of the adipokines has provided insight into the mechanisms and complexity of these interactions. Various adipokine receptors are expressed in follicles and the corpus luteum, though regulation of the specific types can vary by species. Granulosa and theca cells have also been reported to produce some adipokine proteins endogenously, such as adiponectin and leptin. The expression patterns of the adipokines and their receptors are stage- and sometimes species-specific. Leptin, adiponectin, chemerin, and resistin have all been reported to regulate human ovarian steroid production.[111–114] They each can modulate insulin or IGF effects in granulosa or theca cells *in vitro*. Leptin and adiponectin have both been shown to modify progesterone production by luteal cells.[114]

Altered serum levels of adipokines have been described in women with PCOS, and follicular fluid levels of adipokines have been associated with egg retrieval and fertilization rates. Leptin is negatively correlated with embryo number, whereas resistin is positively correlated with it, but the correlations are not definitive enough to be used as predictors of oocyte health.[115] The exact roles of the various adipokines in human ovarian function are still not entirely clear, but there is ample evidence that this family of factors is involved in the intersection of fertility and energy balance and perhaps in the endocrine dysfunction of PCOS and obesity.

KEY POINTS

- Ovarian-derived steroids and peptide hormones function as key paracrine and autocrine regulators of ovarian function.

HORMONE PRODUCTION DURING THE OVARIAN LIFE CYCLE

The Fetal and Prepubertal Ovary

The human fetal ovary displays very limited steroidogenic activity. It contains no detectable StARD1 protein in theca or granulosa cells, and expresses very low levels of P450scc. Homogenates of fetal ovaries convert pregnenolone into progesterone, 17α-hydroxyprogesterone, and 5α-reduced compounds. However, androgens are not formed in any significant quantity, indicating the lack of 17,20-lyase activity. Consistent with these observations, P450c17 mRNA in the fetal ovary is less than 2% of the levels found in fetal testis. Homogenates of fetal ovaries can, however, aromatize androstenedione and testosterone at low rates. The paucity of 17,20-lyase activity, as well as low aromatase expression, would restrict fetal estradiol synthesis.[116]

The Ovary During Reproductive Life

The plasma levels of the two major female sex steroid hormones, estradiol and progesterone, change dramatically during the menstrual cycle, primarily reflecting their rate of secretion from the ovary. During the follicular phase, the theca of maturing follicles contributes 30% of the circulating androstenedione, increasing to 60% immediately prior to ovulation. The rising estradiol levels in the follicular phase are derived almost exclusively from the follicle selected for ovulation. Following ovulation, follicular concentrations of steroids downstream of P450c17 (estrogen and testosterone) are decreased, as is the capacity for conversion of androgens to estradiol.[117] The luteinization of the theca and granulosa cells gives rise to the corpus luteum, comprising granulosa-lutein (large) cells and theca-lutein (small) cells, each displaying a different repertoire of steroidogenic enzymes. Capable of progesterone production, both cell types express StARD1, P450scc, and 3β-HSD type II. Theca-lutein cells express P450c17 and appear to be the primary site of 17α-hydroxyprogesterone production as well as the source of androgen substrates to be converted to estrogens by the aromatase-expressing granulosa-lutein cells,[118] a compartmentalization of steroidogenic enzymes reminiscent of the theca-granulosa cooperation in estrogen synthesis during the follicular phase. Granulosa-lutein cells and theca-lutein cells have an intimate relationship in the formed corpus luteum, and their interactions may be promoted by the action of locally produced hormones, including oxytocin, which stimulates the formation of gap junctions between these cells, facilitating intercellular communication.

The steroidogenic activity of the corpus luteum is dependent upon LH stimulation.[119] Administration of a GnRH antagonist to nonhuman primates results in an immediate and profound decline in plasma progesterone levels and sharp reduction in ovarian StARD1 mRNA expression but without significant changes in P450scc and 3β-HSD mRNA levels. In monkeys with hypothalamic lesions that destroy GnRH-producing neurons, luteal function can be maintained by pulsatile administration of GnRH. Cessation of GnRH infusion results in a rapid fall in progesterone levels that can be restored by reinitiation of infusion. Interestingly, the functional life span of the corpus luteum in this experimental paradigm cannot be extended beyond the usual life of 14 to 16 days despite continued administration of GnRH. Only treatment with progressively increasing amounts of LH or hCG can "rescue" the corpus luteum from luteolysis.[120] These observations suggest that luteolysis in primates is the result of a loss of sensitivity to LH, presumably as a consequence of intraluteal events, and that decreased luteal sensitivity to LH can be overcome with higher levels of gonadotropin stimulation.

Excessive ovarian enlargement and increased production of estrogens occurs in ovarian hyperstimulation syndrome (OHSS), a condition most commonly associated with ovulation induction with exogenous gonadotropins. Very rarely, OHSS occurs spontaneously as a consequence of FSH receptor mutations in the transmembrane domains that increase sensitivity to hCG[121] or a consequence of excessive hCG levels (e.g., hydatidiform mole) or thyrotropin production secondary to profound hypothyroidism.

During early pregnancy, the corpus luteum, which is essential for maintenance of pregnancy during the initial 7 weeks, produces large amounts of progesterone. Secretory function then declines, as assessed by the measurement of 17α-hydroxyprogesterone levels, a steroid that is not secreted by the placenta and is, therefore, assumed to provide an index of luteal function (Fig. 119.4). In the absence of an hCG rescue signal, the corpus luteum undergoes a process of functional and structural regression that has been likened to an immune-mediated event characterized by apoptotic death of luteal cells as well as luteal cell necrosis/autophagy, processes that both contribute to structural luteolysis.[122]

Abnormal production of ovarian steroid hormones, particularly androgens, occurs rarely during gestation as a result of luteoma of pregnancy and hyperreactio luteinalis. Maternal virilization and occasional virilization of female fetuses may occur. Both conditions are benign and resolve during the postpartum period.

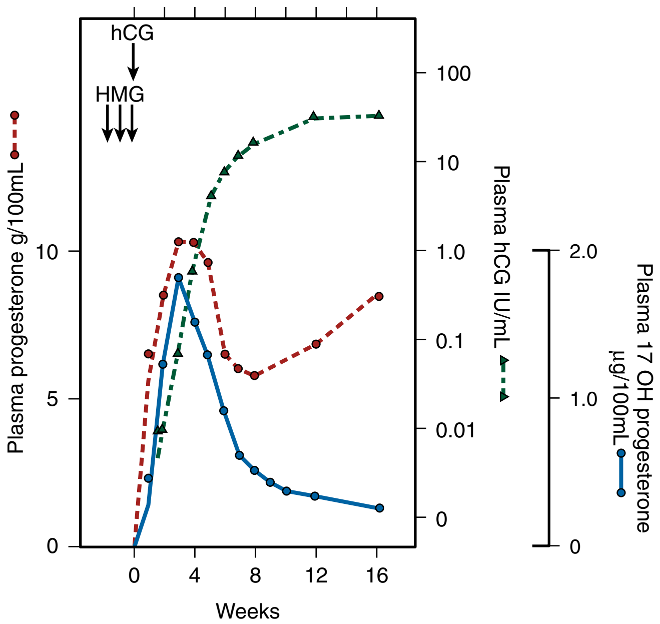


Figure 119.4 Mean Plasma Levels of Progesterone, 17-Hydroxy-Progesterone, and Human Chorionic Gonadotropin (hCG) Following Induction of Ovulation With Human Menopausal Gonadotropins (HMG) and hCG. (Data from Yoshimi T, Strott CA, Marshall JR, et al. Corpus luteum function in early pregnancy. *J Clin Endocrinol Metab.* 1969;29:225–230.)

Endocrine Activity of the Postmenopausal Ovary

The postmenopausal ovary contains two different populations of cells with steroidogenic capacity: hilar cells and stromal cells, including residual thecal elements. *In vitro* studies suggest that the postmenopausal ovary has some steroidogenic potential. Incubation of postmenopausal ovarian stromal slices with pregnenolone yielded progesterone, dehydroepiandrosterone, and testosterone. Incubation of strips of ovarian hilar tissue from postmenopausal women revealed a steroidogenic pattern similar to that of the postmenopausal ovarian stroma. However, the overall amount of steroid produced was substantially greater compared with stroma. Measurable *in vitro* formation of estradiol by postmenopausal cortical stroma and hilar cells has also been reported.[123,124] Immunohistochemical studies found that less than 1% of stromal cells showed evidence of expression of P450scc, 3β-HSD, and P450c17.

The postmenopausal ovary is thought to be a source of circulating testosterone, although there may be considerable variation among individuals with respect to androgen output. The circulating levels of testosterone in postmenopausal women are only slightly lower than those observed in premenopausal women. The postmenopausal ovary probably can contribute no more than 20% of the daily production of androstenedione, and multiple lines of evidence indicate that the adrenal glands are the major source of androstenedione in women. Although some authors have concluded that the postmenopausal ovary is a significant site of androgen biosynthesis, others have questioned this idea. Cauley et al.[125] examined hormone levels in postmenopausal women with and without ovaries and observed no statistically significant differences in circulating levels of testosterone or androstenedione between the two groups. Couzinet et al.[126] reported that plasma androgen levels were very low in all postmenopausal women with adrenal insufficiency and were similar between oophorectomized and non-oophorectomized postmenopausal women with normal adrenal function. Reportedly, P450scc mRNA levels were increased 3-fold in the ovarian stroma of postmenopausal women with endometrial cancer or hyperplasia, compared with postmenopausal women without endometrial pathology.[127] Although P450arom mRNA was not detectable in any of the ovaries examined, the same study found P450c17 mRNA expression by Northern analysis in the ovarian stroma of all women with endometrial cancer or hyperplasia, but only in two of the five control postmenopausal women, suggesting a potential to produce androgens in some postmenopausal ovaries. Collectively, these data suggest that the ovarian contribution to circulating androgens is variable among postmenopausal women.

Estrogens in postmenopausal women appear to arise almost exclusively from extra-glandular aromatization of androstenedione.[128] Oophorectomy results in no significant reduction in urinary estrogen excretion by postmenopausal women. However, adrenalectomy after oophorectomy virtually eliminates measurable estrogens from the urine. *In vitro* studies concluded that the postmenopausal ovarian stroma is unable to aromatize androgens. However, others have suggested that the postmenopausal ovary may synthesize limited amounts of estrogens because the concentrations of estradiol and estrone are two times higher in ovarian venous blood than in peripheral blood of postmenopausal women.[129]

There is some evidence that ovarian androgen production in postmenopausal women can be gonadotropin-dependent. Administration of hCG to postmenopausal women results in a small increase in circulating levels of testosterone.[130] Daily injection of hCG causes hyperplasia of the ovarian hilar cells and histochemical evidence suggestive of active steroidogenesis.[131] Administration of hCG, but not adrenocorticotropic hormone, resulted in increased androgen but not estrogen production by the ovaries.[132] Binding sites for both LH and FSH were identified in the cortical stroma and in hilar cells.[133] Addition of hCG to hilar cells results in increased cAMP formation and steroid biosynthesis, indicating preserved responsiveness to gonadotropins. Taken together, these observations suggest that ovarian androgen biosynthesis of the postreproductive ovary is at least partially gonadotropin-dependent.

The postmenopausal ovary is occasionally involved in pathologic endocrine activity. Stromal hyperplasia can occur, with the ovary enlarging with hyperplastic stromal nodules consisting of lipid-rich luteinized cells that resemble theca interna. The ovaries with stromal hyperplasia produce large amounts of androstenedione, resulting in hirsutism and virilization. Hilar cells can give rise to functional hilar cell tumors that produce excess amounts of androgens, leading to virilization.[134] Signs and symptoms of estrogen excess may also be evident in circumstances of significant peripheral aromatization. For instance, the finding of 3-fold higher P450scc mRNA levels in the ovarian stroma of postmenopausal women with endometrial cancer and endometrial hyperplasia suggests that aberrant steroid biosynthetic activity in the postmenopausal ovary might lead to these two disorders known to be associated with excessive estrogen stimulation.[127]

KEY POINTS

- The ovary is an active endocrine organ throughout the life span. Ovarian-derived products have clinincal utility as markers of ovarian function.

CONCLUSION

The ovary is a dynamic endocrine organ with manifold intrinsic and extrinsic actions depending on a complex interplay of signaling through the lifespan. A principal function in production of important cholesterol-derived reproductive hormones under the direction of gonadotropins is well-studied, though the precision with which they are modulated under specific physiologic and pathological states continues to be elucidated. Further work understanding the complex genetic and epigenetic control over ovarian steroid hormone production and within the various ovarian niches will be crucial in the development of the next generation of personalized therapeutics. New complex diagnostics may also arise to allow us to begin to use the timing, mix, and level of ovarian protein and steroid production to evaluate reproductive and general health as well as measuring more complex responses to the environment.

REFERENCES

1. Bird IM, Conley AJ. Steroid biosynthesis: enzymology, integration and control. In: Mason JI, ed. *Genetics of Steroid Biosynthesis and Function*. London: Taylor & Francis; 2002:1–35.
2. Strauss III JF. The synthesis and metabolism of steroid hormones. In: Barbieri R, Strauss III JF, eds. *Yen & Jaffe's Reproductive Endocrinology*. Philadelphia: Elsevier; 2009.
3. Gwynne JT, Strauss III J. The role of lipoproteins in steroidogenesis and cholesterol metabolism in steroidogenic glands. *Endo Rev*. 1982;3:295–329.
4. Chang TY, Chang CCY, Chen D. Acyl-coenzyme A:cholesterol acyltransferase. *Annu Rev Biochem*. 1997;66:613–638.
5. Brasaemle DL. Thematic review series: adipocyte biology. The perilipin family of structural lipid droplet proteins: stabilization of lipid droplets and control of lipolysis. *J Lipid Res*. 2007;48:2547–2559.
6. Shen W-J, Patel S, Natu V, et al. Interaction of hormone-sensitive lipase with steroidogenic acute regulatory protein: facilitation of cholesterol transfer in adrenal. *J Biol Chem*. 2003;278:43870–43876.
7. Clark BJ. The START-domain proteins in intracellular lipid transport and beyond. *Mol Cell Endocrinol*. 2020;504:110704.
8. Lin D, Sugawara T, Strauss III JF, et al. Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science*. 1995;267:1828–1831.
9. Fujieda K, Tajima T, Nakae J, et al. Spontaneous puberty in 46,XX subjects with congenital lipoid adrenal hyperplasia. Ovarian steroidogenesis is spared to some extent despite inactivating mutations in the steroidogenic acute regulatory protein (StAR) gene. *J Clin Invest*. 1997;99:1265–1271.
10. Tu LN, Zhao AH, Hussein M, et al. Translocator protein (TSPO) affects mitochondrial fatty acid oxidation in steroidogenic cells. *Endocrinology*. 2016;157. 1110–1021.
11. Selvaraj V, Stocco DM, Clark BJ. Current knowledge on the acute regulation of steroidogenesis. *Biol Reprod*. 2018;99:13–26.
12. Obsilova V, Obsil T. A new role for 14-3-3 protein in steroidogenesis. *FEBS J*. 2020;287:3921–3924.
13. Prasad M, Kaur J, Pawlak KJ, et al. Mitochondria-associated endoplasmic reticulum membrane (MAM) regulates steroidogenic activity via steroidogenic acute regulatory protein (StAR)-voltage-dependent anion channel 2 (VDAC2) interaction. *J Biol Chem*. 2015;290:2604–2616.
14. Rajapaksha M, Kaur J, Prasad M, et al. An outer mitochondrial translocase, Tom22, is crucial for inner mitochondrial steroidogenic regulation in adrenal and gonadal tissues. *Mol Cell Biol*. 2016;36:1032–1047.
15. Mason JI, Keeney DS, Bird IM, et al. The regulation of 3 beta- hydroxysteroid dehydrogenase expression. *Steroids*. 1997;62:164–168.
16. Flück CE, Tajima T, Pandey AV, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley Bixler syndrome. *Nat Genet*. 2004;36:228–230.
17. Dean B, Chrisp GL, Quartararo M, et al. P450 oxidoreductase deficiency: a systematic review and meta-analysis of genotypes, phenotypes, and their relationships. *J Clin Endocrinol Metab*. 2020;105:e42–e52.
18. Papadakis GE, Dumont A, Bouligand J, et al. Non-classic cytochrome P450 oxidoreductase deficiency strongly linked with menstrual cycle disorders and female infertility as primary manifestations. *Hum Reprod*. 2020;35:939–949.
19. Tee MK, Miller WL. Phosphorylation of human cytochrome P450c17 by p38α selectively increases 17,20 lyase activity and androgen biosynthesis. *J Biol Chem*. 2013;288:23903–23913.
20. Leung MT, Cheung HN, Iu YP, et al. Isolated 17,20-lyase deficiency in a CYB5A mutated female with normal sexual development and fertility. *J Endocr Soc*. 2020;4:bvz016.
21. Turcu AF, Rege J, Auchus RJ, et al. 11-Oxygenated androgens in health and disease. *Nat Rev Endocrinol*. 2020;16:284–296.
22. Imamichi Y, Yuhki KI, Orisaka M, et al. 11-ketotestosterone is a major androgen produced in human gonads. *J Clin Endocrinol Metab*. 2016;101:3582–3591.
23. Hillier SG, Reichert Jr LE, Van Hall EV. Control of preovulatory follicular estrogen biosynthesis in the human ovary. *J Clin Endocrinol Metab*. 1981;52:847–856.
24. Durham CR, Zhu H, Masters BS, et al. Regulation of aromatase activity of rat granulosa cells: induction of synthesis of NADPHcytochrome P-450 reductase by FSH and dibutyryl cyclic AMP. *Mol Cell Endocrinol*. 1985;40:211–219.
25. Simpson ER, Zhao Y, Argawal VR, et al. Aromatase expression in health and disease. *Recent Prog Hormone Res*. 1997;7:185–213.
26. Praveen VP, Ladjouze A, Sauter KS, et al. Novel CYP19A1 mutations extend the genotype-phenotype correlation and reveal the impact on ovarian function. *J Endocr Soc*. 2020;4. bvaa030.
27. Zhang Y, Word RA, Fesmire S, et al. Human ovarian expression of 17β-hydroxysteroid dehydrogenase types 1, 2, and 3. *J Clin Endocrinol Metab*. 1996;81:3594–3598.
28. Haning Jr RV, Tantravahi U, Zhao Q, et al. 5alpha-reductase 1 and 2 expression and activity in human ovarian follicles, stroma and corpus luteum as compared to neonatal foreskin. *J Steroid Biochem Mol Biol*. 1996;59:199–204.
29. Jakimiuk AJ, Weitsman SR, Magoffin DA. 5α-Reductase activity in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 1999;84:2414–2418.
30. Kuiper GG, Carlsson B, Grandien K, et al. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*. 1997;138:863–870.
31. Tu J, Cheung AH, Chan CL, et al. The role of microRNAs in ovarian granulosa cells in health and disease. *Front Endocrinol*. 2019;10:174.
32. Horie T, Nishino T, Baba O, et al. MicroRNA-33 regulates sterol regulatory element-binding protein 1 expression in mice. *Nat Commun*. 2013;4:2883.
33. Kitahara Y, Nakamura K, Kogure K, et al. Role of microRNA136-3p on the expression of luteinizing hormone-human chorionic gonadotropin receptor mRNA in rat ovaries. *Biol Reprod*. 2013;89:114.
34. Rooda I, Hensen K, Kaselt B. Target prediction and validation of microRNAs expressed from FSHR and aromatase genes in human ovarian granulosa cells. *Sci Rep*. 2020;10:1–3.
35. Walton KL, Kelly EK, Chan KL, et al. Inhibin biosynthesis and activity are limited by a prodomain-derived peptide. *Endocrinology*. 2015;156:3047–3057.
36. McGee EA, Raj RS. An extra arm may impair the open hand. *Endocrinology*. 2015;156:2751–2752.
37. Groome NP, Illingworth PJ, O'Brien M, et al. Measurement of dimeric inhibin B throughout the human menstrual cycle. *J Clin Endocrinol Metab*. 1996;81:1401–1405.
38. Nachtigal MW, Ingraham HA. Bioactivation of Müllerian inhibiting substance during gonadal development by a kex2/subtilisin-like endoprotease. *Proc Natl Acad Sci U S A*. 1996;93:7711–7716.
39. di Clemente N, Jamin SP, Lugovskoy A, et al. Processing of anti-mullerian hormone regulates receptor activation by a mechanism distinct from TGF-beta. *Mol Endocrinol*. 2010;24:2193–2206.
40. Pierre A, Racine C, Rey RA, et al. Most cleaved anti-mullerian hormone binds its receptor in human follicular fluid but little is competent in serum. *J Clin Endocrinol Metab*. 2016;101:4618–4627.
41. Stoelk E, Chegini N, Lei ZM, et al. Immunocytochemical localization of relaxin in human corpora lutea: cellular and subcellular distribution and dependence on reproductive state. *Biol Reprod*. 1991;44:1140–1147.
42. Anand-Ivell R, Ivell R. Regulation of the reproductive cycle and early pregnancy by relaxin family peptides. *Mol Cell Endocrinol*. 2014;382:472–479.
43. McGee EA, Hsueh AJ. Initial and cyclic recruitment of ovarian follicles. *Endocr Rev*. 2000;21:200–214.

44. Zelinski-Wooten MB, Hutchison JS, Hess DL, et al. Follicle stimulating hormone alone supports follicle growth and oocyte development in gonadotrophin-releasing hormone antagonist-treated monkeys. *Hum Reprod*. 1995;10:1658–1666.
45. Sullivan MW, Stewart-Akers A, Krasnow JS, et al. Ovarian responses in women to recombinant follicle-stimulating hormone and luteinizing hormone (LH): a role for LH in the final stages of follicular maturation. *J Clin Endocrinol Metab*. 1999;84:228–232.
46. Filicori M, Cognigni GE, Gamberini E, et al. Efficacy of low-dose human chorionic gonadotropin alone to complete controlled ovarian stimulation. *Fertil Steril*. 2005;84:394–401.
47. Devoto L, Kohen P, Gonzalez RR, et al. Expression of steroidogenic acute regulatory protein in the human corpus luteum throughout the luteal phase. *J Clin Endocrinol Metab*. 2001;86:5633–5639.
48. Bogan RL, Hennebold JD. The reverse cholesterol transport system as a potential mediator of luteolysis in the primate corpus luteum. *Reproduction*. 2010;139:163–176.
49. King SR, LaVoie HA. Gonadal transactivation of StARD1, CYP11A1 and HSD3B. *Front Biosci (Landmark Ed)*. 2012;17:824–846.
50. Hinshelwood MM, Repa JJ, Shelton JM, et al. Expression of LRH-1 and SF-1 in the mouse ovary: localization in different cell types correlates with differing function. *Mol Cell Endocrinol*. 2003;207:39–45.
51. Peng N, Kim JW, Rainey WE, et al. The role of the orphan nuclear receptor, liver receptor homologue-1, in the regulation of human corpus luteum 3 β-hydroxysteroid dehydrogenase type II. *J Clin Endocrinol Metab*. 2003;88:6020–6028.
52. Duggavathi R, Volle DH, Mataki C, et al. Liver receptor homolog 1 is essential for ovulation. *Genes Dev*. 2008;22:1871–1876.
53. Tremblay JJ, Viger RS. Novel roles for GATA transcription factors in the regulation of steroidogenesis. *J Steroid Biochem Mol Biol*. 2003;85:291–298.
54. Law NC, Donaubauer EM, Zeleznik AJ, et al. How protein kinase A activates canonical tyrosine kinase signaling pathways to promote granulosa cell differentiation. *Endocrinology*. 2017;158:2043–2051.
55. Tremblay PG, Sirard MA. Gene analysis of major signaling pathways regulated by gonadotropins in human ovarian granulosa tumor cells (KGN) dagger. *Biol Reprod*. 2020;103:583–598.
56. Puri P, Little-Ihrig L, Law NC, et al. Protein kinase A: a master kinase of granulosa cell differentiation. *Sci Rep*. 2016;6:1–2.
57. Andric N, Ascoli M. Mutations of the lutropin/choriogonadotropin receptor that do not activate the phosphoinositide cascade allow hCG to induce aromatase expression in immature rat granulosa cells. *Mol Cell Endocrinol*. 2008;285:62–72.
58. Sato Y, Suzuki T, Hidaka K, et al. Immunolocalization of nuclear transcription factors, DAX-1 and COUP-TF II, in the normal human ovary: correlation with adrenal 4 binding protein/ steroidogenic factor-1 immunolocalization during the menstrual cycle. *J Clin Endocrinol Metab*. 2003;88:3415–3420.
59. Makrigiannakis A, Amin K, Coukos G, et al. Regulated expression and potential roles of p53 and Wilms' tumor suppressor gene (WT1) during follicular development in the human ovary. *J Clin Endocrinol Metab*. 2000;85:449–459.
60. Hu KL, Zhao H, Chang HM, et al. Kisspeptin/kisspeptin receptor system in the ovary. *Front Endocrinol*. 2018;8:365.
61. Giudice LC, Cataldo NA, van Dessel HJ, et al. Growth factors in normal ovarian follicle development. *Semin Reprod Endocrinol*. 1996;14:179–196.
62. Norman RJ, Brannstrom M. Cytokines in the ovary: pathophysiology and potential for pharmacological intervention. *Pharmacol Ther*. 1996;69:219–236.
63. Zhou J, Bondy C. Anatomy of the human ovarian insulin-like growth factor system. *Biol Reprod*. 1993;48:467–482.
64. Wang C, Xu B, Zhou B, et al. Reducing CYP51 inhibits follicle stimulating hormone induced resumption of mouse oocyte meiosis in vitro. *J Lipid Res*. 2009;50:2164–2172.
65. Palter SF, Tavares AB, Hourvitz A, et al. Are estrogens of import to primate/human ovarian folliculogenesis? *Endocr Rev*. 2001;22:389–424.
66. Rabinovici J, Blankstein J, Goldman B, et al. In vitro fertilization and primary embryonic cleavage are possible in 17 alphahydroxylase deficiency despite extremely low intrafollicular 17 beta-estradiol. *J Clin Endocrinol Metab*. 1989;68:693–697.
67. Zelinski-Wooten MB, Hess DL, Baughman WL, et al. Administration of an aromatase inhibitor during the late follicular phase of gonadotropin-treated cycles in rhesus monkeys: effects on follicle development, oocyte maturation, and subsequent luteal function. *J Clin Endocrinol Metab*. 1993;76:988–995.
68. Henriquez S, Kohen P, Xu X, et al. Significance of pro-angiogenic estrogen metabolites in normal follicular development and follicular growth arrest in polycystic ovary syndrome. *Hum Reprod*. 2020;35:1655–1665.
69. Mabjeesh NJ, Escuin D, LaVallee TM, et al. 2ME2 inhibits tumor growth and angiogenesis by disrupting microtubules and dysregulating HIF. *Cancer Cell*. 2003;3:363–375.
70. Kohen P, Henríquez S, Rojas C, et al. 2-Methoxyestradiol in the human corpus luteum throughout the luteal phase and its influence on lutein cell steroidogenesis and angiogenic activity. *Fertil Steril*. 2013;100:1397–1404.
71. Cordts EB, Santos MC, Peluso C, et al. COMT polymorphism influences decrease of ovarian follicles and emerges as a predictive factor for premature ovarian insufficiency. *J Ovarian Res*. 2014;7:1–4.
72. Hillier SG, Tetsuka M. Role of androgens in follicle maturation and atresia. *Baillieres Clin Obstet Gynaecol*. 1997;11:249–260.
73. Zeleznik AJ, Little-Ihrig L, Ramasawamy S. Administration of dihydrotestosterone to rhesus monkeys inhibits gonadotropin-stimulated ovarian steroidogenesis. *J Clin Endocrinol Metab*. 2004;89:860–866.
74. Lebbe M, Taylor AE, Visser JA, et al. The steroid metabolome in the isolated ovarian follicle and its response to androgen exposure and antagonism. *Endocrinology*. 2017;158:1474–1485.
75. Agarwal SK, Judd HL, Magoffin DA. A mechanism for the suppression of estrogen production in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 1996;81:3686–3691.
76. Lossl K, Frieselben NLC, Wissing ML, et al. Biological and clinical rationale for androgen priming in ovarian stimulation. *Front Endocrinol*. 2020;11:627.
77. Yuan XH, Yang CR, Wang XN, et al. Progesterone maintains the status of granulosa cells and slows follicle development partly through PGRMC1. *J Cell Physiol*. 2018;234:709–720.
78. Iwai T, Nanbu Y, Iwai M, et al. Immunohistochemical localization of oestrogen receptors and progesterone receptors in the human ovary throughout the menstrual cycle. *Virchows Arch A Pathol Anat Histopathol*. 1990;417:369–375.
79. Misao R, Nakanishi Y, Iwagaki S, et al. Expression of progesterone receptor isoforms in corpora lutea of human subjects: correlation with serum oestrogen and progesterone concentrations. *Mol Hum Reprod*. 1998;4:1045–1052.
80. Stouffer RL. Progesterone as a mediator of gonadotrophin action in the corpus luteum: beyond steroidogenesis. *Hum Reprod Update*. 2003;9:99–117.
81. Borman SM, Chaffin CL, Schwinof KM, et al. Progesterone promotes oocyte maturation, but not ovulation, in nonhuman primate follicles without a gonadotropin surge. *Biol Reprod*. 2004;71:366–373.
82. Peng C, Ohno T, Khorasheh S, et al. Activin and follistatin as local regulators in the human ovary. *Biol Signals*. 1996;5:81–89.
83. McGee EA, Raj RS. Regulators of ovarian preantral follicle development. *Semin Reprod Med*. 2015;33:179–184.
84. Schwall RH, Mason AJ, Wilcox JN, et al. Localization of inhibin/ activin subunit mRNAs within the primate ovary. *Mol Endocrinol*. 1990;4:75–79.
85. Teede H, Ng S, Hedger M, et al. Follistatin and activins in polycystic ovary syndrome: relationship to metabolic and hormonal markers. *Metabolism*. 2013;62:1394–1400.
86. Guo Q, Kumar TR, Woodruff T, et al. Overexpression of mouse follistatin causes reproductive defects in transgenic mice. *Mol Endocrinol*. 1998;12:96–106.
87. Ingraham HA, Hirokawa Y, Roberts LM, et al. Autocrine and paracrine Müllerian inhibiting substance hormone signaling in reproduction. *Recent Prog Horm Res*. 2000;55:53–67.
88. Gruijters MJ, Visser JA, Durlinger AL, et al. Anti-Müllerian hormone and its role in ovarian function. *Mol Cell Endocrinol*. 2003;211:85–90.
89. Grossman MP, Nakajima ST, Fallat ME, et al. Müllerian-inhibiting substance inhibits cytochrome P450 aromatase activity in human granulosa lutein cell culture. *Fertil Steril*. 2008;89(5 suppl):1364–1370.

90. Robertson DM, Lee CH, Baerwald A. Interactions between serum FSH, inhibin B and antral follicle count in the decline of serum AMH during the menstrual cycle in the late reproductive age. *Endocrinol Diab Metab*. 2021;4:e00172.
91. Dilaver N, Pellatt L, Jameson E, et al. The regulation and signalling of anti-mullerian hormone in human granulosa cells: relevance to polycystic ovary syndrome. *Hum Reprod*. 2019;34:2467–2479.
92. Kallio PJ, Ruokonen A, et al. Antimüllerian hormone levels decrease in women using combined contraception independently of administration route. *Fert Steril*. 2013;99:1305–1310.
93. Kallio S, Aittomäki K, Piltonen T, et al. Anti-Mullerian hormone as a predictor of follicular reserve in ovarian insufficiency: special emphasis on FSH-resistant ovaries. *Hum Reprod*. 2012;27:854–860.
94. Seifer DB, Maclaughlin DT. Müllerian Inhibiting Substance is an ovarian growth factor of emerging clinical significance. *Fertil Steril*. 2007;88:539–546.
95. Sowers M, McConnell D, Gast K, et al. Anti-Müllerian hormone and inhibin B variability during normal menstrual cycles. *Fertil Steril*. 2010;94:1482–1486.
96. Robertson DM, Hale GE, Fraser IS, et al. Changes in serum antimüllerian hormone levels across the ovulatory menstrual cycle in late reproductive age. *Menopause*. 2011;18:521–524.
97. Knight PG, Glister C. TGF-beta superfamily members and ovarian follicle development. *Reproduction*. 2006;132:191–206.
98. Vitt UA, McGee EA, Hayashi M, et al. In vivo treatment with GDF-9 stimulates primary follicle progression and theca cell marker CYP17 in ovaries of immature rats. *Endocrinology*. 2000;141:3814–3820.
99. Elvin JA, Yan C, Wang P, et al. Molecular characterization of the follicle defects in the growth differentiation factor 9-deficient ovary. *Mol Endocrinol*. 1999;13:1018–1034.
100. Yamamoto N, Christenson LK, McAllister JM, et al. Growth differentiation factor-9 inhibits 3′5′-adenosine monophosphate-stimulated steroidogenesis in human granulosa and theca cells. *J Clin Endocrinol Metab*. 2002;87:2849–2856.
101. Patiño LC, Walton KL, Mueller TD, et al. BMP15 mutations associated with primary ovarian insufficiency reduce expression, activity, or synergy with GDF9. *J Clin Endocrinol Metab*. 2017;102:1009–1019.
102. Renault L, Patiño LC, Magnin F, et al. BMPR1A and BMPR1B missense mutations cause primary ovarian insufficiency. *J Clin Endocrinol Metab*. 2020;105:e1449–e1457.
103. Richards JS. Genetics of ovulation. *Semin Reprod Med*. 2007;25:235–242.
104. Juengel JL, Meberg BM, Turzillo AM, et al. Hormonal regulation of messenger ribonucleic acid encoding steroidogenic acute regulatory protein in ovine corpora lutea. *Endocrinology*. 1995;136:5423–5429.
105. Baumgarten SC, Armouti M, Ko C, et al. IGFR1 expression in ovarian granulosa cells is essential for steroidogenesis, follicle survival, and fertility in female mice. *Endocrinology*. 2017;158:2309–2318.
106. McGee EA, Sawetawan C, Bird I, et al. The effect of insulin and insulin-like growth factors on the expression of steroidogenic enzymes in a human ovarian thecal-like tumor cell model. *Fertil Steril*. 1996;65:87–93.
107. Conover CA, Faessen GF, Ilg KE, et al. Pregnancy-associated plasma protein-a is the insulin-like growth factor binding protein-4 protease secreted by human ovarian granulosa cells and is a marker of dominant follicle selection and the corpus luteum. *Endocrinology*. 2001;142:2155.
108. Terranova PF, Rice VM. Review: cytokine involvement in ovarian processes. *Am J Reprod Immunol*. 1997;37:50–63.
109. Kang SK, Choi KC, Cheng KW, et al. Role of gonadotropin-releasing hormone as an autocrine growth factor in human ovarian surface epithelium. *Endocrinology*. 2000;141:72–80.
110. Metallinou C, Asimakopoulos B, Schröer A, et al. Gonadotropin releasing hormone in the ovary. *Reprod Sci*. 2007;14:737–749.
111. Richards JS, Liu Z, Kawai T, et al. Adiponectin and its receptors modulate granulosa cell and cumulus cell functions, fertility, and early embryo development in the mouse and human. *Fertil Steril*. 2012;98:471–479.
112. Reverchon M, Cornuau M, Ramé C, et al. Resistin decreases insulin-like growth factor I-induced steroid production and insulin-like growth factor I receptor signaling in human granulosa cells. *Fertil Steril*. 2013;100:247–255.
113. Reverchon M, Cornuau M, Ramé C, et al. Chemerin inhibits IGF-1-induced progesterone and estradiol secretion in human granulosa cells. *Hum Reprod*. 2012;27:1790–1800.
114. Wickham 3rd EP, Tao T, Nestler JE, et al. Activation of the LH receptor up regulates the type 2 adiponectin receptor in human granulosa cells. *J Assist Reprod Genet*. 2013;30:963–968.
115. Várnagy A, Bódis J, Kovács GL, et al. Metabolic hormones in follicular fluid in women undergoing in vitro fertilization. *Reprod Med*. 2013;58:305–311.
116. Tapanainen J, Voutilainen R, Jaffe RB. Low aromatase activity and gene expression in human fetal testes. *J Steroid Biochem*. 1989;33:7–11.
117. Poulsen LC, Englund ALM, Andersen AS, et al. Follicular hormone dynamics during the midcycle surge of gonadotropins in women undergoing fertility treatment. *Mol Hum Reprod*. 2020;26:256–268.
118. Sasano H, Suzuki T. Localization of steroidogenesis and steroid receptors in human corpus luteum. Classification of human corpus luteum (CL) into estrogen-producing degenerating CL, and nonsteroid-producing degenerating CL. *Semin Reprod Endocrinol*. 1997;15:345–351.
119. Hutchison JS, Zeleznik AJ. The corpus luteum of the primate menstrual cycle is capable of recovering from a transient withdrawal of pituitary gonadotropin support. *Endocrinology*. 1985;117:1043–1049.
120. Zeleznik AJ. In vivo responses of the primate corpus luteum to luteinizing hormone and chorionic gonadotropin. *Proc Natl Acad Sci U S A*. 1998;95:11002–11007.
121. De Leener A, Montanelli L, Van Durme J, et al. Presence and absence of follicle-stimulating hormone receptor mutations provide some insights into spontaneous ovarian hyperstimulation syndrome pathophysiology. *J Clin Endocrinol Metab*. 2006;91:555–562.
122. Del Canto F, Sierralta W, Kohen P, et al. Features of natural and gonadotropin-releasing hormone antagonist-induced corpus luteum regression and effects of in vivo human chorionic gonadotropin. *J Clin Endocrinol Metab*. 2007;92:4436–4443.
123. Dennefors BL, Janson PO, Knutson F, et al. Steroid production and responsiveness to gonadotropin in isolated stromal tissue of human postmenopausal ovaries. *Am J Obstet Gynecol*. 1980;136:997–1002.
124. Dennefors BL, Janson PO, Hamberger L, et al. Hilus cells from human postmenopausal ovaries: gonadotrophin sensitivity, steroid and cyclic AMP production. *Acta Obstet Gynecol Scand*. 1982;61:413–416.
125. Cauley JA, Gutai JP, Kuller LH, et al. The epidemiology of serum sex hormones in postmenopausal women. *Am J Epidemiol*. 1989;129:1120–1131.
126. Couzinet B, Meduri G, Lecce MG, et al. The postmenopausal ovary is not a major androgen-producing gland. *J Clin Endocrinol Metab*. 2001;86:5060–5066.
127. Nagamani M, Urban RJ. Expression of messenger ribonucleic acid encoding steroidogenic enzymes in post-menopausal ovaries. *J Soc Gynecol Investig*. 2003;10:37–40.
128. Grodin JM, Siiteri PK, MacDonald PC. Source of estrogen production in postmenopausal women. *J Clin Endocrinol Metab*. 1973;36:207–214.
129. Judd HL, Judd GE, Lucas WE, et al. Endocrine function of the postmenopausal ovary: concentration of androgens and estrogens in ovarian and peripheral vein blood. *J Clin Endocrinol Metab*. 1974;39:1020–1024.
130. Vermeulen A. The hormonal activity of the postmenopausal ovary. *J Clin Endocrinol Metab*. 1976;42:247–253.
131. Poliak A, Jones G, Goldberg B. Effect of human chorionic gonadotropin on postmenopausal women. *Am J Obstet Gynecol*. 1968;101:731.
132. Greenblatt RB, Colle ML, Mahesh VB. Ovarian and adrenal steroid production in the postmenopausal woman. *Obstet Gynecol*. 1976;47:383–387.
133. Nakano R, Shima K, Yamoto M, et al. Binding sites for gonadotropins in human postmenopausal ovaries. *Obstet Gynecol*. 1989;73:196–200.
134. Sternberg WH. The morphology, androgenic function, hyperplasia, and tumors of the human ovarian hilus cells. *Am J Pathol*. 1947;25:493.

120

Estrogen and Progesterone Action

Francesco J. DeMayo and Carole R. Mendelson

OUTLINE

Interaction of Steroid Hormones With Target Cells, 2003
Nuclear Receptor Structure, 2004
Coactivators and Corepressors, 2005
  Coactivators, 2005
  Corepressors, 2007
Mechanisms of Estrogen Action, 2007
  Estrogen Receptor Ligands, 2007
  Estrogen Receptor Structure, 2008
  Estrogen Receptor-Coactivator Interactions, 2008
  Molecular Mechanisms of Estrogen Receptor Action, 2009
  Use of Gene-Targeted Mice To Ascertain the Biology of Estrogen Receptor Function, 2011
Mechanisms of Progesterone Action, 2013
  Progesterone Receptor Isoforms, 2013
  Use of Gene-targeted Mice To Evaluate the Biology of PR-A and PR-B Function, 2013
  Molecular Mechanisms of Progesterone Action in the Endometrium and on Blastocyst Implantation, 2014
  Molecular Mechanisms of Progesterone Action in the Pregnant Myometrium, 2015
  Mechanisms for the Loss of PR Function in the Pregnant Myometrium Near Term Leading to Labor, 2015
Conclusions, 2016

The steroid hormones estradiol-17β and progesterone serve key roles in female reproductive function and behavior by binding to their cognate receptors in the uterus, ovaries, mammary glands, and brain. In non-pregnant cycling females, estrogens and progesterone produced by the ovaries serve critical roles in endometrial development and its preparation for blastocyst attachment and implantation and mammary gland development and differentiation. Progesterone is the predominant steroid hormone of pregnancy and is required for implantation of the blastocyst and for the maintenance of myometrial (uterine smooth muscle) quiescence. In rodents, the corpus luteum continues to contribute to circulating levels of steroid hormones throughout pregnancy. However, in humans, certain nonhuman primates, and ruminants, the placenta gains the capacity to serve as the main source of estrogens and progesterone sometime during the first trimester and throughout pregnancy.

Estrogen and progesterone actions are primarily mediated by their nuclear receptors, which are members of the ***nuclear receptor (NR) family***, comprising 48 ligand-activated transcription factors in humans that are present in the chromatin of target cells in their activated states.[1,2] Most NR family members are bound by small lipophilic ligands, including steroid hormones, thyroid hormones, the seco-steroid 1,25-dihydroxyvitamin D_3, retinoids, sterols, phospholipids, heme, and fatty acids; however, this receptor family also contains "orphan" members for which no ligand has yet been identified. Members of the NR family can be placed into subgroups based on similarities in structure and the ligands that they bind. The estrogen (ER) and progesterone receptors (PR) belong to the ligand-activated steroid hormone receptor subgroup. The ERs comprise two closely related receptors for estradiol-17β, ERα and ERβ, encoded by the *ESR1* (*NR3A1*) and *ESR2* (*NR3A2*) genes, respectively. The PR consists of two major isoforms, PR-A and PR-B, and two minor splice forms, PR-C and PR-M, resulting from differential transcription initiation from a single *PGR* gene (*NR3C3*). This family also consists of receptors for the glucocorticoids, cortisol and corticosterone (GR, encoded by the *NR3C1* gene); the mineralocorticoids, aldosterone and deoxycorticosterone (DOC) (MR, encoded by the *NR3C2* gene); and the androgens, testosterone and 5α-dihydrotestosterone (AR, encoded by the *NR3C4* gene).

INTERACTION OF STEROID HORMONES WITH TARGET CELLS

Steroids travel in the blood predominantly bound to several classes of serum proteins. Within the circulation, estrogens are bound to sex hormone-binding globulin, which binds estradiol-17β, as well as testosterone, with relatively high affinity ($K_d \sim 10^{-9}$ M). Circulating estrogens also are weakly bound to serum albumin. Progesterone is bound in the circulation to corticosteroid-binding globulin, which also binds circulating glucocorticoids. The majority of circulating estrogen and progesterone is bound to these globulins and is in equilibrium with a relatively small amount of free steroid, which can enter cells. Because of their lipophilic nature, estrogens and progesterone can freely diffuse across plasma and nuclear membranes of all cells but are sequestered only within target cells that contain their specific receptors (Fig. 120.1). In the absence of hormone binding, ER and PR exist primarily in the nucleus as monomers that are associated with a complex of proteins, including two molecules of a 90,000-Dalton phosphoprotein, heat shock protein (hsp) 90, one molecule of hsp70, immunophilins, and a number of other proteins.[3] The hsp complex prevents the monomeric receptors from binding to DNA and folds them in such a way that they have high affinity and specificity for their corresponding hormones. The binding of hormone to its receptor causes a conformational change in the receptor that results in its dissociation from the hsp complex.[3] The hormone-bound receptor monomer then interacts with another hormone-bound

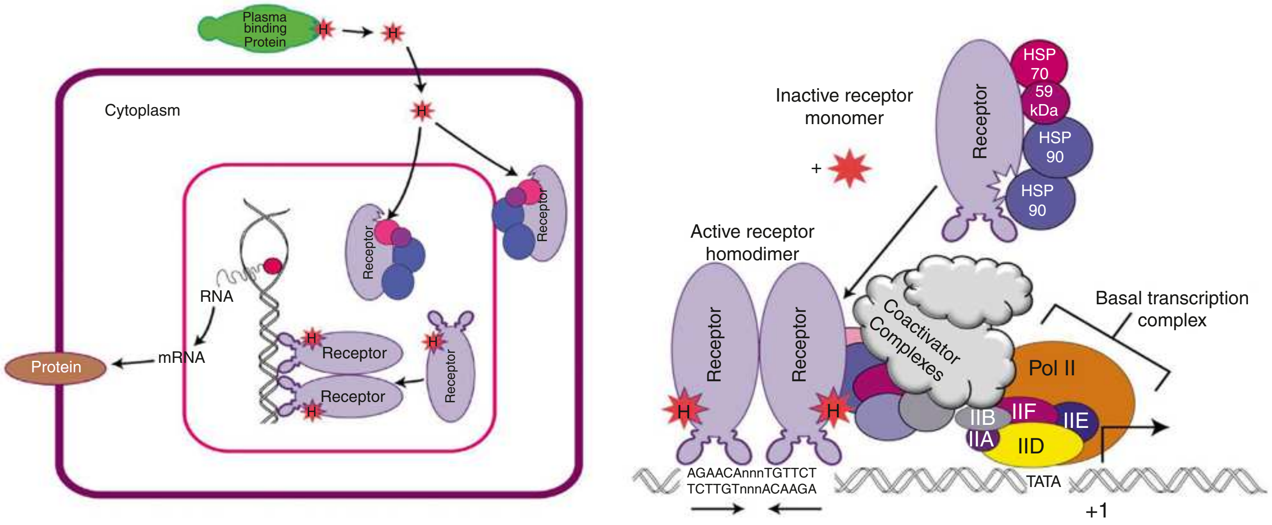


Figure 120.1 Interaction of steroid hormones with their target cells, receptor activation, DNA binding, and recruitment of coregulatory complexes.

receptor monomer to form a homodimer, which has a greatly increased affinity for hormone response elements (HREs) in DNA (see Fig. 120.1). Phosphorylation of the receptor homodimer, either just before or after binding to DNA, likely contributes to the transformation of the receptor dimer to a transcription modulating state.

KEY POINTS

- The estrogen and progesterone receptors are ligand-activated transcription factors that belong to the NR family.

The consensus HREs for ER and PR are inverted repeats/palindromes that can lie near the promoter, or hundreds to thousands to hundreds of thousands of bp upstream. HREs also can lie within the transcribed region, or even downstream of the target gene. The estrogen response element, which specifically binds ERα and ERβ, has the consensus sequence AGGTACnnnTGACCT, where "n" is any nucleotide, whereas the progesterone response element (PRE) has the consensus sequence AGGACAnnnTGTTCT. Notably, the naturally occurring response elements may differ considerably from the consensus sequence. Upon binding to their corresponding response elements in DNA, the ER and PR homodimers recruit a complex of coregulators (see Fig. 120.1). Coregulators do not bind to DNA directly, but rather serve as scaffolds that recruit large protein complexes containing histone modifying enzymes, such as histone acetyltransferases (HATs) or histone methyltransferases. Such histone modifications facilitate the opening of chromatin, making it accessible to additional regulatory proteins and enable recruitment of general transcription factors to form a stable preinitiation complex at the gene promoter. Coactivators and interacting proteins also bridge transcription factors with the basal transcription complex, resulting in further stabilization, recruitment of RNA polymerase II (RNA Pol II) and initiation of gene transcription.

NUCLEAR RECEPTOR STRUCTURE

Through sequencing and x-ray crystallographic studies, great insight has been gained into the structure of members of the NR family and the structural basis for ligand interaction, dimerization, DNA binding, and coactivator interaction.[4] All members of the steroid receptor family have the same modular multidomain structure (Fig. 120.2). The portion of these receptors from the N-terminus to the DNA-binding domain (DBD, **A/B domain**) is the most variable in length and sequence. This region is intrinsically disordered and has not been amenable to x-ray crystallographic analysis. The A/B domain contains an activation function domain (AF-1) that binds coactivators in a ligand-independent manner. It also contains sequences that are posttranslationally modified by phosphorylation. Just downstream of the A/B region is the highly ordered **DBD (Domain C)**, which is the most conserved domain among all members of the NR family, and especially among the steroid receptors. The DBD is composed of ~70 amino acids and contains nine perfectly conserved cysteine residues. This domain contains two repeated units enriched in the amino acids Cys, Lys, and Arg. Each of these units is folded into a finger-like structure containing four cysteines that coordinate a zinc ion (see Fig. 120.2). Amino acids at the base of the more N-terminal Zn-finger within the DBD comprise the "proximal" or P-box, which is critical for DNA sequence recognition by making base-specific contacts with the major groove of DNA. The amino acids at the base on the more C-terminal Zn-finger comprise the "distal" or D-box, which is important for receptor dimerization and recognition of spacing of the two half-sites of the DNA response element. All NRs also contain a flexible "hinge region" (**Domain D**), which lies just C-terminal to the DBD. The hinge region contains a nuclear localization sequence, as well as sites for posttranslational modification. Near the carboxy-terminal end of the receptor molecules is the conserved **ligand-binding domain** (**LBD, Domain E**), composed of largely hydrophobic amino acids. This domain also contains a ligand binding-activated transcription activation domain, termed ***activation function*** **2** (AF-2). The LBD contains 12 α-helices and 4 β-strands that fold to form a ligand-binding pocket that varies

KEY POINTS

- Estrogen and progesterone travel in the circulation largely bound to serum proteins. The small amounts of circulating free steroids can enter cells and specifically bind to their cognate receptors within the nuclei of target cells.
- Upon hormone binding, ER and PR dissociate from a complex of hsps and bind to their HREs in DNA as homodimers.
- Once bound to DNA, ER and PR homodimers recruit a complex of coregulators that contain histone-modifying activities, which relax the chromatin structure to allow assembly of a basal transcription complex and of RNA Pol II to catalyze transcription initiation (see Fig. 120.1).

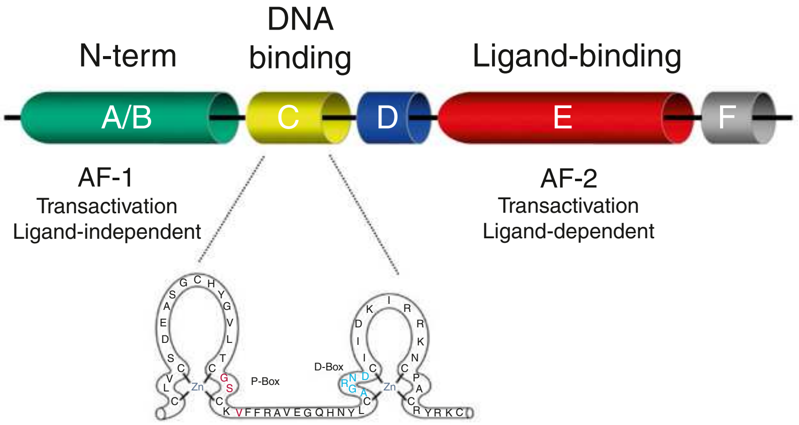


Figure 120.2 Functional domains of steroid hormone receptors, with focus on the transcription activation domains, the highly conserved DNA-binding domain and the amino acids that contribute to receptor recognition of the hormone response elements and to the two Zn-finger structures.

greatly in size among NR family members to accommodate their specific ligands. Upon ligand binding, the position of Helix-12 (H12) within the AF-2 domain is dynamically altered to provide a binding surface for coactivator molecules via their LXXLL motifs, contributing to transcriptional activation. The LBD also contains sequences for interaction with hsp90. C-terminal to the LBD is the **F domain**, which is present only in ERα and ERβ (see Fig. 120.2). The F domain of ERα appears to serve a role in transcriptional activation, coactivator interactions, dimerization, and receptor stability.

KEY POINTS

- All members of the steroid receptor family have the same modular multi-domain structure (see Fig. 120.2).
- The most highly conserved domain is the DBD (Domain C), which contains two repeated units enriched in the amino acids Cys, Lys, and Arg. Each of these units is folded into a finger-like structure containing four cysteines that coordinate a zinc ion.
- In general, there are two transcription activation domains (AF-1 and AF-2). AF-1, near the N-terminus, participates in transcriptional activation in the absence of hormone binding. AF-2, within the LBD, activates transcriptional activity in a ligand-dependent manner. PR-B contains an addition AF domain (AF-3) near its extended amino terminus.

COACTIVATORS AND COREPRESSORS

Receptors for estrogen and progesterone and other members of the NR family undergo conformational changes upon binding their respective ligands but do not have the capacity to activate or inhibit transcription in and of themselves. Rather, upon hormone binding and dimerization, the receptors bind to their HREs in DNA and recruit scaffolding proteins, termed coactivators and corepressors, with histone-modifying activities that alter the transcriptional activities of specific genes. Since the original characterization of the first steroid receptor coactivator, SRC-1, 25 years ago,[5] >400 coregulators have been isolated and characterized.[6] Based on function, coregulators can be grouped into two major categories: ***coactivators***, which associate with agonist-bound NRs to upregulate gene expression, and ***corepressors***, which selectively repress gene expression through interaction either with unliganded or antagonist-bound NRs.

Coactivators

Within the eukaryotic nucleus, DNA is tightly packaged in chromatin as a repeating array of nucleosomes and as higher-order structures, comprising condensed chromatin.[7,8] Nucleosomes consist of ~146 bp of DNA wound around a histone octamer, containing two copies of the histone proteins H2A, H2B, H3, and H4. The favorable electrostatic interactions between the negatively charged phosphodiester backbone of DNA and positively charged lysines and arginines in the histone tails contribute to chromatin condensation, which is generally repressive to transcription.[9] While steroid hormone receptors can bind to their response elements within nucleosomal DNA,[7] the surrounding genomic regions are unavailable for binding and assembly of the transcriptional machinery required for transcription activation. This requires the recruitment of coactivators with histone-modifying activities and of ATP-dependent remodeling complexes.

A cDNA encoding the first NR coactivator to be identified and characterized, steroid receptor coactivator 1 (SRC-1), was isolated from a yeast two-hybrid system using the LBD and hinge region of the human PR as bait.[5] Transient transfection assays in mammalian cells indicated that SRC-1 activation of PR transcriptional activity was agonist-dependent and required agonist-bound PR binding to its cognate DNA response element. Moreover, SRC-1 was found to mediate transcriptional activation by other members of the steroid receptor family bound to their respective ligands and response elements. Furthermore, ligand-dependent transcriptional activation of ERα was observed to be dependent upon the interaction of key hydrophobic residues within H12 of ERα with short sequences within SRC-1 containing the amino acids LXXLL (where L = leucine and X = any amino acid).[10] It was proposed that LXXLL motifs were required for interaction of a number of different coactivator proteins with NRs.[10]

SRC-1 is one of three members of the p160 coactivator family, which also includes SRC-2/TIF-2 and SRC-3/AIB1. Although the p160 coactivators interact with a variety of other transcription factors, their actions are strongly tied to steroid receptor activation and are closely linked to reproductive function.[11] Studies using knockout mice indicate that although structurally similar, each SRC family member has specific physiological functions.[12] For example, although *SRC-1*$^{-/-}$ mice were found to be fertile, females manifested reduced estrogen-stimulated uterine growth, and a decreased progesterone-dependent uterine decidual response.[13] Female *SRC-2*$^{-/-}$ mice displayed placental hypoplasia and hypofertility,[14] as well as an impaired decidual response, further exacerbated by the combined deficiency of SRC-1.[15] *SRC-2*$^{-/-}$ male mice manifested infertility due to defects in spermatogenesis.[16] Notably, mice that were double KO for *Src-1* and *Src-2* died at birth of respiratory distress,[17] suggesting a deficiency of pulmonary surfactant production. Female *Src-1*$^{+/-}$/*Src-2*$^{+/-}$ crossed with genetically like males manifested remarkably delayed parturition,

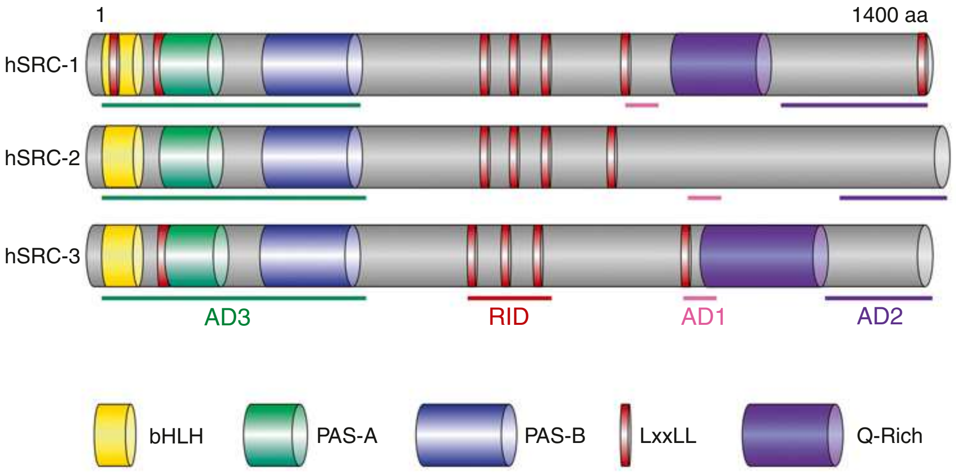


Figure 120.3 Steroid receptor coactivator (SRC) family members, SRC-1, SRC-2, and SRC-3, serve as modular scaffolds for binding of transcription factors, such as steroid receptors, and other coregulators with histone-modifying activities. For abbreviations, see text.

associated with decreased myometrial nuclear factor-κB (NF-κB) activation and contractile gene expression (i.e., connexin 43 [CX43] and oxytocin receptor), impaired luteolysis, and a failure of circulating progesterone levels to decline at term.[18] Interestingly, WT females bred to *Src-1/-2* double-deficient males exhibited similarly delayed parturition, molecular and hormonal changes, indicating that the defect causing delayed labor was fetal in origin.[18] In this regard, lungs of *Src-1/-2* double-deficient fetuses exhibited markedly decreased expression of surfactant protein-A (SP-A) and lysophosphatidylcholine acyltransferase-1, a key enzyme for synthesis of surfactant dipalmitoylphosphatidyl choline (a highly surface-active glycerophospholipid) and platelet-activating factor (PAF), a proinflammatory glycerophospholipid.[18] The effects of *Src-1/-2* double-deficiency in the fetal lung were mediated by impaired glucocorticoid and cytokine signaling through glucocorticoid receptor (GR) and CCAAT enhancer-binding protein transcription factors.[18,19] These findings reveal that fetal lungs when mature produce SP-A and PAF signals to initiate labor; moreover, SRC-1 and SRC-2 coactivators serve crucial roles in the regulation of these fetal parturition signals. SRC-3, which is amplified in breast and ovarian cancer,[20] serves as a key modulator in the upregulation of ERα transcriptional activity.[21] SRC-3 levels also are elevated in endometrial hyperplasia and adenocarcinoma; high expression levels were correlated with poor prognosis.[22]

As mentioned, SRC family members serve as modular scaffolds for binding of transcription factors, such as steroid receptors, which bind to three LXXLL motifs in the receptor interaction domain (RID) within the central regions of the SRCs (Fig. 120.3). SRCs also interact with other transcription factors, including NF-κB, which binds to the C-terminal activation domain 1 (AD1), and with members of the signal transducer and activator of transcription (STAT) family and tumor protein P53 (TP53), which bind to the N-terminal basic helix-loop-helix-Per-Arnt-Sim (bHLH-PAS) domain (AD3) (see Fig. 120.3). Although bHLH-PAS domains present in a variety of transcription factors are known to bind to DNA, SRC family members have no apparent DNA-binding ability. Rather, SRCs also serve as scaffolds for recruitment and binding to other coregulators. For example, the AD1 domain binds to other SRCs and to the HATs, p300 and CREB-binding protein (CBP), which catalyze acetylation of lysine residues in the N-terminal tails of nucleosomal histones. As mentioned, electrostatic interactions between the negatively charged phosphodiester backbone of DNA and positively charged lysines and arginines in the histone tails contributes to chromatin condensation. Coactivator-mediated acetylation and or methylation of these lysine and arginine residues, respectively, results in a neutralization of these interactions and a relaxation and "opening" of chromatin structure.

SRC-1 and SRC-3 also have intrinsic HAT activity, whereas SRC-2 does not.[11,23] The region of SRC-1 that contains HAT activity also binds the HAT, p300/CBP-associated factor (PCAF). The subsequent acetylation of key lysine residues in positively charged histones and relaxation of chromatin structure facilitates the binding of other transcription factors and coregulatory complexes, and assembly of the RNA Pol II transcription machinery to promote transcriptional activation of specific genes, as described above. Notably, p300/CBP also catalyze acetylation of lysine residues on non-histone proteins, which include a variety of transcription factors. For example, acetylation of lysine residues within the ERα hinge region,[24] as well as NF-κB, STAT3, and TP53,[25] increase their DNA-binding and/or transcriptional activity.[25] The AD1 domains of SRCs also bind the coactivator-associated arginine methyltransferase-1 (CARM1/PRMT4) and PRMT1,[26] which methylate histones on arginine residues to promote recruitment of p300/CBP and increase target gene transcription.

KEY POINTS

- Coregulators comprise ***coactivators***, which associate with agonist-bound NRs to upregulate gene expression, and ***corepressors***, which selectively repress gene expression through interaction either with unliganded or antagonist-bound NRs.
- Coactivators and corepressors do not bind to DNA directly; rather, they are recruited to ligand-bound dimeric ER or PR and other transcription factors bound to DNA. Once bound, coregulators can posttranslationally modify histones to relax or condense chromatin structure, act as modular scaffolds (see Fig. 120.3) to recruit other coregulators containing histone-modifying activities, and bridge to members of the basal transcription complex at gene promoters.
- Activated NRs bound to their HREs also can recruit ATP-dependent DNA remodeling complexes that directly break histone–DNA contacts via energy generated by ATP hydrolysis to allow repositioning of nucleosomes.
- The p160 family of coactivators (SRC-1, SRC-2, and SRC-3) are strongly associated with steroid receptor activation and closely linked to reproductive and developmental biology (including fertility, fetal lung development, and the initiation of parturition) and cancer.

While the histone modifications catalyzed by various coactivators recruited by steroid receptors contribute to activation of gene expression, they do not substantially affect the core particle structure of nucleosomes. However, the modification of nucleosomal histones often precedes the binding of large, multisubunit ATP-dependent remodeling complexes (e.g., SWI/SNF) that directly break histone–DNA contacts via energy generated by ATP hydrolysis. This, in turn, allows the sliding and repositioning of nucleosomes.[27] BRG1 comprises the central ATPase of SWI/SNF and other ATP-dependent remodeling complexes. BRG1 and associated factors have the capacity to directly bind to steroid hormone receptors. In studies of glucocorticoid activation of the mouse mammary tumor virus promoter, it was observed that after glucocorticoid binding to the GR and recruitment of histone-modifying coactivators, SWI/SNF is recruited to the promoter through an interaction of BRG1 with GR. This results in the ATP-dependent repositioning of nucleosomes allowing recruitment of Mediator proteins and assembly of the preinitiation complex with RNA Pol II binding and transcription initiation.[27]

Corepressors

Unlike the steroid hormone receptors, NRs for retinoic acid and thyroid hormone are bound to HREs within the regulatory regions of their target genes in the absence of hormone binding. As such, the unliganded receptors recruit the corepressors, nuclear receptor corepressor (NCoR) and silencing mediator of retinoic acid and thyroid hormone receptors (SMRT) to repress target gene expression. NCoR and SMRT, in turn, serve as docking platforms for enzymes that maintain chromatin in a "closed" repressed state.[28,29] In the case of NCoR and SMRT, a motif called the corepressor nuclear receptor (CoRNR) box, interacts with a hydrophobic groove that is exposed within the unliganded receptor. The predominant histone-modifying enzyme associated with NCoR and SMRT is histone deacetylase 3 (HDAC3), which catalyzes histone deacetylation. This maintains electrostatic interactions between the positively charged core histones and negatively charged nucleosomal DNA, resulting in a repressed higher-order chromatin structure. Upon hormone binding, a conformational change in the receptor causes the corepressor complex to dissociate and be replaced by a complex of coactivators, which interacts with the receptor through an LXXLL motif, and catalyzes histone acetylation or other posttranslational modifications that result in an opening of chromatin structure, as described above.

With regard to ER and PR, NCoR and SMRT are recruited to ERα and PR *in vivo* when they are bound to antagonists or tissue-selective estrogen receptor modulators (SERMs). Tamoxifen, a SERM, antagonizes estrogen action in the mammary gland, but serves as an agonist in the uterus. Tamoxifen treatment of MCF-7 human breast cancer cells was found to cause rapid recruitment of endogenous ERα to estrogen-responsive regions of the estrogen target genes *pS2* and *c-Myc*.[30] This occurred with recruitment of NCoR complexes containing HDAC3 and was temporally associated with deacetylation of the N-terminal tails of histones H3 and H4 and dissociation of RNA Pol II from the *pS2* and *c-Myc* promoters, contributing to the inhibition of gene expression.[30] Mifepristone, also known as RU486, has partial agonist/antagonist activity for PR and GR. Like progesterone itself, RU486 binds to PR with high affinity, promotes PR dissociation from the complex of hsps, dimerization, and binding to its HREs in DNA. However, modeling of the PR LBD crystal structure bound to RU486 suggested that H12 is displaced in a conformation that blocks coactivator binding to the AF2 domain.[31,32] While NCoR does not interact with agonist-bound PR, it has high affinity for PR bound to RU486, both *in vitro* and *in vivo*. Thus, both NCoR and SMRT bind preferentially to antagonist-bound, compared with agonist-bound, ER and PR. Moreover, overexpression of these corepressors antagonizes the partial agonist activities of SERMs like tamoxifen and of RU486 bound to ERα and PR, respectively. It has been suggested that cell-specific concentrations of corepressors may help to determine whether tamoxifen or RU486 acts as an agonist or an antagonist of receptor-mediated gene transcription.[33]

Glucocorticoids exert anti-inflammatory actions, in part, via tethering to the Jun subunit of AP-1 or to NF-κB transcription factors bound to the promoters of genes encoding proinflammatory cytokines. In lieu of recruitment of the NCoR and SMRT corepressor complexes, as described above, glucocorticoid receptor (GR)-mediated suppression of AP-1 or NF-κB transcriptional activity requires the recruitment of SRC-2/Glucocorticoid receptor-interacting protein 1, GRIP1.[34] Accordingly, conditional GRIP1-deficient mice manifested increased circulating levels of proinflammatory cytokines, TNFα, IL-1β, and IFNγ in response to a lipopolysaccharide (LPS) challenge. As previously described, SRC-2 has been characterized as NR coactivator; however, the mechanisms for its role in corepression are unclear and may depend upon the target gene and its transcriptional state.[35]

Progesterone/PR also has anti-inflammatory actions and is believed to promote myometrial quiescence during pregnancy by inhibiting the expression of proinflammatory (e.g., *IL-8* and *COX-2*) and contractile genes, including *CX43*/*GJA1* and the oxytocin receptor (*OXTR*).[36,37] This occurs, in part, via the tethering of PR to AP-1[38] and to NF-κB[36] transcription factors bound to the promoters of *CX43* and *OXTR* genes, and is associated with progesterone/PR-mediated recruitment of corepressors p54Nrb[38] and GATAD2B,[36] respectively. Interestingly, GATAD2B, which is known to interact with the repressive nucleosome remodeling and deacetylase (NuRD) complex,[39] was observed to decline in the myometrium of women in labor and in the pregnant mouse myometrium during late gestation and in labor. The corepressor p54nrb, which binds to PR tethered to AP-1 bound to the *CX43* promoter and recruits the multiprotein repressive Sin3A complex containing HDAC1 and HDAC2, declines markedly in pregnant rat myometrium toward term.[38]

MECHANISMS OF ESTROGEN ACTION

Estrogen actions are predominately mediated via the binding of estradiol-17β (E_2) to two closely related members of the NR family, ERα (ESR1)[40] and ERβ (ESR2),[41] which are encoded by separate genes. ERβ shares ~84% homology with ERα in its DBD but only 58% homology in its LBD and lesser amounts of similarity in its other domains. These structural differences are reflected in different binding affinities of ERα and ERβ for various estrogenic steroids and non-steroidal molecules.

Estrogen Receptor Ligands

As mentioned, estrogens are synthesized in the ovaries of cycling mammals and in the placentas of humans, certain nonhuman primates, and ruminants during pregnancy. Estrogens also are synthesized in the hypothalamus, amygdala, and hippocampus of the brain, in the Leydig and Sertoli cells of the testis, in bone, and adipose tissues of humans, and in endometriosis and adenomyosis lesions (see Chapters 125, 126). The key enzyme in estrogen biosynthesis is aromatase (encoded by the *CYP19A1* gene),[42,43] which converts C_{19}-androgens (e.g., testost erone, androstenedione, dehydroepiandrosterone) to C_{18}-estrogens (e.g., estradiol-17β (E_2), estrone (E_1), estriol (E_3)). In most tissues, the androgen substrate is synthesized in tissues and cells distal to the site of aromatization. For example, in the ovary, estrogens are synthesized in the granulosa cells of preovulatory follicles from androstenedione and testosterone produced by cells of the theca interna and ovarian stroma. In the placenta, estrogens are synthesized from circulating 16α-OH-dehydroepi androsterone synthesized by the fetal adrenal. In adipose tissue, estrogens are synthesized by aromatization of circulating androstenedione secreted by the adrenal cortex.[44] Estetrol (E_4) is a relatively weak estrogen synthesized from estradiol and estriol exclusively by the fetal liver during pregnancy.[45] Its function during pregnancy is essentially unknown.

KEY POINTS

- The corepressors NCoR and SMRT interact with unliganded thyroid hormone receptor and retinoic acid receptor bound to their HREs in DNA and recruit HDAC3 to deacetylate histones and promote chromatin condensation.
- These corepressors also bind to antagonist-bound ER and PR to promote chromatin condensation, as above.

The four naturally occurring estrogens, estradiol-17β (E_2, E_1, E_3, and E_4, are 18-carbon steroids with a phenolic A ring. E_2 binds with similarly high affinities to ERα and ERβ. E_2 binds to ERα with a ~25-fold higher affinity than does E_1, and a ~9-fold higher affinity than does E_3.[46] While the relative binding affinities of E_1 for ERα and ERβ are comparable, E_3 has a relatively higher binding affinity for ERβ than for ERα.[46] E_4 binds preferentially to ERα, compared with ERβ, and binds to ERα with a 4-fold lower affinity than does E_2.[45] The plant sterols/phytoestrogens, genistein, coumestrol, and daidzein, and commercially available diarylpropionitrile (DPN) have a considerably higher binding affinity for ERβ than for ERα. On the other hand, 4-propyl-1,3,5-tris (4-hydroxyphenyl) pyrazole (PPT) is highly selective for ERα and has no transactivation potential for ERβ when tested in intact cells.[46]

KEY POINTS

- Estrogen actions are primarily mediated by two members of the NR family, ERα and ERβ.
- The three naturally occurring estrogens—E_2, E_1, and E_3—are synthesized by aromatization of C_{19}-steroid androgens.
- E_2 binds with much higher affinity to ERα and ERβ than do E_1 and E_3. ERβ binds E_3 and plant sterols with relatively high binding affinity compared with ERα.

Estrogen Receptor Structure

Because of intrinsically disordered regions within the N-terminal A/B domain and the hinge region (D domain), the full-length ERα and ERβ have not been amenable to structural studies by x-ray crystallography. Instead, co-crystals of the ERα-DBD bound to an ERE, as well as ERα-LBD bound to estradiol-17β, or to other agonists and antagonists have been analyzed. Combined data from x-ray crystallography, biochemical, biophysical, and cell-based studies have provided important insight into structure-function relationships of ERα. As is the case for other steroid receptors, the LBD of ERα comprises 12 α-helices (H1–H12). In the absence of hormone binding, the LBD of the ERα monomer is partially disordered and bound by hsp90 and other immunophilins, as described above. Upon E_2 binding, the hsps dissociate, ERα dimerizes and is stabilized in a conformation in which H12 folds over the ligand-binding pocket to form an active AF2 conformation, comprising a hydrophobic groove into which coactivators can bind. By contrast, when an SERM, such as tamoxifen, binds to the LBD, the tamoxifen side chain prevents H12 from forming an active AF2 conformation. Thus, H12 docks in the AF2 hydrophobic coactivator-binding groove and blocks coactivator binding. Notably, while tamoxifen can act as an ERα antagonist in the breast, it has ERα agonist activity in the uterus, liver, and bone. This has led to the concept that the cellular environment plays an important role in determining whether an SERM acts as an ER agonist or antagonist.[47]

ERα plays a major role in the growth and pathogenesis of human breast cancer. In estrogen-dependent breast cancer, treatment with SERMs, such as tamoxifen, and with aromatase inhibitors has proved efficacious in enhancing survival; however, treatment resistance frequently develops. In fact, with prolonged treatment, tamoxifen action can be switched from that of an antagonist to that of an agonist.[47] This may be due to alterations in the cellular environment. In ~40% of recurrent ERα-positive breast cancers, mutations in the LBD have been detected, resulting in a constitutively active state of ERα in the absence of agonist binding and a loss of responsiveness to ERα SERMS and antagonists or aromatase inhibitor therapy. Structural analyses of wild-type (WT) versus various mutant forms of ERα have provided important insight into the mechanisms of estrogen-dependent activation of WT ERα and for the constitutive activation of the mutant receptor. The most highly constitutively active forms of ERα harbor mutations in tyrosine-537 (e.g., Y537S) within the loop between H11 and H12; this prevents the receptor from adopting an "off" state in the absence of ligand. When WT ERα binds agonist ligand, H12 bends over the ligand-binding pocket to expose the coactivator-binding groove. This creates a spring-like strain within the loop that enables it to snap back into the inactive conformation in the absence of bound ligand. The Y537S mutation markedly alters the spring-like properties of the H11-H12 loop, keeping H12 in the agonist-bound conformation in the absence of estrogen binding. The effects of various mutations within the ERα LBD resulting in various degrees of constitutive activation are described in great detail by Katzenellenbogen et al.[48]

KEY POINTS

- As is the case for other steroid receptors, the LBD of ERα comprises 12 α-helices (H1–H12).
- Upon E_2 binding, ERα dimerizes and H12 folds over the ligand-binding pocket to form a hydrophobic groove into which coactivators can bind.
- By contrast, when a SERM such as tamoxifen binds, the tamoxifen side chain prevents H12 from forming an active AF2 conformation and coactivator binding is blocked.
- In a significant proportion of ERα-positive breast cancers, there are mutations in the loop between H11 and H12 that keep the receptor in an active conformation in the absence of ligand.
- Tamoxifen acts as an ERα antagonist in the breast, but has ERα agonist activity in the uterus, liver, and bone, suggesting that the cellular environment plays a role in determining its action as an ER agonist or antagonist.

Estrogen Receptor-Coactivator Interactions

Because of multiple intrinsically disordered regions within full-length NRs and coregulators, it has not been possible to solve the x-ray crystal structure of full-length NRs bound to full-length primary (e.g., SRCs) and secondary (e.g., p300/CBP) coregulators. Using the powerful technique of cryo-electron microscopy (cryo-EM) together with biochemical studies, Chiu, O'Malley and colleagues revealed the quaternary structure of full-length ERα bound to a 700-bp ERE in a complex with SRC-3 and p300.[49] Their intriguing findings indicate that ERα binds to the ERE as a dimer, with each molecule of ERα recruiting one molecule of SRC-3 to its AF-2 domain (Fig. 120.4)[162]. Interestingly, the AF-1 domain near the ERα N-terminus aligns in the structure in proximity with the C-terminal ERα AF-2 domain and participates in SRC-3 recruitment, providing support for prior biochemical findings suggesting that transcriptional activation of ERα is facilitated by interaction of N- and C-terminal domains of ERα.[50] Both molecules of SRC-3, in turn, act together to recruit a single molecule of p300. The p300 did not interact directly with ERα or with DNA in the cryo-EM structure.[49] The p300 in the complex exhibited conformational variability which may contribute to its ability to interact with other coregulators.[49] In this regard, subsequent cryo-EM studies by this group revealed that there is a sequential interaction of ER and different primary (SRC-3) and secondary (p300 and CARM1) coactivators at the ERE in DNA. Prior studies indicated that the arginine methyltransferase, CARM1,[51] acted synergistically with p300/CBP to enhance ER-mediated gene transcription.[52] The cryo-EM and associated biochemical studies indicated that following the binding of the ERα dimer to DNA, there is binding of two molecules of SRC-3 and p300, as

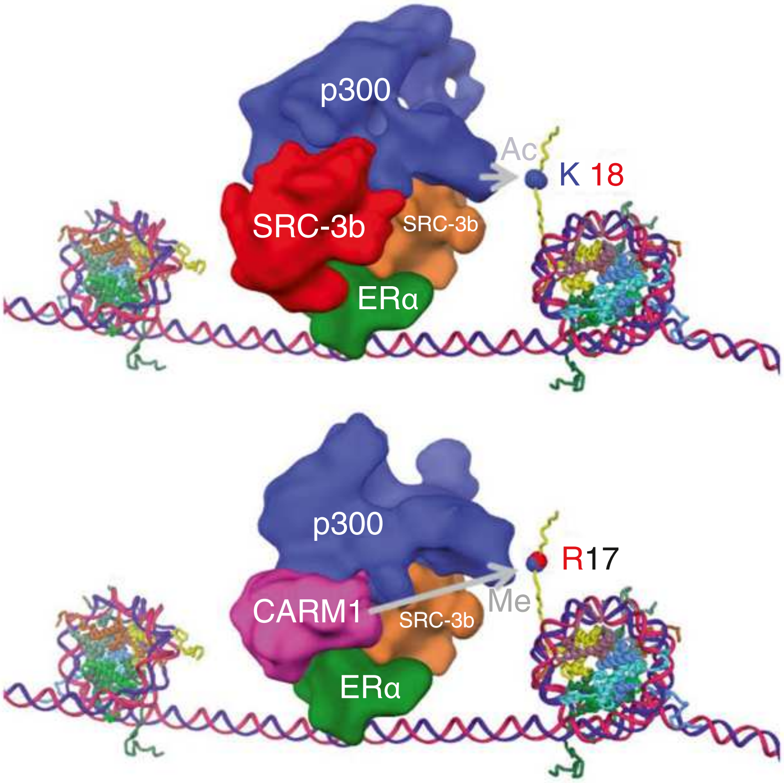


Figure 120.4 **Cryo-electron microscopy structure of ERα-coactivator interactions.** Shown is the quaternary structure of full-length ERα bound to a 700-bp estrogen response element in a complex with SRC-3, p300, and CARM1. The sequential interaction of the hormone response element-bound ERα dimer with SRC-3, p300, and CARM1 is described.[53] (From Yi P, Wang Z, Feng Q, et al. Structure of a biologically active estrogen receptor-coactivator complex on DNA. *Mol Cell.* 2015;57:1047–1058.)

reported previously.[49] Upon binding of p300, interaction of the second molecule of SRC-3 is destabilized and replaced by the binding of CARM1 (see Fig. 120.4), which now has access to methylation sites in ERα, SRC-3, and p300.[53] This causes a conformational change in p300 with further induction of its HAT activity and a cooperative increase in the histone H3 methyltransferase activity of CARM1. Thus, the cooperative interaction of sequentially recruited primary and secondary coactivators to the ERα dimer bound to DNA contributes to estrogen activation of gene expression.[53]

Molecular Mechanisms of Estrogen Receptor Action

In addition to its genomic effects, described directly below, estrogen also has rapid and acute effects on cells that appear to be mediated by a number of alternative signaling pathways. These pathways can be activated by E_2 (***ligand-dependent***) or by other signaling mechanisms within the cell (***ligand-independent***) (Fig. 120.5).

Genomic Actions by Direct ER Binding to EREs in DNA. The nuclear forms of ERα and ERβ mediate genomic actions of estrogens to alter transcription of specific genes. As mentioned, the binding of E_2 to ERα or ERβ promotes its dissociation from a complex of hsps and binding of the ER homodimer to EREs, which can be localized from hundreds to hundreds of thousands of bp upstream of their target genes. In ChIP-microarray studies of ERα binding to chromosomes 21 and 22 in the MCF-7 human breast cancer cell line, it was observed that while a relatively small number of functional EREs were localized proximal to the promoters of their target genes, the majority existed hundreds of thousands of base pairs away.[54] Importantly, binding motifs for the Forkhead protein FoxA1 were found proximal to a large proportion of the ER binding regions that contained a canonical ERE. Notably, FOXA1 was found to serve an essential role in the interactions of ERα with chromatin and the subsequent expression of estrogen target genes.[54] Thus, FOXA1, which can bind to condensed chromatin via its winged-helix structure, allows the local opening of chromatin structure to facilitate ER binding to its response elements in DNA. Moreover, chromosome capture assays indicated that estrogen-mediated gene transcription was associated with physical interaction of these distal EREs with the promoters of the estrogen-regulated genes.[54]

KEY POINTS

- Because of several intrinsically disordered domains, the structure of ER and other members of the NR family have not been solvable by x-ray crystallography.
- Instead, cryo-EM has been used to solve the quaternary structure of ERα bound to its HRE in DNA, in association with SRC-3, CBP, and CARM1.
- This confirms that ERα binds to its HRE as a homodimer, and that the AF-1 and AF-2 domains participate in recruitment of one molecule of SRC-3 each, which in turn together, recruit p300.
- One molecule of SRC-3 is then destabilized and replaced by coactivator CARM1 (see Fig. 120.4).

As described above, the binding of ER to its response elements is followed in time by the recruitment of primary and secondary coactivators, which posttranslationally modify histones to cause a further opening of chromatin structure to facilitate the binding of ATP-dependent remodeling complexes to break histone-DNA contacts, allowing repositioning of nucleosomes and assembly of the general transcription machinery and RNA Pol II, resulting in transcription initiation.

Genomic Actions by ER Tethering to Other Transcription Factors Bound to DNA. Ligand-dependent alternative mechanisms include the E_2-induced tethering of ERα or ERβ to other transcription factors bound

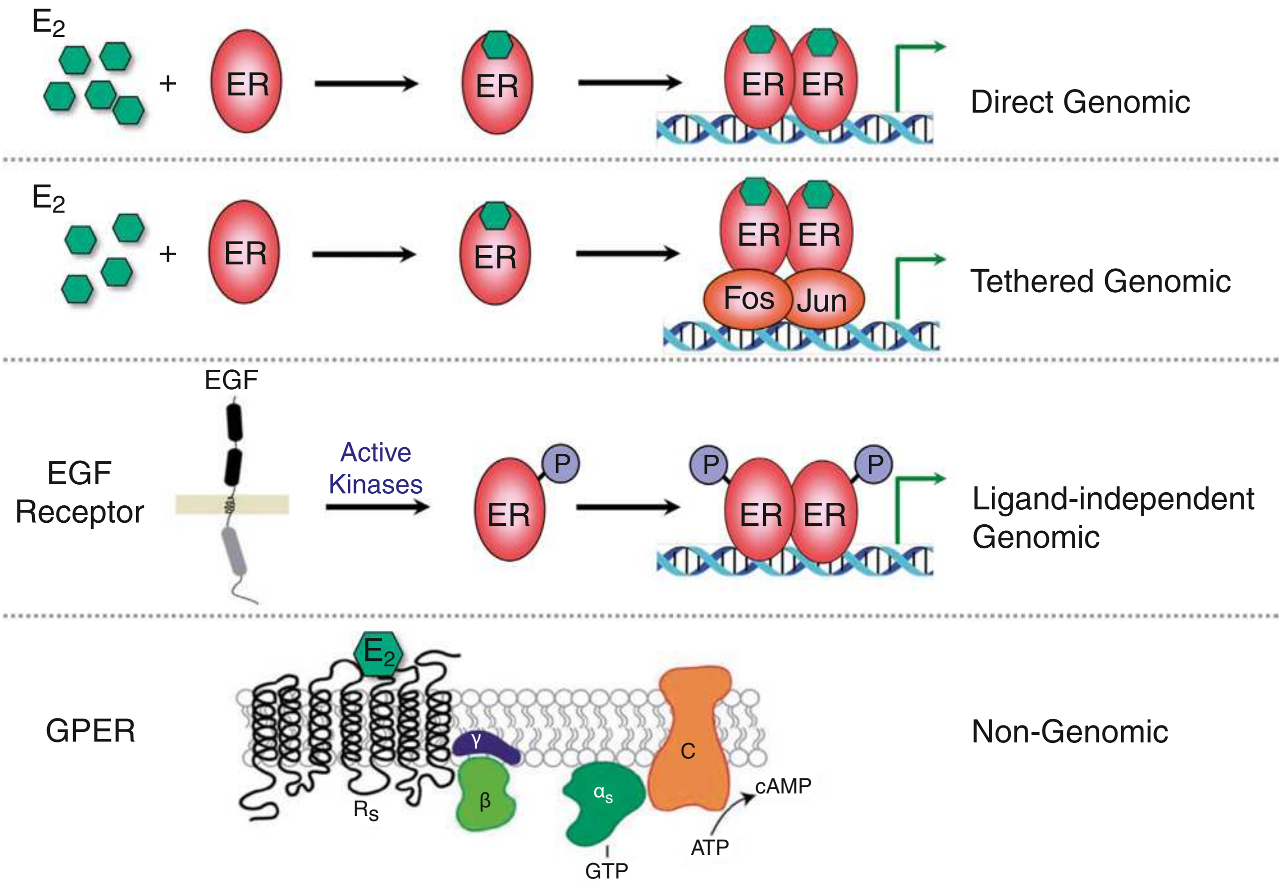


Figure 120.5 Molecular pathways in estrogen receptor action. These actions include direct binding of ER homodimers to DNA and tethering to Sp1 or AP-1 transcription factors bound to DNA, ligand-independent activation of ER bound to HREs in response to growth factor-mediated phosphorylation and rapid non-genomic effects.

to their response elements in DNA, as has also been described for GR and PR (see Fig. 120.5). These include ERα tethering to specificity protein 1 (Sp1) bound to G/C-rich response elements, as well as ERα or ERβ interaction with Fos/Jun bound to AP-1 response elements. Sp1 is required for E_2 regulation of many genes involved in cell proliferation, differentiation, and angiogenesis in cancer cell lines.[55] Genes found to be E_2-responsive via the interaction of ERα with Sp1 bound to GC-rich sites include: *ESR1*, *PGR*, *PRLR*, *EGFR*, *IGF1R*, *HOXA10*, *cMYC*, *cathepsin D*, and *telomerase*, to name a few.[55,56] Notably, transcription of these genes is activated both by binding of E_2 or ER antagonists. ERα binding to Sp1 was found to increase the stability of Sp1-DNA interactions. Whereas ERα can interact with Sp1 in a ligand-independent manner, enhanced transcription of target genes is driven by coactivator recruitment, which is E_2-dependent. Notably, in cancer cell lines expressing ERβ, E_2 did not activate promoters driven by ERβ-Sp1 interactions. From domain-swapping experiments, it appeared that the AF-1 domain of ERα was critical for E_2-dependent transactivation, whereas the corresponding AF-1 domain of ERβ was inactive.[57]

Both ERα and ERβ can tether to Fos/Jun bound to AP-1 sites in the promoters of a number of genes involved in cell growth, motility, and differentiation, including *MMP1*, *FN1*, *cyclin D1*, and *IGFBP4*. Recruitment of ERα to AP-1 bound to gene promoters stabilized the binding of p160 coactivators and p300 to the complex.[58] Gene activation by ERα tethered to AP-1 was stimulated both by E_2 and by antiestrogens. In the case of *cyclin D* promoter, wherein E_2 binding to ERα tethered to AP-1 had a stimulatory effect, E_2 bound to ERβ was inhibitory.[59] In contrast to findings of ER/Sp1 interactions, wherein ERβ was relatively inactive, in the case of ER/AP-1, antiestrogens enhanced ERβ-AP-1 transcriptional activation, whereas E_2 had little effect.[55]

Ligand-Independent ER Activation. ERs can be activated in a ligand-independent manner via phosphorylation mediated by a number of signaling cascades (see Fig. 120.5). For example, EGF was one of the first growth factors found to enhance ERα transcriptional activity in an E_2-independent manner.[60] As opposed to E_2 activation of ERα, which occurs via the AF-2 domain, EGF signaling activates ERα through its AF-1 domain. EGF induces phosphorylation of specific serine residues of ERα and ERβ by activation of MAPK and protein kinase B (Akt). ER phosphorylation may enhance recruitment of p160 coactivators. Insulin-like growth factor 1 (IGF-1) also causes ligand-independent activation of ERα through Akt signaling. As reviewed in detail previously, ligand-independent ER activation occurs in response to a variety of growth factor receptors, cytokines, second messenger pathways, and kinases.[60]

Non-Genomic Signaling. Rapid effects of E_2 also can be mediated by non-genomic signaling, whereby E_2 can bind to ER that is associated with the plasma membrane through interaction with other membrane proteins (e.g., caveolin-1, Shc, and insulin-like growth factor receptor)[61] and/or via lipid modification of ER by palmitoylation.[62] Rapid non-genomic estrogen effects also may be mediated by a G protein-coupled receptor, termed G protein-coupled ER1

(GPER1), also known as GPR30, which is coupled to $G_s\alpha$ in the plasma membrane (see Fig. 120.5).[63] E_2 binding to GPER1 is associated with rapid activation of intracellular signaling pathways, such as adenylyl cyclase, AKT, and MAPK. Notably, the SERM 4-OH tamoxifen and the selective estrogen receptor downregulator (SERD) ICI182,780 also act as agonists of GPER1.[64] Rapid non-genomic effects of E_2 appear to serve functional roles in the brain, testes, and cardiovascular system.

KEY POINTS

- E_2 has genomic and non-genomic actions (see Fig. 120.5).
- Genomic actions include: (1) direct binding of ER homodimers to HREs, recruitment of coactivators, posttranslational modifications of histones and relaxation of chromatin structure; (2) Tethering of ER to other transcription factors (e.g., Sp1, AP-1) bound to response elements in DNA, with recruitment of coactivators and modification of histones.
- Ligand-independent activation of ER bound to HREs is induced by growth factor and cytokine signaling.
- Rapid, non-genomic actions are mediated by E_2 binding to ER associated with the plasma membrane, or with GPER1.

Use of Gene-Targeted Mice To Ascertain the Biology of Estrogen Receptor Function

Genetically engineered mice have been used extensively to investigate the role of ERα and ERβ in the regulation of mammalian physiology *in vivo*. Not only has this approach been used to dissect the function of the ERs in the regulation of tissue homeostasis, but also to dissect the functional role of the domains of the ERs (reviewed in Hamilton et al.[65]). The ablation and mutagenesis of ERα and ERβ resulting in the αERKO and βERKO mouse, respectively, have revealed both distinct and overlapping functions in animal physiology. The phenotypes of the individual and combined ablations of ERα and ERβ are discussed below.

Phenotype of the αERKO Mouse. The phenotypic analysis of the αERKO mouse revealed both predictable and surprising results. As expected, the αERKO female mice were sterile. Although female reproductive tract development was normal, the females displayed failure to mate, hypoplastic uteri, hyperemic ovaries with failure to undergo natural ovulation, no detectable corpora lutea and rudimentary mammary gland development.[66–68] The uteri and vaginas were not responsive to estrogen stimulation with no increase in uterine weight and no vaginal cornification. The impact of ERα ablation on ovulation was due to a combination of the disruption of the hypophyseal-ovarian axis, as well as a direct effect on the ovary.[69] Surprisingly, αERKO males were sterile. The male sterility was not due to a direct role of ERα in spermatogenesis, but due to an impairment of fluid resorption in the rete testis that indirectly reduced spermatogenesis with time.[70] Phenotypes of the αERKO mouse have been seen in humans with familial mutations in ERα.[71,72] In addition to investigating the role of ERα in the reproductive system, the αERKO mouse has been used to investigate its role in the skeletal system, cardiovascular system, and metabolism (reviewed in Hewitt et al.[73]).

Phenotype of Mice With Tissue-Specific Ablation of ERα. The global knockout models for ERα action have defined the role of the loss of estrogen responsiveness in the whole animal. Since estrogen impacts many organ systems in the hypothalamo-hypophyseal-gonadal-uterine axis, some of the phenotypes of ERα ablation on specific target organs are confounded by indirect endocrine effects. Conditional ablation of the ERα has been used to shed light on the role of ERα not only in a specific tissue but also within compartments of that tissue. Mice lacking ERα in the pituitary gonadotrophs have impaired fertility. These mice displayed higher serum levels of luteinizing hormone (LH) but lacked the LH surge required for ovulation. This demonstrated that ERα regulation of the control of gonadotropin feedback is in part dependent on ERα in the pituitary.[74] Ablation of ERα in the mammary epithelium demonstrated that ERα regulated the establishment of the mammary gland architecture and lobuloalveolar gland development[75] (see Chapter 127). Conditional ablation of ERα in the skeletal system demonstrated that ERα is critical for the maintenance of bone mass.[76] Liver-specific ablation of ERα demonstrated that ERα plays a key role in hepatic gluconeogenesis and lipid accumulation in the liver.[77] However, the most in-depth investigation of ERα has been on the role of ERα in the regulation of female reproductive tract function.

The female reproductive tract serves as a conduit for gamete transport to the site of fertilization in the oviduct and the subsequent preparation of the uterine tissue for implantation and fetal development. Prior to embryo implantation, the uterus undergoes specific changes that allow the embryo to attach, establish appropriate contacts with maternal tissue, and receive proper nutrition during pregnancy. The process of embryo implantation is precisely coordinated and is limited to a specific time period termed the “window of receptivity.”[78] The window of receptivity and the ability of the uterus to support implantation are under the control of the ovarian steroids E_2 and P_4. In the uterus, after ovulation, estrogen acting through ERα regulates the proliferation of the uterine epithelium prior to implantation. ERα also regulates the expression of the cytokine leukemia inhibitory factor 1, LIF.[79] This is a cytokine expressed in the uterine glandular epithelium that signals to the endometrial luminal epithelium to regulate embryo attachment.[80] Genetically engineered mouse models have demonstrated that the processes regulated by ERα are through paracrine, as well as direct, mechanisms.

Uterine epithelial proliferation in response to estrogen stimulation is one of the classical examples of stroma to epithelial paracrine signaling. The initial observation that ERα in the uterine stroma regulated epithelial cell proliferation was accomplished by tissue reconstitution experiments using αERKO and WT uterine epithelia and stroma.[81] These reconstitution studies demonstrated that ERα in the stroma was critical for the regulation of uterine epithelial proliferation, while ERα in the epithelial compartment was not able to induce epithelial proliferation. These reconstitution assays were later confirmed when ERα was specifically ablated in the uterine epithelium and estrogen was still capable of inducing epithelial proliferation.[82] Stromal cell production of IGF-1 and fibroblast growth factors (FGFs) are critical in regulating epithelial cell proliferation.[83,84] The role of ERα in the uterine epithelium was found to be important for the regulation of expression of LIF in the uterine glands.[82,85] This analysis showed that ERα acting in both the epithelium and stroma compartments was critical for the regulation of uterine function. The role of ERα signaling in the mouse uterus is shown in Fig. 120.6.

In addition to the regulation of uterine epithelial cell proliferation and embryo implantation, conditional ablation of the ERα in the female reproductive tract uncovered novel findings with respect to the role of ERα in pregnancy. Epithelial specific ablation demonstrated that ERα regulated the many processes supporting fertilization and embryo transport. ERα regulates the ability of the uterus to liquify the seminal fluid that is necessary for sperm to reach the site of fertilization. ERα also regulates the transport of the embryos through the oviduct and protects the embryo from protease-mediated embryo degradation.[86–88] This work demonstrates that ERα is critical for all aspects of preimplantation development of the embryo.

***In Vivo* Dissection of the Functional Domains of ERα.** In addition to identifying the role of the ERs *in vivo*, gene-targeted mice have been generated to identify the roles of the functional domains of ERα in

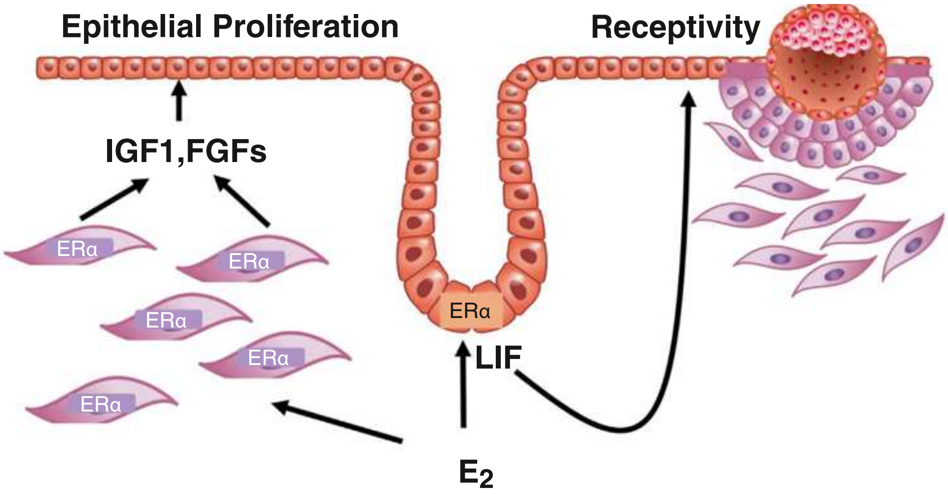


Figure 120.6 A schematic of the pathways regulated by ERα during the preimplantation period governing embryo implantation and cell proliferation.

order to determine the relative importance of the signaling pathways by which ERα regulates hormone responsive processes. Mice have been generated with mutations in activation domains, AF-1 and AF-2, the DBD, and LBD. These mutations not only defined the role of each domain in determining ERα activity but also the relative importance of the specific signaling strategies, such as classical, non-classical, and ligand-independent signaling in the physiology of an organ system *in vivo*.[73]

As mentioned above, ERα contains two transactivation domains. The AF-1 domain located in the amino terminus of ERα and the AF-2 domain located within the LBD. Mutations in the activation domains of ERα identified their relative importance in the regulation of the transactivation of gene expression and the ability of ligands to regulate transcriptional activity of ERα. Mice with mutations in the AF-1 domain showed that this domain was important for the hypertrophic response of the uterus to estrogen. Mice with mutations in this domain did not show an increase in stroma and epithelial proliferation in response to estrogen stimulation. However, the uteri of these mice did develop uterine edema in response to estrogen stimulation. This was due to the induction of vascular endothelial growth factor-A, Vegf-A.[89,90] Systematic analysis of these mice showed that the vascular system was intact, demonstrating that the AF-1 domain was not necessary for estrogen action on the vascular system.[89]

In contrast to mice with AF-1 mutations, mice with AF-2 mutations elicited a similar phenotype as the ERKO mouse, with female mice showing hypoplastic uteri, hemorrhagic and cystic ovaries, impaired mammary gland development, and high levels of serum LH and estradiol.[91] Male mice also showed a progressive sterility phenotype due to efferent duct fluid resorption.[92] Interestingly, ERα antagonists could induce a uterotropic response in female mice and reversed the male sterility phenotype. The antagonists did not impact serum levels of LH, demonstrating that while the antagonists could activate the mutant receptor in the uterus, it has no impact on the pituitary. This demonstrated the tissue-specific role of the AF-2 domain in regulating ligand specificity in ERα action.[91,92]

In order to define the *in vivo* role of ligand-dependent and ligand-independent activation of ERα, mice were generated with a mutation in the LBD. These mice showed a similar phenotype to the αERKO mice with respect to elevated serum LH levels, lack of ovulation, hemorrhagic cystic follicles, hypoplasia of the uterus, and underdeveloped mammary glands. This demonstrated that the majority of actions of ERα was dependent on ligand-dependent activation. However, when assaying the ability of growth factors to stimulate uterine proliferation, mice with the mutation in the LBD were responsive to growth factor stimulation whereas αERKO mice were not. This demonstrated the importance of ligand-independent ERα signaling in sensitizing the uterus to growth factor stimulation.[93]

Finally, genetically engineered mice were generated to investigate the role of classical versus non-classical ERα signaling. To investigate the role of ERα action by binding its DNA response element in the genome, or its non-genomic actions through tethering to other transcription factors, mice were generated with a mutation in the DBD. Two strategies were employed. The first strategy generated mice in which alanine was substituted for the glutamic acid at position 207 and for the glycine at position 208 in the first zinc finger of the DBD. The phenotype of these mice was surprising. Female mice heterozygous for these mutations in the DBD were sterile. They failed to ovulate and develop a corpus luteum. The mice had cystic uteri with endometrial hyperplasia and the mammary glands were hypoplastic. These ovarian and uterine phenotypes were independent of that observed in the αERKO.[94] Genome-wide studies of this ERα mutant revealed that this mutation of the DBD did not ablate DNA binding, but changed the specificity of the mutant ERα to bind HREs of other NRs. This gain of function phenotype could be due to a novel interaction of the DBD of the ERα mutant with the WT ERα or to a novel function generated by this mutation.[95] A second mouse model was generated with four amino acid substitutions at Tyrosine 201, Lysine 210, Lysine 214, and Arginine 215 in the DBD. The phenotype of these mice had a closer resemblance to the αERKO phenotype. These mice showed that the major activity of ERα is through the classical signaling.[96] These studies clearly demonstrate the importance of *in vivo* studies. DBD mutations that have been evaluated in a limited manner *in vitro* may not identify more globally important functions that would be identified *in vivo*.

Phenotype of the βERKO Mouse. Several lines of mice have been generated with different modifications of the ERβ gene locus.[67,97,98] Depending on the mouse model, males and females have either impaired fertility or are sterile. However, it is clear from these mouse models that the major impact of ERβ is on ovarian function and ovulation.[98] The most interesting phenotype is the result of the combined knockout of ERα and ERβ. The phenotype of this double knockout mouse results in the ovarian follicles which have granulosa cells that resemble the male Sertoli cell. This demonstrates that both ERs are critical for maintaining the phenotype of the granulosa cells in the ovaries.[67,69]

KEY POINTS

- Genetically engineered mice have been a useful tool in the structure and function analysis of ER activities.
- ERα is critical in regulating both male and female reproduction as well as other systems including metabolism, cardiovascular, and skeletal muscular systems.
- ERα regulates uterine function by directly activating the expression of specific target genes as well as coordinating the paracrine crosstalk between the cellular compartments of the endometrium.
- Mutations in the functional domains of ERα demonstrated that DNA binding in response to ligand binding is responsible for the majority of actions in the regulation of reproduction.
- ERβ regulates ovarian function and ERα and ERβ work together to maintain the sexually dimorphic phenotype of the ovarian follicles.

MECHANISMS OF PROGESTERONE ACTION

Progesterone Receptor Isoforms

Progesterone acts via binding to its NRs within reproductive tissues, including the uterus, mammary gland, ovary, and brain, and in nonreproductive tissues, including the cardiovascular system, bone, and the central nervous system.[99] In humans, there are two major PR isoforms, PR-A (94 kDa) and PR-B (114 kDa), which are alternative transcripts of a single gene[100,101] and differ only in the lengths of their A/B domains. PR-A and PR-B have identical DBD and LBD, but PR-B contains an additional 164 bp at its NH_2-terminus. Both PR-A and PR-B bind to PREs in DNA. PR-B contains the two major transcriptional activation domains found in PR-A and in other steroid hormone receptors (AF-1, AF-2), but has an additional AF domain within its extended NH_2-terminus (AF-3). Thus, in certain cell- and gene-specific contexts, PR-B is more transcriptionally active than PR-A, and PR-A can repress PR-B transcriptional activity by competing for binding to PREs and coregulators.[102,103]

A third N-terminally truncated PR isoform, PR-C, has been characterized, which lacks part of the DBD and is restricted primarily to the cytoplasm.[104,105] Because PR-C cannot bind to DNA but can bind to progesterone, it may inhibit PR function by sequestering available hormone away from PR-B. PR-C also can bind to PR-B and reduce its capacity to bind to DNA.[104]

A truncated mitochondrial form of PR (PR-M) encoded from the *PGR* gene has been identified. PR-M transcription initiates in the third intron of the *PGR* gene, contains a putative translation start site followed by 16 unique hydrophobic amino acids at its NH_2-terminus similar to that of import sequences of outer mitochondrial membrane proteins. This is followed by sequence identical to exons 4 to 8 of *PGR*. PR-M lacks the A/B and C domains and is missing a complete nuclear localization sequence.[106] In MCF-10A cells, which lack nuclear PR but express PR-M, progestin treatment was found to increase mitochondrial membrane potential and ATP production.[107] It has been suggested that PR-M mediates P_4 action to increase cellular energy production.[106]

Membrane G protein–coupled PR isoforms (mPRα and β), linked to $G_i\alpha$ adenylyl cyclase inhibition, p38 MAPK activation, and increased myosin light-chain phosphorylation also have been characterized in human myometrial cells.[108] Increased expression of mPR in the myometrium of women in labor, compared with those not in labor, was suggested to contribute to increased myometrial contractility.[108] Five subtypes of mPR have been described in neural tissues (α, β, γ, δ, ε), which bind P_4 and the neurosteroid allopregnanolone. Allopregnanolone, which binds with highest affinity to mPRδ, acts through Gαs to inhibit apoptosis and promote neuroprotection.[109]

KEY POINTS

- In humans, there are two major PR isoforms, PR-A and PR-B, which are formed by alternate transcription initiation from a single gene.
- PR-A and PR-B have identical DBD and LBD, but PR-B contains an additional 164 bp at its NH_2-terminus, which contains a third activation domain.
- Two other truncated PR isoforms are produced by alternative transcription initiation: (1) PR-C, which lacks part of the DBD, is primarily cytoplasmic, and may be an inhibitor of PR function; and (2) a mitochondrial form, PR-M.
- There are also several membrane-associated G protein-coupled PR isoforms.

Use of Gene-targeted Mice To Evaluate the Biology of PR-A and PR-B Function

The PR, by its name, is the NR of pregnancy. PR has been shown to regulate all aspects of female reproduction, from female sexual behavior to processes regulating pregnancy and maternal care of newborns through its actions in the central nervous system, pituitary, ovary, female reproductive tract, and mammary gland. Much of the *in vivo* analyses of PR action has been accomplished by the use of gene ablation, site-directed mutagenesis, and transgene expression of the receptors in mouse models.

Phenotypes of the Ablation of PR and Its Specific Isoforms. Using standard homologous recombination in embryonic stem cells, both PR-A and PR-B isoforms were ablated in the mouse, generating the PRKO model. The PRKO mouse demonstrated that PR regulated female sexual behavior, maternal behavior, ovulation, all aspects of uterine function, embryo implantation, immune function, and mammary gland development.[110–113] In addition to these functions, the PRKO mouse demonstrated that PR was responsible for inhibiting the mitogenic effects of E_2 on uterine epithelial proliferation.[110]

Following the ablation of both PR isoforms, site-directed mutagenesis was conducted to determine the individual functions of the PR-A and PR-B isoforms *in vivo*.[114,115] The specific ablation of PR-A *in vivo* demonstrated that in the mouse reproductive tract, PR-A was the dominant functional isoform. PR-AKO mice did not ovulate, and the uterus was not able to support pregnancy. However, PR-A ablation had no effect on mammary gland development. These isoform-specific ablated mice demonstrated that PR-B induced uterine epithelial proliferation in response to P_4 treatment. This is a unique role for the PR-B isoform since PR serves an important role in preventing uterine epithelial proliferation in response to E_2 stimulation.[115] Mice with specific ablation of PR-B showed normal ovarian and uterine function. However, mammary gland development was impaired.[114] Mice with targeted ablation of PR-A and PR-B demonstrated the specific roles of these isoforms in mouse reproduction, with PR-A being the major regulator of ovarian and uterine function, while PR-B regulates mammary gland development. These tissue-specific actions of the PR-A and PR-B isoforms in the mouse may be either due to their tissue-specific activation of target genes, or to the stoichiometry of the expression of these isoforms in different tissues. Such findings in mouse models do not address the role of these isoforms in human reproduction or that of other species where the stoichiometry of these isoforms may be different.

In addition to the investigation of the role of PR-A and PR-B by isoform ablation, *in vivo* site-directed mutagenesis has been used to investigate the role of phosphorylation on the activation of PR. Using site-directed mutagenesis, serine 191 of PR was mutated to an alanine. The resulting mouse showed a 19% decrease in litter size and alteration in the expression of a subset of PR target genes and of PR binding to selective enhancers.[116] This analysis demonstrated the *in vivo* significance of the activation of PR by phosphorylation signaling cascades. The gene-targeted mice, described above, have elucidated the role of

PR in the whole animal. Cell-specific knockouts and gene activation have also been used to decipher the role of PR *in vivo*.

Molecular Mechanisms of Progesterone Action in the Endometrium and on Blastocyst Implantation

Preimplantation Mouse Uterus. PR regulates transcription in both epithelial and stromal cell compartments of the endometrium at specific stages during the preimplantation period. PR functions to directly regulate the transcription of genes required for cell autonomous function, as well as to coordinate the paracrine crosstalk between the epithelial and stromal cells.[81] The expression of PR in the epithelium is spatially and temporally regulated. PR expression increases in the uterine epithelium during the preimplantation period and is lost from the epithelial compartment prior to embryo attachment.[117] The loss of PR in the epithelial compartment marks the window of receptivity.[118] Expression of PR in the uterine epithelium during the preimplantation period functions to inhibit the postovulatory cell proliferation caused by the ovulatory E_2 surge and primes the uterine epithelium and stroma for embryo implantation and the support of the developing embryo.[119] In the endometrial stroma, PR expression increases prior to embryo implantation and is maintained in the stroma throughout pregnancy. PR expression in the stromal fibroblasts promotes their proliferation and allows these cells to undergo decidualization. Decidualization of these stromal cells involves a change in morphology from a fibroblastic phenotype to an epithelioid-like phenotype. Decidualization supports embryo development by supplying nutrition to the embryo and protecting the embryo by modulating maternal inflammation and immune response.[79] The signaling pathways required for PR regulation of uterine function have been identified using the PRKO mouse, as well as mice with compartment-specific ablation of PR, as described above.

KEY POINTS

- Ablation of PR demonstrates that this receptor is critical for all aspects of reproduction including postpartum behavior.
- Specific ablation of the PR-A and PR-B isoforms demonstrated that PR-A is the major regulator of uterine function during the preimplantation period, whereas PR-B is critical for mammary gland development.

PR Regulated Epithelial-Stroma Crosstalk. PR regulates the paracrine crosstalk that controls epithelial cell proliferation and stromal cell decidualization. This crosstalk is initiated by PR induction of the epithelial morphogen Indian hedgehog (IHH) (Fig. 120.7).[119–121] IHH then signals to the endometrial stromal cells to regulate the activity of the orphan nuclear receptor COUP-TFII.[121] COUP-TFII regulates the expression of PR in the stromal cells.[122] Stromal PR in turn regulates the proliferation of the uterine epithelium by regulating the transcription of the helix-loop-helix transcription factor Hand2 in stromal cells.[84] Hand2 serves to inhibit the expression of FGFs by the stromal cells. The decrease in FGF signaling halts epithelial cell proliferation. PR also regulates the growth factor sensitivity of stromal cells to facilitate the decidual reaction. Stromal PR increases the levels of epidermal growth factor receptor 1 (EGFR1) in these cells.[123] Upon embryo attachment, the EGFR1-sensitive stromal cells increase the expression of bone morphogenetic protein 2 (BMP-2), and Wingless-related integration site (Wnt) family member Wnt4.[124] BMP-2 and Wnt4 expression is maintained in the stromal cells by PR. EGFR, BMP-2, and Wnt4 function to regulate the proliferation, cell survival, and differentiation of the stromal cells required for decidualization.[122,124,125] This signaling pathway is shown in Fig. 120.7.

Loss of Epithelial PR Is Required for Embryo Implantation. Not only is the presence of PR action critical for the establishment and maintenance of pregnancy, but also the loss of *PGR* expression in the murine uterine epithelium during the preimplantation period is critical for attachment.[118,126] Mice generated with the constitutive expression of PR-A in the uterine epithelium manifested defects in embryo attachment and stromal decidualization. Not only were PR-regulated genes expressed through the window of receptivity, but also two genes critical for implantation were deregulated. The ERα regulated gene *Lif* whose expression is critical for embryo implantation was inhibited by PR-A constitutive expression prior to implantation. In addition, the cellular localization of the transcription factor FOXO1 was altered. FOXO1 is localized in the nucleus of the uterine epithelium during the window of receptivity. Loss of FOXO1 results in the inability of the embryo to invade the uterine epithelium.[127] Continued expression of PR-A through the window of receptivity in the uterine epithelium caused the retention of FOXO1 in the cytoplasm, impairing embryo invasion. Therefore, in addition to its role in activation of genes required for embryo implantation, the timing of PR expression in the uterus also is of critical importance.

PR Interacting Transcription Factors. PR regulation of uterine function depends on the induction and activation of other uterine transcription factors. Chromatin immunoprecipitation, combined with whole genome sequencing, ChIP-seq, in mouse uterus and in primary human endometrial stromal cells was employed to identify these factors.[128] ChIP-seq analysis not only identified the direct targets of PR binding to response elements in the chromatin of the mouse uterus and human endometrial stromal cells but also identified motifs for other potential interacting transcription factors. These factors function to regulate gene expression in the endometrial cells. In mouse uteri, GATA2 and SOX17 motifs were identified as serving a role in PR binding with in the mouse genome.[128]

Gata2 and Sox 17 are transcription factors expressed under PR control in the epithelial compartment of the mouse uterus. Gata2 belongs to the six-member family of zinc-finger transcription factors that bind to the DNA consensus sequence (T/A)GATA(A/G) via two zinc-finger domains.[129] Sox17, a key regulator of embryo implantation, is a transcription factor belonging to the family of sex-determining region Y (SRY)-related high-mobility group (HMG) box, namely SOX.[130] Deletion of *Gata2* and *Sox17* in the uterus resulted in the inability of the uterus to support embryo implantation. GATA2 was demonstrated to regulate the expression of PR and cooperate with PR in the regulation of target gene transcription.[131] Sox17 was shown to regulate uterine epithelial gland development and the expression of *PGR* target genes. Analysis of PR, Sox17, and Gata2 regulation of the morphogen IHH showed that they bind to an enhancer located 19 kb distal to the *Ihh* promoter. CRISPR/Cas9 gene editing of this enhancer demonstrated that this region was critical for uterine-specific expression of Ihh. Thus, PR, Gata2, and Sox17 bind to a distal enhancer to regulate the expression of *Ihh* in the uterine epithelium. Analysis of the PR, Gata2, and Sox17 ChIP-seq data identified other genes containing overlapping binding sites for these transcription factors, many of which were associated with the regulation of uterine function. This indicates that regions of the PR interact with specific combinations of transcription factors to regulate uterine epithelial gene expression.[132]

KEY POINTS

- PR regulates embryo implantation by controlling epithelial proliferation and stromal cell differentiation during the preimplantation period.
- The timing and cell-specific expression of PR is precisely coordinated to render the uterus receptive for embryo implantation and support of the developing embryo.
- PR regulates a complicated paracrine signaling network to prepare the endometrium for embryo implantation.
- PR works with several other transcription factors in regulating uterine epithelial gene expression.

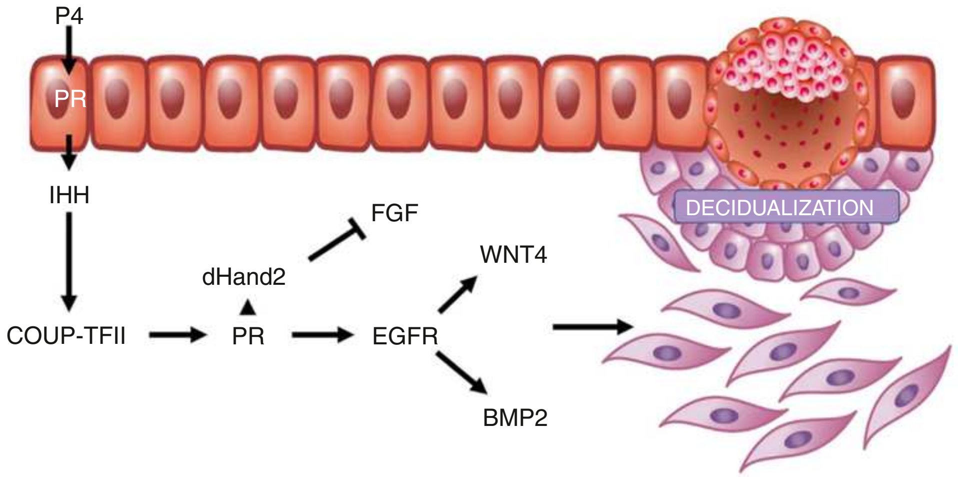


Figure 120.7 Pathways regulated by the PR in controlling epithelial to stroma two-way crosstalk in the regulation of embryo implantation.

Molecular Mechanisms of Progesterone Action in the Pregnant Myometrium

Throughout most of pregnancy, P_4 produced by the ovarian corpus luteum in rodents, and initially by the corpus luteum and subsequently by the placenta in humans, certain non-human primates, sheep, horses, cats, and guinea pigs, maintains myometrial quiescence.[133] The human placenta produces ~300 mg of P_4 per day,[133] resulting in circulating levels of ~10^{-7} M, which are far in excess of the K_d for P_4 binding to PR in target cells. The mechanisms whereby P_4/PR maintains myometrial quiescence are multifactorial and cooperative. One of the most important ways by which P_4/PR suppresses myometrial contractility is via its anti-inflammatory actions. The anti-inflammatory and immunosuppressive actions of P_4 during pregnancy were originally proposed by Siiteri et al. more than 40 years ago.[134,135] We now know that P_4 exerts these effects in the pregnant myometrium, in part, by: inhibiting transcriptional activity of the proinflammatory transcription factors, NF-κB[37] and activating protein 1 (AP-1),[38] via tethering of PR and recruitment of corepressors;[36,38,136]; upregulating expression of IκBα, which sequesters NF-κB in the cytoplasm and prevents its activation and nuclear translocation[37]; and by enhancing expression of MAP kinase phosphatase-1 (MKP-1)/dual specificity phosphatase-1 (DUSP1), an inhibitor of MAPK signaling.[137,138] Inhibition of MAPK, in turn, inhibits activation of NF-κB and AP-1 and their inductive effects on contractile gene expression (Fig. 120.8).[139]

An additional mechanism whereby P_4/PR maintains myometrial quiescence during pregnancy is by upregulating expression of the transcriptional repressor, ZEB1/TCF8/δEF1, which binds to the promoters of the contractile genes *OXTR* and *CX43* to inhibit their expression (see Fig. 120.8).[140–142] Furthermore, P_4 induces expression of the transcription factor STAT5b,[143,144] a negative regulator of the P_4-metabolizing enzyme 20α-hydroxysteroid dehydrogenase (20α-HSD) in reproductive tissues.[145] Thus, in the presence of elevated P_4/PR, STAT5b expression is increased, resulting in inhibited expression of 20α-HSD in the myometrium. This allows local levels of P_4 to remain elevated to maintain PR activation of myometrial quiescence.

Mechanisms for the Loss of PR Function in the Pregnant Myometrium Near Term Leading to Labor

In women and guinea pigs, circulating levels of P_4 remain elevated throughout pregnancy and into labor because of persistent high levels of placental P_4 production. Even in rodents, where circulating levels of P_4 decline precipitously at term, plasma levels remain in excess of the K_d for binding of P_4 to PR.[146] Thus, near term in all species, mechanisms must be in place to inhibit PR function and/or reduce tissue levels of P_4 leading to the initiation of labor. These include the following:

KEY POINTS

- PR acts in the myometrium to maintain quiescence during pregnancy by increasing expression of transcriptional repressors, such as ZEB1 and STAT5b.
- ZEB1 inhibits contractile genes (e.g., *OXTR*, *CX43*); STAT5b inhibits 20α-HSD.
- PR also maintains myometrial quiescence by tethering to and inhibiting activation of transcription factors (e.g., AP-1 and NF-κB) that increase transcription of contractile genes.

Altered Expression of PR Coregulators. Expression of selected SRCs and CBP was found to decline in the myometrium of women in labor, compared with non-laboring myometrium, and was decreased in the myometrial tissues of pregnant mice at term, along with decreased tissue levels of acetylated histone H3.[147] Thus, decreased expression of PR coactivators and a decline in histone acetylation in the myometrium during late gestation may impair the capacity of P_4/PR to transactivate genes that maintain myometrial quiescence. Accordingly, treatment of pregnant mice with the HDAC inhibitor, trichostatin A, during late gestation increased myometrial histone acetylation and delayed parturition by 24–48 hours.[147] The decline in PR coactivators and HAT activity may contribute to the decreased repressive activity of the inhibitory transcription factors ZEB1 and STAT5b near term. As mentioned, the decline in ZEB1 near term likely results in upregulation of the contractile genes *CX43* and *OXTR*. The decrease in STAT5b may promote increased expression of the P_4 metabolizing enzyme 20α-HSD, decreased myometrial P_4 levels and reduced PR activation.

In addition, corepressors, GATAD2B[36] and p54nrb,[38] which interact with PR and mediate its capacity to repress proinflammatory and contractile gene expression by tethering to NF-κB and AP-1, respectively, were found to decrease in rodent myometrium at term, contributing to the increase in myometrial contractility.

Increased Expression of Truncated PR Isoforms. As described above, in certain cell- and gene-specific contexts, PR-B is more transcriptionally active than PR-A. In this regard, PR-A was found to inhibit PR-B transcriptional activity in cultured human myometrial cells,[148] suggesting a potential antagonistic role of PR-A on PR-B action in the myometrium. Notably, PR-A and PR-B were found to be differentially

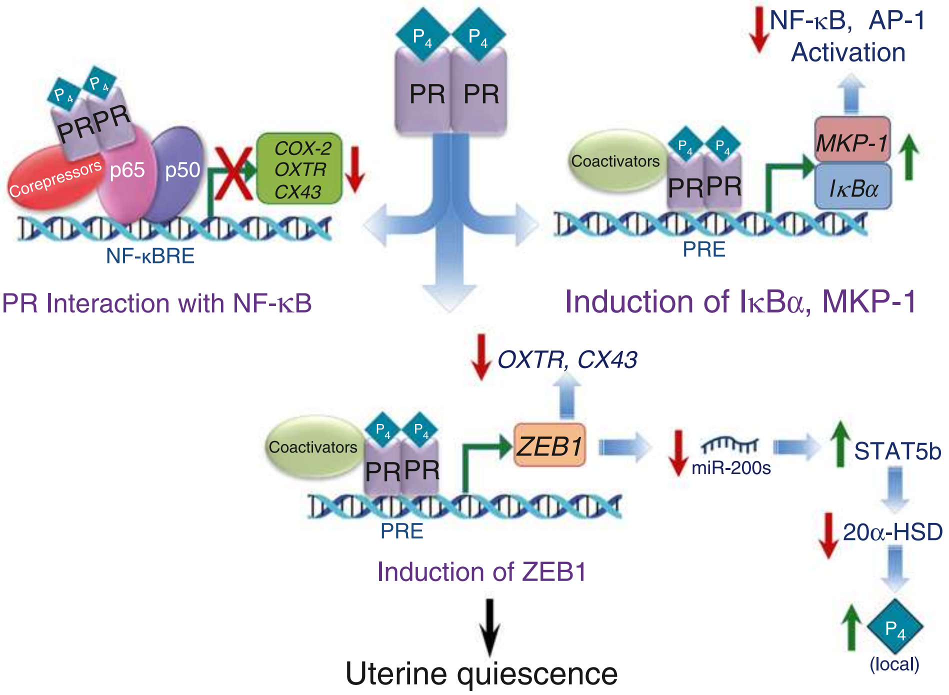


Figure 120.8 Mechanisms whereby P_4/PR mediates myometrial quiescence. These include: enhanced transcription of repressors of contractile genes (*e.g.* ZEB1) and of STAT-5b, which represses the gene encoding the P_4-metabolizing enzyme 20α-HSD; tethering to and repressing transcriptional activators of contractile genes (e.g., AP-1 and NF-κB). (Reproduced with the permission of Cold Spring Harbor Laboratory Press. Renthal NE, Williams KC, Montalbano AP, Chen CC, Gao L, Mendelson CR. Molecular regulation of parturition: a myometrial perspective. Cold Spring Harb Perspect Med. (2015) 5:181–96. doi: 10.1101/cshperspect.a023069.)

regulated in the human myometrium during pregnancy[149]; the ratio of PR-A to PR-B mRNA[150] and protein[148] was observed to be increased significantly in the myometrium from women in labor compared with those not in labor at term.

Moreover, in term fundal myometrium from women before and after the initiation of labor, a marked labor-associated increase in protein and mRNA expression of a truncated (~60 kDa) PR isoform, similar in size to PR-C was observed[151]; this was spatially and temporally associated with increased activation of myometrial NF-κB.[152] Similarly, a temporal increase in the expression of a truncated PR isoform of ~60 kDa also was observed in mouse uterus during late gestation.[151]

Increased Local Metabolism of P_4 to Inactive Steroids. Increased local metabolism of P_4 to inactive products in the uterus near term by 20α-HSD[145] serves a critical role in the decline in PR function that contributes to the initiation of parturition in all mammals. In the myometrium of pregnant women at term, a pronounced decrease in the ratio of P_4 to 20α-dihydroprogesterone (20α-OHP), an inactive metabolite of P_4 generated by 20α-HSD, was observed.[153] As discussed above, the gene encoding 20α-HSD is negatively regulated by STAT5b, which decreases in the pregnant mouse and human myometrium near term and in labor.[143]

Importantly, many of these mechanisms are driven by an increased inflammatory response with activation of NF-κB and AP-1 transcription factors, which is caused by increased uterine stretch, cooperative proinflammatory signals from mother and fetus, and a rise in circulating E_2 and ERα expression in the myometrium.[154]

Increased Levels of Circulating Estradiol-17β and Increased ERα expression in myometrium. Together with the decline in circulating P_4 levels and/or PR function, in most placental mammals, increased uterine contractility near term is associated with an increase in circulating estradiol-17β (E_2)[155,156] and/or increased ERα expression and activity.[150,157,158] Estrogens induce immune cell migration into the uterus and antagonize anti-inflammatory actions of P_4/PR.[113] Furthermore, ERα activation enhances transcription of *OXTR*,[159] *CX43*,[160] and cyclooxygenase 2 (*COX-2*/*PGHS*),[150] which catalyzes the production of contractile prostaglandins.[161]

KEY POINTS

- Multiple mechanisms contribute to a decline in PR function near term (see Fig. 120.8). These include:
- Decreased expression of PR coactivators;
- Increased expression of truncated PR isoforms; for example, increased expression of PR-A and PR-C, relative to PR-B;
- Increased metabolism of PR to inactive steroids within myometrium and cervix;
- Increased production of E_2 and expression of ERα in myometrium.

CONCLUSIONS

Estradiol-17β and progesterone serve critical roles in reproductive function that are primarily mediated by ERα/ERβ and by PR-A/PR-B, respectively. The majority of PR and ER actions are ligand-dependent, mediated by binding of E_2 or P_4 to their respective receptors, dissociation of the complex of hsps, receptor homodimerization, binding to their respective HREs in DNA or tethering to other transcription factors bound to their corresponding response elements. This, in turn, results in recruitment of coactivators or corepressor complexes, and changes in chromatin modifications to activate or inhibit target gene transcription. Through sequencing, biochemical analyses, x-ray crystallographic, and cryo-EM studies, the structures of members of the NR family and coregulators and the structural basis for ligand binding, dimerization, DNA binding, and coactivator and corepressor interactions have been elucidated.

Our understanding of the functional roles of ERα and ERβ and of PR-A and PR-B isoforms has largely been gained from studies of gene-targeted mice. While ERα and ERβ have rather wide-ranging roles in mediating estrogen regulation of female and male reproductive functions, as well as cardiovascular, bone, and metabolic homeostasis, PR primarily mediates progesterone actions in the initiation and maintenance of pregnancy. Mouse studies have demonstrated that PR regulates sexual behavior, ovulation, transport of zygotes through the oviducts into the uterine cavity, embryo implantation, and maintenance of myometrial quiescence during pregnancy. In addition, PR modulates maternal behavior, mammary gland development, and lactation. Like the ERs, ERα and ERβ, which are products of different genes, PR has two major isoforms (produced by alternative transcription initiation of a single gene) that synergize or antagonize one another's actions in a tissue-/cell-specific manner. Understanding the roles of these isoforms and the cell-specific factors that impact their different biological functions in normal and cancer cells will lead to better diagnostics and therapies to improve conditions that impact women's health.

REFERENCES

1. Mangelsdorf DJ, Thummel C, Beato M, et al. The nuclear receptor superfamily: the second decade. *Cell*. 1995;83:835–839.
2. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: a structural perspective. *Protein Sci*. 2018;27:1876–1892.
3. Pratt WB, Toft DO. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev*. 1997;18:306–360.
4. Rastinejad F, Huang P, Chandra V, et al. Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol*. 2013;51:T1–t21.
5. Onate SA, Tsai SY, Tsai MJ, et al. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science*. 1995;270:1354–1357.
6. Dasgupta S, Lonard DM, O'Malley BW. Nuclear receptor coactivators: master regulators of human health and disease. *Annu Rev Med*. 2014;65:279–292.
7. Urnov FD, Wolffe AP. A necessary good: nuclear hormone receptors and their chromatin templates. *Mol Endocrinol*. 2001;15:1–16.
8. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*. 2000;64:435–459.
9. Grunstein M. Nucleosomes: regulators of transcription. *Trends Genet*. 1990;6:395–400.
10. Heery DM, Kalkhoven E, Hoare S, et al. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*. 1997;387:733–736.
11. Szwarc MM, Lydon JP, O'Malley BW. Steroid receptor coactivators as therapeutic targets in the female reproductive system. *J Steroid Biochem Mol Biol*. 2015;154:32–38.
12. Xu J, Wu RC, O'Malley BW. Normal and cancer-related functions of the p160 steroid receptor co-activator (SRC) family. *Nat Rev Cancer*. 2009;9:615–630.
13. Xu J, Qiu Y, DeMayo FJ, et al. Partial hormone resistance in mice with disruption of the steroid receptor coactivator-1 (SRC-1) gene. *Science*. 1998;279:1922–1925.
14. Gehin M, Mark M, Dennefeld C, et al. The function of TIF2/GRIP1 in mouse reproduction is distinct from those of SRC-1 and p/CIP. *Mol Cell Biol*. 2002;22:5923–5937.
15. Mukherjee A, Soyal SM, Fernandez-Valdivia R, et al. Steroid receptor coactivator 2 is critical for progesterone-dependent uterine function and mammary morphogenesis in the mouse. *Mol Cell Biol*. 2006;26:6571–6583.
16. Ye X, Han SJ, Tsai SY, et al. Roles of steroid receptor coactivator (SRC)-1 and transcriptional intermediary factor (TIF) 2 in androgen receptor activity in mice. *Proc Natl Acad Sci U S A*. 2005;102:9487–9492.
17. Mark M, Yoshida-Komiya H, Gehin M, et al. Partially redundant functions of SRC-1 and TIF2 in postnatal survival and male reproduction. *Proc Natl Acad Sci U S A*. 2004;101:4453–4458.
18. Gao L, Rabbitt EH, Condon JC, et al. Steroid receptor coactivators 1 and 2 mediate fetal-to-maternal signaling that initiates parturition. *J Clin Invest*. 2015;125:2808–2824.
19. Chen J, Mishra R, Yu Y, et al. Decreased 11b-hydroxysteroid dehydrogenase 1 in lungs of steroid receptor coactivator *(Src)-1/-2* double-deficient fetal mice is caused by impaired glucocorticoid and cytokine signaling. *FASEB J*. 2020;34:16243–16261.
20. Anzick SL, Kononen J, Walker RL, et al. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science*. 1997;277:965–968.
21. Shao W, Keeton EK, McDonnell DP, et al. Coactivator AIB1 links estrogen receptor transcriptional activity and stability. *Proc Natl Acad Sci U S A*. 2004;101:11599–11604.
22. Balmer NN, Richer JK, Spoelstra NS, et al. Steroid receptor coactivator AIB1 in endometrial carcinoma, hyperplasia and normal endometrium: correlation with clinicopathologic parameters and biomarkers. *Mod Pathol*. 2006;19:1593–1605.
23. Spencer TE, Jenster G, Burcin MM, et al. Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature*. 1997;389:194–198.
24. Kim MY, Woo EM, Chong YT, et al. Acetylation of estrogen receptor a by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. *Mol Endocrinol*. 2006;20:1479–1493.
25. Bondy-Chorney E, Denoncourt A, Sai Y, et al. Nonhistone targets of KAT2A and KAT2B implicated in cancer biology. *Biochem Cell Biol*. 2019;97:30–45.
26. Koh SS, Chen D, Lee YH, et al. Synergistic enhancement of nuclear receptor function by p160 coactivators and two coactivators with protein methyltransferase activities. *J Biol Chem*. 2001;276:1089–1098.
27. King HA, Trotter KW, Archer TK. Chromatin remodeling during glucocorticoid receptor regulated transactivation. *Biochim Biophys Acta*. 2012;1819:716–726.
28. Mottis A, Mouchiroud L, Auwerx J. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev*. 2013;27:819–835.
29. Perissi V, Jepsen K, Glass CK, et al. Deconstructing repression: evolving models of co-repressor action. *Nat Rev Genet*. 2010;11:109–123.
30. Liu XF, Bagchi MK. Recruitment of distinct chromatin-modifying complexes by tamoxifen-complexed estrogen receptor at natural target gene promoters in vivo. *J Biol Chem*. 2004;279:15050–15058.
31. Leonhardt SA, Edwards DP. Mechanism of action of progesterone antagonists. *Exp Biol Med*. 2002;227:969–980.
32. Allan GF, Tsai SY, Tsai MJ, et al. Ligand-dependent conformational changes in the progesterone receptor are necessary for events that follow DNA binding. *Proc Natl Acad Sci U S A*. 1992;89:11750–11754.
33. Zhang X, Jeyakumar M, Petukhov S, et al. A nuclear receptor corepressor modulates transcriptional activity of antagonist-occupied steroid hormone receptor. *Mol Endocrinol*. 1998;12:513–524.
34. Chinenov Y, Gupte R, Dobrovolna J, et al. Role of transcriptional coregulator GRIP1 in the anti-inflammatory actions of glucocorticoids. *Proc Natl Acad Sci U S A*. 2012;109:11776–11781.
35. Gupte R, Muse GW, Chinenov Y, et al. Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. *Proc Natl Acad Sci U S A*. 2013;110:14616–14621.
36. Chen CC, Montalbano AP, Hussain I, et al. The transcriptional repressor GATAD2B mediates progesterone receptor suppression of myometrial contractile gene expression. *J Biol Chem*. 2017;292:12560–12576.
37. Hardy DB, Janowski BA, Corey DR, et al. Progesterone receptor (PR) plays a major anti-inflammatory role in human myometrial cells by antagonism of NF-kB activation of cyclooxygenase 2 (COX-2) expression. *Mol Endocrinol*. 2006;20:2724–2733.
38. Dong X, Yu C, Shynlova O, et al. p54nrb is a transcriptional corepressor of the progesterone receptor that modulates transcription of the labor-associated gene, connexin 43 (Gja1). *Mol Endocrinol*. 2009;23:1147–1160.
39. Brackertz M, Gong Z, Leers J, et al. p66a and p66b of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res*. 2006;34:397–406.

40. Greene GL, Gilna P, Waterfield M, et al. Sequence and expression of human estrogen receptor complementary DNA. *Science*. 1986;231:1150–1154.
41. Kuiper GG, Enmark E, Pelto-Huikko M, et al. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A*. 1996;93:5925–5930.
42. Means GD, Mahendroo MS, Corbin CJ, et al. Structural analysis of the gene encoding human aromatase cytochrome P-450, the enzyme responsible for estrogen biosynthesis. *J Biol Chem*. 1989;264:19385–19391.
43. Simpson ER, Zhao Y, Agarwal VR, et al. Aromatase expression in health and disease. *Recent Prog Horm Res*. 1997;52:185–213.
44. Kamat A, Hinshelwood MM, Murry BA, et al. Mechanisms in tissue-specific regulation of estrogen biosynthesis in humans. *Trends Endocrinol Metab*. 2002;13:122–128.
45. Valéra MC, Noirrit-Esclassan E, Dupuis M, et al. Effect of estetrol, a selective nuclear estrogen receptor modulator, in mouse models of arterial and venous thrombosis. *Mol Cell Endocrinol*. 2018;477:132–139.
46. Escande A, Pillon A, Servant N, et al. Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor a or b. *Biochem Pharmacol*. 2006;71:1459–1469.
47. McDonnell DP, Wardell SE. The molecular mechanisms underlying the pharmacological actions of ER modulators: implications for new drug discovery in breast cancer. *Curr Opin Pharmacol*. 2010;10:620–628.
48. Katzenellenbogen JA, Mayne CG, Katzenellenbogen BS, et al. Structural underpinnings of oestrogen receptor mutations in endocrine therapy resistance. *Nat Rev Cancer*. 2018;18:377–388.
49. Yi P, Wang Z, Feng Q, et al. Structure of a biologically active estrogen receptor-coactivator complex on DNA. *Mol Cell*. 2015;57:1047–1058.
50. Dutertre M, Smith CL. Ligand-independent interactions of p160/steroid receptor coactivators and CREB-binding protein (CBP) with estrogen receptor-alpha: regulation by phosphorylation sites in the A/B region depends on other receptor domains. *Mol Endocrinol*. 2003;17:1296–1314.
51. Chen D, Ma H, Hong H, et al. Regulation of transcription by a protein methyltransferase. *Science*. 1999;284:2174–2177.
52. Chen D, Huang SM, Stallcup MR. Synergistic, p160 coactivator-dependent enhancement of estrogen receptor function by CARM1 and p300. *J Biol Chem*. 2000;275:40810–40816.
53. Yi P, Wang Z, Feng Q, et al. Structural and functional impacts of ER coactivator sequential recruitment. *Mol Cell*. 2017;67:733–743.e734.
54. Carroll JS, Liu XS, Brodsky AS, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*. 2005;122:33–43.
55. Safe S, Kim K. Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. *J Mol Endocrinol*. 2008;41:263–275.
56. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. *Adv Protein Chem Struct Biol*. 2019;116:135–170.
57. Saville B, Wormke M, Wang F, et al. Ligand-, cell-, and estrogen receptor subtype (a/b)-dependent activation at GC-rich (Sp1) promoter elements. *J Biol Chem*. 2000;275:5379–5387.
58. DeNardo DG, Kim HT, Hilsenbeck S, et al. Global gene expression analysis of estrogen receptor transcription factor cross talk in breast cancer: identification of estrogen-induced/activator protein-1-dependent genes. *Mol Endocrinol*. 2005;19:362–378.
59. Paech K, Webb P, Kuiper GG, et al. Differential ligand activation of estrogen receptors ERa and ERb at AP1 sites. *Science*. 1997;277:1508–1510.
60. Bennesch MA, Picard D. Minireview: tipping the balance: ligand-independent activation of steroid receptors. *Mol Endocrinol*. 2015;29:349–363.
61. Chambliss KL, Yuhanna IS, Mineo C, et al. Estrogen receptor a and endothelial nitric oxide synthase are organized into a functional signaling module in caveolae. *Circ Res*. 2000;87:E44–E52.
62. Boonyaratanakornkit V, Edwards DP. Receptor mechanisms mediating non-genomic actions of sex steroids. *Semin Reprod Med*. 2007;25:139–153.
63. Filardo E, Quinn J, Pang Y, et al. Activation of the novel estrogen receptor G protein-coupled receptor 30 (GPR30) at the plasma membrane. *Endocrinology*. 2007;148:3236–3245.
64. Prossnitz ER, Barton M. Estrogen biology: new insights into GPER function and clinical opportunities. *Mol Cell Endocrinol*. 2014;389:71–83.
65. Hamilton KJ, Hewitt SC, Arao Y, et al. Estrogen hormone biology. *Curr Top Dev Biol*. 2017;125:109–146.
66. Lubahn DB, Moyer JS, Golding TS, et al. Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. *Proc Natl Acad Sci U S A*. 1993;90:11162–11166.
67. Dupont S, Krust A, Gansmuller A, et al. Effect of single and compound knockouts of estrogen receptors alpha (ERa) and beta (ERb) on mouse reproductive phenotypes. *Development*. 2000;127:4277–4291.
68. Bocchinfuso WP, Korach KS. Mammary gland development and tumorigenesis in estrogen receptor knockout mice. *J Mammary Gland Biol Neoplasia*. 1997;2:323–334.
69. Couse JF, Bunch DO, Lindzey J, et al. Prevention of the polycystic ovarian phenotype and characterization of ovulatory capacity in the estrogen receptor-a knockout mouse. *Endocrinology*. 1999;140:5855–5865.
70. Eddy EM, Washburn TF, Bunch DO, et al. Targeted disruption of the estrogen receptor gene in male mice causes alteration of spermatogenesis and infertility. *Endocrinology*. 1996;137:4796–4805.
71. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med*. 1994;331:1056–1061.
72. Bernard V, Kherra S, Francou B, et al. Familial multiplicity of estrogen insensitivity associated with a loss-of-function ESR1 mutation. *J Clin Endocrinol Metab*. 2017;102:93–99.
73. Hewitt SC, Korach KS. Estrogen receptors: structure, mechanisms and function. *Rev Endocr Metab Disord*. 2002;3:193–200.
74. Singh SP, Wolfe A, Ng Y, et al. Impaired estrogen feedback and infertility in female mice with pituitary-specific deletion of estrogen receptor a (ESR1). *Biol Reprod*. 2009;81:488–496.
75. Feng Y, Manka D, Wagner KU, et al. Estrogen receptor-a expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proc Natl Acad Sci U S A*. 2007;104:14718–14723.
76. Rooney AM, van der Meulen MCH. Mouse models to evaluate the role of estrogen receptor a in skeletal maintenance and adaptation. *Ann N Y Acad Sci*. 2017;1410:85–92.
77. Qiu S, V JT, Boulger E, et al. Hepatic estrogen receptor a is critical for regulation of gluconeogenesis and lipid metabolism in males. *Sci Rep*. 2017;7:1661.
78. Yoshinaga K. Uterine receptivity for blastocyst implantation. *Ann N Y Acad Sci*. 1988;541:424–431.
79. Wang H, Dey SK. Roadmap to embryo implantation: clues from mouse models. *Nat Rev Genet*. 2006;7:185–199.
80. Stewart CL, Kaspar P, Brunet LJ, et al. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature*. 1992;359:76–79.
81. Kurita T, Young P, Brody JR, et al. Stromal progesterone receptors mediate the inhibitory effects of progesterone on estrogen-induced uterine epithelial cell deoxyribonucleic acid synthesis. *Endocrinology*. 1998;139:4708–4713.
82. Winuthayanon W, Hewitt SC, Orvis GD, et al. Uterine epithelial estrogen receptor a is dispensable for proliferation but essential for complete biological and biochemical responses. *Proc Natl Acad Sci U S A*. 2010;107:19272–19277.
83. Klotz DM, Hewitt SC, Ciana P, et al. Requirement of estrogen receptor-a in insulin-like growth factor-1 (IGF-1)-induced uterine responses and in vivo evidence for IGF-1/estrogen receptor cross-talk. *J Biol Chem*. 2002;277:8531–8537.
84. Li Q, Kannan A, DeMayo FJ, et al. The antiproliferative action of progesterone in uterine epithelium is mediated by Hand2. *Science*. 2011;331:912–916.
85. Pawar S, Laws MJ, Bagchi IC, et al. Uterine epithelial estrogen receptor-a controls decidualization via a paracrine mechanism. *Mol Endocrinol*. 2015;29:1362–1374.
86. Herrera GGB, Lierz SL, Harris EA, et al. Oviductal retention of embryos in female mice Lacking estrogen receptor a in the isthmus and the uterus. *Endocrinology*. 2020:161.
87. Winuthayanon W, Bernhardt ML, Padilla-Banks E, et al. Oviductal estrogen receptor a signaling prevents protease-mediated embryo death. *Elife*. 2015;4:e10453.

88. Li S, Garcia M, Gewiss RL, et al. Crucial role of estrogen for the mammalian female in regulating semen coagulation and liquefaction in vivo. *PLoS Genet*. 2017;13:e1006743.
89. Billon-Gales A, Fontaine C, Filipe C, et al. The transactivating function 1 of estrogen receptor a is dispensable for the vasculoprotective actions of 17b-estradiol. *Proc Natl Acad Sci U S A*. 2009;106:2053–2058.
90. Abot A, Fontaine C, Raymond-Letron I, et al. The AF-1 activation function of estrogen receptor a is necessary and sufficient for uterine epithelial cell proliferation in vivo. *Endocrinology*. 2013;154:2222–2233.
91. Arao Y, Hamilton KJ, Ray MK, et al. Estrogen receptor alpha AF-2 mutation results in antagonist reversal and reveals tissue selective function of estrogen receptor modulators. *Proc Natl Acad Sci U S A*. 2011;108:14986–14991.
92. Arao Y, Hamilton KJ, Goulding EH, et al. Transactivating function (AF) 2-mediated AF-1 activity of estrogen receptor a is crucial to maintain male reproductive tract function. *Proc Natl Acad Sci U S A*. 2012;109:21140–21145.
93. Sinkevicius KW, Burdette JE, Woloszyn K, et al. An estrogen receptor-a knock-in mutation provides evidence of ligand-independent signaling and allows modulation of ligand-induced pathways in vivo. *Endocrinology*. 2008;149:2970–2979.
94. O'Brien JE, Peterson TJ, Tong MH, et al. Estrogen-induced proliferation of uterine epithelial cells is independent of estrogen receptor a binding to classical estrogen response elements. *J Biol Chem*. 2006;281:26683–26692.
95. Hewitt SC, Li L, Grimm SA, et al. Novel DNA motif binding activity observed in vivo with an estrogen receptor a mutant mouse. *Mol Endocrinol*. 2014;28:899–911.
96. Ahlbory-Dieker DL, Stride BD, Leder G, et al. DNA binding by estrogen receptor-a is essential for the transcriptional response to estrogen in the liver and the uterus. *Mol Endocrinol*. 2009;23:1544–1555.
97. Krege JH, Hodgin JB, Couse JF, et al. Generation and reproductive phenotypes of mice lacking estrogen receptor b. *Proc Natl Acad Sci U S A*. 1998;95:15677–15682.
98. Antal MC, Krust A, Chambon P, et al. Sterility and absence of histopathological defects in nonreproductive organs of a mouse ERb-null mutant. *Proc Natl Acad Sci U S A*. 2008;105:2433–2438.
99. Scarpin KM, Graham JD, Mote PA, et al. Progesterone action in human tissues: regulation by progesterone receptor (PR) isoform expression, nuclear positioning and coregulator expression. *Nucl Recept Signal*. 2009;7:e009.
100. Kastner P, Bocquel MT, Turcotte B, et al. Transient expression of human and chicken progesterone receptors does not support alternative translational initiation from a single mRNA as the mechanism generating two receptor isoforms. *J Biol Chem*. 1990;265:12163–12167.
101. Conneely OM, Maxwell BL, Toft DO, et al. The A and B forms of the chicken progesterone receptor arise by alternate initiation of translation of a unique mRNA. *Biochem Biophys Res Commun*. 1987;149:493–501.
102. Giangrande PH, Kimbrel EA, Edwards DP, et al. The opposing transcriptional activities of the two isoforms of the human progesterone receptor are due to differential cofactor binding. *Mol Cell Biol*. 2000;20:3102–3115.
103. Vegeto E, Shahbaz MM, Wen DX, et al. Human progesterone receptor A form is a cell- and promoter-specific repressor of human progesterone receptor B function. *Mol Endocrinol*. 1993;7:1244–1255.
104. Wei LL, Norris BM, Baker CJ. An N-terminally truncated third progesterone receptor protein, PR(C), forms heterodimers with PR(B) but interferes in PR(B)-DNA binding. *J Steroid Biochem Mol Biol*. 1997;62:287–297.
105. Wei LL, Gonzalez-Aller C, Wood WM, et al. 5'-Heterogeneity in human progesterone receptor transcripts predicts a new amino-terminal truncated "C"-receptor and unique A-receptor messages. *Mol Endocrinol*. 1990;4:1833–1840.
106. Price TM, Dai Q. The role of a mitochondrial progesterone receptor (PR-M) in progesterone action. *Semin Reprod Med*. 2015;33:185–194.
107. Behera MA, Dai Q, Garde R, et al. Progesterone stimulates mitochondrial activity with subsequent inhibition of apoptosis in MCF-10A benign breast epithelial cells. *Am J Physiol Endocrinol Metab*. 2009;297:E1089–E1096.
108. Karteris E, Zervou S, Pang Y, et al. Progesterone signaling in human myometrium through two novel membrane G protein-coupled receptors: potential role in functional progesterone withdrawal at term. *Mol Endocrinol*. 2006;20:1519–1534.
109. Thomas P, Pang Y. Anti-apoptotic actions of allopregnanolone and ganaxolone mediated through membrane progesterone receptors (PAQRs) in neuronal cells. *Front Endocrinol*. 2020;11:417.
110. Lydon JP, DeMayo FJ, Funk CR, et al. Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev*. 1995;9:2266–2278.
111. Chappell PE, Lydon JP, Conneely OM, et al. Endocrine defects in mice carrying a null mutation for the progesterone receptor gene. *Endocrinology*. 1997;138:4147–4152.
112. Schneider JS, Burgess C, Sleiter NC, et al. Enhanced sexual behaviors and androgen receptor immunoreactivity in the male progesterone receptor knockout mouse. *Endocrinology*. 2005;146:4340–4348.
113. Tibbetts TA, Conneely OM, O'Malley BW. Progesterone via its receptor antagonizes the pro-inflammatory activity of estrogen in the mouse uterus. *Biol Reprod*. 1999;60:1158–1165.
114. Mulac-Jericevic B, Lydon JP, DeMayo FJ, et al. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc Natl Acad Sci U S A*. 2003;100:9744–9749.
115. Mulac-Jericevic B, Mullinax RA, DeMayo FJ, et al. Subgroup of reproductive functions of progesterone mediated by progesterone receptor-B isoform. *Science*. 2000;289:1751–1754.
116. Grimm SL, Ward RD, Obr AE, et al. A role for site-specific phosphorylation of mouse progesterone receptor at serine 191 in vivo. *Mol Endocrinol*. 2014;28:2025–2037.
117. Tan J, Paria BC, Dey SK, et al. Differential uterine expression of estrogen and progesterone receptors correlates with uterine preparation for implantation and decidualization in the mouse. *Endocrinology*. 1999;140:5310–5321.
118. Lessey BA, Yeh I, Castelbaum AJ, et al. Endometrial progesterone receptors and markers of uterine receptivity in the window of implantation. *Fertil Steril*. 1996;65:477–483.
119. Franco HL, Rubel CA, Large MJ, et al. Epithelial progesterone receptor exhibits pleiotropic roles in uterine development and function. *FASEB J*. 2012;26:1218–1227.
120. Takamoto N, Zhao B, Tsai SY, et al. Identification of Indian hedgehog as a progesterone-responsive gene in the murine uterus. *Mol Endocrinol*. 2002;16:2338–2348.
121. Lee K, Jeong J, Kwak I, et al. Indian hedgehog is a major mediator of progesterone signaling in the mouse uterus. *Nat Genet*. 2006;38:1204–1209.
122. Kurihara I, Lee DK, Petit FG, et al. COUP-TFII mediates progesterone regulation of uterine implantation by controlling ER activity. *PLoS Genet*. 2007;3:e102.
123. Franco HL, Lee KY, Broaddus RR, et al. Ablation of Indian hedgehog in the murine uterus results in decreased cell cycle progression, aberrant epidermal growth factor signaling, and increased estrogen signaling. *Biol Reprod*. 2010;82:783–790.
124. Large MJ, Wetendorf M, Lanz RB, et al. The epidermal growth factor receptor critically regulates endometrial function during early pregnancy. *PLoS Genet*. 2014;10:e1004451.
125. Franco HL, Dai D, Lee KY, et al. WNT4 is a key regulator of normal postnatal uterine development and progesterone signaling during embryo implantation and decidualization in the mouse. *FASEB J*. 2011;25:1176–1187.
126. Wetendorf M, Wu SP, Wang X, et al. Decreased epithelial progesterone receptor A at the window of receptivity is required for preparation of the endometrium for embryo attachment. *Biol Reprod*. 2017;96:313–326.
127. Vasquez YM, Wang X, Wetendorf M, et al. FOXO1 regulates uterine epithelial integrity and progesterone receptor expression critical for embryo implantation. *PLoS Genet*. 2018;14:e1007787.
128. Rubel CA, Lanz RB, Kommagani R, et al. Research resource: genome-wide profiling of progesterone receptor binding in the mouse uterus. *Mol Endocrinol*. 2012;26:1428–1442.
129. Lee ME, Temizer DH, Clifford JA, et al. Cloning of the GATA-binding protein that regulates endothelin-1 gene expression in endothelial cells. *J Biol Chem*. 1991;266:16188–16192.

130. Schepers GE, Teasdale RD, Koopman P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell*. 2002;3:167–170.
131. Rubel CA, Wu SP, Lin L, et al. A gata2-dependent transcription network regulates uterine progesterone responsiveness and endometrial function. *Cell Rep*. 2016;17:1414–1425.
132. Wang X, Li X, Wang T, et al. SOX17 regulates uterine epithelial-stromal cross-talk acting via a distal enhancer upstream of Ihh. *Nat Commun*. 2018;9:4421.
133. Strauss 3rd JF, Martinez F, Kiriakidou M. Placental steroid hormone synthesis: unique features and unanswered questions. *Biol Reprod*. 1996;54:303–311.
134. Siiteri PK, Febres F, Clemens LE, et al. Progesterone and maintenance of pregnancy: is progesterone nature's immunosuppressant? *Ann N Y Acad Sci*. 1977;286:384–397.
135. Siiteri PK, Stites DP. Immunologic and endocrine interrelationships in pregnancy. *Biol Reprod*. 1982;26:1–14.
136. Dong X, Shylnova O, Challis JR, et al. Identification and characterization of the protein-associated splicing factor as a negative co-regulator of the progesterone receptor. *J Biol Chem*. 2005;280:13329–13340.
137. Chen CC, Hardy DB, Mendelson CR. Progesterone receptor inhibits proliferation of human breast cancer cells via induction of MAPK phosphatase 1 (MKP-1/DUSP1). *J Biol Chem*. 2011;286:43091–43102.
138. Vicent GP, Ballare C, Nacht AS, et al. Induction of progesterone target genes requires activation of Erk and Msk kinases and phosphorylation of histone H3. *Mol Cell*. 2006;24:367–381.
139. Vanden Berghe W, Plaisance S, Boone E, et al. p38 and extracellular signal-regulated kinase mitogen-activated protein kinase pathways are required for nuclear factor-kB p65 transactivation mediated by tumor necrosis factor. *J Biol Chem*. 1998;273:3285–3290.
140. Renthal NE, Chen CC, Williams KC, et al. miR-200 family and targets, ZEB1 and ZEB2, modulate uterine quiescence and contractility during pregnancy and labor. *Proc Natl Acad Sci U S A*. 2010;107:20828–20833.
141. Renthal NE, Williams KC, Mendelson CR. MicroRNAs-mediators of myometrial contractility during pregnancy and labour. *Nat Rev Endocrinol*. 2013;9:391–401.
142. Spoelstra NS, Manning NG, Higashi Y, et al. The transcription factor ZEB1 is aberrantly expressed in aggressive uterine cancers. *Cancer Res*. 2006;66:3893–3902.
143. Williams KC, Renthal NE, Condon JC, et al. MicroRNA-200a serves a key role in the decline of progesterone receptor function leading to term and preterm labor. *Proc Natl Acad Sci U S A*. 2012;109:7529–7534.
144. Richer JK, Lange CA, Manning NG, et al. Convergence of progesterone with growth factor and cytokine signaling in breast cancer. Progesterone receptors regulate signal transducers and activators of transcription expression and activity. *J Biol Chem*. 1998;273:31317–31326.
145. Piekorz RP, Gingras S, Hoffmeyer A, et al. Regulation of progesterone levels during pregnancy and parturition by signal transducer and activator of transcription 5 and 20a-hydroxysteroid dehydrogenase. *Mol Endocrinol*. 2005;19:431–440.
146. Challis JRG, Matthews SG, Gibb W, et al. Endocrine and paracrine regulation of birth at term and preterm. *Endocr Rev*. 2000;21:514–550.
147. Condon JC, Jeyasuria P, Faust JM, et al. A decline in progesterone receptor coactivators in the pregnant uterus at term may antagonize progesterone receptor function and contribute to the initiation of labor. *Proc Natl Acad Sci U S A*. 2003;100:9518–9523.
148. Pieber D, Allport VC, Hills F, et al. Interactions between progesterone receptor isoforms in myometrial cells in human labour. *Mol Human Reprod*. 2001;7:875–879.
149. Merlino AA, Welsh TN, Tan H, et al. Nuclear progesterone receptors in the human pregnancy myometrium: evidence that parturition involves functional progesterone withdrawal mediated by increased expression of progesterone receptor-A. *J Clin Endocrinol Metab*. 2007;92:1927–1933.
150. Mesiano S, Chan EC, Fitter JT, et al. Progesterone withdrawal and estrogen activation in human parturition are coordinated by progesterone receptor A expression in the myometrium. *J Clin Endocrinol Metab*. 2002;87:2924–2930.
151. Condon JC, Hardy DB, Kovaric K, et al. Upregulation of the progesterone receptor (PR)-C isoform in laboring myometrium by activation of NF-kB may contribute to the onset of labor through inhibition of PR function. *Mol Endocrinol*. 2006;20:764–775.
152. Condon JC, Jeyasuria P, Faust JM, et al. Surfactant protein secreted by the maturing mouse fetal lung acts as a hormone that signals the initiation of parturition. *Proc Natl Acad Sci U S A*. 2004;101:4978–4983.
153. Runnebaum B, Zander J. Progesterone and 20a-dihydroprogesterone in human myometrium during pregnancy. *Acta Endocrinol Suppl (Copenh)*. 1971;150:3–45.
154. Mendelson CR, Gao L, Montalbano AP. Multifactorial regulation of myometrial contractility during pregnancy and parturition. *Front Endocrinol (Lausanne)*. 2019;10:714.
155. Challis JR. Sharp increase in free circulating oestrogens immediately before parturition in sheep. *Nature*. 1971;229:208.
156. Buster JE, Chang RJ, Preston DL, et al. Interrelationships of circulating maternal steroid concentrations in third trimester pregnancies. II. C_{18} and C_{19} steroids: estradiol, estriol, dehydroepiandrosterone, dehydroepiandrosterone sulfate, D^5-androstenediol, D^4-androstenedione, testosterone, and dihydrotestosterone. *J Clin Endocrinol Metab*. 1979;48:139–142.
157. Wu WX, Myers DA, Nathanielsz PW. Changes in estrogen receptor messenger ribonucleic acid in sheep fetal and maternal tissues during late gestation and labor. *Am J Obstet Gynecol*. 1995;172:844–850.
158. Gao L, Wang G, Liu W, et al. Reciprocal feedback between miR-181a and E_2/ERa in myometrium enhances inflammation leading to labor. *J Clin Endocrinol Metab*. 2016;101:3646–3656.
159. Murata T, Narita K, Honda K, et al. Differential regulation of estrogen receptor a and b mRNAs in the rat uterus during pregnancy and labor: possible involvement of estrogen receptors in oxytocin receptor regulation. *Endocr J*. 2003;50:579–587.
160. Piersanti M, Lye SJ. Increase in messenger ribonucleic acid encoding the myometrial gap junction protein, connexin-43, requires protein synthesis and is associated with increased expression of the activator protein-1, c-fos. *Endocrinology*. 1995;136:3571–3578.
161. Tsuboi K, Iwane A, Nakazawa S, et al. Role of prostaglandin H2 synthase 2 in murine parturition: study on ovariectomy-induced parturition in prostaglandin F receptor-deficient mice. *Biol Reprod*. 2003;69:195–201.
162. Yi P, Wang Z, Feng Q, et al. 2017, Structural and functional impacts of ER coactivator sequential recruitment. *Mol Cell*. 2017;67:733–743.

121

Contraception

Sharon Cameron and Sarah Hardman

OUTLINE

Contraceptive Use, 2021
Methods and Their Effectiveness, 2021
Drug Interactions, 2022
Medical Eligibility Criteria, 2023
  High-Effectiveness Methods, 2023
  Intrauterine Devices, 2023
  Progestogen-Only Implant, 2024
Medium Effectiveness Methods, 2024
  Progestogen-Only Injectable, 2024
  Combined Hormonal Contraception, 2024
  CHC Pill, 2025
  CHC Patch and Vaginal Ring, 2025
  Progestogen-Only Pill, 2026
Low Effectiveness Methods, 2026
  Condoms, 2026
  Diaphragm and Cap, 2026
  Fertility Awareness-Based Methods, 2026
  Lactational Amenorrhea, 2026
  Sterilization, 2026
  Emergency Contraception, 2026
  Non-Contraceptive Benefits of Contraception, 2027
Summary and Future Directions, 2027

CONTRACEPTIVE USE

Correct and consistent use of effective methods of contraception can prevent *most* unintended pregnancies. It is estimated that in low- and middle-income countries, approximately 218 million women want to avoid a pregnancy but are not using effective methods. This is mainly due to lack of supplies, political and cultural barriers, and limited and/or poor quality services.

Abortion is a common consequence of unintended pregnancy. While abortion is a very safe procedure when conducted in a legal setting, in countries where abortion is illegal or restricted, it is often performed under unsafe conditions and can result in women dying or suffering serious injuries. Globally, it is estimated that there are over 25 million unsafe abortions performed each year, that between 30 and 200 women die per 100,000 of these unsafe procedures, and that 7 million women require hospital treatment for complications of unsafe abortion.[1,2] These are preventable.

Unintended pregnancy can also lead to delayed or no antenatal care, with resulting maternal and neonatal health risks. In addition, if conception occurs within a year of giving birth, the newborn is at higher risk of preterm birth and neonatal death.[3]

Contraception allows women to choose when to have a baby and to space births. It can enable them to complete education and play an equal role in society. Women's lives may also be improved by non-contraceptive benefits, such as reduced menstrual blood loss, or through reduction in the risk of reproductive tract cancers (see below).

KEY POINTS

- Contraception saves the lives of millions of women.
- Contraception improves interpregnancy intervals and outcomes for babies.

METHODS AND THEIR EFFECTIVENESS

Contraceptive methods used by women (Table 121.1) include combined hormonal (estrogen/progestogen) contraceptive pills, patch, or vaginal ring, progestogen-only pills, implants, injections, hormone-releasing intrauterine devices (IUDs), copper intrauterine devices (Cu-IUDs) and barrier methods (condoms, diaphragms, and cervical caps), and sterilization. Hormonal contraceptives inhibit ovulation by suppressing luteinizing hormone (LH), thicken cervical mucus (inhibiting sperm penetration), and may affect development of the endometrium (Fig. 121.1). None of these methods is 100% effective at preventing pregnancy, and the effectiveness of a given contraceptive method depends on both mechanism of action (see Fig. 121.1) and correct, consistent, continued use. Uptake, compliance, and continuation depend on acceptability of the method to the user. Acceptability of a method may be influenced by age, fertility intentions, perceptions of effectiveness and safety, familiarity and experience of friends and/or family, ease of use, side effects, and non–contraceptive benefits.[4]

The effectiveness of a method of contraception is expressed by the failure rates associated with its use. Contraceptive failure rates during perfect use (defined as use according to instructions) demonstrate how effective methods can be. Failure rates during typical use show how effective methods are during actual use (including inconsistent or incorrect use).[5] Estimated perfect– and typical–use failure rates of the different methods (based on percentage of users experiencing an unintended pregnancy during the first year of use) are shown in Table 121.1. Long-acting reversible contraceptive methods such as the subdermal implant and intrauterine contraceptives which do not require any action by the user, once inserted, have typical-use failure rates within the first 12 months of use that are similar to those of perfect-use failure rates. In contrast, typical-use failure rates for methods that require compliance by the user (e.g., daily pill to be taken) are much higher than perfect-use failure rates.

TABLE 121.1 Contraceptive Methods and Percentage of Women Experiencing an Unintended Pregnancy Within the First Year of Use With Typical Use and Perfect Use

Method	Typical Use	Perfect Use
None	85	85
Fertility awareness-based methods	15	
Male condom	13	2
Female diaphragm	17	16
Combined pill and progestogen-only pill	7	0.3
Combined hormonal contraceptive patch and vaginal ring	7	0.3
Progestogen-only injectable	4	0.2
Cu-IUD	0.8	0.6
52-mg LNG-IUD	0.1	0.1
Progestogen-only implant	0.1	0.1
Female sterilization	0.5	0.5

From Trussell J, Aiken ARA. Contraceptive efficacy. In: Hatcher RA, Nelson AL, Trussell J, et al., eds. *Contraceptive Technology*. 21st ed. New York: Ayer Company Publishers, Inc., 2018.

Cu-IUD, Copper intrauterine device; *LNG-IUD*, levonorgestrel intrauterine device.

Discontinuation rates are higher for methods that do not require removal by a healthcare professional. Data indicate that approximately 50% of women discontinue a pill or injectable within the first year of use, whereas approximately 80% of women continue the implant or IUD at 12 months.[6] Reasons for discontinuation are often associated with perceived risks and real or perceived side effects.

KEY POINTS

- Effectiveness of a contraceptive method depends on its mode of action and the need for user compliance.

DRUG INTERACTIONS

Certain medications may decrease or increase serum levels of contraceptive hormones with resulting potential for contraceptive failure or adverse effects; likewise, hormonal contraception may alter exposure to other medications. Prescribers should ask women about all drug use and where necessary consult up–to–date and reputable online drug interaction information.

Some of the most common interactions affecting hormonal contraceptives are with medications such as carbamazepine and rifampicin that induce hepatic cytochrome P450 enzymes. These drugs could reduce the effectiveness of combined hormonal contraception (CHC) and the progestogen–only pill and implant, by their metabolism in the liver via

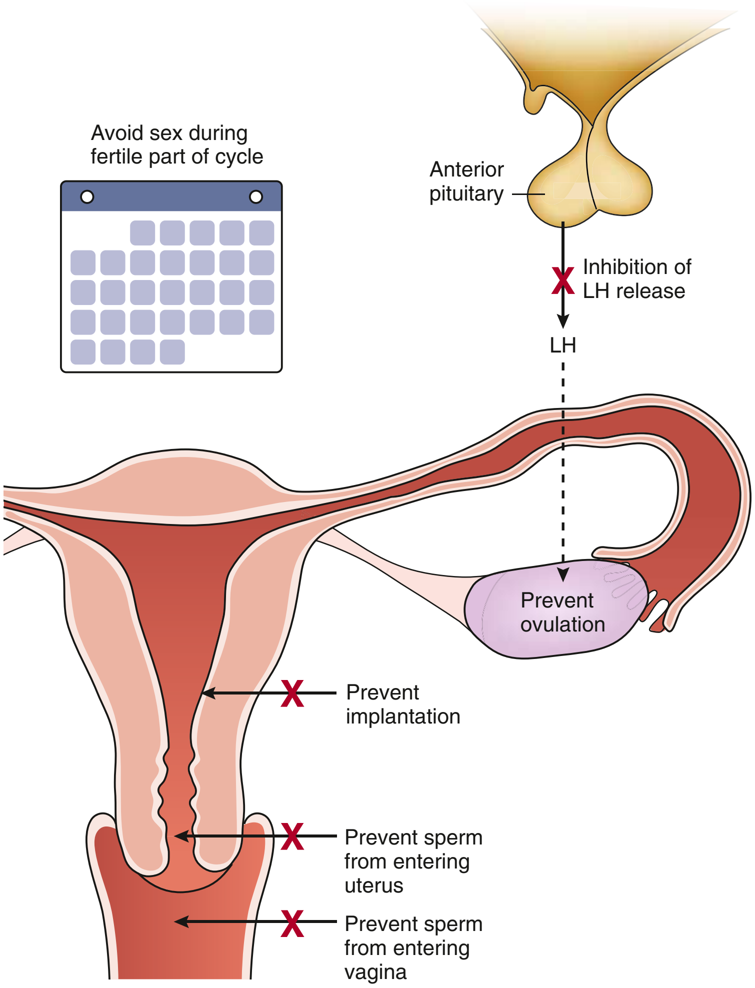


Figure 121.1 Mechanism of action of contraceptives.

this pathway. Enzyme inducers are not considered to affect the contraceptive effectiveness of injectable and intrauterine contraceptives.[7]

Conversely, as an example, use of CHC increases the clearance of the antiepileptic lamotrigine and reduces serum levels of this drug. Seizure frequency could therefore increase when initiating CHC, and lamotrigine side effects may increase in the pill–free interval or when discontinuing this method of contraception.

Contraceptive effectiveness of the combined oral contraceptive pill (and all other methods) is not considered to be affected by coadministration of nonenzyme-inducing broad–spectrum antibiotics.[7]

MEDICAL ELIGIBILITY CRITERIA

Many women who use contraception are fit and healthy and can safely use any effective contraceptive method. However, for women with specific medical conditions, use of particular contraceptive methods could be associated with increased risk of adverse health events. In an attempt to produce a set of international norms for providing contraception to individuals with a range of medical conditions and characteristics that could contraindicate a contraceptive method, the World Health Organization (WHO) developed a system addressing medical eligibility criteria (MEC) for contraceptive use.[8] The WHO MEC comprise a set of evidence-based recommendations for global use that help guide safe contraceptive prescribing decisions. The United States and United Kingdom have adapted these for use in their settings.[9,10] For each of a number of characteristics or medical conditions, the MEC tables indicate a category 1, 2, 3, or 4 for each contraceptive method. The four categories are defined in Table 121.2. Category 1 indicates that there is no restriction to use of that method by a woman with the condition, while category 4 indicates that use of that method by a woman with the condition represents an unacceptable health risk.

The MEC category only reflects the *safety* of use of the method. It is important to note that MEC do not indicate the *best* or *most effective* contraceptive method for a woman; such an evaluation must also take into account the fact that effectiveness of the method may be affected by the condition or the medication required for the condition as well as the woman's preferences and requirements. In addition, MEC categories relate to use of contraceptive methods *for contraception*, but not for other indications. So, for example, if a woman is using a contraceptive method for a non-contraceptive indication (e.g., management of symptoms of polycystic ovarian syndrome), then the risk-benefit balance could differ and the MEC category may not apply.

KEY POINTS

- Effective contraceptive methods are generally very safe (they are extremely safe compared to pregnancy), but assessment of medical eligibility is recommended, particularly for women with preexisting medical conditions.

TABLE 121.2 Medical Eligibility Criteria (MEC) for Contraceptive Use[8]

MEC Categories for Contraceptive Eligibility	
1	A condition for which there is no restriction for the use of the contraceptive method
2	A condition where the advantages of using the method generally outweigh the theoretical or proven risks
3	A condition where the theoretical or proven risk usually outweighs the advantages of using the method
4	A condition which represents an unacceptable health risk if the contraceptive method is used

High-Effectiveness Methods

The most effective methods of contraception are the long-acting reversible methods of contraception (LARC) such as the Cu-IUD, levonorgestrel intrauterine device (LNG-IUD), and progestin-only implant. Their contraceptive failure rate in the first year of use is under 1% with both perfect and typical use (see Table 121.1). (Note that enzyme-inducing drugs can reduce contraceptive effectiveness of the implant.)

Unfortunately, myths and misconceptions amongst both women and health care professionals surrounding use of LARC are major factors that limit their uptake.[11] Education, dispelling myths, and promoting the significant non-contraceptive benefits of LARC methods could improve uptake and continuation as these LARC methods have the potential to prevent many more unintended pregnancies for more women.

Intrauterine Devices

Intrauterine methods of contraception include Cu-IUD and LNG-IUD.

Available Cu-IUDs vary in size, shape, copper content, and duration of use. Most consist of a plastic frame with copper wire wound around the stem; some also have copper on the arms of the device. Cu-IUDs are effective for between 3 and 10 years depending on the device used.

LNG-IUD consists of an elastomere frame with a reservoir on the stem containing LNG. There are devices containing 52 mg, 19.5 mg, and 13.5 mg of LNG. The 52-mg device provides at least 6 years of contraception, and the 19.5-mg device provides 5 years of contraception, and the 13.5-mg device is effective for 3 years.

Cu-IUD and LNG-IUD work primarily by inhibiting fertilization due to the toxic effect of copper ions on ova and sperm or adverse effects of LNG on sperm penetrability and transport.[12,13] Both copper and hormonal IUDs also prevent implantation; Cu-IUD causes a local inflammatory reaction in the endometrium and LNG induces endometrial atrophy.[12,13] In addition, LNG causes thickening of the cervical mucus which prevents sperm from entering the uterus, although it is not known how quickly such changes are established.[14] With LNG-IUD, serum levels of LNG are detectable but much lower than peak levels observed with other LNG-containing contraceptives, and most women (80%) will continue to ovulate.[14a] Levels are also influenced by body mass index (BMI).[15]

Healthcare professionals can be reluctant to offer an IUD to nulliparous women or adolescents due to misconceptions that it will be difficult to insert or carry higher risks. However, the evidence to support use in these groups is reassuring. The WHO, Faculty of Sexual and Reproductive Healthcare UK, and American College of Obstetricians and Gynecologists all support the use of IUDs in young women and nulliparous women.[16–19]

All LNG-IUDs reduce menstrual blood loss[20,21]; there is approximately 90% reduction by 1 year of use of the 52-mg LNG-IUD. Unscheduled bleeding is, however, also common. In contrast, users of Cu-IUD may expect increased menstrual loss. Concurrent use of nonsteroidal anti-inflammatory drugs or an antifibrinolytic can reduce menstrual blood loss.[22]

Risks associated with IUD use are mainly related to the insertion procedure.[19] Perforation of the uterus at insertion is rare (1 per 1000); factors associated with increased risk of perforation include breastfeeding and being less than 6 months postpartum.[23] Around 1 in 20 IUDs are expelled. Expulsion is most likely to happen in the first 3 months after insertion.[24] The overall risk of pelvic inflammatory disease following insertion of IUD is low (about 0.5%). It is more likely in the first 3 weeks after insertion and in women at risk of sexually transmitted infections (STIs).[25] A sexual history should be obtained from every woman before IUD insertion with testing for *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in high-risk groups and populations where prevalence of infection is high.[19]

Women may have concerns surrounding the effects of IUDs on their fertility. However, the available evidence indicates that IUDs do not cause a delay in return to fertility or increase the risk of infertility.[26]

IUD users should be informed that their overall risk of ectopic pregnancy is much reduced compared with women who are using no contraception. Although the absolute risk of ectopic pregnancy is low, if a pregnancy does occur with an IUD *in situ* then the "relative" risk of that pregnancy being ectopic is higher. In a multicountry European prospective study (European Active Surveillance Study), approximately 20% of all pregnancies occurring during IUD use were ectopic.[27]

Progestogen-Only Implant

A single rod subdermal implant containing the progestogen etonogestrel (68 mg) is the currently available implant in the United States and Europe (Nexplanon). Implants in use in other parts of the world include: Uniplant (single rod, containing nomegestrol, and effective for 1 year) and Jadelle/Sino implant II (two rods, containing LNG, and effective for 5 years).

The progestogen-only implant works principally by inhibition of ovulation and provides contraception for 3 years.[28,29] Implants also prevent sperm penetration by altering the cervical mucus and may prevent implantation by thinning the endometrium. Nexplanon is a non-biodegradable, flexible rod, similar in size to a matchstick (40 mm × 2 mm). It is inserted subdermally using a preloaded inserter. The insertion site, 8–10 cm above the medial epicondyle over the triceps muscle, is chosen to minimize the risk of damage to underlying neurovascular structures.[30] The insertion device is designed to facilitate correct insertion in the subdermal layer. Nevertheless, poor insertion technique can still result in deep insertion with consequent difficult removal or (occasionally) neurovascular damage or intravascular insertion. Nexplanon insertion should therefore be conducted only by clinicians with appropriate training.

The implant is not usually visible, but should be easily palpable. If it is not palpable, it should be located. Nexplanon contains a small quantity of barium, which permits it to be visualized by x-ray. It can also be localized using low-frequency ultrasound probes, which can help aid removal of implants that are not easily palpable. Once inserted, there is no need for any routine follow-up until the device is due for replacement or the user wishes it removed. After removal, the contraceptive effect ceases immediately after removal and serum etonogestrel levels are undetectable within 1 week.[30]

Bleeding patterns are variable and unpredictable, and dissatisfaction with bleeding pattern is the most common reason for implant removal requests. While many women bleed less often than once a month and some are amenorrhoeic during Nexplanon use, bleeding episodes are often irregular, unpredictable, and can also be frequent and sometimes prolonged.[31]

MEDIUM EFFECTIVENESS METHODS

Progestogen-Only Injectable

Both depot medroxyprogesterone acetate (DMPA) and norethisterone enanthate (NET-EN) are long-acting progestogens used as contraceptive injectables. Worldwide, the most commonly used injectable is DMPA, with NET-EN less commonly used. Progestogen-only injectables work primarily through inhibition of ovulation. There is also a progestogen effect on cervical mucus that makes it less penetrable to sperm and on the endometrium, creating an unfavorable environment for implantation.[32]

DMPA is available as both an intramuscular (IM) preparation (150 mg) and a micronized subcutaneous (SC) formulation (104 mg). The IM injection is administered to the buttock or upper arm, and the SC injection to the abdomen or anterior thigh. Since SC injections are easier to give, the SC preparation can be administered by a wide range of trained providers and is also licensed for self-administration, thus expanding access for women to this method.[33,34] Injection site reactions are common with the SC preparation; users should be warned that approximately 1 in 10 women develop lipoatrophy (nodule or dimpling of the skin) at the injection site.[32]

IM and SC preparations are equally effective for contraception (estimated typical use failure 4%), and have the same mode of action (inhibition of ovulation), frequency of administration (every 13 weeks), bleeding pattern, and side-effect profile.[32] The injectable causes profound ovarian suppression, resulting in high rates of amenorrhea (over 50% at 1 year) although a small proportion of users experience heavy, persistent bleeding. The injectable is the only hormonal contraceptive method that may delay return of fertility after discontinuation—in some cases, ovulation may not return for a year or more after the last injection. This makes it unsuitable for short-term contraception use. DMPA may cause weight gain in some women, particularly adolescents with obesity.[35] DMPA is associated with loss of bone mineral density (BMD) (5% loss of BMD at lumbar spine) in the first few years of use and may be best avoided by individuals with other risk factors for osteoporosis. However, BMD loss seems to plateau, has not been associated with osteoporotic fractures, and recovers following discontinuation.[36]

There was formerly concern because observational studies in countries with high prevalence of human immunodeficiency virus (HIV) suggested that DMPA users might be at higher risk of acquisition of HIV than users of other hormonal methods. However, a large, robust multisite randomized controlled trial in sub-Saharan Africa recently demonstrated that use of DMPA does not increase the risk of acquisition of HIV among high-risk women.[37] The WHO therefore recommended in 2019 that the injectable can be safely used by women living with HIV or at high risk of HIV infection.[38]

Combined Hormonal Contraception

CHC contains both an estrogen and a progestogen, delivered as a pill, patch, or vaginal ring. The primary mechanism of contraceptive action is to suppress ovarian activity and prevent ovulation.

Contraceptive effectiveness of CHC is highly dependent on the user. When used perfectly, the risk of unintended pregnancy in the first year of use of the pill is < 1%, but with typical use, it has been estimated to be about 7%. Severe vomiting or diarrhea, use of weight loss drugs that cause malabsorption, and bariatric surgery could all potentially reduce absorption (and thus effectiveness) of combined oral contraception. Drugs that induce the enzymes that metabolize estrogens and progestogens reduce contraceptive effectiveness of all combined contraceptive methods.

Current use of CHC is associated with increased risk of venous thromboembolism (VTE).[39] Although the absolute risk of VTE remains low during CHC use (estimated at 5–12 per 10,000 users per year), CHC should be avoided by at-risk individuals including those with thrombophilia or significant additional risk factors such as obesity and immobility. CHC users must be aware of increased thrombotic risk associated with immobility during travel and illness, and CHC should be stopped (and replaced with other effective contraception) several weeks ahead of major surgery.[40] Individuals with risk factors for cardiovascular disease such as hypertension, older age, smoking, and

TABLE 121.3 Examples of World Health Organization Medical Eligibility Criteria Category 4 Conditions for Use of Combined Hormonal Contraception[8]

- Smoking after age 35 years
- Blood pressure ≥ 160/100
- Migraine with aura
- Multiple risk factors for cardiovascular disease
- Cerebrovascular accident (current or past)
- Ischemic heart disease (current or past)
- Venous thrombosis (current or past)
- Known thrombogenic mutations
- Current breast cancer

migraine with aura should avoid the use of CHC because of the associated increased risk of thrombotic stroke and myocardial infarction.[41,42] Absolute risk of these arterial events in women of reproductive age is low even during use of CHC, with approximately two thrombotic strokes and one myocardial infarction per 10,000 CHC users per year. Current CHC use is associated with a very low increased risk of breast cancer and cervical cancer that reduces over 10 years after stopping CHC.[43] Use of CHC is contraindicated after a breast cancer diagnosis. The MEC reflect these potential CHC-associated health risks (Table 121.3).

Ever use of CHC is, however, associated with a significant reduction in risk of endometrial, ovarian, and colorectal cancer[44](Table 121.4). The risk reduction for endometrial and ovarian cancer increases with the duration of CHC use and may persist for 30 years after CHC cessation.

CHC offers additional non-contraceptive benefits. Withdrawal bleeding is generally regular, lighter than menstruation, and may be less painful. Thus, CHC may be a good option for individuals with heavy menstrual bleeding and dysmenorrhea. By elevating serum hormone-binding globulin, CHC reduces free testosterone levels and may improve acne and hirsutism related to androgen activity, including in polycystic ovarian syndrome.[45] CHC may be used to provide both contraception and hormone replacement in premature ovarian insufficiency, although menopausal symptoms often recur during any hormone-free interval.

CHC Pill

The most familiar and widely used combined contraceptive is the combined oral contraceptive pill. Most combined oral contraceptive pills contain a low dose (between 20 and 35 mcg) of the potent synthetic estrogen, ethinylestradiol, in combination with one of a range of progestogens. Most are monophasic preparations (all pills in the pack contain the same dose of estrogen and progestogen) and are designed to be taken for 21 consecutive days followed by a 7-day pill-free interval before restarting. The break in pill taking triggers a regular, 4-weekly withdrawal bleed. This bleeding pattern is acceptable to many women because they are familiar with the concept of monthly menstruation. During the pill-free interval, however, ovarian suppression is reduced, and there is a risk of escape ovulation if pill taking is not restarted on time.

Users of a monophasic preparation, designed to be taken as a regular regimen of 21 days of pills followed by 7 hormone-free days, may choose to shorten the pill-free interval to 3 or 4 days or omit it altogether. This might reduce the risk of escape ovulation and can reduce or avoid the withdrawal bleed[46] as well as withdrawal-associated symptoms such as pelvic pain and headache. Some pill preparations have a sequential change in ethinylestradiol and progestogen dose across the cycle. Such biphasic or triphasic formulations have not been demonstrated to have significant benefit over monophasic pills.[40]

The different progestogens in CHC may offer individual users different side-effect profiles. There is no one combination that is consistently associated with fewest side effects for all users. Risk of VTE is increased during use of all CHC. Absolute risk of a VTE still remains very low, but is lower for CHC pills containing norethisterone, LNG, or norgestimate than for those containing newer progestogens.[47]

With cyclical 21/7 pill use, CHC pills containing 30 or 35 mcg ethinylestradiol are generally associated with better cycle control than 20-mcg pills. Risk of venous and arterial thrombosis appears to increase with increasing ethinylestradiol dose.[47,48]

CHC pills containing 17β-estradiol and estradiol valerate rather than ethinylestradiol have been developed. These appear to have favorable metabolic profiles relative to ethinylestradiol-containing pills, but current evidence is too limited to confirm or exclude a superior cardiovascular risk profile.[49]

CHC Patch and Vaginal Ring

The combined transdermal patch delivers ethinylestradiol and norelgestromin. For 3 consecutive weeks, a patch is applied to the skin and worn for 7 days. After the third week of patch use, a 7-day patch-free interval is taken.[40]

The combined vaginal ring delivers ethinylestradiol and etonogestrel; a flexible ring is worn inside the vagina for 21 days, then removed by the user for a 7-day hormone-free interval before inserting a new ring. As with pill taking, patch and ring users may opt to use these preparations in a continuous regimen or shorten the hormone-free interval.[40]

TABLE 121.4 Risk of Cancer Among "Ever" and "Never" Users of Oral Contraceptives in the Royal College of General Practitioners' Oral Contraception Study[43]

	Cases Observed (total number, ever and never users)	Incidence Rate Ratio (99% CI), Ever Versus Never Use of OC	Risk Attributable to Ever OC Use (number per 100,000 woman years)
Breast cancer	2071	1.04 (0.91–1.17)	+4.78
Invasive cervical cancer	192	1.31 (0.84–2.04)	+3.89
Endometrial cancer	295	0.66 (0.48–0.89)	–10.14
Ovarian cancer	336	0.67 (0.50–0.89)	–11.17
Colorectal cancer	688	0.81 (0.66–0.99)	–11.31
Any cancer	7002	0.96 (0.90–1.03)	–23.65

46,022 subjects observed for up to 44 years. 884,895 woman-years of observation of ever users of OC. 388,505 woman-years of observation of never users of OC.
CI, Confidence interval; *OC*, oral contraception.

Progestogen-Only Pill

The progestogen-only pill (POP) is used less commonly than the combined oral contraceptive pill. Yet, it is as effective as CHC and is extremely safe, being devoid of the cardiovascular risks of estrogen-containing pills. There are very few contraindications to POP use. Its favorable safety profile was important in the decision to make an interim "bridging" supply possible from pharmacies in the United Kingdom along with emergency contraception (EC).[50] The POP containing 75 mcg desogestrel has been shown to inhibit ovulation; most other "traditional" POPs, however, inhibit ovulation in less than one half of cycles, relying on the progestogenic effect of thickening cervical mucus (impedes entry of sperm) for contraception.[51] Until 2021, all POP pills needed to be taken every day (without a pill-free interval) and (in order to maintain effectiveness) within a narrow time window (3 to 12 hours). A new drospirenone POP that inhibits ovulation is taken in a regimen of 24 days of pills, followed by a 4-day pill-free interval. The regimen is designed to reduce unscheduled bleeding. The drospirenone POP is not yet widely available.[52]

Irregular bleeding is a common reason for discontinuation of POP. Almost half of POP users experience prolonged bleeding and up to 70% report irregular bleeding or spotting in one or more cycles.[53] Other reported POP side effects include persistent ovarian follicles (simple cysts), worsening of preexisting acne, and mood change (a causal association between POP use and depression is not, however, established).[54]

LOW EFFECTIVENESS METHODS

Condoms

Male condoms protect against STIs including HIV.[55] Typical contraceptive failure rates are in the region of 13% (see Table 121.2) since they rely on correct, consistent use. The female condom is a lubricated polyurethane condom that is inserted into the vagina. It also protects against STIs.

Diaphragm and Cap

These are latex or silicone devices that are inserted into the vagina to prevent passage of sperm through the cervix.[55] They are inserted in advance of sex. Caps fit over the cervix, whereas the diaphragm forms a hammock between the post-fornix and the symphysis pubis. They are usually used in conjunction with a spermicide, commonly nonoxyl-9 (gel, cream, or pessary). Potential users need to be comfortable to self-examine and insert the device correctly. Typical failure rates of about 17% are reported (see Table 121.2). In some women, use may be associated with increased vaginal discharge and urinary tract infections.

Fertility Awareness-Based Methods

Formerly known as "natural family planning" methods, fertility awareness-based contraceptive methods rely on the signs and symptoms that reflect physiological changes during the menstrual cycle to identify the fertile period.[56,57] Intercourse is then avoided or a barrier method used during that fertile time. With ovarian follicular growth, circulating estrogens increase the cervical mucus which becomes thin and stretchy, facilitating the passage of sperm into the uterus. Following ovulation, progesterone secreted by the corpus luteum thickens the mucus, impeding sperm entry. Progesterone secretion results in a 0.5°C rise in basal body temperature (BBT) that identifies the end of the fertile period. By monitoring cervical mucus quality and BBT, motivated women with regular menstrual cycles can identify their fertile window.

An increasing number of electronic devices and phone apps are available that allow users to record cycle length, BBT, or the results of urinary test dipsticks that detect metabolites of LH. As well as recording these data, some apps will also use the information to help predict the timing of ovulation and thus fertile window in future cycles.

Lactational Amenorrhea

Lactation and suckling raise prolactin, disturb the pulsatile release of gonadotrophins, and so inhibit ovulation. The lactational amenorrhea method can be used for contraception with 98% protection from pregnancy if a woman is within 6 months of childbirth, and is fully (or almost fully) breastfeeding and remains amenorrheic.[58]

Sterilization

This is a permanent method of contraception that prevents sperm reaching the oocyte by obstructing the fallopian tubes.[59] It is most commonly performed laparoscopically but may be carried out at laparotomy or at the time of a cesarean section. Laparoscopic sterilization most commonly occludes the fallopian tube with Filshie clips. Follow–up data available for the Filshie clip suggests a failure rate of 2–5 per 1000 procedures after 10 years.[59] Sterilization failure may increase the risk of ectopic pregnancy. Regret rates are particularly high in women under 30 years of age (up to 30% in some studies).[60] Risks of laparoscopic sterilization include failure and injuries to the bowel, bladder, and blood vessels. Women considering sterilization should be advised that there are subsequent changes in the menstrual bleeding pattern are not due to the tubal occlusion, but due to advancing age or stopping hormonal contraception.[61] There is epidemiological evidence suggesting that bilateral salpingectomy may protect against developing high–grade serous ovarian cancer.[62] It is postulated that this may be because some epithelial cancers may originate in tubal epithelium.

Emergency Contraception

EC is defined as any drug or device used after intercourse to prevent pregnancy. When contraception has not been used or fails, the use of EC can substantially reduce a woman's risk of an unintended pregnancy. The WHO lists oral EC as an essential medicine.[63] EC is extremely safe, and, therefore, countries should have legislation and health policies that facilitate easy access to oral EC. The recommended methods of EC that are in common use include the Cu-IUD and the oral methods LNG and ulipristal acetate (UPA).

Estimates of effectiveness of EC are made from clinical trials comparing the number of observed pregnancies after EC with expected pregnancy rates in the absence of EC (based on likelihood of conception at the time of the menstrual cycle when sex occurred).[64] In recent clinical trials of LNG and UPA, the observed pregnancy rates have been low (1%–2%) and expected pregnancy rates have been calculated at around 5%.[65] One might therefore conclude that oral EC prevents around one-half to two-thirds of pregnancies.[65,66]

Availability of oral EC varies widely across the world, and access to EC is often restricted due to myths about its mechanism of action. The most common is the belief that EC is a form of abortion, rather than a contraceptive, contrary to scientific evidence.[65] Myths also exist that improving access to EC will lead to more STIs or unintended pregnancies. However, here is ample evidence that this is not the case. Adolescents can have difficulty negotiating safe sex, and it is therefore important that they also have good access to EC.

The oral EC methods UPA, a progesterone receptor modulator, and LNG, a progestogen, both work by delaying or inhibiting ovulation, until sperm are no longer viable (presumed lifespan of sperm is 5 days). UPA is licensed to be used within 120 hours of sexual intercourse and LNG within 72 hours. In a meta-analysis of comparative clinical trials, UPA oral EC was shown to be more effective than LNG.[66] This is attributed to the ability of UPA to delay ovulation even after the onset of the LH surge, whereas LNG is no longer able to delay ovulation once the surge (ovulatory trigger) has commenced.[67]

Biomedical studies have demonstrated that the EC dose of LNG or UPA does not exert significant effects on a secretory endometrium or prevent blastocyst implantation *in vitro*.[68–70] LNG and UPA use is not associated with any higher risk of miscarriage, ectopic pregnancy, pregnancy complications, or adverse effects on children conceived after exposure.[71–73]

The type and frequency of reported side effects are similar with both UPA and LNG.[65] There are no absolute contraindications to oral EC, and repeat administration in the same cycle is safe. A Cochrane review of EC reported that women who had further episodes of sex in the same cycle after EC had a 4-fold increased risk of pregnancy.[74] It is, therefore, clearly important that an ongoing method of effective contraception is commenced following EC.

In clinical trials, women with obesity were at higher risk of pregnancy after EC compared with women with a normal BMI,[75] and that in this group the risk of pregnancy was greater amongst those receiving LNG than those using UPA. Pharmacokinetic studies support this finding and have demonstrated that the EC dose of LNG takes a longer time to reach steady-state levels in women with obesity; however, this is not the case for UPA.[76,77]

Cu-IUD is by far the most effective method of EC. The effect of the device on the endometrium is thought to prevent implantation if fertilization has occurred.[12] A systematic review of 42 studies with more than 7000 women calculated that the pregnancy rate was just 1 in 1000 when a Cu-IUD was used for EC.[78] In addition, it has a longer timeframe of use after sex than oral EC. Most guidelines advise that it can be inserted up to 5 days after unprotected sex or 5 days after the earliest estimated date of ovulation, but some studies suggest that it can be inserted much later than this.[79,80] It is an underutilized method of EC as it requires an invasive procedure, and both access to the method and the availability of a healthcare professional trained in insertion.

Non-Contraceptive Benefits of Contraception

Many women will use contraception for over 30 years. Additional health benefits beyond pregnancy prevention offer significant advantages and influence acceptability. Hormonal methods in particular confer non-contraceptive benefits,[81] the commonest being reduced menstrual blood loss (Table 121.5). The 52–mg LNG–IUD is highly effective for management of heavy menstrual bleeding and is more cost–effective than other medical therapies and endometrial ablation.[80] It reduces the pain of endometriosis and adenomyosis.[81] There is evidence that the 52–mg LNG–IUD prevents and causes regression of endometrial hyperplasia and protects the endometrium against the effects of tamoxifen among women with breast cancer.[81]

TABLE 121.5 **Non-Contraceptive Uses and Benefits of Hormonal Contraception**

Method	Management of Condition	Potential Health Benefit
Levonorgestrel intrauterine device	Heavy menses Adenomyosis Endometriosis	Reduced risk of endometrial cancer
Combined hormonal contraception	Heavy menses Irregular menses Dysmenorrhea Hirsutism and acne Premenstrual syndrome	Reduced risk of ovarian cancer Reduced risk of endometrial cancer
Progestogen-only injectable	Heavy menses Endometriosis Dysmenorrhea	Reduced risk of endometrial cancer

There is good evidence that CHC pills can help reduce menstrual bleeding, acne and hirsutism, and symptoms of premenstrual syndrome.[81] The progestogen-only injectable is effective in managing the pain of endometriosis and, given the high incidence of amenorrhea, is also used for treating heavy menstrual bleeding.[81] Hormonal contraception can protect against certain cancers, the most important being the significant reduction in ovarian and endometrial cancer associated with use of the CHC pill.[44] DMPA confers a high degree of protection against endometrial carcinoma and should theoretically also protect against ovarian cancer.[81]

KEY POINTS

- Hormonal contraception can have significant non-contraceptive benefits.

SUMMARY AND FUTURE DIRECTIONS

Although there is a range of hormonal and non-hormonal methods of contraception and delivery systems, the search for new sophisticated methods should continue. Unintended pregnancy rates remain high globally, and current methods are limited by side effects, accessibility, acceptability, or convenience to the user. Additionally, it seems only natural that with advances in our understanding of reproductive biology, women should benefit from new methods of contraception that can be taken more flexibly, have fewer side effects, or confer greater non-contraceptive benefits.

REFERENCES

1. Singh S, Maddow-Zimet I. Facility-based treatment for medical complications resulting from unsafe pregnancy termination in the developing world, 2012: a review of evidence from 26 countries. *BJOG An Int J Obstet Gynaecol*. 2016;123:1489–1498.
2. Ganatra B, Gerdts C, Rossier C, et al. Global, regional, and subregional classification of abortions by safety, 2010-14: estimates from a Bayesian hierarchical model. *Lancet Lond Engl*. 2017;390:2372–2381.
3. Smith GCS, Pell JP, Dobbie R. Interpregnancy interval and risk of preterm birth and neonatal death: retrospective cohort study. *Br Med J*. 2003;327:313.
4. Campbell M, Sahin-Hodoglugil NN, Potts M. Barriers to fertility regulation: a review of the literature. *Stud Fam Plann*. 2006;37:87–98.
5. Hatcher RA. *Contraceptive Technology*. 21st ed. New York: Ayer Company Publishers Inc; 2018.
6. Wellings K, Brima N, Sadler K, et al. Stopping and switching contraceptive methods: findings from Contessa, a prospective longitudinal study of women of reproductive age in England. *Contraception*. 2015;91:57–66.
7. Faculty of Sexual & Reproductive Healthcare. *Drug Interactions with Hormonal Contraception (January 2017, Last Reviewed January 2019)*; 2017. https://www.fsrh.org/standards-and-guidance/current-clinical-guidance/drug-interactions/.
8. World Health Organization (WHO). In: *WHO Medical Eligibility Criteria for Contraceptive Use*. 5th ed. Published Online First; 2015. www.who.int/reproductivehealth/publications/family_planning/MEC-5/en/.
9. Curtis KM, Tepper NK, Jatlaoui TC. US medical eligibility criteria for contraceptive use. *MMWR Recomm Rep*. 2016;65:1–104.
10. Faculty of Sexual & Reproductive Healthcare. *UK Medical Eligibility Criteria for Contraceptive Use (UKMEC)*; 2016. http://www.fsrh.org/standards-and-guidance/uk-medical-eligibility-criteria-for-contraceptive-use/.
11. Glasier A, Scorer J, Bigrigg A. Attitudes of women in Scotland to contraception: a qualitative study to explore the acceptability of long-acting methods. *J Fam Plann Reprod Health Care*. 2008;34:213–217.
12. Stanford JB, Mikolajczyk RT. Mechanisms of action of intrauterine devices: update and estimation of postfertilization effects. *Am J Obstet Gynecol*. 2002;187:1699–1708.

13. Ortiz ME, Croxatto HB. Copper-T intrauterine device and levonorgestrel intrauterine system: biological bases of their mechanism of action. *Contraception*. 2007;75:S16–S30.
14. Apter D, Gemzell-Danielsson K, Hauck B, et al. Pharmacokinetics of two low-dose levonorgestrel-releasing intrauterine systems and effects on ovulation rate and cervical function: pooled analyses of phase II and III studies. *Fertil Steril*. 2014;101:1656–1662.
15. Seeber B, Ziehr SC, GschlieBer A, et al. Quantitative levonorgestrel plasma level measurements in patients with regular and prolonged use of the levonorgestrel releasing intrauterine system. *Contraception*. 2012;86:345–349.
16. Deans EI, Grimes DA. Intrauterine devices for adolescents: a systematic review. *Contraception*. 2009;79:418–423.
17. World Health Organization (WHO). *Selected Practice Recommendations for Contraceptive Use. 3rd ed*. 2016. http://www.who.int/reproductive-health/publications/family_planning/SPR-3/en/.
18. Curtis KM, Jataoui T, Tepper NK. U.S. selected practice recommendations for contraceptive use. *MMWR Recomm Rep*. 2016; 65(No.RR-4):1–66.
19. Faculty of Sexual & Reproductive Healthcare. Intrauterine contraception (April 2015, amended September 2019). http://www.fsrh.org/documents/ceuguidanceintrauterinecontraception/.
20. Nelson A, Apter D, Hauck B, et al. Two low-dose levonorgestrel intrauterine contraceptive systems: a randomized controlled trial. *Obstet Gynecol*. 2013;122:1205–1213.
21. Gemzell-Danielsson K, Schellschmidt I. Apter DA randomized. phase II study describing the efficacy, bleeding profile, and safety of two low-dose levonorgestrel-releasing intrauterine contraceptive systems and Mirena. *Fertil Steril*. 2012;97:616–622.
22. Grimes D, Hubacher D, Lopez LM, et al. Non-steroidal anti-inflammatory drugs for heavy bleeding or pain associated with intrauterine-device use. *Cochrane Database Syst Rev*. 2006;4.
23. Kaislasuo J, Suhonen S, Gissler M, et al. Intrauterine contraception: incidence and factors associated with uterine perforation—a population-based study. *Hum Reprod*. 2012;27:2658–2663.
24. Jatlaoui TC, Riley HEM, Curtis KM. The safety of intrauterine devices among young women: a systematic review. *Contraception*. 2017;95:17–39.
25. Sufrin CB, Postlethwaite D, Armstrong MA, et al. *Neisseria gonorrhea* and *Chlamydia trachomatis* screening at intrauterine device insertion and pelvic inflammatory disease. *Obstet Gynecol*. 2012;120:1314–1321.
26. Hubacher D, Lara-Ricalde R, Taylor DJ, et al. Use of copper intrauterine devices and the risk of tubal infertility among nulligravid women. *N Engl J Med*. 2001;345:561–567.
27. Heinemann K, Reed S, Moehner S. Ectopic pregnancies under IUD use: interim results from the EURAS-IUD study. *Pharmacoepidemiol Drug Saf*. 2013;22:430.
28. Davies GC, Feng LX, Newton JR, et al. Release characteristics, ovarian activity and menstrual bleeding pattern with a single contraceptive implant releasing 3-ketodesogestrel. *Contraception*. 1993;47:251–261.
29. Croxatto HB. Mechanisms that explain the contraceptive action of progestin implants for women. *Contraception*. 2002;65:21–27.
30. Faculty of Sexual & Reproductive Healthcare. *Progestogen-only Implants*; 2014. http://www.fsrh.org/documents/cec-ceu-guidance-implants-feb-2014/. (Accessed January 21, 2016).
31. Mansour D, Korver T, Marintcheva-Petrova M, et al. The effects of Implanon on menstrual bleeding patterns. *Eur J Contracept Reprod Health Care Off J Eur Soc Contracept*. 2008;1(suppl 13):13–28.
32. Cameron S. Subcutaneous depo-medroxyprogesterone acetate. *J Fam Plann Reprod Health Care*. 2013;39:75–77.
33. Cameron ST, Glasier A, Johnstone A. Pilot study of home self-administration of subcutaneous depo-medroxyprogesterone acetate for contraception. *Contraception*. 2012;85:458–464.
34. Heller R, Johnstone A, Cameron ST. The feasibility of contraceptive injections at the community pharmacy. *Eur J Contracept Reprod Health Care Off J Eur Soc Contracept*. 2017;22:327–333.
35. Lopez LM, Ramesh S, Chen M, et al. Progestin-only contraceptives: effects on weight. *Cochrane Database Syst Rev*. 2016:CD008815.
36. Lopez LM, Grimes DA, Schulz KF, et al. Steroidal contraceptives: effect on bone fractures in women. *Cochrane Database Syst Rev*. 2014:CD006033.
37. Evidence for Contraceptive Options and HIV Outcomes (ECHO) Trial Consortium. HIV incidence among women using intramuscular depot medroxyprogesterone acetate, a copper intrauterine device, or a levonorgestrel implant for contraception: a randomised, multicentre, open-label trial. *Lancet Lond Engl*. 2019;394:303–313.
38. World Health Organization (WHO). *Contraceptive Eligibility for Women at High Risk of HIV Guidance Statement—Recommendations on Contraceptive Methods Used by Women at High Risk of HIV*; 2019. https://www.who.int/reproductivehealth/publications/contraceptive-eligibility-women-at-high-risk-of-HIV/en/.
39. Sitruk-Ware R. Hormonal contraception and thrombosis. *Fertil Steril*. 2016;106:1289–1294.
40. Faculty of Sexual & Reproductive Healthcare. *Combined Hormonal Contraception*; 2019. http://www.fsrh.org/documents/combined-hormonal-contraception/.
41. Roach REJ, Helmerhorst FM, Lijfering WM, et al. Combined oral contraceptives: the risk of myocardial infarction and ischemic stroke. *Cochrane Database Syst Rev*. 2015:CD011054.
42. Tepper NK, Whiteman MK, Zapata LB, et al. Safety of hormonal contraceptives among women with migraine: a systematic review. *Contraception*. 2016;94:630–640.
43. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*. 1996;347:1713–1727.
44. Iversen L, Sivasubramaniam S, Lee AJ, et al. Lifetime cancer risk and combined oral contraceptives: the Royal College of General Practitioners' oral contraception study. *Am J Obstet Gynecol*. 2017;216:580.e1–580.e9.
45. Goodman NF, Cobin RH, Futterweit W, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and androgen excess and PCOS society disease state clinical review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome-part 1. *Endocr Pract Off J Am Coll Endocrinol Am Assoc Clin Endocrinol*. 2015;21:1291–1300.
46. Edelman A, Micks E, Gallo MF, et al. Continuous or extended cycle vs. cyclic use of combined hormonal contraceptives for contraception. *Cochrane Database Syst Rev*. 2014:CD004695.
47. Dragoman MV, Tepper NK, Fu R, et al. A systematic review and meta-analysis of venous thrombosis risk among users of combined oral contraception. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet*. 2018;141:287–294.
48. Lidegaard Ø, Milsom I, Geirsson RT, et al. Hormonal contraception and venous thromboembolism. *Acta Obstet Gynecol Scand*. 2012;91:769–778.
49. Lete I, Chabbert-Buffet N, Jamin C, et al. Haemostatic and metabolic impact of estradiol pills and drospirenone-containing ethinylestradiol pills vs. levonorgestrel-containing ethinylestradiol pills: a literature review. *Eur J Contracept Reprod Health Care Off J Eur Soc Contracept*. 2015;20:329–343.
50. Cameron ST, Glasier A, McDaid L, et al. Use of effective contraception following provision of the progestogen-only pill for women presenting to community pharmacies for emergency contraception (Bridge-It): a pragmatic cluster-randomised crossover trial. *Lancet*. 2020;396:1585–1594.
51. Faculty of Sexual & Reproductive Healthcare. *Progestogen-only Pills*; 2015. https://www.fsrh.org/standards-and-guidance/documents/ceuguidanceprogestogenonlypills/.
52. Duijkers IJM, Heger-Mahn D, Drouin D, et al. Maintenance of ovulation inhibition with a new progestogen-only pill containing drospirenone after scheduled 24-h delays in pill intake. *Contraception*. 2016;93:303–309.
53. Kovacs G. Progestogen-only pills and bleeding disturbances. *Hum Reprod Oxf Engl*. 1996;2(suppl 11):20–23.
54. Robakis T, Williams KE, Nutkiewicz L, et al. Hormonal contraceptives and mood: review of the literature and implications for future research. *Curr Psychiatry Rep*. 2019;21:57.
55. Faculty of Sexual & Reproductive Healthcare. *Barrier Methods: Contraception and STI Prevention*; 2012. http://www.fsrh.org/documents/ceuguidancebarriermethodscontraceptionsdi/.
56. Faculty of Sexual & Reproductive Healthcare. *Fertility Awareness Methods*; 2015. http://www.fsrh.org/documents/ceuguidancefertilityawarenessmethods/.

57. Knight J. *The Complete Guide to Fertility Awareness*. Florence: Taylor and Francis; 2016. http://public.eblib.com/choice/publicfullrecord.aspx?p=4732805. (Accessed 12 Dec 2016).
58. Kennedy KI. Efficacy and effectiveness of LAM. *Adv Exp Med Biol*. 2002;503:207–216.
59. Lawrie TA, Kulier R, Nardin JM. Techniques for the interruption of tubal patency for female sterilisation. *Cochrane Database Syst Rev*. 2015:CD003034.
60. Wilcox LS, Chu SY, Eaker ED, et al. Risk factors for regret after tubal sterilization: 5 years of follow-up in a prospective study. *Fertil Steril*. 1991;55:927–933.
61. Faculty of Sexual & Reproductive Healthcare. *Male and Female Sterilisation*; 2014. https://www.fsrh.org/standards-and-guidance/documents/cec-ceu-guidance-sterilisation-cpd-sep-2014/.
62. McAlpine JN, Hanley GE, Woo MMM, et al. Opportunistic salpingectomy: uptake, risks, and complications of a regional initiative for ovarian cancer prevention. *Am J Obstet Gynecol*. 2014;210:471.e1–e11.
63. World Health Organization (WHO). The 2019 Expert Committee on the Selection and Use of Essential Medicines. *WHO Model Lists of Essential Medicines. 21st ed*. 2019. https://www.who.int/medicines/publications/essentialmedicines/en/.
64. Trussell J, Ellertson C, von Hertzen H, et al. Estimating the effectiveness of emergency contraceptive pills. *Contraception*. 2003;67:259–265.
65. Cameron ST, Li H, Gemzell-Danielsson K. Current controversies with oral emergency contraception. *BJOG An Int J Obstet Gynaecol*. 2017;124:1948–1956.
66. Glasier AF, Cameron ST, Fine PM, et al. Ulipristal acetate versus levonorgestrel for emergency contraception: a randomised non-inferiority trial and meta-analysis. *Lancet Lond Engl*. 2010;375:555–562.
67. Brache V, Cochon L, Jesam C, et al. Immediate pre-ovulatory administration of 30 mg ulipristal acetate significantly delays follicular rupture. *Hum Reprod Oxf Engl*. 2010;25:2256–2263.
68. Stratton P, Levens ED, Hartog B, et al. Endometrial effects of a single early luteal dose of the selective progesterone receptor modulator CDB-2914. *Fertil Steril*. 2010;93:2035–2041.
69. Lalitkumar PGL, Lalitkumar S, Meng CX, et al. Mifepristone, but not levonorgestrel, inhibits human blastocyst attachment to an in vitro endometrial three-dimensional cell culture model. *Hum Reprod Oxf Engl*. 2007;22:3031–3037.
70. Berger C, Boggavarapu NR, Menezes J, et al. Effects of ulipristal acetate on human embryo attachment and endometrial cell gene expression in an in vitro co-culture system. *Hum Reprod Oxf Engl*. 2015;30:800–811.
71. Shen J, Che Y, Showell E, et al. Interventions for emergency contraception. *Cochrane Database Syst Rev*. 2017;8:CD001324.
72. Levy DP, Jager M, Kapp N, et al. Ulipristal acetate for emergency contraception: postmarketing experience after use by more than 1 million women. *Contraception*. 2014;89:431–433.
73. Cleland K, Raymond E, Trussell J, et al. Ectopic pregnancy and emergency contraceptive pills: a systematic review. *Obstet Gynecol*. 2010;115:1263–1266.
74. Zhang L, Ye W, Yu W, et al. Physical and mental development of children after levonorgestrel emergency contraception exposure: a follow-up prospective cohort study. *Biol Reprod*. 2014;91:27.
75. Glasier A, Cameron ST, Blithe D, et al. Can we identify women at risk of pregnancy despite using emergency contraception? Data from randomized trials of ulipristal acetate and levonorgestrel. *Contraception*. 2011;84:363–367.
76. Edelman AB, Cherala G, Blue SW, et al. Impact of obesity on the pharmacokinetics of levonorgestrel-based emergency contraception: single and double dosing. *Contraception*. 2016;94:52–57.
77. Praditpan P, Hamouie A, Basaraba CN, et al. Pharmacokinetics of levonorgestrel and ulipristal acetate emergency contraception in women with normal and obese body mass index. *Contraception*. 2017;95:464–469.
78. Cleland K, Zhu H, Goldstuck N, et al. The efficacy of intrauterine devices for emergency contraception: a systematic review of 35 years of experience. *Hum Reprod Oxf Engl*. 2012;27:1994–2000.
79. Faculty of Sexual & Reproductive Healthcare. Emergency contraception (March 2017, amended December 2017). https://www.fsrh.org/standards-and-guidance/current-clinical-guidance/emergency-contraception/.
80. Turok DK, Sanders JN, Thompson IS, et al. Preference for and efficacy of oral levonorgestrel for emergency contraception with concomitant placement of a levonorgestrel IUD: a prospective cohort study. *Contraception*. 2016;93:526–532.
81. Cameron S. Contraception and gynaecological care. *Best Pract Res Clin Obstet Gynaecol*. 2009;23:211–220.

122

Menopause and Perimenopause

John Rushing, Malgorzata E. Skaznik-Wikiel, and Nanette Santoro

OUTLINE

Mechanisms of Menopause and the Transition, 2030
Staging Reproductive Aging, 2030
  Late Reproductive Stage (Stage –3), 2030
  Menopausal Transition, 2031
  Menopause, 2031
  Postmenopause, 2031
Epidemiology, 2032
  Age of Menopause, 2032
  Population-Level Determinants of Menopause, 2032
  Determinants of Early Menopause and Primary Ovarian Insufficiency, 2032
Pathophysiology of Menopause, 2033
  Ovarian Follicle Depletion—Current State of Knowledge, 2033
  Follicular Growth Activation and Atresia, 2033
  Ovarian and Pituitary Hormones, 2033
Clinical Presentation, 2034
  The "Core Four" Symptoms: Hot Flashes, Vaginal Dryness, Adverse Mood, and Poor Sleep, 2034
  Vasomotor Symptoms (Hot Flashes and Night Sweats), 2034
  Vaginal Dryness, 2034
  Adverse Mood, 2034
  Sleep Disturbances, 2034
  Other Symptoms, 2035
Reproductive and Somatic Aging, 2035
  Hypothalamic-Pituitary-Ovarian Axis, 2035
  Somatotropic Axis, 2035
  Thyroid, 2035
  Adrenal, 2035
  Bone, 2035
  Cognition, 2035
  Cardiovascular, 2036
  Estrogen-Dependent Disorders, 2036
Diagnosis, 2037
Treatment, 2037
  Hormone Therapy Overview, 2037
  Vasomotor Symptoms, 2037
  Urogenital Symptoms (GSM), 2038
  Sleep Disturbances, 2038
  Adverse Mood, 2038
  Other Common Issues, 2038
  Complementary and Alternative Medicine in the Treatment of Menopausal Symptoms, 2039
  Long-Term Use for Prevention of Chronic Conditions, 2039
  Cancer Risks and Mitigation of Risk With Hormone Therapy, 2041
Summary, 2042

KEY POINTS

- Menopause is associated with multiple potentially debilitating symptoms including, but not limited to, vasomotor symptoms (VMS), urogenital symptoms, mood changes, sleep disturbances, and decrease in bone density.
- Hormone therapy (HT) remains the most effective treatment modality for severe menopausal symptoms.
- The current recommendation is to use HT in the lowest effective dose for the shortest time needed to control symptoms.

MECHANISMS OF MENOPAUSE AND THE TRANSITION

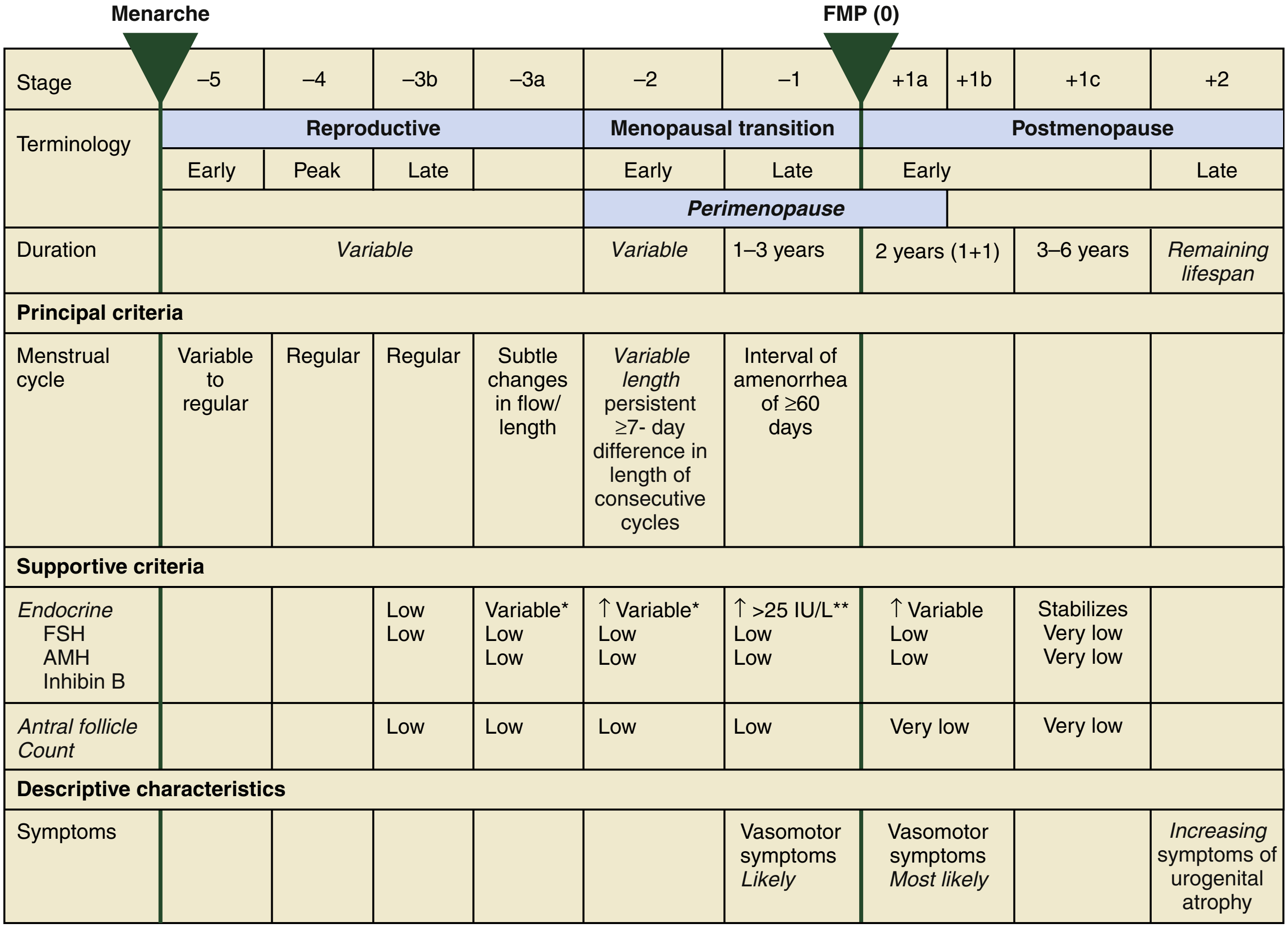
Why do human females among almost all other species survive long after menopause? In 1997, Kristen Hawkes first proposed the grandmother hypothesis of menopause,[1] an evolutionary explanation of the menopause, which supports the idea that the fitness of a population is maximized if women stop reproducing and instead help to raise their grandchildren. Continued investment in already existing children and grandchildren could increase average lifetime reproductive success to a larger degree than a woman's increasingly risky own reproduction as she grows older. Women's attitudes toward menopause range from neutral to positive, and ethnic groups within the United States vary only slightly in their attitudes toward menopause and aging.[2] Although menopause is widely viewed in a negative fashion with an excessive focus on adverse symptoms, it is important to remember that it can be a positive milestone for women.

STAGING REPRODUCTIVE AGING

The Staging Reproductive Aging workshops (STRAW) in 2001 and 2011 provided nomenclature to help standardize research and clinical care in the field.[3] The revised STRAW+10 Staging System (Fig. 122.1) is described in the following sections.

Late Reproductive Stage (Stage –3)

This stage marks the time when fecundability begins to deteriorate and a woman may begin to notice changes in her menstrual cycles. It is further subdivided into two substages (–3b and –3a). In stage –3b, periods remain regular without a change in follicle-stimulating hormone (FSH) levels; however, anti-Müllerian hormone (AMH) and antral



Menarche

FMP (0)

Stage	−5	−4	−3b	−3a	−2	−1	+1a	+1b	+1c	+2
Terminology	**Reproductive**				**Menopausal transition**		**Postmenopause**			
	Early	Peak	Late		Early	Late	Early			Late
					Perimenopause					
Duration	*Variable*				*Variable*	1–3 years	2 years (1+1)		3–6 years	*Remaining lifespan*
Principal criteria										
Menstrual cycle	Variable to regular	Regular	Regular	Subtle changes in flow/ length	*Variable length* persistent ≥7- day difference in length of consecutive cycles	Interval of amenorrhea of ≥60 days				
Supportive criteria										
Endocrine FSH AMH Inhibin B			Low Low	Variable* Low Low	↑ Variable* Low Low	↑ >25 IU/L** Low Low	↑ Variable Low Low		Stabilizes Very low Very low	
Antral follicle Count			Low	Low	Low	Low	Very low		Very low	
Descriptive characteristics										
Symptoms						Vasomotor symptoms *Likely*	Vasomotor symptoms *Most likely*			*Increasing* symptoms of urogenital atrophy

* Blood draw on cycle days 2–5 ↑ = elevated

** Approximate expected level based on assays using current international pituitary standard[67–69]

Figure 122.1 Working model of the female reproductive life span, including but not limited to the menopausal transition. The final menstrual period (FMP) is indicated at time 0, and the menopausal transition (stages –2 through 0) encompasses the onset of increased menstrual irregularity and/or skipped menses up to the FMP. Note that intermittent elevations in early follicular phase follicle-stimulating hormone (FSH) are detectable before the onset of any clinical sign of the transition (stage –3a), and that the FMP is not defined until 1 year without menses has elapsed. *AMH,* Anti-Müllerian hormone. (From Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Fertil Steril.* 2012;97:843–851.)

follicle counts (AFC) are declining. In stage –3a, shorter cycles begin and FSH levels becomes more erratic.

Menopausal Transition

Stages –2 (early) and –1 (late) encompass the menopausal transition, which is defined by menstrual cycle and further endocrine changes. Stage –2 is defined by a persistent difference in the menstrual cycle length of 7 days or more between consecutive cycles. It also includes variable early follicular phase FSH levels and lower AMH and AFC compared with prior stages. Stage –1 is manifested by the occurrence of amenorrhea of 60 days or longer. FSH levels are sometimes elevated into the menopausal range and sometimes within normal limits for earlier reproductive years. This stage is estimated to last 1 to 3 years. The menopausal transition ends with the final menstrual period (FMP).

Menopause

The anchor point of menopause (Stage 0) is defined after 12 months of amenorrhea following the FMP and reflects a near-complete but natural decrease in ovarian hormone secretion. AMH is typically undetectable by the FMP; and AMH levels, if the assay is sufficiently sensitive, can be used to predict the FMP, especially for women aged 48 and older.[4]

Postmenopause

Stages +1 (early) and +2 (late) encompass the postmenopause. Stage +1 is further divided into stages +1a, +1b, and +1c. Stage +1a indicates the end of the 12-month period of amenorrhea required to define that the FMP has occurred. Stage +1b includes the remainder of the period of rapid change in FSH and estradiol levels. Stages +1a and +1b together are estimated to last about 2 years. Menopausal symptoms are most likely to occur during this stage. Stage +1c represents the period of stabilization of high FSH and low estradiol levels that is estimated to last 3 to 6 years. Stage +2 represents the period in which further changes in reproductive endocrine function are more limited and processes of somatic aging predominate. The duration of this stage varies since it ends with death.

TABLE 122.1 **Etiologic Factors in Early Menopause**

Race/ethnicity
Parity
Prior oral contraceptive use
Socioeconomic status
Lower educational attainment
Marital status
Stress
Familial/genetic factors
Blepharophimosis gene
POF 1/POF 2 gene
Fragile X syndrome
PVU II polymorphic allele
Environmental toxins
Smoking
Chemotherapy
Irradiation
Galactose consumption
Body mass index (BMI)
Depression

EPIDEMIOLOGY (TABLE 122.1)

Age of Menopause

The Study of Women's Health Across the Nation (SWAN), a multicenter multiethnic, community-based cohort study of women and the menopausal transition, reported the overall median age at natural menopause to be 52.5 years, after adjustment for other factors.[5] Studies performed outside the United States suggest that Africans,[6] Black Americans,[7] and Hispanics of Mexican descent[8] experience menopause at an earlier age than White women, as opposed to Japanese[9] and Malaysian[10] women, who report a similar median age of menopause to women of European descent. However, findings from a recent analysis of SWAN found that there were no significant differences in the age of natural menopause in women of different racial backgrounds living in the United States, when adjusted for other known demographic factors that influence the age at FMP.[5] Women who are separated, divorced, or widowed have an earlier menopause than women who are married, as do women who smoke or who report difficulty paying for the basic necessities of life.[5] Early menopause is defined as an FMP at or before 45 years of age and late menopause is defined as an FMP that occurs after 54 years of age. Early menopause is present in about 12% of women.[11] Approximately 3% of women experience hypergonadotropic amenorrhea or primary ovarian insufficiency (POI) before 40 years of age.[11] In SWAN, the median age at FMP was significantly older in women who did not smoke, reported better health at baseline, had more education, had higher baseline weight, or reported prior use of oral contraceptives.[5,12] Age at natural menopause for parous women has been reported to occur significantly later than that for nulliparous women.[5,12] A trend of increasing age at menopause with increasing number of live births has also been observed. The proposed mechanism by which parity and use of oral contraceptives may result in later age at natural menopause involves reducing ovulatory cycles earlier in life and thus preserving the oocyte pool for a longer period of time.[5]

Population-Level Determinants of Menopause

Women are born with a finite number of oocytes and go through progressive follicular loss throughout their lifespan. The primary drivers of the menopause transition are alterations in follicle number as well as follicle dynamics. AMH, a transforming growth factor-β superfamily glycoprotein, is produced by granulosa cells of developing primary, secondary, and early antral follicles and has emerged as a marker of ovarian reserve as well as timing of future menopause.[4,13] Nair et al.[13], from a heterogenous population-based sample using premenopausal women from CARDIA Women's Study, found that AMH is independently associated with age at natural menopause. A 0.5 ng/dL AMH reduction correlated with higher risk of menopause ($P < .001$) in fully adjusted discrete-time hazard models, and AMH appears most valuable in recognizing women at risk of menopause in the proximate future.[13] These findings were confirmed and refined further using the SWAN study cohort, with the observation that AMH was a better predictor for menopause than FSH, and AMH measurements could help in estimating when a woman will undergo her FMP.[4] The likelihood of a woman with an AMH < 10 pg/mL would experience her FMP within the next year ranged from 51% at < 48 years old to 79% at ≥ 51 years old. The likelihood of a woman with an AMH > 100 pg/mL would not experience her FMP within the next year ranged from 97% in women < 48 years old to 90% in women ≥ 51 years old.[4] Thus, both a low and high AMH level can be informative for clinical management. The rate of change in AMH may provide further valuable information about the timing of menopause, in that women in the PENN Ovarian Aging Study who had evidence of a more rapid decline were found to be more likely to undergo menopause sooner than those who displayed a lower rate of a decline in serial AMH measurements over time.[14]

Genetic factors have been proven to contribute to the onset of menopause as evidenced by studies involving family history and twin registries. Heritability estimates from twin and family studies for age at natural menopause range from 44% to 63%.[15,16] The association of menopause with heritability has led to many genome-wide association studies (GWAS) that have identified loci for age at menarche and age at natural menopause. To date, 44 genetic loci have been identified for age at natural menopause, but these likely only account for a small portion of overall heritability.[17] Many of these genes are linked to premature menopause (see below) and have been found to play a fundamental role in genome stability and DNA repair.[17]

Determinants of Early Menopause and Primary Ovarian Insufficiency

Single gene mutations, chromosomal deletions, and genetic variants have been shown to influence the timing of menopause, and many of these genetic alterations are associated with early menopause (less than age 45) and POI (less than age 40). The blepharophimosis gene, located on chromosome 3, X-chromosome deletions such as *POF 1* and *POF 2* genes, and mutations in the forkhead transcription factor FOX have been shown to predispose women to an earlier menopause or POI. POI and early menopause share the same genetic features and appear to represent variable expression of the same disease.[17] Fragile X syndrome premutation carriers are prone to premature menopause. The *PvuII* polymorphic allele for estrogen receptor alpha (ERα) has been associated with a slightly earlier age at natural menopause.[18] DNA mismatch repair and other genome maintenance genes and immune modulators are associated with POI as well as being determinants of age at natural menopause.[19] Several of the genes identified to date as responsible for POI and early menopause appear to play a critical regulatory role in primordial follicle activation (see below).

Environmental toxicants may also play a role in early menopause. Current smokers tend to experience menopause at an earlier age (1 to 2 years) than nonsmokers,[5] and may have a shorter menopausal transition.[12] Polycyclic hydrocarbons in cigarette smoke are believed to be toxic to ovarian follicles[20] and may lead to their loss, and thus an earlier menopause in smokers. Irradiation and chemotherapy with alkylating cytotoxics are iatrogenic causes of earlier menopause.[21] Evidence suggests that galactose consumption through the ingestion of high-lactose dairy foods may be a dietary risk factor, and that galactose metabolism, as measured by galactose-1-phosphate uridyl transferase, is a genetic risk factor for early menopause.[22]

Women with a history of medically treated depression have a 20% increased rate of entering perimenopause sooner than women with no depression history after adjustment for age, parity, age at menarche, education, cigarette smoking, and body mass index (BMI).[23]

PATHOPHYSIOLOGY OF MENOPAUSE

Ovarian Follicle Depletion—Current State of Knowledge

The basis of reproductive senescence in women is oocyte/follicle depletion in the ovary. The most massive wave of atresia (follicle loss) that a woman ever experiences happens before she is born. Developmentally, a woman attains her peak oocyte complement at 20 weeks' gestational age. Between 20 and 40 weeks' gestation, the total oocyte count drops from a mean of about 6 to 8 million to 1 to 2 million. By the onset of puberty, germ cell mass has been reduced to 300,000 to 500,000 units. Subsequent reproductive aging consists of a steady loss of oocytes through atresia or ovulation and does not necessarily occur at a constant rate. Atresia is an apoptotic process. During the 35 to 40 years of reproductive life, 400 to 500 oocytes will be selected for ovulation. By menopause, only a few hundred follicles will remain.[24] The relatively wide age range (42 to 58 years) for menopause in normal women seems to indicate that initial oocyte endowment in women is highly variable, or that the rate of oocyte loss varies greatly. The concept that women are endowed with a finite oocyte number that can only be lost and never replenished has been challenged by several studies and is an active area of investigation. Current evidence suggests that mammalian females may generate new oocytes during adult life through the activity of germline stem cells.[25,26] Whether or not human females can regenerate oocytes from oogonial stem cells is highly controversial. Current research directed towards repopulating the human ovary with mesenchymal stem cells is underway, but convincing medical evidence has not yet been forthcoming.

Follicular Growth Activation and Atresia

Developing follicles can be divided into gonadotropin-independent preantral follicle growth and gonadotropin-dependent antral follicle growth (Fig. 122.2). Primordial follicle growth activation (PFGA) occurs when dormant individual follicles are recruited to a growth phase that can result in the development of a preovulatory follicle hosting an immature oocyte. The rate of PFGA is a key factor of ovarian aging. c-Kit/Kit ligand (KL) signaling was one of the initial pathways identified as a PFGA regulatory mechanism. Interruption of this pathway results in heightened arrest of primordial follicles and a decrease in the amount of preantral and antral follicles.[27] A second key regulatory molecule is FoxO3. *FoxO3* knockout mice exhibit a full commitment of primordial follicles to growth early in life that results in the exhaustion of the entire primordial follicle pool by 2–3 months of age. Other central negative regulators of PFGA include phosphatase and tensin homolog (Pten), protein kinase Akt, and mechanistic target of rapamycin (mTOR). High Pten activity downregulates Akt resulting in low downstream mTOR activity; this interaction culminates in impeded cell cycle progression. Absence of negative regulators such as *Pten*, *FoxO3*, and mTOR in mouse models results in vastly accelerated PFGA and early life depletion of the primordial follicle pool.[27] Further understanding of factors that regulate entry into the growing follicle pool is likely to inform treatment targets to prevent early or premature menopause, as many of the genes and gene pathways identified herein are associated with POI and early menopause.

Ovarian and Pituitary Hormones

Concurrent with the loss of ovarian follicles as a woman transitions to menopause are hormonal changes in the hypothalamic-pituitary-ovarian axis. FSH is an indirect marker of follicular activity; as follicle numbers decline, FSH levels increase.[28] In the late reproductive years, initial elevations in FSH are most prominent in the early follicular phase of the menstrual cycle but are intermittent, i.e., they do not occur in every cycle, and they are clinically silent.[28] The rise in FSH appears to be the result of a decline in ovarian inhibin B, a dimeric protein secreted by the growing granulosa cells that surround the oocyte. Inhibin serves to selectively inhibit FSH by binding to receptors on the anterior pituitary gonadotrophs.[28] As the pool of growing follicles shrinks with the overall decline in follicle numbers that occurs with aging, inhibin production is decreased. Circulating estradiol is stable or even elevated during the earlier menopause transition, because the dominant follicle still matures monthly and secretes over 90% of the estradiol produced at midcycle; closer to the FMP, a decline in estradiol can be clearly observed.[28,29] The decline in inhibin B precedes the increase in FSH and the decline in estradiol that occur later in the

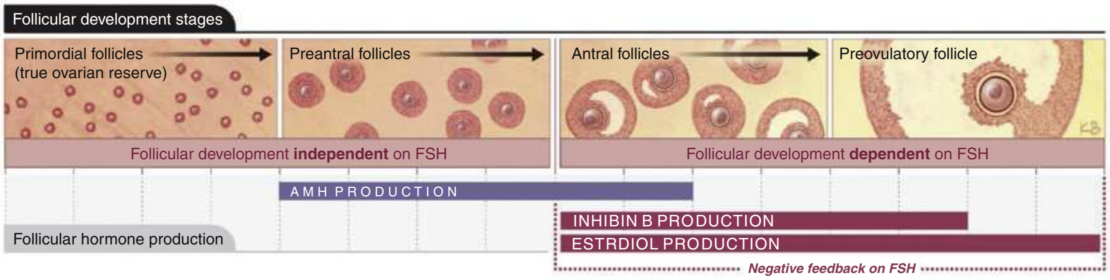

Follicular development stages
Primordial follicles (true ovarian reserve)
Preantral follicles
Antral follicles
Preovulatory follicle
KB
Follicular development **independent** on FSH
Follicular development **dependent** on FSH
AMH PRODUCTION
INHIBIN B PRODUCTION
ESTRDIOL PRODUCTION
Follicular hormone production
Negative feedback on FSH

Figure 122.2 Ovarian follicles at various stages of development. Follicular growth can be divided into gonadotropin-independent and gonadotropin-dependent phases. Primordial follicles represent the ovarian reserve pool from which all other follicles are derived, and the rate of primordial follicle growth activation (PFGA) is a key factor of ovarian aging. c-Kit/Kit Ligand (KL) signaling is a key regulatory pathway identified in activating primordial follicle growth, and interruption of this pathway results in heightened arrest of primordial follicles and a decrease in the amount of preantral and antral follicles. Other central negative regulators of PFGA include forkhead transcription factor 3 (FoxO3), phosphatase and tensin homolog (Pten), protein kinase Akt, and mechanistic target of rapamycin (mTOR). Absence of these negative regulators results in vastly accelerated PFGA and early life depletion of the primordial follicle pool. *AMH,* Anti-Müllerian hormone. (From Santoro N, Johnson J. Diagnosing the onset of menopause. *JAMA.* 2019;322:775–776.)

transition.[29] AMH, produced by preantral and antral follicles, inhibits initial follicular recruitment and inhibits FSH-dependent growth and selection of preantral and small antral follicles.[30] As the follicular pool decreases in patients approaching menopause, AMH levels will also decrease which can contribute to more rapid activation and depletion of primordial follicles.

Although circulating FSH and estradiol vary near menopause, steroidogenic enzymes are completely absent in the postmenopausal ovary after all functional follicles are lost.[31] Between 20 and 40 years of age, concentrations of total testosterone fall by about 50%.[32] This age-related decline does not change further during the transition years. Similarly, dehydroepiandosterone (DHEA) and its sulfate, DHEAS, decline with age.[33] Because circulating sex hormone-binding globulin (SHBG) decreases across the menopausal transition, free androgen levels rise, as indicated by a small increase in free androgen index (T ÷ SHBG × 100).[34] Androstenedione, which remains relatively stable during the transition, is converted to estrone in extraglandular tissue. This accounts for almost all the circulating estrogen after menopause. When ovulation stops, concurrent with a woman's FMP, serum progesterone levels are invariably low.

CLINICAL PRESENTATION

The "Core Four" Symptoms: Hot Flashes, Vaginal Dryness, Adverse Mood, and Poor Sleep

Impending menopause is clinically evident when a woman observes menstrual cycle changes. A normal menstrual cycle ranges from every 21 to 35 days. Anovulatory cycles become more common as women progress through the menopause transition, and cycle lengths become increasingly variable. By the late menopausal transition, that ultimately ends with FMP, amenorrhea of 60 days or longer occurs.[3] These menstrual cycle changes are related to the increase in overall symptoms. Entry into the early menopause transition is accompanied by generally mild but increased symptoms, and by the late transition symptoms usually peak, and thereafter most will gradually subside (except for vaginal dryness, which will not improve without treatment). It is sometimes difficult to differentiate symptoms or clinical complaints that are truly associated with menopause from those due to aging. Hot flashes, night sweats, and vaginal dryness are most clearly tied to the menopausal transition, and there is some positive evidence of a menopausal link for sleep disturbance. Mood symptoms, such as anxiety and major depression, have been tied to the menopausal transition.[35]

Vasomotor Symptoms (Hot Flashes and Night Sweats)

Hot flashes are considered to be a hallmark symptom of the menopausal transition, and approximately 60% to 80% of menopausal women experience hot flashes at some point during the menopausal transition.[36] The extent to what genetic factors play a role in who will experience VMS is unknown, but recent GWAS studies have shown that genetic variations in tachykinin receptor 3 *(TACR3)* may contribute to the risk of VMS.[37] Hot flashes are commonly defined as transient periods of intense heat in the upper arms and face, which often are followed by flushing of the skin and profuse sweating.[38] Many hot flashes are followed by chills and can be accompanied by palpitations and a sense of anxiety. Almost two thirds of women seek medical attention, usually those who are more symptomatic, with VMS and adverse mood being the biggest motivators.[39] Hot flashes negatively affect quality of life for women by causing sleep disturbances, which often result in fatigue, irritability, forgetfulness, acute physical discomfort, and negative effects on work.[39] The natural history of hot flashes has been examined in a longitudinal sample of women from SWAN; the median total duration of hot flashes was 7.4 years, with a median persistence of 4.5 years past the FMP. Women with onset of bothersome hot flashes earlier in the transition tended to have a longer total duration. Median hot flash duration was longest for Black American women (10.1 years).[40]

Although the etiology of hot flashes is not entirely clear, studies suggest that estrogen levels play a role in inhibiting the thermoregulatory center and expression of VMS. KNDy neurons, located within the arcuate nucleus of the hypothalamus, project to thermoregulatory areas in the median preoptic nucleus that express neurokinin 3 receptors (NK_3R). The median preoptic nucleus is vital for heat dissipation via cutaneous vasodilation. ERα is required for estrogen negative feedback and is located on KNDy neurons. A decrease in E_2 levels seen in menopause stimulates KNDy neurons to active NK_3R on the median preoptic nucleus which alters the thermoregulatory response giving rise to hot flashes.[41]

Vaginal Dryness

Vaginal dryness, due to decreased estrogen exposure, is another common symptom associated with the menopause transition. Transudative fluid leaked from capillaries across the vaginal epithelial cells is responsible for vaginal lubrication, and estrogen is a major contributing factor for this process. In particular for menopause, low estrogen levels results in atrophic epithelial cells, and thus a decrease in transudative fluid.[42] In SWAN participants, the prevalence of vaginal dryness increased from 19.4% among all women at baseline (ages 42–53 years) to 34.0% at the ages 57–69 years at baseline,[43] and vaginal dryness was most prevalent in Hispanic women.[44] Early perimenopausal women reported greater pain with intercourse than premenopausal women,[45] and a decline in sexual function was observed 20 months prior to the FMP.[46] Over 75% of the middle-aged women in the SWAN study reported that sex was moderately to extremely important to them, underscoring the clinical importance of evaluating this complaint

Adverse Mood

Data from SWAN showed that in a large sample of middle-aged women, during 10 years of follow-up, women who had low anxiety at baseline were significantly more likely to develop high-anxiety symptoms at perimenopause or postmenopause than at premenopause, independent of multiple risk factors, including upsetting life events, financial strain, fair/poor perceived health, and VMS. However, among women who started the study with high anxiety, menopausal stage was not associated with an increase in anxiety whether or not VMS were included in the analyses. In addition, SWAN and other cohort studies have clearly shown that the late perimenopause is a period of increased risk for depressive symptoms as well as clinical and subclinical depressive disorders.[35]

Sleep Disturbances

Poor sleep is associated with natural menopause, and the highest rates of sleep disturbances are reported in late perimenopausal women. There are many variables that contribute to sleep difficulty during menopause rather than a single primary determinant.

While age and VMS are contributing factors to poor sleep, SWAN data implies that the menopausal transition independent of age is linked to self-reported sleep difficulty, and subgroup analyses limited to women without VMS continue to find a relationship between the perimenopause and diminished sleep quality. Many additional contributing factors that were associated with sleep difficulty included depression, anxiety, quality of life, self-perceived health, current smoker, sedentary lifestyle, and arthritis.[47] In another population-based study of perimenopausal women that included 14 years of follow-up, a high prevalence of poor sleep was reported in middle-aged women; premenopausal baseline sleep status and simultaneous hot flashes strongly and regularly predicted poor sleep in the menopause transition.[48]

Other Symptoms

Menstrual migraines peak during the transitional years, as do often unrecognized hot flashes around menses.[49] In the early menopause transition, breast soreness-tenderness is significantly more common, but as the menopause transition progresses, more hypoestrogenic symptoms such as hot flashes, vaginal dryness, and night sweats occur with greater frequency.[50] Neuromuscular symptoms have been found to be stable throughout the menopausal transition, suggesting that other events in a midlife woman's life (i.e., life stress and acute and chronic illnesses) account for the common complaint of joint pain.[51]

KEY POINTS

- Menopause is associated with multiple potentially debilitating symptoms including, but not limited to, vasomotor symptoms (VMS), urogenital symptoms, mood changes, sleep disturbances, and decrease in bone density.

REPRODUCTIVE AND SOMATIC AGING

Hypothalamic-Pituitary-Ovarian Axis

Reproductive aging is confounded by somatic aging. All women do not age at the same rate. Menopause has been regarded historically as a primary ovarian event, with associated changes in pituitary gonadotropin secretion occurring secondary to the decline in ovarian sex steroid and protein production. However, increasing evidence suggests that aging is associated with dynamic changes in the hypothalamic and pituitary components of the reproductive axis that are independent of changes in gonadal hormone secretion. Using gonadotropin free α-subunit as a surrogate marker of gonadotropin-releasing hormone (GnRH) pulse frequency in postmenopausal women, Hall et al.[52] demonstrated a 35% decline in GnRH pulse frequency between the fifth and eighth decades of life, providing evidence of slowing of hypothalamic GnRH pulse generator activity. Other studies in which LH was used as a marker for GnRH secretion have shown variable results. Santoro et al.[53] did not observe a different rate of LH secretion in postmenopausal women who were compared with young women with premature ovarian failure, but older women had less overall organization of their LH secretion, and LH was much more readily suppressed by exogenous estradiol. Taken together, there appears to be an overall decline in hypothalamic-pituitary organization and output in association with aging per se apart from menopause.

Somatotropic Axis

The "somatopause" refers to the age-related drop in growth hormone (GH) and it begins prior to the menopause. The 24-hour GH secretion decreases in women before the onset of estrogen deficiency, and insulin-like growth factor-1 (IGF-1) levels in adults progressively decline with age.[54] Because IGF-1 is a cofactor for FSH action, additional gonadal compromise may be imposed by reduced somatotropic axis function.[55] Animal studies demonstrate the necessity of interactions between estrogen and IGF-1 for reproductive and menstrual functioning, suggesting direct mechanisms by which somatopause may hasten reproductive aging.[56] The known association of increased levels of physical activity with elevated IGF-1 levels in premenopausal women raises the possibility that decreased physical activity associated with aging hastens natural somatopause and may also hasten reproductive aging.[57] The somatotropic axis may be amenable to modulation using relatively simple modalities. One small study of progesterone supplementation in postmenopausal women observed a restoration of more youthful GH secretion. Progesterone has been shown by others to prevent sleep disturbances and to positively modulate GH, thyroid-stimulating hormone (TSH), and melatonin secretion in postmenopausal women.[58] Estrogen also modulates the somatotropic axis. Women with POI have reduced circulating IGF-1 compared with age-matched eugonadal women, supporting age- and estrogen-dependent changes.[59] Body composition changes, including an increase in body fat in the visceral/abdominal compartment and a decrease in lean body mass, as well as adverse changes in lipoproteins, a relative increase in insulin resistance, and a reduction in aerobic capacity, are thought to be secondary to changes in the somatotropic axis.[60]

Thyroid

As with other endocrine systems, aging is associated with significant changes in the anatomy of the thyroid gland and modifications in the physiology of the hypothalamic-pituitary-thyroid axis. Subtle changes in the hypothalamic-pituitary-thyroid axis suggest decreased hypothalamic stimulation of thyroid function with aging.[61] The TSH nocturnal surge may be lost with increasing age.[62] Moreover, thyroid hormone clearance decreases with age, which explains the reduced dosing needs for daily replacement of thyroxine in hypothyroid older adults.[63] Progesterone supplementation in postmenopausal women has been found to decrease TSH levels, implying that loss of endogenous progesterone is associated with the age-related rise in TSH.[58]

Adrenal

With aging, minimal and mean cortisol plasma concentrations increase, with no alteration noted in pulsatile frequency, and the diurnal amplitude of cortisol and adrenocorticotropic hormone relative to the 24-hour mean of hormones shows an age-related decline.[64] In addition, the evening cortisol quiescent period is shortened in older adults, suggesting an increasingly impaired circadian function with aging. Despite maintenance of corticosteroid secretion, DHEAS levels decline with age,[65] while testosterone levels are preserved.[33] However, DHEAS levels are not consistently lower in younger women with POI,[66] suggesting that these changes are not completely estrogen dependent. Declining DHEAS in menopausal women has been associated with bone loss[67] and a higher risk for cardiovascular mortality.[68] However, during the late menopausal transition, DHEAS rises transiently. During this transient rise in DHEAS, higher DHEAS was positively associated with depression in one cohort study.[69] The complexities of age- and hormone-dependent changes in adrenal function in women to date are incompletely understood.

Bone

In relation to bone and aging around the menopausal transition, the late menopausal transition is clearly the point in time at which menopause-associated bone resorption begins. Results from SWAN showed that bone mineral density (BMD) loss began 1 year before the FMP and decelerated (but did not cease) 2 years after the FMP (termed the transmenopause), at both the lumbar spine (LS) and femoral neck sites.[70] SWAN has also observed that in premenopausal and early perimenopausal women, higher FSH concentrations were positively associated with greater bone turnover before the last menstrual period, as assessed by higher serum osteocalcin and urinary N-telopeptide of type I collagen concentrations.[70] Serum FSH was inversely correlated with BMD, and BMD tended to be lower in women in the late perimenopause or early postmenopause versus those in the early perimenopause.[70] Finally, SWAN found that more sport and home physical activity, compared with work and active living activity, were statistically significantly associated with greater BMD,[70] suggesting the importance of preserving activity in middle-aged women.

Cognition

Cognitive function declines with age. Leisure activity may mitigate those effects. In a prospective, cohort study, Verghese et al.[71] found that reading and playing board games were associated with reduced risks for dementia and Alzheimer's disease. Combined clinical and functional MRI assessment of cognition has found age-independent effects of menopause on cognition, suggesting that at least some contributions to cognitive decline may have a hormonal basis.[72] This is particularly evident in short-term memory and learning, which have been shown to decrease during the late menopausal transition, but come back to near-premenopausal levels after menopause.[73] The PENN Ovarian Aging study also showed a decrement in verbal memory throughout the menopausal transition, independent of age.[74]

Epidemiological and observational studies support a cognitive benefit for hormones, but clinical trials have not supported these findings. For example, women in the Wisconsin Registry for Alzheimer's Prevention who used hormonal contraceptives had higher cognitive scores than nonusers, and the association was stronger with the duration of use.[75] In contrast, the Women's Health Initiative (WHI) memory study demonstrated a detrimental effect of estrogen + progesterone on verbal memory and a trend for positive impact on figural memory,[76] whereas estrogen alone was associated with overall worse cognition, especially among women with impaired cognitive function at baseline.[77] When the data were pooled between the estrogen + progesterone and estrogen alone trials, and the end points of dementia and mild cognitive impairment were assessed, HT overall increased the risk for dementia.[78] Given the hierarchy of evidence favoring clinical trials over observational data, HT should not currently be given with the expectation that it will prevent cognitive decline. To reconcile the divergent findings between the observational studies and randomized clinical trials, a "critical window" hypothesis has been advanced that posits that the effects of HT depend upon the timing of initiation with respect to age and the menopausal transition. While this hypothesis may be correct, a recent review strongly suggests that a definitive test of this hypothesis may not be feasible.[79] The Kronos Early Estrogen Prevention Study (KEEPS) included an exhaustive cognitive battery of tests on > 700 recently menopausal women who were randomized to transdermal or oral estrogen, both with intermittent progesterone, or to placebo and observed an overall neutral effect of hormones on cognition.[80] Henderson et al.[81] found similar neutral findings in their RCT, with a total of 567 healthy women randomly assigned to oral 17β-estradiol 1 mg/day or placebo within 6 years of menopause or 10+ years after menopause. There were no differences in executive function, verbal memory, or global cognition when estradiol was initiated within 6 years of menopause or if begun 10+ years after menopause.

Cardiovascular

The aging process is associated with predictable anatomic and physiologic alterations in the cardiovascular system. A gradual increase in the circumference of all four cardiac valves occurs with age.[82] Histologic changes detected in the vasculature include increased intimal thickness, elastin fragmentation, and increased collagen content of the arterial wall. Finally, the natural history of heart disease is adversely affected by age.[83] Atherosclerosis progression is especially prominent in late postmenopausal women. HT to prevent cardiovascular disease in menopausal patients has been extensively studied, but the data remain controversial. The WHI demonstrated a negative overall risk/benefit ratio of HT for cardiovascular protection in menopausal women; however, when the data were analyzed post-hoc by decade (50–59, 60–69, and 70–79 years), a small benefit was observed in the youngest group of women aged 50–59, HT appeared risk-neutral in women aged 60–69, and only women aged 70–79 had evidence of harm. Thus, the "timing hypothesis" has again been invoked to explain how HT may have multiple beneficial effects when administered during the early period after menopause, but not during the late postmenopausal period. The Early versus Late Intervention Trial of Estradiol (ELITE) study, a single-site RCT of 643 healthy postmenopausal women, stratified the sample according to time since menopause (early postmenopause as < 6 years or late postmenopause as ≥ 10 years). Women were randomly assigned to either oral 17β-estradiol or placebo (with intermittent progesterone or placebo progesterone). The ELITE study found that oral estradiol therapy initiated within 6 years after menopause was associated with a decreased development of subclinical atherosclerosis (measured as carotid artery intima-media thickness [CIMT]) compared with placebo but not when it was initiated ≥ 10 years after menopause.[84] The KEEPS studied > 700 menopausal women who were within 3 years of their FMP who were randomized to transdermal or oral estrogen, both with intermittent progesterone, or to placebo and did not find a difference in progression of CIMT, but also did not find any evidence for harm of HT.[80]

Some data suggest that women with menopause symptoms have a worse cardiovascular profile[85] and a recent study suggests that severe VMS may be related to increased epigenetic aging.[86] However, data from the Rancho Bernardo cohort study of 867 postmenopausal women indicates the opposite, in that women who suffer more severe hot flashes and especially night sweats had fewer cardiovascular events.[87] A lack of association between menopausal symptoms and subclinical atherosclerosis was confirmed by Wolff et al.,[88] who found that, among recently menopausal women screened for KEEPS, neither circulating estradiol nor menopausal symptoms predicted the degree of subclinical atherosclerosis. In SWAN, women with either surgical or natural menopause demonstrated decreased physical activity.[89] This is another factor that could mediate cardiovascular risk.

Estrogen-Dependent Disorders

Abnormal uterine bleeding is common during the menopausal transition, and largely attributable to the increase in anovulatory cycles and the appearance of anatomic disorders. In the Seattle Midlife Women's Health Study, the most common subjective menstrual cycle changes prior to the menopause transition included a lighter menstrual flow (32%), a heavier menstrual flow (29%), a shorter duration of flow (24%), and a longer duration of flow (20%).[90] When compared to perimenopausal women with no menstrual complaints, subjective reports of heavy flow in perimenopausal women were associated with higher levels of estradiol throughout the menstrual cycle.[91] Besides anovulatory cycles, common causes of abnormal uterine bleeding in perimenopausal women include leiomyomata and adenomyosis, emphasizing the importance of correct diagnosis to inform treatment.[92]

Leiomyomata uteri are the most common benign neoplasia of the uterine musculature in women.[93] They are clinically troublesome throughout a woman's reproductive years and typically regress and become clinically quiescent after menopause. However, for some women, the erratic estrogen levels that accompany the early menopausal transition lead to leiomyomata growth, necessitating intervention. It has also been hypothesized that local dysregulation of growth factors within leiomyoma leads to increased vasculogenesis and precipitates bleeding.[94] Polyps can also present with abnormal uterine bleeding in the transition. Relatively little is known about the natural history of these two common conditions throughout menopause. One longitudinal study of women, enrolled at a mean age of 44, indicated that 70% of women with polyps, more than twice the number of women without polyps, reported abnormal uterine bleeding.[95] The behavior of leiomyomata was more variable, with some large increases in size observed in asymptomatic women, but a general regression in size as women aged.

The menopausal transition is associated with the development of stress urinary incontinence.[96] This relationship does not appear to be attributable to estrogen withdrawal, as Waetjen et al.[97] recently found that estradiol levels had no association with the development or worsening of incontinence in women transitioning through menopause in SWAN.

DIAGNOSIS

Despite the population trend of elevated FSH and decreased estradiol with progression through the transition, measurement of FSH, inhibin, or estradiol provide at best an unreliable guide to the menopausal status of an individual woman.[29] Although there is accumulating evidence supporting AMH as a stronger predictor of the FMP than age, AFC, or ovarian volume, an ultrasensitive ELISA is required to provide a reliable index of ovarian aging in the clinical setting, and confirmation of current evidence is needed before recommending AMH as the "menopause test."[4] For most clinical circumstances in women who present with perimenopausal complaints, the most rational approach to diagnosing menopause can be based on an assessment of menstrual cycle patterns and of the longitudinal symptoms. Hormone profiles correlate well with symptoms and cycle features.[98] Thus, if a woman is > 45 years old and has had a recent disruption in her menstrual pattern and symptoms suggestive of transient hypoestrogenemia, it is likely that she has entered her menopausal transition.[55] That being said, the clinician should take care to rule out other pathologies that can be masked by common complaints associated with the menopausal transition. At minimum, a screening TSH level should be performed, as menstrual irregularity may be the only manifestation of thyroid dysfunction.[99]

Many women experience a loss of "wellness" and dysphoric mood during the menopausal transition; these syndromes should be formally distinguished from the more severe and life-threatening risk that clinical depression may pose. Cohen et al.[100] reported that women with no lifetime history of depression, who enter the menopausal transition earlier, have a significant risk for a new onset of major depression as compared with women who remain premenopausal. Analysis of women participating in SWAN revealed that 15.8% of participants with no history of major depressive disorder developed symptoms of major depression over the course of 7 years. Lifetime history of an anxiety disorder and a very stressful life event prior to depression onset also predicted a first episode of major depression.[101]

Atypical hot flashes can be due to non–estrogen-related causes, including new-onset diabetes with autonomic dysfunction, carcinoid, pheochromocytoma, and medication side effects.[102] Women who appear to be entering the menopausal transition earlier than 40 years of age should have a more extensive workup, including screening appropriate for patients with POI.[103]

TREATMENT

Hormone Therapy Overview

It is important to separate short-term and long-term therapeutic goals with patients when discussing treatment options during and after the menopausal transition because short-term (<5 years) and long-term (≥5 years) benefits and harms differ.[104,105] It is challenging for both physicians and patients to interpret the literature, which is often based on aggregate epidemiologic, observational, or clinical trial data, and to apply it to the individual to counsel appropriately. The role of short-term therapy for symptom relief is considerably less complicated because the benefits are immediate and often large.

Some women go through their menopausal transition without notice, others are severely affected, and most have symptoms that are somewhat bothersome. Distinguishing symptoms related to the menopause transition versus those related to the normal aging process is a helpful first step in determining the usefulness of HT. In general, VMS, sleep disturbances, and vaginal symptoms show the strongest and most consistent association with the transition into menopause, and mounting supportive evidence suggests that depression and sexual dysfunction may be independently related to the menopause as well.[106] Choosing the best treatment option for an individual woman involves knowledge of the risks and benefits of each treatment.

The North American Menopause Society (NAMS) describe the comprehensive benefits compared with risks of HT for both treatment of menopause-related symptoms and its consideration for longer-term use based on the time since the FMP.[106] According to the statement, "HT remains the most effective treatment for VMS and the genitourinary syndrome of menopause (GSM) and has been shown to prevent bone loss and fracture. The risks of HT differ depending on type, dose, duration of use, route of administration, timing of initiation, and whether a progestogen is used. Treatment should be individualized to identify the most appropriate HT type, dose, formulation, route of administration, and duration of use, using the best available evidence to maximize benefits and minimize risks, with periodic reevaluation of the benefits and risks of continuing or discontinuing HT. For women aged younger than 60 years or who are within 10 years of menopause onset and have no contraindications, the benefit-risk ratio is most favorable for treatment of bothersome VMS and for those at elevated risk for bone loss or fracture. For women who initiate HT more than 10 or 20 years from menopause onset or are aged 60 years or older, the benefit-risk ratio appears less favorable because of the greater absolute risks of coronary heart disease (CHD), stroke, venous thromboembolism (VTE), and dementia."[106] NAMS as well as other organizations that have issued guidelines on HT (the American College of Obstetricians and Gynecologists and the Endocrine Society) endorse non-oral estrogen to minimize the risk for VTE.[105]

KEY POINTS

- Hormone therapy (HT) remains the most effective treatment modality for severe menopausal symptoms.

Vasomotor Symptoms

Behavioral techniques can be beneficial for mild-moderate hot flashes, but none are evidence-based. Encouraging patients to wear layered clothing that can be added or removed easily can help a woman maintain control of her core body temperature. Changes in dietary habits such as avoiding spicy foods and reducing caffeine and excess alcohol intake, which can cause flushing, can be beneficial. Patients should be told that regular physical activity, a balanced diet, and stress reduction (particularly paced breathing) may be of additional help in decreasing vasomotor flushing.

HT remains the most effective pharmacologic option for the treatment of VMS. More than 40 RCTs of oral and transdermal estrogen have reported a reduction in the severity of VMS, and estrogen was effective in doses lower than the usual 0.625 mg of equine estrogen or equivalent. Transdermal estradiol, transdermal estradiol spray, as well as gels, vaginal rings, and intranasal 17β-estradiol spray are all approved by the Food and Drug Administration (FDA) with demonstrated efficacy in reducing hot flashes. A systematic review of 32 included trials comparing conjugated equine estrogens (CEE) versus 17β-estradiol (oral and transdermal) for reducing hot flashes found them to be comparable in effectiveness.[107]

To optimize compliance and patient satisfaction, selection of the most appropriate regimen of HT (dose, route of administration, and schedule) for the needs of each individual is important. The risk for endometrial hyperplasia and cancer, associated with unopposed estrogen use in perimenopausal and postmenopausal women who have a uterus, is decreased by supplementing the patient with progestogen. Progestogen can be administered orally, transdermally, or vaginally, or even intrauterine, via a levonorgestrel-impregnated intrauterine device (LNG-IUD).[108]

The largest systematic review and meta-analysis to date of double-blind, placebo-controlled RCTs of non-hormonal therapies available for the treatment of VMS showed supportive evidence for the efficacy of selective serotonin reuptake inhibitors (SSRIs), clonidine, and gabapentin in reducing the frequency and severity of menopausal hot flashes.[109] On average, SSRIs provide a 60% reduction in hot flash frequency and severity, but carry significant side effects.[110] Gabapentin, at doses between 300 and 900 mg/day, reduced hot flashes by as much as 45% in two double-blind, placebo-controlled RCTs, with a chief limiting side effect of sedation.[109] Both oral and transdermal clonidine may be beneficial for some patients, but can be associated with postural hypotension.[109] Depot medroxyprogesterone acetate (MPA),[111] oral megestrol acetate,[112] and micronized progesterone at a 300-mg nightly dose are effective in the treatment of severe VMS.[113]

The tissue-selective estrogen complex which combines a selective estrogen receptor modulator (SERM) with estrogen is a relatively new treatment for menopausal symptoms. A combination of a daily dose of 20 mg of the SERM bazedoxifene (BZA) with 0.45 mg or 0.625 mg of CEE is a non-progestogen therapy that has been approved for the prevention of osteoporosis and treatment of VMS. The SMART trial, a double-blinded RCT, showed a low incidence (<1%) of endometrial hyperplasia with BZA/conjugate estrogen (CE), and the BZA/CE group showed significantly greater increases in LS and total hip BMD than placebo.[114] This new form of treatment can be considered in patients with progestin intolerance or those with breakthrough bleeding unresponsive to hormonal manipulation.

On the horizon for treatment of VMS are NK3R antagonists which are currently in clinical trial. NK3R activation stimulates the thermoregulatory center in the hypothalamus which contributes to VMS of menopause. Early data has had favorable results showing promise as a future non-hormonal treatment for VMS of menopause.[115]

Urogenital Symptoms (GSM)

Treatment of vulvovaginal atrophy should focus on relieving atrophic anatomic changes resulting from low estrogen levels. Over-the-counter vaginal lubricants and moisturizers are commonly used as a first-line treatment of vaginal atrophy–related symptoms.

Estrogen by any route, whether oral, transdermal, or vaginal, improves urogenital symptoms. Topical estrogens (cream, tablet, ring, or insert) avoid the higher circulating levels of estradiol found with systemic preparations, and appear to be safe for long-term use.[116] Topical estrogen can relieve sexual dysfunction if dyspareunia is the underlying deterrent.

Data regarding the influence of HT on urinary incontinence symptoms are mixed. A Cochrane Database meta-analysis, which covered a total of 19,313 incontinent women among whom 9417 were treated with HT in 33 studies, reported worsening urinary incontinence symptoms for systemic HT. However, topical vaginal estrogen was associated with improvement in urinary incontinence primarily by reducing urinary urge symptoms.[117]

SERMs have been studied as treatments for vaginal atrophy. Ospemifene at a 60-mg dose was effective and well tolerated for vaginal dryness and dyspareunia and has no demonstrable proliferative effect on the endometrium.[118]

The prevalence of vaginal dryness from participants in the SWAN study ranged from 30% to 40%,[43] which is similar to most studies. One of the main consequences of vaginal dryness is dyspareunia which can contribute to a decrease in sexual desire over the menopause transition.[119,120] However, the relationship between sexual dysfunction and hormone levels, specifically estrogens and androgens, is relatively weak, although consistent across studies. The most dramatic change in androgen levels is seen after surgical menopause rather than natural menopause. RCTs of testosterone have provided evidence that low sexual desire can be improved by use of supraphysiologic androgen supplementation.[121,122] Although short-term safety appears reassuring, long-term safety data are not available, there is a lack of consensus who represents the ideal candidate for therapeutic intervention, and there are no FDA-approved formulations of testosterone appropriate for women at the present time.[123]

A prospective, phase III RCT suggested a significant benefit of vaginally administered DHEA on sexual desire, arousal, orgasm, and pain at sexual activity.[124] Daily intravaginal administration of 0.50% (6.5 mg) DHEA reduces the percentage of parabasal cells, increases the percentage of superficial cells, and reduces vaginal pH, all consistent with improved vaginal health.[125]

Women who present with specifically sexual complaints should undergo a thorough evaluation into their exact nature and potential causes, including medication usage, psychological stressors, and comorbidities. Cognitive behavioral therapy has demonstrated effectiveness in clinical trials, but at present the two neuroactive, FDA-approved compounds for treatment of low sexual desire in women, flibanserin and bremelanotide, are not approved for use in postmenopausal women.

Sleep Disturbances

Sleep disturbances are often treated in a multifactorial manner, as HT alone might be insufficient for improving sleep quality in perimenopausal women. Cognitive behavioral interventions, such as stimulus control, sleep hygiene education, and relaxation training can be very helpful in treatment of insomnia. Empiric HT, along with behavioral techniques, may provide short-term support, with the understanding that sleep should improve after the menopausal transition.[126,127]

Adverse Mood

Depressive and/or worsening anxiety symptoms are experienced among 15%–50% of women during the menopausal transition, and fluctuations in hormones are thought to contribute. Several studies have evaluated the effectiveness of HT on menopausal mood symptoms. A recent review evaluating HT and mood in perimenopausal and postmenopausal women concluded that estrogen was effective in improving mood in perimenopausal women, but the progestogenic component in combined HT could potentially offset the beneficial influence of estrogens by inducing adverse mood symptoms. HT is generally considered to be useful along in mild-to moderate depression, but in more severe cases of depression, a combination of antidepressant and HT must be considered.[128] The ancillary Cognitive and Affective Study (KEEPS-Cog) of the KEEPS examined the effects of up to 4 years of HT on cognition and mood in recently postmenopausal women and found that women treated with oral CEE plus micronized progesterone had a decrease in anxiety and depression symptoms over the 48 months of treatment compared with those taking placebo.[129] Transdermal estradiol at a dose of 100 mcg/day for 12 months prevented clinical depression and reduced depressive symptoms in perimenopausal and postmenopausal women.[130]

Other Common Issues

Menstrual Irregularities. Menstrual irregularities are common during the menopausal transition and are used to mark its progression. Overall, menstrual bleeding tends to become initially more frequent, and then becomes progressively less frequent and lighter as women go from the early to the late transition.[131] Women who report persistent, heavy bleeding require a gynecologic evaluation for anatomic pathology, as uterine polyps, fibroids, or adenomyosis may be present and require surgical management. The unpredictability of bleeding episodes can present an inconvenience for women. Management options include cycling with oral contraceptive pills (OCPs) as long as there are no contraindications (i.e., smokers, migraines with aura[132]). Birth control pills (or patches or rings) are effective for regularization of withdrawal bleeding and can reduce mood lability, while also providing contraception. Women who use hormonal contraception during the menopausal transition may experience adverse symptoms during the pill-free interval of 7 days, and limiting or eliminating the pill-free interval may result in improved symptom profiles.[133] Another option includes adding a low dose of transdermal (i.e., 25 to 50 mcg) or oral estradiol during the pill-free week. In women who do not need contraception, HT can be considered. With combined continuous administration of HT, breakthrough bleeding may occur. With cyclic regimens, women will continue to experience monthly bleeding but breakthrough bleeding may be minimized. In women whose chief complaint is menstrual irregularities, LNG-IUD can decrease menstrual flow as long as there is no uterine pathology.[134] Estrogen-only HT can be used with an LNG-IUD.

Complementary and Alternative Medicine in the Treatment of Menopausal Symptoms

Estimates on the prevalence of the use of complementary and alternative medicine (CAM) in women aged 45 to 54 years are 40% or more, based on the SWAN cohort, with many women using more than one method. Ethnic differences are seen in their usage, with Black American women less likely to use CAM, and usage of specific therapies also differs significantly by ethnic background.[135] Some popular treatments are listed below with limited data supporting their use.

In one small (n = 51) study of soy/isoflavone supplementation, women given a soy protein supplement with 40 mg of protein and 76 mg of isoflavones had a 45% reduction in VMS, compared with 30% in the placebo group.[136] It is important to note that soy supplementation is required to achieve the high levels of isoflavones observed in participants in the latter study; Western diets do not normally contain foods rich in these substances. Although black cohosh continues to be marketed in various forms, a large, well-conducted RCT did not demonstrate efficacy.[137] Evening primrose oil (gamma linolenic acid [GLA]) was not found to be superior to placebo in the treatment of menopausal VMS.[138] Finally, an RCT of 4.5 g of dong quai daily versus placebo observed a 25% reduction in hot flashes.[139] A review of the literature related to specific forms of CAM found the overall evidence to be mixed, of low quality, and insufficient to allow specific recommendations.[140] However, given the relatively large placebo effect size seen in most RCTs for hot flashes, clinicians and patients may wish to balance the practicalities of cost versus potential for harm when considering such treatments.

Long-Term Use for Prevention of Chronic Conditions

Bone. Many factors may increase a woman's risk for bone fracture. Non-modifiable risk factors include age, family history, personal history of osteoporotic fracture, and early menopause. Modifiable risk factors include weight (increased risk in thin women), excess alcohol consumption, current smoking, low calcium intake, vitamin D deficiency, sedentary lifestyle, and low bone density. Based on the National Academy of Science, Engineering and Medicine (NASEM) recommendations released in 2011, calcium supplementation should be given up to a dose of 1000 mg/day in women aged 19 to 50 years and 1200 mg/day in women aged > 50 years who are not taking hormones. Ideally, calcium should be taken via consumption of calcium-rich foods; if supplementation is required based on dietary intake, calcium supplements should be taken with a meal to optimize absorption. The vitamin D recommended daily allowance is 600 IU, although individuals aged > 70 years may need more. NASEM recommends a serum vitamin D level no lower than 20 ng/mL (50 nmol/L) for good bone health because that is the level required by 97.5% of the population; however, routine screening of vitamin D levels is not recommended.[141]

Weight-bearing exercise stimulates osteoblasts to form new bone.[142] In menopausal women, weight-bearing exercise for 22 months led to a 6.1% increase in bone density of the LS[143]; the benefits of exercise have been demonstrated into the ninth decade of life and persist for as long as it is continued.[144] Because lifestyle modifications have been shown to improve bone density in young women and to prevent fracture in older women, they are inexpensive and safe treatment recommendations and should be included in every routine health maintenance examination.

Pharmacotherapy is appropriate for women who are at high risk for fracture in the future and in women with established osteoporosis. The fracture risk assessment tool (FRAX), developed in collaboration with the WHO, can be used to predict a 10-year risk for osteoporotic fracture.[145] FRAX has been widely used to aid clinical decision making regarding treatment initiation when the patient's bone density indicates osteopenia. Treatment should be considered when there is a 3% risk for hip fracture and/or a 20% risk for a major osteoporotic fracture, in the next 10 years. Clinical risk factors used in FRAX include, age, sex, BMI, previous fragility fracture, parental hip fracture, current smoking status, corticosteroid use, alcohol intake ≥ 3 units (a unit of alcohol is equivalent to a glass of beer [285 mL], an ounce [30 mL] of spirits, or a medium-sized glass of wine [120 mL]) per day, rheumatoid arthritis, and other secondary causes of osteoporosis (Table 122.2).[146] The list of available preventive and treatment options is presented in Table 122.3.

Estrogen has both a preventive and therapeutic role for hypoestrogenic premenopausal women with osteoporosis because such women are young enough that they may build new bone if their estrogen deficiency is corrected. After menopause, however, estrogen's antiresorptive

TABLE 122.2 Risk Factors Used in Fracture Risk Assessment (FRAX) Tool

- Age
- Sex
- Weight
- Height
- Previous fracture
- Parent fractured hip
- Current smoking
- Glucocorticoid use
- Rheumatoid arthritis
- Secondary osteoporosis
- Alcohol ≥ 3 units/day
- Femoral neck bone mineral density

TABLE 122.3 Agents Used in Prevention/Treatment of Osteoporosis

Agent	Prevention/Treatment	Site of Action
Bisphosphonates	Prevention + treatment	Vertebrae and hip
Selective estrogen receptor modulators	Prevention + treatment	Vertebrae
Calcitonin	Treatment (>5 years postmenopause)	Vertebrae
Parathyroid hormone	Treatment	Vertebrae and hip
RANK ligand inhibitor	Treatment	Vertebrae and hip
Estrogens	Prevention	

action will only prevent further bone loss, as demonstrated by many RCTs. No HT product currently has government approval for the treatment of osteoporosis; however, most are approved for osteoporosis prevention. Observational studies and RCTs have demonstrated reductions in fractures of the vertebrae and possibly the hip in women on HT. A meta-analysis of 22 trials of estrogen reported an overall 27% reduction in non-vertebral fractures, while taking into account that the quality of individual studies varied.[147] The Heart and Estrogen/Progestin Replacement Study (HERS) and HERS II found no reduction in hip, wrist, vertebral, or total fractures with HT, but the follow-up interval was relatively short.[148] The WHI is the first RCT to demonstrate reduction of hip fracture risk with estrogen use[149]; risk reductions were also reported for vertebral and other osteoporotic fractures. The US Preventive Services Task Force concluded that good evidence suggests that HT prevents bone loss, and fair to good evidence indicates that it reduces fractures.[150] Women experiencing an early menopause who require prevention of bone loss are probably best served by the administration of HT or oral contraceptives, rather than bone specific treatments, until they reach the median age of natural menopause (51.4 years), at which time treatment needs may be reassessed.[106]

For postmenopausal women with established osteoporosis without a symptom indication for HT, non-hormonal modalities are available. Alendronate, a bisphosphonate that is a powerful antiresorptive agent and reduces fractures in menopausal women with osteoporosis, has been approved by the US FDA for the prevention (5 mg daily) and treatment (10 mg daily or 70 mg weekly) of osteoporosis. Risedronate is more potent and has fewer upper gastrointestinal side effects than alendronate; it reduces the incidence of fracture in osteoporotic women. Raloxifene is the only SERM approved for long-term treatment in the prevention of osteoporotic fractures in postmenopausal women. It has been demonstrated to reduce vertebral fractures.[151] Denosumab is a human monoclonal antibody to the receptor activator of nuclear factor-κB (NF-κB) ligand, which binds to the receptor activator of NF-κB on the surface of osteoclasts blocking proliferation and differentiation of the osteoclasts. Approved in 2010 for treatment of postmenopausal women with osteoporosis who are at high risk for fracture, it is administered subcutaneously every 6 months.[152] Finally, recombinant parathyroid hormone, administered subcutaneously, daily, is approved for the treatment of osteoporosis in postmenopausal women who are at high risk for fractures, and can increase BMD substantially.[153]

Cardiovascular Disease. Observational studies have long suggested that combined HT is beneficial for the primary prevention of heart disease.[154] The WHI, a primary prevention trial of 16,608 healthy postmenopausal women after 5.2 years of daily HT or placebo, initially reported a small but significantly increased risk for nonfatal heart attacks. Although CHD mortality was not increased, the risk for heart disease was 29% higher for women taking combined HT than for those on placebo.[149] The WHI findings do not support the use of HT for primary prevention of heart disease in women.

The HERS indicated that women with known CHD who were randomized to CEE and MPA had a 52% increased risk for myocardial infarction during the first year of use and experienced increased CHD deaths during the first 3 years of use.[155] After 6.8 years of follow-up in HERS II, no difference in primary or secondary CHD events was reported among estrogen users.[156] Thus, for secondary prevention of heart disease, no clinical benefit is derived from HT. Furthermore, the American Heart Association has long recommended discontinuing treatment after an acute event in a woman who is currently using HT.[157]

Since the publication of the two major primary and secondary prevention trials, multiple observational cohorts and smaller RCTs have been performed to attempt to reconcile the discrepancy between the apparent benefit of HT in cohort studies and a neutral to harmful outcome in RCTs. Unfortunately, the divide has only deepened with time, with further observational studies such as the Danish Osteoporosis Prevention Study supporting a benefit of HT,[158] despite no clear benefit demonstrated from randomized clinical trials.

The "timing hypothesis" seeks to explain this discrepancy by theorizing that HT has multiple long-term beneficial effects when initiated before or during the early period after menopause, but not during the late postmenopausal period (i.e., more than 10 years after FMP). The ELITE study, a single-site RCT of 643 healthy postmenopausal women, stratified the sample according to time since menopause (early postmenopause as < 6 years or late postmenopause as ≥ 10 years). Patients were randomly assigned to either oral 17β-estradiol or placebo (with intermittent progesterone or placebo progesterone). ELITE found that oral estradiol therapy initiated within 6 years after menopause was associated with a decreased development of subclinical atherosclerosis (measured as CIMT) than placebo but not when it was initiated ≥ 10 years after menopause.[84] At the same time, KEEPS, a multicenter RCT of 727 women, was performed to assess the effect of HT in women who started within 3 years of FMP. Treatment arms included (1) oral conjugated equine estrogens (O-CEE) 0.45 mg/day, (2) transdermal estradiol (t-E2) 50 mcg/day, and (3) placebo. Active arms received cyclical micronized progesterone 200 mg/day for 12 days/month, with a progesterone placebo for the placebo arm. After 48 months of follow-up, O-CEE and t-E2 had no adverse effects on atherosclerosis progression or coronary calcium. O-CEE and t-E2 had neutral effects on blood pressure and mainly favorable or neutral effects on CHD biomarkers. These findings support a lack of harm, and imply the potential for benefit to early intervention with HT for symptomatic women, but they are a far cry from supporting widespread preventive use of HT.

Aside from heart disease, HT may play a role in recurrent thrombotic events. The WHI reported an increased risk for stroke after 5 years of HT use[149] and, in agreement with previous studies, a higher risk for thromboembolism in the first year of HT use, which diminishes thereafter to approximately a 2-fold increased risk.[149] Follow-up analysis of the WHI that examined both coronary and venous thromboembolic events showed that the increased risk for both events in the combined HT group subsided over the 2.4-year postintervention follow-up period.[159] Taking all of the data into account, guidelines on HT have evolved in their recommendations to include the case for "extended use." The American College of Obstetrics and Gynecologists,[105] the NAMS,[106] and the American Heart Association[160] continue to appropriately recommend against the use of HT for primary or secondary prevention of cardiovascular disease. Because cardiovascular disease is a more common cause of morbidity and mortality for women than are osteoporosis and cancer combined, efforts to identify patients at risk and recommendations on lifestyle modifications, diet, and other pharmacotherapy (i.e., beta blockers, aspirin, angiotensin-converting enzyme inhibitors, statins) are more beneficial and probably less harmful than HT for long-term treatment. However, for women who remain symptomatic and are unable to utilize non-hormonal alternatives to HT, longer-term use can and should be considered using shared decision making that takes into account the patient's perceptions of benefit.

Cognition. Similar to the data on cardiovascular disease and HT, observational studies demonstrate a benefit of hormones, whereas RCTs demonstrate a neutral to negative effect. Thus, the timing of initiation of HT may be relevant for both cardiovascular and cognitive outcomes. While this hypothesis may be correct, a recent review by Maki[79] strongly suggests that a definitive test of this hypothesis may not be feasible. KEEPS included an exhaustive cognitive battery of tests on > 700 recently menopausal women randomized to transdermal or oral estrogen, both with intermittent progesterone, or to placebo, and found an overall neutral effect of hormones on cognition.[80] KEEPS is in the process of repeating the cognitive battery on participants from the original cohort and may be able to address the timing hypothesis. Henderson et al.[81] found similar neutral findings in their RCT, with a total of 567 healthy women randomly assigned to oral 17β-estradiol 1 mg/day or placebo within 6 years of menopause or 10+ years after menopause. There were no differences in executive function, verbal memory, or global cognition when estradiol was initiated within 6 years of menopause or if begun 10+ years after menopause.

Primary Ovarian Insufficiency and Early Menopause. POI can result from genetic factors, medical conditions, iatrogenic causes, or idiopathic factors. All patients diagnosed with POI (excluding iatrogenic causes) should be evaluated with a karyotype, adrenal antibodies, *FMR1* premutation, pelvic ultrasound, and bone density.[161] Patients with idiopathic POI have a higher prevalence of autoimmune disorders, and all should be screened regularly for thyroid dysfunction, diabetes, and adrenal insufficiency. Ongoing screening should continue even when initial screening is negative as these conditions can present later. Other conditions such as pernicious anemia, myasthenia gravis, rheumatoid arthritis, and systemic lupus erythematous are associated with POI, and should be screened with onset of symptoms.[103] HT is particularly beneficial for patients with POI or early menopause that occurs from either natural or surgical causes. These patients are at increased risk for osteoporosis, adverse cardiovascular effects, neurological disease, impaired sexual function, and psychiatric conditions, and the risk of poor outcomes increases with earlier age of loss of ovarian function.[162] Estrogen replacement can prevent many of these conditions or reduce symptom severity,[162] and it should be initiated immediately upon the diagnosis of POI and continued until the natural age of menopause for the highest benefit.[161] HT should attempt to mimic physiological estradiol levels with estradiol-progestin preferred over OCPs due to a higher risk of VTE and less proven benefits of bone health in OCPs. It is important to remember that younger patients with POI have a 5%–10% chance of spontaneous conception, and patients should be counseled appropriately about this risk. Those not interested in pregnancy should be on some form of contraception since HT alone does not prevent pregnancy.[161]

Cancer Risks and Mitigation of Risk With Hormone Therapy

Breast Cancer. In the WHI study, an additional 8 breast cancer cases per 10,000 women taking continuous combined CEE + MPA per year was observed, with a HR of 1.24 (95% CI 1.01-1.53) after a median 5.6 years of randomized treatment.[163] Somewhat surprisingly, a reduced risk of breast cancer was observed in women who had a hysterectomy in the CEE-alone arm.[164] The combined lack of cardioprotection and potential risk of breast cancer associated with HT led to widespread discontinuation of its use. Unfortunately, the reassuring message of reduced risk of breast cancer in association with estrogen-only therapy was essentially unheard by the public. Yet, the findings raise fundamental questions about the biology of breast cancer and how hormone exposure contributes to risk. Women who were HT naïve when randomized in the WHI had an overall similar breast cancer incidence whether they took CEE+MPA therapy or placebo for up to 11 years of follow-up; women who had previously taken HT had an overall lower incidence of breast cancer when they were randomized to placebo.[165] The most recent WHI data containing more than 20 years of median follow-up confirms that women randomized to CEE+MPA had a higher long-term risk of breast cancer, but no increase in breast cancer mortality, and that women randomized to CEE alone had a lower risk of both breast cancer and breast cancer mortality.[166] Two more recent RCTs that initiated HT within 6 years of FMP, with a primary outcome being the effects of HT on cardiovascular disease, showed no increased risk of breast cancer compared with placebo; however, these trials were not sufficiently powered to provide definitive data for a breast cancer outcome.[84,158] The Collaborative Group on Hormonal Factors in Breast Cancer's recent meta-analysis found that all forms of HT, excluding vaginal estrogen, were associated with an increased risk of breast cancer, and this risk progressively increased with the duration of HT use and was higher for estrogen plus progestin than estrogen alone. These excess risks were detectable even during years 1–4 of use (estrogen-progestogen RR 1.60, [95% CI 1.52–1.69]; estrogen-only RR 1.17, [1.10–1.26]), and increased incrementally with longer duration of use (years 5–14 estrogen-progestogen RR 2.08, [2.02–2.15]); estrogen-only RR 1.33, [1.28–1.37]).[167] The French E3N-EPIC cohort study found an increased risk of breast cancer in postmenopausal women on HT compared with non-users (RR 1.2 [1.1–1.4]), with most risk attributable to HT containing synthetic progestins compared with micronized progesterone, the RRs being 1.4 [1.2–1.7] and 0.9 [0.7–1.2], respectively.[168] For women without a uterus who do not need a progestin, the overall data favor that HT with estrogen alone is risk neutral to minimal. However, for women who require a progestin to prevent endometrial hyperplasia, breast cancer risk appears to increase with duration of use, and this risk may be mitigated by use of micronized progesterone, but currently there are no randomized clinical trials to support this notion. Breast cancer risk is a serious adverse effect and therefore counseling is recommended to be conservative (i.e., reliant on the best available RCT data) and to involve shared decision making. Although the overall risk of breast cancer attributable to menopausal HT with E+P equates to less than 1 additional case per 1000 woman-years of use,[169] risk accumulates with duration of use and is one of the principal drivers of the dictum: "lowest possible dose for shortest possible time."

KEY POINTS

- The current recommendation is to use HT in the lowest effective dose for the shortest time needed to control symptoms.

Other Cancers. Unopposed estrogen is a well-known risk factor for endometrial hyperplasia and endometrial cancer, and the major method to protect the endometrium in women taking HT with an intact uterus is the addition of progestin. Multiple studies have found an increased risk of endometrial cancer among users of ET only compared with age-matched controls,[170] but this risk does not accrue to women using vaginal estrogen alone.[116] Concomitant estrogen plus progestin (or BZA) is required for women with a uterus who are taking systemic HT. Progestin may be given either continuously or sequentially. Sequential use may be more palatable for women who have adverse effects from progestin use, but it will result in continued menstruation. Sequential combined estrogen plus progestin therapy is also less effective in eliminating the risk of endometrial cancer compared with continuous combined estrogen plus progestin, which decreases endometrial cancer risk below that of never users.[170]

In regards to ovarian cancer risk, several meta-analyses[171,172] have shown an increased risk of ovarian cancer associated with HT use. The pooled RR of ovarian cancer was 1.29 (95% CI 1.19–1.40), and increased risk was only found for the two common types, serous and endometrioid tumors. A recent prospective Swedish population–based matched–cohort study included all women aged ≥ 40 years having used systemic HT between 2005 and 2012 compared with non-users. These investigators found that current estrogen plus progestin HT use correlated with a moderately higher risk of ovarian cancer (OR = 1.38, 95% CI 1.18–1.62), with no increased risk among past users, and a reduced risk was observed for estrogen-only use (OR = 0.25, 95% CI 0.22–0.29).[173] These findings are in direct contradiction of those of the WHI, in which E alone but not E+P, was associated with increased risk of ovarian cancer. Current data regarding association of HT and ovarian cancer is therefore somewhat unclear. Incorporation of a woman's background risk of ovarian cancer is recommended to be part of the risk/benefit discussion of menopausal HT for symptomatic women.

Risk of VTE in patients taking HT is another potential risk, and has been evaluated in many studies, including the WHI. Women randomized to combined E+P HT had double the risk of VTE compared with those randomized to placebo; this risk fell to background levels shortly after discontinuation of HT. Many observational studies have reported similar results in women taking oral estrogen, but no elevated risk of VTE has been reported in patients taking transdermal estrogen.[106] Lower doses of estrogen, and the use of oral estradiol may reduce or avert the increased risk of VTE,[114] and use of vaginal estrogen has not been shown to increase the risk of VTE.[116]

SUMMARY

In conclusion, menopausal HT remains the most effective treatment modality for severe menopausal symptoms. In addition, it protects against bone resorption and improves vaginal atrophy. It is beneficial for adverse mood and may be helpful for sleep problems, particularly during the transition. Current recommendations are to use the lowest effective dose for the shortest time needed to control the symptoms. Long-term outcomes of women from the WHI who have now been followed for over 20 years[174] confirm the overall safety of this intervention for a mean period of use of 5 to 7 years. Ongoing work will help increase understanding of the mechanisms by which menopausal changes cause symptoms, refine our ability to identify women at risk for more severe and prolonged symptoms, and target hormonal and non-hormonal treatments that will be most effective for maintaining the vigor of women traversing this reproductive milestone.

REFERENCES

1. Hawkes K, O'Connell JF, Jones NB, et al. Grandmothering, menopause, and the evolution of human life histories. *Proc Natl Acad Sci U S A*. 1998;95:1336–1339.
2. Sommer B, Avis N, Meyer P, et al. Attitudes toward menopause and aging across ethnic/racial groups. *Psychosom Med*. 1999;61:868–875.
3. Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop+ 10: addressing the unfinished agenda of staging reproductive aging. *J Clin Endocrinol Metab*. 2012;97:1159–1168.
4. Finkelstein JS, Lee H, Karlamangla A, et al. Antimullerian hormone and impending menopause in late reproductive age: the Study of Women's Health across the Nation. *J Clin Endocrinol Metab*. 2020;105:e1862–e1871.
5. Gold EB, Crawford SL, Avis NE, et al. Factors related to age at natural menopause: longitudinal analyses from SWAN. *Am J Epidemiol*. 2013;178:70–83.
6. Frere G. Mean age at menopause and menarche in South Africa. *S Afr J Med Sci*. 1971;36:21.
7. Bromberger JT, Matthews KA, Kuller LH, et al. Prospective study of the determinants of age at menopause. *Am J Epidemiol*. 1997;145:124–133.
8. García VA, Nava L, Malacara J. The age of menopause in the urban population of the city of Leon. *Guanajuato. Rev Invest Clin*. 1987;39:329.
9. Tamada T, Iwasaki H. Age at natural menopause in Japanese women. *Maturitas*. 1996;24:122.
10. Ismael NN. A study on the menopause in Malaysia. *Maturitas*. 1994;19:205–209.
11. Golezar S, Ramezani Tehrani F, et al. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric*. 2019;22:403–411.
12. SM M, Brambilla DJ, Posner JG. The normal menopause transition. *Maturitas*. 1992;14:103–115.
13. Nair S, Slaughter JC, Terry JG, et al. Anti-mullerian hormone (AMH) is associated with natural menopause in a population-based sample: the CARDIA Women's Study. *Maturitas*. 2015;81:493–498.
14. Freeman EW, Sammel MD, Lin H, et al. Contribution of the rate of change of antimüllerian hormone in estimating time to menopause for late reproductive-age women. *Fertil Steril*. 2012;98:1254–1259. e2.
15. Snieder H, MacGregor AJ, Spector TD. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J Clin Endocrinol Metab*. 1998;83:1875–1880.
16. van Asselt KM, Kok HS, Pearson PL, et al. Heritability of menopausal age in mothers and daughters. *Fertil Steril*. 2004;82:1348–1351.
17. Genetics of reproductive aging from gonadal dysgenesis through menopause. In: Desai S, Rajkovic A, eds. *Seminars in Reproductive Medicine*. Thieme Medical Publishers; 2017.
18. Weel AE, Uitterlinden AG, Westendorp IC, et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. *J Clin Endocrinol Metab*. 1999;84:3146–3150.
19. França MM, Mendonca BB. Genetics of primary ovarian insufficiency in the next-generation sequencing era. *J Endocr Soc*. 2020;4:bvz037.
20. Mattison D, Thorgeirsson S. Smoking and industrial pollution, and their effects on menopause and ovarian cancer. *Lancet*. 1978;311:187–188.
21. Thomas-Teinturier C, El Fayech C, Oberlin O, et al. Age at menopause and its influencing factors in a cohort of survivors of childhood cancer: earlier but rarely premature. *Hum Reprod*. 2013;28:488–495.
22. Cramer DW. Epidemiologic aspects of early menopause and ovarian cancer. *Ann N Y Acad Sci*. 1990;592:363–375.
23. Harlow BL, Wise LA, Otto MW, et al. Depression and its influence on reproductive endocrine and menstrual cycle markers associated with perimenopause: the Harvard Study of Moods and Cycles. *Arch Gen Psychiatry*. 2003;60:29–36.
24. Strauss III JF, Williams CJ. *Ovarian Life Cycle. Yen and Jaffe's Reproductive Endocrinology*. Philadelphia, PA Elsevier; 2019:167–205. e9.
25. Johnson J, Skaznik-Wikiel M, Lee H-J, et al. Extra view setting the record straight on data supporting postnatal oogenesis in female mammals. *Cell Cycle*. 2005;4:1471–1477.
26. Niikura Y, Niikura T, Tilly JL. Aged mouse ovaries possess rare premeiotic germ cells that can generate oocytes following transplantation into a young host environment. *Aging (Albany NY)*. 2009;1:971.
27. Kallen A, Polotsky AJ, Johnson J. Untapped reserves: controlling primordial follicle growth activation. *Trends Mol Med*. 2018;24:319–331.
28. Burger HG, Dudley EC, Robertson DM, et al. Hormonal changes in the menopause transition. *Recent Prog Horm Res*. 2002;5:257–275.
29. Burger HG, Hale G, Robertson D, et al. A review of hormonal changes during the menopausal transition: focus on findings from the Melbourne Women's Midlife Health Project. *Hum Reprod Update*. 2007;13:559–565.
30. Visser JA, Themmen AP. Role of anti-Müllerian hormone and bone morphogenetic proteins in the regulation of FSH sensitivity. *Mol Cell Endocrinol*. 2014;382:460–465.
31. Couzinet B, Meduri G, Lecce MG, et al. The postmenopausal ovary is not a major androgen-producing gland. *J Clin Endocrinol Metab*. 2001;86:5060–5066.
32. Zumoff B, Strain GW, Miller LK, et al. Twenty-four-hour mean plasma testosterone concentration declines with age in normal premenopausal women. *J Clin Endocrinol Metab*. 1995;80:1429–1430.
33. Burger HG, Dudley EC, Cui J, et al. A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition. *J Clin Endocrinol Metab*. 2000;85:2832–2838.
34. Rannevik G, Jeppsson S, Johnell O, et al. A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG and bone mineral density. *Maturitas*. 1995;21:103–113.

35. Bromberger JT, Kravitz HM, Chang Y, et al. Does risk for anxiety increase during the menopausal transition? Study of Women's Health Across the Nation (SWAN). *Menopause*. 2013;20:488.
36. Gold EB, Colvin A, Avis N, et al. Longitudinal analysis of the association between vasomotor symptoms and race/ethnicity across the menopausal transition: study of Women's Health across the Nation. *Am J Public Health*. 2006;96:1226–1235.
37. Crandall CJ, Manson JE, Hohensee C, et al. Association of genetic variation in the tachykinin receptor 3 locus with hot flashes and night sweats in the Women's Health Initiative Study. *Menopause*. 2017;24:252.
38. Kronenberg F, Downey JA. Thermoregulatory physiology of menopausal hot flashes: a review. *Can J Physiol Pharmacol*. 1987;65:1312–1324.
39. Williams RE, Kalilani L, DiBenedetti DB, et al. Healthcare seeking and treatment for menopausal symptoms in the United States. *Maturitas*. 2007;58:348–358.
40. Avis NE, Crawford SL, Greendale G, et al. Duration of menopausal vasomotor symptoms over the menopause transition. *JAMA Intern Med*. 2015;175:531–539.
41. Rance NE, Dacks PA, Mittelman-Smith MA, et al. Modulation of body temperature and LH secretion by hypothalamic KNDy (kisspeptin, neurokinin B and dynorphin) neurons: a novel hypothesis on the mechanism of hot flushes. *Front Neuroendocrinol*. 2013;34:211–227.
42. Gold JM, Shrimanker I. *Physiology, Vaginal. StatPearls [Internet]*. StatPearls Publishing; 2019.
43. Waetjen LE, Crawford SL, Chang PY, et al. Factors associated with developing vaginal dryness symptoms in women transitioning through menopause: a longitudinal study. *Menopause*. 2018;25:1094.
44. Green R, Polotsky A, Wildman R, et al. Menopausal symptoms within a Hispanic cohort: SWAN, the study of Women's health across the nation. *Climacteric*. 2010;13:376–384.
45. Avis NE, Zhao X, Johannes CB, et al. Correlates of sexual function among multi-ethnic middle-aged women: results from the Study of Women's Health across the Nation (SWAN). *Menopause*. 2018;25:1244–1255.
46. Avis NE, Colvin A, Karlamangla AS, et al. Change in sexual functioning over the menopause transition: results from the Study of Women's Health across the Nation (SWAN). *Menopause*. 2017;24:379.
47. Kravitz HM, Kazlauskaite R, Joffe H. Sleep, health, and metabolism in midlife women and menopause: food for thought. *Obstet Gynecol Clin*. 2018;45:679–694.
48. Freeman EW, Sammel MD, Gross SA, et al. Poor sleep in relation to natural menopause: a population-based 14-year follow-up of mid-life women. *Menopause*. 2015;22:719.
49. Allshouse A, Pavlovic J, Santoro N. Menstrual cycle hormone changes associated with reproductive aging and how they may relate to symptoms. *Obstet Gynecol Clin*. 2018;45:613–628.
50. Dennerstein L, Dudley EC, Hopper JL, et al. A prospective population-based study of menopausal symptoms. *Obstet Gynecol*. 2000;96:351–358.
51. Mitchell ES, Woods NF. Symptom experiences of midlife women: observations from the Seattle midlife Women's health study. *Maturitas*. 1996;25:1–10.
52. Hall JE, Lavoie HB, Marsh EE, et al. Decrease in gonadotropin-releasing hormone (GnRH) pulse frequency with aging in postmenopausal women. *J Clin Endocrinol Metab*. 2000;85:1794–1800.
53. Santoro N, Banwell T, Tortoriello D, et al. Effects of aging and gonadal failure on the hypothalamic-pituitary axis in women. *Am J Obstet Gynecol*. 1998;178:732–741.
54. Wilshire GB, Loughlin JS, Brown JR, et al. Diminished function of the somatotropic axis in older reproductive-aged women. *J Clin Endocrinol Metab*. 1995;80:608–613.
55. Santoro N. The menopause transition: an update. *Hum Reprod Update*. 2002;8:155–160.
56. Della Torre S, Rando G, Meda C, et al. Amino acid-dependent activation of liver estrogen receptor alpha integrates metabolic and reproductive functions via IGF-1. *Cell Metab*. 2011;13:205–214.
57. Ardawi M-SM, Rouzi AA, Qari MH. Physical activity in relation to serum sclerostin, insulin-like growth factor-1, and bone turnover markers in healthy premenopausal women: a cross-sectional and a longitudinal study. *J Clin Endocrinol Metab*. 2012;97:3691–3699.
58. Caufriez A, Leproult R, L'Hermite-Balériaux M, et al. Progesterone prevents sleep disturbances and modulates GH, TSH, and melatonin secretion in postmenopausal women. *J Clin Endocrinol Metab*. 2011;96:E614–E623.
59. Milewicz T, Krzysiek J, Sztefko K, et al. 17β-estradiol regulation of human growth hormone (hGH), insulin-like growth factor-I (IGF-I) and insulin-like growth factor binding protein-3 (IGFBP-3) axis in hypoestrogenic, hypergonadotropic women. *Endokrynol Pol*. 2005;56:876–882.
60. Cummings DE, Merriam GR, eds. *Age-related Changes in Growth Hormone Secretion: Should the Somatopause Be Treated? Semin Reprod Endocrinol*. New York, NY Medical Publishers, Inc; 1999. Copyright© 1999 by Thieme.
61. Leitolf H, Behrends J, and Brabant G. "The thyroid axis in ageing." *Novartis foundation symposium*. Chichester; New York; John Wiley; 1999, 2002.
62. Monzani F, Del Guerra P, Caraccio N, et al. Age-related modifications in the regulation of the hypothalamic-pituitary-thyroid axis. *Horm Res Paediatr*. 1996;46:107–112.
63. Mooradian AD. Normal age-related changes in thyroid hormone economy. *Clin Geriatr Med*. 1995;11:159–169.
64. Deuschle M, Gotthardt U, Schweiger U, et al. With aging in humans the activity of the hypothalamus-pituitary-adrenal system increases and its diurnal amplitude flattens. *Life Sci*. 1997;61:2239–2246.
65. Parker Jr CR, Slayden SM, Azziz R, et al. Effects of aging on adrenal function in the human: responsiveness and sensitivity of adrenal androgens and cortisol to adrenocorticotropin in premenopausal and postmenopausal women. *J Clin Endocrinol Metab*. 2000;85:48–54.
66. Doldi N, Belvisi L, Bassan M, et al. Premature ovarian failure: steroid synthesis and autoimmunity. *Gynecol Endocrinol*. 1998;12:23–28.
67. Ghebre MA, Hart DJ, Hakim AJ, et al. Association between DHEAS and bone loss in postmenopausal women: a 15-year longitudinal population-based study. *Calcif Tissue Int*. 2011;89:295.
68. Shufelt C, Bretsky P, Almeida CM, et al. DHEA-S levels and cardiovascular disease mortality in postmenopausal women: results from the National Institutes of health—National heart, Lung, and blood Institute (NHLBI)-sponsored Women's Ischemia syndrome evaluation (WISE). *J Clin Endocrinol Metab*. 2010;95:4985–4992.
69. Morrison MF, Freeman EW, Lin H, et al. Higher DHEA-S (dehydroepiandrosterone sulfate) levels are associated with depressive symptoms during the menopausal transition: results from the PENN Ovarian Aging Study. *Arch Womens Ment Health*. 2011;14:375.
70. Karlamangla AS, Burnett-Bowie S-AM, Crandall CJ. Bone health during the menopause transition and beyond. *Obstet Gynecol Clin*. 2018;45:695–708.
71. Verghese J, Lipton RB, Katz MJ, et al. Leisure activities and the risk of dementia in the elderly. *N Engl J Med*. 2003;348:2508–2516.
72. Berent-Spillson A, Persad CC, Love T, et al. Hormonal environment affects cognition independent of age during the menopause transition. *J Clin Endocrinol Metab*. 2012;97:E1686–E1694.
73. Greendale G, Huang M, Wight R, et al. Effects of the menopause transition and hormone use on cognitive performance in midlife women. *Neurology*. 2009;72:1850–1857.
74. Epperson CN, Sammel MD, Freeman EW. Menopause effects on verbal memory: findings from a longitudinal community cohort. *J Clin Endocrinol Metab*. 2013;98:3829–3838.
75. Egan KR, Gleason CE. Longer duration of hormonal contraceptive use predicts better cognitive outcomes later in life. *J Womens Health*. 2012;21:1259–1266.
76. Resnick SM, Maki PM, Rapp SR, et al. Effects of combination estrogen plus progestin hormone treatment on cognition and affect. *J Clin Endocrinol Metab*. 2006;91:1802–1810.
77. Espeland MA, Rapp SR, Shumaker SA, et al. Conjugated equine estrogens and global cognitive function in postmenopausal women: Women's Health Initiative Memory Study. *JAMA*. 2004;291:2959–2968.
78. Shumaker SA, Legault C, Kuller L, et al. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA*. 2004;291:2947–2958.
79. Maki PM. The critical window hypothesis of hormone therapy and cognition: a scientific update on clinical studies. *Menopause*. 2013;20:695.

80. Miller VM, Naftolin F, Asthana S, et al. The kronos early estrogen prevention study (KEEPS): what have we learned? *Menopause*. 2019;26:1071.
81. Henderson VW, John JAS, Hodis HN, et al. Cognitive effects of estradiol after menopause: a randomized trial of the timing hypothesis. *Neurology*. 2016;87:699–708.
82. Kitzman DW, Scholz DG, Hagen PT, et al. Age-related changes in normal human hearts during the first 10 decades of life. Part II (Maturity): A quantitative anatomic study of 765 specimens from subjects 20 to 99 years old. Mayo Clin Proc. 1988 Feb;63(2):137-46.
83. Duncan AK, Vittone J, Fleming KC, et al. Cardiovascular disease in elderly patients. Mayo Clin Proc. 1996 Feb;71(2):184-96.
84. Hodis HN, Mack WJ, Henderson VW, et al. Vascular effects of early versus late postmenopausal treatment with estradiol. *N Engl J Med*. 2016;374:1221–1231.
85. Cagnacci A, Cannoletta M, Palma F, et al. Menopausal symptoms and risk factors for cardiovascular disease in postmenopause. *Climacteric*. 2012;15:157–162.
86. Thurston RC, Carroll JE, Levine M, et al. Vasomotor symptoms and accelerated epigenetic aging in the Women's Health Initiative (WHI). *J Clin Endocrinol Metab*. 2020;105:1221–1227.
87. Svartberg J, von Mühlen D, Kritz-Silverstein D, et al. Vasomotor symptoms and mortality: the Rancho Bernardo study. *Menopause*. 2009;16:888–891.
88. Wolff EF, He Y, Black DM, et al. Self-reported menopausal symptoms, coronary artery calcification, and carotid intima-media thickness in recently menopausal women screened for the Kronos early estrogen prevention study (KEEPS). *Fertil Steril*. 2013;99:1385–1391.
89. Tseng LA, El Khoudary SR, Young EA, et al. The Association of menopausal status with physical function: the Study of Women's Health across the Nation (SWAN): menopausal status and physical function. *Menopause*. 2012;19:1186.
90. Cray L, Woods NF, Mitchell ES. Symptom clusters during the late menopausal transition stage: observations from the Seattle Midlife Women's Health Study. *Menopause*. 2010;17:972–977.
91. Moen M, Kahn H, Bjerve K, et al. Menometrorrhagia in the perimenopause is associated with increased serum estradiol. *Maturitas*. 2004;47:151–155.
92. Donnez J. Menometrorrhagia during the premenopause: an overview. *Gynecol Endocrinol*. 2011;27:1114–1119.
93. Blake RE. Leiomyomata uteri: hormonal and molecular determinants of growth. *J Natl Med Assoc*. 2007;99:1170.
94. Stewart EA, Nowak RA. Leiomyoma-related bleeding: a classic hypothesis updated for the molecular era. *Hum Reprod Update*. 1996;2:295–306.
95. DeWaay DJ, Syrop CH, Nygaard IE, et al. Natural history of uterine polyps and leiomyomata. *Obstet Gynecol*. 2002;100:3–7.
96. Mishra GD, Cardozo L, Kuh D. Menopausal transition and the risk of urinary incontinence: results from a British prospective cohort. *BJU Int*. 2010;106:1170–1175.
97. Waetjen LE, Johnson WO, Xing G, et al. Serum estradiol levels are not associated with urinary incontinence in mid-life women transitioning through menopause. *Menopause*. 2011;18:1283.
98. Burger HG, Dudley EC, Hopper JL, et al. The endocrinology of the menopausal transition: a cross-sectional study of a population-based sample. *J Clin Endocrinol Metab*. 1995;80:3537–3545.
99. Doufas AG, Mastorakos G. The hypothalamic–pituitary–thyroid axis and the female reproductive system. *Ann N Y Acad Sci*. 2000;900:65–76.
100. Cohen LS, Soares CN, Vitonis AF, et al. Risk for new onset of depression during the menopausal transition: the Harvard study of moods and cycles. *Arch Gen Psychiatry*. 2006;63:385–390.
101. Bromberger J, Kravitz H, Matthews K, et al. Predictors of first lifetime episodes of major depression in midlife women. *Psychol Med*. 2009;39:55–64.
102. Santoro N, Miller Valery T. Menopause and the postmenopausal state. *Curr Med*. 1998;1:423–428.
103. Kim TJ, Anasti JN, Flack MR, et al. Routine endocrine screening for patients with karyotypically normal spontaneous premature ovarian failure. *Obstet Gynecol*. 1997;89:777–779.
104. Nelson HD, Humphrey LL, Nygren P, et al. Postmenopausal hormone replacement therapy: scientific review. *JAMA*. 2002;288:872–881.
105. ACo O, Gynecologists. ACOG committee opinion no. 556: postmenopausal estrogen therapy: route of administration and risk of venous thromboembolism. *Obstet Gynecol*. 2013;121:887.
106. Schnatz PF. The 2017 hormone therapy position statement of the North American Menopause Society. *Menopause*. 2017;24:728.
107. Nelson HD. Commonly used types of postmenopausal estrogen for treatment of hot flashes: scientific review. *JAMA*. 2004;291:1610–1620.
108. Somboonporn W, Panna S, Temtanakitpaisan T, et al. Effects of the levonorgestrel-releasing intrauterine system plus estrogen therapy in perimenopausal and postmenopausal women: systematic review and meta-analysis. *Menopause*. 2011;18:1060–1066.
109. Nelson HD, Vesco KK, Haney E, et al. Nonhormonal therapies for menopausal hot flashes: systematic review and meta-analysis. *JAMA*. 2006;295:2057–2071.
110. Shams T, Firwana B, Habib F, et al. SSRIs for hot flashes: a systematic review and meta-analysis of randomized trials. *J Gen Intern Med*. 2014;29:204–213.
111. Bullock JL, Massey F, Gambrell JR. Use of medroxyprogesterone acetate to prevent menopausal symptoms. *Obstet Gynecol*. 1975;46:165–168.
112. Quella SKLC, Sloan JA, et al. Long term use of megestrol acetate by cancer survivors for the treatment of hot flashes. *Cancer*. 1998:165–168.
113. Prior JC, Hitchcock CL. Progesterone for hot flush and night sweat treatment–effectiveness for severe vasomotor symptoms and lack of withdrawal rebound. *Gynecol Endocrinol*. 2012;28:7–11.
114. Pinkerton JV. Hormone therapy for postmenopausal women. *N Engl J Med*. 2020;382:446–455.
115. Depypere H, Timmerman D, Donders G, et al. Treatment of menopausal vasomotor symptoms with fezolinetant, a neurokinin 3 receptor antagonist: a phase 2a trial. *J Clin Endocrinol Metab*. 2019;104:5893–5905.
116. Crandall CJ, Hovey KM, Andrews CA, et al. Breast cancer, endometrial cancer, and cardiovascular events in participants who used vaginal estrogen in the Women's Health Initiative Observational Study. *Menopause*. 2018;25:11.
117. Cody JD, Jacobs ML, Richardson K, et al. Oestrogen therapy for urinary incontinence in postmenopausal women. *Cochrane Database Syst Rev*. 2012;10.
118. Bachmann GA, Komi JO, Group OS. Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: results from a pivotal phase 3 study. *Menopause*. 2010;17:480–486.
119. Avis NE, Brockwell S, Randolph Jr JF, et al. Longitudinal changes in sexual functioning as women transition through menopause: results from the Study of Women's Health across the Nation (SWAN). *Menopause*. 2009;16:442.
120. Dennerstein L. Sexuality, midlife, and menopause. *Menopause*. 2008;15:221–222.
121. El-Hage G, Eden J, Zoa Manga R. A double-blind, randomized, placebo-controlled trial of the effect of testosterone cream on the sexual motivation of menopausal hysterectomized women with hypoactive sexual desire disorder. *Climacteric*. 2007;10:335–343.
122. Braunstein GD, Sundwall DA, Katz M, et al. Safety and efficacy of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo-controlled trial. *Arch Intern Med*. 2005;165:1582–1589.
123. Wierman ME, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:3489–3510.
124. Labrie F, Archer D, Bouchard C, et al. Effect of intravaginal dehydroepiandrosterone (Prasterone) on libido and sexual dysfunction in postmenopausal women. *Menopause*. 2009;16:923–931.
125. Labrie F, Archer DF, Koltun W, et al. Efficacy of intravaginal dehydroepiandrosterone (DHEA) on moderate to severe dyspareunia and vaginal dryness, symptoms of vulvovaginal atrophy, and of the genitourinary syndrome of menopause. *Menopause*. 2018;25:1339–1353.
126. Al-Safi ZA, Santoro N. Menopausal hormone therapy and menopausal symptoms. *Fertil Steril*. 2014;101:905–915.
127. Kravitz HM, Zhao X, Bromberger JT, et al. Sleep disturbance during the menopausal transition in a multi-ethnic community sample of women. *Sleep*. 2008;31:979–990.

128. Toffol E, Heikinheimo O, Partonen T. Hormone therapy and mood in perimenopausal and postmenopausal women: a narrative review. *Menopause*. 2015;22:564–578.
129. Gleason CE, Dowling NM, Wharton W, et al. Effects of hormone therapy on cognition and mood in recently postmenopausal women: findings from the randomized, controlled KEEPS–cognitive and affective study. *PLoS Med*. 2015;12:e1001833.
130. Gordon JL, Rubinow DR, Eisenlohr-Moul TA, et al. Efficacy of transdermal estradiol and micronized progesterone in the prevention of depressive symptoms in the menopause transition: a randomized clinical trial. *JAMA Psychiatr*. 2018;75:149–157.
131. Van Voorhis BJ, Santoro N, Harlow S, et al. The relationship of bleeding patterns to daily reproductive hormones in women approaching menopause. *Obstet Gynecol*. 2008;112:101.
132. Tietjen GE. The relationship of migraine and stroke. *Neuroepidemiology*. 2000;19:13–19.
133. Sulak PJ, Cressman BE, Waldrop E, et al. Extending the duration of active oral contraceptive pills to manage hormone withdrawal symptoms. *Obstet Gynecol*. 1997;89:179–183.
134. Yoo HJ, Lee MA, Ko YB, et al. The efficacy of the levonorgestrel-releasing intrauterine system in perimenopausal women with menorrhagia or dysmenorrhea. *Arch Gynecol Obstet*. 2012;285:161–166.
135. Bair YA, Gold EB, Zhang G, et al. Use of complementary and alternative medicine during the menopause transition: longitudinal results from the Study of Women's Health across the Nation. *Menopause*. 2008;15:32–43.
136. Albertazzi P, Pansini F, Bonaccorsi G, et al. The effect of dietary soy supplementation on hot flushes. *Obstet Gynecol*. 1998;91:6–11.
137. Newton KM, Reed SD, LaCroix AZ, et al. Treatment of vasomotor symptoms of menopause with black cohosh, multibotanicals, soy, hormone therapy, or placebo: a randomized trial. *Ann Intern Med*. 2006;145:869–879.
138. Chenoy R, Hussain S, Tayob Y, et al. Effect of oral gamolenic acid from evening primrose oil on menopausal flushing. *BMJ*. 1994;308:501–503.
139. Hirata JD, Swiersz LM, Zell B, et al. Does dong quai have estrogenic effects in postmenopausal women? A double-blind, placebo-controlled trial. *Fertil Steril*. 1997;68:981–986.
140. Nedrow A, Miller J, Walker M, et al. Complementary and alternative therapies for the management of menopause-related symptoms: a systematic evidence review. *Arch Intern Med*. 2006;166:1453–1465.
141. Ross AC, Manson JE, Abrams SA, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab*. 2011;96:53–58.
142. Gutin B, Kasper M. Can vigorous exercise play a role in osteoporosis prevention? A review. *Osteoporos Int*. 1992;2:55–69.
143. Dalsky GP, Stocke KS, Ehsani AA, et al. Weight-bearing exercise training and lumbar bone mineral content in postmenopausal women. *Ann Intern Med*. 1988;108:824–828.
144. Fiatarone MA, Marks EC, Ryan ND, et al. High-intensity strength training in nonagenarians: effects on skeletal muscle. *JAMA*. 1990;263:3029–3034.
145. Kanis J, McCloskey E, Johansson H, et al. Development and use of FRAX® in osteoporosis. *Osteoporos Int*. 2010;21:407–413.
146. Kanis J, Odén A, Johnell O, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int*. 2007;18:1033–1046.
147. Torgerson DJ, Bell-Syer SE. Hormone replacement therapy and prevention of vertebral fractures: a meta-analysis of randomised trials. *BMC Musculoskelet Disord*. 2001;2:7.
148. Hulley S, Furberg C, Barrett-Connor E, et al. Noncardiovascular disease outcomes during 6.8 years of hormone therapy: heart and Estrogen/progestin Replacement Study follow-up (HERS II). *JAMA*. 2002;288:58–64.
149. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–333.
150. Berg A, Allan J, Frame P, et al. Postmenopausal hormone replacement therapy for primary prevention of chronic conditions: recommendations and rationale. *Ann Intern Med*. 2002;137:834–839.
151. Gizzo S, Saccardi C, Patrelli TS, et al. Update on raloxifene: mechanism of action, clinical efficacy, adverse effects, and contraindications. *Obstet Gynecol Surv*. 2013;68:467–481.
152. Cummings SR, Martin JS, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*. 2009;361:756–765.
153. Neer RM, Arnaud CD, Zanchetta JR, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med*. 2001;344:1434–1441.
154. Grodstein F, Manson JE, Colditz GA, et al. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med*. 2000;133:933–941.
155. Hulley S, Grady D, Bush T, et al. Heart, Group EpRSR. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA*. 1998;280:605–613.
156. Grady D, Herrington D, Bittner V, et al. Cardiovascular disease outcomes during 6.8 years of hormone therapy: heart and Estrogen/progestin Replacement Study follow-up (HERS II). *JAMA*. 2002;288:49–57.
157. Mosca L, Collins P, Herrington DM, et al. Hormone replacement therapy and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. *Circulation*. 2001;104:499–503.
158. Schierbeck LL, Rejnmark L, Tofteng CL, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. *BMJ*. 2012;345.
159. Heiss G, Wallace R, Anderson GL, et al. Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. *JAMA*. 2008;299:1036–1045.
160. Mosca L, Banka CL, Benjamin EJ, et al. Evidence-based guidelines for cardiovascular disease prevention in women: 2007 update. *J Am Coll Cardiol*. 2007;49:1230–1250.
161. Rafique S, Sterling EW, Nelson LM. A new approach to primary ovarian insufficiency. *Obstet Gynecol Clin*. 2012;39:567–586.
162. Shuster LT, Rhodes DJ, Gostout BS, et al. Premature menopause or early menopause: long-term health consequences. *Maturitas*. 2010;65:161–166.
163. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–333.
164. Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291:1701–1712.
165. Hodis HN, Sarrel P. Menopausal hormone therapy and breast cancer: what is the evidence from randomized trials? *Climacteric*. 2018;21:521–528.
166. Chlebowski RT, Anderson GL, Aragaki AK, et al. Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up of the Women's Health Initiative randomized clinical trials. *JAMA*. 2020;324:369–380.
167. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet*. 2019;394:1159–1168.
168. Fournier A, Berrino F, Riboli E, et al. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N−EPIC cohort. *Int J Cancer*. 2005;114:448–454.
169. De Villiers T, Hall J, Pinkerton J, et al. Revised global consensus statement on menopausal hormone therapy. *Maturitas*. 2016;91:153–155.
170. Sjögren LL, Mørch LS, Løkkegaard E. Hormone replacement therapy and the risk of endometrial cancer: a systematic review. *Maturitas*. 2016;91:25–35.
171. Shi LF, Wu Y, Li CY. Hormone therapy and risk of ovarian cancer in postmenopausal women: a systematic review and meta-analysis. *Menopause*. 2016;23:417–424.
172. Liu Y, Ma L, Yang X, et al. Menopausal hormone replacement therapy and the risk of ovarian cancer: a meta-analysis. *Front Endocrinol*. 2019;10:801.
173. Simin J, Tamimi RM, Callens S, et al. Menopausal hormone therapy treatment options and ovarian cancer risk: a Swedish prospective population−based matched−cohort study. *Int J Cancer*. 2020;147:33–44.
174. Manson JE, Aragaki AK, Rossouw JE, et al. Menopausal hormone therapy and long-term all-cause and cause-specific mortality: the Women's Health Initiative randomized trials. *JAMA*. 2017;318:927–938.

123

Female Infertility: Evaluation and Management

Bart C.J.M. Fauser and Frank J. Broekmans

OUTLINE

Introduction, 2046
History, 2046
Epidemiology, 2047
Pathogenesis and Clinical Features, 2048
  Ovulatory Dysfunction, 2048
  Anatomic Dysfunction, 2048
  Ovarian Aging, 2049
  Unexplained Infertility, 2049
Evaluation of the Female in the Infertile Couple, 2050
  Assessment of the Normal Menstrual Cycle, 2050
  Assessment of Anatomic Dysfunction, 2050
  Assessment of Cervical Mucus–Sperm Interaction, 2051
  Assessment of Ovarian Aging, 2052
  From Diagnosis to Prognosis, 2052
Expectant Management of the Infertile Couple, 2053
  Ovarian Stimulation, 2054
  Intrauterine Insemination, 2055
  In Vitro Fertilization, 2055
Summary and Future Developments, 2056

KEY POINTS

- This chapter reviews current knowledge on the evaluation of infertility, with emphasis on the balance between timing and extent of diagnostic procedures.
- Current treatment options for specific diagnostic categories are discussed, as is the management of the large category of unexplained infertility, for which the importance of the right balance between assisted reproduction technology and expectant management is acknowledged.
- Recent developments in assisted reproduction are outlined.

INTRODUCTION

Infertility is defined as the inability to conceive within 1 year of regular unprotected intercourse. *Fecundability* is the probability of achieving a pregnancy within a given menstrual cycle, and *fecundity* is the capacity of a couple to achieve a live birth.[1] One in every six couples visits a physician because of an unfulfilled wish for children. The 1-year period is chosen arbitrarily and is based mainly on assessment of the likelihood for pregnancy without medical intervention (Fig. 123.1). Infertility should be viewed as a problem of the couple; therefore, medical attention should focus on both partners of a couple. Chances are roughly similar for finding some form of reproductive dysfunction in the male as in the female. Abnormalities may also be found in both the male and the female partner. Still, in a considerable number of couples, no obvious abnormalities will be found during the initial infertility workup, and it is not always clear whether an observed abnormality is indeed causally related to reduced fertility. Notably, single individuals and same-sex couples seek fertility services to fulfill their procreative goals. They may also have infertility, warranting infertility evaluation if pregnancies have not been achieved in a 1-year period. As age of the female partner is a major contributor to infertility, shorter time periods for couples or individuals to pursue infertility evaluation and treatment are recommended (see below).

Primary (as opposed to secondary) infertility involves individuals (or couples) with no previous pregnancy. Hence, for a given couple—depending on the individual reproductive history—infertility can be primary for one, and secondary for the other. Decreased fertility (referred to by some as subfertility) of a given individual may be compensated for by enhanced fertility of the partner, and the clinical challenge is to assess chances for pregnancy, either spontaneously or after infertility treatment, of a given couple.

Overall, the proposed management algorithm may include distinctly different options, ranging from expectant management without intervention, to causal therapy for those with distinct forms of infertility, to assisted reproduction as an empirical approach, to increasing chances for offspring in unexplained infertility.

KEYPOINTS

- Infertility is an ill-defined condition, with often no obvious causal explanation.

HISTORY

Infertility care has witnessed rapid growth since the early 1960s. The first two decades were dominated by the possibilities of endocrine diagnosis and treatments and the introduction of reproductive surgical procedures. Gonadotropin and steroid hormone assays were developed for widespread clinical use, and the synthetic antiestrogen clomiphene citrate (CC) and urinary gonadotropins (human menopausal gonadotropin [hMG]) obtained from postmenopausal women containing excreted pituitary follicle-stimulating hormone [FSH] and luteinizing hormone [LH]), and subsequently recombinant hormones became available for stimulating gonadal function in both females and males.

Initial surgical approaches in the female involved wedge resection of polycystic ovaries, microsurgery of endometriotic pelvic disease and uterine myomas, adhesions, and tubal disease[2] and, more recently, by endoscopic pelvic surgery and operative hysteroscopy on uterine anomalies, endometrial abnormalities, submucous myomas, and adhesions. Surgical approaches in the male were dominated by varicocele correction to improve sperm quality, although most of these initial surgical approaches have not proven efficacious after rigorous assessment.[2,3] Yet, the surgical recovery of sperm from the epididymis (percutaneous epididymal sperm extraction) or testis (testicular sperm

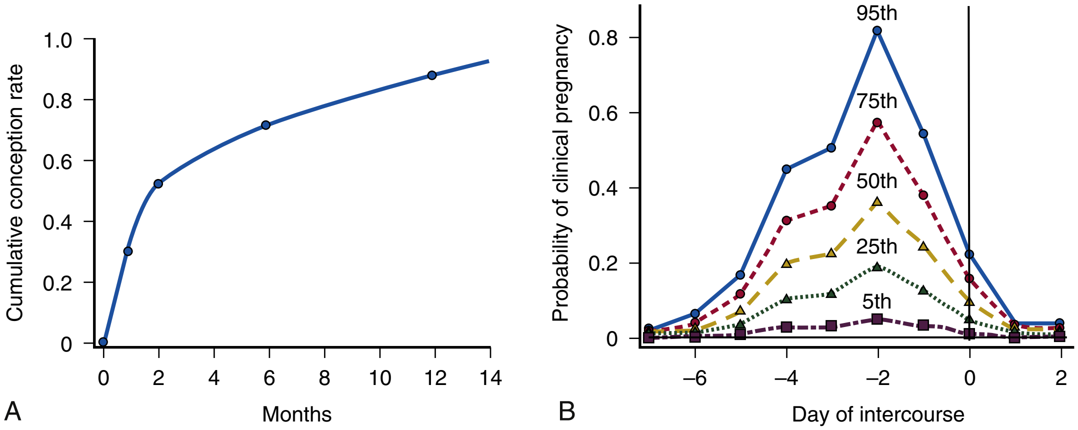


Figure 123.1 A, Cumulative conception rates during the first 12 months of unprotected intercourse. B, Probability of a clinical pregnancy relative to the day of intercourse. *Day 0* depicts the day of the serum luteinizing hormone surge. *Individual lines* depict 5th to 95th percentiles of the total population. (A, from Taylor A. ABC of subfertility: the extent of the problem. *BMJ.* 2003;327:434–436. B, from Evers JL. Female subfertility. *Lancet.* 2002;360:151–159.)

extraction) has greatly changed the perspective of the azoospermic male due to obstructive and spermatogenic causes.

The past two decades have been dominated by the development of assisted reproductive technology (ART), intrauterine insemination (IUI), *in vitro* fertilization (IVF) in 1978,[4] and intracytoplasmic sperm injection (ICSI) in 1992.[5] These novel therapeutic strategies, combined with complex ovarian stimulation (OS) protocols, created new possibilities for infertile couples and individuals. Recent decades have focused on improved methods of embryo cryopreservation which have enabled the large-scale use of single-embryo transfer in IVF, thereby reducing rates of multiple pregnancy and associated medical risks and cost.[6] Other major advances include: mild stimulation protocols using lower doses of drugs, resulting in best embryos to transfer in a given cycle[7,8]; vitrification of oocytes (from stimulated ovaries), which opened the field for oocyte banking for egg donation programs, fertility preservation for cancer treatment patients, and pregnancy postponers[9]; *in vitro* maturation and cryopreservation of oocytes from unstimulated ovaries, which currently lacks sufficient efficiency and safety to allow for large-scale application[10]; and cryopreservation of ovarian tissue, for example, cancer patients before undergoing gonadotoxic chemotherapy and subsequent transfer into the abdominal cavity, once a pregnancy is desired. Currently, over 100 babies have been born worldwide following this procedure.[11] Finally, embryo selection for transfer has been aided by cleavage or blastocyst stage embryo biopsy to detect transmitted hereditary disease (preImplantation genetic testing for disease) such as Huntington's disease, cystic fibrosis or polycystic kidney disease, or embryo aneuploidies (preImplantation genetic testing for aneuploidies).[12] Overall, these advances have enabled millions of children born to infertile couples and individuals, although the cost of IVF, ICSI, TESE, and other technologies can be prohibitive, along with disparities of availability of fertility services in certain regions across the globe, still leaving challenges to be addressed.

KEYPOINTS

- In vitro fertilization and related technologies have been the significant steps forward in infertility management.

EPIDEMIOLOGY

In the general population, the highest chance of pregnancy occurs in the first cycle of unprotected intercourse. The maximal probability for a clinical pregnancy is ~30% per cycle, and chances for pregnancy are highest when intercourse takes place 2 days before ovulation[13] (see Fig. 123.1). The likelihood of pregnancy slowly decreases in subsequent cycles, generating a cumulative chance for pregnancy of 60% within three cycles, 70% within six cycles, and ~90% within 1 year[14] (see Fig. 123.1). Very fertile couples are likely to achieve a pregnancy within the first few cycles, whereas others will conceive within 12 months (Table 123.1). Moreover, a considerable proportion of couples will still conceive spontaneously after the 1-year period or even later (Figs. 123.2 and 123.3). The 1-year arbitrary cutoff period is to define infertility and justify an infertility workup, as the chance of finding an exclusive cause like severe male factor or tubal obstructive pathology is realistic. Infertility occurs in ~10% to 15% of the population.

TABLE 123.1 Estimated Spontaneous Cumulative Pregnancy Rates in Normal Fertile and Infertile Couples*

	Monthly Fecundity Rate (%)	CUMULATIVE PREGNANCY RATE (%)			
		6	12	24	60
Superfertile	60	100			
Normally fertile	20	74	93	100	
Moderately subfertile	5	26	46	71	95
Severely subfertile	1	6	11	21	45
Infertile	0	0	0	0	0

*Rates over a given period of time, under conditions with different monthly fecundity rates. The data indicate that women with moderate fertility, which applies to many in their 30s, may have a considerable chance (≈50%) of not having achieved a pregnancy within 1 year's time. From Evers JL. Female subfertility. *Lancet.* 2002;360:151–159.

Multiple societal tendencies interfering with normal fertility have been described, especially (1) early sexarche and promiscuous sexual behavior (resulting in increased chances for tubal disease due to sexually transmitted disease), (2) postponement of pregnancy and consequently an increased proportion of women older than 30 years of age with already decreased natural fertility (due to diminished ovarian reserve) when offspring is desired, (3) female lifestyle factors (especially obesity and resulting ovarian dysfunction), and (4) environmental and toxic agents (e.g., cigarette smoking inducing accelerated follicle pool depletion). Consequently, more couples will not conceive spontaneously within 1 year of unprotected intercourse and, therefore,

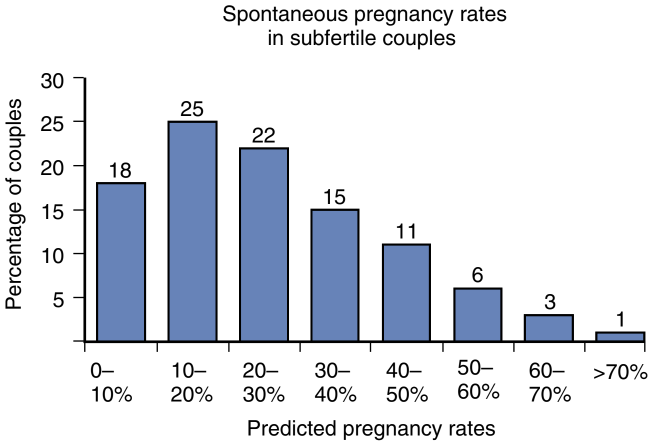


Figure 123.2 The distribution of the probability to become pregnant among untreated infertile couples within the year after intake for infertility workup (defined as having had at least 1 year of unprotected intercourse). (Data from Eimers JM, te Velde ER, Gerritse R, et al. The prediction of the chance to conceive in subfertile couples. *Fertil Steril.* 1994;61:44–52.)

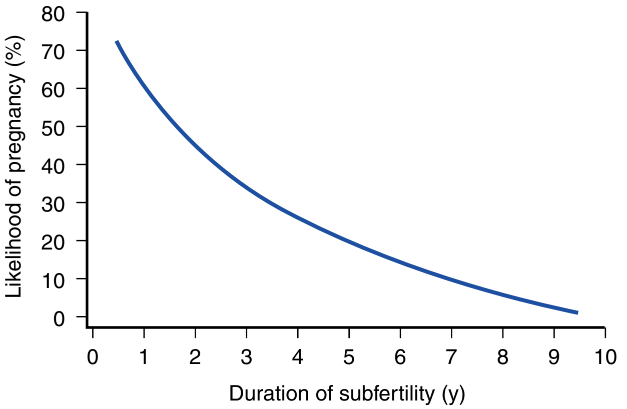


Figure 123.3 The likelihood of pregnancy in couples relative to the duration of their infertility. Note the considerable chance to conceive within the first 3 years of unprotected intercourse. For instance, the probability for a couple to become pregnant in the course of the second year of trying is still 50%. (Data from Evers JL. Female subfertility. *Lancet.* 2002;360:151–159.)

will visit a healthcare professional for infertility care. It appears that the prevalence of infertility in the Western world has remained unaltered over the past 50 years or may even have reduced,[14–16] although more couples have been seeking access to infertility services.[17]

KEYPOINTS

- Postponing family building creates higher risks of infertility occurring, in relation to the female age.

PATHOGENESIS AND CLINICAL FEATURES

According to textbooks, the female factor accounts for approximately 35% of infertility, the male factor for 35%, and in about 30% of cases, both male and female factors are involved. In female infertility, approximately 30% to 40% involve ovulatory dysfunction, 30% to 40% involve tubal and pelvic pathology, and 30% are attributed to "unexplained" causes, of which advanced reproductive age may be the most important contributor.[18] Clinical data supporting these figures are not robust. However, differences among populations and different referral characteristics may exist. Overall, a full workup of the infertile couple will yield a cause leading to infertility in approximately 23% (bilateral tubal obstruction, amenorrhea, azoospermia); in the remainder, only factors that will reduce the probability of conception (i.e., subfertility) are identified of which female age is the factor with the highest impact.[19]

Couples who suffer from infertility usually do not have associated complaints or symptoms. Exceptions include abnormal or absent menstrual bleeding, a history of pelvic inflammatory disease, ruptured appendix, or maldescent of one or both testes. The general rule that the more tests conducted, the more likely it is that at least one test will be abnormal by chance also holds true for infertility. It may not be easy to identify a disease state that is causally related to reproductive failure in a given couple or individual. Classic examples include conditions such as endometriosis or uterine leiomyomas, which were widely believed to induce infertility, until it was found that both these conditions are also prevalent in the general, fertile population.

Similar findings apply to male infertility. Poor sperm quality (by World Health Organization [WHO] criteria) was widely believed to be the cause of infertility in many cases, until more recent studies demonstrated that diminished sperm quality is often observed in males with children.[20,21] Although criteria for defining abnormal sperm have been revised repeatedly, the predictive power of sperm quality assessment should be viewed with caution in an infertile population. A clear effect upon pregnancy rates has been demonstrated only in males with severely reduced sperm density, or azoospermia.[22]

Several classes of disease conditions that may underlie female infertility can be identified, notably ovulatory dysfunction, mechanical dysfunction, ovarian aging, and unexplained infertility.

Ovulatory Dysfunction

Changes in ovarian function usually are presented by extended bleeding intervals (oligomenorrhea) or absence of bleeding (amenorrhea). Amenorrhea is usually associated with a hypoestrogenic status, whereas bleeding in oligomenorrhea may represent occasional ovulation or breakthrough bleeding due to continued estrogen exposure. In primary amenorrhea, with or without disturbed sexual maturation, one should consider rare conditions involving abnormal gonadal or uterine development due to (inherited or spontaneous) gene mutations (Chapter 117).

Ovulatory dysfunction can be classified into three categories based on FSH and estradiol (E_2) hormone assays in peripheral blood.[23] Low FSH and E_2 indicate a central origin of the abnormality at the level of the hypothalamus or pituitary gland. Hence, ovaries fail to function normally owing to insufficient stimulation by pituitary gonadotropins (WHO group 1). In contrast, the combination of high FSH and low E_2 concentrations indicates that the primary defect is at the ovarian level (WHO group 3, i.e., premature ovarian insufficiency [POI]), with the ovaries failing to produce estrogens despite maximal stimulation by endogenous FSH. The third group of oligo/anovulatory patients (WHO group 2) is represented by FSH and E_2 levels within the normal range constituting approximately 80% to 85% of all women presenting with ovulatory dysfunction. This heterogeneous group especially includes polycystic ovarian syndrome (PCOS) also characterized by clinical or biochemical androgen excess and/or polycystic ovaries (see Chapter 117).[24]

Luteal-phase insufficiency (short luteal phase) as an expression of ovulatory dysfunction has long been believed to be an important factor in infertility, occurring in up to 14% of infertile patients. However, controlled studies conducted to establish its role in infertility or the value of any treatment in improving pregnancy rates are currently still lacking.[25]

Anatomic Dysfunction

Normal functioning of the reproductive tract is important for fertility. Uterine abnormalities, including congenital uterine malformations,

malformations due to fetal exposure to diethylstilbestrol, fibroids (in particular, the submucosal or cornual myomas), polyps, and adhesions related to previous intrauterine procedures, are associated with infertility or early miscarriage.[26]

Normal fertility may also be affected by pelvic adhesions (especially involving the ovaries and fallopian tubes) caused by pelvic inflammatory disease, complicated appendicitis, or intraabdominal surgery. Abnormal tubal function may be due to isthmic or terminal occlusion or phimosis of the fimbriae due to salpingitis. Peritubal adhesions may interfere with normal tubal motility and, hence, with gamete and zygote transport. Severely damaged tubes may fill with fluid, giving rise to hydrosalpinges.

Endometriosis (see Chapter 125) is a topic of continued controversy related to disease conditions associated with infertility. It is the ectopic localization of the endometrial tissue, most commonly in the pelvis; retrograde menstruation may be the important causal factor, but metaplastic errors during Mullerian tube development may also contribute to endometrial tissue in the myometrium or rectovaginal septum.[27] Severe forms of endometriosis give rise to adhesions, cyst formation, and reduced fertility. In addition, abdominal discomfort, specifically dysmenorrhea, chronic pelvic pain, and dyspareunia are frequent complaints. Limited data suggest that minimal and mild disease are prevalent in the general population, without a clear relation to infertility or pain. Moreover, positive effects on subsequent fertility of intervention studies using hormonal medication or surgical procedures are still questionable.[28,29]

Uterine myomas (see Chapter 126) occur in 1% to 2.5% of infertile patients without other evident cause of infertility,[30] and removal has long been thought to improve subsequent pregnancy chances. A recent meta-analysis has shown that infertility patients with uterine cavity-distorting fibroids have poorer reproductive outcomes than infertile patients without fibroids, and removal therefore may be considered justified despite a lack of comparative trials. Cases with intramural myomas may have a poorer prospect for pregnancy occurring, but the absence of quality studies prevents advocating their removal.[31] Subserosal fibroids are not associated with fertility issues.[32]

As with many other factors associated with infertility, comparative data on the frequency of similar abnormalities in a matched, fertile population are lacking. In addition, randomized follow-up studies of pregnancy rates comparing intervention versus expectant management are rare, so the efficacy of treatment is uncertain.

Ovarian Aging

In many societies, increased participation of females in extended education and careers has resulted in postponement of childbearing to a later age. An increased proportion of women make their first attempts to conceive when they are beyond 30 years of age, when natural fertility starts to decrease, and with a major subsequent decline in fertility after 37 years of age.[33] Beyond 41 years of age, chances of having a spontaneously conceived baby have decreased on average to close to zero.

At around 20 weeks' gestation, in the female fetus, initial germ cell mitosis is arrested, and the primordial follicle pool reaches its maximum size of around 8 million per ovary (see Chapter 122). Thereafter, this stock of resting primordial follicles is gradually depleted (by the onset of growth of follicles), falling to 2 million at birth and 0.5 million at menarche (first menstrual period), with final exhaustion of the pool with the last spontaneous menstrual period. Thus, only a small proportion (~10% to 15%) of the initial primordial follicle pool is present at the time that many women consider conceiving their first child.

Menopause, which marks the definitive end of the reproductive life span, occurs at the average age of 51, with a range from 40 to 60 years of age (see Chapter 122). This indicates that exhaustion of the follicle pool takes 50% more time (60 years rather than 40 years) in some women than in others. The association between chronologic age and ovarian aging therefore may be complex. Indirect evidence indicates that a fixed time interval exists between the transition from normal fertility into decreased and finally into absent fertility (sterility).[34] Some women at 35 years of age may already have severely compromised fertility due to poor-quality oocytes. These women may also present with an earlier than average menopause, although the relation between the quantitative and qualitative aspects of the ovarian ageing process are far from elucidated. For other women of identical age (35 years), natural fertility is still completely normal or may even be above normal. The latter group of women is often believed to experience menopause at a later age, although the evidence is minimal. The clinical challenge is to assess the reproductive aging status in terms of fecundability of a given woman at a given age, which will help to predict her pregnancy chances more accurately than will chance assessment based on chronologic age alone. This may help the clinician in counseling the patient regarding expectant management or therapeutic strategies.[34] Studies on genetics of cell cycle and immune system maintenance may potentially identify relevant factors in this context. "Early" ovarian aging can be linked to genetic as well as environmental factors (see Chapters 122 and 137).

The mechanisms underlying reduced natural fertility in relation to ovarian aging have not yet been fully elucidated. It has been clearly established that the size of the cohort of growing follicles is diminished, and this coincides with an increased proportion of chromosomally abnormal oocytes. The latter is exemplified by increased rates of aneuploid (e.g., trisomy 21) pregnancy, increased rates of early pregnancy loss (most likely based on an aneuploid conceptus), and increased numbers of aneuploid oocytes and embryos with increasing female age.[35] Changes in follicle cohort size can be assessed by transvaginal ultrasound or by hormone estimates in the peripheral blood (see later). Unfortunately, as explained in the aforementioned, assessment of average oocyte quality on an individual basis so far has not proven feasible.

Unexplained Infertility

The indicated percentage of infertility of unknown origin is ~30% but may be higher depending on the criteria applied. The reported incidence is very much dependent on the extent of the infertility evaluation, characteristics of the population studied, female age, and the reported incidence of other "known" factors related to infertility. Again, a clear-cut causal relationship between an abnormal condition as established during initial evaluation and infertility is often absent. A clear tendency towards less diagnostic evaluation resulting in an increased proportion of infertility remaining "unexplained" can be observed. This implies that several factors are known that do not per se exclude the possibility of achieving a pregnancy (e.g., mild or moderate semen abnormalities, a suboptimal or negative postcoital test, mild-stage endometriosis, elevated basal FSH or low age-specific anti-Müllerian hormone [AMH], or subtle uterine abnormalities like fibroids), but they only contribute to a reduction in monthly fecundity. In assessing the remaining chances for success in these unexplained infertile couples or individuals, key factors such as female age, a previous pregnancy, duration of the current infertility status, and semen quality have been identified as crucial factors for the individualized estimates of pregnancy chances.[36,37]

On the other hand, many processes are involved in natural fertility that cannot be tested properly in a routine clinical setting. These involve oocyte maturation and fertilization, embryo transport through the tube and uterine cavity, chromosomal constitution of embryos, apposition and implantation of embryos, and, finally, endometrial receptivity, including regulation of endometrial maturation by steroids produced by the corpus luteum,[38] and gene expression variation indicative of changes in the window of implantation.[39]

Moreover, many pregnant women, under natural conditions or in an infertile population, may miscarry very early, before the occurrence of her expected, regular menses.[38]

KEYPOINTS

- The true key causes for female infertility are uterine/tubal pathology and oligo-/an-ovulation. However, age related limitations at the level of the oocyte may comprise a contribution of yet unknown magnitude.
- Female age effects on fertility are likely to be related to oocyte quality. An individual specific test to assess average ovulated oocyte quality is urgently needed.

EVALUATION OF THE FEMALE IN THE INFERTILE COUPLE

The extent to which basic infertility testing should be performed is extensively debated in the current era, dominated by ART. Consequently, marked differences exist with regard to the numbers and types of tests performed.[40] Tests are time-consuming and costly, and some are not performed without risks or high patient burden. According to some clinical investigators, results may have little bearing on ART outcomes. In a commercial environment where patients have to pay for services, "customers" may prefer to spend their money on treatment with some chance for pregnancy than on diagnostic procedures that will have no impact on their chances to conceive.

Many tests are not clearly standardized, and therefore reproducibility is limited. The usefulness of infertility tests should be based on the proper evaluation of (1) sensitivity (the capacity of a test to detect abnormalities), (2) specificity (the chance that an abnormality is actually present if the test yields an abnormal result), (3) invasiveness, complexity, and cost of the test, and, finally, (4) the clinical utility of the test in terms of how result will influence the management or treatment options for the infertile couple or individual. Indeed, few studies have been undertaken to evaluate the practical value of tests despite their widespread use in clinical practice.[41]

During the initial consultation, the physician should take a full medical history from both partners. Relevant information includes duration of infertility, past sexual development, children from current or previous relationships, general health and use of medication, current or past smoking or use of drugs, including alcohol, obstetric history, previous surgery, family history, abdominal or menstrual discomfort, and sexually transmitted diseases. Moreover, the clinician should counsel a couple regarding timing of intercourse and fertility chances, the use of LH detection tests, the effects of lifestyle factors (such as excessive smoking, alcohol intake, caffeine consumption, and exercise), food intake and obesity, and occupational exposures. A general physical examination usually is advised, with focus on height, weight, and blood pressure.[17,42,43]

Diagnostic tests that may be performed for the evaluation of infertile couples include assessment of the normal menstrual cycle, assessment of anatomic factors, and assessment of ovarian aging status.

Assessment of the Normal Menstrual Cycle

Regular menstrual periods strongly suggest normal ovulatory cycles. Ovulation and normal corpus luteum function can be assessed in different ways.[25] However, the tests are not optimal in terms of accuracy and reproducibility, or in the detection of a meaningful and treatable disease condition relevant to infertility.[44] Applications for mobile phones are being developed and marketed successfully aiming to monitor the fertile period more closely, which can be used both to improve pregnancy chances when desired but also to avoid pregnancy as a means of contraception. These programs improve over time, also by applying artificial intelligence using big data sources.

Menstrual Period Chart. This represents the simplest way to document the menstrual cycle pattern. The days of onset and cessation of menses should be recorded on a calendar for several months, or may be kept in one of several apps.[45] A normal menstrual cycle ranges between 25 and 35 days, with a median duration of 28 days.

Basal Body Temperature Chart. Progesterone produced by the corpus luteum has many effects on the central nervous system, including changes in the body temperature set point. Consequently, daily assessment of basal body temperature (BBT) can be used to document a persistent increase in temperature (typically from 36.6°C to 37.1°C) and therefore can provide "proof" of ovulation. A rise in temperature indicates that ovulation and corpus luteum formation have taken place. The length of the second half of the cycle should be at least 10 days. Daily temperature assessments should be performed at the same time during the early morning before the activities are begun. Although performing these tests for several months can be quite stressful to women, it may be helpful for the couple to be aware of the day of ovulation, so they can time intercourse to achieve optimal pregnancy chances. However, the BBT is notoriously unreliable for the timing of ovulation, and LH surge tests and menstrual and fertile period apps are taking over fast.[45]

Urinary LH Surge. When LH concentrations are high in serum during the midcycle surge, sufficient quantities are excreted into urine to allow its detection. Kits are commercially available for use at home. On average, LH appears in the urine around 12 hours after onset of the LH surge in the blood, and this can accurately predict ovulation (see Chapters 117 and 120). This method may help with optimal timing of intercourse, although evidence for a beneficial effect upon pregnancy rates has so far not been found.[45,46]

Transvaginal Ultrasonography. The preovulatory, Graafian follicle usually attains a diameter of around 20 mm (ranging from 16 to 24 mm) before rupture and subsequent release of the oocyte. Rupture of the follicle can be established by visualizing the disappearance or decrease in size of the dominant follicle by transvaginal ultrasound. This method usually is reserved for monitoring of OS during infertility therapy or for proper timing of a postcoital test or midluteal phase progesterone assay (see the following section).

Progesterone Assay. Appropriate timing of the assessment of progesterone serum concentrations is crucial and should always be checked in retrospect, especially in women whose cycle length may vary. A level > 2 ng/mL indicates ovulation, although levels between 6 and 25 ng/mL are considered normal and are associated with chances for pregnancy. Some investigators advocate that multiple hormone assays should be performed around the midluteal phase to document insufficient corpus luteum function. The luteal phase may be reduced in length, or insufficient quantities of progesterone may be produced during the luteal phase. However, the exact role of these variations in terms of female fecundity remains a matter of debate.[47]

Assessment of Anatomic Dysfunction

Some type of mechanical factor relevant to infertility is reported to occur in 10% to 30% of infertile couples. Several methods exist to establish abnormalities at the level of the uterus, fallopian tubes, or ovary/ampulla complex relevant for *in vivo* oocyte pickup. Until recently, diagnosis of mechanical abnormalities by hysterosalpingography (HSG) during the early infertility workup and by laparoscopy later in the evaluation has been the standard procedure for each patient. Currently, the tendency is toward the use of less invasive screening tests to identify women at high risk for tubal disease and the performance of more invasive diagnostic procedures only in those women at increased risk for such abnormalities.[48] The question remains as to what extent any mechanical deviation from normal truly contributes to the infertility. The clinician should take this reservation seriously, especially in cases of minor abnormalities of the uterine cavity and of presumed unilateral tubal damage.

History Taking. The guidelines of the National Institute for Clinical Excellence (NICE) in the United Kingdom advise clinicians to use the past medical history in the woman to decide whether invasive diagnostics should be applied to assess the presence of tubal pathology (especially previous deliveries, previous pelvic inflammatory disease, or previous pelvic surgery) enables the identification of such couples.[49] The addition of *Chlamydia* antibody testing will further enhance the correct identification of couples at such low risk for anatomic dysfunction that invasive diagnostics can well be avoided.

Transvaginal Ultrasonography. Today, transvaginal ultrasonography (TVS) is standard in the infertility workup, as it is simple, highly informative and allows for a test with low patient burden. With its use, several relevant abnormalities can be assessed. These comprise uterine cavity abnormalities such as thin endometrium lining, endometrial polyps, submucous myoma, and intrauterine adhesions. Secondly, the status of the myometrium can be assessed allowing for the detection of intramural or subserous myomas, adenomyosis, or congenital anomalies such as a uterine septum, a bicornate or a didelphys uterus. Only pathologic conditions of the fallopian tubes, e.g., hydrosalpinx, can be diagnosed by TVS, due to the presence of intraluminal fluid trapped in a terminally occluded salpinx. Finally, ovarian imaging will reveal information on the presence of cysts (dermoid, endometriosis), polycystic ovarian morphology, as well as the total antral follicle count (AFC), which may indicate the ovarian reserve status. In typical cases of dysmenorrhea, indirect signs of adhesions based on endometriosis can be identified between the rectum and uterus and uterus and bladder.

Screening by *Chlamydia* Antibody Testing in Serum. *Chlamydia trachomatis* represents the most common sexually transmitted pathogen in the Western world, with a reported prevalence of infection of around 5% to 10%. Most of these infections are asymptomatic, but ~10% can cause pelvic inflammatory disease.[50] *Chlamydia* acts as an immunogen that gives rise to antibody formation. Immunoglobulin G (IgG) antibodies remain in the circulation for years and therefore can be used as a marker for past infection. IgG concentrations are reportedly elevated in 30% to 60% of women from infertile couples. The capacity of IgG antibody levels to predict tuboperitoneal disease (as assessed by laparoscopy) has been evaluated prospectively, with acceptable receiver operator characteristic (ROC) curves and the use of titers between 16 and 32.[51] In infertile couples with a relatively good fertility prognosis based on the Hunault prediction rule,[52] a workup strategy starting with history taking and screening for *Chlamydia* antibodies is an effective approach,[53,54] and for many women with only a low risk, performance of more invasive tubal patency evaluating procedures can be prevented.

Hysterosalpingography. HSG represents the radiographic visualization of the endocervical canal, uterine cavity, and lumina of the fallopian tubes. Both oil- and water-based contrast media are used. The oil-based medium may provide the advantage of a minor subsequent increase in spontaneous pregnancy chances,[55,56] whereas the water-based medium is faster and less painful and results in improved visualization of abnormalities. Abnormal uterine features include filling defects (related to submucous fibroids, polyps, or intrauterine adhesions), and abnormal shape (congenital uterine anomalies such as unicornuate, bicornuate, or septate uterus, uterus didelphys, fetal exposure to diethylstilbestrol, or intrauterine adhesions). In addition, unilateral or bilateral tubal occlusion, with or without hydrosalpinges, may be diagnosed. The occurrence of peritubal adhesions may be suspected on the basis of an abnormal intraabdominal distribution pattern of the contrast medium (pocketing, especially noted on the post-filling x-ray).

The HSG procedure should be planned during the follicular phase of the menstrual cycle to avoid an unexpected early pregnancy. The procedure can be quite painful, and an ascending infection with acute salpingitis is the most significant complication, occurring in ~1% to 2% of patients. Preventive antibiotic prophylaxis should be advised in all patients for whom tubal pathology has been diagnosed or suspected. HSG is reasonably accurate in diagnosing tubal occlusion; abnormal results are associated with reduced fecundity rates, especially if abnormalities are double-sided.[57,58] An important drawback of this procedure remains the subjective nature of the evaluation, which renders the reproducibility of this test rather poor.

Hysterosalpingo Contrast Sonography (HyCoSy). Hysterosalpingo contrast sonography (HyCoSy) is a simple and well-tolerated outpatient ultrasound procedure and requires the use of a contrast agent to visualize patency of the fallopian tubes. With a small balloon catheter as filling device and using a transvaginal ultrasound, the contrast agent is instilled into the uterine cavity. The contrast medium may be an agitated saline/air mixture or a non-iodinated contrast agent (ExEm Foam). The contrast flow from the uterine cavity into the fallopian tubes can be visualized by the operator and allows judging rapid passage through the length of the fallopian tubes, indicative of tubal patency. Although HyCoSY may be a promising alternative for HSG with regard to accuracy and effectiveness, proper assessment studies are yet to appear.[59]

Laparoscopy. This procedure is considered the gold standard for the diagnosis of tubal disease and intraabdominal disorders such as endometriosis and adhesions that interfere with normal fertility.[57,58] Some authorities believe that laparoscopy should be performed before unexplained infertility is diagnosed. More recent data suggest that the added value of such a procedure in women without a history suggestive of tubal disease and a normal HSG is very limited.[60]

Laparoscopy involves the intraabdominal inspection of the internal genital organs after abdominal insufflation with CO_2 of a patient under general anesthesia. Next to the reproductive organs, the appendix should be inspected for infection and the perihepatic region for adhesions suggestive of past chlamydial infections (Fitz-Hugh-Curtis syndrome). In experienced hands, risks associated with this technique are minimal. Attention should be focused toward standardizing the evaluation of the abdominal and pelvic cavities.

Other Techniques. Recently introduced, more sophisticated techniques that can be used to visualize intrauterine, intratubal, or intraabdominal abnormalities include ultrasound imaging of the uterine cavity after fluid lavage and endoscopic techniques such as saline infusion sonography, hysteroscopy, salpingoscopy, falloposcopy, transvaginal hydrolaparoscopy, and microlaparoscopy.[61] Although some of these techniques have seemed promising, with their focus on reduced patient discomfort and costs, the usefulness of these tools in a standard infertility workup remains to be established.

Assessment of Cervical Mucus–Sperm Interaction

In humans, semen is deposited in the posterior fornix of the vagina, and the cervix via its mucus secretion is believed to play an important role in allowing sperm to enter the uterine cavity. Several functions of cervical mucus include protecting sperm from the acidic milieu in the vagina, allowing the sperm to enter the cervix, supporting energy requirements for sperm motility, and filtering morphologically abnormal sperm. Before ovulation, cervical mucus is produced in large quantities, allowing easy sperm access at the most appropriate moment during the menstrual cycle. The prognostic performance of cervical mucus tests, the incidence of cervical factor infertility ("cervical hostility"), and the validity of treatment options to correct disturbed cervical mucus–sperm interactions remain poor.

Postcoital Test. This test, which involves appraisal of the presence of motile sperm in cervical mucus, after intercourse during the late follicular phase of the menstrual cycle, was originally described in 1866 by Sims and was reemphasized in 1913 by Huhner (reviewed in ref 62). There is no agreement regarding the preferred interval between intercourse and the test, nor the cutoff between a normal and an abnormal test result. The threshold is 1–5 spermatozoa per high-power magnification (i.e., 10 × 40) field. A systematic review of 11 well-designed studies indicated that the discriminating ability of the postcoital test is poor, regardless of the applied criteria for normality.[62,63] A subsequent systematic prospective evaluation indicated that inclusion of a postcoital test in routine evaluation of new couples only results in additional interventions.[62] Several national and international bodies (such as the American Society of Reproductive Medicine, the European Society of Human Reproduction, and the Royal College of Obstetrics and Gynaecology) currently advise against performing the postcoital test in routine infertility evaluation. In addition, earlier tests used to evaluate the quality of the cervical mucus, along with several *in vitro* tests of interactions between cervical mucus and sperm, have been abandoned.

Assessment of Ovarian Aging

It is well established that the female partner's age is the most prominent factor in determining chances of spontaneous pregnancy[33] or after infertility therapy (such as insemination with donor sperm[64] or with *in vitro* fertilization–embryo transfer [IVF-ET]).[65] As discussed previously, however, chronologic age has limitations in that the degree of ovarian aging varies considerably among women in the same age group. This is true for the quantitative aspects of ovarian reserve, expressed by the AFC and AMH levels. Some 35-year-old women may be normally fertile (i.e., have the fecundity of a 25-year-old), and expectant management may be the most appropriate option. In contrast, natural conception chances may already be severely compromised in other women of the same age (with fecundity of a 41-year-old), and the clinician would best offer the most effective treatment options without delay.

The major challenge of today is to assess the magnitude of ovarian aging on an individual basis.[66,67] Endocrine and ultrasound changes go along with ovarian aging. Recent efforts have focused on the clinical implications of these changes, applying test systems that aim to predict chances for pregnancy, either spontaneously or after infertility therapy. Some factors indeed may have some capacity to predict fertility chances, although studies on the added value of these tests on top of female age are either scarce or inconclusive as to their value in patient management.

While initial physiologic studies in normo-ovulatory women have emphasized that the length of the follicular phase is shortened with increasing age and with elevated early follicular phase serum FSH concentrations, baseline FSH has limited capacity to predict IVF outcome, independent of chronologic age.[68] In addition, a decreased cohort of AFC can be observed by ultrasound and used as a marker of ovarian aging.[69] However, a meta-analysis on the predictive value of baseline FSH and AFC for IVF outcome concluded that their clinical value in pregnancy prediction is restricted to only a small group of patients.[70]

In recent years, AMH has become the marker of promise in ovarian reserve assessment. Serum levels are proportional to the number of antral follicles in the size range of 1 to 8 mm, and indirectly represent the size of the pool of primordial follicles.[71,72] Compared with AFC and basal FSH, the current AMH test has unfortunately failed to significantly improve prediction of pregnancy after IVF-ET.[73,74]

For the prediction of poor response to OS, especially AMH has proven quite adequate. A poor response may indicate advanced ovarian aging, with inherent reduced prospects for a live birth, but underdosing or decreased sensitivity of the ovary may also play a role. In cases where poor ovarian response is due to exhaustion of the ovarian follicle pool, the prediction of poor response has no therapeutic implications because the prognosis for successful treatment will be poor, regardless of the chosen intervention. It needs to be emphasized, however, that especially in younger patients, oocyte quality may still be quite good in spite of low numbers obtained by OS, even if this is caused by low ovarian reserve status.[75] At the other end of the spectrum, the capacity of these tests to predict excessive response to ovarian hyperstimulation allows treatment schedules to be adjusted in order to mitigate the number of stimulated follicles and thereby limit the risk for ovarian hyperstimulation syndrome (OHSS).[76–78]

Today, initial testing of ovarian reserve may be useful especially for managing high responses,[79–81] as a reduced gonadotropin dose in high responder patients will generate less oocytes with adequate IVF success rates; this approach reduces chances for developing OHSS. However, with the use of gonadotropin-releasing hormone (GnRH) antagonists for LH peak suppression, high responders can also be managed by GnRH agonist trigger of oocyte maturation and a "freeze all" embryos strategy.

From Diagnosis to Prognosis

In many couples, a clear cause for infertility remains unknown with currently available diagnostic tools. Various prospective follow-up studies evaluated whether specific characteristics assessed upon initial evaluation could predict the chance for a spontaneous pregnancy in these couples. Three studies[82–84] identified similar factors predicting pregnancy chances for couples with unexplained infertility. The Hunault prediction model[36,37] based on duration of infertility, female age, previous pregnancies, sperm motility, and referral status for couples with unexplained infertility has been externally validated, demonstrating its accuracy in prognosis assessment.[85] Application of this model[37] to assess chances for spontaneous pregnancy will help counseling couples with unexplained infertility regarding the preferred strategy, i.e., expectant management versus initiation of (empirical) treatment. Today, the option has been further developed to estimate the changes in the prospects for an individual couple over time, thereby facilitating potentially the right timing of initiation of infertility treatment. This approach may reduce overtreatment, in that invasive therapies may be avoided in couples who have a still favorable likelihood of spontaneous pregnancy. Notably, a well-designed cohort follow-up study concluded that 61% of couples achieved pregnancy independent of treatment.[85,86] Similar prediction models have been developed aiming to assess cumulative pregnancy changes throughout an entire IVF treatment.[87]

KEYPOINTS

- The purpose of the infertility work-up in the female is to identify clear causes, such as tubal pathology and/or oligo/an-ovulation.
- Invasive evaluation of uterine-tubal abnormalities is best reserved for cases with a positive history (STDs and pelvic surgery), abnormalities at transvaginal ultrasonography and/or presence of Chlamydia antibodies in the serum.
- Assessment of ovarian reserve status is not meaningful in terms of diagnosis or prognosis in the infertility work-up. Its value stems from the prediction of ovarian response to ovarian hyperstimulation in assisted reproduction technology.

EXPECTANT MANAGEMENT OF THE INFERTILE COUPLE

The chance that a spontaneous pregnancy will occur after unprotected intercourse for 1 year is around 90%[88] (see Fig. 123.1. This cumulative pregnancy rate is highly influenced by female age. When a couple seeks evaluation of infertility, it is important first to rule out causes of absolute infertility (i.e., tubal factor, anovulation, and severe sperm abnormalities). Thereafter, the likelihood of a spontaneous pregnancy should be assessed, before any type of empirical therapy is begun. As mentioned previously, the age of the woman and the duration of infertility are the most important factors affecting the future likelihood of conception (see Figs. 123.3 and 123.4, and Table 123.2). With the use of prognostic models applied in unexplained infertile couples, a predicted chance of > 30% live birth in the year to come may urge physicians to promote expectant management as the first-line treatment, specifically as studies have shown that compared to treatment with mild OS combined with IUI no improvement in live birth rates will be obtained. Whether in such cases direct application of IVF will make a true benefit still remains to demonstrated. Comparing different treatment options in couples with an unfavorable prognosis (< 30% predicted one year live birth) revealed that IVF did not result in improved outcomes compared to IUI with mild stimulation.[89]

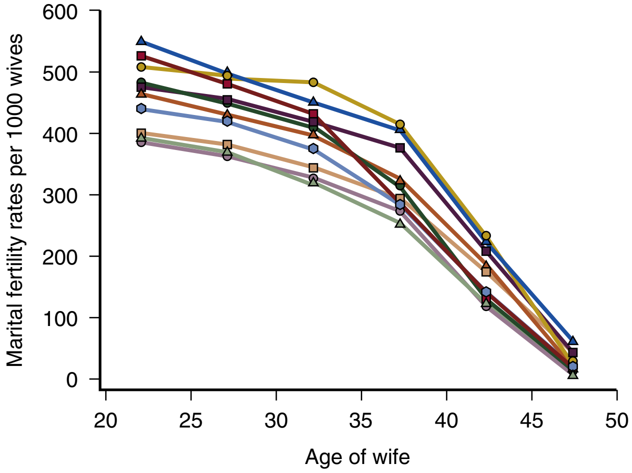


Figure 123.4 **Natural marital fertility rates as a function of age of the woman, indicated for different populations from the sixteenth until the twentieth century (Hutterites, Geneva bourgeoisie, Canada, Normandy, Tunis, Norway, and Iran).** (Data from Menken J, Trussell J, Larsen U, et al. Age and infertility. *Science*. 1986;233:1389–1394; Schwartz D, Mayaux MJ. Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS, *N Engl J Med*. 1982;306:404–406.)

Increasing patient pressure and the financial interests of some infertility centers have prompted starting infertility treatment prematurely. The approach of comparing spontaneous pregnancy chances with chances following intervention may aid in the proper timing to initiate therapy, which may prevent overtreatment. This approach is distinctly different from one that focuses on the highest chances for pregnancy per treatment cycle, which is often used.

The definition of cost-effective infertility care is achieving the desired end point at the lowest possible expenditure of resources.[90] Analysis modeling in the United Kingdom confirmed that IVF is a cost-effective treatment option for severe tubal factor infertility.[91] Initial cost-effectiveness studies also support the use of IUI and mild ovarian hyperstimulation before IVF is initiated, although this may apply only for couples with unexplained infertility who have a 1-year prognosis of at best 30% for an ongoing pregnancy.[92–94] Gradually, a consensus is evolving that a healthy singleton live birth per started treatment should be the preferred end point.[95,96] OS in normo-ovulatory women is inherently associated with multiple follicle development and increased chances for multiple pregnancies and should therefore be applied with great caution. In addition, a systematic evaluation documented that success rates of infertility therapies are often overstated and that the quality of relevant clinical studies is often questionable.[97]

Anovulation usually can be treated effectively with ovulation induction agents (such as CC, the aromatase inhibitor letrozole [LTZ], or exogenous gonadotropins) aimed at restoring normal ovarian function (see Chapter 117). In addition, mechanical factors involving the uterine cavity, tubal function, or oocyte pickup can be restored by endoscopic surgery (such as hysteroscopic myomectomy, adhesiolysis, or fimbrioplasty).

The diagnosis of unexplained infertility can be made only after an appropriate diagnostic infertility workup has established the absence of any detectable abnormalities.[98] Because no consensus has been reached regarding the extent of testing that should be performed during an

TABLE 123.2 Summary of Cochrane Review Concerning the Efficacy of *In Vitro* Fertilization (IVF) for Unexplained Infertility

Comparison	Number of Cases	Ongoing Pregnancy	Rate per Couple	Odds Ratio	95% CI
IVF vs. Expectant Management		**IVF**	**Expectant**		
Hughes 2004[143a]	51				
Soliman 1993[13]	35				
Overall	86	29%	12%	3.24	1.07–9.80
IVF versus IUI					
		IVF	**IUI**		
Goverde 2000[144]	113				
Overall	113	41%	26%	1.96	0.88–4.36
IVF vs. IUI/Stimulation					
		IVF	**IUI Stimulation**		
Goverde 2000[144]	118				
Overall	118	41%	37%	1.15	0.55–2.42

The table shows that randomized clinical trials demonstrating the efficacy of IVF as first-line therapy in unexplained infertility are scarce and cover only limited numbers of cases. The application of intrauterine insemination (IUI) with or without stimulation as first-line treatment is to be recommended in view of lower costs and patient discomfort.

Data from Pandian Z, Bhattacharya S, Vale L, et al. In vitro fertilisation for unexplained subfertility. *Cochrane Database Syst Rev*. 2005;2:CD003357.

infertility workup, cases of unexplained infertility involve a notoriously heterogeneous group of patients. In many cases of unexplained infertility or male factor infertility, patients undergo empirical treatment to improve pregnancy chances. Assisted reproduction encompasses various techniques used to increase the chances of pregnancy per cycle by basically bringing more gametes closer together. This goal can be achieved by OS, combined with IUI or IVF.

Ovarian Stimulation

Under normal conditions, a cohort of follicles is recruited for ongoing gonadotropin-dependent development during the inter-cycle rise in FSH. During a limited time interval (the "*FSH window*"), FSH concentrations rise above the threshold before declining[99] (see Chapter 118). The aim of OS, with exogenous FSH as the most frequently used drug, is to intervene in mechanisms that regulate single dominant follicle development, in order to mature multiple follicles via generation of multiple oocytes for fertilization *in vivo* (after intercourse or IUI) or *in vitro* (IVF).

OS in IUI cycles can be achieved using CC, LTZ, or exogenous gonadotropins. Medication is usually administered during the early or midfollicular phase (cycle days 2 through 7 or 5 through 9) in daily dosages ranging between 100 and 150 mg for CC and 2.5–5 mg for LTZ. Exogenous gonadotropins are used in a fixed dose, starting with 50–75 IU/day on day 2 to 3 of the cycle. Depending on the number of dominant follicles > 12 mm that develop, this dosage can be adjusted in a subsequent cycle, using steps of 25–37.5 IU. The use of GnRH agonist cotreatment to avoid a premature LH rise generally is not recommended in IUI cycles.

Along with a possible beneficial effect on the ongoing pregnancy rate, OS may lead to multiple pregnancies and OHSS, both of which should be considered preventable complications of treatment (Fig. 123.5).[100] When multiple oocytes are released *in vivo*, the number of oocytes fertilized cannot be controlled, and a marked increase in multiple (especially higher-order multiple) pregnancies has occurred over the past two decades.[101,102] Development of more than two dominant follicles has been shown to enhance the rate of multiple pregnancies, without a gain in overall ongoing pregnancy rates.[103] The results of OS are often combined with those of IUI, and few studies allow differentiation between the independent effects of the two.[102–104] Specifically, when judging the efficacy of these two treatment components, several factors need to be considered. The infertility diagnosis (unexplained, male, or cervical factor), the age of the female, and the prognosis level for spontaneous pregnancy according to the Hunault model will all affect the decisions on which treatment modality to apply. Currently, initiating treatment in couples with unexplained or moderate to mild male factor infertility and a 1-year prognosis of ≥ 30% is not advocated. If treatment is indicated, IUI in natural cycles will increase conception rates in couples with cervical factor infertility, preventing exposure to multiple pregnancy risks. In unexplained, but even in mild male factor infertility (total motile sperm count > 10 million) or minimal endometriosis, applying only IUI in the natural cycle may not be sufficiently effective over expectant management,[104] implying that only the combination with OS will clearly benefit over timed intercourse. The combined use of IUI and OS with exogenous gonadotropins or CC for unexplained infertility may bypass various subtle barriers, such as minimal sperm abnormalities, sperm–cervical mucus interaction, and the timing of sperm delivery, and may further enhance fertility through the ovulation of more than one follicle. Notably, success in these treatment modalities is further affected by female age such that applying IUI with OS may be omitted in women aged > 40 years.

The procedures during the treatment itself may be also considered as to their role in the achieved efficacy. OS should lead to the release of two to three oocytes maximally to maintain the proper balance between success and complications. Because of low cost and ease of administration, the antiestrogen CC could be recommended as first-choice medication for OS, although gonadotropins have appeared to be more effective.[105] It should be noted that limited or even absent efficacy has been suggested recently[106] for any drug (CC, gonadotropins, or aromatase inhibitors) that is used as monotherapy without IUI in unexplained infertility.[107] In stimulated cycles, the additional use of GnRH antagonists or agonists for LH peak suppression have shown not to be cost-effective, while the use of luteal support with vaginally applied progesterone may be of benefit but is awaiting larger trials to demonstrate a true effect.

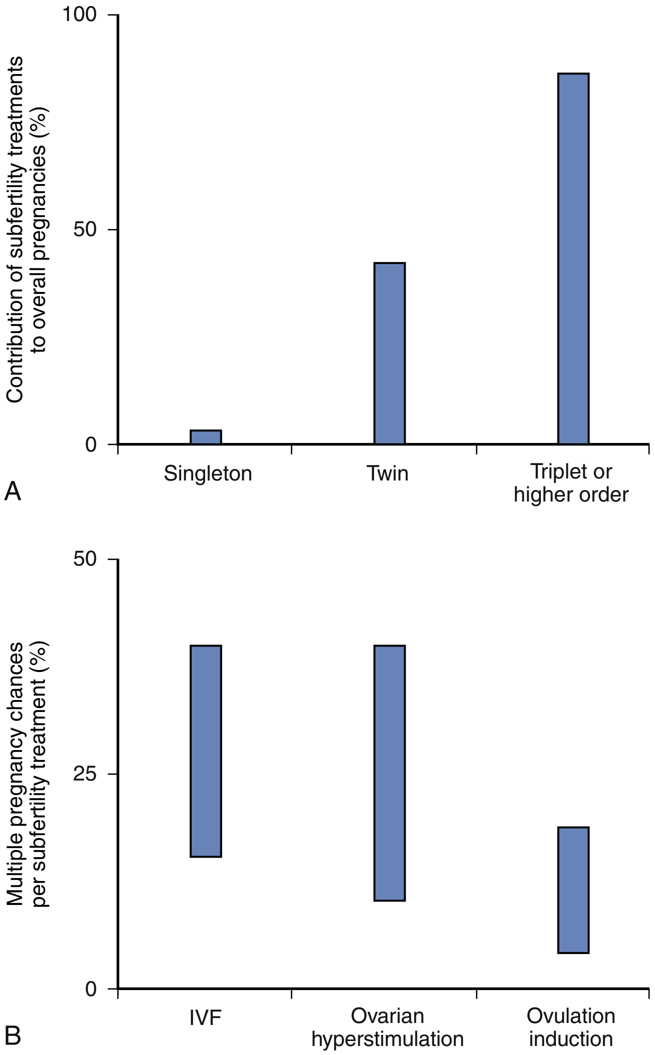


Figure 123.5 Contribution of infertility treatments to overall pregnancies (**A**) and reported frequency of multiple pregnancy in relation to *in vitro* fertilization (IVF), ovarian hyperstimulation, and ovulation induction (**B**). The graphs show that assisted reproduction contributes considerably to the rate of multiple pregnancies in society and affects singleton rates only minimally. Also, ovarian (hyper)stimulation without access to the choice of how many embryos may be implanting contributes heavily to the totals of multiples pregnancies from infertility treatments. (Data from Fauser BC, Devroey P, Macklon NS. Multiple birth resulting from ovarian stimulation for subfertility treatment *Lancet.* 2005;365:1807–1816.)

In case the total number of motile sperm inseminated is < 5 million, IUI treatment will hardly lead to pregnancies and is best abandoned. With sufficient sperm quantity, acceptable cumulative ongoing pregnancy rates after six cycles of IUI with OS can be obtained, with the highest yield in the first three cycles and stable cycle rates up until the ninth cycle. Timing of the insemination using LH surge detection in natural cycles is preferred over the use of human chorionic gonadotropin (hCG) triggering, and generally a time interval of 12 to 36 hours is optimal, depending on the triggering mode applied.

Intrauterine Insemination

IUI refers to placing sperm, after a washing procedure in the laboratory, in the uterine cavity just before ovulation. The rationale for IUI combined with OS is to increase gamete density at the site of fertilization (i.e., the fallopian tube). Hence, the classic indications for IUI are male (erectile and ejaculatory) or female (vaginismus) sexual dysfunction and cervical hostility. IUI combined with mild OS has been applied in couples suffering from male factor and unexplained infertility.

Renewed interest in IUI reflects the evolution of better washing techniques (i.e., gradient or swim-up) to improve the quality of the initial sperm sample by removing prostaglandins, infectious agents, and antigenic proteins, and nonmotile cells (leukocytes) or immotile and immature spermatozoa and thus decreased cytokines and lymphokines and subsequent reduction in free oxygen radicals.[108] IUI is usually performed within 30–36 hours after hCG administration or 20–28 hours after the LH surge.[109] It is performed using a small flexible catheter through which a small volume (0.3 to 0.5 mL) of culture medium containing the washed spermatozoa is delivered into the uterine cavity. When IUI is used to manage fertility problems, single rather than double insemination should be used.[110]

IUI combined with OS in the treatment of infertility remains a topic of ongoing debate. A recent literature review has made clear that treatment with IUI with OS as well as IUI alone probably results in a higher cumulative live birth rate than expectant management in couples with a low prognosis for natural conception. IUI in a stimulated cycle may outperform IUI in a natural cycle.[104] Pregnancy chances as such may increase, but inherent associated risks include a considerable chance of multiple pregnancies, additional costs, and increased patient discomfort and chances for complications.[103]

A large retrospective analysis in a single European center, which assessed 1878 pregnancies attained through IUI combined with gonadotropin stimulation, reported a 16% twin and a 6% higher-order multiple pregnancy rate[111] (see Fig. 123.5). Also, a 2-year follow-up in a US infertility center involving 3347 OS cycles found a high multiple pregnancy rate, with an overall pregnancy rate of 30%.[112] Reducing the risk for a multiple pregnancy may result from restricting the number of dominant follicles allowed at the time of ovulation trigger. From a nationwide survey in the Netherlands, twin rates appeared to be as low as 9% per pregnancy, albeit at the cost of some reduction in the overall chance of ongoing pregnancy.[89,113,114]

Although births resulting from assisted reproduction represent only 1% of all live births in the United States, they now account for 35% of twin births and for more than 40% of triplet and higher-order multiple births.[101] The adverse impact of multiple births on perinatal morbidity and mortality has been clearly documented and should lead to management of the infertile couple that increases chances for pregnancy without causing complications that have lifetime consequences.[115,116]

Cost-effectiveness analyses so far have led to the conclusion that IUI, with or without OS, should precede IVF.[93,117] These assessments were made, however, before the advent of current approaches to minimize costs and minimizing complications in IVF through minimal OS and transfer of a reduced number of embryos.[95]

As pregnancy chances per cycle remain low, it can be calculated that the number needed to treat for patients to benefit from any additional intervention is high. For instance, when the combination of FSH and IUI is compared with IUI alone, 31 treatment cycles are required before there would be one more singleton live birth.[118] Offset against the additional pregnancy are the added costs of medications and of frequent ovarian response monitoring. When costs related to multiple pregnancies are also taken into consideration, the conclusion seems justified that widespread use of OS with exogenous gonadotropins combined with IUI should be discouraged. The real question remains whether society (especially patients, physicians, and health insurance companies) considers the high rate of multiple pregnancies an acceptable price to be paid for increased pregnancy chances per cycle. Thus, current international guidelines (NICE) recommend the use of IUI in the natural cycle as standard first-line treatment for couples with unexplained or mild to moderate male factor infertility, although the efficacy of such treatment clearly may be doubted. As the risk for a multiple pregnancy may become reduced by applying CC or LTZ as an OS agent, the widespread use of FSH OS in IUI treatment may need to be reconsidered. Recent and currently ongoing studies will provide answers to the preferred treatment modality.[119] Beginning treatment with CC/IUI will result in pregnancy in nearly one fourth of the couples, with minimal risk for multiple births. With IVF procedures using a high proportion of single-embryo transfer, eliminating gonadotropin/IUI will result in pregnancies with the lowest possible risk for multiple births. Compared with conventional infertility treatment, an accelerated approach to IVF after the initial CC/IUI treatment results in a shorter time to pregnancy with fewer treatment cycles, and cost savings.[120]

In Vitro Fertilization

IVF involves OS, oocyte retrieval, fertilization of oocytes *in vitro* (either spontaneously or by ICSI), and subsequent transfer of embryos into the uterine cavity. Aspects of IVF relevant to infertility management are discussed here.

IVF was originally designed for treatment of bilateral tubal occlusion, but currently the "tubal factor" represents only a small proportion of indications for IVF. At present, it is applied most often in cases of mild to moderate male factor and unexplained infertility. A meta-analysis of five relevant trials in unexplained infertility confirmed that no difference in pregnancy rates could be observed when IVF was compared with IUI, with or without ovarian hyperstimulation in treatment-naïve infertile couples (odds ratio: 1.96; 95% CI: 0.88–4.36).[102,121] This implies that first-line treatment may best be carried out using mild OS combined with IUI across a series of 6–9 cycles. The next step would then be IVF, although the added benefit has not been exactly assessed. Two separate studies indicated that initial treatment with IVF is not cost-effective when compared with a conventional treatment algorithm.[121–123]

Many advances in IVF, such as improved cryopreservation programs, blastocyst culture, techniques of the embryo transfer, and the widened application of GnRH antagonists cotreatment may have improved the efficiency over the past decade, along with improved safety by applying single-embryo transfer and better OHSS management.[124]

Tubal occlusion can give rise to hydrosalpinges. In women undergoing IVF, the presence of hydrosalpinges is associated with early pregnancy loss and poor implantation due to altered diminished endometrial receptivity. Thus, salpingectomy or at least occlusion of the isthmic part of the tubes should be offered before IVF when hydrosalpinges are present.[125]

Similarly, endometriosis-associated infertility reduces pregnancy rates. IVF represents an effective means of bypassing the hostile peritoneal environment and anatomic distortion associated with this disease. Pretreatment with GnRH agonists immediately before IVF initiation appears to be beneficial. It is uncertain whether only a specific subset or all patients with endometriosis would benefit from this approach; the use of endometrial implantation markers may be helpful in this regard.[28,126]

The incidence of multiple gestations in IVF is easier to control than OS with or without IUI, because in case of IVF the occurrence of multiples is primarily dependent on the number of embryos transferred. Increased emphasis is being directed toward decreasing the number of embryos for transfer and, therefore, diminishing the chances of (higher-order) multiple pregnancies. In a comparison of conventional maximal OS and replacement of two embryos, the application of milder OS with single-embryo transfer resulted in the same number of ongoing pregnancies leading to term life birth after a 1-year treatment period. Although on average one additional treatment cycle was required, the long-term medical costs and patient burden were the same.[96,127]

KEYPOINTS

- In unexplained infertility with a prognosis for spontaneous ongoing pregnancy of over 30% for the subsequent year, expected management is justified as infertility treatment is not likely to yield better prospects for the couple.
- In couples with unexplained infertility and a prognosis below 30% for the subsequent year, and in couples that have applied a period of 6-12 month of expected management without achieving an ongoing pregnancy, mild ovarian stimulation combined with intra uterine insemination is the preferred first line treatment.
- In vitro fertilization and related technology is considered the last resort in treatment of couples with unexplained infertility. It is however the first line treatment in couples with a clear cause such as tubal pathology. In couples with oligo-/an-ovulation, failed treatment by ovulation induction may be a justified indication to apply in vitro fertilization and related technology.

SUMMARY AND FUTURE DEVELOPMENTS

Infertility represents a serious health issue and is considered as a disease by WHO. Notably, most people worldwide have insufficient access to infertility and fertility care. Postponement of pregnancy beyond 30 years of age, as is observed in the Western world, is giving rise to an increased demand for infertility services, because more couples will fail to conceive spontaneously within 1 year of unprotected intercourse. The major clinical challenge remains to assess the reproductive aging status of a given woman. This will help to predict her pregnancy chances more accurately compared with assessment based on chronologic age alone. The use of multivariate models may further improve the prediction capacity of initial screening for a given couple. The duration of infertility and a woman's age constitute the most important predictors of chances for achieving a spontaneous pregnancy. This approach should help when couples are counselled regarding the likelihood of success of expectant management, or whether (empirical) treatment should be initiated.

The diagnosis of unexplained infertility can be made only after a proper diagnostic infertility workup has established the absence of any abnormalities. Because no consensus exists regarding the extent of testing that should be performed, unexplained infertility represents a notoriously heterogeneous group of patients. This condition may include couples with undiagnosed infertility and those with normal fertility for whom pregnancy, purely by chance, has not yet occurred. These patients often undergo empirical treatment to improve pregnancy chances per cycle. The inherent disadvantage of such an approach is decreased interest in the diagnostic workup, which impairs progress in our understanding of the underlying pathophysiology.

When costs and prematurity of newborns related to multiple pregnancies are also taken into consideration, the conclusion is justified that widespread use of OS with exogenous gonadotropins with or without IUI should be discouraged. With IVF, increased emphasis is directed toward decreasing the number of embryos for transfer and, therefore, diminishing the chances of (higher-order) multiple pregnancies. Additional cohort follow-up studies using appropriate end points, such as singleton, term delivered, and live birth rate, are required to elucidate further the optimal approach to empirical treatment of patients with unexplained infertility.

Numerous innovations in ART have emerged and are currently being evaluated for its clinical efficacy and safety. However, the stakes are high and novel technologies are often introduced prematurely in everyday patient care without sufficient scientific proof of safety and efficacy.

The most important recent developments in ART may be listed as follows:

1. Improved vitrification technology, allowing oocytes, embryos, and gonadal tissue to be cryopreserved with increased efficacy.
2. Blastocyst transfer; transition from the transfer of embryos into the uterus after 2 or 3 days of culture toward development up to the blastocyst stage on day 5 or 6 generating improved implantation and ongoing pregnancy rates.
3. Time laps monitoring; embryos are constantly being monitored for development potential in special incubators, using camera images and sophisticated software technology. Its clinical usefulness remains doubtful.
4. PGT-A; preimplantation genetic testing for aneuploidy. Aneuploidy screening can now be performed on a large scale using sophisticated next-generation sequencing of a number of cells removed from an early stage embryo before transfer, so only euploid embryos can be transferred. Potentially, this technology could also improve implantation rates, decrease miscarriage rates following IVF and therefore improve IVF success rates and time to pregnancy. However, again its clinical usefulness remains highly controversial.
5. A GnRH agonist bolus rather than hCG to stimulate final oocyte maturation at the end of the stimulation phase and prior to oocyte retrieval. This approach considerably reduced the risk for OHSS but can only be performed with GnRH antagonist cotreatment during OS.
6. Single embryo transfer; this policy initially developed in Northern Europe and is gradually being applied by the rest of the IVF community. Multiple pregnancies can be prevented by the transfer of one embryo at a time, significantly reducing pregnancy complications and improving perinatal outcomes. Supernumerary embryos can be cryopreserved and individually transferred in subsequent cycles. National data from various countries now demonstrate multiple pregnancy rates following IVF can be < 5%, with similar overall pregnancy rates. At present, in some countries approximately 50% of all children born are from cryopreserved embryos.
7. Freeze all embryos; OS for IVF gives rise to a supraphysiological endocrine milieu during the follicular phase of the cycle. Increasing evidence suggests that this may also impact endometrial receptivity and hence embryo implantation rates may be reduced. For these reasons, increasingly, IVF centers are applying "freeze all" strategies, where only embryos are transferred in subsequent cycles after freezing and thawing. The jury is still out whether this strategy will truly improve IVF outcomes.
8. Widespread use of fertility preservation strategies in oncology patients aiming to preserve their fertility potential once cured from the primary disease. Despite the fact that this technology is more and more available, there still seems much room for growth in the area. Even today, many young cancer patients are not being made aware of this option.

9. Nonmedical oocyte freezing; programs are offered worldwide to cryopreserve oocytes from young potentially fertile women who would like to have children later in life once circumstances are right. It is heavily debated whether this option is actually helping women by increasing their reproductive choices, or whether this development represents yet another example of exploitation of reproductive care.

KEYPOINTS

- Assisted Reproduction Technology as treatment for female infertility has evolved by creating numerous technical improvements in order to make this treatment more effective, safer and more bearable for the patient.
- A special note are the possibilities that this technology offers for fertility preservation strategies as well as the growing role in the prevention of hereditary disease through preimplantation genetic testing of embryos.

REFERENCES

1. Zegers-Hochschild F, Adamson GD, Dyer S, et al. The international glossary on infertility and fertility care: 2017. *Hum Reprod*. 2017;32:1786–1801.
2. Chua SJ, Akande VA, Mol BW. Surgery for tubal infertility. *Cochrane Database Syst Rev*. 2017;1:CD006415.
3. Evers JL, Collins JA, Clarke J. Surgery or embolisation for varicoceles in subfertile men. *Cochrane Database Syst Rev*. 2008;3:CD000479.
4. Edwards RG, Steptoe PC, Purdy JM. Establishing fullterm human pregnancies using cleaving embryos grown in vitro. *Br J Obstet Gynaecol*. 1980;87:737–756.
5. Palermo G, Joris H, Derde MP, et al. Sperm characteristics and outcome of human assisted fertilization by subzonal insemination and intracytoplasmic sperm injection. *Fertil Steril*. 1993;59:826–835.
6. Bhattacharya S, Kamath MS. Reducing multiple births in assisted reproduction technology. *Best Pract Res Clin Obstet Gynaecol*. 2014;28:191–199.
7. Baart EB, Martini E, Eijkemans MJ, et al. Milder ovarian stimulation for in-vitro fertilization reduces aneuploidy in the human preimplantation embryo: a randomized controlled trial. *Hum Reprod*. 2007;22:980–988.
8. Nargund G, Fauser BCJM. Mild ovarian stimulation for IVF is the smartest way forward. *Reprod Biomed Online*. 2020;S1472–6483:30280–30287.
9. Cil AP, Seli E. Current trends and progress in clinical applications of oocyte cryopreservation. *Curr Opin Obstet Gynecol*. 2013;25:247–254.
10. Rao GD, Tan SL. In vitro maturation of oocytes. *Semin Reprod Med*. 2005;23:242–247.
11. Donnez J, Dolmans MM. Fertility preservation in women. *N Engl J Med*. 2017;377:1657–1665.
12. Sermon K. Novel technologies emerging for preimplantation genetic diagnosis and preimplantation genetic testing for aneuploidy. *Expert Rev Mol Diagn*. 2017;17:71–82.
13. Evers JL. Female subfertility. *Lancet*. 2002;360:151–159.
14. Bagshawe A, Taylor A. ABC of subfertility: counselling. *BMJ*. 2003;327:1038–1040.
15. Stephen EH, Chandra A. Updated projections of infertility in the United States: 1995–2025. *Fertil Steril*. 1998;70:30–34.
16. Joffe M, Key J, Best N, et al. Human fertility decline? *Epidemiology*. 2006;17:238–239.
17. Crosignani PG, Rubin BL. Optimal use of infertility diagnostic tests and treatments. The ESHRE Capri Workshop Group. *Hum Reprod*. 2000;15:723–732.
18. Templeton A. Infertility and the establishment of pregnancy—overview. *Br Med Bull*. 2000;56:577–587.
19. Smith S, Pfeifer SM, Collins JA. Diagnosis and management of female infertility. *J Am Med Assoc*. 2003;290:1767–1770.
20. Bostofte E, Bagger P, Michael A, et al. Fertility prognosis for infertile couples. *Fertil Steril*. 1993;59:102–107.
21. Ombelet W, Bosmans E, Janssen M, et al. Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing. *Hum Reprod*. 1997;12:987–993.
22. van der Merwe FH, Kruger TF, Oehninger SC, et al. The use of semen parameters to identify the subfertile male in the general population. *Gynecol Obstet Invest*. 2005;59:86–91.
23. Urman B, Yakin K. Ovulatory disorders and infertility. *J Reprod Med*. 2006;51:267–282.
24. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Fertil Steril*. 2004;81:19–25.
25. Soliman S, Daya S, Collins J, et al. The role of luteal phase support in infertility treatment: a meta-analysis of randomized trials. *Fertil Steril*. 1994;61:1068–1076.
26. Yu D, Wong YM, Cheong Y, et al. Asherman syndrome—one century later. *Fertil Steril*. 2008;89:759–779.
27. Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. *Fertil Steril*. 2012;98:511–519.
28. Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med*. 2020;382:1244–1256. PubMed PMID: 32212520.
29. Brown J, Farquhar C. Endometriosis: an overview of Cochrane reviews. *Cochrane Database Syst Rev*. 2014;3:CD009590.
30. Donnez J, Jadoul P. What are the implications of myomas on fertility? A need for a debate? *Hum Reprod*. 2002;17:1424–1430.
31. Sunkara SK, Khairy M, El-Toukhy T, et al. The effect of intramural fibroids without uterine cavity involvement on the outcome of IVF treatment: a systematic review and meta-analysis. *Hum Reprod*. 2010;25:418–429.
32. Pritts EA, Parker WH, Olive DL. Fibroids and infertility: an updated systematic review of the evidence. *Fertil Steril*. 2009;91:1215–1223.
33. Menken J, Trussell J, Larsen U. Age and infertility. *Science*. 1986;233(4771):1389–1394.
34. Broekmans FJ, Knauff EA, te Velde ER, et al. Female reproductive ageing: current knowledge and future trends. *Trends Endocrinol Metab*. 2007;18(2):58–65.
35. Hunt PA, Hassold TJ. Human female meiosis: what makes a good egg go bad? *Trends Genet*. 2008;24(2):86–93.
36. Hunault CC, Laven JS, van Rooij IA, et al. Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Hum Reprod*. 2005;20(6):1636–1641.
37. Hunault CC, Habbema JD, Eijkemans MJ, et al. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod*. 2004;19(9):2019–2026.
38. Macklon NS, Fauser BC. Mild stimulation in in vitro fertilization. *ANN NY Acad Sci*. 2003;997:105–111.
39. Garrido-Gómez T, Ruiz-Alonso M, Blesa D, et al. Profiling the gene signature of endometrial receptivity: clinical results. *Fertil Steril*. 2013;99(4):1078–1085.
40. Guzick DS. Do infertility tests discriminate between fertile and infertile populations? *Hum Reprod*. 1995;10(8):2008–2009.
41. Mustafa RA, Wiercioch W, Cheung A, et al. Decision making about healthcare-related tests and diagnostic test strategies. paper 2: a review of methodological and practical challenges. *J Clin Epidemiol*. 2017;92:18–28.
42. Healy DL, Trounson AO, Andersen AN. Female infertility: causes and treatment. *Lancet*. 1994;343(8912):1539–1544.
43. Cahill DJ, Wardle PG. Management of infertility. *BMJ*. 2002;325(7354):28–32.
44. van Zonneveld P, Koppeschaar HP, Habbema JD, et al. Diagnosis of subtle ovulation disorders in subfertile women with regular menstrual cycles: cost-effective clinical practice? *Gynecol Endocrinol*. 1999;13(1):42–47.
45. Ali R, Gurtin B, Harper JC. Do fertility tracking applications offer useful information about the fertile window? *Reprod Biomed Online*. 2021;42:273–281.
46. Cervinski MA, Gronowski AM. Reproductive-endocrine point-of-care testing: current status and limitations. *Clin Chem Lab Med*. 2010;48(7):935–942.

47. Fehring RJ, Schneider M. Variability in the hormonally estimated fertile phase of the menstrual cycle. *Fertil Steril*. 2008;90(4):1232–1235.
48. Fiddelers AA, Land JA, Voss G, et al. Cost-effectiveness of *Chlamydia* antibody tests in subfertile women. *Hum Reprod*. 2005;20(2):425–432.
49. Coppus SF, Opmeer BC, Logan S, et al. The predictive value of medical history taking and *Chlamydia* IgG ELISA antibody testing (CAT) in the selection of subfertile women for diagnostic laparoscopy: a clinical prediction model approach. *Hum Reprod*. 2007;22(5):1353–1358.
50. Land JA, Evers JL. *Chlamydia* infection and subfertility. *Best Pract Res Clin Obstet Gynaecol*. 2002;16(6):901–912.
51. Land JA, Evers JL, Goossens VJ. How to use *Chlamydia* antibody testing in subfertility patients. *Hum Reprod*. 1998;13(4):1094–1098.
52. van Eekelen R, Scholten I, Tjon-Kon-Fat RI, et al. Natural conception: repeated predictions over time. *Hum Reprod*. 2017;32(2):346–353.
53. Mol BW, Collins JA, van der Veen F, et al. Cost-effectiveness of hysterosalpingography, laparoscopy, and *Chlamydia* antibody testing in subfertile couples. *Fertil Steril*. 2001;75(3):571–580.
54. den Hartog JE, Lardenoije CM, Severens JL, et al. Screening strategies for tubal factor subfertility. *Hum Reprod*. 2008;23(8):1840–1848.
55. van Rijswijk J, Pham CT, Dreyer K, et al. Oil-based or water-based contrast for hysterosalpingography in infertile women: a cost-effectiveness analysis of a randomized controlled trial. *Fertil Steril*. 2018;110(4):754–760.
56. Johnson N, Vandekerckhove P, Watson A, et al. Tubal flushing for subfertility. In: *The Cochrane Library*. Vol. 4. Chicester, UK: John Wiley & Sons; 2020:1–46. 1–1–0003.
57. Mol BW, Collins JA, Burrows EA, et al. Comparison of hysterosalpingography and laparoscopy in predicting fertility outcome. *Hum Reprod*. 1999;14(5):1237–1242.
58. Verhoeve HR, Coppus SF, van der Steeg JW, et al. Collaborative effort on the clinical evaluation in reproductive medicine. The capacity of hysterosalpingography and laparoscopy to predict natural conception. *Hum Reprod*. 2011;26(1):134–142.
59. Exalto N, Emanuel MH. Clinical aspects of HyFoSy as tubal patency test in subfertility workup. *BioMed Res Int*. 2019;2019:4827376.
60. Fatum M, Laufer N, Simon A. Investigation of the infertile couple: should diagnostic laparoscopy be performed after normal hysterosalpingography in treating infertility suspected to be of unknown origin? *Hum Reprod*. 2002;17(1):1–3.
61. Surrey ES. Endoscopy in the evaluation of the woman experiencing infertility. *Clin Obstet Gynecol*. 2000;43(4):889–896.
62. Oei SG, Helmerhorst FM, Bloemenkamp KW, et al. Effectiveness of the postcoital test: randomised controlled trial. *BMJ*. 1998;317(7157):502–505.
63. Oei SG, Helmerhorst FM, Keirse MJ. Routine postcoital testing is unnecessary. *Hum Reprod*. 2001;16(5):1051–1053.
64. Schwartz D, Mayaux MJ. Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands, Federation CECOS. *N Engl J Med*. 1982;306(7):404–406.
65. Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment [see comments]. *Lancet*. 1996;348(9039):1402–1406.
66. Fauser BC. Follicle pool depletion: factors involved and implications. *Fertil Steril*. 2000;74(4):629–630.
67. Lobo RA. Early ovarian ageing: a hypothesis. What is early ovarian ageing? *Hum Reprod*. 2003;18(9):1762–1764.
68. Scott MG, Ladenson JH, Green ED, et al. Hormonal evaluation of female infertility and reproductive disorders. *Clin Chem*. 1989;35(4):620–629.
69. Chang MY, Chiang CH, Chiu TH, et al. The antral follicle count predicts the outcome of pregnancy in a controlled ovarian hyperstimulation/intrauterine insemination program. *J Assist Reprod Genet*. 1998;15(1):12–17.
70. Hendriks DJ, Mol BW, Bancsi LF, et al. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. *Fertil Steril*. 2005;83(2):291–301.
71. Hansen KR, Hodnett GM, Knowlton N, Craig LB. Correlation of ovarian reserve tests with histologically determined primordial follicle number. *Fertil Steril*. 2011;95(1):170–175.
72. Jeppesen JV, Anderson RA, Kelsey TW, et al. Which follicles make the most anti-Mullerian hormone in humans? Evidence for an abrupt decline in AMH production at the time of follicle selection. *Mol Hum Reprod*. 2013;19(8):519–527.
73. Broer SL, Mol BW, Hendriks D, et al. The role of antimüllerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. *Fertil Steril*. 2009;91(3):705–714.
74. Hamdine O, Eijkemans MJC, Lentjes EGW, et al. Antimullerian hormone: prediction of cumulative live birth in gonadotropin-releasing hormone antagonist treatment for in vitro fertilization. *Fertil Steril*. 2015;104(4):891–898.e2.
75. Leijdekkers JA, Eijkemans MJC, van Tilborg TC, et al. Cumulative live birth rates in low-prognosis women. *Hum Reprod*. 2019;34(6):1030–1041.
76. La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. *Hum Reprod Update*. 2014;20(1):124–140.
77. Broer SL, Dólleman M, van Disseldorp J, IPD-EXPORT Study Group, et al. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. *Fertil Steril*. 2013;100(2):420–429.e7.
78. Broer SL, van Disseldorp J, Broeze KA, et al. IMPORT study group. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. *Hum Reprod Update*. 2013;19(1):26–36.
79. Lensen SF, Wilkinson J, Leijdekkers JA, et al. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI). *Cochrane Database Syst Rev*. 2018;2018(2). https://doi.org/10.1002/14651858.CD012693.pub2.
80. Nyboe Andersen A, Nelson SM, Fauser BC, García-Velasco JA, Klein BM, Arce JC, ESTHER-1 study group. Individualized versus conventional ovarian stimulation for in vitro fertilization: a multicenter, randomized, controlled, assessor-blinded, phase 3 noninferiority trial. *Fertil Steril*. 2017;107(2):387–396.
81. Fernández-Sánchez M, Visnova H, Yuzpe A, Klein BM, Mannaerts B, Arce JC, ESTHER-1, ESTHER-2 Study Group. Individualization of the starting dose of follitropin delta reduces the overall OHSS risk and/or the need for additional preventive interventions: cumulative data over three stimulation cycles. *Reprod Biomed Online*. 2019;38(4):528–537.
82. Eimers JM, te Velde ER, Gerritse R, et al. The prediction of the chance to conceive in subfertile couples. *Fertil Steril*. 1994;61:44–52.
83. Collins JA, Burrows EA, Wilan AR. The prognosis for live birth among untreated infertile couples. *Fertil Steril*. 1995;64:22–28.
84. Snick HK, Snick TS, Evers JL, et al. The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod*. 1997;12(7):1582–1588.
85. van der Steeg JW, Steures P, Eijkemans MJ, et al. CECERM Study Group (Collaborative Effort for Clinical Evaluation in Reproductive Medicine): pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Hum Reprod*. 2007;22(2):536–542.
86. Collins JA, Wrixon W, Janes LB, et al. Treatment-independent pregnancy among infertile couples. *N Engl J Med*. 1983;309(20):1201–1206.
87. Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod*. 2020;35(1):100–116.
88. Bagshawe A, Taylor A. ABC of subfertility: counselling. *BMJ*. 2003;327(7422):1038–1040.
89. Bensdorp AJ, Tjon-Kon-Fat RI, Bossuyt PM, et al. Prevention of multiple pregnancies in couples with unexplained or mild male subfertility: randomised controlled trial of in vitro fertilisation with single embryo transfer or in vitro fertilisation in modified natural cycle compared with intrauterine insemination with controlled ovarian hyperstimulation. *BMJ*. 2015;350:g7771.
90. Gleicher N. Cost-effective infertility care. *Hum Reprod Update*. 2000;6(2):190–199.

91. Philips Z, Barraza-Llorens M, Posnett J. Evaluation of the relative cost-effectiveness of treatments for infertility in the UK. *Hum Reprod*. 2000;15(1):95–106.
92. Steures P, van der Steeg JW, Hompes PG, et al. Intrauterine insemination with controlled ovarian hyperstimulation versus expectant management for couples with unexplained subfertility and an intermediate prognosis: a randomised clinical trial. *Lancet*. 2006;368(9531):216–221.
93. Van Voorhis BJ, Sparks AE, Allen BD, et al. Cost-effectiveness of infertility treatments: a cohort study. *Fertil Steril*. 1997;67(5):830–836.
94. Garceau L, Henderson J, Davis LJ, et al. Economic implications of assisted reproductive techniques: a systematic review. *Hum Reprod*. 2002;17(12):3090–3109.
95. Fauser BC, Bouchard P, Coelingh Bennink HJ, et al. Alternative approaches in IVF. *Hum Reprod Update*. 2002;8(1):1–9.
96. Heijnen EM, Macklon NS, Fauser BC. What is the most relevant standard of success in assisted reproduction? The next step to improving outcomes of IVF: consider the whole treatment. *Hum Reprod*. 2004;19(9):1936–1938.
97. Johnson NP, Proctor M, Farquhar CM. Gaps in the evidence for fertility treatment—an analysis of the Cochrane menstrual disorders and subfertility group database. *Hum Reprod*. 2003;18(5):947–954.
98. Aboulghar MA, Mansour RT, Serour GI, et al. Diagnosis and management of unexplained infertility: an update. *Arch Gynecol Obstet*. 2003;267(4):177–188.
99. Fauser BC, Van Heusden AM. Manipulation of human ovarian function: physiological concepts and clinical consequences. *Endocr Rev*. 1997;18(1):71–106.
100. Fauser BC, Devroey P, Macklon NS. Multiple birth resulting from ovarian stimulation for subfertility treatment. *Lancet*. 2005;365(9473):1807–1816.
101. Hogue CJ. Successful assisted reproductive technology: the beauty of one. *Obstet Gynecol*. 2002;100(5 Pt 1):1017–1019.
102. Zolton JR, Lindner PG, Terry N, DeCherney AH, Hill MJ. Gonadotropins versus oral ovarian stimulation agents for unexplained infertility: a systematic review and meta-analysis. *Fertil Steril*. 2020;113(2):417–425.e1.
103. van Rumste MM, Custers IM, van der Veen F, et al. The influence of the number of follicles on pregnancy rates in intrauterine insemination with ovarian stimulation: a meta-analysis. *Hum Reprod Update*. 2008;14(6):563–570.
104. Ayeleke RO, Asseler JD, Cohlen BJ, Veltman-Verhulst SM. Intra-uterine insemination for unexplained subfertility. *Cochrane Database Syst Rev*. 2020;3:CD001838. https://doi.org/10.1002/14651858.CD001838.pub6 [doi].
105. Danhof NA, van Wely M, Repping S, et al. Gonadotrophins or clomiphene citrate in couples with unexplained infertility undergoing intrauterine insemination: a cost-effectiveness analysis. *Reprod Biomed Online*. 2019. doi: S1472-6483(19)30740-0 [pii].
106. Bhattacharya S, Harrild K, Mollison J, et al. Clomifene citrate or unstimulated intrauterine insemination compared with expectant management for unexplained infertility: pragmatic randomised controlled trial. *BMJ*. 2008:337:a716.
107. Athaullah N, Proctor M, Johnson NP. Oral versus Injectable Ovulation Induction Agents for Unexplained Subfertility. *Cochrane Database Syst Rev*. 2002;3:CD003052.
108. Aitken RJ, Clarkson JS. Cellular basis of defective sperm function and its association with the genesis of reactive oxygen species by human spermatozoa. *J Reprod Fertil*. 1987:81:459–469.
109. Duran HE, Morshedi M, Kruger T, et al. Intrauterine insemination: a systematic review on determinants of success. *Hum Reprod Update*. 2002;8(4):373–384.
110. Boomsma CM, Heineman MJ, Cohlen BJ, et al. Semen preparation techniques for intrauterine insemination. *Cochrane Database Syst Rev*. 2007;4:CD004507.
111. Tur R, Barri PN, Coroleu B, et al. Use of a prediction model for high-order multiple implantation after ovarian stimulation with gonadotropins. *Fertil Steril*. 2005;83(1):116–121.
112. Gleicher N, Oleske DM, Tur-Kaspa I, et al. Reducing the risk of high-order multiple pregnancy after ovarian stimulation with gonadotropins. *N Engl J Med*. 2000;343(1):2–7.
113. Steures P, van der Steeg JW, Hompes PG, et al. Intrauterine insemination in The Netherlands. *Reprod Biomed Online*. 2007;14(1):110–116.
114. Custers IM, Steures P, Hompes P, et al. Intrauterine insemination: how many cycles should we perform? *Hum Reprod*. 2008;23(4):885–888.
115. Suri K, Bhandari V, Lerer T, et al. Morbidity and mortality of preterm twins and higher-order multiple births. *J Perinatol*. 2001;21(5):293–299.
116. Stromberg B, Dahlquist G, Ericson A, et al. Neurological sequelae in children born after in-vitro fertilisation: a population-based study. *Lancet*. 2002;359(9305):461–465.
117. Goverde AJ, McDonnell J, Vermeiden JP, et al. Intrauterine insemination or in-vitro fertilisation in idiopathic subfertility and male subfertility: a randomised trial and cost-effectiveness analysis. *Lancet*. 2000;355(9197):13–18.
118. Collins J. Stimulated intra-uterine insemination is not a natural choice for the treatment of unexplained subfertility: current best evidence for the advanced treatment of unexplained subfertility. *Hum Reprod*. 2003;18(5):907–912.
119. van Rumste MM, Custers IM, van Wely M, et al. IVF with planned single-embryo transfer versus IUI with ovarian stimulation in couples with unexplained subfertility: an economic analysis. *Reprod Biomed Online*. 2014;28(3):336–342.
120. Reindollar RH, Regan MM, Neumann PJ, et al. A randomized clinical trial to evaluate optimal treatment for unexplained infertility: the fast track and standard treatment (FASTT) trial. *Fertil Steril*. 2010;94(3):888–899.
121. Pandian Z, Bhattacharya S, Vale L, et al. In vitro fertilisation for unexplained subfertility. *Cochrane Database Syst Rev*. 2005;(2):CD003357.
122. Soliman S, Daya S, Collins J, et al. A randomized trial of in vitro fertilization versus conventional treatment for infertility. *Fertil Steril*. 1993;59:1239–1244.
123. Karande VC, Korn A, Morris R, et al. Prospective randomized trial comparing the outcome and cost of in vitro fertilization with that of a traditional treatment algorithm as first-line therapy for couples with infertility. *Fertil Steril*. 1999;71(3):468–475.
124. Dosouto C, Haahr T, Humaidan P. Gonadotropin-releasing hormone agonist (GnRHa) trigger - state of the art. *Reprod Biol*. 2017;17(1):1–8.
125. Kontoravdis A, Makrakis E, Pantos K, et al. Proximal tubal occlusion and salpingectomy result in similar improvement in in vitro fertilization outcome in patients with hydrosalpinx. *Fertil Steril*. 2006;86(6):1642–1649.
126. Lessey BA, Kim JJ. Endometrial receptivity in the eutopic endometrium of women with endometriosis: it is affected, and let me show you why. *Fertil Steril*. 2017;108(1):19–27. https://doi.org/10.1016/j.fertnstert.2017.05.031.
127. Min JK, Breheny SA, MacLachlan V, et al. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. *Hum Reprod*. 2004;19(1):3–7.

124

Androgen Excess Disorders in Women

Zi-Jiang Chen, Richard S. Legro, David A. Ehrmann, and Daimin Wei

OUTLINE

Androgen Physiology and Pathophysiology, 2060
  Androgen Biosynthesis, 2060
  Regulation of Androgen Secretion, 2062
  Blood Levels and Transport of Androgens, 2064
  Mechanisms of Androgen Action, 2066
Clinical Manifestation of Androgen Excess, 2066
  Cutaneous Manifestations, 2066
  Polycystic Ovary and Polycystic Ovary Morphology, 2067
  Metabolic Dysfunction, 2067
Common Androgen Excess Disorders, 2068
  Polycystic Ovary Syndrome, 2068
  Ovarian Neoplasms, 2070
  Other Ovarian Hyperandrogenic Disorders, 2071
  Congenital Adrenal Hyperplasia, 2071
  Other Adrenal Hyperandrogenic Disorders, 2072
  Idiopathic Hyperandrogenism, 2072
Diagnosis of Hyperandrogenism, 2072
  Clinical Hyperandrogenism, 2072
  Biochemical Hyperandrogenism, 2073
Treatment of Polycystic Ovary Syndrome and other Androgen Excess Disorders, 2074
  Androgen Suppression, 2074
  Ovulation Induction and Assisted Reproduction, 2075
  Prevention of Long-Term Complications, 2076
Summary and Future Directions, 2076

In this chapter, we first summarize the normal physiology of androgen production in women. This is followed by a review of the pathogenesis and clinical presentation of disorders leading to androgen excess. Finally, an approach to the diagnosis, differential diagnosis, and therapy of these disorders is provided.

ANDROGEN PHYSIOLOGY AND PATHOPHYSIOLOGY

Androgen Biosynthesis

Androgens (from the Greek word *andre*-, the stem of the word meaning "man") comprise a group of steroid hormones that regulate the developmental process and maintain male and female characteristics by binding to their cognate receptors. Classically, four main androgens have been identified including testosterone, dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEAS), and androstenedione, which are secreted from endocrine glands through similar pathways and are released into the circulation to exert a number of physiological functions (Table 124.1). More recently, other androgens, such as androsterone, have been found to play a key role in male sexual development and female hyperandrogenic disorders. Although androgens are essential during childhood, subsequently, their synthesis is important for pubertal development, sexual life, and improving fertility in males and females. Thus, androgen biosynthesis is regulated both in fetal and pubertal life in males and females.

Pathways of Androgen Biosynthesis. The early steps of steroid synthesis are shared between glucocorticoids, mineralocorticoids, and sex steroids; however, defects in the initial step will definitely affect the final product. Two possible pathways are described for androgen biosynthesis: the "classic" and the "backdoor" pathways[1] (Fig. 124.1). Animal studies, analyses of human tissues, and clinical studies have converged to form a classic model of androgen synthesis in a zone of the primate adrenal cortex called the zona reticularis.

The Classic Pathway of Androgen Biosynthesis. Cholesterol is the main precursor for the formation of all steroid end products, and the first step is transport to the inner mitochondrial membrane. Acute stimulation of androgen biosynthesis is mediated by steroidogenic acute regulatory protein (StAR), which is an active transporter of cholesterol through the inner mitochondrial membrane. Notably, any defect or mutation in StAR is accompanied by severe loss of adrenal steroidogenesis and lack of virilization.[2] StAR mutations, along with all other genetic defects of androgen biosynthesis, are inherited in an autosomal recessive fashion, and both types of genetic male and female individuals can be affected.[3]

The first enzymatic and rate-limiting step for the steroid cascade inside the mitochondrial inner membrane is supported by side-chain cleavage (SCC), a mitochondrial cytochrome enzyme (P450scc) that mediates the conversion of cholesterol into pregnenolone (Preg). Two cytochrome P450 enzymes (P450scc and cytochrome P450c17) are involved. P450scc stimulates the chemical reaction with the help of StAR protein in preparation for steroid hormone biosynthesis in the adrenal glands and the gonads.

In contrast, cytochrome P450c17 is important for cortisol and sex hormone biosynthesis. This enzyme is a qualitative regulator of steroid synthesis which has two distinct activities: 17α-hydroxylase and 17,20-lyase activity. These are essential for synthesis of cortisol, 17-ketosteroids DHEA, and androstenedione, precursors for steroid hormones.

For the production of testosterone through the classic pathway, first, pregnenolone is converted to 17-hydroxypregnenolone (17OH-preg) and then the 17,20-lyase activity of P450c17 converts 17OH-preg to DHEA, supported by cofactors P450 oxidoreductase and cytochrome

TABLE 124.1 Major Types of Androgen Source, Functions, and Metabolism in Women

Androgen Types	Site of Production	Production Rate/Day (mg)	Adrenal (%)	Ovarian (%)	Functions
Testosterone	Ovarian stroma, adrenal zona reticularis	0.1–0.4	25	25	Gonadotropin regulation, estradiol precursor, primordial follicle growth
DHEA-S	Peripheral tissue, adrenal zona reticularis	3.5–20	95–100	0–5	Preantral follicle development, androgen precursor, oocyte development
DHEA	Ovarian theca, adrenal zona reticularis	6–8	50	20	Testosterone precursor, anti-aging effect, improve sexual functions, endothelial proliferation
Androstenedione	Ovarian stroma, adrenal zona fasciculata	1.4–6.2	50	50	Testosterone precursor, follicle growth

DHEA, Dehydroepiandrosterone; *DHEA-S,* dehydroepiandrosterone sulphate.

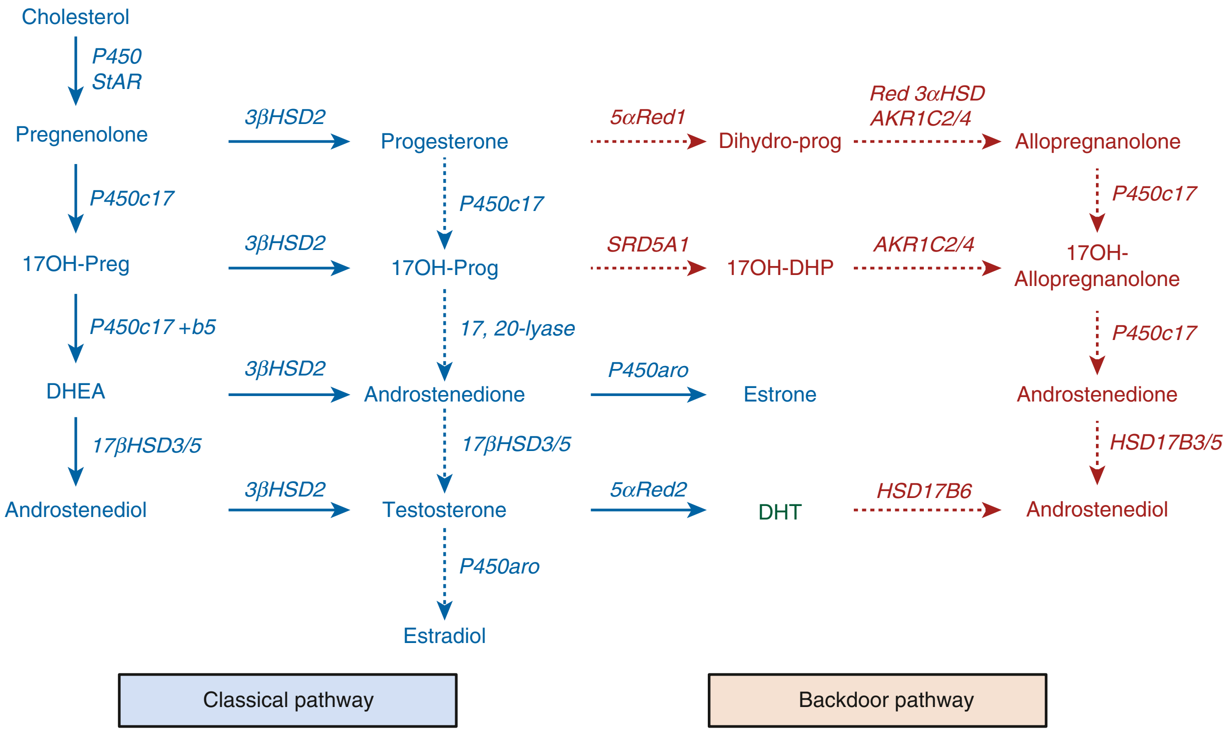

Cholesterol
P450
StAR
Pregnenolone
3βHSD2
Progesterone
5αRed1
Dihydro-prog
Red 3αHSD
AKR1C2/4
Allopregnanolone
P450c17
P450c17
P450c17
17OH-Preg
3βHSD2
17OH-Prog
SRD5A1
17OH-DHP
AKR1C2/4
17OH-
Allopregnanolone
P450c17 +b5
17, 20-lyase
P450c17
DHEA
3βHSD2
Androstenedione
P450aro
Estrone
Androstenedione
17βHSD3/5
17βHSD3/5
HSD17B3/5
Androstenediol
3βHSD2
Testosterone
5αRed2
DHT
HSD17B6
Androstenediol
P450aro
Estradiol
Classical pathway
Backdoor pathway

Figure 124.1 Outlines of classic and backdoor pathway of androgens biosynthesis. Adrenal and gonadal pathways are combined in the figure. The left hand column indicates the Δ5 classic pathway, in which steroids retain the double bond between carbons 5 and 6 in cholesterol's B-ring. The StAR mediates the transport of cholesterol to the inner membrane of mitochondria, where P450 facilitates to cleave off the side chain and resultant pregnenolone (Preg) formation. Δ4 steroids are converted from Δ5 steroid by using 3βHSD2 (HSD3B2) within the adrenal and gonadal tissue. Progesterone (Prog) is converted into 17OH-prog in the zona fasciculate and gonads by P450c17. 17OH-prog and 17OH-preg is converted to androstenedione and DHEA by human P450c17 activity, followed by testosterone production in stepwise manner. DHEA is converted into androstenediol and then androstenedione to testosterone by the activity of 17βHSD3. Right hand side in red arrows indicates the backdoor pathways of adrenal biosynthesis. In alternative backdoor pathway, progesterone is converted into neuroactive steroid, allopregnanolone. Note that in the alternative pathway, 17OH-prog is 5α, 3α reduced and then converted into androstenedione. Androsterone is the principle androgen in fetal circulation which is converted into androstenediol by 17HSD17B3/5 and then further oxidized by HSD17B6 to yield into potent androgen DHT.*5-αRed2,* 5α-reductase type 2; *AKR1C2/4,* aldo-keto reductase family 1, member C2/4; *Aro,* aromates; *DHEA,* dehydroepiandrosterone; *DHT,* dihydrotestosterone; *HSD3B2,* 3β-hydroxysteroid dehydrogenase type 2; *Preg,* pregnenolone; *Prog,* progesterone; *StAR,* steroidogenic acute regulatory protein.

b5. P450c17 is encoded by a single copy gene on chromosome 10q24.3, and mutations within this gene can inhibit either all functions or selectively the 17,20-lyase activity of the resulting protein.[4]

DHEA is a C19 steroid and androgen precursor, which is then either converted to androstenedione (Δ4A) or androstenediol and finally to testosterone by the enzymes 3β-hydroxysteroid dehydrogenase type 2 (HSD3B2) and 17β-hydroxysteroid dehydrogenase (HSD17B3) that exhibit tissue-specific expression. Approximately 50% of testosterone comes from direct secretions from steroid-producing endocrine glands, whereas the remaining 50% is produced by different organs including the adipose tissue, liver, and skin which contain 3β-HSD type 1, 17β-HSD type 5, and P450arom enzymes.

Testosterone is further converted to the most potent androgen, dihydrotestosterone (DHT), by 5α-reductase type 2 (SRD5A2). Two isoforms of 5α-reductase are expressed in different tissues; however, 5α-reductase type 2 is the more abundant form in the gonads. With regard to the degree of androgen action at the androgen receptor (AR), DHT is the most potent androgen followed by testosterone, whereas weak stimulation of AR has been observed by androstenedione and androstenediol.[2]

The Backdoor Pathway of Androgen Biosynthesis. For the synthesis of DHT, an alternative "backdoor" pathway was first discovered in the tammar wallaby pouch young and then was identified in mouse pups.[5] It provided evidence that testosterone is the androgen in the circulation, while DHT exerts its actions locally on genital skin. This pathway bypasses the usual intermediate steroids DHEA, androstenedione, and testosterone. This alternative backdoor pathway diverges from the classic pathway with further conversion of 17-hydroxyprogesterone (17OH-prog) to 17OH-dihydroprogesterone (17OH-DHP; 5α-pregnane-17α-ol-3,20-dione), 17OH-allopregnanolone (17OH-Allo; 5α-pregnane-3α,17α-ol-3,20-dione), androsterone, and androstanediol (3αDiol) to DHT. The stepwise conversion of 17OH-prog to DHT is supported by the enzymes 5α-reductase type 1 (SRD5A1), AKR1C2/4, P450c17, HSD17B3/5, and AKR1C2/4. This novel alternate pathway to DHT thus circumvents the usual steroid intermediates and is thus considered as "the backdoor pathway."

The backdoor pathway plays an important role in a number of hyperandrogenic disorders including PCOS,[6] congenital adrenal hyperplasia (CAH),[7] some virilized female newborns with P450 oxidoreductase deficiency,[8] and mini puberty of infancy.[9]

Regulation of Androgen Secretion

Organ-Specific Androgen Secretion Regulation.

In women, the ovaries and adrenal glands are the prime source of androgen secretion stimulated by luteinizing hormone (LH) and adrenocorticotropic hormone (ACTH), respectively. ACTH is a 39-amino acid peptide synthesized and secreted by the anterior pituitary (Fig. 124.2). Both corticotropin-releasing hormone and arginine-vasopressin work synergistically to regulate the synthesis of ACTH. The secretion and levels of adrenal androgens are controlled by intraglandular paracrine and autocrine mechanisms, rather than direct negative feedback by the pituitary gland.

Circulating androgens are predominantly produced in the ovaries, adrenal glands, and from peripheral tissue conversion. Androgen precursors, including DHEA, DHEAS, and androstenedione, and the active androgen testosterone are the major androgens secreted from the adrenal glands. Two potent bioactive androgens, testosterone and DHT, directly bind to the receptors, whereas DHEA, DHEAS, and androstenedione are pro-androgens that exert androgenic effects after conversion to testosterone and DHT. DHT, the most potent androgen, circulates at very low levels due to its production from local metabolism within tissues.

The synthesis and production of androgens vary by age; for example, DHEAS levels are high in newborns and then precipitously decline until adrenarche, which commences approximately at 6–8 years of age. High levels of DHEA and DHEAS are the hallmark of adrenarche. Increased levels of DHEAS are noted prior to the increase in the circulating concentrations of estrogens and androgens, whose synthesis and increasing blood levels are directly associated with puberty. Circulating levels of DHEAS are high during the second decade of life, and higher levels are noted in males than in females. However, no significant increase in the concentration of cortisol and ACTH has been observed during this period, indicating that adrenarche is not dependent on activation by the pituitary axis. The following sections describe the organs specific for androgen production and metabolism.

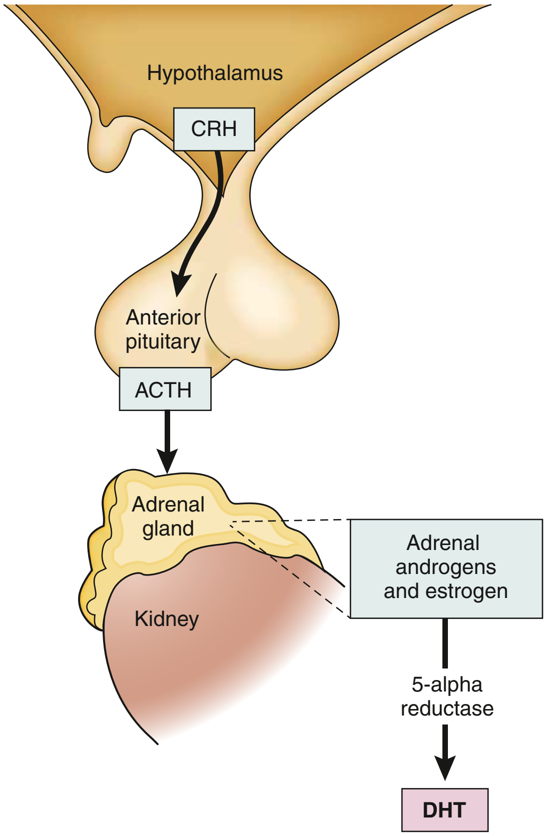


Figure 124.2 Schematic outlines of regulation for androgen secretion by adrenal structure. *ACTH,* Adrenocorticotropin; *CRH,* corticotrpin-releasing hormone; *DHT,* dihydrotestosterone.

Ovary. The principal source of ovarian androgens is theca interna cells and the ovarian stroma of the developing follicles. The mechanism of regulation of androgen secretions is similar to Leydig cell regulation in the testes. The combined activity of LH and follicle-stimulating hormone (FSH) is required to stimulate ovarian steroidogenesis, with LH stimulating the theca-interstitial (thecal) cells to produce androgen precursor for subsequent conversion by ovarian granulosa cells under the action of FSH (Fig. 124.3). In the adult human ovary, progesterone and estradiol are produced from the androgen precursor cholesterol under the control of LH and FSH. The production of progesterone occurs during the luteal phase, while estrogens are produced in the follicular phase. Human granulosa cells convert cholesterol to pregnenolone (due to lack of P450c17) which is further converted into progesterone and then ultimately to androstenedione. Androstenedione secretion from the theca cells represents approximately 50% of circulating androgens in women. Androgenic precursors (especially androstenedione) are produced by thecal cells under LH stimulation and aromatized to estrogen by FSH-stimulated granulosa cell aromatase activity. During the preovulatory phase, the high level of LH inhibits the activity of 17,20-lyase and 17α-hydroxylase, thereby ultimately reducing testosterone and androstenedione concentrations. The growth of antral stage follicles requires stimulation of granulosa cells via FSH action. Increased androstenedione and estradiol secretion accompanies the development of the dominant follicle. Androgens are obligate intermediates for the synthesis of estradiol and play an important role in follicle selection, follicle recruitment, and ultimately follicular growth (see Chapters 118 and 119). Therefore, the overall functional activity of the ovarian follicle structure is highly associated with the level of estradiol. High

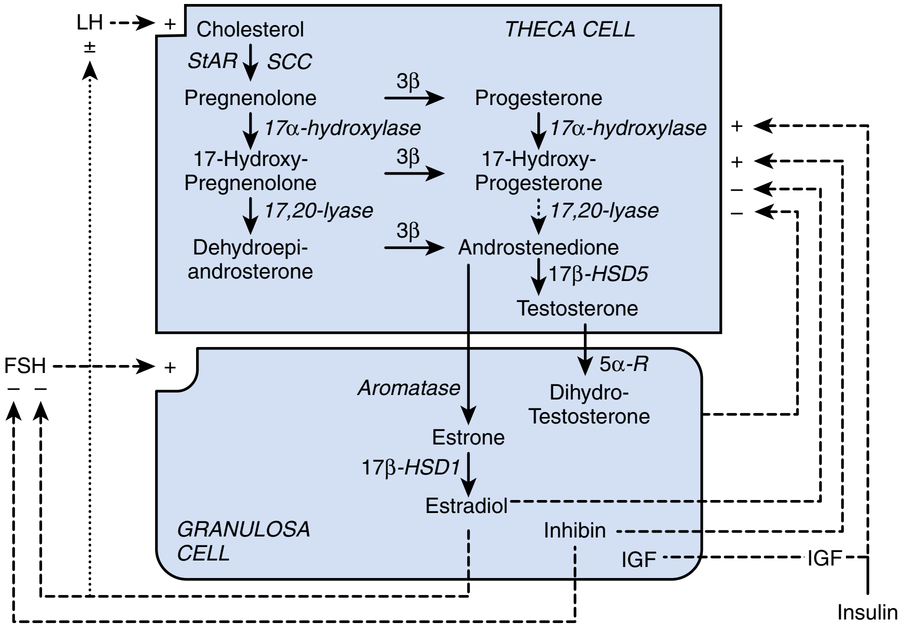


Figure 124.3 Outline of the organization of the major steroid biosynthetic pathways in the small antral follicle of the ovary, depicted according to the two-gonadotropin, two-cell model of ovarian steroidogenesis. Luteinizing hormone (LH) stimulates androgen formation within theca cells via the steroidogenic pathway common to the gonads and adrenal glands. Follicle-stimulating hormone (FSH) regulates estradiol biosynthesis from androgen by granulosa cells. Long-loop negative feedback of estradiol on gonadotropin secretion does not readily suppress LH at physiologic levels of estradiol and stimulates LH under certain circumstances. Androgen formation in response to LH appears to be modulated by intraovarian feedback at the levels of 17-hydroxylase and 17,20-lyase, both of which are activities of cytochrome P450c17. The quantitative importance of androstenedione formation from 17 OH-Prog *(dotted arrow)* in the intact follicle is unknown. Androgens and estradiol inhibit *(minus signs)* and inhibin, insulin, and insulin-like growth factor-1 (IGF-1) stimulate *(plus signs)* 17-hydroxylase and 17,20-lyase activities. Pertinent enzymes are italicized.*3β,* Δ5isomerase-3β-hydroxysteroid dehydrogenase type 2; *17βHSD,* 17β-hydroxysteroid dehydrogenase types 1 and 5; *5α-R,* 5α-reductase; scc, side-chain cleavage; *StAR,* steroidogenic acute regulatory protein. (Modified with permission from Ehrmann DA, Barnes RB, Rosenfield RL. Polycystic ovary syndrome as a form of functional ovarian hyperandrogenism due to dysregulation of androgen secretion. *Endocr Rev.* 1995;16:322–353. Copyright © 2007 The Endocrine Society.)

levels of androgens suppress ovulation and cause the development of the histologic and gross anatomic features of the polycystic ovary (PCO, see below).

Ovarian thecal androgen production is regulated by the functional activity of LH. Desensitization in part is mediated by downregulation of the LH receptor binding sites by LH itself (homologous desensitization). The main product obtained from the theca interna and stroma cells of the ovary is androstenedione, which ultimately is transformed into testosterone by the activity of theca and granulosa cells; however, the ability of the granulosa cells to synthesize testosterone is relatively high compared with theca cells. These androgens transport to the granulosa cells by the process of diffusion through the basal membrane of follicles, and inside the granulosa cells, the androgen precursor is aromatized to 17β estradiol by the stimulatory action of FSH on the aromatase enzyme activity. FSH in combination with estradiol increase the division of granulosa cells and ultimately the volume of the antral cavity. Regulation of androgen secretion is a complex process in which different hormones and growth factors are involved that mediate ovarian functional activity that ultimately increases steroidogenesis by activation of P450c17 (see Fig. 124.3). The activity of P450c17 is inhibited by estrogen through a negative feedback mechanism using both paracrine and autocrine routes. These inhibiting activities are balanced by the action of different hormones and growth factors including insulin, insulin-like growth factor-1 (IGF-1), and inhibin that ultimately amplify P450c17 action (see Fig. 124.3 and see Chapter 119).

Adrenal Cortex. The adult human cortex is composed of three layers: the zona glomerulosa (ZG) for mineralocorticoid production, the zona fasciculata (ZF) for glucocorticoid production, and the zona reticularis (ZR) for androgen production. ZG and ZF play an important role during fetal life, while the ZR production of steroid hormones begins slowly after birth and becomes active at adrenarche, approximately at 6–8 years of age. Differentiation of the centripetal progenitor cells into ZR, then ZG, and finally again conversion into ZR cells underlies the development of the adrenal zones. In terms of androgen production, the activity of ZR reaches a maximum at the age of 30 years and then declines thereafter. The activity of ZR is low for HSD3B2 and high for P450C17-20 lyase, and sulfonyl-transferase (SULT2A1), ultimately yielding massive production of DHEA and DHEAS.[10]

Adrenal 17-ketosteroid is an important androgen whose secretion commences in mid-childhood and reaches its maximum at adrenarche. Adrenarche is characterized by a change in the pattern of the adrenal secretory response to ACTH. ACTH elicits cortisol production prior to adrenarche, although both cortisol and 17-ketosteroid secretion begins after adrenarche. Factors contributing to this change at adrenarche are nutritional status, somatic growth, secretion of leptin, insulin, and IGFs; however, further investigation is needed to highlight the role(s) of each of these factors. There are marked increases in the production of 17OH-preg and DHEA that cause increased transformation of DHEAS, the predominant androgen secreted by the adrenal gland. DHEAS is a unique secretory product released from the ZR at 3.5–20

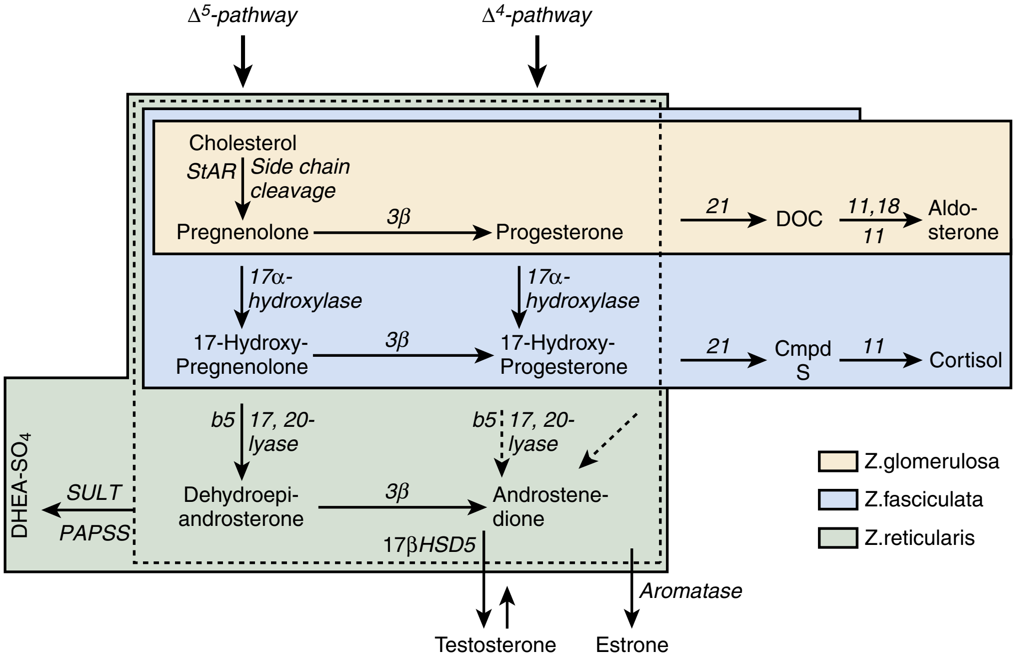


Figure 124.4 Outline of the organization of the major steroid biosynthetic pathways in the adrenal cortex. The area within the dotted square contains the core steroidogenic pathways, of which the *left column* shows the Δ5-pathway and the *right column* shows the Δ4-pathway, also utilized by ovarian theca cells. The *top yellow row* shows the pathway to aldosterone; the *middle blue row* shows the zona fasciculata pathway to cortisol. The *lower green row* shows the zona reticularis steps to DHEA-sulfate (SO_4) and other 17-ketosteroids. The *dotted box* encompasses steps common to all zones and the gonads. *Dotted pathways* are considered to be relatively minor. Cmpd S = 11-deoxycortisol; the 11-deoxy intermediate to aldosterone (deoxycorticosterone) is not shown. The steroidogenic enzymes are *italicized*. Cytochrome P450 enzyme steps are: side-chain cleavage (scc); 17α-hydroxylase/17,20-lyase; 21-hydroxylase (21); 11β-hydroxylase/18-hydroxylase-dehydrogenase (11,18); and aromatase. Non-P450 enzyme steps are steroidogenic acute regulatory protein (StAR); Δ5-isomerase-3β-hydroxysteroid dehydrogenase type 2 (3β); 17β-hydroxysteroid dehydrogenase type 5 (17βHSD5); and sulfotransferase 2A1 (SULT). Clinically relevant electron and sulfate transfer enzymes are, respectively, P450-oxidoreductase (POR), cytochrome b5 (b5), and 3′-phosphoadenosine-5′-phosphosulfate synthase 2 (PAPSS). (Modified with permission from Rosenfield RL. Identifying children at risk for polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2007;92:787. Copyright © 2007 The Endocrine Society.)

mg per day in reproductive age women. Changes during adrenarche are highly associated with the development of the ZR and the adrenal zone that produces large amounts of DHEA and DHEAS (Fig. 124.4).[2] The activity of 3β-HSD and cytochrome b5 increases the production of DHEA that is further converted into DHEAS in the adrenal gland by the action of high zonal expression of *SULT2A1*. *SULT2A1* (steroid sulfotransferase or cholesterol sulfotransferase) is highly expressed in the cytoplasm of adrenocortical cells in the reticularis and its substrates include pregnenolone, 17α hydroxypregnenolone, and DHEA. During fetal development, these steroids are used as substrates resulting in high levels of sulfated products that are then transported to the placenta for further biosynthesis. However, the predominant substrate in adults is DHEA which ultimately is transformed into DHEAS. The quantity of DHEAS varies across the life span, however, although it is the dominant androgen released from the adrenal gland. Similarly, in the ovary different factors, including insulin and IGFs, contribute to the production of adrenal steroids in response to ACTH. Infusion of insulin to hyperandrogenic women modestly potentiates the 17-ketosteroid response to ACTH in a manner that is compatible with increases in 17-hydroxylase and 17,20-lyase activities. *In vitro* studies have directly shown that insulin and IGFs upregulate adrenal 17-hydroxylase, 17,20-lyase, and 3β-HSD activities.

Blood Levels and Transport of Androgens

Serum concentrations of androgens and intermediate products in their biosynthesis and metabolism are shown in Table 124.2. Androgens in women are mainly derived from the adrenal gland and ovary. While serum DHEAS levels usually reflect adrenal androgen production, testosterone, androstenedione, and DHEA are secreted by both the ovary and the adrenal gland. Among circulating androgens, testosterone is the most important biologically and clinically because of its relatively high serum concentration and potent effect on target organs.

Androgens exist in the blood in conjugated or unconjugated forms. Over 96% of plasma testosterone binds to carrier proteins including sex hormone-binding globulin (SHBG) and albumin. SHBG is a glycoprotein synthesized in the liver and secreted into the circulation.[11] It has a high binding affinity for testosterone, leading to a small fraction of total testosterone as free and bioavailable. The binding of testosterone to albumin is of much lower affinity, and therefore SHBG concentration is the major determinant of the testosterone remaining free and biologically active in the blood. SHBG serum levels are influenced by metabolic and hormonal factors in various physiological and pathological conditions. Therefore, they not only change (usually decreased) in metabolic diseases such as obesity, type 2 diabetes mellitus, and non-alcoholic fatty liver disease, but also in thyroid disorders, pituitary diseases, and breast and

TABLE 124.2 Typical Normal Ranges for Ovarian and Adrenal Function Tests in Children and Reproductively Mature Females*

	17PREG (ng/dL)	17OHP (ng/dL)	11-Deoxycortisol (ng/dL)	Cortisol (µg/dL)	DHEAS (µg/dL)	DHEA (ng/dL)	Androstenedione (ng/dL)	Testosterone (ng/dL)	Estradiol (pg/mL)
Baseline									
Children 1–5 years of age	10–105	5–115	<25–160	3–20	5–35	20–130	10–50	<20	<10
Children 6–10 years of age	10–200	5–115	<25–160	3–20	10–115	20–345	10–75	<20	<10
Mature 11–40 years of age	55–360	≤130	<25–160	3–20	35–355	100–1020	50–265	10–60	20–85
Post-Dexamethasone (0.5 mg PO four times a day × 4 days)									
Mature 11–40 years of age	<25	<25–45†	<30	≤1.0	10–50	50–115	30–150	11–27†	25–100
Post-ACTH (0.5–1.0 hour post-IV ≥ 10 µg/m² $ACTH_{1\text{-}24}$)									
Children 1–5 years of age	45–350	50–270	95–300	17–45	5–35	25–100	15–70	<20	—
Children 6–10 years of age	60–650	85–300	95–300	17–45	10–115	70–320	25–100	<20	—
Mature 11–40 years of age	150–1070	35–140	95–300	17–45	75–255	250–1470	60–250	20–60	—
Post-GnRH agonist (20–24 hour post-leuprolide acetate 10 µg/kg SC)‡									
Girls 6–9 years of age	—	<25	—	≤5.0	—	25–70	25–40	<20	<10–50
Mature 11–40 years of age	—	30–145	—	<5.0	—	70–195	50–180	<10–60	65–260
Conversion									
Multipliers to SI units	0.0316 (nmol/L)	0.0303 (nmol/L)	0.0289 (nmol/L)	0.0276 (µmol/L)	0.0271 (µmol/L)	0.0347 (nmol/L)	0.0349 (nmol/L)	0.0347 (nmol/L)	3.67 (pmol/L)

DHEA, dehydroepiandrosterone; *17PREG,* 17-hydroxypregnenolone; *17OHP,* 17OH-prog.

*Range for high-specificity radioimmunoassays after preparatory chromatography, except for cortisol, DHEAS, and estradiol. Values differ slightly among laboratories. Mature females ≥ 1 year postmenarcheal, early follicular phase.

†Short dexamethasone test (4 hours after noontime dexamethasone 0.5 mg/m^2) yields similar testosterone suppression.

‡At 1600-hour after dexamethasone administration (0.5 mg PO at 1200-hour) to blunt coincidental adrenocortical secretion.

Data from Rosenfield RL, Barnes RB, Ehrmann DA. 1994 Studies of the nature of 17 OH-Prog hyperresponsiveness to gonadotropin releasing hormone agonist challenge in functional ovarian hyperandrogenism. *J Clin Endocrinol Metab*. 1994;79:1686–1692; Rosenfield RL. Identifying children at risk of polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2007;92:787–796; Rosenfield RL, Mortensen M, Wroblewski K, et al. Determination of the source of androgen excess in functionally atypical polycystic ovary syndrome by a short dexamethasone androgen-suppression test and a low-dose ACTH test. *Hum Reprod*. 2011;26:3138–3146; Rosenfield RL, Bordini B, Yu C. Comparison of detection of normal puberty in girls by a hormonal sleep test and a gonadotropin-releasing hormone agonist test. *J Clin Endocrinol Metab*. 2013;98:1591–1601; Trabert B, Xu X, Falk RT, et al. Assay reproducibility of serum androgen measurements using liquid chromatography-tandem mass spectrometry. *J Steroid Biochem Mol Biol*. 2016;155:56–62; O'Reilly MW, Kempegowda P, Jenkinson C, et al. 11-Oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2017;102:840–848.

prostate cancer.[12] Reportedly, tumor necrosis factor (TNF)-α downregulates SHBG in chronic low-grade inflammatory diseases such as obesity and type 2 diabetes mellitus.[13] PCOS is also considered to be a low-grade chronic inflammatory state, and SHBG levels are often decreased in the PCOS patients (see below), which may be due to hyperinsulinemia as well as elevated TNF-α.[14] Therefore, in some PCOS patients whose total testosterone levels are still in the normal range, their serum *bioavailable* testosterone levels are often elevated. This elevation of free or bioavailable testosterone is the best single indicator of hyperandrogenism. However, direct free testosterone assays, such as radiometric or enzyme-linked assays, are not recommended to assess biochemical hyperandrogenism in PCOS.[15] Instead, there has been increasing reliance and standardization of measuring testosterone levels through liquid chromatography with tandem mass spectrometry (LC-MS/MS) due to high sensitivity, accuracy, and precision.

Mechanisms of Androgen Action

Androgens exert their effects by binding to the AR in target cells. The human *AR* gene is located on the X chromosome and encodes a nuclear hormone receptor. After binding to testosterone, AR binds to DNA and results in activation or repression of target genes. However, the biological activity of testosterone is dependent largely on its conversion to DHT. DHT has greater biological activity than testosterone due to its higher affinity for AR.[16] It is primarily formed from testosterone and other precursors by the 5α-reductase activity in peripheral tissues including skin and liver, which largely determines the androgen action. Skin is not only a main site for androgen conversion but also a target organ of androgen, where AR is expressed in the hair follicle and sebaceous gland. In the ovary, androgen stimulates follicle development through binding to its receptor which is expressed in granulosa cells, theca cells, and oocytes.[17] Polymorphisms and skewed X-inactivation of the AR have been associated with sensitivity to androgens. In addition, androgens can also exert effects via rapid non-genomic mechanisms.[18]

KEY POINTS

- The major types of androgen in women are testosterone, DHEAS, DHEA, and androstenedione.
  Androgens in women are biosynthesized via the classic pathway and the backdoor pathway.
- Circulating androgens are predominantly produced in the ovaries, adrenal glands, and from peripheral tissue conversion.
- Androgen exit in the blood in conjugated or unconjugated forms.
- Over 96% of plasma testosterone binds to SHBG and albumin.

CLINICAL MANIFESTATION OF ANDROGEN EXCESS

Cutaneous Manifestations

Hair follicles are associated with sebaceous glands to form pilosebaceous units (PSUs) (Fig. 124.5). Hair is categorized as either *vellus* or *terminal*. Vellus hair is fine, soft, and not pigmented, whereas terminal hair is long, coarse, and pigmented. Growth of the hair follicle of PSUs is controlled by an as yet uncharacterized induction factor(s) elaborated by the dermal papilla. Androgen promotes sebum production and induces transformation of vellus to terminal hair. In target areas, androgens cause the prepubertal PSUs to differentiate into either a sexual hair follicle (in which the vellus hair transforms into a terminal hair) or into a sebaceous gland (in which the sebaceous component proliferates and the hair remains vellus). Anti-androgens reverse this process, causing PSUs to revert toward the prepubertal state. The effect of androgens on PSUs varies by body location and largely depends on local concentrations of active androgen (testosterone and DHT) and abundance of ARs.

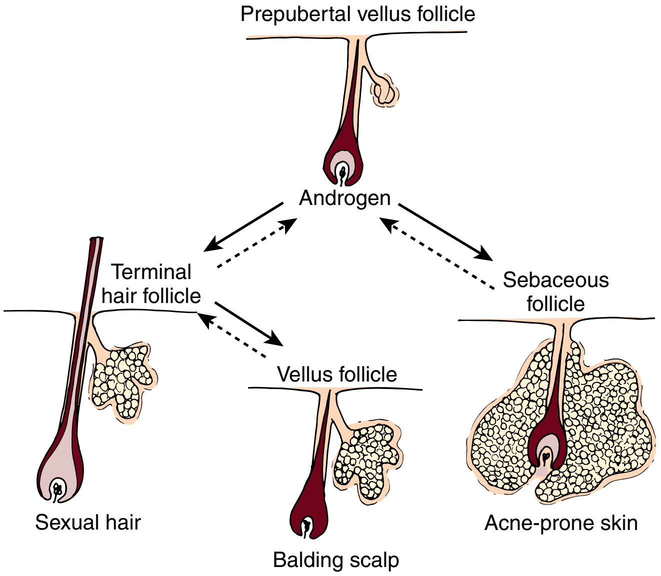


Figure 124.5 **Role of androgen in the development of the pilosebaceous unit.** *Solid lines* indicate effects of androgens, and *dotted lines* indicate the effects of anti-androgens. Hairs are depicted only in the anagen (growing) phase of the hair cycle. In conjunction with other regulatory factors, androgens cause the prepubertal pilosebaceous unit in androgen-dependent areas to differentiate into either a sexual hair follicle, in which the vellus hair transforms into a terminal hair, or into a sebaceous follicle, in which the sebaceous component proliferates and the hair remains vellus. In balding scalp *(bottom center),* terminal hairs not previously dependent upon androgen regress to vellus hairs under its influence. (Reproduced with permission from Deplewski D, Rosenfield RL. Role of hormones in pilosebaceous unit development. *Endocr Rev.* 2000; 21:363-392. Copyright ©2000 The Endocrine Society.)

Hirsutism is defined as excessive male pattern hair growth in women and is the most common clinical diagnostic criterion of hyperandrogenism. Hirsutism must be distinguished from *hypertrichosis*, i.e., excessive growth of androgen-independent hair which is vellus, prominent in nonsexual areas, and most commonly familial or caused by metabolic disorders (e.g., thyroid abnormalities, anorexia nervosa) or medications (e.g., phenytoin, minoxidil, or cyclosporine). The clinical presence of hirsutism is usually determined by a scoring scale of hair growth in nine body areas, i.e., the modified Ferriman-Gallwey (mFG) scale (Fig. 124.6). There are significant ethnic and racial differences in the definition of hirsutism. In adult White American and Black women, hirsutism is defined by a total score of 8 or more on the mFG scale. Thus, it is normal for these women to have a few sexual hairs in many "male" areas (upper lip and chin). Notably, scores ≥2 to 3 in Han Chinese and Thai women, and ≥9 to 10 in Mediterranean, Hispanic, and Middle Eastern women are excessive.[19] However, there are limitations in this scale, mainly due to the subjective nature of assessment, interobserver variance, and problems in evaluating women who had cosmetic therapies.

Acne and male pattern balding (androgenic alopecia) are other cutaneous manifestations of androgen excess.[20] Acne is characterized by increased sebum production and obstructed sebaceous follicle by keratinization of epithelial cells. Androgenic alopecia in women can display both a male pattern and a female pattern of hair loss. The male pattern is characterized by hair loss mainly in the sagittal scalp accompanied by recession of the frontal hairline and can be associated with virilizing levels of circulating androgens. The female pattern of androgenic alopecia is characterized by progressive thinning of the crown area of the scalp and is often characterized clinically by the Ludwig

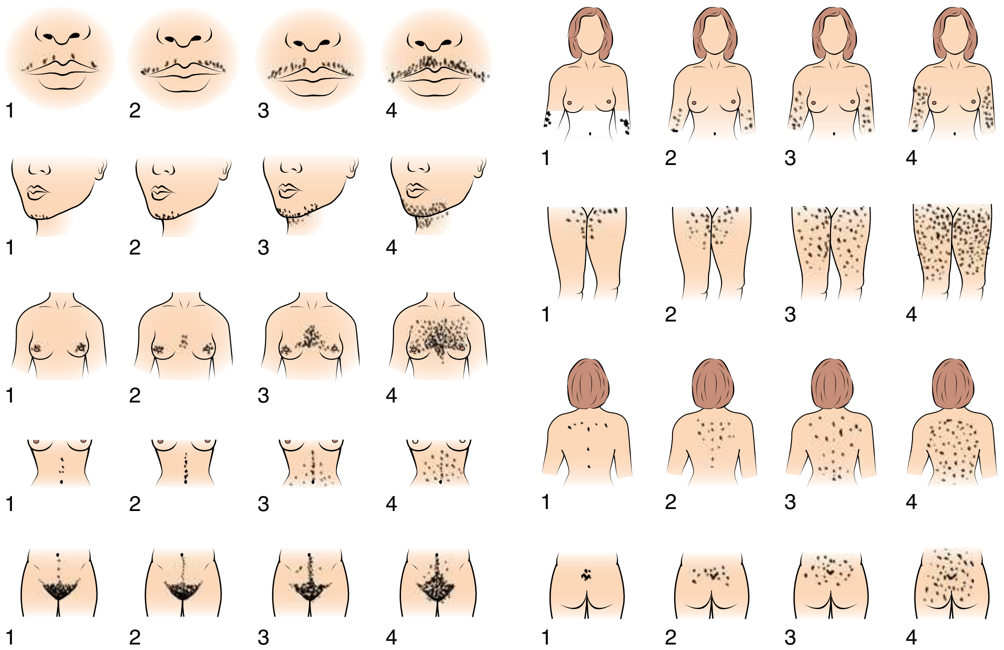


Figure 124.6 **Hirsutism scoring scale of Ferriman and Gallwey.** The nine body areas possessing androgen-sensitive pilosebaceous units are graded from 0 (no terminal hair) to 4 (frankly virile) and the sum totaled. Normal hirsutism score is less than 8 in adult White American and Black women. (Modified with permission from Hatch R, Rosenfield RL, Kim MH, Tredway D. Hirsutism: implications, etiology, and management. *Am J Obstet Gynecol.* 1981;140:815–830.)

scale.[21] Androgenic alopecia is largely the result of conversion of terminal hair to sebaceous follicles under the impact of androgen excess. The mechanisms for the different PSUs' response to androgen are still unclear. The cutaneous signs vary with the degree of androgen excess and are also dependent on the sensitivity of the PSUs to androgens. Some hyperandrogenic patients have hirsutism or acne alone, others both, yet others neither ("cryptic hyperandrogenemia").

Elevated serum total or bioavailable testosterone concentration underlies half of the cases of mild hirsutism, and most women with a 2-fold or greater elevation of androgen levels have some degree of hirsutism.[22] Moderately severe hirsutism or cystic acne is even more likely to be hyperandrogenic. However, the amount of sexual hair and sebaceous gland development seems to depend as much upon the factors determining sensitivity of the PSUs to androgens as it is upon the serum androgen level itself. At one end of the normal spectrum are women whose PSUs seem "hypersensitive" to normal blood free androgen levels; this seems to account for "idiopathic" hirsutism and acne. At the other end of the spectrum are women whose PSUs are relatively insensitive to androgen, accounting for "cryptic" hyperandrogenemia (hyperandrogenism without skin manifestations).

Polycystic Ovary and Polycystic Ovary Morphology

Polycystic ovary morphology (PCOM) describes a state of excessive numbers of small antral follicles (2–9 mm in diameter) present in the ovaries. With advances in sonography, antral follicles can be easily and precisely counted, although the threshold in the number of follicles for the diagnosis of PCOM is still controversial and is subject to the ever-greater resolution of transvaginal ultrasonography.[23] An ovary containing ≥ 12 antral follicles or an ovarian volume of ≥ 10 mL is diagnostic for PCOM by the Rotterdam criteria.[24] PCOM is tightly associated with androgen excess in patients with PCOS, CAH, and in prenatally androgenized nonhuman primate models, suggesting a causal relationship. Exposing rhesus female fetuses to high intrauterine concentrations of testosterone induces features of PCOM. The AR is highly expressed in preantral follicles and early antral follicles, declines with follicle development, and is nearly absent in preovulatory follicles. Although the precise mechanisms are not known, androgen excess promotes initiation of primordial follicle growth and accelerates progression of primordial follicles to primary and early antral follicles.[25] Meanwhile, androgen excess may also reduce atresia of early antral follicles. These two effects combine and lead to excess numbers of small antral follicles and PCOM on ultrasonography. Additionally, androgen excess also stimulates hyperplasia of the ovarian cortex and stroma in studies of female-to-male transsexuals with excessive androgen exposure.[26]

The large number of small antral follicles produce excessive amounts of anti-Müllerian hormone (AMH), which counteracts the action of FSH on follicle development. Together with the accompanying dysfunctions in secretion of gonadotropins and altered intraovarian paracrine factors, the development of early antral follicles arrest and oligo-/an-ovulation occurs in PCOS (see below).

Metabolic Dysfunction

The effect of androgen on metabolism differs between males and females due to sex-specific activation of ARs in peripheral tissues and in the central nervous system. In men, androgen deficiency contributes to metabolic disturbance. However, in women androgen excess has adverse effects on metabolic homeostasis. Androgens widely act on peripheral metabolic organs and tissues, such as the adipose tissue, liver, skeletal muscle, and pancreas.[27] Androgen excess alters the pattern of fat distribution by promoting visceral fat accumulation. Abdominal obesity is closely correlated with hyperandrogenism, as evidenced by studies in women with PCOS, female-to-male transsexuals, and in androgenic animal models.[27] Abdominal adiposity in turn is an important risk factor for metabolic syndrome and cardiovascular disease (CVD).[28] Androgen excess also has a deleterious impact on glucose uptake and metabolism by adipocytes and skeletal muscle cells, which consequently impairs insulin sensitivity. Insulin resistance is the central pathophysiology of type

TABLE 124.3 Causes of Androgen Excess in Women

- Functional Gonadal Hyperandrogenism
  - Primary functional ovarian hyperandrogenism (common form of polycystic ovary syndrome)
  - Secondary polycystic ovary syndrome
  - Virilizing congenital adrenal hyperplasia
  - Ovarian steroidogenic blocks
  - Syndromes of severe insulin resistance
  - Acromegaly
  - Epilepsy ± valproic acid therapy
  - Disorders of sex development
  - Pregnancy-related hyperandrogenism
- Functional Adrenal Hyperandrogenism
  - Primary functional adrenal hyperandrogenism (uncommon form of polycystic ovary syndrome)
  - Congenital adrenal hyperplasia and related adrenal steroid metabolism disorders
  - Other glucocorticoid-suppressible hyperandrogenism
  - Prolactin excess
  - Cortisone reductase deficiency
  - Glucocorticoid-nonsuppressible functional adrenal hyperandrogenism
  - Cushing's syndrome
  - Glucocorticoid resistance
- Peripheral Androgen Overproduction
  - Idiopathic hyperandrogenism
  - Obesity
  - Portohepatic shunting
  - Androgenic drugs
- Tumoral Hyperandrogenism

Modified with permission from: Buggs C, Rosenfield RL. Polycystic ovary syndrome in adolescence. *Endocrinol Metab Clin North Am.* 2005;34:677–705.

2 diabetes. In women with PCOS, studies suggest that PCOS subtypes with hyperandrogenism have an increased risk of metabolic syndrome, insulin resistance, and hyperlipidemia compared with other subtypes.[29]

KEY POINTS

- The common clinical manifestations of androgen excess include hirsutism, polycystic ovary, and metabolic dysfunction.
- Mild hirsutism is unrelated to hyperandrogenemia in half of cases.
- Hirsutism, acne vulgaris, and male pattern balding are variably expressed in women with androgen excess.

COMMON ANDROGEN EXCESS DISORDERS

Polycystic Ovary Syndrome

Most women with androgen excess have ovarian hyperandrogenism as demonstrated by the results of ovarian vein catheterization, long-term administration of gonadotropin-releasing hormone (GnRH) agonists, dexamethasone suppression testing, or the acute ovarian steroidogenic response to the administration of a GnRH agonist or human chorionic gonadotropin (hCG).[30] The causes of gonadal androgen excess are listed in Table 124.3.

PCOS is the most common endocrine disorder in reproductive age women, with a prevalence between 5.5% and 19.9%.[31] Variance in reported prevalence may be due to the inconsistent diagnostic criteria and reflect the impact of ethnic origin, race, and environmental factors that can modify expression of the PCOS phenotype.[31] Since its first description in 1935 by Stein and Leventhal, diagnostic criteria for PCOS have evolved. Three international conferences have developed diagnostic criteria based on various combinations of otherwise unexplained clinical or biochemical evidence of hyperandrogenism, evidence of anovulation, and a PCO: National Institutes of Health (NIH) conference diagnostic criteria (1992), Rotterdam consensus criteria (2004), and Androgen Excess-PCOS Society (AE-PCOS) consensus criteria (2009).[24,25,32]

These criteria must be adapted to age, as adolescents display a high prevalence of physiologic anovulation and polycystic appearing ovaries.[31] Endocrine Society clinical guidelines suggest the criteria of persistent hyperandrogenism and oligomenorrhea.[33]

Reproductive Abnormalities. The signs of PCOS usually begin around menarche and persist thereafter. Anovulatory symptoms (oligomenorrhea, dysfunctional uterine bleeding, and/or infertility) occur in about 75%, and the remainder may have subclinical anovulation with eumenorrhea (phenotype 3).[31]

About 60% to 80% of PCOS patients have elevated levels of circulating androgens. DHEAS levels are elevated in about 25% of PCOS women and occasionally may be the sole abnormality of hyperandrogenemia.

The vast majority of PCOS patients have FOH, which can be identified after functional challenge. About 90% of PCOS women have elevated testosterone after adrenocortical function is suppressed by dexamethasone, and two thirds have a characteristic 17OH-prog hyperresponsiveness to GnRH agonist–induced endogenous gonadotropin release in the absence of evidence of a steroidogenic block.[34] *In vivo* and *in vitro* challenges with LH or the LH analog hCG result in markedly elevated 17OH-prog responses, and, to a lesser extent, elevated androgens, in PCOS versus non-PCOS women.

Functional adrenal hyperandrogenism (FAH) is also common in PCOS. Approximately 30% to 50% of patients with PCOS also have a characteristic adrenal dysfunction. This usually coexists with FOH but in 5% occurs independently as the adrenal form of PCOS.[34] This primary FAH is characterized by moderately excessive responses of DHEA and 17OH-preg to ACTH without evidence of a block in steroidogenesis. Elevated DHEAS is found in 20% to 60% of these cases. This has been postulated to represent the adrenal manifestation of a generalized steroidogenic dysregulation in PCOS.[35]

Excessive LH secretion relative to FSH was the first laboratory abnormality identified in classic PCOS.[34] Hypersecretion of LH occurs in about 90% of PCOS subjects as evidenced by elevated mean LH levels, an exaggerated LH response to GnRH, increased LH pulse amplitude, and increased LH pulse frequency. However, obesity suppresses LH levels, and most obese women have normal baseline and stimulated LH levels.[36] In addition, because of the pulsatile nature of LH secretion, a random LH level may be well within the normal range in a woman with PCOS.

Metabolic Abnormalities. About two-thirds of women with PCOS are insulin resistant, and the degree of insulin resistance in PCOS exceeds that of women without PCOS who are matched by body mass index (BMI) or degree of adiposity.[37] Women with PCOS are at increased risk for developing type 2 diabetes. Impaired glucose tolerance (IGT) or frank diabetes is evident in approximately 45% of women with PCOS by their fourth decade. Long-term follow-up of women with PCOS substantiates an increased prevalence of type 2 diabetes compared with that of appropriate control groups. The conversion rates from normal glucose tolerance to either IGT or type 2 diabetes are also high in women with PCOS.[38,39]

The development of elevated fasting or glucose challenge glucose levels is evidence of compromised ability of the pancreatic beta cells to compensate for insulin. Alterations in insulin secretion are demonstrable in PCOS women, which may have a genetic basis, since they are present more often in women with PCOS who have a first-degree relative with type 2 diabetes.[38]

Women with PCOS are also at increased risk for metabolic syndrome. Up to 43% of nondiabetic women with PCOS meet diagnostic criteria for metabolic syndrome well before the end of their fourth decade, and most of these meet the criteria before the end of their third decade of life.[28] Indeed, the prevalence of metabolic syndrome in young PCOS women is similar to that in women aged 50–60 years.[28] The metabolic syndrome and its individual components are particularly common among those PCOS women with the highest insulin levels and BMI.[28]

Women with PCOS are also at increased risk of cardiometabolic disease.[40] Dyslipidemia is reported in up to 70% of PCOS patients.[41] Various lipid patterns have been described, the most common being increased levels of low-density lipoprotein (LDL) cholesterol, very-low-density lipoprotein (VLDL) cholesterol, and triglycerides, and reduced levels of high-density lipoprotein (HDL) cholesterol (particularly decreased HDL2, the most anti-atherogenic HDL subtype) compared with control women. Qualitative disorders of LDL cholesterol have also been described and are characterized by an increased amount of small, dense LDL cholesterol, which is an independent predictor of cardiovascular morbidity and mortality. The latest systematic review and meta-analysis including 23 cohort studies found that women with PCOS had increased risks of hypertension (risk ratio: 1.75) and nonfatal cerebrovascular disease events (risk ratio: 1.41). Over 10 years of follow-up, the frequency of hypertension of women with PCOS has increased from 16.1% (95% CI: 11.0%–21.2%) to 34.7% (95% CI: 28.1%–41.3%). However, no significant differences were found for coronary disease events (risk ratio: 1.78, 95% CI: 0.99–3.23). Due to the paucity of mortality data, the meta-analyses could not be performed for fatal CVD events.[40] Consequently, prognosticating events from these findings should be done with caution. During 24 years of follow-up, the Nurses' Health Study II found that irregular and long menstrual cycles in adolescence and adulthood are associated with a greater risk of premature mortality (age < 70 years) from cancer and CVD. This relation is slightly stronger among women who smoke.[42] Future clinical trials are needed to assess the ability of cardiometabolic screening and management in women with PCOS.

Women with PCOS have an increased rate of nonalcoholic fatty liver disease (NAFLD) (hazard ratio: 2.23). In addition to increased BMI and dysglycemia, androgen excess contributes to the development of NAFLD in women with PCOS.[43] NAFLD is also associated with insulin resistance. Ultrasound evidence of fatty liver occurs significantly more frequently in obese PCOS subjects than in obese controls and is associated with hyperinsulinemia and insulin resistance. NAFLD was present in 41% of obese PCOS subjects in one series, and half of those had elevations of ALT, suggesting the possibility of the more severe nonalcoholic steatohepatitis.

Endometrial Carcinoma. Women with PCOS have about a 2.7-fold increased risk for endometrial cancer. Most endometrial cancers in this population are well-differentiated adenocarcinomas and have a good prognosis.[44] Although 5% or fewer endometrial cancers occur in women younger than 40 years of age, the majority have PCOS. PCOs are present in the pathologic specimens of 62% of women younger than 50 years of age with endometrial carcinoma.

Pathogenesis

Ovarian Dysfunction. FOH accounts for the vast majority of PCOS. Early *in vitro* studies showed that the ovary was the primary androgen source in PCOS.[34] Women with classic PCOS have a characteristic pituitary-ovarian response to testing with an acute GnRH agonist challenge. Their gonadotropin responses resemble those of males in that they have significantly greater early LH and lesser FSH responses than normal females. They have greater 17OH-prog and androstenedione, without evidence of a block in the subsequent steroidogenic steps or consistent abnormalities in earlier steps in the steroidogenic pathway, suggesting increased ovarian 17-hydroxylase and 17,20-lyase activities as part of a generalized overactivity of steroidogenesis. This dysregulation of ovarian steroidogenesis can manifest in the absence of LH excess or sonographic abnormalities.

Dysregulation of ovarian steroidogenesis appears to account for most FOH. Dysregulation in turn appears to result from resetting of the LH steroidogenic dose-response curve, "escape from downregulation," rather than from excessive LH stimulation. Both *in vivo* and *in vitro* studies indicate that the steroid responses of PCOS women do not fall along the normal LH steroid dose-response curve, but rather seem displaced upward and to the left,[34,44] indicating hypersensitivity to LH stimulation.

The cause of dysregulation of theca cell function could be intrinsic or extrinsic to the ovary. IGFs and inhibins are FSH-inducible ovarian factors that are capable of interfering with downregulation of ovarian steroidogenesis. Insulin is a key endocrine factor because it amplifies the effect of gonadotropins on steroidogenesis. The abnormality of thecal function in PCOS resembles the escape from downregulation that occurs when normal theca cells are treated with insulin or IGFs.[34] Studies of theca cell cultures support an intrinsic thecal abnormality in PCOS.[45,46] In primary and long-term cultures, PCOS theca cells secrete excessive amounts of androgens and androgen precursors, basally and after c-AMP stimulation. These findings persist when insulin is absent from the culture medium. Thus, the persistent increased androgen production of the PCOS ovary is a result of a generalized upregulation of specific steroidogenic enzymes and is consistent with an intrinsic theca cell abnormality rather than an effect of chronically excessive LH stimulation. This concept is supported by the finding of persistent thecal hypersensitivity to gonadotropins after long-term gonadotropin suppression in functionally typical types of PCOS.

Disturbed granulosa cell function underlies PCOM and arrested ovulation. Granulosa cells from patients with PCOS are hyperresponsive to gonadotropins. They overexpress the LH receptor and secrete estradiol in response to LH at an inappropriately early stage, which indicates premature luteinization. These small antral follicles also manifest inhibin B hyperresponsiveness to FSH. Granulosa cells furthermore overproduce estradiol in response to high levels of FSH, likely contributing to the increased risk for developing ovarian hyperstimulation syndrome (OHSS) during fertility treatment in women with PCOS.

Excessive folliculogenesis results from an increased proportion of follicles leaving the resting (primordial) phase to begin growing and then having a prolonged life span. It is related to disturbed elaboration of AMH, which regulates ovarian follicle growth by a paracrine negative feedback effect on the recruitment of primordial follicles.[47] It is possible that these granulosa cell abnormalities reflect an intrinsic defect in intraovarian regulation of early folliculogenesis. Complex mesenchymal–epithelial interactions and endocrine, paracrine, and metabolic influences affect follicular development.[47] An independent folliculogenesis disturbance is supported by evidence of persistently abnormal granulosa cell function after long-term suppression of ovarian hyperandrogenism and by the presence of PCOM in the absence of ovarian hyperandrogenism.

Oocyte gene expression is also dysregulated and the possibility has been raised that this might be related to high pregnancy loss of some PCOS patients. Good morphological oocytes have abnormal gene expression profiles that are associated with defective meiosis or early embryonic development.[48] As upregulated differentially expressed genes at the germinal vesical stage of PCOS mainly related to mitochondrial function, it suggests that aberrant mitochondrial function at the germinal vesical stage may contribute to a decline in oocyte quality of PCOS patients.[49] These data are compatible with excessive androgen and epigenetic metabolic signals contributing to the reduced developmental competency of PCOS oocytes.

Neuroendocrine Dysfunction. Elevated LH was initially thought to cause PCOS by increasing ovarian androgen production. However, the typical ovarian dysfunction of PCOS is often found in the absence of gonadotropin abnormalities. Furthermore, the normal process of desensitization to LH would seem to preclude LH excess from being a primary cause of androgen excess. It now appears that elevated LH may be the result rather than the cause of androgen excess. A modest rise in androgen levels normally stimulates LH pulsatility rather than suppresses it, which requires a substantial elevation. Patients with PCOS are less sensitive to suppression of the LH pulse frequency by sex steroids, compared with controls. Furthermore, anti-androgen treatment normalizes the inhibitory effect of progesterone on elevated LH pulse frequency. These results suggest that androgen excess increases LH pulse frequency by interfering with the hypothalamic inhibitory feedback of progesterone and that LH hypersecretion in PCOS results from androgen excess interfering with the inhibitory feedback of estradiol and progesterone.[50]

Insulin Resistance and Hyperinsulinism. Insulin resistance and compensatory hyperinsulinemia also play an important role in the pathogenesis of PCOS, a state in which steroidogenic and adipose tissues are paradoxically sensitive to insulin in the setting of overall resistance to the glucose metabolic effects of insulin.[37] Distinct postreceptor-binding defects in insulin signaling have been described in fibroblasts, adipocytes, and skeletal muscle cells in PCOS. These defects are specifically associated with impaired glycogen synthesis and glucose uptake and do not involve mitogenic signaling. In contrast to skeletal muscle and skin fibroblasts, there is no evidence for intrinsic defects in insulin signaling in the PCOS adipose cell lineage, although *in vivo* environmental effects could not be excluded.[37]

The relationship between insulin resistance and hyperandrogenism appears to primarily result from the effect of compensatory hyperinsulinemia on steroidogenesis. In the ovary, insulin normally acts in conjunction with LH to upregulate LH binding and enhance the androgenic response to LH (see Fig. 124.3). Clinical evidence uniformly indicates that steroidogenesis in PCOS proceeds as if sensitive to insulin, although it has been difficult to directly demonstrate an effect of insulin on ovarian function. All insulin-resistant states are associated with PCOS, as discussed later (see "Other Ovarian Hyperandrogenic Disorders" below). In rare instances, severe insulin resistance can be the primary pathogenic event in PCOS. Furthermore, any treatment that lowers insulin levels reduces androgen levels in PCOS. Insulin may act via its own receptor in theca cells by a signaling pathway separate from that mediating its metabolic effects. Insulin-resistant states also suppress SHBG, which increases testosterone bioavailability.

Androgen excess has been reported to cause insulin resistance *in vitro* in a minority of studies.[51]

In one of these, high-dose androgen induced resistance to insulin as stimulation of glucose uptake in cultured adipocytes from normal women. Treatment with the anti-androgen flutamide for 12 months significantly reduces visceral fat mass and improves insulin sensitivity.[52] Consequently, the potential exists for a vicious cycle between insulin resistance/hyperinsulinism and hyperandrogenism. Androgens may contribute to insulin resistance indirectly.[53] Visceral fat, which is characterized by increased lipolytic activity, is thought to be important to the pathogenesis of insulin resistance and the increased free fatty acid levels in PCOS women. Free fatty acids cause insulin resistance in the skeletal muscle. Thus, hyperandrogenemia may affect adipocyte insulin sensitivity so as to favor visceral body fat distribution, which results in relative increases of free fatty acids, which, in turn, induce skeletal muscle insulin resistance.

Etiology. The cause of PCOS is unknown, but considerable evidence suggests that it is a complex, heritable trait that first manifests at the onset of puberty and interacts with other congenital or environmental factors resulting in the PCOS phenotype.[38] About 50% of sisters of PCOS probands have hyperandrogenemia, and about one half of those will also be anovulatory. Furthermore, about 70% of sisters have PCOM, and one quarter are anovulatory. Diabetes-related defects in insulin secretion and action are common in primary relatives of women with PCOS. In a Dutch twin-family study, the correlation between monozygotic twins for PCOS was 0.71, and for dizygotic twin or non-twin sister pairs, the correlation was 0.38.[54] These family studies demonstrate a significant genetic component in the pathogenesis of PCOS. While the candidate gene approach has yielded no consistent association, genome-wide association screening has identified at least 16 candidate genes closely related to insulin signaling, sexual hormone function, and other.[55,56] Among them is the DENND1A gene that upregulates steroidogenesis by unrecognized mechanisms.[57]

The intrauterine environment may play a role in the pathogenesis of PCOS. Congenital virilization is well-established to cause PCOS in humans and in animal models. CAH is the most common, but not the exclusive, virilizing disorder causing this outcome. Untreated women with nonclassic CAH also are insulin resistant. As in multiple species, congenitally virilized rhesus monkeys have ovarian and adrenal hyperandrogenism, oligomenorrhea, polyfollicular ovaries, and elevated LH levels; they also have abdominal obesity, insulin resistance, IGT, and dyslipidemia. The LH hypersecretion seen in congenitally virilized animal models appears to result from congenital androgen exposure interfering with the negative feedback effect of ovarian steroids on LH, as in PCOS.[58] Besides this, elevated prenatal AMH can also reprogram the fetus and induce PCOS in adulthood.[59]

Inconsistent associations with low or high birth weight have been reported to be associated with PCOS. In some populations, intrauterine growth restriction is associated with premature adrenarche and PCOS risk. A meta-analysis including 42 studies in 7041 women with PCOS and 63,722 women without PCOS found that PCOS was associated with higher gestational weight gain, higher risks of preterm birth and large for gestational age, and lower birth weight, while this association varied by geographic continent, PCOS phenotypes, and study quality. However, PCOS was associated with none of these outcomes on BMI-matched studies.[60]

Postnatal environmental risk factors for PCOS include obesity and other insulin-resistant states, which can be viewed as precipitating latent susceptibility traits to become manifest. Premature adrenarche appears to pose about a 2-fold increase in PCOS risk in some populations. It is likely that this represents the first manifestation of dysregulated steroidogenesis in these cases.

Ovarian Neoplasms

Androgen-secreting ovarian tumors are very rare causes of androgen excess. About 1% of ovarian tumors that comprise testicular cell types can cause hyperandrogenism followed by characteristic virilization.[61] Virilizing ovarian tumors may be malignant or benign and are frequently suspected by abrupt onset, usually distinct from menarche, of rapidly progressive virilization, testosterone concentrations > 200 ng/dL, and characteristic ultrasound or computed tomography (CT) findings. However, androgen elevations are sometimes modest, in which case the presentation will be indolent. The most common virilizing ovarian tumor is the Sertoli-Leydig cell tumor, which may be gonadotropin-responsive. Ovarian lipid cell tumors are also typically gonadotropin-responsive and also dependent, in part, on ACTH. Ovarian virilizing tumors typically secrete predominantly androstenedione and are thus characterized by a disproportionate elevation of serum androstenedione relative to testosterone. Mild elevation of urinary 17-ketosteroid excretion characteristically results. Ovarian androgen-secreting tumors represent a highly heterogenous group of rare endocrine tumors with possible genetic penetrance including a *FOXL2* mutation. Metabolic implications and long-term outcomes of these kind of tumors are unknown and require more detailed follow-up studies.[61]

Other Ovarian Hyperandrogenic Disorders

Symptoms consistent with PCOS are present in about half of women with classic CAH and about one fifth of women with nonclassic CAH.[62] It is associated with a higher degree of LH hyperresponsiveness to GnRH agonist testing than in PCOS. Adrenal rests of the ovaries, as are found in rare cases of CAH, may also mimic PCOS.

Ovarian steroidogenic blocks in the pathway to estradiol are rare causes of hyperandrogenism and multicystic ovaries. This can result from congenital defects in the activity of ovarian 3β-HSD, 17β-HSD, or aromatase (see Fig. 124.3). The severity of hypoestrogenism depends upon the degree of deficiency of enzyme activity. Aromatase gene mutations reported to date have resulted in congenital virilization with ambiguous genitalia.

Subjects with “true hermaphroditism” (ovotesticular disorder of sex development) are characterized by the presence of functional testicular and ovarian tissue (see Chapter 102). They occasionally present as menstruating, phenotypic females with signs of androgen excess such as clitoromegaly. The diagnosis can be made by the response to gonadotropin stimulation.

All extreme states of insulin-resistant hyperinsulinemia, such as congenital diabetes mellitus caused by insulin-receptor mutations (e.g., Donohue’s syndrome or leprechaunism) or lipodystrophy, are accompanied by PCOS.[34] Less extreme, but nevertheless severe, insulin resistance (pseudo-Cushing’s syndrome and pseudo-acromegalic gigantism) in childhood may precede PCOS. In addition, modest forms of insulin resistance also are associated with PCOS, including type 1 and type 2 diabetes mellitus. Like IGF-1, elevated insulin levels seem to cause PCOS by increasing the activity of steroidogenic enzymes in the ovaries and adrenal glands.

Acromegaly may also be associated with PCOS, possibly because of the steroidogenic effects of elevated IGF-1 on the ovaries and adrenal cortex.[63] Its presentation is usually insidious.

Epilepsy is significantly associated with PCOS independent of drug treatment, although valproic acid directly augments transcription of the steroidogenic gene CYP17 and can cause elevated serum testosterone.[64]

The PCOS phenotype can occur as a complication of portal hypertension or portosystemic shunting.[65,66] This has been attributed to impaired steroid metabolism. Virilization during pregnancy may be due to a luteoma or hyperreactio luteinalis.[67] These conditions represent benign hyperplastic thecal luteinization of the ovary, solid and cystic respectively, which spontaneously regress postpartum. Luteomas appear not to be associated with excessive levels of hCG; in contrast, hyperreactio luteinalis is almost always associated with hCG excess and may be due to defective regulation of its placental secretion. Maternal virilization occurs in 10% to 50% of women with luteomas and in 25% of those with hyperreactio luteinalis. Fetal virilization is rare owing to placental aromatization of androgen to estrogen.

Congenital Adrenal Hyperplasia

CAH is a term historically applied to disorders that arise from an autosomal recessive deficiency in the activity of an adrenocortical enzyme step necessary for corticosteroid biosynthesis (see Fig. 124.4).[2] Because of inefficient cortisol secretion, there is diminished negative feedback inhibition of ACTH secretion, which markedly increases its secretion and leads to compensatory adrenocortical hyperplasia. The result is accumulation of androgen precursors proximal to the enzyme defect. CAH is potentially life-threatening in its classic (severe) form and may be asymptomatic or cause female infertility in its nonclassic (mild) form.[68] It is now appreciated that CAH can occur as the result of enzyme deficiency in any step of corticosteroid biosynthesis: 21-hydroxylase, 3β-HSD type 2, 11β-hydroxylase, and P450 oxidoreductase. About 95% of CAH is due to 21-hydroxylase deficiency and other forms of CAH are very uncommon.[69] Each enzyme deficiency leads to characteristic clinical syndromes and biochemical abnormalities, as described below.

21-Hydroxylase Deficiency. Classic CAH due to 21-hydroxylase deficiency occurs in 1 in 10,000 to 1 in 20,000 live births based on data from millions of newborns screened worldwide. It is subclassified into salt-wasting and simple virilizing forms diagnosed during infancy, but affected individuals may develop signs and symptoms resembling PCOS during adolescence. Classic 21-hydroxylase deficiency typically presents with genital ambiguity due to congenital virilization. Screening for CAH due to 21-hydroxylase deficiency by 17OH-prog assays are recommended in all newborn screening programs.[70]

Nonclassic (“late-onset” or “attenuated”) 21-hydroxylase deficiency, a milder form of CAH with residual 21-hydroxylase enzyme activity, can present as premature pubarche or with peripubertal or postpubertal onset of hirsutism, acne, and/or amenorrhea. After PCOS, it is the next most common cause in women of androgen excess presenting in adolescence and early adulthood and accounts for about 2%–3% of hirsutism in the general population[32] and 4.2% of women with androgen excess.[62] Some ethnic groups, e.g., Ashkenazi Jews (prevalence 1:27), Hispanics (1:40), and Slavs (1:50), are at much higher risk than White Americans (1:200).[71] They can exhibit oligo-ovulation, polycystic-appearing ovaries, and elevated serum LH levels. Some cases are entirely cryptic with no clinical manifestations, discovered serendipitously in family studies. The degree of clinical severity is correlated with the degree of 21-hydroxylase deficiency elucidated by ACTH testing. ACTH response patterns and molecular genetic techniques have shown that the rapid 17OH-prog response to an intravenous bolus of ACTH distinguishes nonclassic cases and heterozygotes, from normals better than baseline steroid levels.[72]

Mutations in *CYP21A2* gene, located on the short arm of chromosome 6, result in lack of 21-hydroxylase. A continuum of disease phenotypes generally correlates with *CYP21A2* genotypes. Almost 300 *CYP21A2* mutations are known, most with defects of gene conversions and large deletions.[68,70]

3β-Hydroxysteroid Dehydrogenase Type 2 Deficiency. Deficiency of 3β-HSD type 2 due to *HSD3B2* gene mutations classically presents in infancy with genital ambiguity due to congenital virilization, accompanied by decreased cortisol, aldosterone, and androstenedione concentrations, but increases in renin, ACTH, and DHEA. Rarely, it presents later in its nonclassic form as hyperandrogenic anovulation. The phenotype varies according to different genetic defects with *HSD3B2* mutations.[73]

11β-Hydroxylase Deficiency. Deficiency of 11β-hydroxylase (CYP11B1) classically presents with genital ambiguity caused by congenital virilization, accompanied by hypertension and advanced bone age. Late-onset and the nonclassic form of hyperandrogenism is rare, and hypertension and hypokalemia are not necessarily present.[2] This disease is more common in Middle East and North America. The accumulated precursor substrate 11-deoxycortisol, although rarely measured routinely, may serve as a serum marker for 11β-hydroxylase deficiency.[74]

Cytochrome P450 Oxidoreductase Deficiency. Cytochrome P450 oxidoreductase deficiency (PORD), a rare, autosomal recessive form of CAH, is commonly diagnosed during infancy or childhood. Based on the severity of *POR* gene mutations and subsequent impairment of POR-dependent enzymes, PORD manifests with a broad spectrum of clinical phenotypes from menstrual cycle abnormalities, infertility, and polycystic-appearing ovaries, to ambiguous genitalia in both 46, XX and 46, XY women.[75] Some nonclassic PORD without genital and skeletal malformations in postpubertal females are very rare and remain poorly characterized.

Other Adrenal Hyperandrogenic Disorders

Adrenal Neoplasms. Adrenal tumors are a very rare cause of hyperandrogenism. Rapid progression of androgenic symptoms, signs of virilization, and elevated androgen levels usually make the tumor diagnosis straightforward. Very high testosterone levels (>200 ng/dL) suggest an androgen-secreting tumor, but some tumors present with lower levels, and only about half of patients with virilizing adrenal tumors have a marked elevation of DHEAS (>700 µg/dL). If an androgen-secreting tumor is suspected, pelvic ultrasound should be done. If negative, it should be followed by an adrenal CT scan.

Glucocorticoid-Suppressible Hyperandrogenism. Hyperprolactinemia may present as hyperandrogenic oligo-anovulation; galactorrhea is typical, although not necessarily present. Polycystic-appearing ovaries and LH hyperresponsiveness to GnRH may be associated with a prolactinoma and acromegaly. DHEA and DHEAS levels are elevated, as is the serum free testosterone, in part because of a direct effect of high levels of prolactin on adrenal cortical function. Mild to modest elevations in prolactin levels can also be found in otherwise classic PCOS. In these instances, it is important to establish that the prolactin elevation is not a consequence of primary hypothyroidism, prolactinoma, or medications.

Cortisone reductase deficiency is a rare autosomal recessive disorder that results from a defect in the peripheral metabolism of cortisol. Increased ACTH production is necessary to compensate for the excessively rapid cortisol turnover caused by failure to regenerate cortisol from cortisone in peripheral tissues. This disorder is caused either by deficiency of 11β-HSD type 1 (HSD11B1), which is functionally the major cortisone reductase, or more commonly by "apparent" cortisone reductase deficiency, inactivating mutation of hexose-6-phosphate dehydrogenase (H6PDH), which is a cofactor for cortisone reductase. Establishing this diagnosis requires urinary steroid profiling with increased urinary 17α-hydroxycorticoids, but low cortisol excretion.

Apparent DHEA sulfotransferase deficiency is due to a genetic defect in DHEA metabolism that does not affect corticosteroid secretion. Inactivating mutations of the sulfate donor to SULT2A1 (3′-phosphoadenosine-5′-phosphosulfate synthase 2, PAPSS2) prevent sulfation of DHEA, yielding high DHEA but undetectable DHEAS levels. Premature pubarche preceded hyperandrogenic anovulation in two reported girls, who also suffered skeletal growth defects.[77]

Glucocorticoid-Nonsuppressible Hyperandrogenism. Glucocorticoid resistance is a rare congenital form of ACTH-dependent adrenal hyperandrogenism. It results from defects in glucocorticoid receptor signaling, resulting in inadequate negative feedback by cortisol, with consequent excessive ACTH release and adrenal hyperplasia, elevated cortisol, and DHEAS secretion. Cushing's syndrome due to diverse forms of adrenal hyperplasia and neoplasm is associated with hyperandrogenic anovulation, although this is seldom the presenting feature. Moreover, polycystic-appearing ovaries can occur in Cushing's syndrome.

Idiopathic Hyperandrogenism

Approximately 8% of hyperandrogenic patients have no ascertainable glandular source of androgen despite thorough testing. Patients who have normal ovulatory cycles and normal ovaries, independently of having elevated androgen levels or with hirsutism, are traditionally diagnosed as having *idiopathic hyperandrogenism* (IHA). IHA is a diagnosis of exclusion in hyperandrogenic patients who do not fulfill the diagnosis of PCOS, nonclassic adrenal enzymatic deficiencies, or other uncommon hyperandrogenic disorders. The pathogenesis of IHA is largely unknown.

Obesity can modify the androgenic phenotype, accounting for most cases of IHA and can mimic PCOS by causing hyperandrogenic anovulation. In PCOS patients who meet NIH diagnostic criteria, the IHA subset without an identifiable ovarian and adrenal source for androgen excess were obese. These patients were characterized by mild hyperandrogenemia, and most had normal size ovaries and normal LH and AMH levels.[78] Adipose tissue is a major site for the formation of testosterone from the circulating precursor androstenedione. Thus, obesity is accompanied by both increased testosterone production and suppressed SHBG. Gonadotropins are suppressed in obesity, in part from increased clearance, and possibly partly related to abnormal estrogen metabolism. Estrone formation from androstenedione is increased, and estradiol metabolism is diverted to active rather than inactive metabolites.

Other contributors to IHA include increased peripheral metabolism of inactive steroid precursors to active androgens by hereditary defects and anabolic steroid use. The latter can cause virilization in women and present with features similar to those of virilizing tumors. Anabolic steroids should be suspected in athletes and women with sexual dysfunction (see Chapter 128). Importantly, many steroids used for these purposes are not detected by routine androgen assays.

KEY POINTS

- PCOS is the most common cause of hyperandrogenism in women.
- Funcional ovarian hyperandrogenism due to a unique type of steroidogenic dysfunction is the essential feature of most PCOS.
- PCOS is associated with obesity and insulin resistance, and the ovarian hyperandrogenism and metabolic abnormalities require distinct treatment considerations.

DIAGNOSIS OF HYPERANDROGENISM

Clinical Hyperandrogenism

Clinical hyperandrogenemism should be considered in women with hirsutism, moderately severe acne vulgaris, or other cutaneous manifestations including seborrhea and alopecia. Severe excess androgen can result in virilization manifesting as clitoromegaly, masculinization, and other features described above. Hirsutism is the most common feature of hyperandrogenism. Over 70% of hirsute women showed increased androgens and are often accompanied with other pathophysiological features.[31] Most women with focal hirsutism and regular menses who have no other evidence to suggest an underlying cause have a very low likelihood of excess androgen production and do not require an endocrine workup. Similarly, women with isolated mild hirsutism (score 8 to 15) are unlikely to have a medical disorder that would change management. However, in women with symptomatic hirsutism of any degree, or acne that is poorly responsive to topical treatment, androgen levels should be assessed. Rapid development or progression of hirsutism, progression despite therapy, or evidence of virilization (such as clitoromegaly, increasing muscularity, deepening of the voice, or balding) increase the likelihood of an androgen-secreting neoplasm or exogeneous androgen use. However, tumors producing only moderate amounts of androgen have indolent progression of signs and symptoms.[20] Evaluation of a woman for hirsutism includes assessment of the pattern of hair growth, so as to differentiate it from hypertrichosis (see earlier), and its quantity. The mFG is the most common visual assessment tool.

Biochemical Hyperandrogenism

Markers for Diagnosis. Although hirsutism constitutes clinical evidence of hyperandrogenism in the consensus diagnostic criteria for PCOS, it is not necessarily indicative of androgen overproduction, as discussed earlier. Guidelines suggest that testing for hyperandrogenemia center around measurement of serum testosterone. Unfortunately, no optimal marker or test method for diagnosis can be recommended based on current evidence.

Testosterone was suggested as the key androgen in past decades. High-quality assays, such as LC-MS/MS and extraction/chromatography immunoassays, are suggested to assess total or free testosterone. However, these reliable testosterone assays are not available to many physicians. The automated assays used in most laboratories are generally not suitable for clinical assessment. Systematic differences between assays and excessively broad normal ranges derived from populations of apparently normal women with unrecognized androgen excess further complicate interpretation of testosterone levels provided by many laboratories.

An elevation in serum bioavailable (includes loosely bound to albumin) testosterone is relatively the single most sensitive test to establish the presence of hyperandrogenemia compared with total testosterone, DHEAS, and androstenedione. The serum bioavailable testosterone concentration is about 50% more sensitive for the detection of hyperandrogenemia than the total testosterone concentration. This results from SHBG, which is the main determinant of the bioactive portion of serum testosterone, the fraction that is bioavailable, often being low in hyperandrogenic women. However, there are no reliable direct assays for bioavailable testosterone. Calculated bioavailable testosterone, calculated free testosterone, and the free androgen index (FAI = 100 × (total testosterone/SHBG) are recommended according to the guideline to assess biochemical hyperandrogenism in PCOS.[79] Considering the secretion of testosterone, serum bioavailable testosterone is best assessed in the early morning in amenorrheic women or in the early follicular phase of regularly cycling women. Patients with an initial normal testosterone level in whom clinical features consistent with hyperandrogenism are present or emerge should have repeat testing, preferably an early morning serum bioavailable testosterone level in a reliable specialty laboratory.

Routine testing for other androgens, such as DHEAS and androstenedione, can be considered when bioactive testosterone is not raised, but provides limited additional information in the diagnosis of PCOS. The level of DHEAS is increased in approximately 15% of women who have normal levels of total and bioavailable testosterone. These two markers are more useful in differential diagnosis of specific androgen excess disorders. DHEAS which is of adrenal origin can be elevated slightly in some PCOS patients. But if it increases significantly, sometimes along with virilization, an adrenal tumor should be suspected. Increased androstenedione is a feature of 21-hydroxylase–deficient nonclassic CAH.

Evaluation. It is because reliable assays are not available that there are no precise cutoff values of the above markers for diagnosis of hyperandrogenism. A mildly elevated level in a woman with a normal bioavailable testosterone level is unlikely to be clinically relevant aside from being associated with acne. Very high levels of total testosterone (>200 ng/dL [6.94 nmol/L]) or of DHEAS (> 700 µg/dL [19 µmol/L]) heighten the likelihood of an underlying ovarian or adrenal neoplasm. However, such high levels usually do not indicate an androgen-secreting neoplasm, and lesser elevations may occur with androgen-secreting tumors. If a neoplasm is suspected, a transvaginal ultrasound should be performed. If no ovarian tumor is identified, further imaging studies (e.g., abdominal CT for an adrenal neoplasm) may be warranted. Nearly one third of women with PCOS can be identified as having hyperandrogenemia by total testosterone measurement, while more than half can be identified by bioavailable testosterone. Once hyperandrogenemia is documented, the diagnosis of PCOS depends on exclusion of other hyperandrogenic disorders. The laboratory evaluation to determine the cause of hyperandrogenism varies among specialists, and the effect of various approaches on outcomes is uncertain.

Ultrasonography, performed initially, can exclude rare but serious ovarian (and sometimes adrenal[80]) tumors. In addition to identifying ovarian changes consistent with PCOS, other ovarian pathology such as hermaphroditism can be detected. Those whose ultrasound shows a tumor or other explanation for hyperandrogenism should be referred for treatment of the underlying disorder. Although the presence of polycystic-appearing ovaries on ultrasound is one of the diagnostic criteria for PCOS, ovaries from women with PCOS are sonographically normal in about 25% of cases. Conversely, PCOs may be found on ultrasound examination in up to 50% of women without PCOS. Therefore, the following endocrine studies can be considered in women with hyperandrogenism regardless of whether or not the ultrasound shows normal or polycystic ovaries: prolactin, thyroid function, 17OH-prog, and possibly a measure of cortisol production. All of these endocrine studies will be normal in patients with PCOS. An early morning non-luteal phase 17OH-prog value below 170 to 200 ng/dL excludes nonclassic CAH with ~95% specificity.[81] An abnormal result in urine cortisol, late night salivary cortisol, or 1 mg overnight dexamethasone suppression test would provide evidence for endogenous Cushing's syndrome,[82] although a more definitive workup is indicated if Cushingoid features are present (**see Chapter 97**). An abnormality of any of these tests suggests that the androgen excess has a cause other than PCOS, which should be pursued accordingly.

More thorough endocrine testing, as recommended by the American Association of Clinical Endocrinologists,[83] will detect the small fraction of patients who have rare disorders that mimic PCOS and are missed by the above approach. The cost-effectiveness of such an approach is unknown, however. Further workup to identify the origin of androgen excess may also include: (1) assessment of androstenedione, the immediate precursor for testosterone, or other steroid intermediates; (2) assessment of the response to cosyntropin of 17OH-prog, DHEA, and other steroid intermediates as appropriate, and/or genotyping to exclude forms of CAH; (3) assessment of urinary corticoid metabolites to exclude cortisone reductase deficiency; (4) dexamethasone suppression test to suppress androgens arising from a functional adrenal source; (5) acute GnRH agonist administration to distinguish the steroidogenic response of PCOS/hyperthecosis from an ovarian androgen-secreting tumor; (6) CT or more specialized imaging studies if there is reason to suspect an androgen-secreting tumor; and (7) assessment of the suppressive response to combined oral contraceptive (COC) or GnRH agonist treatment.

KEY POINTS

- Hyperandrogenism can be diagnosed clinically and/or biochemically.
- Clinical hyperandrogenism should be considered in women with hirsutism, moderately severe acne vulgaris, seborrhea, and alopecia.
- Biochemical hyperandrogenism is frequently determined by testing serum bioavailable testosterone, however, no optimal marker or test method for diagnosis can be recommended based on current evidence.

TREATMENT OF POLYCYSTIC OVARY SYNDROME AND OTHER ANDROGEN EXCESS DISORDERS

Androgen Suppression

Androgen suppression generally involves reduction of androgen levels or their end-organ effects. This is usually accomplished by (1) suppression of ovarian androgen production, (2) alteration of binding of androgens to their serum binding proteins, (3) impairment of the peripheral conversion of androgen precursors to active androgen, and (4) inhibition of androgen action at the target tissue level. With all medical treatments, there appears to be an inverse association between efficacy and BMI.

Combined Oral Contraceptives. COCs alone are recommended in PCOS adult patients for management of hyperandrogenism. Adolescents with a clear diagnosis of PCOS can be prescribed with COCs as well. The estrogenic component of most COCs currently in use is ethinyl estradiol. It is responsible for the suppression of LH and thus serum androgen levels and results in a dose-related increase in SHBG, thus lowering the free fraction of serum testosterone. COC therapy has also been demonstrated to decrease DHEAS levels, perhaps by reducing ACTH levels.

In theory, the choice of a specific COCs should be predicated on the progestational component, as each progestin has a variably suppressive effect on SHBG levels and variable androgenic effect on metabolic biomarkers. Progestins promiscuously signal through diverse steroid pathways. First-generation progestins such as norethindrone, norgestrel, and levonorgestrel are mildly androgenic, as judged from their attenuation of the estrogen-induced increase in SHBG. Norgestimate and desogesterel are virtually nonandrogenic; cyproterone is anti-androgenic. Despite these subtle differences in progestin, common symptoms in PCOS can be controlled with COCs that contain lower doses of ethinyl estradiol combined with most progestins.[84] Thus, there is no specific type of COC or dose of estrogen or progestin that is recommended.

Assessment of adequacy of ovarian suppression can be made as early as the end of the third week after starting treatment; by this time, ovarian androgen suppression is complete. The effect on acne can be expected to be maximal in 1 to 2 months. In contrast, the effect on hair growth may not be evident for 6 months, and the maximum effect requires 9 to 12 months.[85]

With absolute and relative contraindications as well as side effects of COCs, detailed assessments for PCOS patients can help to weigh the risks and benefits of using them for individuals. Elevated risks of venous thromboembolism, CVD, and metabolic morbidities of COCs should be taken into consideration, especially for PCOS patients with specific risk factors such as high BMI, hyperlipidemia (including hypertriglyceridemia), and hypertension. According to the World Health Organization (WHO), absolute contraindications for COCs include deep vein thrombosis (DVT)/pulmonary emboli (PE), known thrombogenic mutations, multiple risk factors for arterial CVD, history of ischemic heart disease or stroke, complicated valvular heart disease, breast cancer, and others.[86] COCs with 35 μg ethinyl-estradiol plus cyproterone acetate (not available in the United States) are not recommended as the first-line medication for PCOS and should only be used for moderate to severe hirsutism or acne, whereas levonorgestrel, norethindrone, and norgestimate are associated with relatively low risk of DVT. Anti-androgens can be considered for hirsutism and androgen-related alopecia when COCs are contraindicated or poorly tolerated, and no anti-androgen has been shown to be more efficacious than another. If 6 months or more of COCs and cosmetic therapy fail to improve hirsutism, anti-androgens can be added.

Anti-Androgens. Cyproterone acetate (not available in the United States) is the prototypic anti-androgen.[22] It acts mainly by competitive inhibition of the binding of testosterone and DHT to the AR. In addition, it may act to enhance the metabolic clearance of testosterone by inducing hepatic enzymes. By means of its progestogenic properties, cyproterone acetate can also suppress the secretion of gonadotropins and result in decreased ovarian androgen synthesis. And because of its potent progestational activity and its prolonged half-life, it is administered in a “reverse sequential” manner: cyproterone acetate 50 to 100 mg/day from day 5 to 15 of the cycle, with ethinyl estradiol in a dose of 35 to 50 mcg/day from day 5 to 26 of the cycle. The dose of cyproterone acetate may be reduced incrementally at 6-month intervals after maximum improvement has occurred. Side effects of cyproterone acetate include irregular uterine bleeding, nausea, headache, fatigue, weight gain, and decreased libido. To prevent potential feminization of a male fetus, effective forms of contraception should be used with anti-androgens.

Spironolactone in doses of 100 to 200 mg/day is as effective an anti-androgen as cyproterone acetate.[22] It is a potent antimineralocorticoid developed as a progestational analog that has diverse effects on steroid metabolism. Its side effects are remarkably few; however, patients must be monitored for hyperkalemia and hypotension. Its most frequent side effect is menstrual irregularity when used alone.

Flutamide is a potent nonsteroidal anti-androgen without progestational, estrogenic, corticoid, antigonadotropic, or androgenic activity that is efficacious in the treatment of hirsutism.[22] Because of the risk for significant hepatotoxicity, however, flutamide should not be used as initial therapy for hirsutism. If flutamide is needed to treat severe hirsutism unresponsive to other anti-androgens, it should be used at low doses (no more than 375 mg/day).[22] Liver function tests should be monitored periodically on therapy.

Other Agents for Treatment of Androgen Excess. Finasteride, a 4-aza-steroid, is a competitive inhibitor of type 2 5α-reductase. Although beneficial effects on hirsutism have been reported,[22] the prominence of type 1 5α-reductase in the PSU makes it unlikely to be an optimal form of treatment. Insulinsensitizing agents such as metformin and thiazolidinediones can modestly improve hyperandrogenism.[38] Metformin is not as effective as COCs in the management of hirsutism in adult PCOS patients with patients with BMI between 25 and 30 kg/m^2.[87] Additionally, regarding treatment for hirsutism, either metformin alone or COCs alone may be less effective than the two combined. PCOS patients with high metabolic risk, such as diabetes risk or IGT, may benefit most from the combination of COCs with metformin.

Chronic administration of GnRH agonist suppresses pituitary-ovarian function, thus inhibiting both ovarian androgen and estrogen secretion. Uncontrolled trials of GnRH agonists in ovarian hyperandrogenism have reported significant reduction in the mFG score. Because of the reduction of serum estrogen levels and the concomitant reduction in bone mineral density with GnRH agonist therapy, an agonist should always be combined with estrogen-progestin add-back for long-term use. This regimen is more effective for hirsutism than are COCs alone.[22] GnRH agonist therapy may be most useful in women with severe forms of ovarian hyperandrogenemia such as ovarian hyperthecosis, but the effect on hirsutism is modest.[22] In 2018, a selective oral GnRH antagonist, elagolix, has been approved alone or in combination with progestin add-back therapy to treat endometriosis and leiomyomas, respectively. These therapies may also potentially offer benefit to women with PCOS.

Ovulation Induction and Assisted Reproduction

Infertility is a major health complication of PCOS and mainly caused by oligo/anovulation. Approximately 40% of PCOS women present with infertility, requiring fertility treatment. Ovulation induction and assisted reproduction are crucial managements for infertile women with PCOS. Ovulation and ovulation induction are adversely affected by overweight and obesity status. Healthy lifestyle is recommended in all PCOS patients and weight loss is recommended for obese PCOS women, as first-line treatment for infertility. Modest weight loss of 5% to 10% may result in return to regular menses or unassisted pregnancy in many women with PCOS.

Ovulation Induction. Letrozole, the best characterized aromatase inhibitor for ovulation induction, is the first-line agent an/oligo-ovulatory infertile PCOS women, although they are not approved for this indication by the US Food and Drug Administration. Aromatase inhibitors prevent conversion of androgens to estrogens. The usual letrozole dose is 2.5 to 5 mg/day for 5 days. According to a Cochrane review, compared to clomiphene citrate (CC), the chance of live birth increased 40%–60% for letrozole.[88] Moreover, ovulation failure and risk of multiple gestation are favored in letrozole versus CC. A daily dose of 5 mg of letrozole resulted in a similar pregnancy rate to 75 IU FSH, despite a significantly lower rate of ovulation in a large study of women not ovulating on 100 mg of CC.[89] A randomized controlled trial including clomiphene-resistant PCOS patients demonstrated that letrozole was comparable with CC in combination with metformin with regard to ovulation rate, pregnancy rate, and miscarriage rate per patient.[90] When letrozole is not available or not approved, other medications for ovulation induction can be used, including CC.

CC, a selective estrogen receptor modulator, can be used alone for ovulation induction for infertile women with PCOS. The usual regimen is 50 mg/day for 5 days. The dose can be increased by 50 mg/day to a maximum dose of 150 mg/day in subsequent cycles if ovulation does not occur. Using a graduated regimen of 50 to 150 mg daily for 5 days, ovulation can be induced in about 80% of women. The probability of ovulation and live birth decreases with increasing androgen levels, BMI, and age. There is a 5%–7% incidence of twin pregnancy, a 0.3% incidence of triplet pregnancy, and a < 1% OHSS rate associated with the use of CC. According to a 2019 systematic review, CC plus metformin increased the ovulation rate and clinical pregnancy rate compared with CC used alone.[91] If PCOS patients are clomiphene-resistant, metformin can be added in ovulation induction instead of persisting with CC alone.

Metformin can also be used alone for anovulatory infertility in PCOS women; however, clinicians should inform patients of other medications that are superior in ovulation induction. A systematic review suggested that metformin achieved increased ovulation, pregnancy, and live birth rates versus placebo, and was associated with a high rate of gastrointestinal side effects.[91] Another Cochrane systematic review compared effects of CC and metformin on pregnancy outcomes and found no significant differences in ovulation rate, pregnancy, and live birth rates.[92]

For those who fail first-line oral agents, gonadotrophins can be used as second-line agent for ovulation induction. A systematic review demonstrated equivalence of urinary-derived gonadotrophins and recombinant FSH with clinical pregnancy, live birth, multiple pregnancy, and miscarriage rates.[93] In PCOS, stimulation is initiated with 37.5 to 75 IU/day, instead of the standard 150 IU, and patients are maintained on this dose for 1 to 2 weeks. This low-dose regimen decreases risk for OHSS and multiple gestation. Gonadotropin dose is stepped up at 37.5 IU increments every 7 days if no follicular growth occurs at the initial dose as determined by ultrasound. Alternatively, treatment can be initiated at higher doses, 100 IU, and "stepped down" when follicular growth is detected by ultrasound. The step-up protocol is more likely to result in monofollicular development and less likely to result in OHSS versus the step-up protocol. Low-dose regimens result in monofollicular development in about 70% of cycles, pregnancy rates of ~20%, multiple gestation rates of ~6%, and < 1% risk for OHSS. About 95% of patients will ovulate with a 55% cumulative pregnancy rate after six cycles. To bring down the multiple pregnancy rate, only when fewer than three mature follicles are developed can ovulation be triggered; otherwise canceling the cycle is recommended.

Gonadotrophins can be considered as first-line treatment for ovulation induction if a conservative stimulation protocol is used along with ultrasound monitoring and counselling on cost and the possibility of multiple pregnancy. Compared with CC plus metformin, use of gonadotrophins is preferable for clomiphene-resistant women with PCOS.[15] Gonadotropins can also be used with metformin since preliminary evidence indicated that adding metformin for PCOS patients undergoing ovulation induction with gonadotropins might improve the live birth rate.[94]

Assisted Reproductive Technology. For PCOS patients who fail to conceive with ovulation induction or have other concomitant causes of infertility (e.g., severe male factor), assisted reproductive technology including *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI) and embryo transfer (ET) can be considered. Although IVF/ICSI is more inconvenient and costly, it provides PCOS patients with good chances for pregnancy and minimal multiple pregnancy risk with elective single ET. However, the elevated risk of OHSS during controlled ovarian hyperstimulation in IVF/ICSI still a major clinical concern. To mitigate this risk, GnRH antagonists have been included in the protocol for ovarian stimulation. Compared with the GnRH agonist long protocol, GnRH antagonist protocol reduces the total gonadotropin dose, the stimulation days, and the risk of OHSS. A systematic review and meta-analysis revealed a significantly decreased OHSS risk in the GnRH antagonist group (RR: 0.53, 95% CI: 0.30–0.95) compared with the long agonist protocol group, and no difference in the number of oocytes retrieved, ongoing pregnancy rates, and live birth rates.[95] As for triggering oocyte maturation, when hCG is used, it is recommended to use at the lowest dose in order to prevent OHSS. Patients undergoing GnRH antagonist protocol can administer a single bolus of GnRH agonist to induce the endogenous LH surge, leading to lower incidence of OHSS than hCG to trigger oocyte maturation and ovulation. However, a Cochrane review suggests that GnRH agonist trigger is associated with lower ongoing birth rate and live birth rate, as well as a higher miscarriage rate after fresh ET, which can be overcome in cycles of frozen ET. So, for women undergoing cycles with a GnRH antagonist protocol who plan to cancel fresh ET or who are at a high risk of OHSS, trigger with GnRH agonist can be implemented. Some clinicians also add low amounts of hCG to the agonist trigger to achieve a better mature oocyte yield. An elective "freeze-all" strategy can be an option for women with PCOS as well. A randomized controlled trial reported a higher live birth rate from frozen versus fresh ET (49.3% vs. 42.0%) (95% CI: 1.05–1.31; $P = .004$) and lower OHSS risk [RR 0.19 (95% CI: 0.10–0.37; $P < .001$)]; however, the risk of preeclampsia was also elevated in the frozen ET group (RR 3.12, 95% CI: 1.26–7.73; $P = .009$).[96] A 2017 Cochrane systematic review showed comparable cumulative live birth rates with fresh and frozen ET.[97] *In vitro* maturation of immature oocytes retrieved from antral follicles is a treatment option for PCOS patients, which eliminates the hCG trigger in a stimulated or unstimulated cycle. However, it is not widely used.

Prevention of Long-Term Complications

Since PCOS is a disease with endocrine and metabolic disorders affecting millions of women, long-term health of PCOS patients is an important issue that prompts public caution. Although the long-term CVD risk independent of BMI in women with PCOS is unclear,[39] risk of CVD requires screening, as the prevalence of CVD among PCOS patients is elevated. The risk for CVD should be evaluated for all women with PCOS, including obesity, smoking, dyslipidemia, hypertension, IGT, and lack of physical activity.[39] Regular monitoring of weight change and excess weight at least 6–12 monthly and blood pressure testing annually or more frequently should be offered to women with PCOS. Women with PCOS who are overweight or obese should get a fasting lipid profile tested.[39] Lifestyle interventions such as weight loss, exercise, and smoke cessation are suggested for PCOS patients who are overweight or obese. Studies across different countries and populations demonstrate enhanced risk of type 2 diabetes in the perimenopause and beyond for PCOS patients,[39] and women with PCOS have a higher risk of gestational diabetes.[97]

Assessment of glycemic status should be monitored every 1–3 years depending on the presence of risk factors for diabetes. As for endometrial cancer, women with PCOS have a 2- to 6-fold increased risk during the premenopausal period, and transvaginal ultrasound scan to investigate the endometrium should be conducted for women with PCOS; however, optimal prevention for endometrial cancer remains uncertain.

KEY POINTS

- The principles for treating androgen excess disorders include suppressing androgen production and activity, solving infertility, and preventing long-term complications.
- The common agent used for androgen suppression are combined oral contraceptives.
- Letrozole and clomiphene citrate are the most common agents for ovulation induction.
- The long-term health of PCOS patients is an important issue that prompts public caution.

SUMMARY AND FUTURE DIRECTIONS

Androgens are mainly produced by the ovary, adrenal gland, and via peripheral tissue conversion in women, regulated by tropic hormones and intraglandular paracrine and autocrine mechanisms. The common clinical manifestations of androgen excess comprise cutaneous abnormalities (hirsutism, acne, and patten balding), polycystic ovarian morphology on ultrasonography, and metabolic dysfunctions. Clinical hyperandrogenism is currently diagnosed by a hirsutism score determined by the mFG scale, while the diagnosis of biochemical hyperandrogenism is questionable, and measurement of free testosterone using high-quality radioimmunoassay or tandem mass spectrometry assay is recommended. PCOS is the most common cause of androgen excess disorders in reproductive age women. However, other rare causes of hyperandrogenism should always be considered in the differential diagnosis. There is still controversy regarding the etiology, pathophysiology, and targeted treatment of androgen excess disorders, especially PCOS. In the future, validated methods for the diagnosis of hyperandrogenism, well-designed studies for evaluating the best treatment for androgen excess and resultant disturbances, as well as further basic studies for underlying mechanisms of pathophysiology of hyperandrogenism are warranted.

REFERENCES

1. Miller WL, Auchus RJ. The "backdoor pathway" of androgen synthesis in human male sexual development. *PLoS Biol*. 2019;17:e3000198.
2. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32:81–151.
3. Miller WL. Disorders in the initial steps of steroid hormone synthesis. *J Steroid Biochem Mol Biol*. 2017;165:18–37.
4. Auchus RJ. Steroid 17-hydroxylase and 17,20-lyase deficiencies, genetic and pharmacologic. *J Steroid Biochem Mol Biol*. 2017;165:71–78.
5. Auchus RJ. The backdoor pathway to dihydrotestosterone. *Trends Endocrinol Metab*. 2004;15:432–438.
6. Shackleton C, Marcos J, Malunowicz EM, et al. Biochemical diagnosis of Antley–Bixler syndrome by steroid analysis. *Am J Med Genet*. 2004;128:223–231.
7. Kamrath C, Hochberg Z, Hartmann MF, et al. Increased activation of the alternative "backdoor" pathway in patients with 21-hydroxylase deficiency: evidence from urinary steroid hormone analysis. *J Clin Endocrinol Metab*. 2012;97:E367–E375.
8. Krone N, Reisch N, Idkowiak J, et al. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2012;97:E257–E267.
9. Dhayat NA, Dick B, Frey BM, et al. Androgen biosynthesis during minipuberty favors the backdoor pathway over the classic pathway: insights into enzyme activities and steroid fluxes in healthy infants during the first year of life from the urinary steroid metabolome. *J Steroid Biochem Mol Biol*. 2017;165:312–322.
10. Udhane SS, Flück CE. Regulation of human (adrenal) androgen biosynthesis—new insights from novel throughput technology studies. *Biochem Pharmacol*. 2016;102:20–33.
11. Hammond GL. Plasma steroid-binding proteins: primary gatekeepers of steroid hormone action. *J Endocrinol*. 2016;230:R13–R25.
12. Mueller NT, Liu T, Mitchel EB, et al. Sex hormone relations to histologic severity of pediatric nonalcoholic fatty liver disease. *J Clin Endocrinol Metab*. 2020;105:3496–3504.
13. Simó R, Barbosa-Desongles A, Lecube A, et al. Potential role of tumor necrosis factor-α in downregulating sex hormone–binding globulin. *Diabetes*. 2012;61:372–382.
14. Zhu JL, Chen Z, Feng WJ, et al. Sex hormone-binding globulin and polycystic ovary syndrome. *Clin Chim Acta*. 2019;499:142–148.
15. Teede HJ, Misso ML, Costello MF, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod*. 2018;33:1602–1618.
16. Pihlajamaa P, Sahu B, Janne OA. Determinants of receptor- and tissue-specific actions in androgen signaling. *Endocr Rev*. 2015;36:357–384.
17. Prizant H, Gleicher N, Sen A. Androgen actions in the ovary: balance is key. *J Endocrinol*. Sep 2014;222(3):R141–R151.
18. Franks S, Hardy K. Androgen action in the ovary. *Front Endocrinol (Lausanne)*. 2018;9:452.
19. Escobar-Morreale HF, Carmina E, Dewailly D, et al. Epidemiology, diagnosis and management of hirsutism: a consensus statement by the androgen excess and polycystic ovary syndrome society. *Hum Reprod Update*. 2012;18:146–170.
20. Rosenfield RL. Clinical practice. Hirsutism. *N Engl J Med*. 2005;353:2578–2588.
21. Ludwig E. Classification of the types of androgenetic alopecia (common baldness) occurring in the female sex. *Br J Dermatol*. 1977;97:247–254.

22. Martin KA, Chang RJ, Ehrmann DA, et al. Evaluation and treatment of hirsutism in premenopausal women: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1105–1120.
23. Dewailly D, Lujan ME, Carmina E, et al. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update*. 2014;20:334–352.
24. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*. 2004;19:41–47.
25. Zawadski JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif A, Givens JR, Haseltine F, eds. *Polycystic Ovary Syndrome*. Boston: Blackwell Scientific; 1992:377–384.
26. Ikeda K, Baba T, Noguchi H, et al. Excessive androgen exposure in female-to-male transsexual persons of reproductive age induces hyperplasia of the ovarian cortex and stroma but not polycystic ovary morphology. *Hum Reprod*. 2013;28:453–461.
27. Sanchez-Garrido MA, Tena-Sempere M. Metabolic dysfunction in polycystic ovary syndrome: pathogenic role of androgen excess and potential therapeutic strategies. *Mol Metab*. 2020;35:100937.
28. Ehrmann DA, Liljenquist DR, Kasza K, et al. Prevalence and predictors of the metabolic syndrome in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2006;91:48–53.
29. Yang R, Yang S, Li R, et al. Effects of hyperandrogenism on metabolic abnormalities in patients with polycystic ovary syndrome: a meta-analysis. *Reprod Biol Endocrinol*. 2016;14(1):67.
30. Rosenfield RL, Ehrmann DA, Littlejohn EE. Adolescent polycystic ovary syndrome due to functional ovarian hyperandrogenism persists into adulthood. *J Clin Endocrinol Metab*. 2015;100(4):1537–1543.
31. Lizneva D, Suturina L, Walker W, et al. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril*. 2016;106(1):6–15.
32. Azziz R, Carmina E, Dewailly D, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril*. 2009;91(2):456–488.
33. Legro RS, Arslanian SA, Ehrmann DA, et al. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2013;98(12):4565–4592.
34. Rosenfield RL, Ehrmann DA. The pathogenesis of polycystic ovary syndrome (PCOS): the hypothesis of PCOS as functional ovarian hyperandrogenism revisited. *Endocr Rev*. 2016;37(5):467–520.
35. Moran C, Azziz R. The role of the adrenal cortex in polycystic ovary syndrome. *Obstet Gynecol Clin North Am*. 2001;28(1):63–75.
36. Lee HS, Yoon JS, Hwang JS. Luteinizing hormone secretion during gonadotropin-releasing hormone stimulation tests in obese girls with central precocious puberty. *J Clin Res Pediatr Endocrinol*. 2016;8(4):392–398.
37. Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr Rev*. 2012;33(6):981–1030.
38. Azziz R, Carmina E, Chen Z, et al. Polycystic ovary syndrome. *Nat Rev Dis Primers*. 2016;2:16057.
39. Cooney LG, Dokras A. Beyond fertility: polycystic ovary syndrome and long-term health. *Fertil Steril*. 2018;110(5):794–809.
40. Wekker V, van Dammen L, Koning A, et al. Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. *Hum Reprod Update*. 2020;26(6):942–960.
41. Legro RS, Kunselman AR, Dunaif A. Prevalence and predictors of dyslipidemia in women with polycystic ovary syndrome. *Am J Med*. 2001;111(8):607–613.
42. Wang YX, Arvizu M, Rich-Edwards JW, et al. Menstrual cycle regularity and length across the reproductive lifespan and risk of premature mortality: prospective cohort study. *BMJ*. 2020;371:m3464.
43. Kumarendran B, O'Reilly MW, Manolopoulos KN, et al. Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: a longitudinal study based on a United Kingdom primary care database. *PLoS Med*. 2018;15(3):e1002542.
44. Dumesic DA, Oberfield SE, Stener-Victorin E, et al. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocr Rev*. 2015;36(5):487–525.
45. McAllister JM, Legro RS, Modi BP, et al. Functional genomics of PCOS: from GWAS to molecular mechanisms. *Trends Endocrinol Metab*. 2015;26(3):118–124.
46. Richards JS, Ren YA, Candelaria N, et al. Ovarian follicular theca cell recruitment, differentiation, and impact on fertility: 2017 Update. *Endocr Rev*. 2018;39(1):1–20.
47. Hsueh AJ, Kawamura K, Cheng Y, et al. Intraovarian control of early folliculogenesis. *Endocr Rev*. 2015;36(1):1–24.
48. Palomba S, Daolio J, La Sala GB. Oocyte competence in women with polycystic ovary syndrome. *Trends Endocrinol Metab*. 2017;28(3):186–198.
49. Qi L, Liu B, Chen X, et al. Single-cell transcriptomic analysis reveals mitochondrial dynamics in oocytes of patients with polycystic ovary syndrome. *Front Genet*. 2020;11:396.
50. Walters KA, Gilchrist RB, Ledger WL, et al. New perspectives on the pathogenesis of PCOS: neuroendocrine origins. *Trends Endocrinol Metab*. 2018;29(12):841–852.
51. Moghetti P, Tosi F. Insulin resistance and PCOS: chicken or egg? *J Endocrinol Invest*. 2021;44(2):233–244.
52. Gambineri A, Patton L, Vaccina A, et al. Treatment with flutamide, metformin, and their combination added to a hypocaloric diet in overweight-obese women with polycystic ovary syndrome: a randomized, 12-month, placebo-controlled study. *J Clin Endocrinol Metab*. 2006;91(10):3970–3980.
53. Goodarzi MO, Dumesic DA, Chazenbalk G, Azziz R. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol*. 2011;7(4):219–231.
54. Vink JM, Sadrzadeh S, Lambalk CB, et al. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*. 2006;91(6):2100–2104.
55. Shi Y, Zhao H, Shi Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*. 2012;44(9):1020–1025.
56. Dapas M, Lin FTJ, Nadkarni GN, et al. Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. *PLoS Med*. 2020;17(6):e1003132.
57. McAllister JM, Modi B, Miller BA, et al. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. *Proc Natl Acad Sci U S A*. 2014;111(15):E1519–E1527.
58. Stener-Victorin E, Padmanabhan V, Walters KA, et al. Animal models to understand the etiology and pathophysiology of polycystic ovary syndrome. *Endocr Rev*. 2020;41(4):538–576.
59. Tata B, Mimouni NEH, Barbotin AL, et al. Elevated prenatal anti-Müllerian hormone reprograms the fetus and induces polycystic ovary syndrome in adulthood. *Nat Med*. 2018;24(6):834–846.
60. Bahri Khomami M, Joham AE, Boyle JA, et al. The role of maternal obesity in infant outcomes in polycystic ovary syndrome-A systematic review, meta-analysis, and meta-regression. *Obes Rev*. 2019;20(6):842–858.
61. Macut D, Ilic D, Mitrovic Jovanovic A, et al. Androgen-secreting ovarian tumors. *Front Horm Res*. 2019;53:100–107.
62. Carmina E, Dewailly D, Escobar-Morreale HF, et al. Non-classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency revisited: an update with a special focus on adolescent and adult women. *Hum Reprod Update*. 2017;23(5):580–599.
63. Kaltsas GA, Androulakis II , Tziveriotis K, et al. 146 Polycystic ovaries and the polycystic ovary syndrome phenotype in women with active acromegaly. *Clin Endocrinol (Oxf)*. 2007;67(6):917–922.
64. Gotlib D, Ramaswamy R, Kurlander JE, et al. Valproic acid in women and girls of childbearing age. *Curr Psychiatry Rep*. 2017;19(9):58.
65. Speiser PW, Susin M, Sasano H, et al. Ovarian hyperthecosis in the setting of portal hypertension. *J Clin Endocrinol Metab*. 2000;85(2):873–877.
66. Satoh M, Yokoya S, Hachiya Y, et al. Two hyperandrogenic adolescent girls with congenital portosystemic shunt. *Eur J Pediatr*. 2001;160(5):307–311.
67. Hakim C, Padmanabhan V, Vyas AK. Gestational hyperandrogenism in developmental programming. *Endocrinol*. 2017;158(2):199–212.
68. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *N Engl J Med*. 2020;383(13):1248–1261.
69. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390(10108):2194–2210.

70. Speiser PW, Arlt W, Auchus RJ, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103(11):4043–4088.
71. Hannah-Shmouni F, Morissette R, Sinaii N, et al. Revisiting the prevalence of nonclassic congenital adrenal hyperplasia in US Ashkenazi Jews and Caucasians. *Genet Med*. 2017;19(11):1276–1279.
72. New MI. Extensive clinical experience: nonclassical 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2006;91(11):4205–4214.
73. Al Alawi AM, Nordenstrom A, Falhammar H. Clinical perspectives in congenital adrenal hyperplasia due to 3beta-hydroxysteroid dehydrogenase type 2 deficiency. *Endocrine*. 2019;63(3):407–421.
74. Khattab A, Haider S, Kumar A, et al. Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11beta-hydroxylase deficiency. *Proc Natl Acad Sci U S A*. 2017;114(10):E1933–E1940.
75. Papadakis GE, Dumont A, Bouligand J, et al. Non-classic cytochrome P450 oxidoreductase deficiency strongly linked with menstrual cycle disorders and female infertility as primary manifestations. *Hum Reprod*. 2020;35(4):939–949.
76. Bas S, Guran T, Atay Z, et al. Premature pubarche, hyperinsulinemia and hypothyroxinemia: novel manifestations of congenital portosystemic shunts (Abernethy malformation) in children. *Horm Res Paediatr*. 2015;83(4):282–287.
77. Rosenfield RL, Mortensen M, Wroblewski K, et al. Determination of the source of androgen excess in functionally atypical polycystic ovary syndrome by a short dexamethasone androgen-suppression test and a low-dose ACTH test. *Hum Reprod*. 2011;26(11):3138–3146.
78. Goodman NF, Cobin RH, Futterweit W, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and androgen excess and PCOS society disease state clinical review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome--part 1. *Endocr Pract*. 2015;21(11):1291–1300.
79. Gong X, Yu Y, Zhan W. Ultrasonographic findings of 1385 adrenal masses: a retrospective study of 1319 benign and 66 malignant masses. *J Ultrasound Med*. 2019;38(9):2249–2257.
80. Bidet M, Bellanné-Chantelot C, Galand-Portier MB, et al. Clinical and molecular characterization of a cohort of 161 unrelated women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency and 330 family members. *J Clin Endocrinol Metab*. 2009;94(5):1570–1578.
81. Nieman LK, Biller BM, Findling JW, et al. The diagnosis of cushing's syndrome: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93(5):1526–1540.
82. Goodman NF, Bledsoe MB, Cobin RH, et al. American Association of Clinical Endocrinologists medical guidelines for the clinical practice for the diagnosis and treatment of hyperandrogenic disorders. *Endocr Pract*. 2001;7(2):120–134.
83. Dokras A. Noncontraceptive use of oral combined hormonal contraceptives in polycystic ovary syndrome-risks versus benefits. *Fertil Steril*. 2016;106(7):1572–1579.
84. Eichenfield LF, Krakowski AC, Piggott C, et al. Evidence-based recommendations for the diagnosis and treatment of pediatric acne. *Pediatrics*. 2013;131(suppl 3):S163–S186.
85. Serfaty D. Update on the contraceptive contraindications. *J Gynecol Obstet Hum Reprod*. 2019;48(5):297–307.
86. Fraison E, Kostova E, Moran LJ, et al. Metformin versus the combined oral contraceptive pill for hirsutism, acne, and menstrual pattern in polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2020;8(8):Cd005552.
87. Franik S, Eltrop SM, Kremer JA, et al. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2018;5(5):Cd010287.
88. Ganesh A, Goswami SK, Chattopadhyay R, et al. Comparison of letrozole with continuous gonadotropins and clomiphene-gonadotropin combination for ovulation induction in 1387 PCOS women after clomiphene citrate failure: a randomized prospective clinical trial. *J Assist Reprod Genet*. 2009;26(1):19–24.
89. Abu Hashim H, Shokeir T, Badawy A. Letrozole versus combined metformin and clomiphene citrate for ovulation induction in clomiphene-resistant women with polycystic ovary syndrome: a randomized controlled trial. *Fertil Steril*. 2010;94(4):1405–1409.
90. Sharpe A, Morley LC, Tang T, et al. Metformin for ovulation induction (excluding gonadotrophins) in women with polycystic ovary syndrome. Cochrane Database. *Syst Rev*. 2019;12:CD013505.
91. Morley LC, Tang T, Yasmin E, et al. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev*. 2017;11:CD003053.
92. Weiss NS, Kostova E, Nahuis M, et al. Gonadotrophins for ovulation induction in women with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2019;1(1):Cd010290.
93. Bordewijk EM, Nahuis M, Costello MF, et al. Metformin during ovulation induction with gonadotrophins followed by timed intercourse or intrauterine insemination for subfertility associated with polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2017;1(1):Cd009090.
94. Lambalk CB, Banga FR, Huirne JA, et al. GnRH antagonist versus long agonist protocols in IVF: a systematic review and meta-analysis accounting for patient type. *Hum Reprod Update*. 2017;23(5):560–579.
95. Chen ZJ, Shi Y, Sun Y, et al. Fresh versus frozen embryos for infertility in the polycystic ovary syndrome. *N Engl J Med*. 2016;375(6):523–533.
96. Wong KM, van Wely M, Mol F, et al. Fresh versus frozen embryo transfers in assisted reproduction. *Cochrane Database Syst Rev*. 2017;3:CD011184.
97. Bahri Khomami M, Boyle JA, Tay CT, et al. Polycystic ovary syndrome and adverse pregnancy outcomes: current state of knowledge, challenges and potential implications for practice. *Clin Endocrinol (Oxf)*. 2018;88(6):761–769.

125

Endometriosis

Richard O. Burney and Linda C. Giudice

OUTLINE

Epidemiology, 2079
Pathogenesis, 2079
  Genetic Basis, 2081
  Environment, 2081
  Somatic Mutations, 2082
Pathophysiology, 2082
  Cell Survival and Adhesion, 2082
  Matrix Degradation and Invasion, 2082
  Cell Growth and Neoangiogenesis, 2083
  Altered Endocrine Activity, 2083
  Immunity and Inflammation, 2084
  Mechanisms Underlying Pain and Infertility, 2085
Diagnosis, 2085
Therapies, 2086
  Medical, 2087
  Surgical, 2088
  Integrated Approach to the Patient With Endometriosis, 2088
  Endometriosis-Associated Cancer Risk, 2088
Summary, 2089

Endometriosis is a common, chronic, estrogen-dependent, inflammatory disorder defined by the presence of endometrial glands and stroma in extrauterine ("ectopic") sites (Fig. 125.1A). It affects 6% to 10% of women of reproductive age and clinically can manifest with dysmenorrhea, dyspareunia, chronic pelvic pain, infertility, and poor pregnancy outcomes.[1] Endometriosis is most commonly found on the ovaries, pelvic peritoneum, and the bowel, and less commonly on the pleura of the lungs, the pericardium, and in previous abdominal and/or vaginal incisions. The prevalence of this condition in women experiencing pain, infertility, or both is as high as 50%.[1] Staging is based on extent of disease and associated adhesions[2] or on anatomic location (superficial peritoneal disease, ovarian endometriomas, deep infiltrating disease). No correlation exists between disease stage and symptomatology, although pain does correlate with site of disease and type of lesion, and staging does not predict chance of pregnancy or outcomes related to medical or surgical therapies.[2] The steroid hormone dependence of this disorder is striking: prolonged periods of unopposed estrogen predispose women, female primates, and men to the disorder, and natural or surgical menopause generally results in suppression.[3] Via upregulation of aromatase, endometriosis lesions can synthesize estradiol,[3] thereby bypassing the need for ovarian estradiol. Also, resistance to progesterone action, resulting in persistence of estrogen responsiveness, is now a well supported pathophysiologic feature of endometriosis.[4] An improved understanding of the molecular basis has led to new therapies and new therapeutic candidates (see later). This debilitating condition also poses quality-of-life issues for the individual patient and requires a multidisciplinary approach to achieve patient goals and well-being.[5,6] Diagnosis and treatment of endometriosis account for a significant portion of healthcare costs in the United States, estimated to approach $69 billion in 2009.[7] In this chapter, the epidemiology, pathogenesis, pathophysiology, and genetics of this disorder are reviewed, along with approaches to diagnosis and therapeutics to relieve pain and achieve fertility.

EPIDEMIOLOGY

Although typically diagnosed during the reproductive years, endometriosis is the leading cause of secondary dysmenorrhea in adolescents,[8] with a reported prevalence of endometriosis as high as 64% in this population.[9] Supporting data from the Nurses' Health Study II demonstrate that increased risks to developing endometriosis include menarche at less than or equal to 10 years of age, low birth weight (<5.5 lbs), body mass index (BMI) less than or equal to 25, nulliparity, and a family history of endometriosis.[10,11] Rates of endometriosis are greater among White women, followed by Asian, then Black, and then Latina women[12], although access to care has brought these estimates into question. Rates are nearly 3-fold higher in women with waist-to-hip ratios of less than 0.60 compared with 0.70 to 0.79 (relative risk [RR] = 2.78, 95% confidence interval [CI]: 1.38–5.60).[13] Interestingly, the rate of endometriosis is significantly lower in women who experienced vaginal parturition than in women who delivered by cesarean section or did not deliver at all, and the rate of endometriosis is inversely related to the sonographically-measured diameter of the internal cervical os.[14] This finding highlights the potential contribution of vaginal delivery in mitigating retrograde menstruation, the predominant theory in the pathogenesis of endometriosis (see later). Endometriosis has been observed in postmenopausal women, with nearly all cases occurring in women taking some form of hormone replacement therapy, further highlighting the estrogen-dependence of this condition.[15]

KEY POINTS

- Endometriosis is a chronic, estrogen-dependent, progesterone-resistant, inflammatory disorder affecting up to 10% of women of reproductive age and up to 50% of women with pelvic pain and/or infertility.

PATHOGENESIS

Though originally described by Shroen in 1690, the precise pathogenesis of this enigmatic disorder remains unknown.[16] Theories proposing

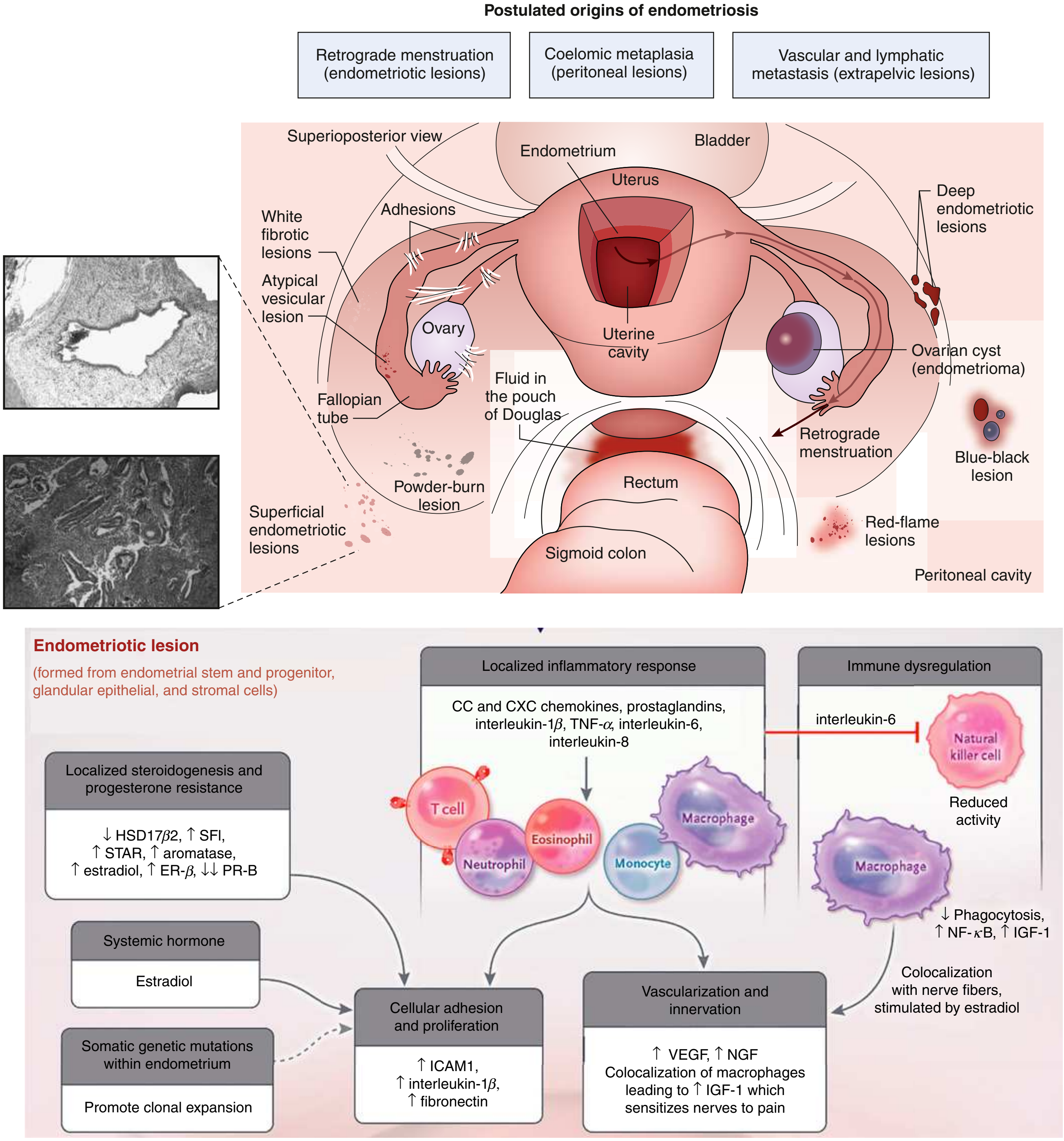

Postulated origins of endometriosis
Retrograde menstruation (endometriotic lesions)
Coelomic metaplasia (peritoneal lesions)
Vascular and lymphatic metastasis (extrapelvic lesions)
Superioposterior view
Endometrium
Bladder
Uterus
Adhesions
White fibrotic lesions
Atypical vesicular lesion
Ovary
Deep endometriotic lesions
Uterine cavity
Ovarian cyst (endometrioma)
Fluid in the pouch of Douglas
Fallopian tube
Retrograde menstruation
Blue-black lesion
Powder-burn lesion
Rectum
Superficial endometriotic lesions
Red-flame lesions
Sigmoid colon
Peritoneal cavity
Endometriotic lesion
(formed from endometrial stem and progenitor, glandular epithelial, and stromal cells)
Localized inflammatory response
CC and CXC chemokines, prostaglandins, interleukin-1β, TNF-α, interleukin-6, interleukin-8
Immune dysregulation
interleukin-6
Natural killer cell
Reduced activity
Localized steroidogenesis and progesterone resistance
↓ HSD17β2, ↑ SFI, ↑ STAR, ↑ aromatase, ↑ estradiol, ↑ ER-β, ↓↓ PR-B
T cell
Eosinophil
Neutrophil
Monocyte
Macrophage
Macrophage
↓ Phagocytosis, ↑ NF-κB, ↑ IGF-1
Colocalization with nerve fibers, stimulated by estradiol
Systemic hormone
Estradiol
Cellular adhesion and proliferation
↑ ICAM1, ↑ interleukin-1β, ↑ fibronectin
Vascularization and innervation
↑ VEGF, ↑ NGF Colocalization of macrophages leading to ↑ IGF-1 which sensitizes nerves to pain
Somatic genetic mutations within endometrium
Promote clonal expansion

Figure 125.1 **Pathogenesis and pathophysiology of endometriosis.** Inset, 10% hematoxylin and eosin stain of a peritoneal lesion of endometriosis and late proliferative endometrium taken from the same patient at the time of surgery. Note the presence of glands and stroma surrounded by fibrosis. (Adapted from Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med.* 2020;382:1244–1256, used with permission.)

a nonendometrial origin for lesions are supported by histologically-confirmed endometriosis in patients with Müllerian agenesis and in men with prostate cancer undergoing high-dose estrogen treatment.[17-19] On the other hand, the theory of benign metastasis holds that ectopic endometrial lesions result from lymphatic or hematogenous dissemination of endometrial cells.[20-22] Microvascular studies demonstrate lymph flow from the uterus into the ovary, supporting this hypothesis.[23] Endometriosis within lymph nodes has been documented in 6.5% of women at lymphadenectomy and 6.7% of women at autopsy.[24] The strongest evidence for the theory derives from reports of histologically proven endometriotic lesions occurring in sites distant from the uterus or pelvis, including bone, brain, and lung.[25]

The most widely accepted description of endometriosis pathogenesis is Sampson's theory of retrograde menstruation and transplantation (Fig. 125.1A), which postulates that sloughed endometrial tissue, displaced in a retrograde manner through the fallopian tubes during menses, contains viable, steroid-sensitive cells that are capable of implanting outside the uterus.[26] Although this hypothesis is attractive, it does not explain why retrograde menstruation, a physiologic event,[27] does not result in endometriosis in all women. The volume of displaced menstrual debris has been considered to be relevant to establishment of the disorder, as women with endometriosis have a higher volume of refluxed menstrual blood than those without the disorder.[27] In addition, the occurrence during the menstrual phase of subendometrial myometrial contractile waves exhibiting a retrograde pattern in women with endometriosis as opposed to an antegrade pattern in women without disease further supports Sampson's theory and suggests that endometriosis may have some of its origin in myometrial dysfunction.[28] Endometriosis can be induced in baboons by cervical occlusion[29] and is uniformly observed in adolescent girls with primary amenorrhea caused by cervical hypoplasia or aplasia, vaginal agenesis, or other Müllerian anomalies associated with outflow tract obstruction.[30] Even subtle compromise of antegrade menstruation may predispose to endometriosis, as evidenced by a higher prevalence of disease in women with a uterine septum.[31]

The inability of any single theory to account for the variety of clinical presentations and pathological features of endometriosis may derive from its clinical heterogeneity, with peritoneal endometriosis, ovarian endometriosis, and deep infiltrating endometriosis (DIE) possibly developing via different mechanisms.[32] Indeed, epithelial cells isolated by laser capture microscopy from eutopic endometrium and ectopic endometrium (peritoneal and ovarian lesions) and subsequent analysis using cDNA microarrays with 9600 genes revealed that samples cluster by site (ovarian endometrioma vs. nonovarian disease), underscoring that ovarian and peritoneal endometriosis have different complements of gene expression and may represent different disorders.[33] Whereas peritoneal and ovarian endometriosis are plausibly accounted for by retrograde menstruation, DIE is less so, as it is typically located in deeper subperitoneal locations as the intestine muscularis or bladder. The endometrial stem cell theory of endometriosis[34] aligns with the widely accepted retrograde menstruation model, while also providing an explanation for the pathogenesis of the DIE lesional subtype. Next-generation sequencing studies of passenger mutations within the epithelial and stromal cells of DIE lesions support clonality of the epithelial compartment, likely reflecting an epithelial stem cell origin.[35] Though it is possible that these stem cells derive from the bone marrow, somatic mutations in endometriosis-associated genes have been observed in the epithelial compartment of normal endometrium but not the bone marrow, consistent with an endometrial origin for these progenitors.[36]

Genetic Basis

A genetic predisposition to the development of endometriosis is well supported. Monozygotic twin studies observed high concordance rates for histologically confirmed endometriosis and disease stage.[37] An increased risk of endometriosis is seen among first-degree relatives of affected women.[38] Abundant genetic evidence supports endometriosis as a multifactorial, polygenic disease, with the proportion of disease variance due to genetic factors estimated at 26%.[39] Additional support for polygenic inheritance is earlier age of onset for familial versus nonfamilial cases and similar age of onset among affected relatives.[40] In addition, a strong familial tendency prevails in nonhuman primates.[41]

Genome-wide association studies (GWAS) provide high-resolution haplotypes of single nucleotide polymorphism (SNPs) in cohorts of women with and without endometriosis toward the identification of genes associated with endometriosis risk. A metaanalysis of ten GWAS datasets involving 17,045 cases and 191,596 controls identified directional association of SNPs in 14 genetic risk loci across populations of European and Japanese ancestry.[42] Though positional evidence of genes near risk loci implicates involvement of hormonal pathways, WNT signaling, cell adhesion, and inflammation-related pathways, functional involvement of these variants in mediating disease pathogenesis awaits multiomic integration. Recently, a large-sale GWAS metaanalysis revealed 27 genetic loci associated with endometriosis ($P < 5\times10^{-8}$), of which 78% had greater effect sizes in stage III/VI versus stage I/II disease.[43] Additional clinical disorders significantly positively correlated with endometriosis included uterine fibroids, excessive and irregular menstrual bleeding, osteoarthritis, diabetes, menstrual cycle length and age at menarche.

Environment

Developmental Origins of Endometriosis Risk. The pre- and perinatal periods are susceptible windows during which exposure to xenoestrogens and other endocrine-disrupting chemicals (EDCs)[44] can induce developmental programming and increase risk for female reproductive tract disorders, including endometriosis.[45] As endometriosis is an estrogen-dependent disorder, it is biologically plausible that EDCs mimicking estrogen actions may reprogram uterine endometrium for endometriosis development later in life and that EDCs that can alter immune function can predispose to subsequent disease. Uterine morphogenesis is a genetically controlled process, conserved across species, and involves ERα and *Wnt* and *Hox* gene family members.[46] Animal studies show *in utero* exposures to some estrogen-like EDCs (e.g., diethystilbestrol [DES] and bisphenol A [BPA]) result in abnormal female reproductive tract development[47] and endometriosis-like tissue in adult female offspring.[48] *In utero* exposure of female mice to tetrahydrochlorodibenzo-p-dioxin (TCDD) augments growth of surgically-induced endometriosis in response to later TCDD exposure.[49] Moreover, adult female mice directly exposed *in utero* (F1, F2 generation) or indirectly exposed (F3 generation, transgenerational inheritance) to TCDD display uterine progesterone receptor (PR) promoter hypermethylation, decreased fecundity, and preterm birth versus unexposed controls,[50] similar to the uterine phenotype and clinical outcomes of women with endometriosis.

In women, the most compelling evidence of development in adulthood after *in utero* exposure to DES derives from the Nurses' Health Study II.[11] This prospective study of 116,678 female nurses 25 to 42 years old revealed 5% endometriosis prevalence at baseline and an incidence of 2941 laparoscopically confirmed cases, with pain symptoms prompting diagnosis in 77% of subjects and an infertility work-up prompting diagnosis in the remaining 33%. A history of *in utero* exposure to DES increased by 80% a woman's chance of having laparoscopically proven endometriosis as an adult. Other risk factors were low birth weight and preterm birth, suggesting developmental effects, which are supported by other studies demonstrating Müllerianosis and endometriosis tissue in human female fetuses.[51] The relationship of these with adult disease has not (yet) been established.

Exposures in Adulthood and Endometriosis Risk. Exposures to EDCs during adulthood, with or without prior exposure, may also predispose to developing or promoting endometriosis and result in abnormal PR signaling, characteristic of endometrium and lesions of women with endometriosis (see later).[45] There is abundant evidence in adult nonhuman primates and rodents that organochlorines, including TCDD, pesticides such as methoxychlor and DDT, and polychlorinated biphenyls promote endometriosis.[51] For example, adult rhesus monkeys exposed to TCDD at 5 to 25 ppt daily for 4 years developed symptomatic endometriosis with dose-dependent severity.[52] In mouse models of surgically induced disease (autologous uterine tissue transplantation to peritoneal structures), TCDD and other polyhalogenated hydrocarbons increased lesion sizes in a dose-dependent manner relative to vehicle controls.[53] Furthermore, in a model wherein endometriosis-like lesions in adult female nude mice were induced by seeding human endometrial tissue fragments onto the peritoneum, TCDD resulted in loss of uterine PR and prevented progesterone-mediated suppression of disease establishment seen in controls.[54] TCDD blocked progesterone inhibition of matrix metalloproteinase (MMP) expression in endometrial stromal cells *in vitro*, similar to the human endometrial and lesion phenotype[55] (see later). Interestingly, the EDC BPA promoted cell proliferation of surgical endometriosis in a mouse model,[56] and di(2-ethylhexyl)phthalate (DEHP) increased surgically-induced endometriosis invasion and growth in adult female mice, accompanied by upregulation of MMP-2, MMP-9, Ki67, and Pax4 and decreased cleaved capsase.[57] Overall, the data strongly support EDCs increasing disease burden and severity in animal models.

With regard to women, four recent systematic reviews and meta-analyses of epidemiologic studies demonstrated associations of persistent and nonpersistent EDCs and endometriosis risk. Cano-Sancho et al.[58] analyzed 17 studies that provided effect sizes and metrics sufficient for pooling estimates through metaanalysis and reported: for dioxins: odds ratio (OR) 1.65 (95% CI: 1.14–2.39); polychlorinated biphenyls (PCBs): OR 1.70 (95% CI: 1.20–2.39); and organochlorine pesticides (OCPs): OR 1.23 (95% CI: 1.13–1.36). Wen et al.[59] analyzed 30 epidemiologic studies and reported: for PCBs: OR 1.58 (95% CI: 1.18–2.21); OCPs: OR 1.40 (95% CI: 1.02–1.92); and DEHP: OR 1.42 (95% CI: 1.19–1.70). Overall OR across all exposures (including phthalate esters as a class) was 1.41 (95% CI: 1.23–1.60). They found no significant associations with BPA, consistent with data from Smarr et al.,[51] and they did not report on dioxins. Sirohi et al.[60] reported on 29 studies and found that phthalate esters were positively associated with endometriosis prevalence, and the majority of studies (71%) revealed significant association between BPA, organochlorinated EDCs (dioxins, dioxin-like compounds, OCPs, PCBs) and endometriosis prevalence. Notably, they found one study with positive association of two heavy metals (copper and chromium), but not cadmium, lead, or mercury, with endometriosis prevalence. Overall, despite statistical significance, the data should be interpreted with caution because of heterogeneity of disease phenotypes (stage, disease type), small effect sizes, differences in study designs, varied definitions of endometriosis, confounders of age, BMI, parity, lactation history, exposure to mixtures versus one chemical at a time, varied analytical methodologies, and definition of controls.[51] That said, epidemiologic data support associations of EDCs with endometriosis, and experimental data from animal studies support such associations and underlying mechanisms. Currently, the precautionary principle is to advise minimizing EDC exposures in all women with and without endometriosis and during pregnancy.

Somatic Mutations

Increased somatic mutations, particularly cancer driver mutations, are commonly seen in DIE and ovarian endometriosis.[61,62] Somatically mutated genes in endometriotic lesions include *KRAS* (oncogene), *ARID1A* (tumor suppressor gene), *PIK3CA* (oncogene), and *PPP2R1A* (tumor suppressor gene). These genes are also frequently mutated in uterine endometrioid carcinoma, ovarian endometrioid carcinoma, and ovarian clear cell carncioma.[63] Not only are these somatic mutations a potential factor in the multistep process of malignant transformation, but they may also contribute to disease pathogenesis. Interestingly, the observed somatic mutations appear to be confined to glandular epithelial cells of endometriotic lesions.[61] In a classic series of experiments involving a murine model of endometriosis, *KRAS* activation in peritoneally transferred endometrial cells resulted in endometriotic lesions of the peritoneal lining and ovaries.[64]

KEY POINTS

- Genetics (including somatic mutations), retrograde menstruation, immune dysfunction in pelvic debris clearance, and epigenetically driven endocrine dysregulation are likely components of endometriosis pathogenesis.

PATHOPHYSIOLOGY

For tissue to implant and thrive in an ectopic location, cell(s) must survive after release from their originating niche, attach and invade into a new microenvironment, proliferate, and establish a new blood supply. Escape from immune clearance, attachment to peritoneal epithelium, invasion of the epithelium, establishment of local neurovascularity, and continued growth and survival are necessary if endometriosis is to develop from retrograde displacement of endometrium. As with normal endometrium, endometriosis tissue responds to cyclic changes in steroid hormones by proliferation and glandular secretion and by production of autocrine and paracrine factors that affect these processes[65] (Fig. 125.1A and B).

Cell Survival and Adhesion

Menstrual blood is found in peritoneal fluid (PF) of up to 90% of healthy women with patent fallopian tube(s) undergoing laparoscopy during menses.[27] The discrepancy between the nearly 90% rate of retrograde menstruation and the 6% to 10% prevalence of endometriosis remains an area of active study. Hereditary or acquired properties of the endometrium and/or the peritoneal epithelium, and/or defective immune clearance have all been proposed as factors influencing survival and implantation of ectopic endometrium. In the formation of peritoneal lesions, retrograde menstrual tissue must execute a molecular strategy to adhere to serosal surfaces. Confocal microscopy studies using an *in vitro* model of endometriosis observed the adherence and invasion of endometrial cells through peritoneal mesothelium.[66] Eutopic and ectopic endometrium from women with endometriosis evidence upregulation of the antiapoptotic gene *BCL-2*.[67] In addition to decreased apoptosis, enhanced proliferation may confer a selective survival advantage to the endometrium of women predisposed to endometriosis.[68] Importantly, peritoneal lesions demonstrate increased expression of estrogen receptor beta (ERβ),[3] shown in a mouse model to promote lesion development by inhibiting tumor necrosis factor alpha (TNF-α)-mediated apoptosis and increasing cellular adhesion and proliferation,[69] exemplifying a key molecular hallmark in endometriosis: estrogen dependence (Fig. 125.1B). Altered expression of integrins by endometrial stromal cells from women with endometriosis contribute to their increased capacity to adhere.[70] It is possible that proteases, growth factors, altered cytokine profiles, or a localized inflammatory microenvironment in women with endometriosis set conditions for the selective adhesion of endometrial cells to serosal surfaces.[71]

Matrix Degradation And Invasion

Endometriosis, a noncancerous disorder, can invade tissues and surfaces in a manner similar to that exhibited by malignancies. The balance between MMPs and their inhibitors (tissue inhibitors of metalloproteinases) is critical in maintaining the appropriate level of MMP activity, and failure to maintain this balance may contribute to matrix breakdown and cellular invasion, as observed in endometriosis. In menstrual phase eutopic endometrium, mRNAs for all the MMPs have been detected, with matrilysin (MMP-7) localized to the epithelium, and all other MMPs expressed in the stroma.[72] Epithelial-specific MMP-7 is expressed in endometriotic lesions during the secretory phase, but not in eutopic endometrium, which suggests that progesterone suppression of MMP-7 is dysregulated in endometriosis, and that this MMP may play a role in growth or progression of the disorder. Foci of MMP-1 expression closely correlate with matrix breakdown and with the absence of PR in adjacent epithelial cells.[73] Indeed, suppression of PR-B in eutopic and ectopic endometrium, possibly due to differential methylation of PR-B, HOX, and GATA transcription factors, may contribute to altered progesterone responsiveness.[74] *In vitro* treatment of eutopic endometrium acquired from affected women with progesterone failed to fully suppress pro–MMP-7 secretion and prevent the implantation of transplanted endometrium to establish experimental disease in mice.[75] Collectively, these results illustrate another molecular hallmark—progesterone resistance—in the pathophysiology of endometriosis (Fig. 125.B).

Cell Growth and Neoangiogenesis

Peritoneal fluid from women with endometriosis reflects a unique microenvironment characterized by increased mitogenic, angiogenic, morphogenic, and chemoattractant activities that are collectively conducive to lesional growth and survival. Growth factors and cytokines are released from a variety of cellular sources, including monocytes/macrophages, as well as endometrial, ovarian, and mesothelial cells within the peritoneal cavity.[76] Growth factors likely play a fundamental role in the pathophysiology of endometriosis by independently stimulating cell survival, growth, or differentiation. Macrophage-conditioned medium contains mitogenic activity, some of which is attributable to TGF-β,[77] platelet-derived growth factor (PDGF),[78] and basic fibroblast growth factor.[79] Epidermal growth factor,[79] insulin-like growth factors,[80] and PDGF[78] are potent mitogens for endometrial stromal cells, and hepatocyte growth factor (HGF) is a mitogen and morphogen for endometrial epithelial cells when cocultured with stromal cells.[81] HGF is also an angiogenic factor. Establishment of vascular supply is a necessary step in the growth and survival of endometriotic lesions, particularly in the peritoneal microenvironment, which is relatively avascular compared with the intrauterine environment. In fact, increased vascularity in the vicinity of peritoneal lesions is leveraged in the surgical identification of lesions at laparoscopy.[82] Angiogenic activity is increased in PF from women with endometriosis. Vascular endothelial growth factor (VEGF), a potent angiogenic factor, has been consistently detected in high concentrations in PF from women with endometriosis, and VEGF levels in the PF correlate directly with the stage of disease[83] (Fig. 125.1B). The consistent detection of VEGF expression across the spectrum of lesions provides compelling evidence for VEGF as the prominent angiogenic factor in endometriosis pathophysiology and a potential target for therapeutic strategies.[84]

Altered Endocrine Activity

The survival and growth of ectopically displaced endometrium is estrogen-dependent and regulated by estradiol released systemically from the ovary and locally from lesions themselves. Lesional production of estradiol is favored by the imbalance resulting from the increased expression of aromatase[85] and steroidogenic acute regulatory protein and decreased expression of 17β-hydroxysteroid dehydrogenase 2[86] in ectopic endometrial stromal cells (ESCs) (Fig. 125.2). Interestingly, increased expression of the transcription factors GATA6 and SF1 in ectopic ESCs results in in an endometriotic phenotype characterized by estradiol synthesis and impaired decidualization.[87] The observed increase in GATA6 and SF1 in these cells is in part a function of promoter hypomethylation.[74] ESCs from endometriotic lesions evidence increased levels of ERβ.[88] By increasing IL-1β, ERβ is a key mediator in the link between the molecular hallmarks of estrogen dependence and inflammation in the pathophysiology of endometriosis. The ERβ pathway inhibits TNF-α–induced cellular apoptosis, enhances cellular invasion, and downregulates PR expression, and suppression of ERβ activity via administration of a selective antagonist inhibited the growth of endometriosis in a mouse model.[69] This has important consequences with regard to therapies (see section on treatment) that are aimed primarily at decreasing circulating estradiol as a treatment strategy.

Support for progesterone resistance as a molecular hallmark in endometriosis is derived from results of global transcriptome analysis of eutopic endometrium in the late proliferative and early and mid-secretory phases in women with moderate to severe endometriosis compared with women without disease.[4] Of note was the marked persistence of estrogen regulated genes in early secretory endometrium (ESE) from women with disease at a time when progesterone action should inhibit their expression. In the proliferative-secretory transition, the fingerprint of persistence of cellular mitosis and minimal responsiveness of classically progesterone-regulated genes in ESE was also apparent in midsecretory endometrium. ESE specimens from women with disease clustered closer to proliferative endometrium specimens than ESE from women without disease, confirming the

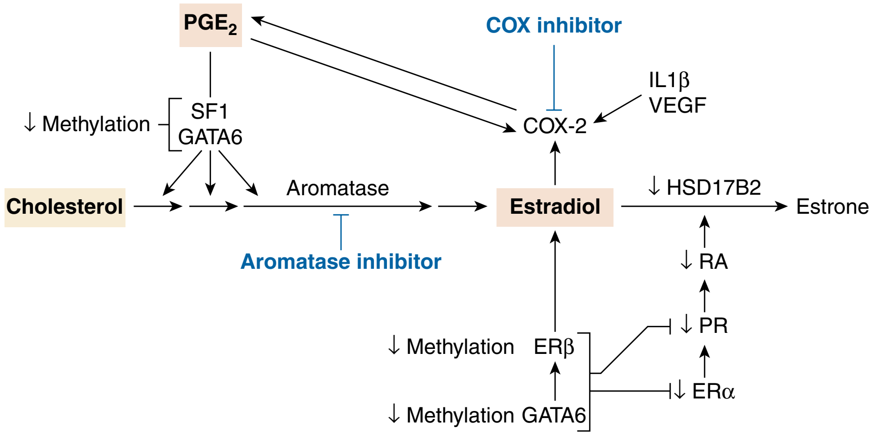


Figure 125.2 Summary of aromatase and other enzymes and the concept of local estradiol biosynthesis and decreased metabolism in endometriotic lesions and the centrality of estrogen in the pathogenesis of endometriosis. (Adapted from Bulun SE, Yilmaz BD, Sison C, et al. Endometriosis. *Endocr Rev.* 2019;40:1048–1079.)

observed resistance to progesterone action. These data underscore progesterone resistance in the endometrium of women with endometriosis and are consistent with inadequate clinical response to progestins in women with this disorder.[4]

MicroRNAs (miRNAs) are key regulators of gene expression through mRNA degradation or repression. Numerous miRNAs are expressed in the endometrium and serum of women with versus without endometriosis that predict target genes, including *TGF-βR2*, *ERα*, *ERβ*, and *PR*.[89] Steroid hormone treatment of human endometrial cells *in vitro* revealed differentially regulated expression of hsa-miR20a, hsa-miR21, and hsa-miR26a, which were partly reversed following cotreatment with ICI-182780 and RU486.[90] Differential expression of a number of miRNAs in eutopic and ectopic endometrium and isolated endometrial cells suggests that aberrant or altered expression of some miRNAs regulated by ovarian steroids may influence the expression of specific target genes with key roles in endometrial cellular activities and in the pathophysiology of endometriosis. Circulating miRNAs have recently been reported to diagnose endometriosis with high accuracy and specificity.[91]

Immunity and Inflammation

Abundant evidence indicates that endometriosis is a chronic inflammatory condition with associated immune dysfunction, and central to endometriosis pathophysiology are enhanced estrogen signaling, disrupted progesterone signaling, and inflammation in the disease tissue and eutopic endometrium[3,92-94] (Fig. 125.3). These processes are interrelated and largely contribute to the pain and infertility in affected women. Recent mass cytometry analysis of PF demonstrated a heterogeneous and complex inflammatory microenvironment with specific endometriosis-associated immune signatures within the innate and adaptive immune lineages, distinct phenotypes marked by CD69 versus controls without disease, and different signatures across disease stages.[95] There are increased numbers of PF activated macrophages involving c-Jun signaling,[96] proinflammatory reactive oxygen species

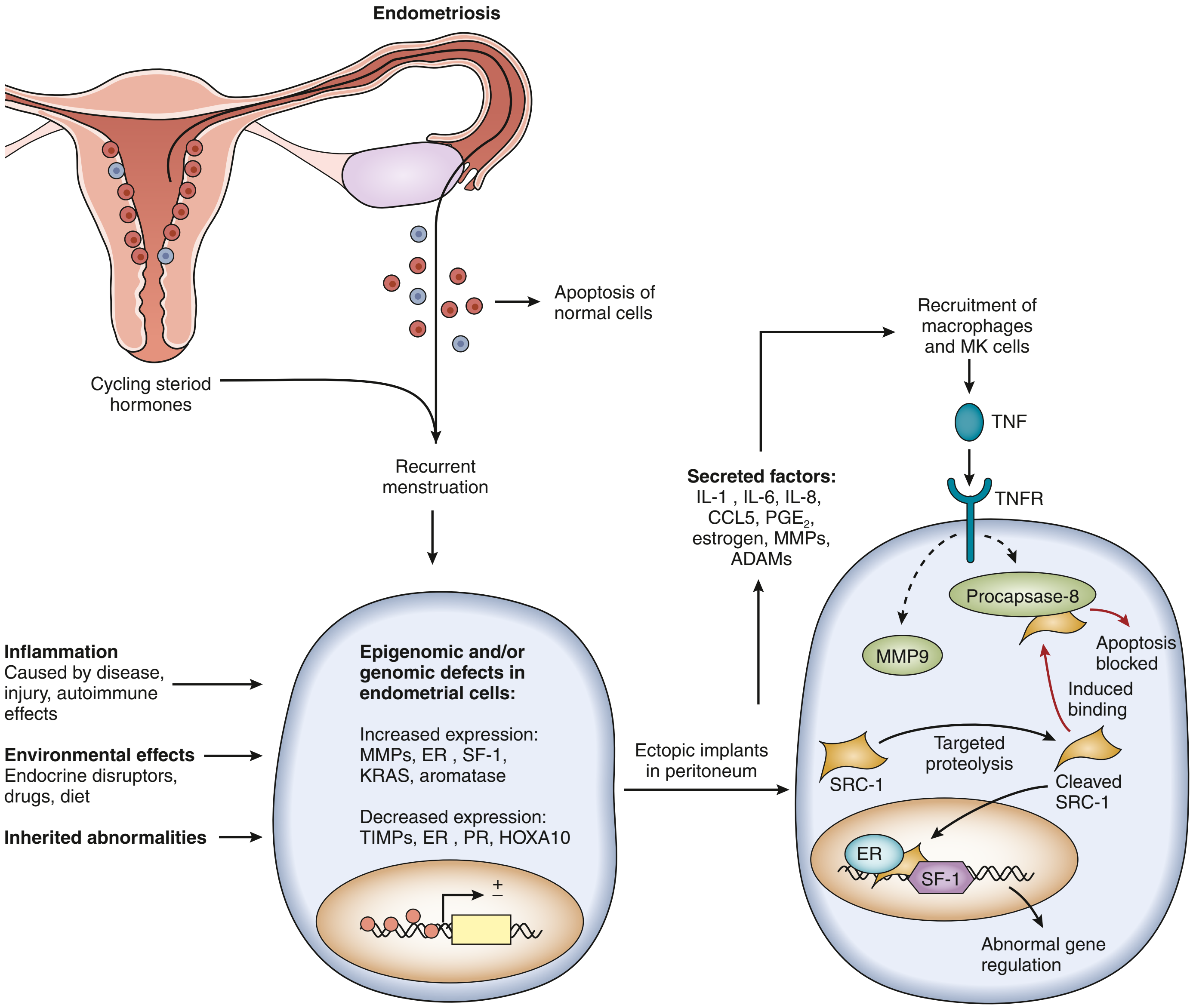

Endometriosis
Apoptosis of normal cells
Cycling steriod hormones
Recurrent menstruation
Recruitment of macrophages and MK cells
TNF
TNFR
Secreted factors:
IL-1 , IL-6, IL-8, CCL5, PGE2, estrogen, MMPs, ADAMs
Procapsase-8
MMP9
Apoptosis blocked
Induced binding
Targeted proteolysis
SRC-1
Cleaved SRC-1
ER
SF-1
Abnormal gene regulation
Ectopic implants in peritoneum
Inflammation
Caused by disease, injury, autoimmune effects
Environmental effects
Endocrine disruptors, drugs, diet
Inherited abnormalities
Epigenomic and/or genomic defects in endometrial cells:
Increased expression: MMPs, ER , SF-1, KRAS, aromatase
Decreased expression: TIMPs, ER , PR, HOXA10
+
–

Figure 125.3 Molecular aberrations resulting in enhanced estrogen signaling, abrogated progesterone signaling, and inflammation in endometriosis. (Adapted from Dyson MT, Bulun SE. Cutting SRC-1 down to size in endometriosis. *Nature Med.* 2012;18:1016–1018.)

(ROS) generated from the breakdown of blood, and increased cytokines and chemokines, including IL-1β, IL-6, IL-8, TNF-α, monocyte chemotactic protein, and regulated on activation normal T expressed and secreted.[92] Interestingly, higher plasma IL-1β levels are observed years before the diagnosis of endometriosis.[97] IL-1β activation of cyclo-oxygenase-2 increases production of prostaglandin-E2, which in turn activates steroidogenic acute regulatory protein and aromatase to increase local estradiol production.[3] This pathway highlights the interplay of estrogen dependence and inflammation in endometriosis. The NF-κB pathway plays a central role in endometriosis-associated inflammation[92] and is activated in stromal fibroblasts by ROS resulting from iron overload, possibly due to impaired clearance of retrograde menstruation, lesion-associated bleeding,[98] or endometrioma cyst fluid accumulation. In a rat model of endometriosis, a candidate NF-κB inhibitor (disulfiram) prevented the growth of endometriosis lesions.[99]

In addition to inflammation, a growing body of evidence supports relative immune dysfunction in women with endometriosis. For instance, natural killer cell activity is impaired in women with endometriosis, which may set conditions for immune evasion of retrogradely menstruated endometrial cells.[100] A recent systematic review and metaanalysis revealed an association between endometriosis and autoimmune disorders, including thyroid disease.[101] Whether it is an autoimmune disorder with accompanying increased generalized B-cell activity and association with autoantibodies to phospholipids and endometrium is yet to be determined. Increased antibody production may be the consequence of a general inflammatory response initiated by activated macrophages and other antigen-presenting cells in response to the ectopic endometrial tissue, rather than a cause of the disease. Alterations in immune function have been described in endometriosis,[92] but the changes are complex, and none independently or satisfactorily explains the pathogenesis of this disorder. In addition, it is not clear whether the alteration is the cause of endometriosis or secondary to the inflammatory reaction that endometriosis induces.

Mechanisms Underlying Pain and Infertility

Pelvic pain associated with endometriosis may be manifest as dysmenorrhea, dyspareunia, dyschezia, and/or chronic, noncyclic pelvic pain.[102] Pain symptoms vary with the location of the disease, stage, type of lesion, and depth of invasion. The pathophysiology of endometriosis-associated pain is believed consequent to inflammation, adhesions, lesional innervation, and increased prostaglandin production by ectopic implants.[103] Red vesicular lesions, for example, produce higher amounts of prostaglandins.[104] Adhesions may cause pain by damage to tissues and subsequent scarring or damage to nerves and/or blood supply resulting in ischemia or tissue hypoxia. Understanding the etiology of endometriosis-associated pain is prerequisite to the design of novel therapies (see treatment section later). Important contributions include a murine model of innervation of endometriosis lesions[105] and the demonstration of nerve fibers in ectopic and eutopic endometrium (functionalis and basalis) from women with endometriosis (Fig. 125.4).[106,107] Although transcriptomic analysis of eutopic endometrium had revealed neurotransmitter receptors and ligands using a variety of approaches,[108] demonstration of unmyelinated nerves (nociceptors) in eutopic and ectopic endometrium suggests mechanisms involved in the pathogenesis of pain. Finally, ascending nociceptive signals in the central nervous system may result in central sensitization and alterations in pain processing, brain chemistry, and function that are unique in women with endometriosis and may explain pain that is refractory to medical and surgical treatments.[105,109]

Endometriosis is strongly associated with infertility. A metaanalysis of 22 studies revealed a significant reduction in *in vitro* fertilization (IVF) pregnancy rates in women with endometriosis relative to women with tubal factor infertility.[110] A recent metaanalysis of 27 studies comparing IVF outcomes in women with and without endometriosis confirmed a decrease in clinical pregnancy rates, but only in women with advanced (stage III–IV) disease (RR 0.79, 95% CI: 0.69–0.91).[111] Endometriosis-associated infertility may be due to mechanical factors affecting gamete/embryo transit, sperm/oocyte/embryo toxicity, endocrine disturbances, immunologic abnormalities, and/or implantation defects. Notably, the proinflammatory PF from women with endometriosis adversely affects sperm motility, sperm velocity,[112] and sperm binding to the zona pellucida in a hemizona assay[113] and significantly reduces mouse embryo blastocyst development rates, suggesting a direct embryotoxic effect.[114] There is an abundance of data demonstrating eutopic endometrial abnormalities (progesterone resistance, inflammatory phenotype) leading to an inhospitable environment for embryo implantation (attachment and stromal decidualization).[115] Dysregulated genes in women with disease include $\alpha_v\beta_3$ integrin, *HOXA10*, *MUC1*, glycodelin, and *BCL6*. Recently, Lessey and colleagues demonstrated that inflammatory cytokines promote *BCL6* expression, which correlates with progesterone resistance, and high *BCL6* expression in the implantation window is associated with endometriosis, unexplained infertility, and implantation failure.[116] These observations and others are opening the field to novel diagnostic and therapeutic development.

DIAGNOSIS

Endometriosis may be suspected but cannot be reliably diagnosed on the basis of presenting symptoms or physical exam, although recently, clinical algorithms to predict disease based on these parameters and family history have been developed.[102] Rather, the gold standard for diagnosing endometriosis is histologic confirmation of endometrial glands and stroma in laparoscopically directed biopsies of suspected lesions.[117] Peritoneal endometriosis is heterogeneous in appearance, with a variety of lesional subtypes including red vesicular, powder burn, or white fibrotic lesions. Recognition of endometriosis varies among surgeons, and inaccuracy has been documented in visual diagnosis.[118] Confounding presentations with other disorders and the requirement for a surgical diagnosis result in an average latency of 8 to 11 years from symptom onset to definitive diagnosis.[119] Unfortunately, disease may progress during this interval, as longitudinal placebo-controlled trials with second-look laparoscopy demonstrate that 71% to 83% of untreated lesions will progress or remain stable over a 12-month period.[120,121]

Though unreliable for the diagnosis of peritoneal endometriosis and associated adhesions, imaging techniques such as ultrasonography and magnetic resonance imaging (MRI) perform well in the detection of ovarian endometriomas, with 80% to 90% sensitivity and 60% to 98% specificity.[122] Also known as a “chocolate cyst,” an endometrioma is sonographically characterized by a hazy appearance on ultrasound and histologically by the presence of endometrial glands and stroma in the cyst wall. MRI is increasingly being used to evaluate deep infiltrating disease, with high accuracy (74% sensitivity and 95% specificity), and to plan the surgical approach.[123]

Noninvasive biomarker discovery for diagnosing endometriosis is an active area of study. Two recent systematic reviews of all proposed endometriosis-related biomarkers (proteins, cytokines, angiogenic and growth factors, miRNAs) over the previous 25 years in serum, plasma, urine, and endometrium were unable to identify unequivocally clinically useful biomarkers, due to low numbers of subjects, small populations for

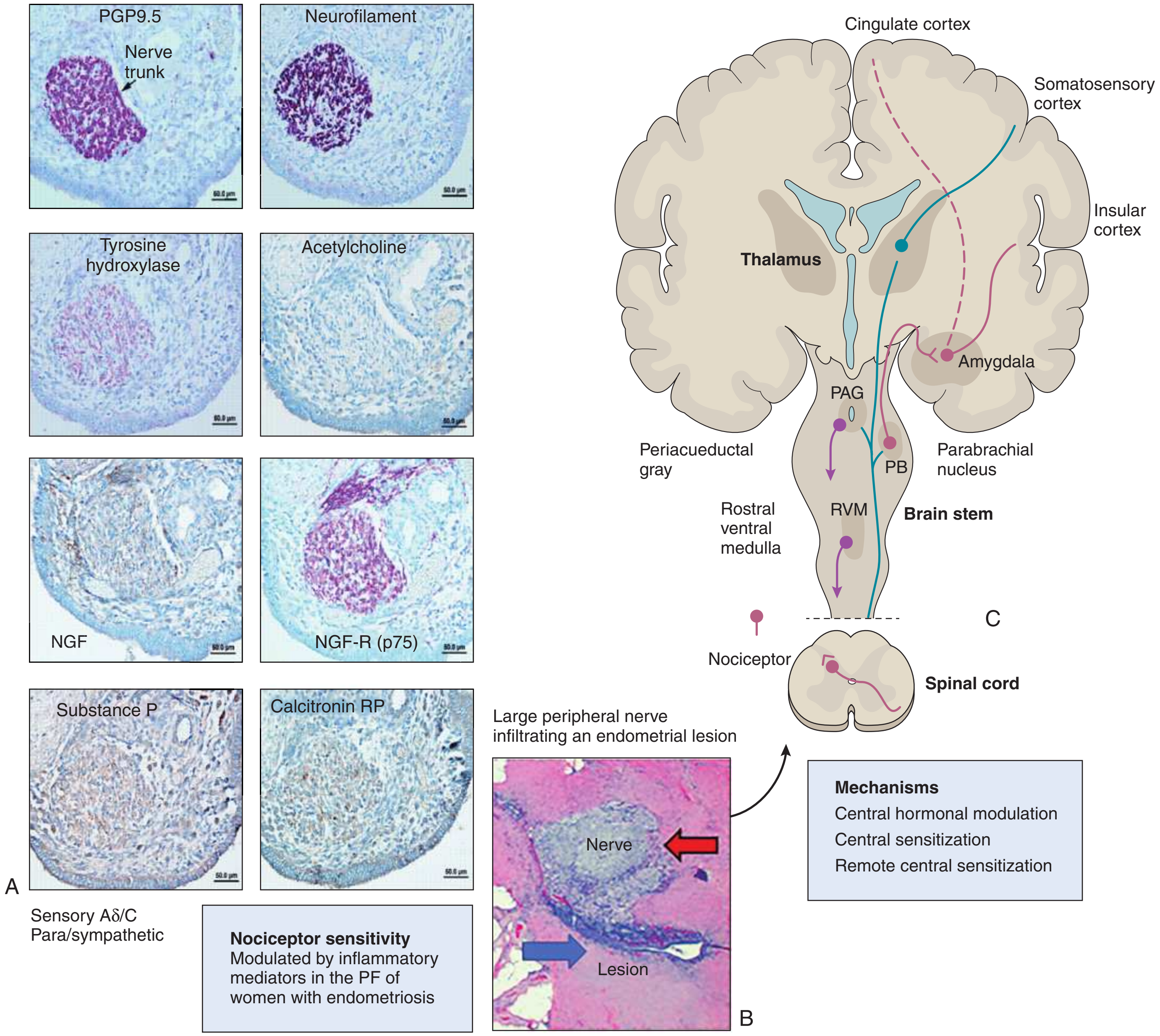

PGP9.5
Nerve trunk
Neurofilament
Tyrosine hydroxylase
Acetylcholine
NGF
NGF-R (p75)
Substance P
Calcitronin RP
A
Sensory Aδ/C
Para/sympathetic
Nociceptor sensitivity
Modulated by inflammatory mediators in the PF of women with endometriosis
Large peripheral nerve infiltrating an endometrial lesion
Nerve
Lesion
B
Cingulate cortex
Somatosensory cortex
Insular cortex
Thalamus
Amygdala
PAG
Periacueductal gray
PB
Parabrachial nucleus
Rostral ventral medulla
RVM
Brain stem
C
Nociceptor
Spinal cord
Mechanisms
Central hormonal modulation
Central sensitization
Remote central sensitization

Figure 125.4 Nerve fibers in endometriosis and pain pathways in endometriosis. **A,** Nerve fibers in endometriosis peritoneal lesions stain positively for protein gene product 9.5 (PGP9.5), neurofilament (NF), tyrosine hydroxylase (TH), nerve growth factor (NGF), NGF receptor p75 (NGF-Rp75), substance P (SP), and calcitonin gene-related peptide (CGRP). **B–C,** Nerve fibers (protein gene product 9 [PGP9] immunoreactivity) in eutopic endometrium of women with versus without endometriosis. (**A**, From Tokushige N, Markham R, Russell P, et al. Nerve fibres in peritoneal endometriosis. *Hum Reprod.* 2006;21:3001–3007, used with permission; **B**, Courtesy: J Rabban, MD, MPH, UCSF; **C**, From Basbaum AL, et al. Cellular and molecular mechanisms of pain. *Cell.* 2009;139:267–284.)

validations, cycle- and stage-dependence, poorly defined controls, and low sensitivity and specificity.[124,125] Recently, however, two commercial tests have been developed towards endometriosis diagnosis based on circulating miRNAs[91] and endometrial *BCL6* expression.[116] Though promising with regards to developing a noninvasive approach, surgical diagnosis for peritoneal disease remains the gold standard.

KEY POINTS

- Diagnosis of peritoneal disease is currently surgical, while deep infiltrating endometriosis and ovarian endometrioma can be detected radiologically.

THERAPIES

Therapy for endometriosis is individualized to the patient's presenting symptoms and treatment goals. The majority of medical treatments for endometriosis-associated pain are incompatible with concomitant fertility interests, and therefore it is important to establish the patient's treatment priorities. The management approach includes treatment of existing disease and adhesions and prevention of disease recurrence. The latter consideration is underscored by the high rate of pain recurrence within 6 to 12 months of treatment completion.[126] Treatment strategies include surgical removal or ablation of existing

TABLE 125.1 Therapies for Endometriosis

Medical
Hormonal:
Oral contraceptives (cycling, noncycling)
Long-acting progestins (levonorgestrel intrauterine system, Depo-Provera, other)
Danazol
GnRH agonists (nafarelin, leuprolide acetate)
GnRH antagonists (elagolix)
Selective progesterone receptor modulators
Other:
Nonsteroidal antiinflammatory drugs
Immunotherapy
Surgical
Conservative therapy:
Laparoscopic ablation
Laparoscopic excision
Abdominal presacral neurectomy
Laparoscopic presacral neurectomy
Definitive therapy:
Hysterectomy/bilateral salpingo-oophorectomy

GnRH, Gonadotropin-releasing hormone.

endometriotic disease, suppression of lesional activity, or empiric pain and fertility approaches (Table 125.1). For prevention of endometriosis recurrence, current strategies are based on reduction or elimination of menstruation. Generally, the therapeutic approaches to endometriosis can be categorized as medical or surgical.

Medical

Pain. Medical approaches are thought to reduce endometriosis-associated pelvic pain through a variety of mechanisms, including suppressing cyclic ovarian and/or lesional production of estradiol, inhibiting the action of estradiol, minimizing inflammation, and reducing or eliminating menses.[127,128] Combined oral contraceptives, oral progestins, long-acting progestins (implants, intrauterine device), gonadotropin-releasing hormone (GnRH) analogs, danazol, and aromatase inhibitors are among the most commonly used medical treatments for endometriosis pain management. These medications reduce estradiol bioavailability via inhibition of ovulation. Many may also prevent disease recurrence by minimizing menstrual blood volume and flow. Nonsteroidal antiinflammatory drugs (NSAIDs) can be a useful adjunct in pain management because of their inhibition of prostaglandin synthesis.[127,128]

Use of a combined oral contraceptive pill with an NSAID represents the first-line approach to cyclic pelvic pain and is often attempted prior to surgery.[117] A randomized controlled trial demonstrated significant reduction in dysmenorrhea pain scores in women using oral contraceptive pill relative to placebo.[129] In patients with dysmenorrhea refractory to a cyclic oral contraceptive regimen, continuous oral contraceptive pill use is more effective in relieving pelvic pain.[130]

Long-term hypoestrogenism is beneficial for relief of pain symptoms and can be achieved with GnRH agonists, such as leuprolide acetate. A systematic review of 15 randomized trials involving 1821 patients observed a 60% to 100% improvement in dysmenorrhea pain scores with use of GnRH agonists.[131] As this treatment can result in a mostly reversible 13% decrease in bone density over a 6-month interval, "add-back" therapy with low-dose estrogens and progestins has been advocated to reduce the side effects of hypoestrogenism and decreased bone density. Although most women continue to experience significant relief of pain, approximately 30% have recurrence of symptoms within 6 months after therapy is discontinued.[132]

By inducing significant endometrial atrophy and subsequent amenorrhea, the levonorgestrel intrauterine system (LNG-IUS) is a promising approach to the prevention of disease recurrence. Indeed, randomized studies have demonstrated that LNG-IUS is more effective than no treatment in the reduction of endometriosis-associated pain after conservative surgery.[133]

Danazol, a synthetic androgen, is approved by the US Food and Drug Administration for the treatment of endometriosis-associated pain. The immunosuppressive effects of danazol render it unusual among the hormonal treatments, and this feature may add to its efficacy in the treatment of endometriosis. Though relatively effective with respect to pain control, oral danazol was associated with androgenic side effects (weight gain, acne, changes in serum lipoprotein cholesterol profile) that limited patient acceptance, leading to a search for alternative routes of administration. Several nonrandomized prospective studies investigating vaginal danazol have demonstrated treatment efficacy without significant androgenic side effects.[134]

Aromatase inhibitors such as letrozole have proven effective in the control of pain associated with endometriosis. The application of these agents is based on the discovery of high levels of aromatase expression in endometriotic lesions relative to eutopic endometrium from women without endometriosis.[3] This enzyme catalyzes the conversion of androgens to estrogens, thereby increasing local estradiol bioavailability. They are best coupled with a progestin to suppress ovulation, because an aromatase inhibitor administered alone in a premenopausal woman can induce ovulation.[135]

Newer medical treatments for endometriosis-associated pain that have demonstrated promise in preclinical and/or small clinical studies include PR modulators, GnRH antagonists, kinase signaling pathway inhibitors, TNF-α inhibitors, angiogenesis inhibitors (e.g., anti-VEGF, dopamine agonists), antioxidants, niclosamide, antiinflammatory agents, and dual suppression of estrogenic and inflammatory activities.[84,136-138]

Fertility. In contrast to their beneficial role in pain management, hormonal therapies are not beneficial in the treatment of endometriosis-associated infertility.[127,128] A Cochrane analysis revealed that ovulation suppression with GnRH analogs, oral contraceptives, medroxyprogesterone acetate, or danazol, compared with placebo in women with endometriosis, has no advantage regarding clinical pregnancy rates.[139] Surgical reduction of endometriotic lesional burden avoids delay associated with ovarian suppression and may improve the pregnancy rate, particularly in patients with stage I to II endometriosis (see later). Historically, endometriosis-associated infertility has been treated with controlled ovarian hyperstimulation coupled with intrauterine insemination, or IVF. The finding of a nearly 50% reduction in IVF pregnancy rates in women with advanced (stage III–IV) endometriosis[110] suggests that IVF, as traditionally conducted, incompletely addresses the pathophysiology of endometriosis-associated infertility and led to the search for adjuncts to optimize fertility in this cohort. GnRH agonist (i.e., leuprolide acetate) treatment for 3 to 6 months preceding IVF improved pregnancy rates in women with endometriosis, and this initial finding was confirmed in a subsequent metaanalysis of three randomized trials involving a total of 165 women.[140] Minimizing chronic niche inflammation has been proposed to improve fertility.[141] Notably, a recent cohort study of 85 women with unexplained infertility and elevated *BCL6* expression in the implantation window were treated with no additional therapy before an IVF cycle or with endometriosis therapies (2 months of GnRH agonist + norethindrone acetate or laparoscopic ablation of disease). Those undergoing medical suppression or laparoscopic treatment for endometriosis had a significantly higher live birth rate, at 50% (95% CI: 23.7–76.3) and 52.4% (95% CI: 32.4–71.7), respectively, compared with controls, at 7.4% (95% CI: 2.9–17.6).[142] While these data are impressive, a randomized controlled trial is needed to determine the efficacy of this screening method and medical (and surgical) treatment in women with elevation of the endometrial inflammatory biomarker *BCL6* in the implantation window for pregnancy success.

Surgical

Pain. Laparoscopy affords unique advantages in detecting and treating endometriosis-associated pain, and laparoscopic-assisted ablation or excision of endometriotic lesions is the most prevalent surgical approach. Laparoscopic surgery improves pelvic pain after 6 months (OR 4.97; 95% CI: 1.85–13.9).[126] A Cochrane review[143] demonstrated that laparoscopic excision was superior to drainage/ablation of endometriomas larger than 3 cm, with significant reductions in the rates of endometrioma recurrence (OR 0.41 [0.18–0.93]), requirement for further surgery (OR 0.21 [0.05–0.79]), dysmenorrhea recurrence (OR 0.15 [0.06–0.38]), dyspareunia recurrence (OR 0.08 [0.01–0.51]), and baseline pain recurrence (OR 0.10 [0.02–0.56]).

Interruption of nerve pathways represents an alternative approach to controlling pelvic pain. A review of randomized clinical trials[144] reported that laparoscopic uterine nerve ablation (LUNA)—interruption of nerves along the uterosacral ligaments—versus no LUNA showed no difference in secondary dysmenorrhea pain scores. In contrast, laparoscopic presacral neurectomy (LPSN)—in which the nerve bundle is removed within the boundaries of the interiliac triangle—plus treatment of endometriosis versus treatment of endometriosis alone showed improvement in secondary dysmenorrhea pain scores at the 6-month (OR 4.52 [1.84–11.09]) and 12-month (OR 3.14 [1.59–6.21]) postoperative timepoints.[145] However, there are increased risks of adverse effects with presacral neurectomy, including urinary urgency and constipation. Definitive surgery for recalcitrant pelvic pain is total hysterectomy and bilateral salpingo-oophorectomy, performed by laparotomy or laparoscopy or as a laparoscopically-assisted vaginal procedure. In women who have successful pain relief, however, symptoms recur in 12% to 54% within a year of therapy.[146] This observation suggests that adjunctive medical therapy may be helpful, and indeed, both preoperative and postoperative medical therapies have been efficacious.[147]

Fertility. Laparoscopic assessment and treatment of endometriosis-associated infertility is a viable option, particularly for patients with coexisting pelvic pain symptoms. Laparoscopy offers the opportunity to stage endometriosis toward triaging further infertility treatment, as patients with more advanced (stage III–IV) disease may need to consider IVF, particularly if concomitant ovarian reserve concerns are present. A controlled trial involving 341 women with infertility undergoing laparoscopy showed a significantly higher cumulative pregnancy rate in women with stage I to II endometriosis randomized to ablation of lesions as compared with untreated patients (30.7% vs. 17.7%).[148] A second smaller study did not confirm these results, although a subsequent metaanalysis of the two trials favored surgical treatment (OR for pregnancy = 1.7; 95% CI: 1.1–2.5).[149] The management of infertile patients with endometrioma(s) is less clear, as surgical removal with respect to reduction of inflammation must be weighed against the deleterious impact on ovarian reserve owing to collateral damage to adjacent healthy ovarian tissue.[150] Surgical removal of endometriomas is recommended if they are large (>3 cm), or if concern exists regarding the exact nature of the ovarian lesion. In general, by reducing endometriotic disease, the inflammatory response within the pelvis is reduced. Based on known actions of the mitogens, morphogens, chemoattractants, and angiogenic factors within the PF, it is likely that a reduction in these factors will result in an environment that is more favorable for fertility.

KEY POINTS

- Treatments for endometriosis-related pelvic pain focus on surgery, anti-inflammatory drugs, and hormonal therapies minimizing or inhibiting menstrual flow and blocking estrogen synthesis and action, with newly identified signaling pathways holding promise for future therapeutic targets.

Integrated Approach to the Patient With Endometriosis

Management of patients with endometriosis requires a multidisciplinary team that includes expertise in gynecology, general surgery, reproductive endocrinology and infertility, urology, immunology, genetics, oncology, psychology, nursing, nutrition, pain management, and complementary medicine, as well as patient support groups.[5] A unilateral approach to achieving patient goals is unwise, as most patients with endometriosis have multiple issues that warrant careful, integrated, and sensitive attention.

Endometriosis-Associated Cancer Risk

The first report on the morphologic relationship between endometriosis and ovarian cancer was presented by Sampson in 1925.[151] Two case-controlled studies by Ness and colleagues on ovarian cancer revealed endometriosis as a risk factor,[152,153] and the coexistence of endometriosis and ovarian cancer has been reported to range between 0.7% and 5% of all cases with ovarian endometriosis.[154] Berqvist and colleagues, using the National Swedish Inpatient Register, identified 64,492 women discharged from hospital with a diagnosis of endometriosis over a 31-year period (between 1969 and 2000) and linked the data to the National Swedish Cancer Register to evaluate standardized incidence ratios (SIR) of cancer.[155] Elevated risks were found for ovarian cancer (SIR 1.43; 95% CI: 1.19–1.71), and women with early diagnosed and longstanding endometriosis had a higher risk of ovarian cancer (SIR of 2.01 and 2.23, respectively). A large pooled analysis of case control studies established a history of endometriosis to be associated with a nearly 50% increase in ovarian cancer risk (OR 1.46; 95% CI: 1.31–1.63).[156]

Important clinical and molecular differences exist among the histologic subtypes of ovarian cancer. In the general population, the serous subtype is most common, representing approximately 70% of cases, whereas endometrioid (20%) and clear cell (5%) carcinomas are less prevalent. Compared with serous carcinoma, endometrioid and clear cell tumors are associated with younger age of onset, presentation at an earlier stage, and reduced sensitivity to common ovarian cancer chemotherapeutic regimens.[218] A link between the endometrioid and clear cell histologic subtypes of ovarian cancer and endometriosis is well supported. In women undergoing surgery for ovarian cancer, the rate of endometriosis is significantly higher with clear cell (26.3%) and endometrioid (21.1%) subtypes than with mucinous (5.6%) or serous (3.6%) subtypes.[157] Synchronous incidence of endometriosis with clear cell (41%) and endometrioid (38%) ovarian cancer suggests malignant transformation,[158] and conceptualization of endometriosis as a precursor lesion for these histologic subtypes is supported by molecular evidence. A mouse model of *K-RAS* protooncogene activation in the ovarian epithelium coupled with targeted deletion of the *PTEN* tumor suppressor gene resulted in development of endometriosis and endometrioid ovarian cancer,[64] highlighting the potential for malignant transformation of endometriosis to ovarian cancer. *ARID1A* has been strongly implicated in the malignant transformation of endometriosis, with somatic mutations observed in 46% of clear cell carcinomas and 30% of endometrioid carcinomas.[63] In several patients, *ARID1A* somatic mutations were evident in tumor and contiguous atypical endometriosis, but not in distant lesions, suggesting malignant transformation from local ovarian endometriotic tissue. The iron-rich microenvironment within ovarian endometriomas is associated with an increase in ROS, which may predispose to cancer-inducing somatic mutations in these lesions.[159]

A study of Swedish women with endometriosis cross-referenced against the National Swedish Cancer registry demonstrated a protective effect against ovarian cancer for both one-sided oophorectomy (OR 0.19; 95% CI: 0.08–0.46) and excision of all visible endometriosis (OR 0.30; 95% CI: 0.12–0.74).[160] Oral contraceptive pills represent a more conservative chemoprophylactic approach, as use for greater than 10 years was associated with an 80% reduction in epithelial ovarian cancer risk among women with endometriosis.[161] Further study is needed to accurately identify women with endometriosis at risk for tumorigenesis and to validate appropriate preventative measures.

SUMMARY

Endometriosis may be a result of modern advances in the twentieth and twenty-first centuries, including contraception (decreased frequency of pregnancy, and thus greater numbers of menses during a woman's reproductive lifetime), hormonal replacement therapy, and modern industrial development with its attendant immunotoxicants and estrogen-like compounds. Advances in basic and clinical research regarding the pathophysiology of endometriosis have begun to identify therapies to minimize pain or maximize fertility. The ultimate therapy for this disorder, however, is prevention. Hysterectomy for endometriosis increased dramatically between 1965 and 1984,[226] probably because of an increased awareness of the presence of endometriosis and perhaps because of an increasing incidence of the disorder. A major global effort to minimize environmental toxins, coupled with pharmacologic suppression of menses, would ideally reduce the incidence of this disease. Although this result may be long in coming, patient advocacy, education, and support, coupled with basic and clinical research in endometriosis, should also be high priorities in a national women's health agenda.

REFERENCES

1. Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med*. 2020;382:1244–1256.
2. Revised American Society for Reproductive Medicine classification of endometriosis. *Fertil Steril*. 1997;67:817–821.
3. Bulun SE, Yilmaz BD, Sison C, et al. Endometriosis. *Endocr Rev*. 2019;40:1048–1079.
4. Burney RO, Talbi S, Hamilton AE, et al. Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. *Endocrinology*. 2007;148:3814–3826.
5. D'hooghe T, Hummelshoj L. Multi-disciplinary centres/networks of excellence for endometriosis management and research: a proposal. *Hum Reprod*. 2006;21:2743–2748.
6. Nnoaham KE, Hummelshoj L, Webster P, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil Steril*. 2011;96:366–373.e8.
7. Simoens S, Dunselman G, Dirksen C, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Hum Reprod*. 2012;27:1292–1299.
8. ACOG Committee Opinion No. 760. Dysmenorrhea and endometriosis in the adolescent. *Obstet Gynecol*. 2018;132:e249–e258.
9. Hirsch M, Dhillon-Smith R, Cutner AS, et al. The prevalence of endometriosis in adolescents with pelvic pain: a systematic review. *J Pediatr Adolesc Gynecol*. 2020;33:623–630.
10. Missmer SA, Hankinson SE, Spiegelman D, et al. Reproductive history and endometriosis among premenopausal women. *Obstet Gynecol*. 2004;104:965–974.
11. Missmer SA, Hankinson SE, Spiegelman D, et al. In utero exposures and the incidence of endometriosis. *Fertil Steril*. 2004;82:1501–1508.
12. Bougie O, Yap MI, Sikora L, et al. Influence of race/ethnicity on prevalence and presentation of endometriosis: a systematic review and meta-analysis. *BJOG*. 2019;126:1104–1115.
13. Shah DK, Correia KF, Vitonis AF, et al. Body size and endometriosis: results from 20 years of follow-up within the Nurses' Health Study II prospective cohort. *Hum Reprod*. 2013;28:1783–1792.
14. Bulletti C, Montini A, Setti PL, et al. Vaginal parturition decreases recurrence of endometriosis. *Fertil Steril*. 2010;94:850–855.
15. Gemmell LC, Webster KE, Kirtley S, et al. The management of menopause in women with a history of endometriosis: a systematic review. *Hum Reprod Update*. 2017;23:481–500.
16. Shroen D. Disputatio inauguralis medica de ulceribus uteri. *Jena Krebs*. 1690:6–17.
17. Rosenfeld DL, Lecher BD. Endometriosis in a patient with Rokitansky-Kuster-Hauser syndrome. *Am J Obstet Gynecol*. 1981;139:105.
18. Oliker AJ, Harris AE. Endometriosis of the bladder in a male patient. *J Urol*. 1971;106:858–859.
19. Schrodt GR, Alcorn MO, Ibanez J. Endometriosis of the male urinary system: a case report. *J Urol*. 1980;124:722–723.
20. Halban J. Metastatic hysteroadenosis. *Wien Klin Wochenschr*. 1924;37:1205–1206.
21. Javert CT. Observations on the pathology and spread of endometriosis based on the theory of benign metastasis. *Am J Obstet Gynecol*. 1951;62:477–487.
22. Sampson JA. Metastatic or embolic endometriosis, due to the menstrual dissemination of endometrial tissue into the venous circulation. *Am J Pathol*. 1927;3:93–110.
23. Ueki M. Histologic study of endometriosis and examination of lymphatic drainage in and from the uterus. *Am J Obstet Gynecol*. 1991;165:201–209.
24. Javert CT. The spread of benign and malignant endometrium in the lymphatic system with a note on coexisting vascular involvement. *Am J Obstet Gynecol*. 1952;64:780–806.
25. Jubanyik KJ, Comite F. Extrapelvic endometriosis. *Obstet Gynecol Clin North Am*. 1997;24:411–440.
26. Sampson JA. Peritoneal endometriosis due to menstrual dissemination of endometrial tissue into the peritoneal cavity. *Am J Obstet Gynecol*. 1927;14:422–469.
27. Halme J, Hammond MG, Hulka JF, et al. Retrograde menstruation in healthy women and in patients with endometriosis. *Obstet Gynecol*. 1984;64:151–154.
28. Salamanca A, Beltran E. Subendometrial contractility in menstrual phase visualized by transvaginal sonography in patients with endometriosis. *Fertil Steril*. 1995;64:193–195.
29. D'hooghe TM, Bambra CS, Suleman MA, et al. Development of a model of retrograde menstruation in baboons (papio anubis). *Fertil Steril*. 1994;62:635–638.
30. Olive DL, Henderson DY. Endometriosis and mullerian anomalies. *Obstet Gynecol*. 1987;69:412–415.
31. Demir B, Dilbaz B, Karadag B, et al. Coexistence of endometriosis and uterine septum in patients with abortion or infertility. *J Obstet Gynaecol Res*. 2011;37:1596–1600.
32. Nisolle M, Donnez J. Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. *Fertil Steril*. 1997;68:585–596.
33. Wu Y, Strawn E, Basir Z, et al. Genomic alterations in ectopic and eutopic endometria of women with endometriosis. *Gynecol Obstet Invest*. 2006;62:148–159.

34. Gargett CE, Schwab KE, Brosens JJ, et al. Potential role of endometrial stem/progenitor cells in the pathogenesis of early-onset endometriosis. *Mol Hum Reprod*. 2014;20:591–598.
35. Noe M, Ayhan A, Wang TL, et al. Independent development of endometrial epithelium and stroma within the same endometriosis. *J Pathol*. 2018;245:265–269.
36. Wang Y, Nicholes K, Shih IM. The origin and pathogenesis of endometriosis. *Annu Rev Pathol*. 2020;15:71–95.
37. Hadfield RM, Mardon HJ, Barlow DH, et al. Endometriosis in monozygotic twins. *Fertil Steril*. 1997;68:941–942.
38. Moen MH, Magnus P. The familial risk of endometriosis. *Acta Obstet Gynecol Scand*. 1993;72:560–564.
39. Lee SH, Harold D, Nyholt DR, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet*. 2013;22:832–841.
40. Kennedy S, Hadfield R, Mardon H, et al. Age of onset of pain symptoms in non-twin sisters concordant for endometriosis. *Hum Reprod*. 1996;11:403–405.
41. Hadfield RM, Yudkin PL, Coe CL, et al. Risk factors for endometriosis in the rhesus monkey (macaca mulatta): a case-control study. *Hum Reprod Update*. 1997;3:109–115.
42. Sapkota Y, Steinthorsdottir V, Morris AP, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun*. 2017;8:15539.
43. Rahmioglu N, Banasik K, Christofidou P, et al. Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions. *bioRxiv*. 2018. https://doi.org/10.1101/406967.
44. Zoeller RT, Brown TR, Doan LL, et al. Endocrine-disrupting chemicals and public health protection: a statement of principles from the Endocrine Society. *Endocrinology*. 2012;153:4097–4110.
45. Ho SM, Cheong A, Adgent MA, et al. Environmental factors, epigenetics, and developmental origin of reproductive disorders. *Reprod Toxicol*. 2017;68:85–104.
46. Spencer TE, Hayashi K, Hu J, et al. Comparative developmental biology of the mammalian uterus. *Curr Top Dev Biol*. 2005;68:85–122.
47. Couse JF, Korach KS. Estrogen receptor-alpha mediates the detrimental effects of neonatal diethylstilbestrol (DES) exposure in the murine reproductive tract. *Toxicology*. 2004;205:55–63.
48. Signorile PG, Spugnini EP, Mita L, et al. Pre-natal exposure of mice to bisphenol A elicits an endometriosis-like phenotype in female offspring. *Gen Comp Endocrinol*. 2010;168:318–325.
49. Cummings AM, Hedge JM, Birnbaum LS. Effect of prenatal exposure to TCDD on the promotion of endometriotic lesion growth by TCDD in adult female rats and mice. *Toxicol Sci*. 1999;52:45–49.
50. Bruner-Tran KL, Ding T, Osteen KG. Dioxin and endometrial progesterone resistance. *Semin Reprod Med*. 2010;28:59–68.
51. Smarr MM, Kannan K, Buck Louis GM. Endocrine disrupting chemicals and endometriosis. *Fertil Steril*. 2016;106:959–966.
52. Rier SE, Martin DC, Bowman RE, et al. Endometriosis in rhesus monkeys (macaca mulatta) following chronic exposure to 2,3,7,8-tetrachloro dibenzo-p-dioxin. *Fundam Appl Toxicol*. 1993;21:433–441.
53. Johnson KL, Cummings AM, Birnbaum LS. Promotion of endometriosis in mice by polychlorinated dibenzo-p-dioxins, dibenzofurans, and biphenyls. *Environ Health Perspect*. 1997;105:750–755.
54. Bruner-Tran KL, Gnecco J, Ding T, et al. Exposure to the environmental endocrine disruptor TCDD and human reproductive dysfunction: translating lessons from murine models. *Reprod Toxicol*. 2017;68:59–71.
55. Bruner-Tran KL, Rier SE, Eisenberg E, et al. The potential role of environmental toxins in the pathophysiology of endometriosis. *Gynecol Obstet Invest*. 1999;48:45–56.
56. Jones RL, Lang SA, Kendziorski JA, et al. Use of a mouse model of experimentally induced endometriosis to evaluate and compare the effects of bisphenol A and bisphenol AF exposure. *Environ Health Perspect*. 2018;126:127004.
57. Kim SH, Cho S, Ihm HJ, et al. Possible role of phthalate in the pathogenesis of endometriosis: in vitro, animal, and human data. *J Clin Endocrinol Metab*. 2015;100:E1502–E1511.
58. Cano-Sancho G, Ploteau S, Matta K, et al. Human epidemiological evidence about the associations between exposure to organochlorine chemicals and endometriosis: systematic review and meta-analysis. *Environ Int*. 2019;123:209–223.
59. Wen X, Xiong Y, Qu X, et al. The risk of endometriosis after exposure to endocrine-disrupting chemicals: a meta-analysis of 30 epidemiology studies. *Gynecol Endocrinol*. 2019;35:645–650.
60. Sirohi D, Al Ramadhani R, Knibbs LD. Environmental exposures to endocrine disrupting chemicals (EDCs) and their role in endometriosis: a systematic literature review. *Rev Environ Health*. 2021;36:101–115.
61. Anglesio MS, Papadopoulos N, Ayhan A, et al. Cancer-associated mutations in endometriosis without cancer. *N Engl J Med*. 2017;376:1835–1848.
62. Suda K, Nakaoka H, Yoshihara K, et al. Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. *Cell Rep*. 2018;24:1777–1789.
63. Wiegand KC, Shah SP, Al-Agha OM, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med*. 2010;363:1532–1543.
64. Dinulescu DM, Ince TA, Quade BJ, et al. Role of K-ras and Pten in the development of mouse models of endometriosis and endometrioid ovarian cancer. *Nat Med*. 2005;11:63–70.
65. Burney RO, Giudice LC. Reprint of: pathogenesis and pathophysiology of endometriosis. *Fertil Steril*. 2019;112:e153–e161.
66. Witz CA, Cho S, Centonze VE, et al. Time series analysis of transmesothelial invasion by endometrial stromal and epithelial cells using three-dimensional confocal microscopy. *Fertil Steril*. 2003;1:770–778.
67. Jones RK, Searle RF, Bulmer JN. Apoptosis and bcl-2 expression in normal human endometrium, endometriosis and adenomyosis. *Hum Reprod*. 1998;13:3496–3502.
68. Wingfield M, Macpherson A, Healy DL, et al. Cell proliferation is increased in the endometrium of women with endometriosis. *Fertil Steril*. 1995;64:340–346.
69. Han SJ, Jung SY, Wu SP, et al. Estrogen receptor beta modulates apoptosis complexes and the inflammasome to drive the pathogenesis of endometriosis. *Cell*. 2015;163:960–974.
70. Klemmt PA, Carver JG, Koninckx P, et al. Endometrial cells from women with endometriosis have increased adhesion and proliferative capacity in response to extracellular matrix components: towards a mechanistic model for endometriosis progression. *Hum Reprod*. 2007;22:3139–3147.
71. Symons LK, Miller JE, Kay VR, et al. The immunopathophysiology of endometriosis. *Trends Mol Med*. 2018;24:748–762.
72. Rodgers WH, Matrisian LM, Giudice LC, et al. Patterns of matrix metalloproteinase expression in cycling endometrium imply differential functions and regulation by steroid hormones. *J Clin Invest*. 1994;94: 946–953.
73. Kokorine I, Nisolle M, Donnez J, et al. Expression of interstitial collagenase (matrix metalloproteinase-1) is related to the activity of human endometriotic lesions. *Fertil Steril*. 1997;68:246–251.
74. Dyson MT, Roqueiro D, Monsivais D, et al. Genome-wide DNA methylation analysis predicts an epigenetic switch for GATA factor expression in endometriosis. *PLoS Genet*. 2014;10:e1004158.
75. Bruner-Tran KL, Eisenberg E, Yeaman GR, et al. Steroid and cytokine regulation of matrix metalloproteinase expression in endometriosis and the establishment of experimental endometriosis in nude mice. *J Clin Endocrinol Metab*. 2002;87:4782–4791.
76. Oral E, Arici A. Peritoneal growth factors and endometriosis. *Semin Reprod Endocrinol*. 1996;14:257–267.
77. Oosterlynck DJ, Meuleman C, Waer M, et al. Transforming growth factor-beta activity is increased in peritoneal fluid from women with endometriosis. *Obstet Gynecol*. 1994;83:287–292.
78. Overton C, Fernandez-Shaw S, Hicks B, et al. Peritoneal fluid cytokines and the relationship with endometriosis and pain. *Hum Reprod*. 1996;11:380–386.
79. Huang JC, Papasakelariou C, Dawood MY. Epidermal growth factor and basic fibroblast growth factor in peritoneal fluid of women with endometriosis. *Fertil Steril*. 1996;65:931–934.
80. Giudice LC, Dsupin BA, Gargosky SE, et al. The insulin-like growth factor system in human peritoneal fluid: its effects on endometrial stromal

cells and its potential relevance to endometriosis. *J Clin Endocrinol Metab*. 1994;79:1284–1293.
81. Sugawara J, Fukaya T, Murakami T, et al. Increased secretion of hepatocyte growth factor by eutopic endometrial stromal cells in women with endometriosis. *Fertil Steril*. 1997;68:468–472.
82. Levey KA. Use of fluorescence imaging technology to identify peritoneal endometriosis: a case report of new technology. *Surg Laparosc Endosc Percutan Tech*. 2014;24:e63–e65.
83. Shifren JL, Tseng JF, Zaloudek CJ, et al. Ovarian steroid regulation of vascular endothelial growth factor in the human endometrium: implications for angiogenesis during the menstrual cycle and in the pathogenesis of endometriosis. *J Clin Endocrinol Metab*. 1996;81:3112–3118.
84. Liu S, Xin X, Hua T, et al. Efficacy of anti-VEGF/VEGFR agents on animal models of endometriosis: a systematic review and meta-analysis. *PLoS One*. 2016;11:e0166658.
85. Zeitoun KM, Bulun SE. Aromatase: a key molecule in the pathophysiology of endometriosis and a therapeutic target. *Fertil Steril*. 1999;72:961–969.
86. Zeitoun K, Takayama K, Sasano H, et al. Deficient 17beta-hydroxysteroid dehydrogenase type 2 expression in endometriosis: failure to metabolize 17beta-estradiol. *J Clin Endocrinol Metab*. 1998;83:4474–4480.
87. Bernardi LA, Dyson MT, Tokunaga H, et al. The essential role of GATA6 in the activation of estrogen synthesis in endometriosis. *Reprod Sci*. 2019;26:60–69.
88. Xue Q, Lin Z, Cheng YH, et al. Promoter methylation regulates estrogen receptor 2 in human endometrium and endometriosis. *Biol Reprod*. 2007;77:681–687.
89. Nothnick WB. MicroRNAs and endometriosis: distinguishing drivers from passengers in disease pathogenesis. *Semin Reprod Med*. 2017;35:173–180.
90. Pan Q, Luo X, Toloubeydokhti T, et al. The expression profile of microRNA in endometrium and endometriosis and the influence of ovarian steroids on their expression. *Mol Hum Reprod*. 2007;13:797–806.
91. Moustafa S, Burn M, Mamillapalli R, et al. Accurate diagnosis of endometriosis using serum microRNAs. *Am J Obstet Gynecol*. 2020;223:557.e1–557e11.
92. Patel BG, Lenk EE, Lebovic DI, et al. Pathogenesis of endometriosis: interaction between Endocrine and inflammatory pathways. *Best Pract Res Clin Obstet Gynaecol*. 2018;50:50–60.
93. Vallve-Juanico J, Houshdaran S, Giudice LC. The endometrial immune environment of women with endometriosis. *Hum Reprod Update*. 2019;25:564–591.
94. Ahn SH, Khalaj K, Young SL, et al. Immune-inflammation gene signatures in endometriosis patients. *Fertil Steril*. 2016;106:1420–1431.
95. Guo M, Bafligil C, Tapmeier T, et al. Mass cytometry analysis reveals a distinct immune environment in peritoneal fluid in endometriosis: a characterisation study. *BMC Med*. 2020;18:3.
96. Beste MT, Pfaffle-Doyle N, Prentice EA, et al. Molecular network analysis of endometriosis reveals a role for c-Jun-regulated macrophage activation. *Sci Transl Med*. 2014;6:222ra16.
97. Mu F, Harris HR, Rich-Edwards JW, et al. A prospective study of inflammatory markers and risk of endometriosis. *Am J Epidemiol*. 2018;187:515–522.
98. Burney RO, Lathi RB. Menstrual bleeding from an endometriotic lesion. *Fertil Steril*. 2009;91:1926–1927.
99. Celik O, Ersahin A, Acet M, et al. Disulfiram, as a candidate NF-kappaB and proteasome inhibitor, prevents endometriotic implant growing in a rat model of endometriosis. *Eur Rev Med Pharmacol Sci*. 2016;20:4380–4389.
100. Kang YJ, Jeung IC, Park A, et al. An increased level of IL-6 suppresses NK cell activity in peritoneal fluid of patients with endometriosis via regulation of SHP-2 expression. *Hum Reprod*. 2014;29:2176–2189.
101. Shigesi N, Kvaskoff M, Kirtley S, et al. The association between endometriosis and autoimmune diseases: a systematic review and meta-analysis. *Hum Reprod Update*. 2019;25:486–503.
102. Agarwal SK, Chapron C, Giudice LC, et al. Clinical diagnosis of endometriosis: a call to action. *Am J Obstet Gynecol*. 2019;220:354.e1–e12.
103. Morotti M, Vincent K, Becker CM. Mechanisms of pain in endometriosis. *Eur J Obstet Gynecol Reprod Biol*. 2017;209:8–13.
104. Vernon MW, Beard JS, Graves K, et al. Classification of endometriotic implants by morphologic appearance and capacity to synthesize prostaglandin F. *Fertil Steril*. 1986;46:801–806.
105. Berkley KJ, Rapkin AJ, Papka RE. The pains of endometriosis. *Science*. 2005;308:1587–1589.
106. Tokushige N, Markham R, Russell P, et al. High density of small nerve fibres in the functional layer of the endometrium in women with endometriosis. *Hum Reprod*. 2006;21:782–787.
107. Tokushige N, Markham R, Russell P, et al. Nerve fibres in peritoneal endometriosis. *Hum Reprod*. 2006;21:3001–3007.
108. Popovici RM, Kao LC, Giudice LC. Discovery of new inducible genes in in vitro decidualized human endometrial stromal cells using microarray technology. *Endocrinology*. 2000;141:3510–3513.
109. As-Sanie S, Kim J, Schmidt-Wilcke T, et al. Functional connectivity is associated with altered brain chemistry in women with endometriosis-associated chronic pelvic pain. *J Pain*. 2016;17:1–13.
110. Barnhart K, Dunsmoor-Su R, Coutifaris C. Effect of endometriosis on in vitro fertilization. *Fertil Steril*. 2002;77:1148–1155.
111. Harb HM, Gallos ID, Chu J, et al. The effect of endometriosis on in vitro fertilisation outcome: a systematic review and meta-analysis. *BJOG*. 2013;120:1308–1320.
112. Burke RK. Effect of peritoneal washings from women with endometriosis on sperm velocity. *J Reprod Med*. 1987;32:743–746.
113. Coddington CC, Oehninger S, Cunningham DS, et al. Peritoneal fluid from patients with endometriosis decreases sperm binding to the zona pellucida in the hemizona assay: a preliminary report. *Fertil Steril*. 1992;57:783–786.
114. Heitmann RJ, Tobler KJ, Gillette L, et al. Dexamethasone attenuates the embryotoxic effect of endometriotic peritoneal fluid in a murine model. *J Assist Reprod Genet*. 2015;32:1317–1323.
115. Lessey BA, Kim JJ. Endometrial receptivity in the eutopic endometrium of women with endometriosis: it is affected, and let me show you why. *Fertil Steril*. 2017;108:19–27.
116. Evans-Hoeker E, Lessey BA, Jeong JW, et al. Endometrial BCL6 overexpression in eutopic endometrium of women with endometriosis. *Reprod Sci*. 2016;23:1234–1241.
117. Giudice LC. Clinical practice. Endometriosis. *N Engl J Med*. 2010;362:2389–2398.
118. Walter AJ, Hentz JG, Magtibay PM, et al. Endometriosis: correlation between histologic and visual findings at laparoscopy. *Am J Obstet Gynecol*. 2001;184:1407–1411; discussion 11–13.
119. Hadfield R, Mardon H, Barlow D, et al. Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. *Hum Reprod*. 1996;11:878–880.
120. Sutton CJ, Ewen SP, Whitelaw N, et al. Prospective, randomized, double-blind, controlled trial of laser laparoscopy in the treatment of pelvic pain associated with minimal, mild, and moderate endometriosis. *Fertil Steril*. 1994;62:696–700.
121. Telimaa S, Ronnberg L, Kauppila A. Placebo-controlled comparison of danazol and high-dose medroxyprogesterone acetate in the treatment of endometriosis after conservative surgery. *Gynecol Endocrinol*. 1987;1:363–371.
122. Brosens I, Puttemans P, Campo R, et al. Diagnosis of endometriosis: pelvic endoscopy and imaging techniques. *Best Pract Res Clin Obstet Gynaecol*. 2004;18:285–303.
123. Tang X, Ling R, Gong J, et al. Deep infiltrating endometriosis MR imaging with surgical correlation. *Quant Imaging Med Surg*. 2018;8:187–195.
124. May KE, Conduit-Hulbert SA, Villar J, et al. Peripheral biomarkers of endometriosis: a systematic review. *Hum Reprod Update*. 2010;16:651–674.
125. May KE, Villar J, Kirtley S, et al. Endometrial alterations in endometriosis: a systematic review of putative biomarkers. *Hum Reprod Update*. 2011;17:637–653.
126. Jacobson TZ, Duffy JM, Barlow D, et al. Laparoscopic surgery for pelvic pain associated with endometriosis. *Cochrane Database Syst Rev*. 2009(4):CD001300.
127. Practice Committee of the American Society for Reproductive M. Treatment of pelvic pain associated with endometriosis: a committee opinion. *Fertil Steril*. 2014;101:927–935.

128. Dunselman GA, Vermeulen N, Becker C, et al. ESHRE guideline: management of women with endometriosis. *Hum Reprod*. 2014;29:400–412.
129. Harada T, Momoeda M, Taketani Y, et al. Low-dose oral contraceptive pill for dysmenorrhea associated with endometriosis: a placebo-controlled, double-blind, randomized trial. *Fertil Steril*. 2008;90:1583–1588.
130. Vercellini P, Frontino G, De Giorgi O, et al. Continuous use of an oral contraceptive for endometriosis-associated recurrent dysmenorrhea that does not respond to a cyclic pill regimen. *Fertil Steril*. 2003;80:560–563.
131. Prentice A, Deary AJ, Goldbeck-Wood S, et al. Gonadotrophin-releasing hormone analogues for pain associated with endometriosis. *Cochrane Database Syst Rev*. 2000(2):CD000346.
132. Dlugi AM, Miller JD, Knittle J. Lupron depot (leuprolide acetate for depot suspension) in the treatment of endometriosis: a randomized, placebo-controlled, double-blind study. Lupron Study Group. *Fertil Steril*. 1990;54:419–427.
133. Abou-Setta AM, Al-Inany HG, Farquhar CM. Levonorgestrel-releasing intrauterine device (LNG-IUD) for symptomatic endometriosis following surgery. *Cochrane Database Syst Rev*. 2006(4):CD005072.
134. Razzi S, Luisi S, Calonaci F, et al. Efficacy of vaginal danazol treatment in women with recurrent deeply infiltrating endometriosis. *Fertil Steril*. 2007;88:789–794.
135. Olive DL, Lindheim SR, Pritts EA. New medical treatments for endometriosis. *Best Pract Res Clin Obstet Gynaecol*. 2004;18:319–328.
136. Becker CM, Gattrell WT, Gude K, et al. Reevaluating response and failure of medical treatment of endometriosis: a systematic review. *Fertil Steril*. 2017;108:125–136.
137. Nothnick WB, Marsh C, Alali Z. Future directions in endometriosis research and therapeutics. *Curr Womens Health Rev*. 2018;14:189–194.
138. Zhao Y, Gong P, Chen Y, et al. Dual suppression of estrogenic and inflammatory activities for targeting of endometriosis. *Sci Transl Med*. 2015;7:271ra9.
139. Hughes E, Brown J, Collins JJ, et al. Ovulation suppression for endometriosis. *Cochrane Database Syst Rev*. 2007(3):CD000155.
140. Sallam HN, Garcia-Velasco JA, Dias S, et al. Long-term pituitary downregulation before in vitro fertilization (IVF) for women with endometriosis. *Cochrane Database Syst Rev*. 2006(1):CD004635.
141. Lin YH, Chen YH, Chang HY, et al. Chronic niche inflammation in endometriosis-associated infertility: current understanding and future therapeutic strategies. *Int J Mol Sci*. 2018;19:2385.
142. Likes CE, Cooper LJ, Efird J, et al. Medical or surgical treatment before embryo transfer improves outcomes in women with abnormal endometrial BCL6 expression. *J Assist Reprod Genet*. 2019;36:483–490.
143. Hart R, Hickey M, Maouris P, et al. Excisional surgery versus ablative surgery for ovarian endometriomata: a Cochrane Review. *Hum Reprod*. 2005;20:3000–3007.
144. Proctor ML, Latthe PM, Farquhar CM, et al. Surgical interruption of pelvic nerve pathways for primary and secondary dysmenorrhoea. *Cochrane Database Syst Rev*. 2005(4):CD001896.
145. Vercellini P, Fedele L, Bianchi S, et al. Pelvic denervation for chronic pain associated with endometriosis: fact or fancy? *Am J Obstet Gynecol*. 1991;165:745–749.
146. Candiani GB, Fedele L, Vercellini P, et al. Repetitive conservative surgery for recurrence of endometriosis. *Obstet Gynecol*. 1991;77:421–424.
147. Yap C, Furness S, Farquhar C. Pre and post operative medical therapy for endometriosis surgery. *Cochrane Database Syst Rev*. 2004(3):CD003678.
148. Marcoux S, Maheux R, Berube S. Laparoscopic surgery in infertile women with minimal or mild endometriosis. Canadian Collaborative Group on Endometriosis. *N Engl J Med*. 1997;337:217–222.
149. Jacobson TZ, Barlow DH, Koninckx PR, et al. Laparoscopic surgery for subfertility associated with endometriosis. *Cochrane Database Syst Rev*. 2002(4):CD001398.
150. Benaglia L, Bermejo A, Somigliana E, et al. In vitro fertilization outcome in women with unoperated bilateral endometriomas. *Fertil Steril*. 2013;99:1714–1719.
151. Sampson JA. Endometrial carcinoma of the ovary arising in endometrial tissue in that organ. *Arch Surg*. 1925;10:1–72.
152. Ness RB, Cramer DW, Goodman MT, et al. Infertility, fertility drugs, and ovarian cancer: a pooled analysis of case-control studies. *Am J Epidemiol*. 2002;155:217–224.
153. Ness RB, Grisso JA, Cottreau C, et al. Factors related to inflammation of the ovarian epithelium and risk of ovarian cancer. *Epidemiology*. 2000;11:111–117.
154. Nishida M, Watanabe K, Sato N, et al. Malignant transformation of ovarian endometriosis. *Gynecol Obstet Invest*. 2000;1:18–25.
155. Melin A, Sparen P, Persson I, et al. Endometriosis and the risk of cancer with special emphasis on ovarian cancer. *Hum Reprod*. 2006;21:1237–1242.
156. Pearce CL, Templeman C, Rossing MA, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol*. 2012;13:385–394.
157. Yoshikawa H, Jimbo H, Okada S, et al. Prevalence of endometriosis in ovarian cancer. *Gynecol Obstet Invest*. 2000;1:11–17.
158. Van Gorp T, Amant F, Neven P, et al. Endometriosis and the development of malignant tumours of the pelvis. A review of literature. *Best Pract Res Clin Obstet Gynaecol*. 2004;18:349–371.
159. Yamada Y, Shigetomi H, Onogi A, et al. Redox-active iron-induced oxidative stress in the pathogenesis of clear cell carcinoma of the ovary. *Int J Gynecol Cancer*. 2011;21:1200–1207.
160. Melin AS, Lundholm C, Malki N, et al. Hormonal and surgical treatments for endometriosis and risk of epithelial ovarian cancer. *Acta Obstet Gynecol Scand*. 2013;92:546–554.
161. Modugno F, Ness RB, Allen GO, et al. Oral contraceptive use, reproductive history, and risk of epithelial ovarian cancer in women with and without endometriosis. *Am J Obstet Gynecol*. 2004;191:733–740.

126

Uterine Fibroids and Adenomyosis

Molly B. Moravek, Serdar E. Bulun, Elizabeth A. Stewart, and Minerva Orellana

OUTLINE

Uterine Fibroids, 2093
  Introduction, 2093
  Epidemiology and Classification, 2093
  Pathogenesis, 2094
  Diagnosis, 2097
  Therapies, 2097
Adenomyosis, 2099
  Overview, 2099
  Epidemiology and Classification, 2099
  Pathogenesis, 2099
  Diagnosis, 2099
  Treatment, 2099
  Summary, 2100

KEY POINTS

- Uterine fibroids are the leading indication for hysterectomy and a cause of significant morbidity and healthcare expenditures for reproductive age women.
- Fibroids are more common and more severe in women of African descent, and the cause of this disparity is not clear.
- A salient feature of fibroids is their responsiveness to steroid hormones, and most current treatments target steroid action.
- The discovery of stem cells in leiomyoma may lead to better understanding of their pathophysiology and the development of new treatments.
- Adenomyosis is a disease that often coexists with uterine fibroids and endometriosis and causes heavy menstrual bleeding, pain, and subfertility, but has been less well understood because until recently it required hysterectomy for pathologic identification.

UTERINE FIBROIDS

Introduction

Uterine fibroids are common noncancerous neoplasms that arise from the myometrium, but their impact on women is not benign. They are the leading indication for hysterectomy and disproportionately impact women of African descent. While new medical, surgical, and procedural therapies are providing many women with alternatives to hysterectomy, a limited understanding of their pathophysiology limits new approaches. The understanding of this disease is broadening beyond the gonadal steroids estrogen and progesterone to include the role that stem cells, growth factors, and the active extracellular matrix (ECM) play in this disease.

Epidemiology and Classification

The prevalence of uterine fibroids, often referred to as uterine leiomyomas or myomas, ranges from 4.5% to 68.6%, with the incidence of ultrasound-confirmed fibroids in a US population being lower in White women than Black women, at 70% and 80%, respectively.[1,2] Uterine fibroids are symptomatic in women of reproductive age, and symptoms often wane when women reach menopause, both due to the lack of menstruation and the lower levels of the gonadal steroids estrogen and progesterone.[3]

Women with uterine fibroids experience a range of symptoms depending on the number, size, and location of their fibroid(s). Fibroid size ranges from a few millimeters to over 30 centimeters.[4] The most common symptom is abnormal uterine bleeding (AUB), designated as AUB-L when due to fibroids (leiomyomas), which is primarily heavy or prolonged menstrual bleeding and can often result in iron-deficiency anemia. Other fibroid symptoms include bulk symptoms such as bowel dysfunction, protruded abdomen, frequent urination, and painful intercourse. The presence of uterine fibroids may contribute to infertility, miscarriage, and pregnancy complications.[3]

Symptomatic uterine fibroids also impact a women's emotional and psychosocial well-being. Women have reported feelings of helplessness, fear, and anxiety surrounding their uterine fibroid diagnosis, and many report decreased quality of life. Women also reported missing work and social activities due to uterine fibroid symptoms.[5] Expense to the US healthcare system for uterine fibroid diagnosis and treatment has been estimated at approximately $34 billion dollars annually.[3]

Risk Factors. There are a few nonmodifiable risk factors that contribute to the development of uterine fibroids. Race is a risk factor seen in all studies. Black women have a higher prevalence of uterine fibroids and have a higher disease burden (larger fibroid volumes, increased number of fibroids) and more severe symptoms at time of diagnosis compared with White women. Black women also have a lower age at diagnosis compared with their White counterparts.[1,2] Furthermore, they are three times as likely to be hospitalized and seven times more likely to undergo uterine-sparing surgical myomectomies than women of other races.[6] Newer research suggests that Latina women may also have more severe symptoms compared with Black and White women in the United States.[7] Other nonmodifiable risk factors linked to fibroids include early menarche and polycystic ovarian syndrome. Nulliparous women are also at higher risk for uterine fibroids.[8] Family history of fibroids and diabetes may also be linked to the disease.[1,8]

Some modifiable factors such as diet are linked to uterine fibroid development. Women with a higher intake of fruits and vegetables are less likely to develop uterine fibroids, while women who consume a diet rich in red meats are at higher risk. High intake of long-chain omega-3 fatty acids is positively correlated with uterine fibroids.[8] A diet deficient in vitamin D and dairy is also positively correlated with uterine fibroids.[8,9] Alcohol and caffeine consumption are also a moderate risk factor.[8]

Protective Factors. Parity appears protective for fibroids and may be due to the uterine remodeling that occurs after delivery. Smoking cigarettes has an inverse relationship with the development of fibroids.[8] There is contradictory information about hormonal contraceptives. Most studies suggest that oral contraceptives (OCs) and depot forms of progestins may decrease the risk for uterine fibroids, but data from the Nurses' Health Study suggest that OCs started before age 16 may be associated with increased risk.[1]

Classification. All classification systems for uterine fibroids are based on anatomical location, and specifically the relationship between the fibroid and the serosal (outer) and mucosal (inner) surfaces of the uterus (Fig. 126.1). There are three general types of fibroids, with the most common being intramural (within the myometrium) and not reaching either uterine surface. Submucosal fibroids are inside the uterus, in contact with the endometrium. Subserosal fibroids are at the outer uterine surface. An updated classification system is provided by the International Federation of Gynecology and Obstetrics (FIGO).[10] The FIGO classification system uses numerical values to expand on the classic system to clarify extent of protrusion from the mucosal or serosa surfaces and describes fibroids that span the uterus. However, there are still clinical limitations in using the FIGO classification for diagnosis.[11]

Typically, the type of fibroid impacts the patient's symptoms. Both submucosal and intramural fibroids are linked to abnormal uterine bleeding. Subserosal fibroids show no effect on fertility, while submucosal fibroids may affect fertility, and their removal may improve fertility outcomes. Intramural fibroid position impacts fertility; however, their removal is associated with conflicting fertility outcomes.[12] Subserosal fibroids are also associated with bulk symptoms.

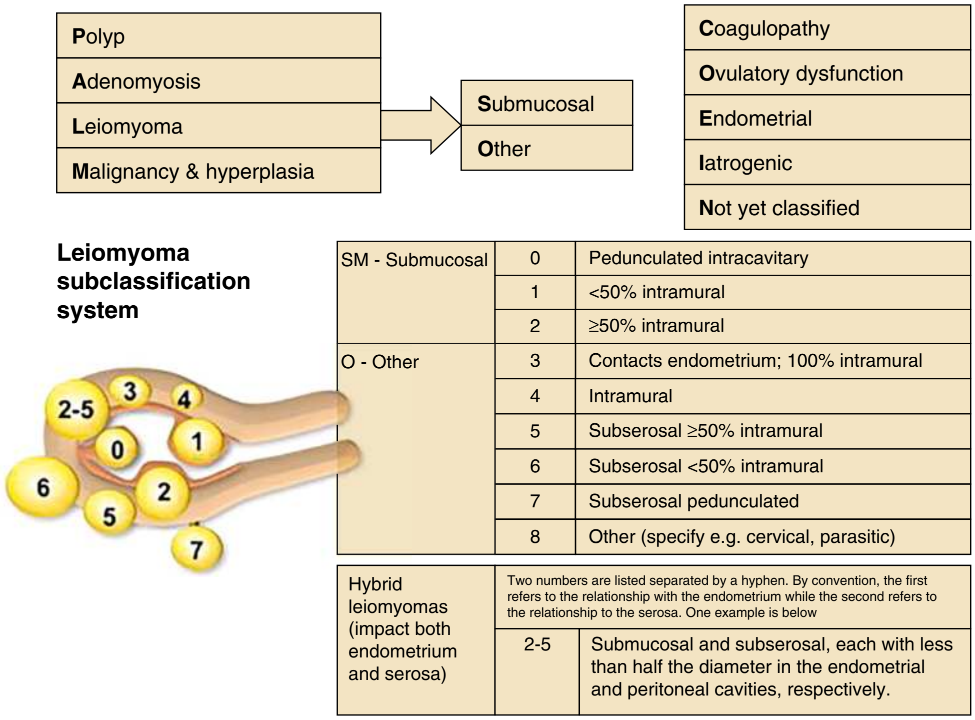
KEY POINTS

- Uterine fibroids are benign tumors arising from the myometrium. They are the leading indication for hysterectomy and a cause of significant morbidity and healthcare expenditures for reproductive age women. They are more prevalent and have greater disease burden in Black women, but the cause of this disparity is unclear.

Pathogenesis

Despite the public health impact of these tumors, their specific etiology remains unclear. Early genetic studies suggest that fibroids are monoclonal tumors, derived from a single myocyte. Neoplastic transformation of a myocyte to a leiomyoma cell is likely due to a single mutation, with mutations in the mediator complex subunit 12 (*MED12)* gene reported in approximately 70% of uterine fibroids.[13,14] Mutations affecting the expression of the high mobility AT-hook 2 (*HMGA2*) gene have also been reported in a large number of fibroids, and appear to be mutually exclusive with *MED12* mutations.[15] Regardless of the nature of the genetic hit, a common salient feature of fibroids is their responsiveness to steroid hormones, thus providing an opportunity for intervention. Estrogen and progesterone lead to tumor expansion by stimulating a modest rate of cellular proliferation and the production of copious amounts of ECM, predominantly collagen.[16,17]

Role of Progenitor Cells. A small population of stem-progenitor cells has been discovered in fibroids and appears to be important in fibroid pathophysiology.[18,19] Somatic stem cells, or tissue-specific stem cells, are a small group of cells present throughout the body that asymmetrically divide, resulting in self-renewal and daughter cells that differentiate into tissue-specific cell types. These cells, which are key for tissue regeneration and repair, are critical in maintaining normal organ function.[20,21] Similarly, tumor-initiating cells, or tumor progenitor cells, are a small group of cells within a tumor that can undergo



Polyp
Adenomyosis
Leiomyoma
Malignancy & hyperplasia

Submucosal
Other

Coagulopathy
Ovulatory dysfunction
Endometrial
Iatrogenic
Not yet classified

Leiomyoma subclassification system

SM - Submucosal	0	Pedunculated intracavitary
	1	<50% intramural
	2	≥50% intramural
O - Other	3	Contacts endometrium; 100% intramural
	4	Intramural
	5	Subserosal ≥50% intramural
	6	Subserosal <50% intramural
	7	Subserosal pedunculated
	8	Other (specify e.g. cervical, parasitic)
Hybrid leiomyomas (impact both endometrium and serosa)	Two numbers are listed separated by a hyphen. By convention, the first refers to the relationship with the endometrium while the second refers to the relationship to the serosa. One example is below	
	2-5	Submucosal and subserosal, each with less than half the diameter in the endometrial and peritoneal cavities, respectively.

Figure 126.1 **International Federation of Gynecology and Obstetrics (FIGO) classification system of leiomyomas.** (Adapted with permission from Munro 2010).

asymmetric division and thereby have the capacity for self-renewal and tumor maintenance and growth.[22,23] A subset of cells characteristic of somatic stem cells exist in human and mouse myometrial tissues, and these cells self-renew and produce daughter cells in a hormone-dependent manner (Fig. 126.2).[24-26] Additionally, a small population of cells consistent with undifferentiated somatic stem cells or tumor progenitor cells has been isolated from fibroid tissue.[18,19] It is possible that, during myometrial stem cell regeneration, a mutation may transform this population to tumor stem cells (Fig. 126.2). Subsequently, the tumor stem cells, via clonal expansion, would give rise to a fibroid (Fig. 126.2). Stem cells derived from fibroid tissue, but not myometrium, carry *MED12* mutations, suggesting that *MED12* may be one of the genes in which a mutation transforms a myometrial stem cell into a fibroid progenitor cell, which then interacts with the surrounding myometrial tissue to give rise to a fibroid tumor.[18] Uterine hypoxia and aberrant estrogen signaling have also been implicated in the transformation of a myometrial stem cell into a fibroid progenitor cell.[27,28]

When cell suspensions containing fibroid progenitor cells and mixed myometrial cells are injected under the kidney capsules of immunodeficient mice treated with estrogen and progesterone, they grow into significantly larger tumors than those containing differentiated fibroid cells and mixed myometrial cells. The tumors derived from fibroid progenitor cells also have a higher proliferation index than tumors without these cells.[18] Intriguingly, fibroid progenitor cells are relatively deficient of estrogen receptors (ERs) and progesterone receptors (PRs) but have tumorigenic capabilities when stimulated by estrogen and progesterone. Moreover, fibroid progenitor cells require the presence of either mature fibroid or myometrial cells for proliferation and growth. It is possible that these cells rely on strikingly higher levels of steroid hormone receptors in surrounding differentiated myometrial and fibroid cells to mediate estrogen and progesterone action via paracrine signaling (Fig. 126.2).[18] Moreover, multiple studies have suggested a role for paracrine pathways in the communication between fibroid stem-progenitor cells and the surrounding differentiated cells, involving the wingless-type (WNT)/β-catenin, insulin-like growth factor (IGF) 2/insulin receptor-A, and receptor activator of nuclear factor κB ligand (RANKL)/RANK pathways.[29-31] These promising results may indicate future therapeutic targets, although much remains to be explored in fibroid progenitor cells and the role they play in fibroid pathogenesis.

Steroid Hormone Action—Estrogen. Estrogen acts primarily via two nuclear receptors, ERα and ERβ, which are expressed in both the myometrium and in fibroids,[32-34] and estrogen binding upregulates expression of several genes thought to play a role in fibroid pathogenesis, including multiple growth factors, collagen, and ER and PR (Fig. 126.3).[35-37] Conflicting results have been reported regarding levels of ER expression in fibroids compared with the surrounding myometrium.[38-40] Although the exact roles of ERs in fibroid development have not been fully elucidated, it is thought that ERβ regulates the transcriptional activity of ERα, which is the stronger transcriptional activator of the two.[40] Additionally, several studies have shown that ERα

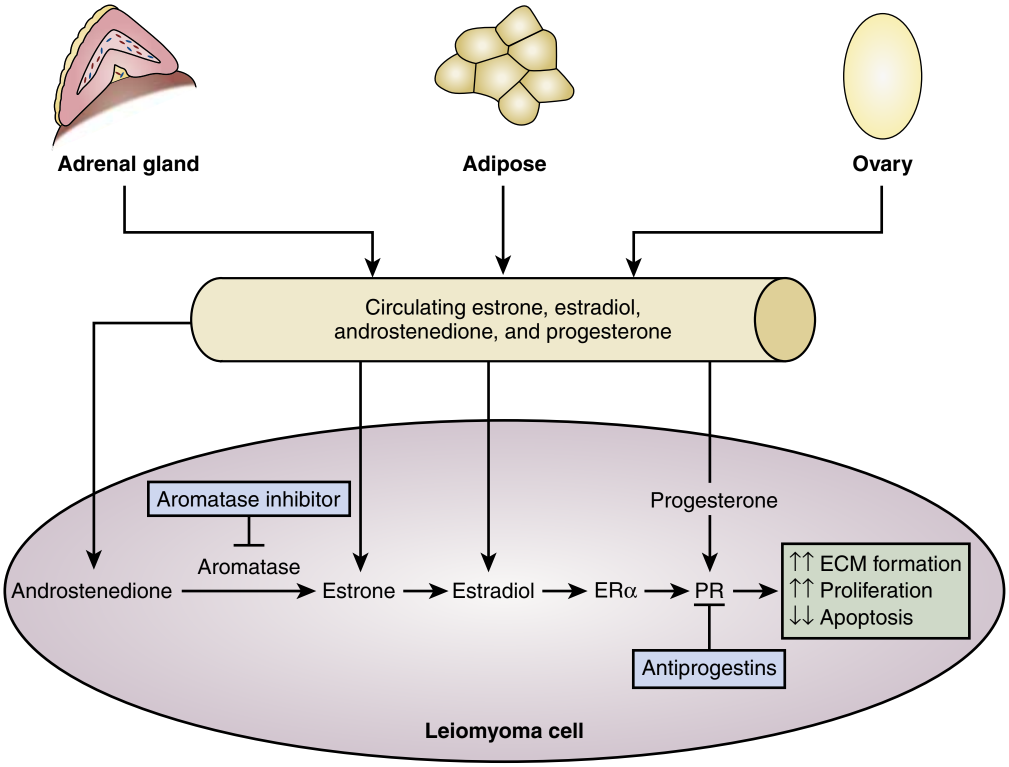


Figure 126.2 **A,** Normal myometrial tissue goes through cyclic growth under the influence of ovarian progesterone secreted in a cyclic fashion. Normal myometrial stem cells undergoing self-renewal and proliferation possibly play a central role in this process. Because stem cells are deficient in estrogen receptor α (ERα) or progesterone receptor (PR), surrounding mature myometrial cells with high steroid receptor expression mediate steroid action via paracrine mechanisms. **B,** If one of the myometrial stem cells is affected by a mutation, this cell is transformed. It may then continue clonal expansion at a rate higher than that of a normal myometrial stem cell. **C,** The mutated myometrial cell gives rise to a leiomyoma via increased self-renewal and proliferation. It eventually differentiates into mature leiomyoma cells with high ERα and PR expression, which, via paracrine signaling, stimulates further stem cell proliferation. Mature leiomyoma cells also lay down extracellular matrix (ECM) under the influence of estrogen and progesterone. This eventually gives rise to a clinically recognizable uterine fibroid.

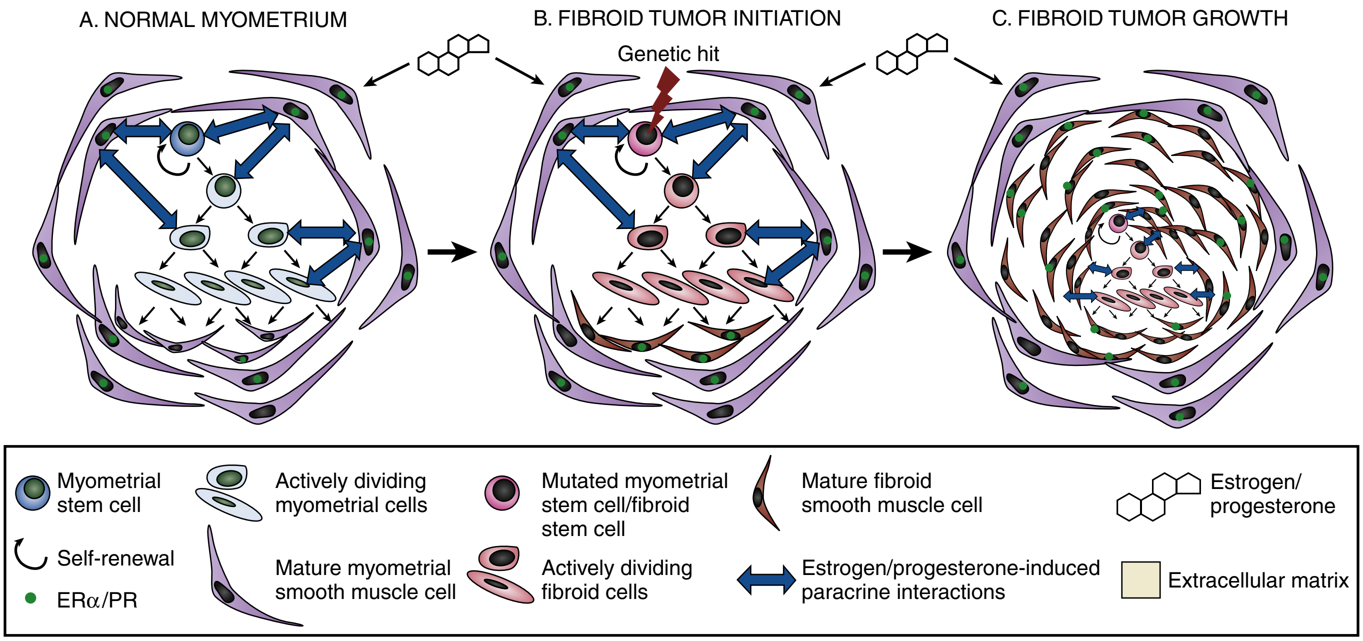

A. NORMAL MYOMETRIUM
B. FIBROID TUMOR INITIATION
C. FIBROID TUMOR GROWTH
Genetic hit
Myometrial stem cell
Self-renewal
ERα/PR
Actively dividing myometrial cells
Mature myometrial smooth muscle cell
Mutated myometrial stem cell/fibroid stem cell
Actively dividing fibroid cells
Mature fibroid smooth muscle cell
Estrogen/progesterone-induced paracrine interactions
Estrogen/ progesterone
Extracellular matrix

Figure 126.3 **Role of sex steroids in leiomyoma pathology.** The ovary, adrenal gland, and adipose tissue produce the precursors, androstenedione and estrone, of the biologically active estrogen, i.e., estradiol. Ovary also secretes progesterone. These steroids reach leiomyoma tissue in circulating blood. Aromatase converts androstenedione to estrone that is further reduced to form estradiol in leiomyoma. Estradiol via estrogen receptor α (ERα) induces the expression of progesterone receptor (PR) in leiomyoma tissue. Progesterone of ovarian origin acts via PR to increases the formation of extracellular matrix (ECM) and cell proliferation and decreases apoptosis.

polymorphisms increase susceptibility to fibroids.[41,42] ERα is also activated via phosphorylation (without estrogen binding) by the mitogen-activated protein kinase (MAPK) pathway, leading to the hypothesis that estrogen-bound ERα induces growth factor expression, which can then stimulate the MAPK pathway and further activate ERα via phosphorylation in an autocrine fashion.[43]

While estrogen has traditionally been thought of as the primary hormone responsible for fibroid growth, *in vivo* and *in vitro* studies have shown that progesterone is required for estrogen-driven fibroid growth, suggesting that estrogen is necessary, but not sufficient, for proliferation.[44-46] Moreover, estrogen/ERα regulates expression of PR, and disruption of the estrogen-signaling pathway by transfection significantly decreases both *ER* and *PR* gene expression.[44,47] Finally, in an *in vivo* human fibroid xenograft model, treatment with estrogen and progesterone stimulated not only fibroid cell proliferation, but also ECM formation, and these effects were abolished by cotreatment with the PR antagonist mifepristone (RU486).[44] These findings suggest a more permissive role for estrogen, acting via induction of PR expression, and thereby allowing fibroid responsiveness to progesterone.[44,48]

Fibroids are exposed to estrogen not only through ovarian steroidogenesis, but also through local conversion of androgens by aromatase within the tumors themselves.[49] In cultured fibroid cells, the addition of androstenedione leads to production of estrone, which is then converted to the more potent estradiol by 17β-hydroxysteroid dehydrogenase (17β-HSD).[50] Furthermore, incubation of cultured fibroid cells with androstenedione leads to similar rates of cellular proliferation as estradiol, suggesting that fibroids are capable of producing enough estrogen to sustain their own growth.[50] The addition of aromatase inhibitors then decreases proliferation in cultured cells and decreases fibroid size clinically, further supporting that aromatase is the key enzyme mediating in situ estrogen production.[48,50] Fibroid tissue has remarkably higher levels of aromatase and 17β-HSD type 1 compared with adjacent normal myometrium, which, accordingly, leads to higher tissue levels of estrogens in fibroids as well.[49-52] Moreover, aromatase is not found in the myometrium of uteri without fibroids.[49] Finally, there is increased aromatase expression in fibroids of Black American women, who are disproportionately affected by fibroids.[48] Although the molecular mechanisms underlying regulation of aromatase are not completely understood, the transcription factor CCAAT/enhancer binding protein β is a key inducer of aromatase expression via regulating its proximal promoter I.3/II region, the predominant promoter utilized by fibroid cells.[45,53]

Steroid Hormone Action—Progesterone. Clinical and experimental studies confirm that progesterone plays a key role in uterine fibroid growth and development (Fig. 126.3).[54] Progesterone acts via PR, a member of the nuclear hormone superfamily of ligand-activated transcription factors.[55,56] The two predominant PR isoforms, PR-A and PR-B, are transcribed from the same gene by two distinct promoters, resulting in differently sized proteins and different transcriptional activities.[57-61] Selective ablation of PR isoforms in mice revealed that PR-A is necessary for ovulation and modulates the antiproliferative effects of progesterone in the uterus, and that PR-B is required for normal mammary gland development and function.[62,63] Several studies have reported increased expression of both PR-A and PR-B in fibroid tissue compared with adjacent normal myometrial tissue, but the specific role of each PR isoform in uterine fibroids has yet to be fully elucidated.[64-66]

Clinical studies in postmenopausal women have shown that fibroid proliferation increases significantly with hormone replacement therapy with combined estrogen and progestin, but not with estrogen replacement alone, supporting a key role for progesterone signaling in fibroid growth.[46] Additionally, proliferation markers and mitotic counts are highest in fibroid tissue during the luteal/secretory phase, when circulating progesterone is highest.[46,67] Despite the belief that progesterone is the primary hormone driving the growth of uterine fibroids, its specific role in the pathogenesis of fibroids is not completely understood. In cultured fibroid cells, treatment with antiprogestins decreases proliferation by downregulating several growth factors and their receptors, including IGF-I, epidermal growth factor (EGF), transforming growth factor (TGF) B3, vascular endothelial growth factor (VEGF)-A, and VEGF-B.[68-72] These treatments also induce apoptosis through activation of multiple differential apoptotic pathways, including the tumor necrosis factor–related apoptosis-inducing ligand-mediated signaling pathway and the endoplasmic reticulum stress-induced pathway,[73,74] and by reducing expression of the antiapoptotic protein BCL2 and stimulating expression of cleaved caspase-3 and cleaved PARP.[75] Furthermore, antiprogestins decrease ECM formation by increasing ECM metalloproteinase inducer, MMP-1, MMP-8, and membrane type 1-MMP protein contents and decreasing TIMP-1, TIMP-2, and type I and type III collagen protein levels.[72,76] Most recently, in an *in vivo* human fibroid xenograft model, Qiang et al.[77] demonstrated that estrogen plus progesterone, but not estrogen alone, induces ECM production via downregulation of miR-29b. Although grafts containing control vector-transduced cells gave rise to typical solid fibroid tumors, fibroid grafts with ectopic mir-29b expression failed to form solid tumors in the presence of estrogen plus progesterone,[77] supporting the hypothesis that dysregulation of mir-29b plays an important role in tissue fibrosis and tumor formation.[78-81]

Genomic studies have also shed some light on progesterone's role in fibroid pathogenesis. Expression of the tumor suppressor gene Kruppel-like transcription factor 11 (*KLF11*) is significantly lower in fibroid tissues compared with adjacent myometrial tissues, and the CpG island in the *KLF11* promoter is hypermethylated in fibroid tissue.[82] Therefore, dysregulation of *KLF11* may be a key factor involved in uterine fibroids. Using a genome-wide approach, a PR-binding site was found 20.5 kb upstream of the *KLF11* promoter, and *KLF11* knockdown markedly increased fibroid cell proliferation.[82] In primary cultures of uterine fibroid cells, treatment with the antiprogestin mifepristone (RU486) robustly regulates the protein and mRNA levels of KLF11 by enhancing recruitment of Sp1, RNA polymerase II, PR, and its coactivator SRC-2 to both the distal enhancer and basal promoter region of the *KLF11* gene. It is unclear whether these activities can alter *KLF11* methylation status. Additionally, using a chromatin immunoprecipitation-sequencing assay, another novel PR target gene, perilipin 2 (*PLIN2*), was recently discovered in fibroid cells.[83] *PLIN2* normally plays an important role in lipid droplet formation and stabilization, but its loss is linked to the expression of fibrogenic genes in hepatic fibrosis.[84-86] Moreover, in clear cell renal carcinoma, higher *PLIN2* expression is associated with better cancer-specific survival and cancer-free survival.[87] These findings strongly suggest that progesterone-mediated expression of *KLF11* and *PLIN2* may regulate both fibroid cell proliferation and ECM formation.

Fibrosis. Another hallmark of uterine fibroid pathogenesis is dysregulation of the ECM, resulting in excessive ECM deposition. Fibroids exhibit overexpression of multiple ECM glycoprotein components, including collagens, fibronectin, fibulin-3, laminins, and proteoglycans.[88,89] In particular, the ECM in uterine fibroids has been shown to have increased collagen types I, III, and V compared with the surrounding myometrium.[89] Overexpression of several growth factors in uterine fibroids, most notably activin-A, TGF-B, and platelet-derived growth factor, likely contribute to overproduction of ECM components.[88,89] Moreover, it has been hypothesized that the ECM itself may protect these growth factors from degradation, thus compounding the dysregulated growth.[88]

KEY POINTS

- A salient feature of fibroids is their responsiveness to steroid hormones, but genetic mutations, progenitor cells, and fibrosis from dysregulation of the extracellular matrix also play a role.

Diagnosis

Despite the high incidence, only 25% to 50% of women report fibroid-related symptoms.[1] Some fibroids are truly asymptomatic, and other symptoms are likely attributed to either normal function or other processes such as aging or musculoskeletal pain. Uterine fibroids are often first suspected when a mass is felt during a pelvic exam due to subserosal fibroids.[3] Fibroids are sometimes an incidental finding on an ultrasound ordered for another clinical indication.[8] The presence of uterine fibroids is often confirmed by ultrasonography, either transabdominally or transvaginally, based on the size of the uterus. Transvaginal ultrasonography provides more resolution and detail on size, number, and location of fibroids.[90] Serial ultrasounds may be useful for monitoring the size of uterine fibroids over time.[4] Instilling saline into the endometrial cavity while using ultrasound imaging for a sonohysterogram aids in defining submucosal fibroids and differentiates them from endometrial polyps and clot.[90]

Magnetic resonance imaging (MRI) is a more expensive imaging modality that is reserved for special uses for fibroid imaging. This technique is usually warranted when ultrasonography is inconclusive, uterine sarcoma is suspected, or the precise location or vascularity of a fibroid is important to determine use of specific treatment options.[4]

KEY POINTS

- Despite the high incidence, only 25%-50% of women report fibroid-related symptoms. Fibroids are often first suspected when a mass is felt during a pelvic exam due to subserosal fibroids but can be confirmed by ultrasonography.

Therapies

Treatment for uterine fibroids (Table 126.1) depends on symptoms, anatomy (including size, number, and location), and the reproductive goals of the patient. There are several medical and surgical procedures that should be individualized to the patient. Expectant management for asymptomatic patients and for symptomatic patients for 6 months is another option.

Expectant Management. Women with asymptomatic fibroids usually undergo a "watch and wait" approach. While data suggest that fibroids can decrease in size, there is no known way to naturally trigger fibroid regression. However, symptomatic women may wish to be more engaged with their providers regarding treatment. Patient decision aids (PDAs) can help facilitate a discussion between women and their health care providers to decide on the best course of treatment. PDAs can provide all the information of different treatment options to women. A study is currently underway providing PDA for uterine fibroid diagnosis.[91]

Medical Therapies. Most medical treatments focus on improving heavy menstrual bleeding. Contraceptive steroids are widely used for this noncontraceptive indication. The most widely used off-label medical treatment is OCs, in one study accounting for almost 80% of medical therapy for fibroids.[92] The progestin-releasing intrauterine device (IUD) is another effective off-label contraceptive steroid to decrease menstrual bleeding, but risk of expulsion is increased in a fibroid

TABLE 126.1 Surgical and Medical Treatments of Uterine Fibroids and Adenomyosis

Treatments	Uterine Fibroids	Adenomyosis
Medical		
Oral contraceptives	**X** [92]	**X** [120]
Progestin-releasing intrauterine device (IUD)	**X** [93]	**X** [120]
Nonsteroidal antiinflammatory drugs (NSAIDS)	**X** [93]	**X** [120]
Gonadotropin-releasing hormone (GnRH) agonists	**X** [93]	**X** [120]
Selective progesterone receptor modulators (SPRMs)	**X** [93]	
Tranexamic acid	**X** [93]	
Surgical		
Hysterectomy	**X** [8]	**X**
Uterine artery embolization (UAE)	**X** [93]	**X** [121]
MRI-guided focused ultrasound surgery (MRgFUS)	**X** [93]	**X**
Myomectomy	**X** [93]	
Adenomyomectomy		**X** [115]
Radiofrequency ablation (RFA)	**X** [93]	

uterus. Nonsteroidal antiinflammatory drugs (NSAIDs) are used to decrease blood loss and treat menstrual pain as a result of fibroids, but high-quality evidence is lacking.[93]

There is moderate consistent evidence that gonadotropin-releasing hormone (GnRH) agonists decrease AUB-L, improve quality of life, and shrink fibroids.[93] GnRH agonist is thus used as primary treatment *and* as an adjuvant preoperative medical therapy, where either the absence of menses or volume reduction will greatly aid surgical treatment. The agonist depot leuprolide acetate is approved for preoperative therapy. GnRH agonists first lead to release of stored pituitary gonadotropins, producing a clinical flare, and then to castrate levels of gonadal steroids, thus placing women at risk for hypoestrogenic symptoms, including bone loss. It is these side effects that typically limit therapy to 6 months of use. A new generation of oral GnRH antagonists, in combination with steroidal add-back to limit hypoestrogenic side effects, is showing both safety and efficacy for fibroid treatment. Elagolix with add-back is now US Food and Drug Administration–approved,[94] and relugolix combination therapy is likely to be approved soon.[95]

Aromatase inhibitors act by competitively binding the aromatase enzyme and blocking its access to androgens and are used for off-label treatment.[96,97] The aromatase inhibitors currently on the market, such as anastrozole and letrozole, have been shown to result in greater than 98% inhibition of the enzyme.[97,98] The efficacy of letrozole and anastrozole has been established in multiple clinical trials, resulting in up to 52.5% reduction in fibroid volume and improved symptomatic control.[99-102] Importantly, letrozole has been shown to have similar efficacy as GnRH agonist in reducing fibroid size, but with the advantage of avoiding the profound hypoestrogenism caused by GnRH agonists and resultant side effects, although this is controversial.[100] Some studies suggest a more systemic effect of aromatase inhibitors, and others more localized effects.[100,103] Bone loss and increased cardiovascular risk, particularly in younger patients, are the most concerning potential consequences of prolonged hypoestrogenism from prolonged treatment with aromatase inhibitors.[98] As with all medical treatments for fibroids, once aromatase inhibitors are discontinued, fibroid growth rebounds.[100]

Selective PR modulators are also effective treatment options for uterine fibroids without the hypoestrogenic side effects of GnRH agonists.[93] Four of these agents—mifepristone (RU486), asoprisnil (J867), ulipristal acetate (CDB2914), and telapristone acetate (CDB4124)—reduce fibroid size and improve quality of life in women with uterine fibroids.[69,73,74,76,104-106] Unfortunately, severe liver toxicity led ulipristal to be significantly limited in the market internationally, and the others never received approval for treatment in the United States.[107]

Finally, tranexamic acid is an antifibrinolytic treatment that is used to decrease bleeding during surgery and during menses in women with hereditary bleeding disorders. In women with AUB-L undergoing surgical myomectomy, tranexamic acid as an off-label treatment reduced blood loss compared with placebo.[93] Likewise, it lessened AUB-L during menses, with mild adverse effects such as menstrual cramps, back pain, and sinus headaches. An advantage of tranexamic acid for AUB-L is that it only needs to be taken during the days of heavy menstrual bleeding, although increased risk of deep vein thrombosis can occur when combined with estrogen-containing contraceptive steroids.[108]

Surgical and Interventional Therapies. In contrast to medical therapies, most surgical and interventional therapies are reserved for women who have symptoms related to the size of the fibroids, or so-called bulk symptoms, those with a combination of AUB-L and bulk symptoms, or those who have failed medical therapies. However, new fibroids commonly form following conservative treatment, often incorrectly labeled as fibroid recurrence. Candidates for surgical and interventional therapy should be counseled about this risk and the risk of adhesion formation following surgery, which can impact fertility and pain.

The one surgical therapy that is often performed in women with isolated AUB-L or in asymptomatic women seeking fertility is hysteroscopic myomectomy, which only treats submucosal fibroids. It is performed on an outpatient basis with an intrauterine endoscope, and patients can typically return to work in less than a week.[3] Recent findings suggest that women have relief of some of their bulk symptoms and improvement of fertility with this approach.[109]

The most common procedure for treating uterine fibroids is hysterectomy, accounting for approximately 79.2% of fibroid procedural therapies.[8] In the United States, over 90% of hysterectomies are performed for noncancerous indications, and fibroids account for one half to three quarters of all hysterectomies.[3] The advantage of hysterectomy is that it prevents new fibroid formation and improves the decreased quality of life many women with fibroids experience. Women undergoing hysterectomy for uterine fibroids typically can retain their ovaries.

However, there is a growing literature that details long-term effects associated with hysterectomy even with bilateral ovarian conservation, including significantly increased rates of hyperlipidemia, hypertension, obesity, cardiac arrhythmias, congestive heart failure, and coronary artery disease.[110] Furthermore, many women who have a hysterectomy report increased *de novo* mental health concerns like depression and anxiety.[111]

Myomectomy is a uterine-preserving technique used to surgically remove uterine fibroids. There is low quality of evidence on myomectomy improving quality of life, and insufficient evidence on resolving AUB-L.[93] Depending on fibroid size and location, as well as surgical skill, myomectomy can be performed laparoscopically or robotically, with a shorter recovery time.[112] Abdominal myomectomy is more invasive but treats more extensive disease such as fibroids larger than 9 cm or if there are more than three to four. This method does have a longer recovery time compared to the laparoscopic technique,[113] and there is a risk that uterine fibroids may reoccur.

Radiofrequency ablation uses radiofrequency energy to shrink uterine fibroids via a laparoscopic or hysteroscopic approach. Small comparative trials have shown advantages compared with the excisional approaches, although data are limited.[93]

Uterine artery embolization (UAE) is a uterine-preserving procedure with strong and consistent evidence for improving quality of life for women with fibroids and reducing fibroid volume.[93] UAE blocks blood flow from uterine arteries, causing fibroid shrinkage, and is performed under sedation and local anesthesia, with short recovery.[4] It has the same success and quality-of-life improvement as myomectomy and hysterectomy, but has a higher reintervention rate than hysterectomy. Based on limited evidence, it is not considered first-line therapy for women who wish future pregnancy.

MRI-guided focused ultrasound surgery (MRgFUS) is a noninvasive procedure available in the United States that directs ultrasound ablation of uterine fibroids with the guidance of MRI, with low-level evidence for decreasing fibroid size.[93] A study on MRgFUS efficacy reported that women who underwent the procedure experienced decreased bleeding ($P = 0.001$) and pelvic symptoms ($P = 0.004$). There were no serious adverse effects, and two pregnancies were reported during follow-up.[114] There are also devices for ultrasound-guided focused ultrasound that are available primarily in China.[93]

KEY POINTS

- Treatment for fibroids depends on symptoms, anatomy including size, number, and location, and the reproductive goals of the patient. Many medical treatments focus on improving AUB-L, but there are a few that have shown evidence in decreasing fibroid size. Most surgical therapies are reserved for women who have bulk symptoms, the combination of AUB-L and bulk symptoms, or those who have failed medical therapies. The most common procedure being hysterectomy, accounting for approximately 79.2% of fibroid procedural therapies.

ADENOMYOSIS

Overview

Adenomyosis is a disease that commonly coexists with uterine fibroids and endometriosis and is a cause of heavy menstrual bleeding, pelvic pain, and reproductive dysfunction. Until the past decade, adenomyosis was typically only identified in women who were undergoing hysterectomy. The recognition that this disease is able to be identified by both ultrasound and MRI is changing the epidemiology of the disease and raising the question of whether symptoms previously attributed to endometriosis and fibroids are actually due to coexisting adenomyosis.

Epidemiology and Classification

Adenomyosis is a disorder where endometrial glands and stroma grow within the myometrium. There is resulting hyperplasia of the normal myometrial tissue, which thickens, causing an enlarged, boggy, globular uterus. It has a prevalence of 20% of women.[115] Symptoms include painful menses, heavy menstrual bleeding, pelvic pain, and possible infertility. These symptoms are similar to other gynecological conditions and make understanding adenomyosis difficult, because many women have endometriosis and/or uterine fibroids with adenomyosis.

Adenomyosis comes in several forms. Diffuse adenomyosis has endometrial implants scattered throughout the myometrium. Focal adenomyosis or an adenomyoma is a spherical collection that resembles a fibroid. Thickening or irregularity of the junctional zone is also a type of adenomyosis. Stated risk factors for adenomyosis include multiparity, older reproductive age, more miscarriages, abortions, and more uterine surgery. However, because adenomyosis is only diagnosed by hysterectomy, these risk factors may instead reflect the epidemiology of hysterectomy.[116]

KEY POINTS

- Adenomyosis is a disorder where endometrial glands and stroma grow into the myometrium, causing subfertility, pelvic pain, dysmenorrhea and heavy menstrual bleeding. It affects ~20% women, and can be diffuse or focal.

Pathogenesis

The specific pathogenesis of adenomyosis is unknown, but two prevailing theories exist (Fig. 126.4). The first is that adenomyosis is caused by invagination of the endometrial basalis into the myometrium, potentially due to a disrupted junctional zone.[117,118] The migration of the endometrium is thought to be supported by local hyperestrogenism, disordered uterine peristalsis leading to microtrauma, and activation of tissue injury and repair mechanism with proliferation and inflammation. The second commonly held theory is that displaced embryonic remnants or adult stem cells undergo metaplasia within the myometrium. A related theory, referred to as "outside-to-inside invasion," hypothesizes that retrograde menstruation enables ectopic endometrial cells to invade the myometrium.[117-119] Given the diversity in adenomyosis phenotypes observed, it is quite possible that a single theory cannot explain all subtypes.[119] Regardless of the inciting event, progression of adenomyosis has been hypothesized to be related to sex steroid hormone levels, genetic alterations, inflammatory cells, changes in the ECM, growth factors, and/or gain in migration of adenomyotic cells.[117,118]

KEY POINTS

- The pathogenesis of adenomyosis is unknown, but there are two prevailing theories: 1) A disrupted junctional zone leads to invagination of the endometrial basalis into the myometrium, or 2) displaced embryonic remnants or adult stem cells undergo metaplasia withing the myometrium.

Diagnosis

Transvaginal ultrasound can be used to diagnose adenomyosis, but is more user-dependent. Key signs include asymmetry of uterine walls, heterogeneity or cystic spaces within the myometrium, thickening or discontinuity of endometrial/myometrial border, and linear striation radiating out from the endometrium. Color-flow doppler can differentiate adenomyosis from symptomatic uterine fibroids by looking at vascular flow. MRI is used to confirm the diagnosis if the ultrasound was not clear. However, the only definitive diagnosis is by histological verification of endometrial islets within the myometrium. There is still an ambiguous classification for the different subtypes of adenomyosis.[116]

KEY POINTS

- Until the past decade, adenomyosis was typically only identified in women who were undergoing hysterectomy. The recognition that this disease is able to be identified by both ultrasound and MR imaging is changing the epidemiology of the disease, and raising the question of whether symptoms previously attributed to endometriosis and fibroids are actually due to coexisting adenomyosis.

Treatment

There is no medical therapy specifically for the treatment of adenomyosis[120]; thus, medical treatments used to treat uterine fibroids and endometriosis are often appropriated for adenomyosis (Table 126.1).

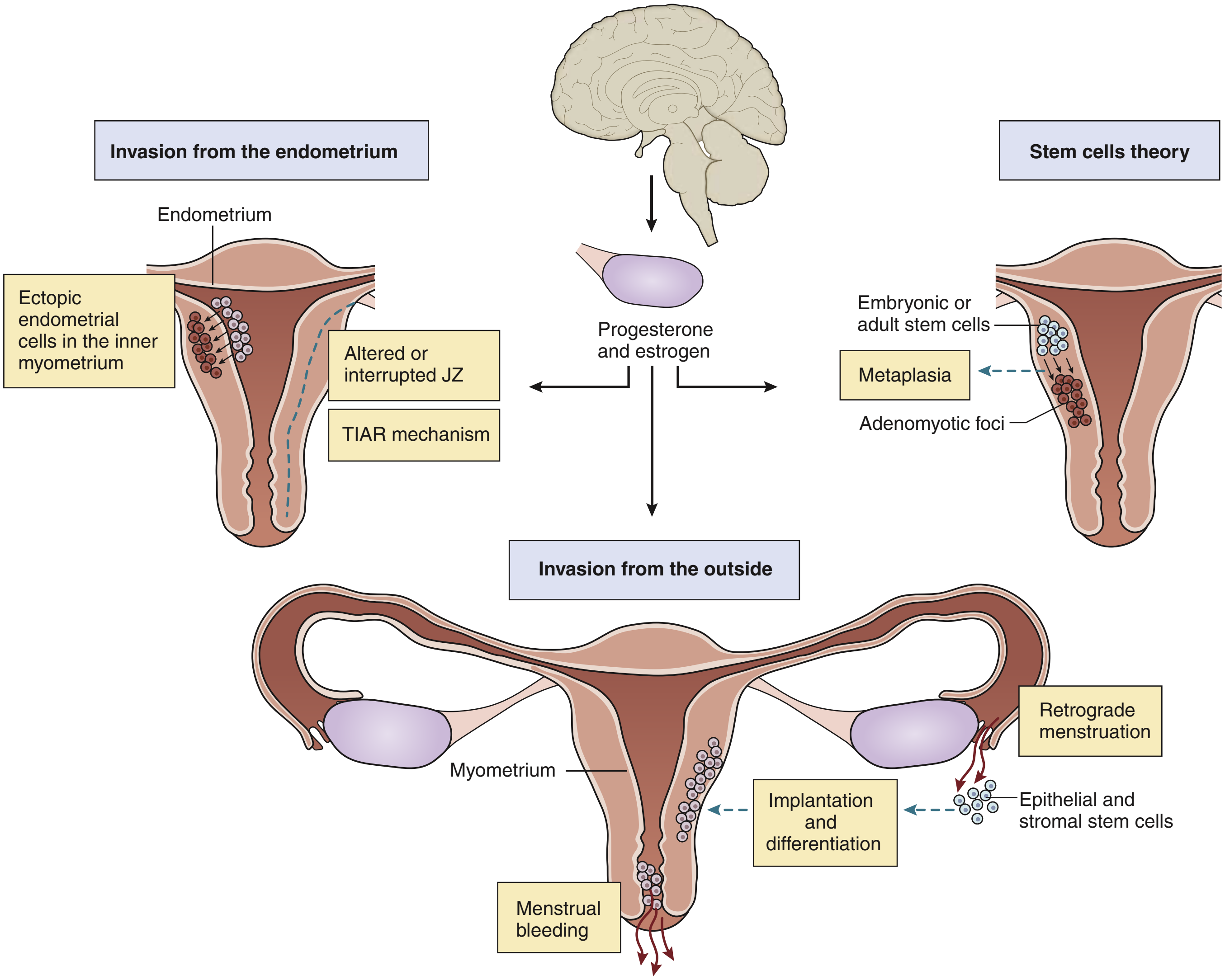

Invasion from the endometrium
Endometrium
Ectopic endometrial cells in the inner myometrium
Altered or interrupted JZ
TIAR mechanism
Progesterone and estrogen
Stem cells theory
Embryonic or adult stem cells
Metaplasia
Adenomyotic foci
Invasion from the outside
Myometrium
Retrograde menstruation
Epithelial and stromal stem cells
Implantation and differentiation
Menstrual bleeding

Figure 126.4 **Pathophysiology of adenomyosis.** Adenomyosis pathogenesis appears to be influenced both directly by the brain through peptides such as prolactin and oxytocin and indirectly through the gonadal steroids estrogen and progesterone. There are several theories of adenomyosis pathogenesis. **A,** The invasion hypothesis, where endometrium traverses an altered or interrupted junctional zone (JZ) leading to tissue injury and repair (TIAR). **B,** Metaplasia of resident stem cells from within the myometrium. **C,** Migration and differentiation of adult stem cells from the endometrium following retrograde menstruation. (Adapted with permission from Chapron, et al. *Hum Repro Update.* 2020;26:392–411. The following copyright applies to the content of this slide: © The Author(s) 2020. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com. This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model (https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model)

NSAIDs, OCs, and especially progestin-releasing IUDs can be used as a therapy for pain relief and abnormal uterine bleeding (AUB-A). Progestins are known to decrease the expression of the PR in the adenomyotic lesions in the myometrium. GnRH analogs are used to stop proliferation within the myometrium, and may induce apoptosis and reduce inflammation.

Surgical treatment varies. Hysterectomy is currently the only surgical procedure that permanently resolves symptoms of adenomyosis, but it is clearly not acceptable to many women. However, there is an increasing variety of other conservative procedures that can be done to improve some symptoms. Adenomyomectomy is a cytoreductive surgery that removes adenomyotic nodules, but has been associated with early uterine rupture during pregnancy, even with expert surgeons.[115] UAE and MRgFUS have been shown to improve symptoms of adenomyosis in some studies.[121] However, there is no comparative effectiveness research for uterine-sparing treatments for adenomyosis.[122]

KEY POINTS

- There is no medical therapy specifically for adenomyosis, but treatments for fibroids and endometriosis are often employed with mixed results. Less invasive surgical procedures, such as adenomyomectomy, can be attempted but the only definitive treatment remains hysterectomy.

Summary

Uterine fibroids are extremely common and are the leading indication for hysterectomy and a cause of significant morbidity and health care expenditures for reproductive-age women. For reasons that remain to be fully elucidated, fibroids are more common and more severe in women of African descent. The discovery of stem cells and genetic mutations in leiomyoma may lead to a better understanding of their pathophysiology and the development of new, potentially individualized treatments. A salient feature of fibroids is their responsiveness to steroid hormones, and most medical treatments, both established and experimental, target steroid action. There are also surgical and radiological treatment options for fibroids. Adenomyosis is a disease that often coexists with uterine fibroids and endometriosis and can cause heavy menstrual bleeding, pelvic pain, and subfertility. The exact pathophysiology of adenomyosis is unknown, but several theories exist. As such, there is currently no medical therapy specifically for adenomyosis, but many of the treatments for fibroids and/or endometriosis have been employed with varying success. The only definitive treatment for adenomyosis remains hysterectomy.

REFERENCES

1. Stewart EA. Epidemiology of uterine fibroids: a systematic review. *BJOG*. 2017;124:1501–1512.
2. Baird D. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. *Am J Obstet Gynecol*. 2003;188:100–107.
3. Stewart EA. Clinical practice. *Uterine Fibroids N Engl J Med*. 2015;372:1646–1655.
4. Lumsden MA. Fibroids: diagnosis and management. *BMJ*. 2015;351:4887.
5. Ghant M. Beyond the physical: a qualitative assessment of the burden of symptomatic uterine fibroids on women's emotional and psychosocial health. *J Psychosom Res*. 2015;78:499–503.
6. Eltoukhi HM. The health disparities of uterine fibroid tumors for African American women: a public health issue. *Am J Obstet Gynecol*. 2014;210:194–199.
7. Marsh EE. Prevalence, and treatment of uterine fibroids: a survey of U.S. women. *J Womens Health Larchmt*. 2018;27:1359–1367.
8. Wise LA, Laughlin-Tommaso SK. Epidemiology of uterine fibroids: from menarche to menopause. *Clin Obstet Gynecol*. 2016;59:2–24.
9. Wise LA. Polymorphisms in vitamin D-related genes and risk of uterine leiomyomata. *Fertil Steril*. 2014;102:503–510e1.
10. Munro MG. The two FIGO systems for normal and abnormal uterine bleeding symptoms and classification of causes of abnormal uterine bleeding in the reproductive years: 2018 revisions. *Int J Gynaecol Obstet*. 2018;143:393–408.
11. Laughlin-Tommaso SK. Clinical limitations of the International federation of Gynecology and Obstetrics (FIGO) classification of uterine fibroids. *Int J Gynaecol Obstet*. 2017;139:143–148.
12. Al-Hendy A, Myers ER, Stewart E. Uterine fibroids: burden and unmet medical need. *Semin Reprod Med*. 2017;35:473–480.
13. Makinen N, Mehine M, Tolvanen J. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science*. 2011;334:252–255.
14. McGuire MM, Yatsenko A, Hoffner L, et al. Whole exome sequencing in a random sample of North American women with leiomyomas identifies MED12 mutations in majority of uterine leiomyomas. *PLoS One*. 2012;7:e33251.
15. Bertsch E, Qiang W, Zhang Q. MED12 and HMGA2 mutations: two independent genetic events in uterine leiomyoma and leiomyosarcoma. *Mod Pathol Off J US Can Acad Pathol*. 2014;27(8):1144–1153.
16. Kim JJ, Kurita T, Bulun SE. Progesterone action in endometrial cancer, endometriosis, uterine fibroids, and breast cancer. *Endocr Rev*. 2013;34:130–162.
17. Flake GP, Moore AB, Sutton D. The natural history of uterine leiomyomas: light and electron microscopic studies of fibroid phases, interstitial ischemia, inanosis, and reclamation. *Obstet Gynecol Int*. 2013;2013:528376.
18. Ono M, Qiang W, Serna VA. Role of stem cells in human uterine leiomyoma growth. *PLoS One*. 2012;7:e36935.
19. Mas A, Cervello I, Gil-Sanchis C. Identification and characterization of the human leiomyoma side population as putative tumor-initiating cells. *Fertil Steril*. 2012;98:741–751.
20. Weissman IL. Stem cells: units of development, units of regeneration, and units in evolution. *Cell*. 2000;100:157–168.
21. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. *Science*. 2010;327:542–555.
22. Jordan CT, Guzman ML, Noble M. Cancer stem cells. *N Engl J Med*. 2006;355:1253–1261.
23. Schofield R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cell*. 1978;4:7–25.
24. Ono M, Maruyama T, Masuda H. Side population in human uterine myometrium displays phenotypic and functional characteristics of myometrial stem cells. *Proc Natl Acad Sci Unit States Am*. 2007;104:18700–18705.
25. Szotek PP, Chang HL, Zhang L. Adult mouse myometrial label-retaining cells divide in response to gonadotropin stimulation. *Stem Cell*. 2007;25:1317–1325.
26. Arango NA, Szotek PP, Manganaro TF, et al. Conditional deletion of beta-catenin in the mesenchyme of the developing mouse uterus results in a switch to adipogenesis in the myometrium. *Dev Biol*. 2005;288:276–283.
27. Zhou S, Yi T, Shen K, et al. Hypoxia: the driving force of uterine myometrial stem cells differentiation into leiomyoma cells. *Med Hypotheses*. 2011;77:985–986.
28. Maruyama T, Miyazaki K, Masuda H, et al. Review: human uterine stem/progenitor cells: implications for uterine physiology and pathology. *Placenta*. 2013:34.
29. Ono M, Yin P, Navarro A. Paracrine activation of WNT/beta-catenin pathway in uterine leiomyoma stem cells promotes tumor growth. *Proc Natl Acad Sci Unit States Am*. 2013;110:17053–17058.
30. Moravek MB, Yin P, Coon JS, et al. Paracrine pathways in uterine leiomyoma stem cells involve insulinlike growth factor 2 and insulin receptor A. *J Clin Endocrinol Metab*. 2017;102:1588–1595.
31. Ikhena DE, Liu S, Kujawa S, et al. RANKL/RANK pathway and its inhibitor RANK-Fc in uterine leiomyoma growth. *J Clin Endocrinol Metab*. 2018;103:1842–1849.
32. Pedeutour F, Quade BJ, Weremowicz S, et al. Localization and expression of the human estrogen receptor beta gene in uterine leiomyomata. *Genes Chromosomes Cancer*. 1998;23:361–366.
33. Benassayag C, Leroy MJ, Rigourd V. Estrogen receptors (ERalpha/ERbeta) in normal and pathological growth of the human myometrium: pregnancy and leiomyoma. *Am J Physiol*. 1999;276:E1112–E1118.
34. Andersen J, Barbieri RL. Abnormal gene expression in uterine leiomyomas. *J Soc Gynecol Investig*. 1995;2:663–672.
35. Andersen J, DyReyes VM, Barbieri RL, et al. Leiomyoma primary cultures have elevated transcriptional response to estrogen compared with autologous myometrial cultures. *J Soc Gynecol Investig*. 1995;2:542–551.
36. Maruo T, Ohara N, Wang J, et al. Sex steroidal regulation of uterine leiomyoma growth and apoptosis. *Hum Reprod Update*. 2004;10:207–220.
37. Li S, McLachlan JA. Estrogen-associated genes in uterine leiomyoma. *Ann NY Acad Sci*. 2001;948:112–120.
38. Tamaya T, Fujimoto J, Okada H. Comparison of cellular levels of steroid receptors in uterine leiomyoma and myometrium. *Acta Obstet Gynecol Scand*. 1985;64:307–309.
39. Wilson EA, Yang F, Rees ED. Estradiol and progesterone binding in uterine leiomyomata and in normal uterine tissues. *Obstet Gynecol*. 1980;55:20–24.
40. Jakimiuk AJ, Bogusiewicz M, Tarkowski R. Estrogen receptor alpha and beta expression in uterine leiomyomas from premenopausal women. *Fertil Steril*. 2004;3:1244–1249.
41. Feng Y, Lin X, Zhou S, et al. The associations between the polymorphisms of the ER-alpha gene and the risk of uterine leiomyoma (ULM). *Tumour Biol J Int Soc Oncodevelopmental Biol Med*. 2013;34:3077–3082.
42. Al-Hendy A, Salama SA. Ethnic distribution of estrogen receptor-alpha polymorphism is associated with a higher prevalence of uterine leiomyomas in black Americans. *Fertil Steril*. 2006;86:686–693.
43. Hermon TL, Moore AB, Yu L, et al. Estrogen receptor alpha (ERalpha) phospho-serine-118 is highly expressed in human uterine leiomyomas compared to matched myometrium. *Virchows Arch Int J Pathol*. 2008;453:557–569.
44. Ishikawa H, Ishi K, Serna VA, et al. Progesterone is essential for maintenance and growth of uterine leiomyoma. *Endocrinology*. 2010;151:2433–2442.

45. Ishikawa H, Fenkci V, Marsh EE. CCAAT/enhancer binding protein beta regulates aromatase expression via multiple and novel cis-regulatory sequences in uterine leiomyoma. *J Clin Endocrinol Metab*. 2008;93:981–991.
46. Lamminen S, Rantala I, Helin H, et al. Proliferative activity of human uterine leiomyoma cells as measured by automatic image analysis. *Gynecol Obstet Invest*. 1992;34:111–114.
47. Hassan MH, Salama SA, Arafa HM, et al. Adenovirus-mediated delivery of a dominant-negative estrogen receptor gene in uterine leiomyoma cells abrogates estrogen- and progesterone-regulated gene expression. *J Clin Endocrinol Metab*. 2007;92:3949–3957.
48. Bulun SE. Uterine fibroids. *N Engl J Med*. 2013;369:1344–1355.
49. Bulun SE, Simpson ER, Word RA. Expression of the CYP19 gene and its product aromatase cytochrome P450 in human uterine leiomyoma tissues and cells in culture. *J Clin Endocrinol Metab*. 1994;78:736–743.
50. Sumitani H, Shozu M, Segawa T. In situ estrogen synthesized by aromatase P450 in uterine leiomyoma cells promotes cell growth probably via an autocrine/intracrine mechanism. *Endocrinology*. 2000;141:3852–3861.
51. Folkerd EJ, Newton CJ, Davidson K, et al. Aromatase activity in uterine leiomyomata. *J Steroid Biochem*. 1984;20:1195–1200.
52. Shozu M, Murakami K, Inoue M. Aromatase and leiomyoma of the uterus. *Semin Reprod Med*. 2004;22:51–60.
53. Imir AG, Lin Z, Yin P. Aromatase expression in uterine leiomyomata is regulated primarily by proximal promoters I.3/II. *J Clin Endocrinol Metab*. 2007;92:1979–1982.
54. Cermik D, Arici A, Taylor HS. Coordinated regulation of HOX gene expression in myometrium and uterine leiomyoma. *Fertil Steril*. 2002;78:979–984.
55. Mangelsdorf DJ, Thummel C, Beato M. The nuclear receptor superfamily: the second decade. *Cell*. 1995;83:835–839.
56. Robinson-Rechavi M, Escriva Garcia H, Laudet V. The nuclear receptor superfamily. *J Cell Sci*. 2003;116:585–586.
57. Lessey BA, Alexander PS, Horwitz KB. The subunit structure of human breast cancer progesterone receptors: characterization by chromatography and photoaffinity labeling. *Endocrinology*. 1983;112:1267–1274.
58. Kastner P, Krust A, Turcotte B. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J*. 1990;9:1603–1614.
59. Gronemeyer H, Meyer ME, Bocquel MT, et al. Progestin receptors: isoforms and antihormone action. *J Steroid Biochem Mol Biol*. 1991;40:271–278.
60. Vegeto E, Shahbaz MM, Wen DX, et al. Human progesterone receptor A form is a cell- and promoter-specific repressor of human progesterone receptor B function. *Mol Endocrinol*. 1993;7:1244–1255.
61. McDonnell DP, Shahbaz MM, Vegeto E, et al. The human progesterone receptor A-form functions as a transcriptional modulator of mineralocorticoid receptor transcriptional activity. *J Steroid Biochem Mol Biol*. 1994;48:425–432.
62. Mulac-Jericevic B, Lydon JP, DeMayo FJ, et al. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc Natl Acad Sci Unit States Am*. 2003;100:9744–9749.
63. Mulac-Jericevic B, Mullinax RA, DeMayo FJ, et al. Subgroup of reproductive functions of progesterone mediated by progesterone receptor-B isoform. *Science*. 2000;289:1751–1754.
64. Brandon DD, Bethea CL, Strawn EY. Progesterone receptor messenger ribonucleic acid and protein are overexpressed in human uterine leiomyomas. *Am J Obstet Gynecol*. 1993;169:78–85.
65. Englund K, Blanck A, Gustavsson I. Sex steroid receptors in human myometrium and fibroids: changes during the menstrual cycle and gonadotropin-releasing hormone treatment. *J Clin Endocrinol Metab*. 1998;83:4092–4096.
66. Nisolle M, Gillerot S, Casanas-Roux F, et al. Immunohistochemical study of the proliferation index, oestrogen receptors and progesterone receptors A and B in leiomyomata and normal myometrium during the menstrual cycle and under gonadotrophin-releasing hormone agonist therapy. *Hum Reprod*. 1999;14:2844–2850.
67. Kawaguchi K, Fujii S, Konishi I, et al. Mitotic activity in uterine leiomyomas during the menstrual cycle. *Am J Obstet Gynecol*. 1989;160:637–641.
68. Chen W, Ohara N, Wang J. A novel selective progesterone receptor modulator asoprisnil (J867) inhibits proliferation and induces apoptosis in cultured human uterine leiomyoma cells in the absence of comparable effects on myometrial cells. *J Clin Endocrinol Metab*. 2006;91:1296–1304.
69. Luo X, Yin P, Coon VJ, et al. The selective progesterone receptor modulator CDB4124 inhibits proliferation and induces apoptosis in uterine leiomyoma cells. *Fertil Steril*. 2010;93:2668–2673.
70. Maruo T, Ohara N, Matsuo H. Effects of levonorgestrel-releasing IUS and progesterone receptor modulator PRM CDB-2914 on uterine leiomyomas. *Contraception*. 2007;75:S99–S103.
71. Wang J, Ohara N, Wang Z. A novel selective progesterone receptor modulator asoprisnil (J867) down-regulates the expression of EGF, IGF-I, TGFbeta3 and their receptors in cultured uterine leiomyoma cells. *Hum Reprod*. 2006;21:1869–1877.
72. Xu Q, Ohara N, Liu J. Progesterone receptor modulator CDB-2914 induces extracellular matrix metalloproteinase inducer in cultured human uterine leiomyoma cells. *Mol Hum Reprod*. 2008;14:181–191.
73. Xu Q, Ohara N, Liu J. Selective progesterone receptor modulator asoprisnil induces endoplasmic reticulum stress in cultured human uterine leiomyoma cells. *Am J Physiol Endocrinol Metab*. 2007;293:E1002–E1011.
74. Sasaki H, Ohara N, Xu Q. A novel selective progesterone receptor modulator asoprisnil activates tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-mediated signaling pathway in cultured human uterine leiomyoma cells in the absence of comparable effects on myometrial cells. *J Clin Endocrinol Metab*. 2007;92:616–623.
75. Xu Q, Takekida S, Ohara N. Progesterone receptor modulator CDB-2914 down-regulates proliferative cell nuclear antigen and Bcl-2 protein expression and up-regulates caspase-3 and poly(adenosine 5'-diphosphate-ribose) polymerase expression in cultured human uterine leiomyoma cells. *J Clin Endocrinol Metab*. 2005;90:953–961.
76. Morikawa A, Ohara N, Xu Q. Selective progesterone receptor modulator asoprisnil down-regulates collagen synthesis in cultured human uterine leiomyoma cells through up-regulating extracellular matrix metalloproteinase inducer. *Hum Reprod*. 2008;23:944–951.
77. Qiang W, Liu Z, Serna VA. Downregulation of miR-29b is essential for pathogenesis of uterine leiomyoma. *Endocrinology*. 2014.
78. Qin W, Chung AC, Huang XR. TGF-beta/Smad3 signaling promotes renal fibrosis by inhibiting miR-29. *J Am Soc Nephrol*. 2011;22:1462–1474.
79. Cushing L, Kuang PP, Qian J. miR-29 is a major regulator of genes associated with pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2011;45:287–294.
80. Roderburg C, Urban GW, Bettermann K. Micro-RNA profiling reveals a role for miR-29 in human and murine liver fibrosis. *Hepatology*. 2011;53:209–218.
81. E R, LB S, JE T. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc Natl Acad Sci Unit States Am*. 2008:13027–13032.
82. Yin P, Lin Z, Reierstad S. Transcription factor KLF11 integrates progesterone receptor signaling and proliferation in uterine leiomyoma cells. *Cancer Res*. 2010;70:1722–1730.
83. Yin P, Roqueiro D, Huang L. Genome-wide progesterone receptor binding: cell type-specific and shared mechanisms in T47D breast cancer cells and primary leiomyoma cells. *PLoS One*. 2012;7:e29021.
84. Chang BH, Li L, Paul A. Protection against fatty liver but normal adipogenesis in mice lacking adipose differentiation-related protein. *Mol Cell Biol*. 2006;26:1063–1076.
85. Lee TF, Mak KM, Rackovsky O. Downregulation of hepatic stellate cell activation by retinol and palmitate mediated by adipose differentiation-related protein (ADRP). *J Cell Physiol*. 2010;223:648–657.
86. Chong BM, Reigan P, Mayle-Combs KD, et al. Determinants of adipophilin function in milk lipid formation and secretion. *Trends Endocrinol Metab*. 2011;22:211–217.
87. Yao M, Huang Y, Shioi K. Expression of adipose differentiation-related protein: a predictor of cancer-specific survival in clear cell renal carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2007;13:152–160.
88. Islam MS, Ciavattini A, Petraglia F, et al. Extracellular matrix in uterine leiomyoma pathogenesis: a potential target for future therapeutics. *Hum Reprod Update*. 2018;24:59–85.

89. Rafique S, Segars J, Leppert P. Mechanical signaling and extracellular matrix in uterine fibroids. *Semin Reprod Med*. 2017;35:487–493.
90. Shwayder J, Sakhel K. Imaging for uterine myomas and adenomyosis. *J Minim Invasive Gynecol*. 2014;21:362–376.
91. Scalia P. Implementation of the uterine fibroids Option Grid patient decision aids across five organizational settings: a randomized stepped-wedge study protocol. *Implement Sci*. 2019;14:88.
92. Yao X, Stewart EA, Laughlin-Tommaso SK, et al. Medical therapies for heavy menstrual bleeding in women with uterine fibroids: a retrospective analysis of a large commercially insured population in the USA. *Int J Obstet Gynaecol*. 2017;124:322–330.
93. Hartmann KE. Management of uterine fibroids. *Comp Eff Rev*. 2017:195.
94. Schlaff WD. Elagolix for heavy menstrual bleeding in women with uterine fibroids. *N Engl J Med*. 2020;382:328–340.
95. Al-Hendy A, Lukes AS, Poindexter AN, et al. Treatment of uterine fibroid symptoms with relugolix combination therapy. *N Engl J Med*. 2021;384:630–642.
96. Michaud LB, Buzdar AU. Risks and benefits of aromatase inhibitors in postmenopausal breast cancer. *Drug Saf Int J Med Toxicol Drug Exp*. 1999;21:297–309.
97. Chumsri S, Howes T, Bao T, et al. Aromatase, aromatase inhibitors, and breast cancer. *J Steroid Biochem Mol Biol*. 2011;125:13–22.
98. Lonning PE, Eikesdal HP. Aromatase inhibition 2013: clinical state of the art and questions that remain to be solved. *Endocr Relat Cancer*. 2013;20:R183–R201.
99. Parsanezhad ME, Azmoon M, Alborzi S. A randomized, controlled clinical trial comparing the effects of aromatase inhibitor (letrozole) and gonadotropin-releasing hormone agonist (triptorelin) on uterine leiomyoma volume and hormonal status. *Fertil Steril*. 2010;93:192–198.
100. Duhan N, Madaan S, Sen J. Role of the aromatase inhibitor letrozole in the management of uterine leiomyomas in premenopausal women. *Eur J Obstet Gynecol Reprod Biol*. 2013;171:329–332.
101. Hilario SG, Bozzini N, Borsari R, et al. Action of aromatase inhibitor for treatment of uterine leiomyoma in perimenopausal patients. *Fertil Steril*. 2009;91:240–243.
102. Varelas FK, Papanicolaou AN, Vavatsi-Christaki N, et al. The effect of anastrazole on symptomatic uterine leiomyomata. *Obstet Gynecol*. 2007;110:643–649.
103. Shozu M, Murakami K, Segawa T, et al. Successful treatment of a symptomatic uterine leiomyoma in a perimenopausal woman with a nonsteroidal aromatase inhibitor. *Fertil Steril*. 2003;79:628–631.
104. Bouchard P, Chabbert-Buffet N, Fauser BC. Selective progesterone receptor modulators in reproductive medicine: pharmacology, clinical efficacy and safety. *Fertil Steril*. 2011;96:1175–1189.
105. Donnez J, Tomaszewski J, Vazquez F. Ulipristal acetate versus leuprolide acetate for uterine fibroids. *N Engl J Med*. 2012;366:421–432.
106. Talaulikar VS, Manyonda IT. Ulipristal acetate: a novel option for the medical management of symptomatic uterine fibroids. *Adv Ther*. 2012;29:655–663.
107. Mahase E. Uterine fibroid drug is recalled after case of liver failure requiring transplant prompts EU review. *BMJ*. 2020;368:1112.
108. Bradley LD, Gueye NA. The medical management of abnormal uterine bleeding in reproductive-aged women. *Am J Obstet Gynecol*. 2016;214:31–44.
109. Nicholson WK. Short-term health-related quality of life after hysterectomy compared with myomectomy for symptomatic leiomyomas. *Obstet Gynecol*. 2019;134:261–269.
110. Laughlin-Tommaso SK, Cardiovascular and metabolic morbidity after hysterectomy with ovarian conservation: a cohort study. *Menopause*. 2018;25:483–492.
111. Laughlin-Tommaso S, Long-term risk of de novo mental health conditions after hysterectomy with ovarian conservation: a cohort study. *Menopause*. 2019;26:1395–1404.
112. Buckley VA. Laparoscopic myomectomy: clinical outcomes and comparative evidence. *J Minim Invasive Gynecol*. 2015;22:11–25.
113. Owen C, Armstrong AY. Clinical management of leiomyoma. *Obstet Gynecol Clin North Am*. 2015;42:67–85.
114. Thiburce AC. Magnetic resonance-guided high-intensity focused ultrasound for uterine fibroids: mid-term outcomes of 36 patients treated with the Sonalleve system. *Int J Hyperth*. 2015;31:764–770.
115. Donnez J, Donnez O, Dolmans MM. Introduction: uterine adenomyosis, another enigmatic disease of our time. *Fertil Steril*. 2018;109:369–370.
116. Gordts S, Grimbizis G, Campo R. Symptoms and classification of uterine adenomyosis, including the place of hysteroscopy in diagnosis. *Fertil Steril*. 2018;109:380–388e1.
117. Vannuccini S, Petraglia F. Recent advances in understanding and managing adenomyosis. *F1000Research*. 2019;8:398–405.
118. García-Solares J, Donnez J, Donnez O, et al. Pathogenesis of uterine adenomyosis: invagination or metaplasia? *Fertil Steril*. 2018;109:371–379.
119. Guo S-W. The pathogenesis of adenomyosis vis-à-vis endometriosis. *J Clin Med*. 2020;9:485.
120. Vannuccini S. Role of medical therapy in the management of uterine adenomyosis. *Fertil Steril*. 2018;109:398–405.
121. Keung JJ, Spies JB, Caridi TM. Uterine artery embolization: a review of current concepts. *Best Pract Res Clin Obstet Gynaecol*. 2018;46:66–73.
122. Osada H. Uterine adenomyosis and adenomyoma: the surgical approach. *Fertil Steril*. 2018;109:406–417.

127

Hormonal Control of the Breast*

Jose Russo[†]

OUTLINE

Hypothalamic-Pituitary Influences, Ovarian Function, and Breast Development, 2104
  Adolescence, 2104
The Mature Breast, 2105
  Nulliparous Women, 2106
  Parous Women, 2106
Distribution Pattern of Estrogen Receptor– And Progesterone Receptor–Positive and Proliferating Cells in the Mammary Gland, 2106
Stem Cells in the Normal Human Breast and Their Implications in Breast Cancer, 2108
  Genomic Signatures of Breast Changes After the First Full-Term Pregnancy in Premenopausal Women, 2110
The Menopausal Breast, 2112
  Genomic Differences in Nulliparous and Parous Breasts at Menopause, 2113
  Defining the Genomic Signature of Breast Cancer Prevention in Parous Women, 2113
  Chromatin Remodeling and Prevention of Human Breast Cancer, 2117

HYPOTHALAMIC-PITUITARY INFLUENCES, OVARIAN FUNCTION, AND BREAST DEVELOPMENT

KEY POINTS

- Breast development begins in the fetus with the mammary ridge and a rudimentary milk-duct system by birth.
- Main changes in mammary gland development are initiated at puberty and are dynamic across the life course and differ from woman to woman.
- Peri-pubertal changes in mammary duct elongation and branching are largely programmed by estradiol, progesterone, growth hormone (and prolactin) and growth factors.
- Age of clinical appearance of breast buds (thelarche) varies with race/ethnicity and BMI, with earlier thelarche onset in African American > Hispanic > White non-Hispanic > Asian girls) and with higher BMI.

In the nonpregnant female, development of the mammary gland is rigorously controlled by the ovary. Although puberty is often considered to be the point of initiation of ovarian function, the development of the ovary, in fact, is a gradual process (see Chapters 102, 117, and 120).[1] The complex regulatory mechanisms that control ovarian function involve gonadotropin-releasing hormone; luteinizing hormone (LH); follicle-stimulating hormone (FSH); the ovarian peptides inhibin, activin, and follistatin; growth hormone (GH); and prolactin (see Chapters 102 and 105).[2] The ovary is active during late fetal development and for a short time after birth but becomes relatively quiescent until the onset of puberty, when mammary ductal elongation and branching begin under the influence of estradiol (E2) and GH.[3] Although its exact mechanism is unclear, GH directly stimulates ductal growth in hypophysectomized-ovariectomized rats and may also act through its local mediator, insulin-like growth factor-1 (IGF-1). Normal duct development, however, requires the presence of E2 and progesterone acting, respectively, through the estrogen receptor (ER) and the progesterone receptor (PR), both of which are present in the mammary gland. E2 acts locally to stimulate DNA synthesis and promote bud formation, likely ER-mediated. Prolactin plays an additional role, but its specific actions are not fully understood. The response of the mammary gland to these complex hormonal and metabolic interactions results in developmental changes that permanently modify the architecture and biological characteristics of the gland.[3] The mammary gland, in turn, responds selectively to given hormonal stimuli, depending on specific topographic differences in gland development that modulate the expression of cell proliferation or differentiation. Hormone receptor knockout studies and tissue recombination and proteomic approaches have revealed sequential activation of hormone signaling in the breast epithelium, acting through coregulators, that govern morphogenesis and cell proliferation, depending on developmental stage (see Chapter 102).[5]

Adolescence

Although the main changes in the mammary gland are initiated at puberty, development continues during adulthood, and the final histomorphometry of the gland varies greatly from woman to woman.[3] Mammary gland development can be defined by the external appearance of the breast or by determination of mammary gland area, volume, degree of branching, or level of differentiation of the gland, such as lobule-type formation.[5]

The adolescent period begins with the first signs of sexual change at puberty and terminates with sexual maturity. Age at thelarche (initial clinical appearance of a breast bud) varies by race/ethnicity and BMI.[6,7] Median age at onset is 8.8, 9.3, 9.7, and 9.7 years for African American, Hispanic, White non-Hispanic, and Asian participants, respectively.[6] Girls with greater BMI reached breast stage 2 at younger ages, especially among girls with BMI >85th percentile.[7] With the approach of puberty, the rudimentary mammae begin to show growth activity in the glandular tissue and the surrounding stroma.[3-5] This glandular increase is as a result of the growth and division of small bundles of primary and secondary ducts (Fig. 127.1). These grow and divide in two ways. One

---

*This chapter is in memory of my long-time collaborator and companion, Dr. Irma H Russo, who passed away on June 25, 2013.

[†]deceased

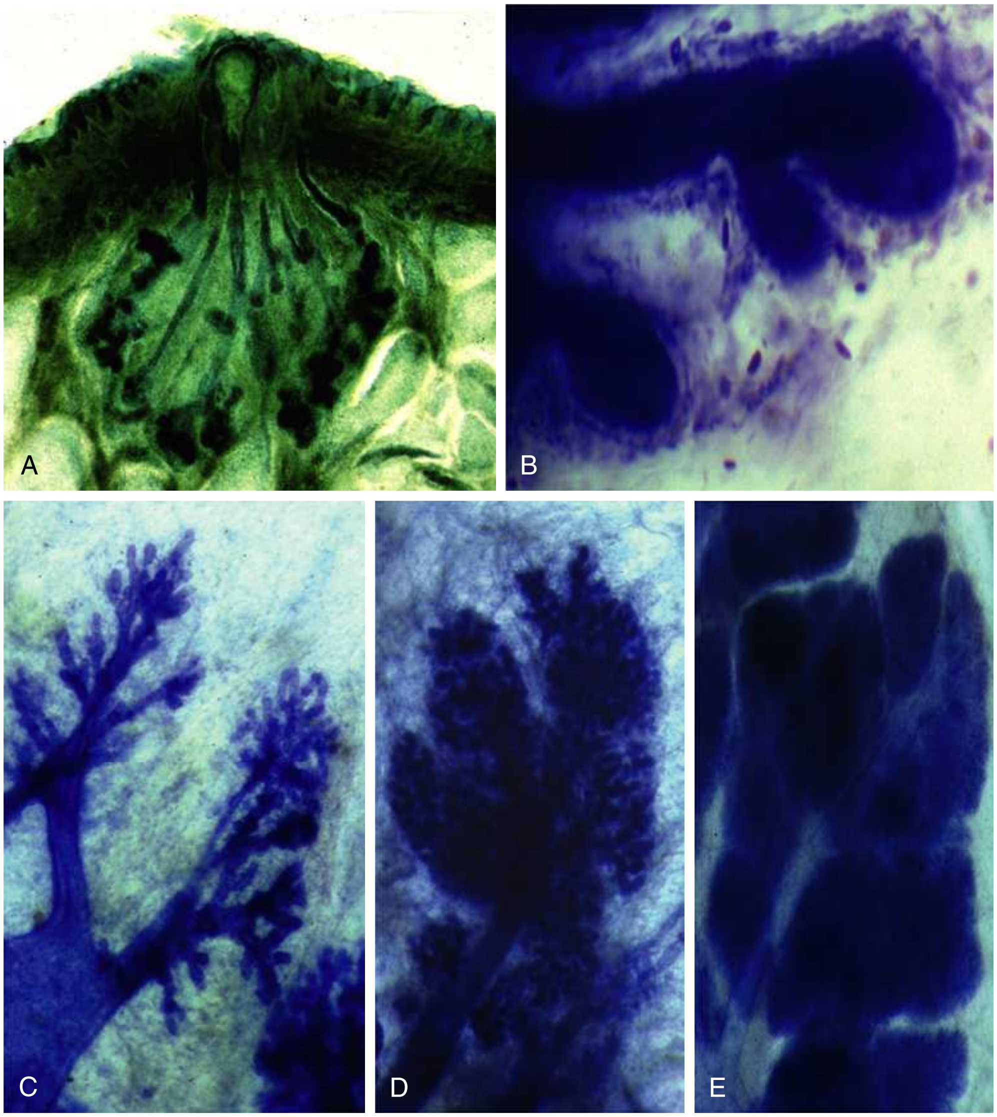


Fig. 127.1 Mammary gland development. **A,** Mammary gland of human female at birth formed by several excretory ducts, ending in terminal ducts. Detail of the inset showing the club-shaped terminal end bud from which lengthening and further divisions of the virginal ducts originate. Toluidine blue ×25. **B,** Before the onset of puberty in human females, the ducts grow and divide on a dichotomous and sympodial basis, a ball-shaped end bud sprouts from the duct, and new branches and twigs develop from the terminal and lateral end buds. Toluidine blue ×25. **C,** Whole mount preparation of breast tissue from an 18-year-old nulliparous woman showing type 1 lobules. Toluidine blue ×25. **D,** Whole mount preparation of human breast tissue from a 24-year-old nulliparous woman showing type 2 lobules. Toluidine blue ×25. **E,** Whole mount preparation of human breast tissue from a 35-year-old parous woman containing type 3 lobules. Toluidine blue ×25.

involves repeated dichotomous (Greek *dichomotos*, "divide into two parts") bifurcation of existing ducts, and the other involves localized buds on an expanding ductular base, a process called "sympodial" from the Greek *syn* + *podion* base. The ducts grow, divide, and form club-shaped terminal end buds (TEBs), giving rise to new branches, twigs, and small ductules or *alveolar buds* (Fig. 127.2). The "alveolar bud" identifies structures that are morphologically more developed than the TEB but are more primitive than the terminal structure of the mature structure (an acinus). Alveolar buds cluster around a terminal duct, forming the type 1 lobule or virginal lobule (Figs. 127.1 and 127.2). Each cluster comprises approximately 11 alveolar buds. Lobule formation in the female breast occurs within 1 to 2 years after onset of the first menstrual period. Full differentiation of the mammary gland is a gradual process that takes many years; in some cases, if pregnancy does not supervene, it is never attained.

THE MATURE BREAST

KEY POINTS

- Growth and differentiation of the breast continue beyond puberty and throughout the life course, and are programmed by a woman's pregnancy history.
- The morphologic and structural changes that occur in the breast in nulliparous versus parous women underscore profound alterations that reflect the function of the gland, as well as potential susceptibilities to breast cancer later in life.

Study of normal breast tissue in adult women identifies two additional and more mature types of lobule, designated types 2 and 3 (see Figs. 127.1 and 127.2). The transition from type 1 to the more mature types 2 and 3

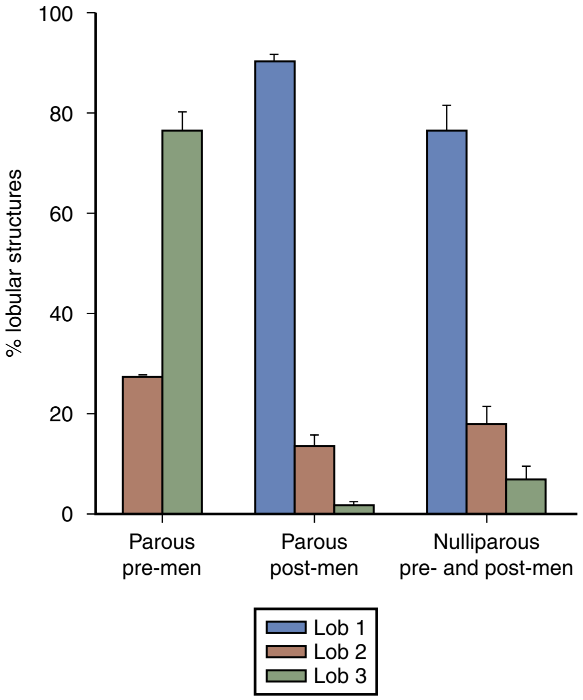


Fig. 127.2 Percentage of lobule-type structures in the breast of premenopausal (pre-men) and postmenopausal (post-men) parous women, and premenopausal and postmenopausal (pre-men and post-men) nulliparous women. Lobules type 1 (lob 1), lobules type 2 (lob 2), and lobules type 3 (lob 3).

TABLE 127.1 Characteristics of the Lobular Structures of the Human Breast

Structure	Lobular Area,[a] μm^2	No. of Ductules/ Lobules[b]	No. of Cells/ Cross Section[c]
Lob 1	48 ± 44	11.2 ± 6.3	32.4 ± 14.1
Lob 2	60 ± 26	47.0 ± 11.7	13.1 ± 4.8
Lob 3	129 ± 49	81.0 ± 16.6	11.0 ± 2.0

[a] Student's t-tests were done for all possible comparisons. Lobular areas showed significant differences between lob 1 vs. lob 3 and for lob 2 vs. lob 3 ($P < .005$).

[b] The number of ductules per lobule was different ($P < .01$) in all comparisons.

[c] The number of cells per cross-section was significantly different in ductules of lob 1 vs. lob 2 and lob 3 ($P < .01$).

From Russo J, Rivera R, Russo IH. Influence of age and parity on the development of the human breast. *Breast Cancer Res Treat.* 1992;23:211–218.

represents a gradual process, with sprouting of an increased number of new alveolar buds. In type 2 and 3 lobules, these are now called ductules; they increase in number from approximately 11 in the type 1 lobule to 47 in the type 2 lobule and 80 in type 3, respectively (see Fig. 127.1; Table 127.1). The increase in number results in a concomitant increase in the size of the lobules and a reduction in the size of each individual structure.[3,8]

Nulliparous Women

In these women, breast tissue contains more undifferentiated structures, such as terminal ducts and type 1 lobules, although occasional type 2 and 3 lobules are seen (Fig. 127.3). This pattern remains constant throughout the reproductive years unless pregnancy ensues. Type 2 lobules are present in moderate numbers during the early reproductive years but sharply decrease after age 23 years, whereas the number of type 1 lobules remains significantly higher. This observation suggests that some type 1 lobules may progress to type 2 lobules, but the number of type 2 lobules that progress to type 3 is significantly lower than in parous (P) women.[9]

Parous Women

In this group, the predominant structure is the most differentiated (type 3) lobule. The number of type 3 structures peaks during the early reproductive years and decreases after the fourth decade of life. Parity between the ages of 14 and 20 years correlates with a significant increase in the number of type 3 lobules that remain as the predominant structure until the age of 40 years, when the number of type 3 lobules decreases owing to their involution to predominantly type 1 lobules[9] (Fig. 127.3).

DISTRIBUTION PATTERN OF ESTROGEN RECEPTOR– AND PROGESTERONE RECEPTOR–POSITIVE AND PROLIFERATING CELLS IN THE MAMMARY GLAND

Even though the breast is influenced by numerous hormones and growth factors,[10-17] estrogens play a major role in promoting proliferation of normal (and neoplastic) breast epithelium.[18-24] They stimulate the proliferative activity of mammary epithelial cells by one of at least three different mechanisms: direct receptor-mediated stimulation,[19,25-32] indirect autocrine/paracrine loops,[22-23] or interruption of negative feedback factors.[21,33]

ERa, encoded by *ESR1*, and ERb, encoded by *ESR2*, with similar affinity for E2, are key mediators of E2 actions on the breast.[34-36] The essential roles of ER and PR on cell proliferation, essential for normal growth and development of the breast, derive from mouse knockout studies. ERa knockout mice display grossly impaired ductal epithelial proliferation and branching, and PR knockout mice display significant ductal development but decreased arborization and absence of alveolar differentiation.[37,38]

Our studies[10-13,18,39-41] and others[42] of mammary epithelial proliferation in rodents and humans have demonstrated that cell division varies with degree of differentiation of mammary parenchyma. In women, the highest level of cell proliferation is observed in the undifferentiated type 1 lobules (lob 1) of young nulliparous (NP) women.[10-13,18] The progressive differentiation of type 1 lobules into type 2 and 3 results in a concomitant reduction of mammary epithelial proliferative activity.[9-13] Further differentiation into type 4 lobules, characteristic of breast tissue during the end of pregnancy and during lactation, further lowers the rate of proliferation.

Our studies have also provided an understanding of the relationship between lobular differentiation, cell proliferation, and hormone responsiveness of the mammary epithelium. The ERa and PR content of the lobular structures is directly proportional to the rate of cell proliferation, both being maximal in the undifferentiated type 1 lobule and decreasing progressively in types 2, 3, and 4. The highest percentage of Ki67-positive cells occurs in type 1 lobules (Fig. 127.4 and Table 127.2), with a 3-fold lower percentage in type 2 lobules and a 10-fold decrement in type 3 lobules. Proliferating cells are found almost exclusively in epithelium-lining ducts and lobules, and occasional positive cells are found in the myoepithelium and the intra- and interlobular stroma. ER- and PR-positive cells are found exclusively in the epithelium, with type 1 lobules containing the highest number, decreasing progressively in type 2 and 3 lobules (see Fig. 127.4 and Table 127.2).

A double immunostaining technique for Ki67 and ER or PR has allowed quantitation in the same tissue sections of the spatial relationship between proliferating and receptor-positive cells. The percentages of ER- and PR-positive cells in type 1 lobules do not differ significantly

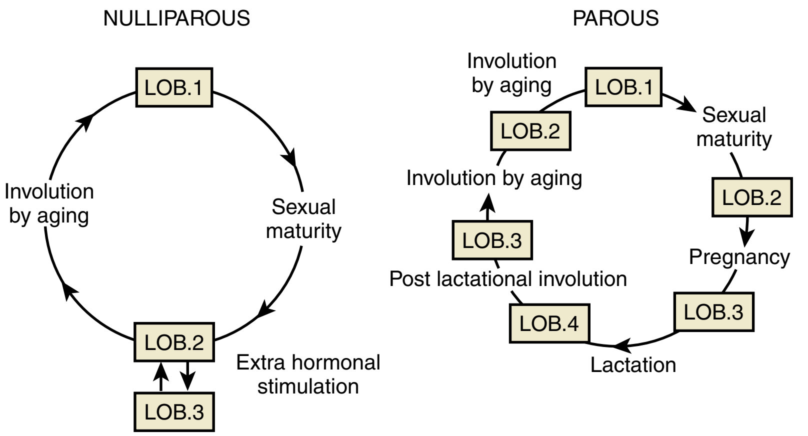


Fig. 127.3 Schematic representation of breast development based upon the relative percentage of lobules present. Nulliparous women's breasts contain primarily lobules type 1 (lob 1), with some progression to type 2 (lob 2) and only minimal formation of lobules type 3 (lob 3). Parous women undergo a complete cycle of development through the formation of lobules type 4 (lob 4), which later regress. (From Russo J, Rivera R, Russo IH. Influence of age and parity on the development of the human breast. *Breast Cancer Res Treat.* 1992;23:211–218.)

(7.5% and 5.7%, respectively, see Table 127.2). In type 2 lobules, they are reduced to 3.8% and 0.7%, respectively, and in type 3, they are negligible (see Table 127.2). Similar relative percentages of Ki67-, ER-, and PR-positive cells and the progressive decrease of positive cells have been observed as lobular differentiation progresses. However, cells positive for Ki67 are not the same as those that express ER or PR (see Fig. 127.4 and Table 127.2). Few cells (<0.5%) in type 1 lobules, and even fewer in type 2 and 3 lobules, are positive for both Ki67 and ER (Ki67+ER) (see Table 127.2). The percentage of cells exhibiting double labeling with Ki67 and PR antibodies (Ki67+PR) in type 1 lobules is lower than the percentage of double-labeled ER-positive cells (see Table 127.2).

KEY POINTS

- ER alpha and PR are key players in lobular differentiation, cell proliferation, and hormone responsiveness of the mammary epithelium.
- Evidence supports 3 types of mammary epithelial cells: ER+/Ki67+ that proliferate and are most abundant in type I lobules; ER+ that do not proliferate, and a few ER-/ Ki67+ cells.
- Growth factors likely produced by estrogen signaling through ERa can stimulate neighboring ER- cells to proliferate.
- ER- cells may convert to ER+ cells during neoplastic transformation.

That proliferating cells differ from those that are ER- and PR-positive supports an indirect mechanism for estrogen regulation of proliferation. A likely explanation is that estrogen acts on ER-positive cells, which respond to the increase in growth factors by acting in a paracrine fashion on neighboring ER-negative cells. This was confirmed by the finding that supernatants from estrogen-treated ER-positive cells stimulate the growth of ER-negative cell lines in culture and *in vivo* in nude mice bearing ER-negative breast tumor xenografts.[43,44] A paracrine mechanism can also explain inhibition of cellular proliferation. ER-positive cells treated with antiestrogens secrete transforming growth factor alpha, which inhibits proliferation of ER-negative cells.[29]

The findings that proliferative activity and the percent of ER- and PR-positive cells are highest in type 1 lobules compared with types 2 through 4 provide a mechanistic explanation for higher susceptibility of these structures to transformation by chemical carcinogens *in vitro*[45,46] and support the observations that type 1 lobules are the site of origin of ductal carcinomas.[47] However, the relationship between ER-positive and ER-negative breast cancers is not clear.[48-50] It has been suggested that ER-negative breast cancers may result from loss of the ability of cells to synthesize ER during clinical evolution of ER-positive cancers, or that ER-positive and ER-negative cancers may be different entities.[48,50] It is postulated that type 1 lobules contain at least three cell types: ER-positive cells that do not proliferate, ER-negative cells that are capable of proliferating, and a small proportion of ER-positive cells that can proliferate as well (Fig. 127.5). Therefore, estrogen may stimulate ER-positive cells to produce a growth factor that in turn stimulates neighboring ER-negative cells capable of proliferating (see Fig. 127.5). In the same fashion, the small proportion of ER-positive cells that can proliferate could be the source of ER-positive tumors. Also, conversion of ER-negative to ER-positive cells has been reported.[51,52] The presence of ERb opens the possibility that cells traditionally considered to be ERa-negative may be ERb-positive.[34,35] It has been found that ER is expressed during immortalization and transformation of ER-negative human breast epithelial cells, supporting the hypothesis of conversion from a negative to a positive receptor cell.[53]

Findings that proliferating cells in the human breast differ from those that contain steroid hormone receptors explain some *in vitro* data.[54-57] Notably, estrogen stimulates ER-positive MCF-7 cell proliferation, and enhanced expression of the receptor by transfection also increases the proliferative response.[54,58] However, when ER-negative cells such as MDA-MB 468 cells are transfected with ER, estrogen inhibits cell growth.[55-59] Although the negative effect of estrogen on ER-negative cells transfected with the receptor has been interpreted as an interference of the transcription factor used to maintain estrogen-independent growth,[58] no definitive explanation has been put forth for their lack of survival, although it can be explained by proliferating and ER-positive cells being two separate populations. Further support is provided by the finding that, when type 1 lobules of normal breast tissue are placed in culture, they lose their ER-positive cells, indicating that only proliferating cells that are also ER-negative can survive. It has been suggested that these constitute the stem cells.[59,60]

Although estrogens act through nuclear ERs that transcriptionally activate specific target genes, a membrane receptor coupled to alternative second messenger signaling mechanisms[61,62] is also operational and may stimulate a cascade of events leading to proliferation. This suggests that ERa-negative cells in the human breast may respond to estrogens through this or other pathways. It is clear that paracrine mechanisms are important in mediating cell proliferation. In the

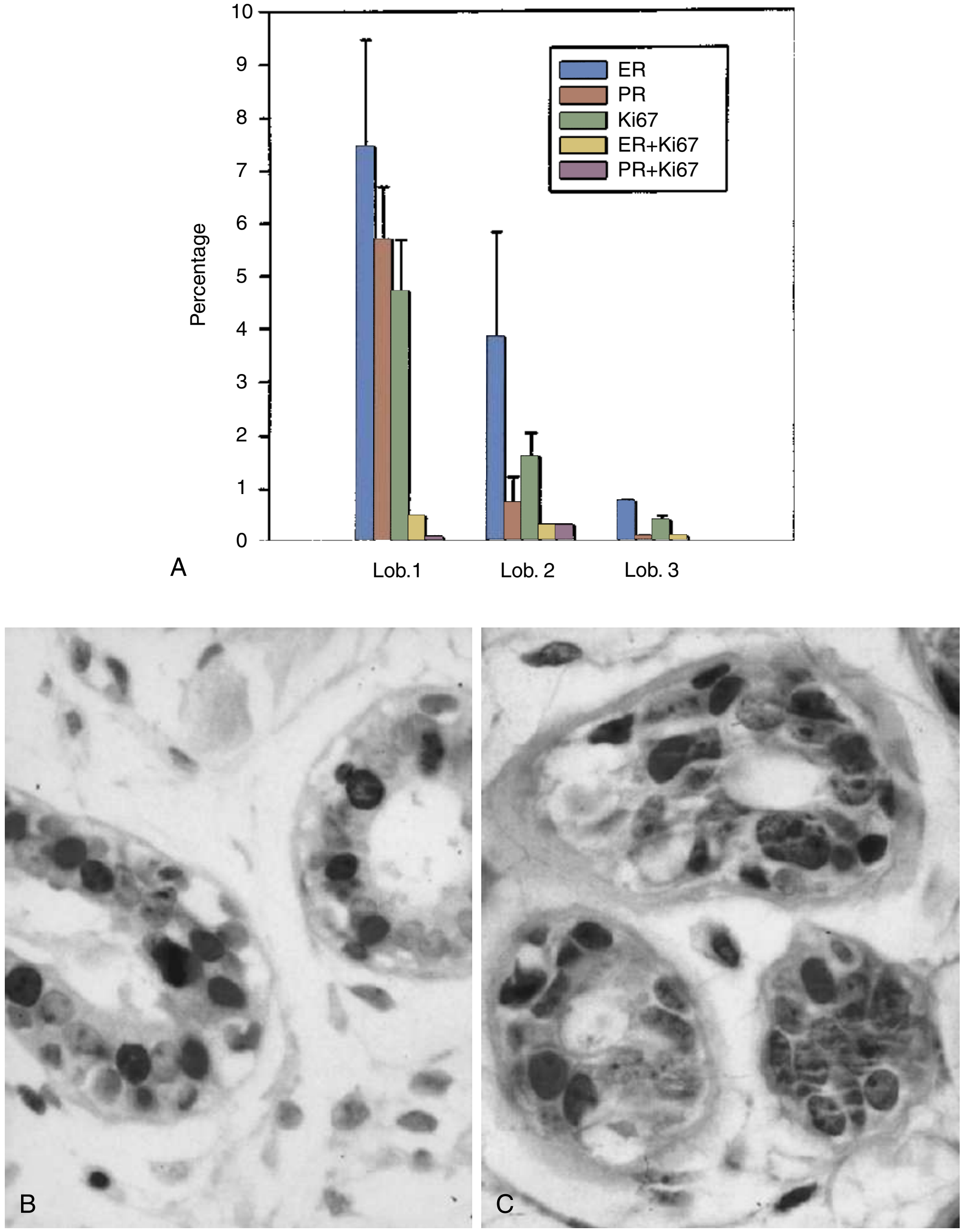


Fig. 127.4 Percentage of cells positive for estrogen receptor (*ER*), progesterone receptor (*PgR*), Ki67, and both ER and Ki67 (ER+Ki67) or PgR and Ki67 (PgR+Ki67) *(ordinate)*. Cells were quantitated in lobules type 1 (lob 1), type 2 (lob 2), and type 3 (lob 3) of the breast *(abscissa)*. Lob 1 ductules of the human breast. **A,** The single-layered epithelium lining the ductule contains Ki67-positive cells *(brown nuclei)* and ER-positive cells *(red-purple nuclei)* (×40). **B,** The single-layered epithelium lining the ductule contains brown Ki67-positive cells and red-purple PgR-positive cells. Sections were stained with 3,3′-diaminoben-zidine (DAB)/alkaline phosphatase-vector red, with light hematoxylin counterstain, and were photographed at ×40.

normal breast, proliferating and steroid hormone receptor–positive cells are different. This finding opens new possibilities to clarify the mechanisms through which estrogens may act on proliferating cells to initiate the cascade of events leading to cancer.

STEM CELLS IN THE NORMAL HUMAN BREAST AND THEIR IMPLICATIONS IN BREAST CANCER

Stem cells in adult structures are defined by their ability to self-renew *and* generate differentiated progeny. In the mammary gland, DeOme et al.[63] demonstrated that fragments of different parenchymal portions were able to generate fully functional mammary outgrowths in mice, forming ductal and lobuloalveolar structures composed of epithelial and myoepithelial cells. This concept was further developed by Kordon and Smith,[64] who demonstrated that the progeny from a single cell could comprise the epithelial population of a fully developed lactating mammary outgrowth in mice. Thus, development of the complete mammary tree from a small portion of a duct or from single cells attests to their multifaceted potential. However, it was not known whether these progenitor/stem cells would be capable of initiating cancer when

TABLE 127.2 Distribution of Ki67-, Estrogen Receptor (ER)–, and Progesterone Receptor (PR)–Positive Cells in the Lobular Structures of the Human Breast

Lobule Type	No. of Cells	Ki67	ER	PR	Ki67+ER	Ki67+PR
Lob l	19,339[a]	4.72 ± 1.00[de]	7.46 ± 2.88[h]	5.70 ± 1.36[k]	0.48 ± –0.28[n]	0.09 ± 0.01[o]
Lob 2	8490[b]	1.58 ± –0.45[f]	3.83 ± 2.44[i]	0.73 ± –0.57[l]	0.31 ± –0.21	0.28 ± –0.27
Lob 3	l7,750[c]	0.40 ± 0.18[g]	0.76 ± –0.04	0.09 ± 0.04[m]	0.01 ± 0.01	0.01 ± 0.01

[a] Total number of cells counted in lobules type 1 (lob l) in breast tissue samples from 12 donors.
[b] Total number of cells counted in lobules type 2 (lob 2) in breast tissue samples from five donors.
[c] Total number of cells counted in lobules type 3 (lob 3) in breast tissue samples of from three donors.
[d] Proliferative activity determined by the percentage of Ki67-positive cells, expressed as the mean ± standard deviation. Differences were significant in [e]lob 1 vs. [f]lob 2 (t = 1.98; $P < .05$), [f]lob 2 vs. [g]lob3 (t = 2.27; $P < .04$), and [e]lob 1 vs. [g]lob 3 (t = 2.56; $P < .01$).
ER-positive cells were significantly different in [h]lob 1 vs. [i]lob 2 (t = 2.04; $P < .05$) and [i]lob 3 (t = 4.03; $P < .001$).
PR-positive cells were significantly different in [k]lob 1 vs. [l]lob 2 (t = 2.27; $P < .05$) and [k]lob 1 vs. [m]lob 3 (t = 2.60; $P < .031$).
[n] Percentage of cells positive for both Ki67 and ER, expressed as the mean ± standard deviation.
[o] Percentage of cells positive for both Ki67 and PR, expressed as the mean ± standard deviation.

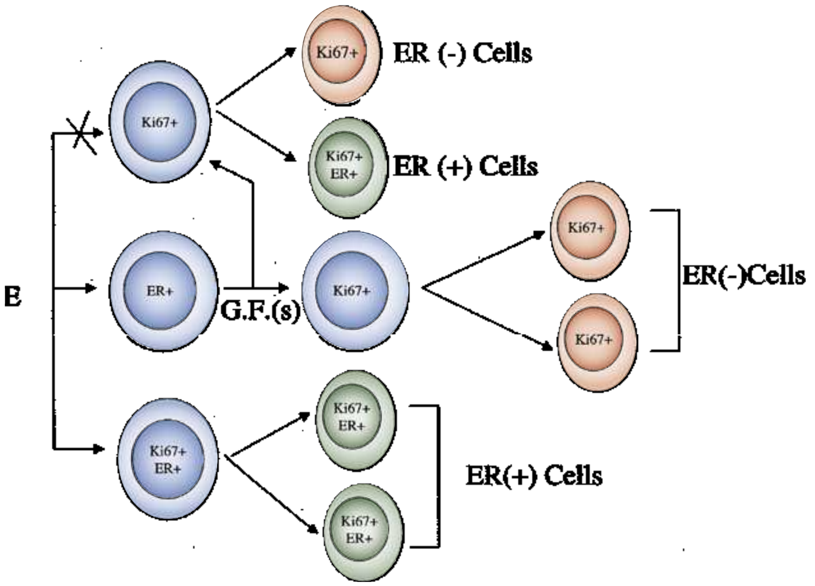


Fig. 127.5 Schematic representation of the postulated pathways of estrogen actions on breast epithelial cells. Cells expressing three different phenotypes might be present in the epithelium: estrogen receptor (*ER*)-negative Ki67-positive cells that are capable of proliferating, ER-positive cells that do not proliferate, and a small proportion of ER- and Ki67-positive cells. Estrogen may stimulate ER-positive cells to produce a growth factor that in turn stimulates neighboring ER-negative cells capable of proliferating. ER+Ki67–positive cells can proliferate and could be stimulated by estrogen to originate ER-positive daughter cells and probably tumors. ER-negative cells may convert to ER-positive cells during neoplastic transformation.

exposed to a carcinogenic agent. This issue was addressed by Russo and coworkers,[65-67] who demonstrated that cancer started in TEBs present in the mammary gland of young virgin rats. Analysis of these structures by electron microscopy allowed characterization of their cellular composition based on cell and nuclear size, nuclear-cytoplasmic ratio, amount of chromatin condensation, electron density of the cytoplasm, organelle number and distribution, and presence or absence of magnesium- and sodium/potassium-dependent ATPases. Based upon these criteria, in addition to myoepithelial cells, three types of epithelial cells were identified: light, intermediate, and dark.[66,67] Dark cells were the predominant type in TEBs; intermediate and myoepithelial cells were present in significantly lower percentages; and light cells were only occasionally seen. Analysis of the DNA labeling index revealed that all cell types proliferated, although at different rates, depending on their location within the mammary gland tree. Cell proliferation was maximal in intermediate cells located in TEBs and was significantly lower in dark and myoepithelial cells in the same location. High proliferation was associated with greater incorporation of 7,12-dimethylbenz(alpha) anthracene (DMBA) and a progressive dominance of intermediate cells in DMBA-induced intraductal proliferations and in ductal carcinomas.[8,67] These results indicated that intermediate cells were not only the targets of the carcinogen but also the stem cells of mammary carcinomas. Further work by Bennett et al.[68] demonstrated that intermediate cells isolated from DMBA-induced mammary tumors resulted in two cell types in culture: the dark cell, representing a terminally differentiated cell or a class in transition to differentiation, and intermediate cells, which could represent an undifferentiated, or stem, cell, a progenitor of dark and myoepithelial cells. Rudland and coworkers[69] isolated and characterized, from the normal rat mammary gland and from DMBA-induced mammary adenocarcinomas, epithelial cells that were cuboidal and gave rise to a mixture of cuboidal and spindle-shaped cells resembling fibroblasts. In confluent cultures, cuboidal cells acquired the morphology of a third type of cell, which was dark, polygonal, and with many small vacuoles, resembling dark cells ultrastructurally described by Russo et al.[67] Chepko and Smith[70] differentiated division-competent cell populations in murine mammary epithelium, including a subset of "large light cells" that were structurally and functionally compatible with early stages of secretory differentiation and "small light cells" that were the least differentiated, suggesting that large light cells were a direct precursor to terminally differentiated secretory and myoepithelial cells. These results confirmed Russo's work.[67]

A shift from the pioneering studies characterizing progenitor/stem cells by morphology and *in vitro* behavior began with the search for immunocytochemical and genomic markers.[71] Expression of keratins 6 and 14 defined subsets of morphologically distinct luminal mammary epithelial cells with kinetic properties expected for latent mammogenic stem cells. Keratin 6 was confined to a few mammary epithelial cells, the growing end buds, and the luminal epithelium, whereas keratin 14 was expressed in basally located fusiform cells such as the myoepithelial cells.[71] The authors emphasized the usefulness of these markers in identifying mammary epithelium–specific primordial cells. Stingl et al.[72,73] used new molecular markers to select subpopulations of cells with distinct differentiation potential. They described bipotent human mammary epithelial progenitor cells based on the expression of epithelial-specific antigen (ESA), sialomucin 1 (MUC1), common acute lymphoblast antigen (CALLA/CD10,) and a-integrin, in combination with exclusion of rhodamine dye. Hebbard et al.[74] observed that CD44, a member of the family of cell surface proteins that is expressed in breast carcinomas, is also expressed in normal mammary gland. In rodents, CD44 expression is first detected at puberty, and thereafter is regulated by the estrous cycle, disappears during lactation, and reappears during involution, suggesting that it is a stem cell marker. Novel studies in mouse mammary gland[75] have identified stem cells in TEBs and ducts by pulse-labeling HC-11 primary mammary epithelial cells with fluorescent TRITC-cell linker membrane label and BrdU. The cells were then transplanted into cleared juvenile syngenic mammary fat pads, where they were identified as long-lived, label-retaining mammary epithelial cells (LRCs) in mammary ducts that were actively growing or static. This study demonstrated that LRCs are stem cells, that their progeny (transitional cells) are arranged as transitional units (TUs), and that both express Zonula Occludens-1 and a-catenin, suggesting that TUs retain stem cells.

The study of markers for other stem cells has been useful in identifying mammary stem/progenitor cells. Sca1 (stem cell antigen 1), a hematopoietic stem cell antigen,[76] is expressed in mouse breast luminal epithelium enriched for functional stem/progenitor cells. These cells are BrdU label–retaining, lack expression of differentiation markers, and are PR-negative. The Sca1-positive population also exhibits "side population" (SP) properties, and it has been proposed that breast cancer resistance protein 1 (BCRP1) is responsible for this phenotype and may serve as a marker for stem cells from various sources.[77,78] Mammary epithelial cells with SP properties were also identified in human mammary gland. Alvi et al.[79] showed that 0.2% to 0.45% of both human and mouse epithelia were formed by distinct SP cells. These cells generated ductal and lobuloalveolar structures when transplanted into murine cleared mammary fat pads. The SP cells highly expressed BCRP1, Sca1, telomerase catalytic subunit, and low levels of differentiated markers for luminal (epithelial membrane antigen and cytokeratin 19) and myoepithelial cell types (cytokeratin 14). These cells were detected in all human breast samples studied, but their presence was not correlated with age, parity, contraceptive use, or day of menstrual cycle.

Further studies identified new markers that may be specific for human stem/progenitor cells. Gudjonsson et al.[80] isolated a cell line derived from human mammary cells expressing ESA and lacking MUC1 expression that gave rise to luminal epithelial and myoepithelial cells in culture. One ESA+/MUC1– cell generated a terminal ductal-lobular unit-like structure in basement membrane gel, similar to that formed when the cell line was implanted in mice. In contrast, an ESA+/MUC1+ subpopulation was differentiated and luminal epithelial–restricted, without stem cell properties. Dontu and coworkers[81] enriched the population of human mammary progenitor/stem cells by culturing them in suspension where they formed "nonadherent mammospheres." These structures differentiated along all three mammary epithelial lineages and clonally generated complex functional structures in three-dimensional culture systems. Cytological and immunocytochemical analysis of secondary mammospheres revealed they contained cells positive for a6 integrin, cytokeratin 5, and CD10; ESA-positive and cytokeratin 14–positive cells were less frequent. Muc1, a-smooth muscle actin, and cytokeratin 18 were not detected. In addition to cells, mammospheres contained extracellular material (ECM). However, immunostains for fibronectin and collagen IV, classical components of adult gland ECM, were negative, although approximately 20% of the mammospheres stained positive for laminin. In contrast, abundant expression of the embryonic ECM components tenascin and decorin was detected in mammospheres.[81] Moreover, comparing genomic profiles of undifferentiated cells from mammospheres with differentiated cells cultured on collagen revealed genes candidates for stem/progenitor cell markers. Some of these were already described in stem/progenitor cell–specific functions or regulation of self renewal, and abnormal expression of some has been correlated with breast cancer development, including proliferation, cell survival, and invasion. Other studies showed that the null mutation of the peroxisome proliferator-activated receptor-binding protein (PBP) resulted in defective mammary gland development and inability of mammary epithelial cells to form mammospheres, suggesting a role of PBP in the survival of mammary stem cells or in mammosphere formation.[82]

Genomic Signatures of Breast Changes After the First Full-Term Pregnancy in Premenopausal Women

Recently, we reported parity-associated gene expression differences in the breast of premenopausal women.[83] Because the breast undergoes involution after full-term pregnancy (FTP) (see Fig. 127.3), we examined gene expression in P versus NP women as a function of time since last FTP. Women ($n = 109$ [$n = 30$ NP, $n = 79$P]), 29 to 47 years old, were recruited.[83] Forty-three genes were differentially expressed (40 up- and three downregulated) between P and NP women.[83] Analyses of gene expression according to the time since last pregnancy (TSLP) identified 286 genes differentially expressed between P and NP samples (Fig. 127.6). Three expression patterns as a function of TSLP were noted (Fig. 127.7). The first consisted of 83 genes upregulated following the last FTP, but whose expression progressively returned to the level of expression observed in NP women (see Fig. 127.7A); these were named "transient" genes. The second cluster consisted of genes for which P–NP differences were highest for women with less than 5 years since last FTP, decreased with increasing TSLP, and remained elevated versus NP women. The 95 genes in this cluster were called "long-term changing" (see Fig. 127.7B). The third cluster included 60 genes whose fold changes between P and NP were constant, regardless of the TSLP. These were called "long-term constant" genes (see Fig. 127.7C).

Gene Ontology (GO) enrichment analyses showed an abundance of terms associated with developmental processes or immune response (Fig. 127.8), with less enrichment related to proliferation, intracellular transport, and cell death. Among the genes whose expression was transiently affected by parity, 55% were related to immune response. Among "long-term changing" genes, 32% were related to immune response (e.g., CD2), and 24% were involved in developmental processes. Notably, of the "long-term constant" genes, 56% were related to developmental/differentiation processes. These data indicate that the proportion of genes involved in immune response decreases among the genes that are upregulated irrespective of TSLP, while the proportion of genes related to developmental processes increases (chi-squared $P < .001$). Evaluation of

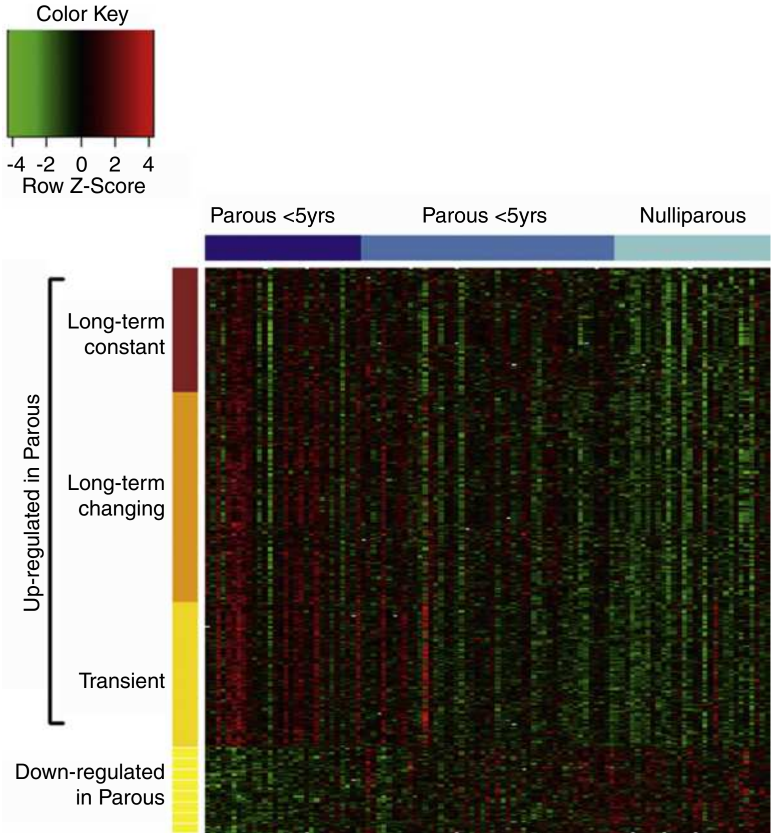


Fig. 127.6 Heatmap of all 416 probesets found differentially expressed between parous and nulliparous women. Each column represents a volunteer. Dark blue represents parous women within 5 years since last full-term pregnancy, blue represents parous women with more than 5 years since pregnancy, and cyan represents nulliparous women. Each line corresponds to a probeset, which, according to its pattern of expression within the parous breast samples compared with nulliparous, was classified as downregulated *(light yellow)*, transiently upregulated *(yellow)*, long-term changing *(orange)*, and long-term constant upregulated *(brown)*. (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res*. 2019;21:46.)

the canonical pathways represented by the differentially expressed genes also showed that pathways involved in signaling and activation of T-cells were enriched by both transient and long-term genes (Fig. 127.9.[83]

The 48 genes downregulated in the P breast fell into two patterns. Twenty-two genes, including *WIF1, EDN1, CXCL1*, and *FOXQ1*, were downregulated in the P breast and remained low compared with NP breast, irrespective of years since last FTP. The remaining 26 genes were downregulated in women with TSLP less than 5 years; however, expression increased, reaching similar or higher levels than in NP women, as TSLP increased. Limited numbers precluded conducting a GO enrichment analysis.[83]

We also analyzed 11 selected genes from our dataset in 5796 patients with breast cancer available in cBioPortal (www.cbioportal.org). One or more of these genes were altered in 770 (13%) breast cancer patients. *BHLHE22* was altered in the largest percentage of patients (6.9%); the GO alterations included amplification, deep deletion, and missense mutations. The other ten genes were altered in 1.5% or less of patients. An overall Kaplan–Meier survival curve generated to compare patients with versus without these 11 altered genes revealed shorter overall survival in the former group (median 143.13 vs. 168.3 months; logrank test $P = 9.09e{-}5$ (Fig. 127.10).

Gene expression was confirmed by reverse transcription polymerase chain reaction (RT-PCR) in a subset of 37 samples (20 P and 17 NP) of the NP versus P study.[83] Also, immunohistochemical (IHC) analyses validated protein expression of some genes in the microarray and RT-PCR analyses. Markers for dendritic cells, important in antigen presentation for T-cell activation and T-cell activation markers *per se*, were studied. CD123 and LAMP3 were detected in a few cells, with no differences between P and NP. CD2 and CD3D, markers of T-cell activation (Fig. 127.11), were expressed at low levels in both groups, with a slight, nonsignificant increase in the percent of CD2-positive cells in P women (1.6 fold; P median = 0.59%, NP median = 0.31%; $P = .156$). The overall percentage of CD3D-positive cells was higher (0.6%–9.4%), with a statistically significant 2-fold increase in the percent of CD3D-positive cells in the P breast (P median = 3.28%, NP median = 1.62%, $P = .006$) (see Fig. 127.11). Comparison of CD2- and CD3D-positive cells between P women with TSLP of 5 or less and greater than 5 revealed a slightly larger number in TSLP of 5 or less, but this was not statistically significant (CD2 $P = .16$, CD3D $P = .09$).

KEY POINTS

- Gene expression differences in the parous (P) and nonparous (NP) breast revealed 43 differentially expressed genes.[83] Evaluation of expression differences between NP and P as a function of time since last pregnancy identified 238 up- and 48 downregulated genes in the P breast. Downregulated genes fell into two patterns (transient and long-term), while upregulated genes fell into three patterns (transient, long-term changing, and long-term constant). Through Gene Ontology enrichment analyses, genes whose expression was transiently increased after pregnancy were mainly related to immune response. Long-term changing genes included immune- and development-related genes, while genes categorized as long-term constant were mainly involved in cell differentiation and developmental processes.

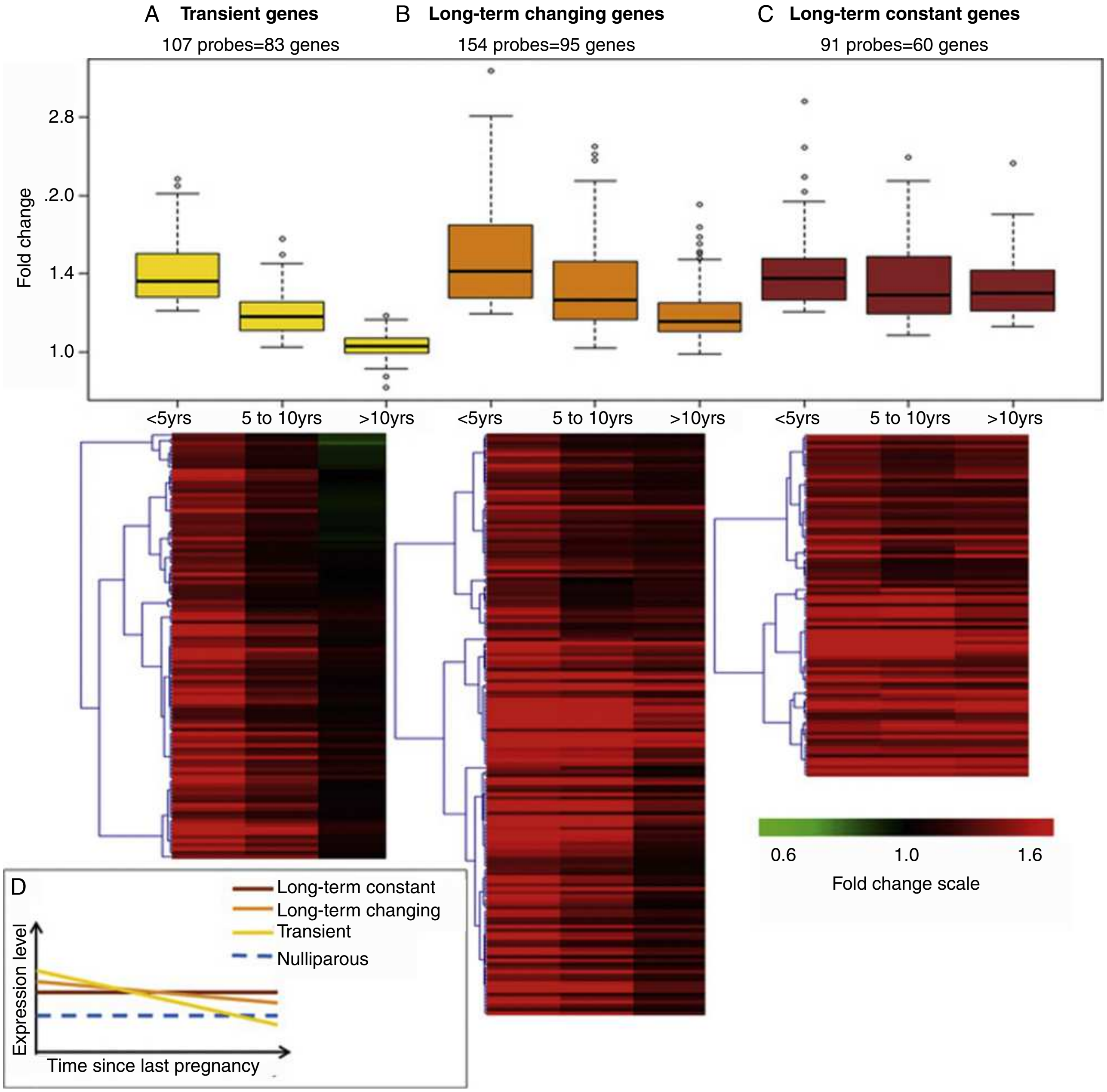


Fig. 127.7 Three expression patterns were identified among the upregulated genes. In the top, the graph shows the fold change ranges of each cluster: **A,** transient; **B,** long-term changing; and **C,** long-term constant, according to time since last full-term pregnancy (<5 years, 5–10 years, and >10 years). Heatmaps show the expression patterns of each cluster. **D,** Schematic representation of the expression levels of each cluster in function of time since last pregnancy. (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* 2019;21:46.)

THE MENOPAUSAL BREAST

Menopause occurs at an average age of 51 years, when the ovaries stop synthesizing E2 and amenorrhea ensues, and the years leading up to the final menstrual period constitute the perimenopause, wherein irregular ovulation is common (see Chapter 122). After menopause, the breast regresses in both NP and P women.[10] This involutionary process is manifested by an increase in the number of type 1 lobules and a concomitant decline in type 2 and type 3 lobules. At the end of the fifth decade of life, the breasts of both NP and P women contain predominantly type 1 lobules (see Fig. 127.3). Understanding breast development requires a longitudinal timeline in which all the phases of growth are considered, as analysis of breast structures at a single point (e.g., age 50 years) may lead to the conclusion that breasts of NP and P women are identical (see Fig. 127.3). From a quantitative perspective, however, breast regression at menopause differs in NP and P women. The predominant structure in NP breasts is the type 1 lobule (65%–80% of the total number of lobule-type components), independent of age. Type 2 lobules represent 10% to 35%, and type 3, 0% to 5%. In breasts of premenopausal P women, however, the type 3 lobule comprises 70% to 90% of the total lobule component. At menopause, type 3 lobules have declined, and the relative proportions of the three lobule types are similar to those in NP women. These observations led us to conclude that early P women undergo lobule differentiation, which was evident at a younger age; whereas NP women seldom reached the lobule type 3 stage and never reached the lobule type 4 stage (see Fig. 127.3).

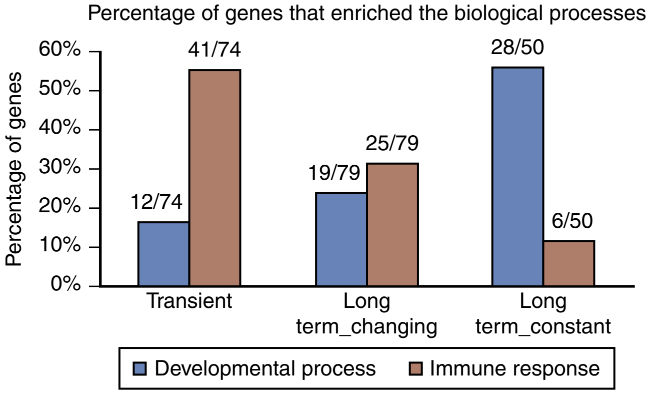


Fig. 127.8 Graph shows the percentage of genes associated with Gene Ontology (GO) terms related to developmental processes *(red)* and immune response *(blue)* for each of the three clusters. The values 74, 79, and 50 are the total number of genes associated at least with one GO term for each cluster. (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res*. 2019;21:46.)

Notably, beyond morphology, accumulating genomic evidence (see later) suggests that prior pregnancy can permanently alter breast biology, not morphologically identifiable, and predispose to neoplastic development.

Genomic Differences in Nulliparous and Parous Breasts at Menopause

KEY POINTS

- Premenopausal and menopausal breast transcriptomic signatures differ from one another.
- Genomic signatures are distinct in the premenopausal breast after first full-term pregnancy compared with nulliparous women
- Pregnancy history (parity and nulliparity) profoundly affects these signatures at menopause, which is related to breast cancer risk.

An early first FTP imprints a specific transcriptomic signature in involuted breast lobules of postmenopausal P women free of breast cancer that significantly differs from P women with cancer and NP women with or without the disease. cDNA microarray analysis of epithelial RNA of completely involuted lobules (lob 1), obtained by laser capture microdissection, revealed that these cells express a signature that comprises 232 deregulated genes, representing 18 functional categories. Compared with gene expression in NP control and case groups and P case groups, the P control group had 126 genes upregulated, verified by RT-PCR. In all, 103 genes were downregulated.[84] Unsupervised hierarchical clustering using expression profiles of 2541 globally varying genes across the NP and P data sets revealed that samples clustered primarily on the basis of parity status. This observation suggested that parity-induced gene expression changes conserved among these highly divergent groups could represent a powerful approach to define parity- and cancer risk–related gene expression signatures. Results of clustering the combined dataset indicate that combined parity and absence of breast cancer generated a distinct transcriptomic profile that differed versus the breast cancer groups, irrespective of parity, and from the NP cancer-free group, which has been identified traditionally as a high-risk group.[84]

We measured the relevance of GO terms in the breast epithelium of P women and analyzed the biological significance of those terms found to be deregulated in response to an early reproductive event.[84] Among the 18 categories identified to contain deregulated genes, the most highly represented and statistically significant biological process was gene transcription, in which 21 genes (64%) were upregulated, and 12 genes (36%) were downregulated. Higher gene expression was observed in 11 processes that included proteolysis and ubiquitination, cell adhesion, response to exogenous agents, metabolism, DNA repair and replication, RNA processing, apoptosis, miscellaneous processes, antiapoptosis, and chromatin modification, in which the ratios of upregulated to downregulated genes ranged from 1.75 to 11.[84]

Numerous genes with low expression were observed in seven processes: cell transport, protein biosynthesis and metabolism, cell signaling–signal transduction, biological process unknown, and biological process and molecular function unknown.[84] Among genes significantly upregulated in the P breast epithelium, 23 belonged to the categories of transcription and chromatin modification—an indication that modifications in epigenetic and transcriptional activity during pregnancy play an important role in and become a permanent component of the genomic signature imprinted by this physiologic process in the postmenopausal breast epithelium. Altogether, an early first FTP induces in the breast epithelium a specific genomic profile in the postmenopausal breast, indicating that these epithelial cells differ from those in NP women and P women with breast cancer. This signature allowed us to evaluate at a genomic level the degree of mammary gland differentiation induced by pregnancy. We postulate that it identifies Stem Cell 2,[85-88] can predict in which women parity has been protective and could be used as an intermediate surrogate biomarker to evaluate preventive strategies. Data obtained to date support our working hypothesis that the lifetime reduction in breast cancer incidence conferred by an early pregnancy is the result of complete differentiation of the breast. This process leads to the replacement of Stem Cell 1, which is a component of the lob 1 present in the NP breast and is the site of origin of ductal breast carcinoma by Stem Cell 2, which is characterized by a specific genomic signature.[86-89] It is important to emphasize that, even though differentiation significantly reduces cell proliferation in the mammary gland, the mammary parenchyma that has regressed after delivery and weaning remains capable of responding with proliferation to given stimuli, such as a new pregnancy. Stem Cell 2 is hypothesized to fulfill these characteristics, because it has been primed by the first cycle of differentiation and thus is able to metabolize carcinogens and repair DNA damage more efficiently than Stem Cell 1 of the NP gland, and therefore is less susceptible to carcinogenesis. However, if the shift from Stem Cell 1 to Stem Cell 2 has not been completed before a carcinogenic stimulus affects the breast, a neoplastic process could be initiated successfully. This may explain cases in which breast cancer develops in spite of the history of early pregnancy, which could be represented by the P cases in our study.[84] The greater breast cancer risk associated with late pregnancy, especially that occurring after 30 years of age, may result from an accumulation of genetic or epigenetic damage during the broadened “high susceptibility” or “high-risk” window noted between menarche and first FTP, as demonstrated in experimental animal models.[80]

Defining the Genomic Signature of Breast Cancer Prevention in Parous Women

More than 300 years have elapsed since a striking excess in breast cancer mortality was reported in nuns, in whom the increased risk was attributed to their childlessness[90] until MacMahon et al.,[91] in a landmark case-control study, found an almost linear relationship between a woman’s risk and the age at which she bore her first child. Numerous studies have examined how dramatic modifications that occur during pregnancy in the pattern of lobular development and differentiation, cell proliferation, and steroid hormone receptor content of the breast[92] influence cancer risk. Global genome-wide analyses, using different platforms, have confirmed the universality of this phenomenon in various strains of rats and mice.[92-97] We recently performed cDNA array

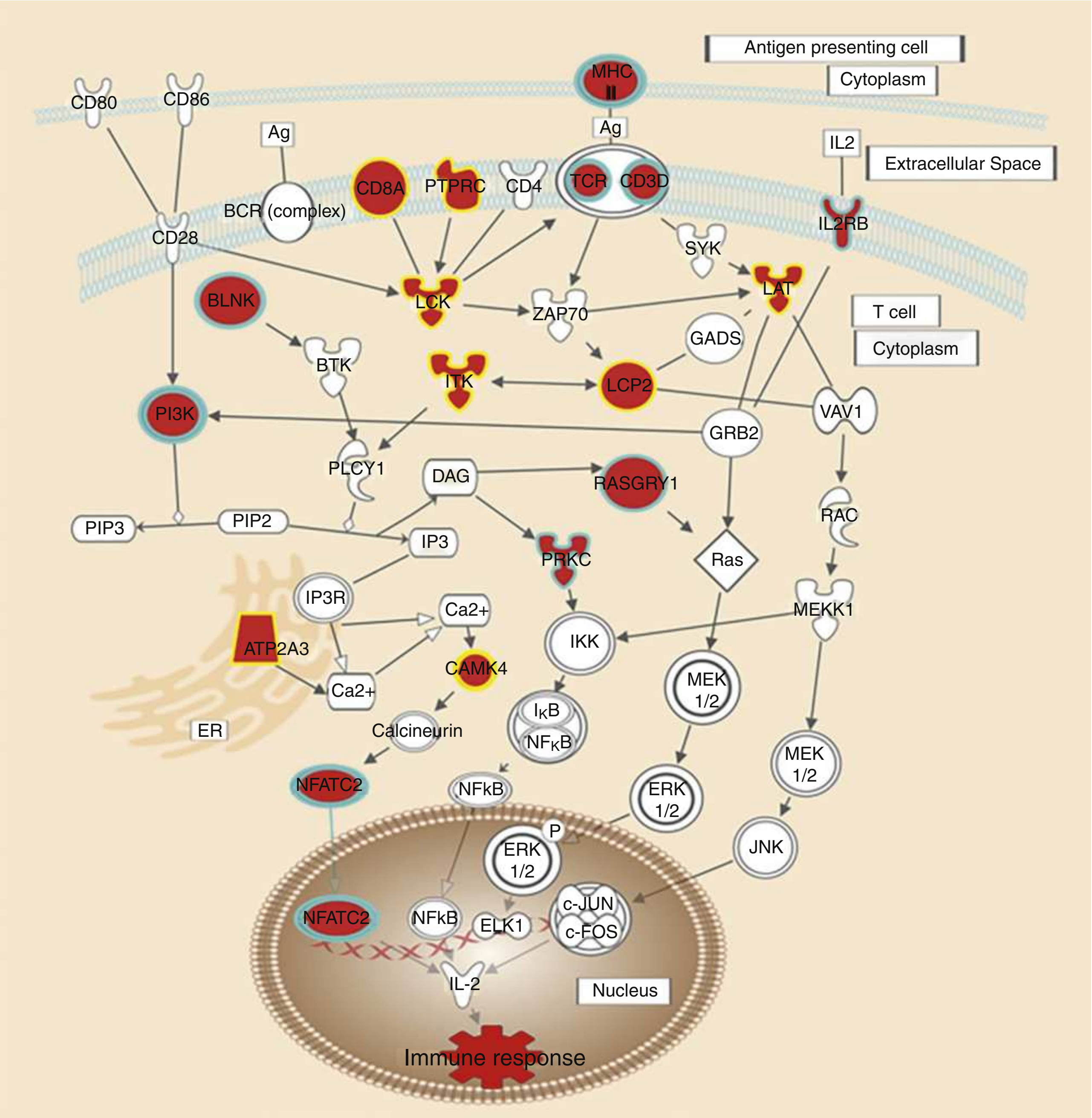

Antigen presenting cell
Cytoplasm
MHC II
CD80
CD86
Ag
Ag
IL2
Extracellular Space
CD8A
PTPRC
CD4
TCR
CD3D
BCR (complex)
IL2RB
CD28
SYK
LAT
BLNK
LCK
ZAP70
T cell
GADS
Cytoplasm
BTK
ITK
LCP2
PI3K
VAV1
GRB2
PLCY1
DAG
RASGRY1
RAC
PIP3
PIP2
IP3
PRKC
Ras
IP3R
Ca2+
MEKK1
ATP2A3
IKK
CAMK4
MEK 1/2
Ca2+
IκB
NFκB
ER
Calcineurin
MEK 1/2
NFATC2
NFkB
ERK 1/2
P
ERK 1/2
JNK
c-JUN
c-FOS
NFATC2
NFkB
ELK1
IL-2
Nucleus
Immune response

Fig. 127.9 Pathway representing T cell-signaling. Genes in red were upregulated in the parous breast. Genes with a blue border had a long-term effect, while genes with a yellow border were transiently upregulated. Pathway was modified from IPA software (QIAGEN) using Path Designer. (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* 2019;21:46.)

analyses of breast core biopsies of postmenopausal NP and P women[84,98-99] and identified 305 differentially expressed probe sets (208 genes) between P and NP women. Of these, 267 were upregulated, and 38 were downregulated. GO enrichment analysis revealed biological processes including RNA metabolic processes, differentiation and development of epidermis and ectoderm, and cell-substrate junction assembly, in agreement with pregnancy hormone–induced differentiation of mammary epithelial cells.[100] We found that genes pivotal to anchoring epithelial cells to the basement membrane, hemidesmosome, and cell-substrate junction assembly, such as *KRT5, LAMA3,* and *LAMC2*, were upregulated in the P group.

Among the downregulated genes, *IGF-1* was enriched in 19 biological processes that comprised cell proliferation, regulation of IGF-like growth factor receptor signaling, somatic stem cell maintenance, muscle cell differentiation, and apoptosis, among others.[98]

We used gene set enrichment analysis to compare differentially expressed genes from this study of 3717 curated gene sets of specific pathways, processes, and profiles from previous profiling experiments obtained through MsigDB.[100] Pathways enriched in upregulated genes included breast cancer estrogen signaling, cell communication, and mRNA processing.[99] The breast cancer estrogen signaling pathway encompassed a set of genes dysregulated in ER-dependent breast

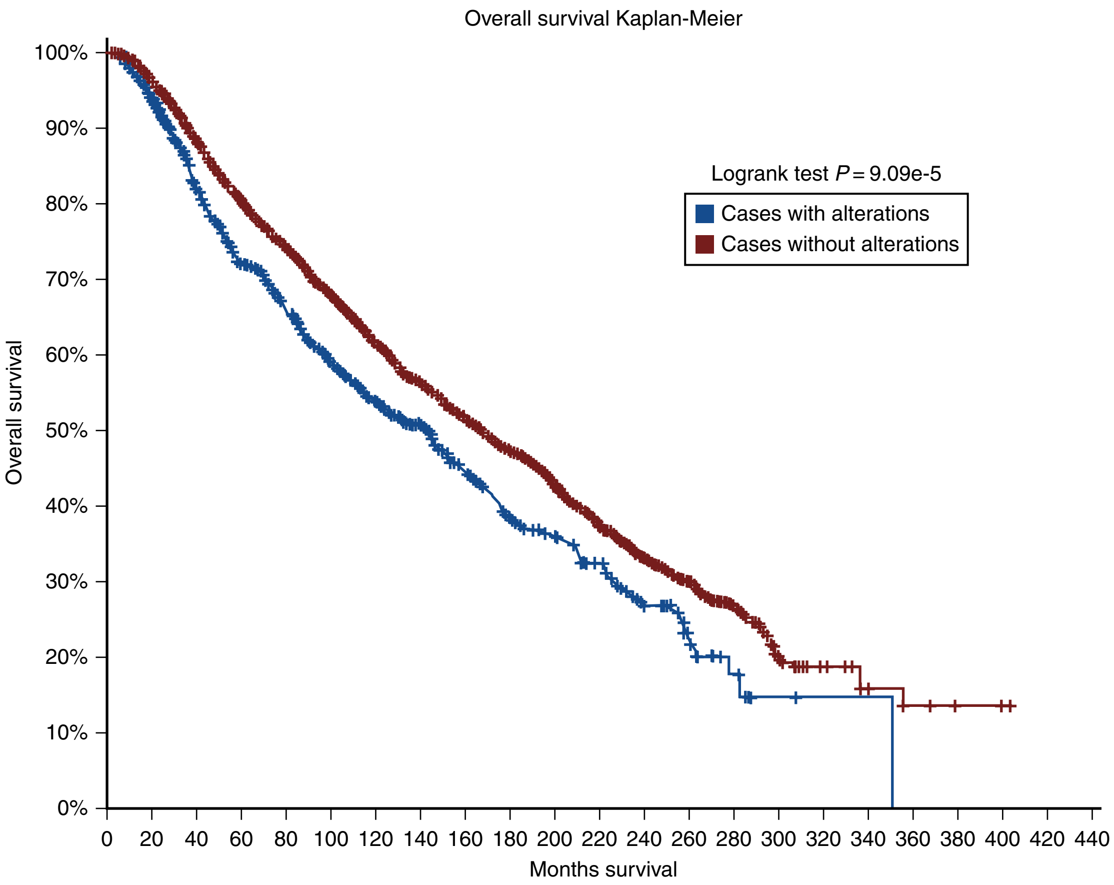


Fig. 127.10 Analyses of the parous differentially expressed genes in breast cancer patients. Genes that were selected for real-time polymerase chain reaction validation were also used for evaluation of breast cancer patients using the cBioPortal tool (http://www.cbioportal.org/). Overall Kaplan–Meier survival curve comparing cases with and without alterations in these genes. Patients who carried alterations had a median of 143.13 months' overall survival, while patients without alteration in these genes had median overall survival of 168.3 months (logrank test $P = 9.09e{-5}$). (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* 2019;21:46.)

cancers. Among these were *SCGB2A1, SCGB2A2, GATA3, TP53, TFF1, STC2, SerpinB5,* and *SerpinA3*. Because FTP involves influx of several hormones, including estrogens, we postulated that several downstream targets of estrogen would be coregulated in P subjects. Other pathways enriched by upregulated genes were cell communication and the mRNA processing reactome.[99] Among key pathways enriched by downregulated genes, the most significant were insulin signaling, MAPK, cytokine-cytokine receptor interaction, and Wnt signaling.[98,99] Using RT-PCR we validated numerous genes ($P < .05$) upregulated in P women, including *CCNL2*. IHC confirmed expression and localization of *CCNL2* in NP and P breasts,[99] with CCNL2 protein significantly overexpressed in the nucleus of epithelial cells of type 1 lobules of the P breast compared with similar structures in the breast of NP women. These observations confirm the localization of this protein in the splicing factor compartment (nuclear speckles).[101,102]

The main components of the spliceosome machinery, including RNA and proteins that undergo dynamic changes during the splicing reaction, were upregulated in the P breast. Among them were the heterogeneous nuclear ribonucleoproteins (HNRPs) that include HNRPA3, HNRPA2B1, HNRPD, and HNRPU,[103] which are implicated in the regulation of mRNA stability, as well as other functions, such as mammary gland involution,[104] negative regulation of telomere length maintenance,[105] and regulation of mRNA trafficking from the nucleus to distal processes in neural cells.[106] Although further studies are needed to define their precise functional roles in the postmenopausal breast, we postulate they play an important function as transcriptional regulators. In addition, posttranscriptional methylation of internal adenosine residues in eukaryotic mRNAs by methyltransferase like 3, which is upregulated in the P breast, could play a role in the efficiency of mRNA splicing, transport, or translation in the differentiated breast epithelium. Other members of the spliceosome complex are the proteins encoded by the genes *SF3B1, SFRS2, SFRS7, SFRS8, SFRS14, SFRS16, SNRP70, SNRPB, SNRPA1, PRF3*, and *PHF5A*, all of which are overexpressed in the P breast. In the case of the small nuclear ribonucleoproteins, there is evidence they suppress tumor cell growth and may have major implications as cancer therapeutic targets. Pre-mRNA splicing factors are enriched in nuclear domains termed interchromatin granule clusters or nuclear speckles. *CCNL1* and *CCNL2* are upregulated in the P breast, and the CCNL2 protein is also overexpressed in the nucleus of breast epithelial cells, as described earlier. CCNL1 and CCNL2 are transcriptional regulators[25-28] that modulate the expression of critical factors leading to cell apoptosis through the Wnt signal transduction pathway,[107-109] which is enriched by downregulated genes in the P breast. In our previously published preclinical and clinical studies,[39,99,110] we reported that pregnancy confers protection from breast cancer development by inducing gland differentiation, which imprints a specific and permanent genomic signature in this organ. A similar phenomenon was demonstrated in the involuted breast of postmenopausal P women.[100] We previously reported a small case-control study of transcriptomic analysis of normal breast tissues obtained from P and

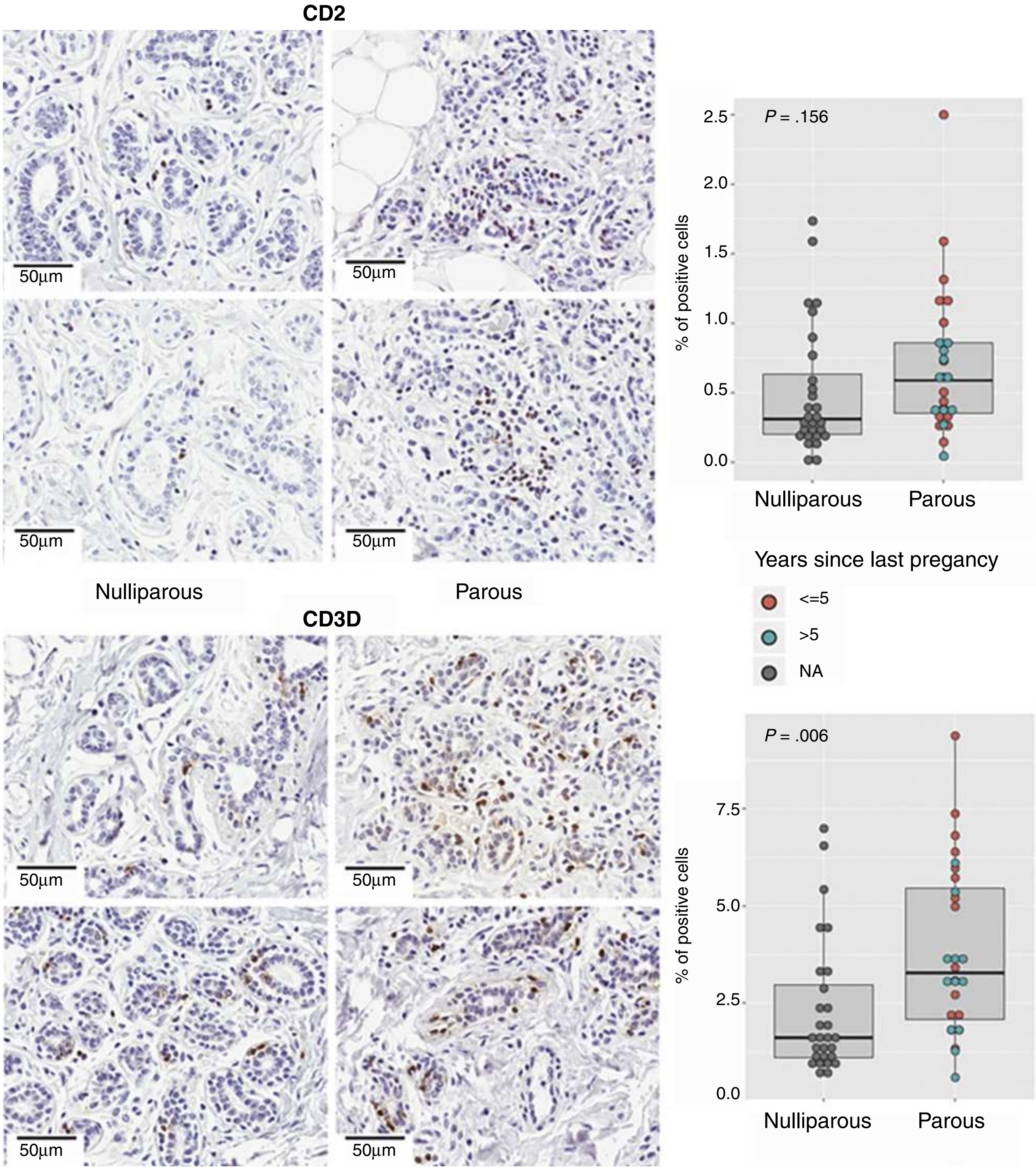

CD2
50µm
50µm
50µm
50µm
Nulliparous
Parous
CD3D
50µm
50µm
50µm
50µm
P = .156
% of positive cells
2.5
2.0
1.5
1.0
0.5
0.0
Nulliparous
Parous
Years since last pregancy
<=5
>5
NA
P = .006
% of positive cells
7.5
5.0
2.5
0.0
Nulliparous
Parous

Fig. 127.11 Immunohistochemical evaluation of CD2 and CD3D for T-cell activation. On the top are results regarding CD2, and on the bottom are the results for CD3D. The panels on the left show the imunohistochemical re=actions (×400): each panel contains two nulliparous samples *(left)* and two parous samples *(right)*. On the right, boxplots show the percentage of positive cells. We observed an increase in the percentage of CD3D-positive cells in the parous group (*P*= .006). (From Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* 2019;21:46.)

NP women free of breast pathology and P and NP women with history of breast cancer that served as controls and cases, respectively.[84] To investigate commonality between the previous case-control and the subsequent study, we applied GO enrichment analysis to the gene lists generated from both studies and found that processes involved in RNA metabolism and RNA processing were similar in both studies.[99]

A number of noncoding RNAs, including XIST, MALAT-1 (NEAT2), and NEAT1, were upregulated in the P breast. Upregulation of XIST occurs upon cell differentiation, whereas failure to express XIST is seen in malignancies and in early embryogenesis.[111] Our findings are supported by recent reports suggesting that XIST is expressed in well-differentiated adult cells to maintain gene repression.[111-115]

Oxytocin, a neurotransmitter that acts through its specific receptor OXTR and is overexpressed during lactation, upregulates expression of MALAT-1, a highly conserved noncoding RNA.[116,117] Interestingly, both MALAT1 and OXTR remain overexpressed in the breast of postmenopausal P women. NEAT1 and NEAT2 localize to the periphery and to the interior of the spliceosome assembly factor SC35 domains or speckles. Our observation that, in breast epithelial cells, CCNL2 is highly enriched in nuclear speckles, indicates that CCNL2 may colocalize with NEAT1 and NEAT2. Downregulation of NEAT1, NEAT2, and XIST in the breast of NP women, in whom this organ never reaches a stage of complete differentiation similar to that achieved after completion of pregnancy and lactation, suggests that the undifferentiated breast is not actively involved in the RNA metabolism that is necessary for maintaining a state of differentiation.

Several genes directly or indirectly regulated by ER were up- or downregulated in the P breast and were enriched in the breast cancer estrogen signaling gene set. Among them, *GATA3*, an important component of this set, is crucial to mammary gland morphogenesis and differentiation of progenitor cells. GATA3 has been suggested to be a tumor suppressor,[118] a fact that is supported by observations that induction of its expression in *GATA3*-negative undifferentiated carcinoma cells is sufficient to induce tumor differentiation and inhibition of tumor dissemination.[118-121] Downregulation of RAS, dexamethasone-induced 1 (RASD1), a potential miR-375 target that negatively regulates ERa expression in breast cancer,[122] further confirms that the genes involved in the ER-regulated pathways could be under permanent transcriptional modification as a manifestation of a higher degree of cell differentiation of the P breast, in spite of the lack of transcriptomic differences in this receptor's levels between P and NP breast tissues.

Cell communication, a key element in cell and organ differentiation, is well represented in the breast of P women. The P breast exhibits upregulation of desmocollin (DSC3), a calcium-dependent glycoprotein that is a member of the desmocollin subfamily of the cadherin superfamily. These members of the desmosomal family, along with the desmogleins, are found primarily in epithelial cells, where they constitute the adhesive proteins of the desmosome cell-cell junction and are required for cell adhesion and desmosome formation. In addition, upregulation of matrix Gla proteins (MGP), laminins (LAMA3 and LAMC2), and keratin 5 (KRT5) in the P breast is an indication that they reflect a greater differentiated state of the breast epithelial cells. This concept is supported by the observation that the loss of MGP expression is associated with tumor progression and metastasis.[123,124]

Our findings that IGF-1 is downregulated in the P breast are consistent with published data reporting overall lower levels of IGF-1 in P than in NP women[125] and support the association of IGF-1 with increased breast cancer risk.[126] It is known that IGF-1 stimulates mitosis and inhibits apoptosis, playing a significant role in signaling pathways involved in the pathogenesis of breast cancer. Downregulation of IGF-1 in the P breast, in association with the significant downregulation of SOX6, early B-cell factor 1 (EBF1), ABHD5, RASD1, and RALGAPA2, could represent a significant driving force in the reduction of breast cancer risk conferred by pregnancy.

Chromatin Remodeling and Prevention of Human Breast Cancer

We have demonstrated, using detailed histological, cytological, IHC, and transcriptomic analyses of breast samples from NP and P postmenopausal women, that differentiation of the breast induced by an early pregnancy imprints a specific phenotypic and genotypic signature that can be detected in postmenopausal women.[127,128] Our observations that, during the postmenopause, the breast of both P and NP women contains preponderantly lob 1, and the fact that NP women are at higher risk of developing breast cancer than P women indicate that lob 1 in these two groups of women either differ biologically or exhibit different susceptibility to carcinogenesis.[127] The breast tissues of P and NP women contain ducts and lob 1, which have been reported to contain 11.2 ± 6.3 ductules per lobular unit. Each ductule is composed of an external layer of myoepithelial cells and a monolayer of cuboidal epithelial cells lining a lumen that contains proteinaceous material. IHC staining revealed that epithelium lining both ducts and ductules consisted of basal cells showing positive reactivity for keratin 5/6 and myoepithelial cells reacting positively with SMA; these two markers had similar reactivity in both P and NP breasts. Hematoxylin and eosin-stained tissues reveal two populations of luminal cells lining ducts and lob 1 characterized by their nuclear appearance: one containing large and palely stained nuclei with prominent nucleoli, and another containing small, hyperchromatic nuclei. The pale staining of the large nuclei in the former population is a feature indicative of a high content of noncondensed euchromatin; these nuclei were called euchromatin-rich nuclei (EUN) or Stem Cell 1. The hyperchromasia observed in the latter nuclei was indicative of chromatin condensation and high content of heterochromatin; these nuclei were identified as heterochromatin-rich nuclei (HTN) or Stem Cell 2.[128] The analysis of the distribution of HTN and EUN cells in histological sections of the breast core biopsies revealed that EUN are more abundant in NP than in P breast tissues, whereas the inverse was true for the HTN; these differences were statistically significant[128] (Fig. 127.12). We have confirmed the differences between the HTN and EUN using a quantitative image analysis system. The nuclear size (diameter, area, and perimeter) of the EUN as a whole was significantly higher ($P < .05$) than that of the HTN in both NP and P women.[119] Differences were also found to be statistically significant ($P < .05$) regarding the nuclear shape (nuclear feret ratio) in the breast of NP women, indicating that, in these breasts, the nuclei of the HTN had a more elongated ellipsoidal shape than the EUN. The light absorbance (mean gray values/nucleus) was always greater for EUN than for HTN of both NP and P breasts, either considered as two groups or individually, an indication that, in densitometric terms, HTN are always more densely stained than EUN. Comparison of the EUN of NP versus P breasts revealed significant differences in nuclear size, stainability, and densitometric energy, leading us to conclude that epithelial cell nuclei were larger, less stainable, and with smaller regions with uniform densitometric intensity in NP breasts. Comparison of the HTN of NP versus P breasts revealed significant differences in nuclear diameter, perimeter, shape and stainability; cell nuclei showed larger contours and more elongated ellipsoidal shape and they were more stainable in NP breasts. These observations indicated that a shift of the EUN cell population to a more densely packed chromatin cell (HTN) had occurred in association with the history of pregnancy as a distinctive pattern of the postmenopausal P breast.[127]

Because chromatin condensation is part of the process of chromatin remodeling towards gene silencing that is highly regulated by methylation of histones, we verified this phenomenon by IHC, incubating NP and P breast tissues with antibodies against histone 3 dimethylated at lysine 9 (H3K9me) and trimethylated at lysine 27 (H3K27me[3]). The IHC staining revealed that methylation of H3 at both lysine 9 and 27 was increased in the HTN epithelial cells of the P breast[128] compared with the EUN of the NP breast. In the NP breast the reactivity in individual cells was less intense, and the number of positive cells was significantly lower (Fig. 127.13). These variations in chromatin reorganization were supported by the upregulation of the *CBX3*, *CHD2*, *L3MBTL*, and *EZH2* genes controlling this process.[128]

Transcription regulation and chromatin organization were also highly represented in the P breast by the upregulation of *CBX3*, *EBF1*, *GATA3*, *RBBP8*, *CCNL1*, *CCNL2*, *CDCA7*, *EZH2*, *FUBP1*, *NFKBIZ*,

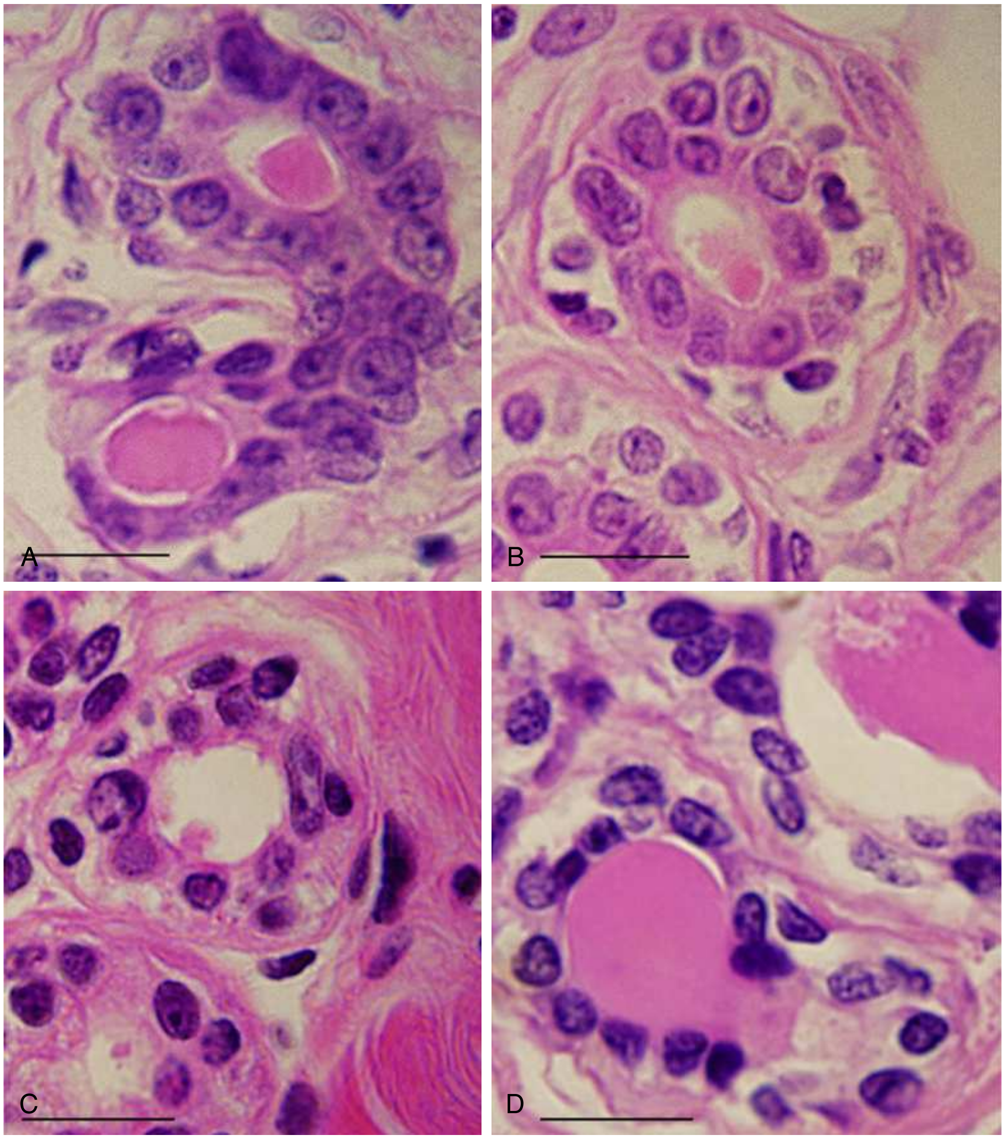


Fig. 127.12 Histologic sections of hematoxylin and eosin–stained ductules in type 1 lobules (Lob 1). **A** and **B**, from nulliparous, and **C** and **D**, from parous women's breast tissues; Magnification bar: 100 μm. (From Russo J, Santucci-Pereira J, De Cicco Lopez R, et al. Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int J Cancer.* 2012;131:1059–1070.)

RUNX3, *ZNF107*, *ZNF207*, *ZNF692*, *ZNF711*, *ZNF789*, *CDCA7*, and *ZNF692.*[128] The P breast also exhibits upregulation of expression of six noncoding regions that included *XIST*, *MALAT-1* (*NEAT2*), and *NEAT1*[127] (Fig. 127.14).

KEY POINTS

- Transcriptionally active chromatin is mianly expressed in the euchromatin-rich nuclei of the breast of nulliparous women.
- Transcriptionally inactive chromatin is more frequently found in heterochromatin-rich nuclei of the parous breast and is associated with histone 3 methylation at lysines 9 and 27, as well as transcriptional silencing in heterochromatin complexes.
- Mechanisms for decreased risk of developing breast cancer in parous versus nulliparous women are complex and likely involve molecular, structural, genomic, and epigenomic changes that have the potential to be mined for future breast cancer prevention strategies.

We demonstrated that the changes in cell types and increases in chromatin condensation are novel markers for defining the concept of differentiation in the adult breast. These findings confirm the universality of histone 3 methylation at lysines 9 and 27 during differentiation, as a similar phenomenon has been described to occur during embryonic stem cell differentiation.[129,130] The observed chromatin changes in P epithelial cells are complemented by the expression of genes related to increasing cell adhesion and differentiation. In contrast to the findings of Asztalos et al.[131] of downregulation of the expression of ERa following recent (<2 years) and distant (5–10 years) pregnancies in premenopausal women, our current genomic and IHC study did not reveal differences in ERa expression in epithelial cells of ducts and lob 1 between P and NP postmenopausal women. Nevertheless, numerous genes that are regulated downstream by the ERa were found to be upregulated in the P breast, supporting a parity-mediated protective effect evident in younger P women[129] but lasting until menopause, as described earlier.

Studies of breast development under the influence of parity in women and in animal models are in agreement on the pregnancy-induced

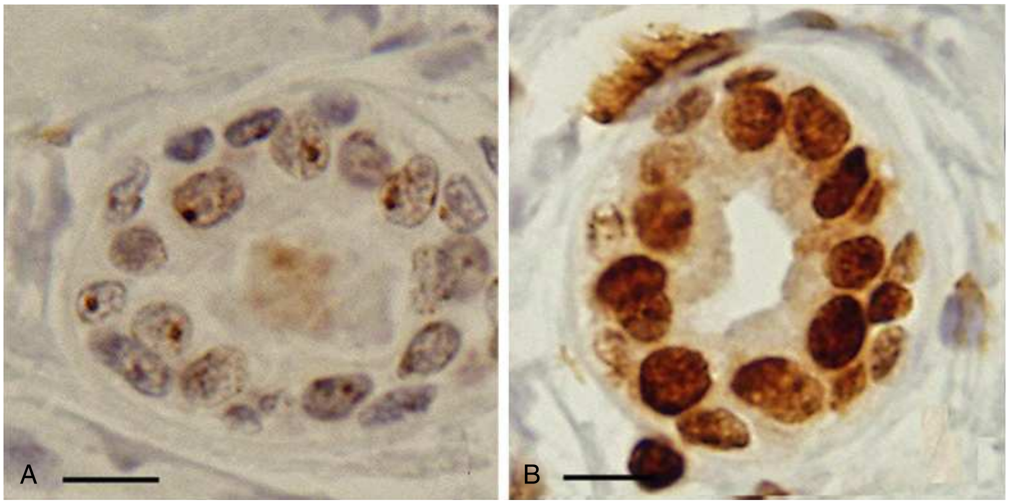


Fig. 127.13 The immunohistochemical reaction for H3K27(me^3) is of higher (+) intensity in the nuclei of the parous (**B**) breast than of the nonparous breast (**A**), in which the nuclear stain is faint and finely granular, being mostly circumscribed by the nucleoli. DAB-H counterstain. Magnification bar: 100 μm. (From Russo J, Santucci-Pereira J, De Cicco Lopez R, et al. Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int J Cancer.* 2012;131:1059–1070.)

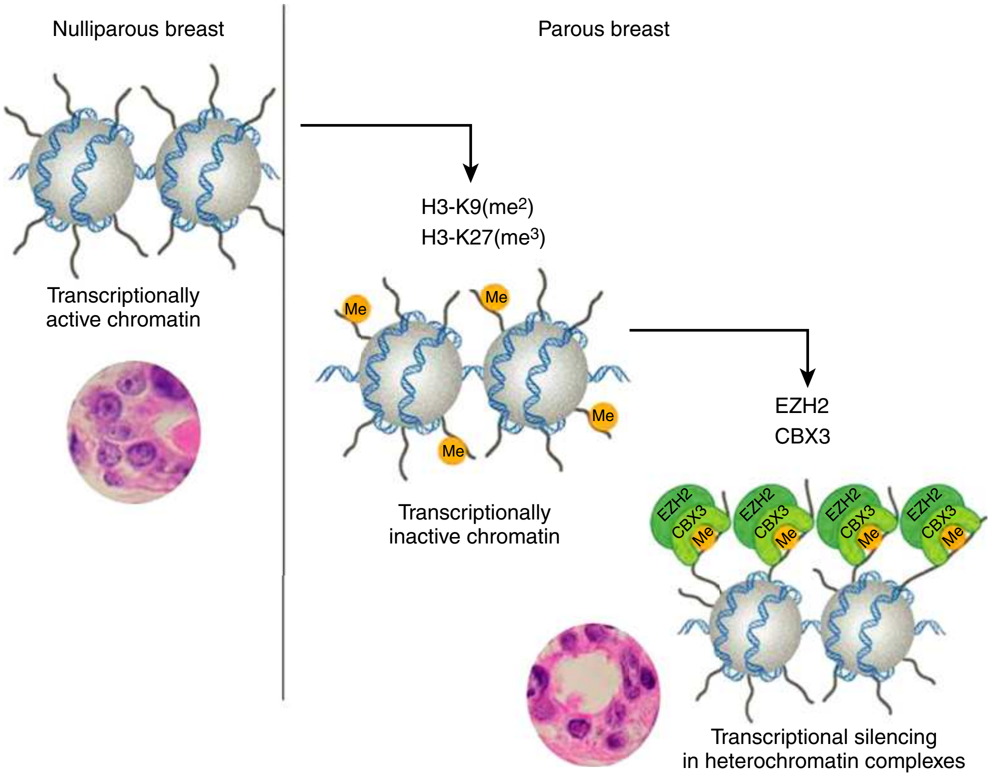

Nulliparous breast
Parous breast
H3-K9(me2)
H3-K27(me3)
Me
Transcriptionally active chromatin
Transcriptionally inactive chromatin
EZH2
CBX3
Transcriptional silencing in heterochromatin complexes

Fig. 127.14 Transcriptionally active chromatin is predominantly expressed in the euchromatin-rich nuclei of the breast of nulliparous women. Transcriptionally inactive chromatin is more frequently found in the heterochromatin-rich nuclei of the parous breast; its presence is associated with histone 3 methylation at lysines 9 and 27, as well as transcriptional silencing in heterochromatin complexes. (From Russo J, Santucci-Pereira J, De Cicco Lopez R, et al. Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int J Cancer.* 2012;131:1059–1070.)

differentiation of the breast, a process that ultimately manifests as a specific genomic signature in the mammary gland.[84,97,131-134] Although variations in gene expression among different studies and species are expected, an increase in immune activity, including overexpression of lipopolysaccharide-binding protein has been reported in the postpregnancy breast of premenopausal women[131] and in the mammary gland of four different strains of rats.[97] Interestingly, this response observed in recently pregnant and distant pregnant groups was not observed in the postmenopausal group in this study. These discrepancies may indicate that upregulation of immune response–related genes persists during postpartum involution but wanes after menopause begins. However, the possibility that cyclic hormonal changes occurring during the menstrual/estrus cycle that influence the genomic profile of the breast activating pathways that greatly differ from those expressed in the postmenopausal P women cannot be ruled out.

We found a shift in the cell population of the postmenopausal breast as a manifestation of the reprogramming of the organ after pregnancy. These observations are in agreement with what is observed in the rat mammary gland, which also contains two types of luminal epithelial cells, designated dark (DC) and intermediate (IC) cells, in addition to the myoepithelial cells (see earlier).[132] The DC and IC cells are equivalent to the HTN and EUN cells described earlier. DCs increase after pregnancy and lactational involution (Stem Cell 2); whereas the ICs (Stem Cell 2) significantly outnumber the DCs in ductal hyperplasias and ductal carcinomas.[134,135] Our analysis of nuclear ultrastructural and morphometric parameters of rodent ICs have allowed the differentiation of mammary progenitor stem cells from cancer stem cells.[128,134,135] Nuclear morphometric analysis of breast and ovarian carcinomas has confirmed the predictive value of nuclear grade for the progression of premalignant lesions to invasiveness.[136-138] Our present findings of a significant decrease in the number of EUN (Stem Cell 1) with a subsequent increase in the number of HTN cells (Stem Cell 2) expressing specific biomarkers identified at the chromatin and transcriptional levels support the value of morphometric analysis as an adjunct to molecular studies. These data clearly indicate there are morphological indications of chromatin remodeling in the P breast, such as the increase in the number of epithelial cells with condensed chromatin and increased reactivity with anti-H3K9me and anti-H3K27me[3] antibodies. Histone methylation is a major determinant for the formation of active and inactive regions of the genome and is crucial for the proper programming of the genome during development.[139] In the P breast, there is upregulation of transcription factors and chromatin remodeling genes such as *CHD2* and *CBX3*, whose products are required for controlling protein/protein or DNA/protein interactions. *CBX3* is involved in transcriptional silencing in heterochromatin-like complexes and recognizes and binds H3 tails methylated at lysine 9, leading to epigenetic repression (see Fig. 127.14). Two other important genes related to the polycomb group (PcG) protein that are upregulated in the P breast are *L3MBTL* and *EZH2*. Members of the PcG form multimeric protein complexes that maintain the transcriptional repressive state of genes over successive cell generations. *EZH2* encodes an enzyme that acts mainly as a gene silencer, performing this role by the addition of three methyl groups to lysine 27 of histone 3, a modification that leads to chromatin condensation.[128,138-141]

RNA molecules recruit PcG complexes to the locus of transcription or to sites located elsewhere in the genome. An important role has been attributed to noncoding RNAs (ncRNAs).[142] It can be postulated that the increased chromatin condensation in the P breast reported herewith could have been initiated by ncRNAs, which is supported by the observed upregulation of several ncRNAs that included nuclear paraspeckle assembly transcript 1 (*NEAT1*), *MALAT-1* (*NEAT2*), and *XIST*,[143,144] all critical components of the speckles. These observations indicate that, in P women, the breast remains actively involved in the RNA metabolism that is necessary to maintain a state of differentiation. A ncRNA transcribed from a portion of the *XIST* gene locus forms hairpin structures that recruit the PRC2 complex to the X-inactivation center *X(ic)*,[145] and transcription of full-length *XIST* RNA, which forms the same hairpin structures, leads to further PRC2 recruitment and the spread of PcG-mediated repression across the inactive X chromosome. *XIST* repression, often seen in malignancies, also occurs in early embryogenesis and during the acquisition of pluripotency in undifferentiated embryonic stem cells by the binding of *NANOG*, *OCT4*, and *SOX2* directly to the chromatin of the *XIST* gene.[110] It is possible that, in the postmenopausal NP breast, the upregulation of *DDX*, *SOX1*, *SOX6*, and *SOX17* play a direct pivotal role in the repression of *XIST* transcription. Although this needs to be functionally verified, it is possible that these genes play a role in controlling *XIST* in the P breast. Upregulation of *XIST* has important implications in the understanding of the differentiation pattern of the P breast. In recent studies it has been shown that the reprogramming of X-chromosome inactivation during the acquisition of pluripotency *in vivo* and *in vitro* is accompanied by the repression of *XIST*.[141-143] Reprogramming experiments have further reinforced the concept that X-chromosome inactivation is intimately linked to differentiation and support our findings that *XIST* is expressed in well-differentiated adult cells, participating in the maintenance of gene repression.[112,143-146]

Our work demonstrates that the breast of P postmenopausal women exhibits a specific signature that has been induced by a FTP. This signature reveals, for the first time, that the differentiation process is centered on chromatin remodeling and the mRNA processing reactome, which emerge as important regulatory pathways induced by pregnancy. The biological importance of the pathways identified in this specific population cannot be sufficiently emphasized, because of the fact that the upregulation of the noncoding sequences that control gene repression and of the genes that control the spliceosomes could represent a safeguard mechanism at genomic and posttranscriptional levels that maintain the fidelity of the transcription process, a phenomenon that could be the ultimate step mediating protection of the breast from malignant transformation conferred by FTP.

REFERENCES

1. Edwards RG, Howles CM, Macnarnee C. Clinical endocrinology of reproduction. In: Baulieu E-E, Kelly PA, eds. *Hormones: From Molecules to Disease*. New York and London: Chapman and Hall; 1990:457–476.
2. *Chapter 105 in this 8th edition of DeGroots: Endocrinology of Sexual Maturation and Puberty by Styne and Colleagues*.
3. Russo J, Russo IH. Development of the human mammary gland. In: Neville MC, Daniel C, eds. *The Mammary Gland*. New York: Plenum; 1987:67–93.
4. Brisken J, O'Malley B. Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol*. 2010;2:a003178.
5. Tanner JM. *Growth at Adolescence; with a General Consideration of the Effects of Hereditary and Environmental Factors upon Growth and Maturation from Birth to Maturity*. 2nd edition. Oxford: Blackwell Scientific Publications; 1962.
6. Biro FM, Greenspan LC, Galvez MP, et al. Onset of breast development in a longitudinal cohort. *Pediatrics*. 2013;132(6):1019–1027.
7. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*. 2009;123(1):84–88.
8. Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility to carcinogenesis. *Breast Cancer Res Treat*. 1982;2:5–73.
9. Russo J, Rivera R, Russo IH. Influence of age and parity on the development of the human breast. *Breast Cancer Res Treat*. 1992;23:211–218.
10. Russo J, Russo IH. Role of hormones in human breast development: the menopausal breast. In: Wren BG, ed. *Progress in the Management of the Menopause*. London: CRC Press; 1997:184–193.
11. Russo IH, Russo J. Role of hormones in cancer initiation and progression. *J Mammary Gland Biol Neoplasia*. 1998;3:49–61.
12. Russo J, Russo IH. Role of differentiation in the pathogenesis and prevention of breast cancer. *Endocr Relat Cancer*. 1997;4:7–21.
13. Calaf G, Alvarado ME, Bonney GE, et al. Influence of lobular development on breast epithelial cell proliferation and steroid hormone receptor content. *Int J Oncol*. 1995;7:1285–1288.
14. Lippman ME, Dickson RB, Gelmann EP, et al. Growth regulation of human breast carcinoma occurs through regulated growth factor secretion. *J Cell Biochem*. 1987;35:1–16.
15. Meyer JS. Cell proliferation in normal human breast ducts, fibroadenomas, and other duct hyperplasias, measured by nuclear labeling with tritiated thymidine. *Hum Pathol*. 1977;8:67–81.

16. Masters JRW, Drife JO, Scarisbrick JJ. Cyclic variations of DNA synthesis in human breast epithelium. *J Natl Cancer Inst*. 1977;58:1263–1265.
17. Ferguson DJP, Anderson TJ. Morphologic evaluation of cell turnover in relation to the menstrual cycle in the "resting" human breast. *Br J Cancer*. 1981;44:177–181.
18. Russo J, Russo IH. Estrogens and cell proliferation in the human breast. *J Cardiovasc Pharmacol*. 1996;28:19–23.
19. Kumar V, Stack GS, Berry M, et al. Functional domains of the human estrogen receptor. *Cell*. 1987;51:941–951.
20. King RJB. Effects of steroid hormones and related compounds on gene transcription. *Clin Endocrinol*. 1992;36:1–14.
21. Soto AM, Sonnenschein C. Cell proliferation of estrogen-sensitive cells: the case for negative control. *Endocr Rev*. 1987;48:52–58.
22. Huseby RA, Maloney TM, McGrath CM. Evidence for a direct growth-stimulating effect of estradiol on human-MCF-7 cells in vitro. *Cancer Res*. 1987;144:2654–2659.
23. Huff KK, Knabbe C, Lindsey R, et al. Multihormonal regulation of insulin-like growth factor-1-related protein in MCF-7 human breast cancer cells. *Mol Endocrinol*. 1988;2:200–208.
24. Dickson RB, Huff KK, Spencer EM, et al. Introduction of epidermal growth factor related polypeptides by 17beta-estradiol in MCF-7 human breast cancer cells. *Endocrinology*. 1986;118:138–142.
25. Katzenellenbogen BS, Kendra KL, Norman MJ, et al. Proliferation, hormonal responsiveness and estrogen receptor content of MCF-7 human breast cancer cells growth in the short-term and long-term absence of estrogens. *Cancer Res*. 1987;47:4355–4360.
26. Petersen OW, Hoyer PE, van Deurs B. Frequency and distribution of estrogen receptor positive cells in normal, nonlactating human breast tissue. *Cancer Res*. 1987;47:5748–5751.
27. Jacquemier JD, Hassouin J, Torente M, et al. Distribution of estrogen and progesterone receptors in healthy tissue adjacent to breast lesions at various stages: immunohistochemical study of 107 cases. *Breast Cancer Res Treat*. 1990;15:109–117.
28. McGuire W, Carbone P, Vollmer R, eds. *Estrogen Receptors in Human Breast Cancer*. New York: Raven Press; 1975.
29. Dickson R, Lippman M. Control of Human Breast Cancer by Estrogen, Growth Factors and Oncogenes. In: Lipman ME, Dickson RB, eds. *Estrogen Receptors in Human Breast Cancer*. New York: Raven Press; 1975.
30. Wittliff IL. Steroid-hormone receptors in breast cancer. *Cancer*. 1984;53:630–643.
31. Watts CKW, Handel ML, King RJB, et al. Oestrogen receptor gene structure and function in breast cancer. *J Steroid Biochem Mol Biol*. 1992;41:5293–5536.
32. Aakvaag A, Utaacker E, Thorsen T, et al. Growth control of human mammary cancer cells (MCF-7 cells) in culture: effect of estradiol and growth factors in serum containing medium. *Cancer Res*. 1990;50:7806–7810.
33. Dell'aquilla ML, Pigott DA, Bonaquist DL, et al. A factor from plasma derived human serum that inhibits the growth of the mammary cell line MCF-7: characterization and purification. *J Natl Cancer Inst*. 1984;72:291–298.
34. Kuiper GG, Enmark E, Pelto-Huikko M, et al. Cloning of a novel estrogen receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci USA*. 1996;93:5925–5930.
35. Byers M, Kuiper GGJM, Gustaffson J-A, et al. Estrogen receptor beta mRNA expression in rat ovary down-regulation by gonadotropins. *Mol Endocrinol*. 1997;11:172–182.
36. Vladusic EA, Homby AE, Guerra-Vladusic FK, et al. Expression of estrogen receptor beta messenger RNA variant in human breast cancer. *Cancer Res*. 1998;58:210–214.
37. Lubahn DB, Moyer JS, Golding TS, et al. Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. *Proc Soc Natl Acad Sci USA*. 1993;90:11162–11166.
38. Bocchinfuso WP, Korach KS. Mammary gland development and tumorigenesis in estrogen receptor knockout mice. *J Mammary Gland Biol Neoplasia*. 1997;2:323–334.
39. Russo J, Russo IH. Influence of differentiation and cell kinetics on the susceptibility of the rat mammary gland to carcinogenesis. *Cancer Res*. 1980;40:2677–2687.
40. Russo J, Romero AL, Russo IH. Architectural pattern of the normal and cancerous breast under the influence of parity. *Cancer Epidemiol Biomarkers Prev*. 1994;3:219–224.
41. Russo J, Russo IH. Biology of disease: biological and molecular bases of mammary carcinogenesis. *Lab Invest*. 1987;57:112–137.
42. Clark RB, Howell A, Potter CS, et al. Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res*. 1997;57:4987–4991.
43. Clarke R, Dickson RB, Lippman ME. Hormonal aspects of breast cancer: growth factors, drugs and stromal interactions. *Crit Rev Oncol Hematol*. 1992;12:1–23.
44. Knabbe C, Lippman ME, Wakefield LM, et al. Evidence that transforming growth factor beta is a hormonally regulated negative growth factor in human breast cancer cells. *Cells*. 1987;48:417–428.
45. Russo J, Reina D, Frederick J, et al. Expression of phenotypical changes by human breast epithelial cells treated with carcinogens In Vitro. *Cancer Res*. 1988;48:2837–2857.
46. Russo J, Calaf G, Russo IH. A critical approach to the malignant transformation of human breast epithelial cells. *Crit Rev Oncog*. 1993;4:403–417.
47. Russo J, Gusterson BA, Rogers A, et al. Biology of the disease: comparative study of human and rat mammary tumorigenesis. *Lab Invest*. 1990;62:244–278.
48. Habel LA, Stamford JL. Hormone receptors and breast cancer. *Epidemiol Rev*. 1993;15:209–219.
49. Harlan LC, Coates RJ, Block G. Estrogen receptor status and dieting intakes in breast cancer patients. *Epidemiology*. 1993;4:25–31.
50. Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. *J Natl Cancer Inst*. 1980;65:559–569.
51. Kodama F, Green GL, Salmon SE. Relation of estrogen receptor expression to clonal growth and antiestrogen effects on human breast cancer cells. *Cancer Res*. 1985;45:2720–2724.
52. Podhajcer OL, Bravo-AL SL, et al. Determination of DNA synthesis, estrogen receptors, and carcinoembryonic antigen in isolated cellular subpopulations of human breast cancer. *Cancer*. 1986;58:720–729.
53. Hu YF, Lau KM, Ho SM, et al. Increased expression of estrogen receptor beta in chemically transformed human breast epithelial cells. *Int J Oncol*. 1998;12:1225–1228.
54. Foster JS, Wimalasena J. Estrogen regulates activity of cyclin-dependent kinases and retinoblastoma protein phosphorylation in breast cancer cells. *Mol Endocrinol*. 1996;10:488–498.
55. Wang W, Smith R, Burghardt R, et al. 17β-estradiol-mediated growth inhibition of MDA-MB 468 cells stably transfected with the estrogen receptor: cell cycle effects. *Mol Cell Endocrinol*. 1997;133:49–62.
56. Izvenson AS, Jordan VC. Transfection of human estrogen receptor (ER) CDNA into ER negative mammalian cell lines. *J Steroid Biochem Mol Biol*. 1994;51:229–239.
57. Weisz A, Bresciani F. Estrogen regulation of proto-oncogenes coding for nuclear proteins. *Crit Rev Oncog*. 1993;4:361–388.
58. Zajchowski DA, Sager R, Webster L. Estrogen inhibits the growth of estrogen receptor negative, but not estrogen receptor-positive, human mammary epithelial cells expressing a recombinant estrogen receptor. *Cancer Res*. 1993;53:5004–5011.
59. Pilat MJ, Christman JK, Brooks SC. Characterization of the estrogen receptor transfected MCF-10A breast cell line 139B6. *Breast Cancer Res Treat*. 1996;37:253–266.
60. Calaf G, Tahin Q, Alvarado ME, et al. Hormone receptors and cathepsin D levels in human breast epithelial cells transformed by chemical carcinogens. *Breast Cancer Res Treat*. 1993;29:169–177.
61. Aronica SM, Kraus WL, Katzenellenbogen BS. Estrogen action via the CAMP signaling pathway: stimulation of adenylate cyclase and CAMP regulated gene transcription. *Proc Natl Acad Sci USA*. 1994;91:8517–8521.
62. Pappos TC, Gametahu B, Watson CS. Membrane estrogen receptors identified by multiple antibody labeling and impeded-ligand binding. *FASEB J*. 1994;9:404–410.
63. Deome KB, Faulkin Jr LJ, Bern HA, et al. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res*. 1959;19:515–520.

64. Kordon EC, Smith GH. An entire functional mammary gland may comprise the progeny from a single cell. *Development*. 1998;125:1921–1930.
65. Russo J, Saby J, Isenberg WM, et al. Pathogenesis of mammary carcinomas induced in rats by 7,12-dimethylbenz(a)anthracene. *J Natl Cancer Inst*. 1977;59:436–445.
66. Russo J, Isenberg W, Ireland M, et al. Ultrastructural changes in the mammary epithelial cell population during neoplastic development induced by a chemical carcinogen. *Electron Micro Soc Am Proc*. 1976:250–251.
67. Russo J, Tait L, Russo IH. Susceptibility of the mammary gland to carcinogenesis. III. The cell of origin of rat mammary carcinoma. *Am J Pathol*. 1983;113:50–66.
68. Bennett DC, Peachey LA, Durbin H, et al. A possible mammary stem cell line. *Cell*. 1978;15:283–298.
69. Rudland PS, Bennett DC, Warburton MJ. Isolation and characterization of epithelial stem-cell cell lines from the rat mammary gland. *Br J Cancer*. 1980;41:666–668.
70. Chepko G, Smith GH. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell*. 1997;29:239–253.
71. Smith GH, Mehrel T, Roop DR. Differential keratin gene expression in developing, differentiating, preneoplastic, and neoplastic mouse mammary epithelium. *Cell Growth Differ*. 1990;1:161–170.
72. Stingl J, Eaves CJ, Kuusk U, et al. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. *Differentiation*. 1998;63:201–213.
73. Stingl J, Eaves CJ, Zandieh I, et al. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat*. 2001;67:93–109.
74. Hebbard L, Steffen A, Zawadzki V, et al. CD44 expression and regulation during mammary gland development and function. *J Cell Sci*. 2000;113:2619–2630.
75. Kenney NJ, Smith GH, Lawrence E, et al. Identification of stem cell units in the terminal end bud and duct of the mouse mammary gland. *J Biomed Biotechnol*. 2001;1:133–143.
76. Spangrude GJ, Aihara Y, Weissman IL, et al. The stem cell antigens Sca-1 and Sca-2 subdivide thymic and peripheral T lymphocytes into unique subsets. *J Immunol*. 1988;141:3697–3707.
77. Welm BE, Tepera SB, Venezia T, et al. Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev Biol*. 2002;245:42–56.
78. Zhou S, Schuetz JD, Bunting KD, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med*. 2001;7:1028–1034.
79. Alvi AJ, Clayton H, Joshi C, et al. Functional and molecular characterisation of mammary side population cells. *Breast Cancer Res*. 2003;5:R1–R8.
80. Gudjonsson T, Villadsen R, Nielsen HL, et al. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev*. 2002;16:693–706.
81. Dontu G, Abdallah WM, Foley JM, et al. *In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev*. 2003;17:1253–1270.
82. Jia Y, Qi C, Zhang Z, et al. Peroxisome proliferator-activated receptor-binding protein null mutation results in defective mammary gland development. *J Biol Chem*. 2005;280:10766–10773.
83. Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res*. 2019;28:21–46.
84. Russo J, Balogh GA, Russo IH, et al. Full term pregnancy induces a specific genomic signature in the human breast. *Cancer Epidemiol Biomarkers Prev*. 2008;17:51–66.
85. Balogh GA, Heulings R, Mailo DA, et al. Genomic signature induced by pregnancy in the human breast. *Int J Oncol*. 2006;28:399–410.
86. Russo J, Balogh GA, Heulings R, et al. Molecular basis of pregnancy induced breast cancer protection. *Eur J Cancer Prev*. 2006;15:306–342.
87. Russo J, Balogh GA, Chen J, et al. The concept of stem cell in the mammary gland and its implication in morphogenesis, cancer and prevention. *Front Biosci*. 2006;11:151–172.
88. Russo J, Russo IH. Role of differentiation in the pathogenesis and prevention of breast cancer. *Endocr Rel Cancer*. 1997;4:7–21.
89. Russo J, Mailo D, Hu YF, et al. Breast differentiation and its implication in cancer prevention. *Clin Cancer Res*. 2005;11:931s–936s.
90. Clarke CA, Purdie DM, Glaser S. Population attributable risk of breast cancer in white women associated with immediately modifiable risk factors. *BMC Cancer*. 2006;6:170–181.
91. MacMahon B, Cole P, Lin TM, et al. Age at first birth and breast cancer risk. *Bull World Health Organ*. 1970;43:209–221.
92. Russo J, Russo IH. Endocrine Control of Breast Development. In: Russo J, Russo IH, eds. *Molecular Basis of Breast Cancer: Prevention and Treatment*. 1st ed. Berlin: Springer; 2004:64–67.
93. Henry MD, Triplett AA, Oh KB, et al. Parity-induced mammary epithelial cells facilitate tumorigenesis in MMTV-neu transgenic mice. *Oncogene*. 2004;23:6980–6985.
94. Srivastava P, Russo J, Russo IH. Chorionic gonadotropin inhibits rat mammary carcinogenesis through activation of programmed cell death. *Carcinogenesis*. 1997;18:1799–1808.
95. Medina D. Breast cancer: the protective effect of pregnancy. *Clin Cancer Res*. 2004;10:380S–384S.
96. Ginger MR, Gonzalez-Rimbau MF, Gay JP, et al. Persistent changes in gene expression induced by estrogen and progesterone in the rat mammary gland. *Mol Endocrinol*. 2001;15:1993–2009.
97. D'Cruz CM, Moody SE, Master SR, et al. Persistent parity-induced changes in growth factors, TGF-beta3, and differentiation in the rodent mammary gland. *Mol Endocrinol*. 2002;16:2034–2051.
98. Belitskaya-Levy I, Zeleniuch-Jacquotte A, Russo J, et al. Characterization of a genomic signature of pregnancy identified in the breast. *Cancer Prev Res*. 2011;4:1457–1464.
99. Peri S, De Cicco RL, Santucci-Pereira J, et al. Defining the genomic signature of the parous breast. *BMC Med Genomics*. 2012;5:46–57.
100. Russo J, Balogh GA, Russo IH. Full term-pregnancy induces a specific genomic signature in the human breast. *Cancer Epidemiol Biomrkers Prev*. 2008;17:51–68.
101. Chang PH, Hwang-Verslues WW, Chang YC, et al. Activation of Robo1 signaling of breast cancer cells by Slit2 from stromal fibroblast restrains tumorigenesis via blocking PI3K/Akt/beta-catenin pathway. *Cancer Res*. 2012;72:4652–4661.
102. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–15550.
103. de Graaf K, Hekerman P, Spelten O, et al. Characterization of cyclin L2, a novel cyclin with an arginine/serine-rich domain: phosphorylation by DYRK1A and colocalization with splicing factors. *J Biol Chem*. 2004;279:4612–4624.
104. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136:701–718.
105. Taga Y, Miyoshi M, Okajima T, et al. Identification of heterogeneous nuclear ribonucleoprotein A/B as a cytoplasmic mRNA-binding protein in early involution of the mouse mammary gland. *Cell Biochem Funct*. 2010;28:321–328.
106. Huang PR, Hung SC, Wang TC. Telomeric DNA-binding activities of heterogeneous nuclear ribonucleoprotein A3 in vitro and in vivo. *Biochim Biophys Acta*. 2010;1803:1164–1174.
107. Han SP, Friend LR, Carson JH, et al. Differential subcellular distributions and trafficking functions of hnRNP A2/B1 spliceoforms. *Traffic*. 2010;11:886–898.
108. Loyer P, Trembley JH, Grenet JA, et al. Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors: influence of cyclin L isoforms on splice site selection. *J Biol Chem*. 2008;283:7721–7732.
109. Li HL, Wang TS, Li XY, et al. Overexpression of cyclin L2 induces apoptosis and cell-cycle arrest in human lung cancer cells. *Chin Med J (Engl)*. 2007;120:905–909.
110. Zhuo L, Gong J, Yang R, et al. Inhibition of proliferation and differentiation and promotion of apoptosis by cyclin L2 in mouse embryonic carcinoma P19 cells. *Biochem Biophys Res Commun*. 2009;390:451–457.
111. Russo J, Moral R, Balogh GA, et al. The protective role of pregnancy in breast cancer. *Breast Cancer Res*. 2005;7:131–142.

112. Erwin JA, Lee JT. Characterization of X-chromosome inactivation status in human pluripotent stem cells. *Curr Protoc Stem Cell Biol*. 2010 (Chapter 1):Unit 1B 6.
113. Do JT, Han DW, Gentile L, et al. Reprogramming of Xist against the pluripotent state in fusion hybrids. *J Cell Sci*. 2009;122:4122–4129.
114. Vincent-Salomon A, Ganem-Elbaz C, Manie E, et al. X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res*. 2007;67:5134–5140.
115. Xiao C, Sharp JA, Kawahara M, et al. The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell*. 2007;128:977–989.
116. Silver DP, Dimitrov SD, Feunteun J, et al. Further evidence for BRCA1 communication with the inactive X chromosome. *Cell*. 2007;128:991–1002.
117. Breton C, Di Scala-Guenot D, Zingg HH. Oxytocin receptor gene expression in rat mammary gland: structural characterization and regulation. *J Mol Endocrinol*. 2001;27:175–189.
118. Koshimizu TA, Fujiwara Y, Sakai N, et al. Oxytocin stimulates expression of a noncoding RNA tumor marker in a human neuroblastoma cell line. *Life Sci*. 2010;86:455–460.
119. Wilson BJ, Giguere V. Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer*. 2008;7:49–57.
120. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: cells GATA have it!. *J Cell Physiol*. 2010;222:42–49.
121. Pei XH, Bai F, Smith MD, et al. CDK inhibitor p18(INK4c) is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell*. 2009;15:389–401.
122. Key TJ, Appleby PN, Reeves GK, et al. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *Lancet Oncol*. 2010;11:530–542.
123. Kouros-Mehr H, Bechis SK, Slorach EM, et al. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell*. 2008;13:141–152.
124. Fischer J, Klein PJ, Farrar GH, et al. Isolation and chemical and immunochemical characterization of the peanut-lectin-binding glycoprotein from human milk-fat-globule membranes. *Biochem J*. 1984;224:581–589.
125. Chen L, O'Bryan JP, Smith HS, et al. Overexpression of matrix Gla protein mRNA in malignant human breast cells: isolation by differential cDNA hybridization. *Oncogene*. 1990;5:1391–1395.
126. Holmes MD, Pollak MN, Hankinson SE. Lifestyle correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. *Cancer Epidemiol Biomarkers Prev*. 2002;11:862–867.
127. de Souza Rocha Simonini P, Breiling A, Gupta N, et al. Epigenetically deregulated microRNA-375 is involved in a positive feedback loop with estrogen receptor alpha in breast cancer cells. *Cancer Res*. 2010;70:9175–9184.
128. Russo IH, Russo J. Pregnancy-induced changes in breast cancer risk. *J Mammary Gland Biol Neoplasia*. 2011;16:221–233.
129. Russo J, Santucci-Pereira J, De Cicco Lopez R, et al. Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int J Cancer*. 2012;131:1059–1070.
130. Golob JL, Paige SL, Muskheli V, et al. Chromatin remodeling during mouse and human embryonic stem cell differentiation. *Dev Dyn*. 2008;237:1389–1398.
131. Asztalos S, Gann PH, Hayes MK, et al. Gene expression patterns in the human breast after pregnancy. *Cancer Prev Res*. 2010;3:301–311.
132. Russo IH, Koszalka M, Russo J. Comparative study of the influence of pregnancy and hormonal treatment on mammary carcinogenesis. *Br J Cancer*. 1991;64:481–484.
133. Russo J, Russo IH, eds. *Molecular Basis of Breast Cancer: Prevention and Treatment*. Berlin: Springer-Verlag; 2004:447.
134. Medina D. Breast cancer: the protective effect of pregnancy. *Clin Cancer Res*. 2004;10:380S–384S.
135. D'Cruz CM, Moody SE, Master SR, et al. Persistent parity-induced changes in growth factors, TGF-beta 3, and differentiation in the rodent mammary gland. *Mol Endocrinol*. 2002;16:2034–2051.
136. Russo IH, Russo J. Mammary gland neoplasia in long-term rodent studies. *Environ Health Perspect*. 1996;104:938–967.
137. Russo J, Tait L, Russo IH. Susceptibility of the mammary gland to carcinogenesis. III. The cell of origin of rat mammary carcinoma. *Am J Pathol*. 1983;113:50–66.
138. Bussolati G, Marchio C, Gaetano L, et al. Pleomorphism of the nuclear envelope in breast cancer: a new approach to an old problem. *J Cell Mol Med*. 2008;12:209–218.
139. Tan PH, Goh BB, Chiang G, et al. Correlation of nuclear morphometry with pathologic parameters in ductal carcinoma in situ of the breast. *Mod Pathol*. 2001;14:937–941.
140. Palmer JE, Sant Cassia LJ, Irwin CJ, et al. The prognostic value of nuclear morphometric analysis in serous ovarian carcinoma. *Intl J Gynecol Cancer*. 2008;18:692–701.
141. Cao R, Wang L, Wang H, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*. 2002;298:1039–1043.
142. Kubicek S, Schotta G, Lachner M, et al. The role of histone modifications in epigenetic transitions during normal and perturbed development. *Ernst Schering Res Found Workshop*. 2006:1–27.
143. Lin W, Dent SY. Functions of histone-modifying enzymes in development. *Curr Opin Genet Dev*. 2006;16:137–142.
144. Guenther MG, Young RA. Transcription. Repressive transcription. *Science*. 2010;329:150–151.
145. Navarro P, Chambers I, Karwacki-Neisius V, et al. Molecular coupling of Xist regulation and pluripotency. *Science*. 2008;321:1693–1695.
146. Surani MA, Hayashi K, Hajkova P. Genetic and epigenetic regulators of pluripotency. *Cell*. 2007;128:747–762

128

Sexual Dysfunction in the Female

Tami S. Rowen and James A. Simon

OUTLINE

Normal Female Sexual Response, 2124
  Neurotransmitters of Desire and Arousal, 2124
  Endocrinology of Desire and Subjective Arousal, 2124
Endocrinology of Objective Arousal, 2125
  Normal Vulvar and Vaginal Anatomy and Role of Hormones, 2125
Definitions and Diagnosis of Sexual Dysfunction, 2126
Pathophysiology of Female Sexual Dysfunction, 2127
  Hypoactive Sexual Desire Disorder, 2127
  Sexual Arousal Disorders, 2128
  Orgasmic Disorder, 2128
  Sexual Pain Disorders, 2128
Treatments For Sexual Dysfunction, 2129
  Hypoactive Sexual Desire Disorder, 2129
  Nonhormonal Treatments, 2129
  Hormonal Management of Hypoactive Sexual Desire Disorder, 2130
  Genital Arousal and Sexual Pain, 2131
Persistent Genital Arousal Disorder, 2132
Conclusion, 2132

Female sexual health is complex and involves numerous physiologic pathways that are relevant to the practicing endocrinologist. Female sexual problems associated with distress include problems with desire, arousal, and orgasm. They affect up to 12% of adult women, while sexual pain prevalence has been reported to be as high as 33%.[1,2] While the seat of reproduction is in the reproductive tract, external genitalia, and gonads, the locus of nearly all sexual response is in the brain, and, as such, psychosocial, developmental, behavioral, and cultural factors all play important roles. The pathways that mediate these brain responses involve neurotransmitters and hormones and can be impacted by nearly every endocrinologic system. Furthermore, female sexuality, while rooted in biology, cannot exist in the absence of the psychological and social context. In this chapter, however, we will focus primarily on the biology of female sexual health and its relevance to endocrinology.

NORMAL FEMALE SEXUAL RESPONSE

Neurotransmitters of Desire and Arousal

Normal female sexual response is a complex process that starts in the brain and is mediated by many well-known neurotransmitters. What is key about these neurotransmitters is that they may play an excitatory or inhibitory role in the promotion of desire and arousal and, as will be discussed later, are strongly influenced by hormones.

Evidence supports that the main sexual excitatory neurotransmitter is dopamine. Exposure to dopamine has been shown in animal models to enhance desire, excitement, and continuation of sexual activity.[3] Neuroimaging and animal modeling have shown multiple brain regions responsible for desire and arousal, including the ventral tegmental area, which is the primary source of dopamine to the mesolimbic and mesocortical pathways. The nigrostriatal system promotes somatomotor activity; the mesolimbic system subserves numerous types of motivation; and the medial preoptic area (MPOA) focuses the motivation onto specifically sexual targets, increases copulatory rate and efficiency, and coordinates genital reflexes. These actions are universally mediated by dopamine (Fig. 128.1).[4] Clinically, this is evidenced by data showing that excess dopamine can lead to increased and hypersexual behaviors.[5]

Another excitatory neurotransmitter involved in female sexual response is norepinephrine. Norepinephrine has been shown to play a key role in both central and peripheral regulation of sexual response. Centrally, it has been shown to promote sexual arousal, while peripherally it controls the autonomic components of sexual response (Fig. 128.1).[3]

The main inhibitory neurotransmitter in sexual desire and arousal is serotonin.[6] Pharmacologic data have demonstrated repeatedly that excess serotonin can lead to diminished sexual response, and lab data suggest this is likely mediated by inhibition of dopaminergic pathways.[4,7]

Other neurotransmitters purported to affect desire and arousal include endogenous opioids, melanocortins, and oxytocin, which more likely modulate the intensity of a perceived sexual experience and play a large role in the rewards centers that are central in mediating sexual response, in addition to pair bonding (oxytocin) (Figs. 128.1 and 128.2).[3]

Endocrinology of Desire and Subjective Arousal

The two best-known hormones related to the sexual response are estrogen (specifically, estradiol) and testosterone. Whether these hormones act individually or synergistically, or actually mediate the same actions in both the female and male brain, has been the subject of debate for many years.[8]

Among the four forms of estrogen (estrone, estradiol, estriol ,and estetrol),[9] the most abundant and relevant to the female sexual response is estradiol. Estradiol has been shown to have a direct role in the female sexual response, mainly for its effects on genital vasodilation and lubrication. Centrally, however, estrogen's main role is to stimulate the key neurotransmitters involved in sexual response at the level of the hypothalamus and limbic system, as noted earlier (Fig. 128.2).[8] The role of testosterone and other androgens has long been known to be important

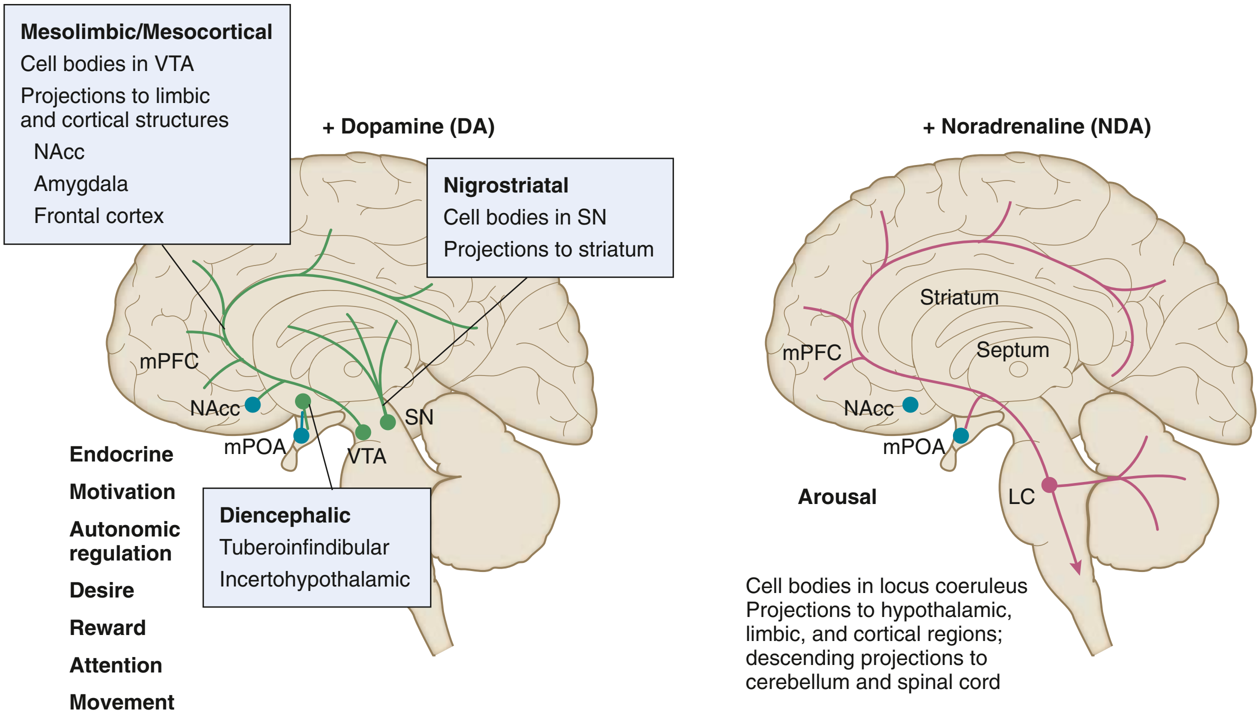

Mesolimbic/Mesocortical
Cell bodies in VTA
Projections to limbic and cortical structures
NAcc
Amygdala
Frontal cortex
+ Dopamine (DA)
Nigrostriatal
Cell bodies in SN
Projections to striatum
mPFC
NAcc
SN
mPOA
VTA
Endocrine
Motivation
Autonomic regulation
Desire
Reward
Attention
Movement
Diencephalic
Tuberoinfindibular
Incertohypothalamic
+ Noradrenaline (NDA)
Striatum
mPFC
Septum
NAcc
mPOA
Arousal
LC
Cell bodies in locus coeruleus
Projections to hypothalamic, limbic, and cortical regions; descending projections to cerebellum and spinal cord

Figure 128.1 Excitatory systems of the brain.

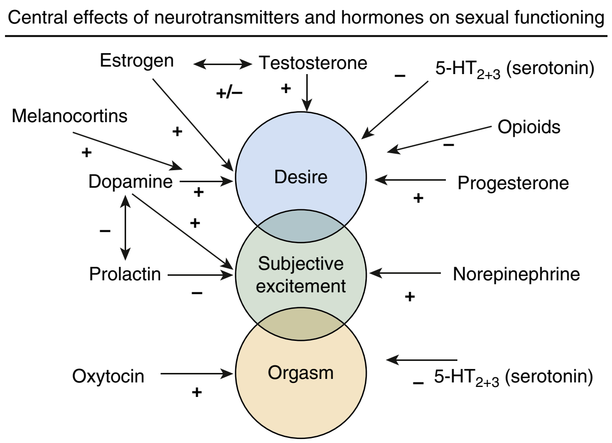

Central effects of neurotransmitters and hormones on sexual functioning
Estrogen
Testosterone
+/–
+
–
5-HT$_{2+3}$ (serotonin)
Melanocortins
+
+
Opioids
–
Dopamine
+
Desire
Progesterone
+
–
+
Prolactin
Subjective excitement
Norepinephrine
–
+
Oxytocin
+
Orgasm
– 5-HT$_{2+3}$ (serotonin)

Figure 128.2 Central effects of neurotransmitters and hormones on sexual functioning.

in the male sexual response. This is also true for women. Testosterone likely influences the expression of nitric oxide synthase, which increases the production of nitric oxide, which then leads to the release of the neurotransmitters central to the female sexual response.[6] Animal studies demonstrate that androgens rather than estrogens are linked to targeting of dopaminergic pathways in the prefrontal cortex.[10] Human studies demonstrating the sexual benefits of testosterone are limited to clinical supplementation with androgens in women with sexual dysfunction.[11] These studies suggest that it is androgens alone, rather than estrogens, that mediate sexual response in women (Fig. 128.2).[6]

It is well established that androgen levels decrease across the lifetime of women.[12] While there are weak associations between lower androgen levels and lower sexual desire, there is no defined cutoff concentration associated with low sexual desire.[13] There has been significant research into the effects of supplemental testosterone on female sexual dysfunction, the details of which will be discussed later in this chapter.

Interestingly, animal research (primarily in rodents) has indicated that, in females only, circulating estradiol is predictive of sexual response.[3] In contrast in humans, there is considerable debate regarding the specific role(s) of the different sex steroids and their effects on sexual desire and arousal, namely whether androgens or estrogens are responsible for these responses. In a well-designed study examining daily sexual desire and hormone levels, only estradiol reliably predicted desire, while progesterone had a negative effect, and testosterone concentrations were not as predictive as estradiol.[14] Additionally, it has been demonstrated that testosterone levels decrease more rapidly in the fourth decade of life in women, rather than in menopause, and circulating estradiol levels in this age group have been shown to be more relevant to lower sexual desire than circulating androgens.[15] However, clinical data have not reliably demonstrated that supplementing women with estradiol leads to increased desire.[8] This is not the case for testosterone, as will be addressed in later sections.

KEY POINTS

- Dopamine is the key neurotransmitter regulating sexual desire.
- Both estradiol and testosterone modulate dopamine and improve all aspects of the female sexual response.
- Both local and systemic estrogens and testosterone can maintain vulvar, vestibular and vaginal health.

ENDOCRINOLOGY OF OBJECTIVE AROUSAL

Normal Vulvar and Vaginal Anatomy and Role of Hormones

Historically, the hormone focus regarding vulvar and vaginal health has been on the degree of estrogenic stimulation. It is well established that there are receptors for estradiol throughout the vagina, vulva, and urethra, and even in the trigone of the bladder.[15,16] The roles of estrogen are numerous, including maintaining blood flow to the vulvovaginal tissue, the collagen within the epithelium, and the hyaluronic acid

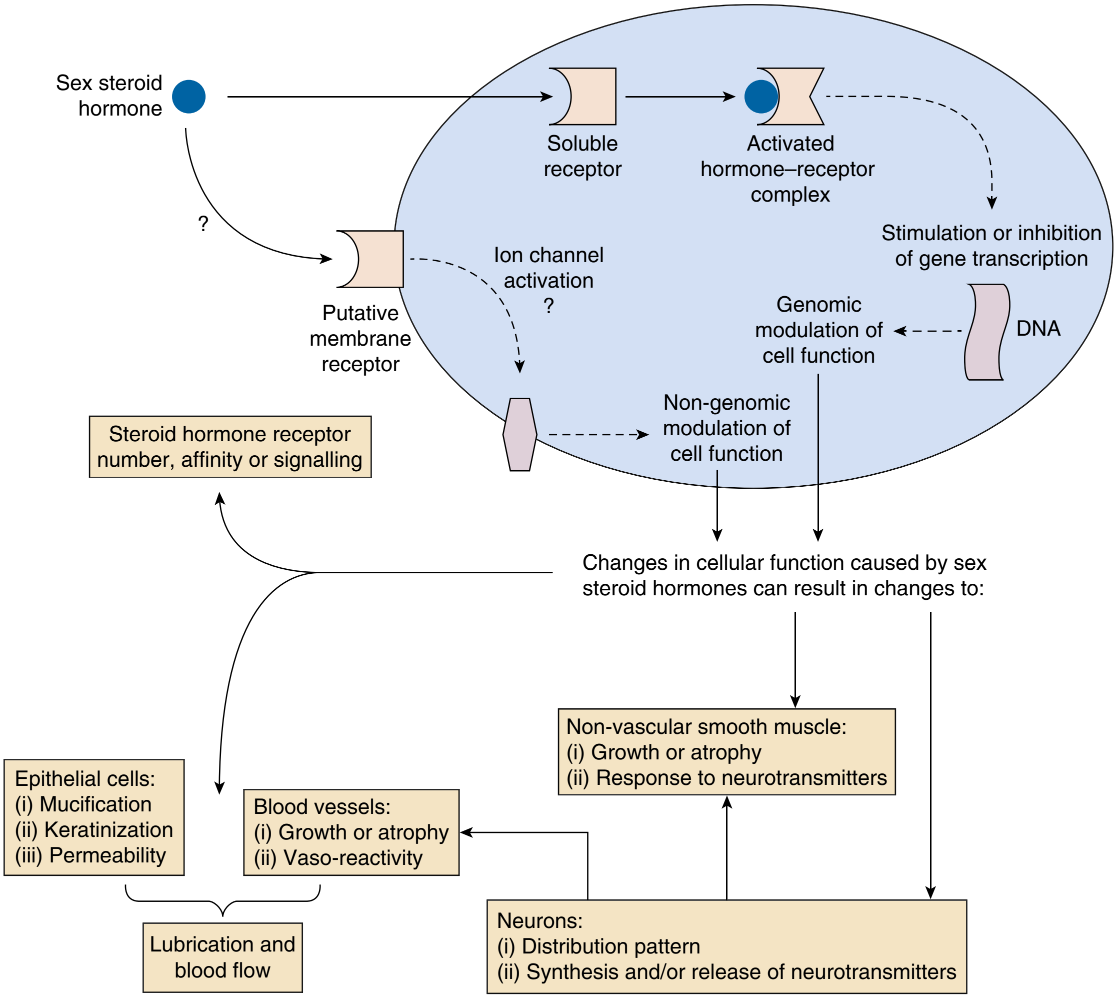

Sex steroid hormone
Soluble receptor
Activated hormone–receptor complex
?
Putative membrane receptor
Ion channel activation
?
Stimulation or inhibition of gene transcription
Genomic modulation of cell function
DNA
Non-genomic modulation of cell function
Steroid hormone receptor number, affinity or signalling
Changes in cellular function caused by sex steroid hormones can result in changes to:
Non-vascular smooth muscle:
(i) Growth or atrophy
(ii) Response to neurotransmitters
Epithelial cells:
(i) Mucification
(ii) Keratinization
(iii) Permeability
Blood vessels:
(i) Growth or atrophy
(ii) Vaso-reactivity
Lubrication and blood flow
Neurons:
(i) Distribution pattern
(ii) Synthesis and/or release of neurotransmitters

Figure 128.3 Estrogen plays a role in supporting the microbiome and protecting the tissue from pathogens.

and mucopolysaccharides within the moistened epithelial surfaces. Estrogen also plays a role in supporting the microbiome and protecting the tissue from pathogens (Fig. 128.3).[17] Although estrogen is important in maintaining vulvovaginal health, there are emerging data that androgens also have an important role.[16] All three layers of the vagina have been shown to have androgen receptors; in addition, androgens have been shown in animal models to stimulate nerve fibers, as well as mucification of the vaginal epithelium.[16] Furthermore. it appears that androgens are likely derived from the conversion of circulating dehydroepiandrosterone (DHEA) entering the cells of the vulvar and vaginal tissue.[18] Androgens maintain nonvascular smooth muscle function in the vagina and vascular smooth muscle in the clitoris and also have a stimulatory effect on vaginal nerve density.[16,19,20]

This critical role of sex steroids in the vulva and vagina is important for normal female sexual response. Physiologic arousal derives from increased blood flow to the clitoris, labia, vulva, and vagina, which results in vasocongestion as well as lubrication.[21,22] As in male genitals, the basic mechanisms of these processes are mediated through adrenergic and nitric oxide signaling, although modulation by the sex steroid hormones themselves is also involved.[19,23]

Testosterone and estradiol regulate their own receptors and can also cross-regulate the other receptors.[24] Estradiol modulates genital blood flow and is related to the density of collagen fibers in the lamina propria. In addition, estradiol moderates muscular thickness in the vagina and vasorelaxant mechanisms of the clitoris involved in arousal.[15,25] Vaginal lubrication is a combination of basal mucin production and vaginal vascular transudate.[20] Sex steroid hormones regulate mucification, keratinization, and the permeability of the vaginal epithelium (Fig. 128.3).[19] Mucification of the vaginal epithelium is a result of androgens, while proliferation of the vaginal epithelium is related to the effects of estrogens.[15,20] Therefore, mucin production is regulated by androgen, whereas vaginal vascular transudation is an estrogen-dependent effect. Thus, together both androgens and estrogens play a role in objective arousal physiology.

DEFINITIONS AND DIAGNOSIS OF SEXUAL DYSFUNCTION

Definitions of female sexual dysfunctions are based primarily on clinical manifestations and patient report rather than pathophysiological etiology. Current definitions are closely tied to the evolution of the Diagnostic and Statistical Manual of Mental Disorders (DSM), the diagnostic classification system of the American Psychiatric Association.[26] The DSM-IV definition of hypoactive sexual desire disorder (HSDD) states that a woman has to have persistently or recurrently deficient (or absent) sexual fantasies and desire for sexual activity. The judgment of deficiency or absence is made by the clinician, taking into account factors that affect sexual functioning, such as age and, importantly, the context of an individual's life. The disturbance must cause marked distress or interpersonal difficulty in order to be considered a disorder, and the sexual dysfunction *cannot* be better accounted for by another axis 1 disorder (except another sexual dysfunction). Most critical, especially for the non–sexual medicine expert, is that the disorder is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication such as an selective serotonin reuptake inhibitors [SSRI], an antihypertensive, etc.) or a general medical condition. The diagnosis additionally segregates the disorder into the categories of acquired versus lifelong, and situational versus generalized.[27] The other categories of female sexual dysfunction in the DSM-IV included arousal disorder, aversion disorder, orgasmic disorder, and two different sexual pain disorders: vaginismus and dyspareunia.

The more recent DSM-V, published in 2013, eliminated many of these definitions.[28] It merged "desire" and "arousal" into one entity called "sexual interest/arousal disorder." The diagnosis depends on a patient having three of the following symptoms:

a. Absent or decreased interest in sexual activity
b. Absent or decreased sexual or erotic thoughts and fantasies
c. No or decreased initiation of sexual activity and lack of receptivity to a partner's attempts to initiate
d. Absent or decreased sexual excitement or pleasure during sexual activation in almost all or all (75%–100%) sexual encounters
e. Absent or decreased sexual interest or arousal in response to any internal or external sexual or erotic cues
f. Absent or decreased genital or nongenital sensations during sexual activity in almost all or all (75%–100%) sexual encounters

In DSM-V, the female orgasmic disorder was defined as "marked delay in, marked inadequacy of, or absence of orgasm" and eliminated the prior requirement of "following a normal sexual excitement phase."

Furthermore, the sexual pain disorders were combined into "genito-pelvic pain/penetration disorder," defined as persistent or recurrent difficulties with at least one of the following characteristics:

a. Vaginal penetration during intercourse
b. Marked vulvovaginal or pelvic pain during vaginal intercourse or penetration attempts
c. Marked fear or anxiety about vulvovaginal or pelvic pain during or in anticipation of vaginal penetration
d. Marked tensing or tightening of the pelvic floor muscles during attempted vaginal penetration

The DSM definitions are a mainstay of mental health practice in the United States and around the world, and thus play a large role in research, evaluation, and clinical management of women with female sexual dysfunction. However, there is a great deal of controversy regarding these definitions, especially regarding combining desire and arousal disorders, and genital pain and penetration disorders. Additionally, no compelling evidence exists that suggests that a duration of 6 months has better diagnostic validity than a duration of 3 months, but the former would certainly delay the establishment of a diagnosis and the availability of reimbursement for therapy. Similarly, the DSM-V rules also require a rubric that the symptoms must be present "in almost all or all (75%–100%) occasions of sexual activity." These definitions and percentages are not evidence-based and may not correlate in any way to a patient's distress level.[29]

International Classification of Diseases. The International Classification of Diseases (ICD) codes comprise a worldwide health information standard developed and published by the World Health Organization for coding mortality and morbidity, signs, symptoms, reasons for encounter, and external causes of both injury and disease. This classification system has been used throughout the world as a standard diagnostic and coding tool and incorporates many of the diagnoses seen in the DSM.[30] The most current version is the ICD-11, published in 2018; however, the ICD-10 is still in worldwide use and includes disorders of desire, aversion, genital response, dyspareunia, and vaginismus. The new ICD-11 has expanded the definitions greatly and addressed some of the concerns with the DSM-V definitions noted previously. For example, ICD-11 takes into account associated conditions and nonnormative sexual practices, reflecting a broader understanding of sexual dysfunctions, and continues to separate desire, arousal, and genital pain disorders as the DSMIV-TR did.[31]

Other Schemata. Given ongoing developments in sexual health research and knowledge, there have been additional approaches to nomenclature and diagnosis proposed by leading researchers and academic groups to expand the nomenclature and definitions of female sexual dysfunction, including desire, arousal, and sexual pain.[29,32-34] These definitions take into account the many associated factors, associated pathophysiology, and nuances in the diagnoses of sexual disorders.[35-37]

Screening Tools. Most screening tools focus on HSDD. Provider interviews are necessary to elucidate many of the nuances as well as the psychosocial factors that are critically involved in all sexual dysfunctions. However, there can be much benefit from selected screening tools and surveys focusing on patient-reported outcomes (PROs). Most clinical studies on treatment for HSDD require specific endpoints related to these PROs, but they can also be used in the office by clinicians and include the Elements of Desire Questionnaire and Decreased Sexual Desire Screener.[33] The Female Sexual Function Index (FSFI)[38] and Brief Index of Sexual Functioning for Women (BISF-W)[39] have domains related to pain, orgasm, and arousal and can thus help incorporate all aspects of female sexual dysfunction into one screening tool, although these are most commonly used in research studies.[38,39]

Similarly, for pain disorders, there are numerous tools to assess the degree of pain and its clinical manifestations. These also are different tools focused for the research versus clinical environments.[40] For many clinicians, using a visual analog scale can be most useful in a busy clinical setting, and testing external and internal genitalia is necessary to elucidate the source and severity of the pain.

Distress. As noted previously, reduced scores on tests of desire, arousal, pain, and so on do not rise to the diagnosis of a disorder unless accompanied by clinically meaningful and measurable distress. Most patients do not use the word "distress" in their vocabulary defining their sexual dysfunction. One of the most used terms closest in sentiment to distress is "bother." Distress is commonly measured by a short questionnaire, the Female Sexual Distress Scale–Revised.[33] This validated tool has been paired with the FSFI and the BISF-W in multiple investigations and is commonly used by regulators in their assessment of new therapies.

PATHOPHYSIOLOGY OF FEMALE SEXUAL DYSFUNCTION

Hypoactive Sexual Desire Disorder

Sexual desire is regulated by key regions in the brain through the action of various effects on resident neurotransmitters. Dopamine, norepinephrine, melanocortin, and oxytocin mediate sexual excitation, whereas opioid, serotonin, endocannabinoid, and prolactin mediate sexual inhibition. HSDD occurs in premenopausal women with normal menstrual function, despite the hormonal fluctuations in the cycle. It happens in women who are menopausal, with or without hormone "replacement" therapy, including with or without the addition of testosterone. Although the underlying biological causes of HSDD remain unknown, HSDD likely involves the effects of inhibitory processes or neuroadaptations that result in decreased excitation, increased inhibition, or a mixture of the two.[32] Based on this understanding, there are numerous medical conditions that would thus affect and be associated with HSDD. The most well-known and understood comorbidity in HSDD is depression, but other medical conditions like diabetes, obesity, incontinence, and especially surgical menopause play a role.[32] Furthermore, there are numerous medications that can also affect sexual desire, based on known alterations in excitatory and inhibitory neurotransmitters and hormones. SSRIs are the most commonly recognized medications, but other psychiatric medication, blood pressure medication, and especially hormonal contraception can contribute as well.[32] A thorough psychosexual history is also necessary, given that the psychosocial component of HSDD cannot be ignored. It should be noted that there are neuroimaging studies that demonstrate differential brain activity patterns in women with and without HSDD,[41] suggesting that alterations in brain function and structure may be additionally modulated or reinforced by experience and behavior, which has been characterized as experience-based neuroplasticity.[42] This can then further propagate the HSDD condition.

Sexual Arousal Disorders

We have already established that sexual arousal is distinct from desire, and there are also differences between genital arousal and subjective, or cognitive, arousal.

In terms of genital arousal disorder, the pathophysiology is often related to changes in the local hormonal and/or vascular environment, as seen in the menopause, pregnancy, and postpartum. It should be noted that there are medications that can alter the hormonal environment of the vagina and vulva.[43] Any assessment of a woman with genital complaints requires a thorough history and medication history to assess for risk factors for hormonal or vascular changes. The most common are aromatase inhibitors for the treatment and prevention of hormone-sensitive breast cancer, but an often-overlooked group of medications are combined hormonal contraceptives or any contraception that decreases follicle-stimulating hormone and luteinizing hormone, and thus estradiol and testosterone.[44,45] There are other etiologies that can affect local genital response as well. These include nerve injuries, as well as medications that can blunt or overstimulate the local genital response.[46,47] Vascular disorders can also blunt the female sexual response due to changes in blood flow and sensation.[48] There are data that suggest sustained exercise, presumably leading to local nerve injury, can also blunt the genital response, as bicycle riders report significantly higher rates of genital numbness.[49]

Female cognitive arousal disorder is a distinct entity characterized by a lack of mental engagement and focus in response to sexual stimuli.[37] There is much research demonstrating that subjective and genital arousal may not happen simultaneously and may in fact be dyssynchronous.[50] Most research on subjective arousal is focused on use of PROs and is often combined with HSDD research. Evidence distinguishing the pathophysiology of subjective arousal from that of HSDD is limited. Many of the same neurotransmitters, hormones, medical conditions, and medications involved in the pathophysiology of desire play a central role in arousal, but their effect on a woman's subjective arousal is often distinct from her subjective desire. However, the ultimate effect is an imbalance of inhibition and excitation that leads directly to alterations in subjective arousal.

Orgasmic Disorder

Among all established sexual health conditions, orgasmic disorder is likely the least well understood, likely because of the lack of knowledge of the fundamental physiology, purpose, and control of female orgasmic function. Women who have difficulty reaching orgasm sometimes show decreased sexual desire and/or inadequate arousal, suggesting a problem with the desire and arousal phases rather than with the orgasm phase *per se*.[51,52] Numerous biological factors can account for disorders of orgasm in women. General medical conditions, such as heart or kidney disease, multiple sclerosis, fibromyalgia, thyroid disorders, weakened pelvic floor, and menopause have been associated with all female sexual dysfunctions, including anorgasmia. The literature on the orgasm-inhibiting effects of SSRIs is particularly abundant and dates back almost two decades. Anorgasmia is reported in at least one third of patients who receive SSRIs, in a dose-dependent manner.[53] Alcohol and other drugs also have been shown to interfere with orgasm. Most importantly, psychosocial factors cannot be overlooked and likely play a significant role; these factors include, but are not limited to, age, education, social class, cultural background, relationship conflicts, and shame or embarrassment about sexuality due to religious beliefs or familial inhibitions.[54] A distinction between primary/lifelong orgasmic dysfunction and secondary/acquired dysfunction may provide the practitioner with both a clue to the cause and the pathway toward treatment. Generally, primary/lifelong orgasmic dysfunction has been the purview of psychosocial practitioners, while secondary/acquired dysfunction has been attributed to a biological source.

Sexual Pain Disorders

Sexual pain disorders encompass a wide variety of conditions. As with all sexual disorders, the importance of psychological and social assessments/treatments must be taken into account, but it is important to remember that women who are experiencing pain will likely have a physical/biological cause underpinning their pain. Often, it is important to assess the psychological and social effects of that biological condition. For example, the patient with deep endometriosis may have severe sexual pain, but the barriers caused by the endometriosis pain's impact on intimacy may create psychological and social effects for the patient, the partner, or the couple taken together.

Common gynecological causes of sexual pain include endometriosis and pelvic floor muscle dysfunction.[55] Pelvic floor muscle overactivity or hypertonicity are commonly overlooked causes of sexual pain. Prolonged use of these muscles, "clenching against the pain," can result in decreased tissue oxygenation, muscle overactivity, shortening of sarcomeres, and the development of myofascial trigger points. Hypertonicity of the muscles that insert at the posterior vestibule—the pubococcygeus, puborectalis, and superficial transverse perinei—can lead to allodynia (as seen in vestibulodynia) in the posterior vestibule. Hypertonicity of deeper muscles (e.g., ileococcygeus, coccygeus, and obturator internus) can lead to vaginal or deep thrusting dyspareunia. In addition, overactivity of the bulbocavernosus and ischiocavernosus is associated with clitorodynia. Many women with pelvic floor hypertonus may also have chronic constipation and elements of urinary retention.[55]

Many women with genital pain have a local component confined to the vulvar vestibule, and this is mediated by hormonal, neurological, and inflammatory conditions,[55] which we will focus on here. As noted earlier, the tissues of the vulva and vagina are both responsive and dependent on sex steroid hormones for proper health and function (Fig. 128.3). There are many causes of decreased sex steroids, both natural and iatrogenic, that can lead to dyspareunia. The most common cause of decreased sex steroids in women is menopause. Other physiologic causes include anovulation secondary to lactation, anorexia, hypothalamic amenorrhea, excessive physical activity or physiological stress, and hyperprolactinemia. Iatrogenic causes of decreased circulating sex steroids include oophorectomy, which leads to immediate menopause, hysterectomy (without oophorectomy), and commonly prescribed medications, such as combined hormonal contraception (CHC).[55] CHC use leads to a reduction in serum estradiol and free testosterone by decreasing ovarian production of estrogen and total testosterone based on negative feedback and by inducing the liver to produce increased levels of sex hormone–binding globulin (SHBG). SHBG binds both free circulating testosterone and estradiol, with higher affinity for testosterone. In addition, some CHCs contain synthetic progestogens that act as testosterone antagonists at the androgen receptor, most notably drospirenone, which is an analog of spironolactone. It has been shown that CHCs cause histopathologic changes in the vestibular epithelium (the embryologic equivalent of the male penile urethra), thereby increasing vulnerability to mechanical strain and decreasing mechanical pain thresholds. Preliminary evidence suggests that women with provoked vestibulodynia may have alterations in their androgen receptor leading to reduced sensitivity to lowered androgen environments.[56] Given their effects on sex steroids, it is perhaps not surprising that studies have shown that CHCs are associated with an increased risk of vestibulodynia.[57]

Inflammation plays a central role in vestibulodynia as well. Although both women with vestibulodynia and healthy women have inflammatory cells in the vestibular epithelium, the relative abundance and organization of these cells may differ between women with and without vestibulodynia. Several controlled studies have demonstrated that women with vulvodynia have evidence of central sensitization.[58] These results suggest that women with vulvodynia have an increased perception of nonpainful and painful stimulation to the vestibule.

KEY POINTS

- Sexual desire disorder is distinct from sexual arousal disorder.
- Sexual pain can be related to conditions that affect the vulva, vulvar vestibule, vagina and internal pelvic organs.
- Most female sexual dysfunctions are associated with other comorbidities and can be caused by both endocrinopathies and their treatments.

TREATMENTS FOR SEXUAL DYSFUNCTION

Hypoactive Sexual Desire Disorder

HSDD is the most common sexual dysfunction in women, and, as outlined earlier, is often associated with poor medical and psychological states. All treatments for HSDD should begin with education, adjustment of potentially modifiable factors, and, if needed, provision of additional therapeutic interventions. The choice of therapeutic intervention should be made in conjunction with the patient (and potentially her partner), based on her/their preferences and goals. The International Society for the Study of Women's Sexual Health has developed a comprehensive process of care for management for women with HSDD. Many of their suggestions are incorporated in this chapter (Fig. 128.4).[35]

Nonhormonal Treatments

The first US Food and Drug Administration–approved treatment for generalized, acquired HSDD in women was flibanserin (August 18, 2015). Flibanserin (100 mg nightly) is a nonhormonal, centrally acting, daily, oral, serotonin agonist and antagonist. Flibanserin works by enhancing 5-HT1A activity and inhibiting 5-HT2A activity. Between the two actions, serotonin production is heavily downregulated. This, in turn, leads to changes in dopamine and norepinephrine in the brain.[59] Over 3000 women were initially studied in the three original, pivotal trials, which revealed changes in sexual desire (increase), sexual distress (decrease), and satisfying sexual events (increase) versus placebo as endpoints.[60] There is evidence that demonstrates the efficacy and safety of flibanserin in postmenopausal women as well, but flibanserin is not approved for postmenopausal women in the United States.[60] Approximately 50% of women with HSDD respond to flibanserin, and it may take up to 8 weeks for efficacy to emerge. The most common adverse events (AEs) in premenopausal women are dizziness (9.2%), somnolence (8.3%), nausea (6.5%), and fatigue (3.7%)· It should be noted that placebo-corrected efficacy rates are similar to those seen with other central nervous system (CNS)-active agents (i.e., for depression). In the pivotal clinical trials, the rate of discontinuation due to AEs was 13% in premenopausal women treated with flibanserin compared with 6% treated with placebo.[62]

Flibanserin's original labeling carried a boxed warning that concomitant alcohol use is contraindicated, and providers had to become certified to prescribe the medication as part of a risk evaluation and mitigation strategy (REMS) with elements to assure safe use program. This was based on the results of an alcohol challenge study that found an increase in sedation, syncope, and hypotension in the treatment group (23 men and two women) when flibanserin was coadministered with alcohol. Of note, alcohol use was not restricted and did not increase such AEs over placebo in the three major pivotal trials, which were limited to premenopausal women. A series of postapproval alcohol interaction studies, for example one performed in 96 women (≤45 years old), revealed no effect of a combination of ethanol for blood pressure, vertigo, or syncope.[63] The boxed warning has since been removed, and there is no longer a REMS certification requirement for prescribers and pharmacies to educate patients to avoid alcohol. The remaining warning is to take flibanserin at bedtime and 2 or more hours following the last consumption of alcohol. If that timing is not possible, then patients are advised to skip that night's dose of flibanserin.

The second nonhormonal medication to be FDA-approved for premenopausal women is bremelanotide (1.75 mg subcutaneous as needed), a melanocortin-4 receptor agonist. Its mechanism of action is also believed to be downstream stimulation of dopamine receptors in the MPOA following melanocortin-4 receptor binding (Figs. 128.1 and 128.2).[64] The efficacy of bremelanotide was shown in two phase III randomized placebo-controlled trials showing improvement in sexual desire and decrease in sexual distress, but notably not affecting the number of satisfying sexual events.[65] Its use "as desired" is seen as a possible mechanism for the absence of this agent's impact on satisfying sexual events. Bremelanotide comes in a single-use insulin-like pen autoinjector and is approved for "as needed" use only. Bremelanotide's main AEs (vs. placebo) are nausea (40.0% vs. 1.3%), flushing (20.3% vs. 0.3%), injection site reactions (13.2% vs. 8.4%), headache (11.3% vs. 1.9%), and vomiting (4.8% vs. 0.2%). All side effects (except for injection site reactions) tended to decrease with additional use beyond the first dose. Whether this improvement in AE experience was due to reduced associated anxiety or tachyphylaxis is unknown. There is no boxed warning or REMS program for bremelanotide. While not tested in its current formulation, a nasal formulation (not FDA-approved) demonstrated efficacy in postmenopausal women.

There are other CNS agents that are used off label for HSDD. Bupropion (300–400 mg/day) is the only one studied with sexual desire as a clinical endpoint, and did show efficacy, but the one study of this subject did not enroll enough patients to meet statistical significance.[66] The evidence for adding bupropion for people with antidepressant-induced HSDD is mixed.[67,68] Buspirone is another antidepressant and anxiolytic agent that has been used for HSDD, and did show efficacy in improving desire in clinically depressed patients, but there is no study on this drug where HSDD is the primary clinical endpoint.[69] Trazodone, a phenylpiperazine compound of the serotonin antagonist and reuptake inhibitor class, is commonly prescribed as a sedative, anxiolytic, and antidepressant. However, it also has efficacy in post-traumatic stress disorder, alcohol dependence, and so on. It is known to bind to several serotonin receptors in the 5HT1 and 5HT2 families, like flibanserin, as noted earlier. Trazodone is known to have sexual side effects at high doses that include priapism/clitorism. Recent research demonstrates efficacy for decreased sexual desire (i.e., HSDD) without priapism/clitorism when used at doses as low as 3.5 mg/d.[70]

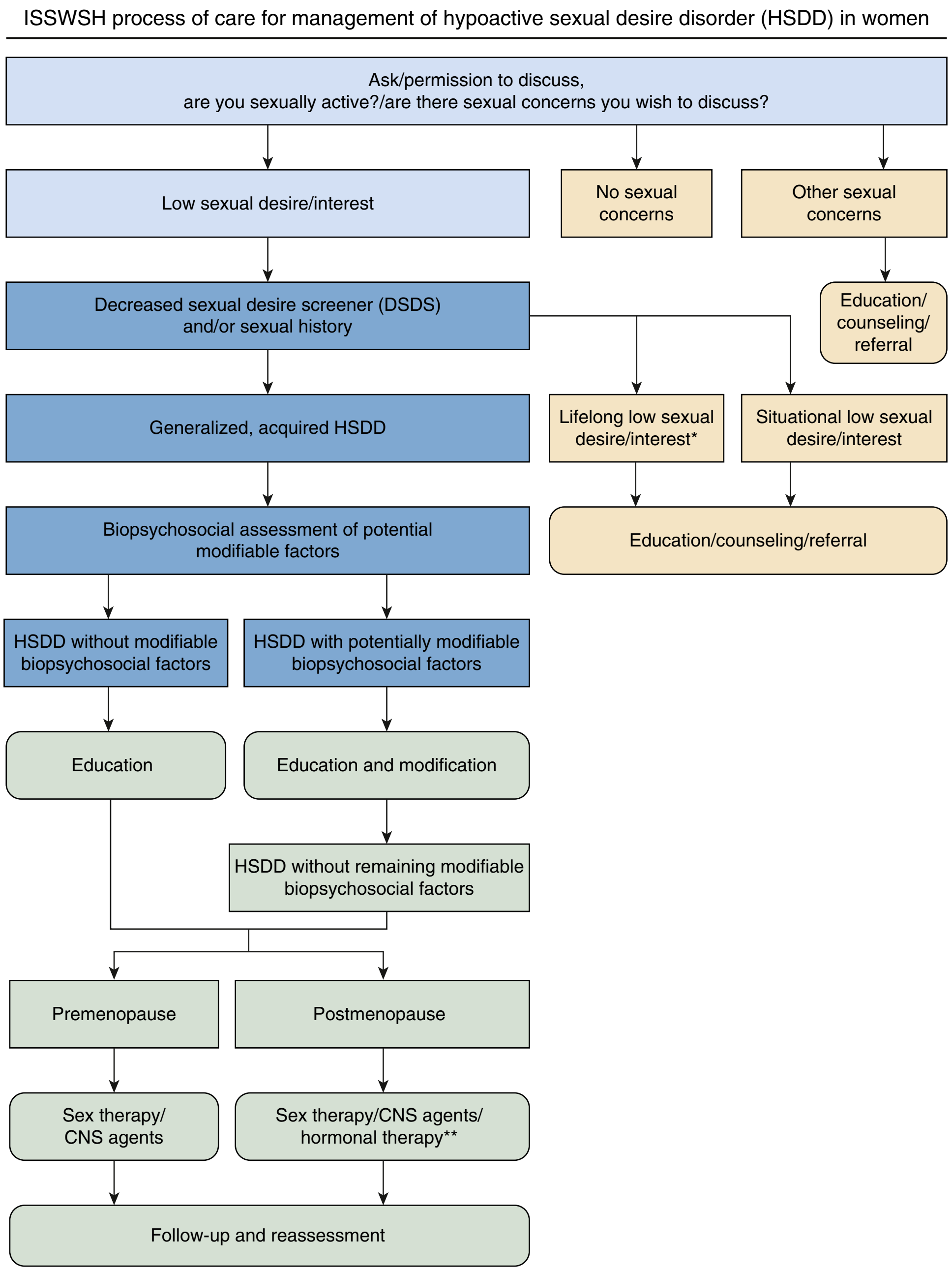

ISSWSH process of care for management of hypoactive sexual desire disorder (HSDD) in women
Ask/permission to discuss, are you sexually active?/are there sexual concerns you wish to discuss?
Low sexual desire/interest
No sexual concerns
Other sexual concerns
Education/ counseling/ referral
Decreased sexual desire screener (DSDS) and/or sexual history
Generalized, acquired HSDD
Lifelong low sexual desire/interest*
Situational low sexual desire/interest
Biopsychosocial assessment of potential modifiable factors
Education/counseling/referral
HSDD without modifiable biopsychosocial factors
HSDD with potentially modifiable biopsychosocial factors
Education
Education and modification
HSDD without remaining modifiable biopsychosocial factors
Premenopause
Postmenopause
Sex therapy/ CNS agents
Sex therapy/CNS agents/ hormonal therapy**
Follow-up and reassessment

Figure 128.4 **The International Society for the Study of Women's Sexual Health process of care for management for women with hypoactive sexual desire disorder.** (Adapted from Mayo Clinic Proceedings, Vol 93, Issue 4, Clayton AH et al. The International Society for the Study of Women's Sexual Health Process of Care for Management of Hypoactive Sexual Desire Disorder in Women. Copyright Elsevier 2018.)

Hormonal Management of Hypoactive Sexual Desire Disorder

Estrogens. While most focus on hormones and HSDD is on the effects of testosterone, we have already demonstrated that some testosterone is converted to estradiol, and there is evidence that estradiol mediates sexual desire in women. It would thus make sense that estradiol supplementation may improve HSDD in women. However, nearly all research focused on this subject has taken place in postmenopausal women who are being given estrogen supplementation for bothersome menopausal symptoms (vasomotor, sleep, and vulvovaginal complaints), including low desire. It is not recommended to treat premenopausal women with estrogen for low sexual desire.

There is research that demonstrates that women on menopausal hormone therapy with estradiol have improved sexual function compared with placebo.[71] Other data have demonstrated that ultralow dose supplementation does not improve function, along with other menopausal symptoms.[72] Further evidence shows that the route of estrogen administration is likely most critical when using postmenopausal estrogen therapy for HSDD. A well-designed randomized clinical trial (RCT) demonstrated that transdermal, but not oral, estrogen replacement was associated with improved sexual function, primarily in the pain and lubrication domains.[73] This difference related to route of estrogen administration may be mediated by the liver "first-pass" increase in SHBG that occurs with oral estrogens, but not transdermal preparations. Increased SHBG during oral estrogen administration has the potential to decrease circulating free testosterone, again suggesting the role of testosterone in sexual function. Thus, if considering systemic estrogen treatment for low sexual desire in postmenopausal women, preference should be focused on nonoral delivery (patches, lotions, gels, vaginal rings, etc.)

Testosterone and Low Desire. We have already established how testosterone mediates sexual desire through effects on the neurotransmitters responsible for excitation and inhibition (Figs. 128.1 and 128.2). As such, there has been consistent evidence from placebo-controlled RCTs that testosterone therapy is more effective than placebo for the treatment of HSDD in postmenopausal women[74,75] when delivered at physiologically normal doses. Furthermore, such doses have demonstrated safety when used short-term, typically 6 months. Initially, testosterone treatments were in the form of oral testosterone and testosterone implants.[76-79] In order to maintain a more consistent and safer level of testosterone while continuing to avoid first-pass metabolism in the liver, transdermal formulations were developed. These included a transdermal testosterone patch releasing 300 mcg of testosterone/day or a cream delivering 5 mg testosterone in 0.5 mL (10 mg/mL) daily.

There are data on the efficacy of using patch-delivered transdermal testosterone for menopausal women with or without concomitant estrogen and progesterone/progestogen for low desire. It also can be used as a single treatment, improving satisfying sexual events, arousal, orgasm frequency, pleasure, responsiveness, and self-image.[80,81] There is likewise a small dataset demonstrating efficacy in late premenopasual women, but this is mainly limited to women closer to the menopasual transition (mean age 39.7 years).[82] A recent metaanalysis demonstrated the need for further research to determine the efficacy of testosterone therapy for premenopausal women with HSDD.[81]

The main concern with testosterone is in regard to safety, both in the short and long term.

Testosterone preparations, specifically intramuscular preparations, subcutaneous implants, and oral formulations (methyltestosterone 1.25–2.5 mg or testosterone undecanoate 40 mg), have wide variation in serum concentrations and limited short- and long-term safety data in women,[11] most of whom were postmenopausal in studies that have been completed. The main side effects of testosterone supplementation (all forms) are acne, hirsutism, and balding, as well as weight gain. Lipid and lipoprotein effects included reduction in total cholesterol, high-density lipoprotein cholesterol, and triglyceride levels along with a statistically significant increase in low-density lipoprotein cholesterol with oral, but not transdermal, testosterone preparations.[81,83]

A major concern is the potential for testosterone to increase the risk of breast cancer, principally through its aromatization to estradiol. Although the data are limited, a metaanalysis of data from over 3000 women using transdermal testosterone, with or without concomitant estrogen and progestin hormone therapy, found that breast density (a surrogate marker for cancer risk) did not change with transdermal testosterone treatment in several small trials with insufficient data to assess long-term breast cancer risk.[81,83]

Long-term safety data are limited. One 4-year study of open-label extension of two, 6-month clinical trials of 300 mcg/day transdermal testosterone patch in 967 women with oophorectomy treated with estrogen therapy showed no increase in the rate of serious AEs.[84] In 2008 a testosterone gel phase II cardiovascular and breast safety RCT required by the FDA to establish safety enrolled 3656 postmenopausal participants over the age of 50 years who had at least two cardiovascular risk factors at baseline (e.g., hypertension, dyslipidemia) and a clinical diagnosis of HSDD.[85] After 4 years, with over 7300 women-years of exposure, the company (BioSante Pharmaceuticals, Inc, Lincolnshire, IL) reported 53 adjudicated cardiovascular events, a lower than anticipated rate, and a low breast cancer rate, which is reassuring. However, although the study was reviewed by an independent data safety and monitoring board that allowed the trial to continue, the data have never been peer reviewed or published.[86] It has been determined that, despite these reassuring outcomes, long-term (beyond 2 years) safety data, particularly regarding breast cancer and cardiovascular events (acute myocardial infarction, stroke, deep vein thrombosis, and death), are limited and inconclusive.[11,81,87,88] To fully satisfy the medical and regulatory agencies' concerns with these important questions would likely require a billion dollars of funding over a decade, which thus severely limits the ability of any naturally occurring product (i.e., testosterone) to be brought to market.[89]

It is recommended that, should a patient and her healthcare provider decide on a therapeutic trial of testosterone, baseline total testosterone should be evaluated.[90] Normal testosterone ranges have been reported for women of different age groups, but there is no minimal value for any androgen that can be used to identify women with HSDD.[13,35] Women using testosterone should have regular follow-up blood testosterone measurements to ensure that supraphysiologic therapy and serum levels are avoided.[11]

Testosterone formulations for women are not globally available, so clinicians are commonly limited to prescribing compounded formulations of 1%, or testosterone formulations for men modified to much lower administered doses (usually one tenth of the male dose). Current guidelines recommend using male formulations, given the reliability of their concentrations.[11]

Other hormonal treatments have been proposed and developed for HSDD. DHEA is a prohormone that is widely available as a supplement. It is converted into testosterone and estrogen *in vivo*.[90] While it has been promoted as improving sexual well-being, and small studies have suggested beneficial effects, systematic reviews and metaanalyses have failed to show this effect.[91-93] Tibolone is a synthetic orally active steroid that is weakly androgenic and lowers SHBG, resulting in an increase in endogenous free testosterone. The data for its efficacy in treating low sexual desire are somewhat limited, with small clinical trials showing efficacy for this endpoint but a larger metaanalysis demonstrating that it is not more effective that placebo.[94]

Genital Arousal and Sexual Pain

Genital arousal and sexual pain are often mediated by local hormonal and inflammatory markers, as mentioned earlier. There is a paucity of data showing the efficacy of any individual treatment for these conditions, with the exception of local hormone therapy for genitourinary syndrome of menopause (GSM).[95,96] GSM is covered elsewhere in this textbook, and thus this section will focus on treatments related to sexual arousal and pain.

There are several nonhormonal treatments that can be used to address sexual arousal disorders. Many of the HSDD medications mentioned earlier have been shown to improve arousal disorders in women, likely through central mechanisms affecting subjective desire. For genital arousal disorders, the main focus of research has been in increasing blood flow and sensation to the local tissue. This can be accomplished via vasodilators, including L-arginine, yohimbe, and sildenafil.[97,98] While these products usually can be compounded, there are over-the-counter formulations that have shown modest benefits (Zestra)[99]

For sexual pain, nonhormonal treatments usually include local neuropathic pain medication, such as lidocaine 2% to 5%, compounded gabapentin 6%, ketamine 10%, and local tricyclic antidepressants (amitriptyline or nortriptyline, 2%–5%), as well as capsaicin.[100] It should be noted that women with sexual pain disorders that are neuropathic in nature also may benefit from systemic treatment with gabapentin, pregabalin, or tricyclic antidepressants, but the data for efficacy are lacking.[101,102]

All women should be counseled about proper use of lubricants for sexual activity. Water-based lubricants are the most common, but may not be the most effective, particularly in circumstances where prolonged duration of sexual activity may require reapplication. Silicone-based lubricants last longer, as do oil-based lubricants, although they may not be safely combined with latex condoms, depending upon the specific active agents. Additionally, there are numerous vaginal moisturizers (typically hydrophilic agents) that purport to improve symptoms of dryness, and thus pain, and while few have actually been studied in rigorous trials, it is well-established that there is no regulation, and indeed very little consistency, in ensuring that both lubricant and moisturizers are pH- and osmolality-balanced for the vagina. [03]

Hormones for Local Vaginal Pain. Given the known sensitivity of vaginal and vulvar tissue to hormonal fluctuations, it is not surprising that local hormone therapy can be effective for sexual pain disorders that are not related to menopause. Separate from menopause-related GSM, many women, both pre- and postmenopausal, suffer from provoked vestibulodynia. This disorder, confined to the vaginal vestibule, is thought to arise in tissues derived from embryonic endoderm. Such tissues may lack androgen sensitivity.[104] As such, this vestibular tissue may require the addition of androgens for pain relief.[44] There are published data supporting the use of both estradiol (0.0%–-0.03%) and testosterone (0.1%) for vestibulodynia.[44] This is first-line therapy for provoked vestibulodynia that can be attributed to hormonal changes, such as that produced by long-term combined hormonal contraceptives, but can also be used for women with other types of provoked vestibulodynia. There is significant evidence for the use of prasterone (DHEA) for dyspareunia, but the research has focused on dyspareunia due to GSM rather than provoked vestibulodynia.[16]

KEY POINTS

- There are two FDA approved treatments for sexual desire disorder in premenopausal women, both targeting neurotransmitters. These treatments also benefit postmenopausal women.
- Testosterone has strong published data documenting benefits for low sexual desire for postmenopausal women, side effects are mainly limited to acne and hirsutism.
- Long-term safety data are limited There are no commercially available testosterone treatments for sexual dysfunctions in women outside Australia; treatment must be with male formulations at reduced doses or compounded therapies.

PERSISTENT GENITAL AROUSAL DISORDER

The final and most unrecognized sexual dysfunction is an entity known as persistent genital arousal disorder (PGAD)/genitopelvic dysethesia (GD). This condition is characterized by persistent sensations of genital arousal in the absence of sexual stimulation or emotion that are considered unwanted and cause the patient at least moderate distress.[46] The proposed etiologies of this condition are plentiful and may involve a range of psychologic, pharmacologic, neurologic, and vascular causes, including medication cessation, genital hormonal changes, local genital masses, and spinal masses.[46] The treatments depend on the etiology, and a detailed history and physical exam, as well as imaging that may involve spinal MRI, are warranted. This condition can lead to significant distress, is often misdiagnosed, and has even been associated with suicide, and should be taken seriously by the evaluating clinician.

CONCLUSION

Female sexual dysfunction is a highly complex set of disorders with multiple etiologies. The practicing endocrinologist may see a range of female sexual health complaints. Each woman should be evaluated from a biopsychosocial perspective to assure incorporation of the many factors that may play a role in a particular patient's presenting symptoms. Once a detailed history is obtained, many women will require a physical exam. The most common sexual dysfunctions are low desire and sexual pain. Sexual arousal and orgasmic disorders and PGAD/GD should be considered. Each of these has specific treatment options depending on a patient's specific complaints, exam findings, and comorbidities. Female sexual health is highly correlated to overall health and well-being and should be considered crucial to any clinical evaluation, and promotion of sexual wellness should be incorporated into any treatment plan for endocrinologic disease.

REFERENCES

1. Koops TU, Briken P. Prevalence of female sexual function difficulties and sexual pain assessed by the female sexual function index: a systematic review. *J Sex Med*. 2018;15:1591–1599.
2. Shifren JL, Monz BU, Russo PA, et al. Sexual problems and distress in United States women: prevalence and correlates. *Obstet Gynecol*. 2008;112:970–978.
3. Kingsberg SA, Clayton AH, Pfaus JG. The female sexual response: current models, neurobiological underpinnings and agents currently approved or under investigation for the treatment of hypoactive sexual desire disorder. *CNS Drugs*. 2015;29:915–933.
4. Hull EM, Lorrain DS, Du J, et al. Hormone-neurotransmitter interactions in the control of sexual behavior. *Behav Brain Res*. 1999;105:105–116.
5. Hinnell C, Hulse N, Martin A, et al. Hypersexuality and compulsive over-eating associated with transdermal dopamine agonist therapy. *Parkinsonism Relat Disord*. 2011;17:295–296.
6. Clayton AH. The pathophysiology of hypoactive sexual desire disorder in women. *Int J Gynaecol Obstet*. 2010;110:7–11.
7. Bijlsma EY, Chan JS, Olivier B, et al. Sexual side effects of serotonergic antidepressants: mediated by inhibition of serotonin on central dopamine release? *Pharmacol Biochem Behav*. 2014;121:88–101.
8. Cappelletti M, Wallen K. Increasing women's sexual desire: the comparative effectiveness of estrogens and androgens. *Horm Behav*. 2016;78:178–193.
9. Thomas MP, Potter BV. The structural biology of oestrogen metabolism. *J Steroid Biochem Mol Biol*. 2013;137:27–49.
10. Locklear MN, Michaelos M, Collins WF, et al. Gonadectomy but not biological sex affects burst-firing in dopamine neurons of the ventral tegmental area and in prefrontal cortical neurons projecting to the ventral tegmentum in adult rats. *Eur J Neurosci*. 2017;45:106–120.
11. Davis SR, Baber R, Panay N, et al. Global consensus position statement on the use of testosterone therapy for women. *J Clin Endocrinol Metab*. 2019;104:4660–4666.
12. Davison SL, Bell R, Donath S, et al. Androgen levels in adult females: changes with age, menopause, and oophorectomy. *J Clin Endocrinol Metab*. 2005;90:3847–3853.
13. Davis SR, Davison SL, Donath S, al at. Circulating androgen levels and self-reported sexual function in women. *JAMA*. 2005;294:91–96.
14. Roney JR, Simmons ZL. Hormonal predictors of sexual motivation in natural menstrual cycles. *Horm Behav*. 2013;63:636–645.
15. Dennerstein L, Randolph J, Taffe J, et al. Hormones, mood, sexuality, and the menopausal transition. *Fertil Steril*. 2002;77:S42–S48.
16. Simon JA, Goldstein I, Kim NN, et al. The role of androgens in the treatment of genitourinary syndrome of menopause (GSM): International Society for the Study of Women's Sexual Health (ISSWSH) expert consensus panel review. *Menopause*. 2018;25:837–847.
17. Muhleisen AL, Herbst-Kralovetz MM. Menopause and the vaginal microbiome. *Maturitas*. 2016;91:42–50.
18. Labrie F, Martel C, Belanger A, et al. Androgens in women are essentially made from DHEA in each peripheral tissue according to intracrinology. *J Steroid Biochem Mol Biol*. 2017;168:9–18.

19. Traish AM, Botchevar E, Kim NN. Biochemical factors modulating female genital sexual arousal physiology. *J Sex Med*. 2010;7:2925–2946.
20. Traish AM, Kim N, Min K, et al. Role of androgens in female genital sexual arousal: receptor expression, structure, and function. *Fertil Steril*. 2002;77:S11–S18.
21. Levin RJ, Both S, Georgiadis J, et al. The physiology of female sexual function and the pathophysiology of female sexual dysfunction (committee 13A). *J Sex Med*. 2016;13:733–759.
22. Park K, Kim NN. Anatomy and physiology of arousal. In: Goldstein ICA, Goldstein AT, Kim NN, et al., eds. *Textbook of Female Sexual Function and Dysfunction: Diagnosis and Treatment*. Oxford: Wiley Blackwell; 2018:107–125.
23. Levin R. The pharmacology of the human female orgasm - its biological and physiological backgrounds. *Pharmacol Biochem Behav*. 2014;121:62–70.
24. Traish A, Vignozzi L, Goldstein I, et al. Role of androgens and estrogens in female genitourinary tract structure and function: implications in the genitourinary syndrome of menopause. *Sex Med Rev*. 2018:558–571.
25. Comeglio P, Cellai I, Filippi S, et al. Differential effects of testosterone and estradiol on clitoral function: an experimental study in rats. *J Sex Med*. 2016;13:1858–1871.
26. *Diagnostic and Statistical Manual*. Washington, DC: American Psychiatric Association; 1952;1.
27. *Diagnostic and Statistical Manual IV*. Washington, DC: American Psychiatric Association; 2004.
28. *Diagnostic and Statistical Manual V*. Washington, DC: American Psychiatric Association; 2013.
29. Derogatis LR, Sand M, Balon R, et al. Toward a more evidence-based nosology and nomenclature for female sexual dysfunctions-part I. *J Sex Med*. 2016;13:1881–1887.
30. Campbell MM, Artz L, Stein DJ. Sexual disorders in DSM-5 and ICD-11: a conceptual framework. *Curr Opin Psychiatr*. 2015;28:435–439.
31. WHO. International. Classification of diseases 2018. Available from: https://icd.who.int/dev11/.
32. Goldstein I, Kim NN, Clayton AH, et al. Hypoactive sexual desire disorder: International Society for the Study of Women's Sexual Health (ISSWSH) expert consensus panel review. *Mayo Clin Proc*. 2017;92(1):114–128.
33. Derogatis LR, Revicki DA, Clayton AH. Instruments for screening, diagnosis, and management of patients with generalized acquired hypoactive sexual desire disorder. *J Womens Health (Larchmt)*. 2020;29:806–814.
34. Bornstein J, Goldstein AT, Stockdale CK, et al. 2015 ISSVD, ISSWSH and IPPS consensus terminology and classification of persistent vulvar pain and vulvodynia. *Obstet Gynecol*. 2016;127:745–751.
35. Clayton A, Goldstein I, Kim NN, et al. The International Society for the Study of Women's Sexual Health process of care for management of hypoactive sexual desire disorder in women. *Mayo Clin Proc*. 2018;93:467–487.
36. Parish SJ, Goldstein AT, Goldstein SW, et al. Toward a more evidence-based nosology and nomenclature for female sexual dysfunctions-part II. *J Sex Med*. 2016;13:1888–1906.
37. Parish SJ, Meston CM, Althof SE, et al. Toward a more evidence-based nosology and nomenclature for female sexual dysfunctions-part III. *J Sex Med*. 2019;16:452–462.
38. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther*. 2000;26:191–208.
39. Taylor JF, Rosen RC, Leiblum SR. Self-report assessment of female sexual function: psychometric evaluation of the Brief Index of Sexual Functioning for Women. *Arch Sex Behav*. 1994;23(6):627–643.
40. Rosen NO, Bergeron S, Pukall CF. Recommendations for the study of vulvar pain in women, part 1: review of assessment tools. *J Sex Med*. 2020;17:180–194.
41. Bianchi-Demicheli F, Cojan Y, Waber L, et al. Neural bases of hypoactive sexual desire disorder in women: an event-related FMRI study. *J Sex Med*. 2011;8:2546–2559.
42. Cacioppo S. Neuroimaging of female sexual desire and hypoactive sexual desire disorder. *Sex Med Rev*. 2017;5:434–444.
43. Aalto A, Huotari-Orava R, Luhtala S, et al. Expression of estrogen-related receptors in localized provoked vulvodynia. *Biores Open Access*. 2020;9:13–21.
44. Burrows LJ, Goldstein AT. The treatment of vestibulodynia with topical estradiol and testosterone. *Sex Med*. 2013;1:30–33.
45. Pessina MA, Hoyt Jr RF, Goldstein I, et al. Differential regulation of the expression of estrogen, progesterone, and androgen receptors by sex steroid hormones in the vagina: immunohistochemical studies. *J Sex Med*. 2006;3:804–814.
46. Pfaus JG. Persistent genital arousal disorder-fact or fiction? *J Sex Med*. 2017;14:318–319.
47. Sipski ML, Alexander CJ, Rosen R. Sexual arousal and orgasm in women: effects of spinal cord injury. *Ann Neurol*. 2001;49:35–44.
48. Nascimento ER, Maia AC, Pereira V, et al. Sexual dysfunction and cardiovascular diseases: a systematic review of prevalence. *Clinics*. 2013;68:1462–1468.
49. Gaither TW, Awad MA, Murphy GP, et al. Cycling and female sexual and urinary function: results from a large, multinational, cross-sectional study. *J Sex Med*. 2018;15:510–518.
50. Meston CM, Stanton AM. Understanding sexual arousal and subjective-genital arousal desynchrony in women. *Nat Rev Urol*. 2019;16:107–120.
51. Ishak WW, Bokarius A, Jeffrey JK, et al. Disorders of orgasm in women: a literature review of etiology and current treatments. *J Sex Med*. 2010;7:3254–3268.
52. Rowland DL, Kolba TN. Understanding orgasmic difficulty in women. *J Sex Med*. 2016;13:1246–1254.
53. Montejo-González AL, Llorca G, Izquierdo JA, et al. SSRI-induced sexual dysfunction: fluoxetine, paroxetine, sertraline, and fluvoxamine in a prospective, multicenter, and descriptive clinical study of 344 patients. *J Sex Marital Ther*. 1997;23:176–194.
54. Meston CM, Levin RJ, Sipski ML, et al. Women's orgasm. *Annu Rev Sex Res*. 2004;15:173–257.
55. Sorensen J, Bautista KE, Lamvu G, et al. Evaluation and treatment of female sexual pain: a clinical review. *Cureus*. 2018 Mar 27;10:e2379.
56. Goldstein AT, Belkin ZR, Krapf JM, et al. Polymorphisms of the androgen receptor gene and hormonal contraceptive induced provoked vestibulodynia. *J Sex Med*. 2014;11:2764–2771.
57. Goldstein A, Burrows L, Goldstein I. Can oral contraceptives cause vestibulodynia? *J Sex Med*. 2010;7:1585–1587.
58. Sutton KS, Yessick LR, Wild CJ, et al. Exploring the neural correlates of touch and pain in women with provoked vestibulodynia. *Pain*. 2020;161:926–937.
59. Stahl SM. Mechanism of action of flibanserin, a multifunctional serotonin agonist and antagonist (MSAA), in hypoactive sexual desire disorder. *CNS Spectr*. 2015;20:1–6.
60. Simon JA, Thorp J, Millheiser L. Flibanserin for premenopausal hypoactive sexual desire disorder: pooled analysis of clinical trials. *J Womens Health (Larchmt)*. 2019;28:769–777.
61. Portman DJ, Brown L, Yuan J, et al. Flibanserin in postmenopausal women with hypoactive sexual desire disorder: results of the PLUMERIA study. *J Sex Med*. 2017;14:834–842.
62. Clayton AH, Brown L, Kim NN. Evaluation of safety for flibanserin. *Expert Opin Drug Saf*. 2020;19:1–8.
63. Simon JA, Clayton AH, Parish SJ, et al. Effects of alcohol administered with flibanserin in healthy premenopausal women: a randomized, double-blind, single-dose crossover study. *J Sex Med*. 2020;17:83–93.
64. Pfaus J, Giuliano F, Gelez H. Bremelanotide: an overview of preclinical CNS effects on female sexual function. *J Sex Med*. 2007;4:269–279.
65. Kingsberg SA, Clayton AH, Portman D, et al. Bremelanotide for the treatment of hypoactive sexual desire disorder: two randomized phase 3 trials. *Obstet Gynecol*. 2019;134:899–908.
66. Segraves RT, Clayton A, Croft H, et al. Bupropion sustained release for the treatment of hypoactive sexual desire disorder in premenopausal women. *J Clin Psychopharmacol*. 2004;24:339–342.
67. Clayton AH, McGarvey EL, Abouesh AI, et al. Substitution of an SSRI with bupropion sustained release following SSRI-induced sexual dysfunction. *J Clin Psychiatr*. 2001;62:185–190.

68. Masand PS, Ashton AK, Gupta S, Frank B. Sustained-release bupropion for selective serotonin reuptake inhibitor-induced sexual dysfunction: a randomized, double-blind, placebo-controlled, parallel-group study. *Am J Psychiatr*. 2001;158:805–807.
69. Landén M, Eriksson E, Agren H, et al. Effect of buspirone on sexual dysfunction in depressed patients treated with selective serotonin reuptake inhibitors. *J Clin Psychopharmacol*. 1999;19:268.
70. Pyke RE. Trazodone in sexual medicine: underused and overdosed?. *Sex Med Rev*. 2020;8. 206e216.206e271.
71. Cayan F, Dilek U, Pata O, et al. Comparison of the effects of hormone therapy regimens, oral and vaginal estradiol, estradiol + drospirenone and tibolone, on sexual function in healthy postmenopausal women. *J Sex Med*. 2008;5:132–138.
72. Huang A, Yaffe K, Vittinghoff E, et al. The effect of ultralow-dose transdermal estradiol on sexual function in postmenopausal women. *Am J Obstet Gynecol*. 2008;198:265. e1–e7.
73. Taylor HS, Tal A, Pal L, et al. Effects of oral vs transdermal estrogen therapy on sexual function in early postmenopause: ancillary study of the Kronos Early Estrogen Prevention Study (KEEPS). *JAMA Intern Med*. 2017;177:1471–1479.
74. Somboonporn W, Davis S, Seif M, et al. Testosterone for peri- and postmenopausal women. *Cochrane Database Syst Rev*. 2005;4:CD004509. updated 2009.
75. Achilli C, Pundir J, Ramanathan P, et al. Efficacy and safety of transdermal testosterone in postmenopausal women with hypoactive sexual desire disorder: a systematic review and meta-analysis. *Fertil Steril*. 2017;107:475–482.
76. Burger HG, Hailes J, Nelson J, et al. Effect of combined implants of estradiol and testosterone on libido in postmenopausal women. *BMJ*. 1987;294:936–937.
77. Davis SR, McCloud PI, Strauss BJG, et al. Testosterone enhances estradiol's effects on postmenopausal bone density and sexuality. *Maturitas*. 1995;21:227–236.
78. Lobo R, Rosen RC, Yang H-M, et al. Comparative effects of oral esterified estrogens with and without methyl testosterone on endocrine profiles and dimensions of sexual function in postmenopausal women with hypoactive sexual desire. *Fertil Steril*. 2003;79:1341–1352.
79. Floter A, Nathorst-Boos J, Carlstrom K, et al. Addition of testosterone to estrogen replacement therapy in oophorectomized women: effects on sexuality and well-being. *Climacteric*. 2002;5:357.
80. Davis SR, Moreau M, Kroll R, et al. Testosterone for low libido in menopausal women not taking estrogen therapy. *N Eng J Med*. 2008;359:2005–2017.
81. Islam RM, Bell RJ, Green S, et al. Safety and efficacy of testosterone for women: a systematic review and meta-analysis of randomised controlled trial data. *Lancet Diabetes Endocrinol*. 2019;7:754–766.
82. Goldstat R, Briganti E, Tran J, et al. Transdermal testosterone therapy improves well-being, mood, and sexual function in premenopausal women. *Menopause*. 2003;10:390–398.
83. Elraiyah T, Sonbol MB, Wang Z, et al. Clinical review: the benefits and harms of systemic testosterone therapy in postmenopausal women with normal adrenal function: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2014;99:3543–3550.
84. Nachtigall L, Casson P, Lucas J, et al. Safety and tolerabililty of testosterone patch therapy for up to 4 years in surgically menopausal women receiving oral or transdermal oestrogen. *Gynecol Endocrinol*. 2011;27:39–48.
85. White WB, Grady D, Giudice LC, et al. A cardiovascular safety study of LibiGel (testosterone gel) in postmenopausal women with elevated cardiovascular risk and hypoactive sexual desire disorder. *Am Heart J*. 2012;163:27–32.
86. Pharmaceuticals B. *BioSante Pharmaceuticals Announces Positive LibiGel Phase III Safety Data Review and Decision to Conclude the Safety Study*; 2012. [cited 2019 June 21]. Available from: https://www.businesswire.com/news/home/20120904005404/en/BioSante-Pharmaceuticals-Announces-Positive-LibiGel%C2%AE-Phase-III.
87. Wierman M, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:2489–3510. 88.
88. Jayasena CN, Alkaabi FM, Liebers CS, et al. A systematic review of randomized controlled trials investigating the efficacy and safety of testosterone therapy for female sexual dysfunction in postmenopausal women. *Clin Endocrinol*. 2019;90:391–414.
89. Simon JA, Kapner MD. The saga of testosterone for menopausal women at the Food and Drug Administration (FDA). *J Sex Med*. 2020;17:826–829.
90. Zhou Y, Kang J, Chen D, et al. Ample evidence: dehydroepiandrosterone (DHEA) conversion into activated steroid hormones occurs in adrenal and ovary in female rat. *PLoS One*. 2015;10:e0124511.
91. Kushnir VA, Darmon SK, Barad DH, et al. Effects of dehydroepiandrosterone (DHEA) supplementation on sexual function in premenopausal infertile women. *Endocrine*. 2019;63:632–638.
92. Panjari M, Bell RJ, Jane F, et al. A randomized trial of oral DHEA treatment for sexual function, well-being, and menopausal symptoms in postmenopausal women with low libido. *J Sex Med*. 2009;6:2579–2590.
93. Panjari M, Davis SR. DHEA therapy for women: effect on sexual function and wellbeing. *Hum Reprod Update*. 2007;13:239–248.
94. Modelska K, Cummings S. Tibolone for postmenopausal women: systematic review of randomized trials. *J Clin Endocrinol Metab*. 2002;87:16–23.
95. Lethaby A, Ayeleke RO, Roberts H. Local oestrogen for vaginal atrophy in postmenopausal women. *Cochrane Database Syst Rev*. 2016;8:CD001500.
96. Lev-Sagie A. Vulvar and vaginal atrophy: physiology, clinical presentation, and treatment considerations. *Clin Obstet Gynecol*. 2015;58:476–491.
97. Leddy LS, Yang CC, Stuckey BG, et al. Influence of sildenafil on genital engorgement in women with female sexual arousal disorder. *J Sex Med*. 2012;9:2693–2697.
98. Meston CM, Worcel M. The effects of yohimbine plus L-arginine glutamate on sexual arousal in postmenopausal women with sexual arousal disorder. *Arch Sex Behav*. 2002;31:323–332.
99. Ferguson DM, Hosmane B, Heiman JR. Randomized, placebo-controlled, double-blind, parallel design trial of the efficacy and safety of Zestra in women with mixed desire/interest/arousal/orgasm disorders. *J Sex Marital Ther*. 2010;36:66–86.
100. Stenson AL. Vulvodynia: diagnosis and management. *Obstet Gynecol Clin North Am*. 2017;44:493–508.
101. Bachmann GA, Brown CS, Phillips NA, et al. Effect of gabapentin on sexual function in vulvodynia: a randomized, placebo-controlled trial. *Am J Obstet Gynecol*. 2019;220:89. e1–e8.
102. van Beekhuizen HJ, Oost J, van der Meijden WI. Generalized unprovoked vulvodynia; A retrospective study on the efficacy of treatment with amitriptyline, gabapentin or pregabalin. *Eur J Obstet Gynecol Reprod Biol*. 2018;220:118–121.
103. Edwards D, Panay N. Treating vulvovaginal atrophy/genitourinary syndrome of menopause: how important is vaginal lubricant and moisturizer composition? *Climacteric*. 2016;19:151–161.
104. Goldstein AT, Belkin ZR, Krapf JM, et al. Polymorphisms of the androgen receptor gene and hormonal contraceptive induced provoked vestibulodynia. *J Sex Med*. 2014;11:2764–2771.

129

Hormonal Control of Human Pregnancy and Parturition

Sam Mesiano and Roger Smith

OUTLINE

Establishment of Pregnancy, 2135
  Preparation of the Endometrium, 2135
  Embryo Competency, 2135
  Decidualization of the Endometrium, 2136
  Implantation and Placentation, 2136
  Immune Tolerance of the Conceptus, 2138
Pregnancy Maintenance, 2138
  Chorionic Gonadotropin, 2138
  Gonadotropin-Releasing Hormone, 2139
  Proopiomelanocortin-Derived Hormones, 2139
  Corticotropin-Releasing Hormone, 2139
  Placental Lactogen and Placental Growth Hormone, 2140
  Modulation of Maternal Energy Homeostasis, 2140
  Growth Factors and Cytokines, 2141
  Steroid Hormones, 2142
Parturition and Birth Timing, 2143
  Evolutionary Perspective, 2143
  Fetal Maturation, 2144
  Progesterone Withdrawal, 2144
  Stress and Tissue-Level Inflammation, 2145
  Progestin Therapy to Prevent Preterm Birth, 2145
  Estrogens, 2146
  Integrative Role of Placental Corticotropin-Releasing Hormone, 2146
Conclusion/Future Directions, 2147

Pregnancy is a state in which one or more fertilized eggs is retained within the uterus and provisioned with oxygen and nutrients for the specific amount of time required to complete the embryonic/fetal development program. Pregnancy ends by the process of parturition, whereby the products of conception (i.e., the embryo/fetus and extraembryonic tissues) are expelled from the uterus. Hormone signaling networks control pregnancy and parturition by:

- Creating a suitable intrauterine environment to enable attachment and implantation of the embryo.
- Signaling the presence of a viable embryo to sustain a propregnancy uterine phenotype and maternal physiology.
- Controlling implantation of the embryo into the uterine wall and the formation of a placenta.
- Modulating maternal physiology to supply oxygen and nutrients to the conceptus.
- Blocking the process of parturition and physically accommodating the conceptus for a gestational period needed for fetal growth and maturation.
- Triggering parturition at a time when the fetus is sufficiently mature to survive as a neonate.
- Establishing maternal lactation and behaviors required to protect and nurture the newborn.

The hormones of pregnancy include multiple proteins, lipids, and steroids acting in endocrine and/or paracrine modes. This chapter will focus on key hormonal signals that control and coordinate the establishment and maintenance of human pregnancy and the timing and process of parturition.

KEY POINTS

- Pregnancy follows a maturational process leading to delivery of a fetus able to survive in the extrauterine environment.

ESTABLISHMENT OF PREGNANCY

Pregnancy is established by implantation of the developing embryo into the uterine wall. This requires synchrony between a receptive endometrium and a competent embryo. For implantation, the competent embryo, usually at the blastocyst stage, first adheres to luminal epithelial cells of the endometrium and then invades into the uterine wall, where it forms a physical and functional attachment. The process includes development of the placenta to establish a biochemical and physiologic link with the mother.

Preparation of the Endometrium

The uterus is prepared for implantation by the ovarian steroid hormones estradiol and progesterone that induce cyclic changes in morphology and function of the endometrium as part of the menstrual cycle. During the follicular phase of the ovarian cycle, estradiol promotes proliferation and thickening of the functionalis layer of the endometrium. During the postovulatory luteal phase, progesterone induces differentiation of the functionalis into a glandular, secretory, and highly vascularized tissue that is receptive to embryo implantation. In women, the progesterone/estradiol-primed endometrium is most receptive to embryo implantation between days 20 and 24 of the menstrual cycle (i.e., 6–10 days after ovulation).[1] The probability of embryo implantation before and after this time window is low.

KEY POINTS

- Ovarian estrogen and progesterone prepare the endometrium for implantation of the blastocyst 6–10 days after ovulation.

Embryo Competency

Upon fertilization and syngamy, the nascent embryo, while transiting the fallopian tube toward the uterus, undergoes an intrinsic program of cell division and differentiation to become a ball of cells known as a morula, encapsulated by the remnant of the oocyte zona pellucida. At around the fifth day postsyngamy (around day 19 of the menstrual cycle) the morula enters the uterine cavity, where it floats freely for several days. As the number of cells in the embryo increases, differentiation occurs to produce two distinct cell types: (1) the outer trophectoderm composed of trophoblast cells that will differentiate into the extraembryonic tissue (placenta and chorionic membrane), and (2) a cluster of inner cells, referred to as the inner cell mass, that will give rise to the embryo/fetus. Differentiation of the two cell types is influenced by multiple transcription factors, especially members of the homeobox family. A critical step is expression of the NANOG homeobox protein, which drives the differentiation of inner cells to acquire a pluripotency phenotype necessary to give rise to the embryo/fetus.[2] Cells on the surface of the morula do not express NANOG and differentiate to form trophoblast cells that will give rise to the placenta and chorionic membrane.

KEY POINTS

- Cells on the surface of the morula do not express the homeobox protein NANOG and become the trophectoderm that forms the placenta. Inner cells do express NANOG and form the inner cell mass that develops into the embryo.

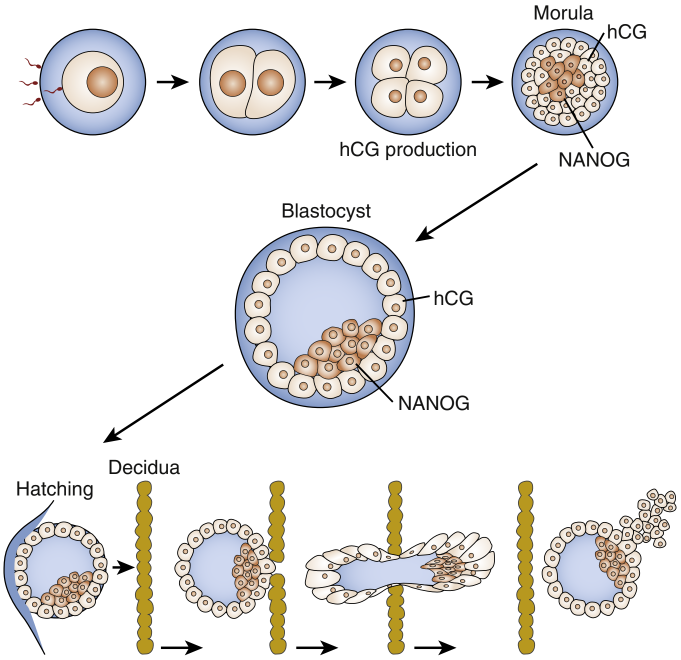


Fig. 129.1 **Formation of the morula with production of human chorionic gonadotropin (*CG*), the first signal to the mother of the presence of pregnancy. As cell division occurs, CG production increases.** The outer cells cease to express homeobox protein NANOG, which is required for totipotency, and form the trophectoderm and later the placenta. These cells transport fluid in an inward direction, leading to the formation of the blastocyst. The central cells, which continue to express NANOG, retain totipotency and form the inner cell mass of the blastocyst and later the embryo. Before attachment of the blastocyst to the endometrial lining of the uterus, the blastocyst hatches from the zona pellucida. The blastocyst then penetrates the decidualized endometrium, and the trophectoderm begins to invade.

At around day 7 the zona pellucida degenerates, and the embryo "hatches" to allow direct contact of trophectoderm with endometrial epithelial cells. At this stage the embryo develops a fluid-filled cavity and is referred to as a blastocyst. At the same time luminal endometrial epithelial cells produce chemokines and cell adhesion molecules that attract the blastocyst to domelike structures known as pinopodes, to which trophoblast cells adhere and produce proteases that degrade the endometrial extracellular matrix (ECM), opening a path for the embryo to migrate into the uterine wall (Fig. 129.1).

Decidualization of the Endometrium

Late in the luteal phase, in fertile and nonfertile cycles, progesterone and increased intracellular cyclic adenosine monophosphate (cAMP) in stromal cells of the endometrium induce changes collectively referred to as decidualization, which includes proliferation and hypertrophy of stromal fibroblasts, formation of a network of contact processes between neighboring stromal fibroblasts, shrinkage and altered secretory activity of glandular epithelial cells, and infiltration of specific leukocytes including natural killer (NK) cells, T-cells, and macrophages.[3-5] Increased cAMP is stimulated by chorionic gonadotropin (CG) secreted by the trophoblast cells of the blastocyst.[6,7] Endometrium in this state is referred to as the decidua. Decidualization begins during the latter part of the luteal phase at endometrial stromal cells adjacent to spiral arterioles. If embryo implantation occurs, the decidualization reaction spreads to include the entire endometrium.

KEY POINTS

- Chorionic gonadotropin secreted by the trophectoderm and progesterone secreted by the maternal ovary transform the endometrium into the decidua that is receptive to implantation.

The decidua facilitates deep implantation and placentation of the embryo and is essential for the establishment and maintenance of human pregnancy. ECM produced by decidualized endometrial stromal cells guides and promotes trophoblast invasion, and the stromal cells provide nutritional support for the implanting blastocyst. However, the dense cellular matrix of the decidua also restricts *de novo* embryo implantation, thus closing the implantation window.[3,5] During pregnancy the decidua exists as three types, determined mainly by anatomy: (1) the basalis that lies directly adjacent to the implantation site and, eventually, the placental disc; (2) the capsularis that surrounds the embryo to separate it from the uterine cavity; and (3) the parietalis that comprises the remainder of the endometrium. The capsularis eventually fuses with the parietalis when the embryo/fetus grows to occupy the entire uterine cavity. Formation and maintenance of the decidua is dependent on progesterone in conjunction with intracellular cAMP in decidual stromal cells.

Implantation and Placentation

The process of implantation and placentation is controlled primarily by paracrine signaling between the embryo and endometrial cells involving multiple hormones including chemokines, cytokines, cell adhesion molecules, matrix metalloproteinases and their inhibitors, *HOX* gene products, and various growth factors and angiogenic factors.[8] These molecules are thought to mediate the effects of progesterone and CG to produce a receptive endometrium and an intrauterine milieu conducive to embryo survival. Locally acting cytokines, especially in response to progesterone, are also thought to condition maternal immune cells resident in the decidua to tolerate the semiallogeneic conceptus.[9]

Trophoblast cells, upon adhesion to pinopodes, produce proteases that degrade the proximate ECM to produce a pathway for invasion of the embryo into the endometrial stroma. Trophoblast cells then divide

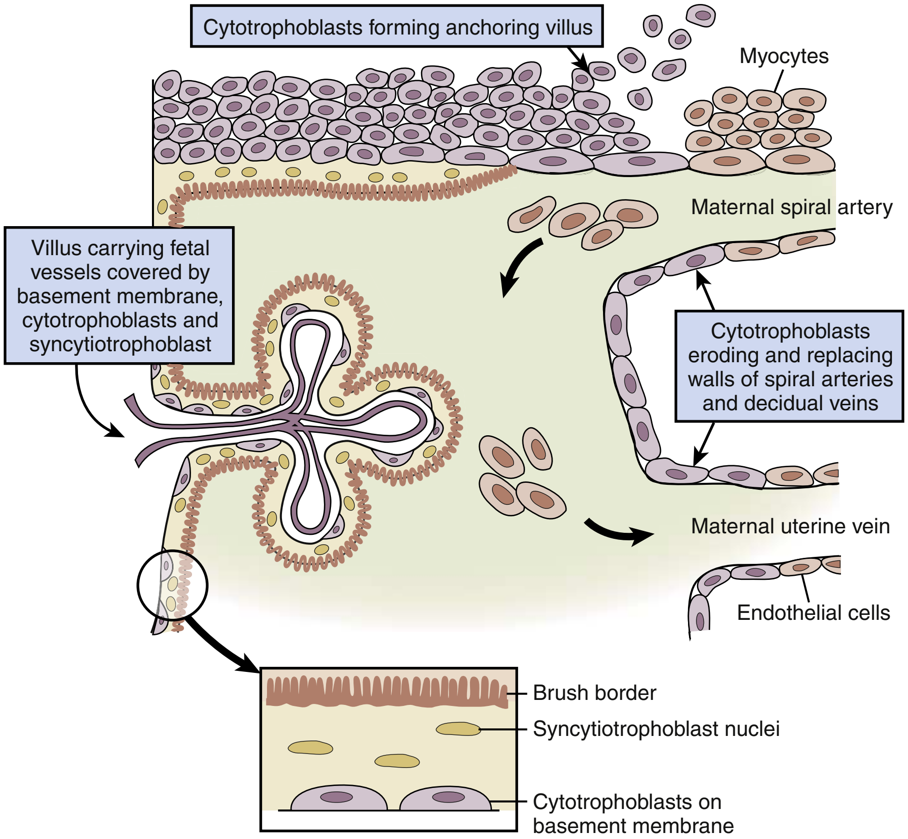

Cytotrophoblasts forming anchoring villus
Myocytes
Maternal spiral artery
Villus carrying fetal vessels covered by basement membrane, cytotrophoblasts and syncytiotrophoblast
Cytotrophoblasts eroding and replacing walls of spiral arteries and decidual veins
Maternal uterine vein
Endothelial cells
Brush border
Syncytiotrophoblast nuclei
Cytotrophoblasts on basement membrane

Fig. 129.2 Cytotrophoblasts grow into the maternal decidua from a basement membrane. The cytotrophoblasts differentiate into several cell types. Some form dense columns of cells that form anchoring villi that attach the placenta to the lining of the uterus, others erode into the maternal spiral arteries and transform into an endothelial phenotype that lines these vessels, and an additional group fuses to form the syncytiotrophoblast layer that lines the villus space.

and differentiate into an invasive phenotype, referred to as cytotrophoblasts, that form columns that penetrate into the uterine wall to anchor the blastocyst deep in the endometrial stroma. Some cytotrophoblasts invade beyond the endometrial basement membrane and through to the inner third of the myometrium toward uterine spiral arterioles, where changes in oxygen tension cause expression of genes that alter their cell phenotype toward that of endothelial cells. The modified cytotrophoblast cells then invade into the arterioles, where they replace the vascular endothelial lining and destroy the arterial smooth muscle. Elimination of endothelial and musculoelastic tissue increases the diameter of the spiral arteries, which significantly increases flow of maternal arterial blood to the conceptus (based on Poiseuille's law: flow through a cylinder is proportional to the fourth power of the radius of the cylinder). This process is critical to ensure an adequate supply of maternal oxygen and nutrients to the developing conceptus, especially in the later stages of pregnancy. Shallow or defective cytotrophoblast invasion into spiral arterioles during implantation is thought to restrict subsequent blood flow to the conceptus and induce a compensatory increase in maternal blood pressure to ensure adequate flow to the placenta. This compensatory mechanism may lead to preeclampsia and/or intrauterine growth restriction.[10,11]

At the same time as vascular cytotrophoblasts invade into the spiral arterioles and disrupt the arterial-venous connection, other cytotrophoblasts fuse to form a multinucleated cell, referred to as the syncytiotrophoblast. At this stage the developing inner cell mass is surrounded by an outer syncytiotrophoblast with a layer of underlying proliferating trophoblast cells, and the embryo is anchored to the uterine vasculature and completely surrounded by endometrial stroma. During embryo invasion, the syncytiotrophoblast forms trabeculae that create fluid-filled spaces, known as lacunae, between the eroded spiral arterioles and associated veins. At the same time cytotrophoblast cells underlying the syncytiotrophoblast proliferate to form finger-like protrusions containing mesenchymal cells from the inner cell mass that eventually form interconnected blood vessels linked to the developing fetal cardiovascular system. The finger-like chorionic villi are the basic units for gas exchange between maternal and fetal blood and for nutrient transfer from the maternal compartment. The entire maternal-facing surface of the chorionic villi network is covered by the syncytiotrophoblast (Fig. 129.2).

During the first trimester vascular cytotrophoblasts occlude the uterine spiral arteries and limit maternal blood flow to the developing embryo. Consequently, early embryo development and organogenesis, especially fetal blood vessels forming in the chorionic villi, occur in a relatively hypoxic state. At weeks 10 to 12 of gestation the vascular plugs of cytotrophoblast cells deteriorate, leading to maternal blood entering the lacunae, where it bathes the network of chorionic villi. At this stage maternal blood directly contacts the syncytiotrophoblast. This form of placentation is referred to as a hemochorial, denoting direct contact of maternal blood with placental cells. The hemochorial arrangement of the human placenta allows the syncytiotrophoblast to secrete hormones directly into maternal blood. Consequently, the endocrine milieu of human pregnancy is dominated by hormones produced by the placental syncytiotrophoblast.[12,13] Cytotrophoblast cells in the amnion-chorion membrane also produce hormones that affect adjacent maternal cells and tissue. Most placental hormones are identical to the non-placental hormones that affect major homeostatic systems. These include hypothalamic-pituitary-adrenal and gonadal hormones, hormones that control metabolism and energy utilization, and hormones involved in the response to stress and the control of lactation. Some placental hormones are close analogs of adult hormones and serve functions specific to pregnancy. In general, placental hormones act to modulate maternal physiology in favor of maintaining pregnancy and provisioning the fetus with oxygen and nutrients.

KEY POINTS

- In the first trimester, cytotrophoblast plugs in maternal arteries maintain the developing embryo in a hypoxic environment. At the end of the first trimester these plugs dissolve, allowing maternal arterial blood to come into direct contact with the syncytiotrophoblast of the placenta. The syncytiotrophoblast layer secretes multiple hormones directly into the maternal blood that reprogram maternal metabolism to direct nutrients to the fetus.

Immune Tolerance of the Conceptus

As an immunologically foreign allograft, the conceptus is potentially at risk from attack by the maternal immune system. For the establishment and maintenance of pregnancy the maternal immune system must tolerate the semiallogeneic conceptus.[14] Several mechanisms have been proposed to explain this effect, each involving paracrine signaling by hormones and cell surface proteins between the syncytiotrophoblast, extravillous trophoblasts, decidual cells, and maternal immune cells at the maternal-fetal interface.

The syncytiotrophoblast and extravillous trophoblasts may camouflage themselves from maternal immune attack. Major histocompatibility complex (MHC) molecules are required for immune recognition and activation of a response to antigens. The usual MHC class I molecules, HLA-A and HLA-B and MHC class II antigens, are not expressed on the syncytiotrophoblast and cytotrophoblast cell surface.[15] However, trophoblasts express classic HLA-C and nonclassic HLA-E and HLA-G. The absence of MHC class I HLA-A and -B molecules prevents cytotoxic CD8 T-cell responses. HLA-C molecules down-regulate NK cell responses towards a tolerogenic state,[16-18] and HLA-G, a unique nonpolymorphic HLA, inhibits decidual NK (dNK) cell activation[19-22]. Syncytiotrophoblast and extravillous cytotrophoblasts also may inhibit proliferation of maternal T-cells in the decidua by catabolizing tryptophan, an essential amino acid for rapidly dividing T-cells.[23,24]

Activity of maternal immune cells at the maternal-fetal interface is also thought to be influenced by progesterone as part of its progestation actions, although the mechanism is not clearly understood. Maternal immune cells have low or undetectable expression of the nuclear progesterone receptor (PR), and as such progesterone actions are not considered to be mediated directly by PRs in immune cells. Instead, progesterone could act indirectly via PR-mediated effects in decidual stromal cells. For example, in decidual stromal cells progesterone stimulates the production of Fas-ligand that induces apoptosis in activated T-cells.[25-27] Progesterone may also induce chemokine activity in decidual cells to recruit dNKs with attenuated cytotoxic activity and increased immunosuppressive activity to the maternal-fetal interface.[28-30] It is also possible that progesterone affects immune cells directly via its interaction with the glucocorticoid receptor (GR). Indeed, the level of progesterone at the maternal-fetal interface (decidua basalis) at term is high (~5 μM) and likely sufficient to activate the GR.[31] Progesterone, likely via the GR, may increase the pool of regulatory T-cells (Tregs) that inhibit inflammatory Th17 cells to promote tolerance to nonself antigens.[32]

The human placental trophoblast cells are remarkable in that they expresses multiple segments of endogenous retrovirus DNA that incorporated into the hominid genome millions of years ago.[33] One particular endogenous retrovirus gene encodes syncytin-1, a placental protein that promotes syncytialization and formation of the syncytiotrophoblast that is crucial for development of the placenta.[34] Placental syncytin-1 protein contains an immunosuppressive domain that may also promote immune tolerance at the maternal-fetal interface.[35,36]

KEY POINTS

- The placental syncytiotrophoblast creates maternal immune tolerance for the semiallogeneic fetus through multiple mechanisms, including production of endogenous immunosuppressive retroviral proteins.

PREGNANCY MAINTENANCE

Placental hormones drive dramatic changes in maternal physiology that favor the developing fetus. For the fetus, early gestation is characterized by differentiation of tissues and organogenesis. From midgestation onward, the focus shifts to accumulation of nutrients, further tissue growth, and the functional maturation of organ and regulatory systems in preparation for its transition to postnatal life. In the mother, early pregnancy is a time of nutrient storage, whereas following midgestation, provision of nutrients to the fetus becomes the primary goal. In addition, the uterus must be maintained in a relaxed, closed, and quiescent state to accommodate and gestate the conceptus. These maternal adaptations are controlled mainly by hormones produced by the fetal-placental unit, the net effect of which is to favor pregnancy and supply the fetus with oxygen and nutrients to complete its developmental program. In this context fetal-placental hormones are allocrine factors, because they are produced by one organism, the fetus, to elicit a response in another organism, the mother.

The human fetal-placental unit produces a wide range of hormones, most of which are identical or close analogs of neuropeptides, hypothalamic-pituitary hormones, steroid hormones, and growth factor and cytokine protein hormones (Table 129.1).[13,37] The hormones are synthesized and released by the syncytiotrophoblast directly into the maternal circulation to act on the same receptors as their maternal counterparts. By this arrangement the placenta becomes an endocrine control center. Some hormones produced by the placenta such as corticotropin-releasing hormone (CRH) are also released into the fetal circulation.

Chorionic Gonadotropin

One of the most important hormones produce by the early embryo that is essential for the establishment of pregnancy is CG (also known as hCG). In a nonfertile cycle the maternal corpus luteum (CL) regresses late in the luteal phase, leading to a decrease in circulating progesterone. Loss of progesterone causes inflammation, vascular remodeling, and necrosis and breakdown in the luteal phase endometrium, which leads to its shedding as menses. The endometrium is then renewed from the remaining endometrial layer in response to estradiol in the following cycle. For pregnancy to occur menstruation must be prevented. To this end, in a conception cycle the early embryo produces CG, a glycoprotein hormone that is structurally and functionally analogous to pituitary luteinizing hormone (LH) (the β-subunits of LH and CG differ, and the α-subunits are identical). CG prevents regression of

TABLE 129.1 Hormones Produced by the Human Placenta

Neuropeptides	Corticotropin-releasing hormone, thyrotropin-releasing hormone, gonadotropin-releasing hormone, melatonin, cholecystokinin, metenkephalin, dynorphin, neurotensin, vasoactive intestinal peptide, galanin, somatostatin, calcitonin gene–related peptide, neuropeptide Y, substance P, endothelin, atrial natriuretic peptide, renin, angiotensin, urocortin, kisspeptin
Pituitary hormones[46]	Adrenocorticotropic hormone, thyroid-stimulating hormone, placental growth hormone, placental lactogen, chorionic gonadotropin, lutenizing hormone, follicle-stimulating hormone, β-endorphin, prolactin, oxytocin
Adipokines	Adiponectin, leptin, resistin, visfatin, ghrelin
Growth factors[46]	Insulin-like growth factor-1, insulin-like growth factor-2, vascular endothelial growth factor, epidermal growth factor, activins, inhibins, follistatin
Cytokines	Interleukin-1, interleukin-2, interleukin-6, interleukin-8, interferon-α, interferon-β, interferon-γ, tumor necrosis factor
Steroids	Progesterone, estradiol, estrone, estriol, estetrol, 2-methoxyestradiol, allopregnanolone, pregnenolone, 5α-dihydroprogesterone
Monoamines	Epinephrine, norepinepherine, dopamine, serotonin, adrenomedullin
Lipids	Prostaglandin-F$_{2\alpha}$, prostaglandin-E$_2$, prostaglandin-I, thromboxane$_{A2}$

the CL and stimulates its continued production of progesterone, maintaining the endometrium and thus preventing menstruation.

Production of CG by the embryo begins at the eight-cell stage under drive from the endogenous retrovirus protein syncytin-1 and is subsequently produced by the syncytiotrophoblast.[38] Blastocyst implantation is associated with the first detectable levels of CG (indicated by levels of the unique β-subunit of CG) in the maternal circulation and urine, and is used as a diagnostic test for pregnancy. During the first trimester, CG in maternal blood increases almost exponentially, doubling every 2 to 3 days in the first 3 weeks of pregnancy. CG levels peak by week 10 to 12 of gestation, and thereafter maternal CG levels decrease to a plateau for the remainder of pregnancy.

As well as affecting the maternal CL, CG also acts on endometrial stromal cells via its interaction with LH/CG receptors to increase intracellular cAMP and sensitize the cells to progesterone. Progesterone in concert with cAMP induces expression of multiple genes in the endometrial stroma to induce decidualization. Thus, one of the earliest endocrine/paracrine signaling events in pregnancy is the production of CG that serves to maintain exposure of the uterus to progesterone, prevent menstruation, and promote decidualization.

Gonadotropin-Releasing Hormone

The human placenta also produces gonadotropin-releasing hormone (GnRH) that is identical to that produced by the hypothalamus.[39] Placental GnRH may regulate CG—as hypothalamic GnRH regulates LH production by the anterior pituitary—in an autocrine/paracrine manner. Levels of circulating GnRH correlate with CG and are highest during the first trimester, and *in vitro* studies show that GnRH stimulates CG production by placenta cells.[40,41]

Proopiomelanocortin-Derived Hormones

The human syncytiotrophoblast expresses proopiomelanocortin (POMC) that serves, as it does in the pituitary, as a precursor via its enzymatic cleavage, for the adrenocorticotropin (ACTH)-endorphin family of peptides: ACTH, β-lipotrophic hormone (β-LPH), α-melanocyte-stimulating hormone (α-MSH), and β-endorphin (β-EP).[42-45] Enkephalins and dynorphins are also produced by the human placenta.[46] The role of placental POMC-derived proteins is unclear. Marked changes in the maternal hypothalamic-pituitary-adrenal (HPA) axis occur during pregnancy, mediated mainly by placental ACTH and possibly other POMC derivatives. Placenta ACTH appears to function primarily in the maternal compartment, because is not sufficient to prevent adrenal hypoplasia in fetal hypopituitarism due to anencephaly. Production of α-MSH is thought to increase skin pigmentation and induce melasma, the mask of pregnancy.[47]

Corticotropin-Releasing Hormone

Production of CRH by the placental syncytiotrophoblast is unique to primates and, within the order, the pattern of production across pregnancy varies. In New World and Old World monkeys, placental CRH production peaks at midpregnancy.[48,49] In contrast, in great apes, including humans, placental CRH production progressively increases around the end of the first trimester and follows an exponential trajectory thereafter, peaking at term.[50] The contribution of the maternal hypothalamus to circulating CRH levels during pregnancy is negligible compared with that of the placenta, and as such maternal CRH levels directly reflect placental CRH synthetic activity. Great apes, including humans, also produce a circulating CRH-binding protein (CRHBP) that is thought to limit CRH bioavailability.[51] For most of pregnancy, CRHBP is present in excess of CRH in maternal blood.[52] A fall in CRHBP levels late in pregnancy is thought to increase bioactive CRH.[53,54] Actions of CRH are mediated by two CRH receptors, CRH-R1 and CRH-R2.[55,56]

CRH attributable to placental secretion can be detected in the maternal blood beginning at around the fifteenth week of gestation. Placental CRH production is modified by estrogen, progesterone, and nitric oxide, which are inhibitory, and by a range of neuropeptides, which are stimulatory.[57-60] Placental CRH production may also be linked to placental development and syncytiotrophoblast size and intrinsic and extrinsic stressors. Placental CRH levels in the maternal circulation increase almost exponentially after the fifteenth week of gestation, with the steepest incline occurring in the 5 weeks before parturition.[61] Late in pregnancy CRH levels in the maternal circulation are approximately 1000-fold higher than in nonpregnant women,[62] and placental CRH is principally responsible for the hyperactivity of the maternal HPA axis and the hypercortisolism that occurs at this time. Maternal hypercortisolism could contribute to the changes in maternal metabolism that promote nutrient transfer, notably by increasing production of glucose transporters in the placenta.[63] The fetus is protected from maternal hypercortisolemia by the 11β-hydroxysteroid dehydrogenase enzyme in the placenta, which converts cortisol to the inactive glucocorticoid cortisone.[64] Alterations in the expression of this enzyme may modulate fetal growth and fetal stress.

The control of placental CRH secretion is complex and includes factors that stimulate CRH expression and release by the hypothalamus.[57] A major difference between placental and hypothalamic CRH secretion is that, unlike the hypothalamus, production of CRH by placental cells is increased by cortisol.[65-67] *In vitro* studies suggest that a positive feedback loop exists whereby placental CRH stimulates ACTH production by the fetal pituitary that increases cortisol production by the fetal adrenal glands that then stimulates further placental CRH production. This positive feedback effect is thought to produce the exponential rise in placental CRH secretion and maternal CRH levels during human pregnancy.[52,68,69] CRH also preferentially increases the production of dehydroepiandrosterone sulfate (DHEA-S) by fetal and

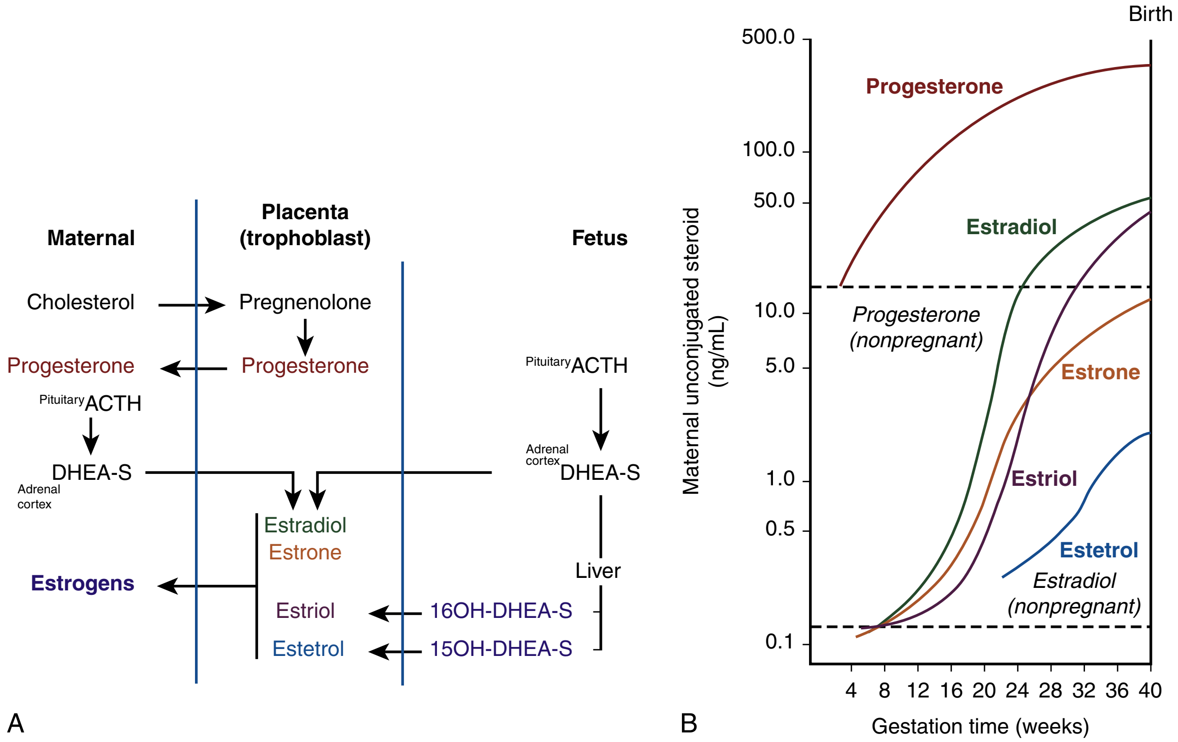


Fig. 129.3 A, Biosynthesis of progesterone and estrogens by the human fetal-placental unit. Progesterone is produced mainly from maternal cholesterol. Estrogens are biosynthesized from C19 androgen precursors (mainly DHEA) produced by the maternal and fetal adrenals. B, Levels of progesterone and estrogens (estrone, estradiol, estriol, and estetrol) in maternal blood during human pregnancy compared with average levels in normal cycling women. Estriol and estetrol are derived from 16OH- and 15OH-DHEA produced by the fetal liver.

adult adrenal cortical cells.[66,70-72] Fetal adrenal DHEA-S is converted to estrogens by the placental cells (see later). Interestingly, the marked rise in placental CRH during the final 10 weeks of pregnancy coincides with increased cortisol production by the fetal adrenals and correlates with increases placental estrogen production during the third trimester. This endocrine loop is thought to play a role in the hormonal control of parturition (discussed later; Fig. 129.3).

Placental Lactogen and Placental Growth Hormone

The human placenta expresses genes encoding members of the growth hormone (GH)/prolactin family. The proteins are encoded by a gene cluster on chromosome 17 that comprises *GH1*, *GH2*, *CSH1*, *CSH2*, and *CSHL1*.[73] *GH1* encodes pituitary GH and is expressed exclusively by pituitary somatotrophs. *GH2* is expressed only by the syncytiotrophoblast and extravillous cytotrophoblasts and encodes a placental GH (PGH) variant that differs from pituitary GH at 13 amino acids. *CSH1* and *CSH2* are duplicated genes exclusively expressed by the syncytiotrophoblast and cytotrophoblasts that encode placental lactogen (PL) that has homology with prolactin and shares some functions with GH. *CSHL1* is a pseudogene that does not produce a protein product. Thus, PL is derived from expression of *CSH1* and *CSH2*.[74]

PGH and PL exert somatogenic and lactogenic activity mediated by the GH and prolactin receptors, respectively. PL is predominantly a lactogen acting via the prolactin receptor, and has weak GH receptor–mediated activity.[74-76] Like pituitary GH, PGH and PL stimulate insulin-like growth factor-1 (IGF-1) production and have similar effects on growth and metabolism.[77]

PL is preferentially secreted into the maternal compartment, is detectable in the maternal circulation early in gestation and rises up to 6000 ng/mL late in the third trimester. Levels of PL in the fetus are 1/100 of the maternal concentration. PGH is only secreted into the maternal compartment, where it becomes detectable at the seventh week of gestation, peaks by the thirty-sixth week, and replaces pituitary GH as the major stimulus to maternal IGF-1 production.[78]

The roles of PGH and PL in the physiology of human pregnancy and parturition are not clearly defined. Fetuses with mutations in the GH/prolactin cluster that causes complete deficiency in PL and GH (PGH and pituitary GH) have severe growth restriction in an otherwise normal pregnancy. In contrast, fetal growth and development, and pregnancy in general, are normal in fetuses with mutations causing complete PL deficiency.[79] Pregnancies bearing mutations causing deficiency in PGH alone have not been identified. Taken together, the experiments of nature suggest that PGH is necessary for fetal growth and that PL is not necessary for normal pregnancy and may function in a redundant manner with PGH. As PGH is secreted into the maternal compartment, it is likely to affect fetal development via actions on maternal cells and tissues. A likely mechanism is that PGH (and PL) modulate maternal energy homeostasis to meet fetal needs, and that disruption of this endocrine axis starves the fetus of energy sources and nutrients, leading to growth restriction. PL and PGH also may affect maternal behavior to facilitate protection and nurturing of the neonate.[80,81]

Modulation of Maternal Energy Homeostasis

The developing fetus must signal its nutrient needs to the mother in step with its growth and development trajectory. This requires changes in maternal food intake and energy handling. In early pregnancy, maternal insulin secretion increases, with minimal alteration in insulin sensitivity. In contrast, insulin resistance is a characteristic feature of late pregnancy and changes maternal metabolism from fuel storage to fuel availability. In normal pregnancy a 50% reduction in insulin-mediated glucose disposal is observed in late gestation.[82,83] The placenta plays a key role in establishing appropriate maternal metabolic adaptations to sustain pregnancy and provide the fetus with nutrient and energy resources. Upon delivery of the placenta, maternal insulin sensitivity reverts to the nonpregnant state.

The genesis of insulin resistance during pregnancy has been attributed to placental hormones; although the exact contribution of specific placental hormones to maternal metabolism is not fully defined. PGH appears to be a major player. In mice, overexpression of human PGH to levels comparable with those in women during the third trimester induces significant peripheral insulin resistance.[84] PGH increases the expression of the p85α subunit of PI3 kinase in skeletal muscle. This acts as a dominant-negative competitor, preventing p85-p110 heterodimers from binding to insulin receptor substrate 1 (IRS-1), effectively producing a postreceptor defect in insulin signaling.[85] Levels of p85α are elevated in human pregnancy.[86]

The extent of insulin resistance during pregnancy does not correlate with maternal plasma levels of estradiol, progesterone, prolactin, or PL.[87] The strongest predictor of insulin sensitivity in pregnant women is instead tumor-necrosis factor-α (TNF-α), a potent proinflammatory cytokine that negatively correlates with insulin sensitivity in early and late pregnancy, even after adjusting for maternal fat mass and body mass index.[88] TNF-α is thought to induce insulin resistance by increasing phosphorylation of IRS-1, impairing downstream signaling from the insulin receptor. The placenta synthesizes and releases TNF-α predominantly into the maternal circulation, and plasma TNF-α concentrations rise by approximately 45% during late pregnancy.[87]

The human placenta also produces the adipokines leptin, adiponectin, resistin, ghrelin, and visfatin, each of which contributes to the control of maternal metabolism, and especially insulin sensitivity.[89] Leptin produced by the placenta is secreted mainly into the maternal circulation to achieve relatively high levels. Pregnancy is, however, associated with leptin resistance, with effects of leptin on eating behavior, satiety, and metabolic activity becoming uncoupled from its circulating levels. Nonetheless, it appears to promotes insulin resistance and fatty acid mobilization from adipose tissue, each of which would favor mobilization of energy sources for fetal use.[90] Leptin levels decline to normal nonpregnant values within 24 hours of delivery.[91] Adiponectin, resistin, ghrelin, and visfatin produced by the placenta and secreted into the maternal compartment likely have similar effects.

Growth Factors and Cytokines

Multiple growth factors and cytokines are produced by the placenta and affect fetal and maternal physiology. The IGFs are important regulators of fetal and placental growth. The IGF system is composed of three ligands (IGF-1, IGF-2, and insulin) that bind to four receptors: the type 1 IGF receptor (IGF-1R), the hybrid insulin/IGF-1R receptor, the mannose-6 phosphate/IGF-2 receptor, and the insulin receptor. Six IGF-binding proteins (IGFBPs) further modulate IGF activity. In particular, IGFBP-1 regulates the action of IGFs in pregnancy by limiting their availability for binding to cognate receptors. Proteases that cleave IGFBP-1 (e.g., pregnancy-associated plasma protein-A) therefore have promitogenic effects by increasing IGF bioavailability.[92] In addition, phosphorylation of IGFBP-1 increases its affinity for IGF-1. Levels of nonphosphorylated IGFBP-1 increase during pregnancy, and large amounts of nonphosphorylated IGFBP-1 in fetal serum have growth-promoting effects.[92] IGFBP-3 is the major carrier of IGF-1 and IGF-2 in the circulation to form a reservoir for IGFs. IGFBP-3 is regulated by GH and increases during pregnancy.[93]

Synthesis of maternal IGFs is modulated by PL and GH. PGH becomes the primary regulator of maternal IGF production as pregnancy advances. Although there is no placental transfer of IGFs, maternal IGFs affect fetal growth by influencing placental structure and function and diverting maternal nutrients to the fetus. Fetal pituitary GH does not control fetal growth, but growth is directly influenced by fetal IGFs and binding proteins, which in turn are regulated by fetal nutrition, insulin, and PL levels. Fetal tissue production of IGFs regulates placental growth.[94] IGF-1 and IGF-2 are expressed in the fetus, but IGF-1 predominates after birth under the stimulus of neonatal GH.[95] Fetal pituitary GH may have a small impact on fetal growth late in the third trimester, as evidenced by the lower birth size of neonates with hypopituitarism or GH insensitivity.[96,97] The placenta is exposed to both maternal and fetal IGFs and synthesizes many components of the IGF system that have roles in trophoblast growth, turnover, and placental exchange.[98]

The syncytiotrophoblast and cytotrophoblast cells express the epidermal growth factor (EGF) family of hormones that includes EGF, heparin-binding EGF-like growth factor, amphiregulin, betacellulin, epiregulin, and transforming growth factor-α.[99] The EGF family of growth factors play key paracrine roles in implantation and trophoblast invasion. Deficiency in EGF expression and/or signaling is associated with preeclampsia and fetal growth restriction.[100,101]

The fibroblast growth factor (FGF) family comprises 23 members (FGF-1 to FGF-23). The human placenta expresses FGF-2 and FGF-10, which are produced by cytotrophoblasts and connective tissue stroma of placental villi.[102,103] FGFs stimulate angiogenesis by affecting endothelial cell proliferation and migration via four high-affinity cell surface FGF receptors (FGFR-1 to -4). In the human placenta, FGFRs are expressed by the mononuclear phagocytic Hofbauer cells,[103,104] and FGF induces Hofbauer cells to produce multiple growth factors and cytokines involved in tissue repair. FGF via Hofbauer cells is thought to promote placental tissue repair and syncytiotrophoblast bud formation that is essential for placental development and growth.

Growth of the placenta with advancing gestation requires coordinated angiogenesis and remodeling of uterine vasculature, and especially the blood supply to the decidua basalis. These processes are affected by the vascular endothelial growth factor (VEGF) family of angiogenic growth factors. Cytotrophoblasts, the syncytiotrophoblast, and villous stromal cells produce VEGF (five isoforms are generated by alternative splicing from a single gene) and placenta growth factor (PLGF; four splice variants)[105-109] that promote formation of new blood vessels and increase vasodilatation and microvascular permeability to enhance tissue perfusion.[110,111] The actions of these factors are mediated by two receptors: FMS-like tyrosine kinase-1 (FLT1) and kinase insert domain receptor (KDR),[112-114] which are present on endothelial cells.[115-117] FLT1 also circulates as a soluble form, sFLT1, that binds and inhibits the angiogenic actions of circulating VEGFs and PLGFs.[118] The etiology of preeclampsia is thought to involve disruption of the VEGF/PLGF system, leading to adverse effects on maternal hemodynamics.[119] Preeclampsia is thought to derive from shallow invasion of trophoblasts into the maternal spiral artery endothelium and muscle walls during early placental development that decreases placental perfusion later in pregnancy.[120,121] Decreased perfusion causes hypoxia that increases VEGF and PLGF production by cytotrophoblasts and the syncytiotrophoblast that, via KDR on endothelial cells, increases release of sFLT1 that is thought to disrupt function of the maternal endothelium and cause hypertension and proteinuria, the typical signs of preeclampsia.[122]

KEY POINTS

- Peptide hormones released by the placenta promote maternal nutrient transfer to the fetus by increasing maternal insulin resistance, but this process predisposes the pregnant woman to gestational diabetes.
- Abnormalities of trophoblast invasion leads to increased release of vasotoxic factors that predispose to preeclampsia.

Steroid Hormones

In all viviparous species examined so far, progesterone is essential for the establishment and maintenance of pregnancy, and its withdrawal or disruption of its biological activity ends pregnancy by triggering parturition.[123] It is aptly referred to as the hormone of pregnancy. During the luteal phase of the menstrual cycle, progesterone prepares the uterus for pregnancy by affecting endometrial growth and function to create an intrauterine environment and endometrial phenotype conducive to embryo survival and implantation. Progesterone induces decidualization of the endometrial stroma that is critical for implantation. It also promotes growth and relaxation of the myometrium and maintains rigidity and closure of the cervix.

Progesterone needed to establish and maintain human pregnancy is initially provided by the maternal CL in response to CG. Removal of the ovaries during the first 6 to 10 weeks of pregnancy leads to abortion.[124] However, at around week 10 of gestation the syncytiotrophoblast of the placenta begin producing progesterone *de novo* from maternal cholesterol, and thereafter become a major source of progesterone. Initiation of the placenta as a source of progesterone is referred to as the luteal-placental shift. After this occurs, pregnancy persists even if the maternal ovaries are removed. Placental progesterone production increases for the remainder of pregnancy and late in third trimester is around 250 mg per day.[125] Consequently, levels of progesterone in maternal blood increase with advancing gestation and during the second and third trimesters are 10- to 20-fold higher than during the luteal phase.[126]

Actions of progesterone to maintain pregnancy are mediated by the nuclear PR isoforms, PR-A and PR-B, present in myometrial, cervical, and decidual cells. The PRs function mainly as ligand-activated transcription factors to affect target cell function by regulating expression of specific genes. That elimination of progesterone (e.g., by ovariectomy before the luteal/placental shift) or treatment with the PR antagonist mifepristone (also known as RU486)[127-129] end pregnancy by inducing parturition reflects the importance of progesterone/PR signaling for the maintenance of pregnancy. Through this signaling pathway, progesterone is thought to affect the expression of genes in myometrial, cervical, and decidual cells whose products promote a uterine phenotype—myometrial relaxation and quiescence, cervical rigidity and closure, and inflammatory quiescence at the chorion-decidual interface—that sustains pregnancy. In the myometrium, progesterone/PR signaling affects expression of genes encoding factors collectively referred to as contraction-associated proteins (CAPs). Progesterone inhibits expression of stimulatory CAPs and increases expression of relaxatory CAPs. Important stimulatory CAP genes include *GJA1,* which encodes connexin-43, a gap junction protein that forms connections between myometrial cells to synchronize contractions, *PTGS2,* which encodes prostaglandin (PG) synthase type 2 that increases local levels of the potent uterotonin, $PGF_{2\alpha}$, and *OXTR,* which encodes the oxytocin receptor to increase responsiveness of myometrial cells to oxytocin, another potent uterotonin. Progesterone/PR signaling also decreases responsiveness of uterine cells to estrogens by inhibiting expression of *ESR1*, which encodes estrogen receptor-α (ERα).[130] This indirectly promotes quiescence by inhibiting estrogen/ERα-induced expression of genes encoding stimulatory CAPs. Progesterone/PR signaling in uterine cells also exerts antiinflammatory effects by decreasing responsiveness to proinflammatory/prolabor stimuli.[131] This is important because proinflammatory stressors may induce parturition via tissue-level inflammation in the myometrium, cervix, and decidua.[132] Thus, progesterone acting via the PRs in myometrial, cervical, and decidual cells maintains pregnancy by promoting a quiescent and compliant uterine phenotype conducive to housing the conceptus, blocking parturition by inhibiting the expression of genes encoding stimulatory CAPs, suppressing responsiveness to estrogen/ERα, and decreasing responsiveness to proinflammatory/prolabor stimuli.

During the menstrual cycle, estrogens (mainly estradiol) stimulate regrowth of the functionalis layer of the endometrium and increase myometrial cell proliferation. Estradiol also increases uterine growth and blood flow to the endometrium and promotes milk gland development in the mammary ductal epithelium in preparation for lactation. During pregnancy estrogens (estrone, estradiol, estriol, and estetrol) increase uteroplacental blood flow via actions mediated by locally produced growth factors and angiogenic factors.[133] Thus, as with progesterone, estrogenic actions also favor pregnancy.

The syncytiotrophoblast of the human placental produces estradiol, estrone, estriol, and estetrol beginning at around week 10 to 12 of gestation.[134] Placental estrogen production increases almost exponentially for the remainder of pregnancy. The syncytiotrophoblast lacks the full set of steroidogenic enzymes to produce estrogens *de novo* from cholesterol but has the capacity to produce estrogens from C19 steroids. The main substrate for placental estrogen formation is DHEA-S produced by the maternal and fetal adrenals in response to ACTH and placental CRH. The syncytiotrophoblast has high levels of sulfatase activity, which converts DHEA-S to DHEA, 3β-hydroxysteroid dehydrogenase activity, which converts DHEA to androstenedione, aromatase activity, which converts androstenedione to estrone, and 17β-hydroxysteroid dehydrogenase activity, which converts estrone to estradiol. The fetal liver also metabolizes DHEA-S to 16-hydroxy-DHEA-S and 15-hydroxy-DHEA-S, which serve as substrates for the synthesis of estriol and estetrol, respectively. Thus, the large amounts of estrogens produced by the human placenta for most of pregnancy are dependent on a supply of C19 steroid, mainly from the fetal adrenal cortex (see Fig. 129.3).

The role of placental estrogens in the physiology of human pregnancy and parturition remains unclear. One conundrum is why large amounts of estrogens are produced by the placenta. Pregnancy and parturition appear normal in cases where placental estrogen production is significantly increased (10- to 100-fold) above the normal range. This is apparent in congenital adrenal hyperplasia, in which the fetal adrenal glands fail to produce cortisol, leading to increases ACTH release (due to loss of cortisol-mediated negative feedback) that increases the supply of DHEA-S to the placenta. Pregnancy and parturition also are normal in cases in which placental estrogen production is significantly decreased (e.g., placental aromatase deficiency; placental sulfatase deficiency; anencephaly; congenital adrenal lipoid hyperplasia). Apart from virilization in the mother and female fetus due to adrenal androgen excess, pregnancy appears to be unaffected by placental estrogen deficiency. This, however, does not exclude a role for estrogens in the control of human pregnancy and parturition, because estrogens from maternal sources (e.g., aromatization of adrenal androgens by adipocytes) are unaffected in cases of placental estrogen deficiency. In fact, in cases of placental aromatase deficiency, maternal estrogen levels were at a physiologically significant range (1–1.6 nmol/L) comparable with levels during the mid and luteal phases of the menstrual cycle.[135,136]

Conditions in which human pregnancy exists in the complete absence of estrogens have not been identified, suggesting that a minimal level of estrogenic drive is necessary for human pregnancy and parturition. In this case the remarkable capacity of the human placenta to convert C19 androgens to estrogens may be important to prevent virilization of the female fetus.

Under the influence of rising estrogen levels produced by the placenta, maternal pituitary lactotrophs increase prolactin production leading to concentrations in maternal serum of 200 to 300 ng/mL at term.[137] This is thought to contribute to mammary gland development in preparation for lactation.

KEY POINTS

- Progesterone is the primary hormone that maintains pregnancy by suppressing the expression of genes encoding contraction-associated proteins such as connexin 43 and cyclooxygenase 2, and by maintaining a closed rigid cervix.

PARTURITION AND BIRTH TIMING

Parturition, the process of birth, is controlled by hormonal signals that affect the contractile state of the uterine smooth muscle, as well as the mechanical integrity of the uterine cervix and the fetal membranes Clinical labor is defined as painful and rhythmic uterine contractions, softening and dilation of the cervix, and weakening/rupture of the fetal membranes. For parturition, the myometrium transforms from a quiescent, relaxed, and compliant phenotype to a highly contractile and excitable state to become the engine for birth, producing the rhythmic, coordinated, and forceful contractions of labor. At the same time, the uterine cervix, which is closed and rigid for most of pregnancy, softens and dilates in response to increased intrauterine pressure caused by the contractions of labor, to eventually expand enough to allow the fetus to be forced into and through the birth canal. In addition, tissue level inflammation in the decidua causes weakening of the adjacent amnion-chorion that leads to rupture of the fetal membranes to liberate the fetus from the amniotic sac. Labor is also associated with tissue-level inflammation that involves the infiltration of activated immune cells into the myometrium and cervix. Thus, the cell targets for the hormonal control of parturition are the smooth muscle cells in the myometrium, cervical stroma cells, and decidual stromal cells at the maternal-fetal interface, and, in each tissue, infiltrating and resident immune cells. In this arrangement, labor status is controlled by the balance between hormonal signals affecting each uterine tissue/cell type that promote labor (i.e., induce myometrial contraction, cervical softening, decidual inflammation/membrane weakening, and tissue-level inflammation) and those that block labor (i.e., promote myometrial relaxation and compliance, cervical rigidity and closure, and decidual immune quiescence and tolerance and prevent tissue-level inflammation). In addition, the hormonal control of human parturition involves factors that initiate the process at a specific stage of gestation that is usually linked with fetal-placental development. Although the hormonal control of this complex process involving multiple tissue and cell types and two organisms, the fetus and the mother, is not yet fully elucidated, several hormones are known to play pivotal roles, including: (1) progesterone, and specifically withdrawal of its propregnancy and labor-blocking effects; (2) estrogens and activation of prolabor estrogenic actions; (3) placental CRH and its effects on fetal adrenal steroidogenic and placental estrogen production; and (4) uterine tissue-level inflammation and net inflammatory stress on the gravid uterus leading to increased production of locally acting prolabor cytokines.

Evolutionary Perspective

Evolution is driven by the natural selection of traits based on the extent to which they benefit reproductive efficiency. The success of pregnancy significantly impacts reproductive efficiency, and therefore the hormonal control of pregnancy and parturition was likely subjected to intense selective pressure. This would explain much of the variation around the common theme of viviparity; e.g., singleton versus polytocus pregnancy; seasonal/cyclic versus continuous fertility; and altricial versus precocial neonates. The appropriate timing of birth is critical for the success of pregnancy, and therefore reproduction, in viviparous species. Preterm birth of an immature fetus imparts major risk for pregnancy failure, with the burden assumed by the immature neonate. Prolonged pregnancy and postterm parturition imparts risk for maternal and fetal survival due to the physical inability to birth an overgrown fetus. Pregnancy in general also confers survival risk on the female, and as such a maternal trigger for parturition likely evolved to end pregnancy to favor survival of the mother and her potential for future pregnancies. In this way, the interests of the female to pass her genes to the next generation via future pregnancies are balanced with those of the fetus, which include the interests of paternal genes, in the current pregnancy. Thus, it has been proposed that extant physiology is the result of evolutionary conflict between the interests borne in maternal and fetal/paternal genomes that significantly impacted the selection of hormonal control mechanisms to regulate the pregnancy and the timing of parturition.[138]

This concept is exemplified by the obstetric dilemma hypothesis that posits that hominid birth timing was impacted by the selective pressures caused by the evolution of bipedalism (obligate bipedal locomotion) and encephalization (increased complexity and relative size of the brain).[139] Obligate bipedalism in hominids began around 3 million years ago and involved changes in the size and shape of the pelvis to increase the efficiency of locomotion. Although this change decreased the size of the obstetric outlet, it likely had little impact on the success of pregnancy until brain and head size began increasing through the process of encephalization around 1 million years ago.[140] Human brain size has remained relatively constant for at least the last 100,000 years.[141] The obstetric dilemma hypothesis posits that increased fetal head size eventually confounded the birth process, requiring selection of traits that trigger parturition before the fetal head becomes larger than the pelvic outlet, a trait that appears to be unique to human lineages. The benefits of continued encephalization would eventually be limited by the reproductive cost of failed pregnancy (i.e., maternal and fetal loss) associated with the difficulty of birthing a large-headed fetus. In this context, encephalization likely exerted strong selective pressure on the hormonal control of birth timing such that gestation was shortened. Among primates, the problems imposed by cephalopelvic disproportion appear to be unique to humans. Comparative studies of neonatal brain size to body size across primate species, and the fact that human neonates are altricial, with a significant amount of brain growth occurring in the first 6 to 12 months after birth, suggest that evolution favored the shortening of human gestation.[142,143] This may have occurred by the selection of specific hormonal interactions to control the timing of parturition leading to birth of an altricial neonate. Successful encephalization also required sufficient energy to facilitate rapid growth of the fetal brain. This could have been achieved by the evolution of deep placentation and invasion of the uterine spiral arteries to maximize access to maternal blood and the secretion of placental hormones such as PL and PGH that induce maternal insulin resistance to increase the supply of glucose for the fetus. A metabolic basis for human altriciality has been proposed that argues that birth occurs when the metabolic demands of the fetus exceed maternal capacity to supply oxygen and nutrients.[144] Although this model challenges the obstetric dilemma, it is feasible that both hypotheses are valid.

Framing the hormonal control of human pregnancy and parturition in the context of the unique natural history puts into perspective extant pathophysiology. Natural selection operates at the population level to reconcile the benefits and costs of traits over multiple generations. With regard to the obstetric dilemma and metabolic constraint hypotheses, both solved the problem of favoring the benefits of encephalization in the context of obligate bipedalism with maternal survival and successful pregnancy. Extant pathophysiologies of pregnancy (e.g., preterm birth, preeclampsia), most of which are unique to humans occur at the individual level and may arise from inefficiencies in hormonal systems that were selected based on net effects at the population level and over multiple generations.

KEY POINTS

- In humans the evolution of bipedalism led to a narrowing of the pelvis. Concurrently the ecological niche occupied by humans favored increased brain size, which became limiting for delivery of the fetus. This resulted in shortening of gestation so that birth occurs before the fetal head becomes larger than the pelvic outlet, resulting in a relatively immature neonate compared to other primates.

Fetal Maturation

Under normal circumstances birth occurs when the fetus is sufficiently mature to survive as a neonate. This involves coordination between functional development of vital organ systems such as the lungs, the immune system, neuroendocrine control axes (especially the HPA axis), and the capacity to regulate energy use and storage and nutrient absorption and utilization. To this end, hormonal signals from the fetal-placental unit feed into the parturition control mechanism such that initiation of labor is synchronized with development. This is achieved by glucocorticoid (cortisol in the human) and is exemplified by parturition in sheep that is triggered by a surge of cortisol produced by the fetal adrenals 1 to 2 weeks before labor that stimulates functional maturation of fetal organs (especially the lungs) and triggers parturition by inducing progesterone withdrawal. Thus, parturition in sheep is controlled by activity of the fetal HPA axis and its production of cortisol, ensuring that birth and fetal maturation are linked. Promotion of fetal maturation by glucocorticoid appears to be a conserved trait across viviparous species, with variation in the source (e.g., fetal or maternal) and type of glucocorticoid (e.g., cortisol or corticosterone), the upstream hormonal axes that control its production, and the stringency with which it is linked to the parturition trigger mechanism. Synthetic glucocorticoid therapy is widely used to promote fetal organ (especially the lungs) maturation in women presenting with preterm labor and threatened preterm birth.

Unlike in sheep, cortisol does not appear to be a trigger for human parturition. Parturition in women occurs normally at term in pregnancies with fetuses that cannot produce cortisol due to anencephaly or the inability to synthesize cortisol due to deficiency in the 21-hydroxylase enzyme.[145,146] Also unlike sheep, administration of synthetic glucocorticoid (usually compounds that are not metabolized by the placenta) to women presenting with threatened preterm birth does not advance the parturition process. Nonetheless, as with other species, cortisol plays a key role in maturing the fetal lungs; hence the success of glucocorticoid therapy to promote fetal lung maturation in cases of threatened preterm birth. However, the fetal adrenals are not the exclusive source of cortisol for fetal maturation. In human pregnancy cortisol from the maternal compartment contributes to promoting fetal maturation. Evidence for this is that late in the third trimester maternal estriol levels are inversely associated with circadian maternal cortisol levels, suggesting that some maternal cortisol leaks through the placenta to suppress the fetal HPA.[147] Thus, the hormonal control of human parturition is not dependent on activity of the fetal HPA axis and may involve redundancy to ensure that parturition and fetal maturation occurs even if the fetal HPA axis, and especially production of fetal cortisol, is disrupted.

Progesterone Withdrawal

Clinical and animal data support the general consensus that, for most of pregnancy, progesterone actively blocks parturition, and loss of this activity induces changes in the uterine tissues to initiate the parturition cascade. The progesterone withdrawal trigger for parturition can occur by two mechanisms: (1) systemic or local decreases in bioavailable progesterone, and (2) desensitization of uterine target cells to the labor-blocking actions of progesterone. In animals in which the maternal CL is the exclusive source of progesterone throughout pregnancy (e.g., mouse, rat, rabbit), parturition is triggered by CL regression leading to systemic progesterone withdrawal. In other animals (e.g., sheep) in which progesterone is supplied by the placenta, parturition is induced by a prepartum surge in fetal adrenal cortisol that decreases placental progesterone secretion by inducing its conversion to estradiol.[148] The coordinated systemic progesterone withdrawal and increased estradiol transforms the uterine tissues to the labor state.

Human parturition occurs without a systemic progesterone withdrawal. Circulating levels of progesterone are high (compared with midluteal-phase levels) for most of pregnancy and persist during parturition, decreasing only after delivery of the placenta.[149] In this scenario progesterone withdrawal could be induced by loss or disruption of progesterone/PR signaling. Indeed, clinical and animal data show that the PR antagonist mifepristone promotes labor and, in most cases, induces the full parturition cascade at any stage of pregnancy, despite high circulating levels of progesterone.[150] Thus, human parturition is likely induced by functional rather than systemic progesterone withdrawal that involves physiological controlled changes in PR transcriptional activity in uterine cells.

Uterine cell responsiveness to progesterone is mediated by the relative abundance and transcriptional activities of the PR isoforms, PR-A and PR-B, that function as ligand activated transcription factors. *In vitro* studies have revealed possible molecular mechanisms by which progesterone/PR signaling in myometrial cells blocks labor. The data suggest that progesterone via PR-B decreases responsiveness to estrogens by inhibiting estrogen receptor expression,[130] decreases expression of genes encoding prolabor CAPs by increasing expression of transcriptional repressors such as ZEB1,[151] and inhibits responsiveness to proinflammatory/prolabor stimuli.[131,132] The PR-mediated functional progesterone withdrawal trigger for parturition centers on inhibiting these, and likely other, actions of progesterone/PR-B signaling in uterine cells. Several mechanisms have been proposed (Fig. 129.4). In most progesterone-responsive cells, PR-A inhibits the transcriptional activity of PR-B and this is considered to be a mechanism for target cell autoregulation of progesterone responsiveness, such that progesterone/PR transcriptional activity is inversely related to the PR-A/PR-B ratio.

KEY POINTS

- The onset of human labor is associated with a functional withdrawal of progesterone action caused by various mechanisms including increased transrepressive activity of PR-A and increased intracellular metabolism of progesterone.

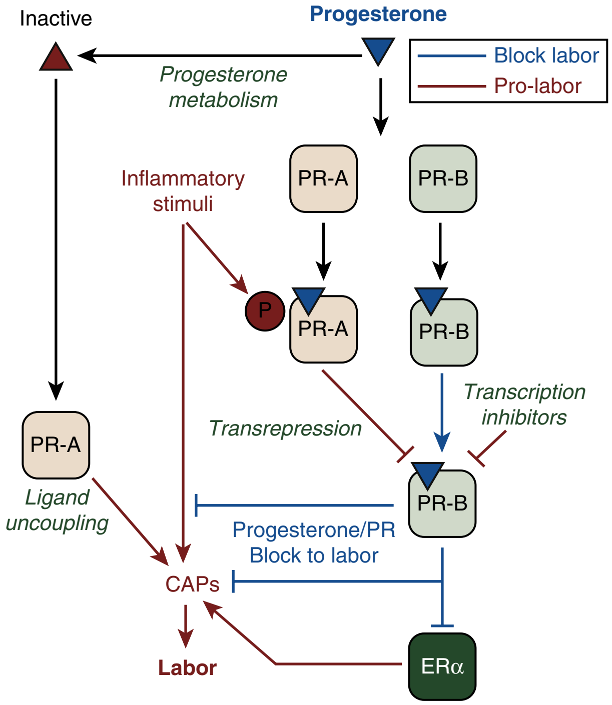


Fig. 129.4 Possible mechanisms for progesterone withdrawal in human pregnancy and mechanism for the progesterone block to labor. Upon ligand binding, progesterone receptors (PRs) dimerize and become activated as transcription factors. For most of pregnancy, progesterone acting via PR-B exerts anti-inflammatory activity to block inflammation-induced labor, inhibits contraction-associated protein gene expression, and inhibits responsiveness to estradiol by inhibiting expression of estrogen receptor-α. Functional progesterone withdrawal could be mediated by (1) increased PR-A transrepressive activity via phosphorylation of PR-A; (2) increased transcriptional repressor activity; (3) local metabolism of progesterone to an inactive form; that (4) generates unliganded PR-A, which increases expression of prolabor-associated genes in myometrial cells.

The onset of labor in women at term is associated with increased abundance of PR-A compared with PR-B in the myometrium,[152] and PR-A transrepressive activity in myometrial cells is induced by proinflammatory/prolabor stimuli via site-specific serine phosphorylation of PR-A.[153] This effect of PR-A may inhibit the capacity for progesterone/PR-B to block labor. PR-B transcriptional activity could also be inhibited by increased expression and activity of corepressors and reduced interaction of PR-B with DNA elements.[154-156] It is also possible that target cells metabolize progesterone to a form that fails to activate PRs causing a localized progesterone withdrawal.[157,158] *In vitro* studies in myometrial cells suggest that progesterone metabolism produces unliganded PR-A that directly induces expression of genes encoding stimulatory CAPs.[159] Thus, a significant body of data suggests redundancy in the mechanisms for the functional progesterone withdrawal trigger for human parturition. This is consistent with the importance of this process for the success of pregnancy.

Stress and Tissue-Level Inflammation

Human parturition is considered an inflammatory event that involves localized inflammation in the uterine tissues characterized by activation of resident tissue immune cells and the infiltration of activated immune cells.[160-162] Studies in mice and rhesus macaques show that the onset of labor is preceded by inflammation in the uterine tissues,[163,164] and parturition is induced by intrauterine administration of proinflammatory stimuli.[165-167] In women, preterm birth is closely associated with intrauterine infection, clinically silent upper genital tract infection, and bacterial vaginosis.[168-175] One hypothesis derived from those observations is that human parturition is triggered by inflammatory stress load impacting the uterine effector tissues (myometrium, decidua, and cervix). In this scenario, an inflammatory load threshold exists, above which various mechanisms for localized progesterone withdrawal are induced (Fig. 129.5). One mechanism by which progesterone/PR signaling maintains pregnancy is to inhibit the responsiveness of uterine cells to proinflammatory stimuli (i.e., an antiinflammatory effect). Studies in myometrial cells suggest that the inhibitory actions of PR-A are induced by site-specific serine phosphorylation in response to proinflammatory stimuli.[153] Based on those data it is proposed that for most of pregnancy progesterone/PR-B signaling in uterine cells blocks labor by exerting anti-inflammatory effects and that an inflammatory load threshold exists, above which specific protein kinase activity is activated that phosphorylates PR-A to induce its transrepressive effect on progesterone/PR-B anti-inflammatory activity. Inflammatory stimuli coupled with loss of progesterone/PR-B anti-inflammatory activity causes tissue-level inflammation and the production of cytokines, especially PGs, that transform the myometrium and cervix to the labor state. Inflammatory drivers of human parturition could be intrinsic (e.g., fetal maturation, uterine wall distention, maternal nutritional stress) or extrinsic (e.g., intrauterine infection; maternal environmental, and psychosocial stress) and involve numerous local and endocrine hormones that converge on the activation of PR-A transrepressive activity in uterine cells to cause progesterone withdrawal.

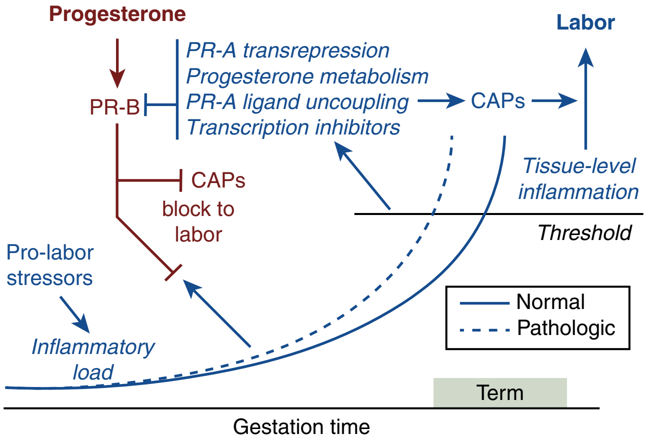


Fig. 129.5 Threshold-limited mechanism for inflammation-induced parturition by induction of progesterone withdrawal in response to net inflammatory load impacting progesterone/PR target cells in the gravid uterus. According to this model the timing of parturition is determined by the inflammatory load trajectory and the threshold for inflammation-induced progesterone withdrawal.

Progestin Therapy to Prevent Preterm Birth

The use of progesterone supplementation as a strategy to prevent preterm labor and preterm birth was proposed soon after progesterone was identified and isolated in the early 1930s. The core assumption is that risk for preterm birth derives from a deficiency of progesterone, and that this can be corrected by progesterone supplementation. The fact that circulating levels of endogenous progesterone in women throughout pregnancy are high and remain elevated even during labor and delivery (this was not known in the 1930s) challenges the biological plausibility for progesterone supplementation.

Nonetheless, in the 1950 (when sufficient stocks of purified progesterone became available for clinical studies) trials of progesterone were conducted. Administration of progesterone as a high-concentration (~1 g) intravenous or intraamniotic bolus in laboring women had no tocolytic activity.[176,177] Subsequent trials focused on prophylactic progesterone supplementation to reduce risk for preterm birth. During the 1960s to 1980s, small clinical trials of progesterone and 17-hydroxyprogesterone-caproate (17HPC) prophylaxis indicated a beneficial effect in reducing risk for preterm birth.[178,179] This led to larger trials in the 1990s that also indicated that progestin supplementation reduces the risk for preterm birth; but only in women with elevated risk for preterm birth.[180,181] More recent clinical trials,[182,183] however, have produced negative outcomes, and consequently progestin prophylaxis to prevent preterm birth remains controversial.[178-180] However, subsequent larger trials[182,183] however, have produced negative outcomes, and consequently progestin prophylaxis to prevent preterm birth remains controversial.

Estrogens

Estrogenic actions on the gravid uterus are generally thought to promote labor by increasing expression of genes encoding prolabor CAPs (e.g., *GJA1*, *OXTR*). In most species maternal estrogen levels increase during the prelude to parturition. Human pregnancy is characterized by a highly estrogenic state beginning at around the end of the first trimester, when the fetal adrenal-placenta steroidogenic axis produces large amounts of estrone, estradiol, estriol, and estetrol (see earlier). Thus, for most of human pregnancy the uterine tissues are exposed to high levels of estrogen, yet effects to promote labor appear to be blocked. One possible reason for this is that progesterone inhibits expression of *ESR1* (which encodes ERα) in uterine cells, thus suppressing the capacity for estrogens to promote labor. This effect decreases upon progesterone withdrawal, which releases the inhibition of *ESR1* expression allowing ERα to mediate prolabor estrogenic effects. Interestingly, abnormalities that compromise placental estrogen production (e.g., placental aromatase deficiency) do not affect the timing of birth, suggesting that the high estrogen levels are not necessary for normal parturition.[135,136,184] In all cases of placental estrogen deficiency, however, the gravid uterus was exposed to estrogens produced in the maternal compartment. Experiments of nature in which human pregnancy exists in the complete absence of estrogens have not been reported. Thus, it appears that relatively little estrogen is required for normal human parturition. Parturition is also normal in pregnancy with abnormally high estrogen levels due to congenital adrenal hyperplasia.[145] These observations suggest that estrogen effects in human parturition are not dependent on circulating levels, provided a minimal amount exists, but are instead controlled by increased uterine tissue estrogen responsiveness (i.e., functional estrogen activation) that may be induced by progesterone withdrawal.

KEY POINTS

- Increased estrogenic activity at term drives increased expression of genes encoding contraction-associated proteins. The increased estrogenic activity is mediated by alterations in estrogen receptor–mediated signaling pathways.

Integrative Role of Placental Corticotropin-Releasing Hormone

The timing of human parturition is associated with the rate of increase in circulating CRH production during the course of pregnancy[52,68] (Fig. 129.6). Levels of placental CRH in maternal blood follow an exponential function. The rate of exponential increase in circulating CRH beginning at around the end of the first trimester correlates with birth timing. The rate of increase in maternal CRH during the second and third trimesters is greater in women destined to deliver before term compared with women who deliver at term, and lowest in women who deliver postterm. Small changes in the exponential function early in pregnancy lead to large differences between different women later in pregnancy that are associated with pre- and postterm birth. Given the large variations among individuals, it is likely that the rate of rise of maternal concentrations of CRH reflects physiologic changes that lead to labor, and as such is a predictor of birth timing, and that this is determined during the first trimester.[52,185] The exponential increase in maternal plasma CRH also reflects the increasing volume of the syncytiotrophoblast and the exponential increase that occurs in syncytiotrophoblast nuclei with advancing gestation that is in part controlled by syncytin-1–induced fusion of cytotrophoblasts with the syncytiotrophoblast. Thus, the rise in maternal CRH concentrations may also be viewed as a developmental process within the placenta leading to delivery. In this context, placental CRH production may be part of a clock mechanism that links placental development to gestational length,[52] all driven by an endogenous retrovirus.

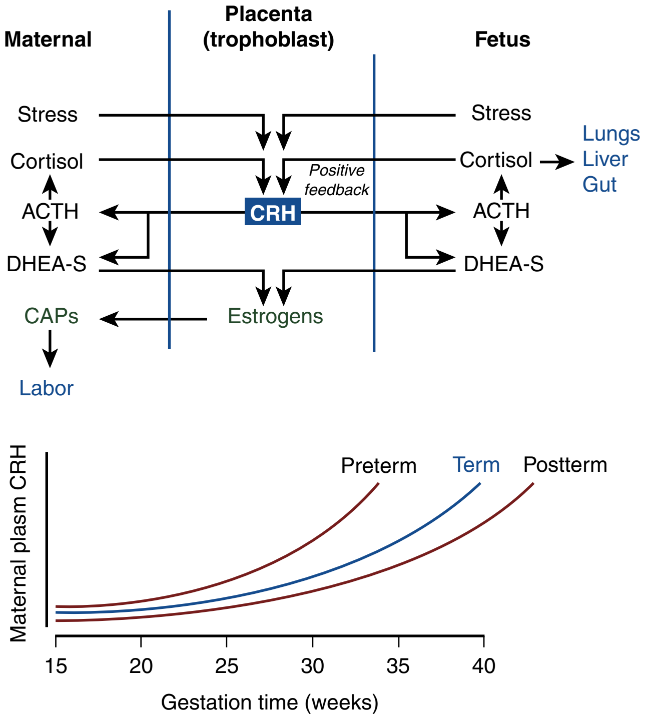


Fig. 129.6 Top: Central role of placental corticotropin-releasing hormone (*CRH*) in controlling fetal organ system maturation via fetal cortisol and parturition via estrogen production by the fetal-placental unit. Placental CRH stimulates adrenal cortical steroidogenic activity either directly (for DHEA) or via adrenocorticotropin. Cortisol exerts positive feedback on placental CRH production and stimulates functional maturation of fetal organ systems. Bottom: Maternal blood CRH levels in human pregnancy increase exponentially after midgestation. The rate of increase is associated with the timing of parturition.

Placenta CRH may influence birth timing by affecting maternal and fetal physiology. Although it is secreted predominately into the maternal blood, physiologically relevant amounts are also released into the fetal circulation. In the mother, CRH receptors are present in the pituitary and the myometrium. In the fetus, CRH receptors are present in the pituitary and in the adrenal cortex and lung.[186] In the mother, placental CRH increases ACTH production by the pituitary, which in turn increases cortisol production by the adrenals. Consequently, maternal cortisol levels increase during the course of pregnancy in concert with CRH levels, although the effect of CRH is moderated by the circulating CRHBP and desensitization of CRH receptors by continuous exposure to high concentrations of CRH.[187,188] Nonetheless, increased levels of CRH and ACTH promote cortisol and DHEA-S production by the maternal adrenals, with DHEA-S serving as a maternal-derived substrate for placental estrogen synthesis. CRH may also directly affect contractility of the myometrium. For most of pregnancy, myometrial cells express mainly CRHR1, a Gs-coupled receptor linked to adenylate cyclase that, in response to CRH, increases formation of cAMP, which promotes relaxation of the myometrium.[189]

Placental CRH is also released into the fetus, and although circulating fetal CRH concentrations are lower than in the maternal circulation, they still rise with advancing gestation. In the fetus, CRH receptors are present on the pituitary[190] and on the cells that form the fetal zone of the adrenal cortex.[191] Stimulation of the fetal pituitary by CRH increases ACTH production and consequently increases the synthesis of cortisol and DHEA-S by the fetal adrenal glands. One possible hormonal cascade initiated by placental CRH is that rising fetal cortisol concentrations induce maturation of the fetal lung. This in turn is associated with increased production of surfactant protein A and phospholipids that have proinflammatory actions in the myometrium and fetal membranes that increases production of prostaglandins in those tissues that act locally to induce the labor phenotype.[192] CRH may also directly stimulate fetal lung development and surfactant phospholipid synthesis.[186] CRH also directly promotes the production of DHEA-S by fetal zone cells in the fetal adrenals, therefore increasing placental estrogen production.[191]

In summary, positive feedback systems in the mother and fetus drive an exponential increase in placental CRH production as pregnancy advances. Increasing placental CRH affects fetal and maternal adrenal steroidogenesis to increase cortisol and estrogen (via increasing adrenal DHEA-S) production. Cortisol promotes fetal organ maturation, especially the lungs (see Fig. 129.6). Thus, placental CRH may be a link between placental development and its response to physiologic stressors that influences multiple hormonal pathways, each contributing to the coordination of fetal maturation with the timing of parturition.

KEY POINTS

- Placental CRH production is linked to maturational processes that lead to parturition. Increased maternal CRH concentrations are linked to preterm delivery while low levels are associated with post-dates delivery. Placental CRH production is linked to fetal adrenal androgen production and increased placental estrogen synthesis.

CONCLUSION/FUTURE DIRECTIONS

Despite meeting the significant metabolic and physical demands of the conceptus, females survive the stress of pregnancy and usually have multiple pregnancies. This finely tuned biological state is critical for species propagation, and as such extant physiology is the product of millions of years of natural experimentation and selection. The leading edge of this process is hormonal signaling between the fetal/placental and maternal compartments. Maternal adaptations to establish and maintain pregnancy, and then end pregnancy at a specific stage of gestation by the process of parturition, are mediated by multiple fetal and maternal hormones acting via endocrine and paracrine modes. We have described some of the key hormones controlling these processes. Continued research in this field will eventually reveal the specific roles of hormones and hormonal systems in the physiology and pathophysiology (e.g., pre- and postterm birth; preeclampsia; gestational diabetes) of human pregnancy and parturition and provide clinical tools (e.g., hormone levels as risk biomarkers) and therapeutic targets to promote pregnancy wellness and prevent adverse pregnancy outcomes.

REFERENCES

1. Nikas G. Endometrial receptivity: changes in cell-surface morphology. *Semin Reprod Med*. 2000;18:229–235.
2. Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res*. 2007;17:42–49.
3. Gellersen B, Brosens IA, Brosens JJ. Decidualization of the human endometrium: mechanisms, functions, and clinical perspectives. *Semin Reprod Med*. 2007;25:445–453.
4. Herington JL, Bany BM. Do molecular signals from the conceptus influence endometrium decidualization in rodents? *J Exp Zool B Mol Dev Evol*. 2009;312:797–816.
5. Gellersen B, Brosens JJ. Cyclic decidualization of the human endometrium in reproductive health and failure. *Endocr Rev*. 2014;35:851–905.
6. Schumacher A, Zenclussen AC. Human chorionic gonadotropin-mediated immune responses that facilitate embryo implantation and placentation. *Front Immunol*. 2019;10:2896.
7. Gridelet V, Perrier d'Hauterive S, Polese B, et al. Human chorionic gonadotrophin: new pleiotropic functions for an "old" hormone during pregnancy. *Front Immunol*. 2020;11:343.
8. Guzeloglu-Kayisli O, Kayisli UA, Taylor HS. The role of growth factors and cytokines during implantation: endocrine and paracrine interactions. *Semin Reprod Med*. 2009;27:62–79.
9. Lissauer D, Eldershaw SA, Inman CF, et al. Progesterone promotes maternal-fetal tolerance by reducing human maternal T-cell polyfunctionality and inducing a specific cytokine profile. *Eur J Immunol*. 2015;45:2858–2872.
10. Fisher SJ. Why is placentation abnormal in preeclampsia? *Am J Obstet Gynecol*. 2015;213:S115–S122.
11. Kim SM, Kim JS. A review of mechanisms of implantation. *Dev Reprod*. 2017;21:351–359.
12. Newbern D, Freemark M. Placental hormones and the control of maternal metabolism and fetal growth. *Curr Opin Endocrinol Diabetes Obes*. 2011;18:409–416.
13. Costa MA. The endocrine function of human placenta: an overview. *Reprod Biomed Online*. 2016;32:14–43.
14. Medawar PB. Some immunological and endocrinological problems raised by the evolution of viviparity in vertebrates. *Symp Soc Exp Biol*. 1953;7:320–338.
15. Apps R, Murphy SP, Fernando R, et al. Human leucocyte antigen (HLA) expression of primary trophoblast cells and placental cell lines, determined using single antigen beads to characterize allotype specificities of anti-HLA antibodies. *Immunology*. 2009;127:26–39.
16. Leong AS, Norman JE, Smith R. Vascular and myometrial changes in the human uterus at term. *Reprod Sci*. 2008;15:59–65.
17. Thomas LM. Current perspectives on natural killer cell education and tolerance: emerging roles for inhibitory receptors. *ImmunoTargets Ther*. 2015;4:45–53.
18. Papuchova H, Meissner TB, Li Q, et al. The dual role of HLA-C in tolerance and immunity at the maternal-fetal interface. *Front Immunol*. 2019;10:2730.
19. Fisher SJ, Damsky CH. Human cytotrophoblast invasion. *Semin Cell Biol*. 1993;4:183–188.
20. Rouas-Freiss N, Goncalves RM, Menier C, et al. Direct evidence to support the role of HLA-G in protecting the fetus from maternal uterine natural killer cytolysis. *Proc Natl Acad Sci U S A*. 1997;94:11520–11525.
21. van Beekhuizen HJ, Joosten I, Lotgering FK, et al. Natural killer cells and HLA-G expression in the basal decidua of human placenta adhesiva. *Placenta*. 2010;31:1078–1084.

22. Liu KJ, Wang CJ, Chang CJ, et al. Surface expression of HLA-G is involved in mediating immunomodulatory effects of placenta-derived multipotent cells (PDMCs) towards natural killer lymphocytes. *Cell Transplant*. 2011;20:1721–1730.
23. Munn DH, Zhou M, Attwood JT, et al. Prevention of allogeneic fetal rejection by tryptophan catabolism. *Science*. 1998;281:1191–1193.
24. Mellor AL, Sivakumar J, Chandler P, et al. Prevention of T cell-driven complement activation and inflammation by tryptophan catabolism during pregnancy. *Nat Immunol*. 2001;2:64–68.
25. Coumans B, Thellin O, Zorzi W, et al. Lymphoid cell apoptosis induced by trophoblastic cells: a model of active foeto-placental tolerance. *J Immunol Methods*. 1999;224:185–196.
26. Makrigiannakis A, Zoumakis E, Kalantaridou S, et al. Corticotropin-releasing hormone promotes blastocyst implantation and early maternal tolerance. *Nat Immunol*. 2001;2:1018–1024.
27. Kalantaridou SN, Zoumakis E, Makrigiannakis A, et al. The role of corticotropin-releasing hormone in blastocyst implantation and early fetal immunotolerance. *Horm Metab Res*. 2007;39:474–477.
28. van den Heuvel MJ, Chantakru S, Xuemei X, et al. Trafficking of circulating pro-NK cells to the decidualizing uterus: regulatory mechanisms in the mouse and human. *Immunol Invest*. 2005;34:273–293.
29. Carlino C, Stabile H, Morrone S, et al. Recruitment of circulating NK cells through decidual tissues: a possible mechanism controlling NK cell accumulation in the uterus during early pregnancy. *Blood*. 2008;111:3108–3115.
30. Santoni A, Carlino C, Stabile H, et al. Mechanisms underlying recruitment and accumulation of decidual NK cells in uterus during pregnancy. *Am J Reprod Immunol*. 2008;59:417–424.
31. Benassayag C, Souski I, Mignot TM, et al. Corticosteroid-binding globulin status at the fetomaternal interface during human term pregnancy. *Biol Reprod*. 2001;64:812–821.
32. Mao G, Wang J, Kang Y, et al. Progesterone increases systemic and local uterine proportions of CD4+CD25+ Treg cells during midterm pregnancy in mice. *Endocrinology*. 2010;151:5477–5488.
33. Rote NS, Chakrabarti S, Stetzer BP. The role of human endogenous retroviruses in trophoblast differentiation and placental development. *Placenta*. 2004;25:673–683.
34. Urnovitz HB, Murphy WH. Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clin Microbiol Rev*. 1996;9:72–99.
35. Tolosa JM, Schjenken JE, Clifton VL, et al. The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. *Placenta*. 2012;33:933–941.
36. Tolosa JM, Parsons KS, Hansbro PM, et al. The placental protein syncytin-1 impairs antiviral responses and exaggerates inflammatory responses to influenza. *PLoS One*. 2015;10:e0118629.
37. Waddell BJ. The placenta as hypothalamus and pituitary: possible impact on maternal and fetal adrenal function. *Reprod Fertil Dev*. 1993;5:479–497.
38. Lin L, Xu B, Rote NS. The cellular mechanism by which the human endogenous retrovirus ERV-3 env gene affects proliferation and differentiation in a human placental trophoblast model, BeWo, Placenta. 2000;21:73–78.
39. Gibbons JM, Mitnick M, Chieffo V. In vitro biosynthesis of TSH- and LH-releasing factors by the human placenta. *Am J Obstet Gynecol*. 1975;121:127–131.
40. Siler-Khodr TM, Khodr GS, Valenzuela G. Immunoreactive gonadotropin-releasing hormone level in maternal circulation throughout pregnancy. *Am J Obstet Gynecol*. 1984;150:376–379.
41. Sasaki K, Norwitz ER. Gonadotropin-releasing hormone/gonadotropin-releasing hormone receptor signaling in the placenta. *Curr Opin Endocrinol Diabetes Obes*. 2011;18:401–408.
42. Liotta AS, Houghten R, Krieger DT. Identification of a beta-endorphin-like peptide in cultured human placental cells. *Nature*. 1982;295:593–595.
43. Chen CL, Chang CC, Krieger DT, et al. Expression and regulation of proopiomelanocortin-like gene in the ovary and placenta: comparison with the testis. *Endocrinology*. 1986;118:2382–2389.
44. Raffin-Sanson ML, Massias JF, Ankotche A, et al. High precursor level in maternal blood results from the alternate mode of proopiomelanocortin processing in human placenta. *Clin Endocrinol*. 1999;50:85–94.
45. Grigorakis SI, Anastasiou E, Dai K, et al. Three mRNA transcripts of the proopiomelanocortin gene in human placenta at term. *Eur J Endocrinol*. 2000;142:533–536.
46. Petraglia F, Florio P, Nappi C, et al. Peptide signaling in human placenta and membranes: autocrine, paracrine, and endocrine mechanisms. *Endocr Rev*. 1996;17:156–186.
47. Clark D, Thody AJ, Shuster S, et al. Immunoreactive alpha-MSH in human plasma in pregnancy. *Nature*. 1978;273:163–164.
48. Goland RS, Wardlaw SL, Fortman JD, et al. Plasma corticotropin-releasing factor concentrations in the baboon during pregnancy. *Endocrinology*. 1992;131:1782–1786.
49. Smith R, Chan EC, Bowman ME, et al. Corticotropin-releasing hormone in baboon pregnancy. *J Clin Endocrinol Metab*. 1993;76:1063–1068.
50. Smith R, Wickings EJ, Bowman ME, et al. Corticotropin-releasing hormone in chimpanzee and gorilla pregnancies. *J Clin Endocrinol Metab*. 1999;84:2820–2825.
51. Bowman ME, Lopata A, Jaffe RB, et al. Corticotropin-releasing hormone-binding protein in primates. *Am J Primatol*. 2001;53:123–130.
52. McLean M, Bisits A, Davies J, et al. A placental clock controlling the length of human pregnancy. *Nat Med*. 1995;1:460–463.
53. Linton EA, Behan DP, Saphier PW, et al. Corticotropin-releasing hormone (CRH)-binding protein: reduction in the adrenocorticotropin-releasing activity of placental but not hypothalamic CRH. *J Clin Endocrinol Metab*. 1990;70:1574–1580.
54. Linton EA, Perkins AV, Woods RJ, et al. Corticotropin releasing hormone-binding protein (CRH-BP): plasma levels decrease during the third trimester of normal human pregnancy. *J Clin Endocrinol Metab*. 1993;76:260–262.
55. Chen R, Lewis KA, Perrin MH, et al. Expression cloning of a human corticotropin-releasing-factor receptor. *Proc Natl Acad Sci U S A*. 1993;90:8967–8971.
56. Perrin M, Donaldson C, Chen R, et al. Identification of a second corticotropin-releasing factor receptor gene and characterization of a cDNA expressed in heart. *Proc Natl Acad Sci U S A*. 1995;92:2969–2973.
57. Petraglia F, Sutton S, Vale W. Neurotransmitters and peptides modulate the release of immunoreactive corticotropin-releasing factor from cultured human placental cells. *Am J Obstet Gynecol*. 1989;160:247–251.
58. Ni X, Chan EC, Fitter JT, et al. Nitric oxide inhibits corticotropin-releasing hormone exocytosis but not synthesis by cultured human trophoblasts. *J Clin Endocrinol Metab*. 1997;82:4171–4175.
59. Ni X, Nicholson RC, King BR, et al. Estrogen represses whereas the estrogen-antagonist ICI 182780 stimulates placental CRH gene expression. *J Clin Endocrinol Metab*. 2002;87:3774–3778.
60. Ni X, Hou Y, Yang R, et al. Progesterone receptors A and B differentially modulate corticotropin-releasing hormone gene expression through a cAMP regulatory element. *Cell Mol Life Sci*. 2004;61:1114–1122.
61. Frim DM, Emanuel RL, Robinson BG, et al. Characterization and gestational regulation of corticotropin-releasing hormone messenger RNA in human placenta. *J Clin Invest*. 1988;82:287–292.
62. Goland RS, Wardlaw SL, Stark RI, et al. High levels of corticotropin-releasing hormone immunoactivity in maternal and fetal plasma during pregnancy. *J Clin Endocrinol Metab*. 1986;63:1199–1203.
63. Hahn T, Barth S, Graf R, et al. Placental glucose transporter expression is regulated by glucocorticoids. *J Clin Endocrinol Metab*. 1999;84:1445–1452.
64. Benediktsson R, Calder AA, Edwards CR, et al. Placental 11 beta-hydroxysteroid dehydrogenase: a key regulator of fetal glucocorticoid exposure. *Clin Endocrinol*. 1997;46:161–166.
65. Robinson BG, Emanuel RL, Frim DM, et al. Glucocorticoid stimulates expression of corticotropin-releasing hormone gene in human placenta. *Proc Natl Acad Sci U S A*. 1988;85:5244–5248.
66. Korebrits C, Yu DH, Ramirez MM, et al. Antenatal glucocorticoid administration increases corticotrophin-releasing hormone in maternal plasma. *Br J Obstet Gynaecol*. 1998;105:556–561.

67. Marinoni E, Korebrits C, Di Iorio R, et al. Effect of betamethasone in vivo on placental corticotropin-releasing hormone in human pregnancy. *Am J Obstet Gynecol*. 1998;178:770–778.
68. Leung T, Chung T, Madsen G, et al. Rate of rise in maternal plasma corticotrophin-releasing hormone and its relation to gestational length. *BJOG*. 2001;108:527–532. Int J Obstet Gynaecol.
69. McGrath S, McLean M, Smith D, et al. Maternal plasma corticotropin-releasing hormone trajectories vary depending on the etiology of preterm delivery. *Am J Obstet Gynecol*. 2002;186:257–260.
70. Smith R, Mesiano S, Chan EC, et al. Corticotropin-releasing hormone directly and preferentially stimulates dehydroepiandrosterone sulfate secretion by human fetal adrenal cortical cells. *J Clin Endocrinol Metab*. 1998;83:2916–2920.
71. Chakravorty A, Mesiano S, Jaffe RB. Corticotropin-releasing hormone stimulates P450 17alpha-hydroxylase/17,20-lyase in human fetal adrenal cells via protein kinase C. *J Clin Endocrinol Metab*. 1999;84:3732–3738.
72. Ibanez L, Potau N, Marcos MV, et al. Corticotropin-releasing hormone as adrenal androgen secretagogue. *Pediatr Res*. 1999;46:351–353.
73. Anthony RV, Limesand SW, Fanning MD, et al. Placental lactogen and growth hormone. In: Bazer FW, ed. *The Endocrinology of Pregnancy*. Totowa: Humana Press Inc.; 1998:461–490.
74. Handwerger S, Freemark M. Role of placental lactogen and prolactin in human pregnancy. *Adv Exp Med Biol*. 1987;219:399–420.
75. Handwerger S. Clinical counterpoint: the physiology of placental lactogen in human pregnancy. *Endocr Rev*. 1991;12:329–336.
76. Lowman HB, Cunningham BC, Wells JA. Mutational analysis and protein engineering of receptor-binding determinants in human placental lactogen. *J Biol Chem*. 1991;266:10982–10988.
77. Fleenor D, Oden J, Kelly PA, et al. Roles of the lactogens and somatogens in perinatal and postnatal metabolism and growth: studies of a novel mouse model combining lactogen resistance and growth hormone deficiency. *Endocrinology*. 2005;146:103–112.
78. Wu Z, Bidlingmaier M, Friess SC, et al. A new nonisotopic, highly sensitive assay for the measurement of human placental growth hormone: development and clinical implications. *J Clin Endocrinol Metab*. 2003;88:804–811.
79. Rygaard K, Revol A, Esquivel-Escobedo D, et al. Absence of human placental lactogen and placental growth hormone (HGH-V) during pregnancy: PCR analysis of the deletion. *Hum Genet*. 1998;102:87–92.
80. Saltzman W, Maestripieri D. The neuroendocrinology of primate maternal behavior. *Prog Neuro-Psychopharmacol Biol Psychiatry*. 2011;35:1192–1204.
81. Bridges RS. Neuroendocrine regulation of maternal behavior. *Front Neuroendocrinol*. 2015;36:178–196.
82. Catalano PM, Tyzbir ED, Roman NM, et al. Longitudinal changes in insulin release and insulin resistance in nonobese pregnant women. *Am J Obstet Gynecol*. 1991;165:1667–1672.
83. Catalano PM, Huston L, Amini SB, et al. Longitudinal changes in glucose metabolism during pregnancy in obese women with normal glucose tolerance and gestational diabetes mellitus. *Am J Obstet Gynecol*. 1999;180:903–916.
84. Barbour LA, Shao J, Qiao L, et al. Human placental growth hormone causes severe insulin resistance in transgenic mice. *Am J Obstet Gynecol*. 2002;186:512–517.
85. Barbour LA, Mizanoor Rahman S, Gurevich I, et al. Increased P85alpha is a potent negative regulator of skeletal muscle insulin signaling and induces in vivo insulin resistance associated with growth hormone excess. *J Biol Chem*. 2005;280:37489–37494.
86. Barbour LA, McCurdy CE, Hernandez TL, et al. Cellular mechanisms for insulin resistance in normal pregnancy and gestational diabetes. *Diabetes Care*. 2007;30(suppl 2):S112–S119.
87. Kirwan JP, Hauguel-De Mouzon S, Lepercq J, et al. TNF-alpha is a predictor of insulin resistance in human pregnancy. *Diabetes*. 2002;51:2207–2213.
88. Wang SL, Liu PQ, Ding Y, et al. [Maternal serum tumor necrosis factor-alpha concentration and correlation with insulin resistance in gestational diabetes]. *Zhonghua Fu Chan Ke Za Zhi*. 2004;39:737–740.
89. Briana DD, Malamitsi-Puchner A. Reviews: adipocytokines in normal and complicated pregnancies. *Reprod Sci*. 2009;16:921–937.
90. Mukherjea R, Castonguay TW, Douglass LW, et al. Elevated leptin concentrations in pregnancy and lactation: possible role as a modulator of substrate utilization. *Life Sci*. 1999;65:1183–1193.
91. Hardie L, Trayhurn P, Abramovich D, et al. Circulating leptin in women: a longitudinal study in the menstrual cycle and during pregnancy. *Clin Endocrinol (Oxf)*. 1997;47:101–106.
92. Westwood M. Role of insulin-like growth factor binding protein 1 in human pregnancy. *Rev Reprod*. 1999;4:160–167.
93. Wang HS, Cheng BJ, Soong YK. Serum levels of insulin-like growth factor-binding protein-3 in normal pregnancy. *Eur J Obstet Gynecol Reprod Biol*. 1995;61:157–160.
94. Kavitha JV, Rosario FJ, Nijland MJ, et al. Down-regulation of placental mTOR, insulin/IGF-I signaling, and nutrient transporters in response to maternal nutrient restriction in the baboon. *FASEB J*. 2014;28:1294–1305.
95. Randhawa R, Cohen P. The role of the insulin-like growth factor system in prenatal growth. *Mol Genet Metab*. 2005;86:84–90.
96. Gluckman PD, Gunn AJ, Wray A, et al. Congenital idiopathic growth hormone deficiency associated with prenatal and early postnatal growth failure. The International Board of the Kabi Pharmacia International Growth Study. *J Pediatr*. 1992;121:920–923.
97. Savage MO, Burren CP, Blair JC, et al. Growth hormone insensitivity: pathophysiology, diagnosis, clinical variation and future perspectives. *Horm Res*. 2001;55:32–35.
98. Han VK, Bassett N, Walton J, et al. The expression of insulin-like growth factor (IGF) and IGF-binding protein (IGFBP) genes in the human placenta and membranes: evidence for IGF-IGFBP interactions at the feto-maternal interface. *J Clin Endocrinol Metab*. 1996;81:2680–2693.
99. Jessmon P, Leach RE, Armant DR. Diverse functions of HBEGF during pregnancy. *Mol Reprod Dev*. 2009;76:1116–1127.
100. Leach RE, Romero R, Kim YM, et al. Pre-eclampsia and expression of heparin-binding EGF-like growth factor. *Lancet*. 2002;360:1215–1219.
101. Moll SJ, Jones CJ, Crocker IP, et al. Epidermal growth factor rescues trophoblast apoptosis induced by reactive oxygen species. *Apoptosis*. 2007;12:1611–1622.
102. Di Blasio AM, Carniti C, Vigano P, et al. Basic fibroblast growth factor messenger ribonucleic acid levels in human placentas from normal and pathological pregnancies. *Mol Hum Reprod*. 1997;3:1119–1123.
103. Anteby EY, Natanson-Yaron S, Hamani Y, et al. Fibroblast growth factor-10 and fibroblast growth factor receptors 1-4: expression and peptide localization in human decidua and placenta. *Eur J Obstet Gynecol Reprod Biol*. 2005;119:27–35.
104. Arany E, Hill DJ. Fibroblast growth factor-2 and fibroblast growth factor receptor-1 mRNA expression and peptide localization in placentae from normal and diabetic pregnancies. *Placenta*. 1998;19:133–142.
105. Ferrara N, Houck K, Jakeman L, et al. Molecular and biological properties of the vascular endothelial growth factor family of proteins. *Endocr Rev*. 1992;13:18–32.
106. Ferrara N. Role of vascular endothelial growth factor in regulation of physiological angiogenesis. *Am J Physiol Cell Physiol*. 2001;280:C1358–C1366.
107. Ferrara N. Vascular endothelial growth factor: basic science and clinical progress. *Endocr Rev*. 2004;25:581–611.
108. Ferrara N. Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy. *Nat Med*. 2010;16:1107–1111.
109. Shibuya M. Vascular endothelial growth factor and its receptor system: physiological functions in angiogenesis and pathological roles in various diseases. *J Biochem*. 2013;153:13–19.
110. Andraweera PH, Dekker GA, Roberts CT. The vascular endothelial growth factor family in adverse pregnancy outcomes. *Hum Reprod Update*. 2012;18:436–457.
111. De Falco S. The discovery of placenta growth factor and its biological activity. *Exp Mol Med*. 2012;44:1–9.

112. Shibuya M, Yamaguchi S, Yamane A, et al. Nucleotide sequence and expression of a novel human receptor-type tyrosine kinase gene (flt) closely related to the fms family. *Oncogene*. 1990;5:519–524.
113. Olofsson B, Korpelainen E, Pepper MS, et al. Vascular endothelial growth factor B (VEGF-B) binds to VEGF receptor-1 and regulates plasminogen activator activity in endothelial cells. *Proc Natl Acad Sci U S A*. 1998;95:11709–11714.
114. Holmes K, Roberts OL, Thomas AM, et al. Vascular endothelial growth factor receptor-2: structure, function, intracellular signalling and therapeutic inhibition. *Cell Signal*. 2007;19:2003–2012.
115. Ahmed A, Li XF, Dunk C, et al. Colocalisation of vascular endothelial growth factor and its Flt-1 receptor in human placenta. *Growth Factors*. 1995;12:235–243.
116. Clark DE, Smith SK, Sharkey AM, et al. Localization of VEGF and expression of its receptors flt and KDR in human placenta throughout pregnancy. *Hum Reprod (Oxford, England)*. 1996;11:1090–1098.
117. Torry DS, Torry RJ. Angiogenesis and the expression of vascular endothelial growth factor in endometrium and placenta. *Am J Reprod Immunol*. 1997;37:21–29.
118. Kendall RL, Thomas KA. Inhibition of vascular endothelial cell growth factor activity by an endogenously encoded soluble receptor. *Proc Natl Acad Sci U S A*. 1993;90:10705–10709.
119. Young BC, Levine RJ, Karumanchi SA. Pathogenesis of preeclampsia. *Annu Rev Pathol*. 2010;5:173–192.
120. Zhou Y, Damsky CH, Chiu K, et al. Preeclampsia is associated with abnormal expression of adhesion molecules by invasive cytotrophoblasts. *J Clin Invest*. 1993;91:950–960.
121. Meekins JW, Pijnenborg R, Hanssens M, et al. A study of placental bed spiral arteries and trophoblast invasion in normal and severe preeclamptic pregnancies. *Br J Obstet Gynaecol*. 1994;101:669–674.
122. Zhou Y, McMaster M, Woo K, et al. Vascular endothelial growth factor ligands and receptors that regulate human cytotrophoblast survival are dysregulated in severe preeclampsia and hemolysis, elevated liver enzymes, and low platelets syndrome. *Am J Pathol*. 2002;160:1405–1423.
123. Young IR, Renfree MB, Mesiano S, et al. The comparative physiology of parturition in mammals: hormones and parturition in mammals. In: Norris D, Lopez K, eds. *Hormones and Reproduction in Vertebrates*. Vol. 5. London: Academic Press; 2010.
124. Csapo AI, Pulkkinen MO, Wiest WG. Effects of luteectomy and progesterone replacement therapy in early pregnant patients. *Am J Obstet Gynecol*. 1973;115:759–765.
125. Johansson E. Plasma levels of progesterone in pregnancy measured by a rapid competitive protein binding technique. *Acta Endocrinol*. 1979;61:607.
126. Tulchinsky D, Hobel CJ, Yeager E, et al. Plasma estrone, estradiol, estriol, progesterone, and 17-hydroxyprogesterone in human pregnancy. I. Normal pregnancy. *Am J Obstet Gynecol*. 1972;112:1095–1100.
127. Avrech OM, Golan A, Weinraub Z, et al. Mifepristone (RU486) alone or in combination with a prostaglandin analogue for termination of early pregnancy: a review. *Fertil Steril*. 1991;56:385–393.
128. Frydman R, Lelaidier C, Baton-Saint-Mieux C, et al. Labor induction in women at term with mifepristone (RU 486): a double-blind, randomized, placebo-controlled study. *Obstet Gynecol*. 1992;80:972–975.
129. Cadepond F, Ulmann A, Baulieu EE. RU486 (mifepristone): mechanisms of action and clinical uses. *Annu Rev Med*. 1997;48:129–156.
130. Haluska GJ, West NB, Novy MJ, et al. Uterine estrogen receptors are increased by RU486 in late pregnant rhesus macaques but not after spontaneous labor. *J Clin Endocrinol Metab*. 1990;70:181–186.
131. Tan H, Yi L, Rote NS, et al. Progesterone receptor-A and -B have opposite effects on proinflammatory gene expression in human myometrial cells: implications for progesterone actions in human pregnancy and parturition. *J Clin Endocrinol Metab*. 2012;97:E719–E730.
132. Hardy DB, Janowski BA, Corey DR, et al. Progesterone receptor plays a major antiinflammatory role in human myometrial cells by antagonism of nuclear factor-kappaB activation of cyclooxygenase 2 expression. *Mol Endocrinol*. 2006;20:2724–2733.
133. Mandala M. Influence of estrogens on uterine vascular adaptation in normal and preeclamptic pregnancies. *Int J Mol Sci*. 2020:21–39.
134. Siiteri PK, MacDonald PC. Placental estrogen biosynthesis during human pregnancy. *J Clin Endocrinol Metab*. 1966;26:751–761.
135. Shozu M, Akasofu K, Harada T, et al. A new cause of female pseudohermaphroditism: placental aromatase deficiency. *J Clin Endocrinol Metab*. 1991;72:560–566.
136. Harada N, Ogawa H, Shozu M, et al. Biochemical and molecular genetic analyses on placental aromatase (P-450AROM) deficiency. *J Biol Chem*. 1992;267:4781–4785.
137. Ben-Jonathan N, LaPensee CR, LaPensee EW. What can we learn from rodents about prolactin in humans? *Endocr Rev*. 2008;29:1–41.
138. Haig D. Genetic conflicts in human pregnancy. *Q Rev Biol*. 1993;68:495–532.
139. Rosenberg KR, Trevathan WR. The evolution of human birth. *Sci Am*. 2001;285:72–77.
140. Rightmire GP. Brain size and encephalization in early to Mid-Pleistocene Homo. *Am J Phys Anthropol*. 2004;124:109–123.
141. Rightmire GP. Brain size and encephalization in early to mid–Pleistocene Homo. *Am J Phys Anthropol*. 2004;124:109–123.
142. Plunkett J, Doniger S, Orabona G, et al. An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet*. 2011;7:e1001365.
143. Plunkett J, Doniger S, Orabona G, et al. An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet*. 2011;7:e1001365.
144. Dunsworth HM, Warrener AG, Deacon T, et al. Metabolic hypothesis for human altriciality. *Proc Natl Acad Sci U S A*. 2012;109:15212–15216.
145. Price HV, Cone BR, Keogh M. Length of gestation in congenital adrenal hyperplasia. *J Obstet Gynaecol Br Commonw*. 1971;78:430–434.
146. Honnebier WJ, Swaab DF. The influence of anencephaly upon intrauterine growth of fetus and placenta and upon gestation length. *J Obstet Gynaecol Br Commonw*. 1973;80:577–588.
147. Patrick J, Challis J, Natale R, et al. Circadian rhythms in maternal plasma cortisol, estrone, estradiol, and estriol at 34 to 35 weeks' gestation. *Am J Obstet Gynecol*. 1979;135:791–798.
148. Liggins GC, Fairclough RJ, Grieves SA, et al. The mechanism of initiation of parturition in the Ewe. *Recent Prog Horm Res*. 1973;29:111–159.
149. Walsh S, Kittinger G, Novy M. Maternal peripheral concentrations of estradiol, estrone, cortisol, and progesterone during late pregnancy in rhesus monkeys (Macaca mulatta) and after experimental fetal anencephaly and fetal death. *Am J Obstet Gynecol*. 1979;135:37–42.
150. Neilson JP. Mifepristone for induction of labour. *Cochrane Database Syst Rev*. 2000:CD002865.
151. Renthal NE, Chen CC, Williams KC, et al. miR-200 family and targets, ZEB1 and ZEB2, modulate uterine quiescence and contractility during pregnancy and labor. *Proc Natl Acad Sci U S A*. 2010;107:20828–20833.
152. Merlino AA, Welsh TN, Tan H, et al. Nuclear progesterone receptors in the human pregnancy myometrium: evidence that parturition involves functional progesterone withdrawal mediated by increased expression of progesterone receptor-A. *J Clin Endocrinol Metab*. 2007;92:1927–1933.
153. Amini P, Michniuk D, Kuo K, et al. Human parturition involves phosphorylation of progesterone receptor-a at serine-345 in myometrial cells. *Endocrinology*. 2016;157:4434–4445.
154. Henderson D, Wilson T. Reduced binding of progesterone receptor to its nuclear response element after human labor onset. *Am J Obstet Gynecol*. 2001;185:579–585.
155. Condon JC, Jeyasuria P, Faust JM, et al. A decline in the levels of progesterone receptor coactivators in the pregnant uterus at term may antagonize progesterone receptor function and contribute to the initiation of parturition. *Proc Natl Acad Sci U S A*. 2003;100:9518–9523.
156. Dong X, Shylnova O, Challis JR, et al. Identification and characterization of the protein-associated splicing factor as a negative co-regulator of the progesterone receptor. *J Biol Chem*. 2005;280:13329–13340.
157. Mahendroo MS, Porter A, Russell DW, et al. The parturition defect in steroid 5alpha-reductase type 1 knockout mice is due to impaired cervical ripening. *Mol Endocrinol*. 1999;13:981–992.

158. Nadeem L, Shynlova O, Matysiak-Zablocki E, et al. Molecular evidence of functional progesterone withdrawal in human myometrium. *Nat Commun*. 2016;7:11565.
159. Nadeem L, Shynlova O, Mesiano S, et al. Progesterone via its type-a receptor promotes myometrial gap junction coupling. *Sci Rep*. 2017;7:13357.
160. Thomson AJ, Telfer JF, Young A, et al. Leukocytes infiltrate the myometrium during human parturition: further evidence that labour is an inflammatory process. *Hum Reprod*. 1999;14:229–236.
161. Osman I, Young A, Ledingham MA, et al. Leukocyte density and pro-inflammatory cytokine expression in human fetal membranes, decidua, cervix and myometrium before and during labour at term. *Mol Hum Reprod*. 2003;9:41–45.
162. Osman I, Young A, Jordan F, et al. Leukocyte density and proinflammatory mediator expression in regional human fetal membranes and decidua before and during labor at term. *J Soc Gynecol Invest*. 2006;13:97–103.
163. Gravett MG, Witkin SS, Haluska GJ, et al. An experimental model for intraamniotic infection and preterm labor in rhesus monkeys. *Am J Obstet Gynecol*. 1994;171:1660–1667.
164. Baggia S, Gravett MG, Witkin SS, et al. Interleukin-1 beta intra-amniotic infusion induces tumor necrosis factor-alpha, prostaglandin production, and preterm contractions in pregnant rhesus monkeys. *J Soc Gynecol Invest*. 1996;3:121–126.
165. Romero R, Mazor M, Tartakovsky B. Systemic administration of interleukin-1 induces preterm parturition in mice. *Am J Obstet Gynecol*. 1991;165:969–971.
166. Hirsch E, Muhle R. Intrauterine bacterial inoculation induces labor in the mouse by mechanisms other than progesterone withdrawal. *Biol Reprod*. 2002;67:1337–1341.
167. Sadowsky DW, Adams KM, Gravett MG, et al. Preterm labor is induced by intraamniotic infusions of interleukin-1beta and tumor necrosis factor-alpha but not by interleukin-6 or interleukin-8 in a nonhuman primate model. *Am J Obstet Gynecol*. 2006;195:1578–1589.
168. Romero R, Mazor M. Infection and preterm labor. *Clin Obstet Gynecol*. 1988;31:553–584.
169. Romero R, Mazor M, Wu YK, et al. Infection in the pathogenesis of preterm labor. *Semin Perinatol*. 1988;12:262–279.
170. Gibbs RS, Romero R, Hillier SL, et al. A review of premature birth and subclinical infection. *Am J Obstet Gynecol*. 1992;166:1515–1528.
171. Andrews WW, Hauth JC, Goldenberg RL. Infection and preterm birth. *Am J Perinatol*. 2000;17:357–365.
172. Goldenberg RL, Hauth JC, Andrews WW. Intrauterine infection and preterm delivery. *N Engl J Med*. 2000;342:1500–1507.
173. Goldenberg RL, Culhane JF. Infection as a cause of preterm birth. *Clin Perinatol*. 2003;30:677–700.
174. Boggess KA. Pathophysiology of preterm birth: emerging concepts of maternal infection. *Clin Perinatol*. 2005;32:561–569.
175. Romero R, Espinoza J, Goncalves LF, et al. The role of inflammation and infection in preterm birth. *Semin Reprod Med*. 2007;25:21–39.
176. Hendricks CH, Brenner WE, Gabel RA, et al. *The Effect of Progesterone Administered Intra-amniotically in Late Human Pregnancy*. Augusta, Michigan: Paper presented at Brook Lodge Symposium: Progesterone; 1961.
177. Pinto RM, Montuori E, Lerner U, et al. Effect of progesterone on the oxytocic action of estradiol-17-beta. *Am J Obstet Gynecol*. 1965;91:1084–1089.
178. Goldstein P, Berrier J, Rosen S, et al. A meta-analysis of randomized control trials of progestational agents in pregnancy. *Br J Obstet Gynaecol*. 1989;96:265–274.
179. Keirse MJ. Progestogen administration in pregnancy may prevent preterm delivery. *Br J Obstet Gynaecol*. 1990;97:149–154.
180. da Fonseca EB, Bittar RE, Carvalho MH, et al. Prophylactic administration of progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth in women at increased risk: a randomized placebo-controlled double-blind study. *Am J Obstet Gynecol*. 2003;188:419–424.
181. Meis PJ, Klebanoff M, Thom E, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate, *N Engl J Med*. 2003;348:2379–2385.
182. Norman JE, Marlow N, Messow CM, et al. Vaginal progesterone prophylaxis for preterm birth (the OPPTIMUM study): a multicentre, randomised, double-blind trial. *Lancet*. 2016;387:2106–2116.
183. Blackwell SC, Gyamfi-Bannerman C, Biggio Jr JR, et al. 17-OHPC to prevent recurrent preterm birth in singleton gestations (PROLONG Study): a multicenter, international, randomized double-blind trial. *Am J Perinatol*. 2020;37:127–136.
184. Harada N, Ogawa H, Shozu M, et al. Genetic studies to characterize the origin of the mutation in placental aromatase deficiency. *Am J Hum Genet*. 1992;51:666–672.
185. McLean M, Bisits A, Davies J, et al. Predicting risk of preterm delivery by second-trimester measurement of maternal plasma corticotropin-releasing hormone and alpha-fetoprotein concentrations. *Am J Obstet Gynecol*. 1999;181:207–215.
186. Emanuel RL, Torday JS, Asokananthan N, et al. Direct effects of corticotropin-releasing hormone and thyrotropin-releasing hormone on fetal lung explants. *Peptides*. 2000;21:1819–1829.
187. Thomson M, Smith R. The action of hypothalamic and placental corticotropin releasing factor on the corticotrope. *Mol Cell Endocrinol*. 1989;62:1–12.
188. Livesey JH, Evans MJ, Mulligan R, et al. Interactions of CRH, AVP and cortisol in the secretion of ACTH from perifused equine anterior pituitary cells: "permissive" roles for cortisol and CRH. *Endocr Res*. 2000;26:445–463.
189. Grammatopoulos DK, Hillhouse EW. Role of corticotropin-releasing hormone in onset of labour. *Lancet*. 1999;354:1546–1549.
190. Asa SL, Kovacs K, Singer W. Human fetal adenohypophysis: morphologic and functional analysis in vitro. *Neuroendocrinology*. 1991;53:562–572.
191. Smith R, Mesiano S, Chan EC, et al. Corticotropin-releasing hormone directly and preferentially stimulates dehydroepiandrosterone sulfate secretion by human fetal adrenal cortical cells. *J Clin Endocrinol Metab*. 1998;83:2916–2920.
192. Condon JC, Jeyasuria P, Faust JM, et al. Surfactant protein secreted by the maturing mouse fetal lung acts as a hormone that signals the initiation of parturition. *Proc Natl Acad Sci U S A*. 2004;101:4978–4983.

130

Fetal and Neonatal Endocrinology

Harshini Katugampola, Ranna El Khairi, and Mehul T. Dattani

OUTLINE

Overview, 2152
Placental Hormone Production and Transfer, 2153
Ectopic Fetal Hormone Production, 2153
Fetal Endocrine Systems, 2154
The Hypothalamic-Pituitary Systems, 2155
  Fetal Adrenal, 2163
  Fetal Gonads, 2166
  Fetal Thyroid, 2168
The Fetal Autonomic Nervous System, 2173
Fetal Pancreas and Glucose Homeostasis, 2174
Fetal Parathyroid and Bone Metabolism, 2178
  Parathyroid Development, 2178
  Transcription Factors Involved in the Development of Parathyroid Glands, 2178
  Fetal Calcium and Phosphate Metabolism, 2179
  Regulation of Calcium Homeostasis by PTH/PTHrP, 2179
  Neonatal Disorders of Calcium and Phosphate Homeostasis, 2181
Fetal Growth, 2181
  Insulin-Like Growth Factors, 2181
  Insulin, 2183
  Fibroblast Growth Factor, 2183
  Other Growth Factors and Signaling Pathways, 2183
Adaptations to Extrauterine Life, 2183
  Cortisol-Mediated Adaptations, 2183
  Catecholamine-Mediated Adaptations, 2184
  BAT Thermogenesis, 2184
  Glucose Homeostasis, 2185
  Calcium Homeostasis, 2185
  Other Hormonal Adaptations, 2186
Future Directions, 2186

Successful pregnancy involves complex genetic, cellular, and hormonal interactions facilitating implantation, placentation, fetal development, parturition, and fetal adaptation to extrauterine life. An array of signaling molecules, transcription factors, and epigenetic events program embryogenesis and fetal development, in concert with autocrine, paracrine, and endocrine networks of hormones that coordinate maternal-placental-fetal interactions. A growing understanding of the spectrum of hormones and their mechanisms of action, as well as advances in molecular genetics, has helped to elucidate the physiology of fetal endocrine systems and the pathological basis of many neonatal disorders. It is becoming clear that the intrauterine environment can also have a significant impact on the lifelong health of an individual, as well as exerting multigenerational effects. This chapter reviews our current understanding of the transplacental passage of hormones, development of fetal endocrine systems, fetal growth, and adaptations of the fetal endocrine systems to extrauterine life and endocrine disorders affecting the neonate.

OVERVIEW

The unfolding of our understanding of mammalian pregnancy and fetal development represents one of the dramatic chapters of scientific progress during the past half-century. Successful pregnancy involves complex genetic, cellular, and hormonal interactions facilitating implantation, placentation, embryonic and fetal development, parturition, and fetal adaptation to extrauterine life. The evolution of pregnancy is accompanied by dramatic changes in the hormonal milieu. This is a time of not only potential vulnerability but also tremendous plasticity. An array of signaling molecules, transcription factors, and epigenetic events program embryogenesis and fetal development, in concert with autocrine, paracrine, and endocrine networks of hormones and growth factors that coordinate maternal-placental-fetal interactions.

The placental-fetal endocrine environment includes a wide spectrum of placental hormones and growth factors and a variety of fetal endocrine adaptations to the intrauterine environment. The fetal adrenal cortex, the para-aortic chromaffin system including the paired organs of Zuckerkandl, and the intermediate lobe of the pituitary are prominent among the unique fetal endocrine tissues. Cortisol and thyroid hormones are largely inactivated during much of fetal life but are programmed to serve critical maturational events during fetal development. Vasotocin, the parent neurohypophyseal hormone in submammalian species, is expressed transiently during fetal life, and calcitonin, a largely vestigial hormone in adult mammals, plays a significant role in fetal calcium and bone metabolism. Other hormones and growth factors that play prominent roles in the fetus include catecholamines, parathyroid hormone (PTH)–related protein (PTHrP), anti-Müllerian hormone (AMH), insulin-like growth factor 2 (IGF-2), transforming growth factor-α (TGF-α), and the neuroregulins. Furthermore, hormonal programming during the fetal-perinatal period conditions the adult functional characteristics of selected endocrine systems.

At parturition, the fetus is faced with the abrupt termination of placental support and rapid transition to the extrauterine environment, in which breathing, circulatory adjustments, endogenous thermogenesis, free water conservation, and mobilization of energy substrates are required for survival. The adrenal cortex and the autonomic nervous system (ANS) play key roles in neonatal transitional physiology, but other endocrine systems are required for optimal homeostasis.

A growing understanding of the spectrum of hormones and their mechanisms of action, as well as advances in molecular genetics, has helped to elucidate the physiology of fetal endocrine systems and the pathological basis of many congenital endocrine disorders. It is becoming increasingly clear that the intrauterine environment not only enables fetal development but can also have a significant impact on the lifelong health of an individual, as well as exerting multigenerational effects.

This chapter reviews our current understanding of the transplacental passage of hormones, development and programming of the fetal endocrine systems, fetal growth, and adaptations of the fetal endocrine systems to extrauterine life and endocrine disorders affecting the neonate.

PLACENTAL HORMONE PRODUCTION AND TRANSFER

The fetal endocrine milieu is largely independent of maternal hormones because the placenta acts as a selective barrier and is impermeable to most peptide hormones. Hormones larger than 0.7 to 1.2 kD have limited or no access to the fetal compartment.[1] Steroid (cortisol), thyroid hormones (T_3, T_4), catecholamines, and estradiol do cross the placenta but are metabolized *en route*.[2–5] The placental cells contain an active 11β-hydroxysteroid dehydrogenase type 2 (11βHSD2) that catalyzes the conversion of maternal cortisol to inactive cortisone.[3,6] This is important considering the steep gradient between maternal cortisol concentration and that of the fetus, amounting to almost a 10-fold difference. Synthetic glucocorticoids (betamethasone, dexamethasone) bypass this protective mechanism. This is utilized acutely in preterm labor to aid lung maturation; however, chronic use in rodent models has been shown to adversely affect fetal growth, as well as postnatal effects on blood pressure, glucose homeostasis, and memory.[7–10] Antenatal steroid treatment for indications other than the obstetric management of threatened preterm delivery lacks robust evidence (e.g., to reduce virilization of a fetus with congenital adrenal hyperplasia [CAH]).[11,12]

Placental production of estrogens in the form of estradiol, estrone, estriol, and estetrol is vast, and by term, the estradiol concentration is 100-fold higher and estriol is 1000-fold higher than in the nonpregnant state.[13] Placental 17βHSD catalyzes inactivation of estradiol to estrone, preventing the passage of excessive estrogens to the fetus.[4]

Transplacental passage of thyroid hormones to the fetus, particularly in the first trimester, is important for neurodevelopment. Placental tissue contains an iodothyronine inner-ring monodeiodinase, which deiodinates most of the T_4 to inactive reverse triiodothyronine (rT_3) and converts active 3,5,3'-triiodothyronine (T_3) to inactive diiodothyronine.[5,14] Nonetheless, there is some transplacental passage of T_4 to the fetus in early pregnancy and this may be of importance as observational studies have suggested altered intellectual function, albeit mild, in infants born to mothers with mild untreated or subclinical hypothyroidism (although this may also be due to placental transfer of thyroid peroxidase TPO antibodies).[15]

Catecholamine-degrading enzymes in placental tissue include both monoamine oxidase and catechol *O*-methyltransferase,[3,9] and both metanephrine and dihydroxymandelic acid metabolites of catecholamines are present in placental homogenates.[16]

In addition to this role in hormone metabolism, the human placenta subserves an important endocrine role during fetal development (see Chapter 129). The placenta is a major source of many of the hypothalamic neuropeptides. Table 130.1 summarizes the menu of placental hormones and growth factors, similar or identical to those secreted from the anterior pituitary (AP, such as placental growth hormone [GH] and placental lactogen) or other nonplacental organs. The significance of many of these remains unclear, but the overall contribution to the unique endocrine environment is substantial.

KEY POINTS

- Pregnancy produces a unique endocrine milieu.
- The fetal endocrine environment is largely independent of maternal hormones because the placenta acts as a selective barrier, impermeable to most peptide hormones.

ECTOPIC FETAL HORMONE PRODUCTION

Human chorionic gonadotropin (hCG) is produced abundantly in the first trimester by the placenta and is preferentially secreted into the maternal circulation.[17] hCG and luteinizing hormone (LH; a structural and functional homolog of hCG produced by the AP) bind to the same receptor, although hCG has a higher binding affinity. hCG is present in the fetal circulation. Kidney, liver, and testes from 16- to 20-week-old human fetuses produce immunoreactive and bioactive hCG *in vitro*.[18,19] Kidney tissue produces nearly half as much hCG per milligram of protein as placenta; liver activity is lower. hCG is also found in the fetal ovary, lung, adrenal, thymus, spleen, and muscle.[20] hCG is able to stimulate the production of dehydroepiandrosterone (DHEA) and testosterone by Leydig cells, which is important for masculinization of fetal tissues. hCG signaling in other tissues may be important for their growth and differentiation.[21,22]

Adrenocorticotropic hormone (ACTH)–like immunoreactivity is present in relatively high concentrations in neonatal rat pancreas and kidneys.[23] This hormone is presumably derived from a pro-opiomelanocortin (POMC) parent molecule. POMC is also secreted by the fetal lung and is known to alter adrenal sensitivity to ACTH. The interplay between fetal pituitary, lung, and adrenal may regulate response to stress or the onset of parturition.[24]

Ghrelin is a hormone with potent GH-releasing activity produced by human gastric endocrine cells and by the pituitary, hypothalamus, and placenta. Fetal lung is an additional source of circulating ghrelin, whose function is as yet unknown.[25,26]

Hypothalamic neuropeptides (thyrotropin-releasing hormone [TRH], corticotrophin-releasing hormone [CRH], GH-releasing hormone [GHRH]) are present in a variety of adult tissues, particularly in the pancreas and gut.[27,28] In the fetus, hypothalamic neuropeptides are also present in the gut and tissues derived from it. High concentrations of TRH and somatostatin immunoreactivity have been reported in neonatal rat pancreas and gastrointestinal tract tissues, whereas hypothalamic concentrations are low.[29,30] These neuropeptides have immunoreactive and chromatographic properties similar to those of the synthetic hypothalamic peptides. Encephalectomy does not alter the circulating TRH levels in the neonatal rat, whereas significant reductions are produced by pancreatectomy. In the sheep fetus, thyroid hormones modulate pancreatic and gut TRH concentrations, which suggests thyroid hormone control of extrahypothalamic *TRH* gene transcription or translation in the fetus.[31] TRH and somatostatin are also present in the human neonatal pancreas and in the blood of the human newborn where both hormones are derived mostly from extrahypothalamic sources.[32–35] The presence of TRH at high concentrations in fetal ovine blood and the modulation of fetal pancreatic, placental, and blood TRH levels by thyroid hormones suggests a role for extrahypothalamic TRH in the control of fetal pituitary thyrotropin secretion before the near-term maturation of hypothalamic TRH.[31] The role of extraneural somatostatin in the fetus is undefined. CRH is produced by the placenta and is identical to hypothalamic CRH. The circulating maternal concentration of CRH increases 1000-fold as pregnancy progresses, reaching 0.5 to 1 nmol/L at term. Fetal CRH concentrations are lower (50 pmol) but are still substantial compared to nonpregnant women and men (<0.01 nmol/L). Placental *CRH* expression and production are stimulated by glucocorticoids (in

TABLE 130.1 Hormones and Growth Factors Produced by the Placenta

Steroid Hormones

Estradiol
Estriol
Estrone
Progesterone

Protein and Peptide Hormones

Activin
Adrenomedullin
Angiotensinogen
Calcitonin
Chorionic gonadotropin
Chorionic somatomammotropin
Chorionic thyrotropin
Follistatin
Placental growth hormone
Inhibin
Leptin
Oxytocin
Pit-1
Placental prolactins
Proliferin
Pro-opiomelanocortin
  ACTH
  β-Endorphin
  α/MSH
  α/β-Lipotropin
PTH-related protein
Relaxin
Renin
Sex hormone–binding globulin
Urocortin

Neuropeptides

CRH
GHRH
GnRH
Ghrelin
Neuropeptide Y
Somatostatin
TRH
PrRP

Growth Factors

Colony-stimulating factors
EGF
Endothelin I
Erythropoietin
Fibroblast growth factor
Hepatocyte growth factor
IGF-1
IGF-2
IGF-binding proteins 1-6
Interleukins
Nerve growth factor
Oncomodulin
PDGF
Placental growth factor
TGF-α
TGF-β
Tumor necrosis factor-α
Vascular endothelial growth factor

ACTH, Adrenocorticotropic hormone; *CRH,* corticotrophin-releasing hormone; *EGF,* epidermal growth factor; *GHRH,* growth hormone–releasing hormone; *GnRH,* gonadotropin-releasing hormone; *IGF,* insulin-like growth factor; *MSH,* melanocyte-stimulating hormone; *PDGF,* platelet-derived growth factor; *PTH,* parathyroid hormone; *PrRP,* prolactin-releasing peptide; *TGF,* transforming growth factor; *TRH,* thyrotropin-releasing hormone.
Data from Fisher DA. Fetal and neonatal endocrinology. In: DeGroot LJ, Jameson JL, eds. *Endocrinology*, 5th ed. Philadelphia: Elsevier Saunders; 2006:3369–3386; Sodha RJ, Proegler M, Schneider H. Transfer and metabolism of norepinephrine studied from maternal to fetal and fetal to maternal sides in the in vitro perfused human placental life. *Am J Obstet Gynecol.* 1984;148:474–481.

contrast to hypothalamic CRH).[36] This positive feed-forward system for placental CRH is unique to pregnancy. *In vitro* placental CRH stimulates pituitary ACTH from cultured primary adrenal cells. This suggests that *in vivo* placental CRH may stimulate pituitary ACTH production, promoting fetal adrenal steroidogenesis.[37–39] The fetal adrenal is an androgen factory, producing substrate for placental estrogen synthesis. The exponential increase in maternal CRH from 35 weeks' gestation to term has led to the proposal that this acts as a biological clock, initiating the onset of parturition by increasing estrogen production.[40]

GH-V encodes placental GH, and *hPL-A*, *hPL-B*, and *hPL-L* encode placental lactogens (hPL; also called chorionic somatomammotropins [CSH]). The major circulating placental lactogens are derived from the *hPL-A* and *hPL-B* genes. *GH-V* rises sharply after midgestation to a peak at 34 to 37 weeks and, within 1 hour after delivery of the placenta, it disappears from the circulation.[41] *GH-V* is not secreted into the fetal circulation and exclusively plays a role in maternal and possibly uteroplacental physiology. hPL is structurally homologous to GH but functionally closer to prolactin (PRL) and is secreted directly into both fetal and maternal circulations. hPL is first detected in the mother at 6 weeks of gestation to reach a peak of 5000 to 7000 ng/mL at 32 to 35 weeks; fetal hPL concentrations, however, approximate 20 to 50 ng/mL at term. hPL concentrations increase with placental mass and are higher in twin than in singleton pregnancies. In the mother, lactogens affect insulin production, hypothalamic gene expression, and leptin action and so maintain metabolic homeostasis while providing the substrates for nutrition for the fetus and newborn infant. Placental GH-V is secreted into the maternal circulation and reduces insulin sensitivity in the mother and so spares glucose and other nutrients for transplacental delivery and fetal growth. As a counterbalance, insulin production and beta-cell mass increase during pregnancy.[42] At least in rodents, evidence suggests that PRL and placental lactogen signaling through the PRL receptor are essential for the increase in beta-cell mass. A surge in PRL concentration in the perinatal period parallels an increase in beta-cell replication, and a whole-body knockout of the PRL receptor reduces beta-cell mass and insulin secretion in the perinatal period.[43]

KEY POINTS

- Ectopic hormone production by the fetus is evident but often has uncertain functional significance.

FETAL ENDOCRINE SYSTEMS

Fetal endocrine systems can be classified as shown in Table 130.2. The neuroendocrine hormone systems transduce or convert neural into

TABLE 130.2 Features of the Fetal Endocrine Environment

- Placental Hormone Production
  - Estrogens
  - Progesterone
  - Neuropeptides
  - Growth factors
- Neutralization of Hormone Actions
  - Growth hormone
  - Cortisol
  - Thyroxine
  - Catecholamines
- Unique Fetal Endocrine Systems
  - Fetal adrenal cortex
  - Para-aortic chromaffin system
  - Intermediate lobe of the pituitary gland
- Prominent Fetal Hormones or Metabolites
  - Vasotocin
  - Calcitonin
  - Cortisone
  - Reverse triiodothyronine (rT$_3$)
  - Sulfated iodothyronines
  - Ectopic neuropeptides
- Fetal Endocrine System Adaptations
  - Adrenal-placental interactions
  - Testicular control of male phenotypic differentiation
  - Developmentally regulated growth factor control of fetal growth
  - Neuropeptides and fetal water metabolism
  - Parathyroid glands and placental calcium transport
  - Catecholamine and vasopressin responses to hypoxia
  - Cortisol programming for extrauterine exposure
  - Catecholamine and cortisol control of extrauterine adaptation
  - Perinatal hormonal programming

endocrine information. These include the hypothalamic-AP system and the hypothalamic-posterior pituitary (PP) system. The insulin-glucagon and PTH-calcitonin systems function autonomously, with hormone secretion regulated largely by local feedback mechanisms. The events of parturition dispel the period of protected fetal development and precipitate a series of profound metabolic stresses. The neuroendocrine transducer systems are well developed at birth and function smoothly to defend against these stresses. The autonomous endocrine systems regulating blood glucose and calcium concentrations, however, are relatively suppressed *in utero* and must adapt rapidly to the extrauterine environment. Abnormalities related to homeostasis of blood glucose and calcium concentrations are relatively frequent in the early neonatal period, particularly in premature babies.

THE HYPOTHALAMIC-PITUITARY SYSTEMS

The pituitary is located in the midline of the brain within the sella turcica and consists of three lobes of dual embryologic origin. The adenohypophysis (anterior and intermediate lobes) originates from Rathke's pouch, an invagination of the oral ectoderm. The neurohypophysis (posterior lobe) develops from the neural ectoderm of the ventral diencephalon.[44–47]

The AP consists of five different cell lineages producing six hormones: somatotrophs (GH), gonadotrophs (follicle-stimulating hormone [FSH] and LH), corticotrophs (ACTH), thyrotrophs (thyroid-stimulating hormone [TSH]), and lactotrophs (PRL).

The intermediate lobe contains melanotrophs, which secrete POMC, a major precursor to endorphins, and α-melanocytic–stimulating hormone (α-MSH).[48] Cleavage of POMC in the anterior lobe produces corticotropin and β-lipoprotein. The human fetal pituitary contains high concentrations of compounds resembling α-MSH and corticotropin-like intermediate lobe peptide (CLIP).[49,50] The role of these peptides in the human fetus remains unclear. CLIP and α-MSH may both play a role in fetal adrenal activation, and α-MSH may influence fetal growth.[51,52] Intermediate lobe cells begin to disappear near term and are virtually absent in the adult human pituitary.[53]

The PP releases two hormones, oxytocin and antidiuretic hormone (ADH; also known as arginine vasopressin [AVP]), which are produced in the hypothalamus (supraoptic and paraventricular nuclei) and transported axonally via the pituitary stalk to be stored and released from the PP.[54]

Pituitary Development. The human fetal forebrain is identifiable by 3 weeks' gestation. The diencephalon and telencephalon can be seen by 5 weeks, and by this point, Rathke's pouch separates from the primitive pharyngeal stomodeum.[55–57] The hypothalamus, pituitary stalk, and PP are largely developed by 7 weeks' gestation, and the bony floor of the sella turcica is present separating the adenohypophysis from the primitive gut.

Pituitary development results from temporospatial interactions of transcription factors and signaling molecules. These occur in distinct and sequential developmental steps, which are critical for pituitary organogenesis. Although direct evidence in humans is lacking, the process of pituitary development is highly conserved across all vertebrate species including rodents; development of the mouse pituitary, in particular, is very well-characterized.[55]

In the mouse, the pituitary placode appears at embryonic (E) day 7.5, located ventrally to the anterior neural ridge and next to the future hypothalamo-infundibular region, which will give rise to the roof of the oral cavity (Fig. 130.1). During the next 24 hours, as the anterior neural tube bends and rapidly expands, the hypophyseal placode is displaced ventrally to become situated under, and in contact with, the prospective forebrain within the ectoderm at the roof of the future oral cavity. By E9.0, it invaginates dorsally and the rudimentary Rathke's pouch is formed.[58] The definitive Rathke's pouch, formed by E10.5, gives rise to the anterior and intermediate lobes. The posterior lobe is derived from the posterior part of the developing diencephalon. The overlying neural ectoderm evaginates to form the infundibulum (from which the PP and pituitary stalk will derive), which comes into direct contact with Rathke's pouch. The juxtaposition of Rathke's pouch and the diencephalon is maintained throughout the early stages of pituitary organogenesis. This close relationship is required for tissue interactions between neural and oral ectoderm, which are critical for the initial stages of pituitary specification. By E12.5, the pouch is completely separated from the underlying oral ectoderm; precursors of the hormone-secreting cells start to proliferate ventrally from the pouch to constitute the future AP. The lumen of the pouch persists as the pituitary cleft, separating the anterior and intermediate lobes in the mature gland. The iterative nature of the inductive interactions required for pituitary morphogenesis makes it very sensitive to both loss- and gain-of-function mutations.[59]

Hormone-producing terminally differentiated cells in the adult AP are organized in three-dimensional homotypic cell networks. In the embryo, the first differentiated cells appear as isolated cells, but rapidly aggregate with other cells producing the same hormone into strands of cells and then into three-dimensional networks. The first hormone-producing network to appear is the POMC lineage at E13 in the mouse, followed by the somatotrope/lactotrope and gonadotrope

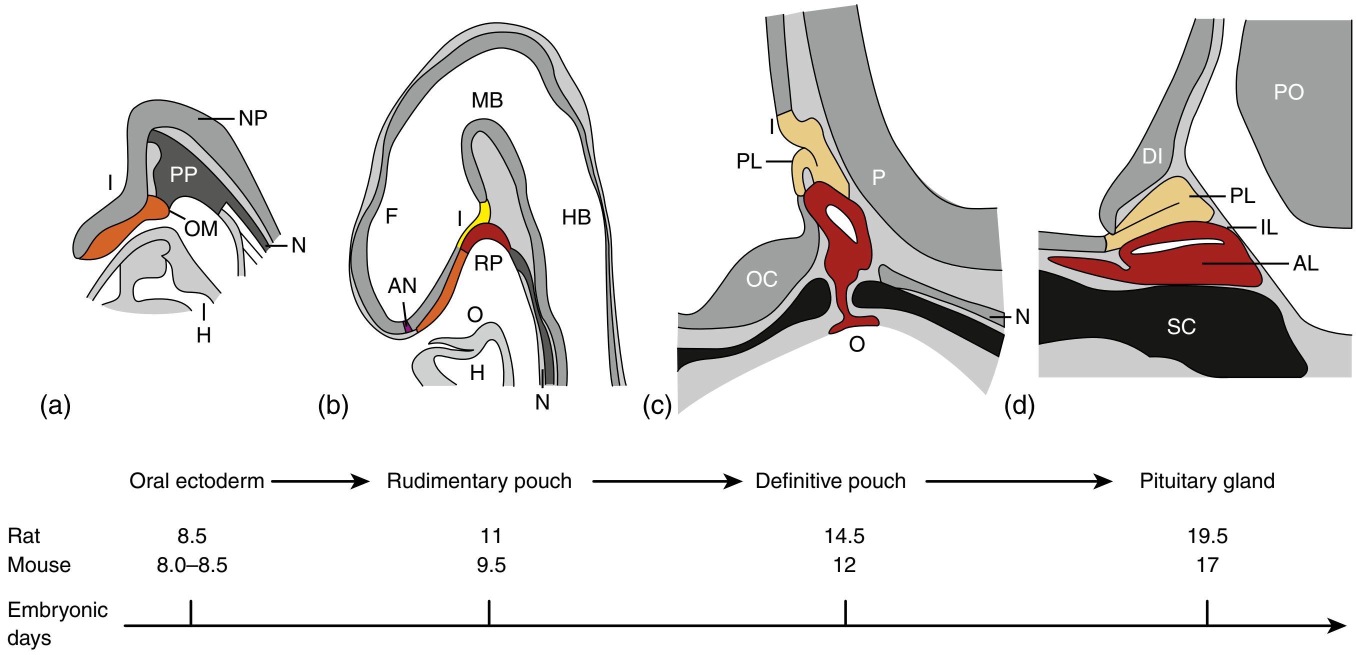


Figure 130.1 Schematic representation of the developmental processes implicated in pituitary development. (a) Oral ectoderm. (b) Rudimentary pouch. (c) Definitive pouch. (d) Adult pituitary gland. *AL,* Anterior lobe; *AN,* anterior neural pore; *DI,* diencephalon; *F,* forebrain; *H,* heart; *HB,* hindbrain; *I,* infundibulum; *IL,* intermediate lobe; *MB,* midbrain; *N,* notochord; *NP,* neural plate; *O,* oral cavity; *OC,* optic chiasma; *OM,* oral membrane; *P,* pontine flexure; *PL,* posterior lobe; *PO,* pons; *PP,* prechordal plate; *RP,* Rathke's pouch; *SC,* sphenoid cartilage. (Adapted from Sheng HZ, Westphal H. Early steps in pituitary organogenesis. *Trends Genet.* 1999;15:236–240, with permission from Elsevier.)

network. A network of thyrotrope cells has not been visualized thus far.[60] Specialized AP cell types, including lactotropes, somatotropes, corticotropes, thyrotropes, and gonadotropes, can be recognized in the human AP between 7 and 16 weeks.[57]

The fetal pituitary can synthesize and secrete GH by 8 to 10 weeks' gestation. Pituitary GH increases from ~1 nmol (20 ng) at 10 weeks to 45 nmol (1000 ng) at 16 weeks. The high plasma GH concentrations after the development of the pituitary portal vascular system at midgestation may reflect unrestrained secretion.[57] Control of GH secretion matures progressively during the last half of gestation and during the first few weeks of postnatal life; the mechanisms are not yet fully understood, however, mature responses to sleep, glucose and l-dopa are present by 3 months of age. GH pulsatility is evident soon after birth in humans, but trough concentrations are higher than in later life so random GH sampling can be used to screen for GH deficiency (GHD) in the neonatal period.[61]

Fetal plasma PRL is low until 25 to 30 weeks' gestation and increases to 11 nmol/L at term. PRL increases in response to TRH and decreases in response to dopamine. Brain and hypothalamic control of PRL matures late in gestation and during the first few months of postnatal life.[57]

There is a general tendency toward hypersecretion of fetal pituitary hormones during the last half of gestation. Pituitary hormones found at high concentrations in cord blood from aborted human fetuses and premature human infants include GH, TSH, ACTH, β-endorphin, β-lipotropin, LH, and FSH.[57,62] Development of hypothalamic-pituitary control involves maturational events in the cortex and midbrain, the hypothalamus and hypothalamic-pituitary portal vascular system, peripheral endocrine systems, and perhaps placental hormone and neuropeptide production. The fetal pituitary hypersecretion appears to be related to relatively delayed maturation of the central nervous system (CNS) and hypothalamic control of secretion of the hypothalamic-releasing hormones.[57]

The fetal neurohypophysis is well developed by 10 to 12 weeks' gestation. Fetal pituitary oxytocin concentration, detectable by 11 to 15 weeks, exceeds AVP concentration by 19 weeks. In the fetus, AVP appears to function as a stress-responsive hormone. The major potential stress for the fetus is hypoxia. Plasma AVP concentrations in human cord blood are elevated in association with hypoxia, bradycardia, and meconium passage.[63] Water channel receptors (aquaporins 1, 2, 3) are present in the newborn infant to regulate free water clearance in response to volume and osmolar stimuli.[64] Lack of AVP is associated with diabetes insipidus (DI) with a failure to retain water resulting in polyuria and polydipsia postnatally. DI may be due to congenital causes such septo-optic dysplasia (SOD; a combination of pituitary, eye, and midline forebrain defects) or holoprosencephaly (HPE), or rarely be due to mutations in the *AVP-neurophysin* gene.

Hypothalamus and Pituitary Stalk Development. The anatomy of the developed hypothalamus is well understood. It extends from the anteriorly located optic chiasm to the posteriorly located mammillary body and is organized into distinct rostral-to-caudal regions: preoptic, anterior, tuberal, and mammillary. The hypothalamus is also subdivided into three medial-to-lateral regions: periventricular, medial, and lateral.[65] Contained within the medial region is the medial preoptic nucleus, anterior hypothalamus, the dorsomedial nucleus, the ventromedial nucleus, and the mammillary nuclei. The lateral zone consists of the preoptic area and hypothalamic area.[65]

As Rathke's pouch invaginates, part of the ventral diencephalon evaginates ventrally to form the infundibulum and later the PP lobe and pituitary stalk. The pituitary stalk acts as a physical connection between the pituitary gland and brain and contains the hypophyseal (hypothalamo-pituitary) portal system, as well as the neuronal connections traversing across the hypothalamic median eminence. The neurons originate from hypothalamic magnocellular bodies termed the supraoptic, suprachiasmatic, and paraventricular nuclei. The former two nuclei release AVP, while the paraventricular nuclei release oxytocin.[66] The median eminence contains a capillary bed into which parvocellular neurons, located in a variety of hypothalamic nuclei, secrete hypophysiotropic hormones, which then reach the anterior and intermediate pituitary lobes via the hypophyseal portal system. These stimulate the release of the seven anterior/intermediate pituitary lobe hormones via the hypophyseal portal system. The parvocellular neurons also secrete oxytocin and AVP, although at much lower concentrations than the magnocellular neurons, with the parvocellular-derived AVP acting synergistically with CRH in regulating ACTH release. It is therefore evident that it is the hypothalamus that is the central mediator of growth, reproduction, and homeostasis, acting through the pituitary gland.

Deciphering embryonic hypothalamic development has proved problematic, due to its anatomical complexity and highly diverse collection of cell groups and neuronal subtypes.[67] Furthermore, genetic expression studies within the hypothalamus have knock-on effects on multiple neuronal subtypes and downstream physiological processes. However, studies are slowly elucidating hypothalamic development.

The hypothalamic cell condensations, which represent the hypothalamic nuclei, and the interconnecting fiber tracts are demonstrable by 15 to 18 weeks.[57,68] Hypothalamic neurons and diencephalic fiber tracts for the neuropeptides somatostatin (SS), GHRH, TRH, and gonadotropin-releasing hormone (GnRH) are visible by this time. Concentrations of dopamine, TRH, GnRH, and SS are significant in hypothalamic tissue by 10 to 14 weeks of gestation.

Maturation of the pituitary portal vascular system continues through 30 to 35 weeks, and the system becomes functional with portal vascular extension into the hypothalamus. The definitive Rathke's pouch comprises proliferative progenitors that will gradually relocate ventrally, away from the lumen as they differentiate. A proliferative zone containing SRY (sex-determining region Y)-box 2 [SOX-2]–expressing progenitors is maintained in the mouse embryo in a periluminal area and was recently found to persist in the adult.[69–71] These progenitors are capable of giving rise to all of the cell types within the AP.

Secretory granules are present within AP cells by 10 to 12 weeks, and pituitary hormones, including GH, prolactin (PRL), TSH, LH, FSH, ACTH, oxytocin (OT), vasotocin (VT), and arginine vasopressin (AVP), can be identified by immunoassay between 10 and 17 weeks.[57,62]

Thus the anatomy and biosynthetic mechanisms that make up the hypothalamic-pituitary neuroendocrine transducer appear functional by 12 to 17 weeks of gestation.

Genes Involved in Pituitary Development and Disease. Normal pituitary development is complex, involving a cascade of signaling molecules and transcription factors that play a crucial role in organ commitment, cell proliferation, cell patterning, and terminal differentiation (Fig. 130.1).

Cells within the primordium of the pituitary are initially competent to differentiate into all cell types. Following expression of the earliest markers of pituitary gland development (e.g., homeobox gene expressed in embryonic stem cells [*Hesx1*]), further signaling pathways are established that direct these cells toward terminal differentiation. Signaling molecules and transcription factors are expressed sequentially at critical periods of pituitary development and many are subsequently attenuated (Fig. 130.2A,B). Genes that are expressed early are implicated in organ commitment but are also implicated in repression and activation of downstream target genes that have specific roles in directing the cells toward a particular fate. Extrinsic molecules within the ventral diencephalon and surrounding structures, such as bone morphogenetic proteins 2 and 4 (B*mp 2, 4*), fibroblast growth factor 8 (F*gf8*), sonic hedgehog (S*hh*), wingless (W*nt4*), thyroid transcription factor 1 (T*tf1*; also called N*kx2.1*), and molecules involved in Notch signaling play critical roles in early organogenesis.[55,56,72] A close interaction between oral ectoderm and neural ectoderm is critical for the initial development of the pituitary gland. Formation of the rudimentary Rathke's pouch is dependent upon *Bmp4*, and *Fgf8* activates two key regulatory genes, LIM homeobox 3 (*Lhx3*) and LIM homeobox 4 (*Lhx4*), that are essential for subsequent development of the rudimentary pouch into a definitive pouch. Both *Bmp4* and *Fgf8* are present only in the diencephalon and not in Rathke's pouch. Murine mutations within the gene encoding TTF1/Nkx2-1 (also known as thyroid-specific enhancer-binding protein), which is expressed only in the presumptive ventral diencephalon, can cause severe defects in the development of the diencephalon and AP. Conditional deletion of *RBPJ* that encodes RBPJ (recombination signal binding protein for immunoglobulin kappa J region), the major mediator of the Notch pathway, leads to the conversion of the late pituitary-specific transcription factor 1 (*Pit1*) lineage into the early corticotroph lineage. Notch signaling is required for maintaining the expression of prophet of Pit1 (*PROP1*), which is required for the generation of the *Pit1* lineage. Attenuation of Notch signaling is necessary for terminal differentiation in *Pit1* cells and for maturation and proliferation of the GH-producing somatotrophs.[73]

Generally, mutations in genes involved in early development (*HESX1*, *LHX3*, *LHX4*, *SOX2*, *SOX3*, *GLI2*, and *OTX2*) tend to be part of a syndrome that includes extrapituitary defects and midline abnormalities, such as cleft lip and/or palate, and eye abnormalities, as well as congenital hypopituitarism. Conversely, mutations in genes implicated in later stages (*PROP1* and *POU1F1*) result in variable phenotypes of combined pituitary hormone deficiencies (CPHD) without any extrapituitary defects.[74]

The variety of phenotypes seen in patients with CPHD is a reflection of the close developmental relationships seen during organogenesis of the pituitary gland, eye, optic nerve, ear, nose, and cranial nerve ganglia. Different gene mutations can result in the same phenotype, and different phenotypes can be secondary to the same single genetic mutation. Therefore the clinical phenotype and associated morphological findings are crucial in the investigation of the underlying genetic mutations of cases of CH.

Spontaneous or artificially induced mutations in the mouse have led to significant insights into human pituitary disease, and identification of mutations associated with human pituitary disease has in turn been invaluable in defining the genetic cascade responsible for the development of this embryological tissue.[55,75] No genetic etiology has been identified in the vast majority of cases of CH to date, suggesting a role for other unidentified genes, oligogenic disease, or environmental factors. However, it is also likely that further genes implicated in hypothalamo-pituitary disease will be identified that play a role in hypothalamic development. Mutations involved specifically in human hypothalamo-pituitary disease[74] are listed in Table 130.3 and will be briefly discussed below.

Syndromic Forms of CH. Syndromic forms of CH are mainly due to mutations in transcription factors implicated during early pituitary development.

SOD and Its Variants. SOD (de Morsier syndrome) is a heterogeneous and complex disorder comprising at least two of the following: (1) optic nerve hypoplasia (ONH), (2) midline abnormalities seen in brain and pituitary magnetic resonance imaging (MRI) (mainly agenesis of the corpus callosum (ACC) and absence of the septum pellucidum), or (3) pituitary hypoplasia with hypopituitarism. The prevalence of SOD is 1 in 10,000 live births.[74,76–78] Transcription

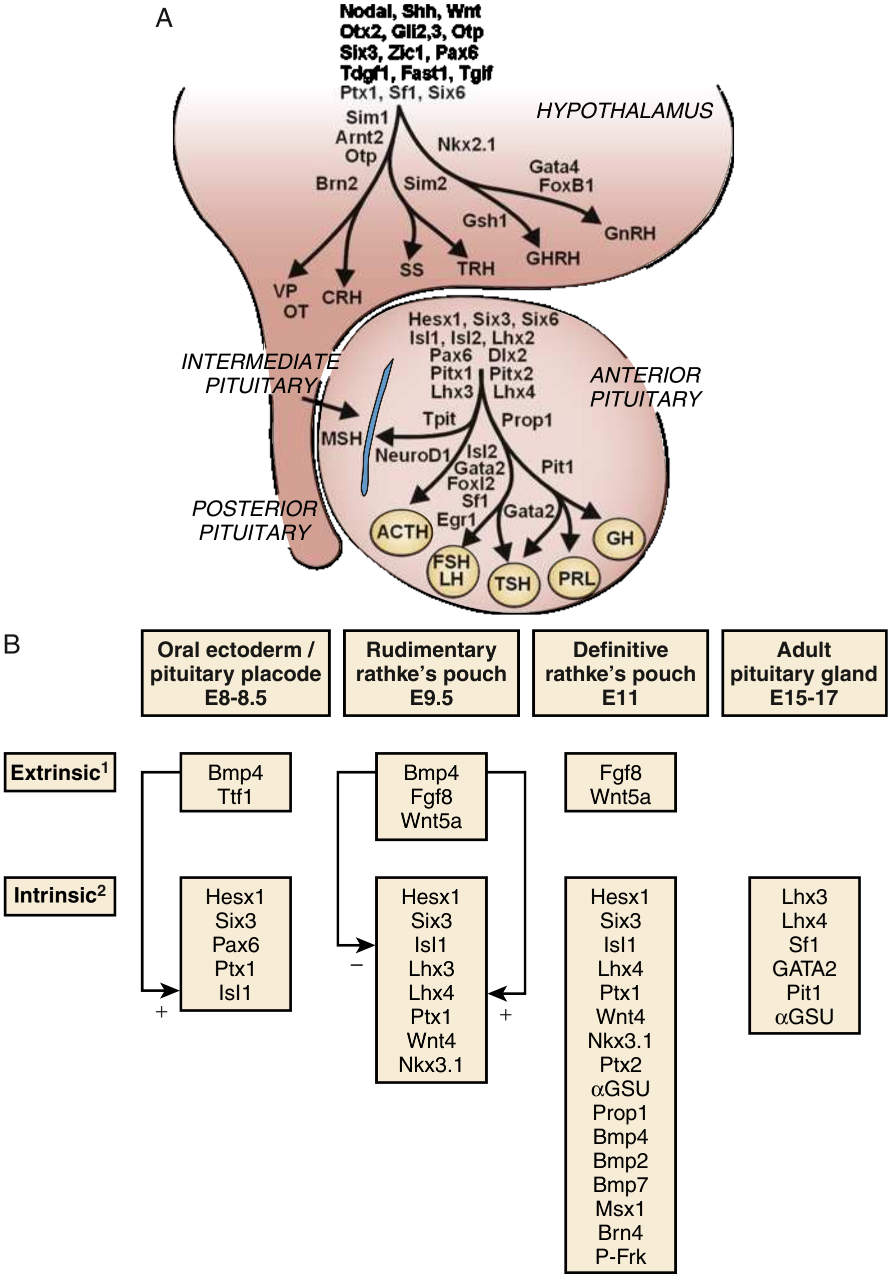

A
Nodal, Shh, Wnt
Otx2, Gli2,3, Otp
Six3, Zic1, Pax6
Tdgf1, Fast1, Tgif
Ptx1, Sf1, Six6
HYPOTHALAMUS
Sim1
Arnt2
Otp
Nkx2.1
Brn2
Sim2
Gata4
FoxB1
Gsh1
GnRH
GHRH
SS
TRH
VP
OT
CRH
Hesx1, Six3, Six6
Isl1, Isl2, Lhx2
Pax6 Dlx2
Pitx1 Pitx2
Lhx3 Lhx4
INTERMEDIATE PITUITARY
ANTERIOR PITUITARY
Tpit
Prop1
MSH
NeuroD1
Isl2
Gata2
Foxl2
Sf1
Egr1
Pit1
Gata2
POSTERIOR PITUITARY
ACTH
FSH
LH
TSH
PRL
GH
B
Oral ectoderm / pituitary placode E8-8.5
Rudimentary rathke's pouch E9.5
Definitive rathke's pouch E11
Adult pituitary gland E15-17
Extrinsic¹
Bmp4 Ttf1
Bmp4 Fgf8 Wnt5a
Fgf8 Wnt5a
Intrinsic²
Hesx1 Six3 Pax6 Ptx1 Isl1
Hesx1 Six3 Isl1 Lhx3 Lhx4 Ptx1 Wnt4 Nkx3.1
Hesx1 Six3 Isl1 Lhx4 Ptx1 Wnt4 Nkx3.1 Ptx2 αGSU Prop1 Bmp4 Bmp2 Bmp7 Msx1 Brn4 P-Frk
Lhx3 Lhx4 Sf1 GATA2 Pit1 αGSU
+
−
+

Figure 130.2 A, Transcription factors and signaling molecules involved in anterior pituitary development. See text for details. B, Transcription factors and signaling molecules expressed at different stages of murine pituitary development. *ACTH,* Adrenocorticotropic hormone; *CRH,* corticotropin-releasing hormone; *FSH,* follicle-stimulating hormone; *GH,* growth hormone; *GnRH,* gonadotropin-releasing hormone; *GHRH,* growth hormone–releasing hormone; *LH,* luteinizing hormone; *OT,* oxytocin; *PRL,* prolactin; *SS,* somatostatin; *TRH,* thyrotropin-releasing hormone; *TSH,* thyroid-stimulating hormone; *VP,* vasopressin.

factors, such as *HESX1, SOX2, SOX3,* and *OTX2*, are the most common genes to date implicated in the etiology of SOD.

Genetic mutations implicated in Kallmann syndrome (KS; a disorder characterized by hypogonadotropic hypogonadism, anosmia/hyposmia, cleft lip/palate, sensorineural hearing loss, dental anomalies, synkinesia, and renal abnormalities), such as *KAL1, FGFR1, PROKR2,* and *FGF8*, have also been recently identified in patients with SOD.[44,74,79]

HESX1: *HESX1* is a member of the paired-like class of homeodomain proteins, the initial activation of which may be dependent upon *LHX1* and *OTX2*. *Hesx1*, one of the earliest markers of the pituitary primordium, can be detected in the anterior forebrain from E7.5 to E8.5 and in Rathke's pouch from E8.5 to E135.[80] *Hesx1* is rapidly downregulated from E12 and becomes undetectable by E13.5.[81] *Hesx1* is a transcriptional repressor and its downregulation activates other downstream genes such as *Prop1*, suggesting that both function as opposing transcription factors. Targeted disruption of *Hesx1* in mice results in anophthalmia or microphthalmia and midline neurological defects (such as absent septum pellucidum and pituitary hypoplasia), reminiscent of SOD.[82] The first homozygous missense mutation reported in *HESX1* (p.R160C) was described in two siblings with SOD, born to consanguineous parents, who presented with a hypoplastic AP gland and panhypopituitarism, ACC, and ONH.[83] Since then, several homozygous and heterozygous *HESX1*

TABLE 130.3 Genes With Reported Pathogenic Variants Known to Cause Hypothalamo-Pituitary Disease

Gene With Reported Variants	Phenotype	Mode of Inheritance
ARNT2	CPHD, congenital abnormalities of the kidneys and urinary tract	Recessive
CDON	PSIS	Dominant
EIF2S3	GHD, TSHD, glucose dysregulation, MEHMO syndrome	X-linked
FGF8	HH/KS; HPE	Dominant
FGFR1	HH/KS, SOD	Dominant
FOXA2	CPHD, HI, childhood-onset diabetes, choroidal coloboma, biliary atresia (cardiac/endoderm-derived organ abnormalities)	Dominant
GH1	IGHD Type IA	Recessive
	IGHD Type IB	Recessive
	IGHD Type II	Dominant
GHRHR	IGHD Type IB	Recessive or dominant (rare)
GLI2	HPE, IGHD/CPHD, polydactyly, single central incisor	Dominant: haploinsufficiency
GPR161	PSIS	Recessive
HESX1	IGHD, CPHD, SOD	Dominant or recessive
IFT172	GHD, retinopathy, metaphyseal dysplasia, renal failure (ciliopathies)	Compound heterozygous
IGSF1	TSHD, hypoprolactinemia, transient GHD; usually with macroorchidism	X-linked
KAL1	HH/KS	X-linked
KCNQ1	GHD, maternally inherited gingival fibromatosis	Dominant
LHX3	CPHD, short neck with limited rotation	Recessive
LHX4	CPHD, Chiari malformation, cerebellar abnormalities, respiratory distress	Dominant or recessive
MAGEL2	Arthrogryposis with hypopituitarism (anterior and posterior pituitary hormone deficiencies, learning difficulties, scoliosis, dysmorphic features, central sleep apnoea, ONH	Maternally imprinted: paternal or de novo mutation
OTX2	IGHD, CPHD, SOD, anophthalmia/microphthalmia, retinal dystrophy	Dominant: haploinsufficiency or dominant negative
PCSK1	IAD, GHD, TSHD, DI, malabsorption	Dominant, compound heterozygous
PNPLA6	Oliver-McFarlane and Laurence-Moon syndrome; GH and gonadotrophin deficiencies	Recessive
POMC	IAD; early-onset obesity and red hair pigmentation	Recessive
POU1F1	GH, TSH, and ACTH deficiencies	Dominant or recessive
PROKR2	HH/KS	Recessive
PROP1	CPHD, pituitary tumors	Recessive
RAX	Anophthalmia/microphthalmia, CPHD, DI, and cleft palate	Recessive or compound heterozygous
RNPC3	IGHD	Recessive
ROBO1	PSIS	Dominant
SOX2	HH, anophthalmia/microphthalmia, learning difficulties, hypothalamo-pituitary tumors	Dominant
SOX3	GHD, CPHD, absent infundibulum, persistent craniopharyngeal canal	X-linked
TBC1D32	Craniofacial dysmorphism, cleft palate, post-axial polydactyly, hypopituitarism with EPP, Joubert-like findings on MRI	Recessive
TBL1X	TSHD, ASD	X-linked
TBX19	IAD	Recessive
TCF7L1	SOD	Dominant
TRHR	TSHD	Recessive
TSHB	TSHD	Recessive

ASD, Autism spectrum disorder; *CPHD,* combined pituitary hormone deficiency; *DI,* diabetes insipidus; *EPP,* ectopic posterior pituitary; *GHD,* growth hormone deficiency; *HH,* hypogonadotropic hypogonadism; *HI,* congenital hyperinsulinism; *HPE,* holoprosencephaly; *IAD,* isolated adrenocortical deficiency; *IGHD,* isolated growth hormone deficiency; *KS,* Kallmann syndrome; *MEHMO,* mental retardation, epileptic seizures, hypogonadism with hypogenitalism, microcephaly and obesity; *PSIS,* pituitary stalk interruption syndrome; *SOD,* septo-optic dysplasia; *TSHD,* thyroid-stimulating hormone deficiency.

mutations have been described. The majority of cases are sporadic and about 1% of patients with SOD present with genetic mutations in *HESX1*. In some patients, the penetrance is variable, suggesting the impact of additional genetic or environmental factors. There is no clear genotype-phenotype correlation, and clinical features range from isolated GHD to CPHD, associated in some cases with SOD and pituitary malformations.[84–87] MRI findings are variable, including a hypoplastic or aplastic AP and an ectopic PP.

SOX2: *SOX2* is a transcription factor member of the SRY-related HMG box (SOX) family. Expression of *SOX2* occurs at 4.5 to 9 weeks' gestation during human development within Rathke's pouch and is maintained throughout AP development as well as in the diencephalon. *SOX2* is expressed throughout the developing CNS as well as in sensory placodes, inner ear, cochlea, and in the developing lens, retina, and optic nerve.[88,89] *SOX2* plays an important role in the maintenance of pituitary progenitor cells and their differentiation into all hormone-producing lineages.[90] Murine *Sox2* loss-of-function mutations result in a phenotype that includes GH, TSH, and gonadotrophin deficiencies.[91] Heterozygous *de novo* mutations in humans have been observed in patients with hypogonadotropic hypogonadism, anophthalmia/microphthalmia (bilateral, often severe), small corpus callosum, hippocampal abnormalities, and variable learning difficulties.[92] Esophageal atresia has also been reported.[93] The pituitary phenotype occasionally includes GHD. Imaging of the hypothalamo-pituitary region can show hippocampal abnormalities, hypoplasia of the corpus callosum, hypothalamic hamartoma, and pituitary enlargement that is reminiscent of tumors.[94]

SOX3: *SOX3* is also a member of the SOX family. It is located on the X chromosome (Xp27.1) and is expressed along the full length of the CNS including the brain and spinal cord. *SOX3* dosage is critical for normal hypothalamic-pituitary development and both under- and overdosage of the gene can lead to hypopituitarism.[95,96] Male patients present with variable hypopituitarism (CPHD or idiopathic GHD) and infundibular hypoplasia, an ectopic/undescended PP and abnormalities of the corpus callosum. Learning difficulties are also frequently reported in these patients.[95] Patients with duplication of *SOX3* can present with GHD without other pituitary defecits.[97] Loss-of-function polyalanine expansions and gene deletions are associated with hypopituitarism including GH, TSH, ACTH, and gonadotrophin deficiencies. Neuroradiological features including a hypoplastic AP, absent pituitary stalk, and ectopic PP are associated with *SOX3* sequence variants or whole gene deletions/duplications. Persistence of the craniopharyngeal canal has been reported in association with a *SOX3* deletion.[98]

OTX2: Orthodenticle homeobox 2 (*OTX2*) is a transcription factor gene involved in the brain, eye, nose, and ear development.[99] *Otx2* is expressed from E10.5 to E14.5 in the ventral diencephalon, from E10.5 to E12.5 in Rathke's pouch. *Otx2* becomes undetectable at both sites from E16.5. *OTX2* regulates various transcription factors implicated in the human brain, eye, and pituitary development, including *RX1*, *PAX6*, *SIX3*, *LHX2*, *MITF*, *GBX2*, and *HESX1* in order to coordinate cell determination and differentiation. As *OTX2* is essential in retinal development, many patients with *OTX2* mutations also present with a variety of ocular abnormalities. In humans, heterozygous *OTX2* mutations or gene deletions have been implicated in the etiology of 2% to 3% of anophthalmia/microphthalmia syndromes.[100] There is no clear genotype-phenotype correlation, even among patients with the same mutation. The pituitary phenotype ranges from partial to complete GHD and brain imaging reveals a normal or hypoplastic AP, normal or ectopic PP, and Chiari malformation. Mutations in *OTX2* have rarely been reported in patients with no ocular involvement, who have CPHD and a small AP, with or without an undescended PP.[101]

PAX6: Paired box protein-6 (*PAX6*) is an early dorsal marker of early AP gland formation, and its expression is required for somatotrope, lactotrope, and thyrotrope development. It is also an important regulator of eye development, and heterozygous mutations in humans cause congenital eye anomalies.[102] *PAX6* mutations have been reported to be associated with impaired pituitary function (ACTH deficiency, hypogonadotropic hypogonadism, and GHD).[103]

BMP4: Bone morphogenetic protein 4 (Bmp4) is the first secreted molecule detected in the prospective infundibulum at E8.5 and is essential for Rathke's pouch formation and maintenance. It is expressed in the optic vesicle, in the diencephalic floor, and in the medial ganglionic eminence and in developing limbs. A study of over 200 individuals with eye abnormalities identified *BMP4* mutations in a familial case of anophthalmia, retinal dystrophy, brain malformation, and poly/syndactyly.[104] Deletions in *BMP4* were associated with bilateral anophthalmia/microphthalmia, in association with hypothyroidism, deafness, developmental delay, and cerebellar and pituitary abnormalities.

FGFR1: FGF receptor 1 (FGFR1) is a tyrosine kinase receptor that acts as a cell-surface receptor for FGF. It is the main receptor mediating FGF8 signaling. Mutations in *FGFR1* have previously been reported in patients with KS; more recently, *FGFR1* variants have been associated with CPHD, absent corpus callosum, SOD, and midline defects.[105]

ARNT2: Aryl hydrocarbon receptor nuclear translocator 2 (*ARNT2*) belongs to the HLH-PAS (Per-ARNT-Sim homology) subfamily of transcription factors. ARNT2 is a basic helix-loop-helix transcription factor that is critical for normal development of the paraventricular and supraoptic nuclei.

In the mouse, *Arnt2* is expressed in hypothalamus, neural retina, kidney, and urinary tract, and this expression pattern is recapitulated in humans. Mutations in *ARNT2* are associated with severe hypothalamo-pituitary insufficiency, including GH, TSH, and ACTH deficiencies as well as DI, together with progressive microcephaly, seizures, severe visual impairment, and severe learning difficulties and abnormalities of the renal and urinary tracts.[106] Webb et al. reported *ARNT2* mutations in six children from a large, consanguineous Saudi Arabian family, who presented with central DI and ACTH deficiency after birth: four had TSH deficiency, two had GHD, and of these two, the single affected male also had gonadotrophin deficiency with undescended testes. Pituitary and extrapituitary anomalies were found on neuroimaging: hypoplastic AP, thin pituitary stalk, hypoplastic frontal and temporal lobes, thin corpus callosum, and a global delay in brain myelination. Postretinal eye anomalies were also reported. All children had dysmorphic features with microcephaly, prominent forehead, deep-set eyes, a well-grooved philtrum, and retrognathia. Hydronephrosis, vesicoureteral reflux, and neurogenic bladder were also found. All carried a homozygous c.1373_1374dupTC *ARNT2* truncating variant.[106]

Holoprosencephaly (HPE). Mutations of the sonic hedgehog (*SHH*) signaling pathway components (*SHH, TGIF, ZIC2, PTCH1, GLI2*) and the transcription factors sine oculis–related homebox 3 (*SIX3*), teratocarcinoma-derived growth factor 1 (*TDGF1*), and *FOXH1/FAST1*, among several other genes, have been identified in patients with HPE, with and without hypothalamo-pituitary defects.[107–112]

***GLI2*:** This mediator of SHH signaling is expressed in the ventral diencephalon. *GLI2* induces *BMP4* and *FGF8* expression. It is also expressed in the oral ectoderm, inducing pituitary progenitors.[113] *GLI2* mutations are associated with HPE or HPE-like features, with craniofacial anomalies, pituitary abnormalities, and polydactyly.[107,114] Variants are frequently identified in patients with isolated hypopituitarism.

FGF8: Fibroblast growth factor 8 (*FGF8*) is a member of the FGF family of signaling molecules that are involved in pituitary organogenesis. *Fgf8* expression is seen in the infundibulum at E9.5, 1 day after the expression of *Bmp4*, and is important in midbrain development.[115] *FGF8* mutations are associated with KS and have more recently been described in association with recessive HPE, craniofacial defects, and hypothalamo-pituitary dysfunction.[44]

Wolfram Syndrome. ***WFS1:*** Mutations in *WFS1* are associated with autosomal recessive Wolfram syndrome includes DI, diabetes mellitus, optic atrophy, and sensorineural hearing loss.[116]

Hypopituitarism With Spine Abnormalities. ***LHX3:*** Expression of the LIM homeobox 3 (*LhX3*) gene, a member of the LIM class of homeodomain proteins, is detected early during AP development at E9.5 (Rathke's pouch, ventral hindbrain, and spinal cord) and persists in the mature pituitary gland. It is one of the earliest markers implicated in the anterior and intermediate lobes development and its expression plays an important role in the formation of gonadotrophs, thyrotrophs, somatotrophs, and lactotrophs.[117] Mice with homozygous mutations of *Lhx3* die soon after birth as a result of pituitary aplasia, whereas those with heterozygous mutations have no abnormalities. In humans, several homozygous or compound heterozygous *LHX3* mutations and a heterozygous variant have been reported to date. Patients with *LHX3* mutations commonly present with GH, TSH, and FSH/LH deficiencies, while ACTH deficiency is reported in 50% of cases.[118] The phenotype varies depending on which part of the gene is affected.[119] If the mutation affects the entire gene or protein, the LIM domains or the homeodomain, patients will present with syndromes involving the nervous and skeletal systems,[120] whereas if the mutation affects the carboxyl terminus of LHX3 protein alone, isolated pituitary dysfunction will be present. *LHX3* is also required for inner ear development. Extrapituitary phenotypes include a short neck with abnormal head and neck rotation (70% of cases), vertebral abnormalities (50% of cases) including a rigid cervical spine, flattened lumbar vertebrae, thoracic kyphosis and progressive scoliosis, and variable degrees of sensorineural hearing loss (50% of cases).[121,122] Developmental delay and learning difficulties have also been reported in nearly 40% of the patients. Two patients have been reported to have respiratory distress. Heterozygous family members are largely unaffected, although a recent publication has described a mild limitation of neck movement in a heterozygous carrier.[123] MRI findings vary from normal (10%) to aplasia or hypoplasia of the AP, a hypointensity suggestive of microadenoma, and enlargement with a hyperintense signal.

Hypopituitarism With Cerebellar Abnormalities. ***LHX4:*** LIM homeobox 4 (*LHX4*) gene is closely related to *Lhx3* and is expressed in the developing brain and spinal cord.[124] *Lhx4* is also detected during early stages (E9.5 in Rathke's pouch and E12.5 in the anterior part of the pituitary) and is subsequently found in the future anterior lobe, with a decrease in expression by E15.5. The AP in patients with *LHX4* mutations is hypoplastic, containing all the differentiated cell types but in reduced numbers. Other brain abnormalities can also be present such as an ectopic PP and a hypoplastic sella turcica as well as corpus callosum hypoplasia or Chiari syndrome.[125,126] Several sporadic or familial *LHX4* mutations have been reported to date. The phenotype in patients ranges from isolated GHD to complete panhypopituitarism.[127] Of note, four patients also presented with respiratory distress, and one presented with a cardiac defect.[128] Several of the variants are variably penetrant, although the underlying mechanism remains to be established. A lethal neonatal phenotype (severe panhypopituitarism associated with AP aplasia and ectopic PP, mild facial hypoplasia, undescended testes, and severe respiratory distress) has been recently described, secondary to a homozygous mutation.[127]

Pituitary Stalk Interruption Syndrome (PSIS). PSIS is a congenital defect of the pituitary gland characterized by the triad of (1) a thin pituitary stalk, (2) an ectopic PP gland, and (3) hypoplasia or aplasia of the AP gland identified by neuroimaging. Patients with PSIS may present with either isolated GHD or CPHD.[129] Genetic alterations in *HESX1*, *LHX4*, *OTX2*, *SOX3*, and *PROKR2* have been reported in patients with PSIS, and among others.

PROKR2: PROKR2, a G-protein–coupled receptor essential for proper neuronal migration and angiogenesis, is involved in sex development and olfactory bulb development in mice. In humans, patients with mutations in *PROKR2* can present with hypogonadotropic hypogonadism or KS. More recently, mutations have been associated with variably penetrant hypopituitarism, including SOD.[129,130]

OTX2: As described previously, *OTX2* gene defects have also been reported in patients with no ocular abnormalities.[131]

GPR161: This orphan member of the G-protein–coupled receptor family has been recently identified in patients with PSIS. *GPR161* is widely expressed in both mouse and human during the early stages of embryogenesis including the neural folds, the pituitary and the hypothalamus. It is a key negative regulator of the SHH pathway, the pituitary target of which is GLI2. It has been suggested that gain-of-function mutations of *GPR161* could lead to abnormal pituitary development by repressing the SHH pathway. A homozygous missense mutation p.L19Q in *GPR161* has been recently described in two female siblings with short stature due to GHD associated with AP hypoplasia and an empty sella with an ectopic PP. Additional clinical features included fifth finger brachydactyly, congenital alopecia, and ptosis of the left eye.[132]

Robo: Roundabout genes (*Robo*), first identified in Drosophila, encode Robo proteins that belong to the immunoglobulin superfamily of proteins secreted by midline glial cells and play a role in axon guidance. Whole-exome sequencing recently identified *ROBO1* variants in 25 patients with PSIS; however, functional studies were not performed to support causality.[133]

Deficient AP Function With Variable Immune Deficiency (DAVID) Syndrome. ***NFKB2:*** This belongs to the NF-κB family, which consists of a collection of evolutionarily conserved transcription factors involved primarily in development including the AP, immunity, and oncogenesis. Patients with *NFKB2* mutation present with deficit in AP function and common variable immune deficiency, a novel disorder called DAVID syndrome.[134] However, the precise mechanism underlying endocrine deficits remains largely unclear.

Axenfeld-Rieger Syndrome. ***PITX2:*** This homeobox gene is detected in the stomodeum at E8, Rathke's pouch at E10.5, and 2 days later in the anterior and intermediate lobes. Patients affected with *PITX2* mutations present with Axenfeld-Rieger syndrome, which is characterized by eye, craniofacial, dental, cardiac, and umbilical anomalies.[135] *Pitx-2* null mice manifest pituitary hypoplasia and decreased *GHRH receptor, GH, FSH, LH,* and *TSH* gene expression.[136]

Pallister Hall Syndrome. ***GLI3:*** This is a component of the SHH signaling pathway. Haploinsufficiency of *GLI3* results in Pallister Hall syndrome, which is associated with polydactyly, hypothalamic disorganization, hypothalamic hamartoma, and hypopituitarism.[137,138]

CHARGE Syndrome. ***CHD7***: CHARGE syndrome is a rare autosomal-dominant disorder that affects multiple organs. It is characterized by ocular coloboma, cardiac defects, choanal atresia, growth and developmental delay, and ear abnormalities. Ectopic PP and hypopituitarism in patients with CHARGE syndrome have been reported in association with two novel *CHD7* variants.[139]

Other CPHD. ***PCSK1***: In a study of 13 children with proprotein convertase 1/3 (PC1/3) deficiency, an autosomal-recessive disorder caused by rare mutations in the *PCSK1* gene, obesity and severe diarrhea were associated with central DI, GHD, ACTH deficiency, hypothyroidism, and primary male hypogonadism.[140]

IGSF1: *IGSF1* is located at Xq26 and is expressed in Rathke's pouch, pituitary gland (present in GH, prolactin, and TSH-secreting cells), and testis. *IGSF1* mutations (loss-of-function or deletions) cause an X-linked syndrome. Male patients with mutations in *IGSF1* present with a characteristic phenotype, which consists of congenital central hypothyroidism, delayed puberty, and adult macroorchidism. PRL and/or GH deficiencies have also been reported in some cases. Some female patients with heterozygous mutations in *IGFS1* present with central congenital hypothyroidism, PRL deficiency, and delayed puberty.[141,142] Recently, *IGSF1* mutations have been associated with acromegaloid features with elevated concentrations of GH and IGF-1 in humans and mice.[143]

Other: More recently, mutations have been described in *FOXA2* (CPHD, eye abnormalities, and congenital hyperinsulinism [CHI]), *EIF2S3* (CPHD with glucose dysregulation), *TCF7L1* (SOD), *KCNQ1* (CPHD, maternally inherited gingival fibromatosis), and *RAX* (anophthalmia, CPHD, and cleft lip and palate). Mutations in *MAGEL2* have been associated with both anterior and PP hormone deficiencies associated with a range of clinical features, including arthrogryposis within the Schaaf-Yang and Chitayat Hall syndrome phenotypes. Mutations in *TBC1D32* have also been associated with CH within a ciliopathy phenotype.[74] Mutations in *BRAF* have also recently been identified in patients with cardio-facio-cutaneous syndrome and SOD.[144]

Syndromic Forms of Isolated GnRH Deficiency. This includes normosmic isolated GnRH deficiency causing hypogonadotropic hypogonadism, and KS. A number of genes have been reported to date in KS, including *KAL1, CCDC141, FEZF1, IL17RD, SEMA3A, SEMA3E*, and *SOX10*.[145–148] Pathogenic variants in *CHD7, FGF8, FGF17, FGFR1, DUSP6, SPRY4 and FLRT3, HS6ST1, NSMF (NELF), PROK2, PROKR2*, and *WDR11* cause both KS and normosmic hypogonadotropic hypogonadism.[148,149]

Variants in PROKR2 and PROK2, encoding prokineticin receptor-2 and prokineticin-2, respectively, have been identified in approximately 9% of patients with KS.[150] PROK1 and PROK2 act through their G-protein–coupled receptors, PROKR1 and PROKR2, which are expressed in the olfactory bulbs. PROK2 functions as a chemoattractant for neuronal progenitors, which follow a rostral migratory stream. *Prokr2*$^{-/-}$ mice have reduced LH in the pituitary, small gonads, and abnormal olfactory bulb formation but *Prokr2* seems to be dispensable for normal pituitary formation.[130] Variations in *PROKR2* have also been associated with hypopituitarism, PSIS, and SOD, but are unlikely to be causative in isolation and likely only contribute to the phenotype, including KS, in combination with other genetic mutations and/or environmental factors.

Mutations in a number of genes have also been identified in association with normosmic hypogonadotrophic hypogonadism, and these include *GNRH1* and its receptor *GnRHR*, *Kisspeptin* (*KISS1*) and its receptor *KISS1R*, and *Neurokinin B* (encoded by *Tachykinin 3* (*TAC3*)) and its G-protein–coupled receptor encoded by *TACR3*.[151–154]

Recently mutations in genes such as *RNF216, OTUD4, STUB1*, and *PNPLA6* have been found in syndromic forms of GnRH deficiency that have associated features such as ataxia and dementia as part of Gordon Holmes syndrome (*RNF216, OTUD4, STUB1*) and Boucher-Neuhauser syndrome (*PNPLA6*).[155–157]

Nonsyndromic CPHD. Mutations in *PROP1* and *POU1F1* constitute the main genetic cause found in patients with nonsyndromic GHD or CPHD.

PROP1: *PROP1* (prophet of PIT-1) is a member of the paired-like family of homeodomain transcription factors and is the earliest expressed pituitary-specific transcription factor. *Prop1* is detected within Rathke's pouch by E10, peaks at E12, and disappears by E15.5.[158] *Prop1* can act as both a transcriptional repressor (for *Hesx1* expression) or a transcriptional activator (for *Pou1f1*).[158,159] Mice with *Prop1* overexpression present with delayed puberty as a consequence of a delay in the differentiation of gonadotrophs.[160] The most frequent genetic cause of CPHD is recessive mutations in *PROP1*.[161–163] The most common of these is a 2-base pair deletion within exon 2, which results in a frameshift at codon 101 and introduction of a termination codon at position 109.[164] Recessive *PROP1* mutations are associated with GH, TSH, PRL, and gonadotrophin deficiencies, which can vary in onset and severity. It is therefore important to have ongoing clinical surveillance. GHD and growth delay usually present during the first years of life in patients with *PROP1* mutations; however, there is a case report of a patient who had normal growth and achieved a normal adult height without GH treatment.[165] Both TSH and gonadotropin deficiency can appear at birth or later in life.[166,167] Some patients may present with micropenis and undescended testes at birth and some others with delayed puberty. Spontaneous puberty can also be seen.[168,169] The majority of patients do not have ACTH deficiency during the first years of life but this can evolve later and ongoing surveillance is needed.[168,170] Neuroimaging shows variable pituitary morphology. The commonest finding is a normal pituitary stalk and posterior lobe with a small or normal AP gland. An enlarged AP gland has also been described with waxing and waning of size and posterior regression.[171]

POU1F1: *POU1F1* (previously known as *PIT-1*) is a member of the POU family. It is expressed later during pituitary organogenesis (E14.5) and persists during adulthood.[172] *POU1F1* plays a crucial role in the regulation of the genes encoding GH, PRL, and TSH-β. Patients with *POU1F1* mutations tend to present first with GH and prolactin deficiencies during the first years of life, whereas TSH deficiency tends to present later.[173] Gonadotroph and corticotroph axes usually remain functional. *POU1F1* mutations are mainly recessive. However, dominant mutations have been described, the most frequent being p.R271W mutation.[174] MRI reveals a normal or small AP gland; the PP and infundibulum are normal and no midline abnormalities have been reported.[173,175]

Isolated Pituitary Hormone Deficiencies. Rare isolated pituitary hormone deficiencies have been associated with mutations in the respective hypothalamic-releasing hormones or the releasing hormone receptors (e.g., familial GHD due to *GHRH-R* mutations, TSH deficiency due to *TRHR* mutations, and gonadotropin deficiency due to *GNRHR* mutations).[152,176,177]

Heterozygous mutations in *HESX1* have been identified in patients with isolated GHD.[82,178] The phenotype of those patients with heterozygous mutations in *HESX1* tends to be milder presenting with isolated GHD and an ectopic or undescended PP, although midline forebrain abnormalities can also be seen. Patients with duplication of *SOX3* can present with isolated GHD without other pituitary deficiencies.[97]

Congenital isolated ACTH deficiency is mainly due to recessive mutations in *TBX19* (formerly *TPIT*), which are responsible for approximately 65% of the cases of isolated ACTH deficiency diagnosed during the first month of life.[179] Neonates present with severe hypoglycemia that can result in seizures and prolonged cholestatic jaundice. Biochemical findings include low basal ACTH and cortisol concentrations and a poor ACTH response to CRH. It is extremely important to diagnose this as it can be a cause of neonatal death if replacement treatment is not started promptly.[180]

CCH may be more common than previously thought, with up to 1 in 16,000 neonates reported to be affected in the Netherlands. Rare recessive biallelic inactivating mutations in *TRHR* have been reported in three affected individuals from two unrelated pedigrees with CCH.[176,181] More recently, a novel homozygous mutation that decreases TRH affinity has been identified in an overweight patient with CCH and normal stature.[182] The X-linked transducing β-like protein 1 (*TBL1X*) gene is a component of the thyroid receptor-corepressor complex. Like *IGSF1*, *TBL1X* is associated with an X-linked form of TSH deficiency.[183,184]

KEY POINTS

- The pituitary consists of three lobes of dual embryologic origin. Pituitary hormone secretion is evident from 8 weeks' gestation.
- Mutations in genes involved in early development tend to be part of a syndrome that includes extrapituitary defects and midline abnormalities, as well as congenital hypopituitarism. Mutations implicated in later developmental stages result in variable phenotypes of CPHD without any extrapituitary defects.

Fetal Adrenal

The fetal adrenal is a remarkable organ, but its uniqueness was largely overlooked until 1911 when its large fetal zone (FZ) was first described as an immature cortex that undergoes involution postnatally.[185] The fetal adrenal is rapidly emerging as a key component of the fetoplacental unit in pregnancy and understanding its development may help to elucidate the pathophysiology of parturition, as well as furthering insight into many congenital adrenal disorders.

Adrenal Development. The adrenal appears as the bipotential adrenogonadal primordium (AGP) at 28 to 30 days post conception (dpc) in humans[186] due to expression of the transcription factor, steroidogenic factor 1 (SF1, NR5A1). The adrenal cortex is derived from a thickening of the intermediate mesoderm, in contrast to the adrenal medulla, which derives from the ectoderm. This region, known as the "gonadal ridge," contains AGP cells that give rise to the steroidogenic cells of the adrenal gland as well as those of the gonad. Cells expressing the highest levels of *NR5A1* are destined to become adrenal tissue and migrate retroperitoneally to the upper pole of the mesonephros to form the adrenal primordium at 33 dpc. At about 48 dpc, sympathetic cells derived from the neural crest infiltrate this area. These cells coalesce and differentiate after birth into the catecholamine-producing chromaffin cells of the adrenal medulla. Following the neural crest invasion, the adrenal primordium becomes encapsulated. The cells of the developing adrenal cortex segregate into two distinct zones by 50 to 52 dpc. The large, inner eosinophilic cells of the FZ have high levels of expression of steroidogenic enzymes, whereas the outer definitive zone (DZ) cells have lower levels of expression. A third transitional zone (TZ) becomes identifiable from about 12 weeks' gestation. The TZ lies between the DZ and FZ and has histological appearances of both.[187]

As pregnancy progresses the fetal adrenal grows tremendously, due largely to an increase in the FZ, which accounts for 80% to 90% of adrenal mass by midgestation,[188,189] with a relative size 10- to 20-fold that of the adult adrenal.[18,190,191] Growth is paralleled by increasing functional development. The fetal adrenal produces C_{19} androgens such as DHEA and DHEA sulfate (DHEAS) for the majority of pregnancy, with brief periods of cortisol production. These steroids maintain intrauterine homeostasis and prepare the fetus for adaptation to extrauterine life. FZ cells robustly express cytochrome P450 17 alpha (P450c17α), a bifunctional enzyme with both 17-hydroxylase and 17,20 lyase activities, that converts pregnenolone to DHEA. DHEA is sulfated to DHEAS, which is subsequently aromatized by the placenta in order to produce estrogens.[187] The FZ is identified in humans and a few higher primates only and the adrenal gland undergoes rapid involution postnatally, largely due to regression of the FZ. Immediately after birth, the FZ involutes and remodels by a process involving apoptosis of cells with a concomitant decrease in androgen secretion.[186] The weight of the adrenal drops by 50% within 2 weeks after birth,[192] and the FZ is absent by 6 months in the majority of cases. Whether the timing of FZ involution is determined by birth or gestational age remains controversial.[187,193] Postnatally, the DZ and TZ give rise to the zona glomerulosa and zona fasciculate, respectively, of the definitive adrenal cortex. Cells with zona reticularis morphology are detectable from around 3 years of age until a continuous ZR forms at around 6 years of age and androgen synthesis recommences, a stage referred to as adrenarche.[194]

Genetic Regulation of Adrenal Development. Fetal adrenal cortical development is programmed by a cascade of genetic transcription factors (Fig. 130.3) (e.g., ODD1, SALL1, FOXD1/FOXD2, WT1, NR5A1, NR0B1), coregulators (e.g., CITED2), signaling molecules (e.g., SHH/GLI3, WNT3/WNT4/WNT11), matrix proteins (e.g., SPARC), and regulators of telomerase activity (e.g., ACD). *Wt1* knockout mice have renal and gonadal abnormalities and lack adrenal glands.[195] Several other transcription factors, notably Pbx1, Lim1, and Wnt4, are involved earlier in the complex genetic cascade programming adrenal gland organogenesis from the coelomic epithelium and urogenital ridge.[195,196] *Pbx1* knockout mice die *in utero* with multiple organ defects that include adrenal agenesis and impaired testis development while *Wnt4* gene disruption leads to abnormal adrenal development, masculinization of XX females, and Müllerian-duct agenesis.

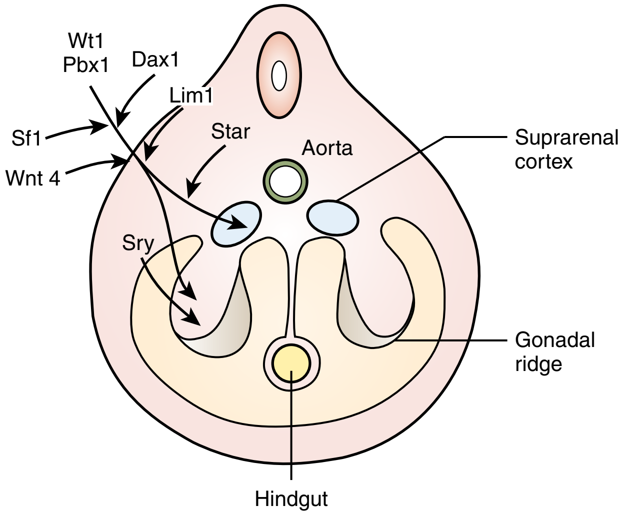


Figure 130.3 Cartoon showing a cross section of a 5-week human embryo, with location of the adrenal primordia (suprarenal cortices) and gonadal ridges. The most important genes known to be implicated in adrenal development and function are indicated. SF1 is required for adrenal and gonadal development, while SRY is the single critical regulator of testicular embryogenesis. DAX1 gene inactivation leads to adrenal hypoplasia. The steroidogenic acute regulatory protein *(StAR)* is the rate-limiting factor for adrenal steroidogenesis.

The most significant transcription factors for early AFP development are NR5A1 and NR0B1.[195,197,198] As well as expression in the adrenal cortex, these genes are expressed in the gonads, hypothalamus, and pituitary. *Nr5a1* gene knockout mice manifest adrenal and gonadal agenesis, gonadotropin deficiency, and absence of the hypothalamic ventromedial nucleus.[195] In the absence of *NR5A1* expression, the adrenal does not form.[186] Severe disruption of *NR5A1* in humans can cause adrenal dysfunction, although most pathogenic variants cause impaired testicular development and Leydig cell function. Heterozygous loss-of-function mutations in *NR5A1* in humans result in 46,XY and 46,XX DSD, complete gonadal dysgenesis and primary adrenal failure.[195]

NR5A1 upregulates *NR0B1*, which is itself a negative regulator of *NR5A1* transcriptional activity, and therefore steroidogenesis. Inactivating *Nr0b1/NR0B1* mutations in mice and humans are associated with gonadotropin deficiency and adrenal hypoplasia.[195] Pathogenic variants or deletions of *NR0B1* are well established as the cause of X-lined AHC.

Signaling Pathways in Adrenal Development. SHH, a member of the vertebrate hedgehog (Hh) family of secreted ligands, performs a number of crucial roles in embryogenesis. *SHH* is expressed in AP subcapsular cells at a later stage than *NR5A1* and *NR0B1*.[199–201] *Shh* expression marks cortical progenitors and Shh-expressing cells are thought to give rise to all steroidogenic cells in cortical zones,[199,200] and *Shh* mutant mice have a thin capsule and a small cortex.

FGF signaling controls early developmental processes[202] such as anterior/posterior patterning and organogenesis and interacts with NR5A1 and SHH signaling. FGFs bind to four receptors, FGFR1 to 4. *Fgfr2* and *Fgfr4* are expressed in the developing adrenal cortex, and deletion of *Fgfr2-IIIb* leads to reduced adrenal growth and impaired expression of NR5A1 and steroidogenic enzymes.[203]

Epidermal growth factor (EGF) stimulates proliferation of both the FZ and DZ. The fetal adrenal expresses high levels of IGF-2 mRNA and protein, which are responsive to ACTH.[204] IGF-2 augments ACTH-stimulated expression of steroidogenic enzymes and stimulates steroid hormone production in fetal adrenal cortical cells.

CDKN1C (Cyclin-dependent kinase inhibitor 1 C, *P57KIP2*) is a paternally imprinted gene, located on chromosome 11p.15 and encoding for the CDKN1C protein, an inhibitor of cell cycle progression. Variations in *CDKN1C* or its genomic imprinting can lead to adrenal pathology. Loss-of-function of CDKN1C results in Beckwith-Wiedemann syndrome (BWS), an overgrowth syndrome with increased susceptibility to adrenal carcinoma. IMAGe (intrauterine growth restriction, metaphyseal dysplasia, AHC, and genital anomalies) syndrome is a rare multisystem disorder that mirrors the features of BWS and is caused by gain-of-function mutations in *CDKN1C*. The resultant adrenal insufficiency can be life threatening shortly after birth secondary to salt loss, hypoglycemia, and shock due to loss of mineralocorticoid and glucocorticoid synthesis.[205]

Fetal Adrenal Steroidogenesis. The maintenance of intrauterine homeostasis and fetal maturation depends on fetal adrenal steroidogenesis. Cholesterol, the major substrate for fetal adrenal steroidogenesis, is derived from circulating low-density lipoprotein (LDL) and from *de novo* adrenal synthesis. LDL cholesterol, largely of fetal liver and testicular origin, contributes 70% of the total. The FZ contains more LDL binding sites and manifests a greater rate of *de novo* cholesterol synthesis than does the DZ, in keeping with its greater steroidogenic activity. Cholesterol availability is tightly controlled by the expression of steroidogenic acute regulatory protein (StAR), which regulates the transport of cholesterol to the inner mitochondrial membrane. StAR expression therefore is a rate-limiting factor in adrenal steroidogenesis.[206] *StAR* knockout mice manifest glucocorticoid and mineralocorticoid deficiency and female genitalia in XY animals. In humans, inactivating *StAR* mutations cause adrenal hypoplasia and adrenal insufficiency.[206] Recently, 46,XY males with normal male genitalia have also been described in association with *StAR* mutations.[206,207]

Sphingosine-1-phosphate lyase (SGPL1) is an endoplasmic reticulum enzyme, ubiquitously expressed in tissues, which mediates cleavage of the lipid-signaling molecule sphingosine-1-phosphate. Several recent studies report mutations in *SGPL1* resulting in primary adrenal insufficiency and steroid-resistant nephrotic syndrome.[208,209]

The fetal adrenal expresses the same five steroidogenic apoenzymes as the adult gland: two microsomal enzymes with 17-hydroxylase and 17,20-desmolase (CYP17 or P450c17) and 21-hydroxylase (CYP21A2 or P450c21) activities, plus two mitochondrial cytochrome P450 enzymes providing cholesterol side-chain cleavage (CYP11A1 or P450scc) and C11/C18 hydroxylation of the parent steroid structure (CYP11B1/CYP11B2 or P450c11/aldosterone synthase).[190,210] A fifth enzyme, expressed by the smooth endoplasmic reticulum, exhibits both 3βHSD and Δ4,Δ5-isomerase activities.[190,210] Quantitative differences in the relative activities of these enzymes are found, however, between cells derived from the FZ versus the DZ, and these differences are largely due to regulated steroidogenic gene transcription.[190] The FZ has relatively high steroid sulfotransferase (*SULT2A1*) activity, and because of the low 3βHSD and high sulfotransferase activities, the major steroid products of the fetal adrenal are DHEA, DHEAS, pregnenolone sulfate, several Δ5 3β-hydroxysteroids, and limited amounts of Δ5 3-ketosteroids, including cortisol and aldosterone.[190,210]

The paradox of human fetal adrenal function is that steroidogenesis is programmed largely to the production of inactive products.[190] The gland is maximally stimulated to maintain fetal cortisol levels and ACTH feedback homeostasis but is programmed by the steroidogenic enzyme expression pattern (e.g., relative 3βHSD deficiency) to produce inactive DHEA and pregnenolone and their sulfate conjugates. Much of the DHEA is converted to 16-hydroxy-DHEAS by fetal adrenal and liver. This programming is designed to provide DHEA substrate for placental estrone and estradiol production; 16-hydroxy-DHEA undergoes metabolism to estriol in the placenta. Fetal DHEAS production and maternal estriol concentrations increase progressively to term; DHEAS production approximates 200 mg/day near term.[204] In the anencephalic fetus, placental estrogen production is reduced to about 10% of normal, and estrogen is known to suppress FZ growth during the second half of primate pregnancy. It is proposed that estrogen feedback regulates the secretion of fetal adrenal DHEA to maintain normal fetal-placental function and development.[211]

Glucocorticoids are synthesized in the TZ in the first trimester (6 to 7 weeks' gestation) due to transient expression of 3βHSD2, and this is maximal between 8 and 9 months of gestation.[212] The hypothalamic-pituitary-adrenal (HPA) axis is sensitive to glucocorticoid-mediated feedback at this time; 46XX fetuses with steroidogenic defects such as CYP21A2 or CYP11B1 lack cortisol and have an elevated ACTH drive that results in excess production of fetal androgens at a time when the genital and scrotal folds are sensitive to androgen exposure, resulting in virilization of female genitalia.[212] Cortisol can retard fetal growth and is converted to inactive cortisone by the enzyme 11βHSD2. In this way, the fetus is protected from the high concentrations of cortisol *in utero*. The concentration of circulating cortisone in the fetus at midgestation is 4- to 5-fold higher than cortisol concentration (Fig. 130.4). In the third trimester, expression and activity of 11βHSD2 decreases and selected fetal tissues, including lung and liver, express 11βHSD1 so that glucocorticoids become available to promote maturation of organs in preparation for

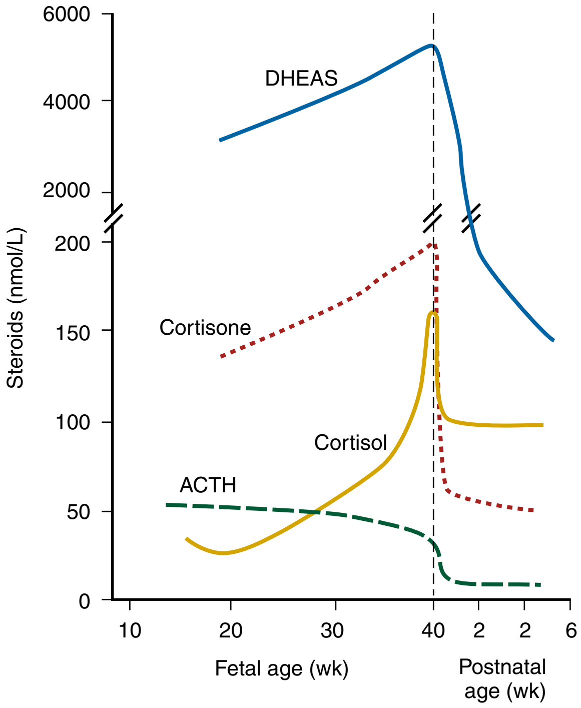


Figure 130.4 The pattern of maturation of serum cortisol, cortisone, and dehydroepiandrosterone *(DHEA)* in the human fetus during midgestation and the perinatal period. *ACTH,* Adrenocorticotropic hormone.

transition to extrauterine life.[190,213] Near-term fetal cortisol production per unit body weight is similar to that in the adult.[190] About two-thirds of fetal cortisol is derived from the fetal adrenal glands and one-third from placental transfer.[190] The metabolic clearance of cortisol in the fetus is rapid; 80% is oxidized in fetal tissues or placenta to cortisone or further metabolites.[190] Glucocorticoid receptors (GRs) are present at birth and are probably present at midgestation in most tissues, including placenta, lung, brain, liver, and gut.[190,210,214,215] Mice lacking GR function manifest enlarged and disorganized adrenal cortices, adrenal medullary atrophy, lung hypoplasia, and defective gluconeogenesis.[214]

In developing mammalian fetal models, all components of the renin-angiotensin system (RAS), including renin, angiotensinogen, angiotensin, angiotensin-converting enzyme, and angiotensin II receptor, are present. The genes are generally expressed at higher levels in fetal than in postnatal life.[190] Both ACTH and angiotensin II receptors (AT_1 and AT_2) are present on fetal adrenal cells early in gestation to stimulate adrenal steroid production.[190] AT_1 receptor mRNA expression in the fetal sheep kidney is low, early in gestation, increases in the latter third of pregnancy, and decreases postnatally; AT_2 mRNA levels, in contrast, are high at midgestation and decrease during the third trimester.[216] These changes reflect growth factor–mediated changes in AT-containing cells in various tissues. Hormonal factors modulate fetal renal AT gene expression in sheep; angiotensin II suppresses both AT_1 and AT_2; cortisol increases AT1 gene expression in kidneys and lungs.[216] The human fetal adrenal is able to produce aldosterone near term. Mineralocorticoid receptors (MRs) are present in fetal tissues (including fetal kidney, skin, hair follicles, trachea, and bronchioles) from 12 to 16 weeks' gestation.[217] The role of these MRs is unclear. MR knockout mice appear normal at birth but demonstrate defects in mineralocorticoid and RAS functions postnatally.[218] The role of the fetal RAS is not clear; rather than modulating renal sodium excretion through aldosterone, it may maintain renal excretion of salt and water into amniotic fluid to prevent oligohydramnios.[219]

Hormonal Regulation of Adrenal Development. A major stimulus to fetal adrenal functional development is fetal pituitary ACTH.[190,204] ACTH is a 39-amino acid peptide secreted from the AP under the control of CRH. ACTH binds to the melanocortin 2 receptor, MC2R, and exerts downstream effects via the cAMP and Ras/MEK/ERK signaling pathways. Although placental hCG may support early adrenal growth, the involution of the adrenal gland that occurs after 15 weeks in the anencephalic fetus suggests a role for pituitary-derived factors.[198] At midgestation, fetal plasma ACTH concentrations average 250 pg/mL, and although concentrations decline toward term, fetal ACTH concentrations exceed those later in life.[204] It does appear that ACTH-dependent and ACTH-independent mechanisms regulate fetal adrenal steroidogenesis.

CRH protein has been demonstrated in fetal baboon pituitary, adrenal, liver, kidney, and lung tissues during the last third of gestation. Concentration is highest in the pituitary (300 to 500 pg/mg protein); concentrations in adrenal, lung, liver, and kidney tissues average 20 to 30 and 5 to 10 pg/mg protein.[220] Circulating CRH concentration is high in the fetus and probably results largely from placental production.[220,221] Maternal concentration of CRH is elevated during the last trimester of gestation and reaches values of 0.5 to 1 nmol/L at term; normal values in nonpregnant women are <0.01 nmol/L.[222] At the end of pregnancy, there is increased bioavailability of CRH due to a fall in its binding protein, and this results in an exponential increase in maternal CRH from 35 weeks' gestation to term.[40,223] This placental CRH is bioactive, and concentrations correlate with maternal cortisol concentrations, suggesting that this circulating placental CRH plays a role in stimulating maternal corticotropin release. Fetal plasma CRH concentration at term, however, is approximately 0.03 nmol/L and, relative to the presumably high concentration in pituitary portal blood, probably has little role in modulating fetal corticotropin release. Midgestation fetal plasma ACTH concentration averages about 55 pmol/L (250 pg/mL), a concentration that maximally stimulates fetal adrenal steroidogenesis. Concentration is higher throughout gestation than in postnatal life, although it falls near term.[190,204] AVP and catecholamines are also significant stimuli for fetal ACTH secretion.[224]

The corticotropin feedback control system matures progressively during the second half of gestation and the early neonatal period. Dexamethasone can suppress the human fetal pituitary-adrenal axis at term but not at 18 to 20 weeks of gestation.[190] In the fetal sheep, hypothalamic and pituitary GRs are present at midgestation and corticotropin suppressibility can be demonstrated by the midpoint of the third trimester of gestation.[215] The number of GRs in the sheep fetal hypothalamus increases at term at the time of increasing glucocorticoid concentrations, suggesting that some process in the fetus allows the normal autoregulation of GRs to be overridden at term.[225]

Neonatal Adrenal Insufficiency. Adrenal insufficiency may be secondary to ACTH deficiency or primary due to abnormal adrenal development or resistance to ACTH.

The mature ACTH peptide is cleaved from the larger precursor molecule, POMC, together with other small peptides such as β-endorphin and α- and β-melanocyte–stimulating hormone (MSH). Given that mineralocorticoid secretion is largely independent of ACTH secretion, it is preserved and salt loss is therefore unusual. Nevertheless, mineralocorticoid deficiency can be an issue in some patients with ACTH insensitivity. ACTH deficiency can be part of MPHD or may be due to isolated ACTH deficiency. A number of single gene defects have now been associated with MPHD that include ACTH deficiency (e.g., mutations in *HESX1*, *SOX3*, *LHX3*, *LHX4*, *PROP1*, *ARNT2*).[178] Occasionally, ACTH insufficiency may not be present at the time of diagnosis but may develop progressively with time.

Isolated ACTH deficiency can be due to recessive mutations in the T-box factor *TPIT* (*TBX19*). TPIT is required for the specification, maturation, and maintenance of both precorticotrope and premelanotrope populations. It is also required to activate the expression of POMC in conjunction with the transcription factor PTX1. *TPIT* deficiency in humans leads to severe, early-onset isolated ACTH deficiency with profound hypoglycemia and prolonged jaundice and may result in sudden neonatal death.[226]

POMC/PC1 mutations lead to ACTH deficiency combined with pale skin, red hair, diarrhea, and obesity. Recently, *POMC* mutations have been found that result in immunoreactive but bioinactive ACTH and MSH; patients manifest cortisol deficiency, obesity, and red hair.[227]

ACTH resistance leads to isolated glucocorticoid deficiency (although salt loss can occasionally occur), hyperpigmentation and markedly elevated concentrations of ACTH.[228,229] ACTH resistance can occur in a number of well-defined entities such as defects in the ACTH receptor (*MC2R*, familial glucocorticoid deficiency [FGD] type 1), MC2R accessory protein (*MRAP*, FGD type 2), or as part of the Triple A syndrome (Alacrimia, Achalasia, Addison; also known as Allgrove syndrome, due to defects in ALADIN/AAAS).[230] Mutations in NNT (nicotinamide nucleotide transhydrogenase), GPX1 (glutathione peroxidase 1), and MCM4 (mini-chromosome maintenance-deficient-4 homolog) have also been found as a cause of FGD.[231–233] Mutations in MCM4 were found in an Irish traveler community, and result in late onset, less severe glucocorticoid deficiency, short stature and NK cell deficiency, due to increased chromosomal breakage. Mutations in GPX1and NNT, involved in detoxification of reactive oxygen species (ROS), can also result in FGD.[234]

Primary adrenal failure may be due to AHC or various forms of CAH. AHC results in severe salt-losing primary adrenal failure in early infancy or childhood, although milder, delayed onset forms of the condition exist.[235] The commonest form of AHC is X-linked and due to mutations in the nuclear receptor encoded by *NR0B1*. In addition to adrenal failure, the males also suffer from hypogonadotropic hypogonadism.[236,237] Heterozygous and homozygous mutations in *NR5A1* can lead to adrenal failure and 46,XY disorders of sexual development (DSD), or isolated adrenal insufficiency[235] or isolated gonadal dysgenesis in 46,XY DSD.[238] Additionally, *NR5A1* mutations have been associated with male infertility[239] and primary ovarian failure and infertility in 46,XX females.[240]

Syndromic adrenal insufficiency due to adrenal hypoplasia includes IMAGe syndrome, Triple A syndrome, Irish Travelers syndrome, and recently MIRAGE syndrome. MIRAGE syndrome comprises a striking phenotype of a multisystem disorder including myeloproliferative dysplasia, infection, growth restriction, adrenal hypoplasia, genital anomalies, and enteropathy, and mutations in *SAMD9* have been identified by whole-exome sequencing.[241,242] Up to 30% of cases of syndromic adrenal hypoplasia have no cause identified.

CAH can lead to a variable degree glucocorticoid and mineralocorticoid deficiency and varying degrees of genital ambiguity and may be due to mutations in *CYP11A1*, *StAR*, *3βHSD2*, *CYP17*, *CYP21A2*, *CYP11B1*. Isolated mineralocorticoid deficiency is due to deficiency of the enzyme P450c11AS, which has 11β hydroxylase, 18-hydroxylase, and 18 methyl-oxidase activities, and catalyzes all the reactions needed to convert deoxycortisol to aldosterone.[243] Functional mineralocorticoid deficiency with severe salt loss resulting in hyponatremia and hyperkalemia may also arise as a result of mutations in the *MR* or in the genes encoding the three subunits of the epithelial sodium channel, ENac (*SCNN1A*, *SCNN1B*, *SCNN1G*).[244]

Manifestations of mineralocorticoid deficiency in the newborn term infant can occur because of aldosterone deficiency or competition for binding to renal MRs by other steroids. Relatively reduced glomerular filtration in the newborn limits sodium loss initially but by 1 week of age, aldosterone deficiency produces the characteristic manifestations of hyponatremia, hyperkalemia, and volume depletion.

KEY POINTS

- Paradoxically fetal adrenal steroidogenesis is programmed largely to produce inactive products.
- The fetal adrenal is a key component of the fetoplacental unit, producing androgens that are aromatized to estrogens by the placenta.

Fetal Gonads

A series of molecular events result in sex determination, directing the bipotential gonad to become a testis or ovary. Factors produced by the gonad then drive sex differentiation and the development of phenotypic sex.

Gonadal Embryology. The mammalian gonad is derived from two tissue anlagen: the primordial germ cells that have migrated from the yolk sac wall and somatic stromal cells that migrate from the primitive mesonephros.[245,246] The germ cells are incorporated into the developing gonadal ridge during the sixth week, Until 6 weeks, the fetal gonads are bipotential.

SRY is the critical regulator of gonadal differentiation; its presence results in differentiation of the gonad into a testis and its absence leads to ovarian differentiation. *WT1*, the Wilms tumor suppressor gene located at 11p13, *SOX9*, and *NR5A1* are expressed in the bipotential gonadal ridge and are critical for normal male sexual differentiation.[247,248] WT1 is expressed in both the primitive kidney and genital ridge and plays a role in mesenchymal-epithelial transition[249] WT1 is a zinc finger DNA-binding protein that acts as a transcriptional activator or repressor depending on the cellular or chromosomal context. Alternative splicing results in several WT1 isoforms that have separate functions. WT1 isoforms associate and synergize with *NR5A1* to promote the expression of AMH. *WT1* missense mutations, associated with 46XY DSD in Denys-Drash syndrome, fail to synergize with *NR5A1*.[250] In addition, *WT1* binds to and activates the *SRY* promoter. NR5A1 is required for both testicular and ovarian development and mediates *AMH* gene expression and gonadotropin production.

NR5A1 and NR0B1 are orphan receptors of the steroid-thyroid hormone family of nuclear receptors and interact as heterodimers in the regulation of target genes in the ventromedial hypothalamic nucleus, the hypothalamic gonadotroph cells, and the adrenal glands.[247,248] *NR5A1* also regulates the expression of *SRY*. *Sry* binds to multiple elements within a *Sox9* gonad-specific enhancer in mice, along with *Nr5a1*. *NR5A1* and *SRY* cooperatively upregulate *SOX9* and then, together with *NR5A1*, *SOX9* also binds to the enhancer to help maintain its own expression after that of *SRY* has declined.[251]

The expression of *SRY* is transcriptionally regulated by *WT1*, *NR5A1*, *GATA4* and its cofactor zinc finger protein *FOG-2* (also called *ZFPM2*), and chromobox protein homolog 2 (*CBX2*).[252] SRY is thought to have evolved from SOX3. Loss-of-function mutations of *SOX3* do not effect sex determination, but *Sox3* overexpression in murine XX gonads leads to testis differentiation, and in humans, *SOX3* duplication or rearrangements of the *SOX3* regulatory region have been found in 46,XX individuals with dysgenetic testes.[253]

INSL3 and its G-protein–coupled receptor relaxin family peptide receptor 2 (RXFP2) play a role in testicular descent. INSL3 is produced in Leydig cells, is secreted in the circulation, and may prove to be a biomarker of Leydig cell function. Rare mutations have been described in cryptorchidism.[254,255]

A menu of genes identified so far in gonadal development is shown in Fig. 130.5. Fetal pituitary gonadotropins are not required for gonadal development or sexual differentiation; lh or fsh receptor knockout mice are born phenotypically normal.[256] Human mutations in several of the genes programming gonadal differentiation have been described.[246] Loss-of-function mutations of *SRY* or *SOX9* lead to 46,XY DSD and sex reversal, while gain-of-function mutations lead to 46, XX DSD and sex reversal. *SOX9* mutations are also associated with campomelic dysplasia. *NR5A1* mutations have been associated with 46,XY DSD with gonadal dysgenesis with or without adrenal failure. More recently, the phenotypic spectrum of *NR5A1* mutations has widened to encompass gonadal dysgenesis, premature ovarian failure, male infertility, and the vanishing testis syndrome.[238,240,257] Mutations may be heterozygous or recessive, the latter being likely to be associated with a more severe loss of function.

WT1 mutations are found in several syndromes associated with abnormal testicular embryogenesis and renal abnormalities such as Wilms tumor or glomerulosclerosis (WAGR, Denys-Drash, and Frasier syndromes). *WNT4* gain-of-function mutations result in 46,XX DSD with sex reversal, and *DAX1* duplication results in 46,XY DSD with sex reversal.53 Mutations in *CBX2* are also associated with 46XY DSD with gonadal dysgenesis and ovarian development.[258]

Testicular Development. In the presence of SRY, male gonadal differentiation begins at 7 weeks' gestation with organization of the gonadal blastema into interstitium and germ cell–containing testicular cords.[259] The primitive cords lose their connections with the epithelium, primitive Sertoli cells and spermatogonia become visible

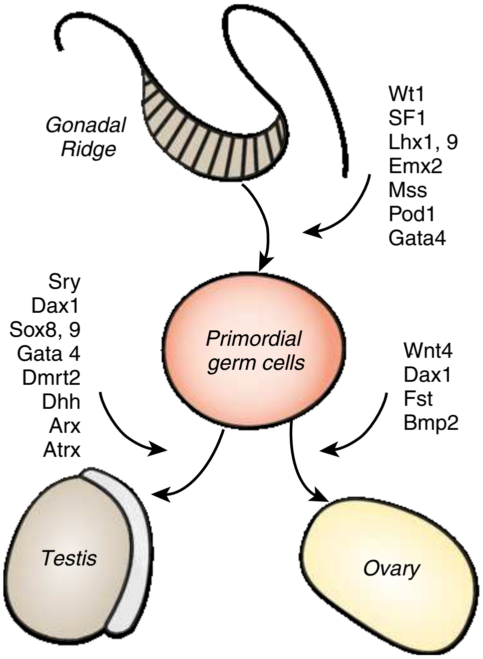


Figure 130.5 Summary of the molecular and cellular events of gonadal differentiation.

within the cords, and the epithelium differentiates to form the tunica albuginea. Leydig cells, derived from the undifferentiated interstitium, are visible by the end of the eighth week of gestation and are capable of androgen synthesis at this time. By 14 weeks, these cells make up as much as 50% of the cell mass, but as the tubules develop, they account for a smaller percentage of the tissue.

The fetal testes grow from approximately 20 mg at 14 weeks of gestation to 800 mg at birth; at 5 to 6 months, they descend into the inguinal canal in association with the epididymis and the ductus deferens.[259] The gonad, adrenal, and kidney all initially develop in close proximity, and as the testes descend, they may carry rests of adrenocortical cells with them. Testicular descent is regulated by Insl3, which is a member of the Insulin-like family and is secreted by Leydig cells.[254,255] Targeted disruption of the *Insl3* gene is associated with maldevelopment of the gubernaculum with bilateral cryptorchidism[151]. *Insl3* is also secreted by theca cells of the postnatal ovary, and females homozygous for *Insl3* mutations are subfertile with deregulation of the estrous cycle. Sertoli cells in the male gonad produce AMH, also called Müllerian-inhibiting substance (MIS), which dedifferentiates the primitive Müllerian system.[260] AMH, a 12-kD glycoprotein, is a member of the TGF-β family. *AMH* gene expression, stimulated by SRY and NR5A1, continues throughout gestation at the time of Müllerian-duct regression and then decreases after birth.68 In the female fetus, Müllerian-duct differentiation proceeds in the absence of AMH, and the primitive mesonephric duct system dedifferentiates in the absence of testosterone. In the absence of dihydrotestosterone (DHT), the urogenital sinus and external genitalia differentiate to the female phenotype. Mutation of the *AMH* gene results in a persistent Müllerian-duct syndrome in the 46,XY fetus.[260] Changes in AMH concentrations throughout the lifespan in males and females, as well as patients with Turner syndrome, have been documented.[261,262]

In the male Leydig cells, there is an increase in testicular testosterone production stimulated by the high levels of circulating placental hCG between 10 and 20 weeks.[263,264] Fetal pituitary LH contributes to fetal Leydig cell stimulation, but the predominant gonadotropic effect at this time is hCG. Fetal testosterone stimulates male sexual differentiation of the primitive mesonephric ducts (to form the ductus deferens, epididymis, seminal vesicles, and ejaculatory ducts) and masculinizes the urogenital sinus and external genitalia (to form the prostate gland, phallus, and penile urethra).

Ovarian Development. In females, differentiation of ovaries begins during the seventh week of gestation, in the absence of SRY.[265] The gonadal blastema differentiates into interstitium and medullary cords containing the primitive germ cells now referred to as *oogonia*. The cords degenerate, and cortical layers of surface epithelium containing individual small oogonia appear. By 11 to 12 weeks, clusters of dividing oogonia are surrounded by cord cells within the cortex.[265] The medulla at this time consists largely of connective tissue.[265] At 12 weeks of gestation, primitive granulosa cells begin to replicate and many of the large oogonia in the deepest layers of the cortex enter their first meiotic division. Primordial follicles appear at about 18 weeks, and the number increases rapidly thereafter.[266] The number of oocytes progressively declines from a peak of 3 to 6 million at 5 months to some 2 million at term.[204,266] Germ cell proliferation and apoptosis are ongoing simultaneously; only those oocytes enfolded by developing granulosa cells (as primordial follicles) survive.[204,266] By 5 to 7 months, stroma-derived thecal cells develop around the primordial follicles as they mature to primary follicles. The number of surviving primary follicles at birth correlates with the duration of postpubertal ovulation.

Each fetal ovary weighs about 15 mg at 14 weeks of gestation and 300 to 350 mg at birth.[265] Interstitial steroid-producing cells are present after 12 weeks, and during the third trimester, theca cells with steroidogenic capacity surround the developing follicles.[204] Significant aromatase activity is also present, but few, if any, steroids are produced by the ovary during development.[204,265]

In the 46,XX female, the absence of androgens leads to development of female genitalia; the Wolffian duct regresses and the Müllerian is maintained and forms the oviduct, uterus, cervix, and upper part of the vagina.[267]

The specific genetic mechanisms dictating ovarian development are being unraveled and some of the most powerful regulators include the WNT/FZD/β-catenin, FOXO/FOXL2, and TGFβ/SMAD, pathways.[252,268] *FOXL2*, encoding a forkhead transcription factor, is required for ovarian development.[269] *DMRT1*, expressed in testis, prevents expression of *FOXL2*, and hence female programming, in the postnatal testis.[270] In the XX gonad, the absence of SRY results in the inability of SOX9 expression to reach a critical threshold, and together with the expression of *RSPO1*, *WNT4*, *CTNNB1*, *FOXL2*, and *FST* lead to formation of the ovary. In humans and mice, R-spondin-1 (encoded by *RSPO1*) augments β-catenin signaling, possibly via WNT4.[271] Homozygous mutations in *RSPO1* result in a syndrome with palmoplantar hyperkeratosis and 46,XX DSD with sex reversal and dysgenetic testes or ovotestes.[272,273] Aberrations of the *WNT4* gene cause Mayer-Rokitanski syndrome or SERKAL syndrome (SEx Reversal, dysgenetic Kidneys, Adrenals and Lung).[274,275]

Gonadal Steroid Action. Androgen receptors appear in the mesenchyme of urogenital structures at 8 weeks of gestation, followed by the appearance of the receptors in the epithelium at 9 to 12 weeks.[276] There is no difference in receptor expression in male and female fetuses. DHT stimulates male differentiation of the urogenital sinus and external genitalia, including differentiation of the prostate, growth of the genital tubercle to form a phallus, and fusion of the urogenital folds to form the penile urethra. DHT is formed from testosterone by the 5α-reductase enzyme within the urogenital sinus and urogenital tubercle and acts through the same androgen receptor that mediates the action of testosterone in the Wolffian ducts.

Both androgens and estrogens are involved in the structural development of the rodent brain.[277,278] Gonadal hormones control gonadotropin production in the hypothalamus, programming cyclic ovarian function, and function of the testes.[278,279] Testosterone administration to neonatal female rats produces permanent inhibition of cyclic hypothalamic control through local aromatization to estradiol. In primates and humans, estrogens seem to be more effective in this regard. However, there is no evidence for permanent programming in the primate, and there appear to be no major tissue biochemical differences between the sexes *in utero* to account for sexual dimorphic behavioral or gonadotropic programming.[280]

Estrogen effects are mediated by cognate receptors, members of the large family of steroid and thyroid hormone, vitamin D, and retinoid receptors.[281,282] Two receptors, estrogen receptors α and β (ERα and ERβ).3,60,61 have been identified with 96% and 58% homology in the DNA-binding and ligand-binding domains, respectively. Expression profiles for mRNAs for both receptors have been characterized in the 16- to 23-week human fetus. One or both receptor mRNAs are present in most tissues. *ERβ* message is predominant, particularly in testis, ovary, spleen, thymus, adrenal, brain, kidney, and skin. *ERα* message is prominent in the uterus, with relatively low levels in most other tissues.[281,282] The significance of ERs in fetal development remains unclear. Knockout of the *ERα* gene in mice does not impair fetal development of any tissue, but adult females are infertile, with hypoplastic uteri and polycystic ovaries; adult males manifest decreased fertility.[282] *ERβ* knockout mice develop normally, and female adults are fertile with normal sexual behavior; adult males reproduce normally but have prostate and bladder hyperplasia.[281] Knockout of both *ERα* and *ERβ* genes also has little impact on fetal development, but after birth, the uterus, fallopian tubes, vagina, and cervix in females are hypoplastic and unresponsive to estrogen.[282] In humans, ERα mutations in males are associated with tall stature and insulin insensitivity.[283]

Disorders of Sex Development. Elucidating the mechanisms underlying sex determination and differentiation, together with society's evolving fluidity regarding gender and sexual identities, has led to rapid changes in the approach to the management of infants with disorders of sex development (DSD) (see Chapter 101). Optimal care requires an experienced multidisciplinary team, which transitions into adolescence and adulthood.[284] There is a need for data sharing and research, as highlighted by the Chicago Consensus in 2005.[285]

KEY POINTS

- In the presence of SRY, male gonadal differentiation begins at 7 weeks' gestation. Ovarian development is not a "default" or passive outcome of the bipotential gonad.

Fetal Thyroid

Thyroid Embryology. The thyroid gland, one of the first endocrine organs to develop, is a derivative of the primitive buccopharyngeal cavity and is composed of two sets of anlagen, which fuse together during development. A median anlage is derived from a midline thickening of the pharyngeal floor in the region of the first pharyngobranchial pouch, consisting of foregut endoderm cells, which give rise to precursors of the T_3- and T_4-producing follicular cells. Paired lateral anlagen (or ultimobranchial bodies), originating from caudal extensions of the fourth pharyngobranchial pouches, give rise to the parafollicular calcitonin (C)–secreting cells. Recent lineage-tracing studies in mice have shown that both medial and lateral anlagen are of endodermal origin rather than the earlier theory of a neural crest origin of thyroid C cells.[286,287]

In humans, the thyroid anlagen are discernible by 20 to 22 dpc, and between 24 and 48 dpc thyroid precursor cells undergo caudal migration, proliferation, and lateral expansion. This whole process is known as lobulation. At 24 dpc, the median anlage develops a thin, flask-like diverticulum extending from the floor of the buccal cavity at a point that is later known as the foramen caecum on the developing tongue to the fourth branchial arch. A persisting epithelial stalk, the thyroglossal duct, connects the thyroid primordium to the pharynx during descent through the anterior midline of the neck to its definitive location in the anterior neck. During normal development, the thyroglossal duct has usually disappeared by 37 dpc leaving behind only the foramen caecum remnant. If the thyroid descent is abnormal, this may result in an ectopic thyroid gland or a persistent thyroglossal duct.[288] Fusion of the lateral and median anlagen occurs at about 44 dpc. By 51 dpc, the gland has formed into two lateral lobes with a connecting isthmus and has descended below the thyroid cartilage, its final position, by 7 to 9 weeks. At the same time, connection of the median anlage with the ultimobranchial body, developed from the endoderm of the fourth pharyngeal pouch, occurs, resulting in the incorporation of the C cells into the thyroid.

The terminal differentiation of thyroid follicular cells and the formation of follicles occur only once migration is complete, at around 60 dpc. By 10 weeks, colloid is visible histologically and thyroglobulin synthesis and iodide accumulation/uptake can be demonstrated within the gland. During the final follicular phase of development, colloid spaces increase in size and there is progressive cell growth and accumulation of thyroid hormones. Terminal differentiation of the human thyroid is characterized by the onset of follicle formation and thyroid hormone synthesis at 11 weeks and is marked by expression of the TSH receptor (TSHR), the sodium-iodide symporter (NIS), thyroglobulin (Tg), and thyroperoxidase (TPO). Thyroid hormones can be detected in the thyroid from 9 weeks post conception (wpc) and in fetal blood from 10 wpc.[289–292] At 12 weeks of gestation, the fetal thyroid gland weighs about 80 mg, and at term, it weighs 1 to 1.5 g. The parathyroid glands develop between 5 and 12 weeks of gestation from the third and fourth pharyngeal pouches.

Genetic Regulation of Thyroid Development. Several developmental genes are involved in programming thyroid and parathyroid gland development. Genes encoding the transcription factors Hhex, Nkx2-1 (Titf1/ Ttf1), Foxe1 (Titf2/ Ttf2), and Pax8 (Fig. 130.6) are coexpressed in the thyroid anlage and are involved in a transcriptional network of interactions of mutual dependence.[293] Deletion of any one of the genes encoding Hhex, Nkx2-1, Pax8, or Foxe1 inevitably results in thyroid agenesis or severe thyroid hypoplasia despite the fact that the other factors are still being expressed in the thyroid placode.[293] In addition to mediating the formation of the thyroid bud, these transcription factors drive functional differentiation, regulate the expression of genes involved in thyroid hormone biosynthesis, and play a role in the maintenance of the mature thyroid gland. Nkx2-1 and Pax8 also play a role in the survival of thyroid cell precursors and regulate thyroid-specific gene expression, while Foxe1 is critical for cellular migration.[290] It is not known if Hhex directly influences specific steps in thyroid morphogenesis, although it might play a role acting indirectly through Pax8 in thyroid progenitor cell survival.[293]

Evidence from mouse and zebrafish studies suggests that thyroid morphogenesis is guided by cues from extrinsic factors such as permissive signals from cardiac mesoderm and blood vessels.[294] Genes known to contribute to thyroid morphogenesis in animal models include the Notch ligand *Jag1*, the thyroid and cardiac transcription factor *Nkx2*–5, and *Ntn1*, which is implicated in zebrafish aortic arch artery formation and thyroid morphogenesis.[295–298]

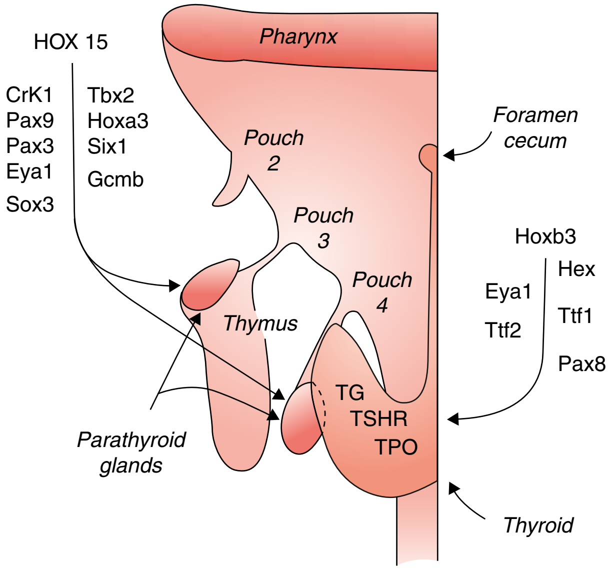


Figure 130.6 Cartoon showing genes implicated in programming the development of the thyroid and parathyroid glands. Most of this information was derived from genetic studies in mice.

Other autonomous factors that may play a role in thyroid morphogenesis include *Tbx1*, a transcriptional regulator of the Fgf signaling pathway.[299] *TBX1* is the prime candidate gene missing in 22q11.2 deletion or DiGeorge syndrome. In mice, Tbx1 is expressed in both the pharyngeal endoderm and subpharyngeal mesoderm but only its mesodermal activity promotes the generation of Nkx2-1$^+$ progenitors in the thyroid placode. In *Tbx1*$^{-/-}$ mice, the downward translocation of *Nkx2.1* expressing thyroid progenitor cells is much delayed. In late mutant embryos, the thyroid gland fails to form symmetric lobes but persists as a single mass that is approximately a quarter of the normal size. The hypoplastic gland mostly attains a unilateral position resembling thyroid hemiagenesis. The effect of *Tbx1* on the embryonic thyroid is mediated by *Fgf8*, which is also produced in the mesoderm. Conditional ablation of *Fgf8* in *Tbx1*-expressing cells caused an early thyroid phenotype similar to that of *Tbx1*-mutant mice. In addition, the expression of an *Fgf8* cDNA in the *Tbx1* domain rescued the early-size defect of the thyroid primordium in *Tbx1* mutants.[300,301] The Shh signaling pathway has also been implicated in murine thyroid development.[302] Shh appears to have an important role in the patterning and process of symmetrical bilobulation late in organogenesis, as well as suppressing ectopic expression of thyroid follicular cells.[303]

Mutations in *NKX2-1*, *PAX8*, and *FOXE1* are all associated with thyroid dysgenesis in humans.[290,292,304–306] *Hhex* gene knockout in mice is associated with thyroid agenesis or severe hypoplasia; however, no mutations have been identified in *HHEX* in humans as yet. Gene expression studies in thyroid tissues from human embryos showed that *NKX2-1*, *FOXE1*, *PAX8*, *TSHR*, and *DUOX2* were stably expressed from 7 to 33 gestational weeks. Genes encoding Tg, TPO, and pendrin (PDS) were expressed as early as 7 weeks but increased significantly before 11 weeks. *NIS* is the last gene to appear, and it shows the strongest upregulation at 10 weeks. Its expression is the most strongly correlated with the sudden onset of thyroid hormone synthesis at 11 weeks.[307]

Fetal Thyroid Hormone Biosynthesis. Although the onset of fetal thyroid hormone synthesis starts at the end of the first trimester, concentrations are low and necessitate a high dependency on maternal thyroid hormones during the first half of pregnancy. Overt maternal hypothyroidism has been shown to be associated with significant complications including miscarriage, premature delivery, and placental abruption, as well as adverse effects on fetal brain development and neurocognitive outcomes in offspring.[308] The effects of subclinical maternal hypothyroidism are less clear. Using a zebrafish model, the ontogeny of the hypothalamic-pituitary-thyroid axis and its negative feedback regulation by thyroid hormones has been characterized. Embryonic exposure to excess thyroid hormone during a developmental time window resulted in long-term dysregulation of pituitary function by apoptosis of developing thyrotrope cells.[309] These findings are relevant in the context of maternal hyperthyroidism due to maternal Graves disease or over-replacement of pregnant hypothyroid women in the first trimester with levothyroxine.

The fetal hypothalamus secretes TRH from 8 weeks' gestations, resulting in fetal pituitary TSH secretion. Despite the high TRH concentration, the TSH concentration in the fetal circulation remains low until 18 to 20 weeks' gestation when the gland functionally matures.[310] Pituitary and plasma TSH concentrations increase during the second trimester in the human fetus, coincident with the development of the pituitary-portal circulation.[311] In the rat, expression of the *TSHR* gene is significantly upregulated on fetal day 17 and this is accompanied by significant growth and rapid development in terms of structure and function. The activation of the TSH/TSHR signaling pathway is concurrent with the expression of genes needed for thyroid hormone synthesis and secretion. The expression of *Tg* and *TPO* increases and thyroid follicles are seen with thyroid hormonogenesis, suggesting that the TSHR must play an important role in these events. Murine mutation of the *TSHR* gene is associated with the *hyt/hyt* phenotype, which exhibits severe hypothyroidism and a hypoplastic but normally located thyroid gland with a poorly developed follicular structure.[312,313] In humans, a similar phenotype is observed in babies of mothers with potent TSHR-blocking antibodies and in babies with severe loss-of-function mutations in *TSHR*.[313]

Plasma concentrations of thyroxine-binding globulin (TBG) and total T_4 increase progressively, from low concentrations at 16 to 18 weeks of gestation to maximal concentrations at 35 to 40 weeks. Free T_4 concentrations increase with the rise in T_4 production. The increases in plasma TSH and T_4 concentrations during the third trimester reflect progressive maturation of hypothalamic pituitary control and of thyroid gland responsiveness to TSH, respectively. A peak T_4 of 6 to 8 μU/mL (6 to 8 mU/L) occurs by term.[310] Pituitary TSH secretion is responsive to hypothyroxinemia and to TRH early in the third trimester. Hypothyroxinemia results in an elevated fetal serum TSH, whereas hyperthyroidism due to maternal Graves disease is associated with a suppressed fetal TSH concentration.[290]

Premature infants born at 26 to 28 weeks of gestation respond to exogenous TRH with an increase in plasma TSH concentration comparable to that in adults.[314] However, their TSH response to extrauterine exposure is reduced, indicating hypothalamic immaturity, and their T_4 and T_3 increments are reduced relative to term infants. Compared to infants born at term, neonatal free T_4 increments are reduced in premature infants born at 31 to 34 weeks' gestation, attenuated at 28 to 30 weeks and absent in 23 to 27-week infants.[315] This reflects hypothalamic-pituitary-thyroid system immaturity inversely related to gestational age.[316]

Hypothalamic-pituitary-thyroid control matures during an interval corresponding to the late third trimester and early neonatal period of human development.[311,317] The period of concomitant increases in fetal TSH and free T_4 concentrations during the latter half of gestation is followed by sequential surges of TSH, free T_4, and T_3 in the early neonatal period and a final slow equilibration of the TSH/free T_4 ratio to adult values during infancy and childhood.[311,316,318,319] This sequence requires coordinated maturation of hypothalamic TRH secretion, pituitary TRH sensitivity, thyrotropin negative-feedback control, and thyroid follicular cell responsiveness to TSH. Fetal serum TRH concentrations are higher than in maternal blood and are a consequence of extrahypothalamic (placenta and pancreas) TRH production and decreased TRH degradation in fetal serum. Functionally, the fetus progresses from a state of both primary (thyroidal) and tertiary (hypothalamic) hypothyroidism at midgestation, through a state of mild tertiary hypothyroidism during the final weeks *in utero*, to a fully mature hypothalamic-pituitary-thyroid axis by 2 months after birth.[317] At birth, several changes occur to stimulate neonatal thyroid hormone production (see "Adaptations to Extrauterine Life"). In response to a cool extrauterine environment, the neonatal hypothalamus secretes TRH, which stimulates a pituitary TSH surge. The TSH concentrations reach a peak of 60 to 80 μU/mL (60 to 80 mU/L) by 30 minutes after delivery and fall to normal-term infant concentrations by 3 to 5 days of life.[310]

The adult thyroid follicular cell can modify iodine transport or uptake with changes in dietary iodine intake, independent of variations in serum TSH concentrations.[320] Before 36 to 40 weeks of gestation, the thyroid gland lacks this autoregulatory mechanism and is susceptible to iodine-induced inhibition of thyroid hormone synthesis.[314,321] The fetal thyroid follicular cell, when exposed to high circulating levels of iodide, is unable to reduce iodide trapping and prevent the high intracellular iodide concentrations that produce the blockade of hormone synthesis referred to as the *Wolff-Chaikoff effect*. Failure of the immature thyroid to exhibit autoregulation is probably due to failure of downregulation of thyroid cell membrane NIS units, which may be related to the absence or reduced iodination of an 8- to 10-kDa protein in the thyroid follicular cell.[320,321] In addition to maturation of autoregulation, thyroidal responsiveness to TSH increases during the last trimester.[322]

Thyroid hormones are metabolized by different pathways: glucuronidation, sulfation, and deiodination, the latter being the most important. The metabolism of thyroid hormones occurs through a progressive series of monodeiodinations.[311,323] Three deiodinase enzymes act to remove an iodine atom from the outer (phenolic) ring or the inner (tyrosyl) ring of the tetraiodothyronine (T_4) molecule, thus, respectively, activating or inactivating the hormone. The deiodinases are encoded by separate genes and share sequence homology. Most of the circulating biologically active T_3 in adults is derived by outer-ring monodeiodination of T_4 in the liver and other nonthyroidal tissues; biologically inactive rT_3 derives from inner-ring deiodination of T_4 in peripheral tissues. The type I enzyme (D1), an outer-ring monodeiodinase, is a high Michaelis constant (K_m) enzyme inhibited by propylthiouracil and stimulated by thyroid hormone. It deiodinates T_4 to T_3 and rT_3 to T_2. D1 also has inner-ring deiodinase activity, converting T_3 to T_2. The activity of D1 is low throughout gestation. The type 2 outer-ring monodeiodinase (D2) is a low-K_m enzyme that is insensitive to propylthiouracil and inhibited by thyroid hormone. It deiodinates T_4 to T_3 and rT_3 to T_2, and it is highly expressed in the brain and pituitary. Type 3 monodeiodinase (D3) inactivates T_4 and T_3 via inner-ring deiodination of T_4 to rT_3 and of T_3 to T_2; it is highly expressed in fetal tissues and placenta. D1 is largely responsible for the production of T_3 that escapes from the cells, especially liver and kidney, into the circulation, whereas D2 is responsible for the production of local tissue T_3. Inactive rT_3 also diffuses out of most tissues to appear in plasma.

Iodothyronine deiodinases belong to a family of selenoproteins. Selenium is an essential trace element required for the biosynthesis of selenoproteins. Selenocysteine insertion sequence (SECIS)–binding protein 2 (SBP2) represents a key factor for the insertion of selenocysteine into selenoproteins. Recently, mutations in the *SBP2* gene have been described that lead to a multisystem disorder with deficiencies of multiple selenoproteins, resulting in growth retardation, myopathy and skin photosensitivity, nonketotic hypoglycemia, colitis, and infertility.[324,325] Due to abnormal D2 activity, T_3 is low, T_4 and rT_3 are increased, and TSH is slightly elevated. As selenoproteins also serve as antioxidants, tissue damage may be related to increased concentrations of ROS.[325,326]

The distribution of the deiodinases has been characterized in rodent and human tissues (Table 130.4).[327] D2 is detectable by midgestation and plays an important role in supplying T_3 to developing brain tissue, regulating thermogenesis in brown adipose tissue (BAT) in the neonatal period, and regulating pituitary TSH secretion. D3 activity is present in the placenta, liver, and perhaps fetal skin, accounting for the higher concentrations of rT_3 in the fetus and limiting the metabolic effects of thyroid hormones during much of fetal life. There is little conversion of T_4 to circulating T_3 via D1 deiodination until midgestation in the human fetus; plasma T_3 concentrations are low (<0.2 nmol/L or <15 ng/dL) until 30 weeks of gestation, after which the mean value increases to 0.7 nmol/L (50 ng/dL) at term (Fig. 130.7).[328] On the other hand, fetal brain T_3 concentrations are 60% to 80% of those in the adult by fetal age 20 to 26 weeks owing to D2 activity. In the presence of fetal hypothyroidism, D2 increases while D3 decreases in an effort to maintain near-normal brain T_3 concentrations.

TABLE 130.4 Deiodinase Expression in Human and Rodent Tissues

Tissue	D1	D2	D3
Brain	X	X	X
Pituitary		X	X
Thyroid		X	X[a]
Liver	X		X[b]
Kidney	X		
Ovary	X		X
Ear	X[a]		
Heart	X[a]		
Muscle	X[b]		
Skin	X	X	
Testes	X	X	
Uterus	X	X	
Brown fat		X	

[a] Expressed only in humans.

[b] Expressed only in fetus.

From Miller WL. Steroid hormone biosynthesis and actions in the materno-feto-placental unit. *Clin Perinatal* 1988;25:799–817.

Sulfation is active in fetal tissues, and the predominant thyroid hormone metabolites in the fetus are iodothyronine sulfates.[311,329,330] High levels of phenolsulfotransferases (SULT) have been characterized in the fetal liver, lung, and brain by midgestation. SULT activities decrease rapidly in the neonatal period.[330] In the last third of gestation in fetal sheep, the mean plasma production rates for T_4 and metabolites (in micrograms per kilogram of body weight per day) are as follows: T_4, 40; T_4 sulfate (T_4S), 10; rT_3, 5; rT_3S, 12; T_3, 2; and T_3S, 2. All metabolites are biologically inactive except T_3 and perhaps T_3S, so 90% of the T_4 metabolites in the fetus are biologically inactive.[329] The sulfated metabolites accumulate in fetal serum as a result of the low D1 activity in fetal tissues and because the sulfated iodothyronines are not substrates for D3.[320,331] The production rate of T_3 increases progressively between 30 weeks of gestation and term because of maturation of D1 activity in the liver and other tissues and because of decreasing D3 activity in the placenta.[322,330] In fetal sheep, hepatic D1 activity increases progressively during the last trimester.[332]

It has long been assumed that thyroid hormones passively diffuse into cells. However, several classes of cell membrane iodothyronine transporters have been described, questioning this hypothesis.[333–335] These transporters belong to different families of organic anion, amino acid, and monocarboxylate solute carriers, including the organic anion transporting polypeptide (OATP) family and the solute carrier family 21 (SLC21).[333–335] The significance of these transporters is not yet clear, but mutations of the human monocarboxylate transporter 8 (MCT8), a member of the SLC21 family shown to be a specific thyroid hormone transporter present in developing brain, leads to X-linked Allan-Herndon-Dudley syndrome (AHDS), which is characterized by thyroid hormone deprivation in the brain resulting in severe psychomotor retardation and elevated T_3 concentration in peripheral tissues.[336–338] A

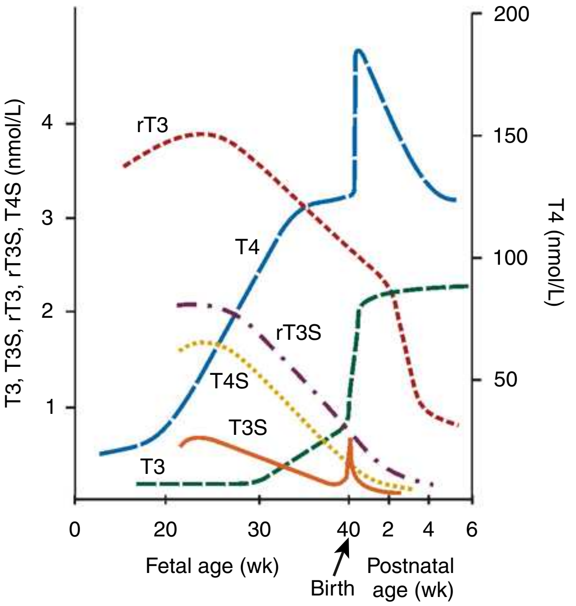


Figure 130.7 The pattern of maturation of thyroid hormone levels in the human fetus from midgestation through the perinatal period. The thyroid-stimulating hormone surge *(not shown)* at birth and the prenatal increase in fetal hepatic monodeiodinase type 1 activity stimulate marked increases in T_3 and T_4, facilitating the transition to postnatal life. The concentrations of bioinactive sulfated iodothyronine analogues T_3S, T_4S, reverse T_3 (rT_3)S, and T2S *(not shown)* peak during midgestation and fall progressively to term.

recent study has shown that brain damage in human fetuses is already induced *in utero*.[339] *Mct8* expression in neonatal mice has been localized to neurons in the olfactory bulb, cerebral cortex, hippocampus, and amygdala. Presumably, all thyroid hormone–sensitive cell populations express iodothyronine membrane transporters. A cell surface T_4 receptor has been characterized as an $\alpha_V\beta_3$ integrin that serves as the initiation site for T_4-induced activation of the mitogen-activated protein kinase (MAPK) pathway for angiogenesis and perhaps actin polymerization and neuronal migration.[340] The ontogenesis and significance of these cell surface receptors and membrane transporters in fetal development remain to be further defined. Recent studies have suggested that MCT8 is also required for normal thyroid hormone secretion, in addition to its role as a thyroid hormone transporter.[341] Treatment options for patients with MCT8 deficiency are limited. PTU blocks thyroid hormone production and inhibits the conversion of T_4 to T_3. PTU in combination with levothyroxine has shown limited metabolic benefits but no effect on neurological features. To date, two thyromimetics, DITPA (diiodothyropropionic acid) and Triac (triiodothyroacetic acid), are under investigation.[342,343] DITPA has been used in a small number of patients from age 9 months, and it has been shown that DITPA crosses the placenta and reaches neuronal target cells in mice inducing a T_3-like effect.[342,343] DITPA may therefore be a future candidate for *in utero* treatment of MCT8 deficiency. Triac has been used in a phase I clinical trial evaluating its effectiveness and safety for treatment of peripheral thyrotoxicosis in MCT8-deficient patients (both pediatric and adult).[337] This study is currently in phase II to evaluate the action of Triac in young boys <30 months with AHDS, in particular its effect on neurological symptoms and neurodevelopment.

The role of maternal thyroid hormone during fetal development remains controversial. The high placental concentration of D3 inactivates most of the thyroid hormone presented from the maternal circulation. The iodide released in this manner is used for fetal thyroid hormone synthesis. Despite the limited maternal-fetal placental transfer of T_4 and the predominant production of inactive thyroid hormone metabolites in the human fetus, significant concentrations of free T_4 are present in fetal fluids, from placental transfer early in gestation and from fetal thyroid production during later gestation.[322,344] Early in gestation, placental transfer is the only source of T_4 in fetal fluids and is essential for normal fetal neurodevelopment. T_4 is detectable in human coelomic fluid at concentrations of 0.5 to 2 nmol/L between 6 and 11 weeks of gestation, before the onset of fetal thyroid function.[345] Low concentrations of T_4 are detectable in the fetal brain at about 10 weeks of gestation. Significant placental transfer continues to term, when serum T_4 concentrations in the athyroid fetus range from 30 to 70 nmol/L (2.3 to 5.4 μg/dL).[346] Isotopic equilibrium studies with pregnant rats at term suggest that 15% to 20% of the T_4 in fetal tissues is of maternal origin.[347] As indicated, thyroid hormones cross the placenta early in gestation, supplying the low concentrations of free T_4 that are essential for brain development between 12 and 20 weeks, before the onset of fetal thyroid hormone production.[344] Most thyroid hormone in the fetal compartment is inactivated to sulfated and deiodinated analogues until the perinatal period.[311,329] This neutralization of active circulating thyroid hormone maintains the low T_3 metabolic state, facilitating fetal growth and programmed tissue maturation.

Thyroid Hormone Action. Classic thyroid hormone actions are mediated via functional thyroid hormone nuclear receptors (TRs), members of the steroid/retinoid/vitamin D family of nuclear transcription factors. Two genes code for the receptors: *THRA* on chromosome 17 encodes TRα and *THRB* on chromosome 3 encodes TRβ.[348] These genes code for four classic receptor isoforms (TRα1, TRα2, TRβ1, and TRβ2), three of which bind thyroid hormones (T_3/T_4 affinity 10 : 1) and bind to DNA to effect gene transcription. The TRα2 isoform does not bind thyroid hormone but binds to DNA and can inhibit binding of other TRs. The TRs exist as monomers, homodimers, and heterodimers with other nuclear receptor family members such as retinoid X (RXR). Other TR transcripts, including TRΔα1 and TRΔα2, have been characterized; they do not bind DNA or T_3 but can inhibit TR and retinoid receptor activities.[349]

The TRs are expressed developmentally and differentially in various fetal and adult tissues (Table 130.5). TRα proteins are present in most tissues. TRβ1 is expressed in the liver, kidney, and lung and in developing brain, cochlea, and pituitary. TRβ2 expression is restricted largely to the pituitary gland, retina, and cochlea.[322,348] The receptors function redundantly, as indicated by knockout studies in mice, but predominant effects of one or other TR have been characterized. Knockout of both the *TRα* and *TRβ* genes in mice is not lethal but results in elevated TSH concentrations, deafness, bradycardia, and decreased postnatal growth with delayed bone maturation.[322,348] Lethality occurs because of improper intestinal development associated with persistent TRΔα isoforms in TRα or combined TRα and TRβ knockout mice.[349] In the fetal rat brain, TRα1 mRNA and receptor binding are detectable by 12 to 14 days of gestation (term is 21 days), increasing to maximal levels at birth. The TRβ1 isoform is detected at birth and increases approximately 40-fold in the early postnatal period.[322,350] In the human fetal brain, TRα1 and TRβ1 isoforms and receptor binding are present by 8 to 10 weeks of gestation; TRα1 transcripts and receptor occupancy increase 8- to 10-fold by 16 to 18 weeks.[322,351,352] Liver, heart, and lung receptor binding can be identified by 13 to 18 weeks.[322,352,353]

Heterozygous missense mutations in *THRA,* coding for the TRα1, have recently been described in association with short stature, developmental delay, and chronic constipation. The mutations act in a dominant-negative fashion. The clinical phenotype is in line with the TRα1 being the predominant subtype receptor in bone, the gastrointestinal tract, cardiac and skeletal muscle, and the CNS.[354,355]

Thyroid hormone–programmed development of selective fetal tissues requires the interaction of local tissue D1, D2, thyroid receptors, receptor coactivators, and thyroid-responsive genes.

TABLE 130.5 Predominant Thyroid Hormone Receptor Subtype Functions in Developing Mice

Brain	Thermogenesis
TRα1, TRβ1	TRα1, TRβ1, TRβ2
Pituitary TSH secretion	Inner ear
TRβ2, TRα1	TRα1, TRβ2
Pituitary GH secretion	Retina
TRα1	TRβ2
Bone maturation	Intestine
TRα1	TRα1
Liver	Heart
TRβ1	TRα1

GH, Growth hormone; *TR,* thyroid hormone receptor; *TSH,* thyroid-stimulating hormone.
From Rivier C, Vale W: Neuroendocrine interactions between corticotrophin releasing factor and vasopressin on adrenocortical hormone secretion in the rat. In: Schrier RW, ed. *Vasopressin.* New York: Raven Press; 1985:181-188 and Rose JC, Turner CS, Ray DeW, et al. Evidence that cortisol inhibits basal adrenocorticotropin secretion in the sheep fetus by 0.70 gestation. *Endocrinology* 1988;123:1307–1313.

There is growing evidence that thyroid hormones act on embryological and fetal tissues early in development. Thyroid hormones regulate gene transcription mostly through the T_3 form, which binds to nuclear receptors, encoded by the *TRα* and *TRβ* genes. TRs are transcription factors with ligand-regulated activity. In most responsive tissues, the timing of maturation events is controlled by the state of the thyroid receptors acting as a molecular switch.[349,356] In the absence of T_3, the unliganded receptor (aporeceptor) recruits corepressors, repressing gene transcription. Non–T_3-binding receptors also can repress transcription by inhibiting receptor DNA binding. Local tissue maturation events are initiated by the coincident availability of T_3, liganded T_3 receptor, T_3-mediated receptor exchange of co-repressor with coactivators for creation of an active holoreceptor, and activation of responsive gene transcription. The iodothyronine deiodinases initiate or terminate thyroid hormone action and are therefore critical for the biological effects mediated by thyroid hormone. The activating deiodinase (D2) and the inactivating deiodinase (D3) can locally increase or decrease thyroid hormone signaling in a tissue- and temporal-specific fashion independently of changes in thyroid hormone serum concentrations.

These programming events have been investigated in studies of transgenic mice, including brain, liver, heart, intestine, and bone tissues, thermogenesis, and spleen erythropoiesis.[356–362] The timing of these events in the mouse ranges from early midbrain neuronal development at gestational day 15, through perinatal activation of hepatic enzymes, cardiac ion channels, and spleen erythropoiesis, to postnatal brain, intestinal, and bone maturation and thermogenesis. Parturition occurs at a gestational age equivalent to human midgestation. In hypothyroid mice, repressive effects of aporeceptors have been shown to delay tissue maturation in the brain, bone, intestine, spleen, and heart.[360] The increase in circulating T_3 levels associated with parturition in mice and humans normally triggers the development of tissue functions essential to postnatal metabolism and homeostasis (e.g., hepatic, intestinal, and cardiac functions and brown fat thermogenesis). Thyroid hormone–stimulated maturation of vision and hearing appears to be triggered by the local expression of D2, mediating local T_3 production, postnatally in the mouse and probably toward the end of the second midtrimester in the human fetus.[363]

In humans, T_3-mediated maturation of fetal tissues, including liver, heart, BAT, and bone, allows them to become thyroid hormone responsive during late gestation and in the perinatal period. Paracrine actions of thyroid hormone are critical for normal fetal development—for example, in the cochlea, where D2 is expressed in connective tissue immediately adjacent to the sensory epithelium and spiral ganglion, where thyroid hormone receptors are located. This implies that D2-containing cells in the connective tissue take up T_4 from the circulation, convert it to T_3, and then release D3 to adjacent responsive cells. Similarly, in the brain, D2 is expressed predominantly in glial cells, whereas TRs are expressed in adjacent neurons and oligodendrocytes. In other areas of the brain, such as the pituitary gland, hippocampus, and caudate nucleus, there is coexpression of D2 and TRs. On the other hand, D3 is coexpressed with TRs in neurons, thereby protecting sensitive tissues from the effects of excess thyroid hormone.

The actions of thyroid hormones and their developmental regulation in the brain are complex. Functionally, thyroid hormones are critical for the establishment of neural circuits during a critical window of brain development. They provide inductive cues for the differentiation and maturation of a number of processes such as neurogenesis and neural cell migration (occurring between 5 and 24 weeks of gestation), neuronal differentiation, dendritic and axonal growth, synaptogenesis, gliogenesis (late fetal period to 6 months postpartum), myelination (second trimester to 24 months postpartum), and neurotransmitter enzyme synthesis. TRs are found in highest concentration in developing neurons and in multiple areas of the fetal brain, including the cerebrum, the cerebellum, and the auditory and visual cortices. The hormones bind to receptors and stimulate a number of genes such as myelin, neurotrophins and their receptors, cytoskeletal components, transcription factors, extracellular matrix proteins and adhesion molecules, intracellular signaling molecules, and mitochondrial and cerebellar genes.

Thyroid hormone is also important for normal bone growth. T_3 regulates endochondral ossification and controls chondrocyte differentiation in the growth plate both *in vivo* and *in vitro*. TRs are expressed on osteoblasts and growth plate chondrocytes, and T_3 target genes have been identified in bone. T_3 stimulates skull suture closure *in vivo*.[364]

In the perinatal period, thyroid hormone stimulates the transcription of thermogenin, also known as uncoupling protein 1 (UCP1), a protein that uncouples nucleotide phosphorylation and the storage of energy as adenosine triphosphate (ATP), actions that are important for nonshivering thermogenesis by BAT.

Congenital Hypothyroidism. Congenital hypothyroidism (CH) is a common endocrine disorder with an incidence of 1:3000 to 4000 at birth. It is typically reported to be sporadic. Pathogenic variations in candidate genes have been found in <10% of the cases; however, genetic studies targeting specific phenotypes or particular ethnic groups yielded higher mutation detection rates suggesting that CH can be considered as a disease with a strong genetic component, but with a largely missing explanation for its heritability.[365]

Thyroid dysgenesis is causal in the majority (85%) of children with permanent primary CH; however, mutations in known genes account for less than 5% of patients. The defects in thyroid morphogenesis can be classified as athyreosis (lack of thyroid tissue), ectopy, hemiagenesis (hypoplasia of one thyroid lobe), or hypoplasia. Genes associated with thyroid hormone dysgenesis and/or syndromic congenital primary hypothyroidism account for less than 5% of patients and include *NKX2-1, FOXE1, PAX8, NKX2-5, GLIS3, JAG1, TBX1, NTN1,* and *CDCA8.*[365] Carré et al. also recently described the impact of *BOREALIN* mutations as a genetic cause of thyroid dysgenesis in humans. *BOREALIN* expression in human embryonic thyrocytes was shown before and after the onset of thyroid hormone synthesis and *BOREALIN* is involved in the adhesion and migration of the thyrocytes, in keeping with a phenotype of thyroid ectopy.[366] Known causative genes in congenital hypothyroidism associated with dyshormonogenesis and/or with gland *in situ* include *TSHR, GNAS, TG, TPO, DUOX1/DUOX2, DUOXA2, SLC5A5, SLC26A4/PDS, IYD,* and the molecular spectrum is increasing. Table 130.6 shows mechanisms resulting in congenital hypothyroidism due to abnormal thyroid

TABLE 130.6 Mechanisms Involved in Congenital Hypothyroidism

Thyroid Dysgenesis (1 in 4500)

- Isolated thyroid aplasia, hemiagenesis, hypoplasia, or ectopy
  - Transcription factor defect (PAX-8)
  - Unknown[a]
- Associated with other developmental abnormalities
  - Transcription factor defect (TTF-1, FOXE1 [TTF-2], NKX2-5, SHH, Tbx1)

Inborn Errors of Thyroid Hormonogenesis (1 in 35,000)

- Abnormal iodide uptake (NIS, SLC5A5)
- Failure to concentrate iodide
- Abnormal organification of iodine
- Abnormal TPO enzyme
- Abnormal H_2O_2 generation (THOX, DUOX2)
- Pendred syndrome (SLC26A4)
- Defective Tg synthesis or transport
- Abnormal iodotyrosine deiodinase (DEHAL1)
- Abnormal thyroid hormone transport in the brain (MCT8)
- Abnormal selenium incorporation (SECISBP2)

Secondary and/or Tertiary Hypothyroidism (1 in 50,000 to 100,000)

- Hypothalamic abnormality
  - Isolated TSH deficiency
    - TRH deficiency
  - Multiple hypothalamic hormone deficiencies
    - Isolated hypothalamic defect
    - Associated with other midline facial/brain dysmorphic features (e.g., SOD, cleft lip/palate)
- Pituitary abnormality
  - Isolated TSH deficiency (IGSF1)
    - TRH resistance
    - Abnormal TSHβ molecule
  - Multiple pituitary hormone deficiencies
    - Ectopic posterior pituitary gland (transcription factor defect, e.g., POU1F-1, PROP1, LHX3)
    - Ectopic posterior pituitary gland (idiopathic, transcription factor defect, e.g., HESX1, SOX3, LHX4, OTX2)

TSH Resistance

- TSH receptor gene mutation
- Post-receptor defect?
- Gsα gene mutation

Thyroid Hormone Resistance (1 in 100,000)

SOD, Septo-optic dysplasia; *TRH,* thyrotropin-releasing hormone; *TSH,* thyroid-stimulating hormone

[a] Most common.

development, dyshormonogenesis, abnormal thyroid hormone transport or action, and the genes involved in these mechanisms.[367] Lower neonatal screening thresholds have increased the incidence of congenital hypothyroidism with a normally located gland *in situ*, and recent studies reveal that a significant proportion of patients have multiple gene mutations in more than one thyroid-specific gene, supporting the hypothesis of an oligogenic etiology of congenital hypothyroidism.[368]

Classical signs of CH (jaundice, lethargy, feeding difficulties, macroglossia, myxedema, hypothermia, growth restriction, progressive developmental delay, and IQ deterioration) appear during the early weeks and months of extrauterine life as maternal T_4 becomes unavailable and the non-CNS tissues become responsive to thyroid hormone.[369] The period of brain dependency for thyroid hormone extends postnatally to 2 to 3 years of age, but the early weeks and months of life are most critical. Untreated thyroid agenesis is associated with a loss of 5 to 7 IQ points monthly during the first months of postnatal life and over 6 to 8 months can amount to a 30- to 40-point IQ deficit.[370,371] Most countries therefore have a neonatal screening program to ensure early diagnosis and treatment and international guidelines for the management of CH are available.[372,373]

In congenital hypothyroidism, there is increasing evidence of maternal-fetal T_4 transfer in the first half of pregnancy, resulting in cord T_4 concentrations that are 25% to 50% of normal.[374] The transplacental passage of thyroid hormone, in conjunction with adjustments in brain deiodinase activity, has a critical role in minimizing the adverse effects of fetal hypothyroidism and helps explain the normal or near-normal outcome of hypothyroid fetuses (provided that prompt and adequate treatment of hypothyroidism ensues postnatally) as well as the relatively normal clinical appearance of the majority of babies with congenital hypothyroidism at birth. On the other hand, in the presence of both maternal and fetal hypothyroidism (such as that observed in the presence of potent TSHR-blocking antibodies, maternal and fetal *POU1F1* deficiency, and severe iodine deficiency), there is severe neurocognitive impairment despite early and adequate commencement of thyroid replacement. Importantly, the presence of inadequately controlled maternal hypothyroidism is also associated with significant neurocognitive deficit in the offspring that is not reversible by early postnatal therapy.[375]

KEY POINTS

- Thyroid hormone synthesis begins at 11 weeks' gestation. Maturation of the hypothalamo-pituitary control of the fetal thyroid develops in the third trimester.

THE FETAL AUTONOMIC NERVOUS SYSTEM

The ANS controls many physiological functions including heart rate, blood pressure, perspiration, and salivation and is composed in part of sympathetic neurons, parasympathetic neurons, and chromaffin cells of the adrenal medulla. The ability of the ANS to regulate the fetal CNS) and adapt to perturbations in homeostasis is a key necessity during human fetal maturation. The interplay between the sympathetic and parasympathetic nervous systems is achieved via the cerebral cortex, the medulla oblongata, the sympathetic ganglia, and the vagus nerve.

There is an increasing realization that the process of fetal autonomic regulation *in utero* serves specialized needs particular to development and that the onset of maturity represents the loss of these unique patterns of neural activity. Nowhere is this more evident than in the sympathetic nervous system and its endocrine counterpart, the adrenal medulla.

Embryology. The primordia of the sympathetic trunk ganglia are visible in the human fetus from as early as 4 to 5 wpc. The sympathoadrenal (SA) cell lineage is a derivative of the neuroectodermal crest (NC), which gives rise to sympathetic neurons, neuroendocrine chromaffin cells of the adrenal medulla, and extra-adrenal chromaffin cells. Signals that are important for the differentiation of SA cells into sympathetic neurons and chromaffin cells are largely unknown. In mice, the SA progenitor cells delaminate from the dorsal neural tube and migrate ventrally through the sclerotome toward the dorsal aorta to form numerous extramedullary sympathetic paraganglia,[376] which are scattered throughout the abdominal and pelvic sympathetic plexuses.[377] Each of the sympathetic paraganglia may reach a maximal diameter of 2 to 3 mm by 28 wpc and the largest of these, the organs of Zuckerkandl near the origin of the inferior mesenteric arteries, may reach 10 to 15 mm in length at term. After birth, the paraganglia gradually atrophy and disappear by 3 years of age.

The SA cells that migrate ventrally to reach the dorsal aorta[378] give rise to tyrosine hydroxylase-expressing, catecholaminergic neuronal progenitor cells, in response to bone morphogenetic protein cues from cells of the wall of the dorsal aorta and the surrounding mesenchyme.[379] Some cells migrate further ventrally around the dorsal aorta along nerves and blood vessels to enter the adrenal primordium at the cranial end and acquire a chromaffin cell phenotype.[380,381] This occurs from 6 wpc. It was thought that these were SA cells that lost expression of neuronal genes; however, more recently, it has been shown that at least some of the presumptive chromaffin cell population enters the adrenal primordium still expressing neural crest markers and then gains tyrosine hydroxylase and dopamine β-hydroxylase expression but not neuronal markers.[382] A recent study has shown that preganglionic axons and chromaffin cell precursors require class 3 semaphorin (SEMA3) signaling through neuropilins (NRP) to target the adrenal medulla.[383]

A variety of transcription factors have been identified in the development of the SA cell lineage including PHOX2B (paired-like homeobox 2b), MASH1 (mammalian achaete-scute homolog-1), PHOX2A (paired-like homeobox 2a), and HAND2 (heart- and neural crest derivatives-expressed protein 2).[376]. The *PHOX2B* gene is pivotal in the development of the ANS. *Phox2b* null mice die *in utero* because ANS neurons either fail to develop or degenerate. Mutations in human *PHOX2B* cause a rare syndrome of dysautonomia (congenital central hypoventilation syndrome), Hirschsprung disease, and a predisposition to neuroblastoma.[384]

Nerve growth factor (NGF) is required for the survival of developing sympathetic neurons. In the rodent, NGF antiserum injected into neonatal rats resulted in degeneration of immature chromaffin cells, sympathetic cells, and phaeochromoblasts.[385] It is unclear whether NGF and other growth factors are involved in the maturation of the human fetal ANS.

Functional Development of the SA System. Chromaffin cells are located in adrenal and extra-adrenal sites during fetal development, although the function of the extra-adrenal tissue is unknown and these chromaffin cells regress postnatally. Chromaffin cells are initially scattered within the adrenal cortex and are mostly noradrenergic. At around 10 wpc, small islets of chromaffin cells express PNMT (phenylethanolamine *N*-methyltransferase), the enzyme required to catalyze the methylation of noradrenaline (NA) into adrenaline, to become adrenergic.[386] With increasing gestational age, there is progressive growth of the adrenal medullae, increasing catecholamine content, and progressive maturation of medullary functional capacity. Histologically, the adrenal medullae are somewhat immature at birth but by 1 year they resemble the adult glands.

Catecholamines are present in the para-aortic chromaffin tissue by 8 to 13 wpc and concentrations increase until term. The predominant catecholamine is NA, presumably because of low activity of PNMT in para-aortic chromaffin tissue. In contrast, the adrenal medulla has high levels of PNMT activity. Adrenal PNMT is activated by the high levels of glucocorticoids that diffuse into the adrenal medulla from the adrenal cortex.[377,387] In contrast, cortisol concentrations in extramedullary chromaffin tissue are low. Cortisol and catecholamine concentrations in the fetus increase significantly in the third trimester, playing a critical role during transition of the fetus to an *ex utero* environment by priming the fetal nervous system.

How the fetus withstands an *in utero* environment of relative hypoxia has been a vibrant area of research for over a century. The ovine model provides a comparative system for physiological and developmental research in the fetus that is applicable to humans. This model has been used to study fetal adaptive responses to chronic oxygen and nutrient deprivation during placental insufficiency. Basal plasma adrenaline, NA, and dopamine concentrations decrease during the last trimester in sheep due to the maturation of clearance mechanisms.[388] The fetal sheep responds to hypoxia with increased catecholamine concentrations.[389] This response is also present in human fetuses during the third trimester of gestation.[390] There is evidence that during the basal state, circulating catecholamines are predominantly derived from sympathetic neurons, whereas the fetal adrenal is the major source of catecholamines during acute hypoxia.[391,392]

Catecholamines are critical for fetal cardiovascular function and fetal survival. Gene knockout studies in mice, targeting either tyrosine hydroxylase or dopamine β-hydroxylase, produced fetal catecholamine deficiency and midgestation fetal death in 90% of the mutant embryos.[393,394] In addition, fetal catecholamines are the major stress hormones in the fetus.[387,390,395,396] The cells of the adrenal medulla and para-aortic tissue can respond directly to fetal hypoxia by secreting large amounts of catecholamines directly into the circulation.[387] Moreover, the defense against fetal hypoxia involves catecholamine actions mediated through cardiac α-receptors that are unique to immature animals. α-Adrenergic receptors predominate in immature cardiac tissue and gradually decline in number as β-adrenergic receptors increase with maturation. Chromaffin tissue in the fetus is also innervated by opiate receptors and contains relatively large amounts of opiate peptides or enkephalins that appear to be cosecreted with the catecholamines. Pretreatment with naloxone potentiates and methadone inhibits the catecholamine response to hypoxia.[397] The extent to which these peptides or pituitary endorphins are involved in modulating fetal catecholamine secretion remains unclear.

The human neonate responds to parturition with an increase in plasma adrenaline and NA concentrations, and these responses are augmented by hypoxia and acidosis.[389,393,394] In the newborn infant, catecholamine secretion also increases after cold exposure and hypoglycemia.[387,390] Labor-induced catecholamine release may provide an important mechanism underlying the neonate's process of adapting to extrauterine conditions and the recruitment of vital postnatal adaptations.

KEY POINTS

- Fetal autonomic regulation *in utero* serves specialized needs particular to development, and this is clearly evident in the interplay between the sympathetic nervous system and its endocrine counterpart, the adrenal medulla.

FETAL PANCREAS AND GLUCOSE HOMEOSTASIS

Pancreas Development. Pancreatic development is a complex, multistep process, in which morphologically and functionally distinct exocrine and endocrine tissue types are derived from common endodermal progenitor cells. The exocrine tissue consists of acinar cells that secrete digestive fluid, centroacinar cells, and a branched ductal network that drains the fluid into the gastrointestinal tract. The endocrine cells are located in clusters called the islets of Langerhans, scattered throughout pancreatic parenchyma. They include insulin-secreting β cells (50% to 60%), glucagon-secreting α cells (30% to 45%), somatostatin-secreting δ cells (5%), pancreatic polypeptide (PP)–secreting γ cells (<5%), and ghrelin-secreting ε cells (<1%). Unlike in mice, where there is clear segregation of cell types in different regions of the islet, for example, β cells are central, surrounded by α, δ, and γ cells, in humans, β cells are intermingled with α and δ cells, throughout the islet.[398]

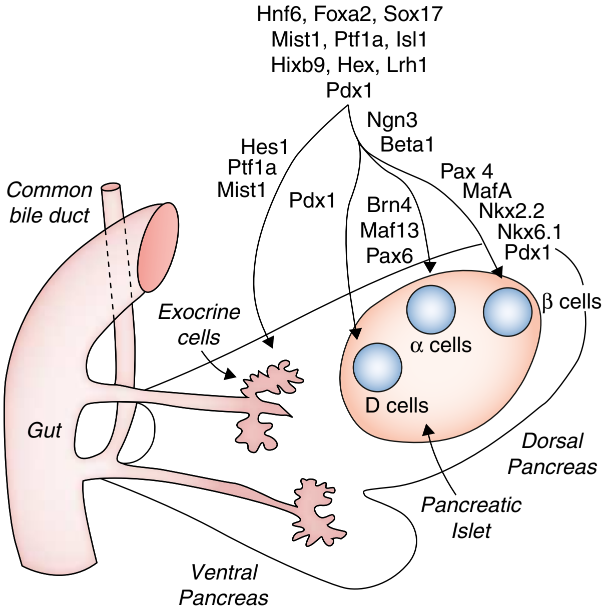


Figure 130.8 Genes implicated in programming development of the pancreas and pancreatic islets. Knockout of *Pdx1* or *Isl1* is associated with early arrest of pancreatic development. *Hes1*, *Ptf1A*, or neurogenin 3 (*Ngn3*) disruption leads to aplasia or hypoplasia of the islets of Langerhans. Disruption of the downstream transcription factors impacts formation of the beta cells or alpha cells. *Sox9* and *Hnf3b* (not shown) are required for early foregut formation and pancreas specification. (Data from Habener JF, Kemp DM, Thomas MK. Minireview: transcriptional regulation in pancreatic development. *Endocrinology*. 2005; 146: 1025–1034.)

Our understanding of human pancreas development has been extrapolated from data obtained in other species, particularly the mouse (Fig. 130.8). Specification of the mouse pancreas is initiated from the distal part of the endodermal foregut and proximal part of the midgut beginning at E8.5 and is first indicated by expression of the transcription factor Pdx1.[399,400] The pancreas arises from distinct dorsal and ventral buds, which are induced by different mesodermal structures and signaling pathways. The pancreas can only develop in an area free of Shh. Coincident with the start of pancreas development, Shh expression is repressed in the dorsal and ventral pancreatic buds, dependent on signals from the notochord that include FGF2 and the TGFβ signaling molecule, activin.[401,402]

In addition to *Pdx1*, which is necessary for normal pancreas development, there are several other transcription factor genes involved in foregut patterning and early pancreas development, including *Hnf4a*, *Hnf6 (Onecut1)*, *Hnf1b*, *Mnx1* (also known as *Hlxb9*), and *Ptf1a*.[403] Both the dorsal and ventral bud, pancreatic progenitors express the essential transcription factor *Pdx1* and *Ptf1a*, but only the dorsal bud expresses *Hoxb9*.[404] Genetic lineage-tracing in mice indicates that essentially all mature pancreatic cells, including acinar, ductal, and endocrine cells, arise from *Pdx1*$^+$/*Ptf1a*$^+$ progenitor cells.[399,405–407]

After pancreatic specification and the onset of bud formation, there are continued signaling processes that regulate the proliferation and differentiation of progenitor cells for the endocrine, ductal, and exocrine lineages. Multipotent progenitor cells initially go through a period of rapid proliferation under the influence of *Fgf10*, forming a branching epithelial network of the pancreas.[408] At around E13, the epithelium becomes segregated into distinct "tip" and "trunk" domains, while the tip domains contain multipotent pancreatic progenitor cells, which are capable of differentiating into proacinar cells, the adjacent trunk epithelial region consists of a duct/endocrine bipotential progenitor pool.[407] *Ptf1a* appears to be the key player in acinar formation and its loss results in acinar agenesis.[409] The β-catenin/Wnt signaling pathway is required for the maintenance of Ptf1a exocrine progenitor cells.[410,411]

Loss-of-function experiments in mice have identified several transcription factors that play a role in specifying the multiple endocrine lineages in the developing pancreas. *Ptf1a* and *Nkx6-1* have an antagonistic relationship, and the segregation of *Nkx6-1* and *Ptf1a* expression into exclusive trunk and tip compartments, respectively, appears to correlate with the progressive restriction of trunk and tip cells toward the future duct/endocrine or acinar lineages.[412] In the mouse, endocrine specification, via the generation of Neurog3 endocrine precursors, occurs in two major waves termed the primary and secondary transition. The primary transition takes place during the bud stages of early pancreas development, where *Neurog3*$^+$ transcripts can be detected in the presumptive pancreatic anlage as early as E8.5 to E9.5. The majority of endocrine precursors are however generated from the trunk plexus during the secondary transition, peaking in absolute numbers at around E15.5.[413]

The level of Notch signaling has been demonstrated to exert antagonistic effects on *Neurog3* transcription. Low levels of Notch signaling induce expression of *Sox9*, which positively regulates *Neurog3* expression, whereas high levels of Notch signaling induce both *Sox9* and *Hes1* leading to repression of *Neurog3* expression.[413] The remaining nondelaminating *Hnf1b*$^+$/*Sox9*$^+$/*Nkx6-1*$^+$ trunk cells are fated to become duct cells. Precocious endocrine differentiation, for example, from deficiency in Notch pathway components, such as *Rbpj*, *Dll1*, and *Hes1*, results in pancreatic hypoplasia due to *Ptf1a*$^+$ progenitor pool depletion.[414–416] *NeuroD1*, *Insm1*, and *Isl1* are all required for the development of multiple endocrine cell lineages and are thought to act downstream of *Neurog3*.[417–419]

Neurog3 expression defines the development of a progenitor population for all endocrine lineages, including α cells, β cells, δ cells, γ cells, and ε cells.[405] Endocrine progenitor cells delaminate from the trunk epithelium and coalesce to form islets. With further maturation, complex networks of factors function to establish the different endocrine lineages. Expression of transcription factor genes in a combinatorial manner further restricts cells into distinct monohormonal endocrine lineages; α cells express *Pax6*, *Nkx2.2*, *Foxa2*, *Arx*, *MafB*; β cells express *Pax4*, *Pax6*, *Nkx2.2*, *Nkx6-1*, *Pdx1*, *Mnx1*, and *MafA*; δ cells express *Pax4* and *Pax6*.[420,421]

Most of our understanding of the molecular mechanisms involved in human pancreas development is derived from studies initiated at 7 to 8 weeks of fetal development, by which time the two pancreatic buds have coalesced as a single organ, and endocrine commitment has already started and later gestational-aged human pancreatic tissue (>22 weeks' gestation).[422] More recent studies, examining specimens obtained from human embryos from as early as human embryonic stage Carnegie Stage 9 (CS9) or 22 to 26 days post conception (dpc), have provided important insights into earlier development and pancreas specification.

In humans, as in mice, the pancreas develops from dorsal and ventral pancreatic buds, which emerge as distinct outgrowths from the posterior foregut endoderm. Foregut endoderm lying adjacent to the notochord can be observed at 25 to 27 dpc (CS9). *SHH*, which must be excluded for pancreatic buds to develop, can be detected at CS10 and *PDX1* expression is first seen at CS12 (29 to 31 dpc).[423–425] It appears that human foregut and pancreatic patterning is similar to those of mice. In mice, a "primary transition" stage occurs between E9.5 and E12.5. The first differentiated endocrine cells, mainly glucagon-producing α cells, begin to appear in the dorsal bud at this phase of development, in a cluster-budding process, rather than by individual delamination of epithelial cells. By contrast, there has been no evidence of early pancreatic endocrine differentiation in human embryos possibly due to a lack of proximity between the paired dorsal aortae and the early pancreatic endoderm.[424]

The ventral and dorsal pancreatic buds at CS13 (30 to 33 dpc) are marked by *SOX9*, *PDX1,* and *GATA4*, all of which are required for human pancreas development based on the phenotype of patients with loss-of-function mutations in these genes.[426–428] Significant growth of the pancreas and proliferation of MPCs occur between CS13 and CS19. Unlike mice, *NKX2.2* is not detected in these cells. Further differences in pancreatic progenitor cells are seen from CS19 (45-47 dpc), equivalent to E14 to E14.5 in mice. *SOX9*$^+$/*NKX6-1*$^+$ "trunk" cells show low expression of *GATA4*, whereas peripheral "tip" cells are *SOX9*$^+$/*NKX6-1*$^+$/*GATA4*$^+$. By 10 wpc, proacinar "tip" cells no longer express *NKX6-1*, similar to mice. Loss of *SOX9* expression in acinar cells is delayed in humans (at 10-14 wpc) compared with mice, in which *Sox9* is lost from peripheral tip cells shortly following their formation, at around E14.5 to E15.[425]

Detection of *NEUROG3* (or *NGN3*) first occurs at CS21 (49-52 dpc) closely followed by detection of insulin-positive cells, which is the first islet cell type to appear.[423,424] *SOX9* expression persists in pancreatic duct cells but is absent in *NEUROG3*-positive endocrine progenitors. Transient expression of *NEUROG3* is also seen in humans as in mice.[425] The ventral pancreatic bud migrates toward the dorsal bud and both buds fuse together resulting in a single pancreatic organ by the end of embryonic development (CS23, eighth week). Between 10 and 12 weeks, the proliferation of branched tubules into primitive acini, islets, and ducts occurs with limited differentiation.[420,424,429]

Functional Development of the Endocrine Pancreas. Insulin, glucagon, somatostatin, and pancreatic polypeptide are measurable by 8 to 10 weeks of gestation.[430] Endocrine cell clusters, in which β cells form a core with a periphery of α cells and δ cells, similar to mouse islets are observed by 12 to 14 weeks. By 18 weeks, the peripheral α cells and δ cells dissociate from the β-cell core to form juxtaposed homogenous islets and these homogenous islets reintegrate after week 22 to form the configuration seen in adult islets, with a heterogeneous distribution of β, α, and δ cells.[431] Early in gestation, α cells predominate, whereas β-cell proliferation is programmed during the latter half of gestation. At term, the ratio of α and β cells is approximately 1:1.[423,432] α and β cells have opposing roles in glucose homeostasis, although they derive from a common progenitor and share many proteins important for glucose sensing and hormone secretion. Mouse models have underlined the similarities between the two cell types by showing that β-to-α as well as α-to-β transdifferentiation can take place under certain experimental circumstances, revealing a remarkable degree of plasticity among the different pancreatic endocrine cells.[433]

In rodents, a period of a rapid increase in β-cell mass around and shortly after birth has been noted and this rapid change is attenuated by a period of apoptosis before adult mass is arrived at. β-cell mass is more difficult to determine from a developmental standpoint in humans. At birth in the human, there are approximately 200 to 300 × 10^6 β cells, which is approximately one-third of the population present in adulthood. However, most of the actual mass change takes place in the newborn period and is associated with changes in β-cell size rather than number.[434] Thereafter there is rapid further expansion in terms of cell numbers, but the fluctuations in β-cell mass particularly with pregnancy is poorly understood. Progression to type 2 diabetes mellitus may depend on the number and functionality of β cells at birth and how these are affected by interaction with environmental factors throughout life.

The β cell is functional by 14 weeks of human gestation, and fetal pancreatic insulin content exceeds that of the adult throughout most of gestation. Fetal glucose metabolism is largely independent of the glucoregulatory hormones insulin and glucagon,[435–437] and acute hypoglycemia or hyperglycemia is not associated with significant alterations in either insulin or glucagon levels. Insulin release from the fetal rat pancreas *in vitro* in response to glucose or pyruvate is minimal but can be stimulated by leucine, arginine, tolbutamide, or potassium chloride, indicating that parts of the secretory mechanism are functional in the fetus.[435,438] Insulin secretion in adult islets is mediated by two or more mechanisms, including stimulation of the adenylate cyclase system with production of cAMP and inhibition of potassium efflux, which leads to depolarization of the cell membrane and opening of voltage-dependent calcium channels. The former mechanism, although suppressed in the fetal islets, can be augmented by theophylline, but calcium channel activation does not occur in fetal islets in response to initiators of insulin release that cause depolarization of adult islet cells.[438]

Infusion of glucose or arginine in pregnant women before hysterotomy fails to provoke fetal insulin secretion at midgestation or near term, and plasma insulin levels in the late human fetus are relatively unresponsive to high glucose concentrations before the onset of labor.[439] Similar observations have been made in the monkey: neither glucose nor arginine stimulated fetal insulin release near term but glucagon evoked prompt insulin production.[439] Late in gestation in the ovine fetus, epinephrine inhibits insulin release through a receptor pathway.[439]

In the anencephalic human fetus, the endocrine pancreas develops normally if maternal carbohydrate metabolism is not impaired; β-cell hypertrophy and hyperplasia do not occur in the anencephalic fetus nor in decapitated fetal rabbits exposed to chronic hyperglycemia. This lack of β-cell response to hyperglycemia may be the result of deficiency of GH and/or IGF-1. GH stimulates insulin gene expression and may play a permissive role in β-cell hyperplasia and hypertrophy.[440] In contrast to acute changes in glucose concentration, chronic fetal hyperglycemia does evoke hyperinsulinemia and glucagon suppression, and chronic hypoglycemia may inhibit fetal insulin and promote fetal glucagon release.[435,439]

Pancreatic glucagon concentrations are relatively high in fetal plasma and increase progressively with fetal age.[435,439] As is true for insulin, the capacity for glucagon secretion is blunted in the fetus. Hyperglycemia does not suppress fetal plasma glucagon concentrations in rats, monkeys, or sheep, and acute hypoglycemia does not evoke glucagon secretion in the rat fetus. Amino acids, which are important secretagogues for insulin and glucagon in the adult, probably have little role in modulating insulin and glucagon secretion in the preterm fetus but infusion of alanine into women at term increases both maternal and cord blood glucagon concentrations, indicating a fetal glucagon response to amino acids in the term fetus. Catecholamines also evoke glucagon release in the near-term ovine fetus.[439]

Thus the fetal pancreatic islet cells, although histologically mature and capable of hormone synthesis and hyperplasia, are relatively immature functionally at birth with regard to the capacity to secrete both insulin and glucagon. The rapid maturation of responsiveness to glucose in the neonatal period in both premature and mature infants suggests that this blunted state may be a secondary result of the relatively stable fetal serum glucose concentrations maintained by placental transfer of maternal glucose rather than a primary, temporally fixed maturation process. Alternatively, the lack of any enteric signal to the pancreas from feeding via the release of incretins may also account for this stability. The blunted capacity for insulin and glucagon secretion has been related to a deficient capacity of the fetal pancreatic islet cells to generate cAMP or a rapid destruction of cAMP by phosphodiesterase, or both.[439]

Insulin and glucagon are normally not necessary for substrate metabolism in the fetus.[435] Fetal energy needs are met by a continuous intravenous supply of glucose across the placenta, fetal glucose uptake being directly related to both the maternal blood glucose concentration and the transplacental gradient. There is little endogenous glucose production. The fetal respiratory quotient is approximately 1, which suggests that glucose is the primary energy substrate for the fetus. Other substrates such as amino acids and lactate may also be utilized in the human, as in the sheep fetus. However, at least early in gestation, hepatic metabolism and substrate utilization appear to be independent of insulin and to be modulated in an autoregulatory fashion by glucose.[439] In addition, the constant supply of glucose normally precludes the necessity for endogenous gluconeogenesis, and gluconeogenic enzyme activities are low in the fetal liver.

Glycogen storage in the fetus is modulated by fetal glucocorticoids and probably by hPL. Fetal insulin plays a role near term, when insulin also has the capacity to increase fetal glucose uptake and lipogenesis.[435,439] Insulin receptors are present in a variety of fetal tissues at levels generally exceeding those of the adult. However, downregulation of insulin receptor binding does not occur during fetal hyperinsulinemia, in contrast with observations in adult animals.[439] Fetal hepatic glucagon receptors, in contrast, are reduced in number, and the fetal liver is therefore relatively resistant to the glycemic effect of glucagon.[439] These conditions tend to potentiate the fetal anabolic milieu during the period of rapid growth in the last trimester of gestation.

Congenital Hyperinsulinism. CHI represents a group of clinically, genetically, and morphologically heterogeneous disorders and is the most frequent cause of persistent and recurrent hypoglycemia in neonates and infants. It is due to inappropriate and unregulated secretion of insulin and can be transient or permanent. Transient forms associated with maternal diabetes mellitus, rhesus isoimmunization, IUGR, perinatal asphyxia, and BWS can last for days to months and may require diazoxide treatment.

Genetic forms of hyperinsulinism (HI) can histologically be divided into focal and diffuse forms that are inherited in a sporadic or autosomal manner, respectively. The estimated incidence is 1 in 30,000 births and a genetic diagnosis is possible for 40% to 50% of individuals. At present, mutations have been described in 14 genes, which include genes encoding the following: ion channels: *ABCC8, KCNJ11, CACNA1D*; enzymes: *GLUD1, GCK, HADH, PGM1, HK1 PMM2*; transporters: *SLC16A1, UCP2*, and transcription factors: *HNF1A, HNF4A, FOXA2*.[441–443]

ABCC8 and *KCNJ11* encode for the subunits (Kir6.2 and SUR1) of the pancreatic β-cell ATP-sensitive K^+ channel. Due to its sensitivity to ATP, K^+ channels convert intracellular metabolic signals into membrane excitability in the following cascade: glucose is transported into the cell by glucose transporters; intracellular glucokinase catalyzes an essential irreversible step for glycolysis; intracellular glucose metabolism raises the cytosol ATP/ADP ratio, which inhibits the SUR1 component of the K^+ channel, resulting in K^+ channel closure, which subsequently results in membrane depolarization, opening of voltage-gated Ca^{2+} channels, and Ca^{2+} influx, which triggers the release of insulin from stored granules. Mutations in *ABCC8* and *KCNJ11* can thus result in oversecretion of insulin.[444]

The HI/hyperammonemia syndrome (HI/HA) is due to heterozygous gain-of-function mutations in *GLUD*, which encodes the glutamate dehydrogenase (GDH) enzyme, and is characterized by both fasting hypoglycemia and hyperinsulinemic hypoglycemia due to protein ingestion. Mutations in the enzyme hydroxyacyl coenzyme A dehydrogenase (*HADH*) also cause protein-sensitive HI and are relatively common in consanguineous populations after *ABCC8/KCNJ11* mutations have been excluded.[445]

Certain syndromes have been associated with CHI with varying mechanisms underlying the HI. These include overgrowth syndromes (BWS, Sotos, Simpson-Golabi-Behmel, Perlman), chromosomal abnormalities (Trisomy 13, mosaic Turner syndrome), growth failure syndromes (Kabuki, Costello, Rubenstein-Taybi), syndromes with abnormal calcium homeostasis (Timothy syndrome), congenital disorders of glycosylation, insulin resistance/Donohue syndrome (*INSR*), congenital hypoventilation syndrome (*PHOX2B*), and a contiguous gene deletion including *ABCC8* (Usher syndrome).[443,446]

At the time of hypoglycemia, a hypoglycemia screen including glucose, insulin, C-peptide, acetoacetate, and β-hydroxybutyrate will easily identify CHI. Diagnostic criteria are a glucose infusion rate of >8 mg/kg/minute, a laboratory blood glucose <3 mmol/L with detectable serum insulin or C-peptide, low serum ketone bodies, and low serum fatty acids.[447] First-line treatment is diazoxide and chlorothiazide, and second-line treatment is octreotide or glucagon. Patients who are diazoxide insensitive will need further investigation with a combined computed tomography (CT)/18F-DOPA positron emission tomography (PET) scan to assess for focal or diffuse lesions. Focal lesions can be cured with surgery, whereas diffuse lesions require near-total pancreatectomy, which often does not result in a cure and may lead to diabetes mellitus.[448]

Neonatal Diabetes. Neonatal diabetes mellitus (NDM) is defined as diabetes with onset before 6 months of age and has an approximate incidence of 1:100,000 live births. There are many different subtypes of neonatal diabetes defined by different genetic causes. We now know of 30 different genetic subtypes of NDM and a genetic diagnosis can be found for 82% of patients.[449–452]

Neonatal diabetes may be transient (TNDM) or permanent (PNDM) or part of a syndrome, with extrapancreatic features. TNDM accounts for 60% of NDM and normally spontaneously remits before 18 months of age or is so mild as not to require treatment. After remission, diabetes mellitus often recurs, usually during adolescence.[453] The majority of cases are due to overexpression of an imprinted gene on chromosome 6q24 (*PLAGL1/HYMAI* genes). Activating mutations of *KCNJ11* and *ABCC8*, encoding for the two subunits (Kir6.2 and sulfonylurea receptor 1 [SUR1]) of the pancreatic β-cell ATP-sensitive inward rectifier potassium (K-ATP) channel, account for 25% of cases.[454] Other rarer causes include mutations in the *HNF1B* gene (renal cyst and diabetes syndrome), encoding a transcription factor, *INS* gene, encoding preproinsulin and *ZPF57*, which regulates gene methylation.

PNDM is most often due to activating mutations in *KCNJ11* or *ABCC8*, resulting in overactivation of potassium channels and thus impaired insulin secretion. These patients are responsive to sulfonylurea therapy and do not require insulin treatment. Other nonsyndromic causes of PNDM include *INS* and *GCK* mutations. In approximately 50% of PNDM cases, the diabetes-causing gene encodes a transcription factor, and the mutation leads to defects in pancreas development and a decreased number of β cells with or without extrapancreatic features. To date, mutations in 12 different genes encoding transcription factors important for pancreatic development have been identified in patients with neonatal diabetes including *PDX1*, *PTF1A* (coding and distal enhancer regions), *GLIS3*, *PAX6, RFX6, NEUROD1, NEUROG3, GATA4, GATA6, NKX2-2, MNX1,* and *CNOT1*.[449,455,456]

Impaired cellular stress response is a known β-cell dysfunction mechanism and cause of early-onset diabetes. The most common PNDM syndromic subtype is Wolcott-Rallison syndrome, which is caused by autosomal recessive mutations in the EIF2AK3 gene (eukaryotic translation initiation factor 2 alpha kinase 3), encoding the pancreatic eIF2-alpha kinase (PERK). Wolcott-Rallison syndrome is characterized by neonatal/early-onset diabetes, skeletal dysplasia, liver dysfunction, cardiac and renal abnormalities, developmental delay, epilepsy, and neutropenia. More recently heterozygous *de novo* variants in EIF2B have been found to cause PNDM and transient liver dysfunction. These mutations disrupt the ability of the eIF2B complex to sense eIF2 phosphorylation, likely resulting in severe cellular/ER stress.[457]

Other recessively inherited syndromes with neonatal-onset diabetes as a component include Wolfram syndrome or DIDMOAD (DI, diabetes mellitus, optic atrophy, and deafness) caused by *WFS1* mutations, thiamine–responsive megaloblastic anemia syndrome (PNDM, thiamine–responsive megaloblastic anemia, sensorineural deafness) caused by mutations in a thiamine transport gene (*SLC19A2*), Roger syndrome and Fanconi-Bickel syndrome (PNDM, hypergalactosemia, liver dysfunction) caused by mutations in *SLC2A2* encoding GLUT2 transporter, and microcephaly with simplified gyration, epilepsy, and PNDM (MEDS) caused by *IER3IP1* mutations.[449]

A growing list of genes causing monogenic autoimmune syndromes, in which diabetes due to β-cell destruction is a common feature, have been identified: *FOXP3* mutations result in IPEX (immune dysregulation, polyendocrinopathy, enteropathy, on X chromosome) syndrome; biallelic LRBA mutations cause common variable immunodeficiency-8, characterized by early-onset severe autoimmune manifestations, enteropathy, lymphoproliferation, hypogammaglobulinemia, and recurrent respiratory tract infections; *IL2RA* mutations cause immunodeficiency 41 with lymphoproliferation and autoimmunity; heterozygous activating *STAT3* mutations result in early-onset polyautoimmunity; T1D, autoimmune hypothyroidism and celiac disease, and mutations in the nucleoside transporter *SLC29A3* cause histiocytosis-lymphadenopathy plus syndrome.[458]

Neonatal diabetes associated with monogenic severe insulin resistance and or lipodystrophy is far less common than monogenic β-cell failure and is caused by genetic defects affecting insulin receptor signaling (*INSR*; Donahue syndrome) or adipocyte function, for example, congenital lipodystrophy syndrome due to AGPAT2, BSCL2, CAV1, and PTRF mutations.[459]

Next-generation sequencing (NGS), using targeted gene panels, enables simultaneous analysis of multiple genes, vastly reducing costs and leading to even higher rates of mutation detection.[451,460] Whole-exome or whole-genome sequencing is subsequently being performed in cases when a pathogenic variant is not identified using the targeted NGS approach. Identifying the genetic causes of neonatal diabetes is not only relevant for patients but also provides great insights into the human pancreas and β-cell development and function, and the clinical heterogeneity of diabetes. Neonatal diabetes that does not respond to sulfonylurea therapy is best treated with continuous subcutaneous insulin infusion via a pump using rapid-acting insulin analogues rather than multiple daily injections as insulin doses are so small, and feeding is so frequent.[461,462]

KEY POINTS

- Pancreatic development is a complex process, in which morphologically and functionally distinct exocrine and endocrine tissue types are derived from common endodermal progenitor cells
- Insulin, glucagon, somatostatin, and pancreatic polypeptide are measurable by 8 to 10 weeks' gestation. Fetal energy needs are met by a continuous intravenous supply of glucose across the placenta
- Insulin, glucagon, somatostatin, and pancreatic polypeptide are measurable by 8 to 10 weeks of gestation

FETAL PARATHRYOID AND BONE METABOLISM

Parathyroid Development

During the fourth to fifth week of intrauterine development, the embryonic pharynx is marked externally by four branchial clefts of ectoderm origin and internally by five branchial pouches of endoderm origin. The branchial arches of mesoderm origin are in between them. Together these structures comprise the branchial apparatus, which normally undergoes involution leaving behind derivatives including the thyroid gland, parathyroid glands, thymus, ultimobranchial body, Eustachian tube, middle ear, and external auditory canal.[463]

The parathyroid glands develop as epithelial thickenings of the dorsal endoderm of the third and fourth branchial pouches. The superior parathyroid glands are derived from the fourth branchial pouch, which also gives rise to the thyroid gland. The third branchial pouch gives rise to the inferior parathyroid glands and the thymus. The parathyroid glands remain intimately connected with their respective branchial pouch derivatives and parathyroid gland development from the third and fourth pharyngeal pouches proceeds in synchrony with thyroid embryogenesis.[322,464]

The parathyroid glands develop between 5 and 15 weeks of gestation. The third pouches encounter the migrating thyroid anlage and the parathyroid anlagen are carried caudally with the thyroid gland, finally coming to rest at the lower poles of the thyroid lobes as the inferior parathyroid glands.[322,464] The fourth pouches encounter the thyroid anlage later and come to rest at the upper poles of the thyroid lobes as the superior parathyroid glands.[322,464] The fifth pouches contribute paired ultimobranchial bodies that are incorporated into the developing thyroid gland as the parafollicular or C cells that secrete calcitonin. The individual parathyroid glands increase in diameter from less than 0.1 mm at 14 weeks of gestation to 1 to 2 mm at birth and adult glands measure 2 to 5 mm in width and 3 to 8 mm in length.

Most of the parathyroid gland parenchyma is composed of chief cells (also known as the parathyroid gland principal cells). These play a critical role in calcium homeostasis by sensing changes in extracellular calcium concentration and releasing the appropriate amount of PTH to correct or maintain normal blood calcium concentrations. The chief cells spend most of their time inactive due to normal calcium homeostasis. Near term, fetal parathyroid cells are largely composed of inactive chief cells, with only a few intermediate chief cells containing occasional secretory granules. The calcitonin content of C cells in the neonatal thyroid gland is as high as 540 to 2100 mU/g of tissue, values as much as 10 times those observed in the normal adult gland.[465] Both endocrine systems are functional during the second and third trimesters.

Transcription Factors Involved in the Development of Parathyroid Glands

The molecular signaling pathways and transcription factors that are involved in determining the differentiation of the pharyngeal pouch endoderm into parathyroid cells are being elucidated by studies of appropriate mouse models and patients with hypoparathyroidism. Fig. 130.6 shows the gene cascade programming normal thyroid-parathyroid development. Deletions of genes encoding several transcription factors in mice result in developmental anomalies that include parathyroid defects, although such abnormalities have not yet been reported in men. The transcription factors include members of the Hox (Hoxa3 and Hox15, which result in parathyroid agenesis) and Pax (Pax1, Pax3 and Pax9) families, Pbx1 (a protein that acts as a cofactor of Hox transcription factors), Eya1, Six1, Gcm2, Crk1, and Tbx1 and Tbx2.[466,467]

In humans, the most common cause of isolated parathyroid aplasia and severe hypoparathyroidism is loss-of-function mutations in the *GCM2* (previously *GCMB*) gene, which is expressed exclusively in the parathyroid glands.[468,469] A study of 20 patients who had childhood-onset, permanent hypoparathyroidism, which was not due to a chromosome 22q11 deletion, showed that 35% had a mutation in one of three genes (*GATA3*, *GCM2*, or *CASR*). Among the 13 patients with syndromic hypoparathyroidism, which was associated with deafness and/or renal dysplasia, six (>45%) had a mutation involving *GATA3*, *GCM2*, or *CASR*, and among seven patients with nonsyndromic hypoparathyroidism, two (>25%) had a mutation involving *GCM2*.[470] These studies indicate that the likelihood of a genetic etiology in hypoparathyroidism is likely to be high. Gain-of-function mutations of *GCM2* are a cause of parathyroid hyperplasia and familial isolated hyperparathyroidism.[471,472] X-linked recessive hypoparathyroidism with complete agenesis of the parathyroid glands has been reported in two multigenerational, related kindreds with a deletion-insertion downstream of *SOX3* likely to exert a position effect on *SOX3* expression.[473] These findings point to a potential role for the *SOX3* gene in the embryologic development of the parathyroid glands from the pharyngeal pouches. Isolated hypoparathyroidism due to disorders of PTH synthesis or secretion can be found in association with mutations in genes expressing PTH, calcium-sensing receptor (CaSR), and GNA11 (α subunit of the G protein, Gα11 that couples the CSR to activation of intracellular signaling pathways in the parathyroid cell).[467,468]

Human disorders with congenital hypoparathyroidism as a component have been found in association with mutations and deletions in *GATA3* (hypoparathyroidism, deafness, renal dysplasia [HDR] syndrome), *AIRE1* (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy [APECED] syndrome or autoimmune polyglandular syndrome type 1 [APS1]), *CHD7* and *SEMA3E* (CHARGE syndrome), *TBCE* (Kenny Caffey syndrome), *PTH/PTHrP* (Blomstrand chondrodysplasia), and *GNAS1* (pseudohypoparathyroidism). *TBX1* (a DNA-binding transcription factor of the T-box family that is known to have an important role in vertebrate and invertebrate organogenesis and pattern formation) and *CRKL* mutations have been associated with 22q11 deletion syndrome.[474,475]

Fetal Calcium and Phosphate Metabolism

The endochondral skeleton begins forming in the embryo, but it is not until the third trimester that substantial bone is formed and mineralized. Studies in fetal sheep and monkeys and measurements in human preterm and term infants indicate that the majority (~80%) of calcium is transported/accreted during the third trimester.[476] The kidneys, intestines, and skeleton (through bone turnover) are not dominant sources of mineral supply for the fetus as they are for children and adults. Instead, the placenta meets the fetal mineral need by actively transporting calcium, phosphorus, and magnesium from the maternal circulation.[476]

Calcium and phosphate are maintained at a higher set point in the fetus than in the pregnant mother, especially in the third trimester, to meet the demands of fetal bone mineralization.[477] In late gestation, fetal serum concentrations of calcium and phosphate are maintained above the maternal concentrations by approximately 1.2 to 2.0 mg/dL (0.30 to 0.50 mmol/L) and 1.5 mg/dL (0.5 mmol/L), respectively.[478] Although the fetal concentration of Mg is also increased compared with the maternal value, the gradient is less (~0.12 mg/dL, 0.05 mmol/L) than those of calcium and phosphate.

Active transport of calcium across the placenta occurs against a concentration gradient and involves an ATP-dependent calcium pump.[479,480] Calcium enters the placental trophoblast at the brush-border membrane primarily via the TRPV6 channel. Transport is facilitated by calbindin D9K and calcium is released to the fetal circulation through the calcium pump PMCA3. Radiolabeled P^{32} transport studies utilizing rat placenta have revealed increasing inorganic phosphate (Pi) transfer across the placenta until term gestation. The placenta expresses phosphate-regulating genes, including sodium-phosphate channels (NaPi2a, NaPi2b, and NaPi2c) and FGF-23 co-receptor Klotho. Little is known of the mechanism of placental magnesium transport, although an *in vitro* perfusion study of isolated human placental lobules has suggested active transport as the major mechanism.[481]

The calcium concentration in the fetus appears to be largely independent of maternal calcium concentration. In animal models, despite maternal hypocalcemia induced by a calcium-restricted diet, vitamin D deficiency, parathyroidectomy, or a vitamin D receptor deletion model, fetuses were reported to be normocalcemic.

Regulation of Calcium Homeostasis by PTH/PTHrP

PTH is an 84-amino acid protein, produced by chief cells within the parathyroid glands in response to hypocalcemia. It stimulates bone resorption, increases renal tubular absorption, and promotes the synthesis of calcitriol (active 1,25$(OH)_2$D) in the kidneys, which in turn stimulates calcium uptake from the gut. A negative feedback loop occurs with rising ionized calcium and calcitriol concentrations regulating the release of PTH. The fetal sheep has low circulating concentrations of PTH but can increase serum PTH concentrations in response to a fall in serum calcium concentration induced by ethylenediaminetetraacetic acid (EDTA) and can respond promptly to infused calcium with increased serum calcitonin concentrations.[482] In this model, fetal parathyroidectomy decreases placental calcium transport and lowers fetal serum calcium.

The parathyroid glands are active in the human fetus and produce PTH from about 10 to 12 weeks of gestation. Maternal PTH concentrations are suppressed during pregnancy compared with those in the nonpregnant adult and fetal PTH concentrations appear to be strongly suppressed. This is due to the high fetal ionized calcium, which acts through the CaSR to suppress PTH. Despite this suppression, the release of fetal PTH can be stimulated by acute or chronic (e.g., from hypoparathyroidism) maternal hypocalcemia.[483] The placenta is impermeable to PTH, PTHrP, and calcitonin, but 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D (calcitriol) are transported across the placenta.[479] It is unknown whether the human placenta produces PTH.

PTHrP has PTH-like activity and has a major role in fetal bone development and metabolism as well as fetal calcium homeostasis. It is expressed in the uterus, placenta, amnion, umbilical cord, lactating breast, and fetal parathyroid glands. In the human fetus, the effect of PTH on placental calcium transport is uncertain; however, PTHrP is produced by many tissues, including the placenta and parathyroid, and acts on PTH receptors

The principal PTH receptor is PTHR1, which has an equal affinity for PTH and PTHrP. A second receptor PTHR2 is present in the CNS and binds PTH but not PTHrP, whereas PTHrP action in the placenta occurs likely through a distinct receptor that binds PTHrP but not PTH.[484,485]

The PTH concentration in the human fetus near the end of gestation is lower (<4.72 pg/mL) than maternal or nonpregnant adult concentrations. In cord blood, the concentration of PTHrP is 15-fold higher than that of PTH at term and plays a critical role in perinatal calcium homeostasis and maintenance of the higher fetal calcium set point.[480,486] Little PTH is detectable in fetal plasma using immunoassays, whereas bioassays showed significant bioactivity.[487] The principal factor responsible for this bioactivity, however, appears to be PTHrP rather than PTH. Although the fetal PTH concentrations are very low, fetal PTH is thought to be important because studies in fetal mice, rats, and sheep indicate that the absence of PTH causes hypocalcemia, hyperphosphatemia, low amniotic fluid mineral content, and reduced skeletal mineral content.

PTHrP knockout mice display increased ossification of the basal portion of the skull, long bones, vertebral bodies, pelvic bones, and mineralization of the normally cartilaginous portions of the ribs and sternum;

as a result of the cartilaginous mineralization, the animals die of asphyxiation in the early neonatal period.[480,486,488,489] They also display hypocalcemia, hypophosphatemia, reduced fetal-placental calcium transfer, normal amniotic fluid calcium content, and increased PTH.[486,490,491] The increase in PTH partly compensates for lack of PTHrP because the absence of both hormones leads to more marked hypocalcemia and a more substantial reduction in skeletal mineral content.[490] The placental calcium pump is stimulated by a mid-molecule portion of the PTHrP secreted by the fetal parathyroid gland and by other fetal tissues and the placenta, where it may exert a paracrine effect.[486,492] In mice, knockout of the gene for PTHrP abolishes the maternal-fetal calcium gradient, and placental transport of calcium is reduced. Placental calcium transport in these models is restored by the mid-molecule fragment of PTHrP (amino acids 67 to 86) but not by PTH or PTHrP fragment (1 to 34), which activate the PTH/ PTHrP receptor.[486,492]

Overall, it is clear that PTH and PTHrP both play roles in fetal bone and mineral homeostasis because the absence of either hormone leads to adverse effects. Fetal mice lacking parathyroid glands, PTH, or the PTH/PTHrP receptor are hypocalcemic and have undermineralized skeletons.[477,490,493] *Hoxa3* null mice, which lack the parathyroids and thymus, and are thus completely devoid of PTH,[303] had more profound hypocalcemia than in *Pthrp* null fetuses such that the blood calcium level was reduced well below the ambient maternal calcium concentration.[490,491] *Hoxa3* nulls also had hypomagnesemia, hyperphosphatemia, a normal rate of placental calcium transfer, low amniotic fluid mineral content, and reduced skeletal calcium and magnesium content. Animals with combined *PTH* plus *PTHrP* (*Hoxa3/ Pthrp* double mutants) or *PTH/PTHrP receptor* gene ablation have much lower blood calcium concentrations and a more severe skeletal chondrodysplasia than the single mutants.[493,494]

To determine the specific role of PTH in fetal calcium regulation, two different genetic models of PTH deficiency were studied. *Pth*-null mice have enlarged parathyroids that are incapable of making PTH, whereas *Gcm2*-null mice lack parathyroids but have PTH that arises from the thymus. *Pth*-nulls served as a model of complete absence of PTH, whereas *Gcm2*-nulls are a model of severe hypoparathyroidism. The *Pth*–null model demonstrates that PTH contributes to fetal mineral homeostasis because a fetal hypoparathyroid phenotype results, with hypocalcemia, hypomagnesemia, hyperphosphataemia, low amniotic fluid mineral content, and reduced skeletal mineral content in its absence.[477] PTH was found to regulate the placental expression of genes involved in calcium and other solute transfer and may contribute to the regulation of placental calcium transfer. These actions may be through both endocrine/systemic (parathyroid-derived) and paracrine (placental-derived) pathways.

Although hypocalcemia occurs when PTH is absent (*Pth, Gcm2,* and *Hoxa3* null mice), there is no compensatory increase in PTHrP and no alteration in placental calcium transfer.[477] These findings indicate that PTHrP must be regulated differently from PTH during fetal life. The only situation in which PTHrP has been found to be increased is in Pthr1 null fetuses, which lack the PTH/PTHrP receptor.[490] PTHrP may be produced autonomously by the placenta or regulated by other factors, such as the sensing of the calcium content exchanged across the trophoblasts and intraplacental yolk sac.[495]

Knockout of the mouse gene for PTH/PTHrP also results in hypocalcemia in the presence of normal or increased placental calcium transport,[479,492] suggesting that other factors are also involved in the maintenance of fetal serum calcium concentrations. PTH and PTHrP, through the PTH/PTHrP receptor, presumably modulate fetal skeletal calcium flux, calcium excretion through the fetal kidney, and perhaps reabsorption of calcium from amniotic fluid.

The fetal parathyroid-placental axis promotes maternal-fetal transfer of bone mineral and accretion of fetal bone mineral. It seems likely that fetal PTH and presumably PTHrP act on the fetal kidney to stimulate 1α-hydroxylation of 25-hydroxyvitamin D (25OHD) and that 1,25-dihydroxyvitamin D (1,25(OH)$_2$D) participates in modulating placental calcium transport. 1,25(OH)$_2$D or 24,25(OH)$_2$D also play a role in fetal cartilage growth and bone mineral accretion.[496] Thus PTHrP and to a lesser extent PTH in the ovine fetus appear to augment maternal-to-fetal calcium transport across the placenta and thus provide for the high rate of bone mineral accretion in the latter half of pregnancy.

1,25-Dihydroxyvitamin D. During pregnancy, maternal 25OHD crosses the placenta, so the fetal concentrations of 25OHD reach 75% to 100% of the maternal concentration at term.[497] Fetal 1,25(OH)$_2$D concentrations, however, are lower than that in the mother (<50%).[498] The maternal concentration of 1,25(OH)$_2$D during pregnancy is 2- to 3-fold higher than that in the nonpregnant adult. It is suggested that the maternal kidneys contribute to the abundance of 1,25(OH)$_2$D in the pregnant mother.[499] Nephrectomy in the fetal sheep reduces fetal serum calcium concentrations, and this can be prevented by prior administration of 1,25-dihydroxyvitamin D. Moreover, infusion into the sheep fetus of antibody to 1,25(OH)$_2$D reduced the placental calcium gradient.[479] The fetal kidney can produce 1,25(OH)$_2$D and the placenta contains 1,25(OH)$_2$D receptors, as well as a vitamin D–dependent calcium-binding protein.[489] It is thought that the synthesis of 1,25(OH)$_2$D in the fetus is suppressed by the high concentrations of calcium and phosphate and low concentrations of PTH.

Calcitonin. The high blood concentrations of calcitonin in the fetus, probably resulting from the chronic stimulation by fetal hypercalcemia, are thought to contribute to the fetal bone mineral accretion.[479,493] The predominant effect of calcitonin is to inhibit bone resorption and the high fetal serum calcium concentrations coupled with high circulating calcitonin promote bone mineral anabolism.[493] Placental calcitonin production may contribute to the calcitonin in fetal plasma, but the persistence of high plasma concentrations in neonatal plasma argues for predominant fetal production. Thyroid C cells begin to differentiate at around the 12th week of gestation,[500] and calcitonin is detectable at around the 15th week of gestation. Fetal circulating calcitonin concentrations are ~2-fold higher than those in the mother.[478,501,502] The trophoblasts of the placenta also produce calcitonin and supply it to the fetus.[503] Because calcitonin does not cross the placenta in the mouse or rat,[504] fetal circulating calcitonin is derived from fetal sources. The major source of fetal circulating calcitonin has not been determined. In the mouse, knockout of the calcitonin gene or the vitamin D receptor gene leads to postnatal osteosclerosis or osteomalacia, respectively; the pups appear normal at birth.[505]

Calcium-Sensing Receptor and FGF23. The CaSR plays a pivotal role in systemic calcium metabolism by modulating the production and secretion of calcium-regulating hormones such as PTH, calcitonin, FGF23, vitamin D, and urinary calcium excretion.[506,507] It responds to the high fetal serum calcium concentration to keep PTH suppressed. The CaSR is present in parathyroid glands, renal tubules, bone and cartilage, and many other tissues. Magnesium binds to the CaSR and influences PTH secretion. Mutations within the *CaSR* result in either inactivation or overactivation of the receptor resulting in hypercalcemia or hypocalcemia, respectively.[508]

FGF23 is the principal hormone that regulates phosphate transport. It is primarily produced and secreted into the circulation by osteocytes and, after its activation through glycosylation by GALNT3, acts through FGFR1c and FGFR4 on renal tubules to increase phosphate excretion through the sodium-phosphate exchanger (NaPi-IIC, SLC34A3).[509]

Alpha-Klotho acts as a cofactor for FGF23 to bind to FGFR1c, increasing the specificity of FGFR1c for FGF23. FGF23 is inactivated by cleavage by subtilisin/furin-like enzyme. PHEX (phosphate regulating gene with homologies to endopeptidases on the X chromosome) regulates cleavage of FGF23, and *PHEX* mutations therefore render FGF23 constitutively active. Besides increasing renal tubular phosphate excretion, FGF23 inhibits 1α-hydroxylase, thus reducing 1,25 vitamin D activity. Although FGF23 is present in the fetal circulation at concentrations that may equal adult values and there is robust expression of FGF23 target genes in the placenta and fetal kidneys, FGF23 itself may not be an important regulator of fetal phosphorous metabolism. In a study by Ma et al., FGF23 deficiency (*Fgf23*-null) and FGF23 excess (*Phex*-null) did not alter fetal phosphorus or skeletal parameters.[510]

Neonatal Disorders of Calcium and Phosphate Homeostasis

Some conditions affecting calcium homeostasis are particularly important in the neonatal period. When a neonate presents with abnormal calcium concentrations, assessment of calcium, albumin, phosphate, creatinine, alkaline phosphatase, vitamin D, and urine Ca-creatinine ratio and phosphate reabsorption in both the child and parents will help formulate the underlying diagnosis.

Maternal hypocalcemia can result in increased PTH secretion from the fetal parathyroids. This will maintain a normal serum calcium in the fetus but can lead to hyperplasia of the parathyroids, skeletal resorption, loss of mineral content, and fractures occurring *in utero* or during delivery. In contrast, maternal hypercalcemia results in increased placental calcium transport and suppression of the fetal parathyroids. This may have negative impacts after birth, with hypoparathyroidism of the neonate that can be prolonged or permanent.[476]

Hypocalcemia resulting from fetal and maternal vitamin D deficiency is relatively common. Occasionally, severe neonatal vitamin D deficiency can be associated with a dilated cardiomyopathy, which is reversible.[511,512] Vitamin D–dependent rickets type I (VDDR-I) is an autosomal recessive disease characterized by low or undetectable levels of $1,25(OH)_2D_3$ resulting in hypocalcemia, rickets, osteomalacia, growth retardation, and failure to thrive. The disease is caused by mutations in the *1α-OHase* gene.[513] In clinical practice, patients are treated with vitamin D analogues and/or an enriched Ca^{2+} diet to prevent rickets, which correct the major perturbations in Ca^{2+} homeostasis.

Heterozygous activating mutations of the *CASR* gene lead to autosomal-dominant hypocalcemia (ADH) type 1. In this condition, the sensitivity of the CASR to extracellular ionized calcium is increased. Consequently, PTH synthesis and secretion are suppressed at normal ionized calcium concentrations. Patients present with hypocalcemia, hyperphosphatemia, low magnesium levels, and low or low-normal levels of PTH increased urinary calcium excretion and sometimes with pseudo-Bartter syndrome. The differential diagnosis consists of the hypercalciuric hypocalcemia syndrome type 2, involving *GNA11* and other hypoparathyroidism etiologies.

Hypercalcemia disorders related to inactivating mutations of the *CASR* gene are either heterozygous (autosomal-dominant familial hypocalciuric hypercalcemia [FHH] type 1) or homozygous (severe neonatal hyperparathyroidism). FHH can present in neonates with mild to moderately elevated calcium concentrations, high normal (nonsuppressed) PTH, and a low urine calcium/creatinine ratio.[514] FHH may also present as neonatal hyperparathyroidism leading to bone disease and variable symptoms of hypercalcemia. This occurs especially if an affected infant has inherited the inactivating *CaSR* mutation from the father while their mother is normal. The fetus will consider the mother as hypocalcemic and therefore overproduce PTH. Cinacalcet therapy may help bind calcium and settle the hyperparathyroidism but parathyroidectomy is occasionally necessary. The differential diagnosis consists of the hypocalciuric hypercalcemia syndrome, types 2 (involving *GNA11*) and 3 (involving *AP2S1*), other hyperparathyroidism etiologies and abnormalities of vitamin D metabolism involving *CYP24A1* and *SLC34A1* genes.[515,516] Recently, mutations in the gene encoding the calcium channel TRPV6 were identified as a cause of transient neonatal hyperparathyroidism (TNHP).[517]

Neonatal subcutaneous fat necrosis is a rare granulomatous condition that occurs following perinatal asphyxia. It can result in extrarenal synthesis and release of $1,25(OH)_2D$ from macrophages, leading to hypercalcemia.[518] Hyperhydration, steroid therapy, low calcium milk, and occasionally bisphosphonate treatment may be needed to reduce calcium concentrations.

Metabolic bone disease (MBD) of prematurity remains a significant comorbidity in preterm and low birth weight neonates. It most commonly occurs as a result of inadequate calcium and phosphorus stores exacerbated by inadequate intake and the high degree of skeletal growth occurring in the weeks following birth. Biochemical changes commonly include hypophosphatemia, hyperphosphatasia, and secondary hyperparathyroidism, which may be accompanied by rachitic changes and/or fractures.

Loss of FGF23 or components of the FGF23 network (GALNT3, FGF23, alpha-Klotho) cause hyperphosphataemia, extraskeletal calcifications, and early mortality; excess FGF23 or pathway components (FGF23, PHEX, SLC34A3, subtilisins [PCSKs], phosphate transporters NPT2a and NPT2c) cause hypophosphatemia with rickets or osteomalacia.[519]

KEY POINTS

- Active calcium transport across the placenta occurs in the third trimester to maintain fetal calcium concentrations and is dependent on PTHrP.

FETAL GROWTH

The hormones most important for postnatal growth, including T_4, GH, and gonadal steroids, have a limited role in fetal growth even though TR and GHR are expressed in many embryonic tissues including the growth plates.[520–522] Placental hormones including the human GH variant and hPL play a limited role; hPL may promote early embryonic growth and may stimulate IGF and insulin production.[522] IGF-1 and IGF-2 are produced by the placenta and may exert autocrine-paracrine actions on placental growth. IGF-1, IGF-2, and the IGF-1 and IGF-2 receptors are widely expressed in fetal tissues of mesenchymal, ectodermal, and endodermal origin and play a critical role in modulating normal fetal growth, including the nervous system.[521,523]

Insulin-Like Growth Factors

The IGFs are involved in regulation of uterine and placental growth during pregnancy; in early embryonic and fetal development, IGF-1, EGF, and estrogens are mitogens for endometrial stromal cells, and the endometrial contents of IGF-1 and *IGF1* mRNA are high at implantation and during early embryogenesis in the sow.[524] Uterine IGF-1 and *IGF1* mRNA levels decrease progressively with advancing gestation.[524] Placental tissue also contains *IGF1* and *IGF2* mRNAs, significant concentrations of the respective proteins, and IGF-1 receptors.[204] Autocrine and paracrine roles for the IGFs in uterine and placental tissues are postulated. IGF-1 and insulin are produced by embryonic tissues during the prepancreatic stage of mouse development, and both factors stimulate the growth of embryonic mouse cells.[525]

IGF2 is genomically imprinted and paternally expressed in the fetus and placenta. The mature IGF-2 protein is generated from the biologically inactive pro–IGF-2 peptide by the action of proprotein convertase 4. Recent studies have shown the role of IGF-2 in determining placental nutrient supply and hence fetal growth.[526] In mutant mice lacking the imprinted placental-specific *IGF2* transcript, growth of the placenta is altered from early gestation but fetal growth is normal until late gestation, suggesting functional adaptation of the placenta to meet the fetal demands. It is believed that this adaptation may be mediated by the altered expression of placental transporters GLUT3 and Slc38a4.[527]

Studies of transgenic mice with null mutations of the genes encoding IGF-1, IGF-2, or the IGF-1 receptor have defined the role of the somatomedins; the birth weight of the embryos lacking IGF-1 or IGF-2 was 65% of that of the control mice. When both genes were inactive, birth weight was reduced another 35%, and mice lacking IGF-1 receptor had birth weights averaging 55% of control values.[528] IGF-2–deficient mice also manifest intrauterine growth retardation in association with a small placenta. They have near-normal postnatal growth but delayed bone development.[528] IGF-2 receptor knockout fetal mice are 30% overweight, suggesting a negative growth modulating effect of this receptor.[523]

The normal growth in fetuses with both IGF-1 and IGF-2 receptor knockout is due to IGF-1 signaling via the insulin receptor; combined IGF-1, IGF-2, and insulin receptor knockout results in severe intrauterine growth retardation and fetal death. Knockout of individual IGF binding proteins has little effect on fetal or placental growth. In humans, mutations in *IGFI* or *IGF1R* are associated with intrauterine growth retardation, suggesting that IGF-1 signaling also contributes significantly to fetal growth in humans.[529,530] Apart from growth retardation, these children also suffer from developmental delay, microcephaly, hypoglycemia, and sensorineural hearing loss.[529,530] Indeed, there is a relation between umbilical IGF-1 concentration and birth weight in humans; maternal smoking reduces both umbilical IGF-1 and birth weight.[531,532] Studies in mice suggest that IGF-1 and IGF1R are required for late gestational lung maturation.[533]

Hypomethylation of the *11p15* imprinted region is associated with the phenotype of Silver-Russell syndrome (SRS).[534] This results in the relaxation of imprinting and biallelic expression of *H19* and downregulation of *IGF2*. Additionally, abnormal processing of IGF-2 by proprotein convertase 4 in the placenta has been implicated in the etiology of fetal growth restriction.[535] Pregnant women carrying IUGR fetuses had higher concentrations of pro–IGF-2 compared with controls. Severe intrauterine growth retardation and atypical diabetes mellitus secondary to insulin resistance were indeed associated with disruption of the regulation of the *IGF2* gene due to a human translocation.[536] On the other hand, overexpression of *IGF2* as a result of loss of imprinting associated with paternal uniparental disomy, *CDKN1C* loss of function, alteration in the *KvLQT1* DMR or microdeletions in the human *H19* DMR are associated with overgrowth in BWS.[537]

IGF-binding proteins are present as early as 5 weeks of gestation, and prenatally as well as postnatally, the IGFs circulate associated with binding proteins.[538] High concentrations of circulating IGFBP-1 are associated with fetal growth restriction in the mouse, as is overexpression of fetal human IGFBP-1.[539,540] High concentrations of IGFBP-4 have been shown to be related to fetal growth restriction.[541] IGFBP-4 is expressed exclusively in the maternal decidua and inhibits IGF action. Thus during fetal and postnatal life, plasma concentrations of IGFs are relatively high compared with tissue concentrations. In the fetus, IGF-2 concentrations are five to six times higher than those of IGF-1, in contrast to these concentrations in children and adults, and concentrations of both increase progressively throughout gestation.[542] Fetal concentrations of both peptides at term are 30% to 50% of adult concentrations. In most studies, cord blood IGF-1 concentrations correlate with birth size.[523]

Despite the fetal growth-enhancing effects of IGF-2, blood concentrations are only weakly related to size at birth, largely because of the inhibiting effect of soluble IGF-2 receptor (IGF2R) but also because IGF-2 appears to exert most of its growth effects in the earlier part of gestation.[523] Soluble IGF2R is derived through proteolytic cleavage of the transmembrane region of the receptor in many tissues. IGF receptors have been identified as early as 5 weeks of gestation and are widespread in fetal tissues.[541] IGF-1 stimulates glycogenesis in cultured fetal rat hepatocytes and induces the formation of myotubes in cultured myoblasts. IGF-2 is active in cultured muscle and neonatal rat astroglial cells. Insulin receptors are increased in fetal cells and are resistant to downregulation; no similar data are available for the IGF-1 receptor.

The control of IGF production differs in fetal and postnatal life (Fig. 130.10). GH, which stimulates IGF-1 production after birth, has a limited role in fetal IGF production.[520] GH does play a minor role in fetal growth as reflected in the low IGF concentrations and slight reduction in birth weight in infants with GH resistance (Laron dwarfism) and GHD.[543]

Receptors for hPL are abundant in fetal tissues.[522,544] hPL stimulates IGF-1 production and augments amino acid transport and DNA synthesis in human fetal fibroblasts and muscle cells.[522,544] IGF-1 and IGF-2 concentrations are reduced in fetuses of protein-starved pregnant rats and the low IGF-2 concentrations are reversed by hPL.[545] Thyroidectomy of the third-trimester sheep fetus impairs skeletal muscle growth in association with a decrease in muscle *GH receptor* mRNA and *IGF1* mRNA without an effect on *IGF2* concentrations.[546] Glucocorticoids can inhibit fetal growth, presumably by inhibiting *IGF* gene transcription.[546] Evidence also indicates that insulin may play a

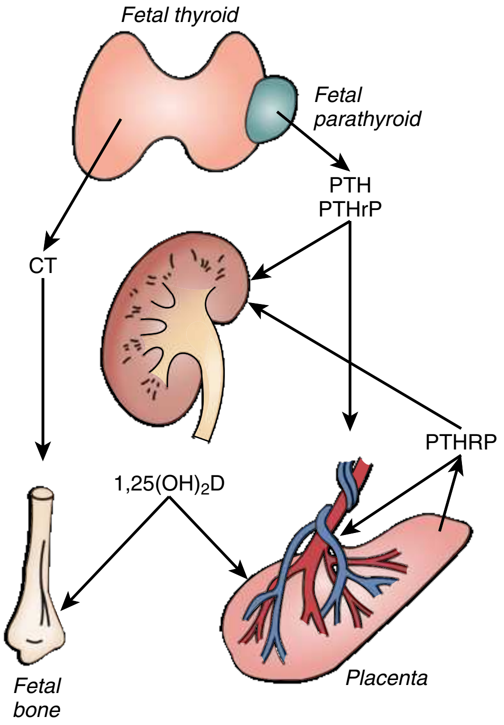


Figure 130.9 Hormonal modulation of fetal calcium homeostasis. Parathyroid hormone *(PTH)* and PTH-related peptide *(PTHrP)* are derived from both fetal parathyroid and placental tissues and stimulate maternal-to-fetal calcium transport. PTHrP also may stimulate renal 1-hydroxylation of 25(OH) vitamin D, which also acts to promote placental calcium transport. *CT,* Calcitonin.

role in modulating IGF synthesis in the fetus.[522,547] Nutrition is the major factor modulating IGF production in the fetus. IGF concentrations fall in suckling rats deprived of milk, and IGF-1 and IGF-2 concentrations are reduced in fetuses of protein-starved pregnant rats and placentally restricted sheep.[522,548] Light stimuli may alter circulating and brain IGF-1 concentration and control neuronal migration through increased IGF-1 signaling.[549] These data support the view that the IGFs are important in embryonic and fetal growth and that in the fetus they are regulated, at least in part, by hPL and by nutritional substrate derived transplacentally. The high concentrations of IGF-2 in fetal rat serum, the high levels of *IGF2* mRNA in fetal tissues, and the presence of a truncated form of IGF-1 in human fetal brain tissue suggest unique developmental actions of these peptides.

Insulin

Insulin has a growth-promoting effect, particularly late in gestation.[521,550] The hyperinsulinemia commonly observed in the fetus of a diabetic mother is associated with augmented body weight and increased body fat deposition. A fetus with pancreatic agenesis and hypoinsulinemia is small and relatively devoid of body fat.[521,547] Insulin infusion to fetal monkeys near term increases body weight, with prominent anabolic effects on the heart, liver, and spleen.[550] Interestingly, homozygosity for a null mutation of the insulin receptor gene in fetal mice leads to early neonatal death with hyperglycemia and ketonemia; birth weight, however, is normal.[551] A homozygous insulin receptor mutation in the human fetus results in severe insulin resistance in association with intrauterine growth retardation, lack of subcutaneous fat, decreased muscle mass, hirsutism, and early death.[459,552] Presumably the difference in growth effect relates to the relatively longer pregnancy in the human species. Treatment with IGF-1 improves the clinical condition to some extent.[459]

Fibroblast Growth Factor

The FGF family has been characterized and includes at least 18 different mammalian FGFs and 4 receptor species.[553] Knockout of Fgf4 and Fgfr1 is lethal. Knockout of Ffg10 leads to pulmonary agenesis and neonatal death. *Fgf2* or *Fgf3* gene knockout in mice produces limb abnormalities, but fetal growth is not impaired. In mice, Fgfr3 knockout leads to chondrocyte hypertrophy and increased bone length.[554] Infusion of FGF into the renal arteries of rats containing a renal subcapsular transplant of a 10-day rat embryo has been shown to stimulate the growth of the embryo.[555,556] The major effects were on tissues of endodermal and mesodermal origin; FGF antiserum inhibited the growth of endodermal tissue in the transplanted embryos, as well as in some of the mesodermal-derived tissues. Thus the FGFs appear to have autocrine-paracrine roles in localized organ growth and differentiation, including an essential role in bone growth and maturation.[553,557]

In humans, a variety of gain-of-function *FGFR* mutations are associated with achondroplasia (*FGFR3*) and craniosynostosis (*FGFR2*).[521] Loss-of-function mutations in *FGFR1* and *FGF8* are associated with KS or hypogonadotropic hypogonadism, and FGF8 mutations are also associated with HPE.[44,558,559]

During calorie restriction in mice, increased FGF21 causes growth attenuation and GH insensitivity. High concentrations of FGF21 may directly suppress growth plate chondrocyte proliferation and differentiation. A negative correlation between FGF21 concentration and linear growth has been reported, and high concentrations of FGF21 were described in association with growth failure in preterm infants.[560]

Other Growth Factors and Signaling Pathways

Additional growth factors involved in fetal growth and organ maturation include the TGF-β superfamily, the vascular endothelial growth factor (VEGF) family, and the platelet-derived growth factor (PDGF) family.[561–564]

The TGF-β superfamily of extracellular growth factors comprises more than 35 members, including TGF-β, the bone morphogenic proteins (BMPs), growth and differentiation factors, activins, inhibins, MIS, nodal, and leftys.[561,562] These ligands activate some 12 transmembrane serine/threonine kinase receptors expressed in a variety of tissues. The family is critical for early embryonic development, left-right asymmetry, heart and vascular system development, craniofacial development, nervous system development, and skeletal morphogenesis and plays an important role in body composition and growth.[561,562] Receptor knockouts in mice generally are lethal *in utero* or during the perinatal period.

The PDGF family is composed of homo- and heterodimers of PDGF A and B chains encoded by two genes. The ligands activate two receptors, α and β, distributed in fibroblasts and in smooth muscle, glial, and endothelial cells.[564] *PDGFα* and/or *β* knockout in mice usually is lethal or results in postnatal death associated with facial anomalies, spina bifida, skeletal anomalies, lack of pulmonary alveolar septa, and epidermal blistering.[563]

During development, VEGF is expressed in almost all tissues.204 The family includes five isoforms from a single gene and two receptor tyrosine kinases widely distributed in vascular endothelium. Receptor gene knockout of either receptor in mice is lethal.[563]

It is of interest that a growing number of these factors have been shown to be produced by the placenta. Thus as is the case for the IGFs, both endocrine and autocrine-paracrine routes of production and action of a variety of growth factors may prove to be important in individual fetal organ development and in overall fetal growth and differentiation.

Lastly, Wnt signaling, Notch signaling, BMP signaling, and hedgehog signaling play major roles in embryogenesis and fetal organ growth and development. These signaling pathways are also involved in bone development and growth and thus have a major effect on fetal size.[565]

KEY POINTS

- Fetal growth involves complex interactions between environmental, genetic, and epigenetic factors. IGFs are important for fetal growth and are regulated by transplacentally derived nutritional substrate. Hormones (T_4, GH, sex steroids), which are so important postnatally for growth, have a limited role in fetal growth.

ADAPTATIONS TO EXTRAUTERINE LIFE

Successful transition to extrauterine life requires that the fetus, previously poikilothermic, with the placenta subserving most of its metabolic needs, becomes homeothermic and self-sustaining. Crucial integration of a variety of neural and endocrine events is a prerequisite for this transition. The coordinated actions of adrenal cortical, adrenal medullary, and thyroid hormones facilitate lung function, cardiovascular adaptation, glucose homeostasis, and thermogenesis (Fig. 130.10). A longer-term extrauterine transition requires adaptation to an intermittent substrate supply, which requires activation of the hormonal control systems for maintenance of serum glucose and calcium homeostasis.

Cortisol-Mediated Adaptations

A modest increase in prenatal near-term circulating cortisol concentrations occurs in a variety of primates and other mammalian species as a result of increased fetal adrenal cortisol production and decreased metabolism of cortisol to cortisone.[190,213] This has been referred to as the *prenatal cortisol surge*. Plasma ACTH concentrations tend to remain stable, although the pulsatility of ACTH secretion may increase

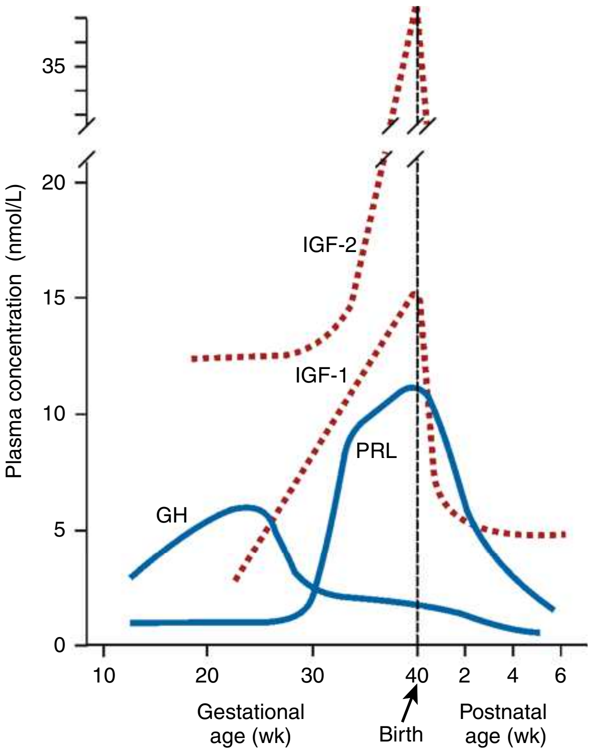


Figure 130.10 Pattern of developmental change in human fetal serum growth hormone *(GH)*, prolactin *(PRL)*, insulin-like growth factor-1 *(IGF-1)*, and IGF-2 levels during midgestation and the perinatal period. The major stimulus to the increasing IGF levels is nutrition. GH, PRL, and placental lactogen *(not shown)* have limited roles in fetal growth.

during this period. The increase in circulating cortisol concentrations is associated with a variety of physiologic responses. These include[213,214,566–571]:

- Increased conversion of T_4 to T_3 by stimulation of hepatic outer-ring iodothyronine deiodinase, leading to increased circulating T_3 concentrations
- Increased epinephrine secretion by sympathetic chromaffin tissue (including adrenal medulla) because of augmented phenylethanolamine transferase (PNMT) activity (which converts norepinephrine to epinephrine)
- Augmented synthesis and secretion of surfactant and maturation of surfactant composition; increases lung liquid reabsorption
- Stimulation of maturation of hepatic gluconeogenic enzyme activities
- Increased β-adrenergic receptor density in a variety of tissues, including lung, heart, and BAT
- Induction and maturation of a variety of gut enzymes for nutrient absorption and maturation of gut transport processes, motility, and structure.

The adaptational effects of the prenatal cortisol surge have led to the current recommendation for prenatal corticosteroid therapy in pregnancies threatened by the risk of preterm delivery. Generally, preterm infants prenatally exposed to augmented glucocorticoid concentrations have lower overall morbidity and mortality than untreated infants.[571] Although antenatal steroids for threatened preterm delivery have clear beneficial effects with a reduction in morbidity and mortality, the long-term consequences of such early high steroid exposure are unknown. The HPA axis is highly susceptible to programming during development and, unlike endogenous glucocorticoids, which are inactivated by placental 11βHSD2, synthetic glucocorticoids readily cross. Longitudinal follow-up of infants of women in preterm labor treated with betamethasone has shown evidence of insulin resistance particularly in female offspring. Antenatal treatment of the fetus of a mother whose previous child had CAH was with dexamethasone administered orally to the mother from 8 to 12 weeks' gestation. This was said to reduce fetal androgen production and hence virilization of female fetuses. There is however a paucity of high-quality evidence with respect to safety and efficacy and hence this should be regarded as experimental, and adverse associations have been reported including neurological abnormalities.[12]

Postnatally, biologically active cortisol predominates over cortisone in the circulation. Healthy term infants have initial pulsatility of cortisol; however, the circadian rhythm of ACTH and cortisol release in neonates is not established until about 3 months of age.[572] Because the circadian system develops *in utero* through exposure to maternal cues, preterm infants lose those cues at an earlier developmental stage than term-born babies.[21] These infants are then exposed to noisy and brightly lit neonatal intensive care units that contrasts markedly with the environment of the term-born infant discharged home within hours to a few days of birth.[573] Preterm infants are said to develop their rhythm around 8 to 12 weeks after being discharged from hospital[574] or at approximately 1 month corrected postnatal age.[575] Use of steroids postnatally to prevent or treat chronic lung disease (CLD) in preterm infants is associated with hyperglycemia, gastrointestinal bleeding, hypertrophic cardiomyopathy, growth failure, and cerebral palsy, and thus cannot be recommended.[576] Trials involving more physiological doses of steroids are underway. Early low-dose hydrocortisone treatment has been reported to improve survival in very preterm infants without pulmonary dysplasia. In an analysis of secondary outcomes of the PREMILOC (Early Low-Dose Hydrocortisone to Improve Survival Without Bronchopulmonary Dysplasia in Extremely Preterm Infants) randomized clinical trial, early low-dose hydrocortisone was not associated with a statistically significant difference in neurodevelopmental outcomes at 2 years of age.[577]

Catecholamine-Mediated Adaptations

Parturition is associated with enormous increases in circulating catecholamine concentrations, which are important for critical extrauterine transitional events. Cord-blood norepinephrine concentrations of 15 nmol/L (2500 pg/mL) and epinephrine values of 2 nmol/L (370 pg/mL) are common after spontaneous delivery of term infants.[578,579] These catecholamine concentrations increase cardiac output and neonatal blood pressure, evoke cardiac ventricular inotropic effects, increase glucagon and decrease insulin secretion, stimulate hepatic glycogenolysis, stimulate BAT thermogenesis, increase free fatty acid (FFA) concentrations, and promote pulmonary adaptation by mobilizing pulmonary fluid and stimulating surfactant synthesis (see Fig. 130.11).

BAT Thermogenesis

BAT is the major site for newborn thermogenesis.[566,580–582] Fetal adipogenesis involves proliferation of preadipocytes and differentiation into mature white- and brown-fat tissues.[583] This process is regulated by several transcription factors activating adipocyte-specific genes, and thyroid hormones regulate many of these genes.[583] Extirpation of this tissue during fetal life is associated with profound neonatal hypothermia. BAT function is coordinated through several endocrine systems. Catecholamines, via β-adrenergic receptors, stimulate fatty acid oxidation and thermogenesis. Mitochondria in BAT express a unique protein, thermogenin (or UCP), which uncouples oxidation from subsequent adenosine diphosphate phosphorylation and thereby promotes the release of heat. Thermogenin is T_3 dependent; BAT expresses MDI-2, which provides for local T_3 production from T_4.223 It was previously thought that BAT involutes soon after birth,

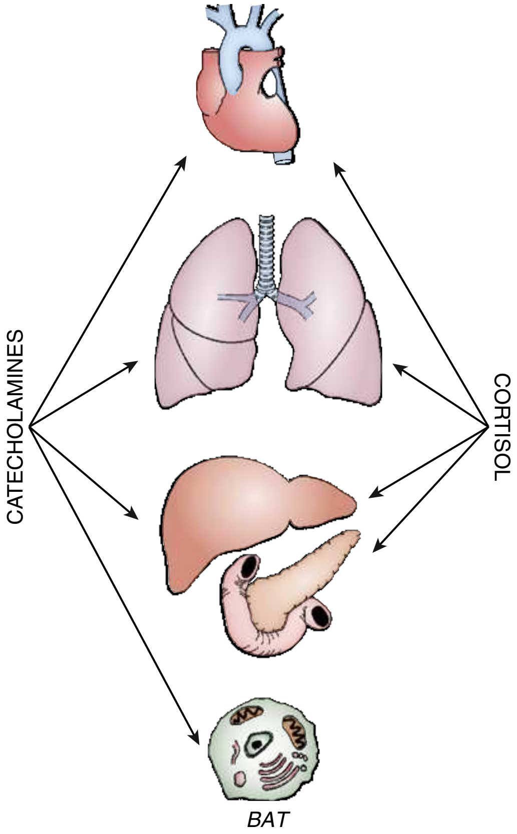


Figure 130.11 Coordinated effects of cortisol and catecholamines mediate the transition to extrauterine life, promoting extrauterine adaptational changes in cardiovascular, pulmonary, hepatic, and pancreatic function and stimulating brown adipose tissue *(BAT)* thermogenesis.

but combined ^{18}F-fluorodeoxyglucose PET (^{18}F-FDG-PET) and CT scans have identified active BAT in adults and shown a strong positive correlation between BAT activity and the basal metabolic rate.[584–586]

Glucose Homeostasis

During the newborn period, the placental source of glucose is abolished. A normal-term infant has an immediate postnatal fall in plasma glucose concentration by 1.4 to 1.7 mmol/L (25 to 30 mg/dL) to a nadir of about 3 to 3.3 mmol/L (55 to 60 mg/dL) by 1 to 2 hours of age; glucose concentrations then steadily rise over the first few days of life to return to the normal range for infants, children, and adults (3.9 to 5.6 mmol/L [70 to 100 mg/dL]).[587] Low glucose concentrations are usually transient, asymptomatic and part of the normal adaptation to extrauterine life. This relative hypoglycemia, along with the augmented catecholamine levels largely caused by the absence of placental degradation, stimulates pancreatic glucagon secretion. These early glucagon and catecholamine surges rapidly deplete available hepatic glycogen stores, so maintenance of blood glucose concentrations in the 12 to 24 hours after birth requires increased hepatic gluconeogenesis. Counterregulatory hormones rapidly become active with high catecholamine, glucagon, GH, and cortisol concentrations, and a fall in insulin, and these changes result in stabilization of blood glucose concentrations by approximately 48 to 72 hours of life.[587–589]

Glucagon increases in the first few hours of life, particularly with protein feeding, which stimulates gut glucagon release and pancreatic glucagon secretion.[588,589] In human infants, glucagon concentrations remain at around 100 ng/L for the first 12 to 24 hours and the glucagon:insulin ratio is high enough to stabilize blood glucose concentrations in the range 2.8 to 4 mmol/L (50 to 70 mg/dL).[589] The high insulin:glucagon ratio stimulates hepatic gluconeogenesis. Secretion of both glucagon and epinephrine appears to be activated in transitional neonatal hypoglycemia in normal neonates. It has been suggested that normal and small for gestational age (SGA) neonates activate neuroendocrine responses to hypoglycemia at plasma glucose concentrations in the range of 50 to 60 mg/dL (2.8 to 3.3 mmol/L).

Plasma glucagon concentrations rise in association with a rapid increase in functionally coupled glucagon receptors. The increase in plasma catecholamines coincident with parturition may be responsible for some of these changes; catecholamines both stimulate glucagon release and inhibit insulin release.[437,590,591] Phosphoenolpyruvate carboxykinase (PEPCK) activity also increases during this period. Thus gluconeogenesis is readily demonstrable in the newborn, where nearly 10% of glucose utilization is accounted for by gluconeogenesis from alanine.[439] The insulin-secretory response to hyperglycemia matures in the neonatal period, and healthy-term infants achieve adult-like regulation of glucose homeostasis within 5 to 7 days of birth.

Low glucose concentrations on the first few days of life are associated with suppressed levels of both ketones and FFA (range 0.3 to 0.5 mmol/L) in normal newborns, and these low levels of ketones and FFA appear to be explained by incomplete suppression of insulin release. The metabolic and hormonal profile of transitional neonatal hypoglycemia indicates that this is a form of hyperinsulinemic hypoglycemia in which the glucose set point for suppression of insulin secretion is reduced to 55 to 65 mg/dL (2.8 to 3.6 mmol/L).[587] Plasma FFA concentrations rise postnatally as a result of the effects of catecholamines and chemical thermogenesis. Oxidation of fatty acids probably provides cofactors (acetyl coenzyme A and the reduced form of nicotinamide adenine dinucleotide) required for gluconeogenesis, as well as sparing glucose for utilization by crucially dependent tissues such as the brain.[435] Failure of this sequence of physiological changes can lead to hypoglycemia, which is most common in the first few hours after birth. It has been reported that lactate may be neuroprotective in at-risk babies with hypoglycemia within the first 48 hours after birth, a time period when ketone bodies are unlikely to provide protection.[592] Premature infants may require up to 2 weeks to achieve glucose homeostasis. Relative hyperinsulinemia may occur in the neonatal period in an infant of a diabetic mother, in SGA and/or asphyxiated or severely stressed infants, or in infants with islet β-cell hyperplasia or CHI.

Neonatal hypoglycemia is common and is usually treated with supplemental feeds or intravenous glucose. Early feeding for prevention of hypoglycemia in at-risk babies (including infants of mothers with diabetes, small and large for gestational age, and preterm) is the mainstay of management. The Sugar Baby study showed that dextrose gel is effective in treating neonatal hypoglycemia without affecting breastfeeding.[593] The pre-hPOD study (hypoglycemia prevention with oral dextrose) reported the first randomized control trial of prophylactic oral dextrose in at-risk babies, showing reduced incidence of neonatal hypoglycemia.[594,595]

Calcium Homeostasis

In utero, active calcium transport from the mother to the fetus takes place through the placenta, responsible for approximately 20% to 30% of calcium stores in the fetus. Neonates must adjust rapidly from a high-calcium fetal environment regulated by PTHrP and calcitonin to a low-calcium environment regulated by PTH and vitamin D. With placental separation, newborn plasma calcium concentrations fall to values approximating 2.3 mmol/L (9 mg/dL) for total calcium and 1.2 mmol/L (4.8 mg/dL) for ionized calcium.[596,597] Plasma PTH concentrations are relatively low and minimally responsive to hypocalcemia during the first 2 to 3 days of life. Calcitonin concentrations, already high in cord blood, increase further and remain high for several days after birth.[597–599] The relatively obtunded PTH response and the high calcitonin concentrations lead to a 2- to 3-day period of transient neonatal hypocalcemia that inhibits calcitonin and stimulates PTH secretion.[598,600]

Calcium homeostasis is also affected in the newborn by a low glomerular filtration rate and reduced renal responsiveness to PTH for several days after birth. These factors limit phosphate excretion and, with high-phosphate milk intake, predispose the neonate to hyperphosphatemia.[598,600] Premature infants tend to have lower PTH and higher calcitonin concentrations, along with more immature kidney function, so neonatal hypocalcemia may be more marked and prolonged. Infants born to mothers with hypercalcemia due to hyperparathyroidism have a high incidence of symptomatic hypocalcemia. These infants have a more marked suppression of parathyroid function and a more prolonged period of transient hypoparathyroidism in the neonatal period. PTH secretion and calcium homeostasis usually normalize within 1 to 2 weeks in full-term infants, but normalization may require 2 to 3 weeks in small, premature infants.

Other Hormonal Adaptations

Delivery of the infant and severing of the placental circulation result in abruptly decreased concentrations of estrogens, progesterone, hCG, and hPL in the newborn.

The decrease in estrogens removes a major stimulus to newborn prolactin release, and PRL concentrations remain low over the next several weeks of life.

GH concentrations are also low in the postnatal period because of delayed hypothalamic-pituitary feedback and control.[62,601] These events probably involve changes in GH sensitivity to both GHRH and somatostatin. Newborn IGF-1 concentrations fall to low values, presumably because of removal of hPL stimulation and placental IGF production.

In male infants, after a transient fall in testosterone as the hCG stimulus decreases, pituitary LH secretion rebounds and there is a secondary surge in testosterone that persists for several weeks.[57,602] In females, a transient secondary surge in FSH may elevate estrogen concentrations.

The dramatic increase in neonatal TSH secretion during the first hour is largely due to the effect of body cooling with extrauterine exposure.[580] In term infants, the TSH surge peaks at 30 minutes at a concentration of about 70 mU/L. The increase in TSH in turn increases thyroidal T_4 and T_3 secretion and enhances thyroidal conversion of peripheral T_4 to T_3. Increased outer-ring (type I) iodothyronine deiodinase activity increases peripheral T_4-to-T_3 conversion.[569,580] T_3 concentrations are maintained in the range 1.6 to 3.4 nmol/L (105 to 220 ng/dL) postnatally. Neonatal serum TSH concentrations are reset to lower prevailing extrauterine values, which requires resetting of the thyroid hormone—TSH negative-feedback control system. T_3 and sulfated iodothyronine production decrease to extrauterine levels by 3 to 4 weeks of age, at which time these hormone concentrations approximate those in adults.[311,317]

KEY POINTS

- Immediate transition to extrauterine life involves the adrenal cortex and ANS. Longer-term transition to an intermittent nutrient supply and transient substrate deficiency requires maturation of secretory control mechanisms for the PTH-calcitonin system and the endocrine pancreas.

FUTURE DIRECTIONS

The management of fetal and neonatal endocrine disorders is an exploding field due to rapid advances in fetal imaging, minimally invasive techniques and genomics. These have revolutionized understanding of the unique endocrine milieu of pregnancy and the natural history of endocrinopathies. Additionally, insight into the mechanisms and pathways involved in developmental endocrinology offers management strategies for preterm infants and those with intrauterine growth retardation. There is growing experience with gene therapy in animals[603] and intraplacental gene therapy, for example, may be effective for growth restriction.[604] Technological progress will undoubtedly have a positive impact on clinical outcomes for patients in the future. This will inevitably bring with it medicolegal and ethical challenges, as well as considerations as to what constitutes innovation versus experimentation, and these complexities must be considered in order to deliver safe and excellent patient care.

REFERENCES

1. Sibley CP, Boyd RDH. Mechanism of transfer across the human placenta. In: Polin EA, Fox WW, eds. *Fetal and Neonatal Physiology*. 3rd ed. Philadelphia, PA: WB Saunders; 2003:111–122.
2. Sodha R, Proegler M, Schneider H. Transfer and metabolism of norepinephrine studied from maternal to fetal and fetal to maternal sides in the in vitro perfused human placental life. *Am J Obs Gynecol*. 1984;148:474–481.
3. Benediktsson R, Calder A, Edwards C, Seckl JR. Placental 11 beta-hydroxysteroid dehydrogenase: a key regulator of fetal glucocorticoid exposure. *Clin Endocrinol*. 1997;46(2):161–166.
4. Takeyama J, Sasano H, Suzuki T, Iinuma K, Pediatrics JT. 17β-Hydroxysteroid dehydrogenase types 1 and 2 in human placenta: an immunohistochemical study with correlation to placental development. *J Clin Endocrinol Metab Metab*. 1998;83(10):3710–3715.
5. Krysin E, Brzezin E. Divergent deiodination of thyroid hormones in the separated parts of the fetal and maternal placenta in pigs. *J Endocrinol*. 1997;155(2):295–303.
6. Murphy BEP. Cortisol and cortisone in human fetal development. *J Steroid Biochem*. 1979;11:509–513.
7. Roberts D, Brown J, Medley N, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev*. 2017;3:CD004454.
8. Benediktsson R, Lindsay RS, Noble J, Seckl JREC. Glucocorticoid exposure in utero: new model for adult hypertension. *Lancet*. 1993;341(8841):339–341.
9. Yau J, Olsson T, Morris R, Meaney M, Seckl J. Glucocorticoids, hippocampal corticosteroid receptor gene expression and antidepressant treatment: relationship with spatial learning in young and aged rats. *Neuroscience*. 1995;66(3):571–581.
10. Garite TJ, Kurtzman J, Maurel K, Clark R, Obstetrix Collaborative Research Network. Impact of a “ rescue course ” of antenatal corticosteroids : a multicenter randomized placebo-controlled trial. *YMOB*. 2009;200(3):248.e1–248.e9.
11. Speiser P, Azziz R, Baskin L, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95(9):4133–4160.
12. Hindmarsh P. Endocrine society congenital adrenal hyperplasia guidelines: great content but how to deliver? *Clin Endocrinol*. 2012;76(4):465–466.
13. Kaludjerovic J, Ward WE. The interplay between estrogen and fetal adrenal cortex. *J Nutr Metab*. 2012;2012:837901.
14. Roti E, Gnudi A, Braverman L. The placental transport, synthesis and metabolism of hormones and drugs which affect thyroid function. *Endocr Rev*. 1983;4:131–149.
15. Ghassabian A, Bongers-schokking JJ, de Rijke YB, et al. Maternal thyroid autoimmunity during pregnancy and the risk of attention deficit/hyperactivity problems in children. *The Generation R Study*. 2012;22(2):178–186.
16. Iisalo E, Castren O. The enzymatic inactivation of noradrenaline in human placental tissue. *Ann Med Exp Biol Fenn*. 1967;45(2):253–257.
17. Lei ZM, Rao C. Endocrinology of the trophoblast tissue. In: Becker K, ed. *Principles and Practice of Endocrinology and Metabolism*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:1096–1102.

18. Mesiano S, Jaffe R. Neuroendocrine-metabolic regulation of pregnancy. In: Strauss J, Barbieri R, eds. *Reproductive Endocrinology*. Philadelphia, PA: WB Saunders; 2004:327–366.
19. Goldsmith PC, McGregor WG, Raymoure WJ, Kuhn RW, Jaffe RB. Cellular localization of chorionic gonadotropin in human fetal kidney and liver. *J Clin Endocrinol Metab*. 1983;57(3):654–661.
20. Huhtaniemi IT, Korenbrot CC, Jaffe R. Content of chorionic gonadotropin in human fetal tissues. *J Clin Endocrinol Metab*. 1978;46:994–997.
21. Seron-Ferre M, Lawrence CC, Jaffe R. Role of hCG in regulation of the fetal zone of the human fetal adrenal gland. *J Clin Endocrinol Metab*. 1978;46:834–837.
22. Molsberry RL, Carr BR, Mendelson CR, Simpson E. Human chorionic gonadotropin binding to human fetal testes as a function of gestational age. *J Clin Endocrinol Metab*. 1982;55:791–794.
23. Kapcala L. Immunoassayable adrenocorticotropin in peripheral organs: concentrations during early development. *Life Sci*. 1985;37:2283–2290.
24. Reece EA, Hobbins J. *Clinical Obstetrics: The Fetus and Mother*. Wiley-Blackwell; 2007.
25. Volante M, Fulcheri E, Allìa E, et al. Ghrelin expression in fetal, infant, and adult human lung. *J Histochem Cytochem*. 2002;50(8):1013–1021.
26. Steculorum SM, Bouret S. Developmental effects of ghrelin. *Peptides*. 2011;32(11):2362–2366.
27. Martino E, Lernmark A, Seo H, Steiner DF, Refetoff S. High concentration of thyrotropin-releasing hormone in pancreatic islets. *Proc Natl Acad Sci U S A*. 1978;75(9):4265–4267.
28. Shibasaki T, Kiyosawa Y, Masuda A, et al. Distribution of growth hormone-releasing hormone-like immunoreactivity in human tissue extracts. *J Clin Endocrinol Metab*. 1984;59(2):263–268.
29. Wu P, Jackson I. Identification, characterization and localization of thyrotropin-releasing hormone precursor peptides in perinatal rat pancreas. *Regul Pept*. 1988;22:347–360.
30. Koshimizu T. The development of pancreatic and gastrointestinal somatostatin-like immunoreactivity and its relationship to feeding in neonatal rats. *Endocrinology*. 1983;112:911–916.
31. Polk D, Reviczky A, Lam R, Fisher DA. Thyrotropin releasing hormone in the ovine fetus: ontogeny and effect of thyroid hormone. *Am J Physiol Endocrinol Metab*. 1991;260(23):E53–E58.
32. Rahier J, Wallon J, Henquin J. Abundance of somatostatin cells in the human neonatal pancreas. *Diabetologia*. 1980;18:251–254.
33. Leduque P, Aratan-spire S, Czemichow P, Dubois P. Ontogenesis of thyrotropin-releasing hormone. *J Clin Investig Clin Investig*. 1986;78:1028–1034.
34. Saito H, Saito S, Sano T, Hosoi E, Saito H. Fetal and maternal plasma levels of immunoreactive somatostatin at delivery: evidence for its increase in the umbilical artery and its arteriovenous gradient in the fetoplacental circulation. *J Clin Endocrinol Metab*. 1983;56(3):567–571.
35. Koshimizu T, Ohyama Y, Yokota Y, Ohtsuka K. Peripheral plasma concentrations of somatostatin-like immunoreactivity in newborns and infants. *J Clin Endocrinol Metab*. 1985;61(1):78–82.
36. Robinson BG, Emanuel RL, Frim DM, Majzoub JA. Glucocorticoid stimulates expression of corticotropin-releasing hormone gene in human placenta. *Proc Natl Acad Sci U S A*. 1988;85(14):5244–5248.
37. Sirianni R, Rehman KS, Carr BR, Parker Jr CR, Rainey W. Corticotropin-releasing hormone directly stimulates cortisol and the cortisol biosynthetic pathway in human fetal adrenal cells. *J Clin Endocrinol Metab*. 2005;90(1):279–285.
38. Chakravorty A, Mesiano S, Jaffe RB. Corticotropin-releasing hormone stimulates P450 17alpha-hydroxylase/17,20-lyase in human fetal adrenal cells via protein kinase C. *J Clin Endocrinol Metab*. 1999;84(10):3732–3738.
39. Rehman KS, Sirianni R, Parker CR, Rainey WE, Carr BR. The regulation of adrenocorticotrophic hormone receptor by corticotropin-releasing hormone in human fetal adrenal definitive/transitional zone cells. *Reprod Sci*. 2007;14(6):578–587.
40. McLean M, Bisits A, Davies J, et al. A placental clock controlling the length of human pregnancy. *Nat Med*. 1995;1(5):460–463.
41. Ho Y, Liebhaber S, Cooke N. Activation of the human GH gene cluster: roles for targeted chromatin modification. *Trends Endocrinol Metab*. 2004;15:40–45.
42. Lacroix MC, Guibourdenche J, Frendo JL, Muller F, Evain-Brion D. Human placental growth hormone - a review. *Placenta2*. 2002;23:S87–S94.
43. Auffret J, Freemark M, Carre N, et al. Defective prolactin signaling impairs pancreatic beta-cell development during the perinatal period. *Am J Physiol Endocrinol Metab*. 2013;305:E1309–E1318.
44. McCabe MJ, Gaston-Massuet C, Tziaferi V, et al. Novel FGF8 mutations associated with recessive holoprosencephaly, craniofacial defects, and hypothalamo-pituitary dysfunction. *J Clin Endocrinol Metab*. 2011;96(10):E1709–E1718.
45. McCabe MJ, Dattani M. Genetic aspects of hypothalamic and pituitary gland development. *Handb Clin Neurol*. 2014;124:3–15.
46. Solov'ev GS, Bogdanov AV, Panteleev SM, Yanin V. Embryonic morphogenesis of the human pituitary. *Neurosci Behav Physiol*. 2008;38(8):829–833.
47. Ooi GT, Tawadros NER. Pituitary cell lines and their endocrine applications. *Mol Cell Endocrinol*. 2004;228(1):1–21.
48. Visser M, Swaab D. Life span changes in the presence of alpha-melanocyte-stimulating-hormone-containing cells in the human pituitary. *J Dev Physiol*. 1979;1(2):161–178.
49. Perry RA, Mulvogue HM, McMillen IC, Robinson P. Immunohistochemical localization of ACTH in the adult and fetal sheep pituitary. *J Dev Physiol*. 1985;7(6):397–404.
50. Facchinetti F, Panerai AE, Petraglia F, et al. Proopiomelanocortin-related peptides in feto-placental structures throughout pregnancy. *Exp Clin Endocrinol*. 1987;89(1):48–54.
51. Glickman JA, Carson GD, Challis J. Differential effects of synthetic adrenocorticotropin1-24 and alpha-melanocyte-stimulating hormone on adrenal function in human and sheep fetuses. *Endocrinology*. 1979;104:34–39.
52. Swaab DF, Martin J. Functions of alpha-melanotropin and other opiomelanocortin peptides in labour, intrauterine growth and brain development. *Ciba Found Symp*. 1981;81:196–217.
53. Sheng HZ, Westphal H. Early steps in pituitary organogenesis. *Trends Genet*. 1999;15(6):236–240.
54. Pearson CA, Placzek M. Development of the medial hypothalamus: forming a functional hypothalamic-neurohypophyseal interface. *Curr Top Dev Biol*. 2013;106(49–88).
55. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson I, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev 2009*. 2009;30:790–829.
56. Cohen LE, Radovick S. Molecular basis of combined pituitary hormone deficiencies. *Endocr Rev*. 2002;23(4):431–442.
57. Grumbach M, Gluckman P. The human fetal hypothalamus and pituitary gland: the maturation of neuroendocrine mechanisms controlling secretion of fetal pituitary growth hormone, prolactin, gonadotropins, adrenocorticotropin-related peptides and thyrotropin. In: Tulchinsky D, Little A, eds. *Maternal Fetal Endocrinology*. Philadelphia, PA: WB Saunders; 1994:193–261.
58. Treier MRM. The hypothalamic-pituitary axis: co-development of two organs. *Curr Opin Cell Biol*. 1996;8:833–843.
59. Davis S, Ellsworth B, Perez Millan M, et al. Pituitary gland development and disease: from stem cell to hormone production. *Curr Top Dev Biol 2013*. 2013;106:1–47. 1061-47.
60. Mollard P, Hodson DJ, Lafont C, Rizzoti KDJ. A tridemensional view of pituitary development and function. *Trends Endocrinol Metab*. 2012;23(6):261–269.
61. Binder G, Weidenkeller M, Blumenstock G, et al. Rational approach to the diagnosis of severe growth hormone deficiency in the newborn. *J Clin Endocrinol Metab*. 2010;95(5):2219–2226.
62. Gluckman P, Pinal C. Growth hormone and prolactin. In: Polin RA, Fox W, Abman S, eds. *Fetal and Neonatal Physiology*. Philadelphia, PA: WB Saunders; 2004:1891–1895.
63. DeVane GW, Porter J. An apparent stress-induced release of arginine vasopressin by human neonates. *J Clin Endocrinol Metab*. 1980;51:1412–1416.
64. Baum MA, Ruddy MK, Hosselet CA, Harris H. The perinatal expression of aquaporin-2 and aquaporin-3 in developing kidney. *Pediatr Res*. 1998;43(6):783–790.

65. Szarek E, Cheah P, Schwartz J, Thomas P. Molecular genetics of the developing neuroendocrine hypothalamus. *Mol Cell Endocrinol*. 2010;323:115–123.
66. Ezzat S, Mader R, Yu S, Ning T, Poussier P, Asa SL. Ikaros integrates endocrine and immune system development. *J Clin Invest*. 2005;115(4):1021–1029.
67. Blackshaw S, Scholpp S, Placzek M, Ingraham H, Simerly R, Shimogori T. Molecular pathways controlling development of thalamus and hypothalamus: from neural specification to circuit formation. *J Neurosci*. 2010;30(45):14925–14930.
68. Mulchakey JJ, DiBiasio AM, Martin MC, Blumenfeld Z, Jaffe RB. Hormone production and peptide regulation of the human fetal pituitary gland. *Endocr Rev*. 1987;8:406–425.
69. Rizzoti K, Lovell-Badge R. Early development of the pituitary gland: induction and shaping of Rathke's pouch. *Rev Endocr Metab Disord*. 2005;6:161–172.
70. Fauquier T, Rizzoti K, Dattani M, Lovell-Badge R, Robinson I. SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland. *Proc Natl Acad Sci U S A*. 2008;105:2907–2912.
71. Castinetti F, Davis SW, Brue T, Camper SA. Pituitary stem cell update and potential implications for treating hypopituitarism. *Endocr Rev*. 2011;32:453–471.
72. Dasen JS, Rosenfeld M. Signaling and transcriptional mechanisms in pituitary development. *Annu Rev Neurosci*. 2001;24:327–355.
73. Zhu X, Zhang J, Tollkuhn J, et al. Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis. *Genes Dev*. 2006;20(19):2739–2753.
74. Gregory LC, Dattani MT. The molecular basis of congenital hypopituitarism and related disorders. *J Clin Endocrinol Metab*. 2020;105(6):184.
75. Fang Q, George AS, Brinkmeier ML, et al. Genetics of combined pituitary hormone deficiency: roadmap into the genome era. *Endocr Rev*. 2016;37(6):636–675.
76. McCabe MJ, Hu Y, Gregory L, et al. Novel application of luciferase assay for the in vitro functional assessment of KAL1 variants in three females with septo-optic dysplasia (SOD). *Mol Cell Endocrinol*. 2015;417:63–72.
77. Morishima A, Aranoff GS. Syndrome of septo-optic-pituitary dysplasia: the clinical spectrum. *Brain Dev*. 1986;8(3):233–239.
78. Webb EA, Dattani MT. Septo-optic dysplasia. *Eur J Hum Genet*. 2010;18(4):393–397.
79. Xatzipsalti M, Voutetakis A, Stamoyannou L, Chrousos GP, Kanaka-Gantenbein C. Congenital hypopituitarism: various genes, various phenotypes. *Horm Metab Res*. 2019;51(2):81–90.
80. Dasen JS, Barbera JP, Herman TS, et al. Temporal regulation of a *paired*-like homeodomain repressor/TLE corepressor complex and a related activator is required for pituitary organogenesis. *Genes Dev*. 2001;15:3193–3207.
81. Alatzoglou KS, Gregory LC, Dattani M. Development of the pituitary gland. *Compr Physiol*. 2020;10:389–413.
82. Dattani MT, Martinez-Barbera JP, Thomas PQ, et al. Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat Genet*. 1998;19:125–133.
83. Thomas PQ, Dattani MT, Brickman JM, et al. Heterozygous HESX1 mutations associated with isolated congenital pituitary hypoplasia and septo-optic dysplasia. *Hum Mol Genet*. 2001;10:39–45.
84. Fang Q, George A, Brinkmeier M, et al. Genetics of combined pituitary hormone deficiency: roadmap into the genome era. *Endocr Rev*. 2016;37(6):636–675.
85. Fang Q, Benedetti AF, Ma Q, et al. HESX1 mutations in patients with congenital hypopituitarism: variable phenotypes with the same genotype. *Clin Endocrinol (Oxf)2*. 2016;85:408–414.
86. McNay DE, Turton JP, Kelberman D, et al. HESX1 mutations are an uncommon cause of septo-optic dysplasia and hypopituitarism. *J Clin Endocrinol Metab*. 2007;92:691–697.
87. Parks JS. Congenital hypopituitarism. *Clin Perinatol*. 2018;45(1):75–91.
88. Kiernan AE, Pelling AL, Leung KK, et al. Sox2 is required for sensory organ development in the mammalian inner ear. *Nature*. 2005;434:1031–1035.
89. Taranova OV, Magness ST, Fagan BM, et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev*. 2006;20:1187–1202.
90. Kelberman D, Rizzoti K, Avilion A, et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest*. 2006;116:2442–2455.
91. Kelberman D, de Castro SC, Huang S, et al. SOX2 plays a critical role in the pituitary, forebrain and eye during human embryonic development. *J Clin Endocrinol Metab*. 2008;93:1865–1873.
92. Stark Z, Storen R, Bennetts B, Savarirayan R, Jamieson RV. Isolated hypogonadotropic hypogonadism with SOX2 mutation and anophthalmia/microphthalmia in offspring. *Eur J Hum Genet*. 2011;19:753–756.
93. Williamson KA, Hyer AM, Rainger J, et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. *Hum Mol Genet*. 2006;15:1413–1422.
94. Alatzoglou KS, Andoniadou CL, Kelberman D, et al. SOX2 haploinsufficiency is associated with slow progressing hypothalamo-pituitary tumours. *Hum Mutat*. 2011;32:1376–1380.
95. Alatzoglou KS, Dattani MT. Genetic forms of hypopituitarism and their manifestation in the neonatal period. *Early Hum Dev*. 2009;85:705–712.
96. Giordano M. Genetic causes of isolated and combined pituitary hormone deficiency. *Best Pr Res Clin Endocrinol Metab*. 2016;30:679–691.
97. Stankiewicz P, Thiele H, Schlicker M. Duplication of Xq26.2-q27.1 including SOX3 in a mother and daughter with short stature and dyslalia. *Am J Hum Genet*. 2005;138:11–17.
98. Alatzoglou KS, Azriyanti A, Rogers N, et al. SOX3 deletion in mouse and human is associated with persistence of the craniopharyngeal canal. *J Clin Endocrinol Metab*. 2014;99(12):E2702–E2708.
99. Kurokawa D, Kiyonari H, Nakayama R, et al. Regulation of Otx2 expression and its functions in mouse forebrain and midbrain development. *Development*. 2004;131:3319–3331.
100. Ragge NK, Brown AG, Poloschek CM, et al. Heterozygous mutations in OTX2 cause severe ocular malformations. *Am J Hum Genet*. 2005;76:1008–1022.
101. Tajima T, Ohtake A, Hoshino M, et al. OTX2 loss of function mutation causes anophthalmia and combined pituitary hormone deficiency with a small anterior and ectopic posterior pituitary. *J Clin Endocrinol Metab*. 2009;94(1):314–319.
102. Shimo N, Yasuda T, Kitamura T, et al. Aniridia with a heterozygous PAX6 mutation in which the pituitary function was partially impaired. *Intern Med*. 2014;53:39–42.
103. Takagi M, Nagasaki K, Fujiwara I, et al. Heterozygous defects in PAX6 gene and congenital hypopitutarism. *Eur J Endocrinol*. 2015;172:37–45.
104. Bakrania P, Efthymiou M, Klein JC, et al. Mutations in BMP4 cause eye, brain and digit developmental anomalies: overlap between BMP4 and hedgehog signaling pathways. *Am J Hum Genet*. 2008;82:304–319.
105. Raivio T, Avbelj M, McCabe MJ, et al. Genetic overlap in Kallmann syndrome, combined pituitary hormone deficiency, and septo-optic dysplasia. *J Clin Endocrinol Metab*. 2021;97:E694–E699.
106. Webb E, AlMutair A, Kelberman D, et al. ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. *Brain*. 2013;136:3096–3105.
107. Roessler E, Muenke M. The molecular genetics of holoprosencephaly. *Am J Med Genet C Semin Med Genet*. 2010;154C(1):52–61.
108. Roessler E, Belloni E, Gaudenz K, et al. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet*. 1996;14(3):357–360.
109. Brown SA, Warburton D, Brown LY, et al. Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. *Nat Genet*. 1998;20(2):180–183.
110. Cohen Jr MM. Holoprosencephaly: clinical, anatomic, and molecular dimensions. *Birth Defects Res A Clin Mol Teratol*. 2006;76:658–673.
111. Demurger F, Ichkou A, Mougou-Zerelli S, et al. New insights into genotype-phenotype correlation for GLI3 mutations. *Eur J Hum Genet*. 2015;23(1):92–102.
112. Mercier S, Dubourg C, Garcelon N, et al. New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *J Med Genet*. 2011;48:752–760.

113. Treier M, O'Connell S, Gleiberman A, et al. Hedgehog signaling is required for pituitary gland development. *Development*. 2001;128(3):377–386.
114. Roessler E, Du YZ, Mullor JL, et al. Loss of function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proc Natl Acad Sci U S A*. 2003;100(23):13424–13429.
115. Treier M, Gleiberman AS, O'Connell SM, et al. Multistep signaling requirements for pituitary organogenesis in vivo. *Development*. 2001;128:377–386.
116. Cryns K, Sivakumaran TA, Van den Ouweland JM, et al. Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease. *Hum Mutat*. 2003;22(4):275–287.
117. Sheng HZ, Zhadanov AB, Mosinger Jr B, et al. Specification of pituitary cell lineages by the LIM homeobox gene Lhx3. *Science*. 1996;272:1004–1007.
118. Netchine I, Sobrier ML, Krude H, et al. Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat Genet*. 2000;25(2):182–186.
119. Sobrier ML, Brachet C, Vié-Luton MP, et al. Symptomatic heterozygotes and prenatal diagnoses in a nonconsanguineous family with syndromic combined pituitary hormone deficiency resulting from two novel LHX3 mutations. *J Clin Endocrinol Metab*. 2012;97(3):E503–E509.
120. Pfaeffle RW, Savage JJ, Hunter CS, et al. Four novel mutations of the LHX3 gene cause combined pituitary hormone deficiencies with or without limited neck rotation. *J Clin Endocrinol Metab*. 2007;92:1909–1919.
121. Rajab A, Kelberman D, de Castro SC, et al. Novel mutations in LHX3 are associated with hypopituitarism and sensorineural hearing loss. *Hum Mol Genet*. 2008;17(14):2150–2159.
122. Kriström B, Zdunek AM, Rydh A, et al. A novel mutation in the LIM homeobox gene is responsible for combined pituitary hormone deficiency, hearing impairment and vertebral malformations. *J Clin Endocrinol Metab*. 2009;94:1154–1161.
123. Jullien N, Romanet P, Philippon M, et al. Heterozygous LHX3 mutations may lead to a mild phenotype of combined pituitary hormone deficiency. *Eur J Hum Genet*. 2018;27(2):216–225.
124. Mullen RD, Colvin SC, Hunter CS, et al. Roles of the LHX3 and LHX4 LIM-homeodomain factors in pituitary development. *Mol Cell Endocrinol*. 2007;265–266:190–195.
125. Tajima T, Hattori T, Nakajima T, et al. A novel missense mutation (P366T) of the LHX4 gene causes severe combined pituitary hormone deficiency with pituitary hypoplasia, ectopic posterior lobe, and a poorly developed sella turcica. *Endocr J*. 2007;54:637–641.
126. Pfaeffle RW, Hunter CS, Savage JJ, et al. Three novel missense mutations within the LHX4 gene are associated with variable pituitary hormone deficiencies. *J Clin Endocrinol Metab*. 2008;93:1062–1071.
127. Gregory LC, Humayun KN, Turton JP, et al. Lethal form of congenital hypopituitarism associated with the first recessive LHX4 mutation. *J Clin Endocrinol Metab*. 2015;100:2158–2164.
128. Filges I, Bischof-Renner A, Röthlisberger B, et al. Panhypopituitarism presenting as life-threatening heart failure caused by an inherited microdeletion of 1q25 including LHX4. *Pediatrics*. 2012;129:e529–e534.
129. Reynaud R, Jayakody SA, Monnier C, et al. PROKR2 variants in multiple hypopituitarism with pituitary stalk interruption. *J Clin Endocrinol Metab*. 2021;97:E1068–E1073.
130. McCabe MJ, Gaston-Massuet C, Gregory LC, et al. Variations in PROKR2, but not PROK2, are associated with hypopituitarism and septo-optic dysplasia. *J Clin Endocrinol Metab*. 2013;98(3):E547–E557.
131. Dateki S, Fukami M, Sato N, et al. OTX2 mutation in a patient with anophthalmia, short stature, and partial growth hormone deficiency: functional studies using the IRBP, HESX1, and POU1F1 promoters. *J Clin Endocrinol Metab*. 2008;93(10):3697–3702.
132. Karaca E, Buyukkaya R, Pehlivan D, et al. Whole-exome sequencing identifies homozygous GPR161 mutation in a family with pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*. 2015;100:E140–E147.
133. Bashamboo A, Bignon-Topalovic J, Moussi N, McElreavey K, Brauner R. Mutations in the human ROBO1 gene in pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*. 2017;102(7):2401–2406.
134. Quentien MH, Delemer B, Papadimitriou DT, et al. Deficit in anterior pituitary function and variable immune deficiency (DAVID) in children presenting with adrenocorticotropin deficiency and severe infections. *J Clin Endocrinol Metab*. 2012;97:E121–E128.
135. Hassed SJ, Li S, Xu W, Taylor AC. A novel mutation in *PITX2* in a patient with Axenfeld-Rieger syndrome. *Mol Syndromol*. 2017;8:107–109.
136. Semina EV, Reiter R, Leysens NJ, et al. Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nat Genet*. 1996;14(4):392–399.
137. Johnston J, Olivos-Glander I, Killoran C, et al. Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of GLI3 mutations. *Am J Hum Genet*. 2005;76:609–622.
138. Démurger F, Ichkou A, Mougou-Zerelli S, et al. New insights into genotype–phenotype correlation for GLI3 mutations. *Eur J Hum Genet*. 2015;23(1):92–102.
139. Gregory LC, Gevers EF, Baker J, et al. Structural pituitary abnormalities associated with CHARGE syndrome. *J Clin Endocrinol Metab*. 2013;98(4):E737–E743.
140. Martín MG, Lindberg I, Solorzano-Vargas RS, et al. Congenital proprotein convertase 1/3 deficiency causes malabsorptive diarrhea and other endocrinopathies in a pediatric cohort. *Gastroenterology*. 2013;145(1):138–148.
141. Sun Y, Bak B, Schoenmakers N, et al. Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat Genet*. 2012;44:1375–1381.
142. Joustra SD, Schoenmakers N, Persani L, et al. The IGSF1 deficiency syndrome: characteristics of male and female patients. *J Clin Endocrinol Metab*. 2013;98:4942–4952.
143. Joustra SD, Roelfsema F, van Trotsenburg P, et al. IGSF1 deficiency results in human and murine somatotrope neurosecretory hyperfunction. *J Clin Endocrinol Metab*. 2020;105(3):e70–e84.
144. Gualtieri A, Kyprianou N, Gregory LC, et al. Activating mutations in BRAF disrupt the hypothalamo-pituitary axis leading to hypopituitarism in mice and humans. *Nat Commun*. 2021;12(1):2028.
145. Pingault V, Bodereau V, Baral V, et al. Loss-of-function mutations in SOX10 cause Kallmann syndrome with deafness. *Am J Hum Genet*. 2013;92:707–724.
146. Hutchins BI, Kotan LD, Taylor-Burds C, et al. CCDC141 mutation identified in anosmic hypogonadotropic hypogonadism (Kallmann syndrome) alters GnRH neuronal migration. *Endocrinology*. 2016;157(5):1956–1966.
147. Watterberg KL, Scott SM. Evidence of early adrenal insufficiency in babies who develop bronchopulmonary dysplasia. *Pediatrics*. 1995;95:120–125.
148. Balasubramanian R, Crowley WJ. Isolated gonadotropin-releasing hormone (GnRH) deficiency. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews® [Internet]*. Seattle; 2017.
149. Miraoui H, Dwyer AA, Sykiotis GP, et al. Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. *Am J Hum Genet*. 2013;92:725–743.
150. Dodé C, Rondard P. PROK2/PROKR2 signaling and Kallmann syndrome. *Front Endocrinol*. 2013;4:19.
151. Brioude F, Bouligand J, Francou B, et al. Two families with normosmic congenital hypogonadotropic hypogonadism and biallelic mutations in KISS1R (KISS1 receptor) : clinical evaluation and molecular characterization of a novel mutation. *PLoS One*. 2013;8(1):e53896.
152. de Roux N, Young J, Misrahi M, et al. A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *N Engl J Med*. 1997;337:1597–1602.
153. Topaloglu A, Tello J, Kotan L, et al. Inactivating KISS1 mutation and hypogonadotropic hypogonadism. *N Engl J Med*. 2012;366:629–635.
154. Topaloglu AK, Reimann F, Guclu M, et al. TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat Genet*. 2009;41:354–358.
155. Synofzik M, Gonzalez MA, Lourenco CM, et al. PNPLA6 mutations cause Boucher-Neuhauser and Gordon Holmes syndromes as part of a broad neurodegenerative spectrum. *Brain*. 2014;137:69–77.

156. Margolin DH, Kousi M, Chan YM, et al. Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. *N Engl J Med*. 2013;368:1992–2003.
157. Shi CH, Schisler JC, Rubel CE, et al. Ataxia and hypogonadism caused by the loss of ubiquitin ligase activity of the U box protein CHIP. *Hum Mol Genet*. 2014;23(4):1013–1024.
158. Vesper AH, Raetzman LT, Camper SA. Role of prophet of Pit1 (PROP1) in gonadotrope differentiation and puberty. *Endocrinology*. 2006;147:1654–1663.
159. De Rienzo F, Mellone S, Bellone S, et al. Frequency of genetic defects in combined pituitary hormone deficiency: a systematic review and analysis of a multicentre Italian cohort. *Clin Endocrinol*. 2015;83:849–860.
160. Nose O, Tatsumi K, Nakano Y, Amino N. Congenital combined pituitary hormone deficiency attributable to a novel PROP1 mutation (467insT). *J Pediatr Endocrinol Metab*. 2006;19:491–498.
161. Abrao MG, Leite MV, Carvalho LR, et al. Combined pituitary hormone deficiency (CPHD) due to a complete PROP1 deletion. *Clin Endocrinol*. 2006;65:294–300.
162. Lebl J, Vosáhlo J, Pfaeffle RW, et al. Auxological and endocrine phenotype in a population-based cohort of patients with PROP1 gene defects. *Eur J Endocrinol*. 2005;153:389–396.
163. Reynaud R, Chadli-Chaieb M, Vallette-Kasic S, et al. A familial form of congenital hypopituitarism due to a PROP1 mutation in a large kindred: phenotypic and in vitro functional studies. *J Clin Endocrinol Metab*. 2004;61:635–640.
164. Cogan JD, Wu W, Phillips JA, et al. The PROP1 2-base pair deletion is a common cause of combined pituitary hormone deficiency. *J Clin Endocrinol Metab*. 1998;83:3346–3349.
165. Arroyo A, Pernasetti F, Vasilyev VV, et al. A unique case of combined pituitary hormone deficiency caused by a PROP1 gene mutation (R120C) associated with normal height and absent puberty. *Clin Endocrinol*. 2002;57:283–291.
166. Voutetakis A, Maniati-Christidi M, Kanaka-Gantenbein C, et al. Prolonged jaundice and hypothyroidism as the presenting symptoms in a neonate with a novel PROP1 gene mutation (Q83X). *Eur J Endocrinol*. 2004;150:257–264.
167. Asteria C, Oliveira JH, Abucham J, Beck-Peccoz P. Central hypocortisolism as part of combined pituitary hormone deficiency due to mutations of PROP1 gene. *Eur J Endocrinol*. 2000;143:347–352.
168. Pernasetti F, Toledo SP, Vasilyev VV, et al. Impaired adrenocorticotropin-adrenal axis in combined pituitary hormone deficiency caused by a two-base pair deletion (301-302delAG) in the prophet of Pit-1 gene. *J Clin Endocrinol Metab*. 2000;85:390–397.
169. Voutetakis A, Sertedaki A, Livadas S, et al. Pituitary size fluctuation in long-term MR studies of PROP1 deficient patients: a persistent pathophysiological mechanism? *J Endocrinol Invest*. 2006;29:462–466.
170. Vallette-Kasic S, Barlier A, Teinturier C, et al. PROP1 gene screening in patients with multiple pituitary hormone deficiency reveals two sites of hypermutability and a high incidence of corticotroph deficiency. *J Clin Endocrinol Metab*. 2001;86:4529–4535.
171. Andersen B, Rosenfeld MG. POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease. *Endocr Rev*. 2001;22:2–35.
172. Turton JP, Reynaud R, Mehta A, et al. Novel mutations within the POU1F1 gene associated with variable combined pituitary hormone deficiency. *J Clin Endocrinol Metab*. 2005;90:4762–4770.
173. Cohen LE, Wondisford FE, Salvatoni A, et al. A “hotspot” in the Pit-1 gene responsible for combined pituitary hormone deficiency: clinical and molecular correlates. *J Clin Endocrinol Metab*. 1995;80:679–684.
174. Pfäffle RW, DiMattia GE, Parks JS, et al. Mutation of the POU-specific domain of Pit-1 and hypopituiarism without pituitary hypoplasia. *Science*. 1992;257:1118–1121.
175. Cherella CE, Cohen LE. Congenital hypopituitarism in neonates. *NeoReviews*. 2018;19(12):e742–e752.
176. Collu R, Tang J, Castagn J, et al. A novel mechanism for isolated central hypothyroidism: inactivating mutations in the thyrotropin-releasing hormone receptor gene. *J Clin Endocrinol Metab*. 1997;82:1561–1565.
177. Alatzoglou KS, Turton JP, Kelberman D, et al. Expanding the spectrum of mutations in GH1 and GHRHR: genetic screening in a large cohort of patients with congenital isolated growth hormone deficiency. *J Clin Endocrinol Metab*. 2009;94(9):3191–3199.
178. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson IC, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev*. 2009;30(7):790–829.
179. Couture C, Saveanu A, Barlier A, et al. Phenotypic homogeneity and genotypic variability in a large series of congenital isolated ACTH-deficiency patients with TPIT gene mutations. *J Clin Endocrinol Metab*. 2012;97(3):E486–E495.
180. Vallette-Kasic S, Brue T, Pulichino AM, et al. Congenital isolated adrenocorticotropin deficiency: an underestimated cause of neonatal death explained by TPIT gene mutations. *J Clin Endocrinol Metab*. 2005;90:1323–1331.
181. Bonomi M, Busnelli M, Beck-Peccoz P, et al. A family with complete resistance to thyrotropin-releasing hormone. *N Engl J Med*. 2009;360(7):731–734.
182. García M, González de Buitrago J, Jiménez-Rosés M, et al. Central hypothyroidism due to a TRHR mutation causing impaired ligand affinity and transactivation of Gq. *J Clin Endocrinol Metab*. 2017;10217:2433–2442.
183. Bassi MT, Ramesar RS, Caciotti B, et al. X-linked late-onset sensorineural deafness caused by a deletion involving OA1 and a novel gene containing WD-40 repeats. *Am J Hum Genet*. 1999;64(6):1604–1616.
184. Somsen D, Davis-Keppen L, Crotwell P, et al. Congenital nasal pyriform aperture stenosis and ocular albinism co-occurring in a sibship with a maternally-inherited 97 kb Xp22.2 microdeletion. *Am J Med Genet*. 2014;164A(5):1268–1271.
185. Malendowicz LK. 100th anniversary of the discovery of the human adrenal fetal zone by Stella Starkel and Lesław Węgrzynowski: how far have we come? *Folia Histochem Cytobiol*. 2010;48(4):491–506.
186. Xing Y, Lerario AM, Rainey W, Hammer GD. Development of adrenal cortex zonation. *Endocrinol Metab Clin North Am*. 2015;44(2):243–274.
187. Ishimoto H, Jaffe RB. Development and function of the human fetal adrenal cortex: a key component in the feto-placental unit. *Endocr Rev*. 2011;32(3):317–355.
188. Keene M. Observations on the development of the human suprarenal gland. *J Anat*. 1927;61(3):302–324.
189. Spencer SJ, Mesiano S, Lee JY, Jaffe RB. Proliferation and apoptosis in the human adrenal cortex during the fetal and perinatal periods: implications for growth and remodeling. *J Clin Endocrinol Metab*. 1999;84(3):1110–1115.
190. Winter J. Fetal and neonatal adrenocortical physiology. In: Polin R, Fox W, Abman S, eds. *Fetal and Neonatal Physiology*. Philadelphia, PA: WB Saunders; 2004:1915–1925.
191. Mesiano S, Jaffe RB. Developmental and functional biology of the primate fetal adrenal cortex. *Endocr Rev*. 1997;18(3):378–403.
192. Jirasek J. *Human Fetal Endocrines*. Dordrecht: Springer Netherlands; 1980.
193. Midgley PC, Russell K, Oates N, Shaw JC, Honour JW. Activity of the adrenal fetal zone in preterm infants continues to term. *Endocr Res*. 1996;22:729–733.
194. Hui XG, Akahira J, Suzuki T, et al. Development of the human adrenal zona reticularis: morphometric and immunohistochemical studies from birth to adolescence. *J Endocrinol*. 2009;203(2):241–252. doi:10.1677/JOE-09-0127.
195. Hammer GD, Parker KL, Schimmer BP. Minireview: transcriptional regulation of adrenocortical development. *Endocrinology*. 2005;146(3):1018–1024.
196. Fujieda K, Tajima T. Molecular basis of adrenal insufficiency. *Pediatr Res*. 2005;57:62R–69R.
197. Ikeda Y, Swain A, Weber TJ, et al. Steroidogenic factor 1 and Dax-1 colocalize in multiple cell lineages: potential links in endocrine development. *Mol Endocrinol*. 1996;10(10):1261–1272. doi:10.1210/mend.10.10.9121493.
198. Yates R, Katugampola H, Cavlan D, et al. Adrenocortical development, maintenance, and disease. *Curr Top Dev Biol*. 2013;106:239–312.

199. Guasti L, Paul A, Laufer E, King P. Localization of Sonic hedgehog secreting and receiving cells in the developing and adult rat adrenal cortex. *Mol Cell Endocrinol*. 2011;336(1–2):117–122. doi:10.1016/j.mce.2010.11.010.
200. King P, Paul A, Laufer E. Shh signaling regulates adrenocortical development and identifies progenitors of steroidogenic lineages. *Proc Natl Acad Sci U S A*. 2009;106(50):21185–21190.
201. Huang CC, Miyagawa S, Matsumaru D, Parker KL, Yao HH. Progenitor cell expansion and organ size of mouse adrenal is regulated by sonic hedgehog. *Endocrinology*. 2010;151(3):1119–1128. doi:10.1210/en.2009-0814.
202. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer*. 2010;10(2):116–129.
203. Guasti L, Candy Sze WC, McKay T, Grose R, King PG. FGF signalling through Fgfr2 isoform IIIb regulates adrenal cortex development. *Mol Cell Endocrinol*. 2013;37(1–2):182–188.
204. Mesiano S, Jaffe R. Neuroendocrine-metabolic regulation of pregnancy. In: Strauss J, Barbieri R, eds. Reproductive Endocrinology. Philadelphia, PA: WB Saunders;327–366.
205. Arboleda VA, Lee H, Parnaik R, et al. Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. *Nat Genet*. 2012;44(7):788–792.
206. Miller WL. Steroid hormone biosynthesis and actions in the materno-feto-placental unit. *Clin Perinatol*. 1988;25:799–817.
207. Baker BY, Lin L, Kim CJ, et al. Non-classic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*. 2006;91(2):4781–4785.
208. Prasad R, Hadjidemetriou I, Maharaj A, et al. Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest*. 2017;127(3):942–953.
209. Lovric S, Goncalves S, Gee HY, et al. Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest*. 2017;127(3):912–928.
210. Geller DH, Miller WL. Molecular development of the adrenal gland. In: Pescovitz OH, Eugster EA, eds. *Pediatric Endocrinology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004:548–567.
211. Albrecht ED, Aberdeen GW, Pepe GJ. Estrogen elicits cortical zone-specific effects on development of the primate fetal adrenal gland. *Endocrinology*. 2005;146(4):1737–1744.
212. Goto M, Piper Hanley K, Marcos J, et al. In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development. *J Clin Invest*. 2006;116(4):953–960. doi:10.1172/JCI25091.
213. Liggins GC. The role of cortisol in preparing the fetus for birth. *Reprod Fertil Dev*. 1994;6(2):141–150.
214. Cole TJ, Blendy JA, Monaghan AP, et al. Targeted disruption of the glucocorticoid receptor gene blocks adrenergic chromaffin cell development and severely retards lung maturation. *Genes Dev*. 1995;9:1608–1621.
215. Rose JC, Turner CS, Ray D, Rawashdeh N. Evidence that cortisol inhibits basal adrenocorticotropin secretion in the sheep fetus by 0.70 gestation. *Endocrinology*. 1988;123(3):1307–1313.
216. Robillard JE, Page WV, Mathews MS, et al. Differential gene expression and regulation of renal angiotensin II receptor subtypes (AT1 and AT2) during fetal life in sheep. *Pediatr Res*. 1995;38(6):896–904.
217. Berger S, Bleich M, Schmid W, et al. Mineralocorticoid receptor knockout mice: pathophysiology of Na+ metabolism. *Proc Natl Acad Sci U S A*. 1998;95(16):9424–9429.
218. Hirasawa G, Sasono H, Suzuki T, et al. 11Beta-hydroxysteroid dehydrogenase type 2 and mineralocorticoid receptor in human fetal development. *J Clin Endocrinol Metab*. 1999;84(4):1453–1458.
219. Lumbers E. Functions of the renin-angiotensin system during development. *Clin Exp Pharmacol Physiol*. 1995;22(8):499–505.
220. Dotzler SA, Digeronimo RJ, Yoder BA, Siler-Khodr TM. Distribution of corticotrophin releasing hormone in the fetus, newborn, juvenile and adult baboon. *Pediatr Res*. 2004;55:120–125.
221. Thomson M, Smith R. The action of hypothalamic and placental corticotroin releasing factor on the corticotrope. *Mol Cell Endocrinol*. 1989;62:1–12.
222. Goland RS, Wardlow SL, Blum M, Tropper PJ, Stark RI. Biologically active corticotrophin-releasing hormone in maternal and fetal plasma during pregnancy. *Am J Obs Gynecol*. 1988;159:884–890.
223. Smith R, Nicholson RC. Corticotrophin releasing hormone and the timing of birth. *Front Biosci*. 2007;12:912–918.
224. Rivier C, Vale W. Neuroendocrine interactions between corticotrophin releasing factor and vasopressin on adrenocorticotropic hormone secretion in the rat. In: Schrier W, ed. *Vasopressin*. ; 1985:181–188. New York.
225. Yang K, Jones SA, Challis JR. Changes in glucocorticoid receptor number in the hypothalamus of the sheep fetus with gestational age and after adrenocorticotrophin treatment. *Endocrinology*. 1990;126:11–17.
226. Pulichino AM, Vallette-Kasic S, Couture C, et al. Human and mouse TPIT gene mutations cause early onset pituitary ACTH deficiency. *Genes Dev*. 2003;17(6):711–716.
227. Samuels ME, Gallo-Payet N, Pinard S, et al. Bioinactive ACTH causing glucocorticoid deficiency. *J Clin Endocrinol Metab*. 2013;98(2):736–742.
228. Clark AJ, Chan LF, Chung TT, Metherell L. The genetics of familial glucocorticoid deficiency. *Best Pr Res Clin Endocrinol Metab*. 2009;23(2):159–165.
229. Cooray SN, Chan L, Metherell L, Storr H, Clark AJL. Adrenocorticotropin resistance syndromes. *Endocr Dev*. 2008;13:99–116.
230. Meimaridou E, Hughes CR, Kowalczyk J, et al. Familial glucocorticoid deficiency: new genes and mechanisms. *Mol Cell Endocrinol*. 2013;371(1–2):195–200.
231. Gineau L, Cognet C, Kara N, et al. Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. *J Clin Invest*. 2012;122(3):821–832.
232. Hughes CR, Guasti L, Meimaridou E, et al. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. *J Clin Invest*. 2012;122(3):814–820.
233. Meimaridou E, Kowalczyk J, Guasti L, et al. Mutations in NNT encoding nicotinamide nucleotide transhydrogenase cause familial glucocorticoid deficiency. *Nat Genet*. 2012;44(7):740–742.
234. Kowalczyk JC, Meimaridou E, Novoselova T, et al. Digenic inheritance of mutations in antioxidant pathway genes causing familial glucocorticoid deficiency?. *Endocr Rev*. 2013;34. SAT-49.
235. Ferraz-de-Souza B, Achermann JC. Disorders of adrenal development. *Endocr Dev*. 2008;13:19–32.
236. Muscatelli F, Strom T, Walker A, et al. Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature*. 1994;372(6507):672–676.
237. Achermann JC, Ito M, Ito M, Hindmarsh PC, Jameson JL. A mutation in the gene encoding steroidogenic factor 1 causes XY reversal and adrenal failure in humans. *Nat Genet*. 1999;22(2):125–126.
238. Köhler B, Lin L, Ferraz-de-Souza B, et al. Five novel mutations in steroidogenic factor 1 (SF1/Ad4BP, NR5A1) are associated with 46XY disorders of sex development with normal adrenal function. *Hum Mutat*. 2008;29:59–64.
239. Bashamboo A, Ferraz-de-Souza B, Lourenço D, et al. Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *Am J Hum Genet*. 2010;87(4):505–512.
240. Lourenço D, Brauner R, Lin L, et al. Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med*. 2009;360:1200–1210.
241. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet*. 2016;48(7):792–797.
242. Buonocore F, Kühnen P, Suntharalingham JP, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. *J Clin Invest*. 2017;127(5):1700–1713.
243. Pascoe L, Curnow KM, Slutsker L, Rösler A, White PC. Mutations in the human CYP11B2 (aldosterone synthase) gene causes corticosterone methyloxidase II deficiency. *Proc Natl Acad Sci U S A*. 1992;89(11):4996–5000.
244. Furgeson SB, Linas S. Mechanisms of type I and type II pseudohypoaldosteronism. *J Am Soc Nephrol*. 2010;21(11):1842–1845.
245. Erickson R, Blecher S. Genetics of sex determination and differentiation. In: Polin R, Fox W, Abman S, eds. *Fetal and Neonatal Physiology*. 3rd ed. 2004:1935–1941. Philadelphia, PA.

246. Lee M. Molecular genetic control of sex differentiation. In: Pescovitz O, Eugster E, eds. *Pediatric Endocrinology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2004:231–242.
247. Harley VR, Clarkson MJ, Argentaro A. The molecular action and regulation of the testis-determining factors, SRY (sex-determining region on the Y chromosome) and SOX9 [SRY-related high-mobility group (HMG) box 9]. *Endocr Rev*. 2003;24(4):466–487.
248. Park SY, Jameson J. Minireview: transcriptional regulation of gonadal development and differentiation. *Endocrinology*. 2005;146(3):1035–1042.
249. Chau YY, Hastie ND. The role of Wt1 in regulating mesenchyme in cancer, development and tissue homeostasis. *Trends Genet*. 2012;28(10):515–524.
250. Larson A, Nokoff NJ, Travers S. Disorders of sex development: clinically relevant genes involved in gonadal differentiation. *Discov Med*. 2012;14(78):301–309.
251. Sekido R, Lovell-Badge R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*. 2008;453:930–934.
252. Ono M, Harley V. Disorders of sex development: new genes, new concepts. *Nat Rev Endocrinol*. 2013;9:79–91.
253. Sutton E, Hughes J, White S, et al. Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J Clin Invest*. 2011;121(1):328–341.
254. Bay K, Main KM, Toppari J, Skakkebæk NE. Testicular descent: INSL3, testosterone, genes and the intrauterine milieu. *Nat Rev Urol*. 2011;8:187–196.
255. Ivell R, Anand-Ivell R. Biological role and clinical significance of insulin-like peptide 3. *Curr Opin Endocrinol Diabetes Obes*. 2011;18:210–216.
256. Zhang FP, Poutanen M, Wilbertz J. Normal prenatal but arrested postnatal sexual development of luteinizing hormone receptor knockout (LURKO) mice. *Mol Endocrinol*. 2001;15:172–183.
257. Philibert P, Zenaty D, Lin L, et al. Mutational analysis of steroidogenic factor 1 (NR5a1) in 24 boys with bilateral anorchia: a French collaborative study. *Hum Reprod*. 2007;22:3255–3261.
258. Biason-Lauber A, Konrad D, Meyer M, DeBeaufort C, Schoenle E. Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the CBX2 gene. *Am J Hum Genet*. 2009;84:658–663.
259. Aslan A, Kogan B, Gondos B. Testicular development. In: Polin R, Fox W, Abmans S, eds. *Fetal and Neonatal Physiology*. Philadelphia, PA: WB Saunders; 2004:1950–1955.
260. Josso N, Belville C, Dicard J. Mutations in AMH and its receptors. *Endocrinology*. 2003;13:247–251.
261. Hagen CP, Aksglaede L, Sørensen K, et al. Serum levels of anti-Mullerian hormone as a marker of ovarian function in 926 healthy females from birth to adulthood and in 172 Turner syndrome patients. *J Clin Endocrinol Metab*. 2010;95(11):5003–5010.
262. Aksglaede L, Sørensen K, Boas M, et al. Changes in anti-Mullerian hormone (AMH) throughout the life span: a population based study of 1027 healthy makes from birth (cord blood) to the age of 69 years. *J Clin Endocrinol Metab*. 2010;95(12):5357–5364.
263. Rey R, Picard JY. Embryology and endocrinology of genital development. *Baillieres Clin Endocrinol Metab*. 1998;12:17–33.
264. Pelliniemi LJ, Dym M. The fetal gonad and sexual differentiation. In: Tulchinsky D, Little AB, eds. *Maternal Fetal Endocrinology*. 2nd ed. Philadelphia, PA: WB Saunders; 1994:298–320.
265. Byskov A, Westergaard L. Differentiation of the ovary. In: Polin R, Fox W, Abman S, eds. *Fetal and Neonatal Physiology*. Philadelphia, PA: WB Saunders; 2004:1941–1949.
266. Fulton N, da Silva S, Bayne R. Germ cell proliferation and apoptosis in the developing human ovary. *J Clin Endocrinol Metab*. 2005;90:4664–4670.
267. Kyriakou A, Lucas-Herald A, McGowan R, Tobias E, Ahmed F. Disorders of sex development: advances in genetic diagnosis and challenges in management. *Adv Genomics Genet*. 2015;5:165–177.
268. Richards JS, Pangas SA. The ovary: basic biology and clinical implications. *J Clin Invest*. 2010;120(4):963–972.
269. Uhlenhaut N, Jakob S, Anlag K, et al. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell*. 2009;139:1130–1142.
270. Matson CK, Zarkower D. Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat Rev Genet*. 2012;13:163–174.
271. Biason-Lauber A. WNT4, RSPO1, and FOXL2 in sex development. *Semin Reprod Med*. 2012;30(5):387–395.
272. Tomaselli S, Megiorni F, De Bernardo C, et al. Syndromic true hermaphroditism due to an R-spondin1 (RSPO1) homozygous mutation. *Hum Mutat*. 2008;29(2):220–226.
273. Parma P, Radi O, Vidal V, et al. R-spondin 1 is essential in sex determination, skin differentiation and malignancy. *Nat Genet*. 2006;38(11):1304–1309.
274. Biason-Lauber A, Konrad D, Navratil F, Schoenle EJ. A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman. *N Engl J Med*. 2004;351:792–798.
275. Mandel H, Shermer R, Borochowitz ZU, et al. SERKAL syndrome: an autosomal-recessive disorder caused by a loss of function mutation in WNT4. *Am J Hum Genet*. 2008;82:39–47.
276. Sajjad Y, Quenby S, Nickson P, Lewis-Jones DI, Vince G. Immunohistochemical localization of androgen receptors in the urogenital tracts of human embryos. *Reproduction*. 2004;128:331–339.
277. Garcia-Falgueras A, Pinos H, Collado P, et al. The role of the androgen receptor in CNS masculinization. *Brain Res*. 2005;1035:13–23.
278. Kudwa AE, Bodo C, Gustafsson J, Rissman EF. A previously uncharacterized role for estrogen receptor β: defeminization of male brain and behavior. *Proc Natl Acad Sci U S A*. 2005;102:4608–4612.
279. Naftolin F, Brawer JR. The effect of estrogens on hypothalamic structures and function. *Am J Obstet Gynecol*. 1978;132:758–765.
280. Sholl SA, Goy RW, Kim KL. 5 alpha-reductase, aromatase, and androgen receptor levels in the monkey brain during fetal development. *Endocrinology*. 1989;124:627–634.
281. Brandenberger AW, Tee MK, Lee JY, Chao V, Jaffe RB. Tissue distribution of estrogen receptors alpha (ERα) and beta (ERβ) mRNA in the midgestation human fetus. *J Clin Endocrinol Metab*. 1997;82:3509–3512.
282. Couse JF, Korach KS. Estrogen receptor null mice: what have we learned and where will they lead us? *Endocr Rev*. 1999;20(3):358–417.
283. Smith EP, Boyd J, Frank GR, et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med*. 1994;331:1056–1061.
284. Brain CE, Creighton SM, Mushtaq I, et al. Holistic management of DSD. *Best Prac Res Clin Endocrinol Metab*. 2010;24(2):335–354.
285. Lee PA, Nordenström A, Houk CP, et al. Global disorders of sex development update since 2006: perceptions, approach and care. *Horm Res Paediatr*. 2016;85:158–180.
286. Johansson E, Andersson L, Örnros J, et al. Revising the embryonic origin of thyroid C cells in mice and humans. *Development*. 2015;142(20):3519–3528. doi:10.1242/dev.126581.
287. Nilsson M, Williams D. On the origin of cells and derivation of thyroid cancer: C cell story revisited. *Eur Thyroid J*. 2016;5(2):79–93. doi:10.1159/000447333.
288. Sprinzl GM, Koebke J, Wimmers-Klick J, Eckel HE, Thumfart WF. Morphology of the human thyroglossal tract: a histologic and macroscopic study in infants and children. *Ann Otol Rhinol Laryngol*. 2000;109(12 Pt 1):1135–1139. doi:10.1177/000348940010901210.
289. Szinnai G. Genetics of normal and abnormal thyroid development in humans. *Best Pract Res Clin Endocrinol Metab*. 2014;28(2):133–150.
290. De Felice M, Di Lauro R. Thyroid development and its disorders: genetics and molecular mechanisms. *Endocr Rev*. 2004;25(5):722–746.
291. Fagman H, Nilsson M. Morphogenesis of the thyroid gland. *Mol Cell Endocrinol*. 2010;323(1):35–54.
292. Fernández LP, López-Márquez A, Santisteban P. Thyroid transcription factors in development, differentiation and disease. *Nat Rev Endocrinol*. 2015;11(1):29–42.
293. Nilsson M, Fagman H. Development of the thyroid gland. *Development*. 2017;144(12):2123–2140. doi:10.1242/dev.145615.
294. Alt B, Elsalini OA, Schrumpf P, et al. Arteries define the position of the thyroid gland during its developmental relocalisation. *Development*. 2006;133(19):3797–3804. doi:10.1242/dev.02550.
295. Hermanns P, Grasberger H, Refetoff S, Pohlenz J. Mutations in the NKX2.5 gene and the PAX8 promoter in a girl with thyroid dysgenesis. *J Clin Endocrinol Metab*. 2011;96(6):E977–E981.

296. Opitz R, Hitz MP, Vandernoot I, et al. Functional zebrafish studies based on human genotyping point to netrin-1 as a link between aberrant cardiovascular development and thyroid dysgenesis. *Endocrinology*. 2015;156(1):377–388. doi:10.1210/en.2014-1628.
297. de Filippis T, Marelli F, Nebbia G, et al. JAG1 loss-of-function variations as a novel predisposing event in the pathogenesis of congenital thyroid defects. *J Clin Endocrinol Metab*. 2016;101(3):861–870. doi:10.1210/jc.2015-3403.
298. Porazzi P, Marelli F, Benato F, et al. Disruptions of global and JAGGED1-mediated notch signaling affect thyroid morphogenesis in the zebrafish. *Endocrinology*. 2012;153(11):5645–5658. doi:10.1210/en.2011-1888.
299. Nilsson M, Fagman H. Development of the thyroid gland. *Development*. 2017;144:2123–2140.
300. Lania G, Zhang Z, Huynh T, et al. Early thyroid development requires a Tbx1-Fgf8 pathway. *Dev Biol*. 2009;328(1):109–117. doi:10.1016/j.ydbio.2009.01.014.
301. Fagman H, Liao J, Westerlund J, et al. The 22q11 deletion syndrome candidate gene Tbx1 determines thyroid size and positioning. *Hum Mol Genet*. 2007;16(3):276–285. doi:10.1093/hmg/ddl455.
302. Alt B, Elsalini O, Schrumpf P, et al. Arteries define the position of the thyroid gland during its developmental relocalization. *Development*. 2006;133:3797–3804.
303. Chisaka O, Capecchi MR. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5. *Nature*. 1991;350:473–479.
304. Kimura S, Hara Y, Pineau T, et al. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev*. 1996;10(1):60–69.
305. Mansouri A, Chowdhury K, Gruss P. Follicular cells of the thyroid gland require Pax8 gene function. *Nat Genet*. 1998;19(1):87–90. doi:10.1038/ng0598-87.
306. Martinez Barbera JP, Clements M, Thomas P, et al. The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development*. 2000;127(11):2433–2445.
307. Szinnai G, Lacroix L, Carré A, et al. Sodium/iodide symporter (NIS) gene expression is the limiting step for the onset of thyroid function in the human fetus. *J Clin Endocrinol Metab*. 2007;92(1):70–76. doi:10.1210/jc.2006-1450.
308. Milanesi A, Brent GA. Management of hypothyroidism in pregnancy. *Curr Opin Endocrinol Diabetes Obes*. 2011;18(5):304–309. doi:10.1097/MED.0b013e32834a91d1.
309. Tonyushkina KN, Shen MC, Ortiz-Toro T, Karlstrom RO. Embryonic exposure to excess thyroid hormone causes thyrotrope cell death. *J Clin Invest*. 2014;124(1):321–327.
310. Eng L, Lam L. Thyroid function during the fetal and neonatal periods. *NeoReviews*. 2020;21(1):e30–e36.
311. Burrow G, Fisher D, Larsen P. Maternal and fetal thyroid function. *N Engl J Med*. 1994;331:1072–1078.
312. Postiglione MP, Parlato R, Rodriguez-Mallon A, et al. Role of the thyroid-stimulating hormone receptor signaling in development and differentiation of the thyroid gland. *Proc Natl Acad Sci U S A*. 2002;99(24):15462–15467. doi:10.1073/pnas.242328999.
313. Davies TF, Ando T, Lin RY, Tomer Y, Latif R. Thyrotropin receptor-associated diseases: from adenomata to Graves disease. *J Clin Invest*. 2005;115(8):1972–1983. doi:10.1172/JCI26031.
314. Roti E. Regulation of thyroid stimulating hormone (TSH) secretion in the fetus and neonate. *J Endocrinol Invest*. 1988;11:145–158.
315. Mercado M, Yu V, Francis I, Szymonowicz W, Gold H. Thyroid function in very preterm infants. *Early Hum Dev*. 1988;16(2):131–141.
316. Williams F, Simpson J, Delahunty C, et al. Developmental trends in cord and postpartum serum thyroid hormones in preterm infants. *J Clin Endocrinol Metab*. 2004;89(11):5314–5320.
317. Fisher DA, Nelson JC, Carlton EI, Wilcox RB. Maturation of human hypothalamic-pituitary-thyroid function and control. *Thyroid*. 2000;10(3):229–234.
318. Murphy N, Hume R, van Toor H, et al. The hypothalamic-pituitary-thyroid axis in preterm infants; changes in the first 24 hours of postnatal life. *J Clin Endocrinol Metab*. 2004;89(6):2824–2831.
319. Lee DK, Nguyen T, O'Neill GP, et al. Discovery of a receptor related to the galanin receptors. *FEBS Lett*. 1999;446(1):103–107. Accessed March 23, 2014. http://www.ncbi.nlm.nih.gov/pubmed/10100623.
320. Eng PH, Cardona GR, Fang SL, et al. Escape from the acute Wolff-Chaikoff effect is associated with a decrease in thyroid sodium/iodide symporter messenger ribonucleic acid and protein. *Endocrinology*. 1999;140:3404–3410.
321. Sherwin JR. Development of regulatory mechanisms in the thyroid: failure of iodide to suppress iodide transport activity. *Proc Soc Exp Biol Med*. 1982;169:458–462.
322. Brown R, Huang S, Fisher D. The maturation of thyroid function in the perinatal period and during childhood. In: Braverman L, Utiger R, eds. *The Thyroid*. 9th ed. Philadelphia, PA: JB Lippincott; 2005:1013–1028.
323. Wassner AJ, Brown RS. Hypothyroidism in the newborn period. *Curr Opin Endocrinol Diabetes Obes*. 2013;20(5):449–454.
324. Schoenmakers E, Agostini M, Mitchell C, et al. Mutations in the selenocysteine insertion sequence-binding protein 2 gene lead to a multisystem selenoprotein deficiency disorder in humans. *J Clin Invest*. 2010;120(12):4220–4235.
325. Dumitrescu AM, Refetoff S. Inherited defects of thyroid hormone metabolism. *Ann Endocrinol*. 2011;72:95–98.
326. Dumitrescu AM, Di Cosmo C, Liao XH, Weiss RE, Refetoff S. The syndrome of inherited partial SBP2 deficiency in humans. *Antioxid Redox Signal*. 2010;12(7):905–920.
327. St Germain DL, Hernandez A, Schneider MJ, Galton VA. Insights into the role of deiodinases from studies of genetically modified animals. *Thyroid*. 2005;15(8):905–916.
328. Hume R, Simpson J, Delahunty C, et al. Human fetal and cord serum thyroid hormones: developmental trends and interrelationships. *J Clin Endocrinol Metab*. 2004;89(98):4097–4103.
329. Polk DH, Reviczky A, Wu SY, Huang WS, Fisher DA. Metabolism of sulfoconjugated thyroid hormone derivatives in developing sheep. *Am J Physiol*. 1994;266:E892–E896.
330. Richard K, Hume R, Kaptein E, et al. Ontogeny of iodothyronine deiodinases in human liver. *J Clin Endocrinol Metab*. 1998;83:2868–2874.
331. Polk DH. Thyroid hormone effects on neonatal thermogenesis. *Semin Perinatol*. 1988;12:151–156.
332. Polk DH, Wu SY, Wright C, Reviczky AL, Fisher D. Ontogeny of thyroid hormone effect on tissue 5'-monodeiodinase activity in fetal sheep. *Am J Physiol*. 1988;254(3):E337–E341.
333. Jansen J, Friesema EC, Milici C, Visser TJ. Thyroid hormone transporters in health and disease. *Thyroid*. 2005;15(8):757–768.
334. Heuer H, Maier MK, Iden S, et al. The monocarboxylate transporter 8 linked to human psychomotor retardation is highly expressed in thyroid hormone-sensitive neuron populations. *Endocrinology*. 2005;146(4):1701–1706.
335. Dumitrescu AM, Liao XH, Best TB, Brockmann K, Refetoff S. A novel syndrome combining thyroid and neurological abnormalities is associated with mutations in a monocarboxylate transporter gene. *Am J Hum Genet*. 2004;74(1):168–175.
336. Groeneweg S, van Geest FS, Abacı A, et al. Disease characteristics of MCT8 deficiency: an international, retrospective, multicentre cohort study. *Lancet Diabetes Endocrinol*. 2020;8:594–605.
337. Groeneweg S, Peeters RP, Moran C, et al. Effectiveness and safety of tri-iodothyronine analogue Triac in children and adults with MCT8 deficiency: an international, single-arm, open label, phase 2 trial. *Lancet Diabetes Endocrinol*. 2019;7:695–706.
338. Visser WE. Therapeutic applications of thyroid hormone analogues. *Ann Endocrinol*. 2021;82(3–4):170–172. https://doi.org/10.1016/j.ando.2020.03.004.
339. López-Espíndola D, Morales-Bastos C, Grijota-Martínez C, et al. Mutations of the thyroid hormone transporter MCT8 cause prenatal brain damage and persistent hypomyelination. *J Clin Endocrinol Metab*. 2014;99(12):E2799–E2804.
340. Bergh JJ, Lin HY, Lansing L, et al. Integrin alphaVbeta3 contains a cell surface receptor site for thyroid hormone that is linked to activation of mitogen-activated protein kinase and induction of angiogenesis. *Endocrinology*. 2005;146(7):2864–2871.

341. Di Cosmo C, Liao XH, Dumitrescu AM, et al. Mice deficient in MCT8 reveal a mechanism regulating thyroid hormone secretion. *J Clin Invest*. 2010;120(9):3377–3388.
342. Ferrara AM, Liao XH, Gil-Ibáñez P, et al. Placenta passage of the thyroid hormone analog DITPA to male wild-type and Mct8-deficient mice. *Endocrinology*. 2014;155(10):4088–4093.
343. Kersseboom S, Horn S, Visser WE, et al. In vitro and mouse studies supporting therapeutic utility of triiodothyroacetic acid in MCT8 deficiency. *Mol Endocrinol*. 2014;28(12):1961–1970.
344. Morreale de Escobar G, Obregon MJ, Escobar del Rey F. Role of thyroid hormone during early brain development. *Eur J Endocrinol*. 2004;151:U25–U37.
345. Contempré B, Jauniaux E, Calvo R, et al. Detection of thyroid hormones in human embryonic cavities during the first trimester of pregnancy. *J Clin Endocrinol Metab*. 1993;77(6):1719–1722.
346. Vulsma T, Gons MH, de Vijlder JJ. Maternal-fetal transfer of thyroxine in congenital hypothyroidism due to a total organification defect or thyroid agenesis. *N Engl J Med*. 1989;321:13–16.
347. Morreale de Escobar G, Calvo R, Obregon MJ, Escobar Del Rey F. Contribution of maternal thyroxine to fetal thyroxine pools in normal rats near term. *Endocrinology*. 1990;126(5):2765–2767.
348. Yen P. Genomic and nongenomic actions of thyroid hormones. In: Braverman LE, Utiger R, eds. *The Thyroid*. Philadelphia, PA: Lippincott Williams & Wilkins; 2005:135–150.
349. Flamant F, S.J.T. Thyroid hormone receptors: lessons from knockout and knockin mutant mice. *Trends Endocrinol Metab*. 2005;14:85–90.
350. Morreale de Escobar GOM, Escobar del Rey F. Role of thyroid hormone during early brain development. *Eur J Endocrinol*. 2004;151:U25–U37.
351. Kilby MD, Gittoes N, McCabe C, Verhaeg J, Franklyn JA. Expression of thyroid receptor isoforms in the human fetal central nervous system and the effects of intrauterine growth restriction. *Clin Endocrinol*. 2000;53(4):469–477.
352. Iskaros J, Pickard M, Evans I, et al. Thyroid hormone receptor gene expression in first trimester human fetal brain. *J Clin Endocrinol Metab*. 2000;85(7):2620–2623.
353. Rajatapiti P, Kester MH, de Krijger RR, et al. Expression of glucocorticoid, retinoid, and thyroid hormone receptors during human lung development. *J Clin Endocrinol Metab*. 2005;90(7):4309–4314.
354. Bochukova E, Schoenmakers N, Agostini M, et al. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*. 2012;366:243–249.
355. van Mullem A, van Heerebeek R, Chrysis D, et al. Clinical phenotype and mutant TRα1. *N Engl J Med*. 2012;366(15):1451–1453.
356. Sadow PM, Chassande O, Koo EK, et al. Regulation of expression of thyroid hormone receptor isoforms and coactivators in liver and heart by thyroid hormone. *Mol Cell Endocrinol*. 2003;203:65–75.
357. Angelin-Duclos C, Domenget C, Kolbus A, et al. Thyroid hormone T3 acting through the thyroid hormone α receptor is necessary for implementation of erythropoiesis in the neonatal spleen environment in the mouse. *Development*. 2005;132:925–934.
358. Quignodon L, Legrand C, Allioli N, et al. Thyroid hormone signaling is highly heterogeneous during pre- and postnatal brain development. *J Mol Endocrinol*. 2004;33:467–476.
359. Plateroti M, Gauthier K, Domon-Dell C, et al. Functional interference between thyroid hormone receptor α (TRα) and natural truncated TRΔα isoforms in the control of intestinal development. *Mol Cell Biol*. 2001;21:4761–4772.
360. Mai W, Janier M, Allioli N, et al. Thyroid hormone receptor α is a molecular switch of cardiac function between fetal and postnatal life. *Proc Natl Acad Sci U S A*. 2004;101:10332–10337.
361. Harvey CB, O'Shea PJ, Scott AJ, et al. Molecular mechanisms of thyroid hormone effects on bone growth and function. *Mol Genet Metab*. 2002;75:17–30.
362. Marrif H, Schifman A, Stepanyan Z, et al. Temperature homeostasis in transgenic mice lacking thyroid receptor α gene products. *Endocrinology*. 2005;146:2872–2884.
363. St Germain DL, Hernandez A, Schneider MJ, Galton VA. Insights into the role of deiodinases from studies of genetically modified animals. *Thyroid*. 2005;15:905–916.
364. Williams GR. Thyroid hormone actions in cartilage and bone. *Eur Thyroid J*. 2013;2(1):3–13.
365. Persani L, Rurale G, de Filippis T, et al. Genetics and management of congenital hypothyroidism. *Best Pract Res Clin Endocrinol Metab*. 2018;32(4):387–396. doi:10.1016/j.beem.2018.05.002.
366. Carré A, Stoupa A, Kariyawasam D, et al. Mutations in BOREALIN cause thyroid dysgenesis. *Hum Mol Genet*. 2017;26(3):599–610.
367. Grasberger H, Refetoff S. Genetic causes of congenital hypothyroidism due to dyshormonogenesis. *Curr Opin Pediatr*. 2011;23:421–428.
368. de Filippis T, Gelmini G, Paraboschi E, et al. A frequent oligogenic involvement in congenital hypothyroidism. *Hum Mol Genet*. 2017;26(13):2507–2514.
369. Burrow GN, Fisher DA, Larsen PR. Maternal and fetal thyroid function. *N Engl J Med*. 1994;331(16):1072–1078. doi:10.1056/nejm199410203311608.
370. Fukushi M, Honma K, Fujita K. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*. 1999;341(26):2016–2017. doi:10.1056/nejm199908193410801.
371. Mitchell ML, Klein RZ. The sequelae of untreated maternal hypothyroidism. *Eur J Endocrinol*. 2004;151(suppl 3):U45–U48.
372. Léger J, Olivieri A, Donaldson M, et al. European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. *Horm Res Paediatr*. 2014;81(2):80–103.
373. Spencer L, Bubner T, Bain E, Middleton P. Screening and subsequent management for thyroid dysfunction pre-pregnancy and during pregnancy for improving maternal and infant health. *Cochrane Database Syst Rev*. 2015;9:CD011263. doi:10.1002/14651858.CD011263.pub2.
374. Morreale de Escobar G, Obregon MJ. Escobar del Rey F. Is neuropsychological development related to maternal hypothyroidism or to maternal hypothyroxinemia? *J Clin Endocrinol Metab*. 2000;85(3975–3987).
375. Pop VJ, Brouwers EP, Vader HL, et al. Maternal hypothyroxinaemia during early pregnancy and subsequent child development: a 3-year follow-up study. *Clin Endocrinol*. 2003;59(3):282–288.
376. Huber K, Karch N, Ernsberger U, Goridis C, Unsicker K. The role of PHOX2B in chromaffin cell development. *Dev Biol*. 2005;279:501–508.
377. Padbury JF. Functional maturation of the adrenal medulla and peripheral sympathetic nervous system. *Baillieres Clin Endocrinol Metab*. 1989;33:689–705.
378. Anderson DJ, Carnahan JF, Michelsohn A, Patterson PH. Antibody markers identify a common progenitor to sympathetic neurons and chromaffin cells in vivo and reveal the timing of commitment to neuronal differentiation in the sympathoadrenal lineage. *J Neurosci*. 1991;11(11):3507–3519.
379. Reissmann E, Ernsberger U, Francis-West P, et al. Involvement of bone morphogenetic protein-4 and bone morphogenetic protein-7 in the differentiation of the adrenergic phenotype in developing sympathetic neurons. *Development*. 1996;122(7):2079–2088.
380. Ehrhart-Bornstein M, Breidert M, Guadanucci P, et al. 17 alpha-Hydroxylase and chromogranin A in 6th week human fetal adrenals. *Horm Metab Res*. 1997;29(1):30–32. doi:10.1055/s-2007-978976.
381. Yamamoto M, Yanai R, Arishima K. Study of migration of neural crest cells to adrenal medulla by three-dimensional reconstruction. *J Vet Med Sci*. 2004;66(6):635–641.
382. Ernsberger U, Esposito L, Partimo S, et al. Expression of neuronal markers suggests heterogeneity of chick sympathoadrenal cells prior to invasion of the adrenal anlagen. *Cell Tissue Res*. 2005;319(1):1–13.
383. Lumb R, Tata M, Xu X, et al. Neuropilins guide preganglionic sympathetic axons and chromaffin cell precursors to establish the adrenal medulla. *Development*. 2018;145(21):dev162552. doi:10.1242/dev.162552.
384. Gaultier C, Trang H, Dauger S, Gallego J. Pediatric disorders with autoimmune dysfunction: what role for PHOX2B? *Pediatr Res*. 2005;58:1–6.
385. Aloe L, Levi-Montalcini R. Nerve growth factor-induced transformation of immature chromaffin cells in vivo into sympathetic neurons: effect of antiserum to nerve growth factor. *Proc Natl Acad Sci U S A*. 1979;76:1246–1250.

386. Jozan S, Aziza J, Chatelin S, et al. Human fetal chromaffin cells: a potential tool for cell pain therapy. *Exp Neurol*. 2007;205(2):525–535.
387. Slotkin TA, Seidler FJ. Adrenomedullary catecholamine release in the fetus and newborn: secretory mechanisms and their role in stress and survival. *J Dev Physiol*. 1988;10:1–16.
388. Palmer SM, Oakes GK, Lam RW, Hobel CJ, Fisher DA. Catecholamine physiology in the ovine fetus. II. Metabolic clearance rate of epinephrine. *Am J Physiol*. 1984;246:E350–E355.
389. Palmer SM, Oakes GK, Champion JA, Fisher DA, Hobel CK. Catecholamine physiology in the ovine fetus. III. Maternal and fetal response to acute maternal exercise. *Am J Obs Gynecol*. 1984;149:426–434.
390. Pryds O, Christensen NJ, Friis-Hansen B. Increased cerebral blood flow and plasma epinephrine in hypoglycemic, preterm neonates. *Pediatrics*. 1990;85:172–176.
391. Jones CT, Roebuck MM, Walker DW, Johnston BM. The role of the adrenal medulla and peripheral sympathetic nerves in the physiological responses of the fetal sheep to hypoxia. *J Dev Physiol*. 1988;10(1):17–36.
392. Jones CT, Roebuck MM, Walker DW, Lagercrantz H, Johnston BM. Cardiovascular, metabolic and endocrine effects of chemical sympathectomy and of adrenal demedullation in fetal sheep. *J Dev Physiol*. 1987;9(4):347–367.
393. Zhou QY, Quaife CJ, Palmiter RD. Targeted disruption of the tyrosine hydroxylase gene reveals that catecholamines are required for mouse fetal development. *Nature*. 1995;374:640–643.
394. Thomas SA, Matsumoto AM, Palmiter RD. Noradrenaline is essential for mouse fetal development. *Nature*. 1995;374:643–646.
395. Cohen WR, Piasecki GJ, Cohn HE, Jackson BT. Plasma catecholamines in the hypoxaemic fetal rhesus monkey. *J Dev Physiol*. 9:507–515.
396. Stonestreet BS, Piasecki GJ, Susa JB, Jackson BT. Effects of insulin infusion on catecholamine concentration in fetal sheep. *Am J Obs Gynecol*. 1989;160:740–745.
397. Padbury JF, Agata Y, Polk DH, Wang DL, Callegari CC. Neonatal adaptation: naloxone increases the catecholamine surge at birth. *Pediatr Res*. 1987;21(6):590–593. doi:10.1203/00006450-198706000-00017.
398. Cabrera O, Berman DM, Kenyon NS, et al. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A*. 2006;103(7):2334–2339. doi:10.1073/pnas.0510790103.
399. Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature*. 1994;371(6498):606–609. doi:10.1038/371606a0.
400. Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J*. 1993;12(11):4251–4259. doi:10.1002/j.1460-2075.1993.tb06109.x.
401. Hebrok M, Kim SK, Melton DA. Notochord repression of endodermal sonic hedgehog permits pancreas development. *Genes Dev*. 1998;12(11):1705–1713. doi:10.1101/gad.12.11.1705.
402. Kim SK, Hebrok M, Melton DA. Notochord to endoderm signaling is required for pancreas development. *Development*. 1997;124(21):4243–4252.
403. Grapin-Botton A. Antero-posterior patterning of the vertebrate digestive tract: 40 Years after Nicole Le Douarin's PhD thesis. *Int J Dev Biol*. 2005;49(2–3):335–347. doi:10.1387/ijdb.041946ag.
404. Sherwood RI, Chen TY, Melton DA. Transcriptional dynamics of endodermal organ formation. *Dev Dyn*. 2009;238(1):29–42. doi:10.1002/dvdy.21810.
405. Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development*. 2002;129(10):2447–2457.
406. Kawaguchi Y, Cooper B, Gannon M, et al. The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat Genet*. 2002;32(1):128–134. doi:10.1038/ng959.
407. Zhou Q, Law AC, Rajagopal J, et al. A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell*. 2007;13(1):103–114. doi:10.1016/j.devcel.2007.06.001.
408. Bhushan A, Itoh N, Kato S, et al. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development*. 2001;128(24):5109–5117.
409. Fukuda A, Kawaguchi Y, Furuyama K, et al. Reduction of Ptf1a gene dosage causes pancreatic hypoplasia and diabetes in mice. *Diabetes*. 2008;57(9):2421–2431. doi:10.2337/db07-1558.
410. Murtaugh LC, Law AC, Dor Y, Melton DA. Beta-catenin is essential for pancreatic acinar but not islet development. *Development*. 2005;132(21):4663–4674. doi:10.1242/dev.02063.
411. Wells JM, Esni F, Boivin GP, et al. Wnt/β-catenin signaling is required for development of the exocrine pancreas. *BMC Dev Biol*. 2007;7:4. doi:10.1186/1471-213X-7-4.
412. Schaffer AE, Freude KK, Nelson SB, Sander M. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Dev Cell*. 2010;18(6):1022–1029. doi:10.1016/j.devcel.2010.05.015.
413. Larsen HL, Grapin-Botton A. The molecular and morphogenetic basis of pancreas organogenesis. *Semin Cell Dev Biol*. 2017;66:51–68. doi:10.1016/j.semcdb.2017.01.005.
414. Apelqvist A, Li H, Sommer L, et al. Notch signalling controls pancreatic cell differentiation. *Nature*. 1999;400(6747):877–881. doi:10.1038/23716.
415. Fujikura J, Hosoda K, Iwakura H, et al. Notch/Rbp-j signaling prevents premature endocrine and ductal cell differentiation in the pancreas. *Cell Metab*. 2006;3(1):59–65. doi:10.1016/j.cmet.2005.12.005.
416. Jensen J, Heller RS, Funder-Nielsen T, et al. Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. *Diabetes*. 2000;49(2):163–176. doi:10.2337/diabetes.49.2.163.
417. Ahlgren U, Pfaff SL, Jessell TM, Edlund T, Edlund H. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature*. 1997;385(6613):257–260. doi:10.1038/385257a0.
418. Gierl MS, Karoulias N, Wende H, Strehle M, Birchmeier C. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic β cells and intestinal endocrine cells. *Genes Dev*. 2006;20(17):2465–2478. doi:10.1101/gad.381806.
419. Naya FJ, Huang HP, Qiu Y, et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev*. 1997;11(18):2323–2334. doi:10.1101/gad.11.18.2323.
420. Conrad E, Stein R, Hunter CS. Revealing transcription factors during human pancreatic β cell development. *Trends Endocrinol Metab*. 2014;25(8):407–414. doi:10.1016/j.tem.2014.03.013.
421. Oliver-Krasinski JM, Stoffers DA. On the origin of the beta cell. *Genes Dev*. 2008;22(15):1998–2021.
422. Pan FC, Brissova M. Pancreas development in humans. *Curr Opin Endocrinol Diabetes Obes*. 2014;21(2):77–82. doi:10.1097/MED.0000000000000047.
423. Piper K, Brickwood S, Turnpenny LW, et al. Beta cell differentiation during early human pancreas development. *J Endocrinol*. 2004;181(1):11–23. doi:10.1677/joe.0.1810011.
424. Jennings RE, Berry AA, Kirkwood-Wilson R, et al. Development of the human pancreas from foregut to endocrine commitment. *Diabetes*. 2013;62(10):3514–3522. doi:10.2337/db12-1479.
425. Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA. Human pancreas development. *Development*. 2015;142(18):3126–3137. doi:10.1242/dev.120063.
426. Piper K, Ball SG, Keeling JW, et al. Novel SOX9 expression during human pancreas development correlates to abnormalities in Campomelic dysplasia. *Mech Dev*. 2002;116(1–2):223–226. doi:10.1016/S0925-4773(02)00145-4.
427. Shaw-Smith C, De Franco E, Lango Allen H, et al. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes*. 2014;63(8):2888–2894. doi:10.2337/db14-0061.
428. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet*. 1997;15(1):106–110. doi:10.1038/ng0197-106.
429. Pan FC, Wright C. Pancreas organogenesis: from bud to plexus to gland. *Dev Dyn*. 2011;240(3):530–565. doi:10.1002/dvdy.22584.

430. Edlund H. Pancreatic organogenesis - developmental mechanisms and implications for therapy. *Nat Rev Genet*. 2002;3(7):524–532. doi:10.1038/nrg841.
431. Steiner DJ, Kim A, Miller K, Hara M. Pancreatic islet plasticity: Interspecies comparison of islet architecture and composition. *Islets*. 2010;2(3):135–145. doi:10.4161/isl.2.3.11815.
432. Riedel MJ, Asadi A, Wang R, et al. Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. *Diabetologia*. 2012;55(2):372–381. doi:10.1007/s00125-011-2344-9.
433. van der Meulen T, Huising MO. Role of transcription factors in the transdifferentiation of pancreatic islet cells. *J Mol Endocrinol*. 2015;54(2):R103–R117. doi:10.1530/JME-14-0290.
434. Meier JJ, Butler AE, Saisho Y, et al. β-cell replication is the primary mechanism subserving the postnatal expansion of β-cell mass in humans. *Diabetes*. 2008;57(6):1584–1594. doi:10.2337/db07-1369.
435. Girard J. Control of fetal and neonatal glucose metabolism by pancreatic hormones. *Baillieres Clin Endocrinol Metab*. 1989;3(3):817–836. doi:10.1016/S0950-351X(89)80055-2.
436. Rao PN, Shashidhar A, Ashok C. In utero fuel homeostasis: lessons for a clinician. *Indian J Endocrinol Metab*. 2013;17:60–68.
437. Aldoretta PW, Hay Jr WW. Metabolic substrates for fetal energy metabolism and growth. *Clin Perinatol*. 1995;22:15–36.
438. Ammon HP, Glocker C, Waldner RG, Wahl MA. Insulin release from pancreatic islets of fetal rats mediated by leucine b-BCH, tolbutamide, glibenclamide, arginine, potassium chloride, and theophylline does not require stimulation of Ca2+ net uptake. *Cell Calcium*. 1989;10(6):441–450.
439. Sperling M. Carbohydrate metabolism: insulin and glucagons. In: Tulchinsky D, Little A, eds. *Maternal-Fetal Endocrinology*. Philadelphia, PA: WB Saunders; 1994:380–400.
440. Formby B, Ullrich A, Coussens L, Walker L, Peterson CM. Growth hormone stimulates insulin gene expression in cultured human fetal pancreatic islets. *J Clin Endocrinol Metab*. 1988;66(5):1075–1079. doi:10.1210/jcem-66-5-1075.
441. Vajravelu ME, De León DD. Genetic characteristics of patients with congenital hyperinsulinism. *Curr Opin Pediatr*. 2018;30(4):568–575. doi:10.1097/MOP.0000000000000645.
442. Banerjee I, Estebanez MS, Cosgrove KE, Dunne MJ. Congenital hyperinsulinism. In: Huhtaniemi I, Martini L, Eds., *Encyclopedia of Endocrine Diseases*. Philadelphia: Elsevier; 2018:602–617.
443. Galcheva S, Demirbilek H, Al-Khawaga S, Hussain K. The genetic and molecular mechanisms of congenital hyperinsulinism. *Front Endocrinol*. 2019;10:111. doi:10.3389/fendo.2019.00111.
444. Pinney SE, MacMullen C, Becker S, et al. Clinical characteristics and biochemical mechanisms of congenital hyperinsulinism associated with dominant KATP channel mutations. *J Clin Invest*. 2008;118(8):2877–2886. doi:10.1172/JCI35414.
445. Palladino AA, Stanley CA. The hyperinsulinism/hyperammonemia syndrome. *Rev Endocr Metab Disord*. 2010;11(3):171–178. doi:10.1007/s11154-010-9146-0.
446. Kapoor RR, James C, Hussain K. Hyperinsulinism in developmental syndromes. *Endocrine Development*. 2009;14:95–113.
447. Senniappan S, Shanti B, James C, Hussain K. Hyperinsulinaemic hypoglycaemia: genetic mechanisms, diagnosis and management. *J Inherit Metab Dis*. 2012;35(4):589–601. doi:10.1007/s10545-011-9441-2.
448. Adzick NS, De Leon DD, States LJ, et al. Surgical treatment of congenital hyperinsulinism: results from 500 pancreatectomies in neonates and children. *J Pediatr Surg*. 2019;54(1):27–32. doi:10.1016/j.jpedsurg.2018.10.030.
449. Rubio-Cabezas O, Hattersley AT, Njølstad PR, et al. ISPAD Clinical Practice Consensus Guidelines 2014. The diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes*. 2014;15(suppl 20):47–64. doi:10.1111/pedi.12192.
450. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic β-cell diabetes. *Nat Clin Pract Endocrinol Metab*. 2008;4(4):200–213. doi:10.1038/ncpendmet0778.
451. De Franco E, Flanagan SE, Houghton JA, et al. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet*. 2015;386(9997):957–963. doi:10.1016/S0140-6736(15)60098-8.
452. Hattersley AT, Patel KA. Precision diabetes: learning from monogenic diabetes. *Diabetologia*. 2017;60(5):769–777. doi:10.1007/s00125-017-4226-2.
453. Kanakatti Shankar R, Pihoker C, Dolan LM, et al. Permanent neonatal diabetes mellitus: prevalence and genetic diagnosis in the SEARCH for Diabetes in Youth Study. *Pediatr Diabetes*. 2013;14(3):174–180. doi:10.1111/pedi.12003.
454. Flanagan SE, Patch AM, Mackay DJ, et al. Mutations in ATP-sensitive K+ channel genes cause transient neonatal diabetes and permanent diabetes in childhood or adulthood. *Diabetes*. 2007;56(7):1930–1937. doi:10.2337/db07-0043.
455. Flanagan SE, De Franco E, Lango Allen H, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell Metab*. 2014;19(1):146–154. doi:10.1016/j.cmet.2013.11.021.
456. De Franco E, Watson RA, Weninger WJ, et al. A specific CNOT1 mutation results in a novel syndrome of pancreatic agenesis and holoprosencephaly through impaired pancreatic and neurological development. *Am J Hum Genet*. 2019;104(5):985–989. doi:10.1016/j.ajhg.2019.03.018.
457. De Franco E, Caswell R, Johnson MB, et al. De novo mutations in eIF2B1 affecting eIF2 signaling cause neonatal/early-onset diabetes and transient hepatic dysfunction. *Diabetes*. 2020;69(3):477–483. doi:10.2337/db19-1029.
458. Johnson MB, Hattersley AT, Flanagan SE. Monogenic autoimmune diseases of the endocrine system. *Lancet Diabetes Endocrinol*. 2016;4(10):862–872. doi:10.1016/S2213-8587(16)30095-X.
459. Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic syndromes of severe insulin resistance. *Endocr Rev*. 2011;32(4):498–514. doi:10.1210/er.2010-0020.
460. Ellard S, Lango Allen H, De Franco E, et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia*. 2013;56(9):1958–1963. doi:10.1007/s00125-013-2962-5.
461. Karges B, Meissner T, Icks A, Kapellen T, Holl RW. Management of diabetes mellitus in infants. *Nat Rev Endocrinol*. 2011;8(4):201–211.
462. Vaxillaire M, Bonnefond A, Froguel P. The lessons of early-onset monogenic diabetes for the understanding of diabetes pathogenesis. *Best Pract Res Clin Endocrinol Metab*. 2012;26:171–187.
463. Frisdal A, Trainor PA. Development and evolution of the pharyngeal apparatus. *Wiley Interdiscip Rev Dev Biol*. 2014;3:403–418.
464. Santisteban P. Development and anatomy of the hypothalamic-pituitary axis. In: Braverman L, Utiger R, eds. *The Thyroid*. 9th ed. Philadelphia, PA: JB Lippincott; 2005:8–25.
465. Wolfe HJ, DeLellis RA, Voelkel EF, Tashjian Jr AH. Distribution of calcitonin-containing cells in the normal neonatal human thyroid gland: a correlation of morphology with peptide content. *J Clin Endocrinol Metab*. 1975;41:1076–1081.
466. Peissig K, Condie BG, Manley NR. Embryology of the parathyroid glands. *Endocrinol Metab Clin North Am*. 2018;47(4):733–742. doi:10.1016/j.ecl.2018.07.002.
467. Grigorieva IV, Thakker RV. Transcription factors in parathyroid development: lessons from hypoparathyroid disorders. *Ann N Y Acad Sci*. 2011;1237:24–38.
468. Gordon RJ, Levine MA. Genetic disorders of parathyroid development and function. *Endocrinol Metab Clin North Am*. 2018;47(4):809–823. doi:10.1016/j.ecl.2018.07.007.
469. Ding C, Buckingham B, Levine MA. Familial isolated hypoparathyroidism caused by a mutation in the gene for the transcription factor GCMB. *J Clin Invest*. 2001;108(8):1215–1220. doi:10.1172/JCI200113180.
470. Mitsui T, Narumi S, Inokuchi M, et al. Comprehensive next-generation sequencing analyses of hypoparathyroidism: identification of novel GCM2 mutations. *J Clin Endocrinol Metab*. 2014;99:E2421–E2428.
471. Maret A, Bourdeau I, Ding C, et al. Expression of GCMB by intrathymic parathyroid hormone-secreting adenomas indicates their parathyroid cell origin. *J Clin Endocrinol Metab*. 2004;89(1):8–12.

472. Guan B, Welch JM, Sapp JC, et al. GCM2-activating mutations in familial isolated hyperparathyroidism. *Am J Hum Genet*. 2016;99(5):1034–1044.
473. Bowl MR, Nesbit MA, Harding B, et al. An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. *J Clin Invest*. 2005;115(10):2822–2831.
474. Papangeli I, Scambler P. The 22q11 deletion: DiGeorge and velocardiofacial syndromes and the role of TBX1. *Wiley Interdiscip Rev Dev Biol*. 2013;2:393–403.
475. Guris DL, Fantes J, Tara D, Druker BJ, Imamoto A. Mice lacking the homologue of the human 22q11.2 gene CRLK phenocopy neurocristopathies of DiGeorge syndrome. *Nat Genet*. 2001;27(3):293–298.
476. Kovacs CS. Bone development and mineral homeostasis in the fetus and neonate: roles of the calciotropic and phosphotropic hormones. *Physiol Rev*. 2014;94(4):1143–1218.
477. Simmonds C, Karsenty G, Karaplis A, Kovacs C. Parathyroid hormone regulates fetal-placental mineral homeostasis. *J Bone Miner Res*. 2010;25(3):594–605.
478. Pitkin R, Cruikshank D, Schauberger C, et al. Fetal calcium, calcitropic hormones and neonatal homeostasis. *Pediatrics*. 1980;66:77–82.
479. Prada J. Calcium-regulating hormones. In: Polin R, Fox W, eds. *Fetal and Neonatal Physiology*. 2nd ed. Philadelphia, PA: WB Saunders; 1998:2287–2296.
480. Kovaks C, Kronenberg H. Maternal-fetal calcium and bone metabolism during pregnancy, puerperium and lactation. *Endocr Rev*. 1997;18:832–872.
481. Sethi A, Priyadarshi M, Agarwal R. Mineral and bone physiology in the foetus, preterm and full-term neonates. *Semin Fetal Neonatal Med*. 2020;25(1):101076.
482. Care AD. Development of endocrine pathways in the regulation of calcium homeostasis. *Baillieres Clin Endocrinol Metab*. 1989;3(3):671–688.
483. Ryan BA, Kovacs CS. Calciotropic and phosphotropic hormones in fetal and neonatal bone development. *Semin Fetal Neonatal Med*. 2020;25(1):101062.
484. Vilardaga J, Romero G, Friedman P, Gardella TJ. Molecular basis of parathyroid hormone receptor signaling and trafficking: a family B GPCR paradigm. *Cell Mol Life Sci*. 2011;68:1–13.
485. Mitchell D, Juppner H. Regulation of calcium homeostasis and bone metabolism in the fetus and neonate. *Curr Opin Endocrinol Diabetes Obes*. 2010;17:25–30.
486. Kovacs C, Lanske B, Hunzelman J. Parathyroid hormone-related peptide (PTHrP) regulates fetal-placental calcium transport through a receptor distinct from the PTH/PTHrP receptor. *Proc Natl Acad Sci U S A*. 1996;93:15233–15238.
487. Allgrove J. Cytochemical bioassay of parathyroid hormone in maternal and cord blood. *Arch Dis Child*. 1985;60(2):110–115.
488. Karaplis A, Luz A, Glowacki J. Lethal skeletal dysplasia from targeted disruption of the parathyroid hormone-related protein gene. *Genes Dev*. 1994;8:277–289.
489. Kovacs C. Bone metabolism in the fetus and neonate. *Pediatr Nephrol*. 2014;29:793–803.
490. Kovacs C, Chafe L, Fudge N, Friel J, Manley N. PTH regulates fetal blood calcium and skeletal mineralization independently of PTHrP. *Endocrinology*. 2001;142:4983–4993.
491. Kovacs CS, Manley NR, Moseley JM, John Martin T, Kronenberg HM. Fetal parathyroids are not required to maintain placental calcium transport. *J Clin Invest*. 2001;107(8):1007–1015.
492. Care A, Abbas S, Pickard D. Stimulation of ovine placental transport of calcium and magnesium by mid-molecule fragments of human parathyroid hormone-related protein. *J Exp Physiol*. 1990;75:605–608.
493. Miao D, He B, Karaplis C. Parathyroid hormone is essential for normal fetal bone formation. *J Clin Invest*. 2002;109:1173–1182.
494. Lanske B, Amling M, Neff L, Guiducci J, Baron R, Kronenberg H. Ablation of the PTHrP gene or the PTH/PTHrP receptor gene leads to distinct abnormalities in bone development. *J Clin Invest*. 1999;104:399–407.
495. Kovacs CS, Chafe LL, Woodland ML, McDonald KR, Fudge NJ, Wookey PJ. Calcitropic gene expression suggests a role for the intraplacental yolk sac in maternal-fetal calcium exchange. *Am J Physiol Endocrinol Metab*. 2002;282(3):E721–E732.
496. Ross R, Halbert K, Tsang R. Determination of the production and metabolic clearance rates of 1, 25-dihydroxyvitamin D3 in the pregnant sheep and its chronically catheterized fetus by primed infusion technique. *Pediatr Res*. 1989;26:633–638.
497. Hillman L, Haddad J. Human perinatal vitamin D metabolism. I. 25-Hydroxyvitamin D in maternal and cord blood. *J Pediatr*. 1974;84:742–749.
498. Steichen J, Tsang R, Gratton T, Hamstra A, DeLuca H. Vitamin D homeostasis in the perinatal period: 1,25-dihydroxyvitamin D in maternal, cord, and neonatal blood. *N Engl J Med*. 1980;302:315–319.
499. Turner M, Barré P, Benjamin A, Goltzman D, Gascon-Barré M. Does the maternal kidney contribute to the increased circulating 1,25-dihydroxyvitamin D concentrations during pregnancy? *Min Electrolyte Metab*. 1988;14:246–252.
500. Chan A, Conen P. Ultrastructural observations on cytodifferentiation of parafollicular cells in the human fetal thyroid. *Lab Invest*. 1971;25:249–259.
501. Wieland P, Fischer J, Trechsel U, et al. Perinatal parathyroid hormone, vitamin D metabolites, and calcitonin in man. *Am J Physiol*. 1980;239:E385–E390.
502. Woloszczuk W, Kovarik J, Pavelka P. Calcitonin in pregnant women and in cord blood. *Gynecol Obs Invest*. 1981;12:272–276.
503. Balabanova S, Kruse B, Wolf A. Calcitonin secretion by human placental tissue. *Acta Obs Gynecol Scand*. 1987;66:323–326.
504. McDonald K, Fudge N, Woodrow J, et al. Ablation of calcitonin/calcitonin gene-related peptide-alpha impairs fetal magnesium but not calcium homeostasis. *Am J Physiol Endocrinol Metab*. 2004;287:E218–E226.
505. Hoff A, Catala-Lehnen P, Thomas P, et al. Increased bone mass is an unexpected phenotype associated with deletion of the calcitonin gene. *J Clin Invest*. 2002;110:1849–1857.
506. Riccardi D, Brennan S, Chang W. The extracellular calcium-sensing receptor, CaSR, in fetal development. *Best Pr Res Clin Endocrinol Metab*. 2013;27:443–453.
507. Brown E. Role of the calcium-sensing receptor in extracellular calcium homeostasis. *Best Pr Res Clin Endocrinol Metab*. 2013;27:333–343.
508. Hannan F, Thakker R. Calcium-sensing receptor (CaSR) mutations and disorders of calcium, electrolyte and water metabolism. *Best Pr Res Clin Endocrinol Metab*. 2013;27:359–371.
509. Hu M, Shiizaki K, Kuro-o M, Moe OW. Fibroblast growth factor 23 and Klotho: physiology and pathophysiology of an endocrine network of mineral metabolism. *Annu Rev Physiol*. 2013;75:503–533.
510. Ma Y, Kirby B, Fairbridge N, et al. FGF23 is not required to regulate fetal phosphorus metabolism but exerts effects within 12 hours after birth. *Endocrinology*. 2017;158(2):252–263.
511. Maiya S, Sullivan I, Allgrove J, et al. Hypocalcaemia and vitamin D deficiency: an important, but preventable, cause of life-threatening infant heart failure. *Heart*. 2008;94:581–584.
512. Bansal B, Bansal M, Bajpai P, Garewal H. Hypocalcemic cardiomyopathy-different mechanisms in adult and pediatric cases. *J Clin Endocrinol Metab*. 2014;99:2627.
513. Kitanaka S, Takeyama K, Murayama A, et al. Inactivating mutations in the 25-hydroxyvitamin D 3 1α-hydroxylase gene in patients with pseudovitamin D–deficiency rickets. *N Engl J Med*. 1998;338:653–662.
514. Hannan F, Nesbit M, Zhang C, et al. Identification of 70 calcium-sensing receptor mutations in hyper- and hypo-calcaemic patients: evidence for clustering of extracellular domain mutations at calcium-binding sites. *Hum Mol Genet*. 2012;21:2768–2778.
515. Schlingmann KP, Kaufmann M, Weber S, et al. Mutations in CYP24A1 and idiopathic infantile hypercalcemia. *N Engl J Med*. 2011;365:410–421.
516. Schlingmann KP, Ruminska J, Kaufmann M, et al. Autosomal-recessive mutations in SLC34A1 encoding sodium-phosphate cotransporter 2a cause idiopathic infantile hypercalcemia. *J Am Soc Nephrol*. 2016;27(2):604–614.
517. Suzuki Y, Chitayat D, Sawada H, et al. TRPV6 variants interfere with maternal-fetal calcium transport through the placenta and cause transient neonatal hyperparathyroidism. *Am J Hum Genet*. 2018;102(6):1104–1114.
518. Farooque A, Moss C, Zehnder D, Hewison M, Shaw NJ. Expression of 25-hydroxyvitamin D3-1α-hydroxylase in subcutaneous fat necrosis. *Br J Dermatol*. 2009;160(2):423–425.

519. Christov M, Juppner H. Insights from genetic disorders of phosphate homeostasis. *Semin Nephrol*. 2013;33:143–157.
520. Russo V, Gluckman P, Feldman E, Werther G. The insulin-like growth factor system and its pleiotropic functions in brain. *Endocr Rev*. 2005;26:916–943.
521. DeLeon D, Cohen P, Katz L. Growth factor regulation of fetal growth. In: Polin RA, Fox WW, Abman S, eds. *Fetal and Neonatal Physiology*. 3rd ed. Philadelphia, PA: WB Saunders; 2004:1880–1890.
522. Murphy VE, Smith R, Giles WB, Clifton V. Endocrine regulation of human fetal growth: the role of the mother, placenta, and fetus. *Endocr Rev*. 2006;27(2):141–169.
523. Ong K, Kratzsch J, Kiess W, et al. Size at birth and cord blood levels of insulin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-1 (IGFBP-1), IGFBP-3, and the soluble IGF-II/mannose-6-phosphate receptor in term human infants. The ALSPAC Study Team. Avon Longitudinal Stu. *J Clin Endocrinol Metab*. 2000;85:4266–4269.
524. Simmen F, Simmon R, Letcher L. IGFs in pregnancy: developmental expression in uterus and mammary gland and paracrine actions during embryonic and neonatal growth. In: LeRoith D, Raizada M, eds. *Molecular and Cellular Biology of Insulin-like Growth Factors and Their Receptors*. New York: Plenum; 1989.
525. Spaventi R, Antica M, Pavelić K. Insulin and insulin-like growth factor I (IGF I) in early mouse embryogenesis. *Development*. 1990;108(3):491–495.
526. Sibley C, Coan P, Ferguson-Smith A. Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional exchange characteristics of the mouse placenta. *Proc Natl Acad Sci U S A*. 2004;101:8204–8208.
527. Constância M, Angiolini E, Sandovici I, et al. Adaptation of nutrient supply to fetal demand in the mouse involves interaction between the Igf2 gene and placental transporter systems. *Proc Natl Acad Sci U S A*. 2005;102(52):19219–19224.
528. Yakar S, Adamo ML. Insulin-like growth factor-1 physiology: lessons from mouse models. *Endocrinol Me*. 2012;41(2):231–247.
529. Woods KA, Camacho-Hübner C, Savage MO, Clark A. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med*. 1996;335(18):1363–1367.
530. Abuzzahab MJ, Schneider A, Goddard A, et al. Intrauterine Growth Retardation (IUGR) Study Group. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med*. 2003;349(23):2211–2222.
531. Geary MP, Pringle PJ, Rodeck CH, Kingdom JC, Hindmarsh P. Sexual dimorphism in the growth hormone and insulin-like growth factor axis at birth. *J Clin Endocrinol Metab*. 2003;888:3708–3714.
532. Pringle P, Geary M, Rodeck C, et al. The influence of cigarette smoking on antenatal growth, birth size, and the insulin-like growth factor axis. *J Clin Endocrinol Metab*. 2005;90:2556–2562.
533. Epaud R, Aubey F, Xu J, et al. Knockout of insulin-like growth factor-1 receptor impairs distal lung morphogenesis. *PLoS One*. 2012;7:e48071.
534. Gicquel C, Rossignol S, Cabrol S, et al. Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. *Nat Genet*. 2005;37(9):1003–1007.
535. Qiu Q, Basak A, Mbikay M, Tsang BK, Gruslin A. Role of pro-IGF-II processing by proprotein convertase 4 in human placental development. *Proc Natl Acad Sci U S A*. 2005;102(31):11047–11052.
536. Murphy R, Baptista J, Holly J. Severe intrauterine growth retardation and atypical diabetes associated with a translocation breakpoint disrupting regulation of the insulin-like growth factor 2 gene. *J Clin Endocrinol Metab*. 2008;93:4373–4380.
537. Sparago A, Cerrato F, Vernucci M, et al. Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome. *Nat Genet*. 2004;36(9):958–960.
538. Asvold BO, Askild A, Jenum PVL. Maternal concentrations of insulin-like growth factor-1 and insulin-like growth factor binding protein 1 during pregnancy and birth weight of offspring. *Am J Epidemiol*. 2011;174(2):129–135.
539. Watson CS, Bialek P, Anzo M, et al. Elevated circulating insulin-like growth factor binding protein-1 is sufficient to cause fetal growth restriction. *Endocrinology*. 2006;147(3):1175–1186.
540. Ben Lagha N, Seurin D, Le Bouc Y, et al. Insulin-like growth factor binding protein (IGFBP-1) involvement in intrauterine growth retardation: study on IGFBP-1 overexpressing transgenic mice. *Endocrinology*. 2006;147(10):4730–4737.
541. Qiu Q, Bell M, Lu X, et al. Significance of IGFBP4 in the development of fetal growth restriction. *J Clin Endocrinol Metab*. 2012;97(8):E1429–E1439.
542. Forhead AJ, Li J, Gilmour RS, Dauncey MJ, Fowden AL. Thyroid hormones and the mRNA of the GH receptor and IGFs in skeletal muscle of fetal sheep. *Am J Physiol Endocrinol Metab*. 2002;282(1):E80–E86.
543. Mehta A, Hindmarsh P, Stanhope R. The role of growth hormone in determining birth size and early postnatal growth, using congenital growth hormone deficiency (GHD) as a model. *Clin Endocrinol*. 2005;63:223–231.
544. Handwerger SFM. The roles of placental growth hormone and placental lactogen in the regulation of human fetal growth and development. *J Pediatr Endocrinol Metab*. 2000;13(4):343–356.
545. Pilistine S, Moses A, Munro HN. Placental lactogen administration reverses the effect of low protein diet on maternal and fetal somatomedin levels in the pregnant rat. *Proc Natl Acad Sci U S A*. 1984;81:5853–5857.
546. Gohike BC, Fahnenstich HDC. Longitudinal data for intrauterine levels of fetal IGF1 and IGFII. *Horm Res*. 2004;61:200–204.
547. Hill DJ, Crace CJSA. Regulation of amino acid uptake and dexyribonucleic acid synthesis in isolated human fetal fibroblasts and myoblasts: effects of human placenta lactogen, somatomedin-C, multiplication stimulating activity and insulin. *J Clin Endocrinol Metab*. 1986;62:753–760.
548. Sferruzzi-Perri A, Sandovici I, Constancia M, Fowden AL. Placental phenotype and the insulin-like growth factors: resource allocation to fetal growth. *J Physiol*. 2017;595(15):5057–5093.
549. Li Y, Komuro Y, Fahrion J, et al. Light stimuli control neuronal migration by altering of insulin-like growth factor 1 (IGF-1) signaling. *Proc Natl Acad Sci U S A*. 2012;109:2630–2635.
550. Susa JB, McCormick KL, Widness JA. Chronic hyperinsulinemia in the fetal rhesus monkey. *Diabetes*. 1979;28:1058–1063.
551. Accili D, Drago J, Lee EJ, et al. Early neonatal death in mice homozygous for a null allele of the insulin receptor gene. *Nat Genet*. 1996;12(1):106–109.
552. Krook A, Brueton L, O'Rahilly S. Homozygous nonsense mutation in the insulin receptor gene in an infant with leprechaunism. *Lancet*. 1993;342:277–278.
553. Goetz R, Mohammadi M. Exploring mechanisms of FGF signalling through the lens of structural biology. *Nat Rev Mol Cell Biol*. 2013;14(3):166–180.
554. Su N, Jin M, Chen L. Role of FGF/FGFR signaling in skeletal development and homestasis: learning from mouse models. *Bone Res*. 2014;2:14003.
555. Bottcher RT, Niehrs C. Fibroblast growth factor signaling during early vertebrate development. *Endocr Rev*. 2005;26(63–77).
556. Liu LM, Nicoll CS. Evidence for a role of basic fibroblast growth factor in rat embryonic growth and differentiation. *Endocrinology*. 1988;123:2027–2031.
557. Dorey K, Amaya E. FGF signalling: diverse roles during early vertebrate embryogenesis. *Development*. 2010;137(22):3731–3742.
558. Falardeau J, Chung W, Beenken A, et al. Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *J Clin Invest*. 2008;118:2822–2831.
559. Dodé C, Levilliers J, Dupont JM, et al. Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat Genet*. 2003;33(4):463–465.
560. Guasti L, Silvennoinen S, Bulstrode NW, et al. Elevated FGF21 leads to attenuated postnatal linear growth in preterm infants through GH resistance in chondrocytes. *J Clin Endocrinol Metab*. 2014;99(11):E2198–E2206.
561. Chang H, Brown CW, Matzuk MM. Genetic analysis of the mammalian transforming growth factor B superfamily. *Endocr Rev*. 2002;23:787–823.
562. Janssens K, Dijke P, Janssens S, Van Hul W. Transforming growth factor-B1 to the bone. *Endocr Rev*. 2005;26:743–774.

563. Stiver S, Dvorak H. Vascular permeability factor/vascular endothelial growth factor (VPF/VEGF). *J Clin Ligand Assay*. 2000;23:193–205.
564. Betsholtz C. Functions of platelet-derived growth factor and its receptors deduced from gene inactivation in mice. *J Clin Ligand Assay*. 2000;223:206213.
565. Dwivedi PP, Lam N, Powell B. Boning up on glypicans-opportunities for new insights into bone biology. *Cell Biochem Funct*. 2013;31(2):91–114.
566. Gluckman PD, Sizonenko SV, Bassett NS. The transition from fetus to neonate: an endocrine perspective. *Acta Paediatr Suppl*. 1999;428:7–11.
567. Birk E, Iwamoto HS, Heymann M. Hormonal effects on circulating changes during the perinatal period. *Baillieres Clin Endocrinol Metab*. 1989;3:795–815.
568. Wallace M, Hooper S, Harding R. Effects of elevated fetal cortisol concentrations on the volume, secretion, and reabsorption of lung liquid. *Am J Physiol*. 1995;269:R881–R887.
569. Forhead AJ, Curtis K, Kaptein E, Visser TJ, Fowden AL. Developmental control of iodothyronine deiodinases by cortisol in the ovine fetus and placenta near term. *Endocrinology*. 2006;147:5988–5944.
570. Reik W, Constância M, Fowden A, et al. Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *J Physiol*. 2003;547(1):35–44.
571. Ballard PL. Antenatal glucocorticoid therapy: clinical effects. *Monogr Endocrinol*. 1986;28:137–172.
572. Vermes I, Dohanics J, Toth G, et al. Maturation of the circadian rhythm of the adrenocortical functions in human neonates and infants. *Horm Res*. 1980;12(5):237–244.
573. Sorensen N, Maloney SK, Illow JJ, Mark PJ. Endocrine consequences of circadian rhythm disruption in early life. *Curr Opin Endocr Metab Res*. 2020;11:65–71.
574. Antonini SRR, Jorge SM, Moreira AC. The emergence of salivary cortisol circadian rhythm and its relationship to sleep activity in preterm infants. *Clin Endocrinol*. 2000;52(4):423–426.
575. Ivars K, Nelson N, Theodorsson A, et al. Development of salivary cortisol circadian rhythm in preterm infants. *PLoS One*. 2017;12:1–15.
576. Halliday H. Update on postnatal steroids. *Neonatology*. 2017;111(4):415–422.
577. Baud O, Trousson C, Biran V, et al. Association between early low-dose hydrocortisone therapy in extremely preterm neonates and neurodevelopmental outcomes at 2 years of age. *JAMA*. 2017;317(13):1329–1337.
578. Padbury JF. Functional maturation of the adrenal medulla and peripheral sympathetic nervous system. *Baillieres Clin Endocrinol Metab*. 1989;3(3):689–705.
579. Padbury JF, Agata Y, Ludlow J, et al. Effect of fetal adrenalectomy on catecholamine release and physiologi adaptation at birth in sheep. *J Clin Invest*. 1987;80:1096–1103.
580. Polk DHFDA. Fetal and neonatal thryoid physiology. In: Polin RA, Abman S, Rowitch D, Benitz W, eds. *Fetal and Neonatal Physiology*. 3rd ed. Philadelphia, PA: WB Saunders; 2004:1926–1933.
581. Polk D. Thyroid hormone effects on neonatal thermogenesis. *Semin Perinatol*. 1988;12(2):151–156.
582. Gunn T, Gluckman P. Perinatal thermogenesis. *Early Hum Dev*. 1995;42:169–183.
583. Obregon M-J. Thyroid hormone and adipocyte differentiation. *Thyroid*. 2008;18:185–195.
584. Cypress AM, Kahn CR. The role and importance of brown adipose tissue in energy homeostasis. *Curr Opin Pediatr*. 2010;22(4):478–484.
585. van Marken Lichtenbelt W, Vanhommerig J, Smulders N, et al. Cold-activated brown adipose tissue in healthy men. *N Engl J Med*. 2009;360:1500–1508.
586. Lidell M, Betz M, Dahlqvist Leinhard O, et al. Evidence for two types of brown adipose tissue in humans. *Nat Med 2013*. 2013;19:631–634. 19631-634.
587. Stanley CA, Rozance PJ, Thornton PS, et al. Re-evaluating "transitional neonatal hypoglycemia": mechanism and implications for management. *J Pediatr*. 2015. doi:10.1016/j.jpeds.2015.02.045.
588. Sperling MA. Carbohydrate metabolism: insulin and glucagons. In: Tulchinsky D, Little A, eds. *Maternal-Fetal Endocrinology*. 2nd ed. Philadelphia, PA: WB Saunders; 1994:380–400.
589. Girard J. Control of fetal and neonatal glucose metabolism by pancreatic hormones. *Baillieres Clin Endocrinol Metab*. 1989;3:817–836.
590. Girard J, Ferre P, Pegorier J, Duée PH. Adaptations of glucose and fatty acid metabolism during perinatal period and suckling-weaning transition. *Physiol Rev*. 1992;72:507–562.
591. Guenther M, Bruder E, Raff H. Effects of body temperature maintenance on glucose, insulin, and corticosterone responses to acute hypoxia in the neonatal rat. *Am J Physiol Regul Integr Comp Physiol*. 2012;302:R627–R633.
592. Harris DL, Weston PJ, Harding JE. Lactate, rather than ketones, may provide alternative cerebral fuel in hypoglycaemic newborns. *Arch Dis Child Fetal Neonatal Ed*. 2015. doi:10.1136/archdischild-2014-306435.
593. Harris DL, Weston PJ, Signal M, Chase JG, Harding JE. Dextrose gel for neonatal hypoglycaemia (the Sugar Babies Study): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2013.
594. Hegarty JE, Harding JE, Gamble GD, et al. Prophylactic oral dextrose gel for newborn babies at risk of neonatal hypoglycaemia: a randomised controlled dose-finding trial (the Pre-hPOD Study). *PLoS Med*. 2016;13(10):e1002155.
595. Rozance PJ, Hay WW. New approaches to management of neonatal hypoglycemia. *Matern Heal Neonatol Perinatol*. 2016;2:3.
596. Longhead J, Minouni F, Tsang R. Serum ionized calcium concentrations in normal neonates. *Am J Dis Child*. 1988;142:516–518.
597. Cooper MS. Disorders of calcium metabolism and parathyroid disease. *Best Pr Res Clin Endocrinol Metab*. 2011;25(6):975–983.
598. Venkataraman P, Tsang R, Chen I. Pathogenesis of early neonatal hypocalcemia: studies of serum calcitonin, gastrin and plasma glucagon. *J Pediatr*. 1987;110:599–603.
599. Kovaks C, Kronenberg H. Maternal-fetal calcium and bone metabolism during pregnancy, puerperium and lactation. *Endocr Rev*. 1997;18:832–872.
600. Mimoumi F, Tsang R. Perinatal mineral metabolism. In: Tulchinsky D, Little A, eds. *Maternal-Fetal Endocrinology*. Philadelphia, PA: WB Saunders; 1994:402–417.
601. Grumbach M, Gluckman P. The human fetal hypothalamus and pituitary gland: the maturation of neuroendocrine mechanisms controlling secretion of fetal pituitary growth hormone, prolactin, gonadotropins, adrenocorticotropin-related peptides, and thyrotropin. In: Tulchinsky D, Little A, eds. *Maternal Fetal Endocrinology*. 2nd ed. Philadelphia, PA: WB Saunders; 1994:193–261.
602. Penny R, Parlow AF, Frasier S. Testosterone and estradiol concentrations in paired maternal and cord sera and their correlation with the concentration of chorionic gonadotropin. *Pediatrics*. 1979;64(5):604–608.
603. Waddington S, Kennea N, Buckley S. Fetal and neonatal gene therapy: benefits and pitfalls. *Gene Ther*. 2004;11:592–597.
604. Keswani SG, Balaji S, Katz AB, et al. Intraplacental gene therapy with Ad-IGF-1 corrects naturally occurring rabbit model of intrauterine growth restriction. *Hum Gene Ther*. 2015;26(3):172–182.

131

Hormonal Changes and Pregnancy Testing

Denis A. Vaughan, Katherine Koniares, Mark E. Molitch, and Errol R. Norwitz

OUTLINE

Introduction, 2200
Pituitary Gland, 2201
  Hypopituitarism, 2201
Prolactin, 2202
  Changes in Prolactin Physiology, 2202
  Prolactinomas in Pregnancy, 2202
  The Effect of Pregnancy on a Preexisting Prolactinoma, 2202
  The Effect of Dopamine Agonists on a Developing Fetus, 2202
  Prolactin and Antipsychotic Medication, 2203
  Breastfeeding and Prolactin, 2203
Growth Hormone, 2203
  Changes in GH Physiology During Pregnancy, 2203
  GH Deficiency, 2203
  GH Excess, 2203
  Vasopressin, 2203
Thyroid Gland, 2205
  Thyroid Hormones and Thyroid Regulation, 2205
  Hyperthyroidism, 2207
  Hypothyroidism, 2208
Adrenal Gland, 2209
  Regulation of Glucocorticosteroid Synthesis, 2209
Parathyroid Glands, 2213
  Regulation of Calcium During Pregnancy, 2213
  Parathyroid Hormone–Related Protein, 2214
  Vitamin D Deficiency, 2214
  Hypercalcemia, 2214
Conclusion, 2215

INTRODUCTION

Pregnancy alters normal physiology and thereby endocrine testing in a number of ways (Table 131.1). The normal physiologic adaptations of pregnancy include an expanded plasma and red blood cell volume. The disproportionate increase in plasma volume results in a mild generalized hemodilution. The increase in plasma volume begins at 6 to 8 weeks of gestation and reaches its zenith at a 40% increase at 30 to 34 weeks.[1] This increase in plasma volume is accompanied by an increase in cardiac output (CO) that appears to peak in the midgestation. This increase is initially due to an increase in stroke volume and subsequently in heart rate.[2] In the first half of pregnancy, there is a decrease in systemic vascular resistance and blood pressure. These changes serve to optimize blood flow to facilitate fetal growth and development. The increase in CO leads to a comparable increase in renal plasma flow and glomerular filtration rate (GFR). The increase in GFR is apparent by 4 weeks of gestation and increases to 140% to 150% of the nonpregnant state by 6 to 12 weeks.[2,3] GFR increases by another 10% over the remainder of the second trimester, stabilizes in the third trimester, and falls in the last 4 weeks of gestation.[3,4] Changes in serum creatinine concentrations show a similar pattern. In a retrospective study including over 240,000 pregnancies in Canada, the mean serum creatinine concentration dropped in the first trimester (beginning at 4 weeks of gestation), leveled off in the second trimester, and then gradually rose in the third trimester to prepregnancy levels.[5] These changes in GFR alter the clearance of various hormones, thereby changing their serum and urinary concentrations.

In addition to GFR, renal tubular function is also altered. For example, tubular reabsorption of glucose is decreased leading to significant glycosuria despite normal plasma glucose levels.[6] In contrast, tubular reabsorption of sodium is increased due to increased GFR, increased placental progesterone production, increased concentrations of aldosterone and cortisol, and exaggerated postural effects.[3]

Hormone production by the placenta may also profoundly affect circulating levels of other hormonal systems. Estradiol levels increase 100-fold in pregnancy, beginning at 6 to 8 weeks of gestation[7] (Fig. 131.1). These elevated estrogen levels cause an increase in circulating concentrations of a number of hormone-binding globulins produced by the liver, including thyroxine-binding globulin (TBG), corticosteroid-binding globulin (CBG), and sex hormone–binding globulin (SHBG). In such instances, measurements of total (bound plus free) hormones are increased, although free hormone levels are largely unaffected (discussed later). Estrogens also have a direct stimulatory effect on pituitary lactotrophs, with increased production of prolactin (PRL). A number of peptide hormones are produced directly by the placenta, including human chorionic gonadotropin (hCG), chorionic somatomammotropin (previously referred to as placental lactogen), placental growth hormone (GH), chorionic thyrotropin, chorionic adrenocorticotropic hormone (ACTH), chorionic corticotropin-releasing hormone (CRH, corticotropin), urocortin, chorionic gonadotropin-releasing hormone (GnRH), and chorionic somatostatin. Some of these peptides interfere with normal circulating hormone levels. For example, placental GH suppresses pituitary GH secretion indirectly through the stimulation of insulin-like growth factor-1 (IGF-1).[8] Although hCG does not directly affect the pituitary gland, its structure is similar to that of other heterodimer glycoprotein hormones with identical alpha and distinct beta subunits (chains), such as luteinizing hormones (LH) and thyroid-stimulating hormone (TSH, thyrotropin). As such, very high levels of hCG (such as that seen in complete molar pregnancies) may activate the TSH receptor resulting in hyperthyroidism. The placenta also produces

TABLE 131.1 Changes in Normal Physiology That May Affect Endocrine Tests

Cardiovascular

- Increased plasma and RBC volume
- Increased cardiac output
  - Lowered "osmostat" for vasopressin release and thirst

Renal

- Increased GFR and renal plasma flow
  - Increased hormone and substrate clearance
- Altered renal tubular function
  - Decreased Tm for glucose

Placental Hormone Production

- Increased estrogen and progesterone production
  - Increased hormone-binding globulin production
- Stimulation of pituitary lactotrophs
- Production of peptide hormones
  - ACTH, CRH, GnRH, hCG, GH variant, hPL, cTSH, PRL

Placental Enzyme Production

- Vasopressinase

ACTH, Adrenocorticotropic hormone; *CRH*, corticotropin-releasing hormone; *cTSH*, chorionic TSH; *GH*, growth hormone; *GnRH*, gonadotropin-releasing hormone; *hCG*, human chorionic gonadotropin; *hPL*, human placental lactogen; *PRL*, prolactin.

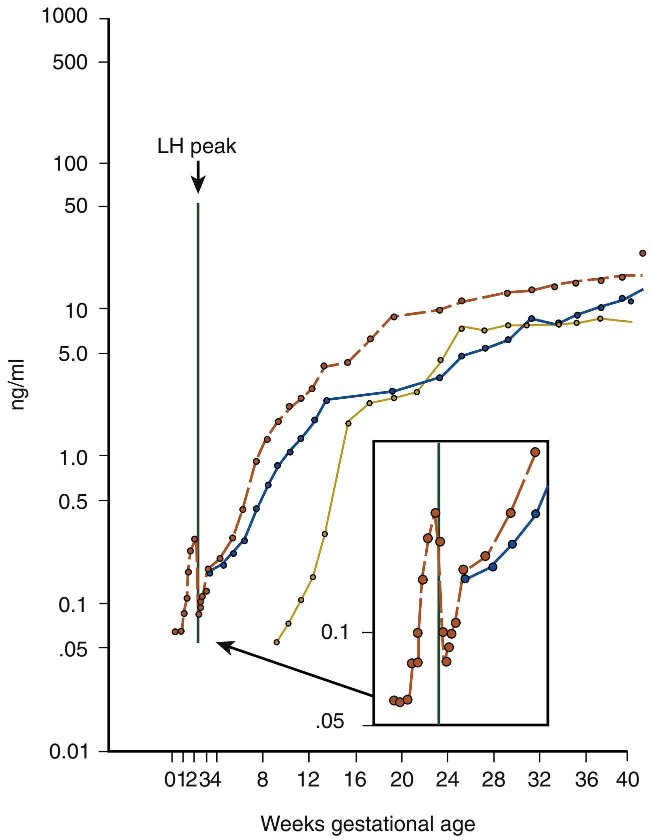


Figure 131.1 **Maternal serum levels of unconjugated estrone (E_1), estradiol (E_2), and estriol (E_3) from last menstrual flow through a conceptual cycle to term pregnancy.** (Data from Buster JE, Abraham GE. The applications of steroid hormone radioimmunoassays to clinical obstetrics. *Obstet Gynecol.* 1975;46:489–499. Reprinted with permission from the American College of Obstetricians and Gynecologists.)

1,25-dihydroxyvitamin D (DHCC), which regulates calcium homeostasis during pregnancy. Placental CRH has numerous actions within the fetoplacental unit, but is likely not involved in maternal ACTH regulation.[9] Placental estrogen, progesterone, lactogen, and GH also function as counterregulatory hormones and cause insulin resistance and accelerated lipolysis. In addition to hormones, the placenta produces enzymes that may affect normal endocrine function. Placental vasopressinase activity leads to the increased production of maternal vasopressin to maintain normal circulating levels.

With this background in mind, the current chapter reviews the changes in normal physiology that occur during pregnancy and the resulting alternations in circulating hormone levels that may create diagnostic difficulties when attempting to evaluate endocrine disorders in pregnant patients.

PITUITARY GLAND

Hypopituitarism

Lymphocytic Hypophysitis. Lymphocytic hypophysitis is a rare primary autoimmune endocrinopathy, more common in women than men, that results in pituitary enlargement. It is characterized histologically by massive infiltration of the pituitary gland by lymphocytes and plasma cells, with destruction of the normal parenchyma. Patients commonly present with symptoms of pituitary mass effect such as headache or visual field changes secondary to compression of the optic chiasm.[10] Traditionally, up to 50% of such cases were thought to affect pregnant women, particularly in the last month of pregnancy and the first 2 months postpartum.[11] More recent studies suggest that far fewer cases are associated with pregnancy.[12] Lymphocytic hypophysitis may be associated with signs of diabetes insipidus in up to 20% cases with or without PRL abnormalities.[11] In more serious cases, it may present with corticotropin deficiency, presenting as fatigue, pallor, or even a life-threatening adrenal crisis.[13] Any symptoms or signs of pituitary mass effect or hypopituitarism in pregnancy warrant serum hormone evaluations and magnetic resonance imaging (MRI). If PRL levels are only modestly elevated (<150 ng/mL) in the presence of a large mass, the diagnosis is unlikely to be a prolactinoma and more likely to be hypophysitis or a nonfunctional pituitary adenoma. Ultimately, the diagnosis may require histological examination to differentiate it from a tumor.[13] Lymphocytic hypophysitis in the postpartum period may be confused with Sheehan syndrome, although such women typically do not have a history of obstetrical hemorrhage.

Sheehan Syndrome. Sheehan syndrome refers to infarction of the pituitary gland following obstetrical hemorrhage. In most cases, presenting symptoms are nonspecific. Some patients present with symptoms of adrenal insufficiency (AI) such as lethargy, anorexia, and/or an inability to lactate in the first few days and weeks after delivery.[14] If the pituitary infarction is incomplete, the development of hypopituitarism may be more insidious and difficult to diagnose. A complete pituitary hormone workup should be performed in any patient with a history of obstetric hemorrhage with signs/symptoms of hypopituitarism. This is a permanent injury. Treatment is lifelong hormonal replacement.

Pituitary Apoplexy. Pituitary apoplexy refers to an acute hemorrhage into the pituitary gland that occurs most often in the setting of a pituitary adenoma. Clinical presentation may include sudden onset of severe headache, diplopia due to pressure on the oculomotor nerves, and panhypopituitarism. Surgical decompression may be required for severe or progressive neurological or visual symptoms.[15] Symptoms may improve immediately after surgical decompression or gradually with expectant management over a period of weeks to months in milder cases as the blood slowly gets reabsorbed.

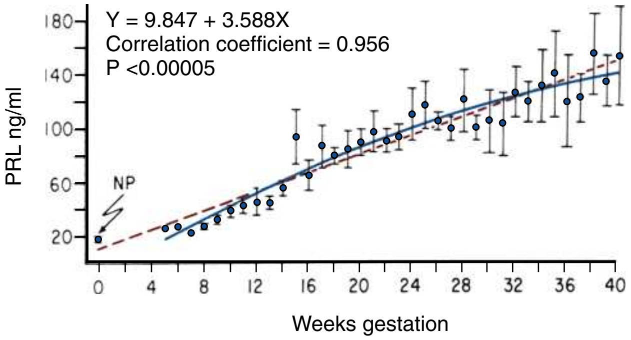


Figure 131.2 Mean ± SEM (*n* = 4) serum prolactin (PRL) concentrations measured serially at weekly intervals as a function of the duration of gestation. (Data from Rigg LA, Lein A, Yen SS. Pattern of increase in circulating prolactin levels during human gestation. *Am J Obstet Gynecol.* 1977;129:454–456.)

Nonfunctional Pituitary Adenoma. These are benign tumors arising from the adenohypophyseal cells and are characterized by an absence of hormonal hypersecretion, both functionally and biochemically. Nonfunctional pituitary adenomas constitute approximately 25% of all pituitary adenomas.[16] The tumor's size or surgical treatment may lead to functional hypopituitarism, which is associated with increased mortality rates.[17]

PROLACTIN

Changes in Prolactin Physiology

During pregnancy, estrogen production by the placenta stimulates lactotroph hyperplasia resulting in an increase in PRL gene expression and PRL protein synthesis.[18] The pituitary volume increases up to 136%.[19] Excessive estrogen production has been implicated in the pathophysiology of hyperprolactinemia and prolactinomas.[20] Progesterone also stimulates PRL secretion.[21] Together, these changes result in a progressive rise in serum PRL levels[22] (Fig. 131.2) that parallels the increase in the size and number of pituitary lactotroph cells within the pituitary. By term, PRL levels may be increased 10-fold relative to prepregnancy levels, with absolute circulating concentrations over 200 ng/mL.[23] These elevated PRL levels are thought to prepare the breast for lactation.

PRL identical to that made by the maternal pituitary is produced also by the decidua during pregnancy, resulting in high amniotic fluid PRL levels.[24] This decidual PRL is not present in maternal blood and its physiologic role may be limited to regulating water and ion transport across extraembryonic membranes.[24] The placenta also produces chorionic somatomammotropin, previously known as placental lactogen. Although it has lactogenic activity, it is structurally more closely related to GH than to PRL. It does not contribute to blood PRL levels during pregnancy.[25]

Prolactinomas in Pregnancy

A prolactinoma (or lactotroph adenoma) is the most common type of pituitary adenoma.[16] There are two major concerns in patients with hyperprolactinemia who conceive:

1. The effect of pregnancy on a preexisting prolactinoma
2. The effect of dopamine agonists on the developing fetus

The Effect of Pregnancy on a Preexisting Prolactinoma

In women with a preexisting prolactinoma, there is a risk of tumor enlargement during pregnancy not only because of estrogen lactotroph hyperplasia but also because of cessation of dopamine agonist treatment. It is estimated that the risk of symptomatic tumor enlargement in pregnancy is 2% for microadenomas but can be as high as 23% in macroadenomas that have not been surgically treated or irradiated (Table 131.2). Monitoring of PRL levels in pregnancy is controversial, but such patients should be closely monitored for clinical symptoms of pituitary enlargement (headaches and visual changes). If a patient develops symptoms, a pituitary MRI without contrast should be performed (Fig. 131.3). If the prolactinoma is noted to be enlarged, then dopamine agonist treatment is recommended as first line with surgery typically reserved for failed medical management.[26]

TABLE 131.2 Effect of Pregnancy on Prolactinomas

Tumor Type	Prior Therapy	Number	Symptomatic Enlargement
Microadenomas	None	658	18 (2.7%)
Macroadenomas	None	214	49 (22.9%)
Macroadenomas	Yes	148	7 (4.8%)

Data from Molitch ME. Prolactinoma in pregnancy. *Best Pract Res Clin Endocrinol Metab.* 2011;25:885–896.

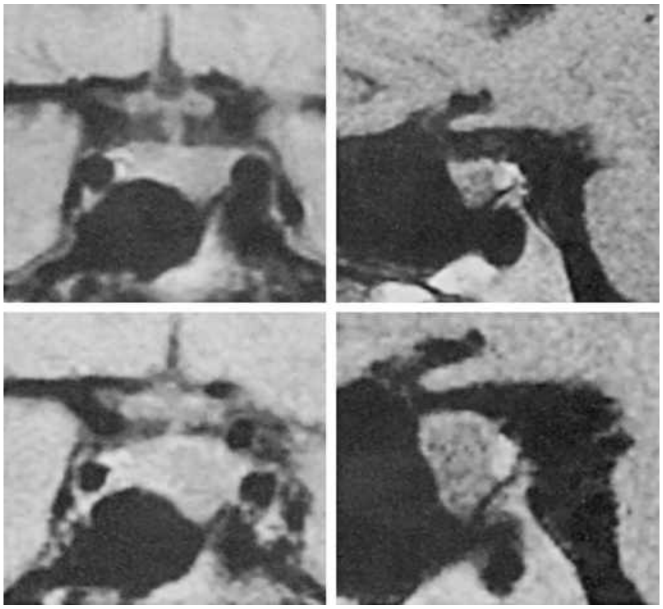


Figure 131.3 Magnetic resonance imaging scans with coronal *(left)* and sagittal *(right)* views demonstrating enlargement of a prolactin secreting macroadenoma before pregnancy *(above)* and in the third trimester of pregnancy *(below)*. The patient had been complaining of increasing headaches.

The Effect of Dopamine Agonists on a Developing Fetus

Maternal and fetal hormone levels are derived from their respective pituitary glands. As such, dopamine agonist treatment will suppress PRL levels in both the fetal and maternal compartment, although such medications have not been shown to adversely affect fetal growth and development. Decidual PRL production is not affected by dopamine agonist treatment and amniotic fluid PRL levels remain unchanged. Although dopamine agonists are typically stopped once pregnancy is confirmed, the existing safety data are reassuring. Data from over 6000 pregnancies suggest that the administration of bromocriptine during the first month of pregnancy does not harm the fetus.[27] In another study of children followed up to 9 years after *in utero* exposure to bromocriptine, no adverse effects were noted.[28] Although there are more limited data available for cabergoline, the reports that do exist are reassuring with no apparent increased risks to the pregnancy or the fetus.[26]

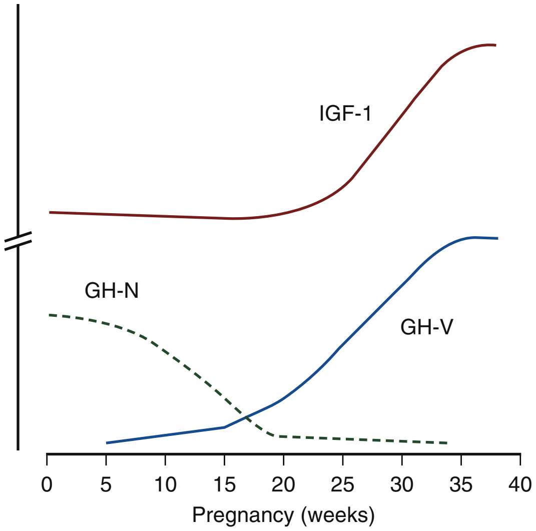


Figure 131.4 Relative changes of GH-V, GH-N, and insulin-like growth factor-1 (IGF-1) in the maternal circulation during pregnancy.

Prolactin and Antipsychotic Medication

The high prevalence of hyperprolactinemia with the use of antipsychotic medications relates to the blockade of D2 receptors. Compared with second-generation antipsychotics, first-generation medications are associated with higher PRL levels, with the exception of risperidone and paliperidone.[29,30] Antipsychotic use in pregnancy is uncommon and is seen in <1% of pregnant women in the first trimester.[31] These medications do not seem to confer an additional risk of congenital fetal malformations, with the possible exception of risperidone that has been associated in some studies with an increased risk of congenital anomalies, in particular cardiac defects.[31]

Breastfeeding and Prolactin

Nipple stimulation during breastfeeding increases serum PRL secretion, likely via a neural pathway. This effect is proportional to the increase in lactotroph hyperplasia that occurs during pregnancy. In the first few weeks postpartum, serum PRL concentrations may increase up to 300 ng/mL above baseline in response to suckling. This response is blunted over time such that, by 6 months after delivery, the increase in PRL in response to suckling is typically less than 10 ng/mL above baseline.[32]

KEY POINTS

- Prolactinoma is the most common type of pituitary adenoma seen in pregnancy.
- The risk of symptomatic tumor enlargement in pregnancy is 2% for microadenomas and up to 23% in macroadenomas.

GROWTH HORMONE

Changes in GH Physiology During Pregnancy

In humans, there are two separate genes coding for different variants of GH that differ by 13 amino acids. GH-N is produced by the pituitary and is responsible for postnatal growth. GH-V is released from the placenta and is the predominant form of GH in the maternal circulation during pregnancy[33] (Fig. 131.4). It can be detected in the maternal circulation as early as 5 weeks' gestation.[34] Beginning at 15 to 20 weeks, GH-V levels rapidly increase and GH-N levels become undetectable. Levels of GH-V peak at 35 to 37 weeks' gestation. The increase in hepatic production of IGF-1 in pregnancy is attributed to high levels of GH-V.[35] The decrease in GH-N is a result of negative feedback from GH-V and IGF-1.[36] In contrast to the pulsatile release of GH-N by the anterior pituitary, GH-V release by the syncytiotrophoblast and extravillous trophoblast cells is continuous in nature.[33]

Both GH-N and GH-V are inhibited by hyperglycemia.[37] In pregnancies complicated by gestational diabetes, GH-V (but not GH-N) levels decrease significantly following an oral glucose load.[35] GH-V relaxes arteries including those supplying the uterus, which results in an increased blood supply to the fetus.[38] Interestingly, GH-V levels are significantly lower in pregnancies complicated by intrauterine growth restriction,[39] but significantly elevated in women with severe preeclampsia.[53] Initial studies were unable to detect GH-V in the fetal circulation, suggesting an indirect role in regulating fetal growth.[35] However, more recent studies were able to measure GH-V in umbilical cord blood samples, but at levels much lower than in the maternal circulation.[39] The likely source is directly from the syncytiotrophoblast rather than transplacental passage from the maternal circulation.[35]

GH Deficiency

The effect on the fetus of low levels of GH is less concerning after 15 to 20 weeks' gestation when GH-V levels increase.[40] The rise in GH-V is unaffected by GH supplementation in GH-deficient patients.[41] Interestingly, women with underlying GH-N deficiency have normally rising levels of IGF-1 during pregnancy.[42] Successful pregnancies have been reported with GH replacement in the first trimester with tapering in the second trimester as well as with no GH replacement in pregnancy.[43] As such, GH replacement during pregnancy is not routinely recommended.[40]

GH Excess

In patients with GH excess, GH-N secretion is not subject to negative feedback and circulating levels do not decrease during pregnancy as GH-V levels rise.[40] It may be difficult to diagnose GH excess during pregnancy because commonly used diagnostic assays do not differentiate GH-N from GH-V. If a specific assay is unavailable, the definitive diagnosis of GH excess may have to be deferred to the immediate postpartum period, since GH-V becomes undetectable within 30 to 60 minutes of delivery.[34] Continued pulsatile secretion of GH during pregnancy suggests elevated GH-N, although this can be difficult to detect.[36] Of note, thyrotropin-releasing hormone (TRH) stimulates secretion of GH-N, whereas GH-V is not responsive to TRH in pregnant patients with acromegaly.[36] Elevated GH in pregnancy has been associated with gestational diabetes and hypertension.[44] An additional concern regarding acromegaly in pregnancy is enlargement of the causal pituitary adenoma resulting in neurological and/or visual symptoms.[45] Bromocriptine and octreotide have been safely used to treat acromegaly in pregnancy.[46]

KEY POINTS

- GH-V is released from the placenta and is the predominant form of GH in the maternal circulation during pregnancy. GH-N is produced by the pituitary and is responsible primarily for postnatal growth.
- In contrast to the pulsatile release of GH-N by the anterior pituitary, GH-V release by trophoblast cells is continuous.

Vasopressin

Changes in Vasopressin Physiology During Pregnancy. Plasma osmolality is the main driver of vasopressin release (see Chapter 12). In nonpregnant women, normal plasma osmolality is 275 to 290 mOsm/kg and the osmotic threshold for vasopressin release is 280 to 285 mOsm/kg. Thirst is triggered at a plasma osmolality approximately 5 mOsm/kg above the threshold for vasopressin release, at which time maximal urine concentration will have already occurred. In normal pregnancy,

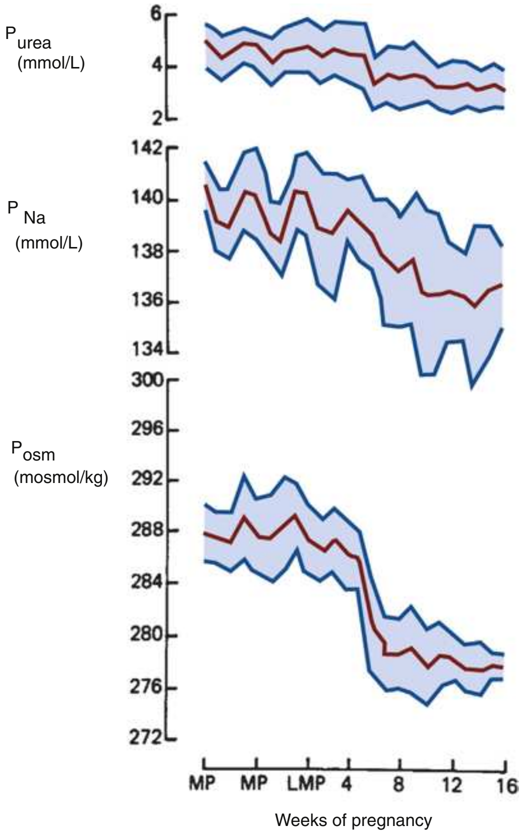


Figure 131.5 **Mean values (±SD) for plasma urea *(Purea)*, sodium *(PNa)*, and osmolality *(Posm)* measured at weekly intervals from before conception to the first trimester in nine women with a successful obstetric outcome.** *LMP,* Last menstrual period; *MP,* menstrual period. (Data from Davison JM, Vallotton MB, Lindheimer MD. Plasma osmolality and urinary concentration and dilution during and after pregnancy: evidence that lateral recumbency inhibits maximal urinary concentrating ability. *Br J Obstet Gynaecol.* 1981;88:472–479.)

the average plasma osmolality is decreased by 5 to 10 mOsm/kg.[47] A decrease in plasma osmolality can be detected as early as the first missed menstrual period and a further decrease by 5 to 10 mOsm/kg is typically seen by the 10th week of gestation, at which time it plateaus for the remainder of pregnancy[48] (Fig. 131.5). The decrease in plasma osmolality results from an increase in total body water and plasma volume expansion of greater than 1.5-times normal.[49] The threshold for vasopressin release is about 6 mOsm/kg lower during pregnancy and thirst is triggered at a plasma osmolality about 10 mOsm/kg lower during pregnancy compared to the nonpregnant state[50] (Fig. 131.6).

Vasopressinase is a cysteine aminopeptidase produced by trophoblast cells during pregnancy that rapidly metabolizes vasopressin.[51,52] Vasopressinase production begins in the seventh week of gestation and levels rise 1000-fold between 7 and 40 weeks, coincident with the increasing placental mass.[52] In normal pregnancies, increased synthesis and release of vasopressin by the posterior pituitary is sufficient to overcome the effects of the vasopressinase.[52] As such, plasma levels of vasopressin during pregnancy are similar to that in nonpregnant women at 1 to 2 pg/mL, despite the 3- to 4-fold increase in clearance.[51] Vasopressinase is metabolized by the liver, which accounts for the higher circulating levels seen in patients with acute or chronic liver dysfunction.[49] Circulating vasopressinase levels become undetectable 5 to 6 weeks after delivery.[52]

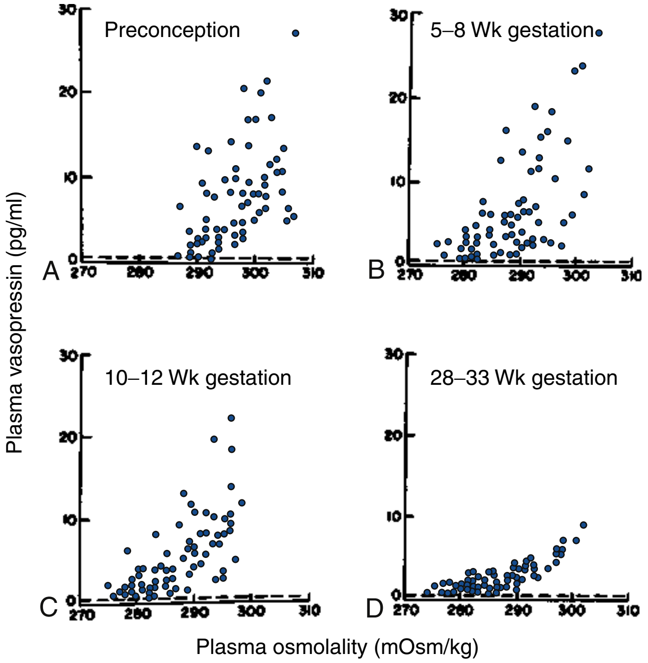


Figure 131.6 **Relationship of plasma arginine vasopressin to plasma osmolality during a 2-hour hypertonic saline infusion in eight women studied serially before and during pregnancy.** Each point is an individual plasma determination. *Dashed lines* indicate the lower limit of detection for arginine vasopressin. (Data from Lindheimer MD, Barron WM, Davison JM. Osmoregulation of thirst and vasopressin release in pregnancy. *Am J Physiol.* 1989;257:F159–F169.)

Diabetes Insipidus of Pregnancy. Gestational diabetes insipidus (GDI) occurs when the increased release of vasopressin from the posterior pituitary is not sufficient to overcome the elevated levels of vasopressinase resulting in a functional deficiency of vasopressin, the hormone required to reabsorb water from the renal glomerular filtrate and concentrate the urine. Patients present with hypotonic polyuria and excessive water consumption.[51,52] GDI affects 2 to 4 per 100,000 pregnancies and is seen most commonly at the end of the second trimester or during the early third trimester.[53] Additional risk factors include multifetal gestations (because there is a larger placental mass to synthesize vasopressinase) and pregnancies complicated by preeclampsia/hemolysis, elevated liver enzymes, low platelet count [HELLP] syndrome or acute fatty liver of pregnancy (because liver dysfunction decreases the metabolism of vasopressinase).[54] GDI usually resolves spontaneously within 4 to 6 weeks of delivery.[53] Acute postpartum DI is rare but has been described in association with placental abruption and after delivery with a sudden release of placental vasopressinase into the bloodstream.[55] The treatment for GDI is desmopressin, a synthetic form of vasopressin that has been structurally modified to make it resistant to degradation by vasopressinase.[56] The dose of desmopressin is similar to or slightly higher than that required for nonpregnant patients. The medication can often be discontinued 4 to 6 weeks after delivery when vasopressinase levels decrease and GDI resolves clinically.[56] Desmopressin appears to be safe in pregnancy and breastfeeding.[57]

Pregnancy-Induced Exacerbation of Preexisting Diabetes Insipidus. Vasopressinase levels increase 1000-fold between 7 and 40 weeks' gestation, which may unmask a preexisting predisposition towards DI.[40] Patients with a history of neurosurgery, head trauma, or a family history of DI are at elevated risk.[51] A water deprivation test, which compares urinary osmolality after dehydration and after administration of vasopressin, may be useful to confirm the diagnosis in nonpregnant patients, but is not recommended during pregnancy.

Diagnostic tests should therefore be limited to measurements of urinary and plasma osmolality. Treatment should be initiated early and not delayed unnecessarily[51] (see Chapter 12 for further discussion of the management of DI).

Syndrome of Inappropriate Antidiuretic Hormone Secretion. Syndrome of inappropriate antidiuretic hormone secretion (SIADH) is rare in pregnancy. Of 18 reported cases of SIADH and hyponatremia in pregnancy, 13 were associated with preeclampsia.[58] Women with preeclampsia are total body fluid overloaded, but intravascularly depleted because of a loss of vascular integrity and third spacing of fluid. The resultant low circulating blood volume may lead to a nonosmotic release of ADH.[59] This may be compounded by the resetting of the osmostat during pregnancy whereby the average plasma osmolality is decreased by 5 to 10 mOsm/kg.[47] Hyponatremia may increase the risk of seizures in patients with preeclampsia[59] (see Chapter 18 for the treatment of SIADH).

KEY POINTS

- In normal pregnancy, the average plasma osmolality is decreased by 5 to 10 mOsm/kg.
- Vasopressinase is produced by placental trophoblasts during pregnancy and rapidly metabolizes vasopressin.
- Vasopressinase levels increase 1000-fold between 7- and 40-weeks' gestation, which may unmask preexisting DI.

THYROID GLAND

Thyroid Hormones and Thyroid Regulation

Changes in Thyroid Physiology During Pregnancy. Pregnancy is associated with a number of changes in the thyroid gland and in levels of circulating thyroid hormones[60] with the ultimate objective of ensuring normal fetal growth and development.[61] Thyroid hormone is critically important in regulating placentation and fetal neurodevelopment, although the relative contributions of the maternal and fetal thyroid axes remain uncertain.[61,62]

Pregnancy is associated with a 10% to 40% increase in the size of the maternal thyroid gland.[63] Increased estrogen levels during the first trimester increase hepatic production of thyroid-binding globulins (TBGs) and decrease their metabolism.[64] This results in a 2-fold increase in circulating TBG levels. Most of the circulating thyroid hormone is bound to TBG. In addition, pregnancy is associated with an increase in iodine utilization, and in thyroid hormone production, which is driven in part by the thyroid stimulatory effects of hCG. hCG is a member of the same family of glycoprotein hormones as TSH, with a common alpha subunit and a unique but homologous beta subunit. As a result, hCG has weak thyroid-stimulating activity.[65] This effect along with the increase in volume of distribution during the first half of gestation[66] results in a decrease in circulating thyroid hormone levels before a steady state is reached around midgestation. In order to maintain adequate circulating levels of biologically active–free (unbound) thyroid hormone, the thyroid gland undergoes substantial changes in thyroid hormone metabolism, iodine uptake, and hypothalamic-pituitary-thyroid axis regulation.[60] The ability to compensate for the increased demands of pregnancy relies on adequate iodine intake and on an intact functional thyroid gland.

Iodine is an absolute requirement for thyroid hormone synthesis and pregnant women have increased dietary iodine requirements. Although dietary intake in Western countries is generally considered adequate in terms of iodine repletion, some studies suggest that over 50% of US women may have suboptimal urinary iodine concentrations.[67] Even the United Kingdom was recently classified as a "mildly iodine deficient" region by the World Health Organization (WHO).[68] The WHO, American Thyroid Association, and Endocrine Society recommend an intake of 250 mcg iodine daily during pregnancy and lactation. For this reason, prenatal vitamins containing 150 to 250 mcg iodine are recommended in the preconception period as well as throughout pregnancy and lactation.[69] Excessive iodine ingestion (defined as >500 mcg/day) should be avoided as this may cause fetal or neonatal goiter.[70]

Severe iodine deficiency may result in fetal and/or maternal goiter and has been associated with increased rates of pregnancy loss, stillbirth, and increased perinatal mortality[71] as well as an adverse effect on fetal and neonatal neurodevelopment. Indeed, iodine deficiency is the leading cause of preventable intellectual deficits globally.[72] Women with mild-to-moderate iodine deficiency are at increased risk of developing a goiter. A clinical diagnosis of goiter in pregnancy should prompt further investigation. Use of radioactive iodine to evaluate a goiter may be useful in the nonpregnant patient, but can be difficult to interpret as iodine uptake is markedly increased in pregnancy.[73] More importantly, use of radioactive iodine is absolutely contraindicated in pregnancy as it freely crosses the placenta and will irreversibly damage or destroy the fetal thyroid, which has an avidity for iodine 20 to 50 times more than the maternal thyroid. Notably, the small tracer dose is not sufficient to cause significant injury if given inadvertently to a pregnant patient.[74] The treatment of women with thyroid cancer using radioactive iodine does not seem to have a detrimental effect on subsequent pregnancies, provided that the pregnancy occurs more than 6 months following the I^{131} treatment.[98]

The fetal thyroid gland and hypothalamic-pituitary-thyroid axis develop independently of maternal thyroid status. At 11 to 12 weeks of gestation, the fetal thyroid gland begins to concentrate iodine. In addition to iodine, the placenta is freely permeable to medications used to treat hyperthyroidism, such as propylthiouracil (PTU), methimazole, and propranolol. In contrast, very little TSH and triiodothyronine (T3) is able to cross the placenta.[60] The prohormone, thyroxine (T4), crosses the placenta in somewhat larger (although still very small) amounts such that, in late gestation, large doses of T4 have been shown to ameliorate the adverse effects of congenital hypothyroidism caused by an enzyme defect or agenesis of the thyroid gland.[60] More commonly, intraamniotic injection of T4 may be used to treat fetal goiter associated with congenital hypothyroidism.[75]

As noted previously, thyroid-stimulating activity is increased in the serum of women in the first trimester of pregnancy as a result of the intrinsic thyroid-stimulating activity of hCG.[76] This may result in elevated circulating T4 and T3 levels as well as a transient clinical hyperthyroidism, which may present as hyperemesis gravidarum, severe nausea, and vomiting of pregnancy with associated weight loss and electrolyte disturbances. The markedly elevated levels of hCG seen with gestational trophoblastic neoplasms may also cause hyperthyroidism, a phenomenon first described by Tisne et al. in 1955.[77]

Alterations in Thyroid Function Tests During Pregnancy. Most thyroid hormone in the circulation is bound and therefore biologically inactive. T4 is 75% bound to TBG, 10% to 15% bound to albumin, and 10% to 15% bound to transthyretin (prealbumin); only about 0.04% is unbound (free) and thus biologically active. Increased estrogen levels in pregnancy result in increased production of TBG by the liver and in a commensurate increase in serum TBG levels, which plateaus in mid-gestation. In addition to the increased production of TBG by hepatocytes, there is also an increase in TBG sialylation resulting in a decrease in plasma clearance.[78] This increase in TBG results in higher

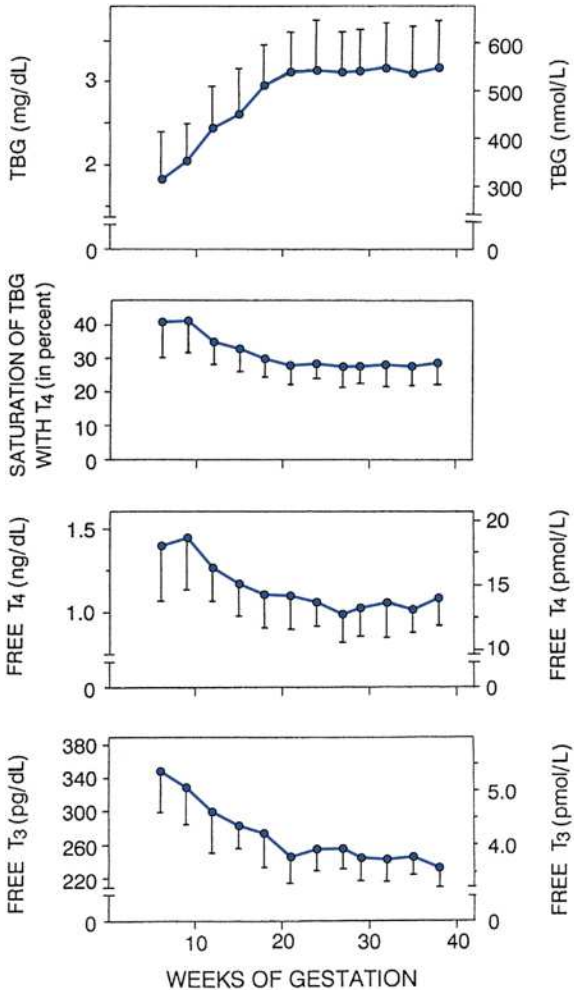


Figure 131.7 **Data collected in 606 normal pregnancies in Brussels are illustrated, showing the progressive rise in serum thyroxine-binding globulin *(TBG)* in the first part of gestation, accompanied by a progressive decrease in the free T index (saturation level of TBG by *T4*), and free T4 and *T3* concentrations.** Brussels is an area with restricted iodine intake, so the quantitative reduction in free hormone concentration observed in the second part of pregnancy is more pronounced than in areas without iodine deficiency. (Data from Glinoer D. Thyroid regulation during pregnancy. In Delange F, Dunn JT, Glinoer D ed. *Iodine Deficiency in Europe: a Continuing Concern.* NATO ASI Series (vol 241). New York: Plenum. 1993, pp. 181–190. © Plenum Publishing Corporation.)

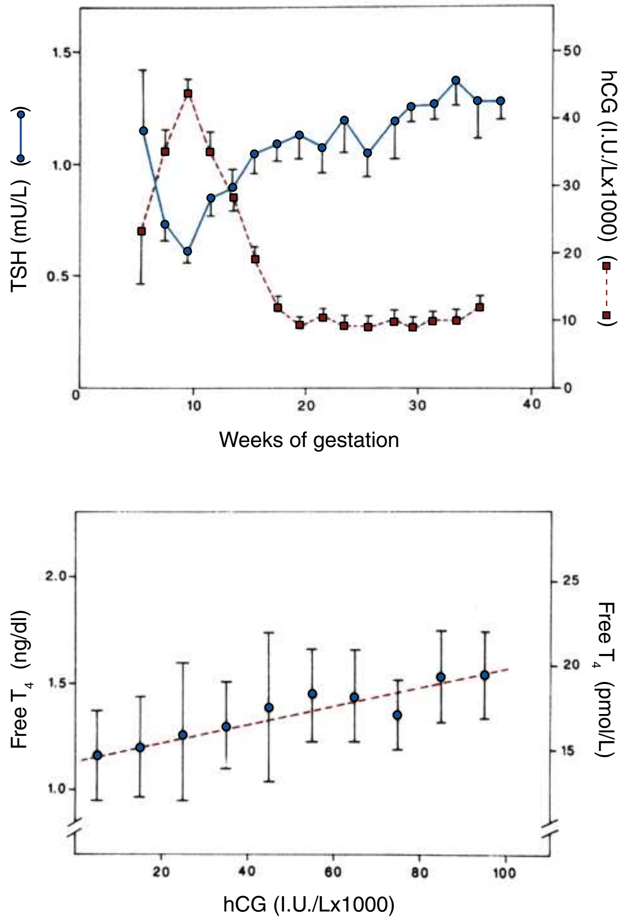


Figure 131.8 **Upper graph, serum thyroid-stimulating hormone (TSH) and human chorionic gonadotropin (hCG) as a function of gestational age in 606 healthy pregnant women.** Between 8 and 14 weeks of gestation, the changes in hCG and TSH levels are mirror images of each other and there is significant negative correlation between the individual TSH and hCG levels ($P < 0.001$). Each point gives the mean value (±SE) of individual determinations pooled for 2 weeks. Lower graph, scattergram of free thyroxine (T4) levels in relation to hCG concentrations in the first half of gestation. Each point represents the mean (±1 SD) free T4 values, determined between 6 and 20 weeks = plotted for 10,000 IU/L increments in hCG. The *dashed line* represents the linear regression curve ($P < 0.001$). (Data from Glinoer D, DeNayer P, Bourdoux P, et al. Regulation of maternal thyroid during pregnancy. *J Clin Endocrinol Metab*. 1990;71:276–287. Copyright © The Endocrine Society.)

total T4 measurements beginning at 4 to 6 weeks of gestation[60] (Fig. 131.7). The metabolic activity of thyroid hormone correlates with the free hormone concentration. Free T4 (fT4) levels have been reported to be unchanged, increased, or decreased during pregnancy, although they usually remain within the normal range[60,79] (Fig. 131.7). The differences between these reports may relate to relative iodine deficiencies in the populations studied,[55] to the stage of pregnancy at which the measurements were made,[60,79] or to the assay being used. fT4 levels are typically minimally elevated or in the high-normal range in the first trimester and then return to baseline and even below baseline in the latter half of pregnancy, in parallel with changes in hCG[60] (Fig. 131.8). Of note, some of the commercially available fT4 immunoassays are not accurate during pregnancy and should not be used in this population.[70] Some investigators recommend using the total T4 assay and then adjust by multiplying by 1.5 during the second and third trimesters.[70] If circulating fT4 levels are measured in pregnancy, it is recommended that assay-specific and trimester-specific pregnancy reference ranges be applied.[69]

Under physiological conditions (preserved thyroid function and adequate iodine availability), the increase in TBG concentrations is almost double the magnitude of the increase in maternal thyroxine. This discrepancy (i.e., a greater increase in TBG than in T4 levels) results in a progressively decreasing T4/TBG ratio over time, resulting in a physiological decrease in fT4 and fT3 concentrations of 10% to 15%.[80] T3 is also bound to TBG, but with somewhat less affinity than T4. The end result is the same pattern of increase in total T3 levels as seen with T4 during pregnancy.[80] TSH levels are decreased during the first trimester of pregnancy (Fig. 131.8), likely due to the increased fT4 levels as a result of hCG directly stimulating the TSH receptor.[60] In the first trimester, the normal range for TSH is 0.1 to 2.5 mU/L.[70] TSH levels have been shown to vary slightly depending on different

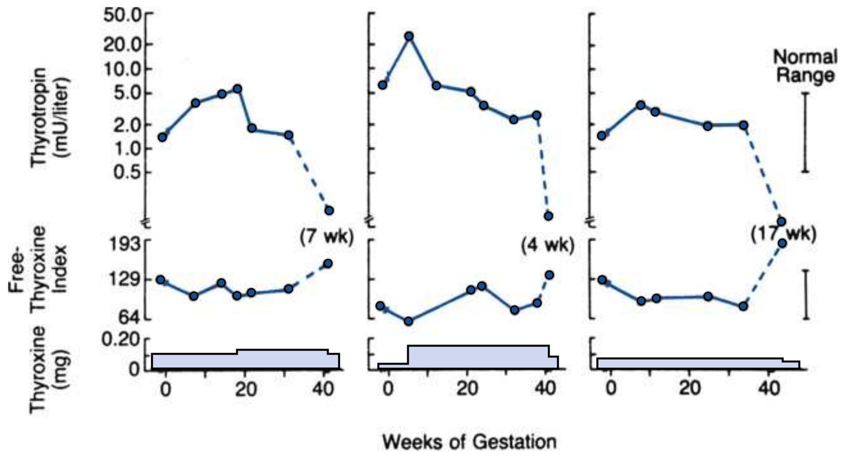


Figure 131.9 **Representative patterns of changes in thyroid function and thyroxine dose during pregnancy in women with hypothyroidism.** The left and middle panels show the results for two women who required an increase dose of thyroxine in pregnancy. The right panel shows the results for a woman whose thyroxine dose was not changed during pregnancy. The normal ranges for serum thyrotropin and free thyroxine index are indicated by the vertical bars on the right side of the figure. The discontinuity marks indicate the beginning and end of pregnancy. The *dashed line* represents the change between the last gestation and the first postpartum values (number of weeks postpartum is noted in parentheses). (Data from Mandel SJ, Larsen PR, Seely EW, Brent GA. Increased need for thyroxine during pregnancy in women with primary hypothyroidism. *N Engl J Med*. 1990;323:91–96; © 1990, Massachusetts Medical Society. All rights reserved.)

methods of analysis, although this variation does not appear to be clinically significant.[81] One approach to reducing the variability related to different assays and different laboratories is to use the multiple of medians (MoM) calculation, which divides an individual value by the population median.[82]

How often to measure thyroid hormones levels during pregnancy is controversial. In women with significant iodine deficiency, close monitoring may be indicated as fT4 levels may fall and TSH levels rise in the last trimester.[83] T4 turnover is increased during pregnancy.[79] In patients receiving thyroxine replacement, TSH levels should be checked every 1 to 2 months and the medication dose adjusted accordingly.[70] In one study, 75% of patients needed to increase their thyroxine dose during pregnancy[84] (Fig. 131.9), although studies have failed to demonstrate whether such medication adjustments actually translate into improved outcomes compared to increasing the dosing empirically as pregnancy progresses.[85] Checking TSH levels more frequently than every 1 to 2 months is generally unhelpful as it takes several weeks for changes in medication dosing to be reflected in realignment of the thyroid axis.

Hyperthyroidism

Diagnosis of Hyperthyroidism in Pregnancy. The clinical diagnosis of hyperthyroidism may be more difficult in pregnancy since many of the symptoms and signs are nonspecific. Tachycardia, warm skin, systolic flow murmurs, and heat intolerance are common to both hyperthyroidism and pregnancy. A goiter is a structural enlargement of the thyroid gland, but does not in of itself tell you about thyroid function. Features that are more suggestive of hyperthyroidism include weight loss, marked tachycardia[86] (particularly a heart rate greater than 120 beats/min), eye signs (proptosis, lid lag), and an audible bruit over the thyroid gland.[70,79] Infiltrative dermopathy and ophthalmopathy are specific for Graves disease. The changes in circulating thyroid hormone levels discussed earlier may cause further diagnostic difficulty. Total T4 and T3 levels are generally elevated, because of the presence of excess TBG. TSH concentrations are generally reduced during the first trimester of pregnancy because of the effect of hCG and negative feedback from elevated levels of fT4.[60] Despite these limitations, it is important to make the diagnosis of hyperthyroidism promptly and accurately to allow timely treatment, since untreated or poorly controlled hyperthyroidism has detrimental effects on fetal and neonatal outcome.[69,87]

Graves Disease. Graves disease is an autoimmune disease caused by circulating anti-TSH-receptor antibodies (TRAb) that activate the receptor, thereby stimulating thyroid hormone synthesis and secretion as well as thyroid growth, resulting in goiter. Other symptoms may include eye disease (orbitopathy) and occasionally a dermopathy referred to as pretibial myxedema. The presence of TRAb in serum and orbitopathy on clinical examination distinguishes this disorder from other causes of hyperthyroidism.

Most cases of hyperthyroidism in pregnancy are due to Graves disease.[70,79] Graves hyperthyroidism affects 0.2% of all pregnancies.[79] Fetal/neonatal hyperthyroidism occurs in 1% to 5% of women with active or a history of Graves disease and is associated with adverse fetal and neonatal outcomes if untreated.[88] Maternal serum TRAb can be measured in the first and third trimesters and may be helpful in predicting which infants are at high risk of developing fetal/neonatal Graves disease.[89] TRAb are immunoglobulin G (IgG) antibodies that can cross the placenta and cause fetal hyperthyroidism, although this is rare. Features of intrauterine hyperthyroidism include fetal tachycardia, fetal goiter, poor growth, and craniosynostosis.[90] Thioamide medications used to treat the mother may also cross the placenta and cause fetal hyperthyroidism.

All fetuses of women with Graves disease should be closely monitored in pregnancy for signs of fetal hyperthyroidism and by ultrasound to evaluate growth and fetal heart rate.[91] In women with active Graves disease in pregnancy and serum TRAb levels greater than 2- to 3-fold the upper limit of normal, fetal thyroid ultrasound should be considered in the second and third trimesters to evaluate for fetal goiter.[89] In equivocal cases, fetal blood sampling may be required to confirm the diagnosis.

Hyperthyroidism and Hyperemesis Gravidarum. Pregnant patients with hyperthyroidism may present with nausea and vomiting. Hyperemesis gravidarum may also be associated with biochemical evidence of hyperthyroidism in up to two-thirds of cases.[70,79] Only

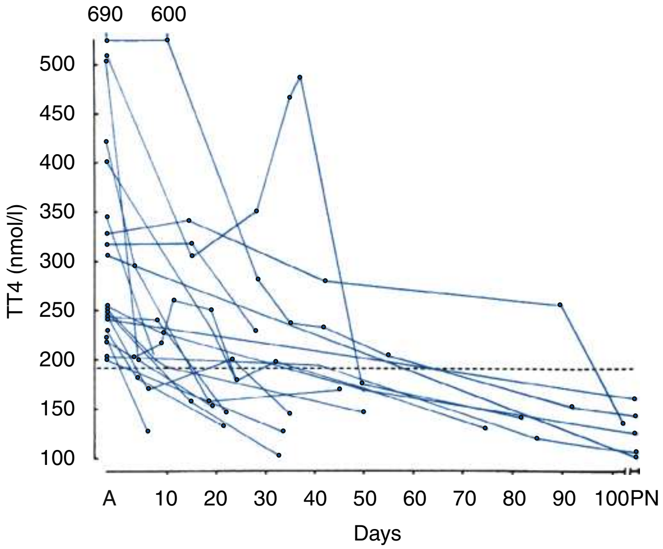


Figure 131.10 **Changes in plasma total thyroxine concentration in 20 subjects with hyperemesis on admission (A) and during the course of their pregnancy.** The interrupted line represents the mean (±SD) value seen in normal pregnancy. (Data from Swaminathan R, Chin RK, Lao TT et al. Thyroid function in hyperemesis gravidarum. *Acta Endocrinol* [Copenh]. 1989;120:155–160.)

10% to 20% of these patients will manifest clinically with thyrotoxicosis. The severity of hyperemesis appears to correlate directly with the degree of thyroid dysfunction.[92] In most patients, fT4 levels decrease to normal or near normal levels within 35 days (range, 6 to 133 days) of pregnancy, generally in parallel with the decrease in circulating hCG levels[93] (Fig. 131.10). Most patients with hyperthyroidism-associated hyperemesis have no goiter and no eye findings; as such, the clinical diagnosis can be challenging.[70]

Treatment. Women with moderate-to-severely symptomatic hyperthyroidism due to Graves disease, toxic adenoma, toxic multinodular goiter, or gestational trophoblastic disease require treatment. Such patients will almost always have TSH values below 0.05 mU/L and trimester-specific total T4/T3 and fT4 concentrations that exceed 1.5 times the upper limit of normal for nonpregnant patients. Treatment should be focused on controlling symptoms as well as decreasing thyroid hormone synthesis and release. Symptoms of hyperthyroidism result primarily from increased beta-adrenergic tone and include such clinical features as palpitations, tachycardia, tremulousness, and anxiety. Treatment with beta-adrenergic blocking agents (beta-blockers) such as metoprolol or propranolol can be highly effective in managing such symptoms. Although generally regarded as safe in pregnancy, beta-blockers have been weakly associated with fetal growth restriction and hypoglycemia; as such, they can be weaned once the hyperthyroidism is well controlled with thionamides.[94]

Thioamides are the treatment of choice for hyperthyroidism in pregnancy. They act at the level of the thyroid gland to inhibit the synthesis, storage, and release of thyroid hormone. Specifically, they inhibit the organification of iodine to tyrosine residues within thyroglobulin and the coupling of iodotyrosines. The two most commonly used thioamides are methimazole and PTU. Both cross the placenta and appear to have similar effects on the fetal thyroid.[95] In general, methimazole is preferred to PTU, except during the first trimester. Both drugs are weakly associated with an increased risk of birth defects of 2% to 3%.[96] The embryopathy associated with methimazole appears to be more severe (including aplasia cutis and dysmorphic faces) and therefore PTU is preferred for treatment in the first trimester.[69] However, PTU has been associated with hepatotoxicity and so its use is often limited to the first trimester.[97] A thyroidectomy can be performed if the hyperthyroidism is refractory to medical management. If necessary, this is best done in the second trimester.

Thyroid Storm. Thyroid storm is a rare, life-threatening condition characterized by severe clinical manifestations of thryotoxicosis.[98] It is seen most commonly in patients with known hyperthyroidism and may be precipitated by an acute event such as trauma, infection, or parturition. The maternal and perinatal mortality rate can be as high as 10% to 30%. Treatment involves both supportive care (often in an intensive care unit [ICU] setting) as well as the administration of beta-blockers, a thioamide, and iodine.[99] Glucocorticoids are also useful as they reduce peripheral conversion of the prohormone (T4) to the active hormone (T3), promote vasomotor stability, and may improve the autoimmune process of Graves disease.[100]

Hypothyroidism

Diagnosis of Hypothyroidism in Pregnancy. Subclinical hypothyroidism (defined as biochemical evidence of hypothyroidism without clinical features) has been reported in 1.7% of pregnant women, whereas overt hypothyroidism complicates 2 to 10 per 1000 pregnancies.[101] Pregnancy may confound the typical clinical manifestations of hypothyroidism such as fatigue, cold intolerance, constipation, and weight gain.[69] Other symptoms such as muscle cramps or dry skin may be more suggestive of the diagnosis.[70,102] Because of the changes in thyroid physiology during pregnancy and the substantial population differences in reference ranges, thyroid function tests should be interpreted using population-based, trimester-specific reference ranges for TSH and using assay- and trimester-specific reference ranges for serum fT4. If internal or transferable trimester-specific reference ranges are not available, then an upper limit for TSH of 4.0 mU/L may be used.[69] Overt primary hypothyroidism is defined as an elevated trimester-specific TSH concentration with a decreased fT4 concentration (using reference range for pregnant women). In patients with hypothyroidism, total T4 levels in pregnancy should be interpreted with caution since the increase in circulating TBG concentrations may cause the total T4 levels to be in the normal range.[102] Established replacement doses of thyroxine may need to be increased, because of the increased clearance of thyroxine during pregnancy.[84]

Indications for Treatment and Management. Overt maternal hypothyroidism is associated with an increased risk of adverse pregnancy outcomes[103] as well as detrimental effects on fetal neurocognitive development.[104] As such, early identification and treatment is important. The data regarding subclinical hypothyroidism are more controversial. Subclinical hypothyroidism has been inconsistently associated with adverse pregnancy outcomes including preterm delivery[105] and infant neurodevelopment.[104] There is also a weak and inconsistent association with miscarriage, although the presence of antithyroid peroxidase antibodies (TPOAb) in this population appears to strengthen this association.[106,107] In a large prospective study of more than 4600 subjects, Negro et al.[106] reported that subclinical hypothyroidism increased the risk of pregnancy complications in TPOAb-positive women, and that treatment of TPOAb-positive women with a TSH >2.5 mU/L resulted in a significant reduction in pregnancy complications. A separate randomized controlled trial (RCT) reported a decrease in preterm delivery and pregnancy loss rates in euthyroid (TSH <4.2 mU/L). TPOAb-positive women were treated with thyroid replacement beginning in the first trimester of pregnancy.108 However, other studies have failed to show an association between subclinical hypothyroidism and adverse pregnancy outcomes.[62] Because of the risk of adverse pregnancy outcomes with subclinical hypothyroidism, some, but not all, national consensus bodies have advocated for routine first trimester screening and treatment.[70,109]

Hypothyroidism and Infertility. There has been considerable debate regarding the association of subclinical hypothyroidism and TPOAb status and pregnancy loss in the infertile population. While one systematic review suggested a strong association between TPOAb positivity and miscarriage,[107] two recent large, well-designed RCTs failed to confirm this association.[110,111]

Fetal Thyroid Gland. The thyroid gland is formed from all three germ cell layers. The majority of critical events related to thyroid morphogenesis happen within the first 60 days postconception. Consequently, most developmental thyroid abnormalities result from morphogenetic errors during this time frame. Ectopic thyroid tissue can result from abnormal thyroid migration due to abnormal morphogenesis of the heart or from anomalous interactions between the thyroid primordium and the heart.[112] Thyroglobulin synthesis can be detected in the fetal thyroid by 4 to 6 weeks of gestation, TRH synthesis by 6 to 8 weeks, iodine trapping by 8 to 10 weeks, and TSH release with T4 synthesis by 12 weeks.[113]

KEY POINTS

- Subclinical hypothyroidism has been reported in 1.7% of pregnant women while overt hypothyroidism complicates 2 to 10 per 1000 pregnancies.
- Poorly controlled maternal hypothyroidism is associated with an increased risk of adverse pregnancy outcomes as well as detrimental effects on fetal neurocognitive development.
- The association between subclinical hypothyroidism and adverse pregnancy events is more variable; as such, universal screening in pregnancy is controversial.

ADRENAL GLAND

Regulation of Glucocorticosteroid Synthesis

Changes in the Adrenal Glands. The demand for steroid production increases during pregnancy. Maternal estrogen and cortisol levels are elevated in pregnancy and the fetus requires steroids for normal growth and development.[114] Although the maternal adrenal glands do not increase in size or weight during pregnancy, the size of the zona fasciculata—that region of the adrenal glands responsible for glucocorticoid synthesis—does increase.[114]

Corticotropin-Releasing Hormone Physiology During Pregnancy. The high levels of free cortisol in the maternal circulation during pregnancy are the result of elevated levels of CRH. As pregnancy progresses, CRH levels rise exponentially to concentrations that are 1000 times higher than in nonpregnant women and peak during labor.[115] The source of this CRH is 2-fold. First, the maternal hypothalamus releases excess CRH in response to the physiological and psychological stresses of pregnancy.[116] Second, CRH is produced by the placenta and released into the maternal (and fetal) compartments in increasing amounts as placental size increases over time.[117] The effects of elevated CRH levels are attenuated in part by an increase in CRH-binding protein in the plasma and by negative feedback and desensitization of the anterior pituitary to high levels of CRH.[118] Despite these attenuations, the 1000-fold rise in CRH leads to a 1.5-fold rise in plasma free cortisol.[115]

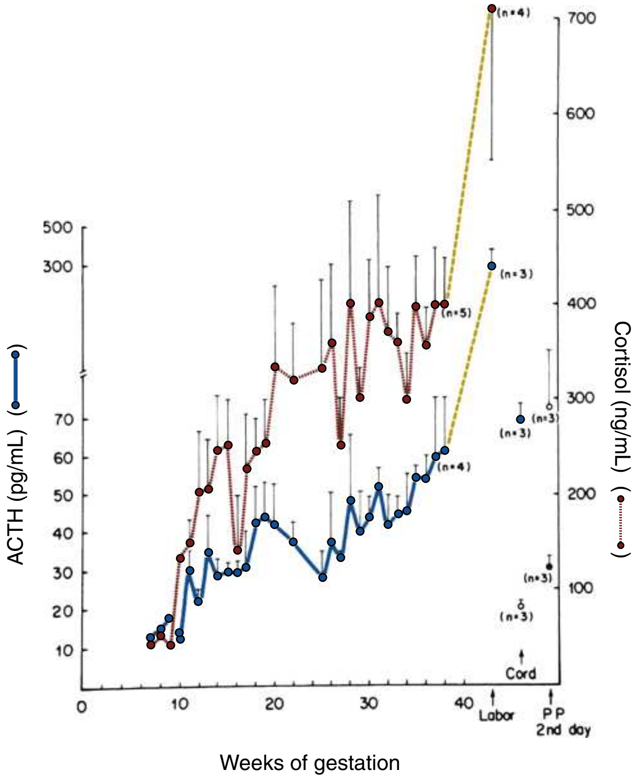


Figure 131.11 Plasma concentration of adrenocorticotropic hormone (ACTH) and cortisol during normal pregnancy. Blood samples were obtained weekly at 8 AM to 9 AM from five normal pregnant women and from three women during labor and on the second postpartum day. In addition, umbilical cord plasma was obtained from the newborn infants of three of these subjects. The mean plasma concentrations for ACTH are denoted by *blue circles*, whereas plasma cortisol levels are denoted by *red circles*. The vertical bars correspond to the magnitude of the standard error of the mean. (Data from Carr BR, Parker CR Jr, Madden JD, et al. Maternal plasma adrenocorticotropin and cortisol relationships throughout human pregnancy. *Am J Obstet Gynecol*. 1981;139:416–422.)

CRH has also been implicated in myometrial contractility and the timing of labor. At term, while circulating CRH levels continue to rise, CRH-binding protein levels decrease, leading to an increase in bioavailable (free) CRH. This switch occurs at an earlier time point in women destined to deliver preterm and later in women who deliver postterm.[119]

Adrenocorticotropic Hormone Physiology During Pregnancy. ACTH, a 39-amino acid polypeptide, is secreted by the anterior pituitary in response to CRH from the hypothalamus. ACTH is formed from cleavage of a precursor peptide, proopiomelanocortin (POMC), along with several other related peptides, β-endorphin and α-melanocyte-stimulating hormone (MSH).[120] During pregnancy, there is increased production of ACTH and β-endorphin from both the anterior pituitary and the placenta.[114,121] Plasma ACTH levels rise 3-fold over the course of pregnancy and peak during labor and delivery.[122] The rise in plasma ACTH occurs despite rising levels of plasma cortisol, suggesting that ACTH secretion in pregnancy is not subject to the normal negative feedback mechanisms.[121,122] In the third trimester, there is no apparent correlation between ACTH and CRH, but ACTH and cortisol levels have a strong positive correlation and maintain a diurnal pattern with ACTH pulses leading cortisol pulses by about 30 minutes.[123] In pregnancy, the ACTH/cortisol ratio is lower than in nonpregnant women. Interestingly, in women with preeclampsia, the ACTH/cortisol ratio is elevated and mirrors that in nonpregnant women.[124]

Cortisol Physiology During Pregnancy. High levels of estrogen during pregnancy stimulate hepatic production of CBG.[114] Circulating levels of CBG are 2-fold higher during pregnancy than in the nonpregnant state, which results in increased total plasma cortisol.[114,125] By the end of pregnancy, total plasma cortisol levels are nearly 3-fold higher than nonpregnant levels[125] (Fig. 131.11). However, only free cortisol, that fraction not bound to CBG, is metabolically active. In the third trimester, plasma free cortisol levels are elevated 1.5- to 4-times that in nonpregnant women.[125] Twenty-four–hour urinary free cortisol levels increase throughout pregnancy and are 3-fold higher in the third trimester than in the nonpregnant state.[125] The diurnal cortisol pattern observed outside of pregnancy is maintained during pregnancy,[126] although pregnancy is associated with an upward shift in the curve with the nadir of cortisol secretion higher than the nonpregnant peak of cortisol secretion.[126] Clinically, the physiological hypercortisolism of pregnancy presents as maternal fatigue, weight gain, hyperglycemia, edema, and emotional lability.[114] Because so many of these features can be considered normal during pregnancy, it can be challenging to detect symptoms of pathologic hypercortisolism.[114]

Cushing Syndrome. Cushing syndrome is rare in pregnancy with only approximately 214 cases in the published literature. Of these, 28% were due to pituitary adenomas, 44% to adrenal adenomas, and 10% to adrenal carcinoma.[127] Several cases of transient pregnancy-induced Cushing syndrome have been reported due to overexpression of LH/hCG receptors in the adrenal glands.[128]

Diagnosing Cushing syndrome during pregnancy can be challenging because many of the symptoms—fatigue, weight gain, peripheral edema—are common in pregnancy.[129] Additionally, laboratory tests traditionally used to confirm the diagnosis are not valid in pregnancy.[129] Moreover, some of the screening tests such as the low-dose dexamethasone suppression test and 24-hour urinary free cortisol level can be difficult to interpret during pregnancy, although monitoring of the diurnal free plasma cortisol pattern may have more utility.[129–131] To illustrate this point, the 1-mg overnight dexamethasone suppression test causes a mean reduction in serum cortisol of 87.1% in nonpregnant patients.[130] However, in pregnancy, the mean cortisol reduction is 82.9% during the first trimester, 43.5% during the second trimester, and 37.3% during the third trimester.[130] These data suggest that, while the overnight dexamethasone suppression test may have some utility in the evaluation of pregnant patients suspected of having Cushing syndrome, the normal values implemented outside of pregnancy must be adjusted.[130] Similarly, in nonpregnant patients, 24-hour urinary free cortisol levels greater than 4-fold the upper limit of normal are suggestive of Cushing syndrome.[129] In pregnancy, however, 24-hour urinary free cortisol levels increase to 3-fold normal and patients with Cushing syndrome in pregnancy have a mean 24-hour urinary free cortisol level that is typically 8-fold normal (range 2- to 22-fold).[132] Because of the overlap in 24-hour urinary free cortisol levels among pregnant patients with and without Cushing syndrome, especially in early pregnancy, the recommendation is to use a 24-hour urinary free cortisol level threshold in the second and third trimesters of greater than 3-fold the upper limit of normal to diagnose Cushing syndrome.[129] The nightly nadir in plasma cortisol is maintained during pregnancy, although at a higher value, but is lost in patients with Cushing syndrome.[133] Elevated evening plasma cortisol levels and loss of the diurnal cortisol pattern may be helpful in suggesting the diagnosis during pregnancy.[131] However, no reliable thresholds for evening cortisol levels have been developed for the diagnosis. Although there is no universally agreed-upon diagnostic algorithm for Cushing syndrome in pregnancy, it has been suggested that women who screen positive for Cushing syndrome using 24-hour urinary free cortisol or midnight cortisol level and have a borderline plasma ACTH proceed with diagnostic testing to differentiate Cushing syndrome from Cushing disease (pituitary adenoma), whereas those with a low ACTH level undergo imaging of the adrenal glands.[129]

Diagnostic tests include the 8-mg dexamethasone suppression test, CRH stimulation test, and inferior petrosal sinus sampling (IPSS). In nonpregnant patients with elevated cortisol levels, a low ACTH level is suggestive of an adrenal etiology while patients with ACTH-secreting tumors have normal or elevated levels of ACTH.[129] In pregnancy, however, adrenal causes of Cushing syndrome are not regularly associated with a decrease in ACTH levels likely due to placental CRH that is not suppressed by hypercortisolism.[129] The 8-mg dexamethasone suppression test may help differentiate an adrenal etiology of Cushing syndrome from Cushing disease in pregnancy. In a study by Lindsay et al., pregnant patients were administered an 8-mg dexamethasone suppression test. All patients with an adrenal cause of Cushing syndrome showed at least an 80% decrease in plasma cortisol, whereas 3/7 patients with Cushing disease did not suppress plasma cortisol.[132] Outside of pregnancy, ACTH-secreting pituitary tumors respond to CRH stimulation while adrenal adenomas and ectopic ACTH-secreting tumors do not.[134] Several studies have demonstrated a rise in plasma cortisol in pregnant patients with Cushing disease during a CRH stimulation test without any adverse effects.[129,131] There is limited research on IPSS in pregnancy likely due to the risk of ionizing radiation. However, the high diagnostic accuracy of IPSS may prove useful considering the challenges with interpreting laboratory values of ACTH and cortisol during pregnancy.[129] With regard to imaging, MRI has been used successfully in pregnancy to localize adenomas in the adrenal glands and pituitary without adverse consequences.[131] Ideally, gadolinium should be used to obtain optimal images. Failure to use this contrast medium will decrease the sensitivity of MRI for detecting pituitary adenomas from 52% to 38%.[135]

Despite the diagnostic challenges, it is imperative to diagnose Cushing syndrome during pregnancy because of the severe maternal and fetal morbidity associated with this condition. Maternal complications include hypertension, preeclampsia, diabetes, wound infection, fractures, cardiac failure, and psychiatric disorders.[132] Fetal complications include spontaneous abortion, preterm birth, intrauterine growth restriction, and stillbirth. In one case series, the live birth rate was 76% (59/78) when no treatment was administered and increased to 89% (50/56) for women who received treatment at an average gestational age of 20 weeks.[132] For women with an adrenal etiology, surgery in the form of unilateral or bilateral adrenalectomy is more successful than medical therapy with an overall fetal loss rate of 12.5% with surgery compared with 20.8% with medical treatment.[127] Patients who decline surgery can be treated with metyrapone, which should be titrated to achieve a 24-hour urinary cortisol level within the normal pregnancy range.[136] Although metyrapone is effective for treating hypercortisolism, it may exacerbate hypertension.[137] Ketoconazole has also demonstrated success in lowering cortisol levels during pregnancy without detrimental effects to the fetus.[138] However, ketoconazole has known teratogenicity in animal models and should be reserved for patients who require emergency treatment and are intolerant of metyrapone.[129] Women with a pituitary adenoma (Cushing disease) can be successfully treated with transsphenoidal surgery and/or medical management.[127]

KEY POINTS

- Screening and diagnostic tests for Cushing syndrome are confounded by the normal physiological changes in ACTH, CRH, and cortisol levels and their respective regulatory pathways.
- Elevated evening plasma cortisol levels and loss of the diurnal cortisol pattern may be helpful in diagnosing Cushing syndrome in pregnancy.

Adrenal Insufficiency. AI (Addison disease) is rare in pregnancy with only about 100 cases reported in the literature.[129] The most common cause in developed countries is autoimmune adrenalitis, whereas destruction of the adrenal glands bilaterally by tuberculosis is more common worldwide.[139] At least seven pregnancies associated with autoimmune polyglandular syndrome type 2 (APS 2) have also been reported.[129] Administration of exogenous corticosteroids is the most common cause of secondary AI, and there have been several reports of maternal adrenal suppression after administration of betamethasone to women at risk of preterm birth.[129,140] The diagnosis of I during pregnancy may be challenging because many of the symptoms—fatigue, vomiting, and depression—are attributed to the pregnancy itself.[139] It is common in pregnancy for serum sodium levels to decrease by up to 5 mEq, but any further hyponatremia should prompt an evaluation for AI.[141] Normal potassium levels should not preclude a workup for AI.[141,142] Pregnant patients may also present with persistent vomiting that may be attributed to hyperemesis gravidarum rather than AI.[141] There should be a low threshold to evaluate patients for AI who have a personal or family history of autoimmune disease and signs or symptoms consistent with the diagnosis.[142]

One of the laboratory tests for diagnosing AI is a plasma cortisol level drawn between 8:00 and 9:00 AM. In the nonstressed general and pregnant populations, an early morning plasma cortisol level <3 μg/dL is indicative of AI in the appropriate clinical context and a value >19 μg/dL effectively excludes the diagnosis.[140] Patients with morning cortisol levels between 3 and 19 μg/dL should undergo low-dose corticotropin stimulation testing.[140] The low-dose corticotropin test has successfully diagnosed AI in pregnant patients using a cortisol threshold of >30 μg/dL.[140] A rise in ACTH to 200 to 2000 pg/mL is also suggestive of AI.[143] The CRH stimulation test, metyrapone stimulation test, and insulin tolerance test are not recommended for use in pregnancy.[129]

Patients with AI should be treated throughout pregnancy. Labor and delivery are stressful events with increased steroid requirements. During pregnancy, patients should receive glucocorticoid supplementation with cortisone acetate 25 to 37.5 mg/day or hydrocortisone 20 to 30 mg/day in two or three daily doses as well as mineralocorticoid replacement with fludrocortisone 0.05 to 0.1 mg/day.[143] For those with severe nausea and vomiting, hydrocortisone 50 mg IM or dexamethasone 1 mg IM should be administered.[143] During labor, "stress-dose steroids" in the form of hydrocortisone 50 mg IV should be administered. Patients who have a vaginal delivery should receive an additional dose of oral cortisone acetate 25 to 50 mg.[143] In the event of cesarean delivery, patients should receive up to 100 mg of hydrocortisone every 6 hours for at least 24 hours with the dose decreased by half on the second day postoperatively followed by a taper to an oral regimen within 3 days.[143]

Changes to the Renin-Angiotensin-Aldosterone System in Pregnancy. The renin-angiotensin system (RAS) adapts to changes in the maternal cardiovascular system during pregnancy. As early as 6 weeks' gestation, systemic vascular resistance decreases and heart rate increases.[144] Pregnancy is associated with a 47% increase in blood volume but a decrease in blood pressure, despite a 20% increase in CO.[144,145] GFR increases 40% to 60% during pregnancy and corresponds to an increase of 20,000 to 30,000 mEq/day of sodium filtration.[146] Overall, total body sodium increases 3 to 4 mEq/day leading to a net gain of 900 to 1000 mEq, and total body potassium increases by 300 to 320 mEq by the end of pregnancy.[146] Progesterone levels increase throughout pregnancy and, by term, circulating concentrations range from 100 to 200 ng/mL. Much of this comes from the placenta, which produces approximately 250 mg of progesterone daily.[147] The effects of progesterone are multiple, including smooth muscle relaxation, which accounts for the generalized decrease in vascular tone, and decreased sodium reabsorption by competitively inhibiting aldosterone receptors in the renal tubule.[148,149] Estrogen levels also increase during pregnancy, which stimulates hepatic synthesis of angiotensinogen and offsets the natriuretic effects of progesterone and elevated GFR.[150] In response to these physiological changes associated with pregnancy, the RAS must adjust to maintain sodium and volume homeostasis.

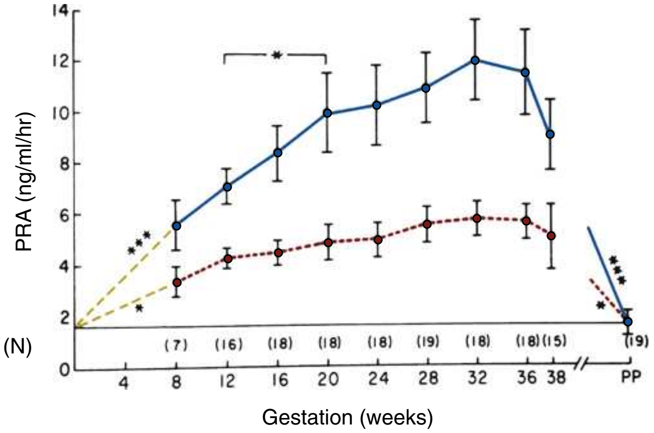


Figure 131.12 Sequential changes in plasma renin activity (PRA) (*blue line*) and in PRA normalized to the postpartum substrate values (*red line*) (mean ± SE) throughout pregnancy. $^{*}P < 0.05$; $^{***}P < 0.001$. (Data from Wilson M, Morganti AA, Zervondakis I et al. Blood pressure, the renin-aldosterone system and sex steroids throughout normal pregnancy. *Am J Med.* 1980;68:97–104.)

Renin is produced predominantly in the kidneys under the influence of estrogen.[129] Plasma renin activity (PRA) increases 7-fold during pregnancy and almost half of this increase is attributed to an increase in renin substrate.[151] By the end of pregnancy, mean plasma renin substrate increases 4-fold[151] (Fig. 131.12). hCG stimulates prorenin secretion by the placenta and ovary.[152] Moreover, studies have shown that, in response to an IV saline infusion and change to an upright posture, pregnant subjects exhibited a change in circulating renin concentrations consistent with that in nonpregnant subjects, but with significantly less sodium excretion. This suggests that estrogen-mediated stimulation of renin synthesis is not sufficient to account for the increased PRA observed in pregnancy, and that a physiological need for increased sodium reabsorption in pregnancy likely also has a role to play.[153] The uteroplacental unit expresses components of the RAS and disruption of the uteroplacental RAS system may contribute to reduced placental blood flow in pregnancies complicated by preeclampsia and intrauterine growth restriction.[154]

Estrogen stimulates hepatic production of angiotensinogen.[150] As estrogen levels rise in pregnancy, so do circulating levels of angiotensinogen. The correlation coefficient of angiotensinogen and estradiol-17β is 0.60 and that for estriol is 0.68.[155] Increased angiotensinogen leads to higher circulating levels of angiotensin II and diminished negative feedback of angiotensin II production.[150] Interestingly, oxidized angiotensinogen has a higher affinity for renin than reduced angiotensinogen.[156] In patients with preeclampsia, the oxidized:reduced angiotensinogen ratio increases above the baseline of 6:4, which may explain in part the pathogenesis of hypertension in this population.[156] Monomeric angiotensinogen (molecular weight [MW] 61.5 kDa) is the predominant form in nonpregnant women. In pregnancy, angiotensinogen can form high MW polymers.[150] Such high MW polymers comprise 4% of circulating angiotensinogen in reproductive age women, 16% in pregnant women, and up to 28% in pregnant women with hypertension.[157] The interaction between renin and

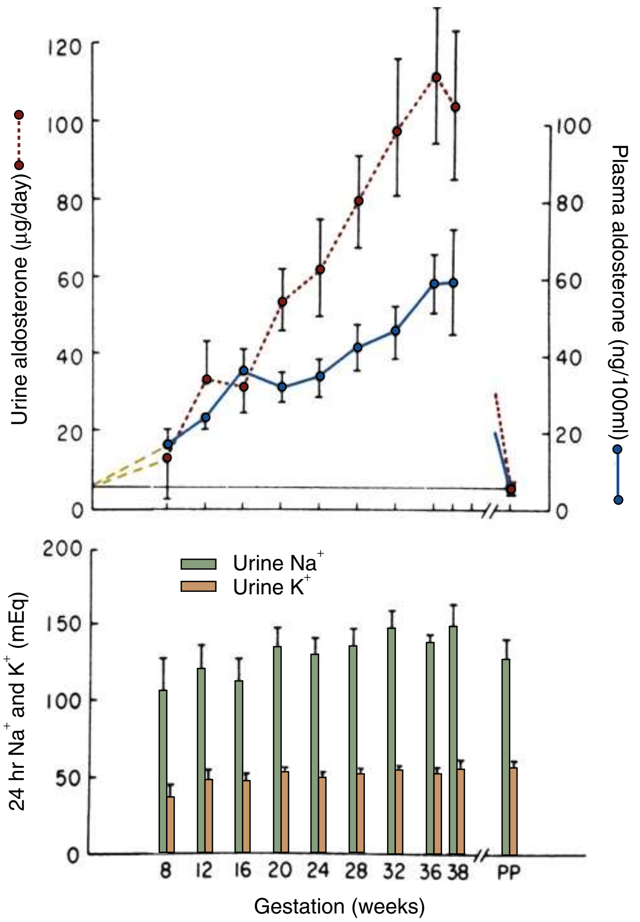


Figure 131.13 Sequential changes throughout pregnancy in plasma aldosterone, urine aldosterone, urinary sodium, and potassium (mean ± SE). (Data from Wilson M, Morganti AA, Zervondakis I et al. Blood pressure, the renin-aldosterone system and sex steroids throughout normal pregnancy. *Am J Med*. 1980;68:97–104.)

angiotensinogen to produce angiotensin I is reduced when angiotensinogen is in its multimeric form.[158] Therefore although circulating angiotensinogen levels are increased in pregnancy, the rate at which angiotensinogen reacts with renin depends both on the redox state of angiotensinogen and on its polymerization with other proteins.

Angiotensin I is converted to angiotensin II by angiotensin-converting enzymes (ACE). ACE activity remains stable throughout pregnancy at approximately 30.6 nmol/min/mL.[159] The isoform, ACE2, converts angiotensin II into angiotensin-(1-7).[160] Levels of ACE2 in plasma are difficult to measure, because it is located primarily within endothelial cells and its catalytic activity in plasma is inhibited.[150] ACE2 is also present in syncytiotrophoblasts and may regulate the release of angiotensin-(1-7) from the placenta into the maternal circulation, thereby contributing to maternal vasodilation.[160] In the third trimester, plasma angiotensin-(1-7) is increased by 34% and plasma angiotensin II by 50% compared to nonpregnant women.[161] In preeclampsia, plasma angiotensin-(1-7) and angiotensin II levels are reduced, but angiotensin II levels are higher than in nonpregnant women and 50% higher than angiotensin-(1-7), suggesting a shift in the balance of hormones that regulate vasodilation in such women.[161]

During pregnancy, elevated progesterone levels cause plasma aldosterone concentrations to increase 5- to 7-fold in the first trimester and up to 10- to 20-fold in the third trimester[162] (Fig. 131.13). Aldosterone is not tightly bound to proteins such that the increase in plasma aldosterone corresponds to an increase in bioactive steroid.[162] Pregnant women demonstrate an amplified response to renin compared with nonpregnant women and have a significant increase in plasma aldosterone concentration:plasma renin concentration (PAC:PRC) ratio.[163] Of note, women with gestational hypertension have lower renin and aldosterone levels than normotensive women, but a 2-fold increase in the PAC:PRC ratio.[163] The amount of aldosterone synthesized is inversely proportional to dietary sodium intake in diets containing 25 to 225 mEq sodium per day, whereas diets containing <10 mEq sodium per day are associated with a further increase in aldosterone secretion.[164] It is estimated that 0.6% to 0.8% of pregnant women have primary hyperaldosteronism, but it is a challenging diagnosis to make due to changes in the RAS during pregnancy and the fact that up to 10% of women develop pregnancy-induced hypertension, which alters RAS signaling.[114] Low renin levels in conjunction with elevated aldosterone, hypokalemia, hypertension, and/or an adrenal mass on imaging suggest primary hyperaldosteronism.[114] Progesterone is a competitive antagonist of aldosterone in the distal convoluted tubule of the kidney. As such, primary hyperaldosteronism can be masked in pregnancy and only become clinically apparent in the postpartum period.[114] Options for the treatment of primary hyperaldosteronism in pregnancy include potassium supplementation, antihypertensive medication, and adrenalectomy if a localized adenoma is identified.[165] Calcium channel blockers are the first choice for antihypertensive therapy. Spironolactone and ACE inhibitors are best avoided due to their potential teratogenic effects on the fetus.[165]

KEY POINTS

- Plasma renin activity increases 7-fold during pregnancy.
- Elevated levels of progesterone cause aldosterone production to increase 5- to 7-fold in the first trimester and 10- to 20-fold in the third trimester.

Adrenal Medulla

Catecholamine Physiology During Pregnancy and Pheochromocytoma. Pheochromocytoma is a rare but potentially fatal neuroendocrine tumor of chromaffin cells of the adrenal medulla with an incidence of 1 in 50,000 pregnancies.[166–168] Chromaffin cells originate from neural crest tissue and are capable of synthesizing, storing, and secreting catecholamines. Neuroendocrine chromaffin cell tumors may also be found in sympathetic ganglia, where they are known as paragangliomas. Approximately 40% of these tumors occur in young women and are associated with pathogenic germline genetic mutations, including multiple endocrine neoplasia type 2 (MEN 2), von Hippel-Lindau syndrome, neurofibromatosis, or succinate dehydrogenase subunit gene mutations.[169] Catecholamines including norepinephrine, epinephrine, and dopamine are produced from tyrosine precursors. They function as neurotransmitters in the sympathetic nervous system and are potent systemic cardiac and vasoactive compounds. Catecholamines play a crucial role in the maternal adaption to pregnancy and especially labor and delivery. Whether circulating catecholamine levels are increased in pregnancy is unclear,[170] even in the setting of preeclampsia.[171] What is clear is that urinary catecholamine levels increase during labor and remain elevated for the first 2 days postpartum.[172]

Only a small fraction of maternal catecholamines cross the placental barrier, with levels in umbilical cord blood less than 10% of that in the maternal circulation.[173] The placenta therefore plays a key role in protecting the fetus from high levels of circulating catecholamines during pregnancy, including in the setting of a pheochromocytoma. It is hypothesized that the high concentrations in the maternal circulation in labor result in high catecholamine levels in the fetus at the time of birth, which may assist the fetus in adapting to early postnatal life.[174]

Maternal mortality from pheochromocytoma has decreased dramatically over the past four decades, due in large part to improve diagnosis. If the diagnosis is made in the antepartum period, maternal mortality rate is 4% compared with 25% in cases diagnosed postpartum.[175] A recent systematic review including 135 cases reported maternal and fetal mortality rates of 8% and 17%, respectively.[176] As in nonpregnant patients, symptoms may be vague or more classic with intermittent paroxysms of hypertension, flushing, tremors, and anxiety due to episodic release of catecholamines (see Chapter 99). The first suspicion of pheochromocytoma may occur with severe hypertension or shock during induction of anesthesia, labor, or surgery. Failure to consider this diagnosis can have serious consequences.[167,176,177] Diagnosis in pregnancy can be particularly challenging since it can be confused with gestational hypertension or preeclampsia. At the onset of hypertension prior to 20 weeks, the absence of proteinuria, edema, or hyperuricemia as well as evidence of episodic signs/symptoms is suggestive of pheochromocytoma.[177] Once the diagnosis is suspected clinically, appropriate biochemical testing should be performed.

Measurement of 24-hour urinary fractionated metanephrines and catecholamines or plasma metanephrines can be used to confirm the diagnosis as for nonpregnant patients[166–168] (see Chapter 99). If there is a low index of suspicion, then urinary fractionated catecholamines and metanephrines would be appropriate. If the index of suspicion is high, plasma fractionated metanephrines should be measured. To decrease the false-positive rate, the plasma should be drawn supine using an indwelling cannula placed at least 30 minutes prior.[178] Elevated urinary or plasma metanephrines or normetanephrines have a 98% to 99% sensitivity but a lower specificity.[167] Antihypertensive medications such as methyldopa and labetalol can interfere with the measurement if a liquid chromatography (LC)-electrochemical detection assay is used, but not if the measurement is performed using LC-MS/MS (liquid chromatography tandem mass spectrometry) technology. Other medications, such as tricyclic antidepressants, can interfere with neuronal norepinephrine uptake resulting in a false-positive result.[177]

Once the diagnosis has been confirmed biochemically, every effort should be made to localize the tumor. Both computed tomography (CT) and MRI are suitable, but MRI is generally preferred in pregnancy because it avoids exposing the fetus to ionizing radiation.[166,179] Once the tumor has been located, surgical removal should be considered. This is best done antenatally in the second trimester or in the postpartum period.[177,178] All such patients should be referred for genetic counseling.

KEY POINTS

- Approximately 40% of patients with a neuroendocrine chromaffin cell tumor (pheochromocytoma or paraganglioma) have an underlying pathogenic germline genetic mutation.
- At the onset of hypertension prior to 20 weeks, the absence of proteinuria, edema, or hyperuricemia as well as evidence of episodic signs/symptoms is suggestive of a pheochromocytoma.
- Measurement of 24-hour urinary fractionated metanephrines and catecholamines or plasma metanephrines can be used to confirm the diagnosis as for nonpregnant patients.

PARATHYROID GLANDS

Regulation of Calcium During Pregnancy

Calcium is actively transferred from the maternal circulation to the fetoplacental unit throughout pregnancy, with 80% transferred in the third trimester and an estimated 300 to 350 mg/day transferred in the final 6 weeks of gestation.[180] In healthy adults, approximately 25% of daily calcium intake is absorbed by the gastrointestinal tract. During pregnancy, the efficiency of intestinal calcium absorption doubles in order to meet the demands of the fetus and palcenta.[180] The expansion of intravascular volume during pregnancy leads to a decrease in serum albumin and hence a decrease in serum total calcium concentration.[180] However, ionized calcium and albumin-corrected serum calcium levels do not change in pregnancy and it is these measurements that should be used to accurately assess maternal calcium status during pregnancy. The increased GFR in pregnancy leads to an increase in urinary calcium excretion.[180] Urinary calcium increases significantly from 4.3 ± 2.2 mmol/day prior to pregnancy to 6.2 ± 3.7 mmol/day in the third trimester.[181] Despite demands of the fetoplacental unit for calcium and the significant rise in urinary calcium excretion, maternal bone mineral density does not change significantly during pregnancy.[181]

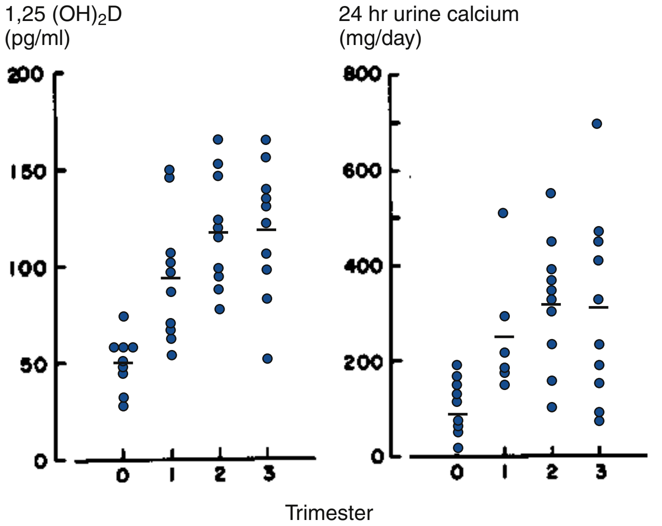


Figure 131.14 Circulating 1,25-dihydroxyvitamin D concentration and 24-hour calcium excretion during pregnancy by trimester. The values for both determinations were significantly increased ($P < 0.05$) during pregnancy (trimesters 1 to 3) when compared with postpartum values (trimester 0). (Data from Gertner JM, Couston DR, Kliger AS, et al. Pregnancy as state of physiologic absorptive hypercalciuria. *Am J Med.* 1986;81:451–456.)

Longitudinal studies of women in North America and Europe have demonstrated that circulating parathyroid hormone (PTH) levels in pregnancy are low-normal in the first trimester (1.3 ± 0.3 pmol/L) and either increase to mid-normal in the second trimester (2.3 ± 0.4 pmol/L) before decreasing to first trimester levels by term or remain suppressed throughout pregnancy.[182,183] Serum 1,25-dihydroxycholecalciferol (DHCC, calcitriol) levels increase 2- to 3-fold in pregnancy[182] (Fig. 131.14). High levels of DHCC explain in part the increase in intestinal absorption of calcium seen during pregnancy, which in turn suppresses the release of PTH.[180] Outside of pregnancy, PTH is the main regulator of renal Cyp27b1 gene expression, which codes for the enzyme 1α-hydroxylase that catalyzes the conversion of vitamin D to its active form, DHCC. Without sufficient PTH, DHCC levels remain low.[180] However, the low- to mid-normal levels of PTH in pregnancy suggest that PTH is not the primary factor responsible for the upregulation of Cyp27b1 expression and the subsequent increase in DHCC[180] (Fig. 131.15). Animal models suggested that the hormones estrogen, PRL, and placental lactogen also contribute to the increase in Cyp27b1 gene expression.[184]

25-Hydroxyvitamin D levels remain unchanged in pregnancy despite greater conversion to DHCC and significant transplacental

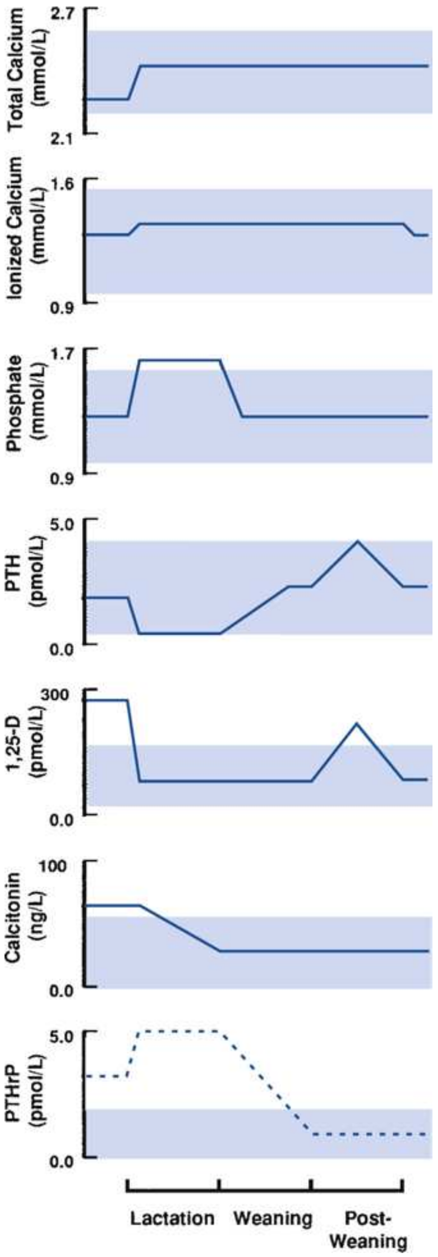


Figure 131.15 **Schematic illustration of the longitudinal changes in calcium, phosphate, and calcitropic hormone levels that occur during human pregnancy.** Normal adult ranges are indicated by the *shaded areas. PTH,* parathyroid hormone; *PTHrP,* parathyroid hormone-related peptide. (Data from Kovacs CS, Kronenberg HM. Maternal-fetal calcium and bone metabolism during pregnancy, puerperium, and lactation. *Endocr Rev.* 1997;18:832–872; Copyright © The Endocrine Society.)

transfer to the fetus.[180] Supplementation with up to 5000 IU of vitamin D during pregnancy does not change maternal serum calcium, albumin-corrected calcium, PTH, or DHCC levels.[180] Interestingly, 25-hydroxyvitamin D levels are approximately 30% lower in twin pregnancies.[185]

High estrogen levels in pregnancy stimulate calcitonin production by the parafollicular cells (C cells) of the thyroid gland. Calcitonin regulates calcium and phosphate levels in the blood by opposing the action of PTH. Circulating calcitonin levels are 3- to 5-fold higher in pregnant than in nonpregnant women.[186] In addition to C cells of the thyroid, the placenta is also a significant source of calcitonin.[187] In women with a history of total thyroidectomy, serum calcitonin levels are undetectable outside of pregnancy, but increase to 16 pg/mL in pregnancy, which exceeds the normal level of 11 pg/mL in nonpregnant women with an intact thyroid gland.[188]

Parathyroid Hormone–Related Protein

Parathyroid hormone–related protein (PTHrP) is produced by the parathyroid glands and placenta and acts to increase circulating calcium levels by promoting osteoclast-induced bone resorption.[189] Other tissues have been proposed as a potential source of PTHrP during pregnancy including the breasts, amnion, decidua, myometrium, umbilical cord, and other fetal tissues, but the evidence is inconclusive.[180] Measuring circulating PTHrP levels can be challenging because it is rapidly degraded. As such, plasma samples for PTHrP measurements should be collected in chilled ethylenediaminetetraacetic acid (EDTA) tubes containing a protease inhibitor such as aprotinin, transported immediately to the laboratory on ice, spun in a refrigerated centrifuge within 15 minutes, separated, and stored frozen until assayed.[190] Improper handling may lead to inaccurately low measured levels of PTHrP.[190] There are three full-length isoforms of PTHrP that result from alternative splicing of a single gene, referred to as $\text{PTHrP}^{1\text{-}139}$, $\text{PTHrP}^{1\text{-}141}$, and $\text{PTHrP}^{1\text{-}173}$.[191] A fourth isoform ($\text{PTHrP}^{1\text{-}36}$) arises from posttranslational processing of any of the three full-length molecules.[191] The most widely used commercial assays measure $\text{PTHrP}^{1\text{-}86}$, which will detect all three full-length isoforms but not $\text{PTHrP}^{1\text{-}36}$.[180] Despite these technical challenges, PTHrP levels increase during pregnancy from a mean of 0.8 ± 0.1 pmol/L in the first trimester to 2.0 ± 0.2 pmol/L at term.[182] Interestingly, this increase is significantly blunted in patients with preeclampsia.[192] Elevated levels of estradiol promote PTHrP production through a direct effect on gene transcription.[193]

Vitamin D Deficiency

The optimal level of 25-hydroxyvitamin D in pregnancy is unknown. The observed decrease in PTH and increase in DHCC appear to be programmed responses to pregnancy and not secondary to 25-hydroxyvitamin D deficiency.[180] Indeed, vitamin D deficiency does not decrease intestinal calcium absorption until the serum 25-hydroxyvitamin D concentration falls below 10 nM.[194] This is because calcium absorption is regulated by DHCC and not 25-hydroxyvitamin D, and the secondary hyperparathyroidism of vitamin D deficiency stimulates renal CYP27b1 gene expression to increase DHCC.[194] Dawodu et al. investigated women with vitamin D deficiency (with a mean baseline 25-hydroxyvitamin D level of 20.5 nM) who were randomized to receive either 400 IU, 2000 IU, or 4000 IU of vitamin D daily throughout pregnancy. At delivery, circulating 25-hydroxyvitamin D levels were 40 nM, 65 nM, and 90 nM, respectively, although no differences were noted in maternal serum calcium or birth weight.[195] There are conflicting data regarding vitamin D deficiency and preeclampsia.[180] Studies have not demonstrated a consistent difference in cord blood calcium or phosphorus, PTH, or birth weight in vitamin D-supplemented mothers compared with placebo-supplemented mothers.[180]

Hypercalcemia

Primary hyperparathyroidism occurs in approximately 0.03% of pregnancies.[196] Symptoms of the resulting hypercalcemia include fatigue, nausea, and constipation, which can be attributed to pregnancy.[180] Severe hypercalcemia can present as hyperemesis gravidarum, acute pancreatitis, renal stones, or recurrent urinary tract infections.[197] Due to the dilutional decrease in calcium and albumin levels along with the reduction in PTH production, pathologic hypercalcemia can be masked in pregnancy.[180] A diagnosis of primary hyperparathyroidism can be made with an elevated ionized or albumin-corrected calcium level in conjunction with a nonsuppressed PTH level.[180] The hypercalcemia caused by primary hyperparathyroidism may be exacerbated by pregnancy because both conditions increase intestinal absorption of calcium, hypercalciuria, and some, albeit mild, bone resorption.[180] Transfer of excess calcium across the placenta and incorporation into the developing fetal skeleton can reduce the maternal complications of hypercalcemia.[180] After delivery, mothers may develop an acute hypercalcemic crisis.[198] The increase in circulating calcium in patients with hypercalcemia leads to suppression of the fetal parathyroid glands, which predisposes to neonatal hypocalcemic tetany once the prenatal supply of calcium is removed.[198] Primary hyperparathyroidism is not associated with adverse pregnancy events, including miscarriage, pregnancy-induced hypertension, polyhydramnios, and/or large-for-gestational age infants.[196] The recommended treatment for asymptomatic women with primary hyperparathyroidism, regardless of pregnancy status, is parathyroidectomy.[199] The decision of whether to perform surgery antenatally (ideally in the second trimester) or postpartum should be individualized.[180] In addition to primary hyperparathyroidism, other potential causes of hypercalcemia that should be considered include excessive production of PTHrP by the placenta[200] or mammary tissues.[201]

KEY POINTS

- A diagnosis of primary hyperparathyroidism can be made with an elevated ionized or albumin-corrected calcium level in conjunction with a nonsuppressed PTH level.
- The transplacental transfer of calcium to the fetoplacental unit and incorporation of calcium into the developing fetal skeleton serve to decrease the maternal complications of hypercalcemia.

CONCLUSION

A thorough knowledge of the many complex physiologic and endocrine changes of pregnancy is fundamental to ensure a timely diagnosis of the pathologies described in this chapter. Ultimately this will allow for expeditious initiation of treatment and optimal maternal and fetal outcomes.

REFERENCES

1. Brown MA. Sodium and plasma volume regulation in normal and hypertensive pregnancy: a review of physiology and clinical implications. *Clin Exp Hypertens B*. 1988:265–282.
2. Mahendru AA, Everett TR, Wilkinson IB, Lees CC, McEniery CM. A longitudinal study of maternal cardiovascular function from preconception to the postpartum period. *J Hypertens*. 2014;32:849–856.
3. Lindheimer MD, Davison JM, Katz AI. The kidney and hypertension in pregnancy: twenty exciting years. *Semin Nephrol*. 2001;21:173–189.
4. Davison JM, Dunlop W, Ezimokhai M. 24-hour creatinine clearance during the third trimester of normal pregnancy. *Br J Obstet Gynaecol*. 1980;87:106–109.
5. Harel Z, McArthur E, Hladunewich M, et al. Serum creatinine levels before, during, and after pregnancy. *J Am Med Assoc*. 2019;321:205–207.
6. Coolen JC, Verhaeghe J. Physiology and clinical value of glycosuria after a glucose challenge during pregnancy. *Eur J Obstet Gynecol Reprod Biol*. 2010;150:132–136.
7. Buster JE, Abraham GE. The applications of steroid hormone radioimmunoassays to clinical obstetrics. *Obstet Gynecol*. 1975;46:489–499.
8. Mukherjee A, Alzhanov D, Rotwein P. Defining human insulin-like growth factor I gene regulation. *Am J Physiol Endocrinol Metab*. 2016;311:E519–E529.
9. Ishimoto H, Jaffe RB. Development and function of the human fetal adrenal cortex: a key component in the feto-placental unit. *Endocr Rev*. 2011;32:317–355.
10. Gubbi S, Hannah-Shmouni F, Stratakis CA, Koch CA. Primary hypophysitis and other autoimmune disorders of the sellar and suprasellar regions. *Rev Endocr Metab Disord*. 2018;19:335–347.
11. Caturegli P, Newschaffer C, Olivi A, et al. Autoimmune hypophysitis. *Endocr Rev*. 2005;26:599–614.
12. Honegger J, Schlaffer S, Menzel C, et al. Diagnosis of primary hypophysitis in Germany. *J Clin Endocrinol Metab*. 2015;100:3841–3849.
13. Foyouzi N. Lymphocytic adenohypophysitis. *Obstet Gynecol Surv*. 2011;66:109–113.
14. Diri H, Tanriverdi F, Karaca Z, et al. Extensive investigation of 114 patients with Sheehan's syndrome: a continuing disorder. *Eur J Endocrinol*. 2014;171:311–318.
15. Capatina C, Inder W, Karavitaki N, Wass JA. Management of endocrine disease: pituitary tumour apoplexy. *Eur J Endocrinol*. 2015;172:R179–R190.
16. Raappana A, Koivukangas J, Ebeling T, Pirila T. Incidence of pituitary adenomas in Northern Finland in 1992-2007. *J Clin Endocrinol Metab*. 2010;95:4268–4275.
17. Olsson DS, Nilsson AG, Bryngelsson IL, et al. Excess mortality in women and young adults with nonfunctioning pituitary adenoma: a Swedish nationwide study. *J Clin Endocrinol Metab*. 2015;100:2651–2658.
18. Tyson JE, Friesen HG. Factors influencing the secretion of human prolactin and growth hormone in menstrual and gestational women. *Am J Obstet Gynecol*. 1973;116:377–387.
19. Gonzalez JG, Elizondo G, Saldivar D, et al. Pituitary gland growth during normal pregnancy: an in vivo study using magnetic resonance imaging. *Am J Med*. 1988;85:217–220.
20. Heaney AP, Horwitz GA, Wang Z, Singson R, Melmed S. Early involvement of estrogen-induced pituitary tumor transforming gene and fibroblast growth factor expression in prolactinoma pathogenesis. *Nat Med*. 1999;5:1317–1321.
21. Rakoff JS, Yen SS. Progesterone induced acute release of prolactin in estrogen primed ovariectomized women. *J Clin Endocrinol Metab*. 1978;47:918–921.
22. Rigg LA, Lein A, Yen SS. Pattern of increase in circulating prolactin levels during human gestation. *Am J Obstet Gynecol*. 1977;129:454–456.
23. Kletzky OA, Marrs RP, Howard WF, McCormick W, Mishell Jr DR. Prolactin synthesis and release during pregnancy and puerperium. *Am J Obstet Gynecol*. 1980;136:545–550.
24. Soares MJ, Faria TN, Roby KF, Deb S. Pregnancy and the prolactin family of hormones: coordination of anterior pituitary, uterine, and placental expression. *Endocr Rev*. 1991;12:402–423.
25. Ho Yuen B, Cannon W, Lewis J, Sy L, Woolley S. A possible role for prolactin in the control of human chorionic gonadotropin and estrogen secretion by the fetoplacental unit. *Am J Obstet Gynecol*. 1980;136:286–291.
26. Molitch ME. Endocrinology in pregnancy: management of the pregnant patient with a prolactinoma. *Eur J Endocrinol*. 2015;172:R205–R213.
27. Molitch ME. Prolactinoma in pregnancy. *Best Pract Res Clin Endocrinol Metab*. 2011;25:885–896.
28. Raymond JP, Goldstein E, Konopka P, et al. Follow-up of children born of bromocriptine-treated mothers. *Horm Res*. 1985;22:239–246.
29. Kinon BJ, Gilmore JA, Liu H, Halbreich UM. Prevalence of hyperprolactinemia in schizophrenic patients treated with conventional antipsychotic medications or risperidone. *Psychoneuroendocrinology*. 2003;28(suppl 2):55–68.
30. Knegtering R, Baselmans P, Castelein S, et al. Predominant role of the 9-hydroxy metabolite of risperidone in elevating blood prolactin levels. *Am J Psychiatry*. 2005;162:1010–1012.
31. Huybrechts KF, Hernandez-Diaz S, Patorno E, et al. Antipsychotic use in pregnancy and the risk for congenital malformations. *JAMA Psychiatr*. 2016;73:938–946.
32. Tyson JE, Hwang P, Guyda H, Friesen HG. Studies of prolactin secretion in human pregnancy. *Am J Obstet Gynecol*. 1972;113:14–20.
33. Liao S, Vickers MH, Stanley JL, Baker PN, Perry JK. Human placental growth hormone variant in pathological pregnancies. *Endocrinology*. 2018;159:2186–2198.

34. Frankenne F, Closset J, Gomez F, et al. The physiology of growth hormones (GHs) in pregnant women and partial characterization of the placental GH variant. *J Clin Endocrinol Metab*. 1988;66:1171–1180.
35. Alsat E, Guibourdenche J, Couturier A, Evain-Brion D. Physiological role of human placental growth hormone. *Mol Cell Endocrinol*. 1998;140:121–127.
36. Beckers A, Stevenaert A, Foidart JM, Hennen G, Frankenne F. Placental and pituitary growth hormone secretion during pregnancy in acromegalic women. *J Clin Endocrinol Metab*. 1990;71:725–731.
37. Patel N, Alsat E, Igout A, et al. Glucose inhibits human placental GH secretion, in vitro. *J Clin Endocrinol Metab*. 1995;80:1743–1746.
38. Schiessl B, Strasburger CJ, Bidlingmaier M, et al. Role of placental growth hormone in the alteration of maternal arterial resistance in pregnancy. *J Reprod Med*. 2007;52:313–316.
39. Mittal P, Espinoza J, Hassan S, et al. Placental growth hormone is increased in the maternal and fetal serum of patients with preeclampsia. *J Matern Fetal Neonatal Med*. 2007;20:651–659.
40. Karaca Z, Kelestimur F. Pregnancy and other pituitary disorders (including GH deficiency). *Best Pract Res Clin Endocrinol Metab*. 2011;25:897–910.
41. Lonberg U, Damm P, Andersson AM, et al. Increase in maternal placental growth hormone during pregnancy and disappearance during parturition in normal and growth hormone-deficient pregnancies. *Am J Obstet Gynecol*. 2003;188:247–251.
42. Vila G, Luger A. Growth hormone deficiency and pregnancy: any role for substitution? *Minerva Endocrinol*. 2018;43:451–457.
43. Sakai S, Wakasugi T, Yagi K, et al. Successful pregnancy and delivery in a patient with adult GH deficiency: role of GH replacement therapy. *Endocr J*. 2011;58:65–68.
44. Caron P, Broussaud S, Bertherat J, et al. Acromegaly and pregnancy: a retrospective multicenter study of 59 pregnancies in 46 women. *J Clin Endocrinol Metab*. 2010;95:4680–4687.
45. Karaca Z, Tanriverdi F, Unluhizarci K, Kelestimur F. Pregnancy and pituitary disorders. *Eur J Endocrinol*. 2010;162:453–475.
46. Fassnacht M, Capeller B, Arlt W, Steck T, Allolio B. Octreotide LAR treatment throughout pregnancy in an acromegalic woman. *Clin Endocrinol*. 2001;55:411–415.
47. Harris K, Shankar R, Black K, Rochelson B. Reset osmostat in pregnancy: a case report. *J Matern Fetal Neonatal Med*. 2014;27:530–533.
48. Davison JM, Vallotton MB, Lindheimer MD. Plasma osmolality and urinary concentration and dilution during and after pregnancy: evidence that lateral recumbency inhibits maximal urinary concentrating ability. *Br J Obstet Gynaecol*. 1981;88:472–479.
49. Marques P, Gunawardana K, Grossman A. Transient diabetes insipidus in pregnancy. *Endocrinol Diabetes Metab Case Rep*. 2015;2015:150078.
50. Davison JM, Gilmore EA, Durr J, Robertson GL, Lindheimer MD. Altered osmotic thresholds for vasopressin secretion and thirst in human pregnancy. *Am J Physiol*. 1984;246:F105–F109.
51. Kalelioglu I, Kubat Uzum A, Yildirim A, et al. Transient gestational diabetes insipidus diagnosed in successive pregnancies: review of pathophysiology, diagnosis, treatment, and management of delivery. *Pituitary*. 2007;10:87–93.
52. Schrier RW. Systemic arterial vasodilation, vasopressin, and vasopressinase in pregnancy. *J Am Soc Nephrol*. 2010;21:570–572.
53. Chong PL, Pisharam J, Abdullah A, Chong VH. Gestational diabetes insipidus. *QJM*. 2019;112:123–124.
54. Gambito R, Chan M, Sheta M, et al. Gestational diabetes insipidus associated with HELLP syndrome: a case report. *Case Rep Nephrol*. 2012;2012:640365.
55. Wallia A, Bizhanova A, Huang W, et al. Acute diabetes insipidus mediated by vasopressinase after placental abruption. *J Clin Endocrinol Metab*. 2013;98:881–886.
56. Oiso Y, Robertson GL, Norgaard JP, Juul KV. Clinical review: treatment of neurohypophyseal diabetes insipidus. *J Clin Endocrinol Metab*. 2013;98:3958–3967.
57. Ray JG. DDAVP use during pregnancy: an analysis of its safety for mother and child. *Obstet Gynecol Surv*. 1998;53:450–455.
58. Nawathe A, Govind A. Pregnancy with known syndrome of inappropriate antidiuretic hormone. *J Obstet Gynaecol*. 2013;33:9–13.
59. Wilson HJ, Shutt LE. Syndrome of inappropriate ADH secretion in a woman with preeclampsia. *Int J Obstet Anesth*. 2007;16:360–362.
60. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocr Rev*. 1997;18:404–433.
61. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*. 2010;31:702–755.
62. Lazarus JH, Bestwick JP, Channon S, et al. Antenatal thyroid screening and childhood cognitive function. *N Engl J Med*. 2012;366:493–501.
63. FIGO Working Group on Good Clinical Practice in Maternal-Fetal Medicine. Good clinical practice advice: thyroid and pregnancy. *Int J Gynaecol Obstet*. 2019;144:347–351.
64. Yalamanchi S, Cooper DS. Thyroid disorders in pregnancy. *Curr Opin Obstet Gynecol*. 2015;27:406–415.
65. Ballabio M, Poshychinda M, Ekins RP. Pregnancy-induced changes in thyroid function: role of human chorionic gonadotropin as putative regulator of maternal thyroid. *J Clin Endocrinol Metab*. 1991;73:824–831.
66. Bartalena L. Recent achievements in studies on thyroid hormone-binding proteins. *Endocr Rev*. 1990;11:47–64.
67. Caldwell KL, Pan Y, Mortensen ME, et al. Iodine status in pregnant women in the national Children's study and in U.S. Women (15-44 years), National Health and Nutrition examination Survey 2005-2010. *Thyroid*. 2013;23:927–937.
68. Zimmermann MB, Gizak M, Abbott K, Andersson M, Lazarus JH. Iodine deficiency in pregnant women in Europe. *Lancet Diabetes Endocrinol*. 2015;3:672–674.
69. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;27:315–389.
70. De Groot L, Abalovich M, Alexander EK, et al. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2012;97:2543–2565.
71. Delange FMDJ. In: *Iodine Deficiency*. 9th ed. Lippincott, Williams and Wilkins; 2005.
72. World Health Organization. In: *Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination: A Guide for Programme Managers*. 3rd ed. Geneva: World Health Organization; 2007.
73. Halnan KE. The radioiodine uptake of the human thyroid in pregnancy. *Clin Sci*. 1958;17:281–290.
74. Brent RL. The effect of embryonic and fetal exposure to x-ray, microwaves, ultrasound, magnetic resonance and isotopes. In: Baron WM, Lindheimer MD, eds. *Medical Disorders of pregnancy*. St Louis: Mosby;1991:568–604.
75. Ferianec V, Papcun P, Grochal F, Schenkova K, Bartova M. Prenatal diagnosis and successful intrauterine treatment of severe congenital hypothyroidism associated with fetal goiter. *J Obstet Gynaecol Res*. 2017;43:232–237.
76. Haddow JE, McClain MR, Lambert-Messerlian G, et al. Variability in thyroid-stimulating hormone suppression by human chorionic [corrected] gonadotropin during early pregnancy. *J Clin Endocrinol Metab*. 2008;93:3341–3347.
77. Tisne L, Barzelatto J, Stevenson C. [Study of thyroid function during pregnancy-puerperal state with radioactive iodine]. *Bol Soc Chil Obstet Ginecol*. 1955;20:246–251.
78. Ain KB, Mori Y, Refetoff S. Reduced clearance rate of thyroxine-binding globulin (TBG) with increased sialylation: a mechanism for estrogen-induced elevation of serum TBG concentration. *J Clin Endocrinol Metab*. 1987;65:689–696.
79. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:238–249.
80. Glinoer D, de Nayer P, Bourdoux P, et al. Regulation of maternal thyroid during pregnancy. *J Clin Endocrinol Metab*. 1990;71:276–287.
81. Thienpont LM, Van Uytfanghe K, Beastall G, et al. Report of the IFCC Working Group for standardization of thyroid function tests; part 1: thyroid-stimulating hormone. *Clin Chem*. 2010;56:902–911.

82. Bestwick JP, John R, Maina A, et al. Thyroid stimulating hormone and free thyroxine in pregnancy: expressing concentrations as multiples of the median (MoMs). *Clin Chim Acta*. 2014;430:33–37.
83. Glinoer D. Regulation of thyroid function in pregnancy: maternal and neonatal repercussions. *Adv Exp Med Biol*. 1991;299:197–201.
84. Mandel SJ, Larsen PR, Seely EW, Brent GA. Increased need for thyroxine during pregnancy in women with primary hypothyroidism. *N Engl J Med*. 1990;323:91–96.
85. Sullivan SD, Downs E, Popoveniuc G, et al. Randomized trial comparing two algorithms for levothyroxine dose adjustment in pregnant women with primary hypothyroidism. *J Clin Endocrinol Metab*. 2017;102:3499–3507.
86. Tan JY, Loh KC, Yeo GS, Chee YC. Transient hyperthyroidism of hyperemesis gravidarum. *BJOG*. 2002;109:683–688.
87. Andersen SL, Laurberg P, Wu CS, Olsen J. Attention deficit hyperactivity disorder and autism spectrum disorder in children born to mothers with thyroid dysfunction: a Danish nationwide cohort study. *BJOG*. 2014;121:1365–1374.
88. Zimmerman D. Fetal and neonatal hyperthyroidism. *Thyroid*. 1999;9:727–733.
89. Luton D, Le Gac I, Vuillard E, et al. Management of Graves' disease during pregnancy: the key role of fetal thyroid gland monitoring. *J Clin Endocrinol Metab*. 2005;90:6093–6098.
90. Segni M, Leonardi E, Mazzoncini B, Pucarelli I, Pasquino AM. Special features of Graves' disease in early childhood. *Thyroid*. 1999;9:871–877.
91. Polak M, Le Gac I, Vuillard E, et al. Fetal and neonatal thyroid function in relation to maternal Graves' disease. *Best Pract Res Clin Endocrinol Metab*. 2004;18:289–302.
92. Mori M, Amino N, Tamaki H, Miyai K, Tanizawa O. Morning sickness and thyroid function in normal pregnancy. *Obstet Gynecol*. 1988;72:355–359.
93. Swaminathan R, Chin RK, Lao TT, et al. Thyroid function in hyperemesis gravidarum. *Acta Endocrinol*. 1989;120:155–160.
94. Lydakis C, Lip GY, Beevers M, Beevers DG. Atenolol and fetal growth in pregnancies complicated by hypertension. *Am J Hypertens*. 1999;12:541–547.
95. Mortimer RH, Cannell GR, Addison RS, et al. Methimazole and propylthiouracil equally cross the perfused human term placental lobule. *J Clin Endocrinol Metab*. 1997;82:3099–3102.
96. Andersen SL, Olsen J, Wu CS, Laurberg P. Birth defects after early pregnancy use of antithyroid drugs: a Danish nationwide study. *J Clin Endocrinol Metab*. 2013;98:4373–4381.
97. Bahn RS, Burch HS, Cooper DS, et al. The role of propylthiouracil in the management of Graves' disease in adults: report of a meeting jointly sponsored by the American Thyroid Association and the Food and Drug Administration. *Thyroid*. 2009;19:673–674.
98. Sarlis NJ, Gourgiotis L. Thyroid emergencies. *Rev Endocr Metab Disord*. 2003;4:129–136.
99. Angell TE, Lechner MG, Nguyen CT, et al. Clinical features and hospital outcomes in thyroid storm: a retrospective cohort study. *J Clin Endocrinol Metab*. 2015;100:451–459.
100. Tsatsoulis A, Johnson EO, Kalogera CH, Seferiadis K, Tsolas O. The effect of thyrotoxicosis on adrenocortical reserve. *Eur J Endocrinol*. 2000;142:231–235.
101. Casey BM, Leveno KJ. Thyroid disease in pregnancy. *Obstet Gynecol*. 2006;108:1283–1292.
102. Teng W, Shan Z, Patil-Sisodia K, Cooper DS. Hypothyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:228–237.
103. van den Boogaard E, Vissenberg R, Land JA, et al. Significance of (sub) clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. *Hum Reprod Update*. 2011;17:605–619.
104. Haddow JE, Palomaki GE, Allan WC, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*. 1999;341:549–555.
105. Casey BM, Dashe JS, Spong CY, et al. Perinatal significance of isolated maternal hypothyroxinemia identified in the first half of pregnancy. *Obstet Gynecol*. 2007;109:1129–1135.
106. Negro R, Schwartz A, Gismondi R, et al. Increased pregnancy loss rate in thyroid antibody negative women with TSH levels between 2.5 and 5.0 in the first trimester of pregnancy. *J Clin Endocrinol Metab*. 2010;95:E44–E48.
107. Thangaratinam S, Tan A, Knox E, et al. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. *BMJ*. 2011;342:d2616.
108. Negro R, Formoso G, Mangieri T, et al. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. *J Clin Endocrinol Metab*. 2006;91:2587–2591.
109. American College of Obstretrics & Gynecology. Practice Bulletin No. 148: thyroid disease in pregnancy. *Obstet Gynecol*. 2015;125:996–1005.
110. Dhillon-Smith RK, Middleton LJ, Sunner KK, et al. Levothyroxine in women with thyroid peroxidase antibodies before conception. *N Engl J Med*. 2019;380:1316–1325.
111. Wang H, Gao H, Chi H, et al. Effect of levothyroxine on miscarriage among women with normal thyroid function and thyroid autoimmunity undergoing in vitro fertilization and embryo transfer: a randomized clinical trial. *J Am Med Assoc*. 2017;318:2190–2198.
112. Santisteban PBL, Utiger RD. *Werner and Ingbar's the Thyroid: A Fundamental & Clinical Text*. Philadelphia: Lippincott, Williams and Wilkins; 2005:9–23.
113. Raymond J, LaFranchi SH. Fetal and neonatal thyroid function: review and summary of significant new findings. *Curr Opin Endocrinol Diabetes Obes*. 2010;17:1–7.
114. Levin G, Elchalal U, Rottenstreich A. The adrenal cortex: physiology and diseases in human pregnancy. *Eur J Obstet Gynecol Reprod Biol*. 2019;240:139–143.
115. Campbell EA, Linton EA, Wolfe CD, et al. Plasma corticotropin-releasing hormone concentrations during pregnancy and parturition. *J Clin Endocrinol Metab*. 1987;64:1054–1059.
116. Thomson M. The physiological roles of placental corticotropin releasing hormone in pregnancy and childbirth. *J Physiol Biochem*. 2013;69:559–573.
117. Smith R, Nicholson RC. Corticotrophin releasing hormone and the timing of birth. *Front Biosci*. 2007;12:912–918.
118. Thomson M. Does the CRH binding protein shield the anterior pituitary from placental CRH? *Endocrine*. 1998;9:221–226.
119. McLean M, Bisits A, Davies J, et al. A placental clock controlling the length of human pregnancy. *Nat Med*. 1995;1:460–463.
120. Waddell BJ. The placenta as hypothalamus and pituitary: possible impact on maternal and fetal adrenal function. *Reprod Fertil Dev*. 1993;5:479–497.
121. Demura R, Odagiri E, Yoshimura M, et al. Placental secretion of prolactin, ACTH and immunoreactive β-endorphin during pregnancy. *Acta Endocrinol*. 1982;100:114–119.
122. Carr BR, Parker CR, Madden JD, MacDonald PC, Porter JC. Maternal plasma adrenocorticotropin and cortisol relationships throughout human pregnancy. *Am J Obstet Gynecol*. 1981;139:416–422.
123. Magiakou MA, Mastorakos G, Rabin D, et al. The maternal hypothalamic-pituitary-adrenal axis in the third trimester of human pregnancy. *Clin Endocrinol*. 1996;44:419–428.
124. Mukherjee K, Swyer GIM. Plasma cortisol and adrenocorticotrophic hormone in normal men and non–pregnant women, normal pregnant women and women with pre–eclampsia. *J Obstet Gynaecol Br Commonw*. 1972;79:504–512.
125. Jung C, Ho JT, Torpy DJ, et al. A longitudinal study of plasma and urinary cortisol in pregnancy and postpartum. *J Clin Endocrinol Metabol*. 2011;96:1533–1540.
126. Cousins L, Rigg L, Hollingsworth D, et al. Qualitative and quantitative assessment of the circadian rhythm of cortisol in pregnancy. *Am J Obstet Gynecol*. 1983;145:411–416.
127. Caimari F, Valassi E, Garbayo P, et al. Cushing's syndrome and pregnancy outcomes: a systematic review of published cases. *Endocrine*. 2017;55:555–563.
128. Herman Chui M, Özbey NC, Ezzat S, et al. Case report: adrenal LH/hCG receptor overexpression and gene amplification causing pregnancy-induced Cushing's syndrome. *Endocr Pathol*. 2009;20:256–261.

129. Lindsay JR, Nieman LK. The hypothalamic-pituitary-adrenal axis in pregnancy: challenges in disease detection and treatment. *Endocr Rev*. 2005;26:775–799.
130. Odagiri E, Ishiwatari N, Abe Y, et al. Hypercortisolism and the resistance to dexamethasone suppression during gestation. *Endocrinol Jpn*. 1988;35:685–690.
131. Ross RJM, Chew SL, Perry L, et al. Diagnosis and selective cure of Cushing's disease during pregnancy by transsphenoidal surgery. *Eur J Endocrinol*. 1995;132:722–726.
132. Lindsay JR, Jonklaas J, Oldfield EH, Nieman LK. Extensive clinical experience: Cushing's syndrome during pregnancy: personal experience and review of the literature. *J Clin Endocrinol Metabol*. 2005;90:3077–3083.
133. Nolten WE, Lindheimer MD, Rueckert PA, Oparil S, Ehrlich EN. Diurnal patterns and regulation of cortisol secretion in pregnancy. *J Clin Endocrinol Metabol*. 1980;51:466–472.
134. Nieman LK, Chrousos GP, Oldfield EH, et al. The ovine corticotropin-releasing hormone stimulation test and the dexamethasone suppression test in the differential diagnosis of Cushing's syndrome. *Ann Intern Med*. 1986;105:862–867.
135. Tabarin A, Laurent F, Catargi B, et al. Comparative evaluation of conventional and dynamic magnetic resonance imaging of the pituitary gland for the diagnosis of Cushing's disease. *Clin Endocrinol*. 1998;49:293–300.
136. Wallace C, Toth EL, Lewanczuk RZ, Siminoski K. Pregnancy-induced Cushing's syndrome in multiple pregnancies. *J Clin Endocrinol Metabol*. 1996;81:15–21.
137. Connell JMC, Cordiner J, Davies DL, et al. Pregnancy complicated by Cushing's syndrome: potential hazard of metyrapone therapy. Case report. *Br J Obstet Gynaecol*. 1985;92:1192–1195.
138. Amadol JA, Pesquera C, Gonzalez EM, et al. Successful treatment with ketoconazole of Cushing's syndrome in pregnancy. *Postgrad Med J*. 1990;66:221–223.
139. Ten S, New M, Maclaren N. Addison's disease 2001. *J Clin Endocrinol Metabol*. 2001;86:2909–2922.
140. McKenna DS, Wittber GM, Nagaraja HN, Samuels P. The effects of repeat doses of antenatal corticosteroids on maternal adrenal function. *Am J Obstet Gynecol*. 2000;183:669–673.
141. George LD, Selvaraju R, Reddy K, Stout TV, Premawardhana L. Vomiting and hyponatremia in pregnancy. *Br J Obstet Gynaecol*. 2000;107:808–809.
142. Gradden C, Lawrence D, Doyle PM, Welch CR. Addison's disease in pregnancy. *Lancet*. 2001;357(9263):1197.
143. Ambrosi B, Barbetta L, Morricone L. Diagnosis and management of Addison's disease during pregnancy. *J Endocrinol Invest*. 2003;26:698–702.
144. Mahendru AA, Everett TR, Wilkinson IB, Lees CC, McEniery CM. Maternal cardiovascular changes from pre-pregnancy to very early pregnancy. *J Hypertens*. 2012;30:2168–2172.
145. Nelson DB, Stewart RD, Matulevicius SA, et al. The effects of maternal position and habitus on maternal cardiovascular parameters as measured by cardiac magnetic resonance. *Am J Perinatol*. 2015;32:1318–1323.
146. Odutayo A, Hladunewich M. Obstetric nephrology: renal hemodynamic and metabolic physiology in normal pregnancy. *Clin J Am Soc Nephrol*. 2012;7:2073–2080.
147. Kumar P, Magon N. Hormones in pregnancy. *Niger Med J*. 2012;53:179–183.
148. Quinkler M, Meyer B, Oelkers W, Diederich S. Renal inactivation, mineralocorticoid generation, and 11β-hydroxysteroid dehydrogenase inhibition ameliorate the antimineralocorticoid effect of progesterone in vivo. *J Clin Endocrinol Metabol*. 2003;88:3767–3772.
149. Landau RL, Lugibihl K. Inhibition of the sodium-retaining influence of aldosterone by progesterone. *J Clin Endocrinol Metabol*. 1958;18:1237–1245.
150. Lumbers ER, Pringle KG. Roles of the circulating renin-angiotensin-aldosterone system in human pregnancy. *Am J Physiol Regul Integr Comp Physiol*. 2014;306:R91–R101.
151. Wilson M, Morganti AA, Zervoudakis I, et al. Blood pressure, the renin-aldosterone system and sex steroids throughout normal pregnancy. *Am J Med*. 1980;68:97–104.
152. Itskovitz J, Sealey JE, Glorioso N, Rosenwaks Z. Plasma prorenin response to human chorionic gonadotropin in ovarian-hyperstimulated women: correlation with the number of ovarian follicles and steroid hormone concentrations. *Proc Natl Acad Sci U S A*. 1987;84:7285–7289.
153. Weinberger MH, Kramer NJ, Grim CE, Petersen LP. The effect of posture and saline loading on plasma renin activity and aldosterone concentration in pregnant, non-pregnant and estrogen-treated women. *J Clin Endocrinol Metabol*. 1977;44:69–77.
154. Nielsen AH, Schauser KH, Poulsen K. The uteroplacental renin-angiotensin system. *Placenta*. 2000;21:468–477.
155. Immonen I, Siimes A, Stenman UH, Kärkkäinen J, Fyhrquist F. Plasma renin substrate and oestrogens in normal pregnancy. *Scand J Clin Lab Invest*. 1983;43:61–65.
156. Zhou A, Carrell RW, Murphy MP, et al. A redox switch in angiotensinogen modulates angiotensin release. *Nature*. 2010;468:108–111.
157. Tewksbury DA, Dart RA. High molecular weight angiotensinogen levels in hypertensive pregnant women. *Hypertension*. 1982;4:729–734.
158. Weyer K, Glerup S. Placental regulation of peptide hormone and growth factor activity by proMBP1. *Biol Reprod*. 2011;84:1077–1086.
159. Parente JV, Franco JG, Greene LJ, Meirelles RS, Martinez AR. Angiotensin-converting enzyme: serum levels during normal pregnancy. *Am J Obstet Gynecol*. 1979;135:586–589.
160. Pringle KG, Tadros MA, Callister RJ, Lumbers ER. The expression and localization of the human placental prorenin/renin-angiotensin system throughout pregnancy: roles in trophoblast invasion and angiogenesis? *Placenta*. 2011;32:956–962.
161. Brosnihan KB, Neves LAA, Anton L, et al. Enhanced expression of Ang-(1-7) during pregnancy. *Braz J Med Biol Res*. 2004;37:1255–1262.
162. Nolten WE, Ehrlich EN. Sodium and mineralocorticoids in normal pregnancy. *Kidney Int*. 1980;18:162–172.
163. Brown MA, Zammit VC, Miltar DA, Whitworth JA. Renin-aldosterone relationships in pregnancy-induced hypertension. *Am J Hypertens*. 1992;5:366–371.
164. Watanabe M, Meeker CI, Gray MJ, Sims EA, Solomon S. Secretion rate of aldosterone in normal pregnancy. *J Clin Invest*. 1963;42:1619–1631.
165. Matsmumoto J, Miyake H, Isozaki T, Koshino T, Araki T. Primary aldosteronism in pregnancy. *J Nippon Med Sch*. 2000;67:275–279.
166. Geller DS, Farhi A, Pinkerton N, et al. Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science*. 2000;289:119–123.
167. Sarathi V, Lila AR, Bandgar TR, Menon PS, Shah NS. Pheochromocytoma and pregnancy: a rare but dangerous combination. *Endocr Pract*. 2010;16:300–309.
168. Lenders JW. Pheochromocytoma and pregnancy: a deceptive connection. *Eur J Endocrinol*. 2012;166:143–150.
169. Favier J, Amar L, Gimenez-Roqueplo AP. Paraganglioma and phaeochromocytoma: from genetics to personalized medicine. *Nat Rev Endocrinol*. 2015;11:101–111.
170. Natrajan PG, McGarrigle HH, Lawrence DM, Lachelin GC. Plasma noradrenaline and adrenaline levels in normal pregnancy and in pregnancy-induced hypertension. *Br J Obstet Gynaecol*. 1982;89:1041–1045.
171. Oian P, Kjeldsen SE, Eide I, Maltau JM. Increased arterial catecholamines in pre-eclampsia. *Acta Obstet Gynecol Scand*. 1986;65:613–617.
172. Katz VL, Jenkins T, Haley L, Bowes Jr WA. Catecholamine levels in pregnant physicians and nurses: a pilot study of stress and pregnancy. *Obstet Gynecol*. 1991;77:338–342.
173. Dahia PL, Hayashida CY, Strunz C, Abelin N, Toledo SP. Low cord blood levels of catecholamine from a newborn of a pheochromocytoma patient. *Eur J Endocrinol*. 1994;130:217–219.
174. Slotkin TA, Seidler FJ. Adrenomedullary catecholamine release in the fetus and newborn: secretory mechanisms and their role in stress and survival. *J Dev Physiol*. 1988;10:1–16.
175. Keely E. In: Lee RVGP, Barron WM, et al., eds. *Current Obstetrics Medicine*. St Louis: Mosby; 1986.
176. Biggar MA, Lennard TW. Systematic review of phaeochromocytoma in pregnancy. *Br J Surg*. 2013;100:182–190.

177. Lenders JWM, Langton K, Langenhuijsen JF, Eisenhofer G. Pheochromocytoma and pregnancy. *Endocrinol Metab Clin North Am*. 2019;48:605–617.
178. Lenders JW, Duh QY, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:1915–1942.
179. De Wilde JP, Rivers AW, Price DL. A review of the current use of magnetic resonance imaging in pregnancy and safety implications for the fetus. *Prog Biophys Mol Biol*. 2005;87:335–353.
180. Kovacs CS. Maternal mineral and bone metabolism during pregnancy, lactation, and post-weaning recovery. *Physiol Rev*. 2016;96:449–547.
181. Ritchie LD, Fung EB, Halloran BP, et al. A longitudinal study of calcium homeostasis during human pregnancy and lactation and after resumption of menses. *Am J Clin Nutr*. 1998;67:693–701.
182. Ardawi MS, Nasrat HA, BA'Aqueel HS. Calcium-regulating hormones and parathyroid hormone-related peptide in normal human pregnancy and postpartum: a longitudinal study. *Eur J Endocrinol*. 1997;137:402–409.
183. Black AJ, Topping J, Durham B, Farquharson RG, Fraser WD. A detailed assessment of alterations in bone turnover, calcium homeostasis, and bone density in normal pregnancy. *J Bone Miner Res*. 2010;15:557–563.
184. Spanos E, Brown DJ, Stevenson JC, MacIntyre I. Stimulation of 1,25-dihydroxycholecalciferol production by prolactin and related peptides in intact renal cell preparations in vitro. *Biochim Biophys Acta*. 1981;672:7–15.
185. Nakayama S, Yasui T, Suto M, et al. Differences in bone metabolism between singleton pregnancy and twin pregnancy. *Bone*. 2011;49:513–519.
186. Hillyard CJ, Stevenson JC, MacIntyre I. Relative deficiency of plasma-calcitonin in normal women. *Lancet*. 1978;311:961–962.
187. Balabanova S, Kruse B, Wolf AS. Calcitonin secretion by human placental tissue. *Acta Obstet Gynecol Scand*. 1987;66:323–326.
188. Bucht E, Telenius-Berg M, Lundell G, Sjoberg HE. Immunoextracted calcitonin in milk and plasma from totally thyroidectomized women. Evidence of monomeric calcitonin in plasma during pregnancy and lactation. *Acta Endocrinol*. 1986;113:529–535.
189. Leere JS, Vestergaard P. Calcium metabolic disorders in pregnancy: primary hyperparathyroidism, pregnancy-induced osteoporosis, and vitamin D deficiency in pregnancy. *Endocrinol Metab Clin North Am*. 2019;48:643–655.
190. Hutchesson ACJ, Hughes SV, Bowden SJ, Ratcliffe WA. In vitro stability of endogenous parathyroid hormone-related protein in blood and plasma. *Ann Clin Biochem*. 1994;31:35–39.
191. Orloff JJ, Reddy D, De Papp AE, et al. Parathyroid hormone-related protein as a prohormone: posttranslational processing and receptor interactions. *Endocr Rev*. 1994;15:40–60.
192. Yadav S, Yadav YS, Goel MM, et al. Calcitonin gene- and parathyroid hormone-related peptides in normotensive and preeclamptic pregnancies: a nested case–control study. *Arch Gynecol Obstet*. 2014;290:897–903.
193. Casey ML. Regulation of parathyroid hormone-related protein gene expression in human endometrial stromal cells in culture. *J Clin Endocrinol Metab*. 1993;77:188–194.
194. Need AG, O'Loughlin PD, Morris HA, et al. Vitamin D metabolites and calcium absorption in severe vitamin D deficiency. *J Bone Miner Res*. 2008;23:1859–1863.
195. Dawodu A, Saadi HF, Bekdache G, et al. Randomized controlled trial (RCT) of vitamin D supplementation in pregnancy in a population with endemic vitamin D deficiency. *J Clin Endocrinol Metab*. 2013;98:2337–2346.
196. Hirsch D, Kopel V, Nadler V, et al. Pregnancy outcomes in women with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2015;100:2115–2122.
197. Kort KC, Schiller HJ, Numann PJ. Hyperparathyroidism and pregnancy. *Am J Surg*. 1999;177:66–68.
198. Nilsson IL, Adner N, Reihner E, et al. Primary hyperparathyroidism in pregnancy: a diagnostic and therapeutic challenge. *J Womens Health (Larchmt)*. 2010;19:1117–1121.
199. Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99(10):3561–3569.
200. Eller-Vainicher C, Ossola MW, Beck-Peccoz P, Chiodini I. PTHrP-associated hypercalcemia of pregnancy resolved after delivery: a case report. *Eur J Endocrinol*. 2012;166:753–756.
201. Jackson IT, Saleh J, van Heerden JA. Gigantic mammary hyperplasia in pregnancy associated with pseudohyperparathyroidism. *Plast Reconstr Surg*. 1989;84:806–810.

132

Autoimmune Polyglandular Syndromes

Eystein Husebye and Anthony P. Weetman

OUTLINE

Background, 2222
The Basis of Autoimmune Disease, 2222
  The Normal Immune Response, 2223
  Self-Tolerance, 2225
  Central Tolerance, 2225
  Peripheral Tolerance, 2225
  Pathogenesis of Autoimmune Disease, 2225
Autoimmune Polyglandular Syndrome Type 1, 2226
  Definition, 2226
  Pathogenesis, 2226
  Clinical Features, 2227
  Chronic Mucocutaneous Candidiasis, 2228
  Hypoparathyroidism, 2229
  Adrenal Insufficiency, 2229
  Gonadal Insufficiency, 2229
  Other Endocrinopathies, 2229
  Gastrointestinal Manifestations, 2229
  Ectodermal Manifestations, 2230
  Other Manifestations, 2230
  Natural Course and Mortality, 2230
  Diagnosis and Management, 2230
Immumodysregulation Polyendocrinopathy Enteropathy X-Linked, 2231
Autoimmune Polyglandular Syndrome Type 2, 2231
  Definition, 2231
  Pathogenesis, 2231
  Clinical Features and Diagnosis, 2233
  Management, 2233
Immune Checkpoint Inhibitor-Induced Polyendocrinopathy, 2233

BACKGROUND

Autoimmune polyglandular syndromes (APSs) are characterized by functional impairment of multiple glands owing to a targeted immune-mediated inflammation and destruction characterized by insidious presentation, circulating autoantibodies, and lymphocytic infiltration into the affected organ, eventually leading to organ failure. Presentation can occur from early infancy to old age, and new components can appear throughout life. There is a huge variation in patterns and frequencies of autoimmunity in affected individuals, and APS tends to aggregate in families, pointing to strong heritability. A number of immune-related genes associate with APS, but environmental factors also play a decisive role.

Endocrine glands are commonly affected, but organ-specific autoimmunity can affect almost any tissue of the body, including skin, liver, muscles, and joints. The combination of multiple insufficiencies was described already in 1908 by Claude and Gourgerot who reported a patient with insufficiencies involving the pancreatic islets, thyroid gland, gonads, adrenal glands, and anterior pituitary.[1] In 1926 Schmidt documented the association between adrenocortical failure and chronic lymphocytic thyroiditis,[2] later extended to type 1 diabetes mellitus (T1DM).[3] In the 1950s autoantibodies were described, the first being those directed against thyroglobulin. Later, important diagnostic biomarkers were described, including autoantibodies directed against thyroid peroxidase,[4] glutamic acid decarboxylase-65,[5] and 21-hydroxylase,[6] all tissue-specific enzymes with key functions in the affected organs.

Scrutiny of a large case series led to subclassification of polyglandular syndromes in 1980, when Neufeld and Blizzard defined three types of APS based on clinical features[7] (Table 132.1). APS type 1 (APS-1) was characterized by the combination of Addison disease, hypoparathyroidism, and chronic mucocutaneous candidiasis and showed autosomal recessive inheritance. APS type 2 (APS-2) was coined for Addison disease combined with autoimmune thyroid disease and/or T1DM, whereas in APS type 3 (APS-3) autoimmune thyroid disease was the defining component in combination with other organ-specific insufficiencies, including T1DM, pernicious anemia, and alopecia. We have now come to appreciate that APS can be broadly separated into monogenic forms, such as APS-1, immune dysregulation polyendocrinopathy enteropathy X-linked (IPEX), and others to be described in this chapter, and APS-2, which includes the other combinations.

KEY POINTS

- Autoimmune polyglandular syndromes are defined as conditions with more than one endocrinopathy in an individual. These conditions tend to aggregate in families, pointing to strong heritability, and can be broadly separated into monogenic forms (e.g., autoimmune polyglandular syndrome type 1 and immune dysregulation polyendocrinopathy enteropathy X-linked) and those where several genes are involved (autoimmune polyglandular syndrome type 2).

THE BASIS OF AUTOIMMUNE DISEASE

This section gives a brief synopsis of the normal immune response, the mechanisms underlying tolerance to self, and the causes of autoimmunity; more comprehensive reviews can be found elsewhere.[8-10]

TABLE 132.1 Classification of the Main Autoimmune Polyendocrine Syndromes

	Autoimmune Polyglandular Syndrome Type 1	Autoimmune Polyglandular Syndrome Type 2	Immune Dysregulation Polyendocrinopathy Enteropathy X-Linked
Main components	Chronic mucocutaneous candidiasis, primary adrenal insufficiency, hypoparathyroidism	Primary adrenal insufficiency, autoimmune thyroid disease, T1DM	Autoimmune enteropathy neonatal T1DM, eczema
Other associated manifestations	Primary ovarian insufficiency, autoimmune thyroid disease, T1DM Gastritis, enteritis with malabsorption, hepatitis, pancreatitis, pneumonitis, nephritis Vitiligo, alopecia, nail dystrophy, enamel hypoplasia, keratitis, retinitis	Autoimmune gastritis, alopecia, vitiligo, celiac disease, primary ovarian insufficiency	Autoimmune thyroid disease, hemolytic anemia, thrombocytopenia, nephritis
Prevalence	1:100,000	1:100	1:1,000,000
Typical age of onset	Childhood, adolescence	Adolescence to adulthood	Infancy
Treatment	Replacement of hormones, antifungal therapy, immunosuppressive therapy for hepatitis, malabsorption, nephritis, pneumonitis, and keratitis	Replacement of hormones	Replacement of hormones, bone marrow transplantation
Complications and mortality	Adrenal and hypocalcemic crises, malignancies in mouth and esophagus	Adrenal crises, diabetic complications	Infections
Genes and inheritance	*AIRE*, autosomal recessive and dominant	Polygenic, MHC, and others	*FOXP3*, X-linked
Immune phenotype	Autoantibodies against interferon omega and alpha (>95%), and organ-specific intracellular proteins	Autoantibodies against 21-hydroxylase, GAD-65, IA-2, TSH receptor, and thyroid peroxidase	Autoantibodies against GAD-65, villin, hormonin, interferon omega, and nuclear receptors, lymphocytosis, eosinophilia, overproduction of cytokines, elevated IgE level

AIRE, Autoimmune regulator; *APS-1*, autoimmune polyglandular syndrome type 1; *APS-2*, autoimmune polyglandular syndrome type 2; *FOXP3*, forkhead box P3; *GAD*, glutamic acid decarboxylase-65; *IA-2*, protein tyrosine phosphatase-like protein islet antigen-512; *IPEX*, immune dysregulation polyendocrinopathy enteropathy X-linked; *MHC*, major histocompatibility complex; *T1DM,* type 1 diabetes mellitus; *TSH*, thyroid-stimulating hormone.

The Normal Immune Response

There are two broad types of immune response: innate and adaptive. The innate response is nonspecific and allows the body to respond immediately to a wide range of pathogens through pattern recognition receptors and complement activation. The adaptive response is directed against specific antigens on a pathogen by clonal expansion of T and B lymphocytes. This response is slower, but has the capacity for long-term memory of the antigen through the generation of memory lymphocytes. Although the adaptive response is primarily directed against pathogens, occasional errors may result in a response against self-antigens and cause autoimmune disease.

The first step in an adaptive immune response involves presentation of the antigen by an antigen-presenting cell (APC), such as a dendritic cell or macrophage, to a CD4$^+$ T helper (Th) cell, which recognizes the antigen by means of a T-cell receptor (TCR) with unique antigen specificity (Fig. 132.1). TCR diversity is generated by complex gene recombination from a large number of potential segments. APCs take up antigen by simple phagocytosis or via specific surface receptors and process them into peptides of eight to 12 amino acids; these antigenic epitopes bind to major histocompatibility complex (MHC) class II molecules expressed constitutively by the APC. The MHC (termed HLA in humans) is a complex of genes that have fundamental immunologic importance because they encode polymorphic molecules capable of binding a large range of T-cell epitopes, and the particular MHC class II alleles that are inherited determine which epitopes an APC can present to a T-cell.

Many other molecules are involved in antigen presentation and act as stabilizers, signal transducers, or second signals (Fig. 132.2). Interaction of a number of adhesion molecule ligands and receptors, such as intercellular adhesion molecule-1 (ICAM-1, CD54)/lymphocyte function-associated antigen-1 (LFA-1, CD11a/CD18) and LFA-3 (CD58)/CD2, allows initial binding of the T-cell to an APC. CD4 binds to a nonpolymorphic region on the MHC molecule to stabilize the complex formed with the TCR. CD3 is expressed by all T-cells and serves to initiate the intercellular events after TCR ligation. Finally, second signals determine whether the T-cell response proceeds, or is terminated because of the absence of an essential costimulator or the presence of an inhibitory signal. The best defined costimulatory pathway involves the binding of CD80 or CD86 on the APC with CD28 on the T-cell, which results in a stimulatory second signal. In contrast, TCR engagement (signal 1) without CD28 engagement (signal 2) results in T-cell anergy, a process in which the T-cell is inactivated by an antigen; it remains alive, but is no longer responsive. Cytotoxic T lymphocyte antigen type-4 (CTLA-4, CD152) is expressed by activated T-cells and can outcompete CD28 for binding CD80 or CD86, resulting in an inhibitory second signal and T-cell anergy. Another important regulatory checkpoint is via programmed death protein 1 (PD1, CD279) expressed on T-cells; engagement with the PD-ligand 1 (CD274) expressed by APCs and other cells also results in T-cell inactivation through anergy. Which of these signal pathways are engaged depends on the maturity of the T-cell and the type of APC involved, and this helps determine the exact outcome from antigen presentation.

Activated CD4$^+$ T-cells can follow one of two main functional pathways, Th1 or Th2, which are characterized by their pattern of cytokine production (Fig 132.1). Th1 cells mediate delayed-type hypersensitivity

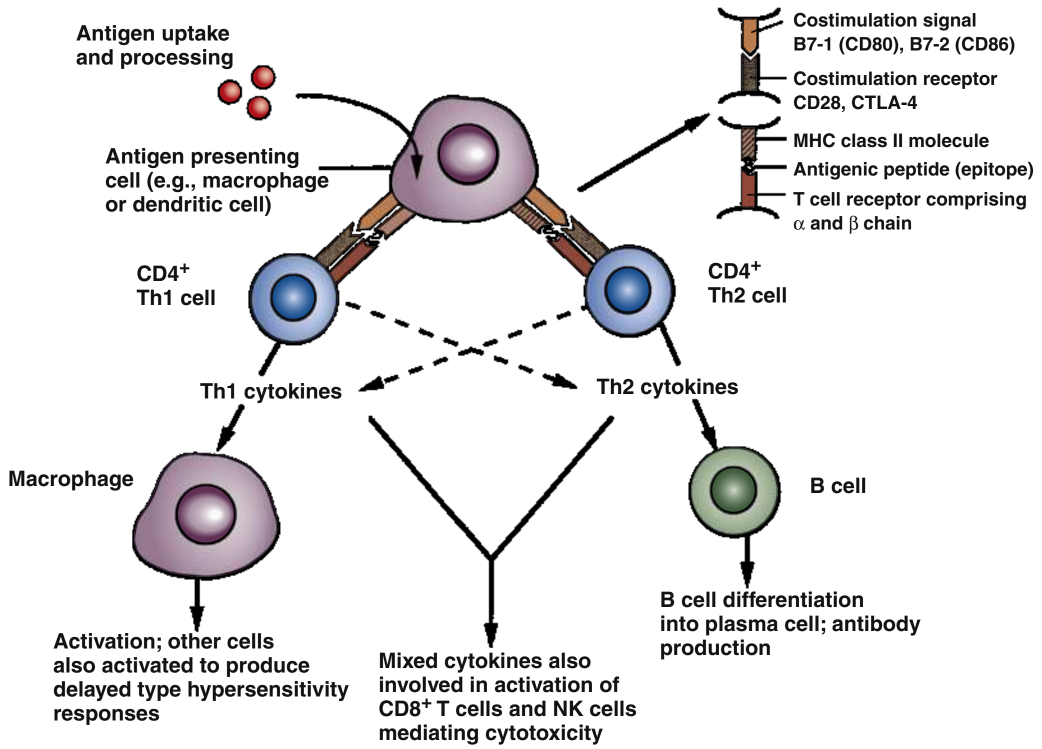

Antigen uptake and processing
Antigen presenting cell (e.g., macrophage or dendritic cell)
Costimulation signal B7-1 (CD80), B7-2 (CD86)
Costimulation receptor CD28, CTLA-4
MHC class II molecule
Antigenic peptide (epitope)
T cell receptor comprising α and β chain
CD4+ Th1 cell
CD4+ Th2 cell
Th1 cytokines
Th2 cytokines
Macrophage
B cell
Activation; other cells also activated to produce delayed type hypersensitivity responses
Mixed cytokines also involved in activation of CD8+ T cells and NK cells mediating cytotoxicity
B cell differentiation into plasma cell; antibody production

Fig. 132.1 **Simplified diagram of the key elements in a normal immune response.** The type of immune response which is induced is dependent on the cytokines produced by the T helper cell activated by antigen presentation. *MHC,* Major histocompatibility complex; *NK,* natural killer; *Th,* T helper. (From Weetman AP: Recent progress in thyroid autoimmunity: an overview for the clinician. Thyroid Today. 1996;19[2]:1–9.)

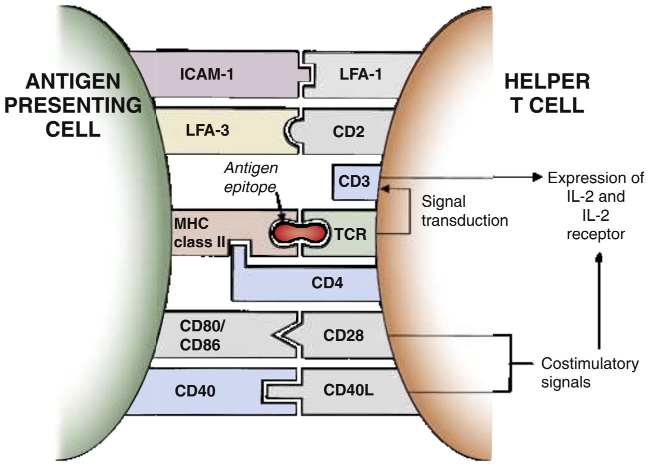

ANTIGEN PRESENTING CELL
HELPER T CELL
ICAM-1
LFA-1
LFA-3
CD2
Antigen epitope
CD3
MHC class II
TCR
CD4
Signal transduction
Expression of IL-2 and IL-2 receptor
CD80/ CD86
CD28
CD40
CD40L
Costimulatory signals

Fig. 132.2 **Diagram of the key molecular interactions in CD4⁺ helper T-cell activation by an antigen-presenting cell.** *MHC,* Major histocompatibility complex; *IL,* interleukin.

responses through their interaction with macrophages and cytotoxic CD8$^+$ T-cells. The latter are activated through TCR binding to epitopes synthesized within the target cell and presented by MHC class I molecules. CD8$^+$ T-cell–mediated cytotoxicity is typically directed against viral or malignancy-related antigens expressed by target cells, but can also be involved in the destruction of healthy cells in autoimmune diseases. Th2 cells promote antibody production by B-cells. Th1 and Th2 responses are mutually inhibitory through the action of their respective cytokines. TCR affinity and ligand density, the nature of the APC, and cytokines like interleukin (IL)-4 and IL-12 determine which pathway is followed. Although this Th1/Th2 paradigm has been useful, it is too simplistic, and many Th cells cannot be classified by such a simple dichotomy. For instance, the Th17 subset produces IL-17, which helps initiate and mediate inflammatory responses, including autoimmunity. IL-23 derived from dendritic cells and macrophages is critical to the activation of Th17 cells, as well as having additional proinflammatory effects of its own.[11] Another subpopulation of T-cells has an important role in controlling and terminating the immune response. The first such regulatory T-cell (Treg) subset to be described is generated in the thymus and is critical to prevention of autoimmunity. These "natural," or thymic, Tregs are CD4$^+$ CD25$^+$ and also express the transcription factor forkhead box P3 (Foxp3), as well as CTLA-4. A second, induced Treg subset differentiates in the periphery in response to specific antigen and is involved in mucosal immunity and also in controlling autoimmune responses. Other Treg populations have been reported, including a FoxP3$^-$ subset and a CD8$^+$ subset.[9]

B-cells produce antibodies after differentiation into plasma cells, a process that is regulated by Th2 cytokines and CD40/CD40 ligand interaction with a helper T-cell. In the secondary lymphoid organs, the follicular helper subset of CD4$^+$ T-cells primes B-cells, and this pathway is critical to germinal center formation and antibody affinity maturation. Up to 10^7 different antibody specificities can be generated in humans by recombination and somatic mutation of the immunoglobulin genes, an even greater diversity than that seen in TCRs. During an immune response, these processes ensure selection of the most appropriate antibodies so that IgG molecules with the highest affinity for antigen are produced. Antibodies generally recognize conformational rather than linear epitopes on intact antigen. As well as antibody production, B-cells are important APCs because they are able to take up specific antigens via surface-bound immunoglobulin and enter into cognate recognition of appropriate T-cells. This type of antigen presentation may be important in diversifying the T-cell response and may also help prevent autoimmunity, as an autoantibody-producing B-cell will not become activated unless there also an autoantigen-specific Th cell present.

Natural killer cells are the third type of lymphocyte; they do not express TCRs or surface immunoglobulins, and their function is to destroy cells with altered surface antigens, for example, tumor cells. Through expression of CD16, a receptor for the Fc region of immunoglobulins, natural killer cells can also kill target cells that have antibody bound to cell surface antigens, a process called antibody-dependent cell-mediated cytotoxicity. Such cytotoxicity can be an effector mechanisms in the destruction of autoantibody-coated tissue cells in autoimmunity.

Self-Tolerance

The adaptive immune system provides a specific way of responding to a pathogen that is retained through memory T- and B-cells to allow a rapid response in future to the same pathogen. At the same time, the immune system must not respond to the body's own autoantigens, which is achieved through a series of self-tolerance mechanisms discussed later. The trade-off for the immune system's astonishing ability to respond to a universe of exogenous antigens is that these self-tolerance mechanisms may occasionally fail, resulting in autoimmune disease.

Central Tolerance

During T-cell maturation, the thymus is responsible for positively selecting T-cells to constitute the immune repertoire and negatively selecting and removing autoreactive T-cells.[12] Positive selection depends on immature T-cell recognition of MHC molecules expressed by thymic cortical epithelial cells. The T-cells then become either CD4$^+$ or CD8$^+$, depending on which MHC class they bind most strongly to, whereas failure to bind leads to apoptosis. As the T-cells mature, they move to the thymic medulla, where they encounter medullary epithelial cells and dendritic cells that, remarkably, can express and present an array of self-antigens that are otherwise tissue-specific. High-affinity TCR engagement at this stage results in apoptosis, and hence the elimination of self-reactive T-cells; self-reactive T-cells with lower TCR affinity differentiate into natural Treg cells. This process cannot operate for autoantigens, which are exclusively expressed in the periphery or appear later in development, nor is it always successful. Thus, the self-reactive T-cell repertoire is thinned out rather than eliminated. Thymic expression of a cluster of self-antigens is controlled by the transcription factor autoimmune regulator (AIRE). The critical part played by AIRE in self-tolerance is demonstrated by the effect of mutations in the *AIRE* gene, which result in APS-1, as discussed in detail later.

Peripheral Tolerance

Four broad mechanisms, classed as ignorance, deletion, anergy, and suppression, ensure that the self-reactive T-cells that escape central tolerance are held in check to prevent autoimmunity.[13] The relative importance of these mechanisms varies according to self-antigen.

1. In immunological ignorance, T-cells reactive to certain self-antigens are normally harmless because they cannot physically come into contact with a sequestered autoantigen, or because localized expression of Fas ligand induces apoptosis in potentially auto-aggressive lymphocytes entering the site (also called immunologic privilege). Such Fas ligand expression is seen in Sertoli cells, corneal cells, and the placental trophoblast.
2. Immature dendritic cells can take up self-antigens and present these to naïve T-cells in lymph nodes, resulting in either their deletion or conversion to Tregs. Lymph node stromal cells may play a similar regulatory role.
3. Anergy can occur when self-antigen presentation occurs without a costimulatory second signal. Aberrant expression of MHC class II molecules in the absence of costimulatory expression occurs in tissue cells in response to cytokines released during an inflammatory response, thereby preventing the initiation of an autoreactive T-cell response. Anergy mediated through the activation of CTLA-4 or PD1 pathways is also important, as shown by the development of autoimmunity in some patients with cancer who are treated with inhibitors of these immune checkpoints (see later).
4. Active T-cell-mediated suppression of self-reactivity was first clearly demonstrated in animal studies: autoimmune thyroiditis or diabetes can be induced in some strains of mice and rats by experimental T-cell depletion and can be prevented by the transfer of T-cells from healthy animals. Characterization of the T-cell subset responsible for suppressing these autoimmune responses led to the discovery of Tregs. The importance of these cells in human self-tolerance is shown by loss-of-function mutations in the *FOXP3* gene expressed by Tregs, which results in multiple autoimmune diseases called IPEX, discussed later. Tregs exert their suppressive effects through several pathways, including cytokines (especially transforming growth factor-β, IL-10, and IL-35), cytolysis, metabolic effects, and modulation of dendritic cells.

This overview has focused on T-cell tolerance, as these cells lie at the heart of the immune response and have been the most studied in the context of autoimmunity, but B-cell self-tolerance is also important. The B-cell repertoire is controlled by both central tolerance during development in the bone marrow and peripheral tolerance in the lymph nodes, involving both deletion and anergy. Self-reactive B-cells that evade tolerance will not initiate an autoreactive response unless the tolerance of corresponding self-reactive CD4$^+$ T-cells fails. Defects in B-cell tolerance may underlie some of the features of autoimmune diseases like systemic lupus erythematosus (SLE).

KEY POINTS

- Autoimmune disease arises when the normal mechanisms to maintain tolerance to self-antigens fail. This failure may be either at the level of central (or thymic) tolerance or peripheral tolerance, in which a combination of T-cell ignorance, deletion, anergy, and suppression normally ensures that any autoreactive T-cells that escape central tolerance are held in check to prevent autoimmunity.

Pathogenesis of Autoimmune Disease

There are over 100 autoimmune diseases that together affect 5% to 10% of the population, and each one is the result of a breakdown of in normal self-tolerance. Loss of tolerance is complex, with a wide variety of both genetic and environmental factors impinging adversely on one or more of the tolerance mechanisms discussed earlier. Some autoimmune diseases show a rising incidence, which is thought to be the result of changes in environmental factors. Many of the genes that are associated with susceptibility to autoimmunity control antigen presentation (MHC genes) or immunoregulatory responses, although some genes are tissue-specific and determine the site of autoimmune disease, for instance through altering expression of specific autoantigens. Around two thirds of the susceptibility loci so far identified are common to multiple autoimmune diseases.[14] The clustering of diseases in APS-2, discussed later, is primarily the result of this sharing of genetic predisposition. Genetic and environmental factors also act together with the existential factors of age, sex, and parity. Growing older results in lowered T- and B-cell production and function, as well as an increased chance of encountering etiological environmental factors. Immune responses including autoimmunity are generally stronger in women than men because of the effects of sex hormones and immune response genes on the X chromosome, and pregnancy causes a variety of altered immune responses to tolerate the fetal semiallograft. As a result, most autoimmune diseases typically occur in adulthood and are more common in women. Particular genetic and environmental factors explain exceptions to this, such as T1DM.

A feature of most autoimmune diseases is a long prodromal phase before disease becomes evident, and there may be intermittent periods of spontaneous remission, indicating that loss of self-tolerance is not absolute, at least initially. This prodromal phase can often be identified by the presence of circulating autoantibodies in the absence of clinical features, although there may be biochemical evidence of tissue injury, for example, elevation in thyroid-stimulating hormone (TSH) levels in subclinical autoimmune hypothyroidism despite normal thyroid hormone levels. The possibility of an initial phase of reversibility has led to attempts to improve prediction of those at risk and the development of new treatments for some severe autoimmune diseases using monoclonal antibodies targeting key cytokines and cells, aimed at tipping the balance back to normal self-tolerance.

Most autoimmune diseases are caused by tissue-specific destruction, using the same pathways that are involved in the normal immune response to a pathogen. Once an autoimmune response is initiated, aberrant expression of chemokines and adhesion molecules help localize autoreactive lymphocytes to the target organ, and tissue injury occurs through T-cell–mediated cytotoxicity and cytokine release, ultimately leading to atrophy and fibrosis. In multisystem diseases like SLE, immune complex deposition with complement fixation is a particularly important pathway of autoimmune damage. Autoantibody production may also contribute to tissue injury through complement activation or antibody-dependent cell-mediated cytotoxicity, although the absence of tissue injury in neonates born to women with autoantibodies that are placentally transferred indicates that any such role lies in exacerbating rather than initiating an autoimmune disease. In normal circumstances many autoantigens are inaccessible to autoantibodies; such sequestration only disappears once inflammation begins. The exception is a small group of primarily autoantibody-mediated diseases like Graves disease and myasthenia gravis, in which the functional effects of the antibody on an accessible receptor (the TSH receptor or the acetylcholine receptor, respectively) can be seen in neonates born to affected mothers as well.

KEY POINTS

- Genetic and environmental factors, together with aging, sex, and parity, contribute to the loss of self-tolerance in autoimmunity. Sharing of genetic factors involved in susceptibility to such diseases leads to clustering within an individual and within families. In most autoimmune disorders there is a long prodromal phase before disease becomes evident, which can often be identified by the presence of circulating autoantibodies in the absence of clinical features and sometimes by biochemical evidence of subclinical tissue injury.

AUTOIMMUNE POLYGLANDULAR SYNDROME TYPE 1

Definition

APS-1, also called autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED) (OMIM 240300), is a monogenic autosomal recessive disease characterized by early-onset organ-specific autoimmunity.[15,16] Traditionally the diagnosis was made when two of the three main components—primary adrenocortical insufficiency, hypoparathyroidism, and chronic mucocutaneous candidiasis—were present (diagnostic dyad), or one component was present and there was a sibling with APS-1. After identification of the causative gene, *AIRE*,[17,18] the diagnosis can be made by uncovering pathogenic variants. Using sequencing in diagnostics has revealed that the phenotypic variation is much wider than earlier appreciated.[19]

Inheritance is usually autosomal recessive, requiring two disease-causing variants in the gene. However, autosomal dominant inheritance has been reported.[20,21] In these cases, one mutation is sufficient to cause the syndrome, often with later onset and milder phenotype without the diagnostic dyad.

Pathogenesis

AIRE was named before its function had been studied, but the name turned out to be fitting. AIRE is mainly expressed in thymic medullary epithelial cells and has two major functions. It enhances the expression of tissue-specific proteins, and so displays such peptides on MHC class II for developing T-cells. T-cells that recognize self-peptides will either undergo apoptosis or become Tregs that have an important role in harnessing autoreactivity.[22] When AIRE is missing or functions abnormally because of a mutation, autoreactive T-cells can escape into the periphery and cause autoimmunity later in life. AIRE may also have a similar role in lymph nodes,[23] creating another layer of immunological tolerance.

Currently more than 130 different mutations are known, spread throughout the gene[24] (Fig. 132.3). The most common is the so-called Finnish major mutation, a nonsense variant p.R258X (c.769C>T) leading to premature mRNA decay and no protein expression. It is the most common mutation worldwide and is especially prevalent in the Finnish, Russian, and Polish populations. The 13-bp deletion in exon 8 (p.L323SfxX51, c.967–979del13) is more common in Norway, the British Isles, and North America. Founder mutations occur among Persian Jews, Sardinians, and Sicilians. Most patients have two disease-causing mutations, and carriers with one mutation do not have excessive autoimmunity as far as we know. However, some mutations can cause APS-1 in heterozygosity by a dominant negative effect, causing the disease to be transmitted in a dominant fashion.[20,21] Intriguingly, these mutation are located in certain domains of AIRE, particularly in the first PHD-zinc finger.

Mutation carriers develop a number of autoantibodies that can be broadly divided into cytokine- and organ-specific autoantibodies. Autoantibodies against type 1 interferons including interferon-omega and several interferon-alpha species are present in almost all patients and develop very early in life, even before 1 year of age.[25,26] Thus, detection of such antibodies is an effective screening tool to diagnose patients. The majority of patients also display autoantibodies against IL-22, IL-17A, and IL-17F.[27,28] A correlation between these autoantibodies and chronic mucocutaneous candidiasis has been reported, but is uncertain because most patients have had periods with chronic candidiasis, and the frequency of anti-IL-22 is more than 90%. Another group of affected enzymes are neurotransmitter-synthesizing enzymes such as glutamic acid decarboxylase-65, tyrosine hydroxylase, tryptophan hydroxylase, and aromatic L-amino acid decarboxylase (DOPA decarboxylase), which are associated with alopecia, enteritis, and hepatitis, respectively (Table 132.2). NACHT leucine-rich-repeat protein 5 (NALP5) is highly expressed in parathyroid cells and is associated with hypoparathyroidism.[29] The mouse ortholog, called maternal antigen that embryos require (MATER) is also expressed in ovaries and is associated with premature ovarian insufficiency. KCNRG, a potassium

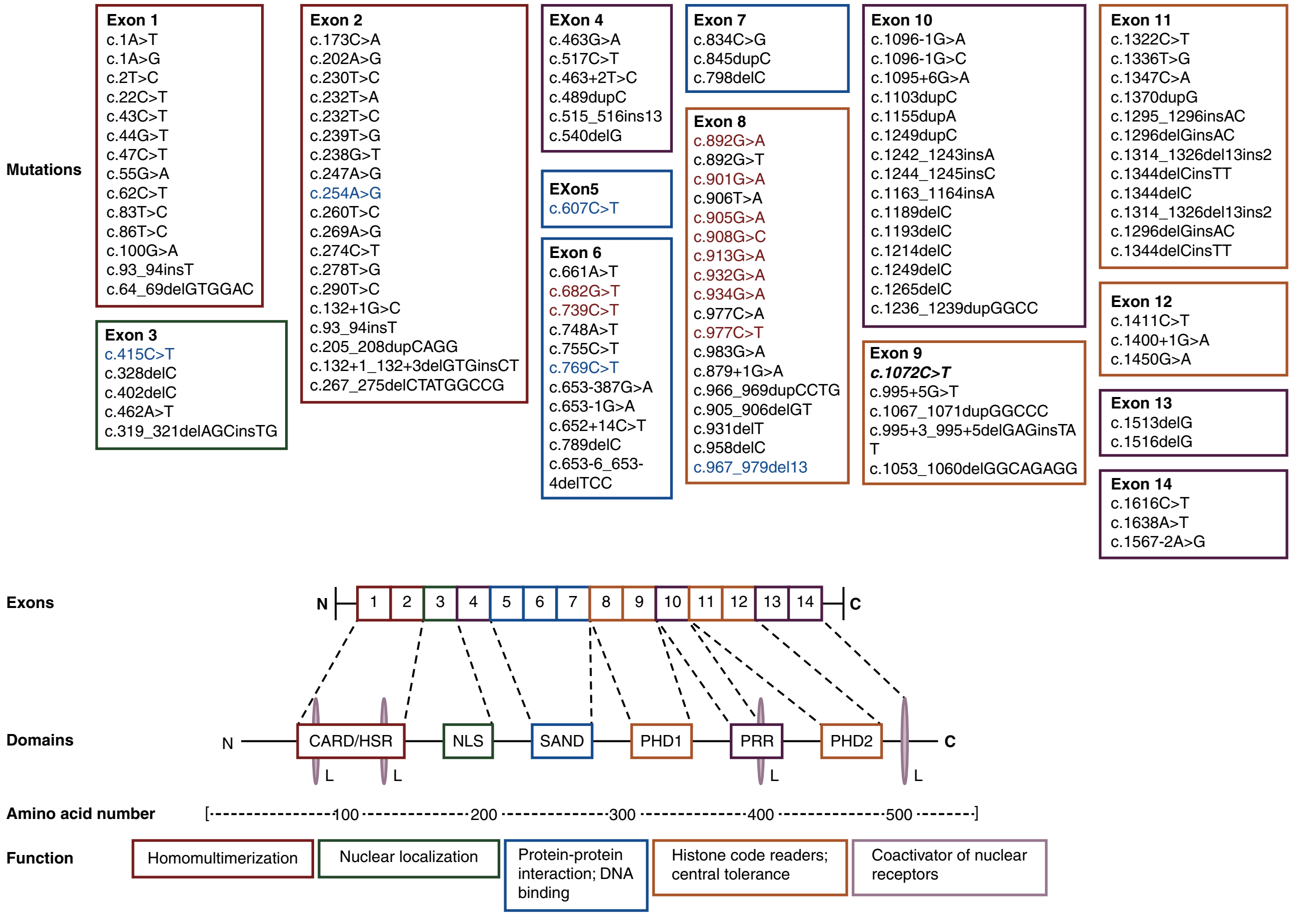

Mutations
Exon 1
c.1A>T
c.1A>G
c.2T>C
c.22C>T
c.43C>T
c.44G>T
c.47C>T
c.55G>A
c.62C>T
c.83T>C
c.86T>C
c.100G>A
c.93_94insT
c.64_69delGTGGAC
Exon 3
c.415C>T
c.328delC
c.402delC
c.462A>T
c.319_321delAGCinsTG
Exon 2
c.173C>A
c.202A>G
c.230T>C
c.232T>A
c.232T>C
c.239T>G
c.238G>T
c.247A>G
c.254A>G
c.260T>C
c.269A>G
c.274C>T
c.278T>G
c.290T>C
c.132+1G>C
c.93_94insT
c.205_208dupCAGG
c.132+1_132+3delGTGinsCT
c.267_275delCTATGGCCG
EXon 4
c.463G>A
c.517C>T
c.463+2T>C
c.489dupC
c.515_516ins13
c.540delG
EXon5
c.607C>T
Exon 6
c.661A>T
c.682G>T
c.739C>T
c.748A>T
c.755C>T
c.769C>T
c.653-387G>A
c.653-1G>A
c.652+14C>T
c.789delC
c.653-6_653-4delTCC
Exon 7
c.834C>G
c.845dupC
c.798delC
Exon 8
c.892G>A
c.892G>T
c.901G>A
c.906T>A
c.905G>A
c.908G>C
c.913G>A
c.932G>A
c.934G>A
c.977C>A
c.977C>T
c.983G>A
c.879+1G>A
c.966_969dupCCTG
c.905_906delGT
c.931delT
c.958delC
c.967_979del13
Exon 10
c.1096-1G>A
c.1096-1G>C
c.1095+6G>A
c.1103dupC
c.1155dupA
c.1249dupC
c.1242_1243insA
c.1244_1245insC
c.1163_1164insA
c.1189delC
c.1193delC
c.1214delC
c.1249delC
c.1265delC
c.1236_1239dupGGCC
Exon 9
c.1072C>T
c.995+5G>T
c.1067_1071dupGGCCC
c.995+3_995+5delGAGinsTAT
c.1053_1060delGGCAGAGG
Exon 11
c.1322C>T
c.1336T>G
c.1347C>A
c.1370dupG
c.1295_1296insAC
c.1296delGinsAC
c.1314_1326del13ins2
c.1344delCinsTT
c.1344delC
c.1314_1326del13ins2
c.1296delGinsAC
c.1344delCinsTT
Exon 12
c.1411C>T
c.1400+1G>A
c.1450G>A
Exon 13
c.1513delG
c.1516delG
Exon 14
c.1616C>T
c.1638A>T
c.1567-2A>G
Exons
N
1
2
3
4
5
6
7
8
9
10
11
12
13
14
C
Domains
N
CARD/HSR
NLS
SAND
PHD1
PRR
PHD2
C
L
L
L
L
Amino acid number
100
200
300
400
500
Function
Homomultimerization
Nuclear localization
Protein-protein interaction; DNA binding
Histone code readers; central tolerance
Coactivator of nuclear receptors

Fig. 132.3 The autoimmune regulator (AIRE) protein with its functional domains and reported mutations. AIRE harbors four major subdomains: the homogeneously staining region or caspase recruitment domain/homodimerisation domain (CARD/HSR, amino acids 1–105); the SAND domain (for Sp100, AIRE-1, NucP41/75 or NucP41/75, DEAF-1, amino acids 181–280); and two plant homeodomain (PHD)–type zinc fingers (amino acids 296–343 and 434–475). In addition, the AIRE protein contains four LXXLL domains that are found on coactivators of nuclear receptors (amino acids 7–11, 63–67, 414–418, 516–520) and a nuclear localization signal (amino acids 100–189). AIRE acts in a multiprotein complex and was early recognized as a transcription factor, as it localizes to the cell nucleus. The functions of the different domains of AIRE are shown in the boxes. Founder and common mutations are shown in blue, including c.254A>G (Persian Jewish), c.415C>T (Sicilian), c.607C>T (Sardinian), c.769C>T (Finnish), and c.967-979del13 (Norwegian, British, North American). Dominant negative mutations are shown in red. Gross deletions and gross insertions are not included. (Slightly modified from Bruserud O, Oftedal BE, Wolff AB, et al. AIRE-mutations and autoimmune disease. *Curr Opin Immunol.* 2016;43:8–15.)

channel regulator expressed in the terminal alveoli, is associated with bronchiolitis,[30] as is BPIFB1, a protein involved in innate immunity in the airways.[31] Another member of this protein family, BPIFA2, was recently found to be associated with dry mouth and "Sjøgren-like" symptoms in patients with APS-1.[32] Reactivity to surface receptors is also seen, such as the calcium-sensing receptor, which can result in hypoparathyroidism.[33]

KEY POINTS

- Autoimmune polyglandular syndrome type 1 is defined by the presence of two of the three manifestations—adrenal insufficiency, chronic candidiasis, and hypoparathyroidism—or the presence of a disease-causing variant in the autoimmune regulator gene. Interferon type I autoantibodies are a highly specific diagnostic biomarker. Inheritance is autosomal recessive or dominant; the latter forms are generally milder, with similarities to autoimmune polyglandular syndrome type 2.

Clinical Features

Typically, APS-1 manifests itself in childhood with chronic mucocutaneous candidiasis followed by hypoparathyroidism and adrenal insufficiency.[34] However, in recent years we have come to appreciate that the presentation is highly variable, even within the same family.[19,35,36] A major explanation is that screening for interferon autoantibodies and the development of next-generation sequencing have allowed diagnosis of patients with atypical disease, patients with late debut disease, and sometimes patients who are lacking the diagnostic dyad.

The prevalence of APS-1 is approximately 1:100,000 in most countries, and epidemiological surveys are scarce. In some populations with founder mutations, higher prevalence is observed, including Finns (1:25,000), Sardinians (1:14,000), and Persian Jews (1:9000). The natural history is highly variable, with between one to over 20 different components in an individual patient. Some features, such as chronic diarrhea, keratitis, autoimmune hepatitis, periodic rash with fever, or

TABLE 132.2 Characteristics of Autoimmune Polyendocrine Syndrome Type 1

Components	Clinical Features	Diagnosis	Key Autoantigens
Chronic mucocutaneous candidiasis	Chronic recurrent candidiasis of mucous membranes, skin and nails	Detection of *Candida* in culture	Interleukin-22
Primary adrenal insufficiency	Hyperpigmentation, salt craving, hyponatremia, hyperkalemia	Low cortisol, elevated ACTH, failed synacthen testing	21-hydroxylase
Primary ovarian insufficiency	Early menopause (<40 years of age), primary amenorrhea	Elevated FSH and LH, low estrogen	Side-chain cleavage enzyme, 17-hydroxylase, NALP-5
Prostatitis	Perineal pain, discharge	Clinical signs, imaging	Transglutaminase-4
Autoimmune thyroid disease	Hypothyroidism (Graves disease extremely rare)	Elevated TSH, low free thyroxine	Thyroid peroxidase
Type 1 diabetes mellitus	Hyperglycemia, lack of insulin	Fasting serum glucose >7 mmol/L, HbA1c >48 mmol/mol	IA-2
Hypophysitis	Secondary adrenal insufficiency and hypothyroidism; hypogonadotropic hypogonadism, growth failure	Decreased pituitary hormones	Tudor domain-containing protein 6
Gastritis	Achlorhydria, vitamin B12 deficiency, iron deficiency	Low vitamin B_{12}, abnormal FBC and iron levels	Intrinsic factor and parietal cell (H^+/K^+-ATPase)
Enteritis with malabsorption	Chronic steatorrhea, obstipation	Increased fat in stools	Tryptophan hydroxylase
Autoimmune hepatitis	Jaundice	Abnormal liver function tests	CYP1A2 and AADC
Exocrine pancreatitis	Chronic steatorrhea, malabsorption	Decreased fecal elastase	
Pneumonitis	Dyspnea	Peribronchiolar infiltrates on computed tomography, respiratory failure	KCNRG and BPIFB2
Nephritis	Reduced kidney function, proteinuria	Interstitial nephritis	Aquaporin-2, HOXB7, NFAT5
Vitiligo	Ranging from patchy to total nonsegmental distribution	Clinical signs	Tyrosinase, tyrosinase-related proteins, AADC, transcription factors (SOX9, SOX10)
Alopecia	Ranging from patchy to total or universal distribution	Clinical signs	Tyrosine hydroxylase
Enamel dysplasia	Ranging from small pits on some teeth to lack of enamel	Clinical signs	Protein phosphatase 4
Retinitis	Reduced vision	Ophthalmological signs of retinitis	Interphotoreceptor retinoid-binding protein
Keratitis	Extreme light sensitivity	Clinical signs	
Nail dystrophy		Clinical signs	
Skeletal metaplasia		Clinical signs	

AADC, Aromatic L-amino acid decarboxylase; *ACTH,* adrenocorticotropic hormone; *CYP1A2*, cytochrome P450 family 1 subfamily A member 2; *BPIFB2*, bactericidal/permeability-increasing fold-containing B1FSH protein, follicle-stimulating hormone; *FBC,* full blood count; *FSH,* follicle-stimulating hormone; *HbA1c* , glycated hemoglobin; *HOXB7*, human homeobox B7 transcription factor; *IA-2,* protein tyrosine phosphatase-like protein islet antigen-512; *KCNRG*, putative potassium channel regulator; *LH*, luteinizing hormone; *MHC*, major histocompatibility complex; *NALP-5*, NACHT, LRR and PYD domains-containing protein 5; *NFAT5*, nuclear factor of activated T-cell 5 transcription, *SOX*, SRY-related HMG-box; *TSH,* thyroid-stimulating hormone.

severe constipation, may dominate the initial picture before any of the major components develop.[36] Thus, a low threshold for autoantibody testing or sequencing of *AIRE* is recommended.

Chronic Mucocutaneous Candidiasis

Chronic superficial infection with *Candida albicans* is usually the first of the major components of APS-1 to appear.[34] It affects mucous membranes, skin, and nails, but never causes deep, systemic *Candida* infections. In mild forms only angular cheilitis is seen, but in more severe cases the whole mouth is involved, with white or grey plaques of yeast and hyperkeratosis. In the atrophic form the mucosa is erythematous with leukoplakic areas or nodules that may develop into squamous cell carcinomas; this is a significant cause of death, especially if the patient smokes and overuses alcohol.[37,38] Candidiasis may also only be confined to the esophagus, causing problems with swallowing caused by overgrowth or strictures, sometimes requiring dilation. A reddish and watery rash may affect the skin of the hands and face. The fingernails may be affected and females may suffer from genital candidiasis.

The diagnosis is based on the signs and symptoms, as well as culture of *C. albicans*. It is very important to treat oral candidiasis to prevent cancer development, which includes practicing good oral hygiene and refraining from smoking and excessive alcohol. This is also important because the majority of patients develop enamel hypoplasia of the permanent teeth,[19] and many have a mild Sjøgren-like syndrome with reduced saliva production.[32] Infrequent bouts of *Candida* can be effectively treated with azole preparations such as fluconazole and ketoconazole, but with regular use there is a risk of drug resistance. Thus, for recurrent and chronic cases, treatment with nystatin and amphoterizin B in combination is an alternative, with pulsed prophylaxis with either of these drugs; alternatively, chlorhexidine mouthwash can be applied. A biopsy should be taken if mucositis with an ulceration fails to respond to treatment within 2 weeks. Angular cheilitis can be treated with azole, chlorhexidine, or amorolfine creams. *Candida* esophagitis usually heals with the discussed treatment regimen in 1 to 2 weeks. If not, endoscopy should be considered. Female genital candidal infection usually responds to a short course of vaginal fluconazole.

Hypoparathyroidism

Hypoparathyroidism is the second most common major manifestation and can be the first endocrine component. The combination of hypocalcaemia, hyperphosphatemia with normal or low parathyroid hormone, and normal plasma creatinine is diagnostic. Symptoms of hypocalcaemia may be vague in the beginning, with muscle cramps during infections, intermittent mild paraesthesia, and clumsiness. Frank hypocalcemia, sometimes with seizures, can be triggered by a febrile illness and fasting. APS-1 should be considered in every patient with acquired hypoparathyroidism because it may be the only component in some patients. Half of patients harbor autoantibodies against NALP5,[29] and some also have autoantibodies against the calcium-sensing receptor.[33]

Hypoparathyroidism is treated with vitamin D derivatives, calcium, and magnesium. Alphacalcidol ($T_{½}$ 2 days) and calcitriol ($T_{½}$ 1 day) are recommended because of their relatively short half-lives. Calcium supplementation in divided doses of 100 to 500 mg elementary calcium outside meals is suggested, preferably calcium citrate, as it helps calcium stay in solution in the urine. Likewise, a daily supplement of magnesium and cholecalciferol should be given to avoid deficiencies and to cover other functions of vitamin D. Therapy is aimed at maintaining plasma calcium in the lower half or slightly below the normal range (total 2.10–2.30 mmol/L, ionized 1.05–1.15 mmol/L), urinary calcium in the normal range, and plasma magnesium in the upper half of its normal range (0.85–1.00 mmol/L). Hypercalcemia may damage the kidneys, particularly if the patient is not well hydrated. Serum calcium should be checked quarterly and when symptoms indicate hypo- or hypercalcemia.

Patients with hypoparathyroidism and adrenal insufficiency may develop hypercalcemia as a result of sodium deficiency and/or dehydration. Increase in glucocorticoid dose tends to cause hypocalcaemia, and vice versa. Increased calcium intake or additional alphacalcidol or calcitriol may be required if glucocorticoid doses remain high for more than 24 hours.

Adrenal Insufficiency

Adrenocortical insufficiency most commonly appears after mucocutaneous candidiasis and hypoparathyroidism, typically between 5 and 15 years of age, but may appear later. The prevalence in different published cohorts is around 70%.[19,37] The characteristic symptoms are fatigue, salt craving, hypotension, weight loss, and an increased pigmentation of the skin and mucous membranes. Biochemically, symptomatic patients display low basal plasma cortisol, elevated adrenocorticotropic hormone (ACTH), low aldosterone, and elevated plasma renin activity. Hyponatremia is common, sometimes in combination with hyperkalemia. Of note, hypocortisolism and hypoaldosteronism may appear dissociated years apart. Just like in the more common form of autoimmune Addison disease, most patients display autoantibodies against 21-hydroxylase.[6,39] In addition, many APS-1 patients also have antibodies against side-chain cleavage enzyme.[39] Autoantibodies can be present years before biochemical or clinical evidence of adrenal insufficiency and call for intensified follow-up of adrenal function. Adrenal insufficiency should be diagnosed before a potentially fatal adrenal crisis develops.

Replacement therapy with hydrocortisone and fludrocortisone should be individualized to provide optimal well-being and safety in the patient's life situation, including incidents of sickness and stress. All patients should be provided with a steroid acute card; a common European card is available for adults and children. The patient and her/his close family should be equipped with syringes for self-injection of hydrocortisone. Glucocorticoid treatment influences calcium and glucose levels. Thus, treatment of hypoparathyroidism and diabetes mellitus may have to be adjusted. For details, see Chapters 39 and 52.

Gonadal Insufficiency

Primary ovarian insufficiency manifests particularly early, and some patients have primary amenorrhea, with either a complete failure of, or arrested, pubertal development. Others develop premature menopause, with a median age of 16 years.[40] Adrenal insufficiency was more common in APS-1 patients with ovarian insufficiency then those without.[40] The biochemical findings are typically elevated plasma follicle-stimulating hormone and luteinizing hormone and low estrogen levels. Ovarian insufficiency is associated with presence of autoantibodies against side-chain cleavage enzyme and NALP5 and are typically present before clinical symptoms develop.[39] Estrogen replacement should be started at pubertal age, with gradually increasing doses to maintain a normal pace of female development. Periodic gestagen should be added at the appropriate stage. There are reports of cases where pregnancies have occurred even several years after menopause, but patients should be encouraged not to delay plans if they wish to have a child. Storing of ovarian tissue may be an option for future pregnancies.

Testicular insufficiency does occur, albeit more seldom and mostly at an adult age, with a maximum prevalence of approximately 25%, as seen in the Finnish series.[39] The blood-testis barrier may protect the Leydig cells from an autoimmune attack, but lack of *AIRE* may affect fertility by disrupting scheduled apoptosis of testicular germ cells.[41] Recently, male APS-1 patients were found to harbor autoantibodies against transglutaminase-4.[42] *Aire* knockout mouse models display severe prostatitis; but so far there are no reports of autoimmune prostatitis in APS-1.

Other Endocrinopathies

The prevalence of diabetes mellitus varies between different populations, from a few percent to one-third, being more common in countries and populations with high prevalence of T1DM. The presence of autoantibodies against the IA-2 tyrosine phosphatase-like protein, but not glutamic acid decarboxylase (GAD), is predictive for the development of diabetes.[39] Thyroid autoantibodies are common, and hypothyroidism is relatively common after puberty, reaching a prevalence of one third by middle age.[37] Graves disease is very rare. Autoimmune hypophysitis is seen with lack of different pituitary hormones, including growth hormone, which may need replacement to ensue normal growth.[35]

Gastrointestinal Manifestations

Autoimmune gastritis with pernicious anemia is a common entity that affects approximately 30% by middle age,[37] associated with autoantibodies to parietal cells and/or blocking intrinsic factor. Vitamin B12 and iron status should be checked, and replacement given if indicated. A definite diagnosis can be made by gastroduodenoscopy and biopsy. Celiac disease has been reported in a few patients[43] but is uncommon. Conversely, autoimmune enteropathy is common and manifests with chronic or recurrent diarrhea and steatorrhea, as well as periodic or chronic constipation. Absence of enterochromaffin cells is typical, and patients have autoantibodies against tryptophan hydroxylase, the rate-limiting enzyme in serotonin biosynthesis.[44] Aberration in serotonin levels could explain changes in intestinal motility. Paneth cells are also affected, and levels of defensins are decreased,[45] perhaps explaining the alterations in the microbiome observed in these patients.[46] Exocrine pancreatitis occurs in a small percentage of patients, usually associated with insulin deficiency. Typically, subnormal levels of secretin-stimulated lipase are found. Replacement of pancreatic enzymes may improve diarrhea, and immunosuppression has been reported to improve exocrine pancreatic function.[47,48] Malabsorption can seriously disturb replacement therapy with steroids, vitamin D preparations, and calcium, and requires close follow-up.

Autoimmune hepatitis is a serious complication that is reported to affect around one fifth of the patients and usually manifests in childhood.[37] After careful scrutiny, a recent survey from the National Institutes of Health revealed that approximately 40% of patients are affected by autoimmune hepatitis.[49] Severity ranges from slightly elevated plasma alanine amino transferase levels to fulminant hepatitis. High levels should prompt liver biopsy. Autoantibodies against CYP1A2 correlate with the presence and severity of the autoimmune hepatitis.[35] CYP2D6, the autoantibody marker for isolated autoimmune hepatitis type 2, is not seen in APS-1.[50] If a biopsy reveals active hepatitis, immunosuppressive therapy with steroids, azathioprine, or other immunosuppressants is indicated.[16,49]

Ectodermal Manifestations

There are several ocular manifestations associated with APS-1, such as chronic keratitis, dry eye, ptosis, cataract, iridocyclitis, retinal detachment, and optic atrophy. Keratitis appears early and can be the first manifestation, and can lead to blindness if not treated appropriately. Symptoms include intense photophobia, blepharospasm, and reduced lacrimation. The etiology is presumably autoimmune, and treatment consists of immunosuppression with local glucocorticoids, or topical cyclosporin A. Sometimes corneal transplantation is warranted. Retinitis is rarer and linked to the presence of autoantibodies to interphotoreceptor retinoid-binding protein.[51]

Alopecia is the most frequent skin manifestation and is usually more severe than in isolated cases. Approximately 40% are affected by middle age. Loss of all hair on the head (alopecia totalis) or even the whole body (alopecia universalis) is not uncommon. It can sometimes be reversible, especially when immunosuppressive therapy is given for other components. Alopecia correlates with the presence of autoantibodies against tyrosine hydroxylase, which is highly expressed in hair follicles.[52]

Vitiligo is also common, and its prevalence approached one-third of the Finnish cohort.[37] It is of variable extent and distribution, but tends to include most of the skin. The contrast to surrounding normal skin may fade, yet the patients often feel it to be disfiguring on the face and hands. Vitiligo correlates with autoantibodies against transcription factors SOX9 and SOX10.[53] Periods with maculopapular, morbilliform, or urticarial rash with fever come early.[36,37] Skin biopsies have reveal lymphoplasmacytic vasculitis.[43] Punctate nail dystrophy is common, without evidence of *Candida* infections.[34]

Enamel hypoplasia of permanent teeth is one of the most common manifestations of APS-1 and is present in 70% to 80% of patients.[19] It can be mild, with transverse grooves or rows of pits alternating with normal enamel. In severe cases, all enamel is hypoplastic. The enamel defects predispose to further damage and caries. Extensive dental repair is often necessary and should be promptly performed. Even this manifestation is probably autoimmune, and recently autoantibodies against protein phosphatase-4, involved in odontogenesis, have been detected in APS-1 patients.[54]

Other Manifestations

Aplasia or hypoplasia of the spleen occurs in up to 20% of patients[37,55] and renders them susceptible to bacterial infections. Vaccination for *Pneumococcus pneumonia*, *Haemophilus influenzae*, and meningococcus is recommended. Autoimmune hemolytic anemia,[37,43] hypoplastic anemia,[35,37] and autoimmune thrombocytopenia[56] have been reported in a few patients. Autoimmune bronchiolitis is an underappreciated life-threatening component of APS-1. Careful pulmonary investigation in a single center revealed that approximately 40% of patients had pneumonitis.[57] Lung biopsy revealed interstitial lymphoid infiltrates around bronchioles. Patients display autoantibodies against the BPIFB2 and KCNRG proteins. Beneficial effects of steroid and rituximab treatment have been reported. Reversible metaphyseal dysplasia with growth retardation and severe progressive myopathy have been observed in isolated cases.

Natural Course and Mortality

The overall mortality is increased,[58] but the course of the disease varies widely. Some patients have very mild disease and are not diagnosed before adult age, whereas others develop multiple and severe manifestations early on. Life expectancy depends on the severity of the disease, the quality of management, including psychosocial support, and the ability of the patient and family to assume responsibility. There is an increased risk of death from oral/esophageal squamous cell carcinoma, acute adrenal and hypocalcemic crises, complications such as pneumonitis, hepatitis, and renal failure, and accidents.[19,35,37,57,59] A recent Finnish survey revealed that standard mortality rates were increased approximately ten times, more so for younger age groups. The cumulative mortality rate up to 60 years of age was over 80%, compared with less than 10% for the general population.[59]

Diagnosis and Management

The traditional diagnostic criterion of the diagnostic dyad remains valid[34] (Table 132.2), but leaves many cases unrecognized. Some patients never develop more than one of these components, or the components manifest at a late stage. Initial presenting components can be chronic diarrhea, keratitis, periodic rash with fever, severe constipation, autoimmune hepatitis, pneumonitis, alopecia, or vitiligo[34,37] Any of these "minor" components in a child should alert clinicians to look for other manifestations and check autoantibodies directed against interferon-omega or -alpha, which are positive in almost all patients. We recommend securing the diagnosis by sequencing *AIRE*.

Every patient should, in addition to a local doctor's follow-up, be seen by appropriate specialists, particularly an endocrinologist and an oral specialist, at least once annually, and know to contact them immediately when new symptoms appear. Ideally, the follow-up should be sufficiently centralized so that the responsible specialists have adequate experience with several APS-1 patients. The goal is to safeguard the patient's physical and psychosocial well-being, as well as recognize and treat components at an early stage. Patient education to avoid and treat acute complications such as adrenal and hypocalcemic crises is important. All patients should be equipped with an acute care card or bracelet and steroids for self-injections and be well informed about components that might develop in the future.

KEY POINTS

- Autoimmune polyglandular syndrome type 1 often develops in childhood. Manifestations other than the main can appear first, such as intermittent fever with rash, hepatitis, intestinal malabsorption, keratitis, enamel hypoplasia, alopecia, vitiligo, and premature ovarian insufficiency. Treatment is replacement of missing hormones, antifungal treatment, and immunosuppression for some components. Mortality is increased because of adrenal insufficiency, hypocalcemic crises, complications owing to organ-specific autoimmunity (liver, lungs), and carcinoma of the mouth and esophagus.

TABLE 132.3 Immune Dysregulation Polyendocrinopathy Enteropathy X-Linked and Immune Dysregulation Polyendocrinopathy Enteropathy X-Linked-like Syndromes

Gene	Manifestations	Pathway	Inheritance
FOXP3	IPEX—T1DM, enteropathy, eczema, cytopenias, nephritis, autoimmune hypothyroidism, recurrent infections	Defective IL-2 signaling and Treg function	X-linked
CD25 (IL2R)	Type 1 diabetes, thyroiditis, enteropathy, eczema, cytopenias, recurrent infections	Defective IL-2 signaling and Treg function	Autosomal recessive
CTLA4	Enteropathy, hemolytic anemia, pneumonitis, lymphadenopathy, and hypogammaglobulinemia, autoimmune thyroid disease	Defective CTLA-4 function in Tregs	Autosomal dominant, haploinsufficiency, dominant negative effect
LRBA	Enteropathy, hemolytic anemia, splenomegaly, T1DM	Defective CTLA-4 cycling in the cell increases lysosomal degradation	Autosomal recessive
STAT1	CMC, thyroiditis, T1DM, eczema, enteropathy	Inhibited Th17 cell differentiation	Autosomal dominant, gain of function
STAT3	T1DM, enteropathy, celiac disease, autoimmune thyroid disease, hemolytic anemia, autoimmune thrombocytopenia	Increased Th17 cells, reduced Foxp3$^+$ Tregs	Autosomal dominant, gain of function
STAT5b	Growth hormone insensitivity, thyroiditis, enteropathy, eczema, cytopenias, recurrent infections	Defective IL-2 signaling pathway	Autosomal recessive
ITCH	Developmental delay, dysmorphic features, autoimmune hypothyroidism, hepatitis, enteropathy	Reduced TGF-β–mediated Foxp3 expression	Autosomal recessive
BACH2	Enteritis, pneumonitis, infections	Defective immunoglobulin class switching, lymphocyte maturation defect	Autosomal dominant, haploinsufficiency

CMC, Chronic mucocutaneous candidiasis; *CTLA-4*, cytotoxic T lymphocyte antigen-4; *Foxp3*, forkhead box P3; *IL*, interleukin; *IPEX*, immumodysregulation polyendocrinopathy enteropathy X-linked; *T1DM*, type 1 diabetes mellitus; *TGF*, transforming growth factor; *Th*, T helper; *Treg*, T regulatory cell.

IMMUMODYSREGULATION POLYENDOCRINOPATHY ENTEROPATHY X-LINKED

Immumodysregulation polyendocrinopathy enteropathy X-linked (IPEX; OMIM 304790) is an exceedingly rare syndrome characterized by early-onset T1DM, autoimmune enteropathy with diarrhea and malabsorption, and dermatitis.[60,61] Eosinophilia and elevated IgE levels are common, and some patients develop renal disease with membranous glomerulonephritis or interstitial nephritis. Other manifestations are alopecia, autoimmune thyroiditis, hepatitis, exocrine pancreatitis, and different autoimmune cytopenias.[62] Many of the components overlap with APS-1 but in general develop much earlier. IPEX is caused by mutations in the *FOXP3* gene, located on the X chromosome, which is a key transcription factor needed for Treg development.

IPEX patients develop a number of autoantibodies, some of which overlap with those found in APS-1. Examples are glutamic acid decarboxylase and islet cell antibodies associated with T1DM. Many develop autoantibodies against harmonin and villin,[63] proteins involved in anchoring intestinal villi, also expressed in the proximal tubules of the kidney. Intriguingly, they also display autoantibodies against interferon-omega and a number of nuclear receptors, including HNFA1, mutations of which cause maturity-onset diabetes of the young.[64]

High-throughput DNA sequencing technology has uncovered other unique monogenic syndromes with endocrine manifestations. Examples are loss-of-function mutations in genes such as *CD25*, which encodes the receptor for IL-2, as well as *STAT5b*,[65] *ITCH*,[66] *BACH2*,[67] *CTLA-4*, and *LRBA*, and gain-of-function mutations in *STAT1*[68] and *STAT3*,[69] all of which affect Treg function (Table 132.3).

KEY POINTS

- Immumodysregulation polyendocrinopathy enteropathy X-linked is caused by mutations in the forkhead box protein 3 gene and lack of functional T regulatory cells. The disease starts early, often in infancy, with components such as type 1 diabetes, enteritis, and dermatitis. Eosinophilia, elevated IgE, and autoantibodies against villin and harmonin are common. Bone marrow transplantation is often needed.

AUTOIMMUNE POLYGLANDULAR SYNDROME TYPE 2

Definition

As originally described, Schmidt syndrome comprises the occurrence of Addison disease together with autoimmune thyroid disease and/or T1DM and has been regarded as synonymous with APS-2 (OMIM 269200). Addison disease was the index disorder in originally separating APS-1 from APS-2 and has been regarded as an essential component.[7] A broader definition has since developed, with the description of multiple autoimmune diseases in patients with autoimmune thyroid disease, T1DM, or Addison disease. Attempts to subdivide patients have been made, such as those with "incomplete" forms of APS-2 or additional types like APS-3 (autoimmune thyroid disease plus another organ-specific autoimmune disease excluding Addison disease) and APS- 4 (Addison disease plus another organ-specific autoimmune disease excluding autoimmune thyroid disease and T1DM), but no classification has been universally adopted. The picture is rendered still more complex by the existence of subclinical autoimmune disease and the temporal evolution of the various components, and the past division of autoimmune diseases into organ-specific or non–organ-specific is artificial. In our view, the term APS-2 should be reserved for the occurrence of at least two autoimmune endocrinopathies—Addison disease, autoimmune thyroid disease, T1DM—in the same patient.[16] There is an increased risk of developing another autoimmune disease with any of these components in isolation, and so APS-2 represents a dense and distinctive clustering of autoimmunity within a wider set of disease associations. The key features are summarized in Table 132.4.

Pathogenesis

APS-2 is primarily caused by genetic polymorphisms controlling antigen presentation and immunoregulation, which are associated with susceptibility to multiple autoimmune diseases. Tissue-specific gene polymorphisms, genetic heterogeneity in susceptibility to the underlying components, and environmental factors, which are also associated with these diseases, determine the exact pattern of autoimmunity in an individual patient. Although APS-2 is often multigenerational, the

TABLE 132.4 **Characteristics of Autoimmune Polyendocrine Syndrome Type 2**

Components	Clinical Features	Diagnosis	Autoantigens
Primary adrenal insufficiency	Hyperpigmentation, salt craving, hyponatremia, hyperkalemia	Low cortisol, elevated ACTH, failed synacthen testing	21-hydroxylase
Autoimmune thyroid disease	Hypothyroidism and Graves disease	Elevated TSH, low free thyroxine Suppressed TSH, elevated free thyroxine	Thyroid peroxidase TSH receptor
Type 1 diabetes mellitus	Hyperglycemia, lack of insulin	Fasting serum glucose >7 mmol/L, HbA1c >48 mmol/mol	Glutamic decarboxylase-65, islet cell antigen-2, zinc transporter-8
Primary ovarian insufficiency	Early menopause (<40 years of age), primary amenorrhea	Elevated FSH and LH, low estrogen	21-hydroxylase
Gastritis	Achlorhydria, vitamin B12 deficiency, iron deficiency	Autoimmune gastritis type A on biopsy	Intrinsic factor and parietal cell (H^+/K^+-ATPase)
Celiac disease	Malabsorption, iron and vitamin B12 deficiency	Villous atrophy and lymphocytic infiltration on biopsy	Transglutaminase-2
Exocrine pancreatitis,	Chronic steatorrhea, malabsorption	Decreased fecal elastase	
Vitiligo	From patchy to total nonsegmental distribution	Clinical features	Tyrosinase, tyrosinase-related proteins, SOX10
Alopecia	From patchy to total or universal distribution	Clinical features	Tyrosine hydroxylase
Others (<1%)	Autoimmune hypophysitis, chronic autoimmune hepatitis, myasthenia gravis, primary biliary cirrhosis, rheumatoid arthritis, serositis, Sjögren syndrome, systemic lupus erythematosus	As for isolated disorders	As for isolated disorders

ACTH, Adrenocorticotropic hormone; *FSH,* follicle-stimulating hormone; *HbA1c*, glycated hemoglobin; *LH,* luteinizing hormone; *SOX,* SRY-related HMG-box; *TSH,* thyroid-stimulating hormone.

pattern of heritability is unclear. The components of APS-2 differ in their relative strength of heritability, with Addison disease being the most strongly based on additive genetic effects and autoimmune thyroid disease the least, especially in the case of Graves disease.[70] Almost all genetic studies to date have been limited to White populations.

The HLA genotype DR3-DQ2, DR4-DQ8 has been individually associated with type 1 diabetes, Addison disease, and autoimmune thyroid disease, as well as with celiac disease, and therefore predisposes to APS-2.[16] More detailed analysis has been performed on APS-2 cohorts that have been defined through a particular combination of clinical features. In one large German cohort, the DR3-DQ2/DRB1*04:04-DQ8 genotype was associated with classical Schmidt syndrome, whereas the DR3-DQ2/DRB1*04:01-DQ8 genotype was associated with APS-3; protective HLA genotypes also differed between these two subdivisions.[71] DRA1*0101/DRB1*0301 shows the strongest association in susceptibility to the combination of type 1 diabetes and autoimmune thyroid disease, and it has been suggested that this allele encodes an HLA molecule with a flexible pocket that can bind both islet cell and thyroid autoantigens.[72]

A number of non-HLA genes involved at multiple points in the immune response have been associated with each of the three components of APS-2. These include polymorphisms in *CTLA4, PTPN22* (encoding lymphoid tyrosine phosphatase), *CD25* (encoding the IL-2 receptor), *NALP1* (encoding NACHT leucine-rich repeat protein), *BACH2* (encoding BTB domain and CNC homolog 2), and *STAT4* (encoding signal transducer and activator of transcription 4).[15,73-75] Attempts have been made to determine whether there is an exaggerated association when the components are present in APS-2 rather than occurring in isolation, and this may be the case for *CTLA-4* and *PTPN22* polymorphisms.[76] However, in the case of *BACH2* polymorphisms, a stronger association with APS-2 over isolated Addison disease appeared in Norwegian, but not British, patients, and significant heterogeneity has also been seen in the association of *STAT4* polymorphism with Addison disease in various European populations. Polymorphisms in these genes have also been associated in various combinations with the other autoimmune components seen in APS-2 (Table 132.4). As one example, of 55 genes known to be associated with vitiligo, 14 have also been implicated in type 1 diabetes and/or autoimmune thyroid disease, and 17, which partially overlap, have been implicated in at least one of the other autoimmune components seen in APS-2.[77] Over 100 genes may contribute to vitiligo, and similar numbers are likely to be involved in the other autoimmune disorders, helping explain the variability of the autoimmune phenotype of APS-2.

Further variability is contributed by environmental factors, but no studies have yet looked at APS-2 systematically, as it has been assumed that genetic factors play a more dominant role in etiology than in the individual diseases. Some environmental factors are component-specific (e.g., iodine intake and autoimmune thyroid disease). Alcohol consumption and smoking are examples of more general effects, as alcohol is protective in autoimmune thyroid disease, SLE, and rheumatoid arthritis, whereas smoking is a risk factor for Graves disease, SLE, and rheumatoid arthritis, as well as for celiac disease. The autoimmune response in APS-2 is directed against the same autoantigens as in the isolated forms of each disease: the key autoantigens are 21-hydroxylase in Addison disease, thyroid peroxidase, and thyroglobulin in autoimmune hypothyroidism, the TSH receptor in Graves disease, and glutamic acid decarboxylase 65 and the protein tyrosine phosphatases IA-2 and IA-2β in T1DM. The pathogenetic mechanisms are also identical. Autoantibodies against interferons and cytokines are rare.[78]

KEY POINTS

- Autoimmune polyglandular syndrome type 2 is the occurrence of at least two autoimmune endocrinopathies—Addison disease, autoimmune thyroid disease, type 1 diabetes mellitus—in the same patient, and often aggregates in families. Multiple genes are involved, in particular human leukocyte antigen variants and immune genes such as *CTLA-4, PTPN-22,* and *BACH2.*

Clinical Features and Diagnosis

The exact prevalence of APS-2 is unknown, owing to the use of different classifications and the existence of subclinical disease, but it is much more common than APS-1. APS-2 occurs in 60% to 70% of patients with autoimmune Addison disease,[79,80] whereas around 15% of type 1 diabetes patients have autoimmune thyroid disease.[81] APS-2 is two to three times more common in women, and the age of onset is typically between 20 and 60 years old, with T1DM usually occurring earliest and autoimmune hypothyroidism occurring latest. The clinical features and diagnosis of each component of APS-2 are identical to those of the individual diseases. Although a family history of autoimmune disease is more common in APS-2, it is often absent, with some family members having autoantibodies but no clinically evident disease. The occurrence of nonendocrine autoimmune conditions in a patient is also a clue (Table 132.4).

Premature ovarian insufficiency in APS-2 occurs only in those with Addison disease and is associated with P450 side-chain cleavage enzyme and/or 17α-hydroxylase antibodies directed against shared steroid cell antigens in the adrenal and ovary; these antibodies are much less frequent in APS-2 than in APS-1.[79,80] An autoimmune basis for premature ovarian insufficiency not related to steroid cell autoimmunity is currently unclear, and although there is an association of this condition with autoimmune thyroid disease and other autoimmune diseases like SLE, a distinct association with APS-2 is not firmly established.[82] In contrast to APS-1, male hypogonadism is not a feature of APS-2, and autoimmune hypoparathyroidism also appears to be absent, although 2% of APS-2 patients have autoantibodies to the calcium-sensing receptor.[78] Autoimmune hypophysitis is a rare disorder, but 2% of such patients have APS-2, and a further 10% have autoimmune thyroid disease, Addison disease, or T1DM in isolation.[83] Vitiligo occurs in 5% to 10% of APS-2 patients and alopecia in 3% to 10%, a lower frequency than in APS-1, and chronic autoimmune hepatitis is much less common.[15,84] Pernicious anemia is also less common in APS-2; it occurs in 0.5% to 5% of patients, particularly in association with autoimmune thyroid diseases.[85] In contrast, celiac disease, Sjögren syndrome, rheumatoid arthritis, and SLE have been associated with APS-2 but not with APS-1.[86]

Management

The treatment of each disease component in APS-2 follows standard lines. There may be a greater risk of adrenal crisis in Addison disease occurring in the setting of APS-2, more often in those with T1DM than in isolated Addison disease, suggesting the need for enhanced advice on preventative measures in this group.[87] Unusual hypoglycemia in a patient with T1DM may herald Addison disease in a patient with unsuspected APS-2, and care is needed when initiating levothyroxine replacement in undiagnosed or inadequately treated Addison disease to avoid precipitating an adrenal crisis. Concurrent celiac disease may be the cause of fluctuating levothyroxine requirements or unexplained vitamin D or iron deficiency, as well as increased need for hydrocortisone and fludrocortisone.

The main importance of recognizing APS-2 lies in the approach to screening both patients (Table 132.5) and their families. The usually long prodromal phase in the development of autoimmune disease means that screening can identify those at risk and lead to treatment at an early stage. In patients with APS-2, follow-up depends on whether Addison disease is present. Annual screening of all patients with Addison disease should include testing for thyroid peroxidase antibodies and TSH, glucose and HbA1c, and full blood count and vitamin B12.[88] In patients with APS-2, but without Addison disease, 21-hydroxylase antibodies should be measured every 3 to 5 years. Biochemical testing should be undertaken with a Synacthen test or basal cortisol and ACTH level if 21-hydroxylase antibody-positive, with regular follow-up thereafter. If overt Addison disease does not develop within 20 years, the risk for future disease is minimal.[89] In all types of APS-2, follow-up should include a careful history and examination to detect any newly emerging autoimmune condition (Table 132.5), with appropriate testing as required. Given the frequency of asymptomatic celiac disease and the risk of malignancy, IgA tissue transglutaminase-2 antibodies should be measured at the time of diagnosis in APS-2.[85,90] There is no well-established protocol for family screening in APS-2, as the inheritance is so variable. First-degree family members should be made aware of the possibility of developing an autoimmune condition, and screening should be offered to those who wish. Such screening should include measurement of thyroid peroxidase antibodies and TSH, basal glucose and HbA1c, and 21-hydroxylase antibodies if the proband has Addison disease. High-throughput multiplexed autoantibody assays are being developed that should simplify and improve screening in the future.[91]

TABLE 132.5 Screening Schedule in Autoimmune Polyendocrine Syndrome Type 2

Frequency	Evaluation
Annual	Physical exam Hemoglobin, electrolytes, creatinine, cholesterol, triglyceride, HbA1c, TSH, thyroid peroxidase antibodies, ferritin, plasma renin activity
Every third year	*Addison disease* (if not present): 21-hydroxylase antibodies *Celiac disease*: transglutaminase-2 autoantibodies *Premature ovarian insufficiency*: side-chain cleavage autoantibodies (females <40 years of age) *Autoimmune gastritis*: parietal cell and intrinsic factor–blocking autoantibodies, full blood count, vitamin B_{12}

HbA1c, glycated hemoglobin; *TSH,* thyroid-stimulating hormone.

KEY POINTS

- Management of autoimmune polyglandular syndrome type 2 consists of standard hormone replacement for each component, together with annual screening to detect any emerging asymptomatic autoimmune disorder (Table 132.5) and ensuring that first-degree family members are aware of their possible risk.

IMMUNE CHECKPOINT INHIBITOR-INDUCED POLYENDOCRINOPATHY

Immune checkpoint inhibitors (ICIs) are therapeutic monoclonal antibodies that inhibit the CTLA-4 or PD1 pathways. Ipilimumab is a CTLA-4 inhibitor, nivolumab and pembrolizumab inhibit PD1 and atezolizumab, avelumab and duvalumab inhibit PD-ligand 1. ICIs inhibit T-cell activation, resulting in enhanced T-cell responses against tumor cells, but with the unwanted effect of reducing the capacity for these pathways to maintain self-tolerance. In practice, these agents are very effective in a range of different cancers, but they also cause a variety of autoimmune disorders, including those affecting the endocrine system.[92,93] Thyroid dysfunction occurs in 5% to 10% of ICI-treated patients. The frequency is greatest with the anti-PD1 agents, especially when given in combination with CTLA-4 inhibition. It presents most commonly as hypothyroidism, often preceded by destructive thyroiditis; Graves disease is much less common and needs to be distinguished from the thyrotoxic phase of destructive thyroiditis before any treatment is started. Hypophysitis occurs in around 1% of patients following ICI treatment, and is associated most strongly with ipilimumab, especially when given in combination with PD1 inhibitors. T1DM after anti-PD1 treatment is uncommon, and rare after CTLA-4 inhibition. A few cases of T1DM have been reported in patients with preexisting autoimmune thyroid disease, as has the concurrent development of T1DM together with thyroid disease, hypophysitis, vitiligo, or other autoimmune disorders. Addison disease is rare after ICI treatment, but in five such cases one had preexisting autoimmune thyroid disease, and two developed concurrent hypophysitis.[92] Polyendocrinopathy after ICI treatment is rare, but could be a model for APS-2, in which some shared genetic predisposition leads to disease, and may offer fresh insights into understanding why particular types of autoimmune disease arise.

KEY POINTS

- Immune checkpoint inhibitors are therapeutic monoclonal antibodies used in cancer treatment that inhibit the CTLA-4 or PD1 immunoregulatory pathways and can cause a variety of endocrine disorders by removing the brakes on T-cell activation.

REFERENCES

1. Claude M, Gougerot M. *J Physiol Pathol Gen*. 1908;10:469–480.
2. Schmidt MB. [Eine biglandulare Erkrankung (Nebennieren und Schilddrüse) bei Morbus Adisonii.] *Verh Dtsch Ges Pathol*. 1926;21:212–221.
3. Carpenter CC, Solomon N, Silverberg SG, et al. Schmidt's syndrome (thyroid and adrenal insufficiency). A review of the literature and a report of fifteen new cases including ten instances of coexistent diabetes mellitus. *Medicine (Baltim)*. 1964;43:153–180.
4. Czarnocka B, Ruf J, Ferrand M, et al. Purification of the human thyroid peroxidase and its identification as the microsomal antigen involved in autoimmune thyroid diseases. *FEBS Lett*. 1985;190:147–152.
5. Baekkeskov S, Aanstoot HJ, Christgau S, et al. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature*. 1990;347:151–156.
6. Winqvist O, Karlsson FA, Kampe O. 21-Hydroxylase, a major autoantigen in idiopathic Addison's disease. *Lancet*. 1992;339:1559–1562.
7. Neufeld M, Maclaren NK, Blizzard RM. Two types of autoimmune Addison's disease associated with different polyglandular autoimmune (PGA) syndromes. *Medicine (Baltim)*. 1981;60:355–362.
8. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol*. 2010;125:S3–S23.
9. Bluestone JA, Bour-Jordan H, Cheng M, et al. T cells in the control of organ-specific autoimmunity. *J Clin Invest*. 2015;125:2250–2260.
10. Theofilopoulos AN, Kono DH, Baccala R. The multiple pathways to autoimmunity. *Nat Immunol*. 2017;18:716–724.
11. Raphael I, Nalawade S, Eagar TN, et al. T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. *Cytokine*. 2015;74:5–17.
12. Cheng M, Anderson MS. Thymic tolerance as a key brake on autoimmunity. *Nat Immunol*. 2018;19:659–664.
13. Mueller DL. Mechanisms maintaining peripheral tolerance. *Nat Immunol*. 2010;11:21–27.
14. Marson A, Housley WJ, Hafler DA. Genetic basis of autoimmunity. *J Clin Invest*. 2015;125:2234–2241.
15. Betterle C, Dal Pra C, Mantero F, et al. Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocr Rev*. 2002;23:327–364.
16. Husebye ES, Anderson MS, Kampe O. Autoimmune polyendocrine syndromes. *N Engl J Med*. 2018;378:2543–2544.
17. Nagamine K, Peterson P, Scott HS, et al. Positional cloning of the APECED gene. *Nat Genet*. 1997;17:393–398.
18. Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*. 1997;17:399–403.
19. Bruserud O, Oftedal BE, Landegren N, et al. A longitudinal follow-up of autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*. 2016;101:2975–2983.
20. Cetani F, Barbesino G, Borsari S, et al. A novel mutation of the autoimmune regulator gene in an Italian kindred with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy, acting in a dominant fashion and strongly cosegregating with hypothyroid autoimmune thyroiditis. *J Clin Endocrinol Metab*. 2001;86:4747–4752.
21. Oftedal BE, Hellesen A, Erichsen MM, et al. Dominant mutations in the autoimmune regulator AIRE are associated with common organ-specific autoimmune diseases. *Immunity*. 2015;42:1185–1196.
22. Anderson MS, Venanzi ES, Klein L, et al. Projection of an immunological self shadow within the thymus by the aire protein. *Science*. 2002;298:1395–1401.
23. Gardner JM, Devoss JJ, Friedman RS, et al. Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science*. 2008;321:843–847.
24. Bruserud O, Oftedal BE, Wolff AB, et al. AIRE-mutations and autoimmune disease. *Curr Opin Immunol*. 2016;43:8–15.
25. Meager A, Visvalingam K, Peterson P, et al. Anti-interferon autoantibodies in autoimmune polyendocrinopathy syndrome type 1. *PLoS Med*. 2006;3:e289.
26. Toth B, Wolff AS, Halasz Z, et al. Novel sequence variation of AIRE and detection of interferon-omega antibodies in early infancy. *Clin Endocrinol*. 2010;72:641–647.
27. Kisand K, Boe Wolff AS, Podkrajsek KT, et al. Chronic mucocutaneous candidiasis in APECED or thymoma patients correlates with autoimmunity to Th17-associated cytokines. *J Exp Med*. 2010;207:299–308.
28. Puel A, Doffinger R, Natividad A, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J Exp Med*. 2010;207:291–297.
29. Alimohammadi M, Bjorklund P, Hallgren A, et al. Autoimmune polyendocrine syndrome type 1 and NALP5, a parathyroid autoantigen. *N Engl J Med*. 2008;358:1018–1028.
30. Alimohammadi M, Dubois N, Skoldberg F, et al. Pulmonary autoimmunity as a feature of autoimmune polyendocrine syndrome type 1 and identification of KCNRG as a bronchial autoantigen. *Proc Natl Acad Sci U S A*. 2009;106:4396–4401.
31. Shum AK, Alimohammadi M, Tan CL, et al. BPIFB1 is a lung-specific autoantigen associated with interstitial lung disease. *Sci Transl Med*. 2013;5:206ra139.
32. Burbelo PD, Ferre EMN, Chaturvedi A, et al. Profiling autoantibodies against salivary proteins in sicca conditions. *J Dent Res*. 2019;98:772–778.
33. Gavalas NG, Kemp EH, Krohn KJ, et al. The calcium-sensing receptor is a target of autoantibodies in patients with autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*. 2007;92:2107–2114.
34. Ahonen P, Myllarniemi S, Sipila I, et al. Clinical variation of autoimmune polyendocrinopathy-candidiasis- ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med*. 1990;322:1829–1836.
35. Orlova EM, Sozaeva LS, Kareva MA, et al. Expanding the phenotypic and genotypic landscape of autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*. 2017;102:3546–3556.
36. Ferre EM, Rose SR, Rosenzweig SD, et al. Redefined clinical features and diagnostic criteria in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *JCI Insight*. 2016;1:e88782.
37. Perheentupa J. Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *J Clin Endocrinol Metab*. 2006;91:2843–2850.
38. Bruserud O, Costea DE, Laakso S, et al. Oral tongue malignancies in autoimmune polyendocrine syndrome type 1. *Front Endocrinol*. 2018;9:463.
39. Soderbergh A, Myhre AG, Ekwall O, et al. Prevalence and clinical associations of 10 defined autoantibodies in autoimmune polyendocrine syndrome type I. *J Clin Endocrinol Metab*. 2004;89:557–562.
40. Saari V, Holopainen E, Makitie O, et al. Pubertal development and premature ovarian insufficiency in patients with APECED. *Eur J Endocrinol*. 2020;183:513–520.

41. Schaller CE, Wang CL, Beck-Engeser G, et al. Expression of Aire and the early wave of apoptosis in spermatogenesis. *J Immunol*. 2008;180:1338–1343.
42. Landegren N, Sharon D, Shum AK, et al. Transglutaminase 4 as a prostate autoantigen in male subfertility. *Sci Transl Med*. 2015;7:292ra101.
43. Betterle C, Greggio NA, Volpato M. Clinical review 93: autoimmune polyglandular syndrome type 1. *J Clin Endocrinol Metab*. 1998;83:1049–1055.
44. Ekwall O, Sjoberg K, Mirakian R, et al. Tryptophan hydroxylase autoantibodies and intestinal disease in autoimmune polyendocrine syndrome type 1. *Lancet*. 1999;354:568.
45. Dobes J, Neuwirth A, Dobesova M, et al. Gastrointestinal autoimmunity associated with loss of central tolerance to enteric alpha-defensins. *Gastroenterology*. 2015;149:139–150.
46. Bruserud O, Siddiqui H, Marthinussen MC, et al. Oral microbiota in autoimmune polyendocrine syndrome type 1. *J Oral Microbiol*. 2018;10:1442986.
47. Ward L, Paquette J, Seidman E, et al. Severe autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy in an adolescent girl with a novel AIRE mutation: response to immunosuppressive therapy. *J Clin Endocrinol Metab*. 1999;84:844–852.
48. Padeh S, Theodor R, Jonas A, et al. Severe malabsorption in autoimmune polyendocrinopathy-candidosis-ectodermal dystrophy syndrome successfully treated with immunosuppression. *Arch Dis Child*. 1997;76:532–534.
49. Chascsa DM, Ferre EMN, Hadjiyannis Y, et al. APECED-associated hepatitis: clinical, biochemical, histological and treatment data from a large predominantly American cohort. *Hepatology*. 2020;73:1088–1104.
50. Obermayer-Straub P, Perheentupa J, Braun S, et al. Hepatic autoantigens in patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *Gastroenterology*. 2001;121:668–677.
51. DeVoss J, Hou Y, Johannes K, et al. Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. *J Exp Med*. 2006;203:2727–2735.
52. Hedstrand H, Ekwall O, Haavik J, et al. Identification of tyrosine hydroxylase as an autoantigen in autoimmune polyendocrine syndrome type I. *Biochem Biophys Res Commun*. 2000;267:456–461.
53. Hedstrand H, Ekwall O, Olsson MJ, et al. The transcription factors SOX9 and SOX10 are vitiligo autoantigens in autoimmune polyendocrine syndrome type I. *J Biol Chem*. 2001;276:35390–35395.
54. Vazquez SE, Ferre EM, Scheel DW, et al. Identification of novel, clinically correlated autoantigens in the monogenic autoimmune syndrome APS1 by proteome-wide PhIP-Seq. *Elife*. 2020;9:e55053.
55. Pollak U, Bar-Sever Z, Hoffer V, et al. Asplenia and functional hyposplenism in autoimmune polyglandular syndrome type 1. *Eur J Pediatr*. 2009;168:233–235.
56. Ulinski T, Perrin L, Morris M, et al. Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome with renal failure: impact of posttransplant immunosuppression on disease activity. *J Clin Endocrinol Metab*. 2006;91:192–195.
57. Ferre EMN, Break TJ, Burbelo PD, et al. Lymphocyte-driven regional immunopathology in pneumonitis caused by impaired central immune tolerance. *Sci Transl Med*. 2019;11:eaav5597.
58. Bensing S, Brandt L, Tabaroj F, et al. Increased death risk and altered cancer incidence pattern in patients with isolated or combined autoimmune primary adrenocortical insufficiency. *Clin Endocrinol (Oxf)*. 2008;69:697–704.
59. Borchers J, Pukkala E, Makitie O, et al. Patients with APECED have increased early mortality due to endocrine causes, malignancies and infections. *J Clin Endocrinol Metab*. 2020;105:e2207–e2213.
60. Powell BR, Buist NR, Stenzel P. An X-linked syndrome of diarrhea, polyendocrinopathy, and fatal infection in infancy. *J Pediatr*. 1982;100:731–737.
61. Wildin RS, Smyk-Pearson S, Filipovich AH. Clinical and molecular features of the immunodysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. *J Med Genet*. 2002;39:537–545.
62. Bacchetta R, Barzaghi F, Roncarolo MG. From IPEX syndrome to FOXP3 mutation: a lesson on immune dysregulation. *Ann N Y Acad Sci*. 2018;1417:5–22.
63. Lampasona V, Passerini L, Barzaghi F, et al. Autoantibodies to harmonin and villin are diagnostic markers in children with IPEX syndrome. *PLoS One*. 2013;8:e78664.
64. Eriksson D, Bacchetta R, Gunnarsson HI, et al. The autoimmune targets in IPEX are dominated by gut epithelial proteins. *J Allergy Clin Immunol*. 2019;144:327–330.e8.
65. Kofoed EM, Hwa V, Little B, et al. Growth hormone insensitivity associated with a STAT5b mutation. *N Engl J Med*. 2003;349:1139–1147.
66. Lohr NJ, Molleston JP, Strauss KA, et al. Human ITCH E3 ubiquitin ligase deficiency causes syndromic multisystem autoimmune disease. *Am J Hum Genet*. 2010;86:447–453.
67. Afzali B, Gronholm J, Vandrovcova J, et al. BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. *Nat Immunol*. 2017;18:813–823.
68. Uzel G, Sampaio EP, Lawrence MG, et al. Dominant gain-of-function STAT1 mutations in FOXP3 wild-type immune dysregulation-polyendocrinopathy-enteropathy-X-linked-like syndrome. *J Allergy Clin Immunol*. 2013;131:1611–1623.
69. Flanagan SE, Haapaniemi E, Russell MA, et al. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat Genet*. 2014;46:812–814.
70. Skov J, Eriksson D, Kuja-Halkola R, et al. Co-aggregation and heritability of organ-specific autoimmunity: a population-based twin study. *Eur J Endocrinol*. 2020;182:473–480.
71. Flesch BK, Matheis N, Alt T, et al. HLA class II haplotypes differentiate between the adult autoimmune polyglandular syndrome types II and III. *J Clin Endocrinol Metab*. 2014;99:E177–E182.
72. Li CW, Osman R, Menconi F, et al. Flexible peptide recognition by HLA-DR triggers specific autoimmune T-cell responses in autoimmune thyroiditis and diabetes. *J Autoimmun*. 2017;76:1–9.
73. Magitta NF, Boe Wolff AS, Johansson S, et al. A coding polymorphism in NALP1 confers risk for autoimmune Addison's disease and type 1 diabetes. *Genes Immun*. 2009;10:120–124.
74. Mitchell AL, Macarthur KD, Gan EH, et al. Association of autoimmune Addison's disease with alleles of STAT4 and GATA3 in European cohorts. *PLoS One*. 2014;9:e88991.
75. Pazderska A, Oftedal BE, Napier CM, et al. A variant in the BACH2 gene is associated with susceptibility to autoimmune Addison's disease in humans. *J Clin Endocrinol Metab*. 2016;101:3865–3869.
76. Houcken J, Degenhart C, Bender K, et al. PTPN22 and CTLA-4 polymorphisms are associated with polyglandular autoimmunity. *J Clin Endocrinol Metab*. 2018;103:1977–1984.
77. Jin Y, Andersen G, Yorgov D, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet*. 2016;48:1418–1424.
78. Kemp EH, Kahaly GJ, Porter JA, et al. Autoantibodies against the calcium-sensing receptor and cytokines in autoimmune polyglandular syndromes types 2, 3 and 4. *Clin Endocrinol (Oxf)*. 2018;88:139–145.
79. Erichsen MM, Lovas K, Skinningsrud B, et al. Clinical, immunological, and genetic features of autoimmune primary adrenal insufficiency: observations from a Norwegian registry. *J Clin Endocrinol Metab*. 2009;94:4882–4890.
80. Dalin F, Nordling Eriksson G, Dahlqvist P, et al. Clinical and immunological characteristics of autoimmune Addison disease: a nationwide Swedish multicenter study. *J Clin Endocrinol Metab*. 2017;102:379–389.
81. van den Boom L, Buchal G, Kaiser M, et al. Multimorbidity among adult outpatients with type 1 diabetes in Germany. *J Diabetes Sci Technol*. 2020 [Epub ahead of print].
82. Sharif K, Watad A, Bridgewood C, et al. Insights into the autoimmune aspect of premature ovarian insufficiency. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101323.
83. Caturegli P, Newschaffer C, Olivi A, et al. Autoimmune hypophysitis. *Endocr Rev*. 2005;2:599–614.
84. Kahaly GJ, Frommer L. Polyglandular autoimmune syndromes. *J Endocrinol Invest*. 2018;41:91–98.
85. Lahner E, Conti L, Cicone F, et al. Thyro-entero-gastric autoimmunity: pathophysiology and implications for patient management. *Best Pract Res Clin Endocrinol Metab*. 2020;34:101373.
86. Kluger N, Krohn K, Ranki A. Absence of some common organ-specific and non-organ-specific autoimmunity in autoimmune polyendocrinopathy candidiasis ectodermal dystrophy. *Endocr Connect*. 2013;2:61–68.

87. Meyer G, Badenhoop K, Linder R. Addison's disease with polyglandular autoimmunity carries a more than 2.5-fold risk for adrenal crises: German Health insurance data 2010-2013. *Clin Endocrinol (Oxf)*. 2016;85:347–353.
88. Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2016;101:364–389.
89. Naletto L, Frigo AC, Ceccato F, et al. The natural history of autoimmune Addison's disease from the detection of autoantibodies to development of the disease: a long-term follow-up study on 143 patients. *Eur J Endocrinol*. 2019;180:223–234.
90. Lebwohl B, Sanders DS, Green PHR. Coeliac disease. *Lancet*. 2018;391:70–81.
91. Gu Y, Zhao Z, Waugh K, et al. High-throughput multiplexed autoantibody detection to screen type 1 diabetes and multiple autoimmune diseases simultaneously. *EBioMedicine*. 2019;47:365–372.
92. Chang LS, Barroso-Sousa R, Tolaney SM, et al. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. 2019;40:17–65.
93. Barroso-Sousa R, Barry WT, Garrido-Castro AC, et al. Incidence of endocrine dysfunction following the use of different immune checkpoint inhibitor regimens: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4:173–182.

133

Multiple Endocrine Neoplasia Type 1

Paul J. Newey and Rajesh V. Thakker

OUTLINE

Historical Aspects, 2237
Clinical Findings and Treatment, 2239
  Parathyroid Tumors, 2240
  Pancreatic Tumors, 2241
  Pituitary Tumors, 2244
  Associated Tumors, 2245
  Carcinoid Tumors, 2245
  Other Clinical Considerations, 2247
Molecular Genetics, 2247
  The *MEN1* Gene, 2247
  MEN1 Germline Mutations, 2249
  MEN1 Germline Coding Variants, 2250
  MEN1 Somatic Mutations, 2250
  MEN1 Phenocopies and Mutations in Other Genes, 2250
  MEN1 Mutational Analysis in Clinical Practice, 2251
Surveillance of Individuals at Risk of Multiple Endocrine Neoplasia Type 1 Tumors, 2252
Function of the Multiple Endocrine Neoplasia Type 1 Protein, Menin, and Future Targeted Treatments, 2252
Conclusion, 2254

Multiple endocrine neoplasia (MEN) is characterized by the occurrence of tumors involving two or more endocrine glands within a single patient.[1] The disorder has previously been referred to as *multiple endocrine adenopathy* or *pluriglandular syndrome*. However, glandular hyperplasia and malignancy may also occur in some patients, and the term *multiple endocrine neoplasia* is preferred. Four major forms of MEN, referred to as MEN types 1 to 4 (MEN1 to MEN4) are recognized, and each form is characterized by the development of tumors within specific endocrine glands (Table 133.1).[1] Thus, combined occurrence of tumors of the parathyroid glands, the pancreatic islet cells, and the anterior pituitary gland is characteristic of MEN1, which is also referred to as *Wermer syndrome*. In addition to these tumors, adrenal cortical tumors, thymic, bronchial, and gastric carcinoid tumors, meningiomas, facial angiofibromas, collagenomas, lipomas, and paraganglioma have been described in patients with MEN1.[1,2,2a] MEN1 is caused by mutations of the tumor suppressor *MEN1* gene, which encodes menin, which functions as a scaffold protein.[3,4] In MEN2, which is also called *Sipple syndrome*, medullary thyroid carcinoma (MTC) occurs in association with pheochromocytoma, and two main forms are recognized: MEN2A (now referred to as MEN2) and MEN2B (also referred to as MEN3). MEN2 and MEN3 are caused by specific mutations of the rearranged during transcription (*RET*) protooncogene, which encodes a receptor tyrosine kinase (RTK).[4a] MEN2, which is the most common form, is characterized by the development of MTC in association with pheochromocytoma and parathyroid tumors. Three variants of MEN2 are reported, including MEN2 with Hirschsprung disease, MEN2 with cutaneous lichen amyloidosis, and familial MTC only, in which MTC is the sole manifestation of the syndrome. However, in MEN3, parathyroid involvement is rare, and the occurrence of MTC and pheochromocytoma is found in association with a marfanoid habitus, mucosal neuromas, medullated corneal fibers, and intestinal autonomic ganglion dysfunction leading to megacolon. MEN4 is associated with mutations of the gene encoding the cyclin-dependent kinase inhibitor (CK1) p27 kip1 (*CDNK1B*), and patients may develop parathyroid adenomas, pituitary adenomas, duodenopancreatic neuroendocrine tumors, and tumors of the gonads, adrenal, kidneys, and thyroid.[5,6] All of these forms of MEN may be inherited as autosomal dominant syndromes or may occur sporadically (i.e., without a family history), although this distinction between sporadic and familial cases may sometimes be difficult.[1]

KEY POINTS

- Multiple endocrine neoplasia (MEN) syndromes (types 1–4, MEN1–4) are a group of heterogeneous disorders characterized by occurrence of specific tumors that may be inherited as autosomal dominant traits.

This chapter discusses the clinical aspects of the tumors arising in the context of MEN1 and reviews the molecular genetic basis of this syndrome.

KEY POINTS

- MEN1 is characterized by the occurrence of tumors of the parathyroid glands, pancreatic islets (e.g., gastrinomas and insulinomas), and anterior pituitary gland (e.g., prolactinomas and somatotropinomas), with some patients also developing adrenal and carcinoid tumors; in the absence of treatment, these endocrine tumors are associated with an earlier mortality in patients with MEN1.

HISTORICAL ASPECTS

Patients with MEN1 are characterized by the combined occurrence of tumors of the parathyroid glands, pancreatic islet cells, and anterior pituitary gland. The first occurrence of such multiple endocrine tumors was reported in 1903, in a patient with acromegaly whose autopsy revealed the presence of an anterior pituitary tumor and enlarged

TABLE 133.1 Multiple Endocrine Neoplasia Syndromes and Their Characteristic Tumors and Associated Genetic Abnormalities

Type (Chromosomal Location)	Tumors (Estimated Penetrance)	Gene; Most Frequently Mutated Codons
MEN1 (11q13)	Parathyroid adenoma (90%)	*MEN1*[c]
	Enteropancreatic tumor (30% to 70%)	83/84, 4-bp del (≈4%)
	Gastrinoma (40%)	119, 3-bp del (≈3%)
	Insulinoma (10%)	210–211, 4-bp del (≈8%)
	Nonfunctioning and PPoma (20%–55%)	418, 3-bp del (≈4%)
	Glucagonoma (<1%)	514–516, del or ins (≈7%)
	VIPoma (<1%)	Intron 4 ss (≈10%)
	Pituitary adenoma (30%–40%)	
	Prolactinoma (15%–20%)	
	Somatotropinoma (5%–10%)	
	Corticotropinoma (<5%)	
	Nonfunctioning (10%–15%)	
	Associated Tumors	
	Adrenal cortical tumor (20%–40%)	
	Pheochromocytoma (<1%)	
	Bronchopulmonary NET (4%–13%)	
	Thymic NET (2%–8%)	
	Gastric NET (10%)	
	Lipomas (30%)	
	Angiofibromas (85%)	
	Collagenomas (70%)	
	Meningiomas (8%)	
	Breast cancer[b]	
MEN2 (10 cen-10q11.2)		
MEN2A[a]	MTC (90%)	RET
	Pheochromocytoma (50%)	634, e.g., Cys→Arg (~85%)
	Parathyroid adenoma (20% to 30%)	
MEN2B (also known as MEN3)	MTC (>90%)	RET 918, Met→Thr (>95%)
	Pheochromocytoma	
	Associated abnormalities (40%–50%)	
	Mucosal neuromas	
	Marfanoid habitus	
	Medullated cornel nerve fibers	
	Megacolon	
MEN4 (12p13)	Parathyroid adenoma[d]	*CDKN1B*, no common mutations identified to date
	Pituitary adenoma (e.g., acromegaly, Cushing disease)[d]	
	Gastro-duodeno-pancreatic NETs (e.g., gastrinoma, nonfunctional pancreatic NET)[d]	
	Reproductive organ tumors[d] (e.g., testicular cancer, neuroendocrine cervical carcinoma)	
	Adrenal + renal tumors[d]	
	Bronchial carcinoid tumor[d]	
	Thyroid tumors[d]	

Autosomal dominant inheritance of MEN1 syndrome has been established.

[a] MEN2A also includes three variant disorders: familial-MTC only, MEN2A with cutaneous lichen amyloidosis, and MEN2A with Hirschsprung disease.

[b] An increased relative risk of breast cancer is reported (relative risk 2.8). The diagnosis of breast cancer in female MEN1 patients is also reported to occur at a younger age compared with the general population (i.e., median/mean age of diagnosis of 45–50 years in different series).

[c] *MEN1* mutation codon location reported according to the transcript encoding the 610–amino acid menin protein (NM_130799)

[d] Insufficient numbers reported to provide prevalence information.

del, Deletion; *ins,* insertion; *PPoma,* pancreatic polypeptide–secreting tumor; *VIPoma,* vasoactive intestinal polypeptide–secreting tumor; *MTC,* medullary thyroid cancer; *NET* neuroendocrine tumor; *RET,* rearranged during transcription; *ss,* splice site.

(Modified from Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab.* 2012;97:2990–3011.)

parathyroid glands.[7] Pancreatic islet cell tumors were subsequently observed in association with parathyroid and pituitary tumors, and a familial occurrence was suggested by the finding of these tumors in two sisters. Further case reports revealing the triad of parathyroid, pancreatic islet cell, and anterior pituitary tumors in individual patients, led to the recognition of a unifying disorder,[8] and a familial basis was demonstrated by documenting its occurrence in a father and daughter of one family and in a father and four daughters of another family. An autosomal dominant mode of inheritance was proposed and established by further family studies that demonstrated inheritance of the syndrome in five generations, with equal frequency in male and female subjects. Molecular genetic linkage studies mapped the *MEN1* locus to chromosome 11q13, and further positional cloning studies identified the *MEN1* gene in 1997[9,10] and revealed that it encoded a 610–amino acid ubiquitously expressed protein, designated *menin*.[9] Menin has been shown to be predominantly a nuclear protein that acts as a scaffold protein, interacting with a number of proteins involved in transcriptional regulation, genome stability, cell division, proliferation, and epigenetic regulation.[1,3,11] Elucidation of the crystal structure of menin has revealed that it contains a deep pocket that binds to the mixed lineage leukemia protein (MLL1) and transcription factor JunD, and that pharmacologic agents can inhibit these interactions.[12]

CLINICAL FINDINGS AND TREATMENT

The estimated prevalence of MEN1 is 1:30,000 of the population. The clinical manifestations of MEN1 are related to the sites of tumor development and to their products of secretion. In addition to the triad of parathyroid, pancreatic, and pituitary tumors (Fig. 133.1), which constitute the major components of MEN1, adrenal cortical tumors, carcinoid tumors (e.g., of the thymus, bronchi, and stomach), facial angiofibromas, collagenomas, meningiomas, and lipomas have also been described (see Table 133.1),[1,13] together with an increased risk of breast cancer in women with MEN1.[14] MEN1 is highly penetrant, such that virtually all patients develop clinical or biochemical evidence of tumor development by the age of 50 years. Clinical disease expression is unusual in early childhood (i.e., <5 years of age), but more than 20% of patients will have at least one clinical or biochemical manifestation by the age of 20 years.[13] Parathyroid tumors are the first clinical manifestation of MEN1 in 75% to 90% of patients.[13] The pancreatic tumors, which are neuroendocrine tumors (NETs) and occur in 40% to 70% of MEN1 patients, consist of gastrinomas, insulinomas, glucagonomas, vasoactive intestinal polypeptidomas (VIPomas), and nonfunctioning pancreatic NETs[4,15-18]; whereas anterior pituitary tumors, consisting of prolactinomas, somatotropinomas, corticotropinomas, and nonfunctioning adenomas, occur in 30% to 40% of patients.[19] The frequency of other endocrine tumors is variable, with 20% to 55% of MEN1 patients developing adrenal tumors[13,20] and less than 15% of patients manifesting thymic or bronchial carcinoid tumors.[21,22] Combinations of these tumors have been reported to differ in members of the same family and even between identical twins.[23] In children, presentations with pituitary adenomas (e.g., prolactinomas, Cushings disease), insulinomas, and nonfunctioning NETs are not infrequent, and some may occasionally present with thymic or adrenal malignancy. MEN1 is inherited as an autosomal dominant disorder, but nonfamilial (i.e., sporadic) forms account for 10% to 15% of patients with MEN1, with *de novo MEN1* mutations occurring in approximately 10% of MEN1 patients.[1]

A diagnosis of MEN1 may be established in an individual by one of three criteria (Fig. 133.2)[1]: on the basis of occurrence of two or more primary MEN1-associated endocrine tumors (i.e., parathyroid, enteropancreatic, and pituitary tumors); occurrence of one of the MEN1-associated tumors in a first-degree relative of a patient with a clinical diagnosis of MEN; and identification of a germline *MEN1* mutation in an individual who may be asymptomatic and has not yet developed serum biochemical or radiologic evidence of tumor development.

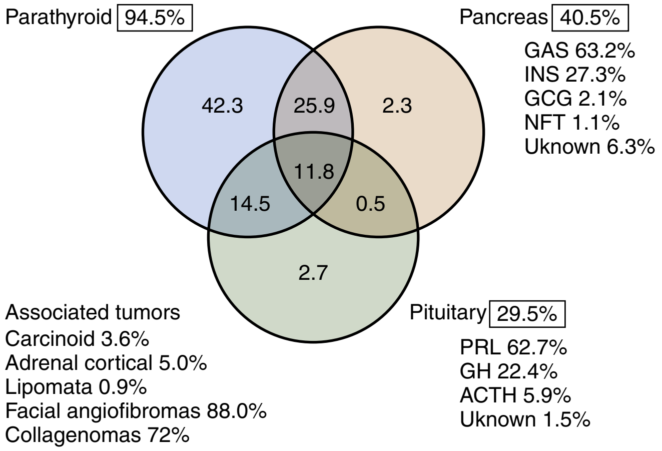


Fig. 133.1 Schematic representation of the distribution of 384 multiple endocrine neoplasia type 1 (MEN1) tumors in 220 patients with MEN1. The proportions of patients in whom parathyroid, pancreatic, or pituitary tumors occurred are shown in the respective boxes; for example, ~95% of patients had a parathyroid tumor. The Venn diagram indicates the proportions of patients with each combination of tumors; for example, ~35% (~25% + ~10%) of patients had both a parathyroid and a pancreatic tumor, whereas ~2% of patients had a pancreatic tumor only. In addition to the tumors reported in this series, a high frequency of facial angiofibromas (>80%) and collagenomas (>70%) have been reported in MEN1 patients in other series. The hormones secreted by each of these tumors are indicated: *ACTH,* adrenocorticotropic hormone; *GAS,* gastrin; *GCG,* glucagon; *GH,* growth hormone; *INS,* insulin; *NFT,* nonfunctioning tumor; *PRL,* prolactin. Parathyroid tumors represent the most common form of MEN1 tumors and occur in approximately 95% of patients, with pancreatic islet cell tumors occurring in approximately 40% and anterior pituitary tumors occurring in around 30%. However, use of radiologic surveillance methods has resulted in identification of nonfunctioning pancreatic NETs in 15%–55% of MEN1 patients, and thus these represent the most prevalent pancreatic NETs. (Modified from Trump D, Farren B, Wooding C, et al. Clinical studies of multiple endocrine neoplasia type 1 [MEN1]. *QJM.* 1996;89:653–669.)

In the absence of treatment, endocrine tumors are associated with an earlier mortality in patients with MEN1[1]; for example, malignant pancreatic NETs and thymic carcinoid tumors are associated with a marked increase in risk for death (hazard ratio [HR] >3, $P < .005$). However, the prognosis of MEN1 patients with gastrinomas, which are associated with severe multiple peptic ulcers, has improved considerably following the introduction of acid-suppressive therapies, although 30% to 70% of individuals with MEN1 still die at a mean age of 50 to 60 years, of causes directly related to MEN1.[24] This earlier mortality, despite treatment, which for each type of MEN1-associated endocrine tumor is generally similar to that for the respective tumors occurring in non-MEN1 patients,[1] may be because of the occurrence of multiple synchronous or asynchronous tumors in the same or different tissues in MEN1 patients. For example, MEN1 patients often develop multiple submucosal duodenal gastrinomas, such that achieving biochemical remission is difficult without extensive surgical resection, and the cooccurrence of multiple synchronous pancreatic NETs may complicate the interpretation of imaging and biochemical investigations, as well as the planning of surgical intervention.

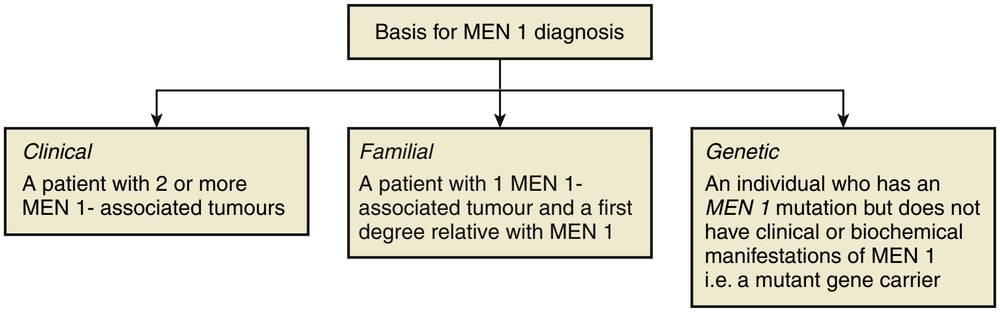

Basis for MEN 1 diagnosis
Clinical
A patient with 2 or more MEN 1- associated tumours
Familial
A patient with 1 MEN 1- associated tumour and a first degree relative with MEN 1
Genetic
An individual who has an *MEN 1* mutation but does not have clinical or biochemical manifestations of MEN 1 i.e. a mutant gene carrier

Fig. 133.2 Basis for a diagnosis of multiple endocrine neoplasia type 1 (*MEN1*) in individuals. A diagnosis of MEN1 based on clinical and familial criteria may be confounded by the occurrence of phenocopies. (Reproduced from Turner JJ, Christie PT, Pearce SH, et al. Diagnostic challenges due to phenocopies: lessons from Multiple Endocrine Neoplasia type 1 [MEN1]. *Hum Mutat.* 2010;31:E1089–E1101.)

Finally, it is important to recognize that there is an absence of high-quality evidence (i.e., randomized controlled trials) to guide the treatment of MEN1-associated tumors, and that the recommendations are typically based on the treatment of their sporadic counterparts or on limited case series of MEN1 patients, which may be prone to ascertainment and reporting biases.[1] It is therefore advised that decisions regarding treatment are made by a multidisciplinary team with relevant experience in the management of MEN1 that also takes into account the views of the patient. In particular, the potential risks and benefits of any intervention need to be balanced, with the aims of treatment being to minimize disease-associated morbidity and mortality, avoid iatrogenic harms, and preserve the patient's quality of life.[1]

KEY POINTS

- Malignant pancreatic neuroendocrine tumors and thymic carcinoid tumors are the leading causes of premature mortality. In the absence of treatment, multiple endocrine neoplasia type 1–associated tumors are associated with significant morbidity and an earlier mortality.

Parathyroid Tumors

Primary hyperparathyroidism occurs in approximately 95% of MEN1 patients and is the first manifestation of disease in 75% to 90% of patients.[1,13,13a] Patients may have asymptomatic hypercalcemia, nephrolithiasis, osteitis fibrosa cystica, vague symptoms associated with hypercalcemia (e.g., polyuria, polydipsia, constipation, or malaise), or occasionally peptic ulcers. Biochemical investigations reveal hypercalcemia, in association with raised or inappropriately normal circulating parathyroid hormone (PTH) concentrations.[1] The hypercalcemia is usually mild, and severe hypercalcemia resulting in crisis is rare, as is the occurrence of parathyroid carcinoma. Additional differences in the primary hyperparathyroidism of patients with MEN1, as opposed to those without MEN1, include an earlier age at onset (e.g., 15–25 years vs. 55 years), an equal male/female ratio (1:1 vs. 1:3), synchronous or asynchronous involvement of all four parathyroid glands, and a greater reduction in bone mineral density (BMD),[1] which particularly affects the lumbar spine, femoral neck, and distal radius, and is postulated to be attributed to the earlier age of disease onset and chronicity of disease. Primary hyperparathyroidism in patients with MEN1 is unusual before 10 years of age, although asymptomatic and symptomatic cases have been reported at as early as 4 and 8 years of age, respectively.[13]

Treatment. Surgical removal of the abnormally overactive parathyroid glands is the definitive treatment and is recommended for MEN1 patients with primary hyperparathyroidism who have symptomatic disease, severe hypercalcemia (i.e., corrected serum calcium >3.0 mmol/L), and/or those with evidence of end-organ damage (e.g., renal stones, hypercalciuria, or osteoporosis on BMD imaging (i.e., T score < –2.5). For those without a clear indication for surgery, regular clinical and biochemical monitoring, together with periodic reevaluation for end organ damage (e.g., renal ultrasound scan, dual-energy x-ray absorptiometry), is suggested.[1]

However, there is controversy regarding the timing (i.e., at an early or late stage[1]) and extent of parathyroidectomy (subtotal parathyroidectomy [SP; three and a half glands] or total parathyroidectomy [TP]) that should be undertaken, as all four (or more) parathyroid glands are typically affected, and preoperative imaging (e.g., neck ultrasound, Tc99m-sestamibi parathyroid scintigraphy, computed tomography [CT], magnetic resonance imaging [MRI], and (^{18}F)-fluorodeoxyglucose (FDG) positron emission tomography [PET]/CT) is reported to be of limited benefit in surgery-naïve MEN1 patients. For example, preoperative localization studies in MEN1 have been reported to: not alter the surgical approach in greater than 90% of MEN1 patients; be of only limited value in identifying ectopic parathyroid glands (correctly identified in only ~40% of patients); and only correctly identify the largest parathyroid in approximately 70% of cases, while failing to identify enlarged contralateral parathyroid glands in approximately 85% of these cases.[25] Thus, bilateral neck exploration is usually required, and the surgical approach needs to balance the likelihood of persistent or recurrent hyperparathyroidism with the risks of permanent hypoparathyroidism. In most centers, SP of at least three and a half glands is the recommended surgical approach. However, some other centers advocate TP with or without autotransplantation of fresh or cryopreserved parathyroid tissue (e.g., to the forearm), even though the risk of permanent hypoparathyroidism may be high and ranges from less than 15% to greater than 65%,[26,27] and recurrent hypercalcemia from the transplanted tissue is frequently observed, necessitating its surgical removal, which is not always successful. Moreover, an autotransplanted parathyroid gland has been reported to contain a metastatic neuroendocrine thymic carcinoma, thereby indicating further caution in adopting this approach. Concurrent bilateral transcervical thymectomy is also recommended at the time of surgery to remove parathyroid tumors that may be embedded within the thymus.[1] Such extensive parathyroidectomy is recommended, because lesser surgical approaches (i.e., removal of less than three glands) are associated with higher persistence and recurrence rates, of 15% to 65% at 15 years postsurgery. For example, a study of 99 MEN1 patients reported persistent hyperparathyroidism in approximately 70% of patients whose initial operation comprised removal of only one or two glands, compared with 20% in those in whom two and a half to three glands were removed, and only approximately 5% in those with removal of three and a half glands.[25] Persistent hypocalcemia following such extensive parathyroid surgery is treated with oral calcitriol

(1,25-dihydroxyvitamin D), although management of hypoparathyroidism can be challenging in some patients, even with the use of vitamin D analogs and calcium replacement.[1]

However, recent studies have reported more favorable short- to medium-term outcomes for MEN1 patients treated with less extensive surgical approaches (i.e., unilateral parathyroidectomy or minimally invasive parathyroidectomy).[28,29] For example, a study of 89 MEN1 patients treated for primary hyperparathyroidism with gland excision (GE) of one or two glands ($n = 28$), SP ($n = 23$), or TP with autotransplantation ($n = 38$) reported disease persistence and recurrence rates at a median follow-up of 112 months in the GE group of 15% and greater than 20%, respectively, which, although significantly higher than for the SP and TP groups, were associated with 0% of permanent hypoparathyroidism that was greater than 15% and greater than 30% in the SP and TP groups, respectively.[29] Thus, some centers advocate GE, which acknowledges the likely need for future surgery but avoids the immediate risks of permanent hypoparathyroidism in some patients (e.g., younger patients with apparent evidence of single or unilateral disease) who may benefit from a limited parathyroidectomy. Such individualization of the extent of parathyroid surgery in some MEN1 patients may be facilitated by the use of: preoperative parathyroid venous sampling, which has been reported to reduce the rates of permanent hypoparathyroidism and recurrent hyperparathyroidism; and of intraoperative monitoring of PTH by rapid assay that has been used to confirm the successful removal of hyperfunctioning parathyroid tissue, and the identification of ectopic and/or supernumerary hyperfunctioning parathyroid glands, located predominantly in the thymus but also at other locations in the mediastinum or carotid sheath.

These uncertainties in parathyroid surgery for MEN1 patients reflect the lack of high-quality evidence to guide clinical recommendations, and at present the timing and extent of surgery requires careful consideration of the surgical experience, availability of facilities for long-term regular serum calcium monitoring, accessibility of calcitriol (or other vitamin D analogs), and patient preferences.[1] Calcimimetics (e.g., Cinacalcet), which act via the calcium-sensing receptor, may also have a role to reduce or normalize serum calcium and circulating PTH levels in some MEN1 patients, in whom surgery has either failed or is contraindicated.

KEY POINTS

- Primary hyperparathyroidism occurs in >95% of multiple endocrine neoplasia type 1 patients and is the first manifestation of disease in the majority of patients. Most centers recommend subtotal parathyroidectomy of at least three and a half glands because typically there is synchronous or asynchronous involvement of all four parathyroid glands. Cinacalcet may be used for those in whom surgery has failed or is contraindicated.

Pancreatic Tumors

The incidence of clinically apparent pancreatic NETs in patients with MEN1 varies from 30% to 80% in different series, although microscopic islet tumors are found in almost all MEN1 patients. Malignant pancreatic NETs remain the leading cause of premature death in MEN1 patients.[15-17,30,31] MEN1-associated pancreatic NETs (Table 133.1) may produce excessive amounts of hormone (e.g., gastrin, insulin, glucagon, or vasoactive intestinal polypeptide [VIP]) and are associated with distinct clinical syndromes, although some (e.g., those secreting pancreatic polypeptide [PP]) may not be associated with clinical manifestations or may be nonsecretory (i.e., nonfunctioning) (see Fig. 133.1).[1] Indeed, nonfunctioning NETs are now reported to be the most common pancreatic NET in MEN1.[4,15-17,31,32] MEN1 patients may have more than one synchronous pancreatic NET, and it cannot be assumed that tumor visualization on imaging studies correlates with the site of hormone excess in functioning tumor syndromes (e.g., gastrinoma, insulinoma).[1,16] The goal of treatment of MEN1 pancreatic NETs is to reduce the morbidity and mortality associated with these tumors, which may be associated with hormone hypersecretion (e.g., gastrin, insulin) and/or their malignant behavior (i.e., development of metastases).[31] Surgery is generally recommended for functioning pancreatic NETs associated with localized disease (i.e., within the pancreas), as it offers a potentially curative approach.[1,4,17] For nonfunctioning pancreatic NETs, surgery is typically reserved for localized tumors greater than 2 cm, as the risk of metastasis and death is observed to increase significantly above this size.[1,4,15-17,33] Indeed, when evaluated as a single group, the risk of metastases in MEN1 duodenopancreatic NETs is associated with size greater than 2cm and age greater than 40 years, and surprisingly not with hypersecretion (or nonsecretion) of hormones.[34]

Gastrinoma. Zollinger and Ellison initially described two patients in whom non–β islet cell tumors of the pancreas were associated with recurrent peptic ulceration and marked gastric acid production, and gastrin was subsequently extracted from such tumors. The association of recurrent peptic ulceration, marked gastric acid production, and non–β islet cell tumors of the pancreas is referred to as the Zollinger–Ellison syndrome (ZES).[1] ZES patients may also suffer from esophageal symptoms (e.g., related to esophageal stricture, or Barrett's esophagus), as well as diarrhea and steatorrhea.

The diagnosis is established by demonstration of a raised fasting serum gastrin concentration in association with increased basal gastric acid secretion.[1] Occasionally, intravenous provocative tests with either secretin or calcium infusion are required to distinguish patients with ZES from other causes of hypergastrinemia, which include achlorhydria, antral G-cell hyperplasia, *Helicobacter pylori* infection, renal failure, hypercalcemia, and use of proton pump inhibitors (PPIs).[1] Investigations that may help to localize the gastrinomas include ultrasonography, endoscopic ultrasonography (EUS), CT, MRI, selective abdominal angiography, or somatostatin receptor scintigraphy (SRS).[1] In addition, the combined use of intraarterial calcium injections with hepatic venous gastrin sampling has been shown to regionalize the gastrinomas, while the selective arterial secretagogue injection test may also help localize the tumor.[1] However, MEN1-associated gastrinomas frequently occur with concomitant pancreatic NETs, most frequently nonfunctioning tumors, and this failure to distinguish between tumor types may have confounded earlier studies. Thus, it is likely that in many MEN1 patients with ZES, large (>2 cm) pancreatic tumors that were assumed to be gastrinoma were instead nonfunctioning NETs, and that the gastrinomas, which occurred within the duodenal wall, were unrecognized.[35] Moreover, in patients with MEN1, ZES does not appear to occur in the absence of primary hyperparathyroidism, and hypergastrinemia has also been reported to be associated with hypercalcemia. The successful treatment of the primary hyperparathyroidism with restoration of normocalcemia may ameliorate the clinical symptoms and biochemical abnormalities in approximately 20% of patients.[1] Thus, the diagnosis of ZES may be difficult in some MEN1 patients.

Gastrin-secreting tumors (gastrinomas) are reported to occur in 20% to 60% of MEN1 patients (see Fig. 133.1), and approximately 20% of patients with gastrinomas will have MEN1.[1,22] The majority of MEN1-associated gastrinomas occur as multiple small tumors (<0.5 cm in size) located within the duodenal mucosa and occur only rarely as macroscopic pancreatic tumors.[18,36] Despite their typical small size, local lymph node metastases appear to occur early in the disease, while hepatic metastases occur less frequently. For example, 70% to 80% of surgically treated patients have evidence of local lymph node metastases, while synchronous hepatic metastases are reported in approximately 10%.[37] Gastrinomas occur more often in male MEN1 patients great than 30 years of age, and past *H. pylori* infection is associated with an increased prevalence and severity of hypergastrinemia in MEN1 patients.[38] Gastrinomas are a major cause of morbidity and mortality in patients with MEN1, and before the advent of H2 receptor blockers and PPI therapy, recurrent severe multiple peptic ulcers (with associated perforation) and cachexia were major contributors to the high mortality. In addition, a recent study of 63 patients with MEN1-associated gastrinomas reported 5- and 10-year overall survival rates of approximately 80% and approximately 65%, respectively, with high gastrin levels (i.e., >20 times the upper limit of normal), a pancreatic NET greater than 2 cm, synchronous liver metastases, multiple concurrent NETs, and gastroduodenoscopy suspicious of gastric NET, each contributing to the reduced survival.[18]

KEY POINTS

- Gastrinomas occur in 20% to 60% of MEN1 patients, and when untreated are associated with the Zollinger–Ellison syndrome and increased mortality. Insulinomas occur in 10% to 30% of MEN1 patients, typically presenting at under 40 years of age, and are the first manifestation of disease in 10% of MEN1 patients.

KEY POINTS

- Clinically apparent nonfunctioning pancreatic neuroendocrine tumors occur in 15% to 55% of patients and are the leading cause of premature death in MEN1. The risk of malignancy is partially dependent on tumor size, and most centers recommend surgery for tumors over 2 cm, although some advocate lower thresholds.

Treatment. The aims of treatment are to control symptoms and/or sequelae associated with hypergastrinemia (e.g., ZES) and to reduce the mortality associated with the development of metastatic disease.[39] However, the inability to accurately predict disease course makes decisions regarding the optimal treatment challenging, with some centers recommending medical treatment and others favoring early surgical intervention.[36,39] Medical treatment of MEN1 patients with ZES is directed toward reducing basal acid output to less than 10 mmol/L, and this may be achieved by high-dose parietal cell H^+/K^+-adenosine triphosphatase (ATPase) inhibitors (i.e., PPI therapy including omeprazole or lansoprazole). Some patients may also require additional treatment with the histamine H2 receptor antagonists cimetidine or ranitidine.[1] The introduction of acid-suppressive therapies has markedly reduced the morbidity and associated mortality from the direct effects of hypergastrinemia, such that the overall prognosis of gastrinoma in MEN1 patients is now generally favorable, with 5-, 10-, and 20-year survival rates of approximately 80% to 95%, approximately 65% to 95%, and approximately 60% to 90%, respectively,[18,36] However, the effects of medical treatment on disease course (i.e., tumor growth, risk of metastases) and long-term consequences with continuous use for greater than 10 to 20 years are not known. The value of somatostatin analogues (SSAs), which have been used in a limited number of patients for the treatment of MEN1-associated gastrinoma, remains uncertain. The benefits of surgery for the majority of MEN1 patients with multiple duodenal gastrinomas are not established. For example, extensive duodenal resection aiming to remove all gastrinomas and thereby achieve persistent postoperative biochemical remission (i.e., eugastrinemia) is rarely possible, and duodenal-preserving procedures are associated with persistent or recurrent hypergastrinemia in 60% to 100% of patients.[1,17,35] However, some studies have reported improved remission rates of 75% to 100%, with short- to medium-term follow up,[35] following extensive surgical approaches including: duodenotomy with local removal of tumors in the duodenal wall, coupled with enucleation/resection of tumors in the pancreatic head and peripancreatic lymph node removal; partial pancreaticoduodenectomy; and pancreas-preserving duodenectomy.[1,35,36] Whipple pancreatoduodenectomy has been reported to achieve a greater than 65% cure rate but is associated with a higher operative mortality and long-term complications, which include weight loss, diabetes mellitus, and malabsorption, such that it is reserved for patients with large tumors, and total duodenopancreatectomy is very rarely performed because of the high risk of long-term sequelae (i.e., brittle insulin-dependent diabetes).[1] Laparoscopic surgery is not recommended, while total gastrectomy is rarely undertaken and only employed in very limited situations (e.g., a persistently noncompliant patient). Currently, most centers undertake nonsurgical management for the majority of MEN1 patients with duodenal gastrinomas, although some specialized centers with experience in preoperative and intraoperative imaging (e.g., transillumination of the duodenum) and regionalization advocate surgery for gastrinoma early in the disease course.[1] However, surgery is recommended for the rare patients (<1%) with a pancreatic gastrinoma greater than 2 cm, as it results in improved disease-related survival, and also for the minority of MEN1 gastrinoma patients who develop aggressive gastrinomas.[1,17,18] In summary, the role of surgery in the treatment of gastrinomas in patients with MEN1 is controversial, and this is partly because of an absence of randomized studies comparing medical and surgical approaches and the availability of a wide range of potential surgical techniques that have been evaluated only in the short to medium term.[17,18,35,40]

Treatment of disseminated gastrinomas is difficult, and chemotherapy with streptozotocin and 5-fluorouracil, capecitabine and temozolomide, cisplatin and etoposide, hormonal therapy with the SSAs octreotide or lanreotide, selected internal radiation therapy, radiofrequency ablation, peptide radioreceptor therapy (PRRT), hepatic artery embolization, administration of human leukocyte interferon, and removal of all resectable tumor, and hepatic transplantation have all been successful occasionally.[1]

Insulinoma. Insulinomas, β islet cell tumors that secrete insulin, occur in 10% to 30% (see Fig. 133.1) of MEN1 patients[1,37] and may occur in association with other pancreatic NETs (e.g., gastrinomas) in approximately 10% of patients. Insulinomas occur more often in patients with MEN1 who are younger than 40 years of age, whereas in patients without MEN1, insulinomas generally occur in those over 40 years.[1] Insulinomas may occur in childhood; in a study of 160 MEN1 patients presenting with manifestations before the age of 21 years, more than 10% of the cohort had insulinoma, with the earliest presentation at 5 years of age.[13] Indeed, insulinomas are the first manifestation of MEN1 in approximately 10% of patients, and approximately 5% of patients with insulinomas will have MEN1.[1,2,13] Patients typically present with hypoglycemic symptoms that develop after a fast or exertion and improve after glucose intake.[1] The most reliable test is a supervised 72-hour fast with biochemical investigations revealing increased plasma insulin, C-peptide, and proinsulin concentrations in association with hypoglycemia.[1] During this test, it is important to demonstrate an absence of sulfonylureas in the plasma and/or urine. The majority (65%–70%) of insulinomas in MEN1 occur as solitary pancreatic lesions, with most occurring in the body or tail, although multiple or multicentric tumors occur in 35%–40% of patients. However, patients diagnosed to have multifocal insulinomas have been reported to not have multiple insulin-staining tumors on histology, but instead to be likely have other pancreatic NET subtypes (e.g., nonfunctioning tumors).[37] Preoperative localization with EUS, CT scanning, MRI or celiac axis angiography, preoperative and perioperative percutaneous transhepatic portal venous sampling, selective intraarterial stimulation by calcium with hepatic venous sampling, and intraoperative direct pancreatic ultrasonography have been undertaken to improve the success rate of surgery.[1] SRS with ^{111}In-octreotide is associated with low sensitivity for insulinoma detection, although ^{68}Ga-DOTATATE PET/CT may improve detection rates. Scintigraphy based on glucagon-like peptide-1 has also been reported to be a highly sensitive modality for the detection of benign sporadic insulinomas, and its value in MEN1 patients with insulinoma has been demonstrated when used in conjunction with MRI.[41] The utility of FDG-PET/CT for insulinoma is typically limited to high-grade metastatic disease.

Treatment. Medical treatment, which consists of frequent carbohydrate meals and diazoxide or octreotide, is not always successful, and surgery is the treatment of choice in those with localized disease.[1,37] Surgical treatment, which ranges from enucleation of a single tumor to distal pancreatectomy or partial pancreatectomy, or excision of all the macroscopic pancreatic tumors with enucleation of nodules in the remaining pancreas, has been curative in many patients.[37] A recent retrospective study reported a 10-year hypoglycemia-free survival rate of greater than 90% following surgery for nonmetastatic disease.[37] Monitoring the insulin/glucose ratio during surgery has been reported to be of value in assessing successful removal of the insulinoma, as well as preventing intraoperative hypoglycemia and postoperative complications and infections. Insulinomas in MEN1 typically occur as benign tumors with highly favorable outcomes, although metastatic disease does occur in a minority of patients. Chemotherapy consisting of streptozotocin, 5-flurouracil, and doxorubicin, PPRT with ^{177}Lu-DOTATATE, or hepatic artery embolization have been used for metastatic disease.[1,4,17]

Glucagonoma. Glucagonomas, α islet cell glucagon-secreting pancreatic tumors, occur in less than 3% of patients with MEN1, although some nonfunctioning pancreatic NETs may immunostain for glucagon.[1] The characteristic clinical manifestations of a skin rash (necrolytic migratory erythema), weight loss, anemia, and stomatitis may be absent, and the presence of the tumor may have been detected in an asymptomatic patient with MEN1 undergoing pancreatic imaging or detected by glucose intolerance and hyperglucagonemia.[1]

Treatment. The tail of the pancreas is the most frequent site for glucagonomas, and surgical removal is the treatment of choice.[1,4] However, treatment may be difficult, because approximately 50% to 80% of patients have metastases at the time of diagnosis. Medical treatment with the SSAs octreotide and lanreotide or chemotherapy with streptozotocin and 5-fluorouracil or dimethyltriazeno-imidazole carboxamide has been successful in some patients, and hepatic artery embolization has been used to treat metastatic disease.[1,4]

Vasoactive Intestinal Polypeptidoma. Patients with VIPomas develop watery diarrhea, hypokalemia, and achlorhydria. This clinical syndrome has been referred to as Verner–Morrison syndrome, WDHA syndrome, or VIPoma syndrome. VIPomas have been reported in only a few patients with MEN1,[1] and the diagnosis is established by excluding laxative and diuretic abuse, confirming a stool volume in excess of 0.5 to 1.0 L/day during a fast, and documenting a markedly increased plasma VIP concentration.[1]

Treatment. Surgical management of VIPomas, which are mostly located in the tail of the pancreas, has been curative. However, in patients with unresectable tumor, treatment with SSAs, streptozotocin with 5-fluorouracil, corticosteroids, indomethacin, metoclopramide, and lithium carbonate has proved beneficial, and hepatic artery embolization has been useful for the treatment of metastases.[1]

Nonfunctioning Pancreatic Neuroendocrine Tumors. Nonfunctioning pancreatic NETs likely represent a heterogeneous group with distinct subtypes, with some nonfunctioning pancreatic NETs not secreting any hormones (also referred to as nonsecreting NETs) and therefore not associated with a clinical syndrome, and others associated with minor elevations (e.g., less than two times the upper limit of normal) of circulating pancreatic hormones (e.g., PP or glucagon) but without clinical features.[1] Thus, MEN1 patients with nonfunctioning pancreatic NETs may be asymptomatic, with the tumor detected during imaging tests, or they may present with symptoms related to local mass effects (i.e., pain and/or compression of adjacent structures) or metastatic disease (e.g., cachexia, jaundice, hepatomegaly). Currently, there are no reliable biomarkers for the detection of nonfunctioning pancreatic NETs, and several studies have reported a low diagnostic accuracy of PP, glucagon, and chromogranin A (CgA).[16,30,42,43] However, a recently developed test based on a multitranscript molecular signature for polymerase chain reaction–based blood analysis (referred to as the NETest) was reported in a study of patients with gastroenteropancreatic NETs to be a better predictor than CgA for progressive disease, tumor grade, and liver metastases.[44] A separate validation of the NETest is required in MEN1, because MEN1 patients often have multiple concomitant NETs that may be from different cellular origins.[16,30] The increasing sensitivity of radiologic surveillance methods has resulted in a greater identification of nonfunctioning pancreatic NETs. Thus, they are now the most prevalent pancreatic NET in MEN1 patients, with macroscopic tumors occurring in 15% to 55% of individuals,[4,16,17] although histopathological studies have reported that almost all MEN1 patients harbor nonfunctioning microadenomas.[17] Nonfunctioning pancreatic NETs, which are most frequently diagnosed in adult MEN1 patients, have also been identified in 5% to 40% of children and young adults aged 12 to 20 years.[13,45]

The identification of nonfunctioning pancreatic NETs is of particular clinical importance for several reasons. For example, malignant pancreatic NETs are now reported to be the most common cause of death in individuals with MEN1,[4,16,17] and have a worse prognosis than other functioning tumors, including insulinoma and gastrinoma. In addition, the absence of a clinical syndrome and specific biochemical abnormalities may result in a delayed diagnosis of nonfunctioning pancreatic NETs, unless a radiologic assessment is undertaken. However, the optimum screening method(s) for the detection of MEN1-associated pancreatic nonfunctioning NETs has not been established. A comparison of imaging modalities has demonstrated that EUS is likely to represent the most sensitive method for detecting small pancreatic tumors, while ^{68}Gallium-DOTA PET/CT is the most reliable method for detecting metastatic disease.[16,17] The use of EUS for screening is limited by its availability and patient acceptability, as well as the marked heterogeneity in sensitivity throughout the pancreas.[30] Furthermore, the value detecting small tumors (e.g., <1 cm) in asymptomatic individuals has yet to be fully evaluated. Based on the available evidence, as well as practical considerations and avoidance of ionizing radiation, MRI every 1 to 3 years is the preferred modality for nonfunctioning pancreatic NET surveillance in most centers, with EUS being reserved for additional characterization, as it allows the acquisition of tumor material via fine-needle aspirate for the analysis of tumor grade and/or additional molecular markers.[17,30]

Treatment. Management of asymptomatic nonfunctioning pancreatic NETs is controversial.[1,16,17] The goal of treatment is to reduce morbidity and mortality associated with metastatic disease while preserving pancreatic tissue and avoiding complications associated with surgery. Pancreatoduodenal surgery, which may be successful in tumor excision in approximately 80% of patients, may also result in complications that include diabetes mellitus, frequent steatorrhea, and early and late dumping syndromes. For example, in a study of 61 MEN1 patients who underwent surgery for nonfunctioning pancreatic NETs, approximately 35% were reported to develop major early complications, whereas approximately 25% developed endocrine or exocrine pancreatic insufficiency, with potential impacts on quality of life.[46] Other factors that should also be considered before undertaking surgery include the presence of occult metastatic disease (i.e., tumors not detected by imaging investigations) that may be present at the time of initial presentation, and that tumors may recur in remnant pancreatic tissue.[1]

Treatment recommendations are currently based on tumor size, as an increased rate of metastases is observed in patients with larger tumors.[15] For example, one study reported the presence of synchronous metastases in approximately 45% of patients with nonfunctioning NETs greater than 3 cm, approximately 20% of patients with NETs 2.1 to 3 cm, and less than 5% of patients with tumors less than 1 cm.[1] However, other studies have not confirmed this association, and tumor size is not universally correct at predicting metastatic risk, as a small percentage of patients with small tumors develop metastatic disease, thereby supporting the importance of World Health Organization (WHO) tumor-grading when considering management of these tumors. However, the majority of MEN1-associated nonfunctioning pancreatic NETs less than 2 cm seem to have an indolent behavior, with slow growth rates of 0.1 to 1.3 mm/year,[16,30,32,42,47] although some tumors will have faster growth rates, thereby suggesting that tumors can be stratified according to growth rates.[30,42,48] Biomarkers accurately predicting tumor behavior and/or disease course are currently unavailable. However, for tumors greater than 2 cm, a higher WHO tumor grade is reported to be associated with an increased risk of liver metastases (WHO G2 vs. G1 tumors = ~80% vs. ~25%).[49] Enhancer signatures that map gene regulatory networks have also been reported to be able to differentiate nonfunctioning pancreatic NETs into subtypes, based on their resemblance to islet α cells (defined by ARX expression) or β-cells (defined by PDX1 expression), with distant relapses being associated almost exclusively with ARX$^+$PDX$^-$ tumors.[50] This finding, which was observed in 61 MEN1 patients, indicated that tumor profiling may facilitate stratification of MEN1-associated pancreatic NETs and guide clinical follow-up.[50]

Currently, most centers advocate surgery for nonfunctioning pancreatic NETs greater than 2 cm (although some centers use a lower threshold of >1 cm) or tumors demonstrating rapid growth (e.g., doubling of size over a 3- to 6-month interval and exceeding 1 cm in diameter). Several studies have indicated that tumor size greater than 2 cm is associated with hepatic metastases and worse outcomes,[15,51] and surgery for nonfunctioning pancreatic NETs greater than 3 cm in MEN1 patients has been reported to be beneficial, as it reduced the occurrence of hepatic metastases or death to approximately 40% (i.e., six of 16 patients), compared with 80% (five of six patients) in the nonsurgery group.[32] However, such benefits of surgery for nonfunctioning pancreatic NETs 2 to 3 cm in size could not be established.[32] For smaller nonfunctioning pancreatic NETs (<2 cm) in MEN1 patients, a conservative (medical) approach is adopted by most centers, because studies have reported that approximately 70% of nonfunctioning pancreatic NETs less than 2 cm occurring in MEN1 patients remain stable over a median follow-up of 3 years[48]; approximately 60% (28 of 46) of MEN1 patients with nonfunctioning tumors less than 2 cm not undergoing surgery had stable disease for 10 years, while in the remaining approximately 40% who had an increase in size or number of tumors, only approximately 5% required surgery, and only one patient died from metastatic disease[47]; and progression-free survival (PFS) rates in MEN1 patients with nonfunctioning pancreatic NETs are not altered by surgery, as evidenced by similar PFS rates of approximately 65%, approximately 40%, and approximately 10% at 5, 10, and 15 years in the surgery (n = 27 patients) and nonsurgery (n = 33 patients) groups.[52] Medical therapies are therefore used to treat MEN1 patients with nonfunctioning pancreatic tumors less than 2 cm in size.

Medical therapies for small (i.e., <2 cm) nonfunctioning tumors include SSAs and use of long-acting octreotide, which has been reported to reduce or stabilize tumor load in approximately 10% and approximately 80% of MEN1 patients (n = 20), respectively; and lanreotide, which, when compared with active surveillance, has been reported in 42 MEN1 patients to be associated with improvements in objective tumor responses (lanreotide vs. active surveillance = ~15% [four of 23 patients] vs. 0%), stable disease (~65% [15 of 23 patients] vs. ~30% [6 of 19 patients]), and PFS over a median period of 6 years.[53] However, further studies are required to validate these findings and to determine whether SSA therapy reduces the risk of developing metastatic disease. For patients with advanced disease, treatments include locoregional and systemic approaches and are similar to those used for sporadic NETs.[1,4,17] However, the evidence base for the use of such treatments has typically arisen from the study of non-MEN1 patients, and extrapolation to the MEN1 population requires caution. For example, a study that demonstrated a significant benefit of SSA therapy in the treatment of metastatic enteropancreatic NETs excluded patients with MEN1; while studies demonstrating the efficacy of sunitinib, an RTK inhibitor, and everolimus, a mammalian target of rapamycin (mTOR) inhibitor in the treatment of advanced pancreatic NETs, included very few (n = 2), who were in the nontreatment arm, or unknown numbers of MEN1 patients, respectively. Additional treatments for advanced, metastatic disease include PRRT using ^{177}Lu-DOTATATE, systemic chemotherapy, and a variety of locoregional approaches (e.g., radiofrequency ablation, transarterial chemoembolization, and selective internal radiation therapy).[4,17]

Somatostatinoma. Pancreatic tumors secreting somatostatin are associated with the somatostatinoma syndrome, which is characterized by hyperglycemia, cholelithiasis, low acid output, steatorrhea, diarrhea, abdominal pain, anemia, and weight loss. Although approximately 5% of pancreatic islet cell tumors in MEN1 secrete somatostatin, the somatostatinoma syndrome has not been reported to date in MEN1.

GHRHoma. GHRHomas, or tumors that secrete growth hormone–releasing hormone (GHRH), have been reported in some patients with MEN1.[1] GHRHomas may be diagnosed by finding elevated serum concentrations of growth hormone and GHRH. In the context of MEN1, GHRHomas occur predominantly within the pancreas, although sporadic GHRHomas may arise in the lung or small intestine. Surgical removal is the treatment of choice for these tumors.

Pituitary Tumors

The incidence of pituitary tumors in patients with MEN1 varies from 15% to 50% in different series (see Table 133.1), and they may be the first presentation of MEN1 in 10% to 20% of patients.[1] Approximately 65% of MEN1-associated pituitary adenomas occur in female patients[19,54] and present in early adulthood, with a mean age of onset of 30 to 40 years, but they may occur earlier. For example, the youngest reported case of a MEN1 pituitary tumor was a prolactinoma in a 5-year-old boy, and in a study of 160 children presenting with manifestations of MEN1 before the age of 21 years, approximately 35% of the children had pituitary adenomas, with the majority arising between 15 and 20 years of age.[13] The majority of pituitary adenomas in MEN1 are microadenomas, with macroadenomas representing 25% to 35% of the tumors.[19,54] Pituitary tumors in MEN1 patients are reported to have a higher occurrence of invasive features but without a greater prevalence of pituitary carcinoma, and an increased prevalence of plurihormonal expression, when compared with their non-MEN1 counterparts, although no specific histological features are established to differentiate MEN1 from non-MEN1 pituitary tumors.[1]

Approximately 60% of the pituitary tumors in MEN1 patients are hormone-secreting, and the remaining approximately 40% are nonsecreting (nonfunctional). Among the hormone-secreting tumors, approximately 75% are prolactinomas, 5% to 15% are growth hormone (GH)-secreting tumors, approximately 5% are adrenocorticotrophic hormone (ACTH)-secreting tumors, less than 5% are gonadotroph-secreting tumors, and 5% to 10% are cosecreting tumors, for example, somatolactotrophinomas.[19,54] Approximately 75% of nonfunctioning pituitary adenomas, which may secrete glycoprotein subunits, are microadenomas, and the majority of these will remain stable for more than 5 years.[19,54] The clinical manifestations of each of the pituitary tumor types in patients with MEN1 are similar to those in patients with sporadic pituitary tumors and depend on the hormone secreted and the size of the pituitary tumor. Thus, patients may have symptoms related to hyperprolactinemia, acromegaly, or Cushing disease, while enlarging pituitary tumors may compress the optic chiasm or adjacent normal pituitary tissue resulting in visual disturbances, and/or hypopituitarism. The diagnosis and clinical investigation of MEN1 pituitary tumors is similar to that for sporadic pituitary tumors, including dynamic pituitary function testing and imaging to characterize the nature and extent of the tumor and determine any compromise on pituitary function.

KEY POINTS

- Pituitary tumors occur in 15% to 50% of multiple endocrine neoplasia type 1 patients and are the first presentation of disease in 10% to 20% of cases.

Treatment. Overall, the prognosis of pituitary tumors in patients with MEN1 is favorable, reflecting their typical benign features. Treatments of pituitary tumors in MEN1 patients are similar to those in patients without MEN1 and consist of appropriate medical therapy (e.g., bromocriptine or cabergoline for prolactinoma; or SSAs or GH receptor antagonists for somatotropinoma) or selective transsphenoidal adenomectomy if feasible, with radiotherapy reserved for residual unresectable tumor tissue.[1] Recent studies have reported that approximately 90% of MEN1 patients with prolactinomas respond well to dopamine agonist therapy,[19] and that the majority of asymptomatic nonfunctioning tumors managed conservatively remain stable (or demonstrate minimal growth) on serial imaging, with only a minority progressing and requiring surgical intervention.[19,54]

Associated Tumors

Carcinoid Tumors

Thymic Carcinoids. Thymic carcinoids occur in 2% to 8% of MEN1 patients and are associated with an aggressive disease course and an increased risk of death (HR = 4.29), such that they account for up to 20% of premature deaths in MEN1 patients.[1] The median age of diagnosis for thymic carcinoids in MEN1 patients is 40 to 45 years, and the median survival following diagnosis is approximately 9.5 years, with approximately 70% of patients dying as a direct result of the thymic tumor.[22,55] Earlier presentations of thymic carcinoids are reported and include a 16 year-old boy who died approximately 4 years following diagnosis with local and distant metastases.[13,22,55] Thymic carcinoids in European patients with MEN1 occur predominantly in men (male:female ratio = 20:1), with an increased prevalence in cigarette smokers; whereas in Japanese and Chinese MEN1 patients, thymic carcinoids have less marked gender differences (e.g., male:female ratio Japan = 2:1; China = 1:1). The majority of thymic carcinoids are nonsecretory, such that the features of carcinoid syndrome are absent, and the presentation is usually with clinical features resulting from the local effects of an anterior mediastinal mass (e.g., chest pain, vena caval obstruction), or from symptoms related to metastatic disease (e.g., to liver, lung, bone). However, the majority of patients are asymptomatic at diagnosis, with most tumors detected radiologically. Biochemical abnormalities (e.g., raised CgA or urinary 5-hydroxyindoleacetic acid) are insufficiently sensitive for tumor detection, and as such radiological surveillance is recommended, although the optimal screening methods have not been established, and there are also concerns regarding the repeated exposure to ionizing radiation that is associated with CT scans, which are considered to be more reliable for tumor detection (see later).[1]

The treatment of thymic carcinoids in MEN1 is dependent on the stage of presentation, as older age of presentation, larger tumor diameter, and presence of metastases are associated with worse outcome.[22,55] Complete surgical resection is the treatment of choice where possible, although recurrence rates following surgery are high. Adjuvant radiotherapy may be used where resection is considered incomplete. For patients with advanced disease, the potential treatment options are based on those used for sporadic thymic carcinoids and include chemotherapy (e.g., fluorouracil/temozolomide, etoposide/cisplatin), SSAs, radiotherapy, mTOR inhibitors (everolimus), and PRRT. Finally, it should be noted that, although transcervical thymectomy is recommended at the time of parathyroid surgery in MEN1 patients, this procedure does not reliably prevent thymic carcinoid development.

KEY POINTS

- Thymic carcinoid tumors occur in only 2% to 8% of multiple endocrine neoplasia type 1 patients but are associated with an aggressive disease course and account for 20% of premature deaths. Bronchopulmonary carcinoid tumors occur in 5% to 30% of patients, with equal sex distribution, and are often small (<1 cm) and multiple. Surgical resection is recommended, although in some settings tumor surveillance may be appropriate.

Bronchopulmonary Carcinoids. Bronchopulmonary carcinoid tumors occur in less than 5% to greater than 30% of MEN1 patients,[56,57] and it is important to distinguish these from alternative etiologies, including metastases from other tumors.[21,22,58] The median age of onset of bronchopulmonary carcinoids in MEN1 is 40 to 45 years,[56,57] with an approximately equal sex distribution and no reports of childhood presentations. Bronchopulmonary carcinoids are rarely hormonally-active, and patients may present with respiratory symptoms (e.g., cough, dyspnea), although increasingly they are diagnosed radiologically in asymptomatic patients and are frequently small (e.g., <1 cm), and may be multiple and/or bilateral.[57] Histologically, 50% to 75% of tumors have features of typical carcinoids, 25% to 30% are atypical carcinoids, approximately 5% are large cell neuroendocrine carcinomas (LCNECs), and approximately 5% are small cell neuroendocrine carcinomas (SCLCs).[22,56-58] The majority of bronchopulmonary NETs appear to have an indolent course and are not associated with increased mortality. For example, the 10-year survival was greater than 70% in one study of 42 MEN1 patients with bronchopulmonary NETs, and none of the deaths were directly related to the bronchopulmonary NET.[57] Another study of 51 MEN1 patients with bronchial NETs reported seven disease-related deaths, and three of these were associated with atypical carcinoids, three with SCLC, and one with LCNEC, thereby suggesting that patients with atypical carcinoids had a worse prognosis than those with typical carcinoids,[56] although this observation has not been confirmed in other studies.[57]

The diagnosis of bronchopulmonary NETs relies on imaging, together with histological confirmation undertaken with biopsy or at time of surgical resection. Thoracic imaging by FDG-PET/CT in MEN1 patients revealed that greater than 25% of patients had pulmonary nodules, and that these increased in size during follow-up in approximately 30% of patients with one or more FDG-avid lesions, with the tumor doubling time of the FDG-avid nodules being significantly less than that of non–FDG-avid nodules (24 months vs. 48 months, respectively).[58] Tumor growth rates have also been reported to be higher in males than females (tumor doubling time 2.5 vs. 5.5 years, respectively).[57] Histologic analysis is important, as it allows tumor characterization and grading (i.e., differentiating typical/atypical lesions) of the tumor and facilitates exclusion of other diagnoses (e.g., metastases from other sites). At present, surgical resection is the recommended treatment, although the increasing detection of multiple small nodules and evidence of slow growth rates indicate that tumor surveillance may be appropriate in some individuals. However, future studies are required to determine the optimal treatment of these tumors and to develop improved prognostic markers to guide decision-making.

Gastric Carcinoids. Type II gastric carcinoids (also referred to as enterochromaffin-like cell carcinoids (ECLomas) are observed in 15% to 70% of MEN1 patients with coexistent hypergastrinemia and are frequently detected incidentally at the time of upper gastrointestinal endoscopy. The tumors are usually small (e.g., <1.5 cm) and multiple, and SRS may demonstrate increased uptake in the stomach. Little is known about the malignant potential of ECLomas and where appropriate surgical resection may be appropriate, although treatment with SSAs such as octreotide or lanreotide has resulted in regression of these tumors.[1]

Adrenocortical Tumors. The incidence of adrenocortical tumors in patients with MEN1 is reported to be 20% to 55%, although a higher frequency of approximately 75% has been reported when sensitive imaging modalities (e.g., EUS) are used.[13,20] Most patients are asymptomatic, as the majority of these tumors are nonfunctioning, but manifestations of hormonal hypersecretion occur in less than 10% of patients, and among these primary hyperaldosteronism and ACTH-independent Cushing syndrome are the most common.[20] Occasionally, hyperandrogenemia may occur in association with adrenocortical carcinoma, and the occurrence of pheochromocytoma is rare (<2% of lesions).[20] Adrenal tumors are typically diagnosed in adulthood, with a mean age of 46 years (range: 2–78 years),[20] although severe presentations in childhood are reported and include a 3-year old boy with adrenocortical carcinoma manifesting as precocious puberty.[20] Some MEN1 kindreds are reported to have increased occurrence of adrenal tumors, including adrenal carcinoma, and in such families increased vigilance in screening for adrenal tumors is required. Biochemical investigation (e.g., plasma renin and aldosterone concentrations, low-dose dexamethasone suppression test, and plasma or urinary catecholamines, and/or metanephrines) should be undertaken in patients with symptoms or signs suggestive of functioning adrenal tumors or with tumors greater than 1 cm.[1] The incidence of adrenocortical carcinoma is approximately 1% in MEN1 patients, but this is higher, at 10% to 15%, in MEN1 patients with adrenal tumors greater than 1 cm,[20] and interval surveillance imaging at least annually is therefore recommended in these patients (see Table 133.2).

Treatment. Consensus has not been reached about the management of MEN1–associated nonfunctioning tumors, as the majority are benign.[1] However, the risk for malignancy is increased if the tumor has a diameter greater than 4 cm, although adrenocortical carcinomas have been identified in tumors less than 4 cm in MEN1 patients.[1,20] Surgery is recommended for adrenal tumors that are greater than 4 cm in diameter, have atypical or suspicious radiologic features (e.g., increased Hounsfield units on unenhanced CT) and are 1 to 4 cm in diameter, or show significant measurable growth over a 6-month interval.[1,20] The treatment of functioning (i.e., secreting) adrenal tumors in MEN1 patients is similar to that for tumors occurring in non-MEN1 patients.

Central Nervous System Tumors. Central nervous system tumors including ependymomas, schwannomas, and meningiomas have been reported in MEN1 patients (see Table 133.1). Meningiomas are found in less than 10% of MEN1 patients; the majority of these were not associated with symptoms, and approximately 60% did not enlarge.[1] The treatment of patients with MEN1-associated meningiomas is similar to of non-MEN1 patients.

Lipomas. Subcutaneous lipomas may occur in 15% to 35% of patients with MEN1,[1] and frequently they are multiple (see Table 133.1). In addition, visceral, pleural, or retroperitoneal lipomas may occur in patients with MEN1. Management is conservative. However, when surgically removed for cosmetic reasons, they typically do not recur.

Facial Angiofibromas and Collagenomas. Studies of patients with MEN1 have revealed that multiple facial angiofibromas occur in approximately 20% to 90%, and collagenomas in 0% to great than 70%, of patients (see Table 133.1).[1] MEN1 angiofibromas are clinically and histologically identical to those observed in patients with tuberous sclerosis, with the exception that, in patients with MEN1, angiofibromas are also present on the upper lip and vermilion border of the lip, which are areas not involved in tuberous sclerosis. These cutaneous findings, which occur with a higher frequency in patients with MEN1, may provide a useful means for possible presymptomatic diagnosis of MEN1 in the relatives of a patient with MEN1. Treatment for these cutaneous lesions is usually not required.

Thyroid Tumors. Thyroid adenomas, colloid goiters, and carcinomas have been reported to occur in more than 25% of patients with MEN1.[1] However, the prevalence of thyroid disorders in the general population is high, and it has been suggested that the association of thyroid abnormalities in patients with MEN1 may be incidental and not significant.[1] The treatment of thyroid tumors in MEN1 patients is similar to that for non-MEN1 patients.

Breast Cancer. Female MEN1 patients are reported to have an increased risk of breast cancer (relative risk 2.3–2.8).[14,59] The majority of breast tumors are of ductal type, with mixed hormone receptor (estrogen receptor [ER[, progesterone receptor and human epidermal growth factor receptor) status.[14,59,60] Some of these tumors have reduced menin expression and loss of heterozygosity (LOH) of the *MEN1* locus, in keeping with a tumor suppressor function, and this is supported by *in vitro* studies demonstrating that menin may exert an antiproliferative effect in

TABLE 133.2 **Suggested Biochemical and Radiologic Screening in Individuals at High Risk for Developing Multiple Endocrine Neoplasia Type 1**

Tumor	Age to Begin (Years)	Biochemical Test (Plasma or Serum) Annually	Imaging Test (Time Interval)
Parathyroid	8	Calcium, PTH	None
Pancreatic NETs			
Gastrinoma	20	Fasting gastrin	None
Insulinoma	5	Fasting glucose, insulin	None
Other pancreatic NET	<10	Chromogranin A; gastrointestinal hormone profile[a] (e.g., pancreatic polypeptide, glucagon, vasoactive intestinal peptide)	MRI, or EUS (annually)
Anterior pituitary	5	Prolactin, IGF-1	MRI (every 3 years)
Adrenal	<10	None unless symptoms or signs of functioning tumor and/or tumor >1 cm on imaging	MRI or CT (annually with pancreatic imaging)
Thymic and bronchial carcinoid	15	None	CT[b] or MRI (every 1–2 years)

[a] Chromogranin A, pancreatic polypeptide, and glucagon concentrations can be elevated with nonfunctioning pancreatic NETs, but they have low sensitivity and specificity, and their value is debatable.

[b] CT scanning has been suggested for the detection of thymic and or bronchial carcinoids, but there is a concern over the potential for high cumulative doses of ionizing radiation, such that some have advocated utilizing "low-dose" CT protocols or MRI.

CT, Computed tomography; *EUS,* endoscopic ultrasound; *MRI,* magnetic resonance imaging; *NET*, neuroendocrine tumor; *PTH*, parathyroid hormone; *IGF-1*, insulin-like growth factor-1.

(Modified from Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab.* 2012;97: 2990–3011.)

normal mammary epithelium, although paradoxically it is reported to promote proliferation in sporadic ER-positive breast cancers.[60] The increase in breast cancer risk is not related to other known risk factors, and some centers now advocate breast screening in female MEN1 patients older than 40 years.[59]

Other Clinical Considerations

Gestational diabetes, hypertension, low birth weight, increased childhood mortality, and higher risk of venous thromboembolism (VTE) are reported to be associated with MEN1.[61-64] Thus, analysis of 96 pregnancies in 26 women with MEN1 revealed that maternal MEN1 was associated with increased prevalence of gestational diabetes, hypertensive disorders and low neonatal birth weight that resulted in admission to a neonatal unit (independent of maternal antenatal hypercalcemia), and that parathyroidectomy or medical management of hypercalcemia in women (n = 16) with confirmed hyperparathyroidism was only required in one and four pregnancies, respectively.[61,62] In addition, children born to a parent with MEN1 had an excess mortality by 15 years of age.[61] Patients did not have subfertility or infertility (except perhaps a subset of mothers with pituitary disease) or increased miscarriage rates.[63] Finally, MEN1 patients have been reported to have increased risk of VTE, as approximately 13% of 286 MEN1 patients had experienced a VTE event, which represents a lifetime incidence twice that of the general population, and approximately 80% of these events occurred in patients with pancreatic NETs, with only approximately 40% occurring in the perioperative period.[64]

MOLECULAR GENETICS

The *MEN1* Gene

The *MEN1* gene is located on chromosome 11q13 and consists of 10 exons that encode *menin,* a 610–amino acid protein[9,10] that regulates transcription, genome stability, cell division, and proliferation through direct associations with interacting protein partners or indirectly via the modulation of multiple cellular signaling pathways (Figs. 133.3 and Fig.133.4).[2a,3,4,11,65] Patients with MEN1 typically harbor a germline heterozygous *MEN1* mutation, which predisposes an individual to tumor development; however, tumorigenesis requires somatic inactivation of the wild-type *MEN1* allele, resulting in biallelic inactivation of the *MEN1* gene, consistent with Knudson's "two-hit" model of inherited tumorigenesis and a tumor suppressor function for menin in endocrine tissues. Somatic inactivation of the wild-type *MEN1* allele results from a large-scale deletion at the 11q13 locus, which manifests as LOH of the tumor DNA in greater than 90% of tumors, and in the remaining approximately 10% of tumors this inactivation may be caused by single nucleotide variants (SNVs) resulting in nonsense or missense amino acid substitutions, or small insertions or deletions (indels).[1,11]

The *MEN1* gene spans 7.7 kb of genomic DNA, and up to 19 different transcripts have been identified, of which 15 are predicted to result in protein products ranging from 146 to 652 amino acids. The main transcript of the *MEN1* gene is reported to be a 2.76-kb mRNA encoding a 610–amino acid isoform of menin, although the canonical transcript, which is 3.16 kb in length, elongates the reading frame by 15 nucleotides at the exon 2/intron 2 boundary and results in a protein isoform of 615 amino acids. *MEN1* transcripts and menin protein have been detected in all human tissues examined, with expression regulated at least in part by a 1400-bp region upstream of exon 2 that displays strong promoter activity, with additional mechanisms including an autoregulatory negative feedback loop, and microRNAs (e.g., miR24-1).[11] The *MEN1* gene and menin protein are highly evolutionary conserved in mammalian species (e.g., ~89% and ~97% DNA and protein identity, respectively, between human and mouse), while orthologs of menin are found in distant vertebrate (e.g., zebrafish) and invertebrate (e.g., *Drosophila*) species, although not in nematodes (e.g., *Caenorhabditis elegans*) or yeast (e.g., *Saccharomyces cerevisiae*). Recent analysis of population-level genetic data indicates that the *MEN1* coding region is under evolutionary constraint against genetic variation.[66]

KEY POINTS

- The *MEN1* gene encodes a 610–amino acid tumor suppressor protein, menin, which has tissue-specific and context-dependent roles in transcription regulation, genome stability, cell division, proliferation, and epigenetic regulation.

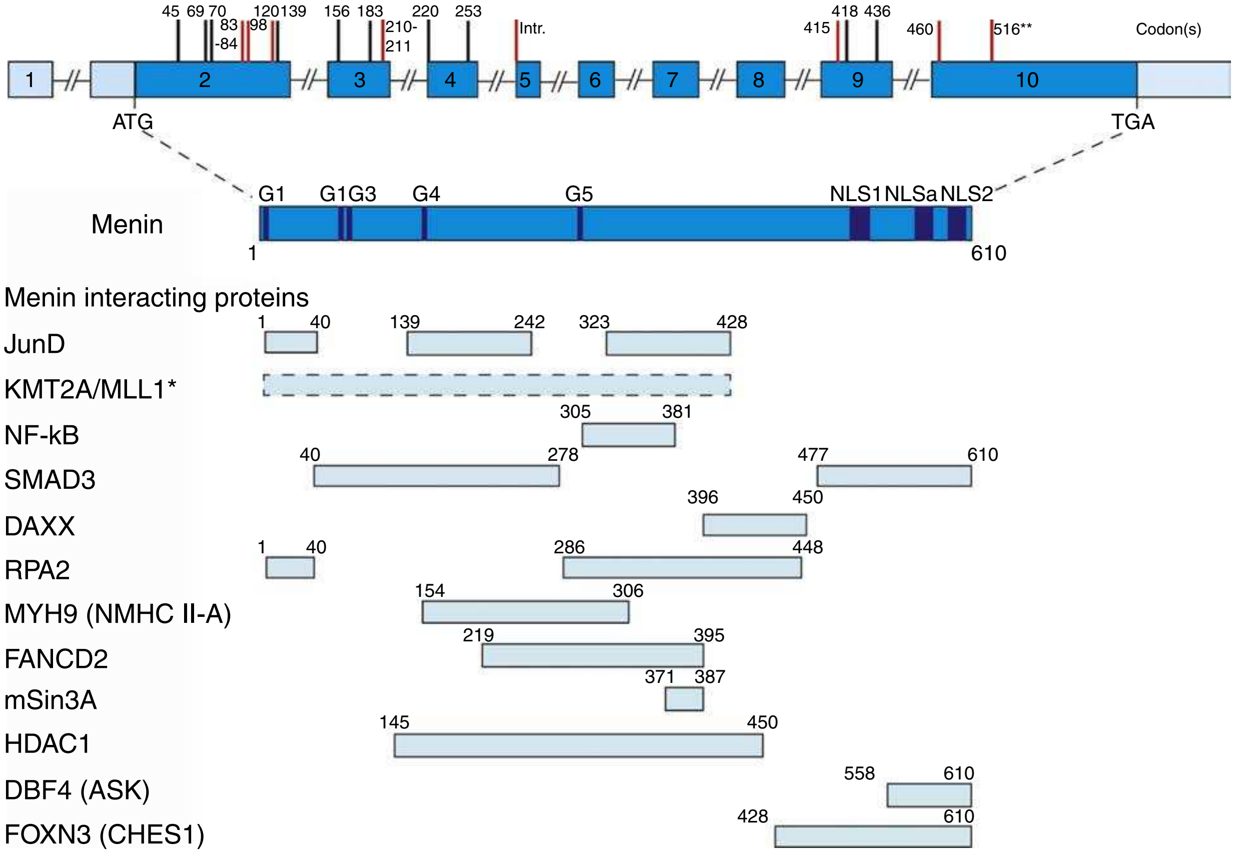

45 69 70 83 -84 98 120 139 156 183 210- 211 220 253 Intr. 415 418 436 460 516** Codon(s)
1 2 3 4 5 6 7 8 9 10
ATG
TGA
Menin
G1 G1 G3 G4 G5 NLS1 NLSa NLS2
1
610
Menin interacting proteins
JunD 1 40 139 242 323 428
KMT2A/MLL1*
NF-kB 305 381
SMAD3 40 278 477 610
DAXX 396 450
RPA2 1 40 286 448
MYH9 (NMHC II-A) 154 306
FANCD2 219 395
mSin3A 371 387
HDAC1 145 450
DBF4 (ASK) 558 610
FOXN3 (CHES1) 428 610

Fig. 133.3 **Schematic representation of the genomic organization of the *MEN1* gene, its encoded protein (menin), and regions that interact with other proteins. A,** The human *MEN1* gene consists of 10 exons that span more than 9 kb of genomic DNA and encodes a 610–amino-acid protein. The 1.83-kb coding region *(dark blue)* is organized into nine exons (exons 2 to 10) and eight introns (indicated by a *line* but not to scale). The sizes of the exons *(boxes)* range from 41 to 1297 bp, and that of the introns range from 80 to 1564 bp. The start (ATG) and stop (TGA) codons in exons 2 and 10, respectively, are indicated. Exon 1, the 5′ part of exon 2, and the 3′ part of exon 10 are untranslated *(light blue).* The promoter region I is located within a few 100 bp upstream of exon 2. The sites of nine germline mutations which have been reported to occur with a frequency >1.5% are represented by red bars with the affected codon number(s) shown. These germline mutations, which collectively represented approximately 20% of all reported mutations include: c.249_252delGTCT (codons 83–84); c.292C>T (codon 98); c.358_360delAAG (codon 120); c.628_631delACAG (codon 210-211); c.784-9G>A (intron 4); c.1243C>T (codon 415); c.1378C>T (codon 460); c.1546delC (codon 516); and c.1546_1547insC (codon 516).** Indicates the two different recurrent germline mutations affecting codon 516. The respective frequency of these mutations is dependent on the populations studied and varies by geographical region. The frequent occurrence of the same germline mutations in apparently unrelated kindreds may indicate the presence of founder mutations or regions of genomic DNA more prone to mutation. Codons reported to be affected by more than four different germline mutations are represented by black bars with respective codon numbers. The locations of these mutations are reported with reference to the transcript encoding the 610–amino acid menin protein (i.e., NM_130799), while it is important to note that the majority of contemporary variant databases report such *MEN1* variants according to the longer transcript encoding the 615–amino acid protein (i.e., NM_000244.3). **B,** Menin has three nuclear localization signals (NLSs) at codons 479–497 (NLS1), 546–572 (NLSa), and 588–608 (NLS2) indicated by *dark blue bars* and five putative guanosine triphosphatase (GTPase) sites (G1–G5) indicated by *dark blue.* **C,** Menin regions that have been implicated in the binding to different interacting proteins are indicated by *light blue boxes.* Illustrative examples of some key interactions are shown. *Recent crystal structure studies indicate that MLL1 and JunD utilize the same binding pocket of menin; the individual domains required for the JunD/menin interaction have been characterized, whereas those for the MLL1/menin interaction remain to be elucidated. (Modified and updated from Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 [MEN1]: Analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat.* 2008;29:22–32.)

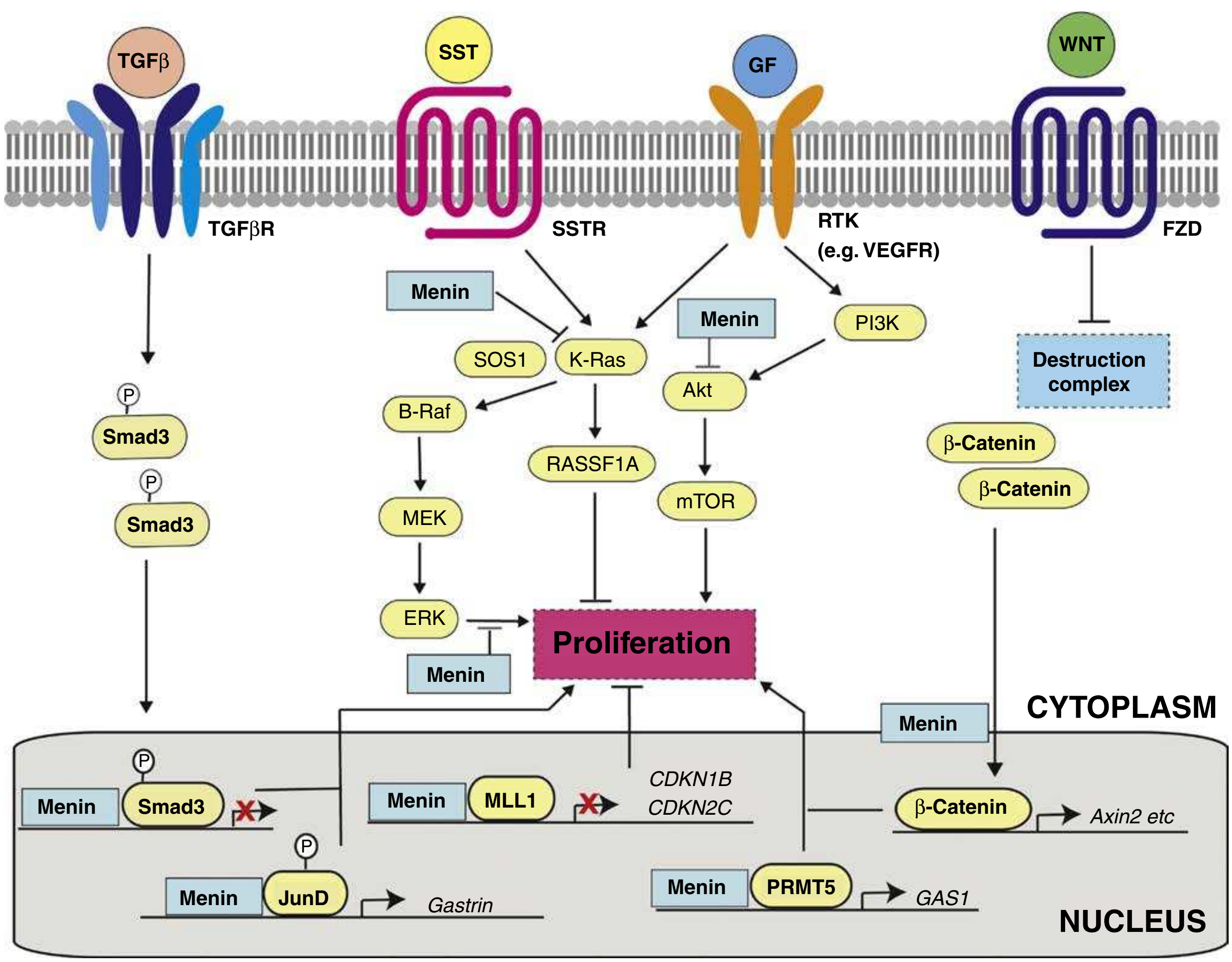


Fig. 133.4 Roles of menin in cell proliferation. Loss of menin expression *(blue boxes)* in endocrine tissues may result in increased cell proliferation by multiple pathways. In the nucleus, loss of menin results in disruption of its interaction with transcription factors JUND and PRMT5, which lifts the transcriptional repression of target genes including *Gastrin* and *GAS1*, respectively; binding to MLL1, MLL2, and SMAD3 (a TGFβ signaling component) to promote transcription of target genes; and ability to regulate the WNT pathway, as β-catenin is no longer prevented from entering the nucleus by menin, thereby enabling transcription of WNT pathway target genes. Interactions with additional transcription factors and chromatin-modifying protein complexes may further modulate oncogenic signaling pathways. In the cytoplasm, loss of menin reduces its inhibitory actions on the mTOR pathway by binding to AKT (downstream of PI3K, part of the RTK signaling pathway) and preventing its translocation to the plasma membrane; and KRAS-induced proliferation (by possible inhibition of ERK-dependent phosphorylation [P] and prevention of the interaction between SOS1 and KRAS). All pathways affect proliferation, which involves both nuclear and cytoplasmic mechanisms (shown in the cytoplasm only). Examples of ligands and their respective receptors are shown. *FZD*, Frizzled; *GF*, growth factor; *TGFβR*, transforming growth factor β receptor; *RTK*, receptor tyrosine kinase; *SST*, somatostatin; *SSTR*, somatostatin receptor. (Modified from Frost M, Lines KE, Thakker RV. Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol* (2018); 14;216–227.)

MEN1 Germline Mutations

To date, more than 1200 germline mutations of the *MEN1* gene (see Fig. 133.3) have been reported in patients with MEN1, or those with associated tumors, and amongst these greater than 600 different mutations are reported.[47] The majority of pathogenic variants (~70%) are predicted to result in a loss of function through premature truncation of the menin protein (i.e., frameshift indels (35%–45%), nonsense SNVs (~15%–20%), or splice site variants (~10%), whereas the remainder result from missense SNVs (20%–25%), in-frame indels (~5%), or gross deletions of all or part of the *MEN1* gene (1%–2.5%).[11,67] The mechanisms leading to large-scale deletions are poorly understood, although a study of a patient with an entire *MEN1* gene deletion noted *Alu* repetitive elements flanking the respective breakpoints, indicating that an *Alu/Alu*-mediated rearrangement may have been responsible for the deletion.[68] Deletions may occur in the noncoding regions of the *MEN1* gene, as illustrated by a study of identical twins that identified a 596-bp deletion involving the 5′ untranslated region (UTR).[23] The majority (~90%) of *MEN1* mutations are inherited from a parent, but approximately 10% occur *de novo*, and studies using next-generation sequencing (NGS) have also identified individuals who are mosaic for *MEN1* mutations.[69]

The mutations affecting the *MEN1* coding region occur throughout the length of the gene, with no obvious clustering, although several variants are reported multiple times in apparently independent kindreds. In addition, at least 11 codons (45, 69, 70, 139, 156, 183, 220, 253, 418, 436, 516) are reported to be affected by more than four different *MEN1* pathogenic variants, indicating that specific regions of the *MEN1* gene may be more prone to mutation, and it is notable that some of these mutations occur within repetitive DNA sequences, consistent with a replication-slippage model of mutagenesis.[11]

A consistent genotype-phenotype correlation has not been observed in MEN1, and studies of several large MEN1 families each harboring the same *MEN1* mutation have shown that members of the respective families develop a different range of tumors.[1,2,11] However, a correlation between *MEN1* mutations affecting the menin checkpoint kinase 1 interaction domain (i.e., codons 428–610) and a higher frequency of aggressive pancreatic NETs and death is reported, while those involving the JunD-interacting domains are reported to be associated with an increased risk of premature death, although these studies require validation. Some kindreds with *MEN1* mutations do not appear to manifest the full clinical phenotype; for example, kindreds harboring the Burin (or "prolactinoma") *MEN1* variant (i.e., Tyr312Ter) are reported to have a high frequency of prolactinoma but a low incidence of gastrinoma; whereas an absence of GH-secreting pituitary adenomas is observed in a large Tasmanian kindred with a splice site variant (c.444-CC>G), thereby indicating that additional genetic factors may influence MEN1 clinical expression. A minority of families with *MEN1* mutations are reported to only manifest parathyroid tumors, referred to as familial isolated hyperparathyroidism (FIHP). A higher prevalence of missense variants are reported in FIHP kindreds compared with MEN1 (~40% vs. ~25%, $P < .01$), with several of these FIHP-associated missense mutations reported to retain menin protein stability and biological activity, potentially consistent with a milder phenotype.[11] However, some FIHP kindreds harbor protein-truncating mutations (e.g., nonsense, frameshift), indicating that genetic modifying influences are also likely to be present. Other studies have reported decreases in the strength of intrafamilial correlation for pituitary, adrenal, and thymic NETs with genetic dilution, thereby further supporting the role of likely genetic modifiers.[56,70]

Between 5% and 25% of patients with a clinical diagnosis of MEN1 do not harbor mutations in the *MEN1* coding region, and a minority of these individuals may have mutations in the promoter or UTRs, or alternatively represent phenocopies with mutations in other genes (see later). However, the possibility that *MEN1* mutations may not have been detected by use of conventional methods should be considered, as a study using whole-genome sequencing (WGS) has reported identifying *MEN1* mutations in greater than 20% of such patients ($n = 14$).[71] Recent studies indicate that patients with a clinical diagnosis of MEN1 (i.e., at least two MEN1 manifestations), but negative *MEN1* genetic testing, have a different clinical course to those with *MEN1* mutations, including a later age of first tumor diagnosis, a higher prevalence of single-gland parathyroid disease, the rare occurrence of three MEN1 manifestations, and a greater life expectancy,[72,73] which all suggest that a substantial proportion of such patients may have the cooccurrence of two sporadic tumors rather than a hereditary MEN1 syndrome.

MEN1 Germline Coding Variants

Five common (allele frequency [AF] >0.5%) germline coding variants are observed in the *MEN1* gene, including three synonymous and two nonsynonymous variants that result in p.Thr546Ala and p.Arg176Gln in the canonical transcript. However, a wide spectrum of rare coding variants (AF <0.5%) are observed in the background population, and greater than 200 nonsynonymous *MEN1* SNVs are reported in the GnomAD dataset (http://gnomAD.broadinstitute/org/), with the majority of these being unlikely to be of clinical significance, although their presence may confound interpretation of genetic test results.[66,74]

MEN1 Somatic Mutations

The role of the *MEN1* gene in sporadic endocrine tumorigenesis has been investigated, and somatic biallelic inactivation of the *MEN1* gene is reported at varying frequencies in other endocrine tumors. For example, somatic *MEN1* mutations are reported in 20% to 35% of sporadic parathyroid tumors, approximately 45% of nonfunctioning pancreatic NETs, 40% of gastrinomas, 0% to 15% of insulinomas, 15% to 20% of bronchopulmonary carcinoids, 3% to 5% of anterior pituitary adenomas, less than 3% of small intestinal NETs, and less than 3% of adrenocortical tumors. These somatic mutations occur throughout the coding region and comprise protein-truncating (i.e., nonsense, frameshift) and missense variants. Currently, identifying somatic *MEN1* mutations in tumor tissue has little clinical utility, but in the future it may help for prognostication, for example, the presence of somatic *MEN1* mutations in pancreatic NETs and absence of chromosomal instability gene mutations (e.g., *DAXX*, *ATRX*), have been reported to be associated with favorable outcomes[75] or stratification for personalized treatment approaches, for example, as reported for selpercatinib in *RET*-altered thyroid cancers.[76]

MEN1 Phenocopies and Mutations in Other Genes

Phenocopies are reported in 5% to 10% of MEN1 kindreds and may occur in different clinical settings. For example, phenocopies have been reported in the context of familial MEN1, in which a member of the kindred has a clinical manifestation of MEN1 (e.g., pituitary adenoma or primary hyperparathyroidism) but is subsequently found not to harbor the familial *MEN1* mutation (i.e., likely representing a coincidental sporadic occurrence of the tumor). Patients with a clinical diagnosis of MEN1 (i.e., with two or more MEN1-associated tumors) may also represent phenocopies, whereby they do not manifest a *MEN1* mutation, but instead have a mutation in another gene, more typically associated with a different disease.[1] Such genes include *CDC73*, which encodes the tumor suppressor parafibromin, mutations of which are associated with the hyperparathyroidism-jaw tumor syndrome,[71,77] *CASR*, which encodes the calcium-sensing receptor, mutations of which are associated with familial hypocalciuric hypercalcemia,[71] and *AIP* (also located at chromosome 11q13), which encodes the aryl hydrocarbon interacting protein, mutations of which are associated with familial isolated pituitary adenoma. Less than 5% of patients with a clinical diagnosis of MEN1 who do not harbor a mutation in the *MEN1* gene may instead harbor mutations in *CDKN1B*, located on chromosome 12p13, which encodes the cyclin-dependent kinase inhibitor p27$^{\mathrm{kip}}$ and results in the associated disorder MEN4.[5,6] To date, fewer than 50 MEN4 patients (from <20 kindreds) have been reported, such that the clinical phenotype and its associated penetrance are not fully characterized.[6] Patients with MEN4 are reported to predominantly express parathyroid and pituitary tumors in association with pancreatic, gastrointestinal, bronchial, and cervical NETs, nonfunctional adrenal tumors, papillary thyroid cancer, lipoma, and breast cancer.[6] Primary hyperparathyroidism occurs in approximately 80% of MEN4 cases, but at a later age than in MEN1 (e.g., typically after age 40 years, but as young as 15 years). Pituitary adenomas, which include GH- and ACTH-secreting adenomas, prolactinomas, and nonfunctioning adenomas, are reported in 30% to 40% of MEN4 patients, and duodenal-pancreatic NETs in 5% to 30% of MEN4 cases.[6] The investigation and treatment of these tumors are similar to those described for MEN1. The germline *CDKN1B* mutations observed in MEN4 include protein-truncating (i.e., nonsense, frameshift, splice site) and missense variants, with the majority resulting in reduced levels of $p27^{kip}$ protein, or altered protein function, consistent with a tumor suppressor function. Germline missense *CDKN1B* variants have also been reported in patients with apparently sporadic parathyroid adenomas, although the relative high background frequency of rare nonsynonymous variants in the Cyclin Dependent Kinase Inhibitor (CDKI) genes suggests that some of these variants represent benign alleles.[66] The $p27^{kip}$ protein is involved in cell cycle regulation, and particularly in controlling the G1/S-phase checkpoint by binding to cyclin E/cdk2, thereby inhibiting cell cycle progression. Downregulation of $p27^{kip}$ is observed in several cancers (e.g., stomach, breast, colon), while somatic *CDKN1B* mutations are observed in a small percentage of small intestinal NETs, in keeping with a tumor suppressor function.

Thus, the possibility of a phenocopy or alternate genetic diagnosis should be considered in patients presenting with typical or atypical manifestations of MEN1 in whom a *MEN1* mutation is not found, although it is recommended that all kindred members in families with a known *MEN1* mutation undergo genetic testing, irrespective of the presence or absence of clinical features.[1]

MEN1 Mutational Analysis in Clinical Practice

MEN1 mutational analysis is helpful in clinical practice in several ways, including: (1) confirmation of the clinical diagnosis; (2) identification of family members who harbor the *MEN1* mutation and require screening for tumor detection and early/appropriate treatment; (3) identification of family members who do not harbor the familial germline *MEN1* mutation and can therefore be reassured and alleviated of the anxiety burden of developing future tumors; and (4) facilitation of preconception genetic counseling and, in some settings, preimplantation genetic diagnosis (PGD) (Table 133.3).

TABLE 133.3 Suggested Approach for Multiple Endocrine Neoplasia Type 1 Mutation Analysis in a Clinical Setting

Value in Clinical Setting

- Aid in confirming the diagnosis
- Identify mutation carriers in a family for screening and development of tumors, thereby facilitating early treatment
- Identify the 50% of family members who do not harbor the *MEN1* mutation, thereby alleviating the anxiety and burden of disease from them and their progeny
- Facilitate preconception genetic counseling (and the potential for preimplantation genetic diagnosis)

Who should be tested?

In an index case

- Meeting the clinical criteria for multiple endocrine neoplasia type 1 (i.e., two or more MEN1-associated tumors or a diagnosis of familial MEN1)
- Suspicious (i.e., multiple parathyroid adenomas before 40 years of age; recurrent hyperparathyroidism; gastrinoma or multiple pancreatic neuroendocrine tumors at any age) or atypical for MEN1 (i.e., development of two nonclassical MEN1-associated tumors, e.g., parathyroid and adrenal tumor)

A first-degree relative of family member with known MEN1 mutation

- Asymptomatic first-degree relative
- First-degree relative with familial MEN1 (i.e., one MEN1-associated tumor)

When should testing be undertaken?

- As early as possible (e.g., <5 years of age for asymptomatic individuals)

Where should testing be performed?

- In an accredited department/laboratory undertaking DNA testing of *MEN1* gene

(Modified from Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab.* 2012;97: 2990–3011.)

Current guidelines[1] recommend that *MEN1* mutational analysis should be undertaken in: (1) an index case with two or more of the major MEN1-associated endocrine tumors (i.e., parathyroid, pancreatic, or pituitary tumors); (2) all first-degree relatives of a known *MEN1* mutation carrier, irrespective of whether they are symptomatic or manifest associated clinical features; (3) patients with suspicious or atypical MEN1, which includes individuals with parathyroid adenomas occurring before 30 years of age or multigland parathyroid disease, gastrinoma, or multiple pancreatic NETs at any age, or individuals who have two or more MEN1-associated tumors that are not part of the classical triad of parathyroid, pancreatic islet, and anterior pituitary tumors (e.g., parathyroid tumor plus adrenal tumor) (see Table 133.3).[1] Several studies have evaluated these recommendations in different clinical settings and confirmed that tumor multiplicity, a family history of MEN1, and/or young age of tumor onset are the strongest predictive indicators for the individual to have an *MEN1* mutation. Thus, one study of 200 patients with endocrine tumors has reported that *MEN1* mutations were present in greater than 70% of individuals with two or more of the major MEN1-associated tumors and a family history of relevant tumors; approximately 60% of patients with at least one of the major endocrine tumors and a first-degree relative with one of the major tumors; and only approximately 5% of patients with sporadic (i.e., nonfamilial) MEN1-associated tumors, although all of these patients had an onset of disease before the age of 30 years, and/or multiple tumors, thereby highlighting the importance of genetic testing in younger patients. Another study of 292 patients reported identification of *MEN1* mutations in 80% of index cases who presented with three MEN1-related tumors, and this increased to greater than 90% if there was a history of a relative with an associated MEN1 tumor; but only 15% of patients who presented with a single MEN1-associated tumor, and this decreased to 0% if family history of endocrine neoplasia was absent. Furthermore, the strongest predictors for the presence of an *MEN1* mutation from a study of 365 patients with sporadic endocrine tumors (i.e., patients without a clinical diagnosis of MEN1, or known family history of MEN1), were reported to be recurrent or multigland primary hyperparathyroidism; nonrecurrent primary hyperparathyroidism with an additional risk factor, such as relevant family history or young age of onset; and occurrence of NETs of the gastrointestine, pancreas, bronchi, or thymus.[78] In addition, a study of 205 patients with primary hyperparathyroidism and an associated family history of primary hyperparathyroidism, which identified a *MEN1* mutation in approximately 45% of cases, found that multigland disease, male sex, and age less than 45 years were each independent predictors for the presence of a germline *MEN1* mutation (odds ratios of 14, ~1.5, and ~8, respectively).[78] However, patients with acromegaly and coexistent primary hyperparathyroidism[79] or patients presenting with pituitary adenomas and pheochromocytoma/paraganglioma (i.e., "atypical" MEN1) have been reported to have a low occurrence (0%–5%) of germline *MEN1* mutations.[80]

Once a genetic diagnosis of MEN1 has been established in an index case, predictive testing should be offered to first-degree relatives, and current guidelines recommend that this should be undertaken at the earliest opportunity.[1] Delays in diagnosis of MEN1 are common, and a median lag time of approximately 3.5 years has been reported between index case diagnosis and that of subsequent family members, leading to higher rates of metastatic disease in family members.[81,82] The delay in diagnosis may be because of multiple reasons, which include the relevant information not being available to the patient (or doctors) because they may be unaware of relevant family history because of medical information not being shared between family members (and/or medical teams); geographical or social separation of family members; relatives dying prematurely of causes unrelated to the disorder; or incomplete or variable penetrance of manifestations.[74] There are also potential ethical issues to consider for predictive testing, and therefore individuals should receive appropriate genetic counseling, especially as many MEN1 patients are fearful of disease occurrence, both for themselves and their relatives, and this may be compounded by investigations and regular surveillance screening.[83]

Predictive testing in young children, which requires parental/guardian consent, is important because MEN1-associated tumors may develop at an early age, and the current guidelines therefore recommend that genetic testing of asymptomatic children should be undertaken in the first decade of life, and ideally before 5 years of age. Preconception genetic counseling can also be provided to individuals at risk of transmitting MEN1 to their future offspring, and advances in genetic testing combined with *in vitro* fertilization now facilitate PGD, in which direct sequencing of DNA from a small number of cells removed from the early blastocyst facilitates implantation of embryos without the *MEN1* mutation.[74]

Genetic testing for MEN1 should be undertaken in accredited genetic diagnostic laboratories and include DNA sequence analysis of all exonic and splice site regions of the *MEN1* gene, and, in those without *MEN1* mutations, multiplex ligation-dependent probe amplification (MLPA) should be considered to detect partial or whole gene deletions. The identified DNA sequence variants (i.e., SNVs, indels) should then be assessed for variant pathogenicity using validated methods; many centers adopt the American College of Medical Genetics and Genomics (ACMG) guidelines.[84] These ACMG guidelines consider multiple variant- and gene-specific factors in categorizing variants into benign or pathogenic categories, but this may not be able to achieve a definitive classification (i.e., benign/likely benign or pathogenic/likely pathogenic), but instead yield a "variant of uncertain significance" status.[84] This situation mostly arises for rare missense variants for which there is either insufficient or conflicting evidence to support a definitive classification.[66,74] A recent report that places additional weight on phenotype data to improve the accuracy of genetic test reporting for MEN1 has recommended a MEN1-specific adjustment to the ACMG criteria.[85] Other approaches aimed at improving *MEN1* variant interpretation include *in silico* and/or bioinformatic analysis modeling; for example, thermodynamic analysis of protein structure was used to evaluate the impact of 338 missense *MEN1* variants, and this facilitated discrimination between pathogenic and benign variants.[86] Finally, if initial genetic testing by DNA sequence analysis and MLPA has failed to identify a *MEN1* mutation in a patient for whom there is a strong clinical suspicion for MEN1, then *MEN1* mosaicism and repeat testing should be considered using a different method such as NGS, for example, WGS with a greater depth of DNA sequencing, which has been reported to successfully identify *MEN1* mutations in greater than 20% of cases (three of 14).[69,71]

SURVEILLANCE OF INDIVIDUALS AT RISK OF MULTIPLE ENDOCRINE NEOPLASIA TYPE 1 TUMORS

Current guidelines recommend that all individuals at high risk of MEN1 (i.e., *MEN1* mutation carriers) undergo periodic clinical, biochemical, and radiologic screening to facilitate the early detection and treatment of tumors, with the aim of reducing their associated morbidity and mortality (Table 133.2, Fig. 133.5), although high-quality evidence to support this approach remains to be established.[1] However, most clinicians implement such surveillance protocols, although there are debates about the extent and frequency of screening, as well as the age at which it should commence. These controversies reflect many factors, including the variable clinical expression and age-related penetrance of MEN1-associated tumors; the lack of genotype-phenotype correlations; the observations that some manifestations cluster at higher frequency in some individual kindreds[70]; a possible genetic anticipation for clinical manifestations[87]; a lack of high-quality evidence to guide treatment decisions; the possibility of iatrogenic harms from screening methods; and the psychological burdens associated with frequent investigation.[16,83] The current guidelines for surveillance suggest that individuals at higher risk for MEN1 (i.e., *MEN1* mutation carriers) undergo clinical and biochemical screening at least once per annum and also have baseline pituitary and abdominal imaging (e.g., MRI or CT scans), which should be repeated at 1- to 3-year intervals (Table 133.2). A similar schedule is likely to be appropriate for MEN4. Screening should be repeated annually, because the disease may not manifest in some individuals until the eighth decade. Screening should commence in early childhood, because the disease has developed in some individuals by 5 years of age, with imaging for pancreatic and pituitary lesions from around 10 years of age. Some centers consider that radiological screening in asymptomatic *MEN1* mutation carriers for pancreatic NETs should be delayed until 16 years of age, although children with nonfunctioning NETs great than 2 cm have been identified at as young as 12 years.[45] Screening for thymic carcinoids is likely to require cross-sectional imaging with CT every 1 to 2 years, thereby resulting in high cumulative doses of ionizing radiation, and the risks of this need to be appraised and discussed with the patients, a minority of whom will develop thymic carcinoids with a potentially rapid and aggressive course. MRI and "low-dose" CT have been suggested as alternative approaches, but their suitability for screening thymic carcinoids is not established. Screening for breast cancer in women with MEN1 from age 40 years has also been proposed, although it should be noted that this is undertaken for all women aged 50 to 71 years in some countries. Further studies are required to validate these screening approaches, and it may be appropriate to individualize clinical follow-up based on the availability of local resources and patient-specific factors (e.g., age, comorbidities, family history of early-onset malignant disease).

KEY POINTS

- Inheritance of a germline *MEN1* mutation predisposes an individual to developing a tumor that arises after a somatic mutation, and genetic testing combined with periodic clinical, biochemical, and radiological screening has facilitated earlier detection and treatment of MEN1-associated tumors.

FUNCTION OF THE MULTIPLE ENDOCRINE NEOPLASIA TYPE 1 PROTEIN, MENIN, AND FUTURE TARGETED TREATMENTS

Menin is a 67-kDa ubiquitously expressed nuclear protein that can also be found in the cytoplasm of dividing cells, where it localizes to the mitotic spindle and spindle poles in early mitosis. Menin is reported to interact or associate with greater than 75 proteins and to function as a scaffold protein, which is a type of protein that simultaneously interacts (or binds) with two or more member proteins of a signaling pathway, thereby tethering them into larger regulatory complexes that enhance the efficiency and fidelity of the pathways.[3,4,88,89] Furthermore, analysis of the crystal structure of menin has revealed that it adopts a "curved left hand" structure, with a deep pocket within

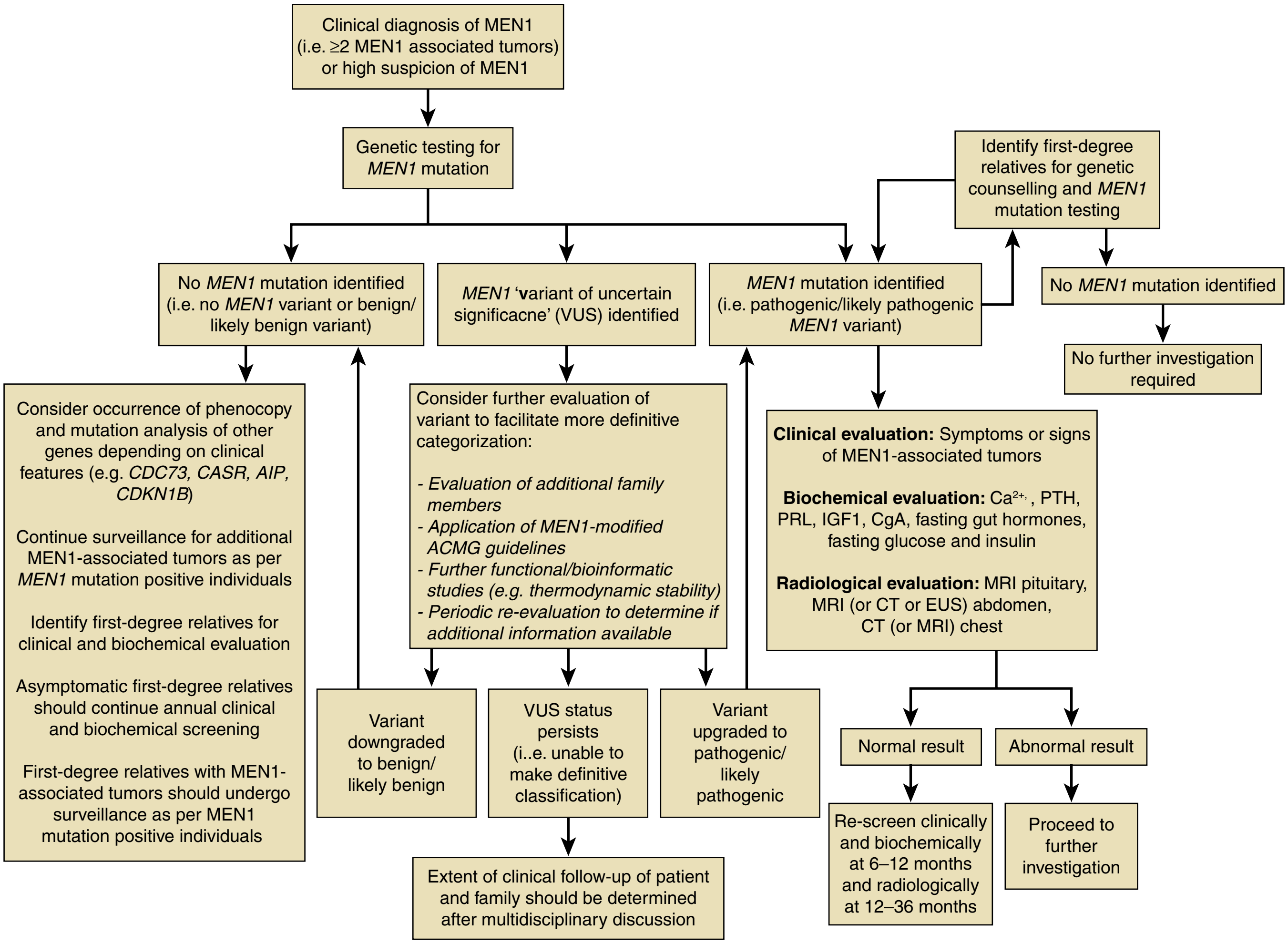

Clinical diagnosis of MEN1 (i.e. ≥2 MEN1 associated tumors) or high suspicion of MEN1
Genetic testing for MEN1 mutation
Identify first-degree relatives for genetic counselling and MEN1 mutation testing
No MEN1 mutation identified (i.e. no MEN1 variant or benign/ likely benign variant)
MEN1 'variant of uncertain significacne' (VUS) identified
MEN1 mutation identified (i.e. pathogenic/likely pathogenic MEN1 variant)
No MEN1 mutation identified
No further investigation required
Consider occurrence of phenocopy and mutation analysis of other genes depending on clinical features (e.g. CDC73, CASR, AIP, CDKN1B)
Continue surveillance for additional MEN1-associated tumors as per MEN1 mutation positive individuals
Identify first-degree relatives for clinical and biochemical evaluation
Asymptomatic first-degree relatives should continue annual clinical and biochemical screening
First-degree relatives with MEN1-associated tumors should undergo surveillance as per MEN1 mutation positive individuals
Consider further evaluation of variant to facilitate more definitive categorization:
- Evaluation of additional family members
- Application of MEN1-modified ACMG guidelines
- Further functional/bioinformatic studies (e.g. thermodynamic stability)
- Periodic re-evaluation to determine if additional information available
Clinical evaluation: Symptoms or signs of MEN1-associated tumors
Biochemical evaluation: Ca2+, PTH, PRL, IGF1, CgA, fasting gut hormones, fasting glucose and insulin
Radiological evaluation: MRI pituitary, MRI (or CT or EUS) abdomen, CT (or MRI) chest
Variant downgraded to benign/ likely benign
VUS status persists (i..e. unable to make definitive classification)
Variant upgraded to pathogenic/ likely pathogenic
Normal result
Abnormal result
Re-screen clinically and biochemically at 6–12 months and radiologically at 12–36 months
Proceed to further investigation
Extent of clinical follow-up of patient and family should be determined after multidisciplinary discussion

Fig. 133.5 **An approach to screening in multiple endocrine neoplasia type 1 (*MEN1*).** Index cases, or individuals in whom there is a high suspicion of clinical MEN1 (e.g., multigland parathyroid disease, parathyroid + adrenal tumor), should be offered genetic counseling and *MEN1* mutation testing. Mutation testing should also be offered to those with familial MEN1 (i.e., individuals with one MEN1-associated tumor and a first-degree relative with a known *MEN1* mutation). The identification of a germline *MEN1* mutation should prompt entry into a periodic clinical, biochemical, and radiologic screening program. At the same time, first-degree relatives should be identified and offered genetic counseling and *MEN1* mutation testing. Individuals who have inherited the *MEN1* mutation should enter periodic screening, even if asymptomatic. First-degree relatives who have not inherited the *MEN1* mutation require no further follow-up and may be alleviated of the anxiety associated with the development of MEN1-associated tumors. Identification of a variant of uncertain significance (VUS) represents a clinical challenge, whose resolution to establish it as a pathogenic/likely pathogenic variant or benign/likely benign variant requires further evaluation that may include assessment of the strength of the clinical phenotype in the patient; evaluation of additional family members for relevant clinical features to determine if the variant cosegregates with affected individuals; and/or additional bioinformatic analyses. The management of patients in whom the VUS cannot be established to be pathogenic/likely pathogenic or benign/likely benign requires careful consideration that should involve the multidisciplinary team, patients, and family members. For index cases in whom a *MEN1* mutation, which includes testing for partial or whole gene deletions *(asterisk),* is not identified, additional genetic testing may be required, depending on the specific clinical features. This may include examination for mutations in genes associated with familial parathyroid syndromes, including *CDC73* associated with the hyperparathyroidism-jaw tumor syndrome syndrome and *CASR* associated with familial hypocalciuric hypercalcemia; or *CDKN1B* and *AIP*, which are rarely identified in those with clinical MEN1. Up to 10% of kindreds with clinical MEN1 may harbor phenocopies, emphasizing the importance of accurate genetic evaluation. For MEN1 kindreds in whom no *MEN1* mutation is identified, a pragmatic approach is to offer clinical, biochemical, and radiologic screening to those with clinical manifestations of disease and to offer annual clinical and biochemical screening to asymptomatic first-degree relatives. *ACMG*, American College of Medical Genetics and Genomics; *Ca*2+, calcium; *PTH,* parathyroid hormone; PRL, prolactin; IGF-1, insulin-like growth factor-1; *CgA*, chromogranin A; *MRI,* magnetic resonance imaging; *CT,* computer tomography; *EUS,* endoscopic ultrasound. (Modified from Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 [MEN1]. *J Clin Endocrinol Metab.* 2012;97:2990–3011.)

the central "palm" region that acts as a binding site for several interacting partners, such as the transcription factor JunD and the histone methyltransferase MLL1.[3,12] Thus, menin, by its direct interactions with transcription factors or membership of epigenetic regulatory complexes, has central roles in transcription of genes involved in cell proliferation.[3,4] Menin is also reported to be involved in additional cellular functions, including regulation of genome stability, DNA repair, cell division, cell signaling, and cell motility, that are cell-type and tissue-specific and enable it to modulate the activity of several canonical signaling pathways, which are implicated in endocrine and nonendocrine tumorigenesis (Fig. 133.4).[4,89] For example, menin has been reported to act as both a negative and positive regulator of Wnt/β-catenin signaling.[4,89] Thus, in embryonic fibroblasts and insulinoma cells from *Men1* knockout mice, menin acts as a negative regulator of Wnt signaling, facilitating export of β-catenin out of the nucleus, such that menin depletion results in nuclear accumulation of β-catenin and Wnt signaling activation.[4] Similar observations (i.e., Wnt activation following menin loss) are reported in human and murine pancreatic NETs and provide the rationale for evaluating small-molecule Wnt inhibitors in preclinical MEN1 studies.[88] In contrast, menin is also reported to augment Wnt signaling in pancreatic endocrine cells, via a direct interaction with β-catenin, revealing apparent paradoxical activities.[4] Menin is reported to have similarly complex and tissue-specific roles in regulating the K-RAS signaling pathways in the pancreas.[4,90] K-RAS activation is generally considered to be an oncogenic signaling pathway that is associated with cellular proliferation, as observed in pancreatic ductal adenocarcinoma, which arises from pancreatic exocrine cells.[90] However, in pancreatic endocrine cells K-RAS activation is associated with suppression of proliferation, which is mediated by a menin-dependent activation of the Ras effector RASSF1A pathway, with simultaneous suppression of the proproliferative RAF/MAPK cascade.[90] In this setting, menin loss results in a shift in balance of these opposing pathways, leading to preferential activation of RAF/MAPK signaling, thereby releasing the "brake" on proliferation.[90] Additional signaling pathways modulated by menin include TGFβ, BMP, AKT, Hedgehog, and MYC (Fig 133.4).[4]

Men1 mouse models generated as conventional and conditional "knockout" murine models have also been used to study these tumorigenic pathways *in vivo* and also in preclinical studies.[89,91-93] *Men1* mouse models have been used to investigate novel therapies for MEN1-associated NETs, including SSAs, epigenetic modifying agents, *Men1* gene therapy, and Wnt antagonists, as well as to evaluate the potential for drug repurposing.[4,89] The efficacy of SSAs has been assessed in different models. For example, pasireotide, a pan-SSTR agonist targeting SSTR1, 2, 3, and 5, was demonstrated to have antiproliferative and proapoptotic activity against pancreatic NETs and pituitary adenomas in conventional and conditional (Pdx-1-Cre) *Men1* models,[89,94] while both pasireotide and lanreotide had chemoprophylactic activity in *Men1* models that reduced tumor development.[94,95] The epigenetic modifying agent JQ1, a member of the bromo and extra terminal domain family of proteins that bind to acetylated histone residues, has been demonstrated to have antiproliferative, antisecretory, and proapoptotic activity in NET cell lines *in vitro,* and also *in vivo* on pancreatic NETs that developed in β cell–specific *Men1* knockout mice.[93,96] Proof-of-concept for *Men1* gene replacement therapy has been established for the treatment of MEN1-associated tumors. The administration of a replication-incompetent adenoviral vector expressing murine *Men1* cDNA directly into pituitary tumors of female conventional *Men1*$^{+/-}$ mice resulted in restoration of menin expression and decreased tumor proliferation[97]; and use of an adeno-associated virus and phage vector displaying octreotide to deliver a tumor necrosis factor transgene to pancreatic NETs in *Pdx1-Cre-Men1* mice resulted in decreased tumor volume and improved survival.[98] Drugs targeting the WNT and KRAS pathways have also been assessed, and a small-molecule β-catenin antagonist (PKF115-584) was shown to reduce proliferation of pancreatic NETs in *RIP-Cre-Men1* mice and decrease excess insulin secretion[88]; whereas MEK1/2 inhibitors targeting KRAS signaling via the RAF/MEK/ERK pathway were demonstrated to have antiproliferative activity and result in synthetic lethality in pancreatic islets from *Ins2-Cre-Men1*$^{+/-}$ and *Ins2-Cre-Men1*$^{-/-}$ mice, respectively.[90] Finally, *Men1* models have been used to investigate repurposing of drugs used for other ailments and diseases. For example, the combination of aspirin and enalapril was reported to reduce the size and number of pancreatic NETs in conventional *Men1*$^{+/-}$ mice, indicating a potential chemo-preventive effect,[99] while the autophagy inhibitor hydroxychloroquine accelerated apoptosis in pancreatic NETs in conventional *Men1*$^{+/-}$ mice.[100]

KEY POINTS

- Menin, which acts as a scaffold protein, is involved in proliferative pathways that represent targeted approaches for future treatments, and use of *MEN1* mouse knockout models that develop the same endocrine tumors observed in patients has facilitated *in vivo* evaluations of several potential novel treatments that include gene therapy and epigenetic modifying drugs.

CONCLUSION

Combined clinical, genetic, and laboratory investigations of MEN1 have resulted in an increased understanding of this autosomal dominant disorder, which has improved patient management and treatment. Germline genetic testing facilitates the identification of mutant *MEN1* gene carriers, who are at high risk for developing this disorder and are likely to benefit from interval biochemical and radiologic surveillance, which aims for earlier detection and treatment of endocrine tumors. Application of the techniques of molecular biology has facilitated important insights into the function of the *MEN1* gene and its encoded protein, menin, which has context-specific pleotropic activities involved in the regulation of transcription, genome stability, cell division, and the epigenome. These advances have provided new avenues for exploring future targeted treatments for the different endocrine tumors that occur in patients with this disorder.

Acknowledgments

PJN is grateful to the Chief Scientist Office, NHS Research Scotland, and University of Dundee. RVT is grateful to the Wellcome Trust; National Institute of Health Oxford Biomedical Research Centre Programme, and Medical Research Council (United Kingdom). The authors thank Tracey Walker for her assistance in preparing this chapter.

REFERENCES

1. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab.* 2012;97:2990–3011.
2. Trump D, Farren B, Wooding C, et al. Clinical studies of multiple endocrine neoplasia type 1 (MEN1). *QJM.* 1996;89:653–669.

2a. Brandi ML, Agarwal SK, Perrier ND et al. Multiple Endocrine Neoplasia Type 1: Latest Insights. *Endocr Rev.* 2021;42(2):133–170.

3. Matkar S, Thiel A, Menin HX. A scaffold protein that controls gene expression and cell signaling. *Trends Biochem Sci.* 2013;38:394–402.
4. Frost M, Lines KE, Thakker RV. Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol.* 2018;14:216–227.

4a. Wells SA et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25(6):567–610.
5. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, et al. Germ-line mutations in p27kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *Proc Natl Acad Sci U S A*. 2006;103:15558–15563.
6. Frederiksen A, Rossing M, Hermann P, et al. Clinical features of multiple endocrine neoplasia type 4: novel pathogenic variant and review of published cases. *J Clin Endocrinol Metab*. 2019;104:3637–3646.
7. Erdheim J. [Zur normalen und pathologischen histologie der glandula thyreoidea, parathyreoidea und hypophysis.]. *Beitr Pathol Anat*. 1903;33:158–263.
8. Underdahl LO, Woolner LB, Black BM. Multiple endocrine adenomas; report of 8 cases in which the parathyroids, pituitary and pancreatic islets were involved. *J Clin Endocrinol Metab*. 1953;13:20–47.
9. Chandrasekharappa SC, Guru SC, Manickam P, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science*. 1997;276:404–407.
10. Lemmens I, Van de Ven WJ, Kas K, et al. Identification of the multiple endocrine neoplasia type 1 (MEN1) gene. The European consortium on MEN1. *Hum Mol Genet*. 1997;6:1177–1183.
11. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat*. 2008;29:22–32.
12. Huang J, Gurung B, Wan B, et al. The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature*. 2012;482:542–546.
13. Goudet P, Dalac A, Le Bras M, et al. MEN1 disease occurring before 21 years old: a 160-patient cohort study from the groupe d'etude des tumeurs endocrines. *J Clin Endocrinol Metab*. 2015;100:1568–1577.
13a. Newey PJ, Hannan FM, Wilson A, Thakker RV. Genetics of monogenic disorders of calcium and bone metabolism. *Clin Endocrinol (Oxf)*. 2021; https://doi:10.1111/cen.14644.
14. Dreijerink KM, Goudet P, Burgess JR, et al. Breast-cancer predisposition in multiple endocrine neoplasia type 1. *N Engl J Med*. 2014;371: 583–584.
15. Sadowski SM, Pieterman CRC, Perrier ND, et al. Prognostic factors for the outcome of nonfunctioning pancreatic neuroendocrine tumors in MEN1: a systematic review of literature. *Endocr Relat Cancer*. 2020;27:R145–R161.
16. Challis BG, Casey RT, Grossman A, et al. What is the appropriate management of nonfunctioning pancreatic neuroendocrine tumours disclosed on screening in adult patients with multiple endocrine neoplasia type 1? *Clin Endocrinol (Oxf)*. 2019;91:708–715.
17. Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. *Lancet Diabetes Endocrinol*. 2015;3:895–905.
18. van Beek DJ, Nell S, Pieterman CRC, et al. Prognostic factors and survival in MEN1 patients with gastrinomas: results from the Dutchmen study group (dmsg). *J Surg Oncol*. 2019;120:966–975.
19. de Laat JM, Dekkers OM, Pieterman CR, et al. Long-term natural course of pituitary tumors in patients with MEN1: results from the DutchMEN1 study group (dmsg). *J Clin Endocrinol Metab*. 2015;100:3288–3296.
20. Gatta-Cherifi B, Chabre O, Murat A, et al. Adrenal involvement in MEN1. Analysis of 715 cases from the Groupe D'Etude des Tumeurs Endocrines Database. *Eur J Endocrinol*. 2012;166:269–279.
21. Bartsch DK, Albers MB, Lopez CL, et al. Bronchopulmonary neuroendocrine neoplasms and their precursor lesions in multiple endocrine neoplasia type 1. *Neuroendocrinology*. 2016;103:240–247.
22. Pieterman CR, Conemans EB, Dreijerink KM, et al. Thoracic and duodenopancreatic neuroendocrine tumors in multiple endocrine neoplasia type 1: natural history and function of menin in tumorigenesis. *Endocr Relat Cancer*. 2014;21:R121–R142.
23. Kooblall KG, Boon H, Cranston T, et al. Multiple endocrine neoplasia type 1 (MEN1) 5'UTR deletion, in MEN1 family, decreases menin expression. *J Bone Miner Res*. 2021;36:100–109.
24. Norton JA, Krampitz G, Zemek A, et al. Better survival but changing causes of death in patients with multiple endocrine neoplasia type 1. *Ann Surg*. 2015;261:e147–e148.
25. Nilubol N, Weinstein LS, Simonds WF, et al. Limited parathyroidectomy in multiple endocrine neoplasia type 1-associated primary hyperparathyroidism: a setup for failure. *Ann Surg Oncol*. 2016;23:416–423.
26. Fyrsten E, Norlen O, Hessman O, et al. Long-term surveillance of treated hyperparathyroidism for multiple endocrine neoplasia type 1: recurrence or hypoparathyroidism? *World J Surg*. 2016;40:615–621.
27. Choi HR, Choi SH, Choi SM, et al. Benefit of diverse surgical approach on short-term outcomes of MEN1-related hyperparathyroidism. *Sci Rep*. 2020;10:10634.
28. Kluijfhout WP, Beninato T, Drake FT, et al. Unilateral clearance for primary hyperparathyroidism in selected patients with multiple endocrine neoplasia type 1. *World J Surg*. 2016;40:2964–2969.
29. Manoharan J, Albers MB, Bollmann C, et al. Single gland excision for MEN1-associated primary hyperparathyroidism. *Clin Endocrinol (Oxf)*. 2020;92:63–70.
30. Pieterman CRC, Sadowski SM, Maxwell JE, et al. Hereditary endocrine tumours: current state-of-the-art and research opportunities: MEN1-related pancreatic NETs: identification of unmet clinical needs and future directives. *Endocr Relat Cancer*. 2020;27:T9–T25.
31. Niederle B, Selberherr A, Bartsch D, et al. Multiple endocrine neoplasia type 1 (MEN1) and the pancreas—diagnosis and treatment of functioning and non-functioning pancreatic and duodenal neuroendocrine neoplasia within the MEN1 syndrome - an international consensus statement. *Neuroendocrinology*. 2020;111:609–630.
32. Nell S, Verkooijen HM, Pieterman CRC, et al. Management of MEN1 related nonfunctioning pancreatic NETs: a shifting paradigm: results from the DutchMEN1 study group. *Ann Surg*. 2018;267:1155–1160.
33. Nell S, Verkooijen HM, Pieterman CR, et al. Management of MEN1 related nonfunctioning pancreatic NETs: a shifting paradigm: results from the dutchmen1 study group. *Ann Surg*. 2018;267:1155–1160.
34. Vinault S, Mariet AS, Le Bras M, et al. Metastatic potential and survival of duodenal and pancreatic tumors in multiple endocrine neoplasia type 1: a GTE and AFCE cohort study (Groupe D'Etude des Tumeurs Endocrines and Association Francophone de Chirurgie Endocrinienne). *Ann Surg*. 2018;272:1094–1101.
35. Albers MB, Manoharan J, Bartsch DK. Contemporary surgical management of the zollinger-ellison syndrome in multiple endocrine neoplasia type 1. *Best Pract Res Clin Endocrinol Metab*. 2019;33:101318.
36. Norton JA, Foster DS, Ito T, et al. Gastrinomas: medical or surgical treatment. *Endocrinol Metab Clin North Am*. 2018;47:577–601.
37. van Beek DJ, Nell S, Verkooijen HM, et al. Surgery for multiple endocrine neoplasia type 1-related insulinoma: long-term outcomes in a large international cohort. *Br J Surg*. 2020;107:1489–1499.
38. Endall R, Thompson M, Parameswaran V, et al. The relationship of gastrinoma in MEN1 to *Helicobacter pylori* infection. *J Clin Endocrinol Metab*. 2020;105:dgaa004.
39. Sadowski SM, Cadiot G, Dansin E, et al. The future: surgical advances in MEN1 therapeutic approaches and management strategies. *Endocr Relat Cancer*. 2017;24:T243–T260.
40. van Leeuwaarde RS, de Laat JM, Pieterman CRC, et al. The future: medical advances in MEN1 therapeutic approaches and management strategies. *Endocr Relat Cancer*. 2017;24:T179–T193.
41. Antwi K, Nicolas G, Fani M, et al. 68Ga-exendin-4 PET/CT detects insulinomas in patients with endogenous hyperinsulinemic hypoglycemia in MEN-1. *J Clin Endocrinol Metab*. 2019;104:5843–5852.
42. van Treijen MJC, van Beek DJ, van Leeuwaarde RS, et al. Diagnosing nonfunctional pancreatic NETs in MEN1: the evidence base. *J Endocr Soc*. 2018;2:1067–1088.
43. Qiu W, Christakis I, Silva A, et al. Utility of chromogranin a, pancreatic polypeptide, glucagon and gastrin in the diagnosis and follow-up of pancreatic neuroendocrine tumours in multiple endocrine neoplasia type 1 patients. *Clin Endocrinol (Oxf)*. 2016;85:400–407.
44. van Treijen MJC, van der Zee D, Heeres BC, et al. Blood molecular genomic analysis predicts the disease course of GEP NET patients: a validation study of the predictive value of the NETest(r). *Neuroendocrinology*. 2021;111:586–598.
45. Newey PJ, Jeyabalan J, Walls GV, et al. Asymptomatic children with multiple endocrine neoplasia type 1 mutations may harbor nonfunc-

tioning pancreatic neuroendocrine tumors. *J Clin Endocrinol Metab*. 2009;94:3640–3646.

46. Nell S, Borel Rinkes IHM, Verkooijen HM, et al. Early and late complications after surgery for MEN1-related nonfunctioning pancreatic neuroendocrine tumors. *Ann Surg*. 2018;267:352–356.
47. Triponez F, Sadowski SM, Pattou F, et al. Long-term follow-up of MEN1 patients who do not have initial surgery for small </=2 cm nonfunctioning pancreatic neuroendocrine tumors, an AFCE and GTE study: association Francophone de Chirurgie Endocrinienne & Groupe D'Etude des Tumeurs Endocrines. *Ann Surg*. 2018;268:158–164.
48. Pieterman CRC, de Laat JM, Twisk JWR, et al. Long-term natural course of small nonfunctional pancreatic neuroendocrine tumors in MEN1-results from the Dutch MEN1 study group. *J Clin Endocrinol Metab*. 2017;102:3795–3805.
49. Conemans EB, Brosens LAA, Raicu-Ionita GM, et al. Prognostic value of who grade in pancreatic neuro-endocrine tumors in multiple endocrine neoplasia type 1: results from the dutchmen1 study group. *Pancreatology*. 2017;17:766–772.
50. Cejas P, Drier Y, Dreijerink KMA, et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. *Nat Med*. 2019;25:1260–1265.
51. Deleted in review.
52. Partelli S, Tamburrino D, Lopez C, et al. Active surveillance versus surgery of nonfunctioning pancreatic neuroendocrine neoplasms </=2 cm in MEN1 patients. *Neuroendocrinology*. 2016;103:779–786.
53. Faggiano A, Modica R, Lo Calzo F, et al. Lanreotide therapy vs active surveillance in MEN1-related pancreatic neuroendocrine tumors < 2 centimeters. *J Clin Endocrinol Metab*. 2020;105.
54. Cohen-Cohen S, Brown DA, Himes BT, et al. Pituitary adenomas in the setting of multiple endocrine neoplasia type 1: a single-institution experience. *J Neurosurg*. 2020;133:1132–1138.
55. Ye L, Wang W, Ospina NS, et al. Clinical features and prognosis of thymic neuroendocrine tumours associated with multiple endocrine neoplasia type 1: a single-centre study, systematic review and meta-analysis. *Clin Endocrinol*. 2017;87:706–716.
56. Lecomte P, Binquet C, Le Bras M, et al. Histologically proven bronchial neuroendocrine tumors in MEN1: a GTE 51-case cohort study. *World J Surg*. 2018;42:143–152.
57. de Laat JM, Pieterman CR, van den Broek MF, et al. Natural course and survival of neuroendocrine tumors of thymus and lung in MEN1 patients. *J Clin Endocrinol Metab*. 2014;99:3325–3333.
58. So A, Pointon O, Hodgson R, et al. An assessment of (18) F-FDG PET/CT for thoracic screening and risk stratification of pulmonary nodules in multiple endocrine neoplasia type 1. *Clin Endocrinol (Oxf)*. 2018;88:683–691.
59. van Leeuwaarde RS, Dreijerink KM, Ausems MG, et al. MEN1-dependent breast cancer: indication for early screening? Results from the Dutch MEN1 study group. *J Clin Endocrinol Metab*. 2017;102:2083–2090.
60. Dreijerink KMA, Groner AC, Vos ESM, et al. Enhancer-mediated oncogenic function of the menin tumor suppressor in breast cancer. *Cell Rep*. 2017;18:2359–2372.
61. Thompson M, Hogg P, De Paoli A, et al. Parental multiple endocrine neoplasia type 1 (MEN1) is associated with increased offspring childhood mortality. *J Clin Endocrinol Metab*. 2020;105:dgz231.
62. Hogg P, Thompson M, Burgess J. The clinical expression and impact of multiple endocrine neoplasia 1 during pregnancy. *Clin Endocrinol (Oxf)*. 2020;93:429–438.
63. Thompson M, Burgess J. No evidence of adverse fertility and pregnancy outcomes in patients with unrecognized and untreated multiple endocrine neoplasia type 1. *Clin Endocrinol*. 2019;90:312–319.
64. Lee ME, Ortega-Sustache YM, Agarwal SK, et al. Patients with MEN1 are at an increased risk for venous thromboembolism. *J Clin Endocrinol Metab*. 2020;106:e460–e468.
65. Agarwal SK. The future: Genetics advances in MEN1 therapeutic approaches and management strategies. *Endocr Relat Cancer*. 2017;24:T119–T134.
66. Newey PJ, Berg JN, Zhou K, et al. Utility of population-level DNA sequence data in the diagnosis of hereditary endocrine disease. *J Endocr Soc*. 2017;1:1507–1526.
67. Romanet P, Mohamed A, Giraud S, et al. Umd-MEN1 database: an overview of the 370 MEN1 variants present in 1676 patients from the French population. *J Clin Endocrinol Metab*. 2019;104:753-–64.
68. Yoshiji S, Iwasaki Y, Iwasaki K, et al. Alu-mediated MEN1 gene deletion and loss of heterozygosity in a patient with multiple endocrine neoplasia type 1. *J Endocr Soc*. 2020;4:bvaa051.
69. Beijers H, Stikkelbroeck NML, Mensenkamp AR, et al. Germline and somatic mosaicism in a family with multiple endocrine neoplasia type 1 (MEN1) syndrome. *Eur J Endocrinol*. 2019;180:K15–K19.
70. Thevenon J, Bourredjem A, Faivre L, et al. Unraveling the intrafamilial correlations and heritability of tumor types in MEN1: a Groupe D'Etude des Tumeurs Endocrines study. *Eur J Endocrinol*. 2015;173:819–826.
71. Backman S, Bajic D, Crona J, et al. Whole genome sequencing of apparently mutation-negative MEN1 patients. *Eur J Endocrinol*. 2020;182:35–45.
72. Pieterman CRC, Hyde SM, Wu SY, et al. Understanding the clinical course of genotype-negative MEN1 patients can inform management strategies. *Surgery*. 2021;169:175–184.
73. de Laat JM, van der Luijt RB, Pieterman CR, et al. MEN1 redefined, a clinical comparison of mutation-positive and mutation-negative patients. *BMC Med*. 2016;14:182.
74. Newey PJ. Clinical genetic testing in endocrinology: current concepts and contemporary challenges. *Clin Endocrinol (Oxf)*. 2019;91:587–607.
75. Lawrence B, Blenkiron C, Parker K, et al. Recurrent loss of heterozygosity correlates with clinical outcome in pancreatic neuroendocrine cancer. *NPJ Genom Med*. 2018;3:18.
76. Wirth LJ, Sherman E, Robinson B, et al. Efficacy of selpercatinib in RET-altered thyroid cancers. *N Engl J Med*. 2020;383:825–835.
77. Lines KE, Nachtigall LB, Dichtel LE, et al. Multiple endocrine neoplasia type 1 (MEN1) phenocopy due to a cell cycle division 73 (CDC73) variant. *J Endocr Soc*. 2020;4:bvaa142.
78. El Lakis M, Nockel P, Gaitanidis A, et al. Probability of positive genetic testing results in patients with family history of primary hyperparathyroidism. *J Am Coll Surg*. 2018;226:933–938.
79. Nachtigall LB, Guarda FJ, Lines KE, et al. Clinical men-1 among a large cohort of patients with acromegaly. *J Clin Endocrinol Metab*. 2020;105:e2271–2281.
80. Denes J, Swords F, Rattenberry E, et al. Heterogeneous genetic background of the association of pheochromocytoma/paraganglioma and pituitary adenoma: results from a large patient cohort. *J Clin Endocrinol Metab*. 2015;100:E531–E541.
81. de Laat JM, van Leeuwaarde RS, Valk GD. The importance of an early and accurate MEN1 diagnosis. *Front Endocrinol*. 2018;9:533.
82. van Leeuwaarde RS, van Nesselrooij BP, Hermus AR, et al. Impact of delay in diagnosis in outcomes in MEN1: results from the Dutch MEN1 study group. *J Clin Endocrinol Metab*. 2016;101:1159–1165.
83. van Leeuwaarde RS, Pieterman CRC, Bleiker EMA, et al. High fear of disease occurrence is associated with low quality of life in patients with multiple endocrine neoplasia type 1 (MEN1): results from the Dutch MEN1 study group. *J Clin Endocrinol Metab*. 2018;103:2354–2361.
84. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet Med*. 2015;17:405–424.
85. Romanet P, Odou MF, North MO, et al. Proposition of adjustments to the ACMG-AMP framework for the interpretation of MEN1 missense variants. *Hum Mutat*. 2019;40:661–674.
86. Caswell RC, Owens MM, Gunning AC, et al. Using structural analysis in silico to assess the impact of missense variants in MEN1. *J Endocr Soc*. 2019;3:2258–2275.
87. van den Broek MFM, van Nesselrooij BPM, Pieterman CRC, et al. Clues for genetic anticipation in multiple endocrine neoplasia type 1. *J Clin Endocrinol Metab*. 2020;105:dgaa257.
88. Jiang X, Cao Y, Li F, et al. Targeting beta-catenin signaling for therapeutic intervention in MEN1-deficient pancreatic neuroendocrine tumours. *Nat Commun*. 2014;5:5809.
89. Grozinsky-Glasberg S, Lines KE, Avniel-Polak S, et al. Preclinical drug studies in MEN1-related neuroendocrine neoplasms (MEN1-nens). *Endocr Relat Cancer*. 2020;27:R345–R355.

90. Chamberlain CE, Scheel DW, McGlynn K, et al. Menin determines k-ras proliferative outputs in endocrine cells. *J Clin Invest*. 2014;124:4093–4101.
91. Lines KE, Vas Nunes RP, Frost M, et al. A MEN1 pancreatic neuroendocrine tumour mouse model under temporal control. *Endocr Connect*. 2017;6:232–242.
92. Mohr H, Pellegata NS. Animal models of MEN1. *Endocr Relat Cancer*. 2017;24:T161–T177.
93. Lines KE, Javid M, Reed AAC, et al. Genetic background influences tumour development in heterozygous Men1 knockout mice. *Endocr Connect*. 2020;9:426–437.
94. Walls GV, Stevenson M, Soukup BS, et al. Pasireotide therapy of multiple endocrine neoplasia type 1-associated neuroendocrine tumors in female mice deleted for an Men1 allele improves survival and reduces tumor progression. *Endocrinology*. 2016;157:1789–1798.
95. Lopez CL, Joos B, Bartsch DK, et al. Chemoprevention with somatuline(c) delays the progression of pancreatic neuroendocrine neoplasms in a mouse model of multiple endocrine neoplasia type 1 (Men1). *World J Surg*. 2019;43:831–838.
96. Lines KE, Stevenson M, Filippakopoulos P, et al. Epigenetic pathway inhibitors represent potential drugs for treating pancreatic and bronchial neuroendocrine tumors. *Oncogenesis*. 2017;6:e332.
97. Walls GV, Lemos MC, Javid M, et al. Men1 gene replacement therapy reduces proliferation rates in a mouse model of pituitary adenomas. *Cancer Res*. 2012;72:5060–5068.
98. Smith TL, Yuan Z, Cardo-Vila M, et al. AAVP displaying octreotide for ligand-directed therapeutic transgene delivery in neuroendocrine tumors of the pancreas. *Proc Natl Acad Sci U S A*. 2016;113:2466–2471.
99. Manoharan J, Fendrich V, Di Fazio P, et al. Chemoprevention with enalapril and aspirin in Men1(+/t) knockout mouse model. *Neuroendocrinology*. 2018;107:257–266.
100. Nakano K, Masui T, Yogo A, et al. Chloroquine induces apoptosis in pancreatic neuroendocrine neoplasms via endoplasmic reticulum stress. *Endocr Relat Cancer*. 2020;27:431–439.

134

Multiple Endocrine Neoplasia Types 2 and 3, and Medullary Thyroid Carcinoma

Matti L. Gild, Venessa Tsang, Roderick J. Clifton-Bligh, and Bruce G. Robinson

OUTLINE

Prevalence and Epidemiology, 2258
Molecular Pathogenesis: *RET* Protooncogene, 2259
  RET Structure and Function, 2259
  RET Ligands, 2259
  Functional Effects of *RET* Variants, 2260
Histology, 2261
  C-Cells, 2261
  C-Cell Hyperplasia, 2261
Clinical Relevance of Germline *RET* Variants in Multiple Endocrine Neoplasia Type 2 and Multiple Endocrine Neoplasia Type 3, 2261
  Multiple Endocrine Neoplasia Type 2 and Familial Medullary Thyroid Carcinoma, 2261
  Multiple Endocrine Neoplasia Type 3, 2261
Specific Genotype-Phenotype Correlations, 2261
Clinical Presentation, 2261
  Medullary Thyroid Carcinoma, 2261
  Pheochromocytoma, 2262
  Hyperparathyroidism, 2262
Screening in Multiple Endocrine Neoplasia Type 2 and Multiple Endocrine Neoplasia Type 3, 2262
  Biochemical Screening, 2262
  Genetic Screening, 2263
Management Implications of Genetic Results, 2263
  Timing and Benefits of Prophylactic Thyroidectomy, 2263
  Strategy and Extent of Thyroidectomy, 2263
The Role of Somatic Mutations, 2264
  Sporadic Medullary Thyroid Carcinoma, 2264
  Low-Penetrance Phenotypes, 2264
  Hereditary Medullary Thyroid Carcinoma, 2264
Germline Screening, 2264
  Sporadic Medullary Thyroid Carcinoma, 2264
Natural Progression and Behavior of Multiple Endocrine Neoplasia Type 2 and Multiple Endocrine Neoplasia Type 3, 2264
  Staging of Medullary Thyroid Carcinoma, 2264
  Hereditary Versus Sporadic Medullary Thyroid Carcinoma, 2266
  Paraneoplastic Features of Medullary Thyroid Carcinoma, 2266
  Pheochromocytoma, 2267
  Hyperparathyroidism, 2267
Management of Medullary Thyroid Carcinoma, 2267
  Localized Medullary Thyroid Carcinoma, 2267
  Recurrent Medullary Thyroid Carcinoma, 2267
  Persistent and Metastatic Medullary Thyroid Carcinoma, 2267
Management of Pheochromocytoma, 2270
Management of Hyperparathyroidism, 2270
Conclusion, 2270

KEY POINTS

- Medullary thyroid carcinoma (MTC) presents sporadically in 75% of cases. In 25% of cases, MTC shows autosomal dominant inheritance as part of the multiple endocrine neoplasia type 2 (MEN2) and 3 (MEN3) syndromes.
- MEN2 and MEN3 syndromes are associated with mutations in the *RET* protooncogene. Affected patients may also develop pheochromocytomas and primary hyperparathyroidism, in addition to other clinical features.
- All patients who develop MTC, including apparently sporadic cases, require germline *RET* mutation analysis.
- Genotype-phenotype correlations have allowed familial cases to be stratified according to the level of risk associated with the specific *RET* variant. Guidelines published by the American Thyroid Association guide clinical management and help to determine the age at which prophylactic thyroidectomy should be performed in asymptomatic *RET* mutation carriers.
- MTC is characterized by early spread to local lymph nodes, and nodal disease is evident in up to 50% of patients at diagnosis. Surgery remains the mainstay of therapy; total thyroidectomy with comprehensive clearance of central lymph nodes is performed as a minimum in index cases.
- Treatment of persistent and metastatic disease has evolved, with targeted kinase therapy demonstrating effective control in patients harboring *RET* germline or somatic variants.

PREVALENCE AND EPIDEMIOLOGY

Multiple endocrine neoplasia type 2 (formerly MEN2A) refers to the occurrence of *medullary thyroid carcinoma* (MTC) in association with pheochromocytoma in 50% of cases, and in association with hyperparathyroidism in 20% to 30% of cases.[1] *Multiple endocrine neoplasia type 3* (formerly MEN2B) consists of MTC in conjunction with pheochromocytoma and several additional features, including mucosal neuromas, ganglioneuromatoses of the gastrointestinal tract, and a marfanoid habitus (Fig. 134.1B). Parathyroid disease is very rare in MEN3. *Familial medullary thyroid carcinoma* (FMTC) describes the autosomal dominant inheritance of MTC alone and was previously a freestanding disease but is now considered a variant of MEN2. Clinical presentation is usually delayed, and prognosis is more favorable, compared with MEN2 and MEN3.[2]

Rare kindreds have been described with the combination of MEN2 and a pruritic skin lesion over the upper back with a lichenoid appearance (Fig. 134.1A). This cutaneous lichen amyloidosis (CLA) may be preceded by pruritus alone for several years.[3]

Although 25% of cases of MTC occur in well-defined familial endocrine syndromes with autosomal dominant inheritance, 75% of MTC cases are sporadic. MTC may also present with ectopic hormonal

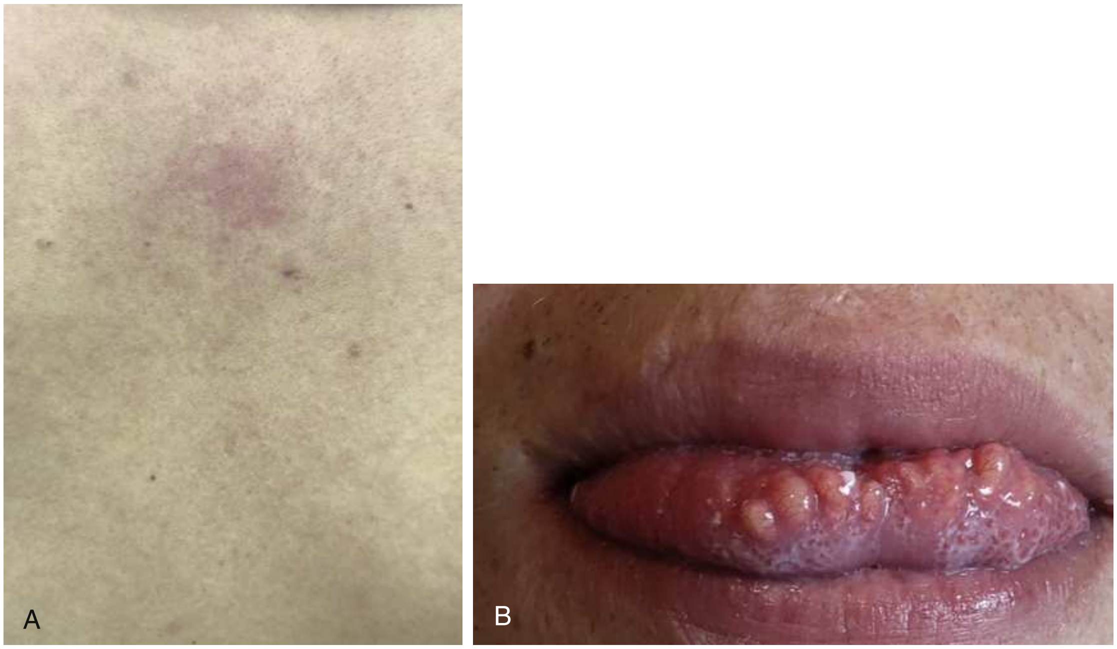


Fig. 134.1 Clinical Images in Multiple Endocrine Neoplasia Type 2 (MEN2) and Multiple Endocrine Neoplasia Type 3 (MEN3). **A**, Cutaneous lichen amyloidosis in a MEN2 patient. **B**, Mucosal neuromas in a MEN3 patient.

syndromes such as ectopic adrenocorticotropic hormone (ACTH; discussed later). Although MTC accounts for only 5% to 10% of all thyroid cancers, its management can be very challenging, and early diagnosis is essential for good outcomes.

The overall prevalence of MEN2 is approximately 1:80,000.[4] The penetrance is approximately 70% at 70 years of age.[5] It has been reported that 90% of MEN2 carriers will eventually show evidence of MTC.6 This contrasts with multiple endocrine neoplasia type 1 (MEN1), in which penetrance is 94% by 50 years of age.[7] For further detailed discussion of pheochromocytoma, see Chapter 110. For primary hyperparathyroidism, see Chapter 63. MEN1 is discussed in Chapter 148.

MOLECULAR PATHOGENESIS: *RET* PROTOONCOGENE

The causative gene for the MEN2 syndromes was identified in 1993.[8,9] The gene is the rearranged during transfection (*RET*) protooncogene located on chromosome 10q11.2.[10]

The *RET* protooncogene encodes a tyrosine kinase cell surface receptor. Point mutations in *RET* segregating with the disease phenotype were identified in MEN2 and FMTC in 1993.[9] Germline variants of *RET* at two specific codons were subsequently identified in MEN3, at a different site.[11] Genetic screening has revolutionized the management of MEN2 families.

The International *RET* Mutation Consortium performed an extensive study of 477 MEN2 families around the world and established important genotype-phenotype correlations.[12] *RET* mutations have been characterized in approximately 98% of MEN2 families overall. In MEN2, approximately 97% of families will have an identifiable *RET* germline variant, and in MEN3, approximately 98% of families are mutation-positive. *RET* variants are summarized in Table 134.1.

RET Structure and Function

RET consists of 21 exons spanning a minimum of 30 kb of genomic DNA, encoding a cell surface receptor. RET consists of an extracellular domain (exons 1–10), a transmembrane domain (exon 11), and an intracellular domain (exon 12 and onward) (Fig. 134.2).

The RET protein is membrane-bound, with its intracellular portion having tyrosine kinase activity. The extracellular domain extends for the initial 635 amino acids and includes an area of cadherin homology and a conserved cysteine-rich region immediately adjacent to the transmembrane domain. The cysteine-rich domain is typical of the transforming growth factor beta superfamily.[13] The protein is glycosylated to produce receptors with molecular weights of 150 and 170 kD. The 170-kD form is present in the cell membrane, whereas the 150-kD form is an immature form present in the cell cytoplasm.[14] Phosphorylation of RET tyrosine residues results in activation of several intracellular second messenger systems and signaling pathways. These include phospholipase Cγ/protein kinase C, c-Jun N-terminal kinases, products of the protooncogene Src-related kinases, and nuclear factor κB. Other downstream targets include the RAS/RAF/ERK (extracellular regulated kinase) and phosphatidyl-inositol-3-kinase/AKT pathways. Through these signaling cascades, *RET* appears to have a central role in cellular proliferation, differentiation, and migration.[15-17] Interestingly, in addition to its role in endocrine neoplasia, aberrant *RET* signaling has been associated with other distinct phenotypes. Hirschsprung disease has been described together with MEN2 in subjects with codon 609, 611, 618, and 620 variants,[18] and CLA with the codon 634 variant.[19]

RET Ligands

There are four known RET ligands: glial cell line–derived neurotrophic factor (GDNF)[20], neurturin,[21] artemin,[22] and persephin.[23] The link between *RET* and GDNF was suggested by studying mice that were deficient in GDNF expression. These mice were very similar to *RET* knockout mice, with absent or limited development of the kidney and an absence of enteric neurons in the colon and small intestine.[24] Interestingly, genitourinary abnormalities are not a feature of MEN2, despite the fact that *RET* is expressed in human fetal kidney during development. The three other ligands have sequences related to GDNF. RET activation by ligand occurs via membrane-bound

TABLE 134.1 Genotype-Phenotype Correlations and Risk Levels for Medullary Thyroid Carcinoma Behavior

RET Mutation	Affected Exon	ATA Level Risk	Phenotype: MEN2 or 3	Other
G533C	8	MOD	2	
C609F/R/G/S/Y	10	MOD	2	Hirschsprung's
C611R/G/F/S/W/Y	10	MOD	2	Hirschsprung's
C618R/G/F/S/Y	10	MOD	2	Hirschsprung's
C620R/G/F/S/W/Y	10	MOD	2	Hirschsprung's
C630R/F/S/Y	11	MOD	2	
D631Y	11	MOD	2	
C634F/G/S/W/Y	11	HIGH	2	
K666E	11	MOD	2	
E768D	13	MOD	2	
L790F	13	MOD	2	
V804L	14	MOD	2	
V804M	14	MOD	2	
A883F	15	HIGH	3	
S891A	15	MOD	2	
R912P	16	MOD		MTC rare
M918T	16	HIGHEST	3	

Rare mutations based on very few families with genetic variants of unknown significance are not included.

ATA, American Thyroid Association; *MEN*, multiple endocrine neoplasia; *MTC*, medullary thyroid carcinoma.

(From Wells SA Jr, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.)

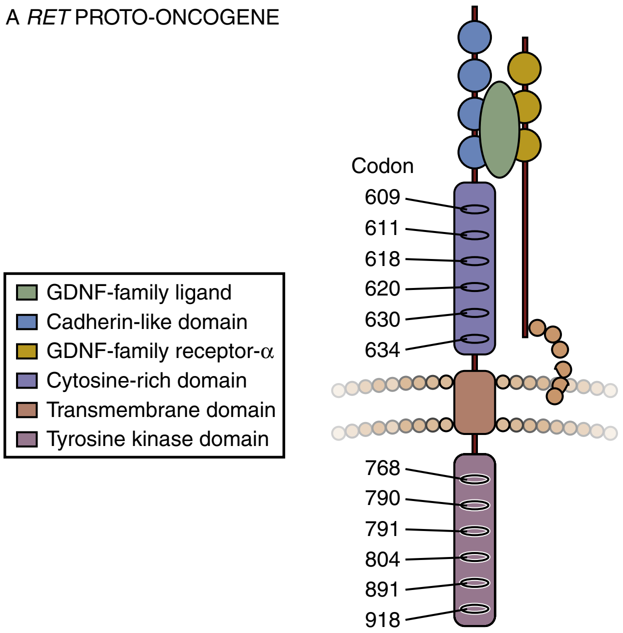


Fig. 134.2 Structure of the *RET* Protooncogene. *GDNF*, glial cell line–derived neurotrophic factor. (From Cote GJ, Gagel RF. Lessons learned from the management of a rare genetic cancer. *N Engl J Med*. 2003;349:1566–1568.)

proteins that function as the ligand-binding domain of a ligand/receptor complex. GDNF-family ligands bind to specific GDNF-family receptor proteins, all of which form receptor complexes and signal through RET. Additional proteins, GDNF-family receptor alpha (GFRα) or GDNF-family receptor alpha 1 (GFRα-1), were found to be essential for the high-affinity binding of GDNF to RET and for its consequent functional effects.[25] A single molecule of GDNF binds two GFRα-1 molecules that interact with two molecules of RET and lead to activation of RET tyrosine kinase. Similarly, neurturin, persephin, and artemin require the binding of different GFRαs (GFRα1–4) as coreceptors before downstream receptor activation occurs. It was suggested that MEN2 families with no identifiable *RET* variants may have *GDNF* variants, but so far no such variants have been described in MEN2. However, somatic *GDNF* variants have been found in sporadic pheochromocytomas.[26]

KEY POINTS

- Multiple endocrine neoplasia type 2 is a hereditary syndrome referring to the occurrence of medullary thyroid carcinoma (MTC) in association with pheochromocytoma (50% of cases) and hyperparathyroidism (20%–30% of cases). Multiple endocrine neoplasia type 3 refers to a hereditary syndrome involving MTC in conjunction with pheochromocytoma, together with several additional features including mucosal neuromas, ganglioneuromatoses of the gastrointestinal tract, and marfanoid habitus.

Functional Effects of *RET* Variants

Although the genotypes do overlap, in general, different codons are affected in either MEN2 or MEN3. These mutations in different regions of the *RET* protooncogene increase the transforming activity of the kinase by distinct mechanisms. In MEN2, the extracellular cysteine-rich or transmembrane domains are predominantly affected. Such mutations lead to a loss of cysteine residues. As a result, two RET molecules may undergo ligand-independent dimerization, with resultant constitutive activation of tyrosine kinase activity.[27] A variant in codon 634, usually a cysteine-to-arginine substitution (C634R), occurs in 85% of cases of MEN2.[17] In contrast, MEN3 usually is caused by a mutation in the intracellular tyrosine kinase domain. Codon 918 is involved in 95% of cases, in a methionine-to-threonine substitution (M918T).[17] This alters the kinase specificity, with alternative substrate phosphorylation and altered downstream signaling. The rarer intracellular RET mutations in codons 768 and 804 have been shown to produce “gain of function” of the receptor, although the precise mechanism by which tumor formation occurs is not clear.[28]

KEY POINTS

- *RET* mutations are found in nearly all multiple endocrine neoplasia type 2 (MEN2) and multiple endocrine neoplasia type 3 (MEN3) kindreds. In MEN2, codon 634 is involved in >85% of cases, and in MEN3 codon 918 is involved >95% cases.

HISTOLOGY

C-Cells

Calcitonin-secreting C-cells of the thyroid gland arise from pharyngeal endoderm, and their development is significantly influenced by the neural crest cells.[29,30] Located mainly in the upper and middle thirds of the thyroid gland, C-cells are members of the amine-precursor uptake and decarboxylation family and express a variety of neuroendocrine markers, including calcitonin, chromogranin A, neuron-specific enolase, and serotonin.[31] Calcitonin is a highly sensitive tumor marker for monitoring of MTC. Carcinoembryonic antigen (CEA) is another C-cell marker used in clinical practice. C-cells usually display an endocrine phenotype, but may revert to a neuronal phenotype following malignant transformation, at which point they predominantly express alternately spliced calcitonin gene-related peptide (CGRP).[31] The biological role of C-cells is not precisely clear. Calcitonin may be important for calcium homeostasis *in utero*, and C-cells may contribute to thyroid follicular microenvironment. Pharmacologically, calcitonin inhibits calcium resorption from bone, opposing the effect of parathyroid hormone. It has been used therapeutically for treatment of hypercalcemia and osteoporosis.

C-Cell Hyperplasia

C-cell hyperplasia (CCH) is the pathologic description of an increase in C-cell numbers that can precede malignancy. Criteria for the diagnosis of CCH are variable in the literature, although consensus suggests that there should be more than 50 C-cells per field at 100× magnification.[32] C-cells may increase in number with age, and CCH is a common finding in normal thyroid glands at autopsy. CCH is no longer considered to be a specific marker of familial MTC. Multifocal CCH is regarded as a precursor lesion to hereditary MTC. Its progression to microscopic MTC is variable and may take many years.[33]

CLINICAL RELEVANCE OF GERMLINE *RET* VARIANTS IN MULTIPLE ENDOCRINE NEOPLASIA TYPE 2 AND MULTIPLE ENDOCRINE NEOPLASIA TYPE 3

Multiple Endocrine Neoplasia Type 2 and Familial Medullary Thyroid Carcinoma

Extensive studies of MEN2 families from around the world have now clarified that nearly all will have a germline *RET* mutation, usually in the cysteine-encoding codons 609, 611, 618, 620, or 634. The codon 634 variant in exon 10 predominates, usually with an arginine substitution, although other substitutions, such as by tyrosine or glutamine, are described.[34] In the subcategory FMTC, 88% of families have a germline mutation in one of the listed cysteine codons or, more rarely, in intracellular codons, including 768, 790, and 791. *RET* mutations in the intracellular exons 13, 14, and 15 are less common in MEN2, although they will be detected by routine sequencing of these exons. Isolated cases of kindreds with germline mutations in other exons, namely exons 1, 5, and 8 have also been reported.[17,35]

Multiple Endocrine Neoplasia Type 3

In 95% of MEN3 families, the germline RET variant at codon 918 in exon 16 occurs, causing a mutation of methionine to threonine (M918T).[36] This mutation is in the tyrosine kinase domain and is specific to MEN3.[11] Of note, this *RET* codon 918 variant is also found in the tumor tissue in some sporadic MTCs. The majority of germline *RET* codon 918 variants occur *de novo*, where MEN3 patients have no family history of the disease, and the pathogenic variant is mostly paternal in origin.[37] However, they still have a 50% chance of passing this variant on to their offspring. Rarer tyrosine kinase domain mutations occur in codons 883 and 891, and the MEN3 phenotype has also been reported when a germline codon 804 variant is coupled with a second germline *RET* mutation (see Table 134.1).[38]

SPECIFIC GENOTYPE-PHENOTYPE CORRELATIONS

There has been much interest in the prediction of phenotypic behavior from genotype analysis. As published by the International *RET* Consortium, the presence of any *RET* codon 634 variant correlates significantly with the presence of both pheochromocytoma and hyperparathyroidism. One study suggested more aggressive behavior of MTC with the common arginine substitution at codon 634 (C634R), compared with the C634Y variant.[39] The C634R variant has not been associated with FMTC. FMTC was reclassified as a variant of MEN2 to avoid failure to identify pheochromocytoma, which can occur later in life.[34] The occurrence of different phenotypes in different families with the same germline *RET* variant suggests a role for more complex gene interactions. The rarer phenotypes associated with MEN2 include Hirschsprung disease with *RET* codon 609, 618, and 620 variants,[18] and CLA with codon 634 mutations.[19] The codon 918 mutation in exon 16 is MEN3-specific.[17] The rarer codon 883 MEN3 variant in exon 15 appears to be associated with less aggressive MTC than does the codon 918 variant.[40]

KEY POINTS

- There is a strong genotype-phenotype correlation in multiple endocrine neoplasia type 2. Each mutated codon is correlated with a specific phenotype.

CLINICAL PRESENTATION

Medullary Thyroid Carcinoma

MTC is the initial manifestation in the majority of cases of MEN2 and occurs in over 95% of these patients. MTCs are usually multifocal and have spread to adjacent cervical lymph nodes in approximately 40% of “index” patients at the time of diagnosis, making cure difficult. Patients with MTC may present with a palpable thyroid lump or with the symptoms of calcitonin excess—usually diarrhea and flushing. Diarrhea does not usually occur until the calcitonin level is very high (at least 10-fold above the upper limit of the normal range). Unfortunately, by the time the diarrhea has occurred, regional lymph node spread is likely, together with distant metastases. Aggressive tumors spread early to bone, lung, and/or liver, and may lose expression of the calcitonin gene, switching to production of CGRP. Fortunately, only a minority of MTC patients (5%–10%) display an aggressive course, whereas many show an indolent course, surviving for decades in the presence of disease. A clinical presentation with MTC (i.e., by palpable neck lump or elevated calcitonin), as opposed to a preclinical diagnosis by prospective screening, is associated with higher rates of residual or recurrent disease. Nonsecretory MTCs are rare (0.83% in one MTC cohort), but absence of raised calcitonin and CEA lead to their diagnosis at a later clinical stage.[41]

Familial cases of MTC will have different presentations, depending on the specific *RET* variant, as described earlier. MEN3 presents with aggressive MTC at a young age. It accounts for 5% of familial MTC cases overall. Early recognition is crucial but challenging, as *de novo* germline mutations account for greater than 50% of MEN3 cases, and thus the parents of an affected child may not provide phenotypic clues.[42] Patients may present in early childhood with diarrhea from very high calcitonin levels or with other features of MEN3 syndrome.[1] Affected individuals may have a marfanoid habitus, thickened corneal nerves, musculoskeletal abnormalities, and intestinal ganglioneuromatosis. Interestingly, some patients with MEN2 also show corneal nerve thickening. Mucosal neuromas in MEN3 can occur on the tongue, in the lips, and on the eyelids, giving a distinctive facial phenotype (see Fig. 134.1B). Gastrointestinal symptoms are universal and varied. Neuromas in the gastrointestinal tract can cause abdominal pain, gaseous distension, and even pseudoobstruction. Hirschsprung disease can have a similar presentation in children and is as a result of inactivating mutations in the extracellular domain of RET, as distinct from the activating mutations of the tyrosine kinase domain seen in MEN3 patients.[1] *RET* variants are found in 20% of sporadic and 50% of familial Hirschsprung disease cases.[43] *RET* mutation carriers may develop simultaneous Hirschsprung disease and MEN2, or either syndrome in isolation. The mechanism by which *RET* mutations are simultaneously activating (to cause MEN2) and inactivating (to cause Hirschsprung disease) is complex and fascinating.[44]

CLA is seen in approximately 10% of MEN2 families and may precede the onset of MTC, with skin lesions seen in the interscapular region of the back (see Fig. 134.1A). Pruritus leads to scratching and amyloid deposition in the skin.[45]

KEY POINTS

- Medullary thyroid carcinoma (MTC) is the most common manifestation of multiple endocrine neoplasia type 2 and multiple endocrine neoplasia type 3. Presentation with MTC, as opposed to preclinical presentation through screening, is associated with higher rates of residual and recurrent disease. Pheochromocytomas are often bilateral and usually follow a benign course. Hyperparathyroidism is multiglandular and is frequently asymptomatic.

Pheochromocytoma

Pheochromocytoma occurs in approximately 50% of MEN2 patients, although it is less common in those with moderate risk variants. Pheochromocytoma is occasionally the initial manifestation of MEN2 and typically presents at an older age (>35 years) or concurrently. The exception is MEN3, in which pheochromocytoma may occur at an early age.[38] Knowledge of the hereditary pheochromocytoma and paraganglioma syndromes has increased in the last decade and is discussed in detail in Chapter 110. Sporadic pheochromocytomas present later than those associated with MEN2. Pheochromocytoma in MEN2 is usually benign and develops on a background of adrenal medullary hyperplasia. A unilateral pheochromocytoma may be followed by pheochromocytoma in the contralateral gland some years later.[46] Bilateral adrenal pheochromocytomas are seen in 63% of cases overall; coexistent paragangliomas are rare.[47] Metastatic pheochromocytoma is rare (<2.9%) in MEN2, despite the fact that capsular invasion is commonly seen.[47] The hormonal profiles of MEN2 and MEN3 pheochromocytomas are similar, and epinephrine secretion generally predominates, leading to elevated plasma and urinary epinephrine and metanephrines. This pattern of secretion may relate to regulation of adrenomedullary phenylethanolamine N-methyltransferase by RET. By contrast, norepinephrine and normetanephrine metabolite secretion predominate in both sporadic and von Hippel-Lindau disease–associated pheochromocytomas.[48,49]

Hyperparathyroidism

Hyperparathyroidism is usually asymptomatic and occurs in only 20% of MEN2 patients, and is absent in MEN3 patients. It is a much less challenging disease in MEN2 than in MEN1. Hyperparathyroidism is usually multiglandular and rarely occurs at an early age. Recurrence after initial surgery is unusual. It is often diagnosed at the time of MTC surgery, where parathyroid glands may have an appearance similar to nodular hyperplasia. Enlarged parathyroid glands are concurrently removed along with the thyroid gland, even if the patient is asymptomatic and normocalcemic.[50-52]

SCREENING IN MULTIPLE ENDOCRINE NEOPLASIA TYPE 2 AND MULTIPLE ENDOCRINE NEOPLASIA TYPE 3

The purpose of biochemical and genetic screening in MEN2 and 3 families is to detect MTC well before it becomes clinically apparent, as, once clinically evident, it may have spread to the lymph nodes. Early detection of pheochromocytoma is important to reduce the risk for catecholamine-induced hypertensive crises.

Biochemical Screening

Medullary Thyroid Carcinoma. Historically, stimulation with pentagastrin, a calcitonin secretagogue, was used to screen members of MEN2 families biochemically for CCH or MTC. However, the advent of genetic screening has rendered the test largely redundant.

Calcitonin values are generally higher in males (normal range [NR] <10 pg/mL) than in females (NR <5 pg/mL), and ranges are more varied in children. Reference ranges vary in different laboratories, and there can be variation even in the same patient. Chemiluminescent immunometric assays are preferred, and clinicians should use the same laboratory for serial measurement in a particular patient.

Pentagastrin is not available in the United States, and calcium may be used as an alternative calcitonin secretagogue if needed. One such rare indication may be when parents are reluctant to allow prophylactic thyroidectomy in a *RET* mutation–positive child with normal basal calcitonin. A stimulated calcitonin level significantly above normal may help to reinforce the need for timely thyroidectomy.

Pheochromocytoma. Annual measurement of fractionated metanephrines in plasma, rather than urine, is now the method of choice for biochemical pheochromocytoma screening in MEN2. The age at which this should commence will depend on the *RET* variant in the family. Screening generally should commence by 11 years of age for patients with MEN3 and those with codon 634 or 630 variants (American Thyroid Association [ATA] "high" and "highest" risk categories) with MEN2, and by 16 years of age for patients with all other *RET* variants in the ATA "moderate" risk category.[38] In addition, preoperative screening for pheochromocytoma should occur in all MTC patients, with baseline plasma metanephrines demonstrating the highest sensitivity.

Hyperparathyroidism. Biochemical screening for hyperparathyroidism is straightforward, in that serum calcium (albumin-corrected or ionized), and usually serum parathyroid hormone, are measured annually from the time that calcitonin testing is implemented.

Genetic Screening

RET variants are clustered tightly, and the analysis can be confined to specific exons. This compares favorably with genetic analysis in MEN1, where the variants are scattered. Once a pathogenic variant is identified, other family members are screened with a straightforward analysis for that specific variant. Various techniques have been used, and direct polymerase chain reaction–based sequencing is currently commonly performed by most laboratories, although next-generation sequencing is becoming more widespread. Testing should ideally be performed on two different blood samples collected on separate occasions to avoid risk of sample error. *RET* variant–negative members of known MEN2 families (i.e., where a pathogenic *RET* variant has been identified in the family) can be reassured that their risk for developing MTC is no greater than for the general population. Genetic screening must therefore be comprehensive and include sequencing of exons 5, 8, 10, 11, 13, 14, 15, and 16.[35,53]

MANAGEMENT IMPLICATIONS OF GENETIC RESULTS

Timing and Benefits of Prophylactic Thyroidectomy

The advent of genetic screening has allowed prophylactic thyroidectomy to be performed earlier than it would be by traditional biochemical screening. The specific mutated *RET* codon may be correlated with the degree of aggressiveness of the MTC. The three-level risk stratification system developed at the Seventh International Workshop on Multiple Endocrine Neoplasia in 1999[53] was revised in the guidelines published by the ATA in 2015 to include three levels of risk (see Tables 134.1 and 134.2).[38] The highest risk group includes all MEN3-associated mutation carriers harboring the M918T variant. Such patients should have thyroidectomy within the first year of life, because metastases within the first year of life have been described. The second group is the "high risk" category, which includes patients with *RET* codon C634 and A833F variants, where thyroidectomy before 5 years of age is recommended depending on basal calcitonin level. MTC with the codon 634 variant has been reported as early as age 2 years, although cases of MTC are generally rare before the age of 5 years in MEN2 and FMTC. Thyroidectomy at this age is technically possible without significant morbidity. The final cohort has combined category A and B into a moderate risk category that includes mutations at a number of *RET* codons, and surgery should be considered, but the timing is not always clear. Patient/parent preference and calcitonin levels can be used to help guide timing (see Table 134.2). It should be noted that a single case of metastatic MTC in a 6-year-old patient was reported in a family with the codon 804 *RET* variant.[54] Recent data suggesting that p.V804M may be a low-penetrance variant[55] makes this timing and recommendation even more challenging, as discussed later. Fig. 134.1 describes the earliest reported age at onset of MTC for each of the known *RET* variants. Early prophylactic surgery has also alleviated the need for ongoing biochemical stimulation testing being performed in a child.

Calcitonin measurement may assist timing prophylactic thyroidectomy in combination with baseline genetic status. If baseline calcitonin is elevated, then thyroidectomy should be performed immediately, regardless of the *RET* variant found. Histology in this setting is likely to show MTC rather than CCH, with the associated risk for lymph node spread.[56] In one series, all patients with a preoperative basal calcitonin level less than 30 ng/L were cured with prophylactic thyroidectomy.[57] In summary, an ideal protocol for timing includes following early genetic testing and risk stratification, depending on the specific *RET* variant in combination with calcitonin measurement.

KEY POINTS

- Prophylactic thyroidectomy is recommended at specific ages, depending on *RET* genotype. Each codon mutation is divided into three risk categories. Calcitonin measurement may assist in determining optimal timing of thyroidectomy for patients harboring codons in the high or moderate risk category.

Strategy and Extent of Thyroidectomy

In young children with MEN2, the MTC that occurs is usually small; lymph node metastases are rare if the tumor is less than 10 mm, and some groups have reported nodal disease being uncommon under 11 years of age.[38] Hence, although there has been a general consensus regarding the timing of prophylactic thyroidectomy based on risk stratification, the extent of central lymph node dissection at prophylactic thyroidectomy is controversial. Most advocate thyroid surgery with central node dissection for MEN3. The practice of routine central-node clearance is most important in asymptomatic gene

TABLE 134.2 American Thyroid Association Risk Level and Prophylactic Thyroidectomy Testing and Therapy

American Thyroid Association Risk Level	Age of First Investigation	Age of Prophylactic Surgery	Surgical Comments
Highest	Soon after birth, calcitonin levels are high in first few months of life and limited value in timing thyroidectomy	Within the first few months of the first year of life	Central neck dissection considered if parathyroid glands can be identified and left *in situ* or autotransplanted
High	From 3 years of age physical examination, cervical US, and serum calcitonin	Before 5 years of age	Central neck dissection if serum calcitonin >40 pg/mL
Moderate	From 5 years of age physical examination, cervical US, and serum calcitonin	Timing should be based on elevated calcitonin glevel if parental concern recommend surgery at 5 years	

ATA, American Thyroid Association; *US,* ultrasound.

(From Wells SA Jr, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid.* 2015;25:567–610.)

carriers with the higher-risk variants but can be modified for lower-risk variants with normal preoperative calcitonin. ATA guidelines suggest that, if basal calcitonin is less than 40 pg/mL preoperatively, a central neck dissection could be avoided.[38] Considering the need for lateral lymph node dissection is also challenging. Basal calcitonin has not been shown to predict the need for prophylactic lateral lymph node dissection.[58]

THE ROLE OF SOMATIC MUTATIONS

Sporadic Medullary Thyroid Carcinoma

Somatic mutations of *RET* are observed in around 60% of sporadic MTC cases. The most common somatic *RET* mutation is the M918T variant.[59] The reported prevalence of this mutation may vary based on geographic differences in population studies or perhaps methodologic differences. Other somatic mutations reported in sporadic MTC include codon 883 in exon 15 and codon 634 in exon 11. The somatic *RET* codon 918 variant is associated with a poorer prognosis in sporadic MTC, with several series correlating the mutation with distant metastases and tumor recurrence.[60-62] Dual *RET* mutations have been correlated with poorer outcome.[63] Somatic mutations of the protooncogene *RAS* (predominantly *HRAS* and *KRAS*) occur in a proportion of sporadic MTCs, nearly exclusively those tumors that are *RET* mutation–negative, suggesting that *RAS* mutations provide an alternative pathogenic pathway in sporadic MTC. In contrast to *RET* mutations, there is no evidence that harboring a *RAS* mutation predicts poorer outcome.[64] *RAS* mutations in sporadic MTC may even procure a positive outcome, as these tumors demonstrate a less aggressive phenotype.[65]

Low-Penetrance Phenotypes

With the advent of large, population-based genetic databases, it has become apparent that some RET variants formerly thought of as rare are in fact more common in the community than expected.

These variants thus have a lower penetrance for MEN2, and this is particularly obvious for *RET* p.Val804Met.[55] Using the ExAC database, the penetrance of p.Val804Met has been estimated at 46%, which was corroborated by a German cohort where only 51% of patients with this variant developed MTC.[66] One theory for the low penetrance for this particular *RET* variant is the modification by coinheritance of other *RET* variants in *cis*. Ongoing research is needed when considering deferring prophylactic thyroidectomy and replacing it with close clinical surveillance for carriers of p.Val804Met.

Hereditary Medullary Thyroid Carcinoma

Somatic *RET* mutations have also been described in hereditary MTC. The somatic codon 918 *RET* variant has been reported, albeit rarely, in association with germline *RET* mutations in exons 10 or 11.[67] This important observation negated the theory that the codon 918 somatic *RET* variant could be used as a specific marker of sporadic disease. Others have observed the somatic codon 918 *RET* variant occurring in patients with a germline codon 768 *RET* variant.[68] Microdissection techniques have shown that somatic *RET* mutations may be found in some regions of an MTC tumor but not in others.[12] It has been suggested that the presence of a somatic mutation in addition to the germline *RET* mutation provides a "second hit" that may explain the more aggressive behavior of some familial cases of MTC.

Whole-exome studies have identified a novel germline *ESR2* frameshift mutation that predispose patients to CCH and MTC. Smith et al. demonstrated the mechanism of upregulation of RET activity through constitutional activation of ESR2.[69] A possible link between germline *MET* variants and MTC has also been reported.[70]

GERMLINE SCREENING

Sporadic Medullary Thyroid Carcinoma

It is now recommended that all new cases of sporadic MTC should have germline screening for *RET* variants. Unsuspected familial cases are uncovered in approximately 6% of such patients, despite an apparently negative family history.[71] This was confirmed in a series of apparently sporadic cases of MTC, in which 7.3% of subjects were found to possess a germline mutation. Lower-risk *RET* mutations in the noncysteine codons were most commonly discovered, including codons 804, 891, and 768, although cysteine codon variants were also found. Genetic screening of first-degree relatives should be undertaken if a germline *RET* mutation is identified in a subject with MTC.

Given that more than 97% of MEN2 families have identifiable *RET* mutations, a negative germline *RET* mutation test provides a high degree of certainty that a particular individual has sporadic MTC. This is assuming that the genetic analysis is comprehensive. It is important to include exons 13, 14, and 15, because mutations in these exons are most likely to manifest late, with a lower prevalence of pheochromocytoma, and are more likely to escape detection as familial.

Whole-gene expression profiling of both hereditary and sporadic MTC showed no distinct differences between each group, suggesting a homogenous nature of MTC.[72] *NTRK3* dysregulation was disproportionally observed in in MEN3 over MEN2 samples, which may be a component of its clinical aggressiveness.

NATURAL PROGRESSION AND BEHAVIOR OF MULTIPLE ENDOCRINE NEOPLASIA TYPE 2 AND MULTIPLE ENDOCRINE NEOPLASIA TYPE 3

Staging of Medullary Thyroid Carcinoma

Once the initial diagnosis of MTC has been made, either through the presentation of thyroid nodule or screening of familial presentations, both biochemical testing and imaging should be performed to stage the extent of disease, as this may alter the goals of management and the prognosis. Once surgery has been performed, certain histological features may also be useful in determining prognosis of MTC.

KEY POINTS

- Calcitonin is a sensitive tumor marker for medullary thyroid carcinoma (MTC) and can be used in staging. Structural imaging (ultrasound scanning and magnetic resonance imaging), together with functional imaging, are both advantageous in MTC staging. (18F)-dihydroxyphenylalanine positron emission tomography (PET) demonstrates the highest sensitivity of the PET tracers for MTC.

Biochemical Markers. Calcitonin is a sensitive tumor marker for MTC. Rarely, other non–C-cell tumors may secrete calcitonin, including carcinoid tumors and small cell lung carcinomas. Calcitonin elevation is useful in confirming early diagnosis of MTC, and monitoring can indicate persistent or recurrent disease. Levels above 100 ng/mL confirm diagnosis, with the level corresponding with the volume of disease. Lower levels (10–40 ng/mL) may be found in nonmalignant conditions, such as autoimmune thyroiditis, chronic renal failure, chronic pulmonary disease, trauma, pregnancy, hypercalcemia, hypergastrinemia, pheochromocytoma, mastocytosis, and sepsis.[73] A metaanalysis in 2020 looked at 16 studies with 72,368 participants who had had a calcitonin level assessed during investigation of thyroid nodule. Results showed a calculated sensitivity of 100% (range within studies of 83%–100%) and specificity of 97.2% (range within studies of 94%–100%) with a threshold of 10 pg/mL.[74] However, 15 of the 16 studies lacked long-term follow-up of calcitonin participants, likely causing bias towards use of calcitonin in nodule diagnosis. Calcitonin levels may be measured within fine-needle aspirate biopsy washout to determine the risk of MTC.[75] The preoperative calcitonin level can also determine the risk of metastatic disease, with lymph node metastases very unlikely if below 53 ng/mL but likely if greater than 500 ng/mL.[73,75] Calcitonin levels above 1000 ng/mL are highly suggestive of distant metastases. The level of calcitonin may guide surgical management of localized disease, with calcitonin levels greater than 200 ng/mL suggesting a need for contralateral lateral neck dissection in addition to total thyroidectomy, bilateral central neck dissection, and ipsilateral central neck dissection.[38] Postoperative levels less than 10 ng/mL are suggestive of "biochemical remission." Nonsecretory MTC can evidently not be biochemically staged with calcitonin.[41] CEA, along with other less used markers including neuron-specific enolase, chromogranin A, and procalcitonin, have also been useful in staging the extent of MTC, particularly in those with calcitonin-negative disease.

Structural Imaging. Structural imaging of MTC should include imaging for MTC, phaeochromocytoma, and hyperparathyroidism. The current staging for MTC is via the eighth edition of the American Joint Committee on Cancer staging system. Unlike differentiated thyroid carcinoma, age is not considered in staging. Stage I denotes T1 tumor, which is less than 2 cm and confined to the thyroid, without nodal or metastatic spread. Stage II denotes a T2 tumor, between 2 and 4 cm, or a T3 tumor greater than 4 cm but confined to the thyroid, without nodal or distant metastatic spread. Stage III is any size tumor with local nodal spread, without distant metastases. Stage IV is where the cancer has grown to nearby tissues of the neck (IVA), or tumor has grown to invade local structures (IVB), or there is distant metastases (IVC).

MTC is most often detected with sonographic examination, with an often calcified hypoechoic, sometimes cystic, nodule, which can also be seen on computed tomography (CT) and magnetic resonance imaging (MRI).[76] The size of the primary tumor correlates with the likelihood of nodal metastases, with tumors less than 1 cm predicting 20% to 30% nodal metastases and tumors greater than 4 cm predicting 90% nodal metastases.[77]

The intensity of search for metastases will be guided by the calcitonin levels. There is a paucity of systematic studies assessing various radiographic imaging protocols. One study compared ultrasonography, spiral CT scan, MRI of the liver and whole body, and (^{18}F)-fluorodeoxyglucose (FDG) positron emission tomography (PET) in a group of 55 patients with MTC and persistently elevated calcitonin.[78] They found neck ultrasound to be the most sensitive tool in detecting lymph node metastases. Liver metastases are best seen in dynamic contrast-enhanced MRI, appearing as hyperechoic. In a study comparing MRI liver to functional imaging ^{18}F-DOPA-PET/CT, MRI was superior and more precise in detecting liver lesions.[79] Lung metastases are best seen on CT or FDG-PET. Bone scan complemented MRI in the detection of axial skeletal metastases but was superior in assessing peripheral skeletal lesions.

Functional Imaging. A variety of nuclear imaging methods have been used in an attempt to take advantage of the specific expression of hormonal receptors or transporters by endocrine tumors to allow disease localization. Unfortunately, no single method is sufficiently sensitive or specific for definitive diagnosis. Radionuclide agents that have been previously investigated include ^{201}T1-chloride, ^{131}I-metaiodobenzylguanidine (MIBG), ^{99m}Tc pentavalent-dimercapto-succinic acid, ^{131}I anti-CEA or anticalcitonin antibodies or fragments, labeled somatostatin receptors, and technetium-99m-sestamibi (^{99m}Tc-sestamibi).[80]

In recent years, the utility of PET imaging has been increasingly studied in MTC, initially using FDG and (^{18}F)-dihydroxyphenylalanine (DOPA) and more recently the somatostatin analogs ^{68}Ga-DOTATATE and ^{68}Ga-DOTATOC. FDG-PET has been widely used in oncology, and its combination with CT imaging has enhanced diagnostic accuracy. Initial results using FDG as a tracer in MTC suggested disappointingly low sensitivities; however, small studies suggest that FDG-PET positivity may be more useful for identifying metastatic deposits in patients who have shorter calcitonin doubling times, and perhaps more aggressive disease.[78] A retrospective series of 29 patients with MTC found calcitonin doubling time to be shorter in PET-positive patients (n = 14) than those who were FDG-PET–negative (n = 15), and ten of the 11 PET-positive patients for whom clinical data were available had either died from metastatic disease or experienced disease progression, compared with 11 of the 12 PET-negative patients for whom data were available, who remained free from clinical disease.[81] Another recent study suggested that FDG-PET was useful in the small subset of study patients who had shorter CEA doubling times.[82]

DOPA is a positron-emitting analog of dopamine that has shown particular promise in imaging the sympathetic nervous system (e.g., pheochromocytomas). It may have greater sensitivity than FDG as a PET tracer in MTC: a metaanalysis of eight studies reported a pooled detection rate of 66% per patient and 71% per lesion, increasing to 86% in those patients with calcitonin levels great than or equal to 1000 ng/L or doubling times less than24 months.[83] There is interest in whether DOPA- and FDG-PET/CT studies may be complementary.[82] A retrospective study of 47 patients with recurrent MTC found FDG-PET positivity to be associated with progressive disease and reduced survival, while DOPA-PET showed greater sensitivity in detecting lesions, correlating with calcitonin and CEA levels as an indicator of tumor burden.[84]

The radiolabelled somatostatin analog ^{68}Ga-DOTATATE has a broader role in diagnosis and staging of neuroendocrine tumors, and small studies have examined its role in MTC. One small retrospective study compared ^{68}Ga-DOTATATE and FDG-PET/CT in 18 patients with recurrent MTC, reporting a similar sensitivity of 72.2% for ^{68}Ga-DOTATATE and 77.8% for FDG; however, FDG-PET/CT identified a greater number of lesions, particularly in the liver.[85] Another retrospective study of 18 patients with MTC and elevated serum calcitonin levels compared DOPA, FDG, and two different ^{68}Ga-somatostatin analogs, DOTANOC (14 patients) and DOTATOC (four patients), as PET/CT tracers. It reported a sensitivity of 72.2% for DOPA, significantly greater than 16.7% for FDG, and 33.3% for ^{68}Ga-somatostatin analogs, with the latter two sensitivities not being significantly different. Imaging was negative for all tracers in five patients who had calcitonin doubling times greater than 24 months (i.e., indolent disease).[86] This study is the only multicenter trial comparison trial looking at all three tracers. Somatostatin receptor–PET was studied in a metaanalysis and showed a detection rate per patient-based analysis of 63.5%.[87] The detection rate rose to 83% in those with serum calcitonin greater than 500 ng/L, although faster doubling time did not correlate with improved diagnostic performance, as seen using DOPA-PET. Most studies have small numbers and variable diagnostic performance, and overall these analogs seem inferior to DOPA-PET, but may complement the role of FDG-PET.

A metaanalysis of different PET radiopharmaceuticals in 2020 included 14 trials with 306 patients using five different PET radiopharmaceuticals (FDG, DOPA, ^{68}Ga-somatotatin analogs, (3-)-methyl-6-(F) fluoro-DOPA, and C-methionine) for the detection of MTC.[88] Of these, DOPA-PET had the best performance for detection of metastatic MTC on both patient- and lesion-based analysis (odds ratios 2.44 and 5.74, respectively), and was found to be superior to FDG-PET.

Histological Grading. The diagnosis of MTC is confirmed with surgical pathology, with characteristic features including solid and nesting growth patterns, along with amyloid deposits. Some variants may mimic papillary thyroid cancer (PTC), and immunohistochemistry for calcitonin is required for confirmation.[38] A number of histological, clinical, and genetic features were interrogated in 76 patients in a retrospective cohort in an attempt to create a histological grading system that could predict prognosis.[89] The clinical features of age and MEN2, as well as histological features of Ki-67 proliferative index, mitotic count, and coagulative necrosis, were predictive of prognosis. This grading system used three histological features and divided MTC into low-grade (82%), intermediate-grade (12%), and high-grade (7%). Overall survival for each grade was 193, 146, and 45 months respectively, and the difference was highly significant even after correction for age and MEN2 syndrome (P = .0001). A second study looking at 144 patients with MTC also found that tumor necrosis and high mitotic rate (as determined by greater than five mitoses per ten high-powered fields) correlated with disease-specific survival, whereas tumor necrosis only correlated with local recurrence–free survival and distant metastasis–free survival. These appeared to outperform calcitonin level and stage.[90]

Future Directions for Staging. Liquid biopsy is the detection of mutations in circulating plasma cell-free DNA and is emerging as a less invasive tool in diagnoses and treatment outcome in many cancers. Cote et al. demonstrated they could detect the *RET* codon 918 variant in patient plasma with high specificity and low sensitivity, and that it correlated with worse overall survival (OS).[91] MicroRNAs are small noncoding RNAs that function in RNA silencing and posttranscriptional regulation of gene expression. They promote chemoresistance, can predict prognoses, and may assist in staging in the future. MicroRNA expression profiling has demonstrated that miR-375 expression is a negative prognostic marker for MTC,[92] and miR-153-3p has been shown to be a RET-regulated tumor suppressor microRNA in MTC.[93] These assays remain largely in research settings for now but will likely infiltrate clinical practice in the near future.

Hereditary Versus Sporadic Medullary Thyroid Carcinoma

Hereditary MTC is characterized by the early onset of bilateral and multifocal disease. Sporadic MTC may present as a palpable thyroid mass or a nodule on ultrasound, and lymph node involvement is commonly present at diagnosis. Local lymph node spread occurs early in the course of the disease in both hereditary and sporadic forms and can be present with tumors as small as 1 cm in diameter.

A study by Raue and colleagues evaluating prognostic factors in MTC found that the stage of disease at diagnosis, age, sex, and type of disease (sporadic vs. hereditary) were relevant prognostic factors in a univariate analysis, with better prognosis for young female patients with familial disease diagnosed at an early stage.[94] The difference in the survival rate of patients with hereditary disease versus those with the sporadic form disappeared in multivariate proportional hazards analysis, but the prognostic information provided by age and sex was still significant. The overall adjusted survival rate was 86.7% at 5 years and 64.2% at 10 years. These rates are comparable to those quoted by other studies.[95,96] Another study found only the stage of disease and presence of extrathyroidal growth at presentation with MTC to be independent predictors of clinical behavior and life expectancy. In this study, increased age at diagnosis was found to be associated with a reduction in OS; however, when adjusted for the baseline mortality rate of the general population on multivariate analysis, this effect was lost.[96] Disease-free survival and OS were no different in hereditary versus sporadic MTC in the analysis by Abraham and colleagues, but there was a significant reduction in the rates of local MTC recurrence in the hereditary group compared with the sporadic MTC patients.[97]

A retrospective analysis of MEN2 kindreds with the codon 634 mutation looked at whether the presence of pheochromocytomas in these patients was associated with a more aggressive MTC. Those who had MTC without pheochromocytoma did not differ significantly from those who had MTC and pheochromocytoma in terms of the stage of MTC or survival.[46]

Paraneoplastic Features of Medullary Thyroid Carcinoma

Although MTC presents with the presence of a thyroid mass or distant particularly liver metastases, systemic symptoms may occur as a result of secretion of hormones from the tumors, particularly with increased tumor load with metastatic disease. High levels of calcitonin can cause flushing, particularly of the face, and refractory diarrhea. In some patients this may be the presenting symptom, and measurement of calcitonin should be performed as part of initial investigation. Diarrhea is caused by a combination of hypersecretion and increased gastrointestinal mobility.[38] Management is based upon reducing MTC burden, both surgically and with systemic therapies, including tyrosine kinase inhibitors (TKIs; discussed later). Nutritional status should be monitored, with avoidance of alcohol and increased intake of high-fiber foods. Antimotility agents (loperamide and codeine) have limited benefit, as do somatostatin analogs,[98,99] even in combination with interferon alpha.[100] There are reports of the anticholinergic agent glycopyrrolate assisting in resistant cases.[101] The serotonin-3 receptor antagonist ondansetron was reported in a nonrandomized series to relieve diarrhea in six patients with metastatic small bowel neuroendocrine tumors, and anecdotally can improve MTC-associated diarrhea.[102] Local therapy to large liver metastases may also improve diarrhea.[103]

Ectopic Cushing syndrome is an uncommon but debilitating paraneoplastic manifestation of MTC and accounts for 1% to 3% of Cushing syndrome cases and 2.2% to 7.5% of all ectopic ACTH cases.[104] It features in only 0.7% of cases of MTC but is known particularly for its association with MEN2-related MTC.[105,106] Patients will often have metastatic disease, with significant morbidity secondary to the Cushing's, including weight gain, abdominal striae, lethargy, diabetes, and myopathy. Histological confirmation may be challenging, as immunohistochemistry staining for ACTH is not always positive because of reduced stored granules and increased secretion of the hormone from the tumors.[107] Investigations include confirmation of cortisol excess, ACTH elevation, and localizing studies including MRI pituitary and[68] Ga-DOTATATE PET. Patients who develop Cushing syndrome have significantly increased mortality, accounting for 50% contribution to overall mortality in MTC with Cushing syndrome. Again, the mainstay of treatment includes reducing tumor burden surgically but also includes antiadrenal therapies, including medical therapy (ketoconazole, mitotane, and metyrapone) or surgical intervention with bilateral adrenalectomy, which can anecdotally significantly improve quality of life. TKIs can offer disease control in difficult cases. Vandetanib has been shown to control ACTH secretion and symptoms of Cushing syndrome,[108-110] although clinicians must be mindful of amplifying adverse effects (AEs), as ketoconazole, mitotane, and vandetanib all cause QTc prolongation. A case report of sorafenib resolving hypercortisolism has also been published.[111]

KEY POINTS

- Paraneoplastic features of medullary thyroid carcinoma are often attributed to high levels of calcitonin causing refractory diarrhea. Ectopic Cushing syndrome is a rare but debilitating complication. Bilateral adrenalectomy or treatment with the tyrosine kinase inhibitor vandetanib can improve quality of life.

Pheochromocytoma

Pheochromocytomas related to MEN2 tend to be bilateral and of adrenal origin, best visualized on CT or MRI structurally and by functional imaging using ^{68}Ga-DOTATATE PET/CT scan, which has replaced the role of MIBG scan.[112] Structural imaging often reveals areas of calcification and/or necrosis, with avid enhancement on dynamic contrast-enhanced MRI. Pheochromocytomas almost always have a precontrast CT density of at least 10 Hounsfield units, although low-density areas can occur, with intratumoral fat or necrosis.[76]

Hyperparathyroidism

Imaging of hyperparathyroidism may be via structurally via ultrasound, with a hypoechoic oval shape, and asymmetric hypervascularity. ^{99m}Tc-sestamibi uptake scan can show parathyroid adenomas, assisting with perioperative staging.[76]

MANAGEMENT OF MEDULLARY THYROID CARCINOMA

Localized Medullary Thyroid Carcinoma

Complete surgical removal of all neoplastic tissue offers the only potential for cure, and alternative forms of treatment using radiotherapy or chemotherapy provide little benefit. As mentioned earlier, whenever a patient is assessed for thyroid surgery, fractionated metanephrines should be measured before undergoing general anesthesia to exclude unrecognized pheochromocytoma. Serum calcium and parathyroid hormone levels should ideally also be measured preoperatively. Clinical guidelines for surgical decision-making include those published by the ATA in 2015.[38] Overall, 40% to 50% of MTC patients will have lymph node involvement at presentation. Patients with MEN3 and sporadic MTC have a higher rate of extrathyroidal spread than do those with MEN2. The primary neck operation should include total thyroidectomy and removal of all lymph nodes in the paratracheal central neck compartment (level VI), bounded by the hyoid bone, sternal notch, and internal jugular veins. Lymph nodes in the upper mediastinum and lateral neck regions (levels II to V) should be removed if clinically or radiologically abnormal, and some authors advocate prophylactic lateral neck dissection if lymph node metastases are present in the adjacent central compartment. In individuals with advanced local or extensive distant disease, less aggressive neck surgery may be considered, with the goal of obtaining local disease control and symptom palliation with less morbidity to speech, swallowing, and parathyroid function.[38]

Recurrent Medullary Thyroid Carcinoma

Persistent elevation of serum calcitonin and CEA are useful markers for the presence of residual or recurrent MTC. The most common sites of distant metastases from MTC are the liver, lung, bone, and brain. Regional spread in the neck and chest is very common and can occur early.

Localization of recurrence is often very difficult, particularly when calcitonin levels are only modestly elevated. One commonly faces the problem of failure of detection of metastatic disease, even when its presence is suggested by persistently elevated serum calcitonin or CEA. The biological behavior of MTC varies between subjects. In addition to being tumor markers, the doubling times of calcitonin and CEA are prognostic factors for progressive disease and poor survival.[113,114] These doubling times may be used to guide the frequency and intensity of serial imaging in most patients. One study compared calcitonin and CEA doubling times in 55 patients with MTC to disease progression as measured by Response Evaluation Criteria in Solid Tumors (RECIST). The authors found the doubling times of the two markers to be strongly correlated in 80% of patients, reporting progressive disease in 94% of patients who showed doubling times of less than 24 months and stable disease in 86% of those with doubling times of greater than 24 months.[115] A subsequent metaanalysis by Meijer and coworkers of 10 retrospective studies confirmed calcitonin and CEA doubling times to be significant indicators for both clinical disease recurrence and survival, reporting a higher hazard ratio (HR) for recurrence in those patients with a calcitonin doubling time of less than 12 months (HR 5.33) compared with those with a doubling time of less than 24 months (HR 2.93). In this study, the 5-year and 10-year survival rates for patients with a calcitonin doubling time of less than 1 year was 36% and 18%, respectively, compared with 98% and 95%, respectively, for those with a calcitonin doubling time of greater than 1 year; the corresponding CEA analysis found 5-year and 10-year survival rates of 43% and 21%, respectively, for a doubling time of less than 1 year, and 100% at both time points for a doubling time of greater than 1 year.[114] Of note, calcitonin and CEA markers can fluctuate significantly in the short term, so caution in assessing trends is required.[116] In both hereditary and sporadic MTC, the calcitonin level may accelerate and decelerate over time, reflecting the unpredictability of MTC disease.[117]

Selective Venous Catheterization. Historically, selective venous catheterization was regarded as a sensitive tool for the localization of residual disease and early detection of distant metastases in MTC.[118] However, in light of newer imaging modalities, it now has limited practical utility.

Structural and Functional Imaging. The large number of imaging modalities used to investigate persistently elevated calcitonin reflects the inadequacies of each modality. Reports in the literature are limited by the lack of use of a gold standard of histologic confirmation. In considering restaging with structural and functional imaging, refer to the description discussed in primary staging section earlier.

Persistent and Metastatic Medullary Thyroid Carcinoma

Local Therapies

Reoperation. Elevated levels of calcitonin persist after primary operation in many patients with MTC. Failure of normalization of calcitonin is more likely in subjects with a higher preoperative calcitonin level and in those in whom nodal metastases were confirmed at surgery.[119] Because of the lack of a highly sensitive and specific diagnostic tool to detect micrometastases postoperatively, one is often confronted with the unique situation of whether to reoperate without definite localization and, if so, to what extent. There have been no clinical trials performed to date in which patients have been randomized to observation or reoperation.[38]

Radiotherapy. Routine adjuvant neck irradiation postsurgery confers little demonstrated benefit and may be associated with significant morbidity.[38] A retrospective study by Brierley and colleagues in 73 patients demonstrated no overall difference in the local or regional relapse-free rate between those who received external irradiation and those who did not.[120] A systematic review in 2019 suggested that the risk of locoregional relapse was minimized in patients at high risk, including those with extrathyroidal extension and nodal involvement, but of concern is the potential for tissue damage complicating future surgery.[121] Radiotherapy may have a role in controlling local disease, which is an important quality-of-life endpoint in MTC.[38] If used, it appears that 60 Gy given within the first 2 months of surgery was most effective.[121]

Systemic Therapy

Somatostatin Analogues. Therapy with different formulations of somatostatin analogues does not demonstrate antitumor activity in MTC, in contrast to various neuroendocrine tumors. No significant decrease in the size of metastases has been found with the use of these somatostatin analogues in patients with advanced MTC.[122] However, such agents have a role in symptomatic control of symptoms of hormonal excess.[123]

Multikinase Inhibitors. Recent decades have seen an increased understanding of the role of receptor tyrosine kinases in a number of human malignancies, including the role of oncogenic RET in MTC, as discussed in this chapter. These cell surface transmembrane proteins, activated by hormones, growth factors, or, in disease states, genetic mutation, induce intracellular signaling pathways to modulate cellular activities, including proliferation, migration, and invasion. Inhibition of these cascades in neoplastic tissue has formed the basis of research into targeted molecular therapies. Most of the agents developed to date lack specificity for a single kinase and are best described as multikinase inhibitors (MKIs). Earlier studies of MKIs demonstrated modest benefit, but more targeted TKIs with greater specificity are proving useful.[38,124] The constitutive expression of the active RET receptor tyrosine kinase in MTC offers an enticing therapeutic target. The MKIs that have been assessed to date in phase III clinical trials are listed in Table 134.3, and phase I/II/III trial data are discussed in the following section.

Axitinib, a selective inhibitor of vascular endothelial growth factor receptor (VEGFR) 1, 2, and 3, but not RET, was evaluated in a phase II study of advanced thyroid cancer patients, including 11 patients with MTC. The small study suggested modest clinical efficacy: two patients demonstrated a partial response, and three patients demonstrated stable disease.[125] In a later study, a cohort of 13 MTC patients who received compassionate access showed median progression-free survival (PFS) of 9.4 months (95% confidence interval [CI]: 4.8–13.9).[126] Motesanib, an oral agent that targets VEGFR1, 2, and 3, platelet-derived growth factor receptor (PDGFR), and c-KIT (stem cell factor receptor), as well as wild type RET *in vitro*, was studied in 91 patients with locally advanced or metastatic MTC, of whom just 2% demonstrated a partial response and 48% demonstrated stable disease beyond 24 weeks.[127] Vandetanib, an oral inhibitor of RET, VEGFR, and epidermal growth factor receptor signaling, has been assessed in two phase II trials with patients with advanced hereditary MTC, at doses of 300 mg/day (n = 30) and 100 mg/day (n = 19).[128,129] Disease control was achieved in 73% and 68% of patients, respectively, and a phase III study in patients with advanced hereditary or sporadic disease has subsequently been reported, as discussed later.[130] The MKI sunitinib, targeting VEGFR1 and 2, PDGFR, c-KIT, Fms-related tyrosine kinase (FLT3), and RET, was assessed in a phase II study of advanced iodine-refractory differentiated thyroid carcinoma and metastatic MTC. The MTC cohort included seven patients, of whom six were able to be evaluated for clinical response, and three achieved an objective partial response by RECIST.[131] Twenty-six MTC patients received sunitinib as first-line antiangiogenic therapy in another phase II trial and showed an objective response rate (ORR) of 38.5% (95% CI: 22.6–56.4) with median PFS 16.5 months.[132] Of note, this study also had considerable AEs, with 14.1% of all patients (combined PTC, anaplastic thyroid cancer, and MTC) having a cardiac event and frequent grade III AEs.

Sorafenib, an orally available MKI that targets VEGFR2 and 3, RAF kinases C-RAF and B-RAF, PDGFR, c-KIT, FLT3, and RET, has had limited success in MTC treatment. In a phase II study that included 16 patients with advanced sporadic MTC, one individual achieved a partial response, 14 demonstrated stable disease, and one individual was not evaluable by RECIST and demonstrated clinical progression; *post hoc* analysis of the 10 patients who had demonstrated progressive disease before study entry included one partial response and four patients with stable disease beyond 15 months on study drug.[133] A subsequent study evaluating sorafenib in advanced thyroid carcinoma of any histopathologic type reported a partial response at 12 months in 25% of MTC patients.[134] Ito et al. showed an ORR of 25% and disease control rate of 75% in eight MTC patients.[135] Cabozantinib, an oral inhibitor of MET (also known as *hepatocyte growth factor receptor,* a receptor

TABLE 134.3 Phase III Trials and US Food and Drug Administration-Approved Tyrosine Kinase Inhibitors in Medullary Thyroid Carcinoma

Tyrosine Kinase Inhibitor	Target	Phase	Population	Result	Common Adverse Events
Cabozantinib[142]	VEGFR, c-Kit, ERT, MET, RET	III	Progressive metastatic disease	PFS 11.2 months, OS 44.3 months[a143]	[b]Diarrhea, PPE, weight loss, anorexia, nausea, fatigue, dysgeusia, hair color changes, HTN, constipation, hemorrhage
Vandetinib[130]	VEGFR, c-Kit, RET, EGFR, RET	III	Stable metastatic disease	PFS 30.5 months	[b]Diarrhea, rash, nausea, hypertension, fatigue, headache, anorexia, acne, asthenia, vomiting, back pain, dry skin, insomnia, abdominal pain, dermatitis acneiform
Selpercatinib[124]	RET	II	Stable or progressive metastatic disease	69% objective response at 1 year, 86% response ongoing	Hypertension, increased alanine aminotransferase and aspartate aminotransferase, hyponatremia, and diarrhea

[a]Not significant.

[b]AEs were observed in >25% of patients treated with vandetanib or cabozantinib. In patients treated with selpercatinib, even the most frequent AE was only found in 21% of cases.

FDA, US Food and Drug Administration; *AE*, adverse event; *TKI*, tyrosine kinase inhibitor; *MTC*, medullary thyroid carcinoma; *PPE*, palmar-plantar erythrodysethesia; HTN, hypertension; PFS, progression-free survival; OS, overall survival.

RECIST or Response Evaluation Criteria in Solid Tumours is a very standard description for measuring response see: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/recist.

tyrosine kinase involved in invasive growth, metastasis, and angiogenesis[136]), VEGFR2, and RET kinase, was suggested to demonstrate clinical benefit in the MTC subset of a phase I study of advanced solid malignancies: a partial response was observed in 29% of the 37 patients with MTC, and 41% achieved stable disease of at least 6 months. Lenvatinib, another antiangiogenic MKI that inhibits VEGFR1 to 3, FGFR1 to 4, PDGFRα, RET, and v-kit Hardy–Zuckerman 4 feline sarcoma viral oncogene homolog (KIT), is approved for radioiodine-refractory differentiated thyroid cancer and was assessed in a phase II study of MTC patients.137 In the cohort of 59 patients, ORR was 36% (95% CI: 24–49), with all partial responses. Median PFS was 9 months (95% CI: 7–not evaluable).

The combination of sorafenib and tipifarnib (a farnesyltransferase inhibitor that inactivates RAS and other farnesylated proteins) has been reported in a phase I study of advanced thyroid malignancy. Of the 13 patients with MTC, 38% had a partial response (duration ranging from 9+ to 34+ months), and 31% had stable disease of at least 6 months.[138] There is also evidence that the mammalian target of rapamycin (mTOR) pathway is activated in MTC tissue,[139] and *in vitro* data suggest that dual blockade of RET and mTOR kinase activities results in suppression of MTC cell growth.[45] Everolimus, an mTOR inhibitor, has been shown to be antiproliferative to MTC cells *in vitro*,[140] and a small ($n = 7$) phase II trial demonstrated clinical efficacy, with 71% showing stable disease.[141]

Two phase III studies of MKIs in MTC have been reported, with the resultant approval of vandetanib and cabozantinib by the US Food and Drug Administration (FDA) for use in progressive metastatic disease in adults. The vandetanib study included 331 patients with sporadic or hereditary advanced MTC and reported prolonged PFS with vandetanib compared with placebo (HR 0.46, 95% CI: 0.31–0.69); predicted median PFS was 30.5 months with vandetanib and 19.3 months with placebo.[130] Cabozantinib was assessed in 330 patients with progressive, locally advanced, or metastatic MTC: median PFS with cabozantinib was 11.2 months, compared with 4.0 months with placebo (HR 0.28, 95% CI: 0.19–0.40).[142] OS was assessed in 2017 and showed a statistically nonsignificant increase in OS of 26.6 versus 21.1 months in patients treated with cabozantinib versus placebo (HR 0.8595%, 95% CI: 0.64–1.12; $P = 0.24$).143 The above MKIs are all associated with AEs, including diarrhea, hypertension, fatigue, and skin rashes, often necessitating dose reductions or interruptions. Vandetanib and cabozantinib are also associated with prolonged QT intervals, palmar-plantar erythrodysesthesia, and mucositis.

The development of the specific RET kinase inhibitors LOXO292 (selpercatinib) and BLU667 (pralsetinib) has created a new paradigm in systemic therapy for metastatic MTC. These more specific RET kinase inhibitors have shown not only a better disease control rate, albeit in phase I/II trials, but also fewer and less severe AEs.[144] Among 49 response-evaluable MTC patients, pralsetinib demonstrated an ORR of 47% (95% CI: 33–62).[145] In a phase I/II trial, selpercatinib demonstrated a 1-year PFS rate of 82% (95% CI: 62–82) in patients who had been previously treated with vandetanib and cabozatinib.[124] Those who had not been previously treated had an even higher 1-year PFS rate of 92% (95% CI: 82–97). AEs included hypertension (21%), transaminase rise (9%–11%), hyponatremia (8%), and diarrhea (6%). However, in contrast to the other TKIs mentioned previously, adverse events were far less debilitating, and withdrawal from drug only occurred in 2% of patients. In May 2020, the FDA has approved selpercatinib for *RET*-fusion radioiodine-refractory differentiated thyroid cancer and for patients 12 years and over with *RET*-mutant MTC.

Resistance to TKIs, particularly vandetanib, in patients harboring the *RET* codon 804 variant is well described.[146] Selpercatinib has been reported to overcome both hereditary and acquired V804M resistance, but a recent report showed the development of solvent-front mutations leading to acquired resistance in two MTC patients receiving selpercatinib.[147]

There is much interest in whether tumor genotype may determine response to drug in an individual. Patients on vandetanib harboring the M918T mutation exhibited a higher response rate.[130] Analysis of the cabozantinib study found the presence of any *RET* mutation to be associated with a longer median PFS on drug compared with wild-type *RET* (60 weeks vs. 25 weeks, $P = 0.0001$), and, in addition, found mutation M918T to show the longest median PFS on cabozantinib treatment compared with any other mutation (61 weeks vs. 36 weeks, $P = .009$). The small number of patients found to harbor a *RAS* gene mutation showed similar treatment responses to *RET*-mutated patients when treated with cabozantinib.[148] Lenvatinib disease response rates showed no difference between *RET* variants.137

KEY POINTS

- Traditional treatment, including chemotherapy and radiation, has been surpassed by targeted treatment with tyrosine kinase inhibitors. Multikinase inhibitors have shown increased progression-free survival in phase III trials, but adverse effects can impair quality of life. Early trials of more targeted RET kinase inhibitors have shown promising overall response rates and appear to be better tolerated.

Radionuclide Therapy. Radioactive iodine (^{131}I) has not demonstrated efficacy in treatment of MTC.[149] This is attributed to the fact that MTC arises from parafollicular C-cells of the thyroid gland, which do not concentrate iodine. There has been limited experience with therapeutic ^{131}I-MIBG in MTC. A recently reported retrospective series of ^{131}I-MIBG therapy in 13 patients with metastatic MTC described a partial response in four individuals and stable disease in four individuals.[150] Randomized, prospective data are lacking. Other attempts to deliver targeted radioactivity are under ongoing investigation. One approach has been to administer monoclonal antibodies coupled with radioisotopes as pretargeted radioimmunotherapy. A recent phase II study reported disease stabilization and prolonged survival with two-part radioimmunotherapy against CEA in patients with rapidly progressing metastatic MTC, although hematologic toxicity was observed.[151] Other groups have used yttrium-90–labeled somatostatin analogues for targeted radiotherapy, with modest success. A phase II trial evaluating ^{90}Y-DOTATOC reported a partial response in 29% of the 13 patients.[152] Forty-three somatostatin receptor–positive metastatic MTC patients were treated with ^{177}Lu–DOTATATE peptide receptor radionucleotide therapy (PRRT), and 61% were responders based on molecular imaging (^{68}Ga–DOTATATE) with prognostic variables relating to size of lesion (<2 cm) and lower FDG uptake (maximum standardized uptake value <5). Some 62% were responders on anatomical imaging using RECIST criteria, demonstrating that PRRT has the utility as a therapeutic tool in metastatic MTC.[153]

Chemotherapy. Conventional cytotoxic chemotherapy has been of limited value and is not recommended as first-line therapy for patients with persistent MTC.[38] Doxorubicin has been the most widely used single chemotherapeutic agent in the treatment of MTC, and therefore is the one associated with most reported responses. Dacarbazine is the most widely used agent in combination therapy. Petursson reported a patient with metastatic MTC who demonstrated a complete response with combination chemotherapy of dacarbazine and 5-fluorouracil.[154] Combined chemotherapy has been tried in patients with advanced

MTC using cyclophosphamide, vincristine, and dacarbazine; this yielded only moderate antitumor activity.[155] Small studies have investigated selective arterial chemoembolization to treat hepatic metastases, with some tumor response and symptom palliation.[156,157]

MANAGEMENT OF PHEOCHROMOCYTOMA

Early identification of pheochromocytoma is important to prevent the associated morbidity and the risks associated with unsuspected pheochromocytoma in pregnancy or with intercurrent surgery. When pheochromocytoma is confirmed, the patient will require alpha blockade, likely plus beta blockade, as discussed in Chapter 110. Molecular imaging modalities such as ^{68}Ga–DOTATATE localize pheochromocytomas at a smaller size than in the past,[112] and surgical techniques have also advanced. Confirmed pheochromocytomas should be removed before thyroidectomy for MTC.[38] Laparoscopic adrenalectomy, ideally with the posterior retroperitoneal approach, is now the operation of choice, where available.[158] There is interest in cortical-sparing adrenalectomy, particularly when bilateral adrenalectomy is required and when patient compliance with steroid replacement is likely to be problematic. However, the cortical-sparing approach may be associated with higher recurrence rates, and there is a lack of long-term studies.[159]

MANAGEMENT OF HYPERPARATHYROIDISM

The surgical approach to hyperparathyroidism is less aggressive than in MEN1 owing to the different behavior of the parathyroid disease in the two conditions. Total parathyroidectomy is not recommended, and recurrences are uncommon. However, an open neck exploration is preferred to minimally invasive surgery.[51,52] Preoperative localization studies are only recommended in patients who have had previous neck surgery. Only enlarged parathyroid glands should be removed, and forearm parathyroid autografting is recommended at the initial parathyroid surgery. This is because the risk for MTC recurrence is such that further neck operations are likely to be needed in the future, resulting in parathyroid removal or devascularization.[38] Normal parathyroid glands may also be resected or devascularized at the time of initial thyroidectomy for MTC. Parathyroid "normal" tissue should be transplanted into the sternomastoid muscle in the neck in the case of MEN3 or FMTC patients, in whom there is a low risk for hyperparathyroidism. Devascularized parathyroid glands from patients with the codon 634 mutation or a strong family history of hyperparathyroidism should be autografted into the forearm, where portions of the graft can be removed if hyperparathyroidism develops.[38]

CONCLUSION

Therapeutic modalities for advanced MTC have improved over the past decade. Cure is still only possible with early diagnosis followed by surgery. Genetic screening for *RET* variants has revolutionized the management of familial MTC and MEN2 by enabling prophylactic thyroidectomy or early diagnosis of established disease. The more common sporadic disease remains a much greater challenge. Phase II clinical trials of the more selective RET kinase inhibitors have demonstrated efficacy, with tolerable side effects, and may offer a better prognosis and quality of life for patients with advanced MTC.

Acknowledgments

The authors wish to acknowledge A/Professor Diana Learoyd and Dr. Lyndal Tacon, who wrote earlier versions of this chapter.

REFERENCES

1. Marsh DJ, Learoyd DL, Robinson BG. Medullary thyroid carcinoma: recent advances and management update. *Thyroid*. 1995;5:407–424.
2. Ceolin L, et al. Medullary thyroid carcinoma beyond surgery: advances, challenges, and perspectives. *Endocr Relat Cancer*. 2019;26:R499–R518.
3. Gagel RF, et al. Multiple endocrine neoplasia type 2a associated with cutaneous lichen amyloidosis. *Ann Intern Med*. 1989;111:802–806.
4. Machens A, et al. Molecular epidemiology of multiple endocrine neoplasia 2: implications for RET screening in the new millennium. *Eur J Endocrinol*. 2013;168:307–314.
5. Komminoth P. The RET proto-oncogene in medullary and papillary thyroid carcinoma. Molecular features, pathophysiology and clinical implications. *Virchows Arch*. 1997;431:1–9.
6. Easton DF, et al. The clinical and screening age-at-onset distribution for the MEN-2 syndrome. *Am J Hum Genet*. 1989;44:208–215.
7. Chandrasekharappa SC, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science*. 1997;276:404–407.
8. Donis-Keller H, et al. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum Mol Genet*. 1993;2:851–856.
9. Mulligan LM, et al. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature*. 1993;363:458–460.
10. Pasini B, et al. The physical map of the human RET proto-oncogene. *Oncogene*. 1995;11:1737–1743.
11. Hofstra RM, et al. A mutation in the RET proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature*. 1994;367:375–376.
12. Eng C, et al. Heterogeneous mutation of the RET proto-oncogene in subpopulations of medullary thyroid carcinoma. *Cancer Res*. 1996;56:2167–2170.
13. Ullrich A, Schlessinger J. Signal transduction by receptors with tyrosine kinase activity. *Cell*. 1990;61:203–212.
14. Takahashi M, et al. Characterization of the ret proto-oncogene products expressed in mouse L cells. *Oncogene*. 1993;8:2925–2929.
15. Schlumberger M, et al. New therapeutic approaches to treat medullary thyroid carcinoma. *Nat Clin Pract Endocrinol Metab*. 2008;4:22–232.
16. Santoro M, Carlomagno F. Drug insight: small-molecule inhibitors of protein kinases in the treatment of thyroid cancer. *Nat Clin Pract Endocrinol Metab*. 2006;2:42–52.
17. de Groot JW, et al. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. *Endocr Rev*. 2006;27:535–560.
18. Mulligan LM, et al. Diverse phenotypes associated with exon 10 mutations of the RET proto-oncogene. *Hum Mol Genet*. 1994;3:2163–2167.
19. Donovan DT, et al. Familial cutaneous lichen amyloidosis in association with multiple endocrine neoplasia type 2A: a new variant. *Henry Ford Hosp Med J*. 1989;37:147–150.
20. Jing S, et al. GDNF-induced activation of the ret protein tyrosine kinase is mediated by GDNFR-alpha, a novel receptor for GDNF. *Cell*. 1996;85:1113–1124.
21. Klein RD, et al. A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. 1997;387:717–721.
22. Baloh RH, et al. Artemin, a novel member of the GDNF ligand family, supports peripheral and central neurons and signals through the GFRalpha3-RET receptor complex. *Neuron*. 1998;21:1291–1302.
23. Milbrandt J, et al. Persephin, a novel neurotrophic factor related to GDNF and neurturin. *Neuron*. 1998;20:245–253.
24. Schuchardt A, et al. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. *Nature*. 1994;367:380–383.
25. Treanor JJ, et al. Characterization of a multicomponent receptor for GDNF. *Nature*. 1996;382:80–83.
26. Woodward ER, et al. Genetic predisposition to phaeochromocytoma: analysis of candidate genes GDNF, RET and VHL. *Hum Mol Genet*. 1997;6:1051–1056.
27. Asai N, et al. Mechanism of activation of the ret proto-oncogene by multiple endocrine neoplasia 2A mutations. *Mol Cell Biol*. 1995;15:1613–1619.

28. Pasini A, et al. Oncogenic activation of RET by two distinct FMTC mutations affecting the tyrosine kinase domain. *Oncogene*. 1997;15:393–402.
29. Johansson E, et al. Revising the embryonic origin of thyroid C cells in mice and humans. *Development*. 2015;142:3519–3528.
30. Kameda Y. Cellular and molecular events on the development of mammalian thyroid C cells. *Dev Dyn*. 2016;245:323–341.
31. Russo AF, Clark MS, Durham PL. Thyroid parafollicular cells. An accessible model for the study of serotonergic neurons. *Mol Neurobiol*. 1996;13:257–276.
32. Rosai J. Carcangiu ML. De Lellis RA. Tumors of the thyroid gland. In: Atlas of tumor pathology, 1992.
33. Papotti M, et al. Poorly differentiated thyroid carcinomas with primordial cell component. a group of aggressive lesions sharing insular, trabecular, and solid patterns. *Am J Surg Pathol*. 1993;17:291–301.
34. Eng C, et al. The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2. International RET mutation consortium analysis. *J Am Med Assoc*. 1996;276:1575–1579.
35. Sarika HL, et al. High prevalence of exon 8 G533C mutation in apparently sporadic medullary thyroid carcinoma in Greece. *Clin Endocrinol*. 2012;77:857–862.
36. Carlson KM, et al. Single missense mutation in the tyrosine kinase catalytic domain of the RET protooncogene is associated with multiple endocrine neoplasia type 2B. *Proc Natl Acad Sci U S A*. 1994;91:1579–1583.
37. Castinetti F, et al. Natural history, treatment, and long-term follow up of patients with multiple endocrine neoplasia type 2B: an international, multicentre, retrospective study. *Lancet Diabetes Endocrinol*. 2019;7:213–220.
38. Wells Jr SA, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567–610.
39. Punales MK, et al. RET codon 634 mutations in multiple endocrine neoplasia type 2: variable clinical features and clinical outcome. *J Clin Endocrinol Metab*. 2003;88:2644–2649.
40. Jasim S, et al. Multiple endocrine neoplasia type 2B with a RET proto-oncogene A883F mutation displays a more indolent form of medullary thyroid carcinoma compared with a RET M918T mutation. *Thyroid*. 2011;21:189–192.
41. Frank-Raue K, et al. Prevalence and clinical spectrum of nonsecretory medullary thyroid carcinoma in a series of 839 patients with sporadic medullary thyroid carcinoma. *Thyroid*. 2013;23:294–300.
42. Carlson KM, et al. Parent-of-origin effects in multiple endocrine neoplasia type 2B. *Am J Hum Genet*. 1994;55:1076–1082.
43. Bolk S, et al. A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. *Proc Natl Acad Sci U S A*. 2000;97:268–273.
44. Asai N, et al. RET receptor signaling: dysfunction in thyroid cancer and Hirschsprung's disease. *Pathol Int*. 2006;56:164–172.
45. Gild ML, et al. Targeting mTOR in RET mutant medullary and differentiated thyroid cancer cells. *Endocr Relat Cancer*. 2013;20:659–667.
46. Thosani S, et al. The characterization of pheochromocytoma and its impact on overall survival in multiple endocrine neoplasia type 2. *J Clin Endocrinol Metab*. 2013;98:E1813–E1819.
47. Welander J, Soderkvist P, Gimm O. Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. *Endocr Relat Cancer*. 2011;18:R253–R276.
48. Eisenhofer G, et al. Measurements of plasma methoxytyramine, normetanephrine, and metanephrine as discriminators of different hereditary forms of pheochromocytoma. *Clin Chem*. 2011;57:411–420.
49. Eisenhofer G, Tischler AS, de Krijger RR. Diagnostic tests and biomarkers for pheochromocytoma and extra-adrenal paraganglioma: from routine laboratory methods to disease stratification. *Endocr Pathol*. 2012;23:4–14.
50. Raue F, et al. Primary hyperparathyroidism in multiple endocrine neoplasia type 2A. *J Intern Med*. 1995;238:369–373.
51. Yoshida S, et al. Long term parathyroid function following total parathyroidectomy with autotransplantation in adult patients with MEN2A. *Endocr J*. 2009;56:545–551.
52. Scholten A, et al. Evolution of surgical treatment of primary hyperparathyroidism in patients with multiple endocrine neoplasia type 2A. *Endocr Pract*. 2011;17:7–15.
53. Brandi ML, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab*. 2001;86:5658–5671.
54. Frohnauer MK, Decker RA. Update on the MEN 2A c804 RET mutation: is prophylactic thyroidectomy indicated?. *Surgery*. 2000;128:1052–1057; discussion 1057–1058.
55. Loveday C, et al. p.Val804Met, the most frequent pathogenic mutation in RET, confers a very low lifetime risk of medullary thyroid cancer. *J Clin Endocrinol Metab*. 2018;103:4275–4282.
56. Learoyd DL, Robinson BG. Do all patients with RET mutations associated with multiple endocrine neoplasia type 2 require surgery? *Nat Clin Pract Endocrinol Metab*. 2005;1:60–61.
57. Rohmer V, et al. Prognostic factors of disease-free survival after thyroidectomy in 170 young patients with a RET germline mutation: a multicenter study of the Groupe Francais d'Etude des Tumeurs Endocrines. *J Clin Endocrinol Metab*. 2011;96:E509–E518.
58. Opsahl EM, et al. The role of calcitonin in predicting the extent of surgery in medullary thyroid carcinoma: a nationwide population-based study in Norway. *Eur Thyroid J*. 2019;8:159–166.
59. Fussey JM, et al. The role of molecular genetics in the clinical management of sporadic medullary thyroid carcinoma: a systematic review. *Clin Endocrinol (Oxf)*. 2019;91:697–707.
60. Zedenius J, et al. Mutations of codon 918 in the RET proto-oncogene correlate to poor prognosis in sporadic medullary thyroid carcinomas. *J Clin Endocrinol Metab*. 1995;80:3088–3090.
61. Romei C, et al. Somatic mutations of the ret protooncogene in sporadic medullary thyroid carcinoma are not restricted to exon 16 and are associated with tumor recurrence. *J Clin Endocrinol Metab*. 1996;81:1619–1622.
62. Marsh DJ, et al. Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinoma. *Clin Endocrinol*. 1996;44:249–257.
63. Romei C, et al. New insights in the molecular signature of advanced medullary thyroid cancer: evidence of a bad outcome of cases with double RET mutations. *J Med Genet*. 2016;53:729–734.
64. Vuong HG, et al. Clinical significance of RET and RAS mutations in sporadic medullary thyroid carcinoma: a meta-analysis. *Endocr Relat Cancer*. 2018;25:633–641.
65. Cavedon E, et al. Prognostic impact of miR-224 and RAS mutations in medullary thyroid carcinoma. *Int J Endocrinol*. 2017;2017:4915736.
66. Machens A, et al. Genotype-specific progression of hereditary medullary thyroid cancer. *Hum Mutat*. 2018;39:860–869.
67. Marsh DJ, et al. Germline and somatic mutations in an oncogene: RET mutations in inherited medullary thyroid carcinoma. *Cancer Res*. 1996;56:1241–1243.
68. Miyauchi A, et al. A novel somatic mutation in the RET proto-oncogene in familial medullary thyroid carcinoma with a germline codon 768 mutation. *Jpn J Cancer Res*. 1997;88:527–531.
69. Smith J, et al. Germline ESR2 mutation predisposes to medullary thyroid carcinoma and causes up-regulation of RET expression. *Hum Mol Genet*. 2016;25:1836–1845.
70. Sponziello M, et al. Whole exome sequencing identifies a germline MET mutation in two siblings with hereditary wild-type RET medullary thyroid cancer. *Hum Mutat*. 2018;39:371–377.
71. Wohllk N, et al. Relevance of RET proto-oncogene mutations in sporadic medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 1996;81:3740–3745.
72. Oczko-Wojciechowska M, et al. Differences in the transcriptome of medullary thyroid cancer regarding the status and type of RET gene mutations. *Sci Rep*. 2017;7:42074.
73. Danila R, Livadariu R, Branisteanu D. Calcitonin revisited in 2020. *Acta Endocrinol (Buchar)*. 2019;15:544–548.
74. Verbeek HH, et al. Calcitonin testing for detection of medullary thyroid cancer in people with thyroid nodules. *Cochrane Database Syst Rev*. 2020;3:Cd010159.
75. Costante G, Meringolo D. Calcitonin as a biomarker of C cell disease: recent achievements and current challenges. *Endocrine*. 2020;67:273–280.
76. Grajo JR, et al. Multiple endocrine neoplasia syndromes: a comprehensive imaging review. *Radiol Clin North Am*. 2016;54:441–451.

77. Kim MJ, et al. US-guided fine-needle aspiration of thyroid nodules: indications, techniques, results. *Radiographics*. 2008;28:1869–1886; discussion 1887.
78. Giraudet AL, et al. Imaging medullary thyroid carcinoma with persistent elevated calcitonin levels. *J Clin Endocrinol Metab*. 2007;92:4185–4190.
79. Puhr-Westerheide D, et al. The added diagnostic value of complementary gadoxetic acid-enhanced MRI to (18)F-DOPA-PET/CT for liver staging in medullary thyroid carcinoma. *Cancer Imag*. 2019;19:73.
80. Lebouthillier G, et al. Tc-99m sestamibi and other agents in the detection of metastatic medullary carcinoma of the thyroid. *Clin Nucl Med*. 1993;18:657–661.
81. Bogsrud TV, et al. The prognostic value of 2-deoxy-2-[18F]fluoro-D-glucose positron emission tomography in patients with suspected residual or recurrent medullary thyroid carcinoma. *Mol Imaging Biol*. 2010;12:547–553.
82. Kauhanen S, et al. Complementary roles of 18F-DOPA PET/CT and 18F-FDG PET/CT in medullary thyroid cancer. *J Nucl Med*. 2011;52:1855–1863.
83. Treglia G, et al. Detection rate of recurrent medullary thyroid carcinoma using fluorine-18 dihydroxyphenylalanine positron emission tomography: a meta-analysis. *Acad Radiol*. 2012;19:1290–1299.
84. Verbeek HH, et al. Clinical relevance of 18F-FDG PET and 18F-DOPA PET in recurrent medullary thyroid carcinoma. *J Nucl Med*. 2012;53s:1863–1871.
85. Conry BG, et al. Comparison of (68)Ga-DOTATATE and (18)F-fluorodeoxyglucose PET/CT in the detection of recurrent medullary thyroid carcinoma. *Eur J Nucl Med Mol Imaging*. 2010;37:49–57.
86. Treglia G, et al. Comparison of 18F-DOPA, 18F-FDG and 68Ga-somatostatin analogue PET/CT in patients with recurrent medullary thyroid carcinoma. *Eur J Nucl Med Mol Imaging*. 2012;39:569–580.
87. Treglia G, Tamburello A, Giovanella L. Detection rate of somatostatin receptor PET in patients with recurrent medullary thyroid carcinoma: a systematic review and a meta-analysis. *Hormones (Basel)*. 2017;16:362–372.
88. Lee SW, et al. Comparison of 5 different PET radiopharmaceuticals for the detection of recurrent medullary thyroid carcinoma: a network meta-analysis. *Clin Nucl Med*. 2020;45:341–348.
89. Fuchs TL, et al. A proposed grading scheme for medullary thyroid carcinoma based on proliferative activity (Ki-67 and mitotic count) and coagulative necrosis. *Am J Surg Pathol*. 2020;44:1419–1428.
90. Alzumaili B, et al. Grading of medullary thyroid carcinoma on the basis of tumor necrosis and high mitotic rate is an independent predictor of poor outcome. *Mod Pathol*. 2020;33:1690–1701.
91. Cote GJ, et al. Prognostic significance of circulating RET M918T mutated tumor DNA in patients with advanced medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2017;102:3591–3599.
92. Romeo P, et al. Circulating miR-375 as a novel prognostic marker for metastatic medullary thyroid cancer patients. *Endocr Relat Cancer*. 2018;25:217–231.
93. Joo LJS, et al. RET kinase-regulated microRNA-153-3p improves therapeutic efficacy in medullary thyroid carcinoma. *Thyroid*. 2019;29(6):830–844.
94. Raue F, et al. Prognostic factors in medullary thyroid carcinoma: evaluation of 741 patients from the German Medullary Thyroid Carcinoma Register. *Clin Investig*. 1993;71:7–12.
95. Saad MF, et al. Medullary carcinoma of the thyroid. A study of the clinical features and prognostic factors in 161 patients. *Medicine (Baltim)*. 1984;63:319–342.
96. de Groot JW, et al. Determinants of life expectancy in medullary thyroid cancer: age does not matter. *Clin Endocrinol (Oxf)*. 2006;65:729–736.
97. Abraham DT, et al. Medullary thyroid carcinoma: long-term outcomes of surgical treatment. *Ann Surg Oncol*. 2011;18:219–225.
98. Mahler C, et al. Long-term treatment of metastatic medullary thyroid carcinoma with the somatostatin analogue octreotide. *Clin Endocrinol (Oxf)*. 1990;33:261–269.
99. Vainas I, et al. Somatostatin receptor expression in vivo and response to somatostatin analog therapy with or without other antineoplastic treatments in advanced medullary thyroid carcinoma. *J Exp Clin Cancer Res*. 2004;23:549–559.
100. Wu Q, et al. A systematic review and meta-analysis of gastrointestinal events associated with nonoperative therapies for neuroendocrine tumors. *OncoTargets Ther*. 2018;11:7655–7668.
101. Alawneh A, et al. Successful palliation of octreotide-resistant diarrhea using glycopyrrolate in a patient with medullary thyroid cancer: a case report. *J Pain Symptom Manage*. 2015;50:e4–e6.
102. Kiesewetter B, Raderer M. Ondansetron for diarrhea associated with neuroendocrine tumors. *N Engl J Med*. 2013;368:1947–1948.
103. Fromigue J, et al. Chemoembolization for liver metastases from medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2006;91:2496–2499.
104. Ilias I, et al. Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institutes of Health. *J Clin Endocrinol Metab*. 2005;90:4955–4962.
105. Kasturi K, et al. Cushing disease in a patient with multiple endocrine neoplasia type 2B. *J Clin Transl Endocrinol Case Rep*. 2017;4:1–4.
106. Barbosa SL, et al. Ectopic adrenocorticotropic hormone-syndrome in medullary carcinoma of the thyroid: a retrospective analysis and review of the literature. *Thyroid*. 2005;15:618–623.
107. Wijewardene AA, Glastras SJ, Learoyd DL, Robinson BG, Tsang VHM. ACTH-secreting medullary thyroid cancer: a case series. *Endocrinol Diabetes Metab Case Rep*. 2017;2017:16–0110. http://dx.doi.org/10.1530/EDM-16–0110. PMID: 28567290; PMCID: PMC5445425.
108. Nella AA, et al. Vandetanib successfully controls medullary thyroid cancer-related Cushing syndrome in an adolescent patient. *J Clin Endocrinol Metab*. 2014;99:3055–3059.
109. Pivovarova AI, Patrick S, Reddy PJ. Sporadic medullary thyroid carcinoma with paraneoplastic Cushing syndrome. *Case Rep Endocrinol*. 2019;2019:6414921.
110. Paepegaey AC, et al. Long-term control of hypercortisolism by vandetanib in a case of medullary thyroid carcinoma with a somatic RET mutation. *Thyroid*. 2017;27:587–590.
111. Barroso-Sousa R, et al. Complete resolution of hypercortisolism with sorafenib in a patient with advanced medullary thyroid carcinoma and ectopic ACTH (adrenocorticotropic hormone) syndrome. *Thyroid*. 2014;24:1062–1066.
112. Gild ML, et al. Role of DOTATATE-PET/CT in preoperative assessment of phaeochromocytoma and paragangliomas. *Clin Endocrinol (Oxf)*. 2018;89:139–147.
113. Barbet J, et al. Prognostic impact of serum calcitonin and carcinoembryonic antigen doubling-times in patients with medullary thyroid carcinoma. *J Clin Endocrinol Metab*. 2005;90:6077–6084.
114. Meijer JA, et al. Calcitonin and carcinoembryonic antigen doubling times as prognostic factors in medullary thyroid carcinoma: a structured meta-analysis. *Clin Endocrinol (Oxf)*. 2010;72:534–542.
115. Laure Giraudet A, et al. Progression of medullary thyroid carcinoma: assessment with calcitonin and carcinoembryonic antigen doubling times. *Eur J Endocrinol*. 2008;158:239–246.
116. de Groot JW, et al. Biochemical markers in the follow-up of medullary thyroid cancer. *Thyroid*. 2006;16:1163–1170.
117. Miyauchi A, et al. Spontaneous deceleration and acceleration of growth rate in medullary thyroid carcinomas suggested by changes in calcitonin doubling times over long-term surveillance. *World J Surg*. 2019;43:504–512.
118. Abdelmoumene N, et al. Selective venous sampling catheterisation for localisation of persisting medullary thyroid carcinoma. *Br J Cancer*. 1994;69:1141–1144.
119. Machens A, et al. Prospects of remission in medullary thyroid carcinoma according to basal calcitonin level. *J Clin Endocrinol Metab*. 2005;90:2029–2034.
120. Brierley J, et al. Medullary thyroid cancer: analyses of survival and prognostic factors and the role of radiation therapy in local control. *Thyroid*. 1996;6:305–310.
121. Rowell NP. The role of external beam radiotherapy in the management of medullary carcinoma of the thyroid: a systematic review. *Radiother Oncol*. 2019;136:113–120.
122. Diez JJ, Iglesias P. Somatostatin analogs in the treatment of medullary thyroid carcinoma. *J Endocrinol Invest*. 2002;25:773–778.
123. Kebebew E, Clark OH. Medullary thyroid cancer. *Curr Treat Options Oncol*. 2000;1:359–367.

124. Wirth LJ, et al. Efficacy of selpercatinib in RET-altered thyroid cancers. *N Engl J Med*. 2020;383:825–835.
125. Cohen EE, et al. Axitinib is an active treatment for all histologic subtypes of advanced thyroid cancer: results from a phase II study. *J Clin Oncol*. 2008;26:4708–4713.
126. Capdevila J, et al. Axitinib treatment in advanced RAI-resistant differentiated thyroid cancer (DTC) and refractory medullary thyroid cancer (MTC). *Eur J Endocrinol*. 2017;177:309–317.
127. Schlumberger MJ, et al. Phase II study of safety and efficacy of motesanib in patients with progressive or symptomatic, advanced or metastatic medullary thyroid cancer. *J Clin Oncol*. 2009;27:3794–3801.
128. Robinson BG, et al. Vandetanib (100 mg) in patients with locally advanced or metastatic hereditary medullary thyroid cancer. *J Clin Endocrinol Metab*. 2010;95:2664–2671.
129. Wells SA, et al. Vandetanib for the treatment of patients with locally advanced or metastatic hereditary medullary thyroid cancer. *J Clin Oncol*. 2010;28:767–772.
130. Wells SA, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol*. 2012;30:134–141.
131. Carr LL, et al. Phase II study of daily sunitinib in FDG-PET-positive, iodine-refractory differentiated thyroid cancer and metastatic medullary carcinoma of the thyroid with functional imaging correlation. *Clin Cancer Res*. 2010;16:5260–5268.
132. Ravaud A, et al. A multicenter phase II study of sunitinib in patients with locally advanced or metastatic differentiated, anaplastic or medullary thyroid carcinomas: mature data from the THYSU study. *Eur J Cancer*. 2017;76:110–117.
133. Lam ET, et al. Phase II clinical trial of sorafenib in metastatic medullary thyroid cancer. *J Clin Oncol*. 2010;28:2323–2330.
134. Ahmed M, et al. Analysis of the efficacy and toxicity of sorafenib in thyroid cancer: a phase II study in a UK based population. *Eur J Endocrinol*. 2011;165:315–322.
135. Ito Y, et al. Sorafenib in Japanese patients with locally advanced or metastatic medullary thyroid carcinoma and anaplastic thyroid carcinoma. *Thyroid*. 2017;27:1142–1148.
136. Stellrecht CM, Gandhi V. MET receptor tyrosine kinase as a therapeutic anticancer target. *Cancer Lett*. 2009;280:1–14.
137. Schlumberger M, et al. A phase II trial of the multitargeted tyrosine kinase inhibitor lenvatinib (E7080) in advanced medullary thyroid cancer. *Clin Cancer Res*. 2016;22:44–53.
138. Hong DS, et al. Phase I trial of a combination of the multikinase inhibitor sorafenib and the farnesyltransferase inhibitor tipifarnib in advanced malignancies. *Clin Cancer Res*. 2009;15:7061–7068.
139. Tamburrino A, et al. Activation of the mTOR pathway in primary medullary thyroid carcinoma and lymph node metastases. *Clin Cancer Res*. 2012;18:3532–3540.
140. Grozinsky-Glasberg S, et al. The rapamycin-derivative RAD001 (everolimus) inhibits cell viability and interacts with the Akt-mTOR-p70S6K pathway in human medullary thyroid carcinoma cells. *Mol Cell Endocrinol*. 2010;315:87–94.
141. Schneider TC, et al. Beneficial effects of the mTOR inhibitor everolimus in patients with advanced medullary thyroid carcinoma: subgroup results of a phase II trial. *Int J Endocrinol*. 2015;2015:348124.
142. Elisei R, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol*. 2013;31:3639.
143. Schlumberger M, et al. Overall survival analysis of EXAM, a phase III trial of cabozantinib in patients with radiographically progressive medullary thyroid carcinoma. *Ann Oncol*. 2017;28:2813–2819.
144. Hu M, et al. Clinical activity of selective RET inhibitor, BLU-667, in advanced RET-altered thyroid cancers: updated results from the phase 1 arrow study. *Thyroid*. 2018;28:A170.
145. Taylor MH, Gainor JF, I-Nan Hu M, et al. Activity and tolerability of BLU-667, a highly potent and selective RET inhibitor, in patients with advanced RET-altered thyroid cancers. *J Clin Oncol*. 2019;37:6018.
146. Carlomagno F, et al. Disease associated mutations at valine 804 in the RET receptor tyrosine kinase confer resistance to selective kinase inhibitors. *Oncogene*. 2004;23:6056–6063.
147. Solomon BJ, et al. RET solvent front mutations mediate acquired resistance to selective RET inhibition in RET-driven malignancies. *J Thorac Oncol*. 2020;15:541–549.
148. Sherman SI, et al. Efficacy of cabozantinib in a phase 3 study of medullary thyroid cancer (MTC) patients with RET or RAS mutations. *Eur Thyroid J*. 2013;2:76.
149. Meijer JA, et al. Radioactive iodine in the treatment of medullary thyroid carcinoma: a controlled multicenter study. *Eur J Endocrinol*. 2013;168:779–786.
150. Castellani MR, et al. MIBG for diagnosis and therapy of medullary thyroid carcinoma: is there still a role? *Q J Nucl Med Mol Imaging*. 2008;52:430–440.
151. Salaun PY, et al. Phase II trial of anticarcinoembryonic antigen pretargeted radioimmunotherapy in progressive metastatic medullary thyroid carcinoma: biomarker response and survival improvement. *J Nucl Med*. 2012;53:1185–1192.
152. Iten F, et al. Response to [90Yttrium-DOTA]-TOC treatment is associated with long-term survival benefit in metastasized medullary thyroid cancer: a phase II clinical trial. *Clin Cancer Res*. 2007;13:6696–6702.
153. Parghane RV, et al. Clinical utility of (177) Lu-DOTATATE PRRT in somatostatin receptor-positive metastatic medullary carcinoma of thyroid patients with assessment of efficacy, survival analysis, prognostic variables, and toxicity. *Head Neck*. 2020;42:401–416.
154. Petursson SR. Metastatic medullary thyroid carcinoma. Complete response to combination chemotherapy with dacarbazine and 5-fluorouracil. *Cancer*. 1988;62:1899–1903.
155. Wu LT, et al. Treatment of advanced medullary thyroid carcinoma with a combination of cyclophosphamide, vincristine, and dacarbazine. *Cancer*. 1994;73:432–436.
156. Hughes P, et al. Treatment of hepatic metastases from medullary thyroid cancer with transarterial embolisation. *Eur Radiol Exp*. 2017;1:9.
157. Lorenz K, et al. Selective arterial chemoembolization for hepatic metastases from medullary thyroid carcinoma. *Surgery*. 2005;138:986–993; discussion 993.
158. Walz MK, et al. Posterior retroperitoneoscopic adrenalectomy--results of 560 procedures in 520 patients. *Surgery*. 2006;140:943–948; discussion 948–950.
159. Asari R, et al. Estimated risk of pheochromocytoma recurrence after adrenal-sparing surgery in patients with multiple endocrine neoplasia type 2A. *Arch Surg*. 2006;141:1199–1205; discussion 1205.

135

Neuroendocrine Tumor Syndromes

Jessica E. Maxwell, Joseph S. Dillon, and Daniel M. Halperin

OUTLINE

Introduction, 2274
Epidemiology, 2274
Etiology, 2275
Pathologic Classification and Prognosis, 2275
Diagnostic Approach and Clinical Presentation, 2276
  Laboratory Tests and Biomarkers, 2276
  Imaging, 2276
  Biopsy, 2278
Clinical Features of Functional net and Their Syndromes, 2278
  Insulinoma, 2278
  Gastrinoma, 2279
  VIPoma, 2279
  Glucagonoma, 2279
  Somatostatinoma, 2279
  PPoma, 2279
  Serotoninoma, 2279
  Carcinoid Syndrome, 2279
Management Strategies for Functional Neuroendocrine Tumors, 2283
  Surgical Resection, 2283
  Ablation of Primary Pancreas Neuroendocrine Tumors, 2283
  Hepatic Arterial Embolization, 2284
  Selective Internal Radiation Therapy, 2285
  Systemic Therapies, 2285
  Palliative Symptom Management, 2287
Conclusion, 2287

INTRODUCTION

In 1907, Siegfried Oberndorfer selected the term *karzinoide* (“carcinoma-like”) to refer to the group of neoplasms he discovered within the small intestine that simultaneously shared the histological appearance of adenocarcinoma but yet had no apparent ability to invade or metastasize. This latter observation would prove false, but Oberndorfer’s name for the pathologic entity would be maintained and later used to name the syndrome for which these tumors would become best known.[1]

Neuroendocrine tumors (NETs) have since been found in a variety of sites. They are most common in the small and large intestine, pancreas, lungs, and bronchi, and less common in the thyroid, parathyroid, pituitary gland, adrenal glands, breast, ovary, and thymus. NETs produce and secrete a variety of peptide hormones and bioactive amines. Many of these factors can be measured in blood or urine, and some aid in diagnosis, prognostication, and disease surveillance. When hypersecreted, these factors can produce clinical symptoms and physiologic derangements that adversely affect a patient’s quality of life or even prove to be life-threatening. Tumors manifesting with such symptoms are termed “functional.” Nonfunctional tumors may secrete factors that do not result in a clinical syndrome.[2] The hormonal syndromes produced by these tumors are distinct but rare and typically require multimodal therapy for palliation or cure.

This chapter will briefly cover the epidemiology and pathogenesis of well-differentiated NETs of the gastrointestinal and aerodigestive tracts but focuses more specifically on the diagnosis and management of the clinical syndromes produced by these unique tumors. Ectopic hormone syndromes and other NETs (e.g., pheochromocytoma) are covered elsewhere in this text and will not be discussed here.

EPIDEMIOLOGY

Neuroendocrine neoplasms (NENs) are rare but increasing in incidence, affecting 6.98 per 100,000 individuals according to the Surveillance Epidemiology End Results (SEER) database. The rise is seen across all NET primary sites and disease stages. Lung NETs have the highest incidence, followed by ileal/jejunal, rectal, and pancreas. The rising incidence is likely multifactorial and related not only to a true increase in incidence and growing recognition of the disease, but also greater use of cross-sectional imaging and endoscopy that may detect these tumors during the course of surveillance and diagnosis of other conditions. Patients are typically 65 years old or greater at diagnosis, with women only slightly more commonly affected (52.7%).[3]

The heterogeneity of primary tumor sites, functional status, and propensity for metastasis requires careful communication with subtly different terminology. NEN is a generic term used to encompass all NETs, regardless of grade, differentiation, or metastatic potential. A group of well-differentiated NENs with metastatic potential are termed NETs. The term “neuroendocrine carcinoma” (NEC) specifically refers to the group of poorly differentiated NETs with Ki-67 greater than 20%.

Despite their infamy, NET syndromes are only seen in a fraction of patients. The most common is carcinoid syndrome (most frequently found in patients with jejunal/ileal NETs), followed by those caused by hormones hypersecreted by pancreas/duodenum NETs. The proportion of patients affected by these clinical presentations is difficult to ascertain. It ranges widely in the literature, as most series are single-institution and retrospective, but is often estimated at 25% to 30% in gastroenteropancreatic (GEP) NETs[4] and less than 10% of lung NETs.[5] A 2017 study using the SEER-Medicare database estimated that 19% of NEN patients suffered from carcinoid syndrome

TABLE 135.1 **Pathologic Classification of Gastroenteropancreatic Neuroendocrine Tumors**

Terminology	Differentiation	Grade	Mitotic Rate (mitoses/mm^2)	Ki-67 Index (%)
NET G1	Well-differentiated	Low	<2	<3
NET G2	Well-differentiated	Intermediate	2–20	3–20
NET G3	Well-differentiated	High	>20	>20
NEC G3	Poorly differentiated	High	>20	>20

at diagnosis. These patients were more likely to have regional or distant metastases, grade 1 or 2 tumors, and primary sites either in the jejunum/ileum or cecum.[6] In pancreatic NENs, an analysis of SEER data revealed that the majority were not classified as functional (90.8%). Of the functional neoplasms, gastrinomas are most common, followed by insulinomas, glucagonomas, and vasoactive intestinal peptide (VIP) tumors.[7]

KEY POINTS

- The terminology of neuroendocrine neoplasms (NENs) varies by metastatic potential, grade, and differentiation. Neuroendocrine tumors (NETs) are well-differentiated NENs with metastatic potential, and a Ki-67 ≤20%. The term “neuroendocrine carcinoma” should only be applied to poorly differentiated NETs with Ki-67 >20%. NENs are further defined by their function, or lack thereof.

ETIOLOGY

NETs arise from secretory cells of the diffuse neuroendocrine system. Gastrointestinal NETs derive from enterocytes, goblet cells, Paneth cells, or endocrine omnipotent stem cells at the base of the crypts of Lieberkühn, with differences in transcription and signaling driving ultimate neoplastic differentiation.[8] Pancreas NETs (PNETs) originate from the islets of Langerhans, as well as from embryonal cells within the duct/acinar system.[2,9]

The majority of NETs are sporadic and not related to an inherited genetic factor. Neither GEP nor lung NETs are associated with significant behavioral or environmental factors (i.e., smoking). Genetic predisposition accounts for less than 10% of GEP NETs, with multiple endocrine neoplasia type I (MEN1) the most common of the inherited cancer syndromes.[10] MEN1 is inherited in an autosomal dominant fashion. Patients develop multigland hyperparathyroidism (95%–100%), PNETs (80%–100%), and pituitary adenomas (54%–65%). Most MEN1-associated PNETs are nonfunctional (70%). Gastrinomas and insulinomas are the most common functional tumors in hereditary syndromes, as in sporadic disease. Less common familial NET syndromes include neurofibromatosis-1, tuberous sclerosis, Cowden syndrome, von Hippel Lindau,[4] and inositol polyphosphate multikinase.[11]

PATHOLOGIC CLASSIFICATION AND PROGNOSIS

Gastrointestinal NENs are classified as either well-differentiated NETs or poorly differentiated NECs, depending on the extent to which the tumor cells resemble their nonneoplastic counterparts. Well-differentiated tumors have uniform, “organoid” cellular arrangements, whereas poorly differentiated tumors have a more diffuse sheet-like architecture, with irregular nuclei and scant cytoplasmic granularity.[12]

Gastrointestinal NETs are also described by their grade, which is determined by the cellular proliferation rate (Table 135.1). It is most common to define a GEP NET’s grade via Ki-67, an immunohistochemical (IHC) method that reports the percentage of neoplastic cells that stain for the proliferation antigen, although calculation of the mitotic rate (mitoses per 2 mm^2)[13] is also acceptable. Low-grade (G1) NETs are defined by a Ki-67 index of less than 3%, intermediate-grade (G2) 3% to 20%, and high-grade (G3) greater than 20%.[14] Ki-67 IHC is highly reproducible, and concordance between biopsy and surgical specimens is greater than 90%.[15]

Clinical behavior can be predicted by a tumor’s cellular differentiation and grade. Well-differentiated, low-grade (G1 and G2) tumors behave indolently and tend to be slower to metastasize. Low- (G1) and intermediate- (G2) grade GEP NETs are associated with a median overall survival (OS) of 16 and 8 years, respectively. Examined by primary site (but not controlled for grade or stage), appendiceal NETs have the most favorable OS, estimated at greater than 30 years, followed by rectal NET, with a median OS of 24.6 years. PNETs have the shortest median OS of GEP NETs, at only 3.6 years. Stage is also predictive of prognosis, but also depends on the primary site. Localized small intestine NETs have a median OS of 14 years. This is reduced to 8.6 years in advanced cases. Pancreas NETs diagnosed when still localized have a median OS of approximately 20 years. Survival decreases to approximately 5 years in advanced disease.[3]

Historically, shorter OS was reported for patients with functional tumors.[16-18] This observation is likely related to two factors—the disproportionate effect carcinoid heart disease had on NET-related mortality, and methodological errors in the retrospective studies performed. The pathophysiology of carcinoid heart disease is incompletely understood but is thought to relate to chronically elevated levels of circulating serotonin causing fibrous endocardial thickening of heart valves (typically right-sided), heart failure, and death.[19] In the 1990s, carcinoid heart disease accounted for approximately 50% of NET-related deaths, but aggressive surgical management has improved disease-related survival, and these patients now have an OS that approximates that of nonsyndromic patients.[20] The second factor relating to this observation was that many studies failed to control for hepatic tumor burden when determining factors associated with shorter OS in NET patients. Carcinoid syndrome is associated with high metastatic tumor burdens, and a significant portion of these patients will ultimately die of liver failure. Correction of this factor eliminated the survival disadvantage seen in patients with elevated levels of serotonin and urinary 5-hydroxyindoleacetic acid (u5-HIAA).[21,22]

Poorly differentiated, high-grade NEC have a poor prognosis and progress rapidly, and are therefore managed differently than well-differentiated, lower-grade NETs.[23] In this group, median OS is a mere 10 months.[3] Sorbye et al. analyzed a large group of NECs and found that survival outcomes and response to platinum-based chemotherapy were associated with Ki-67 index. NECs with Ki-67 less than 55% had low response rates to platinum-based chemotherapy, but significantly longer OS (14 months vs. 10 months with Ki-67 index >55%).[24] Bridging these two categories of GEP NENs are the newly described well-differentiated G3 GEP NETs, which seem to have a better prognosis than poorly differentiated G3 NEC. In one series of 31 GEP NETs, median disease-specific survival was 55 months for well-differentiated G3 NET compared with 11 months for G3 NEC.[25]

Lung NETs have retained the term "carcinoid" and are categorized as typical or atypical carcinoids, based on their grade and histologic appearance. Typical lung carcinoids have fewer than 2 mitoses/mm^2 and lack necrosis, whereas atypical lung carcinoids have two or more mitoses/mm^2 or have focal necrosis. The use of Ki-67 to differentiate these entities is controversial, and not currently recommended.[26] Typical lung carcinoids behave similarly to G1 GEP NETs, whereas atypical lung carcinoids more closely approximate G2 GEP NETs. Small and large cell lung cancers, classified morphologically, are analogous to poorly differentiated G3 NECs. Typical lung carcinoids are associated with a median OS of approximately 20 years, whereas patients with high-grade lung carcinomas demonstrate a median OS of 1 to 2 years.[3] Lung NETs do not have an equivalent well-differentiated, high-grade entity.

KEY POINTS

- All neuroendocrine neoplasms should be described by the location of the primary tumor, cellular differentiation and grade, and function, as each of these factors contributes to management decisions and discussions of prognosis. Functional tumors are not necessarily associated with poor prognosis. Rather, it is the underlying tumor burden and physiologic sequelae of hypersecreted hormones and bioactive amines that affect outcome.

DIAGNOSTIC APPROACH AND CLINICAL PRESENTATION

Whether functional or nonfunctional, a high index of suspicion is required to correctly diagnose a NET. The symptoms produced by hypersecreting NETs may not manifest at once, can vary in severity, and also mimic symptoms of common diseases like inflammatory bowel disease, diabetes, or dermatitis. Patients with nonfunctional tumors may suffer from vague abdominal pain or inflammatory bowel disease–like symptoms, allowing the tumor to go undetected for many years. These difficulties support the recommendation that patients diagnosed with NET are best served at high-volume NET centers with specialized multidisciplinary teams familiar with the nuances of diagnosing, staging, and treating these unique tumors.

Laboratory Tests and Biomarkers

Nonspecific biochemical testing is most useful to monitor response to therapy or detect disease progression and recurrence. Current biomarkers lack the sensitivity and specificity to be used as screening tools and are more appropriately obtained to confirm or rule out a diagnosis when pretest probability is high.[27] Historically, guidelines from both the North American and European Neuroendocrine Tumor Societies have recommended measuring chromogranin A (CgA) and 24-hour u5-HIAA for extrapancreatic NET,[28,29] although more recent guidelines have deemphasized their use and utility.[30] In PNETs, institutional practices vary. Some initially order a broad spectrum of tests and then narrow their scope once the secretory behavior of the tumor is ascertained, whereas others order only those labs that will confirm the diagnosis of a specific functional syndrome (e.g., gastrin in a patient with symptoms of Zollinger–Ellison syndrome). Most data used to guide practice are retrospective and can be difficult to interpret or apply broadly. The lack of consensus emphasizes the importance of specialized, multidisciplinary care delivered by practitioners well-versed in the limitations of each biomarker, as well as their utility when applied consistently.

CgA may be a useful aid to assess the completeness of surgical cytoreduction, or to indicate disease recurrence and progression, as its levels can correlate with tumor burden. It is widely distributed in neuroendocrine tissues, present in normal islet cells, and also cosecreted with serotonin. Its sensitivity and specificity range between 67% and 93% and 85% and 96%, respectively. Caution must be used when interpreting results, as the levels can be falsely elevated by proton pump inhibitors, or in patients with atrophic gastritis, renal or hepatic failure, or cardiac dysfunction.[31]

5-HIAA is a metabolite of serotonin with high sensitivity (>95%) but low specificity for GEP NET. It is measured in a 24-hour urine collection or via plasma assays. Levels do not correlate with either tumor burden or symptom severity,[20] so its utility is limited to diagnosis, rather than surveillance. As with CgA, its levels can be falsely elevated by a number of substances—medications (acetaminophen, monoamine oxidase inhibitors, 5-fluorouracil), caffeine, wine, or tryptophan-rich foods (mockernuts, cheese, pineapple, tomato).[31]

Patients with a suspected functional NET in the duodenum or pancreas may require a more intensive laboratory workup to confirm a diagnosis, with selection of tests guided by symptoms. In most cases, demonstration of elevated circulating levels of the hormone responsible for the symptoms is sufficient. When these tests are equivocal, diagnosis of insulinomas or gastrinomas can be pursued with provocative testing. Patients with a suspected insulinoma may undergo a monitored 48-hour fast. Plasma glucose, insulin, and proinsulin levels are assessed at admission, and then serially every 4 to 6 hours. Patients are monitored for symptoms of hypoglycemia, which, in combination with low blood glucose and high proinsulin and insulin levels, clinch the diagnosis.[32] When a fasting gastrin level is equivocal for diagnosis of a gastrinoma, a secretin stimulation test may be helpful. Patients are administered intravenous secretin, and gastrin levels are measured approximately every 5 minutes for 30 minutes. An increase of serum gastrin levels greater than 200 ng/L confirms a gastrinoma.[33] In patients treated with proton pump inhibitors who are unable to stop therapy during their diagnostic workup, integration of imaging, biopsy, and IHC are useful.[34] Suggested laboratory tests for various NET tumor types are outlined in Table 135.2.

Imaging

After a thorough history, imaging is essential for NET diagnosis. Standard modalities such as endoscopy, computed tomography (CT) and magnetic resonance imaging (MRI) can localize the primary tumor, define anatomic relationships crucial for surgical planning, and monitor for progression or recurrence. Somatostatin-targeted imaging is a modality unique to NETs that can localize tumors in multiple tissue types and identify individuals who may benefit from somatostatin receptor (SSTR)-targeted therapy. Frequently, multiple modalities are used throughout the care cycle.

Endoscopy. Endoscopy is helpful in a variety of settings for GEP NET of the foregut and hindgut. Esophagogastroduodenoscopy (EGD) can detect primary tumors, treat small tumors using snare or

TABLE 135.2 Suggested Baseline Laboratory Tests for Neuroendocrine Tumor Types

Tumor Type	
Insulinoma	Fasting: Proinsulin Insulin C-peptide Blood glucose
Gastrinoma	Gastrin (fasting) Gastric pH
Glucagonoma	Glucagon (fasting)
Somatostatinoma	Somatostatin
VIPoma	VIP
Gastric NET	Gastrin (fasting) Chromogranin A or u5-HIAA
Jejunoileal, colorectal NET	Chromogranin A or u5-HIAA
Nonfunctional PNET	No NET-specific biomarker recommended
Lung NET	No NET-specific biomarker recommended

The suggestions follow recommendations set forth by the North American Neuroendocrine Tumor Society[30,114] and the European Neuroendocrine Tumor Society. In small intestine and nonfunctional pancreas NETs, obtaining pancreastatin and neuron-specific enolase levels can be considered. Institutional practices vary.

VIP, Vasoactive intestinal peptide; *u5-HIAA,* urinary 5-hydroxyindoleacetic acid.

electrocautery, and provide tissue for diagnosis. When combined with endoscopic ultrasound (EUS), it may provide information about T and N stage, and allows for ultrasound-guided fine-needle aspiration of PNETs or regional lymph nodes. In PNETs, EUS localizes tumors with 93% sensitivity and 95% specificity. Duodenal NETs are often more difficult to detect, as they are small and submucosal. EGD detects these tumors in only 40% to 60% of cases.[35] Double balloon enteroscopy or capsule endoscopy may be employed to detect small bowel primaries, but this is not part of the standard workup, as detection rates are less than 50%.[36,37] Colonoscopy is used analogously to EGD for NETs of the colon, rectum, and terminal ileum. Bronchoscopy, with or without ultrasound, can be used for detection, biopsy, local staging, and surveillance of lung and bronchial NETs.

Computed Tomography. Computed tomography is the most frequently used imaging modality in the workup and surveillance of NETs. In GEP NETs, CT demonstrates a mean sensitivity of 73% and 96% specificity for localizing the primary tumor, with an overall detection rate of 73%. It performs better for hepatic metastases, with a mean sensitivity of 82% and specificity of 92%. Efficacy in soft tissue metastases is similar to that of hepatic metastases. CT is useful for detecting typical lung carcinoids, although the positive predictive value of CT in detecting pathological regional lymphadenopathy is a disappointing 20% to 45%.[5] CT enteroclysis, which involves distending the small bowel with 2 L of tap water before imaging, can be used to detect small intestine NETs and increases the detection rate of these often small tumors, with a mean sensitivity and specificity of 85% and 97%, respectively.[38]

Regardless of primary tumor site, CT should be obtained with intravenous contrast (and oral contrast in cases of GEP NETs). NETs are typically hypervascular and detected in the early arterial phase of a triple phase (early arterial, late arterial, portal venous phases) scan.

Many institutions have dedicated CT protocols for various NET primary tumor sites (e.g., pancreas or appendix NET protocol), which differ from generic CT protocols by the types and timing of contrast administered.

Magnetic Resonance Imaging. MRI is less commonly used to detect sporadic primary tumors, but rather used for detection and surveillance of hepatic or, rarely, brain metastases.[5,38] In this latter context, the mean detection rate ranges from 82% to 95%. Occasionally, PNETs that escape detection by CT are detected by MRI. In these cases, MRI demonstrates an average sensitivity of 93% and specificity of 88%.[38] This modality is also useful in patients who cannot undergo disease surveillance with CT, such as those with renal failure or an allergy to iodinated contrast. As with CT, optimal imaging requires intravenous contrast, in this case gadolinium. NET are hyperintense on arterial phases. Magnetic resonance cholangiopancreatography can be used to define biliary ductal anatomy, especially in patients who cannot be imaged with endoscopy.

Somatostatin Receptor Imaging. Somatostatin is an inhibitory hormone produced by the delta cells of the gastric antrum, duodenum, and pancreas. The hormone binds to SSTR, a receptor with five different subtypes in humans (SSTR1–5).[39,40] NETs may express a combination of these receptors, commonly with a greater proportion of SSTR2, and it is these receptors that serve as the basis for SSTR imaging and treatment. SSTR expression declines with advancing tumor grade and loss of differentiation, and SSTR-based imaging is generally helpful in grade 1 and 2 (and some well-differentiated grade 3) NETs, but less frequently in NECs.[41] The newest iteration of SSTR imaging is a type of positron emission tomography (PET). The most common modalities use a ^{68}Ga-labelled radioligand conjugated to a somatostatin analog (SSA) (^{68}Ga-DOTATOC, ^{68}Ga-DOTANOC, ^{68}Ga-DOTATATE).[42] This novel PET scan is fused with either CT or MRI to aid anatomic localization. It has enhanced sensitivity (97%) compared with its predecessor, the OctreoScan (sensitivity of 85%), which is a scintographic imaging modality that uses ^{111}In-labeled diethylenetriamine penta-acetic acid octreotide to detect tumors expressing SSTR2 and SSTR5.[43]

DOTA PET is most appropriately used for initial staging after the histologic diagnosis of a NET, for localization of a primary tumor in patients with metastatic disease but unknown primary, and to select patients for SSTR-targeted peptide receptor radionuclide therapy (PRRT). It is not an absolute requirement in the workup and staging, but an adjunct when standard imaging cannot identify suspected primary or metastatic lesions, and for determination of SSTR status.[42] These studies are useful for both functioning and nonfunctioning NET, although insulinomas may demonstrate low expression of SSTR2 and thus localize poorly.[44,45] Other subtypes of functional NETs typically express high levels of SSTR, and thus detection may only be limited by their size, as tumors less than 1 cm are often below the limit of resolution.

(^{18}F)-Fluorodeoxyglucose Positron Emission Tomography. (^{18}F)-fluorodeoxyglucose (FDG) is a synthetic analog of glucose that is taken up by metabolically active cells, and thus FDG-PET is used to detect distant metastases in NECs. It has poor sensitivity for well-differentiated and low-grade NETs.[41] Metabolic intensity should not be confused with tumor function, as low- or intermediate-grade hypersecreting NETs are no more likely to be detected with FDG-PET than their nonfunctional counterparts. The extent to which cells take up the radioactive tracer is indicated by the maximum standard uptake value (SUV_{max}). Small, retrospective series have correlated an SUV

TABLE 135.3 **Syndromic Presentations of Functioning Neuroendocrine Tumors**

Syndrome	Primary Tumor Location	Hormones Secreted	Signs and Symptoms
Carcinoid syndrome	Small intestine, less commonly lung or pancreas	Serotonin Variety of bioactive amines (bradykinin, tachykinins, kallikrein)	Facial flushing Diarrhea Bronchospasm/wheezing Carcinoid heart disease
Insulinoma	Pancreas	Insulin	Hypoglycemia Autonomic or neuroglycopenic symptoms: shakiness, confusion, sweating
Gastrinoma	Pancreas Duodenum	Gastrin	Gastric reflux Peptic ulceration Gastrointestinal bleeding
VIPoma/watery diarrhea hypokalemia achlorhydria (WDHA) syndrome	Pancreas	Vasoactive intestinal peptide	Watery diarrhea Hypokalemia Achlorhydria
Glucagonoma	Pancreas	Glucagon	Glucose intolerance Necrolytic migratory erythema Stomatitis Glossitis Hypoaminoacidemia Thromboembolism
Somatostatinoma	Pancreas	Somatostatin	Hyperglycemia Cholestasis Steatorrhea Achlorhydria

The primary tumor location refers to the most common locations of these tumors. There are a number of case reports of ectopic tumors that may hypersecrete hormones and cause a clinical syndrome.

ratio (tumor SUV_{max}/normal SUV_{max}) of 2.5 or more with poor survival and a 4-year survival rate of 0%.[46]

Biopsy

Histopathologic evaluation is essential for a formal diagnosis of NET, although there are circumstances under which biopsy can be avoided. NETs commonly diagnosed with EGD or colonoscopy (gastric, duodenal, or colorectal) are discovered and biopsied in the patient's initial encounter. As well, appendiceal NETs are often diagnosed during the course of an appendectomy for appendicitis. Decisions to biopsy other NET subtypes depend on primary site and the likely course of management. Often the combination of high-quality imaging and clinical history strongly suggests a diagnosis of NET. If the disease is amenable to surgical resection for cure, preoperative biopsy is unnecessary. In cases where the disease has metastasized, but may still be amenable to surgical cytoreduction, biopsy of an easily accessible metastasis may help the team decide whether surgical resection or systemic therapy is most appropriate, based on the histologic characteristics of the tumor and the Ki-67 index. Patients with nonfunctional PNETs that are small (<1 cm) and are observed do not require histologic confirmation of the disease unless imaging characteristics and clinical history are discordant. Similarly, MEN1 patients with small, multifocal PNETs can be observed without undergoing an invasive biopsy. Biopsy is most helpful in patients who have an unknown primary tumor, or who have disease that is not amenable to resection, in which case biopsy is the only opportunity for pathologic diagnosis to guide further treatment. In patients where the location of the primary tumor is unknown, biopsy samples may be tested with a panel of IHC stains to identify the tissue of origin.[47]

KEY POINTS

- A thorough history and physical examination will aid in differentiating a functional neuroendocrine tumor (NET) from a more common condition, such as diabetes mellitus or inflammatory bowel disease. Laboratory studies are useful to confirm hypersecretion of peptide hormones, with selection of labs driven by a patient's symptomatology. Contrast-enhanced computed tomography or magnetic resonance imaging are required for initial workup and staging. Somatostatin receptor imaging is an adjunct to standard imaging when the extent of metastatic disease burden is unclear, if the diagnosis of NET is suspected but the primary tumor site is unknown, or to determine candidacy for peptide receptor radionuclide therapy. Biopsy can provide useful information such as cellular grade and differentiation, but is unnecessary if a patient is destined for surgical resection.

CLINICAL FEATURES OF FUNCTIONAL NET AND THEIR SYNDROMES

Insulinoma

In 1935, Dr. A. O. Whipple and Dr. V. K. Franz detailed six cases of symptomatic sporadic hyperinsulinism apparently caused by a pancreas adenoma and laid out an algorithm for their diagnosis and treatment. Although 85 years have passed, many of their original management guidelines are used today.[48]

Insulinomas are typically localized (90%) and located in the body or tail of the pancreas.[49] They are the most common type of functional PNET, and approximately 5% are found in patients with MEN1. The clinical syndrome caused by these tumors is characterized with the “Whipple triad” of (1) hypoglycemia, (2) autonomic or neuroglycopenic symptoms of shakiness, confusion, or sweating, and (3) resolution of these symptoms with eating (Table 135.3). As mentioned above, a monitored fast may be required to provoke the syndrome. Insulinomas are often less than 2 cm in diameter and escape detection by imaging, making surgical planning difficult. In these situations, arterial calcium stimulation with hepatic venous sampling is an invasive procedure that has been shown to localize radiographically invisible tumors with 93% sensitivity and 86% specificity.[50] Patients with poorly controlled symptoms but occult primaries can be taken to the operating room for exploration, as some tumors may be identified with palpation or with intraoperative ultrasound. This latter procedure has a sensitivity of 89% in this context.[45]

Metastatic insulinomas are more likely intermediate-grade and have elevated CgA in more than 50% of cases.[51,52] These tumors, like most NET subtypes, most commonly metastasize to the liver. Only 20% of malignant insulinomas present with only locoregional metastases. Reports of patients with malignant disease suggest that Whipple's triad is present and ultimately helpful for diagnosis in the majority of cases. Interestingly, even in floridly symptomatic patients, the diagnosis can be delayed for many months after the clinical syndrome manifests.[52]

Gastrinoma

Dr. Zollinger and Dr. Ellison described their eponymous syndrome, and its unfortunate consequences (Table 135.3), in a 1955 report of two patients who underwent repeated gastric resections to treat recurrent peptic ulcers. They suspected an "ulcerogenic humoral factor of pancreatic islet origin," as both patients were ultimately found to have pancreas tumors—one uncovered at autopsy, and the other at the time of their total gastrectomy.[53] These tumors are found in the pancreas, duodenum, and stomach, with their most common locations defined by the gastrinoma, or Passaro's, triangle (Fig. 135.1).[54] These tumors are often very small and submucosal, making their detection by conventional imaging or EGD difficult. Approximately 50% to 60% of patients present with metastases at diagnosis.[55]

VIPoma

The clinical syndrome caused by vasoactive intestinal peptide–secreting PNETs described by Dr. Verner and Dr. Morrison in 1958 features progressive, large-volume watery diarrhea that results in profound hypokalemia, achlorhydria, and dehydration.[56] It is colloquially referred to as "pancreatic cholera" or "watery diarrhea hypokalemia achlorhydria syndrome," and can also include hypercalcemia, facial flushing, or a skin rash. Eight percent of patients report intense facial flushing during episodes of diarrhea (Table 135.3).[57] When refractory to treatment, patients can require at-home intravenous fluids or inpatient admission for resuscitation. They may ultimately succumb to multiorgan failure and death from the electrolyte derangements. These rare tumors are solitary, located in the body and tail of the pancreas, and typically larger than 3 cm at diagnosis. The majority are malignant, and 60% to 80% have metastasized on diagnosis. Approximately 5% are found in patients with MEN1.[58]

Glucagonoma

Glucagonomas typically arise from the alpha cells of the islets of Langerhans in the tail of the pancreas. Glucagonoma syndrome is associated with diabetes, necrolytic migratory erythema (NME), diarrhea, neuropsychiatric disturbances, anemia, venous thromboembolism, and weight loss. Most symptoms are a result of the elevated glucagon levels, although the cutaneous manifestations likely result from poor nutrition and amino acid deficiency secondary to glucagon- and tumor-induced catabolism. The rash typically involves the face, abdomen, perineum, or extremities. It appears reddish-brown and can form erythematous bullae that eventually open and encrust. Depending on the case series, approximately 50% to 80% of patients will present with NME, and 25% to 60% with diabetes (Table 135.3). Diagnosis is often delayed, and more than 50% of patients present with distant metastases.[59]

Somatostatinoma

Somatostatinomas are exceedingly rare and associated with a clinical syndrome of diabetes, cholestasis, and steatorrhea as a result of somatostatin's inhibitory effect on insulin, cholecystokinin, and pancreatic exocrine secretion (Table 135.3). More than 75% of somatostatinomas present with distant metastases. The majority of pancreatic somatostatinomas are sporadic. However, some of these tumors arise from the duodenum and, in this context, are associated with neurofibromatosis in up to 50% of cases and are less likely to have metastasized at presentation.[60]

PPoma

Classification of a pancreatic polypeptide (PP)-secreting PNET as functional is controversial, as these tumors are not associated with a distinct clinical syndrome. PP is secreted as part of the parasympathetic response to meals, and high levels are associated with increased satiety, inhibition of gallbladder contraction, and delayed gastric emptying. There are reports of PP hypersecretion and new-onset glucose intolerance, although the rarity of the tumor, retrospective nature of the case series, and prevalence of metabolic syndrome in the general population limit the conclusions that can be drawn.[61,62]

Serotoninoma

A small proportion of PNETs secrete serotonin, cause carcinoid syndrome, and are referred to as "pancreatic carcinoids." The incidence is estimated at between 1% and 4% of all PNETs, but given the low number of patients with PNETs tested for serotonin or u5-HIAA, this may be an underestimation. In a series of 255 sporadic PNET cases, 22 (8.6%) had u5-HIAA levels three times the normal limit, but only two (0.8%) of these patients had symptoms compatible with carcinoid syndrome. It is not uncommon for these patients to cosecrete another hormone, most often insulin. Of the 22 patients with elevated u5-HIAA, two patients had insulinomas, and one a gastrinoma. The majority of patients (81.8%) presented with elevated levels of serotonin or u5-HIAA at presentation, although four went on to develop elevated u5-HIAA later in their disease. High levels of u5-HIAA correlated with high levels of CgA. Patients with serotonin-secreting PNETs often presented with liver metastases (95%), and their median survival was significantly shorter (42 months) than those patients with normal u5-HIAA (116 months).[22]

Carcinoid Syndrome

The pathophysiology of carcinoid syndrome is incompletely understood, but the most common symptoms of diarrhea, facial flushing, and bronchospasm are associated with elevated circulating levels of serotonin and other bioactive amines such as histamine, prostaglandins, tachykinins, and kallikrein (Table 135.3). Patients can also have elevated levels of u5-HIAA, although a proportion of patients manifest the syndrome with normal levels of these factors, while patients with elevated levels may be completely asymptomatic. The syndrome is frequently found in patients with hepatic metastases. When these metastases secrete high concentrations of bioactive substances, they enter directly into the systemic circulation, avoiding breakdown in the liver.

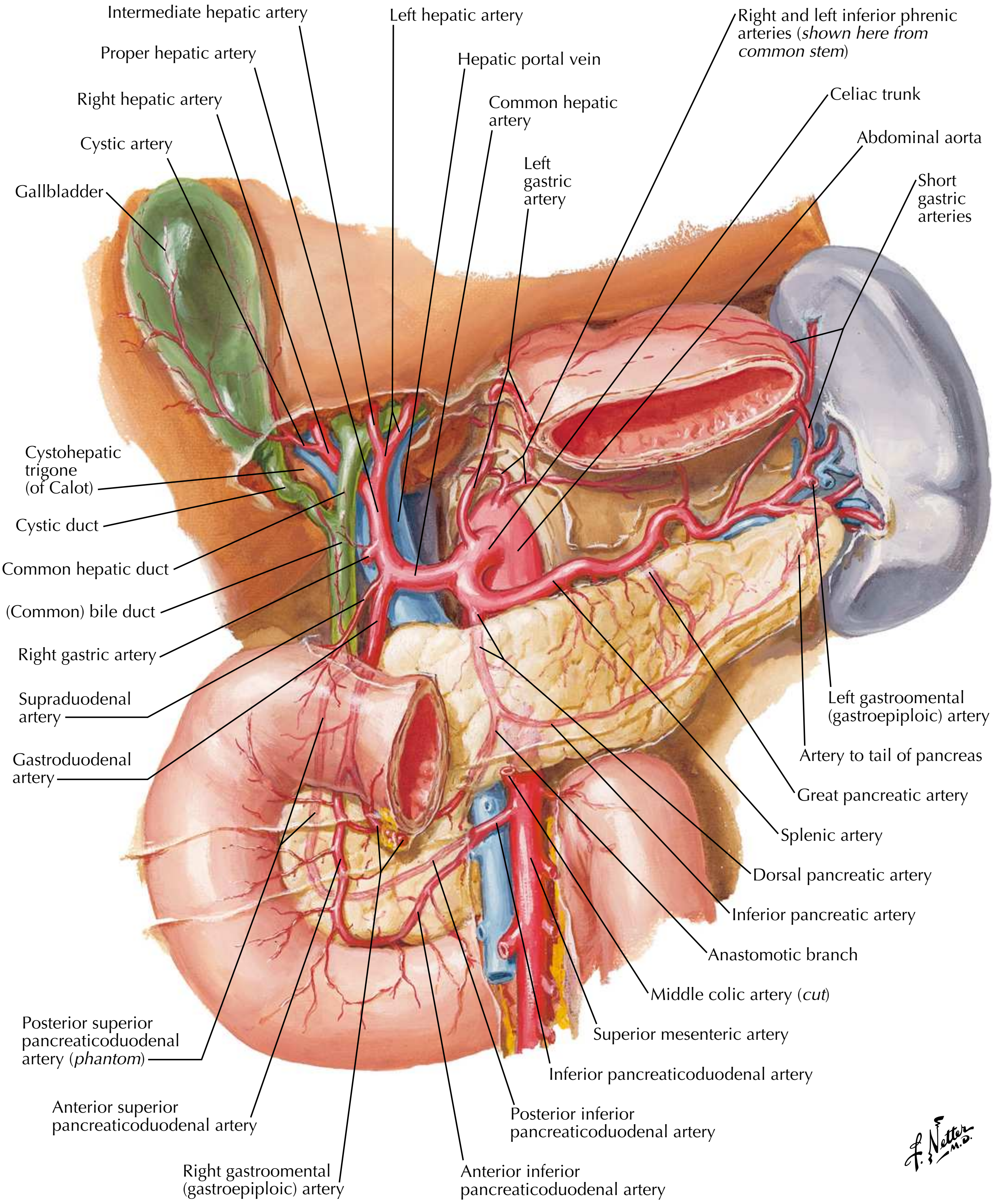

Intermediate hepatic artery
Left hepatic artery
Right and left inferior phrenic arteries (*shown here from common stem*)
Proper hepatic artery
Hepatic portal vein
Celiac trunk
Right hepatic artery
Common hepatic artery
Abdominal aorta
Cystic artery
Left gastric artery
Short gastric arteries
Gallbladder
Cystohepatic trigone (of Calot)
Cystic duct
Common hepatic duct
(Common) bile duct
Right gastric artery
Supraduodenal artery
Gastroduodenal artery
Left gastroomental (gastroepiploic) artery
Artery to tail of pancreas
Great pancreatic artery
Splenic artery
Dorsal pancreatic artery
Inferior pancreatic artery
Anastomotic branch
Middle colic artery (*cut*)
Superior mesenteric artery
Inferior pancreaticoduodenal artery
Posterior superior pancreaticoduodenal artery (*phantom*)
Anterior superior pancreaticoduodenal artery
Posterior inferior pancreaticoduodenal artery
Right gastroomental (gastroepiploic) artery
Anterior inferior pancreaticoduodenal artery

Fig. 135.1 The gastrinoma triangle. The anatomic region in which most gastrinomas are found is indicated by the circle. The vertices of passaro's triangle are marked by stars and located at the junction of the cystic and common bile ducts, the junction of the neck and body of the pancreas, and the junction of the second and third portion of the duodenum. The triangle is best visualized intraoperatively, after the second portion of the duodenum and head of the pancreas have been reflected medially (a kocher maneuver). Kocherization of the duodenum and pancreas is required to thoroughly palpate the duodenum during exploration for the primary tumor(S). (From Netter F: Atlas of Human Anatomy, 7e. ISBN: 9780323393225. Plate 291. "Arteries of Liver, Pancreas, Duodenum, and Spleen.)

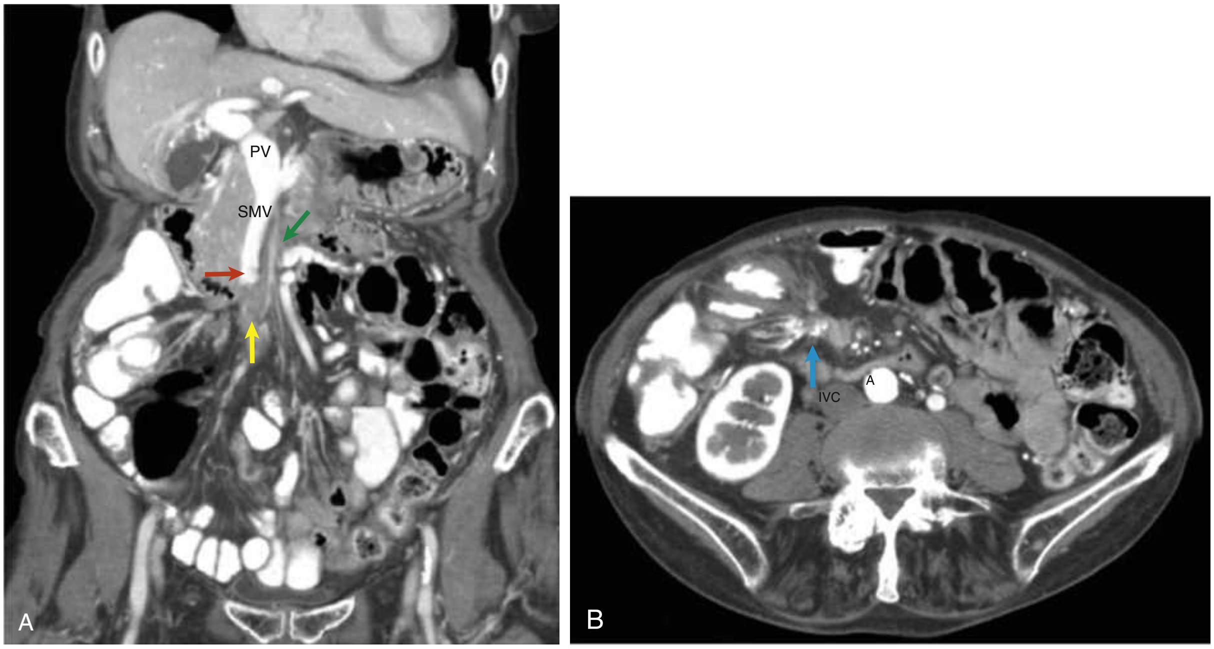


Fig. 135.2 **Consequences of serotonin hypersecretion, mesenteric fibrosis.** An 81-year-old woman with a grade ii ileal neuroendocrine tumor, metastatic to the liver and bone. She presented with surgically unresectable mesenteric fibrosis that encased her superior mesenteric vein (*SMV*) and superior mesenteric artery (SMA). Fibrosis will appear grey and hazy on arterial and portal venous phases of computed tomography scans. (**A**) Coronal view, portal venous phase. The red arrow shows a tumor thrombus within the smv. The green arrow shows the SMA encased by fibrotic tissue. The fibrosis narrowed the proximal mesenteric vessels and obliterated sma branches to a portion of the terminal ileum *(yellow arrow).* (**B**) Axial view, portal venous phase. The blue arrow demonstrates the same anatomic region as the yellow arrow in the coronal image. The desmoplastic reaction causes tethering of the mesentery, which can lead to kinking of the bowel and obstruction in severe cases. As the fibrosis progresses, inflow and venous drainage are impaired, and the bowel wall will thicken and can progress to frank necrosis and perforation. *PV*, portal vein; *IVC,* inferior vena cava; *A*, aorta.

TABLE 135.4 **Triggers of Facial Flushing or Carcinoid Syndrome**

Medications	Foods	Other Factors
Bromocriptine	Spicy food	Physical activity
Tamoxifen	Glutamate	Stress
Nicotinic acid	Monosodium glutamate	
Opiates	Meat	
Calcium channel blockers	Eggs	
Ketoconazole	Cheese	
Chlorpromazine	Sodium nitrates	
Cephalosporin	Smoked meats	
Epinephrine	Beef jerky	
	Sulfites	
	Wine	
	Pickled foods	
	Hot beverages	
	Tyramines	
	Hard/aged cheeses	
	Cured meats	
	Nuts	
	Tryptophan	
	Dark chocolate	

Chronically elevated levels of these factors are thought to contribute to fibrotic complications such as valvular heart disease (predominantly right-sided) and mesenteric fibrosis (Fig. 135.2).[63] Carcinoid syndrome is most commonly associated with small bowel NET, and as an homage to Dr. Oberndorfer's original discovery, small bowel NETs are often referred to as carcinoid tumors, regardless of whether they are associated with carcinoid syndrome. Carcinoid syndrome is also associated with colorectal, lung, and pancreas NETs, albeit at much lower frequencies.

Flushing is reported in 90% of patients with carcinoid syndrome. It is episodic and can be triggered by amine-rich foods such as chocolate, avocadoes, and nuts; alcohol; and physical activity or stress (Table 135.4). Carcinoid-associated diarrhea is also intermittent, secretory, and usually accompanied by abdominal cramping, as serotonin functions to increase intestinal motility. Bronchospasm is found in approximately 15% of patients with carcinoid syndrome, and patients also often suffer from concurrent flushing, sneezing, and dyspnea. Pellagra is found in a minority of carcinoid syndrome patients and is the result of tryptophan metabolism to serotonin rather than nicotinic acid. In extreme cases, peripheral tryptophan consumption may deprive the brain of adequate levels of centrally-synthesized serotonin, which can result in neurocognitive dysfunction.[63] In a 2014 study, NET patients with carcinoid syndrome and hepatic metastases were compared to a group of patients with hepatic metastases from other primary sources. Patients

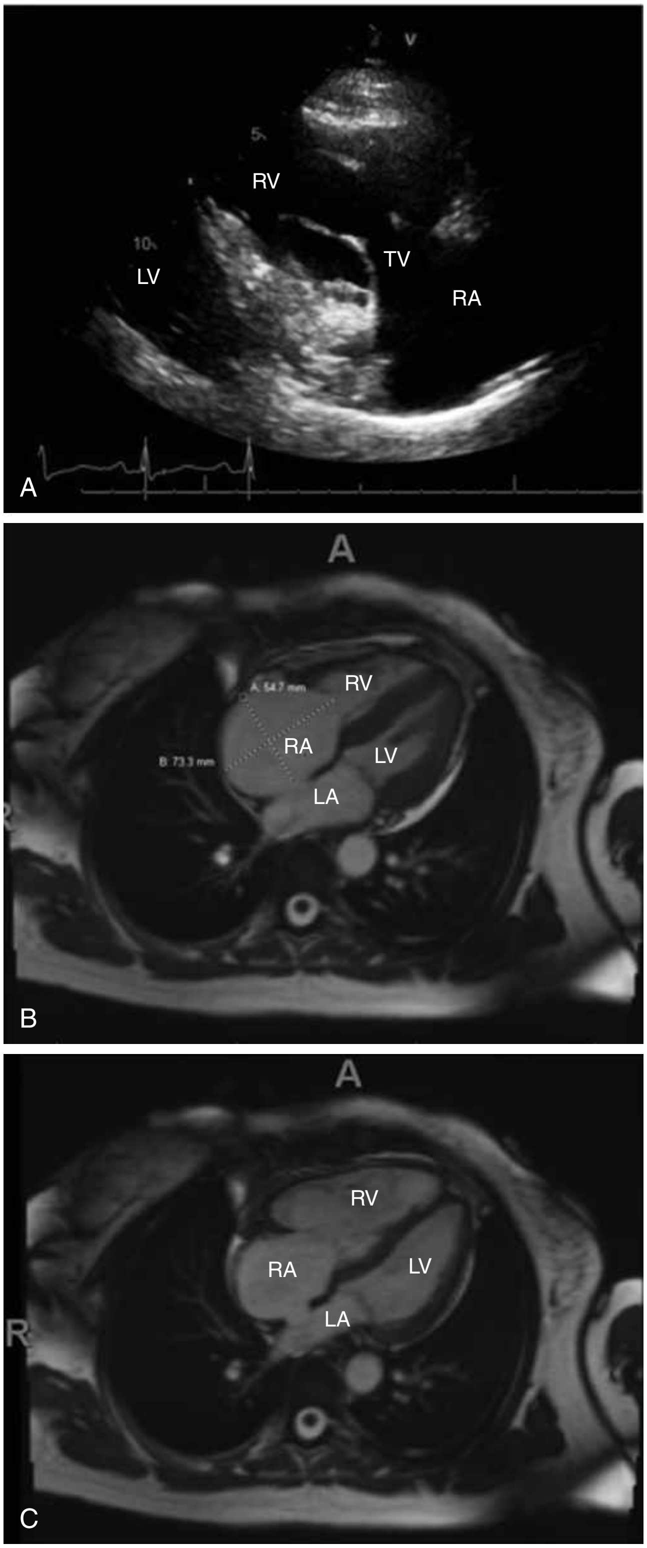


Fig. 135.3 **Carcinoid heart disease.** A 49-year-old woman with a jejunal neuroendocrine tumor, large-volume liver metastases, and carcinoid syndrome. Her serotonin hypersecretion resulted in carcinoid heart disease requiring replacement of both her pulmonic and tricuspid valves. (**A**) transthoracic echocardiogram, midsystole, demonstrating a lack of tricuspid valve coaptation during systole. (**B**) cardiac magnetic resonance imaging (mri), end systole. Right atrial enlargement as a result of severe tricuspid regurgitation. A normal atrium measures approximately 4–5 cm. (**C**) cardiac mri, systole. Severe, prolonged pulmonic and tricuspid regurgitation ultimately resulted in right heart failure. The right ventricle chamber is dilated, and right ventricular wall hypertrophy is evident. *RA,* right atrium; *RV*, right ventricle; *LA,* left atrium; *LV*, left ventricle; *TV*, tricuspid valve.

were matched for age, sex, and educational levels. All NET patients were currently being treated with SSAs, and although all had presented with elevated levels of u5-HIAA, 33% had levels in the normal range at the time of the study. Patients were surveyed across five cognitive domains. NET patients self-reported greater cognitive dysfunction than the non-NET cancer group and also performed worse on formal neurocognitive testing, particularly in the delayed recall domain.[64]

Carcinoid heart disease is a result of plaque-like deposits on the endocardial surfaces of the valve leaflets, cardiac chambers, and occasionally the intima of the aorta, pulmonary arteries, and inferior vena cava. Patients may present with dyspnea, fatigue, and a systolic murmur that is best heard at the sternal border. Left uncorrected, patients may develop right heart failure (Fig. 135.3) and manifest ascites, peripheral edema, and pleural effusions. Ninety percent of cases involve the right side of the heart. Left heart involvement is seen in situations of a patent foramen ovale with right-to-left shunt, with bronchial carcinoids, and when circulating concentrations of bioactive amines overwhelm the degradative capacity of the hepatic and pulmonary systems.[19]

Carcinoid crisis is a controversial hemodynamic phenomenon that is often considered in patients with carcinoid syndrome, as the underlying pathophysiology of the two entities is thought to be similar.[65] There is no consensus definition for carcinoid crisis, and thus its true prevalence is unknown. In patients undergoing surgical resection, rates are reported to be between 0% and 35%.[66,67] Crisis is generally described as an acute, prolonged episode of hemodynamic instability temporally related to an invasive procedure that cannot be ascribed to other periprocedural factors such as medication administration, hypovolemia, hemorrhage, or decreased venous return. Patients diagnosed with crises manifest cutaneous flushing and diarrhea and may also describe chest or abdominal pain. In addition to surgical procedures, carcinoid crises have been described in patients undergoing hepatic artery embolization, selective internal radiation therapy, peptide receptor radionuclide therapy, and tumor biopsy. As with carcinoid syndrome, therapy includes a bolus injection of 500 to 1000 ug of octreotide, or a continuous infusion of 50 to 100 ug/hour. Some practitioners administer prophylactic doses of octreotide to patients at risk for developing crisis, and others use intraprocedural infusions that are tapered over the next 8 to 24 hours after surgery. The efficacy of these treatments has been called into question, and institutional practices vary widely.[65,67]

KEY POINTS

- Functional neuroendocrine tumors (NETs) hypersecrete hormones and/or bioactive amines that result in a clinical syndrome. Although each described syndrome has hallmark features, the spectrum of signs and symptoms experienced by the patient can vary in number and severity. Insulinomas are the most common functional pancreas NET. Carcinoid syndrome is most often associated with small bowel NET. A fraction of these patients will develop carcinoid heart disease, typically manifest by right heart valvular dysfunction, and thus cardiac testing should play a role in these patients' ongoing disease surveillance.

MANAGEMENT STRATEGIES FOR FUNCTIONAL NEUROENDOCRINE TUMORS

Management of NET syndromes is achieved through cytoreduction of the secreting tumor bulk, antineoplastic agents that control cellular proliferation, or symptom management with medication. Cytoreduction can be attempted by a number of means, with surgical extirpation of the primary and all metastatic disease offering the only chance of cure. Unfortunately, metastasectomy is associated with high rates of recurrence. There is no consensus on the appropriate ordering of the various modalities, and medical management of symptoms can be difficult and require disease-specific expertise. Therefore, management discussions amongst members of a multidisciplinary team are crucial in the design of an optimal strategy (Fig. 135.4).

Surgical Resection

Regardless of functionality, patients presenting with a unifocal primary, limited regional disease, and good performance status should undergo resection of their primary tumor and locoregional lymph nodes. Patients with pancreas primaries are treated with pancreaticoduodenectomy when the tumor is located in the head, or distal pancreatectomy when the tumor(s) is located in the body or tail of the pancreas. Every effort should be made to operate upon patients with functional tumors to prevent progression of their symptoms. Some subtypes of PNET, especially gastrinomas, are often multifocal, which can complicate resection. Tumors distributed throughout the gland can be treated with total pancreatectomy. This is often reserved for only the most symptomatic patients refractory to other management because the operation is extremely morbid and can have a significant effect on postoperative quality of life. It is common for patients considering this option to meet with an endocrinologist before surgery to discuss the resulting brittle diabetes, and its management with an insulin pump, so that the full range of risks and benefits of the surgery are considered. In response to this issue, and the fact that some functional tumors seem to be more exophytic and encapsulated, “sitting” on the surface of the pancreas, enucleation is another option for tumor extirpation. Institutional practices of this technique vary, but it is a reasonable option when formal resection is suboptimal. Enucleation does spare pancreas parenchyma, but also increases the risk of postoperative pancreatic fistula.

Small intestine NETs are most appropriately approached in an open fashion (Fig. 135.5). This allows for resection of the dominant primary tumor, often located within 100 cm of the terminal ileum, regional lymphadenectomy that may include removal of bulky lymph nodes near the root of the mesentery, and careful palpation of the entire small bowel.[68] High-volume NET centers report small bowel tumor multifocality rates of up to 45%.[69]

The presence of distant metastases does not preclude surgical intervention, especially in cases of functional tumors, but it does make the risk-benefit calculus more difficult. Regardless of primary tumor site, the most common site for metastases is the liver. It is well accepted that surgery is the optimal management strategy for distant disease when removal of all microscopic and macroscopic disease (R0 resection) can be achieved. Survival benefit is hotly debated and has not been demonstrated prospectively, although retrospective series support the assertion that patients able to undergo aggressive resection have longer survival compared with those unable to have surgery (whether because of disease burden or fitness). Metastasectomy or parenchymal-sparing techniques (such as radiofrequency ablation) do confer symptomatic relief when 70% to 90% of the hepatic tumor bulk can be treated. This strategy has been used successfully for initial resections and in patients who present with recurrence of their symptoms years after their primary was removed.[70-72] There is only anecdotal evidence supporting isolated lymphadenectomy for symptom control, and it is only occasionally done in cases of recurrence of symptoms after resection of a functional tumor where the lymph nodes appear to be the only site of disease.

Ablation of Primary Pancreas Neuroendocrine Tumors

A variety of ablative techniques have demonstrated efficacy treating functional PNETs in patients with multifocal tumors or in those who

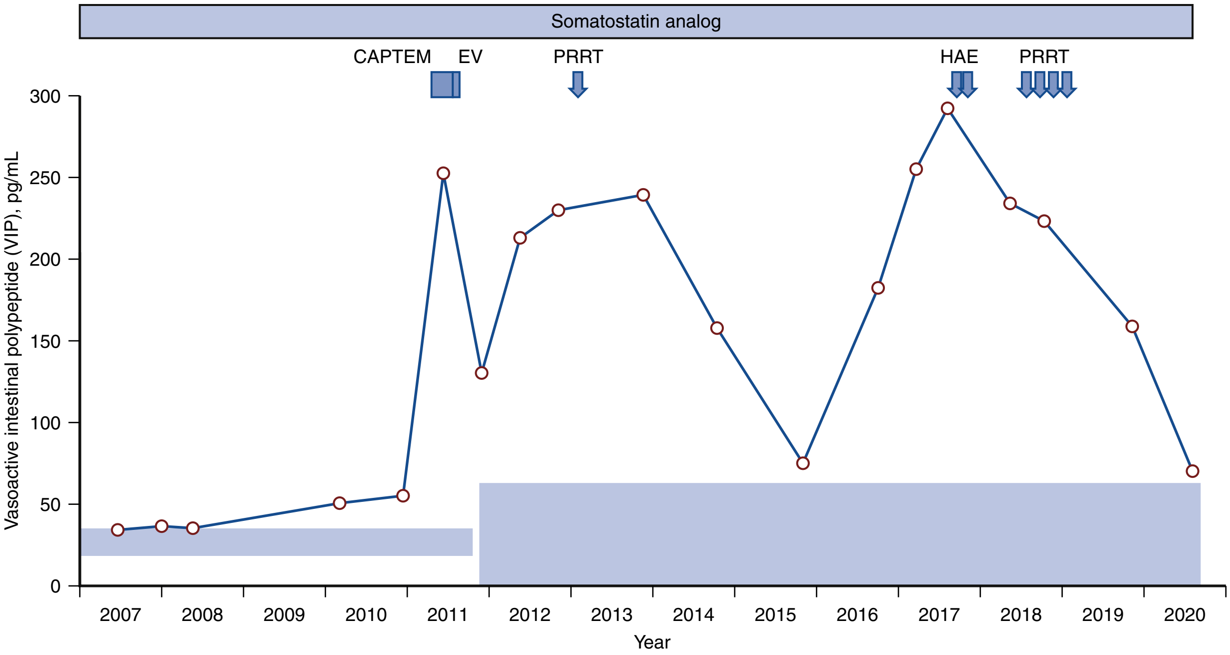


Fig. 135.4 **Treatment strategy in a patient with a well-differentiated, vasoactive intestinal peptide (*vip*)-secreting pancreas neuroendocrine tumor that was metastatic to the regional lymph nodes and liver.** The patient was diagnosed in 1996 and underwent distal pancreatectomy and liver debulking in 1998. Sequential ordering of the treatments required multiple discussions with a specialized, multidisciplinary team. The patient's clinical, radiographic, and biomarker responses to each therapy were monitored. Clinical features, including severe diarrhea, tracked closely with the vip levels. Therapies included distal pancreatectomy and liver debulking (1998), streptozocin and 5-fluorouracil (1996–2000), liver transplantation (2001), liver debulking surgery (2007), capecitabine and temozolomide (*Captem*, poorly tolerated, 2011), everolimus (*Ev*, not tolerated, 2011). The patient had one cycle of yttrium-90 peptide receptor radionuclide therapy (*Prrt*, 2013), two hepatic artery bland embolizations (hae, 2017), and four cycles of lutecium-177 prrt (2018). The patient was treated with a somatostatin analog throughout his disease course. Normal vip range indicated by the hatched bars.

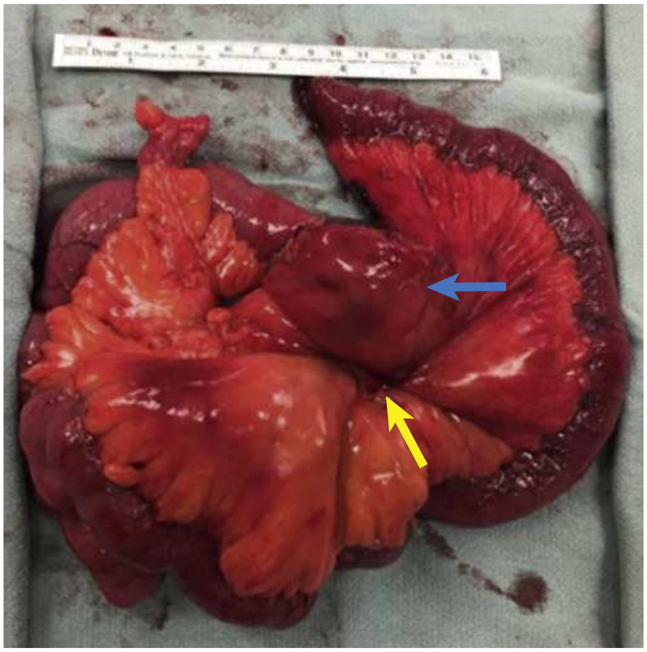


Fig. 135.5 **Mesenteric fibrosis in a surgical specimen, small bowel resection.** This patient had a large mesenteric mass *(blue arrow)*. The desmoplastic reaction caused puckering of the mesentery *(yellow arrow)*. The fibrosis occluded many venous branches, causing impaired drainage and congestion of the nearby small intestine. The patient complained of abdominal pain that was occasionally worse after meals.

are poor candidates for formal pancreatectomy. Tumors can be targeted laparoscopically or endoscopically with ultrasound guidance. Ablation may be achieved with ethanol, radiowave, or microwave ablation, or with electroporation. Treatment success is gauged by radiographic and symptomatic response. The majority of series report amelioration of symptoms and normalization of hormone levels within 24 hours after treatment. Patients can undergo multiple treatments of the same tumor, either in short succession if complete ablation is not achieved, or in cases of tumor recurrence. The most common complication is pancreatitis and most patients are observed for 1 to 2 days after the procedure. Late complications, such as duct stricture, are rare (<5%).[73-75]

Hepatic Arterial Embolization

NET liver metastases are supplied by the hepatic arteries, rather than the portal venous circulation. Embolization of the supplying artery results in targeted ischemia of the tumor and leaves normal parenchyma relatively spared. There are three techniques available—bland embolization, chemoembolization, and drug-eluting bead transarterial chemoembolization—although bland and chemoembolization are the most widely used. A prospective study is underway to determine the optimal embolization technique to treat NET liver metastases, and was written to compare all three modalities (www.clinicaltrials.gov; NCT02724540). However, the DEB-TACE arm closed early because of toxicity concerns.[76]

Retrospective studies in patients with functional NETs report objective tumor responses of approximately 50% and symptomatic responses ranging from 9% to 100% (average 56%), but

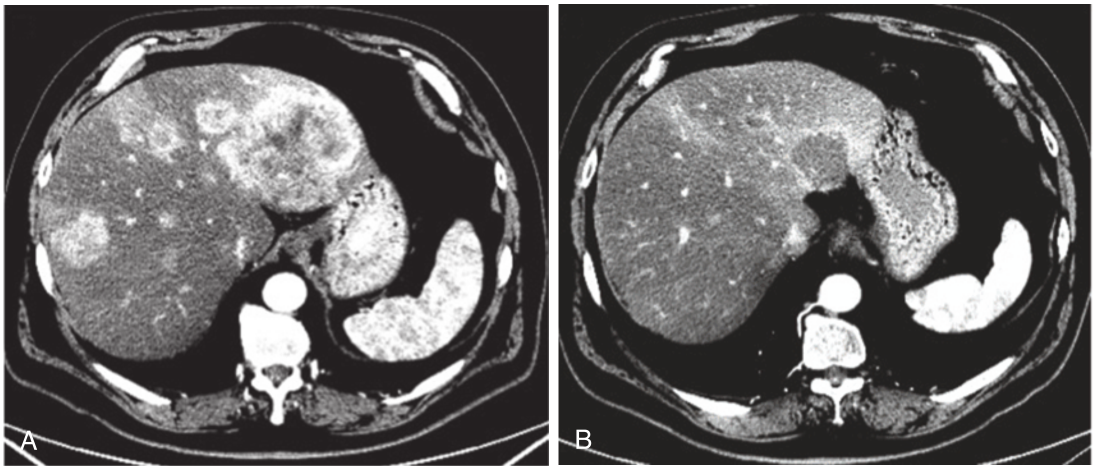


Fig. 135.6 **Radiographic response to the combination regimen of capecitabine and temozolomide.** A 70-year-old man with metastatic grade i pancreatic neuroendocrine tumor. Serial computed tomography scans of the abdomen, axial view, arterial phase. Images obtained (**A**) 2 weeks prior and (**B**) 2 months following 12 cycles of capecitabine 750 mg/m^2 by mouth daily on days 1–14 and temozolomide 200 mg/m^2 by mouth daily on days 10–14 of a 28-day cycle.

mean clinical response rates do not seem to differ by technique.[77] Biochemical responses are noted in more than 50% of patients, and are enhanced in patients with carcinoid syndrome when simultaneously treated with an SSA.[78] Smaller case series report symptom relief in patients treated with hepatic arterial embolization (HAE), even in situations where patients had other hepatic or distant metastases left untreated.[79] A patient with hepatic, bone, and mediastinal metastases from a malignant insulinoma underwent bland HAE for symptoms refractory to multiple treatment modalities. The patient's hypoglycemic symptoms resolved within 24 hours. Symptoms recurred 5 months later.[80]

Survival correlates with objective tumor response but has only been shown in one retrospective study to correlate with biochemical or symptom response.[78] The most common complication is postembolization syndrome, where patients develop abdominal pain, nausea, fevers, leukocytosis, and transient transaminitis that resolve with time. Serious complications such as cholangitis, gastrointestinal bleeding, hepatic abscess, gallbladder necrosis, and renal failure are rare (<5%).[78,81] Late complications such as biliary ischemia and hepatic failure occur in fewer than 2% of patients.[78]

Selective Internal Radiation Therapy

Selective internal radiation therapy is another technique for the treatment of unresectable hepatic metastases that uses yttrium-90–coated microspheres. Delivery of this agent to nontarget organs can cause disastrous complications, and thus each procedure is preceded by an angiogram of extrahepatic arteries and determination of lung shunt fraction. This has not shown superiority in survival over other embolization techniques, but is known to induce radiographic objective tumor responses, increases in PFS and OS, and symptomatic improvement.[82-84] This procedure is safe, with low rates of postembolization syndrome and postprocedural infection.[82] More serious complications, such as hepatic cirrhosis, are only rarely reported.[85]

Systemic Therapies

Somatostatin Analogs. SSAs are synthetic versions of the endogenous hormone somatostatin. There are a number of formulations that differ in their duration of action or SSTR specificity, with octreotide and lanreotide being the best known. These medications demonstrate efficacy in controlling symptoms and slowing tumor proliferation.[86-88] They are typically the first agents selected in patients with advanced disease and hormonally driven symptoms negatively impacting quality of life. Octreotide is available in intermediate-acting and long-acting forms. Lanreotide is exclusively a long-acting formulation that is administered once per month.

Meta-analyses detailing the efficacy of intermediate-acting octreotide in patients with functional GEP NETs demonstrate resolution or improvement of diarrhea and flushing in 71% of patients, with biochemical responses in 77%. Long-acting forms of octreotide are similarly efficacious in controlling symptoms in patients with functional tumors, with the majority of data derived from studies examining use in small intestine NET and carcinoid syndrome.[89] Episodic symptom control can be achieved by using intermediate-acting octreotide in addition to monthly injections. The utility of lanreotide for symptom control was most recently demonstrated in the 2016 ELECT trial. In the 115 patients with carcinoid syndrome studied, those randomized to lanreotide used breakthrough octreotide significantly less frequently for symptom control compared with the control group.[88] A follow-up study of patient-reported outcomes found that patients in the lanreotide group had improved diarrhea and flushing compared with the placebo group.[90] Lanreotide can also induce a biochemical response in patients with elevated CgA or u5-HIAA, regardless of whether the tumor is functional. Pooled analysis of the patients enrolled in the CLARINET[87] and ELECT trials showed biochemical responses in 43.1% of patients with elevated u5-HIAA and 38.5% of those with elevated CgA.[91]

Despite the efficacy of these drugs, a subset of treated patients will either develop resistance to the medication or not respond at all to standard doses. In these situations, escalated doses of SSA have been tried. Escalated doses are safe, and approximately one-half to three-quarters of patients reported at least some improvement in frequency of diarrhea and flushing.[92] A multinational study of escalating dose lanreotide reported a decrease in bowel movements from 4.7 to 2.6 per day, with 76% of patients reporting that they were "completely or rather satisfied" with their diarrhea control at

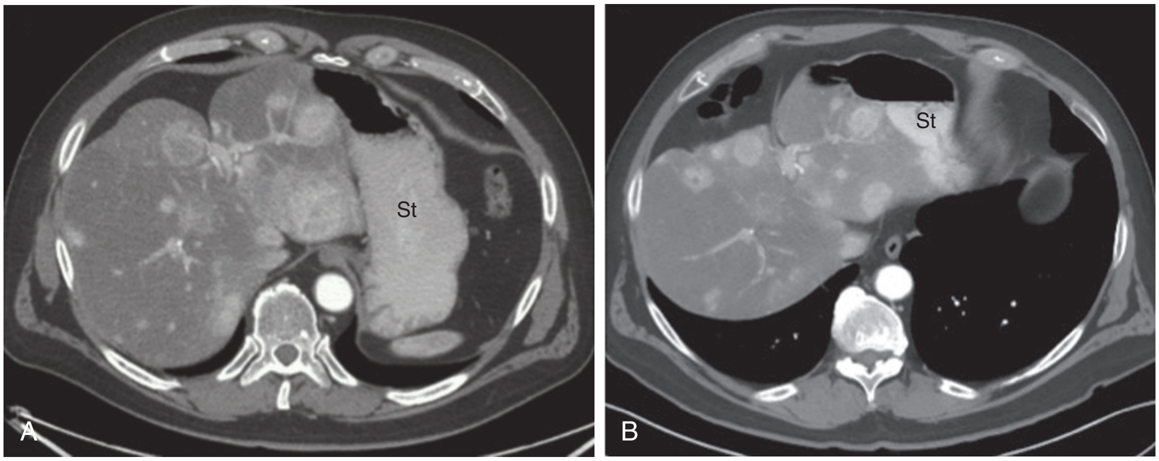


Fig. 135.7 **Radiographic response to peptide receptor radionuclide therapy.** A 65-year-old man with metastatic grade II pancreatic neuroendocrine tumor. Serial computed tomography scans, axial view, arterial phase. Images taken (**A**) 1 month before and (**B**) 1 month following four doses of 200 mci of ^{177}lu-dotatate, at 8-week intervals. *St*, stomach.

these doses. Investigators noted that patients reporting the greatest satisfaction often reported subjectively worse symptoms before treatment.[93]

Chemotherapy. Choice of chemotherapeutic regimen depends on tumor histology and grade. In general, well-differentiated extrapancreatic NETs respond poorly to cytotoxic chemotherapy, with PNETs responding better. The first chemotherapy regimens were streptozocin-based, and randomized studies showed improvements in survival and biochemical responses.[94,95] Experience with 5-fluorouracil, doxorubicin, and streptozocin was described in a group of 84 patients with PNETs, of which 20 were functioning. Median PFS was estimated at 18 months, and median OS at 37 months. Patients with radiographic responses to treatment were most likely to demonstrate a 30% or greater decrease in their CgA levels. Symptomatic response was not documented.[96] Temozolomide combined with capecitabine is another commonly used regimen in PNET (Fig. 135.6). A prospective study comparing this combination of drugs to temozolomide alone found that patients treated with the combination had improved PFS (22.7 vs. 14.4 months) and OS (not reached vs. 38 months).[97] Platinum-based regimens are most common for NECs[24] and lung NETs. In lung NETs, the overall response rate is only 8%.[98]

Targeted Therapy. Targeted therapies such as sunitinib and everolimus have been approved for use in NET and demonstrate efficacy in a variety of histologies. Sunitinib is a vascular endothelial growth factor inhibitor approved for use exclusively in patients with advanced PNET. In the study that led to the drug's approval, patients randomized to the treatment arm demonstrated a hazard ratio (HR) for progression or death of 0.42, favoring sunitinib. The drug was used in both functioning and nonfunctioning PNETs, but biochemical and symptomatic responses were not detailed.[99] Sunitinib has limited proven benefit in extrapancreatic gastrointestinal NET.[100]

The mammalian target of rapamycin pathway regulates cell growth, proliferation, and metabolism and is implicated in NET pathogenesis. The RADIANT-2 study was a phase III, placebo-controlled study investigating the effects of everolimus plus octreotide long-acting release in a group of patients with metastatic, functional GEP and lung NET. The study failed to demonstrate a statistically significant improvement in the time to disease progression or death. The participants' subjective responses to treatment were not characterized, but a greater proportion of patients in the treatment group had biochemical responses in u5-HIAA and CgA compared with the placebo group.[101] RADIANT-3 randomized patients with advanced, progressive PNET to everolimus or placebo. The HR for progression or death was 0.35, favoring everolimus.[102]

Everolimus has demonstrated efficacy in treating symptoms associated with a variety of functional NETs. One retrospective study including patients with small intestine, pancreas, and lung NETs reported that 70% of patients started on everolimus for the indication of symptom control had significant improvement of their diarrhea and flushing for a median duration of 13.9 months. These patients had simultaneous radiologically stable or responsive disease.[103] The drug has unique efficacy in insulinoma and demonstrates simultaneous antitumor and symptom effects. Everolimus can improve or resolve hypoglycemic episodes in patients with metastatic disease and refractory hypoglycemia.[104] In a series of 12 patients, 11 had near-immediate improvement in symptoms and ability to discontinue other medications used to control their hypoglycemia. The median duration of symptomatic response was 6.5 months.[105]

Peptide Receptor Radionuclide Therapy. PRRT is a treatment option for patients with metastatic disease (Fig. 135.7). The radiolabeled SSA 177Lu-DOTATATE was approved for use in SSTR-positive GEP NETs in 2018. In 2017, the phase III NETTER-1 trial compared patients with progressive, well-differentiated, metastatic small intestine NETs treated with PRRT versus high-dose octreotide. The study demonstrated an improvement in PFS of 79% in the treatment arm, a statistically significant survival advantage, although a response rate of only 18%. Patient-reported outcomes have not yet been reported.[106]

A number of retrospective studies have cataloged patient response to PRRT. The patient populations are heterogeneous and include both functional small intestine NETs and PNETs. Biochemical responses and symptomatic improvements correlate with radiographic responses, as only 30% of patients with stable or progressive disease reported a reduction in symptoms after treatments. In patients who eventually progressed, loss of symptomatic control occurred after or concurrent with radiographic progression.[107] The majority of patients are treated with four cycles of PRRT, but biochemical

TABLE 135.5 Adjunctive Treatments for Functional Neuroendocrine Tumors

Tumor Type/ Syndrome	Symptom	Treatment
Insulinoma	Hypoglycemia	Diazoxide Infusion of crystalloid dextrose Corticosteroids Everolimus SSA
Gastrinoma	Reflux Peptic ulceration	High-dose proton pump inhibitor SSA
Carcinoid diarrhea	Secretory diarrhea	Telotristat SSA

SSA, Somatostatin analog.

and symptom responses may begin to manifest after the first one to two cycles. Most patients reported clinical response to treatment, with response rates greater than 50%.[107,108] When patients' reported responses were broken down by individual symptom type, those with poorly controlled diarrhea experienced the greatest benefits (70.7% had improvement), whereas the majority of patients struggling with weight saw little improvement (48% reported no change, and 22.9% had worsening of their weight loss).[108] Fifty percent of patients presenting with elevated serum serotonin had a decrease of greater than 50% in these levels after treatment, with only 25.2% and 15.6% experiencing biochemical responses in CgA and u5-HIAA, respectively.[108]

Palliative Symptom Management

Patients with advanced disease may require palliation of their symptoms beyond the control that may be provided by the cytoreductive or cytotoxic treatments previously discussed (Table 135.5).

Insulinoma. Patients with insulinomas can experience profound episodes of hypoglycemia. In early stages of the disease, patients may be able to maintain normoglycemia with frequent meals. Many will experience significant weight gain, as the frequency of meals and caloric intake required exceed a patient's basal metabolic rate. As the disease progresses, patients develop nocturnal hypoglycemia and risk developing severe side effects such as seizure. These episodes can be treated with nocturnal parenteral or enteral nutrition, which will require the patient have an enteral feeding tube placed, or a tunneled central venous line. Diazoxide is a medication that controls insulin secretion by opening potassium channels. It is administered daily in doses ranging from 50 to 1500 mg. Unfortunately, it is only effective in 50% of cases. Side effects, experienced by approximately 50% of patients, include palpitations, nausea, anorexia, hirsutism, and peripheral edema. Corticosteroids can also be used for symptom control. They are fast-acting, but also cause immunosuppression.[51]

Gastrinoma. Zollinger–Ellison syndrome because of hypersecretion of gastrin can cause recurrent, devastating peptic ulcer disease. Symptoms can be controlled with proton pump inhibitors. Administered at high doses (omeprazole 60 mg three times daily, either intravenously or orally), studies have shown increases in gastric pH and a decrease in the volume of gastric fluid secreted. Patients report improvement of regurgitation and reflux within 2 weeks of treatment initiation. Endoscopy performed 1 month after treatment initiation demonstrated healed ulcers.[109,110]

Carcinoid Syndrome Diarrhea. Elevated levels of serotonin are associated with diarrhea. Recently, telotristat ethyl was approved by the US Food and Drug Administration for use in patients with carcinoid syndrome diarrhea refractory to treatment with an SSA. This drug inhibits tryptophan hydroxylase, the rate-limiting step in serotonin synthesis. The phase 3 TELESTAR study reported a reduction in bowel movement (BM) frequency of 2.1 BM/day from baseline, but –0.81 BM/day compared with placebo. Responders did have a significant decrease in u5-HIAA levels.[111] TELECAST, a phase III companion study to TELESTAR, demonstrated similar results, with treated patients reporting fewer BMs per day and a decrease in u5-HIAA levels.[112] Patients in a phase II study of telotristat ethyl were interviewed about their experiences taking the medication approximately 29 months after study initiation. Their reactions were favorable, as 82% reported improvements in diarrhea, 45% noted improved abdominal pain, 36% suffered less abdominal cramping, and 36% had fewer episodes of flushing.[113]

KEY POINTS

- Optimal treatment of hypersecreting neuroendocrine tumors (NETs) involves cytoreduction of the secreting tumor bulk, control of cellular proliferation, and symptom management. There are a variety of means available to practitioners to achieve these goals, but no consensus on appropriate ordering, thus multidisciplinary discussion amongst NET specialists is crucial to optimize outcome. Surgical resection is helpful in patients with unifocal primary tumors, limited regional or distant disease, and good performance status. Liver-directed therapies may induce symptomatic responses as well as radiologic responses. Systemic treatments include peptide receptor radionuclide therapy, targeted therapies, cytotoxic chemotherapy, and somatostatin analogs. These are typically employed when the disease has spread beyond what is reasonable to resect or treat with more directed therapies. Medications for symptom management can be employed at any point during a patient's treatment course, but will rarely durably ameliorate symptoms in isolation.

CONCLUSION

NETs are unique in their ability to produce hormonal syndromes that can significantly impact a patient's quality of life. The rarity of these tumors and their associated clinical syndromes can make diagnosis difficult, and it is often delayed by many years. Once a diagnosis is made, treatment selection and ordering can present challenges, and optimal care requires a multidisciplinary team of surgeons, endocrinologists, medical oncologists, nurse clinicians, pathologists, nuclear medicine physicians, and interventional radiologists experienced in managing these unique diseases. Recent trials in targeted therapies and inhibition of serotonin synthesis have led to improvements in the quality of life of NET patients. Ongoing and future studies of agents and techniques aimed toward cytoreduction and symptom control will deepen the armamentarium available to treat this unique set of patients.

REFERENCES

1. Tsoucalas G, Karamanou M, Androutsos G. The eminent German pathologist Siegfried Oberndorfer (1876-1944) and his landmark work on carcinoid tumors. *Ann Gastroenterol*. 2011;24:98–100.
2. Cives M, Strosberg JR. Gastroenteropancreatic neuroendocrine tumors. *CA Cancer J Clin*. 2018;68:471–487.
3. Dasari A, Shen C, Halperin D, et al. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*. 2017;3:1335–1342.

4. Fottner C, Ferrata M, Weber MM. Hormone secreting gastro-entero-pancreatic neuroendocrine neoplasias (GEP-NEN): when to consider, how to diagnose? *Rev Endocr Metab Disord*. 2017;18:393–410.
5. Hann CL, Forde PM. Lung and thymic carcinoids. *Endocrinol Metab Clin North Am*. 2018;47:699–709.
6. Halperin DM, Shen C, Dasari A, et al. Frequency of carcinoid syndrome at neuroendocrine tumour diagnosis: a population-based study. *Lancet Oncol*. 2017;18:525–534.
7. Halfdanarson TR, Rabe KG, Rubin J, et al. Pancreatic neuroendocrine tumors (PNET): incidence, prognosis and recent trend toward improved survival. *Ann Oncol*. 2008;19:1727–1733.
8. Grande E, Capdevila J, Barriuso J, et al. Gastroenteropancreatic neuroendocrine tumor cancer stem cells: do they exist? *Cancer Metastasis Rev*. 2012;31:47–53.
9. Vortmeyer AO, Huang S, Lubensky I, et al. Non-islet origin of pancreatic islet cell tumors. *J Clin Endocrinol Metab*. 2004;89:1934–1938.
10. Anlauf M, Garbrecht N, Bauersfeld J, et al. Hereditary neuroendocrine tumors of the gastroenteropancreatic system. *Virchows Arch*. 2007;451:S29–S38.
11. Sei Y, Zhao X, Forbes J, et al. A hereditary form of small intestinal carcinoid associated with a germline mutation in inositol polyphosphate multikinase. *Gastroenterology*. 2015;149:67–78.
12. Klimstra DS. Pathologic classification of neuroendocrine neoplasms. *Hematol Oncol Clin North Am*. 2016;30:1–19.
13. Nagtegaal ID, Odze RD, Klimstra D, et al. The 2019 WHO Classification of tumours of the digestive system. *Histopathology*. 2020;76:182–188.
14. Rindi G, Klimstra DS, Abedi-Ardekani B, et al. A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. *Mod Pathol*. 2018;31:1770–1786.
15. Nadler A, Cukier M, Rowsell C, et al. Ki-67 Is a reliable pathological grading marker for neuroendocrine tumors. *Virchows Arch*. 2013;462:501–505.
16. Turner GB, Johnston BT, McCance DR, et al. Circulating markers of prognosis and response to treatment in patients with midgut carcinoid tumours. *Gut*. 2006;55:1586–1591.
17. Formica V, Wotherspoon A, Cunningham D, et al. The prognostic role of WHO classification, urinary 5-hydroxyindoleacetic acid and liver function tests in metastatic neuroendocrine carcinomas of the gastroenteropancreatic tract. *Br J Cancer*. 2007;96:1178–1182.
18. van der Horst-Schrivers AN, Post WJ, Kema IP, et al. Persistent low urinary excretion of 5-HIAAiaa is a marker for favourable survival during follow-up in patients with disseminated midgut carcinoid tumours. *Eur J Cancer*. 2007;43:2651–2657.
19. Davar J, Connolly HM, Caplin ME, et al. Diagnosing and managing carcinoid heart disease in patients with neuroendocrine tumors: an expert statement. *J Am Coll Cardiol*. 2017;69:1288–1304.
20. Zandee WT, Kamp K, van Adrichem RC, et al. Effect of Hormone secretory syndromes on neuroendocrine tumor prognosis. *Endocr Relat Cancer*. 2017;24:R261–R274.
21. Janson ET, Holmberg L, Stridsberg M, et al. Carcinoid tumors: analysis of prognostic factors and survival in 310 patients from a referral center. *Ann Oncol*. 1997;8:685–690.
22. Zandee WT, van Adrichem RC, Kamp K, et al. Incidence and prognostic value of serotonin secretion in pancreatic neuroendocrine tumours. *Clin Endocrinol*. 2017;87:165–170.
23. Klimstra DS, Modlin IR, Coppola D, et al. The Pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas*. 2010;39:707–712.
24. Sorbye H, Welin S, Langer SW, et al. Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the Nordic NEC study. *Ann Oncol*. 2013;24:152–160.
25. Tang LH, Untch BR, Reidy DL, et al. Well differentiated neuroendocrine tumors with a morphologically apparent high-grade component: a pathway distinct from poorly differentiated neuroendocrine carcinomas. *Clin Cancer Res*. 2016;22:1011–1017.
26. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol*. 2015;10:1243–1260.
27. Aluri V, Dillon JS. Biochemical testing in neuroendocrine tumors. *Endocrinol Metab Clin North Am*. 2017;46:669–677.
28. Anthony LB, Strosberg JR, Klimstra DS, et al. The NANET consensus guidelines for the diagnosis and management of gastrointestinal neuroendocrine tumors (NET): well differentiated NETs of the distal colon and rectum. *Pancreas*. 2010;39:767–774.
29. Boudreaux JP, Klimstra DS, Hassan MM, et al. The NANET consensus guideline for the diagnosis and management of neuroendocrine tumors: well differentiated neuroendocrine tumors of the jejunum, ileum, appendix, and cecum. *Pancreas*. 2010;39:753–766.
30. Halfdanarson TR, Strosberg JR, Tang L, et al. The North American Neuroendocrine Tumor Society consensus guidelines for surveillance and medical management of pancreatic neuroendocrine tumors. *Pancreas*. 2020;49:863–881.
31. Maxwell JE, O'Dorisio TM, Howe JR. Biochemical diagnosis and preoperative imaging of gastroenteropancreatic neuroendocrine tumors. *Surg Oncol Clin N Am*. 2016;25:171–194.
32. Hirshberg B, Livi A, Bartlett DL, et al. Forty-Eight-hour fast: the diagnostic test for insulinoma. *J Clin Endocrinol Metab*. 2000;85:3222–3226.
33. Lamers CBH, van Tongeren JHM. Comparative study of the value of the calcium, secretin, and meal stimulated increase in serum gastrin to the diagnosis of the Zollinger-Ellison syndrome. *Gut*. 1977;18:128–134.
34. Metz DC, Cadiot G, Poitras P, et al. Diagnosis of Zollinger-Ellison syndrome in the era of ppis, faulty gastrin assays, sensitive imaging and limited access to acid secretory testing. *Int J Endocr Oncol*. 2017;4:167–185.
35. Anderson MA, Carpenter S, Thompson NW, et al. Endoscopic ultrasound is highly accurate and directs management in patients with neuroendocrine tumors of the pancreas. *Am J Gastroenterol*. 2000;95:2271–2277.
36. Bellutti M, Fry LC, Schmitt J, et al. Detection of Neuroendocrine tumors of the small bowel by double balloon enteroscopy. *Dig Dis Sci*. 2009;54:1050–1058.
37. van Tuyl SA, van Noorden JT, Timmer R, et al. Detection of small-bowel neuroendocrine tumors by video capsule endoscopy. *Gastrointest Endosc*. 2006;64:66–72.
38. Sundin A, Vullierme MP, Kaltsas G, et al. ENET consensus guidelines for the standards of care in neuroendocrine tumors: radiological examinations. *Neuroendocrinology*. 2009;90:167–183.
39. Reubi JC. Somatostatin and other peptide receptors as tools for tumor diagnosis and treatment. *Neuroendocrinology*. 2004;80:51–56.
40. Reubi JC, Krenning E, Lamberts SWJ. Distribution of somatostatin receptors in normal and tumor tissue. *Metabolism*. 1990;39:78–81.
41. Binderup T, Knigge U, Loft A, et al. Functional imaging of neuroendocrine tumors: a head-to-head comparison of somatostatin receptor scintigraphy, 123i-MIBG scintigraphy, and 18F-FDG PET. *J Nucl Med*. 2010;51:704–712.
42. Hope TA, Bergsland E, Bozkurt MF, et al. Appropriate use criteria for somatostatin receptor PET imaging in neuroendocrine tumors. *J Nucl Med*. 2018;59(1):66–74.
43. Krenning EP, Kwekkeboom DJ, Bakker WH, et al. Somatostatin receptor scintigraphy with [111In-DTPA-D-Phe1]- and [123I-Tyr3]-octreotide: the rotterdam experience with more than 1000 patients. *Eur J Nucl Med*. 1993;20:716–731.
44. Pattison DA, Hicks RJ. Molecular imaging in the investigation of hypoglycaemic syndromes and their management. *Endocr Relat Cancer*. 2017;24:R203–R221.
45. Andreassen M, Ilett E, Wiese D, et al. Surgical management, preoperative tumor localization, and histopathology of 80 patients operated on for insulinoma. *J Clin Endocrinol Metab*. 2019;104:6129–6138.
46. Bahri H, Laurence L, Edeline J, et al. High prognostic value of 18F-FDG PET for metastatic gastroenteropancreatic neuroendocrine tumors: a long-term evaluation. *J Nucl Med*. 2014;55:1786–1790.
47. Bellizzi AM. Assigning site of origin in metastatic neuroendocrine neoplasms: a clinically significant application of diagnostic immunohistochemistry. *Adv Anat Pathol*. 2013;20:285–314.
48. Whipple AO, Frantz VK. Adenoma of islet cells with hyperinsulinism. *Ann Surg*. 1935;101:1299–1335.

49. Okabayashi T, Shima Y, Sumiyoshi T, et al. Diagnosis and management of insulinoma. *World J Gastroenterol*. 2013;19:829–837.
50. Wang H, Ba Y, Xing Q, et al. Diagnostic value of ASVS for insulinoma localization: a systematic review and meta-analysis. *PLoS One*. 2019;14:e0224928.
51. Baudin E, Caron P, Lombard-Bohas C, et al. Malignant insulinoma: recommendations for characterisation and treatment. *Ann Endocrinol*. 2013;74:523–533.
52. Veltroni A, Cosaro E, Spada F, et al. Clinico-pathological features, treatments and survival of malignant insulinomas: a multicenter study. *Eur J Endocrinol*. 2020;182:439–446.
53. Zollinger RM, Ellison EH. Primary peptic ulcerations of the jejunum associated with islet cell tumors of the pancreas. *Ann Surg*. 1955;142:709–723.
54. Stabile BE, Morrow DJ, Passaro E. The gastrinoma triangle: operative implications. *Am J Surg*. 1984;147:25–31.
55. Gibril F, Doppman JL, Reynolds JC, et al. Bone metastases in patients with gastrinomas: a prospective study of bone scanning, somatostatin receptor scanning, and magnetic resonance image in their detection, frequency, location, and effect of their detection on management. *J Clin Oncol*. 1998;16:1040–1053.
56. Verner JV, Morrison AB. Islet cell tumor and a syndrome of refractory watery diarrhea and hypokalemia. *Am J Med*. 1958;25:374–380.
57. Schizas D, Mastoraki A, Bagias G, et al. Clinicopathological data and treatment modalities for pancreatic VIPomas: a systematic review. *J BUON*. 2019;24:415–423.
58. Angelousi A, Koffas A, Grozinsky-Glasberg S, et al. Diagnostic and management challenges in vasoactive intestinal peptide secreting tumors: a series of 15 patients. *Pancreas*. 2019;48:934–942.
59. Eldor R, Glaser B, Fraenkel M, et al. Glucagonoma and the glucagonoma syndrome - cumulative experience with an elusive endocrine tumour. *Clin Endocrinol*. 2011;74:593–598.
60. Williamson JM, Thorn CC, Spalding D, et al. Pancreatic and peripancreatic somatostatinomas. *Ann R Coll Surg Engl*. 2011;93:356–360.
61. Maxwell JE, O'Dorisio TM, Bellizzi AM, et al. Elevated pancreatic polypeptide levels in pancreatic neuroendocrine tumors and diabetes mellitus: causation or association? *Pancreas*. 2014;43:651–656.
62. Kuo SC, Gananadha S, Scarlett CJ, et al. Sporadic pancreatic polypeptide secreting tumors (PPomas) of the pancreas. *World J Surg*. 2008;32: 1815–1822.
63. Rubin de Celis Ferrari AC, Glasberg J, Riechelmann RP. Carcinoid syndrome: update on the pathophysiology and treatment. *Clinics (Sao Paulo)*. 2018;73:e490s.
64. Pasieka JL, Longman RS, Chambers AJ, et al. Cognitive Impairment associated with carcinoid syndrome. *Ann Surg*. 2014;259:355–359.
65. Condron ME, Pommier SJ, Pommier RF. Continuous infusion of octreotide combined with perioperative octreotide bolus does not prevent intraoperative carcinoid crisis. *Surgery*. 2016;159:358–365.
66. Fouche M, Bouffard Y, Le Goff MC, et al. Intraoperative carcinoid syndrome during small-bowel neuroendocrine tumour surgery. *Endocr Connect*. 2018;7:1245–1250.
67. Kinney MAO, Nagorney DM, Clark DF, et al. Partial hepatic resections for metastatic neuroendocrine tumors: perioperative outcomes. *J Clin Anesth*. 2018;51:93–96.
68. Howe JR, Cardona K, Fraker DL, et al. The surgical management of small bowel neuroendocrine tumors: consensus guidelines of the North American Neuroendocrine Tumor Society. *Pancreas*. 2017;46:715–731.
69. Choi AB, Maxwell JE, Keck KJ, et al. Is multifocality an indicator of aggressive behavior in small bowel neuroendocrine tumors? *Pancreas*. 2017;46:1115–1120.
70. Wilson H, Butterick OD. Massive liver resection for control of severe vasomotor reactions secondary to malignant carcinoid. *Ann Surg*. 1959;149:641–647.
71. Maxwell JE, Sherman SK, O'Dorisio TM, et al. Liver-directed surgery of neuroendocrine metastases: what is the optimal strategy? *Surgery*. 2016;159:320–333.
72. Chambers AJ, Pasieka JL, Dixon E, et al. The palliative benefit of aggressive surgical intervention for both hepatic and mesenteric metastases from neuroendocrine tumors. *Surgery*. 2008;144:645–653.
73. Yao C, Wang X, Zhang Y, et al. Treatment of insulinomas by laparoscopic radiofrequency ablation: case reports and literature review. *Open Med (Wars)*. 2020;15:84–91.
74. Chen OT, Dojki FK, Weber SM, et al. Percutaneous microwave ablation of an insulinoma in a patient with refractory symptomatic hypoglycemia. *J Gastrointest Surg*. 2015;19:1378–1381.
75. Park DH, Choi JH, Oh D, et al. Endoscopic ultrasonography-guided ethanol ablation for small pancreatic neuroendocrine tumors: results of a pilot study. *Clin Endosc*. 2015;48:158–164.
76. Chen JX, Wileyto EP, Soulen MC. Randomized embolization trial for neuroendocrine tumor metastases to the liver (RETNET): study protocol for a randomized controlled trial. *Trials*. 2018;19:390.
77. Pericleous M, Caplin ME, Tsochatzis E, et al. Hepatic artery embolization in advanced neuroendocrine tumors: efficacy and long-term outcomes. *Asia Pac J Clin Oncol*. 2016;12:61–69.
78. Del Prete M, Fiore F, Modica R, et al. Hepatic arterial embolization in patients with neuroendocrine tumors. *J Exp Clin Cancer Res*. 2014;33:43.
79. Case CC, Wirfel K, Vassilopoulou-Sellin R. Vasoactive intestinal polypeptide-secreting tumor (VIPoma) with liver metastases: dramatic and durable symptomatic benefit from hepatic arterial embolization, a case report. *Med Oncol*. 2002;19:181–187.
80. Kyriacou A, Mansoor W, Lawrance J, et al. Intractable hypoglycemia in a patient with advanced carcinoid syndrome successfully treated with hepatic embolization. *Hormones (Basel)*. 2015;15:118–121.
81. O'Toole D, Ruszniewski P. Chemoembolization and other ablative therapies for liver metastases of gastrointestinal endocrine tumours. *Best Pract Res Clin Gastroenterol*. 2005;19:585–594.
82. Dermine S, Palmieri LJ, Lavole J, et al. Non-pharmacological therapeutic options for liver metastases in advanced neuroendocrine tumors. *J Clin Med*. 2019;8:1907.
83. Kennedy AS, Dezarn WA, McNeillie P, et al. Radioembolization for unresectable neuroendocrine hepatic metastases using resin 90Y-microspheres: early results in 148 patients. *Am J Clin Oncol*. 2008;31:271–279.
84. Kennedy A, Bester L, Salem R, et al. Role of hepatic intra-arterial therapies in metastatic neuroendocrine tumours (NET): guidelines from the NET-Liver-Metastases Consensus Conference. *HPB*. 2015;17:29–37.
85. Loree JM, Hiruki T, Kennecke HF. Case report of cirrhosis following yttrium-90 radioembolization for pancreatic neuroendocrine liver metastases. *Case Rep Oncol*. 2016;9:76–82.
86. Rinke A, Muller HH, Schade-Brittinger C, et al. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*. 2009;27:4656–4663.
87. Caplin ME, Pavel M, Cwikla JB, et al. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*. 2014;371:224–233.
88. Vinik AI, Wolin EM, Liyanage N, et al. Evaluation of Lanreotide Depot/Autogel Efficacy and Safety as A Carcinoid Syndrome Treatment (ELECT): a randomized, double-blind, placebo-controlled trial. *Endocr Pract*. 2016;22:1068–1080.
89. Anthony L, Freda PU. From somatostatin to octreotide LAR: evolution of a somatostatin analogue. *Curr Med Res Opin*. 2009;25:2989–2999.
90. Blot K, Duchateau L, Lescrauwaet B, et al. A patient-reported outcomes analysis of lanreotide in the treatment of NETs patients with carcinoid syndrome: evidence from the ELECT trial. *Patient Relat Outcome Meas*. 2019;10:335–343.
91. Mirakhur B, Pavel ME, Pommier RF, et al. Biochemical responses in symptomatic and asymptomatic patients with neuroendocrine tumors: pooled analysis of 2 phase 3 trials. *Endocr Pract*. 2018;24:948–996.
92. Broder MS, Beenhouwer D, Strosberg JR, et al. Gastrointestinal Neuroendocrine tumors treated with high dose octreotide-LAR: a systematic literature review. *World J Gastroenterol*. 2015;21:1945–1955.
93. Ruszniewski P, Valle JW, Lombard-Bohas C, et al. Patient-reported outcomes with lanreotide autogel/depot for carcinoid syndrome: an international observational study. *Dig Liver Dis*. 2016;48:552–558.
94. Moertel CG, Hanley JA, Johnson LA. Streptozocin alone compared with streptozocin plus fluorouracil in the treatment of advanced islet cell carcinoma. *N Engl J Med*. 1980;303:1189–1194.

95. Moertel CG, Lefkopoulo M, Lipsitz S, et al. Streptozocin-doxorubicin, streptozocin-fluorouracil, or chlorozotocin in the treatment of advanced islet-cell carcinoma. *N Engl J Med*. 1992;326:519–523.
96. Kouvaraki MA, Ajani JA, Hoff P, et al. Fluorouracil, doxorubicin, and streptozocin in the treatment of patients with locally advanced and metastatic pancreatic endocrine carcinomas. *J Clin Oncol*. 2004;22:4762–4771.
97. Kunz PL, Catalano PJ, Nimeiri H, et al. A randomized study of temozolomide or temozolomide and capecitabine in patients with advanced pancreatic neuroendocrine tumors: a trial of the ECOG-ACRIN Cancer Research Group (E2211). *J Clin Oncol*. 2018;36:4004-4004.
98. Robelin P, Hadoux J, Forestier J, et al. Characterization, prognosis, and treatment of patients with metastatic lung carcinoid tumors. *J Thorac Oncol*. 2019;14:993–1002.
99. Raymond E, Dahan L, Raoul J, et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*. 2011;364:501–513.
100. Kulke MH, Lenz HJ, Meropol NJ, et al. Activity of sunitinib in patients with advanced neuroendocrine tumors. *J Clin Oncol*. 2008;26:3403–3410.
101. Pavel ME, Hainsworth JD, Baudin E, et al. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, phase 3 study. *Lancet*. 2011;378:2005–2012.
102. Yao JC, Shah MH, Ito T, et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*. 2011;364:514–523.
103. Bainbridge HE, Larbi E, Middleton G. Symptomatic control of neuroendocrine tumours with everolimus. *Horm Cancer*. 2015;6:254–259.
104. Kulke MH, Bergsland EK, Yao JC. Glycemic control in patients with insulinoma treated with everolimus. *N Engl J Med*. 2009;360:195–197.
105. Bernard V, Lombard-Bohas C, Taquet MC, et al. Efficacy of everolimus in patients with metastatic insulinoma and refractory hypoglycemia. *Eur J Endocrinol*. 2013;168:665–674.
106. Strosberg J, El-Haddad G, Wolin E, et al. Phase 3 trial of (177) Lu-DOTATATE for midgut neuroendocrine tumors. *N Engl J Med*. 2017;376:125–135.
107. Zandee WT, Brabander T, Blazevic A, et al. Symptomatic and radiological response to 177Lu-DOTATATE for the treatment of functioning pancreatic neuroendocrine tumors. *J Clin Endocrinol Metab*. 2019;104:1336–1344.
108. Hamiditabar M, Ali M, Roys J, et al. Peptide receptor radionuclide therapy with 177Lu-octreotate in patients with somatostatin receptor expressing neuroendocrine tumors: six years' assessment. *Clin Nucl Med*. 2017;42:436–443.
109. Lamers CBHW, Lind T, Moberg S, et al. Omeprazole in Zollinger-Ellison syndrome. *N Engl J Med*. 1984;310:758–761.
110. Blanchi A, Delchier JC, Soule JC, et al. Control of acute Zollinger-Ellison syndrome with intravenous omeprazole. *Lancet*. 1982:1223–1224.
111. Kulke MH, Horsch D, Caplin ME, et al. Telotristat ethyl, a tryptophan hydroxylase inhibitor for the treatment of carcinoid syndrome. *J Clin Oncol*. 2017;35:14–23.
112. Pavel M, Gross DJ, Benavent M, et al. Telotristat Ethyl in carcinoid syndrome: safety and efficacy in the telecast phase 3 trial. *Endocr Relat Cancer*. 2018;25:309–322.
113. Gelhorn HL, Kulke MH, O'Dorisio T, et al. Patient-reported symptom experiences in patients with carcinoid syndrome after participation in a study of telotristat etiprate: a qualitative interview approach. *Clin Ther*. 2016;38:759–768.
114. Singh S, Bergsland EK, Card CM, et al. Commonwealth Neuroendocrine Tumour Research Collaboration and the North American Neuroendocrine Tumor Society guidelines for the diagnosis and management of patients with lung neuroendocrine tumors: an international collaborative endorsement and update of the 2015 ENET expert consensus guidelines. *J Thorac Oncol*. 2020;15:1577–1598.

136

Ectopic Hormone Syndromes

David W. Ray

OUTLINE

History, 2291
Theoretical Considerations, 2291
  Theories of the Origin of Ectopic Hormones, 2292
  Hormones in Small-Cell Lung Carcinoma, 2292
Ectopic Adrenocorticotropic Hormone Syndrome, 2293
  The *POMC* Gene, 2293
  Regulation of *POMC* Gene Expression, 2293
  Ectopic Adrenocorticotropic Hormone Syndrome, 2293
  Pro-Opiomelanocortin Processing, 2293
  Dysregulation of *POMC* Gene Expression in Extrapituitary Tumors, 2294
  Diagnosis, 2294
  Treatment, 2294
Ectopic Corticotropin-Releasing Hormone Secretion, 2295
Syndrome of Inappropriate Antidiuretic Hormone Secretion, 2295
  Diagnosis, 2296
  Management, 2296
Humoral Hypercalcemia of Malignancy, 2297
  Parathyroid Hormone–Related Protein, 2297
  Hypercalcemia in Hematologic Malignancy, 2297
  Diagnosis, 2298
  Treatment, 2298
Oncogenic Osteomalacia, 2298
Ectopic Growth Hormone–Releasing Hormone, 2298
Non–Islet Cell Tumor Hypoglycemia, 2299
  Diagnosis, 2299
  Management, 2299
Other Pituitary Ectopic Hormones, 2300
  Gonadotropins, 2300
Ectopic Gut Hormone Syndromes, 2300
  Ectopic Renin Secretion, 2300

HISTORY

Tightly regulated production of hormones from specialized endocrine glands is usually under the control of higher centers, ultimately the brain, and is normally subject to negative feedback. This affords a mechanism for influencing diverse tissue function throughout the body. Apparently, inappropriate hormone production by nonendocrine tissue causes a spectrum of rare syndromes. These disorders are important because they may be the first manifestation of an underlying tumor, and the hormonal manifestations can induce morbidity and affect quality of life. Ectopic hormone secretion also sheds light on mechanisms of tissue-specific gene expression and regulation. Ectopic hormone production reveals important linkage between cell differentiation and gene expression. Ectopic expression of a peptide may also be useful as a tumor marker (e.g., human chorionic gonadotropin, α-fetoprotein).

Initial reports of endocrine activity associated with tumors appeared in the 1920s. These examples included features of parathyroid-like activity with disseminated carcinoma and adrenocortical activity in patients with lung and thymic tumors.[1,2] At the same time, initial reports of tumor-associated hypoglycemia were documented.[3,4] Later, based on the metabolic similarities to hyperparathyroidism, Albright and Reifenstein[5] proposed the secretion of parathyroid-like hormone activity in a patient with renal cell carcinoma as the cause of hypercalcemia.

Proof of tumor-derived endocrine factors had to await the development of accurate and sensitive hormone assays. These assays allowed identification of adrenocorticotropic hormone (ACTH), and antidiuretic hormone secretion from tumors causing ectopic ACTH syndrome and the syndrome of inappropriate antidiuretic hormone secretion (SIADH), respectively.[6–9] Since these breakthroughs, further advances in cell and molecular biology have sought to explain how and why such ectopic hormone production occurs and its consequences.

THEORETICAL CONSIDERATIONS

Circulating hormones are almost exclusively derived from specialized cells in endocrine glands. However, low-level hormone production or gene expression, at the very least, can be found much more widely. For example, the gene encoding ACTH is expressed at a high level in anterior pituitary corticotroph cells, but its expression has also been detected in a wide array of normal and tumoral extrapituitary tissues. This “ectopic” expression can give rise to the clinical syndrome of ectopic ACTH production, a cause of Cushing’s syndrome. However, it is also possible that POMC gene expression may have a role in healthy tissue, especially in the placenta, lymphocytes, testes, and lungs.

There must be compelling evidence that the hormone is produced by the tissue or tumor in question for the diagnosis of ectopic hormone production to be secure. Production is the most fundamental level of hormone control. Thus mRNA or the peptide should be demonstrated in tissue sections or hormone secretion can be shown to persist in vitro in primary cell culture or there may be an arteriovenous gradient of hormone across the tissue *in vivo*. Other criteria, such as reduction in hormone level postresection, are weaker because the tumor may express a hormone secretogogue (e.g., growth hormone–releasing hormone [GHRH] rather than growth hormone [GH] itself).

Hormones most frequently associated with ectopic production tend to be those with the most widespread distribution. The *POMC* gene is widely expressed in nontumorous, nonendocrine tissues and is frequently overexpressed in tumors, giving rise to the ectopic ACTH syndrome. Insulin is almost never produced ectopically by tumors, consistent with its restricted expression only in specialized β cells within the pancreatic islets.

Peptide hormones are often processed differently in tumors compared with their usual organs of production. They may undergo incomplete proteolytic cleavage or altered glycosylation, but the peptides themselves are the products of the same genes responsible for their production in their normal cells. Tumors that produce one hormone ectopically frequently express other hormones and proteins inappropriately; however, it is rare to have clinically significant secretion of more than one peptide.

Theories of the Origin of Ectopic Hormones

With the development of sophisticated molecular techniques, low levels of hormone gene expression have been documented in a wide range of normal tissues.[10] Therefore it is not surprising that tumors arising from these tissues can give rise to ectopic hormone-producing syndromes. If hormone production in the tumor remained under physiologic control, ectopic hormone production would not pose a clinical problem. Sometimes aberrant control has an obvious mechanism (e.g., lack of specific neural connection, physical distance from a portal system, absence of receptors for hormonal modulators of gene expression).

Several theories have been proposed to explain ectopic hormone production.

Amine Precursor Uptake and Decarboxylation Hypothesis[11,12]. A number of endocrinologically active cells are dispersed through normal tissues, but they are not usually considered to be endocrinologically active. These cells share amine precursor uptake and decarboxylation (APUD) properties. At one time, these cells were thought to derive from a common source in the embryonic neural crest. Although an appealing concept, evidence of this common origin is lacking. Nonetheless, their common features suggest that they may respond to transformation in a similar way. Many ectopic hormone-producing tumors have features of APUD cells, and these tumors may originate from APUD precursor cells. However, other ectopic hormone-producing tumors have no features of APUD tissue, and these tumors are often the most aggressive, secreting high levels of hormone, evidence that contradicts the importance of APUD characteristics for hormone production.

Dysdifferentiation Theory[13–15]. Problems with the previous theories have led to the dysdifferentiation hypothesis. It proposes that neoplastic changes occur in progenitor cells rather than in terminally differentiated cells. As a result of one or more mutations, the transformed progenitor cell undergoes variable differentiation. If this theory holds true, it predicts a tumor with a mixed population of cells at different stages of development. Depending on the blocks to normal differentiation, a subset of cells may develop into an endocrine cell type and give rise to an endocrine-type tumor. This model explains why some tissues produce particular hormones because the tumor would tend to transcribe the same genes as its parent tissue, although these may be expressed aberrantly (Fig. 136.1).

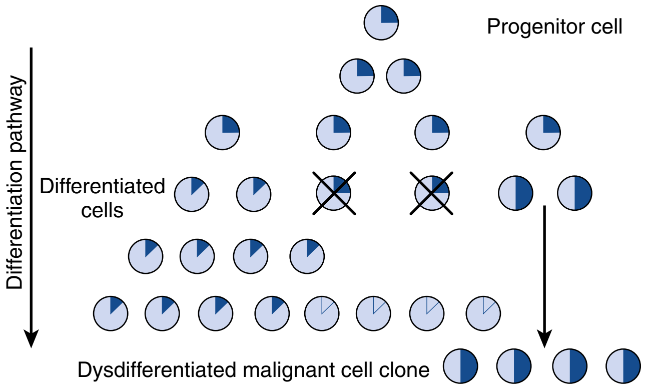


Figure 136.1 Dysdifferentiation results in aberrant gene expression. The malignant transformation of a progenitor cell results in a clonal population of cells that have undergone atypical differentiation. These cells share some features with their progenitor cells *(shaded areas)* and may continue to express genes associated with these immature, incompletely differentiated cells. Malignant transformation of a cell can lead to persistent, or enhanced, hormone gene expression in the clone of proliferating cells *(heavily shaded area).*

Extrapituitary expression of the ACTH gene (pro-opiomelanocortin [*POMC*]) is described in detail in the following section. Ectopic ACTH production shares many features of other ectopic hormone syndromes and may provide a paradigm for understanding general mechanisms of ectopic hormone production.

Hormones in Small-Cell Lung Carcinoma

Small-cell lung carcinoma (SCLC) is the most common and aggressive neuroendocrine tumor. Patients with SCLC seldom present to the endocrinologist, but because there is evidence of hormone expression, it is a useful model from which to extrapolate to other, less common, syndromes.

The normal adult lung contains a diffuse population of cells that synthesize very low levels of peptide hormones including ACTH, GHRH, and gastrin-releasing hormone. In fetal life, these cells are more densely represented and produce higher concentrations of hormones, suggesting a possible role in differentiation. These scattered endocrine cells have APUD properties and comprise part of a diffuse neuroendocrine system.

SCLC, a bronchogenic tumor, produces most of the hormones found in normal bronchial epithelium. ACTH, vasopressin, and calcitonin secretion have been studied most extensively because of the clinical syndromes that they cause and because they produce potential disease markers that can be used for diagnosis or response to treatment. In a study of 157 primary lung tumor extracts, 83% expressed at least one peptide hormone. Multiple hormone production is common and includes ectopic ACTH syndrome in 20% of cases, although production of one hormone tends to dominate the clinical picture.[16,17] However, only approximately 1% of patients with SCLC have clinical features of cortisol excess, possibly because of the short duration of disease.[18] Peptide hormones often undergo incomplete processing. In particular, POMC is not efficiently cleaved to ACTH, resulting in high–molecular-weight forms of ACTH in tumor extracts and in the circulation.[16,19] It is now clear that these alternate forms represent the prohormones POMC and pro-ACTH. The abnormal protein processing may result from a lack of specific cleavage enzymes, prohormone convertases 1 and 2, which are expressed only in specialized endocrine tissue or from a switch from the regulated secretory pathway to the constitutive one.[20]

KEY POINTS

- Ectopic hormone production likely results from abnormal differentiation of a stem cell population lying within the organ.

ECTOPIC ADRENOCORTICOTROPIC HORMONE SYNDROME

The *POMC* Gene

Pituitary corticotroph cells are the only cells that express the *POMC* gene at a high level. The human *POMC* gene is encoded in three exons on chromosome 2. The first exon is noncoding. The second exon contains the signal peptide, which targets the protein product to the regulated secretion pathway. The third exon encodes the majority of the mature protein, including ACTH.[21] The mature mRNA from the *POMC* gene is 1200 nucleotides. In addition, a short form of the mRNA has been found at low levels in most tissues analyzed. This arises from a transcription start site 5′ to exon 3 and thus only includes the coding sequence for exon 3.[22] Therefore this transcript lacks a signal peptide and does not give rise to the mature POMC molecule. There is no evidence that this transcript produces a peptide product, and its physiologic role is unclear. A third POMC transcript has also been described that is longer than the pituitary form (~1500 nucleotides). It arises from a site, or multiple sites, within the 5′ flanking region of the human *POMC* promoter.[23,24] This mRNA species includes the entire coding region of the peptide and does give rise to a secreted peptide product. This "long" form of the POMC mRNA is particularly found in extrapituitary tissues and tumors.

Regulation of *POMC* Gene Expression

Expression of the *POMC* gene appears to be predominantly controlled at the level of gene transcription.[25] The rat *POMC* gene has been most extensively studied, and pituitary expression is conferred by the 5′ flanking region of the gene. It has recently been found that pituitary corticotroph expression of POMC requires the action of a tightly restricted transcription factor, a member of the T-box family, termed Tpit.[26] This factor acts with the homeodomain protein PitX1 and promotes recruitment of SRC family coactivators to the *POMC* promoter, leading to enhanced gene transcription.[27]

Corticotropin-releasing hormone (CRH) acts on pituitary corticotroph cells to increase cyclic adenosine monophosphate (cAMP) accumulation and activates mitogen-activated protein kinases. There is also evidence of activation of the orphan nuclear receptor nerve growth factor–induced clone B (or Nur 77)[28–30] leading to enhanced *POMC* transcription through the recruitment of SRC coactivators to nerve growth factor–induced clone B.[27,31] As nerve growth factor–induced clone B and Tpit act synergistically, this suggests the formation of a regulatory complex on the *POMC* promoter with Tpit, NGFI-B, and SRC coactivators.[27] It is important that expression of Tpit promotes corticotroph cell differentiation and that its expression is more limited than that of POMC. Therefore there is no Tpit expression in hypothalamic POMC-expressing neurons, suggesting that Tpit is specific for corticotroph-specific expression of POMC; other mechanisms are responsible for expression elsewhere. Tpit expression has been found specifically in human pituitary corticotroph adenomas.[26]

Glucocorticoids repress transcription of the *POMC* gene by binding to two DNA elements in the 5′ flanking region of the promoter. The more proximal element, an imperfect palindrome 63 nucleotides upstream from the transcription start site, is thought to bind three glucocorticoid receptor molecules in an unusual trimer formation.[32–34] This conformation of receptors on DNA directs repression of transcription rather than enhancement. Further upstream, between −480 and −320, there is another glucocorticoid-regulated element, suggesting that these two DNA elements interact to achieve the full effect of glucocorticoid repression.[35] It is interesting that Tpit expression, essential to the corticotroph cell type and to POMC expression, is not affected by glucocorticoids, in contrast to POMC, which is repressed.[36] However, it now appears clear that the activated glucocorticoid receptor can antagonize the actions of Nur 77, and so by removing a positive transactivator lead to *POMC* gene repression.[37] The mechanism of glucocorticoid receptor interaction with Nur 77 requires the chromatin remodeling protein Brg1, which stabilizes their binding. Furthermore, Brg1 assists recruitment of a further chromatin remodeling enzyme HDAC2 to the glucocorticoid receptor bound to the *POMC* gene. The pathophysiological importance of Brg1 and HDAC2 is illustrated by finding deficient expression of one or both in 50% of glucocorticoid-resistant pituitary corticotroph adenomas.[38] The expression of Brg1 and HDAC2 in ectopic ACTH syndrome has not been explored.

A number of other hypothalamic factors act on the pituitary corticotroph to influence POMC expression. However, their modes of action are not well defined. In particular, arginine vasopressin stimulates POMC expression rather weakly but augments CRH action. The intracellular pathways activated by arginine vasopressin appear to be protein kinase C dependent, but arginine vasopressin also potentiates the action of CRH on cAMP generation.[25,39]

However, many other peptide growth factors and cytokines are capable of activating cAMP, mitogen-activated protein kinase, and Janus kinase/signal transducers and activators of transcription signaling cascades and thus are potentially capable of regulating POMC expression in nonpituitary tissue. Although extrapituitary tissues lack expression of corticotroph-specific transcription factors, activation of common signaling cascades might be expected to result in *POMC* gene expression. In extrapituitary tissues, the *POMC* gene may be modified to render it transcriptionally silent. One such irreversible modification is DNA methylation. The loss of methylation in tumor tissue may allow transcription of the gene to be activated by the common signaling pathways described previously. There is some evidence that such changes in DNA methylation do occur in cell line models of ectopic ACTH syndrome.[40,41] It seems likely that POMC expression per cell is less in most extrapituitary tumors compared with the pituitary corticotroph, but this relative inefficiency of expression is compensated for by the greater number of cells expressing the gene in extrapituitary tumors.

Ectopic Adrenocorticotropic Hormone Syndrome

ACTH immunoreactivity has been recognized to show size heterogeneity for many years, with the presence of high–molecular-weight forms detected in human plasma.[42,43] Ectopic ACTH syndrome was the first of the ectopic hormone syndromes to be recognized. In its most florid form, it is rare, in one study affecting 4.5% of patients with SCLC, although there is evidence of derangement in the hypothalamic-pituitary-adrenal axis in the majority of patients with SCLC[42,44] (Table 136.1). Analysis of tumor tissue surprisingly suggests the presence of immunoreactive ACTH, even in the absence of clinical features of hormone excess. ACTH is present predominantly in a high–molecular-weight form of approximately 20 kD, but this purified material can be cleaved to mature ACTH (4.5 kD) by the action of trypsin. Further work identified the presence of immunoreactive ACTH-like peptide in a variety of normal tissues, suggesting that extrapituitary ACTH expression is less ectopic than it is inappropriately regulated. The ACTH immunoreactivity was found to have no biologic activity and was assumed to be "big" ACTH. However, identification of predominantly high–molecular-weight forms of ACTH in the circulation of patients with clinically apparent Cushing's syndrome suggests that the precursors of ACTH may have some activity at the ACTH receptor.

Pro-Opiomelanocortin Processing

The *POMC* gene leads to the generation of a pre-pro hormone, POMC. This protein undergoes a series of proteolytic cleavages at dibasic

TABLE 136.1 Types of Tumors Causing Ectopic Adrenocorticotropic Hormone Secretion

Small-cell lung carcinoma
Carcinoid tumors (bronchus, thymus, small intestine)
Pancreatic islet cell tumor
Pheochromocytoma
Medullary carcinoma of the thyroid
Carcinomas (breast, gastrointestinal tract [esophageal, gastric, colorectal], ovarian, cervical, prostate)

amino acid residues to give rise to a series of small molecules, including ACTH, melanocyte-stimulating hormone, and β-endorphin.[45–47] In the anterior pituitary, ACTH is cleaved by the action of a specific protease, termed PC1 (for prohormone convertase type 1).[48] In the rodent intermediate lobe melanotroph, the POMC molecule undergoes more comprehensive digestion to give smaller fragments, melanocyte-stimulating hormone, β-endorphin, and corticotropin-like intermediate lobe peptide as a result of cleavage by prohormone convertase 2. Expression of prohormone convertase 2 and thus detection of circulating ACTH fragments have been described in ectopic ACTH syndrome tumors.[49]

In the majority of extrapituitary tumors that cause ectopic ACTH syndrome, processing of the preprohormone is incomplete. Therefore ectopic ACTH syndrome is characterized by high–molecular-weight forms of ACTH in the circulation.[17,19].It is likely that the extent of processing correlates with the degree of neuroendocrine differentiation of the tumor, and hormonal manifestations are probably only seen in tumors with significant hormone-processing capacity. A number of small, highly differentiated, slow-growing tumors (typically bronchial carcinoid) have been shown to process POMC in the neurointermediate lobe manner, giving rise to small fragments in the circulation, such as corticotropin-like intermediate lobe peptide and α-melanocyte-stimulating hormone. These have been used to aid diagnosis in some cases of Cushing's syndrome, although the series are small.[8,50]

Dysregulation of *POMC* Gene Expression in Extrapituitary Tumors

In contrast to *POMC* gene expression in pituitary corticotroph cells, expression in extrapituitary tumors is characteristically resistant to glucocorticoid.[7,51] This is the basis of the high-dose glucocorticoid suppression test used to distinguish eutopic from ectopic sources of ACTH in Cushing's syndrome. As the test has approximately 10% false-positive and 10% false-negative rates, it has largely been superseded by sophisticated imaging and inferior petrosal sinus sampling for differential diagnosis.[52] With the availability of recombinant CRH, responses of extrapituitary tumors to this peptide have been measured. In general, only pituitary corticotrophs stimulate POMC expression in response to CRH, but exceptions are increasingly being identified.[53]

As pituitary expression of POMC was defined with the aid of a cell line model, so a cell line model was sought for extrapituitary expression. To this end, a panel of human SCLC cell lines was established. These cell lines express the *POMC* gene, and expression is resistant to glucocorticoid suppression.[16,54–58] Further, the cell lines secrete predominantly unprocessed POMC and partially processed forms, again reflecting the pattern of activity characterized *in vivo*.[54]

It is intriguing that the majority of extrapituitary tumors are resistant to glucocorticoid inhibition of POMC expression. Receptors for glucocorticoids are present in most cells, including malignant cells, so exploring the mechanisms of glucocorticoid resistance was important. Using the panel of cell lines, expression of glucocorticoid receptor was identified using both Western blot with a polyclonal anti–glucocorticoid receptor antibody, and ligand-binding assays using tritiated dexamethasone.[56–58] To determine whether the receptors were sufficient for glucocorticoid signaling, a synthetic, glucocorticoid-responsive gene was used. This gene was transfected into the cells, and the effects of glucocorticoid incubation on expression of the reporter measured. In contrast to the brisk induction of expression seen in control pituitary cells, none of the human SCLC cells responded to either natural or synthetic glucocorticoids.[56,58] Thus resistance of the POMC gene to glucocorticoid is only part of a global resistance of malignant cells to glucocorticoid action. High concentrations of wild-type receptor in the cells were found to be sufficient to restore glucocorticoid signaling, thereby suggesting that resistance resides at the level of the endogenous receptor.[56] Because one of the actions of glucocorticoids on pituitary corticotrophs is to inhibit proliferation and in the developing lung glucocorticoids act to promote differentiation, it is possible that evasion of glucocorticoid signaling confers a survival advantage on the malignant cells. Indeed, it has recently been shown that overexpressing the wild-type glucocorticoid receptor in human SCLC cells powerfully induces apoptosis. Intriguingly, this effect occurs even in the absence of added glucocorticoid.[59] As yet, a single, unifying molecular mechanism of glucocorticoid resistance in the human SCLC cell lines remains to be defined.

It is interesting that well-differentiated carcinoid tumors causing ectopic ACTH syndrome sometimes show appropriate POMC suppression to supraphysiologic glucocorticoid levels, as in pituitary-dependent Cushing's disease, but these tumors express high levels of the glucocorticoid receptor.[60]

Diagnosis

The diagnosis of Cushing's syndrome and the differential diagnosis of ACTH-dependent Cushing's syndrome are described elsewhere (see Chapter 7). Dynamic endocrine testing is required to diagnose Cushing's syndrome, and detection of ACTH using a sensitive two-site immunoradiometric assay is useful for making the diagnosis of ACTH-dependent Cushing's syndrome. A variety of dynamic endocrine and imaging protocols is used to identify a pituitary or extrapituitary source of the ACTH excess (Tables 136.2 and 136.3). These all have variable sensitivity and specificity, but combined tests afford near 100% diagnostic accuracy.[61]

Most occult tumors are carcinoid, pheochromocytoma, or medullary thyroid carcinoma and originate in the neck, chest, or abdomen. Computed tomography or magnetic resonance imaging can be used to detect chest tumors in patients with a normal chest radiograph. There has been some success in using indium-labeled octreotide scanning to identify occult neuroendocrine tumors, although its diagnostic performance is poor, tending only to confirm the presence of tumors found with conventional imaging.[61]

Treatment

Treatment is based on two objectives: (1) to control endocrine manifestations and (2) to manage the underlying tumor. Individual patients will present with different priorities. The ideal treatment is curative resection of the primary tumor, achieving both objectives. If this is not possible, patients with small, occult primary tumors may be managed by chemical or surgical adrenalectomy; in most cases, the primary tumor is not life-threatening. Patients with extensive carcinoma, for example, small-cell carcinoma where ACTH excess coexists, may be best managed by chemotherapy, which indirectly reduces ACTH expression. Chemotherapy should be tailored for the cell type and tumor stage, regardless of hormone excess.

TABLE 136.2 Tumors Associated With Ectopic Corticotropin-Releasing Hormone Secretion

Pancreatic tumors
Small-cell lung carcinoma
Prostate carcinoma
Hypothalamic gangliocytoma
Medullary thyroid carcinoma
Bronchial carcinoid

TABLE 136.3 Diagnostic Criteria for the Syndrome of Inappropriate Antidiuretic Hormone Secretion

Hyponatremia
Inappropriately increased urine osmolality (>100 mOsm/kg)
Persistent sodium excretion in urine (>20 mmol/L)
Normal renal, thyroid, and adrenal function
No hypovolemia, edema, or diuretic use

In some patients with florid Cushing's syndrome, control of cortisol production may help in preparation for surgery. In these cases, treatment with metyrapone, mitotane, or ketoconazole, in combination or individually, is helpful.[62] Side effects are common with these agents in the doses required, and so combinations (e.g., ketoconazole and metyrapone) may allow more effective control. Rapid control of hypercortisolemia may require intravenous etomidate.[61] There are case reports of good responses to long-acting somatostatin analogues[63] (Table 136.4).

KEY POINTS

- Ectopic ACTH syndrome is a frequent cause of Cushing's syndrome, can be hard to detect, but must be excluded before assuming a pituitary cause.
- Treatment is of the underlying tumor, which can be very small and hard to detect. Adrenalectomy may be necessary to save the patient from Cushing's.

ECTOPIC CORTICOTROPIN-RELEASING HORMONE SECRETION

It has been more than 30 years since the original description of ectopic CRH secretion, and it is now clear that true, isolated secretion of CRH is very rare (Table 136.5). In tumors secreting ACTH peptides, there are frequent reports of CRH immunoreactivity, which may play a paracrine role in the development of the hormone syndrome, but such a role has not been defined.[64,65] CRH is expressed outside the central nervous system, particularly in sites of inflammation, and may subserve other roles, including vasodilatation. CRH is seldom measured in the peripheral blood during the workup of Cushing's syndrome or in inflammatory disease, and so evidence is lacking for a true endocrine role of this hormone in the peripheral circulation.

The clinical features are typical of Cushing's syndrome, and the hormonal features may resemble either pituitary secretion of ACTH (if the ectopic source secretes purely CRH) or ectopic ACTH syndrome (if the tumor cosecretes ACTH-related peptides). Measurement of CRH is probably best left to those cases with definitive pituitary ACTH production and confirmed corticotroph hyperplasia by histology.

TABLE 136.4 Tumors Associated With Syndrome of Inappropriate Antidiuretic Hormone Secretion

Tumor
Small-cell lung carcinoma
Pancreas
Duodenum
Urethra
Prostate
Bladder
Lymphoma and other hematologic malignancies

TABLE 136.5 Clinical Features of Hyponatremia

Headache
Lethargy
Weakness
Nausea/vomiting
Mood swings
Confusion
Drowsiness
Hyporeflexia
Positive Babinski's sign
Convulsions
Coma

SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE SECRETION

SIADH is the most common cause of hyponatremia and one of the most frequent hormone syndromes associated with malignant disease (Table 136.6). It may be caused by a wide range of underlying disorders that can be categorized into four broad groups: malignancies, neurologic disorders, and lung disease; and drugs. These conditions cause hyponatremia as a result of abnormal hypothalamic vasopressin, when secretion comes under aberrant control from either neuronal inputs or circulating humoral factors.

The *vasopressin* gene is expressed in a number of separate neuronal nuclei and peripheral tissues. Regulation of vasopressin expression is dependent on the site. For example, hyperosmolality increases vasopressin expression in the supraoptic nucleus and the magnocellular division of the paraventricular nucleus, but vasopressin mRNA in other sites is unaltered. For example, vasopressin expression in the suprachiasmatic nucleus is under diurnal regulation. Androgens upregulate expression of vasopressin in the striae terminalis, and glucocorticoids suppress expression in the parvocellular division of the paraventricular nucleus. Differential regulation, even within anatomically related sites, likely results from differential expression of hormone receptors in the cells and different neuronal afferents. Vasopressin gene transcription is under positive regulation by cAMP and protein kinase C pathways. Little is known about regulation of vasopressin outside the central nervous system, but glucocorticoids have been shown to suppress its expression in an SCLC cell line.

Ectopic secretion of vasopressin occurs in squamous cell carcinoma; small-cell carcinoma; neuroblastoma; pancreatic, duodenal, prostatic, and urothelial tumors; and undifferentiated carcinoma[65–68] (Table 136.7). In one series, 16% of patients with SCLC had hyponatremia (<130 mmol/L) at diagnosis compared with 0% of patients with

TABLE 136.6 Mechanisms of Malignancy-Associated Hypercalcemia

Mechanism	Agent	Tumor Type
Lytic metastases	TGF-β	Squamous cell carcinoma of the lung
	IL-1	Breast
	TNF	Kidney
	Lymphotoxin	Myeloma
	PTHrP	
Humoral effects	PTHrP	Solid tumors, particularly squamous cell carcinoma of the skin, lung, kidney, head, and neck
	PGE	Solid tumors
	TNF	Multiple myeloma
	TGF-β	
	IL-1	
	Lymphotoxin	
	1,25-Dihydroxyvitamin D	T-cell lymphoma
		Non-Hodgkin's lymphoma
		Hodgkin's lymphoma
		Melanoma
		Small-cell lung carcinoma
	Ectopic PTH	Small-cell lung carcinoma (very rare)
Coexistent other causes of hypercalcemia	Primary hyperparathyroidism	Ovarian cancer
	Sarcoidosis	
	Vitamin D–mediated	

1,25-DHCC, 1,25-Dihydroxycholecalciferol; *IL-1,* interleukin 1; *PGE,* prostaglandin E; *PTH,* parathyroid hormone; *PTHrP,* parathyroid hormone–related protein; *TGF-β,* transforming growth factor β; *TNF,* tumor necrosis factor.

TABLE 136.7 Tumors That Produce Ectopic Calcitonin

- Pheochromocytomas
- Pancreatic neuroendocrine tumors
- Adrenocortical carcinoma
- Neuroendocrine tumor of esophagus
- Acute leukemia
- Lung cancer (27% small-cell lung carcinoma or adenocarcinoma)
- Cervical carcinoma
- Prostatic carcinoma
- Breast cancer
- Renal carcinoma
- Gastrointestinal tract tumors

TABLE 136.8 Symptoms and Signs of Hypercalcemia

- Polyuria
- Thirst
- Nausea
- Anorexia
- Constipation
- Confusion
- Drowsiness
- Headache
- Coma

non-SCLC. Hyponatremia was found to be an independent predictor of poor prognosis in extensive stage disease. *In vitro* studies found that 7 of 11 tumors in culture produced vasopressin, 9 of 11 tumors produced atrial natriuretic factor, and 5 of 11 tumors produced both hormones. All the cells studied from patients with hyponatremia produced one of the two hormones.[65]

The active hormone, vasopressin, is the product of a precursor peptide cleavage, which also gives rise to neurophysin II, and a C-terminal glycopeptide. Similar to the identification of partially processed forms of ACTH in the circulation in ectopic ACTH syndrome, the vasopressin-neurophysin precursor has been found in plasma from patients with SIADH due to SCLC.[69] This is in contrast to patients with SIADH caused by central nervous system disease. Differential hormone processing may provide an additional diagnostic test for the underlying cause of SIADH.

Diagnosis

Hyponatremia presents with features of neuropsychiatric dysfunction in most cases (Table 136.8). The elderly and the young are more likely than others to be symptomatic. The absolute sodium concentration is less reliable as a predictor of symptoms than the rate of decrease in sodium concentration, although almost all symptomatic patients will have plasma sodium concentration of less than 120 mmol/L. Clinical features include lethargy, fatigue, impaired conscious level, coma, seizures, and psychosis. Hyponatremia may cause death as a result of cerebral edema, uncontrolled seizures, and the consequences of coma. Although mild hyponatremia (>125 mmol/L) is usually regarded as a straightforward condition, which may not require specific treatment, hyponatremia should not be regarded as benign.

A set of diagnostic criteria must be fulfilled before a secure diagnosis is reached (see Table 136.8). The underlying cause is then sought. Neurologic, lung-related, drug-related, and miscellaneous causes can result in dysregulation of vasopressin regulation in the hypothalamus and should not, therefore, be regarded as ectopic hormone secretion states. In contrast, a variety of tumors (see Table 136.7) has been shown to aberrantly secrete vasopressin and express the vasopressin gene inappropriately. There is evidence from T_1-weighted magnetic resonance imaging scans of the pituitary that such ectopic vasopressin secretion results in central suppression of vasopressin synthesis.

Management

The management of this disorder falls into two parts. The first is to diagnose and treat the underlying cause, and the second is to remove excess, free body water. Discussion of specific therapy on the variety of underlying tumors is beyond the scope of this chapter, but surgical cure or debulking, chemotherapy, and radiotherapy have all been applied. In general, the circulating vasopressin concentration bears a direct relationship to tumor bulk in an individual patient, but there is a low correlation across a patient cohort, presumably reflecting intertumoral differences in cellular differentiation and hormone production. Decisions about the acute correction of hyponatremia are complicated by the occurrence of both pontine and extrapontine myelinolysis as consequences of therapy. The risk of myelinolysis is linked to the rate of change in sodium concentration. Therefore a prudent approach is always justified, with an increase of between 0.5 and 1.0 mmol/L sodium per hour and a maximum of 8 mmol/L over 24 hours. This requires close monitoring every 2 to 3 hours[70]. In symptomatic patients, treat with furosemide and hypertonic saline until convulsions cease and the level of consciousness improves. This is usually achieved by an increase in sodium concentration of 10% (~10 mmol/L) and subsequent water restriction. In asymptomatic patients, the condition is almost always chronic; these patients should be treated by water deprivation initially.

Specific approaches to antagonize the action of vasopressin usually rely on demeclocycline, in divided doses up to 1200 mg daily. These agents induce a state of nephrogenic diabetes insipidus. Alternatively, oral sodium supplementation, up to 3 g daily, with furosemide, 40 to 80 mg, results in net loss of free water. More recently, a vasopressin receptor antagonist (Vaprisol, Astellar Pharma) has been approved. The indications for intravenous use are currently limited to moderate to severe hyponatremia which would warrant the use of hypertonic saline.[70] In the future chronic treatment with orally available agents seems likely, although the precise indications remain undefined.

KEY POINTS

- Hyponatremia is common in hospital inpatients and is associated with poor outcomes. Apply simple diagnostic tests to determine the cause, and aim for gentle correction. The outlook depends on the underlying cause

HUMORAL HYPERCALCEMIA OF MALIGNANCY

Hypercalcemia is a common complication of malignancy (Table 136.9). It may result from the direct lytic effect of bony metastases or the action of tumor-derived humoral factors, although it is now clear that a spectrum of disorders lies between these two extremes, and, in most cases, there is a humoral component. Calcitonin has been detected in peripheral blood in a number of patients with malignancy and appears to be incompletely processed, in a manner similar to ACTH, with higher–molecular-weight variants.[71] However, there is no clinical syndrome ascribed to such ectopic production. Calcitonin secretion appears to be most commonly associated with a multihormonal secretory phenotype, with other peptide hormones including ACTH and gastrin[55,71–74] (Table 136.10).

Parathyroid Hormone–Related Protein

The humoral syndrome has been explained by the isolation and characterization of a peptide hormone, parathyroid hormone–related protein (PTHrP). This hormone is closely related in amino acid sequence to parathyroid hormone (PTH) in its N-terminal region (amino acids 1–34), but after residue 34, the two peptides have unique sequences. The discovery of PTHrP as the circulating mediator of hypercalcemia in malignancy ruled out earlier theories that ectopic production of PTH was the cause. There have been isolated reports of ectopic PTH production by tumors, but these are rare. There is convincing evidence of PTH production in SCLC,[75] an ovarian tumor,[76] thymoma,[77] and neuroectodermal tumor.[78] In most cases, patients with malignancy and hypercalcemia with increased PTH concentrations will have coincident primary hyperparathyroidism rather than ectopic PTH secretion. As many as 70% of cases of hypercalcemia in malignancy are caused by excess PTHrP secretion.[79–84]

PTHrP is seldom detectable in the circulation of normal subjects, but its expression has been shown in a number of normal tissues. PTHrP may be regarded as a product of the diffuse paracrine system and may have evolved to perform quite different physiologic roles compared with the structurally related PTH. Under these circumstances, it is hard to describe production of PTHrP from a tumor arising from any tissue as truly ectopic because no definite eutopic source for the peptide has been defined. However, humoral hypercalcemia of malignancy is most conveniently considered with the group of ectopic hormone secretion syndromes.

TABLE 136.9 Tumors Associated With Growth Hormone–Releasing Hormone Secretion

Carcinoid tumors (e.g., bronchial)	Pheochromocytomas
Pancreatic islet tumors	Adrenal adenomas
Small-cell lung carcinoma	Hypothalamic gangliocytomas

TABLE 136.10 Tumors Associated With Non–Islet Cell Tumor Hypoglycemia

Carcinoma	Mesenchymal
Hepatocellular, hepatoma	Fibroma, fibrosarcoma
Adrenocortical	Mesothelioma
Pancreatic	Rhabdomyosarcoma
Gastric	Neurofibroma, neurofibrosarcoma
Colon	Leiomyosarcoma
Lung (small cell, squamous)	Others
Kidney	Hemangiopericytoma
Prostate	Hematologic
	Lymphoma

PTH and PTHrP share a common receptor that they recognize through their homologous N-terminal sequences. The PTH/PTHrP receptor mediates the action of both peptides in bone and kidney and is a member of the G-protein–coupled seven-transmembrane receptor family. The common receptor explains how PTHrP is able to generate cAMP in membrane preparations of PTH-sensitive renal tubule and, further, why the humoral hypercalcemia of malignancy syndrome resulted in hypercalcemia with hypophosphatemia. In addition, there are receptors that specifically recognize PTHrP, and these are presumed to be the natural targets for the peptide within the tissues in which it is produced.[85–86] In the past, there was controversy about 1,25-dihydroxyvitamin D levels in primary hyperparathyroidism versus humoral hypercalcemia of malignancy syndrome. PTHrP and 1,25-dihydroxyvitamin D appear to be loosely correlated in humoral hypercalcemia of malignancy syndrome, suggesting that PTHrP shares with PTH the capacity to induce 1α-hydroxylase. Occasionally, reported discrepancies may stem from the action of other tumor-derived, circulating factors, from the metabolic consequences of malignancy, or from some other mechanism related to hypercalcemia.[87,88]

PTHrP may also play a role in hypercalcemia related to osseous metastases in that the hypercalcemia associated with bony metastases has a significant humoral component. Furthermore, expression of PTHrP by primary tumors is a predictor of bony, metastatic disease development. This suggests that the local production of PTHrP by bone micrometastases may facilitate bony invasion and destruction.[89–92] PTHrP expression is under complex regulation. A number of cytokines and growth factors have been shown to induce its expression, and both glucocorticoids and 1,25-dihydroxycholecalciferol, which have an antiproliferative effect on epithelial cells, inhibit expression.[93–96]

Hypercalcemia in Hematologic Malignancy

Hypercalcemia occurs in as many as 30% of patients with multiple myeloma. Skeletal involvement causes extensive bone destruction with pain and risk of pathologic fracture. Histologic evidence suggests that the bone disease is caused by increased osteoclastic activity in the absence of significant osteoblastic activity. Loss of osteoblastic activity is also supported by the characteristically negative bone scan and suppressed circulating osteocalcin concentration. A number of cytokines, produced by activated immune cells, have been shown to have direct effects promoting bone resorption. Such cytokines include tumor necrosis factor α, tumor growth factor β, interleukin-1, and leukemia inhibitory factor. A superseded generic term for these factors was osteoclast-activating factor. However, it is noteworthy that in three of nine patients with multiple myeloma complicated by hypercalcemia, there was an elevation in circulating PTHrP, suggesting that a mechanism similar to humoral hypercalcemia of malignancy syndrome may be operating in at least some patients with hematologic malignancy-associated hypercalcemia.[83,97,98]

Generally, hypercalcemia is rare in lymphoma, with the exception of adult T-cell leukemia/lymphoma. This disease occurs in Japan and the West Indies and is caused by infection with the human T-cell lymphotrophic virus type I. At least one-fourth of patients will develop hypercalcemia, which is associated with suppressed 1,25-dihydroxyvitamin D. Hypercalcemia predicts outcome and is implicated in causing patient mortality. There is strong evidence that the hypercalcemia is mediated by PTHrP.[99–102]

Diagnosis

Hypercalcemia is the most common metabolic complication of malignant disease and the cause of much morbidity (Table 136.11). Most cases are due to humoral mechanisms, principally PTHrP, rather than direct damage of bone by malignant cells. This is clear from the observation that even patients who have bone metastases and hypercalcemia have a poor correlation between extent of skeletal involvement and calcium concentration in the circulation.

The presentation of hypercalcemia may be confusing and may be attributed to the underlying disease process itself. Patients may be nonspecifically unwell and may complain of constipation, nausea, vomiting, confusion, or dehydration. Hypercalcemia induces a diuresis and may cause profound dehydration, particularly in association with vomiting or drowsiness. Further manifestations include lethargy, depression, poor concentration, and drowsiness.

It is sometimes necessary to distinguish malignancy-induced hypercalcemia from other causes of hypercalcemia, such as hyperparathyroidism. PTH levels are low in hypercalcemia of malignancy. An elevated PTHrP level confirms the diagnosis, and it is increased in approximately 80% of patients with hypercalcemia with cancer. 1,25-Dihydroxyvitamin D levels may be increased in patients with lymphoma. Recognition of hypercalcemia and its appropriate treatment allow retrospective allocation of symptoms in an individual patient. Therefore it is almost always worth treating such patients, irrespective of their underlying disease prognosis.

Treatment

There is little evidence that hypercalcemia is a significant cause of premature mortality in cancer, but it is a significant cause of morbidity.[103] Even if the underlying malignancy is beyond cure, effective relief of hypercalcemia is a most useful palliative intervention. There are few patients in whom treatment should not be seriously considered because intervention is quick, easy, and largely free of serious complications. In the past, calcitonin and/or mithramycin were used, but these have been replaced by the bisphosphonate drugs, in conjunction with rehydration with intravenous saline.[104] The saline promotes a calcium diuresis. Usually, single infusions of pamidronate or clodronate are sufficient, with a suggestion that pamidronate is slightly more effective.[105,106] There are newer bisphosphonates that have become available (e.g., ibandronate), but further head-to-head comparisons are lacking.[107] These drugs are usually administered by slow intravenous infusion. There may be a transient febrile reaction within the first 24 hours, but this is usually self-limiting. The calcium response is typically rapid and may last as long as 1 month. In patients in whom a more rapid acute effect is desired, bisphosphonates can be combined with calcitonin.[108]

The bisphosphonates are rapidly cleared from the circulation and concentrated in the bone. They appear to inhibit osteoclast activity and may induce osteoclast apoptosis. Their duration of action is significantly longer than predicted by their plasma half-life, reflecting their distribution and mode of action.[107]

TABLE 136.11 Non–Islet Cell Tumor Hypoglycemia vs Insulinoma

	Non–Islet Cell	Insulinoma
IGF-1	↔	↔
IGF-2	↑	↔
IGFBP-3	↓	↔
Insulin	↓	↑
Glucose	↓	↓
Growth hormone	↓	↔ or ↑
β-Hydroxybutyrate	↓	↓

↑, increased; ↓, decreased; ↔, equivocal. *IGF,* Insulin-like growth factor; *IGFBP,* insulin-like growth factor–binding protein.

KEY POINTS

- Humoral hypercalcemia of malignancy is commonly seen, and simple tests can be used to confirm the diagnosis, the main differential being coincident primary hyperparathyroidism. Bisphosphonates are effects agents.

ONCOGENIC OSTEOMALACIA

Osteomalacia associated with malignancy was first described in 1947.[109] It is characterized by hypophosphatemia, normal or low calcium, elevated alkaline phosphatase, and suppressed 1,25-dihydroxyvitamin D. It is usually found with benign mesenchymal tumors, of which hemangiopericytomas are the most common. The disorder exhibits biochemical features similar to those seen with inactivating mutations in the PHEX gene, the cause of hereditary X-linked hypophosphatemia. The PHEX gene encodes a protease that activates phosphatonins, including the protein fibroblast growth factor type 23, which seems to be the phosphaturic factor produced by tumors that cause oncogenic osteomalacia.

ECTOPIC GROWTH HORMONE–RELEASING HORMONE

Ectopic GHRH-releasing tumors, often the bronchial or upper gastrointestinal carcinoid cell type, are a very rare cause of acromegaly (Table 136.12). In patients who present with acromegaly, the syndrome is usually of several years' duration to allow development of the typical clinical features. Accordingly, it is not surprising that the underlying tumors are usually small and benign. Clinical features are similar to those of typical acromegaly. Pituitary imaging is unhelpful because the radiographic features reveal an enlarged pituitary or an asymmetric mass that can be confused with an adenoma.[110] The histologic identification of true pituitary somatotroph hyperplasia, with a preserved reticulin network, led to the identification of GHRH in extracts from islet cell tumors.[111–114] Since then, many tumors and tumor cell lines have been shown to express or secrete GHRH.[115–119] For example, 25

TABLE 136.12 Tumors Associated With Ectopic Renin Secretion

Kidney (Wilms' tumor, renal cell carcinoma)
Lung (small-cell lung carcinoma, adenocarcinoma, leiomyosarcoma)
Pancreatic carcinoma
Ovarian tumors
Liver (hepatocellular carcinoma, hamartoma)
Ileal carcinoma
Adrenal paraganglioma
Orbital hemangiopericytoma

of 97 carcinoid tumors expressed GHRH in one study,[120] 17% of gastrointestinal and pancreatic tumors stained for GHRH in another,[121] and 63% of endometrial adenocarcinomas expressed GHRH.[122] In addition, subtle abnormalities in GH secretion and regulation may be found in patients harboring carcinoid tumors.[123] GH secretion has been associated with the development of several malignant tumors, notably prostate and colon adenocarcinoma. Therefore a complex long feedback loop may confer survival advantage on these tumors.

It has been suggested that all patients with acromegaly should be screened for ectopic GHRH because measurements in the plasma of greater than 0.3 ng/mL are virtually diagnostic of GHRH production by a tumor.[124] However, the assays are not widely available and currently remain research tools. In addition, the incidence is rare. In one series of 177 unselected patients with acromegaly, only one had detectable GHRH concentrations in plasma.[113,125]

Most commonly, the diagnosis is made after hypophysectomy, when histology shows somatotroph hyperplasia. In these cases, measurement of GHRH is necessary, as is the search for its source. The best treatment is curative resection of the primary tumor, but in cases in which the tumor is elusive or metastatic, success has been achieved using long-acting somatostatin analogues.[110,126–128]

NON–ISLET CELL TUMOR HYPOGLYCEMIA

Fasting hypoglycemia may arise as a consequence of non–islet cell tumor formation (see Chapter 38). Such tumors do not express insulin, which appears to be very tightly regulated in its tissue distribution, but rather the insulin-related molecules insulin-like growth factor (IGF)-1 or, more commonly, IGF-2. These two growth factors are members of the insulin family of peptide hormones along with relaxin, and both are capable of signaling through the IGF-1 receptor or the insulin receptor. IGF-1 is the liver-derived circulating mediator of GH action, whereas IGF-2 has a more important role in development. In normal subjects, IGF-1 and IGF-2 circulate at much higher concentrations than insulin and, if unopposed, cause profound hypoglycemia due to their actions through the insulin receptor. That this does not occur due to the presence of high-affinity, high-capacity, circulating IGF-binding proteins.[129-130] These result in very low concentrations of free IGFs, and they act to chaperone the IGFs and deliver them to their tissue beds of action.

A number of tumors, typically of mesenchymal origin, have been identified as the cause of non–islet cell hypoglycemia. The apparent mechanism is overproduction of IGF-2.[131] In addition to aberrant overproduction of IGF-2, there is also secretion of a partially processed form of IGF-2, "big" IGF-2, which includes a C-terminal extension peptide and frequently an abnormal glycosylation pattern.[132,133] The circulating IGF-2 would be expected to cause few problems if it were effectively sequestered by IGF-binding proteins, but this does not occur.[134,135] IGF-2 exerts a negative feedback at the pituitary somatotroph to suppress GH secretion. GH is the key regulator of hepatic IGF-binding protein 3, a component of the principal, ternary circulating IGF complex (Fig. 136.2). Therefore in the absence of effective GH drive to IGF-binding protein 3 production, the tumor-derived IGF-2 is left in a free state and is thus capable of acting through the insulin receptor and causing hypoglycemia.[136]

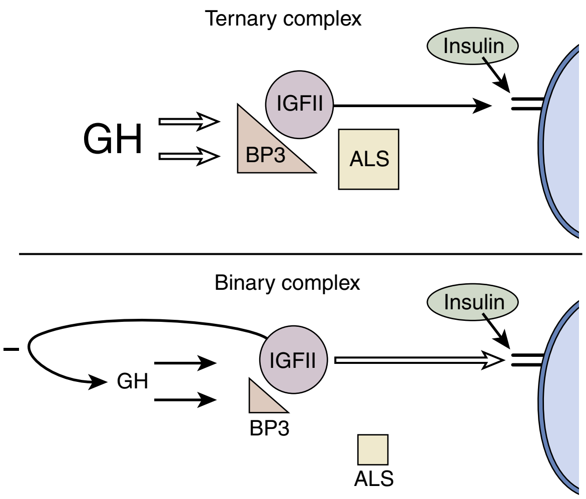


Figure 136.2 Formation of the ternary complex between insulin-like growth factors *(IGFs)*, IGF-binding protein 3 *(BP3)*, and the acid-labile subunit *(ALS)*. Normal physiology *(top)* in ectopic production of IGF (usually IGF-2) by tumor *(bottom)*. Increased activity of IGF inhibits pituitary production of growth hormone *(GH)* and so reduces hepatic production of ALS and IGF-BP3. The resulting binary complex has increased insulin-like activity and, acting through the insulin receptor, causes hypoglycemia.

Diagnosis

The diagnosis is based on the recognition that the patient's symptoms are due to hypoglycemia, which involves a detailed history and confirmation of either hypoglycemia occurring during a symptomatic episode or fasting hypoglycemia. The diagnosis should be considered in all unconscious cancer patients without signs of vascular events or brain metastases. The typical biochemical accompaniment to hypoglycemia is suppressed insulin, suppressed ketone bodies (β-hydroxybutyrate), and suppressed GH. In addition, IGF-2 is usually elevated or normal, with suppression of IGF-binding protein 3 as a consequence of low GH.[131,137]

Most tumors are large and intraabdominal, although thoracic tumors have also been reported.[138] The underlying tumor may be benign, in which case curative resection is the aim. Histologic confirmation of the diagnosis is by specific immunostaining of a biopsy for IGF-2.

Management

Effective management requires either surgical excision or debulking of the tumor. This is best achieved by close consultation between oncologist, surgeon, and interventional radiologist, who may be able to perform tumor embolization. These tumors tend not to be radiosensitive, although there are isolated case reports of therapeutic response. Because these tumors are rare, management is tailored to the individual patient and is usually based on pragmatic approaches. In patients whose tumors are inoperable, relief of hypoglycemia can be achieved by using glucocorticoid, usually dexamethasone, either alone or with recombinant human GH.[136,139,140]. Glucocorticoids exert a direct anti-insulin and anti-IGF effect, and GH acts both by its anti-insulin action and by increasing the IGF-binding protein concentration, which then acts to "mop up" the excess IGF.

KEY POINTS

- Spontaneous hypoglycemia can be hard to diagnose, but non–islet cell tumors should be considered. Typically tumors are large, and abdominal, and the biochemical features are low insulin with raised IGF2.

OTHER PITUITARY ECTOPIC HORMONES

The two pituitary hormones prolactin and GH are of interest because they have wide extrapituitary expression and yet are very seldom the cause of clinically significant ectopic hormone syndromes. Prolactin is expressed in decidualized endometrium, T lymphocytes, mammary epithelial cells, skin, sweat glands, and the brain. It is the same gene that is transcribed in all these cases, rather than the related placental lactogen-type gene, but the regulation of the gene appears completely different.[141] Whereas in the pituitary lactotroph cell, prolactin is under the transcriptional control of the pituitary-specific factor Pit1, in extrapituitary tissues, Pit1 is not expressed, and the pituitary promoter of the prolactin gene is silent. The gene is transcribed from an upstream promoter, which gives rise to a slightly longer mRNA, with a unique 5′ end but, after processing, results in a protein with the same amino acid sequence. Because the gene is transcribed from a different promoter, the control of gene transcription, its basal rate, and regulation by external signals is different. For example, in T-lymphocyte prolactin, gene transcription is responsive to the immunophilins, including cyclosporin A. The function of this extrapituitary prolactin is subject to debate, and it is not clear why such widespread expression is accompanied by such rarity of overexpression in malignant disease, in contrast to ACTH or vasopressin expression. Because prolactin receptors are found in a variety of tissues that cannot be reconciled with an exclusive action on mammary milk production, prolactin may have a more diverse role than so far determined.

Ectopic prolactin secretion is a rare occurrence.[142] There are reports of ectopic prolactin secretion from a bronchial tumor,[143] a gonadoblastoma,[144] a renal cell carcinoma,[145] and undifferentiated lung cancer.[145] However, breast cancers have been shown to express both prolactin and prolactin receptors.[146–148] There is some evidence of a weak tumor-promoting effect of prolactin on breast adenocarcinoma cells[149,150] that has led to the suggestion of a paracrine role for prolactin in the development or progression of breast malignancy. Prolactin may have other activities in addition to its best characterized function on lactation, and, in particular, a 16-kD fragment of prolactin has antiangiogenic potential and may influence tumorigenesis.[151,152] In addition, prolactin can activate nuclear factor-κB, a survival-promoting transcription factor.[153]

GH is also found in extrapituitary tissues, again in cells of hematopoietic lineage. This expression has been suggested to result in paracrine signaling, although hard data are lacking.

Ectopic GH secretion was first described in 1968 in a man with lung cancer. Resection of the tumor caused a decrease in GH concentrations, although GH was not measured in the primary tumor.[154] There are very few cases in which the criteria are met for true ectopic secretion of GH. One such case was described by Melmed and coworkers,[155] who carefully showed GH secretion from a pancreatic islet tumor, with demonstration of a gradient of GH across the tumor *in vivo* and resolution of abnormal GH and IGF-1 after tumor resection. They also showed GH expression in tumor sections, at both the mRNA and protein level. It is interesting that there is hypertrophic osteoarthropathy in cases of ectopic GH secretion, although this is not a feature of acromegaly; it may result from other humoral factors secreted by the tumors or from neural influences.[155–157]

Gonadotropins

Human chorionic gonadotropin is expressed in trophoblast and from both germ cell and trophoblast tumors. The most frequent nontrophoblast tumor associated with human chorionic gonadotropin overproduction is lung,[158] but other tumors have been described including adrenal, breast, bladder, maxilla, hepatoblastoma, osteosarcoma, and lymphoma.[159–163]

The production of human chorionic gonadotropin in men causes gynecomastia as a result of increased estrogen secretion and may cause precocious puberty in children.[164]

ECTOPIC GUT HORMONE SYNDROMES

Clinical syndromes associated with ectopic production of gut hormones by tumors are very rare, but vasoactive intestinal polypeptide causing typical watery diarrhea has been described (see Chapter 135). Tumors of the lung, medullary thyroid carcinoma, pheochromocytoma, and neuroendocrine tumors of the kidney have all been reported to produce vasoactive intestinal polypeptide.[165–167]

Ectopic Renin Secretion

Renin production is usually tightly restricted to the juxtaglomerular apparatus of the kidney, and true ectopic secretion by nonrenal tumors is very rare. In cases described, hypertension is a feature, along with hypokalemia.[168] As is often the case with ectopic production of usually processed hormones, renin precursors are described, with an increased prorenin-renin ratio.[169–171] Resection of the primary tumor affords cure, if identified, but medical therapy with angiotensin-converting enzyme inhibitors or with angiotensin-blocking agents may be beneficial in noncurable disease.[172]

Extrarenal tumors secreting renin include lung,[173–175] pancreas,[171] ovary,[176,177] liver,[178] ileum,[179] adrenal,[180] and orbital hemangiopericytoma.[181]

KEY POINTS

- Almost all hormones have been reported as ectopically produced in cancer but for most experience is limited to case reports. Key considerations are proven tumor production and loss of hormone excess after tumor resection.

ACKNOWLEDGMENTS

The author is grateful to Professor Wass and Dr. Turner, who wrote the previous version of this chapter, for inspiration and for the use of tables.

REFERENCES

1. Meador CK, Liddle GW, Island DP, et al. Cause of Cushing's syndrome in patients with tumors arising from "nonendocrine" tissue. *J Clin Endocrinol Metab*. 1962;22:693–703.
2. Orth DN, Nicholson WE, Mitchell WM, et al. Biologic and immunologic characterization and physical separation of ACTH and ACTH fragments in the ectopic ACTH syndrome. *J Clin Invest*. 1973;52:1756–1769.
3. Amatruda Jr TT, Mulrow PJ, Gallagher JC, Sawyer WH. Carcinoma of the lung with inappropriate antidiuresis. Demonstration of antidiuretic-hormone-like activity in tumor extract. *N Engl J Med*. 1963;269:544–549.

4. DeBold CR, Menefee JK, Nicholson WE, Orth DN. Proopiomelanocortin gene is expressed in many normal human tissues and in tumors not associated with ectopic adrenocorticotropin syndrome. *Mol Endocrinol*. 1988;2:862–870.
5. Albright F, Reifenstein ED, eds. *The Parathyroid Glands and Metabolic Bone Disease-Selected Studies*. Baltimore: Williams and Wilkins; 1948.
6. Newell-Price J, King P, Clark AJ. The CpG island promoter of the human proopiomelanocortin gene is methylated in nonexpressing normal tissue and tumors and represses expression. *Mol Endocrinol*. 2001;15:338–348.
7. Isidori AM, Kaltsas GA, Mohammed S, et al. Discriminatory value of the low-dose dexamethasone suppression test in establishing the diagnosis and differential diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*. 2003;88:5299–5306.
8. Newell-Price J, Morris DG, Drake WM, et al. Optimal response criteria for the human CRH test in the differential diagnosis of ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*. 2002;87:1640–1645.
9. Ray DW, Littlewood AC, Clark AJ, et al. Human small cell lung cancer cell lines expressing the proopiomelanocortin gene have aberrant glucocorticoid receptor function. *J Clin Invest*. 1994;93:1625–1630.
10. Clark AJ, Stewart MF, Lavender PM, et al. Defective glucocorticoid regulation of proopiomelanocortin gene expression and peptide secretion in a small cell lung cancer cell line. *J Clin Endocrinol Metab*. 1990;70:485–490.
11. Ray DW, Davis JR, White A, Clark AJ. Glucocorticoid receptor structure and function in glucocorticoid-resistant small cell lung carcinoma cells. *Cancer Res*. 1996;56:3276–3280.
12. Sommer P, Le Rouzic P, Gillingham H, et al. Glucocorticoid receptor overexpression exerts an antisurvival effect on human small cell lung cancer cells. *Oncogene*. 2007;26(5):7111.
13. Florkowski CM, Wittert GA, Lewis JG, et al. Glucocorticoid responsive ACTH secreting bronchial carcinoid tumours contain high concentrations of glucocorticoid receptors. *Clin Endocrinol (Oxf)*. 1994;40:269–274.
14. Isidori AM, Kaltsas A, Pozza C et al. The ectopic adrenocorticotropin syndrome: clinical features, diagnosis, management and long-term follow-up. *J Clin Endocrinol Metab*. 2006;91:371–377.
15. Tabarin A, Navarranne A, Guerin J, et al. Use of ketoconazole in the treatment of Cushing's disease and ectopic ACTH syndrome. *Clin Endocrinol (Oxf)*. 1991;34:63–69.
16. De Rosa G, Testa A, Liberale I, et al. Successful treatment of ectopic Cushing's syndrome with the long-acting somatostatin analog octreotide. *Exp Clin Endocrinol (Oxf)*. 1993;101:319–325.
17. Young J, Deneux C, Grino M, et al. Pitfall of petrosal sinus sampling in a Cushing's syndrome secondary to ectopic adrenocorticotropin-corticotropin releasing hormone (ACTH-CRH) secretion. *J Clin Endocrinol Metab*. 1998;83:305–308.
18. Gross AJ, Steinberg SM, Reilly JG, et al. Atrial natriuretic factor and arginine vasopressin production in tumor cell lines from patients with lung cancer and their relationship to serum sodium. *Cancer Res*. 1993;53:67–74.
19. Ghandur-Mnaymneh L, Satterfield S, Block NL. Small cell carcinoma of the prostate gland with inappropriate antidiuretic hormone secretion: morphological, immunohistochemical and clinical expressions. *J Urol*. 1986;135:1263–1266.
20. Moses AM, Notman DD. Diabetes insipidus and syndrome of inappropriate antidiuretic hormone secretion (SIADH). *Adv Intern Med*. 1982;27(73–100).
21. Ellison DH, Berl T. The syndrome of inappropriate antidiuresis. *N Engl J Med*. 2007;356:2064–2072.
22. Zajac JD, Martin TJ, Hudson P, et al. Biosynthesis of calcitonin by human lung cancer cells. *Endocrinology*. 1985;116:749–755.
23. Asa SL, Kovacs K, Killinger DW, et al. Pancreatic islet cell carcinoma producing gastrin, ACTH, alpha-endorphin, somatostatin and calcitonin. *Am J Gastroenterol*. 1980;74:30–35.
24. Yoshimoto K, Yamasaki R, Sakai H, et al. Ectopic production of parathyroid hormone by small cell lung cancer in a patient with hypercalcemia. *J Clin Endocrinol Metab*. 1989;68:976–981.
25. Nussbaum SR, Gaz RD, Arnold A. Hypercalcemia and ectopic secretion of parathyroid hormone by an ovarian carcinoma with rearrangement of the gene for parathyroid hormone. *N Engl J Med*. 1990;323:1324–1328.
26. Rizzoli R, Pache JC, Didierjean L, et al. A thymoma as a cause of true ectopic hyperparathyroidism. *J Clin Endocrinol Metab*. 1994;79:912–915.
27. Strewler GJ, Budayr AA, Clark OH, et al. Production of parathyroid hormone by a malignant nonparathyroid tumor in a hypercalcemic patient. *J Clin Endocrinol Metab*. 1993;76:1373–1375.
28. Budayr AA, Nissenson RA, Klein RF, et al. Increased serum levels of a parathyroid hormone-like protein in malignancy-associated hypercalcemia. *Ann Intern Med*. 1989;111:807–812.
29. Broadus AE, Mangin M, Ikeda K, et al. Humoral hypercalcemia of cancer. Identification of a novel parathyroid hormone-like peptide. *N Engl J Med*. 1988;319:556–563.
30. Burtis WJ, Wu T, Bunch C, et al. Identification of a novel 17,000-dalton parathyroid hormone-like adenylate cyclase-stimulating protein from a tumor associated with humoral hypercalcemia of malignancy. *J Biol Chem*. 1987;262:7151–7156.
31. Burtis WJ, Brady TG, Orloff JJ, et al. Immunochemical characterization of circulating parathyroid hormone-related protein in patients with humoral hypercalcemia of cancer. *N Engl J Med*. 1990;322:1106–1112.
32. Firkin F, Schneider H, Grill V. Parathyroid hormone-related protein in hypercalcemia associated with hematological malignancy. *Leuk Lymphoma*. 1998;29:499–506.
33. Grill V, Ho P, Body JJ, et al. Parathyroid hormone-related protein: elevated levels in both humoral hypercalcemia of malignancy and hypercalcemia complicating metastatic breast cancer. *J Clin Endocrinol Metab*. 1991;73:1309–1315.
34. Karaplis AC. PTHrP: novel roles in skeletal biology. *Curr Pharm Des*. 2001;7:655–670.
35. Wysolmerski JJ, Philbrick WM, Dunbar ME, et al. Rescue of the parathyroid hormone-related protein knockout mouse demonstrates that parathyroid hormone-related protein is essential for mammary gland development. *Development*. 1998;125:1285–1294.
36. Ralston SH, Cowan RA, Robertson AG, et al. Circulating vitamin D metabolites and hypercalcaemia of malignancy. *Acta Endocrinol (Copenh)*. 1984;106:556–563.
37. Fraher LJ, Hodsman AB, Jonas K, et al. A comparison of the in vivo biochemical responses to exogenous parathyroid hormone-(1-34) [PTH-(1-34) and PTH-related peptide-(1-34) in man. *J Clin Endocrinol Metab*. 1992;75:417–423.
38. Bundred NJ, Walls J, Ratcliffe WA. Parathyroid hormone-related protein, bone metastases and hypercalcaemia of malignancy. *Ann R Coll Surg Engl*. 1996;78:354–358.
39. Bundred NJ, Walker RA, Ratcliffe WA, et al. Parathyroid hormone related protein and skeletal morbidity in breast cancer. *Eur J Cancer*. 1992;28:690–692.
40. Chirgwin JM, Guise TA. Molecular mechanisms of tumor-bone interactions in osteolytic metastases. *Crit Rev Eukaryot Gene Expr*. 2000;10:159–178.
41. Guise TA, Yin JJ, Thomas RJ, et al. Parathyroid hormone-related protein (PTHrP)-(1-139) isoform is efficiently secreted in vitro and enhances breast cancer metastasis to bone in vivo. *Bone*. 2002;30:670–676.
42. Ikeda K, Lu C, Weir EC, et al. Transcriptional regulation of the parathyroid hormone-related peptide gene by glucocorticoids and vitamin D in a human C-cell line. *J Biol Chem*. 1989;264:15743–15746.
43. Lu C, Ikeda K, Deftos LJ, et al. Glucocorticoid regulation of parathyroid hormone-related peptide gene transcription in a human neuroendocrine cell line. *Mol Endocrinol*. 1989;3:2034–2040.
44. Eto M, Akishita M, Ishikawa M, et al. Cytokine-induced expression of parathyroid hormone-related peptide in cultured human vascular endothelial cells. *Biochem Biophys Res Commun*. 1998;249:339–343.
45. Urena P, Iida-Klein A, Kong XF, et al. Regulation of parathyroid hormone (PTH)/PTH-related peptide receptor messenger ribonucleic acid by glucocorticoids and PTH in ROS 17/2.8 and OK cells. *Endocrinology*. 1994;134:451–456.
46. Seymour JF, Grill V, Martin TJ, et al. Hypercalcemia in the blastic phase of chronic myeloid leukemia associated with elevated parathyroid hormone-related protein. *Leukemia*. 1993;7:1672–1675.
47. Zeimer H, Firkin F, Grill V, et al. Assessment of cellular expression of parathyroid hormone-related protein mRNA and protein in multiple myeloma. *J Pathol*. 2000;192:336–341.

48. Richard V, Lairmore MD, Green PL, et al. Humoral hypercalcemia of malignancy: severe combined immunodeficient/beige mouse model of adult T-cell lymphoma independent of human T-cell lymphotropic virus type-1 tax expression. *Am J Pathol*. 2001;158:2219–2228.
49. Obagi S, Derubertis F, Brown L, Deng JS. Hypercalcemia and parathyroid hormone related protein expression in cutaneous T-cell lymphoma. *Int J Dermatol*. 1999;38:855–862.
50. Moseley JM, Danks JA, Grill V, et al. Immunocytochemical demonstration of PTHrP protein in neoplastic tissue of HTLV-1 positive human adult T cell leukaemia/lymphoma: implications for the mechanism of hypercalcaemia. *Br J Cancer*. 1991;64:745–748.
51. Bunn Jr PA, Schechter GP, Jaffe E, et al. Clinical course of retrovirus-associated adult T-cell lymphoma in the United States. *N Engl J Med*. 1983;309:257–264.
52. Ralston SH, Gallacher SJ, Patel U, et al. Cancer-associated hypercalcemia: morbidity and mortality. Clinical experience in 126 treated patients. *Ann Intern Med*. 1990;112:499–504.
53. Ralston SH, Gardner MD, Dryburgh FJ, et al. Comparison of aminohydroxypropylidene diphosphonate, mithramycin, and corticosteroids/calcitonin in treatment of cancer-associated hypercalcaemia. *Lancet*. 1985;2:907–910.
54. Gallacher SJ, Ralston SH, Fraser WD, et al. A comparison of low versus high dose pamidronate in cancer-associated hypercalcaemia. *Bone Miner*. 1991;15:249–256.
55. Ralston SH, Gallacher SJ, Patel U, et al. Comparison of three intravenous bisphosphonates in cancer-associated hypercalcaemia. *Lancet*. 1989;2:1180–1182.
56. Ralston SH, Thiebaud D, Herrmann Z, et al. Dose-response study of ibandronate in the treatment of cancer-associated hypercalcaemia. *Br J Cancer*. 1997;75:295–300.
57. Ralston SH, Alzaid AA, Gardner MD, et al. Treatment of cancer associated hypercalcaemia with combined aminohydroxypropylidene diphosphonate and calcitonin. *Br Med J*. 1986;292:1549–1550.
58. McCance RA. Osteomalacia with Looser's nodes (Milkman's syndrome) due to raised resistance to vitamin D acquired about the age of 15 years. *Q J Med*. 1947;16:33–47.
59. Drange MR, Melmed S. Long-acting lanreotide induces clinical and biochemical remission of acromegaly caused by disseminated growth hormone-releasing hormone-secreting carcinoid. *J Clin Endocrinol Metab*. 1998;83:3104–3109.
60. Guillemin R, Brazeau P, Bohlen P, et al. Growth hormone-releasing factor from a human pancreatic tumor that caused acromegaly. *Science*. 1982;218:585–587.
61. Rivier J, Spiess J, Thorned M, Vale W. Characterization of a growth hormone-releasing factor from a human pancreatic islet tumour. *Nature*. 1982;300:276–278.
62. Thorne MO, Perryman RL, Cronin MJ, et al. Somatotroph hyperplasia. Successful treatment of acromegaly by removal of a pancreatic islet tumor secreting a growth hormone-releasing factor. *J Clin Invest*. 1982;70:965–977.
63. thorned MO, Rivier J, Spiess J, et al. Human pancreatic growth-hormone-releasing factor selectively stimulates growth-hormone secretion in man. *Lancet*. 1983;1:24–28.
64. Sano T, Saito H, Yamasaki R, et al. Production and secretion of immunoreactive growth hormone-releasing factor by pheochromocytomas. *Cancer*. 1986;57:1788–1793.
65. Othman NH, Ezzat S, Kovacs K, et al. Growth hormone-releasing hormone (GHRH) and GHRH receptor (GHRH-R) isoform expression in ectopic acromegaly. *Clin Endocrinol (Oxf)*. 2001;55:135–140.
66. Sano T, Yamasaki R, Saito H, et al. Growth hormone-releasing hormone (GHRH)-secreting pancreatic tumor in a patient with multiple endocrine neoplasia type I. *Am J Surg Pathol*. 1987;11:810–819.
67. Schulte HM, Benker G, Windeck R, et al. Failure to respond to growth hormone releasing hormone (GHRH) in acromegaly due to a GHRH secreting pancreatic tumor: dynamics of multiple endocrine testing. *J Clin Endocrinol Metab*. 1985;61:585–587.
68. Sano T, Asa SL, Kovacs K. Growth hormone-releasing hormone-producing tumors: clinical, biochemical, and morphological manifestations. *Endocr Rev*. 1988;9:357–373.
69. Dayal Y, Lin HD, Tallberg K, et al. Immunocytochemical demonstration of growth hormone-releasing factor in gastrointestinal and pancreatic endocrine tumors. *Am J Clin Pathol*. 1986;85:13–20.
70. Chatzistamou I, Schally AV, Pafiti A, et al. Expression of growth hormone-releasing hormone in human primary endometrial carcinomas. *Eur J Endocrinol*. 2002;147:381–386.
71. Oberg K, Norheim I, Wide L. Serum growth hormone in patients with carcinoid tumours; basal levels and response to glucose and thyrotrophin releasing hormone. *Acta Endocrinol (Copenh)*. 1985;109:13–18.
72. Faglia G, Arosio M, Bazzoni N. Ectopic acromegaly. *Endocrinol Metab Clin North Am*. 1992;21:575–595.
73. Thorner MO, Frohman LA, Leong DA, et al. Extrahypothalamic growth-hormone-releasing factor (GRF) secretion is a rare cause of acromegaly: plasma GRF levels in 177 acromegalic patients. *J Clin Endocrinol Metab*. 1984;59:846–849.
74. Melmed S. Extrapituitary acromegaly. *Endocrinol Metab Clin North Am*. 1991;20:507–518.
75. Melmed S, Ziel FH, Braunstein GD, et al. Medical management of acromegaly due to ectopic production of growth hormone-releasing hormone by a carcinoid tumor. *J Clin Endocrinol Metab*. 1988;67:395–399.
76. Lefebvre S, De Paepe L, Abs R, et al. Subcutaneous octreotide treatment of a growth hormone-releasing hormone-secreting bronchial carcinoid: superiority of continuous versus intermittent administration to control hormonal secretion. *Eur J Endocrinol*. 1995;133:320–324.
77. Marks V, Teale JD. Tumours producing hypoglycaemia. *Diabetes Metab Rev*. 1991;7:79–91.
78. Teale JD. Non-islet cell tumour hypoglycaemia. *Clin Endocrinol (Oxf)*. 1999;51:147.
79. Teale JD, Marks V. Inappropriately elevated plasma insulin-like growth factor II in relation to suppressed insulin-like growth factor I in the diagnosis of non-islet cell tumour hypoglycaemia. *Clin Endocrinol (Oxf)*. 1990;33:87–98.
80. Daughaday WH, Wu JC, Lee SD, et al. Abnormal processing of pro-IGF-II in patients with hepatoma and in some hepatitis B virus antibody-positive asymptomatic individuals. *J Lab Clin Med*. 1990;116:555–562.
81. Daughaday WH, Trivedi B, Baxter RC. Serum "big insulin-like growth factor II" from patients with tumor hypoglycemia lacks normal E-domain O-linked glycosylation, a possible determinant of normal propeptide processing. *Proc Natl Acad Sci U S A*. 1993;90:5823–5827.
82. Frystyk J, Skjaerbaek C, Zapf J, et al. Increased levels of circulating free insulin-like growth factors in patients with non-islet cell tumour hypoglycaemia. *Diabetologia*. 1998;41:589–594.
83. Moller N, Frystyk J, Skjaerbaek C, et al. Systemic and regional tumour metabolism in a patient with non-islet cell tumour hypoglycaemia: role of increased levels of free insulin-like growth factors. *Diabetologia*. 1996;39:1534–1535.
84. Teale JD, Blum WF, Marks V. Alleviation of non-islet cell tumour hypoglycaemia by growth hormone therapy is associated with changes in IGF binding protein-3. *Ann Clin Biochem*. 1992;29:314–323.
85. Daughaday WH, Deuel TF. Tumor secretion of growth factors. *Endocrinol Metab Clin North Am*. 1991;20:539–563.
86. Masson EA, MacFarlane IA, Graham D, et al. Spontaneous hypoglycaemia due to a pleural fibroma: role of insulin like growth factors. *Thorax*. 1991;46:930–931.
87. Teale JD, Marks V. Glucocorticoid therapy suppresses abnormal secretion of big IGF-II by non-islet cell tumours inducing hypoglycaemia (NICTH). *Clin Endocrinol (Oxf)*. 1998;49:491–498.
88. de Groot JWB, Rikhof B, van Doorn J, et al. Non-islet cell tumour-induced hypoglycaemia: a review of the literature including two new cases. *Endocr Relat Cancer*. 2007;14:979–993.
89. Ben Jonathan N, Mershon JL, Allen DL, et al. Extrapituitary prolactin: distribution, regulation, functions, and clinical aspects. *Endocr Rev*. 1996;17:639–669.
90. Molitch ME, Schwartz S, Mukherji B. Is prolactin secreted ectopically? *Am J Med*. 1981;70:803–807.
91. Hoffman WH, Gala RR, Kovacs K, Subramanian MG. Ectopic prolactin secretion from a gonadoblastoma. *Cancer*. 1987;60:2690–2695.

92. Turkington RW. Ectopic production of prolactin. *N Engl J Med*. 1971;285:1455–1458.
93. Jahnke GD, Trempus CS, Kari FW, DiAugustine RP. Expression of a prolactin-like factor in preneoplastic and neoplastic mouse mammary gland and cells. *J Mol Endocrinol*. 1996;17:247–256.
94. Clevenger CV, Chang WP, Ngo W, et al. Expression of prolactin and prolactin receptor in human breast carcinoma. Evidence for an autocrine/paracrine loop. *Am J Pathol*. 1995;146:695–705.
95. Clevenger CV, Furth PA, Hankinson SE, Schuler LA. The role of prolactin in mammary carcinoma. *Endocr Rev*. 2003;24:1–27.
96. Clevenger CV, Plank TL. Prolactin as an autocrine/paracrine factor in breast tissue. *J Mammary Gland Biol Neoplasia*. 1997;2:59–68.
97. Kwa HG, Cleton F, Wang DY, et al. A prospective study of plasma prolactin levels and subsequent risk of breast cancer. *Int J Cancer*. 1981;28:673–676.
98. Clapp C, Martial JA, Guzman RC, et al. The 16-kilodalton N-terminal fragment of human prolactin is a potent inhibitor of angiogenesis. *Endocrinology*. 1993;133:1292–1299.
99. Ferrara N, Clapp C, Weiner R. The 16K fragment of prolactin specifically inhibits basal or fibroblast growth factor stimulated growth of capillary endothelial cells. *Endocrinology*. 1991;129:896–900.
100. Macotela Y, Mendoza C, Corbacho AM, et al. 16K prolactin induces NF-kappaB activation in pulmonary fibroblasts. *J Endocrinol*. 2002;175:R13–R18.
101. Steiner H, Dahlback O, Waldenstrom J. Ectopic growth-hormone production and osteoarthropathy in carcinoma of the bronchus. *Lancet*. 1968;1:783–785.
102. Melmed S, Ezrin C, Kovacs K, et al. Acromegaly due to secretion of growth hormone by an ectopic pancreatic islet-cell tumor. *N Engl J Med*. 1985;312:9–17.
103. Melmed S, Rushakoff RJ. Ectopic pituitary and hypothalamic hormone syndromes. *Endocrinol Metab Clin North Am*. 1987;16:805–821.
104. Greenberg PB, Martin TJ, Beck C, Burger HG. Synthesis and release of human growth hormone from lung carcinoma in cell culture. *Lancet*. 1972;1:350–352.
105. Yokotani T, Koizumi T, Taniguchi R, et al. Expression of alpha and beta genes of human chorionic gonadotropin in lung cancer. *Int J Cancer*. 1997;71:539–544.
106. Broder LE, Weintraub BD, Rosen SW, et al. Placental proteins and their subunits as tumor markers in prostatic carcinoma. *Cancer*. 1977;40:211–216.
107. Scholl PD, Jurco S, Austin JR. Ectopic production of beta-HCG by a maxillary squamous cell carcinoma. *Head Neck*. 1997;19:701–705.
108. McArthur JW, Toll GD, Russfield AB, et al. Sexual precocity attributable to ectopic gonadotropin secretion by hepatoblastoma. *Am J Med*. 1973;54:390–403.
109. Ordonez NG, Ayala AG, Raymond AK, et al. Ectopic production of the beta-subunit of human chorionic gonadotropin in osteosarcoma. *Arch Pathol Lab Med*. 1989;113:416–419.
110. Mendelsohn G, Eggleston JC, Olson JL, et al. Vasoactive intestinal peptide and its relationship to ganglion cell differentiation in neuroblastic tumors. *Lab Invest*. 1979;41:144–149.
111. Said SI, Faloona GR. Elevated plasma and tissue levels of vasoactive intestinal polypeptide in the watery-diarrhea syndrome due to pancreatic, bronchogenic and other tumors. *N Engl J Med*. 1975;293:155–160.
112. Hamilton I, Reis L, Bilimoria S, Long RG. A renal vipoma. *BMJ*. 1980;281:1323–1324.
113. Hollifield JW, Page DL, Smith C, et al. Renin-secreting clear cell carcinoma of the kidney. *Arch Intern Med*. 1975;135:859–864.
114. Soubrier F, Devaux C, Galen FX, et al. Biochemical and immunological characterization of ectopic tumoral renin. *J Clin Endocrinol Metab*. 1982;54:139–144.
115. Atlas SA, Hesson TE, Sealey JE, et al. Characterization of inactive renin ("prorenin") from renin-secreting tumors of nonrenal origin. Similarity to inactive renin from kidney and normal plasma. *J Clin Invest*. 1984;73:437–447.
116. Ruddy MC, Atlas SA, Salerno FG. Hypertension associated with a renin-secreting adenocarcinoma of the pancreas. *N Engl J Med*. 1982;307:993–997.
117. Aurell M, Rudin A, Tisell LE, et al. Captopril effect on hypertension in patient with renin-producing tumour. *Lancet*. 1979;2:149–150.
118. Hauger-Klevene JH. High plasma renin activity in an oat cell carcinoma: a renin-secreting carcinoma? *Cancer*. 1970;26:1112–1114.
119. Kawai K, Fukamizu A, Kawakami Y, et al. A case of renin producing leiomyosarcoma originating in the lung. *Endocrinol Jpn*. 1991;38:603–609.
120. Anderson PW, Macaulay L, Do YS, et al. Extrarenal renin-secreting tumors: insights into hypertension and ovarian renin production. *Medicine (Baltim)*. 1989;68:257–268.
121. Tetu B, Lebel M, Camilleri JP. Renin-producing ovarian tumor. A case report with immunohistochemical and electron-microscopic study. *Am J Surg Pathol*. 1988;12:634–640.
122. Fried G, Wikstrom LM, Hoog A, et al. Multiple neuropeptide immunoreactivities in a renin-producing human paraganglioma. *Cancer*. 1994;74:142–151.
123. Yokoyama H, Yamane Y, Takahara J, et al. A case of ectopic renin-secreting orbital hemangiopericytoma associated with juvenile hypertension and hypokalemia. *Acta Med Okayama*. 1979;33:315–322.
124. Faglia G, Arosio M, Bazzoni N. Ectopic acromegaly. *Endocrinol Metab Clin North Am*. 1992;21:575–595.
125. Thorner MO, Frohman LA, Leong DA, et al. Extrahypothalamic growth-hormone-releasing factor (GRF) secretion is a rare cause of acromegaly: plasma GRF levels in 177 acromegalic patients. *J Clin Endocrinol Metab*. 1984;59:846–849.
126. Melmed S. Extrapituitary acromegaly. *Endocrinol Metab Clin North Am*. 1991;20:507–518.
127. Melmed S, Ziel FH, Braunstein GD, et al. Medical management of acromegaly due to ectopic production of growth hormone-releasing hormone by a carcinoid tumor. *J Clin Endocrinol Metab*. 1988;67:395–399.
128. Lefebvre S, De Paepe L, Abs R, et al. Subcutaneous octreotide treatment of a growth hormone-releasing hormone-secreting bronchial carcinoid: superiority of continuous versus intermittent administration to control hormonal secretion. *Eur J Endocrinol*. 1995;133:320–324.
129. Marks V, Teale JD. Tumours producing hypoglycaemia. *Diabetes Metab Rev*. 1991;7:79–91.
130. Teale JD. Non-islet cell tumour hypoglycaemia. *Clin Endocrinol (Oxf)*. 1999;51:147.
131. Teale JD, Marks V. Inappropriately elevated plasma insulin-like growth factor II in relation to suppressed insulin-like growth factor I in the diagnosis of non-islet cell tumour hypoglycaemia. *Clin Endocrinol (Oxf)*. 1990;33:87–98.
132. Daughaday WH, Wu JC, Lee SD, Kapadia M. Abnormal processing of pro-IGF-II in patients with hepatoma and in some hepatitis B virus antibody-positive asymptomatic individuals. *J Lab Clin Med*. 1990;116:555–562.
133. Daughaday WH, Trivedi B, Baxter RC. Serum "big insulin-like growth factor II" from patients with tumor hypoglycemia lacks normal E-domain O-linked glycosylation, a possible determinant of normal propeptide processing. *Proc Natl Acad Sci U S A*. 1993;90:5823–5827.
134. Frystyk J, Skjaerbaek C, Zapf J, Orskov H. Increased levels of circulating free insulin-like growth factors in patients with non-islet cell tumour hypoglycaemia. *Diabetologia*. 1998;41:589–594.
135. Moller N, Frystyk J, Skjaerbaek C, et al. Systemic and regional tumour metabolism in a patient with non-islet cell tumour hypoglycaemia: role of increased levels of free insulin-like growth factors. *Diabetologia*. 1996;39:1534–1535.
136. Teale JD, Blum WF, Marks V. Alleviation of non-islet cell tumour hypoglycaemia by growth hormone therapy is associated with changes in IGF binding protein-3. *Ann Clin Biochem*. 1992;29:314–323.
137. Daughaday WH, Deuel TF. Tumor secretion of growth factors. *Endocrinol Metab Clin North Am*. 1991;20:539–563.
138. Masson EA, MacFarlane IA, Graham D, Foy P. Spontaneous hypoglycaemia due to a pleural fibroma: role of insulin like growth factors. *Thorax*. 1991;46:930–931.
139. Teale JD, Marks V. Glucocorticoid therapy suppresses abnormal secretion of big IGF-II by non-islet cell tumours inducing hypoglycaemia (NICTH). *Clin Endocrinol (Oxf)*. 1998;49:491–498.

140. de Groot JWB, Rikhof B, van Doorn J, et al. Non-islet cell tumour-induced hypoglycaemia: a review of the literature including two new cases. *Endocr Relat Cancer*. 2007;14:979–993.
141. Ben Jonathan N, Mershon JL, Allen DL, Steinmetz RW. Extrapituitary prolactin: distribution, regulation, functions, and clinical aspects. *Endocr Rev*. 1996;17:639–669.
142. Molitch ME, Schwartz S, Mukherji B. Is prolactin secreted ectopically? *Am J Med*. 1981;70:803–807.
143. Rees LH, Bloomfield GA, Rees GM, et al. Multiple hormones in a bronchial tumor. *J Clin Endocrinol Metab*. 1974;38:1090–1097.
144. Hoffman WH, Gala RR, Kovacs K, Subramanian MG. Ectopic prolactin secretion from a gonadoblastoma. *Cancer*. 1987;60:2690–2695.
145. Turkington RW. Ectopic production of prolactin. *N Engl J Med*. 1971;285:1455–1458.
146. Jahnke GD, Trempus CS, Kari FW, DiAugustine RP. Expression of a prolactin-like factor in preneoplastic and neoplastic mouse mammary gland and cells. *J Mol Endocrinol*. 1996;17:247–256.
147. Clevenger CV, Chang WP, Ngo W, et al. Expression of prolactin and prolactin receptor in human breast carcinoma. Evidence for an autocrine/paracrine loop. *Am J Pathol*. 1995;146:695–705.
148. Clevenger CV, Furth PA, Hankinson SE, Schuler LA. The role of prolactin in mammary carcinoma. *Endocr Rev*. 2003;24:1–27.
149. Clevenger CV, Plank TL. Prolactin as an autocrine/paracrine factor in breast tissue. *J Mammary Gland Biol Neoplasia*. 1997;2:59–68.
150. Kwa HG, Cleton F, Wang DY, et al. A prospective study of plasma prolactin levels and subsequent risk of breast cancer. *Int J Cancer*. 1981;28:673–676.
151. Clapp C, Martial JA, Guzman RC, et al. The 16-kilodalton N-terminal fragment of human prolactin is a potent inhibitor of angiogenesis. *Endocrinology*. 1993;133:1292–1299.
152. Ferrara N, Clapp C, Weiner R. The 16K fragment of prolactin specifically inhibits basal or fibroblast growth factor stimulated growth of capillary endothelial cells. *Endocrinology*. 1991;129:896–900.
153. Macotela Y, Mendoza C, Corbacho AM, et al. 16K prolactin induces NF-kappaB activation in pulmonary fibroblasts. *J Endocrinol*. 2002;175:R13–R18.
154. Steiner H, Dahlback O, Waldenstrom J. Ectopic growth-hormone production and osteoarthropathy in carcinoma of the bronchus. *Lancet*. 1968;1:783–785.
155. Melmed S, Ezrin C, Kovacs K, et al. Acromegaly due to secretion of growth hormone by an ectopic pancreatic islet-cell tumor. *N Engl J Med*. 1985;312:9–17.
156. Melmed S, Rushakoff RJ. Ectopic pituitary and hypothalamic hormone syndromes. *Endocrinol Metab Clin North Am*. 1987;16:805–821.
157. Greenberg PB, Martin TJ, Beck C, Burger HG. Synthesis and release of human growth hormone from lung carcinoma in cell culture. *Lancet*. 1972;1:350–352.
158. Yokotani T, Koizumi T, Taniguchi R, et al. Expression of alpha and beta genes of human chorionic gonadotropin in lung cancer. *Int J Cancer*. 1997;71:539–544.
159. Broder LE, Weintraub BD, Rosen SW, et al. Placental proteins and their subunits as tumor markers in prostatic carcinoma. *Cancer*. 1977;40:211–216.
160. Scholl PD, Jurco S, Austin JR. Ectopic production of beta-HCG by a maxillary squamous cell carcinoma. *Head Neck*. 1997;19:701–705.
161. McArthur JW, Toll GD, Russfield AB, et al. Sexual precocity attributable to ectopic gonadotropin secretion by hepatoblastoma. *Am J Med*. 1973;54:390–403.
162. Ordonez NG, Ayala AG, Raymond AK, et al. Ectopic production of the beta-subunit of human chorionic gonadotropin in osteosarcoma. *Arch Pathol Lab Med*. 1989;113:416–419.
163. Senba M, Watanabe M. Ectopic production of beta-subunit of human chorionic gonadotropin in malignant lymphoma. *Zentralbl Pathol*. 1991;137:402–404.
164. Treves N. Gynecomastia; the origins of mammary swelling in the male: an analysis of 406 patients with breast hypertrophy, 525 with testicular tumors, and 13 with adrenal neoplasms. *Cancer*. 1958;11:1083–1102.
165. Mendelsohn G, Eggleston JC, Olson JL, et al. Vasoactive intestinal peptide and its relationship to ganglion cell differentiation in neuroblastic tumors. *Lab Invest*. 1979;41:144–149.
166. Said SI, Faloona GR. Elevated plasma and tissue levels of vasoactive intestinal polypeptide in the watery-diarrhea syndrome due to pancreatic, bronchogenic and other tumors. *N Engl J Med*. 1975;293:155–160.
167. Hamilton I, Reis L, Bilimoria S, Long RG. A renal vipoma. *BMJ*. 1980;281:1323–1324.
168. Hollifield JW, Page DL, Smith C, et al. Renin-secreting clear cell carcinoma of the kidney. *Arch Intern Med*. 1975;135:859–864.
169. Soubrier F, Devaux C, Galen FX, et al. Biochemical and immunological characterization of ectopic tumoral renin. *J Clin Endocrinol Metab*. 1982;54:139–144.
170. Atlas SA, Hesson TE, Sealey JE, et al. Characterization of inactive renin ("prorenin") from renin-secreting tumors of nonrenal origin. Similarity to inactive renin from kidney and normal plasma. *J Clin Invest*. 1984;73:437–447.
171. Ruddy MC, Atlas SA, Salerno FG. Hypertension associated with a renin-secreting adenocarcinoma of the pancreas. *N Engl J Med*. 1982;307:993–997.
172. Aurell M, Rudin A, Tisell LE, et al. Captopril effect on hypertension in patient with renin-producing tumour. *Lancet*. 1979;2:149–150.
173. Hauger-Klevene JH. High plasma renin activity in an oat cell carcinoma: a renin-secreting carcinoma? *Cancer*. 1970;26:1112–1114.
174. Genest J, Rojo-Ortega JM, Kuchel O, et al. Malignant hypertension with hypokalemia in a patient with renin-producing pulmonary carcinoma. *Trans Assoc Am Physicians*. 1975;88:192–201.
175. Kawai K, Fukamizu A, Kawakami Y, et al. A case of renin producing leiomyosarcoma originating in the lung. *Endocrinol Jpn*. 1991;38:603–609.
176. Anderson PW, Macaulay L, Do YS, et al. Extrarenal renin-secreting tumors: insights into hypertension and ovarian renin production. *Medicine (Baltim)*. 1989;68:257–268.
177. Tetu B, Lebel M, Camilleri JP. Renin-producing ovarian tumor. A case report with immunohistochemical and electron-microscopic study. *Am J Surg Pathol*. 1988;12:634–640.
178. Cox JN, Paunier L, Vallotton MB, et al. Epithelial liver hamartoma, systemic arterial hypertension and renin hypersecretion. *Virchows Arch A Pathol Anat Histol*. 1975;366:15–26.
179. Saito T, Fukamizu A, Okada K, et al. Ectopic production of renin by ileal carcinoma. *Endocrinol Jpn*. 1989;36:117–124.
180. Fried G, Wikstrom LM, Hoog A, et al. Multiple neuropeptide immunoreactivities in a renin-producing human paraganglioma. *Cancer*. 1994;74:142–151.
181. Yokoyama H, Yamane Y, Takahara J, et al. A case of ectopic renin-secreting orbital hemangiopericytoma associated with juvenile hypertension and hypokalemia. *Acta Med Okayama*. 1979;33:315–322.

137

Endocrine-Disrupting Chemicals and Human Health

R. Thomas Zoeller and Jerrold J. Heindel

OUTLINE

The US Population is Less Healthy Today, 2305
Environmental Chemicals Contribute to Human Disease, 2306
Endocrine-Disrupting Chemicals are not Typical Poisons, 2306
Characteristics of Endocrine-Disrupting Chemicals, 2306
Health Effects of Environmental Chemicals, 2307
  The Lessons of Lead, 2307
  The Lessons of PCBs, 2307
  The Lessons of Diethylstilbestrol (DES), 2309
  Windows of Susceptibility, 2309
  Integrating Mechanistic Information to Interpret Exposures and Health Outcomes, 2310
Mechanisms of EDC Actions, 2310
Dose-Response Characteristics of EDCS, 2311
  Chemical Exposures in the Human Population, 2311
Endocrine Systems: Disease Trends and Effects of EDCs , 2312
  Disease Trends and Effects of EDCs, 2312
  Endocrine Disruptors, the Epigenome, and Developmental and Transgenerational Effects, 2313
  Cancer, 2315
  Thyroid Disruption, 2316
  Neurobehavior, 2316
Emerging Issues, 2316
  Endocrine Disruptors, the Epigenome, and Transgenerational Effects, 2316
  Increasing Exposures Worldwide, 2317
  Expanded List of Diseases Related to EDCs, 2317
  Mixtures, 2317
  E-Waste, 2317
  Green Chemistry, 2317
Conclusion and Guidance for Clinicians, 2317

THE US POPULATION IS LESS HEALTHY TODAY

Although the United States is one of the wealthiest countries in the world, it is not the healthiest by far. The United States is below average compared to peer economies in 9 health domains, including adverse birth outcomes, obesity and diabetes, heart disease, and chronic lung disease.[1] While this health disadvantage is greater in socioeconomically disadvantaged populations, wealthy Americans are also less healthy than their counterparts in peer countries. In their extensive analysis of the health disadvantage in America, an Institute of Medicine (now National Academy of Medicine) Panel concluded that while there is no single cause, the environment is an important contributor.[1] These trends have been recently chronicled by front-line scientists in books aimed at informing the public.[2–5]

Complex, noncommunicable (chronic) diseases and disorders today represent the largest burden on the health care system in the United States.[6] (This statistic does not include the human and economic impact of SARS-Cov-2 infections COVID-19]). In fact, nearly half of the adult population in the United States has at least one chronic condition.[7] These diseases include cardiovascular disease, cancer, diabetes, obesity, reproductive disorders, psychiatric diseases and disorders, thyroid disease, and others. By 2030, it is estimated that 50% of Americans will be obese and the prevalence will be greater than 50% in 29 states.[8] Obesity is associated with co-morbidities including cardiovascular disease, hypertension, diabetes mellitus, and other chronic diseases.[7] Note also that these diseases are linked to a poorer outcome of infectious diseases like COVID-19.[9] This national disease profile is not sustainable, so it is important to address not only appropriate therapies, but also ways to reduce this population disease burden.

The increase in chronic disease is not likely to be related to changes in population demographics or to genetic drift. Because of the availability of sequence information for the human genome and of rapid whole-genome platforms for analysis, genome-wide association studies have been performed to explore the heterogeneous nature of these complex chronic diseases and to identify potential genetic risk factors. These studies, in general, identify important single nucleotide polymorphisms (SNPs) that contribute to disease risk. However, these data indicate that the contributions of SNPs to disease are small.[10] Therefore, elements of the environment represent risk factors for many chronic diseases,[11] and gene-environment interactions are becoming increasingly attractive to identify the major contributions to disease risk.[12] Indeed, it is likely that all noncommunicable diseases have both a genetic and an environmental component. Here we focus on the environmental component, which not only has changed over the last few decades but is also the only disease risk that is modifiable on a population level.

The environment in this regard can be broadly defined as the complex array of nongenetic variables, including but not limited to nutrition, stress, drugs, and environmental chemical exposures. Each of these components is complex in nature and can contribute to disease risk independently and in combination with other environmental components and with genetic profiles. However, unlike with other components, chemical exposures can be modified in part by actions on the population and individual level.[13] Studies have documented the degree to which the human population is exposed to environmental chemicals.[14] The human population is exposed to these chemicals through a variety of mechanisms including food and food packaging, water, work, the home, personal care products, and the environment. The Agency for Toxic Substances and Disease Registry has developed a method for taking an exposure history for primary care physicians that addresses this in part.[15]

ENVIRONMENTAL CHEMICALS CONTRIBUTE TO HUMAN DISEASE

In the Centers for Disease Control and Prevention (CDC) fourth national report on human exposure to environmental chemicals,[14] 212 compounds were measured in blood and urine of participants identified in a statistically representative sample of the US population. In this report, acrylamides, cotinine, trihalomethanes, bisphenol A (BPA), phthalates, chlorinated pesticides, pyrethroids, heavy metals, aromatic hydrocarbons, polybrominated diphenyl ethers, benzophenone from sunblock, perfluorocarbons, and a host of polychlorinated biphenyls (PCBs) and solvents were found commonly in the average US resident.[14] This program—the National Health and Nutrition Examination Survey (NHANES)—has been an enormously valuable tool for researchers to identify chemical exposures and relationships between chemical (and chemical mixture) exposures and human health. These studies have provided important evidence that environmental chemical exposure is contributing to human disease. But as important as this program is to characterize human exposures to environmental chemicals, it is also important to recognize that the number of chemicals measured in this program is only about 0.2% of chemicals licensed for use in the United States.[16] Moreover, about 3000 chemicals are produced at over 1 million pounds per year. While some of these chemicals have been tested for safety to various degrees, most have not been tested at all.

The mechanisms by which chemicals produce toxicity are variable, including chemical reactivity resulting in covalent binding to cellular targets, metabolic production of reactive species, induction of oxidative or nitrosative stresses leading to the production of reactive species,[17] or multiple mechanisms of mutagenicity and carcinogenicity.[18,19] There is also another important mechanism—chemicals that exert adverse effects by interfering with hormone actions—endocrine-disrupting chemicals (EDCs).

ENDOCRINE-DISRUPTING CHEMICALS ARE NOT TYPICAL POISONS

Unlike other toxicants, an EDC is defined by its ability to interfere with hormone actions, independent of its chemical structure. The World Health Organization (WHO) defines an EDC as an "…exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse effects in an intact organism, or its progeny, or (sub)populations."[13] This definition was designed to be applied by government regulators. In contrast, from a biological perspective, the Endocrine Society has defined an EDC as "an exogenous chemical, or mixture of chemicals, that interfere(s) with hormone action."[20] It is important to focus on hormone action in this definition rather than "endocrine function," because it is ultimately the interference with hormone action at the receptor that can cause adverse effects, regardless of where the chemical acts in the pathway of the regulation of hormone action. If an environmental chemical interferes with hormone action at a critical time when hormonal control is actively changing, for example, during development, puberty, or aging, adverse health effects in the human population will most certainly be the result and will be permanent.

The science of endocrine disruption has proceeded rapidly since the mid-2000s as evidenced by reports commissioned by the United Nations Environment Programme (UNEP) and the WHO published in 2013[13] and by the European Commission.[21] These publications make it clear that EDCs are widespread globally; in fact, there are no longer any pristine places on Earth. In addition, EDCs are chemically highly variable; some are lipophilic with long residence times in the environment and in the human body, whereas others are hydrophilic and short-lived. Thus, there are few general statements that can be made about the chemistry of this class of chemicals. In this chapter, we provide information about EDCs: what they are, how they interfere with hormone action, the evidence for their effects in humans, and some thoughts about how these insights may affect clinical care.

CHARACTERISTICS OF ENDOCRINE-DISRUPTING CHEMICALS

EDCs are identified according to their ability to interfere with hormone actions, not by their physicochemical characteristics.[22,23] For example, perchlorate (ClO_4) is a highly water soluble anion that inhibits iodide uptake through the sodium/iodide symporter,[24,25] which can interfere with thyroid function. In contrast, PCBs are highly lipophilic compounds[26] that can interact with an array of nuclear receptors and enzyme systems and also interfere with thyroid function.[27] These are two very different types of chemicals with different physicochemical properties and different pathways through which people are exposed. But both are examples of an endocrine disruptor and both are linked to cognitive deficits in the offspring of exposed pregnant women.[28,29]

Likewise, the particular structure of the chemical is not always indicative of its hormone action. For example, BPA is a diphenolic compound that interacts with the estrogen receptor, and to a lesser degree, the androgen and thyroid hormone (TH) receptors.[30–32] Its structure is quite different from that of estradiol 17β, but its structure is such that it binds to the ligand binding domain of the estrogen receptor nonetheless.[33] Finally, because they interfere with hormone action, EDCs are not typical poisons. That is, their effect is not acute but rather is more sustained, contributing to the prevalence of chronic disease. What this means is that identifying EDCs requires an approach that is different from identifying "traditional" toxins.[21] Thus, identifying endocrine-disrupting properties of environmental chemicals is both challenging and debated. Recently, a new approach to identifying EDCs was published that focuses on the key features that characterize the action(s) of EDCs.[34] There are 10 key characteristics of an EDC as shown in Table 137.1.

These key characteristics of EDCs are useful to show the ways/mechanisms that EDCs can interfere with hormone action. They can also guide determinations of which chemicals act via an endocrine mode of action and therefore are EDCs. A short list of common EDCs—those to which human exposure is common and characterized—is provided in Table 137.2.[35,36]

KEY POINTS

- Industrial chemicals are manufactured for commercial purposes but many of these chemicals are known to interact with endocrine systems in ways that are not typical of a "poison".
- Endocrine-disrupting chemicals (EDCs) exert effects on endocrine systems in ways that are typical of hormones and of drugs designed to target hormone systems. However, these mechanisms are not evaluated by industries or regulatory agencies. Moreover, EDCs are not identified by specific physicochemical properties, but by their actions.

TABLE 137.1 **10 Key Characteristics of an EDC**

Key Characteristic	Description
1	Interacts with or activates hormone receptors by binding to or agonism of hormone receptors
2	Antagonism of nuclear or cell surface receptors
3	Alters hormone receptor expression
4	Alters signal transduction in hormone-responsive cells
5	Induces epigenetic modifications in hormone-producing or hormone-responsive cells
6	Alters hormone synthesis
7	Alters hormone transport across cell membranes
8	Alters hormone distribution of circulating hormone levels
9	Alters hormone metabolism or clearance
10	Alters fate of hormone-producing or hormone-responsive cells

EDC, Endocrine-disrupting chemical.
From La Merrill MA, Vandenberg LN, Smith MT, et al. Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. *Nat Rev Endocrinol.* 2019;16:45–47.

HEALTH EFFECTS OF ENVIRONMENTAL CHEMICALS

At first glance, it may seem a relatively simple endeavor to identify the relationship between chemical exposure and adverse outcome. For example, organophosphate pesticides account for about 200,000 deaths per year in developing countries,[37] the symptoms of cholinesterase inhibition are syndromic, and the severity of symptoms is associated with the severity of exposure. Pyrethroid pesticides also are known to produce acute effects when ingested exposure is greater than about 250 mL, although a significant number of patients exhibit atypical symptoms.[38] In both cases, high exposure is immediately linked to adverse effects, so the cause of the symptoms and the treatment regimen for clinical care are clear. It is also clear when well-known, naturally occurring EDCs produce acute effects, such as the relationship between excess licorice consumption and symptoms of mineralocorticoid excess,[39] or cassava consumption and goiter.[40]

However, identifying the relationship between chemical exposure and health outcome is much more complicated when low-dose chronic exposure produces effects observed at the population level or when exposure occurs at a time of life that is different from the manifestation of effect. Knowledge in this area has been hard won, and the human population has suffered to teach us these lessons. The following cases illustrate a few of these lessons and how they inform us about EDCs in general.

The Lessons of Lead

The acute neurotoxicity of lead exposure in adults has been known for thousands of years.[41] However, it was generally believed that only high exposures would produce toxicity, and only in adults. As a result, in 1924, the Ethyl Corporation was founded in the United States to produce tetraethyl lead (TEL) as a gasoline additive to reduce engine knock in cars. Within a few months of operation there were numerous deaths among company workers and other workers with neurological symptoms. Nonetheless the Surgeon General's report of 1926 approved the use of TEL for the purpose of a gasoline additive, in part because the studies at the time were focused on adult men and the end point was acute lead toxicity.[42] During the ensuing years, there was virtually no independent research on the health effects of lead. The only published research was performed or sponsored by industry, and the results of these studies supported the industry position that low-level lead exposure was natural and safe.

The first independent research funds that were awarded to study the health impacts of lead on children occurred in 1970, resulting in a seminal article by Dr. Herb Needleman, who reported that children's IQs were inversely related to measures of lifetime lead exposure.[43] Despite considerable controversy,[44] the U.S. Environmental Protection Agency (EPA) issued new standards for lead including its removal from gasoline. However, lead was removed from gasoline because it damaged catalytic converters, which were required on all new cars beginning in 1975. Lead consumption in gasoline declined by 50% during the next 4 years, and blood lead levels in average Americans declined by 37%, and by 71% by 1984. As blood lead levels have fallen over time, it has been possible to identify the relationship between lead exposure and IQ at lower levels of exposure. It is currently concluded that there is no safe level of lead exposure for children and that the annual economic cost of lead exposure in children worldwide is in the trillions of dollars.[45]

This history shows first that recognition of lead toxicity derived initially from acute high-dose exposure and severe and immediate consequences. Then, new data began to show that lead could affect child health without the acute effects that served as the primary endpoint of harm. Ultimately, it was determined that there is no safe exposure level for children. This path required nearly 100 years, during which generations of children were exposed to lead.

The early work of Needleman also highlights the importance of identifying the correct measure of exposure. Specifically, Needleman recognized that the half-life of lead in blood was about 30 days, but the half-life in bone was about 30 years. Because bone biopsies were out of the question, he collected shed deciduous teeth, which reflect lifetime exposure to lead. Evaluating lead exposure using a metric that reflects lifetime exposure (in a child) was one of the reasons that exposure measurements were so strongly associated with health outcome. Today it is clear that lead exposure *in utero* and in early childhood when the brain is developing will lead to brain damage, that environmental lead—especially in paint—is an important source of exposure to control, and that lead in paint is a worldwide source of adverse public health effects.[46]

The problem of lead exposure is not one relegated to the past only. The water crisis in Flint, Michigan, demonstrated that lead in drinking water is a pervasive problem and that vigilance is required.[47] Moreover, the events leading up to this crisis were not only a lesson in poor governance and unethical behavior, but also a lesson in the ways in which scientific knowledge is understood by various actors.[48] Finally, the magnitude of the crisis was exacerbated by the sheer complexity of the laws regulating how government agencies maintain and monitor safe drinking water.[49] Considering what we know about the health effects of lead exposure with certainty, the failure of public health and government agencies to avoid the drinking water crisis in Flint demonstrates that health care professionals need to remain vigilant for the role of the environment—broadly defined—in the health and care for patients.

The Lessons of PCBs

PCBs are a mixture of synthetic organic chemicals (biphenyls) that have been chlorinated to different degrees (different numbers of chlorine atoms). These chemicals were used in a wide variety of applications from electrical equipment because of their insulating properties and resistance to high temperatures, wood finishes because of their optical properties, caulking because of their ability to soften the material and to remain stable for long periods, and many others.[50,51] In

TABLE 137.2 Selected Common Endocrine Disruptors in the Human Population

Chemical	Population With Measured Levels (%)	Source of Contamination	Effects
Lead	>90	Degradation of lead-based paint and the contaminated dust that it forms.	Effects on brain development and IQ. Multiple endocrine and toxic effects.
Mercury	>90	Diet is a significant source of blood mercury, especially in some seafood species. Some manufacturing can produce atmospheric mercury that enters the air and ultimately the food chain.	Effect on brain development, learning, and memory. Effects on thyroid and other endocrine and toxic effects.
Per- and Polyfluoroalkyl Substances (PFAS)[a]	~100	Used to repel oil and water from clothing, furniture, and food packaging such as pizza boxes and fast-food containers. Firefighting foams, cleaners, some paints, roof treatments, and hardwood floor products. Persist and bioaccumulate in environment and have long half-lives in humans.	Many different congeners. Linked to thyroid disease as well as several cancers.
Polybrominated Diphenyl Ethers (PBDEs)[b]	68–100	Used as flame retardants in foam, furniture, electronic equipment, and high-impact plastics. There are many individual congeners that have different exposure profiles in the population. Prevalent in the built environment, especially in dust. Persist and bioaccumulate in environment and have long half-lives in humans.	Many different congeners. Some interact with estrogen, androgen, and thyroid hormone actions. Metabolites may be most important.
Polychlorinated Biphenyls (PCBs)[c]	>95	Production was banned in the 1970s, but they are still used in a number of commercial and personal applications. They were used as dielectric fluids, wood treatments, printing ink, and in many other products. Because of their persistence and ability to bioaccumulate, they are still widely present in the environment.	Many different congeners. Some interact with estrogen, androgen, and the actions of thyroid hormones. Metabolites may be most important.
1,1-Dichloro-2,2-Bis(p-Chlorophenyl) Ethylene (DDE)	100	Metabolite of the insecticide dicloro-diphenyl-trichloroethane (DDT). Because of previous extensive and wide use, and because of its relatively long environmental half-life, DDE is ubiquitous in human samples.	Interacts with estrogen signaling. Associated with hormone-related cancers and precocious puberty.
BPA	>95	Can linings, epoxy resins, some plastics, thermal receipts, others	Interacts with estrogen receptor and at higher doses also interacts with the androgen and thyroid hormone receptor. Linked to a variety of chronic diseases including obesity, cardiovascular disease, and some cancers.
Triclosan	>85	Antimicrobial soaps and other products, clothing, kitchenware, body washes, and some cosmetics.	Appears to interact with thyroid hormone through several mechanisms.
Parabens	>90	These are esters of parahydroxybenzoic acid. Present as preservative in many consumer products including nail polish, hair spray, lipstick, fragrance, body wash, sunscreen, etc.	Class of chemicals with estrogenic activity.
Phthalates[d]	100	Different phthalates are used in different applications including personal care products, PVC plastics, some flooring, car products, insect repellents, etc.	Can interact with androgen and thyroid hormone systems. Associated with lack of masculinization in boys including reduced anogenital distance and lower sperm count in men.
Perchlorate	100	Widely present in foods including fruits and vegetables, beer and wine, and other products. It is FDA approved for use in food packaging.	Blocks iodide uptake into the thyroid and reduces serum thyroid hormone levels in animals. Linked to lower thyroid hormone levels in some women, and to negative cognitive impacts on children.

[a] Perfluorinated chemicals including perfluorooctanesulfonic acid (PFOS) and perfluorooctanoic acid (PFOA).
[b] Polybrominated diphenyl ethers. Many different congeners based on bromine number and placement.
[c] Polychlorinated biphenyls. Many different congeners based on chlorine number and placement.
[d] There are several variations of chemical structure for phthalates.
BPA, bisphenol A; *FDA*, Food and Drug Administration; *PCBs*, Polychlorinated biphenyls.
Data are derived in part from Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. *Environ Health Perspect.* 2011;119(6):878–885 and from Calafat AM, Ye X, Wong LY, Bishop AM, Needham LL. Urinary concentrations of four parabens in the U.S. population: NHANES 2005-2006. *Environ Health Perspect.* 2010;118(5):679–685.

1899, just a few years after they were first synthesized in the laboratory, a condition called chloracne (a painful skin disease) was noted in PCB production workers. In 1936, workers in a production plant also were affected by chloracne and liver disease, and some subsequently died. Animal studies showed that high doses of PCBs can cause liver damage. In the late 1960s, it was reported that PCBs were found at high concentrations in Baltic Sea animals, and, in 1968, a human population was accidently exposed to high levels of PCBs as a result of rice-oil contamination in Japan.[52] This was followed by studies showing that PCBs were found throughout the environment and were bioaccumulating in a variety of animals including fish. In 1972 Sweden banned the use of PCBs in situations where environmental contamination was likely. In the United States, Congress banned all manufacturing, processing, and distribution of PCBs in 1979. That same year there was another serious accident in which PCBs contaminated rice oil, this time in Taiwan (using the same manufacturing equipment that produced the accident in Japan).[53]

Until the 1990s, the general theory was that PCBs were toxic because they interact with the aryl hydrocarbon receptor (AhR),[54] thereby acting like dioxin. Of the 209 different PCB congeners, some are "dioxin-like", and some are not. This concept led to categorizing PCB mixtures in terms of their "toxic equivalence" (TEQ)[55]; that is, their ability to exert dioxin-like effects. However, in 1990, there was a report in nonhuman primates and in cell culture that nondioxin-like PCB congeners could affect dopamine levels in specific brain areas.[56]. This spawned research focused on congener-specific effects of PCBs on health outcomes including cognitive function. It is now well known that background exposure to PCBs in pregnant women is associated with lower IQ as well as other cognitive deficits in their offspring.[29] Moreover, these studies were done 20 years or more after the ban on PCB production, demonstrating that these environmentally persistent chemicals have long-lasting impacts on human health. Today the Food and Drug Administration (FDA) considers PCB contamination in food to be "unavoidable" and set limits in 2010 for PCBs in various foods. For example, the acceptable limit of total PCBs in milk or cheese is 1.5 mg/kg (fat, not volume).[57] Considering that the half-life of PCBs in the human body is measured in years,[58] these FDA limits do not appear to be health protective.

Different PCB congeners appear to interact with different systems to impact brain development. For example, some PCB congeners can be hydroxylated by the P450 enzyme CYP1A1 and interact with the thyroid hormone (TH) receptor.[59,60] There are other PCB congeners that can displace T4 from the serum transport protein, transthyretin.[61,62] Still others can activate liver enzymes that increase T4 clearance from serum.[63] Finally, some PCB congeners can interact with the ryanodine receptor, an important component of calcium signaling.[64]

Again, an important lesson is that early recognition of toxicity was related to very high exposures and acute symptoms including death. However, the assumption that toxicity was limited to high exposure levels was again mistaken, and this caused a delay in the recognition of more widespread toxicity. We now have very good mechanistic data demonstrating that specific PCB molecules (congeners) can directly interact with, for example, the thyroid system at low concentrations.[65] In addition, other PCB congeners can interact directly with the ryanodine receptor to influence intracellular calcium signaling at very low concentrations.[64] Both of these pathways are important for brain development and may well help explain much of PCB's effects on cognitive function.

The Lessons of Diethylstilbestrol (DES)

DES is a relatively potent nonsteroidal estrogen that was prescribed to pregnant women from the 1940s to the 1970s ostensibly to prevent miscarriages based on the belief that spontaneous abortion was the result of a decrease in estrogen levels (now shown to be a consequence, not a cause). DES was marketed using more than 200 names, and during that time tens of millions of women and their offspring worldwide were exposed to this drug. The first indication of a problem occurred when DES was linked to a rare form of vaginal cancer, vaginal clear cell adenocarcinoma, in a small number of young women exposed *in utero*.[66] It was subsequently shown to be associated with a number of reproductive problems in the DES-exposed daughters including increased uterine fibroids, endometriosis, breast cancer, miscarriage, preterm delivery, and ectopic pregnancies later in life, as well as reproductive tract malformations.[66] Moreover, data have demonstrated abnormalities in adult male offspring of women taking DES during pregnancy.[67–69]

A critical lesson from the DES experience—and one that could not have been gained without formal records of exposure and the astute perception of a physician—was that exposure to a synthetic estrogen during fetal development could increase the risk of various adverse effects in the adult offspring, decades after the chemical has disappeared from the body. We know from many population-based endocrine studies that developmental hormone exposure can have lifelong impacts. For example, serum T_4 levels at age 2 are associated with verbal IQ in adulthood.[70,71] Thus, the findings of DES exposure during fetal development are consistent with our understanding of the impacts of various hormones on development and the manifestation in adulthood of developmental endocrine interference. The effects of DES exposure shown in humans have been reproduced in animal studies; in fact, some effects were first observed in rodent studies and were then discovered in humans. Thus, DES is an example of an estrogenic agent that produces adverse effects in the human population. Not only do these studies demonstrate a proof of principle for developmental exposures to estrogenic compounds leading to birth defects, but they also demonstrate the temporal dissociation between exposure and manifestation of effects—the so-called "developmental origins of health and disease" (DOHaD).[72]

An important lesson common among the three examples is that the risk of adverse outcomes of chemical exposures was initially underestimated in each example. This is also a common feature in a broader analysis of human health effects of chemical exposures.[73,74] This failure to recognize the risk is due in part to a reliance on insensitive measures of chemical effects. In the case of lead and PCB exposures, there was reliance on measures of acute toxicity. In the example of DES, it was a reliance on measures of adverse outcome in the mother, not in the fetus. Therefore, as independent research progressed, more sensitive measures of adverse outcome were implemented. In the case of DES, the recognition that fetal exposure could lead to adverse outcome 20 years later was a novel and unexpected insight. These are not just lessons of academic interest but are chronicles of avoidable harm to fellow humans on a population level and over generations.

KEY POINTS

- We have highlighted the history of 3 chemicals that illustrate important features of EDCs: Lead, Polychlorinated Biphenyls (PCBs), and Diethylstilbestrol (DES). In each case, the risk to human health was dramatically underestimated because the outcome measures evaluated were insensitive. Generations of people have been damaged by these mistakes.

Windows of Susceptibility

Hormones are well known to produce effects on development during very specific developmental times ("windows"). For example, the "critical period" of TH action was initially viewed as the early postnatal period.[75] However, it is now recognized that there are different critical periods of TH action for different development events and cognitive domains.[76,77] Likewise, testosterone acts during specific times to affect male phenotypic development,[78] which is very different from the effects of testosterone in the adult.[79] Moreover, the effect of androgens in adulthood can depend on the action of testosterone during development.[80]

Considering this background of information about critical periods of hormone actions on different organ systems, it is not surprising that EDCs would be found to have different effects when exposures are studied at different life stages. Awareness of the developmental effects of chemical exposures has been heightened by studies of DES exposures (above) and of fetal alcohol exposure.[81] Focus on EDC action has been, for the most part, limited to specific windows of sensitivity including *in utero* exposures, neonatal exposures, and adult exposures. It is not clear what mechanisms account for a sensitive window of exposure to EDCs and thus how many sensitive time points there might be across the lifespan, largely because this issue is not well understood for hormones either. Nonetheless it is likely that preconception, preimplantation, early development, puberty, aging, as well as pregnancy itself are particularly sensitive times for the adverse effects of environmental exposures. Thus, it is clear that an approach is needed that includes not only a lifespan approach but also examination of effects across generations. There are now multiple examples of EDC effects that are heritable across not just one generation but also two, three, and four generations.[82,83] Thus exposures during pregnancy may lead to increased disease susceptibility that may last for generations.

Everyone is exposed across their lifespan starting with fertilization. Actual exposures depend on where one lives and one's lifestyle, such that exposures within and between chemicals can be quite variable. However, it is not just the level of exposures that are key to causing health concerns but perhaps more importantly when the exposure occurs. It is now absolutely clear that *in utero* and/or early in life, when tissues are forming, they are very sensitive to the adverse health effects of EDC exposures leading to an increase in noncommunicable diseases. There are two reasons for this sensitivity to the action of EDCs. First, development is a highly sensitive and orchestrated process controlled by hormones and growth factors. They act by controlling gene expression to stimulate the development of specific tissues. Thus, any EDC that interferes with endogenous hormone signaling will upset the normal control of gene expression leading to the development of a tissue that may "look" normal but that will not function normally. This phenomenon, first observed in the nutrition field, is now guiding research in the environmental chemical field and is called the DOHaD.[72] The second reason for this sensitivity to EDCs is that the enzymes that metabolize chemicals are not fully developed so the EDCs are active for longer than "normal times" along with a lack of DNA repair and a lack of blood/brain/organ barriers.[84] Because tissues develop at different times, the effects of EDCs on tissue development will depend on tissue specificity and timing of their development. Data in support of this DOHaD concept are sufficiently robust and repeatable across species to support incorporation into clinical practice.[85] The difficulty arises in that the exposure *in utero* may not result in increased disease until years or decades later, making it difficult to link the exposure to the disease. Thus, the clinical focus needs to be on reducing the risk of adult disease by limiting exposures before and during pregnancy and early childhood.

Exposures to EDCs occur across the lifespan. Adult exposures to EDCs can also cause health effects, but those effects typically remain only while the exposure exists; for example, the effects are reversible, and they may require higher concentrations due to the fact that the tissues are fully formed.

Integrating Mechanistic Information to Interpret Exposures and Health Outcomes

Mechanistic information from experiments using animal models or cell-based/biochemical studies provide important information to support epidemiologic data linking chemical exposures to human health outcomes at both the individual and population level. This body of research is highly multidisciplinary—from cell-free and cell-based experiments, to animal models across most vertebrate classes and some invertebrates, including the use of various genetic strains of animals. This of course requires high-quality research in these various domains and this level of information is not available for a large number of chemicals. A good example is BPA for which there are well over 1000 publiations.[86] Mechanistic research shows that exogenous chemicals can interact with biological systems much the way drugs do—by interacting with hormone receptors, biosynthetic enzyme systems, catabolic and metabolic pathways, and with various transport processes.[23] The regulation of hormone action is highly complex, and the consequences of interfering with hormone action are likewise highly complex. Mechanistic research since the mid-1990s has provided insight into the many ways in which chemicals can and do interfere with hormone action. We present here examples that provide some insight into these issues.

MECHANISMS OF EDC ACTIONS

Chemicals can interfere with hormones anywhere along the chain of events that control hormone action, from synthesis, release, and transport of the hormone to hormone effects on receptor function or on hormone degradation and elimination (metabolism).[34] In general, this means that chemicals can either interfere with the delivery of the hormone to and from its target tissues and cells or interfere with the action of the hormone on the receptor—or both. There are situations in which EDCs can interfere with the delivery of some hormones to specific target cells; in this case the effects are tissue-specific. Likewise, there are known cases in which EDCs must be metabolized before they can act on a hormone receptor, and this metabolism occurs only in certain tissues.[59,87] Thus, there are few rules about EDC actions that can be generalized to all cells and all organs of the body and all EDCs.

Hormone receptors have a high affinity for their natural ligand, but they often have a much lower affinity for endocrine disruptors—with some exceptions. However, it is important not to confuse *affinity* (ability to bind) for the receptor with *potency* (dose at which a given effect occurs) or *efficacy* (maximum effect). It is clear from clinical chemistry that differences in affinity do not always translate to differences in potency or efficacy. Recently, for example, it was shown that mutations in the gene coding for the estrogen receptor alpha (ESR1) that are associated with estrogen insensitivity in patients alter the effects of ligand binding without affecting affinity.[88] This can be true for exogenous ligands. For example, the endocrine disruptor BPA has the same potency as 17β-estradiol in pancreatic beta cells,[89] although its affinity for endoplasmic reticulum the estrogen receptor alpha (ERalpha) is much lower compared to that of 17β-estradiol. The way in which BPA exerts the same effect on the ER as the natural estrogen in one cell type is not known, but it should be emphasized that affinity measurements usually are made in completely artificial circumstances that do not reflect the environment of the cell. Moreover, these studies also emphasize that endocrine disruptors can act on a hormone receptor in one tissue type without affecting the receptor in another tissue type. Thus, just like the hormones they mimic or disrupt, the effects of endocrine disruptors will be receptor, tissue, and coregulator specific. Some endocrine disruptors actually have an affinity that is similar to or greater than that of the natural ligands. An example of this is tributyltin (TBT), which has an affinity for the nuclear proteins, Retinoid-X-Receptor (RXR) and Peroxisome Proliferator-Activated Receptor gamma (PPAR(gamma symbol)) in the low nanomolar range. Indeed, it is the most potent agonist known for these receptors.[90] See Table 137.2 for a list of examples of EDCs and how they interact with hormone systems.

DOSE-RESPONSE CHARACTERISTICS OF EDCS

A significant challenge in the area of EDCs is to understand fully the dose-response characteristics of EDCs on various elements of hormone action in a manner that provides useful insight into the potential for human health effects. Within this context, the concepts of *homeostasis* and *compensation* need to be considered carefully. The term "homeostasis" refers to the property of a dynamic system that maintains "internal constancy." The term "compensation" is related in that when a homeostatic system is perturbed, various elements within the system are activated or inhibited to compensate for the perturbation. For example, when glucose levels rise, insulin is secreted and glucagon secretion is inhibited such that glucose levels fall again. But homeostatic systems can reset in the face of chronic perturbation, as illustrated by baroreceptor resetting in the face of even short-term changes in mean arterial pressure.[91] Thus, in the face of chronic (lifetime) exposure to low doses of environmental chemicals, we need to be cautious about using the concepts of homeostasis and compensation when no studies have been designed to inform us about whether—or how—homeostatic endocrine systems may change in the face of these chronic exposures.

These issues are particularly important when considering the potential adverse effects of low-dose exposure to the fetus compared to the adult. Fetal development is a time when disruption of hormone action may not be reversible because early developmental events become the substrate for subsequent events (see "Windows of Susceptibility"). In addition, the fetus may have fewer homeostatic and compensatory mechanisms to ameliorate effects of EDCs. Finally, a variety of hormone receptors are much more abundant during development than in the adult, and this may also serve to make the fetus much more sensitive to chemical interference.[92] Thus, adults may be both less sensitive to EDC actions and more able to compensate—if homeostatic set points do not shift with chronic lifetime exposure.

Considering this, it is important to recognize that the field of "endocrine disruption" is focused on a novel area within the established fields of both endocrinology and toxicology, and the dose-response characteristics observed in short-term, high-dose toxicity studies may bear little resemblance to the dose-response characteristics in real life. Moreover, because endogenous hormones exhibit nonlinear dose responses, it is expected that environmental chemicals interacting with hormone systems will also exhibit nonlinear responses. These nonlinear dose responses can take several forms.[13] In its simplest form, because hormones act on receptors, which are limited in number, the response itself is "saturable." That means that there is a dose of hormone—or endocrine disruptor—beyond which there is no further response on a measured outcome. Hormones and EDCs can also produce nonmonotonic dose responses. Technically, the term "nonmonotonic" simply means that the slope of the curve changes sign (positive to negative or vice versa) over the dose-response range.[93] For example, when fetal mice are exposed to low or high doses of DES, a synthetic estrogen, their adult prostate weights are relatively low. However, intermediate doses of DES produced significantly heavier prostates.[94] Research also suggests that nonmonotonic responses can be extended to the population level. For example, individuals in the highest quartile of environmental exposure to 1,1-Dichloro-2,2-Bis(p-Chlorophenyl) Ethylene (DDE) (an estrogenic Dichlorodiphenyltrichloroethane [DDT] metabolite) have decreased body mass index (BMI) and blood triglyceride levels compared to individuals in the third quartile.[95] Moreover, women exposed to the lowest and highest doses of dioxin had no changes to the age at which they entered menopause, whereas those women exposed to intermediate doses showed an increased risk of early menopause.[96] Finally, low or high free T4 in first trimester women were both associated with IQ deficits in the offspring.[97] The mechanisms underlying these effects have not always been defined, but it is important to recognize that these dose-response characteristics are fully within the realm of our understanding of hormone action and endocrine disruption.[93]

Chemical Exposures in the Human Population

If the human population is being affected by exposures to EDCs, then those exposures must be sufficient in abundance and timing to impact human health outcomes. The CDC manages a biomonitoring program in part to characterize human exposures to environmental chemicals in a nationally representative manner. The fourth national report on human exposures to environmental chemicals was published in 2009,[98] and this information is updated regularly on the CDC website (http://www.cdc.gov/exposurereport/). Although the fourth national report makes it clear that the presence of a chemical in the human population does not mean that it causes disease, it is also clear that the US population is simultaneously exposed to a large number of chemicals with a wide range of chemistries and biological actions, and that most of these chemicals are not tested for safety. This is true for the general population as well as for subgroups that might be expected to be particularly vulnerable, including pregnant women[35] and newborns, and people living near areas of intense pollution. In fact, in a 2005 study of cord blood samples, nearly 300 chemicals were identified, with each cord blood sample containing at a minimum more than 100 different chemicals.[99] This means that babies are born pre-polluted, and it is likely that some of those exposures will cause adverse effects that could lead to increased sensitivity or susceptibility to disease later in life. Thus, the human population is exposed to a great many environmental chemicals throughout the life cycle, and most of these chemicals are not tested for safety.

It should also be noted that the chemicals included in the NHANES biomonitoring program are only a subset of chemicals to which the human population is exposed. The reason for this is that analytical methods for identifying chemicals are not available for the large number of chemicals licensed for commerce (more than 70,000), or even the several thousand high-production volume chemicals. In addition, there are estimated to be about 10,000 synthetic chemicals approved as food additives, the majority of which include little or no toxicity data.[100] Thus, we do not know the complete picture about human exposures to anthropogenic chemicals, which increases the difficulty in linking exposures to human health. Nonetheless, it remains important to know that everyone is exposed to uncharacterized chemicals throughout his or her lifespan.

An important lesson that is reflected in the experience of lead toxicity, PCBs, and DES is that ambient low-dose exposure and/or exposures during critical periods of development can produce adverse effects that go unnoticed without considerable and costly study. Furthermore, these adverse effects can be predicted and mechanistically understood by experimental studies in animals as well as cell-free and cell-based biochemical and molecular studies. Therefore, considering this large and largely unregulated chemical exposure pattern, studies designed to identify relationships between chemical exposure and human health outcomes particularly need to be incorporated into a logical framework quickly so that the findings can reliably inform patient care and public health policy.[101]

KEY POINTS

- The human population is heavily contaminated with industrial chemicals, many of which have not been tested for health effects and very few of which have been tested for activities other than acute toxicity. In fact, chemical measurements of cord blood demonstrate that babies are born pre-polluted with at least one hundred known chemicals.
- Regulatory agencies have not evolved to test the ability of chemicals to interact with endocrine systems, nor to test the ability of chemicals to produce chronic disease.
- EDCs can produce effects in animal studies at concentrations consistent with human exposures and with concentration-response characteristics that do not fit the regulatory assumption that the "dose makes the poison."

ENDOCRINE SYSTEMS: DISEASE TRENDS AND EFFECTS OF EDCs

This section focuses on the state of the science linking EDC exposures to human disease. There are a number of lessons that we have learned from this work, which should be considered when reviewing data from human health studies. These are highlighted as follows:

- Outcome measures linked to EDC exposures in humans occur at exposure concentrations lower—often considerably lower—than predictions based on regulatory studies.
- EDC effects in the human population depend on genetic background creating a heterogeneity in responses that require large population studies using a genomic and statistical approach.
- EDC effects are almost certainly the result of mixtures of EDCs, as that is what humans are exposed to, and also are a result of low environmentally relevant concentrations that are difficult to measure that require highly sensitive assays of mixtures.
- The effects will be a result of "multiple hits" across the lifespan with the most sensitive exposure window during development (*in utero* and early life), which will require a lifespan approach.
- EDC effects may occur after a long latent period including effects transmitted across multiple generations (transgenerational), requiring long-term prospective studies.
- Effects will vary depending on other environmental stressors: medications, stress, socioeconomic status, infections, and microbiome, which will require the analysis of confounders.
- Effects can vary according to sex, tissue, dose, and route of exposure. Indeed, effects noted at high doses may be different from those at lower doses.
- Epidemiological studies identify potentially avoidable risks to individual health much like smoking, drinking and dietary choices. This can be applied to adults as well as pregnant women and children.

Disease Trends and Effects of EDCs

Temporal trends in the incidence and prevalence of disease in the human population can provide circumstantial evidence for a causal relationship between chemical exposure and human disease.[13] Although these trends do not constitute proof of causation, they are important elements of the evidence. Chronic disease has increased during the past 50 years and is now more prevalent than infectious disease worldwide, with the exception of COVID-19.[102] Public agencies around the world have invested heavily in research on EDCs in large part because the human population manifests increased chronic diseases such as obesity, diabetes, infertility, and certain cancers.[103,104] increased incidence of anomalies of male reproductive development[105] and various aspects of fertility. In addition, there are also clear increases in the incidence and prevalence of neurobehavioral disorders including attention deficit (AD),[106] autistic spectrum (ASD),[107] and measures of cognitive function. One of the most visible changes in the human population is the increased incidence and prevalence of obesity,[108] especially in the pediatric population,[109] with attendant increases in type 2 diabetes and heart disease. The incidence of some cancers is also increasing, including cancers of the brain, breast, male and female reproductive tract, and thyroid.[110,111] A disease trend that is increasing during just a few decades and that cannot be accounted for by changes in diagnostic criteria is important evidence for a role of environment in disease because genetic drift does not occur during such a short time frame. EDCs play an important role in the etiology of such disease trends.

The Endocrine Society's Second Statement on EDCs,[23] which provides an overview of the effects of EDCs on many of the most prevalent human diseases in 2015 along with a shortened executive summary.[112] This review states, "The Endocrine Society's first Scientific Statement in 2009 provided a wake-up call to the scientific community about how environmental EDCs affect health and disease. Five years later, a substantially larger body of literature has solidified our understanding of plausible mechanisms underlying EDC actions and how exposures in animals and humans—especially during development—may lay the foundations for disease later in life".

Obesity. The prevalence of obesity has increased dramatically both in the United States and in many countries worldwide since the mid-1990s[13] such that there are currently more than one billion overweight adults, with 400 million being classified as obese; 20 millions of these are children. Obesity is an intractable problem in that nearly 90% of those who lose weight gain it back within a year.[113] Obesity is a complex endocrine condition caused by an interaction between genetics and environment with the environmental component consisting of nutrition, drugs, behavior and some EDCs. Conventional wisdom posits that obesity is an energy-balance disorder resulting from excess calories and decreased exercise. However, both the increased incidence during the past few decades and data showing that only a subset of people become obese suggest that there is more to obesity than this simple explanation. Moreover, the incidence of obesity among infants is also increasing, which would appear to counter the energy-balance hypothesis. Hormones such as estrogens, androgens, glucocorticoids, insulin, and thyroid hormones play important roles in controlling adipose tissue development, metabolism and satiety, and body-weight regulation is also sensitive to environmental chemicals and drugs that can affect these hormonal pathways. Indeed, some drugs include a side effect of increasing weight gain, such as some antipsychotics[114] and antidiabetic drugs such as Avandia.[115] The fact that there are drugs with side effects of weight gain provides proof of the principle that the endocrine systems controlling weight gain and metabolism are sensitive to perturbation.

Events occurring in early life—during pregnancy and early childhood—may increase the risk of obesity.[116] There are considerable data linking altered development to increased susceptibility to disease later in life including obesity (DOHaD). Although the initial focus of these studies was on altered nutrition during development, it is now clear that there also are EDCs that can cause altered developmental programming, likely because of altered epigenetic regulation of gene expression, potentially leading to weight gain later in life. These EDCs have been called obesogens.[117] It is key to note that exposures to obesogens during development can permanently alter programming of the system regulating body weight, producing an altered "set point" for adipose tissue formation, regulation of satiety, and metabolic rate that persists throughout life. Tributyl Tin, for example, induces an increase in fat cell number and these fat cells are abnormal in their response to hormones including impaired glucose uptake and insulin signaling.[118] In fact, even after body weight reduction, there are long-term changes in a profile of hormones that support obesity.[119] In addition, EDC exposure can trigger changes in the hypothalamus that play a particularly important role in feeding behaviors. Improper hypothalamic programming may adjust metabolic "set points" in adolescents and adults, and these adjustments may help explain differences between the eating behaviors of lean and obese individuals.[120] Thus, the obesogen hypothesis offers an explanation for both the increase in obesity rates and also the differences between lean and obese individuals.

The obesogen hypothesis was developed in the early 2000s in response to the startling rise in the incidence and prevalence of obesity in the human population, especially in children.[121] Although a relatively new field of study, there is now strong evidence in experimental animals showing that developmental exposure to a variety of environmental chemicals, including those that activate PPARγ such as TBT, and estrogenic chemicals such as BPA, some organochlorine and organophosphate pesticides, and chemicals acting through other mechanisms such as elements of air pollution, lead, perfluoroctanoic acid, phthalates, and nicotine, can lead to weight gain later in life.[13] There are now about 50 chemicals and chemical classes classified as obesogens.[122] These animal studies provide significant evidence for a role of EDCs in affecting specific elements of body weight regulation, including appetitive behavior and metabolism. An important example is that of TBT, which is a potent activator of the nuclear receptor PPARγ, a key receptor involved in the development of new adipocytes.

Human studies are more difficult to perform and interpret in part because humans are exposed to complex chemical mixtures and are more genetically variable as well as variable in their environmental conditions.[123] Nonetheless, animal studies have produced data that are concordant with epidemiological data linking exposures to obesogens during development to increased body weight in infants and later in life. Some of the strongest associations between exposures during pregnancy and obesity in their offspring include smoking, BPA, persistent organic pollutants, DDT/DDE, per- and polyfluoroalkyl substances (PFAS), and air pollution.[120]

It has been proposed that the confluence of developmental programming of metabolic set points by obesogens in association with continued obesogen exposures throughout the lifetime, overconsumption of processed foods containing added sugars, and EDCs from the packaging materials, together with decreased physical activity throughout life, creates an important combination that is driving the obesity epidemic.[124,125]

The obesogen hypothesis proposes both a mechanism for the increased epidemic of obesity and also a solution. If obesity results in part from continuous exposures to obesogens from development and later, then for the first time there is actually a path for individuals and for physicians to reducing the risk of obesity by reducing these exposures. Thus, rather than waiting for obesity to occur and trying with poor success to intervene and cause weight loss, we may be able to reduce the risk of overweight and obesity from the beginning.

Diabetes. Type 2 diabetes—especially in the pediatric population—is becoming recognized as a global epidemic.[126] Diabetes has increased dramatically since the mid-1990s to 347 million in 2011.[13] And, parallel to the incidence of obesity, there has been a significant increase in childhood type 2 diabetes, something that was rare just a few decades ago. Although type 2 diabetes can occur by itself, 70% of the risk associated with it is attributed to weight gain, indicative of the direct link between increased adiposity and insulin resistance. In animal studies, many chemicals that cause weight gain also disrupt glucose management. EDCs including phthalates, BPA, some flame retardants, arsenic, some persistent organic pollutants such as PCBs, and pesticides have been linked to the development of type 2 diabetes in both animal studies and in human epidemiologic studies.[120,127] Overall, there are significant data in both animal and human studies that support concern about the relationship between chemical exposures and the production of type 2 diabetes. In contrast, there is little information about the link between EDC exposures and type 1 diabetes. However, there are environmental chemicals that exhibit both EDC activity and immunotoxicity, which would be prime candidates for EDC effects on this hormonally and immune-mediated, largely childhood, disease.

Endocrine Disruptors, the Epigenome, and Developmental and Transgenerational Effects

Parents pass on their genes to their children and in so doing pass on various traits associated with those genes. The combination of all genes in a species is referred to as the "genome," and "genomic" studies refer to those studies designed to understand how various patterns of genes are controlled; but cells in the body can pass on heritable traits to their cellular progeny without altering their genome. Certainly, EDCs can also act across the lifespan just as hormones do. However, there is a major difference in effects noted after developmental exposures vs adult exposures.

The ability of endocrine disruptors to alter the normal hormonal control of development is perhaps the most significant consequence of exposure, because developmental effects will occur at lower doses than are required for effects in adults. Additionally, the effects of exposure to endocrine disruptors during development will remain throughout life and may have a long latency between exposure and disease outcomes; whereas, in many cases, exposures in adults remain only while the chemical is present. These differences in sensitivity to EDCs and persistence of effects are related to the epigenetic state of the organism.

During development, a single cell—the ovum—will divide, multiply, and differentiate into many cell types and tissues and ultimately will become a person. Development from this perspective is a process of "fate restriction"—permanently turning on or turning off different combinations of genes required for a cell to be a functional cell in the liver, kidney, brain, etc. The traits a cell passes on to its cellular progeny are controlled by "epigenetic" mechanisms. Epigenetics is broadly defined as those heritable changes not dependent on genetic sequences, and it is these epigenetic processes that define and control tissue development by controlling gene expression. Thus, a major route by which hormones act during development is by changing the epigenome and the combination of genes that can or cannot be expressed. The fact that endocrine disruptors can influence epigenomic mechanisms during the times when tissues are forming is likely the mechanism whereby development (*in utero* and early childhood, when tissues are developing) is a sensitive window for EDC action. It is also the reason that developmental exposure leads to effects long after the exposure is gone and may occur after a long latent period across the lifespan. In adults, the tissues are already formed, and the epigenetic marks are relatively stable; thus, EDCs may not act on the epigenome but may act directly on genomic pathways—effects that are more easily reversible when exposure is controlled.

Endocrine disruptors have also been shown to produce transgenerational effects as a result of their ability to alter epigenetic processes. This issue first arose with studies in which an antiandrogenic pesticide (vinclozolin) was administered to developing mice at a single time during development—when the testis was in a critical period of development. Not only did vinclozolin produce adverse effects on the developing testis, but also this was transmitted through four generations of mice. This effect is likely caused by epigenetic changes that were transmitted with high fidelity from one generation to the next via the germ cells (sperm). A number of endocrine disruptors have now been shown to influence epigenetic mechanisms and to produce effects in several generations of animals.[129] Grandmother exposures may be affecting the F2 and subsequent generations.

Reproductive System. Because the reproductive system was an early focus of studies on chemical toxicities in the 1930s and 1940s, it should be no surprise that there is abundant literature on EDCs that interfere with estrogen or androgen actions. In addition, the field of reproductive endocrinology has significantly matured both in terms of our understanding of the medicinal chemistry associated with estrogen and androgen receptors (ARs), but also in terms of the role of sex steroids in development, in puberty, and in the adult. Thus, there is a significant amount of information now about environmental chemicals and different aspects of reproduction, as discussed in the following text.

Male Reproduction. One of the most significant concepts in this field is that of the testicular dysgenesis syndrome (TDS).[130] Skakkebaek et al. proposed that poor semen quality, testis cancer, undescended testis, and hypospadias are symptoms of one underlying entity caused by defects in androgen action during fetal development. One of the most compelling observations in support of this hypothesis is a report of a novel AR mutation in a family that exhibits all of the hallmark characteristics of the TDS.[131] This mutation (c.2214T>G;p.il3738Met) produces an AR protein with only 50% of the transcriptional potency of the wild-type protein, suggesting that a 50% reduction in androgen action can produce all of the elements of TDS. However, the implication is that a reduction in AR action for any reason may also contribute to TDS. Importantly, each of these symptoms is increasing in incidence and prevalence, suggesting an environmental component.[132] In addition, a number of EDCs to which the human population is chronically exposed have antiandrogenic actions, including phthalates, BPA, and some pesticides.[133,134] In addition, evidence generated from controlled animal studies indicates that males exposed to estrogens and antiandrogens develop hypospadias, undescended testis, and low sperm counts.[130] Given the number of estrogenic and antiandrogenic chemicals to which the human population is chronically exposed, it is reasonable to posit that there is a significant contribution of EDCs to the increased incidence of TDS symptoms. These are described as follows.

Sperm Counts and Fertility. Infertility, as defined by the inability to conceive after frequent and unprotected sexual intercourse for 12 months, currently affects 15% of couples worldwide.[135] There has been a significant effort to associate infertility to sperm counts, and to further associate sperm counts to environmental stressors including EDCs. Since the mid-1990s, there has been a growing recognition that sperm count and sperm quality are declining in humans across the globe. The first report of this decline was published in 1992,[136] and since then the original data have been reanalyzed and confirmed. Although there was early and reasonable skepticism about these observations and conclusions, new prospective data have been added on a regional basis, which also demonstrate declines in sperm counts that vary by geographic location. At this time, there is general scientific consensus that sperm counts have declined throughout the past 50 years in the human population, and that there are likely to be both genetic and environmental contributors to this decline. A causal link between exposures to EDCs, especially those that act via the AR or those that decrease testosterone production, and sperm counts and fertility is plausible, and there is evidence that several pesticides as well as polybrominated diphenyl ethers are involved. It is possible that adult sperm production is sensitive to current environmental exposures because it is an ongoing process. However, it also appears that adult sperm counts are sensitive to developmental programming, especially to EDCs with antiandrogenic properties like phthalates. Phthalates can reduce serum androgens in rodents and have been linked to reduced anogenital distance in humans, a marker of fetal testosterone levels.[137] In animal models, there are strong and consistent data showing effects of EDCs, including BPA, vinclozolin, dioxin, and phthalates, on several aspects of male reproduction, including sperm counts. Indeed, all these chemicals also cause reduced sperm counts and fertility across four generations, indicating transgenerational effects in animal models.[138] These new data are significant, as they suggest that exposures during pregnancy may lead to toxic effects in grandchildren and great-grandchildren. Because of the lifelong sensitivity of sperm production to EDCs, along with the fact that there are hundreds of EDCs with varying exposures throughout the lifespan, it will be difficult to ferret out the exact contribution of individual EDC exposures to sperm counts and fertility. Nonetheless, data support a role for EDCs impacting both sperm counts and fertility and possibly the secular decline in these measures globally.

Hypospadias and Cryptorchidism. Normal differentiation and growth of the male external genitalia require normal androgen production and an adequate tissue response to the hormone. Hypospadias can be defined as a consequence of the abnormal development of urethral spongiosum tissue, with a defect of the ventral aspect of the prepuce and the skin of the distal shaft. There are many anatomic variants exhibited; minor forms are characterized by an ectopic urethral opening on the glans penis or more proximally, whereas the more severe forms present a proximal urethral opening even at the perineum. The different forms of hypospadias may represent phenotypes that are linked to different risk factors,[139] and this heterogeneity may be important in identifying specific causes.

The incidence of male urogenital abnormalities has increased during the past few decades.[13] Early estimates of hypospadias incidence may have been underreported in some countries, depending on the way the data were gathered. However, in the case of the more severe proximal phenotype, where surgical correction is required, the reporting metrics have not changed, and incidence has clearly increased in the United States, Australia, and many European countries.[132] There is increasing circumstantial evidence that at least some forms of hypospadias and other TDS symptoms are caused by environmental factors. One thread of evidence for this is that idiopathic partial androgen insensitivity syndrome can produce symptoms of TDS including hypospadias,[140] even in cases where the androgen system appears to be working normally. Despite this, there is no clear relationship between hypospadias and individual EDCs. Given the heterogeneous nature of hypospadias and the large mixture of chemicals known to produce this defect in animals, including those chemicals to which humans are exposed, it should be expected that relationships of interest will be difficult to identify.

Like hypospadias, cryptorchidism presents with different degrees of severity, from palpable undescended testes to abdominal testes that are not palpable. Epidemiologic studies with well-defined criteria for identifying the incidence of cryptorchidism have shown very large differences in the incidence between countries and increasing trends in Denmark and the United Kingdom.[132] Genetic associations with the presence of cryptorchidism do not account for a large proportion of cases.[141] In addition, genetic mouse models have shown important mechanisms of male genital development and provide insight into the mechanisms by which cryptorchidism and hypospadias may be triggered. Normal androgen action and tissue response are essential for penile development and testicular descent. However, estrogens can interfere with normal male genital development by downregulating insulin-like factor 3 and stimulating ATF3, leading to cryptorchidism and hypospadias, respectively. Therefore, endocrine disruptors that alter the estrogen-androgen ratio or act as hormone mimics (agonists or antagonists) can cause maldevelopment in animals, which is highly plausible in humans. Interestingly, BPA exposure is associated with poor outcomes in in vitro fertilization.[142]

Female Reproduction. Conception rates in both Danish and US women have declined 44% since 1960.[143] There are limited prevalence data on women's environmental health diseases including menstrual cyclicity, pubertal development, endometriosis, uterine fibroids, or polycystic ovarian syndrome (PCOS); however, it is clear that these diseases are common. Indeed, PCOS is the most common endocrine disorder in women of reproductive age, although, because the definition of PCOS is still under debate, diagnosis is variable and actual prevalence is unclear, which results in an estimated occurrence between 3% and 15% depending on criteria used (see Chapter 126).[144]

Uterine fibroids (leiomyomas) are common with estimates that they occur in 25% to 50% of all women; however, the effects of EDCs have not been examined in detail for possible impact on these diseases.[145] Interestingly, the incidence of endometriosis is thought to be in the range of 10% to 15% for reproductive-age women (see Chapter 127).[146]

As noted earlier, studies of DES in humans and animals establish the principle that estrogenic EDC exposure during development can produce adverse effects on women's reproductive health. In addition, data from animals models consistently show that developmental or adult exposure to some EDCs including organochlorine pesticides, heavy metals, phthalates, BPA, and dioxin can alter ovarian development including follicle formation, follicle growth, and disruption of steroid hormone levels.[147] Studies in humans show that adult exposure to some of these EDCs is associated with increased risk of miscarriage, time to pregnancy and reduced size for gestational age.[147] For example, BPA has been associated with infertility, shortened gestation, and increased preterm birth in women. These observations are concordant with those in animals showing that BPA affects the morphology and function of the oviduct, ovary, and hypothalamic-pituitary-ovarian axis.[148]

Puberty. Puberty is a complex physiological process involving central activation of the hypothalamic-pituitary-gonad axis. Two hundred years ago, the average age of puberty was around 17 years old, and this declined and stabilized in the 1950s at about 13 years of age.[149,150] However, the age of puberty has declined again during the past few decades; in girls, the average age of pubertal onset is 10 years, and in boys it is around 11.5 years.[151] Improved nutrition is reasonably proposed to contribute to these changes, especially as the average BMI has increased and the relationship between BMI, leptin, and puberty onset is well known (see Chapter 104). However, the incidence of precocious puberty in girls less than 8 years old has increased in general and is geographically variable, indicating that environmental factors can affect both the average age of puberty onset and the incidence of precocious puberty.[152] There are multiple reports that focus on a role of EDCs as an explanation for early puberty.[153]

Because of the role of sex steroids in the onset of puberty, it is reasonable to propose that exposures to a variety of EDCs during development or early life may contribute to the earlier onset of puberty.[154] There is convincing evidence from animal studies that prenatal and/or neonatal treatment with EDCs that act via the estrogen receptor can accelerate pubertal development in females, whereas estrogenic chemicals tend to delay puberty onset in males. EDCs acting by other mechanisms, which are shown to have effects in some animal studies, include lead, styrene, methoxychlor, phthalates, and BPA. Indeed, the data linking estrogenic EDCs to early puberty in animal models are strong. There are some epidemiologic studies linking exposures to EDCs (DDT/DDE, dioxins, PCBs, PBBs, some flame retardants, hexachlorobenzene, BPA, lead, and cadmium);[155] however, overall epidemiologic evidence for a role of EDCs in early puberty is still limited.[156] The importance of changes in puberty is partly a result of the effects of puberty on other health outcomes where early puberty is associated with reproductive tract cancers, depression and eating disorders, diabetes, and cardiovascular disease.[153] Note also that obesogens can increase weight leading to an increase in leptin and puberty advance,[157] and that early puberty is a risk factor for breast cancer.[158]

Endometriosis. Endometriosis is an estrogen-dependent disorder, but the exact etiology is unclear. The strongest data linking EDCs to endometriosis are those of dioxin causing endometriosis in primates,[150] which is supported by experiments in rodents. Although the mechanism for the dioxin effects on endometriosis is unclear, there are data showing that dioxin decreases circulating estradiol and also causes its degradation, and that dioxin reduces the expression of the progesterone receptor[159] as well as acting as an immunosuppressant.

Because of the role of the sex steroids in reproduction, it is biologically plausible that EDCs that mimic or antagonize sex steroids will play a role in the multicausality for female reproductive disorders.[160]

KEY POINTS

- Secular trends in chronic human disease and disorders represent evidence of the impact of EDC exposures on human health in part because these outcomes are endocrine in nature.
- Obesity and diabetes are epidemic in the United States and the world and evidence indicates that a variety of chemicals contribute to these human health problems.
- Infertility and male and female reproductive disorders represent significant human health issues today that are being exacerbated by chemical exposures.

Cancer

A large amount of research during the past several decades has established the involvement of endogenous estrogens and androgens in cancers of reproductive organs and the thyroid.[161] However, the possible contribution of environmental chemicals that interact with these hormone systems has only recently received attention for two primary reasons. First, cancers of the breast, endometrium, ovary, testis, prostate, and thyroid glands continue to rise among populations of Western countries, and more recently also among Asian nations. Second, the involvement of the synthetic estrogen, DES, in vaginal cancers and breast cancer has heightened concerns that a multitude of other hormonally active chemicals in everyday use can cause or increase the risk of these diseases. Most research into associations with endocrine disruptors has been performed with breast, prostate, and testis cancers, whereas other hormone-related cancers such as endometrial, ovarian, and thyroid cancers have received less attention.

Breast cancer is the most common endocrine cancer and is the second leading cause of cancer death. Its incidence has increased since the mid-1980s in almost all environmentalized/ Westernized countries[13] (see also Chapter 128). The genetic and metabolic aspects of breast cancer are responsible for up to 27% of breast cancer, with the other 70% or more related to environmental factors.[162] It is clear that excess estrogen across the lifespan is an important risk factor for breast cancer. Thus, there is an increased risk of breast cancer among current and recent users of oral contraceptives.[163] This has manifested in the effects on increasing breast cancer in offspring of mothers exposed to DES, the effects of hormone replacement therapy on increased breast cancer incidence, and the association of early puberty or increasing age of first pregnancy with increased breast cancer.

It is also clear that the development period is a sensitive time for EDCs to affect breast development, and effects can lead to cancer or precancerous lesions later in life. However, in contrast to most tissues, the breast continues developing until adulthood and even changes slightly with each menstrual cycle.

Polycyclic aromatic hydrocarbons are a class of chemicals that have both genotoxic/mutagenic properties as well as endocrine properties. Other EDCs that alter mammary gland development in animal models by affecting the organization of the tissue include BPA, dioxin, PFOA, the fungicide vincolozolin, polybrominated diphenyl ether (PBDE), nonophenol, and atrazine.[164] Genistein, a plant estrogen found in soy products, can protect against carcinogen-induced mammary gland tumors in some animal studies, canceling out the protective effect of tamoxifen in others, and can actually cause ductal hyperplasia in others depending on the timing and dose. Clearly, assessing the role of EDCs in breast cancer will remain highly complex.

For decades environmental health scientists studied the potential link between adult exposures to a variety of EDCs to breast cancer, but the results were inconsistent. After the animal data showed the importance of developmental exposures in increasing susceptibility to mammary cancer in animals, epidemiologists started to examine developmental exposures and breast cancer. However, because of the long latency to develop cancer, these studies are expensive and time consuming, requiring assessment of developmental exposures and follow-up for 4 to 5 decades to assess the incidence of breast cancer. A classic study examined exposure to DDT and its metabolites during childhood and throughout life and showed that there was no association of exposures in adults with breast cancer but a high correlation of exposure around puberty with adult-onset breast cancer.[165] This study provided the data that focused epidemiological studies on the *in utero* and childhood exposures as the sensitive window of exposure leading to increased susceptibility to breast cancer later in life.

Overall, the growing literature linking developmental exposure to EDCs and later development of breast cancer is particularly strong.[164] Singly and in combination, these EDCs—many of which are present in everyday products—may be contributing significantly to the public health burden of breast cancer. Thus, it is important to prioritize strategies to reduce risk in addition to improving breast cancer therapies.

Thyroid Disruption

TH plays an important role in development and adult physiology.[166] Therefore, it is important to recognize that there are a large number of chemicals that can interfere with thyroid function[167,168] and/or with TH action.[169] These different kinds of chemicals may interact to affect TH action during development. For example, perchlorate can reduce iodide uptake and in the face of low dietary iodine can affect thyroid function and TH levels.[170] In addition, a large number of chemicals can increase the clearance rate of TH, similar to that of phenobarbital.[171] Some chemicals also appear to interact directly with receptors for TH[172,173] and may produce effects that are not fully consistent with hypo- or hyperthyroidism.[174] Finally, there are chemicals that can displace THs from serum transport proteins,[175] and others that may interfere with cellular transport proteins such as MCT8 or OATP1c1, although this is less well-studied. However, because epidemiologic studies investigating the relationship between chemical exposure and the thyroid uniformly evaluate serum hormone levels, it will be a challenge to link exposure to the combination of chemicals that can impact thyroid function and TH action to human disease. The neurobehavioral impacts of TH insufficiency are so well known that it is reasonable to consider the possibility that environmental chemicals may play a role in the increased incidence and prevalence of neurobehavioral disorders. This may best be apparent for PCBs.[29] Likewise, environmental chemical interference with TH action may contribute to adult diseases including cardiovascular, obesity, and metabolic syndrome. These relationships are much less well understood. Given the variability in clinical manifestations of thyroid disease in adults, this will be a challenging topic.

Neurobehavior

Neurodevelopmental disorders have increased in prevalence during the past several decades, and the environment is likely to play a role.[176–178] Since the 1970s, there have been dramatic increases in previously rare neurodevelopmental disorders. For example, in the 1970s the prevalence of autism spectrum disorders (ASD) was estimated to be between 4 and 5 in 10,000 children,[179] but today this value is estimated to be 1 in 54 children.[180] There is significant debate about the role of broadening definitions of ASD related to this increase[181] or the role of other factors that may mask a real increase.[182] Similar trends have been observed for other neurobehavioral problems such as ADHD (attention-deficit hyperactivity disorder), learning disabilities, and childhood and adult depressive disorders. Predominant among these disorders are attention-deficit disorders (ADDs)—with or without hyperactivity—with a worldwide, pooled prevalence estimate of about 5.3%.[106] These are difficult trends to attribute unequivocally to the environment, although it appears clear that the increases are not fully attributable to artifacts related to more aggressive diagnosis and reporting and may involve genetic susceptibility.

KEY POINTS

- Secular trends in the incidence of some cancers (e.g., pediatric brain cancer), neurobehavioral disorders and thyroid dysfunction represent additional impacts of EDCs on the human population.
- In each case, scientific evidence from mechanistic studies support conclusions that associations between chemical exposures and these disease trends are causal. Moreover, these population studies represent evidence that can be employed in the clinic.

EMERGING ISSUES

Endocrine Disruptors, the Epigenome, and Transgenerational Effects

Parents pass on their genes to their children, and in so doing pass on various traits associated with those genes. The combination of all genes in a species is referred to as the "genome," and "genomic" studies refer to those studies designed to understand how various patterns of genes are controlled. But, cells in the body can pass on heritable traits to their cellular progeny without altering their genome. Certainly, EDCs can also act across the lifespan just as hormones do.

Epigenetics is broadly defined as heritable changes not dependent on genetic sequences, and it is these epigenetic processes that define and control tissue development by controlling gene expression. Endocrine disruptors have also been shown to produce transgenerational effects as a result of their ability to alter epigenetic processes. This issue first arose with studies in which an antiandrogenic pesticide (vinclozolin) was administered to developing mice at a single time during development—when the testis was in a critical period of development. Not only did vinclozolin produce adverse effects on the developing testis, but also this was transmitted through four generations of mice.[128] This effect is likely to be caused by epigenetic changes that were transmitted with high fidelity from one generation to the next via the germ cells (sperm). A number of endocrine disruptors have now been shown to influence epigenetic mechanisms and to produce effects in several generations of animals.[129] Grandmother exposures may be affecting the F2 and subsequent generations. Notably, in a mouse model, deleterious effects of EDCs are corrected in the mammalian germline by epigenomic reprogramming.[183] These deeply concerning and contradictory observations in animal studies with profound implications for the human population warrant further study.

Increasing Exposures Worldwide

There are no longer pristine areas worldwide without environmental pollutants.184 EDCs are everywhere; they are present in food, nature, and humans, and there is global transport through ocean, air currents, plants, and animals. Indeed, the Faroe Islands, off the coast of Iceland and Norway, which because of its isolated location should be one of the most pristine places in the world, has a human population that is severely affected by EDCs via the eating of contaminated whale blubber and meat.

Expanded List of Diseases Related to EDCs

Until recently, the focus of EDC research was on chemicals that affected the estrogen, androgen, and thyroid pathways, and diseases that result from their disruption. Endocrine disruption is no longer limited to these pathways, as chemicals have also been shown to interfere with metabolism, fat storage, bone development, glucose homeostasis, glucocorticoid regulatory pathways, and immune systems.[184] It is possible that all the endocrine systems will be impacted by environmental chemicals. Indeed, even the known EDCs may not be representative of the full range of relevant structures and properties of chemicals that may have endocrine-disrupting properties. Because only a limited number of chemicals in commerce have been tested for endocrine-disrupting activity, it is likely that there are many more chemicals with endocrine-disrupting activity than currently believed. In addition, metabolites or environmental transformation products and products formed during waste treatment are usually not examined for endocrine-disrupting activity. Thus, research is needed to identify other EDCs.

Mixtures

The focus of research identifying and characterizing the effects of EDC in animal models and in human populations has been on linking exposure to a single chemical at a single exposure leading to a single disease. However, humans are exposed to mixtures of chemicals throughout their lifetimes, and indeed each chemical has the potential to affect several endocrine diseases. DES is a good example of one exposure leading to many sequelae. A developmental "estrogenization syndrome" has been proposed.[96] In addition, it is now clear that exposures to many chemicals at low levels that individually have no effect can actually produce significant effects because of the additivity of their individual effects.[134] Although mixture studies are difficult to develop in animals and are expensive and complicated to interpret statistically in human studies, there is an urgent need to expand research into the effects of mixtures of EDCs both during development and across the lifespan.

E-Waste

Modern society has developed many electronic devices that are scrapped when broken or obsolete, including computers, televisions, phones, audio components, and refrigerators. Because these electronic components contain many endocrine-disrupting toxicants including lead, cadmium, brominated flame retardants, plastics, and plasticizers, they can involve significant risk to workers and communities that recycle or incinerate or otherwise dispose of the components. Although the majority of e-waste ends up in landfills, recycling involves chemical stripping, open burning, shredding, and melting of the e-waste under conditions where the workers are highly exposed to the toxic chemicals via air emissions, water, and soil contamination as well as direct physical exposures.[185]

Green Chemistry

Green chemistry, also called sustainable chemistry, is a new field that encourages the design and development of chemicals using principles that minimize the use and generation of toxic chemicals. Paul Anastas coined the term "green chemistry," and it consists of 12 principles that he developed.[97] Indeed green chemistry focuses on prevention of toxicity by producing the next generation of chemicals without toxicity. A joint project by green chemists and environmental scientists focused on the development of test methods (TIPED)[98] to detect endocrine activity in new chemicals being developed before their release into the environment. Green chemistry offers an alternative to chemical development that has the potential to reduce the number of new toxic chemicals released into the environment and also to develop nontoxic alternatives to chemicals currently in use.

KEY POINTS

- There are two categories of emerging issues that are important to monitor. The first category is that the current chemical mixture in the environment is being recycled in ways that contaminate workers and the general population. Thus, exposure routes are being identified that were not recognized previously. In addition, chemicals that have been banned (e.g., PCBs) or heavily regulated (e.g., lead) continue to pose health risks both because of their widespread use before being banned and because of their long-term effects on human health.
- A second category of emerging issue is that there is a nascent movement to produce safer chemicals. The traditional strategy is to produce and employ chemicals that have not been proven to cause harm and often are as toxic as the chemical it replaces. Proof of harm requires decades after deployment, harming generations of people. The revised strategy is to avoid the production and use of chemicals that exhibit any mechanistic signs of biological effects. This strategy will gain traction as innovation occurs.

CONCLUSION AND GUIDANCE FOR CLINICIANS

The human population is heavily contaminated with environmental chemicals. Most of those chemicals have not been studied for their impact on human health, but many have been shown to contribute to chronic disease prevalence. This observation has two important implications. The first is that every patient a clinician sees will have a certain burden of toxic chemicals. The second is that the current clinical and scientific evidence can and should be included in clinical recommendations. It is important to recognize that EDC effects on human health are observed at the population level. Thus, these data can inform clinicians about relative risk, just as other health-related domains such as smoking, exercise and nutrition inform clinicians about recommendations for health care.

There are several lines of evidence that support concern that chemicals in the environment are producing adverse health effects in the human population by interfering with hormone action. These are in part articulated and referenced more fully in the UNEP/WHO State of the Science of EDCs.[13] This evidence includes observations that there are a high incidence and increasing secular trends of many endocrine-related disorders in humans; that endocrine-related effects in wildlife populations are prevalent; and that laboratory studies have identified chemicals with endocrine-disrupting properties that can account for some of these observations in humans and wildlife. Considering the evidence, the Endocrine Society has released a Scientific Statement[84] that articulates the concern for the public health impact of chemical exposure. In addition, the American College of Obstetricians and Gynecologists in collaboration with the American Society of Reproductive Medicine has released a statement calling for a reduction in exposure to these chemicals and advising clinicians to be aware of the sources of contamination in women of reproductive age.[85] The Royal Society of Obstetricians and Gynecologists likewise released a statement providing advice on handling the potential consequences of chemical exposures (http://www.rcog.org.uk/files/rcog-corp/5.6.13ChemicalExposures.pdf), as has the International Federation of Gynecologists and Obstetricians.[186] The overriding message from all these society statements is that clinicians must play a larger role in providing information to their patients regarding how to avoid exposure to toxic agents. Although it is not expected that health care professionals be experts in toxic chemicals, they can still provide important information via literature provided to patients, specific focus on environmental health sciences in birth preparation classes and referring patients to specialists when a hazardous exposure is suspected. The noted reports provide information regarding key chemical exposures, sources of exposures at home, work, and in the community, and clinical implications of exposure. With the help of health care professionals, patients can become more aware of the risks to their health and the health of their children, including those unborn, and they can thus take action to reduce their body burden of toxic chemicals, thereby potentially improving overall health throughout the lifespan. An example of an exposure history is available at http://prhe.ucsf.edu/prhe/clinical_resources.html. Websites that contain useful information include: www.healthandenvironment.org, www.ewg.org, www.cehn.org, and www.prhe.ucsf.edu.

Simple steps that patients can take include the following: avoid smoking and secondhand smoke, avoid fast foods and processed foods, limit foods that are high in animal fat, limit fish that are known to contain PCBs and mercury (shark, swordfish, king mackerel, and tilefish), do not microwave items in plastic containers, be mindful of ingredients in cosmetics and skin care products, do not use pesticides around the home, keep the home free of dust that contains lead, pesticides, and flame retardants, use nontoxic cleaning products (www.prhe.ucsf.edu/prhe/tmlinks.html#cleaningproducts), choose plastics carefully to avoid BPA and products made with soft PVC, use VOC-free and water-based paints, use a digital thermometer, wash all fruits and vegetables, and eat organic and fresh products when possible.

REFERENCES

1. Bongaarts J. U.S. health in international perspective: shorter lives, poorer health. *Popul Dev Rev*. 2013;39:165–167.
2. Trasande L. *Sicker, Fatter, Poorer: The Urgent Threat of Hormone-Disrupting Chemicals on Our Health and Future... And What We Can Do About It*. Boston: Houghton Mifflin Harcourt; 2019:1 (online resource).
3. Grandjean P. *Only One Chance: How Environmental Pollution Impairs Brain Development-Aand How to Protect the Brains of the Next Generation*. Oxford: Oxford University Press; 2013.
4. Demeneix B. *Losing Our Minds: How Environmental Pollution Impairs Human Intelligence and Mental Health*. Oxford: Oxford University Press; 2014.
5. Demeneix B. *Toxic Cocktail: How Chemical Pollution Is Poisoning Our Brains*. New York: Oxford University Press; 2017.
6. Furtado KS, Budd EL, Ying X, et al. Exploring political influences on evidence-based non-communicable disease prevention across four countries. *Health Educ Res*. 2018;33:89–103.
7. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Publ Health*. 2009;9:88.
8. Ward ZJ, Bleich SN, Cradock AL, et al. Projected U.S. state-level prevalence of adult obesity and severe obesity. *N Engl J Med*. 2019;381:2440–2450.
9. vom Saal FS, Cohen A. *How Toxic Chemicals Contribute to COVID-19 Deaths*. Environmental Health News; 2020.
10. Jeck WR, Siebold AP, Sharpless NE. Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell*. 2012;11:727–731.
11. Woodruff TJ, Burke TA, Zeise L. The need for better public health decisions on chemicals released into our environment. *Health Aff*. 2011;30:957–967.
12. Bookman EB, McAllister K, Gillanders E, et al. Gene-environment interplay in common complex diseases: forging an integrative model-recommendations from an NIH workshop. *Genet Epidemiol*. 2011;35:217–225.
13. Bergman A, Heindel JJ, Jobling S, Kidd KA, Zoeller RT, eds. *State of the Science of Endocrine Disrupting Chemicals 2012*. Geneva, Switzerland: World Health Organization; 2013.
14. Crinnion WJ. The CDC fourth national report on human exposure to environmental chemicals: what it tells us about our toxic burden and how it assist environmental medicine physicians. *Altern Med Rev*. 2010;15:101–109.
15. ATSDR. ATSDR case studies in environmental medicine: taking an exposure history. In: *Services UDoHaH*. ATSDR; 2019.
16. Dionisio KL, Frame AM, Goldsmith MR, et al. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. *Toxicol Rep*. 2015;2:228–237.
17. Berg EL, Yang J, Melrose J, et al. Chemical target and pathway toxicity mechanisms defined in primary human cell systems. *J Pharmacol Toxicol Methods*. 2010;61:3–15.
18. Koedrith P, Kim H, Weon JI, Seo YR. Toxicogenomic approaches for understanding molecular mechanisms of heavy metal mutagenicity and carcinogenicity. *Int J Hyg Environ Health*. 2013;216:587–598.
19. Benigni R, Bossa C, Tcheremenskaia O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. *Chem Rev*. 2013;113:2940–2957.
20. Zoeller RT, Brown TR, Doan LL, et al. Endocrine-disrupting chemicals and public health protection: a statement of principles from the Endocrine Society. *Endocrinology*. 2012;153:4097–4110.
21. Kortenkamp A, Martin O, Faust M, et al. *State of the Art Assessment of Endocrine Disruptors Final Report*. Brussels: European Commission; 2011:442.
22. Veiga-Lopez A, Pu Y, Gingrich J, Padmanabhan V. Obesogenic endocrine disrupting chemicals: identifying knowledge gaps. *Trends Endocrinol Metab*. 2018;29:607–625.
23. Gore AC, Chappell VA, Fenton SE, et al. EDC-2: the Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36:er20151010.
24. Llorente-Esteban A, Manville RW, Reyna-Neyra A, Abbott GW, Amzel LM, Carrasco N. Allosteric regulation of mammalian Na(+)/I(−) symporter activity by perchlorate. *Nat Struct Mol Biol*. 2020;27:533–539.
25. Dohan O, Portulano C, Basquin C, Reyna-Neyra A, Amzel LM, Carrasco N. The Na+/I symporter (NIS) mediates electroneutral active transport of the environmental pollutant perchlorate. *Proc Natl Acad Sci U S A*. 2007;104:20250–20255.
26. Narizzano R, Risso F, Magherini A, et al. Extensive study on physicochemical properties of polychlorinated biphenyls in a commercial ion trap mass spectrometer, relevance in analytical and environmental chemistry. *J Mass Spectrom*. 2017;52:837–847.
27. Miyazaki W, Iwasaki T, Takeshita A, Kuroda Y, Koibuchi N. Polychlorinated biphenyls suppress thyroid hormone receptor-mediated transcription through a novel mechanism. *J Biol Chem*. 2004;279:18195–18202.
28. Taylor PN, Okosieme OE, Murphy R, et al. Maternal perchlorate levels in women with borderline thyroid function during pregnancy and the cognitive development of their offspring; Data from the Controlled Antenatal Thyroid Study. *J Clin Endocrinol Metab*. 2014;99:jc20141901.
29. Schantz SL, Widholm JJ, Rice DC. Effects of PCB exposure on neuropsychological function in children. *Environ Health Perspect*. 2003;111:357–576.

30. Sun H, Xu LC, Chen JF, Song L, Wang XR. Effect of bisphenol A, tetrachlorobisphenol A and pentachlorophenol on the transcriptional activities of androgen receptor-mediated reporter gene. *Food Chem Toxicol*. 2006;44:1916–1921.
31. Kitamura S, Jinno N, Ohta S, Kuroki H, Fujimoto N. Thyroid hormonal activity of the flame retardants tetrabromobisphenol A and tetrachlorobisphenol A. *Biochem Biophys Res Commun*. 2002;293:554–559.
32. Nagel SC, vom Saal FS, Thayer KA, Dhar MG, Boechler M, Welshons WV. Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. *Environ Health Perspect*. 1997;105:70–76.
33. Delfosse V, Grimaldi M, Pons JL, et al. Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. *Proc Natl Acad Sci U S A*. 2012;109:14930–14935.
34. La Merrill MA, Vandenberg LN, Smith MT, et al. Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. *Nat Rev Endocrinol*. 2019;16:45–47.
35. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. *Environ Health Perspect*. 2011;119:878–885.
36. Calafat AM, Weuve J, Ye X, et al. Exposure to bisphenol A and other phenols in neonatal intensive care unit premature infants. *Environ Health Perspect*. 2009;117:639–644.
37. Chaou CH, Lin CC, Chen HY, Lee CH, Chen TH. Chlorpyrifos is associated with slower serum cholinesterase recovery in acute organophosphate-poisoned patients. *Clin Toxicol*. 2013;51:402–408.
38. Cha YS, Kim H, Cho NH, et al. Pyrethroid poisoning: features and predictors of atypical presentations. *Emerg Med J*. 2013;31:899.
39. Bisogni V, Rossi GP, Calo LA. Apparent mineralcorticoid excess syndrome, an often forgotten or unrecognized cause of hypokalemia and hypertension: case report and appraisal of the pathophysiology. *Blood Press*. 2013;23:189.
40. Delange F. Cassava and the thyroid. In: Gaitan E, ed. *Environmental Goitrogenesis*. Boca Raton: CRC Press, Inc.; 1989:173–194.
41. Needleman H. Low level lead exposure: history and discovery. *Ann Epidemiol*. 2009;19:235–238.
42. Howell WH, Chesley AJ, Edsall DL, et al. Report of surgeon general's committee on tetraethyl lead. *Ind Eng Chem*. 1926;18:193–196.
43. Needleman HL, Tuncay OC, Shapiro IM. Lead levels in deciduous teeth of urban and suburban American children. *Nature*. 1972;235:111–112.
44. Markowitz GE, Rosner D. *Deceit and Denial: The Deadly Politics of Industrial Pollution*. Berkeley, CA: University of California Press; 2002.
45. Attina TM, Trasande L. Economic costs of childhood lead exposure in low- and middle-income countries. *Environ Health Perspect*. 2013;121:1097.
46. Meyer PA, Brown MJ, Falk H. Global approach to reducing lead exposure and poisoning. *Mutat Res*. 2008;659:166–175.
47. Gomez HF, Borgialli DA, Sharman M, et al. Analysis of blood lead levels of young children in Flint, Michigan before and during the 18-month switch to Flint River water. *Clin Toxicol*. 2019;57:790–797.
48. Santucci Jr RJ, Scully JR. The pervasive threat of lead (Pb) in drinking water: unmasking and pursuing scientific factors that govern lead release. *Proc Natl Acad Sci U S A*. 2020;117:23211–23218.
49. Jacobson PD, Boufides CH, Chrysler D, Bernstein J, Citrin T. The role of the legal system in the Flint water crisis. *Milbank Q*. 2020;98:554–580.
50. ATSDR. Toxicological profile for polybrominated Biphnyls and polybrominated diphenyl ethers. In: *Services UDoHaH*. Atlanta, GA: Agency for Toxic Substances and Disease Registry; 2004.
51. ATSDR. Toxicological profile for polychlorinated biphenyls (PCBs). In: *Services UDoHaH, Ed.* Atlanta, GA. 2000.
52. Kashimoto T, Miyata H, Kunita S, et al. Role of polychlorinated dibenzofuran in Yusho (PCB poisoning). *Arch Environ Health*. 1981;36:321–325.
53. Kashimoto T, Miyata H, Kunita S, et al. Role of polychlorinated dibenzofuran in yusho (PCB poisoning). *Arch Environ Health*. 1981;36:321–326.
54. Kafafi SA, Afeefy HY, Ali AH, Said HK, Kafafi AG. Binding of polychlorinated biphenyls to the aryl hydrocarbon receptor. *Environ Health Perspect*. 1993;101:422–428.
55. Kodavanti PR, Kannan N, Yamashita N, et al. Differential effects of two lots of aroclor 1254: congener-specific analysis and neurochemical end points. *Environ Health Perspect*. 2001;109:1153–1161.
56. Seegal RF, Bush B, Shain W. Lightly chlorinated ortho-substituted PCB congeners decrease dopamine in nonhuman primate brain and in tissue culture. *Toxicol Appl Pharmacol*. 1990;106:136–144.
57. FDA. In: FDA US, ed. *Code of Federal Regulations, Title 21 Part 109.3*. Washington, DC: US Govt; 2010.
58. Gao Q, Ben Y, Dong Z, Hu J. Age-dependent human elimination half-lives of dioxin-like polychlorinated biphenyls derived from biomonitoring data in the general population. *Chemosphere*. 2019;222:541–548.
59. Giera S, Bansal R, Ortiz-Toro TM, Taub DG, Zoeller RT. Individual polychlorinated biphenyl (PCB) congeners produce tissue- and gene-specific effects on thyroid hormone signaling during development. *Endocrinology*. 2011;152:2909–2919.
60. Gauger KJ, Giera S, Sharlin DS, Bansal R, Iannacone E, Zoeller RT. Polychlorinated biphenyls 105 and 118 form thyroid hormone receptor agonists after cytochrome P4501A1 activation in rat pituitary GH3 cells. *Environ Health Perspect*. 2007;115:1623–1630.
61. Grimm FA, Lehmler HJ, He X, Robertson LW, Duffel MW. Sulfated metabolites of polychlorinated biphenyls are high-affinity ligands for the thyroid hormone transport protein transthyretin. *Environ Health Perspect*. 2013;121:657–662.
62. Hamers T, Kamstra JH, Sonneveld E, et al. Biotransformation of brominated flame retardants into potentially endocrine-disrupting metabolites, with special attention to 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47). *Mol Nutr Food Res*. 2008;52:284–298.
63. Hood A, Klaassen CD. Differential effects of microsomal enzyme inducers on in vitro thyroxine (T(4)) and triiodothyronine (T(3)) glucuronidation. *Toxicol Sci*. 2000;55:78–84.
64. Sethi S, Morgan RK, Feng W, et al. Comparative analyses of the 12 most abundant PCB congeners detected in human maternal serum for activity at the thyroid hormone receptor and ryanodine receptor. *Environ Sci Technol*. 2019;53:3948–3958.
65. Gilbert ME, Rovet J, Chen Z, Koibuchi N. Developmental thyroid hormone disruption: prevalence, environmental contaminants and neurodevelopmental consequences. *Neurotoxicology*. 2012;33:842–852.
66. Newbold RR. Lessons learned from perinatal exposure to diethylstilbestrol. *Toxicol Appl Pharmacol*. 2004;199:142–150.
67. Troisi R, Palmer JR, Hatch EE, et al. Gender identity and sexual orientation identity in women and men prenatally exposed to diethylstilbestrol. *Arch Sex Behav*. 2020;49:447–454.
68. Titus-Ernstoff L, Troisi R, Hatch EE, et al. Birth defects in the sons and daughters of women who were exposed in utero to diethylstilbestrol (DES). *Int J Androl*. 2010;33:377–384.
69. Klip H, Verloop J, van Gool JD, et al. Hypospadias in sons of women exposed to diethylstilbestrol in utero: a cohort study. *Lancet*. 2002;359:1102–1107.
70. Oerbeck B, Sundet K, Kase BF, Heyerdahl S. Congenital hypothyroidism: influence of disease severity and L-thyroxine treatment on intellectual, motor, and school-associated outcomes in young adults. *Pediatrics*. 2003;112:923–930.
71. Heyerdahl S, Oerbeck B. Congenital hypothyroidism: developmental outcome in relation to levothyroxine treatment variables. *Thyroid*. 2003;13:1029–1038.
72. Heindel JJ, Skalla LA, Joubert BR, Dilworth CH, Gray KA. Review of developmental origins of health and disease publications in environmental epidemiology. *Reprod Toxicol*. 2017;68:34–48.
73. Gee D. Late lessons from early warnings: toward realism and precaution with endocrine-disrupting substances. *Environ Health Perspect*. 2006;114(suppl 1):152–160.
74. Gee D, Krayer von Krauss MP. Late lessons from early warnings: towards precaution and realism in research and policy. *Water Sci Technol*. 2005;52:25–34.
75. Schwartz HL, Ross ME, Oppenheimer JH. Lack of effect of thyroid hormone on late fetal rat brain development. *Endocrinology*. 1997;138:3119–3124.
76. Rovet JF. The role of thyroid hormones for brain development and cognitive function. *Endocr Dev*. 2014;26:26–43.

77. Zoeller RT, Rovet J. Timing of thyroid hormone action in the developing brain: clinical observations and experimental findings. *J Neuroendocrinol*. 2004;16:809–818.
78. Hiort O. The differential role of androgens in early human sex development. *BMC Med*. 2013;11:152.
79. Fernandez-Balsells MM, Murad MH, Lane M, et al. Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2010;95:2560–2575.
80. Zuloaga DG, Heck AL, De Guzman RM, Handa RJ. Roles for androgens in mediating the sex differences of neuroendocrine and behavioral stress responses. *Biol Sex Differ*. 2020;11:44.
81. Johnson VP, II VWS, Sato Y, Andreasen NC. Fetal alcohol syndrome: craniofacial and central nervous system manifestations. *Am J Med Genet*. 1996;61:329–339.
82. Chamorro-Garcia R, Diaz-Castillo C, Shoucri BM, et al. Ancestral perinatal obesogen exposure results in a transgenerational thrifty phenotype in mice. *Nat Commun*. 2017;8:2012.
83. Skinner MK. Endocrine disruptors in 2015: epigenetic transgenerational inheritance. *Nat Rev Endocrinol*. 2016;12:68–70.
84. Heindel JJ, Vandenberg LN. Developmental origins of health and disease: a paradigm for understanding disease cause and prevention. *Curr Opin Pediatr*. 2015;27:248–253.
85. Heindel JJ, Vom Saal FS, Blumberg B, et al. Parma consensus statement on metabolic disruptors. *Environ Health*. 2015;14:54.
86. Vom Saal FS, Vandenberg LN. Update on the health effects of bisphenol A: overwhelming evidence of harm. *Endocrinology*. 2020;162:bqaa171.
87. Giera S, Zoeller RT. Effects and predicted consequences of persistent and bioactive organic pollutants on thyroid function. In: *Effects of Persistent and Bioactive Organic Pollutants on Human Health*. New York: Wiley; 2013:203–236.
88. Li Y, Coons LA, Houtman R, et al. A mutant form of ERalpha associated with estrogen insensitivity affects the coupling between ligand binding and coactivator recruitment. *Sci Signal*. 2020;13:160.
89. Soriano S, Alonso-Magdalena P, Garcia-Arevalo M, et al. Rapid insulinotropic action of low doses of bisphenol-A on mouse and human islets of Langerhans: role of estrogen receptor beta. *PLoS One*. 2012;7:e31109.
90. Tabb MM, Blumberg B. New modes of action for endocrine-disrupting chemicals. *Mol Endocrinol*. 2006;20:475–482.
91. Thrasher TN. Arterial baroreceptor input contributes to long-term control of blood pressure. *Curr Hypertens Rep*. 2006;8:249–254.
92. Bradley DJ, Towle HC, Young WS. Spatial and temporal expression of alpha- and beta-thyroid hormone receptor mRNAs, including the beta-2 subtype, in the developing mammalian nervous system. *J Neurosci*. 1992;12:2288–2302.
93. Vandenberg LN, Colborn T, Hayes TB, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr Rev*. 2012;33:378–455.
94. vom Saal FS, Timms BG, Montano MM, et al. Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. *Proc Natl Acad Sci U S A*. 1997;94:2056–2061.
95. Lee DH, Steffes MW, Sjodin A, Jones RS, Needham LL, Jacobs Jr DR. Low dose organochlorine pesticides and polychlorinated biphenyls predict obesity, dyslipidemia, and insulin resistance among people free of diabetes. *PLoS One*. 2011;6:e15977.
96. Eskenazi B, Warner M, Marks AR, et al. Serum dioxin concentrations and age at menopause. *Environ Health Perspect*. 2005;113:858–862.
97. Korevaar TI, Muetzel R, Medici M, et al. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*. 2016;4:35–43.
98. CDC. *Fourth National Report on Human Exposure to Environmental Chemicals*. Atlanta: Center for Disease Control and Prevention; 2009.
99. Houlihan J, Kropp T, Wiles R, Gray S, Campbell C. *Body Burden: The Pollution in Newborns*. Washington, DC: Environmental Working Group; 2005.
100. Neltner TG, Alger HM, Leonard JE, Maffini MV. Data gaps in toxicity testing of chemicals allowed in food in the United States. *Reprod Toxicol*. 2013;42:85.
101. Woodruff TJ, Sutton P. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff*. 2011;30:931–937.
102. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2197–2223.
103. Kahn LG, Philippat C, Nakayama SF, Slama R, Trasande L. Endocrine-disrupting chemicals: implications for human health. *Lancet Diabetes Endocrinol*. 2020;8:703–718.
104. Ong KK, Ahmed ML, Dunger DB. Lessons from large population studies on timing and tempo of puberty (secular trends and relation to body size): the European trend. *Mol Cell Endocrinol*. 2006;254–255:8–12.
105. Bhasin S. Secular decline in male reproductive function: is manliness threatened? *J Clin Endocrinol Metab*. 2007;92:44–45.
106. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry*. 2007;164:942–948.
107. Maenner MJ, Shaw KA, Baio J, et al. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveill Summ*. 2020;69:1–12.
108. Hales CM, Carroll MD, Fryar CD, Ogden CL. *Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017-2018*. Washington, DC: NCHS Data Brief; 2020:1–8.
109. Hales CM, Carroll MD, Fryar CD, Ogden CL. *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016*. Washington, DC: NCHS Data Brief; 2017:1–8.
110. Henley SJ, Ward EM, Scott S, et al. Annual report to the nation on the status of cancer, part I: national cancer statistics. *Cancer*. 2020;126:2225–2249.
111. Henley SJ, Thomas CC, Lewis DR, et al. Annual report to the nation on the status of cancer, part II: progress toward Healthy People 2020 objectives for 4 common cancers. *Cancer*. 2020;126:2250–2266.
112. Gore AC, Chappell VA, Fenton SE, et al. Executive summary to EDC-2: the Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36:er20151093.
113. Sumithran P, Proietto J. The defence of body weight: a physiological basis for weight regain after weight loss. *Clin Sci*. 2013;124:231–241.
114. Serretti A, Mandelli L. Antidepressants and body weight: a comprehensive review and meta-analysis. *J Clin Psychiatry*. 2010;71:1259–1272.
115. Schug TT, Janesick A, Blumberg B, Heindel JJ. Endocrine disrupting chemicals and disease susceptibility. *J Steroid Biochem Mol Biol*. 2011;127:204–215.
116. Oken E, Gillman MW. Fetal origins of obesity. *Obesity Res*. 2003;11:496–506.
117. Grun F, Blumberg B. Endocrine disrupters as obesogens. *Mol Cell Endocrinol*. 2009;304:19–29.
118. Egusquiza RJ, Blumberg B. Environmental obesogens and their impact on susceptibility to obesity: new mechanisms and chemicals. *Endocrinology*. 2020;161:bqaa024.
119. Sumithran P, Prendergast LA, Delbridge E, et al. Long-term persistence of hormonal adaptations to weight loss. *N Engl J Med*. 2011;365:1597–1604.
120. Heindel JJ, Blumberg B, Cave M, et al. Metabolism disrupting chemicals and metabolic disorders. *Reprod Toxicol*. 2017;68:3–33.
121. Grun F, Blumberg B. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. *Endocrinology*. 2006;147:S50–S55.
122. Heindel JJ, Blumberg B. Environmental obesogens: mechanisms and controversies. *Annu Rev Pharmacol Toxicol*. 2019;59:89–106.
123. Braun JM. Early-life exposure to EDCs: role in childhood obesity and neurodevelopment. *Nat Rev Endocrinol*. 2017;13:161–173.
124. Heindel JJ, Schug TT. The perfect storm for obesity. *Obesity*. 2013;21:1079–1080.

125. Janesick A, Blumberg B. Obesogens, stem cells and the developmental programming of obesity. *Int J Androl*. 2012;35:437–448.
126. Wilson V. Type 2 diabetes: an epidemic in children. *Nurs Child Young People*. 2013;25:14–17.
127. Alonso-Magdalena P, Quesada I, Nadal A. Endocrine disruptors in the etiology of type 2 diabetes mellitus. *Nat Rev Endocrinol*. 2011;7:346–353.
128. Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*. 2005;308:1466–1469.
129. Skinner MK, Haque CG, Nilsson E, Bhandari R, McCarrey JR. Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line. *PLoS One*. 2013;8:e66318.
130. Skakkebaek NE, Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod*. 2001;16:972–978.
131. Lottrup G, Jorgensen A, Nielsen JE, et al. Identification of a novel androgen receptor mutation in a family with multiple components compatible with the testicular dysgenesis syndrome. *J Clin Endocrinol Metab*. 2013;98:2223–2229.
132. Toppari J, Virtanen HE, Main KM, Skakkebaek NE. Cryptorchidism and hypospadias as a sign of testicular dysgenesis syndrome (TDS): environmental connection. *Birth Defects Res A Clin Mol Teratol*. 2010;88:910–919.
133. Luccio-Camelo DC, Prins GS. Disruption of androgen receptor signaling in males by environmental chemicals. *J Steroid Biochem Mol Biol*. 2011;127:74–82.
134. Kortenkamp A. Which chemicals should be grouped together for mixture risk assessments of male reproductive disorders? *Mol Cell Endocrinol*. 2019;499:110581.
135. Breyer BN, Smith JF, Shindel AW, Sharlip ID, Eisenberg ML. The impact of infertility on family size in the USA: data from the National Survey of Family Growth. *Hum Reprod*. 2010;25:2360–2365.
136. Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ*. 1992;305:609–613.
137. Swan SH, Main KM, Liu F, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ Health Perspect*. 2005;113:1056–1061.
138. Doyle TJ, Bowman JL, Windell VL, McLean DJ, Kim KH. Transgenerational effects of di-(2-ethylhexyl) phthalate on testicular germ cell associations and spermatogonial stem cells in mice. *Biol Reprod*. 2013;88:112.
139. van Rooij IA, van der Zanden LF, Brouwers MM, Knoers NV, Feitz WF, Roeleveld N. Risk factors for different phenotypes of hypospadias: results from a Dutch case-control study. *BJU Int*. 2013;112:121–128.
140. Gaspari L, Paris F, Philibert P, et al. Idiopathic' partial androgen insensitivity syndrome in 28 newborn and infant males: impact of prenatal exposure to environmental endocrine disruptor chemicals? *Eur J Endocrinol*. 2011;165:579–587.
141. Dalgaard MD, Weinhold N, Edsgard D, et al. A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *J Med Genet*. 2012;49:58–65.
142. Ehrlich S, Williams PL, Missmer SA, et al. Urinary bisphenol A concentrations and early reproductive health outcomes among women undergoing IVF. *Hum Reprod*. 2012;27:3583–3592.
143. Crain DA, Janssen SJ, Edwards TM, et al. Female reproductive disorders: the roles of endocrine-disrupting compounds and developmental timing. *Fertil Steril*. 2008;90:911–940.
144. Dunaif A, Fauser BC. Renaming PCOS: a two-state solution. *J Clin Endocrinol Metab*. 2013;98:4325–4328.
145. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. *Am J Obstet Gynecol*. 2003;188:100–107.
146. Buck Louis GM, Hediger ML, Peterson CM, et al. Incidence of endometriosis by study population and diagnostic method: the ENDO study. *Fertil Steril*. 2011;96:360–365.
147. Rattan S, Zhou C, Chiang C, Mahalingam S, Brehm E, Flaws JA. Exposure to endocrine disruptors during adulthood: consequences for female fertility. *J Endocrinol*. 2017;233:R109–R129.
148. Ziv-Gal A, Flaws JA. Evidence for bisphenol A-induced female infertility: a review (2007-2016). *Fertil Steril*. 2016;106:827–856.
149. Wolff MS, Britton JA, Boguski L, et al. Environmental exposures and puberty in inner-city girls. *Environ Res*. 2008;107:393–400.
150. Rier S, Foster WG. Environmental dioxins and endometriosis. *Semin Reprod Med*. 2003;21:145–154.
151. DiVall SA, Radovick S. Endocrinology of female puberty. *Curr Opin Endocrinol Diabetes Obes*. 2009;16:1–4.
152. Su PH, Yang SF, Yu JS, Chen SJ, Chen JY. Study of leptin levels and gene polymorphisms in patients with central precocious puberty. *Pediatr Res*. 2012;71:361–367.
153. Zawatski W, Lee MM. Male pubertal development: are endocrine-disrupting compounds shifting the norms? *J Endocrinol*. 2013;218:R1–R12.
154. Fowler PA, Bellingham M, Sinclair KD, et al. Impact of endocrine-disrupting compounds (EDCs) on female reproductive health. *Mol Cell Endocrinol*. 2012;355:231–239.
155. Lucaccioni L, Trevisani V, Marrozzini L, et al. Endocrine-disrupting chemicals and their effects during female puberty: a review of current evidence. *Int J Mol Sci*. 2020;21:2078.
156. Greenspan LC, Lee MM. Endocrine disrupters and pubertal timing. *Curr Opin Endocrinol Diabetes Obes*. 2018;25:49–54.
157. Lee JE, Jung HW, Lee YJ, Lee YA. Early-life exposure to endocrine-disrupting chemicals and pubertal development in girls. *Annu Pediatr Endocrinol Metab*. 2019;24:78–91.
158. Natarajan R, Aljaber D, Au D, et al. Environmental exposures during puberty: window of breast cancer risk and epigenetic damage. *Int J Environ Res Public Health*. 2020;17:493.
159. Vonier PM, Crain DA, McLachlan JA, Guillette Jr LJ, Arnold SF. Interaction of environmental chemicals with the estrogen and progesterone receptors from the oviduct of the American alligator. *Environ Health Perspect*. 1996;104:1318–1322.
160. Smarr MM, Kannan K, Buck Louis GM. Endocrine disrupting chemicals and endometriosis. *Fertil Steril*. 2016;106:959–966.
161. Zahid M, Goldner W, Beseler CL, Rogan EG, Cavalieri EL. Unbalanced estrogen metabolism in thyroid cancer. *Int J Cancer*. 2013;133:2642–2649.
162. Macon MB, Fenton SE. Endocrine disruptors and the breast: early life effects and later life disease. *J Mammary Gland Biol Neoplasia*. 2013;18:43–61.
163. Althuis MD, Brogan DR, Coates RJ, et al. Hormonal content and potency of oral contraceptives and breast cancer risk among young women. *Br J Cancer*. 2003;88:50–57.
164. Gray JM, Rasanayagam S, Engel C, Rizzo J. State of the evidence 2017: an update on the connection between breast cancer and the environment. *Environ Health*. 2017;16:94.
165. Cohn BA, Cirillo PM, Terry MB. DDT and breast cancer: prospective study of induction time and susceptibility windows. *J Natl Cancer Inst*. 2019;111:803–810.
166. DeGroot LJ, Jameson JL. In: *Endocrinology: Adult and Pediatric*. 6th ed. Philadelphia: Saunders/Elsevier; 2010.
167. Howdeshell KL. A model of the development of the brain as a construct of the thyroid system. *Environ Health Perspect*. 2002;110(suppl 3):337–348.
168. Brucker-Davis F. Effects of environmental synthetic chemicals on thyroid function. *Thyroid*. 1998;8:827–856.
169. Zoeller RT. Environmental chemicals targeting thyroid. *Hormones (Basel)*. 2010;9:28–40.
170. Blount BC, Pirkle JL, Osterloh JD, Valentin-Blasini L, Caldwell KL. Urinary perchlorate and thyroid hormone levels in adolescent and adult men and women living in the United States. *Environ Health Perspect*. 2006;114:1865–1871.
171. Hood A, Hashmi R, Klaassen CD. Effects of microsomal enzyme inducers on thyroid-follicular cell proliferation, hyperplasia, and hypertrophy. *Toxicol Appl Pharmacol*. 1999;160:163–170.
172. Zoeller RT. Environmental chemicals as thyroid hormone analogues: new studies indicate that thyroid hormone receptors are targets of industrial chemicals? *Mol Cell Endocrinol*. 2005;242:10–15.

173. Paul-Friedman K, Martin M, Crofton KM, et al. Limited chemical structural diversity found to modulate thyroid hormone receptor in the Tox21 chemical library. *Environ Health Perspect*. 2019;127:97009.
174. Bansal R, Zoeller RT. Polychlorinated biphenyls (Aroclor 1254) do not uniformly produce agonist actions on thyroid hormone responses in the developing rat brain. *Endocrinology*. 2008;149:4001–4008.
175. Ren XM, Guo LH. Assessment of the binding of hydroxylated polybrominated diphenyl ethers to thyroid hormone transport proteins using a site-specific fluorescence probe. *Environ Sci Technol*. 2012;46:4633–4640.
176. Landrigan PJ, Goldman LR. Protecting children from pesticides and other toxic chemicals. *J Expo Sci Environ Epidemiol*. 2011;21:119–120.
177. Landrigan PJ, Goldman LR. Children's vulnerability to toxic chemicals: a challenge and opportunity to strengthen health and environmental policy. *Health Aff*. 2011;30:842–850.
178. Bennett D, Bellinger DC, Birnbaum LS, et al. Project TENDR: targeting environmental neuro-developmental risks the TENDR consensus statement. *Environ Health Perspect*. 2016;124:A118–A122.
179. Wing L, Potter D. The epidemiology of autistic spectrum disorders: is the prevalence rising? *Ment Retard Dev Disabil Res Rev*. 2002;8:151–161.
180. Blumberg SJ, Bramlett MD, Kogan MD, Schieve LA, Jones JR, Lu MC. Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011-2012. *Natl Health Stat Report*. 2013;65:1–12.
181. Wazana A, Bresnahan M, Kline J. The autism epidemic: fact or artifact? *J Am Acad Child Adolesc Psychiatry*. 2007;46:721–730.
182. Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. *Res Autism Spectr Disord*. 2011;5:418–425.
183. Iqbal K, Tran DA, Li AX, et al. Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming. *Genome Biol*. 2015;16:59.
184. Bergman A, Heindel JJ, Jobling S, Kidd KA, Zoeller RT. *State of the Science of Endocrine Disrupting Chemicals 2012*. Washington, DC: United National Environment Programme and World Health Organization; 2013.
185. Lancet. Electronic waste--time to take stock. *Lancet*. 2013;381:2223.
186. Di Renzo GC, Conry JA, Blake J, et al. International Federation of Gynecology and Obstetrics opinion on reproductive health impacts of exposure to toxic environmental chemicals. *Int J Gynaecol Obstet*. 2015;131:219–225.